Đilda Pečarić
The University of Zagreb
dpecaric@ffzg.hr
Miroslav Tuđman
The University of Zagreb
mtudman@ffzg.hr

Understanding the Development of Information Science in Croatia. Co-word Analysis of Doctoral Dissertations from 1978 to 2009

(Rozwój informacji naukowej w Chorwacji. Analiza co-word rozpraw doktorskich za lata 1978-2009)

Słowa kluczowe: analiza co-word, klaster, wizualizacja danych, informacja naukowa, analiza słów kluczowych, rozprawy doktorskie.
Abstrakt: Artykuł przedstawia wyniki analizy rozpraw doktorskich z zakresu informacji naukowej powstałych w Chorwacji w latach 1978-2009. Autorzy pracy skupili się na analizie słów kluczowych, które ich zdaniem dają wgląd w rozwój informacji naukowej na terenie Chorwacji. Analizą została objęta sieć klastrów, która łączy następujące dyscypliny naukowe: archiwistykę i dokumentację naukową, bibliotekoznawstwo, komunikację społeczną, muzealnictwo, informację naukową, systemy informacyjne i leksykografię. Rozwój tematów i korelacji klastrów w informacji naukowej został zaprezentowany dzięki klastrowaniu i wizualizacji danych oraz przeglądowi frekwencji słów kluczowych rozpraw doktorskich powstałych na przestrzeni trzydziestu lat. Wyniki analizy co-word na temat rozwoju informacji naukowej w Chorwacji zostały przedstawione według kryterium czasowego, a także według kryterium przynależności do określonej dyscypliny naukowej.

Keywords: co-word analysis, cluster, data visualization, information science, keywords analysis, doctoral dissertations.
Abstract: For the analysis of doctoral dissertations in information science in the Republic of Croatia (from 1978 to 2009), keywords are used in order to get an insight into the development of information science. By the method of co-word analysis of keywords with which doctoral dissertations are indexed, a network of clusters that match following scientific disciplines is obtained: archival and documentation science, librarianship, communicology, museology, information science, information systems and lexicography. By cluster and data visualization and the overview of keywords frequency, the development of subjects and the correlation of clusters in information science, during the period of thirty years in which doctoral dissertation are made, is shown. The results of the co-word analysis about the

development of information science in the Republic of Croatia are shown according to time periods, but also according to affiliation to certain disciplines inside the information science.

Introduction

The framework for theoretical and practical development of information science in Croatia was founded by Professor Božo Težak in early 1960s (1969, 1971). Božo Težak also established the first postgraduate studies of information science in 1961, with the curriculum based on the idea that different information-documentation disciplines belong to the same field. At the beginning of the 1960s he was convinced that all "traditional information and documentation activities which differ among themselves according to the nature of the objects they deal with (archive documents, publications, museum exhibits, etc.)" can develop as scientific disciplines if they share the information science as a common theory [12]. That is why the domain of information science in Croatia is composed of following disciplines: archivistics and documentation, librarianship, communicology, museology, information science, information systems and lexicography. Doctoral dissertations in these disciplines can be done (in Croatia) on Faculties in Zagreb, Varaždin and Zadar.

First doctoral dissertation in information science was made in 1978. First doctoral dissertation in communicology was made in 1979, then in 1980 in information systems and librarianship. First doctoral dissertation in museology was made in 1988, and in archivistics and documentation in 1991.

Since Težak's time until today, one theoretical and pragmatic question has remained: are all of the above mentioned disciplines a part of information science or are they sciences that jointly make information sciences? This duality, in theory and practice, can be noticed in the usage of both terms (information science and information sciences) in course titles that are studied at the Department of Information Sciences at the Faculty of Humanities and Social Sciences in Zagreb [7].

Our hypothesis is that the verification of the division and understanding of the field of information sciences can be empirically researched with co-word analysis of doctoral dissertations. This quantitative method can be used for mapping the structure of information science. We want to explore *knowledge map*, i.e. structure of research topics and their correlation within broadly defined field of information science without making a priori commitment to any definitions of information sciences. In this way, we want empirical confirmation, not only about the structure and dynamics of the development of information sciences, but also about the correlation of disciplines in information sciences [13].

Methodology

Regarding a larger scientific production, new methods are developed by which we can follow changes in the structure of science, linkage of scientific disciplines, development of new scientific disciplines etc. Beside citation analysis and co-citation analysis, the development of science (i.e. overview of research topics structure and their coherence) could be followed by co-word analysis[20].

The co-word analysis technique was first proposed to map the dynamics of science. Based on the co-occurrence of pairs of words, co-word analysis seeks to extract the themes of science and detect the linkages among these themes directly from the subject content of texts [4].

Co-word analysis is defined as the analysis of the co-occurrence of two or more words in one document or in different documents[21]. Words that are used can be words from text or keywords (word or group of words) taken from the title or from the text of the document, which symbolize the content and enable the search of the document [11].

In our analysis, used key words were taken from abstracts and doctoral dissertation index. A part of dissertations was indexed by key words by PhD candidates themselves. Another part of dissertations was retrospectively indexed. Indexing was done by keyword that best described the topic of research. For the purpose of our analysis key words found was *edited* i.e. synonyms were matched, and the words were stemmed and lemmatized [13, s. 386].

Our research of information science field is done on the corpus of 170 doctoral dissertations in information science at Croatian Universities from 1978 to 2009. The distribution of doctoral dissertations according to disciplines is: information systems 61, information science 35, communicology 28, librarianship 26, archivistics and documentation 10, museology 9, lexicography 1. The majority of doctoral dissertations were made at the Faculty of Organization and Informatics in Varaždin (FOI) – 78, followed by the Faculty of Humanities and Social Sciences in Zagreb – 72 doctoral dissertations. According to the periods of production: 21 doctoral dissertations were made until 1989; 62 doctoral dissertations from 1990 to 1999; 87 doctoral dissertations from 2000 to 2009.

We used co-word method to analyze key words that are retrieved from doctoral dissertations' abstracts and indexes (average number of key words per doctoral dissertation is 9). The total number of key words is 1,462. From that number 515 key words occur only once (35% of all occurrences); 66 key words occur twice (5%); 32 key words occur three times (2%); 19 and 6 key words occur four and

[20] Technique of co-word analysis was for the first time developed in co-operation between the Centre de Sociologie de l'Innovation of the Ecole Nationale Superieure des Mines of Paris and CNRS (Centre National de la Recherche Scientifique) in France during '80s [4].

[21] Two words, W_i and W_j are said to co-occur in corpus if there is at least one document containing both W_i and W_j. The strength of the links between W_i and W_j is given by the number C_{ij} of documents in which the couple (W_i, W_j) appears." [2, s. 144; zob. także 3, s. 54-55].

five times (7%); 31 key words occur six to ten times (15%); 12 key words occur 11 to 19 times (11%); four key words occur 20 to 32 times, and they hold 8% of all key words; and two key words occur 51 times, and they hold 7% of of overall frequency of all key words.

Pairs are made from key words that have frequency higher than 1. The total number of pairs is 5,179, from which 3,938 pairs occur only once (76% pairs), 304 pairs occur twice (12%), 80 pairs occur three times (5%), 24 pairs occur four times (2%), 20 pairs occur five times (2%), and remaining 21 pairs occur from 6 to 21 times (4%).

The programme used for statistical processing of key words, making pairs and clusters is Bibexcel. And for the purpose of visualization of key words' pairs and clusters we used program Pajek.

Mapping the Structure and Dynamics of Information Science Development in Croatia

A co-occurrence matrix of keywords was organized and analysed according to several criteria. Our goal was to extract central themes in the domain of information science in order to detect linkages among these themes. In order to follow the dynamics of information science development, density of certain clusters and their linkages, interaction and transformation of centrality, we mapped key words from doctoral dissertations according to time periods, disciplines and faculties at which dissertations were done.

With such methodological approach, we obtained series of maps that allow us more precise interpretation of empirical domain and the dynamic of information science development in Croatia [13, s. 387].

Figure 1 shows the structure of main topics and domains in information science based on the key words according to which analysed texts are indexed. Dominant cluster has several central nodes, whose hierarchy and linkages can be determined by the frequencies of words, pairs of words, and amount of linkages [13, s. 387]. Central nodes according to the frequency of pairs of words are:

Table 1. Frequency of pairs of key words

Pairs of key words		**Frequency**
Information	System	21
Information System	System	21
Communication	Information	20
Model	System	14
Information	Information System	12
Information	Model	11
Analysis	Model	10

Source: self-elaboration.

In this case linkage density of these words within the cluster is more than obvious and that is the indicator of internal strength of the cluster [13, s. 387]. Likewise, this is the indicator of the existing subject areas[22] within but also among the clusters. Thus, from the map in figure 1, *dominant cluster* with the sub-cluster *archivistic* and isolated cluster *museums* can also be discerned. Sub-clusters are linked with dominant clusters by mediator words.

Central poles of the main cluster are: information, information system, system, model, communication, analysis, media, library. Cluster that occurs on the periphery of the central cluster is archivistic. This cluster (archivistic) is through node protection linked to the dominant cluster and belongs to information science domain.

Cluster *museums* is linked (on the lower level) with mediator words *document* and *information* to the central cluster, and that makes the subject area of *museums* a mutual subject area of the information science research.

From the map presented in figure 1 we can conclude that information science discipline is a homogeneous field. In other words, we can discern central poles that represent different information science disciplines and make one cluster (dominant cluster); even sub-cluster *archivistic* and isolated cluster *museums* are linked by mediator words to the dominant cluster. Also, clusters *museums* and *archivistic* show the uniqueness of these disciplines, although they are linked with dominant clusters by mediator words.

Information science was established in the 1960s and the dynamic of its development can be followed with the development of thematic topics and the growth of knowledge production[23]. Only when we show cumulative knowledge map, shown in figure 1, in temporal series of the production of analysed documents (i.e. according to the publication of doctoral dissertation), we can get an insight into the occurrence and transformation of clusters, and thus also the internal dynamics of the development of thematic topics in information science. Since doctoral dissertations have been made in the time span of 30 years, we observed the development of clusters in ten-year time periods [13, s. 387-388].

During the period of 30 years in which analysed documents were made, key word map shows constant change of central poles and dominant nodes in the clusters. In the first decade (1978-1989) following pairs have dominant nodes and central position in the cluster (figure 2): *information – communication, information – information systems*; dominant nodes are: *system, information science, information, computer, data base, and programming.*

[22] „Subject field or subject area is totality of *concepts* or topics observed like unity of some human activity“ [11].

[23] Overview of the usage of co-word analysis in mapping the dynamic of science and technology [zob. 2, s. 143-179].

In the first period only one cluster is distinguished: dominant cluster with 4-5 central topics: *information, communication, model, computer, information science* [6, s. 218-219].

In the second period (1990-1999) central pairs of clusters shift (figure 3): *information system – system, information – information system, information system – model*. However, central nodes with close linkages still remain: *information, information system, information technology, archive* and *library*. Within the dominant cluster several substructures formed by the nodes with a large number of linkages (as indicated in the brackets) can be recognized: information system (17), systems (17), information (14), model (9), protection (8), library (7), decision making (6), education (6). The density of these linkages refers to the content of information science field in that period. With the dominant cluster in this period *information system*, three other clusters can be recognized: *library, graphical technology* and *information technology.* Isolated clusters in the same period are: *library science, applications* and *public opinion* [6, s. 219-220].

The third period 2000-2009 has following clusters (figure 4): *main clusters, archives, museums*. Even if sub-clusters are noticeable like *e-learning, libraries* and *media*, they are incorporated in the main cluster and make a homogeneous field of information science. The most frequent key words in this period are: system (14), ICT[24] (12), information (9), communication (7), media (7). Here we can see that the frequency of certain words alone is not crucial for the formation of a cluster, but if it co-occurs with other frequent words it determines the thematic area and overall knowledge map [6, s. 220-221].

With cluster comparison in figures from 1 to 4, large differences in number and content of the clusters can be noticed. Although these knowledge maps were done on the same corpus of data, their time sequence warns us that knowledge corpus constantly changes, transforms and re-structures.

Universities and Mapping the Information Science. Following figures (5-7) show maps of information science as they been created at the University, more precisely Faculty of Humanities and Social Sciences (FHSS) and Faculty of Organization and Informatics in Varaždin (FOI). Namely, most doctoral dissertations were done on these two faculties [13, s. 224].

Figure 5 shows knowledge map, the *structure* of information sciences as it has been *produced* at the Faculty of Humanities and Social Sciences at the University of Zagreb by the doctoral dissertations analysis. Clusters were obtained from empirical data, i.e. based on the co-occurrence of pairs of words, extracted subject areas and their linkages. Several clusters can be recognized: dominant cluster, cluster *information technology,* and isolated clusters: *archives* and *museums.* In the dominant cluster we can recognize four thematic topics: *libraries, e-learning, communication, information systems* [6, s. 224].

[24] ICT = Information communication technology

These clusters, or more precisely subject areas that are bound by those clusters, greatly correspond with organizational scheme of the Department of Information Sciences. The department consists of the following sub-departments: archivistics and documentation, librarianship, social-humanistic informatics, museology, knowledge organisation and lexicography. Apropos, these subject areas greatly match the courses offered by the Department of Information Sciences: archival science study, library science study, museum study and information science study [13, s. 390].

Figure 6 shows another evidence that in the last observed period (2000-2009) at the Faculty of Humanities and Social Sciences the same subject areas are repeated: *information technology, e-learning, communication, general theory* and *libraries* in dominant cluster, and as isolated clusters *archives* and *museums*, which are linked on the lower level to the dominant cluster.

Figure 7 shows subject areas that are embedded in the map of information sciences by the Faculty of Organization and Informatics. Central subject area is *information systems.* Information systems is also scientific and teaching area at the Faculty of Organization and Informatics.

The most frequent pairs of words and at the same time the central nodes of clusters are: *information system, system, information, model, decision making, analysis.* These nodes are the center of subnetworks within the large area of *information systems* that belongs to information science [6, s. 226].

According to the cluster density and subject area correlations within this cluster, it can be concluded that the cluster *information system* is a homogeneous area. More detailed analysis, by ranking subject areas in terms of their internal coherence, would give us additional insight into subject areas hierarchy and their correlations [13, s. 391].

For our research it is important to notice the interpolation of this cluster, that is the subject area which it shows, in the overall corpus of information sciences' knowledge map. Or in other words: information science structure shown in figure 1 is the result of integration and linkages of topics and subject areas of information sciences shown in figure 5 and figure 7.

Instead of Conclusion

The goal of this paper was to show how co-word analysis and techniques can be used to study the development and interactions among information science disciplines. Co-word method enabled systematic content analysis of doctoral dissertations in information sciences made at universities in Croatia in a period of thirty years. The result are the maps of dominant subject areas which were of interest for the information scientists in the analysed period. We believe that we also succeeded to describe the dynamics of the development of these subject areas and their transformation during longer time period, as well as interaction among information science disciplines.

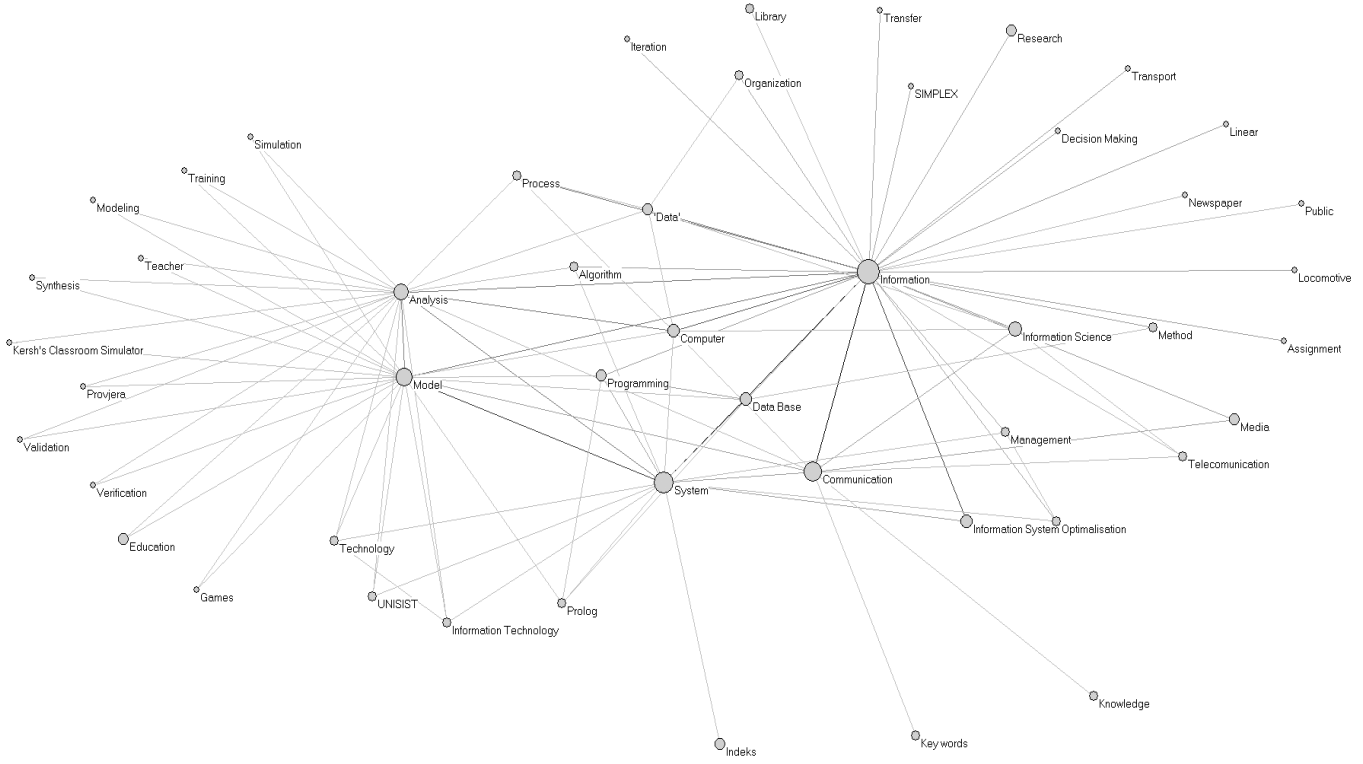
The analysis of *knowledge map*, i.e. thematic topics and their dynamics at the Faculty of Organization and Informatics and at the Faculty of Humanities and Social Sciences indicates the existence of two different areas of research interest and research topics. However, our analysis also indicates the existence of the cooperation and correlations between these two faculties within broadly defined area of information sciences.

We did not get the answer to the question whether we are dealing with information science or information sciences. It is still not clear whether information science is a sum of several disciplines, or information sciences are a common name for several sciences. However, we came to the empirical confirmation that the analysed subject areas share mutual structure and the dynamic of development, and we offer confirmation about these correlations. This points us to the conclusion that differences in the definition of information sciences are derived from the differences in the comprehension of information as a basic subject [8]. But those differences disappear when all subject areas become a part of matrix and maps of the information research area. At the same time these differences are the driving force of the dynamics and development of new subjects and thematic topics in broadly defined area of information science.

Figure 1. Representation of the structure of information science in Croatia (1978-2009)

Source: self-elaboration.

Figure 2. Dynamics of information science development (1st period 1987-1989)

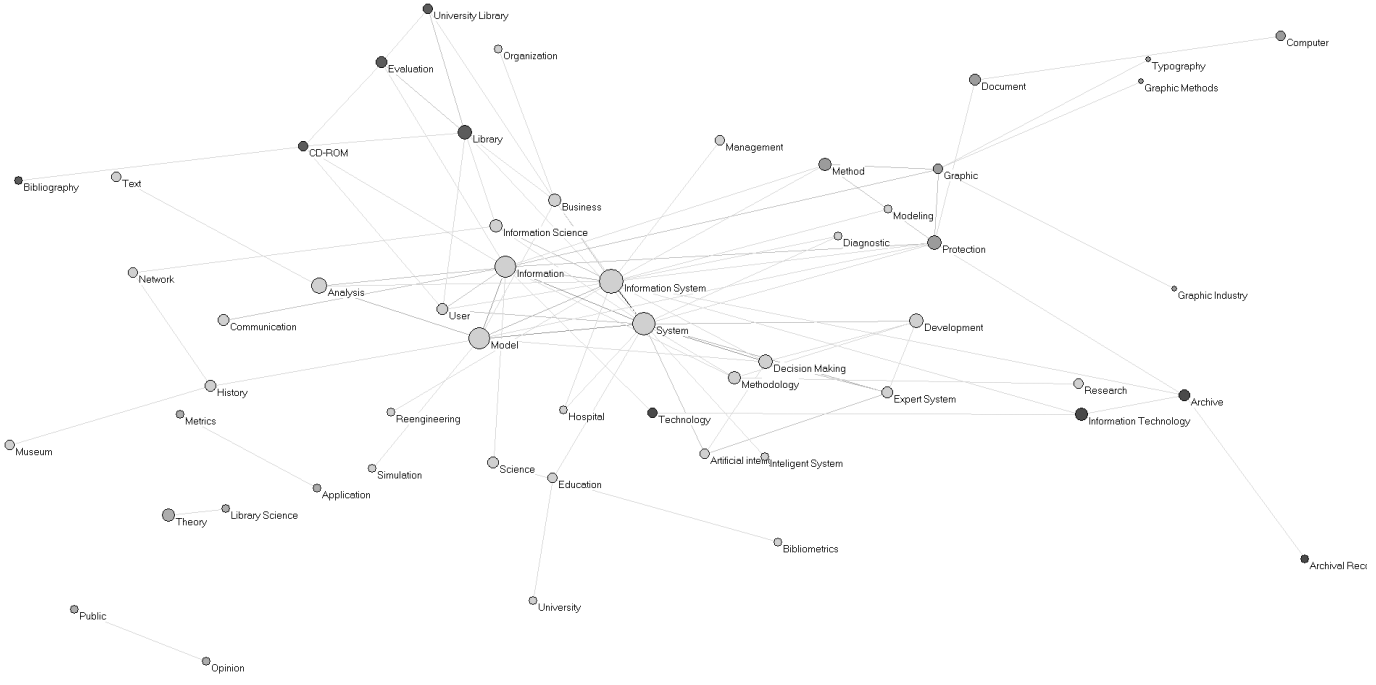


Legend:

dominant cluster

Source: self-elaboration.

Figure 3. Dynamics of information science development (2nd period 1990-1999)

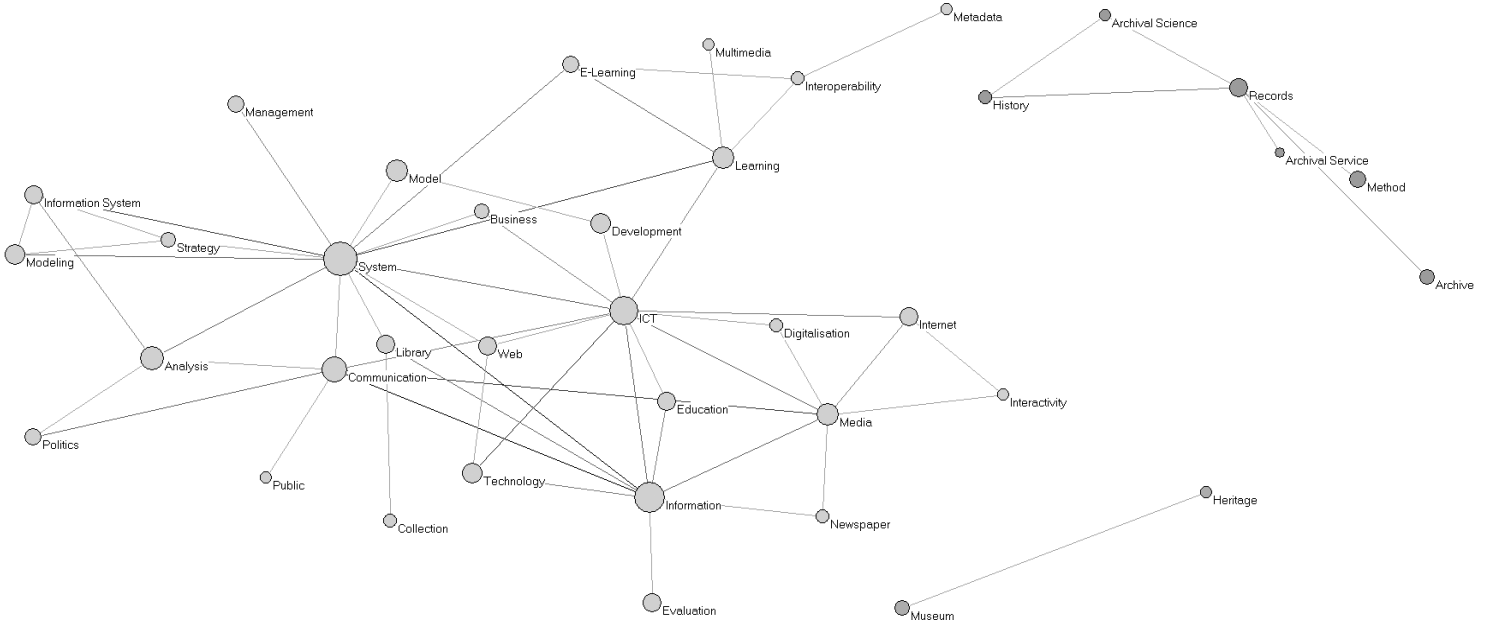


Legend:

- information system
- library
- graphical technology
- information technology
- library science
- applications
- public opinion

Source: self-elaboration.

Figure 4. Dynamics of information science development (3rd period 2000-2009)

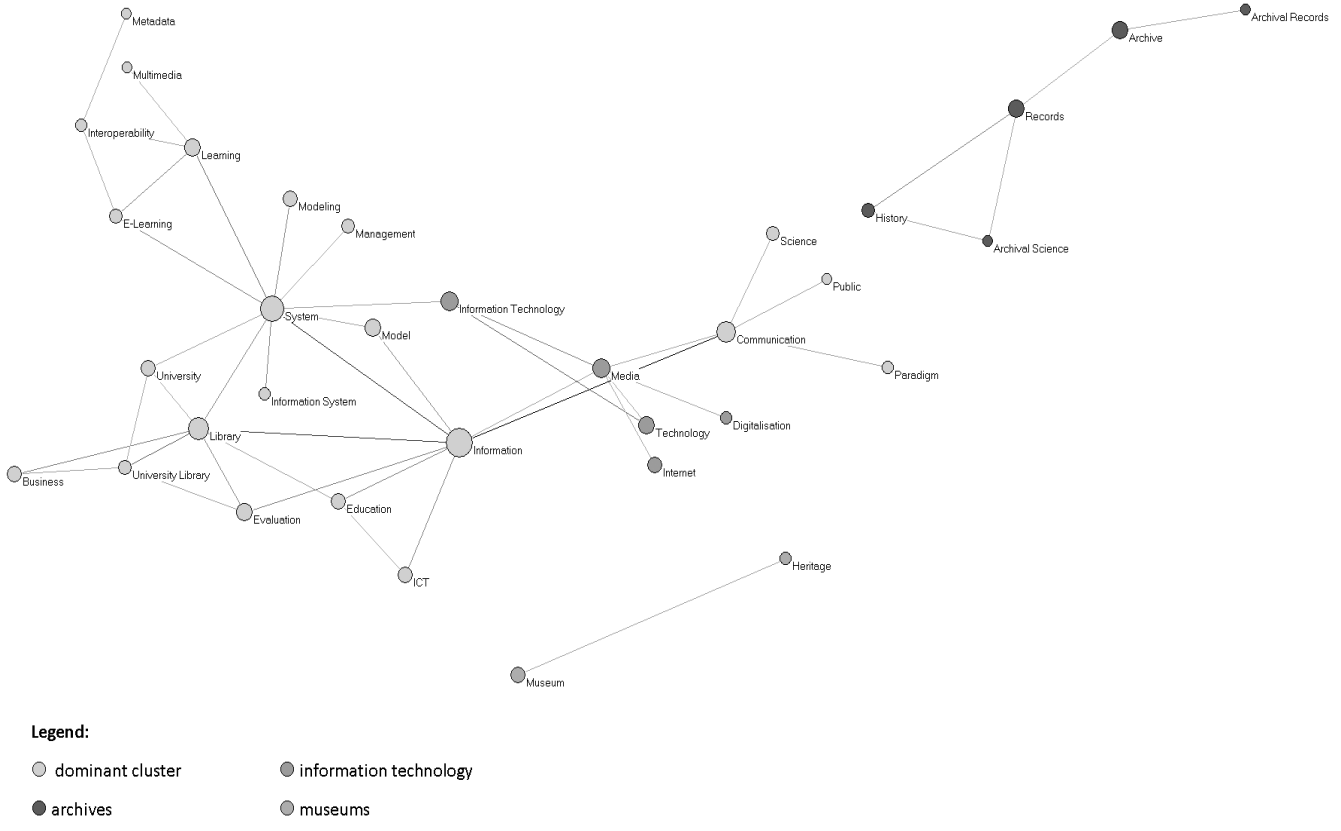


Legend:

main cluster archives museums

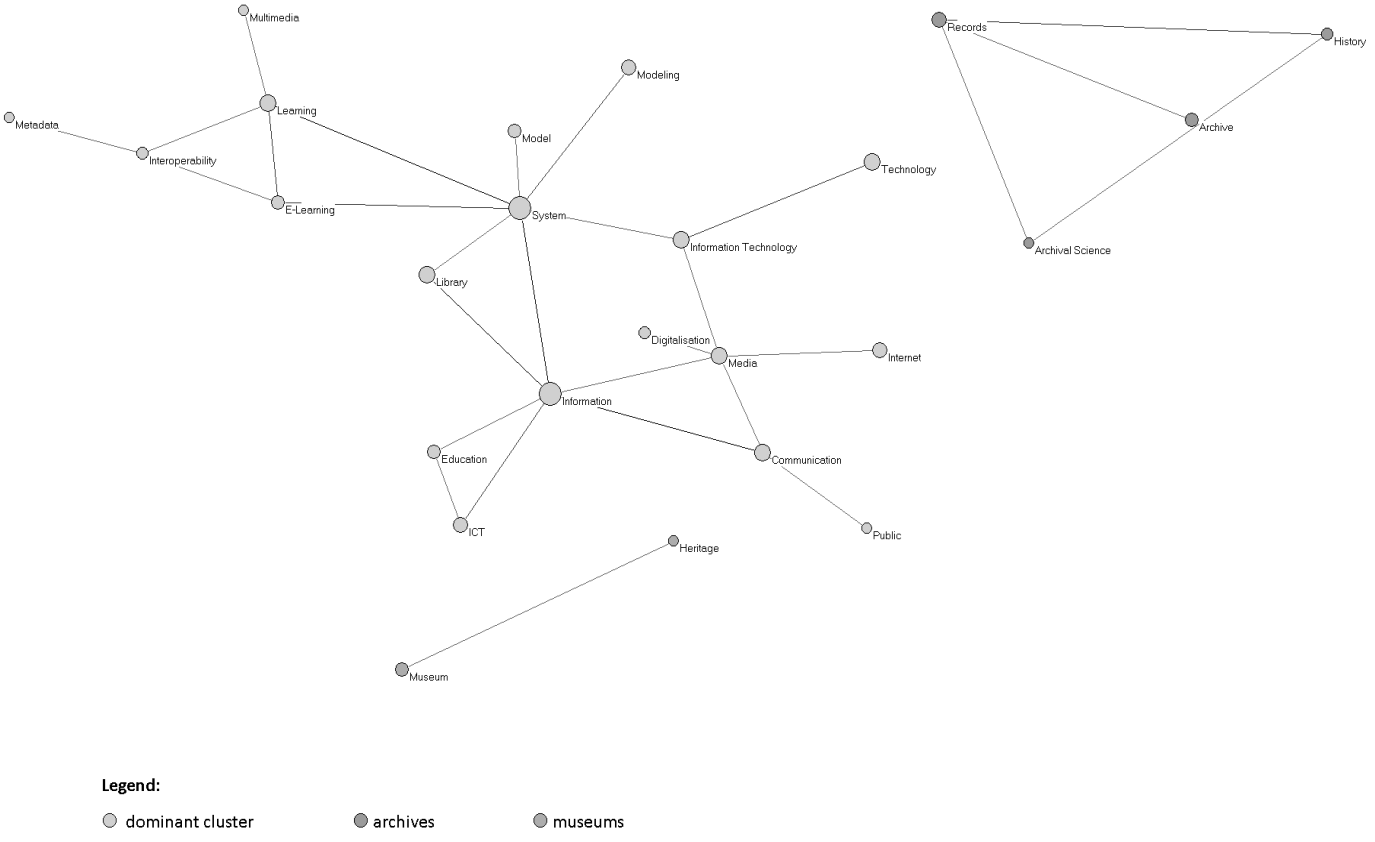
Source: self-elaboration.

Figure 5. The Faculty of Humanities and Social Sciences: information science structure according co-word clusters



Source: self-elaboration.

Figure 6. The dynamics of information science development (Faculty of Humanities and Social Sciences: period 2000-2009)



Source: self-elaboration.

Figure 7. The Faculty of Organization and Informatics: information science structure according co-word clusters

Source: self-elaboration.

References

1. *Bibexcel free-ware for academic non-profit use program* [Dokument elektroniczny]. Tryb dostępu: http://www.umu.se/inforsk/Bibexcel. Stan z dnia 01.03.2010.
2. DE BELLIS Nicola. *Bibliometrics and Citation Analysis*. Lanham, Md. : Scarecrow Press, 2009. ISBN 978-0-8108-6713-0.
3. DIODATO Virgil. *Dictionary of Bibliometrics*. New York: Haworth Press, 1994. ISBN 1-56024-852-1.
4. HE Qin. Knowledge Discovery Through Co-Word Analysis. *Library Trends*. 1999, vol. 48, nr 1, s. 133-159. ISSN 0024-2594.
5. *Pajek – open source program for visualization and clustering* [Dokument elektroniczny]. Tryb dostępu: http://vlado.fmf.uni-lj.si/pub/networks/pajek/. Stan z dnia 01.03.2010.
6. PEČARIĆ Đilda. *Development of Information Sciences in Croatia. Bibliometric Analysis of Doctoral Dissertations in Information Sciences from 1978 to 2009*. Saarbrücken**:** LAP LAMBERT Academic Publishing, 2011. ISBN 978-3846508893.
7. *Program studija Odsjeka za informacijske znanosti. Odsjek za informacijske znanosti Filozofskog fakulteta Sveučilišta u Zagrebu* [Dokument elektroniczny]. Tryb dostępu: http://www.ffzg.hr/programi/preddiplomski.html/. Stan z dnia 16.03.2010.
8. SARACEVIC Tefko. Information Science. *Journal of the American Society for Information Science*. 1999, vol. 50, nr 12, s. 1051-1063. ISSN 1532-2882.
9. TEŽAK Božo. Informacione znanosti i službe: njihova struktura, odnosi i politika – Information sciences and Services: Components, Relationships and Politics. *Informatologia Yugoslavica*. 1969, vol. 1, nr 1/4 , s. 13-31. ISSN 0046-9483.
10. TEŽAK Božo. Informatika – ime za zbunjujući ili razjašnjavajući koncept – Informatics – the Name of Confusing or Clearing Concept. *Infomratologia Yugoslavica*. 1971, vol. 3, nr 1/4, s. 1-13. ISSN 0046-9483.
11. TUĐMAN Miroslav. *Obavijest i znanje: s rječnikom osnovnih pojmova*. Zagreb: Zavod za informacijske studije, 1990. ISBN 8680279374.
12. TUĐMAN Miroslav. Profesor dr. Božo Težak i razvoj informacijske znanosti. In TEŽAK Đurđica (red.) *Profesor Božo Težak, lučonoša znanosti*. Zagreb: Hrvatska sveučilišna naklada, 2007, s. 257-271. ISBN 978-953-169-149-9.
13. TUĐMAN Miroslav, PEČARIĆ Đilda. Co-word analysis of doctoral dissertations in Information Science in the Republic of Croatia from 1978 to 2007. Contribution to research of development of Information Science. In KATSIRIKOU Anthi, SKIADAS Christos. (red.). *Qualitative and quantitative methods in libraries: theory and applications: proceedings of the International Conference on QQML2009, Chania, Crete, Greece, 26-29 May 2009*. New Jersey [i in.]: World Scientific, 2010, s. 385-393. ISBN 978-981-4299-69-5.