Journal of Informetrics 15 (2021) 101214

Contents lists available at ScienceDirect

Journal of Informetrics

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

Understanding interdisciplinary knowledge integration through citance analysis: A case study on eHealth ☆

Shiyun Wang[a,b], Jin Mao[a,b,*], Kun Lu[c], Yujie Cao[d], Gang Li[a]

[a] Center for Studies of Information Resources, Wuhan University, Wuhan, China
[b] School of Information Management, Wuhan University, Wuhan, China
[c] School of Library and Information Studies, University of Oklahoma, Norman, OK, USA
[d] School of Information Management, Central China Normal University, Wuhan, China

ARTICLE INFO

Keywords:
Knowledge flow
Citation age
Interdisciplinary research
Citation context
eHealth

ABSTRACT

Recent research has shifted to investigating knowledge integration in an interdisciplinary field and measuring the interdisciplinarity. Conventional citation analysis does not consider the context of citations, which limits the understanding of interdisciplinary knowledge integration. This study introduces a novel analytical framework to characterize interdisciplinary knowledge integration by both the content, i.e., integrated knowledge phrases (IKPs), and location of citances (i.e., citing sentences) in addition to citations. Seven knowledge categories are used to classify IKPs, including Research Subject, Theory, Research Methodology, Technology, Human Entity, Data, and Others. The eHealth field is explored as an exemplar interdisciplinary field in the case study. The result reveals that the ranks of source disciplines quantified by the integrated knowledge phrases are different from those by citations, especially in terms of average knowledge integration density. The distributions of the IKPs over the knowledge categories differ among source disciplines, indicating their different contributions to knowledge integration of eHealth field. The knowledge from adjacent disciplines is integrated into the field faster than that from other disciplines. Knowledge distributions over sections of articles are also different among source disciplines, and a correlation between knowledge categories and the sections they were used is observed. The analytical framework offers a way to better understand an interdisciplinary field by disclosing the characteristics of interdisciplinary knowledge integration from the perspective of knowledge content and usage.

1. Introduction

The breakthroughs of many major scientific problems have often been made in interdisciplinary fields (Morillo et al., 2003). Recently, interdisciplinary research has attracted extensive attention from researchers due to its important role in the science community. In the Scientometrics field, researchers have identified the key aspect of interdisciplinary research as the process of integrating different bodies of knowledge (Porter et al., 2007; Rafols & Meyer, 2010). The knowledge, recognized as methods, theories, tools, and concepts, is often integrated from multiple disciplines to solve problems in interdisciplinary research (Wagner et al., 2011). Some efforts were made to investigate the social process of interdisciplinary knowledge integration (Miller & Mansilla, 2004). However, in a quantitative manner, more studies measure the results of interdisciplinary knowledge integration as reflected in the publications

☆ This article is an extension of our work (Mao et al., 2020), which was reported in the 1st Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents co-located with the ACM/IEEE Joint Conference on Digital Libraries in 2020.

* Corresponding author.
E-mail address: danveno@163.com (J. Mao).

https://doi.org/10.1016/j.joi.2021.101214
Received 1 September 2020; Received in revised form 11 July 2021; Accepted 19 August 2021

and the authors in the field (Porter et al., 2007; Wagner et al., 2011). To understand the characteristics of interdisciplinary knowledge integration, citation analysis has often been used to quantify knowledge flow among disciplines (Chang & Huang, 2012; Van Leeuwen & Tijssen, 2000; Yan, 2016; Zhu & Yan, 2015).

Traditionally, the knowledge flow to a field is measured by the number of references cited by the papers in the field (Karunan et al., 2017; Rinia et al., 2001). The underlying theory is that citations usually imply knowledge transfer from the cited papers to the citing papers (Yan, 2016). However, treating all citations equal is an oversimplification because the citation behavior is complicated (Zhang et al., 2013). Authors of an article may cite the references for various purposes, such as theoretical construction or methodological reference, along the discourse logic of the article. Measuring interdisciplinary knowledge flow using the frequency of references is far from precise. In addition, the frequency of references does not reveal the content that each related discipline contributes to the interdisciplinary field.

Several studies have attempted to illustrate the knowledge connections between interdisciplinary fields and related disciplines by analyzing their associated topics. Text mining approaches, e.g., keyword mining and topic modeling, have been applied to extract keywords or topics from publications in a field, and then the knowledge connections were revealed by the keywords or topics occurring in related disciplines (Nichols, 2014; Xu et al., 2016). These studies can demonstrate what knowledge is integrated by an interdisciplinary field. However, they purely rely on expert wisdom to discern and illustrate the knowledge connection, but not based on any clues from citations. Furthermore, they neither identify the semantic meaning of the integrated knowledge nor investigate how the knowledge is applied in the interdisciplinary field papers. An approach that reveals *what* and *how* knowledge is integrated into an interdisciplinary field is of interest, which can offer more details beyond citation counts to understand the interdisciplinary knowledge integration.

Recent advances in citation context studies offer a finer-granular approach to distinguish citations from various aspects by digging into the surrounding text of reference mentions. Citation contexts embed the syntactic information, e.g., the location of section and rhetoric style, and the semantic information, e.g., the meaning of citation content (Ding et al., 2014). The combination of them can be applied to characterize the function (Zhang et al., 2013), importance (Hassan et al., 2018), and knowledge contribution (Thelwall, 2019), of different citations. The syntactic information and literal content of citation context could help explore the integrated knowledge of an interdisciplinary field from a more comprehensive perspective.

In this study, we introduce a novel analytical framework to understand the characteristics of knowledge integration in an interdisciplinary field based on citance analysis. A citance is denoted as the sentence that contains in-text citations, which provides the content and location of the citation. We attempt to explore what knowledge has been transferred from source disciplines to the interdisciplinary field through the content of citations by applying natural language processing techniques. And, we further explore how the knowledge is integrated in the field through the locations of citations. The field of *eHealth* is selected as an exemplary interdisciplinary research field to be analyzed, which is an emerging field of medical informatics (Drosatos & Kaldoudi, 2020), referring to all aspects of the intersection of health care and the Internet (Pagliari et al., 2005). Our study is driven by the following research questions:

RQ1. What are the major source disciplines that contribute to the eHealth field as measured by integrated knowledge content in citances? Is the result consistent with that by citation counts, or not?
RQ2. What knowledge is integrated from the source disciplines into the eHealth field? What are their features, e.g., their categories and age?
RQ3. Where is the integrated knowledge content cited in the eHealth articles? Are there any differences in the location distribution of knowledge from different source disciplines?

The answers to these questions can characterize the knowledge integration between the eHealth field and its source disciplines and help to understand the roles of different disciplines in the eHealth field as revealed by the citation content and citation location.

The article is organized as follows: The following section provides the related work, followed by the methodology used in this study. Then, the results section illustrates our findings of the analysis. Afterwards, we provide some further discussion for the results and their implications, as well as the avenues for future research.

2. Related work

2.1. Interdisciplinary knowledge flow

With the benefit of citation databases, such as Scopus, Web of Science (WoS), or Google Scholar, citations are used to form quantitative indicators to measure knowledge transfer between disciplines. A few indicators have been proposed to quantify cross-disciplinary knowledge diffusion, for instance, Brillouin index (Brillouin, 1956; Steele & Stier, 2000), Citation Out Categories (Porter & Chubin, 1985), Rao-Stirling (Stirling, 2007), Integration (Porter et al., 2007), and field diffusion breadth (Liu & Rousseau, 2010). These indicators can reflect the interdisciplinarity of a specific research field by exploring different natures.

Patterns of interdisciplinary knowledge diffusion have also been widely examined at various levels through citation analysis. At the journal level, a study found an increasing trend of citations from journals in Information Science to Communication (Borgman & Rice, 1992). At the field level, it has been found that citation behavior is influenced by the discipline of references in Behavioral Sciences (Lange, 1985). It has also been found that papers in one discipline are more likely to cite papers in nearby disciplines (Van Leeuwen & Tijssen, 2000), and the knowledge transfer across disciplines tends to have a greater time lag than knowledge transfer within a discipline (Rinia et al., 2001). Furthermore, papers and their citations are used to construct networks, then the knowledge diffusion relationships between disciplines can be measured by network analysis (Karunan et al., 2017; Zhu & Yan, 2015). These studies focused on macro-level interdisciplinary knowledge diffusion.

A few recent studies have attempted to analyze interdisciplinary knowledge flow at a much finer granularity by observing the content of diffusion. Engerer (2017) explored the interdisciplinary interactive relationships between Linguistics field and Information Retrieval field by observing linguistic topics involved in recent information science publications. Ba et al. (2019) quantified Computer Science terms integrated by different layers of the co-word network of Medical Informatics to measure the breadth and strength of knowledge integration between the two fields. These studies showed the knowledge association between two disciplines through keyword analysis. However, they did not investigate the features of the spread knowledge content.

The diffusion of specific knowledge units among different disciplines was also analyzed in recent studies. Zhang et al. (2011) traced the spread of “h-index” among different disciplines over a five-year period. Zhai et al. (2018) measured the interdisciplinary diffusion of Latent Dirichlet Allocation (LDA). Pan et al. (2018) explored the diffusion patterns of software across research fields. These studies focused on the diffusion characteristics of specific knowledge units, but did not consider how they are integrated in an interdisciplinary field.

2.2. Citation content analysis

Recently, it is much easier to obtain full text of publications under the support of Open Access initiatives and big scholarly databases. Citation content analysis, which expands citation analysis by using both syntactic information, e.g., the position, and semantic information, e.g., citation function, citation motivation, citation sentiment, of citation sentences, can distinguish the importance and functions of citations (Zhang et al., 2013).

Some studies suggested that the citation location may be able to reflect the importance of citations. Voos and Dagaev (1976) found that the section of Introduction contained more highly cited articles, thus they argued that citation location should be considered as part of the evaluation of citations. In addition, Herlach (1978) suggested that a reference cited in several different sections tends to make a more substantial contribution to citing papers. Early studies were mostly based on manual analysis on small datasets. In recent years, the standardized organization of scientific papers in major journals, e.g., IMRaD structure (Introduction, Methods, Results, and Discussion), and the application of text mining techniques have provided the possibility to automatically determine the citation locations of in-text citations within the full text of papers. The in-text citations distribution over sections have been widely explored (Ding et al., 2013a; Hu et al., 2013; Hu et al., 2017; Poncela-Casasnovas et al., 2019). Furthermore, Halevi and Moed (2013) analyzed the in- and out-disciplinary citations distribution over the sections of articles in Journal of Informetrics (JOI), and demonstrated this analysis could reveal the connections between the issues, methods, and conclusions coming from different disciplines. Liu et al. (2019) evidenced different roles played by various disciplines in the knowledge flow of holistic science according to the distribution of references from various disciplines over sections. Zhang et al. (2021) further characterized the distribution of references from various disciplines in PLoS journals, by analyzing the number of citations, average cited intensity, and average citation length at the section level. Their results illustrated that different disciplines are distributed differently in various sections, which reveals the different influence and contributions of different disciplines. These previous studies revealed a common phenomenon that the distribution of in-text citations in different sections of citing papers is uneven and demonstrated the importance of citation location in analyzing the roles of citations in citing papers.

Researchers have also attempted to classify citation sentences based on their citation contexts, which needs to develop a codebook by expert knowledge to reveal the reasons and functions of citations. Lipetz (1965) developed a citation classification scheme based on the citation context. Subsequent generations have developed citation classification schemes from different dimensions, for example, citation motivation (Case & Higgins, 2000; Chubin & Moitra, 1975), citation function (Garzone, 1997; Oppenheim & Renn, 1978; Radoulov, 2008; Spiegel-Rosing, 1977), and citation sentiment (Frost, 1979; Small, 2011; Teufel, Siddharthan, & Tidhar, 2006a). A few recent studies attempted to automate the citation classification process by using machine learning techniques, for example, IBk k-Nearest Neighbors algorithm, Random Forests algorithm, support vector machine (SVM), and deep learning algorithms (Dong & Schäfer, 2011; Meng et al., 2017; Perier-Camby et al., 2019; Taşkın & Al, 2018; Teufel et al., 2006b). In particular, Cohan et al. (2019) incorporated the structural information of scientific papers to automatic citation intent classification and demonstrated the effectiveness of the scientific discourses in inform citation intent classification. These studies focused on sentences, but not the specific knowledge units in the citation context. Small (1978) noted that the citation context usually contains the citing author's interpretation of the concept, method, and other ideas of the cited work. Following this work, a few researchers used citation content analysis to investigate the impact of particular author' works or certain theories from the content level (González-Teruel & Pérez-Pulido, 2020; Lu et al., 2017; McCain & Salvucci, 2006; Tsay, 2009).

In summary, previous studies in interdisciplinary knowledge flow research typically use citation analysis, which mainly focus on a binary relationship between citing papers and cited papers, but generally ignore the context of citations within the full text of citing papers. In our previous study (Mao et al., 2020), we preliminarily investigated what knowledge is integrated from source disciplines to the interdisciplinary field based on citation content. In this article, we attempt to provide a more comprehensive analytical framework to understand the knowledge integration in an interdisciplinary field, which combines both syntactic and semantic analysis of citations.

3. Methodology

Our analytical framework consists of the following stages: data collection and parsing, pre-processing, and data analysis (Fig. 1).

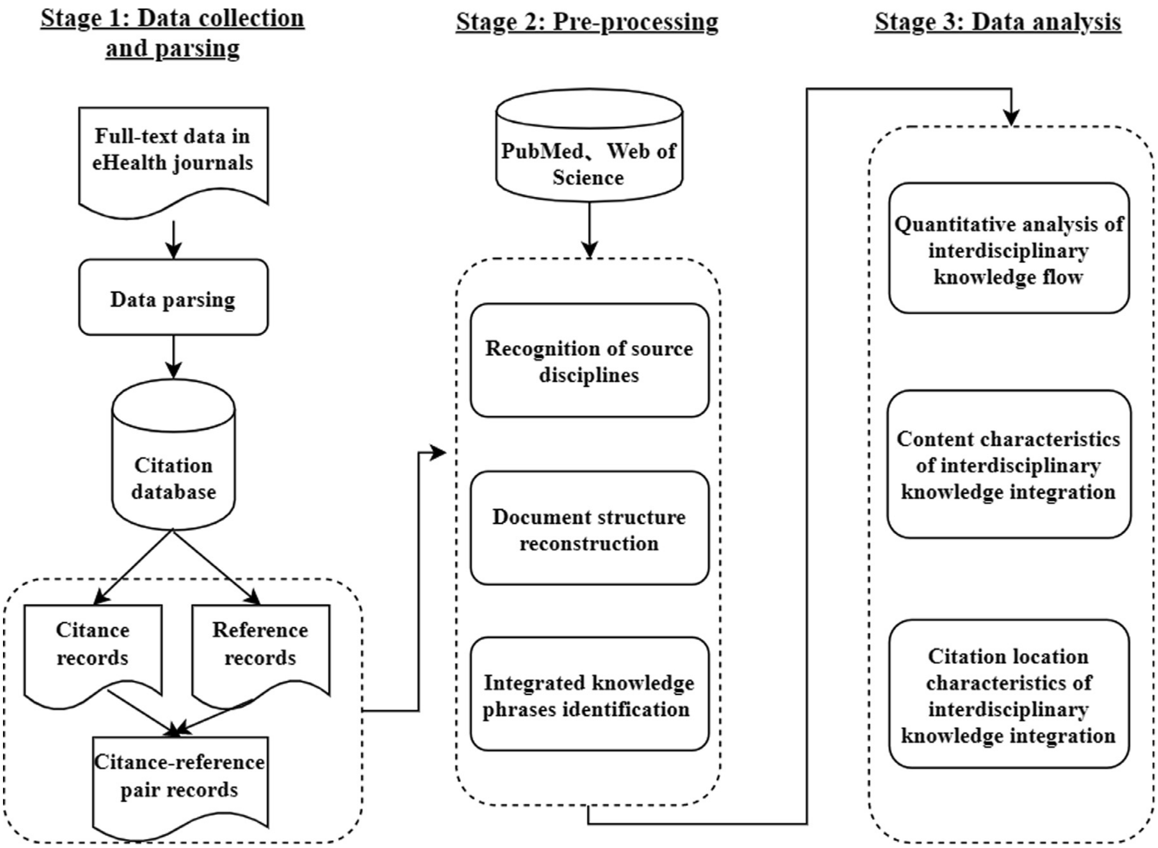

Fig. 1. Overview of our analytical framework to investigate interdisciplinary knowledge integration of the eHealth field.

3.1. Data collection and parsing

In this study, two high impact eHealth journals, the *Journal of Medical Internet Research* (JMIR) and *JMIR mHealth and uHealth*, were selected as data sources. JMIR is the leading journal in eHealth that was established in 1999, when the eHealth field was just emerging (Della Mea, 2001). JMIR mHealth and uHealth, is a spin-off journal of JMIR, which covers papers that are more technical or more developmental. According to the 2018 Journal Citation Reports, these two journals are ranked in the top two out of 26 journals in the subject category of *Medical Informatics*. A survey among 398 health informatics experts has shown that the JMIR was ranked as a top tier (A+) journal among the eHealth management-focused journals, whereas the JMIR mHealth and uHealth was ranked as an A journal (Serenko et al., 2017). In total, 3,416 articles in the format of XML files published from 1999 to 2018 were collected from these two journals. It should be noted that the field of eHealth in this study is only represented by the two journals, which is a narrow scope.

The metadata (article type, publish year, etc.) and bibliographic information (citation type, journal title, DOI, PubMed ID, publish year, etc.) of the articles were parsed according to the XML schema. The sections and paragraphs of the articles were extracted, and then segmented into sentences by using the punctuations (periods, question marks, etc.). From the sentences, we identified the citances that contained in-text reference information, i.e., reference id.

Papers in the types of Corrigenda and Addenda, Editorial, Letter to the Editor, etc., do not contain section type information or have fewer references. We excluded them and only focused on the remaining 3,221 articles (Original papers, Reviews and Viewpoints), accounting for 94.29% of all articles. In total, we obtained 119,598 citances and 140,572 reference records, i.e., bibliographic items, from our corpus.

3.2. Data pre-processing

3.2.1. Recognition of source disciplines

We identified the disciplines of the references according to their journals. We chose the Web of Science (WoS) subject categories classification scheme to determine the disciplines of the references by following many previous studies on interdisciplinary knowledge flow/integration (Porter et al., 2007; Stopar et al., 2016; Shen et al., 2019). WoS classification scheme has 254 subject categories and can assign a journal to more than one subject category due to the interdisciplinarity of journals. In addition, WoS classification

Table 1
Classification rules to match the IMRaD structure in our corpus.

Section label	Section titles in our corpus
Introduction(I)	Intro, introduction
Methods(M)	Materials, materials and methods, method, methods, methods and materials
Results(R)	Results, results
Discussion(D)	Discussion, discussion

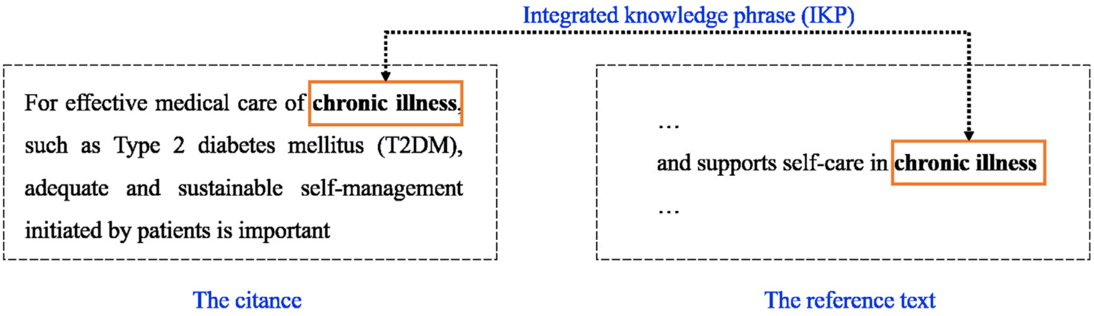

Fig. 2. Example of our definition of integrated knowledge phrases.

scheme is more fine-grained and may be more suitable to analyze the eHealth field by considering the limited amount of eHealth articles in our corpus.

First, we obtained 7,393 distinct journal titles from the 104,891 reference records with the citation type of "journal" and with DOI/PubMed ID. They were matched directly with the journal titles in the 2018 version of WoS journal list, which covers 21,003 journals with full titles. However, only 12,280 reference records of 852 distinct journals were matched. We found that many journal titles in the reference records use abbreviations, or even have wrong forms of full titles. Next, we manually completed and corrected the full titles for 2,466 unmatched journal titles that have more than two reference records in our corpus. Finally, 93.77% journal reference records (98,355 out of 104,891) were matched to WoS journal list, and therefore can be categorized into WoS subject categories.

3.2.2. Document structure reconstruction

To investigate the citation location of the knowledge from source disciplines, we aligned the logical structure of the articles in our corpus with the IMRaD structure, which is a predominant publication style in biomedical domain (Sollaci & Pereira, 2004). Various titles may indicate the same section type, e.g. "materials and methods", and "method". We formulated the classification rules as shown in Table 1 to standardize the IMRaD structure. Some citances do not have section type information or their section type could not be matched to any IMRaD section type. A total of 115,456 citances were tagged with section labels, accounting for 96.54% of all citances.

3.2.3. Integrated knowledge phrases identification

Citation contexts contain information about the cited articles relevant to the citing papers (Elkiss et al., 2008; Small, 1978). Similarly, we assume that the knowledge units occurred in both citation context and the corresponding cited text can reflect the explicit integrated knowledge transferred from the cited papers to the citing papers. We defined the integrated knowledge phrases as follows:

Integrated knowledge phrases (IKPs): The noun phrases extracted from the citances that also appear in the corresponding reference texts. An example of our definition of integrated knowledge phrases is shown in Fig. 2.

To identify the IKPs in our corpus, we first augmented the corresponding reference text for each citance. In this paper, we only used the title and abstract of each reference to represent the cited text due to the difficulty of obtaining full texts of all references. Nonetheless, we consider that the key concepts or topics in a paper would be reproduced in the title and abstract. Therefore, the use of title and abstract to represent the cited text is meaningful although it may ignore some detailed information in the full text. The abstracts of references were obtained by searching PubMed ID in PubMed or DOI in Web of Science. In total, we obtained 89,372 abstracts of reference records.

Then, we extracted noun phrases that carry meaningful concepts from the citances as well as the titles and abstracts of the references by using the spaCy package, an open-source python natural language processing toolkit (Honnibal & Montani, 2017). A few cleaning processes were conducted for the phrases. Noun phrases that in caps were retained the raw forms, while other phrases were converted to lowercase. Noun phrases with a single character or some wildcards (e.g., "#", "*", "@", etc.) were removed, so were those starting or ending with a number. Stop words listed in the NLTK package (Bird et al., 2009) were also eliminated. Acronyms were identified and expanded into their full forms by using the scispaCy package (Neumann et al., 2019).

Table 2
Indicators for quantifying the interdisciplinary knowledge flow.

Indicator	Definition
The number of IKPs	It is the number of integrated knowledge phrases identified from the citances, where the corresponding references belonging to the specific discipline.
The number of CountX citations	The number of total mentions of the references belonging to the discipline in citing papers, i.e., in-text citations (Ding et al., 2013a).
The number of CountOne citations	The CountOne citations are counted by the traditional approach which counts each reference only once in a citing paper (Ding et al., 2013a).
Average knowledge integration density (AKID)	The average knowledge integration density means the average number of IKPs per citation for each source discipline. It is calculated as follows: $AKID = \frac{The\ number\ of\ IKPs}{The\ numer\ of\ CountX\ citations}$

Finally, a stem-matching method was applied to identify the shared knowledge phrases between the citances and the corresponding references. If a citance contains multiple in-text citations, it would be paired with multiple reference texts respectively. Besides the IKPs directly matched between the phrase lists of citances and references, we also included the phrases of reference texts that matched with the substring of the citances by regular expression rules. It should be noted that we used both the acronym and the full form of each noun phrase in the matching process, but only retained the raw forms of the noun phrases in IKPs identification results. This stem-matching method achieved a precision of 0.882, and a recall of 0.827 according to the evaluation on a randomly sampled 100 citances, indicating its good performance.

3.3. Data analysis on interdisciplinary knowledge integration

3.3.1. Quantitative analysis of interdisciplinary knowledge flow

To investigate the first research question RQ1, we proposed several indicators to quantify the interdisciplinary knowledge flow from source disciplines to the eHealth field based on citation counts and integrated knowledge phrases, as defined in Table 2. It should be noted that if a reference belongs to multiple subject categories, it is counted independently in each subject category.

3.3.2. Content characteristics analysis

We analyzed the knowledge categories and age of IKPs from different source disciplines (RQ2).

(1) Knowledge categories. In previous studies, Radoulov (2008) has proposed object components, including concept, method, data, to identify "what" in citations. Wang and Zhang (2018) analyzed what types of Computer Science (CS) knowledge (e.g., Data, Theory) are cited by the method sections in the articles with high interdisciplinary degree through citation content analysis. On the other hand, some researchers have attempted to categorize keywords/phrases in scientific text into multiple semantic functions, for example, scientific problem, solution, method, data, and technique (Gupta & Manning, 2011; Heffernan & Teufel, 2018; Kondo et al., 2009; Mesbah et al., 2017; Pettigrew & McKechnie, 2001; Sahragard & Meihami, 2016; Tsai et al., 2013). Based on the practice of previous researches and the features of our corpus, we designed a classification framework for IKPs that considering the semantic functions in the field. Two graduate students majored in Information Science but familiar with the eHealth field were recruited to annotate the categories of the IKPs. The annotation process included the following steps:

- Initial knowledge classification framework. One author constructed a preliminary classification schema based on literature review. Afterward, the author randomly selected 100 IKPs for trial annotation, then organized the annotation details, and wrote an annotation specification document, which provides a detailed definition of each knowledge category with a few exemplar phrases.
- Pre-annotation. Pre-annotation training was carried out for the two coders. The two coders independently annotated 500 identical IKPs that were randomly selected. The Kappa value of inter-rater agreement was 0.65, which indicated moderate agreement (McHugh, 2012). The two coders then discussed the ambiguous cases with a professional in the eHealth field. We found that some phrases may be related to multiple knowledge categories in the annotation process, especially for phrases related to Theory, Research Methodology, and Technology categories. We determined the categories of these phrases by referring to their entries in Wikipedia. For example, "transtheoretical model" may be considered as a method in some situations, however, it was described as a theory model in the Wikipedia. Therefore, we assigned it to the Theory category. In addition, if a phrase is about physical equipment, medical instrument or wearable device, rather than research method, such as algorithm, data analysis software, or scale, it is considered as a phrase of Technology rather than Research Methodology. Following this idea, the two coders reached a consensus.
- Formal annotation. The two coders annotated all distinct phrases. During the annotation process, two coders maintain communication with the professional in the eHealth field to reach an agreement. Wikipedia entries of the IKPs were also referenced to determine the categories of the phrases.

Our final framework contains seven categories, including Research Subject, Theory, Research Methodology, Technology, Human Entity, Data, and Others, as shown in Table 3.

(2) Age of integrated knowledge phrases. The age of an integrated knowledge phrase is defined as the time interval between the citance and the reference text where the IKP occurs. It is calculated as the difference between the publication year of the citing paper and that of the reference paper. For each source discipline, we calculated the average age of IKPs of each knowledge category.

Table 3
The knowledge classification framework of integrated knowledge phrases.

Category	Description	Exemplar phrases
Research Subject	subject terms related to medical research problems, such as diseases, research areas, also some background information	*information, depression, diabetes, health information*
Theory	theoretical knowledge phrases with specific names, including some theoretical frameworks/models that are widely accepted	*TAM, social cognitive theory, transtheoretical model*
Research Methodology	research methodology, including research methods, analytical techniques, measurement scales, guidelines, evaluation indicators	*systematic review, analysis, meta analysis, randomize control trial*
Technology	medical instruments, physical devices, and medical management systems that referred in research, which were used for treatment, diagnosis, or interventions in health services.	*mobile phone, web, smartphone, app*
Human Entity	people or organization that is the target group of the research, including patient group, health caregiver group and health care related organizations	*patient, woman, child, adolescent*
Data	phrases related to dataset, data source and data material	*tweet, qualitative datum, clinical datum*
Others	other phrases that cannot be included in the above categories, e.g., geolocations, projects, and some meaningless phrases	*study, use, result, outcome, number, canada, project, USA*

Table 4
Brief information of our dataset for analysis.

Characteristics	Statistics
Citing papers	3,221
Citances	115,456
Reference records	89,372
Citation-reference pair records	124,024
Integrated knowledge phrases	215,142

Table 5
Distribution of citations across sections.

Sections	In-text citations
Introduction	55,842 (45.03%)
Methods	16,029 (12.92%)
Results	15,943 (12.85%)
Discussion	36,210 (29.20%)

3.3.3. Citation location analysis

Similar sections of text in scientific papers often perform similar rhetorical functions (Teufel, 1999; Thelwall, 2019), indicating that the citations in the same type of sections may have similar knowledge functions. We further investigated the occurrences of IKPs in the sections of eHealth papers from two aspects (RQ3), i.e., the location distributions of IKPs from different source disciplines and that of different knowledge categories.

4. Results

4.1. Descriptive statistics of the dataset

To measure the interdisciplinarity of the eHealth field, we calculated the Citations Outside Category indicator of our data (Porter & Chubin, 1985). The result shows that the number of references from subject categories other than “Health Care Sciences & Services” and “Medical Informatics”, which are the subject categories of the selected journals, accounts for more than 74.98% of the total references. The score of Brillouin Index (Brillouin, 1956) is 1.566, which also indicates a high degree of the diversity of the references (Huang et al., 2012). The two indicators reflect the interdisciplinarity of the eHealth field.

Brief information of our dataset is shown in Table 4. In total, 115,456 citances with IMRaD sections and 89,372 reference records with discipline and abstract information were linked to 124,024 citance-reference pair records. We extracted 153,013 meaningful noun phrases from the citances, and 6,215,357 noun phrases from the reference texts. From them, we identified 215,142 IKPs, with 24,088 distinct ones.

The distribution of in-text citations over the IMRaD sections is presented in Table 5. Introduction contains the most citations, followed by Discussion. This distribution pattern is consistent with previous reports (Bertin et al., 2016; Poncela-Casasnovas et al., 2019). It could be attributed to different functions of the sections in scientific papers. For example, Introduction sections contain more background information and related prior research, and in Discussion sections, the authors tend to compare their results with the relevant studies to show the significance of findings (Hu et al., 2013; Bertin et al., 2016), thus the citations in the two sections are more than other sections.

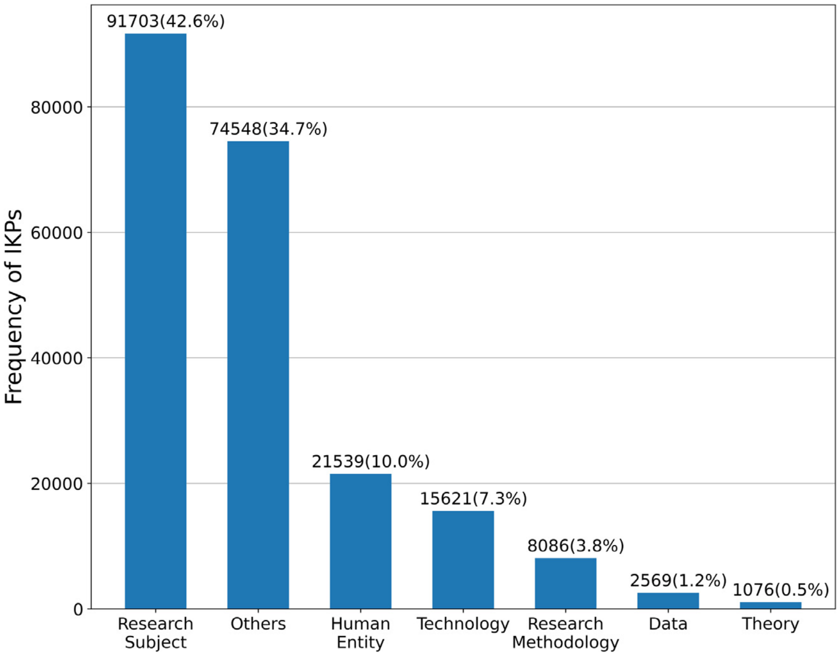

Fig. 3. Integrated knowledge phrases of different categories.

Fig. 3 presents the frequency of IKPs for each category. The phrases of Research Subject category are the most. We observed that the categories of Human Entity and Technology have more phrases than Research Methodology. This could be explained by the domain feature of the field we selected. Many eHealth research topics are related to information technology applied in health services, and eHealth researches are mainly targeting on various groups of human entity, including patients with different diseases and physicians. Data and Theory phrases are the fewest. There are many phrases in Others category, which is the second most. It involves a large number of general phrases, such as “study” (9177 times), “use”(6308 times), and “result”(1688 times), and a variety of phrases that are not the focus of this study, e.g., geolocations. The phrases in the category of Others show few domain features of source disciplines and are not very helpful in understanding the knowledge integration from source disciplines. Therefore, the Others category is excluded in the following analysis.

4.2. Quantitative characteristics of interdisciplinary knowledge flow

To investigate the knowledge flow between the source disciplines and the eHealth field, we analyzed the number of IKPs, CountX citations, CountOne citations, and average knowledge integration density (AKID) for each source discipline.

The references of eHealth field cover 190 WoS subject categories. Table 6 shows the top 20 subject categories in terms of the number of IKPs. Overall, the top contributing disciplines for eHealth field were mostly related to health and medicine (e.g., *Health Care Sciences & Services, Medicine, General & Internal*), psychiatry and psychology (e.g., *Psychiatry, Psychology, Clinical*), and information technology (IT, e.g., *Computer Science, Information Systems, Information Science & Library Science*). Among them, *Health Care Sciences & Services* and *Medical Informatics* provided the largest number of IKPs for the eHealth field, which are the subject categories that the selected eHealth journals belong to.

We observed that the subject category ranks by the number of IKPs, CountX citations, and CountOne citations are slightly different. For example, the CountOne and CountX citations of *Psychology, Multidisciplinary* are more than those of *Nutrition & Dietetics* and *Endocrinology & Metabolism*, but its IKPs are relatively fewer than these two disciplines’, indicating its lower knowledge transfer efficiency per citation. We further calculated the AKID of each discipline to examine the knowledge density of each in-text citation. Interestingly, the disciplines rank in terms of AKID is different from the ranks by the number of IKPs ($\rho = 0.194,\ p-\text{value} = 0.413 > 0.05$), the number of CountX citations ($\rho = 0.149,\ p-\text{value} = 0.531 > 0.05$) and the number of CountOne citations ($\rho = 0.074$, p-value $= 0.758 > 0.05$) through the Spearman’s correlation analysis. The WoS categories of the two journals, i.e., *Health Care Sciences & Services* and *Medical Informatics*, as well as the medical and health related disciplines, have higher AKID, while IT related disciplines have lower AKID. It might be due to the fact that the research topics of a discipline are more similar as the references from the discipline or neighboring disciplines. A Kruskal-Wallis H test shows the knowledge integration density of the 20 source disciplines is significantly different (p-value < 0.001, df=19). The results demonstrate that different disciplines play different roles in transferring knowledge into this interdisciplinary field in terms of the knowledge amount per citation.

In summary, the four metrics can provide different perspectives to understand knowledge flow from source disciplines to eHealth field.

Table 6
Quantitative characteristics of interdisciplinary knowledge flow.

Rank	Subject category	IKPs	CountX citations	CountOne citations	AKID
1	Health Care Sciences & Services	63386	33231	19703	**1.907**
2	Medical Informatics	48030	24754	14324	**1.940**
3	Public, Environmental & Occupational Health	39142	23139	12533	1.692
4	Medicine, General & Internal	27359	15282	10228	1.790
5	Psychiatry	19630	11140	5802	1.762
6	Psychology, Clinical	12685	7789	4862	1.629
7	Substance Abuse	9657	5547	2250	1.741
8	Health Policy & Services	8898	5223	3550	1.704
9	Nursing	8228	4893	1887	1.682
10	Computer Science, Information Systems	7347	4636	2736	1.585
11	Information Science & Library Science	7282	4682	2825	1.555
12	Nutrition & Dietetics	6718	3926	2406	1.711
13	Endocrinology & Metabolism	6566	3835	2159	1.712
14	Psychology, Multidisciplinary	6552	4251	2806	1.541
15	Psychology	5547	3321	2204	1.670
16	Computer Science, Interdisciplinary Applications	5414	3476	2067	1.558
17	Clinical Neurology	4977	2778	1830	1.792
18	Oncology	4717	2579	1637	1.829
19	Multidisciplinary Sciences	4558	2618	1618	1.741
20	Pediatrics	3937	2246	1396	1.753

4.3. Content characteristics of interdisciplinary knowledge integration

We further analyzed the IKPs from each source discipline. Appendix A provides a few examples of citances and the IKPs from the source disciplines. The top 10 most frequent IKPs of each source discipline evidence some differences between the source disciplines (Appendix B). For example, mental and cognitive diseases, such as depression and anxiety, have a greater frequency in psychology related disciplines, while information and technology related topics, e.g., information, technology, internet, and app, appear more frequently in IT related disciplines. It demonstrates the knowledge roles of different source disciplines are various. To understand the content characteristics of interdisciplinary knowledge integration, we further analyzed the knowledge category distribution and age distribution of IKPs for each discipline.

4.3.1. Knowledge category distribution of IKPs from different disciplines

For each source discipline, Fig. 4 presents the proportion of IKPs over the knowledge categories. The frequency distribution is shown in Appendix C. To showcase some integrated knowledge, we also list several IKPs examples for each source discipline in Appendix D.

In general, the Research Subject category accounts for the largest proportion for all disciplines. This is due to the fact that the definition of Research Subject category is broader than other categories in this study, and most phrases are related to the research topics or background information.

However, there are also some differences in the distribution between disciplines. According to Table 6, *Health Care Sciences & Services* is the most contributing discipline for the eHealth field. For this discipline, the rank of knowledge categories by the number of IKPs is Research Subject, Human Entity, Technology, Research Methodology, Data, and Theory, which is the same as the rank for all integrated knowledge phrases (Fig. 3). Most medical and health related disciplines show the same distribution pattern as *Health Care Sciences & Services*, e.g., *Public, Environmental & Occupational Health, Medicine, General & Internal*, and *Health Policy & Services*, with more IKPs of Research Subject and Human Entity categories. While, the phrases of Technology (e.g., app, robot, smartphone, sensor) and Data (e.g., datum) occupy a bit more proportion in IT related disciplines, including *Computer Science, Information Systems, Information Science & Library Science*, and *Computer Science, Interdisciplinary Applications*. For *Psychology and Psychiatry* related disciplines, they have a larger percentage of Research Methodology phrases (e.g., meta-analysis, hospital anxiety and depression scale). This is related to the most popular disease topic in eHealth field: depression (Drosatos & Kaldoudi, 2020). The researches relevant to this topic often investigate the performance of technological interventions on depression and related disorders through meta-analysis and randomize control trial methods and with the help of some measurement scales. In addition, the proportion of Theory phrases in *Psychology, Multidisciplinary* is very large, even exceeding the proportion of Data category, but the discipline is only ranked 14th in Table 6.

The Pearson Chi-Square test on the distributions of the 20 source disciplines shows significant difference (p-value << 0.001, df = 95). And, 0 cells (0.0% < 20%) have expected count less than 5, indicating the result is meaningful. The above analysis shows that different disciplines have different distributions of the IKPs over the six knowledge categories, reflecting their different roles in the knowledge integration of the eHealth field.

4.3.2. The average age of IKPs in different knowledge categories

The average age of IKPs in the six categories for the top 20 disciplines are presented in Fig. 5. Overall, phrases of Research Methodology and Theory category are older than those of other knowledge categories, whereas Technology phrases are relatively new. It could be explained by that a successful method or theory about a certain topic should be verified by the scientific community to

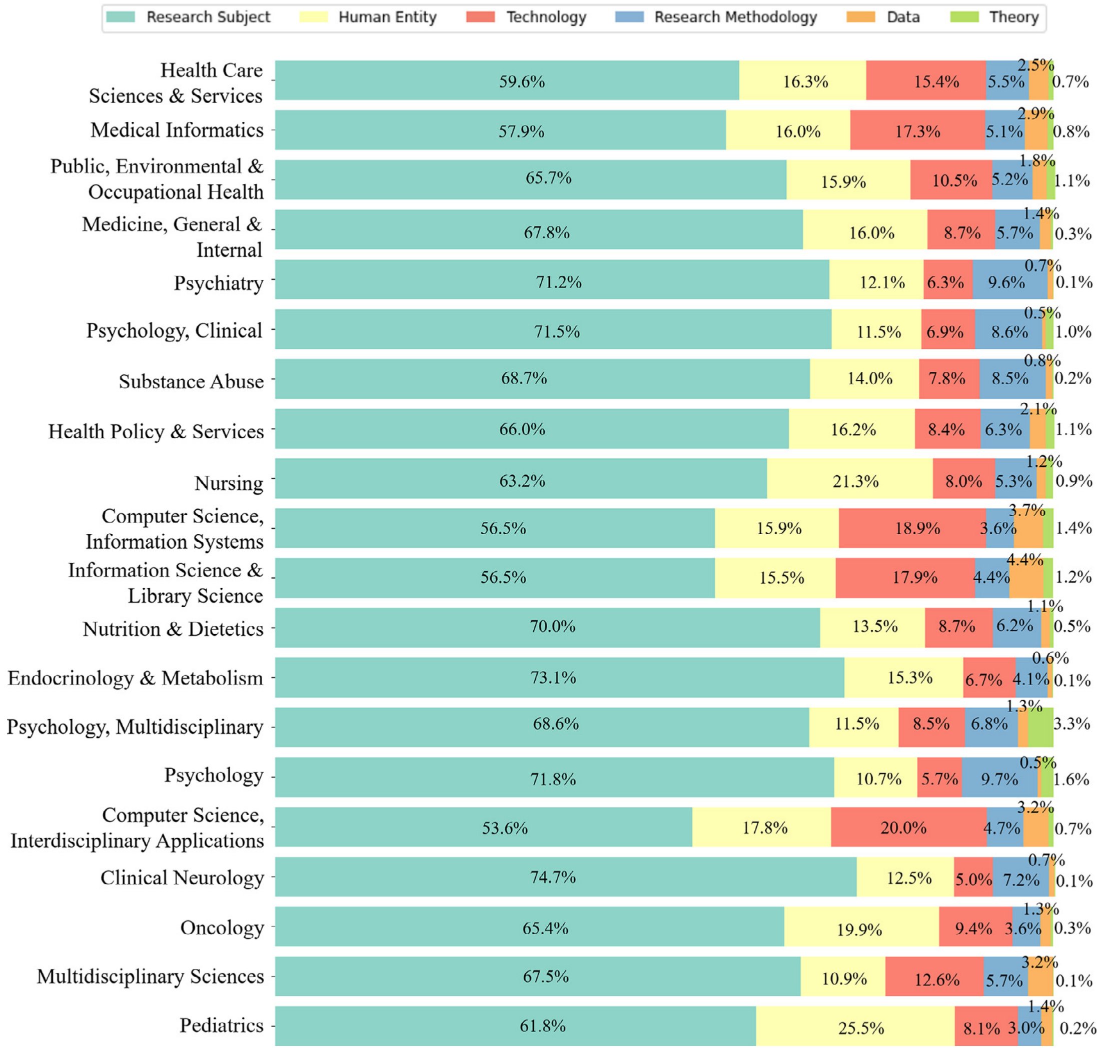

Fig. 4. The proportion of integrated knowledge phrases in the six categories for each source discipline.

achieve the status of little uncertainty (Small, 2018), while the technologies are developing fast in modern society; Recent technology typically show superior performance, thus researchers tend to adopt new technology in their research.

On the other hand, the age distributions of IKPs in various disciplines are different. Research Methodology phrases in *Psychology, Psychiatry*, and *Clinical Neurology* are much older than that in other disciplines. It is because that the highly cited methodology phrases in *Psychology, Psychiatry*, and *Clinical Neurology* involve some classical self-report scales/questionnaires for preliminary diagnosis, such as hospital anxiety and depression scale (Zigmond & Snaith, 1983), AUDIT (Alcohol Use Disorders Identification Test, Bush et al., 1998), and insomnia severity index (Bastien et al., 2001). These scales are widely used even now for measuring popular psychology and neurology diseases, e.g., depression and its related disorders (Drosatos & Kaldoudi, 2020) and unhealthy behaviors (e.g., unhealthy alcohol use) that might lead to neurology diseases. As for the Theory category, *Psychology, Multidisciplinary*, and *Information Science & Library Science* supply more mature theories, e.g., social cognitive theory (Bandura et al., 1986), technology acceptance model (Davis et al., 1989), which form the fundamental theories for applying technology intervention strategies (e.g., SMS text messaging and individual counseling session) to health services in the eHealth field. The Theory phrases of *Multidisciplinary Sciences* are quite old as well according to Fig. 5. The reason is that there are only 4 *Theory* phrases in this discipline, among which a classic theory model, biopsychosocial model (Engel, 1977), increases the average age at large.

By contrast, knowledge phrases in *Medical Informatics* and *Health Care Sciences & Services* are generally younger than the phrases in other disciplines. This is because that the knowledge transfer across disciplines tends to have a greater time lag than knowledge transfer within a discipline (Rinia et al., 2001). The average age of IKPs of most categories (except for *Theory*) in *Multidisciplinary Sciences* are the smallest. It may be attributed to the literature in this subject category are more likely to stay in the research front of science since they have a large proportion of young references (Rinia et al., 2001).

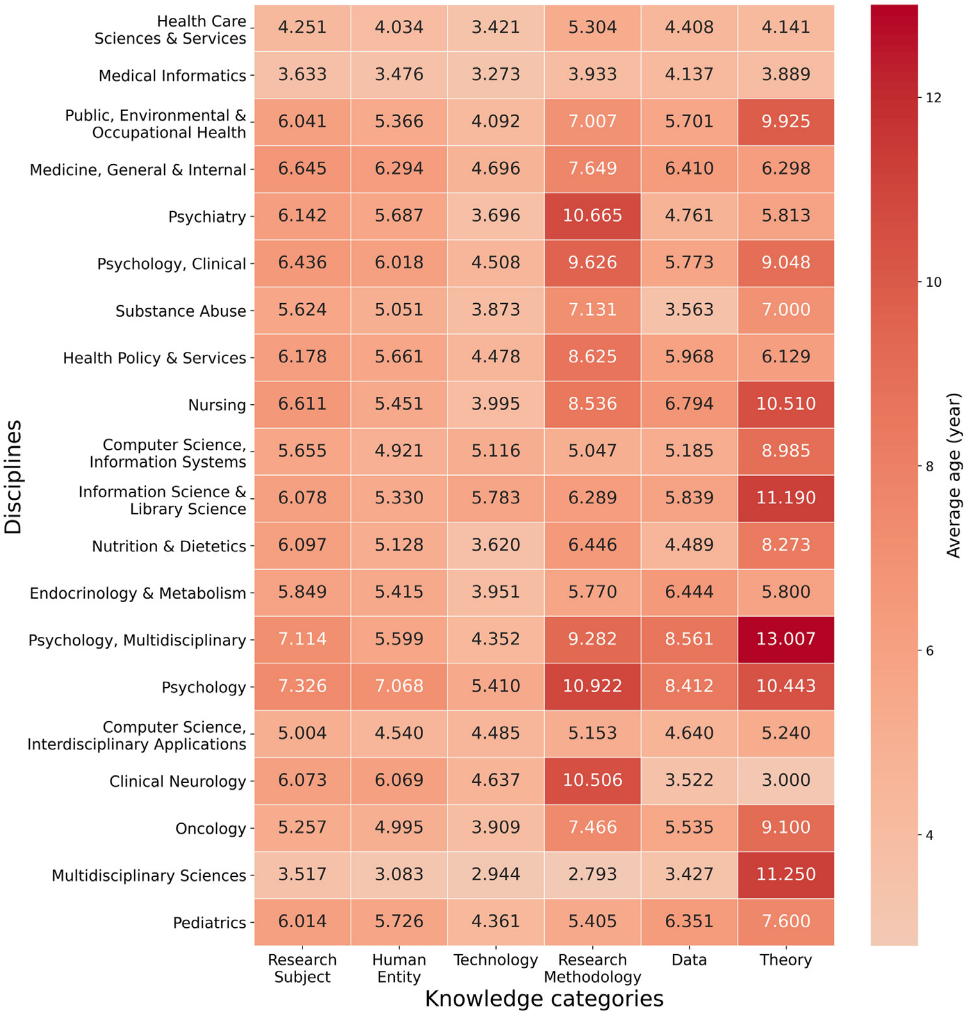

Fig. 5. Average age of IKPs in the six categories for the top 20 disciplines.

We used the Kruskal-Wallis H test to validate if the age difference among the top 20 disciplines in each category is significant, as the assumption of homogeneity of variance is violated in the two-way ANOVA analysis. The result shows that the p-value score of each category is far below 0.001, which indicates that the age distributions of IKPs in different disciplines are statistically different.

4.4. Citation location characteristics of interdisciplinary knowledge integration

To study how integrated knowledge phrases are used by the eHealth papers, we next analyze the IKPs in different sections by considering their source disciplines and knowledge categories.

4.4.1. Section distribution for IKPs of each source discipline

For each discipline, Fig. 6 shows the frequency and proportion of IKPs in the four types of sections. Some examples of IKPs appearing in different sections are shown in Appendix E.

Overall, for all disciplines, the most IKPs were used in Introduction sections, followed by Discussion, while Methods and Results sections were the fewest. This is consistent with the overall distribution of in-text citations over different sections (Table 5), and could also be attributed to the functions of different sections. Furthermore, the section distribution of IKPs over source disciplines is also related to the knowledge category distribution over source disciplines in Fig. 4. It shows that the Research Subject phrases account for the largest percentage for each source discipline, while most Research Subject related phrases would be used in Introduction or Discussion sections due to the purpose of these sections is to describe or compare relevant studies (Hu et al., 2013).

Nevertheless, the location distributions of IKPs also show some differences among the disciplines. Knowledge phrases from *Psychology* and *Psychiatry* are more likely to appear in Methods sections than those from other disciplines. This agrees with the highest proportion of IKPs of Research Methodology category in these disciplines as illustrated in Fig. 4, indicating that the phrases from

Fig. 6. The proportion and frequency of integrated knowledge phrases in different sections for each discipline.

Psychology and *Psychiatry* have a higher probability to play a methodological role in the eHealth field. *Medicine, General & Internal* has the highest proportion of IKPs in Introduction sections, which means that the IKPs from this discipline are more related to the research topics of the eHealth field and provide more relevant background information. *Endocrinology & Metabolism* owns the highest proportion in Results sections since it involves many phrases related to physiological indicators, e.g., blood pressure, which might be compared in the Results section to test the performance of some medicine or intervention strategies for addressing specific disease. As for Discussion section, the proportion in *Nutrition & Dietetics* is the highest. This is reasonable because many articles on diseases and healthy lifestyle related topics discussed more about nutrition and diet in the Discussion section, leading to the important role of *Nutrition & Dietetics* in Discussion section.

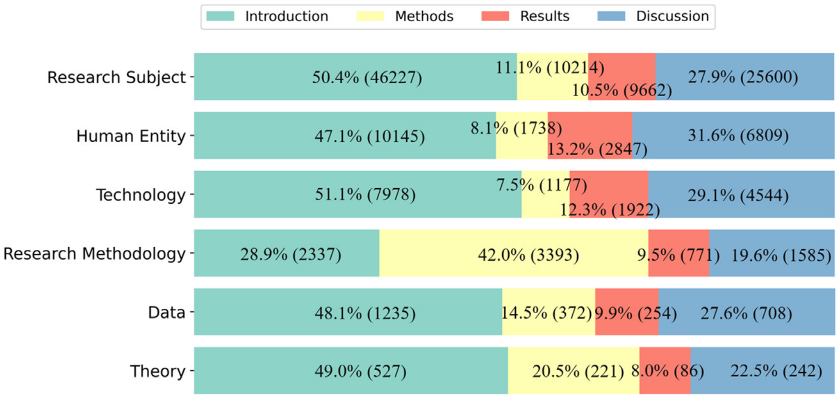

Fig. 7. The proportion and frequency of integrated knowledge phrases in the four sections for each category.

The Pearson Chi-Square test shows that the IKP distributions over the sections are relevant to the source disciplines (p-value << 0.001, df = 57), suggesting the different usage patterns of IKPs from different source disciplines, and indicating the different functions of IKPs from various source disciplines in the eHealth field. 0 cells (0.0% < 20%) have expected count less than 5, thus, the result is meaningful.

4.4.2. Section distribution for IKPs of each knowledge category

Fig. 7 presents the percentage of IKPs in the four types of sections for each knowledge category. Except for Research Methodology, all the knowledge categories have the most IKPs in Introduction sections, followed by Discussion sections, while those in Methods and Results are the fewest. The distribution is similar to the overall location distribution of in-text citations (Table 5) as well as the location distribution of IKPs in different source disciplines (Fig. 6), further indicating different functions of various sections. At the same time, it evidences that it is better to consider the locations when measuring the impact of certain types of knowledge, since different sections might have different functions although with the same type of knowledge.

There are also some distribution differences between the knowledge categories. Research Methodology phrases are mostly cited in Methods sections, reflecting a high correlation between the knowledge category and the citation location. Theory related phrases also account for a slightly larger proportion in Methods section than other categories, showing that theories are also important for developing the research methodology. In addition, the percentage of Technology phrases cited in Introduction section is the highest, over 50%. It is due to that eHealth field is focused on information and technologies application for the improvement of health care, thus there are a lot of research topics related to different technologies, e.g., social media, electronic health record, wearables (Drosatos & Kaldoudi, 2020). As for Human Entity, it plays an important role in Results and Discussion sections. This could be attributed to that most articles in eHealth field involve comparison of different medicine or intervention strategies for human entity groups with specific diseases in Results and Discussion sections.

Similarly, the result of Pearson Chi-Square test shows that the knowledge distribution differences over sections are significant among the knowledge categories (p-value << 0.001, df = 15). And, 0 cells (0.0% < 20%) have expected count less than 5, indicating the result is meaningful. This demonstrates that the use of IKPs in various sections is associated with the knowledge categories, reflecting the different discourse functions of various sections.

5. Discussion and conclusions

In this study, we introduced a novel analytical approach to analyze the characteristics of knowledge integration in an interdisciplinary field based on the content and location of citances. Full text articles in the eHealth field are used in the case study. The knowledge phrases shared between the citances and the corresponding references (i.e., IKPs) were extracted to study the integrated knowledge from source disciplines to eHealth field. Then, we quantified the interdisciplinary knowledge flow and analyzed the knowledge categories and citation locations of IKPs to manifest the characteristics of interdisciplinary knowledge integration. A few interesting findings are obtained in the case study.

5.1. Quantifying interdisciplinary knowledge flow by knowledge units

In addition to measuring the interdisciplinarity of a field through various indicators, some recent studies focused on tracing the source disciplines and quantifying the interdisciplinary knowledge flow to understand the formation of an interdisciplinary field. Traditional approach is based on citations, which does not reveal the content of the knowledge spread from the source disciplines to the interdisciplinary field (Karunan et al., 2017; Zhu & Yan, 2015). Alternatively, we proposed interdisciplinary knowledge phrases that are shared between the citances and the corresponding references to measure the knowledge spread from the source disciplines to the interdisciplinary field from content perspective. Although this approach is essentially based on citations, it considers the knowledge content carried by the citations. Integrated knowledge phrases are observable and countable, based on which the measurement is easy to implement. Our study focuses on the characteristics of the interdisciplinary knowledge integration based on the citance analysis. Furthermore, the proposed framework can be generalized to analyze any knowledge relationship between source and destination fields, if full text data of articles are available.

As for our first research question, the results in Table 6 show that according to IKPs, the top contributing source disciplines for eHealth field are related to health and medicine, psychiatry and psychology, and information technology. We found that the ranks measured by the number of IKPs and the number of citations were slightly different for several disciplines. The IKPs-based method offers more content details than the citation-based approach does to understand the relationships between the source disciplines and the interdisciplinary field according to the amount of integrated knowledge. Furthermore, the knowledge integration density per in-text citation shows that the knowledge flow from neighboring disciplines contained more IKPs per citation than other disciplines. Differentiating citations according to the density of transferred knowledge could be a different perspective than the importance (Hassan et al., 2018) or the intent (Cohan et al., 2019) of citations, which needs further investigation.

5.2. The characteristics of interdisciplinary knowledge integration based on IKPs and citation location

Citation content analysis probes into both the content and the location of citations in the articles. By applying citation content analysis to the interdisciplinary field, the knowledge content and their use in the articles' sections can reveal more details of the roles that the source disciplines play in the interdisciplinary field. Some previous studies found the connections between different categories of knowledge, e.g., issues, methods, analysis, and conclusions, with their source disciplines (Halevi & Moed, 2013). Compared to the traditional citation-count based methods, the proposed analytical framework in this study provides more valuable insights to understand what and how knowledge is integrated from source disciplines to the interdisciplinary field based on the content and location of citances. It can be used to deepen the understanding about the interdisciplinarity of an interdisciplinary field. Specifically, the knowledge categories, citation age, and citation location of IKPs are used to understand the functions of the source disciplines in the knowledge integration of eHealth field.

By analyzing the content characteristics of source disciplines from the perspectives of knowledge categories and citation age, we found some differences among the source disciplines, indicating their different roles in the development of the interdisciplinary field. Regarding the knowledge categories, the distribution is significantly different for the source disciplines (Fig. 4). Health and medicine related disciplines provided more Research Subject and Human Entity phrases, psychology related disciplines have a larger percentage of Research Methodology phrases, while Technology phrases account for a bit more in IT related disciplines. It evidences that the source disciplines may take heterogeneous functions of knowledge contribution to the eHealth field. Regarding the age distribution, different disciplines also display different patterns (Fig. 5). Knowledge phrases in *Medical Informatics* and *Health Care Sciences & Services* are relatively newer. It indicates that the knowledge from adjacent disciplines whose research topics are closer to the interdisciplinary field is absorbed into the field faster than other disciplines. This also reveals different characteristics of knowledge integrated from various source disciplines.

As for the citation location characteristics, a few previous studies have investigated the citation distribution over sections (Bertin et al., 2016; Boyack et al., 2018; Zhang et al., 2021), our results extend these studies from the citation content perspective. The results suggest that the IKPs distribution over the sections are significantly different among the source disciplines (Fig. 6), and are closely related to the knowledge category distributions of different disciplines (Fig. 4). We observed the correlation between the categories of knowledge and the sections they are used in the articles, especially for the IKPs of Research Methodology category (Fig. 7). It evidences different rhetorical functions of various sections in the scientific papers.

5.3. Limitations and future work

There are several limitations in our study, hence a few avenues for future research are suggested. Firstly, the corpus we collected covers only two open access journals due to the lack of full text in the interdisciplinary field and the considerable human efforts involved in the knowledge category annotation. To fully understand the interdisciplinarity of the eHealth, the dataset needs to be enhanced by supplementing more eHealth articles. Other retrieval strategies, e.g., by subject queries, can be used to harvest more articles. At the same time, the full text could be obtained by applying PDF parsing tools. In addition, identifying the disciplines of references can also be improved. For one thing, we did not process the papers from conference proceedings in this study, which account for a large proportion. Many of them are from Computer Science related disciplines, thus it is necessary to alleviate the bias towards these disciplines by identifying the disciplines of the references from conference proceedings. For another thing, the references metadata have a lot of typos and different forms of publication venues. Some techniques, e.g., Edit Distance algorithm (Zhang, et.al. 2021), could be applied to improve the discipline identification of references more easily and efficiently.

Secondly, the identification and categorization of integrated knowledge phrases could also be improved. We applied a stem-matching method to extract noun phrases shared by the citances and the references. To more accurately identify the knowledge spread from the references to citing papers, the text span for identifying IKPs can be extended from the citance to a broader citation context, e.g., the surrounding three sentences (Hernández-Alvarez & Gomez, 2016), and from the title and abstract of the reference to the full text of the reference. And, the cited text identification techniques (Ma et al., 2018) can also be applied to find the exact cited text spans in references. In addition, the stem-matching method fails to identify the phrases with the semantic relations of synonyms and hypernym-hyponym from the citances and the reference texts, which should also be considered as IKPs. It's our future effort to apply word embedding techniques to address this problem. And, future work can attempt to extract knowledge entities in large scale dataset by using existing domain-specific knowledge graph and machine learning methods (Ding et al., 2013b; Heffernan & Teufel, 2018). As for the categorization of IKPs, the manually annotation in this study hinders the application of the proposed analytical framework. We are attempting to use machine learning methods and deep learning methods to automate the categorization of IKPs. To do so, the labelled data in this study will be used as a training set for the supervised methods.

Last but not least, combining our analytical framework with citation network analysis to investigate the evolution of different categories of knowledge in a specific domain may be an exciting prospect, which could offer a novel perspective for understanding how technology or methodology influence the research topics and paradigm of an interdisciplinary field.

Appendix A

Examples of citances and IKPs from the source disciplines.

Disciplines	Citances	IKPs
Health Care Sciences & Services	Their availability, their ease of use, and their interactive and playful interface make them good devices to measure physical activity in patients [50].	physical activity, patient
Medical Informatics	Before participating in an interview, all patients were asked to complete the technology usage questionnaire [14] to obtain further information on the general and technological background of the patients.	patient, technology usage questionnaire
Public, Environmental & Occupational Health	A computer-tailored (but not Web-based) intervention for nutrition and physical activity in a workplace setting demonstrated increases in the frequency of strengthening and flexibility exercise compared to a delayed group [11].	nutrition, delay group, physical activity
Medicine, General & Internal	Additionally, previous studies have suggested that the use of rich media, such as videos, may improve the appeal of health interventions [29-32] and may attract and stimulate comprehension among low health literacy groups [25,33,34].	health intervention
Psychiatry	A growing body of evidence shows that clinician-assisted CCBT results in significant improvements in patients with depressive [2,6] and anxiety disorders [2,7], with results comparable with those obtained from face-to-face treatment.	depressive, patient
Psychology, Clinical	A meta-analysis that reviewed 51 positive psychology interventions across a spectrum of domains, found that positive psychology programs significantly increased well-being (mean r=.29) [29] and led to significant reductions in depressive symptoms (mean r=.31).	meta analysis, positive psychology intervention, depressive symptom
Substance Abuse	A previous study by the present authors [13] identified moderate to high 2-year test-retest reliability of items about lifetime tobacco use.	lifetime tobacco use
Health Policy & Services	Consequently, low levels of health literacy might negatively influence health outcomes, success of treatment, and medical costs [8-10].	health literacy
Nursing	A major challenge of eHealth in care coordination is to make it beneficial and easy to use for both health care providers and patients [27].	care coordination, ehealth
Computer Science, Information Systems	A variety of machine-learning methods have been used for text categorization, including Bayesian classification [6], decision trees [18], cluster classification [15], k-nearest neighbor (k-NN) algorithms [5], and neural nets [20].	machine learn method
Information Science & Library Science	AMOS has been widely used in other health communication studies [30].	AMOS
Nutrition & Dietetics	A similar study in Canada found that 51% of the adult population reported using the Internet for nutrition information in 2008 [6].	internet, nutrition, information
Endocrinology & Metabolism	A program designed specifically for Muslim diabetics was also assessed by retrospective analysis and reported beneficial effects [65].	muslim diabetic
Psychology, Multidisciplinary	Audio recordings of the focus groups were transcribed verbatim and anonymized, and a thematic analysis was undertaken [30].	thematic analysis
Psychology	Anxiety symptoms were assessed with the 7 anxiety items of the Dutch version of the Hospital Anxiety and Depression Scale (HADS) [43,44].	hospital anxiety and depression scale
Computer Science, Interdisciplinary Applications	In order to use free text information for CDSS, natural language processing is applied taking the context into consideration [47].	natural language processing, free text information
Clinical Neurology	A number of large-scale longitudinal studies have shown that several behavioral risk factors are associated with the onset and progression of diabetes, cardiovascular disease, stroke, and cognitive impairment [1-3].	cardiovascular disease, stroke
Oncology	Addressing the needs of this growing population of cancer survivors has been identified as supportive care's new challenge [23,24].	cancer survivor
Multidisciplinary Sciences	Depressive disorders are highly prevalent and are estimated to affect approximately 350 million people worldwide [1-3].	depressive disorder
Pediatrics	In total, 3 studies involved children (range 7-10 years) [19,22,23], 3 studies involved adolescents (range 11-15 years) [17,18,21], and 2 studies included both children and adolescents (range 5-12 years) [20,24].	child

Note: The citances are the original sentences with citation labels.

Appendix B
Top 10 IKPs of each source discipline.

Disciplines	Top 10 IKPs
Health Care Sciences & Services	patient, intervention, information, internet, app, participant, user, care, quality, health information
Medical Informatics	patient, intervention, internet, information, app, participant, user, health information, technology, health
Public, Environmental & Occupational Health	intervention, internet, information, patient, participant, physical activity, app, health, smoking, quality
Medicine, General & Internal	patient, intervention, information, internet, depression, participant, quality, physical activity, care, physician
Psychiatry	depression, intervention, treatment, patient, participant, anxiety, internet, app, meta analysis, disorder
Psychology, Clinical	intervention, depression, treatment, participant, anxiety, patient, internet, app, behavior, eat disorder
Substance Abuse	intervention, smoking, participant, alcohol, smoker, cessation, treatment, app, internet, alcohol consumption
Health Policy & Services	patient, care, information, quality, internet, health, intervention, physician, measure, health information
Nursing	patient, information, intervention, internet, woman, care, cancer, quality, child, participant
Computer Science, Information Systems	patient, information, technology, internet, user, health information, physician, datum, app, quality
Information Science & Library Science	patient, information, health information, internet, technology, health, user, intervention, datum, phr
Nutrition & Dietetics	intervention, physical activity, participant, weight loss, weight, obesity, app, food, child, adolescent
Endocrinology & Metabolism	diabetes, patient, intervention, weight loss, obesity, participant, hba1c, depression, weight, app
Psychology, Multidisciplinary	intervention, internet, information, depression, participant, physical activity, behavior, theory, patient, datum
Psychology	intervention, depression, participant, treatment, behavior, anxiety, patient, meta analysis, eat disorder, taxonomy
Computer Science, Interdisciplinary Applications	patient, information, technology, phr, datum, user, health, app, care, internet
Clinical Neurology	patient, depression, pain, intervention, chronic pain, participant, treatment, insomnia, stroke, dementia
Oncology	patient, cancer, information, internet, treatment, intervention, cancer patient, quality, breast cancer, care
Multidisciplinary Sciences	depression, intervention, patient, twitter, app, datum, anxiety, treatment, information, participant
Pediatrics	child, parent, intervention, adolescent, internet, patient, information, asthma, participant, app

Note: We excluded the knowledge phrases of “Others” type, and then selected the top 10 IKPs for each discipline.

Appendix C
The number of IKPs in the six categories for the top 20 source disciplines.

Subject categories	Research Subject	Human Entity	Techno-logy	Research Methodology	Data	Theory
Health Care Sciences & Services	23898	6548	6165	2217	1000	277
Medical Informatics	17527	4846	5239	1528	873	235
Public, Environmental & Occupational Health	16692	4032	2664	1324	445	267
Medicine, General & Internal	12290	2894	1581	1041	256	57
Psychiatry	9474	1612	843	1272	88	16
Psychology, Clinical	6048	973	587	725	44	83
Substance Abuse	4257	867	482	525	48	14
Health Policy & Services	3833	939	487	363	124	62
Nursing	3444	1161	437	291	68	49
Computer Science, Information Systems	2656	746	887	170	173	68
Information Science & Library Science	2629	722	834	204	205	58
Nutrition & Dietetics	3048	586	379	271	47	22
Endocrinology & Metabolism	3122	655	288	174	27	5
Psychology, Multidisciplinary	2913	489	361	287	57	141
Psychology	2662	396	210	360	17	61
Computer Science, Interdisciplinary Applications	1855	615	693	163	111	25
Clinical Neurology	2530	422	168	243	23	2
Oncology	2126	647	307	118	43	10
Multidisciplinary Sciences	2007	325	374	169	96	4
Pediatrics	1643	678	216	79	37	5

Appendix D

Examples of IKPs in the six categories.

Disciplines	Research Subject	Human Entity	Techno-logy	Research Methodology	Data	Theory
Health Care Sciences & Services	diabetes, ehealth literacy, health care	user, physician, provider	social medium, mobile phone, patient portal	systematic review, questionnaire, randomize control trial	tweet, medical record, health datum	technology acceptance model, plan behavior, health belief model
Medical Informatics	communication, web base intervention, treatment	consumer, internet user, clinician	website, facebook, smartphone	analysis, meta analysis, online trial	tweet, GPS datum, medical record	technology acceptance model, social cognitive theory, behavioral change theory
Public, Environmental & Occupational Health	physical activity, smoking, health	woman, smoker, employee	social medium, e cigarette, cessation app	questionnaire, taxonomy, focus group	tweet, school absenteeism datum, real time datum	transtheoretical model, social cognitive theory, behavioral theory
Medicine, General & Internal	care, diabetes, obesity	physician, researcher, clinician,	patient portal, telemedicine, phr	preferred reporting items, questionnaire, randomize control trial	medical record, pharmacy prescription record	chronic care model, transtheoretical model, share decision making
Psychiatry	depression, treatment, anxiety	adolescent, young people, therapist	social medium, videoconference, computerized CBT	meta analysis, hospital anxiety and depression scale, mini international neuropsychiatric interview	tweet, image, unstructured text	share decision making, social cognitive theory, attachment theory
Psychology, Clinical	anxiety, behavior, eat disorder	child, therapist, caregiver	behavior change technique, intellicare app, mobile phone	ecological momentary assessment, hospital anxiety and depression scale, questionnaire	image, tweet, sensor datum	common sense model, relational frame theory, motivational model
Substance Abuse	smoking, alcohol, alcohol consumption	smoker, user, problem drinker	social medium, e cigarette, search engine	alcohol use disorders identification test, smoking index, control trial	twitter datum, message post, image	transtheoretical model, social cognitive theory, TPB
Health Policy & Services	quality, patient activation, health literacy	physician, provider, consumer	health information technology, PHR, electronic health record	patient activation measure, report qualitative research, content analysis	randomly select narrative comment, claim datum, patient experience datum	CFIR, theoretical consumer choice model, dynamic sustainability framework
Nursing	cancer, symptom, treatment	woman, nurse, caregiver	mobile phone, E learning, telemonitoring	qualitative content analysis, heart failure index, self efficacy scale	qualitative datum, teacher report, contextual and nonverbal datum	OPT model, TPB, quality health osutcomes model
Computer Science, Information Systems	design, acceptance, adoption	user, physician, consumer	PHR, PDA, computer	machine learning, text mining, social network analysis	registry, free text, eye movement datum	technology acceptance model, social interdependence theory, information space model

(*continued on next page*)

Appendix D (*continued*)

Disciplines	Research Subject	Human Entity	Techno-logy	Research Methodology	Data	Theory
Information Science & Library Science	health literacy, acceptance, behavior	user, physician, provider	phr, social medium, information technology	analysis, qualitative content analysis, taxonomy	text datum, eye movement datum, wos	unified theory, technology acceptance model, plan behavior
Nutrition & Dietetics	physical activity, weight loss, obesity	child, adolescent, dietitian	facebook, personal digital assistant, fitbit	questionnaire,international physical activity questionnaire, food frequency questionnaire	sensecam image, phenotypic datum, dietary intake datum	transtheoretical model, self determination theory, technology base behavior change model
Endocrinology & Metabolism	diabetes, weight loss, obesity	woman, adolescent, provider	telemedicine, mobile phone, activity monitor	randomize control trial, remote food photography method, focus group	SMBG datum, self monitor datum, open loop datum	theory
Psychology, Multidisciplinary	depression, physical activity, stress	MSM, student, therapist	social medium, computer tailoring, online community	thematic analysis, taxonomy, meta analysis	qualitative datum, social medium post, behavioral datum	social cognitive theory, self determination theory, goal set theory
Psychology	depression, treatment, behavior	parent, caregiver, therapist	behavior change technique, computer, telephone	meta analysis, hospital anxiety and depression scale, randomize control trial	datum, image	social cognitive theory, control theory, activity theory
Computer Science, Interdisciplinary Applications	design, adoption, internet base intervention	user, provider, online support group	PHR, redcap, cloud computing	user center design, natural language processing, machine learning	medical record, tweet, corpus	technology acceptance model, information systems research ISR framework, information foraging theory
Clinical Neurology	chronic pain, insomnia, dementia	adult, physician, MS patient	telemedicine, mobile device, polysomnography	insomnia severity index, questionnaire, epworth sleepiness scale,	internet post, tweet, suicide death datum	rasch model, health belief model
Oncology	breast cancer, treatment, symptom	cancer patient, cancer survivor, physician	screen, facebook, webchoice	functional assessment, electronic self report assessment, feasibility study	scientifically validate datum, registry, incidence datum	TPB, social cognitive theory, preventive health model
Multidisciplinary Sciences	depression, anxiety, influenza	researcher, MSM, woman	social medium, GPS device, robot	focus group, time series analysis, control trial	search engine query datum, metadata, scientific database	rasch model, biopsychosocial model
Pediatrics	asthma, risk, physical activity	child, parent, adolescent	facebook, social medium, active video game	systematic review, postal survey, longitudinal study	self report datum, demographics, prevalence datum	social cognitive theory, share decision making

Appendix E
Examples of IKPs in the four sections.

Disciplines	Introduction	Methods	Results	Discussion
Health Care Sciences & Services	health information, effectiveness, health care	systematic review, measure, analysis	report, effect, group	finding, quality, improvement
Medical Informatics	health information, health care, ehealth literacy	analysis, datum, trial	effect, outcome, difference	effectiveness, need, development
Public, Environmental & Occupational Health	health, physical activity, smoking	measure, age, questionnaire	life, physical activity, social support	physical activity, smoke cessation, self efficacy
Medicine, General & Internal	diabetes, obesity, physical activity	systematic review, preferred reporting items, measure	physician, outcome, effect	physician, quality, treatment
Psychiatry	depression, anxiety, disorder	measure, hospital anxiety and depression scale, score	report, outcome, cost effectiveness	treatment, anxiety, effectiveness
Psychology, Clinical	depression, anxiety, eat disorder	validity, taxonomy, mini international neuropsychiatric interview	follow up, comparison, control group	self monitoring, behavior change, therapist support
Substance Abuse	smoking, alcohol, substance use	questionnaire, brief intervention, alcohol use disorders identification test	smoking, social medium, recruitment	smoke cessation, brief intervention, alcohol consumption
Health Policy & Services	health care, health literacy, patient activation	patient activation measure, consolidated criterion, short form health survey	health literacy, quality, patient activation	physician, implementation, practice
Nursing	support, physical activity, symptom	self efficacy, qualitative content analysis, clinical decision	empowerment, life, woman	cancer, self efficacy, pregnant woman
Computer Science, Information Systems	technology, health information, phr	PHR, perceive usefulness, acceptance	outcome, effect, system	technology, physician, perceive usefulness
Information Science & Library Science	health information, health literacy, behavior	analysis, qualitative content analysis, taxonomy	outcome, effect, user	health information, health literacy, development
Nutrition & Dietetics	physical activity, obesity, weight loss	physical activity, international physical activity questionnaire, behavior	weight loss, group, facebook	weight loss, physical activity, adolescent
Endocrinology & Metabolism	diabetes, obesity, weight loss	diabetes, self monitoring, age	hba1c, woman, blood pressure	diabetes, weight loss, hba1c
Psychology, Multidisciplinary	physical activity, behavior, anxiety	thematic analysis, datum, effect size	effect, treatment, SMS text message	effectiveness, physical activity, behavior
Psychology	depression, treatment, anxiety	measure, taxonomy, hospital anxiety and depression scale	effect, follow up, behavior change technique	effectiveness, anxiety, attitude
Computer Science, Interdisciplinary Applications	phr, technology, personal health record	redcap, PHR, analysis	outcome, user, development	patient portal, phr, medication
Clinical Neurology	chronic pain, insomnia, dementia	measure, validity, insomnia severity index	control group, bipolar disorder, sleep	chronic pain, treatment, adherence
Oncology	cancer, treatment, life	self efficacy, functional assessment, trial	life, quality, symptom	cancer patient, breast cancer, time
Multidisciplinary Sciences	anxiety, social medium, influenza	health outcome, trial, reliability	anxiety, month, need	anxiety, influenza, twitter
Pediatrics	child, pregnancy, obesity	child, behavior, questionnaire	child, asthma, group	child, asthma, physical activity

CRediT authorship contribution statement

Shiyun Wang: Data curation, Conceptualization, Formal analysis, Writing – original draft. **Jin Mao:** Conceptualization, Formal analysis, Writing – original draft. **Kun Lu:** Conceptualization, Formal analysis, Writing – original draft. **Yujie Cao:** Data curation, Formal analysis, Writing – original draft. **Gang Li:** Conceptualization, Formal analysis.

Acknowledgments

We are very grateful to the anonymous reviewers for their helpful comments. This study was funded by the National Natural Science Foundation of China (NSFC) grant nos. 71804135, 71921002, and 71790612. This research was also supported by the world class discipline project of the Ministry of Education “Library, Information, and Data Science” in China.

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.

Bandura, A. (1986). In *Social foundations of thought and action. Englewood Cliffs, NJ: 1986* (pp. 23–28).

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine, 2*(4), 297–307.

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.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: Analyzing text with the natural language toolkit*. O'Reilly Media, Inc..

Borgman, C. L., & Rice, R. E. (1992). The convergence of information science and communication: A bibliometric analysis. *Journal of the American Society for Information Science, 43*(6), 397–411.

Boyack, K. W., van Eck, N. J., Colavizza, G., & Waltman, L. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. *Journal of Informetrics, 12*(1), 59–73.

Brillouin, L. (1956). *Science and information theory*. New York: Academic Press.

Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. *Archives of Internal Medicine, 158*(16), 1789–1795.

Case, D. O., & Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science, 51*(7), 635–645.

Chang, Y. W., & Huang, M. H. (2012). A study of the evolution of interdisciplinarity in library and information science: Using three bibliometric methods. *Journal of the American Society for Information Science and Technology, 63*(1), 22–33.

Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social studies of science, 5*(4), 423–441.

Cohan, A., Ammar, W., Zuylen, M. V., & Cady, F. (2019). Structural scaffolds for citation intent classification in scientific publications. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota* 2019.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003.

Della Mea, V. (2001). What is e-Health (2): The death of telemedicine? *Journal of Medical Internet Research, 3*(2), e22.

Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013a). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics, 7*(3), 583–592.

Ding, Y., Song, M., Han, J., Yu, Q., Yan, E., Lin, L., & Chambers, T. (2013b). Entitymetrics: Measuring the impact of entities. *PLoS ONE, 8*(8), e71416.

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 American Society for Information Science and Technology, 65*(9), 1820–1833.

Dong, C., & Schäfer, U. (2011). Ensemble-style self-training on citation classification. In *Proceedings of 5th international joint conference on natural language processing* (pp. 623–631).

Drosatos, G., & Kaldoudi, E. (2020). A probabilistic semantic analysis of eHealth scientific literature. *Journal of telemedicine and telecare, 26*(7-8), 414–432.

Elkiss, A., Shen, S., Fader, A., Erkan, G., States, D., & Radev, D. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology, 59*(1), 51–62.

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*(4286), 129–136.

Engerer, V. (2017). Exploring interdisciplinary relationships between linguistics and information retrieval from the 1960s to today. *Journal of the Association for Information Science and Technology, 68*(3), 660–680.

Frost, C. O. (1979). The use of citations in literary research: A preliminary classification of citation functions. *The Library Quarterly, 49*(4), 399–414.

Garzone, M. A. (1997). *Automated classification of citations using linguistic semantic grammars* (master's thesis). London, Ontario, Canada: The University of Western Ontario.

González-Teruel, A., & Pérez-Pulido, M. (2020). The diffusion and influence of theoretical models of information behaviour. *Journal of Documentation*. 10.1108/JD-10-2019-0197.

Gupta, S., & Manning, C. (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).

Halevi, G., & Moed, H. F. (2013). The thematic and conceptual flow of disciplinary research: A citation context analysis of the Journal of Informetrics, 2007. *Journal of the American Society for Information Science and Technology, 64*(9), 1903–1913.

Hassan, S. U., Safder, I., Akram, A., & Kamiran, F. (2018). A novel machine-learning approach to measuring scientific knowledge flows using citation context analysis. *Scientometrics, 116*(2), 973–996.

Heffernan, K., & Teufel, S. (2018). Identifying problems and solutions in scientific text. *Scientometrics, 116*(2), 1367–1382.

Herlach, G. (1978). Multiple mention of a reference as a characteristic of the link between cited and citing article. *Journal of the American Society for Information Science, 29*(6), 308–310.

Hernández-Alvarez, M., & Gomez, J. M. (2016). Survey about citation context analysis: Tasks, techniques, and resources. *Natural Language Engineering, 22*(3), 327–349.

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Hu, Z., Chen, C., & Liu, Z. (2013). Where are citations located in the body of scientific articles? A study of the distributions of citation locations. *Journal of Informetrics, 7*(4), 887–896.

Hu, Z., Lin, G., Sun, T., & Hou, H. (2017). Understanding multiply mentioned references. *Journal of Informetrics, 11*(4), 948–958.

Huang, M. H., & Chang, Y. W. (2012). A comparative study of interdisciplinary changes between information science and library science. *Scientometrics, 91*(3), 789–803.

Karunan, K., Lathabai, H. H., & Prabhakaran, T. (2017). Discovering interdisciplinary interactions between two research fields using citation networks. *Scientometrics, 113*(1), 335–367.

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).

Lange, L. (1985). Effects of disciplines and countries on citation habits. An analysis of empirical papers in behavioral sciences. *Scientometrics, 8*(3-4), 205–215.

Lipetz, B. A. (1965). Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators. *American Documentation, 16*(2), 81–90.
Liu, L., Wang, Y., & Zhang, C. (2019). Does citation content of references from different disciplines demonstrate same distribution pattern? In *Proceedings of the association for information science and technology: 56* (pp. 719–721).
Liu, Y., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology, 61*(2), 340–351.
Lu, C., Ding, Y., & Zhang, C. (2017). Understanding the impact change of a highly cited article: A content-based citation analysis. *Scientometrics, 112*(2), 927–945.
Ma, S., Xu, J., & Zhang, C. (2018). Automatic identification of cited text spans: A multi-classifier approach over imbalanced dataset. *Scientometrics, 116*(2), 1303–1330.
Mao, J., Wang, S., & Shang, X. (2020). Investigating interdisciplinary knowledge flow from the content perspective of citances. *EEKE@JCDL 2020*, 40–44.
McCain, K. W., & Salvucci, L. J. (2006). How influential is Brooks' Law? A longitudinal citation context analysis of Frederick Brooks' the Mythical Man-Month. *Journal of Information Science, 32*(3), 277–295.
McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.
Meng, R., Lu, W., Chi, Y., & Han, S. (2017). Automatic classification of citation function by new linguistic features. *iConference 2017 proceedings*.
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).
Miller, M., & Mansilla, V. B. (2004). *Thinking across perspectives and disciplines*. Boston, MA: Harvard Graduate School of Education.
Morillo, F., Bordons, M., & Gómez, I. (2003). Interdisciplinary in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology, 54*(13), 1237–1249.
Neumann, M., King, D., Beltagy, I., & Ammar, W. (2019). Scispacy: Fast and robust models for biomedical natural language processing. arXiv:1902.07669.
Nichols, L. G. (2014). A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics, 100*(3), 741–754.
Oppenheim, C., & Renn, S. P. (1978). Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science, 29*(5), 225–231.
Pagliari, C., Sloan, D., Gregor, P., Sullivan, F., Detmer, D., Kahan, J. P., et al. (2005). What is eHealth (4): A scoping exercise to map the field. *Journal of Medical Internet Research, 7*(1), e9.
Pan, X., Yan, E., Cui, M., & Hua, W. (2018). Examining the usage, citation, and diffusion patterns of bibliometric mapping software: A comparative study of three tools. *Journal of Informetrics, 12*(2), 481–493.
Perier-Camby, J., Bertin, M., Atanassova, I., & Armetta, F. (2019). A preliminary study to compare deep learning with rule-based approaches for citation classification. In *Proceedings of 8th international workshop on bibliometric-enhanced information retrieval (BIR) co-located with the 41st European conference on information retrieval* (pp. 125–131).
Pettigrew, K. E., & McKechnie, L. (2001). Use of theory in information science research. *Journal of the American Society for Information Science and Technology, 52*(1), 62–73.
Poncela-Casasnovas, J., Gerlach, M., Aguirre, N., & Amaral, L. A. N. (2019). Large-scale analysis of micro-level citation patterns reveals nuanced selection criteria. *Nature Human Behaviour, 3*(6), 568–575.
Porter, A. L., & Chubin, D. E. (1985). An indicator of cross-disciplinary research. *Scientometrics, 8*(3–4), 161–176.
Porter, A. L., Cohen, A. S., David Roessner, J., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics, 72*(1), 117–147.
Radoulov, R. (2008). *Exploring automatic citation classification* (master's thesis). Waterloo, Ontario, Canada: The University of Waterloo.
Rafols, I., & &Meyer, M. (2010). Diversity measures and network centralities as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics, 82*, 263–287.
Rinia, E. J., Leeuwen, T. Van, Leeuwen, T. Van, Bruins, E. E. W., & Raan, T. van (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*, 198–215.
Shen, L., Wang, S., Dai, W., & Zhang, Z. (2019). Detecting the interdisciplinary nature and topic hotspots of robotics in surgery: social network analysis and bibliometric study. *Journal of Medical Internet Research, 21*(3), e12625.
Small, H. (2018). Characterizing highly cited method and non-method papers using citation contexts: The role of uncertainty. *Journal of Informetrics, 12*(2), 461–480.
Small, H. G. (2011). Interpreting maps of science using citation context sentiments: A preliminary investigation. *Scientometrics, 87*(2), 373–388.
Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science, 8*(3), 327–340.
Sollaci, L. B., & Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: A fifty-year survey. *Journal of the medical library association, 92*(3), 364.
Spiegel-Rosing, I. (1977). Science studies: Bibliometric and content analysis. *Social Studies of Science, 7*(1), 97–113.
Steele, T. W., & Stier, J. C. (2000). The impact of interdisciplinary research in the environmental sciences: A forestry case study. *Journal of the American Society for Information Science, 51*(5), 476–484.
Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface, 4*(15), 707–719.
Stopar, K., Drobne, D., Eler, K., & Bartol, T. (2016). Citation analysis and mapping of nanoscience and nanotechnology: Identifying the scope and interdisciplinarity of research. *Scientometrics, 106*(2), 563–581.
Taşkın, Z., & Al, U. (2018). A content-based citation analysis study based on text categorization. *Scientometrics, 114*(1), 335–357.
Teufel, S. (1999). *Argumentative zoning: Information extraction from scientific text*. Edinburgh: University of Edinburgh.
Teufel, S., Siddharthan, A., & Tidhar, D. (2006a). An annotation scheme for citation function. In *Proceedings of the 7th SIGdial workshop on discourse and dialogue* (pp. 80–87).
Teufel, S., Siddharthan, A., & Tidhar, D. (2006b). Automatic classification of citation function. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 103–110).
Thelwall, M. (2019). Should citations be counted separately from each originating section? *Journal of Informetrics, 13*(2), 658–678.
Tsai, C. T., Kundu, G., & Roth, D. (2013). Concept-based analysis of scientific literature. In *Proceedings of the 22nd ACM international conference on Information and knowledge management* (pp. 1733–1738).
Tsay, M. Y. (2009). Citation analysis of Ted Nelson's works and his influence on hypertext concept. *Scientometrics, 79*(3), 451–472.
Van Leeuwen, T., & Tijssen, R. (2000). Interdisciplinary dynamics of modern science: Analysis of cross-disciplinary citation flows. *Research Evaluation, 9*(3), 183–187.
Voos, H., & Dagaev, K. S. (1976). Are all citations equal? Or, did we op. cit. your idem? *Journal of Academic Librarianship, 1*(6), 19–21.
Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., et al. (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.
Yan, E. (2016). Disciplinary knowledge production and diffusion in science. *Journal of the Association for Information Science and Technology, 67*(9), 2223–2245.
Zhai, Y., Ding, Y., & Wang, F. (2018). Measuring the diffusion of an innovation: a citation analysis. *Journal of the Association for Information Science and Technology, 69*(3), 368–379.

Zhang, C., Liu, L., & Wang, Y. (2021). Characterizing references from different disciplines: A perspective of citation content analysis. *Journal of Informetrics, 15*(2), Article 101134.

Zhang, G., Ding, Y., & Milojevic, S. (2013). Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content. *Journal of the Association for Information Science and Technology, 64*(7), 1490–1503.

Zhang, L., Thijs, B., & Glänzel Wolfgang, W. (2011). The diffusion of H-related literature. *Journal of Informetrics, 5*(4), 583–593.

Zhu, Y., & Yan, E. (2015). Dynamic subfield analysis of disciplines: An examination of the trading impact and knowledge diffusion patterns of computer science. *Scientometrics, 104*(1), 335–359.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361–370.