Journal of Informetrics 16 (2022) 101317

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Towards more convergent main paths: A relevance-based approach

Chen-Hao Huang[a], John S. Liu[a,*], Mei Hsiu-Ching Ho[a], Tzu-Chuan Chou[b]

[a] *Graduate Institute of Technology Management, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC*
[b] *Department of Information Management, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC*

ARTICLE INFO

Keywords:
Citation relevance
Literature review
Main path analysis
Health interoperability standards

ABSTRACT

This research proposes a new approach that considers citation relevance in main path analysis (MPA). Traditional MPA assumes that all citations have equal weight, but in practice treating every citation equally may not find the main paths that truthfully reflect the knowledge flow in a target science field. To address the issue, this study suggests taking the level of relevance among documents into consideration. For demonstration purposes, the level of relevance is determined by similarity in both citation structure and key phrases among documents. The approach not only achieves convergence of development trajectories, but also helps frame the topics on the main paths to a specific concept from a wide range of research domains. This study takes health interoperability fields as the demonstration case to show the effects of converging the trajectories toward a target domain.

1. Introduction

Main path analysis (MPA) is a citation-based method commonly used to trace the key development paths of a scientific discipline. First introduced by Hummon and Doreian (1989), it has been applied to a variety of technological and academic fields for the purpose of uncovering their main knowledge flows (Chuang, Liu, Lu, & Lee, 2014; Ho & Liu, 2021; Xiao, Lu, Liu, & Zhou, 2014; Yu & Pan, 2021) as well as mapping technological trajectories (Hung, Lai, & Liu, 2022; Hung, Liu, Lu, & Tseng, 2014; Verspagen, 2007). MPA results are presented as a sequence of articles that usually carry the main knowledge of the target field. Nevertheless, articles addressing an unrelated topic may sometimes appear on the main paths. This is likely due to the fact that traditional MPA, which is the same as most other citation-based analyses, assumes all citations are of direct relevance and have equal weight. In other words, the topics of the citing papers are fully related to the cited papers, and all citations are treated with equal weight.

In reality, the above scenario is not so. Moravcsik and Murugesan (1975) indicate that the nature of a citation varies and can be conceptual or operational, important or perfunctory, evolutional or juxtapositional, or confirmatory or negational. Such a variety in nature leads to different effects and influences by each citation. Furthermore, there can be “strategic” intentions by the authors. For example, they can select to cite (or not to cite) a particular article in order to increase the probability of publishing in a journal (Bhupatiraju, Nomaler, Triulzi, & Verspagen, 2012). In addition, there is a tendency to cite articles that are published in high quality journals (Chan, Önder, & Torgler, 2016) or written by personal friends (Luor, Lu, Yu, & Chang, 2014). All these factors result in different levels of citation relevance. Not considering the relevance of each citation causes the inclusion of low-relevance articles on the paths, which means some non-congruent research focuses may interrupt the observation of the target domain, thus limiting MPA’s capability of converging the trajectories toward a specific topic.

* Corresponding author.
E-mail address: johnliu@mail.ntust.edu.tw (J.S. Liu).

https://doi.org/10.1016/j.joi.2022.101317
Received 5 January 2022; Received in revised form 7 June 2022; Accepted 18 July 2022
Available online 28 July 2022
1751-1577/

MPA has been adopted to extract knowledge paths from academic articles and patent documents since 1990 (Hummon, Doreian, & Freeman, 1990). While the benefits of MPA have been known for decades, only recently has the algorithmic enhancement its effectiveness been proposed. The approach to enhance MPA can be divided into three types. The first type focuses on improving the search method to make the resulting main paths more practical. For example, Liu and Lu (2012) propose key-route search, which ensures that the links with the highest traversal count are included in the main paths as well as adds a mechanism for presenting multiple paths. The second type pays attention to topic modeling with the aim to improve the understanding of topic evolution in scientific literature. For instance, Kim, Baek, and Song (2018) employ a citation influence topic model to identify topic evolution. More recently, Kim, Jeong, Kim, and Song (2022) propose topic-integrated path analysis to uncover the topics from an entire dataset by using latent Dirichlet allocation topic modeling.

The third type concentrates on adjusting the citation weight by taking the level of citation relevance into consideration. For instance, Yu and Pan (2021) propose preference main path, which assigns different weights to each citation so as to enhance the distinguishability of the citation values. The method resolves the problem of discipline bias and time bias in interdisciplinary research. Liu, Chen, Ho, and Li (2014) similarly adopt the idea of adjusting weighting factors to handle citations with different levels of relevance. They propose three treatments, including flat, linear, and exponential, to adjust the relevance level and find that incorporating relevance information helps MPA uncover scientific studies of higher importance. Their study, nevertheless, stops short of uncovering relevance through a more comprehensive approach. For one, the relevance information is provided by a database. Second, only citation relevance is taken into consideration.

This current research deviates from Liu, et al. (2014) by taking both citation and key phrase relevance into consideration and determines them through an automatic algorithm. Doing so lessens the role of low-relevance citations, thus improving the effectiveness of MPA. The level of relevance, as a methodological demonstration, is measured by analyzing the numeric similarity on both citation structure and key phrases and then generating an overall relevance indicator. In this article, similarity refers to closeness in numerical measures of citation structure and key phrases, and relevance emphasizes that the citing and cited articles are relevant in their research topics.

To show how the proposed approach works, this study illustrates it by uncovering the main paths in health interoperability, which refers to the communication capability among health information resources. Health interoperability is about the connection between different health information resources. Therefore, establishing a “standard” to connect these resources has been recognized as a critical sub-concept, which thus allows two heterogeneous information resources to share health information.

The reasons for selecting health interoperability field are twofold. First, the field includes various sub-concepts, such as software components (Jean et al., 1994), mediator concepts (Xu, Sauquet, Degoulet, & Jaulent, 2003), and the usability of emerging technologies (Gordon & Catalini, 2018; Sung, et al., 2020), among others. The diversity of these sub-concepts complicates the citation structure and enlarges the difficulty of focusing on a specific sub-concept. Second, there is a need to trace the knowledge flow of a particular sub-concept of the domain, or the “standard”. Framing the paths to a “standard” topic could help health researchers and practitioners concentrate on essential factors of each standard topic, thus helping the development of a new standard. To sum up, by taking citation and key phrase relevance into consideration, this study allows us to frame the topics on the main paths to the “standard” concept in health interoperability literature.

2. Background

This section discusses prior literature on the relevance between the citing and cited articles as well as on the concepts of health interoperability standards and their roles in the health domain.

2.1. Taking Relevance into Consideration

The blind spot of citation analysis, as stated by Bonzi (1982), “… of course, is that not all citations are equal.” Because authors might cite a paper following a variety of considerations (Moravcsik & Murugesan, 1975), such varieties lead to different degrees of similarity in a topic between pairs of cited and citing documents (Cronin, 1994). Furthermore, there can be perfunctory citations, which cite articles without any further comments to the contents of the source cited (Hu, Chen, & Liu, 2015). Krampen, Becker, Wahner, and Montada (2007) indicate that one-fourth of all citations in psychological publications are perfunctory and have “little information utility”. An irrelevant self-citation may also affect the result of certain citation studies. As Pichappan and Sarasvady (2002) indicate, many author self-citations are intrinsic, maintaining the intellectual link among the scholarly work, yet an irrelevant self-citation makes it extrinsic. All these issues - i.e., ignoring disparity in citation and treating all citations equally - are challenging to most citation analyses. This is the reason why Liu, et al. (2014) suggest taking the relevance level of citations into consideration to improve the legitimacy of a citation-based analysis.

Many studies have sought to take various information to assign a weight to each citation. In regards to text-based similarity, Mei and Zhai (2008) propose an impact-based summarization approach to extract sentences that are capable of rendering the most influential content of the article. Similarly, Ding, et al. (2014) suggest interpreting article content based on its context at both the syntactic and semantic levels to measure the value of a citation. Moreover, Aman (2018) adopts cosine similarity to calculate the similarity of the knowledge base of authors and their co-authors and thus to uncover the potential knowledge transfer. In short, text-related factors help researchers to find high-impact and similar studies on unfamiliar topics.

The terms used in the citing and cited articles are likely to be similar if they discuss the same topic. As Trivison (1987) notes, the term co-occurrence is much greater for cited/citing document pairs than for randomly-selected pairs of documents.

Peters, Braam, and van Raan (1995) show in the chemical engineering literature that publications with a citing relationship have a similar word-profile; i.e., carry a similar set of title-words, keywords, and/or classification codes. Zhu et al. (2017) consider that the semantic relatedness and similarity between biomedical terms have a critical impact on the analysis of biomedical publication data and help researchers retrieve more relevant documents. All these factors highlight that similarity among documents does exist and can be used for identifying relevance among documents.

As the goal of this study is to demonstrate the value of taking relevance into consideration in MPA, we adopt a simple and straightforward approach to measure citation relevance. We assume that the similarity of content-specific key phrases between the citing/cited pair is closely associated with document similarity. In addition, we consider that structural similarity between two articles in a citation network also relates to document similarity and thus include structural similarity in the relevance measure.

2.2. Health Interoperability Standards

Interoperability refers to the capability of providing two heterogeneous objects a way to exchange and share health information. Integrating health information from various medical sources is crucial in providing a comprehensive view on healthcare (Heart, Ben-Assuli, & Shabtai, 2017). However, integrating health information faces several barriers, such as data being generated from various systems or belonging to different health institutions or organizations. To overcome these barriers, the American National Standards Institute (ANSI) established Health Level 7 (HL7) standards to enable semantic interoperability across all platforms (Dolin, et al., 2001). After the establishment of HL7 standards, a growing number of health institutions began to pay attention to the concept of interoperability, thus making HL7 play an increasingly important role in facilitating the sharing and exchange of health information (Yang, Chou, & Chen, 2019).

Along the way in the development of health interoperability, health information technology has produced many heterogeneous objects such as devices, data, information, etc. However, the growing scope of heterogeneous objects includes various sub-concepts in health interoperability and complicates the citation structure of health interoperability. More precisely, such a growing body of health interoperability studies increases the difficulty of tracing the most vital concept of the target domain.

Among the many sub-concepts in health interoperability research, the “standard” concept has been recognized as one of the most critical issues. Health interoperability standards are constantly changing with the development of information technology and related medical regulations. In addition, defining health interoperability standards involves not only the problem of diverse semantic architecture paradigms (BGME Blobel, Engel, & Pharowe, 2006; Garde, Knaup, Hovenga, & Heard, 2007), but also various standardization issues, such as data standards (Gay & Leijdekkers, 2015; Joda, Waltimo, Probst-Hensch, Pauli-Magnus, & Zitzmann, 2019), medical device standards (Hudson & Clark, 2018), system standards (Frontoni, et al., 2019), storage standards (Masud, Hossain, & Alamri, 2012), and prescription transfer standards (Webster & Spiro, 2010). Considering the complexity of standards in health interoperability literature, this study chooses standards as a demonstration case to extract more convergent main paths.

3. Relevance-based MPA

The proposed relevance-based MPA operates in three steps. The process is virtually the same as traditional MPA except that a new step (the second step) is added. The first step assigns a traversal count to each citation link in a citation network. The second step measures citation relevance and adjusts each link’s traversal counts accordingly. The third step searches for the main paths based on the adjusted traversal counts. The following sections mathematically present how the first and second steps are processed.

3.1. Assigning Traversal Counts

The first step assigns a traversal count to each citation link in a citation network. A citation network N = (U, R) is constructed by a set of documents U that are connected by the relationships $R \subseteq$ U × U, and (u, v) indicates a citation relationship whereby article v cites document u. Such a network is direction sensitive and includes four categories of nodes: sources, sinks, intermediates, and isolates. A source represents a node that is cited by other nodes, but cites no others, while a sink has a completely reversed property. An intermediate node cites other nodes and is also cited by others, whereas an isolated node has no referencing relationship with others.

Assuming that an imaginary messenger disseminates knowledge around the citation network though the directional links among documents, to enumerate the number of times a link (u, v) is traversed by the messenger, one follows the steps of the messenger and traces knowledge flows from a specified origin document to another specified destination document. The traversal count $w_{o\to d}(u, v)$ for the link (u, v) can be represented by the multiplication of the number of paths leading from the origin to u by the number of paths leaving from v to the destination, which are denoted as $N^-(u)$ and $N^+(v)$, respectively, as follows:

$$w_{o\to d}(u, v) = N^-(u) \cdot N^+(v), (u, v) \in R \tag{1}$$

where:

$$N^-(u) = \begin{cases} 1, \text{ if } u \text{ is the origin} \\ \text{number of paths leading from the origin to } u, \text{ if otherwise} \end{cases} \tag{2}$$

$$N^+(v) = \begin{cases} 1, \text{ if } v \text{ is the destination} \\ \text{number of paths leaving from } v \text{ to the destination, if otherwise.} \end{cases} \tag{3}$$

According to Hummon and Doreian (1989), the search path link count (SPLC) for a link (*u, v*) is the number of times the link is traversed on the premise that the messenger's mission is to deliver knowledge through all possible paths from all the ancestors of the node *u* (including itself) to all the sinks. In other words, SPLC is the sum of $w_{o\rightarrow d}(u,v)$, where the origins are all the ancestors of node *u* (including *u*), and the destinations are all the sinks. Therefore, we have:

$$w_{\mathrm{SPLC}}(u,v) = \sum\nolimits_{\text{all combinations of ancestors and sinks}} w_{o\rightarrow d}(u,v). \tag{4}$$

The MPA literature defines several other traversal counts such as SPC, SPNP, SPAD, SPGD, and SPHD (Hummon & Doreian, 1989; Batagelj, 2003; Liu & Kuan, 2016). This study chooses the SPLC algorithm, because it better fits the real-world knowledge diffusion scenario than the other traversal counts (Liu, Lu, & Ho, 2019). In addition, we want to avoid traversal counts that consider the factor of knowledge decay, such as SPAD, SPGD, and SPHD, to better demonstrate the effect of citation relevance.

3.2. Measuring Citation Relevance and Adjusting Traversal Counts

The second step measures citation relevance and adjusts each link's traversal counts. 3.2.1 Measuring citation relevance

To measure relevance between two articles, one needs to first select the features with which to compare and then choose a comparison method. Features can be as simple as author, journal, key phrase, citation structure, or as comprehensive as whole text. Commonly-used comparison methods include cosine similarity, Jaccard similarity, Euclidean distance, etc. To demonstrate relevance-based MPA, this study takes key phrase and citation structure as the features and Jaccard similarity as the method for comparison.

Jaccard similarity measures the number of similar elements between two given sets, as follows:

$$r(u,v) = \frac{|\mathbf{A}\cap\mathbf{B}|}{|\mathbf{A}\cup\mathbf{B}|} \tag{5}$$

where $\mathbf{A} = \{a_1, a_2, \ldots\}$ and $\mathbf{B} = \{b_1, b_2, \ldots\}$ are the value set of a selected feature for node *u* and node *v*, respectively, and a_i, b_i, … are feature values that are either 0 (not the same) or 1 (the same).

Key phrase is one of the features selected for similarity analysis. This is based on the assumption that the topics of the two target articles are similar when their key phrases are highly correlated. The largest value of key phrase similarity is 1 when the key phrases of the two target articles are exactly the same. In this study we take key phrases from both the title and abstract of an article, and a multiple appearance of the same key phrase is treated the same as a single appearance. To take care of the situation whereby various terms express the same meaning, we standardize terms with the same meaning into a representative key phrase. For example, "HL7", "health level 7", and "health level seven" carry the same meaning and are all replaced with "HL7" before similarity comparison.

In regards to citation structure, we assume that structural similarity between any two articles in a citation network relates to document similarity and that high structural similarity indicates a high level of relevance. Structural similarity, in the context of this study, measures overlap in neighbors of two nodes in a network and could be recognized as a form of integrating bibliographic coupling and co-citation. As stated in the literature, despite bibliographic coupling and co-citation providing different facets for viewing the field, these two research methods could enhance the accuracy of the research fronts found by leveraging their own characteristics (Huang & Chang, 2015). To the extreme that the neighbors of two articles are completely the same, the two articles are "structurally equivalent" (Lorrain & White, 1971). In other words, two structurally equivalent articles cite the same set of articles and are cited by the same other set of articles. This assumption is in line with Ciotti, Bonaventura, Nicosia, Panzarasa, and Latora (2016), who suggest when two articles have similar bibliographies that they might be engaged with a similar research topic or problem.

Both keyphrase similarity indicator and citation structure similarity indicator play a significant role in evaluating the level of relevance between the citing and cited articles. Correctly balancing these two indicators is a delicate issue, especially when one of the indicators produces a significant influence on overall relevance. To ensure these two indicators have a balanced influence on overall relevance, this study balances their influence by adopting a weighting scheme rather than merely adding the similarities. The aggregated similarity can be expressed as follows:

$$r(u,v) = w_x \cdot r_x + w_y \cdot r_y \tag{6}$$

where w_x is the citation structure weight, r_x is the citation structure similarity, w_y is the key phrase weight, and r_y is the key phrase similarity.

The weighting scheme adopted takes the reverse of the typical values for the two similarity measures as their weight to avoid discounting one of them on the overall relevance. In this study, as will be shown later, the similarity level of key phrases is greater than that of citation structure. Therefore, w_x is set to be greater than w_y - that is, w_x =0.7346, and w_y=0.2654.

3.2.1. Adjusting traversal counts

Liu, et al. (2014) suggest taking the relevance level of citations into consideration when determining traversal counts. Following this idea, we multiply each link's traversal count by an adjustment factor depending on its level of relevance. Assuming that the adjustment factor for the link between documents *u* and *v* is $\alpha(u, v)$, the adjusted traversal counts can be expressed as:

$$w^{*}_{\mathrm{SPLC}}(u,v) = \alpha(u,v)\cdot w_{\mathrm{SPLC}}(u,v) \tag{7}$$

where w_{SPLC} and w^{*}_{SPLC} represent the original and adjusted traversal counts, respectively. For traditional MPA, the adjustment factors are one for all the links. Liu, et al. (2014) propose three ways of adjusting traversal counts, including flat, linear, and exponential. Flat adjustment simply ignores citation relevance; linear adjustment multiplies traversal counts with the relevance level for each citation

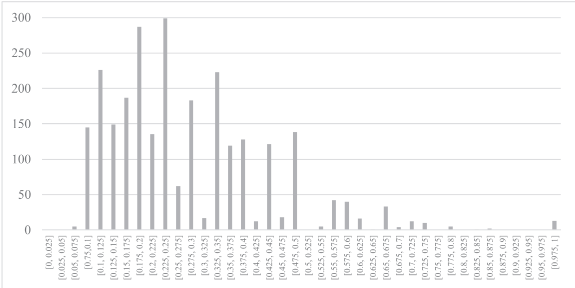

Fig. 1. Key phrase relevance histogram.

link; exponential adjustment factor is defined as Euler's number e (2.71828) to the power of the relevance level. Considering the limited network size of the demonstrative case, this study adopts linear adjustment as it is rather intuitive. The adjustment factor therefore is simply the relevance level, as follows:

$$\alpha = r(u, v) \tag{8}$$

where *r(u,v)* is the level of relevance between nodes *u* and *v*.

4. Results

This section examines the effectiveness of the proposed approach. We first illustrate the method and present the changes in the ranks of significant citations. Next, we make a comparison between the main paths obtained from traditional MPA and relevance-based MPA.

4.1. Applying the Proposed Approach

This section illustrates the relevance-based MPA using bibliographic information of the health interoperability research field. As mentioned earlier, this research field includes various sub-issues that allow relevance-based MPA to show its merit.

4.1.1. Literature Selection

A query string is first designed to search for health interoperability literature. The main term in the string is "interoperab*", where the asterisk at the end of "interoperab" dictates the search system to take "interoperability", "interoperation", and "interoperable" into consideration. In addition, by applying Boolean operator AND, the terms including "health", "hospital", "medical", "patient", and "clinical" limit the search results to the healthcare domain. The search from the Web of Science database results in a total of 2,412 papers in the period from January 1, 1993 to December 31, 2020. The number of articles drops to 1,603 after removing irrelevant articles; i.e., those that are isolated from the others in citation. The beginning year is set to 1993, because it is the year that the first academic health interoperability article was published.

4.1.2. Relevance Statistics

A citation network among the 1,603 papers is established, which contains 3,332 citation links. This section presents the relevance statistics for the cases when adopting only key phrase similarity, only structural similarity, and both.

To find key phrase similarity, one needs to establish a key phrase set that is used as the base for comparison. We select 20 commonly-used standard-related key phrases such as "architecture", "infrastructure", "standard", "model", etc. as the base key phrase set and compare the usages of these base key phrases between each pair of the citing and cited papers. We ignore non-relevant links. Among the 2,670 remaining links, the maximum, minimum, and average are 1, 0.0625, and 0.28626, respectively. Figure 1 offers a histogram. Most scores are within the range of 0.075 to 0.5, and the peak of the distribution falls in the range of 0.225 to 0.25, which is a little below the average relevance.

For structural similarity, Figure 2 shows the resulting histogram, In comparison with the citation structure, we eliminate any article without any same citation and obtain 1,345 citation links. The maximum, minimum, and average relevance are 0.54547, 0.01, and 0.1034, respectively. Most values are within the range of 0.075 and 0.5, and the peak of the distribution is at the fourth bin. The result indicates that structural similarity is relatively low in comparison with key phrase similarity.

Based on the average relevance of 0.2826 and 0.1034, respectively, for key phrase and citation structure similarity, we assign w_x as 0.7346 and set w_y to 0.2654 for the purpose of balancing the impact of these two similarity measures. By checking the key phrase sets and citation similarity, we remove non-relevance articles and obtain 2,688 citation links. Figure 3 presents the histogram of overall relevance. The distribution of overall relevance is similar to that of key phrase relevance. As the figure reveals, the distribution hits the peak in the range of 0.1 to 0.125. The maximum relevance is 1, the minimum relevance is 0.125, and the average relevance is 0.482755.

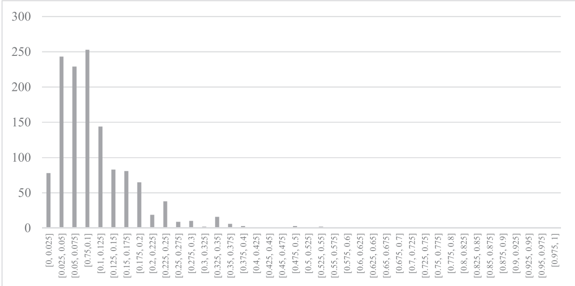

Fig. 2. Structural relevance histogram.

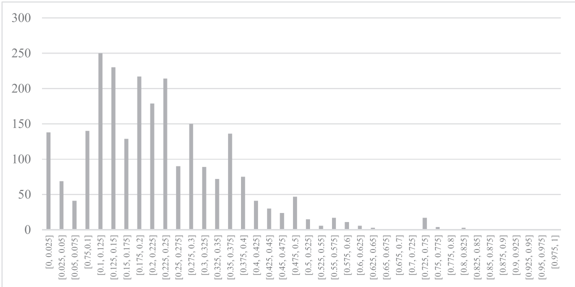

Fig. 3. Overall relevance histogram.

Table 1
Top 20 significant citations with the changes after adjustment.

Rank	SPLC	From => To	Relevance	Adjusted SPLC	Ranking Change
1	280227	GardeKHH2007 => Martinez-costaMTA2009	0.311073772	87171.27	-1
2	250316	JaspersKS2006 => GardeKHH2007	0.3673	91941.07	1
3	169382	BakkenCCHH2000 => DolinABBBEKLM2001	0.265819	45025	-1
4	159509	DolinABBBBS2006 => JaspersKS2006	0.14692	23435.06	-12
5	159292	XuSDJ2003 => GardeKHH2007	0.091825	14626.99	-29
6	144565	GardeKHH2007 => ChenKSKA2009	0.091825	41041.44	1
7	130460	DolinABBBEKLM2001 => DolinABBBBS2006	0.493529	64385.75	4
8	117400	MoriCG1998 => BakkenCCHH2000	0.244867	28747.35	-2
9	113780	BirdGT2003 => GardeKHH2007	0.221564	25209.58	-4
10	112791	PahlZNDWDSI2015 => ChristensenE2016	0.266354	30042.31	1
11	110174	Martinez-costaMTA2009 => MaldonadoMBFAR2009	0.142646	15715.91	-20
12	104764	DolinABBBEKLM2001 => BirdGT2003	0.23399	18260.66	-15
13	103066	GardeKHH2007 => MaldonadoMBFAR2009	0.142265	17566.2	-16
14	91024	DolinABBBEKLM2001 => GardeKHH2007	0.192802	17549.58	-16
15	88245	LiLLAGJ2018 => RoehrsDRDGS2019	0.29384	25929.91	3
16	87720	MartinezcostaMT2011 => Menarguez-tortosaT2013	0.366299	32131.74	9
17	84691	DolinABBBEHLMRSSW1999 => DolinABBBEKLM2001	0.382466	32391.4	11
18	83160	MarcosGPC2015 => RoehrsDR2017	0.227553	18923.31	-5
19	80432	RoehrsDRDGS2019 => HussienYUZZ2019	0.258041	20754.76	1
20	78702	Kopanitsa2017 => WulffHTMBJ2018	0.191456	15067.96	-13

4.2. Changes in The Ranks of Significant Citations

After taking relevance into consideration, the top citations largely shift. Table 1 presents the before and after SPLC values of the top 20 citations (before adjustment). As the table highlights, the rankings of two citations, XuSDJ2003 cited by GardeKHH2007 and Martinez-costaMTA2009 cited by MaldonadoMBFAR2009, drop significantly (by more than 20), thus largely reducing their chances of being put on the main paths.

The citation "XuSDJ2003=>GardeKHH2007" experiences the largest change due to its low relevance score (0.091825). Its rank drops from 5 to 34. Examining further, we find that GardeKHH2007 (Garde, et al., 2007) cites XuSDJ2003 (Xu, et al., 2003) for the reason of supporting the claim that interoperability is important among different information systems. GardeKHH2007 uses mainly mediation-related terms such as middleware and mediator, while XuSDJ2003 elaborates on domain knowledge governance phrases, thus causing a low similarity between their key phrases. Moreover, the citing and cited documents of these two articles vary in research topics and perspectives. More precisely, XuSDJ2003, which discusses component-based mediation services, mainly cites and is cited by engineering-oriented studies published in journals such as IEEE and ACM, while GardeKHH2007, which focuses on Electronic Health Record (EHR) standard establishment, generally cites and is cited by articles published in medical information-related journals such as *Methods of Information in Medicine* and *International Journal of Medical Informatics*.

The ranking of citation "Martinez-costaMTA2009 => MaldonadoMBFAR2009" drops by 20. In analyzing the result, we see that the topic of MaldonadoMBFAR2009 focuses on the development of archetype editing framework, while Martinez-costaMTA2009 targets transforming Archetype Definition Language (ADL) into Ontology Web Language (OWL) (Maldonado, et al., 2009; Martínez-Costa, Menárguez-Tortosa, Fernández-Breis, & Maldonado, 2009). MaldonadoMBFAR2009 regards archetypes as a basic factor of semantic interoperability among EHR systems and suggests investigating similarities and differences between archetypes and ontologies when both of them are treated as representation technologies. Because the comparison between archetypes and ontologies is not the core element of MaldonadoMBFAR2009, the authors cite Martinez-costaMTA2009 for their readers to access more detailed discussion. In regards to citation structure, MaldonadoMBFAR2009 mainly cites journal articles to support the framework that paper purposes, while more than one-third of the citations of Martinez-costaMTA2009 are standard and technical-oriented websites.

In addition to the two highlighted citations, those citations that decrease significantly in adjusted SPLC mostly have a different focus in the citing and cited articles, either discussing remotely related topics or similar topics in different dimensions. In some cases, they cite papers to support their sub-ideas or to compare the features with content-related literature.

4.3. Comparison of MPA Results

This study compares the traditional MPA and relevance-based MPA with both quantitative analysis and qualitative content analysis. For quantitative analysis, MPA for both approaches is applied at key-routes 5, 10, 20, and 50, and the resulting relevance scores for both are compared. As expected, the result shows that relevance-based MPA presents more relevant main paths. This new method is significantly different from the traditional one as the key-route numbers increases. Mean relevance score (MRS) for each path on the traditional main paths and relevance-based main paths at key-routes 5, 10, 20, and 50 are measured, respectively. MRS refers to the average relevance scores of document-pairs on the main paths. MRS of the relevance-based main path is higher at all key-route levels. In addition, *t*-test is conducted to examine if there exists significant difference between MRS of each traditional result and that of the relevance-based one. According to *t*-test results, as the number of key-route main paths and sample size increases, the difference between MRS of each traditional result and that of the relevance-based becomes more discernable and achieve statistical significance when the key-route parameter increases to 50 ($P<0.1$).

Regarding qualitative content analysis, we compare the main paths generated by traditional and relevance-based MPA. As will be discussed later in this section, the articles on the traditional main paths discuss a variety of issues in health interoperability, while those on relevance-based main paths mainly focus on the health interoperability standard.

4.3.1. Traditional MPA

Figure 4 illustrates the traditional main paths at key-route 10, which includes 27 papers. In this figure, arrows point from the cited papers to the citing papers. Each paper is labelled, in sequence, with the last name of the first author, the first initials of the co-authors (in capital letters), and the publishing year. The thickness of a line is drawn based on the paper's SPLC value.

The development of health interoperability research, according to the topics of the articles on the main paths, can be divided into three phases. The first phase includes two research streams. The one on the right-hand side focuses on a series of interoperability standards, while the other stream elaborates on a practical approach of interoperability. The study by Mori, Consorti, and Galeazzi (1998) introduces a series of standardization approaches and characterizes three generations of terminological systems to improve semantic interoperability. The studies that follow either review the semantic interoperability activities of the Vocabulary Technical Committee (Bakken, Campbell, Cimino, Huff, & Ed Hammond, 2000) or discuss release 1 and release 2 of Clinical Document Architecture, which is a standard derived from HL7 (Dolin, et al., 2001; Dolin, et al., 2006). The topics of the other stream (see the top left corner of Figure 4, dotted area) focus on the integration approach of medical applications, such as integration strategy (Jean, et al., 1994), semantic mediation (Degoulet, Sauquet, Jaulent, Zapletal, & Lavril, 1998), and multi-agents service model (Xu, et al., 2003).

At the beginning of the second phase, Garde, et al. (2007) introduce an archetype-based approach - open electronic health records (openEHR). Archetypes refer to the minimal information component, which only enables the formal definition of clinical content by clinicians. It (openEHR) allows different clinical information systems to be interfaced. The subsequent literature pays attention to interoperability components – archetypes. For example, Chen, Klein, Sundvall, Karlsson, and Åhlfeldt (2009) investigate the exchange of openEHR archetypes between systems, while Martínez-Costa, Menárguez-Tortosa, and Fernández-Breis (2010) examine the semantic interoperability of two EHR standards (openEHR and ISO EN 13606). In this phase, aside from archetypes, some studies pay attention to practical web technologies. For example, Martinez-costaMTA2009 suggests transforming Archetype Definition Language (ADL) into the Ontology Web Language (OWL) for EHR architecture (Martínez-Costa, et al., 2009). By doing so, practitioners are able to take advantage of the already existing infrastructure by adding a new layer, thus enabling heterogeneous systems interoperability. Considering the modeling errors in archetypes, the discussions then shift to the topics of establishing formal methods for validating archetypes. Archetype validation plays an increasingly significant role in the achievement of semantic interoperability in healthcare (Braun, Brandt, Schulz, & Boeker, 2014; Menárguez-Tortosa & Fernández-Breis, 2013).

Many studies in the third phase mention the dual-model standard, hinting at the significance of the practice of modeling data. Discussions include modeling Hospital Information Systems on a regional level based on national logical infrastructure (Pahl, et al., 2015), identifying key socio-technical challenges of the openEHR implementation (Christensen & Ellingsen, 2016), and adopting

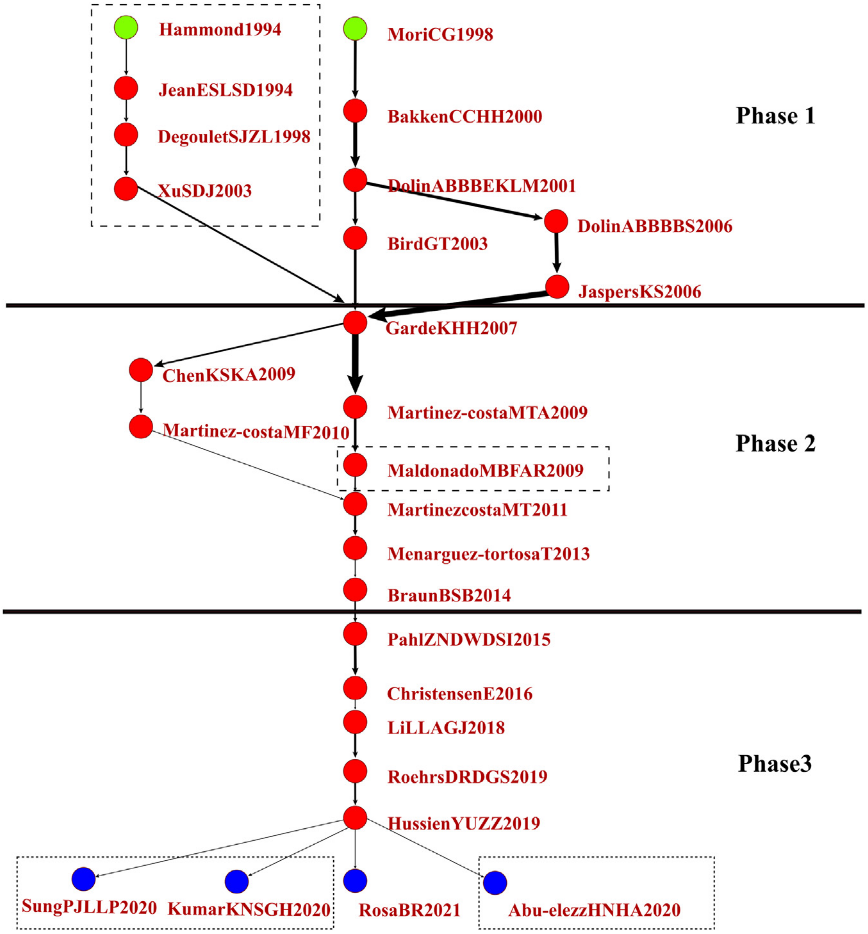

Fig. 4. Traditional main paths of health interoperability studies.

simpler metadata modeling to improve the usability of openEHR (Li, et al., 2018). The studies at the end of main paths pay less attention to standard issues and focus more on emerging technology applications (see the dotted area at the bottom of Figure 4). They suggest deploying blockchain technology applications to improve interoperability standards (Abu-Elezz, Hassan, Nazeemudeen, Househ, & Abd-Alrazaq, 2020; Kumar, et al., 2020; Sung, et al., 2020).

4.3.2. Relevance-based MPA

Despite traditional MPA revealing the importance of a standard in health interoperability main paths, there still are some low-relevance articles that make researchers hard to label the three phases. To trace more convergent main paths, this study adopts relevance-based MPA to eliminate some low-relevance articles, presented in Figure 5. Based on our analysis, the main paths can be divided into three topics: interoperability standard concepts, interoperability standard components, and interoperability standard practices. The discussions begin with issues of standard formation on interoperability and then shift to interoperability components. The most recent articles explore the practical application of interoperability standards.

These relevance-based main paths deviate from the traditional one in three aspects. First, a new source node, which presents the draft proposal of HL7 Document Patient Record Architecture (Dolin, et al., 1999), is added. The addition of this article helps us understand the root of HL7 Architecture and to not take the standard establishment for granted.

Second, a new path with 5 articles that mainly focus on the heterogeneity of clinical data sources, is brought into the picture. The first three of these five articles discuss the interaction of interoperability components, such as transforming source data to meet other health interoperability standards (José Alberto Maldonado, et al., 2012), establishing the interoperability of clinical decision-support systems and EHRs (Marcos, Maldonado, Martínez-Salvador, Boscá, & Robles, 2013), as well as leveraging EHR data for the identification of patient cohorts (Fernández-Breis, et al., 2013). The other two articles, under the context of chronic condition monitoring, suggest that communication practices between the personal health record and heterogeneous data consumers challenge the health interoperability standard (Marcos, González-Ferrer, Peleg, & Cavero, 2015)and stimulate an increased interest in the implementation issue of a distributed and interoperable personal data standard (Roehrs, Da Costa, & da Rosa Righi, 2017).

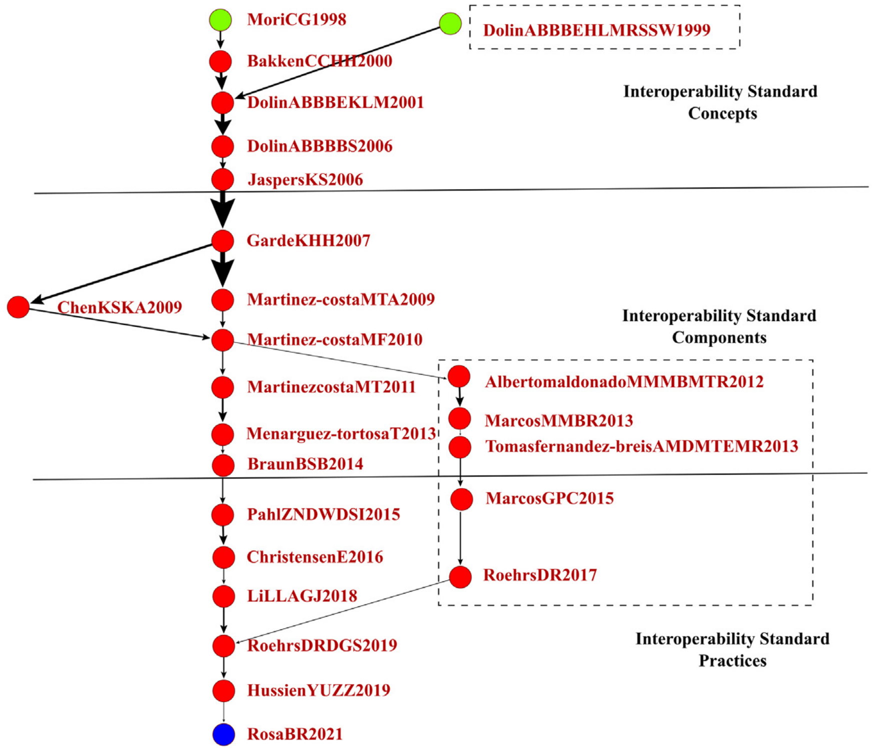

Fig. 5. Relevance-based main paths of health interoperability studies.

Third, a research stream on the traditional main paths (see the dotted area at the top left corner of Figure 4) is eliminated, because this stream does not pay much attention to the standard issues, but focuses on the integration practices of medical applications. By eliminating low-relevance articles, researchers are able to focus on standard issues in health interoperability.

To sum up, such differences help researchers obtain more convergent main paths. These relevance-based main paths not only eliminate low-relevance articles, but also frame the topics on the main paths to a specific concept.

5. Discussion

One of the critical components of relevance-based MPA is the key phrase set. This section investigates the effects of key phrase sets on relevance-based MPA. We first compare the MPA results among three key phrase sets that contain different numbers of terms yet all embed the same concepts. Second, we examine how a key phrase set that contain only one specific concept would affect the MPA results.

5.1. *Comparison of Three Key phrase Sets*

This study uses two article features, citation structure and key phrase, to compare document similarity. While the former is objectively framed by article authors' discretion, the latter is determined subjectively as key phrases are chosen based on one's assessment of the domain knowledge. For example, why does one choose to include the term "medical device" rather than the more general term "device"? Why does one recognize "service provider", "supplier", and "vendor" as synonyms? As such, the similarity results can be sensitive to the chosen key phrase set. In practice, a hybrid paradigm of the subjective and objective approaches is adopted to ensure the credibility of the main paths. Review papers and highly cited papers provide an objective way of identifying and selecting the key phrases while researchers' prior experience play a subjective role in the selection process of key phrases. In order to assess how sensitive relevance-based MPA is to the key phrases, we design two additional sets, which include 15 (small) and 25 (large) key phrases that embed the same concept as the original 20 (medium) set, and compare their main paths with that obtained from the original key phrase set.

Table 2 lists the key phrases in all three sets. The original set includes standard-related key phrases largely used in medical interoperability literature as well as the terms associated with the purpose, concept, and medical informatics of an interoperable standard, such as health portal, medical device, etc. The small set maintains the medical informatics key phrases, but excludes some medical informatics terms. The large set, on the other hand, adds more medical informatics key phrases.

Table 2
Three standard-related key phrase sets.

Set	Number of key phrases	Standard-related key phrases
Small	15	standard, architecture, infrastructure, model, archtype, HL7, semantic, heterogeneous, database, ontology, EHR, EMR, platform, information exchange, data exchange
Medium	20	standard, architecture, infrastructure, model, archtype, HL7, semantic, heterogeneous, database, ontology, EHR, EMR, platform, information exchange, data exchange, health portal, medical portal, medical device, regional health system, integrate
Large	25	standard, architecture, infrastructure, model, archtype, HL7, semantic, heterogeneous, database, ontology, EHR, EMR, platform, information exchange, data exchange, health portal, medical portal, medical device, regional health system, integrate, implement, design, develop, information technology, information system

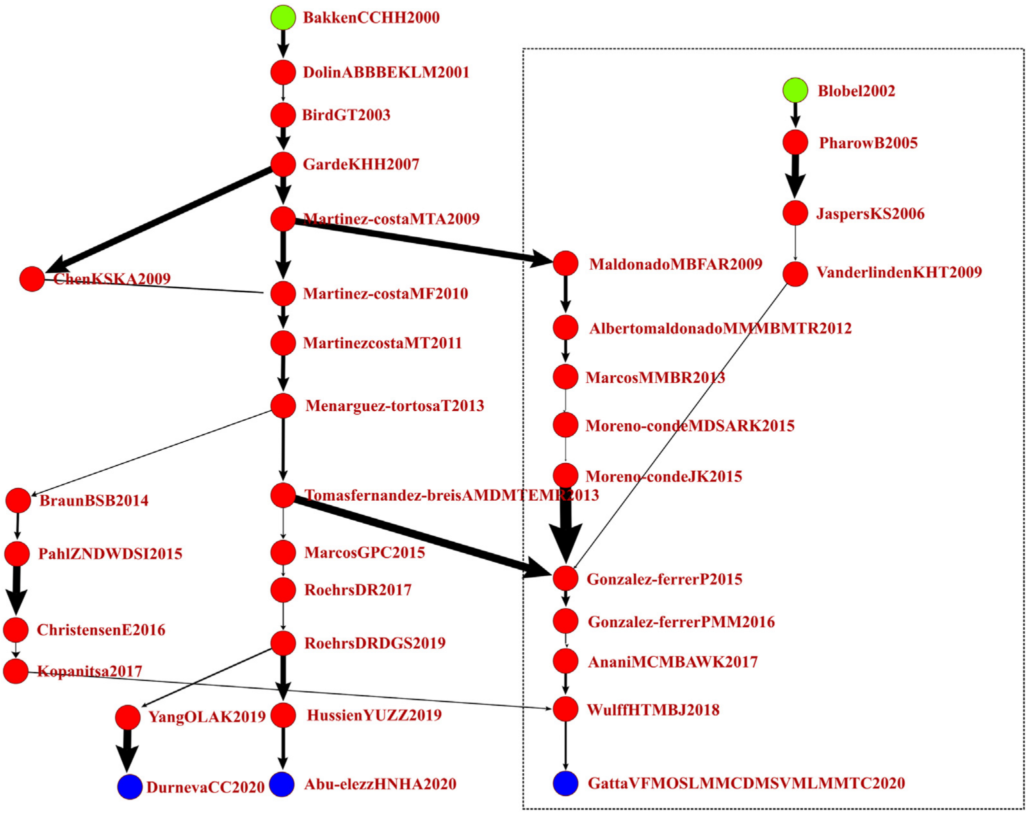

Fig. 6. Privacy-related interoperability MPA.

After applying relevance-based MPA, the main paths obtained from these three key phrase sets turn out to be exactly the same. This result indicates that, with reasonably good judgement on key phrase selection, relevance-based MPA is not that sensitive to the key phrase set.

5.2. Privacy Domain of Relevance-based MPA

The topics of the articles on the main paths can largely focus on the concepts conveyed by the terms in the key phrase set, which can be used as a tool to frame the main paths to a domain that one wishes to observe. For the purpose of demonstration, we set up another key phrase set that transforms standard-related terms into privacy-related terms, such as privacy, security, trust, authentication, and so on. Figure 6 presents the results of relevance-based MPA given this privacy-related key phrase set. The main paths heavily exhibit privacy-related interoperability development as it includes two development paths (see the right of Figrue7) in addition to the existing main paths.

The articles on the path from (Maldonado, et al., 2009) to (Gatta, et al., 2020) discuss security and privacy issues in practical-oriented studies associated with clinical archetypes (Maldonado, et al., 2012), clinical decision-support systems (Marcos, et al., 2013), and the encoding of clinical statements (González-Ferrer, Peleg, Marcos, & Maldonado, 2016). The articles on another path, from

Blobel 2002 (Bernd Blobel, 2002) to VanderlindenKHT2009 (van der Linden, Kalra, Hasman, & Talmon, 2009), focus on security-related interoperability issues, including secure EHR (Blobel, 2002), electronic signatures based on asymmetric cryptography (Pharow & Blobel, 2005), privacy, and security goals of computerized patient records (Jaspers, Knaup, & Schmidt, 2006).

In summary, framing the key phrases not only helps researchers pay attention to a specific concept out of a wide range of research domains, but also helps them to explore the relationship between the specific issue and the domain to where it belongs. For example, focusing on the privacy issue in the health interoperability literature allows us to conduct a comparison between general interoperability articles and the interoperability articles associated with privacy issues. In addition, researchers can dig into privacy issues in the health interoperability literature and obtain the “with privacy” paths within the context of overall paths.

6. Conclusions

Considering the difference between the level of relevance among citations, this study proposes relevance-based MPA as it can provide more convergent main paths. The level of relevance, for the purpose of demonstration, is defined by aggregating structural similarity and key phrase similarity. Traversal counts in traditional MPA are then adjusted according to the level of relevance. Citations are therefore treated unequally as they should be. As a result, articles of low relatedness on a topic can be removed from the main paths. Additionally, one is able to use the approach to frame main paths to address only topics of interest. This study demonstrates the new approach with the case of the health interoperability field, presenting that the new approach is capable of identifying more convergent main paths as well as framing the topics on the main paths to a “standard” concept without being blurred by some low-relevance sub-concepts.

Our analysis is subject to several limitations. First, this study demonstrates the relevance-based approach with a field involving a relatively small citation network and thus adopts flat adjustment. However, the flat adjustment may not be appropriate to a sizable citation network. As such, future studies can examine the appropriateness of these adjustment approaches for different sample sizes. Second, although this study follows a hybrid paradigm of the subjective and objective key phrase selection approaches to improve the main path results, we may still miss some key phrases, because of the large body of knowledge. Third, emphasizing relevance itself makes an assumption that development in science and technology is a continuous process. However, science and technology advances not only through continuous process but also through discontinuous process, which may destroy past achievements and create new paths (Lin, Evans, & Wu, 2022). Discontinuous concepts or disrupting innovations is likely to have low relevance with others in citation structure and key phrases, thus causing these innovative concepts not to be included in the results of this approach. Fourth, there is a need to explore the advantages of similarity algorithms other than Jaccard. While our study does not explicitly elaborate on the potential of various similarity algorithms, it does set up a foundation that future research may build upon. Specifically, future investigations can investigate the feasibility of applying other algorithms, such as cosine similarity, Euclidean distance, etc., to achieve a fairer relevance.

The contribution of this study is twofold. First, we suggest that considering citation relevance provides more convergent MPA results. Low-relevance articles that have low similarity with others in citation structure and key phrases can be removed from the main paths. We demonstrate how this can be done through simple algorithms that compare the similarity in citation structure and key phrases. One can nevertheless apply more sophisticated algorithms in the future to better estimate the relevance between citation pairs. Second, we show that relevance-based MPA can also be used to frame the topics on the main paths to a specific concept through careful selection of the key phrase set. The design of key phrases is in fact a tool for zooming in when examining specific topics or zooming out when observing a broader subject in a field.

CRediT authorship contribution statement

Chen-Hao Huang: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **John S. Liu:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Mei Hsiu-Ching Ho:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Tzu-Chuan Chou:** Writing – review & editing.

Acknowledgments

We thank three anonymous reviewers for their constructive comments, which have greatly improved the accuracy and readability of this article. This work is partially supported by Taiwan’s Ministry of Science and Technology grants MOST 110-2410-H-011-020 and 109-2410-H-011-020-MY2.

References

Abu-Elezz, I., Hassan, A., Nazeemudeen, A., Househ, M., & Abd-Alrazaq, A. (2020). The benefits and threats of blockchain technology in healthcare: A scoping review. *International Journal of Medical Informatics, 142*, Article 104246.

Aman, V. (2018). A new bibliometric approach to measure knowledge transfer of internationally mobile scientists. *Scientometrics, 117*(1), 227–247.

Bakken, S., Campbell, K. E., Cimino, J. J., Huff, S. M., & Ed Hammond, W. (2000). Toward vocabulary domain specifications for Health Level 7—Coded data elements. *Journal of the American Medical Informatics Association, 7*(4), 333–342.

Bhupatiraju, S., Nomaler, Ö., Triulzi, G., & Verspagen, B. (2012). Knowledge flows—Analyzing the core literature of innovation, entrepreneurship and science and technology studies. *Research policy, 41*(7), 1205–1218.

Blobel, B. (2002). Architecture of secure portable and interoperable electronic health records. In *International conference on computational science* (pp. 982–994). Springer.

Blobel, B., Engel, K., & Pharowe, P. (2006). Semantic interoperability. *Methods of information in medicine, 45*(04), 343–353.
Bonzi, S. (1982). Characteristics of a literature as predictors of relatedness between cited and citing works. *Journal of the American Society for Information Science, 33*(4), 208–216.
Braun, M., Brandt, A. U., Schulz, S., & Boeker, M. (2014). Validating archetypes for the multiple sclerosis functional composite. *BMC medical informatics and decision making, 14*(1), 1–12.
Chan, H. F., Önder, A. S., & Torgler, B. (2016). The first cut is the deepest: Repeated interactions of coauthorship and academic productivity in Nobel laureate teams. *Scientometrics, 106*(2), 509–524.
Chen, R., Klein, G. O., Sundvall, E., Karlsson, D., & Åhlfeldt, H. (2009). Archetype-based conversion of EHR content models: Pilot experience with a regional EHR system. *BMC medical informatics and decision making, 9*(1), 1–13.
Christensen, B., & Ellingsen, G. (2016). Evaluating model-driven development for large-scale EHRs through the openEHR approach. *International Journal of Medical Informatics, 89*, 43–54.
Chuang, T. C., Liu, J. S., Lu, L. Y., & Lee, Y. (2014). The main paths of medical tourism: From transplantation to beautification. *Tourism Management, 45*, 49–58.
Ciotti, V., Bonaventura, M., Nicosia, V., Panzarasa, P., & Latora, V. (2016). Homophily and missing links in citation networks. *EPJ Data Science, 5*, 1–14.
Cronin, B. (1994). Tiered citation and measures of document similarity. *Journal of the American Society for Information Science, 45*(7), 537–538.
Degoulet, P., Sauquet, D., Jaulent, M.-C., Zapletal, E., & Lavril, M. (1998). Rationale and design considerations for a semantic mediator in health information systems. *Methods of information in medicine, 37*(04/05), 518–526.
Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology, 65*(9), 1820–1833.
Dolin, R. H., Alschuler, L., Beebe, C., Biron, P. V., Boyer, S. L., Essin, D., et al., (2001). The HL7 clinical document architecture. *Journal of the American Medical Informatics Association, 8*(6), 552–569.
Dolin, R. H., Alschuler, L., Behlen, F., Biron, P. V., Boyer, S., Essin, D., et al., (1999). HL7 document patient record architecture: An XML document architecture based on a shared information model. In *Proceedings of the AMIA symposium* (p. 52). American Medical Informatics Association.
Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., & Shabo, A. (2006). HL7 clinical document architecture, release 2. *Journal of the American Medical Informatics Association, 13*(1), 30–39.
Fernández-Breis, J. T., Maldonado, J. A., Marcos, M., Legaz-García, M. d. C., Moner, D., Torres-Sospedra, J., et al., (2013). Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts. *Journal of the American Medical Informatics Association, 20*(e2), e288–e296.
Frontoni, E., Mancini, A., Baldi, M., Paolanti, M., Moccia, S., Zingaretti, P., … Misericordia, P. (2019). Sharing health data among general practitioners: The Nu. Sa. project. *International Journal of Medical Informatics, 129*, 267–274.
Garde, S., Knaup, P., Hovenga, E. J., & Heard, S. (2007). Towards semantic interoperability for electronic health records. *Methods of information in medicine, 46*(03), 332–343.
Gatta, R., Vallati, M., Fernandez-Llatas, C., Martinez-Millana, A., Orini, S., Sacchi, L., et al., (2020). What role can process mining play in recurrent clinical guidelines issues? A position paper. *International Journal of Environmental Research and Public Health, 17*(18), 6616.
Gay, V., & Leijdekkers, P. (2015). Bringing health and fitness data together for connected health care: Mobile apps as enablers of interoperability. *Journal of medical Internet research, 17*(11), e260.
González-Ferrer, A., Peleg, M., Marcos, M., & Maldonado, J. A. (2016). Analysis of the process of representing clinical statements for decision-support applications: A comparison of openEHR archetypes and HL7 virtual medical record. *Journal of medical systems, 40*(7), 1–10.
Gordon, W. J., & Catalini, C. (2018). Blockchain technology for healthcare: Facilitating the transition to patient-driven interoperability. *Computational and structural biotechnology journal, 16*, 224–230.
Heart, T., Ben-Assuli, O., & Shabtai, I. (2017). A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. *Health Policy and Technology, 6*(1), 20–25.
Ho, M. H.-C., & Liu, J. S. (2021). The swift knowledge development path of COVID-19 research: The first 150 days. *Scientometrics, 126*(3), 2391–2399.
Hu, Z., Chen, C., & Liu, Z. (2015). The recurrence of citations within a scientific article. *ISSI*.
Huang, M.-h., & Chang, C.-P. (2015). A comparative study on detecting research fronts in the organic light-emitting diode (OLED) field using bibliographic coupling and co-citation. *Scientometrics, 102*(3), 2041–2057.
Hudson, F., & Clark, C. (2018). Wearables and medical interoperability: The evolving frontier. *Computer, 51*(9), 86–90.
Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social networks, 11*(1), 39–63.
Hummon, N. P., Doreian, P., & Freeman, L. C. (1990). Analyzing the structure of the centrality-productivity literature created between 1948 and 1979. *Knowledge, 11*(4), 459–480.
Hung, S.-C., Lai, J.-Y., & Liu, J. S. (2022). Mapping technological trajectories as the main paths of knowledge flow: Evidence from printers. *Industrial and Corporate Change, 31*(3), 863–889.
Hung, S.-C., Liu, J. S., Lu, L. Y., & Tseng, Y.-C. (2014). Technological change in lithium iron phosphate battery: The key-route main path analysis. *Scientometrics, 100*(1), 97–120.
Jaspers, M., Knaup, P., & Schmidt, D. (2006). The computerized patient record: Where do we stand. *Yearbook of medical informatics, 1*, 29–39.
Jean, F.-C., Engelmann, U., Sauquet, D., Lavril, M., Schröter, A., & Degoulet, P. (1994). The HELIOS medical connection services. *Computer methods and programs in biomedicine, 45*, S117–S126.
Joda, T., Waltimo, T., Probst-Hensch, N., Pauli-Magnus, C., & Zitzmann, N. U. (2019). Health data in dentistry: An attempt to master the digital challenge. *Public Health Genomics, 22*(1-2), 1–7.
Kim, E. H., Jeong, Y. K., Kim, Y., & Song, M. (2022). Exploring scientific trajectories of a large-scale dataset using topic-integrated path extraction. *Journal of Informetrics, 16*(1), Article 101242.
Kim, M., Baek, I., & Song, M. (2018). Topic diffusion analysis of a weighted citation network in biomedical literature. *Journal of the Association for Information Science and Technology, 69*(2), 329–342.
Krampen, G., Becker, R., Wahner, U., & Montada, L. (2007). On the validity of citation counting in science evaluation: Content analyses of references and citations in psychological publications. *Scientometrics, 71*(2), 191–202.
Kumar, A., Krishnamurthi, R., Nayyar, A., Sharma, K., Grover, V., & Hossain, E. (2020). A novel smart healthcare design, simulation, and implementation using healthcare 4.0 processes. *IEEE Access, 8*, 118433–118471.
Li, B., Li, J., Lan, X., An, Y., Gao, W., & Jiang, Y. (2018). Experiences of building a medical data acquisition system based on two-level modeling. *International Journal of Medical Informatics, 112*, 114–122.
Lin, Y., Evans, J. A., & Wu, L. (2022). New directions in science emerge from disconnection and discord. *Journal of Informetrics, 16*(1), Article 101234.
Liu, J. S., Chen, H. H., Ho, M. H. C., & Li, Y. C. (2014). Citations with different levels of relevancy: Tracing the main paths of legal opinions. *Journal of the Association for Information Science and Technology, 65*(12), 2479–2488.
Liu, J. S., & Lu, L. Y. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology, 63*(3), 528–542.
Liu, J. S., Lu, L. Y., & Ho, M. H.-C. (2019). A few notes on main path analysis. *Scientometrics, 119*(1), 379–391.
Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of mathematical sociology, 1*(1), 49–80.
Luor, T., Lu, H.-P., Yu, H., & Chang, K. (2014). Trends in and contributions to entrepreneurship research: A broad review of literature from 1996 to June 2012. *Scientometrics, 99*(2), 353–369.
Maldonado, J. A., Costa, C. M., Moner, D., Menárguez-Tortosa, M., Boscá, D., Giménez, J. A. M., et al., (2012). Using the ResearchEHR platform to facilitate the practical application of the EHR standards. *Journal of biomedical informatics, 45*(4), 746–762.

Maldonado, J. A., Moner, D., Boscá, D., Fernández-Breis, J. T., Angulo, C., & Robles, M. (2009). LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics. *International Journal of Medical Informatics, 78*(8), 559–570.

Marcos, González-Ferrer, A., Peleg, M., & Cavero, C. (2015). Solving the interoperability challenge of a distributed complex patient guidance system: A data integrator based on HL7's Virtual Medical Record standard. *Journal of the American Medical Informatics Association, 22*(3), 587–599.

Marcos, M., Maldonado, J. A., Martínez-Salvador, B., Boscá, D., & Robles, M. (2013). Interoperability of clinical decision-support systems and electronic health records using archetypes: A case study in clinical trial eligibility. *Journal of biomedical informatics, 46*(4), 676–689.

Martínez-Costa, C., Menárguez-Tortosa, M., & Fernández-Breis, J. T. (2010). An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. *Journal of biomedical informatics, 43*(5), 736–746.

Martínez-Costa, C., Menárguez-Tortosa, M., Fernández-Breis, J. T., & Maldonado, J. A. (2009). A model-driven approach for representing clinical archetypes for Semantic Web environments. *Journal of biomedical informatics, 42*(1), 150–164.

Masud, M., Hossain, M. S., & Alamri, A. (2012). Data interoperability and multimedia content management in e-health systems. *IEEE transactions on Information Technology in biomedicine, 16*(6), 1015–1023.

Mei, Q., & Zhai, C. (2008). Generating impact-based summaries for scientific literature. In *Proceedings of ACL-08: HLT* (pp. 816–824).

Menárguez-Tortosa, M., & Fernández-Breis, J. T. (2013). OWL-based reasoning methods for validating archetypes. *Journal of biomedical informatics, 46*(2), 304–317.

Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social studies of science, 5*(1), 86–92.

Mori, A. R., Consorti, F., & Galeazzi, E. (1998). Standards to support development of terminological systems for healthcare telematics. *Methods of Information in Medicine-Methodik der Information in der Medizin, 37*(4), 551–563.

Pahl, C., Zare, M., Nilashi, M., de Faria Borges, M. A., Weingaertner, D., Detschew, V., … Ibrahim, O. (2015). Role of OpenEHR as an open source solution for the regional modelling of patient data in obstetrics. *Journal of biomedical informatics, 55*, 174–187.

Peters, H., Braam, R. R., & van Raan, A. F. (1995). Cognitive resemblance and citation relations in chemical engineering publications. *Journal of the American Society for Information Science, 46*(1), 9–21.

Pharow, P., & Blobel, B. (2005). Electronic signatures for long-lasting storage purposes in electronic archives. *International Journal of Medical Informatics, 74*(2-4), 279–287.

Pichappan, P., & Sarasvady, S. (2002). The other side of the coin: The intricacies of author self-citations. *Scientometrics, 54*(2), 285–290.

Roehrs, A., Da Costa, C. A., & da Rosa Righi, R. (2017). OmniPHR: A distributed architecture model to integrate personal health records. *Journal of biomedical informatics, 71*, 70–81.

Sung, M., Park, S., Jung, S., Lee, E., Lee, J., & Park, Y. R. (2020). Developing a mobile app for monitoring medical record changes using blockchain: Development and usability study. *Journal of medical Internet research, 22*(8), e19657.

Trivison, D. (1987). Term co-occurrence in cited/citing journal articles as a measure of document similarity. *Information processing & management, 23*(3), 183–194.

van der Linden, H., Kalra, D., Hasman, A., & Talmon, J. (2009). Inter-organizational future proof EHR systems: A review of the security and privacy related issues. *International Journal of Medical Informatics, 78*(3), 141–160.

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in complex systems, 10*(01), 93–115.

Webster, L., & Spiro, R. F. (2010). Health information technology: A new world for pharmacy. *Journal of the American Pharmacists Association, 50*(2), e20–e34.

Xiao, Y., Lu, L. Y., Liu, J. S., & Zhou, Z. (2014). Knowledge diffusion path analysis of data quality literature: A main path analysis. *Journal of Informetrics, 8*(3), 594–605.

Xu, Y., Sauquet, D., Degoulet, P., & Jaulent, M.-C. (2003). Component-based mediation services for the integration of medical applications. *Artificial intelligence in medicine, 27*(3), 283–304.

Yang, C., Chou, T.-C., & Chen, Y.-H. (2019). Bridging digital boundary in healthcare systems—An interoperability enactment perspective. *Computer Standards & Interfaces, 62*, 43–52.

Yu, D., & Pan, T. (2021). Tracing the main path of interdisciplinary research considering citation preference: A case from blockchain domain. *Journal of Informetrics, 15*(2), Article 101136.