Content Characteristics of Knowledge Integration in the eHealth Field: An Analysis Based on Citation Contexts

Shiyun Wang^{1,2}, Jin Mao^{1,2†}, Jing Tang^{1,2}, Yujie Cao³

Citation: Wang, S.Y., Mao, J., Tang, J., & Cao, Y.J. (2021). Content characteristics of knowledge integration in the eHealth field: An analysis based on citation contexts. *Journal of Data and Information Science*, 6(3), 58–74. <https://doi.org/10.2478/jdis-2021-0015>

Received: Nov. 1, 2020
Revised: Dec. 29, 2020;
Jan. 13, 2021;
Feb. 2, 2021
Accepted: Feb. 5, 2021

¹Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China

²School of Information Management, Wuhan University, Wuhan 430072, China

³School of Information Management, Central China Normal University, Wuhan 430079, China

Abstract

Purpose: This study attempts to disclose the characteristics of knowledge integration in an interdisciplinary field by looking into the content aspect of knowledge.

Design/methodology/approach: The eHealth field was chosen in the case study. Associated knowledge phrases (AKPs) that are shared between citing papers and their references were extracted from the citation contexts of the eHealth papers by applying a stem-matching method. A classification schema that considers the functions of knowledge in the domain was proposed to categorize the identified AKPs. The source disciplines of each knowledge type were analyzed. Quantitative indicators and a co-occurrence analysis were applied to disclose the integration patterns of different knowledge types.

Findings: The annotated AKPs evidence the major disciplines supplying each type of knowledge. Different knowledge types have remarkably different integration patterns in terms of knowledge amount, the breadth of source disciplines, and the integration time lag. We also find several frequent co-occurrence patterns of different knowledge types.

Research limitations: The collected articles of the field are limited to the two leading open access journals. The stem-matching method to extract AKPs could not identify those phrases with the same meaning but expressed in words with different stems. The type of Research Subject dominates the recognized AKPs, which calls on an improvement of the classification schema for better knowledge integration analysis on knowledge units.

Practical implications: The methodology proposed in this paper sheds new light on knowledge integration characteristics of an interdisciplinary field from the content perspective. The findings have practical implications on the future development of research strategies in eHealth and the policies about interdisciplinary research.

Originality/value: This study proposed a new methodology to explore the content characteristics of knowledge integration in an interdisciplinary field.

[†] Corresponding author: Jin Mao (E-mail: maojin@whu.edu.cn).

Keywords Knowledge integration; Interdisciplinary research; Citation contexts; eHealth; Knowledge content

1 Introduction

In recent years, many major scientific research problems are complex and cannot be solved by a single field. Interdisciplinary research (IDR) has gradually become an essential mode in modern science, and received extensive attention from researchers and policymakers (Porter et al., 2006; Wagner et al., 2011; Xu et al., 2016; Xu et al., 2018). Interdisciplinary research that integrates knowledge units, such as theories, techniques, and data, from multiple research bodies of specialized knowledge or research practice (Porter et al., 2006), could create a holistic view or stimulate new ideas to solve complicated scientific problems. Knowledge integration is of nature an important phenomenon in IDR. Exploring its characteristics could further our understanding about the mechanism of IDR to facilitate the progress of scientific development.

Current studies have investigated the knowledge integration of interdisciplinary research from various perspectives. Porter et al. (2007) proposed an “integration” metric to measure the interdisciplinarity of a research article according to subject categories of its references. However, they did not consider the content of references. A few recent studies have attempted to discern interdisciplinary topics in an interdisciplinary field by using co-word analysis (Ba et al., 2019) and cluster analysis based on co-citation networks (Chi & Young, 2013). These approaches rely heavily on expert wisdom to determine domain-specific knowledge and to interpret each cluster. Alternatively, text mining methods that could automatically identify interdisciplinary topics from scientific text, such as keyword mining and topic modeling, have gradually attracted a lot of attention (Nichols, 2014; Xu et al., 2016). Nevertheless, these approaches do not reveal explicit evidence about what knowledge from the references is integrated by citing articles.

Citation contexts, which contain contextual information of citations, could provide rich information for the analysis of what knowledge has been integrated through citations. Recently, Mao et al. (2020) proposed a new approach to identify the knowledge phrases shared between citation contexts and their corresponding references in an interdisciplinary field, which can be regarded as explicit symbols of knowledge spread from cited papers to citing papers. By identifying the integrated knowledge units, knowledge integration in an interdisciplinary field could be measured and analyzed quantitatively. In this paper, we take the eHealth field as a case of interdisciplinary field (Eysenbach, 2001). A classification schema that considers the functions of knowledge units in the field is proposed to categorize the

Research Paper

identified AKPs in the eHealth field. We attempt to address the following research questions:

RQ#1 What are the highly contributed disciplines for each knowledge type? Do the disciplines vary among different knowledge types?

RQ#2 What are the integration characteristics of different types of knowledge in the eHealth field? And, how have they been changing over time?

The answers to these questions could offer a fine granular perspective for understanding knowledge functions of source disciplines in the eHealth field as well as the dynamic knowledge integration process in the eHealth field.

2 Methodology

2.1 Data collection

We selected two leading journals in the eHealth field, *Journal of Medical Internet Research* (JMIR) and *JMIR mHealth and uHealth* (JMU), as our data sources. Our reasons are threefold. First, according to an expert survey of 398 active e-health researchers, JMIR and JMU were ranked as top A+ and top A journals out of 63 peer-reviewed eHealth related journals, respectively (Serenko, Dohan, & Tan, 2017). Second, JMIR was established in 1999, when the eHealth field was just emerging (Della Mea, 2001). This could provide us with a comprehensive understanding about the formation and evolution of the eHealth field. JMU is a newer spin-off journal of JMIR, focusing on more technical and developmental papers than JMIR. It covers more frontier scientific and technological contents in the eHealth field. Third, both JMIR and JMU provide open access articles in XML format. Since we aim at investigating the content characteristics of knowledge integration through citation context analysis, the availability of full text articles is helpful for us to obtain citation contexts. Other journals in the eHealth field often provide PDF-format articles, which require heavy and error-prone text processing to obtain the text content of articles (Bertin et al., 2016).

We collected all papers published by the two journals from 1999 to 2018, and selected 3,221 articles with the type of “original papers”, “reviews”, and “viewpoints”. Other types of articles, such as “Corrigenda and Addenda”, “Editorial”, and “Letter to the Editor”, which list fewer references, were excluded.

2.2 Data pre-processing

For each article, we parsed the metadata (DOI, publish year, etc.), bibliography information (title, PMID, journal, publish year, etc.), and citation contexts. The context of a citation in this study is defined as the sentence where the citation occurs rather than a longer text span so that the association between the citation context and its corresponding reference will be closer (Small, Tseng, & Patekc, 2017).

We augmented the metadata information (abstract, keyword, Keyword Plus, MeSH term) of the references by linking them to Web of Science (WoS) and PubMed. The disciplines of the references were determined as the WoS subject categories of the journal where it was published. The references without WoS subject categories were not analyzed.

In total, 119,598 citation sentences were obtained, as well as 101,751 reference records (i.e. bibliographic items) with metadata information, which account for 93.00% of all journal references and 72.38% of all references.

2.3 AKPs identification and classification

Most previous studies used expert knowledge to identify cited objects in citation sentences by human annotation, which were then applied to investigate the domain knowledge used in interdisciplinary research (Wang & Zhang, 2018). In this study, we used an automatic approach proposed in our previous study (Mao, Wang, & Shang, 2020) to identify associated knowledge phrases (AKPs), which can be regarded as explicit integrated knowledge content spread from references to citing papers.

The approach extracts noun phrases from citation sentences as well as titles and abstracts of references by using spaCy, an open-source natural language processing package. Several pre-processing operations were performed before the noun phrases from the two sources were matched. Single characters and the phrases starting or ending with numbers were removed. Author keywords, Keyword Plus terms, and MeSH (Medical Subject Headings) terms in the references are also treated as noun phrases of references. All phrases from the two sources were lemmatized using the NLTK Python package. Next, the noun phrases appearing in each pair of citation sentence and the corresponding reference were compared by our stem-matching approach. The noun phrases between the pair were matched if their stemmed forms were the same. We also matched the stemmed noun phrases extracted from the citation sentence with the stemmed sentences in the corresponding reference (including its title and abstract). Then, we denote the matched noun phrases of the citation sentence as the AKPs. This method recalled 78.57% phrases (209 of all 266 phrases) according to the evaluation on a randomly sampled 100 citation sentences. A total of 246,167 AKPs were extracted from our dataset, with 25,764 distinct ones.

To characterize the knowledge integrated by the interdisciplinary field, we designed a knowledge classification schema to categorize the identified AKPs. Recently, a few studies have attempted to discern the functions of knowledge played in a domain. Ding et al. (2013) pointed out that scientific papers embed many types of micro-level entities, including datasets, methods, and domain-specific entities.

Research Paper

Heffernan and Teufel (2018) focus on the identification of problems and solutions in scientific text. Lu et al. (2019) proposed a classification schema for author selected keywords, reflecting how they function semantically in scientific manuscripts. To favor the investigation of micro-level knowledge integration relationships, we also designed a knowledge classification schema based on the functions of knowledge in scientific articles.

We recruited two graduate students to annotate the types of all distinct AKPs based on the knowledge classification schema in Table 1. Each distinct AKP and one of its citation sentences that was randomly selected were given for the coders. Some examples are given in Table 2. First, two coders independently annotated 500 identical randomly selected knowledge phrases for pre-annotation. However, the kappa coefficient between the annotation of two coders was only 0.65. Therefore, an expert in the eHealth field was invited to guide the annotation work and helped the coders to distinguish the ambiguous cases. We found that some phrases could be labeled into different categories in different contexts. To avoid ambiguity, we only considered the frequently used meaning of the term in our annotation process. After discussion, two coders reached a consensus. Then, they independently annotated all 24,132 unique phrases that are associated with the disciplines of our interests. During the annotation process, two coders kept in communication with each other to reach an agreement. Among all 24,132 distinct phrases annotated in our previous study (Mao, Wang, & Shang, 2020), 24,063 distinct phrases were related to the WoS subject categories of this study’s interest, and another 1,701 distinct AKPs from the remaining references were annotated by the two coders in the same way for this study.

Table 1. The knowledge classification schema for AKPs.

Category	Description	Literature sources
Research Subject	subject terms related to research problems, such as diseases and research areas.	Heffernan & Teufel, 2018; Kondo et al., 2009
Theory	theory related phrases, e.g., specific names of theories, and frameworks	Wang & Zhang, 2018; Pettigrew & McKechnie, 2001
Research Methodology	research methodology, including research methods, scales, guidelines, evaluation indicators, etc.	Sahragard & Meihami, 2016; Heffernan & Teufel, 2018; Mesbah et al., 2017; Radoulov, 2008;
Technology Entity	techniques, devices, and systems people or organizations that are involved in any aspect of the research	Gupta & Manning, 2011; Tsai et al., 2013 Bahadoran et al., 2019
Data	phrases related to datasets, data sources, and data material	Wang & Zhang, 2018; Sahragard & Meihami, 2016; Mesbah et al., 2017; Radoulov, 2008
Others	other phrases that are not included in the above categories, e.g., geolocations, projects, etc.	Kondo et al., 2009

Table 2. Annotation example of each knowledge category.

AKPs	Citation sentences	Knowledge type
<i>chronic illness</i>	<i>For effective medical care of chronic illness, such as Type 2 diabetes mellitus (T2DM), adequate and sustainable self-management initiated by patients is important</i>	Research Subject
<i>social cognitive theory</i>	<i>The intervention, including both the SMS text messaging and individual counseling session, was modeled after national treatment guidelines, and guided by Social Cognitive Theory and the stages of change model</i>	Theory
<i>qualitative research methodology</i>	<i>In recent years, qualitative research methodology has become more recognized and valued in diabetes behavioral research because it helps answer questions that quantitative research might not, by exploring patient motivations, perceptions, and expectations</i>	Research Methodology
<i>SMS text messaging</i>	<i>Consistent with the literature, SMS text messaging was an appropriate and accepted tool to deliver health promotion content</i>	Technology
<i>heart failure patient</i>	<i>De Vries et al (2013) evaluated the actual use and goals of telemonitoring systems, whereas Seto et al (2012) developed a randomized trial of mobile phone-based telemonitoring systems to examine the experience of heart failure patients with these systems</i>	Entity
<i>bacteriology datum</i>	<i>PDA-based technologies were used to develop a PDA-based electronic system to collect, verify, and upload bacteriology data into an electronic medical record system; develop a wireless clinical care management system; and develop a data collection/entry system for public surveillance data collection</i>	Data
<i>low risk</i>	<i>Free et al found that while mHealth studies have been conducted many are of poor quality, few have a low risk of bias, and very few have found clinically significant benefits of the interventions</i>	Others

2.4 Measuring knowledge integration patterns

We introduce several indicators to measure the integration characteristics of different types of knowledge based on the identified AKPs. The indicators are defined as follows:

- Knowledge amount: the number of AKPs.
- Knowledge integration density: the average number of AKPs per reference.
- Number of references: the number of references carrying the AKPs.
- Number of source disciplines: the number of distinct disciplines with references carrying the AKPs.
- Citation interval: the citation interval of the in-text citation where the AKPs appear. It is defined as the time distance between the publication year of the citing paper and the cited paper (Otto et al., 2019), which represents the integration time lag of the knowledge. We calculated the average citation interval for each type of AKPs.

To further understand the relationship of different knowledge in the integration process, we also analyzed the co-occurrence of different types of knowledge in the same citation contexts.

3 Results and discussion

3.1 Identified AKPs

The descriptive information of our dataset is shown in Table 3. From the dataset, 119,598 citation sentences and 101,751 references with metadata information were extracted. Since a citation sentence may contain more than one in-text citation (Small, Tseng, & Patekc, 2017), the number of in-text citations (199,461) exceeds the number of citation sentences. In total, we obtained 246,167 AKPs with 25,764 distinct ones.

Table 3. Brief information of our dataset.

Statistical items	Value
Citing papers	3,221
Citation sentences	119,598
References	101,751
In-text citations	199,461
AKPs	246,167
Distinct AKPs	25,764

3.2 The classification results of AKPs

The annotation results of AKPs classification are shown in Table 4. The number of references and source disciplines, as well as knowledge integration density and average citation interval, are presented for each knowledge type. It is observed that the knowledge amount for different knowledge types is uneven. The phrases in the category of *Research Subject* are the most, followed by *Others*. The category of *Theory* contains the fewest AKPs, however, the knowledge integration density of *Theory* exceeds that of most other knowledge types, ranking the second place among all knowledge types. This indicates that *Theory* related references may carry more phrases of theories in each citation.

The average citation interval shows that different knowledge types have significantly different time lags. As Table 4 presents, *Theory* related phrases have the longest time lag in the knowledge integration, followed by *Research Methodology*, while *Technology* has the shortest time lag. This result could be explained by that theory and methodology need more time to be verified by the scientific community, while technology is updated rapidly.

Table 4. Integration characteristics of different knowledge types.

Knowledge type	Knowledge amount	Distinct AKPs	References	Source disciplines	Knowledge integration density	Average citation interval
Research Subject	104,988	15,324	51,622	187	2.03	5.91
Entity	25,213	1,665	18,219	150	1.38	5.33
Technology	17,945	1,885	13,256	157	1.35	4.22
Research Methodology	9,099	2,079	6,773	144	1.34	7.74
Data	3,297	296	2,822	124	1.17	5.11
Theory	1,315	225	921	88	1.43	10.55
Others	84,310	4,290	44,346	190	1.90	5.50

3.3 Highly contributed disciplines

We next turn our attention to the source disciplines of each type of AKPs. In this paper, we defined the source disciplines of AKPs as the WoS subject categories of the references carrying the AKPs.

Table 5 illustrates the top 10 highly contributed disciplines with the largest number of AKPs for each knowledge type. Overall, except *Theory*, Health Care Sciences & Services is the largest knowledge provider, followed by Medical Informatics. Nonetheless, the top 10 highly contributed disciplines rank significantly different among the knowledge types. Medical, healthcare, and psychology related disciplines provided the eHealth field with more knowledge about *Research Subject*, *Entity*, and *Research Methodology*, while for *Technology* and *Data*, information and computer science related disciplines contributed more. Psychology and management related disciplines supplied the eHealth field with more AKPs of *Theory*. This demonstrates that different disciplines may play different roles in the formation of the interdisciplinary field of eHealth according to their contributions in different knowledge types.

3.4 Integration patterns of each knowledge type

In this section, we present the integration characteristics in terms of the proposed indicators.

3.4.1 Knowledge amount

Fig. 1 displays the knowledge amount of each knowledge type over time. For every type, the number of AKPs remained stable before 2010 and has been rising since then. This trend is along with the increasing publication tendency of the eHealth papers (Fig. 1a), which reveals the emergence of the eHealth field in recent years. It appears that the category of *Research Subject* has grown the fastest, followed by *Entity* and *Technology*, while *Theory* has grown the slowest. It shows the abundance of research subjects in the interdisciplinary field of eHealth. The

Research Paper

Table 5. Top 10 source disciplines for each knowledge type.

Research Subject	Entity	Technology	Research Methodology	Data	Theory
Health Care Sciences & Services	Health Care Sciences & Services	Health Care Sciences & Services	Health Care Sciences & Services	Health Care Sciences & Services	Public, Environmental & Occupational Health
Medical Informatics	Medical Informatics	Medical Informatics	Medical Informatics	Medical Informatics	Health Care Sciences & Services
Public, Environmental & Occupational Health	Public, Environmental & Occupational Health	Public, Environmental & Occupational Health	Public, Environmental & Occupational Health	Public, Environmental & Occupational Health	Medical Informatics
Medicine, General & Internal Psychiatry	Medicine, General & Internal Psychiatry	Medicine, General & Internal	Psychiatry	Medicine, General & Internal	Psychology, Multidisciplinary
		Computer Science, Information Systems	Medicine, General & Internal	Information Science & Library Science	Management
Psychology, Clinical	Nursing	Information Science & Library Science	Psychology, Clinical	Computer Science, Information Systems	Psychology, Applied
Substance Abuse	Psychology, Clinical	Computer Science, Interdisciplinary Application	Substance Abuse	Computer Science, Interdisciplinary Application	Psychology, Social
Health Policy & Services	Health Policy & Services	Psychiatry	Health Policy & Services	Health Policy & Services	Psychology
Nursing	Substance Abuse	Psychology, Clinical	Psychology	Multidisciplinary Sciences	Psychology, Clinical
Endocrinology & Metabolism	Computer Science, Information Systems	Substance Abuse	Psychology, Multidisciplinary	Psychiatry	Computer Science, Information Systems

highly cited research subjects include “*information*”, “*intervention*”, “*depression*”, “*physical activity*”, “*health*”, “*diabetes*”, etc. These research subjects reflect the research hotspots in the eHealth field from the citation content perspective.

To deeply understand the patterns of different knowledge categories, we further analyzed the proportion of each knowledge type in each year, as shown in Fig. 1b. It is observed that the proportion of every knowledge type has gradually remained stable after the fluctuations in the early years. As the knowledge structure of the eHealth field has been formed over time, the integration pattern of different knowledge types has become relatively fixed. Besides, *Technology* was gradually surpassed by *Entity*, which shows that human beings and related organizations are highly involved in the field.

Figure 1. The knowledge amount distribution for each knowledge type from 1999 to 2018. The panel on the left (a) shows the total number of AKPs for each knowledge type over the period, and the inside subgraph in (a) presents the number of eHealth papers in our dataset between 1999 and 2018. The panel on the right (b) shows the proportion of knowledge amount of each knowledge type in each year.

3.4.2 Number of references

As Fig. 2 presents, similar to the growing trend of knowledge amount, the number of references remained stable before 2010 and has been increasing afterward. For the proportion of references (Fig. 2b), it also shows a similar pattern to the knowledge amount, which remained stable in later years after the fluctuations in early years. This further proves the integration patterns of different types of knowledge have gradually remained stable in recent years.

Figure 2. The number of references with the AKPs. (a), The total number of references with the AKPs for each knowledge type from 1999 to 2018. (b), The proportion of references with the corresponding type of AKPs in each year. The ratio of references for each knowledge type in every year was calculated by the references with the corresponding type of knowledge divided by the total number of references with AKPs in that year. Notably, one reference may contain different types of knowledge.

Research Paper

3.4.3 Number of source disciplines

The number of source disciplines involved by each type of AKPs has continued to grow dramatically since 1999, as shown in Fig. 3a, which demonstrates the increase of interdisciplinarity in the eHealth field. The proportion of distinct source disciplines for each knowledge type also shows an upward trend, and the growth rate has slowed down recently.

Figure 3. The number of source disciplines of the AKPs. (a), The total number of distinct source disciplines with AKPs between 1999 and 2018. (b), The proportion of distinct source disciplines with AKPs for each knowledge type in each year. The ratio of disciplines for each knowledge type in every year was calculated by the distinct disciplines containing the corresponding type of knowledge divided by the total number of distinct disciplines with AKPs in that year. Notably, one distinct discipline may contain different types of knowledge.

3.4.4 Citation interval

Fig. 4 presents the average citation interval of AKPs, which represents the time lag that eHealth integrates these types of knowledge. Overall, the citation interval of every knowledge type increased steadily with the development of the field. This may be due to that some classic publications of pioneering research work in the field would increase the citations in the following years (Sun & Latora, 2020). As a result, the average citation age would increase over time. On the other hand, as shown before (Fig. 3a), the interdisciplinary character of the eHealth field has been rising over time. Since the cross-disciplinary knowledge flow often has a longer time lag (Rinia et al., 2001), the citation intervals between cited papers from other disciplines and citing papers in the eHealth field would also increase with the rise of interdisciplinarity.

We notice that there were no *Theory* related AKPs in some early years, therefore, the curve of *Theory* is not continuous. It may be driven by several reasons. First, the early studies in the eHealth field were more focused on the application of

information technology to assist the information acquisition process of medical workers but were concerned less about the theory of interaction between humans and technology. Second, the definition of *Theory* in the present study is very narrow as we only included the phrases with specific theory names due to the operability of annotation. Finally, we only used the metadata of references to do the matching process. However, some references from the early years may not have recorded abstract or the theory related information was not covered in the metadata, which prevents us from annotating the AKPs of theory.

Moreover, we observe that the curve of *Theory* in Fig. 4 has fluctuated during the period. The rapid increase from 2008 to 2010 may be attributed to the rapid growth of the publications in the period, and they cited a few classical theory models (e.g. “social cognitive theory”) which were proposed in the early years. On the other hand, the theories cited by the eHealth field covered both relatively new information technology theories (e.g. “sensor acceptance model”) and classic cognitive theories (e.g. “social cognitive theory”). Therefore, the curve of the *Theory* has fluctuated during the later years. For *Research Methodology*, it shows a relatively long rise before 2007. At the moment, eHealth research absorbed some traditional psychology questionnaires (e.g. “SCL90R”, “CES D”). Then, it experiences a falling interval between 2007 and 2010. In this period, some novel data analysis approaches (e.g. “text mining”, “natural language processing”, “thematic analysis”) were introduced into the eHealth field. As the development of the eHealth field, more and more psychology questionnaires were used to assist the eHealth research, thus, the citation interval was increased again and gradually remained stable.

3.5 Co-occurrence analysis of knowledge types

We further analyze the co-occurrence pattern of knowledge types within citation contexts to disclose their interactions in the knowledge integration process, as shown in Fig. 5. The ratio value in the figure is calculated as twice the co-occurrence frequency divided by the total frequency of the two knowledge types. It is clear that the most frequent pair of knowledge types is *Research Subject* and *Research Subject*, followed by *Research Subject* and *Entity*, then *Research Subject* and *Technology*. It is reasonable because authors often need to describe research subjects related information when citing the references, and it demonstrates *Entity* and *Technology* are two types of knowledge that are often integrated across different research topics. However, the co-occurrence of *Theory* and *Data* is the fewest. This may be due to the fewest total number of theory related knowledge. We also observe an interesting finding that the cells along with the diagonal line exhibit a relatively high ratio value. This phenomenon may be driven by that when we cite a knowledge entity

Research Paper

(e.g. a methodology or a theory), we usually compare it with other similar types of entities. For example, in our dataset, “*TAM*” theory is frequently occurred with “*TPB*” theory.

Figure 4. The average citation interval of AKPs for each knowledge type.

Figure 5. The co-occurrence frequency of knowledge types within citation context and its ratio to the sum of the two knowledge types. The heatmap was drawn based on the ratio value.

4 Conclusion

The study explores the content characteristics of knowledge integration of an interdisciplinary field, eHealth field. We followed our previous study (Mao, Wang, & Shang, 2020) to highlight several new aspects of integration characteristics of knowledge content in the eHealth field. First, associated knowledge phrases between citation contexts and text of corresponding references were extracted and classified to determine the types of explicit integrated knowledge in the eHealth field. For

each knowledge type, we recognized the highly contributed source disciplines to investigate the knowledge contribution roles of different disciplines in the eHealth field. Then, several indicators, as well as co-occurrence analysis, were applied to study the integration pattern of different knowledge types.

Our case study has shown that different disciplines have different knowledge functions in the eHealth field. For example, medical and health related disciplines, supplied more knowledge of *Research Subject*, *Entity*, and *Research Methodology*, while information technology related disciplines played a more prominent role in providing *Technology* and *Data* related knowledge. In addition, the integration characteristics of different knowledge types are significantly different. *Research Subject* related knowledge spread faster than other types of knowledge, and its interdisciplinary characteristics are more significant. For every knowledge type, their integration time intervals have increased throughout the period, while *Theory* and *Research Methodology* have experienced more fluctuations than other knowledge types. Overall, the integration pattern of different knowledge types became stable along with the mature of the eHealth field, which could be revealed by that the proportion of knowledge amount, references, and source disciplines as well as citation interval of different knowledge types were becoming stable in recent years. Finally, we found that the co-occurrence patterns of knowledge pairs between *Research Subject*, *Entity*, and *Technology* appeared frequently, which suggests entity and technology could be easily integrated to different eHealth research subjects. Furthermore, the co-occurrence of each knowledge type with itself is relatively higher than most other knowledge type pairs.

This study has several implications. For the eHealth field, the knowledge relationships between the field and its related disciplines in the aspect of knowledge types are manifested, which could enlighten the researchers to apply potential interdisciplinary knowledge to the studies in the field. The frequent co-occurrence pairs of knowledge types could promote specific research strategies in the eHealth field. In addition, this article provides a holistic view for domain researchers to understand the evolution of the eHealth field from a fine-grained knowledge integration perspective. On the other hand, for Scientometrics field, we provide valuable insight into understanding the interdisciplinarity of a field by analyzing the types of knowledge from source disciplines in the knowledge integration process.

However, there are also some limitations in this study. First of all, our results are limited, which were only based on the articles from two leading journals in the eHealth field. Second, we designed a stem-matching method to find noun phrases appearing in both citation sentences and the corresponding references, which were regarded as knowledge spread from the references to citing papers. The method could be improved by identifying those phrases with the same meaning, but are

Research Paper

represented by different words. Word embedding techniques could be applied to improve the method, which is one of our future attempts. Nonetheless, there was also some integrated knowledge that may not be contained in the metadata of references (Jaidka, Khoo, & Na, 2019). Therefore, more efforts are called to explore the knowledge integration process of an interdisciplinary field by combining cited text identification approaches (Ou & Kim, 2019). Third, the knowledge integration in an interdisciplinary field is essentially shaped by the interactions and integrations among the knowledge units of the field. We only make a shallow analysis on the co-occurrence among different types of knowledge. For the type of Research Subject, the terms could be further partitioned into sub-categories so that a finer granularity analysis on knowledge integration could be performed. It needs to further explore the structure, patterns and underlying mechanisms of knowledge integration from a micro-level perspective. In addition, we recognized the sources of AKPs from the disciplines of references containing the AKPs, but did not track the origins of each distinct AKP. In the future, we will study the knowledge integration characteristics of an interdisciplinary field from more perspectives.

Acknowledgments

This study was funded by the National Social Science Foundation of China with Grant No. 20CTQ024.

Author contributions

Shiyun Wang (563157995@qq.com) analyzed the data and wrote the manuscript; Jin Mao (maojin@whu.edu.cn) performed the research design and helped to edit the text; Jing Tang (1426137493@qq.com) contributed on data processing and annotation; Yujie Cao (cathy0021@163.com) proposed the original idea and reviewed the manuscript.

References

- Ba, Z., Cao, Y., Mao, J., & Li, G. (2019). A hierarchical approach to analyzing knowledge integration between two fields—a case study on medical informatics and computer science. *Scientometrics*, 119(3), 1455–1486.
- Bahadoran, Z., Mirmiran, P., Kashfi, K., & Ghasemi, A. (2019). The principles of biomedical scientific writing: Title. *International Journal of Endocrinology and Metabolism*, 17(4), e98326.
- Bertin, M., Atanassova, I., Gingras, Y., & Larivière, V. (2016). The invariant distribution of references in scientific articles. *Journal of the Association for Information Science and Technology*, 67(1), 164–177.
- Chi, R., & Young, J. (2013). The interdisciplinary structure of research on intercultural relations: A co-citation network analysis study. *Scientometrics*, 96(1), 147–171.

- Della Mea, V. (2001). What is e-Health (2): The death of telemedicine? *Journal of Medical Internet Research*, 3(2), e22.
- Ding, Y., Song, M., Han, J., Yu, Q., Yan, E., Lin, L., & Chambers, T. (2013). Entitymetrics: Measuring the impact of entities. *PloS ONE*, 8(8), e71416.
- Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, 3(2), e20.
- Gupta, S., & Manning, C.D. (2011). Analyzing the dynamics of research by extracting key aspects of scientific papers. In *Proceedings of 5th International Joint Conference on Natural Language Processing* (pp. 1–9). Asian Federation of Natural Language Processing, Chiang Mai.
- Heffernan, K., & Teufel, S. (2018). Identifying problems and solutions in scientific text. *Scientometrics*, 116(2), 1367–1382.
- Jaidka, K., Khoo, C.S., & Na, J.C. (2019). Characterizing human summarization strategies for text reuse and transformation in literature review writing. *Scientometrics*, 121(3), 1563–1582.
- Kondo, T., Nanba, H., Takezawa, T., & Okumura, M. (2009). Technical trend analysis by analyzing research papers' titles. In *Language and Technology Conference* (pp. 512–521). Springer, Berlin, Heidelberg.
- Lu, W., Li, X., Liu, Z., & Cheng, Q. (2019). How do Author-Selected Keywords Function Semantically in Scientific Manuscripts? *Knowledge Organization*, 46(6), 403–418.
- Mao, J., Wang, S., & Shang, X. (2020). Investigating interdisciplinary knowledge flow from the content perspective of citances. *EEKE@JCDL 2020* (pp. 40–44).
- Mesbah, S., Fragkeskos, K., Lofi, C., Bozzon, A., & Houben, G.J. (2017). Facet embeddings for explorative analytics in digital libraries. In *International Conference on Theory and Practice of Digital Libraries* (pp. 86–99). Springer, Cham.
- Nichols, L.G. (2014). A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics*, 100(3), 741–754.
- Otto, W., Ghavimi, B., Mayr, P., Piryani, R., & Singh, V.K. (2019). Highly cited references in PLOS ONE and their in-text usage over time. *arXiv preprint arXiv:1903.11693*.
- Ou, S., & Kim, H. (2019). Identification of citation and cited texts for fine-grained citation content analysis. *Proceedings of the Association for Information Science and Technology*, 56(1), 740–741.
- Pettigrew, K.E., & McKechnie, L. (2001). The use of theory in information science research. *Journal of the American Society for Information Science and Technology*, 52(1), 62–73.
- Porter, A., Cohen, A., David Roessner, J., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics*, 72(1), 117–147.
- Porter, A.L., Roessner, J.D., Cohen, A.S., & Perreault, M. (2006). Interdisciplinary research: Meaning, metrics and nurture. *Research Evaluation*, 15(3), 187–195.
- Radoulov, R. (2008). Exploring automatic citation classification (master's thesis). Waterloo, Ontario, Canada: The University of Waterloo.
- Rinia, E.D., Van Leeuwen, T., Bruins, E., Van Vuren, H., & Van Raan, A. (2001). Citation delay in interdisciplinary knowledge exchange. *Scientometrics*, 51(1), 293–309.
- Sahragard, R., & Meihami, H. (2016). A diachronic study on the information provided by the research titles of applied linguistics journals. *Scientometrics*, 108(3), 1315–1331.
- Serenko, A., Dohan, M.S., & Tan, J. (2017). Global ranking of management- and clinical-centered e-health journals. *Communications of the Association for Information Systems*, 41(1), 9.

Research Paper

- Small, H., Tseng, H., & Patekc, M. (2017). Discovering discoveries: Identifying biomedical discoveries using citation contexts. *Journal of Informetrics*, 11, 46–62.
- Sun, Y., & Latora, V. (2020). The evolution of knowledge within and across fields in modern physics. *Scientific Reports*, 10(1). doi: 10.1038/s41598-020-68774-w.
- Tsai, C.T., Kundu, G., & Roth, D. (2013). Concept-based analysis of scientific literature. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management* (pp. 1733–1738).
- Wagner, C.S., Roessner, J.D., Bobb, K., Klein, J.T., Boyack, K.W., Keyton, J., . . . & Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14–26.
- Wang, Y., & Zhang, C. (2018). What type of domain knowledge is cited by articles with high interdisciplinary degree? *Proceedings of the Association for Information Science and Technology*, 55(1), 919–921.
- Xu, H., Guo, T., Yue, Z., Ru, L., & Fang, S. (2016). Interdisciplinary topics of information science: A study based on the terms interdisciplinarity index series. *Scientometrics*, 106(2), 583–601.
- Xu, J., Bu, Y., Ding, Y., Yang, S., Zhang, H., Yu, C., & Sun, L. (2018). Understanding the formation of interdisciplinary research from the perspective of keyword evolution: A case study on joint attention. *Scientometrics*, 117(2), 973–995.

This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).