Postprint on Knowledge Gene Expression, Inheritance, and Variation in Scientific Literature

Authors: Bai Rujiang, Zhang Qingzhi, Yigang Sun

Date: 2023-04-01T00:00:00+00:00

Abstract

[目的/意义] The accumulation and inheritance of knowledge propel the development of human society. This study proposes employing knowledge genes from scientific literature to investigate the inheritance and variation of knowledge, aiming to provide a more intuitive and comprehensive insight into knowledge transmission and evolutionary changes. [方法/过程] By dissecting the concept of knowledge genes, this research establishes their significance and explores specific research objects; proposes two approaches for content expression of knowledge genes in scientific literature and categorizes their types; analyzes the primary factors influencing the inheritance and variation of knowledge genes and designs methodologies for identifying these genetic and variational processes. [结果/结论] Through the identification of knowledge genes in scientific literature, this approach can effectively reveal knowledge inheritance and iteration across different documents, thereby fostering the development and refinement of the knowledge gene theoretical framework.

Full Text

A Study on the Expression, Inheritance, and Variation of Knowledge Genes in Scientific Literature

Bai Rujiang[1],[2], **Zhang Qingzhi**[3], **Sun Yigang**[1]
[1] National Library of China, Beijing 100081
[2] Institute of Scientific and Technical Information, Shandong University of Technology, Zibo 255049
[3] Department of Information Management, Peking University, Beijing 100871

Abstract: [Purpose/Significance] The accumulation and inheritance of knowledge drive the development of human society. This paper proposes to study the inheritance and variation of knowledge using the knowledge gene concept in scientific literature, aiming to provide a more intuitive and comprehensive perspective on knowledge inheritance and evolution. [Method/Process] By analyzing the narrow and broad definitions of knowledge genes, this study clarifies their research significance and explores specific research objects. Two methods for expressing knowledge genes in scientific literature are proposed, and the types of knowledge genes are summarized. The main factors affecting the inheritance and variation of knowledge genes are analyzed, and a method for identifying such inheritance and variation is designed. [Result/Conclusion] The identification of knowledge genes in scientific literature can effectively reveal knowledge inheritance and iteration between different documents, promote the development and improvement of the knowledge gene theoretical system, and provide intellectual support for knowledge innovation.

Keywords: meme; knowledge gene; knowledge gene inheritance; knowledge gene variation; text content analysis

1. Definition of Knowledge Genes

1.1 Knowledge Gene Concepts Dawkins first established a connection between genes and knowledge, arguing that there exists a "gene" independent of chemical substances, whose "soup" is human culture, whose transmission vehicle is language, and whose spawning ground is the brain. He gave it a new name: "meme." Memes originate in the brain and establish beachheads on information carriers such as paper, film, and silicon chips. Simple ideas like colors and numbers are analogous to single nucleotides and insufficient to constitute memes. Memes should be complex units like genes, possessing certain durability. After Dawkins proposed the meme concept, it received widespread attention and research both domestically and internationally.

In foreign research, early geneticists used genetic inheritance and variation to conclude that class stratification was detrimental to social equity, causing most sociologists to resist genetic theory. Subsequently, memes were applied to cultural studies. S.K. Sen proposed the concept of information genes based on cultural genes, suggesting that information gradually stabilizes and grows through information gene inheritance and trial-and-error testing. S. Blackmore defined knowledge genes in *The Meme Machine* as replicators with high-fidelity copying, high reproductive capacity, and strong vitality. R. Aunger analyzed memes from a linguistic perspective in *The Electric Meme*, arguing that buzzwords are also memes and emphasizing the explanatory role of memes in popular culture. In these studies, scholars failed to delineate research boundaries, lacked strict definitions, and had difficulty mathematizing their research, making cultural genes too abstract, intangible, and unmeasurable—capable only of serving explanatory functions in cultural evolution theory and mental evolution theory, but unable to support practical applications.

Domestically, Li Bowen considered knowledge genes to be scientific concepts. Liu Zhihui argued this definition was too broad and redefined knowledge genes as the smallest functional units of knowledge evolution, possessing stability, heritability and variability, governability, and directionality, with the goal of understanding knowledge evolution patterns. Liu Zhihui elaborated in detail on the definition, characteristics, classification, scientific laws, applications, and genetic and variation movements of knowledge genes. Sun Xiaoling defined frequently occurring words or phrases in literature as important knowledge genes and calculated knowledge gene strength by combining word frequency with propagation degree in citation networks. Liu Zeyuan viewed knowledge genes as self-organizing knowledge systems in specific domains, capable of demonstrating knowledge generation, evolution and recombination, emergence, fault and transformation, dissemination and application. He Jinsheng et al. studied the role and feedback mechanisms of knowledge genes in enterprise innovation processes. Gu Xinjian et al. used the SAO triple method to extract knowledge genes from patent citation networks and establish knowledge evolution trajectories. Ding Kun et al. used knowledge gene discovery algorithms to identify key knowledge genes in knowledge evolution and conflict. Tan Zongying et al. constructed a research model for topic evolution based on knowledge gene dissociation and recombination to understand disciplinary development patterns and research topic layouts.

The above review reveals that knowledge genes are the basic units of knowledge evolution trajectories. Knowledge genes can be used to discover and mine implicit, unknown, and potentially valuable knowledge, providing intellectual support for knowledge innovation. However, conceptual descriptions of knowledge genes remain vague, using only explanatory language to describe what the knowledge gene theory and identification algorithms can accomplish, without explaining what knowledge genes are or why they can complete these tasks. Although articles are labeled with "knowledge genes," they still use subject terms and keywords to represent knowledge genes in practical applications.

The authors argue that scientific literature knowledge genes are one type of meme meaning expression, representing an organic combination of knowledge objects that characterize literature value in scientific text content—the most basic, active, and widely influential knowledge content in scientific innovation. Knowledge genes consist of original literature abstracts and citation content from citing literature. Since citing literature does not necessarily reference the most prominent innovation points of original literature, the organic combination and recombination of both can yield the knowledge genes of scientific literature.

The knowledge gene generation process for scientific literature should include the following steps: first, extract category labels from the original literature abstract and citation content; second, identify the citation position, citation sentiment, citation function, and citation nature; finally, calculate inheritance and variation results based on these annotations [Figure 1: see original paper].

Based on this knowledge gene generation process, comprehensive citation category annotation results, inheritance and variation judgments, and temporal factors are used to complete the extraction of knowledge gene expression for individual scientific literature, as shown in [Figure 2: see original paper]. The horizontal axis represents specific knowledge gene category labels, the vertical axis represents general knowledge gene category labels, yellow indicates original literature, red indicates variation relationships between citing literature citation content and original literature, and blue indicates inheritance relationships. Numbers within the yellow, red, or blue blocks represent publication years of scientific literature.

---

2. Knowledge Genes and Knowledge Elements

Wu Jun noted in *Insights* that great inventions are always based on previous work, completing the process from "N" to "N+1." Given the rapid development of modern technology, achieving the entire process from 0 to N has become increasingly impossible. The current scientific development model involves taking a step further on existing foundations—standing on the shoulders of giants (N) to produce new technologies or tools (1), achieving qualitative leaps and driving scientific development through incremental progress. Similarly, humans do not verify all previous hypotheses and experiments one by one when acquiring knowledge, but instead directly accept some knowledge as theorems, enabling rapid advancement to the frontiers of current technology to focus on generating new knowledge.

In scientific literature, N can be understood as the references of an original paper, while 1 can be understood as the abstract of the original paper. When a paper is cited by subsequent literature, the original paper becomes "N."Scientific literature continuously evolves and develops in the process from "N" to "N+1," so the knowledge genes of a single paper should consist of the original literature' s abstract and its citing literature' s citation content.

Knowledge elements typically refer to concrete terms extracted from scientific literature content using information extraction, topic modeling, and other technologies. These terms may be subject terms or triples. Many scholars in knowledge element research focus on expanding attribute and relationship descriptions of subject terms, such as using literature sources, versions, comments, access records, and other information to represent knowledge elements, primarily for knowledge representation, control, and processing, but ultimately remaining fragmented knowledge.

When representing knowledge, knowledge genes use both general knowledge gene type labels and specific knowledge gene type labels to prevent getting lost in the sea of knowledge due to fragmentation. Specific knowledge gene type labels vary according to different knowledge domains, making them relatively flexible and unrestricted. In addition to describing original literature knowledge genes, both types of labels also describe citing literature citation content. Due to the computable characteristics of citation content, we can determine whether knowledge undergoes inheritance or variation in its development from original to citing literature. Therefore, while representing original literature knowledge, knowledge genes introduce citing literature content and relationships, using two label types to represent content and inheritance or variation to represent relationships. As citing literature continues to emerge, the content of original literature knowledge genes becomes increasingly rich.

---

3. Composition of Knowledge Genes

The human body has 23 pairs of chromosomes in somatic cells, but only one pair—the 23rd—determines sex. When researchers focus on gender and related information, they concentrate on sex chromosomes. Similarly, in the knowledge domain, a single literature involves multiple knowledge points, such as reviews of various problem-solving methods or implementation tools, but the literature' s outstanding contribution lies in methodological innovation built upon previous research. To accurately describe a single literature' s outstanding contribution, the value of knowledge genes should be determined by the "1" in the knowledge output "N+1."

Based on the common phenomenon that researchers cannot quickly locate needed knowledge, combined with the gene and "N+1" principle, the authors argue that knowledge genes, based on the innovation point "1" in the "N+1" principle, represent an organic combination of valuable knowledge units in original papers that characterize original viewpoints, methods, data, conclusions, etc. Knowledge is fluid, and knowledge genes should consist of the author' s subjective and static value expression of literature innovation at publication time and the objective and dynamic value in citing literature. Static expression is reflected in the literature' s abstract, which is fixed at publication time; dynamic objective description is reflected in subsequent citing literature' s citation content, objectively describing the value in the citing author' s research field.

3.1 Static Expression of Knowledge Genes When a paper is published, the value of scientific literature has been recognized by the author, editors, and peer reviewers, constituting its static description. In the abstract, the author provides a refined expression of the literature's subjective value without commentary or supplementary explanation, maximizing the brief, concise, and coherent statement of important information from the single paper and concentrating the essence of the original text. The abstract reflects the value representation of a single paper from the author' s perspective. Through fine-grained, structured, and semantic expression of the abstract, knowledge and its interrelationships can be accurately described. Therefore, static expression of knowledge genes should be extracted from literature abstracts.

3.2 Dynamic Expression of Knowledge Genes Citing literature authors elevate the objective value of cited literature through citation. As subsequent researchers cite and evaluate the literature, it constitutes the literature’ s subsequent dynamic value. Researchers’ research directions and perspectives on articles determine the literature’ s contribution to citing authors, forming the literature’ s subsequent objective description.

Citation content represents citing literature’ s objective judgment of cited literature’ s knowledge innovation points based on current development status, reflecting the incremental nature of scientific knowledge. Compared with cited literature abstracts and full texts, citation content can provide more objective and rich semantic information while representing the main contributions of cited literature to subsequent research, expressing citing authors’ recognition of original literature value. Therefore, as traces of scientific communication and exchange, citation content should be used for dynamic expression of knowledge genes.

The functional performance of static abstracts and dynamic citation content is shown in .

---

4. Categories of Knowledge Genes

Connections between knowledge form a complex knowledge network. Locating a knowledge point in this network requires both horizontal and vertical coordinates. This paper proposes using general knowledge genes as the vertical coordinate and specific knowledge genes as the horizontal coordinate for knowledge gene visualization.

General knowledge genes are universal knowledge gene objects across research fields, such as research methods, purposes, theories, tools, and data. Specific knowledge genes are domain-specific knowledge gene objects that vary by field. Therefore, designing specific knowledge gene objects requires professional expertise, selecting research objects based on domain characteristics, or designing research objects according to the domain knowledge base after understanding the research field. Thesauri and ontologies, as common understandings of domain knowledge, formally describe domain entity concepts, relationships, characteristics, and laws, and have been widely applied in earth science, energy and transportation, geology, meteorology, and biomedicine, providing references for specific knowledge gene object design.

Using Alzheimer’ s disease research as an example, specific knowledge gene objects include disease symptoms, detection methods, treatment options, etc.

4.1 General Knowledge Gene Categories Researchers including M.J. Moravcsik, M. Garzone, I. Spiegel-Rösing, C. Oppenheim et al., R. Radoulou, Lu Wei, and Qin Chunxiu have studied general annotation objects for scientific text content, as shown in . These annotation objects can characterize general knowledge genes in scientific literature.

These general annotation frameworks, in practical application, require overly high discrimination ability from annotators if the distinctions are too detailed, but lose the significance of fine-grained knowledge organization if too general. Therefore, combining previous research, the authors merged annotation objects with low discrimination and selected the following as general knowledge gene expression objects: concept (including explanation, definition, connotation, principle, theory), problem, data, material, equipment, tool, method (including scheme, steps), algorithm (including equations, formulas), model (including systems), application, and result.

4.2 Specific Knowledge Gene Categories Different research fields have different domain-specific knowledge genes. In the medical field, each disease has its causal relationships. When pathological changes occur in cells, tissues, and organs, or when biochemical reactions malfunction, these manifest as patient symptoms and abnormal physical examination results. Therefore, internal abnormalities can explain clinically observed phenomena, and external manifestations can serve as clues to seek pathogenic causes during disease diagnosis.

The causal relationship of Alzheimer' s disease remains a mystery. The combination of clinical manifestations and pathological changes forms the basis for pathophysiology to explore causes and effects. Based on basic knowledge classification introductions of Alzheimer' s disease from medlineplus.gov and nia.nih.gov, combined with electronic medical record object research, the authors designed specific research objects for the Alzheimer' s disease research field. These objects can serve as type expressions for domain-specific knowledge genes, as shown in .

---

5. Inheritance and Variation of Knowledge Genes

J. Monod believed that, like organisms, knowledge maintains its structure and reproduces through fusion, recombination, and separation. The "transmissibility" or "infectivity" of knowledge promotes dissemination through interaction. Elimination also plays an important role in evolution, accelerating social progress through elimination mechanisms.

Knowledge gene inheritance manifests as intergenerational knowledge inheritance, forming fixed values under inertial effects and maintaining the stability of human knowledge. Rapidly and dynamically developing society impacts the solidification of knowledge gene inheritance. Knowledge adapts to new environments through continuous variation—that is, knowledge innovation. Knowledge innovation causes original knowledge to no longer evolve along its original path but instead produces knowledge gene variation. The inheritance and variation of knowledge genes are most evident in interdisciplinary fields, where adjacent disciplines cross-penetrate, mutually absorb, and organically fuse theories. Various knowledge forms collide, producing inheritance and variation that generate new momentum for academic research and nurturing major scientific achievements.

5.1 Analysis of Factors Influencing Knowledge Gene Inheritance and Variation Genes undergo inheritance or variation through natural selection, while knowledge genes are the result of artificial selection during their inheritance or variation. Genes inherit or vary from parents through offspring characteristics and traits, while knowledge genes demonstrate inheritance or variation of original literature through citation content. This artificial selection in scientific literature manifests as basic citation properties: citation position, citation sentiment, citation function, and citation nature.

Researchers interpret original literature based on their knowledge background, research direction, citation motivation, citation purpose, and citation sentiment. Just as “there are a thousand Hamlets in a thousand readers’ eyes,” researchers’ subjective factors such as research themes and emotions toward cited literature make the value of literature in subsequent citing literature non-equivalent. In citation content analysis, multiple factors must be considered: whether the author’ s attitude toward cited literature is positive, neutral, or negative; whether the citation position indicates citation importance; whether the citation references concepts or practical applications. These influencing factors vary in each literature and should be incorporated into the scope of investigation for citation content analysis modeling and calculation.

The main factors include citation position judgment, citation sentiment calculation, citation function, citation nature, and citation category. These content analyses provide the basis for knowledge gene inheritance and variation discrimination. The analysis of influencing factors of citation content on inheritance and variation is shown in , and relationships between different influencing factors are shown in [Figure 4: see original paper].

Specific analysis of different influencing factors:

(1) Citation Position Factor. H. Voos believed that combining citation position and citation function could calculate the actual value of citations. S. Teufel empirically found that 62.7% of references made no substantial contribution to the original text, with only 18.9% providing value output for terminology definitions or tool usage. B.A. Lipeetz empirically studied the contribution of citation positions, arguing that citations related to background introduction have smaller scientific contributions, while citations in literature reviews that identify existing problems have greater contributions and certain citation significance. G. Herlach merged introduction and literature review sections and merged methodology and results discussion sections, finding through analysis that methodology and results discussion sections contribute more to citing literature. X. Zhu used supervised machine learning methods with automatic feature selection to build an academic impact prediction model, where position features played an important role in experiments.

Based on the above research, the authors believe that citations in article introduction sections have smaller scientific contributions to cited literature and require reduced variation calculation weight and increased inheritance calculation weight in knowledge gene inheritance and variation calculations. Citations in methodology sections require reduced inheritance calculation weight and increased variation calculation weight.

(2) Citation Intensity Factor. M.J. Moravcsik empirically studied redundant citations, arguing that consecutive citations are mostly references to classic theories and methods belonging to redundant citations. After analyzing 575 references from 30 papers, he found one-third were redundant citations. Chang Simin identified consecutive citations as patchwork redundant citations, pointing out they provide no 铺垫 for subsequent narratives. M.H. MacRoberts divided citations into influential and non-influential citations, with results showing fewer influential citations.

This paper argues that word citations and consecutive citations with continuity greater than 3 should have reduced weight in inheritance and variation calculations.

(3) Citation Sentiment Factor. D.E. Chubin divided citations into positive and negative citations through manual reading, with results showing 95% were positive citations. M.J. Moravcsik divided citations into positive, neutral, and negative citations through question-based manual 界定, with results showing 84% were positive citations. Subsequently, Moravcsik limited selection to positive and negative citations, with positive citations reaching 92%. Domestically, Liu Shengbo divided citation content into positive, neutral, and negative, conducting empirical research on BMC Bioinformatics journal, with results showing 62.88% were neutral citations and 3.53% were negative citations. These studies show that foreign authors consistently judge most citations as positive, while domestic authors have higher standards for positive citation judgment, categorizing most as neutral. Both domestic and foreign authors agree on the small number of negative citations. Although negatively cited literature is referenced, its main ideas are not inherited and contradict the original literature' s intended meaning.

This paper argues that positive citation types should increase inheritance calculation weight, neutral citations should reduce weight in inheritance and variation calculations, and negative citations should increase variation calculation weight.

(4) Citation Function Factor. S.U. Hassan divided citation functions into four types based on preliminary reading and syntactic structure features: related research, comparative research, existing research, and extended research, summarizing cue words for each type and using regular expressions to classify citations for national and institutional evaluation.

This paper argues that related research and existing research citation functions should be used to design inheritance calculation weight, while extended research and comparative research should be used to design variation calculation weight.

5.2 Identification and Calculation of Knowledge Gene Inheritance and Variation in Scientific Literature Based on the previously analyzed knowledge gene types and factors affecting inheritance and variation, this paper designs a calculation implementation method for knowledge gene inheritance and variation, as shown in [Figure 5: see original paper].

After completing the calculation steps in [Figure 5: see original paper], comprehensive inheritance and variation judgments are made based on calculation results. If citation categories are consistent, it is judged as knowledge gene inheritance; if inconsistent, category correlation analysis is conducted. If correlation is weak, it is judged as knowledge gene variation; if correlation is strong, those below the weight threshold are judged as variation and those above as inheritance. Weight and threshold settings are determined based on literature characteristics in the scientific field.

---

6. Narrow and Broad Knowledge Genes

(1) Narrow Knowledge Genes The authors define knowledge genes based on single literature as narrow knowledge genes, refining and organizing literature value at the micro level through fine-grained structured knowledge gene extraction to improve literature knowledge discovery efficiency. Additionally, scientific literature knowledge genes can be used for single literature knowledge evaluation, with subsequent citing literature’ s inheritance or variation representing the paper’ s content output capability.

(2) Broad Knowledge Genes The above knowledge gene definition and application are based on single scientific literature, representing a narrow definition. Scientific literature carriers facilitate knowledge gene quantification. When inheritance and variation calculations are performed on large amounts of scientific text, the development and change trajectories of a discipline field become clearer, giving rise to broad knowledge genes that can describe macro-level knowledge evolution trajectories.

Broad knowledge genes can control the direction of knowledge development, representing the common knowledge gene embodiment of literature in a discipline’ s research direction, as Foucault stated: “An invisible hand controls the knowledge system through cognitive patterns.” Some literature shares similar knowledge genes at the macro level and connects tightly through knowledge gene nodes, forming its research community.

References

[1] MARTIN W B. *Atoms, bytes and genes public resistance and technoscientific responses*. New York: Routledge, 2015: 1.

[2] MANDELBROT B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*, 1967, 156(3775): 636-638.
[3] SIDDHARTHA M. *The gene: an intimate history*. New York: Scribner, 2017: 121.
[4] Dawkins. *The selfish gene*. Translated by Lu Yunzhong. Changchun: Jilin People' s Publishing House, 1998: 192.
[5] JAMES G. *The information: a history, a theory, a flood*. New York: Books, 2012: 307-320.
[6] HERRNSTEIN R, MURRAY C. *The bell curve: intelligence and class structure in American life*. *Transforming Anthropology*, 2010, 6(1/2): 87-89.
[7] SEN S K. A note on the idea gene and its relevance to information science. *ALIS*, 1981, 28(1/4): 97-102.
[8] BLACKMORE S. *The meme machine*. Oxford: Oxford Paperbacks, 2000.
[9] AUNGER R. *The electric meme: a new theory of how we think*. New York: Simon and Schuster, 2002.
[10] DALTON C, JASON F. *The genome factor: what the social genomics revolution reveals about ourselves, our history, and the future*. Princeton: Princeton University Press, 2017: 4, 284.
[11] Li Bowen. On scientific "inheritance" and "variation" . *Science of Science and Management of S&T*, 1985(10): 21-25.
[12] Liu Zhihui. New developments in knowledge gene theory. *Information Science*, 2003(12): 1243-1245.
[13] Liu Zhihui. The origin, basic content, and development of knowledge gene theory. *Information Studies: Theory & Application*, 1998(2): 8-13.
[14] Liu Zhihui. Knowledge gene exploration (I). *Information Studies: Theory & Application*, 1998(1): 63-66.
[15] SUN X L, DING K. Identifying and tracking scientific and technological knowledge memes from citation networks of publications and patents. *Scientometrics*, 2018, 116(3): 1735-1748.
[16] Liu Zeyuan. Foreword to "Emerging Research Field Identification Metrics" from the perspective of knowledge gene theory. Beijing: Science Press, 2017: i-vi.
[17] He Jinsheng, Lv Wenjuan. The origin, content, and development of knowledge gene theory. *Studies in Science of Science*, 2011, 29(10): 1454-1459.
[18] Xu Qi, Gu Xinjian. A knowledge gene extraction method based on Subject-Action-Object triples. *Journal of Zhejiang University (Engineering Science)*, 2013, 47(3): 385-493.
[19] Sun Xiaoling, Ding Kun. Research on the relationship between science and technology based on knowledge gene discovery. *Information Studies: Theory & Application*, 2017, 40(6): 23-26, 17.
[20] Lu Wanhai, Tan Zongying. Research on domain topic evolution based on knowledge gene dissociation and recombination. *Information Studies: Theory & Application*, 2019, 42(2): 101-107.
[21] Wu Jun. *Insights: the nature of business and the wisdom of life*. Beijing: CITIC Press, 2017: 42-46.
[22] Suo Chuanjun, Gai Shuangshuang. Research on the connotation, structure,

and description model of knowledge elements. *Journal of Library Science in China*, 2018, 44(4): 54-72.
[23] Rice W R. Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, 1984, 38(4): 735-742.
[24] Fu Rongxian. On the document views of ancient 提要 and modern abstracts. *Library and Information Service*, 2016, 60(6): 26-31.
[25] Zhu Qingsong, Leng Fuhai. Review of citation content analysis methods. *Information and Documentation Services*, 2013(5): 39-43.
[26] Xu Lei, Pan Ying. Comparative analysis of knowledge networks and related concepts. *Information Science*, 2017, 35(12): 10-15.
[27] RASKIN R. Enabling semantic interoperability for earth system science. *AGU Fall Meeting Abstracts.* New York: American Geophysical Union, 2004: 11-16.
[28] Zhang Yunliang, Xu Shuo, Zhu Lijun, et al. Chinese scientific and technological word system—a semantic resource for deep content analysis of scientific and technical information resources. *Library and Information Service*, 2011, 55(4): 100-105.
[29] Ma X, Carranza E J M, Wu C, et al. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps. *Computers & Geosciences*, 2011, 37(10): 1602-1615.
[30] Moinet M P, Valcke S, Lawrence B N, et al. Development and exploitation of a controlled vocabulary in support of climate modelling. *Geoscientific Model Development*, 2014, 7(2): 479-493.
[31] Su Y, Andrews J, Huang H, et al. Reengineering of MeSH thesauri for term selection to optimize literature retrieval and knowledge reconstruction in support of stem cell research. *BMC Medical Informatics and Decision Making*, 2016, 16(1): 54.
[32] MORAVCSIK M J, MURUGESAN P. Some results on the function and quality of citations. *Social Studies of Science*, 1975, 5(1): 86-92.
[33] GARZONE M, MERCER R E. Towards an automated citation classifier. *Lecture Notes in Computer Science*, 2000, 1822: 337-349.
[34] SPIEGEL-RÖSING I. Science studies: bibliometric and content analysis. *Social Studies of Science*, 1977, 7(1): 97-113.
[35] OPPENHEIM C, RENN S P. Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science*, 1978, 29(5): 225-231.
[36] RADOULOV R. *Exploring automatic citation classification.* Waterloo, ON, Canada: University of Waterloo, 2008: 33-37.
[37] Lu Wei, Meng Rui, Liu Xingbang. Research on a citation content annotation framework for citation relationships. *Journal of Library Science in China*, 2014, 40(6): 93-104.
[38] Liu Jie, Qin Chunxiu, Zhao Pengwei, et al. A knowledge element-based content organization method for scientific text resources. *Information Studies: Theory & Application*, 2018, 41(4): 128-133.
[39] Sherwin B N. *How we die: reflections of life' s final chapter, New Edition.* New York: Vintage Books, 1995: 89.

[40] MONOD J. A biologist' s world view (Book Reviews: Chance and Necessity. An essay on the natural philosophy of modern biology). *Science*, 1972, 175(4017): 49-50.
[41] DENNETT D C. *Consciousness explained.* New York: Little, Brown and Company, 1991.
[42] VOOS H, DAGAEV K S. Are all citations equal? Or, did we Op. Cit. your idem? *Journal of Academic Librarianship*, 1976, 1.
[43] TEUFEL S, SIDDIHARTHAN A, DAN T. Automatic classification of citation function. *Proc. 2006 conference on empirical methods in natural language processing.* Stroudsburg: Association for Computational Linguistics, 2006: 103-110.
[44] LIPETZ B A. Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators. *Social Studies of Science*, 1974, 4(2): 265-272.
[45] HERLACH G. Can retrieval of information from citation indexes be simplified? Multiple mention of a reference as characteristic of the link between cited and citing article. *Journal of the Association for Information Science & Technology*, 2014, 29(6): 308-310.
[46] ZHU X, TURNEY P, LEMIRE D, et al. Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science & Technology*, 2015, 66(2): 408-427.
[47] Chang Simin. Analysis of redundant references in scientific papers. *Publishing Science*, 2015, 23(1): 43-45.
[48] MACROBERTS M H, MACROBERTS B R. Quantitative measures of communication in science: a study of the formal level. *Social Studies of Science*, 1986, 16(1): 151-172.
[49] CHUBIN D E, MOITRA S D. Content analysis of references: adjunct or alternative to citation counting? *Social Studies of Science*, 1975, 5(4): 423-441.
[50] Liu Shengbo, Ding Kun, Zhang Chunbo. Research on citation evaluation based on citation content nature. *Information Studies: Theory & Application*, 2015, 38(3): 77-81.
[51] HASSAN S U, SAFDER I, AKRAM A, et al. A novel machine-learning approach to measuring scientific knowledge flows using citation context analysis. *Scientometrics*, 2018, 116(4): 1-24.
[52] Foucault. *The archaeology of knowledge.* Translated by Xie Qiang, Ma Yue. Beijing: SDX Joint Publishing Company, 1998: 202-215.

Author Contributions:
Bai Rujiang: Responsible for proposing the paper framework and designing research 思路;
Zhang Qingzhi: Responsible for collecting materials and writing the paper;
Sun Yigang: Responsible for research 思路 review and finalizing the paper.

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

Source: ChinaXiv —Machine translation. Verify with original.