Predecessors, Scholars and Researchers in Information Sciences.
Contribution to Methodology for Bibliometrics Analysis of Scientific Paradigms

Đilda Pečarić
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
dpecaric@ffzg.hr

Miroslav Tuđman
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
mtudman@ffzg.hr

Summary

Development of Information Science paradigm is researched on the corpus of most cited references retrieved from doctoral dissertations in Information Science (from 1978 to 2007). New approach for analysis of scientific paradigm by empirical display of dominant zones within scientific paradigm is proposed: empirical knowledge zone, conceptual knowledge zone and research front zone. Alterations of scientific paradigm are followed across three time periods by display of most cited authors in librarianship, information systems, communicology, archivistics and documentation, museology and information science. Besides the data about most cited authors, the data about most cited references according to periods and disciplines are shown. Analysis of most cited references resulted with discovery of the dominant research topics in particular periods. Based on changes in research topics it can be concluded: a) which research topics were interesting for Information Science researchers in Croatia, and b) changes within Information Science paradigm, by tracking changes of key authors which are cited during period of thirty years. Suggested methodology can serve as a model for tracking the development of scientific paradigm in other research disciplines as well.

Key words: Information Science, Scientific Paradigm, Communication Models, Knowledge Zones, Methodology, Bibliometrics Analysis

Introduction

It is possible to analyze the development of Information Science and the role of key authors and key publications in Information Science community by bibliometrics methods.

We start from the assumption that doctoral dissertations in Information Science are a good sample for the analysis of Information Science development in Croatia, because doctoral dissertations are original scientific publications which are using up to date world key literature.

Methods

We analyzed 134 doctoral dissertations in Information Science done on Universities in Croatia from 1987 to 2007. The doctoral dissertations were done on Croatian Universities that have postgraduate studies in Information Science, i.e. The Senate of the University of Zagreb / Znanstveno-nastavno vijeće Sveučilišta u Zagrebu (from 1978 to 1981), Zajednički studij informacijskih znanosti (from 1985 to 1987)/, Faculty of Organization and Informatics (from 1987), Faculty of Humanities and Social Sciences (from 1990)/ and The University of Zadar (from 2001).

The classification of doctoral dissertations according to disciplines is based on the classification of scientific disciplines and fields used by the Ministry of Science, Education and Sports of The Republic of Croatia. According to that classification Information Science is divided into following disciplines: Archivistics and Documentation, Librarianship, Communicology, Lexicography and Encyclopedics, Museology, Information Science and Information Systems (According to classification of Ministry of Science 'Information Systems and Information Science' are the same discipline, but for the purpose of our analysis we divided them into two disciplines, Information Systems and Information Science, in order to separate doctoral dissertations done on the Faculty of Organization and Informatics and Faculty of Humanities and Social Sciences). For the analysis of citation corpus of 22,210 bibliographic units in 134 doctoral dissertations we used cluster analysis. Clusters are formed according to the frequency of cited authors and titles. The obsolescence of literature was important for our analysis. Therefore we used usual criterion of citation "half-life" which is determined as period of time in which 50% of references are cited.

In previous papers we presented the criteria that can more precisely describe the development of the Information Science. We advocate that is possible to identify dominant fields of scientific influence inside scientific paradigm, i.e. empirical knowledge zone, conceptual knowledge zone and research knowledge zone (M. Tuđman, Đ. Pečarić, 2009.). Further analysis of relationships between authors' in research and in conceptual knowledge zones (Đ. Pečarić, 2009.) indicates that in spite of constantly changing position and role of authors, it is possible, with citation obsolescence criteria, to identify three different groups of authors: group of predecessors, group of scholars and group of researchers.

The development of Information Science in Croatia, ie. Information Science disciplines in the last thirty years will be analyzed by prepared methodology.

The Most Cited Authors in Information Science Disciplines

Tables 1 to 3 show the most cited authors in museology, information science and information systems[1]. Authors of papers written in different languages are not grouped in the same cluster. Why? We wanted to stress the fact that there exists a difference between citation and reference. Although both are formed from the same bibliographic data and 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.). Because of language barrier it is possible to assume and advocate inequality that exists between citation and reference. It is evident that authors of doctoral dissertation acknowledged the authors who published their papers in foreign languages. At the same time, the authors who are not familiar with the "small" languages can not respond in the same way. Because of that asymmetry of citation usage publications published in foreign languages are shown in the right top corner of the table, and publications published in Croatian language are shown in the left bottom corner of the table.
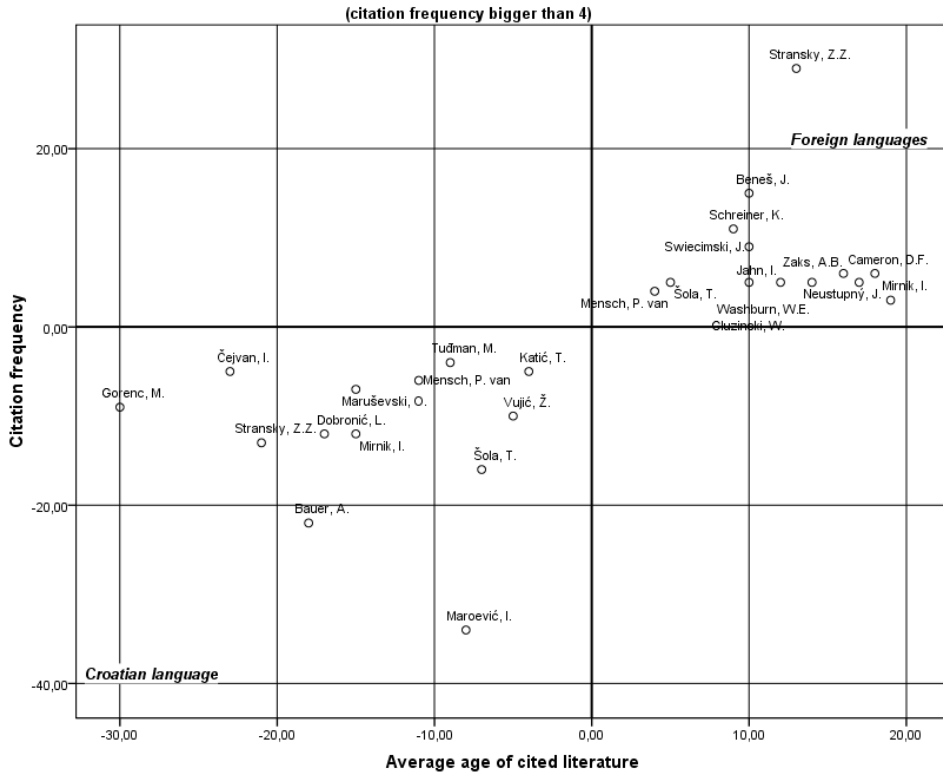
In order to be able to make conclusions about development of Information Science paradigm, it is important to evaluate the sample of the most cited authors in certain disciplines that are shown in tables 1 to 3.

Of the overall number (1279) of all cited authors in museology, 22 of the most cited authors make only 1.7%. However, these 22 authors hold 10.2% of citations from overall number of cited documents in museology. There are 972, or 76%, of authors that are cited only once in museology. But, in order to be precise, these percentages should be corrected, because the number of documents (both anonymous and those having an author) that are cited only once is 51.9%. Therefore, it is more precise to say that almost 1/5 of all multiple citations hold 1.7%, that is, 22 most cited authors.

In other two disciplines frequency of citations behave in a similar manner. In information science, first 32 authors or 1.8% authors (from 1770 most cited authors) hold 7.7% of citations. In information science there are even 80.8% of authors that are cited only once. However, since in this discipline a large number of documents without authors (16%) are cited, the overall number of all documents (with or without authors) cited only once is 62.9%. So, the conclusion is similar to previous one, i.e. a small number of authors (1.8%) holds 1/6 of all multiple citations.

---

[1] Because of the lack of space, in this paper, we are not able to show the most cited authors in all disciplines. In previous paper (M. Tuđman, Đ. Pečarić, 2009.) the most cited authors from librarianship and communicology are shown.

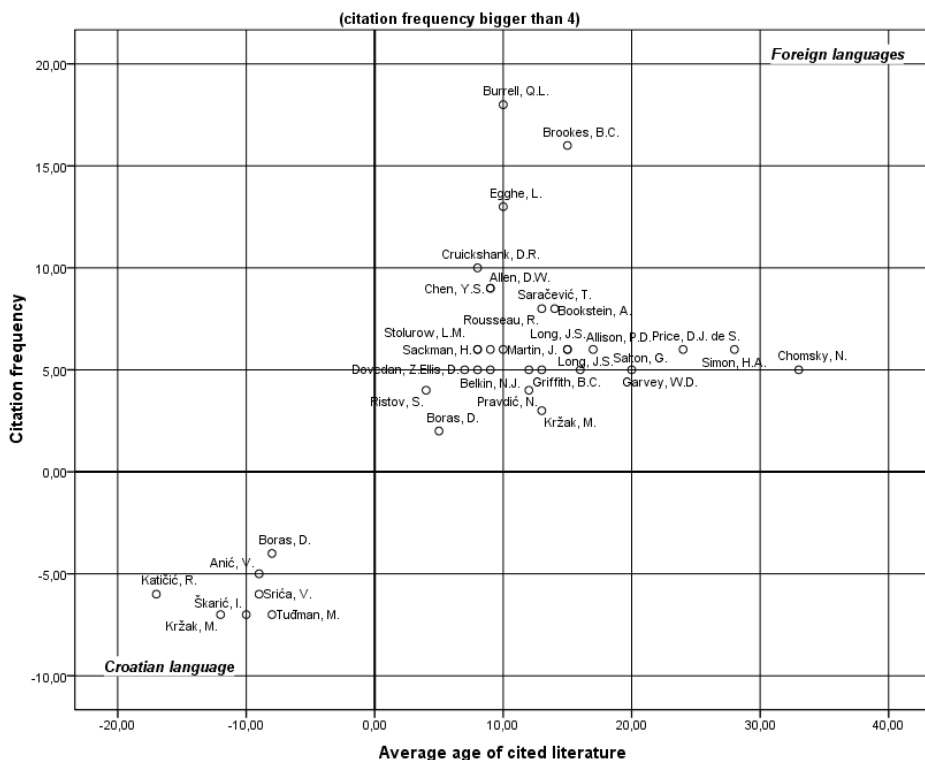Table 1: 27 most cited authors in Museology from 1988 to 2007

Although the number of most cited authors in information systems (table 3) is similar to the previous discipline, the differences are following. These 31 most cited authors make only 0.8% of 3662 authors cited in this discipline. In information systems 2981 authors or 81.4% are cited only once. Also in this discipline 11.6% of cited documents are without authors, so it is more realistic to accept that 61.1% documents are cited only once. But, in comparison with this information, 0.8% authors hold almost 1/6 of multiple citations.

From this data it can be concluded that a small number of authors (in our example between 1.6% and 1.8%) receives between 8% and 12% of all authorial citation. However, it is realistic to start from the fact that in these disciplines about 60% of cited documents are cited only once (regardless of the authorship status), so it can be concluded that 1.6% to 1.8% authors hold 1/5 or 1/6 of all multiple citation.

In three analyzed disciplines 90 authors hold 1/6 of all citation. However, it should be taken into account that out of 90 most cited authors in all three disciplines, 50% of authors is "mutual"; namely, 44 authors are cited in two or three disciplines.

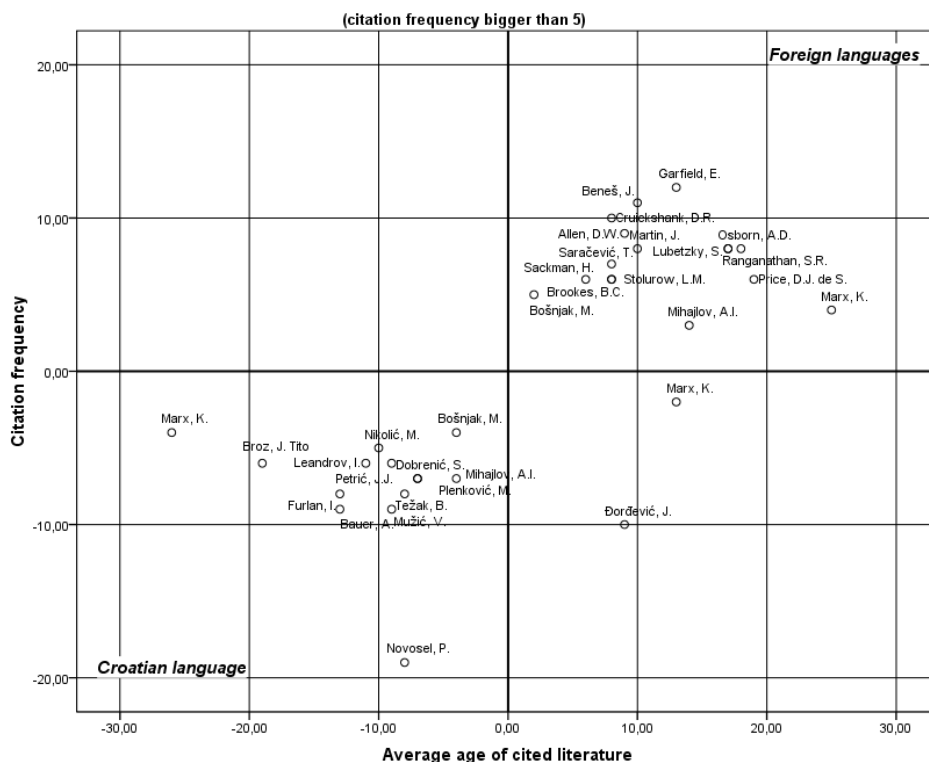Table 2: 32 most cited authors in Information Science from 1978 to 2007

It is also important to know how many cited authors in three disciplines (table 1 to 3) are cited in other Information Science disciplines. In museology only 4 out of 27 authors are cited in other disciplines. However, 26 out of 32 most cited authors in information science are cited in other disciplines, whereas 19 out of 31 most cited authors in information systems are cited in other disciplines.

We can raise the question in how many disciplines are present the most cited authors from museology, information science and information systems[2]? Only one author (M. Tuđman) is cited in all seven disciplines. Four authors (N. J. Belkin, G. Salton, T. Saračević, A. I. Mihajlov) are cited in five different disciplines. Seven authors are cited in four different disciplines (V. Anić, M. Kržak, D. de S. Price, V. Srića, B. Težak, S. Tkalac, M. Žugaj). Nine authors are cited in three, and 23 authors are cited in two different Information Science disciplines.

---

[2] In this analysis we use partition of Information science into following seven disciplines: archivistics and documentation, librarianship, communicology, lexicography, museology, information science and information systems.

Table 3: 31 most cited authors in Information Systems from 1980 to 2007

It is evident that joint core of authors who are often cited in several Information Science disciplines exist. Based on these data we can conclude that joint theoretical baseline in Information Science also exist.

Predecessors, Scholars and Researchers in Librarianship, Communicology and Information Systems

According to the criteria of cited literature obsolescence, and according to location in the cluster of cited authors, we recognize several groups of authors: predecessors, scholars and researchers (Đ. Pečarić, 2009.).

In the group of predecessors we can include authors that are continuously cited after double citation half-life, i.e. those that belong to the last 20% of citations cycle.

The group of scholars form the authors that are cited after the period of citation half-life and until the end of double period of citation half-life. Those two groups of authors (scholars and predecessors) are defining conceptual knowledge zone. According to T. Kuhn these two groups of authors are key representatives of dominant scientific theories.

Researchers are the most cited authors in the first half of citation half-life. They belong to the research front. Their publications are mostly cited immediately after publishing – and if they remain permanently present in scientific community, then during the time they become part of the dominant scientific paradigm.

Table 4: 28 most cited authors from 1978 to 1989

The group of authors that form predecessors in museology are both founders and key authors. According to obsolescence of cited literature the group of predecessors in museology is: R. Horvat, M. Gorenc, I. Čejvan, Z. Z. Strànský, I. Mirnik, A. Bauer[3].

The group of predecessors in information science form: N. Chomsky, H. A. Simon, D. J. de S. Price, W. D. Garvey, K. Katičić, P. D. Allison, G. Salton, J. S. Long, B. C. Brookes.

---

[3] Average obsolescence time in museology is unrealistically high (12.6 years) because it was not possible to discern the citation of documentation's source material from the citation of relevant literature. That is why the authors whose cited literature is around 20 years old are included in this group, and not only those whose cited literature is more than 24 years old.

The group of predecessors in information systems forms: W. D. Garvey, G. Salton, S. Dobrenić, D. Radošević, A. I. Mihajlov, A. V. Aho.
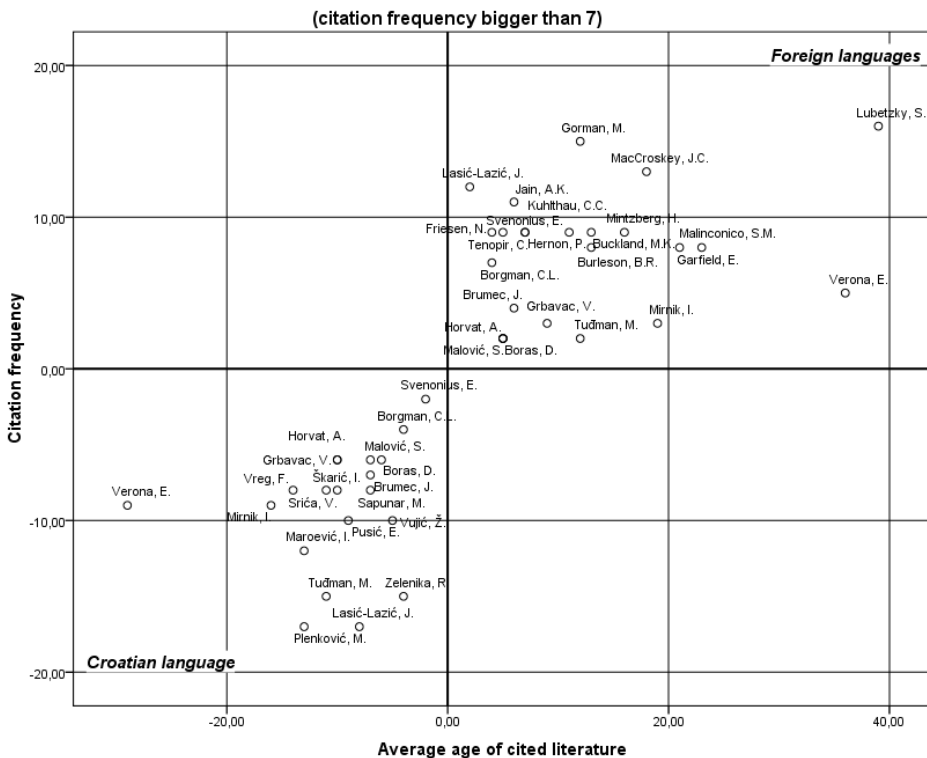
According to formal criteria of most cited authors and literature obsolescence, the group of scholars in museology form: D. F. Cameron, L. Dobronić, J. Neustupný, O. Maruševski, W.E. Washburn.

The group of scholars, according to formal criteria of most cited authors and literature obsolescence, in information science form: A. Bookstein, T. Saračević, N. Pravdić, M. Tuđman, M. Kržak, Q. L. Burrell, L. Egghe, J. Martin, D. W. Allen, Y. S. Chen, V. Anić, N. J. Belkin, V. Srića, D. R. Cruickshank, R. Rousseau, H. Sackman, L. M. Stolurow, D. Boras, Z. Dovedan.

According to same criteria, the group of scholars in information systems form: R. A. Kowalski, M. Tuđman, I. Turk, G. B. Davis, P. F. Drucker, I. Martin, J. J. Petrić, V. Strahonja, V. Srića, V. Čerić, Đ. Deželić, B. Aurer, M. Žugaj, S. Tkalac, J. Brumec, A. K. Jain, V. Lovrek.

The group of researchers in all three disciplines is formed from the remaining authors that we did not list as scholars or predecessors.

Table 5: 33 most cited authors from 2000 to 2007

In this paper we were unable to indicate change of place and authors' role within the paradigm in all Information Science disciplines over time. In tables 4 and 5 we show different positions of certain authors – at the beginning and at the end of analyzed development cycle of Information Science. Data fragmented according to document citation and periods[4] indicates that certain authors are being cited for a long period of time. But usually citation period is not longer than two periods of time[5]. In fact only one author occurs in all three periods - M. Plenković. Authors that are cited in the first and second periods are A. Bauer (13, 22), B. C. Brookes (6, 19), J. Martin (10, 8), P. Novosel (8, 11), D. de S. Price (19, 23), T. Saračević (8, 13)[6].

Authors cited in the second and third periods are: I. Maroević (6, 13), V. Srića (7, 11), M. Tuđman (9, 12). It is interesting that E. Garfield (13, 21) and S. Lubetzky (17, 39) are cited in the first and third, but not in the second period.

With these examples it has to be taken into account that there is approximately the same small number of most cited authors in all three periods[7].

Some of these most cited authors are cited in other periods as well, but with not so high frequency. Therefore, the absence of cited frequency indicates the oscillations of the authors' influence and alterations of authors' position in scientific knowledge zones.

Predecessors, Scholars and Researchers' Key Publications According to Disciplines

We can provide empirical data for qualitative analysis of Information Science development, specifically data about who key authors in specific time periods were, as well as the publications crucial for the education and scientific development of information science. But we have to establish the criteria for the selection of those authors and publications. Only after that we can make conclusions about main topics that were dominant in certain Information Science disciplines during thirty years.

Citation criterion, i.e. insight in most cited authors, is not sufficient alone and can lead to wrong conclusions. For example: among most cited authors there are publications of: T. Mušnjak, P. Strčić in arhivistics; P. Selem, E. Laszowki, G. Novak, I. Uranić, etc. in museology; P. Rudan, A. Sujoldžić, D. Horga, etc. in

---

[4] Analyzed cited literature corpus is divided into three periods: 1st period is from 1978 to 1989; 2nd period is from 1990 to 1999; 3rd period is from 2000 to 2007

[5] What we have in mind here is the "durability" of the most cited author, i.e. on their presence among the most cited authors in empirical and conceptual knowledge zones.

[6] The numbers in brackets symbolize the average age of cited literature in the 1st and 2nd periods.

[7] The first period embraced 28 authors whose citation frequency was bigger than 5; the second period embraced 28 authors, too, but their citation frequency was bigger than 11; the third period embraced 33 authors whose citation frequency was bigger than 7.

information science. However, each of this authors' are cited in only one dissertation and therefore it is realistic to assume that these publications or authors are not crucial for Information Science paradigm.

An overview of key authors and their publications can be presented according to several criteria, or combination of criteria, so far described as:

  a) overview of most cited authors and their publications according to disciplines;
  b) overview of most cited authors and their publications according to periods;
  c) overview of most cited authors according to location and authors' role in scientific community: predecessors, scholars, researchers;
  d) overview of most cited authors and their publications according to the number of disciplines in which they were cited.

Since in this paper is not possible to elaborate the presentation of all these overviews, i.e. implementation of all analysis' criteria, this approach will be illustrated only with a few fragmentary examples.

Overview of the most cited authors and their publications according to disciplines

First five most cited authors and publications in museology:

- Strànský, Z.Z.: *Pojam muzeologije; Temelji opće muzeologije; Prezentacija najnovije historije u čehoslovačkim muzejima.*
- Maroević, I.: *Uvod u muzeologiju; Predmet muzeologije u okviru teorijske jezgre informacijskih znanosti; Sadašnjost baštine.*
- Bauer, A.: *Muzeologija; Mreža muzeja i međumuzejska suradnja.*
- Šola, T.: *Prilog mogućoj definiciji muzeologije; Marketing u muzejima : ili o vrlini i kako je obznaniti; Od obrazovanja do komunikacije.*
- Mirnik, I.: *Numizmatička zbirka; Skupni nalaz novca iz Krupe.*

First five most cited authors and publications in information science:

- Burrell, Q.L.: *The analysis of library data; A note on ageing in a library circulation model.*
- Brookes, B.C.: *The foundations of information science; A New Paradigm for Information Science.*
- Egghe, L.: *Introduction to informetrics: quantitive methods in library, documentation and information science; Consequences of Lotka's law for the law of Bradford.*
- Tuđman, M.: *Teorija informacijske znanosti; Struktura kulturne informacije; Obavijest i znanje.*
- Kržak, M.: *Serbo-Croatian Morpho-spelling; Rječnička baza hrvatskoga književnoga jezika; Opisna, stohastička i relacijska gramatika na primjeru morfologije hrvatskog književnog jezika.*

First five most cited authors and publications in information systems:

- Srića, V.: *Uvod u sistemski inženjering*
- Strahonja, V. M. Varga, M. Pavlić: *Projektiranje informacijskih sustava*
- Lazarević, B., V. Jovanović, M. Vučković: *Projektovanje informacijskih sistema*
- Radovan, M.: *Projektiranje informacijskih sistema*
- Tkalac, S.: *Relacijski model podataka*

It is not hard to conclude that overview based only on citation frequency of authors and publications is not sufficient for conclusions that would make us better to understand key authors in Information Science. This list should be corrected and presented in such way that authors can be grouped, not just according to citation frequency, but according to place and role in scientific community, in order to recognize whether they are researchers, scholars or predecessors.

Overview of most cited authors according to periods
From overall number of cited authors in all disciplines, in the first period (from 1978 to 1989) first five most cited publications are:

- Mihajlov, A.I.: *Uvod u informatiku i dokumentaciju.*
- Vreg, F.: *Društveno komuniciranje.*
- Dworatschek, S.: *Uvod u obradu podataka.*
- Eco, U.: *Kultura, informacija, komunikacija.*
- Novosel, P.: *Delegatsko informiranje.*

In the second period (from 1990 to 1999) first five most cited publications are:

- Tuđman, M.: *Teorija informacijske znanosti.*
- Srića, V.: *Uvod u sistemski inženjering.*
- Plevnik, D.: *lnformacija je komunikacija.*
- Žugaj, M.: *Osnove znanstvenog i stručnog rada.*
- Grad, J., G. Resinović, V. Rupnik: *Ekonomika informacijskih sistema.*

In the third period (from 2000 to 2007) first five most cited publications are:

- Lasić-Lazić, J.: *Znanje o znanju.*
- Tuđman, M.: *Obavijest i znanje.*
- Žugaj, M.: *Temelji znanstvenoistraživačkog rada.*
- Boras, D.: *Teorija i pravila segmentacije teksta na hrvatskom jeziku.*
- Eco, U.: *Kultura, informacija, komunikacija.*

An overview of the most cited publication is also not sufficient for the complete understanding of Information Science development. The reason for that is that the overviews of the most cited authors and the most cited publications often differ. In fact, often the most cited authors are the authors that have large number of publications. That amount of publications is what, in the end, puts them in the leading position on the citation scale. In other words, authors that publish larger number of publications cover larger number of topics, and that is the rea-

son why they get cited more often. Overview based only on citation frequency of single publication does not take into account continuous presence of authors that publish large number of publications and their relevance for the broader field of Information Science.

Overview of most cited authors and their publications according to the number of disciplines in which they are cited

Earlier we stated that a small number of authors are cited in more than three Information Science disciplines. That is why we can also display those authors and their papers which are cited in several disciplines.

Authors cited in five or more Information Science disciplines[8]:

- M. Tuđman (21): *Teorija informacijske znanosti; Struktura kulturne informacije; Obavijest i znanje.*
- N. J. Belkin (12): *Information concepts for information science; The cognitive viewpoint in information science; Information science and the phenomenon of information.*
- G. Salton (10): *On the Development of Information Science.*
- T. Saračević (24): *Relevance. A Review of and a Framework for the Thinking on the Notion in Information Science; An Essay on the Past and Future (?) of In-formation Science Education; The impact of information science on library practice.*
- A. I. Mihajlov (9): *Uvod u informatiku i dokumentaciju; Uvodni tečaj o informatiki i dokumentaciji.*

Authors cited in four different Information Science disciplines:

- V. Anić (5): *Pravopisni priručnik hrvatskoga ili srpskoga jezika*
- M. Kržak (12): *Serbo-Croatian Morpho-spelling; Opisna, stohastička i relacijska gramatika na primjeru morfologije hrvatskog književnog jezika; Rječnička baza hrvatskoga književnoga jezika.*
- D. de S. Price (10): *Little Science, Big Science; Networks of Scientific Papers.*
- V. Srića (21): *Informacijski sistemi; Informatički inženjering i menadžment; Od krize do vizije skice - za jugoslavensku tehnološku utopiju.*
- B. Težak (13): *Informaciono-dokumentaciono-komunikacioni (INDOK) sistem.*
- S. Tkalac (7): *Relacijski model podataka.*
- M. Žugaj (10):*Osnove znanstvenog i stručnog rada.*

It is obvious that citation of a larger number of key authors and their publications in several Information Science disciplines, would make a better foundation for joint theoretical basis, because of the fact that scientific community quotes

---

[8] The number of cited publications is given in brackets behind the authors' name. Further, we give the titles of first or next several titles of most cited publications for each author.

and shares same sources. Nevertheless, even in that case one could perceive a lack of insight into the inner dynamics of Information Science development: according to time periods and according to roles that specific group of authors has in specific time period. The lack of insight into the inner dynamics of Information Science development can be perceived even if we expand roles that specific groups of authors have in specific time period: Are these authors a part of research front? Are these authors scholars that dominate in scientific community? or Are these authors predecessors whose knowledge is the authority, but also a part of historical knowledge?

Instead of conclusion

The task of this paper was not to give precise answer on who were the key authors and what were the key publications in Information Science that the Croatian scientific community from 1978 to 2007. Our intention was to prepare possible methodology for the research of Information science development.

Usage of quantitative bibliometrics methods, to make qualitative conclusions could be rather risky. However, with the combination of a variety of quantitative criteria it is possible to process data in such a way that a large number of data (in our research 22,210 cited documents) can be reduced. Using empirical method to find set of key data (several dozens of key authors and publications) we can provide reliable data for qualitatively analyzed.

In our analysis of Information Science development we advocate several starting points. First of all, we demonstrate how it is possible to identify dominant field of scientific influence inside the scientific paradigm (i.e. we recognized empirical knowledge zone, conceptual knowledge zone and research knowledge zone).

Second, we propose criteria for the recognition of several groups of authors, with different influence and roles in described zones: predecessors, scholars and researchers.

Third, based on the examples given in this paper we uphold the use of several criteria that can serve as a filter for data selection: a) citation of authors according to disciplines; b) citation of authors and their publications according to periods; c) classification of authors according to location and role in scientific community (i.e. on predecessors, scholars, researchers); d) overview of authors and their publications according to the number of disciplines in which they are cited.

We are convinced that with this kind of approach it can be possible to obtain empirical data relevant for research and qualitative analysis not only for Information Science development but also for some other disciplines in social sciences.

References

Capurro, Rafael; Chaim Zins. Knowledge Map of Information Science. Rafael Capurro's responses to Chaim Zins. (2006). http://www.capurro.de (2009.)

Garfield, Eugene. Citation indexes to science: a new dimension in documentation through association of ideas. // Science. (1955.): 122; 108-111. http://garfield.library.upenn.edu/essays/v6p468y1983.pdf. (2005.)

Garfield, Eugene. The History and Meaning of the Journal Impact Factor. //JAMA. 295 (2006.) 1, 90-93

Garfield, Eugene; Sher, Irvin H. Genetics Citation Index. Philadelphia, Pa: Institute for Scientific Information; 1963. http://www.garfield.library.upenn.edu/essays/v7p515y1984.pdf. (2005.)

Jokić, Maja. Bibliometrijski aspekti vrednovanja znanstvenog rada. Zagreb: Sveučilišna knjižara, 2005.

Kuhn, Tomas S. Struktura znanstvenih revolucija. Naklada Jasenski i Turk. Hrvatsko sociološko društvo. Zagreb. 1999.

Pečarić, Đilda. Razvoj informacijske znanosti u Hrvatskoj. Bibliometrijska analiza doktorskih disertacija iz informacijskih znanosti 1978.-2007. Zagreb: Filozofski fakultet, doctoral dissertation, manuscript, 2009.

Petrak, Jelka. Citati i njihova analiza. Vidi: Hrvatsko informacijsko i dokumentacijsko društvo, 2003. http://www.hidd.hr/articles/citati.php (URL 10.4.2009).

Tuđman, Miroslav; Pečarić, Đilda. Prilozi dubinskoj analizi komunikacijskih obrazaca. // Informatologija 42, 2009., 2, 87-92.