

# Information: a question of quality?

---

H. Peter OHLY

ISKO, Prinzenstr. 179, D- 53175 Bonn  
president@isko.org

## Resumen

En este artículo se describe la naturaleza de la información y los criterios para evaluar su calidad. El concepto de calidad en la Internet y en los portales resulta difuso, pues muchos proveedores y usuarios interactúan, multiplicando el triángulo cognitivo. Los estándares de descripción, los métodos bibliométricos y la minería de datos proporcionan cierta ayuda. Por otro lado, los usuarios esperan sistemas transparentes y personalizados, cuyas ventajas y desventajas se discuten. La solución radica en respetar la diferente complejidad de los distintos tipos de usuarios, pero la integración de los diferentes enfoques de la calidad sigue siendo una tarea pendiente y demandante.

**Palabras clave:** Información. Concepto. Calidad. Internet. Complejidad.

## Abstract

This paper describes the nature of information and its quality criteria. But especially in Internet and portals quality is rather diffuse as many suppliers and users come together, thus multiplying the cognitive triangle. Some help is given by description standards, bibliometric methods or information mining. On the other hand, the user expects transparent and customized systems. Some advantages and disadvantages of each approach are discussed. Some solution can be given by respecting the different complexities of user types, whereas a combination of all quality approaches is still a demanding task.

**Keywords:** Information. Concept. Quality. Internet. Complexity.

## 1. What are we speaking about?

There are different views of information. Some use this term like as interpretable coded data, others use it in the sense of integrated, justified knowledge. Here I want to use it in an in between sense: interpreted data in a certain context, but not necessary integrated in a whole universe of scientific knowledge. It can be a fact, a causal relationship or merely a name with a special meaning for the information seeker in an applied context. Gernot Wersig (1973) speaks of the "reduction of uncertainty," i.e. compared to a previous state of indecision, the user may experience a state of more than random deciding. Rainer Kuhlen (1995) calls information "knowledge in action" and expresses thus that a chunk of knowledge is used or reused in an application context and might finally enrich the meaning of the whole knowledge universe. Information is knowledge which is available at the right time and right place for problem-solving.

## Implications of information

Information in the proper sense can only be spoken of when the data actually provided for the information users has a current, specific value and reduces my nescience (and preferably also that of others). Otherwise there will be no obvious advantage of knowing it or even a disadvantage if it proves to be false or known already by every competing player. Prerequisites of Information have to be then: empirical or scientific evidence (validity), relative stability of evidence irrespective of some time or place variation (reliability), advantage with respect to the known knowledge of my community of discourse (peculiarity, uniqueness). In total, Information must per se be high quality information; otherwise, it is not justified to consider it as information. The goal of any specialized information supply, including technically provided information must be to ensure this. Information societies must also be qualitatively secured, which entails a demand for protection and an increase in information quality (Wiethaus, 2001). In 2005 the government of Germany and its Länder founded an excellence initiative to forward the sustainable strengthening of science in Germany and increasing its international visibility and competitiveness (Leibfried, 2010).

## 2. What is quality?

Eppler (2006) writes in his introduction to the book *Managing information quality*:

Information quality is a term that is vague and general, yet promising and pertinent: Amidst the increasing quantity of information available, the quality of information becomes a crucial factor for the effectiveness of organizations and individuals.

Quality, according to Wiethaus (2001) is a relative property which is affected by material and sociological circumstances. How it is determined depends upon, among other things, the development of technological standards; which, of course, vary over time.

DIN ISO 8402 standard defined quality as “the totality of qualities and characteristics of a product or service, which in turn affects their ability to meet fixed or assumed anticipated needs” (Rittberger, 2004). Here quality is seen as sum of relative stable and clear properties that have influence on the applicability of the good, e.g the information. In connection with information services quality may be seen as a distinguishing characteristic (such as scientific, practical), but also as a value scale (cf. Enderle, 2001). The degree of realistic anticipation of influencing factors denotes the quality and should be more or less easy to calculate. But it is dependent on the needs of users, since quality can only result through the participation in the process of information transmission (see Hobohm, 1998). The substituting DIN ISO 9000 (version 2005) defines quality as “degree to which a set of inherent characteristics fulfills requirements”. Though quality is also dependent on characteristics of the product itself more emphasis is given to its nature to be really useful in application situation. Especially with respect to the Web and its multitudinous resources and data types quality is rather diffuse as many suppliers and users come together. E.g. main reasons to use scholarly research databases are: credible content, in-depth information or instructors’ expectations (JISC, 2008). And a study of OCLC finds out that users and librarians have quite different search interests (2008).

We have to take into account then multiple triangles who are constituted by (many) objects (information and its providers), many subjects (users) and context (functionality). The information as object should have a meaningful content. Implied are also questions of the comprehensible presentation of it, ease of use, and confidentiality of the source. The information with respect to the using subject comprises characteristics of this person or his group and even his cultural background. As context determines the kind of

task it implies whether information in a given context reveals as quality.

Moreover I would also stress out the non obvious effect of the application of certain information. Should it help to make profit in monetary terms or as reputation? Or should it help to get new fruitful contacts, to promote a career, to acquire research funds, or to work out an expertise? Expectations like these (the kind of situational quality transfer) are rather implicit and difficult to anticipate. Additionally to the different aspects of information quality the questions of determining the exact quantity of quality remain, such as scaling the quality, how to measure it, and how to cope with dynamic changes.

## 3. Time perspective of information quality

If we take into account time orientations and their contribution to quality we can isolate these with consideration of past, present and future. Evaluations that consider the source of information are caused by the past. Does the information come from a verified corpus? Has the offering source a good reputation? Have other users had good experience with the corpus or the source? Was the method to describe the information rule based and approved? The substantive, authoritative component (manufacturer-process orientated reference) should be represented by absolute criteria, as complete, trustworthy, reliable, balanced, and well represented. While traditional information agencies accrue confidence via their specialization and institutionalization, on the World Wide Web, where there are many technically, highly capable information providers, the quality of the information supplied is in question. Thus, since the existence of the Web attempts have been undertaken to give (only) a filtered offer to the respective users. Search engines already filter a fraction of the sites existing in the Internet using formal criteria before further indexing (visible Web vs. invisible Web). In particular, search engines like Yahoo or Lycos also took advantage of user proposals and editor evaluations. Automatic procedures, like evaluation of word frequencies or number of linking pages have been the usual automatic procedures to weight the universe of Web contents since a long time.

If we are looking at the present situation of retrieving information, we have to judge about the possibilities to discriminate the information for a certain need. At least the classical inverse relation of information recall to precision shows that quality is dependent from its viewpoint: either very selective or either very mean. Moreover the presentation, description, and context relation of

the given information become relevant (Tufte, 1990). Mainly these characteristics are considered as quality aspects in information retrieval. According to Huang et al. (1999), Mandl (2005) emphasizes the importance of context in the user assessment of value, thus being a subjective component (expectation oriented reference). It becomes clear that even with a substantive point of view, no absolute assessment can be made, because it would require a perfect overview of the scientific field and the website supply. Huang et al. also stress the intrinsic aspect of the accessibility of content in addition to the quality of the content. The structuring of the information offered determines the possibility of navigation and finding the desired information. Another important aspect is topicality.

Taking into account the future some questions of usability and implicit expectations, like mentioned before, are of interest. Even questions of reusability for similar problems or a greater community arise. Can this information, or its codification, be easily stored, later retrieved, or connected with other information? This the question of sustainability seen from the side of the user, his group or even the society.

It is important that information is also secured generally and for the longer term (at least during the data availability). On the other hand the user only receives an information advantage through news not already generally well-known. The current emphasis may depend on the action context of the user, which can be of a more politic-strategic or more systematic-sustainable nature. At the same time, the information must be presented in a way that conforms to expectations, such that the user can also process the received information adequately in his context (see Rittberger, 2004). As mentioned earlier, the user will also attach very personal expectations for utilization to the information received, which can be seen rather as side effects in the given context, and can hardly be guaranteed by the information supplier.

## 4. Some approaches

### 4.1. Information creation

One of the everyday life criteria to rate information is the trust in the information provider. Has he a good reputation, has he experience with this kind of information since a long time ago, have I myself or others (whom I trust as well) good experience with him? Is he powerful enough to create good information. Has it appropriate tools to select and handle information? One way a ensure this is a certification process

by authorities, as the user (especially as uninformed information seeker) might not have the background nor the means to ensure the credibility of an information provider. Established big information institutions have in so far a certain credit, as national information providers and their data bases, specialized journals, or well known hosts. Even universities will have a certain credit per se.

A more sophisticated but technical way is the coupling of information. If it would be page rank, which ensures that a source is well linked by other sources, or co-citation of cited and bibliographic coupling of citing papers, or connectivity of authors. It's a quality decision by quantity of links between stake holders in the field of consideration. Sitations are only a slight analogon to citations but with CiteSeer, Scopos and Google Scholar Open Internet published articles become accessible also to bibliometric considerations. Besides Webometrics with the analysis of web growth and hyperlinks, the investigation of user forums enables to take into account the structure and nodes of the Internet and its users.

Further evaluation criteria are the clear definition of the selection processes of the creator, clear shared standards to describe and evaluate the information, and comprehensible ways to present or adapt the information given. Besides subject and type descriptions as well as selection criteria expanded principles for user and supplier-oriented data enrichment have to be considered, like different versions, creator information, access conditions, release information. Based on Dublin Core metadata further foundations for information gateway quality insurance were laid in the DESIRE project as well as in the German Digital Library Project by describing information sources according to different dimensions, like scope, content, form, process, collection management policy (Koch, 2000). With such selection rules and quality criteria the providers of Internet information additionally gain trust by their clientele.

### 4.2. Information processing

As information is rarely bijective it is important to enable the information searcher, to go in more detail or to adapt information outcome to his needs. All kind of sophisticated retrieval techniques can be listed here, as known from the TREC studies (Voorhees, 2005). If we go further we come into the field of information mining, the non-trivial process of identifying valid, novel, potentially useful, and understandable patterns in heterogeneous information sources (Kruse and Borgelt, 2003).

The aim is to retrieve information that is unique enough to fall out of the mainstream of information and to be precise enough to give an clear answer. The problem is “to find a needle in a haystack”, or taking the aim of Data Mining, that refers to large data sets or databases: “...the nontrivial extraction of implicit, previously unknown, and potentially useful information from data” (Frawley et al., 1992). This procedure is a kind of scientific induction, such as generating facts and hypotheses. Aside from the application of statistic procedures to databases, still more aspects such as computer based decision support, machine learning procedures and visualization can come into consideration if we aim at Knowledge Discovery (Schmidt-Thieme, 2002). Examples are learning algorithms to find characteristic relations, like Iterative Dichotomiser 3 and Automatic Interaction Detector (Sonquist et al., 1973), or Kohonen maps, who find different segments in a mass of information.

Scientometric methods to refine and describe the potential information are as well appropriate as far as they enable a user to move around in a collection and to find out combinations that are preferred or even the opposite. Term clouds give a connected but comprehensive overview with individual occurrence information. They are sometimes considered as self-adapting approaches with democratic potentiality. The so called ‘strategic diagram’ is an explorative mean to arrange documents by density and centrality of their clusters, such that unconventional, isolated, main stream or imitational science outcome could be identified (Callo,n 1985). Bibliometric indicators, such as Impact Factor, Hirsch Index and others are based on trust in citation rates with group-dynamic elements. Their interpretation as a measure of scientific productivity, intellectual affinity or user acceptance is not obvious and must be discussed depending upon granularity against the background of the data generation.

Important especially with the Web are semantic or knowledge based techniques. By application of logical or ontological knowledge they make a combination of information from different sources such that the prior information is validated or supplemented. Text Mining as an extension of data mining refers to alphanumeric data. Question answering, information extraction, thesaurus production, content summarizing, categorizing and text grouping are some examples (Hearst, 1999). Visual diagrams are more difficult to produce by the non-numerical form of the data since quantitative characteristics can only be derived indirectly. Here also the objective is more than just optimal retrieval of extant (text) data,

rather, scattered context information has to be evaluated and combined. Compression and selection of information from the Worldwide Web contains principles of text and information mining. “Web Mining,” however, goes even further beyond that, since very many different data types occur, the database is enormously large and information on page linking and access can be included (Hearst, 1997; Ghani et al., 2000). Web mining can thereby operate under different criteria: analysis of the contents (of a web page or a search result), the structures and the general or problem-related usage (Schmidt-Thieme, 2002).

#### 4.3. User orientation

A principle that could have been listed also under of creator quality is how far users were incorporated to develop the information system or whether there is a permanent user feedback or user group. They ensure the influence of the target main user group but may not meet the needs of individual users. Here come to adaptive systems with user models. The user is explicitly asked or implicitly analysed what his idiosyncrasies are with the aim to offer him the information more in his style and to give him user tailored interpretation hints. Surely these systems are limited to only certain types of users. A newer approach is social software. Here information is created by different groups or individuals —often users themselves— and they are free to reference their information to existing material or to create their own universe. The flexibility and user orientation are obvious. On the other hand different semantics might hinder the transfer of information if not pollute the information outcome even more than rigidly build information systems. Probably unconventional items might be found more easy whereas a recall-precision optimized retrieval seems to be less accomplished.

### 5. Summary remarks

Quality is an inherent feature of information and information supply. Though there are ambitious definitions of quality it’s rather difficult to restrict quality to some few criteria. Characteristics belonging to the provider and his offer as well as retrieval sophistication and profit for the user are concerned. Certain solutions favor the one or the other aspect. Surely the best would be a combination of all. But some are contradictory: e.g. a user driven system might be difficult to be enhanced by semantic or knowledge based techniques; user friendly presentation and sophisticated retrieval and mining techniques seem to be opposite positions. Nevertheless there is

hope to get an intelligent combination if the advantage and disadvantage of all proposed solutions is taken into account. Currently information systems that operate on a sound information processing philosophy together with facilities to combine different well described aspects seem to be the second best solution. Here the development of standards and explanations is of special importance, since information can be judged by the (uninformed) customer only after receipt and never exhaustively (cf., Rösch and Weisbrod, 2004). In the early days of the Internet, the quantity of documents found was critical, but now with its rapid growth (e.g., Rüter, 2008) there is a demand for restriction to those documents that most precisely match a given question or information request. Not at least the kind of presenting the information, its internal structure, and its external relations is of importance to make the special quality of information available. Information design, exploratory presentation and intuitive access are part of the quality characteristics of individual information, aggregated overviews and gateway access. Quality is thus not static nor relative, but a question of adaptability to user groups, user needs, and user expectations.

## Referencias

- Head, Alison; Eisenberg, Michael (2009). How College Students Seek Information in the Digital Age, Washington: Project Information Literacy). [http://projectinfo.lit.org/pdfs/PIL\\_Fall2009\\_Year1Report\\_12\\_2009.pdf](http://projectinfo.lit.org/pdfs/PIL_Fall2009_Year1Report_12_2009.pdf).
- Bargheer, Margo (2002). Qualitätskriterien und Evaluierungswege für wissenschaftliche Internetressourcen. Report zum DFG-Projekt. „Datenbankbasierte Clearinghouses im Kontext digitaler Bibliotheken. Niedersächsische Staats- und Universitätsbibliothek Göttingen.
- Callon, M. (1985). Les cartes stratégiques de la recherche: la méthode des mots-associés. // INSERM-Informations. (Décembre).
- Eppler, Martin J. (2006). Managing information quality. New York: Springer. ISBN: 3-540-23571-3.
- Ghani, Rayid; Jones, Rosie; Mladenic, Dunja; Nigam, Kamal; Slattery, Sean (2000). Data mining on symbolic knowledge extracted from the Web. // Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining (KDD-2000) Workshop on Text Mining, Boston, MA, August 2000. 29-36. <http://www.accenture.com/NR/rdonlyres/24FF352B-C886-4D2C-86DA-9DF7795F6DD5/0/7.pdf>.
- Hearst, Marti (1997). Distinguishing between Web data mining and information access: Position statement for Web Data Mining KDD 1997. <http://www.sims.berkeley.edu/~hearst/talks/data-mining-panel/>.
- Hearst, Marti (1999). Untangling Text Data Mining. // Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 1999. 20-26. <http://www.sims.berkeley.edu/~hearst/papers/acl99/acl99-tdm.html>.
- Hearst, Marti (2003). What Is Text Mining?. <http://people.ischool.berkeley.edu/~hearst/text-mining.html>.
- Huang, V.; Lee, Yang W.; Wang, Richard Y. (1999). Quality information and knowledge. Prentice Hall: Englewood Cliffs, NJ.
- JISC (2008). Information behaviour of the researcher of the future. London: JISC. <http://www.ucl.ac.uk/infostudies/research/ciber/downloads/ggexecutive.pdf>.
- Kahn, Beverly K.; Strong, Diane M. (1998). Product and Service Performance Model for Information Quality: An Update. // IQ. (1998) 102-115.
- Koch, Traugott (2000). Quality-controlled subject gateways: definitions, typology, empirical overview. // Online Information Review. 24: 1 (2000) 24-34.
- Kruse, Rudolf; Borgelt, Christian (2003). Information mining. // Int. Journal of Approximate Reasoning. 32:2-3 (2003) 63-65. <http://fuzzy.cs.uni-magdeburg.de/~borgelt/papers/ijar.pdf>.
- Kuhlen, Rainer (1995). Informationsmarkt: Chancen und Risiken der Kommerzialisierung von Wissen. Konstanz: Universitätsverlag Konstanz. <http://www.inf-wiss.uni-konstanz.de/FG/IV/imk/imk.html>.
- Leibfried, Stephan (2010). Die Exzellenzinitiative. Bilanz und Perspektiven, Frankfurt a/M: Campus.
- Mandl, Thomas (2005). The quest for the best pages on the web. // Information Services & Use. 25:2 (2005) 69-76.
- Nohr, Holger (2001). Management der Informationsqualität. // Working Papers Knowledge Management, Nr. 3/2001.
- OCLC (2008). Online Catalogs: What Users and Librarians Want. [Dublin, Ohio]: OCLC. <http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf>.
- Rittberger, Marc. (2004). Vertrauen und Qualität in Informationsdienste. Wo finde ich Vertrauen im Information Quality Framework? // Hammwöhner, R., Rittberger, M. und Semar, W. (eds.). Wissen in Aktion: Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft: Festschrift für Rainer Kuhlen. Konstanz: UVK Verlagsgesellschaft. 153-165.
- Rösch, Hermann; Weisbrod, Dirk (2004). Linklisten, Subject Gateways, Virtuelle Fachbibliotheken, Bibliotheks- und Wissenschaftsportale: Typologischer Überblick und Definitionsvorschlag. // BIT-online. 7:3, 177-188. <http://www.b-i-t-online.de/archiv/2004-03/fach1.htm>.
- Rüter, Martina (2008). Qualität von Webseiten. [http://tools-internetservices.suite101.de/article.cfm/qualitaet\\_von\\_webseiten](http://tools-internetservices.suite101.de/article.cfm/qualitaet_von_webseiten).
- Schmidt-Thieme, Lars (2002). Webmining. Vorlesungsunterlagen vom 15.10.2002. <http://www.informatik.uni-freiburg.de/cgmn/lehre/wm-02w/webmining-1.pdf>.
- Sonquist, John A.; Baker, Elizabeth L.; Morgan, James N. (1973). Searching for Structure. Ann Arbor: Univ. o. Michigan Press.
- Tufte, Edward R. (1990). Envisioning Information. Nuneaton: Graphics Press.
- Voorhees, Ellen M.; Harman, Donna K. (2005). TREC Experiment and Evaluation in Information Retrieval. Sample chapter. [http://mitpress.mit.edu/books/chapters/0262220\\_733chap1.pdf](http://mitpress.mit.edu/books/chapters/0262220_733chap1.pdf).
- Wersig, Gernot (1973). Informationssoziologie: Hinweise zu einem informationswissenschaftlichen Teilbereich. Frankfurt (am Main): Fischer-Athenäum-Taschenbücher.
- Wiethaus, Jörg (2001). Entwicklung und Nutzung eines Information Quality Rating-Tools am Beispiel von Content Management-Systemen. Konstanz. Master Thesis Univ. <http://www.ub.uni-konstanz.de/kops/volltexte/2002/758/pdf/Wiethaus.pdf>.

