Postprint: A Study on the Disciplinary Positioning of Individual Papers from the Perspective of Academic Evaluation

Authors: Fang Shengyu, Xi Yu

Date: 2023-11-07T00:00:00+00:00

Abstract

[Purpose/Significance] This study examines the evolution of disciplinary classifications, analyzes and compares the thematic positioning of individual documents by three systems—CWTS, Citation Topics, and Dimensions, and provides a reference for disciplinary positioning based on research themes of single papers in academic evaluation. [Method/Process] We analyze and compare the principles and construction methods of the three systems, using research papers from a university's school of psychology as the case study, conduct a comparative analysis of their disciplinary topic classifications across the three systems, and analyze the disciplinary themes of the papers in combination with their disciplinary classifications in WOS. [Results/Conclusion] The three classification systems each have their own advantages and limitations: CWTS offers finer-grained thematic classification and provides dynamic mapping; Citation Topics facilitates the export and further analysis of thematic literature; Dimensions provides more comprehensive and extensive coverage of literature sources and document types. CWTS performs better than Citation Topics in thematic description; Citation Topics does not necessarily provide accurate positioning of papers' disciplinary themes; and Dimensions' thematic positioning is relatively broader in granularity. The disciplinary positioning of a paper's research theme requires combining multiple disciplinary classification systems and making comprehensive judgments from the paper's content, references, and citing literature.

Full Text

Preamble

Research on the Discipline Attributes of Individual Papers from the Perspective of Academic Evaluation

Fang Shengyu[1], *Yu Xi*[2]

[1] School of Information and Intelligence Engineering, Tianjin Renai College, Tianjin 301636, China
[2] Library of Tianjin Normal University, Tianjin 300387, China

Abstract:
[Purpose/Significance] This study examines the development and evolution of disciplinary classification, analyzing and comparing how the CWTS, Citation Topics, and Dimensions systems thematically position individual documents. It provides a reference for disciplinary positioning based on research topics of single papers in academic evaluation. [Method/Process] We analyze and compare the principles and construction methods of the three systems, using research papers from a university's psychology school as a case study to conduct comparative analysis of their disciplinary classifications across the three systems. Combined with the disciplinary classification of these papers in WOS, we analyze their subject themes. [Result/Conclusion] Each of the three classification systems has distinct advantages and limitations. CWTS offers finer-grained topic classification with dynamic mapping capabilities. Citation Topics facilitates the export and further analysis of topical literature. Dimensions provides comprehensive and extensive coverage of literature sources and types. CWTS outperforms Citation Topics in topic description, while Citation Topics does not always accurately position papers' disciplinary themes. Dimensions' topic positioning is relatively broad. Determining the disciplinary positioning of a paper's research topic requires integrating multiple disciplinary classification systems and making comprehensive judgments based on the paper's content, references, and citing literature.

Keywords: Discipline attributes; Discipline classification; Topic orientation; CWTS; Citation Topics; Dimensions
Classification Number: G254.2
DOI: 10.31193/SSAP.J.ISSN.2096-6695.2023.03.02

1. Origin and Evolution of Disciplines

The exploration of disciplines began with systematic human production activities. Disciplinary research can be traced back to ancient Greece around 400 BCE, when both Plato and Aristotle expounded upon and preliminarily divided disciplines. Plato believed that all disciplines possessed artistic characteristics, dividing them into "acquisitive arts" and "productive arts." Acquisitive arts produce nothing but conquer through words and actions, or prevent others from conquering existing things, such as mathematics, political science, and dialectics. Productive arts bring into existence what did not exist before, such as agriculture and medicine [1]. Aristotle considered disciplines as the domain of science, encompassing theoretical knowledge, practical knowledge, and productive knowledge. Theoretical knowledge included philosophy, mathematics, and natural sciences; practical knowledge referred to ethics and political science; and

productive knowledge referred to arts that could imitate nature or accomplish what nature could not [2]. Marcus Tullius Cicero of ancient Rome proposed that disciplines included both science and art, arguing that philosophy and oratory were sciences, while most natural sciences belonged to art [3]. Francis Bacon of the Renaissance period suggested that human cognition could be divided into three major categories: history, poetry, and philosophy [4]. In modern times, the Industrial Revolution drove technological progress, and as human cognition developed, disciplinary divisions continuously evolved. Representative scholars of this period include Jean Piaget, who argued that disciplines underwent four evolutionary stages: first logic and mathematics, second physics, third biology, and fourth psychology and physiology, with these stages mutually reinforcing one another [5].

Since the advent of the internet era, online publishing has generated vast quantities of scientific literature, prompting attempts to organize and classify documents by discipline. In 1957, Eugene Garfield of the Institute for Scientific Information (ISI) founded the Science Citation Index in Philadelphia, initiating the scientific organization of literature. Subsequent development of universities worldwide and the establishment of major database providers further promoted disciplinary development and refinement, creating diverse disciplinary classification systems. These include single-discipline systems such as Chemical Abstracts for chemistry, Econlit for economics, and MeSH for medicine; and multi-disciplinary systems such as WOS subject categories, ESI subject categories, and GIPP subject categories from Clarivate Analytics, and Scopus and EI subject categories from Elsevier.

Countries have also developed disciplinary classification systems suited to their own characteristics and development needs. Different literature types have different classification systems for books, patents, and grants. Third-party academic evaluation institutions such as QS, THE, U.S. News, and ARWU have established their own disciplinary classification systems.

2. Discipline Classification in Academic Evaluation

Most disciplinary classification systems developed by database providers use journals as the basic unit. Different databases have different classification systems, and the same journal may have different disciplinary positions across databases. Taking the journal *Spectroscopy and Spectral Analysis* as an example—this journal, sponsored by the Chinese Optical Society and co-organized by the Central Iron and Steel Research Institute, Institute of Physics of the Chinese Academy of Sciences, Peking University, and Tsinghua University, is indexed in SCI, EI, and the Chinese Core Journals—it corresponds to different disciplinary classifications across databases. In ESI, it belongs to “Chemistry”; in WOS, it belongs to “Spectroscopy”; and in Scopus, EI, and CNKI, it belongs to multiple disciplinary categories (Table 1).

University disciplinary divisions and school setups are also inconsistent with

database classifications, further complicating the accurate disciplinary positioning of academic achievements. School establishment relates to national development and institutional characteristics, while database classifications aim to facilitate literature management, with no necessary relationship between the two. When conducting statistical analysis of research outputs, we frequently find that contributions to a particular discipline come from multiple schools. As shown in Table 2 , a university's research output in ESI Materials Science TOP 1% comes from contributions by schools of physics, chemistry, mathematics, and others. Additionally, due to interdisciplinary integration, academic achievements often involve multiple fields, making single-discipline classification difficult. In the Ministry of Education's disciplinary evaluation, the disciplinary definition of academic achievements has been de-emphasized, relying instead on self-reporting by universities, while ignoring the consistency between researchers' disciplinary backgrounds and the disciplinary positioning of their achievements.

The complexity and confusion of disciplinary classification systems seriously affect the accurate positioning of academic achievements and subsequent academic evaluation, thereby influencing discipline construction, talent recruitment, and strategic discipline layout. Assigning scientific literature to appropriate disciplinary fields is a fundamental prerequisite for effective scientometric analysis. Under the national policy of "breaking the five-only" and promoting representative work evaluation systems, the issue of how to achieve disciplinary positioning of papers has become urgent. Consequently, attempts have been made to transition the basic unit of disciplinary positioning from journals to individual papers, and from belonging disciplines to research topics, classifying papers through topic division.

3. Research Topic Positioning of Individual Papers

Recent research on scientific paper topic positioning mainly falls into three categories: Citation-based topic positioning methods, including direct citation, co-citation, indirect citation, and bibliographic coupling, which cluster literature according to algorithms to divide research topics. Klavans et al. conducted literature clustering based on direct citation, bibliographic coupling, and co-citation, comparing the accuracy of topic classification across the three methods [6]; Ahlgren et al. conducted a comparative study of clustering relevance detection for PubMed publications [7]; Wei Ruibin obtained research topics through main path analysis and self-citation networks [8]. Text-based topic positioning methods, which extract information from paper titles and abstracts and use machine learning techniques to automatically classify papers into disciplines, such as through NLP [9-10], LDA [11-13], SAO [14], and machine learning models [15] for research topic identification, frontier prediction, and evolutionary trend analysis. Xia Lei identified interdisciplinary topics through AT topic models and similarity calculations [16]; Yang Jing et al. used the Keygraph algorithm to extract keywords reflecting research topics from papers to position their re-

search themes and evaluate paper innovativeness through comparative analysis [17]. Manual topic positioning methods, where authors provide disciplinary affiliation, such as the Chinese Library Classification numbers provided by authors in Chinese literature. Some scholars have compared the accuracy of the three classification methods for paper topic positioning, such as Zhang et al., who compared Web of Science subject categories based on citation classification, Dimensions' Fields of Research (FoR) classification based on machine learning, and Springer Nature subject classification based on author selection for papers published in *Nature* [18]; Wang Xin conducted a comparative study between the InCites Citation Topics system and journal classification systems [19]; Geng Haiying et al. reviewed algorithmically constructed paper-level disciplinary classification systems [20]. Existing topic positioning systems based on individual papers include the CWTS fields of science from Leiden Ranking (hereinafter "CWTS"), Incites Citation Topics from Clarivate Analytics (hereinafter "Citation Topics"), and the Dimensions system from Digital Science.

3.1 Introduction to CWTS, Citation Topics, and Dimensions Systems

CWTS is developed by the Centre for Science and Technology Studies at Leiden University based on citation relationships between papers, using the Leiden algorithm to divide papers into 5 main fields and 4,159 micro-fields [21]. The 5 main fields include: Biomedical and Health Sciences; Life and Earth Sciences; Mathematics and Computer Science; Physical Sciences and Engineering; Social Sciences and Humanities. For each of the 4,159 micro-fields, CWTS provides: a field identifier; the number of publications in that micro-field; the main field(s) to which it belongs; the top 5 journals with the most publications; and 5 characteristic terms extracted from publication titles (Table 3).

Citation Topics is a disciplinary classification method launched by Clarivate Analytics in the Incites database in December 2020, with dataset updates in April 2023. Citation Topics is also a paper-level classification system that constructs a three-tier classification system of macro, meso, and micro topics, including 10 macro topics, 326 meso topics, and 2,437 micro topics. Each topic is labeled with a permanent numeric prefix and a characteristic term to identify the precise topic [22]. For example, micro topic 1.5.77 Deep Brain Stimulation is a subtopic of meso topic 1.5 Neuroscience and macro topic 1 Clinical & Life Sciences.

Dimensions is a new scientific data retrieval and management platform launched by Digital Science in January 2018 [23]. The platform includes various academic resources such as journals, books, grants, patents, clinical trials, and policy documents. Using machine learning and cloud computing technology, it integrates and transforms data to create consistent models. Through deep indexing of full-text literature, including funders, research institutions, researchers, or category status, it assigns appropriate disciplines or research fields to each document regardless of source. Dimensions adopts the Australian and New Zealand Standard Research Classification (ANZSRC) Field of Research (FOR) as its classification standard. FOR is a component of the ANZSRC system used across

all research and education fields in Australia and New Zealand, replacing the previous Australian Standard Research Classification (ASRC). Initially released in March 2008, the classification was reviewed in 2019, with a new system released in June 2020 that provides more detailed content and clearer descriptions. FOR has a three-level hierarchical structure: Division, Group, and Field, comprising 23 Divisions, 213 Groups, and 1,967 Fields. For example, Field 300101 Agricultural biotechnology diagnostics (incl. biosensors) belongs to Group 3001 Agricultural biotechnology and Division 30 Agricultural, veterinary and food sciences.

3.2 Principles and Construction Methods of CWTS, Citation Topics, and Dimensions

CWTS, Citation Topics, and Dimensions are all paper-level research topic positioning systems, but their principles differ. CWTS and Citation Topics are based on relationships between scientific literature, specifically direct citation relationships, using algorithms to automatically cluster literature into research topics. By adjusting parameters, these topics are aggregated into larger clusters to produce disciplinary fields, which are then labeled to construct a hierarchical classification system [20]. This is a bottom-up, self-organizing disciplinary classification system built from literature that simultaneously achieves disciplinary positioning of documents, rather than relying on existing classification systems. Dimensions is based on existing disciplinary classification systems, using machine learning methods to analyze textual content features and automatically classify documents into disciplinary categories. This is a top-down method for disciplinary topic positioning at the paper level. The specific construction methods are as follows:

CWTS conducts large-scale analysis of citation relationships among hundreds of millions of publications, algorithmically constructing 5 main fields and over 4,000 micro-fields. The latest version (2022) assigns each publication in Web of Science from 2000-2021 (limited to Articles and Reviews) to one of 4,159 fields, determining the overlap between each of the 4,159 micro-fields and each of the 254 journal subject categories defined in Web of Science (excluding "Multidisciplinary Sciences"). Each Web of Science subject category is associated with one of the 5 main fields. Based on the connections between subject categories and main fields, each of the 4,159 micro-fields is assigned to one or more main fields. If at least 25% of publications in a micro-field belong to a particular main field, the micro-field is assigned to that main field. Publications belonging to two main fields are assigned to both with a weight of 0.5. Consequently, every paper in Web of Science is assigned to a micro-field, and each micro-field is assigned to one or more main fields [21].

Citation Topics clusters all documents in the Web of Science Core Collection from 1980 to the present based on direct citation relationships using the Leiden algorithm, which includes forced clustering and minimum cluster size parameters. Citation relationships for publications before 1980 are also considered

in clustering. Each document is assigned to only one topic, though not all documents have been successfully assigned. Approximately 75% of documents from 1980 onward have been assigned to a topic, with over 90% of Articles and Reviews included in clustering calculations. Publications without citations after publication do not participate in clustering and are not assigned to topics. Citation Topics' macro and meso topics are manually labeled by ISI based on their content. Micro topics are labeled with the most important keywords algorithmically derived. Since citation topics are obtained based on citation relationships rather than the main research content of the constituent documents, the provided topic terms may not reflect every document in the topic. New documents are added to existing topics based on their cited references, with data updated monthly. A complete clustering update is performed annually, during which individual documents may move between micro-topics and entirely new micro-topics may emerge, while some micro-topics may change their parent meso-topics [22].

Dimensions uses existing classification systems and machine learning-based methods to automatically assign relevant disciplinary categories to all documents. Classification algorithms are trained using established research classification systems from associated datasets. Using machine learning-based reverse engineering technology, it examines manually coded authorized corpora, with computer algorithms generating FOR codes that are then checked against actual codes and iteratively refined. In Dimensions, classification is simulated to the second level (Group level), meaning Dimensions provides paper-level disciplinary positioning at the Group level. In addition to FOR codes, other classification systems are implemented, primarily driven by research funder requirements, such as the NIH's Research, Condition, and Disease Categorization (RCDC) and Health Research Classification System (HRCS) provided on the platform. To implement these schemes, an appropriate machine learning method has been developed that can similarly generate any other classification system and can classify documents not in Dimensions [24].

Domestic research on CWTS mainly focuses on comparative analysis between universities in the Leiden Ranking [25-26]. Research on Citation Topics primarily uses the system to develop journal topic selection plans [27-28]. Studies on Dimensions stem from the integration of A, B, and C category journals funded by the "Excellence Program" into the Dimensions platform in July 2020, enabling scholars to analyze the influence and article characteristics of these journals [29-30]. Currently, relatively few studies examine paper topic positioning using CWTS, Citation Topics, and Dimensions. This paper conducts a comparative study of these three topic classification systems, analyzing their characteristics in positioning papers' disciplinary themes to provide reference for individual paper disciplinary theme positioning analysis.

4. Comparative Analysis of Paper Topic Positioning Across the Three Systems

To determine which system most accurately positions paper topics, this study uses research papers from a university's psychology school as a case study, identifying their disciplinary themes in Citation Topics, CWTS, and Dimensions for comparative analysis.

4.1 Data Sources and Acquisition

This study retrieved the university's psychology school paper data from 2011-2021 in the Web of Science Core Collection's SSCI and SCIE databases (2022 papers were excluded as they fall outside the latest CWTS statistical range). The retrieval was conducted in February 2023, limited to Articles and Reviews. Due to university mergers and splits, institutional names continuously change, coupled with non-standard or even erroneous writing forms that seriously affect accurate and comprehensive retrieval of institutional research output. Web of Science provides three main search field tags for institutional retrieval: AD, OG, and OO. AD searches the institutional address field, which includes institution, school, street, city, and country. OG searches the institution's preferred organization name (official name), retrieving all records under different historical spelling forms. OO only retrieves records with the exact spelling form entered, missing other spelling variations. To ensure comprehensive and accurate retrieval, the OG field was used to retrieve all research output of the university as a primary institution. The retrieved data were imported into Derwent Data Analyzer (DDA) to identify different writing forms of the secondary institution (psychology school), yielding 269 publications. Since all 269 papers originated from the psychology school, their research topics should fall within the scope of psychology. We then identified these papers' disciplinary positioning in CWTS, Citation Topics, and Dimensions to analyze whether their research topics belonged to psychology and to further assess the accuracy of topic positioning across the three systems. Citation Topics was updated in April 2023, with no changes to psychology-related meso-topics but slight differences in micro-topics. However, since sample data were downloaded and the paper was written during February-March 2023, the topic positioning in Citation Topics is based on the previous version.

4.2 Psychology Field Division in the Three Systems

In Citation Topics, psychology belongs to the "Social Sciences" macro-topic, including 2 meso-topics (6.24 Psychiatry & Psychology and 6.73 Social Psychology) and 28 micro-topics (Figure 1 [Figure 1: see original paper]). In Dimensions, psychology is classified under Division 52 (52 Psychology), including 6 Groups and 36 Fields (Figure 2 [Figure 2: see original paper]).

CWTS only provides main fields and micro-fields. The main fields are broad and have no direct correspondence with psychology, while micro-fields are too

granular. To identify psychology-related micro-fields, manual identification was required based on the top 5 journals and 5 characteristic terms provided for each micro-field. Using the English equivalents of the Chinese term for psychology (心理)—psychology, mentality, mind, psychic, psycho—we searched and judged each micro-field's journal names and characteristic terms. However, we found that micro-fields containing these terms were not necessarily psychology-related. For example, micro-field 1651 contains "psychic distance," but based on its top journals and other characteristic terms, we determined this micro-field concerns "international business" research and should be excluded. Ultimately, 79 micro-fields were identified as belonging to the psychology research scope.

4.3 Paper Topic Positioning in the Three Systems

Among the Web of Science papers published by the psychology school, 267 successfully matched to 22 meso-topics in Citation Topics, with only 2 papers from 2020 failing to match; 222 successfully matched to main fields and micro-fields in CWTS, with unmatched papers primarily from 2020 and 2021; 266 papers could be found in Dimensions for research topic positioning, with the 3 unretrieved papers all lacking DOI numbers. A total of 217 papers were successfully positioned in all three systems. Among these, only 28 were positioned as psychology in Citation Topics, 82 in CWTS, and 192 in Dimensions. Merely 16 papers were classified as psychology across all three systems, 69 in two systems, 116 in only one system, and 16 in none of the systems (Figure 3 [Figure 3: see original paper]).

4.3.1 Papers Classified as Psychology in All Three Systems Since Citation Topics and CWTS position papers based on direct citation relationships, we reasoned that papers classified as psychology in these two systems should have references predominantly from psychology. Dimensions uses training sets and machine learning to automatically classify literature based on content features. As the training sets and models are unknown, we cannot manually determine how these papers were positioned as psychology. Therefore, we analyzed only the references of the 16 papers classified as psychology across all three systems. These 16 papers contained 749 references, of which 275 were from journals with titles containing "psycho-," meaning 275 references were clearly from psychology journals. Additional references came from journals containing psychological terms such as "cognition," "anxiety," "emotion," "happiness," and "PTSD." References from journals containing these terms were also considered psychology references. Statistical analysis of the psychology reference percentage for these 16 papers shows that only one paper had a low percentage of 19.05%, while all others exceeded or approached 50% (Table 4).

4.3.2 Papers Classified as Psychology in Two or One System There were 69 papers classified as psychology in two systems. Notably, zero papers were simultaneously classified as psychology in both Citation Topics and CWTS, indicating that if papers belong to psychology in these two systems, they are

also positioned as psychology in Dimensions. Eleven papers were classified as psychology in both Citation Topics and Dimensions, while 58 were classified as psychology in both CWTS and Dimensions. One paper was positioned as psychology only in Citation Topics, eight only in CWTS, and 107 only in Dimensions (Figure 4 [Figure 4: see original paper]).

The inconsistency in topic positioning across the three systems may be attributed to several factors: Citation Topics positioned the fewest papers as psychology, possibly due to its limited number of psychology-related micro-topics (only 1.15% of all micro-topics). Dimensions contains the most psychology-related micro-topics (1.83% of all micro-topics) and consequently positioned the most papers as psychology. CWTS lacks clear disciplinary divisions, providing only characteristic terms and top journals, requiring manual judgment for disciplinary affiliation. Different disciplinary backgrounds, personal cognitive biases, and varying expressions and meanings of the same terms across disciplines inevitably introduce errors. CWTS paper topic positioning requires further validation by peer experts to improve accuracy. Dimensions positioned the most papers as psychology partly because its classification derives from existing disciplinary systems with clear disciplinary assignment. The new version's Group-level coverage of psychology is comprehensive, containing 6 Groups, while the old version contained only 2. Additionally, Dimensions does not assign single disciplines to papers. Among the 192 papers positioned as psychology by Dimensions, 27 included 2 Divisions, meaning these papers were also assigned to other disciplines besides psychology. Some papers with only 1 Division still included multiple Groups. This assignment principle significantly increases the likelihood of papers being positioned as psychology. The interdisciplinary nature of psychology also contributes to inconsistent positioning across systems. Psychology is a highly interdisciplinary field with connections to multiple disciplines. For multidisciplinary and diverse research fields, boundaries between disciplines are often quite blurred, making proper assignment challenging. In our sample, papers positioned as psychology in one or two systems were primarily assigned to psychiatry, neuroscience, and imaging in other systems, with smaller numbers assigned to biomedicine, health sciences, education, economics, and sports science. As disciplines develop and expand, positioning psychology becomes increasingly difficult.

4.3.3 Papers Not Classified as Psychology in Any System Among the 16 papers not classified as psychology in any system, their disciplines were mainly distributed in biosciences, clinical medicine, and neuroscience. Eight of these papers had consistent topic positioning across all three systems. The topics or disciplines of these papers in Citation Topics, CWTS, Dimensions, and WOS are shown in Table 5 .

Papers 1 and 3, though described from different angles across the three systems, essentially belong to "medical physiology" (Paper 1) and "medical nursing" (Paper 3). Paper 2 shows high consistency across all three systems and

WOS, belonging to "library science." Papers 4 and 5 also show high consistency, belonging to "linguistics." Papers 6, 7, and 8 in CWTS describe three specific mental disorders, consistent with the disciplines or topics provided by the other three systems, belonging to "clinical psychiatry."

The sample data originated from a psychology school and should theoretically fall within psychology. However, the sample institution's research has developed toward medicine, likely due to the close connections between psychiatry in medicine and psychology. Some research topics are quite distant from psychology, such as linguistics and library science. Excluding academic misconduct, possible explanations relate to these topics' connections with educational research in linguistics and library science, such as language teaching, second language acquisition, reader education, and reader psychology.

5. Discussion and Conclusion

CWTS, Citation Topics, and Dimensions are all paper-level topic classification systems, but their different principles and construction methods give each distinct characteristics.

CWTS has over 4,000 micro-fields, offering finer-grained classification than Citation Topics' 2,000+ micro-topics. Each micro-field provides 5 characteristic terms for more specific topic description and 5 top journals to help identify core journals for each topic. It provides complete, dynamically changing topic maps showing temporal changes. However, CWTS has limitations: its topic information for Articles and Reviews in the Web of Science Core Collection is time-limited (only papers published 2000-2021); it only provides science mapping and individual paper topic positioning, making it inconvenient to obtain all papers under a topic for further disciplinary analysis; and it lacks search functionality without a dedicated searchable system.

Citation Topics includes all document types in the Web of Science Core Collection from 1980 to present; offers search and download functions for retrieving all papers under a topic; and enables rapid positioning of papers' research topics. However, Citation Topics also has defects: its topic positioning for papers is not always accurate. For example, micro-topic 6.185.1004 Internet Addiction under meso-topic 6.185 Communication, though not classified under psychology meso-topics (6.24 Psychiatry & Psychology, 6.73 Social Psychology), should belong to psychology. Papers from the target institution classified under 6.185.1004 are positioned as psychology in both CWTS and Dimensions. Micro-topics provide only one characteristic term, which is insufficient for scientific and comprehensive topic labeling and cannot reflect the topic's specific content. Different micro-topics may share the same characteristic term but with different emphases (e.g., micro-topics 3.91.1064 and 3.91.172 both had "Heavy Metals" as characteristic terms, but 3.91.1064 concerns heavy metals in water sediments while 3.91.172 concerns heavy metals in soil; after the 2023 Citation Topics update, 3.91.1064's term changed to "Sediments"). Unlike CWTS, Citation Topics does not provide

complete topic maps, preventing intuitive visualization of relationships between disciplinary topics.

The Dimensions platform uses the Australian and New Zealand Standard Research Classification (ANZSRC, 2020) Field of Research (FOR) as its disciplinary classification basis, achieving paper-level classification through machine learning-based methods and training set construction. Dimensions includes extensive data sources such as DOAJ, PubMed, SciELO, and Nature Index journals, as well as A, B, and C category journals funded by the Excellence Program. It covers rich document types encompassing all stages of the research lifecycle, far exceeding CWTS and Citation Topics in literature type coverage, and can position most papers with DOIs to research topics (some without DOIs can also be positioned). Dimensions allows single paper retrieval through full text, title, abstract, and DOI, and enables filtering of literature by research topic. Additionally, it provides two normalized metrics and altmetrics beyond traditional citation indicators for in-depth topic literature analysis. However, Dimensions has limitations. First, some small disciplinary fields are not represented in Dimensions' FOR classification, leaving some small-discipline papers without disciplinary assignment [31]. Second, Dimensions' classification depends on machine learning methods reliant on established classification systems and training set size/quality. Training set structure and granularity limit classification accuracy, and varying training set sizes and qualities result in incomplete accuracy [32]. Third, Dimensions' positioning does not reach the more detailed sub-topic level (Field level), making its disciplinary positioning insufficiently granular [32]. Finally, while Dimensions has search and download functions, data can only be exported to Google Sheets and requires machine query statements, posing difficulties for beginners [33].

The three systems each have advantages and disadvantages in paper topic positioning. Combined use can better help analyze and position paper topics. If topic positioning is consistent across all three systems, the paper's disciplinary affiliation can be determined. When inconsistencies occur, multiple disciplinary classification systems should be combined for comprehensive judgment based on article content and citation relationships. Content-based positioning can generate cluster labels through similarity clustering analysis of titles and abstracts or extract topic terms from full text using NLP techniques to position article disciplines (topics). Citation-based positioning analyzes the disciplinary affiliation of references and citing literature to position article disciplines (topics). However, when topic positioning is inconsistent across systems, it is difficult to determine which system is absolutely accurate. With increasingly common interdisciplinary and convergent research, some papers involve multiple disciplinary fields with certain relationships. The idea of dividing disciplines into many clearly bounded fields is idealistic, and a completely satisfactory solution does not yet exist. Nevertheless, when topic positioning in all three systems is inconsistent with the corresponding discipline of the institution, this should alert research administrators and information analysts.

References

[1] John M. Cooper. *Plato: Complete Works* [M]. Indianapolis: Hackett Publishing Company, 1997.
[2] Jonathan Barnes. *The Complete Works of Aristotle: The Revised Oxford Translation* [M]. Princeton: Princeton University Press, 2014.
[3] Cicero. *On the Orator: Book 3. On Fate. Stoic Paradoxes: Divisions of Oratory* [M]. Cambridge: Harvard University Press, 1942.
[4] Francis Bacon. *Of the Proficience and Advancement of Learning* [M]. Cambridge: Da Capo Press, 1970.
[5] Jean Piaget. Classification des disciplines et connexions interdisciplinaires [J]. *Revue internationale des sciences sociales*, 1964, 16(4): 598-616.
[6] Klavans R, Boyack K W. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? [J]. *Journal of the Association for Information Science and Technology*, 2017, 68(4): 984-998.
[7] Ahlgren P, Chen Y, Colliander C, et al. Enhancing direct citations: A comparison of relatedness measures for community detection in a large set of PubMed publications [J]. *Quantitative Science Studies*, 2020, 1(2): 714-729.
[8] Wei Ruibin. Empirical Research on Paper Topic Innovation Based on Self-citation Networks and Main Path Analysis [J]. *Library and Information Service*, 2018, 62(3): [page numbers].
[9] Jiang Yanting, Hu Renfen. Investigating the Impact of Natural Language Processing in Other Disciplines: Chinese Literature Mining Based on CNKI [J]. *Journal of Intelligence*, 2021, 40(12): 169-176.
[10] Liu Haiyan, Zhang Zhiyi, Yin Xiaohu. Dynamic Analysis of Disciplinary Research Topics Based on Paper Titles [J]. *Information Science*, 2019, 37(4): 36-43, 136.
[11] Yue Lixin, Zhou Xiaoying, Chen Yini. Research on Identification of Core Research Topics in Journal Articles and Visualization Methods for Their Evolutionary Paths: A Case Study of Journal Articles in China's Medical Health Information Field [J]. *Library and Information Service*, 2020, 64(5): 89-99.
[12] Liu Bowen, Bai Rujiang, Zhou Yanting, et al. Identification of Frontier Themes in Scientific Research from the Perspective of Integrating Fund Project Data and Paper Data: A Case Study of the Carbon Nanotube Field [J]. *Data Analysis and Knowledge Discovery*, 2019, 3(8): 114-122.
[13] Wu Yiping, Yu Chunliang, Qu Jiabin, et al. Research on Frontier Fields and Evolutionary Development Trends of University Papers from the Perspective of Text Topics [J]. *Information Science*, 2021, 39(5): 156-162, 183.
[14] Ma Ming, Wang Chao, Zhou Yong, et al. Research on Methods for Identifying Core Technology Topics and Analyzing Evolutionary Trends Based on Semantic Information [J]. *Information Studies: Theory & Application*, 2021, 44(9): 106-113.
[15] Li Xin, Wen Yang, Huang Lucheng, et al. Research on a Machine Learning-Based Method for Identifying Research Frontiers [J]. *Science Research Management*, 2021, 42(1): [page numbers].

[16] Xia Lei. Research on Identifying Interdisciplinary Topics Based on Journal Papers [J]. *New Century Library*, 2019(12): 62-67.
[17] Yang Jing, Wang Fang, Bai Rujiang. A Method for Evaluating the Innovation of Individual Academic Papers Based on Research Topic Comparison [J]. *Library and Information Service*, 2018, 62(17): 75-83.
[18] Zhang L, Sun B, Shu F, et al. Comparing paper level classifications across different methods and systems: an investigation of Nature publications [J]. *Scientometrics*, 2022, 127: 7633-7651.
[19] Wang Xin. From Disciplinary Classification System to Topic Classification System: Analysis of the InCites Citation Topics Classification System [D]. Dalian University of Technology, [year].
[20] Geng Haiying, Zhang Jiandong, Yang Liying, et al. A Review of Research on Algorithmically Constructed Paper-Level Disciplinary Classification Systems [J/OL]. *Information Studies: Theory & Application* [2023-5-6]. http://kns.cnki.net/kcms/detail/11.1762.G3.20230330.1028.002.html.
[21] CWTS Leiden Ranking [EB/OL]. [2022-3-7]. https://www.leidenranking.com/information/fields/.
[22] Introducing Citation Topics in InCites [EB/OL]. [2022-3-7]. https://clarivate.com/blog/introducing-citation-topics/.
[23] A New Scientific Data Platform—Dimensions [EB/OL]. [2022-5-29]. http://blog.sciencenet.cn/blog-1792012-1113138.html.
[24] Dimensions Data Guide [EB/OL]. [2022-5-6]. https://www.dimensions.ai/resources/a-guide-to-the-dimensions-data-approach/.
[25] Dong Yanbang. Analysis of Characteristics of Research Impact and Collaboration in Chinese Universities: Based on the 2013 Leiden Ranking [J]. *Higher Education Development and Evaluation*, 2014, 30(5): 19-28, 114-115.
[26] Miao Yanwei, Dong Yanbang. Comparative Analysis of Research Impact and Collaboration of Top Universities in Asian Countries and Regions: Based on the 2013 Leiden Ranking [J]. *Modern Educational Science*, 2014(5): 36-41.
[27] Wang Yan, Yao Wei, Du Min, et al. Using InCites' New Feature Citation Topics to Assist Academic Journal Editors in Developing Topic Selection Plans: A Case Study of Horticulture Research [J]. *Chinese Journal of Scientific and Technical Periodicals*, 2021, 32(6): 777-785.
[28] Bao Ying, Cui Yujie, Wen Juan, et al. Research on Topic Planning Paths for Scientific Journals Based on InCites: A Case Study of Soil Science [J]. *Journal of Southwest University (Natural Science Edition)*, 2022, 44(12): 221-231.
[29] Ding Zuoqi. Evaluation of International Influence of Excellence Action Plan Journals Based on the Dimensions Platform [J]. *China Publishing*, 2020(20): 3-8.
[30] Li Chuwei, Ding Zuoqi. Analysis of Characteristics of Top 100 Articles by Altmetrics Score in Journals Funded by the "China Science and Technology Journal Excellence Action Plan" [J]. *Science-Technology & Publication*, 2020(10): 135-140.
[31] Lutz Bornmann. Field classification of publications in Dimensions: a first case study testing its reliability and validity [J]. *Scientometrics*, 2018, 117: 637-640.

[32] Christian Herzog, Brian Kierkegaard Lunn. Response to the letter field classification of publications in dimensions: a first case study testing its reliability and validity [J]. *Scientometrics*, 2018, 117: 641-645.
[33] Li Jie, Meng Ye, Jin Jiali, et al. Comparative Study of Emerging Science Citation Index Databases [J]. *Journal of Academic Libraries*, 2021, 39(6): 48-55, 77.

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