

## **Information Ecology of Bioinformatic in Web of Science with Emphasizing on Articles Thematic Interaction**

### **Masoume Kiani**

PhD Candidate, Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran. ORCID: 0000-0001-7145-0031. E-mail: m.kiani@edu.ui.ac.ir

### **Asefeh Asemi\***

\*Corresponding Author, Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran. E-mail: asefi@edu.ui.ac.ir

Doctoral School of Business Informatics, Corvinus University of Budapest, Hungary. ORCID: 000-0003-1667-4408. E-mail: asefi.asefeh@uni-corvinus.hu

### **Mozaffar CheshmehSohrabi**

Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran. ORCID: 0000-0003-4974-4193. E-mail: mo.sohrabi@edu.ui.ac.ir

### **Ahmad Shabani**

Department of Knowledge and Information Science, University of Isfahan, Isfahan, Iran. ORCID: 0000-0003-0466-6240. E-mail: shabania@edu.ui.ac.ir

*Received January 15, 2020; Accepted March 20, 2020*

---

### **Abstract**

The present study aims to represent the infoecology of bioinformatics, with an emphasis on the topic of relationships between studies in order to provide a scientific framework for infoecological investigations in this field. It was conducted using the co-occurrence analysis of the topic keywords of studies available on the Web of Science (WoS) Database from Clarivate

Analytics and using the social network analysis. The infoecology of the studies' topic relationships indicated that the topics of the studies were inter-related as information elements, also known as infocenoses, in the form of seven basic clusters in the infotope of the WoS. According to the centrality indexes, topics such as proteomics, microarray, MicroRNA, genomics, gene expression, computational biology, database, and data mining had the largest topic relationships. Moreover, as emerging information elements, LNCRNA, big data, differentially expressed genes, Osteosarcoma, next-generation sequencing, protein-RNA-SEQ, protein interaction network, high-throughput sequencing metagenomics, and ITRAQ were found to create new topic relationships.

## Keywords

Infoecology; Information ecology; Bioinformatics; Topic relationships; Clarivate Analytics; Web of Science (WoS); Scientific studies

---

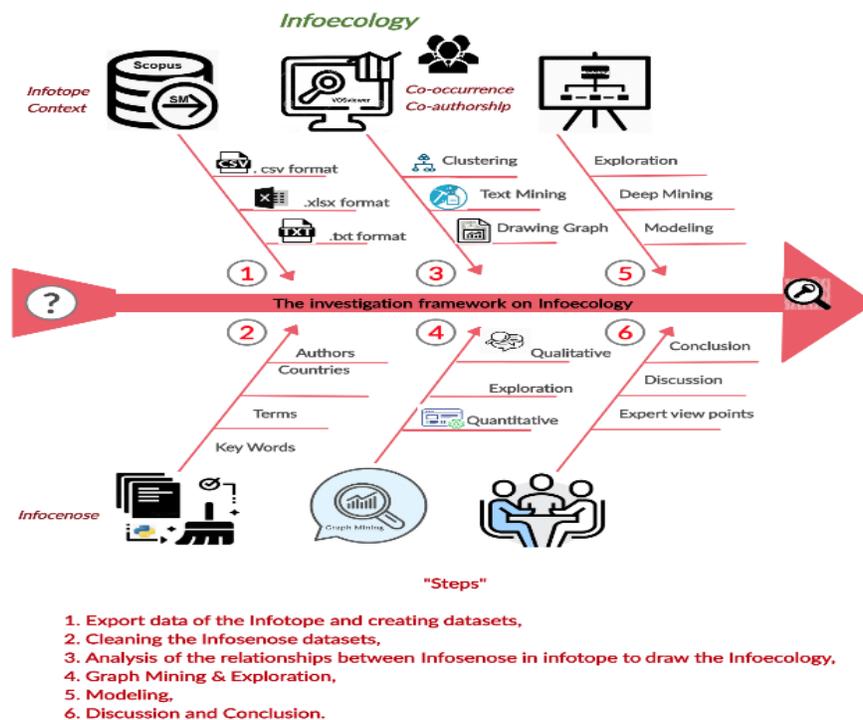
## Introduction

Researchers in different scientific fields are showing interest in finding out the thematic relationships governing different areas of studies. Most of the previous researches rely on citation studies and consist of quantitative data to address this issue. Now, the exploration of the thematic relationships based on information ecology seeks to perform a deeper analysis of the relation among different information elements in an informational ecosystem. In the information ecology perspective, various techniques are applied to explore interrelationships among information elements. This brings forth a pattern that properly defines the thematic relationships in any discipline. Information ecology also aims at describing and analyzing the information systems based on metaphoric uses of such concepts in environmental ecology as biotope, species, evolution, ecosystem, and evolution and equilibrium, to study the distribution and abundance of organisms, relationships among them, and how the organisms can affect their environment or be affected by it (Eryomin 1998). International Encyclopedia of Information and Library Science defines information ecology as: “a study of interpersonal communication, activities, technologies, and information environment” (Feather & Sturges 2003). As noted earlier, the information ecology concept is metaphorically structured which considers environment information as an ecosystem.

The information ecosystem is a type of entity contains communities of diverse living organisms, their interrelationships, and their habitats. There are three main features of information ecology, including environment, information living elements, and interaction between both (Waring 1989). Thus, environment and elements are the two important factors that appear in information ecosystems such as biological ecosystems. The information ecosystem environment and a set of

information elements can be called as ‘infotope’ and infocenose’, respectively. The former comprises of all cultural, economic, and epistemic platforms, in which the information is conveyed, communicated, and evaluated. The latter consists of a set of information elements that depend on each other and can persist in a particular environment due to the interrelationships among them; while the livelong and persistence of the elements are affected by different factors. Infotope and infocenose can, therefore, be seen as an inseparable pair of information ecosystems that create a partially consistent system. This implies that the information ecosystem is an essential part of information ecology in terms of its role and function (Nardi & O’Day 1999). Exploring infocenose thematic relationships of scientific papers in different infotopes, including such scientific databases as Clarivate, will offer patterns of progression and evolution to a particular subject of studies over time. The pattern allows researches in any discipline to compare papers published with a similar subject area for a certain period and analyze the changes that occurred. Researchers, then, plan future researches based on the analysis results.

Asemi and Ko (2019) propose a framework for a study of infoecology in different subject areas (FIG. 1). Based on this framework, they investigated the infoecology of Smart Manufacturing in 6 steps: 1. Export data of the infotope and creating datasets, 2. Cleaning the infocenose datasets, 3. Analysis of the relationships between infocenose in infotope to draw the infoecology, 4. Graph Mining, 5. Modeling, and 6. Discussion and Conclusion. The purpose of this framework is to guide scholars using infoecology analysis as a research method.



**Figure 1. Investigating framework on infoecology**

Across various academic fields, bioinformatics has gained attention in different areas such as information technology, mathematics, and biological science in recent years. This is because of high biological data availability and the need for converting these data into usable information and knowledge (Chen, 2005). Bioinformatics has also gained much attention from researchers in the last few years as an interdisciplinary science that, along with biology, can be applied to other fields of studies (Jiang, Zhang, & Zhang 2013). A definition depicts bioinformatics as “the collection, classification, storage, and analysis of biochemical and biological information using computers especially as applied to molecular genetics and genomics” (Merriam Webster collegiate Dictionary, 1993).

## **Problem Statement**

The main emphasis in the current study was on present fact that, despite the importance of the information ecosystem in bioinformatics, little attention has been paid to this inquiry with respect to previous studies. To conduct a large-scale and effective research in this area, it is necessary to draw a topic map. In doing so, analyses of ecological information are found as a promising method. The analyses are performed in different sizes of information and are applied to determine the relationships among any type of infocenose in a different type of infotope. Our first analysis results showed that there was no determined main (cores) and minor (close to the core) thematic infocenose of bioinformatics published paper in the WoS infotope. Thus, to identify what core and close to core Themes of papers in bioinformatics in the WoS were seemed interesting to researchers and scholars, ecological information analyses were performed.

Present paper also investigated the relations among emerging Themes of the information ecosystem in the field of bioinformatics and identified information ecology for such scientific area. The information ecology was confined to the thematic infocenose of bioinformatics published paper in the WoS infotope. The bioinformatics ecosystem also consisted of the thematic relationship of papers (i.e., infocenose or information elements) in the information environment of the Clarivate database (i.e., infotope). Conducting ecological information analysis of bioinformatics papers only included the publications during 1980-2018 in the WoS. The current study, however, was not undertaken as an investigation of all infocenose in the infotope, and the relations among the elements, including users, technologies, and other elements of the ecosystem. Only the thematic relationships among infocenose and how such relationships influence each other were investigated in-depth. These aims were identified by the following questions:

1. What are the core infocenose in the information ecology of the bioinformatics published papers in the WoS from Clarivate Analytics, and what are their relations to the other Themes?

2. What are the emerging infocenose in the information ecology of the bioinformatics published papers in the WoS from Clarivate Analytics, and what are their relations to the other Themes?
3. How ecological information map of thematic relationships among bioinformatics published papers in the WoS from Clarivate Analytics is constructed?

According to the findings, present study seems important for supporting the identification of important areas of bioinformatics study and the relations among them in the form of an information ecosystem. Also, the results were found to be useful to the planning of education and research and to the development of a framework for collaborative work of experts in researches in order to extend the knowledge of the bioinformatics area of study. The findings also involve providing a good information source for decision-and policy-makers in performing scientific activities in bioinformatics at universities. They, in addition, help researchers determine policies for clinical trials by providing proper guidelines. Thus, these researches will overcome the lack of bioinformatics Themes in their studies and help budget planning for research.

## Literature Review

There have been no researches that specifically address ecological information analysis of bioinformatics published papers. In fact, this section includes only those previous studies with the method similar to that of the current investigation. In 2001, Ding, Chowdhury & Foo study's titled "Bibliometric cartography of information retrieval research by using co-word analysis" explores the structural changes of informational retrieval thematic relationships over a period of 11 years (1987-1997), using co-word analysis of published papers. Their report depicts the thematic patterns and trends for information retrieval and includes all indexed papers in the citation databases of SCIE and SSCI from 1987 to 1997. The total number of 2012 articles included in this study as the selected population. The titles, abstracts, and key terms of all identified articles were reviewed in full by the researchers to gather the required data. Using the method of hierarchical clustering, analyses of the obtained data were also performed and resulted in the identification of five major clusters, and several minor clusters from each major cluster. The first cluster includes definitions such as searching, analysis systems, computerized information storage and retrieval, online information retrieval, and the like. The second cluster consists of online indexes, libraries, information-seeking behavior, and information storage and retrieval. The third cluster provides image storage and retrieval, retrieval algorithms, and so forth. The fourth cluster consists of multimedia, internet, World Wide Web, computer programs, text analysis, etc. In the last cluster, such topics as every data storage, memory, etc. They conclude that the use of co-word analysis is suitable for identifying patterns and trends of a particular academic field for a period of time.

In their article “co-word analysis of keywords in scientific literature produced at the Department of Information Sciences in Zagreb”, Bosanac, Matešić and Tolić (2009) discuss how the concepts are related in studied literature for the year 1995 to 2009. They preferred the method of co-word analysis which gave the identification of the important concepts of information science papers published from 1995 to 2009. These findings were then visualized. The data of this study included 376 published papers from Croatian scientific bibliography which were extracted by 35 faculty members of the Department of Information Sciences in Zagreb through these 14 years period. Out of 689 keywords used in the articles, 64 were analyzed through the method of co-word. In this study, a grid of concepts and the clusters for the whole 14 years period, as well as for the first seven years (1995 to 2002) and the second seven years (2002 to 2009) separately, were constructed. Their finding illustrates 5 major clusters from which Themes of museology, Croatian, scientific relations, and information literacy are depicted. The first seven years consists of 5 major clusters from the analysis of 35 key terms that discusses topics of information technology, education, N.L.P, society, and user interface. The second seven years concludes four clusters from the rest key terms (n=37) with Themes such as education, data warehouse, web 2.0, and society. Their study also offers some other important findings. First, it mentions that education and society play a more decisive role in the grid center. Second, it highlights that new concepts of a grid are used as a replacement for similar concepts.

Co-word analysis method is also addressed in Muñoz-Leiva, Viedma-del-Jesús, Sánchez-Fernández and López-Herrera (2012) study's, titled “An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective” for identifying the main Themes. Also, the scientific mapping techniques are used to underline the linkage among main concepts, depict their thematic structures, and categorize them into more minor domains. 1800 social science articles, all indexed in Web of Science from 1996 to 2008, are reviewed in their paper. The identification of dominant Themes is based on quantitative and qualitative methods. The former and the latter methods are used for developing thematic clusters (i.e., related concepts) and for measuring the quality of the selected clusters from the qualitative indexes (i.e., citations). The findings from this study indicate that the consumer behavior themes increased by an increase in the number of publishers during 1996-2008. Besides, the increased attention to “information and communications technology” in the area of consumer behavior in recent years was reported in these findings. Because of more centrality of this area than during the last 5 years, the result also notes the increased attention to “consumer satisfaction and food products”. In sum, the article concludes that such themes as “food products, perception, attitude, satisfaction, and consumer trust in underlying countries” are the most important subjects in the area of consumer behavior.

The bioinformatics thematic changes from 2000 to 2011 are investigated in the study by Song, Heo & Kim (2014), titled “Analyzing topic evolution in bioinformatics: investigation of dynamics of the field with conference data in DBLP”, to distinguish dynamic bioinformatics

topics. Their study, therefore, uses data from the conferences on bioinformatics from 2000-2011. The topic evolution in their study is presented from the use of a thematic clustering method. The similarities among resulting clusters are based on the thematic similarities. Over 12 years' period, the analysis performs at four points as follows: The first point is from 2000 to 2002, the second from 2003 to 2005, the third from 2006 to 2008, and the fourth point is from 2009 to 2011. It can be seen from the first to third points that new topics are emerged and extended. While in the third to fourth points, these topics are mixed together and each exists in a closer relationship with the other. They furthermore verify such changes on topics by using the patterns of cooperation over time.

A study by Sedighi (2015), “Application of word co-occurrence analysis method in the mapping of the scientific fields (case study: the field of Informetrics)”, identify the thematic subfields of the informatics field and explores how these subfields influence each other, using the method of co-word analysis. This implies that Sedighi’s investigation is an attempt for studying the implementation of the co-word analysis method and its effectiveness in constructing the map of scientific fields. Based on the constructed map on the field literature, the study reports the most frequently used subjects such as information science, library, bibliography analysis, innovation, and text mining. The finding on the mapping of co-words over different time periods shows persistence and changes in the concepts and words related to the informatics area. Some words, including “bibliography analysis”, persist during all year period while others disappear over time. New words also emerge from the recombination of existing words and their interaction with new technologies and changes over time.

Zarei and Babolhavaeji (2016) present the information ecology of the thematic structure in information science journals. The identification of information ecology is based on the interaction of information science thematic structures and other structures, including information and communication technology structures, for all journals in information science during 2002-2012. Results from the information ecology identification demonstrate that there is an interaction between all information science thematic structures and the other structures. Information and communication technology receive a higher proportion of interaction while this between other structures and different areas of science is a lower amount. The result concludes with the impact of the environment (i.e., eco) on the information ecology of the information science thematic structures. Such an influence helps thematic structures develop and have interaction on issues related to the environment, particularly information and communication technology.

Wang, Guo, Yang, Chen and Zhang (2017) examine information ecology published papers from the past to their present time to predict the future trend in this field. To achieve a better understanding of the trends in research on information ecology, they use 138 papers of this field from the Web of Science database published in the years 1992-2013. Their review of articles is performed to inspect any changes in terms of the number of published researches in information

ecology over time, their geographical and country distribution, the topics, the approaches to research, funding resources, and any other considerations indicating the changes in trends in information ecology. The study aims to help researchers to attain a deeper understanding of topics and discussions in the information ecology field and inform any changes in trends in this field. The findings report the countries with the highest publication numbers, including the U.S, England, Canada, France, Germany, Netherland, China, Russia, and Spain, respectively. The United States accounts for more than half of the publications, almost 36 percent. On the other side, the number of publications among Asian countries is limited as China contributing 5% of the total number of publications. Findings also indicate that developed countries such as the U.S are the pioneers in the information ecology research and have beaten their developing countries rivals.

In sum, our literature review reveals that the thematic relationships among different scientific fields have always conveyed the potential of attracting considerable attention from researchers. The section also shows that the number of studies on thematic relationships of different periods of time is relatively large, among which much work undertakes the method of co-word analysis, as the case for our investigation. However, there is no attempt to analyze thematic relationships with the use of information ecology in bioinformatics while it seems reasonable to fill this gap.

## Materials and Methods

Carrying out an information ecology investigation can result in practical benefits. The present research uses an informatics approach to analyze the ecological information of thematic elements of an information environment. The population study includes all papers that have been published in bioinformatics for the years 1980 to 2018 added to the Web of Science (WoS) database from Clarivate Analytics. To exclude the data, a keyword search strategy is used. First, the study interviews bioinformatics experts with the generated keywords. Next, a thematically grouped key terms is created based on the experts' opinions and Medical Subject Headings (MeSH) keywords. After selecting the keywords, the research method is performed through the following steps:

- Chosen keywords are searched in a specific topic area in the Web of Science (WoS) database from Clarivate Analytics with the use of the OR operator. The following strategy depicts the searching process: TS=("Computational Biology" OR "Biology, Computational" OR "Computational Molecular Biology" OR "Biologies, Computational Molecular" OR "Biology, Computational Molecular" OR "Computational Molecular Biologies" OR "Molecular Biologies, Computational" OR "Molecular Biology, Computational" OR "Bio-Informatics" OR "Bio Informatics" OR "Bio-Informatic" OR "Bioinformatics" OR "Bioinformatic" OR "Medical Informatics" OR "Health Informatics" OR "Informatics, Health" OR "Informatics, Medical" OR "Clinical Informatics" OR "Informatics, Clinical" OR "Medical Computer Science" OR "Computer

Science, Medical" OR "Medical Computer Sciences" OR "Science, Medical Computer" OR "Health Information Technology" OR "Health Information Technologies" OR "Information Technology, Health" OR "Technology, Health Information" OR "Information Science, Medical" OR "Medical Information Sciences" OR "Medical Information Science")

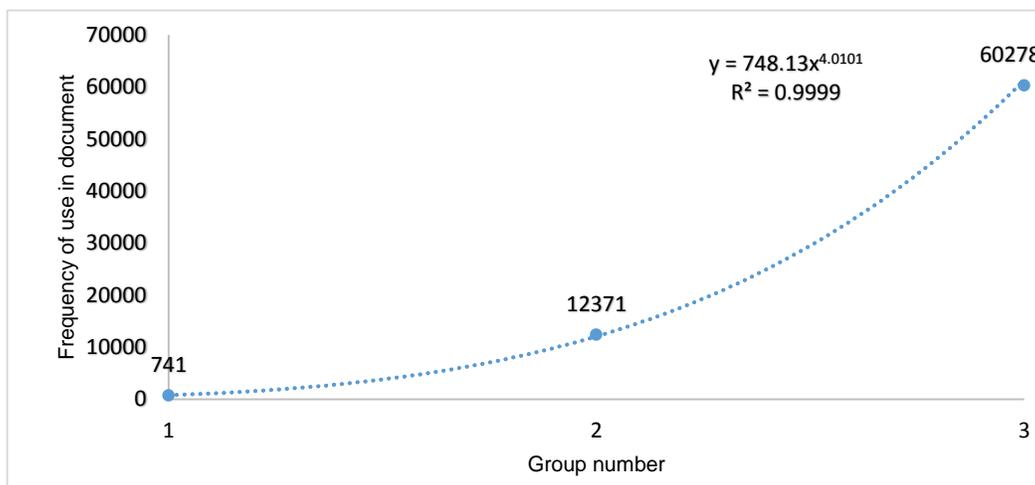
- A total of 53806 papers on bioinformatics are retrieved that were published between 1980 and 2018. The data are stored in the tab-delimited (Win, UTF 8) file format, each of which contains 500 records. After the data extraction, there are a total number of 108 files containing information on the articles. Next, all files of records are stored as a single combination text file with notepad.exe, leading to the creation of the first dataset.
- Keywords unification is carried out after the data collection has been completed. Unifying is performed twice: once by utilizing the software tool of Ravar PreMap, and once based on bioinformatics experts' opinions and the use of MeSh. The entire process of unification in present study includes changing the form of terms from the plural to singular, excluding nonsense words, excluding stoplists, and excluding countries' names as well as numbers and statistical terms. After unification completion, every keyword is classified either into major keyword (i.e. core), the minor keyword (i.e. close to the core), or more minor keyword (i.e. away from the core) based on Bradford's Law.
- Every keyword of the core group is selected for performing thematic analysis. Using the software of BibExcel, 101 X 101 symmetric matrices of the co-occurrence word are created for those keywords with a repetition of at least 100 times.
- In the clustering of the obtained keywords, the SPSS software is used. Next, a total of 7 thematic clusters are generated by the use of Ward's hierarchical clustering method.
- In the last step, the measures of degree, betweenness, and closeness centrality are calculated using the UCINET and NetDraw software programs. An ecological information map of the subjects obtained from the co-word analysis of bioinformatics articles is constructed by the use of techniques for social networking analysis.

## Finding

### 1. Major, Minor, and the More Minor Subject Infocenose

The results obtained with the methods described in previous sections revealed that every subject in three groups of core, close to the core, and away from the core is calculated. This is shown in Figure 2 where the fitting distribution of thematic infocenose in bioinformatics is seen to follow Bradford's Law. To identify the most used words (i.e., the most popular fields of bioinformatics), with the assumption that keywords are the indexes to indicate a subject and content, Bradford's Law was used. In an article published by Bradford (1934), the distribution of articles in the same subject-matter is discussed. As stated by the author, there is a high-rate of manuscripts in the

same subject-matter only in a few publications. Based on such results, the collection of keywords can be classified as core, close to the core, and away from the core groups. The number of keywords from each group was calculated for the current investigation and resulted in 741 items for major subjects (core), 12374 for minor subjects (close to the core), and 60278 for more minor subjects (away from the core). The analysis in the present study was conducted based on the obtained subjects from the major group of keywords as the most popular subjects of bioinformatics.



**Figure 2. Fitting distribution of thematic infocnose in bioinformatics articles published 1980-2018 based on Bradford's Law**

As was already mentioned, a list of the major subjects contains 741 keywords. The most used keywords with high-frequency of occurrence are shown in Table 1 the list is sorted from the highest to the lowest occurrence. This Table represents only 40 words occurring 110 times.

**Table 1. Core thematic infocnose of bioinformatics papers published in the years 1980-2018**

No.	Keyword	Frequency	No.	Keyword	Frequency
1	Bioinformatics	6693	21	miRNA	403
2	MicroRNAs	1246	22	Breast Cancer	384
3	Proteomics	1119	23	Machine Learning	374
4	Medical Informatics	1110	24	Cancer	351
5	Computational Biology	1079	25	Transcriptome	348
6	Microarrays	943	26	Comparative Genomics	321
7	Gene Expression	895	27	Evolution	318
8	Health Information Technology	703	28	RNA-SEQ	312
9	Next-Generation Sequencing	679	29	Invasion	304
10	Genomics	634	30	Differentially Expressed Genes	298
11	Bioinformatics Analysis	603	31	Gastric Cancer	294
12	Biomarkers	589	32	Transcription Factors	286

13	Mass Spectrometry	556	33	Hepatocellular Carcinoma	278
14	Apoptosis	492	34	Colorectal Cancer	262
15	Databases	481	35	Metastasis	249
16	Proliferation	476	36	Genetics	242
17	Electronic Health Records	470	37	Prognosis	240
18	Health Informatics	443	38	Classification	236
19	Data Mining	436	39	Transcriptomics	227
20	Systems Biology	404	40	Inflammation	224

Table 2 reports result for the frequency of usage and lifespan of the emerging thematic infocenose of papers in bioinformatics. In other words, a low amount of keyword searching implies that there is a novelty concerning the selection of that keyword.

**Table 2. Emerging thematic infocenose of bioinformatics papers published in the years 1980-2018**

Keywords	First Year Presence	Last Year Presence	Frequency	Keyword presence life
LNCRNA	2013	2018	80	6
Big Data	2012	2018	76	7
Differentially Expressed Genes	2011	2018	264	8
Osteosarcoma	2011	2018	70	8
Next-Generation Sequencing	2009	2018	326	10
RNA-SEQ	2009	2018	249	10
Protein-Protein Interaction Network	2009	2018	149	10
High-Throughput Sequencing	2009	2018	124	10
Metagenomics	2008	2018	137	11
ITRAQ	2008	2018	121	11
Health Information Technology	2006	2018	516	13
Proliferation	2006	2018	474	13
Migration	2006	2018	187	13
Cell Proliferation	2006	2018	166	13
MicroRNAs	2006	2019	195	14
Colorectal Cancer	2005	2018	154	14
Alzheimer's Disease	2005	2018	73	14
miRNA	2005	2019	279	15
Prognosis	2004	2018	239	15
Gene Ontology	2004	2018	176	15
Prostate Cancer	2004	2018	120	15

## 2. Hierarchical Clustering of Thematic Infocnose of Bioinformatics Papers

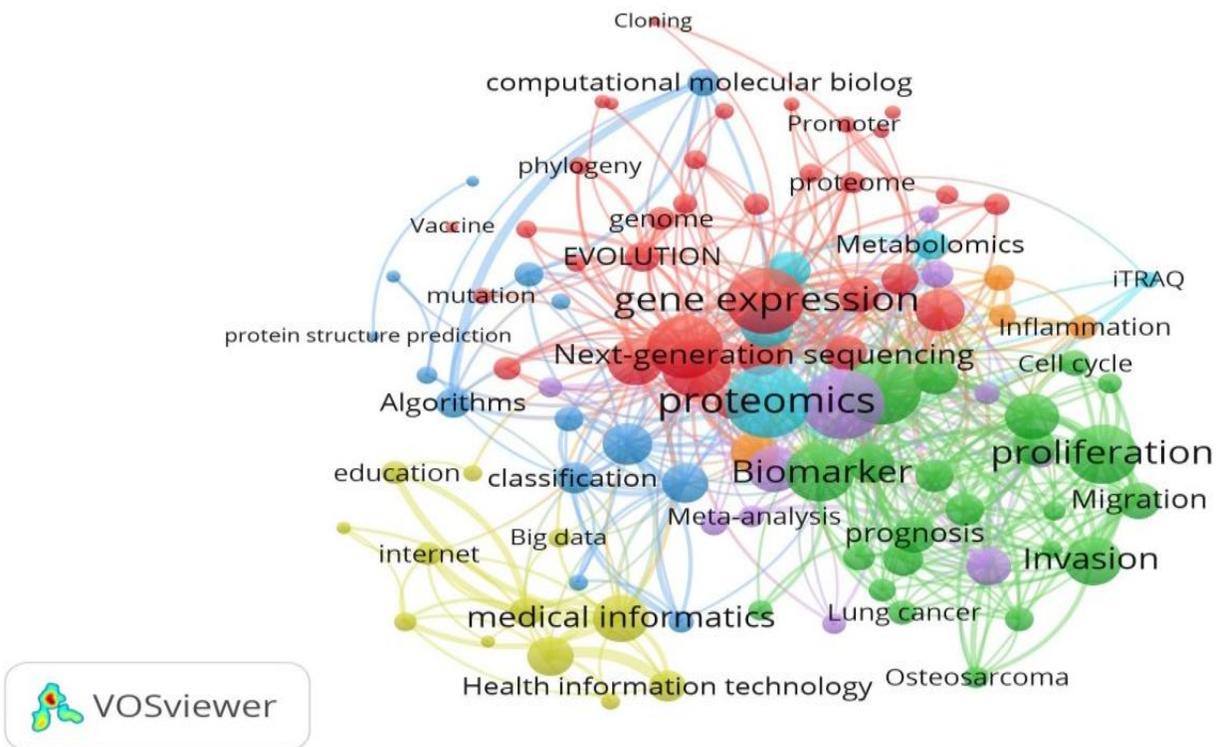
Table 3 displays the clustering results obtained using the analysis of data. From Table 3, it is evident that 7 clusters were generated. The topics from cluster 1 and those from cluster 2 were the most popular subjects with a high frequency of usage, playing a more important role. The researchers in this field mainly focus on the thematic infocnose of clusters 1 and 2.

**Table 3. Clustering results of the thematic infocnose of bioinformatics papers published in the years 1980-2018**

Cluster	Number of Keywords	Keywords in each cluster
1	33	Alternative splicing; Arabidopsis thaliana; Arabidopsis; Bioinformatics analysis; Cloning; Comparative genomics; computational biology; Database; Evolution; functional genomics; gene expression; gene regulation; genetics; genome; genomics; high-throughput sequencing; Metagenomics; mitochondria; mutation; Next-generation sequencing; Phylogenetic analysis; phylogenetics; phylogeny; polymorphism; Promoter; proteome; RNA-Seq; SNP; software; transcription factor; Transcription; Transcriptome; Vaccine
2	23	apoptosis; Biomarker; breast cancer; Cell cycle; cell proliferation; colorectal cancer; diagnosis; gastric cancer; Glioma; Hepatocellular carcinoma; Invasion; lncRNA; Long non-coding RNA; Lung cancer; Metastasis; MicroRNA; Migration; miRNA; Osteosarcoma; ovarian cancer; prognosis; proliferation; Prostate cancer
3	14	Algorithms; classification; Clustering; computational molecular biology; data mining; feature selection; homology modeling; Machine learning; Prediction; protein structure prediction; protein structure; sequence analysis; structural bioinformatics; Support vector machine
4	12	Big data; clinical informatics; education; electronic health records; health informatics; Health information technology; informatics; internet; Medical Informatics Applications; medical informatics; public health informatics; telemedicine
5	11	Alzheimer's disease; Data integration; differentially expressed genes; Gene expression profiling; Gene ontology; Meta-analysis; microarray; Pathway analysis; protein-protein interaction network; Protein-protein interaction; Systems biology
6	5	iTRAQ; mass spectrometry; Metabolomics; proteomics; Transcriptomics
7	4	Cancer; DNA methylation; Epigenetics; Inflammation

Figure 3 presents the ecological information map of topics in bioinformatics study. The map was constructed from the obtained thematic clusters. A total of 7 clusters are depicted in Figure 1, each of which displays a particular color. Every cluster was labeled with the name of its keyword with high-frequency of occurrence in order to develop a classification of clusters into thematic categorizations. It can be seen that the biggest cluster of the thematic map was yielded by the topics of clusters 1 and 2, and other smaller clusters were given other topics of other clusters.

The size of the biggest cluster is particularly relevant to the importance of the thematic infocenose which arises from their high frequency of usage in bioinformatics articles. Each circle in Figure 1 indicates a keyword. And the lines in this figure represent relation to each keyword's relation.



**Figure 3. Ecological information map of topics in bioinformatics articles published during 1980-2018**

### 3. Values for Degree, Betweenness, and Close-ness Centrality of the Thematic Infocenose in Bioinformatics

The values for degree centrality of the thematic infocenose with high popularity in bioinformatics articles published from 1980 to 2018 are reported in Table 4.

**Table 4. A total of 20 thematic infocenose in bioinformatics articles published from 1980 to 2018 with highest degree centrality**

No	Keywords	Degree centrality	Keywords	Closeness centrality	Keywords	Betweenness centrality
1	Proteomics	9.861	Computational Biology	84.167	Computational Biology	3.001
2	Microarray	9.267	Genomics	83.471	Genomics	2.877
3	MicroRNA	8.525	Proteomics	82.787	Database	2.679

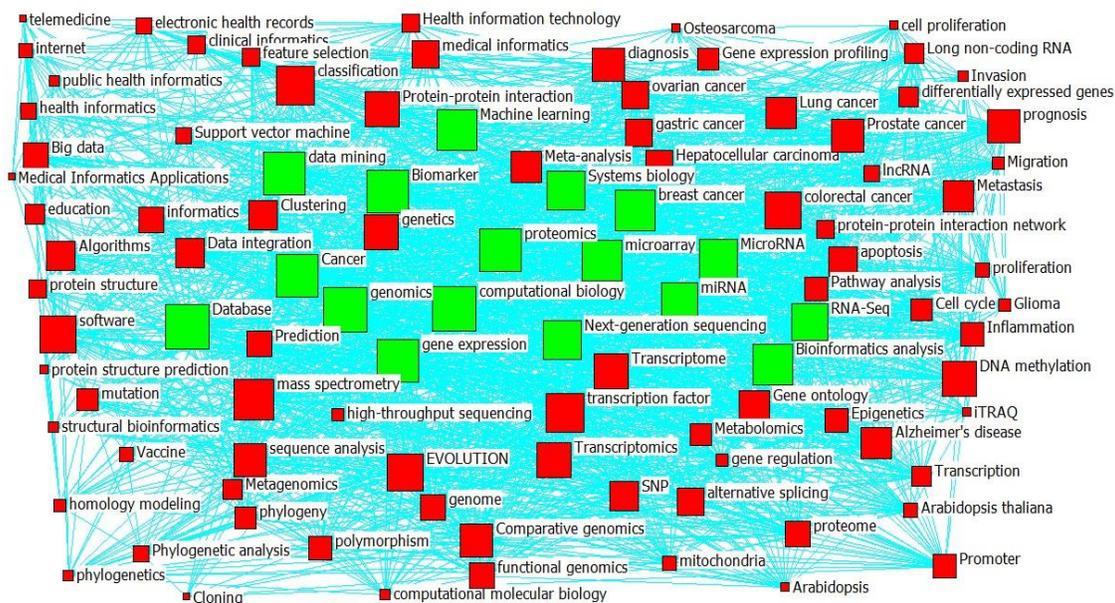
4	Gene Expression	8.436	Gene Expression	82.114	Proteomics	2.363
5	Genomics	7.950	Microarray	80.159	Data Mining	2.320
6	Proliferation	7.059	Database	80.159	Cancer	1.960
7	Computational Biology	6.762	Biomarker	79.528	Gene Expression	1.877
8	Biomarker	6.703	Cancer	78.906	Biomarker	1.666
9	Invasion	5.109	MicroRNA	77.692	Microarray	1.577
10	Database	4.752	Next-generation sequencing	75.940	Machine Learning	1.406
11	Medical Informatics	4.743	data mining	75.373	Breast Cancer	1.379
12	Systems Biology	4.604	RNA-Seq	74.815	Bioinformatics Analysis	1.366
13	Apoptosis	4.535	breast cancer	74.265	Mass Spectrometry	1.279
14	Mass Spectrometry	4.317	Machine learning	74.265	Classification	1.249
15	Bioinformatics Analysis	3.911	miRNA	73.188	Next-generation sequencing	1.193
16	Data Mining	3.842	Systems biology	72.662	MicroRNA	1.181
17	Next-Generation Sequencing	3.673	Bioinformatics analysis	72.662	Systems biology	1.107
18	Cancer	3.644	Mass Spectrometry	72.143	Transcription Factor	1.040
19	Machine Learning	3.614	Transcription Factor	71.631	RNA-Seq	1.015
20	Breast Cancer	3.545	Genetics	71.631	Colorectal Cancer	1.013

#### 4. Mapping Ecological Information Thematic Infocense in Bioinformatics Based on the Measures of Centrality

For a better analysis of co-occurrence word of articles in bioinformatics, the most-used keywords with a nonzero value of the closeness centrality were identified and their ecological information map was constructed. Closeness centrality in the sense used here is the keywords with the shortest distance of connection (minimum number of links) between them (Soheili & Osareh 2013).



Figure 5 shows the measure of degree closeness of the ecological information map of the bioinformatics articles. As can be seen from this figure, such thematic infocenose as *microarray*, *gene expression*, *proteomics*, *genomics*, and *computational biology* (as indicated by green color) are the most influential keywords based on the index of degree centrality. Other parts of the thematic map are related to the other subjects based on their importance ratings.



**Figure 6. Ecological information of the thematic infocenose in bioinformatics articles published from 1980 to 2018 based on the measure of betweenness centrality**

Figure 6 shows the measure of betweenness centrality of the ecological information map of the bioinformatics articles. As can be seen from this figure, such thematic infocenose as *data mining*, *proteomics*, *genomics*, *database*, and *computational biology* (as indicated by green color) are the most influential keywords based on the index of betweenness centrality. Other parts of the thematic map are related to the other subjects based on their importance ratings. In both Figures 3 and 4 (i.e. betweenness and closeness centrality), the thematic infocenose at higher importance levels are distinguished and shown by green color.

## Discussion

The results indicated that in infoecology, bioinformatics, 1) information diversity, 2) thematic evolution of information in the scientific productions of researchers in this field, 3) increased complexity of bioinformatics topics, and 4) the use of such topics demonstrate the change of the attitude of researchers toward this thematic field over time and the appearance of new thematic infocenoses in infotope. The data analysis led to the formation of seven thematic clusters in percent study. With the help of centrality indexes, it was also revealed that thematic infocenoses

such as proteomics, microarray, genomics, gene expression, computational biology, database, and data mining were the main thematic infoceneses of this field. The use of co-word analysis and social media analysis techniques is effective in determining the infoecology of different thematic infoceneses. Using infoecological thematic maps in bioinformatics, the thematic relationships between the studies conducted in this field can be analyzed. In line with present study, Ding, Chowdhury & Foo (2001) showed that the co-word analysis method was an appropriate approach to identify thematic patterns and trends in a scientific field at different time intervals.

According to the results, based on the calculation of the presence time of thematic infoceneses by the first and last presences in studies, infoceneses such as LNCRNA, big data, differentially expressed genes, Osteosarcoma, next-generation sequencing, RNA-SEQ, high-protein-protein interaction network, throughout sequencing, metagenomics, and ITRAQ were identified as emerging thematic infoceneses that appeared in bioinformatics studies during 2009-2018. The results also demonstrated that thematic infoceneses such as alternative splicing, *Arabidopsis thaliana*, *Arabidopsis*, bioinformatics analysis, cloning, comparative genomics, and computational biology had been the most used infoceneses in bioinformatics studies. Thus, they were identified as the most used infoceneses. In this regard, compared to the present study's results, Sedighi (2015) showed that concepts such as information science, library, bibliometrics analysis, innovation, and text mining were reported as the most internationally used topics in informatics. The results of the drawing of co-word maps in different time periods demonstrated changes and stabilities in informatics-related concepts and words. Some words, such as bibliometrics analysis, have been present in the entire studied years, while some others disappeared over time. Bosanac, Matešić and Tolić (2009) indicated that new concepts in the network were independently used as replacements for similar concepts. In this respect, Song, Heo & Kim (2014) demonstrated that new topics emerged and developed during 2000-2008, while thematic topics were integrated and had closer relationships with each other during 2008-2011. This trend was confirmed by cooperation patterns over time. Moreover, in line with the present study, Zarei and Babolhavaeji (2016) showed that the infoecology of the thematic topics of knowledge and information science were influenced by their habitat and move toward evolving and interacting with the environment and different related topics, particularly information and communication technology (ICT). In line with the present study, Muñoz-Leiva, Viedma-del-Jesús, Sánchez-Fernández and López-Herrera (2012) used the co-word analysis technique to identify the main topics. They also employed thematic map-drawing techniques to demonstrate relationships between the main concepts. They determined attitude toward consumer satisfaction and food products based on the rise of their centrality in the recent five years. Overall, their results indicated that food products and consumer understanding, attitude, satisfaction, and trust in the main countries were the most important studies published on consumer behavior.

## Conclusion

The present study attempted to identify the thematic infoceneses of bioinformatics, infoecology, and the infoecological maps of this field's topics by using the co-word and social network analysis techniques. The analysis results demonstrated that the infoecology of bioinformatics studies dynamically developed over time. The continuous production of scientific studies in this field, the same as other scientific fields, creates continuous changes in the infoecology of this field. Considering the extensiveness and interdisciplinarity of this field, it continues to interact with other sciences. Some bioinformatics fields are more considered, considering the requirements of scientific communities and countries, and they are considered as emerging topics. The present study's results also demonstrated that thematic infoecological maps could contribute to understanding the existing knowledge status and guiding scientific policies in this field. The drawn maps provide a clear representation of research topics in bioinformatics and the relationships between different topics. These maps represent changes and stabilities in bioinformatics-related topics and words over time. Moreover, new concepts have appeared as recombination of the existing words in interaction with new evolutions and technologies. Overall, the results of the present study and of similar studies can respond to these questions: What topics are considered by the scientific community? What are the different topics and sub-topics of scientific fields and how have they evolved?

Infoecology attempts to investigate the relationships between ecosystems in information environments and the relationships of information elements with such environments by finding their patterns, processes, and relationships. The same as the ecology of nature and wildlife that investigates processes and relationships governing natural and animal systems, infoecology can investigate patterns and relationships governing information environments (Burgin and Zhong 2018).

Infoecology can provide a new framework for interpreting complicated relationships between organizations, technologies, and information, or between information elements in information environments. In fact, infoecology is clearly capable of identifying the knowledge formed in a field and the relationships between knowledge points, indicating the thematic features of that field. This includes using methods, models, algorithms, and techniques by which researchers can study the collected knowledge and reveal distributed thematic fields, emerging topics, hot topics, and any growing trends in a field. The results of ecological studies conducted in information environments significantly contribute to investigating and understanding the nature and statuses of conducted studies, identifying the research fields of previous studies, modeling study trends, indicating considerable advances, representing the growth and evolution of current studies, and predicting the thematic patterns of future studies (Wang, Guo, Yang, Chen & Zhang 2017).

Studies such as the present study, which investigate the infoecology of a scientific field, are very valuable in identifying the scientific status of that field. Considering that present study used Web of Science (WoS) Database from Clarivate Analytics to obtain resources, it is recommended to conduct a study in this field by other accurate analysis methods on international databases, such as Scopus and MEDLINE. The comparison of such a study to those of the present study would yield valuable results.

## Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments that helped to improve the paper.

## References

- Apostolato, I. A. (2013). An overview of software applications for social network analysis. *International Review of Social Research*. 3(3), 71-77. <https://doi.org/10.1515/irsr-2013-0023>
- Asemi, A., & Ko, A. (2019). An investigation on infoecology of smart manufacturing in Scopus. *16<sup>th</sup> OGIK2019 National Conference on Economic Informatics: New IT Challenges for Economy and Society*. November 8-9, Milton Friedman University, Budapest.
- Bosanac, S., Matešić, M., & Tolić, N. (2009). Telling the future of information sciences: co-word analysis of keywords in scientific literature produced at the department of information sciences in Zagreb. In *Future 2009 Digital Resources and Knowledge Sharing*, 737- 746. Retrieved January 15, 2020, from <http://darhiv.ffzg.unizg.hr/id/eprint/8418>
- Chen, Y. P. P. (2005). *Bioinformatics Technologies*. Berlin Heidelberg: Springer-Verlag.
- Ding, Y. , Chowdhury, G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management*. 37 (6), 817-842. [https://doi.org/10.1016/S0306-4573\(00\)00051-0](https://doi.org/10.1016/S0306-4573(00)00051-0)
- Eryomin, A. L. (1998). Information ecology: a viewpoint. *International Journal of Environmental Studies*. 54(3-4), 241-253. <https://doi.org/10.1080/00207239808711157>
- Feather, J., & Sturges, P. (2003). *International encyclopedia of information and library science*. Routledge. <https://doi.org/10.4324/9780203403303>
- Jiang, R. , Zhang, X. , & Zhang, M. Q. (2013). Basics of bioinformatics. Lecture Notes of the Graduate Summer School on Bioinformatics of China. Berlin: Springer-Verlag. <https://doi.org/10.1007/978-3-642-38951-1>
- Merriam-Webster's collegiate dictionary (1993). Springfield, MA: Merriam-Webster.

- Muñoz-Leiva, F., Viedma-del-Jesús, M. I., Sánchez-Fernández, J., & López-Herrera, A. G. (2012). An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective. *Quality & Quantity*, 46, 1077-1095. <https://doi.org/10.1007/s11135-011-9565-3>
- Nardi, B. A., & O'Day, V. L. (1999). *Information Ecology: Using Technology with Heart*. Cambridge: MIT Press.
- Nicolaisen, J. , & Hjørland, B. (2007). Practical potentials of bradford's law: a critical examination of the received view. *Journal of Documentation*, 63 (3), 359-377. <https://doi.org/10.1108/00220410710743298>
- Sedighi, M. (2015). Using co-word analysis method in mapping of the structure of scientific fields (Case study: the field of informetrics). *Iranian Journal of Information Processing & Management*.30 (2), 373-396. Retrieved January 15, 2020, from <http://jipm.irandoc.ac.ir/article-1-2583-en.html>
- Soheili, F., & Osareh, F. (2013). Concepts of centrality and density in scientific and social networks. *National Studies on Librarianship and Information Organization*, 24 (3), 92-108.
- Song, M. , Heo, G. E. , & Kim, S. Y. (2014). Analyzing topic evolution in bioinformatics: investigation of dynamics of the field with conference data in DBLP. *Scientometrics*, 101(1), 397-428. <https://doi.org/10.1007/s11192-014-1246-2>
- Wang, X. , Guo, Y. , Yang, M. , Chen, Y. , & Zhang, W. (2017). Information ecology research: past, present, and future. *Information Technology and Management*, 18(1), 27-39. <https://doi.org/10.1007/s10799-015-0219-3>
- Waring, R. H. (1989). Ecosystems: fluxes of matter and energy. In J. M. Cherrett (ed.). *Ecological Concepts*. British Ecological Society Symposium 29. Oxford: Blackwell.
- Zarei, A., & Babolhavaeji, F. (2016). Information ecology of subject categories of knowledge and information science (KIS). *Library and Information Sciences*, 19 (2), 79-100.

---

***Bibliographic information of this paper for citing:***

Kiani, M., Asemi, A., CheshmehSohrabi, M., & Shabani, A. (2020). "Information ecology of bioinformatic in Web of Science with emphasizing on articles thematic interaction." *Webology*, 17(1), Article 215. Available at: <http://www.webology.org/2020/v17n1/a215.pdf>

---

Copyright © 2020, Masoome Kiani, [Asefeh Asemi](#), [Mozafar CheshmehSohrabi](#)  
and [Ahmad Shabani](#).