A Review of Automatic Citation Summarization Research (Post-print)

Authors: Liu Tianyi, Step one, Zhao Danqun, Huang Wenbin

Date: 2017-10-11T00:00:00+00:00

Abstract

【Purpose】 To review and analyze mainstream international research methods and procedures in the field of citation-based summarization.

【Literature Scope】 Important research in citation-based summarization since 2007 and prior advances in automatic summarization and citation analysis are selected.

【Method】 Based on literature investigation, the basic concepts of this field and the application of natural language processing methods in citation-based summarization are introduced.

【Results】 Citation sentences play significant summarizing, indicative, and relational roles in summarization practice, offering certain advantages.

【Limitations】 A lack of comparison between existing achievements in citation-based summarization and potentially ideal scenarios.

【Conclusion】 Citation-based summarization expands research directions for both automatic summarization and traditional informetrics, necessitates improvements to existing evaluation schemes for automatic summarization, and generates a series of new issues such as citation window expansion and corpus construction. This paper discusses these issues and provides an outlook on future research development in citation-based summarization.

Full Text

Preamble

ChinaXiv Partner Journal, Issue 270, 2016, Issue 5
A Review of Citation-Based Automatic Summarization Research
Liu Tianyi, Bu Yi, Zhao Danqun, Huang Wenbin
(Department of Information Management, Peking University, Beijing 100871)

Abstract

[**Objective**] This paper provides an in-depth review of popular research methodologies adopted in Citation-Based Summarization (CBS) studies. [**Coverage**] We retrieved scholarly papers on CBS published since 2007, as well as earlier research on automatic summarization and citation analysis. [**Methods**] Based on a comprehensive literature review, we introduce the fundamental concepts of this field and discuss the application of natural language processing methods in citation-based summarization. [**Results**] Citances play crucial summarizing, indicative, and relational roles in summarization practice, demonstrating distinct advantages. [**Limitations**] The review lacks a comparison between current achievements and potential ideal outcomes in the CBS field. [**Conclusions**] Citation-based summarization expands the research directions of both automatic summarization and traditional informetrics. It also calls for improved evaluation schemes for automatic summarization and raises new issues such as citation window expansion and corpus construction. This paper explores these issues and outlines future research directions for citation-based summarization.

Keywords: Automatic summarization; Citation-based summarization; Citance; Natural Language Processing
Classification Number: G350

Automatic summarization is a significant research topic in natural language processing. Traditionally, research has focused primarily on summary generation techniques and methods based on full-text information. While abstracts derived exclusively from main text (including author-written abstracts) can adequately reflect document content, their capacity to summarize the impact of cited works is limited, and they cannot reflect the diachronic changes in a document' s influence.

Citation (and being cited) relationships represent one of the most valuable associative relationships among academic documents, which can be viewed as a form of summary or interpretation of the cited document' s content from other scholars' perspectives, thereby reflecting the cited document' s academic impact or value on other research. Consequently, drawing upon citation analysis concepts from bibliometrics, research on summaries based on citation information (referred to as "citation-based summarization") has gradually emerged as a new exploration direction in the automatic summarization field over the past decade.

The fundamental approach of citation-based summarization research can be summarized as follows: locate all citing documents of a target paper, extract all sentences containing citation markers and other relevant information from these full texts as a collection, then select a subset from this collection to process and generate a summary of the target document, ensuring the subset achieves sufficient compression ratio and good summarization capability. To this end, the main research steps (or key issues) include: selecting an appropriate full-text corpus; identifying and extracting citation sentences (contexts); recognizing

citation types and purposes to classify and filter citation sentences; organizing and ranking citation sentences to form an initial summary; post-processing the summary; and evaluating the final result.

Automatic summarization has long been widely applied in information retrieval. Bradshaw proposed that citation relationships could improve academic retrieval effectiveness, reflecting the superiority of citation-based summarization. Additionally, research shows that for highly-cited documents, citation-based summaries are more objective and diverse, and demonstrate clear advantages in revealing target document information. Compared to main text-based summarization methods, citation-based summaries are not only more generalizable than original sentences but also, having undergone a round of human analysis, possess certain commentary and extension qualities that better reflect the significant parts of the original text.

The rise of citation-based summarization research has primarily benefited from the increasing availability of full-text corpora and advances in natural language understanding technology. It can also be considered an important application of citation context analysis technology. Since Qazvinian et al. first conducted experimental research on citation-based summarization in 2008, this new research direction, developed through the intersection and integration of "automatic summarization" and "citation analysis," has not only attracted widespread scholarly attention and achieved significant research progress but has also gradually shifted from initially over-relying on bibliometric and citation analysis methods to paying more attention to natural language processing technologies, particularly new techniques such as text semantics and sentiment analysis. However, limited by the scarcity of citation corpus information for low-cited documents and the relative difficulty of obtaining structured full-text corpora (except in the medical field), there remains considerable room for expansion in deeper research, promotion, and application of citation-based summarization techniques.

A search on the Web of Knowledge platform using the query TI = (citation* OR reference* OR bibliography) AND TI = (summar* OR survey OR extract* OR abstract), refined by research direction, yielded over 300 documents from 2007-2015—a relatively small number, with only dozens directly related to citation-based summarization. This indicates that academic research on citation-based summarization is still in its infancy. Domestic research in automatic summarization is also limited, and according to search results from CNKI, Wanfang, and other databases, research on citation-based summarization in China is virtually non-existent. Through comprehensive literature review, this paper attempts to systematically analyze and review the progress of foreign research on this emerging topic, starting from the key steps of (single-document) citation-based summarization, clarify some basic concepts, and provide necessary reference and inspiration for the in-depth development of domestic citation-based summarization research.

2. Citation Context Identification and Citation Window Expansion

Citation-based summarization research is primarily built upon the cognitive foundation of mining and utilizing the thematic relevance and value embedded in citation relationships. Therefore, citation context identification and extraction is not only one of the key steps in this research but also the first problem that must be properly addressed, as it provides crucial support for subsequent steps and summary generation.

As early as 2004, Nakov et al. coined the term "Citance" (Citation Sentence), referring to sentences surrounding citation markers in citing documents. In its narrow sense, a "citance" can be understood as the sentence containing the citation marker itself, while in its broad sense, it can be extended to include the sentence containing the citation marker and its surrounding information, i.e., citation context or citation contextual information. Typically, a citation context can be viewed as the passage in a citing document that discusses the cited document, where this discussion should be relatively explicit, meaningful, and have identifiable boundaries or thresholds. The scope of the citation context can be termed the "citation window."

In early research, citation context identification and extraction were mostly based on the narrow understanding, typically extracting the sentence containing the citation marker directly through the identification of characteristic boundaries. For example, Nanba et al. manually identified 86 clue words related to citation context identification and completed the extraction task based on these words. More recently, the common approach has been to use regular expressions to describe patterns of various citation styles.

Recently, researchers have begun to adopt methods of expanding the citation window to extract relevant information, attempting to utilize sentences near the citation sentence or other semantically similar sentences to effectively improve the richness of information for citation-based summarization.

Citation window expansion is important for understanding citation practices, citation motivations, and the characteristics of information flow in scholarly citations. Currently, there are two main methods for citation window expansion: distance-based expansion and similarity-based expansion. Distance-based expansion methods are relatively simple, generally assigning different weights to other sentences based on their physical distance from the citation sentence to determine whether they should be included in the citation context, or directly specifying the number of sentences (or paragraphs) that can be included in the citation window. Similarity-based expansion methods are relatively more complex, aiming to identify and incorporate sentences with high semantic relevance to the citation sentence (not necessarily adjacent in position), making them particularly suitable for "implicit citation" scenarios. For example, when academic documents use the Harvard system for citation or marking, it often results in numerous implicit citations. Athar et al.' s research suggests that implicit cita-

tions often contain richer semantic information and have significant utilization value, making it necessary to expand them into the citation window.

The quality of citation context extraction directly determines the quality of citation-based summarization. Currently, citation context identification primarily relies on extracting boundary features of citations, while citation window expansion depends more on analyzing lexical and semantic features of natural language text. There remains considerable room for improvement and optimization in the future, particularly regarding the identification of topics and viewpoints in citation contexts. Although few studies have combined this with citation-based summarization, it represents a meaningful direction for improvement.

3. Citation Sentence Classification

Citation sentence classification aims to organize and structure the set of citation sentences extracted from all citation contexts for a target document according to certain structural (or logical) criteria. For example, citation sentences can be grouped or classified by research purpose, method, limitation, conclusion, etc., enabling preliminary screening and filtering to ensure the summary' s comprehensiveness and conciseness while facilitating organization according to its internal logical sequence to guarantee readability.

Early citation sentence classification primarily drew upon research findings and approaches regarding author citation behavior or citation motivation, including whether the author' s evaluation of the cited document was positive or negative, which could reflect the relationship between citing and cited documents to some extent. For instance, Nanba et al. summarized citation sentences into three types (Type B/C/O) based on 15 types of author citation behavior identified by Garfield in 1965.

In such research, early studies mostly relied on the identification of characteristic words and clue words, while more recent research has increasingly introduced sentiment analysis techniques.

Emerging abstract structure theories have provided new possibilities for citation sentence classification. These theories initially emerged from research classifying academic abstract sentences according to their article sections. Specific research efforts mainly include: studying the structure of documents or document abstracts to provide guidance for automatic summarization; and studying classification techniques and features under certain abstract structures. The former often serves as the theoretical foundation for the latter to ensure that classification results place citation sentences in appropriate positions in the summary, conforming to readers' reading habits (logic).

Classification based on natural document segmentation is intuitive and has been applied in many studies. Bai Guangzu et al. used a Naive Bayes algorithm to identify academic article section names under small sample conditions, achiev-

ing good results. However, natural segmentation varies across different documents, leading to inherent limitations of this approach. Consequently, new classification theories have been developed. Teufel et al. proposed the rigorous Argumentative Zoning (AZ) model. The AZ theory initially divided the rhetorical status of sentences in articles into seven categories: aim, structure, author' s own claim, background, contrast, basis, and other, and further refined them into 15 components in the subsequent AZ-II version. This refinement was considered more informative and more adaptable to different disciplines. Another important theory is the Core Scientific Concept (CoreSC) structure theory, whose classification is even more detailed than AZ theory. Comparative studies of these three classification theories (natural segmentation, AZ theory, CoreSC) found that using machine learning to classify abstract sentences according to their categories was reliable, and there were potential connections between their categories. Table 1 shows some important recent research on academic abstract structure:

Table 1. Theories Related to Academic Abstract Structure Research

First Author/Year	Zone Classes	Features	Classifier
Mizuta/2006[19]	Group 1: Background, Problem, Author' s own claims (Method, Conclusion, Insight, Implication, Significance, Other)Group 2: Related work, DistinctionGroup 3: Outline	Words, main verbs, tense, modality, chunk or paragraph, sentence, citation, underlined part, sentence object	Human annotator

First Author/Year	Zone Classes	Features	Classifier
Teufel/2006[20]	Argumentative Zoning: Aim, Structure, Author' s own claims, Background, Contrast, Basis, Other	Chunk structure, paragraph structure, title, sentence length, term frequency*inverse document frequency, content, verb tense, verb modality, citation, etc.	Naive Bayes classifier
Ehrler/2005[21]	Intention, Method, Result, Conclusion	Distance, term frequency*inverse document frequency* (*tf*idf)	Vector space classifier, regular expression matching classifier
Hiroakata/2008[22]	Aim, Method, Result, Conclusion, two custom prefix classes	N-grams, related sentences, sentence position, citation context features	Conditional random fields, support vector machine
Teufel/2009[23]	Argumentative Zoning-II (AZ-II)	Refined based on original version	Human annotator
Liakata/2010[24], Liakata/2013[17]	Core Scientific Concepts (CoreSC)	Hypothesis, Motivation, Background, Goal, etc. (18 categories)	Human annotator

Contractor et al. used AZ theory as features for both citation sentence classification and filtering, achieving good results. Overall, research on abstract

structure theories is gradually avoiding subjectivity and moving toward more universal feature selection and processing techniques.

4. Citation Sentence Organization and Sorting

Citation sentence organization and sorting refers to selecting the sentences with the strongest content expression capability from the organized candidate citation sentence set and ranking them according to certain methods to generate a citation-based summary with strong summarization and logical coherence.

Current research on citation sentence organization and sorting mainly includes: (1) **Clustering citation sentences** using certain similarity measures. Clustering is generally based on content similarity between citation sentences to eliminate semantically similar sentences. Common clustering methods include hierarchical clustering; Maximum Marginal Relevance (MMR) and its variants, a commonly used baseline algorithm; and MEAD and its variants. MEAD is an open, free automatic summarization research platform that provides various basic summarization algorithms and is frequently used as a baseline by researchers in evaluations. Its variants include MEAD-Centroid and MEAD-LexRank. Additionally, this system has extensive applications in multi-document summarization. Moreover, because citation sentences and original documents, as well as citation sentences themselves, have inherent citation relationships that can form citation networks, citation sentence clustering is particularly suitable for the Influence Graph Summarization (IGS) problem. Shi et al. compressed large-scale citation network graphs to track research topic evolution and reveal importance, which, although not a typical citation-based summarization problem, offers valuable insights for the field in terms of graph model thinking, visualization methods, and algorithmic improvements.

(2) **Scoring (weighting) citation sentences** for ranking output. For example, Mei et al. designed an influence-based scoring scheme to rank citation sentences that reflect document impact; Qazvinian et al. ranked citation sentences based on keyword extraction and keyword density and importance information, and performed deduplication accordingly. Currently, research relying solely on scoring citation sentences is decreasing. Since scoring generally depends on words in sentences, such as specific named entities or factual occurrence features, it inevitably leads to high-scoring sentences often pointing to similar objects or expressing similar meanings, objectively affecting the diversity of summary sentences.

5. Post-processing

Post-processing refers to examining the selected citation sentences in the initial summary draft, including checking for redundancy, ambiguous references, and incoherence in the use of key terms. Researchers have increasingly focused on post-processing summary results in recent years.

Currently, post-processing addresses two main issues: (1) **Deduplication**. A simple case of semantic repetition in citation sentences is when sentence pairs use basically the same words and express the same meaning. In this case, deduplication is relatively simple and can be solved through word-level similarity calculation. Generally, if corresponding citation classification methods are available, there is often no situation of extremely high sentence similarity. More complex semantic repetition problems in citation sentences are more difficult to deduplicate. For example, two citation sentences both cite an article about "information retrieval," but one uses the term "search engine result" while the other uses "information retrieval." Although the words are completely different, they express similar semantics, which often cannot be discovered through simple similarity comparison and usually requires cooperation with relevant knowledge bases to resolve.

(2) **Coherence**. Generally, citation sentences differ in format from main text. Additionally, citation sentences usually come from different citing documents with significant differences in expression habits. Therefore, compared to main text-based automatic summarization, citation-based summaries have poorer readability and coherence. The main solution is to analyze the structure of citation-based summaries and the lexical characteristics of human-written summaries, then insert or replace appropriate pronouns and conjunctions to improve summary coherence.

6. Evaluation

Evaluation mainly refers to qualitative and quantitative assessment of summaries regarding their summarization capability, coherence, accuracy, grammatical correctness, and readability. Traditional automatic summarization evaluation generally uses Recall-Precision and F-measure methods to measure the comprehensiveness and accuracy of template units designated in the original text that are covered by the summary. Another widely used evaluation scheme is ROUGE, which is primarily based on recall calculation using n-gram models, with variants including ROUGE-P, ROUGE-S, ROUGE-L, etc.

To date, there are almost no evaluation schemes specifically designed for citation-based summarization. Therefore, ROUGE and Precision-Recall are still widely used. However, these two schemes cannot fully meet the evaluation requirements of citation-based summarization. Citation wording may deviate from the original text, making it difficult to judge coverage capability and accuracy. Since citation sentences from multiple documents are independent of each other, citation-based summaries may contain more semantic redundancy than general automatic summaries, inevitably reducing readability, which existing evaluation schemes cannot measure.

Many citation-based summarization evaluations are based on human involvement. For example, comparing with human-generated high-quality summaries for evaluation or inviting domain experts to score generated citation-based sum-

maries. A similar approach is the question-answering method, where professionals read the summary content and then answer questions about the original text to evaluate whether the summary captures key points. These schemes are labor-intensive and difficult to avoid subjectivity in human processing. In a recent study, Christensen et al. used pure human evaluation methods, inviting professional evaluators to conduct "blind evaluation" of automatically generated citation-based summaries against baseline summaries, placing new emphasis on coherence and readability evaluation.

It should be noted that this review focuses on single-document citation-based summarization. In recent years, based on single-document citation-based summarization research, multi-document citation-based summarization and automatic literature review generation have also attracted active scholarly attention. Since the latter two types of summarization are significantly more difficult than single-document summarization, particularly in citation sentence classification, clustering, semantic deduplication, and sorting, they present higher technical difficulty and challenges. Therefore, in future research, how to conduct deeper semantic understanding and sentiment analysis of citation sentence corpora, how to form linkage relationships between single and multi-document summarization based on the indicative and relational functions of citation sentences, and how to integrate with main text-based automatic summarization technology will all become the next directions for citation-based summarization research.

Conclusion

The proposal of citation-based summarization and related research has a history of less than 10 years. Based on content elements and starting from the key steps of citation-based summarization, this paper has conducted comprehensive literature review and analysis of foreign research progress in this emerging field. From early qualitative discussions of citation sentence functions and roles, to in-depth analysis of citation sentence content, viewpoints, and tendencies, to automatic summary generation and evaluation based on citation sentence sets; from single-document summarization to multi-document summarization, and then to automatic generation of literature reviews based on larger-scale documents, research in the citation-based summarization field is gradually maturing and deepening.

However, as an emerging research topic, citation-based summarization research still has many shortcomings and faces numerous difficulties, specifically manifested as: (1) **Lack of applicable full-text corpora**. Currently, except for PubMed, structured full-text corpora based on XML markup are rare in other disciplines, and due to imperfect OCR technology, the noise in obtained full-text texts is still significant. Therefore, citation-based summarization research often relies on manual preprocessing of full-text data in many cases, greatly affecting research efficiency. (2) **Lack of targeted evaluation schemes**. Traditional automatic summarization evaluation schemes cannot directly reflect the quality of citation-based summarization, while some of the most effective evaluation

schemes lack widely recognized theoretical support. Moreover, over-reliance on human scoring makes large-scale evaluation impossible. (3) **Length variability**. Citation-based summarization length varies greatly and lacks flexibility, affected by the number of citation sentences and the document' s discipline. Few studies have discussed the appropriate length for citation-based summarization. (4) **Inapplicability to low-cited documents**. Citation-based summarization methods cannot be applied to the majority of academic literature that is low-cited.

References

[1] Mei Q, Zhai C. Generating Impact-Based Summaries for Scientific Literature [C]. In: Proceedings of ACL-08: HLT, 2008: 816-824.
[2] Bradshaw S. Reference Directed Indexing: Redeeming Relevance for Subject Search in Citation Indexes [C]. In: Proceedings of the 7th European Conference on Research and Advanced Technology on Digital Libraries (ECDL 2003), Trondheim, Norway. Springer, 2003: 499-510.
[3] Elkiss A, Shen S, Fader A, et al. Blind Men and Elephants: What do Citation Summaries Tell Us about a Research Article? [J]. Journal of the American Society for Information Science and Technology, 2008, 59(1): 51-62.
[4] Mohammad S, Dorr B, Egan M, et al. Using Citations to Generate Surveys of Scientific Paradigms[C]. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2009: 584-592.
[5] Kan M-Y, Klavans J L, McKeown K R. Using the Annotated Bibliography as a Resource for Indicative Summarization [C]. In: Proceedings of LREC, Las Palmas, Spain. 2002: Experiments with Relevance and Rhetorical Status [J]. Computational Linguistics, 2002, 28(4): 409-445.
[6] Qazvinian V, Radev D R. Scientific Paper Summarization Using Citation Summary Networks[C]. In: Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1, 2008: 689-696.
[7] 王连喜. 自动摘要研究中的若干问题 [J]. 图书情报工作, 2014, 58(20): 13-22. (Wang Lianxi. Issues in Automatic Summarization Research [J]. Library and Information Service, 2014, 58(20): 13-22.)
[8] Nakov P I, Schwartz A S, Hearst M A. Citances: Citation Sentences for Semantic Analysis of Bioscience Text [C]. In: Proceedings of the SIGIR' 04 Workshop on Search and Discovery in Bioinformatics, 2004: 81-88.
[9] Nanba H, Kando N, Okumura M. Classification of Research Papers Using Citation Links and Citation Types: Towards Automatic Review Article Generation [J]. Advances in Classification Research Online, 2000, 11(1): 117-134.
[10] Nanba H, Okumura M. Towards Multi-paper Summarization Using Reference Information [C]. In: Proceedings of the 16th International Joint Conference on Artificial Intelligence, 1999: 926-931.
[11] 刘洋, 崔雷. 引文上下文在文献内容分析中的信息价值研究 [J]. 图书情报工作, 2014, 58(6): 101-104. (Liu Yang, Cui Lei. The Information Value of Citation Context in Document Content Analysis [J]. Library and Information Service, 2014, 58(6):

101-104.)
[12] Qazvinian V, Radev D R. Identifying Non-explicit Citing Sentences for Citation-based Summarization [C]. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010: 555-564.
[13] Athar A, Teufel S. Detection of Implicit Citations for Sentiment Detection [C]. In: Proceedings of the Workshop on Detecting Structure in Scholarly Discourse, 2012: 18-26.
[14] 白光祖, 何远标, 马建霞, 等. 利用小样本量机器学习实现学术文摘结构的自动识别 [J]. 现代图书情报技术, 2014(7-8): 34-40. (Bai Guangzu, He Yuanbiao, Ma Jianxia, et al. Application of Machine Learning with Limited Corpus to Identify Structure of Scientific Abstracts Automatically [J]. New Technology of Library and Information Service, 2014(7-8): 34-40.)
[15] Teufel S. Argumentative Zoning: Information Extraction from Scientific Text [D]. Edinburgh: University of Edinburgh School of Cognitive Science, 2000.
[16] Teufel S, Moens M. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status [J]. Computational Linguistics, 2002, 28(4): 409-445.
[17] Liakata M, Dobnik S, Saha S, et al. A Discourse-Driven Content Model for Summarising Scientific Articles Evaluated in a Complex Question Answering Task [C]. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, USA. 2013: 747-757.
[18] Guo Y, Korhonen A, Liakata M, et al. Identifying the Information Structure of Scientific Abstracts: An Investigation of Three Different Schemes [C]. In: Proceedings of the 2010 Workshop on Biomedical Natural Language Processing (ACL 2010), 2010: 99-107.
[19] Mizuta Y, Korhonen A, Mullen T, et al. Zone Analysis in Biology Articles as a Basis for Information Extraction [J]. International Journal of Medical Informatics, 2006, 75(6): 468-487.
[20] Teufel S. Argumentative Zoning for Improved Citation Indexing [A]. //Computing Attitude and Affect in Text: Theory and Applications [M]. Netherlands: Springer, 2006: 45-55.
[21] Ehrler F, Geissbühler A, Jimeno A, et al. Data-poor Categorization and Passage Retrieval for Gene Ontology Annotation in Swiss-Prot [J]. BMC Bioinformatics, 2005, 6(S1): S23.
[22] Hirohata K, Okazaki N, Ananiadou S, et al. Identifying Sections in Scientific Abstracts Using Conditional Random Fields [C]. In: Proceedings of the International Joint Conference on Natural Language Processing, 2008: 381-388.
[23] Teufel S, Siddharthan A, Batchelor C. Towards Discipline-independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics [C]. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3, 2009: 1493-1502.
[24] Liakata M, Teufel S, Siddharthan A, et al. Corpora for the Conceptualisation and Zoning of Scientific Papers [C]. In: Proceedings of the International Conference on Language Resources and Evaluation, 2010: 2054-2061.

[25] Contractor D, Guo Y, Korhonen A. Using Argumentative Zones for Extractive Summarization of Scientific Articles [C]. In: Proceedings of the International Conference on Computational Linguistics, 2012: 663-678.
[26] Abu-Jbara A, Radev D. Coherent Citation-based Summarization of Scientific Papers [C]. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011: 500-509.
[27] Carbonell J, Goldstein J. The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries [C]. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1998: 335-336.
[28] Qazvinian V, Radev D R, Özgür A. Citation Summarization Through Keyphrase Extraction [C]. In: Proceedings of the 23rd International Conference on Computational Linguistics, 2010: 895-903.
[29] Mollá D, Jones C, Sarker A. Impact of Citing Papers for Summarisation of Clinical Documents[C]. In: Proceedings of the Australasian Language Technology Association Workshop, 2014: 79.
[30] Jaidka K, Chandrasekaran M K, Jha R, et al. The Computational Linguistics Summarization Pilot Task [C]. In: Proceedings of Text Analysis Conference, 2014.
[31] Radev D, Allison T, Blair-Goldensohn S, et al. MEAD-A Platform for Multidocument Multilingual Text Summarization [C]. In: Proceedings of Conference on Language Resources and Evaluation, 2004: 699-702.
[32] Chen J, Zhuge H. Summarization of Scientific Documents by Detecting Common Facts in Citations [J]. Future Generation Computer Systems, 2014, 32: 246-252.
[33] Galgani F, Compton P, Hoffmann A. Summarization Based on Bi-directional Citation Analysis [J]. Information Processing & Management, 2015, 51(1): 1-24.
[34] Erkan G, Radev D R. LexRank: Graph-based Lexical Centrality as Salience in Text Summarization[J]. Journal of Artificial Intelligence Research, 2004, 22: 457-479.
[35] Shi L, Tong H, Tang J, et al. VEGAS: Visual influEnce GrAph Summarization on Citation Networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(12): 3417-3431.
[36] Christensen J, Mausam S S, Soderland S, et al. Towards Coherent Multi-Document Summarization [C]. In: Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013: 1163-1173.
[37] Barzilay R, Lapata M. Modeling Local Coherence: An Entity-based Approach [J]. Computational Linguistics, 2008, 34(1): 1-34.
[38] Lin C-Y. Rouge: A Package for Automatic Evaluation of Summaries [C]. In: Proceedings of the Workshop on Text Summarization Branches out. 2004.
[39] Nenkova A, Passonneau R. Evaluating Content Selection in Summarization: The Pyramid Method [C]. In: Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,

2004: 145-152.

Chinese Books Join the ProQuest Platform

ProQuest is providing support for libraries specializing in Asian studies or interested in building Chinese language collections, with resources available for permanent access. Until now, building comprehensive, high-quality Chinese collections has been time-consuming and laborious for overseas librarians. Currently, ProQuest has simplified the book procurement process for librarians and improved the research experience for students, researchers, and staff.

Providing Chinese books to readers represents the beginning of ProQuest' s long-term cooperation project with numerous publishers, with new publishers joining as the project progresses. Currently, ProQuest will begin cooperation with the following publishers:

1. **Peking University Press**–A large comprehensive national-level publisher established in 1902, China' s first university press. Its publications cover humanities, social and natural sciences, information science, and other fields.
2. **China Intercontinental Press**–Publishes art, culture, literature, philosophy, religion, and travel titles.
3. **Economic Science Press**–Affiliated with the Ministry of Finance, specializing in economics publications.
4. **Hong Kong University Press**–Publishes over 50 Chinese and English books annually.
5. **Huazhong University of Science and Technology Press**–Huazhong University of Science and Technology is a national key university under the Ministry of Education.
6. **China Translation & Publishing Corporation**–An award-winning publishing company covering social sciences, literature, art, education, and youth publications.

"ProQuest' s ebook resources have always been committed to making content that readers genuinely need but is difficult to discover more accessible and usable, improving workflows for librarians and researchers. This cooperation is an excellent example," said Kevin Sayar, Senior Vice President and General Manager of Books at ProQuest. "We are pleased to see these new publishing partners join and build a solid foundation upon this."

These authoritative Chinese publications will be available to readers on ProQuest' s Ebook Central, EBL, and Ebrary platforms. The Ebook Central platform, launched in early 2016, supports users through intuitive, user-centered design, enabling them to discover, evaluate, and access books more conveniently and quickly. Users can increase engagement through time-saving features like marking, annotating, and instant citation while reading. Users can read directly online or offline through DRM-free chapter downloads and full book downloads,

and the user interface can be directly translated into Traditional or Simplified Chinese.

(Compiled from: http://www.proquest.com/about/news/2016/Key-Chinese-Language-Titles-on-the-ProQuest-Platform.html)
(Journal News)

Author Contribution Statement:
Liu Tianyi: Proposed the research topic, literature reading and organization, drafted the manuscript;
Bu Yi: Literature reading and organization, drafted the manuscript;
Zhao Danqun, Huang Wenbin: Revised the final version of the manuscript.

Conflict of Interest Statement:
All authors declare no conflict of interest.

Support Data:
Support data is available in the online version of the journal at http://www.infotech.ac.cn.
[1] Liu Tianyi, Bu Yi, Zhao Danqun, Huang Wenbin. Raw Data-Search Results.xlsx.
[2] Liu Tianyi, Bu Yi, Zhao Danqun, Huang Wenbin. Raw Data-Relevant Literature in Search Results.xlsx.

Received: October 21, 2015
Revised: January 3, 2016

Note: Figure translations are in progress. See original paper for figures.