Analysis and Development Strategies for Cross-Language Retrieval Functionality in Multilingual Shared Databases of the Belt and Road Initiative: Postprint

Authors: Si Li, Zhou Jing

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

Abstract

[Purpose/Significance] To effectively utilize multilingual shared database resources for the "Belt and Road" Initiative, the cross-language information retrieval problem must be addressed. Based on survey results of existing "Belt and Road" database retrieval functions, this study analyzes the functional requirements for "Belt and Road" multilingual shared databases. From the perspective of investigating cross-language retrieval platforms, it provides references for the design and development of cross-language retrieval functions for "Belt and Road" multilingual shared databases. [Method/Process] Through literature review and web-based investigation, 11 representative cross-language retrieval platforms from domestic and international sources were selected. Analysis was conducted across six dimensions: cross-language retrieval methods, cross-language translation implementation methods, retrieval function configuration, retrieval result presentation, interface design, and supported retrieval languages, thereby summarizing their implementation approaches. [Results/Conclusion] This paper proposes strategies for designing and developing cross-language retrieval functions for "Belt and Road" multilingual shared databases: adopting a query-document translation method based on neural machine translation, implementing diverse retrieval functions, applying visualization technologies for retrieval result presentation, and providing multilingual retrieval interfaces and resources.

Full Text

Analysis and Development Strategies for Cross-Language Retrieval Functions in the "Belt and Road" Multilingual Shared Database

Si Li, Zhou Jing
School of Information Management, Wuhan University, Wuhan 430072

Abstract:
[**Purpose/Significance**] To achieve effective utilization of resources in the "Belt and Road" multilingual shared database, the problem of cross-language retrieval must be solved. Based on survey results of existing "Belt and Road" database retrieval functions, this study analyzes the functional requirements for cross-language retrieval in "Belt and Road" multilingual shared databases. From the perspective of investigating cross-language retrieval platforms, it provides references for the design and development of cross-language retrieval functions for "Belt and Road" multilingual shared databases. [**Method/Process**] Through literature and web surveys, 11 typical cross-language retrieval platforms at home and abroad were selected and analyzed from six aspects: cross-language retrieval methods, cross-language translation implementation methods, retrieval function settings, retrieval result presentation, interface design, and supported retrieval languages. Their implementation approaches were summarized. [**Result/Conclusion**] Strategies are proposed for the design and development of cross-language retrieval functions for "Belt and Road" multilingual shared databases: adopting the query-document translation method based on neural machine translation, implementing multiple retrieval functions, applying visualization technology to present retrieval results, and providing multilingual interfaces and resources.

Keywords: "Belt and Road" database; multilingual; cross-language retrieval

---

2. Related Research

How to implement cross-language retrieval functions for the "Belt and Road" multilingual shared database and solve its multilingual problems is the research objective of this paper. Therefore, relevant research is reviewed from two aspects: cross-language retrieval translation methods and their implementation, and multilingual issues in "Belt and Road" databases.

2.1 Cross-Language Retrieval Translation Methods and Implementation

2.1.1 Cross-Language Retrieval Translation Methods Cross-language retrieval translation methods include query translation, document translation, query-document translation, interlingual translation, and non-translation

methods. Among them, query translation, document translation, and query-document translation are currently the mainstream methods [4-5]. The query translation method converts the query language into the target language of the documents before performing monolingual retrieval. The document translation method converts the source documents into the query language—that is, it does not translate the query but translates the documents in the collection to match the query language [6]. The query-document translation method combines the advantages of both, reducing user translation costs while improving retrieval service quality, making it an ideal choice for implementing cross-language retrieval [7].

2.1.2 Cross-Language Translation Implementation Methods Cross-language translation implementation methods include machine translation, corpus-based methods, dictionary-based methods, hybrid dictionary-corpus methods, and ontology-based methods [6, 8]. The machine translation method uses machine translation systems for translation. P. Iswarya and V. Radha developed a cross-language text retrieval system using a hybrid rule-based and statistical machine translation system, improving translation accuracy and efficiency [9]. Corpus-based methods describe the same information or same-topic information in two or more languages and establish connections between different language versions manually or automatically [10]. R. Rahimi et al. used a probability distribution-based model to extract correspondences between source and target words, building translation models in comparable corpora to provide more reliable translations for low-frequency words [11]. Dictionary-based methods use machine-readable dictionaries to translate user queries into target languages for retrieval [12]. O. F. W. Onifade et al. proposed a fuzzy bilingual dictionary with dual concept-driven document clustering technology to extend dictionary translation models [13]. Hybrid dictionary-corpus methods combine the advantages of both, first using dictionaries to translate queries and then using specialized corpora to disambiguate results. J. Vilares et al. used parallel corpora to automatically generate bilingual machine-readable N-gram dictionaries for translation based on the query translation method, followed by monolingual text retrieval [14]. Ontology-based methods translate queries at the semantic level, perform semantic processing on retrieval objects, analyze the semantic correlation between potential target objects and query requests in semantic passages, and finally perform matching [15]. Sun Yingying et al. proposed a domain knowledge base-based scientific terminology information matching model, combining linguistic features, domain information, and long short-term memory network language models to select the most appropriate translation [16].

2.2 Multilingual Issues in "Belt and Road" Databases

Multilingual issues in "Belt and Road" databases pose new challenges for information resource construction and database retrieval services. Yu Shiyang et al. believe that multilingual issues represent the biggest difference between

"Belt and Road" data collection and ordinary database construction [17]. Existing "Belt and Road" databases generally lack information resources in minor languages and have not yet achieved the organization and integration of multilingual resources [18], which prevents them from providing resource guarantees for cross-language retrieval functions and affects information resource sharing among Belt and Road countries. Yan Dan et al. propose that attention should be paid to introducing original materials in multiple languages, especially from small and medium-sized countries, to build a multilingual, interdisciplinary, and multi-source "Belt and Road" information resource system [19-20]. Liang Haoguang et al. believe that accelerating the construction of a "Belt and Road" multilingual cloud service platform based on language technologies such as multilingual identification and multilingual perception, providing basic data resources and technical support [21], is an important link in the construction and development of "Belt and Road"databases. However, from both theoretical and practical perspectives, cross-language retrieval function analysis and development remain a gap in "Belt and Road" database construction.

3. Functional Requirements Analysis for "Belt and Road" Multilingual Shared Database

3.1 Resource Characteristics and User Needs

In terms of resource characteristics, the "Belt and Road" multilingual shared database involves multilingual, multi-type, multi-domain, and multi-source information resources. Information resources are diverse in language, with some being minor languages with low prevalence, such as Khmer, Hungarian, Lao, Polish, Serbian, and Vietnamese. Resource types cover policies and regulations, statistical data, index data, research and technical reports, news, journals and newspapers, dissertations, books, yearbooks, important reference books for economic management, patents, and standards. Thematic information resources involve multiple research fields including politics, economy, culture, law, and national security. Information sources include government agencies, international organizations, research institutions, enterprises, databases, authoritative media, think tanks, and the Internet.

In terms of user needs, information resource demands for "Belt and Road" research mainly focus on macro information needs of Belt and Road countries and regions, news reports and public opinion information released by national media, multilingual information needs in professional fields, and needs for multilingual and interdisciplinary academic resources and scientific research information [19]. It is evident that the "Belt and Road" multilingual shared database needs to integrate original information resources across different dimensions, including document language, document type, research field, and document source. The Belt and Road initiative involves 138 countries and regions, covering over a hundred languages. Currently, database users can only rely on their own multilingual information literacy and comprehension abilities or external transla-

tion tools to access original information resources in minor languages. To help users understand and access multilingual information resources, cross-language retrieval functions that meet user needs should be configured.

3.2 Database Retrieval Function Survey

To understand user needs for retrieval functions in the "Belt and Road" multilingual shared database, an investigation was conducted on existing "Belt and Road" database retrieval functions. The survey objects are detailed in Table 1 of the special article "Investigation and Development Analysis of 'Belt and Road' Thematic Databases". The survey results are as follows:

3.2.1 Cross-Language Information Retrieval Not Yet Implemented Currently, no "Belt and Road" database provides multilingual information services based on cross-language retrieval. Among existing "Belt and Road" databases, only the Silk Road Science and Technology Knowledge Service System can achieve cross-language retrieval at the metadata level. Its document resources' titles, keywords, and abstracts usually contain English or Chinese-English bilingual versions, such as "Title" and "Alternate Title," "Chinese Abstract" and "Abstract from Author" in both English and source languages. When users submit queries in different source languages, the system uses the query translation method to machine-translate the query and match it with resource metadata. However, this cannot meet the full-text retrieval needs of users from different native language backgrounds for multilingual information resources. Language barriers in information retrieval make it difficult to discover and utilize "Belt and Road" multilingual resources.

3.2.2 Most Only Support Simple Retrieval Functions Most "Belt and Road" databases only support simple retrieval, with only 27% supporting advanced retrieval. Currently, no "Belt and Road" database supports expert retrieval, and available operators and searchable fields are limited, failing to meet the retrieval needs of professional researchers. Regarding retrieval result refinement functions, 52% of "Belt and Road" databases support refining results by document type, theme, country, publication year, etc., while 12% support secondary retrieval. It is evident that "Belt and Road" databases still need improvement in expert retrieval functions, result refinement functions, and secondary retrieval functions.

3.2.3 Single Form of Retrieval Result Presentation Sorting and visualization of retrieval results help users quickly grasp the overview and characteristics of retrieved resources and accurately locate needed resources. Only 24% of "Belt and Road" databases support retrieval result sorting, and these databases can sort by publication time. Additionally, "China Belt and Road Network," "Countries of the World Database," "Belt and Road Resource Center Database" can also sort by relevance. Regarding retrieval result visualization,

the "Belt and Road Statistical Database" has substantial business data and statistical information that users can visualize and customize charts, while other databases cannot visualize results. The current single form of retrieval result presentation reduces retrieval efficiency and user experience.

3.2.4 Most Do Not Support Multilingual Interfaces "Belt and Road" database users have diverse native language backgrounds, exceeding the scope of single-language or mainstream language community information services. However, most current "Belt and Road" databases do not support multilingual interfaces, creating cognitive barriers for users who cannot read or understand other languages. Specifically, 60% of "Belt and Road" databases only support Chinese; 9% only support English; 19% support both Chinese and English. Additionally, four databases support three or more languages including Chinese and English, accounting for 12%. Among them, the U.S. EBSCO company' s database supports 30 language interfaces including English, Japanese, Korean, and German; "China Belt and Road Network" supports six UN official languages (Chinese, English, Russian, French, Spanish, Arabic); "Silk Road Science and Technology Knowledge Service System" supports Chinese, English, Russian, and Spanish; and "Xinhua Silk Road Network" supports Chinese, English, and Italian. Most databases cover few languages and lack multilingual interfaces, hindering users from using familiar languages to understand, browse, or read information resources in other languages and impeding the dissemination and utilization of multilingual information resources [3].

Based on these findings, "Belt and Road" multilingual shared databases need to implement cross-language information retrieval, provide simple, advanced, and expert retrieval functions, support retrieval result sorting, classification refinement, and visualization, and support multilingual interfaces. Among these, multilingual information services based on cross-language retrieval are key and challenging aspects of "Belt and Road" database information services, and currently there are no mature experiences to draw from in existing "Belt and Road" databases. Therefore, it is necessary to learn from the construction experience of cross-language retrieval platforms at home and abroad to build and improve the cross-language retrieval functions of "Belt and Road" multilingual shared databases.

4. Survey and Analysis of Cross-Language Retrieval Platforms

4.1 Selection of Survey Objects

Existing cross-language retrieval platforms at home and abroad are well-established and can provide references for developing cross-language retrieval functions for "Belt and Road" databases. Referring to the multilingual information organization model proposed by Li Yueting and Si Li [22], 11 cross-language retrieval platforms were selected using web and literature

surveys, including 3 cross-language databases, 3 subject information portals, 2 search engines, and 3 digital library projects:

(1) **Cross-language databases**: OECD iLibrary is a database based on information resources provided by the Organisation for Economic Co-operation and Development. IMF eLibrary is a database of economic data and analysis reports. AIpatent is a patent intelligence retrieval system developed by Nanjing Shensite Information Technology Co., Ltd.

(2) **Subject information portals**: WorldWideScience is a cross-language, cross-database scientific literature retrieval platform covering more than 70 countries and regions, approximately 100 databases and portals, and over 500 million web pages of scientific information. The Silk Road Science and Technology Knowledge Service System, developed by the International Knowledge Center for Engineering Sciences and Technology at Xi' an Jiaotong University, is an engineering science and technology knowledge service platform oriented toward Belt and Road needs. The Petroleum and Petrochemical Big Data Knowledge Service Platform is a personalized knowledge service system built for China' s petroleum and petrochemical industry.

(3) **Search engines**: 2lingual Google Search allows users to conduct Google searches in two languages—after entering a query, users can select any two languages to obtain search results. Sogou Overseas Search applies Sogou' s machine translation system to provide users with three versions of search results: English original, Chinese translation, and bilingual.

(4) **Digital library projects**: The World Digital Library (WDL) is a digital library project of human historical and cultural heritage initiated by the U.S. Library of Congress and UNESCO. The International Children' s Digital Library (ICDL), sponsored by the U.S. National Science Foundation and developed by the University of Maryland in cooperation with the Internet Archive, is a children' s digital library project that includes digital literary works reflecting different periods, regions, cultures, and languages. Europeana is a European digital cultural heritage project commissioned by the European Commission and hosted by the European Foundation, covering collections from over 1,500 museums, archives, and libraries, providing access to 53 million digital objects.

4.2 Survey Results of Cross-Language Retrieval Platforms

The core steps for cross-language retrieval platforms to implement are translation and retrieval. The survey of each platform was conducted across six dimensions: cross-language retrieval method, cross-language translation implementation method, retrieval function settings, retrieval result presentation, interface-supported languages, and retrieval-supported languages. The results are shown in Table 1.

Table 1. Survey Results of Cross-Language Retrieval Platforms at Home and Abroad

[The table content would be preserved here with proper formatting, showing the 11 platforms and their characteristics across the six dimensions]

4.2.1 Cross-Language Retrieval Methods Among the surveyed platforms, 7 platforms (64%) adopt the document translation method: OECD iLibrary, IMF eLibrary, Silk Road Science and Technology Knowledge Service System, Petroleum and Petrochemical Big Data Knowledge Service Platform, WDL, ICDL, and Europeana. Among them, IMF eLibrary, WDL, Europeana, and ICDL use the method of translating metadata of documents to be retrieved, facilitating users' understanding of basic information for each resource and enabling retrieval of relevant resources in corresponding languages. WDL, Europeana, and ICDL digital libraries mostly contain non-text collections, requiring only metadata descriptions and their translations. OECD iLibrary resources are mainly statistical data and analysis reports with small translation workloads, multiple language versions, and direct provision of multilingual translations for some full texts. Search engines' resources to be retrieved are mainly web resources with large quantities and diverse types, so they adopt the most economical query translation method with minimal workload. Four platforms (36%) adopt the query translation method: AIpatent, WorldWideScience, 2lingual Google Search, and Sogou Overseas Search.

4.2.2 Cross-Language Translation Implementation Methods Currently, the main cross-language translation implementation method is machine translation. All surveyed platforms use machine translation, which is far faster than human translation but still has relatively high error rates. To improve machine translation accuracy in specific application scenarios, especially for commercial contracts, legal provisions, and patent literature, AIpatent and Sogou Overseas Search use neural machine translation technology, the Petroleum and Petrochemical Big Data Knowledge Service Platform and ICDL use human-assisted translation, and Europeana uses context-based glossary translation. ICDL, targeting child users worldwide, has high requirements for translation accuracy, fluency, and 趣味性, and its resources are mainly children's picture books with small translation workloads, so it has established a dedicated volunteer translation team for human-assisted translation of website interfaces, basic bibliographic information, abstracts, and entire books, with a review volunteer team for proofreading.

4.2.3 Retrieval Function Settings All surveyed platforms provide simple retrieval functions, and 7 platforms (64%) provide advanced retrieval functions: OECD iLibrary, IMF eLibrary, AIpatent, WorldWideScience, Silk Road Science and Technology Knowledge Service System, Petroleum and Petrochemical Big Data Knowledge Service Platform, and ICDL. These platforms provide retrieval function usage guides to facilitate users in performing field-limited retrieval

or using logical operators, positional operators, and truncation operators for combined retrieval. AIpatent also provides concept retrieval, allowing users to independently retrieve patent information by entering invention patent technical disclosure documents or full patent texts in the concept retrieval editing box; the system performs machine translation and keyword extraction, and users can adjust keywords for secondary retrieval to quickly find patents that meet their needs.

4.2.4 Retrieval Result Presentation Sorting and scope adjustment are basic functions of retrieval platforms. Among the surveyed platforms, 9 platforms (82%) support adjusting retrieval scope through dimensions such as "language," "country," "resource type," and "author" : OECD iLibrary, IMF eLibrary, AIpatent, WorldWideScience, Silk Road Science and Technology Knowledge Service System, Petroleum and Petrochemical Big Data Knowledge Service Platform, WDL, ICDL, and Europeana. Five platforms (45%) support sorting retrieval results by "relevance" and "publication time" : OECD iLibrary, IMF eLibrary, WorldWideScience, Silk Road Science and Technology Knowledge Service System, and Petroleum and Petrochemical Big Data Knowledge Service Platform. Three platforms (27%) support secondary retrieval to narrow search scope: IMF eLibrary, AIpatent, and Petroleum and Petrochemical Big Data Knowledge Service Platform.

Two platforms (18%) can visualize retrieval results: WorldWideScience and Petroleum and Petrochemical Big Data Knowledge Service Platform. WorldWideScience can visualize the topic clustering results of retrieval results, revealing the co-occurrence frequency and distribution patterns of topics under the retrieval results. The Petroleum and Petrochemical Big Data Knowledge Service Platform can perform quantitative visualization analysis of selected retrieval results from dimensions such as publication volume, keywords, discipline, research level, literature source, institution, author, and funding.

4.2.5 Interface and Retrieval-Supported Languages Multilingual interfaces include multilingual navigation bars, buttons, lists, pop-ups, and other important page components. To meet the needs of users from different native language backgrounds, cross-language retrieval platforms provide multilingual interfaces, allowing users to directly switch interface languages on the website homepage. Among the surveyed platforms, 7 platforms (64%) support interfaces in more than one language: OECD iLibrary, IMF eLibrary, AIpatent, Silk Road Science and Technology Knowledge Service System, Petroleum and Petrochemical Big Data Knowledge Service Platform, WDL, ICDL, and Europeana. Europeana' s main function is to disseminate European historical, cultural, and scientific knowledge to the European public, and its interface language versions cover most European countries' languages. AIpatent integrates patent resources mainly from Japanese, U.S., and Chinese official patent databases, and its interface language versions are Japanese, English, and Chinese.

Regarding cross-language retrieval, 9 platforms support retrieval in no fewer than 3 languages (82%): OECD iLibrary, IMF eLibrary, AIpatent, World-WideScience, Silk Road Science and Technology Knowledge Service System, 2lingual Google Search, WDL, ICDL, and Europeana. Each platform's retrieval-supported languages mainly focus on commonly used languages such as Chinese, English, Russian, French, Spanish, Arabic, Japanese, and Portuguese. 2lingual Google Search, as an international cross-language search engine, continuously adds retrieval-supported languages—when the platform prototype was released in 2004, it only supported 11 retrieval languages, but currently supports 37 retrieval languages.

The above typical cross-language retrieval platforms mainly adopt metadata-level document translation and query translation methods. Cross-language translation implementation primarily uses machine translation, especially neural machine translation technology. Platforms provide simple and advanced retrieval functions, can sort and adjust retrieval scope, use visualization technology to present retrieval results, and support commonly used languages in interfaces and retrieval, with continuous expansion.

5. Development Strategies for Cross-Language Retrieval Functions in "Belt and Road" Multilingual Shared Database

The "Belt and Road" multilingual shared database is a shared database with multi-participant involvement, multi-source heterogeneous resource collection, and multilingual coverage. The above survey results can provide references for the design and development of cross-language retrieval functions for "Belt and Road" multilingual shared databases.

5.1 Adopt Query-Document Translation Method Based on Neural Machine Translation

In terms of translation methods, "Belt and Road" multilingual shared databases can learn from existing cross-language retrieval platforms' document translation and query translation methods by adopting the combined query-document translation method. First, translate the source language query into the source language form consistent with the documents to be retrieved for monolingual retrieval, then translate all or part of the retrieval results into information described by the source language. This method is currently an ideal approach for implementing cross-language retrieval. In terms of implementation technology, we can learn from Google, Sogou Overseas Search, and AIpatent by adopting neural machine translation as the main technology. As the mainstream artificial intelligence translation technology [34], it can train a neural network mapping from one sequence to another to output variable-length sequences, which is more efficient than other machine translation technologies in translation, dialogue, and text summarization [35]. Meanwhile, open-source tools for neural machine translation are abundant, providing platform foundations and

development specifications for building cross-language translation systems and automatic evaluation [36].

The "Belt and Road" multilingual shared database using neural machine translation can apply more advanced technologies to train models, optimize neural network structures, improve model expression capabilities, increase neural network layers, and further enhance translation quality and efficiency. The document translation method can choose to translate metadata, first two lines, abstracts, or important words in result texts. The "Belt and Road" multilingual shared database has rich resource types, especially non-text collections such as manuscripts, historical materials, videos, pictures, photos, maps, and recordings that only have entries. We can learn from Taiwan Digital Museum' s cross-language information retrieval implementation strategy [37] by translating resource metadata to provide multilingual metadata descriptions, helping users from different native language backgrounds in Belt and Road countries discover, identify, evaluate, select, and use resources, achieving resource integration, sharing, management, and long-term preservation. This method fully utilizes the advantages of both query translation and document translation, simplifying translation processes, reducing user translation costs, and improving retrieval service quality.

5.2 Implement Multiple Retrieval Functions

Most existing "Belt and Road" databases have simple and advanced retrieval functions but rarely provide expert retrieval. The "Belt and Road" multilingual shared database should meet the retrieval needs of users with different professional levels, including government, enterprise, and research users, by providing simple, advanced, and expert retrieval functions, and producing database retrieval guide documents or establishing independent help center columns in the navigation bar. Simple retrieval with a one-stop search entry should be provided on every page, allowing users to search heterogeneous database resources through a unified search entry. After users input queries, they need to select the source language used or the platform can automatically recognize the source language used by users. A link to advanced retrieval should be placed next to the simple search box for users to switch freely. Advanced retrieval functions should provide combinations of various metadata for cross-language retrieval to improve precision. Expert retrieval should be implemented because of its powerful functionality, allowing users to construct Boolean logic expressions for retrieval queries, providing more accurate direction and control over retrieval results. Cross-language expert retrieval facilitates users in constructing queries in their familiar language, thereby improving retrieval efficiency.

5.3 Apply Visualization Technology to Present Retrieval Results

Visualization technology makes the retrieval process more transparent, providing vivid and meaningful classification and organization of retrieval results, and establishing effective user feedback and interaction mechanisms. The "Belt and

Road" multilingual shared database needs to use visualization means to display the correlation relationships and development logic of multi-country, multi-type, and multilingual resources to meet users'deep-level and personalized information needs.

Retrieval results of the "Belt and Road" multilingual shared database can be visualized by analyzing and processing retrieval result data sets through statistical, clustering, and association analysis methods, converting them into two-dimensional or three-dimensional graphics, and using intuitive interactive and dynamic visualization methods to reveal multilingual information resources. This can enhance users' cognition of information and strengthen system affinity, helping users quickly understand foreign language resources and reveal the internal connections and deep meanings of information resources. We can learn from the evolutionary description visualization methods reviewed by Zhou Xiaoying and Wei Dawei to improve retrieval result browsing functions, implementing spatiotemporal narrative visualization of "Belt and Road" information resources on timelines and maps [38]. We can also refer to key information visualization technologies reviewed by Sun Qian, Sun Yusheng, Ruan Guangce, Qiu Junping, et al. to improve retrieval result analysis functions [39-42], providing appropriate view forms and hierarchical structures for visual analysis of all or batch retrieval results to help users quickly grasp the distribution of retrieval results across dimensions such as publication volume, publication time, authors, topics, journals, languages, and resource types, deeply revealing literature knowledge structures and facilitating user browsing selection. Additionally, retrieval process visualization can apply dynamic visualization retrieval and filtering technology to help users execute and track retrieval steps in a visual manner during interaction with the retrieval system, with real-time information feedback and retrieval strategy control support, reducing users' memory burden in cross-language retrieval.

5.4 Provide Multilingual Interfaces and Resources

Providing multilingual interfaces enables users to better adapt to multilingual environments. Considering broad applicability, the "Belt and Road" multilingual shared database can first support interface versions in commonly used languages such as Chinese, English, Russian, French, Spanish, and Arabic, then selectively expand to more minor language interface versions. Alternatively, it can automatically identify users' commonly used languages based on IP addresses and automatically switch database website language interfaces to match user habits. Content with regional and national characteristics (such as numbers, time, currency, etc.) should be displayed in local language formats to reduce potential understanding ambiguity during platform use and maintain cultural neutrality. The "Belt and Road" multilingual shared database implementing multilingual interfaces should maintain a single source program version that is easy to modify, maintain, and upgrade, with consistent structure and business logic between different language versions of web pages. Differences between different language versions should be concentrated at the UI layer [43], requiring no recompila-

tion when adding new language versions, facilitating easy expansion to new languages.

For existing "Belt and Road" multilingual resources, such as government documents, statistical data, survey reports, and books with multiple language versions, all language versions of resources need to be completely collected and provided on retrieval result detail pages for users to directly select and download required language versions. For website descriptions, online exhibition introductions, and information resource items with small translation workloads, full-text translation can be performed using machine translation systems with human-assisted proofreading, eliminating the need for users to select machine translation systems themselves. Considering required costs and translation accuracy, only commonly used language versions such as Chinese and English can be provided to reduce language barriers and promote the "going global" of "Belt and Road" databases.

References

[1] List of countries that have signed cooperation documents with China on jointly building the "Belt and Road"[EB/OL]. [2020-09-22]. https://www.yidaiyilu.gov.cn/gbjg/gbgk/77073.htm
[2] Su Xinning. Information Retrieval Theory and Technology [M]. Beijing: Science and Technology Literature Press, 2004.
[3] Zhao Shenghui, Hu Ying. Research on the hierarchical framework of cross-language information services in digital libraries [J]. Information Science, 2020, 38(12): 63-69.
[4] Wang Hao. Discussion on implementation methods and key technologies of cross-language information retrieval [J]. Journal of Intelligence, 2005(7): 46-49.
[5] Li Pei, Wu Lihui. Cross-language retrieval of online information [J]. Information and Documentation Services, 2004(2): 71-74.
[6] Guo Yufeng, Huang Min. Research on cross-language information retrieval theory and application [J]. Library and Information, 2006(2): 79-81, 84.
[7] Zhang Sufang. Review of research on translation ambiguity in foreign cross-language information retrieval [J]. Library Science Research, 2006(6): 72-75, 78.
[8] Si Li, Jia Huan. Progress and implications of multilingual information organization and retrieval research in China from 2004 to 2014 [J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(6): 662-672.
[9] ISWARYA P, RADHA V. Adapting hybrid machine translation techniques for cross-language text retrieval system [J]. Journal of engineering science and technology, 2017, 12(3): 648-666.
[10] Xu Mingwu, Zhao Chunlong. The name and reality of corpus translation studies in China [J]. Shanghai Journal of Translators, 2018(4): 3-9, 94.
[11] RAHIMI R, SHAKERY A, KING I. Extracting translations from comparable corpora for cross-language information retrieval using the language modeling framework [J]. Information processing & management, 2016, 52(2):

299-318.
[12] Huang Hai, Jiang Liehui, He Hongqi, et al. Design of IDA-based decompilation intermediate language [J]. Computer Engineering and Design, 2009, 30(20): 4734-4737.
[13] ONIFADE OFW, IBITOYE AOJ, MITRA P. Embedded fuzzy bilingual dictionary model for cross-language information retrieval systems [J]. International journal of information technology, 2018, 10(4): 457-463.
[14] VILARES J, VILARES M, ALONSO MA, et al. On the feasibility of character n-grams pseudo-translation for cross-language information retrieval tasks [J]. Computer speech & language, 2016, 36: 136-164.
[15] Guo Huageng, Zhao Ying. Research and application of cross-language information retrieval [J]. Modern Intelligence, 2008(9): 142-145.
[16] Sun Yingying, He Yanqing, Wu Guangyin. Research on domain knowledge base-based scientific terminology information matching model [J]. Information Science, 2019, 37(8): 16-21.
[17] Yu Shiyang, Yang Daoling, Wang Jingxuan, et al. Construction of "Belt and Road" data resource collection system [J]. E-Government, 2017(1): 8-14.
[18] Dai Yanqing, Liu Yangqing. Research on resource organization strategies of "Belt and Road" research and decision support platforms [J]. Library Science Research, 2020(16): 64-70, 80.
[19] Yan Dan, Li Mingyan. Construction of information needs and resource support system for "Belt and Road" research in universities [J]. Library Construction, 2018(8): 56-63.
[20] Yan Dan, Ma Yinxue. Research on construction status and development strategies of "Belt and Road" thematic databases [J]. Library Science Research, 2017(12): 40-47.
[21] Liang Haoguang, Zhang Yaojun. "Belt and Road" language strategic planning and policy practice [J]. People' s Tribune Academic Frontier, 2018(10): 98-105.
[22] Li Yueting, Si Li. Research on semantic-based multilingual information organization model [J]. Library Tribune, 2016, 36(2): 13-19.
[23] OECD iLibrary [EB/OL]. [2020-10-27]. https://www.oecd-ilibrary.org/.
[24] IMF eLibrary [EB/OL]. [2020-10-27]. https://www.elibrary.imf.org/.
[25] AIpatent [EB/OL]. [2020-10-27]. https://www.aipatent.com.
[26] WorldWideScience [EB/OL]. [2020-10-27]. https://world-widescience.org.
[27] Silk Road Science and Technology Knowledge Service System [EB/OL]. [2020-10-27]. http://silkroadst.ikest.org.
[28] Petroleum and Petrochemical Big Data Knowledge Service Platform [EB/OL]. [2020-10-27]. http://oil.cnki.net.
[29] 2lingual Google Search [EB/OL]. [2020-10-27]. https://2lingual.com.
[30] Sogou Overseas Search [EB/OL]. [2020-10-27]. https://overseas.sogou.com.
[31] World Digital Library [EB/OL]. [2020-10-27]. https://www.wdl.org.
[32] International Children' s Digital Library [EB/OL]. [2020-10-27]. http://en.childrenslibrary.org.
[33] Europeana [EB/OL]. [2020-10-27]. https://www.europeana.eu/portal/en.
[34] Lin Qian, Liu Qing, Su Jinsong, et al. Analysis of research hotspots and

frontier trends in neural machine translation [J]. Journal of Chinese Information Processing, 2019, 33(11): 1-14.
[35] Zhang Wen, Feng Yang, Liu Qun. Deep neural machine translation model based on simple recurrent units [J]. Journal of Chinese Information Processing, 2018, 32(10): 36-44.
[36] ZHANG B, XIONG DY, XIE JS. Neural machine translation with GRU-gated attention model [J]. IEEE transactions on neural networks and learning systems, 2020, 31(11): 4688-4698.
[37] CHEN HH. Global digital library development in the new millennium [M]. Beijing: Tsinghua University Press, 2001.
[38] Zhou Xiaoying, Wei Dawei. Research on knowledge visualization methods based on needs in the context of digital humanities—taking video content visualization of National Library Open Courses as an example [J]. Library, 2020(1): 20-28.
[39] Sun Qian. Practical exploration of resource visualization from the perspective of digital library website construction [J]. Library Theory and Practice, 2017(5): 84-87.
[40] Sun Yusheng, Li Wanrong. Research progress on information visualization in domestic digital libraries: architecture and key technologies [J]. Library Science Research, 2019(4): 2-9.
[41] Ruan Guangce, Ren Jinyu. Research on visualization application of literature retrieval results based on thematic hierarchical relationships [J]. Library Journal, 2019, 38(5): 71-78.
[42] Qiu Junping, Yu Houqiang, Lü Hong, et al. Review of foreign research on collection resource visualization [J]. Information and Documentation Services, 2014(1): 12-19.
[43] Hu Zhenning, Yang Wei, Ding Pei, et al. Design and implementation of multilingual interface for SULCMIS OPAC [J]. New Technology of Library and Information Service, 2013(2): 70-76.

Author Contributions:
Si Li: Determined the overall 思路 and framework design, revised the paper;
Zhou Jing: Conducted database surveys and data collection, wrote and revised the draft.

Analysis and Development Strategy of Cross-Language Retrieval Function for "the Belt and Road" Multilingual Shared Database

Abstract: [Purpose/significance] To realize the effective use of "the Belt and Road" multilingual shared database resources, the problem of cross-language retrieval should be solved. Based on the survey results of "the Belt and Road" database retrieval function, "the Belt and Road" multilingual shared database' s retrieval function demand is analyzed. From the perspective of researching on the cross-language retrieval platform, reference for cross-language retrieval function design and development of "the Belt and Road" multilingual shared database can be provided. [Method/process] Through literature and network survey, 11 typical cross-language retrieval platforms at home and abroad were

selected. Analysis was carried out from five aspects: cross-language retrieval method, cross-language translation implementation method, retrieval function, retrieval results, interface and retrieval support language. Then concluded their implementation ways. [Result/conclusion] Based on this, strategies are proposed for the cross-language retrieval function design and development of "the Belt and Road" multilingual shared database: adopting query-document translation method based on neural machine translation, implementing multiple retrieval functions, visualization technology used to present retrieval results, providing multi-language interface and resources.

Keywords: "the Belt and Road"database; multilingual; cross-language retrieval

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