





# ***SCIRE***

Representación y organización del conocimiento



# SCIRE

Representación y organización del conocimiento

Vol. 22, n.º 2, julio-diciembre 2016

ISSN 1135-3716

Scire:

knowledge representation and organization  
Vol. 22, n. 2, July-December 2016

Ibersid:

Red de Investigación  
en Sistemas de Información  
y Documentación

© 2016 Francisco Javier García Marco y los autores.

Reservados todos los derechos. Prohibida la reproducción por cualquier medio de la totalidad o parte de la presente edición sin permiso escrito de los titulares del copyright. Queda, sin embargo, autorizada expresamente la reproducción de los resúmenes y palabras clave de los artículos en inglés y español. También se permite la reproducción de sus textos con objetivos exclusivamente docentes para su uso ocasional en el aula, en cuyo caso se excluye el almacenamiento superior a un mes en un servidor.

Los artículos publicados están sometidos a evaluación científica y editorial en beneficio de la comunidad científica y los autores, pero dicha evaluación se realiza dentro de los límites del desarrollo actual de la disciplina, las limitaciones humanas de editores y revisores, y con una perspectiva ideológica, teórica y metodológica abierta, por lo que la dirección no siempre comparte las opiniones, métodos y procedimientos reflejados en la revista. Los autores se responsabilizan personalmente del cumplimiento de los códigos éticos y la legislación vigente, así como de la obtención de los permisos de copia de materiales ajenos.

ISSN: 1135-3716 = Scire (Zaragoza)

Depósito legal: Z. 1.790 — 1995

Edita: Ibersid® con la colaboración de Prensas Universitarias de Zaragoza

Imprime:

Servicio de Publicaciones. Universidad de Zaragoza.

Edificio de Ciencias Geológicas, C/ Pedro Cerbuna, 12.

50009 Zaragoza, España. Tel.: 976 761 330. Fax: 976 761 063.

---

# Scire

representación y organización  
del conocimiento

---

## Alcance y objetivos

---

*Scire: representación y Organización del Conocimiento* es una publicación semestral de carácter interdisciplinar sobre la representación, normalización, tratamiento, recuperación y comunicación de la información y el conocimiento.

## Difusión

---

Scire tiene difusión internacional. Agradecemos la inclusión en los siguientes servicios de referencia: Scopus, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC y Catálogo Latindex.

## Agradecimientos

---

Este número ha contado con una Ayuda para la edición de revistas científicas del Vicerrectorado de Investigación de la Universidad de Zaragoza, del año 2016.

## Redacción, distribución y canje

---

Revista Scire  
Departamento de Ciencias de la Documentación e Historia de la Ciencia  
Facultad de Filosofía y Letras  
Universidad de Zaragoza  
C/ Pedro Cerbuna 12,  
E-50.009 Zaragoza (Spain)

Tfno: int + 34 976 762239. Fax: 34 976761506.  
E-mail: <mailto:scire@unizar.es>

## Suscripciones y números sueltos

---

Suscripción anual / Annual subscription: 30 €.  
Número suelto: 20 €.  
(IVA incluido)

---

# Scire

knowledge organization  
and representation

---

## Scope and aims

---

*Scire: Representación y Organización del Conocimiento* is an interdisciplinary journal published twice a year that is devoted to the representation, standardization, treatment, retrieval and communication of information and knowledge.

## Dissemination

---

Scire has international distribution. We acknowledge its inclusion in the following reference services: Scopus, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC and Catálogo Latindex.

## Acknowledgments

---

This issue has received a grant for the edition of scientific journals of the Vicerectorate of Research of the University of Saragossa, Spain, for the year 2016.

## Contact address

---

Revista Scire  
Departamento de Ciencias de la Documentación e Historia de la Ciencia  
Facultad de Filosofía y Letras  
Universidad de Zaragoza  
C/ Pedro Cerbuna 12,  
E-50.009 Zaragoza (Spain)

Tel.: int + 34 976 762239. Fax: 34 976761506.  
E-mail: [scire@unizar.es](mailto:scire@unizar.es)

## Subscriptions

---

Annual subscription: 30 €.  
Issue: 20 €.  
(VAT included)

## Editor

Francisco Javier García Marco, Univ. de Zaragoza. E-mail: jgarcia@unizar.es

### Consejo de redacción / Editorial council

Mario Guido Barité Roqueta, Universidad de Montevideo, Uruguay	João Batista Ernesto Moraes, Universidade Estadual Paulista, Brasil	Francisco Javier Martínez Mendez, Universidad de Murcia, España
José Augusto Chaves Guimarães, Universidade Estadual Paulista, Brasil	Francisco Javier García Marco, Universidad de Zaragoza, España	Álvaro Quijano Solís, Colegio de México, México

### Consejo científico / Scientific council

Tomás Baiget, EPI S. A., España Mercedes Caridad Sebastián, Universidad Carlos III de Madrid, España	María Pilar Lasala Calleja, Universidad de Zaragoza, España Alfonso López Yepes, Universidad Complutense de Madrid, España José López Yepes, Universidad Complutense de Madrid, España Pedro Marijuán Fernández, Universidad de Zaragoza, España Bonifacio Martín del Brío, Universidad de Zaragoza, España José Antonio Moreira González, Universidad Carlos III de Madrid, España Purificación Moscoso Castro, Universidad de Alcalá, España Félix Moya Anegón, Universidad de Granada, España María del Carmen Negrete Gutiérrez, Universidad Autónoma de México	María Pinto Molina, Universidad de Granada, España Gloria Ponjuán Dante, Universidad de La Habana, Cuba José Vicente Rodríguez Muñoz, Universidad de Murcia, España Adelaida Román Román, CINDOC (Madrid), España Juan Ros García, Universidad de Murcia, España Francisco José Ruiz de Mendoza Ibáñez, Universidad de La Rioja, España Félix Sagredo Fernández, Universidad Complutense de Madrid, España Elías Sanz Casado, Universidad Carlos III de Madrid, España Carlos Serrano Cinca, Universidad de Zaragoza, España Emilio Setién Quesada, Asociación Cubana de Bibliotecarios, Cuba †Félix del Valle Gastaminza, Universidad Complutense de Madrid, España
Alberto Carreras Gargallo, Universidad de Zaragoza, España		
Emilia Currás Puente, Universidad Autónoma de Madrid, España		
Constança Espelt Busquets, Universidad de Barcelona, España		
María Eulalia Fuentes Pujol, Universidad Autónoma de Barcelona, España		
Fernando Galindo Ayuda, Universidad de Zaragoza, España		
Blanca Gil Urdiciain, Universidad Complutense de Madrid, España		
Alan Gilchrist, Cura Consortium, Reino Unido	José Luis Otal, Universidad Jaume I de Castellón, España	
José María Izquierdo Arroyo, Universidad de Murcia, España	Manuel José Pedraza Gracia, Universidad de Zaragoza, España	

†Dedicamos un recuerdo especial para nuestro querido compañero y amigo, Félix del Valle Gastaminza, que falleció durante la edición de este número. Nuestro más sincero duelo y homenaje.

### Revisores externos del número / External reviewers in this issue

Agradecemos enormemente la colaboración generosa y altruista de María Cláudia Cabrini Gracio, Carlos Cândido de Almeida, Antonio García Jiménez, Cristina Herrero Pascual, Mónica Izquierdo Alonso, Joaquín López del Ramo, Daniel Martínez-Ávila, Mercedes Martínez González, Montserrat Mateos Sánchez, Luis Miguel Moreno Fernández, Javier Noguera Iso, María Olivera Zaldua, Juan Pastor Sánchez, Pepita Raventós, Blanca Rodríguez Bravo, Rodrigo de Sales, Montserrat Salas Valero, José Antonio Salvador Oliván, Juan Miguel Sánchez Vigil, Enric Saperas Lapiedra y Ángel Zazo Rodríguez.

### Candidaturas al consejo científico

Se aceptan candidaturas al consejo científico de especialistas del área de Biblioteconomía y Documentación y de otras disciplinas relacionadas (Informática, Ciencias Sociales, Lingüística, Filosofía, Psicología, etc.) con experiencia profesional e investigadora demostrada. En el sistema público de investigación español, suele ser equivalente al doctorado y dos sexenios de investigación o méritos equivalentes.

### Scientific council membership policy

Candidatures of researchers from LIS and other related disciplines (Computer Science, Social Sciences, Linguistics, Philosophy, Psychology, etc.) with demonstrated professional and research experience are welcomed. In the Spanish public research system, for example, this usually means having a doctorate and two scientific productivity sexennia or equivalent outputs.



---

# Tabla de contenidos en español

Table of contents in Spanish

---

Tabla de contenidos en español .....	9	Joaquín LÓPEZ DEL RAMO, María Luisa HUMANES .....	87
Tabla de contenidos en inglés.....	11		
<b>Artículos</b>		<b>Estudios de caso, investigaciones en curso, artículos breves y comunicaciones</b>	
<i>Información sobre la información</i>		<i>Memoria y representación: reflexiones para la organización del conocimiento</i>	
ALAN GILCHRIST .....	13	Mona Cleide Quirino da Silva FARIAS, María Leandra BIZELLO .....	99
<i>Organización del conocimiento para la documentación en periodismo: situación y prospectiva</i>		<i>Vocabulario controlado para archivos universitarios: la experiencia del VTArq de UNESP (Brasil)</i>	
Antonio GARCÍA JIMÉNEZ.....	21	Walter MOREIRA, Mariângela Spotti Lopes FUJITA, Luciana Beatriz Piovezan dos SANTOS .....	107
<i>Integración de anotaciones espaciales automáticas de diferentes fuentes mediante tecnologías semánticas</i>			
Javier NOGUERAS-ISO, Verónica LÁZARO, Ludovic MONCLA .....	29		
<i>La estructura lógica y jerárquica de lenguajes de indización utilizados por las bibliotecas universitarias</i>		<i>Procedimiento de evaluación .....</i>	113
Mariângela Spotti Lopes FUJITA, Luciana Beatriz Piovezan dos SANTOS .....	37	<i>Instrucciones para la presentación de trabajos.....</i>	115
<i>Estrategias para mejorar la visibilidad y difusión de la revista Fotocinema: creación de un vocabulario controlado multilingüe</i>		<i>Índice de autores.....</i>	118
Rocío PALOMARES-PERRAUT, Carmen GÓMEZ CAMARERO.....	47	<i>Índice de materias en español .....</i>	118
<i>Análisis de las teorías semióticas en la ciencia de la información brasileña: referentes teóricos.....</i>		<i>Índice de materias en inglés.....</i>	119
Carlos Cândido de ALMEIDA, Mona Cleide Quirino da Silva FARIAS.....	57		
<i>Análisis de temas emergentes a través de Twitter</i>			
José Luis ALONSO BERROCAL, Carlos G. FIGUEROLA, Ángel F. ZAZO RODRÍGUEZ .....	67		
<i>Análisis del discurso y ciencia de la información: aportes teóricos para la organización y representación de la información</i>			
João Batista Ernesto MORAES, Larissa Mello LIMA, Mariana Silva CAPRIOLI .....	75		
<i>Análisis de contenido de la representación fotográfica de la crisis de los refugiados sirios y su incidencia en el framing visual</i>			



---

# Tabla de contenidos en inglés

Table of contents in English

---

Table of contents in Spanish .....9

Table of contents in English .....11

## Articles

Information about information

ALAN GILCHRIST .....13

*Knowledge organization  
for journalism documentation:  
state of the art and prospective* .....

Antonio GARCÍA JIMÉNEZ.....21

Integration of automatic spatial  
annotations from different sources  
by means of semantic technologies

Javier NOGUERAS-ISO,  
Verónica LÁZARO,  
Ludovic MONCLA .....29

*The logical and hierarchical structure  
of indexing languages used  
by university libraries*.....

Mariângela Spotti Lopes FUJITA,  
Luciana Beatriz Piovezan dos SANTOS .....37

*Strategies to improve the visibility  
and dissemination of the journal  
Fotocinema: creation of a multilingual  
controlled vocabulary* .....

Rocío PALOMARES-PERRAUT,  
Carmen GÓMEZ CAMARERO.....47

*Analysis of semiotic theories  
in Brazilian information science:  
theoretical referents*

Carlos Cândido de ALMEIDA,  
Mona Cleide Quirino da Silva FARIAS.....57

*Analysis of emerging subjects in Twitter*

José Luis ALONSO BERROCAL,  
Carlos G. FIGUEROLA,  
Ángel F. ZAZO RODRÍGUEZ .....67

*Discourse analysis and Information  
Science: theoretical contributions  
to information organization  
and representation*

João Batista Ernesto MORAES, Larissa Mello  
LIMA, Mariana Silva CAPRIOLI .....75

*Content analysis of the photographic  
representations of the Syrian refugee  
crisis and their visual framing impact*

Joaquín LÓPEZ DEL RAMO,  
María Luisa HUMANES .....87

## Research in advance, brief articles, communications and notes

*Memory and representation: reflections  
for Knowledge Organization*

Mona Cleide Quirino da Silva FARIAS,  
María Leandra BIZELLO .....99

*An archival controlled vocabulary  
for university archives: the experience  
of VTArq (UNESP, Brazil)*

Walter MOREIRA,  
Mariângela Spotti Lopes FUJITA,  
Luciana Beatriz Piovezan dos SANTOS .....107

Evaluation process .....113

Instructions for authors .....115

Author index .....118

Subject index in Spanish .....118

Subject index in English .....119



# Information about information

---

ALAN GILCHRIST

Cura Consortium, Reino Unido, [alangilchrist77@gmail.com](mailto:alangilchrist77@gmail.com)

## Resumen

Se sostiene que, aunque el panorama de la información ha cambiado radicalmente, los principios básicos de la ciencia de la información siguen siendo válidos, y que a menudo inadvertidamente siguen aplicándose, por lo que podrían volver a ser de nuevo más ampliamente reconocidos. Este nuevo escenario se ilustra brevemente con algunos ejemplos seleccionados de su importancia, peligros e historias de éxito actuales. Están empezando a surgir nuevos modelos y enfoques, de los cuales se presentan algunos ejemplos. Es vital que sean alentados y apoyados para que la sociedad en su conjunto y, en particular, sus ciudadanos puedan estar mejor informados.

**Palabras clave:** Ciencia de la información. Organización del conocimiento. Metadatos. Teoría. Perspectivas. Casos de éxito.

## 1. Introduction

The title of this paper is taken from a remark made by Robert Fairthorne many years ago. An Information Scientist, but originally a mathematician, Fairthorne, well known for his commonsensical remarks, once said “Information Scientists do not give information, they give information about information”. While this was true at the time that he said it, the advance of information technology, the interconnectedness of communications technology and the resultant increase of disintermediation have radically changed this basic aspect of information science in practice. This paper argues, however, that basic principles described in the literature of information science and repeated as recently as 2004 are still valid and might be redeployed, though with obvious difficulty, in the new electronic environment. This difficulty should be faced and understood, and tackled with a concerted effort by information scientists and allies in other disciplines.

## 2. Traditional Information Science practice

Expanding the pithy comment of Fairthorne's it may be seen that the task of the information scientist mediating between recipient and source involved:

## Abstract

This paper argues that though the information landscape has radically changed, the basic tenets of information science are still valid, are often unwittingly being applied and may again become more widely recognized. This new landscape is briefly illustrated with some selected examples of its size, dangers and current success stories. New models and approaches are beginning to emerge and some examples of these are presented. It is vital that these are encouraged and supported so that society and its citizens may become better informed.

**Keywords:** Information science. Knowledge organization. Metadata. Theory. Perspectives. Success stories.

1. Understanding the request for information, including the want behind the request, and perhaps the use to which the recipient hopes it may be put.
2. Gauging the likelihood of the ability of the enquirer to assimilate any information contained in the item(s) retrieved. (It would be futile to answer a schoolchild's request for information on the solar system with an academic treatise by the Astronomer Royal of the United Kingdom).
3. Establishing an effective contact with a potential physical or personal source of information.
4. Helping in the formulation of an enquiry.
5. Assessing the relevance of the answers and an informed view of the reliability of the source.

These exchanges between enquirer and mediator and between mediator and source (face to face or via an electronic link) require interpretation, or what might more grandly be called hermeneutics. In the words of Zimmermann (2015) “The goal of interpretation is to make sense of a text or situation, to understand what they mean”, and if that seems obvious, it should be remembered that failures in interpretation are experienced every day, either knowingly or not. Given the five

potentially complex steps above, effective interpretations throughout the sequence may not be straightforward and may require a high degree of knowledge and empathy between the mediator and enquirer.

Zimmermann goes further in his discussion of hermeneutics when he states that “the goal of hermeneutics is understanding, and that although understanding may be guided by analytical principles, it cannot be reduced to them”, and that “Understanding requires art”, and “Understanding is knowledge in the deeper sense of grasping not just facts but their integration into a meaningful whole”. In the case of the mediation as described above this integration is formed through discourse using language to arrive at a shared understanding of the actions to be performed from start to finish. To emphasise this point one must remember that language is complex and that understanding between two people must often depend on reiteration and the use of alternative and explanatory expressions. By replacing a human mediator with a machine we have lost much of this “knowledge in the deeper sense”, and the searcher for information often works alone.

### 3. The basics of Information Science

Though the scope of Information Science may have expanded (in both theory and practice) it is

worth going back to the work of Brian Vickery and Alina Vickery, notably in their seminal book *Information Science in Theory and Practice* (12). In the final chapter of their book, echoing the hermeneutic principle outlined above, they suggest that:

[...] information science should seek to increase our understanding of:

- (1) The behaviour of people as generators, sources, recipients, and users of information, and as channel agents;
- (2) The quantitative study of the population of messages – its size, growth rate, distribution, patterns of production and use;
- (3) The semantic organization of messages and of channels that facilitates their identification by sources and recipients;
- (4) Problems particularly associated with the functions of information storage, analysis and retrieval;
- (5) The overall organization of information systems and their performance in transfer;
- (6) The social context of information transfer, in particular its economics and politics.

This quotation has been reproduced in full because, in every respect, the constituent parts remain valid within the context of the contemporary world of electronic interconnection and social media, though will have to be modified in practice.

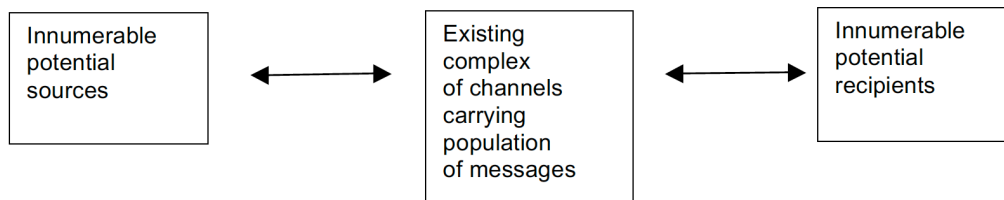


Figure 1 (adapted from Vickery and Vickery, 2004)

### 4. Size and speed

A small Figure from the book by the Vickerys (204) underlined the enormous task faced by information scientists when their book was published. That task facing information workers (and, it must be said, society as a whole) has now increased dramatically:

What is startling about this Figure are the currently enormous numbers reflected in each of the three boxes. It is almost meaningless to quote large numbers concerning the Internet, numbers that are increasing by the minute, but a short survey of some figures proposed by reputable

sources taken at random from the Web may nevertheless serve as a reminder.

1. There were roughly one billion personal computers in the world at March 2014; and almost 48 million have been sold so far this year (at 15 March 2016).
2. There are now more active mobile devices in the world than people, some 7.2 billion gadgets, (though half of the world’s population do not own one).
3. Another source claims there are some 2 billion smart-phones; and these each have the power of a supercomputer of 30 years ago.

4. In September 2014 a milestone was reached when 1 billion websites were recorded.
5. There may, as of 2013, have been some 152 million blogs, though many are abandoned by the user after a few weeks. More meaningfully, 1.13 million postings were recorded on the single day of 21 March 2015.

And searching the Internet produces big numbers too, not always with completely satisfactory results. For example, two searches on Google produced these results:

- Astrobiology: about 529,000 (hits) in 0.55 seconds
- Learning to play the ocarina: about 388,000 (hits) in 0.58 seconds.

The first of these searches produced, not surprisingly, a mixed bag of references on the first page, while the second (again, not surprisingly) produced a majority of relevant items. As an exhibition of enormous computing power and clever algorithms these figures are hugely impressive, but who looks beyond the first few pages to check for the possibility of interesting items, and would it matter if the search took one or two seconds? These may be trivial observations of detail, but they do raise the question of what value these figures have, other than to boast of the power of Google. More importantly, perhaps, they tend to lull the searcher into what might be a feeling of satisfaction that may be unwarranted, or a hurried 'that will do' reaction.

Admittedly, these observations are rather unfair to Google, considering the way in which it has made the Web accessible, but (in passing) it may be added that other search engines are not so impressive: one specialist blog search engine produced a first page of irrelevant hits in answer to the query 'Learning to play the ocarina' all being concerned with online video games – two involving 'Queen Ocarina'.

Another aspect of size is seen in the power and financial might of the GAFA quartet of Google, Apple, Facebook and Amazon, a power that is changing the landscape of large areas of information provision particularly in that of print on paper, affecting the newspaper sector and publishers and sellers of books. The size of the GAFA giants has also, perhaps, inhibited the growth of more specialised services though, as will be shown later in this paper, there are a number of interesting initiatives beginning to appear in the area of information processing and provision.

## 5. Dangers of size, accessibility, speed and time

Now that anybody with access to the Internet can be a 'publisher', there are new dangers of the abuse of free speech, most notably to be found in trolling, the dissemination of misinformation including that based on unfounded opinions and worse, intentional disinformation. Nigel Warburton (2009) (quoting Richard Posner) has noted four dangers relating to "irresponsible speech" on the Internet:

- *Anonymity*: making it easy to hide behind false, or even illegal, information.
- *Lack of quality control*: allowing the spread of inaccurate and misleading information and the publication of often libellous gripes and rumours concerning celebrities.
- *Huge potential audience*: tending to magnify any harm caused by false information.
- *Antisocial people find their soul mates*: possibly leading to antisocial acts by individuals emboldened by their perceived support.

Children are particularly vulnerable and a reputable newspaper in the U.K. has reported that "Up to three quarters of primary school children (i.e. ages 5 to 11) have online connections with people they don't know in the real world".

Even the amount of information churned out by the mass media is indigestible, particularly in the way in which it is used by the media to sell newspapers, and by politicians and others using the media, often guilty of spreading misinformation and even occasionally disinformation. The famous philosopher and author Umberto Eco (2002), in discussing the interplay between politicians and the newspapers, has observed

[...] with statements following one another day after day...the result is that the reader loses count and forgets what has been said ...the price, both for the press and the politician, is unreliability and a 'who cares?' reaction from the man in the street".

These short-comings contribute to widespread assumptions of false information, often about emotive issues. For example, it has been reported in the British newspaper *The Guardian* that "Britons think that £24 out of every £100 spent (by the Government) on (social) benefits is claimed fraudulently when the official estimate is 70p". This sort of misapprehension is not confined to the U.K. The current anti-immigration sentiment leads "Britons and Spaniards to think they have twice as many immigrants in their country as they actually do, the Italians, Belgians and French assume there are three times as

many as there are, and the Poles more than thirty times". The same newspaper, in a short Editorial, draws attention to cyberwarfare and the deliberate intra-national dissemination of misinformation. In one instance the U.S Government discovered that "it had fallen victim to a hack attack on an unprecedented scale, with the security of the details of up to four million former and present employees apparently breached". The same newspaper discovered a building in St Petersburg staffed by paid bloggers to establish inoffensive online personalities in the comment sections of media outlets with the aim of seeding such outlets with pro-Putin or pro-Government remarks.

An article in the *New Scientist* by Baraniuk (2016) is also disturbing. It starts with a report of research undertaken on how the social web works, specifically on how different types of information are spread on Facebook by two different communities – one that shared science news articles and another that shared conspiracy theories. The research found that "science stories received an initial spike of interest and were shared frequently" while "conspiracy theories started off with a low level of interest but sometimes grew to be even more popular than the science stories overall". More importantly, showing how information bias can lead to an "echo chamber effect", it was found that in a study of 55 million Facebook users, "out of 50,000 posts debunking rumours, only 1 in 12 reached people who had shared the rumour in question". Baraniuk (2016), the author of this article puts forward five ways in which research has shown how the Internet twists the truth:

*"Majority illusion*, where groups within social networks tend to be influenced most by a small number of popular individuals.

- *List ordering*, the use of search engine optimisation to ensure that certain hits appear high up in the list of sites retrieved, which results in a feeling in the searcher that sites high on the list are more credible.
- *Astroturfing*, the invention of a wave of support or dissent manufactured by advertisers, political parties, or even governments.
- *BOT democracy*, whereby an army of Bots (web robots – software applications that run automated tasks (scripts)) is deployed to give an impression of, for example, a grass-roots policy (as was admitted by a Parliamentary candidate in London in 2012).
- *Disinformation*, deliberately planted false information. For example in 2014 Cynk Technology briefly became worth \$6 billion, after a barrage of tweets and emails promoting the stock flooded the Web. It turned out that the

firm had no assets, no revenue and one employee.

Swamped with potential information and the pressure of time, and seduced by the advertising of the ubiquitous 'click' many users of IT suffer further problems. Research commissioned by KasperskyLab (2015) in a report under the title *The Rise and Impact of Digital Amnesia. Why we Need to Protect what we no Longer Remember* concluded that there was "a direct link between data available and a failure to commit that data to memory", defining Digital Amnesia as "Experience of forgetting information that you trust a digital device to store for you". Another piece of academic research, this one of the millennial generation by Taylor (2012) observes that "Millennial generation Web searchers proceed erratically through an information search process, make only a limited attempt to evaluate the quality or validity of information gathered, and may perform some level of 'backfilling' or adding sources to a research project before final submission". Clearly, the lack of any expert intermediation is to be regretted, though it is difficult to see how it might be installed.

Finally, in this section concerning some of the worst aspects of current problems in the world of information, it must be noted that even 'big business' is failing to address information problems adequately. After years of reports repeating that while senior management claimed to regard information as vital to their business, their own knowledge workers were dissatisfied with the level and quality of information provision. A report from Gartner, noted by Allan Foster (2014), predicts in a research study that "By 2017, 33% of Fortune's organizations will experience an information crisis due to their inability to effectively value, govern and trust their enterprise information".

## 6. All is not lost!

After the first waves of uncertainty, curiosity, acceptance and delight in the new, it was probably inevitable that such a powerful device as the Internet would attract misuse and carry social problems. The previous section has indicated some of these problems, albeit in a piece-meal and subjective way, but it remains a fact that such problems do exist as individuals, groups and governments strive to sideline or ban the worst features and concentrate on making better sense of the whole. This section briefly reviews some of the important advances as well as some of the new developments.



## 6.1. Current success stories

Despite the sorry state of 'enterprise search' as mentioned above, large databases continue to thrive and be interconnected to others. Medline, Agrovoc, Eurovoc and the shared database supplied by CERN for scientists worldwide working in the field of particle physics are some notable examples. New initiatives are springing up in the area of information provision with wide ambitions, for example the Europeana Project, whose beta-test website (Surowiecky, 2004) invites users "...to explore 52,557,036 artworks, artefacts, books, videos and sounds from across Europe" with growing multilingual access. Europeana is just one example of systems using Linked Open Data (LOD) largely based on a complex array of standards laid down by W3C, the World Wide Web Consortium.

These standards are hugely important, indeed necessary. For example, the basic device for describing a resource known as the RDF (Resource Description Framework) is in the form of what is known as a 'triple', This is a statement based on 'subject-predicate-object', for example 'Michelle – is the wife of – Barack Obama', and each of these three entities must then be given a URI (Universal Resource Indicator). It is clear that the diversity of possible meanings must also be dealt with so that, for example, two URIs containing the word 'Berlin' must be processed to distinguish between 'Berlin', the city and 'Berlin' the songwriter. This is, obviously, detailed and painstaking work.

In addition to LOD projects such as Europeana, other projects are building what might be termed 'hubs'. In technical network terms a hub is 'a common connection point', but this could be extended to include connection points carrying openly available information. One such is the DBpedia, a huge scheme which, in earlier times might have been described as an example of secondary publishing. The DBpedia is described by Lehmann *et al* (2012) as "a knowledge base that extracts structured, multilingual knowledge from Wikipedia and makes it available using Semantic Web and Linked Open Data technologies [...] the largest DBpedia knowledge base, extracted from the English edition of Wikipedia consists of 400 million facts that describe 3.7 million things". Expanding this idea of hubs, Javier Garcia Marco (2016) has called for better devices allowing browsing on the Web through "a semantic network of KOS", with a particular idea of creating such a network around a large facility such as Wikipedia.

## 6.2. Radical movements

With so much advertising and trivia on the surface web it is heartening to see a range of public facing and interactive websites being built by governments and some independent fact checking agencies. However, while this is to be welcomed, the range of sources, quantity and complexity of information can be daunting. For example:

- [www.gov.uk](http://www.gov.uk) is the official UK Government website, combining 24 Ministerial Departments and 331 other agencies and public bodies.
- [www.data.gov.uk](http://www.data.gov.uk) is the official UK Government website containing data sets released under the Government's Transparency Agenda.

But, perhaps somewhat confusingly,

- <https://yougov.uk> is an independent global market research and data company which has established "a community of 4 million people round the world who share their views".

At the time of writing this paper the campaign for the UK referendum on whether to remain in or leave the European Union was in full swing. Though the official Government policy favoured Remain, many Government members and even some Government Ministers favoured Leave. All the political parties were issuing their arguments on their own websites in addition to the websites of the two official campaign organizations 'Britain Stronger in Europe' and 'Vote Leave' (quoting their own versions of 'facts'). Thankfully, there is an independent and independently funded website (<https://fullfact.org>) dedicated to checking 'facts' disseminated on official websites, including all of the websites listed above, currently then with particular reference to the EU referendum debate. There is also a similar American website at ([www.snopes.com](http://www.snopes.com)) that in its own words "debunks urban legends, old wives' tales, fake news, shoddy journalism and political spin". Add to the above examples the plethora of facts, misinformation, disinformation and opinions disseminated on other websites and in the media and it becomes clear that the process of arriving at an objective and informed personal opinion is a daunting task.

In addition to outward looking websites, a number of central and local governments are now operating interactive sites. For example, Madrid (<https://decide.madrid.es>) under the banner "La ciudad que quieres será la ciudad que quieras" elicits responses to plans to remodel the Plaza Espana; while Paris (<https://idee.paris.fr>) with the headline "Madame la Maire j'ai une idée" invites ideas from its citizens.

This attempt to gather comments and opinions has been related to the famous story of Francis Galton (cousin of Charles Darwin) and the country cattle fair where farmers were asked to guess, for a prize, the weight of a displayed ox, butchered and dressed. Galton took the tickets after the event and discovered that the average guess of all the entrants was remarkably close to the actual weight – in fact, one pound weight for an ox that weighed 1,198 pounds. This collective guess was not only better than the actual winner but also better than the guesses made by cattle experts at the fair. This discovery became known as “The wisdom of crowds” and lies behind some of the psychology in a number of websites being developed in a number of countries.

However, attractive as the idea of the wisdom of crowds is, tapping into it must be carefully planned. As the French writer Anatole France said “If fifty million people say a foolish thing, it is still a foolish thing”, and there are more subtle problems to be avoided. For example, a piece of research carried out at ETH in Switzerland (14) asked two groups to suggest the answers to two questions: ‘the length of the Swiss-Italian border’ and ‘the annual number of murders in Switzerland’. The research, not surprisingly, showed that as the amount of information that participants were given about each other’s guesses increased, the range of guesses became narrower; tending towards a consensus to the detriment of accuracy. James Surowiecky has studied such phenomena more widely in his book *The Wisdom of Crowds* (Ball, 2014) in which he proposes “four key qualities that make a crowd smart”. These are:

1. It needs to be diverse, so that people are bringing different pieces of information to the table.
2. It needs to be decentralized, so that no one at the top is dictating the crowd’s answer.
3. It needs a way of summarizing peoples’ opinions into one collective verdict.
4. And the people in the crowd need to be independent, so that they can pay attention mostly to their own information, and not worry about what everyone around them thinks.

One particularly ambitious project is being carried out by GovLab at New York University following encouragement from Barack Obama. This project has carried out an extensive review of other projects such as the ‘Decide.Madrid’ and ‘Idée.Paris’ mentioned above and is now researching new approaches and championing the activities of other agencies, as explained by Beth Simone Noveck in her book ‘Smarter Citizens, Smarter State’ (Noveck, 2015). Noveck wants “citizens with practical experience to sign up to a networked

searchable ‘Brains Trust’”, with accompanying advanced technology for its exploitation. The key to this ambition is to capture details on not only the qualifications, but also the experience of civil servants and to match these with similar databases of citizens. There are examples of the former, established by such agencies as the Department of Defense, but few examples of the latter, one being a New York database of citizens with language skills. Though Noveck’s intentions are no doubt honourable, there are already some sceptics, some being nervous of surveillance, others unconvinced that such an enormous network will actually be capable of producing viable high-level policy decisions.

## 7. Some conclusions

### 7.1. Image and language

Sargon of Akkad, who reigned around 2250 BCE, has been called ‘The Father of Libraries’ and there is certainly much archaeological evidence of large and well-organized libraries in ancient Mesopotamia, particularly at Nineveh. The Assyrian word for librarian is translated as ‘keeper of the tablets’, and this image has persisted through the ages as the keepers of scrolls, codices and books, though librarianship has since embraced many other skills. Nevertheless, in the public mind libraries are usually large, often magnificent buildings full of shelves full of books. This image persists from Alexandria to the huge multilingual collections of 11<sup>th</sup> Century Baghdad and Cordoba and through to the Vatican and today’s National Libraries, not forgetting the University and Public Libraries - all highly visible and recognizable for what they are. The names of famous librarians have also persisted – Callimachus of Alexandria who devised a classified catalogue for the collection; Archibald Macleish, Chief Librarian of the Library of Congress and advisor to Franklin D. Roosevelt, and Jorge Luis Borges, Director of the National Library of Argentina – these last two also being internationally renowned writers and poets. So, the image of the imposing buildings are further enhanced by the renown of their supervisors.

Some 4,200 years after Sargon, the term ‘Information Scientist’ was first coined, later extended to ‘Information Science’. These were not terms that became widely recognized and, indeed, the reaction in conversations that mentioned the terms soon became “Oh. you mean computers”. In addition, in the public mind (except for the users in scientific and industrial establishments) there was no great physical evidence of their activity; and often in an industrial situation, the ‘information department’ was housed with the Research Department often separated on a different

site. Today, it has to be admitted that any image that there might have been has further faded. But A. J. Meadows had this to say in a collection of essays published in 2009 (2009):

Both computer studies and business studies are much larger than information science in terms of number of researchers and financial support. Consequently, this merging of interests is also, to some extent, a submerging as can be seen in some of the departmental reorganizations at universities in the UK... The overall result is that the information science activities developed over the last 50 years have triumphed, but information science as a separate entity may be on the wane. If so, its final epitaph may well be that of (the architect) Sir Christopher Wren in St Paul's Cathedral: 'If you want a monument, look around you'.

## 7.2. A resurgence?

In 2002 in the UK the IIS (Institute of Information Scientists) was merged into the (LA) Library Association to form CILIP, the Chartered Institute of Librarians and Information Professionals. The former influence of the IIS was thus diminished, information science being included in the poorly defined term 'Information Professional'. By coincidence, in the same year, the older International Federation of Information and Documentation (FID) was disbanded. Now, in the UK few groups appear to be carrying on the ethos of the IIS except perhaps for the UK Chapter of the International Society for Knowledge Organization which, despite the somewhat grandiose title and sometimes abstract preoccupations of the parent body can at least boast, like librarians, of having tangible features: classification schemes, thesauri and other retrieval languages. It can also be observed that, in promoting the application of Knowledge Organization, the ISKO UK Chapter is approaching the original scope of information science.

So, while Meadows was essentially correct in his analysis reproduced above, there are some signs that the principles of information science are being (in a sense) recognized and rediscovered mainly, perhaps, in the world of librarianship. Sir Nigel Shadbolt (co-founder with Sir Tim Berners-Lee of the Open Data Institute in the UK) has said, addressing CILIP at its annual conference: "As organizations increasingly recognize the value of open data, so will professional opportunities increase and information workers will be well-placed to take advantage. There is already a lack of skilled people to fill the roles and this will only get worse." There are corresponding signs that librarians are making real efforts to embrace information technology. One American website article (Barclay, 2016) claims, under the title "Has the library outlived its usefulness in the age of the Internet" that, despite severe cuts in funding, the

number of public libraries has increased and that academic libraries are sharpening their services; for example "Grand Valley State University's 'Knowledge Market' provides students with peer consultation services for research, writing, public speaking, graphic design and analysing quantitative data". In a similar, but even more radical development in the UK (Allen, 2016),

Europe's first integrated public and university library is housed in the modern Hive building in Worcester alongside the local archive and archaeology service and the County Council's services hub [...] the library's front line team is completely integrated, with some staff being employed by the University and some by the County Council.

A key feature of this initiative is to engage students in new ways, involving them in shaping library services of the future, and already they are producing guides to library use that differ from those produced by the staff.

These two bottom-up initiatives should be strongly supported by the professional associations which should also collaborate more closely with each other, adapting their own world-views and adopting a more flexible, less conservative approach to the demands of this rapidly changing information environment. The dream would be to have a single over-arching national professional association covering all the closely related information activities, including at least librarianship, information science, records management, archives management and museums documentation, as well as aspects of web-page design and, of course, LOD. There is an interesting potential model for this range of applications and techniques, one provided for self-employed information workers – the Association of Independent Information Professionals (AIIP) which states "What IIPs have in common is the ability to provide timely, accurate, and actionable information, analysis, design, review, and strategy to a diverse set of clients across a wide range of sectors" (Ojala, 2016).

Though, as Meadows said, information science may no longer be a separate entity, its basic tenets are still valid - even in the new and still changing information environment. To a large extent, those tenets are also valid for the work of the other information professionals mentioned above; and for many others who find themselves working as authors, intermediaries and searchers in the information chain. While, for example, academics are well served it is vital that new initiatives based on these tenets are nourished so that a better informed society may emerge, and it is encouraging to see such initiatives as have been discussed above such as Full Fact and Snopes. Increasingly, the individual – as employee, voter,

family member, and member of different economic, cultural and ethnic groups - derives much of his or her information directly or indirectly from the mass media of which the Internet is increasingly dominant. The resultant potential overload of information, much of it poorly accessed and assessed could, with the right efforts, perhaps be improved for the benefit of all.

## References

- Allen, Katherine (2016). Students improving services for themselves. // *Information Today Europe*. 13 May 2016. [www.infotoday.eu/Articles/Editorial/Featured-Articles/Students-improving-services-for-themselves-111071.aspx](http://www.infotoday.eu/Articles/Editorial/Featured-Articles/Students-improving-services-for-themselves-111071.aspx) (last accessed 18 May 2016)
- Ball, Philip (2014). Wisdom of the crowd: the myths and the realities. // *BBC Future*. 8 July 2014. [www.bbc.com/future/story/20140708-when-crowd-wisdom-goes-wrong](http://www.bbc.com/future/story/20140708-when-crowd-wisdom-goes-wrong) (last accessed 12.5.16)
- Baraniuk, Chris (2016). World Wide Warp. // *New Scientist*. 20 (February 2016) 38-41.
- Barclay, Donald A. (2016). Has the library outlived its usefulness in the age of Internet? You'd be surprised. // *The conversation*. April 28, 2016. <https://theconversation.com/has-the-library-outlived-its-usefulness-in-the-age-of-the-internet-you-d-be-surprised-58198>. (2016-05-18)
- Eco, Umberto (2002). *On the Press: Five Moral Pieces*. London: Vintage.
- Europea. [www.europeana.eu/portal](http://www.europeana.eu/portal) (last accessed 20.4.16)
- Foster, Allan (2014). Initiatives. // *Business Information Review*. 31:2 (June 2014) 122-139.
- Garcia-Marco, Francisco-Javier (2016). Enhancing the visibility and relevance of thesauri in the web: searching for a 'hub' in the linked data environment. // *Knowledge Organization*. 43:3 (2016) 193-202.
- Kaspersky Lab (2015). The Rise and Impact of Digital Amnesia. <https://blog.kaspersky.com/files/2015/06/005-Kaspersky-Digital-Amnesia-19.6.15.pdf> (2016-04-12)
- Lehmann, Jens et al. (2012). DBpedia: a large-scale, multilingual knowledge base extracted from Wikipedia. // *Semantic Web*. 1 (2012) 1-5.
- Meadows, A.J. (2009). Fifty years of UK research in information science. // Gilchrist, A. (Editor). *Information Science in Transition*. London: Facet Publishing.
- Noveck, Beth Simone (2015). *Smarter Citizens, Smarter State*. Harvard: Harvard University Press.
- Ojala, Marydee (2016). AIIP explores partnering for success. // *Information Today Europe*. 3 May 2016. [www.infotoday.eu/Articles/Editorial/Featured-Articles/AIIP-explores-partnering-for-success-110803.aspx](http://www.infotoday.eu/Articles/Editorial/Featured-Articles/AIIP-explores-partnering-for-success-110803.aspx). (last accessed 18 May 2016).
- Surowiecky, James (2004). *The Wisdom of Crowds*. New York: Doubleday.
- Taylor, Arthur (2012). A study of the information search behaviour of the millennial generation. // *Information Research*. 17:1 (March 2012).
- Vickery, Brian C.; Alina Vickery (2004). *Information Science in Theory and Practice*. Third revised and enlarged edition. München: K.G.Saur.
- Warburton, Nigel (2009). *Free Speech*. Oxford: Oxford University Press.
- Zimmermann, Jens (2015). *Hermeneutics*. Oxford: Oxford University Press.

---

Enviado: 2016-06-23. Segunda versión: 2016-08-17.  
Aceptado: 2016-09-01.

---

---

# Organización del conocimiento para la documentación en periodismo: situación y prospectiva

*Knowledge organization for journalism documentation: state of the art and prospective*

---

**Antonio GARCÍA JIMÉNEZ**

Facultad Ciencias de la Comunicación, Universidad Rey Juan Carlos, C/ Camino del Molino, s/n, Campus de Fuenlabrada, 28943 Fuenlabrada, Madrid, España; antonio.garcia@urjc.es

## Resumen

Se pretende establecer el estado actual, por un lado, de instrumentos como tesauros, ontologías y folksonomías y, por otro, de los procedimientos y modelos de formalización para el análisis y etiquetado semántico, al objeto de determinar los aspectos más relevantes en su aplicación en el dominio periodístico. Tras esta revisión, y tomando en consideración los cambios que se vienen produciendo en el periodismo, se plantean los puntos sobre los que debe girar futuras reflexiones sobre la organización del conocimiento en conexión con la web semántica, los datos abiertos, las necesidades de información y los nuevos modelos de recuperación de información.

**Palabras clave:** Organización del conocimiento. Tesauros. Ontologías. Folksonomías. Marcado semántico. Documentación. Periodismo.

## 1. Introducción

El objetivo de este trabajo es doble. Por un lado, precisar la realidad actual de las técnicas e instrumentos que afectan a la organización y representación del conocimiento, en especial en lo que tiene que ver como los tesauros, las folksonomías y los modelos de análisis y marcado semántico. Por otro lado, delimitar los aspectos concretos vinculados con los vocabularios y los procesos de análisis semántico en los que hay profundizar para poder evaluar la realidad de la organización del conocimiento para la Documentación periodística, partiendo de la premisa de que se trata de un dominio que ha sufrido profundas transformaciones en los últimos años.

La presencia definitiva de lo digital, el cambio en el trabajo de los periodistas y en su relación con las noticias y las fuentes, la creciente tecnificación, el asentamiento definitivo de las redacciones digitales, el disperso uso de la web semántica, así como la relevancia de los medios sociales son fenómenos que dibujan un escenario diferente y que hacen necesaria una revisión de lo que atañe a la organización y representación del conocimiento en este ámbito.

## Abstract

This paper aims to establish the state of the art, on the one hand, about instruments such as thesauri, ontologies and folksonomies and, secondly, procedures and models of formalization for analysis and semantic tagging in order to determine the most relevant aspects in its application in the journalistic domain. Following this review, and taking into consideration the changes that have taken place in journalism, the study points as future reflections focuses: knowledge organization in connection with semantic web, open data, information needs and new models of information retrieval are considered.

**Keywords:** Knowledge organization. Thesauri. Ontologies. Folksonomies. Semantic markup. Documentation. Journalism.

En primer lugar, se aborda el contexto comunicativo en el que nos situamos, caracterizado por fenómenos como el *open data* y el *big data*, especialmente en los elementos que tienen especial incidencia en el plano documental, en general, y en la organización del conocimiento, en particular. A continuación, se hace una revisión de la literatura existente, por un lado, de los lenguajes controlados y, por otro, de los procedimientos de análisis. En tercer lugar, se revisan experiencias aplicadas y académicas ya centradas en el periodismo escrito. Finalmente, se plantean los aspectos centrales necesarios para avanzar en el conocimiento sobre los instrumentos y procedimientos de organización del conocimiento que se aplican y el modo en que se lleva a cabo.

## 2. El contexto comunicativo y periodístico

Es cada vez más patente el impacto que tienen en el periodismo los nuevos avances y desarrollos derivados del *open data* y del *big data*. Como es sabido, con los datos en abierto se busca la eliminación de las restricciones existentes, en di-

ferentes niveles, en la difusión, acceso y reutilización de los datos y la información. Esta realidad impregna diferentes áreas tanto del ejercicio periodístico como de la realidad documental. En gran medida, esta corriente es deudora del trabajo desplegado por los activistas que defienden el libre acceso al conocimiento y está en sintonía con acciones, incluso con elementos estéticos, *hacker*. Por otra parte, la incorporación paulatina del trabajo con algoritmos que ayuden a la interpretación de la gran cantidad de información existente en la red también es uno de los aspectos clave para el análisis de la actividad documental, en general, y de la organización y representación del conocimiento, en particular.

A juicio de Sugimoto, Ding y Thelwall (2012), la propia Documentación no ha sido ajena a los inicios del *big data*, por ejemplo a partir de las diferentes métricas aplicadas al estudio de la web, en concreto en lo que se refiere a la recopilación y análisis de datos web a gran escala, junto a los programas y proyectos de análisis de enlaces. Además, esto se ha hecho en conexión con las fórmulas de integración de la web semántica y los *linked open data*, junto a las herramientas, técnicas y teorías bibliométricas, profusamente empleadas.

Ya en el plano mediático, conforme al trabajo de Stone (2014), es cada vez mayor el interés aplicado de los medios de comunicación, como CCN, Financial Times o The Huffington Post, por el *big data*. En primer lugar, destacan los formatos de visualización. Por otra parte, se presentan como referentes los modelos de análisis de titulares y de adaptación a las audiencias (a través del análisis de comentarios, de la calidad de los productos periodísticos, de métricas *ad hoc*, de preferencias de contenidos, o de comportamientos de navegación, etc.), así como la personalización y segmentación de contenidos periodísticos, el seguimiento de noticias en tiempo real (mediante la identificación, clasificación y alerta de información relevante a usuarios concretos) y, finalmente, la generación de noticias por medio de la condensación de datos y la posterior presentación en formatos concretos.

Una de las vertientes más destacables que muestra conexiones con el hecho documental es el periodismo de datos. En este sentido, Ferreras Rodríguez (2012) caracteriza al periodista de datos alrededor de cuatro ideas, con claras concomitancias documentales: por un lado, la recopilación de datos e información, teniendo en cuenta tanto los datos en abierto como las técnicas de recuperación más avanzadas. En segundo lugar, la tarea de filtrar información con vistas a aportar calidad al trabajo periodístico. A continuación, se encuentra la contextualización y la combinación,

es decir, el cruce con otras fuentes de información y bases de datos (incluyendo aquí la geolocalización), y finalmente, la difusión del producto, con un fuerte componente visual.

Esta modalidad periodística se encuentra vinculada a otros periodismos, como el de investigación o el de precisión, así como a la documentación, la infografía, el "fact checking", el *open data* y el *big data* (Peiró y Guallar, 2013). Se trata de una práctica que nace del uso de grandes cantidades de información, procedentes de fuentes internas y externas, que pueden ser usadas y difundidas en diferentes formatos, a través de canales diferentes. Y que está en conexión con un gran número de aplicaciones que ejecutan, parcial o completamente, tareas de naturaleza documental, con base en una cultura de datos abiertos, donde la clave no solo es su visualización sino el acceso real.

En el trabajo de García Jiménez y Catalina García (2016), después de analizar diferentes casos de periodismo de datos en *El País*, *ABC* y *Elconfidencial.es*, se observa el número elevado de fuentes de información empleadas, la gran variedad de formatos de documentos, y el trabajo con la base de datos como principal herramienta documental. Destaca el uso de software destinado a la elaboración de mapas y visualizaciones (D3) y de datos geoespaciales (CartoDB), así como el de hojas de cálculos tales como Excel o Google Sreadheets.

En un contexto dominado por la transformación radical hacia redacciones digitales, y de la presencia de los medios sociales, y con una proyección más Search Engine Optimization (SEO) que informativa, se hace más necesario que nunca profundizar en el estudio y prospectiva de la relación entre los periodistas y los documentalistas, en este caso en el dominio de la Documentación Informativa. Y, claro está, hacen necesario conectar las acciones de marcado semántico y los instrumentos como los tesauros con el comportamiento informacional, aspecto tratado en estudios como el trabajo de Pintado (2013).

En definitiva, lo que aquí se plantea es si estos fenómenos pueden conducir a un cambio en el modelo documental existente y, en consecuencia, en los procedimientos y herramientas de la organización del conocimiento. En el sentido que apunta Dowd (2016), la complejidad para la gestión documental se eleva, puesto que los datos, enlazados, están siendo empleados para nuevos cometidos gracias a su capacidad de unir diferentes dominios y diversas plataformas. Una de las posibles respuestas vendría de la mano de lo que Heravi y McGinnis (2015) denominan *Social Semantic Journalism*. Por medio de técnicas como

el procesamiento del lenguaje natural y los metadatos semánticos, se mejora la detección y acceso a la información de actualidad, con capacidad para verificarla y contrastar las fuentes.

Desde nuestro punto de vista, no está en juego el papel de la organización del conocimiento, sino conocer cuáles son las fórmulas adecuadas para enfrentarse no solo al periodismo en red (Sampedro, 2014) sino también al activismo de datos (Milan y Gutiérrez, 2015). En un contexto en el que se deben tener en consideración los cambios en la configuración y estructura de las noticias y la realidad de nuevos espacios por donde éstas se difunden, como los medios sociales. Lo que puede poner en cuestión el paradigma documental tradicional, basado en la transformación en piezas almacenadas derivadas en las bases de datos documentales, que están gestionadas con herramientas documentales al uso.

### 3. Instrumentos y procedimientos

En la línea de lo que señala Moreira (2012), la clave siempre se situará en qué medida se puede representar el conocimiento, a partir de la reducción semántica que se propone, con vistas a la recuperación de información. También sugiere tener presente el trabajo que se hace en la teoría del conocimiento, y añadir los objetivos pragmáticos propios de nuestro dominio. Los cambios que se están produciendo en la formas de producción y difusión de la información y del conocimiento evidentemente afectan a las técnicas y finalidades propias de la representación del conocimiento, tarea que hasta ahora se ha hecho a partir de clasificaciones, tesauros y en algunos casos folksonomías y ontologías.

#### 3.1. Tipos de vocabularios

La primera piedra de toque se circunscribe al tipo de lenguaje que se aplica, ya sea taxonomía, folksonomía, tesoro o incluso ontología. Desde un punto de vista global, hay autores que defienden, dadas las características del ciberespacio, la conveniencia de las folksonomías frente a otras opciones. Es la línea de trabajo defendida por Yedid (2013), que subraya los aspectos positivos de las folksonomías:

- Las etiquetas extraídas se corresponden, en gran medida, con los términos que aparecen en los vocabularios controlados. Globalmente, son similares en los conceptos implicados en las etiquetas únicas y en el uso de sustantivos, así como en la utilización de ortografía reconocida. En cualquier caso, y desde su punto de vista, las inconsistencias, por ejemplo en el plural y singular, no representarían un porcentaje muy alto.

- Se acercan más al vocabulario propio de los usuarios (en lo que se refiere a los términos y a la especificidad deseada), lo que puede tener su influencia positiva en la elección del léxico y la traslación en la recuperación de información de las necesidades de información de los individuos, ya que se convierte en un proceso más intuitivo. A su vez, se suman más voces y puntos de vista a esta actividad: se podría hablar de una mayor democratización al considerar a las etiquetas seleccionadas como conocimiento compartido, y se sustituiría así un enfoque autoritario y centralizado por una visión colaborativa.
- Requiere menos esfuerzo que un vocabulario controlado. Este valor es recomendable para situaciones en las que se dispone de presupuestos bajos. Las folksonomías se muestran más cercanas a lo amateur, si bien podría suponer una penalización para entornos profesionales como el que nos ocupa. En realidad, se recomiendan más cuando no hay profesionales de la documentación cerca, o bien si el número de documentos analizados es considerable. En este último caso, puede incluso ser adecuada la distribución de las actividades indizadoras al mayor número de personas posibles, también a los autores. Aunque puedan surgir problemas a raíz de la falta de conocimiento al respecto del fondo documental de o de los propios objetivos documentales explícitos e implícitos. En definitiva, un recurso puede ser indizado más de una vez por diversos usuarios, lo que da lugar a una descripción intersubjetiva, agregada y quizá más exhaustiva.

En lo que se refiere a las ontologías, y a pesar de las concomitancias existentes con el resto de lenguajes (García Jiménez, 2004), y desde un plano puramente documental, Codina y Pedraza (2011) ya apuntan la consistencia de los tesauros. Son tecnologías ya contrastadas con una trayectoria consistente en los sistemas de información, que se derivan de modelos de trabajo con una implantación positiva desde hace muchos años (Mendes et al., 2015). Esto no quita para que las ontologías, que están detrás de algunos avances en documentación, sean instrumentos consolidados en otras esferas, pero pendientes de desarrollos en nuestro ámbito.

En este punto conviene destacar el uso e implementación de SKOS y su conexión con lenguajes pertenecientes a la web semántica; ya que hace posible que el tesoro se presente en un nivel alto de formalización, lo que además de mejorarlo favorece la interconexión con otros lenguajes y la eficiencia de los sistemas de información. En realidad, en el entorno semántico es cada vez

es mayor la presencia de ontologías como Semantically Interlinked Online Communities (SIOC) o Friend Of a Friend (FOAF). La primera hace posible el uso de datos en línea generados por usuarios y permite nuevas aplicaciones semánticas. Mientras, FOAF permite describir a las personas y las conexiones que tienen entre ellas.

Por su parte, partiendo de una posición relativa a las clasificaciones, Szostak (2014) entiende que las ontologías deben hacer frente a dos problemas serios si quieren una plena integración en la web semántica. La gran cantidad de ontologías, y los diferentes orígenes y supuestos de partida complican su traducción e interoperabilidad. La conexión entre bases de datos por ordenadores a través de ontologías incompatibles se presenta como una dificultad seria. Otra pega viene de la mano de la terminología, puesto que en términos generales es muy vaga. Una posible solución puede provenir del formato de RDF *triplets*, y mediante el esquema de cosas, relaciones y propiedades como base para dar forma a la estructura de la web semántica.

Finalmente, en el trabajo de Soler y Gil (2010) se analizan las diferencias y concomitancias de los diferentes instrumentos señalados. A partir de categorías válidas para el análisis de nuestro objeto de estudio (objetivo, productor, lenguaje con que se trabaja, estructuras que se aplican, los tipos de relaciones, el coste de elaboración y mantenimiento, la actualización, las aplicaciones, los modelos de presentación, la inclusión de definiciones, las propiedades de los términos, el control de la ambigüedad y de la sinonimia, y los editores que facilitan la construcción) concluyen que los diferentes lenguajes, por nivel de complejidad, y de menor a mayor, se escalarían de la siguiente forma: las folksonomías, las taxonomías, los tesauros y las ontologías.

### 3.2. Nuevos espacios

Sánchez, Colmenero y Moreiro (2012) apuntan, durante los últimos años, la relevancia del tesauro y su formalización (lo que condiciona la estructura relacional) en aras de potenciar los vínculos e interoperabilidad de los sistemas de organización del conocimiento (KOS, Knowledge Organization System).

Los instrumentos de organización del conocimiento también son deudores de SKOS en la medida que hacen más fácil su uso y aumentan su eficiencia en conexión con la web semántica. En este sentido, Pastor et al. (2012) analizan el comportamiento de diferentes lenguajes controlados que aplican el modelo de datos Simple Knowledge Organization System (SKOS) en el marco de la web semántica y el *linked open data*. En

cuanto al acceso a los datos, tanto los tesauros como las clasificaciones consiguen buenos resultados en comparación con los encabezamientos de materia y los ficheros de control, que muestran un peor comportamiento. Pero si nos fijamos en el grado de interoperabilidad, los resultados obtenidos por los encabezamientos de materia son mejores, por lo que cabría esperar una mayor integración de SKOS en las bibliotecas. Por último, parece sorprendente el grado bajo de interconexión entre tesauros, lo que quizá se podría solucionar a través de un mayor esfuerzo en las propiedades de mapeado entre conceptos.

Conviene no dejar al margen la potencial relevancia de los gestores de tesauro. En el trabajo de Martínez y Alvite (2014) se aplica un método para evaluar las herramientas de gestión de tesauros a partir de criterios como el objetivo de la aplicación, sus funcionalidades, la capacidad para mantener la integridad y la coherencia del vocabulario, y los aspectos relacionados con la interoperabilidad, especialmente en lo que se refiere a la compatibilidad con estándares propios de la web semántica (RDF/SKOS), la integración con otros sistemas de información, o la capacidad de combinación con otros tesauros. En definitiva, se percibe una clara diferenciación entre las aplicaciones de gestión de tesauros y las aplicaciones orientadas a la web semántica, fundamentalmente a SKOS. Para estos autores, son dignos de mención PoolParty o SKOSed, dada su capacidad para adoptar formatos de web semántica, o bien Turtle y N3, a la hora de importar y exportar de tesauros.

Otro de los aspectos que puede ser de interés es el de la automatización, en diferentes niveles, en la construcción de los tesauros. En el trabajo de Gil y Sánchez (2014) emplean las técnicas de recuperación de información para la generación de tesauros, gracias a las técnicas de cluster para los términos empleados y la categorización, y a las técnicas de similitud para la propuesta de las relaciones semánticas.

### 3.3. Análisis, etiquetado y formalización

En lo que se refiere al análisis y marcado semántico, Pastor (2013) defiende planteamientos que superen el actual papel de Schema.org, siempre teniendo en consideración que el fin último es el de incrementar la efectividad de la recuperación de información. Por ejemplo, el uso de datos estructurados en código X(html) en los sitios web presentan consecuencias positivas tanto en la interoperabilidad semántica, como en la recuperación y la reutilización de la información. De todas formas, lo recomendable sería mirar hacia el modelo que se considere más apropiado tanto para



las tareas analíticas como para su reconocimiento por parte de los motores de búsqueda, ya sea como microdato, microformato o RDFa, o sus combinaciones. Y en cada opción primará la naturaleza semántica de los contenidos, el objetivo de uso, así como la posible reutilización que se plantee.

Según este mismo autor, y conforme al paradigma actual, el posicionamiento de los contenidos está íntimamente ligado al uso de microdatos y al Schema.org. Excepto en aquellos casos que requieren una descripción del contenido muy detallada, en las publicaciones ligadas al *linked open data* y en generación automática de datos a partir de contenidos web, lo que requiere de RDFa. Ante este panorama, posiblemente una solución puede venir de la mano de auténticas redes de equivalencias entre los diferentes modelos de formalización. Son muchos los problemas que se deben solucionar, entre los que destaca la conexión entre el mercado y los vocabularios o la relación con los modelos de consulta y búsqueda.

Por otra parte, y más allá de los modelos de formalización, también destacan fenómenos como el etiquetado social que, a juicio de Pérez et al. (2013), hace progresar la gestión de la información. Desde su punto de vista, este procedimiento permite plantear términos y etiquetas más concretas y con mayor relación con los intereses de los usuarios. Se ajustan más a los entornos compartidos y colaborativos, entendidos como espacios donde se distribuyen libremente etiquetas y recursos. Una de las derivas que tiene este enfoque es la creación de auténticas comunidades o grupos de usuarios, y la combinación con otras formas de comunicación, e incluso de organización social, lo que se puede producir a través de la valoración, positiva o negativa, ante una información o dato.

En cualquier caso, es ya tradición la escasa atención que se presta desde la academia a la selección de términos que, en su momento, servirán de base a los vocabularios y a la recuperación de información. Hecho nunca exento de matices ideológicos y de impacto en la sociedad y los individuos, especialmente en campos como el periodístico.

#### 4. Organización del conocimiento en el periodismo

Las últimas experiencias que se observan en la literatura sobre el particular están referidas, en primer lugar, a nuevos lenguajes documentales. Rubio Lacoba (2012) aborda el caso de *El País*, que goza de una folksonomía controlada, basada en una especie de indexación social que se

asigna a los diferentes productos periodísticos del diario. Sobresale que los periodistas también propongan las etiquetas, con posteriores filtros realizados por los documentalistas al objeto de eliminar tanto la sinonimia como la polisemia y permitir el enriquecimiento con relaciones entre los términos-etiquetas.

La base de trabajo se encuentra en las experiencias del *The New York Times*, *Delicious* y *Soitu*. Se trata de un lenguaje que también se ha nutrido de las base de datos propias, el IPTC en su versión digital, así como de otros tesauros especializados. De igual modo, aplica técnicas SEO para la delimitación de los términos. Con una estructura organizada en las áreas de temas, personajes, organizaciones, lugares y eventos, presentaba 74.000 términos (incluyendo el nombre de personajes) en 2012. El etiquetado se realiza antes de la publicación de la noticias y, por medio de portadillas dinámicas, se vincula de forma automatizada cada noticia a los temas con los pueden plantear algún tipo de conexión. El editor también permite: incluir y vincular nuevas etiquetas; eliminar la ambigüedad fundamentalmente de siglas, cargos y fechas; proponer palabras en comparación con noticias similares; recibir solicitudes de términos por parte de los periodistas; y establecer etiquetas que no se pueden quedar en el lenguaje principal pero que se corresponden con "conceptos editoriales".

En esta misma línea de modificación de los tesauros, destaca la aproximación de Baños (2013), centrada en la capacidad que ofrecen las noticias de divulgación científica para la renovación y actualización de los tesauros. Por medio de técnicas propias de la recuperación de información (N-gramas y Apache Solr), se plantean equivalencias con diferentes macrotesauros. La primera conclusión de utilidad es que estos macrotesauros no son adecuados para el análisis de noticias orientadas a la divulgación científica, fundamentalmente por la ausencia de términos que resultan. No obstante, queda demostrado que se pueden utilizar como base para la elaboración de otros instrumentos de organización del conocimiento. Las noticias de divulgación resultan apropiadas para dicha actualización, y se constituyen como un auténtico yacimiento de conceptos y términos.

Centrados en el campo de los metadatos y el mercado semántico, Baños et al. (2015) revisan, en el dominio periodístico, los diversos de modos de representación de los metadatos, caracterizados por falta de uniformidad, lo que va en contra realmente de los objetivos de estandarización e interoperabilidad. En su estudio se observa que frente a la mayor presencia, en la bibliografía so-

bre el particular, de NewsML y NIFT, lo más habitual es encontrarse con schema.org así como dos esquemas orientados a los medios sociales, que son Twitter Cards y Open Graph Protocol (en Facebook) en los códigos fuente. En menor medida, se acude a normas como rNews, que está preparada para aplicar marcas semánticas a noticias en sistemas de información web, o bien al HTML, el Dublin Core o los microdatos.

Por otra parte, y más allá de estas consideraciones técnicas, García Gutiérrez (2011), que trabaja alrededor de la idea de exomemoria, sostiene la necesidad de aportar una perspectiva pluralista y democrática en las memorias digitales, mediante instrumentos como los tesauros, las ontologías, y las prácticas del análisis documental. Desde un enfoque transcultural y participativo, incide en la relevancia social de las actividades que giran alrededor de las etiquetas, los metadatos, las categorías, las relaciones semánticas. Su mirada se inclina por lo transnacional y transcultural al objeto de superar las limitaciones y enclaustraciones propias de imposiciones positivistas o estereotipadas.

Partiendo de estos postulados, y en el marco del trabajo con mecanismos lógico semánticos y operadores lógico-críticos, propone un análisis documental-semántico en el campo periodístico, capaz también de registrar la información observable y de detectar omisiones, errores y solapamientos (García Gutiérrez, 2013, 2014; García Gutiérrez y Martínez Ávila, 2014)). Su registro modelo se compone de, en primer lugar, una descripción externa, en la que se incluye información sobre el autor, el título, el medio de publicación, la fecha, y la página (o URL). A continuación, plantea la macroproposición, que es una síntesis textual de la acción, considerada la clave de toda la estructura analítica, y planteada mediante un verbo o nombre de proceso. También se incluye el sujeto, que abarca el ser vivo, colectivo o cosa que lleva a cabo la acción que predomina, previamente seleccionada. Y finalmente, el predicado, que se constituye conforme a los aspectos que se deriva de la ejecución de la acción principal definida. Los puntos más relevantes en los que se debe fijar el analista a lo largo del texto son: sujetos, objetos, elementos asociativos, situación, causa, finalidad, consecuencia, modo, instrumento, lugar y tiempo.

## 5. Puntos de análisis

A partir del estado de la cuestión que se ha planteado hasta el momento, proponemos los puntos sobre los que puede realizarse un análisis sobre la organización y representación del conocimiento en la documentación periodística en su apartado

textual. En un primer momento, se debe analizar el contexto teniendo en consideración algunas de las siguientes cuestiones:

- Redacción y convergencia digital: esto tiene que ver tanto con las prácticas periodísticas en sí como con las prácticas digitales cotidianas.
- Características del medio: desde diferentes enfoques que van desde la orientación editorial, el trabajo diferenciado conforme a las plataformas, la presencia en papel, el número de visitas, o las rutinas informativas existentes, entre otros.
- Perfiles profesionales: características de periodistas, programadores, especialistas en medios sociales. Cada vez tienen más influencia las personas encargadas de tareas como la generación de flujo de información o en la aplicación métricas web y su conexión con la documentación y la organización del conocimiento.
- Orientación documental: es relevante detenerse en la existencia de documentalistas y de un servicio de documentación. Interesa detectar funciones y características.
- Usuarios del sistema: es decir, hábitos informacionales globales y necesidades de información.
- Bases de datos: sistema de gestión empleado y diseño de la base de datos.
- Fondo documental: características y naturaleza de la colección de documentos y datos existentes.

En cuanto al vocabulario, los elementos que se tendrán en consideración son los siguientes:

- Tipología: la primera piedra de toque se circunscribe al tipo de lenguaje que se aplica, ya sea taxonomía, folksonomía, tesauro o incluso ontología. O bien combinación de clases.
- Antecedentes y fuentes: se ha de precisar cuáles son los antecedentes, si los tiene, el lenguaje y con qué bases se ha construido.
- Términos: en lo que se refiere a cuestiones como el número, la tipología, el procedimiento de selección, las formas de refinamiento, el nivel de especificidad o el impacto en la recuperación de información.
- Estructura: formas de y existencia de relaciones semánticas y de formas de categorización. Organización conceptual del propio instrumento.

- Productores y usuarios: qué profesionales y en qué forma son los encargados de la elaboración y mantenimiento, y cuáles de su acceso y consulta.
- Actualización: se debe analizar los métodos y periodos de actualización.
- Modelos de interoperabilidad y compatibilidad semántica: conexión en diferentes niveles y sentidos con las normas o propuestas de formalización. También se incluye la conexión con otros vocabularios o instrumentos.
- Gestor: software empleado para la elaboración y mantenimiento del constructo.
- Evaluación y calidad, norma: nivel de seguimiento de las normas vigentes. Seguimiento de la calidad.
- Coste: es importante determinar el coste tanto de la elaboración como del mantenimiento.
- Aplicaciones periodísticas: se trata la virtualidad directa e indirecta en el medio de comunicación.

Por otra parte, y en lo que se refiere al marcado semántico, se debe analizar los siguientes aspectos:

- Elementos y categorías de descripción semántica: cuáles son las categorías que se emplean.
- Tipología de la descripción: con vistas a determinar si se plantea un etiquetado social o un etiquetado profesional.
- Procedimiento de elaboración: análisis, lo más detallado posible, de los mecanismos que se emplean para determinar que un término o etiqueta es válido para la representación.
- Lenguaje: análisis de la relación entre el análisis o marcado y vocabulario correspondiente.
- Recuperación de información: estudio sobre las implicaciones de esta tarea para la recuperación de información y la resolución de las necesidades de información.
- Formalización y codificación: estudio sobre el modelo de formalización que se plantea para esta actividad.
- Calidad y evaluación: hemos de fijarnos en los procedimientos, si existen, para evaluar el resultado del etiquetado o marcado semántico.

## 6. Conclusiones

Tal y como se señalaba al principio de este trabajo, se ha establecido el contexto y la situación de las técnicas e instrumentos que afectan a la organización y representación del conocimiento, en especial a lo que tiene que ver como los tesauros, las folksonomías y los modelos de análisis y marcado semántico. A continuación, se ha analizado la situación de este ámbito en lo que se refiere a la Documentación periodística. Y esto se ha llevado a cabo para delimitar, finalmente, aquellos temas que necesitan de una mayor concreción para dibujar el escenario real de la organización del conocimiento en el cambiante ámbito periodístico.

Como aspectos destacables podemos señalar, por un lado, la continuidad de la creciente tecnificación y de las propuestas vinculadas a la web semántica. No obstante, parece que este proceso choca con la falta de aplicaciones normativas unívocas. En segundo lugar, se debe señalar también el impacto real de lo que podríamos denominar como organización del conocimiento social. No solo se observa en la generalización del uso de los medios sociales, con las consecuencias correspondientes para los sistemas de información, sino también en los lenguajes y el marcado semántico.

## Referencias

- Baños Moreno, María José (2013). Fuentes para la actualización de macrotesauros: noticias de divulgación científica. // Cuadernos de Gestión de Información. 3 (2013) 13-24.
- Baños Moreno, María José; Felipe, Eduardo R.; Pastor Sánchez, Juan Antonio; Martínez Béjar, Rodrigo y Lima, Gercina (2015). Metadatos en noticias: un análisis internacional para la representación de contenidos en periódicos. // XII Congreso ISKO España y II Congreso ISKO España-Portugal, 19-20 de noviembre, 2015.
- Codina, Lluís; Pedraza Jiménez, Rafael (2011). Tesauros y ontologías en sistemas de información documental. // El profesional de la información. 20:5 (sept-oct, 2011) 555-563.
- Dowd, Cate (2016) The new order of news and social media enterprises: visualisations, linked data, and new methods and practices in journalism. // Communication Research and Practice. 2:1 (2016) 97-110. DOI:10.1080/22041451.2016.1155339
- Ferreras Rodríguez, Eva M. (2012). Nuevos perfiles profesionales: el Periodista de Datos. // Actas IV Congreso Internacional Latina de Comunicación Social. [http://www.vistalatinacs.org/12SLCS/2012\\_actas/062\\_Ferreras.pdf](http://www.vistalatinacs.org/12SLCS/2012_actas/062_Ferreras.pdf)
- García Gutiérrez, Antonio (2011). Epistemología de la Documentación. Barcelona: Stonberg, 2011.
- García Gutiérrez, Antonio (2013). Entrevistas periodísticas y textos declarativos: un procedimiento de análisis y recuperación documentales basado en modos de búsqueda pronominal. // El profesional de la información. 22:4 (julio 2013), 315-325

- García Gutiérrez, Antonio (2014). Análisis documental de noticias de prensa en sistemas de información factual. // *Revista Española de Documentación Científica*. 37:2 (abril-junio, 2014) e046.
- García Gutiérrez, Antonio; Martínez Ávila, Daniel (2014). Critical organization of knowledge in mass media information systems. // *Knowledge organization*. 41:3 (2014) 205-216
- García Jiménez, Antonio (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. // *Anales de Documentación*. 7 (2004), 79-95. <http://revistas.um.es/analesdoc/article/view/1691/1741>
- García Jiménez, Antonio y Catalina García, Beatriz (2016). Procesos y herramientas documentales para el periodismo de datos. // 1º Congreso Internacional sobre Periodismo de Datos. Nuevas narrativas para el periodismo especializado. Málaga, 25-26 de febrero de 2016 (en prensa).
- Gil Urdiciain, Blanca; Sánchez Jiménez, Rodrigo (2014). Técnicas de recuperación de información aplicadas a la construcción de tesauros. // *TransInformação*. 26:1 (jan-abr, 2014) 19-26.
- Heravi, Bahareh Rahmanzadeh; McGinnis, Jarred (2015). Introducing Social Semantic Journalism. // *The Journal of Media Innovations*. 2.1 (2015) 131-140
- Martínez González, M. Mercedes; Alvite Díez, M. Luisa (2014). Propuesta metodológica de evaluación de gestores de tesauros compatibles con la web semántica. // *Anales de Documentación*. 17:1 (2014).
- Milan, Stefania y Gutiérrez, Miren (2015). Medios ciudadanos y big data: la emergencia del activismo de datos. // *Mediaciones*. 14 (enero-junio, 2015) 10-26.
- Mendes, Paula Raphisa; Martins dos Reis, Raquel y Coura Moreira dos Santos Maculan, Benildes (2015). Tesauros no acesso à informação: uma retrospectiva. // *Revista ACB: Biblioteconomia em Santa Catarina*. 20:1 (2015) 49-66.
- Moreira, Walter (2012). Relações conceituais como ponto de inflexão entre as linguagens documentais, a terminologia e as ontologías. // *Scire*, 18:2 (2012), 123-127.
- Pastor Sánchez, Juan Antonio (2013). Marcado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data. // *Scire*. 19:2 (2013) 55-68.
- Pastor Sánchez, Juan Antonio; Martínez Méndez, Francisco Javier; Rodríguez Muñoz, José Vicente (2012). Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. // *El profesional de la información*. 21:3 (may-jun, 2012) 245-253.
- Peiró, K. y Guallar, J. (2013). Introducció al periodisme de dades. Característiques, panoràmica i exemples. // *Item*. 57 (2013) 23-36
- Pérez Sanchidrián, Elaine; Campos Posada, Raúl, y Campos Posada, Gloria Elisa (2014). Etiquetado social: un modelo de representación de la información en la blogosfera. // *Biblios*. 56 (2014) 19-27.
- Pintado Navarro, Rocio (2013). El comportamiento informacional de los periodistas en la Región de Murcia. *Cuadernos de Gestión de Información*. 3 (2013) 25-51.
- Rubio Lacoba, María (2012). Nuevas destrezas documentales para periodistas: el vocabulario colaborativo del diario El País. // *Trípodos*. 31 (2012) 65-78.
- Saad Corrêa, Elizabeth y Bertocchi, Daniela (2012). A cena cibercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria. // *MATRIZES*. 5:2 (ene-jun, 2012) 123-144
- Sampedro, Víctor (2014). El cuarto poder en red. Por un periodismo (de código) libre. Barcelona: Icaria, 2014.
- Sánchez Cuadrado, Sonia; Colmenero Ruiz, María Jesús y Moreiro, José Antonio (2012). Tesauros: estándares y recomendaciones. // *El profesional de la información*. 21:3 (may-jun, 2012) 229-235.
- Soler Monreal, Concha y Gil Leiva, Isidoro (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías. // *Revista Interamericana de Biblioteología*. 33:2 (jul-dic, 2010) 361-377.
- Stone, Martha L. (2014). Big data for Media. Reuters Institute for the Study of Journalism. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media\\_0.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media_0.pdf) (23-02-2016)
- Sugimoto, Cassidy R.; Ding, Ying and Thelwall, Mike (2012). Library and Information Science in the Big Data Era: Funding, Projects, and Future [a panel proposal]. // *Proceedings of the ASIST 2012*. 49:1 (2012), 26-31.
- Szostak, Rick (2014). Advances in Classification Research Online 2013. Classification, Ontology, and the Semantic Web. // *Advances In Classification Research Online*. 24:1 (2014) 30-37.
- Yedid, Nadina (2013). Introducción a las folksonomías: definición, características y diferencias con los modelos tradicionales de indización. // *Información, Cultura y Sociedad*. 29 (dic. 2013) 13-2.

---

Enviado: 2016-03-31. Segunda versión: 2016-07-18.  
Aceptado: 2016-07-18.

---

# Integration of automatic spatial annotations from different sources by means of semantic technologies

---

**Javier NOGUERAS-ISO (1), Verónica LÁZARO (2), Ludovic MONCLA (3)**

(1) Aragon Institute for Engineering Research (I3A), Universidad de Zaragoza, C/ María de Luna 1, 50018 Zaragoza, Spain; jnog@unizar.es (2) vlazgar@gmail.com (3) LIUPPA, Avenue de l'Université, BP 1155, 64013 Université Pau Cedex, France; ludovic.moncla@univ-pau.fr

## Resumen

La extracción de información es una de las tareas principales de la minería de textos que resulta de gran utilidad para todo tipo de aplicaciones que exploten la información geográfica ya que hay gran cantidad de información geográfica que no se recopila directamente en formatos reconocibles por Sistemas de Información Geográfica, sino directamente como texto plano. Actualmente existen diversas soluciones informáticas para el procesamiento de textos y la anotación de entidades espaciales. Sin embargo, el problema que tienen estas herramientas es que producen como resultado de este procesamiento un texto anotado con lenguajes de marcado propio, que dificulta su integración en otros sistemas. El objetivo de este trabajo es proponer la conversión de la salida de estas herramientas a un lenguaje de anotación espacial común basado en tecnologías semánticas que facilite su integración e interoperabilidad. Como factor común de los lenguajes de marcado se propone una anotación de textos basada en RDFa y utilizando el vocabulario de la iniciativa internacional Schema.org. Para validar la utilidad de esta propuesta se ha creado la infraestructura necesaria para construir un repositorio semántico de documentos donde se integren y armonicen las anotaciones generadas por diversas herramientas de anotación existentes.

**Palabras clave:** Extracción de información. Integración de datos. Información geográfica. Web semántica. RDFa. Schema.org.

## 1. Introduction

Information Extraction is one of the main tasks in Text Mining, which is essential for all types of applications exploiting geographic information because there is a big volume of geographic information not directly compiled in specific formats proposed by Geographic Information Systems, but just embedded in plain text sources. For instance, in this context we can find different types of narrative structures describing: itineraries in travelogues or travel novels (Leidner and Lieberman, 2011); geographic boundaries in legislative documents; locations through text transcriptions

## Abstract

Information extraction is one of the main tasks in text mining, which is essential for all types of applications exploiting geographic information because there is a big volume of geographic information not directly compiled in specific formats proposed by Geographic Information Systems, but just embedded in plain text sources. Currently, there are several software solutions for the processing of texts and the annotation of spatial named entities. However, the problem of these tools is that their output is based on heterogeneous annotation languages, which make it difficult their integration in other systems. The objective of this work is to propose the conversion of the output of these tools into a common spatial annotation language based on semantic technologies to facilitate their integration and interoperability. As a common annotation language we propose the use of a text annotation based on RDFa and using the vocabulary proposed by the international initiative Schema.org. In order to validate this proposal, we have created the necessary infrastructure to build a semantic repository of documents, where the annotations generated by different annotation tools can be integrated and harmonized.

**Keywords:** Information extraction. Data integration. Geographic information. Semantic web. RDFa. Schema.org.

of calls to emergency services; or any type of official information associated to a geographic location and found in official bulletin documents (López-Pellicer et al, 2012).

For the processing of these information sources and building a structured and simplified view of their content, the recognition of named entities (Named Entity Recognition - NER) is an essential activity (Sekine and Ranchhod, 2009). Nowadays, there are several software solutions like OpenNLP, OpenCalais, CasEN or PERDIDO for the processing of texts and the annotation of spatial entities. However, the problem of these tools

is that their output is based on heterogeneous annotation languages, which makes it difficult to integrate them in more complex systems exploiting these annotations.

The objective of this work is to propose the conversion of the output of these tools into a common spatial annotation language based on semantic technologies to facilitate their integration and interoperability. As a common annotation language we propose the use of a text annotation based on RDFa and using the vocabulary proposed by the international initiative Schema.org. On the one hand, RDFa is a set of extensions to XHTML proposed by the World Wide Web Consortium (W3C) to embed semantics in documents through RDF triples. On the other hand, Schema.org is an initiative promoted by the main search engine companies (e.g. Google, Yahoo, Microsoft, Yandex) to define a common vocabulary of entities (Person, Place, Date, ...) to annotate the semantics of Web resources. The initial hypothesis is that an annotation approach based on standards accepted by the main stakeholders in Web development would increase the general use of these tools for the publication of enriched Web content.

In order to validate this proposal, we have created the necessary infrastructure to build a semantic repository of documents, where the annotations generated by different annotation tools can be integrated and harmonized. Additionally, this repository will provide as well query functionalities for searching documents in terms of spatial filters that take profit of previously annotated spatial entities. Moreover, this repository can be seen as data integration solution (Doan et al, 2012) as annotations obtained from different tools are linked to the same original document.

The rest of the paper is structured as follows. Section 2 describes the annotation tools that have been reviewed for this work. Section 3 introduces our proposal for a common spatial annotation based on semantic technologies. Section 4 describes the design of the infrastructure for a document repository, whose validation is reported in section 5. Last, section 6 concludes the work and presents some issues for future work.

## 2. Tools for spatial named entity recognition

Among the systems for Named Entity Recognition and Classification (NERC) specifically focused on the detection of spatial named entities (place names or toponyms) we can find: approaches based on Natural Language Processing (NLP), in particular using syntactico-semantic rules (Moncla et al, 2014); approaches based on machine learning (Santos et al, 2014); or even

hybrid solutions combining previous approaches (Leidner, 2008). In any case, the most typical feature of this specific NERC systems is the need to use external lexical resources, known as gazetteers, to facilitate the connection between place names and their explicit reference, in terms of geographic coordinates, to the Earth surface.

This section describes briefly the features of four tools for the annotation of spatial named entities and the annotation languages used in their output formats. These tools are OpenNLP, CasEN, PERDIDO and OpenCalais. The reason to study these tools is their availability to be freely accessed and their capability to process texts in English, French or Spanish.

OpenNLP (Apache, 2010) is an Apache Project, available as an Open Source library, which facilitates a set of tools based on machine learning for NLP tasks. In particular, NERC can be performed thanks to a *Tokenizer* tool, which divides sentences into *tokens*, and a tool known as *Name Finder*, which facilitates the named entity recognition. Currently, given the availability of dictionaries, the spatial named entity recognition is only possible for documents in Dutch, Spanish or English. The output generated by OpenNLP is just plain text. For instance, the result of processing the string “*I live in Zaragoza.*” would be the following:

[3..4] *location*

where number 3 indicates the initial token of the place name in the text, number 4 indicates the position of the final token, and *location* indicates the type of the recognized spatial named entity.

CasEN (University of Tours, 2016) is a tool developed at University of Tours (France), also available as Open Source, which offers a solution based on syntactico-semantic rules for the recognition of named entities. In particular, the recognition of entities is based on the definition of cascades of transducers (a specific type of finite state automata) (Maurel and Friburger, 2004) working on top of the Unitex platform, a research project from the University of Paris-Est (France) to facilitate processing text tasks. As the dictionaries for spatial annotation are only available in French, CasEN can be used only for French texts. The annotation output generated by CasEN for a sentence containing the place name *Pau* is the following:

*J'ai visité les villages au sud de {Pau,.entity+loc+adm+town+grftagNToponyme}*

As it can be observed, the information of the recognized named entities is shown within curly brackets: *entity* for indicating the detection of an entity; *loc+adm+town* for indicating the type of entity;

and *grftagNToponyme* for indicating the graph (transducer cascade) that detected the entity.

PERDIDO (Moncla, 2015a) is the result of a research project made in collaboration between the LIUPPA laboratory at University of Pau (France), the Advanced Information Systems Laboratory at University of Zaragoza (Spain), and the COGIT laboratory at National Geographic Institute of France. This tool, accessible as a Web service, is specifically focused on the recognition of spatial named entities. As well as CasEN, NERC in PERDIDO is based on the application of syntactico-semantic rules on top of the Unitex platform, but also using some additional heuristics based on clustering techniques for the resolution of fine-grained toponyms. Currently, the lexical resources and rules used by this tool enable the processing of texts in French, Spanish and Italian, and it generates an XML based annotation based on the TEI (*Text Encoding Initiative*) markup language. Left side of figure 1 shows an example of the annotation produced for the string input “Madrid”, which is recognized as a place name with the *placeName* tag. Additionally, the service provided by PERDIDO facilitates another XML file with additional information about the detected toponym (see right side of figure 1). This additional information is recovered from the gazetteer services used as external knowledge resources and consists of information items like the standard name of the toponym (*name* tag), entity type (*feature* tag), geographic coordinates (*coord* tag), country (*country* tag) or continent (*continent* tag).

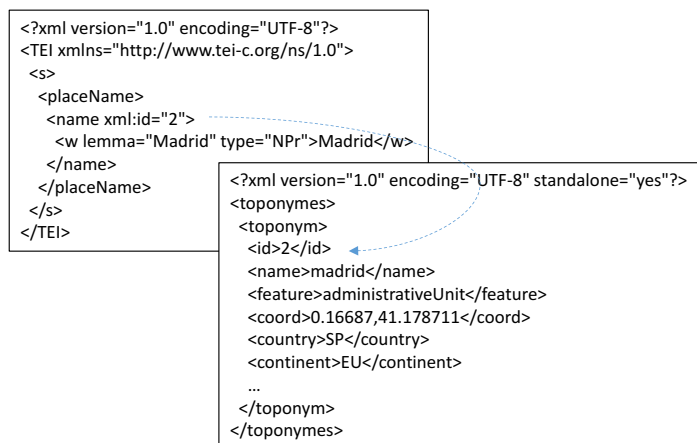


Figure 1. Example of the annotation generated by PERDIDO service

OpenCalais (Thomson Reuters, 2016) is a Web service provided by Thomson Reuters for NERC, which adopts a hybrid approach based on a combination of statistics techniques, machine learning and rule-based methods. Within the scope of spatial annotation, OpenCalais is able to process

and annotate texts in English, French and Spanish for the recognition of city and country names. The output format generated by OpenCalais is RDF serialized as XML (see discussion on RDF in next section). Figure 2 shows an example of this annotation format obtained as the result of processing the input string “London”. In this example it can be observed that tags *c:shortname*, *c:latitude*, *c:longitude* and *c:containedbycountry* inform about the name, the latitude, the longitude and the country to which the detected city belongs respectively.

```
<rdf:type rdf:resource="http://s.opencalais.com/1/type/er/Geo/City" />
<c:docid
  rdf:resource="http://d.opencalais.com/dochash-1/02d32132-5c1d-38a3-8148-2977e6f9a0ef" />
<c:name>London, Greater London, United Kingdom</c:name>
<c:shortname>London</c:shortname>
<c:latitude>51.517124</c:latitude>
<c:longitude>-0.106196</c:longitude>
<c:containedbystate>Greater London</c:containedbystate>
<c:containedbycountry>United Kingdom</c:containedbycountry><!-- London -->
<c:subject
  rdf:resource="http://d.opencalais.com/genericHasher-1/6fda72fd-105c-39ba-bb79-da95785a249f" />
<c:rcscode>G:21P</c:rcscode>
</rdf:Description>
<rdf:Description
  rdf:about="http://d.opencalais.com/genericHasher-1/6fda72fd-105c-39ba-bb79-da95785a249f"
  rdf:type rdf:resource="http://s.opencalais.com/1/type/em/e/City" />
<c:forenduserdisplay>true</c:forenduserdisplay>
<c:name>London</c:name>
</rdf:Description>
```

Figure 2. Example of the annotation generated by OpenCalais service

In summary, after the description of these NERC tools, some of them widely used for English, French and Spanish texts, one can check that there is great heterogeneity with respect to the granularity of the toponyms detected and, overall, the output formats used for the annotation.

### 3. Annotation proposal based on RDFa and Schema.org

RDF (*Resource Description Framework*) is a family of specifications proposed by W3C, which was originally designed as a data model for metadata, but currently it is extensively used to describe semantics of information in a machine-readable way (Cyganiak et al, 2014). The representation model of RDF is based on the use of information triples *subject-predicate-object*: a resource *x* (subject) has a property *y* (predicate) with a value *z* (object).

Apart from expressing RDF semantics in a graphical way thanks to a graph formalism (e.g., see later discussion on figure 5), this semantics can be serialized as plain text (Turtle/N3 format), XML (RDF/XML format), or embedding the triples within XHTML documents using an extension of XHTML specification, which is known as RDFa specification (Herman et al, 2015).



Our proposal for a common language for spatial annotation of texts is to use XHTML with the RDFa extension and using the vocabulary proposed by Schema.org. On the one hand, XHTML is an XML version of the HTML for the construction of web pages that allows to maintain the original text processed by the spatial annotation tools. On the other hand, thanks to the use of RDFa we can benefit from *meta* and *link* XHTML attributes to introduce the semantics defined by the Schema.org vocabulary.

The mission of Schema.org is to create, maintain and promote schemas for structured data on the Internet. The vocabulary of Schema.org covers entities, relations between entities, and actions. Additionally, it can be easily extended thanks to a well-documented extension model.

In particular, taking into account the context of this work where we are going to process mainly texts describing itineraries, we will initially consider the annotation of the full input document as an instance of a *TravelAction*, i.e. an action involving a displacement. Within a document, we propose to annotate every recognized spatial entity using the following entities proposed by Schema.org: *Place* for the general annotation of place names; *City* for the specific annotation of city names; *Country* for country names; *Continent* for continent names; *BodyOfWater* for water bodies like lakes or rivers; and *Mountain* for geographic accidents. These entities cover the types of spatial named entities recognized by the tools reviewed in section 2, but in the future this set of entities could be extended to take into account other types covered by new annotation tools.

Schema.org	PERDIDO	Schema.org	OpenCalais
Place	placeName	Country	Country
	toponym	geo.latitude	c:latitude
name	name	geo.longitude	c:longitude
description	feature	name	c:shortname
geo	coord	City	City
Country		name	c:shortname
name	country	geo.latitude	c:latitude
Continent		geo.longitude	c:longitude
name	continent	Country	c:containedbycountry
		name	

Schema.org	CasEn	Schema.org	OpenNLP
Place	loc.adm.town / loc.adm.reg / loc.adm.nat / loc.phys.sup	Place	location
name		name	
Mountain	loc.phys.geo		
name			
BodyOfWater	loc.phys.hydro		
name			

Figure 3. Mapping between Schema.org and annotation languages of other tools

With respect to the information provided for each annotated entity, all types of entity share a property called *name* to store the recognized place name. Additionally, depending on the detailed information provided by each annotation tool, additional properties like *geo* or *description* can be used to annotate the associated geographic coordinates or the type of entity respectively. Figure 3 shows the mapping between the vocabulary proposed by Schema.org and the annotation languages used by the annotation tools reviewed in section 2.

```

<?xml version="1.0" encoding="UTF-8" ?>
<HTML xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:schema="http://schema.org/"
xmlns="http://www.w3.org/1999/xhtml"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
>
<HEAD/>
<BODY typeof="schema:TravelAction" property="schema:url" content="Near_the_Puerta_del_Sol4">
<p>
  I live in
  <div property="schema:location" typeof="schema:Place">
    <h1 property="schema:name">Madrid</h1>
    <h1 property="rdfs:comment">OpenCalais</h1>
    <div property="schema:geo" typeof="schema:GeoCoordinates">
      <meta property="schema:latitude" content="40.4"/>
      <meta property="schema:longitude" content="-3.6833"/>
    </div>
    <div property="schema:location" typeof="schema:Country">
      <h1 property="schema:name">Spain</h1>
    </div>
  </div>
</p>
</BODY>
</HTML>

```

Figure 4. Example of our annotation proposal (XHTML-RDFa + Schema.org)

Figure 4 shows an example of the annotation for the text "I live in Madrid" in XHTML-RDFa format and using the Schema.org vocabulary. As it can be seen, the full document is a resource of *TravelAction* type and the recognized spatial named entity *Madrid* has been annotated as a *location* property of this resource.

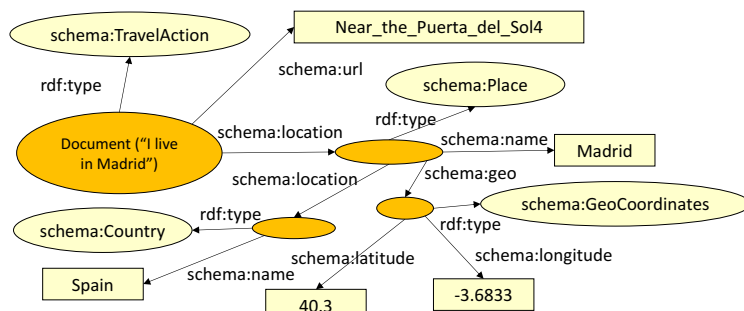


Figure 5. RDF triples under the perspective of a graph model

Additionally, in our annotation proposal we include complementary information about the annotation process. For instance, thanks to the



*schema:url* property we store the identifier of the file containing the original input text that has been processed. We also use the *rdfs:comment* property in each entity of *Place* type (or any of its derived types) to indicate the annotation tool that recognized the toponym, e.g. OpenCalais in this example.

Graphically, the content annotated in figure 4 could be also seen under the perspective of an RDF graph as shown in figure 5. Moreover, this RDF content can be stored as well in a triple store enabling the construction of a document repository, which can be queried by query languages like SPARQL (SPARQL Protocol and RDF Query Language) (Harris and Seaborne, 2013). SPARQL is a standardized language proposed by W3C for the query of RDF graphs. For instance, figure 6 shows a query to retrieve documents containing the place name *Madrid* in a graph model like the one proposed in figure 5.

```

PREFIX schema:<http://schema.org/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?x ?name ?nameFile
WHERE {
  ?x rdf:Description ?id.
  ?x rdf:type <http://schema.org/TravelAction>.
  ?x schema:url ?nameFile.
  ?x schema:location ?lo.
  ?lo rdf:type <http://schema.org/Place>.
  ?lo schema:name ?name.
  {FILTER regex(?name, 'Madrid')}*
}

```

Figure 6. Example of a SPARQL query for documents containing spatial named entities

This duality of RDFa and RDF is useful for the work proposed in this paper as our purpose is twofold. On the one hand, XHTML-RDFa allows to maintain the original text with the in-line annotations. On the other hand, if more specific queries on documents are required, the storage of RDF triples on a triple-store enables the use of powerful query languages like SPARQL.

#### 4. Infrastructure for a document repository

Figure 7 shows the multilayer architecture of the infrastructure we have designed to build a system able to convert the output of annotation tools into our annotation proposal and, at the same time, feed the contents of a document repository, which can be searched in terms of the spatial entities that have been previously recognized.

In the bottom layer of the architecture we can find the external annotation tools and the RDF triple store based on JENA technology, which gives

support to the document repository providing search functionalities. Additionally, it must be noted that we also store the XHTML-RDFa documents to let users see the annotations produced by the annotation tools at the exact position of the original text.

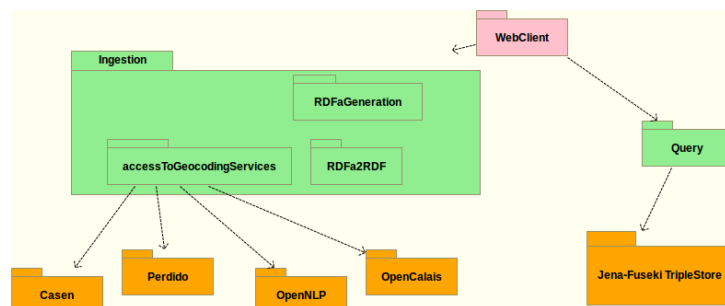


Figure 7. Architecture of the document repository

On top of this bottom layer, there is a business logic layer with the *Ingestion* and *Query* subsystems. The *Ingestion* subsystem enables the access to the external annotation tools, the conversion of their output to our annotation proposal, and the ingestion of documents in the repository. The *Query* subsystem facilitates the services that are necessary to facilitate searches in the document repository using SPARQL.

Last, the upper layer consists of the web application that allows the interaction with the final users for the setup and ingestion of documents in the repository, and the searching of documents according to different spatial filters. The following subsections describe in more detail the *Ingestion* subsystem and the Web application.

##### 4.1. Ingestion subsystem

In order to facilitate the access to the different annotation tools, the *Ingestion* subsystem, developed with Java technology, uses an *AbstractFactory* design pattern to separate the problem of creating components with a standard annotation functionality from the problem of using different access protocols: whereas CasEN and OpenNLP can be integrated as libraries in the source code, PERDIDO and OpenCalais must be accessed as Web services.

Another important aspect in the *Ingestion* subsystem has been the conversion of the outputs of the different annotation tools into our common proposed language. In order to implement the mapping specified in figure 3, specific customizations were required for each tool. In the case of CasEN and OpenCalais, generating plain text as output, it is required to perform a pre-processing to generate

an intermediate XML markup language. Then, this intermediate markup language, as well as the annotation output of PERDIDO (directly available as XML), is transformed into our proposed XHTML-RDFa format using XSLT technology (eXtensible Stylesheet Language Transformations). On the opposite side, in the case of OpenCalais, as this tool generates an RDF annotation, the transformation can be directly programmed using the JENA API for accessing RDF content.

Finally, the *Ingestion* subsystem is also in charge of the transformation of XHTML-RDFa documents into RDF triples for its storage in a triple store (which can be later queried by the *Query* subsystem). This transformation is made thanks to the RDF Translator tool (Stolz, 2014), which is accessible online as a Web service. Figure 8 shows an example of the conversion to RDF of the annotated document in figure 4 (and its graph representation in figure 5).

At this point it must be noted that both XHTML-RDFa annotations and their conversion into RDF maintain a reference to the original document identifier using the *schema:url* property. This facilitates the integration of the input of different annotation tools at the moment of storing them in the same triple-store. Thanks to this, it is possible to query later for place names that have been

identified by different tools on the same document.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:schema="http://schema.org/"
>
  <rdf:Description rdf:nodeID="N6d2e01594db44edf9751591df57c3939">
    <schema:location rdf:nodeID="N6c1e82d03f9d48c1b5cf0db3c918cfc2"/>
    <rdf:type rdf:resource="http://schema.org/TravelAction"/>
    <schema:url>Near_the_Puerta_del_Sol4</schema:url>
  </rdf:Description>
  <rdf:Description rdf:nodeID="N6c1e82d03f9d48c1b5cf0db3c918cfc2">
    <rdf:type rdf:resource="http://schema.org/Place"/>
    <schema:location rdf:nodeID="N351fee839c1247b49be2dec603417deb"/>
    <schema:name>Madrid</schema:name>
    <schema:geo rdf:nodeID="N1bcc407fec644c2088ccd2610148ffe8"/>
    <rdfs:comment>OpenCalais</rdfs:comment>
  </rdf:Description>
  <rdf:Description rdf:nodeID="N1bcc407fec644c2088ccd2610148ffe8">
    <schema:longitude>-3.6833</schema:longitude>
    <schema:latitude>40.4</schema:latitude>
    <rdf:type rdf:resource="http://schema.org/GeoCoordinates"/>
  </rdf:Description>
  <rdf:Description rdf:nodeID="N351fee839c1247b49be2dec603417deb">
    <schema:name>Spain</schema:name>
    <rdf:type rdf:resource="http://schema.org/Country"/>
  </rdf:Description>
</rdf:RDF>
```

Figure 8. Example of RDF triples stored in the semantic document repository

Figure 9. User interface of the Web application

#### 4.2. Web application

The Web application prototype for interacting with the document repository has been developed with AngularJS, Bootstrap and Spring MVC technologies. On the one hand, AngularJS (Google, 2016) is an open source project, based on Ja-

vascript, which contains a set of libraries to facilitate the development of front-ends in web applications. On the other hand, Bootstrap (2016) is an Open Source framework that provides Cascading Style Sheets (CSS) to define the layout of HTML or XML documents (including also XHTML documents). Last, Spring MVC (Pivotal Software, 2016) facilitates the connection between the web

application front-end and the business logic of the *Ingestion* and *Query* subsystems through HTTP GET and POST requests.

Figure 9, in the previous page, shows an overview of the screenshots provided by the web application: an interface for the ingestion of documents or direct typed text (left side of figure 9), and an interface for querying documents in the repository and browse the annotation and toponyms found in each document.

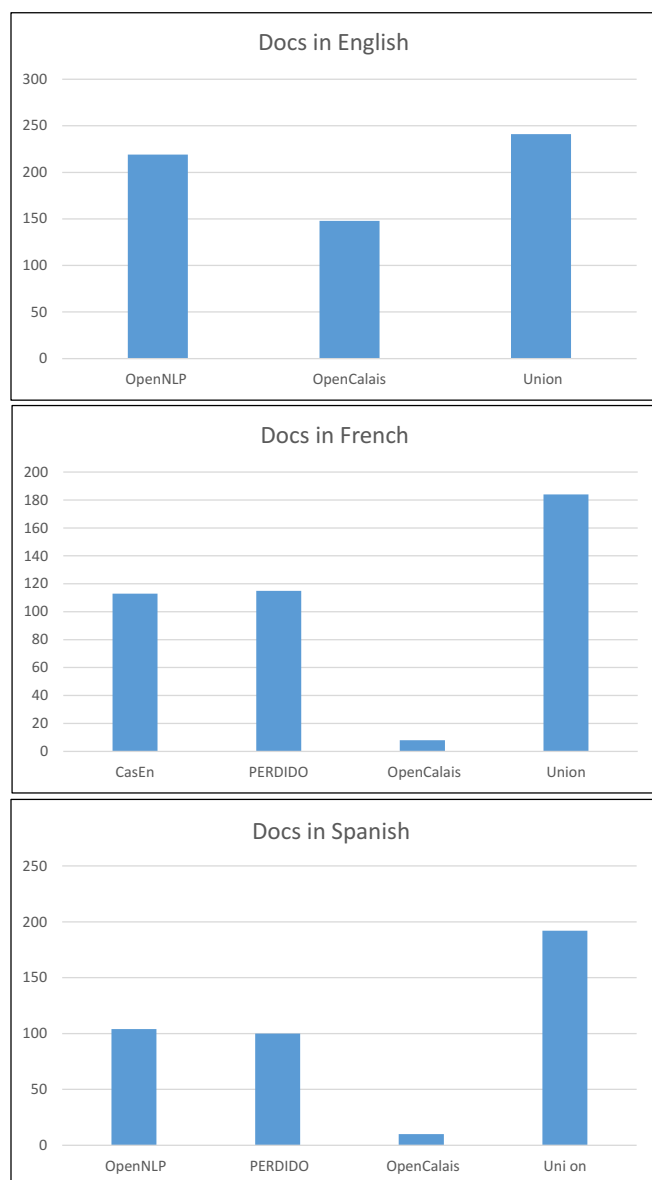


Figure 10. Number of toponyms detected as a result of the ingestion process

## 5. Validation of the approach

In order to validate the infrastructure proposed for a document repository we have uploaded a corpus of documents consisting of: 32 documents in

English selected from a corpus created for the testing of spatial queries in an information retrieval conference (SemEval, 2015); 20 documents in French selected from a Web site of hiking descriptions (Visorando, 2016); and 28 documents in Spanish selected from another Web site with hiking descriptions (Gobierno de Aragón, 2016) (1). Figure 10 shows three bar charts with the results obtained by the ingestion process in each language. Each bar chart shows the number of spatial entities detected by each annotation tool and a final bar with the number of spatial entities detected by all tools excluding duplicates.

Although the results obtained depend greatly on the features of the corpus and the granularity of place names recognized by each tool, it seems that a document repository able to integrate different existing annotation tools provides an exhaustive annotation. In the case of the English experiment, the documents describe cities and countries all over the world. Thus, OpenNLP and OpenCalais, the tools available for processing English texts and focused on the detection of administrative units like cities or countries, provide similar results. In the case of French and Spanish experiments, the documents contain fine-grain toponyms and small to medium size names of administrative units. In these cases, we can see that the annotation provided by PERDIDO annotation tool, and focused on fine-grained toponyms, is an excellent complement to the annotations provided by CasEn or OpenNLP, which are more focused on the detection of names of small/medium administrative units. Additionally, it can be observed that the use of OpenCalais for French and Spanish documents provides little help as there are a very small number of city and country names.

## 6. Conclusions and future work

This work has described an approach to transform the heterogeneous output of different NER tools into a common annotation framework based on the use of RDFa and the vocabulary of Schema.org. In principle, this annotation approach should increase the visibility and recoverability of spatial annotated texts since big web search engine companies have agreed in the use of Schema.org as a de-facto standard for semantic annotation on the Web. There are other more specific application level ontologies for defining spatial concepts such as the ones referred by Lopez-Pellicer et al (2012), but our purpose was to select a common core set of concepts accepted by different communities adopting Semantic Web technologies. Other well-known Web content authoring tools like RDFace (Khalili and Auer,

2015) have also adopted Schema.org for their annotations.

Additionally, this work has also presented the design of the infrastructure for a semantic document repository with a Web interface that facilitates the ingestion of documents and their spatial annotation with the different external NER tools, and the spatial searching of documents based on the annotated spatial entities.

As future lines of this work we plan to study and integrate other existing annotation tools, establishing the mapping to our annotation proposal. Additionally, we would like to improve our annotation approach to take into account other types of named entities like Persons, Organizations or Dates. Furthermore, instead of considering each annotated document as an instance of a *TravelAction*, we would like to provide a finer heuristics to annotate the general intent of the annotated document.

Finally, our goal is also to extend the web application to facilitate the interaction of final users in different aspects: facilitate the ingestion of documents in a wider range of input formats (e.g., PDF, Word, ODF); provide an analytical comparison of the results produced by the different annotation tools; and facilitate the possibility of making a manual revision of these annotations.

## Acknowledgements

This work has been partially funded by Universidad de Zaragoza through Catedra Logisman on Technological Document Management and UZ2016-TEC-05 research project.

## Notes

(1) The documents in Spanish and French form part of the PERDIDO corpus: a collection of documents that are representative of the type of documents available on Web sites of hiking descriptions and have been already manually annotated for the purpose of testing geoparsing algorithms. More details about the corpus can be found at (Moncla, 2015b).

## References

- Apache (2010). The Apache Software Foundation. Apache OpenNLP web site. <https://opennlp.apache.org/> (2016-04-28)
- Bootstrap (2016). Bootstrap web site. <http://getbootstrap.com/> (2016-04-28)
- Cyganiak, R., Wood, D., Lanthaler, M. (eds.) (2014). RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014. <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/> (2016-06-14)
- Doan, A., Halevy, A., Zachary, I. (2012). Principles of Data Integration (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN 9780124160446
- Gobierno de Aragón (2016). Turismo de Aragón. Senderos de Aragón. Excursiones y rutas para GPS. <http://senderos.turismodearagon.com> (2016-04-28)
- Google (2016). AngularJS API Docs. <https://docs.angularjs.org/api> (2016-04-28)
- Harris, S., Seaborne, A. (eds.) (2013). SPARQL 1.1 Query Language. W3C Recommendation 21 March 2013. <http://www.w3.org/TR/2013/REC-sparql11-query-20130321/> (2016-06-14)
- Herman, I., Adida, B., Sporny, M., Birbeck, M. (eds.) (2015). RDFa 1.1 Primer - Third Edition. Rich Structured Data Markup for Web Documents. W3C Working Group Note 17 March 2015 <https://www.w3.org/TR/2015/NOTE-rdfa-primer-20150317/> (2016-06-14)
- Khalili, A., Auer, S. (2015). WYSIWYM-Integrated visualization, exploration and authoring of semantically enriched un-structured content. *Semantic Web*, 6(3), 259-275.
- Leidner, J. L. (2008). *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names*. Universal-Publishers
- Leidner, J. L.; Lieberman, M. D. (2011). Detecting geographical references in the form of placenames and associated spatial natural language. // *SIGSPATIAL Special* 3(2), 5-11.
- López-Pellicer, F. J.; Lacasta, J.; Florczyk, A.; Nogueras-Iso, J.; Zarazaga-Soria, F.J. (2012). An Ontology for the representation of Spatio-Temporal Jurisdictional Domains in Information Retrieval Systems. // *Int. Journal of Geographical Information Science*. 26:4, 579-597.
- Maurel, D.; Friburger, N. (2004). Finite-state transducer cascades to extract named entities in texts. *Theoretical Computer Science*. 313:1, 93-104.
- Moncla, L. (2015a). PERDIDO REST API Specifications. <http://erig.univ-pau.fr/PERDIDO/api.jsp> (2016-04-28)
- Moncla, L. (2015b). Automatic reconstruction of itineraries from descriptive texts. PhD Dissertation, Université de Pau et des Pays de l'Adour & Universidad de Zaragoza. <https://zaguan.unizar.es/record/47425/> (2016-06-16)
- Moncla, L.; Gaio, M.; Mustière, S. (2014). Automatic Itinerary Reconstruction from Texts. Proc. 8th International Conference, GIScience 2014, 253-267.
- Pivotal Software (2016). Spring Framework. <http://projects.spring.io/spring-framework/> (2016-04-28)
- Santos, J.; Anastacio, I.; Martins, B (2014). Using machine learning methods for disambiguating place references in textual documents. // *GeoJournal*. 80:3, 375-392.
- Sekine, S.; Ranchhod, E. (eds.) (2009). *Named Entities: Recognition, Classification and Use*. John Benjamins. ISBN 9789027289223.
- SemEval (2015). International Workshop on Semantic Evaluation (SemEval-2015): Task 8-SpaceEval Test Data. <http://alt.qcri.org/semeval2015/task8/index.php?id=data-and-tools> (2016-04-28).
- Stolz, A. (2014). RDF Translator. <http://rdf-translator.appspot.com/> (2016-04-28).
- Thomson Reuters (2016). Open Calais web site. <http://www.opencalais.com/> (2016-04-28).
- University of Tours (2016) CasEN web site. [http://tln.li.univ-tours.fr/Tln\\_CasEN.html](http://tln.li.univ-tours.fr/Tln_CasEN.html) (2016-04-28).
- Visorando (2016). Visorando web site. <http://www.visorando.com> (2016-04-28).

Enviado: 2016-0-28. Segunda versión: 2016-0-  
Aceptado: 2016-0-

---

# A estrutura lógico-hierárquica de linguagens de indexação utilizadas por bibliotecas universitárias

*La estructura lógica y jerárquica de lenguajes de indización utilizados por las bibliotecas universitarias*

*The logical and hierarchical structure of indexing languages used by university libraries*

---

**Mariângela Spotti Lopes FUJITA (1), Luciana Beatriz Piovezan dos SANTOS (2)**

(1) Universidade Estadual Paulista Júlio de Mesquita Filho, Av. Hygino Muzzi Filho, 737, Campus Universitário, CEP. 17525-900, Marília, SP, Brasil, fujita@marilia.unesp.br, (2) Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista, lbpiovezan@gmail.com

## Resumen

El uso de lenguajes de indización realiza con precisión la mediación entre los vocabularios de los documentos y los de los usuarios durante la recuperación de información. Dichos lenguajes, sin embargo, deben ser actualizados a medida que avanza la generación e innovación del conocimiento. Dicha actualización depende de la estructura jerárquica para facilitar su inclusión y clasificación. Para la investigación propuesta sobre las estructuras jerárquicas de lenguajes de indización se adoptó el método del comité ascendente propuesto en la Directiva ANSI/NISO de 2005 para construir vocabularios controlados usando el modelo metodológico integral, comparando las jerarquías de las Library of Congress Subject Headings (LCSH), la Terminología de Assuntos da Biblioteca Nacional (TBN), el Linguagem BIBLIODATA y el Vocabulário de la Universidad de São Paulo en las áreas de física y matemáticas. Los resultados muestran que el lenguaje Bibliodata no posee algunas relaciones jerárquicas, lo que impide su comparación. El TBN tiene ocho niveles de jerarquía para las matemáticas; y los demás lenguajes tienen jerarquías de cuatro niveles para la física. Se concluye que es posible el desarrollo de jerarquías utilizando los métodos propuestos.

**Palabras clave:** Lenguajes de indización. Bibliotecas universitarias. Relaciones jerárquicas. Estructuras categoriales. Actualización. Brasil.

## 1. Introdução

A linguagem de indexação, como ferramenta de conversão, deve ser utilizada tanto pelo indexador após a atribuição de conceitos extraídos do documento quanto pelo usuário antes da atribuição de conceitos extraídos da necessidade de informação que orientará a estratégia de busca no catálogo da biblioteca. Ocorre que usuários do catálogo bibliográfico *on-line* ATHENA (2) não conhecem e nunca utilizaram linguagem de indexação por não estar disponibilizada na web junto à ferramenta de busca. Nunca existiu, então, a mediação da linguagem no catálogo ATHENA e,

## Abstract

The use of indexing languages mediates accurately between document and user vocabularies during information retrieval. Nevertheless, it should be updated accordingly to the generation and innovation of new knowledge. Updating terms depends on the hierarchical structure that guides the inclusion and classification of terms. With the research proposal about hierarchical structures of indexing languages, the bottom-up committee method proposed in the ANSI/NISO-2005 Guidelines for the compilation of controlled vocabularies using the integrated methodological model was adopted, comparing the hierarchies of the Library of Congress Subject Headings (LCSH), the Subject Terminology of National Library (TBN), the BIBLIODATA Language and the USP Vocabulary in the fields of Physics and Mathematics. The results show that BIBLIODATA language does not have certain hierarchical relationships, making any comparison impossible. TBN has eight levels hierarchy for Mathematics and the other languages have four levels for Physics. The conclusion is that it is possible to develop hierarchies using these methods.

**Keywords:** Indexing languages. University libraries. Hierarchical relations. Categorical structures. Updating. Brasil.

em realidade, não precisaria ser utilizada pelos catalogadores, pois a linguagem natural é a mais utilizada pelos usuários como demonstraram os resultados obtidos no primeiro estudo de avaliação realizado por Boccato, Fujita e Gil Leiva (2011), que verificou o desempenho e o índice de precisão na recuperação de áreas científicas especializadas em 3 áreas de conhecimento no catálogo *on-line* ATHENA do sistema de bibliotecas da UNESP. A análise dos dados coletados utilizou, como indicadores de avaliação da linguagem de indexação, os índices de precisão e revocação pela recuperação da informação no catálogo Athena (Lancaster, 2004; Gil Leiva, 2008).

Essa constatação pode ser a mesma de outras bibliotecas que necessitam aprimorar a recuperação por assuntos de seus catálogos *on-line* e isso poderá ser resolvido pela escolha de uma linguagem de indexação. A decisão de usar uma só linguagem deve ser tomada em função da compatibilização de vocabulário a ser utilizado na indexação e na estratégia de busca para obter resultados de recuperação mais precisos, corretos e específicos.

O uso de uma linguagem de indexação adequada e atualizada para o contexto de contínua geração de novos conhecimentos e de inovação, no caso das bibliotecas da UNESP, foi determinante para a construção da Linguagem de indexação UNESP mediante processo de adequação e compatibilização de vocabulários de três linguagens de indexação de mesma raiz terminológica utilizadas por bibliotecas: Terminologia de assuntos da Biblioteca Nacional, *Library of Congress Subject Headings* (LCSH) e *Medical Subject Headings* (MeSH). O avanço desse processo de adequação e compatibilização exige, contudo, a incorporação de termos novos conforme previsto inicialmente pela própria necessidade do contexto de geração de novos conhecimentos científicos e tecnológicos das bibliotecas universitárias.

Ocorre que a criação de termos novos depende da existência de uma macroestrutura de categorias temáticas que oriente a inclusão e classificação dos termos para que se componha a estrutura lógica e hierárquica dos termos superordenados e subordinados.

Essa macroestrutura para organizar a ordem hierárquica dos assuntos ainda não existe na Linguagem de indexação UNESP e precisa ser criada para orientar a adequação, a compatibilização e a inclusão de termos novos pela organização de registros de autoridade. Aparentemente, a Linguagem da Biblioteca Nacional (BN) e a Linguagem da *Library of Congress Subject Headings* (LCSH) não possuem uma macroestrutura de categorias temáticas que organiza a ordem hierárquica dos assuntos, ainda que, todos os termos possuam uma organização hierárquica.

Considerando-se a observação desse problema de estruturação lógica e hierárquica da Linguagem de indexação UNESP, é importante o desenvolvimento de investigação sobre o processo de elaboração da estrutura lógica e hierárquica de linguagens de indexação na perspectiva de uma política de indexação. Para isso, propõe-se o estudo da macroestrutura de categorias da linguagem LCSH, Terminologia da Biblioteca Nacional, Vocabulário USP e Linguagem do BIBLIODATA para elaboração e desenvolvimento dessa

linguagem de indexação para bibliotecas universitárias.

## 2. Linguagens de indexação utilizadas por bibliotecas: aspectos históricos e de estrutura lógico-hierárquica, semântica e associativa

Para avançarmos no que tange ao uso da linguagem de indexação em bibliotecas examinaremos este instrumento a partir de sua concepção como Lista de Cabeçalhos de Assuntos e sua construção.

Veremos, então, que as listas de cabeçalhos de assunto são linguagens de indexação pelos aspectos estruturais e que, por isso, evoluíram e incorporaram aspectos de tesouros, tendo em vista as grandes mudanças de acesso e de recuperação da informação em sistemas de buscas por assunto.

Tal como o tesouro, as listas de cabeçalhos de assuntos são linguagens de estrutura alfabética, considerando-se que os termos que as integram são ordenados alfabeticamente. Tanto o tesouro como as listas de cabeçalhos de assuntos, possuem entre os termos relações hierárquicas, semânticas e associativas como se observa no Quadro I:

Descrição	Horticultura (subdividido geograficamente)
Remissiva Ver (US/UF)	Hortaliças - Cultivo Hortas
Remissiva Ver Também (TR)	TE: Aclimação (Plantas) TE: Cobertura dos solos TE: Cogumelos - Cultivo TE: Estufas para cultivo TE: Frutas - Cultivo TE: Hidroponia TE: Jardinagem paisagística TE: Olericultura TE: Plantas - Propagação TE: Plantio (Cultivo de plantas) TE: Poda TE: Produtos hortigranjeiros TG: Agricultura TR: Jardinagem
Fonte positiva dos dados	LCSH
Nota geral pública	Usado para obras que tratam do cultivo, para fins comerciais, de frutas, vegetais e plantas ornamentais
Outros vocabulários	
Library of Congress	Horticultura

Quadro I. Ficha da autoridade – Termo tópico “Horticultura” da TBN

O termo “Horticultura” tem relação semântica com os termos “Hortaliças-Cultivo” e “Hortas” como termo preferido ou de equivalência, relação hierárquica como termo superordenado com os termos específicos (TE) subordinados a partir do termo “Aclimação (Plantas)” até o termo “Produtos Hortigranjeiros” e como termo subordinado



com o termo genérico (TG) “Agricultura” e, por último, relação associativa (TR) com o termo “Jardinagem”.

As relações hierárquicas entre termos superordenados e subordinados demonstram a estrutura vertical de relações lógico-hierárquicas de uma linguagem de indexação que se constitui de vértices e níveis. O vértice das cadeias hierárquicas em sistemas de classificação, segundo explica Cintra et al. (1994, p. 32), é “constituído de disciplinas convencionais que se subdividem sucessivamente”. Nos tesouros e listas de cabeçalhos de assunto, os vértices equivalem às classes e são denominados de top terms ou termos genéricos superordenados que poderão ser o gênero ou o todo.

No exemplo do Quadro II, o *top term* da cadeia hierárquica é “Agricultura”, que possui 3 níveis constituídos de 3 vértices:

. AGRICULTURA
.. Horticultura
... Aclimação (Plantas)
... Cobertura dos solos
[...]
... Produtos hortigranjeiros

Quadro II. Vértices e níveis da estrutura lógico-hierárquica do termo “Horticultura” na TBN

Para a construção, atualização e manutenção de linguagens de indexação é fundamental o conhecimento e uso da estrutura lógico-hierárquica representada pelas cadeias hierárquicas de termos superordenados e subordinados compostas de vértices e níveis com a finalidade de representação do domínio de assunto de modo completo.

Para o entendimento das linguagens de indexação utilizadas por bibliotecas passamos a examinar as listas de cabeçalhos de assunto que são continuamente atualizadas e estão disponíveis para uso.

### 2.1. A linguagem da LCSH: história e configuração lógico-hierárquica, semântica e associativa

A *Library of Congress Subject Headings* (Lista de Cabeçalho de assunto da Biblioteca do Congresso) começou a ser construída em 1898 para atender às necessidades de representação dos conteúdos de livros e periódicos custodiados pela Biblioteca do Congresso dos Estados Unidos (*Library of Congress* - LC). Inicialmente intitulada *Subject Headings in the Dictionary Catalogues of the Library of Congress*, o título atual foi

adotado na oitava edição da lista, publicada em 1975 (Hoshy, 2001, p. 135-136).

Torres (2006) ressalta o papel de destaque da LC na disseminação da catalogação de assuntos e das listas de cabeçalhos de assuntos, que, desde sua criação e com a comercialização de seus registros bibliográficos e de autoridade, sedimentara o uso do cabeçalho de assunto na representação do conteúdo de documentos.

Deste modo, a LCSH tem sido amplamente utilizada em bibliotecas para promoção do acesso por assunto, tendo sido traduzida e adaptada para diversos idiomas e servido como base para construção de outras linguagens de indexação.

Considerada como a principal lista de cabeçalhos de assunto existente, a LCSH, embora seja originalmente uma linguagem pré-coordenada, estão hoje organizados em formato de tesouro, com seus cabeçalhos arranjados em uma estrutura de relacionamentos de equivalência, hierárquicos e associativos (Fujita, Boccato e Rubi, 2010).

No âmbito da LCSH, os cabeçalhos de assuntos são construídos a partir da criação de um registro de autoridade de assunto em formato *Machine Readable Cataloging* - MARC21, que será incorporado à lista de autoridades que compõe a LCSH (Hoshy, 2001).

O Registro de autoridades de assunto da LCSH é composto por notas de escopo e indicações de relacionamentos, tendo, como objetivos básicos, prover o cabeçalho de assunto autorizado ou direcionar ao cabeçalho autorizado, indicar se um cabeçalho pode ou não ser subdividido geograficamente, prover uma nota de escopo quando houver necessidade de explicar a cobertura a que o termo se aplica e indicar relacionamentos entre cabeçalhos.

Como dito, a LCSH está organizada em uma estrutura de relacionamentos de equivalência, hierárquicos e associativos, que permitem aos usuários ampliar, especificar ou redirecionar suas buscas. Tais relacionamentos e suas designações são padronizados para construção e gerenciamento de tesouros (Hoshy, 2001).

Desse modo, a estrutura da LCSH compreende uma ordenação sistemática que evidencia as relações que unem os cabeçalhos de assunto que a compõem. Nessa estrutura, as relações hierárquicas, expressas pelo código BT para termo geral e NT para termo específico, são feitas para indicar que há uma relação de superordenação/subordinação entre dois ou mais cabeçalhos (Yi e Chan, 2010).

Nesse tipo de relacionamento, há a possibilidade de um cabeçalho pertencer a mais de uma hierarquia lógica, casos em que serão feitas referências a cada hierarquia à qual o cabeçalho se liga, respeitando-se a ligação apenas ao termo imediatamente superior em cada hierarquia (Hoshy, 2001).

O relacionamento de equivalência, expresso pelos códigos USE e UF (usado para), existe entre termos não-preferidos remetendo para um cabeçalho autorizado, quando dois ou mais termos designam o mesmo conceito, de maneira a atender ao princípio do cabeçalho uniforme e controlar variantes semânticas e ortográficas (Hoshy, 2001).

O relacionamento associativo, expresso pelo código RT (termo relacionado) indica termos que estão ligados de alguma forma diferente das anteriormente expressas, de modo que se relacionem cabeçalhos relativos a assuntos entre os quais há uma associação mental.

## 2.2. Linguagens derivadas da LCSH: terminologia da Biblioteca Nacional e linguagem BIBLIODATA em comparação com a LCSH

A Lista de Cabeçalhos de Assunto da Rede BIBLIODATA começou a ser desenvolvida no âmbito do projeto BIBLIODATA/CALCO (Catalogação Legível por Computador), um esforço conjunto da Fundação Getúlio Vargas (FGV) e do Instituto Brasileiro de Bibliografia e Documentação (IBBD), atual IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia) com o objetivo de padronizar a utilização de cabeçalhos de assunto (Histórico, [2016]).

A Linguagem BIBLIODATA é disponibilizada eletronicamente a todas as instituições credenciadas pela Rede BIBLIODATA acompanhada de um manual de procedimentos que contém a sua estrutura formal e instruções de operacionalização da linguagem.

Verifica-se que a Linguagem BIBLIODATA é uma linguagem controlada e pré-coordenada, cujos termos são arranjados alfabeticamente, composta por meio de tradução e adaptação para a língua portuguesa da LCSH. De acordo com o Manual de cabeçalhos de assunto da Rede BIBLIODATA, a opção pela LCSH deu-se por sua característica multidisciplinar, além de sua confiabilidade (Fundação Getúlio Vargas, 1995).

A lista de cabeçalhos da Linguagem BIBLIODATA é composta por cabeçalhos tópicos e subdivisões de assunto, de forma, geográficas e cronológicas, remissivas (relações de equivalência) e relacionamentos hierárquicos e associativos. O registro de autoridade de assunto da Linguagem

BIBLIODATA compõe-se pelo cabeçalho, relacionamentos, nota de escopo e termo LC – em que se indica o termo correspondente na LCSH (Fundação Getúlio Vargas, 1995).

A Terminologia da Biblioteca Nacional (TBN) remonta a 1945, quando a instituição passou a adotar os princípios da *American Library Association* (ALA) para tratamento da informação, criando catálogos de autoridades de nomes pessoais e coletivos e, conjuntamente, passou a adotar a LCSH como padrão para criação de pontos de acesso por assunto para uso interno da instituição. Foi apenas no ano de 1982, com sua entrada na Rede BIBLIODATA/CALCO que a instituição passou a construir sua própria linguagem baseada na LCSH e cujo uso “foi sistematizado não só em âmbito interno como para utilização pelas bibliotecas participantes da Rede BIBLIODATA, mantendo-se os cabeçalhos originais e gerando, assim, um vocabulário bilíngue” (GRINGS, 2016, no prelo).

Em 1994, quando a FGV optou por uma mudança de *software* operacional para a Rede BIBLIODATA, a BN deixou de participar da Rede e ambas as linguagens passaram a ser desenvolvidas separadamente, mantendo como semelhança apenas o fato de possuírem a LCSH como matriz e base terminológica (GRINGS, 2016, no prelo).

## 2.3. Linguagem de indexação de bibliotecas universitárias: o Vocabulário USP

O Vocabulário USP é uma linguagem construída para a indexação de assuntos da coleção das bibliotecas universitárias da USP em seu Banco de Dados Bibliográficos DEDALUS. Para atender às necessidades relativas às atividades inerentes à universidade, o Vocabulário USP constitui-se de uma linguagem interdisciplinar, que abrange as diversas áreas do conhecimento (Universidade de São Paulo, 2015).

A criação do Vocabulário USP iniciou-se com a composição, em 1993, de uma equipe de bibliotecários da USP, responsável por atualizar e ampliar a Lista de Assuntos que era utilizada até o momento. A equipe de bibliotecários contou com o apoio de professores da Escola de Biblioteconomia da USP e também da UNESP (Universidade de São Paulo, [s.d.]).

Como resultado, em 2003 foi implementado o Vocabulário USP, disponível para consulta *on-line* na página web da universidade. O vocabulário constitui-se de descritores, não-descritores, qualificadores, remissivas e termos denominados como “elos falsos”, que são termos não-descritores que atuam como elemento aglutinador de



descritores mais específicos (Universidade de São Paulo, 2006).

O Vocabulário USP está disponível para consulta externa, podendo ser visualizado por sua apresentação em lista alfabética ou hierárquica, complementada por tabelas de qualificadores, além de poder ser consultado por sua macroestrutura, em que são explicitadas suas relações lógico-semânticas. Estima-se que atualmente o Vocabulário USP contenha aproximadamente 45 mil termos (Universidade de São Paulo, 2015).

Para uso interno, o vocabulário está disponibilizado na base de dados SIBIX 650 integrada ao sistema de gerenciamento para uso na indexação de itens do acervo. O Vocabulário também pode ser utilizado por outras instituições por meio de consulta à página web institucional (Universidade de São Paulo, 2006; Universidade de São Paulo, 2015).

As quatro linguagens de indexação apresentadas foram construídas para uso em bibliotecas, sendo que apenas o Vocabulário USP tem a especificidade de haver sido desenvolvido para bibliotecas universitárias; as demais são utilizadas em diversas modalidades de bibliotecas indistintamente. Tais linguagens apresentam estrutura alfabética de organização de termos, sendo que três são linguagens pré-coordenadas (LCSH, TBN e BIBLIODATA) e uma linguagem é pós-coordenada, o Vocabulário USP. Duas linguagens, LCSH e BIBLIODATA, são denominadas como listas de cabeçalhos de assunto, mas a primeira adota características de tesauro para organização de seus termos, da mesma forma que a TBN, que segue a estrutura e a terminologia da LCSH. As quatro linguagens apresentam complementos para os termos incluídos, como listas de qualificadores, subdivisões e remissivas.

### **3. O estudo das macroestruturas das linguagens de indexação: método e procedimentos**

A construção de linguagens de indexação possui procedimentos específicos abordados por normatizações e literatura da área. As principais diretrizes, em diferentes momentos, abordaram a construção de vocabulários controlados. São elas: a do IBICT (1984), da UNESCO (1996) e da ANSI/NISO (2005) cujas etapas podem ser observadas e analisadas a partir de síntese elaborada por Cervantes (2009, p. 200) no quadro “Etapas de construção de um tesauro de acordo com as diretrizes”, que contém 9 etapas: Trabalho preliminar, métodos de compilação, Registro de termos, Verificação dos termos, Determinação da estrutura e formatos de apresentação, Níveis de Especificidade, Inclusão e Exclusão de

termos, Uso de equipamento automático de processamento de dados (IBICT e UNESCO), Termos não designados (ANSI/NISO).

Conforme análise mais detalhada das etapas no quadro de Cervantes (2009, p. 200), verificou-se que é na etapa de “Métodos de compilação” que as três Diretrizes expõem a construção das classes mais amplas, denominadas de categorias a partir das quais se constroem as hierarquias de termos genéricos e específicos. As Diretrizes do IBICT e da UNESCO dividem os métodos em dedutivo, indutivo e combinação de métodos e a ANSI/NISO amplia para 4 os tipos de métodos, sendo abordagem de comitê, abordagem empírica, combinação de métodos e assistência da máquina.

Os métodos indutivo e dedutivo são considerados pela Diretriz ANSI/NISO (2005) como empíricos e estão incluídos em seu método de abordagem empírica. De forma geral, são descritos com conteúdo semelhante nas três diretrizes com enfoque para a orientação de inclusão de termos em categorias que devem ser identificadas pelos termos genéricos de primeiro nível hierárquico antes (indutivo) ou depois (dedutivo) da compilação dos termos. A combinação de métodos indutivo e dedutivo propõe método que incorpora o dedutivo e o indutivo de modo que sugere o estabelecimento das categorias dos termos de modo indutivo por um grupo editorial composto por indexadores e especialistas e depois de as hierarquias de termos serem examinadas a partir da perspectiva dedutiva.

A abordagem de comitê, sugerida pela diretriz ANSI/NISO (2005), propõe a atuação de um comitê composto por especialistas no domínio dos assuntos e especialistas em criação de vocabulários e identifica dois métodos para a atuação desse comitê: Top-down (descendente) e Bottom-up (ascendente). No descendente, os termos mais genéricos ou superordenados são identificados inicialmente e depois são construídas as hierarquias de termos subordinados ou mais específicos para atingir o nível de profundidade hierárquica necessário. No ascendente, é mais característica a situação de vocabulários controlados já iniciados que precisam incorporar termos derivados de novos conteúdos e, nesse caso, a estrutura de categorias e as relações hierárquicas vão sendo construídas ao contrário: iniciam-se pelos termos mais específicos e avançam para os termos superordenados. Desse modo, o método descendente é mais indicado para a construção de novos vocabulários que resultarão em linguagens de indexação e o ascendente, para linguagens de indexação já publicadas e que necessitam do acréscimo de novos termos para sua atualização e manutenção.

Portanto, para a situação desta proposta de acréscimo de termos novos combinados com a necessidade de criar uma estrutura de categorias e suas hierarquias para o conjunto de termos já existentes da Linguagem de indexação da UNESP, será adotado para a compilação o método de comitê ascendente, considerando-se que os termos específicos já existem e as hierarquias de termos necessitam da identificação das categorias.

Para a identificação das categorias serão analisadas as estruturas de categorias das linguagens de indexação da LCSH, da Terminologia de Assuntos da Biblioteca Nacional, do BIBLIODATA e

do Vocabulário USP utilizando o “Modelo Metodológico Integrado para construção de Tesouro” de Cervantes (2009, p. 164).

Observa-se que na coluna da direita estão os procedimentos terminográficos para cada etapa da construção do tesouro (coluna da esquerda) que aqui generalizamos para linguagem de indexação. Nosso enfoque será a aplicação dos procedimentos terminográficos propostos para o método de compilação de abordagem de comitê ascendente com a finalidade de construir é uma estrutura de categorias para a Linguagem UNESP e, para isso, definimos o desenvolvimento de suas etapas e fases:

---

*Modelo metodológico integrado para construção de tesouro*

---

Sistematização de etapas da construção de tesouros (normalização, literatura e tesouros) -  
Procedimentos terminográficos

---

1. Trabalho preliminar (Orientações gerais/ Uso de equipamento automático de processamento de dados)	Escolha do domínio e da língua do tesouro Delimitação do subdomínio Estabelecimento dos limites da pesquisa terminológica temática; Consulta a especialista do domínio/subdomínio
2. Método de compilação (Abordagem de compilação)	Coleta do corpus do trabalho terminológico Estabelecimento da árvore de domínio Expansão da representação do domínio escolhido
3. Registro de termos	Coleta e classificação de termos.
4. Verificação de termos (Admissão e exclusão de termos /Especificidade)	Verificação, classificação e confirmação de termos Elaboração de definições Uso do vocabulário de especialidade para o estabelecimento de relações entre os descritores e de relações entre descritores e não descritores Organização das relações entre descritores
5. Forma de apresentação de um tesouro	Trabalhos de apresentação do tesouro.

---

*Quadro III. Sistematização das etapas de construção de tesouros*

6. Coleta do *corpus* do trabalho terminológico

- Fase 1: identificação e descrição das linguagens de indexação selecionadas para análise do domínio da Física e da Matemática conforme apresentado nos itens 2.1, 2.2 e 2.3;
- Fase 2: identificação e descrição das estruturas de representação documental das áreas de Física e da Matemática nas linguagens de indexação selecionadas;

7. Estabelecimento da árvore de domínio

- Fase 3: identificação e descrição das hierarquias de termos dos domínios de Física e Matemática das linguagens de indexação;

8. Expansão da representação de domínio

- Fase 4: elaboração da estrutura lógico-hierárquica do domínio da Física e da Matemática, tendo, como parâmetro, as estruturas das linguagens de indexação.

#### **4. Análise e discussão sobre o estudo das macroestruturas das linguagens de indexação**

---

Neste tópico será abordada a aplicação do “Modelo Metodológico Integrado para construção de Tesouro”, proposto por Cervantes (2009) relativamente à etapa Método de compilação, descrevendo o desenvolvimento das fases 2, 3 e 4.

#### 4.1. Coleta do *corpus* do trabalho terminológico

*Fase 2: Identificação e descrição das estruturas de representação documentária das áreas de Física e da Matemática nas linguagens de indexação selecionadas*

Foram analisadas as linguagens de indexação identificadas previamente de modo a compilar e estruturar os termos relativos às áreas de Física e Matemática, agrupados em uma planilha, com o objetivo de identificar a sua estrutura terminológica, visto que apenas o Vocabulário USP disponibiliza de forma estruturada em sua macroestrutura. Por meio da identificação dos termos de primeiro nível hierárquico em ambas as áreas temáticas, discriminaram-se as categorias de assunto de cada linguagem analisada.

Na LCSH, a categoria da área de Física é indicada pelo termo *Physics*, ao qual estão agregados 67 termos imediatamente subordinados, e a categoria da área de Matemática é indicada pelo termo *Mathematics*, ao qual estão relacionados 77 termos imediatamente subordinados.

A TBN identifica a área da Matemática pelo termo Matemática, ao qual foram relacionados 45 termos subordinados, já o termo Física identifica a área de Física, ao qual são subordinados 47 termos.

No Vocabulário USP, a área de Física é identificada pela categoria CE520 – FÍSICA, posteriormente subdividida em 20 termos genéricos, e a área da Matemática é identificada pela categoria CE550 MATEMÁTICA, subdividida em 65 termos genéricos.

Diante do identificado nesta fase, em que se verificou que as três linguagens analisadas indicam as categorias de Física e de Matemática por estes termos, sugere-se que a Linguagem UNESP adote esta nomenclatura como denominação das categorias aqui tratadas.

#### 4.2. Estabelecimento da árvore de domínio

*Fase 3: identificação e descrição das hierarquias de termos dos domínios de Física e Matemática das linguagens de indexação*

A partir do agrupamento dos termos relativos às áreas de Física e Matemática existentes nas linguagens de indexação analisadas, foi possível o estabelecimento de uma estrutura proveniente das categorias principais, identificadas anteriormente.

A estruturação dos termos em uma linguagem tem por objetivo a caracterização do domínio e se faz por meio da classificação dos termos nas

categorias e indicação dos relacionamentos entre estes termos (Fujita, 2011).

Observando-se a estrutura hierárquica das linguagens de indexação, verificou-se que na LCSH foi possível encontrar termos subordinados que avançaram até o quarto nível hierárquico para ambas as categorias analisadas.

LCSH	
TG Physics	TG Mathematics
TE Agricultural physics	TE Algebra
TE Soil physics	TE Axioms
TE Soil absorption and adsorption	TE Axiomatic set theory
TE Soil colloids	TE Axiom of choice
TE Archaeological physics	TE Banach-Tarski paradox

Quadro IV. Exemplo da hierarquia das categorias Física e Matemática na LCSH

Exemplo disso é o termo “*Soil colloids*”, que é um termo subordinado de quarto nível hierárquico, cujo termo imediatamente superordenado é “*Soil absorption and adsorption*”, que, por sua vez, é específico de “*Soil physics*”, todos pertencentes à hierarquia “*Agricultural physics*” na categoria Física. Um exemplo da categoria Matemática é “*Banach-Tarski paradox*” também subordinado de quarto nível hierárquico, que pertence à hierarquia de “*Axioms*”, completado pelos termos “*Axiomatic set theory*” e “*Axiom of choice*”.

Na Terminologia da Biblioteca Nacional, foram identificados termos subordinados que avançaram até o quarto nível hierárquico para a área de Física e até o oitavo nível hierárquico para a área de Matemática.

Na categoria Física, o termo “Articulações – Hipermobilidade” demonstra essa afirmação. Observa-se que é um termo subordinado de quarto nível hierárquico que se relaciona imediatamente ao termo superordenado “Articulações - Amplitude de movimento”, por sua vez termo específico de “Biomecânica” que completa a hierarquia do termo “Biofísica”. Para a categoria de Matemática, tem-se o exemplo da hierarquia do termo “Cálculo vetorial”, composta por oito níveis hierárquicos demonstrados nos termos subordinados “Cálculo tensorial”, “Espaços generalizados”, “Banach, Espaços de”, “Banach, Álgebra de”, “Análise harmônica”, “Harmônicos esféricos” e do termo “Grassmann, Teoria da extensão de”.

TBN	
TG Física	TG Matemática
TE Biofísica	TE Cálculo vetorial
TE Biomecânica	TE Cálculo tensorial
TE Articulações – Amplitude de movimento	TE Espaços generalizados
TE Articulações - Hipermobilidade	TE Banach, Espaços de
TE Calor	TE Banach, Álgebra de
TE Calor animal	TE Análise harmônica
TE Controle de temperatura	TE Harmônicos esféricos
TE Capilaridade	TE Grassmann, Teoria da extensão de

Quadro V. Exemplo da hierarquia das categorias Física e Matemática na TBN

A análise do Vocabulário USP identificou termos subordinados que avançaram até o quinto nível hierárquico na categoria Física e até o quarto nível na categoria referente à Matemática.

Vocabulário USP	
CE520 FÍSICA	CE550 - MATEMÁTICA
CE520.12 - MECÂNICA DOS LÍQUIDOS	CE550.1 – ÁLGEBRA
CE520.12.1 - HIDRÁULICA (MECÂNICA DOS LÍQUIDOS)	550.16 - ANÁLISE FUNCIONAL
CE520.12.2 - HIDROMECAÂNICA	CE550.16.51 - OPERADORES
CE520.12.2.1 - HIDROSTÁTICA	CE550.16.51.17 - OPERADORES LINEARES
CE520.12.2.1.1 - DENSIDADE DOS LÍQUIDOS	CE550.16.51.17.2 - C-SEMIGRUPOS

Quadro VI. Exemplo da hierarquia das categorias Física e Matemática no Vocabulário USP

Na categoria CE520 FÍSICA, tal estrutura pode ser vista no termo “CE520.12.2.1.1 - DENSIDADE DOS LÍQUIDOS”, termo mais específico subordinado à “CE520.12.2.1 – HIDROSTÁTICA”, que é um termo subordinado à “CE520.12.2 - HIDROMECAÂNICA”, cujo termo genérico é “CE520.12.1 - HIDRÁULICA (MECÂNICA DOS LÍQUIDOS)”. Já na categoria “CE550 - MATEMÁTICA”, é possível demonstrar a hierarquia do termo “CE550.16 - ANÁLISE FUNCIONAL”, ao qual o termo “CE550.16.51 - OPERADORES” é subordinado, chegando ao termo “CE550.16.51.17 - OPERADORES LINEARES”, ao qual o termo “CE550.16.51.17.2 - C-SEMIGRUPOS” se relaciona no quarto nível de profundidade.

Verificou-se que a linguagem da Rede BIBLIODATA não possui relacionamentos hierárquicos determinados, constituindo-se efetivamente como uma lista de cabeçalhos de assunto e, portanto, não há como compará-la com as demais

linguagens que possuem essa estrutura hierárquica estabelecida. Dessa forma, de modo a garantir a consistência e adequação da Macroestrutura da Linguagem Unesp, não é indicada a utilização da linguagem da Rede BIBLIODATA neste processo.

#### 4.3. Expansão da representação de domínio

*Fase 4: Elaboração da estrutura lógico-hierárquica do domínio da Física e Matemática tendo como parâmetro as estruturas das linguagens de indexação*

Com as comparações expostas nos Quadros 6 (Física) e 7 (Matemática) em Apêndices na versão eletrônica, é possível observar que, entre as três estruturas, há poucas semelhanças entre os vértices por causa da estrutura do Vocabulário USP, porém entre as estruturas da LCSH e TBN existem várias semelhanças. Nas três estruturas de Física, existe semelhança, como, por exemplo, com o vértice “Física médica” e na Matemática com o vértice “Álgebra”. Com relação às diferenças entre vértices, elas existem somente na LCSH, como, por exemplo, “Física Agrícola” e “Física Arqueológica”, termos genéricos que demonstram a interdisciplinaridade científica de domínios presentes na linguagem. Na Matemática, “Geometria diferencial”, “Geometria diferencial clássica” e “Geometria não euclidiana” são exemplos de vértices que existem somente no Vocabulário USP.

A proposta para elaboração de estruturas lógico-hierárquicas para expansão da representação dos domínios de Física e Matemática será, portanto, a de incluir as semelhanças e as diferenças entre as três estruturas para que a exaustividade e a especificidade sejam contempladas de modo a garantir profundidade no permanente desenvolvimento da linguagem de indexação que se pretende para bibliotecas universitárias conforme Quadro 8 (Apêndice na versão eletrônica).

#### 5. Considerações finais

Após a análise das estruturas de categorias e das hierarquias lógicas das linguagens de indexação, verificou-se que a escolha das linguagens LCSH, TBN, BIBLIODATA e Vocabulário USP foi adequada para o tipo de análise pretendida, considerando que se trata de linguagens desenvolvidas para contextos semelhantes de aplicação.

A linguagem BIBLIODATA, porém, demonstrou não possuir hierarquização de termos, inviabilizando sua utilização nas fases finais da pesquisa e, portanto, decretando que sua utilização para o propósito final desta análise, a construção de

uma proposta de macroestrutura para a Linguagem UNESP, é inviável. Desta verificação, conclui-se que a escolha das linguagens de indexação destinadas à análise de macroestruturas deve considerar linguagens constantemente atualizadas, como o são a LCSH, a TBN e o Vocabulário USP e que sejam direcionadas a contextos de uso semelhantes.

Quanto à estrutura das linguagens, é possível destacar que a LCSH sobressaiu-se por sua exatidão e diversidade terminológica, oferecendo opções para o trabalho do indexador, de modo que este possa atender às necessidades mais atuais de interdisciplinaridade científica.

A TBN, por ser uma versão da LCSH, possui com essa linguagem mais semelhanças, porém, não apresenta a mesma diversidade que a sua matriz, de modo que esta análise não poderia prescindir do uso de ambas as linguagens para a composição da representação documentária dos domínios.

Quanto ao Vocabulário USP, podemos destacar que sua estrutura diferenciada das demais linguagens indica que sua utilização será importante para a elaboração da macroestrutura da Linguagem UNESP, especialmente por haver sido construído no contexto de um sistema integrado de bibliotecas universitárias prezando, em sua constituição, características das áreas de domínio da universidade.

As três linguagens demonstram ser importantes para a construção de estruturas lógico-hierárquicas com profundidade adequada à busca contínua pela geração de novos conhecimentos e inovação no âmbito das bibliotecas universitárias.

Finalmente, a metodologia utilizada mostrou-se apropriada à proposta, destacando-se o uso da Diretriz ANSI/NISO (2005), que contempla e amplia as orientações propostas pelo IBICT (1984) e UNESCO (1996). O uso do método de comitê ascendente associado ao Modelo Metodológico Integrado para construção de Tesouro de Cervantes (2009) foi bem sucedido e conduziu a elaboração da proposta de estrutura lógico-hierárquica para expansão dos domínios de Física e Matemática. Dessa forma, a utilização de procedimentos terminográficos para a construção de macroestrutura lógica de linguagens de indexação permitiu a caracterização dos domínios e indica-se que seja utilizada esta opção metodológica para a construção da macroestrutura para a Linguagem UNESP.

## Notas

- (1) Parte das atividades de pesquisa em desenvolvimento com auxílio regular FAPESP (Processo 2015/13410-8)

- (2) Catálogo *on-line* construído pela rede de bibliotecas universitárias da Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, disponível em: <http://www.athe na.biblioteca.unesp.br>.

## Referências

- American National Standards Institute; National Information Standards Organization (2005). ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, Ma: NISO Press, 2005. 184 p. [http://www.niso.org/apps/group\\_public/project/details.php?project\\_id=46](http://www.niso.org/apps/group_public/project/details.php?project_id=46) (2016-04-31).
- Barbosa, Alice P. (1972). Projeto CALCO: adaptação do MARC II para implantação de uma central de processamento da catalogação cooperativa. 1972. Dissertação (Mestrado em Biblioteconomia e Documentação) – Instituto Brasileiro de Biblioteconomia e Documentação, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Boccatto, V. R. C., Fujita, M. S. L., Gil Leiva, I. (2011). Avaliação comparada do uso de linguagens de indexação em catálogos de bibliotecas universitárias para recuperação por assunto. // *Scire*. 17:1 (2011) 55-64.
- Cervantes, Brígida Maria Nogueira (2009). A construção de tesouros com a integração de procedimentos terminográficos. 2009. 209f. Tese (Doutorado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009.
- Cintra, A. M. M.; et al. (1994) Para entender as linguagens documentárias. São Paulo: Polis, 1994. 72 p.
- Fujita, M. S. L. (2011). Modelos de categorização para a construção de tesouros: metodologia de ensino. // Boccatto, V. R. C.; Gracioso, L. de S. (org.) Estudos de linguagem em Ciência da Informação. Campinas, SP: Alínea, 2011. 35-67.
- Fujita, M. S. L.; Boccatto, V. R. C.; Rubi, M. P. (2010). O contexto da indexação para a catalogação de livros em abordagem sociocognitiva. // *Brazilian Journal of Information Science*. 4:2 (jul./dez. 2010) 22-40. <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/485> (2011-07-16)
- Fundação Getúlio Vargas. Rede BIBLIODATA/CALCO (1995) Manual de cabeçalhos de assunto: normas e procedimentos: versão 1.0. Rio de Janeiro.
- Gil Leiva, I. (2008) Manual de indización: teoría y práctica. Gijón: Ediciones Trea.
- Gil Urdiciain, B. (2004). Manual de lenguages documentales. Gijón: Ediciones Trea. 280 p.
- Grings, L. (2016). Controle de autoridades na Biblioteca Nacional do Brasil: breve histórico e práticas atuais. // *Revista Brasileira de Biblioteconomia e Documentação*. 12:1 (No prelo).
- Guimarães, J. A. C. (2008). A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO) // *Revista Ibero-Americana de Ciência da Informação*. 1:1 (jan./jun. 2008) 77-99.
- Histórico. ([2016]). Rede BIBLIODATA. Rio de Janeiro: IBICT. <http://bibliodata.ibict.br/geral/modelos/historicoantes.htm> (2016-03-05)
- Hoshy, L.M.E. (2001). Relationships in Library of Congress Subject Headings. // Bean, C.A. (ed.) Relationships in the organization of knowledge. Dordrecht: Kluwer Academic Publishers, 2001. 135-152.
- Lancaster, F. W. (2004). Indexação e resumos: teoria e prática. 2. ed. Brasília: Briquet de Lemos.

- Torres, L. M. C. (2006). Sistematização da sintaxe de cabeçalho de assunto. // Gomes, H. E. (coord.) BITI - Biblioteconomia, Informação & Tecnologia da Informação. Rio de Janeiro, 2006. <http://www.conexaorio.com/bitil/lecy/lecy.htm> (2013-10-22).
- UNESCO (1993) Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngües. [2ª ed. rev. por Derek Austin e Peter Dale]; Trad. de Bianca Amaro de Melo; rev. de Lúgia Maria Café de Miranda. Brasília: IBICT; SENAI.
- Universidade de São Paulo (2015) Sistema Integrado de Bibliotecas. Vocabulário Controlado USP. <http://www.sibi.usp.br/produtos/vocabulario-controlado-usp/> (2016-01-16).
- Universidade de São Paulo ([s.d.]). Vocabulário Controlado do SIBi/USP: histórico. <http://143.107.154.62/Vocab/imagens/Historico.htm> (2016-03-27).
- Universidade de São Paulo (2006) Sistema Integrado de Bibliotecas. Manual de indexação de assuntos com uso do Vocabulário Controlado USP: versão preliminar. São Paulo. 72 p. [http://areatecnica.sibi.usp.br/wiki/index.php/Vocabul%C3%A1rio\\_Controlado](http://areatecnica.sibi.usp.br/wiki/index.php/Vocabul%C3%A1rio_Controlado). (2012-11-27)
- Yi, K.; Chan, L. M. (2010). Revisiting the syntactical and structural analysis of Library of Congress Subject Headings for the digital environment. // Journal of the american society for information science and technology. 61:4, 677-687. <http://onlinelibrary.wiley.com/doi/10.1002/asi.v61:4/is-suetoc> (2016-03-16).

---

Enviado: 2016-04-01. Segunda versão: 2016-06-29.  
Aceptado: 2016-09-15.

---

---

# Estrategias para mejorar la visibilidad y difusión de la revista *Fotocinema*: creación de un vocabulario controlado multilingüe

*Strategies to improve the visibility and dissemination of the journal Fotocinema: creation of a multilingual controlled vocabulary*

---

Rocío PALOMARES-PERRAUT, Carmen GÓMEZ CAMARERO

Departamento de Filología Griega, Estudios Árabes, Lingüística General, Documentación y Filología Latina. Facultad de Filosofía y Letras. Universidad de Málaga. Campus de Teatinos, 29071. Málaga. ESPAÑA. perraut@uma.es; gomez@uma.es

## Resumen

Se describen los procedimientos seguidos para la elaboración de un lenguaje controlado multilingüe (español, inglés, francés) para la revista científica de acceso abierto *Fotocinema: revista científica de Cine y Fotografía*, como estrategia para alcanzar la máxima visibilidad y difusión en redes sociales, bases de datos científicas y sistemas de evaluación nacionales e internacionales. Metodológicamente se han empleado técnicas de tipo analítico, descriptivo, cuantitativo y terminológico. A partir de las palabras aportadas por los autores de los diez números publicados de esta revista (2010-2015) se propone un listado abierto de términos controlados con el objetivo de normalizar el vocabulario de los campos temáticos abordados por la revista y facilitar a los autores la tarea de asignación de palabras clave a sus artículos. Para ello se han tenido en cuenta diferentes herramientas terminológicas de consulta e indización. El resultado se plasma en un listado de 338 términos en tres idiomas que, junto a unas recomendaciones a los autores, se incorpora a las normas de publicación de *Fotocinema*. Se pretende, en definitiva, contribuir a aumentar la visibilidad de las publicaciones de la revista y mejorar la indexación, análisis y clasificación de la misma en los índices internacionales, con el fin último de elevar su calidad y su índice de impacto en su área de conocimiento.

**Palabras clave:** Revistas científicas. Calidad. Acceso abierto. Visibilidad. Difusión. Tesoros. Cine. Fotografía. Cinematografía. España. Fotocinema.

## 1. Introducción

La recuperación de información (*Information Retrieval*) es una de las disciplinas científicas del área de la Documentación cuyo objetivo principal es el de favorecer los diálogos que se establecen entre los usuarios que buscan información y los sistemas que almacenan información. Con el desarrollo de las tecnologías de información, estos sistemas han ido evolucionando y, si en sus orígenes eran bases de datos de acceso en línea, ahora podemos hablar de buscadores, arañas o servicios desarrollados en la Web (Baeza-

## Abstract

We describe the procedures followed to create a multilingual controlled vocabulary (Spanish, English, French) for the open access journal *Fotocinema: Scientific Journal of Cinema and Photography*. It has been conceived as a strategy to achieve the maximum visibility and dissemination in social networks, scientific databases and national and international evaluation systems. We have applied, methodologically, analytical, descriptive, quantitative and terminological techniques. We propose an open list of controlled terms taken from those keywords given by authors in the ten issues published by *Fotocinema* from 2010 to 2015. The aim is to standardize the thematic fields addressed by the journal *Fotocinema* and to facilitate to the authors how to choose the accurate keywords of their articles. To do so, we have worked with several terminological tools for querying and indexing. The result is a list of 338 semi-controlled terms in three languages which, together with some recommendations for authors, is incorporated to the "Author Guidelines" of *Fotocinema*. The ultimate aim is to promote the visibility of the articles published in *Fotocinema* and also to improve its indexation, analysis and classification on those international evaluation systems to get better quality and impact factor in their area of expertise.

**Keywords:** Scientific journals. Quality. Open access. Visibility. Dissemination. Thesauri. Cinema. Photography. Cinematography. Spain. Fotocinema.

Yates & Ribeiro-Neto, 1999) o la ahora denominada web social o web 2.0 (O'Really, 2005); nos referimos a redes sociales, repositorios, blogs, espacios de alojamiento de vídeos, wikis, entre otros, que son reflejo de los movimientos de colaboración y de participación de una gran masa de usuarios altruistas (Domingo, González y Lloret, 2008). Este comportamiento del usuario web 2.0 también se refleja en el ámbito académico y científico donde investigadores y creadores de conocimiento se expresan con este nuevo paradigma de comunicación o Ciencia 2.0 (REBIUN, 2010, p. 5):

La innovación aplicada a la investigación también se sirve de tecnologías participativas y recursos abiertos. Es la denominada e-Ciencia o Ciencia 2.0. La ciencia 2.0 es la aplicación de las tecnologías de la web social al proceso científico. La web social, web 2.0 o web participativa se caracteriza por el empleo de tecnologías abiertas, tanto desde el punto de vista de la arquitectura de la información, como de la interconexión de servicios y, sobre todo, del trabajo colectivo que se realiza de forma telemática, colaborativa y desinteresada.

En este contexto de tecnologías colaborativas, de expansión de canales de difusión y de comunicación para la información científica, donde cada vez más la rapidez en la publicación es un aspecto relevante, surge el movimiento de acceso abierto (*Open Access*), una iniciativa que (Suber, 2004 *apud* Melero; Barrueco, 2007)

[...] permite el libre acceso a los recursos digitales derivados de la producción científica o académica sin barreras económicas o restricciones derivadas de los derechos de copyright sobre los mismos (...) El acceso es online a través de Internet, y salvo limitaciones tecnológicas y de conexión a la red del usuario no debería estar restringido por otro tipo de imposiciones.

Varios son los motivos por los que el acceso abierto ha tenido una amplia acogida en la producción y difusión de las revistas científicas; entre ellos, se encuentran el control del *copyright* o derechos de autor, los abusos de las editoriales y el alto coste de las mismas en formato impreso (Melero, 2005). Para obtener información sobre las políticas editoriales de las revistas científicas existe la base de datos RoMEO (*Rights METadata for Open Archiving*) (Sherpa/Romeo, 2015) que registra unas 500 políticas editoriales en relación con la publicación en acceso abierto, y que ofrece información sobre las condiciones de derechos de copia y de autoarchivo establecidas por revistas y editoriales muy prestigiosas a nivel internacional. Los investigadores que quieran poner sus trabajos a disposición pública en la web disponen de esta base de datos para consultar qué tipo de política de acceso siguen las editoriales de las revistas en que publican sus artículos. RoMEO divide las editoriales en cuatro grupos y cuatro colores en función de la política de permisividad o de prohibición en el acceso abierto que sigan. De este modo, los editores calificados como verde, serían los más permisivos, ya que permiten el acceso abierto de cualquier versión de las publicaciones de los investigadores; los editores azules, que permiten la versión *post-print*; los editores amarillos, únicamente la versión *pre-print* de los artículos; y por último, los editores blancos que no tienen ninguna política de acceso abierto y, por lo tanto, no ceden sus derechos a los autores para que puedan publicar

alguna versión en la web (Robinson-García; Delgado López-Cózar & Torres-Salinas, 2011). Las editoriales de revistas científicas españolas están mínimamente representadas en esta base de datos internacional, pero existe una herramienta similar en el ámbito español que sigue los mismos criterios que la anterior. Se trata del portal DULCINEA (2008), cuyo objetivo es identificar y analizar las políticas editoriales de las revistas españolas respecto al acceso a sus textos y archivos, los derechos de *copyright* sobre los mismos y cómo éstos pueden afectar a su posterior auto-archivo o depósito en repositorios institucionales o temáticos. La base de datos de DULCINEA cuenta, en la actualidad, con más de 1723 revistas españolas registradas.

En cualquier caso, la distribución online, abierta, libre y gratuita está permitiendo que nuevas áreas de conocimiento tengan mayor difusión y visibilidad. Es la razón por la que el volumen de revistas científicas de acceso abierto haya crecido rápidamente en todos los ámbitos del conocimiento en los últimos quince años, hasta el punto de convertirse en un nuevo modelo de negocio consolidado (Laakso, 2011). Instituciones científicas de todo el mundo, o directorios como *Directory of Open Access Journals* (DOAJ) se están convirtiendo en opciones de información científica altamente consolidadas entre la comunidad de investigadores, por lo que no dejan de mejorar sus plataformas de acceso y de endurecer los criterios de calidad para la selección de las revistas que indexan. Es el caso de la base de datos de DOAJ, que desde su creación en 2003 no ha dejado de crecer: de 300 revistas indexadas inicialmente a más de 10.000 en 2015 (Olijhoek, Mitchell & Bjørnshauge, 2015, p. 4).

Muy relacionado igualmente con el concepto de acceso abierto, se encuentran las licencias *Creative Commons* (2015), modelos que permiten al autor publicar y depositar sus obras científicas y/o artísticas según los usos que de ellas se proponga. Esta organización sin ánimo de lucro propone cuatro condiciones que, combinadas entre sí, proporcionan seis licencias de usos diferentes. Estas cuatro condiciones son Reconocimiento de la autoría de la obra (*Attribution*); No Comercial (*Non commercial*), que autoriza la obra para usos no comerciales; No derivada (*No Derivated Works*), cuya autorización para explotar la obra no incluye la transformación para crear una obra derivada; y Compartir Igual (*Share alike*), en la que la explotación autorizada incluye la creación de obras derivadas siempre que mantengan la misma licencia al ser divulgadas. La combinación de estas cuatro condiciones generan las seis licencias *Creative Commons*, un



nuevo modelo de derechos de autor más flexible que da más oportunidades a aquellos que quieren compartir o hacer uso de nuevos trabajos (Armstrong, 2012).

### 1.1. Breve descripción de la revista *Fotocinema*

La revista *Fotocinema: Revista Científica de Cine y Fotografía* se crea en el seno de esta filosofía de la política de libre acceso y gratuito, acogién-dose a los principios de licencia de *Creative Commons* del tipo Reconocimiento-No Comercial-Sin obra derivada. Fundada en el seno de la Universidad de Málaga en el año 2010 y con una periodicidad semestral, actualmente, cuenta con once números con trabajos en español, francés, inglés, italiano y portugués, lo que le otorga su carácter internacional. Aunque está representa-da en más de 26 plataformas de evaluación, entre las que destacan DOAJ, Dialnet, Latindex o Scielo, también aspira a alcanzar mayor visibili-dad y difusión si reúne los requisitos de calidad que exigen las bases de datos científicas interna-cionales tales como *Scopus* o *Web of Science*. Así pues, con el objetivo de alcanzar un mejor posicionamiento en los índices internacionales más reconocidos en la comunidad científica y los organismos nacionales de evaluación científica, y más concretamente en *Scopus*, *Fotocinema* ha iniciado una serie de estrategias, entre las que se encuentra la propuesta que describimos en este trabajo, es decir, el proceso de elaboración de un lenguaje controlado multilingüe de cine y fotogra-fía a fin de facilitar a los autores de esta revista la indización o asignación de palabras clave para sus trabajos científicos. Este listado de palabras clave se propone como un elemento imprescindible para aumentar la visibilidad y la calidad de la revista en la sección de “Normas para los auto-res”.

### 1.2. La importancia de las palabras clave para la descripción y recuperación de recursos electrónicos

Hoy por hoy, la mayoría de la literatura científica se sigue expresando en lenguaje escrito, como medio de comunicación entre los científicos que quieren dar a conocer sus investigaciones y que se nutren de otras investigaciones escritas para avalar sus resultados. La unidad tradicional para comunicar las innovaciones científicas y académicas es el artículo. Parte de su estructura la constituyen las palabras clave o *keywords* como elementos obligatorios e indispensables para facilitar y garantizar una identificación y recupera-ción eficaz del conocimiento que se pretende com-unicar en ese artículo (Yang, Chang-Jin, 2010).

Y esto es así, porque es la forma en que se recu-pera el conocimiento en las bases de datos cien-tíficas.

También, en el código fuente de los documentos web, las palabras clave forman parte sustancial de los metadatos Dublin Core, una iniciativa que comienza en 1995 con el objetivo de contribuir al desarrollo de la web semántica o web inteligente (Lamarca, 2006). Los metadatos Dublin Core (DC) se crean con la finalidad de describir, identi-ficar, encontrar y recuperar recursos electróni-cos en la red; son parte de vocabularios y estándar-es más amplios (las ontologías) para que sean las máquinas, y no los humanos, las que distinguen los significados o contextos de los do-cumentos. Se pretende, por tanto, elaborar nor-mas interoperables sobre metadatos y desarrol-lar vocabularios especializados para la descripc-ió de objetos digitales y garantizar la recupera-ción de la información. Uno de los 15 metadatos DC que todo recurso electrónico debería poseer en su cabecera para facilitar su indización, es el metadato *D.C. Subject*, sobre materia y palabras clave. En la descripción que proporciona Dublin Core de este metadato se describe la etiqueta, se ofrece la definición del metadato además de in-cluir un comentario. En consecuencia, para obten-er una mayor visibilidad, difusión e impacto, se procura que las revistas científicas digitales apli-quen los estándares DC en la descripción, no ya sólo de la revista, sino de cada uno de los artícu-los publicados en ella. En concreto, esta última condición, el que cada uno de los artículos de una revista científica proporcione metadatos, es uno de los 7 requisitos indispensables para las revistas indexadas en DOAJ desde 2014 para obtener el denominado Sello DOAJ (2015), que testimonia los más elevados niveles de compro-miso de políticas de acceso abierto, es decir, “la forma más abierta del acceso abierto” ([http://du-blincore.org/](http://dublincore.org/)).

Las plataformas de acceso a revistas digitales, tanto comerciales como de libre acceso, utilizan esta forma estandarizada en mayor o en menor medida, adoptando como obligatorios todos o parte de los metadatos DC (Estivill et al., 2005). Señala Alonso-Arévalo (2015) que “los metada-tos en el ámbito editorial proporcionan un carác-ter comunitario y dinámico, juegan un papel esencial en el control de los útiles de promoción de las obras, dotándolas de mayor visibilidad y por lo tanto de mayor capacidad de proyección”. Aunque se refiere a los metadatos de los libros electrónicos, las afirmaciones de este autor pue-den trasladarse perfectamente al ámbito de las revistas digitales. Y añade, “la visibilidad de un autor, de un título o de un editor, dependerá de la calidad de los metadatos que se incluyan en la

obra, siendo éste un aspecto al que los profesionales de la edición han de prestar un especial cuidado”. La información aportada por los metadatos es fundamental para buscadores, bases de datos, catálogos y demás sistemas de recuperación. En el caso de las revistas digitales y de las plataformas que facilitan su acceso, la presencia de metaetiquetas DC se establece como un criterio de selección, a veces recomendado y otras veces como muy valorado. Así lo hacen, por ejemplo, la Fundación Española para la Ciencia y la Tecnología (FECYT) para revistas científicas electrónicas de España y de la América Latina o

Latindex para revistas electrónicas de libre acceso de Iberoamérica. Para esta última organización, uno de sus criterios de calidad, concretamente el criterio número 34 (metaetiquetas), expresa que “califica positivamente si aparecen metaetiquetas Dublin Core en la página de presentación de la revista (código fuente)” (Latindex, 2015). De ser así, como venimos insistiendo, el buscador localizará con más pertinencia el documento electrónico que el usuario demande al encontrar en el código fuente esos términos o palabras clave que identifiquen su búsqueda (Figura 1).

```

12
13 <meta name="DC.keywords" content=" Cine, cinema, cinematógrafo, cinematografía, celuloide, gran par
14
15 <meta name="DC.Description.Abstract" content="FOTOCINEMA. Revista Científica de Cine y Fotografía,
16
17
18 <meta name="DC.Source" content=" Fotocinema : Revista Científica de Cine y Fotografía. Málaga: Univ
19
20 <meta name="DC.Type" content=" text/>
21
22
23
24 <meta name="DC.Relation" content=" PLATCOM http://plataformarevistascomunicacion.org/ />
25
26
27 <meta name="DC.Coverage" content="artículos científicos y académicos sobre el cine y fotografía de
28
29 <meta name="DC.Creator" content="Universidad de Málaga"/>
30
--

```

Figura 1. Metaetiquetas Dublin Core en el código fuente de la revista Fotocinema

Por otra parte, otros criterios de calidad que debe cumplir una revista científica en sus aspectos formales son: la existencia de palabras clave y de sus equivalentes en inglés y/o francés, el resumen o *abstract* en inglés y/o francés, el sumario que relacione los artículos que se publiquen en un determinado número de ese título, el cumplimiento de la periodicidad, etc. Con todo ello, la finalidad que se persigue es la normalización de las revistas electrónicas de libre acceso, cuestión tratada en numerosos trabajos (Delgado, 1997; Delgado, 2004; Giménez et al., 1999; Román et al., 2002; Abadal & Rius, 2006). Muchos de estos aspectos formales se han elevado a la categoría de normas oficiales; en el caso que nos ocupa, la norma ISO 215 (1986) – UNE 50133:1994, referente a la presencia de palabras clave en los artículos, así como la ISO 5963 – UNE 50121:1991, relativa a los procedimientos para seleccionar los términos de indización. La normalización es indispensable para definir los objetos -artículos científicos-, simplificar los procesos -formas de asignar palabras clave-, y permitir el intercambio -importación o exportación a otros sistemas de recuperación.

### 1.3. La indización en el proceso editorial de las revistas científicas

En el proceso editorial de la mayoría de las revistas científicas, se otorga al/los autor/es la responsabilidad de asignar estos términos en número y forma, de manera que son ellos mismos los que identifican y describen el contenido de sus artículos científicos, aunque existen estudios que demuestran que los autores no conceden la suficiente importancia a la asignación de palabras clave (Vrkic, 2014, p. 850). Pero una incorrecta utilización de las palabras clave puede traducirse en que estos trabajos sean invisibles, no existan a la hora de hacer una búsqueda bibliográfica y/o documental, y consecuentemente, una deficiente visibilidad del autor/es de cara a la comunidad científica (González & Mattar, 2012). Generalmente, los autores utilizan palabras clave libres, extraídas del lenguaje natural y que son utilizadas también en el título y el resumen para darles mayor ponderación e importancia. Es excepcional que estas palabras clave pertenezcan a un lenguaje documental, ya sea un tesoro, un vocabulario controlado o un listado de términos, que suelen ser las herramientas utilizadas por los indizadores profesionales. Gil-Leiva y Alonso-Arroyo (2005) aseguran que no existe una gran

diferencia entre la indización de los autores y aquella que hacen los profesionales de bases de datos como ISOC, aunque Grande et al. (2005), en un análisis de 706 manuscritos originales relacionados con áreas del sistema respiratorio, observaron que sólo el 50% de dichos manuscritos emplearon palabras clave de manera correcta. Según Maltrás (2003), para mejorar la indización y homogeneizar de algún modo la terminología usada para la indización de artículos científicos, sería conveniente que las palabras clave sean asignadas por indizadores profesionales, y que debe tratarse más de una herramienta usada por el editor de la revista que por el propio autor del trabajo (Maltrás, 2003, p. 102, *apud* Sánchez Valle, 2006, p. 142). Indudablemente, se trata de una cuestión que continúa sin estar definida de manera normalizada, por lo cual, nos encontramos con diferentes maneras de abordar el tema por parte de los equipos editoriales y de redacción de las revistas científicas. Así, por ejemplo, en otros casos, la propia revista facilita la labor de asignación a los autores mediante recomendaciones de indización. Esta última eventualidad es defendida por expertos en evaluación de la calidad de las revistas científicas españolas, los cuales han articulado una propuesta de evaluación integral de las mismas, conformada por un conjunto de criterios e indicadores entre los que se encuentra definida la cuestión de las palabras clave y la forma de asignarlas (Delgado López-Cózar, Ruiz-Pérez y Jiménez-Contreras, 2006, p. 72). Al definir este indicador, se considera un requisito normativo básico que las palabras clave sean extraídas de listas de materias, clasificaciones o tesauros propios de la especialidad.

En esta línea se sitúa nuestra hipótesis, puesto que además de proporcionar orientaciones, consideramos que ofrecer un lenguaje controlado de términos clave especializados en las materias de cine y fotografía, elevaría los niveles de calidad y aumentaría la visibilidad y difusión de la revista *Fotocinema*. En este sentido, proponemos que debieran ser los comités de redacción de las revistas científicas las que ofrecieran índices controlados basados en palabras clave del área de conocimiento al que pertenezcan, para así proporcionar una normalización terminológica y conseguir un mejor diálogo entre los usuarios que buscan información y los sistemas de recuperación.

## 2. Metodología

Nuestro punto de partida para recopilar información fue el portal de evaluación de revistas y países *Scimago Journal & Country Rank* (2007-2015), que calcula el factor de impacto *Scimago*

*Journal Rank* (SJR) basándose en la información incluida en la base de datos *Scopus*, de la empresa *Elsevier* (1). Nuestro objetivo inicial consistió en realizar un análisis y diagnóstico de cómo los comités de redacción de las revistas evaluadas por *Scopus* resuelven la cuestión de la indización o asignación de palabras clave. Hemos consultado que *Scopus* dispone de un equipo de indizadores profesionales que utilizan herramientas terminológicas específicas (tesauros y vocabularios controlados) para complementar, de forma manual, la indización realizada por los propios autores de la revista (*Scopus: Content Coverage Guide*, 2014, p. 13). Sin embargo, observamos que todos estos vocabularios controlados están especializados en los ámbitos de la ciencia y la tecnología, pero ninguno en los campos de las ciencias humanas y sociales que son los ámbitos de conocimiento que nos interesan. No obstante, para esos ámbitos, *Scopus* se sirve de clasificaciones que dividen a las revistas de las ciencias humanas y sociales en áreas y categorías temáticas. De ahí que pudiéramos indagar en qué categoría/s estaría clasificada la revista *Fotocinema*.

Identificamos dos categorías *Scopus* para *Fotocinema*: “Comunicación”, porque es una revista de comunicación audiovisual, y “Artes Visuales y Escénicas”, porque es una revista de cine y fotografía. A caballo entre estas dos categorías temáticas, nos inclinamos más por “Artes Visuales y Escénicas”, por considerarla más específica. Así pues, metodológicamente, seleccionamos como muestra de estudio todas las revistas de la categoría de “Artes Visuales y Escénicas” de acuerdo con la información proporcionada por *Scimago* del año 2013.

Según estos datos (Figura 2, en la página siguiente), nos encontramos con un total de 383 revistas en la categoría “Artes Visuales y Escénicas”. A partir de aquí, empezamos a establecer filtros, siempre con el objetivo de conocer cómo estas revistas contemplan la cuestión de la indización. Seguidamente, procedimos a seleccionar aquellas revistas que sólo se dedican al cine y a la fotografía, excluyendo las que tratan otras artes escénicas como son la danza, el teatro, los museos, la arquitectura, entre otros. De manera manual, revisando los objetivos de cada una de ellas, pudimos extraer 124 revistas que abordan los temas de cine y fotografía. Con esta muestra, nos detuvimos en identificar aquellas que proporcionan guías de estilo para los autores, reduciéndose la selección a 103 revistas. A continuación, de ellas, distinguimos las que mencionan expresamente algo sobre la asignación de las palabras clave (guías *keywords* en el gráfico): conforman un total de 64 revistas. De esta muestra de 64

revistas, sólo 36 proporcionan una información más completa y específicamente dedicada a la asignación de palabras clave (guías explícitas para *keywords*). El resto, se limitan a indicar el número de palabras clave que el autor debe proporcionar (número que oscila entre 4 y 6, por regla general, aunque algunas proponen la asignación de hasta 10 palabras clave).

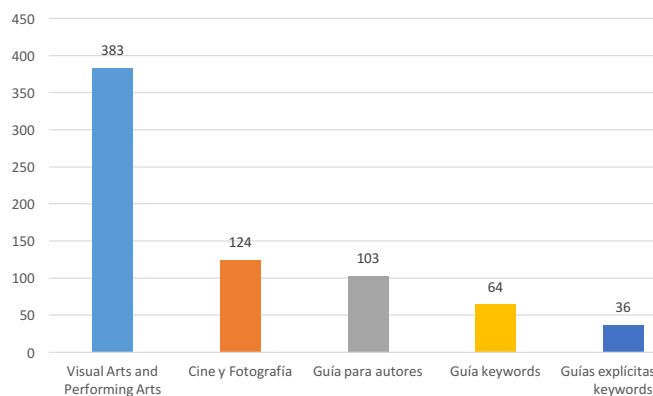


Figura 2: Selección de revistas sobre Cine y Fotografía indexadas por Scopus y presencia de guías sobre keywords

En este último grupo de 36 revistas con guías explícitas para palabras clave, nos parece interesante destacar las recomendaciones que ofrecen las editoriales que por lo general, monopolizan dichas guías:

a) *Taylor & Francis* (2015) destaca en sus *Guías para los autores* la importancia de las palabras clave para que sus trabajos sean más visibles; además, aporta información sobre los métodos de ranking de los buscadores, y la necesidad para el investigador de conocer lenguajes especializados. Asimismo, establece una correlación entre la presencia de pertinentes palabras clave y una mayor posibilidad de citas, tal y como figura en el siguiente texto dirigido a los autores:

“Making your article (and you) more discoverable  
Keywords and identifiers

A huge number of research articles are published every year, so ensuring that others can find your work is essential. As the author, there are a few things you can do to make your work more discoverable (and there are some things we're doing too).

Selecting keywords

When you submit your article you'll need to include keywords. These will be used to index your article on Taylor & Francis Online and on search engines such as Google Scholar™. These keywords will help others find your article quickly and accurately, so think of them as the labels for your article. What's more, a strong correlation exists between online hits and subsequent citations for journal articles.

But how do you choose your keywords? Think about how you search for articles, and what words or phrases you put in. Then think about your own article, and what keywords are most relevant to the focus of your work. Once you've drawn up a shortlist, try searching with them, to ensure the results fit with your article and so you can see how useful they would be to others.

Narrow down your keywords to ensure they are as accurate as possible, and then ensure you also include them in your title and abstract (as some search engines only index these), whilst still making it readable.

(<http://authorservices.taylorandfrancis.com/making-your-article-and-you-more-discoverable/>)

b) Otra editorial destacable en este sentido es *Sage* (2015), cuyas recomendaciones para la indexación son incluso más extensas pues proporcionan ejemplos prácticos sobre cómo hacer una mejor indexación. Explica los métodos de *Google* y *Google Scholar* como buscadores preferidos de los investigadores y estudiantes para localizar información. Destaca, en este sentido, la importancia del título, del resumen y de las palabras clave para hacer más visible un artículo. Igualmente, ofrece consejos para redactar los resúmenes y aconseja sobre el número de palabras clave que hay que asignar (entre 5 a 10). Y lo que es más importante, menciona de nuevo expresamente la correlación existente entre una buena indexación y una mejor visibilidad y factor de impacto:

Help Readers Find Your Article

The importance of search engines

Google and Google Scholar are the principal ways in which people will find your article online today. Between them they account for 60% of referral traffic to SAGE Journals Online. The search engine is now the first port of call for researchers and it is of paramount importance your article can be found easily in search engine results.

By taking some simple steps to optimize your article for search engines it will help your work to be discovered, then read, used and cited in others' work. This helps with the ISI Impact Factor of the journal your article is published in and will further raise the visibility of your article.

SAGE already undertakes many measures to ensure SAGE journals are indexed in the all the major search engines. There are over 100 factors that a search engine will look at before deciding how to rank your article in their search results, but the starting point is the content that you write”

(<https://uk.sagepub.com/en-gb/eur/help-readers-find-your-article>)

c) La tercera editorial importante es *Inderscience Publishers* (s.d.) que califica, en sus recomendaciones, el campo de las palabras clave como esencial, poniendo ejemplos de indexación al igual que hace *Sage*. E incluso incidiendo en las metaetiquetas:

## Keyword requirements

Introduction: the Importance of Keywords

Metadata (title/author details/abstract/keywords) is a vital part of any paper, since, in our online world, it is the metadata which is freely available to all users and which gets sent to many and varied online sources, from which users undertaking online searches will trace the paper. If you want your paper retrieved by the maximum number of people, the metadata is vital.

The Keywords field is essential in that this is the most frequently searched field if search results need to be narrowed, i.e. if the original search has produced too many records.

The title is a very concise indication of the content of a paper; the abstract is a more detailed – but still concise – description of that content; the keywords codify the key words and terms from the title and abstract (and text of the paper, if necessary).

Thus, the keywords should

- a) contain *all* the essential words/terms from the title and abstract and
- b) be in an optimum format, which is ideally of 1-3 words; if more than 1 word, the words should be a phrase, not a description (see below)

*If in doubt, think*

A) *what is the paper about [in terms of key words and phrases]?*

B) *what would i search for if i was trying to retrieve a paper on this topic? [...]*

([http://www.inderscience.com/info/pdf/dl.php?filename=id31\\_keywords.pdf](http://www.inderscience.com/info/pdf/dl.php?filename=id31_keywords.pdf))

El hecho es que, a pesar de todas estas recomendaciones, sólo una revista de nuestro conjunto de 36, ofrece un lenguaje controlado para ayudar al investigador en su labor de indización. Se trata de la revista *Journal of Mathematics and the Arts* (2007-2015), que recomienda y anima a los autores a usar una clasificación específica para las matemáticas: *2010 Mathematics Subject Classification*.

Por lo tanto, hemos observado que si hay propuestas y recomendaciones de diversa índole para que el autor de un trabajo científico asigne palabras clave, pero sólo ocasionalmente se proporciona un listado de términos controlados, tesaurus o algún otro tipo de lenguaje documental que pueda servir de orientación a los autores de los artículos.

### 3. Resultados

Los resultados obtenidos de este análisis nos animaron a seguir con nuestra idea de crear un lenguaje controlado específico para los autores que publican en *Fotocinema*. A partir de las palabras clave aportadas por los autores de los 10 números publicados (2010-2015), hemos realizado un listado abierto de 338 términos controlados, en 3 idiomas, español y sus equivalentes en

inglés y francés, con el objetivo de normalizar el vocabulario de los campos temáticos abordados por la revista. En su elaboración hemos tenido en cuenta como referencia diferentes herramientas terminológicas de consulta e indización para normalizar este lenguaje documental mediante remisiones, control de autores, control de topónimos y periodos históricos, uso del singular y del plural, entre otros, siguiendo las normas ISO y demás recomendaciones de otros tesaurus. Las herramientas terminológicas de referencia utilizadas son las siguientes: los Encabezamientos de Materia y Catálogo de Autoridades de la Biblioteca Nacional Española (BNE, 2013), le *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (RAMEAU) de la Biblioteca Nacional Francesa, *Library of Congress Authorities* de la Biblioteca del Congreso de los Estados Unidos, los encabezamientos de materia del proyecto *Periodical Indexing Project* (PIP) de la Federación Internacional de Archivos Fílmicos (FIAF); tesaurus especializados en cine y fotografía tales como *Art & Architecture Thesaurus* del Getty Research Institute, el *Thesaurus for Graphic Materials* (TGM) de la Biblioteca del Congreso de los Estados Unidos, el glosario especializado "*Movie Terminologie Glossary*" de *Internet Movie Data Base* (IMDb), *Glossary of Film Terms* del Portland Community College, o el *Tesaurus y Diccionario de objetos asociados a la expresión artística*, para la descripción y catalogación de bienes culturales (Trinidad Lafuente, 2012).

Debemos destacar que algunas de las dificultades que se nos han presentado a la hora de normalizar y documentar nuestro lenguaje controlado han sido las características del léxico propio de las materias de cine y fotografía (Hoyos, A., 1969), no sólo por la especialización de este vocabulario sino, sobre todo, por las diferentes perspectivas de análisis desde las que los autores abordan los temas cinematográficos y fotográficos en la revista. Estas perspectivas pueden ir desde la pragmática, estética, simbólica, histórica, cultural, ideológica, filosófica, hasta la lingüística o técnica (Rajas, M., 2005). Así, por ejemplo, la palabra clave "análisis fotográfico" puede referirse a un punto de vista técnico o estético. Serán las otras palabras clave del contexto la que definan la perspectiva. O igualmente, podemos encontrar términos pertenecientes a ámbitos muy dispares, como "fotofija", "éxito", "feminismo", "esperpento", "espejo" o "cine documental". Por esta razón hemos tenido que consultar diferentes fuentes de referencia para abarcar el amplio espectro temático de los artículos publicados en *Fotocinema*.

El control de los términos se ha ejercido en diversos aspectos, tales como:

a) Los topónimos o nombres geográficos; por ejemplo, para indicar “cine musical español”, se recomienda utilizar dos términos, es decir, “cine musical” porque es un género cinematográfico y otro término para indicar el lugar geográfico que en este caso sería “España”. Sin embargo, para referirse al cine creado en un país o región determinada, es recomendable utilizar el gentilicio. Por ejemplo, “Cine argentino” para el cine creado en Argentina;

b) Los periodos históricos: se recomienda utilizar los años que comprenden y/o el término con el que se identifica. Por ejemplo: Guerra civil española, 1936-1939; Cine Documental, 15M usado por Documentales del 15M;

c) Los nombres de persona: se recomienda utilizar primero el/los apellido/s seguido/s del nombre separado por coma. Por ejemplo: Almodóvar, Pedro.

d) Las materias o contenidos a los que se refiere el trabajo científico, se aconseja seguir las siguientes directrices generales:

- Utilizar sustantivos en lugar de adjetivos. Por ejemplo, en lugar de “Metaficcional” usar “Metaficción”.
- Utilizar preferentemente el plural para referirse a las cosas que pueden ser contadas. Por ejemplo, en lugar de “Género cinematográfico” emplear “Géneros cinematográficos”. Y utilizar el singular para los sustantivos no contables. Por ejemplo: Felicidad, Sátira.

Se trata de un listado con un vocabulario especializado, puesto que es el utilizado por los propios autores de los artículos de la revista, sólo que controlado y normalizado, para evitar ambigüedades y repeticiones. El listado se incorpora a las *Normas de publicación para los autores de Fotocinema*, junto con unas recomendaciones para su utilización y para facilitar la asignación de palabras clave. Nuestro propósito, desde el consejo de redacción de la revista, es que este listado se mantenga siempre en continua mejora y actualización, en beta permanente, y se vayan incorporando nuevos términos sugeridos y aportados por los autores a medida que éstos publiquen en los siguientes números. El fin último perseguido es contribuir a que los artículos de *Fotocinema* obtengan una mayor visibilidad, una mayor difusión y en definitiva, un mayor impacto entre la comunidad científica.

#### 4. Conclusiones

El etiquetado, la presencia de metadatos o la existencia de palabras clave complementan y enriquecen a los recursos electrónicos, cualquiera

que sea el formato que posean éstos. Cuando se refiere a información científica, estos elementos proporcionan un valor añadido, puesto que facilitan su identificación, almacenamiento, gestión y recuperación en internet, entre cantidades cada vez más ingentes de información. En el caso de las revistas científicas de acceso abierto, como ocurre con *Fotocinema*, creemos que es un mecanismo para potenciar su visibilidad y su difusión, y por ende, su impacto entre la comunidad científica interesada.

Por lo general, la tarea de asignación de las palabras clave es responsabilidad de los autores de los trabajos científicos, conocedores de los campos temáticos de sus dominios de investigación, pero está demostrado que muy raramente acuden a lenguajes documentales normalizados para hacerlo. Las plataformas de acceso a revistas digitales disponen de sus propios profesionales que completan la indización de los autores, apoyándose para ello en diferentes herramientas terminológicas, aunque todas ellas, curiosamente, pertenecen a los ámbitos de la ciencia y la tecnología. ¿Por qué la ausencia de lenguajes documentales normalizados para las ciencias sociales y humanas?

De 124 revistas indizadas por *Scopus* especializadas en cine y fotografía, sólo 36 (29%) proporcionan una guía explícita sobre palabras clave para ayudar a los autores a asignar palabras clave a sus trabajos, y únicamente una de ellas (0,8%) les recomienda consultar un listado de términos controlados. Sin embargo, en estas mismas guías se insiste sobre la importancia de la indización.

Además, se evidencia la correlación existente entre la presencia de recomendaciones ofrecidas por las revistas sobre indización, y un mayor impacto y visibilidad de las mismas, pues de las 36 revistas que ofrecen estas recomendaciones, 30 están situadas en la primera mitad del ranking de *Scimago* en la categoría de “Artes Visuales y Escénicas”.

Por lo que respecta a *Fotocinema*, con la creación de un lenguaje controlado de palabras clave de su área de conocimiento, el cine y la fotografía, su inclusión en las *Normas para los Autores*, además de unas recomendaciones para asignar las palabras clave, se persigue aumentar los rankings de impacto y poder ser incluida de esa manera, en los índices internacionales como *Scopus* o *Web of Science*. Nuestra propuesta pretende igualmente elevar el nivel de alfabetización de los autores en el ámbito de la indización, para que sean conscientes de la trascendencia de esta operación y de las repercusiones en cuestión de

impacto, visibilidad y difusión de sus trabajos en su ámbito de especialidad.

Las diferentes perspectivas desde las que el cine y la fotografía son analizadas y estudiadas en los trabajos publicados en *Fotocinema* han representado una dificultad a la hora de crear y normalizar los términos de nuestro lenguaje controlado y se ha necesitado acudir a diversas herramientas terminológicas, de diferentes ámbitos temáticos.

En definitiva, creemos que ofrecer en sus “Normas para los Autores” recomendaciones y orientaciones, y por qué no, un lenguaje controlado de referencia para facilitar y ayudar a los autores a indizar sus artículos constituye una buena estrategia para *Fotocinema* y, en general, para todas las revistas científicas de cualquier área de conocimiento.

## Notas

- (1) El factor de impacto *Scimago Journal Rank* (SJR) ha sido desarrollado por *Scimago*, grupo de investigación del Consejo Superior de Investigaciones Científicas (CSIC), y de las universidades de Granada, Extremadura, Carlos III (Madrid) y Alcalá de Henares.

## Referencias

- Abadal Falgueras, Ernest; Rius Alcaraz, Lluís (2006). Revistas científicas digitales: características e indicadores. // Roca, Genís (coord.). La presencia de las universidades en la Red [monográfico en línea], UOC. // Revista de Universidad y Sociedad del Conocimiento (RUSC), 2006, 3:1. [http://www.uoc.edu/rusc/3/1/dt/esp/abadal\\_rius.pdf](http://www.uoc.edu/rusc/3/1/dt/esp/abadal_rius.pdf) (2015-12-27).
- Alonso-Arévalo, Julio. (2015) ¿Cómo establecer buenos metadatos para mejorar la venta de libros?. // Universo Abierto: Blog de la Facultad de Traducción y Documentación de la Universidad de Salamanca. 26 noviembre 2015. <http://www.universoabierto.com/24275/%C2%BF-como-establecer-buenos-metadatos-para-mejorar-la-venta-de-libros/> (2015-12-27).
- Armstrong, Michelle (2012). Creative Commons – Saving the Internet One License at a Time. // The Idaho Librarian. 62:2. [http://works.bepress.com/michelle\\_armstrong/25](http://works.bepress.com/michelle_armstrong/25) (2015-12-27).
- Baeza-Yates, R.; Ribeiro-Neto, B. (1999) Modern Information Retrieval. New York: ACM ; Harlow, Essex : Addison-Wesley Longman, 2012.
- Biblioteca Nacional De España (2013). Manual de indización de Encabezamientos de Materia. Madrid: BNE, 2013.
- Creative Commons (2015). <http://es.creativecommons.org/blog/licencias/> (2015-12-27)
- Delgado López-Cózar, Emilio (1997). Normas ISO de presentación de publicaciones periódicas científicas. // Boletín de la Anabad. 47:2, 169-193.
- Delgado López-Cózar, Emilio (2004). Normalización editorial de las publicaciones científicas. // Rodríguez, Joaquín et al., Ediciencia: Manual de buenas prácticas para la difusión digital de los contenidos científicos de las universidades españolas. Madrid: Fundación Residencia de Estudiantes. Dirección General de Universidades, 2004, 231-282.

- Delgado López-Cózar, Emilio; Ruiz-Pérez, Rafael; Jiménez-Contreras, Evaristo (2006). La Edición de Revistas Científicas: Directrices, Criterios y Modelos de Evaluación. Madrid: FECYT; Granada: EC3, Universidad de Granada, 2006. <http://www.revistacomunicar.com/pdf/2011-04-Delgado.pdf> (19-11-2015).
- DOAJ (2015). 'Indexed in DOAJ' versus 'the DOAJ Seal'. // News Service: News, Updates & Developments from DOAJ. 2015/11/03. <https://doajournals.wordpress.com/2015/11/03/indexed-in-doaj-versus-the-doaj-seal/> (2015-12-27).
- Domingo, C.; González, J.; Lloret, O. (2008) La Web 2.0. Una revolución social y creativa: A Social and Creative Revolution. Web 2.0. // Telos: Cuadernos de Comunicación, Tecnología y Sociedad. 74. <http://telos.fundaciontelefonica.com/telos/anteriores/index.html> (2015-12-27).
- Estivill, Assumpció; Abadal, Ernest; Franganillo, Jorge; Gascón, Jesús; Rodríguez-Gairín, Josep-Manuel (2005). Uso de metadatos Dublin Core en la descripción y recuperación de artículos de revista digitales. // DC-2005 International Conference on Dublin Core and Metadata Applications, Madrid (Spain), 12-15 September 2005.
- Fotocinema: Revista científica de cine y fotografía. 2010-. ISSN 2172-0150. <http://www.revistafotocinema.com/> (2015-12-22).
- González Tous, Marco; Mattar V., Salim (2012). Las claves de las palabras clave en los artículos científicos. // Revista MVZ Córdoba. 17:2, 2955-2956.
- Gil-Leiva, Isidoro; Alonso-Arroyo, Adolfo (2005). La relación entre las palabras clave aportadas por autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT. // Revista Española de Documentación Científica. 28:1, 62-79.
- Giménez, Elea; Román, Adelaida; Sánchez, José María (1999). Aplicación de un modelo de evaluación a las revistas españolas de economía: una aproximación metodológica. // Revista Española de Documentación Científica. 23:3, 309-324.
- Granda Orive, J. I.; et al. (2005). Las palabras clave como herramientas imprescindibles en las búsquedas bibliográficas. Análisis de las áreas del sistema respiratorio a través de Archivos de Bronconeumología. // Archivos de Bronconeumología. 41, 78-83.
- Grupo de investigación “Acceso abierto a la ciencia” (2008). Dulcinea: Derechos de explotación y permisos para el auto-archivo de revistas científicas españolas. 2008. <http://www.accesoabierto.net/dulcinea/> (2015-12-22).
- Hoyos Ruiz, Antonio de (1967). Lingüística y cinema ( Notas para una estilística de la expresión cinematográfica ). // Monteagudo: Revista de literatura española, hispanoamericana y teoría de la literatura. 45, 7-17.
- Inderscience Publishers: Linking academia, business and industry through research. (s.d.). <http://www.inderscience.com> (2015-12-27).
- Journal of the Mathematics and the Arts (2007-2015). Taylor & Francis. ISSN 1751-3472 (Print), 1751-3480 (Online).
- Laakso M, Welling P, Bukvova H, Nyman L, Björk B-C, et al. (2011). The Development of Open Access Journal Publishing from 1993 to 2009. // PLoS ONE. 6:6, e20961. doi:10.1371/journal.pone.0020961
- Lamarca Lapuente, María Jesús (2006). Metadatos Dublin Core. // Hipertexto, el nuevo concepto de documento en la cultura de la imagen. Madrid: Universidad Complutense, 2006. Tesis doctoral.
- Latindex (2012-2015) Documentos de Latindex: Características editoriales para las revistas científicas. Criterio 34. [http://www.latindex.unam.mx/documentos/revistas\\_elec.html](http://www.latindex.unam.mx/documentos/revistas_elec.html) (2015-12-27).



- Maltrás, B. (2003). Los indicadores bibliométricos: Fundamentos y aplicación al análisis de la ciencia. Gijón: TREA, 2003.
- Melero, Remedios. (2005). Acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto. // El Profesional de la Información ( julio-agosto 2005), 14:4, 255-266.
- Melero, Remedios; Barrueco Cruz, José Manuel (2007). Acceso abierto y repositorio de documentos. // SEDIC: Asociación Española de Documentación e Información, 2015. [http://www.sedic.es/autoformacion/acceso\\_abierto/1-Introducci%F3n-acceso-abierto.html](http://www.sedic.es/autoformacion/acceso_abierto/1-Introducci%F3n-acceso-abierto.html) (2015-12-27).
- Melero, Remedios et al. (2009). DULCINEA: Copyright Policies and Type of Access to Spanish Scientific Journals. // ELPUB 2009: 13th International Conference on Electronic Publishing: Rethinking Electronic Publishing: Innovation in Communication Paradigms and Technologies. Milán, 10-12 junio 2009. [http://elpub.scix.net/data/works/att/89\\_elpub2009.content.pdf](http://elpub.scix.net/data/works/att/89_elpub2009.content.pdf) (2015-12-27).
- Miguel, S., Gómez, N-D. Y Bongiovani, P. (2012). Acceso abierto real y potencial a la producción científica de un país. El caso argentino. // El profesional de la información. 21:2, 146-153.
- Rajas, Mario (2005). Introducción al análisis retórico del texto fílmico. // ICONO: Revista de Comunicación y Nuevas Tecnologías. 2005, 14:5.
- Sherpa/Romeo: Políticas de copyright de las editoriales y autoarchivo. University of Nottingham, 2006-2015. <http://www.sherpa.ac.uk/> (2015-12-22).
- Olijhoek, T.; Mitchell, D.; Bjørnshauge, L. (2015). Criteria for Open Access and publishing. // ScienceOpen Research. 2015 [https://www.scienceopen.com/document\\_file/b34b287d-1e7f-4f7c-b5b6-ecdd19bcacef/ScienceOpen/ScienceOpenCleanVersionOct18th.pdf](https://www.scienceopen.com/document_file/b34b287d-1e7f-4f7c-b5b6-ecdd19bcacef/ScienceOpen/ScienceOpenCleanVersionOct18th.pdf) (2015-12-27).
- O'Reilly, T. (2005). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. // O'Reilly Media, 2015. <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html> (2015-12-27).
- REBIUN (2010). Ciencia 2.0: Aplicación de la web social a la investigación. REBIUN, 2010. [http://eprints.rclis.org/3867/1/Ciencia20\\_rebiun.pdf](http://eprints.rclis.org/3867/1/Ciencia20_rebiun.pdf) (2015-12-27).
- Robinson-García, Nicolás; Delgado López-Cózar, Emilio; Torres-Salinas, Daniel (2011). Cómo comunicar y diseminar información científica en Internet para obtener mayor visibilidad e impacto. // Aula Abierta. 39:3, 42-50.
- Román, Adelaida; Vázquez, Manuela; Urdín, Carmen (2002). Los criterios de calidad LATINDEX en el marco de la evaluación de las revistas españolas de humanidades y ciencias sociales. // Revista Española de Documentación Científica. 25:3, 286-307.
- Sage (2015). Resources: Journal Author Gateway. Sage Publications, 2015. <https://uk.sagepub.com/en-gb/eur/page/journal-author-gateway> (2015-12-27).
- Sánchez Valle, Ignacio (2007). Análisis de la indización practicada en la revista Pedagogía Social (1998-2003). // Revista Complutense de Educación. 18:1, 133-158.
- Scimago Lab (2007-2015). SJR: Scimago Journal and Country Rank. 2007-2015. <http://www.scimagojr.com/> (2015-12-27).
- Scopus: Content Coverage Guide (2014) Elsevier, 2014. [https://www.elsevier.com/\\_data/assets/pdf\\_file/0007/69451/sc\\_content-coverage-guide\\_july-2014.pdf](https://www.elsevier.com/_data/assets/pdf_file/0007/69451/sc_content-coverage-guide_july-2014.pdf) (Fecha de consulta: 20-11-2015).
- Suber, P. (2004). Open Access Overview. Richmond (Indiana): Earlham College, 2015. <http://bit.ly/oa-overview> (2015-12-27).
- Taylor & Francis (2015). Author Services: Supporting Taylor & Francis authors. London: Informa UK Limited, 2015. <http://authorservices.taylorandfrancis.com/category/writing-your-paper/> (2015-12-27).
- Trinidad Lafuente, Isabel (2012). Tesouro y Diccionario de objetos asociados a la expresión artística. Madrid: Ministerio de Educación, Cultura y Deportes, 2012.
- Vrkic, Dina (2014). Are they a perfect match? Analysis of usage of author suggested keywords, IEEE terms and social tags. // 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija (Croatia) 26-30 May 2014, 732-737. Doi: 10.1109/MIPRO.2014.6859662.
- Yang, Chan-Jin (2010). Study on Keywords and Their Use of Academic Theses - Focused on Database Development and Information Link. // Humanities Contents. 19, 395-416.

---

Enviado: 2015-12-19. Segunda versión: 2016-07-13.  
Aceptado: 2016-09-15.

---



---

# Análise das teorias semióticas na ciência da informação brasileira: referentes teóricos

*Análisis de las teorías semióticas en la ciencia de la información brasileña: referentes teóricos*

*Analysis of semiotic theories in Brazilian information science: theoretical referents*

---

**Carlos Cândido de ALMEIDA (1), Mona Cleide Quirino da Silva FARIAS (2)**

(1) Departamento de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista - UNESP, Av. Higino Muzzi Filho, 737, Marília, São Paulo, Brasil, CEP: 17525-900, Brasil carlosalmeida@marilia.unesp.br

(2) Programa de Pós-Graduação em Ciência da Informação, monaquirino@gmail.com

## Resumen

Se identifica y discute la presencia de las teorías semióticas en la literatura brasileña sobre información y documentación. Los resultados muestran que la semiótica de Peirce, a pesar de su reciente aparición en Documentación, ha sido consistente y relativamente dominante en los estudios semióticos del área. Consiguientemente, las teorías semióticas están presentes en la composición teórica de la Documentación en Brasil. Asimismo, los trabajos analizados muestran un cambio de paradigma, desde la perspectiva lingüística hacia la semiótica general.

**Palabras clave:** Teorías semióticas. Documentación. Ciencia de la información. Epistemología. Organización del conocimiento. Peirce, Charles Sanders. Brasil.

## 1. Introdução

Dada as dificuldades polissêmicas, natural do processo de evolução da linguagem humana, a compreensão de outrem, mesmo que este compartilhe de signos linguísticos comuns, nem sempre se realiza. Em determinados contextos sentimos que não basta compartilhar expressões do pensamento, mas contar com uma realidade comum. Segundo Flusser (2007, p. 81-82) “[...] a realidade dos dados brutos é apreendida e compreendida por nós em forma de língua”. Em outras palavras, a realidade nos é apresentada pela língua e, segundo o autor, cada língua criaria uma nova realidade. Sem discutir esta tese em termos objetivista ou relativista, identificamos a partir de Flusser um problema ainda maior em se interpretar os signos, mesmo em um contexto científico.

No caso de campos de pesquisa que estudam, organizam e manipulam a linguagem e o conhecimento sistematizado de outras áreas do saber e da ciência, como é o caso da Ciência da Informação (campo que estuda as práticas de informação e os processos envolvidos para isso) as questões da linguagem se agravam. A polissemia e a evolução da interpretação são fenôme-

## Abstract

The presence of semiotic theories in the Brazilian literature on information and documentation is identified and discussed. The results show that Peirce's semiotics, despite its recent appearance on Documentation, has been relatively consistent and dominant in the semiotic studies produced in the discipline. Thus, semiotic theories are present in the theoretical composition of the Brazilian Information Science. Also, the analyzed studies show a paradigm shift from the linguistic perspective towards the general semiotics approach.

**Keywords:** Semiotic theories. Information science. Epistemology. Knowledge organization. Peirce, Charles Sanders. Brazil.

nos reconhecidos pelos pesquisadores e especialistas do campo da Ciência da Informação e talvez as soluções encontradas em uma época, podem não servir a profissionais em outra. Nos ensina Moura (2007, p. 62) que

A estabilidade e a imutabilidade do signo informacional alardeados pelo campo mostrou-se, em termos semióticos, teoricamente falacioso e empiricamente irrealizável.

Os sistemas de organização do conhecimento (1), os sistemas de recuperação da informação, as técnicas adotadas para restringir a busca e a seleção da informação em sistemas automatizados, os métodos de elaboração de esquemas de organização do conhecimento, os métodos de determinação e restrição de significado dos termos, os procedimentos de controle terminológico, o desenvolvimento de sistemas de informação automatizados e, não menos importante, a comunicação da informação em quaisquer ambientes informacionais, supõem todos o tratamento da linguagem e o conhecimento das teorias a ele afeitas, as quais são oriundas da Linguística, Ciências Cognitivas, Filosofia, Sociologia e Semiótica.

Os processos informacionais (2) que envolvem, entre outras coisas, a transferência da informação

entre as pessoas, são problemas constantes do ponto de vista teórico e prático. Tais processos visam, a grosso modo, levar a informação aos sujeitos da forma mais adequada possível, com o objetivo precípua de realizar a interação, seja com o recurso às técnicas tradicionais compostas, principalmente, pela interação face-a-face, seja pelo uso massivo de tecnologias de processamento e transmissão de dados.

Tanto a linguagem quanto os processos informacionais são condições indispensáveis da apropriação da informação que requer a codificação e a decodificação de signos, bem como a transmissão destes no espaço e no tempo. A codificação simbólica está presente em todos os processos interpretativos, e a linguagem não se estruturaria sem os contratos estabelecidos socialmente. Não há fenômeno informacional que possa ser explicado sem a menção dos códigos e da tradução sógnica. Portanto, seria difícil compreendê-lo sem a referência a fatores semióticos.

Face a isso, o exame dos processos informacionais envolvidos nos sistemas de organização do conhecimento e na comunicação da informação dentro do campo da Ciência da Informação leva em conta diversas teorias científicas e abordagens filosóficas. Entre as perspectivas oferecidas para compreender tais fenômenos está o ponto de vista das teorias semióticas (ciência dos signos), em suas mais diversas vertentes. As teorias semióticas contemplam as principais abordagens sobre os signos, a semiose (ação dos signos) e a significação no contexto humano e não humano. Desse modo, teríamos propostas teóricas não antropocêntricas, em que a semiose é estudada de maneira a abarcar o mundo natural, como, segundo Noth (2005, p. 201) o comportamento sógnico de animais na zoossemiótica e dos processos microbiológicos na biossemiótica. Nas aproximações antropocêntricas, o objeto de análise é a semiose no contexto humano e cultural. Entre as diversas teorias semióticas, estão as relacionadas aos teóricos: Charles Peirce (1839-1914), Ferdinand de Saussure (1857-1913), Louis Hjelmslev (1899-1965), Roman Jakobson (1896-1982), Roland Barthes (1915-1980), Algirdas Julien Greimas (1917-1992), Charles William Morris (1901-1979), Yuri Lotman (1922-1993), Umberto Eco (1932-2016), entre outros comentadores contemporâneos.

Para fins deste artigo, as teorias semióticas foram divididas em duas linhas: teorias peirceanas, as quais recorrem a propositura teórica de Charles S. Peirce (1834-1914), e teorias não peirceanas (linguísticas, culturais e semiológicas), as quais não se vinculam à proposta lógica subjacente aos estudos peirceanos. Reconhecemos

que um estudo detido destas configurações teóricas deveria ser empreendido de modo a esclarecer as nuances conceituais e as imbricações existentes, pois estas reúnem um grande complexo de abordagens. Para justificar esta separação mencionamos a influência, no campo semiótico, da obra do filósofo Peirce.

Embora pareça generalizante, e pendendo a uma valorização da teoria semiótica peirceana, essa divisão tem revelado tendências mais globalizantes no trato dos temas. Uma teoria não peirceana, pode valer-se de elementos variados, como do estruturalismo linguístico, do funcionalismo ou da pragmática, sem reconhecer uma continuidade da linha teórica peirceana. Algumas teorias semióticas contemporâneas comportam-se assim, o que ficaria difícil classificá-las como estruturalista ou pragmática. Acreditamos que a divisão entre abordagens peirceanas e não peirceanas é mais abrangente além de abrir espaço para as teorias híbridas, as quais se vinculam a mais de uma matriz.

No âmbito da comunidade brasileira de pesquisadores em Ciência da Informação, a análise semiótica dos processos informacionais, com o intuito de recuperar os registros do conhecimento, são os processos que, desde a década de 1960, valem-se de teorias linguísticas para explicar a dinâmica das linguagens e suas condições para uma tradução.

Outros estudos tiveram como foco também a revisão e análise da presença da teoria semiótica na Ciência da Informação, enfatizando os atores brasileiros, dentro e fora da Ciência da Informação que aportavam elementos à questão. Entre estes, encontramos os trabalhos de Moura (2006), Monteiro (2006), Almeida (2010), Barros e Café (2012) e Vieira (2015).

Este trabalho está relacionado com a pesquisa que investigou a presença das teorias semióticas entre os especialistas em Ciência da Informação no Brasil e teve como objetivo geral a identificação e a análise das teorias semióticas presentes na literatura brasileira de Ciência da Informação. No entanto, o objetivo deste artigo é apresentar, de maneira descritiva, os autores, as fontes de publicação e a filiação teórica de acordo com a abordagem semiótica. Uma futura publicação tratará de versar sobre os trabalhos mais relevantes do ponto de vista semiótico que vem sendo produzidos nos últimos anos.

A pesquisa considerou que a linha semiótica ainda requeria um tratamento amplo, sistematizado e que envolvesse vários elementos do conjunto de processos de informação, da análise à apropriação. No que tange às atividades dos pro-

cessos de análise de informação, merecia investigar os conceitos semióticos que presidem o tratamento da informação. A hipótese central da referida pesquisa sustentou que as teorias semióticas colaboram à constituição epistemológica da Ciência da Informação no Brasil.

A seguir, serão apresentados os procedimentos metodológicos adotados para compilação do material e na sequência serão descritos os resultados da análise dos autores e fontes principais.

## 2. Metodologia

De acordo com o delineamento adotado, entendemos a pesquisa como bibliográfica. Para a consecução dos objetivos elencados foi necessário recorrer ao tipo conhecido como pesquisa bibliográfica, na medida em que se buscou esclarecer problemas examinando a literatura existente sobre o assunto, identificando e consultando teses, dissertações, livros e artigos relevantes ao tema. Esse conjunto de materiais relevantes também pode ser chamado de corpus. Uma pesquisa bibliográfica levanta a literatura pertinente ao assunto, organizando e sistematizando o material para desenvolver uma discussão. Conforme sustenta Dieterich (1999, p. 178), um objetivo e um problema de pesquisa podem ser alcançados mediante a consulta exclusiva a bibliografia. Desse modo, não consideramos a pesquisa bibliográfica como meramente instrumental para atingir uma pesquisa que requer coleta de dados. A pesquisa bibliográfica é um procedimento válido para se confrontar hipóteses, consistindo, então, em um tipo de pesquisa legítima.

A qualidade da pesquisa bibliográfica de, unicamente, verificar, refutar ou retificar uma hipótese, foi a aceção adequada a esta pesquisa. As técnicas aplicadas a pesquisas bibliográficas presumem as etapas: identificação das fontes, localização e obtenção do material, leitura do material, compilação e fichamento com crítica sobre a documentação e bibliografia, construção lógica do trabalho e redação do texto.

O universo da pesquisa bibliográfica foi constituído pelos materiais cujos assuntos se correlacionaram aos temas da pesquisa, e divididos em três grupos: a) artigos de periódicos; b) trabalhos em anais de eventos e c) livros/teses/dissertações. A eleição das fontes de pesquisa relacionadas à Ciência da Informação foram orientadas pelos temas/palavras-chave: semiótica, teorias semióticas, teorias semiológicas, epistemologia da Ciência da informação, organização da informação, organização e representação do conhecimento, e outras expressões correlacionadas. O levantamento foi realizado entre os anos de 2013

e 2014, e ampliado o prazo de conclusão, recuperando assim documentos no ano de 2015.

Para localizar os artigos de periódicos, primeiramente, selecionamos os principais periódicos da Ciência da Informação, no âmbito nacional recorrendo a base de dados BRAPCI (<http://www.brapi.ufpr.br>), além dos periódicos nacionais do campo das teorias semióticas, especificamente *Trans/Form/Ação*, *Cognitio* e *Cognitio-Estudos*, registrando os dados através dos sites. A pertinência dos periódicos justifica-se pela importância acadêmico-científica que possuem em cada uma das comunidades envolvidas, seja entre os pesquisadores em Ciência da Informação e nas teorias semióticas. Em um segundo momento, consultamos os sumários dos volumes e números dos periódicos obtidos e, em seguida, os resumos dos artigos de maior interesse. Com os anais de eventos, adotamos as mesmas estratégias do levantamento realizado com os artigos de periódicos. Esta fonte de informação recolhe trabalhos em sua maioria originais e nem sempre são publicados posteriormente na forma de artigos. Para tanto, consultamos diretamente os sites dos congressos: Portal de eventos da Ancib (<http://enancib.ibict.br/index.php/enancib/index/>), CIFORM (<http://www.ciform2015.ufba.br/>) e ISKO Brasil (<http://isko-brasil.org.br/>). As teses e dissertações referentes às teorias semióticas na Ciência da Informação foram localizadas e obtidas consultando o *site* dos programas de pós-graduação, os quais oferecem uma lista dos trabalhos defendidos.

Na sequência, resumiremos os principais resultados relacionados à análise dos autores e suas vinculações teóricas, segundo os propósitos e categorias utilizadas.

## 3. Autores e teóricos da semiótica no campo da Ciência da Informação no Brasil

Esta seção tem por finalidade analisar os autores, as fontes de publicação e a filiação teórica de acordo com a abordagem semiótica. A fundamentação semiótica, que contempla as teorias peirceanas e não peirceanas, pôde ser constatada no contexto da Ciência da Informação brasileira pelos artigos e trabalhos publicados, assim como pela contagem dos autores. Porém, seria uma visão ainda reduzida para conhecer as formas de inserção dos conceitos semióticos, pois há, certamente, um enraizamento histórico das teorias com determinadas escolas. Por exemplo, as abordagens semióticas classificadas como não peirceanas estão mais evidentes, segundo nossa análise, em razão da tradição do pensa-

mento francês da análise documental de conteúdo na Ciência da Informação no Brasil. A escola de documentação francesa recebeu maior influência da Linguística, como notamos na análise dos trabalhos de Gardin (Almeida, 2011).

A noção de presença é fundamental para conhecer, em um primeiro momento, o surgimento de um assunto na literatura e corresponde à perspectiva da pesquisa qualitativa. A presença de um fenômeno é mais importante que a sua frequência em pesquisas qualitativas. De modo geral, nas pesquisas qualitativas que requerem coleta de dados discursivos, um único depoimento declarado por uma pessoa entrevistada é tratado como fundamental, dada a sua relevância em um contexto, enquanto nem sempre a frequência de um fenômeno ou de uma ideia em um contexto – isto é, a taxa de ocorrência (quantidade de vezes que ocorre) –, será mais importante que a qualidade de sua existência (presença).

Em síntese, o que interessou foi saber fundamentalmente quem são os autores e as ideias semióticas por eles professadas, e a quantificação que se segue é apenas um meio para organizar a exposição, pois não foi nosso interesse identificar padrões de citação, apesar de serem arrolados comentários nesse sentido para identificação do núcleo de autores. Cabe acrescentar que, pelo viés quantitativo, notamos pouca representatividade dos trabalhos de Semiótica, como será observado na sequência. Sendo assim, foram recuperados 115 documentos, divididos da seguinte forma: 49 artigos selecionados provenientes revistas da área e de fora da área (representa 42,60% do total); 41 trabalhos de eventos (35,65) e 25 teses (21,73). Como visto, a maior parte dos trabalhos foi veiculada em artigos científicos, seguida dos anais de congresso.

Nesta análise, destacamos os periódicos, os eventos nacionais e os autores dos trabalhos relacionados à Semiótica. Com esta simples análise, não foi possível aprofundarmos nos temas tratados pelas correntes teóricas. Por um lado, no que tange aos artigos de periódicos nacionais da Ciência da Informação, a lista dos autores mais citados foi produzida partindo dos materiais selecionados. Sendo assim, foram recuperados 23 artigos publicados em 10 periódicos diferentes. Os periódicos com maior destaque foram Encontros Bibli (21,7%) e Informação e Sociedade (17,4%), como podemos notar a seguir, representando juntos 39,1% dos artigos, conforme tabela I.

Assim como as iniciativas de Encontros Bibli, outros periódicos têm fomentado nos últimos anos edições específicas sobre novas temáticas no campo da Ciência da Informação, tal como a Se-

miótica. Supomos que seja uma forma de contemplar temas antes tratados marginalmente pelas revistas tradicionais. Esse fenômeno potencializa a veiculação de ideias e o aprofundamento de um tema recente na literatura brasileira. Devemos salientar que as revistas foram consultadas desde o primeiro número até o ano de 2014. Considerando este conjunto de trabalhos, pudemos elencar os autores com maior número de publicações, em temas estritamente semióticos, na seguinte ordem: Monteiro, S.D. (5 artigos), Almeida, C.C. (4), Moura, M.A. (3), Lara, M.L.G. (3) e Gomes, H. F. (2).

<i>Periódicos</i>	<i>N.</i>	<i>%</i>
Revista Digital de Biblioteconomia & Ciência da Informação	1	4,3
Liinc em Revista	1	4,3
Encontros Bibli	5	21,7
TransInformação	2	8,7
Perspectivas em Ciência da Informação	1	4,3
DataGramaZero	3	13
Ciência da Informação	3	13
Informação e Informação	2	8,7
Informação e sociedade: Estudos	4	17,4
Ponto de Acesso	1	4,3
Total	23	99,7

*Tabela I. Periódicos Nacionais em Ciência da Informação (Almeida, 2016)*

Para a análise de citação referente aos artigos recuperados, seguimos os seguintes procedimentos. Primeiro, a partir dos artigos buscamos todas as referências bibliográficas desses trabalhos e as organizamos em ordem alfabética de autor. Com base nas referências utilizadas pelos trabalhos da lista dos artigos de periódicos nacionais, elaboramos uma tabela na qual arrolamos os autores e o número de vezes que foram citados nos artigos selecionados.

Na sequência, identificamos os autores mais citados e os organizamos por quantidade de citações, considerando todos os textos, dos mais citados aos menos citados. A seguir, apresentamos a tabela II.

<i>Autor</i>	<i>N.</i>	<i>Autor</i>	<i>N.</i>
Santaella, L.	35	Monteiro, S.D	5
Peirce, C. S.	30	Izquierdo Arroyo, J. M.	3
Lara, M. L. G. de	18	IBRI, I. A	3
Eco, U.	14	Azevedo Netto, C. X.	3
Almeida, C. C.	11	Saussure, F. De	2
Moura, M.A	10	Tellefsen, M.M	2

Mai, J.-E.	8	Silveira, L.F.B	2
Pinto, J.	8	Blikstein, I.	2
Nöth, W.	8	Merrel, F	2
Tellefsen, T. L	5	Machado, I.	1

Tabela II. Autores mais citados (Almeida, 2016)

Em terceiro lugar, separamos os autores da Ciência da Informação dos autores vinculados à Semiótica e os organizamos em tabelas distintas, por ordem de citação. O objetivo foi verificar quais os autores com maior número de citações. Nesta fase, os autores citados aparecem conjuntamente, sem a distinção da vinculação teórica e a origem dos autores, se são de Semiótica ou de Ciência da Informação.

Como etapa seguinte, retornamos aos artigos analisados para verificar, com base na lista de referências, quais deles citaram os autores identificados na Tabela II. Além disso, elaboramos as tabelas III e IV com a finalidade de separar os autores pela vinculação disciplinar. Com isso, pretendemos conhecer os autores de Semiótica geral e os autores da Ciência da Informação com maior número de citações.

Com estes dados, foi elaborado uma rede de citação com o auxílio do *Software Ucinet* (3). A rede de citação representa o universo de autores citados a partir da bibliografia dos autores citantes, consultada e analisada através artigos de periódicos nacionais da Ciência da Informação. Os autores citantes foram representados pelos círculos azuis. Os autores citados foram representados pelos círculos de cor vermelha. Os círculos de cor verde representam os autores que foram citantes e citados ao mesmo tempo. Merece destaque a observação de que quanto maior a espessura da flecha, mais vezes o autor foi citado.

Na página seguinte, apresentamos a rede de autores citados e citantes construída a partir dos autores mencionados anteriormente.

Autor	N. citações
Lara, M. L. G. de	18
Almeida, C. C	11
Moura, M. A.	10
Mai, J.-E.	8
Tellefsen, T. L.	5
Monteiro, S. D.	5
Azevedo Netto, C.	3
Tellefsen, M. M.	2

Tabela III. Autores da Ciência da Informação citados (Almeida, 2016)

Autor	N. Citações
Santaella, L.	35
Peirce, C. S.	30
Eco, U.	14
Pinto, J.	8
Nöth, W.	8
Izquierdo A, J. M.	3
Ibri, I. A.	3
Saussure, F. de	2
Silveira, L. F. B	2
Merrel, F.	2
Machado, I.	1

Tabela IV. Autores de Semiótica Geral citados (Almeida, 2016)

Dos autores citados, verificamos duas características: a primeira, a participação efetiva de autores da abordagem peirceana em Semiótica geral (Peirce, Santaella, Pinto, Noth etc.), a segunda, entre os autores da Ciência da Informação citados destacamos os envolvidos com as perspectivas semióticas e linguísticas. Quando observamos a raiz teórica dos autores em Semiótica geral, notamos claramente uma quantidade expressiva de citações no campo da semiótica peirceana em detrimento das abordagens não peirceanas.

Contudo, é com os autores citantes e citados que identificamos um núcleo de especialistas em semiótica atuantes na Ciência da Informação no Brasil. Nesse sentido, figuram os autores Lara, Moura, Monteiro, Azevedo Neto e Almeida. Isso mostra que a comunidade de especialistas em Semiótica na Ciência da Informação no Brasil é relativamente pequena quando comparada aos especialistas atuantes nos núcleos de pesquisa, como organização do conhecimento, gestão do conhecimento ou produção científica. Presumimos que os temas semióticos tenham sido direcionados a aplicações, tal como as questões envolvidas no processo de análise da imagem e análise semiótica de sites.

Há um predomínio de publicações dos autores citados e o maior número de obras foi publicado entre os anos 1990 e 2010. Podemos dizer assim que os anos 1990 foram os que marcaram o surgimento dos estudos semióticos na Ciência da Informação brasileira, primeiramente, tendo seus conceitos traduzidos e compilados através das ideias linguísticas, depois emancipando-se destes e constituindo-se em uma vertente própria revisitando os clássicos da matéria.



boração de autores da área. Dentre os artigos selecionados, 26,92% (7 artigos) foram de autoria de Silveira, 11,53% (3 artigos) de Ibri, e 7,69% (2 artigos) de Santaella. Os demais autores, um total de 17, contaram apenas com um trabalho, sendo que três artigos foram escritos em parceria.

Os anais do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) concentram a maior parte dos trabalhos de semiótica na Ciência da Informação em literatura cinzenta. Dos ENANCIBs, foram analisadas 15 edições (1994, 1995, 1997, 2000, 2003, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013 e 2014), com um total de 2747 trabalhos. Desse total, apenas 21 trabalhos (0,76%) mantêm relação com a Semiótica, seja de maneira específica ou indiretamente sob o arranjo temático geral com linguagem, comunicação e mediação.

Esses 21 trabalhos contaram com a participação de 25 autores, mas apenas 4 autores com mais de um trabalho. Os principais autores foram Almeida (28,56%, 6 trabalhos), Lara (9,52%, 2 trabalhos), Moura (9,52%, 2 trabalhos) e Guimarães (9,52%, 2 trabalhos) e outros autores com apenas um único trabalho cada (Barros, Café, Capolina, Capurro, Coelho, Orrico, Cordeiro, Evangelista, Farias, González Gomez, Gracioso, Assis, Manini, Nakayama, Novellino, Pato, Suescun, Scheiner, Tautain, Weiss, Brascher).

Contudo, restringimos a seleção dos trabalhos que devem tratar de maneira integral de um dos problemas da semiótica e não apenas mencionar conceitos gerais, tal como linguagem. Em razão disso, o número de trabalhos foi reduzido. Observamos que há maior incidência da relação da Semiótica com os temas organização do conhecimento, análise documental de conteúdo, linguística, ciência da informação e epistemologia. Esse dado confirma o encontrado anteriormente sobre os artigos científicos, isto é, a aproximação à semiótica dá-se de maneira teórica e vinculada fundamentalmente à organização do conhecimento. No entanto, necessitamos aprofundar a definição de classes e o lugar teórico-conceitual de cada um destes trabalhos, pois acreditamos que o impacto teórico não é medido unicamente pelo índice de citação, embora seja um indicador válido do potencial de um conceito em um universo.

Os quatro autores citados (Almeida, Lara, Moura e Guimarães) colaboraram com 57,12% dos trabalhos no campo da Semiótica nos eventos do ENANCIB. Confrontando com os demais autores, podemos inferir que o baixo número de trabalhos geral e a concentração dos trabalhos em poucos autores demonstram que a Semiótica, enquanto objeto de estudo, está em plena fase

de desenvolvimento na Ciência da Informação, isto é, a implementação de métodos de análise documental de conteúdo e análise de imagem de base semiótica, bem como a criação de algoritmos para processamento da linguagem que se voltam a Semiótica, são considerados assuntos emergentes. Uma observação adicional é que o número de trabalhos é menor que o número de autores, pois muitos trabalhos foram elaborados em coautoria. Além disso, houve um crescimento dos trabalhos publicados na última década, ainda que isto não represente de maneira alguma uma tendência ascendente.

No Encontro Nacional de Ensino e Pesquisa em Informação (CINFOR), realizado em Salvador, identificamos resultados parecidos com os do ENANCIB, pois apenas quatro trabalhos relacionados à Semiótica, os quais podemos segmentar em dois grupos: um aplicado (um trabalho que aplica as noções semióticas para análise de interfaces digitais) e outro teórico (dois trabalhos que versam sobre problemas da semiótica e um a respeito de sua inserção na formação profissional).

No que tange ainda os anais de eventos, devemos sublinhar outro congresso nacional que mantém contribuições em Semiótica: os encontros da ISKO-Brasil. Nossa avaliação dos anais ISKO-Brasil 2011, 2013 e 2015 selecionou 17 trabalhos relacionados com a pesquisa. Em 23,52% (4 trabalhos) dos trabalhos Moraes figura como autor, em 17,64% (3 trabalhos) aparece Moura, e 11,76% (2 trabalhos) dos trabalhos são de autoria de Alves, Sabbag ou Fujita. Surgiram também mais 14 autores com apenas um trabalho cada (Almeida, Arboit, Bufrem, Freitas, Assis, Azevedo Netto, Brascher, Candido, Guim, Lima, Medeiros, Ortega, Souza J. e Souza R.). Nesses trabalhos há uma notória presença da semiótica não peirceana - fundamentalmente de matriz greimasiana -, haja vista os estudos de Moraes, Alves e Sabbag, seguidos das contribuições na linha da semiótica peirceana refletidas nos trabalhos de Moura e Almeida.

No caso das monografias, acreditamos ser importante indicar, a título ilustrativo, as produções de teses e dissertações pelos programas de pós-graduação brasileiros no campo da Ciência da Informação. Foi constatado que os trabalhos de pesquisa dessa natureza, assim como as demais publicações, sofreram um crescimento na última década, atestando nossa leitura que sugere uma década da Semiótica nos estudos da informação no Brasil, em comparação a outros períodos. De 1996 a 2005 foram apresentadas 3 teses/dissertações, e nos últimos anos, de 2006 a 2015, foram defendidos 22 trabalhos. Isso se explica, em

parte, pelo fato de que até 2004 existiam no Brasil apenas quatro programas de doutorado em Ciência da Informação, mas nos anos que se seguiram foram instalados mais cinco cursos em várias regiões do país, totalizando no ano de 2015 nove programas de doutorado. As teses defendidas nestes programas refletem a diversificação das linhas de pesquisa além da consolidação das já existentes. Assim, as abordagens semióticas puderam ser desenvolvidas em termos qualitativos e quantitativos.

De forma geral, quando comparadas as contribuições dos autores brasileiros com as investigações análogas em outros países, notamos que as bases teóricas peirceana e não peirceana estão presentes também nestes contextos. Em outras palavras, os teóricos citados, especialmente Peirce, Eco e Saussure, são utilizados com o mesmo nível de profundidade em estudos consagrados ao tema nos últimos anos, em especial, nos trabalhos de Izquierdo Arroyo (1990, 1992, 1993, 1995), Izquierdo Alonso (2000, 2004), Izquierdo Alonso e Izquierdo Arroyo (2014), Mai (1997a, 1997b, 2001, 2000), Thellefsen (2002, 2003, 2004), Thellefsen e Thellefsen (2004), Raber e Budd (2003), Friedman e Thellefsen (2011). Em outras palavras, autores representantes das escolas da Ciência da Informação tão diferentes como a espanhola, a americana e a dinamarquesa. A despeito da distância entre as comunidades científicas, o grupo de pesquisadores brasileiros citados demonstra o mesmo nível de profundidade nos estudos semióticos recebido em âmbito internacional.

Seria importante averiguar o conteúdo dessas teses e dissertações, bem como sistematizar os pontos de convergência com as teorias semióticas peirceana e não peirceana. Esse trabalho poderá ser realizado em uma nova pesquisa em que investigaremos os objetos e os métodos das pesquisas semióticas. Contudo, esta análise não foi levada a cabo neste artigo.

No que se refere aos temas, notamos uma preferência pela semiótica peirceana, em detrimento da semiótica greimasiana, semiologia, semiótica da cultura e semiótica da imagem. Embora os dados não visem estimar uma porcentagem entre as abordagens semióticas.

Com base nos levantamentos anteriores, nas tabelas com indicações dos autores mais destacados, construímos um quadro, que ainda merece aperfeiçoamento, pois os autores indicados participam ou utilizam mais de uma abordagem semiótica para dar conta de seus problemas de pesquisa, de tal modo que circunscrever linhas rígidas é uma tarefa temerária, mas é um estágio necessário de formalização.

<i>Teorias e abordagens</i>	<i>Teóricos em semiótica geral</i>	<i>Autores brasileiros em semiótica</i>
<i>Teoria Semiótica Peirceana/Semiótica geral</i>	Peirce, Santaella, Noth, Ibri, Silveira, Merrel, Pinto, Blikstein	Lara, Moura, Monteiro, Assis, Amorim, Gomes Almeida, Azevedo Netto, Barros e Café
<i>Teorias Semióticas Estruturalistas/S emântica narrativa/Semiótica discursiva</i>	Saussure, Hjelmslev, Eco, Greimas, Barthes,	Lara, Moraes, Guimarães, Tálamo, Guarido
<i>Teorias Semióticas Híbridas/Semiótica da Cultura/Semiologia da imagem</i>	Eco, Lotman, Machado, Barthes	Smit, Pato, Toutain, Silva, Fujita, Bizello, Elliott, Mádio, Farias e Almeida

*Quadro I. Linhas em Semiótica na Ciência da Informação no Brasil (Almeida, 2016)*

O referido quadro ainda requer complementação, pois deveríamos indicar ainda em uma última coluna o conjunto de temas apontados pela literatura que conectam pontualmente a Semiótica com a Ciência da Informação. Desse modo, teremos condições de obter um mapa ou diagrama dos principais assuntos investigados que estabelecem a interface com a Semiótica.

#### 4. Considerações finais

Nesta última seção ofereceremos um balanço com as principais conclusões. Considerando os objetivos da pesquisa geral, os quais trataram da análise conceitual e revisão dos trabalhos em periódicos e anais de evento, bem como da identificação dos conceitos semióticos utilizados e principais teóricos, estimamos que estes foram alcançados substancialmente, de modo a facilitar a composição de um diagrama do pensamento da Ciência da Informação brasileira sobre a Semiótica.

Com a classificação adotada das abordagens peirceana e não peirceana foi possível compreender os vínculos teóricos dos autores, bem como os fundamentos utilizados, sinalizando a origem conceitual. Desta forma, respondemos a uma das questões da pesquisa que tratava de apontar a necessidade de saber quais abordagens e teorias semióticas estariam presentes na Ciência da Informação brasileira.

Também aludimos a um núcleo de pesquisadores em Semiótica, relativamente pequeno em comparação à extensão do campo de aplicações e à abrangência de sua teorização.



No que tange a Semiótica Peirceana, a despeito de seu aparecimento tardio na Ciência da Informação em comparação com as abordagens não peirceanas (linguísticas, culturais e semiológicas), esta mostrou-se coerente e majoritária nos estudos na área. Contudo, a hipótese central da pesquisa, que supôs que as teorias semióticas colaborariam para a constituição epistemológica da Ciência da Informação no Brasil, não pôde ser corroborada, pois a quantidade de trabalhos, apesar da riqueza conceitual, não permite ainda constituir um corpo de conhecimentos consistente e que molde a estruturação conceitual do campo. Isso talvez possa ser alcançado nos próximos anos, quando a tendência na produção em Semiótica identificada mostrar-se constante ou, na melhor das hipóteses, ser ampliada.

Este trabalho também acrescenta elementos à análise realizada por Barros e Café (2012) sobre a interdisciplinaridade dos estudos da semiótica na Ciência da Informação.

Os autores arrolados aqui revelaram indícios de uma mudança de paradigma - pelo menos no âmbito discursivo -, em que há um deslocamento de uma perspectiva linguística para uma Semiótica geral, especialmente na compreensão das ferramentas de organização do conhecimento e dos processos de organização da informação.

## Notas

- (1) Segundo Brascher e Café (2008), organização da informação é um processo que envolve a descrição física e de conteúdo de objetos informacionais. Seu objeto de trabalho é formado por registros da informação, mas em seu caráter individual. O produto da organização da informação é a representação da informação, na forma da descrição física e descrição temática, ambas resultantes dos processos correlacionados: catalogação, classificação, redação de resumos e indexação. A organização do conhecimento, por sua vez, é um processo de construção de modelos de representações de conhecimento. Tem por objeto o mundo das ideias, convertido em conceitos ou aglomerado de conceitos, os quais devem ser analisados para verificação de seus vínculos dispostos em um domínio de conhecimento.
- (2) Por processos da informação ou processos informacionais entendemos *latu sensu* o conjunto de processos resultantes de ações, etapas e fases, que fazem com que - especificamente no âmbito da Ciência da Informação - a informação possa ser criada, comunicada e permitir novas apropriações. Desse modo, incluímos os processos de criação (intelectual), produção (registro e edição), seleção, aquisição, tratamento (físico e temático), organização, armazenamento, recuperação, distribuição, mediação, comunicação, utilização e apropriação, realizados nos mais diversos sistemas e ambientes de informação. Cunha e Cavalcanti (2008, p. 293-294) definem processo de informação (*information process*) como o "processo de comunicação que conduz a um conhecimento mais preciso", mas fazem a seguinte distinção: o processo de comunicação considera os aspectos tecnológicos enquanto no processo informativo a ênfase está

na substância das mensagens, ou seja, o conteúdo. A partir disso, também podemos entender como processos da informação os levados a cabo pela cadeia ou ciclo documental, como as operações necessárias para a utilização dos documentos e os pertencentes a cadeia ou ciclo informacional, isto é, "fases pelas quais a informação passa, desde sua criação até sua disseminação." (Cunha, Cavalcanti, 2008, p. 62). A cadeia informacional pressupõe a cadeia documental, pois esta lhe fornece os processos essenciais para a disponibilização da informação. Nesse sentido, os processos elencados por Barreto (2001) nos fluxos interno e fluxos extremos da informação são considerados processos da informação, em outras palavras, processos pelos quais a informação é submetida. As atividades do campo da organização da informação e do conhecimento, bem como as ações que permitem a comunicação da informação em sistemas de informação são consideradas processos informacionais no contexto da Ciência da Informação.

(3) <https://sites.google.com/site/ucinetsoftware/downloads>.

## Agradecimentos

Agradecemos imensamente o CNPq pelo auxílio recebido e registrado com o processo 406021/2012-0 que permitiu a conclusão desta pesquisa.

## Referências

- Almeida, C. C. (2010). Sobre o pensamento de Peirce e a organização da informação. // Anais do XI Encontro Nacional de Pesquisa em Ciência da Informação: Rio de Janeiro, 2010. Rio de Janeiro: PPGCI/UNIRIO, 2010. <http://enancib.ibict.br/index.php/enancib/xienancib/paper/view/3412/2538> (2016-07-15).
- Almeida, C. C. (2011). Elementos de linguística e semiologia na organização da informação. São Paulo: Cultura Acadêmica, 2011.
- Almeida, C. C. (2012). Mediação como processo semiótico: em busca de bases conceituais. // Tendências da Pesquisa Brasileira em Ciência da Informação. 5 (2012) 1-18.
- Almeida, C. C. (2014). Indícios da presença de fundamentos semióticos na literatura da ciência da informação. // Scire. 20:1 (ene.-jun. 2014) 65-71.
- Almeida, C. C. (2016). A presença das teorias semióticas na ciência da informação brasileira. Marília: s. n., 2016. 128 p. (Relatório de pesquisa relativo ao auxílio recebido através do edital MCTI/CNPq/MEC/CAPES Nº 18/2012).
- Barreto, A. A. (2001). A informação em seus momentos de passagem. // DataGramaZero: Revista de Ciência da Informação. 2:4. [http://dgz.org.br/ago01/Art\\_01.htm](http://dgz.org.br/ago01/Art_01.htm) (2009-09-27).
- Barros, C. M. de; Café, L. M. A. (2012). Estudos da semiótica na ciência da informação: relatos de interdisciplinaridades. // Perspectivas em Ciência da Informação. 17: 3 (jul./set. 2012) 18-33. <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1501/1050> (2014-04-20).
- Brascher, M.; Café, L. (2008). Organização da informação ou organização do conhecimento? // Anais do IX Encontro Nacional de Pesquisa em Ciência da Informação: São Paulo, 2008. São Paulo: ANCIB, 2008. <http://www.enancib2008.com.br/> (2008-12-15).
- Cunha, I. M. R. F. Coord. (1989). Análise documental: considerações teóricas e experimentações. São Paulo: FEBAB, 1989.
- Cunha, M. B.; Cavalcanti, C. R. O. (2008). Dicionário de biblioteconomia e arquivologia. Brasília: Briquet de Lemos Livros, 2008.

- Dieterich, H. (1989). *Novo guia para pesquisa científica*. Blumenau: EDFURB, 1999.
- Flusser, V. (2007). *Língua e realidade*. 3. ed. São Paulo: Anablume, 2007.
- Friedman, A.; Thellefsen, M. (2011). Concept theory and semiotics in knowledge organization. // *Journal of Documentation*. 67:4 (2011) 644-674.
- Izquierdo Alonso, M.; Sánchez Domínguez, C. (2011). Los estudios de resumen documental en las ciencias de la documentación: un recorrido histórico desde sus orígenes hasta las teorías actuales. // *Revista de Lingüística y Lenguas Aplicadas*. 6 (2011) 209-227.
- Izquierdo Alonso, M. (2000). Nuevos enfoques en el estudio del tratamiento documental de contenido desde los presupuestos de las ciencias del lenguaje. // *Scire*. 6:1 (ene.-jun. 2000) 143-163.
- Izquierdo Alonso, M. (2004). Nuevos retos en el análisis documental de contenido: gestión de la forma documental del contenido. // *Scire*. 10:1 (ene.-jun. 2004) 31-50.
- Izquierdo Alonso, M.; Izquierdo Arroyo, J. M. (2014). Entrevista a José María Izquierdo Arroyo, realizada por Mónica Izquierdo Alonso em Alcalá de Henares (Madrid), junio de 2014. // *Scire*. 20:1 (ene.-jun. 2014) 109-116.
- Izquierdo Alonso, M.; Moreno Fernandez, L. M. (2009). *El resumen documental: un reto didactico*. Madrid: ANABAD, 2009.
- Izquierdo Arroyo, J. M. (1990). *Esquemas de lingüística documental*. Barcelona: Promociones y Publicaciones Universitarias, 1990. Tomos I a III.
- Izquierdo Arroyo, J. M. (1990). La ciencia de la búsqueda documental secundaria. // *Documentación de las Ciencias de la Información*. 13 (1990) 87-111.
- Izquierdo Arroyo, J. M. (1992). Cuatro trabajos en curso. // *Documentación de las Ciencias de la Información*. 15 (1992) 35-65.
- Izquierdo Arroyo, J. M. (1993). De la semiótica del discurso a la semiótica documental. // Moreno González, J. A. (eds.). *Aplicación de las ciencias del texto al resumen documental*. Madrid: Universidad Carlos III, 1993. 199-216.
- Izquierdo Arroyo, J. M. (1995). *La organización documental del conocimiento*. Madrid: Tecnidoc, 1995.
- Mai, J-E. (1997a). The concept of subject in a semiotic light. // Schwarts, C.; Rorvig, M. (ed.). *Digital collections: implications for users, funders, developers and maintainers*. Medford, NJ: Information Today, 1997a. 54-64.
- Mai, J-E. (1997b). The concept of subject: on problems in indexing. // McIlwaine, I. C. (ed.). *Knowledge organization for information retrieval: 6th International Study Conference on Classification Research*. The Hague: FID, 1997b. 60-67.
- Mai, J-E. (2000). *The subject indexing process: an investigation of problems in knowledge representation*. Austin: Faculty of Graduate School of Library and Information Science, The University of Texas at Austin, 2000. Tesis doctoral.
- Mai, J-E. (2001). Semiotics and indexing: an analysis of the subject indexing process. // *Journal of Documentation*. 57: 5 (Sep. 2001) 591-522.
- Monteiro, S. D. (2006). Semiótica peirciana e a questão da informação e do conhecimento. // *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*. 2: número especial (2º Sem. 2006) 43-57. <http://www.periodicos.ufsc.br/index.php/eb> (2006-10-15).
- Moura, M. A. (2006). Ciência da informação e semiótica: conexão de saberes. // *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*. 2: número especial (2º Sem. 2006) 1-17. <http://www.periodicos.ufsc.br/index.php/eb> (2006-10-15).
- Moura, M. A. (2007). Significa ou signi-vai? as teorias da significação no campo da Ciência da Informação. // Reis, A. S. ; Cabral, A. M. (org.). *Informação, cultura e sociedade: interlocuções e perspectivas*. Belo Horizonte: Novatus, 2007. 61-80.
- Nöth, W. (2005). *A semiótica no século XX*. 3 ed. São Paulo: Anablume, 2005.
- Raber, D.; Budd, J. M. (2003). Information as sign: semiotics and information science. // *Journal of Documentation*. 59:5 (2003) 507-522.
- Smit, J. W. Coord. (1989). *Análise documentária: a análise da síntese*. 2. ed. Brasília: IBICT, 1989.
- Thellefsen, T. L. ; Thellefsen, M. M. (2004). Pragmatic semiotics and knowledge organization. // *Knowledge Organization*. 31:3 (2004) 177-187.
- Thellefsen, T. L. (2002). Semiotic knowledge organization: theory and method development. // *Semiotica*. 142:1/4 (2002) 71-90.
- Thellefsen, T. L. (2003). Pragmaticism and the role of terminology. *Impact: an electronic journal on formalisation in text, media and language*, (abr. 2003). <http://www.impact.hum.auc.dk> (2007-05-05).
- Thellefsen, T. L. (2004). Knowledge profiling: the basis for knowledge organization. // *Library Trends*. 52:3 (winter 2004) 507-514.
- Vieira, F. S. (2015). *Semiótica peirciana e Ciência da Informação: abordagens na publicação brasileira*. Florianópolis: Curso de Biblioteconomia, Universidade Federal de Santa Catarina, 2015. Trabalho de Conclusão de Curso.

---

Enviado: 2016-04-05. Segunda versão: 2016-07-20.  
 Aceptado: 2016-09-15.

---

---

# Análisis de temas emergentes a través de Twitter

*Analysis of emerging subjects in Twitter*

---

José Luis ALONSO BERROCAL, Carlos G. FIGUEROLA, Ángel F. ZAZO RODRÍGUEZ

Departamento de informática y automática, Universidad de Salamanca.  
{berrocal | figue | zazo}@usal.es

## Resumen

El análisis de temas emergentes en las redes sociales se aplica para conocer las opiniones que expresan usuarios individuales, para controlar actividades y actos de asociaciones, analizar las campañas de los políticos o estudiar el impacto de campañas publicitarias por parte de las empresas. Para la detección de dichos temas se aplicó el algoritmo Latent Dirichlet Allocation a un conjunto de perfiles del ámbito de la información y documentación, con el fin de conocer los temas que se tratan en dichos grupos y para evaluar si el sistema de detección es fiable. El sistema funciona correctamente y proporciona resultados fiables.

**Palabras clave:** Twitter. Temas de actualidad. Detección de temas. Latent Dirichlet Allocation.

## Abstract

Analysis of emerging issues in social networks applies to the views expressing individual users, to control activities and acts of associations, analyze political campaigns or study the impact of advertising campaigns by companies. For detection of these issues the algorithm Latent Dirichlet Allocation shall apply to a set of profiles in the field of information and documentation, in order to know the topics covered in these groups and to assess whether the detection system is reliable. The approach works correctly, and provides reliable results.

**Keywords:** Twitter. Trending topics. Topic detection. Latent Dirichlet Allocation.

## 1. Introducción

El crecimiento y diversificación de los recursos informativos, asociado a la popularización de la web ha propiciado un cambio en los hábitos de consumo de información por parte de los ciudadanos. En 2010 la OCDE publica el informe *The Evolution of News and the Internet* (Wunsch-Vincent y Vickery, 2010) en el que se documenta que más de la mitad de la población de los países occidentales se informa habitualmente a través de las ediciones digitales de los periódicos. Ese mismo año el Pew Research Center registra, por primera vez desde el surgimiento de internet, que el número de lectores de la prensa digital supera al de las ediciones en papel de los medios.

Por otra parte, debe considerarse que el cambio descrito ha estado necesariamente acompasado con la creciente implantación de las tecnologías de acceso (más del 68% de los hogares españoles cuenta con acceso a internet), así como con la difusión de nuevas aplicaciones, productos o servicios informativos, entre los que se incluyen desde hace algo más de una década las redes sociales (Pinholster y Ham, 2013).

El uso de las redes sociales se ha extendido entre aquellos agentes más directamente implicados en la difusión de la información. Así para aquellos actores sociales que habitualmente han

ejercido el papel de fuente informativa, las redes se han convertido en herramientas útiles para comunicarse de manera directa y sin intermediarios con aquellos colectivos a los que se desea hacer llegar un mensaje. Redes como Twitter permiten a individuos y organizaciones saltarse los filtros tradicionales y acceder de manera directa a lectores, consumidores, fans, etc.

El estudio de los llamados "Trending Topics" (TT), término que hace referencia a los temas tendencia o a los temas que más se hablan en un determinado momento, se aplica para conocer las opiniones que expresan usuarios individuales, para controlar actividades y actos de asociaciones, analizar las campañas de los políticos o estudiar el impacto de campañas publicitarias por parte de las empresas. Tal es la magnitud que ha adoptado este concepto, que incluso ha traspasado las fronteras del propio mundo de Twitter, para pasar a ser una frase de uso común en los medios de comunicación, siendo noticia aquello que se vuelve tema del momento en esta red social. Esto ha propiciado, por ejemplo, que se creen secciones especializadas en periódicos online que se basen en aquello que es tendencia en Twitter.

Este trabajo está organizado como sigue: en el capítulo siguiente se hace una breve introducción

al estado del arte en este campo, para, a continuación, exponer los principales principios metodológicos seguidos en nuestra investigación. Después se expondrán los principales resultados obtenidos del análisis realizado. Finalmente se ofrecen unas conclusiones.

## 2. Trabajos previos

Topic Detection and Tracking (TDT) es un área que comenzó como track en las Text Retrieval Conferences (TREC) (Allan, 2002). Sin embargo, la aplicación de estas técnicas a Twitter es relativamente reciente. Algunos trabajos reseñables son los de (Sankaranarayanan et al., 2009), que aplicaron técnicas de clustering, o los de Petrovic (Petrović et al., 2010), que además constituyó una colección experimental de tweets que ha sido empleada en otros trabajos, el Edinburgh Twitter Corpus (Petrovic et al., 2010). Mathioudakis y Koudas (Mathioudakis and Koudas, 2010) propusieron un sistema para detectar trending topics a partir de un stream de tweets. También sobre la detección de trending topics han trabajado Shariffi, Hotton y Kalita (Sharifi et al., 2010), al igual que Cheong y Lee (Cheong y Lee, 2009), aunque el trabajo de éstos se centra en la evolución temporal de los trending topics.

La cuestión es, en consecuencia, cómo determinar la similitud entre todos los pares de tweets correspondientes a una entidad dada. La similitud entre dos documentos es uno de los problemas centrales de la Recuperación de Información y puede abordarse de diversas formas.

Podemos trabajar con algoritmos basados en modelos estadísticos. Su objetivo es la identificación de los términos relevantes o términos clave del texto, a través del cálculo de los pesos asociados a dichos términos. A este cálculo se le conoce como ponderación del término. Uno de los más conocidos es el de considerar cada documento (tweet) como un bag of words y aplicar un esquema clásico  $tf \times idf$  (Salton y Buckley, 1988).

Otros algoritmos se basan en la minería de datos que es un campo de las ciencias de la computación que agrupa una serie de técnicas y tecnologías que tienen como misión la extracción de información a partir de patrones contemplados en grandes colecciones de datos estructurados, almacenados en bases de datos. Trata de detectar patrones repetitivos que expliquen el comportamiento de los datos en un determinado contexto. Por ello se trata de una buena técnica para la detección de temas, ya que es capaz de determinar qué elementos son propensos a coocurrir en un conjunto de datos o transacciones. En un contexto de redes sociales y concretamente en Twit-

ter, un elemento  $w$  es cualquier término mencionado en un tweet (sin incluir signos de puntuación, palabras vacías, etc). De esta forma la transacción se corresponde con un tweet y el conjunto de transacciones o datos a analizar son todos los tweets que se producen en un intervalo de tiempo  $T_j$ . El número de veces que un conjunto de términos dado se produce en un intervalo de tiempo se llama soporte, y cualquier conjunto de elementos que cumple un soporte mínimo se llama patrón.

Por otro lado, y entrando en algoritmos de un mayor nivel de dificultad, tenemos los basados en métodos probabilísticos. Según el tipo de salida que produzcan, podemos distinguir varios grupos de algoritmos (Aiello et al., 2013): algoritmos basados en documentos y en funciones.

En el primer grupo destaca el algoritmo LDA. LDA (del inglés, "Latent Dirichlet allocation") es uno de los modelos Bayesianos más conocidos y usados en la detección de TT. Se trata de un modelo de aprendizaje supervisado para caracterizar el contenido de los mensajes (Benhardus and Kalita, 2013). En él, cada documento se representa mediante una bolsa de términos, que son la única variable observada, a través de la cual se intenta extraer los temas tendencia. Es decir, LDA tiene como entrada un conjunto de términos correspondientes a la representación de cada documento de la colección, dando como salida los temas latentes de cada conjunto de palabras, que es lo mismo que los temas latentes de cada documento. Formalmente, un documento se asocia con una distribución multinomial de temas que a su vez son distribuciones multinomiales de palabras. De esta forma podemos definir la siguiente estructura (Blei et al., 2003):

- Palabra, es la unidad básica y discreta de datos. Se define como un conjunto de elemento de un vocabulario de tamaño  $V$
- Documento ( $w = \{w_1, w_2, \dots, w_N\}$ ), es una secuencia de  $N$  palabras
- Cuerpo ( $D = \{w_1, w_2, \dots, w_M\}$ ), es una colección de  $M$  documentos

Al contrario que muchos modelos de agrupación, LDA permite que un documento pueda estar asociado con múltiples temas. Esto se debe a los tres niveles implicados en el algoritmo y en particular, al nodo tema que se muestrea repetidamente en un documento.

Los tweets, sin embargo, son documentos con una serie de características especiales que deberían tomarse en cuenta. Anta y colegas (Anta et al., 2013) mencionan varias de ellas. A las

cuestiones 'clásicas' de la utilización de unigramas o bi, trigramas, etc., o del stemming, en el caso de los tweets debemos añadir los derivados del uso de emoticones, abreviaturas, incluso de un argot propio de ese medio, así como de numerosas anomalías orto-tipográficas. La brevedad de los tweets es otra cuestión importante a considerar (Sriram et al., 2010).

El trabajo de (Mehrotra et al., 2013) utiliza modelos basados en LDA, para trabajar sobre tweets, permitiendo la reconstrucción de clusters y facilitando el análisis de la coherencia de los topics.

El trabajo de Ostrowski (Ostrowski, 2015) analiza diferentes modelos de detección de topics para concluir que el algoritmo LDA, empleado como mecanismo de clasificación, identifica los temas que son dignos de mención y así poder filtrar los mensajes de Twitter.

### 3. Metodología

El trabajo con twitter se ha llevado a cabo trabajando con la API que suministra la red social. Hemos trabajado con la REST API, que nos ofrece el acceso al core de los datos de Twitter. Podemos obtener aproximadamente los últimos 3200 tweets en la primera recogida y a partir de ese momento podemos recoger todo lo nuevo que emite un perfil. Para probar nuestro planteamiento creamos cinco grandes grupos de perfiles relacionados con el campo de la información y documentación, instituciones (26), bibliotecas (33), empresas (25), webs-blogs (14) y profesionales (20), dando un total de 118 perfiles (Tabla II, en apéndice). Hay que indicar que para determinar los perfiles que se iban a analizar se realizó una encuesta a los alumnos del máster en Sistemas de Información Digital, para que nos indicaran aquellos perfiles que consideraban como prioritarios a analizar, y los hemos mantenido tal cual. Con este mecanismo de selección de los perfiles pretendíamos trabajar con perfiles dados por alumnos, que además tienen una procedencia dispar, en cuanto a la titulación de la que proceden. Es evidente que en dicha lista pueden quedar algunos perfiles fueran, que los profesionales pueden echar en falta, pero la visión de los alumnos nos pareció interesante de evaluar.

La serie de resultados obtenida compone el total de los tweets emitidos durante el periodo de recogida para todos los usuarios seleccionados. Esos tweets emitidos incluyen tanto los tweets propios, los escritos directamente por esos usuarios, como sus retweets, tweets escritos por otros, pero difundidos por los usuarios de nuestra lista.

Twitter ofrece una información muy rica para cada tweet descargado. Buena parte de ella es descriptiva: el texto completo del mensaje o datos como el usuario emisor, la fecha y hora de envío, los usuarios mencionados, las direcciones de sitio web citados (url), las referencias a recursos multimedia externos o las etiquetas (hashtags) sirven para caracterizar cada mensaje. Otros campos pueden considerarse más bien metadatos, como los identificadores, las relaciones conversacionales que pueden existir entre los mensajes (respuesta, cita), estados atribuidos al mensaje o número de veces que se ha marcado como favorito, por mencionar algunos. El resto de la información es de tipo administrativo y tiene su papel en la gestión interna de Twitter: preferencias de usuario o atributos para el tratamiento de los mensajes. Toda esta información se recibe en ficheros de texto plano, estructurada según el formato JSON (JavaScript Object Notation). La sintaxis de este formato mediante pares claves-valor que están agrupados en objetos. Conceptualmente se corresponden con los campos y valores de las tablas de una base de datos, pero formalmente son más bien datos brutos que requieren algún tipo de procesamiento para poder abordar su análisis, bien mediante programas escritos ad hoc, bien mediante una conversión de los datos a un formato que permita llevar a cabo consultas normalizadas.

Es importante señalar que hemos trabajado solamente con los tweets originales emitidos por cada uno de los perfiles, prescindiendo de los retweets.

Los tweets se han procesado eliminando de los mismos las palabras vacías en español, inglés y francés. También se han eliminado los enlaces existentes, así como las menciones de usuario. Hemos mantenido los hashtags, pues pueden dar una buena idea temática o por donde van las tendencias en un determinado momento.

Para analizar la información recogida hemos trabajado con la información correspondiente a 2015, de manera que podemos representar la evolución de temas a lo largo del año completo.

Se han separado los tweets correspondientes a cada semana del año y sobre ellos se ha aplicado el algoritmo de detección LDA. Así hemos obtenido los temas de cada grupo de perfiles, para cada una de las semanas del año.

Este sistema nos permite ver la evolución temática semana a semana, y así podemos capturar los eventos que suceden en un periodo corto de tiempo, que si se analizaran en periodos más largos podían pasar desapercibidos.

#### 4. Resultados

El número de tweets analizados en el año 2015 para cada uno de los perfiles fue:

Grupo	Nº de tweets
Instituciones	28.188
Bibliotecas	38.059
Empresas	14.252
Webs-blogs	14.328
profesionales	23.444

Tabla 1. Número de Tweets analizados para cada grupo

Para representar la detección de temas, en el apéndice, hemos puesto las 4 semanas del mes de mayo para cada grupo de perfiles. Podemos así observar la evolución en los temas.

Es significativo cómo uno de los temas emergentes en la última semana de mayo es el hashtag #fesabid (Figuras 16-20), con menor repercusión en webs-blogs que aparece con poco peso, coincidiendo con las jornadas en la última semana del mes. Pero viendo la evolución del hashtag, podemos ver que las instituciones lo incluyen en la tercera semana, las bibliotecas solamente en la 4ª semana, y los profesionales empiezan desde la 2ª semana (Figura 10).

En la primera semana de mayo se celebró la 5ª Conferencia internacional sobre revistas de ciencias sociales y humanidades (CRECS) y tiene su reflejo esencialmente en el grupo webs-blogs y profesionales (Figuras 4-5).

En el grupo webs-blogs aparece todo el mes el hashtag #recbib (Figuras 9-19, con más fuerza), centrado en recursos bibliotecarios, que aparece con fuerza en la última semana del mes de marzo y se mantiene hasta la segunda semana de agosto, desaparece y vuelve a aparecer con fuerza en la primera semana de septiembre.

En bibliotecas se hace eco en la primera semana del centenario de la BID y en empresas las #encuestasneodoc, relacionadas con la #gestiondocumental.

La tercera semana de los profesionales (Figura 15) tiene como tema emergente #cibermetria, coincidiendo con la presentación del primer libro de la colección EPI Scholar *Cibermetría. Midiendo el espacio red*. También aparece el tema #jgic, que refleja las jornadas de gestión de la información científica, que tuvieron lugar el mes de

mayo de 2015. Lo mismo sucede con las jornadas bibliosalud, cuyo primer encuentro tuvo lugar el mes de mayo.

Podemos ir viendo el reflejo de los temas y su evolución con el paso de las semanas, para cada grupo de perfiles.

##### 4.1. Semana 1



Figura 1. Instituciones



Figura 2. Bibliotecas



Figura 3. Empresas



Figura 4. Webs-blogs





Figura 5. Profesionales



Figura 9. Webs-blogs

## 4.2. Semana 2



Figura 6. Instituciones



Figura 10. Profesionales

## 4.3. Semana 3

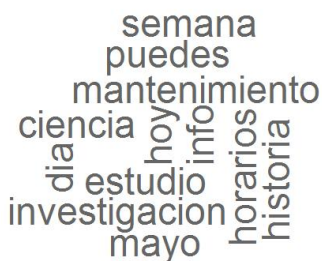


Figura 7. Bibliotecas



Figura 11. Instituciones



Figura 8. Empresas



Figura 12. Bibliotecas



Figura 13. Empresas



Figura 14. Webs-blogs



Figura 15. Profesionales

#### 4.4. Semana 4



Figura 16. Instituciones



Figura 17. Bibliotecas



Figura 18. Empresas



Figura 19. Webs-blogs



Figura 20. Profesionales



## 5. Conclusiones

Planteamos un mecanismo de análisis de los temas emergentes, basado en el algoritmo LDA, que van apareciendo en un conjunto de perfiles y hemos podido ver cómo se refleja la actualidad y el interés en cada uno de dichos grupos.

El sistema funciona correctamente, y nos da unos resultados que creemos muy fiables de las temáticas que se tratan en los grupos de perfiles.

El sistema requiere mejorar los perfiles asignados a los grupos, y probar el empleo de técnicas de stemming, para comprobar si mejora o empeora el sistema.

Este sistema de detección pretendemos implantarlo en tiempo real, de forma que se vayan recogiendo los tweets de forma diaria y al mismo tiempo se vaya construyendo la red de temas emergentes, permitiendo conocer en tiempo real los temas que se están tratando.

## Referencias

- Aiello, L. M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Jaimes, A. (2013). Sensing trending topics in Twitter. // *Multimedia, IEEE Transactions on*. 15:6, 1268-1282.
- Allan, J. (2002). Introduction to topic detection and tracking. *Topic detection and tracking* (pp. 1-16): Springer.
- Anta, A. F., Chiroque, L. N., Morere, P., & Santos, A. (2013). Sentiment analysis and topic detection of Spanish tweets: A comparative study of NLP techniques. // *Procesamiento del lenguaje natural*. 50, 45-52.
- Benhardus, J.; Kalita, J. (2013). Streaming trend detection in twitter. *International Journal of Web Based Communities*. 9:1, 122-139.
- Blei, D. M.; Ng, A. Y.; Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*. 3, 993-1022.
- Cheong, M.; Lee, V. (2009). Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. // *Proceedings of the 2nd ACM workshop on Social web search and mining*.
- Mathioudakis, M.; Koudas, N. (2010). Twittermonitor: trend detection over the twitter stream. // *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*.
- Mehrotra, R.; Sanner, S.; Buntine, W.; Xie, L. (2013). Improving lda topic models for microblogs via tweet pooling and automatic labeling. // *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*.
- Ostrowski, D. A. (2015). Using latent dirichlet allocation for topic modelling in twitter. // *Semantic Computing (ICSC), 2015 IEEE International Conference on*.
- Petrovic, S.; Osborne, M.; Lavrenko, V. (2010). The edinburgh twitter corpus. // *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*.
- Petrović, S.; Osborne, M.; Lavrenko, V. (2010). Streaming first story detection with application to twitter. Paper presented at the *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*.
- Pinholster, G.; Ham, B. (2013). Science communication requires time, trust, and Twitter. // *Science*. 342 (6165), 1464-1464.
- Salton, G.; Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. // *Information processing & management*. 24:5, 513-523.
- Sankaranarayanan, J.; Samet, H.; Teitler, B. E.; Lieberman, M. D.; Sperling, J. (2009). Twitterstand: news in tweets. // *Proceedings of the 17th acm sigspatial international conference on advances in geographic information systems*.
- Sharifi, B.; Hutton, M.-A.; Kalita, J. K. (2010). Experiments in microblog summarization. // *Social Computing (SocialCom), 2010 IEEE Second International Conference on*.
- Sriram, B.; Fuhry, D.; Demir, E.; Ferhatosmanoglu, H.; Demirbas, M. (2010). Short text classification in twitter to improve information filtering. // *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*.
- Wunsch-Vincent, S., & Vickery, G. (2010). *The evolution of news and the Internet: OECD*.

---

Enviado: 2016-04-29. Segunda versión: 2016-09-19.  
Aceptado: 2016-10-10.

---



---

# Análise do discurso e ciência da Informação: aportes teóricos para organização e representação da Informação

*Análisis del discurso y ciencia de la información:  
aportes teóricos para la organización y representación de la información*

*Discourse analysis and Information Science: theoretical contributions to information organization and representation*

---

**João Batista Ernesto MORAES, Larissa Mello LIMA, Mariana Silva CAPRIOLI**

Universidade Estadual Paulista, Av. Hygino Muzzi Filho, 737, Campus Universitário, CEP. 17525900, Marília, SP, Brasil;  
prof.joao@gmail.com, larissademello@gmail.com, mariana.caprioli@gmail.com

## Resumen

Se presenta una perspectiva conceptual y teórica de la escuela francesa del análisis del discurso como una metodología complementaria y válida para ser utilizada en estudios teóricos de organización y representación de la información en la Ciencia de la Información. Se seleccionan autores como Foucault (2010 1986), Orlandi (1999) y Mazière (2007) que ofrecen debates fundamentales sobre el Análisis del Discurso. El segundo paso es discutir el tema de la interdisciplinariedad que se centra en las Ciencias de la Información con el fin de fortalecer las bases para el establecimiento de la relación entre el Análisis del Discurso de la escuela francesa y la Ciencia de la información.

**Palabras clave:** Análisis del discurso. Ciencia de la información. Interdisciplinariedad. Escuela francesa de análisis del discurso. Foucault, Michel. Tratamiento temático de la información.

## 1. Introdução

O objetivo deste trabalho reside na apresentação de alguns aportes teóricos que tornam a Análise do Discurso de Matriz Francesa uma metodologia viável de ser utilizada no âmbito da Ciência da Informação para tanto é retomada a relação proposta por Frohmann (1995); Gonzalez-Gomes (2012) entre discurso e documentação, retomando o conceito de regime de informação.

Desta forma, selecionaram-se, primeiramente, os autores Michel Pêcheux, Michel Foucault e Eni Pulcinelli Orlandi que apresentam um panorama conceitual (Foucault, 1986) e discussões fundamentais (Orlandi, 1999), (Henry, 1997) (Mazière, 2007) acerca da Análise do Discurso de Matriz Francesa.

Em um segundo momento foi selecionado o autor Pinheiro (1995); (1998) para situar a Ciência da Informação expressando o quanto as estruturas desta estão calcadas na interdisciplinaridade. Tal relação interdisciplinar se torna fundamental para

## Abstract

Is intended to bring a brief overview of the conceptual and theoretical discourse analysis of the French Discourse Analysis as a viable methodology to be used in theoretical studies for representation and information organization within the frame of Information Science. For this, selected authors as Foucault (1985; 2010), Orlandi (1999) and Mazière (2007) were reviewed, who offer fundamental discussions on the French Discourse Analysis. The second step was to discuss the issue of the interdisciplinarity nature of Information Science in order to fortify the bases for the constitution of the relationship between the French Discourse Analysis and Information Science.

**Keywords:** Discourse analysis. Information science. French school of discourse analysis. Foucault, Michel. Subject indexing. Interdisciplinarity.

estabelecer-se o paralelo entra a Análise do Discurso de matriz Francesa e a Ciência da Informação. Em seguida, porém não menos importante, retoma-se os estudos de Frohmann (2001) que realizam um paralelo entre a Análise do discurso e a documentação, caminhando para a perspectiva de materialidade da informação.

Através da categorização das contribuições conceituais que cada um deles elencou, se tornará possível traçar um panorama importante da relação interdisciplinar que a Ciência da Informação possui com a Análise do Discurso de Matriz Francesa retomando por fim pesquisa já desenvolvida no grupo de pesquisa Análise Documentária (Lima, 2015).

## 2. Aportes teóricos da Análise do discurso e Ciência da Informação para Organização e Representação da Informação

A Análise do Discurso de Matriz Francesa, segundo Foucault (1986), nasce do estruturalismo especulativo que tem bases em trabalhos com matrizes filosóficas e ideológicas, mas vem para modificar essa estrutura, pois visa a compreender o fenômeno da linguagem não concentrado apenas na língua, mas considerando também os elementos que a extrapolam.

Acredita-se ser importante retomar brevemente o contexto histórico pelo qual a França passava para contextualizar os mecanismos que tornaram possível dar importância para os estudos centrados no sujeito da linguagem e por conseguinte, discurso. Michel Pêcheux, filósofo de formação, começa a realizar uma forte ligação entre a Filosofia e as Ciências Sociais a partir de uma série de estudos centrados na língua. (Henry, 1997)

Tal perspectiva se liga com o movimento que começa a irromper na França de negação ao estruturalismo em meados da década de 1960. Buscava combater o excesso de formalismo e padronização dos elementos da linguagem tratados, por influência dos postulados de Saussure, de maneira estrutural.

Em *Curso de Linguística Geral* Saussure afirma que língua e fala são diferentes, possuindo relação de oposição. Para Saussure a língua é o conjunto de signos estruturados, o todo, que uma comunidade utiliza para se comunicar, ou seja, é algo coletivo e social. Já a fala ele denomina como algo individual, particular, é a maneira como as pessoas usam a língua. Embora diferentes, elas se completam, são dependentes uma da outra. A língua, para que possa existir, se estabelecer, precisa de falantes. Desta maneira, a fala também precisa de um modo para existir, um código, uma ferramenta. Saussure fez seu estudo sobre a língua, pois, segundo ele, era propícia para aprofundamento naquele momento, deixando a fala de lado. (Saussure, 1973)

Ao mesmo tempo em que deixou em segundo plano a fala, Saussure deixa aberto um campo de estudo rico e até então sem aprofundamento teórico.

Ao serem interligados estes fatores listados acima; o excesso de formalismo linguístico somado ao sujeito que conduz a fala sendo deixado em segundo plano, Pêcheux e seus seguidores criam a Análise do Discurso Francesa. O marco inicial da corrente teórica foi fixado no lançamento paralelo em 1969 da obra "Análise Automática do Discurso" de Michel Pêcheux, e da

Revista *Langages* criada por Jean Dubois. (Gregolin, 1996; Maingueneau, 1997).

Torna-se legítimo afirmar que eles possuíam um solo fértil para germinar teorias centradas no sujeito até então esquecido. Ele retoma, por exemplo, elementos da teoria do inconsciente coletivo de Lacan para dar conta de falar sobre o papel desempenhado pelo sujeito nos discursos.

Além disso são colocados em primeiro plano a questão da ideologia, ele afirma que "a ciência é antes de tudo a ideologia com a qual rompe". Para falar de ideologia ele retoma Louis Althusser, de quem foi aluno e a encara enquanto o que "interpela os indivíduos enquanto sujeitos" (Althusser, 1998), sendo reflexo de uma representação entre a realidade e o sujeito. Seu percurso teórico e metodológico na análise do discurso Francesa é sustentada na pirâmide que busca articular o materialismo histórico com a teoria da ideologia, a linguística focando nos processos de enunciação, e a teoria do discurso que coloca em centro a teoria de determinação histórica dos processos semânticos. (Orlandi, 1999)

A Análise do Discurso, então, trabalha na relação entre o linguístico e o ideológico, na medida em que todo e qualquer discurso sofre a ação da ideologia que o atravessa (Foucault, 1986, p. 56).

[...] Gostaria de mostrar que o discurso não é uma estreita superfície, de contato ou de confronto, entre uma realidade e uma língua, ou intrinsecamente entre um léxico e uma experiência, gostaria de mostrar, por meio de exemplos precisos, que analisando os próprios discursos vemos se desfazerem os laços aparentemente tão fortes entre as palavras e as coisas, e destacar-se um conjunto de regras, próprias da prática discursiva. [...] Não mais tratar os discursos como conjunto de signos (elementos significantes que remetem a conteúdos ou a representações), mas com práticas que formam sistematicamente os objetos de que falam. Certamente os discursos são feitos de signos; mas o que fazem é mais que utilizar esses signos para designar coisas. [...] É esse "mais" que é preciso fazer aparecer e que é preciso descrever.

Foucault, na passagem acima, diz que a questão do discurso ultrapassa o paradigma linguístico alertando que todo discurso reflete uma prática própria. Nesta prática entra a perspectiva ideológica; um dos elementos que formam o "mais" dito por Foucault, ou seja, todo e qualquer discurso sofre a ação da ideologia que o atravessa.

A Análise do Discurso apresenta uma definição extremamente estruturada de Orlandi (1999, p. 15) que se trata:

A Análise de Discurso, como seu próprio nome indica, não trata da língua, não trata da gramática, embora todas essas coisas lhe interessem. Ela trata

do discurso. E a palavra discurso, etimologicamente, tem em si a ideia de curso, de percurso, de correr por, de movimento. O discurso é assim palavra em movimento, prática de linguagem: com o estudo do discurso observa-se o homem falando.

De tal forma, a autora nos leva a observar que a Análise do Discurso não trabalha com a língua como um sistema abstrato, mas com ela no mundo, significando para o homem que a fala, levando em consideração a produção desses sentidos em suas vidas, enquanto sujeitos ou enquanto membros de uma sociedade.

Entende-se que a linguagem não é um instrumento que se encontra facilmente visível e transparente, por tal motivo a Análise do Discurso procura detectar como um texto significa indo muito além de padrões e de como cada objeto em seu contexto significa. Então a Análise do discurso busca compreender não somente a prática de linguagem, mas também a parte simbólica e em como isso constrói o ser humano enquanto pessoa. O que leva a refletir que “Conseqüentemente, o discurso é o lugar em que se pode observar essa relação entre língua e ideologia [...]” (Orlandi, 1999, p. 17).

No que tange ao percurso histórico e evolutivo da escola francesa da Análise do Discurso é possível citar Foucault e Pêcheux na década de 1970 como precursores na formação do quadro epistemológico da disciplina. Começou-se a delinear neste período as correntes norteadoras da disciplina com M. Pêcheux (1975) e D. Maingueneau. A denominada Escola Francesa de Análise do Discurso possui três fases fundamentais delimitadas por Pêcheux.

Fregonezi (2002, p. 104) retoma Pêcheux (1975) e explica que a AD da primeira geração, nos anos 60 e início de 70, procurava colocar em evidência os detalhes de formações discursivas (o discurso comunista, socialista, etc.), considerados como espaços relativamente autossuficientes, apreendidos a partir de seu vocabulário. (Fregonezi, 2002, p. 104). A raiz da Análise do Discurso desta corrente Francesa está no entendimento de que é possível intervir na sociedade por meio e através da linguagem, dando destaque para os então discursos políticos. Na obra de Pêcheux (1983) “Análise Automática do Discurso”, a evolução teórica da disciplina é marcada através da categorização de três fases. Na primeira fase AD-1, o corpus de análise era formado por espaços discursivos denominados estáveis e homogêneos: Nesta primeira fase imagina-se que exista um conjunto de enunciados que compõe um discurso idêntico a si mesmo e diferente de outros: o que está contido num discurso está excluído de outro. Por exemplo, o discurso médico é composto de enunciados relativos ao corpo, à

saúde. A expressão “o sujeito é assujeitado” pertence a essa fase. (Fregonezi, 2002, p. 82).

É possível perceber que a questão da ideologia afeta e perpassa os ideais desta fase, na medida em que a questão da enunciação é entendida a partir da lógica de que o discurso está submetido a regras específicas que ultrapassam o indivíduo e sua consciência. Ou seja, a dimensão social e histórica orienta as ideias desta fase. Na segunda fase, a AD-2, Fregonezi (2002) explica que os estudos de Foucault (1985) trouxeram à tona a questão da formação discursiva levando a desconstrução da ideia de homogeneidade das condições de produção. Uma formação discursiva passa a ser caracterizada enquanto dispersão dos enunciados.

A questão do interdiscurso passa a ser elemento que deve ser considerada na constituição de toda análise. Na terceira fase, a AD-3, Pêcheux traz à tona a questão da heterogeneidade discursiva. Neste momento, a polifonia passa a ser discutida na medida em que os enunciados de cada discurso levam a memória de outros discursos (Brandão, 1997, p. 42):

Existe uma negociação entre a heterogeneidade mostrada na linguagem e a heterogeneidade constitutiva da linguagem em que o sujeito, movido pela ilusão de centro, pela ilusão de ser a fonte do discurso, localiza o outro e delimita o seu lugar.

É importante entender a multiplicidade de abordagens que o objeto discurso pode assumir. A vertente centrada nos postulados de Van Dijk (1972), denominada Análise Crítica do Discurso, busca analisar enunciações focando-se nos problemas sociais e mecanismos políticos da atuação discursiva. Pode-se dizer que a escola Francesa da Análise do Discurso e a análise crítica do discurso compartilham o olhar sob a dimensão textual discursiva enquanto uma ruptura com a materialidade linguística.

A abordagem deste trabalho parte da teoria discursiva da Escola Francesa de Análise do Discurso. Pêcheux (1987) refere-se a esta como efeito de sentido entre interlocutores. Retomar Pêcheux e a perspectiva e memória discursiva é, de certa forma, se aliar à perspectiva de Ducrot (1987, p. 94) que em uma de suas hipóteses considera que, para compreender um enunciado, o locutor cumpre “as leis que regulamentam a tomada da palavra na coletividade linguística a que pertence”.

Bronckart (2008, p. 90) segue tal linha de pensamento ao entender o locutor como “textualizador”. “Instância a qual o autor empírico de um texto confia a responsabilidade sobre aquilo que vai ser enunciado.” A partir deste processo o autor explica que são caracterizadas as vozes que

são expressas no texto, ligando-se assim com a perspectiva de Ducrot ao considerar que os sentidos estão sempre carregados de outros significados.

Perspectiva esta que considera as vozes que são expressas no texto é caracterizada por Maingueneau (2008) como heterogeneidade enunciativa. O autor retoma a oposição conceitual desenvolvida por J. Authier (2004) "Hétérogenéité montrée et hétérogenéité constitutive". Maingueneau (2008, p. 31) explica o primeiro exemplo do seguinte modo:

[...] É acessível aos aparelhos linguísticos, na medida em que permite apreender sequências delimitadas que mostram claramente sua alteridade [...]

Esta citação direta é um exemplo de heterogeneidade mostrada pois se trata de uma sequência linguística em que é marcada de forma explícita a presença de outro locutor no texto. Já a heterogeneidade constitutiva acontece quando há uma sequência linguística marcada de forma implícita na podem ser apreendidas por uma abordagem linguística stricto sensu.

Na heterogeneidade constitutiva se encontra o interdiscurso, no sentido em que amarra o Mesmo do discurso e seu Outro, sendo uma relação intrincada (Maingueneau, 2008). Assim, os discursos são colocados em relação com outros, posteriormente, não existindo juntos previamente, ou seja, nascem nessa brecha da rede interdiscursiva.

O interdiscurso possuem uma relação de consonância com os processos de enunciação que segundo Pêcheux e Fuchs (1975, p. 176) "consistem em uma série de determinações sucessivas pelas quais o enunciado se constitui pouco a pouco e que tem como característica colocar o 'dito' e em consequência rejeitar o 'não-dito'."

Pêcheux e Fuchs (1975, p. 176) explicam ainda que o processo da enunciação é formado por fronteiras que situam o que é dito em um campo criado pelo enunciador buscando validar o seu dizer, ou seja, buscando ter domínio do sentido. Este fenômeno é conhecido como esquecimento numero 2:

A enunciação equivale pois a colocar fronteiras entre o que foi "selecionado" e tornado preciso aos poucos através do que se constitui o "universo do discurso", e o que é rejeitado. Desse modo se acha, pois, desenhado num espaço vazio o campo de "tudo o que teria sido possível ao sujeito dizer (mas que não diz)" ou o campo de "tudo a que se opõe o que o sujeito disse". Esta zona do "rejeitado" pode estar mais ou menos próxima da consciência e há questões do interlocutor - visando a fazer, por exemplo, com que o sujeito indique com precisão "o que

ele queria dizer" — que o fazem reformular as fronteiras e re-investigar esta zona. Propomos chamar este efeito de ocultação parcial esquecimento n 2 e de identificar a a fonte da impressão de realidade do pensamento para o sujeito ("eu sei o que eu digo", "eu sei do que eu falo").

Movimento inverso é denominado como esquecimento numero 1 (Pêcheux & Fuchs, 1975, p. 176):

Por oposição, o esquecimento n 1, cuja zona é inacessível ao sujeito, precisamente por esta razão, aparece como constitutivo da subjetividade na língua. Desta maneira, pode-se adiantar que este recalque (tendo ao mesmo tempo como objeto o próprio processo discursivo e o interdiscurso, ao qual ele se articula por relações de contradição, de submissão ou de usurpação) é de natureza inconsciente, no sentido em que a ideologia é constitutivamente inconsciente dela mesma.

Entende-se assim que o esquecimento numero 1 é aquele de natureza inconsciente e sendo assim propriamente ideológica, tendo a ideia ilusória de ser centro de origem do sentido. Situar o processo ideológico na análise é falar sobre formação ideológica que segundo os referidos autores são os elementos externos à formação discursiva que repousam nas relações de poder.

## 2.1. Ciência da Informação e interdisciplinaridade

Um ponto importante a ser tratado é a conceitualização da Ciência da Informação, uma vez que se trata da ciência que usa a informação como objeto de estudo, ou seja (Pinheiro e Loureiro, 1995, p. 2):

[...] ciência da informação é aquela disciplina que investiga as propriedades e o comportamento da informação, as forças que governam o fluxo da informação e os meios de processamento para acesso e uso otimizados. Ela diz respeito àquele corpo de conhecimento ligado à origem, coleta, organização, armazenagem, recuperação, interpretação, transmissão, transformação e utilização da informação... possui um comportamento de ciência pura, que investiga o interior do assunto sem considerar suas aplicações, é um componente de ciência aplicada, que desenvolve serviços e produtos.

Uma das primeiras definições de C.I., exatamente essa: "disciplina que investiga as propriedades e comportamento da informação, as forças que regem o fluxo de informação, a fim de alcançar acessibilidade e utilização ótimas", definição dada por Taylor em 1966 e posteriormente sintetizada por Borko, em 1968.

A C.I. é amplamente conhecida como interdisciplinar por nascença, sendo impossível, desde o seu surgimento, estudá-la como algo isolado, sabendo que a todo o momento existe diálogo entre

ela e alguma outra área do conhecimento, e portanto, acaba se tornando uma área mais rica.

Segundo Pinheiro (1998, p. 133):

Durante vinte anos de estudos de Ciência da Informação, nossa percepção é de que a Ciência da Informação tem seu próprio estatuto científico como ciência social que é, portanto, interdisciplinar por natureza, e apresente interfaces com a Biblioteconomia, Ciência da Computação, Ciência Cognitiva, Sociologia da Ciência e Comunicação, entre outras áreas, e suas raízes, em princípio, vem da bifurcação da Documentação/Bibliografia e da Recuperação da Informação. E seu objeto de estudo, por si mesmo, na complexidade de categoria abstrata, é de difícil apreensão.

A existência da interdisciplinaridade se dá pela relação de constante troca que executa com outras disciplinas, essas trocas precisam ser teóricas e metodológicas sendo importante e indispensável a clareza para identificar onde ocorre o encontro das duas áreas do conhecimento que se relacionam. Segundo Le Coadic (1996), uma série de temas periféricos que se apoiam firmemente na C.I. se destacam, como temas psicológicos, relacionados a comportamento de comunicação, por exemplo, ou linguísticos como a semiótica, e sociológicos como a produtividade científica, entre outros.

Ainda existem definições que sugerem relações interdisciplinares implícitas, pois a C.I. é uma

[...] disciplina que surge de uma fertilização cruzada de ideias que incluem a velha arte da Biblioteconomia, a nova arte da Computação, as artes dos novos meios de comunicação, e aquelas ciências como a Psicologia e Linguística, que em suas formas modernas têm a ver diretamente com todos os problemas da comunicação – a transferência do pensamento organizado. (PINHEIRO, 1998, p. 135).

A partir das discussões que tangenciam a questão da interdisciplinaridade na Ciência da Informação torna-se importante considerar a perspectiva de transdisciplinaridade na Análise Documental de conteúdo (Garcia Marco, 1998; Izquierdo Arroyo, 1993; Gonzalez de Gomes, 1990) na medida em que adentrar na questão da transdisciplinaridade é delimitar ainda mais a preocupação com a formação de um quadro teórico para abarcar a multiplicidade de abordagens teóricas e metodológicas da área. Caracterizam como “momentos históricos sucessivos no programa de sistematização científica” (Sagredo Fernández e Izquierdo Arroyo, 1983, p. 24) os níveis de interação disciplinar em que estão a interdisciplinaridade e transdisciplinaridade.

A questão da transdisciplinaridade se sustenta na discussão dos sistemas complexos (complexidade) que seriam aqueles que possuem uma

série de componentes interagindo de maneiras diversas.

Já a integração de conhecimentos interdisciplinar baseia-se no diálogo entre conceitos e disciplinas na medida em que os mesmos são apropriados e transformados no âmbito de determinado contexto disciplinar.

O contexto disciplinar base deste artigo parte da Ciência da Informação diante da interdisciplinaridade, sendo apresentada a Análise do Discurso como uma metodologia viável, e para tanto, viável contextualização no âmbito da Ciência da Informação. Exemplifica-se o quanto suas estruturas são impregnadas de interdisciplinaridade para se compreender, posteriormente, a relação com a Análise do Discurso.

## 2.2. Foucault e as formações discursivas

Foucault (1986) diz que a questão do discurso transcende o paradigma linguístico alertando que todo discurso reflete uma prática própria. Nesta prática entra a perspectiva ideológica; um dos elementos que formam o “mais” dito por Foucault, ou seja, todo e qualquer discurso sofre a ação da ideologia que o atravessa.

Ligar o linguístico e o ideológico, no âmbito da prática discursiva, é falar de formação discursiva na medida em que esta determina o que pode ou não ser dito em certo campo do conhecimento levando em conta a posição do enunciador.

Em relação à perspectiva de repetição Foucault (2010, p. 23) alerta:

Não há de um lado, a categoria dada de uma vez por todas dos discursos fundamentais ou criadores; e, de outro, a massa daqueles que repetem, glosam e comentam. Muitos textos maiores se confundem e desaparecem, e por vezes comentários vêm tomar o primeiro lugar. Mas embora seus pontos de aplicação possam mudar, a função permanece; e o princípio de um deslocamento encontra-se sem cessar reposto em jogo. (Foucault, 2010, p. 23).

Interligar a perspectiva de formação discursiva com gênero discursivo se torna importante sob tal perspectiva na medida em que o gênero do discursivo científico possui características singulares.

Guimarães (2001) nos explica acima é que o discurso científico é mais fechado, finalizado em si mesmo na medida em que o enunciador leva em conta um viés para se comunicar com o enunciatário que nesse caso são os pares da comunidade científica. No que tange à complexidade das relações enunciativas nessa esfera, Barros (2014, p. 167) expõe que:

[...] ponderação fundamental a respeito do discurso científico é que ele é acima de tudo, polifônico e representado. Quer dizer as relações enunciativas serão profundamente complexas e irão se estabelecer em camadas.

A questão da polifonia se explica no fato do discurso científico ser formado por várias vozes que se relacionam no texto além da do autor.

Estas várias vozes que se relacionam e atestam o discurso científico são explicadas por Bakhtin (2003, p. 281) através do fato de que todos os gêneros do discurso são encarados a partir da marca da heterogeneidade. Ao que parece o discurso científico apresenta uma marca mais expressiva desta heterogeneidade na medida em que:

Os gêneros secundários do discurso – o romance, o teatro, o discurso científico, o discurso ideológico – aparecem em circunstâncias de uma comunicação cultural mais complexa e relativamente mais evoluída, principalmente escrita: artística científica, sociopolítica. Durante o processo de sua formação, esses gêneros secundários absorvem e transmitem os gêneros primários (simples) de todas as espécies, que se constituíram em circunstâncias de uma comunicação verbal espontânea.

Ao descrever o processo de formação do discurso científico, é ressaltado o quanto este enquanto gênero secundário é complexo por ser formado e atestado pelos gêneros primários. Acredita-se que aconteça um fenômeno de engessamento na medida em que a camada da comunicação científica segue regras e padrões argumentativos.

Desta forma, devido a essas características, pode-se compreender que o discurso científico, no qual a Ciência da Informação repousa, existe uma constante redefinição e repetição de posturas anteriores e de deslocamentos provocados pela ação do poder (instituições de pesquisa, pesquisadores, agências de fomento, universidades).

É parte do próprio movimento de validação científica a questão da repetição na medida em que um autor X de menor “impacto” cita autor Y de maior impacto e reconhecido pela comunidade científica (poder) o autor X busca filiar-se ao ponto de “superar” o autor Y, desta forma é possível dizer que a prática científica é refém de suas formulações enunciativas.

Em pesquisa desenvolvida, Lima (2015) caracterizou a presença de duas formações discursivas incidindo sobre a área no Brasil:

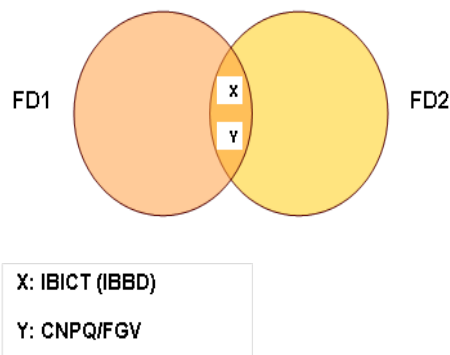


Figura 1. Formações discursivas da Ciência da Informação

Como formação discursiva 1 (FD1) há a caracterização da Ciência da Informação a partir de sua interdisciplinaridade. Como formação discursiva 2 (FD2) há o entendimento de que a Ciência da Informação é uma disciplina recente que carece de balizas teóricas.

No ponto de intersecção entre ambas as formações discursivas (X;Y) temos o IBICT e o CNPq/FGV. O Instituto Brasileiro de Bibliografia e Documentação (IBBD atual Instituto Brasileiro de Informação Científica e Tecnológica; IBICT) foi criado em 1956, como uma iniciativa da Unesco IBBB que sugeriu à fundação Getúlio Vargas (FGV) que promovesse no Brasil a criação de um centro nacional de bibliografia tendo como foco o intercâmbio de documentações técnico científica valorizando a perspectiva tecnológica.

Acredita-se que tais órgãos se configuram como instâncias de validação científica da Ciência da Informação no Brasil, na medida em que segundo Menezes, Silva e Couzinet (2007) o IBBB foi o órgão responsável pela criação de periódicos de importância norteadora para a área em âmbito Nacional; Ciência da Informação e Revista da Escola de Biblioteconomia da UFMG. Este fato não pode ser desvinculado do fato do IBBB ter implementado no país dois anos antes, em 1970, o primeiro curso de mestrado em Ciência da Informação.

O IBBB possui relação direta com o CNPq (Centro Nacional de Desenvolvimento Científico e Tecnológico) e a FGV (Fundação Getúlio Vargas) na medida em que estas se configuram como entidades governamentais específicas para fomentar o desenvolvimento científico no Brasil. Ou seja, são fatores que dentro de uma conjuntura histórico discursiva formam a perspectiva de institucionalização da área.



Interligar a perspectiva institucional com a discursiva, por meio da prática caracterização das formações discursivas (FD1); (FD2), é se ater ao conceito de “Regime de Informação”. Frohmann (1995) sinaliza para o fato de como a informação é materializada através do meio institucional

A partir de Frohmann (1995) González de Gómez (2012) pontua que a criação do então IBBD, atual IBICT, indicou o surgimento de um novo regime de informação no Brasil, uma vez que sua criação foi oriunda de uma parceria da Unesco com a Fundação Getúlio Vargas, por intermédio de seu representante no Brasil, Paulo Carneiro, o qual indicou Lygia de Queiroz Sambaquy, juntamente com Jannice Monte-Mór, para percorrerem as principais bibliotecas e centros de documentação da Europa e Estados Unidos durante quase um ano, viagem essa, que tinha a finalidade de fomentar o estabelecimento de centros bibliográficos nacionais, se beneficiando da política da Unesco.

Tal ação contou com o apoio da Fundação Getúlio Vargas e serviu para ampliar as ideias de Lygia de Queiroz Sambaquy sobre a estruturação que um órgão que fosse dedicado a trabalhos bibliográficos deveria possuir para impulsionar atividades científicas e tecnológicas do país. Tal criação confirma a teoria de regime de informação, que se trata de:

Um regime de informação seria o modo informacional dominante em uma formação social, o qual define quem são os sujeitos, as organizações, as regras e as autoridades informacionais e quais os meios e os recursos preferenciais de informação, os padrões de excelência e os modelos de sua organização, interação e distribuição, enquanto vigentes em certo tempo, lugar e circunstância (Gonzales de Gomez, 2012).

É possível pontuar, a partir das considerações acima, que o conceito de Regime de Informação de Frohmann (1995) relaciona a perspectiva institucional com a discursiva via a relação entre formações discursivas. Ela torna possível a compreensão das instâncias de validação científica da Ciência da Informação no Brasil situadas enquanto atores que relacionam dinâmicas de poder da área no país.

### 2.3. Mazière e Orlandi: Dispositivo de observação da Análise do Discurso

Utilizar a Análise do Discurso enquanto metodologia de pesquisa é trabalhar na construção de um dispositivo de observação (Mazière, 2007).

O referido autor explica que o discurso para o analista do discurso se configura como produto,

na medida em que há a materialidade do enunciado ou de um grupo de enunciados que formam a “corpora” que será submetida à análise.

Mazière (2007, p. 14), além disso, destaca o papel do analista do discurso dizendo que “ele delimita, põe em correspondência, organiza fragmentos de enunciados, mais ou menos longos e mais ou menos homogêneos para submetê-los à análise.”

Para formar este “dispositivo de observação” é necessário se atentar para elementos que interferem na produção de todo e qualquer discurso. Orlandi (1999) explica que a Análise do Discurso possui um estatuto histórico próprio, possuindo contribuições paralelas de J. Dubois e M. Pêcheux, dois nomes configurados como polos na vertente francesa da Análise do Discurso.

Do lado de J. Dubois, a instituição da Análise do Discurso é pensada dentro de um continuum: a passagem do estudo das palavras (lexicologia) ao estudo do enunciado é “natural”, é uma extensão, um progresso permitido pelo linguista. A AD, em suma, não terá sido senão um lance de seu caminho científico. Do lado de M. Pêcheux, ao contrário, a Análise do Discurso é pensada como ruptura epistemológica com a ideologia que domina nas ciências humanas. (especialmente a psicologia) (Orlandi, 1999, p. 18).

O “continuum” de Dubois que faz a Análise do Discurso um reflexo do percurso da linguística, pode ser encarada como uma tentativa de “dominação” por parte deste teórico, no sentido de que se busca enquadrá-la em uma naturalidade extensiva.

Já Pêcheux representa a “resistência” a esta perspectiva na medida em que enxerga a Análise do Discurso sob a ótica do rompimento com a ideologia das Ciências Humanas.

Orlandi (1999, p. 15) diz que o interesse de estudiosos na linguagem de maneira particular com base nas diversas formas de significar, preconizou a origem da Análise do Discurso.

Assim, a primeira coisa a se observar é que a análise do discurso não trabalha com a língua enquanto um sistema abstrato, mas com a língua no mundo, com as maneiras de significar, com homens falando, considerando a produção de sentidos enquanto parte de suas vidas, seja enquanto sujeitos, seja enquanto membros de uma determinada forma de sociedade.

Ou seja, sai do foco a esfera da língua como elemento neutro para a visão de que a linguagem enquanto discurso não é neutra.

Orlandi (1999, p. 19) define o quadro epistemológico da Análise do Discurso relacionando três regiões do conhecimento científico:

- o materialismo histórico, como teoria das formações sociais e suas transformações.
- a linguística, como teoria dos mecanismos sintáticos e dos processos de enunciação.
- a teoria do discurso, como teoria da determinação histórica dos processos semânticos.

Ou seja, a linguagem é dotada de uma materialidade. Dentro desta materialidade está inserido o falante, sujeito parte das formações e transformações sociais, os atos deste falante possuem uma semântica e, por conseguinte fazem parte de um processo histórico.

Mapear o discurso dos sujeitos que atuam em uma ciência é levar em consideração estes elementos que se movimentam em relação à posição de cada enunciado.

#### 2.4. Paralelo entre Análise do Discurso e Ciência da Informação

Ao falar-se da Ciência da Informação e conceitos importantes da Análise do Discurso de matriz francesa, encontra-se a necessidade de relacionar as duas, uma vez que esse trabalho visa a situar a Análise do Discurso na Ciência da Informação, e dessa forma, a pergunta que surge é: O que ambas tem em comum ou como conversem? Para Cordeiro (2004, p. 85):

Análise de Discurso e Ciência da Informação evoluem a partir dos anos 1960 seguindo caminhos semelhantes se consideradas, ambas, como expressão de uma face da ciência pós-moderna. A partir do pós- Guerra, apresentam crises que são comparáveis e em sincronia, apresentam trânsitos por várias espaços epistemológicos: do paradigma físico, ao paradigma cognitivo, daí ao social. Assim se descreve os passos da CI. Se bem que há quem diga que os modelos dos paradigmas já não se aplicam mais. Nem mesmo a epistemologia teria aplicação tão efetiva se comparada à arqueologia do saber.

No que tange a perspectiva de pós modernidade, o paradigma físico citado por Cordeiro (2004) pode ser exemplificado.

Então, tem-se que a Ciência da Informação e Análise do Discurso seguiram caminhos paralelos ou parecidos, considerando de onde são oriundas. Belkin e Robertson (1976) citados por Freire (1995) mostram que a primeira é constituída por um conjunto ou coleção de signos que foram estruturados de tal forma propositalmente ou propositalmente por um emissor que tem a intenção de mudar a imagem para o receptor, ou

seja, se pensa em como o emissor pode estruturar seus pensamentos para que o receptor seja capaz de perceber a mensagem que o emissor desejava passar. O que se torna semelhante com o que a Análise do Discurso se propõe a realizar enquanto metodologia de pesquisa, na medida em que ela trata a linguagem levando em conta a ideologia que incide durante todo o processo histórico que é também propriamente discursivo. Ou seja, se preocupando como a língua como algo concreto, em como ela significa no mundo levando em consideração a produção de sentido enquanto parte na vida dos homens na qualidade de sujeitos ou membros de uma sociedade, mostrando, dessa forma, que a linguagem não é neutra e leva em consideração emissor, receptor e contexto.

Viu-se a necessidade de diferenciar os conceitos de Análise de Conteúdo e de Análise Documental de Conteúdo, além de Análise do Discurso, a fim de verificar a diferença presente entre elas, embora sejam de fato tão próximas. Dessa forma, segundo Bardin (1997) a Análise de Conteúdo se trata de um método de organização e análise dos dados, e para tanto, deve possuir algumas características como se aceitar que o seu foco seja na qualificação de vivências de um sujeito, assim como suas preocupações sobre determinado objeto e seus fenômenos. Ainda segundo Bardin, é constituída por várias técnicas na qual se busca descrever o conteúdo emitido no processo de comunicação, sendo por meio de falas ou de textos, técnicas essas, compostas por procedimentos sistemáticos que levam ao levantamento de indicadores, podendo ou não ser quantitativos, e permitindo a realização de inferências de conhecimentos.

A principal diferença entre a Análise de Conteúdo e a Análise do Discurso, é que não compartilham do mesmo horizonte teórico, nem pressuposto metodológico de que a produção linguística seria reveladora de opiniões e de crenças de um sujeito em relação ao mundo que o rodeia, uma vez que na Análise do Discurso não se pode negligenciar as relações de um texto e seu entorno e, segundo Charaudeau e Maingueneau (2004), deve-se visar predominantemente o modo como a enunciação é capaz de inter-relacionar uma organização textual e um lugar social determinado.

Da mesma forma, também se diferencia de Análise Documental de Conteúdo, que "(...) ao lado da catalogação de assunto e da indexação, uma das correntes teóricas que historicamente permeia a Organização do Conhecimento no âmbito do Tratamento Temático da Informação" (guimarães, 2008, 2009). Esse tipo de Análise Documental foi ao longo de duas décadas, e con-

tinua sendo, amplamente empregada na Espanha, possuindo clareza de situação de referência teórica.

Vale ressaltar que, segundo Pinto Molina (1992) e Ruiz Pérez (1992), diferentemente do Brasil ou da França que se utilizam da Análise Documental para o tratamento temático da informação (dimensão de conteúdo), na Espanha a expressão se trata de uma definição mais abrangente enquanto Tratamento da Informação, se dividindo entre: Análise documental de forma (ADF) e Análise documental de conteúdo (ADC).

Realiza-se nesse momento um breve trajeto do cenário Francófono que abarca de maneira efetiva a ligação entre a Análise Documental de conteúdo e a Análise do Discurso a partir do viés da corrente francesa da pragmática, lingüística textual que culminaram na aplicação das teorias do resumo documental dentro da Análise Documental de Conteúdo. (Izquierdo Alonso e Moreno Fernández, 2009).

A questão da Análise Documental, segundo Guimarães e Sales (2010), encontra bases na tradição Francófona, onde se constata que a construção do conceito de Documentação parte-se das idéias de Paul Otlet (1934), e é nesse contexto que se destaca o papel dos autores como José López Yepes, Izquierdo Arroyo e Felix Sagredo, que introduziram a concepção francesa de documentação no universo biblioteconômico espanhol.

Dessa forma, Sagredo Fernández e Izquierdo Arroyo (1982) realizam um histórico sobre a noção de documento focando no avanço tanto semântico quanto etimológico próprio do termo. Termo para os referidos autores é conceito chave para definir documento na medida em que consideram o documento como palavra (termo) e também como objetivo (entidade).

No contexto da Análise do discurso o documento se enquadra na perspectiva construtiva funcional pois esta destaca a função do documento no que diz respeito aos fundamentos do modelo de comunicação humana que são: Emissão, Receptor, mensagem, código, canal, ruído e contexto em que se insere todo esse processo, ou seja, reflexo de qual formação discursiva e ideológica.

Segundo Izquierdo Alonso (2000) a pragmática documental coloca em relação um elemento lingüístico com seus autores e receptores. Estes em uma situação de comunicação documental concreta e específica. Nesta dimensão é destacada a importância de adequar a representação do conteúdo ao contexto, retomar o contexto é

propriamente discursivo entendendo a língua enquanto elemento concreto, que dá significado e produz sentido.

Segundo Serres (s.d.) e Chaumier (1982), citados por Guimarães et al (2012, p. 47), nessa área

[...] estudos de forte matriz lógico-lingüística foram desenvolvidos, no intuito de melhor compreender como se processa a operação intelectual de representação do conteúdo de um documento de tal modo a propiciar uma consulta e uma recuperação mais efetivas, por meio de produtos oriundos da condensação (tais como resumos) e da indexação (tais como conceitos-chave de conteúdo).

No Brasil, partido de bases Francesas, bem como o cenário Francófono Espanhol apresentado, a Ciência da Informação se articula na tríade *produção, organização e uso* da informação, segundo Guimarães (2008), e a organização da informação se importa com a representação da informação, justifica-se a Análise do Discurso como parte da Análise Documental e “[...] apresenta um conjunto de procedimentos metodológicos voltados à definição do conteúdo temático de documentos de modo a permitir a recuperação, o acesso e o uso da informação neles contida” (Moraes, 2011).

Sendo assim é importante pontuar, primeiramente, que a principal diferença entre análise do discurso e análise de conteúdo está no fato de que a segunda se importa com o que está sendo dito enquanto que para a primeira, interessa o como está sendo dito, ou seja, quais são as filiações teóricas e institucionais que possibilitam aquele dizer. Já a análise documental de conteúdo e a análise documental de forma se caracterizam, segundo (Guimarães, 2008) a partir da diferenciação que coloca a análise documental de forma como a que se preocupa com “os elementos de natureza material do documento” (Guimarães, 2008, p. 198), e a análise documental de conteúdo se preocupa com “os elementos de natureza intelectual do documento.” (Guimarães, 2008, p. 198). o referido autor nos explica que estas dimensões de análise são realizadas de maneira consecutiva caracterizando assim a natureza cognitiva que incide sobre os processos.

### 3. Considerações finais

Retomadas as discussões propostas por Foucault (1986; 2010), Orlandi (1999), Pinheiro (1998) e Cordeiro (2004) é possível dizer que a Análise do discurso de matriz francesa apresenta uma metodologia viável de utilização nos estudos da área, sendo então a Ciência da Informação uma ciência interdisciplinar, a qual se relaciona com diversas outras ciências. recorda-se que a

análise do discurso e a ciência da informação se preocupam com a forma que a informação será recebida a partir de um emissor e em determinado contexto. além disso, a partir da interação entre a perspectiva institucional com a discursiva foi elucidada a importância do regime de informação (Frohmann, 1995; González-Gomez, 2012) no tecido do impacto da análise do discurso na documentação.

Ou seja, pode-se dizer que a utilização justificase através do fato de que a Análise do Discurso articula ferramentas que visam não apenas situar o dito mas relacioná-lo às dinâmicas de poder, perspectiva que tem o potencial de trazer um novo viés para os estudos sobre representação e organização da informação na área.

Ilustrou-se de maneira pontual tal viés através da caracterização das instâncias de validação científica da área no Brasil permitindo a visualização da conjuntura histórico-discursiva que formam a institucionalização da Ciência da Informação no Brasil.

## Referências

- Althusser, L. P. (1998). *Aparelhos Ideológicos de Estado*. 7ª ed. Rio de Janeiro: Graal, 1998.
- Bardin, L. (1977). *Análise de conteúdo*. Lisboa: Edições 70 Ltda, 1977.
- Bakhtin, M. (2003). *Estética da criação verbal*. São Paulo: Martins Fontes, 2003.
- Brandão, N. H. H. (2004). *Introdução a análise do discurso*. 2.ed. Campinas: ed. da unicamp, 2004.
- Barros, T. H. B. (2014). *A representação da informação Arquivística: Uma Análise do discurso teórico e institucional a partir dos contextos Espanhol, Canadense e Brasileiro*. Marília: Unesp. Tesis Doctoral.
- Bronckart, J. P. (2008). *O agir nos discursos: das concepções teóricas às concepções dos trabalhadores*. Campinas: Mercado de Letras, 2008.
- Cavalcante, R. B.; Calixto, P.; Pinheiro, M. M. K. (2014). *Análise de conteúdo: considerações gerais, relações com a pergunta de pesquisa, possibilidades e limitações do método*. // *Inf. & Soc.* 24:1, (abr 2014).
- Charaudeau, P.; Maingueneau, D. (2004). *Dicionário de Análise do Discurso*. São Paulo: Contexto, 2004.
- Cordeiro, P. A. (2004). *Análise do discurso e ciência da informação: Ensaio sobre uma possibilidade metodológica*. Rio de Janeiro: UFRJ. Dissertação. [http://www.cprm.gov.br/publique/media/mestrado\\_pedro.pdf](http://www.cprm.gov.br/publique/media/mestrado_pedro.pdf) (2013-12-03).
- Ducrot, O. (1987). *O dizer e o dito*. Campinas: Pontes, 1987.
- Freire, I. M. (1995). *Informação: consciência possível; campo. Um exercício com construtos teóricos*. // *Ci. Inf. Brasília*. 24:1 (fev 1995).
- Fregonezi, J. A. (2002). "Que casô era pro resto da vida, né?": Um olhar feminino: o processo discurso/memória. Londrina: Edições Humanidades, 2002.
- Frohmann, B. (1995). *Taking information policy beyond information science: applying the actor network theory*. // 23rd Annual Conference: Canadian Association for Information, 1995.
- Foucault, M. (1986). *A arqueologia do saber*. Rio de Janeiro: Forense, 1986.
- Foucault, M. (2010). *A ordem do discurso*. São paulo: Edições Loyola, 2010.
- García Marco, F. J. (1998). *El concepto de información: una aproximación transdisciplinar*. // *Revista geneneral de información y documentación*. 8:1.
- Gonzalez de Gomez, M. N. (1990). *O objeto de estudo da Ciência da Informação : paradoxos e desafios*. // *Ciência da Informação*. 19:2 (dez.1990)
- Gonzalez de Gomez, M. N. (2012). *Regime de informação: construção de um conceito*. // *Informação & Sociedade*. 22: 3 (dez.2012).
- Guimarães, J. A. C.; Oliveira, E. T.; Gracio, M. C; Fernández-Molina, C. (2012). *Referentes teóricos em análise documental de conteúdo no ambiente acadêmico espanhol de Biblioteconomia e Documentação - Universidade da Co-ruña (Espania)*, 2012.
- Guimarães, E. (2001). *Expressão modalizadora no discurso de divulgação científica*. // *Educação e Linguagem*. 4: 5.
- Guimarães, J. A. C. (2008). *A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO)*. // *Revista Ibero-americana de Ciência da Informação*. 1:1.
- Guimarães, J. A. C. (2009). *Abordagens Teóricas em tratamento temático da informação: catalogação de assunto, indexação e análise documental*. // Garcia Marco, Francisco Javier (org.). *Avances y perspectivas en sistemas de información y de documentación*. Zaragoza: Prensas Universitarias de Zaragoza, 2009.
- Guimarães, J. A. C.; Sales, R. (2010). *Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação*. // *DatagramaZero*. 11:1 (fev. 2010).
- Izquierdo Alonso, M. (2000). *Nuevos enfoques en el estudio del tratamiento documental de contenido desde los presupuestos de las ciencias del lenguaje*. // *Scire: representación y organización del conocimiento*. 6:1 (enero-jun. 2000).
- Izquierdo Alonso, M.; Moreno Fernández, L. M. (2009). *El resumen documental: un reto didáctico*. Madrid: Anabad, 2009.
- Lima, L. M. (2015). *O percurso discursivo da Ciência da Informação no Brasil: Uma análise discursiva à partir dos periódicos Ciência da Informação e Revista da Escola de Biblioteconomia da UFMG*. 2015. Marília: Unesp. Trabalho de conclusão de curso. 2015.
- Le Coadic, Yves-François. (1996). *A ciência da informação*. Brasília: Briquet de Lemos/Livros, 1996.
- Mazière, F. (2007). *A análise do discurso: história e práticas*. São paulo: Parábola Editora, 2007.
- Maingueneau, D. (1997). *Novas tendências em análise do discurso*. Campinas: Pontes, 1997.
- Maingueneau, D. (2008). *Gênese dos discursos*. São Paulo: Parábolas Editorial, 2008.
- Menezes, E. M ; Silva, E. L.; Couzinet, V. (2007). *A Ciência da Informação na França e no Brasil*. // *DataGramaZero* 8: 6 (dez 2007).
- Moraes, J.B.E.de (2011). *A questão do aboutness no texto narrativo de ficção*. Marília: Unesp. Tese Livre Docência Linguística e Documentação. 2011.
- Otlet, P. (1943). *Traité de documentation: le livre sur le livre: théorie et pratique*. Bruxelles: Mundaneum, 1934. [http://lib.ugent.be/fulltxt/handle/1854/5612/Traite\\_de\\_documentation\\_ocr.pdf](http://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf) (2016-09-19).
- Orlandi, E. P. (1999). *Análise de discurso: princípios e procedimentos*. Campinas, SP: Pontes, 1999.
- Orlandi, E. P. (2000). *Discurso e leitura*. 5. ed. São Paulo: Editora da Unicamp, 2000.

- Pecheux, M. (1975). Semântica e discurso: uma crítica à afirmação do óbvio. Tradução de Eni P. Orlandi [et al.]. Campinas: Editora da UNICAMP, 1975.
- Pêcheux, M; Fuchs, C.(1997). A propósito da Análise Automática do Discurso: atualizações e perspectivas, 1975. // Por uma análise automática do discurso: uma introdução a obra de Michel Pêcheux / Gadet, Francaise; Hak, Tony (orgs.); tradutores Eni Pulcinelli Orlandi... [et al.]. 3. ed. Campinas, SP: Editora da UNICAMP, 1997.
- Pinheiro, L. V. R. (1998). Campo interdisciplinar da Ciência da Informação: fronteiras remotas e recentes. // Investigación Bibliotecológica. México. 12:25 (jun 1998) file:///C:/Users/Unesp/Downloads/ibi02508.pdf. (2013-08-07).
- Pinheiro, L. V. R; Loureiro, J. M. M. (1995). Traçados e limites da Ciência da Informação// Ciência da Informação. Brasil. 24:1 (fev 1995) <http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/531/483>. (2013-10-18).
- Pinto Molina, M. (1992). El resumen documental: principios y métodos. Madrid; Salamanca: Fundación Germán Sánchez Ruipérez; Madrid: Pirámide, 1992.
- Rocha, D.; Deusdará, B . (2005). Análise de Conteúdo e Análise do Discurso: aproximações e afastamentos na (re)construção de uma trajetória. // Alea (online). 7:2 (fev 2005).
- Ruiz Pérez, R. (1992). El análisis documental: bases terminológicas, conceptualización y estructura operativa. Granada: Ed. Universidad de Granada, 1992
- Saussure, F. (1973). Curso de linguística geral. 5. ed. São Paulo: Cultrix, 1973.
- Sagredo Fernández, F.; Izquierdo Arroyo, J. M. (1983). Concepción lógico-lingüística de la Documentación. Madrid: IBERCOM, 1983.
- Sagredo Fernández, F.; Izquierdo Arroyo, J. M. (1982). Reflexiones sobre "documento": palabra/objeto (I). // Boletín Millares Carlo. 5:1 (fev.1982).
- Van Dijk, T. (1972). Some aspects of text grammar: a study in theoretical linguistics and poetics. La haye, Mouton. 1972.

---

Enviado: 2015-04-01. Segunda versión: 2016-08-03.  
Aceptado: 2016-10-11.

---



---

# Análisis de contenido de la representación fotográfica de la crisis de los refugiados sirios y su incidencia en el framing visual

*Content analysis of the photographic representations of the Syrian refugee crisis and their visual framing impact*

---

Joaquín LÓPEZ DEL RAMO, María Luisa HUMANES

Universidad Rey Juan Carlos, Camino del Molino s/n, 28943 Fuenlabrada, España;  
joaquin.lopezdelramo@urjc.es; marialuisa.humanes@urjc.es

## Resumen

Se examina mediante un análisis de contenido la representación fotográfica de la crisis de los refugiados sirios en Europa central, acaecida entre mediados de 2015 y los primeros meses de 2016. El objetivo básico es identificar y clasificar los componentes denotativos y connotativos presentes en las fotografías, como documentos informativos visuales de especial incidencia en el framing visual. Interesa el tratamiento de los dos aspectos fundamentales para la fijación de marcos perceptivos: el contexto físico y el contexto humano, así como sus efectos cognitivo-afectivos. Sobre una muestra de 196 fotografías procedentes de 13 medios de prensa digital internacionales, se aplica una ficha de codificación original, que proporciona datos significativos. Los resultados caracterizan la representación iconográfica del hecho, marcando la preponderancia de lugares y escenarios de tránsito, ámbitos rurales y episodios de naturaleza estresante, pero sin excesiva carga dramática. Se enfatiza el aspecto grupal o colectivo sobre el individual, con clara mayoría de actores masculinos, jóvenes o de mediana edad. En general, los personajes expresan una actitud más pasiva que activa, con acciones moderadas o estatismo físico. La gestualidad no siempre se muestra explícitamente, pero en gran medida transmite naturalidad, aunque también son relevantes los sentimientos negativos. Se exhibe una imagen de cercanía solidaria con los refugiados, neutra de los cuerpos de seguridad y positiva de los cooperantes. Las conclusiones indican que la visión general del acontecimiento es realista, carente de tintes violentos o dramáticos, pero sin ocultar la dureza propia de la situación. No se detectan estereotipos retóricos en cuanto a la forma en que se exhiben las acciones y personajes.

**Palabras clave:** Análisis de contenido. Documentación fotográfica. Representación de la información visual. Framing visual. Prensa internacional. Refugiados. Fotogalerías. Crisis de los refugiados sirios.

## 1. Introducción

La llegada a Europa de refugiados provenientes de Siria se ha convertido en uno de los episodios más importantes en la vida política y social de los

## Abstract

This paper studies the visual framing components through a photojournalistic analysis applied to digital media outlets on the crisis of Syrian refugees in central Europe, which occurred between mid-2015 and early 2016. The main objective is to identify and classify the denotative and connotative components of the photographs as visual documents with special emphasis on the visual framing. We are interested in the treatment of the two fundamental aspects for fixing perceptual frameworks: the physical context and human context, as well as their cognitive-affective effects. On a sample of 196 photographs from 13 international digital media a content analysis is carried out. The results reveal the preponderance of places and traffic scenarios, rural areas and with certain stressful episodes component, but without excessive dramatic charge. The group or collective aspect is emphasized over the individual, with clear majority of male, young or middle-aged actors. Overall the characters express a more passive than active attitude, with moderate physical actions or statism. Gestures are not always shown explicitly, and largely transmitted naturally, but are also relevant negative feelings. An image of solidarity with refugees, neutral of the security forces and positive for the NGOs-personal is transmitted. The findings indicate that the overall vision of the event is realistic, devoid of violent or dramatic tone, but without obscuring the very harshness of the situation. No rhetorical stereotypes are detected.

**Keywords:** Content analysis. Photographic documentation. Visual representation. Visual framing. International press. Photo galleries. Refugees. Syria refugee crisis.

últimos años en Europa. Los gobiernos y la opinión pública en todos los países europeos se han movilizado ante un éxodo de población sólo comparable al de la Segunda Guerra Mundial.

ACNUR calcula que aproximadamente 5 millones de sirios han huido de la guerra en su país. Muchos de ellos han cruzado Turquía con la intención de iniciar una nueva vida en Europa. La llegada de los refugiados a las playas y puertos de Grecia, y su travesía a través de Macedonia, Serbia, Bulgaria, Rumanía y Hungría hasta sus destinos finales en los países más ricos de la Unión Europea, han sido mostradas por los medios de comunicación, la mayoría de ellos con una amplia cobertura fotográfica. Pero sin duda uno de los momentos más recordados y con más impacto en la opinión pública fue la publicación de las fotografías del niño Aylan Kurdi, de tres años, muerto en una playa turca. Ante cientos de personas habían perdido la vida en su huida hacia algún país de la Unión Europea, pero la publicación de esta imagen provocó la discusión sobre el papel de los medios de comunicación.

El European Journalism Observatory (EJO) ha patrocinado varios estudios sobre la cobertura de la crisis de los refugiados sirios –How Europe's newspapers reported the migration crisis, How media in Greece reported the migration crisis and Reporting hate crime: informing, not inciting- entendiendo la gran influencia que puede ejercer la forma de representar la crisis de los refugiados y sus implicaciones.

En concreto, la representación visual de los refugiados puede tener un importante efecto en la percepción de los mismos por parte de la opinión pública, puesto que la carga emotiva de la fotografía es mucho mayor que la de cualquier texto escrito, y ayuda a humanizar problemas sociales como es el caso que nos ocupa en este estudio.

## 2. Antecedentes

La perspectiva teórica que ha generado más estudios sobre la construcción mediática de los acontecimientos de actualidad es la teoría del framing o encuadre. Esta teoría nos explica el proceso por el cual los medios encuadran los acontecimientos sociales: (1) seleccionando algunos aspectos de una realidad percibida, recibiendo mayor relevancia en un mensaje, (2) asignándoles una definición concreta, (3) una interpretación causal, (4) un juicio moral y/o (5) una recomendación para su tratamiento (Entman, 1993).

Así, cuando aludimos al concepto de encuadre nos referimos a la manera en que el comunicador enfoca un tema y fija una agenda de atributos del mismo que puede influir sobre cómo pensamos acerca del mismo. Los encuadres, los enfoques, los marcos, permiten al periodista encargarse de grandes cantidades de información de una manera rápida y rutinaria y “envasarla” para el con-

sumo público. Las investigaciones experimentales sobre los efectos del framing han señalado que la manera en que se enfoca un tema influye en la percepción que el público desarrolla sobre ese asunto (Humanes, 2001).

Tradicionalmente la investigación se ha centrado en localizar los encuadres en los contenidos noticiosos de carácter textual. Sin embargo, las imágenes son también poderosos vehículos para generar representaciones de la realidad e influenciar en el conocimiento y las opiniones de los individuos a cerca de los temas de actualidad (Entman, 1991). Messaris y Abraham (2001: 220) afirman que:

The special qualities of visuals—their iconicity, their indexicality, and especially their syntactic implicitness – makes them very effective tools for framing and articulating ideological messages' controversial or might meet with greater audience resistance if they were conveyed through words.

Rodríguez y Dimitrova (2011) establecen que existen cuatro niveles de análisis visual para determinar encuadres. El primer nivel es el denotativo, es decir, en el que los encuadres son identificados a través de los objetos y otros elementos que se muestran en la imagen. El segundo nivel está relacionado con los recursos estilísticos. El tercer nivel es el connotativo, en el que además de mostrar qué elementos (personas, objetos, lugares) aparecen en las imágenes se consideran las ideas o conceptos asociados a estos elementos. Por último, las autoras mencionan el nivel de las imágenes como representaciones ideológicas.

En el contexto anglosajón existe ya una importante cantidad de trabajos que han analizado los encuadres visuales sobre diferentes temáticas: conflictos bélicos (Fahmy, 2010; Fahmy y Kim, 2008; Moriarty y Shaw, 1995; Parry, 2010; Pfau et al., 2006; Schwalbe, 2013), desastres naturales (Borah, 2009; Fahmy, Kelly y Kim, 2007), el sistema político (Coleman y Banning, 2006; Grabe y Bucy, 2009), problemas sociales (Heuer, McClure y Puhl, 2011). En España Muñiz, Igartua y Otero (2006) y Muñiz et al. (2008) han analizado los encuadres visuales sobre la inmigración contenidas en las fotografías periodísticas y en las imágenes de los informativos de televisión. López del Ramo ha abordado en distintos trabajos el análisis visual de los contenidos periodísticos en las publicaciones digitales (2010a; 2010b). Estos trabajos de López del Ramo han generado un modelo de análisis que es el utilizado en este trabajo.

Por su dimensión social, política y estratégica, la cobertura mediática de la guerra en Siria ha dado lugar ya algunos trabajos que abordan el framing



visual (Greenwood y Jenkins, 2015; Pantti, 2016). Pero también la crisis de los refugiados sirios en Europa central ha sido un acontecimiento que, además de centrar la atención de la prensa internacional de forma prioritaria durante varios meses, ha sido objeto de diversas investigaciones en el ámbito de las Ciencias de la Comunicación. Giannakopoulos (2016) ha abordado la representación de los refugiados en su llegada a las playas del Egeo. En el caso español Hoyer (2016) ha estudiado el encuadre noticioso de la crisis de los refugiados en los diarios *El Mundo*, *El País* y *ABC*, centrándose en los términos elegidos para designar a los refugiados.

### 3. Objetivos

Frente a los trabajos previos este artículo enfoca su interés no en las fotografías que cumplen una mera función de ilustrar las noticias textuales, sino en los géneros emergentes del fotoperiodismo digital, donde la fotografía es el componente esencial de la narrativa, los cuales marcan los estándares más altos de calidad y expresividad comunicativa a través de la imagen dentro de medios informativos digitales. El informe *Digital News Report* de 2016 indica que uno de los formatos con mayor crecimiento en lectores son las *picture stories* (Newman et. al., 2016:19), de ahí la importancia de su estudio.

Además, respecto a los trabajos que se han centrado en la prensa de un único país, en nuestro caso se ha ampliado el objeto de estudio a los medios internacionales de mayor relevancia.

En consecuencia, el objetivo general que se persigue en este trabajo es analizar el tratamiento de los encuadres visuales en las fotografías de prensa publicadas en medios digitales internacionales sobre la crisis de los refugiados sirios, acaecida en Europa central entre mediados de 2015 y los primeros meses de 2016, así como medir sus posibles efectos cognitivo-afectivos. La intención de fondo es proponer un procedimiento analítico basado en la indagación de un conjunto características objetivamente observables, cuantificables y significativas respecto al estudio de los marcos visuales en acontecimientos de crisis generadas por movimientos poblacionales masivos, en los que también hay un factor humano subyacente, como es el caso de ésta.

De forma más específica, se establecen los siguientes objetivos particulares:

1. Registrar y computar los tipos de episodios y acciones relacionados con la crisis que son prioritario objeto de plasmación en las fotografías y, por exclusión, los que no lo son, así como la localización de los mismos.

2. Identificar las propiedades del marco físico mostrado en las imágenes, que otorgan contexto ambiental a los acontecimientos relacionados con la crisis de los refugiados. Interesa conocer el código escenográfico y los elementos materiales que lo integran, en suma, el "clima" espacial que envuelve los acontecimientos.
3. Realizar una exploración y tipificación de los actantes humanos por rol, grupo de edad y sexo, y averiguar su grado de protagonismo en función de dichos atributos.
4. Comprobar si en las fotografías se reflejan situaciones de tensión o violencia que pudieran estar latentes en este tipo de acontecimientos
5. Analizar cómo se representa el factor humano, es decir la manera en que las personas aparecen reflejadas en las fotografías y las posibles diferencias existentes entre unas y otras en función de sus roles. Para ello se toman en cuenta como aspectos más sustanciales, como la proximidad de los personajes, la gestualidad, la actividad o pasividad y las acciones que realizan.
6. Indagar y determinar la posible existencia de tendencia o patrones fijos que pudieran dar lugar a una imagen estereotipada y retórica de los personajes según su rol. En caso de apreciarse, realizar una caracterización general de los diferentes roles y la tendencia positiva o negativa que transmiten objetivamente las imágenes respecto a cada uno de ellos.

#### 3.1. Hipótesis de investigación

Sobre la base de las consideraciones realizadas en el marco teórico y de los objetivos anteriormente formulados, se plantea como hipótesis de investigación en relación a las imágenes de la crisis de los refugiados sirios en Europa central publicadas en los medios de prensa digital internacional la siguiente:

H1: Las imágenes transmiten una visión solidaria a favor de los refugiados, que, por un lado, enfatiza el sufrimiento de los mismos y su papel como víctimas y, por otro, acentúa las dificultades y trabas que se les pone por parte de los países receptores, sin entrar en otras consideraciones. En tal sentido, es verosímil que se produzca un cierto maniqueísmo y retoricismo en el tratamiento fotoperiodístico de estos acontecimientos.

## 4. Metodología

### 4.1. Muestra de contenidos

Dado que la investigación aborda un asunto de escala internacional, habida cuenta de la enorme cantidad de medios informativos digitales existentes en todo el mundo y en busca de resultados representativos y significativos, el análisis se acota a una serie de medios que cumplen varias premisas. Los criterios para su elección fueron los siguientes:

- Reconocido prestigio internacional y amplia difusión.
- Medios periodísticos pertenecientes a tres categorías básicas: Periódicos, revistas y sitios de noticias generalistas, para contar con una visión más enriquecedora del fenómeno.
- Caracterizados y reconocidos por la importancia que otorgan al tratamiento de la fotografía, plasmada en la inclusión habitual de formatos o géneros específicamente fotoperiodísticos (galerías, foto-reportajes o foto-ensayos, con imágenes de calidad y formato grande) y en el hecho de contar con una sección específica.
- Ubicación en diferentes países de Europa central u occidental, Estados Unidos y/o América del Sur, por tratarse de zonas en las que el periodismo digital se encuentra más avanzado, existen mayores niveles de libertad de prensa y, en el caso del continente europeo, en razón a su mayor cercanía geográfica a los hechos

Con arreglo a estos requerimientos, tras una pre-selección previa, fueron elegidos los siguientes periódicos: *Le Monde*, *New York Times*, *Telegraph*, *The Guardian*, *El País*, *Folha de Sao Paulo*, *Bild* y *Wiener Zeitung*. En cuanto a las revistas, se seleccionaron *Time*, *Paris Match*, *Spiegel* y *Stern*, y como sitios de noticias, por su relevancia especialmente por la calidad de sus trabajos fotográficos, *Msnbc.com*. Como puede apreciarse, fueron ligeramente primados los medios de origen alemán, debido a ser éste el país destino mayoritario del flujo migratorio de refugiados. Dentro de cada medio, se escogió una sola pieza (fuese galería o foto-ensayo/foto-reportaje) y en aquellos en que había varias piezas relativas a las crisis de los refugiados sirios se escogió una al azar. Las URLs, títulos, tipos y número de fotografías de las piezas analizadas se incluyen el Apéndice 1.

Desde el punto de vista cuantitativo, la muestra comprende un total de 195 fotografías correspondientes a 13 piezas individuales, una por cada

medio analizado, siendo éstas de dos tipos: galerías fotográficas y foto-reportajes/foto-ensayos. El mayor número de imágenes corresponde a los periódicos (108), seguidos por las revistas (61) y los portales de noticias (26). Cada fotografía se define como unidad de análisis.

Desde el punto de vista cronológico, los movimientos masivos de refugiados sirios con dirección a Centroeuropa comenzaron en torno al mes de junio de 2015 y se extendieron más o menos hasta comienzos de 2016. Las piezas seleccionadas corresponden a este lapso temporal (junio 2015 a febrero 2016).

### 4.2. Procedimiento e instrumento

En función del objeto de estudio, los objetivos y la hipótesis formulada, se ha escogido como procedimiento más idóneo el análisis de contenido, dada su idoneidad como técnica cuya utilización permite evaluar de forma empírica las características del mensaje informativo (en este caso icónico-visual y de codificación fotográfica) y partir de ello obtener datos objetivos y significativos sobre el mismo (Salvador Benítez y Sánchez Vigil, 2016). Es, por tanto, la vía procedimental apropiada para determinar los atributos que conforman el framing visual.

La herramienta de análisis empleada ha sido una ficha de codificación de diseño original basada en el modelo de López del Ramo (2014), readeuada a los objetivos del presente estudio, siendo seleccionadas únicamente aquellas variables más pertinentes en relación a los objetivos e hipótesis. La ficha está incluida en el Apéndice 2 con el listado de variables y valores correspondientes al completo. Consta de 17 variables significativas para la determinación de los aspectos fundamentales del framing visual, que se presentan articuladas en dos grandes categorías:

a) Datos biográficos (elementos denotativos: lo que aparece fotografiado). Dentro de este apartado figuran diez variables: fecha, lugar, episodio, escenario, elementos escenográficos, momento del día, tipo de actor, rol, edad y sexo. Las seis primeras informan sobre el contexto físico-espacial, y las cuatro últimas sobre el contexto humano.

b) Datos de codificación fotográfica (elementos connotativos: la forma en que el fotógrafo ha captado la escena utilizando los recursos técnicos del lenguaje fotográfico). Las variables aquí incluidas son siete: encuadre, ángulo, luminosidad general de la imagen, equilibrio, objetos interpuestos, estatismo/dinamismo y gestualidad.

### 4.3. Codificación

La codificación de la muestra corrió a cargo de dos investigadores. Al objeto de asegurar la unificación de criterios y la fiabilidad del trabajo, se seleccionó de forma aleatoria una submuestra compuesta por 37 unidades (fotografías), equivalentes al 18,9% de la muestra y pertenecientes a dos piezas de otros tantos medios informativos (*The Guardian* y *Paris Match*). Para el acuerdo interjueces se ha utilizado el alpha de Krippendorff ( $K_a$ ), que ha arrojado un valor total de .85.

El proceso de captura y codificación se extendió entre diciembre de 2015 y febrero de 2016. Para la captura de las piezas completas se utilizó el software: Screen Capture (navegador Chrome).

Para el análisis estadístico de los datos se ha utilizado el programa SPSS. Se han realizado análisis univariados de frecuencias y porcentajes, mientras que los análisis bivariados se han llevado a cabo a través de tablas de contingencia. Para el contraste de hipótesis se ha utilizado la prueba Chi Cuadrado.

## 5. Resultados

A continuación se exponen los resultados, comenzando por los obtenidos del análisis univariado, los cuales a su vez se agrupan en tres grandes categorías correspondientes a las dos dimensiones fundamentales que atañen al framing visual, más una que incumbe simultáneamente a ambas:

1. Contexto físico-espacial en el que transcurren los acontecimientos mostrados en las fotografías.
2. Contexto humano, es decir, aquellos que describen a las personas que aparecen en las imágenes.
3. Rasgos fotográficos que atañen al modo en que se representan tanto del contexto físico-espacial como del humano.

El último apartado se dedica a la exposición de los resultados obtenidos a partir de estadísticos bivariados, a fin de comprobar el grado de correspondencia entre múltiples características que resultan significativas a efectos de cumplir los objetivos y verificar la hipótesis del trabajo.

### 5.1. Análisis univariados

#### 5.1.1. El contexto geográfico-físico

*Lugares.* Respecto a los países donde fueron obtenidas las fotografías que componen la muestra, la distribución es la siguiente, en orden de frecuencia decreciente: Grecia (25%), Turquía

(11,2%), Croacia (9,7%), Frontera Greco-Macedonia (9,7%), Hungría (9,7%), Macedonia (8,2%), Frontera Serbio-Húngara (6,1%), Frontera Húngaro-Croata (4,6%), Serbia (4,6%), Frontera Serbo-Croata (4,1%), Frontera Esloveno-Croata (3,1%), Frontera Greco-Turca (1,5%), Frontera Sirio-Turca (1%) y Alemania, Austria e Italia, cada una de las tres naciones con un 0,5% del total respectivamente. Teniendo en cuenta la procedencia y el destino de esta corriente migratoria excepcional, cabe apreciar que salvo los tres últimos (Alemania, Austria e Italia) y la frontera de Siria-Turquía el resto son países de paso para los refugiados. Apenas hay fotografías de éstos en los puntos de origen y destino del viaje.

*Escenarios.* Han sido registrados 12 tipos de escenarios físicos reconocibles, que de mayor a menor en porcentaje en el total de la muestra son los siguientes: Campos y bosques (20,9%). Estaciones ferroviarias/trenes (13,8%). Campamentos de refugiados (13,3%). Calles y parques urbanos (12,8%). Playas (11,2%). Fronteras (9,7%). Autobuses (3,1%). Comisarias (1,5%). Puertos (2%). Mar (1%). Barcos (0,5%). Domicilios (1,5%). A ellos se añade un grupo de fotos captadas en escenarios indeterminados, que suponen el 8,7% del total. A la vista de estos datos, cabe indicar que en todos los casos se trata de escenarios de tránsito o viaje. Asimismo, la gran mayoría de escenarios son espacios abiertos, prácticamente a la intemperie, y casi en su totalidad pertenecen a lugares de ámbito rural, no a núcleos urbanos grandes.

*Elementos y objetos físicos.* La presencia de determinados elementos u objetos físicos en las imágenes (y en ocasiones también su ausencia) puede constituir uno de los rasgos más significativos en relación al marco visual que se ofrece de un acontecimiento, por cuanto muchas veces incluso encierran una carga connotativa o simbólica. En el caso de esta investigación, se han contabilizado los siguientes tipos de elementos, ordenados de mayor a menor frecuencia y en porcentaje sobre el total de las unidades de análisis: enseres de refugiados (54,6%), tiendas de campaña (8,2%), armas (7,7%), vallas (6,6%), alambradas (6,6%), maletas (5,6%), basura (5,6%), material sanitario (4,6%), símbolos religiosos (2%), juguetes (1,5%), banderas (0,5%) y pancartas (0,5%). Como elementos de connotación no positiva, relacionados con conceptos como represión o precariedad, podrían destacarse las armas, vallas, alambradas y basura. Se constata que el grado de presencia de este tipo de objetos es porcentualmente bajo, por lo que énfasis escénico de los mismos existe, pero es discreto,

predominando más los objetos y elementos de connotación neutra.

*Episodios.* Por lo que respecta al encuadre episódico, las fotografías analizadas se enmarcan en cinco tipologías básicas, cuya frecuencia en términos porcentuales sobre el conjunto de la muestra se expresa entre paréntesis. En la tabla I se muestran los porcentajes.

<i>Episodios</i>	%
Vida en campos y lugares de acogida	32,1
Viajes y desplazamientos	28,6
Esperas en trámites y traslados	22
Control de refugiados	9,7
Asistencia a los refugiados	5,6
Imágenes que no es posible relacionar con ningún episodio	2

Tabla I. Episodios reflejados en la fotografía

Estos episodios componen una especie de estructura temática que define los enfoques que los medios dan a este acontecimiento a través de las fotografías, como aspectos de la crisis de los refugiados sirios que más se destacan, aunque evidentemente no sean los únicos. Sobre la naturaleza de estos episodios, puede estimarse que, salvo uno, el correspondiente a la asistencia a los refugiados, el resto se trata de situaciones con un componente estresante o de tensión más o menos acentuado.

*Momentos del día.* Respecto al momento del día en que fueron obtenidas las fotografías, se constata que el 19,4% de las mismas son nocturnas, y el resto (80,6%) se captó en diferentes franjas horarias diurnas.

### 5.1.2. El contexto humano

La correcta lectura de los resultados de las variables de atañen a actores, lleva a considerar que en una misma fotografía pueden aparecer varios actores y ser estos de diferentes tipos (individuales, grupales, por rol, sexo, edad, etc.), motivo por el cual los recuentos totales no van a coincidir en algunos indicadores con el total de unidades de análisis, sino que en muchos casos serán superiores, es decir, aparecen más casos que unidades tiene la muestra. Se han codificado hasta tres actores principales por fotografía.

Asimismo, es necesario indicar que el análisis de los actores se ha centrado en aquellos que aparecen en las imágenes con la suficiente completitud, cercanía y nitidez como para poder determinar sus rasgos fundamentales, no tomándose en consideración otros personajes que figuren dentro del cuadro de manera parcial (por ejemplo, sólo se aprecia algún miembro del cuerpo) o muy alejados en el campo de visión hasta el punto no se ser posible reconocer su rol u otros detalles identificativos básicos, salvo que éstos sean evidentes por razones contextuales o de otro tipo.

*Actor grupal / individual y roles-* Los resultados obtenidos indican un claro predominio de los actores colectivos o grupales sobre los individuales: un 67,4% de los primeros respecto al 32,6% de los segundos. En el conjunto de la muestra se registra la aparición de cinco grandes tipos o grupos de actores humanos principales, clasificados en función de su rol. El mayoritario es el de los refugiados (79,7%), seguido de fuerzas de seguridad (11,7%), cooperantes (4,7%), habitantes locales de los países donde transcurren los hechos (3,5%) y periodistas (0,4%). Los anteriores porcentajes se refieren a las fotografías en las que hay actores humanos, quedando al margen el 0,4% de las imágenes que componen la muestra por carecer de actores humanos.

*Edad y sexo.* En cuanto a la distribución por edades, se evidencia que el porcentaje más alto corresponde al de las fotografías que muestran personas de diferentes edades (32,9%). Cuando sólo aparecen personas de la misma edad, predominan ligeramente las de jóvenes (27%), seguidas del grupo de personas maduras (26,6%) y niños (10,8%), seguidas, ya a más distancia, de adolescentes (2,3%) y ancianos, con apenas el 0,5% del total. Por sexos, los resultados indican que en el 57,8% de las fotografías únicamente aparecen hombres, y en el 11% sólo mujeres, mientras que en el 31,2% figuran simultáneamente personas de ambos sexos, es decir, cuantitativamente dominan con claridad los actores masculinos.

*Gestualidad.* Tomando como referencia los criterios expuestos por Flora Davis (2005), el análisis de esta variable se centró en chequear la presencia/ausencia de 20 tipos diferentes de expresiones gestuales, a priori acordes y previsibles en un acontecimiento de esta naturaleza. El criterio adoptado fue que si en el proceso de codificación de detectaban otros gestos no contemplados inicialmente, se agregaban a la lista de valores. De la veintena de tipos gestuales considerados con carácter previo, se ha comprobado que cuatro no están presentes en ninguna fotografía: ira o agresividad, aburrimiento, decepción y esfuerzo. Los

resultados ponen de manifiesto que, curiosamente, la tipología gestual con mayor frecuencia de aparición es la indeterminada, es decir, aquella en la que el sujeto aparece en una posición en la que no es posible reconocer o vislumbrar su gesto (por ejemplo, por encontrarse de espaldas), lo cual se observa en el 37,8% de las imágenes.

Entre las categorías que se refieren a un gesto reconocible, su peso porcentual se distribuye de la siguiente manera: naturalidad (30,1%), cansancio (23,5%), preocupación o incertidumbre (15,8%), alegría (8,7%), apelación (6,1%), tristeza (6,1%), desesperación (5,6%), afecto o ayuda (5,6%), indiferencia (4,1%), firmeza o mando (4,1%), desconfianza (2,6%), miedo (2,6%), respeto u obediencia (2%), amenaza (2%) y reivindicación (0,5%).

Puede apreciarse que los tipos gestuales más cargados de connotación negativa (cansancio, preocupación, tristeza, desesperación, indiferencia, desconfianza, miedo y amenaza) suman un 58,2%. Por el contrario, los de connotación positiva (alegría y afecto/ayuda) únicamente suponen el 14,3%, siendo el resto las tipologías de carácter neutro (apelación, indiferencia, firmeza, respeto/obediencia y reivindicación), que totaliza el 10,2%. Así pues, considerando la totalidad de imágenes y de personajes que aparecen en ellas, predominan los gestos que evocan sentimientos negativos. En un epígrafe posterior se expondrán los resultados fruto del cruce entre las variables tipología gestual y rol de los actores.

### 5.1.3. Rasgos de representación fotográfica

*Encuadre.* Tomando en consideración la totalidad de la muestra, los tres encuadres predominantes, con gran diferencia sobre el resto, son el plano grupal (32,7%), seguido del general (32,1%) y del plano medio (20,4%). En un nivel de empleo mucho más reducido figuran el plano corto (7,1%) y el entero (5,1%), y ya a distancia el plano general largo (1,5%) y el plano detalle (0,5%). Estos resultados ponen de manifiesto que el marco de visión ofrecido en las fotografías no es excesivamente corto ni distante, si bien predomina la cercanía a hechos y personajes sobre la lejanía, lo cual posee un sentido connotativo bastante elocuente, en la medida que permite mostrar las circunstancias personales de los afectados, incluso las reacciones y sentimientos, lo que es un factor fuertemente subjetivador.

*Ángulo visual.* En el 85,7% de las fotografías analizadas no se observan efectos de distorsión angular significativos, sino que el ángulo de visión es natural. La utilización de lentes tipo gran an-

gular se aprecia en el 8,7% de la muestra, mientras las imágenes que presentan los efectos de distorsión característico del teleobjetivo suponen el 5,1% del total. Por ello cabe concluir que la tendencia en el tratamiento de este aspecto de codificación fotográfica es claramente neutra, transmitiendo así mayor grado de realismo.

*Luminosidad general.* El tratamiento de la luminosidad de la imagen es mayoritariamente equilibrado (39,8%) lo que implica que predominan las imágenes con tonos equilibrados. En el 24,5% de las unidades que componen la muestra las escenas poseen una luz brillante. No obstante, en el 31,1% de los casos las tonalidades lumínicas de las fotografías son oscuras, y en el 5,1% impera una luz opaca, lo que sumado casi iguala al porcentaje neutro. En suma, se aprecia una presencia significativa de imágenes que reflejan un tono ambiental sombrío, aunque ésta no sea la tónica de la mayoría de fotografías.

*Estatismo/dinamismo.* El indicador estatismo/dinamismo se refiere a la naturaleza de las acciones captadas en las fotografías en función a su grado de movimiento. Se distinguieron cuatro posibilidades: estatismo en posado: los personajes posan ex profeso y de forma manifiesta ante la cámara, a modo retrato preparado; estatismo físico: los personajes son captados en su actitud natural, siendo ésta la quietud; acción moderada: los personajes realizan acciones poco vigorosas; acción rápida: aquella en la que dominan los movimientos intensos, veloces o fuertes.

Los resultados obtenidos en el conjunto de la muestra indican una clara mayoría de fotografías cuya categoría dinámica es la Acción moderada (52,6%), seguida del estatismo físico (43,9%). Ya a gran distancia están las imágenes de estatismo en posado (1,5%) y acción rápida (1%). Otro 1% de las imágenes muestran un tipo de dinamismo indeterminado. Por tanto, puede afirmarse que predominan claramente las imágenes que muestran escaso o muy discreto dinamismo, acciones poco intensas.

## 5.2. Análisis bivariados

En este epígrafe se muestran los resultados de los cuatro cruces bivariados que se han considerado más significativos respecto al objeto y alcance de esta investigación, teniendo presentes la obligada limitación de espacio en una publicación de esta naturaleza, si bien el corpus de datos obtenidos en el proceso de trabajo era susceptible de un desarrollo más amplio.

### 5.2.1. Actores (rol) / Sexo

La asociación entre el rol de los actores y su sexo indica como primer resultado significativo ( $X^2=12.660, .049$ ) que son mayoritarias en todos los roles las fotografías donde únicamente aparecen personajes de sexo masculino. De una manera más pormenorizada, el tipo de rol donde la anterior tendencia se manifiesta más acentuada es el de los periodistas (100%), seguido de los cuerpos de seguridad (62,7%), refugiados (60,1%), cooperantes (47,6%) y habitantes locales (47,1%).

Las fotografías en las que aparecen a la vez personas de ambos sexos son el segundo grupo más numeroso en todos los roles, salvo el de los habitantes locales. Destacada el porcentaje de este tipo de imágenes en el rol de cooperantes (38,1%), seguido de los cuerpos de seguridad (32,2%) y refugiados (28,9%).

En cuanto a las fotografías en las que sólo aparecen mujeres, se constata que es el caso minoritario en todos los roles, salvo en habitantes locales. El porcentaje menor de mujeres se da en los roles de periodista (valor nulo), seguidos de cuerpos de seguridad (5,1%), refugiados (11%) y cooperantes (14,3%), y el mayor en habitantes locales (35,3%).

### 5.2.2. Actores (rol) / Encuadre

El encuadre marca el grado de cercanía de los personajes fotografiados respecto al lector, por lo que es un factor de particular relevancia como indicativo de la proximidad percibida. Por ello se ha realizado un análisis bivariado interrelacionando los encuadres con los roles de los actores, del cual se desprende como resultado más genérico la existencia de una proporcionalidad entre el número de fotografías de cada rol y la variedad de encuadres utilizados. Así, el único rol del cual existen fotografías con todos los tipos de encuadres es el de refugiados, y en el resto la variedad de encuadres disminuye el mismo orden que el número de imágenes: cuerpos de seguridad, cooperantes, habitantes locales y periodistas. La prueba Chi cuadrado no ha resultado estadísticamente significativa.

Respecto a los refugiados, el porcentaje más alto de fotografías corresponde al encuadre grupal (33,2%), seguido del general (30,7%), del plano medio (22,3%), el plano corto (8,4%) y el general largo (0,5%). En relación con los otros roles, a los refugiados les corresponde el mayor porcentaje de planos medios y cortos.

El plano general es el mayoritario en las fotografías protagonizadas por personajes de cuerpos

de seguridad y periodistas (éstos únicamente figuran en tomas de plano general). Respecto a los primeros, cabe señalar que aparecen en todos los tipos de encuadres, salvo el general largo, y asimismo es relevante la cantidad de imágenes de los mismos donde se muestran cercanos: 33,3% en plano grupal, 16,7% en plano medio y 6,7% en plano corto. El rol cooperante registra el mismo porcentaje en encuadres general y grupal (33,3%), y un 24% en encuadre medio, pero posee un valor nulo en plano corto y general largo. Por su parte, los habitantes locales aparecen mayoritariamente en plano grupal (44,4%), aunque también figuran en plano corto (22,2%), general (22,2%) y general largo (11,1%), sin que exista ninguna imagen de ellos en encuadre entero y medio.

Para medir el grado de cercanía se ha empleado como criterio la suma de los porcentajes de los planos que muestran más próximos a los personajes, es decir: grupal + entero + medio + corto. A partir de esta premisa, el resultado indica que el tipo rol presentado en las imágenes como más cercano son los Refugiados (68,9%), seguido de Cooperantes y Habitantes locales (ambos con el 66,6%), Cuerpos de seguridad (56,7%) y Periodistas (0%).

### 5.2.3. Actores (rol) / Estatismo-dinamismo

Se constata que, para todos los tipos de rol, la categoría dinámica mayoritaria en los personajes es acción moderada. La segunda categoría que más aparece es estatismo físico en las fotos de refugiados, cuerpos de seguridad, cooperantes y habitantes locales. En todos los roles las dos dinámicas minoritarias son el estatismo-posado y la acción rápida, es decir, los puntos extremos de la escala. La prueba Chi cuadrado no ha resultado estadísticamente significativa.

De manera más concreta, la categoría Estatismo-posado sólo aparece en unas pocas fotos de refugiados, registrando valor nulo en los roles cuerpos de seguridad, cooperantes, periodistas y habitantes locales. Por lo tanto, se puede afirmar que existe bastante homogeneidad en el tratamiento del factor dinámico entre los diferentes roles.

### 5.2.4. Actores (rol) / Expresión facial

Como se indicó en el apartado correspondiente, en el análisis de muestra fueron detectadas 16 tipologías gestuales distintas convertidas en variables dicotómicas de presencia-ausencia de cada gesto. La lectura cruzada entre este indicador y el tipo de rol ofrece resultados significativos en cuando al encuadre visual dado por la prensa

a la crisis de los refugiados sirios, ya que la expresividad que transmiten los distintos personajes a través de sus gestos es un componente con notable peso específico en la representación fotográfica de un hecho noticioso y puede marcar claras diferencias entre la imagen ofrecida unos u otros. Las pruebas Chi cuadrado no han resultado estadísticamente significativas para las expresiones miedo ( $X^2= 18.989, .000$ ), respeto ( $X^2= 8.494, .037$ ), indiferencia ( $X^2= 26.924, .000$ ) y apelación ( $X^2= 21.051, .001$ ).

Desde el punto de vista de las propias categorías, la gestualidad indefinida es predominante en los roles refugiados, cuerpos de seguridad y habitantes locales. Por su parte, en el rol cooperantes es mayoritario el rasgo gestual de naturalidad, e igualmente ocurre en el caso de los periodistas si bien éste rol tiene escasa significación por su mínima presencia en la muestra (una sola unidad).

Los refugiados son el único rol en el que se aprecian todas las tipologías gestuales, lo cual es lógico por poseer mayor cantidad de imágenes, con bastante diferencia sobre los demás. El rango de gestos observable en los personajes de cuerpos de seguridad es también muy amplio, pues únicamente no se ha hallado el gesto de reivindicación, dándose los 15 restantes.

En las fotografías de cooperantes se constata la ausencia de siete tipologías gestuales: tristeza, desesperación reivindicación, indiferencia, firmeza o mando, desconfianza y amenaza, como puede apreciarse, la mayor parte de ellas con una carga subjetiva negativa.

Respecto a las imágenes donde son protagonistas los habitantes locales, no se registran los gestos de indiferencia, firmeza o mando y desconfianza, y sí todos los demás. En la única fotografía correspondiente al rol periodistas sólo aparece como rasgo gestual la naturalidad.

Ya expuestas las tendencias básicas de este análisis bivariado, más que la mera relación exhaustiva indicador por indicador, se considera más elocuente para la identificación de los marcos cognitivo-afectivos agrupar las tipologías gestuales con las que trabajamos en tres grandes bloques, según su connotación: negativas, positivas y neutras, y comprobar su correlación en cada rol. Las tipologías claramente asociables con sentimientos negativos o problemáticos serían: tristeza, preocupación, desesperación, cansancio, miedo, reivindicación, desconfianza y amenaza. Los tipos de gestos con carga semántica positiva son alegría y afecto/ayuda. Por último, consideramos que tienen una connotación neutra (o al menos no necesariamente negativa) las tipologías naturalidad, respeto, indiferencia y

firmeza. Situaríamos aparte la categoría de gesto indefinido por razones obvias.

## 6. Conclusiones

A continuación, exponemos las conclusiones más importantes del análisis:

1. El contexto geográfico de las fotografías nos sitúa en los puntos más conflictivos de este acontecimiento, que fueron los países balcánicos que atravesó la corriente de refugiados en su camino hacia la Europa centro-occidental. El encuadre episódico se orienta hacia las situaciones álgidas, con mayor carga conflictiva o de tensión (zonas fronterizas o de aglomeración de refugiados), lo que, por otra parte, suele concitar el interés informativo en cualquier hecho noticioso.

2. Apenas se plasman escenarios urbanos, sino fundamentalmente rurales y fronteras secundarias o marginales, lo que refuerza la impresión de irregularidad o éxodo caótico respecto del acontecimiento. Los elementos físicos presentes en las imágenes son los que cabe esperar en este tipo de movimientos masivos, pero a través de los objetos no se muestra una sobrecarga dramática excesiva, más allá de la aparición ocasional de alambradas de espinos. La escenografía y los objetos o enseres que aparecen retratados no presentan apenas vinculación con lo religioso, cuando en realidad es el trasfondo de la crisis.

3. Desde el punto de vista de contexto humano, la noticia es exhibida más en su dimensión de fenómeno colectivo que de problemática individual, como revela el hecho de que en la mayor parte de las imágenes aparecen grupos de personas. Los refugiados son los principales protagonistas humanos, en lógica correspondencia con la naturaleza del acontecimiento. Los actores fundamentales son hombres, tanto jóvenes como de media edad, y ello es así en todos los roles, pero especialmente entre los refugiados.

4. Resulta llamativa la mayoritaria presencia de fotografías donde, por el tipo de encuadre y/o composición, no se aprecian los gestos de las personas, y también son abundantes aquellas en las que los personajes transmiten naturalidad, lo cual aporta un componente de cierta asepsia en la plasmación del acontecimiento. Asimismo, es numéricamente significativa la aparición de gestos que expresan sentimientos negativos, como cansancio o sufrimiento, sobre todo entre los refugiados, pero en pocas imágenes se llega a extremos dramáticos.

5. Los refugiados son el grupo de personajes que aparecen captados en una distancia más cercana al objetivo. Ello permite visualizar sentimiento y emociones, y otorga un mayor grado de

proximidad subjetiva con el lector; por así decirlo, humaniza más la imagen que se ofrece de ellos.

6. La visión general que se transite de este hecho es fundamentalmente tranquila, sosegada y más bien realista en cuanto a su modo de captación fotográfica. Ni siquiera en los episodios más tensos aparecen escenas violentas. Los cuerpos de seguridad manifiestan su papel de mando, pero no de una manera brusca o agresiva. Impera la moderación y la contención como tónica general, sobre el telón de fondo de un acontecimiento que por sí mismo es traumático.

7. No se detectan de manera manifiesta patrones fijos o estereotipos muy exagerados respecto a los personajes según su rol, debido a que, según se ha expuesto, los aspectos que inciden en la manera como están recreados en las fotografías no indican orientaciones exageradas, ni mucho menos caricaturescas. Las imágenes transmiten una visión solidaria a favor de los refugiados, y revelan las dificultades sufridas en los países de tránsito o receptores, pero en absoluto de forma exagerada, y por tanto no se llega a un maniqueísmo retórico en el tratamiento fotoperiodístico de esta crisis, por lo que se confirma solo parcialmente la hipótesis central de investigación.

## Referencias

- Borah, Porismita (2009). Comparing visual framing in newspapers: Hurricane Katrina versus tsunami. // *Newspaper Research Journal* 30 (1). 50.
- Coleman, Renita y Banning, Stephen (2006). Network TV news' affective framing of the presidential candidates: Evidence for a second-level agenda-setting effect through visual framing. // *Journalism & Mass Communication Quarterly* 83 (2). 313-328.
- Entman, Robert (1991). Symposium framing US coverage of international news: Contrasts in narratives of the KAL and Iran air incidents. // *Journal of communication*, 41(4). 6-27.
- Entman, Robert (1993). Framing: Toward clarification of a fractured paradigm. // *Journal of communication*, 43(4). 51-58.
- Fahmy, Shahira (2010). Contrasting visual frames of our times: A framing analysis of English-and Arabic-language press coverage of war and terrorism. // *International Communication Gazette* 72 (8). 695-717.
- Fahmy, Shahira y Kim, Daekyung (2008). Picturing the Iraq War Constructing the Image of War in the British and US Press. // *International Communication Gazette*, 70(6). 443-462.
- Fahmy, Shahira, Kelly, James y Kim, Yung Soo (2007). What Katrina revealed: A visual analysis of the hurricane coverage by news wires and US newspapers. // *Journalism & Mass Communication Quarterly* 84 (3). 546-561.
- Davis, Flora (2005). *La comunicación no verbal*. Madrid: Alianza.
- Giannakopoulos, Georgios (2016). Depicting the pain of others: Photographic representations of refugees in the Aegean Shores. // *Journal of Greek Media & Culture* 2(1). 103-113.
- Grabe, Maria Elizabeth y Bucy, Erik (2009). *Image bite politics: News and the visual framing of elections*. Oxford University Press.
- Greenwood, Keith y Jenkins, Joy (2015). Visual framing of the Syrian conflict in news and public affairs magazines. // *Journalism Studies* 16 (2). 207-227.
- Heuer, Chelsea; McClure, Kimberly y Puhl, Rebecca (2011). Obesity stigma in online news: a visual content analysis. // *Journal of health communication* 16 (9). 976-987.
- Hoyer, Ariana (2016). Spanish News Framing of the Syrian Refugee Crisis. // WWU Honors Program Senior Projects. Paper 26. [http://cedar.wvu.edu/wwu\\_honors/26](http://cedar.wvu.edu/wwu_honors/26)
- Humanes, María Luisa (2001). El encuadre mediático de la realidad social. Un análisis de los contenidos informativos en televisión. // *Zer*, 11. 119-142.
- López del Ramo, J. (2010a). El tratamiento fotoperiodístico en las portadas de los diarios digitales. Propuesta y aplicación de un modelo de análisis. // *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*, (11). 77-99.
- López Del-Ramo, J. (2010b). Configuración y contextualización de las galerías fotográficas en los diarios on-line. Propuesta de analítica aplicada. // *El profesional de la información*, 19(5). 469-475.
- Messarís, Paul y Abraham, Linus (2001). The role of images in framing news stories. // Reese, Stephen D., et al., (eds). *Framing public life: Perspectives on media and our understanding of the social world*. London: Routledge. 215-226.
- Moriarty, Sandra y Shaw, David (1995). An Antiseptic War: Were News Magazine Images of the Gulf War Too Soft? // *Visual Communication Quarterly* 2 (2). 4-11.
- Muñiz, Carlos; Igartua, Juan José y Otero, José Antonio (2006). Imágenes de la inmigración a través de la fotografía de prensa. Un análisis de contenido. // *Comunicación y Sociedad* 19 (1). 103-128.
- Muriel, Carlos Muñiz et al. (2008). Imágenes periodísticas de la inmigración: aportaciones metodológicas al estudio de la comunicación visual. // *Análisi: quaderns de comunicació i cultura* 37. 31-48.
- Newman, Nic, et al. (2016). *The Reuters Institute Digital News Report 2016*. Oxford: Reuters Institute for the Study of Journalism. <https://ora.ox.ac.uk/443/objects/uid:94c3a5e7-8eb0-4841-b7cb-ca659d9c7261>.
- Pantti, Mervi (2016). Seeing and not seeing the Syrian crisis: New visibility and the visual framing of the Syrian conflict in seven newspapers and their online editions. // *JOMEC journal* 4.
- Parry, Katy (2010). A visual framing analysis of British press photography during the 2006 Israel-Lebanon conflict. // *Media, War & Conflict* 3 (1). 67-85.
- Pfau Michael et al. (2006). The effects of print news photographs of the casualties of war. // *Journalism and Mass Communication Quarterly* 83(1). 150-168.
- Rodríguez, Lulu y Dimitrova, Daniela (2013). The levels of visual framing. *Journal of Visual Literacy*, 30(1). 48-65.
- Salvador Benítez, Antonia y Sánchez Vigil, Juan Miguel (2016). La fotografía como documento informativo en las webs de los ministerios españoles. Acceso, visibilidad, funcionalidad y políticas de uso. // *Revista Española de Documentación Científica*, 39 (2). 1-9. DOI: <http://dx.doi.org/10.3989/redc.2016.2.1295>.
- Schwalbe, Carol B (2013). Visually framing the invasion and occupation of Iraq in Time, Newsweek, and US News & World Report. // *International Journal of Communication* 7. 24.



## 7. Apéndice I: Identificación de las piezas fotoperiodísticas objeto de análisis

(Tipo de pieza: FR= Foto-reportaje G= Galería)

### 7.1. Periódicos

[http://www.lemonde.fr/europe/visuel/2015/09/25/face-aux-migrants-les-portes-de-l-europe-verrouillees\\_4768411\\_3214.html](http://www.lemonde.fr/europe/visuel/2015/09/25/face-aux-migrants-les-portes-de-l-europe-verrouillees_4768411_3214.html). Tipo de pieza: FR (35 fotos)

<http://www.nytimes.com/interactive/2015/11/21/opinion/sunday/exposures-uncertain-journeys.html>. Tipo de pieza: FR (14 fotos)

<http://www.telegraph.co.uk/news/picturegalleries/worldnews/9139137/Syrian-refugees-and-rebels-pictures-by-Telegraph-photographer-David-Rose.html>. Tipo de pieza: G (10 fotos)

[http://elpais.com/elpais/2015/08/26/album/1440594076\\_266806.html#1440594076\\_266806\\_1440594385](http://elpais.com/elpais/2015/08/26/album/1440594076_266806.html#1440594076_266806_1440594385). Tipo de pieza: G (13 fotos)

<http://fotografia.folha.uol.com.br/galerias/42604-sirios-cruzam-rio-entre-grecia-e-macedonia>. Tipo de pieza: G (7 fotos)

<https://www.theguardian.com/world/gallery/2016/feb/26/fabio-bucciarelli-europe-refugee-crisis-syria-libya-afghanistan-pictures>. Tipo de pieza: FR (8 fotos)

<http://www.bild.de/politik/inland/fluechtling/damals-und-heute-blickt-auf-diese-bilder-heidenau-42336300.bild.html>. Tipo de pieza: FR (6 fotos)

[http://www.wienerzeitung.at/multimedia/fotostrecken/dossiers/784591\\_Transitlager-Slavonski-Brod.html](http://www.wienerzeitung.at/multimedia/fotostrecken/dossiers/784591_Transitlager-Slavonski-Brod.html). Tipo de pieza: G (15 fotos)

### 7.2. Revistas

<http://time.com/4047547/europe-refugees-migrants-yuri-kozyrev/>. Tipo de pieza: G (26 fotos)

<http://l-instant.parismatch.com/Jour-apres-jour/A-la-recherche-de-la-terre-promise-749916>. Tipo de pieza: G (17 fotos)

<http://www.spiegel.de/fotostrecke/fluechtlinge-aus-syrien-ander-grenze-zur-tuerkei-fotos-fotostrecke-134386.html>. Tipo de pieza: G (8 fotos)

<http://www.stern.de/politik/ausland/syrische-fluechtlinge-erleben-dramatische-szenen-in-mazedonien-6414618.html>. Tipo de pieza: G (10 fotos)

### 7.3. Sitios de noticias

<http://www.msnbc.com/msnbc/sixteen-days-the-life-three-syrian-refugees#slide1>. Tipo de pieza: G (26 fotos)

## 8. Apéndice II: Ficha de codificación

### 8.1. Datos biográficos

Variables	Valores
Medio	
Fecha de obtención	
Lugar de obtención	
Episodio	Se anotó cada episodio y se recodificó posteriormente
Tipo de Escenario	Se anotó cada escenario y se recodificó posteriormente
Elementos escenográficos	1. Vallas 2. Alambradas 3. Basura 4. Enseres refugiados 5. Juguetes 6. Maletas 7. Armas 8. Material sanitario 9. Tiendas de campaña 10. Banderas

	11. Símbolos religiosos 12. Pancartas
Momento del día	1. Mañana 2. Tarde 3. Noche 4. Indeterminado
Tipo de actor principal	1. Individual 2. Colectivo
Rol Actores principales (Se registra individualmente por cada actor principal)	1. Refugiados 2. Cuerpos de seguridad 3. Cooperantes 4. Periodistas 5. Habitantes locales
Edad Actores principales (Se registra individualmente por cada actor principal)	1. Niños 2. Adolescentes 3. Jóvenes 4. Maduros 5. Ancianos 6. Varios grupos de edad
Sexo Actores principales (Se registra individualmente por cada actor principal)	1. Sólo hombre 2. Sólo mujer 3. Ambos

### 8.2. Codificación fotográfica

Encuadre	1. General largo 2. General 3. Grupal 4. Entero 5. Medio 6. Corto 7. Detalle
Ángulo	1. Neutro 2. Teleobjetivo 3. Gran angular 4. Ojo de pez
Luminosidad general de la imagen	1. Oscuridad 2. Brillo 3. Opaca (gris) 4. Neutra 5. Otros
Composición: equilibrio predominante	1. Formal (Simetría) 2. Informal (Asimetría)
Objetos interpuestos	1. Sí (identificarlos) 2. No
Estatismo/dinamismo	1. Estatismo (posado) 3. Estatismo físico 4. Acción moderada 5. Acción rápida
Expresiones gestuales Cada expresión se ha codificado como presencia (1) o ausencia (0). Se registra individualmente por cada actor principal)	1. Naturalidad 2. Alegría 3. Tristeza, pena 4. Preocupación, incertidumbre 5. Desesperación 6. Cansancio 7. Miedo 8. Respeto 9. Reivindicación 10. Ira, agresividad 11. Indiferencia 12. Aburrimiento 13. Decepción 14. Apelación 15. Firmeza, mando, superioridad 16. Desconfianza 17. Esfuerzo 18. Amenaza 19. Afecto, ayuda 20. Indeterminado

Enviado: 2016-10-11. Segunda versión: 2016-10-12.  
Aceptado: 2016-10-12.



---

# Memória e representação: reflexões para a organização do conhecimento

*Memoria y representación: reflexiones para la organización del conocimiento*

*Memory and representation: reflections for Knowledge Organization*

---

**Mona Cleide Quirino da Silva FARIAS, Maria Leandra BIZELLO**

Departamento de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Av. Higino Muzzi Filho, 737, Marília, São Paulo, Brasil, CEP:17525-900; monaquirino@gmail.com; mleandra23@gmail.com

## Resumen

Se discute la concepción de memoria basada en una perspectiva sociocultural en pos de un posible análisis de ésta como elemento de representación del conocimiento en el ámbito de la Organización del Conocimiento. La memoria estudiada desde una perspectiva sociocultural presenta implicaciones para las prácticas y vivencias de los individuos y grupos que comparten experiencias y conllevan significados comunes; y se vuelve de esa forma una representación del conocimiento con características colectivas e híbridas. Nuestro objetivo es retomar algunas concepciones de la memoria para realizar un posterior análisis de la cuestión de la representación del conocimiento. Se presupone que las formas de organizar y representar el conocimiento acarrear implicaciones para los elementos simbólicos cuyo propósito es la representación de la memoria, es decir, una mediación semiótica evidenciada en la medida en que los sujetos comparten un repertorio de signos. Desde este punto de vista, consideramos que la memoria tanto en su dimensión individual como colectiva es un elemento semiótico del conocimiento. Así, la concepción de memoria no se puede reducir apenas a la idea de recuerdo y olvido, ya que el objeto de la memoria se compone de las prácticas individuales y colectivas que se transforman en el camino dinámico en el que los grupos buscan representar su memoria cultural mediada por un contexto histórico y sociocultural dado.

**Palabras clave:** Memoria. Organización del conocimiento. Mediación semiótica. Memoria cultural.

## 1. Introdução

A memória é aqui compreendida como um elemento de representação do conhecimento. Com base nessa perspectiva entendemos que a memória pode ser também um meio de comunicação, em que grupos buscam compartilhar seus conhecimentos no intuito de manter o significado da memória e de sua representação para posteridade. Nesse sentido, a memória possui um caráter da coletividade, e como memória coletiva esta contribui segundo Pollak (1989), para que seja mantida a coesão de grupos e instituições, e ainda para definir seu lugar.

## Abstract

This paper discusses the concept of memory based on a sociocultural perspective for a possible analysis of it as an element of representation in Knowledge Organization. Memory, from a sociocultural perspective, involves the practices and experiences of individuals and groups that share experiences and carry common meanings and become that, the representation of knowledge with collective and hybrid characteristics. We aim to discuss some conceptions of memory for a further analysis of the issue of knowledge representation. We assume that the ways of organizing and representing knowledge imply symbolic elements whose purpose is the representation of memory, that is, a semiotic mediation evidenced in the way that individuals share a repertoire of signs. Within this scope, we consider memory at both its individual and collective dimensions, as a semiotic element of knowledge. Thus, the conception of memory can not be reduced just to the idea of remembering and forgetting, since the object of memory is the individual and collective practices that turn into the dynamic way in which groups seek to represent their cultural memory mediated by a given historical and sociocultural context.

**Keywords:** Memory. Knowledge organization. Semiotic mediation. Cultural memory.

Destacamos que a memória seja ela definida enquanto social, individual(1), coletiva e cultural, é elemento sócio de representação que nos permite analisar a dinâmica sociocultural de uma comunidade, pois a partir da memória podemos identificar o comportamento de um grupo e até mesmo conhecer sua identidade. Vale destacar que, a identidade só poderá ser construída no interior de contextos sociais e, por sua vez, determinam a posição dos sujeitos e os orientam nas suas representações e escolhas (Cuche,2002). Vale destacar que, ainda que as identidades sejam construídas no interior dos contextos sociais, elas tomam como referência o que lhe é externo

a esse espaço social. Nesse sentido, as identidades são formadas por consequência das experiências múltiplas dos sujeitos em interação com demais contextos.

Desse modo, seria interessante compreender sobre códigos culturais, como estruturas complexas que reconhece, armazenam e processam informações com o objetivo de regular e controlar as manifestações da vida, do bio, do socius, do semeion. Constituem um vocabulário mínimo da cultura que se encontra sempre em movimento (Machado, 2003, p. 156), embora muitas vezes não conseguimos interpretá-los em sua plenitude. Nesse sentido, as linguagens necessitam de atenção, pois além de um sistema de comunicação, constituem um sistema modelizante. Enquanto um sistema modelizante e de comunicação, a linguagem/linguagens é(são) responsável(eis) pela transmissão de mensagens representadas na esfera da cultura —Lotman (1978) apud Machado (2007, p. 29)—.

A memória é a manifestação cultural e social do conhecimento construído em uma comunidade que representa um conjunto de saberes e práticas na perspectiva de comunicar ações, atitudes, sentimentos, os quais podem ser rememorados, muitas vezes, por meio de determinados suportes a partir de um contexto escolhido e/ou determinado pelos sujeitos, atores da memória. Vale destacar que quando nos referimos ao termo manifestação, este é compreendido como algo que se apresenta, se revela, se manifesta. Desse modo, entendemos que a memória implica em manifestação de uma vivência, de hábitos, costumes, de fatos, e por isso de uma cultura e sociedade.

Pressupomos que a memória é também motivada por uma variedade de textos, manifestada por múltiplas linguagens que ultrapassam uma representação material, pois a mesma pode ser comunicada por meios visuais, sonoros e arquitetônicos. Os textos são entendidos como um mecanismo semiótico gerador de sentidos (Machado, 2003, p. 169). Para Lotman (1990, p. 8) o texto além de atuar como a matriz de novos significados é também condensador de memória cultural. Para tanto, trazemos também a memória vista na perspectiva da Semiótica da Cultura(2), como algo que é assegurada pela presença de alguns textos e pela unidade de códigos, ou por sua invariância, ou ainda pelo caráter contínuo e regular de sua transformação (Lotman, 1996).

Nesse ínterim, consideramos também que, para comunicar conhecimentos, repassar tradições, hábitos, crenças e desse modo manter o conhecimento do grupo ou comunidade ativo, a memória deveria ser interpretada como um mecanismo

representativo de um povo, uma geração, ou seja, que está em processo de transformação e implica assim em diálogos e trocas simbólicas.

Se entendermos a memória como referência de uma cultura, que para Coelho Netto (2012, p. 114), a cultura, em sua conceituação mais ampla, remete à ideia de uma forma que caracteriza o modo de vida de uma comunidade em seu aspecto global, ou seja, totalizante. Assim, seria possível reconhecer que, assim como a cultura possui traços híbridos, também se faz a memória e, dentro de uma perspectiva semiótica, pode motivar nos grupos e comunidades processos de semiose, pois é um processo de veiculação dos signos e/ou a própria evolução dos signos. Isso quer dizer que a semiose seria a constante no percurso contínuo percorrida pelo signo, isto é de evolução de significados.

Parafraseando Braga (2000), a memória humana é concebida como um processo elaborado no movimento coletivo que emerge nas interações, e está constituído na cultura. Esse processo pode ser entendido do ponto de vista dos signos(3) dados em uma dimensão simbólica (orais e escritos), como também nos signos icônicos (imagens), os quais podem servir de suporte para a construção da memória. Nesse sentido, a memória pode ancorar-se em diversos suportes e/ou linguagens como, por exemplo, em um texto, na comunicação oral, por meio de sons, de imagens, entre outros (Gondar, 2005).

Para tanto, entende-se que a memória não é algo estagnado finalizado, pelo contrário, a memória é dinâmica e essa característica deve ser reconhecida principalmente quando buscamos representar algo, pois representar não condiz em ditar o que uma coisa é ou foi, mas permitir reconhecer tal coisa em sua dinâmica, mas não desconsiderar as peculiaridades signicas da memória.

Nesse sentido, é necessário ter a consciência de possíveis transformações desde um ponto de vista das linguagens adotadas e assimiladas por culturas e sociedades distintas, como do sentido adquirido ao longo de seu percurso representativo e de agregação de sentidos. Daí surge a questão: será que a memória materializada, isto é, enquadrada(4) em um dado suporte informacional, institucionalizada condiz de fato com a vivência, a história de um povo, ou apenas representa um recorte de um acontecimento e /ou invenção?

Tal questão norteia esse artigo e nos possibilita refletir sobre formas de representação da memória e ainda de como a memória de um determinado grupo pode ser representada tendo em vista que a mesma pode sofrer influências externas e por isso requer instrumentos apropriados

que consigam garantir (Beghtol, 2002) uma representação que mais se aproxime da realidade e que acima de tudo reconheça a hibridiz de significados que perpassam uma determinada memória.

Vale ressaltar que já existem estudos(5) que discorrem sobre o conceito de memória na Ciência da Informação, mas este artigo não possui esse intuito. Nossa intenção é refletir sobre a concepção de memória e apresentá-la sob a perspectiva de mecanismo gerador de conhecimentos, isto é, como elemento de organização que pode desencadear em discussões posteriores sobre como a memória de uma comunidade específica pode ser representada em um sistema de organização do conhecimento desde que leve em consideração os processos dialógicos e semióticos da representação e principalmente as peculiaridades simbólicas da cultura e da comunidade que gera tal memória.

## 2. A memória como objeto do conhecimento

Objetiva-se discutir brevemente sobre a memória como um objeto do conhecimento na perspectiva de reconhecer que as comunidades ou grupos com base em suas culturas, costumes e hábitos desenvolvem memória e esta pode ser objeto de conhecimento na medida em que alcança uma representação, seja por meio de um suporte físico/material, imaterial ou por outros meios de expressão e representação da memória em uma dada cultura, como por exemplo, a dança – em suas várias singularidades artísticas, a prática da oralidade em algumas sociedades etc.

Vale dizer que a memória é construída numa dimensão coletiva mas que desfruta de individualidades, e por isso está sujeita a transformações. Assim também entendemos o conhecimento, algo em plena construção. Ademais, a memória nos remete sempre ao passado, mas um passado comentado, ou representado em um dado presente circunstanciado. Automaticamente o passado rememorado nos fornece lembranças e estas são parte de um conhecimento compartilhado a partir de uma memória. O passado rememorado a partir de lembranças coletivas ou individuais, são parte de um conhecimento compartilhado, uma memória ou memórias.

Contudo, Halbwachs (2004, p. 26-34) destaca que:

[...] nossas lembranças permanecem coletivas, e elas nos são lembradas pelos outros, mesmo que se trate de acontecimentos nos quais só nós estivemos envolvidos, e com objetos que só nós vimos. É porque, em realidade, nunca estamos sós. Não é necessário que outros homens estejam lá, que se

distingam materialmente de nós: porque temos sempre conosco e em nós uma quantidade de pessoas que não se confundem[...]. Para que nossa memória se auxilie com a dos outros, não basta que eles nos tragam seus depoimentos: é necessário ainda que ela não tenha cessado de concordar com suas memórias e que haja bastante pontos de contato entre uma e outras para que a lembrança que nos recordam possa ser reconstruída sobre um fundamento comum.

A capacidade de lembrar na concepção de Halbwachs, só é possível se assumimos o ponto de vista de um ou mais grupos e nos situamos em uma ou mais correntes do pensamento coletivo (Halbwachs, 2004, p. 31):

Assim, se encontrarmos mais tarde membros de uma sociedade que se tornou para nós a tal ponto estranha, por mais que nos encontremos no meio deles, não conseguimos reconstituir com eles o grupo antigo. É como se abordássemos um caminho que percorremos outrora, mas de viés, como se o encarássemos de um ponto de onde nunca o vimos.

Para explorar nossa capacidade de lembrar, a linguagem é fundamental, pois pode ampliar nosso repertório de lembranças e auxiliar na representação da memória. Assim entendemos que a linguagem para a memória é um elemento indissociável da representação. Santos (2003, p. 46) considera que “[...] a linguagem é uma forma de memória que nos antecede. Ela está lá, seja na nossa mente em forma de pensamento, seja organizada sob a forma de discursos ou textos produzidos numa determinada cultura.”

Dessa maneira, visualiza-se de um modo geral que memória possui a potencialidade de ser um objeto do conhecimento. Porém é preciso entender que todo e qualquer conhecimento é desenhado e articulado com base em um domínio, o qual irá determinar as formas de análise e projeção de tal conhecimento. Nesse contexto, entendemos os domínios como campos de atividades que compartilham de conhecimentos que se aproximam, tendo em vista interesses comuns como, por exemplo, a Semiótica da Cultura e os campos de investigações da Antropologia, Linguística estrutural, entre outras.

No entanto, o conhecimento implica em operações organizadas e dialógicas em constante processamento, se trata de algo construído não somente pelo indivíduo, mas pelas relações com o coletivo. Sendo assim, entendemos que o caráter dialógico do conhecimento requer a compreensão do contexto no qual foi criado, isto é, necessita conhecer a memória pela qual tal conhecimento foi elaborado.

Com base nisso, compartilhamos da perspectiva do “conhecimento em ação” como destacado por

Dahlberg (1993), e acrescentamos também que “a cultura, assim como a linguagem e possivelmente a memória são elementos em ação” e por conseguinte estão imbuídos na configuração do/de conhecimento(s).

Desse modo, entendemos que o conhecimento pode ser descrito como a soma de experiências, organizadas e sistematizadas, e por isso não é algo acabado. Assim, devemos considerar que em cada contexto, por ocasião das relações estabelecidas entre os sujeitos, numa dada cultura o conhecimento possui características distintas e está se renovando. A memória nesse contexto é elemento determinante para o conhecimento.

No âmbito do campo da Organização do Conhecimento resgatamos de maneira breve algumas concepções sobre conhecimento que nos auxiliam em entender este como potencializado por uma memória cultural. Barité (2001, p. 42, tradução nossa) destaca que o conhecimento implica em um produto social e uma necessidade social, o autor ressalta ainda que

[...], em suma, pode ser visto como o saber acumulado da humanidade, mas também fornecedor permanente que atenda às necessidades sociais e identifica “lacunas” onde novos conhecimentos são necessários para se regenerar.

Fernandez-Molina (1994, p. 328), destaca que o termo conhecimento constitui-se na conversão de informação, em que esta última só pode tornar-se conhecimento na medida em que venha a provocar modificação na estrutura do conhecimento de um determinado receptor. Podemos dizer que o conhecimento implica no processo de alteração sobre um dado conhecimento anteriormente apresentado por um sujeito em um determinado contexto.

O conhecimento é uma construção coletiva e, portanto, reflexo de uma determinada cultura resguardada e revelada em uma memória. Os conhecimentos de uma cultura encontram-se em constante relação de busca por sentido. Ainda, enquanto produto cultural, os conhecimentos são resultantes de inúmeras condições socioculturais que interagem em tempo e espaço da cultura em sua dimensão híbrida e dialógica. Entende-se que “[...] é na interação [...] onde residem o sentido, a cultura, o conhecimento e a memória” (García Gutiérrez, 2008, p. 42).

Destaca-se que a memória consiste também na tecitura de uma cultura a qual explicita aspectos individuais e sociais. Pois, “mesmo no nível individual, o trabalho da memória é indissociável da organização social da vida” (Pollak, 1989, p. 11). Nora (1993, p. 9) considera a memória como um

processo vivido, conduzido por grupos vivos, portanto, em evolução permanente e suscetível a todas as manipulações. Ainda ressalta que:

A memória é vida, sempre carregada por grupos vivos e, nesse sentido, ela está em permanente evolução, aberta à dialética da lembrança e do esquecimento, inconsciente de suas deformações sucessivas, vulnerável a todos os usos e manipulações, susceptível de longas latências e de repentinas reavitalizações.

A memória está atrelada a uma cultura. E a cultura, enquanto universo simbólico e representativo do conhecimento, possui a capacidade de criar e desenvolver linguagens o que favorece a dinâmica da mesma e suas possíveis relações interculturais. Diante desse universo simbólico mergulhada em diversos signos é que a memória passa a ser projetada e posteriormente representada nos ditos “lugares de memória”.

Para Nora (1993, p. 21) esses lugares são identificados por meio de uma acepção triádica, isto é, como “lugares materiais, lugares funcionais e lugares simbólicos”. O autor entende que tais lugares são carregados de necessidade de memória. Contudo, ressaltamos que é importante a reflexão sobre os lugares de memória e a representação como criações necessárias para a existência e/ou consciência de uma memória.

Nessa perspectiva, de lugares de memória, entendemos que estes representam mais que meros monumentos, espaços institucionalizados, são também documentos e portanto requerem atividades interpretativas que possibilitem entender a procedência da memória veiculada. Para tanto, é necessário realizar leituras sobre sua representação assim como das intenções de quem representou tal memória.

Numa breve sistematização, Pollak (1992, p. 200-212) observa que a memória, seja ela individual ou coletiva, tem como elementos constitutivos acontecimentos, pessoas e lugares. Na perspectiva de acontecimentos tem-se que as experiências são as vividas pessoalmente e ainda aquelas que, não vivenciadas particularmente por nós, foram pelo grupo a que se pertence. Aqui entraria a noção de uma memória “herdada”.

Sendo composta por pessoas, a memória agrega personagens, seja do nosso tempo ou não, sobre as quais se sabem coisas como se as conhecêssemos. Numa compreensão dos lugares podemos ter como exemplo – uma escola espaço de uma infância, da qual guardamos registros afetivos, assim como monumentos, documentos, o espaço da casa – ambiente familiar onde compartilhamos saberes e afetividades etc.

De todo modo, Halbwachs (2004, p. 12) comenta que para que nossa memória se beneficie da dos outros, não basta apenas seus testemunhos, é preciso também que haja um acordo com suas memórias e que seja suficiente os pontos de contato entre ela e as outras memórias, para que a lembrança que os outros nos fornecem possam ser reconstruídas sobre uma base comum.

Em linhas gerais, a memória como uma manifestação de um grupo que deseja destacar um conhecimento dentro de uma relação entre passado e presente pode ser compreendida como um dispositivo do conhecimento, pois por meio da memória, no uso de linguagens e dentro de uma dinâmica cultural, é possível desmembrar conteúdos, informações e (re)construir conhecimentos, assim como representá-los.

### 3. A memória e suas múltiplas representações

Nos encontramos imersos em contextos dinâmicos de representação, enxertados por diversos mecanismos, linguagens e suportes que buscam representar um conhecimento, ou mesmo uma memória. Nesse sentido, como compreender a memória dentro desses contextos híbridos de representação?

Um exemplo que podemos mencionar é o ciberespaço, como uma nova forma/função de produção, representação, organização do conhecimento e memória (Monteiro, 2002; 2006). Nessa perspectiva a memória virtual representada no ciberespaço possui, de algum modo, algumas aproximações com a memória da sociedade oral, como por exemplo, no uso do discurso narrativo, no uso das imagens, e as possibilidades de esquecimento.

No entanto, este cenário do ciberespaço está se configurando como um novo tipo de memória que se distancia da característica mnemotécnica, como também não se limita na memória representada a partir da materialidade dos livros, pelo contrário, este novo cenário evidencia a dinâmica do signo em favor do conhecimento em um fluxo contínuo pelo qual a memória também está sujeita.

Aqui tentamos considerar que a memória pode estabelecer diálogos por meio de múltiplas representações em vários ambientes e linguagens, pois a representação se situa entre o vivido e o concebido, atuando como mediadora de um conhecimento que se encontra direta ou indiretamente atrelado a uma determinada memória.

Desse modo, é possível questionar se representar a memória dentro de uma dimensão coletiva e/ou individual requer apenas o uso de uma única

linguagem e ainda se tal linguagem representa de modo coerente a memória de um grupo/comunidade.

Afirmamos que a representação se origina em um sujeito individual ou coletivo imerso em condições específicas de seu espaço e tempo, e refere-se a um objeto. Ainda ao se referir sobre representações tem-se que tais representações atuam como “[...] uma forma de conhecimento socialmente elaborado e compartilhado, com um objetivo prático, e que contribui para a construção de uma realidade comum a um conjunto social” (Jodelet, 2002, p. 22).

No âmbito da Organização do Conhecimento Brascher e Café (2008, p. 6) destacam sobre o ponto de vista da representação do conhecimento que trata de temáticas que nos encaminham ao “mundo dos conceitos” e não meramente ao dos registros de informação. Ao partir dessa compreensão a representação do conhecimento busca refletir uma visão consensual sobre a realidade a ser representada.

Em outras palavras, a representação é um processo reflexivo sobre os contextos em que estão inseridos diversos domínios de conhecimento. Podemos dizer que, as representações em um dado domínio necessita também o reconhecimento da memória mediada nos mesmos e que devem ser levado em conta os aspectos sociais, culturais e também éticos da representação.

Guimarães e Milani (2011, p. 1) consideram a representação do conhecimento como sendo uma tomada de decisão e atenta para que seus produtos devam figurar-se defensáveis. Ainda no contexto da discussão os autores reforçam sobre os problemas éticos das *bias* no universo da representação do conhecimento. Segundo Milani et al (2009) o termo *bias* possui o sentido de desvios em inglês, e *sesgos* em espanhol. Tal termo está presente em discussões e trabalhos de caráter teórico, em sua maioria, que envolvem a representação do conhecimento.

No escopo dessa discussão é necessário ressaltar que “[...] as representações são apenas uma parte: aquela que se cristalizou e se legitimou em uma coletividade”. Ainda nesse raciocínio, a memória como um processo, “[...] é bem mais que um conjunto de representações; ela se exerce também em uma esfera irrepresentável: modos de sentir, modos de querer, pequenos gestos, práticas de si, ações políticas inovadoras” (Gondar e Dodebei, 2005, p. 24).

Contudo, em se tratando não apenas do conhecimento mais também da memória esta não pode ser reduzida à representação. Pois não “Pode-

mos articular o afeto e a representação na produção da memória como partes integrantes de um mesmo processo.” (Gondar e Dodebei, 2005, p. 25). No entanto, uma representação social, por exemplo, “[...] é algo mais que uma idéia genérica e instituída que se impõe a nós: todas as representações são inventadas e somos nós que as inventamos, valendo-nos de uma novidade que nos afeta e de nossa aposta em caminhos possíveis” (Gondar e Dodebei, 2005, p. 25).

Vale destacar que (Oliveira e Rodrigues, 2011, p. 313):

A sociedade do efêmero, na qual se defrontam questões de ordem ideológica, ética, econômica e tecnológica, parece reforçar a necessidade de tornar explícitas as concepções de memória que permeiam o conhecimento produzido pela Ciência da Informação. Isto talvez ajude a entender o papel que a área pretende, ou deveria pretender desempenhar no processo social de construção da memória.

No mais, a memória é uma construção social, produzida pelos homens a partir de suas relações, de seus valores e de suas experiências vividas. Ela sofre transformações à medida que o tempo passa, a história dos indivíduos toma um novo rumo. Assim, pode-se dizer que a memória não se restringe apenas um registro dos fatos, mas uma combinação de construções sociais do passado, que se agregam a fatores significantes da vida social de um presente, sendo permanentemente reconstruída.

Nesse sentido (Halbwachs, 2004, p. 35):

Não é suficiente reconstituir peça por peça a imagem de um acontecimento do passado para se obter uma lembrança; é necessário que esta reconstrução se opere a partir de dados ou de noções comuns que se encontram tanto no nosso espírito como no dos outros, porque elas passam incessantemente desses para aquele e reciprocamente, o que só é possível se fizeram e continuam a fazer parte de uma mesma sociedade.

Pollak (1989) enfatiza que a referência ao passado serve para manter a coesão dos grupos e das instituições que compõem uma sociedade, para definir seu lugar respectivo, sua complementaridade, mas também as posições irredutíveis.

A recordação passa a ser afetada por transformações inconscientes, em função de interesses e sentimentos, individuais ou coletivos, todavia, “[...] esse movimento só é possível se as pessoas fizeram ou ainda fazem parte de um mesmo grupo social” (Braga, 2000, p. 51).

Nessa perspectiva, a memória coletiva só se efetiva na medida em que os sentimentos, pensamentos e ações de cada indivíduo são expressos

nos meios e circunstâncias sociais, onde este possui vínculo, convivência e conhecimento, isto é, onde compartilhe de uma cultura ou ainda de um repertório de signos que marcam a sua identidade a partir de linguagens mediadas.

Consideramos que as linguagens são, dessa maneira, as formas dinâmicas de expressão da cultura, e também da memória, visto que abarcam a diversidade cultural existente nos grupos e comunidades. Nesse sentido, as culturas não devem ser entendidas como estáticas, pois “Entre as culturas se produz uma interação simbólica” (García Gutiérrez, 2004, p. 113, tradução nossa).

Em suma, na ocasião da representação da memória, retomamos brevemente a concepção de exomemória de García Gutiérrez (2002) onde esta constitui nas diversas formas de registrar o conhecimento, como por exemplo, os livros, as fotografias, as gravuras, os quadros, as esculturas, etc. Sobre tal discussão, vale destacar a concepção do autor sobre conhecimento e memória, em que o mesmo trata tanto o conhecimento quanto a memória como sinônimos.

#### 4. Considerações

Entende-se que a discussão da memória é um tema aberto e por isso nos possibilita ampliar nossas concepções. Dessa maneira, reconhecemos que as formas de organizar e representar o conhecimento implicam em elementos simbólicos cujo propósito seja a representação da memória, a qual implica em mediação semiótica evidenciada na medida em que os sujeitos compartilham de um repertório de signos.

Neste escopo, consideramos que a memória tanto em sua dimensão individual e/ou coletiva, pode ser configurada como elemento semiótico do conhecimento e, por conseguinte, da representação. Em alguns aspectos a memória pode vir a ser compreendida como o próprio conhecimento no sentido de compartilhar saberes mediante as práticas culturais de um grupo ou comunidade.

Memória, no ponto de vista dessa discussão, se configura como um dispositivo semiótico articulado e transformado pelos indivíduos nas relações com o coletivo e que não está alheia à um passado que em dimensões múltiplas dialoga com um presente para uma suposta representação. A representação nesse contexto seria a forma de comunicação de uma memória transposta em um determinado suporte materializado ou não, mas que garanta seu significado.

Portanto, a concepção de memória não pode se reduzir apenas a ideia de lembrança e esquecimento, nem mesmo ser reconhecida a partir de



um único modo de representação, pois o objeto da memória constitui das práticas individuais e coletivas que transformam-se no percurso dinâmico em que os grupos buscam representar sua memória cultural condicionada a um dado contexto histórico e sociocultural.

## Notas

- (1) Para Halbwachs (1990), a memória individual não está inteiramente fechada. Ela é limitada no tempo e no espaço. A memória coletiva também sofre os mesmos limites. Os acontecimentos históricos são auxiliares na nossa memória; não desempenham outro papel, senão as divisões do tempo assinaladas em relógio ou determinadas pelo calendário.
- (2) Também conhecida por Semiótica Russa. Suas primeiras perspectivas de estudos tiveram como palco a Escola de Tártu-Moscou, na Estônia, que em meados da década de 1960 desenvolveu as primeiras discussões sobre sistemas de signos: texto cultural, códigos culturais, linguagens etc. Seu campo possui caráter transdisciplinar e dialoga com diversas disciplinas como a Teoria literária, Linguística estrutural, Semiótica, Crítica da arte, Cibernética, Teoria da Informação e da Comunicação, Antropologia, Etnologia etc.
- (3) Peirce (1990, p. 4), diz que o signo é qualquer coisa que conduz alguma outra coisa (seu interpretante). Para Coelho Netto (2010, p. 56), a concepção de signo dada por Peirce é aquilo que sob determinado aspecto representa algo para alguém. Na concepção atribuída por Saussure (1970, p. 81) o signo linguístico constitui de uma unidade psíquica de duas faces que seria a união entre a imagem acústica e o conceito, ou significado e significante.
- (4) O termo exposto, ressalta a noção de “enquadramento de memória” que segundo Michel Pollak (1989), além de servir como um referencial do passado evidencia a disputa em torno da memória e revela um modo de manter a coesão entre os grupos sociais.
- (5) Ver a tese de Oliveira (2010) cujo título é: “O conceito de memória na ciência da informação no Brasil: uma análise da produção científica dos programas de pós-graduação”.

## Referências

- Barité, Mário (2001). Organización del conocimiento: un nuevo marco teórico-conceptual em Bibliotecología y Documentación. // Carrara, Kester. (ed.). Educación, universidad e pesquisa. Marília: Unesp-Marília-Publicações. 2001. 35-60.
- Beghtol, Clare (2002). Universal concepts, cultural warrant, and cultural hospitality. // López-huertas, M. J. (ed.). Challenges in knowledge representation and organization for the 21 century: integration of knowledgeacrouss boundaries. Wurzburg: Ergon-Verlag, 2002. 45-49. (Advances in Knowledge Organization, 8).
- Braga, Elizabeth dos Santos (2000). A Construção Social da Memória: uma perspectiva histórico-cultural. Ijuí: Unijuí, 2000.
- Brascher, Marisa; Café, Lígia (2008). Organização da informação ou organização do conhecimento? // Encontro Nacional de Pesquisa em Ciência da Informação – Enancib, 9. São Paulo: ECA/USP, 2008. <http://www.an-cib.org.br/media/dissertacao/1835.pdf> (2013-05-18).
- Coelho Netto, José Teixeira (2010). Semiótica, Informação e Comunicação. São Paulo: Perspectiva, 2010.
- Coelho Netto, José Teixeira (2012). Dicionário crítico de política cultural: cultura e imaginário. 2.ed. rev.ampl. São Paulo: Iluminuras, 2012.
- Cuche, Denys (2002). A noção de cultura nas ciências sociais. 2. ed. Bauru: EDUSC, 2002.
- Dahlberg, Ingetraut (1993). Knowledge organization: its scopes and possibilities. // Knowledge Organizaton. Frankfurt. 20:4, 211-222.
- Fernandez-molina, Juan Carlos (1994). Enfoques objetivo y subjetivo del concepto de información. // Revista Española de Documentación Científica. 17:3. 320-331.
- García Gutiérrez, Antonio (2002). Exomemoria y cultura de frontera: hacia una ética transcultural de la mediación. // VII Congreso Internacional sobre Organización del Conocimiento celebrado en Granada (España).
- García Gutiérrez, Antonio (2004). Otra memoria es posible: estrategias descolonizadoras del archivo mundial. Sevilla: Universidad de Sevilla. 2004.
- García Gutiérrez, Antonio (2008). Outra memória é possível: estratégias descolonizadoras do arquivo mundial. Petrópolis: Vozes. 2008.
- Gondar, Jô; Dodebei, Vera Lucia Doyle (eds.). (2005) O que é Memória Social? Rio de Janeiro: Universidade Federal do Rio de Janeiro. 2005.
- Guimarães, José Augusto Chaves; Milani, Suellen Oliveira (2011). Problemas éticos em representação do conhecimento: uma abordagem teórica. // DataGramaZero. 12:1, 1-18.
- Halbwachs, Maurice (1990). A memória coletiva. tradução de Laurent Leon Schaffer. São Paulo:Vértice, 1990.
- Halbwachs, Maurice (2004). A memória coletiva. São Paulo: Centauro, 2004.
- Jodelet, Denise (2002). Representações Sociais: um domínio em expansão. // Jodelet, Denise. (ed.). As Representações Sociais. Rio de Janeiro: Eduerj, 2002.
- Lopes, Luís Carlos (1998). A imagem e a sombra da arquivística. Rio de Janeiro: Arquivo Público, 1998.
- Lotman, Iuri. M. (1990). Universe of the mind: a semiotic theory of culture. London: Tauris, 1990.
- Lotman, Iuri. M. (1996). La semiosfera I: semiótica de la cultura y del texto. Tradução de Desiderio Navarro. Valência: Frónesis Cátedra, 1996.
- Machado, Irene (2003). Escola de Semiótica: a experiência de Tártu - Moscou para o estudo da cultura. São Paulo: Ateliê Editorial. 2003.
- Machado, Irene (2007) Semiótica da Cultura e Semiosfera. São Paulo: Annablume/Fapesp, 2007.
- Millani, Suellen Oliveira; Guimarães, José Augusto Chaves; Pinho, Fábio de Assis; Fernández-Molina (2009). Os desvios na representação do conhecimento em um contexto multicultural: abordagens teóricas. // Romero, N. L. (ed.). Nuevas perspectivas para la difusión y organización del conocimiento. Valencia: Universidad Politécnica de Valencia. 2009. 204-218.
- Monteiro, Silvana Drumond (2003). A organização virtual do conhecimento no ciberespaço: os agenciamentos do sentido e do significado. Pontifícia Universidade Católica de São Paulo, São Paulo. 2003. Tese de doutorado.
- Monteiro, Silvana Drumond (2006). O ciberespaço e os mecanismos de busca: novas máquinas semióticas. // Ciência da Informação. Brasília. 53:1, 31-38, (jan./abr.) <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=738&layout=abstract> (2006-03-20).
- Nora, Pierre (1993). Entre memória e história: a problemática dos lugares. // Projeto História. São Paulo. 10, 7-28.
- Oliveira, Eliane Braga de; Rodrigues, Georgete Medleg (2011). O conceito de memória na Ciência da Informação:

- análise das teses e dissertações dos programas de pós-graduação no Brasil. // Liinc em Revista. 7:1. Rio de Janeiro. 311-328 <http://www.ibict.br/liinc>
- Oliveira, Eliane Braga de (2010). O conceito de memória na ciência da informação no Brasil: uma análise da produção científica dos programas de pós-graduação. Universidade de Brasília, Brasília, 2010. Tese de doutorado.
- Peirce, Charles. Sanders (1990). Semiótica. 2. ed. São Paulo: Perspectiva. 1990.
- Pollak, Michael (1992). Memória e identidade social. // Estudos Históricos. Rio de Janeiro. 5:10. 1-15.
- Pollak, Michael (1989). Memória, Esquecimento, Silêncio. // Estudos Históricos, tradução de Dora Rocha Flaksman. Rio de Janeiro. 2:3, 3-15.
- Santos, Myrian Sepúlveda dos (2003). História e memória: o caso do ferrugem. // Revista Brasileira de História. São Paulo. 23:46, 271-295.
- Saussure, Ferdinand. de. (1970). Curso de lingüística geral. 2. ed. São Paulo: Cultrix. 1970.

---

Enviado: 2016-05-14. Segunda versão: 2016-06-29.  
Aceptado: 2016-09-15.

---

---

# Vocabulário controlado para arquivos universitários: a experiência do VTArq-UNESP

*Vocabulario controlado para archivos universitarios: la experiencia del VTArq de UNESP (Brasil)*

*An archival controlled vocabulary for university archives: the experience of VTArq (UNESP, Brazil)*

---

**Walter MOREIRA, Mariângela Spotti Lopes FUJITA, Luciana Beatriz Piovezan dos SANTOS**

Universidade Estadual Paulista, Av. Hygino Muzzi Filho, 737, Campus Universitário, CEP. 17525900, Marília, SP, Brasil;  
{ walter.moreira | mariangelaslf57 | lbpiovezan }@gmail.com

## Resumen

Se presenta la construcción de un vocabulario controlado común para a las tareas de procesamiento de información del archivo de la Universidad Estadual Paulista (VTArq-UNESP). Se consultó con las unidades de archivo de la universidad para la identificación y recopilación de expresiones utilizadas con eficacia en la representación de la documentación. Seguidamente, se investigó la adecuación del vocabulario con programas informáticos de gestión terminológica. La experiencia demostró la necesidad de un vocabulario específico para los archivos de la universidad y de su normalización para la organización y recuperación de información.

**Palabras clave:** Tesaurus. Planificación. Construcción. Archivística. Gestión de documentos. Archivos universitarios.

## Abstract

The experience of construction of a controlled vocabulary for processing the university records of the Universidade Estadual Paulista (UNESP VTArq) is presented. First, the different units of the university were consulted for the identification and collection of the expressions effectively used for representing the documentation. Thereafter, the adequacy of the vocabulary was assessed with the help of terminology management software. The experience shows the need of vocabulary specificity and standardization to support the organization and retrieval of information in the field of university archives.

**Keywords:** Thesauri. Planning. Construction. Archival science. Record management. University archives.

## 1. Introdução

Na área da arquivologia, as discussões sobre aspectos teóricos e metodológicos envolvidos na construção e na aplicação de instrumentos de controle vocabular, denominados genericamente pela expressão “vocabulários controlados”, ainda não são realizadas com a frequência e o aprofundamento com que são encontrados na biblioteconomia e na documentação, disciplinas citadas para fins de comparação (Davanzo; Moreira, 2015; Moreira; Davanzo; Moraes, 2015). Mesmo a terminologia arquivística referente a esses instrumentos de organização e representação do conhecimento ainda é incipiente.

Dentre algumas experiências bem-sucedidas, pode-se destacar o trabalho do *Grupo de Archiveros Municipales de Madrid*, em atividade desde 1981, formado por profissionais de arquivo e de gestão documental municipal da Comunidade de Madri. Esse Grupo organiza, desde 1982 as *Jornadas de Archivos Municipales*. Na sua edição de número dezoito, realizada em 2010, o Grupo publicou, como resultado de uma comissão especialmente criada para este fim, o *Tesouro de archivos municipales* (Grupo de Archiveros de Madrid, 2010), trabalho iniciado em 1997. Ainda

que tal instrumento seja destinado a documentação distinta da que se ocupa o trabalho que ora se apresenta, a experiência relatada e, principalmente, o detalhamento da metodologia, revestem-se de substancial interesse para o projeto que se apresenta ainda em desenvolvimento.

Em vista do volume de informações e de documentos em circulação e da impossibilidade de gerenciá-los em termos de suas representações documentárias sem o recurso de instrumentos adequados, a construção e a utilização de vocabulários controlados tornou-se necessária em quaisquer unidades de informação, incluindo-se, vale a redundância, as unidades de informação arquivística.

Nesse contexto, a efetiva compreensão da complexidade do fluxo documentário e a busca por soluções eficazes para os problemas de representação e recuperação da informação arquivística atuam como fatores motivadores para o investimento que a Universidade Estadual Paulista (UNESP) vem realizando na construção de um vocabulário controlado que possa ser utilizado amplamente pelas diversas unidades que a compõem bem como pelos demais arquivos universitários com o objetivo de compartilhamento

de experiências teóricas e metodológicas e, sobretudo, no aprimoramento do controle de vocabulário arquivístico.

## 2. Vocabulário controlado como instrumento de organização e representação do conhecimento

A organização do conhecimento preocupa-se com os aspectos teóricos e metodológicos referentes ao modo pelo qual o conhecimento será classificado ou estruturado nas diversas áreas. Observando-se este seu caráter aplicado, notadamente na ciência da informação, é mister entrelaçar o conceito de organização do conhecimento com os conceitos de representação do conhecimento e, inclusive, com os aspectos que envolvem a recuperação da informação. A rigor, respeitando-se essa faceta referente à aplicação, criam-se sistemas de organização para que, por seu intermédio, sejam possíveis a organização e a recuperação da informação.

García Marco (1993, p. 100, tradução livre) reconhece o caráter científico-aplicado da organização do conhecimento quando informa que seu objetivo é “melhorar a circulação da informação dentro e por meio de sistemas mediadores – os centros de seleção, armazenamento, recuperação e disseminação de informação – para produzir novo conhecimento e/ou facilitar o acesso ao [conhecimento] existente”.

A temática que ocupa este trabalho não diz respeito, contudo, aos problemas diretamente envolvidos na recuperação da informação, apenas os aborda de modo indireto, com base na compreensão sistêmica das inter-relações apresentadas brevemente no parágrafo anterior. Trata-se mais especificamente nesse trabalho da organização e representação do conhecimento, uma disciplina oriunda da confluência da ciência da informação com as ciências cognitivas, a linguística, a terminologia, a ciência da computação e a lógica.

A institucionalização e a legitimação social e profissional da organização e representação do conhecimento ocorrem também por meio do trinômio desenvolvimento, utilização e aprimoramento dos sistemas de organização do conhecimento (SOC, ou KOS, *knowledge organization systems*, na terminologia mais universalizada adotada na língua inglesa). Dentre os SOC mais conhecidos incluem-se os sistemas de classificação, os tesouros, as taxonomias, as ontologias e os *topic maps*. Considerando-se apenas as características que os aproximam, apresentam elementos que podem ser subsumidos na expressão genérica “vocabulários controlados”. Note-se bem que não se está afirmando que o controle vocabular

é a finalidade ou a função primeira de tais instrumentos, mas sim que todos eles manifestam em algum momento, em maior ou menor grau, algum nível de padronização ou de normalização terminológica no processo.

O nível demandado de controle do vocabulário irá variar, naturalmente, conforme a configuração estrutural do SOC e de acordo com o contexto e os objetivos de seu desenvolvimento, bem como pelo caráter de sua aplicação. Exige-se, por exemplo, maior precisão terminológica num tesouro do que num sistema de classificação, do mesmo modo que se espera de uma ontologia um uso do léxico mais acurado do que aquele revelado por uma taxonomia.

Os vocabulários controlados desempenham um importante papel na redução dos eventuais ruídos presentes na comunicação entre os sistemas de informação documentária e seus usuários. Desse modo atuam como instrumentos auxiliares indispensáveis aos processos de organização, representação e recuperação da informação.

Além de aumentar a eficácia do sistema nas ações de representação da informação, um vocabulário controlado eficaz também amplia, em razão diretamente proporcional, sua confiabilidade. A noção de representação, neste caso, refere-se à representação como processo, traduzido no ato que realizam, de modo assíncrono, o profissional indexador e o usuário. Considerando-se que ambos se interessam pela localização e pela utilização efetiva da informação, ou seja, concorrem para o mesmo fim, a linguagem natural torna-se um obstáculo quando se revela seu caráter naturalmente polissêmico. Assim, o instrumento de controle vocabular funciona como instrumento de tradução, mediador entre as expressões que representam conceitos formulados pelos usuários e pelos profissionais indexadores. O vocabulário controlado permite traduzir a linguagem do usuário e a linguagem do sistema para uma linguagem documentária que possa atuar como *lingua franca* nas situações em que se requer sua interlocução.

A partir dessa função intermediadora, atribui-se também aos vocabulários controlados uma função de protagonismo social na disponibilização e na apropriação da informação, bem como na promoção do conhecimento como dinamismo social (Barité, 2001). O compartilhamento social da informação é, afinal, devedor dos modelos de organização e representação da informação e do conhecimento.

### 3. A proposta do VTArq-UNESP

A UNESP é uma instituição de ensino superior relativamente jovem, completando 40 anos em 2016, que visa a oferecer ensino superior público, gratuito e de qualidade. Está geograficamente localizada de forma descentralizada no estado de São Paulo, no Brasil: são 34 unidades distribuídas em 24 cidades.

Seria utopia imaginar que a cultura organizacional que imprime identidade à universidade, a qualquer universidade, e a identificação e descrição de seu fluxo documentário possam atuar como garantia para a coerência na representação dos seus documentos. Essas dificuldades, no caso da UNESP, acentuam-se em vista do caráter de descentralidade, e até mesmo de certa incomunicação provocada, entre outros, pela barreira geográfica a qual cria, muitas vezes, condições de isolamento.

Nesse contexto, visando a garantir a eficácia necessária ao bom fluxo da informação arquivística, tornou-se imperativo investir na construção de um vocabulário controlado comum que pudesse subsidiar todas as unidades em suas tarefas de tratamento da informação arquivística nas diferentes seções de comunicação que compõem as unidades de ensino.

A construção do referido vocabulário controlado foi iniciada em resposta à necessidade detectada de forma mais premente a partir da implantação do Sistema de Gestão Eletrônica de Documentos (SIGAD). Dentre as novidades e desafios que o SIGAD apresentou, estava a utilização de um campo destinado à inserção de palavras-chave. Bem poucas unidades já utilizavam, no sistema antigo, o recurso das palavras-chave, e esse recurso era realizado sem qualquer elemento de controle de vocabulário.

Para a construção do Vocabulário de Termos Arquivísticos da UNESP (VTArq-UNESP) foram consultadas as seções técnicas de comunicação dos diversos *campi*, visando à identificação e compilação das expressões efetivamente utilizadas na representação da documentação que circulava nas seções técnicas de comunicação. Completada essa fase, realizou-se pesquisa terminológica visando à identificação e padronização dos termos candidatos, distinguindo-os entre termos preferidos e termos não preferidos. Os procedimentos metodológicos utilizados na condução da pesquisa terminológica estão descritos no capítulo 3 deste trabalho.

A subcomissão de atualização e manutenção do VTArq-UNESP iniciou seus trabalhos em março

de 2014. Vincula-se hierarquicamente à Comissão de Avaliação de Documentos e Acesso (CADA), da mesma universidade.

O VTArq-UNESP foi construído em função de solicitação do Grupo Gestor do Sistema de Gestão Documental Eletrônica (SIGAD) à CADA.

CADA é a atual denominação da comissão responsável pela avaliação dos documentos de arquivos conforme o Decreto no 58.052, de 16 de maio de 2012, que regulamenta, no estado de São Paulo, a Lei de Acesso à Informação - Lei Federal no 12.527, de 18 de novembro de 2011. O Decreto também criou em todos os órgãos e entidades de administração pública estadual os Serviços de Informação ao Cidadão (SIC), os quais são responsáveis pelo atendimento presencial ou eletrônico ao cidadão e pelo encaminhamento das solicitações aos SICs setoriais.

Na UNESP, a instalação do SIC, em 2012, ocorreu de forma centralizada junto ao Gabinete do Reitor, mas organizada mediante a determinação de que as seções técnicas de comunicação de cada unidade universitária ou unidade experimental, onde se localizam os arquivos correntes, se tornassem unidades locais do SIC central, funcionando como um sistema de arquivos correntes para provimento de qualquer solicitação de informação. Dessa forma, a gestão documental é necessária nos arquivos universitários porque são, no entendimento da CADA, o local de destinação e organização da documentação produzida pelas atividades meio e fins da UNESP.

O SIGAD é um sistema que instrumentaliza e dá visibilidade à gestão documental das seções técnicas de comunicação da UNESP. Para seu completo funcionamento necessita dos instrumentos de organização e representação da informação da área arquivística que são o plano de classificação de atividades meio e o plano de classificação de atividades fins, com tabela de temporalidade compatíveis com o Plano de Classificação e Tabela de Temporalidade do Sistema de Arquivos do Estado de São Paulo (SAESP), o qual coordena o sistema SIC do estado de São Paulo.

O VTArq-UNESP foi elaborado não em função do Decreto Lei Estadual, que não prevê este instrumento de organização e representação da informação, mas por solicitação das seções técnicas de comunicação, que tinham uma prática comum de fazer a determinação de assuntos em todos os denominados processos que organizam os documentos de uma mesma procedência organizacional.

Assim, a origem do VTArq-UNESP está na compatibilização de três listas alfabéticas de “palavras-chave” compiladas por seções técnicas de

comunicação da UNESP. Essa compilação, sem método ou técnica de elaboração de vocabulários ou mesmo de recursos de pesquisa terminológica, foi realizada ao longo do tempo de prática dessas seções, o que determinou, em alguns casos, a autorização desses termos sob garantia de uso

Para a necessária normalização terminológica das listas alfabéticas oriundas das seções de comunicação desempenhou-se procedimentos de controle de vocabulário a partir das diretrizes estabelecidas pela UNESCO (1993) e com o uso de um modelo de ficha terminológica adaptado aos interesses e particularidades do projeto. A adaptação da ficha terminológica (Quadro I) deu-se com base nos modelos propostos por Cervantes (2004), Pavel e Nolet (2002) e Barros (2004).

<i>Termo:</i>	
Categoria gramatical:	Sigla/Acrônimo:
Unidade UNESP: [Unidade na qual o termo é utilizado, conforme lista fornecida pelas seções técnicas de comunicação]	
Outras designações	
Contexto de ocorrência	
Referência completa	
Conceito institucional	[Observação da ocorrência do termo/conceito em documentos legais da UNESP]
Referência Completa	
Compreensão do servidor	[Informação obtida junto às seções técnicas de comunicação]
Responsável pela coleta:	Data:
Observações:	

*Quadro I. Ficha terminológica de trabalho (adaptado de Cervantes, 2004; Pavel e Nolet, 2002; e Barros, 2004)*

O uso da ficha terminológica teve como objetivo principal, na primeira fase, o controle de sinonímia, de número e gênero e de forma gramatical, além do fornecimento de elementos para a identificação de relações de equivalência entre os termos. Um elemento que se revelou particularmente importante na ficha terminológica utilizada foram os campos denominados “conceito institucional”, referindo-se à representação da linguagem técnica da UNESP, por meio da terminologia empregada na documentação oficial da universidade, e “compreensão do servidor”, referindo-se ao contexto de uso pela comunidade de servidores e docentes da UNESP.

O termo “água”, por exemplo, que aparecia na lista de palavras autorizadas de duas seções técnicas de comunicação e que parece vago por seu amplo grau de generalidade, pôde ser substituído, após a identificação de seu uso efetivo (contexto de uso) pelo termo “conta de água”.

De outro lado, as sutilezas do sentido e das relações semânticas entre termos “admissão” (servidor admitido pelo regime autárquico), “contratação” (servidor admitido pelo regime da Consolidação das Leis do Trabalho – CLT) e “nomeação” (servidor admitido pelo regime efetivo) só puderam ser devidamente captadas com o apoio da legislação específica aplicada na UNESP.

Atualmente, o VTArq-UNESP está em processo final de adequação à terminologia dos Planos de Classificação da UNESP para sua disponibilização final dentro do SIGAD. Nessa fase, também estão sendo realizados os testes de fluxo navegacional do vocabulário no ambiente do *software*. Essa revisão e os testes precedem, necessariamente, o uso, uma vez que sua terminologia precisa estar adequada à produção documental estadual e universitária determinantes da procedência arquivística. Desse modo, a liberação para o uso do vocabulário somente será possível após completadas as fases de revisão, adequação e testes. Posteriormente a essa etapa o corpo técnico responsável pelas seções técnicas de comunicação serão devidamente informados e capacitados para a utilização efetiva do VTArq-UNESP.

Com a finalização do plano de classificação de atividades meio e tabela de temporalidade da UNESP (Troitiño, 2016) foi possível realizar uma adequação com base no vocabulário empregado. Aguarda-se a finalização do plano de classificação de atividades fins para uma segunda adequação de vocabulário.

A elaboração do VTArq-UNESP, tem origem nas listas alfabéticas como primeira etapa que se constituiu da normalização e estruturação lógica e semântica dos termos compilados com uso do *software* TemaTres, versão 1.6. Na sequência, a segunda etapa consistiu da adequação de vocabulário ao Plano de classificação de atividades meio da UNESP.

O TemaTres é uma aplicação *web* utilizada para a gestão de representações linguísticas formais do conhecimento, incluindo vocabulários controlados, taxonomias, tesouros e listas de cabeçalhos de assunto (Gonzales Aguilar, Ramírez Posada, Ferreyra, 2012; Ferreyra, 2016).

Além de ser um *software* reconhecido pela comunidade científica (Scornik; Sosa, 2007; Martínez-

González; Alvite-Díez, 2014;), a escolha do TemaTres se deu também em função de sua gratuidade e por conta dos recursos de promoção de interoperabilidade que a ferramenta oferece, como, por exemplo, a exportação da estrutura do vocabulário controlado em formato *Simple Knowledge Organization System* (SKOS).

O SKOS é uma linguagem formal baseada em *Resource Description Framework* (RDF) que pode ser utilizada para a construção de representações documentárias que incluem os tesouros e os sistemas de classificação (World Wide Web Consortium - W3C, 2012). Trata-se de um esforço de promoção de interoperabilidade baseado na concepção de que os sistemas de organização do conhecimento guardam muitas semelhanças, tanto em relação ao modo como são estruturados quanto às suas aplicações. A identificação dessas similaridades e sua explicitação em linguagem formal compreensível por máquina facilita as ações de interoperabilidade na construção e uso de vocabulários controlados.

#### 4. Observações sobre os resultados do projeto

O VTArq-UNESP possui, após as primeiras revisões, um total de 1.249 termos, sendo 1.087 termos preferidos e 162 termos não preferidos. Por se tratar de um projeto que envolve investigação e execução em médio prazo, os resultados apresentados são parciais. Verificou-se que será necessário, em continuidade, avaliar o uso do VTArq-UNESP no atendimento às necessidades da produção e gestão documental dos arquivos correntes da UNESP no Sistema SIGAD.

A composição atual do VTArq tem a seguinte origem: a) as listas de expressões utilizadas pelas três seções técnicas de comunicação que foram apresentadas aos pesquisadores responsáveis pelo projeto e b) o Plano de Classificação de Atividades Meio da UNESP.

Para composição do conjunto de termos preferidos e não-preferidos, foi realizada a análise terminológica de todos os termos candidatos, isto é, todas as entradas resultantes dos instrumentos citados no parágrafo anterior. Essa decisão foi tomada em observação às garantias necessárias ao funcionamento do vocabulário controlado. Refere-se, especificamente, à garantia de uso (pela incorporação, quando possível, das expressões efetivamente utilizadas pelos profissionais responsáveis nas seções técnicas de comunicação) e à garantia institucional (pelo respeito ao plano de classificação instituído).

Considerando-se as diferentes naturezas e propósitos dos instrumentos, a tarefa de compatibilização terminológica entre o vocabulário controlado e o plano de classificação revelou-se mais complexa do que se poderia inicialmente crer. O recurso utilizado para a compatibilização foi o cotejo direto entre os termos presentes nos dois instrumentos buscando-se identificar sinônimos e quase-sinônimos e verificar quais as entradas elencadas no plano de classificação que poderiam ser utilizadas na construção de novos termos preferidos ou não preferidos no VTArq-UNESP.

Com vistas a organizar hierarquicamente os termos foram utilizadas como proposta de macrocategorização as categorias definidas pelo *Classification Research Group* (CRG), conforme sua apresentação em Frické (2012, p. 214). Em tradução livre, são as seguintes: coisa, tipo, parte (órgão, constituinte), propriedade, material, processo (uma ação interna ao item), operação (uma ação realizada no item), paciente (objeto da ação, matéria-prima), produto (substância), subproduto, agente, espaço, tempo.

Para o desenvolvimento pleno do projeto e considerando-se principalmente a perspectiva de hibridiz nos espaços de informação, há que se observar diretrizes e normas internacionais aplicadas tanto à descrição arquivística quanto à elaboração e manutenção de tesouros (Davanzo, 2016). Neste aspecto, destacam-se, sem prejuízo de outras correlacionadas, as seguintes normas ISAAR(CPF): norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias (Conselho Internacional de Arquivos, 2004); ISO 25964: *information and documentation: thesauri and interoperability with other vocabularies* (International Organization for Standardization, 2011a; 2011b) e as *Guidelines for Authority Records and References* (International Federation of Library Associations and Institutions, 2001).

Em sequência, está prevista a capacitação de todos os responsáveis pelas seções técnicas de comunicação na UNESP para a função e uso do VTArq-UNESP. Nessa oportunidade, será estimulada a criação de um grupo de atualização e manutenção do Vocabulário, que deverá ser permanente e acompanhado pelo grupo responsável pela elaboração.

O Grupo de elaboração será responsável por realizar o monitoramento avaliativo de uso do VTArq-UNESP, com metodologia de avaliação extrínseca e intrínseca aplicadas anualmente além da verificação quantitativa e qualitativa dos termos e sua utilização.

## 5. Considerações finais

Diante do exposto consideramos relevante reafirmar a importância do debate relativo aos instrumentos de controle de vocabulário, seus aspectos teórico-metodológicos e sua aplicação, direcionado às unidades de informação arquivística visando ao desenvolvimento das práticas de organização e recuperação da informação em sua responsabilidade.

No contexto das seções técnicas de comunicação das unidades universitárias da UNESP, com o investimento na construção deste instrumento normalizado de representação da informação espera-se não apenas atender aos dispositivos legais, mas também aperfeiçoar a própria gestão documental ali realizada, vindo a beneficiar a comunidade acadêmica.

## Referências

- Barite, Mario (2001). Organización del conocimiento: un nuevo marco teórico-conceptual em Bibliotecología y Documentación. // Carrara, Kester (Org.). Educação, universidade e pesquisa. Marília: Unesp-Marília, 2001. 35-60.
- Barros, Lidia Almeida (2004). Curso básico de Terminologia. São Paulo: Edusp, 2004.
- Cervantes, Brígida Maria Nogueira (2004). Contribuição para a terminologia do processo de inteligência competitiva: estudo teórico e metodológico. 2004. 209f. Dissertação (Mestrado). Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2004.
- Conselho Internacional de arquivos (2004). ISAAR(CPF): norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias/tradução de Vitor Manoel Marques da Fonseca. Rio de Janeiro: Arquivo Nacional, 2004.
- Davanzo, Luciana; Moreira, Walter (2015). A ISAAR(CPF) e os vocabulários controlados: uma aproximação preliminar. // Zabala Vázquez, J.; Sánchez Jiménez, R.; García Moreno, M.A. (Orgs.). Desafíos y oportunidades de las Ciencias de la Información y la Documentación en la era digital: actas del VII Encuentro Ibérico EDICIC 2015 (Madrid, 16 y 17 de noviembre de 2015). Madrid: Universidad Complutense de Madrid, 2015.
- Ferreira, Diego (2016). TemaTres. <http://www.vocabularyserver.com/index.html> (2016-03-16).
- Frické, Martin (2012). Logic and the organization of information. New York: Springer, 2012.
- García-Marco, Francisco Javier (1993). Paradigmas científicos en representación y recuperación de la información. // García-Marco, Francisco Javier (Ed.). Organización del conocimiento en sistemas de información y documentación: actas del I Encuentro de ISKO-España, Madrid, 4 y 5 nov. 1993. Zaragoza: Universidad de Zaragoza, 1993.
- Gonzales Aguilar, Audilio; Ramírez Posada, María; Ferreyra, Diego (2012). Tematres: software para gestionar tesauros. // El profesional de la información. 21:3 (Mayo-Junio) 319-325.
- Grupo de Archiveros Municipales de Madrid (2010). XVII Jornadas de Archivos Municipales: "Pilares de la e-administración: cuadro de clasificación y tesoro": Tesoro de Archivos Municipales. San Sebastián de los Reys: B.O.C.M., 2010.
- International Federation of Library Associations and Institutions (2001). Guidelines for Authority Records and References. 2.ed. Munchen: K.G. Saur, 2001.
- International Organization for Standardization (2011a). ISO 25964: information and documentation: thesauri and interoperability with other vocabularies - part 1: thesauri for information retrieval. Ginebra, 2011.
- International Organization for Standardization (2011b). ISO 25964: information and documentation- thesauri and interoperability with other vocabularies - part 2: interoperability with other vocabularies. Ginebra, 2011.
- Martínez-González, M. Mercedes; Alvite-Díez, María-Luisa (2014). On the evaluation of thesaurus tools compatible with the Semantic Web. // Journal of Information Science. 40:6 711-722.
- Moreira, Walter; Davanzo, Luciana; Moraes, Isabela S. Abordagem sobre vocabulários controlados para arquivos: conceitos, aplicações e metodologias. // Rodríguez Muñoz, J. V. et al. (Orgs.). Organización del conocimiento: sistemas de información abiertos: actas del XII Congreso ISKO España y II Congreso ISKO España y Portugal. Murcia: Universidad de Murcia, 2015.
- Pavel, Silvia; Nolet, Diane (2002). Manual de terminologia. Canadá: Departamento de Tradução do Governo Canadense, 2002.
- Scornik, Carolina; Sosa, Osvaldo (2007). Análisis de software libre para la gestión de lenguajes documentales: caso de estudio: software TemaTres. Corrientes: Universidad nacional del Nordeste.
- The World Wide Web Consortium (W3C). 2012. Introduction to SKOS. <http://www.w3.org/2004/02/skos/intro>. (2016-03-16).
- Troitiño, Sonia (coord.) (2016). Plano de classificação e tabela de temporalidade de documentos da UNESP: atividades meio. São Paulo: Cultura Acadêmica, 2016.
- Unesco (1993). Diretrizes para o estabelecimento e desenvolvimento de tesauros monolíngües. [2. ed. rev. por Derek Austin e Peter Dale]; Trad. de Bianca Amaro de Melo; rev. de Lígia Maria Café de Miranda. Brasília: IBICT; SENAI.

Enviado: 2016-04-29. Segunda versão: 2016-06-13.  
 Aceptado: 2016-06-04.



---

# Procedimiento de evaluación

*Evaluation process*

---

## 1. Introducción

El procedimiento de evaluación en Scire es el de revisión por pares mediante el sistema de doble ciego —los revisores no conocen el nombre de los autores, ni viceversa— semejante al de otras revistas científicas, aunque tiene algunas peculiaridades que permiten ajustarlo a las condiciones específicas de su campo científico y de sus lectores.

## 2. Objetivos

El objetivo del procedimiento de evaluación es que se puedan admitir con rapidez los trabajos buenos, mejorar los que pueden ser susceptibles de aceptación —esto es, que, siendo interesantes, contienen errores o lagunas que deben ser corregidas— y rechazar los que no son adecuados para la revista por su tema o calidad, redirigiéndolos hacia otras publicaciones adecuadas o haciendo sugerencias al autor para la posible reutilización y reorientación de su trabajo.

El procedimiento de revisión tiene que tener en cuenta a los autores de los artículos que revisa, con los que el revisor debe intentar ser justo, pero también debe responder ante los lectores y los demás autores que compiten por el espacio de la revista, así como, de forma más general, con el proyecto editorial, la propia comunidad científica y el conjunto de la sociedad.

## 3. Procedimiento

Cada artículo requiere al menos la opinión favorable de al menos dos revisores y cada nota breve de al menos uno. Si el director de la revista estima que el artículo entra dentro de los objetivos de Scire —que se encuentran públicamente disponibles en cada número y en la página web de la revista—, lo envía a dos revisores —o, si es necesario, a más— elegidos por el director. Los evaluadores pueden ser miembros del consejo científico, revisores propuestos por los miembros del consejo científico con anterioridad o con motivo del artículo en cuestión, o científicos de reconocido prestigio propuestos por el autor o autores del artículo. La elección se hace según su conocimiento del tema y el número de trabajos que han evaluado hasta el momento (para equilibrar la carga de trabajo entre ellos).

Si el director estima sin lugar a dudas que el tema del artículo no se adecua al de la revista, se lo comunica al autor con una explicación suficiente, y, si le es posible, le sugiere otra revista que pueda ser más adecuada. De esta forma, se ahorra tiempo del autor y de los revisores.

La agilidad del procedimiento de revisión es un aspecto fundamental, por lo que el director solicita a los revisores que envíen su valoración en dos semanas, y que, en caso contrario, renuncien expresamente al proceso de revisión. Sin embargo, si el revisor no realiza ninguna de estas acciones, el proceso se retrasa sin remedio. En este sentido, el autor debe tener en cuenta que el proceso de revisión tiene un carácter voluntario. La decisión de aceptar el trabajo depende, salvo en casos excepcionales, de la opinión de los revisores. En los casos en que no están de acuerdo, la decisión es tomada por el director de la revista, que, excepcionalmente, puede decidir enviar el artículo a un nuevo revisor.

## 4. Criterios de evaluación

Se solicita a cada revisor que indique su valoración sobre los siguientes aspectos en una escala del 1 al 5 (1 muy cuestionable, 2 marginal, 3 suficiente, 4 sólido, 5 excelente):

1. *Tipo de contribución*: Se indica si se trata de una contribución de tipo teórico o metodológico, un estado de la cuestión, una comunicación de resultados de investigación, una experiencia específica, una exposición con valor introductorio o docente sobre el tema o una noticia de interés. Si es una mezcla de varios tipos, es posible marcar más de uno.

2. *Adecuación del tema de la revista*: Aunque al enviar el artículo al revisor el director de la revista ya ha realizado un juicio sobre su adecuación al enfoque de la revista Scire, el revisor puede también valorar este aspecto, comparando el tema del artículo con la declaración de objetivos y alcance de la revista, disponible en cada número y en la página web.

3. *Interés e importancia del tema*: Grado en el que el artículo aborda una cuestión de gran interés para el área científica de Scire, ya sea tradicionalmente, en el momento actual o en el futuro previsible.

4. *Actualidad del tema*: Grado en el que el tema se sitúa en los frentes actuales de investigación. No es un aspecto imprescindible, pero sí se considera informativo.

5. *Novedad y originalidad teórica de la aportación*: Grado en el que el artículo aporta una novedad teórica a su campo de investigación; esto es, un nuevo problema, un nuevo enfoque, un nuevo modelo, una nueva relación o ley, una nueva hipótesis, un nuevo concepto o una interpretación o modificación original de otros anteriores. Debe tratarse lógicamente de una aportación relevante y fundamentada.

5. *Novedad y originalidad metodológica de la aportación*: Grado en el que el artículo aporta una novedad metodológica a su campo de investigación; esto es, un nuevo procedimiento de estudio, de análisis o de prueba, aunque sea a un problema conocido.

7. *Aplicabilidad práctica*: Grado en el que las ideas del artículo pueden servir para la mejora de procedimientos y actividades prácticas y, en definitiva, para el desarrollo y la innovación.

8. *Rigor metodológico y validez de los resultados*: Grado de coherencia, exactitud, precisión y cuidado con el que se aborda el tema de estudio; validez y actualidad de la metodología utilizada; rigor en la selección de las muestras y en la aplicación de la metodología a las mismas; grado de replicabilidad del estudio.

9. *Rigor en el estado de la cuestión y el reconocimiento del trabajo previo*: Hasta qué punto se ha realizado un estado de la cuestión suficiente y actualizado y se han manejado las referencias pertinentes para el problema abordado.

10. *Claridad expresiva, orden y facilidad de lectura*: El artículo debe estar redactado de forma clara y bien ordenada, sin redundancias ni ideas fuera de contexto; y debe ser fácilmente entendible por cualquier lector potencial de Scire. No debe faltar información necesaria ni deben existir lagunas sobre el problema abordado, el método, los resultados, la discusión, las conclusiones o las recomendaciones. Si el artículo contiene aparato matemático, debe ser comprensible por un lector no especialista.

11. *Adecuación de gráficos, ilustraciones, tablas y apéndices*: Los gráficos, ilustraciones, tablas y apéndices deben ser necesarios y relevantes, estar bien resueltos técnicamente y disponer de una leyenda suficientemente explicativa.

13. *Conformidad de las referencias a las normas de la revista*: Grado en el cuál las referencias del artículo son suficientes y se conforman a las normas de la revista Scire.

## 5. Familiaridad del revisor con el tema

Aunque el director de la revista envía los artículos para su revisión intentando que coincidan con el campo de especialidad del revisor, pueden producirse desajustes. El revisor puede rechazar la revisión sugiriendo o no un revisor alternativo, o puede efectuar la revisión indicando de 1 (mínimo) a 5 (máximo) su grado de familiaridad con el tema del artículo evaluado. Si no rellena esta información, se entenderá que su grado de familiaridad es alto (4) o muy alto (5).

Esta información sirve para que el director de la revista pueda estimar la necesidad de enviar el artículo a un nuevo revisor en caso de conflicto de pareceres entre los revisores.

## 6. Valoración final

El revisor puede realizar cinco tipos de propuestas finales:

1. Aceptar el artículo para su publicación sin restricciones ni cambios, porque es un trabajo de alta calidad.

2. Aceptar el artículo para su publicación una vez realizadas las correcciones indicadas, bien sea incondicionalmente, al ser un trabajo de calidad, o bien de forma condicionada a que haya espacio disponible en el número, por ser un trabajo de calidad marginal. Las correcciones solicitadas se indican en las instrucciones para el autor, son de carácter menor y su comprobación puede ser delegada en el comité editorial.

3. Someter a reevaluación el artículo una vez realizadas las correcciones indicadas en las instrucciones para el autor, porque son de importancia y amplitud, y deben ser comprobadas por el revisor original u otro revisor.

4. Rechazar el artículo por las razones indicadas en las instrucciones para el autor, donde el revisor especificará con claridad y amabilidad las razones por las que su trabajo debe ser rechazado y, si le resulta posible, realizará sugerencias sobre su mejora o lugares alternativos de publicación.

5. El revisor rechaza evaluar el artículo por su falta de familiaridad con el tema o por razones que especifica en los Comentarios confidenciales para el equipo directivo.

La labor de los revisores es fundamental para que una publicación científica realice su labor de manera adecuada. Scire agradece de forma muy especial el tiempo y esfuerzo que dedican, y reconoce su trabajo publicando periódicamente sus nombres.

---

# Instrucciones para la presentación de trabajos

*Instructions for authors*

---

## 1. Cuestiones generales

---

Scire admite artículos con estados de la cuestión, discusiones de carácter teórico y metodológico, e informes de investigación sobre proyectos terminados, así como noticias sobre proyectos en curso, experiencias, notas y reseñas, que traten de la representación, normalización, tratamiento, recuperación y comunicación de la información y el conocimiento.

Las lenguas de publicación de Scire son el español, el inglés y el portugués.

Los artículos son sometidos a proceso de revisión por pares según el sistema de doble ciego, aunque los autores pueden proponer revisores libremente. El procedimiento es público, y puede ser leído en el apartado anterior a las instrucciones.

Los autores se comprometen a presentar y discutir públicamente y presencialmente sus trabajos en el congreso Scire correspondiente al año en curso.

Los trabajos deben ser originales y tendrán una extensión máxima de 20.000 caracteres en el caso de los artículos; 12.000 caracteres para los estudios de caso; y 6.000 caracteres para las noticias, proyectos en curso, experiencias, notas y reseñas. Excepcionalmente, estas cifras podrán ajustarse en virtud del interés del trabajo, aunque siempre con el acuerdo previo del editor.

Los autores deben presentar en una hoja aparte sus datos personales completos: nombre, cargo, dirección, teléfono, fax y correo electrónico.

Los artículos han de enviarse por correo electrónico en un documento Word o RTF maquetados según la plantilla disponible en [www.ibersid.org](http://www.ibersid.org) con sus gráficos incluidos. Irán precedidos de sus títulos, resúmenes informativos y palabras clave, todos ellos en inglés y español. Se incluirán en ficheros aparte los gráficos e imágenes en su archivo fuente para su eventual tratamiento, y, en cualquier caso, con resolución de impresión.

Todos los apartados, incluidos las notas, referencias y apéndices irán numerados secuencialmente. Los subapartados —cuyos títulos serán claros y sintéticos— deberán tener la forma “1.1. ”, “1.2. “, “1.2.1. “ y así sucesivamente. No numere los resúmenes en español y en inglés, ni

los apartados de notas, agradecimientos y referencias. Evite un rango de encabezamientos superior a tres.

No se deben relatar datos analíticos repetitivos ni elencos de recursos en el texto principal. Por favor, preséntelos en forma de tablas o, si fuera imposible, en forma de apéndice.

El editor podrá realizar cambios de estilo para adecuar el trabajo a los requisitos de la revista.

Se recuerda al autor que la violación de derechos intelectuales y códigos éticos es un comportamiento inaceptable y puede constituir un grave delito.

## 2. Formatos de texto

---

No utilice subrayados, negritas ni versales.

Utilice mayúsculas sólo para acrónimos. No ponga puntos y espacios entre las letras que los forman.

Las cursivas se deben utilizar tan sólo para resaltar términos en otras lenguas, títulos de obras, palabras clave muy importantes que podrían servir a modo de títulos de un párrafo o conjunto de párrafos, y, con parquedad, para enfatizar los demás términos y frases a gusto del autor.

No utilice listas con salto de línea o de párrafo. Redacte párrafos completos según las reglas y los recursos gramaticales disponibles en su lengua de redacción. Puede indicar orden en las enumeraciones por medio de números o letras entre paréntesis, como, por ejemplo, (1), (2)...

## 3. Figuras y gráficos

---

Las figuras deben presentarse también en ficheros aparte en formato EPS o PDF con resolución de impresión de alta calidad (600 ppp) y, además, en el formato de la aplicación original con la que se creó.

Tenga cuidado si genera imágenes a partir de la pantalla de su ordenador. Si recurre a imprimir pantalla en Windows obtendrá solo una resolución de 72 ppp. En ese caso, utilice una pantalla lo más grande posible, amplíe la ventana al máximo, copie con la tecla “impr pant”, pegue en el programa “Paint” y guarde el fichero; compruebe

luego los resultados. Existen programas que permiten obtener imágenes de pantalla a una resolución aceptable.

En el cuerpo del artículo las imágenes portarán epígrafes indicativos de su contenido precedidos de la palabra "Figura" y un número correlativo (por ejemplo, "Figura 1. "). No encuadre las figuras o gráficos, ni repita dentro de ellas el título.

#### 4. Tablas

Deben realizarse con el mismo procesador de textos que el artículo. En texto portarán epígrafes indicativos de su contenido precedidos de la palabra "Tabla" y un numeral romano correlativo (por ejemplo, "Tabla I. ").

#### 5. Notas

Las notas explicativas deben usarse excepcionalmente. No utilice el sistema automático de su procesador de textos. Refiéralas en texto tecleando su número secuencial entre paréntesis, e inclúyalas al final del trabajo en un apartado titulado "Notas" situado inmediatamente antes del dedicado a las "Referencias", en orden secuencial y precedida cada una de ellas de su correspondiente número entre paréntesis seguido de tabulador. Las citas bibliográficas se realizarán como en el resto del texto; y la referencia completa se incluirá en el apartado de "Referencias". Las páginas web se consideran referencias bibliográficas y deben tratarse como tales, evitando su cita como nota.

#### 6. Citas bibliográficas en texto

Las referencias bibliográficas en texto se denotarán por el apellido del autor, el año del trabajo y, si es necesario, el número de página. Si conviene, el nombre y el apellido del autor podrán quedar fuera del paréntesis. Si dos autores de la lista de referencias poseen el mismo primer apellido, se referirán por sus dos apellidos y, si tienen ambos apellidos iguales, por el nombre también. Si dos o más trabajos del mismo autor son del mismo año, se distinguirán mediante letras del alfabeto secuenciales en minúsculas pospuestas a la fecha sin espacio. Si hay varias referencias dentro del paréntesis, se separarán por punto y coma si son de diferentes autores, y coma si son del mismo.

Así, por ejemplo, "Es una teoría sustentada por varios autores (Gallego, 1975; Fernández y Alonso, 1993) [...]". "Según Jaime Pérez (1993, p. 24) se establecen [...]". "El mismo autor (Pérez, p. 27) hace constar la conveniencia de [...]". "Se han detectado diferencias en la replicación

de la experiencia (Menéndez, 1994a, 1994b; Menéndez y Alonso, 1997)".

#### 7. Formato de las referencias bibliográficas

Las referencias bibliográficas se incluirán por orden alfabético al final del artículo bajo el título "Referencias". Deberán seguir la norma UNE 50104:1996/ISO 690:1987, y, en cuanto a la puntuación, aproximarse lo más posible a las ISBD abreviadas, aunque sin espacio delante de punto, coma y punto y coma. Las referencias a páginas web deben incluirse también en este apartado, no en el texto ni en nota. Refiera adecuadamente en el texto del artículo toda la bibliografía utilizada; no incluya en el apartado referencias documentos que no haya incluido en el texto. Se presenta a continuación el formato para los principales tipos de materiales, con los elementos que pueden darse o no en cursivas, y seguidos de un ejemplo.

##### 7.1. Monografías, congresos, informes, normas y tesis

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. Edición. Lugar de publicación: editor; lugar de publicación: editor. ISBN número. Tesis doctoral.

Delclaux, Isidoro; Seoane, Julio (1982). Psicología cognitiva y procesamiento de la información: teoría, investigación y aplicaciones. Madrid: Ediciones Pirámide.

##### 7.2. Artículos de publicaciones periódicas

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. // Título de la publicación periódica. ISSN número. Volumen:número, primera página-última página.

Ellis, David (1992a). The physical and cognitive paradigms in Information Retrieval Research. // Journal of Documentation. 48:1, 45-46.

##### 7.3. Capítulos de una monografía, de un informe o contribuciones a un congreso

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. // Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. Edición. Lugar de publicación: editor; lugar de publicación: editor. ISBN número. Primera página-última página.

Markey, Karen (1990). Keyword searching in an online catalog enhanced with a library classification. // Bengtson, Betty G.; Hill, Janet Swan (eds.). Classification of library materials: current

and future potential for providing access. New York: Neal-Shuman Publishers. 99-125.

Smith, Ph. J.; Beghtol, C.; Fidel, R.; Kwasnik, B. H. (eds.) (1993). *Proceedings of the 4th ASIS SIG/CR Classification Research Workshop*: Columbus, OH, Oct.24, 1993. Silver Spring, MD.: American Society for Information Science.

#### 7.4. Páginas web y recursos en línea

Se tratarán como monografías, artículos o capítulos según corresponda y se añadirá al final detrás de punto el URL y la fecha de consulta entre paréntesis:

Referencia. URL (Fecha de consulta).

Sagredo Fernández, Félix; Espinosa Temiño, María Blanca (2000). Del libro, al libro electrónico-digital. // Cuadernos de Documentación Multimedia. 9 (2000). <http://www.ucm.es/info/multi/doc/multidoc/revista/num9/cine/sagredo.htm> (2001-01-22).

### 8. Apéndice I. El resumen

Un resumen proporciona información sobre objetivos, alcance, metodología, resultados, conclusiones y recomendaciones. Por alcance se entiende la determinación de hasta qué punto se han tratado los objetivos. Los resultados, conclusiones y recomendaciones se distinguen unos de otros así: los resultados son las informaciones o datos específicos y concretos que se obtienen de la aplicación de la metodología; las conclusiones se derivan de los resultados bien al interpretarlos en el marco de conocimientos teóricos previos o bien al crear un marco de interpretación nuevo a partir de ellos a través de un proceso de abstracción; finalmente, las recomendaciones se refieren a los nuevos problemas y vías de estudio que plantean a la vista de los resultados y conclusiones.

El resumen se debe redactar de forma clara y concisa, sin redundancias ni perífrasis ni información difícil de interpretar, como neologismos o abreviaturas que no se desarrollan.

Es muy conveniente que su primera frase contenga en primer lugar la mención del tipo de documento. Por ejemplo, "Estado de la cuestión sobre la aplicación del modelo europeo de calidad a las bibliotecas universitarias", "Proyecto de aplicación del MARC21 a la catalogación de páginas web institucionales", etc.

### 9. Apéndice II. La redacción del artículo científico: algunos aspectos obvios frecuentemente descuidados

#### 9.1. Sistemática

El artículo científico debe redactarse de forma muy organizada, con una estructura textual firme. En los informes de resultados de proyectos, se suele utilizar el esquema objetivos-metodología-resultados- conclusiones- recomendaciones. En los estados de la cuestión se aborda el asunto en orden cronológico o bien sistemático, según la ontología propia del área científica en cuestión; sigue la discusión, la síntesis, las conclusiones y las recomendaciones. Los artículos críticos suelen seguir el esquema tesis-síntesis-antítesis. Otros tipos documentales pueden exigir otras estructuras textuales, pero, en cualquier caso, dicha estructura debe ser explícita y aparente.

#### 9.2. Claridad y concisión

El lenguaje del artículo científico debe ser claro y conciso. Excepcionalmente, en los trabajos de tipo ensayístico las personas con un alto dominio del lenguaje escrito pueden permitirse la utilización de recursos literarios.

#### 9.3. Pertinencia y parsimonia

Debe evitarse proporcionar información ya conocida, salvo si constituye el punto central de la discusión. A ella cabe referirse por medio de las citas o, si estas no son necesarias, suponerla patrimonio de la comunidad profesional y científica. Sistematizar conocimiento es una función que se realiza solo en un tipo de artículo científico —la revisión o estado de la cuestión—; en los demás casos es contraproducente, aunque se deben citar las fuentes y las obras clave en la introducción del trabajo.

#### 9.4. Civildad

Es necesario citar a otros autores que han trabajado sobre el tema, demostrando que se conoce la literatura previa y que el trabajo se inserta civilizadamente en un determinado contexto científico.

---

# Índice de autores

*Author index*

---

ALMEIDA, Carlos Cândido de, 57  
ALONSO BERROCAL, José Luis, 67  
BIZELLO, Maria Leandra, 99  
CAPRIOLI, Mariana Silva, 75  
FARIAS, Mona Cleide Quirino da Silva, 57, 99  
FIGUEROLA, Carlos G., 67  
FUJITA, Mariângela Spotti Lopes, 37, 107  
GARCÍA JIMÉNEZ, Antonio, 21  
GILCHRIST, Alan, 13  
GÓMEZ CAMARERO, Carmen, 47  
HUMANES, María Luisa, 87  
LÁZARO, Verónica, 29  
LIMA, Larissa Mello, 75  
LÓPEZ DEL RAMO, Joaquín, 87  
MONCLA, Ludovic, 29  
MORAES, João Batista Ernesto, 75  
MOREIRA, Walter, 107  
NOGUERAS-ISO, Javier, 29  
PALOMARES-PERRAUT, Rocío, 47  
SANTOS, Luciana Beatriz, 37, 107  
ZAZO RODRÍGUEZ, Ángel F., 67

---

# Índice de materias en español

*Subject index in Spanish*

---

Acceso abierto, 47  
Actualización terminológica, 37  
Análisis de contenido, 87  
Análisis del discurso, 75  
Archivística, 107  
Archivos universitarios, 107  
Bibliotecas universitarias, 37  
Brasil, 37, 57  
Calidad, 47  
Casos de éxito, 13  
Ciencia de la información, 13, 57, 75  
Cine, 47  
Cinematografía, 47  
Construcción, 107  
Crisis de los refugiados sirios, 37  
Detección de temas, 67  
Difusión, 47  
Documentación, 21, 57  
Documentación fotográfica, 87  
Epistemología, 57  
Escuela francesa de análisis del discurso, 75  
España, 47  
Estructuras categoriales, 37  
Extracción de información, 29  
Folksonomías, 21  
Fotocinema, 47  
Fotogalerías, 87  
Fotografía, 47  
Foucault, Michel, 75  
Framing visual, 87  
Gestión de documentos, 107  
Información geográfica, 29  
Integración de datos, 29  
Interdisciplinariedad, 75  
Latent Dirichlett Allocation, 67  
Lenguajes de indización, 37  
Mercado semántico, 21  
Mediación semiótica, 99  
Memoria, 99  
Memoria cultural, 99  
Metadatos, 13  
Ontologías, 21  
Organización del conocimiento, 13, 21, 57, 99  
Peirce, Charles Sanders, 57  
Periodismo, 21  
Perspectivas, 13  
Planificación, 107  
Prensa internacional, 87  
RDFa, 29  
Refugiados, 87  
Relaciones jerárquicas, 37  
Representación de la información visual, 87  
Revistas científicas, 47  
Schema.org, 29  
Temas de actualidad, 67  
Teorías semióticas, 57  
Tesauros, 21, 47, 107  
Tratamiento temático de la información, 75  
Twitter, 67  
Visibilidad, 47  
Web semántica, 29

---

# Índice de materias en inglés

*Subject index in English*

---

- Archival science, 109
- Brasil, 39
- Brazil., 59
- Categorical structures, 39
- Cinema, 49
- Cinematography, 49
- Construction, 109
- Content analysis, 89
- Cultural memory, 101
- Data integration, 29
- Discourse analysis, 77
- Dissemination, 49
- Documentation, 21
- Epistemology, 59
- Folksonomies, 21
- Fotocinema, 49
- Foucault, Michel, 77
- French school of discourse analysis, 77
- Geographic information, 29
- Hierarchical relations, 39
- Indexing languages, 39
- Information extraction, 29
- Information science, 13, 59, 77
- Interdisciplinarity, 77
- International press, 89
- Journalism, 21
- Knowledge organization, 13, 21, 59, 101
- Latent Dirichlett Allocation, 69
- Memory, 101
- Metadata, 13
- Open access, 49
- Peirce, Charles Sanders, 59
- Perspectives, 13
- Photo galleries, 89
- Photografic documentation, 89
- Photography, 49
- Planning, 109
- Quality, 49
- RDFa, 29
- Record management, 109
- Refugees, 89
- Schema.org, 29
- Scientific journals, 49
- Semantic markup, 21
- Semantic web, 29
- Semiotic mediation, 101
- Semiotic theories, 59
- Spain, 49
- Subject indexing, 77
- Success stories, 13
- Syria refugee crisis, 89
- Thesauri, 21, 49, 109
- Topic detection, 69
- Trending topics, 69
- Twitter, 69
- University archives, 109
- University libraries, 39
- Updating, 39
- Visibility, 49
- Visual framing, 89
- Visual representation, 89