

SCIRE

Representación y organización del conocimiento

SCIRE

Representación y organización del conocimiento

Vol. 24, n.º 1, enero-junio 2018

ISSN 1135-3716

Scire:

knowledge representation and organization

Vol. 24, n. 1, January-June 2018

Ibersid:

Red de Investigación
en Sistemas de Información
y Documentación

© 2018 Francisco Javier García Marco

Reservados todos los derechos. Prohibida la reproducción por cualquier medio de la totalidad o parte de la presente edición con fines comerciales 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.

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.

Se permite expresamente el autoarchivo de la versión electrónica por parte de los autores en sus repositorios personales e institucionales. Los autores pueden reproducir los textos y figuras de sus artículos sin permiso expreso del editor indicando la fuente de publicación, y conservan los derechos de su original con la obligación de citar la presente edición.

ISSN: 1135-3716 = Scire (Zaragoza)

ISBN:

Depósito legal: Z. 1.790 — 1995

Edita: Ibersid® con la colaboración de Prensas de la Universidad 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, ESCI, 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 organización de congresos científicos del Vicerrectorado de Investigación de la Universidad de Zaragoza, del año 2017.

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, ESCI, 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 organization of scientific conferences of the Vicerectorate of Research of the University of Saragossa, Spain, for the year 2017.

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

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

Revisores externos del número / External reviewers in this issue

Agradecemos la colaboración altruista y desinteresada de María Luisa Alvite Díez, José Luis Herrera Morillas, Joana Lemos, Daniel Martínez-Ávila, Javier Nogueras Iso, Adriana Carla de Oliveira, Ely Tannuri de Oliveira, Dunia Padrón, Ana Ríos Hilario, Aires Rover, Blanca Rodríguez Bravo, 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

<i>Tabla de contenidos en español</i>	9	<i>Aproximación bibliométrica a las contribuciones de Hope A. Olson en el campo de la organización del conocimiento</i>	91
<i>Tabla de contenidos en inglés</i>	11	<i>María Camila RESTREPO FERNÁNDEZ</i>	91
Editorial			
<i>De las tarjetas perforadas a Google: una historia esquemática de la ciencia de la información</i>	13	<i>Procedimiento de evaluación</i>	103
<i>Alan GILCHRIST</i>	13	<i>Instrucciones para la presentación de trabajos</i>	105
Artículos		<i>Índice de autores</i>	108
<i>Representación, información, identificación y valoración del libro antiguo: bibliotecarios e investigadores ante un problema común</i>	23	<i>Índice de materias en español</i>	108
<i>MANUEL JOSÉ PEDRAZA GRACIA</i>	23	<i>Índice de materias en inglés</i>	108
<i>Estudios sobre el análisis de contenido en Brasil: estado de la cuestión y perspectivas futuras</i>	35		
<i>Paula Regina DAL'EVEDOVE Roberta Cristina Dal'Evedove TARTAROTTI Mariângela Spotti Lopes FUJITA</i>	35		
<i>La información como tema en el derecho y en la ciencia de la información: un estudio comparativo a partir de la realidad brasileña</i>	45		
<i>João Carlos Gardini SANTOS José Augusto Chaves GUIMARÃES Juan-Carlos FERNÁNDEZ-MOLINA</i>	45		
<i>Confluencias filosóficas para la clasificación de las artes: aproximaciones de Bacon, Hegel y Harris</i>	55		
<i>Verônica de Sá FERREIRA Rodrigo de SALES</i>	55		
<i>Diseño de información, ciencia de la información y organización del conocimiento: un análisis de dominio desde la perspectiva de la complejidad</i>	67		
<i>Natália NAKANO Daniel MARTÍNEZ-ÁVILA María José Vicentini JORENTE Mariana CANTISANI</i>	67		
<i>Del CMBD al Big Data en salud: un sistema de información hospitalaria para el siglo XXI</i>	77		
<i>Gonzalo MARCO CUENCA José Antonio SALVADOR OLIVÁN</i>	77		

Tabla de contenidos en inglés

Table of contents in English

<i>Table of contents in Spanish</i>	9	<i>Evaluation process</i>	103
<i>Table of contents in English</i>	11	<i>Instructions for authors</i>	105
Editorial		<i>Author index</i>	108
From punched cards to Google: an outline history of information retrieval.....	13	<i>Subject index in Spanish</i>	108
<i>Alan GILCHRIST</i>	13	<i>Subject index in English</i>	108
Articles			
<i>Portrayal, information, identification and value of rare book: librarians and researchers facing a shared problem</i>			
<i>MANUEL JOSÉ PEDRAZA GRACIA</i>	23		
<i>Studies about subject analysis in Brazil: research and future perspectives</i>			
<i>Paula Regina DAL'EVEDOVE Roberta Cristina Dal'Evedove TARTAROTTI Mariângela Spotti Lopes FUJITA</i>	35		
<i>Information as a subject in law and in information science: a comparative study based on the Brazilian reality</i>			
<i>João Carlos Gardini SANTOS José Augusto Chaves GUIMARÃES Juan-Carlos FERNÁNDEZ-MOLINA</i>	45		
<i>Philosophical confluences for the classification of Art studies: the approaches of Bacon, Hegel, and Harris</i>			
<i>Verônica de Sá FERREIRA Rodrigo de SALES</i>	55		
<i>Information Design, Information Science, and Knowledge Organization: a domain analysis from the perspective of complexity</i>			
<i>Natália NAKANO Daniel MARTÍNEZ-ÁVILA Maria José Vicentini JORENTE Mariana CANTISANI</i>	67		
<i>From MBDS to big data: a hospital information system for the 21st century</i>			
<i>Gonzalo MARCO CUENCA José Antonio SALVADOR OLIVÁN</i>	77		
<i>Bibliometric approach to the contributions of Hope A. Olson to Knowledge Organization</i>			
<i>María Camila RESTREPO FERNÁNDEZ</i>	91		

From punched cards to Google: an outline history of information retrieval

De las tarjetas perforadas a Google: una historia esquemática de la ciencia de la información

Alan GILCHRIST

Cura Consortium, alangilchrist77@gmail.com

Resumen

Se revisa la historia de la recuperación de información (IR) desde las tarjetas perforadas y la primera computadora programable (el ENIAC de 1945) hasta el actual buscador web Google y la “tecnología cognitiva” Watson de Microsoft. La revisión se basa en tres factores principales en el desarrollo de IR; (1) el enorme aumento en el poder de cómputo en los últimos 72 años, (2) la “competencia” entre el análisis estadístico del texto y el procesamiento del lenguaje natural (NLP) en la que ambos finalmente han convergido en gran medida, y (3) los cambios correspondientes en la intervención humana en el proceso de IR.

Palabras clave: Ciencia de la información. Recuperación de la información. Historia. Tendencias. Prospectiva.

1. Introduction

There have been four great revolutions in the history of human communication: spoken language, the invention of writing in Sumeria in around 2600 BCE, the Gutenberg press which introduced modern printing in Europe in the mid 15th Century; and the present age of the computer starting in earnest in 1945 with the launch of the ENIAC computer. The gap between writing and printing is thus roughly 4000 years and that between printing and the computer about 500 years. Seventy years after ENIAC we are increasingly affected by a rapidly developing revolution in information technology. Whether we think of clay tablets, papyrus scrolls, paper or tapes and disks people have stored these information carrying media in collections and libraries. So, from the earliest times people have ‘retrieved’ information embedded in documents, but it was not until 1950 that the term ‘Information retrieval’ was coined, attributed to Calvin Mooers by *the Oxford English Dictionary* with the words “The requirements of information retrieval, of finding information whose location or very existence is *a priori* unknown...”

This paper is concerned with the development of information retrieval (IR) from that time until the present with three major strands in mind: first, the extraordinary increase in computing power; sec-

Abstract

This paper reviews the history of information retrieval (IR) from punched cards and the first programmable computer (the ENIAC of 1945) to the present day Web searcher Google and Microsoft’s “cognitive technology” Watson. The review is based on three major factors in the development of IR; (1) the enormous increase in computing power over the last 72 years, (2) the “competition” between statistical analysis of text and Natural Language Processing (NLP) in which the two have finally to a large extent converged, and (3) the corresponding changes in human intervention in the IR process.

Keywords: Information science. Information retrieval. History. Trends. Prospective.

ond, the techniques employed in information retrieval, notably statistical analysis of text and competing techniques of Natural Language Processing (NLP); and third the radically changing activities of human intervention in the IR process due to the increasing sophistication brought about by the first two strands.

2. Genesis of the modern computer

Most histories of computers start with descriptions of the early use of punched cards at the end of the 19th and beginning of the 20th Centuries and which can be seen as precursors of the modern computer, but mention should also be made of two early pioneers in their attempts to produce calculating machines. The first is the brilliant French mathematician and philosopher, Blaise Pascal who, at the age of eighteen, devised a machine in 1642 based on cogs and gearwheels to operate addition and subtraction, thus helping his father who was a tax inspector. Some thirty years later the German polymath Gottfried Leibniz devised a similar but more sophisticated machine which was capable of multiplication and division as well as calculating square roots. Whereas these two innovations were born in the field of mathematics, punched cards were first devised in the late 18th Century and refined in 1801 by a French inventor who has given his name to the

'Jacquard loom', used to drive the machines employed for the weaving of complex patterns. His revolving chains of punched cards were made of wood! While the Englishman Charles Babbage was certainly aware of, and influenced by, the use made of punched cards by Jacquard he must also have been aware of the ideas of Leibniz probably because the latter had been elected a member of the Royal Society of London of which Babbage was also a member. Babbage worked on two complicated pieces of machinery, first the 'Difference Engine' and later the 'Analytical Engine', earning the title given by some as the 'Father of Computers'. Neither of these devices were finished before his death in 1871, due to lack of funding and, it is said, his grumpy treatment of colleagues and employees, but replicas have been produced in this century and are on display in London's Science Museum. The Analytical Engine could deal with all four arithmetic operations plus comparison and, optionally, the calculation of square roots. It incorporated an input mechanism using punched cards with output to a printer or a curve plotter, and stored the procedures for calculations by use of pegs inserted into revolving drums (a device used by Leibniz). A close acquaintance of Babbage was Ada, Countess of Lovelace (née Ada Byron, the only legitimate daughter of the famous poet). Lovelace was a competent mathematician who understood the work of Babbage and later said that "machines could never truly think", thus countering the extraordinary question put by an English Member of Parliament with the words "Pray, Mr Babbage, if you put into the machine wrong figures, will the right answers come out?"

Despite these early innovations involving punched cards it was not until 1890 that the first seriously practical application of punched cards sorted by machine was invented by Herman Hollerith who worked for the U.S Census Bureau. Daunted by the time and effort expended on analysing the returns of the census of 1880, he set out to build what was subsequently called a 'tabulating machine', which 'read' the holes and their positions on the cards to arrive at accurate results of the data gathering exercise in one year rather than the eight taken for the previous census. Encouraged by this success Hollerith set up the Tabulating Machine Company, which became the Computing-Tabulating-Recording Company (C-T-R) in 1911 and which, in turn and after a series of mergers and acquisitions, later became the International Business Machines Corporation (IBM) in 1924. The stage was now set for big advances in the technology of computing and these were accelerated by a brilliant young Englishman called Alan Turing who conceived of a theoretical

'Logical Computing Machine'. This was a philosophical leap from the engineering aspect of punched cards into the realms of mathematics and logic and represented a huge advance in the history of computing. His concept included the theoretical possibility of handling any mathematical computation from an input carried on continuous paper tape carrying data in binary code; the calculations being based on a 'table of instructions'. In 1936 Turing continued his studies, started at Cambridge University in England, with a visit to Princeton University in America where he met and discussed his ideas with John von Neumann, who was later to be involved in most important developments some ten years later. This resulted in Turing's seminal paper 'On Computable Numbers', and this fed into his work at the now historically famous Bletchley Park code-breaking centre where, during World War II and with Turing's collaboration, a team built in 1943 the first electronic digital and programmable computer and which, called the Colossus, was used exclusively for the decryption of German signals.

3. The Modern computer

The success at Bletchley Park and a surge of theoretical and practical work in the U.S. and other countries produced significant results which heralded the start of years of concerted research and development. It is generally agreed that the first truly modern computer was the ENIAC which was unveiled in 1945, a colossal machine which, according to Wikipedia, weighed almost 35 tons and occupied a space of 167 square metres. It was equipped with 17,468 vacuum tubes to control currents, 7,200 crystal diodes, 1,500 relays, 70,000 resistors, 10,000 capacitors and 5 million soldered joints. It was kept in a carefully heat and atmosphere controlled environment, but even so, was very vulnerable; several vacuum tubes burnt out every day and had to be replaced. There is a story, perhaps apocryphal, that on one occasion moths got into the machine and short circuited it – giving rise to the now common expression "bugs in the machine". Even with this vast array of components, the input and output was effected by punched cards and had no memory until new equipment was installed in 1953. However, its arithmetic calculations were some thousand times faster than electro-mechanical machines. With these dimensions and performance in mind one can forgive IBM President Thomas J. Watson, for saying in 1943 "I think there is a world market for about five computers". Watson was soon proven wrong by an enormous wave of research and development in a number of countries leading to rapid and important advances in com-

puter technology involving new start-up companies earning large amounts of money, and the establishment of the famous 'Silicon Valley', not to mention a number of Nobel prize winners on the way. From 1945 on, the history of the development of computers, seemingly rather paradoxical, is one where computers got smaller and smaller and at the same time more and more powerful. As early as 1946 the first commercial computer arrived, crafted by two engineers who had been responsible for the design of ENIAC, and named UNIVAC (UNiversal Automatic Computer). In fact, it was not released until 1951 when the U.S Census Bureau purchased a machine even though initially, UNIVAC employed clumsy input and output processes. In the 1950s two famous computer programming languages were released, COBOL (Common Business-Oriented Language) and IBM's FORTRAN (FORmula TRANslation) this latter designed for numerical computation and scientific applications, and these were responsible for starting a revolution in the portability and standardization across different hardware platforms and operating systems. At the end of the 1950s there was another huge advance with the invention and introduction of integrated circuits replacing the cumbersome and unreliable vacuum tubes; this miniaturization of hundreds of tiny transistors onto a small piece of semiconductor material being popularly known as a microchip. These smaller, cheaper and much faster devices paved the way for the introduction and spread of mini- and microcomputers, and later the personal computer, thus confounding another poor prediction made in 1977 by Ken Olson, the founder of the Digital Equipment Company when he said, "There is no reason anyone would want a computer in their home", another piece of evidence that showed how fast things were moving. In fact minis, micros and terminals spread very rapidly as they became cheaper and were used widely in workplaces, then installed in universities and later in schools so that students could be taught how to use, and even program them. The Proceedings of a Conference held in London in 1980, co-organized by the British Computer Society and the Institute of Information Scientists was published under the title *Minis, Micros and Terminals for Libraries and Information Services* (Gilchrist, 1981). In the preface it was noted that a computer specialist attending the Conference asked what was an inverted file: this was, in retrospect, an interesting question because it was the information scientists who had 'invented' the inverted file for searching computer held alphabetical indexes, where computer scientists used data dictionaries for their work in, mainly, data processing. Previously, information scientists had introduced the punched card known as a feature

card which, unlike the edge-notched card, represented a subject and had usually 10,000 positions at which to punch the numbers given to the documents containing that subject – in practice, an inverted file. Since then, computer scientists have embraced text processing and the Institute of Information Scientists no longer exists. Refuting Olson, the computer then invaded the home with, in quick succession, the desktop, the laptop, the portable; and then in 2007 in yet another advance in miniaturization the iPhone was launched, a final combination of computer, screen and keyboard, and also the first device that combined these with the telephone (whose development is briefly reviewed below). These advances confirmed what became known as 'Moore's Law' which stated that his observation that the number of transistors per square inch on a microchip had doubled every two years and was likely to continue to do so, and consequently that computer power would double at the same rate. This has proved to be the case though it is recognized that being an exponential growth can not continue indefinitely but it is an astonishing fact that with advances in nanotechnology, some transistors are smaller than a virus. In just over 70 years from the ENIAC to the iPhone it is possible to compare the two devices: the iPhone costs 17,000 times less than ENIAC; it is 40,000 times smaller; it uses 400,000 times less power; it is 120,000 times lighter...; but... it is 1,300 times more powerful.

<i>Flops</i>	<i>Computer</i>	<i>Decade</i>
Hundred – ten squared	ENIAC	1940s
Kilo – ten cubed	IBM 704	1950s
Mega - ten to the power ⁶	CDC 6600	1960s
Giga – ten to the power ⁹	Cray 2 (CERN)	1980s
Tera – ten to the power ¹²	ASCI Red	1990s
Peta – ten to the power ¹⁵	Jaguar	2000s
Exa – ten to the power ¹⁸	???	

Table 1. Processing power

What happens next? What happens if you pack large and powerful computers into a single connected system? The answer is 'parallel processing', already widely employed and known as the 'supercomputer', and whose applications will be briefly discussed at the end of this paper. The Table above shows the growth in computing power from the ENIAC onwards, and is measured in the technical term FLOPS standing for Floating

Point Operations per Second. This refers to a method of encoding real numbers within the limits of finite precision available in computers.

Jaguar was developed by Cray at the Oak Ridge National Laboratory in the U.S. In 2010 it was then ranked as the most powerful supercomputer in the world but in the same year was overtaken by a supercomputer developed in China. Since then more supercomputers have arrived in what looks like a race between the U.S and China. Currently the top spot is occupied by China's Sunway Taihu Light with an amazing strength of 93 petaflops and is also the most energy efficient of the top ten supercomputers.

Almost in parallel with the growth of computer technology, telegraphy arrived to join it in a spectacular way leading to today's revolution in information access and mobile computing. But, to complete a trilogy of embarrassingly bad predictions, when Edison visited London to promote his invention of the telephone he was rebuffed by Sir William Henry Preece, Chief Engineer of the Post Office with the words "...there are conditions in America which necessitate the use of instruments of this kind more than here. Here we have a superabundance of messengers, errand boys and things of that kind". Not only did extensive worldwide telephone networks rapidly develop, but by the late 1960s computer networks were developing such as ARPANET (Advanced Research Project Agency Network) which led directly to the Internet as a merger of a number of independent networks of which ARPANET was the leader. Then in the early 1990s Tim Berners-Lee working at CERN in Geneva is credited with inventing the World Wide Web, and in 1996 Serge Brin and Larry Page launched the Google search engine, these two innovations now being used by literally billions of people every day across the world. Information and Communications Technology has since had a fundamental effect on the social, political and commercial spheres and the world is still adjusting to its present impact and considering its future.

4. Retrieval techniques

4.1. Mathematical techniques

In the famous library in Assyrian Nineveh, the librarian was known as "the man of the tablets". Callimachus of the Alexandria Library devised a catalogue (on scrolls) listing the contents of several different subject-based collections, each arranged by author. Since these early endeavours man has diligently stored and arranged documents in various ways, making it easier to find

specific items or documents that might be of interest to an enquirer. Classifications have been devised and catalogue cards created in the increasing endeavour to guide searchers to content 'hidden' within documents not easily detected from the titles or classification, or even multiple catalogue cards. Increasingly, books were being added to by other forms of document, notably scientific and technological reports and articles in journals. Within these documents, the scope, complexity and granularity of the topics discussed began to overwhelm librarians and a new intermediary came into being – the information scientist; initially a person (in the U.K.) officially qualified with a degree in science or technology plus a second language and five years experience in handling complex information. By the time of the arrival of the ENIAC computer in 1945, the problem was becoming worse and was brought to a head by the acquisition by the U.S. of a vast number of scientific and technological reports captured at the end of World War II, and this is when modern computers entered the field of text processing and the term 'Information retrieval' was born. However, before computers began to develop more sophisticated ways of handling text, mechanised sorting of punched cards was developed further by Calvin Mooers (mentioned earlier as the originator of the term 'information retrieval'). In the 1950s Mooers devised a complex system called Zatocoding such that a broad picture of the contents of a document could be stored on a single Hollerith card. The rationale for this was that a Hollerith card contained 960 positions and that with a four punch Zatocode it was possible to store 165 different 'subject ideas'. Mooers described Zatocoding as "the most efficient coding system presently known". It was also made available by his Company for selecting the smaller edge-notch cards using small mechanical sorters. Another primitive approach to the IR problem was initially applied by its inventor Hans Peter Luhn to the rapid indexing of the large quantity of captured reports mentioned earlier. This was the KWIC Index (Key Word In Context), a simple device that rotated the words in a title presenting them in an alphabetical order. For example, after deleting non-informative words, the title "Analysing sentences, an introduction to English syntax" would be presented as:

- Analysing sentences introduction English syntax
- English syntax analysing sentences introduction
- Introduction English syntax analysing sentences
- Sentences introduction English syntax analysing
- Syntax analysing sentences introduction English

The first realistic use of the computer for IR was to use standard Boolean logic (named after George Boole, the 19th Century English mathematician and inventor of Boolean algebra). This used the relationships AND, OR and NOT to combine search terms, usually taken from a list used in indexing the documents in the collection. Initially, complex constructs were fed into the computer which operated overnight in batch processing mode. The three simple combinations A AND B, A OR B, A NOT B were simple, but more complicated search expressions had to use brackets such as {(A OR B) AND C} or {(D AND E) NOT F} which became more complicated when all three Boolean operators were used in longer strings for a single search. Any mistake in the placing of brackets would lead to ambiguity causing false retrieval and necessitating the running of a corrected search again the next night. Online retrieval overcame this by the ability to conduct two consecutive searches, the second using the NOT function to reject unwanted documents. This approach was, of course, the simplest possible and was followed by many new improvements and alternatives. For example, in the late 1970s Stephen Robertson, a mathematician at the U.K. Cambridge University working with his computer scientist colleague Karen Sparck Jones applied probability theory using Bayesian statistics, to tackle the IR problem. This approach posited that the probability of finding a document that was relevant to the query addressed to a collection could be statistically calculated by comparing mathematical representations of the terms in the document and in the query, and even that portions of the document could be identified as being richer in relevance. These two academics later developed and applied a complete IR package under the name of Okapi BM25 (where BM stands for Best Matching). Here, the probabilistic approach was expanded to rank a set of documents based on the query terms appearing in each document without regard to such aspects as their proximity or other relationships. The system was applied successfully at London's City University where Robertson had become Professor of Information Science. In 1983 Gerard Salton of the American Cornell University and one of the most accomplished thinkers in the history of IR research, proposed the 'Extended Boolean model' which introduced the concepts of partial matching and term weights (also later used with the probabilistic approach described above) in an algebraic approach to the IR problem, and which involved more complex mathematics to create a 'vector space'. This used the algebraic approach to the problem that he had first introduced in the 1960s and continuously developed. Salton's full Vector

Space Model (VSM), launched into the public domain as SMART (System for the Mechanical Analysis and Retrieval of Text) is a difficult concept to grasp for anyone not familiar with mathematical abstractions. In technical terms the model is a hypercube: an n-dimensional space in four (or geometrically more) dimensions. Vector quantities have both direction and size and can be used to describe a movement from one point to another. Within this hypercube vectors are stored and used in calculating the relationships between documents, words (as tokens) and queries with reference to the relevance of results from addressing queries to the collection of documents. Not surprisingly, the algebraic equations addressed by the computer are complex but there was more to come, specifically with Latent Semantic Indexing (LSI) introduced in the late 1980s. This approach introduced the idea that patterns between the semantics of terms, expressed as unit words or their compounds used in the texts of a collection could be correlated with their underlying abstract concepts. The mathematical technique used is called Singular Value Decomposition (according to Wikipedia "A technique used in linear algebra...a factorization of a real or complex matrix"), a description which underlines the complexity of the mathematics used in IR!

The techniques described briefly above fall into three mathematical categories: set-theoretic, probabilistic and algebraic and the examples given are the most notable in a wide field incorporating variations of these three approaches. As these techniques developed, commercial software companies added various refinements and further approaches in the hope of making the packages either more powerful or user-friendly, or both. Very many packages were developed and sold in countries world-wide, tailored to specific languages and requirements. In the U.K some of these were commercial off-shoots of software created and used in-house by large companies and institutions, such as the nuclear research laboratory at Harwell and a package called ASSASSIN created by Imperial Chemical Industries Ltd, which initially ran on IBM machines only, but was later developed to operate with the UK ICL computers. Competition was fierce and various ideas, some of which appeared odd, reached the market place. One example was a package called Excalibur., which was based on early experiences of its inventor started by watching an immobile chameleon which had the ability to identify a passing insect and to automatically launch its rolled-up tongue to catch it. The inventor saw this as pattern matching and developed a sensing machine that could distinguish between

different species of felled tree trunks by the patterns visible on the cut planes. More ambitiously he moved on to applying pattern matching to IR.

4.2. Natural Language Processing

Another, initially more ambitious, approach to IR is provided by Natural Language Processing (NLP), one that has not solved the problem by itself but which has made many significant contributions to the IR problem and is likely to continue to do so. Elizabeth Liddy, not only an expert in NLP but a qualified librarian defines NLP in these words “Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications.” This is a careful definition that does not claim that NLP can solve the IR problem by itself but that it can, and has, provided much significant help. The combination of the two different approaches, one involving the rigid discipline of mathematics and the application of the even more complex linguistic analysis inherent in NLP has proved increasingly fruitful in tackling the complex problems of IR, both employing the increasing power of computers. Liddy and others have described the complexity of language by defining six levels of semantic characteristics:

- Phonetic: the way in which words are pronounced.
- Morphological: referring to the smallest piece of a word that has a meaning, for example by ignoring prefixes and suffixes such as un- and -ation (though this can, by itself, lead to confusion in retrieval).
- Syntactic: the structure of a sentence, taking into account the roles played by different types of words (nouns, verbs etc.)
- Semantic: the meanings of words individually or in combination, taking into account synonyms and various combinations.
- Discourse: the different ways in which structure is used in different communications, for example in newspapers or technical articles.
- Pragmatic: the use of what is called ‘synecdoche’, for example the “White House” to indicate the seat of the U.S. Government (rather than a white house).

As evidence of how difficult a problem it is that faces NLP in its application to IR, it has been reported, some years ago, that Liddy and colleagues developed a prototype software IR package based on the six semantic levels above, but

it was discovered that, at the time of its completion, there were too few computers in operation that had sufficient power to run the programs effectively and the plan was discontinued. However, the basic principles have been successfully applied since then, both manually by information scientists and incorporated in machine processing by computer scientists, thereby providing significant power to tackling the IR problem. Broadly speaking NLP can be used to improve relevance ranking by using such devices as automatic stemming of words, identification of proper nouns or, more cleverly identifying related terms at the concept level. Using many analytical tools NLP, it has been said, can extract additional information at all four stages of the IR process: document processing, query processing, query matching, and ranking and sorting: in short, attacking for example the problems of too many synonyms, too many meanings, inability to specify vague concepts, improving indexing consistency and avoiding variations and errors in spelling. As mentioned above, Latent Semantic Indexing has been applied within NLP to analyse the relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. If all of these actions seem (or did seem) ambitious one only has to look at the performance of Google in searching the Web.

4.3. Human intervention

Before the arrival of the modern computer the information scientist’s role was carried out by librarians acting at the interface between the document collections and the readers. With the onset of the new technology, information scientists found themselves working at two interfaces; the first, working sometimes between the collections and the computers; and also at the interface between the computers and the users. In the first intervention the information scientist indexed the documents, later providing abstracts (some of which, in the scientific area were structured in a standard manner). The indexing was also formalised through the application of a standard vocabulary which device became increasingly sophisticated over time. Initially, the words used in indexing were called ‘terms’, though the American Mortimer Taube coined the word ‘Uniterm’ to indicate that it was a simple ‘unit term’. This word did not catch on, but Taube did introduce another word which had a more lasting and fundamental effect – this was the word ‘concept’ which he used to define the basic idea underlying the word (or words); in reality, the actual meaning beneath the choice of words. Other semantic techniques, more concerned with the operations of the computer, consisted of simple NLP with such devices

as stemming, phrase searching and word order, an unlikely example of this last being quoted in the early days of disambiguating between 'Blind Venetians' and 'Venetian blinds' (a type of window covering). Initially, searches came to be conducted by matching search terms with index terms and to facilitate this exercise simple vocabularies were created, many of which were limited in size (initially for use with punched cards). Tricks were used to economise on the number of words used in the vocabulary such as 'Father USE Male + Parent' (a device called Semantic Factoring). But soon, as computers became more widely used more extensive vocabularies were created and called thesauri. This term, from the Greek word 'thesauros' meaning 'treasury' was initially suggested by Helen Brownson of the U.S. National Science Foundation at a Conference in England, borrowing the word from its well-known use created by Peter Mark Roget, compiler in 1852 of the dictionary of synonyms and related terms of the English language, a work that has undergone numerous revisions ever since. The number of thesauri created then grew dramatically so that by the late 1970s the Directorate of the European Commission responsible for matters of information and technology was able to publish a directory containing over 1500 of these vocabularies used for IR.

One of the pioneer thesauri was that created by the Engineers Joint Council in the U.S., which was followed by two more produced by the Armed Services Technical Information Agency and the other by the American Institute of Chemical Engineers. In 1965 it was decided that these agencies should collaborate in combining these three thesauri into what became a huge publication created by 329 scientists working in a number of coordinated committees, whose effort was later calculated to amount to over 23 man-years. The thesaurus, called TEST (Thesaurus of Engineering and Scientific Terms) contained 17,800 descriptors with a further 5554 lead-in terms and cross-references adding up to 162,657 line entries. In addition there were a permuted index of the compound terms, a subject category index and an accompanying hierarchical index. Other thesauri, one which preceded TEST and the other which followed, broke new ground, the first in its display, the second in its structure. In 1965, Euratom (a European Commission research programme for nuclear research) published a thesaurus with graphical displays. These were initially created as sets of Euler circles and then presented to the users in the shapes of abutting polygons containing randomly placed descriptors with the top terms in each polygon underlined. These displays were augmented by others called

'arrowgraphs' in which uni-directional arrows indicated hierarchical relations and two-directional arrows indicated other relationships. This approach was further emulated in Europe and is later to be seen in what were called 'Topic maps' and later again in the more complex diagrams representing ontologies used in the semantic web (see below). A second innovation in thesaurus construction was the intellectual advance pioneered by Jean Aitchison with the Thesaurofacet, so called because it combined the features of a faceted classification and a thesaurus. In fact, the primary construct was a disciplined faceted classification from which the descriptors were extracted and alternatively displayed in thesaurus format employing normal thesaurus features. This thesaurus was published by the English Division of the American giant GE (General Electric), and contained some 16,000 descriptors augmented by a further 7,000 lead-ins, all terms meticulously connected by the faceted classification. Aitchison went on to create other dual classification/thesaurus works, principally the two-volume UNESCO Thesaurus. The thesaurus had now become the standard device for supporting computer-based systems so that, for example, the National Library of Medicine transformed its Index Medicus subject heading system into a more thesaurus-like structure to support its computerised database called MEDLARS (Medical Literature And Analysis Retrieval System). Since those early days many more thesauri have been created including some large international and multilingual schemes such as those created by the Food and Agriculture Organization (called AGRIVOC), another by the European Commission to support parliamentary documentation (called EUROVOC) and the Art and Architecture Thesaurus produced by the Getty Foundation.

While all this was happening a revolution occurred in IR which began to diminish and in many cases removed the need for indexing. This was full text searching, an advance which relied far more on the computer to do the retrieval work supported by more advanced mathematical and NLP techniques, the origins of which have been described earlier in this paper. This wider range of semantic tricks included requests for two words within x words of each other and suggestions of synonyms derived from a dictionary incorporated in the system. The use of thesauri became search aids, not always used by what had become known as the 'end-users'. With many users of large organizations on scattered sites many were unsupported by trained information professionals and research in the private sector has shown a continuing dissatisfaction with systems of information access and provision. One small study in a U.K. government department in the 1980s was

instructive; the information service had set up a template on users' screens that offered a choice of 'simple' or 'advanced' search. The first was a single box for a subject entry while the second offered a menu of different metadata that could be combined including subject but also author, publisher, etc. (The large topic of metadata is not included here). A small study showed that 95% of the users used simple search against 5% that used advanced search – a foretaste of Google.

There had been a number of national standards presenting guidelines for thesaurus construction since the late 1970s but a major new and thorough standard was published by ISO (2011,2013) in two parts with the title *Thesauri and Interoperability*. A major advance in this work was the recognition of the importance of other vocabularies and the desirability of connecting them in shared subject areas. One of the new forms of structured vocabulary was the 'taxonomy' which, though the word means simply classification was a thesaurus-like structure with more elastic construction rules devised mainly as more user-oriented aids, principally to websites and intranets. In this, it followed in spirit the earlier 'folksonomies' created organically by end-users and preceding the now more common tagging. Meanwhile there were separate developments within the Semantic Web of which the most relevant to this paper can be seen in what is known as the 'Semantic Web Stack' showing the interrelated standards used from the underlying bases up to the user interface. At the lowest level are found the standards, Unicode for character sets and URIs (Uniform Resource Indicators). The URIs disambiguate descriptors with the same name, for example: Segovia, the city and Segovia, the musician. Further up is the standard for data interchange, the Resource Description Framework (RDF). This dictates the way in which resources are described using 'triples', composed of subject-predicate-object, for example: "Prado is situated in Madrid" Above the RDF level is found 'Taxonomies' used here to indicate any standardized vocabulary, and 'Ontologies', (another word borrowed from another era and use). This ontology is defined by Tom Gruber as "a formal specification of a shared conceptualization". As a more complex 'topic map', it can be created for describing the contents of a collection of entities, and unlike the thesaurus that proposes strict rules for relationship types (for example genus-species, part-whole, instance for hierarchies). However, the ontology can define such relationships and more, for example as in the Prado example used above in explaining the RDF. While these are powerful devices for relatively small and homogeneous collections problems grow as the collection increases in size and scope.

It might be useful here to mention the work done in previous years on retrieval performance. The influential experiments carried out in the 1960s in the U.K. at what is now Cranfield University. The Cranfield measures of retrieval effectiveness were 'Recall', the number of correct results divided by the number of results that should have been retrieved and 'Precision', the number of correct results divided by the number of all results. Cranfield compared four systems using the same corpus specially indexed for the experiment: an alphabetical subject catalogue, the Universal Decimal Classification, a faceted classification (specially constructed for the experiment) and a 'Uniterm' index. The actual results of the experiment are now a matter of history; what mattered was that the idea of evaluation using these two measures caught on and were used in many more tests. To these two measures the concept of user-judged relevance was added and relevance feedback loops were built into the IR process. More measures and measurement techniques were added and now the world leader in IR evaluation is undoubtedly TREC (Text Retrieval Conference), a series of workshops addressing different areas of IR research (called 'tracks'). TREC celebrated its 25th Anniversary in 2016 and supplies test collections and evaluation software to 93 groups working in 22 countries.

5. Epilogue

What next? First, 'quantum computing'. The brilliant physicist Richard Feynman once said, "I think I can safely say that nobody understands quantum mechanics." Nevertheless, according to the science journal *Nature* Google and Microsoft may release post-research models of quantum computers in 2017. If Feynman is right few people will understand the definition of quantum computing. According to Wikipedia, "Quantum computers make direct use of quantum-mechanical phenomena such as superposition and entanglement to perform operations on data". Meanwhile, in a recently published paper with the title "Combining word semantics within complex Hilbert Space for information retrieval", the introduction says "In quantum theory, states are represented by vectors defined as a complex-valued Hilbert Space...In information retrieval, pioneering work showed that the quantum formalism encompasses many state-of-the-art retrieval models and subsequent works proposed many quantum-like models." If all this sounds like futurology one should look at what Google and Microsoft are currently doing. After some years of operating massive powers of parallel computer processing, employing many of the retrieval models and NLP

techniques (briefly described above, and augmented by “page-ranking”, an algorithm to measure the relative importance of web pages, Google now employs another technique called “looping”, which uses information about each individual’s search patterns to “improve” their search results. This means that two people carrying out identical searches at exactly the same time may retrieve different results. This is intelligent manipulation of “big data”, using NLP and other techniques, an approach which is said to be revolutionizing the older approach to artificial intelligence that started out trying to mimic the workings of the human brain. Microsoft is working in the same direction with its powerful system called Watson (named after IBM’s first CEO). Early experiments with IBM’s Deep Blue chess-playing machine which competed with Gary Kasparov, the chess champion prompted an interesting comment from him when he said, “What if, instead of human versus machine, we played as partners?”. After trying that he commented “We could concentrate on strategy planning instead of spending so much time on calculations. Human creativity was even more paramount under these conditions.” Subsequently, Professor Thomas Poggio of MIT has said “These recent achievements have, ironically, underscored the limitations of computer science and artificial intelligence. We do not understand how the brain gives rise to intelligence, nor do we know how to build machines that are as broadly intelligent as we are.” Nevertheless, Watson has enormous power and it has been reported that it can process 500 gigabytes, the equivalent of one million books, per second. In fact, it has been reported (Otake, 2016) that Watson cross-referenced a patient’s records with 20 million research papers and produced a correct diagnosis within ten minutes! While such power is being deployed successfully in such areas it is not yet replacing human decision-making. What can we expect to see by next year’s Ibersid Conference? And what of Information Science and the application of information retrieval and knowledge organization in social systems? Contrary to some appearances, it has not vanished but is now even more diffuse and necessary and, though it is not fully understood or supported by much of academia and employers, is embedded in the applications of very

many organizations, notably on the Web. The first winner in 1999 of the Tony Kent Strix Annual Award for advancing the art and science of information retrieval was Stephen Robertson (one time Professor of Information Science at City University and part-time Research Associate at Microsoft’s Cambridge Laboratory). In 2015 the Award was presented to Dr. Susan Dumais, Senior Researcher at Microsoft.

References

- Gilchrist, Alan (Editor) (1981). *Minis, Micros and Terminals for Libraries and Information Services*. London. Heyden & Son Ltd., on behalf of the British Computer Society/ 1981.
- International Standards Organization (2011). *ISO 25964. Thesauri and Interoperability. Part 1: Thesauri for Information Retrieval*. Geneva: ISO.
- International Standards Organization (2013). *ISO 25964. Thesauri and Interoperability. Part 2: Interoperability with other Vocabularies*. Geneva: ISO.
- Otake, Tomoko (2016). IBM big data used for rapid diagnosis of rare leukemia case in Japan. // *The Japan Times*. Aug. 11, 2016. <https://www.japantimes.co.jp/news/2016/08/11/national/science-health/ibm-big-data-used-for-rapid-diagnosis-of-rare-leukemia-case-in-japan/#.Wyo-wS0rxSN>

Further reading

This does not purport to be an academic study and material for its compilation was drawn from many places, including Wikipedia and other numerous websites. Consequently, it would be tedious and largely unprofitable for the reader to cite them all. There follows a short list of references that might be useful, but idiosyncratic and by no means comprehensive.

For the first section on Computers:

Isaacson, Walter (2014). *The Innovators: how a Group of Hackers, Geniuses and Geeks created the Digital Revolution*. London: Simon and Schuster.

For IR techniques:

Liddy, E.D. (2001-2). A breadth of NLP applications. // *ELSL-NEWS*, 10.4, Winter2001-2 (Newsletter of the European Network in Human Language Technologies).

Manning, Christopher D., Prabhakar Raghavan and Hinrich Schütze (2008). *Introduction to Information Retrieval*. Cambridge. Cambridge: Cambridge University Press.

For human intervention:

Frické, Martin (2012). *Logic and the Organization of Information*. New York: Springer.

Vickery, Brian C. and Alina Vickery (2004). *Information Science in Theory and Practice*. Munich: K.G. Saur.

Enviado: 2017-03-13. Segunda versión: 2018-01-12.
Aceptado: 2018-01-12.

Representación, información, identificación y valoración del libro antiguo: bibliotecarios e investigadores ante un problema común

Portrayal, information, identification and value of rare book: librarians and researchers facing a shared problem

MANUEL JOSÉ PEDRAZA GRACIA

Universidad de Zaragoza, Facultad de Filosofía y Letras, c/ Pedro Cerbuna, 12 (50009 Zaragoza), pedraza@unizar.es

Resumen

La información que ofrece un libro antiguo o un conjunto coherente de libros y documentos antiguos puede ser muy heterogénea y afectar múltiples aspectos. El documentalista, ante el libro antiguo, está obligado a ofrecer la mejor y más pertinente información que pueda obtenerse de él y el investigador tiene la obligación de extraer de y sobre él nueva información para darla a conocer por él mismo a la sociedad. Se pretende determinar la importancia de la relación entre las informaciones que aportan los documentalistas y los investigadores sobre los documentos antiguos de cara a la identificación de la información que ofrecen esos documentos y su puesta en valor desde los centros de información.

Palabras clave: Libros antiguos. Identificación. RDA. Valoración. Bibliotecarios. Investigadores.

Abstract

Rare books, besides their heritage value, have useful information for research. This paper analyses what information librarians should offer to researchers. It differentiates between the information that researchers extract for themselves and that utility that rare books have for libraries. Finally, it establishes the importance for librarians with rare books of carefully balancing the relationship between the information that rare books provide, its identification and valuation.

Keywords: Rare books. Identification. RDA. Valuation. Librarians. Researchers .

1. Introducción

Por mucho que en los tres últimos decenios debido a la evolución de las tecnologías se haya abierto el abanico de las funciones de la profesión (Doucet y Sánchez García, 2008), en esencia, la principal y la que le dio origen es y ha sido proporcionar al usuario potencial un acceso lo más sencillo, adecuado y completo posible a la información que contiene un documento físico o la que proporciona un mediador tecnológico. De hecho, las sucesivas normas de tratamiento de los documentos han incidido siempre en este principio. Pero con respecto a los libros antiguos es preciso partir de una de esas verdades de Pero Grullo, que muy frecuentemente hay que recordar: ni el documento ni la información que puede obtenerse de él están sujetos a caducidad ya que lo que resulta obsoleto para un investigador puede ser de principal interés para otro. Sin embargo, muchos documentos bibliográficos antiguos se han considerado obsoletos desde la perspectiva informativa o, en todo caso, se ha determinado que pueden ser susceptibles de cierta revitalización mediante un proceso de digitalización tendente a su difusión por medios electrónicos, lo que tiene como errónea consecuencia que

se entienda que en el momento que un ejemplar se ha digitalizado, el investigador ya tiene acceso a toda la información que pueden ofrecer todos los ejemplares de la edición a la que pertenece, lo que obviamente es falso. Solo documentos excepcionales, especialmente si son únicos o singulares, se consideran dotados de un halo distintivo por circunstancias que se encuentran en el entorno histórico, económico o patrimonial; y no precisamente por su valor informativo, que sin duda también poseen.

Siendo evidente el valor patrimonial de este tipo de documentos, no se debe olvidar que, además, mantienen su función y valores informativos que deben ser explotados y salvaguardados por los centros en los que se encuentran custodiados; centros que, con gran acierto, se denominan bibliotecas patrimoniales.

El libro, como tecnología robusta, ha sido el principal soporte de información durante siglos. De hecho, los documentos antiguos (impresos y manuscritos, textuales, gráficos, musicales, cartográficos, objetos...) poseen una información muy rica, compleja y continuada en el tiempo. En algunos ejemplares esa información debe desen-

marañarse de entre un complejo conjunto de textos y elementos que individualmente aportan datos de valor con objeto de ser analizados y estudiados —por esta misma causa, al tratar con estos documentos Martín Abad (2007) ha hablado de *Un enredijo de mil y un diablos*—. Y, además, son objetos manufacturados de forma artesanal y mecánica lo que introduce factores de mutación. A todo esto hay que añadir que el ejemplar, no la edición, con el trascurso del tiempo sufre o se beneficia de las sucesivas actuaciones de poseedores, acontecimientos y agentes externos e internos que lo adornan o demeritan, pero que, a su vez, proporcionan informaciones y datos diversos sobre muchos aspectos no necesariamente contemporáneos a la edición ni relacionados con el resto de los ejemplares de la misma.

Por si la complejidad no fuera ya notable, estos documentos antiguos llegan a los centros en los que se custodian por muy diversas causas: herencia, donación o legado, depósito, compra..., y, aunque en ocasiones el interés que ofrecen radica en el propio documento, como unidad, en otras lo hace como conjunto o fondo bibliográfico, puesto que su evolución y desarrollo es un elemento de análisis para el estudio y conocimiento de su creador o poseedor, de un momento en la historia, de una forma de ver el mundo o de una ideología específica...

Este trabajo intenta ofrecer un análisis sobre la información que ofrece un documento bibliográfico antiguo o un conjunto coherente de libros y documentos antiguos, intentando analizar la posición que tiene el documentalista ante el libro antiguo de cara a ofrecer la mejor y más pertinente información y qué información estudia el usuario investigador para darla a conocer por él mismo; finalmente, en la medida de lo posible, se analizan también las relaciones entre las aportaciones de unos y otros y las posibilidades que ofrece esa interacción en el tratamiento de estos documentos.

2. La información (las informaciones) de los libros antiguos

Si se pretende estudiar el nivel de profundidad a la hora de tratar estos documentos o a la de buscarlos para su investigación es preciso definir qué información, o grupo de datos, se pueden obtener de los libros antiguos.

Claramente se pueden diferenciar cuatro niveles: el nivel del contenido puro, el nivel del contenido formal, el nivel de la materialidad, y el nivel de la singularidad.

Pero, es necesario estudiar el grado de dificultad en la obtención de esas informaciones, dificultad

condicionada por las propias características de estos documentos.

2.1. El contenido puro

El nivel del contenido puro ha de entenderse en su doble vertiente gráfica e iconográfica, puesto que, además del texto, la imagen adquiere una notable importancia en el libro antiguo como transmisor de información en una sociedad en la que los iletrados superan con creces a los alfabetizados y la información iconográfica está muy presente como medio de comunicación de las ideas y conocimientos. Además el precio de los recursos de reproducción de imágenes y la dificultad en su elaboración favorece la reiteración de su uso, muchas veces con intencionalidades similares aunque (y esto es lo interesante) no siempre.

El contenido puro en los libros antiguos casi siempre plantea la dificultad de su interpretación puesto que se aborda desde una perspectiva cronológica diferente de la que fue planificado y elaborado. La percepción sobre un fenómeno o el uso del objeto hacen que la intencionalidad y la información sesguen hacia una u otra dirección; y esos sesgos, perfectamente claros para un contemporáneo, sin embargo, pasan desapercibidos cuando se pierde el contexto. Además —el medio es el mensaje— hay que tener en cuenta los posibles contenidos distintos en función de las posibles lecturas diferentes del libro (en voz alta, para uno mismo o para una comunidad —letrada o analfabeta—, o en silencio, o la consulta puntual o la lectura icónica —la que hace el analfabeto de las ilustraciones o, como se diría en lenguaje coloquial, el leer los santos, frase popular pero, no obstante muy reveladora—).

2.2. El contenido formal

El nivel del contenido formal coadyuva a la identificación del documento (al menos, de la obra, la expresión y la manifestación, en terminología FRBR y RDA). Permiten afrontar de manera justificada las identificaciones sustancial (autor y título, esencialmente), característica (edición, editor e impresor) y, en ocasiones, formal (en lo referido a algunos paratextos y sus responsables). La información de este nivel también es trascendente para establecer el llamado «ejemplar ideal» (aquel que responde a la intención del autor, editor e impresor) fundamental para poder catalogar el documento, definición que viene a ser coincidente con la de manifestación. El contenido formal del libro aporta los datos fundamentales para que el investigador acceda al documento y recupere la información o, lo que es lo mismo, los que debe proporcionarle el documentalista o, de la misma manera, los que busca el

investigador. Estos datos en RDA se transcriben tal y como figuran en el ejemplar tanto por lo que respectan a las erratas, las mayúsculas, las fechas o signos de puntuación (RDA 1.7 y 1.8.1), con lo que con el principio fijado en RDA «toma lo que veas» en realidad tiene como consecuencia la facsimilización de determinadas fuentes.

Naturalmente el libro antiguo ofrece la misma información que el libro contemporáneo: autor, título, editor, impresor (que adquiere una especial importancia en este tipo de documentos), fecha y edición. Pero en muchas ocasiones no resulta tan evidente y estos datos, que constituyen el contenido formal del libro, resultan muy difíciles de localizar en (y fuera) del documento (y autenticar):

1. Lo más frecuente, entre las dificultades con las que se encuentra el catalogador, es simplemente que no figuren los datos (en ocasiones para evitar la censura o la persecución legal, no se indiquen por tradición o se hayan perdido por mutilación...). Si se pretende ofrecer esa información al investigador hay que localizarla fuera del documento pero no siempre es posible acceder a ella.

2. Conceptos como los de autor no se entienden como en la actualidad y la responsabilidad puede corresponder a un traductor u ocultarse a causa de los intereses de un editor (Sánchez Sánchez-Serrano, Prieto de la Iglesia, 2011). Y, aunque el autor sea conocido, no siempre es fácil indicar con corrección su nombre. Los apellidos pueden aparecer alterados, hay homonimias, se usan nombres de religión, títulos nobiliarios, monogramas, seudónimos (RDA 9.2.2.8), sobrenombres —cuya representación queda peor reflejada en RDA (9.2.2.18) a entender de quien esto escribe y de la Biblioteca Nacional de España (2015, 11-12)—, etc. Las entidades de la administración en estructuras pretéritas resultan poco conocidas o muy complejas. Los títulos tampoco son imprescindibles cuando las obras (también las expresiones) se conocen por un «alias» que tiene que ver generalmente con el autor, «Summa Angelica», por ejemplo (Fig. 1).

3. Tampoco el concepto de editor está fijado en ese momento por lo que el nombre del editor puede esconderse tras un mecenas, un librero o una institución. El editor puede ser el propio autor, el impresor (RDA 2.8.4.1) o una entidad sin mayor interés por que esa función figure entre los datos que se ofrecen. Este fenómeno propio de estos documentos ha dotado de especial importancia al conocimiento de estos datos de impresión.

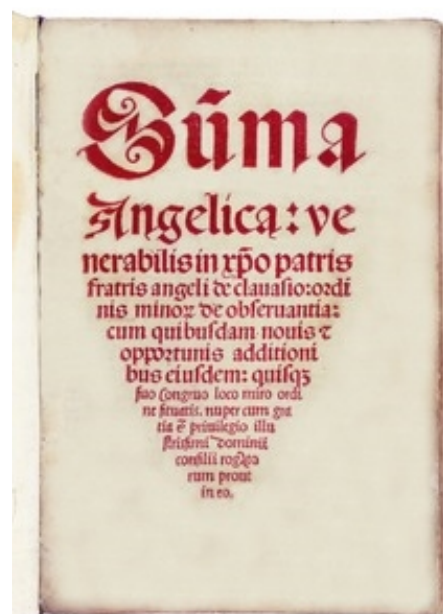


Figura 1. Carletti de Chivasso, Angelo. *Suma Angelica*. Venetij: impressa per Alexandrum de Paganinis, 1511 die VIII Marcii

4. No son infrecuentes los impresos *sine notis* y existen ediciones falsificadas, contrahechas (Fig. 2) y piratas (Moll, 1979) que proporcionan datos que pueden distar mucho de la realidad.



Figura 2. Edición princeps de 1617 y contrahecha (en torno a 1668) de la obra de Miguel de Cervantes, *Los trabajos de Persiles y Sigismunda* pero con el mismo pie de imprenta

5. Los ejemplares que se consideran pertenecientes a una misma edición no tienen por qué ser idénticos. Cualquier variable que pueda concebirse en la elaboración de un objeto complicado con múltiples copias como es el libro se habrá producido, sin ninguna duda, en algún (o más de algún) momento y tendrá como resultado la

percepción de una ausencia, diferencia o presencia extraña en uno o en un conjunto de específico de ejemplares, o en una edición concreta. Además, el éxito o la demanda de una obra puede dar lugar a distintas ediciones con el mismo pie de imprenta, sin que esta circunstancia, la de tratarse de ediciones diferentes, se ponga de manifiesto en la fuente, como ocurre con la primera y la segunda edición del *Quijote* (En Madrid, Por luan de la Cuesta, Vendese en casa de Francisco de Robles, 1605). Hay otros muchos problemas como la intervención de varios actores en la edición, en varios momentos, cambios de portada...

Gran parte de este cúmulo de dificultades y carencias se puede solventar, al menos en parte, mediante el uso de fuentes de referencia (RDA 2.2.4) o siguiendo una serie de procedimientos identificativos de cierto nivel técnico perfectamente conocidos por los profesionales.

2.3. La materialidad

Mediante el análisis de los elementos materiales del objeto libro se puede acceder a la identificación de los responsables de la elaboración del libro y su encuadernación (en la terminología establecida por FRBR y RDA, hacen referencia a la manifestación y al ítem). La identificación de esos elementos materiales también facilita la determinación del «ejemplar ideal». De hecho, los formatos y la estructura del libro son también esenciales a la hora de encuadrar el ejemplar en una edición o variante (RDA 3.21.2.9). Los estudios referidos a aspectos específicos tienen este nivel informativo como base: encuadernación, ilustración, tipografía, papel, entre otros.

2.4. El ejemplar singular

El nivel de la singularidad ofrece información exclusiva del «ejemplar singular» (ítem, según FRBR y RDA). Estos datos permiten, conjuntamente con la identificación de la integridad documental, la identificación singular y son esenciales de cara a la catalogación con objeto de ofrecer al usuario las características propias del ejemplar o ítem, ya que puede ser información de principal importancia para investigaciones relacionadas con los poseedores, con la historia de la lectura y la encuadernación... Además, el cúmulo de anotaciones manuscritas que pueden aparecer tanto sobre otras obras manuscritas como sobre obras impresas ofrecen información en torno a las ideologías (censura y anotaciones de lector), comprensión y relación con otras obras, especialmente sobre procedencia y, en ocasiones, sobre precios.

3. Descripción e identificación para la localización

El bibliotecario adquiere el compromiso de la descripción de los documentos y el de su localización entre un conjunto de documentos, esto es, la de representarlo de manera sintética dando como resultado un asiento o registro bibliográfico. «Los registros bibliográficos sencillamente representan e individualizan a los ejemplares singulares y permiten su localización en un determinado depósito bibliotecario. La localización facilita su consulta» (Martín Abad, Becedas, Lilao, 2008, 13). El registro es un producto bibliográfico derivado de dos acciones diferentes: la descripción, cuyo producto es una reseña bibliográfica, y la extracción y construcción de puntos de acceso preferidos (autorizados o normalizados) que se han denominado tradicionalmente encabezamientos.

Resulta evidente que la descripción del documento se hace a partir de lo que el documentalista-observador aprecia en él. Esta afirmación se hace más patente en el «principio de representación» sustentado por RDA. Esta forma de proceder, que en el trabajo cotidiano con los documentos contemporáneos ofrece resultados óptimos, puede no ofrecerlos de igual manera con los documentos antiguos. En efecto, de esta manera se ofrece la descripción del documento-ejemplar o «ejemplar singular» (ítem); pero no siempre, como ya ha quedado patente, lo que el ejemplar ofrece es suficiente para describir el documento-edición o «ejemplar ideal» (manifestación). Se ha de tener presente que en los manuscritos manifestación e ítem se confunden, pero en el mundo del impreso antiguo son fenómenos muy diferentes.

Una reseña bibliográfica es un documento que debe representar a todos los integrantes de la edición y, por eso mismo, resulta válido para cualquiera de ellos. No ocurre lo mismo con el registro que debe incluir también los datos característicos del ítem de cara a su individualización para su localización. Pero además es preciso previamente identificar el documento como perteneciente a un conjunto dado (de obras de un autor, de una edición de una colección específica) mediante los puntos de acceso que aportan información de la obra y la expresión. El catalogador se encuentra con el problema de la propia tipología documental. Como dice Martín Abad (2004) «difiere totalmente la problemática que plantea la identificación de un manuscrito de la que plantea la de un impreso antiguo. El catalogador y el bibliógrafo deben identificar correctamente la pieza que tienen a la vista. Esta inevitable tarea es siempre previa a la construcción de

cualquier tipo de registro bibliográfico de un catálogo o un repertorio bibliográfico retrospectivo». Cada manuscrito, por definición, es único. En el manuscrito se trata de identificar su singularidad absoluta o cierta «multiplicidad», si es copia. Además, la identificación de los elementos codicológicos permite adscribir el documento a un *scriptorium*, a un taller o a una colección.

Para ser útiles, los datos de la reseña bibliográfica realizada por el documentalista deben, entre otros requisitos permitir al usuario (RDA 0.4.2.1):

- Encontrar recursos que correspondan a los criterios de búsqueda establecidos por el usuario.
- Encontrar todos los recursos que materialicen una obra o una expresión concreta de esa obra.
- Encontrar todos los recursos asociados con una persona, una familia o una entidad corporativa específica.
- Encontrar las personas, las familias y las entidades corporativas que corresponden a los criterios de búsqueda establecidos por el usuario.
- Identificar el recurso descrito (esto es, confirmar que el recurso descrito corresponde al recurso buscado o distinguir entre dos o más recursos con las mismas o similares características).
- Identificar la persona, la familia o la entidad corporativa representada por los datos (esto es, confirmar que la entidad descrita corresponde a la entidad buscada o distinguir entre dos o más entidades con el mismo nombre o con nombres similares, etc.).

Para conseguir estos requisitos los datos se pueden extraer de fuentes externas al recurso con objeto de corregir o aclarar las representaciones dudosas, ambiguas, ininteligibles o engañosas que se extraigan del recurso (RDA 0.4.3.5). Esta posibilidad adquiere mucha importancia cuando se trata de documentos antiguos cuyos datos no figuren en el propio recurso o se hallan mutado intencionadamente (RDA 2.2.4).

La identificación es esencial y como se dijo en otro lugar (Pedraza 2010, 51): «la identificación es múltiple ya que no solamente hay que reconocer el contenido; hay que examinar también el resto de los factores que constituyen el documento, dando lugar a diversas identificaciones (del impresor o del editor, del encuadernador...) entre las que no es desdeñable la identificación de la integridad documental, bien sea por registrar carencias con respecto a su edición o variante, ser un mutilo, o por estar compuesto por

más de un ejemplar conformando un volumen facticio».

En consecuencia, por mucha trascendencia que RDA otorgue al principio: «toma lo que ves» o «acepta lo que tienes», el bibliotecario no puede circunscribirse ahora, como no la hacía antes, únicamente al «ejemplar singular» que tiene en las manos para conseguir que el usuario encuentre e identifique el documento que busca (que no deja de ser un documento que existió, quizás, pero que es posible que ya no exista). Es necesario recurrir a las fuentes de información o a actuaciones técnicas específicas para identificar datos inexistentes, dudosos o falsos. Esas actuaciones conducen a la elaboración de una reseña que representa el «ejemplar ideal», o lo que es lo mismo, una reseña que permita reflejar cualquier ejemplar de una misma edición (manifestación). Y para ello hay que tener en cuenta que «la necesidad de examinar atentamente un ejemplar singular no descarta la (inevitable) necesidad de comparar más de uno de la misma edición» (Martín Abad, Becedas, Lilao, 2008, 18). Sin olvidar, además, las características propias del ejemplar singular o ítem que deben quedar reflejadas, pero perfectamente independizadas y separadas de las propias de ese «ejemplar ideal».

Por consiguiente, en el proceso de catalogación, también desde la filosofía RDA, es imprescindible realizar una identificación correcta mediante procesos de consulta de fuentes y análisis del documento que, a diferencia del libro contemporáneo, son actuaciones concurrentes con labores propias del investigador. El bibliotecario que trabaja con este tipo de fondos ha de ofrecer una información fidedigna que se debe derivar de la correcta identificación de los datos. Y es que el registro debe reflejar la representación que el documento hace de sí mismo, ha de satisfacer al usuario por lo que respecta a la selección del documento apropiado y, para ello, es esencial diferenciar el recurso (documento) de otros (RDA 1.2). De lo contrario el bibliotecario puede crear fantasmas o ediciones imaginarias: entes bibliográficos que no han existido ni se pretendieron realizar y que conducen a que el usuario pierda la posibilidad de encontrar el ejemplar que busca. La dificultad varía mucho a la hora de determinar ciertos datos, pero la información en este ámbito debe ser lo más exhaustiva y correcta posible, de la misma manera que lo es para el tratamiento del libro moderno. Tal es así que se roza en este extremo la línea deontológica: el descuido en la realización de la reseña bibliográfica y la incorrecta identificación pueden perfectamente llegar a ser consideradas como mala praxis.

Pero qué se debe identificar, describir y localizar. En una obra compleja como es un libro antiguo:

¿hay que dar el paso siguiente para ofrecer información que puede resultar trascendente? En primer lugar, el autor y el título, (la obra) lo que condiciona en buena medida la materia. Tras la obra, la expresión, que condiciona otras responsabilidades y la forma de la materia.

Después, la manifestación, que presenta en este tipo de materiales algunas peculiaridades. El libro impreso antiguo puede incluir junto al texto principal, que lo caracteriza, un conjunto de textos y recursos gráficos que le preceden y le siguen. Estos textos e imágenes, paratextos, pueden poseer diferentes funciones (Reyes, 2003) y responden a diversas justificaciones. Los paratextos independientemente de su tipología son fuente principal de identificación de la manifestación, por consiguiente, fuente para la identificación y útil para la investigación, pero, a no ser que se consideren de excepcional valor (informativo, histórico, artístico, literario...) no son incorporados a la reseña bibliográfica. Si su valor fuese extraordinario podría emplearse una descripción analítica (RDA 1.1.4 y 1.5.3). En este mismo ámbito de la manifestación hay una serie de datos que se obtienen en el entorno material que afectan, en primer lugar, al soporte empleado, y, en segundo, a las grafías y los tipos con los que se confecciona el sistema gráfico y a las miniaturas, dibujos, tacos o planchas con los que se elabora el sistema iconográfico del libro. Todos ellos ofrecen información que puede ser relevante (García, 2008). Generalmente el registro bibliográfico indica la presencia de ilustraciones y es posible, si se considera necesario hacer mención a la tipografía (RDA 3.21.2.9).

Finalmente, el ítem, el ejemplar singular, independientemente de la edición a la que pertenezca, cuenta una historia que habla de aspectos que conformaron su entorno y que son de interés básico para determinadas investigaciones. Los *ex libris*, las anotaciones manuscritas, el estado del ejemplar, la censura, los testigos, la incorporación de elementos ajenos a la edición del libro... son esenciales para enmarcar el ítem, para contextualizarlo. La identificación de los aspectos que individualizan los ejemplares no puede, de ninguna manera, ser obviada y debe conducir a su descripción, puesto que el contexto es también información. Estos datos e indicaciones son fundamentales para el investigador y el bibliotecario debe proporcionarlos. La encuadernación (cuando supera la clásica encuadernación de pergamino flexible tan presente en los libros españoles antiguos) es otro de los aspectos que debe quedar reflejado. Es cierto que se ha llegado a pensar que el tratamiento de la encuadernación pudiera ser independiente de la del propio libro tratándola como documento independiente,

pero esta posición queda descartada cuando se entiende que la encuadernación (preindustrial) ha sido preparada en general *ex profeso* para un ejemplar específico, incluso cuando un ejemplar ha sido reencuadernado cientos de años después del momento en el que se confeccionó se habrá tenido en cuenta el ejemplar concreto que protege. Todo ello a pesar de que para que exista libro no es imprescindible la encuadernación y que el mismo ejemplar ha podido ser encuadernado un número indeterminado de veces y se ha podido conservar más de una de ellas. Una vez solventada su función de preservación, la encuadernación posee valores artísticos pero también posee valores informativos sobre el poseedor (super libros, motivos heráldicos, monogramas, inscripciones en los cortes...) y sobre ella misma (hierros o planchas, firma del encuadernador...). Los valores artísticos pueden o no coincidir con los que el ejemplar posee, como ya se ha dicho, debido a la posibilidad de no ser contemporáneos, pero la encuadernación debe ser descrita e identificados los motivos emblemáticos que indiquen quien fue su propietario ya que permiten establecer relaciones entre el ejemplar, sus intervenciones y sus poseedores (particulares o institucionales). Con mucha más razón se han de indicar las referencias a los poseedores, mediante menciones textuales o gráficas. Las anotaciones y la información que incluyen los *ex libris* y otras marcas de propiedad (Carreño, 2008) son otro elemento que debe quedar expresado puesto que es prueba determinante para la elaboración de la historia del ejemplar y su relación con la biblioteca en la que se encuentra y que, normalmente, es la que proporciona la información. Otros elementos como las anotaciones y la censura también deben ser mencionados para que el investigador posea datos específicos sobre las posibilidades de uso del ejemplar para los fines que persigue. Algo similar ocurre con el estado de conservación y si éste condiciona su disponibilidad puesto que un ejemplar digitalizado o reproducido puede carecer de interés para una investigación en concreto, por mucho que determinados contenidos se hayan salvaguardado. Incluso el grado y el tipo de deterioro debe ser considerado como fuente de información (Sánchez Herrador, Asencio y Carrasco 2010). Las anteriores firmas que el ejemplar ha tenido en la biblioteca o en las bibliotecas de las que procede son también menciones que no pueden obviarse, puesto que pueden dar lugar a la existencia de dobles identificaciones del mismo ítem. Existen también otras muchas intervenciones en los libros antiguos que se indicarán a efectos de individualización.

Pero aunque el bibliotecario no siempre pueda llegar a desplegar todo el abanico de informaciones posibles que el manuscrito o impreso ofrece debe saber adentrarse en lo que precisa el investigador si aspira a cumplir con la labor que tienen encomendada por la sociedad de forma eficaz.

4. Localización para su identificación y descripción

El investigador (y el bibliotecario también lo es a estos efectos), cuando se enfrenta a un documento antiguo puede hacerlo desde muy diversos puntos de vista.

Para el investigador del manuscrito el interés generalmente radica en la obra o en la copia (manifestación/ítem) y en menor medida en la morfología (codicología). Para el investigador del impreso, bien esté interesado en el texto o en el producto bibliográfico es esencial la identificación de la edición (manifestación) y para él los datos que se encuentran en torno a ella poseen un interés principal ayudan al estudio de los textos y las diversas modificaciones de los mismos o su proximidad a la obra (también en el sentido FRBR y RDA) o una expresión tal y como la concibió su autor a la par que pueden explicar la génesis del conjunto de ejemplares al que pertenece el ítem en cuestión. Otros investigadores se concentran en infinidad de aspectos: las mentalidades, la ilustración, la biografía (del autor del texto o los paratextos, del poseedor, del impresor...), la encuadernación, etc. Los hay, además, que se centran en la materialidad. Esta información, en tanto pueda ser aportada por el bibliotecario, es trascendente ya que individualiza y parcela aspectos del documento que el investigador debe tener en cuenta a la hora de establecer una selección de su interés a través de las diferencias y semejanzas que proporcionan los registros bibliográficos.

Algunos paratextos poseen valor por sí mismos puesto que son pequeñas joyas firmadas por primeras filas de la literatura o del arte universal que no pueden pasar desapercibidas.

En el entorno material del libro se obtienen también datos de interés. Los investigadores del papel empleando las marcas identificativas y estructurales del mismo establecen secuencias en la utilización del tipo de soporte. El análisis de la impaginación del manuscrito, junto con el sistema de marcado, son también esenciales para la identificación, por analogía, de los *scriptoria* y talleres. Estos aspectos relacionados con el soporte entran de lleno en el ámbito del trabajo de investigación y pueden ser empleados con el fin de avalar o confirmar datos no expresados de forma explícita. Las características del soporte

también pueden ser expresados en la descripción del ítem (RDA 3.22.1.3).

La caligrafía empleada en un manuscrito permite ayudar a datar y a adscribir su confección a un ámbito geográfico aproximado. Este mismo análisis puede resultar útil en el caso de anotaciones realizadas en libros impresos.

La identificación tipográfica es relevante en numerosos aspectos que pueden ayudar a un mejor conocimiento de la obra o de la edición. Los tipos empleados —bien sean tipos metálicos o xilográficos para imprimir letras, o tacos xilográficos y planchas metálicas para imprimir imágenes (Fernández Valladares 2012)— y los usos de composición pueden ayudar a conocer el taller en el que se compuso la edición. De nuevo se trata de datos de gran utilidad para el tratamiento de impresos *sine notis*. Las marcas tipográficas son para este fin perfectas de cara a la identificación y confirmación del editor y del impresor, aunque resulte infrecuente que se omitan indicaciones explícitas textuales cuando se incluyen.

Es muy difícil establecer e identificar los autores responsables de la creación de los tipos, pero las planchas metálicas y los tacos xilográficos empleados, en menor medida, pueden estar firmados, lo que aporta otra información de validez para los investigadores de este tipo de recursos, de su trasmisión y de la evolución de las ideas estéticas. En el caso de la iluminación (del manuscrito y del impreso) este tipo de información es prácticamente inexistente o muy difícil de localizar ya que son extraordinarias las ilustraciones firmadas. Algunas de estas circunstancias relacionadas con la ilustración deben ser tenidas en cuenta y, en la medida de lo posible, descritas, especialmente la presencia de láminas y la presencia de escudos tipográficos. Las ilustraciones más relevantes deberían ser descritas y relacionados sus responsables. Por último, la identificación de las tipografías (de validez contrastada en los impresos más antiguos) debe dejarse al campo de la investigación bibliográfica. Los recursos gráficos se confeccionaban originalmente para ser utilizados en una edición específica y forman parte integrante de la misma, aunque puedan ser empleados en otras producciones posteriores.

Por lo que se refiere al ítem, partiendo del hecho de que la encuadernación suele ser propia del ejemplar, de ella se puede analizar desde los muy diversos sistemas de cosido, hasta el tipo de material empleado en guardas y tapas, sin olvidar los materiales usados en su confección, la decoración de esos materiales y la de los cantos y guardas, entre otros aspectos. Quizás un problema fundamental es que se suelen presentar

de forma anónima ya que son muy pocas las que se firman y, con frecuencia, son las más modernas. En cualquier caso, se constituyen en documentos trascendentes por ellos mismos (aunque sin desligarse del libro) que poseen valores notables —se han llegado a constituir colecciones de encuadernaciones, con o sin libro— y no son demasiado extraños los casos de *remboîtage* (reutilización de una encuadernación confeccionada para un libro en otro diferente). Por esta causa, es preciso proporcionar una información que describa la encuadernación y todos aquellos aspectos que la enmarcan: guardas, tipo de material, decoración... Gran parte de los aspectos relacionados con la materialidad se pueden incluir dentro de la singularidad: la encuadernación editorial es muy infrecuente en el libro antiguo, especialmente en el libro hispano, y la coloración de grabados xilográficos no suele ser decisión editorial, por ejemplo. La encuadernación puede ser descrita en nota de ejemplar (RDA 3.22.1.4).

Los *ex libris*, las anotaciones manuscritas, el estado del ejemplar, la censura, los testigos, la incorporación de elementos ajenos a la edición del libro... son atributos que trascienden al propio ejemplar de cara a su investigación, aunque pueden ser descritos en nota de ejemplar (RDA 3.22.1.4). La presencia de datos de antiguos poseedores, por ejemplo, aporta una información que puede definir no solo la presencia de un libro en unas manos concretas, sino también los intereses de un personaje en un momento específico. Algo parecido ocurre con las anotaciones. La censura ofrece información sobre las mentalidades en un momento específico...

Se han especificado algunas informaciones cuya identificación y explotación corresponde al investigador, pero no obsta para que cualquier información obtenida por esta vía deba relacionarse con el documento que la generó. Se tiene que tener presente que la investigación es la fuente sobre el propio documento menos explotada y, sin embargo, la que más datos fidedignos y contrastados aporta.

El investigador, mediante la difusión de los conocimientos que obtiene en su trabajo aporta datos identificativos cuyo acceso supera en ocasiones las posibilidades que el documentalista posee para identificar el documento.

5. Los libros antiguos como conjunto: su identificación

Como conjunto estructurado y homogéneo los libros antiguos (bibliotecas y archivos personales) también aportan información tremendamente valiosa. Los documentos que han acompañado las circunstancias vitales de un individuo, familia o

institución son fuente esencial para conocer y comprender a ese individuo, a esa familia o a esa institución. Los documentos patrimoniales bibliográficos o de otro tipo poseen una información que resulta matizada por el contexto; es decir el contexto en el que se encuentra un documento cualquiera aporta también información sobre él mismo. Por consiguiente, cuando un documento se desgaja del conjunto al que pertenece se pierden dos tipos de informaciones: las que el conjunto, colección, fondo... aporta al documento (que ha quedado amputado) y el significado del documento en ese conjunto específico (por cierto, algo muy parecido al expolio arqueológico). El conjunto (la biblioteca o fondo) se constituye así en una unidad que debe ser considerada como tal, un documento patrimonial único, sin obviar que también puedan serlo cada una de sus unidades o partes integrantes. Por tanto, la simple dispersión de estos conjuntos impide este tipo de investigaciones.

Generalmente lo que se denomina papeles o archivo personal viene acompañado de una biblioteca personal y de algunos objetos de especial significación entre los que existen numerosos nexos comunes: cartas en las que un autor envía un libro a otro; facturas de adquisición de libros, diplomas de entrega de un placa o condecoración, un objeto de medición científica y su manual de manejo o libros que indican su manejo, árboles genealógicos y ejecutorias de hidalguía... Los puntos de acceso de familias, resuelven algunos de estos casos que trascienden a una única persona (RDA 10).

Resulta evidente que la dispersión de estos conjuntos bien sea por unidades bibliográficas o documentales o por bloques pretendidamente coherentes (por contenido o tipología) lo único que consigue es destruir la información que el fondo o colección ofrece sobre quienes les han dado origen y su recuperación resulta, por mucho que esas actuaciones se hayan documentado, casi imposible. Se argumenta que es imposible mantener su unidad en función de la existencia de duplicados, falta de espacio... Es cierto, pero no puede olvidarse que el centro de interés cambia por completo hacia el quién, el cómo, el cuándo y el por qué se ha constituido ese fondo cuando era una colección. O, hacia el por qué este documento concreto está integrado en él. El conjunto documental se transforma en un documento complejo y único que no se puede ni debe mutilar mediante el expurgo de las partes que lo componen. La información sobre estos fondos debe ser proporcionada de la misma manera que se proporciona información sobre cada una de las unidades que los integran mediante instrumentos descriptivos e identificativos apropiados.

Es cierto que puede haber información que se duplica si se atiende solamente a la parte de la información susceptible de ser ofrecida a los usuarios desde las perspectivas de contenido y el contenido formal (aunque solo en parte). Ya se ha dicho que los ejemplares de la misma edición devienen en objetos informativos y patrimoniales completamente diferentes con el tiempo, el uso, los propietarios, y su conservación. Se obtienen, por esta causa, documentos singulares y, a la par, un conjunto coherente que aporta otros valores e intereses distintos que la simple suma de sus unidades componentes.

6. Información y valoración

El bibliotecario tiene la obligación patrimonial de dar a conocer el documento o, si se quiere desde la perspectiva de los derechos, el ciudadano tiene el derecho de conocer los documentos que conforman el patrimonio bibliográfico y documental (Abad, 2013). El bibliotecario-documentalista debe proporcionar la información necesaria para acceder a todos los contenidos que posee el documento (o puedan obtenerse de él) con objeto de ofrecer acceso al documento independientemente de la faceta que interese al investigador o por la que pueda pretender acceder, cumpliendo así el objetivo del documento patrimonial.

El investigador, por su parte, tiene la obligación social de ofrecer nuevas perspectivas sobre hechos conocidos y avanzar en el conocimiento de acontecimientos, objetos, personas... proporcionando aspectos, facetas, informaciones desconocidas sobre los documentos.

Como ha quedado de manifiesto, la investigación aborda el documento antiguo desde muy diversas perspectivas. Entre ellas las hay que podrían calificarse de fundamentales incidiendo en la obra, la expresión y la manifestación y las hay que podrían calificarse de singulares, incidiendo en el ítem. Se produce de esta manera una simbiosis entre el investigador y el documentalista en torno a las labores de tratamiento del documento. Hay una especie de retroalimentación sobre el descubrimiento de algunas (nuevas) facetas en torno al documento que tienen como resultado la posibilidad de ofrecer una descripción y una identificación más adecuadas del mismo o de aspectos específicos referidos al ítem: poseedores, encuadernadores, ausencias, presencias...

Un mejor conocimiento del documento (manifestación o ítem) tiene como consecuencia la ampliación del abanico de los posibles intereses en él, lo que, a su vez, genera más investigación y conocimiento cerrándose de esta manera un círculo virtuoso de fundamental trascendencia en

la que de forma simbiótica se benefician las descripciones e identificaciones documentales (Fig. 3). En cualquiera de los casos, es de trascendental importancia poner de manifiesto que el desconocimiento de esta información aportada por la investigación cierra la posibilidad de añadir valor a la documentación que se custodia en las bibliotecas. En este punto parece preciso recordar la definición que de expolio hace la legislación vigente «... se entiende por expoliación toda acción u omisión que ponga en peligro de pérdida o destrucción todos o alguno de los valores de los bienes que integran el Patrimonio Histórico Español, o perturbe el cumplimiento de su función social.» (LPHE, 1985, art. 24).



Figura 3. Importancia de la investigación de cara a la valoración

Por el contrario, la difusión de la información (descripción e identificación) localizada produce conocimiento, con lo que se cumple con el principio derivado de la legislación patrimonial. El conocimiento tiene como consecuencia el aumento del uso del documento (localizado e individualizado) y el mejor conocimiento da lugar a una mayor investigación (descripción e identificación) sobre el documento. Esa investigación aporta valores añadidos a los documentos y a los fondos que deben también ser difundidos.

Las nuevas aportaciones de los investigadores se transforman en fuentes informativas relevantes (avaladas) para el documentalista-bibliotecario puesto que profundizan en los aspectos identificativos que se iniciaron en el proceso de catalogación.

Puede decirse que el documentalista-bibliotecario ofrece un panorama identificativo general que pueden concretar los investigadores: la identificación de una variante, del editor oculto, de la fecha de publicación, del nombre del encuadernador, la concreción de la biografía del autor o del poseedor..., por ejemplo, son puntos específicos a los que el documentalista-bibliotecario no siempre ha podido llegar mediante los procedimientos

técnicos de los que dispone pero que son útiles a la hora de proporcionar una identificación más precisa del conjunto documental, de la obra, de la expresión, de la manifestación o del ítem.

La dualidad que puede presentar el documento (singular/integrante de un conjunto) opera en contra de soluciones sencillas descriptivas que deben suplirse con herramientas que se crean y mantienen desde las bibliotecas patrimoniales con objeto de realizar búsquedas más precisas sobre aspectos o facetas concretos: bases de datos de marcas de impresor y editor (Universitat de Barcelona, 1998-), de florones (University of Cambridge, 2013-), de iluminaciones o miniaturas (Institut de Recherche et d'Histoire des Textes, 1979-), de antiguos poseedores (Universidad de Salamanca. Biblioteca General Histórica, 2012-), de encuadernaciones (British Library, ca. 2010-), etc. En otras ocasiones son los investigadores los que las crean: sobre encuadernaciones (Bibliopégia, 2006-), filigranas (Bernstein, ca. 2008-), etc.

Los investigadores que realizan trabajos bibliográficos también confeccionan noticias bibliográficas muy precisas que inciden y profundizan en aspectos representados en las reseñas y registros bibliográficos de los documentalistas. No obstante, RDA abre la puerta para establecer colaboraciones entre los dos extremos del trabajo sobre el libro antiguo. Estas posibilidades de interacción se ven favorecidas desde RDA que tiene como fundamento el facilitar que cualquiera pueda, con una pequeña herramienta de comunicación (interfaz) realizar completamente una noticia bibliográfica o concretar una existente que se puede convertir sin dificultad en un registro bibliográfico. De esta manera se perfecciona esta relación entre documentalista e investigador.

La relación entre la noticia bibliográfica del investigador y el registro bibliográfico del documentalista por medio de registros RDA es una vía que debe explorarse y que la nueva norma facilita cuando permite acomodar la misma información en estructuras diferentes. Todo ello, preferiblemente mediante el establecimiento de relaciones entre unos y otros productos bibliográficos. Pero esta relación puede poseer una segunda vía desde la noticia bibliográfica del investigador también se pueden establecer enlaces con los registros bibliográficos de las bibliotecas en las que es posible encontrar ejemplares, de la misma manera que las bibliotecas mantienen enlaces hacia las digitalizaciones de los documentos catalogados. El resultado sería la obtención de registros bibliográficos enlazados tanto con el documento digitalizado como con las investigaciones realizadas sobre él que han perfeccionado el registro (Fig. 4).

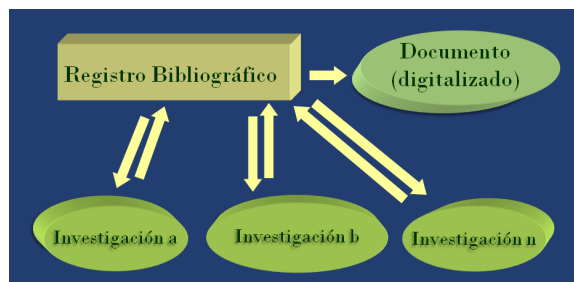


Figura 4. Relación entre el registro bibliográfico, el documento digitalizado y la investigación de cara al acceso y la valoración

En suma, si la premisa que defiende este trabajo: la relación entre el documentalista-bibliotecario y el investigador colaborativa y simbiótica ofrece un mayor beneficio que si ambos mantienen barreras rígidas de separación. La correcta relación valora el documento y un documento con valor añadido aporta patrimonialmente más informaciones y es más útil, por tanto más valorado.

Pero ese nuevo valor, debe ser, primero, difundido y en segundo lugar para hacerlo más útil se debe relacionar con el documento y el registro bibliográfico. Hay que buscar, por tanto, mecanismos para poder alcanzar esta situación óptima de cara al rendimiento eficaz del trabajo de los investigadores y de los bibliotecarios-documentalistas.

7. A modo de conclusión

Se deduce de lo anteriormente expuesto, en primer lugar, que la información que pueden proporcionar los libros y bibliotecas antiguos es mucho más que la evidente procedente de la portada y el colofón y, en segundo lugar, que los investigadores que trabajan con este tipo de documentos buscan informaciones muy variadas sobre los mismos. Esto es, la investigación con y sobre los libros antiguos, además de la información que se extrae de la propia explotación de los contenidos, posee muchas más vertientes: el estudio histórico del libro (edición, técnicas, impresión...), el estudio artístico (ilustración, encuadernación, tipografía), el estudio bibliográfico, el estudio de los materiales (papel, grabado...), el estudio biográfico, el estudio de colecciones, etc.

El investigador debe encontrar en la información que proporcione el centro una serie de datos identificativos que afectan a todos los niveles posibles, pero que tienen una característica clave: deben ser fidedignos y no deben dejar cabida al error de identificación. A la par, el documentalista debe encontrar en las investigaciones publicadas información, también fidedigna, susceptible de coadyuvar a una mejor identificación y descrip-

ción de los documentos. Se produce de esta manera un cruce de compromisos del que se pueden obtener beneficios mutuos y generales.

Las reseñas y registros bibliográficos, productos intrínsecamente relacionados, son hasta la fecha responsabilidad única de los documentalistas, pero pueden y, en la medida de lo posible, deben ser, mediante el establecimiento de las relaciones oportunas, completados por noticias bibliográficas procedentes del campo de la investigación bibliográfica.

Existe un grupo de datos que proporcionan los documentos bibliográficos antiguos que pueden ser ofrecidos por parte de los centros mediante recursos especiales impresos o electrónicos referentes a poseedores, ilustradores, presencia de escudos o marcas tipográficas, anotaciones manuscritas... Pero estas informaciones no se pueden obviar en el tratamiento común de los documentos.

Hay un tercer grupo de datos e informaciones que únicamente se pueden hacer llegar mediante la relación del documento con los estudios llevados a cabo por otros investigadores. Estos estudios deberán tenerse en cuenta y se debe proporcionar información sobre ellos para que aporten otros valores al documento.

Estas investigaciones tienen como consecuencia una puesta en valor del documento debida tanto a la profundización en su conocimiento y al nuevo que se podría aportar sobre él. Estos aspectos tienen siempre otras consecuencias: a) el perfeccionamiento de la identificación de la obra, la expresión, la manifestación y el ítem; b) un mayor interés de los investigadores por el documento o el fondo antiguo en cuestión puesto que se amplían los campos y niveles de investigación; c) nuevos valores de los documentos como objetos patrimoniales susceptibles de ser ofrecidos a la ciudadanía; y d) Mayor proximidad con el ciudadano, cumpliendo con el mandato legislativo, que puede encontrar en las nuevas aportaciones más motivos de interés por los documentos y las bibliotecas patrimoniales.

Referencias

- Abad Hiraldo, Ramón (2013). *Bibliotecas y patrimonio: colección de normas y recomendaciones sobre el acceso, tratamiento y uso de las colecciones de fondo histórico en las bibliotecas españolas*. Madrid: ANABAD, 2013.
- Bernstein, ca. 2008-; Bernstein: the memory of paper. Viena: Austrian Academy of Sciences, ca.2008-. (Consulta: 2017-08-17) http://www.memoryofpaper.eu/BernsteinPortal/app1_start_disp#
- Bibliopégia (2006-); Catálogo colectivo de encuadernaciones artísticas. Madrid: Universidad Complutense, 2006-. (<http://freja.ccdoc.ucm.es:30080/index.php> (2017-08-17)).

- Biblioteca Nacional de España (2015). Grupo de Trabajo sobre RDA. Análisis de diferencias entre RDA, Reglas de catalogación e ISBD. Madrid: Biblioteca Nacional de España, 2015.
- British Library (ca. 2010-); Database of bookbindings. London: British Library, ca. 2010-. (Consulta: 2017-08-17) <https://www.bl.uk/catalogues/bookbindings/>
- Carreño Velázquez, Elvia (2008). *Marcas de propiedad en los libros novohispanos*. México: Apoyo al Desarrollo de Archivos y Bibliotecas de México, 2008.
- Doucet, Anne-Vinciane; Sánchez García, María José (2008). La gestión del conocimiento como estrategia en el trabajo del documentalista // Scire. ISSN 1135-3716 14:1 (en.-jun. 2008) 75-86.
- Fernández Valladares, Mercedes (2012). *Biblioiconografía y literatura popular impresa: la ilustración de los pliegos sueltos burgaleses (o de babuines y estampas celestinescas)*, eHumanista: Journal of Iberian Studies. ISSN 1540-5877 21 (2012). 87-131.
- García, Idalia (2008). Entre páginas de libros antiguos: la descripción bibliográfica material en México. // Investigación bibliotecológica ISSN 0187-358X 22:45 (2008) 13-40.
- Institut de Recherche et d'Histoire des Textes (1979-); Enluminures. Paris: Institut de recherche et d'histoire des textes, 1979-. <http://www.enluminures.culture.fr/documentation/enluminures/fr/> (2017-08-17).
- LPHE (1985); Ley 16/1985, de 25 de junio, del Patrimonio Histórico Español. Madrid: Boletín Oficial del Estado, 1995-. <https://www.boe.es/buscar/pdf/1985/BOE-A-1985-12534-consolidado.pdf> (2017-08-17).
- Martín Abad, Julián (2007). *Un enredijo de mil y un diablos*. Madrid: Ollero y Ramos, 2007.
- Marín Abad, Julián (2004). *La valoración del libro. El punto de vista del bibliotecario de fondo antiguo*. Madrid: Universidad Complutense de Madrid, 2004. (Consulta: 2017-08-17) <http://eprints.ucm.es/5698/1/2004-8.pdf>
- Martín Abad, Julián; Becedas González, Margarita; Lilao Franca, Óscar (2008). *La descripción de impresos antiguos: análisis y aplicación de la ISBD(A)*. Madrid: Arco/libros, 2008.
- Moll, Jaime (1979). *Problemas bibliográficos del Siglo de Oro*. Madrid: Imprenta Aguirre, 1979. Separata de Boletín de la Real Academia Española ISSN 0210-4822 59: 216, (en.-abr. 1979, 49-107. También en: Alicante: Biblioteca Virtual Miguel de Cervantes, 2009. <http://www.cervantes-virtual.com/obra/problemas-bibliograficos-del-libro-del-siglo-de-oro-0/> (Consulta: 2017-08-17)).
- Pedraza-Gracia, Manuel-José. *Deontología profesional en torno al libro antiguo*. // Ibersid. (2010) 49-58.
- Reyes Gómez, Fermín de los (2003). *Los preliminares en la identificación del libro antiguo*. // Pedraza Gracia, Manuel José (ed.) (2003), *Comercio y tasación del libro antiguo: análisis, identificación y descripción* (Textos y materiales). Zaragoza: Prensas Universitarias, 2003, 201-255.
- Sánchez Herrador, Miguel Ángel; Asencio Padilla, David; Carrasco Pleguezuelo, Inmaculada (2010). *El deterioro del libro antiguo como fuente de información histórica*. // Revista general de información y documentación ISSN 1132-1873 20 (2010) 281-296.
- Sánchez Sánchez-Serrano, Antonio; Prieto de la Iglesia, Remedios (2011). «Auctor», «Autor» y otros problemas semánticos concernientes a la autoría, gestación y ediciones de la Celestina. // Celestinesca ISSN 0147-3085 35 (2011) 85-134.
- Universidad de Salamanca. Biblioteca General Histórica (2012-); *Antiguos poseedores (provenance)*. Salamanca: Universidad de Salamanca, 2012-. <http://bibliotecahistorica.usal.es/es/recursos/antiguos-poseedores> (2017-08-17).

Universitat de Barcelona (1998-); Macas de impresor. Barcelona, Universitat de Barcelona, 1998-. http://www.bib.ub.edu/fileadmin/impressors/home_spa.htm (2017-08-17).

University of Cambridge (2013-); Fleuron: A Database of Eighteenth-Century Printers' Ornaments. Cambridge:

University of Cambridge, 2013-. <https://fleuron.lib.cam.ac.uk/> (2017-08-17).

Enviado: 2017-03-31. Segunda versión: 2017-08-18.
Aceptado: 2018-01-11.

Estudos sobre análise de assunto no Brasil: estado da arte e perspectivas futuras

Estudios sobre el análisis de contenido en Brasil: estado de la cuestión y perspectivas futuras

Studies about subject analysis in Brazil: research and future perspectives

Paula Regina DAL'EVEDOVE (1), Roberta Cristina Dal'Evedove TARTAROTTI (2) E Mariângela Spotti Lopes FUJITA (3)

(1) Universidade Federal de São Carlos, Av. Washington Luis, s/n, CEP. 13565-905, São Carlos, SP, Brasil, dalevedove@ufscar.br, (2 e 3) Universidade Estadual Paulista, Av. Hygino Muzzi Filho, 737, Campus Universitário, CEP. 17525900, Marília, SP, Brasil, roberta_tartarotti@yahoo.com.br, mariangelasf57@gmail.com

Resumen

Este estudio pretende trazar un mapa y discutir el conocimiento producido sobre análisis de contenido en la Ciencia de la Información brasileña. Con este objetivo, se llevó a cabo una pesquisa bibliográfica de los artículos científicos publicados en revistas brasileñas en el ámbito de la Ciencia de la Información, utilizando como términos de búsqueda "análisis de contenido" y "análisis conceptual". El análisis del *corpus* de investigación permitió, entre otros resultados, la identificación de las referencias teóricas actuantes en este universo y las tendencias actuales en la literatura nacional de Ciencia de la Información, lo que permitió esbozar una visión general de los principales desafíos y perspectivas de futuro para la pesquisa y desarrollo en análisis de contenido.

Palabras clave: Análisis de contenido. Investigación. Estado de la cuestión. Brasil.

1. Introdução

O desenvolvimento científico e consequente crescimento da produção científica nas últimas décadas levaram à necessidade de criação de instrumentos que possibilitassem tanto a gestão como uma melhor recuperação dos documentos em sistemas de informação, especialmente por assunto, o que nos leva a duas dimensões: a representação e a recuperação temática da informação. Esta atividade, denominada indexação, é percebida como um processo intelectual realizado pelo indexador, que tem por base a compreensão do texto e a representação dos documentos. Como abordagem teórica da área de Organização e Representação do Conhecimento, a indexação viabiliza a representação temática dos documentos, com fins de recuperação por assunto em sistemas de recuperação da informação.

Neves, Dias e Pinheiro (2006) esclarecem que as pesquisas sobre indexação têm evidenciado três pontos: compreensão de texto (identificação do conteúdo de um documento); produção de texto

Abstract

The present study aims to map and discuss the knowledge produced on the subject analysis in Brazilian Information Science. For that, a bibliographic research was carried out by means of the survey of scientific articles published in national journals of the area of Information Science, using search terms: "subject analysis" and "conceptual analysis". The analysis of the research corpus provided, among others, the identification of the theoretical referents in this universe and the current trends in the national literature of Information Science, which allowed to outline a panorama of the main challenges and future perspectives of research and development in subject analysis.

Keywords: Subject analysis. Research. State of the art. Brazil.

(tradução do conteúdo identificado para a linguagem de indexação e/ou sua expressão na forma de um resumo); e a representação do conteúdo (criação de vocabulários controlados, como os tesouros).

Conceitualmente, a indexação de assuntos é um processo formado por subprocessos ou etapas que tem como objetivos identificar o conteúdo de um documento e representá-lo por meio de uma metalinguagem construída, com o intuito de promover a efetiva recuperação da informação (Tartarotti e Boccato, 2013). Nesta concepção, entendemos o processo de indexação mediante determinação do assunto do documento e tradução do assunto por termos representáveis em linguagens de indexação, ou seja, representação por conceitos e representação por linguagens construídas (sistemas de classificação, cabeçalhos de assunto ou tesouros). A divisão em duas etapas não é consensual entre teóricos e pesquisadores da área dedicados à questão, podendo-se encontrar indicações de três, quatro, cinco ou mais etapas compreendidas pelo processo (Chan, 1994; Cleveland e Cleveland, 1990; ISO

5963, 1985; Langridge, 1989; Chu e O'Brien, 1993; Mai, 1997; 2000; Van Slype, 1991; Tamayo e Valdez, 2008; dentre outros).

Apesar das diferenças em termos teóricos e práticas adotadas, Fujita, Rubi e Boccato (2009, p. 24) explicam que as etapas do processo de indexação não precisam ser necessariamente realizadas de maneira sequencial, visto que o indexador profissional, já familiarizado com o processo, pode realizá-las simultaneamente e até de maneira automática. Neste sentido, quanto mais familiarizado com a área científica especializada no qual atua, mais estas etapas são superpostas durante o processo de indexação pelo profissional. De acordo com Mai (2000), a ação tomada por indexadores com diferentes níveis de experiência pode não ser exatamente a mesma, mas as etapas e elementos do processo de indexação de assunto são considerados fundamentais para qualquer processo de indexação.

Como etapa inicial, a análise de assunto, também denominada por *análise temática* ou *análise conceitual*, é considerada na literatura e na prática profissional como sendo a mais importante do processo de indexação, pois determina os resultados das estratégias de busca realizada pelo usuário no momento da recuperação da informação. Sendo, portanto, a “operação base para todo o procedimento de recuperação de informações” (Cesarino e Pinto, 1980, p. 32). Esta etapa se efetiva mediante a leitura documental, momento em que os conceitos tratados em um documento são identificados e selecionados para posterior representação em termos de indexação. Em outras palavras, a análise de assunto decidirá sobre os principais tópicos do assunto de um documento, precedendo a fase de tradução desses tópicos de acordo com a linguagem de indexação adotada pelo sistema de recuperação de informação (Chu e O'brien, 1993).

Considerando-se que a construção de qualquer texto é baseada em um ou mais temas, é por meio da leitura profissional na representação temática da informação que ocorre a busca pela tematicidade documental, revelada na análise de assunto. Assim, a análise de assunto é realizada durante a leitura documental, sendo, por sua vez, subdividida em outros três estágios: compreensão do conteúdo do documento; identificação dos conceitos que representam este conteúdo; e seleção dos conceitos válidos para recuperação. É a leitura do documento que dá início à atividade de indexação, diferindo de uma leitura normal ao exigir procedimentos diferentes, “ainda que os conhecimentos necessários para um bom entendimento de um texto sejam comuns a ambas” (Fujita, 2003, p. 84).

Durante a análise de assunto, o indexador examina o documento procurando compreender o texto por meio da identificação do assunto, com o objetivo posterior de selecionar os conceitos indexáveis. A utilização e o domínio de normas e metodologias de análise pelo indexador possibilitam bons resultados na recuperação, o que pressupõe o conhecimento de abordagens sistematizadas ao texto, cuja eficácia desta atividade pode ser avaliada frente aos resultados obtidos pelo usuário no momento da recuperação da informação (Fujita, 2003). Sendo assim, a análise de assunto implica em determinar a tematicidade (tema, tópico ou atinência) do documento mediante a identificação e seleção dos conceitos que comporão o assunto principal e secundários. Constitui-se, portanto, em uma etapa complexa e subjetiva que exige todo um trabalho intelectual por parte do indexador na construção de sentidos e significados.

Sendo notória a sua importância no processo de indexação, a proposta neste estudo foi avaliar o estado da arte das pesquisas brasileiras sobre análise de assunto. Para tanto, apresenta a trajetória das pesquisas publicadas em formato de artigo científico para compreender a atual constituição e enfoque da temática, de modo a contribuir com o avanço das produções intelectuais dos pesquisadores da Ciência da Informação brasileira.

2. Procedimentos metodológicos

De modo a mapear as pesquisas sobre análise de assunto conduzidas pela comunidade brasileira de Ciência da Informação, o estudo desenvolveu-se como pesquisa exploratória e descritiva, com uso de fontes bibliográficas para aproximação, abordagem e tratamento do objeto de estudo proposto. Portanto, recorreu-se às pesquisas anteriores para verificar a produção de conhecimento em análise de assunto no cenário nacional.

Para uma melhor aproximação com o objeto de estudo, a pesquisa foi pautada em uma sequência ordenada de procedimentos relativos à pesquisa bibliográfica apresentados por Salvador (1986), delimitados em: a) escolha do assunto e formulação do problema de pesquisa; b) investigação das soluções a partir da coleta da documentação; c) análise explicativa das soluções; e d) síntese integradora (análise, reflexão e proposição de soluções mediante leitura, anotações e indagações do material).

Delimitou-se como *corpus* de análise pesquisas publicadas em formato de artigo científico em periódicos brasileiros de Ciência da Informação em razão do importante papel que estes canais de

comunicação formal desempenham para a divulgação e promoção das pesquisas acadêmico-científicas realizadas na área.

Foram coletados todos os artigos que continham os termos “análise de assunto”, “análise de conteúdo”, “análise temática” e “análise conceitual” mediante busca nos campos de pesquisa dos periódicos científicos pertencentes aos estratos A1 até B2, conforme lista de classificação de periódicos *Qualis* 2017 elaborada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). A busca realizada em março de 2017 na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI) viabilizou a localização de 37 artigos que foram analisados por meio da leitura do título, resumo e palavras-chave, com o propósito de verificar a pertinência dos mesmos à temática em estudo. Sendo desconsiderados os trabalhos que faziam alusão de forma muito indireta ao tema, o corpus de pesquisa foi constituído por 31 artigos que datam de 1980 a 2016. Para a elaboração das redes, utilizou-se o *software Ucinet*.

3. Resultados e discussão

No primeiro momento, empregou-se a estatística descritiva para a apresentação dos dados referentes às características gerais do corpus da pesquisa. Os 31 artigos recuperados estão distribuídos em 15 diferentes periódicos. O periódico com maior destaque é *Perspectivas em Ciência da Informação*, concentrando 16% do total de artigos publicados na temática análise de assunto. *Ciência da Informação*, *DataGramaZero* e *Informação & Sociedade* concentram 10% dos estudos cada. Juntos, os periódicos mencionados representam 46% da produção brasileira na temática. Outro grupo formado pelos periódicos *Brazilian Journal of Information Science*, *Informação & Informação*, *PontodeAcesso*, *Revista de Biblioteconomia de Brasília*, *Revista Digital de Biblioteconomia e Ciência da Informação* e *Tendências da Pesquisa Brasileira em Ciência da Informação* concentra 39% da produção total ou 6,5% cada; seguido do grupo constituído pelos periódicos *Biblionline*, *Em Questão*, *Encontros Bibli*, *Revista ACB* e *Transinformação* com apenas 3% dos artigos cada, conforme segue (Tabela I).

Dos 33 autores que publicaram na temática, destaca-se Fujita com 11 artigos, seguida de Dias, Naves, Neves e Redigolo, ambos com 3. Um terceiro grupo de autores composto por Cervantes, Lima, Maculan, Rodrigues e Rubi publicaram 2 artigos cada. Já com apenas 1 artigo aparecem os autores Almeida, Alves, Boccato, Carvalho, Cavalcanti, Cesarino, Cordeiro, Dal'Evedove, Duarte, Ferreira, Gonçalves, Marinho, Moraes,

Moura, Nascimento, Novellino, Oliveira, Pinheiro, Pinho, Pinto, Silva, Sousa e Souza.

<i>Periódicos</i>	<i>N.</i>
Perspectivas em Ciência da Informação	5
Ciência da Informação	3
DataGramaZero	3
Informação & Sociedade	3
Brazilian Journal of Information Science	2
Informação & Informação	2
PontodeAcesso	2
Revista de Biblioteconomia de Brasília	2
Revista Digital de Biblioteconomia e Ciência da Informação	2
Tendências da Pesquisa Brasileira em Ciência da Informação	2
Biblionline	1
Em Questão	1
Encontros Bibli	1
Revista ACB	1
Transinformação	1

Tabela I. Distribuição dos artigos nos periódicos de Ciência da Informação

<i>Autores</i>	<i>N.</i>
Fujita, M. S. L.	11
Dias, E. W.	3
Naves, M. M. L.	3
Neves, D. A. de B.	3
Redigolo, F. M.	3
Cervantes, B. M. N.	2
Lima, G. A. B. de O.	2
Maculan, B. C. M. dos S.	2
Rodrigues, M. R.	2
Rubi, M. P.	2
Almeida, C. C. de	1
Alves, R. C. V.	1
Boccato, V. R. C.	1
Carvalho, M. A.	1
Cavalcanti, H. C.	1
Cesarino, M. A. da N.	1

Cordeiro, R. I. de N.	1
Dal'Evedove, P. R.	1
Duarte, F. A.	1
Ferreira, A. C.	1
Gonçalves, E. F.	1
Marinho, A. C. M.	1
Moraes, J. B. E. de	1
Moura, M. A.	1
Nascimento, F. A.	1
Novellino, M. S. F.	1
Oliveira, R. A. de	1
Pinheiro, A. M. V.	1
Pinho, F. A.	1
Pinto, M. C. M. F.	1
Silva, M. dos R. da	1
Sousa, B. P. de	1
Souza, R. F.	1

Tabela II. Autores mais produtivos

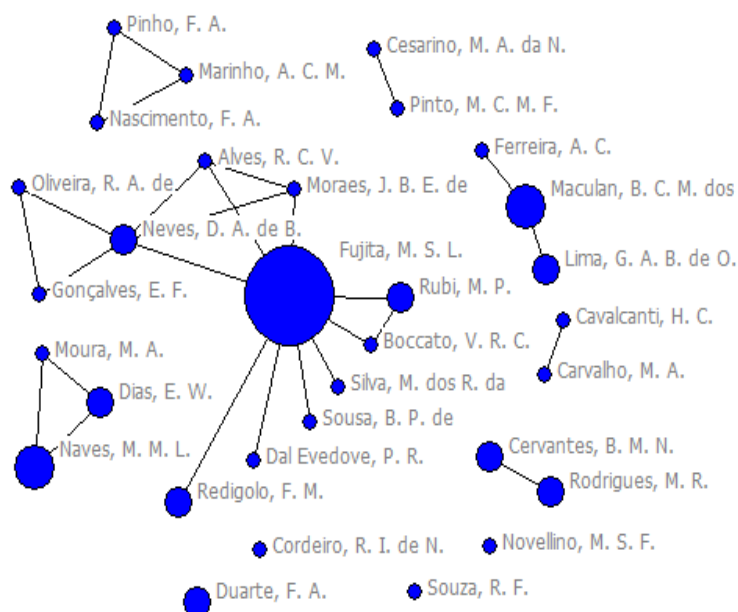


Figura 1. Representação gráfica da rede de coautoria

Ao elaborarmos a rede de coautoria, temos a seguinte representação gráfica (Figura 1). Conforme apresentado, percebe-se uma rede não muito densa, sendo que a maior subrede, encabeçada por Fujita é composta pelos seguintes

autores: Rubi, Boccato, Silva, Sousa, Dal'Evedove, Redigolo, Moraes, Almeida, Alves, Neves, Oliveira e Gonçalves. Outras subredes são compostas, primeiramente, pelos autores Pinho com Marinho e Nascimento; e Dias, Naves e Moura. Os autores Ferreira, Maculan e Lima formam outra subrede. Três subredes aparecem com apenas dois autores, compostas por Cesarino com Pinto; Rodrigues com Cervantes; e Cavalcanti com Carvalho. Dentre os autores que não realizaram colaboração com os demais, aparecem Cordeiro, Novellino, Souza e Duarte. De forma geral, existe uma forte tendência na concentração de colaboração científica entre autores da mesma universidade, indicando subgrupos de pesquisa isolados.

Em relação às temáticas presentes nos artigos, foram identificadas 110 palavras-chave, sendo que um artigo de 1980 não menciona e em um dos artigos as palavras-chave estavam em inglês, tendo sido traduzidas. Ao agruparmos as que se repetiram, no total foram identificadas 100 diferentes palavras-chave nos artigos relacionados à temática análise de assunto.

Palavras-chave	N.
Análise de assunto	14
Indexação	13
Leitura documental	4
Bibliotecas universitárias	3
Catálogo de assunto	3
Indexador	3
Tratamento temático da informação	3
Análise documental	2
Identificação de conceitos	2
Mapas conceituais	2
Modelo de leitura técnica	2
Organização da informação	2
Organização e representação do conhecimento	2
Representação da informação	2

Tabela III. Palavras-chave identificadas nos artigos

De acordo com os dados apresentados, as palavras-chave que mais apareceram foram primeiramente *Análise de assunto*, com 14 ocorrências, seguida por *Indexação* com 13 ocorrências, como esperado. Na sequência, aparecem *Leitura documental*, *Bibliotecas universitárias*, *Catálogo de assunto*, *Indexador* e *Tratamento temá-*

tico da informação com 3 ocorrências cada, demonstrando a tendência de pesquisas sobre análise de assunto em bibliotecas universitárias. Com 2 ocorrências estão *Análise documental*, *Identificação de conceitos*, *Mapas conceituais*, *Modelo de leitura técnica*, *Organização da informação*, *Organização e representação do conhecimento* e *Representação da informação*. As demais palavras-chave apareceram apenas 1 vez nos artigos, não tendo relevância significativa: *Abordagem sociocognitiva*; *Abordagem sociocognitiva qualitativa*; *Análise conceitual*; *Áreas de ciência e tecnologia*; *ARIST*; *Artigo científico*; *Atenção*; *Atuação profissional*; *Catálogos coletivos online*; *Ciclo junino*; *Ciências cognitivas*; *Classificação do conhecimento*; *Contexto profissional*; *Ensino de indexação*; *Estudo bibliométrico*; *Festas populares*; *Fontes de informação visual*; *História da indexação*; *Imagens e filmes*; *Indexação - NBR 12676/1992*; *Indexação – teoria*; *Indexação de assunto*; *Indexação de documentos*; *Indexação de imagens*; *Informação jurídica*; *Jurisprudência*; *Leitura técnica*; *Linguagem documental*; *Mediação da informação*; *Metacognição*; *Metodologia de indexação*; *Modelo de leitura documental*; *Modelo de leitura para indexação*; *Musicologia*; *Normalização*; *Obra musical*; *Partituras*; *Percurso gerativo de sentido*; *Pollissemia*; *Princípios de análise e indexação*; *Processos cognitivos*; *Protocolo verbal*; *Recuperação da informação*; *Recuperação de imagens e filmes*; *Recuperação de informação*; *Superestrutura textual*; *Tesouro*; *Texto científico*; *Transferência de informações*; *Tratamento da informação*; *Tribunal de contas e Usuários*.

Por meio da análise das fontes citadas, obteve-se um conjunto total de 807 referências bibliográficas, o que significa uma média de 26 referências/citações por artigo. A análise das referências não contemplou aquelas que se caracterizavam como normas, pessoas jurídicas e instituições. As referências de autoria múltipla foram desdobradas para contemplar todos os autores do documento. Evidencia-se que 30 autores foram citados pelo menos em 5 diferentes artigos, representando a frente de pesquisa na temática.

Na Tabela IV são indicados os autores mais citados, com os respectivos dados. Neste conjunto de autores mais citados, Fujita aparece com 51 citações distribuídas em 21 trabalhos, seguida por Naves com 25 citações recebidas em 16 artigos; ambas pesquisadoras brasileiras.

Autores	N. de citações	N. de artigos citado
Fujita, M. S. L.	51	21
Naves, M. M. L.	25	16
Lancaster, E. W.	17	13
Hjorland, B.	18	11
Kobashi, N. Y.	14	11
Dias, E. W.	19	10
Pinto Molina, M.	12	10
Guimarães, J. A. C.	17	8
Albrechtsen, H.	10	8
Chaumier, J.	9	8
Beghtol, C.	8	8
Langridge, D.	8	8
Neves, D. A. de B.	10	7
Foskett, A. C.	7	7
Rubi, M. P.	11	6
Moreiro González, J. A.	7	6
Cavalcanti, M. C.	6	6
Chu, C. M.	6	6
Cintra, A. M. C.	6	6
Hutchins, J.	6	6
O'brien, A.	6	6
Pinto, M. C. F.	6	6
Mai, J. E.	9	5
Farrow, J.	7	5
Tálamo, M. de F. G. M.	7	5
Cesarino, M. A. N.	6	5
Gil Leiva, I.	5	5
Pinheiro, A. M. V.	5	5
Šaupert, A.	5	5
Todd, R. T.	5	5

Tabela IV. Autores mais citados

A partir destes autores citados, foi elaborada a rede de citação. Os círculos vermelhos representam os autores citados, enquanto que os círculos azuis representam os autores que foram citantes e citados ao mesmo tempo. Não houve autores que foram apenas citantes. Cabe esclarecer que quanto maior a espessura do seguimento, maior o número de vezes em que o autor foi citado.

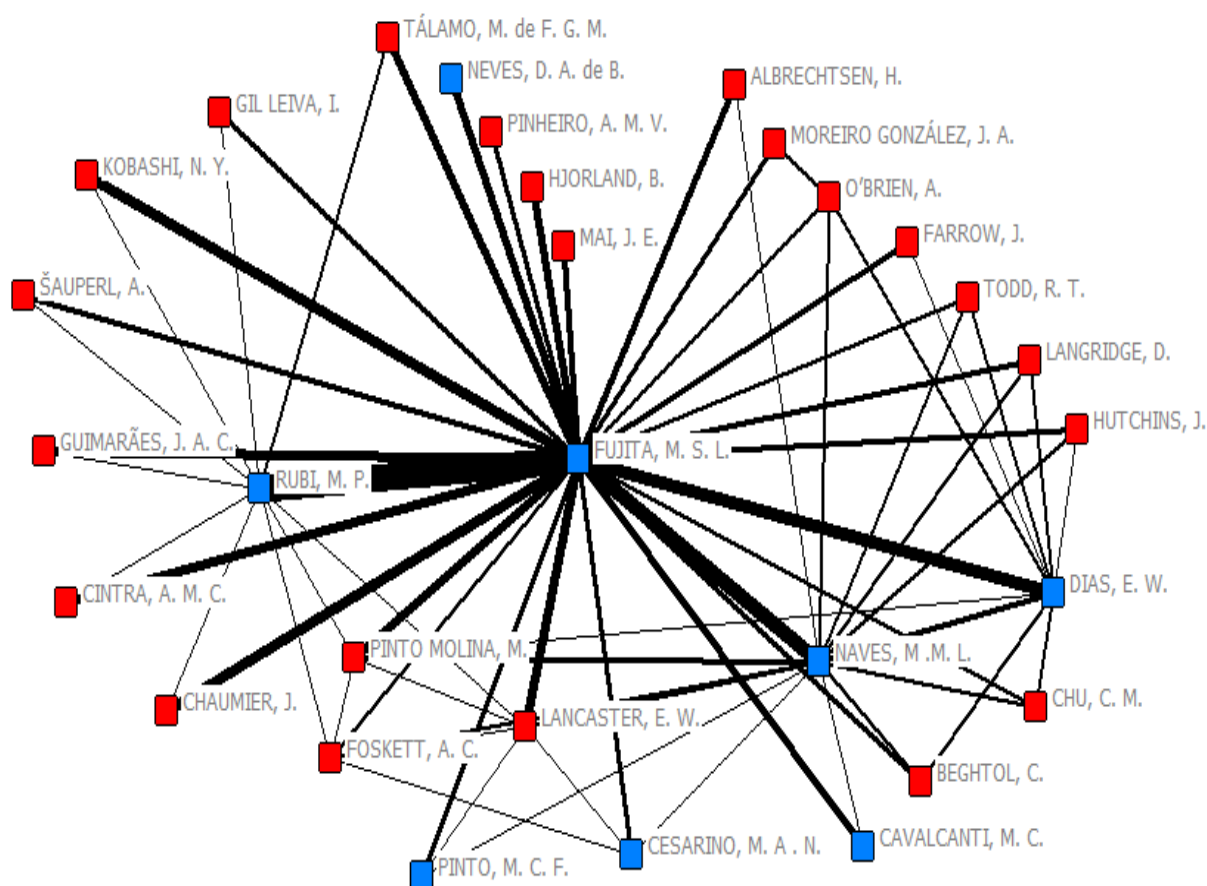


Figura II. Representação gráfica da rede de autores citantes e citados em análise de assunto

Na rede de autores citantes e citados, verifica-se que os autores Fujita, Naves, Dias, Rubi, Cavalcanti, Cesarino, Pinto e Neves formam o grupo de autores que citam e ao mesmo tempo são citados pelos demais autores da rede, podendo ser identificados como o núcleo de especialistas na temática análise de assunto. Dentre os autores brasileiros apenas citados estão: Tálamo, Pinheiro, Cintra, Guimarães e Kobashi. Os autores estrangeiros, maioria dentre os autores citados, são: Lancaster, Hjørland, Albrechtsen, Chaumier, Beghtol, Langridge, Foskett, Moreiro González, Chu, Hutchins, O'Brien, Mai, Farrow, Gil Leiva, Sauperl e Todd. Em sua maioria, estes são autores clássicos da abordagem teórica da indexação no campo da Organização e Representação do Conhecimento.

Na sequência, são dispostos os principais resultados relacionados à categorização do corpus de pesquisa mediante análise interpretativa.

3.1. Panorama das pesquisas no cenário nacional

Dentre os estudos compilados nesta pesquisa, ganham destaque os de natureza conceitual. Ao

consolidarem as concepções teóricas e metodológicas sobre análise de assunto hoje praticadas na área, tornam-se investigações de vasta contribuição para o enriquecimento da temática na Ciência da Informação brasileira.

O primeiro estudo refere-se ao de Cesarino e Pinto (1980), no qual é apresentada uma contextualização geral acerca da análise de assunto como operação fundamental para a recuperação da informação. Os diferentes níveis de análise de assunto são abordados, considerando a determinação do conteúdo informativo do documento; a compreensão das necessidades dos usuários; bem como a escolha ou elaboração de linguagens de indexação. Além de servir de base para outros estudos conduzidos no cenário nacional, apresenta uma das primeiras reflexões acerca da automação para a análise de textos, até então tarefa exclusiva dos indexadores.

Em Naves (1996) são abordadas as várias concepções e abordagens empregadas ao termo, assim como ao conceito de atinência. De modo elucidativo, a discussão estabelecida no estudo trabalha a abordagem conceitual da análise de assunto como etapa do processo de indexação.

Considerada por muitos dos teóricos da área uma atividade intelectual e subjetiva, a imprecisão terminológica da análise de assunto ganha destaque no trabalho da autora, a qual chama a atenção para a influência negativa desta divergência conceitual na determinação de assunto dos documentos por parte dos indexadores.

Tomando como base as dificuldades inerentes à análise de assunto, Fujita (2003) discorre, mediante revisão de literatura, sobre os aspectos que envolvem a identificação de conceitos realizada durante a leitura documental na indexação. Observa que as concepções de análise de assunto orientadas para o conteúdo e para a demanda devem ser privilegiadas pelo indexador para garantir bons resultados na recuperação.

A prática da indexação foi objeto de estudo no trabalho apresentado por Silva e Fujita (2004). Neste, as autoras realizam uma ampla e consistente análise da evolução das tendências teóricas e metodológicas desenvolvidas pela literatura especializada até então. Por conseguinte, o estudo contribui para o entendimento da prática do indexador, com destaque especial para as especificidades do processo e suas etapas inerentes.

Por sua vez, Duarte (2007) estuda os fatores intervenientes no processo de indexação, com ênfase para a etapa da análise de assunto. Ao sistematizar as principais contribuições teóricas acerca da identificação de conceitos, da tematicidade e das concepções de análise de assunto, acentua a importância de se conduzir o processo tendo o usuário como ponto central. Em Neves, Dias e Pinheiro (2006), a intenção foi identificar semelhanças e divergências no uso das principais estratégias metacognitivas entre leitores proficientes indexadores e não-indexadores. O estudo revelou a necessidade de se introduzir o ensino de estratégias metacognitivas no treinamento de leitura do indexador.

Ao discorrerem sobre os aspectos teóricos e normativos da análise de assunto no processo de indexação, o estudo de Sousa e Fujita (2014) oferece à área uma importante contribuição ao apontar lacunas na utilização dos instrumentos que dão subsídios à análise de assunto, com destaque para a norma brasileira NBR 12676/1992 que versa sobre os métodos para análise de documentos e a necessidade de atualização deste repertório.

Com proposta semelhante à realizada aqui, Lima e Maculan (2014) verificaram a literatura existente sobre a análise de assunto, tendo como corpus de análise os capítulos de revisões de literatura do *Annual Review of Information Science and Technology* (ARIST), dos anos de

1977, 1982, 1986 e 1989. Como conclusão, destacam a pouca atenção dada pela área para as questões que cercam a análise de assunto, muito em razão dos procedimentos interpretativos complexos que cercam a atividade ou, ainda, pelo encaminhamento de pesquisas destinadas à automação dos procedimentos.

Em Novellino (1996), o processo de representação da informação é objeto de estudo, com ênfase para a criação de instrumentos e elaboração de metodologias que viabilizem a transferência de Informações. O diferencial está na importância da interpretação do documento como parte de uma área de conhecimento, prerrogativa advinda com a análise de domínio.

Dias, Naves e Moura (2001) realizaram estudo do comportamento de busca de informação do usuário do sistema de informação, como forma de contribuir para o entendimento dos aspectos que envolvem o leitor e seus interesses e, assim, trazer novos entendimentos para as questões relativas à análise de assunto e com a prática do profissional indexador no momento da atribuição de conceitos. Nesta mesma direção, Dias (2004) analisou a percepção dos usuários quanto ao conteúdo dos documentos, tendo como pano de fundo a análise de assunto e o aperfeiçoamento do processo.

Rodrigues e Cervantes (2015) discorrem sobre as semelhanças entre os mapas conceituais e a atividade da análise de assunto, especificamente no que tange às etapas de extração e tradução de conceitos. Para avançar na questão, as autoras realizaram estudo teórico-metodológico para a aplicação dos mapas conceituais no processo de análise, síntese e representação da informação, com especial destaque no auxílio deste instrumento nas estratégias de cognição dos profissionais no tratamento temático da informação (Rodrigues e Cervantes, 2016).

Os temas de estudo com perspectiva mais específica são: observação dos elementos para análise e representação temática de partituras musicais, propondo uma análise conceitual a partir da leitura documental amparada por elementos básicos no âmbito da Musicologia (Cavalcanti e Carvalho, 2011); análise das classificações do conhecimento com base na análise de assunto e outras abordagens teóricas, buscando verificar os fatores intervenientes nos processos de organização e representação da informação no universo da ciência e tecnologia (Souza, 2012); reflexão acerca dos aspectos básicos para a análise de imagens e filmes a serem considerados na determinação de princípios e diretrizes para a elaboração de política de indexação e na recuperação em diversos ambientes informacionais

(Cordeiro, 2013); uso de técnicas de análise de assunto e de conteúdo para a observação da construção imagética e discursiva do ciclo junino, tendo as letras de músicas como recurso informacional (Pinho; Marinho e Nascimento, 2014); investigação das estratégias metacognitivas para análise de assunto em literatura infanto-juvenil (Alves et. al, 2016); e reflexão acerca da contribuição teórica das ciências cognitivas para o entendimento dos processos ligados à mente (cognição humana e suas implicações), com atenção especial para a indexação de imagens e a análise conceitual desses documentos (Gonçalves; Oliveira e Neves, 2016).

A natureza aplicada da análise de assunto é fortemente trabalhada pela literatura a partir do desenvolvimento de metodologias que buscam contribuir para a leitura documental dos diversos tipos de documentos, com fins de identificação e seleção de conceitos representativos. Uma discussão prática sobre a sistematização dos procedimentos foi objeto de estudo de Fujita e Rubi (2006) ao discorrerem sobre os princípios de elaboração e uso de um modelo de leitura documental para a indexação de artigos científicos empregado na formação de indexadores em leitura documental. Essa proposição metodológica alia estratégias de exploração da estrutura textual e a identificação de conceitos de textos científicos por meio de questionamento, o que favorece o ensino da indexação.

Neste viés investigativo, Ferreira e Maculan (2016) apresentam uma proposta de modelo de leitura técnica de acórdãos direcionado para o contexto dos tribunais de contas, como forma de estabelecer parâmetros que auxiliem a prática da indexação de documentos jurisprudenciais. Tem-se, ainda, uma proposta de modelo de leitura técnica para estruturar a identificação dos conceitos representativos do conteúdo de documentos acadêmicos do tipo teses e dissertações (Maculan e Lima, 2011).

Considerando-se a leitura documental como um ato social, Fujita (2004) investigou as variáveis texto, leitor e contexto, o que viabilizou uma maior compreensão das dificuldades inerentes à análise de assunto. O estudo discorreu sobre as principais dificuldades no processo de leitura advindas com a combinação das variáveis, observadas a partir de revisão teórica, entrevistas com indexadores e diagnóstico de sistemas de informação brasileiros. A relevância desta investigação está, sobretudo, com a indicação de novas orientações para a formação do indexador em leitura documental. Em outro estudo, a referida autora observou a tarefa de catalogação de assuntos de livros em bibliotecas universitárias para desenvolver e apresentar o modelo de

leitura documental para a indexação na catalogação de assuntos de livros em bibliotecas, metodologia adaptada do modelo de leitura documental para indexação de textos científicos, apresentado em estudos anteriores. Para tanto, a identificação de conceitos foi combinada com a exploração da estrutura textual de livros no processo de catalogação de assunto (Fujita, 2013). Ainda no contexto da leitura documental, Redigolo e Fujita (2015) debruçaram-se sobre a influência da ação mediadora inerente ao profissional da informação na representação de assunto dos documentos, sendo requerido uma ação consciente na condução do processo e atribuição de conceitos.

Objetivando retratar os diversos fatores que interferem no processo da análise de assunto, a principal contribuição do estudo de Naves (2001) advém com a indicação da necessidade de abordagens interdisciplinares para amenizar e tratar questões inerentes ao processo, principalmente da Linguística, Psicologia Cognitiva e Lógica. Considerando a relevância do estudo semiótico para a atividade de indexação, Redigolo e Almeida (2012) propuseram uma reflexão teórica acerca das relações e contribuições da Filosofia e da Semiótica no tocante às questões que envolvem a análise de assunto, bem como os conceitos de hábito, experiência, representação e tipos de signos gerados nesse processo.

Em Fujita, Boccato e Rubi (2010) os procedimentos, dificuldades e percepções de bibliotecários, usuários e gestores acerca do processo de indexação na catalogação em bibliotecas universitárias foram observadas sob a perspectiva da abordagem sociocognitiva. A ausência de procedimentos para análise e representação do livro, a incompatibilidade da linguagem de indexação e problemas relacionados à recuperação de assuntos nos catálogos, fruto da tendência de atuarem como bases de dados, foram alguns dos resultados identificados.

O estudo apresentado por Dal'Evedove e Fujita (2012) centrou esforços para compreensão do conhecimento profissional acerca da sistematicidade da catalogação de assunto em perspectiva sociocognitiva. A análise do processo em contexto de bibliotecas universitárias pela perspectiva profissional culminou na constatação de que muitas das ações desempenhadas pelos catalogadores de assunto em sua prática cotidiana são alicerçadas em experiências pessoais e profissionais; resultado que reforça a importância que manuais, procedimentos e instrumentos ocupam na prática profissional com fins de viabilizarem a padronização do processo.

Tendo como ponto central a importância do aprimoramento e atualização de normas e procedimentos de análise de assunto, Redigolo (2014) observou a catalogação de assunto praticada em bibliotecas universitárias do Brasil e da Espanha, concluindo que a ausência de instrumentos metodológicos faz com que o profissional desenvolva seus próprios métodos e não se pautem nas concepções trabalhadas e propostas pela literatura especializada, o que acarreta prejuízos na representação de assunto e recuperação da informação.

Apesar do avanço das pesquisas brasileiras acerca da indexação, observa-se que o primeiro estágio do processo, a análise de assunto, apresenta questões que não foram devidamente exploradas, sendo necessário novos esclarecimentos para o avanço da temática no cenário nacional.

4. Tendências de pesquisa na Ciência da Informação brasileira

A categorização das pesquisas brasileiras sobre análise de assunto publicadas em periódicos científicos viabilizou a compreensão do estado atual das discussões na temática, assim como o levantamento das principais lacunas a serem trabalhadas pelos teóricos da Ciência da Informação no Brasil. Na maioria dos casos, as conclusões dos estudos reforçam a escassez de investigações no campo da representação temática da informação e alertam para a necessidade de aprofundamento teórico da etapa de análise de assunto.

Como contributo à questão, são apresentados, de modo sumarizado, alguns dos pontos que merecem destaque e atenção pelos especialistas, quais sejam:

- Maior atenção aos fatores intrínsecos e extrínsecos interferentes no processo de indexação, especificamente na etapa de análise de assunto, dado a sua natureza subjetiva;
- Proposta de estudos que verifiquem a aplicação prática da análise de assunto durante a indexação realizada em diferentes contextos informacionais; assim como a atualização de normas e políticas específicas para oferecer maior suporte à prática profissional;
- A complexidade da sistematização de seus procedimentos exige um olhar verticalizado para os aspectos histórico-culturais envolvidos no processo;
- Abordagens complementares acerca do processamento da informação durante a atividade de análise de assunto, por decorrência

da dificuldade em se compreender como a mente humana realiza determinadas funções cognitivas;

- Maior familiaridade com as atividades cognitivas e metacognitivas do profissional para a criação de novos saberes e conhecimento teórico-aplicado;
- Observações complementares que privilegiem o ponto de vista dos usuários para a criação de arcações teóricas e metodológicas em análise de assunto; e
- A deficiência na formação e falta de especialização do profissional reforçam a condução de estudos específicos e aplicados aos diversos contextos de informação.

5. Considerações finais

Ao objetivar a mediação da informação socializada para o indivíduo, produzindo novos conhecimentos, o ciclo de operações documentais é considerado a base epistemológica da Ciência da Informação, composto por produtores/autores e consumidores/usuários de informação como substrato para a produção de novos conhecimentos. Neste cenário, destaca-se o importante papel da indexação nos sistemas de recuperação da informação.

Considerando-se a relevância de estudos destinados à compreensão do conhecimento produzido em indexação, esta investigação foi pensada e construída de modo a verificar o atual estado da arte sobre a análise de assunto no meio acadêmico-científico da Ciência da Informação brasileira.

Mediante um amplo mapeamento bibliográfico, observou-se a produção do conhecimento na temática divulgada como artigos científicos em periódicos nacionais da área. A sistematização, análise e interpretação dos dados viabilizaram a postulação de hipóteses e recomendações que servirão de ponto de partida para outras pesquisas.

De modo geral, destaca-se a necessidade de pesquisas complementares que busquem subsídios para as múltiplas questões que envolvem a análise de assunto do ponto de vista teórico e prático. Alguns desses caminhos a serem percorridos foram aqui evidenciados, tomando-se como base os resultados e considerações da atual literatura periódica brasileira na temática.

Referências

- Chan, L. M (1994). *Cataloging and classification: an introduction*. 2nd ed. New York: McGraw-Hill, 1994.

- Chu, C. M., O'Brien, A (1993). Subject analysis: the critical first estage in indexing. // *Journal of Information Science*. 1:19.
- Cleveland, D. B.; Cleveland, A. D (1990). *Introduction to indexing and abstracts*. 2nd ed. Englewood: Libraries Unlimited, 1990.
- Cesarino, A. da N.; Pinto, M. C. M. F (1980). Análise de assunto. // *R. Bibliotecon*. Brasília 8:1.
- Fujita, M. S. L. (2003). A identificação de conceitos no processo de análise de assunto para indexação. // *Revista Digital de Biblioteconomia e Ciência da Informação*. 1:1 60-90.
- Fujita, M. S. L.; Rubi, M. P.; Boccato, V. R. C (2009). As diferentes perspectivas teóricas e metodológicas sobre indexação e catalogação de assuntos. // Fujita, M. S. L. (Org.). *A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias*. São Paulo: Cultura Acadêmica, 2009. 19-42.
- ISO 5963 Documentation. *Methods for Examining Documents, Determining their Subjects, and Selecting Indexing Terms*. Geneve: International Organization for Standardization, 1985.
- Lancaster. F. W (2004). *Indexação e resumos: teoria e prática*. Tradução de: Antonio Agenor Briquet de Lemos. 2 ed. Brasília: Briquet de Lemos, 2004.
- Langridge, D (1989). *Subject analysis: principles and practice*. London: Bowker-Sour, 1989.
- Mai, J-E. (1997). The concept of subject: on problems in indexing. // *Knowledge Organization for Information Retrieval. Proceedings of the 6th International Study Conference on Classification Research*. 6, 60-67, 1997.
- Mai, J-E. (2000). Deconstructing the Indexing Process. *Advances in Librarianship* 23: 269-298.
- Neves, D. A. de B.; Dias, E. W.; Pinheiro, A. M. V (2006). Uso de estratégias metacognitivas na leitura do indexador. // *Ci. Inf.* 35:3 (Sep./Dec. 2006) 141-152.
- Salvador, A. D (1986). *Métodos e técnicas de pesquisa bibliográfica*. // Porto Alegre: Sulina, 1986.
- Tamayo, A. M. M.; Valdez, J. C. *Indización y clasificación en bibliotecas*. Buenos Aires: Alfagrama, 2008.
- Tartarotti, R. C. D., Boccato, V. R. C (2013). A abordagem teórica bakhtiniana e o processo de indexação: diálogos. // *Versão Beta: sob o signo da palavra*. 11:74.
- Van Slype, G (1991). *Lenguages de indización: concepción, contrucción y utilización en los sistemas documentales*. Trad. Pedro Hípola e Félix de Moya. Madrid: Fundación Germán Sánchez Ruipérez, 1991.

Enviado: 2017-04-01. Segunda versão: 2017-09-28.
Aceptado: 2018-03-23.

A informação como assunto no direito e na ciência da informação: um estudo comparativo a partir da realidade brasileira

*La información como tema en el derecho y en la ciencia de la información:
un estudio comparativo a partir de la realidad brasileña*

Information as a subject in law and in information science: a comparative study based on the Brazilian reality

**João Carlos Gardini SANTOS (1), José Augusto Chaves GUIMARÃES (1),
Juan-Carlos FERNÁNDEZ-MOLINA (2)**

(1) Universidade Estadual Paulista Júlio de Mesquita Filho – UNESP, Brasil, jcgardini@gmail.com, guima@marilia.unesp.br

(2) Universidad de Granada, España, jcfernan@ugr.es

Resumen

Este artículo busca delinear hasta cómo se presenta el concepto de información en las áreas de Derecho y Ciencias de la Información en Brasil, buscando identificar sus características descriptivas en sus aspectos convergentes y complementarios. Para ello, se realizó una revisión bibliográfica sobre el tema seguida de la aplicación de la metodología de análisis de contenido (Bardin, 2009) en función de las categorías procesos, objetivos, elementos, características y derechos derivados de la información. Los resultados revelan una convergencia entre el Derecho y la Ciencia de la Información en los aspectos relacionados con el acto de informar o informar a alguien, la transmisión de conocimiento como el objetivo de la información y la necesidad de un soporte para dar materialidad a la información. En el caso de las ciencias de la información, se hace mayor hincapié en los procesos informacionales de producción, organización y uso de la información, y sus dimensiones conceptuales (como conocimiento, como proceso y como cosa). En el caso del derecho, se hace hincapié en los efectos o garantías que surgen del acceso a la información y los principios jurídicos inherentes al derecho a la información (por ejemplo, no divulgación, publicidad, universalidad, etc.). Se puede concluir que el uso cada vez más técnico del término en la Ley requiere una referencia teórica a la Ciencia de la Información, para que su dimensión conceptual pueda ser delimitada de forma más específica, de cara a la consolidación de su terminología.

Palabras clave: Ley de acceso a la información. Información. Información jurídica. Análisis de contenido. Organización de la información.

1. Introdução

Após o surgimento do conceito de informação na Lei de Acesso à Informação (Lei nº 12.527/2011 – LAI) o termo passou a ser utilizado, pelo Supremo Tribunal Federal do Brasil (STF), com maior parcimônia e cuidado semântico, razão pela qual tornou-se necessária a identificação das categorias temáticas que o permeiam a fim

Abstract

This paper seeks to delineate the extent to which the concept of information is presented in the areas of Law and Information Science in Brazil, seeking to identify descriptive features in their convergent and complementary aspects. For this, a literature review on the subject followed by the application of the content analysis methodology (Bardin, 2009) was carried out based on the categories processes, objectives, elements, features and rights arising from the information. The results reveal a convergence between the Law and the Information Science in aspects related to the act of informing or informing someone, the transmission of knowledge as the objective of the information and the need for a support to give materiality to the information. In the case of Information Science, there is a greater emphasis on the informational processes of production, organization and use of information and its conceptual dimensions (as knowledge, as a process and as a thing). In the case of Law, the emphasis is on the effects or guarantees arising from access to information and the legal principles inherent in the right of information (for example, non-disclosure, publicity, universality etc.). It can be concluded that the increasingly technical use of the term in Law requires the theoretical reference of Information Science, so that its conceptual dimension can be delimited more specifically, with a view to the sedimentation of its terminology.

Keywords: Access to information act. Information. Legal information. Content analysis. Information organization.

de que ele seja caracterizado como um instituto jurídico no âmbito jurisprudencial e possa ser mais desenvolvido e representado enquanto um termo descritor de um tesouro jurídico.

Nesse sentido, e para que se possa contribuir para um uso cada vez mais técnico do termo, necessário se torna analisar como se dá essa di-

menção conceitual, no âmbito do Direito e da Ciência da Informação, que possa subsidiar uma delimitação mais específica com vistas à sedimentação de sua terminologia. Para que isso ocorra, utilizando a metodologia da análise de conteúdo (Bardin, 2008), analisa-se, no âmbito do Brasil, a literatura do Direito e da Ciência da Informação relativa à questão da informação, notadamente sob o aspecto conceitual, de modo a identificar seus aspectos convergentes e as diversidades de abordagem.

Espera-se, com isso, delinear em que medida o conceito de informação se apresenta nos âmbitos do Direito e da Ciência da Informação, no Brasil, buscando identificar traços descritivos em seus aspectos convergentes e complementares e contribuir para os processos de caracterização da informação como descritor de tesouros, em especial os tesouros jurídicos, uma vez que esse termo vem assumindo, cada vez mais a condição de instituto jurídico (Santos; Guimarães, 2016).

2. Metodologia

Partiu-se de uma revisão de literatura sobre a informação no Direito brasileiro assim como na teoria de Ciência da Informação, uma vez que esse método visa a fornecer fundamentação teórica à pesquisa, além de permitir que seja identificado qual é o atual estágio do tema de pesquisa. Destaca-se, ainda, que a revisão bibliográfica permite “[...] ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que poderia pesquisar diretamente” (Gil, 2010, p. 30).

Uma vez devidamente analisadas as fontes e destacadas as menções à concepção de informação, aplicou-se a metodologia de análise de conteúdo (Bardin, 2008), a qual pressupõe um procedimento de decomposição do todo em suas diversas partes constitutivas a partir de uma sistemática que se desenvolve em três fases: pré-análise, na qual se escolhem os documentos para a formação do corpus, além da formulação das hipóteses e da definição dos objetivos da análise; exploração do material, quando se efetiva o processo de categorização do conteúdo; e, por último, o tratamento dos resultados, quando se realizam as inferências e interpretações (Guimarães; Sales, 2010).

Para tanto, estabeleceram-se as seguintes categorias de análise: Processos, Objetivos, Elementos, Características e Direitos e garantias decorrentes da informação, cujos resultados são apresentados e discutidos na seção 5.

3. A dimensão conceitual da informação no Direito brasileiro

No âmbito do Direito, necessitamos recorrer à doutrina e à legislação para definir e caracterizar a informação. Por se tratar da norma de maior importância no âmbito jurídico nacional, inicialmente, realizaremos uma análise acerca dos aspectos constitucionais que envolvem o direito fundamental de acesso à informação e, em seguida, dos aspectos infraconstitucionais que orbitam esse direito.

De pronto, devemos observar que o direito à informação está previsto na Constituição Federal de 1988 (CF/88), mas tal direito já se encontrava previsto na redação original da Constituição de 1967, assim como na Emenda Constitucional nº 1 de 1969.

A CF/88 garante o direito à informação no artigo 5º, XIV e XXXIII, no artigo 37, § 3º, II e, por fim, no artigo 216, § 2º. O artigo 5º está inserido no Título II – Dos Direitos e Garantias Fundamentais –, o artigo 37 no Título III – Da Organização do Estado – e o artigo 216 no Título VIII – Da Ordem Social.

Por estar previsto no artigo 5º, trata-se precipuamente de um direito fundamental que, como tal, apresenta as seguintes características: a) universalidade (aplica-se a todas as pessoas, sem distinção de qualquer natureza); b) irrenunciabilidade (ninguém pode deles renunciar); c) inalienabilidade (são indisponíveis, não admitindo alienação); d) imprescritibilidade (não se perdem ao longo do tempo). Tais direitos, como destaca Novellino (2011) apresentam caráter vinculante e obrigatório, de tal forma que a ninguém é permitido escusar-se cumprir/respeitar um direito fundamental.

Dentre os diversos vieses que envolvem a questão da informação na CF/88 destaca-se, primeiramente, o fato de que o direito de acesso à informação abriga a questão da liberdade de informação prevista no artigo 5º, XIV, que assegura o amplo acesso à informação resguardando, quando profissionalmente necessário, o sigilo da fonte, e que não deve ser confundido com a liberdade de expressão do pensamento. Por um lado, o direito de acesso à informação abrange o direito de todas as pessoas terem o acesso amplo e geral às informações (o direito de informar e de ser informado é um dos pressupostos básicos de existência das democracias modernas) e, por outro, a liberdade de expressão do pensamento que, além de envolver o direito de emitir uma opinião, está intrinsecamente relacionada à liberdade de imprensa, sendo essa um reflexo daquela.

Em segundo lugar temos o artigo 5º, XXXIII da CF/88 que trata do direito que todas as pessoas têm de “[...] receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado” (Brasil, 1988). Trata-se do direito de acesso à informação por parte dos órgãos públicos, conforme explicado pelo Ministro Roberto Barroso no Mandado de Segurança nº 28.178 do Supremo Tribunal Federal. Observamos, nesse caso, que a regra é a publicidade das informações e que o sigilo é a exceção, conforme disposto na parte final desse mandamento constitucional. A isso acrescenta-se a complementação advinda do direito de obtenção de certidões, previsto no artigo 5º, XXXIV, “b”, do mesmo diploma legal. Essa norma prevê que todas as pessoas têm o direito de obter certidões em repartições públicas para a defesa de direitos e esclarecimentos de situações pessoais independentemente do pagamento de taxas.

Outra citação constitucional referente à informação está prevista no artigo 37, § 3º, II o qual prevê que à lei compete regulamentar os meios de participação dos cidadãos na administração pública direta e indireta, regulamentando o acesso dos usuários aos registros administrativos e às informações sobre os atos de governo, observando o direito disposto no inciso XXXIII do artigo 5º e a inviolabilidade da intimidade, da vida privada, da honra e da imagem das pessoas. Nesse caso, observamos novamente o conflito entre a regra, a publicidade, e a exceção, o sigilo. Lenza (2015, p. 886) explica que nos casos em que se objetiva violar o sigilo, o “ônus argumentativo de demonstrar a caracterização de uma dessas circunstâncias incumbiria a quem pretendesse afastar a regra geral da publicidade”. Destaca-se que a violação indevida das informações de caráter estritamente pessoal é passível de indenização por danos materiais e morais dela decorrente.

Nesse contexto é possível a existência de conflito entre o direito fundamental da privacidade e direito fundamental da liberdade de informação. Apesar da solução do conflito, a fim de verificar qual direito deve prevalecer, somente poder ser analisada no caso concreto, existem alguns critérios de ponderação, criados pela doutrina, que objetivam auxiliar na solução desse conflito. São esses critérios: “i) licitude da informação; ii) forma adequada de transmissão; iii) contribuição para o debate de interesse geral ou relevância para a formação da opinião pública, eixo em torno do qual gira o direito à informação” (Novelino, 2011, p. 455).

A Constituição aborda, ainda sobre a questão da informação, no artigo 216, § 2º, a questão da gestão da documentação governamental. Esse dispositivo prevê que “cabe à administração pública, na forma da lei, a gestão da documentação governamental e as providências para franquear sua consulta a quantos dela necessitem” (Brasil, 1988). Esse dispositivo é regulamentado pela Portaria nº 293 de 01 de abril de 2013 do Ministério da Ciência Tecnologia e Inovação. O Anexo Política de Gestão Documental dessa norma define, no artigo 1º, VI essa atividade como o “conjunto de procedimentos e operações técnicas referentes à produção, classificação, tramitação, uso, avaliação e arquivamento de documentos em fase corrente e intermediária, visando a sua eliminação ou recolhimento para guarda permanente” (Brasil, 2013).

Além disso, essa mesma norma conceitua, no artigo 1º, VII, o acesso à informação como o “ato de franquear o acesso/conhecimento a dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato”. Essa definição possui diversos elementos daqueles encontrados no conceito de informação trazido pela Lei de Acesso à Informação (LAI), abordados a seguir.

Por fim, um último aspecto relativo à informação encontrado na Constituição de 1988 é a possibilidade da impetração do habeas data. Trata-se de uma ação constitucional que permite “o conhecimento de informações relativas à pessoa do impetrante, constantes de registros ou bancos de dados de entidades governamentais ou de caráter público”, “a retificação de dados, quando não se prefira fazê-lo por processo sigiloso, judicial ou administrativo” e “para a anotação nos assentamentos do interessado, de contestação ou explicação sobre dado verdadeiro mas justificável e que esteja sob pendência judicial ou amigável” (artigo 5º, LXXII, “a” e “b” e artigo 7º, III da Lei nº 9.507/1997, respectivamente).

Esse último diploma legal, inclusive, regulamenta o direito de acesso às informações e disciplina o rito processual do habeas data. Ele trata como de caráter público todos os registros ou banco de dados que contenham informações que possam ser transmitidas a terceiros ou que não sejam de uso exclusivo do órgão produtor ou depositário das informações.

O procedimento extrajudicial do habeas data consiste na apresentação de um requerimento ao órgão ou entidade depositária do registro ou banco de dados, o qual será deferido ou indeferido no prazo de quarenta e oito horas. Após esse prazo, a decisão será comunicada ao requerente

em vinte e quatro horas. Se o pedido for deferido, o depositário do registro ou do banco de dados marcará dia e hora para que o requerente tome conhecimento das informações. Se, eventualmente, for verificada a inexatidão de quaisquer dados, o interessado, por meio de petição, poderá requerer a sua retificação, a qual será providenciada em no máximo dez dias a contar da entrada do requerimento. Se, porventura, o interessado desejar apresentar explicações ou contestações sobre os dados, ela será anotada no seu cadastro.

Se o requerimento extrajudicial for indeferido, o interessado poderá ingressar com uma ação judicial de *habeas data* a qual deverá ser devidamente instruída com provas da recusa de algumas das três situações que ensejam o cabimento da ação, a depender do caso. Se for julgada procedente, o juiz marcará data e horário para que o requerido apresente ao requerente as informações e/ou as retificações e/ou as anotações feitas no seu assentamento. Da sentença judicial de improcedência do pedido de *habeas data* cabe apelação. Essa ação é impetrada de forma gratuita e independe da existência de motivação, bastando somente o desejo de ter conhecimento das informações constantes na Administração Pública.

No âmbito da legislação ordinária brasileira, encontra-se em vigor, desde 18 de novembro de 2011, a Lei 12.527/2011, denominada Lei de Acesso à Informação (LAI). O conceito do termo informação está previsto em seu artigo 4º, I e se caracteriza como sendo os “dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato” (Brasil, 2011).

Decompondo a referida definição em seus elementos constitutivos, encontramos os seguintes termos de especial importância para a Ciência da Informação: dados, produção e transmissão de conhecimento, meio, suporte e formato.

Dessa forma, tem-se por dado “algo acontecido, realizado, feito, e que se tornou irrevogável ou necessário” (Semidão, 2014, p. 71), que, ao ser processado, gera informação. Já a produção de conhecimento é a atividade intelectual realizada pelo responsável por manipular os dados enquanto a transmissão do conhecimento é a sua divulgação a terceiros interessados ou não. Tais aspectos levam a concluir que, no contexto da LAI, somente os dados que podem ser utilizados para a produção e transmissão do conhecimento é que são considerados informação.

Relativamente a meio, suporte e formato, tem-se, por meio o “suporte ou material no qual são registrados dados, por exemplo: papel, disquete, disco rígido, CD-ROM, fita magnética” (Cunha; Cavalcanti, 2008, p. 242). O suporte, por sua vez, é o “objeto material, ou dispositivo, sobre o qual, ou no qual se encontram representados os dados ou informações; [...] material (ou dispositivo) ativo ou passivo que pode memorizar uma informação e restituí-la quando necessário” (Cunha; Cavalcanti, 2008, p. 352). Por fim, formato é a “descrição da representação física das informações, independentemente da sua representação codificada, p. ex.: livro, periódico ou mapa” (Cunha; Cavalcanti, 2008, p. 173).

Esse conceito é uma reprodução literal do conceito apresentado pelo artigo 3º, I do Decreto nº 7.724 de 16 de maio de 2012 – regulamento no âmbito federal da LAI – e o objetivo da lei ao conceituar institutos é o de “evitar a negativa de acesso à informação por meio de uma hermenêutica reducionista por parte do agente público obrigado a observar plenamente o direito” (Salgado, 2015, p. 88), já que é possível ocorrer a recusa dos agentes obrigados ao seu cumprimento em razão de uma má interpretação legislativa.

Além disso, e partindo para a análise do termo no âmbito jurisprudencial do Supremo Tribunal Federal (STF), identificamos que a palavra informação possui maior incidência em acórdãos proferidos anteriormente à LAI, o que permite inferir que, após o advento da lei, esse termo passou a ser usado de forma mais específica e parcimoniosa, não sendo mais utilizado como sinônimo indiscriminado de dado, suporte, mensagem ou conhecimento (Santos; Guimarães, 2016).

No mesmo trabalho citado acima observamos, ainda, após análise de conteúdo (Bardin, 2008) que o termo informação é utilizado pelo STF, em ordem decrescente de incidência, nos contextos de informação sobre algo (60 incidências), elementos de informação (41 incidências), liberdade de informação (41 incidências), direito à informação (21 incidências), peças de informação (8 incidências), acesso à informação (7 incidências) e informação jornalística (7 incidências).

4. A dimensão conceitual da informação na Ciência da Informação

A Ciência da Informação, ramo de estudo das Ciências Sociais Aplicadas, surgiu como campo de conhecimento mais especificamente a partir da década de 1960, fruto de todo um avanço tecnológico que se fazia sentir no pós-guerra e em muito influenciada pelas ideias de Vannevar

Bush, em seu legendário artigo *As we may think*, de 1945 (Bush, 1945).

Como destaca Borko (1968, p. 3, tradução nossa), a Ciência da Informação constitui “a disciplina que investiga as propriedades e o comportamento da informação, as forças que governam os fluxos de informação, e os significados do processamento da informação objetivando à acessibilidade e a usabilidade ótimas”. Para tanto, o autor destaca que a Ciência da Informação se vale de conhecimentos que objetivam a organização, coleção, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. Nesse sentido, Tognoli e Guimarães (2011, p. 23) destacam, ainda, que a Ciência da Informação estuda os processos que objetivam a produção, a organização e o uso da informação registrada. Esse registro da informação, como ressaltado por Smit e Barreto (2002, p. 19) serve “para garantir sua permanência no tempo e portabilidade no espaço”.

Ao buscar investigar como a informação se comporta e quais as suas propriedades, a Ciência da Informação, enquanto área de estudos, abrange espaços profissionais distintos, e em sua maioria fortemente ligados a espaços insituacionais específicos, nos quais desenvolve aquilo que Fernandes (1995) denomina como gestão institucional dos saberes: a arquivologia, a biblioteconomia e a museologia. Smit (1993; 1999-2000) a elas metaforicamente se refere como as “3 Marias”, em cujo âmbito as divergências ocorrem mais no nível terminológico que conceitual mas que guardam um importante ponto de convergência: o fato de trabalharem com a produção, a organização e uso de uma informação registrada.

Desde seu nascedouro, como se pode observar no texto seminal de Borko anteriormente citado, a Ciência da Informação se situa em um espaço interdisciplinar que interage com conhecimentos de diversos campos, como a Linguística, a Ciência da Computação, a Administração, a Estatística, o Direito, a Psicologia etc. No entanto, Guimarães (2008, p. 38) elenca algumas balizas que possam evidenciar a disciplinaridade da Ciência da Informação: o objeto, na informação registrada; o objetivo, na geração de conhecimento; os processos de geração (produção), organização e uso da informação registrada; os produtos, a partir da denominada informação documentária (Kobashi, 1994) e os instrumentos, via de regra exteriorizados em códigos, normas, sistemas de organização etc.

A informação registrada como objeto de estudos da área constitui elemento de fundamental

importância, pois é somente a partir dela que o cidadão, e a sociedade como um todo, poderão apoderar-se de forma segura de um conhecimento para que, após internalizado por um indivíduo e socializado com os demais, seja suscetível de constituir matéria-prima para a construção de um novo conhecimento, aspecto que Guimarães (2008) denomina como natureza helicoidal do conhecimento na Ciência da Informação.

Indo além, González de Gomez (1990, p. 121) refere-se à Ciência da Informação como uma área de estudos transdisciplinar cujo objeto reside nas “pragmáticas sociais de informação”, ou, em outras palavras, em “[...] um conjunto de regras e relações tecidas entre agentes, processos e produções simbólicas e materiais”. Essa transdisciplinaridade decorre na necessidade intrínseca de agregar conhecimentos oriundos de diversas disciplinas para, a partir daí, construir um objeto de estudo e análise: a meta-informação e suas relações com a informação.

Cardoso (1996, p. 71) destaca que o termo informação sofreu, ao longo da História, inúmeras modificações quanto a sua acepção, a tal ponto de, em nossos dias, chegar às raias da ambiguidade, pois não raras vezes cai na vala comum dos dados, da comunicação, da instrução e, mais recentemente, do conhecimento.

A partir do fenômeno da explosão informacional, que sucedeu o período pós-guerra e causou forte impacto na sociedade, a informação passou a ocupar papel de centralidade no cotidiano da sociedade, sendo determinante em todos os aspectos da vida cultural, econômica, social etc., razão pela qual cada vez mais ela vem sendo estudada pelos diversos ramos da ciência. Ainda assim, a Ciência da Informação, da mesma forma que os demais ramos em que é utilizada a expressão, encontra dificuldades em obter consenso sobre o que realmente seja informação de tal forma que há vários conceitos ligados a esse termo.

Diretamente relacionada à linguagem, a informação, uma vez transmitida, propicia o processo comunicacional, cujo conteúdo se transmite a partir da linguagem, enquanto um sistema de estruturas em que a exteriorização das ideias dá-se por meio de símbolos (Berlo, 1997, p. 70) que, ao final, buscam organizar e comunicar o pensamento humano (Câmara Junior, 1959; Wurman, 1991).

Essa comunicação, cuja base etimológica remete ao “tornar comum, estabelecer comunhão, participar da comunidade, através do intercâmbio de informações” (Melo, 1977, p. 14),

por sua vez, somente se efetiva a partir da existência de um dado conteúdo – a informação – esta considerada sua matéria-prima (Wurman, 1991) e sua motivação básica para que algo possa ser transmitido de uma mente que, por sua vez, a receberá como informação nova. Semidão (2014, p. 77) ressalta que “o termo informação está invariavelmente relacionado ao conhecimento e à comunicação, independentemente da motivação que suscita o ato de informar”.

Nesse contexto de um conteúdo transmitido em um processo de comunicação, essencial se torna a menção à Teoria Matemática da Comunicação desenvolvida por Shannon e Weaver (1975) para os quais a informação é “uma mensagem enviada por um emissor para um receptor” (Logan, 2012, p. 27). Para tanto, implica um processo de comunicação que pressupõe a existência de um emissor, um sinal, um canal, uma mensagem, um receptor e, ainda os ruídos que daí possam decorrer.

Sem uma preocupação semântica acerca da informação a ser transmitida, a Teoria Matemática da Comunicação centra-se nos aspectos quantitativos relacionados à transmissão, com o objetivo de reduzir os gastos na comunicação eletrônica. Nessa linha de pensamento, Gleick (2013) entende que a informação é contabilizada em bits, os quais entende serem as unidades elementares da informação. Assim, os bits, uma vez assimilados pelas pessoas, configuram uma informação nova. Gleick defende a tese de que a informação está contida em todas as coisas, dentro e fora da ciência, generalizando que “a informação é aquilo que alimenta o mundo: o sangue e o combustível, o princípio vital” (Gleick, 2013, p. 16). Ainda que historicamente importante para a configuração da Ciência da Informação, a Teoria Matemática da Comunicação tem sua aplicabilidade mais diretamente relacionada à Engenharia da Comunicação, utilizando-se de conhecimentos de probabilidade e de estatística para fundamentar seus problemas, de modo que as linguagens ordinárias existentes na comunicação são substituídas por equações matemáticas.

Partindo das definições mais comumente aceitas sobre o termo, em seu sentido geral, como “1. Ação ou efeito de informar-se; 2. opinião sobre alguém; 3. instrução; 4. direção” (Amora, 2009, p. 389), a informação, na Ciência da Informação, vai além do “sinônimo de mensagem, notícias, fatos, eventos e idéias que são adquiridos e passados adiante como conhecimento” (Machado, 2003, p. 15) para assumir uma dimensão ao mesmo tempo complexa e específica.

Cunha e Cavalcanti (2008, p. 201) definem informação como “um conhecimento que pode ser necessário a uma decisão”, de onde decorre a informação registrada como “expressão que em sentido amplo abrange todas as espécies documentárias que possuam conteúdo informativo, independentemente do suporte físico dessas espécies” (Cunha e Cavalcanti, 2008, p. 204).

Etimologicamente destacada por Logan (2012, p. 26) como a ação de “dar uma forma para a mente”, a informação, para Le Coadic (1996) é um conhecimento inscrito que se utiliza de sistemas de signos (a linguagem) gravados em um suporte de modo a comportar um elemento de sentido. Em outras palavras, podemos dizer que a informação é um conhecimento transmitido a um ser consciente através de uma mensagem registrada, de modo a considerá-la, no âmbito da Ciência da Informação, como parte integrante de um processo comunicativo que pressupõe a existência um fluxo helicoidal de produção, organização e uso (Guimarães, 2008).

Nesse sentido, o conhecimento é o objetivo pretendido pela informação, de modo que é inútil referir-se à informação sobre algo que já conhecemos (Rapoport, 1970). Le Coadic (1996, p. 6) explica que a informação se transforma em “desinformação quando o conhecimento inscrito é alterado, falsificado (informação falsa), ou está ausente (não-informação)”.

Nesse sentido, e como já destacado anteriormente, Fernandes (1995) refere-se a um elemento dinâmico, a gestão institucional dos saberes enquanto um conjunto de procedimentos ou ações que pretendem selecionar, sumarizar e organizar o conhecimento contido nas instituições (bibliotecas, arquivos etc.) com o intuito de produzir informação a fim de (re) ligar conhecimentos que, em razão da grande quantidade de informações produzidas, tenham sido separados.

Buckland (1991, p. 351) é quem, de forma mais sistemática, categoriza a informação em três dimensões, nomeadamente: como processo, como conhecimento e como coisa. Assim, para o autor, a *informação-come-processo* refere-se ao ato de informar, de comunicar algo a alguém. Na segunda acepção, a *informação-come-conhecimento*, tem-se a assimilação de algo durante o processo anterior, ou seja, é algo transmitido a alguém que aumenta o conhecimento e, conseqüentemente, diminui a incerteza do sujeito sobre algo. Por fim, a terceira e última acepção, *informação-come-coisa*, compreende tanto objetos quanto documentos que apresentam e exteriorizam conteúdo informacional.

Nesse sentido, para o autor, a *informação-come-processo* refere-se ao ato de informar, de comunicar algo a alguém. Ocorre quando o conhecimento de alguém é modificado em razão da informação recebida. Ela é circunstancial. Isso significa que “quaisquer objetos particulares, documentos, dados ou eventos podem ser considerados como informativos dependendo das circunstâncias” (Buckland, 1991, p. 10, tradução nossa).

Na segunda acepção, a *informação-come-conhecimento*, tem-se a assimilação da informação transmitida durante o processo anterior e, como consequência, há redução de incertezas sobre determinado fato. Outras vezes, porém, a informação gera um aumento de incertezas. Destaca-se que a principal característica da *informação-come-conhecimento* é que ela é intangível – que não se pode tocar ou medir. Todavia, Buckland (1991) ressalta que essas informações devem estar, de alguma forma física, representadas como um sinal, texto ou comunicação, por exemplo. A partir daí é que surge a terceira acepção: a *informação-come-coisa*.

Dessa forma, a *informação-come-coisa*, compreende tanto textos quanto documentos que apresentam e exteriorizam conteúdo informacional. Buckland (1991, p. 2, tradução nossa) ressalta que diversos teóricos não aprovam o uso do termo informação nesse contexto, que também se relaciona diretamente com os sistemas de informação e de recuperação da informação. O propósito da *informação-come-coisa* é:

[...] esclarecer seu significado em relação a outros usos do termo informação; estabelecer a regra fundamental de *informação-come-coisa* no sistema de informação; e considerar o possível uso da noção de *informação-come-coisa* trazendo ordem teórica a campos heterogêneos, mal-ordenados associados com a ‘ciência da informação’.

Silva e Ribeiro (2002, p. 37) vinculam a representação e o registro da informação à aceitabilidade social dos símbolos utilizados e à possibilidade de esses ficarem sedimentados em distintos suportes materiais, tais como papel, disco magnético etc. Isso leva a uma relação necessária entre informação (o conteúdo) e documento (o continente), este último materializando-se, como já destacava Otlet (1934), em uma forma diversificada, como livros, periódicos, peças de conjuntos de arquivos, estampas, fotografias, objetos, partituras, filmes, discos, amostras, modelos, fac-símiles, enfim, “tudo que tenha caráter representativo em três dimensões e, eventualmente, em movimento” (Cunha; Cavalcanti, 2008, p. 132).

Se, por um lado, a Ciência da Informação confere ao documento uma acepção ampla e diversificada, tanto em termos de conteúdo quanto de forma, tal não ocorre no Direito, em que o documento constitui “a prova escrita oferecida em juízo para demonstração do fato ou do direito alegado” (Silva, 2007, p. 493). Dessa forma, o documento insere-se no terreno probatório (a prova documental), em paralelo à prova testemunhal. No entanto, e uma vez constituindo meio de prova, o documento, juridicamente entendido, pode manifestar-se em diferentes continentes, como é o caso da fotografia ou de uma gravação, quando utilizadas como provas em um processo criminal. Assim, e antes de constituir ou não elementos probatórios, o documento, para o Direito, constitui a representação material de um ato/fato, capaz de gerar consequências (a lide ou o conflito de interesses resistidos) na esfera jurídica.

Em síntese, pode-se observar que a informação, “considerada a quinta necessidade do homem, precedida por ar, água, alimentação e abrigo” (Octaviano, Rey e Silva, 1999, p. 175), pressupõe uma ideia (conteúdo), materializada em um suporte (registro) e que serve para fins comunicativos (transmissão) de forma a gerar conhecimento.

5. Apresentação, análise e discussão dos resultados

Apresentam-se, a seguir, as categorias de análise, dentro das quais tem-se os traços descritivos encontrados bem como o campo (Direito - D ou Ciência da Informação - CI) em que são tratados).

<i>Categoria: Processos que incidem sobre a informação</i>
acesso (D)
apropriação (CI)
armazenamento (CI)
colecção (CI)
comunicação (CI)
contabilização em bits (CI)
contestação (D)
geração / produção (CI)
instrução (CI)
intercâmbio (CI)
interpretação (CI)
materialização (CI)
organização (CI)
recuperação (CI)

registro (CI)
representação (CI)
socialização (CI)
transformação (CI)
transmissão (D / CI)
uso (CI)
<i>Categoria: Elementos / componentes da informação</i>
canal (CI)
conteúdo (CI)
dados (CI)
emissor (CI)
fluxo (CI)
formatos (D)
linguagem (CI)
meios (D)
mensagem (CI)
órgão depositário (D)
órgão produtor (D)
registros (D)
ruídos (CI)
sentido (CI)
signos (CI)
símbolos (CI)
sinal (CI)
suporte (D / CI)
<i>Categoria: Características / atributos / propriedades da informação</i>
acessibilidade (CI)
atendência (CI)
ausência (CI)
caráter vinculante (D)
comportamento da informação (CI)
desinformação / não-informação (CI)
falsidade (CI)
imprescritibilidade (D)
inalienabilidade (D)
informação como matéria-prima (CI)
informação como mensagem (CI)
informação documentária (CI)
informação-como-coisa (CI)
informação-como-processo (CI)
licitude (D)
necessidade (CI)
novidade (CI)

obrigatoriedade (D)
permanência (CI)
portabilidade (CI)
publicidade (D)
registro (CI)
sigilo (D)
universalidade (D)
usabilidade (CI)
<i>Categoria: Objetivo</i>
acesso a documentos (D)
acesso ao conhecimento (D)
alteração do conhecimento (CI)
transmissão do conhecimento (D / CI)
geração/ produção de conhecimento (CI)
gestão institucional dos saberes (CI)
produção de documentos (D)
<i>Categoria: Direitos e garantias decorrentes da Informação</i>
direito à privacidade (D)
direito de informar (D)
direito de obtenção de certidões (D)
direito de ser informado (D)
direito fundamental (D)
habeas data (D)
inviolabilidade da honra (D)
inviolabilidade da imagem (D)
inviolabilidade da intimidade (D)
inviolabilidade da vida privada (D)
irrenunciabilidade (D)
liberdade de imprensa (D)
liberdade de informação (D)

Tabela I

Como se pode observar, o Direito atribui poucos processos à informação, se comparado à Ciência da Informação, nomeadamente o acesso, a contestação e a transmissão. Isso revela em que medida essa temática é ainda nova para o Direito, em especial por meio das questões legais relacionadas à transparência administrativa que, por sua vez, impõem um franqueamento do acesso da informação aos cidadãos, o direito do cidadão de contestar uma informação falsa a seu respeito e que possa lhe causar danos, e, ainda, a transmissão da informação, seja como pressuposto do acesso seja, ainda, como processo que, a exemplo da situação anterior, igualmente pode causar

dano ao cidadão. Já no caso da Ciência da Informação, esses processos assumem uma natureza mais operacional, enquanto etapas – muitas vezes sequências – de um continuum ligado à trílogia produção, organização e uso da informação (Guimarães, 2008).

No que tange aos elementos da informação, o Direito volta-se mais especificamente para a informação enquanto documento, ou “informação como coisa” postulada por Buckland (1991), ao se referir a formatos, registros, suportes e aos órgãos produtor e depositário, ao passo que a Ciência da Informação volta-se para aspectos mais intangíveis, seja ligados ao processo comunicacional em si (emissor, receptor, conteúdo, canal, fluxo, mensagem, meios, linguagem, signos, símbolos e sinais), seja a aspectos mais específicos, que se refletem na recuperação tais como os ruídos e a formação de sentido.

A categoria características / atributos / propriedades da informação, por sua vez, assume, no Direito, uma abordagem mais ligada a questões que podem se aquilatar processualmente, tais como caráter vinculante, imprescritibilidade, inalienabilidade, licitude, obrigatoriedade, publicidade, sigilo e universalidade. Já na Ciência da Informação, discute-se se essa informação pode ser, entre outros aspectos, acessível, utilizável, relevante, portátil, inovadora, permanente, necessária, verdadeira ou falsa, presente ou ausente, registrada, ou, reportando-se a Michael Buckland, se pode ser encarada como coisa ou como processo.

Os objetivos da informação, por sua vez, igualmente revelam uma abordagem mais material ou concreta por parte do Direito e mais abstrata da Ciência da Informação, visto que o primeiro se volta prioritariamente aos documentos (produção e acesso) ao passo que a Ciência da Informação centra sua ênfase maior no conhecimento, que pode ser produzido, alterado, gerido e transmitido.

Por último, tem-se, juridicamente, um conjunto de direitos e garantias decorrentes da informação, tais como a privacidade, a obtenção de certidões, o habeas data, a liberdade de imprensa, e a inviolabilidade da honra, da imagem e da intimidade.

Indo além das categorias de análise estabelecidas, observa-se, ainda, na Ciência da Informação, uma recorrente associação da informação com dado (acesso a dados, bancos de dados, dados processados e não processados, assim como inexatidão, retificação e veracidade de dados) e com documento (anotação em assentamentos, espécies documentárias e os processos

de arquivamento, avaliação, classificação, eliminação, gestão, guarda, recolhimento, tramitação e uso de documentos).

Uma análise das categorias apresentadas revela pouca convergência entre o Direito e a Ciência da Informação no que tange à informação. Tal aspecto reitera a complementaridade entre ambas as áreas quando da abordagem dessa temática. Nesse contexto, tem-se, como elementos coincidentes, apenas a informação como processo (o ato de informar-se ou informar a alguém) e a transmissão de conhecimento como objetivo da informação, bem como a necessidade de um suporte para que essa informação tenha materialidade e, portanto, possa ser percebida e utilizada.

No que tange às especificidades de cada área, observa-se, na Ciência da Informação, uma maior ênfase nos processos que incidem sobre a informação, revelando aquilo que Guimarães (2008) denomina como fluxo helicoidal, a partir dos macroprocessos de produção, organização e uso da informação. A isso se alia a presença das dimensões conceituais da informação: como conhecimento, como processo e como coisa (Buckland, 1991).

No caso do Direito, tem-se, por sua vez, uma ênfase nos efeitos ou garantias decorrentes do acesso à informação (por exemplo: *habeas data*, liberdade de imprensa etc.) e a características que, por sua vez, revelam princípios jurídicos inerentes ao direito à informação (por exemplo: irrenunciabilidade, publicidade, universalidade etc.).

6. Considerações finais

Se, por um lado, no âmbito jurídico, a informação é um direito fundamental do ser humano, por outro, o termo é utilizado em diversas acepções, ora referindo-se à informação sobre algo, ora aos elementos de informação ou, ainda, sobre o direito de acesso à informação, dentre outros. Nesse contexto, ainda se mantém uma concepção mais tradicional de informação, eminentemente voltada a sua materialização enquanto documento enquanto a Ciência da Informação vai além, ao abordar questões ligadas ao conteúdo informacional.

Se, no âmbito da Ciência da Informação, ainda não existe consenso sobre o que efetivamente seja informação, alguns elementos comuns podem ser aventados, a partir da pesquisa realizada, tais como o fato de constituir a matéria-prima do conhecimento, de pressupor uma dimensão mais ativa (informar alguém sobre algo) e mais reflexiva (informar-se sobre algo) e, prin-

cipalmente, de pressupor um continuum de processos que vai desde a geração até o uso, tendo na organização seu elemento-ponte.

Referências

- Amora, Antônio Soares (2009). *Minidicionário Soares Amora da língua portuguesa*. São Paulo: Saraiva, 2009.
- Bardin, Laurence (2008). *Análise de Conteúdo*. Lisboa: Edições 70, 2008.
- Berlo, David Kenneth (1997). *O processo da comunicação: introdução à teoria e à prática*. 8ª ed. São Paulo: Martins Fontes, 1997.
- Borko, Harold (1968). *Information Science: what is it?* *American Documentation*. 19:1 (1968) 3-5.
- Brasil (1988). *Constituição Federal (1988)*. Constituição da República Federativa do Brasil. Promulgada em 5 out. 1988. Brasília, DF, Senado Federal, 5 out. 1988. http://www.planalto.gov.br/ccivil_03/constituicao/constituicaoconsolidado.htm (2016-10-14).
- Brasil (2011). Congresso Nacional. Lei nº 12.527 de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. *Diário Oficial da União, Poder Executivo, Brasília, DF, 18 nov. 2011 – edição extra*. http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm (2016-09-18).
- Brasil (2013). Ministério da Ciência, Tecnologia e Inovação. Portaria MCTI nº 293, de 01.04.2013. Institui a Política de Gestão Documental no âmbito do Ministério da Ciência, Tecnologia e Inovação. Brasília, DF, 1 abr. 2013. <http://cnpq.br/gestao-de-documentos> (2016-10-20).
- Buckland, Michael Keeble (1991). *Information as thing*. *Journal of the American Society for Information Science*. 45:5 (1991) 351-360.
- Bush, Vannevar (1945). *As we may think*. *Atlantic Monthly*. 176:1 (1945) 101-108.
- Câmara Junior, Joaquim Mattoso (1959). *Princípios de lingüística geral*. 3ª ed. Rio de Janeiro: Livraria Acadêmica, 1959.
- Cunha, Murilo Bastos da; Cavalcanti, Cordélia Robalinho de Oliveira (2008). *Dicionário de Biblioteconomia e Arquivologia*. Brasília: Briquet de Lemos/Livros, 2008.
- Gil, Antônio Carlos (2014). *Métodos e técnicas de pesquisa social*. 6ª ed. São Paulo: Atlas, 2014.
- Gleick, James (2013). *A informação: uma história, uma teoria, uma enxurrada*. São Paulo: Companhia das Letras, 2013.
- González de Gomez, Maria Néida (1990). *O objeto de estudo da Ciência da Informação: paradoxos e desafios*. *Ciência da Informação*. 19:2 (1990) 117-122.
- Guimarães, José Augusto Chaves (2008). *Ciência da Informação, Arquivologia e Biblioteconomia: em busca do necessário diálogo entre o universo teórico e os fazeres profissionais*. // Fujita, Mariângela Spotti Lopes; Guimarães, José Augusto Chaves (orgs.) (2008). *Ensino e Pesquisa em Biblioteconomia no Brasil: a emergência de um novo olhar*. São Paulo: Cultura Acadêmica, 2008. 33-44.
- Guimarães, José Augusto Chaves; Sales, Rodrigo de (2010). *Análise Documental: concepções do universo acadêmico brasileiro em Ciência da Informação*. *DataGramZero - Revista de Ciência da Informação*. 11:1 (2010) 1-17.
- Kobashi, Nair Yumiko (1994). *A elaboração de informações documentárias: em busca de uma metodologia*. 1994. Tese de doutorado.
- Le Coadic, Yves-François (1996). *A ciência da informação*. Brasília: Briquet de Lemos/Livros, 1996.
- Lenza, Pedro (2015). *Direito Constitucional esquematizado*. 19ª ed. São Paulo: Saraiva, 2015.
- Logan, Robert K. (2012). *Que é informação?: a propagação da organização na biosfera, na simbiosfera, na tecnosfera e na econosfera*. Rio de Janeiro: Contraponto; PUC-Rio, 2012.
- Machado, Ana Maria Nogueira (2003). *Informação e controle bibliográfico: um olhar sobre a cibernética*. São Paulo: UNESP, 2003.
- Melo, José Marques de (1977). *Comunicação social: teoria e pesquisa*. Petrópolis: Vozes, 1977.
- Novelino, Marcelo (2011). *Direito Constitucional*. 5ª ed. São Paulo: Método, 2011.
- Octaviano, Vera Lúcia de Campos; Rey, Carla Monte; Silva, Kelly Cristina da (1999). *A informação na atividade técnico-científica: em enfoque pós-moderno*. *Transinformação*. 11 :2 (1999) 173-184.
- Rapoport, Anatol (1970). *What is information? // Saracevic, Tefko (1970). Introduction to information science*. Nova York: Bowker, 1970.
- Salgado, Eneida Desiree (2015). *Lei de Acesso à Informação (LAI): comentários à Lei nº 12.527/2011 e ao Decreto nº 7.724/2012*. São Paulo: Atlas, 2015.
- Santos, João Carlos Gardini; Guimarães, José Augusto Chaves (2016). *As dimensões temáticas da informação na jurisprudência brasileira: uma análise a partir dos acórdãos do Supremo Tribunal Federal do Brasil*. // Neto, Luísa; Ribeiro, Fernanda (orgs.) (2016). *Direito e Informação na Sociedade em Rede: atas. Faculdade de Direito e Faculdade de Letras da Universidade do Porto: Porto, 2016. 408-418*.
- Semidão, Rafael Aparecido Moron (2014). *Dados, Informação e Conhecimento enquanto elementos de compreensão do universo conceitual da Ciência da Informação: contribuições teóricas*. 2014. Tese de doutorado.
- Silva, De Plácido e (2007). *Vocabulário Jurídico*. 27. ed. Rio de Janeiro: Forense, 2007.
- Smit, Johanna Wilhelmina (1993). *O documento audiovisual ou a proximidade entre as 3 Marias*. *Revista Brasileira de Biblioteconomia e Documentação*. 26:1/2 (1993) 81-85.
- Smit, Johanna Wilhelmina (1999-2000). *Arquivologia, Biblioteconomia e Museologia: o que agrega estas atividades profissionais e o que as separa?* *Revista Brasileira de Biblioteconomia e Documentação*. 1:2 (1999-2000) 27-36.
- Smit, Johanna Wilhelmina; Barreto, Aldo de Albuquerque (2002). *Ciência da Informação: base conceitual para a formação do profissional*. // Valentim, Marta Lígia Pomim (Org.) (2002). *Formação do Profissional da Informação*. São Paulo: Polis, 2002.
- Tognoli, Natália Bolfarini; Guimarães, José Augusto Chaves (2011). *A organização do conhecimento arquivístico: perspectivas de renovação a partir das abordagens científicas canadenses*. *Perspectivas em Ciência da Informação*. 16:1 (2011) 21-44.

Enviado: 2017-03-28. Segunda versão: 2017-07-25.
Aceptado: 2018-09-14.

Confluências filosóficas para os estudos de classificação das artes: aproximações de Bacon, Hegel e Harris

Confluencias filosóficas para la clasificación de las artes: aproximaciones de Bacon, Hegel y Harris
Philosophical confluences for the classification of Art studies: the approaches of Bacon, Hegel, and Harris

Verônica de Sá FERREIRA, Rodrigo de SALES

Programa de Pós-Graduação em Ciência da Informação - PPGCI/UFF, Rua Tiradentes 148, Ingá - Niterói - RJ - Brasil - CEP: 24210-510, saferreira.v@gmail.com. rodrigo.sales.s@gmail.com

Resumen

Analiza las influencias filosóficas de la clasificación bibliográfica de William Torrey Harris, reconocida por la literatura del área, como la fuente más inmediata utilizada por Dewey para la creación de su sistema. Ampliando la reflexión sobre las bases teóricas adoptadas por Harris, se propone cuestionar el lugar ocupado por Francis Bacon en este esquema. El incentivo para la discusión parte del artículo publicado en 1959 por Eugene E. Graziano, que indica el pensamiento hegeliano como el verdadero soporte filosófico del esquema de Harris y, consecuentemente, de la clasificación bibliográfica más utilizada en el mundo, de Dewey. Usando la clase de asuntos destinada a las Artes, Graziano expone evidencias que buscan acercar el pensamiento de Hegel y apartar el de Bacon. En este ensayo, se propone examinar el entendimiento de Arte que sustentó las clasificaciones de Bacon, Hegel y Harris. Los resultados presentados apuntan con claridad a la presencia hegeliana no asumida explícitamente por Harris y Dewey en sus sistemas y aún poco estudiada en el campo de la representación de la información en Brasil. Se concluye, a diferencia de Graziano, que no es posible apartar totalmente el pensamiento baconiano de la estructura de la clasificación bibliográfica de Harris, sin embargo, la presencia de Hegel se confirma como la fuente alimentadora del contenido de la clase Artes. La claridad de la revelación de Graziano, a través de las Artes, amplía las discusiones teóricas dejadas en abierto por Harris y Dewey, evidenciando la importancia del debate.

Palabras clave: Teoría de la clasificación. Organización del conocimiento. Clasificación de las artes. William Torrey Harris. Francis Bacon. G. W. F. Hegel.

1. Introdução

Em tempos de se questionar o determinismo positivista das classificações bibliográficas e de se buscar alternativas teóricas que vislumbrem novas possibilidades de organização e representação do conhecimento, o presente ensaio ousa tomar o caminho inverso e reabrir uma discussão no âmbito das já consagradas perspectivas que

Abstract

This article analyzes the philosophical influences that grounded the William Torrey Harris's bibliographic classification structure, recognized in the area literature as the most immediate source used by Dewey to create his system. Expanding the reflection about the theoretical bases adopted by Harris, it is proposed to question the place occupied by Francis Bacon in this scheme. The encouragement for the discussion starts from the article published in 1959 by Eugene E. Graziano, which points to hegelian's thought as the true philosophical support of the Harris' scheme and, consequently, for the most widely used bibliographic classification in the world, Dewey's. Drawing on the subjects class Arts, Graziano raises evidences that try to approximate the hegelian thought and to remove the one of Bacon. In this essay, it is proposed to examine the understanding of Art that formed the classifications of Bacon, Hegel, and Harris. The results presented clearly indicate Hegel's presence not explicitly assumed by Harris and Dewey in their systems and still little discussed in information representation in Brazil. It is concluded, unlike Graziano, that it is not possible to completely remove Baconian's thought from the structure of Harris's bibliographic classification, however, the presence of Hegel really confirms itself as the feeder source of the Arts class content. The clarity of the revelation pointed out by Graziano, through the Arts, amplifies the theoretical discussions left open by both Harris and Dewey, highlighting the importance of this debate.

Keywords: Classification theory. Knowledge organization. Arts classification. William Torrey Harris. Francis Bacon. G. W. F. Hegel.

servem de aporte para os estudos das classificações de bibliotecas. Neste cenário, a literatura que vem há décadas se ocupando com a história das classificações de bibliotecas no ocidente parece afinada com o discurso de que a classificação das ciências proposta por Francis Bacon no século XVII foi a base forte das principais classificações bibliográficas desenvolvidas no século XIX, especificamente nos Estados Unidos, com

destaque para as classificações de Harris e Dewey (Leidecker, 1946; Sayers, 1955; Eaton, 1959; Mills, 1960; La Montagne, 1961; Barbosa, 1969; Foskett, 1973; Vickery, 1980; Piedade, 1983; Wiegand, 1998). Entretanto, e em que pese estudos desta natureza ocorridos nos Estados Unidos, se por um lado é possível flagrar a hegemonia do crédito dado à classificação de Bacon para as pesquisas relativas às classificações bibliográficas, por outro lado observa-se umas poucas afirmações que questionam o protagonismo da influência de Bacon nas classificações de bibliotecas, e atribuem a Hegel o fundamento mais marcante nas classificações de Harris e Dewey (Graziano, 1959; Olson, 2011).

Dentre estas vozes dissonantes, este ensaio toma como ponto de partida e como inspiração para uma análise mais crítica a afirmação do pesquisador e bibliotecário norteamericano Eugene E. Graziano, para quem a classificação de Harris e a classificação de Dewey pouco ou quase nada tinham a ver com a classificação de Francis Bacon. Segundo Graziano (1959), a classificação de livros de William Torrey Harris, publicada em 1870, fonte principal da classificação de Dewey, era um sistema classificatório de base hegeliana, sobretudo no que se referia às subdivisões das Artes.

Entendendo que a influência de Hegel quase nunca é abordada no ensino e na pesquisa de classificações bibliográficas no Brasil, procura-se aqui examinar mais de perto a pertinência da concepção de Graziano (1959) e avançar um pouco mais na discussão que se refere à classificação dos assuntos atinentes às Artes. Considerando que o texto de Eugene Graziano é praticamente desconhecido na comunidade científica brasileira, adota-se neste ensaio uma estratégia semelhante a que foi empregada pelo autor, ou seja, uma análise comparativa entre as divisões de Bacon, Hegel e Harris para a classe de Artes, no intuito de verificar sua plausibilidade e avançar a discussão em prol dos aportes teóricos que sustentam o debate a respeito das classificações bibliográficas. Toma-se tal postura por entender que a contribuição para o universo da organização e representação do conhecimento pode se dar também pela rediscussão de abordagens já consolidadas pela área, o que, na melhor das hipóteses, pode propiciar uma ampliação no que se refere à compreensão teórica das classificações bibliográficas.

2. Pressupostos de uma comparação

Em artigo de 1959, Eugene E. Graziano afirma que a classificação de livros proposta por William Torrey Harris (1835-1909) – e consequentemente a Classificação Decimal de Dewey –, não

tiveram influência filosófica de Francis Bacon (1561-1626), sendo inspiradas, essencialmente, no pensamento de Hegel (1770-1831). Aponta-se a classificação de Dewey como consequência de Harris, pois estudos acerca das influências sofridas por Dewey na elaboração de seu esquema evidenciam que ordem e lógica de suas classes principais de assunto derivam do sistema de Harris. Segundo Graziano (1959, p.51-52):

The Harris and Dewey Systems have been supposed by some to be related to the classification of Francis Bacon. The evidence is strongly against the validity of such a supposition. Evidence is very strong, if not conclusive, that the philosophy of Hegel is integral to the ordinal relationship of these classes, and that the demand for logic in the order is satisfied only in terms of Hegel's philosophy.

Para assegurar tal conclusão, Graziano (1959) analisa e compara (Figura 1) as divisões das classes Belas Artes, exatamente como ordenadas nas respectivas classificações de Harris, Hegel e Bacon – como afirma.

<i>Harris</i>	<i>Hegel Fine Arts</i>	<i>Bacon Poesy</i>
64. (B) Art		
65. XII. Fine Arts		
a. Architecture	1. Architecture	A. Narrative or Heroic (Epic)
b. Sculpture	2. Sculpture	B. Dramatic
c. Drawing and Painting	3. Painting	C. Allegorical. Fables, mythologies, etc. ²²
d. Engraving and lithography		
e. Pictures		
f. Music	4. Music	
66. XII. Poetry	5. Poetry. ²¹	
69. XIV. Prose Fiction.²⁰		

Figura 1. Comparação de Graziano (1959, p. 51)

A partir desta avaliação, Graziano (1959, p. 51) estrutura o seguinte questionamento:

When we observe further that Bacon makes "Music" and "Architecture" subclasses of "Mathematics"; and that he subsumes most of the fine arts under "Philosophy", there can be little question as to where Harris found his classes. In art, the Hegelian progression is from the art closest to matter, architecture; to poetry and prose, which are the "purest" arts because they are furthest from matter, and the senses, and depend almost entirely upon imagination for medium.

O posicionamento de Graziano, curiosamente, colide de modo significativo com a tradição literária referente à história das classificações de bibliotecas, que defende o pensamento de Bacon como a influência filosófica que deu suporte à organização de livros nestes espaços. De igual maneira, a consideração do autor desafia as próprias afirmações de Harris e Dewey sobre a presença baconiana em seus sistemas.

A citação de Graziano impulsiona este ensaio que buscará compreender o ponto de vista do autor, aproximando as três classes apresentadas na Figura 1. Nosso horizonte serão as ideias expostas nos livros “O progresso do conhecimento” de Francis Bacon (1605 – edição consultada de 2007) e “Curso de Estética: o Sistema das Artes”, de G. W. F. Hegel (1835 – edição consultada de 1997), bem como o artigo “Book Classification”, de William Torrey Harris (1870).

2.1. A arte no sistema de Bacon

Dar-se-á início pelo ponto de divergência apontado por Graziano (1959), a classe Poesia elaborada por Bacon e apresentada da seguinte forma:

<i>Poesia</i>
Narrativa ou Heroica (Épica)
Dramática
Alegórica. Fábulas, Mitologias, etc.

*Tabela 1. Classe Poesia de Bacon.
Trecho da comparação de Graziano*

Entende-se que Graziano (1959), ao elaborar sua análise comparativa sobre a classe Belas Artes (Figura 1), toma como ponto de partida a posição intermediária que estas classes ocupam nos sistemas de Bacon e Harris. Isto explica porque o autor utilizou como referência para a comparação a Poesia de Bacon, e não a Arte ou os assuntos atinentes à esta, que se desmembram predominantemente na Filosofia.

Em sua obra “O progresso do conhecimento”, Bacon explica as divisões da classe Poesia e das demais classes que compõe o esquema por ele criado. Dividida em dois livros, apresenta: no primeiro, “um panorama com questões e entendimento do que é o conhecimento [...] indicando como ele deve ser produzido e para que propósitos devam servir, como parte do ‘progresso da humanidade’” (Shirayama, 2016, p. 46); no segundo livro, traz a classificação dos saberes de maneira detalhada. Realiza um “mapeamento das partes do conhecimento que apontam as condições atuais de desenvolvimento de cada uma. Se as partes estão defectivas, perfeitas ou inacabadas” (ibidem, p. 46).

A classificação de Bacon pretendia ser mais do que uma simples ordenação das ciências e sim um método, guia determinante de como o conhecimento deve ser construído. Isto é, “[...] quais os limites e de que forma deve ser abordada cada parte, pois para ele, o conhecimento, na concepção aristotélica, era algo único e cada parte

formaria um todo amarrado, onde as áreas eram próximas umas das outras” (ibidem, p. 48).

No mapeamento do “aprendizado humano” – como o nomeia –, Bacon estabelece que “as partes do conhecimento humano fazem referência às três partes do entendimento humano, que é a sede do saber: a História à sua Memória, a Poesia à sua Imaginação e a Filosofia à sua Razão” (Bacon, 2007, p. 112). Destas três divisões principais se desdobram as classes que compõe a classificação dos saberes, apresentadas de forma resumida na Figura 2 (em apêndice).

Apesar de Graziano (1959) desenvolver sua análise considerando a classe Poesia de Bacon como equivalente a Arte na classificação de Harris – tendo em comum a posição intermediária que ocupam em ambos os sistemas – observa-se que a palavra “Arte” aparece duas vezes na classificação de Bacon, como destacado na Figura 2. Cabe aqui analisar tais situações.

<p><i>History</i></p> <p>A. Natural History</p> <p> a. Generations</p> <p> b. Praeter</p> <p> c. Arts</p> <p>B. Civil History</p> <p> a. Civil History Proper</p> <p> b. Ecclesiastical History</p> <p> c. Literary History</p> <p>C. Appendix to History</p> <p><i>Poetry</i></p> <p>A. Narrative (Epic)</p> <p>B. Dramatic</p> <p>C. Allegorical. Fables, Mythologies, etc.</p> <p><i>Philosophy</i></p> <p>A. Theology or Divine Philosophy</p> <p>B. Natural Philosophy</p> <p> a. Speculative</p> <p> 1. Physics</p> <p> 2. Metaphysics</p> <p> b. Practical</p>	<p>c. Appendix Mathematics</p> <p> 1. Pure Mathematics</p> <p> 2. Mixed Mathematics</p> <p> (a) Perspectives</p> <p> (b) Music</p> <p> (c) Astronomy</p> <p> (d) Cosmography</p> <p> (e) Architecture</p> <p> (f) Mechanics</p> <p>C. Philosophy of Man</p> <p> a. Human Philosophy</p> <p> 1. Body</p> <p> (a) Medicinal Arts</p> <p> (d) Voluptuary (Liberal) Arts</p> <p> (1) Painting</p> <p> (2) Music</p> <p> 2. Soul and Body related</p> <p> 3. Soul</p> <p> (a) Rational Soul</p> <p> (1) Faculties:</p> <p> a) Logic</p> <p> b) Ethics⁸.</p>
---	---

*Figura 2. Classificação de Bacon resumida
(Graziano, 1959, p. 47, grifo nosso)*

A primeira manifestação da palavra (Figura 2) está subordinada à História (emanação da Memória) e relaciona Arte à História da Natureza.

Bacon divide a História da Natureza em três: natureza em seu curso normal (história das Criaturas), em seus erros ou variações (história das Maravilhas), e alterada ou trabalhada (história das Artes). O filósofo vê a primeira como real e bem feita, mas considera as demais deficientes, pois recebem um tratamento fraco e improdutivo (Bacon, 2007).

Quando discorre sobre o último tipo de História da Natureza – a que interessa a esta análise – Bacon faz uma crítica ao tratamento insuficiente dado a este saber. Ao termo História da Natureza Trabalhada, o autor acrescenta a palavra “Mecânica”, passando a ser “História da Natureza Trabalhada ou Mecânica”.

Sobre o assunto, Bacon afirma existir “compilações de *agricultura*, bem como de *artes manuais*, mas geralmente com desprezo dos experimentos familiares e vulgares” (Bacon, 2007, p. 115, grifo nosso). Isto porque era visto com descrédito “[...] meditar sobre questões mecânicas, a não ser que estas se apresentem como coisas recônditas, raridades e sutilezas especiais” (ibidem, p. 115).

O autor considera este comportamento desdenhoso às questões mecânicas como arrogante e fútil, uma vez que são os exemplos mais simples que proporcionam informações mais seguras e não os mais elevados. Ilustra esta afirmativa apresentando a história do filósofo que “[...] por andar com a vista voltada para as estrelas, caiu na água; pois se tivesse olhado para baixo poderia ter visto as estrelas na água, mas olhando para cima não pôde ver a água nas estrelas” (ibidem, p. 116).

Bacon vê o uso da História da Natureza alterada, trabalhada ou mecânica como primordial à filosofia natural, pois beneficia e enriquece a vida humana, sugerindo soluções “mediante a conexão e transferência das observações de uma arte à prática de outra [...]” (ibidem, p. 117). Tal conduta proporcionaria “uma iluminação mais verdadeira e real sobre as causas e axiomas que até agora se alcançou” (ibidem, p. 117), o que ressalta a crença do filósofo no método empírico-indutivo.

A descrição que oferece desta História, refere-se às Artes Mecânicas, posto que são estas as artes pertinentes ao trabalho humano. De acordo com Bacon (2002, p. 49)

[...] artes mecânicas, que são fundadas na natureza e se enriquecem das luzes da experiência, [...] essas (desde que cultivadas), como que animadas por um espírito, continuamente se acrescentam e se desenvolvem, de início grosseiras, depois cômodas e aperfeiçoadas, e em contínuo progresso.

Antes, vistas como menores e desprestigiadas, as Artes Mecânicas foram resignificadas nos séculos XVI e XVII. “Os procedimentos cotidianos dos artesãos, dos engenheiros, dos técnicos, dos navegantes, dos inventores são elevados à dignidade de fato cultural [...]” (Rossi, 2006, p. 84). A valorização de sua utilidade para o progresso do conhecimento, e o destaque sobre seu valor educativo levaram homens como Bacon, Harvey e Galileu a reconhecerem suas dívidas para com

os artesãos (Rossi, 2006). Isto significa que a colaboração de Bacon não se voltou ao avanço das ciências clássicas, “[...] mas deu origem a um grande número de outros setores científicos que tinham raízes, muitas vezes, em misteres anteriores e estavam ligados a uma nova e diferente avaliação das artes mecânicas e de seu lugar na cultura” (ibidem, p. 37-38). O magnetismo e a eletricidade são exemplos da colaboração de Bacon à ciência.

A segunda manifestação da palavra Arte, está subordinada à parte Filosofia (emanação da Razão).

Inseridas, mais especificamente, dentro de Filosofia Humana, as Artes Voluptuárias (Artes Liberais), estão localizadas hierarquicamente abaixo de Corpo na classificação, conforme Figura 2.

Bacon acredita que a contemplação do homem, em Filosofia, pode ser dirigida a Deus, à Natureza, ou ao próprio Homem, gerando, consequentemente, a filosofia Divina, Natural e Humana ou Humanidade. Isto porque, “todas as coisas estão marcadas e estampadas com este caráter triplice: o poder de Deus, a diferença da natureza e a *utilidade do homem*” (Bacon, 2007, p. 136, grifo nosso).

As Artes Liberais fazem parte da Filosofia humana e, portanto, se voltam ao próprio homem e são úteis a ele. Nesta Filosofia, o indivíduo pode ser visto de forma particular e civil. Bacon (2007, p. 163) acredita que “a Humanidade Particular se compõe das mesmas partes que o homem, isto é, de conhecimentos relativos ao Corpo e [...] ao Espírito”, considerando o corpo tabernáculo do espírito.

O conhecimento concernente ao corpo humano se divide em quatro classes de acordo com seus bens: Saúde, Beleza, Força e Prazer. Assim, os saberes são, respectivamente, “[...] a Medicina ou arte da Cura; a arte do Adorno, que se chama Cosmética; a arte da Atividade, que se chama Atlética; e a arte Voluptuária, que Tácito chama acertadamente *eruditus luxus* [voluptuosidade refinada]” (ibidem, p. 167).

Bacon insere as Artes Liberais (Artes Voluptuárias) especificamente na quarta classe – Prazer – e afirma que sua principal deficiência é a de leis que a reprimam. No trecho abaixo, Bacon (2007, p. 178) deixa explícita sua insatisfação com tal Arte.

Pois, se tem sido bem observado que as artes que florescem naquelas épocas em que cresce a virtude são as marciais, e enquanto a virtude está em pleno vigor são as liberais, e quando a virtude vai declinando as voluptuosas, por isso temo que esta

época presente do mundo esteja um pouco no movimento descendente da roda. Com as artes *voluptuosas* emparelho as práticas *jocosas*, pois o engano dos sentidos é um dos prazeres delas. Quanto aos jogos de entretenimento, os considero incluídos dentro da vida e educação civis.

Também no livro “O progresso do conhecimento”, o filósofo manifesta a injustiça cometida por Platão ao considerar a Retórica como uma arte voluptuosa, “declarando-a semelhante à arte de cozinhar, que estraga as boas carnes saudáveis e as torna insalubres com variedades de molhos para agradar o paladar” (ibidem, p. 218).

Talvez sua crítica se deva ao fato de, atribuindo nova importância às Artes Mecânicas, as Artes Liberais, antes eruditas, tornam-se ilusões.

Finalmente, a Poesia (emanação da Imaginação) – divisão que Graziano (1959) considera como equivalente à Arte em Harris –, de acordo com Bacon (2007, p. 131), é “[...] uma parte do saber quase sempre restringida quanto à medida das palavras, mas em todos os demais aspectos sumamente livre”. Possui o poder de “[...] unir a seu prazer o que a natureza separou, e separar o que a natureza uniu, e desse modo fazer matrimônios ilegais e divórcios das coisas: [...] Os pintores e os poetas sempre tiveram permissão para tomar as liberdades que quisessem” (ibidem, p. 131).

Bacon considera que a Poesia possui dois sentidos: um relacionado às palavras e outro ao conteúdo. O primeiro é, objetivamente, uma característica do estilo e se inclui nas artes retóricas. O segundo constitui “[...] uma das partes principais do saber, e não é outra coisa que história simulada, que pode ser composta tanto em prosa como em verso” (ibidem, p. 131).

O filósofo dá ênfase ao segundo sentido e entende que a Poesia, nesta função de história simulada, possui o papel de satisfazer a mente humana quando a natureza das coisas não oferece esta condição. Isto porque “[...] sendo o mundo em proporção inferior à alma, é agradável ao espírito do homem uma grandeza mais ampla, uma bondade mais perfeita e uma variedade mais completa do que cabe encontrar na natureza das coisas” (ibidem, p. 131). Assim, complementa Bacon (ibidem, p. 132),

[...] a poesia simula fatos e eventos maiores e mais heróicos; [...] porque a história verdadeira apresenta ações e eventos mais comuns e menos permutáveis, a poesia lhes dota de maior raridade e variações mais inesperadas e alternantes. De modo que a poesia serve e contribui à grandeza de ânimo, à moralidade e ao deleite. E por isso sempre se pensou que tivesse em si algo de divina, porque eleva e exalta o espírito, ao submeter as aparências das coisas aos seus desejos enquanto a razão o conforma e o dobra à natureza das coisas.

Bacon (2007, p. 132-133) divide a Poesia em três tipos, conforme seu caráter próprio: Narrativa, Representativa e Alusiva. A primeira se ocupa em contar a história com as licenças da imaginação, “[...] comumente escolhendo como temas as guerras e os amores, raramente os assuntos de Estado, e ocasionalmente o prazeroso e o cômico”. A segunda apresenta a história de forma mais realista, ou seja, “[...] visível, e é uma imagem das ações como se estas estivessem presentes, do mesmo modo que a história o é das ações como são na realidade, isto é, passadas”. Já a terceira tem como objetivo “[...] expressar algum propósito ou ideia particular”. Bacon afirma que a parábola, decorrente da Poesia Narrativa, foi amplamente utilizada no passado, como por exemplo, nas fábulas de Esopo. Era preciso que “todo raciocínio que fosse mais agudo ou sutil que o vulgar se expressasse dessa maneira, porque os homens daqueles tempos queriam ao mesmo tempo variedade de exemplos e sutileza de ideias”. As parábolas tornaram-se, desta forma, as antecessoras das argumentações.

O filósofo observa ainda outra aplicação da poesia parabólica, oposta a anteriormente mencionada. Ao passo que a primeira ocupa-se em expor e exemplificar o que se ensina, a segunda encarrega-se de obscurecer a comunicação. “[...] é quando se envolvem em fábulas ou parábolas os segredos e mistérios da religião, da política ou da filosofia. Vemos este uso autorizado na poesia divina” (Bacon, 2007, p. 133).

No final do século XVI, na Inglaterra, o “[...] tema da poesia criadora do saber se apresenta ligado à visão da poesia e da fábula como *véus* que encobrem verdades muito profundas que os antigos não queriam que aparecessem para evitar que o vulgo as profanasse” (Rossi, 2006, p. 218).

Bacon (2007, p. 135) não assinala na Poesia nenhuma deficiência e acrescenta: “se quisermos ser justos com ela, teremos de dizer que, para a expressão dos afetos, paixões, corrupções e costumes, devemos mais aos poetas do que às obras dos filósofos; e, quanto ao engenho e à eloquência, não muito menos do que às arengas dos oradores”.

Portanto, constata-se que, para Bacon, a Poesia está absolutamente relacionada ao uso da palavra, fazendo parte dela a arte retórica e a história simulada (em prosa e verso). É função da primeira, segundo Rossi (2006), determinar comportamentos, estando rigorosamente ligada à ética e função da segunda, conforme visto, deleitar a mente humana.

Apesar de reforçar a importância que a Poesia possui em contribuir “à grandeza de ânimo, à moralidade e ao deleite” (Bacon, 2007, p. 131) e afirmar que “os pintores e os poetas sempre tiveram permissão para tomar as liberdades que quisessem”, Bacon não une Arte e Poesia na posição intermediária de sua classificação dos saberes – como acontece na classificação de Harris.

O filósofo apresenta o conhecimento como o final de um processo que passa pela Memória (História), Imaginação (Poesia) e Razão (Filosofia). O lugar central, habitado pela Poesia, tem a função de compreender o repositório de experiências obtido pela História, sendo então processado pela Filosofia, gerando assim o conhecimento.

As Artes Mecânicas e Liberais são contempladas no sistema de Bacon subordinadas à História e à Filosofia, respectivamente. Todavia, suas divisões em nada se assemelham às de Belas-Artes determinadas tanto por Hegel quanto por Harris. O mesmo acontece com Poesia, o que confirma, neste ponto, a afirmativa de Graziano (1959) sobre a quase inexistência de influência do pensamento baconiano na classificação de Harris.

2.2. O Sistema das Artes de Hegel

Hegel, diferente de Bacon, não estabeleceu formalmente uma classificação das ciências ou dos conhecimentos conforme consagrados à sua época.

Belas Artes

1. Arquitetura
2. Escultura
3. Pintura
4. Música
5. Poesia

*Tabela II. Divisão das Artes de Hegel.
Trecho da comparação de Graziano*

Para tecer comparação entre as classificações, Graziano (1959) resgatou os cinco gêneros artísticos que Hegel determinou em seu Sistema das Artes, conforme apresentados na Tabela II.

O objetivo de Hegel não era ordenar as ciências de seu tempo decretando-as à “[...] um esquema axiomático de princípios organizadores, pré-concebido e exterior, que as trataria enquanto realidades ‘imóveis’, como se de ‘espécies’ diferentes se tratasse [...]”. Pretendia sim “[...] captar as ciências na sua unidade orgânica, na sua plasticidade dialética, no movimento interno que as

anima enquanto partes necessárias de um todo – a filosofia” (Pombo, 2002).

Poesia e Arte caminham juntas nos gêneros estabelecidos por Hegel, todavia, ao incluí-los na comparação, Graziano (1959) não o faz pela posição intermediária que ocupam em uma classificação, e sim por acreditar que seu conteúdo é o que alimenta a classe Belas Artes de Harris.

O livro “Curso de Estética: o sistema das Artes” apresenta tal Sistema explicando cada gênero e o posicionamento que ocupa na divisão. Esta obra não é um livro de Hegel, mas uma compilação de seus manuscritos e anotações de seus alunos dos quatro cursos que ministrou em Berlim, de 1820 a 1828, sobre Estética. O responsável pela reunião destes textos, em 1835, foi Hotho, aluno de Hegel. No prefácio do livro, Hotho esclarece que “serviu-se em primeiro lugar do manuscrito que Hegel usava para dar aula. Como achou esse material insuficiente, promoveu uma colagem de diversas transcrições de alunos” (Keller, 2011, p. 7).

A edição realizada por Hotho é controversa. Segundo Keller (2011, p. 7) “ele parece ter esquecido a diferença entre editor e autor e ter, com isso, imiscuído pensamentos próprios com os de Hegel”. Keller (2011) elucida que outras transcrições de alunos foram descobertas, e o acesso a elas possibilitou uma visão mais precisa do conteúdo dos cursos ministrados por Hegel. Os novos trabalhos editoriais começaram a surgir apenas na década de 30, todavia a edição de Hotho foi a que obteve maior sucesso. Este ensaio utiliza a tradução realizada em 1997 pela editora Martins Fontes de “Vorlesungen über die Ästhetik”, título original da obra em alemão, editada por Hotho.

Hegel (1997, p. 13) aponta que “a verdadeira classificação deve basear-se [...] na natureza das obras de arte, a qual esgota no conjunto dos gêneros a totalidade dos aspectos e dos momentos inerentes ao conceito de arte”. Sendo assim, cada gênero ocupa uma posição lógica dentro do Sistema.

O primeiro lugar, ocupado pela Arquitetura, a pertence por sua própria natureza. Segundo Hegel (1997, p. 16), a arte tem início por ela, pois em seu princípio, ainda não havia estabelecido, “[...] nem os materiais apropriados nem as formas correspondentes” para representar seu conteúdo espiritual. Isto “[...] a obrigou a limitar-se à simples procura da verdadeira adequação e a contentar-se com o conteúdo e com um modo de representação puramente exteriores”.

O segundo lugar corresponde à Escultura. Para Hegel (1997, p. 17), esta forma artística possui

seu princípio e conteúdo “[...] representados pela individualidade espiritual, tal como ideal clássico, de maneira que a interioridade e a espiritualidade encontrem sua expressão no aspecto corporal imanente ao espírito [...]”. Assim como a arquitetura, a escultura faz uso de materiais pesados em sua totalidade espacial, mas nela, a forma determinada pelo conteúdo representa a vida real, ou seja, é a figura humana “[...] que deve ser a representação exterior adequada da independência divina na sua alma grandiosa e na grandeza tranquila, sem se deixar tocar pelas vicissitudes e limitações da ação, por conflitos e sofrimentos”.

O terceiro e quarto lugar do Sistema, são ocupados pelas artes que Hegel (1997, p. 17) determina como as últimas “[...] que têm por incumbência exteriorizar a interioridade subjetiva”. São elas: Pintura e Música.

A pintura é o primeiro passo do sensível para se aproximar do espírito. Diferente das outras artes, não manipula materiais pesados, suprimindo a realidade da manifestação sensível à aparência, fazendo uso das cores. Isto é, “[...] a pintura para exprimir a profundidade da alma, reduz as três dimensões do volume às duas da superfície e representa as distâncias e as figuras espaciais com o auxílio de aparências produzidas pelas cores” (ibidem, p. 17-18). A figura exterior torna-se a expressão total do interior, mostrando o absoluto não apenas voltado a si, mas também a “[...] subjetividade espiritual, capaz de ver, de sentir, de agir, de contrair relações com o que é diferente dela, e, por conseguinte, exposta à dor, ao sofrimento, acessível, a todas as paixões e a todas as satisfações”. Ao representar este conteúdo, “[...] a pintura deve utilizar, no que concerne à forma, os objetos exteriores em geral, tanto aqueles que encontra na natureza como as manifestações do organismo humano, na medida em que deixem transparecer o espírito”.

A música, diferente da pintura, não possui forma. “O seu elemento próprio é a interioridade como tal [...] que se manifesta não na realidade exterior, mas por uma exteriorização instantânea que se desvanece logo que é surpreendida” (ibidem, p. 18). A subjetividade espiritual, na sua unidade imediata, é o que estabelece seu conteúdo. Utiliza como materiais “[...] os sons e as suas figuras, os seus acordes e separações, as suas combinações, oposições, contradições e mediações, segundo as suas diferenças quantitativas e as suas durações submetidas a medidas especiais, elaboradas com método”. A arte permite que a alma humana absorva música como sentimento puro, desprendida de um suporte.

O gênero que segue a pintura e a música, e ocupa o último lugar no Sistema das Artes, é a Poesia. Descrita por Hegel (1997, p. 18) como a arte da palavra, é compreendida por ele como a “[...] verdadeira arte absoluta do espírito manifestando-se como espírito”. Por sua forma de representação, o filósofo considera a Poesia como a arte total ou a arte em si, pois a palavra é a única com a capacidade de “[...] se apropriar, de exprimir, transformando-o [espírito] em objeto de representação, tudo quanto a consciência concebe e reveste de uma forma que ela encontra em si própria”.

A Poesia vence sob o ponto de vista espiritual, sendo a arte mais rica e ilimitada, mas perde sob o sensível, pois é desprovida “[...] de toda a existência sensível na qual o espírito possa encontrar a forma sensível que lhe convém” (ibidem, p. 19), como fazem as demais artes do Sistema. A Poesia “[...] procura somente tornar acessíveis à intuição e à representação espirituais as significações revestidas de uma forma de proveniência puramente interior”.

Estas cinco artes formam, segundo Hegel (1997, p. 20), “[...] o sistema definido e articulado da arte real e efetiva. Existem certamente, além destas, outras artes, tais como a dança, a arte dos jardins etc., mas artes incompletas”.

Resumidamente, sobre as representações sensíveis, Keller (2011, p. 71) afirma:

A arquitetura é a arte mais pesada, que trabalha a materialidade bruta. A escultura está ainda no domínio das três dimensões, seguindo-se a pintura que já reduz seu material ao plano de duas dimensões. A sequência mostra as duas últimas artes [música e poesia] como artes que se desenrolam precipuamente no tempo.

De acordo com Duarte (2016, p. 6), a música apesar de ser “[...] mais espiritual do que a pintura, porque marca a passagem do domínio da espacialidade para o da temporalidade”, é superada pela poesia, pois, “[...] nela, a sonoridade não vale por si como elemento expressivo, mas está a serviço de um sentido, [...] a sonoridade da língua é apenas o transmissor do elemento espiritual”. Por utilizar a palavra, a quinta arte, Poesia, está menos presa ao sensível tornando-se mais espiritual, sendo “entendida não apenas como a arte mais elevada, mas também como ‘última’ expressão artística, já que no desenrolar dialético da esfera do espírito absoluto a próxima estação já se encontra para além do âmbito artístico [...]” (ibidem, p. 5).

Vale ressaltar que, apesar de Graziano (1959) não incluir em seu quadro comparativo a Prosa após a Poesia na classificação de Hegel, no livro “Curso de Estética: o Sistema das Artes”, a Prosa

está contida no capítulo terceiro, que trata “A Poesia”.

É notório que as Artes se dividem de forma absolutamente diferentes nas classificações de Bacon e Hegel. E, ainda que a Poesia assuma um papel relevante para ambos, isto ocorre de forma distinta. Bacon a utiliza como classe intermediária na progressão do conhecimento, sendo responsável por compreender as experiências da História transmitindo-as para a Filosofia. Em sua classificação, as Artes mecânicas e liberais ficam dispersas e distantes desta função mediadora. Já Hegel, sem pretender classificar o conhecimento, vê a Poesia como a arte mais elevada por sua proximidade com o Espírito, inserindo-a como o último gênero de seu Sistema das Artes, que se trata de “um processo dialético a partir do qual os principais *métiers* artísticos [gêneros da arte], começando pela arquitetura e terminando com a poesia, se interrelacionam, levando em conta características de sua espaço-temporalidade” (Duarte, 2016, p. 5).

2.3. A classe arte de Harris

Em seu quadro comparativo, Graziano (1959) expõe a classe Arte de Harris conforme retirada de seu artigo “Book Classification”, de 1870 (Tabela III):

64. (B) Arte
65. XII. Belas Artes
a. Arquitetura
b. Escultura
c. Desenho e Pintura
d. Gravura e Litografia
e. Imagens
f. Música
66. XII. Poesia
69. XVI Prosa Ficcional

Tabela III. Classe Arte de Harris.
Trecho da comparação de Graziano.

O autor apresenta seu sistema de classificação, elaborado especificamente para organização de livros em bibliotecas, esperando que este seja útil “[...] não apenas aos bibliotecários, mas especialmente aos estudantes de filosofia que desejarem olhar para a gama de inteligência humana contida nos livros” (Harris, 1870, p. 115, tradução nossa).

Harris (1870) afirma que todo esquema de classificação tem como base um sistema filosófico e declara fazer uso das ideias formativas de Bacon no que propõe. Isto significa que, em seu esquema, o autor não aplica

integralmente o mapeamento dos saberes idealizado por Bacon, pois o considera inapto a atender questões específicas da classificação de livros, dado que não foi criado para tal. Ao diferenciar classificação das ciências e classificação de livros, Harris defende que as demandas não são as mesmas para ambas. Enquanto a ciência tem em seu domínio toda existência e, em certo grau, pode ser classificada por seu assunto-objeto, os livros não devem receber o mesmo tratamento, já que podem abordar uma ou mais fases do assunto de uma só vez. “[...] O princípio da classificação [de livros] não é simples, como aquele usado pelos classificadores das ciências – Coleridge, Ampère, Comte, Bacon, e Aristóteles –, mas composto, no qual forma e conteúdo se limitam mutuamente” (Harris, 1870, p. 120, tradução nossa). No entanto, apesar de um esquema de classificação das ciências não atender as necessidades impostas na organização de livros em uma biblioteca, pode sim oferecer base filosófica para um sistema com essa finalidade.

Desta maneira, Harris (1870, p. 115, tradução nossa) utiliza na construção de sua classificação “a distinção fundamental de Bacon [...] das diferentes faculdades da alma [...] Memória, Imaginação e Razão, a partir da qual se realizam os três grandes departamentos de aprendizagem humana, a saber: História, Poesia e Filosofia”. O autor assegura que, apesar de Bacon não ter a intenção de classificar livros, “[...] seu olhar profundo capturou a ideia formativa que distingue as diferentes espécies de livros” (ibidem, p. 115, tradução nossa).

Harris (1870) chama a atenção para mais um fator que impossibilita o uso de forma plena do esquema proposto por Bacon, sua obsolescência em relação ao pensamento vigente no século XIX. Como exemplo, menciona que à época de Bacon “[...] a prosa ficcional tinha se desenvolvido muito pouco, [...] conseqüentemente, Bacon escolheu o nome Poesia para todo o domínio” (ibidem, p. 117, tradução nossa). Já no século XIX, “[...] o domínio [...] da Fantasia ou Imaginação Produtiva é chamado Arte ou Estética. A derivação da palavra Poesia, Poieo [...] contendo significado criativo admiravelmente adaptado para nomear as obras da Imaginação Produtiva” (ibidem, p. 117-118, tradução nossa).

Apesar da inaptidão do sistema de Bacon para classificação de livros, Harris defende seu princípio de divisão como verdadeiro e alega que para aplica-lo em uma biblioteca “[...] é necessário não perder de vista seu espírito, nos detalhes que Bacon dá. Verifica-se que em divisões e seções menores, o conteúdo exerce

uma influência predominante sobre a classificação, enquanto nas divisões principais a forma é o princípio orientador” (ibidem, p. 119, tradução nossa).

As mudanças de pensamentos filosóficos levaram Harris (ibidem, p. 119, tradução nossa) a adaptar as divisões de Bacon ao seu tempo, iniciando pela inversão da ordem por ele proposta. Ao que Bacon ordena História-Poesia-Filosofia, Harris estabelece Ciência-Arte-História. A Ciência ocupa o lugar da Filosofia e vem em primeiro lugar fornecendo “[...] o método e os princípios para o que se segue”.

I. CIÊNCIA dá o departamento de livros em que o sistema consciente prevalece; II. ARTE (Estética) dá o departamento onde a “unidade orgânica” ou sistema inconsciente prevalece; III. HISTÓRIA dá o departamento onde o sistema é determinado pelas relações acidentais, como tempo e lugar.

Enfim, tratando-se da divisão principal Arte, Harris (1870, p. 123, tradução nossa) estabelece o seguinte desdobramento em seu sistema de classificação:

- I. Belas Artes
- II. Poesia
- III. Prosa ficcional

IV. Miscelânea literária, composta por obras retóricas (orações) e ensaios literários que têm uma forma de arte mais ou menos impura, ou estão tão relacionados com obras de arte em seu assunto-objeto que não se separam da classe.

Especificamente, na classe Belas Artes, observa-se a seguinte subdivisão (ibidem, p. 127, tradução nossa):

- a. Arquitetura.
- b. Escultura.
- c. Desenho e Pintura.
- d. Gravura e Litografia, Fotografia.
- e. Imagens (coleções).
- f. Música.

Identifica-se, através do recorte temático da Arte, a expansão da classificação de Harris em relação ao sistema que ele diz se basear, o de Francis Bacon. O autor não apenas reúne os gêneros artísticos sob a classe Belas Artes, como insere a Poesia e a Prosa dentro da divisão principal Arte, o que não ocorre no esquema de Bacon.

Hegel e sua filosofia não são mencionados por Harris em nenhum momento de suas considerações. No entanto, sua proximidade com o filósofo torna-se visível pelo conteúdo que compõe

sua classe Artes, com divisões claramente extraídas do Sistema elaborado por Hegel (Tabela IV), conforme sugerido por Graziano (1959).

Sistema das Artes de Hegel	Classificação bibliográfica de Harris
Belas Artes	65. XII. Belas Artes
1. Arquitetura	a. Arquitetura
2. Escultura	b. Escultura
3. Pintura	c. Desenho e Pintura
4. Música	d. Gravura e Litografia
5. Poesia	e. Imagens (coleções)
	f. Música
	XII. Poesia
	69. XVI Prosa Ficcional

Tabela IV. Comparação Hegel e Harris

3. Avançando um pouco mais a questão

Para sustentar sua proposição e negar a influência de Bacon sob a classificação de Harris, Graziano utiliza a classe Belas Artes de maneira, pode-se dizer, esclarecedora. A presença de Hegel torna-se incontestável em Harris quando são expostas as divisões das classes de ambos os filósofos.

A fim de fortalecer sua justificativa, Graziano (1959) agrega outros fatores. O autor afirma que Harris era hegeliano por admissão e profissão, tendo estudado esta filosofia de 1858 a 1879, tornando-a base de todas as suas atividades. Isto teria levado Harris a agir sob influência de Hegel, mesmo quando, em seu artigo, admite como verdadeiro o princípio da divisão de Bacon, baseado tanto na forma quanto no conteúdo. Ciente da Lógica de Hegel, Harris sabia das considerações do filósofo sobre o assunto. Segundo Graziano (1959, p. 48), Hegel afirma:

Em um livro, por exemplo, certamente não tem influência sobre o conteúdo, se ele for escrito ou impresso, encadernado em papel ou em couro. [...] além de uma forma tão indiferente e externa, o conteúdo do livro é sem forma... O conteúdo da *Iliada*, pode-se dizer, é a guerra de Tróia, e especialmente a ira de Aquiles. Nisto temos tudo, e ainda muito pouco; *Iliada* é a *Iliada* pela forma poética, na qual esse conteúdo é moldado.

Graziano (1959) indica também que as três divisões principais de Harris (Ciência – Arte – História), correspondem e se referem essencialmente, aos mesmos três níveis de conhecimento lógico e ontológico de Hegel (*Begriff – Wesen – Sein*) e não às divisões de Bacon (História – Poesia – Filosofia) invertidas, como declara Harris.

Sobre este ponto de vista, observa-se que Olson (2011) possui um posicionamento diferente. Afirmado que a ordem estabelecida por Hegel é *Sein – Wesen – Begriff*, a autora opõe-se à evidência de Graziano e acrescenta que Harris inverteu não só a ordem de Bacon, como também a de Hegel.

Segundo Olson (2011, p. 7),

a *Lógica* de Hegel enumerava três categorias de existência: Ser (*Sein*), Essência (*Wesen*), e Ideia (*Begriff*). [...] Harris descreveu o conceito de Ser como uma condição ilusória na qual a pessoa percebe a realidade somente como se ela fosse o que aparenta ser – aquilo que é imediatamente percebido e que não requer mediação. [...] o Ser dá espaço à Essência – um estado desiludido no qual uma pessoa reconhece que aquilo que esta aparente não é uma realidade independente, mas os resultados dependentes de várias forças ocultas por trás das ilusões do Ser independente. Assim, a Essência é uma condição na qual o óbvio é mediado por uma compreensão cética. Finalmente, a Essência leva à atividade autônoma da Ideia – uma categoria superior, que é a revelação de um ser superior. [...] A ideia transcende estas formas para se tornar uma revelação da perspectiva teísta que Hegel tinha de Deus.

Apesar do esforço de Graziano, o pensamento baconiano existente em Harris não pode ser absolutamente descartado, ainda que a presença de Hegel seja inegável. Ou seja, é possível que tanto Bacon quanto Hegel tenham proporcionado bases teóricas para elaboração da classificação bibliográfica de Harris.

Ao propor a análise da classe Artes nos três esquemas classificatórios – de Bacon, Hegel e Harris – Graziano (1959) estabelece uma comparação entre três perspectivas distintas. Bacon cria seu esquema filosófico para classificar as ciências, no século XVII; Hegel propõe uma classificação que capte a unidade orgânica e plasticidade dialética das Artes, no século XIX; e Harris elabora um sistema para organização de livros, no mesmo século que o anterior. Tais esquemas podem possuir, portanto, uma estruturação diferente, de acordo com suas funções e épocas em que foram elaborados.

Harris afirma que toda classificação tem base em um sistema filosófico e destaca a importância da forma e conteúdo de Bacon como princípios da classificação. O autor declara fazer uso das ideias formativas de Bacon no esquema que propõe – demarcando as divisões principais que distinguem as diferentes espécies de livros –, mas não esclarece a influência que o direciona quanto ao conteúdo, responsável por alimentar as divisões e seções menores. Graziano (1959) proporciona esta resposta ao abordar em seu

artigo a classificação das Artes, que se apresenta como retrato do Sistema das Artes de Hegel.

Olson (2001) corrobora com este pensamento e menciona que para elaborar seu sistema, Harris interpretou a dialética epistemológica de Bacon e o modelo ontológico da lógica de Hegel.

A Poesia, classe intermediária de Bacon, ocupa no esquema de Harris a mesma posição, todavia aparece acompanhada das Belas Artes, seguindo a mesma ordenação estabelecida por Hegel. Assim como Harris aponta que, no tempo de Bacon, século XVII, a prosa havia se desenvolvido muito pouco e por isso não é mencionada, também as outras Artes não poderiam ocupar o mesmo lugar que dominam na classificação de Harris, século XIX, ficando divididas no sistema de Bacon.

A classe mediana, que para Bacon está ligada à imaginação e para Harris ao sistema inconsciente, conforme afirmam, no esquema de Harris não se limita apenas às palavras, isto é, à Poesia. Percebe-se que as Artes no sistema de Hegel, e conseqüentemente no de Harris, progredem da menor proximidade que possuem com o espírito para a maior, implicando em uma menor participação do sensível.

Segundo Sales e Pires (2017, p. 3, tradução nossa), o reflexo do pensamento baconiano em Harris está na “formação da ideia de conhecimento com base nas faculdades mentais do homem”. O objetivo de Bacon era estruturar um método do conhecimento, e suas divisões principais agem como formas de conhecimento. Para o filósofo inglês, “forma significa caminho e a possibilidade de aproximar o conhecimento”. Por sua vez, o pensamento idealista de Hegel “focado no ‘como’ e no ‘processo’, talvez tenha encontrado na interpretação de Harris o diálogo para a concepção baconiana de ‘forma’ e ‘forma de conhecimento’”.

Sales e Pires (2017) sugerem que um dos motivos de Harris não ter mencionado Hegel como parte de sua influência filosófica se deve ao fato de que a lógica baconiana já havia alcançado amplo reconhecimento acadêmico na época de Harris, tendo sido aplicada, por exemplo, em iniciativas relevantes como o arranjo de classificação da Enciclopédia de Arte e Ciência de D’Alembert e Diderot, na segunda metade do século XVIII, e na classificação de Thomas Jefferson, em 1815.

4. Considerações

A afirmação de que o arranjo das artes na classificação de Harris não seria baseada na classifi-

cação do conhecimento proposta por Francis Bacon, mas sim pela divisão das artes realizada por Hegel (afirmação feita por Eugene Graziano), reabre discussões teóricas que contribuem diretamente para questões de ordem histórica no universo biblioteconômico, visto que o sistema de Harris é uma das portas de entrada para os estudos de classificações bibliográficas no século XIX. A literatura especializada vem ao longo de décadas creditando a Bacon a base teórica que estrutura conceitualmente as classificações de bibliotecas mais influentes na história da biblioteconomia ocidental, tais como a classificação de Harris e as subsequentes classificações decimais de Dewey e Universal. Essa perspectiva tem exercido definitiva influência nos estudos classificatórios levados a cabo nos cursos de biblioteconomia do Brasil, uma vez que é flagrante o fato de que a contribuição de Hegel para a lógica classificatória passa longe das disciplinas vinculadas às classificações bibliográficas.

O presente ensaio, que teve como ponto de partida a afirmação de Graziano (1959), que desloca o referencial teórico das classificações de Bacon para Hegel, procurou trazer à tona o questionamento revelado pelo pesquisador norte-americano e avançar um pouco mais a discussão, de modo a contribuir para a pesquisa na área das classificações no Brasil, visto que nos Estados Unidos, notadamente com Graziano e Olson (2001; 2011), a influência de Hegel já mostra sua presença.

Assim como Graziano (1959), o foco de atenção aqui esteve voltado para a subdivisão da classe Artes nas classificações examinadas. Pode-se constatar que, embora declarada sua base baconiana ao elaborar seu sistema de classificação, Harris não enunciou explicitamente que suas divisões foram fundamentalmente orientadas pela lógica hegeliana. Tal orientação fica clara no fato de que Harris parece ter dado muito mais ênfase ao movimento que vai das artes mais concretas para as mais abstratas, iniciando pela Arquitetura e finalizando com a Poesia, como preferia Hegel, do que privilegiando a função mediadora das artes espalhadas em diversos pontos da classificação de Bacon.

Quando Harris praticamente copia a estrutura das Artes definida por Hegel em seus cursos de Estética, torna evidente que embora não o tenha citado em seu artigo *Book Classification*, Hegel foi sua principal influência para o desdobramento interno dos assuntos de sua classificação. Nesse sentido, é possível afirmar que Harris tomou como ponto de partida as ideias formativas de Bacon para a macroestruturação de sua classificação, organizando-a com base nas faculdades

de Memória-Imaginação-Razão, e sua ordenação interna, ou melhor, o desdobramento de seu arranjo conceitual, se deu essencialmente com base na divisão de Hegel.

É curioso observar que ao desenvolver sua classificação de livros, Harris conseguiu unir duas perspectivas bastante distintas que sequer tinham a proposta de classificar livros. A perspectiva indutiva de Bacon, que tinha como pretensão dar ordem aos saberes produzidos pela humanidade, deu vez a uma dialética empirista que deixava às artes um papel mediador que ligaria a memória e a razão. Bacon fez isso em um período em que a História parecia estar muito mais consolidada como forma de saber do que a própria Ciência. Hegel, por sua vez, trouxe uma perspectiva racionalista fomentadora de uma dialética idealista encabeçada pela razão (pela Ciência e pela Filosofia) no intuito de compreender o processo do saber, do conhecimento.

Bacon está presente na classificação de Harris na definição das três grandes classes – História-Poesia-Filosofia –, ao passo que Hegel está fundamentalmente na ordenação dessas classes – Ciência-Arte-História – e nas subdivisões das artes, da mais concreta (mais sensível) à mais abstrata (mais essencial). Assim, não é descabido afirmar que a classe Arte em Harris se aproxima sensivelmente da categoria Essência de Hegel. Desse modo, não cabe aqui concordar integralmente com Graziano e afirmar que a classificação de Harris “vira as costas” para a lógica de Bacon e se aproxima exclusivamente à lógica de Hegel. Prefere-se aqui entender que Harris conseguiu a proeza de unir filosofias distintas para o universo biblioteconômico das classificações. No entanto, dada a visível convergência entre a classificação de Harris e a divisão proposta por Hegel, entende-se que a alegação de Graziano presta um favor pertinente e relevante aos estudos das classificações bibliográficas, transferindo a esses a responsabilidade de inserir Hegel no ensino e pesquisa da área.

Referências

- Bacon, Francis (2002). *Novum organum*. 2002. <http://www.ebooksbrasil.org/adobeebook/norganum.pdf> (2018-01-11).
- Bacon, Francis (2007). *O progresso do conhecimento*. São Paulo: UNESP, 2007. Tradução Raul Fiker. Título original: *The Proficiency and Advancement of Learning Divine and Humane* (1605).
- Barbosa, A. P. (1969). *Teoria e prática dos sistemas de classificação bibliográfica*. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969. *Obras Didáticas*, 1.
- Duarte, Rodrigo (2016). *Do sistema das artes à ambiência pós-histórica: itinerários da estética contemporânea*. // *Viso - Cadernos de estética aplicada: revista eletrônica*

- de estética. 10:19 (julho-dezembro 2016). http://www.revistaviso.com.br/pdf/Viso_19_RodrigoDuarte.pdf (2018-01-12).
- Eaton, T. (1959). The development of classification in America in The role of classification in the modern American library. Champaign, Illinois: University of Illinois, 1959.
- Foskett, A.C. (1973). A abordagem temática da informação. São Paulo: Polígono; Brasília: UnB, 1973. Tradução de Antônio Agenor Briquet de Lemos.
- Graziano, Eugene E. (1959). Hegel's philosophy as basis for the Dewey Classification Schedule. // *Libri*. 9 (1959) 45-52.
- Harris, WM. T. (1870). Book classification. // *The Journal of Speculative Philosophy*. 4 (1870) 114-128.
- Hegel, G. W. F. (1997). Curso de Estética: o Sistema das Artes. São Paulo: Martins Fontes, 1997. Tradução Álvaro Ribeiro. Título original: *Vorlesungen über die Ästhetik* (1835).
- Hegel, Georg Wilhelm Friedrich (1835). *Vorlesungen über die Ästhetik*. Berlin: Verlag von Dunder und Humboldt, 1835. Edição de D. H. G. Hotho. http://reader.digitale-sammlungen.de/de/fs1/object/display/bsb10045986_00006.html?zoom=0.55 (2018-01-14).
- Keller, Phillip Wilhelm (2011). Estrutura da obra de arte na filosofia de Hegel: análise da estrutura da arte nos cursos de Berlim com relação aos conceitos de organismo, ação e conceito. São Paulo: Universidade de São Paulo, 2011. Tese de Doutorado em Filosofia.
- La Montagne, L. E. (1961). American library classification: with special reference to the Library of Congress. Handen: The Shoe String Press, 1961.
- Leidecker, K. F. (1946). Yankee teacher: the life of William Torrey Harris. New York: The Philosophical Library, 1946.
- Mills, J. (1960). A morden outline of library classification. London: Chapman and Hall, 1960.
- Olson, Hope A. (2011). A potência do não percebido: Hegel, Dewey e seu lugar na corrente principal do pensamento classificatório. // *InCID: R. Ci. Inf. e Doc.* 2:1 (janeiro-junho 2011), 3-15. <http://www.revistas.usp.br/incid/article/view/42331> (2017-05-20).
- Olson, Hope A. (2001). Sameness and difference: a cultural foundation of classification. // *Library Resources & Technical Services*. 45:3 (julho 2001) 115-122. http://polaris.gseis.ucla.edu/gleazer/462_readings/olson_2001.pdf (2017-05-20).
- Piedade, M. A. R. (1983). Introdução à teoria da classificação. 2. ed. Rio de Janeiro: Interciência, 1983.
- Pombo, Olga (2002). O enciclopedismo romântico: Novalis e Hegel. Lisboa, 2002. <http://www.educ.fc.ul.pt/hyper/enc/cap3p7/romantico.htm> (2018-01-13).
- Rossi, Paolo (2006). Francis Bacon: da magia à ciência. Londrina: Eduel; Curitiba: UFPR, 2006. Tradução Aurora Bernardini.
- Sales, Rodrigo de; Pires, Thiago Blanch (2017). The classification of Harris: influences of Bacon and Hegel in the universe of library classification. // *Proceedings of the VI North American Symposium on Knowledge Organization*. Champaign: University of Illinois, 2017. http://www.iskocus.org/NASKO2017papers/NASKO2017_paper_5.pdf (2017-08-16).
- Sayers, W. C. B. (1955). An introduction to library classification. 9. ed. Londres: Grafton, 1955.
- Shirayama, Cristiane de Melo (2016). Francis Bacon e O Progresso do Conhecimento no Início do Século XVII. São Paulo: Universidade de São Paulo, 2016. Dissertação de Mestrado em Filosofia.
- Vickery, Brian C. (1980). Classificação e indexação nas ciências. Rio de Janeiro: BNG/Brasilart, 1980. Tradução de M.C.G. Pirolla.
- Wiegand, W. A. (1998). The Amherst Method: the origins of the Dewey Decimal Classification scheme. // *Libraries & Culture*. 33 (1998) 175-194.

Enviado: 2018-03-30. Segunda versão: 2018-06-09.
Aceptado: 2018-05-28.

Information Design, Information Science, and Knowledge Organization: a domain analysis from the perspective of complexity

*Diseño de información, ciencia de la información y organización del conocimiento:
un análisis de dominio desde la perspectiva de la complejidad*

Natália NAKANO, Daniel MARTÍNEZ-ÁVILA, Maria José Vicentini JORENTE, Mariana CANTISANI

Department of Information Science, São Paulo State University – UNESP, Av. Higyno Muzzi Filho, 737, Marília (17525-900) – São Paulo – Brazil, natinakano@gmail.com, {dmartinezavila|mjjorente}@marilia.unesp.br, mariana.cantisani@gmail.com

Resumen

Este estudio tiene como objetivos analizar las interrelaciones e interdisciplinariedad del Diseño de la Información, Ciencia de la Información y la Organización del Conocimiento; el lugar del Diseño de la Información como disciplina en Brasil; los objetos de estudio comunes y que conectan el Diseño de la Información, la Ciencia de la Información y la Organización del Conocimiento; el área de conocimiento del Diseño de la Información en Brasil según las clasificaciones de CAPES y CNPq; las relaciones con los estudios sociales al igual que con la Ciencia de la Información y la Organización del Conocimiento; y los programas en Ciencia de la Información con cursos sobre Diseño de la Información. Como metodología usamos análisis de dominio y nos centramos especialmente en los estudios históricos y el principio de la división social del trabajo, para analizar el dominio del Diseño de la Información desde la perspectiva de la complejidad. Se concluye que el Diseño de la Información, la Ciencia de la Información y la Organización del Conocimiento comparten algo más que información como objeto común de estudio, ya que comparten contenidos comunes en instituciones de educación superior.

Palabras clave: Diseño de la información. Ciencia de la información. Análisis de dominio. Interdisciplinariedad.

1. Introduction

Historically, human beings have organized and stored information using different media and technologies. The necessity to organize knowledge is an innate urge related to the very process of learning, understanding the world, and survival. As García Marco (1996) pointed out, classification, as an essential part of knowledge organization, is not only a fundamental activity in libraries, archives, and documentation centers. Classification is a key aspect of the human behavior. The activity of organizing and managing information is common ground for different areas of knowledge that were separated in different courses and universities and thus did not escape Descartes' specificity.

Abstract

This paper aims to discuss the interrelations and interdisciplinarity of Information Design, Information Science, and Knowledge Organization; the place of Information Design, as a discipline, in course syllabi and the classifications of sciences in Brazil; the objects of study that are common and connect Information Design, Information Science, and Knowledge Organization; the area of knowledge of Information Design in Brazil according to the CAPES and CNPq classifications; the relation of Information Design to social studies, Information Science, and Knowledge Organization; and programs in Information Science offering courses in Information Design. As a methodology, we use a domain-analytical approach focusing on historical studies and the principle of social division of labor to discuss the Information Design domain from the perspective of complexity. We conclude that Information Design, Information Science, and Knowledge Organization share more than information as a common object of study. They also share common contents in higher education institutions.

Keywords: Information design. Information science. Knowledge organization. Domain analysis. Interdisciplinarity.

According to Morin (1977), a new knowledge of organization is able to create a new organization of knowledge once the explanatory, reductionist, and atomistic paradigm is replaced by a new paradigm of interrelations. In this context, raised by complexity, our paper aims to discuss the interrelations of Information Design (ID), Information Science (IS), and Knowledge Organization (KO) using domain analysis (Hjørland & Albrechtsen, 1995; Hjørland, 2002, 2017a) and complexity theory as a theoretical framework.

The linearity of disciplines in the social division of labor - a recognized approach to establish a knowledge domain - separated the areas of knowledge in institutions and in courses as a didactic way to present knowledge. Previously,

people associated in guilds based on the technologies that were available and the knowledge and practices that were shared. This was before knowledge was gathered, organized, and transmitted in universities and scientific institutions. The division and simplification of knowledge in disciplines served for practical and rational purposes. However, in order to understand knowledge, we need to embrace complexity in opposition to the Cartesian view of the division of knowledge.

The positivist and reductionist paradigm defended by Descartes and Newton prevailed at the historical moment in which assembly lines emerged, when the industrial and mechanist views of the world prevailed in society and these views were important for the advancement of science and philosophy. Then, science progressed at a rapid pace. The specialist Cartesian model privileged the division of the whole to study its parts, as well as the possibility of conducting analyses that are more and more specific. This subdivision assumes that the whole is the sum of all its parts, and that assembling the parts together is enough to make the whole. The perception here is that the complexity of systems for science is not a simple thing, as the term itself is associated with the meaning of something complicated, difficult, and disorganized. However, complexity should be understood as something that is woven together, as a fabric, impossible to be completely understood unless all the parts are associated and considered.

Complexity should be mainly understood by the notion that the sum of the parts (the whole) is always greater than the total of individuals, simply because a new, often unforeseen, phenomenon emerges from this sum. Information is constituted by a network of relations, such as a story that expands from the inside out through the perceptive phenomena. Its theorizations are part of a global movement in science, started by Ludwig von Bertalanffy's theory of systems, in which the interdependent parts create something larger than their sum.

In complexity, there is a process of re-signification, reordering the spheres of information communication in a continuous cycle among order creation, patterns and beliefs, and ruptures of the cultural codes and representations in sign systems. This is of interest as intersections occur among the information and knowledge media (mediators), digital objects, and the subjects of interaction (netizens).

The communicative way introduced by Information and Communications Technology enables a broader sensorial involvement of the indi-

viduals with the medium that prolongs a set of human senses and requires a sensorial interaction between the netizens and the communication channels they engage. This results in cognitive changes that are consequence of the processes of perception and association activated by the convergence of multiple languages.

These multiple languages trigger numerous forms of representation of knowledge and information that are interrelated to the netizens' organizing and interpretative mental sets, which constitute the combinations for situations of varying degrees of complexity.

Information Science alone cannot cope with the new needs that result from the interactions in electronic communications. By appealing to interdisciplinarity with Design, new features emerge, offering solutions to the new problems. This article aims to situate Design and Information Design as an interdisciplinary domain within IS that can provide answers to those problems that IS alone cannot handle.

Although separated areas of expertise, IS is an interdisciplinary science concerned with information, and information is the main object of study for ID. Capurro, in "What is information Science for?" (1992), suggests that the focus of IS studies, based on information technology studies, is closely related to the possibilities that these technologies have in relation to the physical (bodily) capabilities of the users. This does not imply, however, that the focus of IS is just the assessment of the usability and ergonomic design (i.e., the structural aspects) of information systems. Capurro, as well as Orna and Stevens (1991), observed and outlined a relationship/connection between IS and ID, taking into account all the dimensions of human existence going beyond the aesthetics and bodily aspects, also including the perception and the behavior of the individuals as a whole.

ID emerges as a discipline that deals with project issues, including aspects that go beyond the structural ones such as the organization of information in digital and physical spaces, coping with representation in a tridimensional way, producing meaning, and understanding through language, signals, words, and shapes. ID seeks to process vast amounts of information, especially in digital environments.

IS and KO are in the same context of ID: one goal of IS is to organize information resources, so that users will be able to conveniently access relevant information, satisfy their needs, and construct new knowledge. On the other hand, Knowledge Organization (1) has been defined as "the domain in which the order of knowledge is both the primary paradigm for scientific investigation and the

primary application in the development of systems” (Smiraglia 2012, p. 225). The design of systems, not only their development, has been identified as one of the key challenges and opportunities for Knowledge Organization as a field (Martínez-Ávila et al., 2014; Martínez-Ávila, 2015). As design is part of KO, it is reasonable to think that ID can have many other aspects in common with KO too, or that both are subfields of the same field. In this vein, according to Orna and Stevens (1991, p.197), Information Design can be broadly understood as “everything that we do to make ideas visible so that others can make them their own and use them for their own purposes”. This is actually a process that involves signs representing knowledge.

Since information is a common aspect of ID, IS, and KO, it is in this context that we ground the research questions of this study: What is the place of Information Design, as a discipline, in Brazil and in the academic programs? What are the objects of study that connect and are common to ID, IS, and KO? What is the area of knowledge of Information Design in Brazil according to the CAPES (2) and CNPq (3) classifications? Is ID related to Social Studies (such as Information Science and KO)? Are there IS offering courses on ID in Brazil and in the United States? By answering these research questions, we aim to clarify the interrelations between Information Design, Information Science, and Knowledge Organization.

As a methodology, we draw on Hjørland’s domain analysis and the historical studies and the principle of social division of labor to discuss the Information Design domain. Hjørland (2002; 2017a) describes the special competencies of library and information specialists and information scientists from the domain analytic point of view. In these papers, Hjørland presents eleven (plus three) specific approaches to domain analysis that can assist the definition of the specific competencies of information specialists. These approaches are: 1. Producing and evaluating literature guides and subject gateways; 2. Producing and evaluating special classifications and thesauri; 3. Research on and competencies in indexing and retrieving information in specialties; 4. Knowledge about empirical user studies in subject areas; 5 Producing and interpreting bibliometric studies; 6. Historical studies of information structures and services in domains; 7. Studies of documents and genres in knowledge domains, 8. Epistemological and critical studies of different paradigms, assumptions and interests in domains; 9. Knowledge about terminological studies, LSP (languages for special purposes) and discourse analysis in knowledge fields; 10. Knowledge about and studies of structures and institutions in scientific and

professional communication in a domain; 11. Knowledge about methods and results from domain analytic studies about professional cognition, knowledge representation in computer science and artificial intelligence. 12. Database semantics. 13. Discourse analysis. 14. Knowledge about the provenance.

Hjørland (2012) also studied the relation between several “fields” or “tags” in IS, namely Information Organization (IO), Organization of Information (OI), Information Architecture (IA), and Knowledge Organization (KO), using bibliometric methods as part of his domain analytic approach. Following a historical approach, we also aim to reveal the nature and structure of ID within IS through and in relation to KO. Are there overlapping contents? Is ID part of or related to KO? Does ID belong to a different field than IS? To study the ID domain, first, our paper focuses on the history and implementation of courses on Design, especially in Brazil; second, we analyze the contents of courses offered by some IS programs in Brazil and the United States (as a benchmark against the Brazilian courses) in order to establish the interrelation/interdisciplinarity between the courses and the analyses of the domains.

The value of this paper lies in the continuation of a discussion started by Orna and Stevens in 1991 that, although cited by Capurro in 1992, was overlooked within the field, and also in the recognition of the domain of Information Design as an own subfield within Information Science, different from Information Architecture.

2. The history of Design in Brazil, Information Science, and Knowledge Organization

In the 1950s, Europe, the United States, and other countries such as Brazil shared a climate of effervescence and optimism about a promising future after World War II. The economic stability and growth provided a fertile environment for the investment in culture and arts. In Brazil, the São Paulo Museum of Art (MASP) was created in 1947, and in 1951, with the initiative of the same architect responsible for the MASP project, Lina Bo Bardi, the Institute of Contemporary Art (IAC) opened. The IAC, although short-lived for only three years due to a lack of funds, was the seed of the Brazilian higher education in Design and a key element of its history. In this matter, it established the philosophy that prevailed in the formal higher education in the country - the School of Ulm.

The development and implementation of courses in Design in Brazil and in the world is closely linked to the developments in industrial design,

i.e., the production of objects in industry. During the 1950s, influenced by the United States, the activities of industrial design began to be used in Brazil. In this vein, the Portuguese term “desenho industrial” (industrial design) - considered mistaken by many for being too restrictive (Niemeyer, 1998) - was established as a synonym for Design (4). The English term “design” comes from the Latin term ‘designāre’, meaning to designate, to develop, to conceive; it has the sense of “to project” and the same etymological root of the terms “to wish” and “to name”. According to Jorente (2014, p. 117):

The expression design (used with the meaning of creation of objects developed for production through modern industrial means and post industrial contemporary means) appeared in the eighteenth century, in England, as a translation of the Italian term disegno to denote activities linked to the production of objects in the Industrial Revolution. At that moment, design becomes a discipline.

In 1962, the Faculty of Architecture and Urbanism (FAU), at the University of São Paulo (USP) included a course on Design as a “result of an evolutionary process of 14 years” (Niemeyer, 1998, p. 65). This course was taught by João Batista

Vilanova Artigas, a professor of aesthetics, composition, and urban planning (5). Artigas believed that architecture should solve the design problems in a historical moment in which Brazil was establishing nationalistic, developmental, and positivist characteristics of progress.

In this context, the courses Industrial Design and Visual Communication were included, focusing on Industrial Design. FAU's proposal was not followed by other schools and this caused a split among professionals that remains today. On the one hand, architects from São Paulo “advocate for themselves the responsibility for the development of design projects” (Niemeyer, 1998, p. 67), and, on the other hand, designers, who took courses in design all over the country “reject this prerogative of architects” (Niemeyer, 1998, p. 67).

This historical event in the implementation of the higher education in Design in Brazil explains why the CAPES classification places “Desenho Industrial” under the area of Architecture and Urbanism, while the CNPq classification places “Desenho Industrial” as a discipline in the area of Social Sciences, as shown in Figure 1.

CNPq	CAPES
6.00.00.00-7 Ciências Sociais Aplicadas	ÁREA DE AVALIAÇÃO: ARQUITETURA E URBANISMO
6.07.00.00-9 Ciência da Informação	ARQUITETURA E URBANISMO
6.07.01.00-5 Teoria da Informação	60401001 FUNDAMENTOS DE ARQUITETURA E URBANISMO
6.07.01.01-3 Teoria Geral da Informação	60401010 HISTÓRIA DA ARQUITETURA E URBANISMO
6.07.01.02-1 Processos da Comunicação	60401028 TEORIA DA ARQUITETURA
6.07.01.03-0 Representação da Informação	60401036 HISTÓRIA DO URBANISMO
6.07.02.00-1 Biblioteconomia	60401044 TEORIA DO URBANISMO
6.07.02.01-0 Teoria da Classificação	60402008 PROJETO DE ARQUITETURA E URBANISMO
6.07.02.02-8 Métodos Quantitativos. Bibliometria	60402016 PLANEJAMENTO E PROJETOS DA EDIFICAÇÃO
6.07.02.03-6 Técnicas de Recuperação de Informação	60402024 PLANEJAMENTO E PROJETO DO ESPAÇO URBANO
6.07.02.04-4 Processos de Disseminação da Informação	60402032 PLANEJAMENTO E PROJETO DO EQUIPAMENTO
6.07.03.00-8 Arquivologia	60403004 TECNOLOGIA DE ARQUITETURA E URBANISMO
6.07.03.01-6 Organização de Arquivos	60403012 ADEQUAÇÃO AMBIENTAL
6.11.00.00-5 Economia Doméstica	60404000 PAISAGISMO
6.12.00.00-0 Desenho Industrial	60404019 DESENVOLVIMENTO HISTÓRICO DO PAISAGISMO
6.12.01.00-6 Programação Visual	60404027 CONCEITUAÇÃO DE PAISAGISMO E METODOLOGIA DO PAISAGISMO
6.12.02.00-2 Desenho de Produto	60404035 ESTUDOS DE ORGANIZAÇÃO DO ESPAÇO EXTERIOR
6.13.00.00-4 Turismo	60404040 PROJETOS DE ESPAÇOS LIVRES URBANOS
7.00.00.00-0 Ciências Humanas	61200000 DESENHO INDUSTRIAL
7.01.00.00-4 Filosofia	ÁREA DE AVALIAÇÃO: CIÊNCIAS SOCIAIS APLICADAS I
	60700009 CIÊNCIA DA INFORMAÇÃO
	60701005 TEORIA DA INFORMAÇÃO
	60701013 TEORIA GERAL DA INFORMAÇÃO
	60701021 PROCESSOS DA COMUNICAÇÃO
	60701030 REPRESENTAÇÃO DA INFORMAÇÃO

Figure 1. “Desenho Industrial” according to CAPES and CNPq classifications

There was a split among professionals working with ID: architects, on one side, and designers, on the side. Correspondingly, the analysis of the courses in Design offered by higher education institutions in Brazil reveals that, currently, the main courses on ID are focused on project development, on the one side, and contents such as graphic design, visual communication, user experience, and user studies, on the other side. Currently, ID is being concerned with the development of digital environments which are able to

provide effective and pleasurable experiences for the agents interacting with the system. Information retrieval and findability can be seen as consequences of ID features that follow principles and methodologies proposed by the field (Jacobson, 2000; Horn, 2000).

As for IS, both CAPES and CNPq classifications place it under Applied Social Sciences. According to some authors such as Pinheiro and Loureiro

(1995), initial discussions on the origin and foundations of this new field began in the 1960s, when the first concepts and definitions were drawn from important landmarks such as: the conference in Georgia Institute of Technology in 1962; Mikhailov's paper "Informatika" in 1966, and the classic Borko's definition in his article "Information Science: what is it?" in 1968. However, as several authors have pointed out and discussed (Spang-Hanssen, 2001; Hjørland, 2017a; Martínez-Ávila, 2018), already in the late 1940s and 1950s, the popularity of Shannon's Mathematical Theory of Communication (Shannon, 1948; Shannon & Weaver, 1949) - also misnamed theory of information - made people in the field of Library Science/Documentation embrace this new label in an attempt to gain prestige. According to this interpretation, information science would be just a mere new name for a field that had been already fighting for its scientific status for several decades (Martínez-Ávila & Guimarães, 2015).

Referring to universal laws that represent the informational phenomena, and thus seeking to cope with the manifestations of the individuals (users), it is true that it was arguably mainly during the 1970's that IS began to refer to mathematical, physical, and biological models. From the moment Information Science was accepted as a discipline, it is accepted to be based on a body of knowledge and on professional practices that seek to cope with a large volume of information recorded in diverse media. More generally, and over several decades, IS has been based on a wide, scattered, and often contradictory set of academic literature devoted to the information society. Although the premises and interpretative frameworks favored by different scholars and schools of thought, in what is sometimes referred to as the information society studies, vary, the common intention has been to put the spotlight on the social dimensions and the branches of computerization (Cronin, 2008).

As for KO, despite having some autonomy as a discipline, field of study, or a new science (Dahlberg, 1995; 2006; Hjørland, 2008), it is also inserted in the area of Library and Information Science (Hjørland, 2017a), as it is concerned with the order of knowledge and the construction of systems to organize knowledge. KO and classification have played a key role in the birth and transformation of IS at its different moments, from Librarianship as a practice to organize materials and its transformation into a science with Melvil Dewey and his method for the development of the Dewey Decimal Classification, to the adaptation and use of the Universal Decimal Classification by Paul Otlet and Henry LaFountain in the Documentation movement, as well as in the newer and

vague concept of Information Science (Martínez-Ávila & Guimarães, 2015).

According to Hjørland (2002), to solve the problems related to the progression of the studies in IS, it is necessary to formulate theoretical principles or seek empirical justification for the decisions made in the studies related to information seeking, representation, and retrieval, in order to change the frameworks that are normally atomistic.

In this same line, in "Deflating Information: from Science Studies to Documentation," Bernd Frohmann (2004) states that it is necessary to shift the focus from cognitive processes to labor processes, to reveal the scientific work as the construction of sets of things, people, apparatus, and social relations, a mix of material and immaterial elements. It is necessary to have a holistic view of IS, the area needs to build specific theories related to knowledge domains.

IS, and KO within, is an area that deal with the organization and representation of knowledge and takes into account the scientific questions and the professional activities related to the communication of knowledge within the social, institutional, and individual contexts (Buckland, 1999). Therefore, IS is considered an integral part of the social issues. This aspect of IS is also common ground with ID. We consider that the hybridization occurring in types of media and subjects and institutions create systemic emergences of Design.

According to Knemeyer (2003) "information design ostensibly comes down to a broad set of information deliverables, not any single type or particular component of other disciplines. Information Design serves as a resource for other disciplines engaged in the creation of better understanding and the building of human knowledge." As in the cases of KO and IS, ID deals with the problems related to the vast amount of information, as it aims to provide, define, and plan the contents of messages and digital environments. ID focuses on the intention of making agents interact with a digital environment in order to achieve objectives related to the agents' needs, the way they interact and represent information, and the design that supports the goals of the agents and the designers. Given this, our hypothesis is that ID emerges as a discipline that complements IS as an answer to the of the problems of the increasing amount of information caused by the advent of the Internet that Computer Science alone cannot solve, as physical, cognitive (intellectual), and affective aspects must be addressed when dealing with information (aspects that are addressed by ID).

To test our hypothesis, we searched for courses in Computer Science programs that are related to

both Information Science and Information Design. We also searched for IS programs that offer courses related to ID. The results are presented in the next section.

3. Results

In both CAPES and CNPq classifications of areas of knowledge (see Figure 2), the areas of Database (“Banco de dados”), Information Systems (“Sistemas de Informação”), and Graphics (“Pro-

cessamento Gráfico (Graphics)”) are placed under Computer Science. These subareas also share common ground with ID, as these subareas and ID deal with information organization in courses addressing Interaction Interfaces, User Experience, and Ergonomics, among others. Therefore, it is Computer Science who provides IS with the knowledge on informational and organizational systems, although making use of principles and methodological approaches from Information Design, thus revealing the interface of the interrelation between ID and IS.

CAPES	CNPq
ÁREA DE AVALIAÇÃO: CIÊNCIA DA COMPUTAÇÃO	
10300007 CIÊNCIA DA COMPUTAÇÃO 10301003 TEORIA DA COMPUTAÇÃO 10301011 COMPUTABILIDADE E MODELOS DE COMPUTAÇÃO 10301020 LINGUAGEM FORMAIS E AUTÔMATOS 10301038 ANÁLISE DE ALGORÍTMOS E COMPLEXIDADE DE COMPUTAÇÃO 10301046 LÓGICAS E SEMÂNTICA DE PROGRAMAS 10302000 MATEMÁTICA DA COMPUTAÇÃO 10302018 MATEMÁTICA SIMBÓLICA 10302026 MODELOS ANALÍTICOS E DE SIMULAÇÃO 10303006 METODOLOGIA E TÉCNICAS DA COMPUTAÇÃO 10303014 LINGUAGENS DE PROGRAMAÇÃO 10303022 ENGENHARIA DE SOFTWARE 10303030 BANCO DE DADOS 10303049 SISTEMAS DE INFORMAÇÃO 10303057 PROCESSAMENTO GRÁFICO (GRAPHICS) 10304002 SISTEMA DE COMPUTAÇÃO 10304010 HARDWARE 10304029 ARQUITETURA DE SISTEMAS DE COMPUTAÇÃO 10304037 SOFTWARE BÁSICO 10304045 TELEINFORMÁTICA	1.03.00.00-7 Ciência da Computação 1.03.01.00-3 Teoria da Computação 1.03.01.01-1 Computabilidade e Modelos de Computação 1.03.01.02-0 Linguagem Formais e Automatos 1.03.01.03-8 Análise de Algoritmos e Complexidade de Computação 1.03.01.04-6 Lógicas e Semântica de Programas 1.03.02.00-0 Matemática da Computação 1.03.02.01-8 Matemática Simbólica 1.03.02.02-6 Modelos Analíticos e de Simulação 1.03.03.00-6 Metodologia e Técnicas da Computação 1.03.03.01-4 Linguagens de Programação 1.03.03.02-2 Engenharia de Software 1.03.03.03-0 Banco de Dados 1.03.03.04-9 Sistemas de Informação 1.03.03.05-7 Processamento Gráfico (Graphics)

Figure 2. Subareas under Computer Science interrelated with Information Design

Regarding higher education institutions that offer programs in Information Studies and courses related to ID, we selected a sample of eight institutions in the United States (University of Colorado – Boulder, Rutgers – New Jersey, Palmer School, University of Kentucky, University of Pittsburgh,

University of British Columbia – Vancouver Campus, University of Wisconsin – Milwaukee, and University of Oklahoma) and one institution in Brazil (Federal University of Paraná, Brazil) (see Figure 3).

University	Program	Courses that include Design
University of Colorado - Boulder	Information Studies	1) Human centered design 2) Information visualization 3) Visual Design 4) Information Architectures
Rutgers – New Jersey	Information Studies	1) Designing user centered Information services 2) Interface Design 3) Database design management 4) Understanding, designing and building social media
Palmer School	Library and Information Science	1) Instructional Design and leadership 2) Web Design and content management systems 3) Human-computer interaction
University of Kentucky	Communication and Information – School of Information Science	1) Information Systems Design 2) Information Architecture
University of Pittsburgh	Information Studies	1) Human information interaction 2) Information Architecture

		3) Database design and applications 4) Information visualization
University of British Columbia – Vancouver Campus	Library and Information Studies	1) Database Design 2) Information Design I 3) Systems, Information Design II
University of Wisconsin - Milwaukee	School of Information Studies	1) Web design 2) Human Factors in information seeking and use 3) Multimedia web design
University of Oklahoma	Library and Information Studies	1) Knowledge Organization and Information resources (design and structure of information system) 2) Information users in the Knowledge society (practical methodologies for study of uses and for user-centered design of information and knowledge systems and services) 3) Information technology management (Fundamentals of planning, designing, implementing and managing information technology solutions) 4) Design and Implementation of Web-based information services
Federal University of Paraná, Brazil	Information Management	1) Information Ergonomics, 2) Information Design, 3) Infometrics, 4) Data Mining, 5) Database 6) Information and Organizational Culture 7) Knowledge Management

Figure 3. IS programs and courses interrelated with ID

Most American institutions include in their contents technical aspects related to databases, organization of information on the Web, and Information Architecture. University of Colorado – Boulder includes contents on Information Visualization and Information Architecture. It should be noted here that Information Architecture, although not unanimously considered part of Knowledge Organization, it is a discipline that is sometimes taught and included in KO (e.g., Taylor and Joudrey, 2009), and was also discussed by Hjørland (2012) in relation to KO.

Rutgers includes contents on databases and interfaces. Database is also a content commonly included in manuals of KO (e.g., Taylor and Joudrey, 2009; Rowley and Hartley, 2008), while interface design is defined in this context by Taylor and Joudrey as “the part of a system design that controls the interaction between the computer and the user” (p.460). If KO is primarily concerned with the construction of systems as reported by Smiraglia and others, then interface design might also be an aspect related to KO.

The University of Kentucky includes contents on Information Systems Design and Information Architecture. University of Pittsburgh also includes contents on Information Architecture, Database Design and Applications, and Information Visualization. University of British Columbia – Vancouver Campus includes contents on Database Design, and Systems, Information Design.

The course Electronic Publishing & Web Design, offered by the School of Information Studies at the University of Wisconsin-Milwaukee (UWM), presents an introduction to principles of visual communications related to electronic media with emphasis on website development, electronic documents, and production and dissemination of electronic information. In addition to the comprehension of the basic principles of systems development, websites, and electronic publications, the course also includes readings that refer to design principles and elements, and aspects related to data visualization, database, and color theory, among others. In this vein, we also verify a direct relation between IS and ID.

The University of Oklahoma also includes concepts related to IS such as information services and information technologies, and also to knowledge, such as design and structure of information systems, user-centered design of knowledge systems and services.

At the Federal University of Parana, the contents of the course Information Design address the contextualization and assessment of the syntactic and semantic aspects associated to information systems and the processes of information acquisition and dissemination in analog and digital communication media. Therefore, we verify that ID in Brazil emerges as a discipline within IS and a subarea in the social sciences. This course

in Brazil also covers aspects related to databases, data mining, and knowledge management. Although Knowledge Management is not usually included among the objects of study of KO as a domain (at least in the context of ISKO), it is indeed an aspect that it is discussed in some manuals on KO such as Taylor and Joudrey's (2009).

4. Conclusion

As suggested by Popper (1973), the concern of research in science is related not only to the areas of knowledge or the object of study, but also with problem solving. It is important to point out that problems arise from changes that take place in different contexts (historical and cultural) within an area. In this context, with the advent of the Internet, IS is one of the sciences that has been constantly re-shaped and in search for solutions for the problems that arise from the overload of information. Similarly, it has been argued that KO has been, or could be, a fundamental process in every moment and transformation of the World Wide Web discourse, and design might be another key area/opportunity for it (Martínez-Ávila, 2015).

The analysis of some ID courses in Brazil and in the United States suggests that, from the intersections between the domains of IS, ID, and Computer Science, there is an emergence of problems that stems from the complexities of the phenomenon of communication and the objectification of knowledge. This paper proposes that ID acts in the IS domain in the context of these new studies in order to address complex issues, such as the problems that involve human-information interaction, information processing, horizontal indexing, and collaboration, among others. In this sense, some of the contents of Information Studies courses that include characteristics from ID and Computer Science are Database Design, Web Design and Content Management Systems, and Information Systems Design.

Finally, this study concludes that ID, IS, and KO share more things than just information as a common object of study. Further research on the contents that are being taught in the courses might be helpful to determine the characteristics of the domains.

Notes

(1) Knowledge Organization has been discussed both as a subfield of (Library and) Information Science and as an independent field (see Hjørland 2017a, p.453). We are not discarding both possibilities for Information Design as well beforehand, as we are trying to investigate its relationship with both Information Science (perhaps as a subfield) and with Knowledge Organization.

- (2) CAPES stands for Coordination for the Improvement of Higher Education Personnel. It depends on the Brazilian Ministry of Education and it aims to coordinate efforts to improve the quality of Brazilian faculty and staff in higher education.
- (3) CNPq stands for National Council for Scientific and Technological Development. This agency belongs to the Ministry of Science and Technology and it is dedicated to the promotion of scientific and technological research.
- (4) There is not a word for "design" in Portuguese. The Portuguese word "desenhar" means "to draw".
- (5) "Design" is sometimes translated to Portuguese as "projetar" ("to project").

References

- Borko, H. (1968). Information Science: What is it? // *American Documentation* 19:1 (1968) 3-5.
- Buckland, Michael (1991). Information as thing. // *Journal of the American Society of Information Science* 42:5 (1991) 351-360.
- Capurro, Rafael (1992). Foundations of Information Science: review and perspectives. <http://www.capurro.de/tampere91.htm> (2016-03-11).
- Cronin, B. (2008). The sociological turn in information science. // *Journal of Information Science* 34:4 (2008) 465-475.
- Dahlberg, Ingetraut (1995). Current trends in knowledge organization. // García Marco, Francisco Javier (ed.). *Organización del conocimiento en sistemas de información y documentación*. Zaragoza: Universidad de Zaragoza, 1995. 7-25.
- Dahlberg, Ingetraut (2006). Knowledge organization: a new science? // *Knowledge Organization*. 33:1 (2006) 11-19.
- Frohmann, Bernd (2004). *Deflating information: from science to documentation*. Toronto: University Press.
- García Marco, Francisco Javier (1996). Contexto y determinantes funcionales de la clasificación documental. // *Scire* 2:1 (1996) 109-145.
- Hjørland, Birger; Albrechtsen, Hanne (1995). Toward a new horizon in information science: Domain-analysis. // *Journal of the American Society for Information Science*, 46:6 (1995) 400-425.
- Hjørland, Birger (2002). Domain analysis in information science: Eleven approaches - traditional as well as innovative. // *Journal of Documentation* 58:4 (2002) 422-462.
- Hjørland, Birger (2008). What is knowledge organization (KO)? // *Knowledge Organization* 35:2-3 (2008) 86-101.
- Hjørland, Birger (2012). Knowledge Organization = Information Organization? // Neelameghan, A.; Raghavan, K.S. (eds.). *Categories, Contexts and Relations in Knowledge Organization: Proceedings of the Twelfth International ISKO Conference 6-9 August 2012 Mysore, India*. *Advances in knowledge organization* 13. Würzburg: Ergon Verlag, 2012 206-11.
- Hjørland, Birger (2017a). Domain Analysis // *Knowledge Organization* 44:6 (2017) 436-464.
- Hjørland, Birger (2017b). Library and information science (LIS). // *ISKO Encyclopedia of knowledge organization*. <http://www.isko.org/cyclo/lis> (2018-05-16).
- Horn, Robert E. (2000). Information design: Emergence of a new profession. // Jacobson, Robert (ed.). *Information design*. Cambridge, Mass: MIT press, 2000. 15-33.
- Jacobson, Robert (2000). *Information design*. Cambridge, Mass: MIT press, 2000.

- Jorente, Maria Jose Vicentini (2014). Design da Informação, linguagens convergentes e complexidade na rede social e ambiente digital do Facebook. // *Informação & Tecnologia (ITEC)* 1:1 (jan./jun. 2014) 116-129
- Knemeyer, Dirk. (2003). Information Design: The Understanding Discipline. <http://boxesandarrows.com/information-design-the-understanding-discipline/> (2016-07-25)
- Martínez-Ávila, Daniel (2015). Knowledge Organization in the Intersection with Information Technologies. // *Knowledge Organization* 42:7 (2015) 486-98.
- Martínez-Ávila, Daniel (2018). Hacia una base teórica social de la ciencia de la información. // *Anuario ThinkEPI* 12 (2018) 83-89.
- Martínez-Ávila, Daniel; Guimarães, José Augusto (2015). La construcción de la Biblioteconomía como ciencia y su relación con la clasificación. // Rodríguez Muñoz, José Vicente; Gil Leiva, Isidoro; Díaz Ortuño, Pedro M.; Martínez Méndez, Francisco Javier (eds.). XII Congreso ISKO España y II Congreso ISKO España-Portugal, 19-20 de noviembre, 2015, Organización del conocimiento para sistemas de información abiertos. Murcia: Universidad de Murcia, 2015. 533-43.
- Martínez-Ávila, Daniel; San Segundo, Rosa; Zurian, Francisco A. (2014). Retos y oportunidades en organización del conocimiento en la intersección con las tecnologías de la información. // *Revista Española de Documentación Científica* 37:3 (julio-septiembre 2014) e053.
- Mikhailov, A. I. (1967). Informatics - A Scientific Discipline. // *Documentação e Informação Científica* 10:53 (1967) 239-242.
- Morin, E. (1977). Le système, paradigme ou théorie? // *Congres de l'A.F.C.E.T, Versailles, November 21, 1977. Paris: Fayard, 1982, pp. 172-189; and reprinted in Science avec conscience, new edition, Paris: Points/Seuil, 1990, pp. 238-255.*
- Niemeyer, L. (1998). Design no Brasil: origens e instalação. Rio de Janeiro: 2AB Editora, 1998.
- Orna, E.; Stevens, G. (1991). Information Design and Information Science: a new alliance? // *Journal of Information Science* 17 (1991) 197.
- Pinheiro, Lena Vania Ribeiro; Loureiro, José Mauro Matheus (1995). Traçado e limites da ciência da informação. // *Ciência da Informação* 24:1 (jan./abr. 1995) 42-53.
- Popper, K.R. (1973). *Objective Knowledge. An Evolutionary Approach.* Oxford: Clarendon Press. 1973.
- Rowley, Jennifer and Richard Hartley (2008). *Organizing Knowledge: An Introduction to Managing Access to Information.* 4th edition. Hampshire: Ashgate, 2008..
- Shannon, Claude E. (1948). A Mathematical Theory of Communication. // *The Bell System Technical Journal* 27:3 (1948) 379-423, 623-656.
- Shannon, Claude E.; Weaver, Warren. (1949). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press, 1949.
- Smiraglia, Richard P. (2012). Knowledge Organization: Some Trends in an Emergent Domain. // *El Profesional de la Información* 21:3 (mayo-junio 2012) 225-7.
- Spang-Hanssen, Henning (2001). How to teach about information as related to documentation. // *Human IT* 5:1 (2001) 125-143.
- Taylor, Arlene G.; Joudrey, Daniel N. (2009). *The Organization of Information.* 3rd ed. Westport, Conn.: Libraries Unlimited, 2009.

Enviado: 2017-05-23. Segunda versión: 2018-05-17.
Aceptado: 2018-06-19.

Del CMBD al Big Data en salud: un sistema de información hospitalaria para el siglo XXI

From MBDS to big data: a hospital information system for the 21st century

Gonzalo MARCO CUENCA, José Antonio SALVADOR OLIVÁN

Departamento de Ciencias de la Documentación e Historia de la Ciencia,
Universidad de Zaragoza. gmarco@unizar.es, jaso@unizar.es

Resumen

Se analiza la evolución del Conjunto Mínimo Básico de Datos (CMBD), un sistema de información hospitalario desde sus inicios hasta su última versión denominada Registro de Atención Sanitaria Especializada (RAE-CMBD). Su objetivo es normalizar toda la actividad a nivel especializado en hospitalización y atención ambulatoria del sector público y privado; y permite conocer el funcionamiento de los hospitales, su actividad asistencial, las patologías atendidas, su complejidad y abordaje clínico y los costes que suponen. Los registros clínico-administrativos, como es el caso del RAE-CMBD, se presentan como una gran oportunidad para la explotación de datos masivos sanitarios o *big data*, con el objetivo de poder conocer de forma anticipada la incidencia de enfermedades, el manejo de las mismas y los diferentes resultados asistenciales, tanto a nivel de conjunto del SNS, como de forma individual o comparada por hospitales o centros sanitarios. Si se cumplen las previsiones para los próximos años, el manejo de datos a gran escala podría mejorar la eficacia en la toma de decisiones para la planificación, evaluación y desarrollo de los sistemas sanitarios, convirtiéndose el RAE-CMBD en la fuente de datos del Sistema de Información Sanitaria del SNS (SISNS) de mayor relevancia e impacto para la gestión de los servicios, el conocimiento de la salud de la población y la investigación epidemiológica.

Palabras clave: Conjunto Mínimo Básico de Datos. CMBD. Big data. Sistemas de Información hospitalaria. Registro de Actividad de Atención Sanitaria Especializada. RAE-CMBD. Sistema de Información Sanitaria del SNS. SISNS. Gestión de la información sanitaria. Datos en salud.

1. Introducción

Dentro de los sistemas de información del Sistema Nacional de Salud (SNS) de España hay que destacar como una pieza clave y fundamental al *Sistema de Información Sanitaria del SNS (SISNS)*. Este sistema tiene como principal finalidad facilitar información objetiva, fiable y comparable sobre el funcionamiento del sistema sanitario, permitiendo valorar la consecución de sus objetivos, respondiendo a las necesidades de información de los agentes que lo conforman y sir-

Abstract

The Minimum Basic Data Set (MBDS) is a hospital information system of demographic type that has been developed in the National Health System (NHS) for over more than 25 years. The present study analyzes its evolution from its beginnings to latest version called the Register of Specialized Health Care (RSH-MBDS). This new model standardizing all activity at specialized level, both in hospitalization and ambulatory care in public and private sector, allowing to know about the operation of hospitals, care activity, pathologies treated, complexity and clinical approach and the costs involved. Clinical-administrative records, as is the case of RSH-MBDS, are a great opportunity for the exploitation of massive health data or big data, with the aim of being able to know in advance the incidence of diseases, their management and the different healthcare outcomes, in the entire NHS, as a separately or compared by hospitals or health centers. If the forecasts for the next years are fulfilled, large-scale data management could improve decision-making efficiency in the planning, evaluation and development of health systems, becoming the RSH-MBDS in the most relevant data source for the management of services, knowledge of the population health and epidemiological research.

Keywords: Minimum Basic Data Set. MBDS. Big Data. Hospital Information System. Specialized Health Care Activity Register. RAE-CMBD. Health Information System of the NHS. SISNS. Health Information management. Health Data.

viendo de soporte para la generación de conocimiento (Alfaro Latorre, 2012). El SISNS se caracteriza por ser un sistema que reúne información, principalmente estadística, procedente de otros sistemas sanitarios dentro del contexto del SNS (1). Estos datos que proceden de la gestión clínica y administrativa diaria que se produce en los diferentes centros y niveles asistenciales en España, tanto a nivel público, como privado, constituyen una fuente trascendental para comprender la situación actual del SNS, su contexto y su administración. Las Comunidades Autónomas, la

Administración General del Estado y las Entidades Gestoras de la Seguridad Social tienen la obligación de aportar datos al SISNS para favorecer su desarrollo y su mantenimiento. De forma paralela, las administraciones autonómicas y estatales tienen derecho a acceder y disponer de los datos que forman parte del SISNS. Es, por tanto, un sistema basado en la cooperación de las diferentes administraciones sanitarias, que busca rentabilizar el conocimiento que se genera y los recursos de información existentes en beneficio del conjunto.

El SISNS fue definido como un elemento instrumental capaz de medir la efectividad (aquello que funciona y lo que no), la eficiencia (reconocer si los recursos se están utilizando de manera óptima para obtener los resultados deseados) y dar una respuesta sobre el sistema de protección de la salud, en relación con: el nivel de acceso a las prestaciones, la medida en que el sistema sanitario responde a las necesidades de salud y el grado de satisfacción de la población. El SISNS, por tanto, se ha convertido en un recurso estratégico que busca hacer frente a los retos presentes y futuros de la realidad organizativa y de las demandas de información que plantean los tres grandes actores que integran el SNS: ciudadanos, profesionales y administraciones sanitarias (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2014).

Una de las principales fuentes de datos que integran el SISNS es el *Conjunto Mínimo Básico de Datos (CMBD)*. El CMBD es una fuente tradicional y muy arraigada en el SNS, que cuenta con más de 25 años de vida (la cumplimentación del CMBD es obligatoria desde el año 1992). El CMBD ofrece información estadística a nivel estatal relativa a los casos atendidos en los centros sanitarios de Atención Especializada y, en concreto, a los centros hospitalarios, ambulatorios y centros de especialidades dependientes de los mismos. Se trata de un registro de la actividad de la atención sanitaria especializada que reúne información poblacional y que utiliza como unidad de registro cada contacto asistencial con un paciente (2). Los registros se basan en un núcleo común de datos básicos normalizados y anonimizados que integran variables demográficas y clínico-asistenciales, permitiendo conocer la morbilidad atendida en cada hospital, ya sea público o privado. El CMBD es una fuente esencial para la investigación epidemiológica dentro del SNS y un recurso de alto valor organizacional para la planificación y evaluación de los servicios sanitarios (3).

La información como recurso es cuantiosa en un hospital y, tal y como indica Salvador Oliván (1997) “no existe un campo con una necesidad tan grande de proceso automático de datos como

el de la asistencia sanitaria”. En este sentido, el CMBD se constituye como una herramienta imprescindible para el procesamiento de datos de una forma estandarizada. Según Zapatero Gaviña (2010), el CMBD no debe interpretarse solamente como una base de datos administrativa que es utilizada para la gestión sanitaria; es, además, una herramienta muy útil para que los clínicos diseñen proyectos de investigación y, al ser una fuente de obligado cumplimiento (4), aporta abundante información que incluye registros con millones de entradas cada año (casi cuatro millones de altas anuales). El CMBD puede también considerarse como un sistema de información propiamente dicho, entendiendo el mismo como el proceso estadístico de entrada de datos y de salida de resultados en forma de indicadores que permiten medir la actividad realizada por los servicios y la morbilidad atendida en cada uno de ellos, pudiendo incluso detectar problemas potenciales de calidad e ineficiencia (Salvador Oliván, 1997).

Asimismo, es un sistema especializado que facilita el *benchmarking*, pues incluye diversos parámetros asistenciales como, por ejemplo, la duración de la estancia, los diagnósticos principales y secundarios, la continuidad asistencial, etc., que ofrecen la posibilidad de realizar comparaciones entre los distintos hospitales o entre los diversos servicios hospitalarios. Para poder analizar la información que integra el CMBD y poder ser objeto de comparación, ha sido necesaria la normalización de su estructura y la uniformidad de su contenido Carrillo y otros (1991) destacan “el tema de la uniformidad es un punto clave a la hora de tratar la información sanitaria. Además, la eficiencia, eficacia y equidad en la oferta de asistencia sanitaria no podría ser comparada entre distintos sistemas sanitarios en ausencia de una información uniforme de cada uno de ellos”.

El presente artículo mantiene como objetivo principal analizar la consolidación del CMBD como un sistema de información y un recurso estratégico para el conocimiento de la actividad asistencial especializada, a presente y futuro, en el conjunto del SNS. De igual modo, se busca exponer como su crecimiento, dada la magnitud de datos que anualmente se generan, obligará a un tratamiento basado en técnicas para el análisis de datos masivos o *big data*.

2. Origen, antecedentes y evolución del CMBD

El concepto de conjunto mínimo de datos básicos fue formulado por primera vez en 1969 en el contexto de una conferencia sobre sistemas de infor-

mación sanitarios en los Estados Unidos (Treviño, 1988; Foster y Conrick, 1998). Desde entonces hasta la actualidad el CMBD ha ido evolucionando a nivel internacional (ver tabla I), mejorando sus prestaciones, armonizando su información y normalizando sus datos y variables, tanto a nivel de datos administrativos, como clínicos.

Año	Hito cronológico
1969	Primera formulación del concepto CMBD.
1969	La OMS, a través de un Grupo de Estudio, recomienda un conjunto básico de datos con una finalidad estadística.
1972	El National Committee on Vital and Health Statistics de los Estados Unidos pone en marcha el Uniform Hospital Discharge Data Set (UHDDS), primer CMBD al Alta Hospitalaria en ese país, formado por 14 variables.
1974	El UHDDS es adoptado para los programas de salud estadounidenses: Medicare y Medicaid (Treviño, 1988).
1981	Bajo el patrocinio de la Comisión Europea, se define el European Minimum Basic Data Set (EMBDS), tomando como referencia el UHDDS. El EMBDS contenía un total de 13 variables y su cobertura debía contemplar tanto la atención hospitalaria como la ambulatoria.
2002	Primer Programa europeo de acción comunitaria en el ámbito de la salud (2001-2006). Se establece como área prioritaria el desarrollo y coordinación de un sistema de información sanitaria (Calvete, 2003).
2008	Segundo Programa europeo de acción comunitaria en el ámbito de la salud (2008-2013). Se concreta entre sus objetivos la generación y difusión de información, impulsando un sistema de información con mecanismos para recoger datos e información comparables y con indicadores estadísticos.
2008	Se publica el Reglamento (CE) 1338/2008 del Parlamento Europeo y del Consejo sobre estadísticas comunitarias de salud pública y de salud y seguridad del trabajo (5).
2014	Reglamento 282/2014 que define el Tercer Programa europeo de acción comunitaria en el ámbito de la salud (2014-2020). Como prioridad temática se encuentra la promoción de un sistema de información sobre salud que contribuya al proceso de toma de decisiones basado en datos.

Tabla I. Evolución internacional del CMBD (hitos más destacados)

2.1. Antecedentes y evolución del CMBD en España

En España, a mediados de los años 70, por medio del Real Decreto 1360/1976, se hace obligatorio el uso, por parte de los establecimientos sanitarios con régimen de internado de un libro

único de registro de ingresos y altas. Esta normativa buscaba ampliar el conocimiento, existente en esos momentos, sobre el sector hospitalario, planteando la recogida de aquellos datos estimados como fundamentales para comprender las circunstancias de las personas ingresadas. Este libro de registro se creaba como un instrumento de mejora y de apoyo al conocimiento sobre la morbilidad hospitalaria que ya se venía desarrollando, desde los años 50, por parte del Instituto Nacional de Estadística (INE) (6). Esta norma, actualmente derogada (7), se constituye como el primer antecedente legislativo de registro hospitalario en el contexto del SNS.

En 1986, la *Ley 14/1986, General de Sanidad*, en su artículo 23, dispone que las Administraciones Sanitarias, de acuerdo con sus competencias, deben crear registros y elaborar análisis de información para el conocimiento de las distintas situaciones que puedan derivar acciones de intervención de la autoridad sanitaria (Ministerio de Sanidad y Consumo, 2004). En 1987, vista la necesidad de crear una fuente de datos uniforme para cumplir con la *Ley General de Sanidad* y facilitar los procesos de gestión sanitaria, implantación de nuevos sistemas de financiación, elaboración de indicadores de rendimiento, control de la calidad asistencial y elaboración de una base de datos para investigación clínica y epidemiológica se propone, a través del *Grupo de Trabajo de la Ponencia sobre Planificación Sanitaria* del Ministerio de Sanidad, elevar una propuesta para establecer un CMBD al Alta Hospitalaria o CMBD de Hospitalización (CMBD-H) al Consejo Interterritorial del SNS (CISNS) (8). El 14 de diciembre de 1987, el Pleno del CISNS (1987), aprueba un CMBD nacional compuesto por 14 variables que toman como referencia las ya definidas en el UHDDS de los Estados Unidos y en el EMBDS europeo.

En 1988, por medio de otro Pleno del CISNS (1988) se desarrolla la estrategia de implantación del CMBD y se acuerda la creación del *Comité Técnico del Consejo Interterritorial para la Implantación y Evaluación del CMBD (Comité Técnico del CMBD)* (9) que tendrá como principales funciones: la homologación de criterios y definiciones en los datos del CMBD, la homologación de los criterios de codificación (10), la evaluación de la factibilidad y limitaciones del proceso de implantación del CMBD, el análisis de la información suministrada por el CMBD y la ampliación y/o revisión del CMBD.

A partir de esta decisión del CISNS, las Comunidades Autónomas se encargarán de ir implementando a lo largo del tiempo el CMBD, siendo Cataluña la primera Comunidad Autónoma en po-

nerlo en marcha en 1990. En el año 1992, la Secretaría General del SNS del Ministerio de Sanidad y Consumo determina que todos los centros hospitalarios del Instituto Nacional de la Salud (INSALUD) se encuentran obligados al registro y codificación del CMBD-H y su implantación deberá producirse antes del 31 de diciembre de ese mismo año. Los servicios sanitarios de las Comunidades Autónomas, en virtud de sus competencias en materia de gestión sanitaria, irán publicando normas locales para regular la implantación y recogida de datos para el CMBD-H a diferentes ritmos desde 1992 hasta el año 2011. El CMBD-H hay que considerarlo, consecuentemente, como un repositorio nacional que se abastece de una red de sistemas de información autonómicos. Estos últimos se van a encargar de suministrar la información que se registra en cada hospital de forma anual para conformar un producto único para todo el SNS.

Desde el año 1993, se viene regulando el denominado *Plan Estadístico Nacional (PEN)* (Real Decreto 136/1999) como el principal instrumento ordenador de la actividad estadística de la Administración General del Estado. El Real Decreto 1658/2012, por el que se aprueba el PEN 2013-2016, así como el Real Decreto 410/2016, por el que se aprueba el PEN 2017-2020, así como sus planes de ejecución, servirán de marco regulador de todas las operaciones estadísticas producidas con fines estatales, incluido el CMBD (11), garantizando el suministro de una información integral para el conjunto del Estado, la Unión Europea, las instituciones y los usuarios que así la requieran.

En el año 1998, se incluye en el modelo del CMBD-H una nueva variable, el tipo de financiación (12), creciendo su estructura de 14 a 15 variables.

En el año 2003, la *Ley 16/2003 de Cohesión y Calidad del SNS* dispone, a través de los artículos 53 y 55, el establecimiento de un sistema de información sanitaria para el SNS y la realización de estadísticas en materia sanitaria, reforzando aún más la necesidad de producción de un recurso como el CMBD.

Un año después, en el año 2004, se inicia la recogida de datos para el Conjunto Mínimo Básico de Datos de Atención Ambulatoria Especializada (CMBD-AAE). Desplegándose, a partir de este momento, dos productos a nivel nacional relacionados con el CMBD, el CMBD-H, definido en 1987 y obligatorio desde el año 1992, y el nuevo CMBD-AAE que nace para recoger la información clínico asistencial de los pacientes atendidos en área de hospitalización sin ingreso y hospitalización a domicilio. El modelo de datos del

CMBD-AAE será común al del CMBD-H e integrará información del paciente, del proveedor e información clínica. Esta información se integrará en el sistema único de información del CMBD pudiéndose explotar la información de forma desagregada. La actividad ambulatoria se irá integrando poco a poco, de inicio se incluye la actividad relativa a los procedimientos quirúrgicos realizados de forma ambulatoria y, posteriormente, el resto de modalidades asistenciales ambulatorias (procedimientos diagnóstico-quirúrgicos de alta complejidad, hospital de día, urgencias y consultas externas) (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2017a).

En el año 2005, de nuevo se produce un cambio en la estructura del CMBD, con la incorporación de dos nuevas variables, el código de identificación personal del paciente (CIP) y la identificación del servicio clínico responsable del alta del paciente. En estos momentos, el CMBD cuenta con 17 variables (ver tabla II).

Variable	Nombre de la variable	Año
1	Identificación del hospital	1987
2	Identificación del paciente	1987
3	Sexo	1987
4	Fecha de nacimiento	1987
5	Lugar de residencia	1987
6	CIP	2005
7	Financiación	1987
8	Tipo de financiación	1998
9	Fecha de ingreso	1987
10	Circunstancias de ingreso	1987
11	Diagnóstico principal y otros	1987
12	Procedimientos quirúrgicos y otros	1987
13	Procedimientos diagnósticos	1987
14	Fecha de alta	1987
15	Circunstancias al alta	1987
16	Identificación del médico responsables del alta	1987
17	Identificación del servicio clínico responsable del alta	2005

Tabla II. Relación de variables del CMBD por año de incorporación

2.2. El nuevo RAE-CMBD

El 10 de febrero de 2015 se publica el *Real Decreto 69/2015 por el que se regula el Registro de Actividad de Atención Sanitaria Especializada (RAE)*. La publicación de este registro fue la cul-

minación de una larga trayectoria de trabajo llevada a cabo por: *el Comité del CMBD, la Unidad Técnica de la Clasificación Internacional de Enfermedades (CIE) y la Subcomisión de Sistemas de Información del CISNS*. Esta nueva norma reguladora, que toma como base el CMBD existente y que, por este motivo, se va a reconocer por las siglas RAE-CMBD, amplía el alcance del CMBD a nivel de contenido, cobertura y finalidad, abarcando diferentes productos estadísticos y de información en un mismo instrumento, convirtiendo este nuevo registro como el principal referente y fuente de datos para la planificación, gestión sanitaria, evaluación e investigación clínica, epidemiológica y de servicios de salud en la actualidad y a futuro. El registro se implantará en soporte digital y su diseño y estructura deben facilitar el intercambio de datos, así como una explotación de los mismos por medios electrónicos (13).

El RAE-CMBD mantiene el objetivo de regular la actividad de la atención sanitaria especializada, sirviéndose del CMBD y extendiendo su estructura y contenido a un modelo homogéneo y normalizado integrado en el SISNS. Tal y como indica el Real Decreto 69/2015, en su artículo 2, el registro tiene una naturaleza administrativa y, como parte del SISNS, busca garantizar la disponibilidad de la información de la actividad sanitaria especializada, respondiendo a las necesidades de información de los diferentes agentes implicados en el SNS, así como: conocer la demanda asistencial y la morbilidad atendida, proporcionar a los registros autonómicos la información necesaria para la evaluación y control de la atención prestada y facilitar la realización de estadísticas de salud a nivel estatal, así como las que deriven de compromisos con organismos oficiales internacionales.

El RAE-CMBD amplía su modelo de datos de 17 a 31 variables, llevándose a cabo una revisión completa de su estructura y de las características de cada variable, tales como: su descripción, formato, longitud, listas de valores admitidos o de clasificaciones aceptadas, etc. Todo ello, buscando establecer los mecanismos necesarios para obtener un registro estandarizado mucho más eficiente que pueda garantizar una explotación óptima de la información. Para Sanz Sebastián (2017) *“la implantación del RAE-CMBD ha supuesto actuaciones sobre dos pilares fundamentales de la información sanitaria: por una parte, la adopción de la CIE10-ES como clasificación de referencia de codificación de la información clínica y por otra la adopción de un nuevo modelo de datos y ampliación del ámbito de aplicación de su antecesor el CMBD”*.

Hay que tener en cuenta que el 77% de las variables son de tipo administrativo e incluyen datos que hacen referencia a la identificación del paciente, proveedor de servicios y episodio asistencial, frente al 23% de las variables que incluyen información clínica. En la tabla III podemos observar el total de variables según su tipología, grupo (en función del tipo de información o datos que proporcionan) y número de variables relacionadas con cada grupo.

Tipo de variable	Grupo	Nº
Administrativa	Datos del paciente	8
	Datos Proveedor	2
	Datos del episodio asistencial	14
Clínicas	Información clínica	7
	Total	31

Tabla III. Variables del RAE-CMBD según tipología, grupos y número

En el RAE-CMBD se mantiene como unidad de registro el contacto asistencial, pudiendo ser clasificados estos contactos en función de la modalidad asistencial: hospitalización, hospitalización a domicilio, hospital de día médico, cirugía ambulatoria, procedimiento ambulatorio de especial complejidad y urgencia. En cuanto al origen de los datos, la fuente de información puede ser el informe clínico de alta o la historia clínica de cada paciente.

El proceso de implementación de dicho registro comenzó el 1 de enero de 2016 y tiene previsto extenderse hasta el 1 de enero de 2020. En los plazos en la disposición adicional cuarta del Real Decreto 69/2015, se indica que se iniciará la recogida de datos atendiendo a los siguientes criterios por tipo de contacto: hospitalización y cirugía ambulatoria en hospitales de agudos públicos y privados (1 de enero de 2016), hospitalización en otros hospitales públicos y privados (1 de enero de 2017), procedimientos ambulatorios de alta complejidad, hospital de día médico, hospitalización a domicilio y urgencias en hospitales públicos y privados (1 de enero de 2018), cirugía ambulatoria, procedimientos ambulatorios de alta complejidad y hospital de día médico en centros ambulatorios públicos especializados (1 de enero de 2019), procedimientos ambulatorios de alta complejidad, hospital de día médico, hospitalización a domicilio, urgencias y cirugía ambulatoria en hospitales de agudos y centros ambulatorios especializados privados (1 de enero de 2020). Estos plazos, según Sanz Sebastián (2017), del Ministerio de Sanidad, se están desarrollando con cierto retraso.

En julio de 2017 se encontraba a disposición la herramienta del RAE-CMBD para carga y validación por las Comunidades Autónomas. En octubre de 2017, 13 Comunidades Autónomas habían accedido a la aplicación y solamente 9 de ellas habían cargado datos. En esa misma fecha, de octubre de 2017, el nuevo sistema de información contaba ya con aproximadamente 2 millones de registros.

Con relación al sistema de clasificación de los diagnósticos clínicos para el RAE-CMBD, el CISNS fijó el 1 de enero de 2016 como fecha de entrada en vigor de la nueva CIE10-ES. La CIE10 es la décima revisión de la CIE que desde el año 1948 está a cargo de la OMS. La CIE ha sido durante muchas décadas el instrumento esencial que permitía la comparación nacional e internacional en el campo de la salud pública (Ramos Martín-Vegue, Vázquez-Barquero y Herrera Castanedo, 2002).

La transición a la nueva CIE10-ES tiene una gran complejidad, ya que resulta ser un sistema que presenta considerables cambios con respecto a la edición anterior, además de integrar diagnósticos y procedimientos (14). Uno de los principales cambios, y a su vez principal ventaja, afecta al volumen del número de entradas; se ha pasado de unos 14.000 diagnósticos aproximadamente con la CIE9-MC a más de 68.000 con el actual sistema de CIE10-ES.

Con respecto a los procedimientos el cambio ha sido aún mayor, de casi 4.000 a 72.000. También su estructura y sistema de codificación ha cambiado. La CIE10-ES incorpora nuevos capítulos, características y categorías. En cuanto al sistema de codificación pasa de ser un sistema numérico a alfanumérico (15). Todo ello completa un panorama de notables dificultades para su adaptación por los propios sistemas de información, así como para los codificadores, documentalistas, facultativos y gestores. La codificación de los diagnósticos y procedimientos es clave para la explotación de la información del CMBD y, tal y como señala Sierra (2016), *“el paso de CIE-9 a CIE-10 está lastrando el registro de información, pero la exhaustividad del nuevo código permitirá afinar mucho la gestión”*.

3. El ICMBD y el Portal Estadístico del SNS

El Ministerio de Sanidad, Servicios Sociales e Igualdad lanzó en el año 2016 la versión en línea del CMBD, denominada ICMBD (Indicadores y ejes de análisis del CMBD). Esta aplicación se encuentra disponible en la dirección de Internet: <http://icmbd.es>. La versión en línea se define como un modelo para el análisis y explotación de

los datos del registro del CMBD de una forma estructurada, fiable y sencilla (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2017b).

La aplicación ha sido desarrollada, a través de un convenio de colaboración, por el Grupo de Minería de datos de la Universidad de Cantabria (16) para el Ministerio de Sanidad. Este proyecto consistía en la implantación y soporte de un modelo de indicadores y ejes de análisis de los datos del CMBD-H desde el año 2001 hasta el año 2015.

El modelo integra una serie de cinco componentes básicos que pueden ser utilizados para la consulta, agrupación y explotación de los datos. Estos componentes son: 1) Dimensiones: constituyen los objetivos generales de conocimiento (por ejemplo, estilos de práctica clínica, eficiencia, efectividad clínica, frecuentación, resolutivez, etc.); 2) Ejes de análisis: son las líneas para el análisis de las dimensiones (por ejemplo: punto de corte, serie temporal, valor base, mejores prácticas, etc.); 3) Indicadores: son los datos o conjuntos de datos cuyo resultado informa sobre facetas concretas de una o varias dimensiones; 4) Niveles de clasificación de los indicadores y 5) Filtros.

Indicador	Descripción
1	Estancia media
2	Estancia media preoperatoria
3	Frecuentación en hospitalización
4	Hospitalizaciones potencialmente evitables
5	Procedimientos realizados en hospitalización
6	Tasa de ambulatorización quirúrgica
7	Tasa de cesáreas
8	Tasa de complicaciones
9	Tasa de infección nosocomial
10	Tasa de mortalidad
11	Tasa de reingresos
12	Informe resumen de todos los indicadores

Tabla IV. Relación de indicadores ICMBD

El modelo cuenta con un total de siete dimensiones y seis ejes de análisis diferenciados. Del mismo modo, cuenta con la posibilidad de selección de 12 indicadores para consulta y explotación de datos.

Por otra parte, el Portal Estadístico del SNS, una de las principales herramientas en línea del SISNS, incluye dentro de su consulta interactiva, desde el año 2017, la posibilidad de consulta del CMBD-H y del CMBD-AAE. Este recurso del SNS

incluye, además del CMBD, la posibilidad de consulta de una lista de los principales sistemas de información estadística del SNS (Encuesta Nacional de Salud (ENS), Barómetro Sanitario, Sistema de Información de Atención Primaria (SIAP), Sistema de Información de Atención Especializada (SIAE), Base de datos Clínicos de Atención Primaria (BDCAP), etc. (figura 1).



Figura 1. Captura de pantalla de la Consulta Interactiva del Portal Estadístico del SNS (<http://peestadistico.inteligenciadegestion.msssi.es/>)

En el caso del CMBD-H integra la información procedente de las Comunidades Autónomas en su repositorio de datos con una cobertura que va desde el año 1997 hasta el año 2015. La consulta permite cruzar datos por las variables básicas poblacionales (edad y sexo del paciente); las del episodio asistencial (tipo de ingreso y alta) y las variables clínicas (diagnósticos y procedimientos), a las que se añaden otras derivadas del uso de sistemas de clasificación de pacientes y estimaciones de costes. La información puede también obtenerse de una forma desagregada llegando al nivel de Grupo de Hospital o Comunidad Autónoma, para la consulta sobre hospitales, y hasta el nivel de Comunidad Autónoma y provincia, para la consulta sobre pacientes (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2017c).

En cuanto al CMBD-AAE se incluyen datos desde el año 2004 hasta el año 2015, integrando la práctica totalidad de la Cirugía Ambulatoria del SNS y parcialmente hospital de día médico y los hospitales privados, siendo aún muy escasa la información del resto de modalidades ambulatorias. La única variable de análisis es el número de contactos o casos y al igual que el CMBD-H

puede obtenerse información de forma desagregada hasta el nivel de Hospital o Comunidad Autónoma, para hospitales, y hasta el nivel de Comunidad Autónoma y Provincia, para pacientes (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2017d). La aplicación también facilita la reutilización de los datos proporcionando herramientas para la exportación a ficheros estructurados en diferentes formatos (xls, csv, etc.).

4. Más allá de la estadística: conocimiento útil y accesible

Los registros clínicos son la base principal para conocer la realidad asistencial. En ellos se recogen los hechos, incluyendo los datos que se van agregando y contextualizando en los sistemas para que la información se transforme en conocimiento y pueda ser aprovechado por las organizaciones, la sociedad en general y el SNS en su conjunto. Solamente con los registros del CMBD-H, desde el año 1997 hasta el año 2015 ha habido un total de 66.425.949 contactos o casos, con una media de 3.496.103 contactos anuales (ver figura 2).

Un sistema de información como el nuevo RAE-CMBD, doblando en número de variables al histórico CMBD, se transformará a muy corto plazo en un repositorio de datos masivos. Si planteamos una operación simple y multiplicamos las 31 variables del RAE-CMBD por la media de contactos anuales para el CMBD-H, considerando que estas variables sean completadas en su totalidad y que solamente se incluye un valor por variable, la cantidad de datos a registrar es de más de 108 millones de datos al año.

Estas grandes dimensiones hacen que, inevitablemente, el SNS y el Ministerio de Sanidad se planteen una estrategia sanitaria orientada al “big data” y, mucho más, teniendo en cuenta que el RAE-CMBD solamente representa una fuente de datos para el SISNS. Tal es así, que el nuevo Consejo Asesor de Sanidad (17), designado en febrero de 2018, con el cometido de establecer las bases del SNS para las próximas décadas, tenga entre sus encargos prioritarios la realización de “un análisis de las utilidades que tiene que tener el big data como elemento de información y conocimiento” para saber cómo afrontar decisiones para agilizar el diagnóstico precoz y mejorar las decisiones en cuanto a tratamientos, así como para dar una respuesta adecuada y satisfactoria al desafío que representan los retos demográficos (Huertas, 2018; Ministerio de Sanidad, Servicios Sociales e Igualdad, 2018). En este sentido, el Ministerio de Sanidad en el año 2009, fecha en la que ya se planteaba la necesidad de crear un nuevo registro para el CMBD,

comunicaba la necesidad de darle “*utilidad de verdad al CMBD*” indicando que “*una gigantesca base de datos como el CMBD debe servir para algo más que para producir estadísticas*” (Sierra, 2009).

En esa misma fecha, el Ministerio había seleccionado, dentro de la estrategia de explotación del CMBD, diez indicadores básicos con los que se pretendía generar conocimiento útil y accesible

para todos los potenciales usuarios. De igual manera, se señalaba que iba a ser necesario para poder explotar esta información utilizar “*minería de datos, generación de informes y herramientas versátiles para análisis multidimensional*” con objeto de ser capaces de generar estadísticas e informes, y modelizar indicadores de *benchmarking* que generen conocimiento mediante tecnología de *business intelligence* (Sierra, 2009).

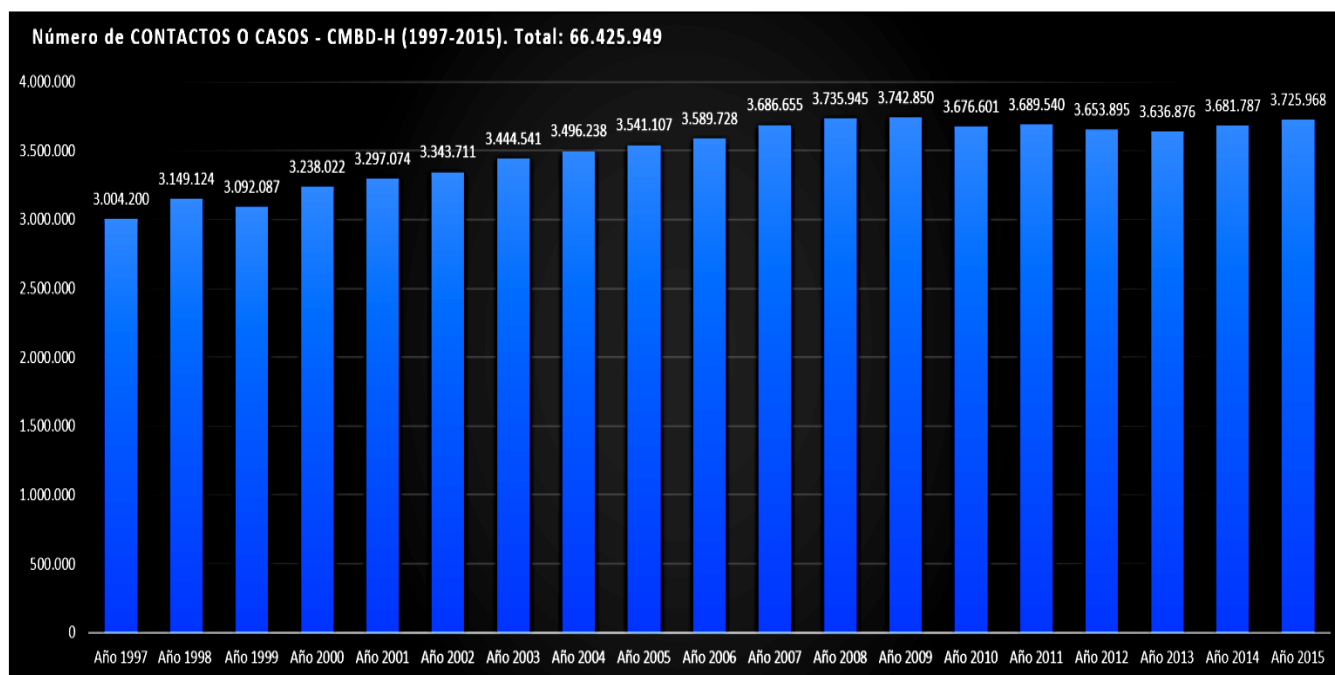


Figura 2. Registro de altas del CMBD-H. Número de contactos (1997-2015) (elaboración propia a partir de los datos del Ministerio de Sanidad, Servicios Sociales e Igualdad, <http://pestadistico.inteligenciadegestion.msssi.es>)

4.1. Una breve aproximación al Big data en salud

La Recomendación UIT-Y.3600, de la Unión Internacional de Telecomunicaciones (UIT), es la primera norma a nivel internacional donde se presentan los requisitos, las capacidades y los casos de uso de los volúmenes masivos de datos (big data). En este documento se define el concepto de *big data* como “*un paradigma para hacer posible la recopilación, el almacenamiento, la gestión, el análisis y la visualización, potencialmente en condiciones de tiempo real, de grandes conjuntos de datos con características heterogéneas*” (International Telecommunication Union, 2015). De una forma más sencilla, podríamos interpretar *big data* como el conjunto de datos e información de gran volumen y variedad, cuyo procesamiento resulta ser tan complejo que no puede realizarse por medio de bases de datos convencionales.

El resultado de toda explotación de datos masivos, más allá de los nuevos procesos, tecnologías o modelos de negocio que están surgiendo para dar soporte a su aprovechamiento, debe ser la obtención del valor o conocimiento que los propios datos encierran. En este sentido, Mayer-Schonberger y Cukier (2013) indican que “*los big data, los datos masivos, se refieren a cosas que se pueden hacer a gran escala, pero no a una escala inferior, para extraer nuevas percepciones o crear nuevas formas de valor*”. Según la Fundación COTEC para la Innovación, en su informe sobre «Generación de talento Big Data en España» (Fundación COTEC, 2017), “*la economía de los datos, implica que industrias, sectores y mercados, operarán completamente apoyados en el uso e intercambio de datos, así como en base a la “inteligencia” generada por su análisis. Como consecuencia, empresas de diversos y diferentes sectores de la industria, tendrán que organizar y orquestar múltiples aplicaciones y pro-*

cesos, para trabajar de forma integrada y eficiente, con el objetivo de optimizar entornos operativos enteros”.

Podríamos concluir, para definir este concepto, que los *big data* se caracterizan por las denominadas 6 V: *volumen, velocidad, variedad, veracidad, variabilidad y valor*. La primera característica, volumen, refleja las grandes cantidades de datos que se manejan; la segunda, velocidad, hace referencia a la frecuencia y rapidez de crecimiento de la información; la tercera, variedad, implica la diversidad de los tipos de información, estructuras y formatos que se pueden integrar; la cuarta, veracidad, expresa la autenticidad de los datos; la quinta, variabilidad, se refiere a los cambios de datos durante su procesamiento y ciclo de vida y, por último, la sexta, valor, hace mención a la calidad de los datos y el valor agregado que pueden aportar para la creación de conocimiento (Wang y Alexander, 2015).

El registro de la información y la normalización de los procesos ha provocado que organizaciones de todo tipo acumulen una ingente cantidad de datos. En el sector sanitario esta situación no ha pasado desapercibida, al contrario, monitorizar la situación de la salud de la población está aportando un volumen enorme de datos (volumen), datos que crecen a gran velocidad y de forma exponencial haciéndose necesario el uso de las nuevas tecnologías avanzadas para su captura y tratamiento con objeto de extraer el valor de los mismos de una forma eficiente (velocidad). Además, en el SNS existen una gran variedad de sistemas de información, que necesitan interrelacionarse e interoperar y que incluyen datos de salud de diferentes contextos de atención y especialización que, en muchas ocasiones, se encuentran sin normalizar, datos no estructurados, y sin codificar (variedad y variabilidad). Los datos de salud deben cumplir con la obligación de ser fidedignos y, para ello, hay que fomentar que la calidad de la información registrada en los sistemas sea, no solamente veraz, sino también útil y significativa para la seguridad y evolución de los pacientes y para la toma de decisiones de los profesionales (veracidad y valor).

Se espera, por tanto, que las aplicaciones de *big data* en el sector de la salud tengan un alto potencial para mejorar la eficiencia y la calidad en la prevención de enfermedades y la provisión de los cuidados. La medicina se está convirtiendo en una ciencia de la información y en pocos años se dispondrán de billones de datos de cada individuo, siendo un desafío el desarrollo de la tecnología de la información que pueda conducir dichos datos a hipótesis reales sobre los propios individuos (Kim, 2017).

En cuanto a la procedencia de los datos de salud, su origen puede encontrarse, entre otras fuentes, en: la historia clínica electrónica, los sistemas de prescripción, los sistemas de almacenamiento y comunicación de imágenes y pruebas de laboratorio, las bases de datos construidas con finalidades clínicas y estadísticas, los dispositivos médicos y *wereables*. En el informe “*Big data en salud digital*” (Fundación Vodafone España y Red.es, 2017) se indica que, de una forma operativa, los datos masivos deberán permitir a los centros sanitarios disponer de información práctica y precisa sobre la demanda de servicios y la disposición y calidad de los mismos, convirtiéndose en un recurso muy valioso para la gestión. Los datos masivos en salud, podrán estar disponibles en diferentes formatos (texto, audio, video u otros registros biométricos) siendo susceptibles de ser analizados para proporcionar información nueva y útil para los sistemas de salud. Del mismo modo, este documento expresa la trascendencia que el *big data* tendrá como elemento clave para el sostenimiento presente y futuro del SNS.

4.2. Analítica predictiva, descriptiva y modelos de decisión para generar conocimiento

La utilización de grandes registros de datos masivos y el empleo de herramientas analíticas, puede aportar información y evidencia para la práctica asistencial (Angus, 2017). Tal y como indica Martínez Sesmero (2015)

[...] la gran mayoría de los agentes que participan en las estructuras de los servicios de salud reconocen que el análisis del big data puede ofrecer nuevas posibilidades en la elaboración de modelos predictivos, patrones de comportamiento, el descubrimiento de nuevas necesidades, reducir riesgos, así como proveer servicios más personalizados, todo ello en tiempo real y teniendo en cuenta toda la información relevante.

Actualmente, se están utilizando tres tipos de modelos para el análisis de datos masivos: 1) los *modelos predictivos* que estiman las probabilidades, a partir de los datos, detectando tendencias y pronosticando posibles situaciones en el futuro; 2) los *modelos descriptivos* que se encargan de representar las relaciones entre individuos, grupos, poblaciones y otros factores; y 3) los *modelos de optimización o decisión*, que son un híbrido entre los dos anteriores y se centran en aportar el valor de los datos para dar soporte a la toma de decisiones. Estos modelos pretenden generar una nueva inteligencia clínica basada en el procesamiento de los datos de forma continuada y en la exploración e interpretación de los mismos, buscando extraer el conocimiento que de otro modo se nos podría haber pasado por alto o podría encontrarse oculto a nuestros ojos.

Para todos estos casos y métodos de análisis, una herramienta como el RAE-CMBD cumple con las características adecuadas para ser la fuente de información principal y mejor candidata para su tratamiento por herramientas y tecnologías *big data*. Hay que tener en cuenta que el RAE-CMBD es un producto consolidado y que presenta grandes ventajas para su análisis, ya que contiene datos normalizados, estructurados e interoperables. Además, los datos son depurados y revisados, mejorando enormemente su calidad y beneficios derivados.

Existe, asimismo, una amplia experiencia por parte de los centros sanitarios en la recolección de datos para el CMBD, en la aplicación de garantías de seguridad y privacidad, en la protección de datos personales y en la clasificación de la información clínica. De igual modo, hay una especial sensibilidad, reflejada en la legislación vigente y transmitida a lo largo del tiempo por medio del trabajo desarrollado por comités, grupos de trabajo en el SNS, así como por el propio Ministerio de Sanidad y las Comunidades Autónomas, de la utilidad e interés que esta fuente de datos tiene para la gestión y la investigación sanitaria a nivel nacional e internacional.

Todos estos motivos hacen que las posibilidades de análisis del RAE-CMBD, a todos los niveles, sean inmensas. El nuevo producto sigue integrando información demográfica del paciente, de su situación clínica, del proceso asistencial, de los procedimientos realizados, del servicio responsable y de los resultados clínicos (mortalidad, complicaciones o reingresos) y, además, mejora la información recogiendo datos de las diferentes modalidades asistencializadas, públicas y privadas. Es cierto que aún queda espacio para su evolución y desarrollo, pues acaba de comenzar una nueva etapa para el registro de datos relacionados con el CMBD y tendrá que observarse su progreso, aunque no cabe duda de que la abundancia de los datos que aportará serán una gran oportunidad para poner a prueba la tecnología y los modelos analíticos del *big data* en el SNS.

5. Conclusiones

El CMBD es una de las principales fuentes de información poblacional del SISNS. Como tal, se constituye como un recurso estratégico para el conocimiento de la actividad asistencial especializada, tanto de hospitalización como ambulatoria, pública y privada. Con la regulación en el año 2015 del RAE-CMBD se incluyen considerables ventajas a este recurso en relación a modelos anteriores, añadiendo una estructura de datos nor-

malizados y homogéneos que proporcionará información sobre la actividad asistencial, las patologías atendidas, su complejidad, su abordaje clínico y los costes que suponen para el sistema. El RAE-CMBD, además, mantiene una hoja de ruta para que su implantación obligatoria se realice de una forma progresiva atendiendo al tipo de hospital y ámbito asistencial; dando comienzo en enero del año 2016 y finalizando en enero de 2020. Se espera que el registro clínico del nuevo RAE-CMBD aporte tal magnitud de datos que generará un repositorio de grandes dimensiones a muy corto plazo.

Para que los datos sean de utilidad deben aportar conocimiento, para ello hay que fomentar el desarrollo de modelos de negocio que sean útiles en el aprovechamiento de los datos masivos. Estos modelos deberían basarse en el consenso, la armonización y normalización de las estructuras y registro de datos, buscando facilitar la interoperabilidad y la explotación de los datos de salud en beneficio del sistema, los ciudadanos, el soporte a la toma de decisiones y la investigación científica.

Potenciar el análisis de los datos masivos o *big data* será una apuesta a largo plazo para las instituciones y autoridades sanitarias, no sin riesgo o complejidad, pues aún se desconocen todas las posibilidades que pueden ofrecer las tecnologías y técnicas alrededor de la gestión y análisis de datos a gran escala. En cualquier caso, una estrategia nacional compartida en el SNS para el buen gobierno de datos y la aplicación de *big data* sobre los recursos de información nacionales, tales como el RAE-CMBD y otras fuentes del SISNS, podría aportar valor y evitar riesgos, detectar patrones de comportamiento sobre la población y las enfermedades, ofrecer una mayor inteligencia y control sobre los procesos sanitarios y ahorrar costes al sistema y, posiblemente, aumentar la eficacia y eficiencia sobre la asistencia, la calidad de la atención y un mayor conocimiento sobre las demandas o necesidades de los diferentes agentes.

Para ello, como primer paso imprescindible, habría que asegurar la calidad de los datos desde su registro, mejorar la coordinación e intercambio de datos e información en el conjunto del SNS, de modo que, posteriormente, se puedan reutilizar los datos y establecer relaciones entre los mismos, trabajar con nuevos indicadores y métodos de análisis más potentes que puedan ofrecer respuestas a las preguntas más complejas, aportando informaciones fiables y valiosas para la interpretación y comprensión del estado de la salud, así como para pronosticar tendencias de futuro de la asistencia sanitaria.

Tal y como indica Kumholz (2014) la incorporación de grandes datos y análisis de próxima generación en la investigación, la práctica clínica y la información sobre salud demográfica requerirá no solo nuevas fuentes de datos, sino también nuevas ideas, capacitación y herramientas. Ya existen herramientas o sistemas, como por ejemplo *Hadoop* (18), que se están utilizando para almacenar, procesar y analizar grandes volúmenes de datos con cientos de terabytes, petabytes e, incluso, exabytes. Se tratan de herramientas de nueva generación que consiguen escalar desde unos pocos servidores a miles de máquinas y proponen soluciones para secuenciación que puede basarse en la información de la historia de salud electrónica de los pacientes y, de este modo, detectar patrones de riesgo relacionados con grupos poblacionales, revelar perfiles (clusters), redes de comorbilidad, eventos adversos de los medicamentos, patrones terapéuticos, etc. También parecen prometedoras las herramientas que aportan soluciones de análisis visual, con posibles aplicaciones para el desarrollo y mejoras en el campo de la salud pública (19).

El RAE-CMBD, dada sus considerables mejoras para facilitar una información estructurada e intercambiable, así como las expectativas que existen sobre su crecimiento y posibilidades de explotación para generar conocimiento, hacen que sea la fuente de información idónea para ser sometida, de forma gradual, a las técnicas y modelos de análisis para datos masivos. Sin embargo, hay que tener en cuenta que aún existen grandes incógnitas por resolver sobre lo que la tecnología *big data* en el dominio de la salud podrá ofrecer, por lo que tendrá que ser asumida por el SNS con cierta prudencia y tiempo.

La ciencia de los datos es una realidad, una oportunidad y un desafío cuyo avance y mejora dependerá de cómo se vayan transformando los sistemas de información en el futuro, facilitando datos en tiempo real, organizados, precisos y procesables. Tampoco hay que olvidar que, muy posiblemente, la mayor parte del conocimiento clínico aún se encuentre oculto, sumergido en la profundidad de las historias clínicas de los pacientes y en la gran cantidad de información conservada, no normalizada y registrada por los profesionales por medio del lenguaje natural. En este sentido, el *big data* tendrá también mucho en lo que contribuir.

Notas

(1) Tal y como indica la Ley de Cohesión y Calidad del SNS (2003), en su artículo 55, "el sistema de información sanitaria contemplará específicamente la realización de estadísticas para fines estatales en materia sanitaria, así como las de interés general supracomunitarias y las que

se deriven de compromisos con organizaciones supranacionales e internacionales, que se llevarán a cabo con arreglo a las determinaciones metodológicas y técnicas que establezca el Ministerio de Sanidad, consultado el Consejo Interterritorial del SNS".

- (2) Según el Real Decreto 69/2015, en su artículo 6, se entiende por contacto la atención sanitaria prestada bajo la misma modalidad asistencial y de forma ininterrumpida por un proveedor sanitario a un paciente.
- (3) Este registro viene a responder a las necesidades de información de los diferentes agentes del SNS, tal y como se establece en los artículos 53 y 55 de la Ley 16/2003, de 28 de mayo, de cohesión y calidad del Sistema Nacional de Salud.
- (4) El artículo 23 de la Ley 14/1986, de 25 de abril, General de Sanidad dispone que las Administraciones Sanitarias, de acuerdo con sus competencias, crearán los registros y elaborarán los análisis de información necesarios para el conocimiento de las distintas situaciones de las que puedan derivarse acciones de intervención de la autoridad sanitaria (Ministerio de Sanidad y Consumo, 2004).
- (5) El Reglamento 1338/2008, en su artículo 1.2, indica que "las estadísticas incluirán, en forma de conjunto de datos armonizados y comunes, la información que requieran la acción comunitaria en el ámbito de la salud pública, el apoyo a estrategias nacionales para el desarrollo de una asistencia sanitaria de alta calidad". Los datos estadísticos serán la fuente principal para la lista de Indicadores de Salud de la Unión Europea (ECHI) que buscan ofrecer una visión general del estado de salud, los determinantes de la salud y los sistemas sanitarios de los estados miembros. Para más información sobre ECHI se puede consultar: http://ec.europa.eu/health/indicators_data/indicators_en
- (6) El INE viene recogiendo desde el año 1951, a través de la *Encuesta de Morbilidad Hospitalaria (EMH)*, la evolución de la morbilidad hospitalaria en España. Esta encuesta, de carácter anual, ofrece información sobre las altas hospitalarias con internamiento y la estancia media de las mismas en función del diagnóstico principal asociado al alta. Se considera alta hospitalaria el procedimiento por el que un paciente ingresado en un centro o establecimiento sanitario deja de ocupar una cama de hospitalización por curación, mejoría, fallecimiento, traslado o alta voluntaria (Instituto Nacional de Estadística, 2016).
- (7) Este Real Decreto 1360/1976 se encuentra derogado por el Real Decreto 69/2015, de 6 de febrero, por el que se regula el Registro de Actividad de Atención Sanitaria Especializada.
- (8) Hay que destacar que el CISNS, como órgano de coordinación del SNS, fue creado en abril de 1987, siendo la propuesta del CMBD-H uno de los temas aprobados durante el mismo año de su creación.
- (9) Este Comité Técnico del CMBD, tal y como se recoge en el acuerdo del CISNS (1988), se constituye como una Comisión delegada del CISNS.
- (10) Dentro de las funciones de este Comité Técnico del CMBD se incluye la de asegurar la traducción completa de la Clasificación Internacional de Enfermedades 9ª edición – Modificación Clínica (CIE-9-MC). La CIE-9-MC ha sido la principal fuente para la codificación de diagnósticos en el CMBD desde sus inicios hasta el año 2016, fecha en la que se decide sustituir por la nueva versión CIE-10-ES.
- (11) En el PEN 2017-2020 figura la operación anual de explotación estadística dentro del Inventario de Operaciones

Estadísticas (IOE) la ficha 54011 – 7734. Explotación estadística del CMBD (RAE-CMBD). Demanda Hospitalaria y Morbilidad Atendida.

- (12) Esta variable, tipo de financiación, hace referencia al responsable de la financiación económica de los costes del contacto: SNS, convenios bilaterales con otros países europeos, mutualidades de regímenes especiales de la Seguridad Social de funcionarios, mutuas colaboradoras de la Seguridad Social, financiación privada directa, etc. Se diferencia de la variable financiación que indica si es pública o privada.
- (13) El fichero del RAE-CMBD se crea por Orden SSI/1885/2015, de 8 de septiembre, por la que se modifica la Orden de 21 de julio de 1994, por la que se regulan los ficheros con datos de carácter personal gestionados por el Ministerio de Sanidad y Consumo.
- (14) En el año 1991, el Ministerio de Sanidad publicó la primera edición completa en castellano de la CIE9-MC y ha realizado desde entonces, con periodicidad aproximada bianual, las sucesivas actualizaciones siguiendo la estela de las publicadas por el Centro Nacional de Estadísticas de Salud (NHSC) del Centro para el Control de las Enfermedades de los Estados Unidos (CDC) (Ministerio de Sanidad, Política Social e Igualdad, 2010).
- (15) La CIE10-ES es una clasificación alfanumérica que contiene entre 3 y 7 dígitos, siendo el primer carácter siempre una letra.
- (16) Para más información sobre este grupo consultar: <http://www.meteo.unican.es/research/datamining>
- (17) El Consejo Asesor de Sanidad es el máximo órgano consultivo y de asistencia a la persona titular del Departamento en la formulación de la Política Sanitaria, en este caso al ministro/a titular del Ministerio de Sanidad.
- (18) Hadoop es un proyecto de código abierto que incluye diferentes soluciones para datos masivos administrado por la Fundación Apache Software. Para más información se puede consultar: <http://hadoop.apache.org/>
- (19) Para más información sobre aplicaciones visuales para big data se puede consultar la revista Forbes (Marr, 2017) donde se ha publicado un listado con las 7 mejores herramientas de visualización de datos.

Referencias

- Alfaro Latorre, M (2012). Utilizando el sistema de información sanitaria del SNS. Madrid: Escuela Nacional de Sanidad. http://e-spacio.uned.es/fez/eserv/bibliuned:500569/n2.5_Utilizando_Sistema_de_Informaci__n_Sanitaria_del_SNS.pdf
- Angus, DC. (2017). Wheter to intubate during cardiopulmonary resuscitation: Conventional wisdom vs Big Data. // JAMA. 317:5, 477-478.
- Calvete, A (2003). El programa de acción de la Unión Europea en el ámbito de la salud pública. // Revista Española de Salud Pública. 77:5, 517-521.
- Carrillo, E; Juncosa, S; Bólibar, B; Parkin, D; Hutchinson, A y Fisher, P (1991). Conjuntos Mínimos de Datos en la Atención Ambulatoria. //Gaceta Sanitaria, 26 (5): 225-230.
- Consejo Interterritorial del Sistema Nacional de Salud (1987). Acuerdo nº 30. Pleno del 14 de diciembre de 1987. Acta nº 5. <https://www.msssi.gob.es/organizacion/consejoInterterri/docs/30.pdf>
- Consejo Interterritorial del Sistema Nacional de Salud (1988). Acuerdo nº 52. Pleno del 13 de julio de 1988. Acta nº 8. <https://www.msssi.gob.es/organizacion/consejoInterterri/docs/52.pdf>
- Decisión Nº 1786/2002/CE del Parlamento Europeo y del Consejo de 23 de septiembre de 2002 relativa a la adopción de un programa de acción comunitario en el ámbito de la salud pública (2003-2008). Diario Oficial de la Unión Europea. 9 de octubre de 2002. L271: 1-11.
- Decisión Nº 1350/2007/CE del Parlamento Europeo y del Consejo de 23 de octubre de 2007 por el que se establece el segundo programa de acción comunitario en el ámbito de la salud pública (2008-2013). Diario Oficial de la Unión Europea. 20 de noviembre de 2007. L301: 3-13.
- Foster, J; Conrick M. (1998). Nursing Minimun Data Sets: Historical Perspective and Australian Development. En: Saba, V; Pocklington DB; Miller KP (ed.). Nursing and Computers. An anthology, 1987-1996. New York: Springer-Verlag.
- Fundación COTEC (2017). Generación de talento Big Data en España. Madrid: Fundación COTEC para la Innovación. <http://cotec.es/media/BIG-DATA-FINAL-web.pdf>
- Fundación Vodafone España y Red.es (2017). Big data en salud digital. Madrid: Fundación Vodafone España. http://www.fundacionvodafone.es/sites/default/files/informe_big_data_en_salud_digital.pdf
- Huertas, J. (2018). Sanidad quiere nuevos roles sanitarios, medicina personalizada y "big data". // Redacción Médica, 14 de febrero de 2018.
- Instituto Nacional de Estadística (2016). Encuesta de Morbilidad Hospitalaria. Año 2015. Nota de prensa de 29 de noviembre de 2016. <http://www.ine.es/prensa/np1005.pdf>
- International Telecommunication Union (2015). Recommendation ITU-T Y.3600. Big data. Cloud computing-based requirements and capabilities. Series Y: Global Information infrastructure, Internet Protocol aspects and next-generation networks. Cloud Computing. Ginebra: ITU. <https://www.itu.int/rec/T-REC-Y.3600-201511-1>
- Kim J. (2017). Big data, health informatics, and the future of cardiovascular medicine. // Journal of the American College of Cardiology. 69:7, 899-902.
- Kumholz, HM (2014). Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. // Health Aff (Millwood). 33:7, 1163-70.
- Ley 16/2003, de 28 de mayo, de cohesión y calidad del Sistema Nacional de Salud. // Boletín Oficial del Estado. 29 de Mayo de 2003. 128; 20567-20588.
- Marr, B. (2017). The 7 Best Data Visualization Tools in 2017. Forbes, 20 de julio. <https://www.forbes.com/sites/bernardmarr/2017/07/20/the-7-best-data-visualization-tools-in-2017/#4374e0f96c30>
- Martínez Sesmero, JM. (2015). "Big Data"; aplicación y utilidad para el sistema sanitario. // Farmacia Hospitalaria. 39:2, 69-70.
- Mayer-Schonberger, V. y Cukier, K. (2013). Big data. La revolución de los datos masivos. Madrid: Turner.
- Ministerio de Sanidad y Consumo (2004). Ley 14/1986, de 25 de abril, General de Sanidad. Colección Textos legales. Serie Sanidad; 1. Madrid : Ministerio de Sanidad y Consumo.
- Ministerio de Sanidad, Política Social e Igualdad (2010). Hacia la CIE-10. Documento Técnico 1. Madrid: Instituto de Información Sanitaria. Agencia de Calidad del SNS. https://www.msssi.gob.es/estadEstudios/estadisticas/normalizacion/CIE10/2010_HaciaCIE10.pdf
- Ministerio de Sanidad, Servicios Sociales e Igualdad (2014). Sistema de Información Sanitaria del Sistema Nacional de Salud. <https://www.msssi.gob.es/fr/estadEstudios/estadisticas/sisInfSanSNS/pdf/SISNS.pdf>
- Ministerio de Sanidad, Servicios Sociales e Igualdad (2017a). Registro de Altas – CMBD estatal. Atención Ambulatoria Especializada CMBD-AAE. Manual de definiciones y glo-

- sario de términos. <http://pestadistico.inteligenciadegestion.msssi.es/publicoSNS/comun/DescargaDocumento.aspx?IdNodo=6435>
- Ministerio de Sanidad, Servicios Sociales e Igualdad (2017b). i-CMBD. Indicadores y ejes de análisis del CMBD de Hospitalización del Sistema Nacional de Salud. <http://icmbd.es/docs/modeloIndicadores.pdf>
- Ministerio de Sanidad, Servicios Sociales e Igualdad (2017c). Portal Estadístico. Conjunto Mínimo Básico de datos – Hospitalización (CMBD-H). Acerca de. <http://pestadistico.inteligenciadegestion.msssi.es/publicoSNS/Comun/ArbolNodos.aspx?idNodo=6383>
- Ministerio de Sanidad, Servicios Sociales e Igualdad (2017d). Portal Estadístico. Conjunto Mínimo Básico de datos – Hospitalización (CMBD-AAE). Acerca de. <http://pestadistico.inteligenciadegestion.msssi.es/publicoSNS/Comun/ArbolNodos.aspx?idNodo=6425>
- Ministerio de Sanidad, Servicios Sociales e Igualdad (2018). Notas de Prensa. Dolors Montserrat pide al Consejo Asesor que trabaje para establecer las bases de Sistema Nacional de Salud de próximas décadas (14 de febrero de 2018). <https://www.msssi.gob.es/gabinete/notas-Prensa.do?id=4297>
- Orden de 6 de septiembre de 1984 por la que se regula la obligatoriedad del informe de alta. Boletín Oficial del Estado. 14 de septiembre de 1984. 221; 26685-26686.
- Orden SSI/1885/2015, de 8 de septiembre, por la que se modifica la Orden de 21 de julio de 1994, por la que se regulan los ficheros con datos de carácter personal gestionados por el Ministerio de Sanidad y Consumo. Boletín Oficial del Estado. 17 de septiembre de 2015. 223; 81999-82001.
- Ramos Martín-Vegue, AJ, Vázquez-Barquero JL y Herrera Castanedo, S. (2002). CIE-10 (I): Introducción, historia y estructura general. // Papeles Médicos. 11 :1, 24-35.
- Real Decreto 1360/1976, de 21 de mayo, por el que se hace obligatorio el uso, por parte los establecimientos sanitarios con régimen de internado, de un libro de registro. //Boletín Oficial del Estado. 23 de junio de 1976. 150, 12200.
- Real Decreto 136/1993, de 29 de enero, por el que se aprueba el Plan Estadístico Nacional 1993-1996. // Boletín Oficial del Estado. 5 de marzo de 1993. 55; 7012-7013.
- Real Decreto 1658/2012, de 7 de diciembre, por el que se aprueba el Plan Estadístico Nacional 2013-2016. // Boletín Oficial del Estado. 8 de diciembre de 2012. 295; 84141-84322.
- Real Decreto 69/2015, de 6 de febrero, por el que se regula el Registro de Actividad de Atención Sanitaria Especializada. // Boletín Oficial del Estado. 10 de febrero de 2015. 35; 10789-10809.
- Real Decreto 410/2016, de 31 de octubre, por el que se aprueba el Plan Estadístico Nacional 2017-2020. // Boletín Oficial del Estado. 18 de noviembre de 2016. 279; 80830-80995.
- Reglamento (CE) N° 1338/2008 del Parlamento Europeo y del Consejo de 16 de diciembre de 2008 sobre estadísticas comunitaria// s de salud pública y de salud y seguridad en el trabajo. Diario Oficial de la Unión Europea. 31 de diciembre de 2008. L354: 70-81.
- Reglamento (CE) N° 282/2014 del Parlamento Europeo y del Consejo de 11 de marzo de 2014 relativo a la creación de un tercer programa de acción de la Unión Europea en el ámbito de la salud para el período 2014-2020 y por el que se deroga la Decisión N° 1350/2007/CE. // Diario Oficial de la Unión Europea. 21 de marzo de 2014. L86: 1-13.
- Salvador Oliván, J.A. (1997). Sistemas de información hospitalarios: el C.M.B.D. Scire. 3:2, 115-130.
- Sanz Sebastián, C. (2017). El Registro de Actividad de Atención Sanitaria Especializada (RAE-CMBD) la experiencia de implantación. // 8º Foro sobre el Sistema de Información del Sistema Nacional de Salud. Madrid, Ministerio de Sanidad, Servicios Sociales e Igualdad. 18 de octubre de 2017.
- Sierra, R. (2009). El Ministerio debe darle utilidad de verdad al CMBD. // Diario Médico, 3 de noviembre de 2009.
- Sierra, R. (2016). Un CMBD de valor incalculable... pero no este año. // Diario Médico, 10 de octubre de 2016.
- Treviño FM (1988). Uniform minimum data sets: in search of demographic comparability. // American Journal of Public Health. 78:2, 126-127.
- Wand, L.; Alexander, CA. (2015). Big data in medical applications and health care. // Current Research in Medicine. 6:1, 1-8.
- Zapatero Gaviria, A (2010). ¿Por qué es importante el informe médico de alta? // Revista Clínica Española. 210:7, 355-358.

Enviado: 2018-05-28. Segunda versión: 2018-06-07.
Aceptado: 2018-06-17.

Aproximación bibliométrica a las contribuciones de Hope A. Olson en el campo de la organización del conocimiento

Bibliometric approach to the contributions of Hope A. Olson to Knowledge Organization

María Camila RESTREPO FERNÁNDEZ

Colombia, camila0330.maria@gmail.com

Resumen

Se presenta la trayectoria académica de Hope A. Olson, cuya producción bibliográfica se caracteriza por una marcada crítica hacia los principios de universalidad y jerarquía presentes en las clasificaciones documentales. Para caracterizar su producción se analizaron sus publicaciones indexadas en Scopus hasta 2015 y, posteriormente, se procesó toda la información por medio de la metodología bibliométrica CAMEO con la asistencia de Pajek, VOSviewer y Nvivo. El resultado de la investigación es una caracterización reticular de las contribuciones teóricas de la autora.

Palabras clave: Olson, Hope A. Organización del conocimiento. Clasificación documental. Lenguajes documentales. Clasificación Decimal Dewey.

Abstract

This research presents the academic trajectory of Hope A. Olson, whose bibliographic production is characterized by a marked criticism of the principles of universality and hierarchy present in library classifications. In order to characterize the academic production, its publications indexed in Scopus up to 2015 were taken as a center of analysis and all the information was then processed through the bibliometric methodology CAMEO with the assistance of Pajek, VOSviewer and Nvivo. The result of the research is a reticulated characterization of the author's theoretical contributions.

Keywords: Olson, Hope A. Knowledge Organization. Classification. Library classification. Indexing language. Decimal Dewey Classification.

1. Introducción

El trabajo de Hope A. Olson versa sobre la aplicación de teorías y metodologías feministas, posestructuralistas y poscolonial en el campo de la organización del conocimiento (OC), desde el redil bibliotecológico, con énfasis en los sistemas de clasificación.

A lo largo de su obra, Olson explora la presunta necesidad de usar lenguajes universales para representar o nombrar el conocimiento con el objetivo de ser recuperado, explica las consecuencias del uso de estos lenguajes y propone técnicas para un cambio paulatino en la representación del conocimiento en los catálogos de las bibliotecas, percibiendo como un asunto problemático la universalización en la representación y organización del conocimiento.

Olson considera que tal como se representa el conocimiento, las bibliotecas funcionan como espejos de la marginalización y exclusión de las sociedades a las que sirven, razón que motiva a que su trabajo en OC se apoye en discursos teóricos, ya mencionados al inicio, que se oponen a la universalidad en las ciencias.

Sus publicaciones, entre las que se encuentran 3 libros completos, 33 capítulos de libro y más de

40 artículos publicados desde 1991, validan las contribuciones de Olson como un cúmulo importante en el campo de estudio de OC, así mismo sus intersecciones con los estudios de género y las Ciencias Sociales enriquecen su trabajo.

Revisitar el trabajo académico de Hope A. Olson posibilita: ver el impacto que ha tenido en otros autores, visibilizar nuevos autores con enfoques teóricos y metodológicos en su misma línea investigativa y evidenciar el poder de las clasificaciones y la indización como mecanismos reproductores de exclusión (en determinados casos), por eso se hace categórico revisar su trabajo.

Así lo prueba el número A Festschrift for Hope A. Olson (2016) de la revista Knowledge Organization, en el que una polifonía de voces, compuesta por especialistas reconocidos en OC, aborda y discute el enfoque teórico y metodológico, así como las cuestiones éticas sobre OC presentes en su obra. Los autores del número conmemorativo, a modo grosso, resaltan la labor de Olson al proveer al campo de OC de un suelo epistemológico renovado, alejado de enfoques tradicionales predominantes en el campo. De los trabajos compilados en el número en mención, sobresale un análisis infométrico (Dietmar, 2016, p. 331-337)

elaborado a partir de citas recolectadas en Google Scholar y Thomson Reuters Web of Science, estudio que concluye en la afirmación de la influencia que puede provocar un individuo, en el caso particular Olson, en el campo y sus investigadores. Otra estudio que expone la relevancia de Olson es Alfaya Lamas (2012) quien califica como representativa la investigación de Olson sobre género y sistemas de organización de conocimiento.

Dentro de este contexto, las inquietudes y cuestionamientos deben dirigirse al futuro de investigación y formación en OC, no sin antes revisar a aquellos autores indispensables para aproximarse a respuestas que abran aún más preguntas de las existentes. Los autores que contribuyen al fortalecimiento de la organización del conocimiento son tantos como diversos por los paradigmas y enfoques que emplean. Autores como Hjørland, Dahlberg, San Segundo Manuel, Cabré, Smiraglia, Izquierdo, Guimarães, Olson (Oliveira, Gracio, & Silva, 2010), entre otros, han acumulado un trabajo valioso y en conjunto heterogéneo, por lo que estudiar a cualquiera de ellos mediante la aplicación de estudios egocéntricos contribuiría benéficamente a la OC como campo de conocimiento.

El propósito de la investigación tuvo tres ejes: identificar los referentes teóricos sobre los que Olson sustenta su trabajo académico, caracterizar los autores que citan las obras de Olson indexadas en Scopus como base para su producción, identificando sus campos disciplinares y procedencia institucional como vía para aproximarse al impacto del trabajo académico de la autora y establecer las relaciones de Olson con otros autores del campo de la organización de conocimiento y conexos, mediante la identificación de las obras en coautoría.

2. Método

La investigación se basa en un enfoque cualitativo, desde el que se pretende explorar el trabajo académico de Hope A. Olson teniendo como insumo principal la producción indexada en Scopus hasta 2015, empleando la metodología CAMEO, que aunque su autor la denomine 'metodología' en este proyecto su empleo es netamente instrumental. La producción de Olson que conforma el universo finito de análisis se seleccionó bajo un muestreo por conveniencia, esto es, un tipo de muestreo no probabilístico, esta clase de muestreo ampara lo aleatorio del periodo temporal seleccionado, permitiendo explorar con mayor rapidez, facilidad y economía (Páramo, 2013) la producción de Olson. La muestra se abordó desde

un estudio exploratorio, este método permite tener un acercamiento cualitativo a los datos extraídos por la aplicación de los CAMEOs y a la descripción relacional de Olson en el campo de la OC.

Bajo el enfoque, método e instrumento descritos, se realizó la investigación en tres fases, la primera de ellas avocada a la lectura y revisión de la bibliografía de la autora, la segunda a la aplicación de CAMEO y Pajek y una última fase para el análisis de los resultados.

Characterizations Automatically Made and Edited Online o CAMEO por sus siglas en inglés, es una técnica de visualización de información científica creada por Howard White, diseñada para crear perfiles bibliométricos, dichos perfiles pueden corresponder a un autor, es decir, su objetivo es crear redes egocéntricas centradas en un autor cuya utilidad es ofrecer una breve caracterización de los autores a partir de su producción científica, en el caso de las redes centradas en sujetos CAMEO posibilita la creación de cuatro tipo de perfiles: *citation identity*, todos los autores citados por el autor focal; *citation image-maker*, todos los autores que citan al autor focal; *citation image*, autores cocitados con el autor focal; y *co-authors*, coautores del autor focal.

Los cuatro perfiles permiten caracterizar a un autor y su entorno de producción más cercano, obteniendo información sobre sus principales campos de interés y los lazos que tiene en el ámbito académico por medio de la identificación de aquellos que cita, lo citan, coautores y niveles de cocitación.

White publicó la metodología CAMEO en 2001 en el N°. 51 de *Scientometrics*, en el artículo específico que la metodología debe ser aplicada en Web of Science debido a que el uso de las aplicaciones Scisearch y el software Dialog era imperativo para obtener de manera automática los perfiles, actualmente la metodología puede aplicarse en otras bases de datos con igual éxito que en Web of Science, mediada por otras herramientas bibliométricas. Aunada a la propuesta metodológica de White, se consideraron las contribuciones de Aji-feruke & Wolfram (2010) como referentes para la etapa de análisis de los resultados.

La investigación en curso aplicó tres perfiles de CAMEO: *citation identity*, *citation image-maker* y *citation image* para caracterizar la producción de Olson, para ello se empleó la base de datos Scopus en la que se albergan 26 artículos de Olson publicados entre 2000–2015 (Ver bibliografía).

Los artículos fueron extractados de la base de datos con la siguiente fórmula de búsqueda:

AUTHOR-NAME (olson AND h. AND a.) AND PUBYEAR > 1999 AND PUBYEAR < 2015.

De los artículos hallados se encontraron 177 artículos publicados bajo el mismo periodo temporal que citan alguna de las veintiséis publicaciones; los artículos y las citas fueron el insumo principal para la elaboración de las gráficas y tablas que están registradas en el tercer apartado del documento en lectura.

Para la construcción de los gráficos se empleó Pajek, un programa de software para el análisis y gráfica de redes ampliamente usado en los estudios bibliométricos que manejan grandes volúmenes de información, su uso es gratuito y la creación de redes es automática. Con Pajek se puede elaborar: genealogías, flujos de datos de programas, bibliografías, redes de citaciones, entre otros. La función de Pajek es soportar la abstracción mediante factorialización de grandes redes en diversas redes de inferior tamaño susceptibles de ser tratadas con métodos más sofisticados; proporcionar herramientas potentes para visualizar estas redes e implementar una selección de algoritmos eficientes para análisis de grandes redes.

Pajek fue crucial para la investigación, porque permitió graficar las redes de Olson de manera automática y sin costo. Para hacerlo se exportaron los resultados obtenidos después de aplicar el parámetro de búsqueda, seleccionando all available information en formato .RIS. El archivo Scopus.ris se colocó en una carpeta con el programa Scopus.exe disponible en <<http://www.leydesdorff.net/scopus/index.htm>> y al procesar la información se generaron tres nuevas bases de datos: Los títulos en TI.DBF, autores en AU.DBF y referencias en CR.DBF.

Para la graficación se importaron las listas relacionales desde un texto plano al programa Pajek disponible en <<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>>. Las gráficas son mapas de redes donde los puntos o nodos son los elementos y las líneas, las relaciones. El tamaño de los nodos representa además la cantidad de relaciones que tiene cada nodo, conocido en el análisis de redes como grado de centralidad. En todas las redes la ruta para obtener la centralidad de grado: en el panel central se seleccionó Network/create vector/centrality/degree/all.

Inicialmente el programa Pajek generó redes de dos modos: a) dónde se relacionan los artículos con alguno de los otros parámetros, en todos los casos se transformaron a redes de un nodo, y b) dónde los elementos se relacionan con los de su mismo tipo (autores con autores, palabras con palabras, referencias con referencias). Para rea-

lizar el cambio se siguió la siguiente ruta: Network/2-mode network/2-mode to 1-mode/columns. Finalmente, las gráficas fueron exportadas desde Pajek a VOSviewer disponible en <<http://www.vosviewer.com/>> para mejorar el aspecto gráfico de la red, aplicativo de uso libre desarrollado por la Universidad de Leiden. Los resultados de la aplicación de Pajek se presentan en el siguiente apartado del texto.

3. Resultados

Los resultados se presentan reticularmente en Red de autores citados, Red de autores que citan, Red coautoría y Redes semánticas.

3.1. Red de autores citados por Olson en los artículos 2000-2015

En los 26 artículos de Olson hay 693 referencias, para un promedio de 27 referencias por artículo. En ellas se citan a 382 autores incluyendo a Olson, y exceptuando la autocitación hay una baja redundancia en las referencias, A continuación, se presenta el listado de los autores referenciados como mínimo 3 veces.

Autores	fi	Autores	fi
Olson, H.A	47	Fraiser, M	3
Cutter, C.A	9	Gorman, M	3
Dewey, M	9	Green, R	3
Markey, K	5	Hegel, G.W.F	3
Mehra, B	5	Hildreth, C.R	3
Nunberg, G	5	Hill, J.S	3
Swanson, D.R	5	Hjorland, B	3
Chan, L.M	4	Iivonen, M	3
Cleverdon, C.W	4	Kipp, M.E.I	3
Harris, W.T	4	Kublik, A	3
Intner, S.S	4	Maron, M.E	3
Smiraglia, R.P	4	Martinez-Ávila, D.	3
Svenonius, E	4	Miksa, F	3
Weinberg, B.H	4	Ranganathan, S.R	3
Wiegand, W.A	4	Sayers, W.C.B	3
Avram, H.D	3	Spivak, G.C	3
Ellis, D	3	Taylor, A.G	3
Fister, B	3	Wolfram, D	3
Foucault, M	3	Yee, M.M	3

Tabla 1. Autores citados por Olson

67 autores son citados en dos artículos diferentes (17,5 % del total de autores) y 277 (72,5 %) una

única vez. La relación es susceptible de ser representada en la red, si dos o más autores figuran en las referencias de un mismo artículo, debido a que el autor los utiliza para una misma investigación. En los artículos de la muestra hay veintisiete referencias por artículo, lo que quiere decir que cada autor referenciado tiene en promedio veintiséis relaciones con otros autores. Cuando de los 27 autores referenciados hay un

par de ellos que aparecen juntos en otros artículos indicaría que la relación entre su investigación podría ser más grande, a mayor co-ocurrencia mayor similaridad entre los trabajos académicos de los autores. El total de relaciones entre referencias fue de 597, generando una red densa e ilegible, por lo que se realizaron operaciones adicionales para transformarla en una red con autores referenciados en 2 o más artículos.

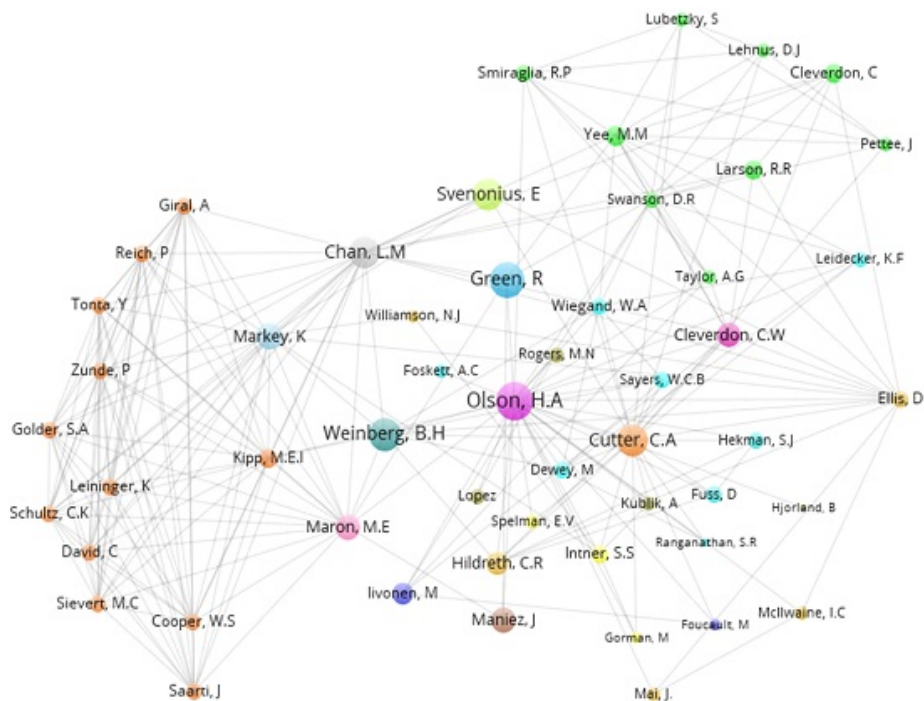


Figura 1. Red de principales autores citados por Olson

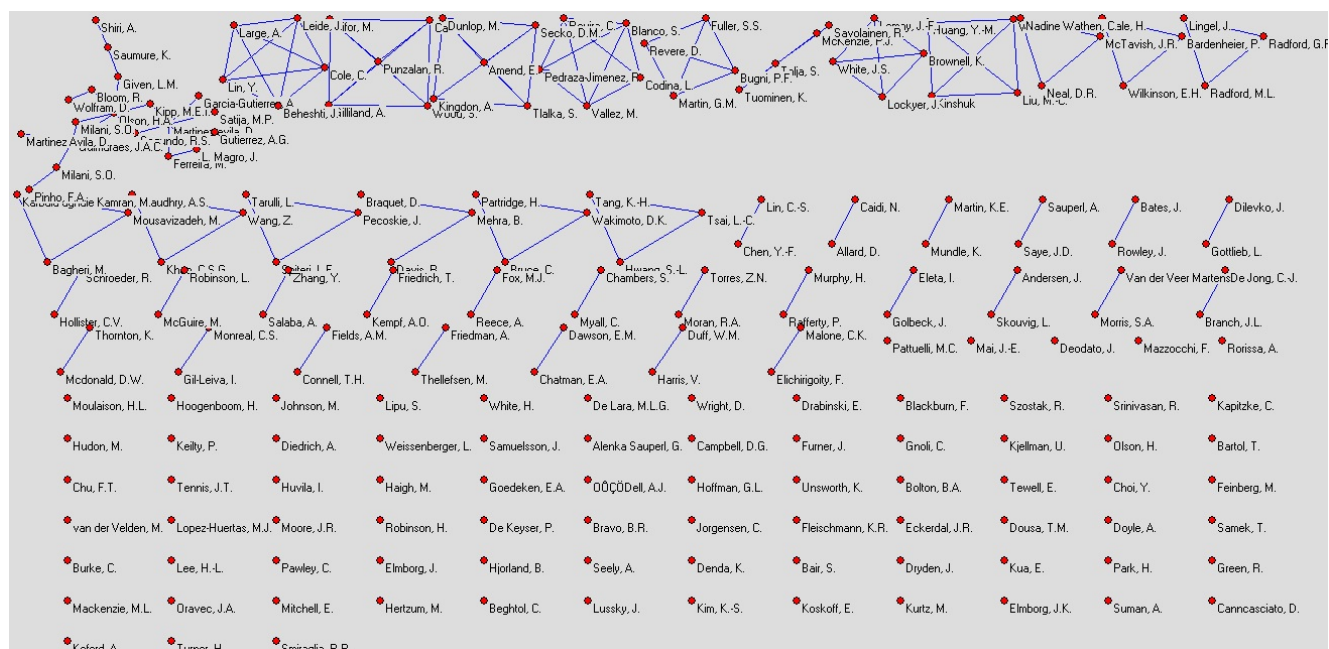


Figura 2. Red total de autores que citan a Olson

3.2. Red de autores que citan a Olson 2000-2015

En total se hallaron 212 autores que citan a Olson, incluyéndola; 80 son autores únicos y 59 escribieron en co-autoría. Es relevante aclarar que, exceptuando la gran red que hay en la parte superior izquierda de la Figura 3, los grupos de autores que figuran en la gráfica escribieron cada uno un solo artículo. Para graficar de forma más clara, sólo se destacó la red de la parte superior izquierda de la Figura 2 (en la página anterior), para ello se ejecutó la ruta: Network/components/weak y se seleccionó que identificara la red con más de diez autores, luego al seleccionar Operation/extract subnetwork y seleccionar “1” se eliminó a todos los autores que no hacen parte de la red principal, dando como resultado la Figura 3, presentada a continuación:

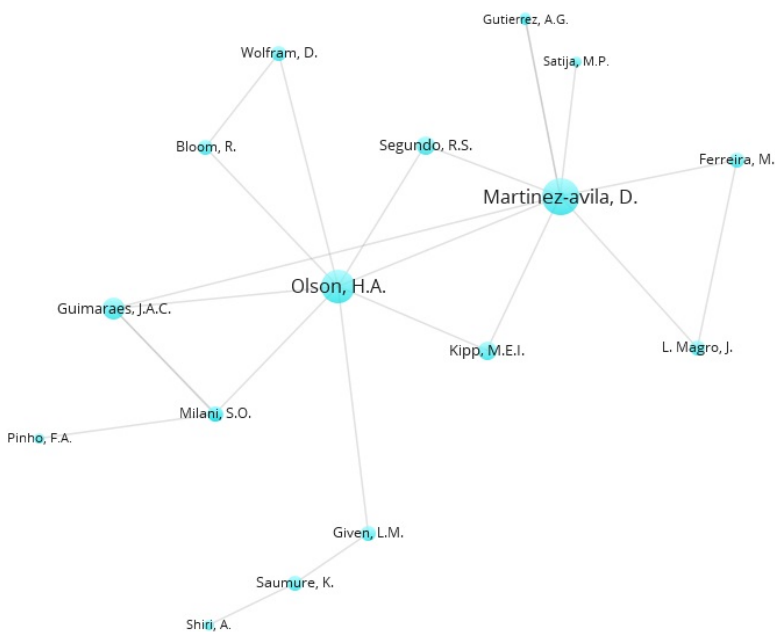


Figura 3. Red de principales de autores que citan a Olson

3.2. Red de coautoría 2000-2015

Esta es una red egocéntrica, lo que explica la centralidad del nodo ‘Olson’, los demás nodos son periféricos, cada nodo numerado representa a uno de los veintiséis artículos de la muestra, la numeración fue aleatoria. Cada línea relaciona al artículo con los autores que lo escribieron, así por ejemplo el artículo cuatro sólo fue escrito por Olson, mientras que el artículo diecinueve fue escrito por Clevette, Ward, Kublik y Olson.

Se tomó la red de relaciones entre autores y artículos y se eliminaron los artículos de la gráfica

para que la red fuera sólo de relaciones entre autores, sin embargo, las líneas siguen representando la co-autoría en los mismos artículos.

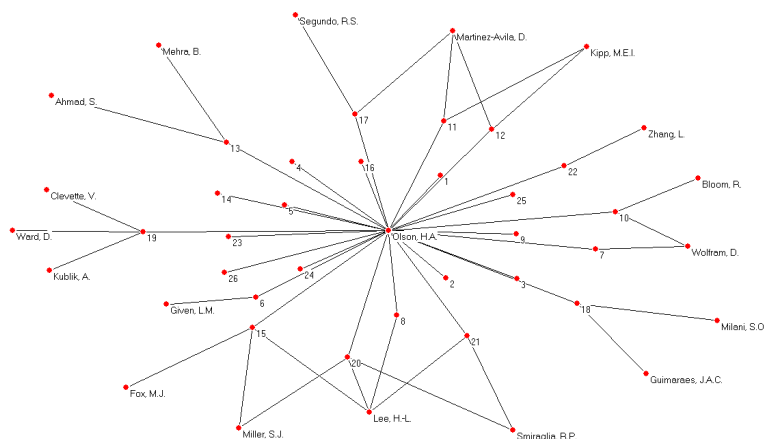


Figura 4. Red de Coautoría

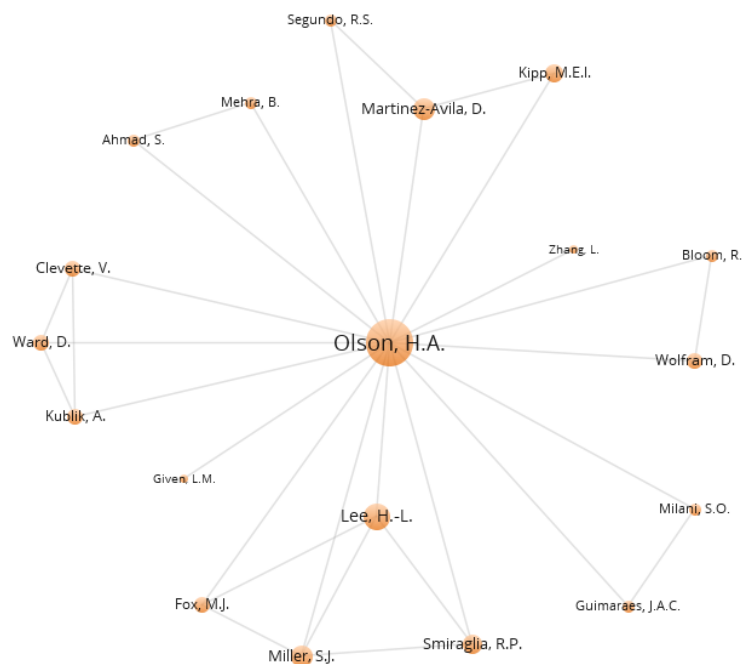


Figura 5. Red de Coautoría

3.3. Red semántica

En los 26 artículos de Olson hay 87 palabras clave, de estas 75 aparecen en un solo artículo, y sólo doce tuvieron aparición en al menos dos artículos. En la Tabla II, en la página siguiente, se presentan las palabras clave que se utilizaron en dos o más artículos.

A pesar de que la baja continuidad entre las palabras, sí se presentan muchas relaciones porque cada palabra se relaciona con las otras palabras clave de su mismo artículo. En la gráfica de redes se puede ver claramente subgrupos de temas relacionados entre sí densamente, que corresponderían a las palabras de un mismo artículo y a unas palabras puente como son las de recurrencia en varios artículos. Para que Pajek identificara los nodos que hacen parte de un mismo campo temático por sus relaciones mutuas se ejecutó la ruta: Network/create partition/communities/Louvain Method/Multi-level coarsening+multi-level refinement, con el parámetro de resolución de '1', dando como resultado la Figura 6.

Palabras clave	fi
Classification	5
Knowledge management	4
Libraries	4
Information organization	3
Organization of knowledge	3
Subject Access	3
Authorship	2
Cataloging codes	2
Cross-cultural studies	2
Dewey decimal classification	2
Feminism	2
Library and information science	2

Tabla II. Palabras clave en los artículos de Olson

3.4. Red semántica de los artículos que citan a Olson 2000-2015

Se hallaron 177 artículos que citan alguno de los artículos de Olson publicados entre 2000-2015 en Scopus, de los cuales se identificaron 574 palabras clave, de las que 472 se utilizaron una sola vez y 59 en, al menos, dos artículos. A continuación, se presenta la Tabla III de las palabras que recurrieron tres o más veces en los artículos:

Palabras clave	fi	Palabras clave	f i	Palabras clave	f i
Classification	1 6	Semantic web	5	Information management	3
Knowledge management	1 3	Digital libraries	4	Information retrieval systems	3
Knowledge organization	1 2	Indexing	4	Internet	3
Information retrieval	9	Information systems	4	Knowledge representation	3
Indexing (of information)	8	Ontology	4	library services	3
Information science	8	article	3	Methodology	3
Libraries	8	Bias	3	Philosophical aspects	3
Classification of information	7	controlled vocabularies	3	PUBLIC libraries	3
Design	6	culture	3	Search engines	3
Ethics	6	Dewey decimal classifications	3	Semantics	3
Information organization	6	Education	3	Students	3
Social tagging	6	human	3	Teaching	3
Thesauri	6	Indexing language	3	Theoretical foundations	3
Cataloging	5	Indigenous knowledge	3	Web 2.0	3
Classification system	5				

Tabla III. Palabras clave en artículos que citan a Olson

La red generada es densa, como se ve en la Figura 7, para obtener una gráfica más clara se eliminaron las palabras clave que tuvieran relación con menos de veintiún nodos. Para obtenerla se creó una partición de grado ejecutando network/create partition/degree/all, luego se extrajeron las palabras con operation/network+partition/extrac subnetwork. Además, para que el programa identificara los nodos que hacen parte de un mismo espacio temático por sus relaciones mutuas, se ejecutó en Pajek la ruta: Network/create partition/communities/Louvain Method/Multi-level coarsening+multi-level refinement, con un parámetro de resolución de '1'. El resultado se exportó al programa VOSviewer y como resultado se obtuvo la Figura 8.

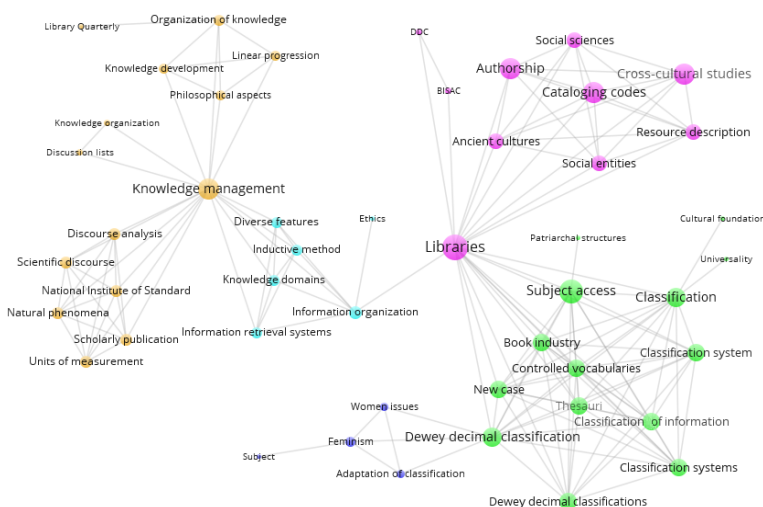


Figura 6. Red semántica Olson



Figura 7. Red semántica Total de artículos que citan a Olson

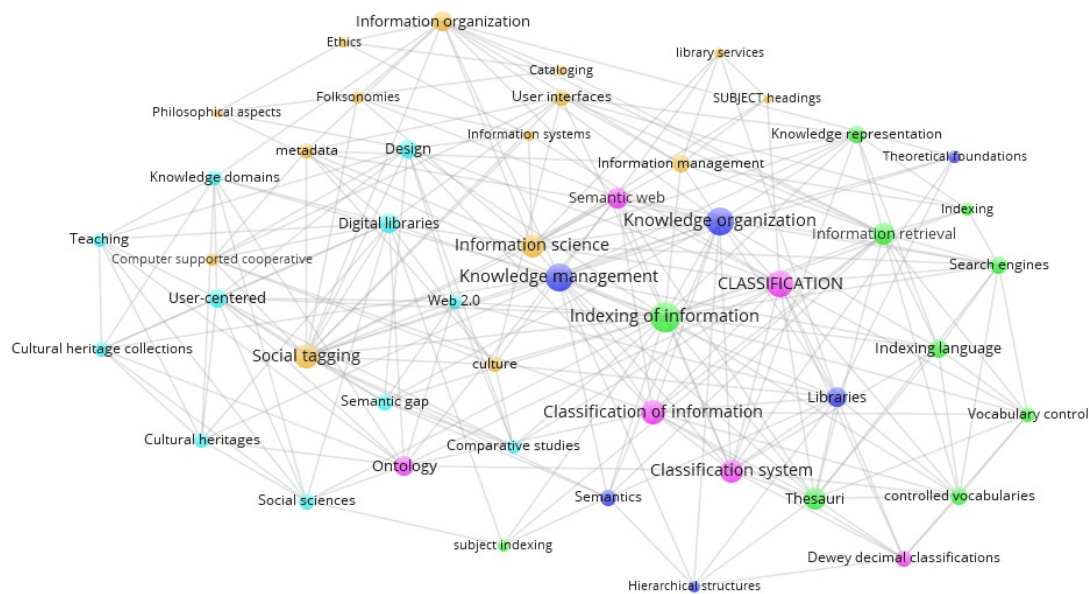


Figura 8. Red semántica de artículos que citan a Olson

4. Conclusiones

El estudio de la obra de Hope A. Olson, mediante la aplicación de la metodología CAMEO con la asistencia del programa Pajek arrojó resultados esperados, debido al conocimiento previo de su obra, pero permitió dilucidar aspectos de suma importancia para comprender sus aportes teóri-

cos al campo de la clasificación desde una perspectiva crítica y subjetivista, concentrando gran parte de su trayectoria al estudio sistemático de la DDC, por sus siglas en inglés, Dewey Decimal Classification y su impacto en la cultura a través de las bibliotecas, especialmente las estadounidenses. Para ello, Olson recurre a referentes conceptuales propios del campo, su referente

más común son sus propios trabajos, de la producción que hizo parte de la muestra de este estudio. Olson se citó a sí misma 47 veces, lo que significa que en cada artículo cita aproximadamente 1.8 de sus artículos, como única autora o coautora, empero el resultado no es sorprendente, porque debido a la originalidad y unicidad de sus trabajos académicos es imperativo que recurra a sus propios comentarios para dar suelo firme a sus nuevas acepciones sobre clasificación y organización del conocimiento.

Luego, entre sus referencias más usuales se encuentran dos clásicos de la clasificación: Dewey y Cutter. Sobre el primero, sus referencias son a la DDC, sus comentarios a la clasificación bibliográfica más ampliamente traducida y usada en el mundo, se enfoca en las implicaciones de los números que empleados para la representación de los grupos sociales y culturas históricamente marginalizados en la clasificación en cuestión, así como en los vocabularios controlados utilizados en el proceso de la indización, especialmente el LCSH. Olson indaga específicamente por el lugar/notación del feminismo entre las notaciones de la clasificación y compara recurrentemente las denominaciones asignadas a las teorías feministas y a la mujer, ambos como conceptos, en los vocabularios controlados más populares en contraste con algunos más específicos como el *Women's Thesaurus* y *Canadian Feminist Thesaurus*. En la DDC, Olson encuentra pobremente representadas las teorías feministas y estudios sobre la mujer por la que propone en su artículo de 2003 *Adapting Dominant Classifications to Particular Contexts* publicado en *Knowledge Organization and Classification in International Information Retrieval*, la creación de una base de datos y herramientas de acceso para solucionar lo que Olson percibe como un problema de representación, para la base de datos Olson y sus coautoras tipifican el tipo de información contenida en: Vocabularios, una mixtura entre Índice Dewey y el *Women's Thesaurus*; las notaciones de la DDC enlazadas al tesoro y las nuevas notaciones de la DDC, nuevamente relacionadas con términos del tesoro ya mencionado. Las herramientas de acceso o herramienta de acceso es un software sobre el que los autores dicen que debe funcionar con una licencia en abierto que permite tener un software barato y cooperativo, por lo tanto ampliamente accesible y cualificable. La propuesta se sintetiza así: Olson y las autoras proponen una expansión a la DDC relacionándola con el tesoro a fin de representar el feminismo y los estudios sobre la mujer sobrepasando el nivel meramente crítico al presentar una solución metodológica viable y posible, sin embargo se queda en nivel de prototipo porque no hay respuesta o réplicas aún identificadas.

A lo largo de su trabajo, Olson no sólo se refiere a la DDC, sino a las clasificaciones en general como representaciones caprichosas que se rigen bajo los principios de universalidad y exclusividad o jerarquía (2002, pp. 382-385), cuyo objetivo es organizar el universo bibliográfico y documental, de modo que define límites para unir y separar en razón de la similitud y diferencia entre conocimiento para asignarle un número o notación única para ser recuperado de un sistema de información. El disgusto de Olson no es la DDC, aunque se enfoque principalmente en ella; el malestar de Olson es que bajo los principios que rigen las clasificaciones, estas fortalecen la parcialidad en favor de algunos grupos sociales representándolos con mayor fuerza en los sistemas de organización del conocimiento. Como oposición usa a Browker y Star (2002, 389 p.) que proponen tres consideraciones al momento de diseñar una clasificación para favorecer la neutralidad: reconocer la polifonía de voces, ser sensible con las exclusiones y reconocer la clasificación como una actividad de balance, de modo que la iniciativa de Olson es el solapamiento de las clasificaciones y vocabularios controlados para ampliar los límites creados bajo los principios de universalidad y jerarquía.

Las disertaciones de Olson sobre Cutter no varían mucho en relación con los comentarios a Dewey, de manera consistente critica la ubicuidad del principio de jerarquía presente en ambas clasificaciones, con la diferencia de que para Cutter, además de la lógica, el absurdo podría servir para estructurar la clasificación, de una parte la estructura lógica facilita la creación de un orden más eficaz. Para Cutter, lógica y absurdo son necesarios para la organización por lo que acepta la dualidad como principio y lo materializa con la creación del catálogo diccionario, en él ideó una organización por entradas estructuradas jerárquicamente incluyendo referencias cruzadas, es decir, aceptando la lógica y el absurdo (2004, pp. 605-606).

Después de los tres referentes conceptuales ya mencionados, Olson usa los aportes de la profesora de la Universidad de Michigan, Karen Markey, en el campo de indización, las contribuciones del profesor de la Universidad de Tennessee, Bharat Mehra, sobre la justicia social aplicada a las bibliotecas y los estudios culturales, y el trabajo del profesor de la School of Information de Berkeley, Geoffrey Nunberg, con su trabajo sobre el concepto de información y su historia. Sobre esto se puede concluir que los principales referentes de Olson son de origen norteamericano (por la afiliación institucional de los autores mencionados), los cinco desarrollaron y desarrollan su trabajo en Estados Unidos denotando que

aunque Olson se contrapone a una visión clásica de la clasificación, está sujeta a una escuela clásica, aunque renovada, de la bibliotecología.

Los demás referentes de Olson incluyen colegas de la Universidad de Milwaukee como Kipp y Smiraglia, autores clásicos de la bibliotecología como Ranganathan, Don Swanson, Cyril Cleverdon, Elaine Svenonius, entre otros, y autores generales como los filósofos franceses Foucault y Derrida.

Del rastreo hecho en Scopus se hallaron 212 citas, en las que predominó Olson como principal citadora de su trabajo. De manera independiente y en coautoría, Daniel Martínez Ávila es el segundo; y Guimarães el tercero. El trabajo de los tres autores está estrechamente relacionado porque sus inquietudes son similares, incluso como se puede ver en la Figura 3, el centro de la red es la relación entre Guimarães-Olson-Martínez, por lo que se puede deducir que entre estos autores hay una fuerte conexión temática que podría calificarse como un colegio invisible. Esta idea se reafirma en las Imágenes 4 y 5, en la que se pueden ver los enlaces de coautoría entre los autores.

En la Figura 3 también se pueden detallar tres triadas relevantes que podrían sumarse al colegio invisible como periferias, la triada Olson-Martínez-Kipp, la triada Olson-Milani-Guimarães y la triada Olson-Martínez-San Segundo. Las trataremos con mayor detalle: en la primera se relacionan la autora focal con Kipp que es una colega directa de Olson y Martínez-Ávila que es el segundo citador de los trabajos de la autora. Lo común entre estos autores es que en algún momento han trabajado en un grupo de investigación que dirigió Olson durante su tiempo como profesora activa de la Universidad de Milwaukee. Actualmente, Kipp es profesora asistente en la Universidad y Martínez-Ávila ha sido pasante de investigación en la alma mater de Olson. En la segunda triada Olson-Milani-Guimarães se demuestra el impacto de Olson en el ámbito bibliotecológico brasilero, concretamente en los estudios desarrollados en la Universidad Estatal Paulista y, la tercera triada Olson-Martínez-San Segundo representa el impacto de Olson en las investigaciones de autores españoles siendo baja respecto a la representación brasilera. También podría sumarse a esta última triada las citas por parte de Antonio García Gutiérrez, sin embargo, este autor sólo está relacionado con Martínez.

De las relaciones presentadas, además de Olson como autora central de la red egocéntrica, destaca la importancia de Martínez-Ávila que juega como intermediación entre los autores brasileros y españoles, por lo que es valioso destacar su papel en la difusión del trabajo de Olson en ambos

escenarios y su esfuerzo por ensanchar el campo de clasificación a partir de los postulados escritos por Olson durante su trayectoria académica.

A modo de resumen, el impacto de Olson se localiza en Brasil, Estados Unidos, Canadá, España e India, este emplazamiento se fijó a partir de la formación académica de los citadores, que además permite determinar las instituciones de estos países que acogen la obra de Olson: en Brasil es la Universidad Estatal Paulista, en Estados Unidos son la Universidad de Milwaukee y la Universidad de Wisconsin, en Canadá la Universidad de Alberta y la Universidad de Ontario, en España la Universidad Carlos III de Madrid y en India la Universidad Guru Nanak Dev.

El CAMEO de coautoría coincide considerablemente con el CAMEO de citación, de los dieciocho coautores de Olson, ocho hacen parte de los principales citadores de Olson, a saber: Daniel Martínez-Ávila (3), Dietmar Wolfram (2), Margaret Kipp (2), Rosa San Segundo (1), Raina Bloom (1), Suellen Milani (1), Lisa M. Given (1), Jose Augusto Guimarães (1).

Los coautores de Olson que no se ubicaron entre los principales citadores de Olson, sí la citan, pero debido a su número reducido de publicaciones en las que citaron algún trabajo de la autora no quedaron representados en la Figura 3, estos coautores son: Hur-Li Lee (4), Richard Smiraglia (2), Steven J. Miller (2), Dennis Ward (2), Melodie J. Fox (1), Angela Kublik (1), Virginia Clevete (1), Suzana Ahmad (1), Bharat Mehra (1), Lei Zhang (1).

Estos diez coautores confirman, como se mencionó en el apartado anterior, que el impacto de Olson se localiza en países específicos, en este caso sus colaboraciones coinciden con los países mencionados, Estados Unidos y Canadá, porque Lee, Smiraglia, Miller, Ahmad, Fox y Bharat están afiliados en calidad de profesores y/o investigadores a universidades norteamericanas y Ward, Kublik, Clevete y Zhang a universidades canadienses. En ambos casos los coautores de Olson son del campo de la ciencia de la información, con énfasis en clasificación, ontologías, indización, metadatos, entre otros. Olson no tiene colaboraciones con investigadores de áreas diferentes a las de la ciencia de la información, cabe aclarar que es posible que las tenga y no estén incluidas en la colección de Scopus por lo que se ignora tal aspecto.

El trabajo de Olson, de acuerdo con la indización temática de sus publicaciones en Scopus, se encamina capitalmente por la clasificación, de igual manera la indización temática de las publicaciones que citan los trabajos de Olson también se

enfocan mayoritariamente en la clasificación, como asunto cardinal de la publicación.

El segundo tema más común entre las publicaciones de Olson y las publicaciones de sus citadores es "Knowledge Management" que en español se traduce como gestión del conocimiento. Es sorprendente, porque en el trabajo de Olson tal enfoque no es fundamental, sin embargo, mediante el análisis de las palabras claves, el término reluce como el segundo más tratado.

Los demás temas encontrados son irrefutables, entre ellos: Information organization, Organization of Knowledge, Subject access, Cataloging codes, DDC, Feminism, Library Science, entre otros. Estas materias distinguen el enfoque de Olson y caracterizan su trayectoria, por lo que es frecuente que los términos que predominan en la red semántica de sus citadores sean Knowledge organization, Classification of information, Libraries, Social tagging, Cataloging, Semantic web, Ontology, entre otros, que pueden ser vistos en la Tabla 3, Figura 7 y Figura 8, lo que implica que el impacto del trabajo de Olson se ha extendido a áreas distintas y conexas a la clasificación.

Las contribuciones de Olson son sobresalientes, por la originalidad y criticismo con los que examina el campo matricial de la OC, la clasificación.

Notas

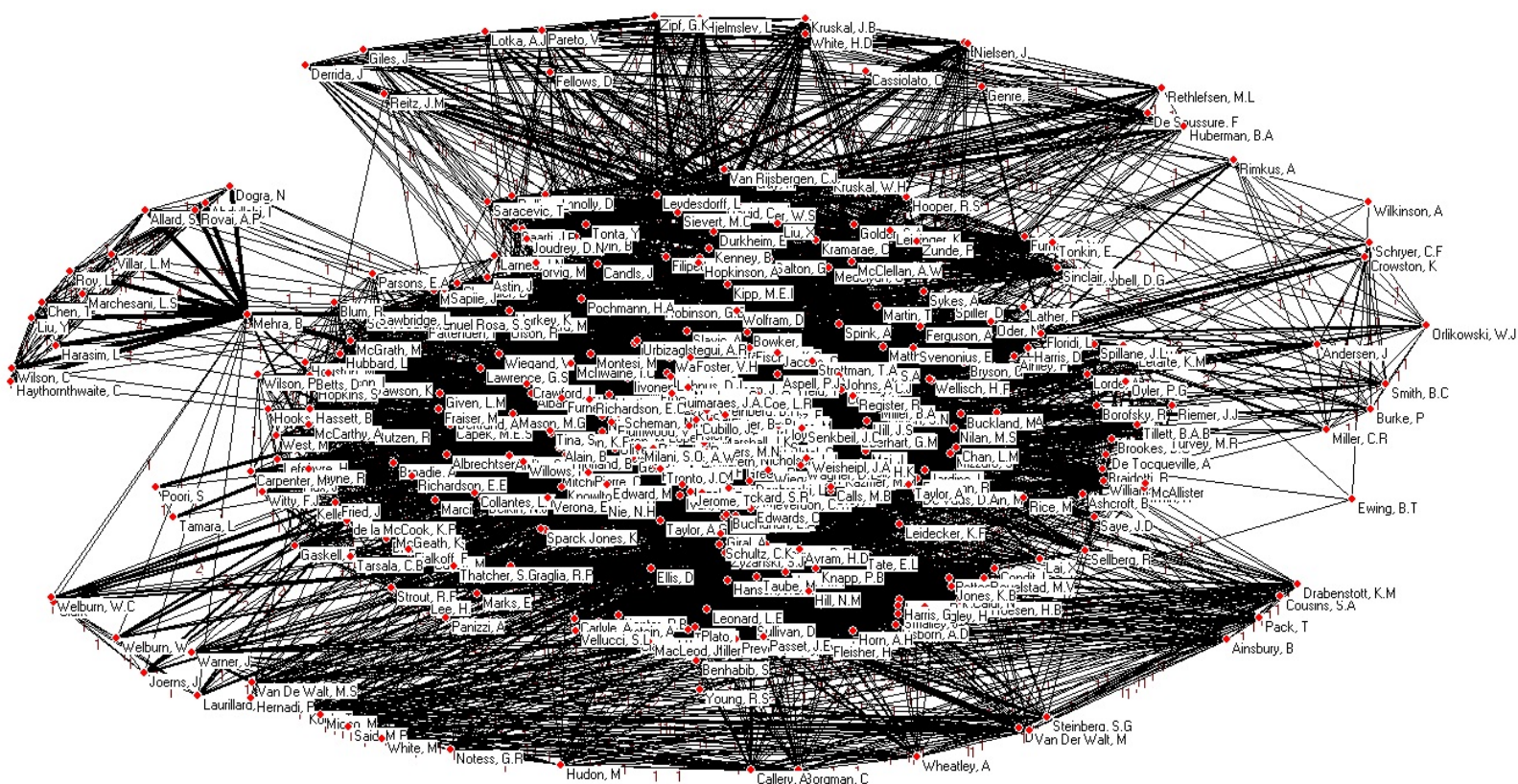
Existen otros estudios métricos sobre el trabajo de Olson, se recomienda la revisión complementaria de D. Wolfram (2016).

Apéndice 1. Bibliografía de Hope A. Olson

- Given, L. M.; Olson, H. A. (2003). Knowledge organization in research: A conceptual model for organizing data. // *Library and Information Science Research*. 25:2, (Summer 2003) 157–176.
- Lee, H.-L.; Olson, H. A. (2005). Hierarchical navigation: An exploration of yahoo! Directories. // *Knowledge Organization*. 32:1 (Enero 2005) 10–24.
- Martínez-Ávila, D.; Kipp, M. E. I.; Olson, H. A. (2012). DDC or BISAC: The changing balance between corporations and public institutions. // *Knowledge Organization*. 39:5 (2012) 309–319.
- Martínez-Ávila, D.; Olson, H. A.; Kipp, M. E. I. (2012). New roles and global agents in information organization in Spanish libraries. // *Knowledge Organization*. 39:2, 125–136.
- Martínez-Ávila, D.; Segundo, R. S.; Olson, H. A. (2014). The use of BISAC in libraries as NEW cases of reader-interest classifications. // *Cataloging and Classification Quarterly*, 52:2 (January 2014) 137–155.
- Mehra, B.; Olson, H. A.; Ahmad, S. (2011). Integrating diversity across the LIS curriculum: An exploratory study of instructors' perceptions and practices online. // *IFLA Journal*. 37:1 (Marzo 2011) 39–51.
- Milani, S. O., Guimarães, J. A. C.; Olson, H. A. (2014). Bias in subject representation: Convergences and divergences in the international literature. // Babik W. (Ed.), *Advances in Knowledge Organization*. Poland: International Society for Knowledge Organization, 2014. 335–342.

- Miller, S. J.; Fox, M. J.; Lee, H.-L.; Olson, H. A. (2006). Great expectations: Professionals' perceptions and knowledge organization curricula. // *In Advances in Knowledge Organization*. Vienna: International Society for Knowledge Organization, 2006. 349–358.
- Miller, S. J., Lee, H.-L., Olson, H. A.; Smiraglia, R. P. (2012). Online cataloging education at the University of Wisconsin-Milwaukee. // *Cataloging and Classification Quarterly*, 50:2-3 (January 2012) 110–126.
- Olson, H. (2007). Cultural infrastructure: The story of how classification came to shape our lives. // *Annual Advances in Classification Research Online*. 18. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84870023860&partnerID=40&md5=ef860544f7342a3e008a9793c1c3eaff>
- Olson, H. A. (2001). Patriarchal structures of subject access and subversive techniques for change. // *Canadian Journal of Information and Library Science*. 26:2-3 (June 2001) 26–29.
- Olson, H. A. (2001). Sameness and difference a cultural foundation of classification. // *Library Resources and Technical Services*. 45:3 (January 2001) 115–122.
- Olson, H. A. (2001). The power to name: Representation in library catalogs. // *Signs*. 26:3 (Springs 2001) 639–668.
- Olson, H. A. (2002). Classification and universality: Application and construction. // *Semiotica*. 139 (March 2002) 377–391.
- Olson, H. A. (2004). The ubiquitous hierarchy: An army to overcome the threat of a mob. // *Library Trends*. 52:3 (Winter 2004) 604–616.
- Olson, H. A. (2006). Codes, costs, and critiques: The organization of information in *Library Quarterly*, 1931-2004. // *Library Quarterly*. 76:1 (January 2006) 19–35.
- Olson, H. A. (2007). How we construct subjects: A feminist analysis. // *Library Trends*. 56:2 (Fall 2007) 509–541.
- Olson, H. A. (2008). Wind and rain and dark of night: Classification in scientific discourse communities. // *Advances in Knowledge Organization*. Montreal: ISKO. 235–241.
- Olson, H. A. (2009). Introduction to the special issue on the ethics of information organization. // *Cataloging and Classification Quarterly*, 47:7 (2009) 609–611.
- Olson, H. A. (2010). Earthly order and the oneness of mysticism: Hugh of saint victor and medieval classification of wisdom. // *Knowledge Organization*. 37:2 (2010) 121–137.
- Olson, H. A. (2010). Hegel's epistemograph, classification, and Spivak's postcolonial reason. // *Advances in Knowledge Organization*. Rome: International Society for Knowledge Organization, 2010. 23–30.
- Olson, H. A.; Wolfram, D. (2008). Syntagmatic relationships and indexing consistency on a larger scale. // *Journal of Documentation*. 64:4 (2008) 602–615.
- Smiraglia, R. P., Lee, H.-L.; Olson, H. A. (2011). Epistemic presumptions of authorship. // *ACM International Conference Proceeding Series*. <http://doi.org/10.1145/1940761.1940780>
- Wolfram, D., Olson, H. A.; Bloom, R. (2009). Measuring consistency for multiple taggers using vector space modeling. // *Journal of the American Society for Information Science and Technology*, 60:10 (October 2009) 1995–2003.
- Zhang, L.; Olson, H. A. (2015). Distilling abstractions: Genre redefining essence versus context. // *Library Trends*, 63:3 (Winter 2015) 540–554.
- Kublik, A., Clevette, V., Ward, D.; Olson, H. A. (2013). Adapting dominant classifications to particular contexts. *Knowledge organization and classification in international information retrieval*. // *Library and Information Science Research*, 25:2 (2013) 13-32.

Apéndice 2: Red de autores citados



Referencias

- Alfaya Lamas, E. (2012). La asunción del género neutro en la teoría y práctica de la organización del conocimiento. // Pérez, Carmen.; González Bonome, María de los Angeles (eds.). 20 años del capítulo español de ISKO. Coruña: Universidad de Coruña, 2012. 249-263.
- Ajiferuke, I.; Wolfram, D. (2010). Comparison of citer and Citation-Based measure outcomes for multiple disciplines. // Journal of the American Society for Information Science and technology. 6: 10, (October 2010) 2086-2096
- Ajiferuke, I.; Wolfram, D. (2010). Citer analysis as a measure of research impact: library and information science as a case study. // Scientometrics. 83: 3, 623-638.
- Frías, J.; Travieso, C. (Eds.) (2003). Tendencias de investigación en organización del conocimiento. Salamanca: Universidad de Salamanca, 2003.
- García Marco, F. J. (2002). Un panorama de la investigación en organización y representación del conocimiento. // Scire. 8:1 (Enero 2002) 9-16.
- García-Gutiérrez, A. (2013). La organización del conocimiento desde la perspectiva poscolonial: itinerarios de la paraconsistencia. // Perspectivas em Ciência da Informação. 18:4 (Octubre 2013) 93-111.
- Knowledge Organization (2016). Special Issue: A Festschrift for Hope A. Olson // Knowledge Organization. 43:5 (2016) 313-389.
- Martínez-Ávila, D.; Guimarães, J. A. (2013). Library classifications criticisms: universality, poststructuralism and ethics. // Scire. 19:2 (Julio 2013) 21-26.

- Oliveira, E.; Gracio, M.; Silva, A. (2010). Investigadores con mayor visibilidad en Organización y Representación del Conocimiento. // Scire. 16:2 (2010) 39-45.
- Olson, H. (s.f.). Cv Hope A. Olson. [archivo de computadora].
- Ovalle-Perandones, M., Olmeda-Gómez, C., & Perianes-Rodríguez, A. (2010). Una aproximación al análisis de Redes egocéntricas de colaboración interinstitucional. Redes. // Revista Hispana para el Análisis de Redes Sociales, 19:1 (Diciembre 2010) 168-190.
- Pajek. (S.f.). Software Pajek. Recuperado de <http://pendientemigracion.ucm.es/info/pecar/Software.htm>
- Páramo, P. (2013). La investigación en ciencias sociales: estrategias de investigación. Bogotá: Universidad Piloto de Colombia, 2013.
- Trillo-Domínguez, M.; de Moya-Anegón, F. (2008). Aproximación cuantitativa a la investigación en comunicación: el caso de Marshall McLuhan A scientometric approach to communication research: the Marshall McLuhan case. // El Profesional de la Información. 17:3 (Mayo 2008) 303-309.
- Vosviewer. (S.f.). Welcome to VOSviewer. <http://www.vosviewer.com/>
- White, H. D. (2001). Author-centered bibliometrics through CAMEOs: Characterizations automatically made and edited online. // Scientometric. 51:3, 607-637.
- Wolfram, D. (2016). The Power to Influence: An Informetric Analysis of the Works of Hope Olson. // Knowledge Organization 43:5, 331-337.

Enviado: 2017-07-31. Segunda versión: 2018-01-18.
Aceptado: 2018-03-23.

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 presentarse en un documento Word o RTF debidamente maquetados según la plantilla disponible en www.iversid.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. ").

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 puestas 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/multidoc/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

Cantisani, Mariana, 67	Martínez-Ávila, Daniel, 67
Dal'Evedove, Paula Regina, 35	Nakako, Natália, 67
Fernández-Molina, Juan-Carlos, 45	Pedraza Gracia, Manuel José, 23
Ferreira, Verônica de Sá, 55	Restrepo Fernández, María Camila, 91
Fujita, Mariângela Spotti Lopes, 35	Sales, Rodrigo de, 55
Gilchrist, Alan, 13	Salvador Oliván, José Antonio, 77
Guimarães, José Augusto Chaves, 45	Santos, João Carlos Gardini, 45
Jorente, María José Vicentini, 67	Tartarotti, Roberta Cristina Dal'Evedove, 35
Marco Cuenca, Gonzalo, 77	

Índice de materias en español

Subject index in Spanish

Análisis de contenido, 35, 45	Interdisciplinariedad, 67
Análisis de dominio, 67	Investigación, 35
Bacon, Francis, 55	Investigadores, 23
Bibliotecarios, 23	Lenguajes documentales, 91
Big data, 77	Ley de acceso a la información, 45
Brasil, 35	Libros antiguos, 23
Ciencia de la información, 13, 67	Olson, Hope A., 91
Clasificación de las artes, 55	Organización de la información, 45
Clasificación Decimal Dewey, 91	Organización del conocimiento, 55, 91
Clasificación documental, 91	Prospectiva, 13
CMBD, 77	RAE-CMBD, 77
Conjunto Mínimo Básico de Datos, 77	RDA, 23
Datos en salud, 77	Recuperación de la información, 13
Diseño de la información, 67	Registro de Actividad de Atención Sanitaria Especializada, 77
Estado de la cuestión, 35	SISNS, 77
Gestión de la información sanitaria, 77	Sistema de Información Sanitaria del SNS, 77
Harris, William Torrey, 55	Sistemas de Información hospitalaria, 77
Hegel, G. W. F., 55	Tendencias, 13
Identificación, 23	Teoría de la clasificación, 55
Información, 45	Valoración, 23
Información jurídica, 45	

Índice de materias en inglés

Subject index in English

Access to information act, 45	Classification theory, 55
Arts classification, 55	Content analysis, 45
Bacon, Francis, 55	Decimal Dewey Classification, 91
Big Data, 77	Domain analysis, 67
Brazil, 35	Harris, William Torrey, 55

- Health Data, 77
- Health Information management, 77
- Health Information System of the NHS, 77
- Hegel, G. W. F., 55
- History, 13
- Hospital Information System, 77
- Identification, 23
- Indexing language, 91
- Information, 45
- Information design, 67
- Information organization, 45
- Information retrieval, 13
- Information science, 13, 67
- Interdisciplinarity, 67
- Knowledge organization, 55, 67
- Knowledge Organization, 91
- Legal information, 45
- Librarians, 23
- Library classification, 91
- MBDS, 77
- Minimum Basic Data Set, 77
- Olson, Hope A., 91
- Prospective, 13
- RAE-CMBD, 77
- Rare books, 23
- RDA, 23
- Research, 35
- Researchers, 23
- SISNS, 77
- Specialized Health Care Activity Register, 77
- State of the art, 35
- Subject analysis, 35
- Trends, 13
- Valuation, 23