From Information Retrieval System Evaluation to Knowledge Service Platform Evaluation (Post-print)

Authors: Li Yuelin, Han Hongliang

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] With the development of knowledge service platforms, how to improve their functionality and design through evaluation to better serve users has become an important issue. This paper aims to explore the evaluation elements and criteria for knowledge service platforms, providing theoretical guidance for their evaluation practice. [Method/Process] Starting from information retrieval system evaluation, this study reviews and organizes relevant research and theoretical models on information retrieval system evaluation and digital library evaluation, as well as the evaluation elements and criteria of digital libraries. [Results/Conclusion] Based on summarizing and generalizing the evaluation elements of previous information service systems, combined with the characteristics of knowledge service platforms, this paper expounds on the evaluation elements and criteria for knowledge service platforms.

Full Text

In early information retrieval system evaluation research, the formal studies were conducted in two distinct phases: the Cranfield 1 test from 1957–1961 and the Cranfield 2 test from 1963–1966. The purpose of Cranfield 1 was to investigate and compare the efficiency of four indexing methods: the Universal Decimal Classification (UDC), an alphabetical subject catalogue, a faceted classification scheme, and the uniterm system of coordinate indexing. The experimental results of Cranfield 1 demonstrated the necessity and feasibility of information retrieval (IR) system evaluation, establishing preliminary methods and models for system assessment. Simultaneously, it revealed the relationship between recall and precision, laying the groundwork for Cranfield 2. Cranfield 2 expanded the sample size, employed retrieval questions that approximated real-world scenarios as closely as possible, and introduced relevance judgments

to obtain more reliable and valid results. In summary, the Cranfield tests, spanning over a decade, left a rich academic legacy and profoundly influenced IR system evaluation, marking a significant milestone in the history of information science development. K.S. Jones [2] summarized the contributions of the Cranfield tests, noting that they demonstrated minimal performance differences among various indexing languages (including natural language), which positively impacted people's understanding of information retrieval systems. More importantly, the Cranfield experiments methodologically established standards for IR experiments, defining the concept of test collections, experimental procedures, and evaluation metrics, thereby laying the theoretical and practical foundation for subsequent system-oriented IR system evaluations. The methodological framework established by Cranfield continues to be widely employed in Text Retrieval Conference (TREC) experiments.

The MEDLARS system evaluation, in contrast to the system-oriented characteristics of the Cranfield tests, exemplified user-oriented IR system assessment. The objectives of the MEDLARS evaluation included: (1) investigating users' search requirements; (2) exploring how efficiently and effectively MEDLARS services met these requirements; (3) identifying adverse factors affecting system performance; and (4) revealing methods for more effectively and economically satisfying MEDLARS users' search requirements, particularly by providing recommendations for more effective utilization of new-generation equipment and programs [3]. The test requests used in the evaluation were derived from real user needs while ensuring representativeness of user search topics, covering various domains included in the MEDLARS system. For the user aspect, twenty medical organizations were selected to form a user group and invited to participate in the evaluation. However, in terms of evaluation metrics, the assessment still focused on recall and precision rates, analyzing in detail the reasons for their failure while noting that interaction barriers between users and the system constituted one of the primary causes of failure. The MEDLARS system evaluation was a rare early assessment project that incorporated users into consideration, particularly its analysis of user-system interaction, which laid the foundation for subsequent user-oriented information retrieval system evaluation practices and development.

The TREC conference, initiated in 1992, represents an important academic conference launched by the U.S. government to promote collaboration among industry, academia, and research institutions. It has since evolved into a global international academic conference, attracting various countries and regions to conduct corresponding research activities. The conference aims to promote and disseminate the application of information retrieval research achievements in society through evaluation, thereby facilitating societal utilization of information and information technology. The first TREC conference was held in November 1992 in Gaithersburg, Maryland. Initially, TREC focused on relatively narrow content, concentrating on text information retrieval. In recent years, with the continuous enrichment of web information resources and the expansion of information retrieval system concepts, TREC has continuously adjusted its themes

to track cutting-edge fields, broaden research scope, and enrich conference content and formats. In addition to traditional documentation, TREC focuses on retrieval research for various information carriers, such as speech and video. Furthermore, TREC addresses non-English document retrieval and cross-language retrieval. Through evaluation, the conference continuously promotes the improvement and advancement of information retrieval technology. The evaluation paradigm and various test collections established by TREC have greatly influenced the development of information retrieval system technology and assessment. Today, TREC has become one of the most important conferences in the information retrieval field.

3. Digital Library Evaluation Theoretical Models

As traditional information retrieval systems evolved toward digital libraries, digital library evaluation became the focus of the IR assessment field. In digital library evaluation, system-oriented assessment still dominates [4]. However, user-oriented digital library evaluation has gained attention for its ability to more effectively support digital library improvement and service capability enhancement. T. Saracevic [5] noted that digital library evaluation lagged behind the vigorous development of digital library construction and practice. Therefore, based on a review of traditional information retrieval system research, he discussed digital library concepts, evaluation content, evaluation criteria, evaluation methods, and reasons for evaluating digital libraries, proposing a framework to help us think about how to evaluate digital libraries [5]. This framework covers evaluation elements and key questions, including: (1) Evaluation concept: What to evaluate? What is a digital library? What does it include? What elements does evaluation comprise? (2) Evaluation environment: Selecting evaluation objectives, frameworks, perspectives, or evaluation levels. What levels does evaluation include? What are the key issues for a particular evaluation level? Finally, what are the purposes of different evaluation levels? (3) Evaluation criteria: Criteria reflect performance related to evaluation objectives. What performance parameters should evaluation focus on? What dimensions and characteristics should be assessed? (4) Measures: Measurement indicators reflect selected criteria to record system performance. What specific measurement indicators should be used for selected criteria? (5) Evaluation methods: What measurement means and tools should be used? How to sample? What procedures should be used for data collection? What methods for data analysis?

The framework proposed by T. Saracevic provides guidance for digital library evaluation theory and practice research. Interaction is an important means for users to access system resources. N. Fuhr et al. [6] proposed a three-dimensional model for digital library evaluation centered on interaction, with the three dimensions referring to the three major interactive elements of user, content, and system, as well as their relationships. They pointed out that digital library evaluation should focus on these three elements and emphasize assessing the relationships among them: usefulness should be the focus of evaluation between

users and content; usability should be the focus between users and systems; and performance should be the focus between systems and content. G. Tsakonas et al. [7] further elaborated on the three-dimensional framework, analyzed the key elements contained in users, systems, and content, and proposed standards and sub-standards for usefulness, usability, and performance evaluation. For usefulness, target attributes and resource attributes are evaluation criteria; usability is evaluated based on effectiveness, efficiency, and user satisfaction; precision, recall, relevance, and response time are the main criteria for evaluating systems. These proposed standards enriched and improved the three-dimensional model for digital library evaluation proposed by N. Fuhr et al., also providing reference for thinking about how to evaluate information service systems.

Based on T. Saracevic's evaluation framework and layered interaction model [8], Y. Zhang [9] conducted digital library evaluation research from the perspective of different digital library stakeholders, proposing a holistic digital library evaluation model. This model emphasizes that the importance of different evaluation criteria varies for different stakeholders—that is, different stakeholders often select different evaluation criteria when evaluating different aspects of digital libraries such as environment, technology, users, interface, content, and services. For example, when evaluating the interface, users only consider effort required and interaction support, whereas when evaluating technology, performance and efficiency are their selected criteria. Y. Zhang's research suggests that holistic evaluation of digital libraries can be considered from different digital library elements while taking into account the needs of different stakeholders.

Building on previous research, Li Yuelin et al. [10] proposed that effective digital library interaction function design is the guarantee of digital library success. Therefore, evaluating digital libraries based on the core element of interaction is necessary. To this end, they integrated tasks into the user interaction model, constructing a three-dimensional interaction model for digital libraries. The three dimensions refer to tasks, digital library support technology, and information resources, with users at the core of interaction. They further explored how different interaction dimensions affect users' evaluation of digital libraries from different perspectives. Through experimental research methods and based on user-digital library interaction, they proposed a digital library interaction function evaluation model, a task-based digital library evaluation model, and a holistic digital library evaluation model. The research found that the theoretical models, ideas, and methods of digital library evaluation can provide reference and 借鉴 for knowledge service platform evaluation.

4. Knowledge Service Platform Evaluation

As T. Saracevic [5] pointed out, the development of digital library evaluation lags behind its practice. Knowledge service platforms have flourished in recent years; however, as mentioned above, research and development on knowledge service platform evaluation have lagged relatively behind. This article attempts to follow T. Saracevic's [5] approach to exploring digital library evaluation to

analyze and construct an evaluation model for knowledge service platforms, starting with "What is a knowledge service platform?" to discuss "Why evaluate knowledge service platforms?" "What to evaluate?" "What are the evaluation criteria?" and "How to apply these criteria in evaluation?"

4.1 What is a Knowledge Service Platform?

Undoubtedly, knowledge service platforms are information service systems that provide knowledge services. Unlike traditional information retrieval systems, search engines, and digital libraries that primarily provide information as their main service content, knowledge service platforms add knowledge services as a core component. Traditional information services focus on information collection, organization, and transmission, with knowledge products typically not within the scope of information service institutions' work. Knowledge services, however, center on providing problem-solving knowledge solutions to help users solve their problems. As Zhang Xiaolin [11] pointed out, information services emphasize "Have I provided the information you need?" whereas knowledge services focus on "Have I solved your problem through my services?" This distinction highlights the difference between information services and knowledge services. Therefore, the core question in knowledge service platform evaluation should differ from that of digital libraries, focusing more on whether the knowledge service platform helps users achieve their original goal of "problem solving."

As mentioned above, knowledge service platforms are, on the one hand, an extension and development of digital libraries—that is, they add knowledge service functions based on digital library information services to upgrade and transform original digital library functions. On the other hand, some knowledge service platforms have established knowledge services as their core function from the outset, adopting a construction policy with knowledge services as primary and information services as supplementary, striving to create new types of information service systems. Currently, knowledge service platforms can be divided into three types: First, enterprise-oriented knowledge service platforms built by enterprises integrating industry resources according to their own status and needs, such as LOTUS, Microsoft, and Tsinghua Tongfang's enterprise-oriented knowledge management platforms, and enterprise knowledge service platforms for technology-based small and medium-sized enterprises. Second, database-type platforms that expand new knowledge service functions based on information service systems and provide knowledge retrieval services based on databases, represented by domestic platforms such as CNKI, Wanfang, and VIP, and international platforms such as Web of Science, Emerald, and Elsevier. Third, encyclopedia and Q&A knowledge service platforms that rely on the Internet for knowledge sharing and serve the general public, such as Wikipedia, Quora, Zhihu, and Dedao.

4.2 Why Evaluate Knowledge Service Platforms?

Like building digital libraries, constructing knowledge service platforms requires substantial investment in construction and maintenance, including hardware and software construction and maintenance, information content purchase and acquisition, and staff recruitment and training. With such significant capital investment, what are its social and economic benefits? Has it achieved the expected results? To answer these questions, evaluating knowledge service platforms is essential. Depending on the nature of the knowledge service platform, for some platforms with public welfare characteristics, social benefits may be a more important aspect requiring focused evaluation.

At the technical level, whether the technical support provided by knowledge service platforms effectively delivers knowledge services is also a question requiring evaluation. With the continuous development and advancement of information technology, big data technology, and artificial intelligence technology, these new technologies, means, and methods will increasingly be utilized in information search, knowledge creation and innovation, and information and knowledge presentation. However, to what extent can these new technologies help us achieve the purpose of knowledge services? Their capabilities in knowledge services require evaluation. The history of information retrieval system and digital library evaluation development reveals that only through evaluation can we identify system and technology deficiencies and problems, providing guidance for further improving knowledge service platform technical support.

Furthermore, the primary service objects of knowledge service platforms are users, including individual and institutional users. How effective are the services provided by knowledge service platforms? Do they meet users' knowledge needs? Can they help users effectively solve problems? Answers to these questions can, on the one hand, help understand whether knowledge service platforms achieve their intended purposes; on the other hand, and more importantly, they can help knowledge service platforms enrich content construction, improve system design, and enhance service quality from the user perspective. The long-term exploration of information retrieval system evaluation theory and practice has spawned user-oriented information system evaluation, establishing a user-centered evaluation philosophy that effectively helps information retrieval systems improve services and interaction with users, contributing to enhanced information retrieval system performance. For knowledge service platform evaluation, while assessing technical capabilities and system performance, focusing on support for user needs through user-oriented evaluation approaches is undoubtedly also important.

4.3 What to Evaluate?

Early information retrieval system evaluation and digital library evaluation undoubtedly laid the foundation for knowledge service platform evaluation. As an information service system, knowledge service platforms share many similarities

with the former two, including elements such as technology, content, interface, service, interaction, and users. Evaluation should also proceed from different elements to achieve partial or holistic assessment of knowledge service platforms.

4.3.1 Technology Knowledge service platforms incorporate many important new technologies based on information service systems, such as using big data technology to improve knowledge service platform operation and management [12]; using grid technology to improve information resource discovery and establish knowledge networks [13-14]; building professional service-oriented knowledge service platforms based on ontology technology [15]; integrating information resources using multilingual information discovery technology and heterogeneous database integration technology; and using push technology, data mining technology, navigation database technology, and intelligent agent technology to analyze user needs and provide personalized and customized services [16]. Integrating so many technologies raises the fundamental question of how to achieve optimal system performance and maximize input-output efficiency, which is a basic problem facing researchers and practitioners and also the significance of evaluation.

4.3.2 Content Content is the core of meeting user needs. To provide solutions to users' problems, knowledge service platforms need to integrate various information resources on the Internet to form knowledge products. Different knowledge platforms have different content emphases due to differences in their target user groups. Knowledge service platforms derived from digital libraries build knowledge warehouses for specific disciplines and specialties based on original functions, construct knowledge networks based on their own and network resources, and provide personalized and customized content according to user characteristics [16]. Public knowledge service platforms serve the general public, offering extensive content covering quality supervision, social security, medical and health care, housing, education, food, transportation, law and finance, employment and entrepreneurship, and other aspects of knowledge, encompassing all facets of daily life and work [17]. Industry-oriented knowledge service platforms provide more targeted and professional content to help professional users solve problems. For example, in the medical industry, knowledge services have broken through traditional medical knowledge base models, forming medical knowledge service platforms represented by Clinical Evidence, DynaMed, Essential Evidence Plus, MDConsult, UpToDate, and ClinicalKey, which serve students, teachers, and both doctors and patients. Based on authoritative medical data, they provide education, patient communication, and problem-solving services while supporting various clinical activities such as disease diagnosis, treatment, disease control, and prevention [18-19]. However, can such rich content meet user needs? Can it help users solve their problems? Can it bring new knowledge or inspiration to users and help them find problem-solving solutions? To answer these questions, conducting knowledge service platform content evaluation is undoubtedly necessary.

4.3.3 Interface User interface has always been a focus of information system design and system evaluation. N.J. Belkin et al. studied four information system interface projects at the European Space Agency and proposed that information system interface design should be oriented toward user goals, tasks, and characteristics [20]. Li Yuelin et al. [10] sorted out and studied evaluation criteria for digital library interfaces, proposing relevance and predictive interaction sub-dimensions that affect user-interface interaction effectiveness, viewing the interface as the main venue for user-technology interaction. Knowledge service platform system interfaces also play the same role, serving as a bridge between users and system content. Therefore, interface functional design is key to whether users can obtain useful knowledge to solve their problems. However, how can we provide users with usable, effective, and easy-to-use interactive interfaces? On the one hand, technical means must be considered; on the other hand, user experience and perception should also be important factors in interface design, and evaluation is an important means of understanding these factors.

4.3.4 Service Service is the core capability and basic requirement of knowledge service platforms. Currently, knowledge service platforms in practical application have different types with distinct service models. Knowledge service platforms derived from traditional digital libraries inherit existing norms and systems in their service models, including autonomous service models where users complete the search process independently, specialized service models, and personalized service models. Autonomous service models mean users complete the search process alone; specialized service models reflect knowledge service functions through reference consultation services, subject services, and portal services; personalized service models push relevant content to users according to their needs and characteristics [21-22]. Internet-born knowledge service platforms such as “Zhihu” and “Dedao” adopt user-based knowledge sharing methods to provide services, offering knowledge services to users through knowledge push, subscription, learning, search, and other functions based on massive data resources [23]. Whether these service models meet user needs and provide users with more convenient, high-quality, and humanized services requires evaluation to answer.

4.3.5 User Users are the target objects of knowledge service platforms. In previous evaluation practices, system evaluation was achieved by measuring user behavior, characteristics, task completion, usage, and acceptance willingness. Li Yuelin et al. [10] proposed relevance and predictive sub-dimensions affecting digital library interaction function evaluation based on user perception. Chang Hong et al. [24] constructed an information push user acceptance model based on the Technology Acceptance Model and Task-Technology Fit Model, and through empirical research, identified factors affecting library information push service user acceptance and proposed improvement suggestions for information push service models. Liu Mingzhu and Yang Jianlin [25] evaluated Sina Weibo search

effectiveness based on seven types of user information needs, identifying deficiencies of Weibo search in retrieving professional websites, knowledge information, and time costs through comparison with Baidu search. Additionally, some studies based on user experience have evaluated health search engines [26], Weibo information quality [27], government website performance [28], and interface usability [29] by measuring user feelings. These studies and previous user-oriented information retrieval system evaluation research provide useful insights for how to conduct user evaluation in knowledge service platform contexts.

4.3.6 Interaction Compared with traditional information systems, interactivity is particularly important in knowledge service platform evaluation. First, scientific and technological development has further expanded human-computer interaction methods. New search methods such as voice search and photo search provide users with more search channels, and information retrieval, browsing, learning, and entertainment based on AR and VR technologies have begun to be applied in practice. The application of new technologies to knowledge service platforms is a trend and inevitability. Second, due to practical demands, academic research on human-computer interaction has continuously strengthened. The shift from a system perspective to a user perspective in evaluation research has deepened, with users becoming the core of system construction, design, and services. Today, users have transformed from passively adapting to actively demanding, and system interactivity directly affects user experience and even usage willingness. Therefore, evaluating interactivity should naturally become an important aspect of knowledge service platform evaluation.

4.4 What Are the Evaluation Criteria?

By drawing on evaluation standards and dimensions from information retrieval system evaluation and digital library evaluation, and combining them with the tasks and purposes of knowledge service platforms, the author constructed evaluation elements, criteria, and measurement indicators for knowledge service platforms as shown in Table 1 :

Table 1. Elements and Criteria for Knowledge Service Platform Evaluation

Evaluation Elements	Evaluation Criteria	Measurement Indicators
Technology	Effectiveness, Efficiency	Average precision, Relative recall, Browsing effectiveness
Content	Usefulness	Problem-solving assistance, New knowledge acquisition, Inspiration and insight

Evaluation Elements	Evaluation Criteria	Measurement Indicators
Interface	Usability	Ease of use, Learnability, Error tolerance, Memorability
Service	Quality	Effectiveness, Efficiency, User satisfaction
User	Task Completion	Completion time, Completion quality
Interaction	Support	Effectiveness, Relevance, Inspiration
Overall	Satisfaction	User satisfaction, Continued use intention, Recommendation intention
	Intelligence	User-friendliness, Effort required

The evaluation elements in Table 1 originate from commonly used elements in information service system evaluation [5, 9]. The evaluation criteria of effectiveness and efficiency are traditional metrics in information retrieval system evaluation; usefulness and usability are important indicators in digital library evaluation [9-10]; task completion degree, interaction support degree, and satisfaction are important indicators in task-based digital library evaluation [10, 30]. The problem resolution degree is determined by the purpose of knowledge service platforms. As mentioned earlier, knowledge services differ from traditional information services, focusing on solving users' problems [11]. Therefore, it is necessary to incorporate measurement standards and evaluation indicators for "problem-solving" degree when evaluating knowledge service platform performance. Furthermore, the focus on service quality originates from the long-standing emphasis on service quality in information service institutions and systems, and transplanting this concept to knowledge service platform evaluation is logical. The intelligence indicator measures the application of artificial intelligence technology in knowledge service platforms and the realization degree of intelligent knowledge services. The inclusion of this indicator responds to the vigorous development of AI technology and applications, and the implantation of intelligent service performance in knowledge service platforms is an inevitable development trend; thus, evaluation should also consider the intelligence degree of knowledge service platforms.

Specific measurement indicators inherit information retrieval system performance evaluation indicators (including average precision, relative recall, browsing effectiveness) and usability evaluation indicators (including ease of use, learnability). Additionally, combining knowledge service platform characteristics and various functional implementation requirements, problem-solving degree, task completion quality, user satisfaction degree, and other

measurement indicators are proposed. For example, knowledge service platform systems emphasize problem solving, thus requiring focus on problem-solving effectiveness, relevance, and inspiration; content usefulness evaluation should also center on problem-solving degree, including the extent of helping problem solving, acquiring new knowledge, and providing inspiration and insight.

These standards and measurement indicators originate from traditional information retrieval system evaluation, digital library evaluation, especially user-oriented system evaluation indicators, combined with interpretation of the purposes and objectives of knowledge service platforms. Whether they can be applied to knowledge service platform evaluation still requires further testing.

4.5 How to Apply These Criteria in Evaluation?

Analogous to digital library evaluation, knowledge service platform evaluation can similarly be divided into two types: holistic evaluation and partial evaluation. Holistic evaluation should include all evaluation elements in Table 1; partial evaluation focuses on one or several components for more in-depth exploration and investigation. These standards and measurement indicators can serve as foundational evaluation metrics for knowledge service platforms, and specific evaluation practices should appropriately adjust them according to the characteristics of evaluation objects, adding, deleting, or refining criteria or measurement indicators based on needs.

The selection of specific measurement indicators should be based on the specific population and scope served by the platform, focusing on specific user tasks and needs, providing solutions for users as the basic orientation, combining the platform's objectives, and considering the scientificity and applicability of evaluation indicators. Simultaneously, we should fully utilize qualitative and quantitative data collection and analysis methods, emphasizing the use of mixed data processing methods to combine qualitative and quantitative data analysis to explain evaluation results. Additionally, the indicator system in Table 1 can be applied to develop knowledge service platform evaluation tools, further determining the effectiveness of evaluation standards and measurement indicators through empirical research, developing evaluation scales, and establishing a complete knowledge service platform evaluation system.

In evaluation methods, experimental research methods are commonly used in information system evaluation. From the Cranfield tests to the TREC conference spanning over 20 years, experimental research methods have played important roles. Questionnaire surveys are also an option. To make evaluation more effective, follow-up interviews after questionnaire surveys are commonly used techniques in evaluation. Whether conducting experimental research or questionnaire surveys, attention should be paid to selecting real users and real contexts. The current trend in system evaluation research is shifting from laboratory experiments to field experiments, aiming to allow users to present the most essential and natural cognitive and behavioral characteristics in the most

authentic contexts and states. In specific operations, tasks can be used to stimulate users' information needs. One approach is to use real work or learning tasks from users' actual lives, which yields authentic and reliable data but has the drawback of diverse user needs that are not conducive to comparison. Another approach is to use simulated work task situations [31], which typically originate from users' actual needs, are standardized and designed by researchers with specific contexts, and have been validated for effectiveness in different studies, making them commonly used task design methods in interactive information retrieval [32-33]. In research, either of these two approaches can be used alone or in combination, depending on research purposes. Furthermore, with the development and application of big data technology, data mining, sentiment analysis, and other technologies have become increasingly mature. How to use the "footprints" left by users during platform usage and interaction—that is, large amounts of data and behavioral trajectories—to conduct evaluation research is a topic worthy of in-depth investigation.

This article starts from the basic viewpoint that knowledge service platforms are information service systems, beginning with early information retrieval system evaluation, reviewing the history and methodology formation of system evaluation, introducing theoretical models, elements, and main standards of digital library evaluation, and discussing evaluation elements and standards for knowledge service platforms based on T. Saracevic's evaluation approach and information retrieval system and digital library evaluation.

Compared with traditional information retrieval system and digital library evaluation, knowledge service platform evaluation demonstrates many differences. Traditionally, we evaluated "information" retrieval systems; today, we have moved to evaluating "knowledge" service platforms. The evaluation has shifted from assessing the efficiency and performance of information retrieval to evaluating the performance and efficiency of knowledge point, knowledge unit, or module retrieval; from evaluating the degree of information retrieval automation to evaluating the degree of knowledge acquisition intelligence; from evaluating fragmented knowledge services in information retrieval systems to evaluating integrated knowledge services in knowledge service platforms; from evaluating universal knowledge provision methods to evaluating more flexible and personalized knowledge provision methods; and from evaluating whether information retrieval systems provide information resources that support users in completing their tasks to evaluating whether knowledge service platforms provide knowledge or solutions that support users in solving task-related problems and completing tasks. These differences contain further improvements and leaps in human information processing and acquisition. How to make these differences more practically meaningful requires continuous evaluation during knowledge service platform development and application practice to provide theoretical and practical support for their continuous improvement and enhancement. From this perspective, knowledge service platform evaluation is undoubtedly crucial, and we still have a long way to go.

References

[1] Hao Meng. Research on knowledge service platform evaluation from the perspective of knowledge construction [D]. Tianjin: Nankai University, 2018.

[2] SPARCK JK. Information retrieval experiment [M]. London: Butterworths, 1981.

[3] LANCASTER FW. MEDLARS: report on the evaluation of its operating efficiency [J]. American documentation, 1969, 20(2): 119-142.

[4] TSAKONAS G, MITRELIS A, PAPACHRISTOPOULOS L, et al. Evaluation of user interaction in digital libraries [OL]. [2018-10-27]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.9642&rep=rep1&type=pdf.

[5] SARACEVIC T. Digital library evaluation: toward an evolution of concepts [J]. Library trends, 2000, 49(3): 350-369.

[6] FUHR N, HANSEN P, MABE M, et al. Digital libraries: a generic classification and evaluation scheme [C]// European conference on research and advanced technology for digital libraries. Berlin: Springer Press, 2001: 187-199.

[7] TSAKONAS G, KAPIDAKIS S, PAPACHRISTOPOULOS C. Evaluating the user interaction in digital libraries [J]. Journal of the American Society for Information Science and Technology, 2013, 64(9): 1914-1926.

[8] SARACEVIC T. Modeling interaction in information retrieval (IR): a review and proposal [J]. Proceedings of the American Society for Information Science and Technology, 1996, 33(1): 3-9.

[9] ZHANG Y. Developing a holistic framework for digital library evaluation [J]. Journal of the American Society for Information Science and Technology, 2010, 61(1): 88-110.

[10] Li Yuelin, Liang Na, Qi Xue. From interaction dimensions to interaction functions: constructing a digital library interaction evaluation theoretical model [J]. Journal of Library Science in China, 2016, 42(1): 66-82.

[11] Zhang Xiaolin. Toward knowledge services: finding growth points for library and information work in the new century [J]. Journal of Library Science in China, 2000, 26(5): 30-35.

[12] Li Chenhui, Cui Jianming, Chen Chaoquan. Research on key technologies for big data knowledge service platform construction [J]. Information and Documentation Services, 2013(2): 29-34.

[13] Wang Zhijin, Yu Xiaoyan. Analysis and evaluation of digital library grid resource discovery methods [J]. Library Journal, 2010, 29(6): 16-20.

[14] Xie Xiaoyang. Grid-based knowledge service platform [J]. Information Studies: Theory & Application, 2006, 29(3): 374-375, 348.

[15] Gao Dongping, Fang An, Li Yang, et al. Design and application of knowledge service platform: taking the major infectious disease information knowledge service platform as an example [J]. Information Studies: Theory & Application, 2011, 34(7): 111-115.

[16] Jiang Yongchang. Research on CNKI digital library knowledge services [J]. Journal of the China Society for Scientific and Technical Information, 2004, 23(3): 265-274.

[17] Ma Jie, Wu Qiong, Qi Lili. Analysis of knowledge service content in public service management platforms [J]. Library and Information Service, 2010, 54(6): 40-43, 101.

[18] Wu Hao, Sun Mengru. Thoughts on the path of knowledge services in China's publishing industry: starting from Elsevier's ClinicalKey super medical platform [J]. Publishing Wide Angle, 2017(13): 13-16.

[19] Yu Tong, Zhang Zhulu, Jia Lirong. Overview of evidence-based medicine-oriented knowledge service platforms [J]. Chinese Journal of Library and Information Science for Traditional Chinese Medicine, 2014, 38(4): 55-57.

[20] BELKIN NJ, MARCHETTI PG, ALBRECHT M, et al. User interfaces for information systems [J]. Journal of information science, 1991, 17(6): 327-344.

[21] Chen Daichun. Analysis of university library knowledge service platform construction [J]. Information Science, 2008, 26(6): 841-844.

[22] He Shaohua, Wang Peilin. Knowledge service models of Chinese libraries under knowledge management environment [J]. Information Science, 2007(5): 619-621, 637.

[23] Li Fanghui, Wang Ling. Exploration of knowledge service platform operation models in the "Internet Plus" era [J]. Library Science Research, 2018(12): 63-67.

[24] Chang Hong, Tian Ye, Sha Shuxin. Research on user acceptance of library information push services [J]. Library Tribune, 2014, 34(11): 106-111.

[25] Liu Mingzhu, Yang Jianlin. Comparative analysis of Weibo search and web search capabilities to meet user information needs: taking Sina Weibo search and Baidu search as examples [J]. Library and Information, 2016(5): 126-136.

[26] Wang Ruojia, Li Yuelin. Usability evaluation of health search engines based on user experience [J]. Library and Information Service, 2016, 60(7): 92-102.

[27] Feng Ying, Zhang Ruiyun. Research on microblog information quality evaluation based on user experience [J]. Library Science Research, 2014(9): 62-67, 101.

[28] Wang Jingxuan, Yang Daoling. Performance evaluation of government websites based on user experience: exploration and practice [J]. E-Government, 2014(5): 35-41.

[29] Chen Congcong. Research on usability evaluation of mobile B2C interfaces oriented to user experience [D]. Chongqing: Southwest University, 2014.

[30] Qi Xue. Research on digital library interaction function evaluation [D]. Tianjin: Nankai University, 2014.

[31] BORLUND P, INGWERSEN P. The development of a method for the evaluation of interactive information retrieval systems [J]. Journal of documentation, 1997, 53(3): 225-250.

[32] Li Yuelin, Xiao Xue, Hu Die. Task design in information retrieval experiments: a comparative study of real and simulated work tasks [J]. Library and Information Service, 2014, 58(16): 5-12.

[33] LI Y, HU D. Interactive retrieval using simulated versus real work task situations: differences in sub-facets of tasks and interaction performance [J/OL]. Proceedings of the American Society for Information Science and Technology, 2013, 50(1): 1-10. [2018-12-06]. http://www.asis.org/asis2013/proceedings/openpage.html.

Author Contributions

Li Yuelin: Responsible for topic selection, paper writing, and revision; Han Hongliang: Responsible for literature collection, organization, and formatting.

Abstract

[**Purpose/significance**] With the development of knowledge platforms, it has become an important issue to improve the function and design of knowledge platforms through system evaluation. The article aims to explore the evaluation elements and criteria for the evaluation of knowledge platforms. [**Method/process**] The article starts with the evaluation of information retrieval. It reviews related studies on the evaluation of information retrieval systems and digital libraries, especially addressing the theoretical models, elements, and criteria of digital library evaluation. [**Result/conclusion**] Based on the review and discussion, the article identifies and addresses the elements and criteria for the evaluation of knowledge platforms.

Keywords: knowledge service; evaluation for knowledge platform; conceptual framework

2018 Excellent Reviewers of *Library and Information Service*

In 2018, nearly 300 reviewers participated in the peer review of manuscripts for *Library and Information Service*, reviewing over 1,400 manuscripts. More than 100 reviewers reviewed six or more manuscripts. Considering the quantity, quality, and timeliness of reviews throughout the year, 50 excellent reviewers for 2018 were selected (listed below). *Library and Information Service* will

issue certificates to the excellent reviewers and provide them with a free one-year electronic subscription to the journal. Thank you to all reviewers for your strong support of *Library and Information Service*!

(The following excellent reviewers are sorted alphabetically by pinyin):

Name	Workplace
Liu Xueli	Xinxiang Medical University Press / Henan Scientific and Technical Journals Research Center
Mao Yihong	Nanjing Agricultural University College of Information Science and Technology
Deng Xiaozhao	Southwest University School of Computer and Information Science
Bai Rujiang	Shandong University of Technology Institute of Science and Technology Information
Feng Jia	Shanghai Academy of Social Sciences Institute of Literature
Ren Shuhuai	Shanghai International Studies University Library
Pei Lei	Nanjing University School of Information Management
Mu Dongmei	Jilin University School of Public Health
Deng Shengli	Wuhan University School of Information Management
Cao Jindan	Jilin University School of Public Health
Liu Ziheng	Peking University Department of Information Management
Qin Hong	University of Electronic Science and Technology Library
Cui Yuhong	Beijing Institute of Technology Library
Ma Jie	Jilin University School of Management
Fang Xiangming	Shanghai University Library
Fan Aihong	Tsinghua University Library
Ma Xueliang	National Library of China
Gan Chunmei	Sun Yat-sen University
Gao Fan	Southwest Jiaotong University Library
Shao Bo	Nanjing University Library

Name	Workplace
Teng Guangqing	Northeast Normal University School of Information Science and Technology
Han Yi	Southwest University School of Computer and Information Science
Wang Lixue	Institute of Scientific and Technical Information of China
Hu Changping	Wuhan University Research Center for Information Resources
Huang Jing	Beijing Normal University School of Government
Huang Guobin	Beijing Normal University School of Government
Huang Linghe	Hebei University School of Management
Wang Xiwei	Jilin University School of Management
Wang Yanfei	Peking University Department of Information Management
Wu Hong	Shandong University of Technology Institute of Science and Technology Information
Wu Zhenxin	National Science Library, Chinese Academy of Sciences
Jiang Chunlin	Dalian University of Technology Faculty of Humanities and Social Sciences, Science and Technology Management
Xiang Guilin	Institute of Biophysics, Chinese Academy of Sciences
Yan Hui	Renmin University School of Information Resources Management
Li Gang	Nanjing University School of Information Management
Li Ming	Nanjing University School of Information Management
Li Wu	Shanghai Jiao Tong University School of Media and Design
Li Shuning	Beijing Normal University Library
Li Zhuozhuo	Soochow University
Liu Bing	Tianjin Normal University School of Management
Yang Xinya	Chongqing University Library

Name	Workplace
Zhang Guangqin	Peking University Department of Information Management
Zhang Pengyi	Peking University Department of Information Management
Zhang Weidong	Jilin University School of Management
Zhao Fei	Peking University Library
Zhao Yuxiang	Nanjing University of Science and Technology
Liu Kan	Zhongnan University of Economics and Law School of Information and Safety Engineering
Zhu Zhongming	Lanzhou Library of Chinese Academy of Sciences / Resources and Environmental Science Information Center, Chinese Academy of Sciences
Liu Chunli	China Medical University Library
Liu Xiaojuan	Beijing Normal University School of Government

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