


## Scientometric Study of Research in Information Retrieval in Medical Sciences

Masoud Mohammadi<sup>1,2</sup>, Gholamreza Roshandel<sup>2</sup>, Seyed Javad Ghazimirsaeid<sup>1</sup>, Marzieh Zarinbal<sup>3</sup>, MolukoSadat Hosseini Beheshti<sup>3</sup>, Fatemeh Sheikhshoei<sup>1\*</sup> 

Received: 12 Mar 2021

Published: 16 Jun 2022

### Abstract

**Background:** Mapping scientific trends is one of the most important missions of scientometric research for effective research. The main goal of this paper was to visualize and draw the intellectual and cognitive structures of information retrieval (IR) in the medical sciences using science mapping.

**Methods:** In this cross-sectional scientometric study, we recruited all documents indexed in the Web of Science database with the topic of storing and retrieval of information in medical sciences. To analyze the results, 3 software, SciMAT-v1.1.04, VOSviewer-v1.6.14, CitNetExplorer\_v1.0.0, were used.

**Results:** Our results showed that most scientific productions in this field fall into 2 categories: (1) effective methods of organizing information and (2) application and operation of the IR system in the process of intelligent questioning and answering, and analyzing information behaviors of physicians and health professionals. The results showed that the similarity index increased over time from 0.43 to 0.71. Analysis of the findings showed that similarity measures, expert systems, concepts, experience, answers, and multimodel IR clusters were considered as mature and completely centralized clusters in the first quarter of the strategic chart.

**Conclusion:** Because of the dramatic approximation of the vocabulary used by researchers and a relative slowdown in the growth rate of the subject's domain in the last decade, it seems necessary to pay attention to the expansion of the fields of IR and the application of its concepts in medical information sciences. Also, it can be recommended that designers of IR systems and techniques in medical information sciences pay more attention to human factors attentively to develop new technologies and tools.

**Keywords:** Medical And Health Science, Information Retrieval, Citation Network, Scientometric, Research Trends

**Conflicts of Interest:** None declared

**Funding:** None

**\*This work has been published under CC BY-NC-SA 1.0 license.**

Copyright© Iran University of Medical Sciences

**Cite this article as:** Mohammadi M, Roshandel G, Ghazimirsaeid SJ, Zarinbal M, Hosseini Beheshti M, Sheikhshoei F. Scientometric Study of Research in Information Retrieval in Medical Sciences. *Med J Islam Repub Iran*. 2022 (16 Jun);36:65. <https://doi.org/10.47176/mjiri.36.65>

### Introduction

In recent years, there has been a drastic change in the way information is been disseminated in the scientific world. Large amounts of scientific evidence make it more difficult for researchers to access the information they need. Therefore, information retrieval (IR) technologies

have been developed to answer the information needs of researchers and scholars and to help them retrieve the most appropriate and diverse scientific resources to the questions provided to fully meet their information needs (1-3).

**Corresponding author:** Dr Fatemeh Sheikhshoei, [fashoei@sina.tums.ac.ir](mailto:fashoei@sina.tums.ac.ir)

<sup>1</sup> Department of Medical Library and Information Sciences, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

<sup>2</sup> Research Center of Gastroenterology and Hepatology, Golestan University of Medical Sciences, Gorgan, Iran

<sup>3</sup> Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran

#### ↑What is “already known” in this topic:

In the field of medical sciences, research on information retrieval was quite diversified and concentrated on the application of data analysis methodologies; nevertheless, it appears that to date, no scientometric study has been conducted to examine research in this subject.

#### →What this article adds:

In the present study, we visualize and draw the intellectual and cognitive structures of IR in the medical sciences using science mapping and draw future perspectives in the highly variable and developing field of IR in medical sciences.

The tools created in the science of data retrieval are used to achieve the maximum content produced, both directly and indirectly. In the past decades, the use of language modeling, filtering, recommendation systems, and answering interactive questions has become the main area of research, and these researches seem to be focusing more on users, including modeling behavior, and fixing user interface has become more important. Nonetheless, information technology, information systems, and data retrieval have changed in ways that cannot even be imagined. These changes occurred so rapidly that it is difficult to predict what will happen in the next 20 years, making it difficult to understand the nature of scientific development in the field of information retrieval, especially in medical sciences (4, 5).

In this regard, the ability to detect subject trends, map out concepts, ideas, and issues in various scientific fields, and explain the status of citation nodes to identify subject areas of greatest interest to the scientific community has become increasingly important for logical reasons that will aid policymakers in achieving their objectives (6-8).

To achieve these goals, focusing on the previous paradigm and the intellectual base of a discipline that is reflected in its existing scientific productions can inform us about future research fronts, analysis of research frontiers, and development of unique research areas or topics. Overall, these connections represent the cognitive structure of the research area, which is usually done through science mapping methods (9, 10). Scientific mapping or bibliographic mapping focuses on determining areas of research in a scientific field to determine the cognitive structure and its evolution (11).

This was a scientometric study of scientific outputs in the field of information retrieval in medical sciences and aimed to provide a knowledge mapping domain in this field. In this regard, we conduct this research to determine the advantages and disadvantages of research in this field based on the identification of high-power and low-power citation nodes. In addition, the purpose of analyzing topics and keywords appearing in scientific resources is to determine topic trends. Finally, it is our research goal to provide correct viewpoints and evidence-based visions for

researchers in this field to determine the position of various research topics in the evolution and strategic themes.

## Methods

In this cross-sectional scientometric study, we recruited all documents indexed in the Web of Science database with the topic of storage and retrieval of information in medical sciences. To achieve maximum comprehensiveness, SCI-EXPANDED and SSCI collections were selected. The search for the documents took place on March 15, 2020. To illustrate the thematic process, all the documents available on the Web of Science database in the search field published from 1968 to 2020 were examined.

## Search Strategy

Searching for resources in the topic field was done with the following strategy: ((Retrieval OR storage\*) AND (information\* OR data\* OR system\* OR article\* OR research\* OR image\*)) AND (health\* OR Medic\*).

## Inclusion and Exclusion Criteria

This study examined all articles on information retrieval in the field of medical sciences. Therefore, the inclusion criteria were as follows: All research articles on the subject of information retrieval or data retrieval or data storage and retrieval systems or retrieval systems, or article retrieval or research retrieval, or image retrieval in the field of medicine and health. Exclusion criteria were as follows: (1) studies that were not research articles; (2) studies whose bibliographic information was not sufficient to obtain standard outputs.

Therefore, out of the total documents retrieved, 8404 articles were included in the study. A total of 4578 unrelated articles were excluded from the study. If after reviewing the abstract or the full text of an article it became clear that its subject was not directly related to information retrieval in medical sciences, it was removed as an irrelevant item. Finally, 3826 articles were analyzed (Fig. 1).

## Data Analysis

To analyze the results, 3 software, SciMAT-v1.1.04,

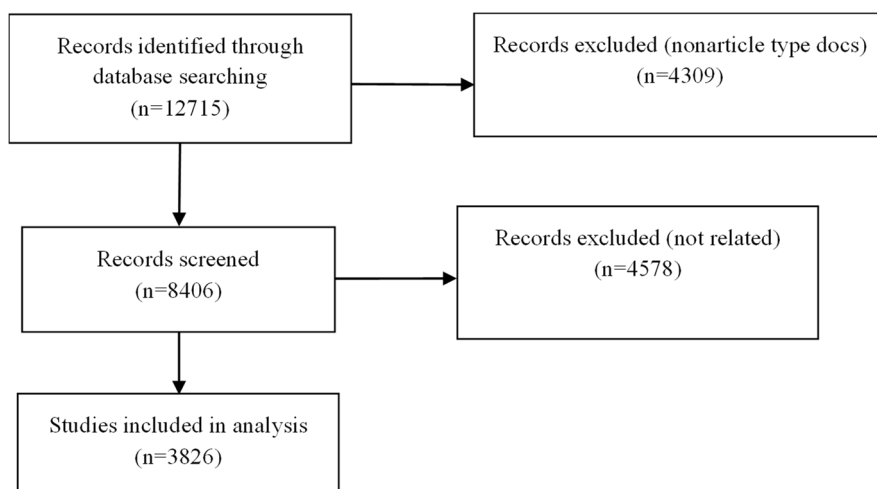


Fig. 1. The flow chart for excluding ineligible articles

Table 1. Features Applied in Each Software to Perform Analyses

Tool	Index	Value
CitNetExplorer	The base of citation analysis	citation sources, predecessors <sup>1</sup> , and successors <sup>2</sup>
	Minimum number of citation links a processor or successor	2
	The maximum distance at which the processor or successors may be located from marked publication	1
VOSviewer	Keywords	author's keywords and keyword plus
SciMAT	Minimum number of occurrence	20
	Repetitions for assigning to include each keyword	3
	Edge value minimum in network reduction	2
	Similarity scale in network normalization	association strength
	Clustering algorithm	simple centers algorithm
	Document mapper section	core mapper
	Similarity criteria	equivalence index

<sup>1</sup> publications cited by marked publications<sup>2</sup> publications citing marked publications

VOSviewer-v1.6.14, CitNetExplorer-v1.0.0, were used. SciMAT (Science Mapping Analysis Software Tool) is a scientific mapping software designed by the University of Granada and is available as an open-source (12). This software allows scientometric analysis based on bibliographic networks such as co-word, co-citation, author co-citation, journal co-citation, coauthor, bibliographic coupling, journal bibliographic coupling, and author bibliographic coupling (13). CitNetExplorer software was used in this study to cluster documents based on citation relationships and analyze results based on individual authors. This software is designed by Leiden University as a tool for illustrating and analyzing citation networks of scientific publications at the level of authors. VOSviewer software was used for thematic cluster analysis. VOSviewer is a software tool for creating and visualizing bibliographic networks. While CitNetExplorer is used to analyze a cluster at the level of separate documents, VOSviewer is used to analyze clustering at the entire level of articles. (13-17).

To extract bibliographic and citation information from documents in a readable format by the software used, the data is exported in the form of full records (covering author and author units, source journal titles, titles, keywords, and abstracts) and cited references in plain text format. Some considerations on how to configure software for analysis are provided in Table 1.

## Results

Citation analysis of the results was performed using CitNetExplorer software. This section's conclusions are based on grouping publications based on their citation relationship and assessing the clustering solutions at the individual publication level. Analyses show that there were 3661 citation links between 3826 scientific productions during this period. Table 2 provides an overview of citation links in 3 time periods. The highest citation links and relative publications are observed in the third 10-year period.

Table 2. Evolution of Citation Links

Block Period	Publications	Citation Links	CI/P*
1960-2000	462	113	0.24
2000-2010	1388	932	0.67
2010-2020	2211	1157	0.52

\* Citation Links/Publications

The chronological citation network is shown in Figure 2. By default, CitNetExplorer displays tags with the first author's last name when visualizing a citation network of documents. In this image, the circles symbolize the documents. The curved lines represent the citation relationships of each document (17).

The map above shows that the main and most cited articles were in 2 main themes. In the first theme (left), the main content of the resources was an effective method of organizing information. Most of the articles in this category deal with the methods of mining, anthologies, and their application and indexing of resources. As time goes on, the topics of the articles move from mining and its methods in retrieving information to anthologies and their application to the meaning of information. In this regard, Wilbur and Yang's article is considered as a basic article. They provided a new information-theoretical interpretation of term strength, reviewed some of its uses in focusing on the processing of documents for IR, and described new results obtained in document categorization (18).

In the second topic, the main content of the documents was the application and performance of IR systems in question and answering forms and the analysis of information behaviors of health professionals. Over time, the thematic content of documents has shifted from search, text browsing, and search tools to topics such as physicians' clinical answers and information-seeking behaviors. The harsh article in this category is considered as a basic article in which it discusses the use of IR systems by physicians to answer clinical questions and physician information behavior. The purpose of this article was to provide a conceptual framework and to apply the results of previous studies to this framework (19).

Thematic clustering of documents based on the CiteNetExplorer analysis is shown in Figure 3. After the analysis, the documents were categorized into 4 thematic clusters. Each cluster contained documents strongly related to each other. The results showed that a total of 136 documents were placed in these clusters and the core clusters were in the form of the following clusters: 48 (35%) in group 1 (blue), 36 (26%) in group 2 (green), 35 (25%) in group 3 (red), and 10 (7%) in group 4 (orange).

The thematic theme of the documents in blue clusters was "the analysis of physicians' information behavior, IR systems, EBM, and CDSSs" and "EHR and Medical Doc-

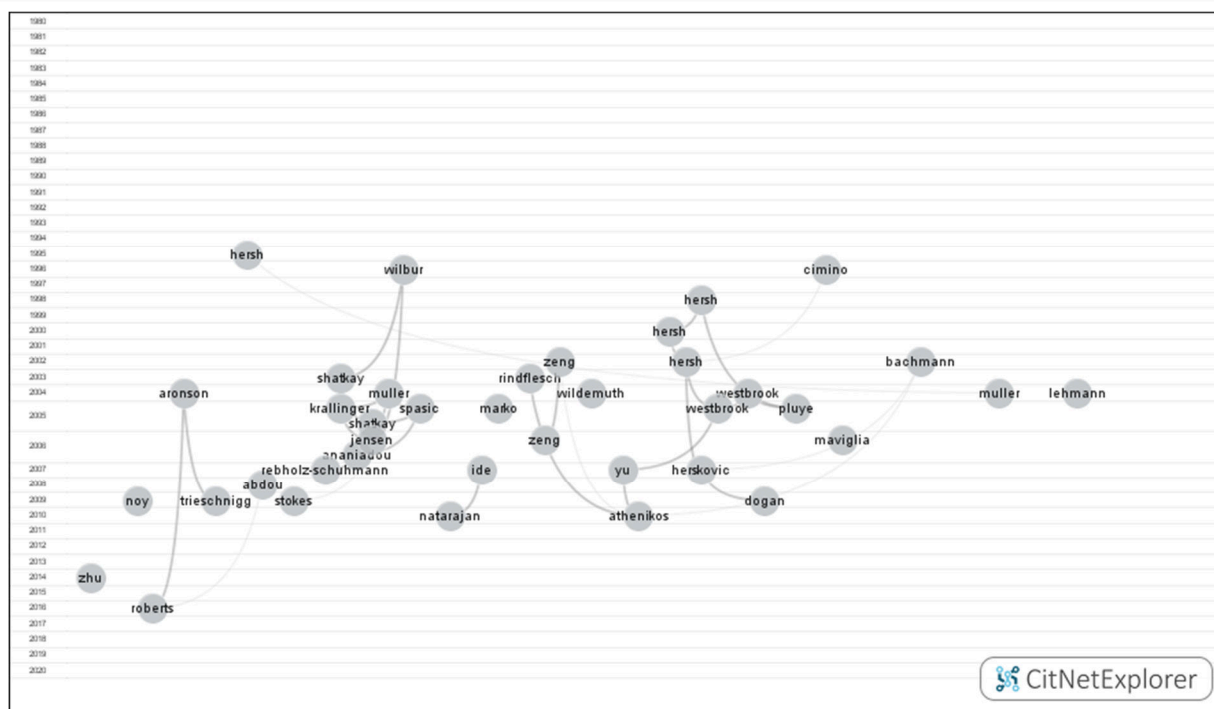


Fig. 2. Chronological citation network for research in IR in medical sciences

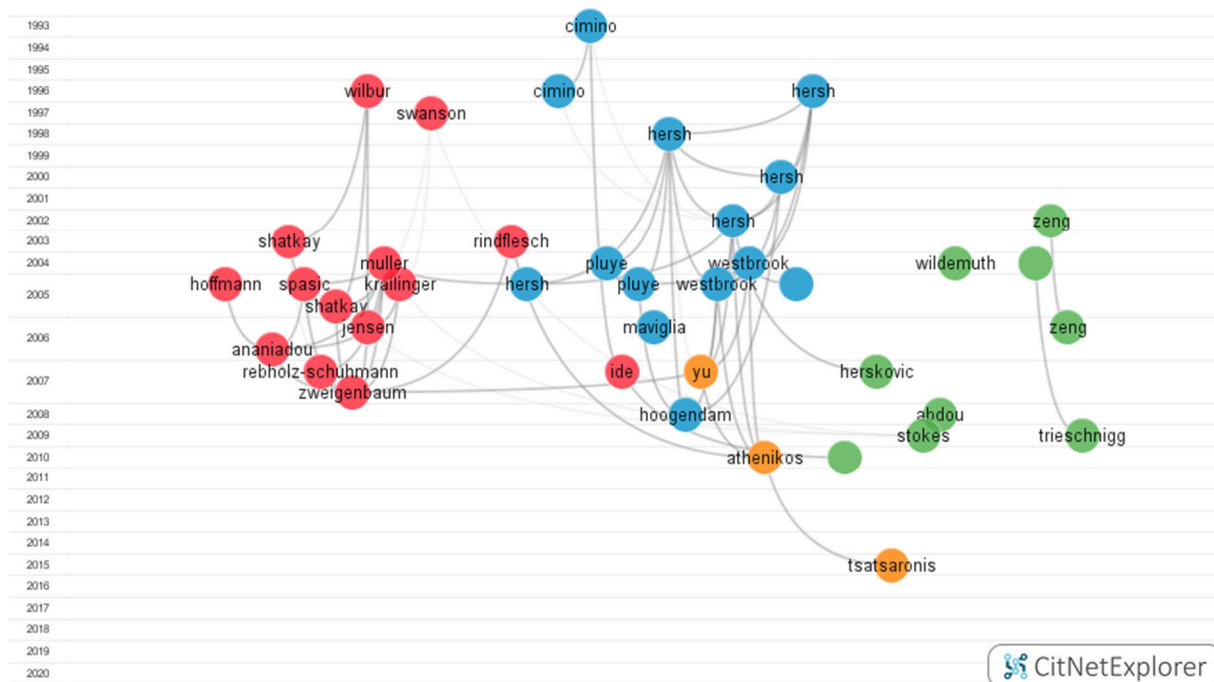


Fig. 3. Thematic clustering of documents based on CiteNetExplorer analysis

uments” were in the green cluster. Also, the thematic theme of the red cluster was “text mining and indexing” and the thematic theme of the orange cluster was “question answering systems.”

Also, as can be seen in the chart above, most of the core publications were published from 2000 and 2010, and this indicates the significant impact of the scientific activities within this decade on the scientific productions of the next decade. In other words, most of the core scientific prod-

ucts that have created the infrastructure for other IR research in medical sciences were produced from 2000 to 2010. As shown in Table 2, the citation link ratio of the scientific production of this period was higher than the number of its publications (0.67).

Topic networks were based on co-occurrence networks and term maps using VOSviewer software. This embodiment shows the most important terms in the publications belonging to a cluster and the correlational relationships of



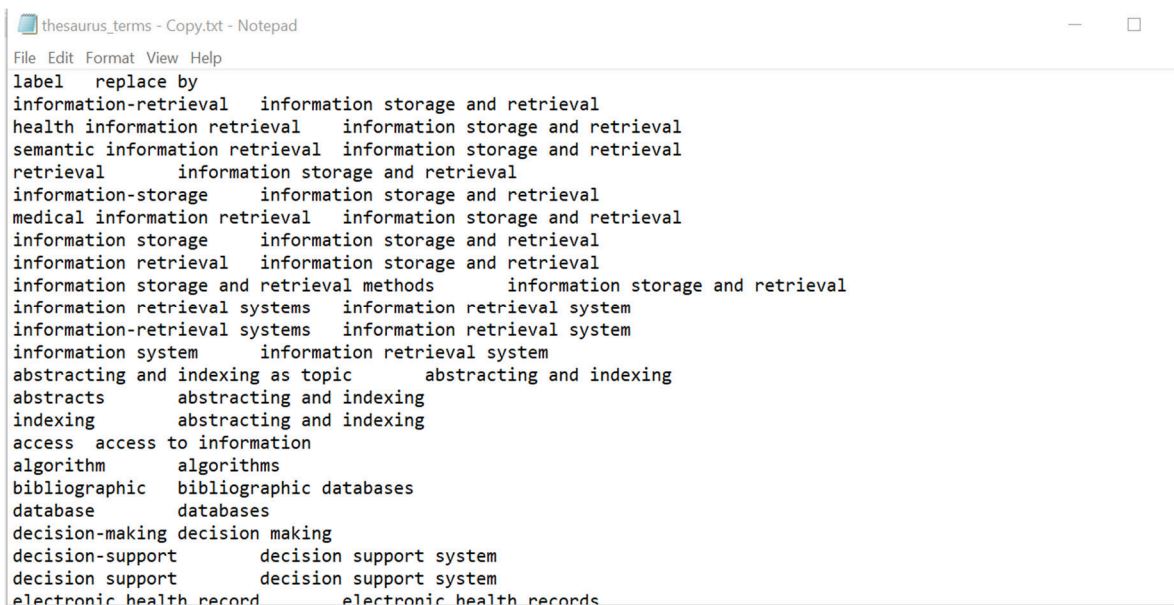


Fig. 4. A picture of designed terminology to be used in analyzing data by VOSviewer

these terms. In this section, the co-occurrence analysis of words for the analysis of thematic trends in the field of IR in medical sciences is examined.

One of the problems of this stage was the existence of different forms of writing, such as singular and plural forms, and synonyms of concepts for drawing lexical maps. Therefore, to unify the concepts and prevent the dispersion of the same concepts, the researchers first designed a specialized thesaurus in IR in medical sciences to be used in the analysis by VOSviewer (Supplementary 1). This is one of the specialized advantages of VOSviewer software analysis. Figure 4 shows a picture of designed terminology to use in analyzing data by VOSviewer.

The results of this section showed that the documents examined had a total of 10,783 keywords. In addition to the author's keywords, a "keyword plus" is provided on the web of science database to provide a more accurate overview of the summary of articles. Therefore, based on the researchers' experience, both options were selected as the criteria for selecting keywords for deeper analysis. For the meaningful drawing of knowledge maps, the minimum number of occurrence conditions was considered to be 20 for analysis, and under these conditions, 116 keywords were selected as frequent keywords for these articles. Then, to increase accuracy, irrelevant keywords such as "medicine" were removed from the selected keywords. In the end, 80 keywords remained. In all maps, we plotted the weight of the words based on the frequency of the events.

The placement of keywords in clusters and the distance between nodes was based on the simultaneous use of 2 or more similar keywords. The size of each circle in the cluster indicated the abundance of that word in that cluster (14, 20).

After drawing the clusters and examining the keywords, it was found that the analyzed documents were in the

themes of IR technologies and techniques (first cluster), information behaviors and CDSS systems (second cluster), indexing and knowledge representation tools (the third cluster), and the knowledge of searching for resources and topics related to databases (fourth cluster), and searching for information as placed on the web (fifth cluster). The first and second clusters had the highest number of keywords with 30 items, and after these clusters, the third clusters with 10, the fourth with 7 items, and the fifth cluster with 4 items.

In terms of all the 3 indicators of links, the total strength link, keyword occurrence, the order of importance of keywords in the 5 clusters are as follows: In the first cluster, the keywords of "information storage and retrieval," "IR system," "natural language processing," and "ontology"; in the second cluster, "knowledge," "models," and "electronic health record"; in the third cluster, "query expansion," "MeSH," "UMLS," and "terminology"; in the fourth cluster, "bibliographic databases," "bibliometric," "databases," and "literature searching" had the most important in their cluster (Fig. 5).

Table 3 provides detailed information on the keywords in each cluster, the number of links per keyword with other concepts, total strength link, and keyword occurrence. Links and the total strength link showed the number of links a keyword had with other keywords and the overall strength of the links a keyword had with other items for each individual term. There can be a link between any pair of items. A link is a relationship between 2 things. In other words, the numbers presented indicate the number of links between each item and other items; that is, the X keyword is related to several other keywords in terms of coincidence. Each link has a strength that is indicated by a positive numerical value, as the higher the value, the stronger the bond. The strength of a link indicates the number of documents in which the 2 terms occur together. Occur-

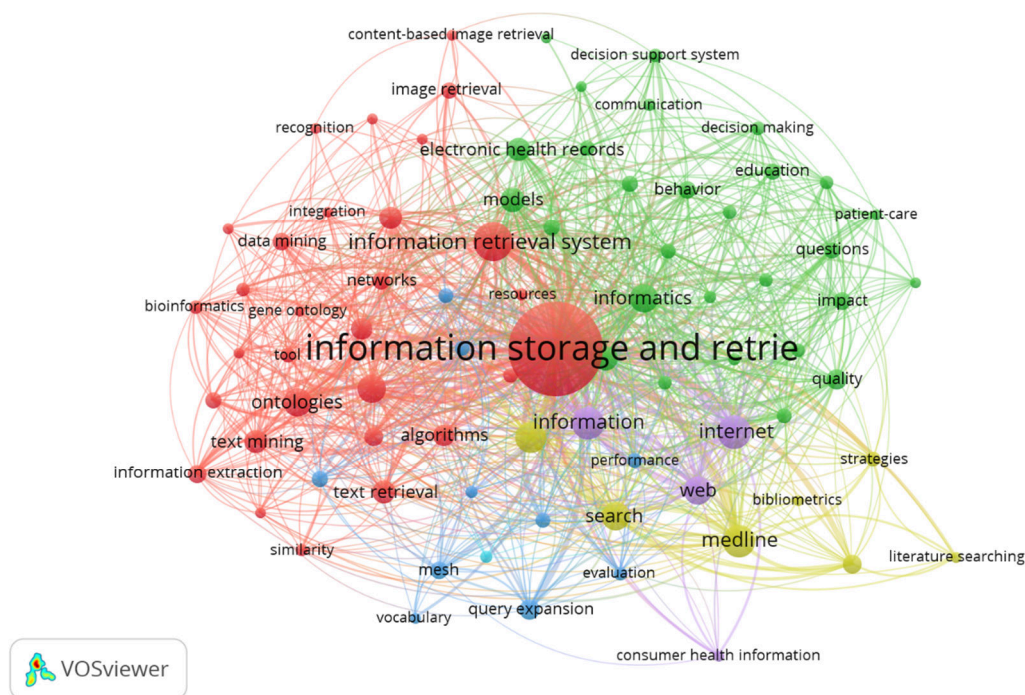


Fig. 5. Visualization of the 5 clusters created based on the analysis of keywords used in IR in medical sciences

rences show the number of documents in which a keyword appears.

Based on these results, the first cluster, "IR technologies and techniques" had the highest link (1176): total strength link (7265) and keyword occurrence (3228). Regarding the link index, the keywords of the fifth cluster "web IR" had the lowest number of links and coincidences with 216 items. However, the keywords of the third cluster with the total strength link equal to 1129 and the keyword occurrence equal to 420 had the lowest indicators.

We used SciMAT to draw a thematic strategic diagram in the field of IR in medical sciences. To do this, after entering the data into the software, 10,530 keywords were recovered. The reason for the difference with VOSviewer is that SciMAT only considers the author's words and not the keyword plus. Then, we cleared the keywords by removing the unrelated ones and replacing the synonyms. Next, 263 items (keywords that have been cleared) remained for analysis.

Figure 5 shows stability measures over 3 consecutive periods. The loops represent the periods and numbers inside each loop, indicating the number of keywords. The number of common keywords in both periods is shown via the horizontal arrows, and the similarity index for them is shown in parentheses. The upper-incoming arrow indicates the number of new keywords within a period and in the period but not in the next period (11).

The results of this section showed that the number of keywords increased significantly over time, and in the 2000-2010 period, compared with the period before 2000, it increased by 2.38 times. Similarly, the number of common keywords between subsets has increased from 94, between the period before 2000 and 2000 to 2010, to 224

during 2000-2010 and 2010-2020. The similarity index has grown over time from 0.43 to 0.71. This means that medical IR researchers have gradually pushed their words closer together. However, the findings revealed that the majority of new keywords ( $n = 91$  keywords) entered the literature and terminology of IR in the medical field during the 2000s and 2010s, indicating the growth of new concepts and dramatic changes in the development of thematic boundaries in this decade. However, between 2010 and 2020, the emergence of new keywords has decreased to nearly half ( $n = 57$  keywords), indicating a relative slowdown in its growth rate (Fig. 6).

Figure 7 shows a strategic chart of scientific topics in a chart. In this diagram, the centrality index is on the x-axis and the density index is on the y-axis. The strategic chart is used to determine and analyze the position of clusters and thematic concepts under each field and to describe the internal relationship and correlation from thematic clusters and the illustration of maturity and the coherence of thematic clusters. Also, in the strategic chart, centrality indicators are used to measure the relationship between one subject area and other thematic areas and the density. Centrality indicates the importance of an issue, and the larger the index, the more important the cluster among the existing issues. The density index indicates the strength of the bonds that connect words in a cluster (13, 21).

Using 2 indicators, centrality and density, the strategic chart is divided into 4 quarters. The topics in the upper right quarter (first quarter) are fully developed and are very important for the development of the main research structure in medical science. They are known as special themes because of their high centrality and density. The placement of the topics in this quarter means that they

Table 3. Thematic clusters in IR in medical sciences and detailed information of keywords based on the 3 attribute

Cluster number (color)	Keyword	Link	TSL*	KOc*	Cluster number (color)	Keyword	Link	TSL	KOc
1 (red)	Algorithms	47	164	83	2 (green)	Access to information	51	130	41
	Annotation	35	98	24		Behavior	40	143	57
	Architecture	33	61	21		Clinical question	44	130	35
	Big data	21	45	29		Communication	28	49	27
	Bioinformatics	30	98	37		Decision making	36	97	37
	Biomedical literature	28	71	22		Decision support systems	35	95	35
	Classification	56	241	100		Design	48	148	51
	Content-based image retrieval	16	32	23		Education	36	90	47
	Data mining	39	134	60		Electronic health record	50	219	97
	Image retrieval	38	118	45		Framework	53	147	74
	Gene ontology	32	74	22		Impact	46	148	56
	Information extraction	39	174	63		informatics	65	381	143
	Information retrieval system	76	723	279		Information management	29	81	29
	Information storage and retrieval	79	2845	1562		Information seeking behavior	24	53	20
	Integration	30	63	20		Information systems	34	97	37
	Machine learning	41	151	63		Knowledge	71	390	110
	Natural language processing	59	401	146		Management	44	128	44
	Networks	54	171	59		Medical records	22	71	25
	Ontologies	50	354	147		Memory	8	14	22
	Patterns	28	61	23		Models	61	261	109
	Recognition	20	36	21		Needs	37	125	37
	Resources	49	103	31		Patient care information	29	103	24
	Search engines	47	116	40		Quality	48	230	78
	Semantic web	49	212	85		Question	42	147	45
	Similarity	27	66	29		Relevance	36	97	35
	Text mining	50	254	93		seeking	41	151	44
	Text retrieval	58	258	58		Support	36	86	26
	Tools	45	141	43		Technology	36	86	32
	Total	1176	7265	3228		Total	1130	3897	1417
3 (blue)	Indexing and abstracting	40	133	40	4 (yellow)	Bibliographic databases	38	202	65
	Controlled vocabularies	35	81	26		Bibliometric	25	49	20
	Evaluation	31	92	33		Databases	68	529	178
	Language	42	115	41		Literature searching	14	62	23
	MeSH	52	160	58		Medline	62	513	185
	Performance	37	86	40		Search	65	419	142
	Query expansion	46	172	71		Strategies	35	125	41
	Terminology	46	127	42		Total	307	1899	654
	UMLS	36	111	49	5 (purple)	Consumer health information	21	71	21
	Vocabulary	26	52	20		Information	72	598	192
Total		391	1129	420		Internet	62	567	206
						Web	61	406	141
						Total	216	1642	560

\* Total Strength Link

\*\* Keyword Occurrence

have the most internal coherence and connection and are conceptually very close and related. Topics in the upper left quadrant (second quarter) are still coherent but decentralized, each of which consists of smaller specialized areas of science. Topics in the lower left quadrant (third quarter) have a low density and centrality, reflecting the emergence or the decrease in scientific disciplines. Topics in the lower right quarter (fourth quarter) are important in a research field but have not yet matured and have the potential to become major topics in the field (11, 21-23) (Fig. 7).

To draw a strategic diagram to explain the situation more accurately, a strategic diagram is presented based on the number of scientific productions and the index of citation to the scientific products of the field under study.

Based on the average of citation to scientific products, the largest clusters include “similarity measures (40.41 citations),” “mechanism” (39.37 citations),” and “barriers (34.82 citations).” In the similarity measures cluster, “similarity measures,” “distance nodes” with 11 documents were the largest nodes, followed by “sets” and “topic models” with 6 documents in the next ranks. In the mechanism, the cluster was “mechanism” nodes with 15 documents and “single-molecule magnet” with 3 documents. In the barriers cluster, there were “complexes” nodes with 14 documents and “barriers” with 4 documents.

Based on the number of documents, “medical informatics (1281 docs),” “experience” (51 docs),” and “expert systems (45 docs)” were the largest clusters. In the medical informatics cluster, “Medline search (224 docs),”

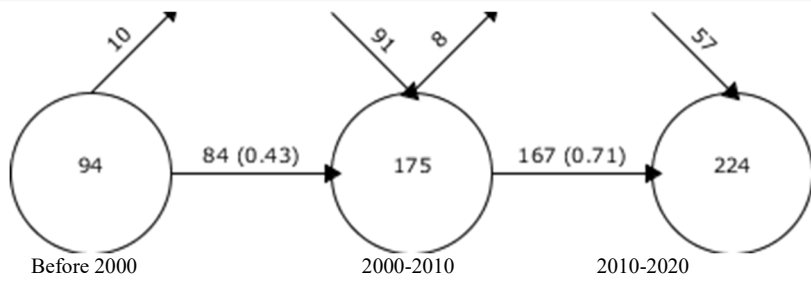


Fig. 6. The stability measures across the 3 consecutive periods in IR in medical sciences

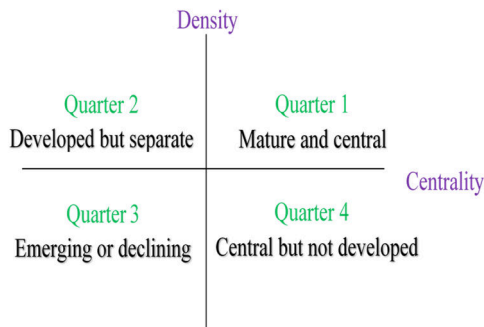


Fig. 7. Distribution of clusters in the strategic chart

“medical informatics (211 docs),” “database management systems (189 docs),” and “ontology (152 docs)” were the largest nodes. In the experience cluster, “methodology and experience (15 docs)” and “university library (12 docs)” nodes were the largest nodes. In expert system clusters, “expert systems (13 docs),” “conceptual graph (12 docs),” “interface (11 docs),” and “cased-based reasoning (11 docs)” were the largest nodes (Fig. 8).

Analysis of these findings showed that in the field of IR in medical sciences, clusters of similarity measures, expert systems, concepts, experience, answers, and multimodel

IR were in the first quarter of the strategic chart. Smartphone, hybrid, decision tree, RFID, and feasibility study clusters were in the second quarter. The third quarter included relational database clusters, mechanism, clinical information systems, medical terminologies, and barriers. The fourth quarter included health information exchanges, metadata, and medical informatics.

## Discussion

Examining the thematic areas of information retrieval in medical sciences and drawing its maps, is one of the most

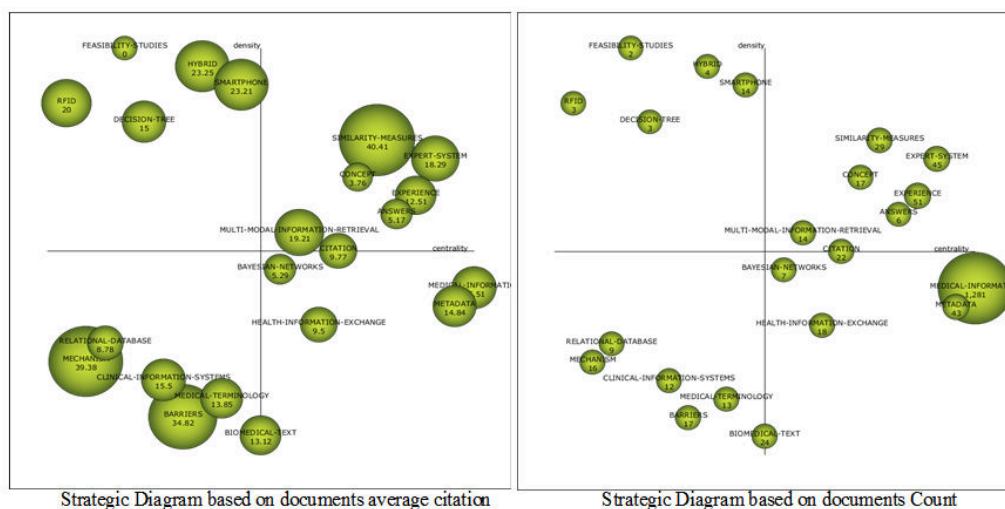


Fig. 8. Strategic diagram based on documents average citation and document count



essential methods for predicting ground research based on the past path. Thus, this study was performed to evaluate the evolution of research and mapping global knowledge domain in works of literature of this field.

Analysis of information based on the effectiveness of research in the field of IR in medical sciences shows that most scientific productions in this field fall into 2 categories: (1) Effective methods in organizing information and (2) applications and operations of IR systems, the process of intelligent questioning and answering and analysis of information behavior of physicians and health professionals. The important point in this regard is to increase the effectiveness of scientific productions in the issues of structuring and organizing knowledge and using tools such as ontologies and other semantic tools in systematizing knowledge compared with methods such as data mining. To put it another way, study and attention to pre-designed tools and semantic tools has grown over time in comparison to automatic data extraction and retrieval approaches. Also, the effectiveness of scientific documents based on answering clinical questions and focusing on health professionals' information behaviors has increased compared with search methods and tools. This situation indicates the conditions in which researchers have focused more on human factors in IR.

Our results showed that the documents were in 4 thematic clusters: "Analysis of Physicians' Information Behavior, IR Systems, EBM, and CDSS", "EHR and Medical Documents," "Text Mining and Indexing" and "Question Answer Systems." In a similar study, Zowj et al identified 10 clusters in a study to identify trends in data retrieval research using the author's citation network, including Library and Information Science, Computer Science, Electrical Engineering, Information Retrieval, Information-seeking Behavior, Psychology, Multimedia Information Retrieval, Software Engineering, Ophthalmology, and Surgery. The reason for this difference, in addition to the focus of the current study on information retrieval articles in the field of medical sciences, was the exclusion of noninformation retrieval articles in our study. Therefore, only articles written directly in the field of information retrieval in medical sciences were included in our cluster mapping. The point is that regarding the information behavior of users, the results of the mentioned research are in line with those of our research. In both studies (information retrieval and information retrieval in medical sciences), attention to human dimensions and user behavior has been one of the most important focuses of research (24).

The analysis of scientific documents published based on keywords in the field of IR research in medical sciences shows that the thematic cluster of "IR technologies and techniques" in terms of all 3 indicators, total strengths link and keywords occurrences has been the strongest and most cohesive cluster. The "Information Behaviors and CDSS Systems" cluster ranks next to all of these indicators with little difference. This situation shows that although the technologies and retrieval techniques are still at the top in terms of the frequency of the research topics, the strengths of human aspects are quite significant. In

other words, the focus and attention on human elements of IR in medical sciences, such as information behavior and use of technology in clinical scientific processes and related clinical domains, has increased in terms of the number of Items. This confirmed the analysis of scientific products based on their effectiveness (eg, the citation status of published documents).

Ding et al in their research on data retrieval research mapping using keyword analysis, identified 5 main clusters in this research and stated that the trend of information retrieval research is moving toward concepts such as the World Wide Web. The Web, information retrieval behaviors, artificial intelligence, online databases, electronic publishing, neural networks, knowledge illustration, data mining and search engines, and topics such as information needs of users in parallel with technical issues of information retrieval have been considered. This research is consistent with ours and indicates the continued focus of researchers in this field on the human aspects of information retrieval. Also, the use of intelligent methods of knowledge organization instead of classical methods such as organization based on traditional methods has received more attention (6). This part of the results is also consistent with the current results.

From another perspective, Zhao and Rui identified cross-language information retrieval research centers. The main centers of research are CLIR techniques, machine translation, query translation, query expansion, and parallel corpora. Similarly, in our study, query expansion was in the third cluster, and this situation shows the importance of query expansion in various areas of data retrieval (25).

The results also showed that the similarity index increased over time from 0.43 to 0.71. This means that researchers in medical IR have, over time, brought their terms closer together. On the other hand, the findings showed that during the 2000s and 2010s, most new keywords ( $n=91$ ) entered the literature and terminology of IR in the medical field, indicating the growth of new concepts and dramatic changes in the development of thematic boundaries in this area. However, between 2010 and 2020, the development of new keywords has decreased by about half ( $n=57$ ) compared with the period 2000-2010, showing a slowing in the subject area's growth rate. Analysis of the findings showed that similarity measures, expert systems, concepts, experience, answers, and multi-model IR clusters were considered as mature and completely centralized clusters in the first quarter of the strategic chart. In other words, these thematic clusters were highly centralized and had the highest internal coherence and communication and were conceptually very close and interconnected. These clusters were quite developed and very important for the development of the main research structure in the scientific field. In the second quarter, which represents cohesive but centralized clusters, each of which consisted of smaller specialized areas of science, the smartphone, hybrid, decision tree, RFID, and feasibility study clusters were included. In the third quarter, clusters of relational database, mechanism, clinical information systems, medical terminologies, and barriers were

clusters of low density and centrality, with most emerging or declining themes. In the fourth quarter, health information exchanges, metadata, and medical informatics were not mature clusters but had the potential to become major research topics in the field of health IR in the future.

In a research by Abdollahzadeh, which drew a thematic map of the field of librarianship and information using the co-occurrence method, it was found that the metadata cluster was one of the central but not developed clusters, which is completely consistent with the results of our research (21).

In terms of the study's findings, it's important to note that some aspects of the study have limitations, which may have hampered the retrieval of all relevant studies. First, there is no vocabulary control tool on the Web of Science to retrieve all related studies, including broader, narrower, and related topics. There may also be restrictions on retrieving resources on the subject of information retrieval in the medical sciences and all related fields. To overcome this problem, searching of documents was run in Topic Field in Web of Science. The search in the Topic Field, in addition to the author's keywords, retrieves the Keyword Plus, which are index terms that are generated automatically from the titles of referenced publications and are useful for retrieving the connected vocabulary network.

### Conclusion

Paying attention to the evolution of various scientific fields is one of the most important prerequisites for research policymaking and predicting the scientific needs of researchers. This study aimed to respond to this goal and draw future perspectives in the highly variable and developing field of IR in medical sciences. The importance of this issue is that the IR and its related subjects in medical sciences need to evaluate IR techniques as a powerful tool for developing research capabilities. As a result, paying close attention to the model, maps, and visualizations in this study, which are the result of a thorough analysis of scientific products published in the world's most famous scientific journals, can be useful in identifying research gaps and future needs in IR. Other considerations include a dramatic approximation of the vocabulary used (in fact, research areas) by researchers and a relative slowdown in the growth rate of the subject's domain in the last decade from 2000 to 2010. Therefore, it seems necessary to pay attention to the expansion of the fields of IR and the application of its concepts in medical information sciences.

In particular, research findings indicate a relative growth in the focus of IR research on the practical and human aspects of IR and information retrieval behaviors. These conditions indicate the specific situation of the application of IR technologies in medical sciences and the focus on human factors along with technological factors. Therefore, it can be recommended that designers of IR systems and techniques in medical information sciences pay more attention to human factors attentively to develop new technologies and tools.

### Acknowledgment

This article was part of a Ph.D. dissertation that was done at Tehran University of Medical Sciences. Researchers thank Dr Alireza Norouzi for providing valuable guidance in conducting this research.

### Conflict of Interests

The authors declare that they have no competing interests.

### References

- Xu B, Lin H, Yang L, Xu K, Zhang Y, Zhang D, et al. A supervised term ranking model for diversity enhanced biomedical information retrieval. *BMC Bioinform.* 2019;20(16):1-11.
- Alves T, Rodrigues R, Costa H, Rocha M. Development of an information retrieval tool for biomedical patents. *Comput Methods Programs Biomed.* 2018;159:125-34.
- Di Girolamo N. Advances in Retrieval and Dissemination of Medical Information. *Vet Clin North Am Exot Anim Pract.* 2019;22(3):539-48.
- Harman D. Information retrieval: the early years. *Found Trends Inf Retr.* 2019;13(5):425-577.
- Li S, Jin Q, Jiang X, Park JJH. Frontier and Future Development of Information Technology in Medicine and Education: ITME 2013: Springer Science & Business Media; 2013.
- Ding Y, Chowdhury GG, Foo S. Bibliometric cartography of information retrieval research by using co-word analysis. *Inf Process Manag.* 2001;37(6):817-42.
- Janssens F, Glenisson P, Glänzel W, De Moor B, editors. Co-clustering approaches to integrate lexical and bibliographical information. Proceedings of the 10th international conference of the International Society for Scientometrics and Informetrics (ISSI); 2005.
- Irmawati S, Cakrawijaya H, Lydia EL, Shankar K, Nguyen PT. Medical Information Retrieval for Healthcare: The Challenges. *Int J Eng Adv Technol.* 2019;8(6S3):811-4.
- Rorissa A, Yuan X. Visualizing and mapping the intellectual structure of information retrieval. *Inf Process Manag.* 2012;48(1):120-35.
- Lu C, Liu M, Shang W, Yuan Y, Li M, Deng X, et al. Knowledge Mapping of Angelica sinensis (Oliv.) Diels (Danggui) Research: A Scientometric Study. *Front Pharmacol.* 2020;11:294.
- Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *J Informetr.* 2011;5(1):146-66.
- Cobo M, López-Herrera A, Herrera-Viedma E, Herrera F. SciMAT Version 1.0 User guide.1-17.
- Cobo MJ, López - Herrera AG, Herrera - Viedma E, Herrera F. SciMAT: A new science mapping analysis software tool. *J Am Soc Inf Sci Technol.* 2012;63(8):1609-30.
- Rezaei L, Mohammadi M. Scientometric analysis of Iranian scientific productions in the field of Ophthalmology. *JCBR.* 2018;2(4):23-32.
- Van Eck NJ, Waltman L. VOS: A new method for visualizing similarities between objects. *Advances in data analysis: Springer;* 2007. p. 299-306.
- Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics.* 2010;84(2):523-38.
- Van Eck NJ, Waltman L. Text mining and visualization using VOSviewer. *arXiv preprint arXiv:11092058.* 2011.
- Wilbur WJ, Yang YM. An analysis of statistical term strength and its use in the indexing and retrieval of molecular biology texts. *Comput Biol Med.* 1996;26(3):209-22.
- Hersh WR, Hickam DH. How well do physicians use electronic information retrieval systems?: A framework for investigation and systematic review. *JAMA.* 1998;280(15):1347-52.
- Mohammadi M, Sheikhshoei F, Banisafar M, Mozafari O. Scientometric Analysis of Scientific Publications on Persian Medicine Indexed in the Web of Science Database. *Webology.* 2019;16(1):151-65.
- Abdollahzadeh P. Mapping Research Topics of Library and Information Sciences based on Co-word Analysis. Tabriz: Tabriz University of Medical Science; 2019.

22. Ke W, Yunjiang X, Xiao L, Weichan L, editors. Analysis on current research of supernetwork through knowledge mapping method. International Conference on Knowledge Science, Engineering and Management; 2013: Springer.
23. Melcer E, Nguyen T-HD, Chen Z, Canossa A, El-Nasr MS, Isbister K. Games research today: Analyzing the academic landscape 2000-2014. network. 2015;17:20.
24. Zowj HA, Ghane MR, Ehsanifar F. Identifying information retrieval research trends using author co-citation network. Int J Inf Sci Manag. 2019;17(2):99-117.
25. Zhao R, Rui C, editors. Visual analysis on the research of cross-language information retrieval. Proceedings of the International Conference on Uncertainty Reasoning and Knowledge Engineering, URKE 2011; 2011.

*Supplementary 1*

label replace by  
 information-retrieval information storage and retrieval  
 health information retrieval information storage and retrieval  
 semantic information retrieval information storage and retrieval  
 retrieval information storage and retrieval  
 information-storage information storage and retrieval  
 medical information retrieval information storage and retrieval  
 biomedical information retrieval information storage and retrieval  
 information storage information storage and retrieval  
 information retrieval information storage and retrieval  
 information storage and retrieval methods information storage and retrieval  
 information retrieval systems information retrieval system  
 information-retrieval systems information retrieval system  
 information system information retrieval system  
 system information retrieval system  
 systems information retrieval system  
 abstracting and indexing as topic abstracting and indexing  
 abstracts abstracting and indexing  
 indexing abstracting and indexing  
 access access to information  
 algorithm algorithms  
 bibliographic bibliographic databases  
 database databases  
 decision-making decision making  
 decision-support decision support system  
 decision support decision support system  
 electronic health record electronic health records  
 electronic medical record electronic health records  
 electronic medical records electronic health records  
 evidence based practice evidence based medicine  
 ebm evidence based medicine  
 information-seeking behaviour information seeking behaviour  
 seeking behaviour information seeking behaviour  
 information seeking information seeking behaviour  
 seeking information information seeking behaviour  
 behaviour information seeking behaviour  
 internet use internet  
 medical informatics applications informatics  
 medical informatics informatics  
 medical subject headings mesh  
 model models  
 network networks  
 ontology ontologies  
 searches search  
 search strategies search  
 search filters search  
 semantics semantic web  
 semantic similarity semantic web  
 text text retrieval  
 world-wide-web web  
 world wide web web  
 web searchweb