Assessment of User Knowledge Level Before and After Search and Analysis of Its Changes - Post-print

Authors: Song Xiaoxuan, Liu Chang

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] Information search is a commonly used method for information retrieval. Current search systems provide effective support for fact-finding tasks, but research on search functionalities for learning purposes remains insufficient. "Search as learning" (search as learning) has emerged as a research hotspot in interactive information retrieval in recent years. Such research conceptualizes search as a learning process and attempts to evaluate users' knowledge acquisition during search, subsequently proposing functional optimization recommendations for systems to support user learning. This paper focuses on addressing how to comprehensively assess users' knowledge levels before and after search, providing a reference for such research. [Method/Process] A user experiment methodology was employed to evaluate knowledge content written by users before and after search, proposing a user knowledge assessment method that integrates both quantitative and qualitative dimensions to assess users' knowledge levels before and after searching on learning-oriented tasks. In the data analysis phase, statistical methods were employed to verify differences in users' knowledge levels after versus before search. [Results/Conclusions] The study found that users' performance in knowledge quantity became more comprehensive and in-depth following search completion, with significant improvements in the number of knowledge points, number of knowledge facets, breadth of knowledge facets, and depth of knowledge facets. Simultaneously, knowledge facets with higher degrees of specialization emerged after search. For certain concepts that were relatively vague before search, expression became clearer and more explicit after search. In terms of quality, the vast majority of users demonstrated improvements in knowledge relevance, level of analysis, and articulation of personal viewpoints after search.

Full Text

Preamble

Vol. 62 No. 2 January 2018
Assessment of User Knowledge Level Before and After Searching and Analysis of Its Changes
Song Xiaoxuan, Liu Chang
Department of Information Management, Peking University, Beijing 100871

Abstract

[**Purpose/Significance**] Information search is a commonly used method for information seeking. While current search systems perform well in supporting factual information retrieval, research on search functions for learning purposes remains insufficient. "Search as learning" has become a hot topic in interactive information retrieval research in recent years. These studies view search as a learning process and attempt to evaluate knowledge acquisition during search, thereby proposing functional optimization suggestions to support user learning. This paper focuses on how to comprehensively assess user knowledge levels before and after searching, providing a reference for such research.

[**Method/Process**] Using the user experiment method, this study evaluates knowledge content written by users before and after searching, proposes a user knowledge assessment method that integrates both quantitative and qualitative dimensions, and evaluates user knowledge levels before and after searching for learning-oriented tasks. Statistical methods are employed in the data analysis phase to verify differences in user knowledge levels after versus before searching.

[**Result/Conclusion**] The study finds that users' performance on knowledge quantity becomes more comprehensive and in-depth with the completion of searching, showing significant improvements in the number of knowledge points, number of knowledge facets, breadth of knowledge facets, and depth of knowledge facets. Simultaneously, knowledge facets with relatively high professionalism emerge after searching. Certain concepts that were vague before searching are expressed more clearly and explicitly after searching. In terms of quality, the vast majority of users show improvements in knowledge relevance, level of analysis, and the articulation of personal viewpoints after searching.

Keywords: search as learning, knowledge assessment, knowledge change

2 Related Research

2.1 Exploratory Search and Knowledge Learning

In recent years, "search as learning" has become a research hotspot in interactive information retrieval, with dedicated workshops at major international conferences such as SWIRL 2012, Dagstuhl 2013, the 77th ASIS&T Annual Meeting in 2014, and SIGIR 2016. Regarding search as learning, scholars have proposed

focusing on the user' s learning process during search, the impact of search tasks and other search contexts on user learning, measurement and evaluation methods for learning outcomes, and how to optimize systems to support user learning during search.

Research on knowledge learning during search is closely related to exploratory search tasks. G. Marchionini proposed exploratory search, distinguishing it from fact-finding search tasks. Beyond basic facts and information, exploratory search emphasizes knowledge acquisition, conceptual understanding, explanation and comparison of viewpoints, or aggregation of data concepts. Exploratory search typically refers to situations where users are unfamiliar with the target domain, unclear about search goals, and uncertain about the path to achieve them, making it impossible to directly obtain needed answers from search systems and requiring continuous exploration and attempts with system assistance to eliminate uncertainty. Related research on exploratory search indicates that user interaction behavior characteristics and performance differ significantly from fact-based search. For example, Yuan Hong and Li Qiu found that for question-answering searches, users tend to use the question itself as search terms, whereas in exploratory search, as tasks become more complex, users are more inclined to use query expansion, with more search terms, lower term frequency, and looser connections between terms. Meanwhile, in exploratory search, the number of browsed web pages is greater than in question-answering search, with deeper link depth. Zhang Yunqiu et al. found that in exploratory search, as search topic difficulty increases, average page browsing time decreases, making it harder to concentrate on intensive reading of a particular page. Zhang Yunqiu et al. also attempted to analyze changes in user knowledge structure during exploratory search, finding through pre- and post-search concept maps drawn by participants that searchers' knowledge structure changes significantly during exploratory search tasks, as they attempt to construct new, complex, and more professional knowledge structures.

Exploratory search encompasses a wide range of activities, with learning-related search being one type that aims at learning and requires deeper understanding of the "how" and "why" behind search. K. Collins-Thompson et al. proposed that we should focus on learning-oriented exploratory search, where learning is an important cognitive process in search and learning outcomes are one of the important indicators for evaluating exploratory search effectiveness.

2.2 User Knowledge Level Assessment

Users' pre-search knowledge level is an important user characteristic in information retrieval, with many studies confirming that it affects information behavior and interaction characteristics. Post-search knowledge level is considered an important indicator for evaluating search effectiveness. Therefore, knowledge level assessment is a core issue in studying user search, particularly learning-oriented search.

In information search behavior research, scholars have attempted to propose various methods for evaluating user knowledge levels. For example, using scales for users to self-assess their knowledge level, which is relatively subjective and largely depends on questionnaire validity or the accuracy of personal evaluation. However, evidence shows that users' self-assessment of knowledge level may not be accurate, with cases of overestimation or underestimation. Some researchers use testing methods, having users answer task-related questions before or after searching. These questions often have clear and unique answers, with test scores representing current knowledge level. For instance, W. R. Hersh et al. had medical students answer medical-related questions with clear answers before and after searching, scoring their responses to demonstrate knowledge improvement through information search. L. Nelson et al., when exploring the impact of social tagging on search learning, had users conduct two-hour searches on "Enterprise 2.0 hybrid" tasks and answer 20 true/false questions designed by experimenters afterward, using scores to represent post-search learning.

Additionally, some studies evaluate knowledge level based on the quality of task summaries completed after searching. Task summaries refer to short text reports or answers completed according to task requirements. For example, Y. Kammerer et al. evaluated users' post-search knowledge states based on summaries written during task completion and provided keywords, counting valuable keywords and calculating the number of arguments found and relevant organizations or individuals listed in task summaries. T. Willoughby et al., when exploring the impact of students' domain knowledge on learning effectiveness, required students to complete a short essay on the task topic after searching, evaluating essays by counting acceptable and correct statements or phrases.

The above studies focus on measuring the quantity of knowledge points. M. J. Wilson and M. L. Wilson proposed that knowledge quality is also an important dimension for assessing user knowledge levels. Based on Bloom' s Taxonomy of Educational Objectives in the cognitive learning domain, they established a depth of learning measurement method for manual evaluation of usefulness (4-point scale), analysis (3-point scale), and evaluation (2-point scale) of task summaries written before and after information retrieval, representing knowledge changes brought by search. They also compared traditional task summary processing methods by counting facts and statements and topic coverage, finding that simply counting facts and statements made knowledge growth largely dependent on summary length, and that counting only facts, statements, and topic coverage poorly reflected information quality. Therefore, Wilson and Wilson argued that knowledge measurement should not one-sidedly use any single method such as fact items or facet coverage, but should measure multiple aspects comprehensively to represent users' current knowledge state. This study synthesizes the above evaluation methods, adopting an assessment system combining quantity (knowledge points, knowledge facets, facet breadth, facet depth) and quality (relevance, user opinions, analysis level) to comprehensively demonstrate user knowledge levels.

3 Research Methods

This study primarily employs experimental methods. This section introduces participant details, experimental procedure design, and experimental task design. The user knowledge level assessment section explains the multi-dimensional knowledge level evaluation system and describes the data coding process and results based on this system.

3.1 User Experiment

This experiment recruited 32 participants from Peking University (12 male, 20 female, aged 18-30). They came from different departments: 14 from humanities and social sciences, 5 from information and engineering, 10 from science, and 3 from medicine, distributed across different academic levels: 20 undergraduates (8 freshmen, 3 sophomores, 4 juniors, 5 seniors), 4 master' s students, and 8 doctoral students. During the experiment, participants were free from any restrictions or interference to naturally simulate their real-life search behavior.

3.1.1 Experimental Procedure Design Before the experiment, participants completed a basic information questionnaire and cognitive style test, then proceeded to search four tasks sequentially. Before each task, participants completed a pre-search questionnaire rating task familiarity and difficulty, while describing in text what they already knew about the task and possible search terms. Participants then formally searched the task without time limits, recording and organizing what they considered reasonable answers in a computer notepad file. When participants believed their notepad file had completely recorded the answer, they could stop searching at any time. After searching, participants closed the browser and notepad file, completed a post-search questionnaire rating satisfaction with task completion, post-search task familiarity, richness of online information, and task difficulty, and described their answer to the search task from memory without referring to external information. After completing all four tasks, participants filled out a final questionnaire evaluating their search experience and process.

The experiment used Morae Recorder 3.3 to record user search interaction behavior in the background, including clicked page URLs, time spent on each page, reading time, mouse and keyboard activity, page view counts, relative activity across different pages, query input and modification, and relationships between queries and page activities.

3.1.2 Experimental Task Design This experiment aims to explore characteristics of user knowledge changes before and after searching. Therefore, task design centers on user knowledge learning, using J. Lee' s cognitive learning model as the theoretical basis. J. Lee proposed three new cognitive learning patterns based on Bloom' s Taxonomy: receptive learning, critical learning, and creative learning. Receptive learning refers to memory, understanding, and expression of learned knowledge, corresponding to Bloom' s remembering and

understanding levels, representing relatively shallow learning. Critical learning involves reflection, criticism, and evaluation of knowledge from different sources to form students' own thinking, corresponding to Bloom' s applying, analyzing, and evaluating levels, representing deeper learning than receptive learning. Creative learning corresponds to Bloom' s creating level, requiring the highest learning mode where students generate new ideas or construct new structures.

This experiment selected topics related to daily health concerns of college students, designing receptive learning tasks and critical learning tasks based on receptive and critical learning levels. Each participant searched both types of learning tasks, with each type including two different search topics, totaling four tasks per participant. Task order used a Latin square rotation method to minimize topic order and fatigue effects. Example task descriptions:

(1) Receptive Learning Task: Beijing' s smog is getting worse, and you feel it' s necessary to wear masks during smoggy weather. Before buying a mask, you want to know what types exist, differences between them, and which type is suitable for young people to wear long-term in Beijing. Please search online and record your answers in the notepad during searching.

(2) Critical Learning Task: Your middle school cousin is considering joining the school soccer team. Most relatives agree with this idea, but some argue soccer is a dangerous sport with potential health risks. His parents also want your opinion: What health hazards does long-term soccer pose for teenagers? Should the child join the team? Please search online and record your answers in the notepad during searching.

3.2 User Knowledge Level Assessment

This study evaluates participant knowledge levels from both quantitative and qualitative dimensions based on their written answers to search task questions before and after searching. Quantitative indicators include knowledge points, knowledge facets, facet breadth, and facet depth. Qualitative evaluation indicators include knowledge relevance, user opinions, and knowledge analysis level. Specific evaluation indicators and explanations are shown in Table 1 .

To obtain information for these variables, this study employed a three-step analysis process:

Step 1: Word Segmentation and Part-of-Speech Tagging. The NLPIR Chinese word segmentation system (also known as ICTCLAS2013) by Dr. Zhang Huaping from Beijing Institute of Technology was used for word segmentation and part-of-speech tagging of pre- and post-search answers.

Step 2: Manual Coding of Knowledge Points and Facets. After segmentation, nouns and quantifiers were selected for manual coding of knowledge points. Coding mainly included two aspects: whether the term could be accepted as a knowledge point, and which knowledge facet the point belonged

to. Two coders completed the manual annotation process, with a third independent coder handling inconsistent cases. Knowledge facets were continuously adjusted during coding. Final inter-coder reliability for knowledge points was high (Kappa=0.841 for mask task, Kappa=0.883 for soccer task, both >0.75).

Based on coding results, the receptive learning task (mask task) formed 8 facets: model, material, function, appearance, property, certification, usage, and price. Table 2 lists sample knowledge points for these 8 facets. The critical learning task (soccer task) formed 12 facets: skeletal muscle, brain, internal organs, soft tissue, system regulation, other bodily injuries, physical function exercise, academics, personality, social interaction, interest, and sports protection. Table 3 lists sample knowledge points for these 12 facets.

Step 3: Coding of Knowledge Relevance, User Opinions, and Analysis Level. Similar to knowledge point and facet coding, two coders evaluated all 128 responses (32 participants × 2 tasks × 2 time points) for knowledge relevance, user opinions, and knowledge analysis level according to Table 1 . Inconsistent scores were judged by a third coder. Final inter-coder reliability was Kappa=0.671 for knowledge relevance, Kappa=0.95 for user opinions, and Kappa=0.735 for analysis level.

4 Data Analysis Results

This section compares differences in user knowledge states before and after searching through statistical analysis, examining knowledge quantity (knowledge points, facets, facet breadth, facet depth) and knowledge quality (relevance, user opinions, analysis level). Since knowledge quantity variables are numerical and meet normality and homogeneity of variance requirements, paired-sample t-tests were used to observe differences. Knowledge quality variables are ordinal, so Wilcoxon signed-rank tests were used. Thirty-two participants completed the experiment. As this knowledge assessment and change analysis is exploratory, this analysis focuses on the mask task (receptive learning) and soccer task (critical learning).

4.1 Changes in User Knowledge Quantity Before and After Search

User knowledge levels showed significant differences in knowledge quantity before and after searching. As shown in Table 4 , information search significantly increased knowledge points, facets, facet breadth, and facet depth. Pre-search, users expressed an average of 3-4 knowledge points per task, covering 1-2 facets, with 1-2 points per facet. Post-search, users expressed over 11 knowledge points on average, covering 3-4 facets, with about 3 points per facet. Subsequent verification for both task types found significant improvements in knowledge quantity after searching for both.

Table 4 Differences in User Knowledge Quantity Before and After Search (t-test)

Knowledge Quantity	Pre-search (Mean (SD))	Post-search (Mean (SD))	p-value
Knowledge Points	3.75 (10.1)	11.16 (47.72)	p<0.0001*
Knowledge Facets	1.92 (1.72)	3.75 (3.11)	p<0.0001*
Facet Breadth	0.21 (0.025)	0.40 (0.047)	p<0.0001*
Facet Depth	1.66 (1.35)	2.85 (2.22)	p<0.0001*

Note: p<0.05*

Analysis at the individual user level shows that among 64 sessions (32 users × 2 tasks), 55 sessions increased in knowledge points, 50 sessions increased in facets and facet breadth, and 48 sessions increased in facet depth. Before searching, 10 sessions had users who felt they had nothing to say about the task (zero knowledge points and facets), showing difficulty in knowledge expression. This improved effectively after searching, decreasing to 2 sessions, indicating that for most users, knowledge levels improved in terms of points and facets.

However, some sessions showed unchanged or even decreased numbers of knowledge points and facets after searching. Case analysis revealed that for some users, although the quantity didn' t increase, the content changed. For example, some users mentioned 3 facets both before and after, but post-search they generated new facets while dropping less relevant pre-search facets. Thus, these users still acquired new knowledge through searching, explaining why some users' absolute knowledge quantities didn' t increase.

From a knowledge perspective, facet coverage and changes were analyzed specifically. In the receptive task (mask task), as shown in Figure 5 [Figure 5: see original paper] and Figure 6 [Figure 6: see original paper], the overall facets mentioned didn' t change, but distribution proportions across 8 facet types differed significantly. Pre-search, most users mentioned mask function (21 users, 57 knowledge points), particularly anti-smog and dust-proof functions. Usage, property, and certification were mentioned by 12 users each with about 20 points. However, professionally strong facets like model and material were mentioned by only 5 and 2 users respectively. With deeper searching, more users realized that mask differences relate closely to model and material. While function remained important (27 users), model and material rose to 27 and 20 users respectively, with significant increases in property and usage attention.

In the critical task (soccer), facet changes were more obvious, as shown in Figure 7 [Figure 7: see original paper] and Figure 8 [Figure 8: see original paper]. Post-search, users generated content related to sports protection, brain, personality, and system regulation facets that were absent pre-search. Particularly for sports protection, no user focused on this before searching, but nearly one-third (10 users, 38 knowledge points) advocated for joining the soccer team after searching,

believing proper protection would mitigate risks. Pre-search, 15 users mentioned skeletal muscle injuries and 18 mentioned vague injury concepts like collision and injury. Post-search, vague injuries became more specific, with about 80% of users (25 users) clearly identifying soccer' s harm to skeletal muscles, and related knowledge points increasing significantly from 36 to 113.

Overall, users' knowledge quantity performance became more comprehensive and in-depth with searching completion, with significant improvements across all variables. As task understanding deepened, professionally strong facets emerged (e.g., material and model for masks, sports protection for soccer). Users could also express previously vague concepts more clearly. For example, in the soccer task, pre-search users focused on vague bodily injury expressions because they were unclear about specific injured body parts, while post-search they concentrated on skeletal muscle injuries.

4.2 Changes in User Knowledge Quality Before and After Search

Analysis revealed significant differences in knowledge quality before and after searching. As shown in Table 5 , pre-search mean scores for knowledge relevance, personal viewpoint articulation, and analysis level were 0.84, 0.13, and 0.39 respectively, indicating most users had weak task knowledge, making it difficult to propose viewpoints and conduct deep analysis. Post-search, knowledge relevance and analysis level rose to 1.63 and 1.17 respectively, with over half of users forming personal viewpoints on task-related topics.

Table 5 Differences in User Knowledge Quality Before and After Search (Wilcoxon Test)

Knowledge Quality	Pre-search (Mean (SD))	Post-search (Mean (SD))	p-value
Knowledge Relevance	0.84 (0.45)	1.63 (0.43)	p<0.0001*
Personal Viewpoint	0.13 (0.11)	0.66 (0.23)	p<0.0001*
Analysis Level	0.39 (0.4)	1.17 (0.68)	p<0.0001*

Note: p<0.05*

Figure 9 [Figure 9: see original paper] shows individual user changes. For knowledge relevance, 43 users improved, including 6 from "0-1" , 12 from "0-2" , and 25 from "1-2" . Seventeen users showed no change in relevance, but only 2 ("0-0") and 6 ("1-1") truly didn' t improve in knowledge quality. The remaining 9 ("2-2") had already achieved good pre-search relevance with no room for improvement.

For personal viewpoint articulation, over half (35 users) formed viewpoints post-search ("0-1"), though 21 still didn' t express viewpoints ("0-0"). What

factors influence viewpoint generation and what characteristics users who form viewpoints possess are questions for future research.

For analysis level, 24 users showed no change, including 15 with almost no analysis before or after (“0-0”). Thirty-six users improved in analysis level, including 11 from “0-1” , 18 from “0-2” , and 7 from “1-2” .

Overall, post-search knowledge quality improved significantly over pre-search. While a few users might experience quality decreases (e.g., from having viewpoints to none “1-0” , or from strong to weak analysis “2-1”), this may result from fatigue after expending excessive effort understanding and analyzing large amounts of information.

5 Discussion and Conclusion

This paper evaluates user knowledge levels before and after searching from both quantity and quality dimensions and analyzes changes. Data analysis reveals significant differences between pre- and post-search knowledge levels. Specifically, all variables in the constructed knowledge assessment system—knowledge quantity (knowledge points, facets, facet breadth, facet depth) and knowledge quality (relevance, user opinions, analysis level)—show significant post-search improvements.

According to Wilson and Wilson, this indicates these assessment dimensions are sensitive to real knowledge growth. Unlike Wilson and Wilson’ s study, this research analyzes answers written by users after searching without referring to any web pages or materials, more objectively demonstrating knowledge states in users’ minds rather than information collection and organization during search. Overall, the knowledge assessment system designed in this study effectively measures user knowledge levels and validates from a more specific level that information retrieval processes are often accompanied by knowledge growth.

Specifically, this study finds that users’ knowledge quantity performance becomes more comprehensive and in-depth with searching completion. As task understanding deepens, professionally strong facets emerge (e.g., material and model for masks, sports protection for soccer). Users can also express previously vague concepts more clearly. In terms of knowledge relevance, analysis level, and viewpoint articulation, the vast majority of users improve after searching.

This study provides researchers with a feasible approach to evaluating user knowledge learning in information search from multiple angles. It also identifies specific changes in user knowledge levels before and after searching. Future research will further explore what factors affect the degree of knowledge level change, such as task type, personal characteristics, interaction behavior, or other contextual factors.

References

[1] RIEH S Y, COLLINS-THOMPSON K, HANSEN P, et al. Towards searching as a learning process[J]. Journal of information science, 2016, 42(1): 19-34.

[2] GWIZDKA J, HANSEN P, HAUFF C, et al. Search as learning (SAL) workshop 2016[C]//Proceedings of International ACM SIGIR conference on research and development in information retrieval. New York: ACM, 2016: 1249-1250.

[3] RIEH S Y, GWIZDKA J, FREUND L, et al. Searching as learning: novel measures for information interaction research[J]. Proceedings of the American Society for Information Science & Technology, 2014, 51(1): 1-4.

[4] MARCHIONINI G. Exploratory search: from finding to understanding[J]. Communications of the ACM, 2006, 49(4): 41-46.

[5] ZHANG Y, AN W, YU S. Experimental study on user cognition in exploratory search[J]. Information Studies: Theory & Application, 2013, 36(6): 73-77.

[6] ALLAN J, CROFT B, MOFFAT A, et al. Frontiers, challenges, and opportunities for information retrieval: report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne[J]. Comptes rendus des séances de la société de biologie et de ses filiales, 2012, 161(12): 2355-2359.

[7] AGOSTI M, FUHR N, TOMS E, et al. Evaluation methodologies in information retrieval dagstuhl seminar 13441[J]. ACM SIGIR forum, 2014, 48(1): 36-41.

[8] WHITE R W, ROTH R A. Exploratory search: beyond the query-response paradigm[M]//Synthesis lectures on information concepts, retrieval, and services. Williston: Morgan & Claypool Publishers, 2009: 1-98.

[9] YUAN H, LI Q. Impact of search tasks and search ability on user exploratory search behavior[J]. Library and Information Service, 2015, 59(15): 94-105.

[10] ZHANG Y, AN W, FENG J. Research on exploratory information search behavior[J]. Library and Information Work, 2012, 56(14): 67-72.

[11] EICKHOFF C, TEE VAN J, WHITE R, et al. Lessons from the journey: a query log analysis of within-session learning[C]//Proceedings of ACM conference on Web search and data mining. New York: ACM, 2014: 223-232.

[12] COLLINS-THOMPSON K, RIEH S Y, HAYNES C C, et al. Assessing learning outcomes in Web search: a comparison of tasks and query strategies[C]//Proceedings of ACM conference on human information interaction and retrieval. New York: ACM, 2016: 163-172.

[13] LAZONDER A W, BIEMANS H J A, WOPEREIS I G J H. Differences between novice and experienced users in searching for information on the World Wide Web[J]. Journal of the American society for information science & technology, 2000, 51(6): 576-581.

[14] SAITO H, MIWA K. A cognitive study of information seeking processes in the WWW: the effects of searcher' s knowledge and experience[C]//Proceedings of the Second international conference on Web information systems engineering (WISE' 01) IEEE Computer Society. Washington: IEEE Computer Society, 2001: 321-327.

[15] WHITE R W, DUMAIS S T, TEE VAN J. Characterizing the influence of domain expertise on Web search behavior[C]//Proceedings of the second ACM international conference on Web search and data mining. New York: ACM, 2009: 132-141.

[16] KOROBILI S, MALLIARI A, ZAPOUNIDOU S. Factors that influence information-seeking behavior: the case of Greek graduate students[J]. Journal of academic librarianship, 2011, 37(2): 155-165.

[17] LIU C, ZHANG L, SONG X. Are self-assessment of search ability and performance reliable?[C]//Proceedings of the 79th ASIS&T annual meeting: creating knowledge, enhancing lives through information & technology. Maryland: American Society for Information Science Silver Springs, 2016: 87.

[18] HERSH W R, ELLIOT D L, HICKAM D H, et al. Towards new measures of information retrieval evaluation[C]//International ACM SIGIR conference on research and development in information retrieval. New York: ACM, 1995: 164-170.

[19] NELSON L, HELD C, PIROLLI P, et al. With a little help from my friends: examining the impact of social annotations in sense-making tasks[C]//Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM, 2009: 1795-1798.

[20] KAMMERER Y, NAIRN R, PIROLLI P, et al. Signpost from the masses: learning effects in an exploratory social tag search browser[C]//Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM, 2009: 625-634.

[21] WILLOUGHBY T, ANDERSON S A, WOOD E, et al. Fast searching for information on the Internet to use in a learning context: the impact of domain knowledge[J]. Computers & education, 2009, 52(3): 640-648.

[22] WILSON M J, WILSON M L. A comparison of techniques for measuring sensemaking and learning within participant-generated summaries[J]. Journal of the American Society for Information Science & Technology, 2013, 64(2): 291-306.

[23] BLOOM B S, ENGELHART M D, FURST E J, et al. Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive domain[M]. London: Longmans, 1956: 58.

[24] LEE J. Does higher education foster critical and creative learners? An exploration of two universities in South Korea and the USA[J]. Higher education research & development, 2015, 34(1): 1-16.

[25] LIU C, ZHAO Y, YANG F. Research on the influence of time limits and task order in information retrieval user experiment design[J]. Library and Information Service, 2015, 59(1): 99-105.

[26] LIU J, BELKIN N J, ZHANG X, et al. Examining users' knowledge change in the task completion process[J]. Information processing & management, 2013, 49(5): 1058-1074.

Author Contributions

Song Xiaoxuan: Responsible for data analysis and paper writing.
Liu Chang: Responsible for research design and determining the overall article structure.

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