Postprint of Research on Faceted Search for Government Open Data

Authors: Lin Xin, Long Cunyu, Luo Yu

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

Abstract

[Purpose/Significance] Construct a facet system and an open government data facet retrieval model for faceted retrieval to optimize user experience. [Method/Process] Based on the characteristics of open government data, construct the open government data facet system; on this basis, build the open government data facet retrieval model and implement a prototype system to verify the model' s feasibility and effectiveness. [Results/Conclusion] The prototype system implementation confirms that the constructed open government data facet system can effectively help users explore and filter search results; the proposed facet retrieval system model has good user experience, is convenient to operate, and can avoid the problem of information overload.

Full Text

Research on Faceted Search for Government Open Data

Lin Xin[1],[2], **Long Cunyu**[1], **Luo Yu**[1]
[1]School of Information Management, Central China Normal University, Wuhan 430079
[2]Hubei Data Governance and Intelligent Decision Research Center, Wuhan 430079

Abstract: [Purpose/Significance] This study constructs a faceted system for faceted search and a faceted search model for government open data to optimize user experience. [Method/Process] Based on the characteristics of government open data, we constructed a faceted system for government open data. On this foundation, we built a faceted search model for government open data and implemented a prototype system to verify the feasibility and effectiveness of the model. [Result/Conclusion] The prototype system implementation demonstrates that the constructed government open data faceted system can effectively help users explore and filter search results. The proposed faceted search system

model provides a good user experience, is convenient to operate, and can avoid information overload problems.

Keywords: Government open data; Faceted search; Faceted system

---

Faceted search significantly improves users' information retrieval efficiency and enhances their overall experience [1-2], and has been widely applied in government open data retrieval practice. The faceted system and its display control are core elements affecting user experience in faceted search. However, investigations of existing government open data websites reveal problems in facet settings and display control that are inconsistent with users' cognitive habits, leading to poor retrieval experiences.

Regarding facet settings, three main issues exist: First, some government websites have overlapping focal points across content-related facets, resulting in poor usability of the faceted system. For example, the Shenzhen government open data platform simultaneously maintains three facets—domain, theme, and industry—with high degrees of focal overlap. Second, some government websites have incomplete faceted systems that fail to comprehensively reveal the characteristics of government open data. Third, some facets have unreasonably configured common focal points, preventing facets from effectively fulfilling their value, or they have too many focal points causing information overload [3]. For instance, on the Zhejiang provincial data open platform, the data format facet includes focal points such as "XSL," "XML," and "JSON," but users are unlikely to forego obtaining a particular open dataset simply because of its format.

In terms of display control, existing government open data websites lack dynamic adjustment mechanisms based on retrieval results, making them prone to information overload due to excessive facets or focal points. Additionally, current government open data is scattered across various national and regional platforms, lacking integrated resource discovery. Based on these observations, this paper proposes to construct a faceted system based on user needs and data characteristics, and then build a government open data integrated retrieval system model for multi-source data to support integrated discovery of government open data.

1. Related Research and Practice

When designing a faceted search model, in addition to incorporating all functional modules of conventional retrieval systems, it is necessary to design facet system construction, facet system application strategies, and interactive interface design. Among these, facet system construction and facet system application strategy design are most relevant to this study. The following discussion will focus on three aspects: facet system construction methods, facet application strategies, and the progress of faceted search practice for government open data.

1.1 Facet System Construction Methods

Research on facet system construction methods primarily focuses on faceted transformation, faceted system improvement, and automated facet construction. Regarding faceted transformation, Hu Changping and Lin Xin proposed constructing scientific literature content facets through faceted transformation based on thesauri [4]. Gao Wenfei and Zhao Xinli proposed a faceted transformation concept based on the "Comprehensive E-Government Thesaurus" [5]. Jia Guojun et al. constructed seven basic facets and thirteen extended facets applicable to e-government thesauri based on the principles and concepts of FAST subject facet application [6].

In terms of faceted system improvement, relevant research addresses the core problem of how to more comprehensively and deeply reveal the content and quality characteristics of resources. Some studies propose incorporating new attribute features such as user-generated content, ratings, and popularity into the faceted system [7-9]. Other studies improve existing facets, such as creating new facets based on the educational institutions of politicians' children in political figure retrieval [10]. Although the expert method remains the mainstream approach for facet system construction to ensure the rationality and comprehensibility of facet and focal point selection [11-12], semi-automated and automated construction strategies have begun to emerge. Semi-automated facet construction involves manually setting the facet framework while automatically constructing the focal set through information extraction, such as F. Abel' s use of entity recognition technology to construct focal sets for location, organization, and personnel facets in fire incidents [13]. Fully automated facet construction aims to automate both facet framework setting and focal set construction, such as Dou Zhicheng et al.' s proposed strategy for automatic facet mining that integrates retrieval results with massive corpus resources [14].

1.2 Facet System Application Strategies

To fully leverage the navigational role of faceted systems while avoiding information overload, it is necessary to design facet sorting strategies, display control strategies, and interaction mechanisms.

1.2.1 Facet Sorting Strategies When there are few facets, the common strategy is manually specifying a fixed order. This approach provides users with stable, predictable sorting while ensuring overall rationality, thereby improving facet utilization efficiency and representing the most popular strategy in practice. However, research has also explored dynamic sorting strategies to accommodate scenarios with numerous facets. For example, E. Kharlamov et al. designed a sorting strategy that comprehensively considers three factors: the number of results obtained after applying a facet, whether unique results can be obtained, and the richness of browsing paths [15]. Kong Shengqiu et al. proposed a relevance judgment method based on retention rate [16]. For focal sorting within facets, common dynamic sorting strategies arrange focal

points in descending order of relevant results quantity, while common static strategies include manual specification, alphabetical order, chronological order, and spatial order [17].

1.2.2 Display Control Strategies Reasonable display control strategies help alleviate information overload. As faceted systems become increasingly rich and complete, the importance of display control has grown further. Regarding whether to display facets, E. Kharlamov et al. suggested collapsing less important focal points and facets [18], while Li Bing proposed determining facet display based on user query intent [19]. For hierarchical facet display forms, Hu Changping et al. suggested dynamically adjusting the number of facet hierarchy levels based on retrieval result distribution to avoid excessive depth [20].

1.2.3 Facet Interaction Mechanisms This aspect primarily concerns the logical operations among multiple focal points within the same facet and across different facets. The prevailing solution in research and practice is to apply logical "OR" operations among focal points within the same facet and logical "AND" operations among focal points across different facets [21]. However, some studies note that when users select multiple focal points within the same facet, their needs may also require logical "AND" operations [22].

1.3 Progress in Related Practice

Currently, China has launched over 130 government data open platforms, most of which provide faceted search services. The following analysis examines four typical platforms: Zhejiang Province, Guizhou Province, Shandong Province, and Guiyang City. The Zhejiang government open data platform contains nearly 10,000 datasets and provides three facets: domain, format, and data source unit, with all focal points displayed regardless of retrieval result distribution. The Guizhou platform offers over 1,400 datasets with six facets: theme, scenario, department, type, format, and open attribute, where focal points for the first three facets are hidden by default while the latter three display all focal points directly. The Shandong platform provides nearly 1,000 datasets with four facets: domain, format, open type, and provincial department, with all focal points displayed directly for each facet, though only the provincial department facet shows the number of datasets per focal point. The Guiyang platform offers nearly 3,000 datasets with seven facets: domain, theme, industry, service, municipal department, district/county list, and data format. Regarding interaction mechanisms, none of these four platforms support selecting multiple focal points within the same facet, and focal points across different facets are all in logical "AND" relationships.

Overall, domestic and international research has explored various aspects of facet system construction methods and application strategies, providing valuable references for government open data faceted search research and practice. However, theoretical research on faceted search specifically for government open

data remains limited. While practical exploration has made initial progress, prominent issues persist: facet systems are mostly constructed from external features of open data, inadequately revealing data content characteristics; some facets offer little value for user exploration and filtering of retrieval results, easily causing information overload; interaction mechanisms lack flexibility, not supporting simultaneous selection of multiple focal points within a facet or dynamic control of facet system display, thereby affecting user experience. To address these problems, this paper first reconstructs the facet system based on users' faceted application needs, then builds a user-oriented government open data faceted search model to optimize the user experience of government open data faceted search services.

2. Government Open Data Faceted System Construction

Currently, government open data in China is published independently by governments at various levels, resulting in highly dispersed data distribution. Therefore, to implement government open data faceted search, it is first necessary to collect, integrate metadata, and annotate it according to the government open data faceted system to establish associations between datasets and the facet system. Users only need to utilize a few characteristics—data theme, type (such as lists, announcements, statistics), publishing organization, and time range—to explore and filter retrieval results and obtain datasets from unique sources.

When retrieving open data, users prioritize data features that can assist their search, generally combining content features and external features for screening. For content features, users prioritize the theme and type of open data, using the former to quickly determine whether the content meets their needs and the latter to locate specific types of open data. For external features, users prioritize the publishing organization and update time, judging data quality through the publishing organization and obtaining information within the required time range through update time. Based on this understanding, the government open data faceted system should include six facets: open data theme, open data type, administrative division level, administrative division name, publishing organization type, and data update time, as shown in Figure 1 [Figure 1: see original paper].

2.1 Open Data Theme Facet

Different government open data websites provide different theme facets, resulting in a chaotic and non-standardized open data theme facet landscape. The "Part 4: Government Information Resource Classification" of the "Government Information Resource Catalog System" (GB/T21063.4-2007) [23] proposes a thematic category system comprising 21 first-level categories and 133 second-level categories, which comprehensively covers government information resource themes and has been widely applied in practice. This study adopts this category system as the basis for the government open data theme facet.

However, some first-level categories are overly generalized and cannot effectively partition themes, while underlying second-level categories vary significantly. For example, the first-level category "Land Resources, Energy" includes nine second-level categories: "Land," "Minerals," "Water Resources," "Ocean," "Coal," "Petroleum," "Fuel, Gas," "Electricity," and "Comprehensive Category," with "Comprehensive Category" also appearing under other first-level categories. Therefore, using only first-level categories as focal points would be too coarse to help users accurately locate needed resources, while using only second-level categories would cause confusion with the "Comprehensive Category" across themes and create excessive focal points, leading to information overload. Based on this analysis, both first-level and second-level categories from GB/T21063.4-2007 are incorporated into the theme facet, forming a hierarchical faceted system with sub-facets.

2.2 Open Data Type Facet

Typically, a dataset covers only one or several aspects of its theme or belongs to different document types. Accordingly, datasets can be categorized into different types to indicate the nature and purpose of the data they contain, such as lists, indicators, details, announcements, and notifications. Given the lack of systematic classification for current government open data types, this study employs an inductive method to configure focal points for the open data type facet.

First, dataset names from nine provincial and municipal government open data platforms, including Guizhou and Guangdong, were collected as foundational data for constructing the focal point vocabulary set. Next, a combination of manual analysis and machine-assisted methods was used to analyze all 21,036 data entries, extracting feature words that reflect data types. Finally, manual normalization, merging, and deduplication were performed on the extraction results, forming a focal point vocabulary set for the data type facet containing terms such as "list."

2.3 Administrative Division Level, Name, and Publishing Organization Type Facets

As government open data continues to develop and improve, an increasing number of government agencies will participate as data publishers. When constructing cross-regional, cross-level, or even national government open data discovery platforms, directly using publishing organizations as focal points would cause severe information overload and usability issues. To address this problem, this study proposes three facets: administrative division level, name, and publishing organization type, enabling both compound selection of multiple organizations through single focal points and precise positioning of publishing organizations through cross-facet focal combinations.

(1) Administrative Division Level Facet. China' s current administrative

divisions have four levels: provincial, prefecture-level city, district/county, and township. In practice, township-level open data is scarce, while national-level open data is abundant and highly open. Therefore, referencing China's four-level administrative structure, the focal points for the administrative division level facet are set as national, provincial, prefecture-level city, and district/county.

(2) Administrative Division Name Facet. Division names refer to the specific administrative divisions to which open data publishing organizations belong, such as Hubei Province, Zhengzhou City, and Yuhang District of Hangzhou City. This facet has a three-level structure comprising province, city, and district/county, enabling hierarchical positioning through affiliation relationships.

(3) Publishing Organization Type Facet. Under China' s administrative system, despite strong consistency in institutional settings across local governments at different levels and between local and central governments—for example, the central government has a Ministry of Education while local governments have education departments and bureaus—differences exist, such as the universal establishment of government service data management bureaus at local levels without corresponding national-level departments. Therefore, considering state organ configurations and based on data from eight provincial and prefecture-level government open data platforms including Shanghai and Shandong, focal points for publishing organization types were inductively identified, resulting in 122 focal points including education departments, public security departments, and cyberspace administration departments.

2.4 Data Update Time Facet

Government open datasets are time-sensitive. When selecting open datasets, users consider update times to obtain information within their required time ranges. Theoretically, update time is a continuous concept whose values change over time. In practical application, precise year and month of data updates can satisfy users' filtering needs when combined with other facets. Therefore, the update time facet includes year and month levels to help users conveniently select specific years and months.

3. Government Open Data Faceted Search System Model

The current publication of government open data in China is completed independently by governments at various levels, resulting in highly dispersed data distribution. Therefore, to implement government open data faceted search, it is first necessary to collect and integrate metadata and annotate it according to the government open data faceted system to establish associations between datasets and the facet system. Subsequently, application strategies for the facet system must be designed, including sorting, display control, and interaction mechanisms.

Based on this process, the architecture of the government open data faceted search system is shown in Figure 2 [Figure 2: see original paper], comprising

government open data metadata collection and integration modules, automatic indexing module, facet application strategy module, retrieval module, and interactive interface module. The metadata collection and integration module provides data support for the automatic indexing module and forms the foundation for facet system design and facet application strategy modules. The automatic indexing module connects the open data integration module with the facet system module. The open data retrieval module and facet application strategy module connect users with computers. Interactive interface design can reference research results from [24-25], while the retrieval module is similar to ordinary retrieval systems and will not be discussed further. The following sections will focus on the open data metadata collection, integration and indexing module, and the facet application strategy module.

3.1 Government Open Data Metadata Collection

To achieve integrated discovery of government open data, metadata must first be collected, including not only the government open data metadata published by governments at various levels but also the URL information of each dataset. Metadata collection can be achieved either through collaboration with government open data platforms, where platforms regularly push metadata information for newly added or updated datasets, or through directed collection using web information collection tools to actively obtain data from various government data open platforms. When using directed collection methods, in addition to collecting new datasets, it is necessary to monitor previously collected dataset information to promptly identify updated data.

3.2 Government Open Data Integration

Since current open data metadata collection uses official platforms as information sources and each dataset is unique, the main tasks of government open data integration are metadata standardization and heterogeneous metadata mapping strategy design. Metadata standardization involves data conversion and information supplementation to enhance data normativity, with particular attention needed for title and publishing organization standardization and theme metadata standardization.

Title standardization requires processing dataset titles that do not explicitly reflect regional information (e.g., "Baiyun District Statistical Monthly Report," which makes it difficult to determine whether it refers to Baiyun District in Guiyang or Guangzhou) to clearly include prefecture-level city and district/county information, thereby avoiding ambiguity and improving title comprehensibility. Publishing organization standardization similarly requires completing incomplete organization names. Theme metadata standardization addresses cases where some government open data platforms do not use GB/T21063.4-2007 for thematic cataloging but instead use data domain or industry vocabularies, requiring conversion to corresponding thematic terms in GB/T21063.4-2007 based on mapping relationships between the two vocabu-

laries. Heterogeneous metadata mapping strategy design involves establishing mapping rules for metadata items across different open data platforms to resolve issues arising from different metadata standards or different naming conventions for the same metadata item.

3.3 Government Open Dataset Annotation

Based on data integration, datasets must be annotated according to the constructed government open data faceted system to lay the data foundation for resource retrieval and facet application. Dataset annotation is primarily achieved through existing metadata extraction, rule-based annotation, and machine learning-based annotation. Metadata extraction involves extracting information already contained in open dataset metadata, applicable mainly for annotating open data themes and update times. Rule-based annotation involves extracting basic information from open dataset-related information using extraction rules and then establishing associations between datasets and corresponding facet focal points through mapping rules, primarily applied for annotating administrative division levels, names, and publishing organization types. Machine learning-based annotation utilizes machine learning methods for model training to achieve annotation of open data themes and types.

3.4 Facet Application Strategy Module

To improve users' efficiency in using facets and reduce information overload caused by facet display, strategies must be designed for facet system sorting, display control, and interaction mechanisms.

(1) Facet System Sorting. Given that the government open data faceted system contains only six facets, all commonly used, this study adopts a fixed sorting strategy. The facet order is: open data theme, open data type, administrative division level, publishing organization type, administrative division name, and data update time. Different facets employ different focal sorting strategies: administrative division level uses fixed order sorted from high to low level; data update time sorts years in descending order and months from December to January; other facets sort focal points by the number of covered open datasets from most to least.

(2) Facet System Display Control. When retrieval results are limited, users do not need facet systems to explore or filter results, so facet systems are not displayed. When facet systems are displayed, display control is achieved through vocabulary level adjustment and controlling the number of default displayed focal points. Vocabulary level adjustment means that when an open data thematic term includes narrower terms and datasets containing the thematic term concentrate in one or a few narrower terms, the thematic term is removed from the facet system and the narrower terms are promoted to simplify the facet hierarchy. Controlling the default number of displayed focal points means that when too many focal points correspond to retrieval results, only the top few

focal points are displayed by default based on sorting results, while supporting users' autonomous viewing of all focal points.

(3) Facet Interaction Mechanisms. To enhance facet interaction flexibility, multiple focal point selection is allowed within the same facet or across different facets. Focal points within the same facet have a logical "OR" relationship, while focal points across different facets have a logical "AND" relationship.

4. Prototype Implementation of Government Open Data Faceted Search System

To verify the feasibility and effectiveness of the government open data faceted search model, this study collected datasets from six provincial government open data platforms (including Guangdong and Guizhou) and 24 prefecture-level city platforms (including Binzhou, Shenzhen, and Hangzhou) as foundational data (collected between December 25-30, 2020) to implement the prototype system. The prototype system adopts a B/S (Browser/Server) architecture, with an Intel(R) Core(TM) m3-8100Y CPU @ 1.10GHz 1.61GHz processor and 4GB RAM. The software operating system platform is Windows 10 for x64-based processors.

4.1 Prototype System Construction Process

Thanks to the emphasis on data standardization across provincial and municipal government open data platforms, internal consistency within each dataset is strong. However, severe data heterogeneity exists between provinces and cities, including significant differences in metadata systems, different naming conventions for the same metadata item, and different value specifications. Taking Guangdong and Hangzhou government open data platforms as examples, the former includes a "keyword" metadata item absent in the latter; the former calls the publishing organization metadata item "data provider" while the latter uses "publishing department" ; the former' s publication date is precise to the day while the latter is precise to the second. Additionally, titles, publishing organizations, and thematic cataloging in metadata across provinces and cities exhibit non-standard phenomena.

Therefore, during data integration, the study first established the prototype system' s metadata system based on existing metadata systems from various open data platforms. Then, mapping relationships were created between provincial and municipal platforms and the prototype system for metadata items that could not be precisely matched, resolving metadata naming heterogeneity. Finally, non-standard or heterogeneous metadata values were processed to maintain consistency across provincial and municipal open data, achieving data integration.

In the data annotation phase, the prototype system focused on reflecting the previously constructed faceted system within datasets, requiring annotation of

dataset themes, types, administrative division levels and names, publishing organization types, and update times. Theme and update time were achieved through metadata extraction. Administrative division name annotation used rule-based extraction from data titles. Administrative division level was obtained through mapping relationships between division names and levels. For publishing organization type annotation, information was first obtained from the publishing organization metadata item, from which organization type feature words were extracted and then annotated according to a mapping vocabulary. Data type annotation similarly employed a rule-based method: if a focal point word from the data type facet appeared in a dataset' s title, that word was used as its data type; otherwise, it was annotated as comprehensive.

Facet application strategy design implemented facet and focal sorting and facet interaction mechanisms as described in the previous model. For display control, when retrieval results were fewer than 10 (the maximum number displayed per page), displaying facets had little value and might disturb users, so facets were not displayed. By default, each facet displayed a maximum of five focal points or sub-facets to avoid information overload. When a sub-facet contained only one or two focal points (except for administrative division names), the sub-facet was not displayed; instead, the focal point level was promoted for direct display. When a facet contained only one focal point, the facet was hidden.

4.2 Prototype System Effectiveness

Overall, the government open data faceted search prototype system provides a good user experience. The system' s facets help users conveniently explore and filter retrieval results. Influenced by display control strategies, the system also avoids information overload. Using the search for COVID-19 designated hospital lists across regions as an example demonstrates the prototype system' s effectiveness.

First, searching with the query "novel coronavirus" yields the interface shown in Figure 3 [Figure 3: see original paper]. With numerous retrieval results, facet system assistance is needed for convenient result filtering. The left-side displayed facet system shows facets such as theme (including healthcare, comprehensive government affairs), type (including list, announcement), and publishing organization (health departments, local governments) that help users quickly filter results. Benefiting from display control strategies, the default number of displayed facets and focal points is limited, maintaining a relatively flat system that avoids information overload and provides good retrieval experience.

Given the need for hospital lists, selecting "healthcare" under open data theme and "list" under data type filters the results. As shown in Figure 4 [Figure 4: see original paper], retrieval results are reduced from 97 to 12, with data relevance significantly improved—four of the top five results are relevant. This interaction process demonstrates the feasibility and effectiveness of the constructed government open data faceted system and faceted search system model, which

can effectively support users in searching government open data, substantially improving their search efficiency and interaction experience.

Conclusion

In current government open data faceted search practice, unreasonable facet system settings and application strategy design lead to poor user experience. To address this issue, this paper constructed a government open data faceted system and designed a government open data faceted search system model, implementing a prototype system based on data from six provincial and 24 prefecture-level city government open data platforms. The results demonstrate that the constructed government open data faceted system can help users quickly filter and locate needed open data, and the proposed government open data integration strategy and facet system application strategy in the faceted search system model are highly effective. They can assist users in exploring and filtering search results while reducing cognitive burden and avoiding information overload problems.

References

[1] KULES B, CAPRA R. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface[J]. Journal of the American Society for Information Science and Technology, 2012, 63(1): 114-138.

[2] MAHDIN M N, AHMAD A R, ISMAIL R. Paradigm extension of faceted search techniques: a review[J]. Journal of telecommunication, electronic and computer engineering, 2017, 9(2/7): 149-153.

[3] LIN X, WU Q. Comparative study on faceted search functions in literature resource discovery systems[J]. Digital Library Forum, 2019, 4(9): 16-23.

[4] HU C, LIN X. Facet construction based on the faceted transformation of thesauri in scientific literature retrieval[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(8): 875-884.

[5] GAO W, ZHAO X. Research on thematic classification of government information resources in China[J]. Information Science, 2008, 26(9): 1340-1343.

[6] JIA G, WU X. Faceted transformation of the comprehensive e-government thesaurus based on FAST[J]. Library and Information Service, 2014, 58(8): 105-110.

[7] CHEN G, XIAO L, SUN J. Construction of a faceted navigation system for online communities: a case study of the DXY cardiovascular forum[J]. Information Studies: Theory & Application, 2017, 40(10): 112-116.

[8] WANG J, ZHOU Y. Research on faceted metadata for e-commerce platforms[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(3): 317-325.

[9] VANDIC D, DAM J W V, FRASINCAR F. Faceted product search powered by the semantic Web[J]. Decision support systems, 2012, 53(3): 425-437.

[10] ARENAS M, GRAU B C, KHARLAMOV E, et al. SemFacet: semantic faceted search over yago[C]//International conference on world wide web. Seoul: ACM, 2014: 123-126.

[11] XIE Z, SHI G, YANG H, et al. Construction and application of a faceted ontology in the international freshwater dispute domain[J]. Journal of Intelligence, 2018, 37(11): 192-196.

[12] KONG W. Extending faceted search to the open-domain Web[J]. ACM SIGIR forum, 2016, 50(1): 90-91.

[13] ABEL F, HAUFF C, HOUBEN G J, et al. Twitcident: fighting fire with information from social Web streams[C]//International conference on world wide web. Lyon: ACM, 2012: 305-308.

[14] DOU Z, JIANG Z, LI J, et al. Query facet mining method based on term graph analysis[J]. Chinese Journal of Computers, 2017, 40(3): 556-569.

[15] KHARLAMOV E, GIACOMELLI L, SHERKHONOV E, et al. SemFacet: making hard faceted search easier[C]//ACM on conference on information and knowledge management. Singapore: ACM, 2017: 2475-2478.

[16] KONG S, FENG J, DU B. Design and implementation of a dynamic faceted search engine for water conservancy metadata[J]. Computer Technology and Development, 2017, 27(10): 151-155.

[17] QIU M. Research on faceted navigation design for information query systems[J]. Modern Intelligence, 2018, 38(10): 78-84.

[18] KHARLAMOV E, GIACOMELLI L, SHERKHONOV E, et al. SemFacet: making hard faceted search easier[C]//ACM on conference on information and knowledge management. Singapore: ACM, 2017: 2475-2478.

[19] LI B. Research on adaptive book faceted search based on query intent recognition[J]. Library Science Research, 2017(15): 57-64.

[20] HU C, LIN X. Facet construction based on the faceted transformation of thesauri in scientific literature retrieval[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(8): 875-884.

[21] DEUSCHEL T, GREPPMAIER C, HUMM B G, et al. Semantically faceted navigation with topic pies[C]//International conference on semantic systems. Leipzig: ACM, 2014: 132-139.

[22] ARENAS M, GRAU B C, EVGENY E, et al. Towards semantic faceted search[C]//International conference on World Wide Web. Seoul: ACM, 2014: 219-220.

[23] National Information Technology Standardization Technical Committee. Government information resource catalog system—Part 4: Government infor-

mation resource classification: GB/T21063.4-2007[S]. Beijing: China Standard Press, 2007: 9.

[24] CAPRA R, MARCHIONINI G. The relation browser tool for faceted exploratory search[C]//Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries. Pittsburgh: ACM, 2008: 420-428.

[25] DANIEL T. Faceted search[J]. Synthesis lectures on information concepts, retrieval, and services, 2009, 1(1): 1-80.

Author Contributions: Lin Xin: Responsible for topic selection, paper writing and revision; Long Cunyu: Responsible for model and prototype implementation, initial draft writing; Luo Yu: Participated in prototype implementation and literature collection.

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