

Theoretical difference between impact factor and influence factor

Đilda Pečarić, Miroslav Tudman

Faculty of Humanities and Social Sciences

Department of Information Sciences

The University of Zagreb

Ivana Lučića 3, 10000 Zagreb, Croatia

dpecaric@ffzg.hr; mtudman@ffzg.hr

Abstract. *Bibliometric constructions of "knowledge maps" and "cognitive structures of science" do not differentiate between impact and influence factors. The difference can be constructed according to different meaning and interpretation of the terms reference and citation. Reference is "acknowledgment which one author gives to another", whereas citation is "acknowledgment which one document receives from another". Development of Information Science according to period and subject area is analyzed on the corpus of citation literature retrieved from doctoral dissertations in Information Science from 1978 to 2007 at Croatian universities. The research aim is to indicate the difference between document impact factor and author's influence factor (i.e. reference ability to produce effects on actions, behavior, and opinions of authors of doctoral theses). The influence factor serves to distinguish the key role of cited authors in time and according to the duration of the influence (the average age for cited papers of dominant authors in different periods is between eight and ten years). The difference between linear and interactive communication seems vital for the interpretation of cited half-life, i.e. the attitude of one science community towards used information resources and cognitive heritage. The analyzed corpus of 22,210 citations can be divided into three communication phases according to influence factor criteria: in the phase of dialogue and interactive communication 25% of bibliographic units are cited in the first four years; in the second phase another 25% of units are cited from the fifth to the ninth year;*

after ten years, in the dominant linear communication phase, approximately 30% of units are cited.

1. Introduction

Impact factor was for the first time mentioned by E. Garfield in 1955 in journal 'Science'. (E. Garfield, 2006). At the beginning of 1960's E. Garfield and Irving H. Sher defined „journal impact factor“ as a criterion for journals selection for **Science Citation Index**, which *Institute for Scientific Information* (ISI) in Philadelphia began publishing in 1961. Since then, the impact factor has become one of the basic, most prevailing and most used criterion in the Information Science for the evaluation of scientific journals and scientific papers.

In 1975 *Institute for Scientific Information* (ISI) started publishing **Journal Citation Reports-a** (JCR), which together with other ISI editions enabled the development of series of statistical analyses based on „journal citation frequency“, „impact factor“ and other statistical indicators. In the last few decades JCR has become „the ideal tool“ for analysis and comparison of scientific journals, and also for the evaluation of scientific production from individual to institutional and national level. Moreover, in many countries scientists have to report about impact factor of their work in order to prove their scientific production efficiency.

It is the fact that „impact factor“ has tremendous but controversial impact on the study and evaluation of published scientific production. Recently, the *Joint Committee on Quantitative Assessment of Research*

(2008) castigates uncritical implementation of impact factor in statistical analysis:

The validity of statistics such as the impact factor and h-index is neither well understood nor well studied. The connection of these statistics with research quality is sometimes established on the basis of "experience." The justification for relying on them is that they are "readily available." The few studies of these statistics that were done focused narrowly on showing a correlation with some other measure of quality rather than on determining how one can best derive useful information from citation data...

Using the impact factor alone to judge a journal is like using weight alone to judge a person's health¹.

Since Joint Committee report is composed by prominent scientific association (International Mathematical Union /IMU/, as well as International Council of Industrial and Applied Mathematics /ICIAM/ and Institute of Mathematical Statistics /IMS/), it is obvious that the usage of "impact factor" is either not unambiguous or criterion and his usage is not theoretically sufficiently explained.

2. Problem

In scientific and subject literature in Croatian language the problem already exists with the translation of the term *impact factor*. It seems that the usage of the term *impact factor* (faktor odjeka) prevails, as M. Jokić uses it in one of the basic bibliometric handbooks in Croatian language. However, the term *influence factor* (faktor utjecaja)² exists as a synonym. This difference, which arose with the translation of the term *impact factor* and the usage of the terms *impact factor* and *influence factor*, is not accidental. It implicates different usage and understanding of the term *impact factor*, and therefore can be methodological instruction for the interpretation of described controversial usage of terms.

In fact, critics of impact factor rightfully claim that this indicator can not be used for qualitative estimation. Most of bibliometric handbooks agree with this, because definition of impact factor is „ratio between citations and recent citable items published in the same period³".

¹ See: Joint Committee on Quantitative Assessment of Research (June 12, 2008).

² "FAKTOR UTJECAJA (impact factor) is a measure of the frequency with which the "average article" in a journal has been cited in a particular year of period". Jelka Petrak "Citati i njihova analiza" (2003.). In literature in English language the usage of the term "influence factor" also exists, but it is inconsistent, and is primarily used to describe trends more then to be specifically defined measure. This term usually does not exist in Information science dictionary (see: Joan M. Reitz: ODLIS — Online Dictionary for Library and Information Science).

³ See M. Jokić, p. 88. with respect to E. Garfieldu "...impact factor is determined by counting citations in the current year's publications to papers published in the previous two years and divided by the number of papers published in the same period" (E. Garfield, Scientist. 1998;12:10-12.)

Recent impact factor definitions are usually generalization, because initial definitions was referring to journal impact factor (JIF): journal impact factor (C) is calculated by dividing the number of current year citations (A) to the source items (B) published in that journal during the previous two years. Therefore, journal impact factor is $C = A/B$.

Gradually impact factor is developed like the measure of journal and author impact. Yet, with time, impact factor has been used more like the measure of journal and author influence. Hence, more as qualitative and less as quantitative indicator of participation in scientific communication. Difference in the usage of the term *impact factor* reveals this connotation: the usage of the term *impact factor* points to quantitative and the term *influence factor* to qualitative value of this indicator. However, this is just semantic indication of connotative difference, which yet has to be theoretically explained.

The research of Đ. Pečarić (2009.) indicates that realistic empirical difference exists between impact factor and influence factor. Citing intensity according to age (in analyzed doctoral dissertation in Information Science at Croatian universities) is the biggest in first four to five years (citing reaches maximum between fourth to fifth year) and then it gradually decreases to cited half-life (which is between ninth and tenth year (Table 1)).

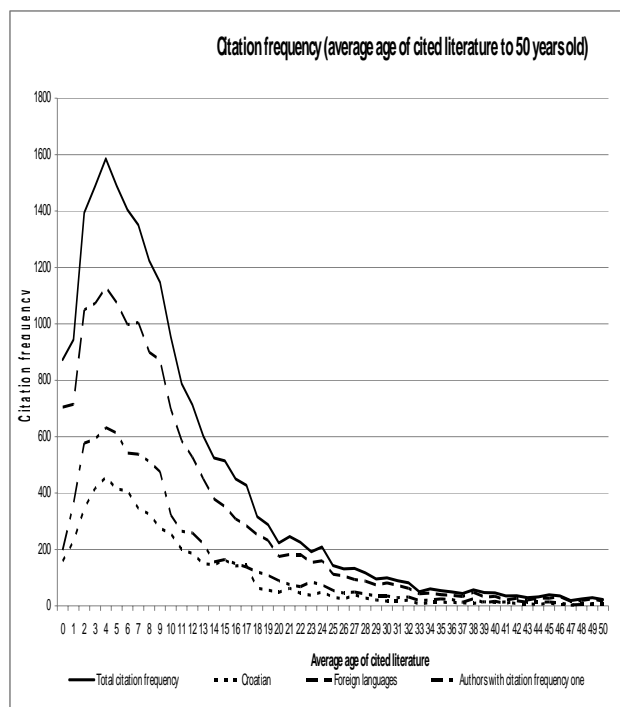


Table 1.

Nevertheless, the most cited authors whose papers obviously have biggest "impact", respectively "influence" are older than citing half-life. According to E. Garfield's thesis (E. Garfield, 2006) "the rankings based on 1-, 7-, or 15-year impact factors do not differ significantly" is hardly sustainable. In the

following table fifty most cited authors and the average age of their most cited articles are presented. It can be seen that the frequency of cited authors significantly differs in respect with the time flow (Table 2).

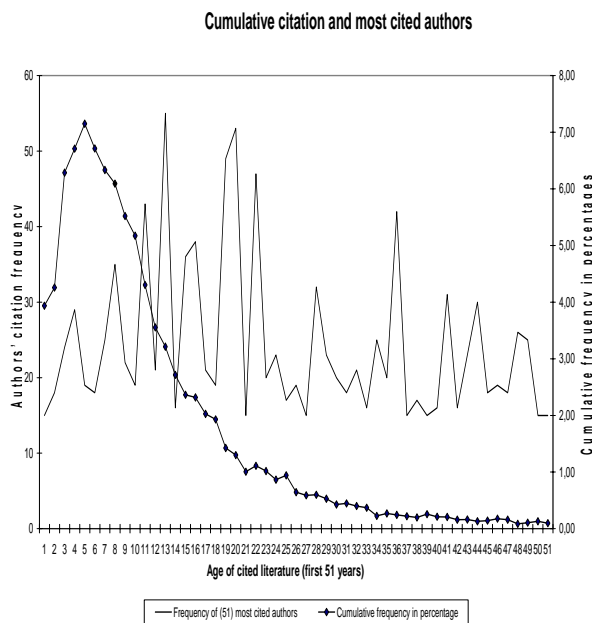


Table 2.

It is evident, that overall number of citations rapidly decreased after cited half-life while the citations of a small number of authors increased. The second phenomenon can be explained by influence factor, i.e. with difference between influence and impact factor.

3. About "unimportant" differences and methodological (in)consistency

What is a citation and what is a reference? Are they the same or different entities? Both, citation and reference are using bibliographic categories, such as author, title, publisher, etc., to describe document. It can be easily seen that citation and reference are the same bibliographic entity, because citation is defined as synonym for bibliographic reference:

The citation for an article includes the author, title of the article, title of the periodical, volume, pages, and date. ... (www.noblenet.org/merrimack/guides/infolit_glossary.htm)

... (synonym for bibliographic reference) identifying information about a publication, used in catalogs and indexes as well as in lists of "literature cited" or "references" in scholarly publications. ... (library.hunter.cuny.edu/lyannott/thesis_guide/libraryterms.html)

Hence, reference is defined similarly as citation:

"Reference - A 'citation' referring to a document or passage" (G. Wersig, U. Neveling, str. 95).

Although both are formed from the same bibliographic data and because of that both can be and are the same, the important difference between citation and reference lies in the manner of their usage: reference is "acknowledgment which one author gives to another", whereas citation is "acknowledgment which one document receives from another" (J. Petrak, 2003.).

Therefore, it is about how we read or interpret the information that is contained in them. Reference is establishing relationship of "acknowledgment" between authors. Citation is establishing relationship of "acknowledgment" between documents. Established relations between documents are presented as statistical information about possible relations between documents, relations that can be, but usually are not, the description of realistic scientific communication's relationships. References describe "acknowledgment" that are established and exist between authors; therefore they are realistic communication messages even if communication is linear: one author gives acknowledgement to another with whom he does not have direct communication, but he received and accepted his message through the document he is referring to.

Described difference between reference and citation reminds us of familiar communication and social fact that the meaning and sense of the message depend on who is sending message to whom: person A to person B or vice-versa. Information science does not take into account this social and communication fact. That is why Information science reduced the message on information: based on the assumption that information is the same regardless of the direction in communication process.

Thereby information science agreed to annul difference between message and information. This approach prevents understanding of the message influence, more precise the influence of cited author, because influence is 'strength, power to act upon someone or something so as to make them change their action, behavior or beliefs' (Anić, 1980., p. 1270). Accordingly, power that sender's message has on the recipient, or "acknowledgment" that one author gives to another, is influence factor. "Acknowledgment" that one document receives from another is impact factor.

Impact factor can be just quantitative indicator, while influence factor can be qualitative indicator of (scientific) communication processes. Having that in mind, we have to understand the difference between message and information, as Rafael Capurro explained in his papers (2006): "In my view information science should take the phenomenon of message as its core perspective..... Information science is ... a science dealing with the phenomenon of messages as part of the phenomenon of communication i.e. including the 'meaning offer', the process of selection ('information') and the process of interpretation ('understanding)".

This approach is closer to the comprehension of Information science as a science about organization and exchange of knowledge; because "knowledge is offered like message" that has meaning, which becomes the information in the process of the selection of meaning. That point of view is opposite to 'information becoming knowledge through a process of linear accumulation'. The message and information are connected, but they are not identical concepts, as Capurro (2006) claims: „the message depends on the sender... which is not the case with information: we are receiving the messages, but we are looking for the information”⁴.

If we wish to study influence factor, then that is the impact of the message on the recipient, respectively "acknowledgment" that one author gives to another. From methodological point of view that means to consistently investigate and respect the difference between the message and information, between qualitative and quantitative indicators.

4. Influence factor and scientific paradigm

Influence factor cannot be a quantitative indicator, but with the help of the quantitative data processes and events can be recognized, which then can be the base for qualitative estimations. On the example of bibliometric analysis of doctoral dissertations in Information Science at Croatian universities (Đ. Pečarić, 2009.) we can discuss the research of influence factor. Research embraced 134 doctoral dissertations in Information Science from 1978 to 2007 at Croatian universities. The basic data about dissertation, citation and authors are the following.

4.1 Data about dissertation

In the period from 1978 to 2007 at Croatian universities 134 doctoral dissertations were done in seven different disciplines in Information science: 20 in librarianship, 21 in information science, 53 in information systems, 22 in communicology, 9 in museology, 8 in archivistics and documentation, 1 in lexicography. The majority of doctoral dissertations were made at the Faculty of organization and informatics in Varaždin (FOI) – 69, followed by the

⁴ Capurro offers following definitions for basic terms: For the purpose of the study let us use the following ad-hoc definitions:

"Data are sets of symbols that represent empirical perceptions (e.g., an image of a chair, the voice of a child pronouncing the word "chair").

"Information" is a set of symbols that represent empirical knowledge (e.g., "The panel is composed of 55 members.").

"Knowledge" is a set of symbols that represent thoughts, which the individual justifiably believes to be true (e.g., "2+2=4", "Cogito ergo sum", "E=MC²").

"Message" is a set of symbols that represent any meaningful content (e.g., "I have 10 fingers", "I have 15 fingers", an image of a chair, the phrase "The White House", the image of the White House, a recording of Beethoven's Piano Concerto n. 5, the musical notes of Beethoven's Piano Concerto n. 5). Note that "message" is defined here in its broadest sense (i.e., as meaningful content) rather than in the narrow sense of a sender-recipient phenomenon.

Faculty of Humanities and Social Sciences in Zagreb – 49 doctoral dissertations. According to the periods of production: 21 doctoral dissertations were made until 1989; 62 doctoral dissertations from 1990 to 1999; 51 doctoral dissertations from 2000 to 2007.

Altogether there were 55 mentors for 134 PhD candidates, 26 mentors were mentors only once, 9 mentors twice, 7 mentors three times, and also another 7 mentors four times. One mentor had five PhD candidates, one had seven, 3 mentors had six candidates. And finally, one mentor was mentor eleven times.

The youngest PhD candidate was 28 years old, and the oldest was 75 years old. By the age of thirty 3 PhD candidates made their doctoral dissertations, by the age of forty 48 candidates made doctoral dissertation, by the age of fifty 53 candidates, and after their sixties 28 PhD candidates (we do not have data for 2 PhD candidates).

4.2 Data about cited documents

In 134 doctoral dissertations the total of cited bibliographic units is 22,210 (thereof 4,443 units are without authors). From 17,767 cited units with authors 590 units are self-citations (3,3%).

If 22,210 cited documents are 100% then: 22.76% of cited documents are without author; 37.35% of documents are cited just once; 39.99% documents are cited more than once (8,881).

Cited half-life from all documents is 7.5 years. Certain differences exist according to the type of cited documents: for monographs' cited half-life is 9.1 years, for journals it is 7.2 years, and for semi-publications it is 9.3 years. The difference exists also among disciplines. The shortest cited half-life is in information systems 5.9 years, and the longest in museology 12.6 years. For information science it is 7.1 years, for librarianship it is 7.8 years, for communicology it is 8.5 years, and for archivistics and documentation it is 8.6 years.

Co-citation analysis is used for the retrieval of citation clusters. Frequency of citation pairs is calculated. Total number of pairs, i.e. total number of co-citation is 108,228 and their distribution is as follows: 2 pairs with frequency of 9; 2 pairs with frequency 8; 11 pairs with frequency 7; 20 pairs with frequency 6; 61 pairs with frequency 5; 139 pairs with frequency 4; 771 pairs with frequency 3; 8,578 pairs with frequency 2; 98,704 pairs with frequency 1.

4.3 Data about cited authors

From the overall number of citations (22,210), there are 17,178 citations with authors; that is, total number of cited authors is 10,683. 8,296 (77.65%) authors are cited only once, and 2,387 (22.34%) authors are cited more than once. Those 2,387 authors that are cited twice or more, cover 51.71% of citations. So, the rule that small number of authors is frequently cited is repeating: 451 authors that are cited 5 or more times cover 23.61% of citations; 118

authors that are cited 10 or more times cover 11.71% of citations. The first 29 of the most frequently cited authors cover 5% of citations, and the first 49 authors cover 7% of citations.

4.4 Seeking influence factor

The citation frequency according to ages is presented in table 1. What can be immediately noticed in this table, and is also confirmed in the analysis of other data in quoted research: the distribution curve is always the same or similar to the one shown in table 1, regardless of the frequency of citation, self-citation, citation according to different languages or citations that are quoted only once. This is also confirmed when the data are fragmented according to scientific discipline, as well as according to time periods or faculties on which doctoral dissertation are made.

The pattern is the following: if we know cited half-life ($t/2$), period in which 50% of documents are cited, then first 25% of documents are cited until half of cited half-life ($t/4$). In time period $t/4$ maximum frequency from overall number of cited documents is reached. Therefore, citation curve has log-normal distribution, with maximum in time period $t/4$.

51 most cited authors in Information Science

(from 1978 to 2007)

(citation frequency bigger than 14)

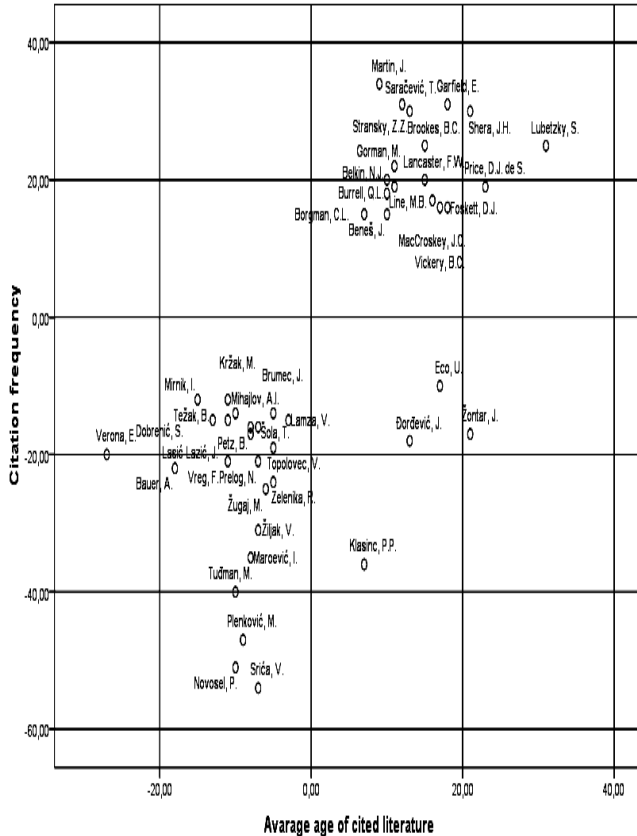


Table 3.

Second, perceived regularity is that in time period between $t/2$ and t , i.e. second time period of cited half-life, following 30% of documents are cited. After double cited half-life, the last 20% of documents are cited, documents which ages cannot be statistically predicted.

The most cited authors in Croatian and in foreign languages are presented in table 3, and their distributions according to frequency and ages of cited literature.

Table 3 is a good indicator about place and space of influence of the most cited authors. The most cited authors are not among the first 50% of cited documents, which are cited until cited half-life. The biggest number of cited authors is, most often, in time period between $t/2$ and t , which means among 30% of documents that are cited after cited half-life. That greatly nullifies enforced opinion that the most valuable authors are the most recent and the most frequent authors in the area of cited half-life. This is the area of the highest density of citation in short time period, the highest intensity and selection of the information, or in Capurro's terminology selection of "empirical knowledge".

According to its own logic, influence is greater when it lasts longer. That is the reason why it is not odd that ages of most cited authors are greater than cited half-life. This also implicates another assumption: influencing authors are not cited because of information, but because of the message they offer. In scientific communication the message of the dominant problem solving theory is the most important. In other words, members of one communication community are bound by mutual scientific paradigm, which is defined by influential scientists.

Considering the time range of almost 30 years in which doctoral dissertation were written, we can follow how some authors from relevant "empirical knowledge" have become influential representatives of scientific paradigm, and also how others have gradually lost their influence.

Table 3 presents authors cited in Croatian language apart from authors cited in other languages. This kind of graphical presentation aims to point out how influence factor is determined by communication processes: PhD candidates "reward" authors in Croatian and in other languages (mostly in English language). However, it would be incorrect to assume that mutual "reward" exists *between* most cited authors, because of insuperable language barriers⁵. The conclusion is that neither the analysis of co-citation nor impact factor is relevant for the analysis of influence factor (real relationships between participants of communication communities).

⁵ Very often they are reality, because it can be assumed that most of the authors communicate for example, in English do not know Croatian language and articles.

5. Instead of conclusion: influence factor and qualitative analysis

Influence factor is not qualitative indicator which can be statistically calculated. By impact factor we detect the cited area and time within which the influence of specific authors can be analyzed. That area is recognized as network of the most influential authors in dominant scientific paradigm.

In the case of doctoral dissertation analysis, influence factor of cited authors will be presented by communication models, i.e. mutual relation of the authors, not only based on statistical indicators but also on qualitative estimation. In concrete research numerical indicators which determine the dimensions of authors' influence factor are: cited authors frequency, number of dissertation in which authors are cited, number of disciplines in which authors are cited, average ages of cited literature, relationship between citing and mentorship, cited mentors, etc. Qualitative estimation of influence can provide analysis of cited papers. The aim of this paper was to determine the difference between influence and impact factor, and resolve how they can be researchable.

Literature:

- [1] Anić, Vladimir. **Rječnik hrvatskoga jezika**. Zagreb: Novi Liber, 1998.
- [2] Capurro, Rafael; Chaim Zins. **Knowldege Map of Information Science. Rafael Capurro's responses to Chaim Zins**. (2006)(URL www.capurro.de/10.4.2009.)
- [3] Garfield E. **Citation indexes to science: a new dimension in documentation through association of ideas**. *Science*. 1955;122:108-111. Available at: <http://garfield.library.upenn.edu/essays/v6p468y1983.pdf>. Accessed October 26, 2005.
- [4] Garfield E, Sher IH. **Genetics Citation Index**. Philadelphia, Pa: Institute for Scientific Information; July 1963. Available at: <http://www.garfield.library.upenn.edu/essays/v7p515y1984.pdf>. Accessibility verified November 29, 2005.
- [5] Garfield E. **Long-term vs. short-term journal impact: does it matter?** *Scientist*. 1998;12:10-12. Available at: [http://www.garfield.library.upenn.edu/commentaries/tsv12\(03\)p10y19980202.pdf](http://www.garfield.library.upenn.edu/commentaries/tsv12(03)p10y19980202.pdf). Accessed October 25, 2005.
- [6] Garfield E. **Long-term vs. short-term journal impact, II: cumulative impact factors**. *Scientist*. 1998;12:12-13. Available at: [http://www.garfield.library.upenn.edu/commentaries/tsv12\(14\)p12y19980706.pdf](http://www.garfield.library.upenn.edu/commentaries/tsv12(14)p12y19980706.pdf). Accessed October 25, 2005.
- [7] Garfield E, **The History and Meaning of the Journal Impact Factor**. *JAMA* Vol. 295 No. 1, 90-93, January 4, 2006.
- [8] Joint Committee on Quantitative Assessment of Research (June 12, 2008). **"Citation Statistics"** (PDF). International Mathematical Union. <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>.
- [9] Jokić, Maja. **Bibliometrijski aspekti vrednovanja znanstvenog rada**. Zagreb: Sveučilišna knjižara, 2005.
- [10] Pečarić, Đilda. **Razvoj informacijskih znanosti u Hrvatskoj. Bibliometrijska analiza doktorskih disertacija iz informacijskih znanosti 1978. - 2007**. Zagreb: doktorska disertacija, rukopis, 2009.
- [11] Reitz, Joan M.: **ODLIS — Online Dictionary for Library and Information Science** (URL <http://lu.com/odlis/about.cfm>, 10.4.2009.)
- [12] Petrak, Jelka. **Citati i njihova analiza**. Vidi: Hrvatsko informacijsko i dokumentacijsko društvo, 2003. <http://www.hidd.hr/articles/citati.php> (URL 10.4.2009).
- [13] Wersig, Gernot; Ulrich Neveling. **Terminology of documentation**. Paris: The Unesco Press, 1976.