

RESEARCH PAPER

How Do People Make Relevance Judgment of Scientific Data?

Jianping Liu^{1,2}, Jian Wang^{1,2}, Guomin Zhou^{2,3}, Mo Wang^{1,2} and Lei Shi⁴

¹ Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, CN

² Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing, CN

³ Department of Science and Technology Management, Chinese Academy of Agricultural Sciences, Beijing, CN

⁴ National Science and Technology Infrastructure Center, Ministry of Science and Technology of the People's Republic of China, Beijing, CN

Corresponding author: Jian Wang (wangjian01@caas.cn)

Many efforts have been made to explore user relevance judgment for documents, images, web pages and music in the field of information retrieval. However, there is a lack of attention to scientific data even when scientists and researchers are facing an increasing data deluge. In this study, we carried out a two-phase (first exploratory and then empirical) research to explore relevance judgment patterns of scientific data users. In the exploratory study, we interviewed 23 subjects who participated in a national competition related to scientific data. Five relevance criteria (RC) and seven paths of their usage were identified by content analysis of the transcribed records of the interview. Based on the results of the first phase, seven hypotheses were proposed and verified by partial least squares structural equation modelling (PLS-SEM). The study identified five RC, i.e. topicality, accessibility, authority, quality and usefulness used by scientific data users. Three patterns were identified including 1) data topicality judgment as the first step or starting point, 2) data reliability judgment as the necessary process and 3) data utility judgment as final purpose. These findings provide new understanding of relevance judgement and behaviours of scientific data users, and could benefit the design for cognitive retrieval systems and algorithms specific to scientific data.

Keywords: scientific data; data retrieval; user relevance; relevance criteria

1 Introduction

With continuous development of data-intensive research and open science, along with remarkable progress of data acquisition technology, there is an increasing demand of scientific data sharing. Hence, finding relevant data within massive scientific data repositories is an urgent need of scientific data users. Such need calls for a scientific data-specific retrieval technology designed and developed based on understanding of how people make judgments on the relevance of scientific data. Data sharing and discovery depend on the development of infrastructure, support systems and data supplies (Borgman, 2013), “it is equally important to understand the behaviours involved in data retrieval. But a user-focused analysis of data retrieval practices is lacking” as Gregory et al. (2017: 1) pointed out. Until now, however, such understanding and the underlying study are still out of the main research interests of the fields of information retrieval (IR) and data sciences.

User relevance has been acknowledged as a basic concept in information science because it contributes to the explanation of ubiquitous behaviours involving information selection and utilization (Borlund, 2003; Ingwersen & Järvelin, 2011; Saracevic, 2016). Many researchers have utilised the concept by investigating various forms of information such as scientific papers or reports, web pages, pictures and images, multimedia, and scientific data (Barry, 1994; Xu and Chen, 2006; Crystal and Greenberg, 2006; Savolainen, 2013; Choi, 2002; Inskip, 2010; Liu et al., 2019). The goal of all these studies is to upgrade traditional IR to be more

interactive, cognition-friendly and highly effective. For scientific data, such upgrade seems more challenging. For example, Google dataset search¹ as a representative data search system follows the traditional IR mechanism and cannot respond to particular requirements of data searchers due to their intensive concerns of data quality and credibility.

1.1 Problem statement

The understanding of how and why people select one dataset rather than another, or specifically, how and why people judge the relevance of a certain dataset, should be an important prerequisite for data retrieval. The study aims at addressing the problem by answering the following three questions:

- What relevance criteria (RC) do scientific data users use to judge relevance, and how do users combine those RC to make the final judgment?
- Can the structure of RC combination be verified by PLS-SEM? If yes,
- Are there any proprietary patterns of relevance judgment that can be summarized from the structure?

2 Literature Review

2.1 Scientific data

Scientific data is a subcategory of data, also known as research data, which is defined as “recorded factual material commonly retrained by and accepted in the scientific community as necessary to validate research findings...”² by Engineering and Physical Sciences Research Council (EPSRC), which is the main funding body for engineering and physical sciences research in the UK. Borgman (2013: 29) defined scientific data as “entities used as evidences of phenomena for the purposes of research or scholarship”. These definitions reveal the essential characteristics of fact-bearing of scientific data and the function of scientific data as “evidences”.

Scientific data is generally considered as fact-bearing information, while documents, the traditional objects of relevance study in IR, are regarded as knowledge-bearing information. Such distinctive feature leads to many differences between scientific data and documents such as formats, forms and communication characteristics etc. Among those the most important is that they have different functions. Data is generally regarded as evidences for reasoning and deciding of scientific data users, while information in documents usually carries various knowledge potentially modifying the receivers' state of cognition. Such differences lead to the necessity of research on how scientific data users make relevance judgment by using different relevance criteria (relevance criteria study will be discussed in section 2.3).

2.2 Scientific data retrieval

The essential feature of fact-bearing of scientific data determines the necessity of developing proprietary scientific data retrieval systems and algorithms. Although information retrieval (IR) has been developed for more than 60 years, data retrieval is still a nascent field (Sanderson & Croft, 2012; Gregory et al., 2017), especially in user-oriented data retrieval. Some progress has been made in scientific data sharing and discovery. Google released Google dataset search in 2018. This has a dataset search engine similar to Google search but was released almost 19 years later. The World Data System (WDS)³ integrated data from over 100 independent data centres to support dataset retrieval. Quandl⁴ focuses on the search of financial and social science data sets. China built 23 scientific data sharing systems aiming at curation and sharing of scientific data in various fields, such as National Earth System Science Data Sharing Infrastructure,⁵ China Earthquake Data Centre,⁶ National Agricultural Scientific Data Centre.⁷

However, at present, these systems are all developed based on system-oriented retrieval methods. System relevance, or algorithm relevance, is a typical objective relevance depended on a given procedure or algorithm without considering the principal position of end users. For example, the relevant items are ranked by calculating the level of term matching between information objects and queries based on a vector space model. If

¹ <https://toolbox.google.com/datasetsearch>.

² <https://epsrc.ukri.org/about/standards/researchdata/scope/>.

³ <http://www.icsu-wds.org/services/data-portal>.

⁴ <http://www.quandl.com>.

⁵ <http://www.geodata.cn/>.

⁶ <http://data.earthquake.cn/>.

⁷ <http://www.agridata.cn/>.

relevance is not objective, then, how do people make relevance of information objects? Therefore, researchers studied the relative concept of system relevance or objective relevance—user relevance, or subjective relevance.

2.3 User relevance

The study on user relevance focuses on describing, interpreting and modelling user relevance judgment process of information objects (Schamber et al., 1990; Cosijn & Ingwersen, 2000; Harter, 1992; Saracevic, 1975), as well as providing new design requirements and directions for IR practice. From the perspective of measurement, user relevance study focuses on the identification and use of RC. Schamber, Eisenberg, and Nilan (1990: 771) stated that “an understanding of relevance criteria, or the reasons underlying relevance judgment, as observed from the user’s perspective, may contribute to a more complete and useful understanding of the dimensions of relevance”.

2.3.1 Identification of relevance criteria

RC identification is the premise and foundation of user relevance judgment research. Judging by existing user RC identification studies, the main results are summed into three points. First, researchers have verified that user relevance judgment not only depends on topicality, but also considers other RC (quality, authority, novelty, accessibility, etc.) by using methods of situational experiments and user interviews (Barry, 1994; Wang and Soergel, 1998; Bateman, 1998). Second, researchers have expanded the research scenarios from relevance judgment of documents to other information types such as images, music, and web pages (Laplante, 2010; Markkula, 2000; Tombros, 2003; Savolainen, 2013; Choi, 2002; Inskip, 2010), which were all prominent information types at that time. Third, user’s RC for inferences are fairly stable, in which a cross-situational RC set exists (Barry and Schamber, 1998; Wang and Soergel, 1998). The differences among RC are manifested in the differences of RC usage according to varying situations and information objects (Saracevic, 2016).

2.3.2 Usage of relevance criteria across different information carriers

The use structure of RC reflects specific information behaviour of users. Greisdorf (2000) proposed the conjunctive and disjunctive rules of RC use in documents relevance judgment. For example, for conjunctive rules, users make relevant decisions based on positive aspects of RC. Xu and Chen (2006) empirically concluded that topicality and novelty were the two most important RC for documents relevance judgment. On this basis, they suggested four types of document retrieval modes to IR practice.

Researchers also explored the RC use of other information types. Choi (2002) found that topicality still played the most fundamental role in image relevance judgment, but image quality and clarity were the most frequently used RC. Crystal and Greenberg (2006) studied RC use modes of health-related web pages of web browsing users in different IR stages. Laplante (2010) and Inskip et al. (2010) studied RC use modes in music relevance judgment. They found that topicality is the most important criterion, but music users use criteria like personal hobbies, personal needs and novelty frequently.

3 Research Framework

Based on the perspective of user relevance research and the essential characteristics of fact-bearing of scientific data, this study carried out a two-phase study. In the first phase (see details in Section 4), by using the methods of situational interviews and content analysis, we conducted exploratory study focusing on what RC do scientific data users use and how they use RC to make relevance judgment? In the second phase of empirical study (see details in Section 5), seven research hypotheses were proposed based on the results of the first phase and the hierarchical structure of user information need provided by Taylor (see details in 5.1). The hypotheses were verified by using PLS-SEM (see details about this method in Section 5.3). The purpose of this study is to explore the proprietary relevance judgment patterns of scientific data users, and then to provide guidance for the development of user-oriented scientific data retrieval systems and algorithms.

4 Phase 1: Exploratory Study

4.1 Data collection

4.1.1 Subjects

The subjects in this study are participants from a national competition (Innovation Competition of Science and Technology Resources Sharing Service for College Student, “Sharing Cup” for short) of scientific data in China. The competition is a national science and technology activity aimed at promoting the reuse and

efficiency of scientific data. Competitors are presumed to submit works in the form of research papers, multimedia presentation, website systems design, business plans, etc., all should base on given scientific data and topics provided by the 23 scientific data sharing platforms.

This study investigated the competitors from the fifth “Sharing Cup” (May 2017–December 2017). By sending emails to the competitors, 23 volunteer competitors were selected as subjects for the interview experiment in the exploratory study. The 23 subjects all used the data from the scientific data sharing platform and completed the competition works before the interview. There were 5 undergraduate students and 18 postgraduate students among these 23 subjects.

4.1.2 Data collection process

Twenty-three subjects were interviewed face-to-face in laboratory or video-meeting. The interviewers were two doctoral students who participated in the whole interview process cooperatively. Semi-structured situational interview method was used to collect data. Subjects answered questions (see Appendix A part one) related to long-term memory by recalling and questions (see Appendix A part two) related to real retrieval scenarios by showing how they judge one dataset relevant or not relevant. The whole interview process was recorded and videotaped. It generally took 30–60 minutes to complete each interview conversation and subjects were given about 50 Ren Min Bi (RMB) as reward for their participation.

4.2 Data analysis process

4.2.1 Content analysis method

The data collected by interview were transcribed into texts. Content analysis (CA) method was applied for this purpose. CA is a method for analyzing written, oral or visual communication messages, to construct a conceptual model to describe the research phenomenon (Downe-Wamboldt, 1992). The core process of CA was divided into three steps: the determination of analysis units, the development of categories and the construction of the relationships of categories. In this study, the coding units are the sentences from scientific data user's descriptions in interviews, while the categories are the RC and corresponding clues mentioned by scientific data users, and the relationships of categories are the different combination usages of RC in data relevance judgment context. The transcribed texts were coded by three coders by using Nvivo11,⁸ and the coding coincidence rate (C.R.)⁹ reached 82%, which was greater than the minimum threshold of 60% (Perreault & Leigh, 1989).

4.2.2 Coding for relevance criteria and corresponding clues

Table 1 shows the main process of extracting key concepts from transcribed texts based on content analysis. As shown in **Table 1**, RC and corresponding stimulus clues were the main concepts extracted in this study. Clues were information features or attributes perceived by users, reflecting the connotation of RC. RC is the “cognitive tool” on which users rely for relevance judgment, and it also represents a certain level of judgment made by users (e.g. to judge the relevance of scientific data, user may judge the authority of the data by taking the producers and affiliation of the authors as the stimulus clues).

4.2.3 Coding for relevance criteria usage paths

According to the interviews, users use a combination of multiple RC, instead of using one RC, to judge the relevance of scientific data. The combination of RC reflects the Scientific data user's relevance judgment patterns. The coding process is shown in **Table 2**.

Table 1: RC and corresponding clues coding examples.

Interview process	Clues	RC
Q1: What is your basis for judging the relevance of data in completing this task?		
A1: Mainly focusing on data keyword	Data keywords	
Q2: So what is the role of data keywords?		
A2: I often use the keywords to determine whether it is the topic I want		Topicality

Note: Q = question from interviewer; A = answer from subject.

⁸ <https://www.qsrinternational.com>.

⁹ C.R. = $3M / (N_1 + N_2 + N_3)$, M = the number of coincident units from three coders, N_n = the number of units from each coders.

4.3 Results

4.3.1 Data clues and corresponding relevance criteria

Table 3 summarizes the coding results of RC and corresponding clues. Five RC (topicality, accessibility, quality, authority and usefulness) and 18 corresponding clues were coded. As shown in **Table 4**, each

Table 2: RC usage paths coding examples.

Interview process	Paths of criteria use
Coding based on direct answers	
Q1: How do you judge the relevance of scientific data?	
A1: First, I judge the topic based on the data keywords, and then check the quality of the data. If the data quality is satisfying, it will be relevant.	Topicality→quality
Coding based on different answers in context	
Q1: How do you judge the relevance of scientific data?	
A1: It is based on data keywords	
Q2: Do you rely solely on data keywords?	
A2: No, I still need to see the data production organization and whether the data can solve my current task."	Topicality→authority→usefulness

Note: Q = question from interviewer; A = answer from subject.

Table 3: The coding results of RC and corresponding clues.

Clues	RC	Freq.	Resp.
	Topicality (TO)	325	20
Data Title (DT)		107	19
Data keywords (DK)		123	20
Data description (DD)		60	14
Data time scope (DTS)		35	12
	Accessibility (AC)	268	19
Data acquisition channel (DAC)		88	19
Data sharing level (DSL)		74	19
Support download? (DSD)		95	19
Data size (DS)		11	8
	Authority (AU)	135	17
Data producer (DP)		44	13
Organization of data producer (DODP)		35	13
Data supply platform (DSP)		56	15
	Quality (QU)	123	16
Data quality illustration (DQI)		54	16
Data producing and processing methods (DPPM)		67	16
Data Searching ranking order (DSRO)		3	2
Data Visiting volume (DVV)		2	2
	Usefulness (US)	293	19
US1: Scientific data as research evidences		44	15
US2: Scientific data can verify research theories		52	19
US3: Scientific data is the basis of my research		68	20

Note: Freq. = number of coding reference nodes; Resp. = number of subjects.

criterion was given a clear definition to fit scientific data research context as one of the important results of exploratory study.

4.3.2 Relevance criteria use paths

Table 5 summarizes the coding results of the RC usage paths. Seven types of RC use paths were coded. The seven usage paths were all started with topicality and combined with other RC. Usefulness, as the user's overall perception of data relevance, was influenced by other RC.

5 Phase 2: Empirical Study

5.1 Research model and hypotheses

The results of phase 1 showed that user's judgment on scientific data relevance do not depend on just one CR, nor can the final decision be made for the first time. It is an interactive process, in which users form different levels of questions and make differential levels of relevance judgment before reaching the final decision.

Firstly, it involves user information question formation process. As Taylor (1968: 182) stated that: "There are four levels of question formation that shade into one another along the question spectrum in user information retrieval". Different levels of questions reflect different needs for information. The original definitions of four questions proposed by Taylor (1968: 182) are as follows:

- “Q₁—the actual, but unexpressed need for information (the visceral need);
- Q₂—the conscious, within-brain description of the need (the conscious need);
- Q₃—the formal statement of the need (the formalized need);
- Q₄—the question as presented to the information system (the compromised need)”

Secondly, it involves different levels of user information relevance judgment. Corresponding to the question spectrum provided by Taylor, the results (**Table 5** of phase 1) showed that users combined different levels of RC to make the final judgment. For example, when a user's information need state changes from Q1 to Q2, the user mainly determines the query and retrieval topics, in which he/she makes topic relevance judgment. When a user's information need state changes from Q2 to Q3, the user understands and infers information based on various aspects of information content, in which he/she makes relevance judgment from the perspective of different aspects (such as quality and authority judgment of information). Finally, when a

Table 4: Definitions of scientific data RC.

Criteria	Definition
Topicality	The consistency between the topic perceived by users and the topic expressed by the data themselves.
Accessibility	The external restriction of the data.
Authority	The source of the data is reliable.
Quality	The data meet the requirements in terms of precision, accuracy, verifiability, etc.
Usefulness	Users perceive the utility of scientific data to solve problems in situations.

Table 5: The coding results of RC use paths.

RC use paths	Mentions	Percent	Respondents
TO → AC	96	20.3	19
TO → QU	65	13.8	19
TO → AU	64	13.6	17
TO → US	58	12.3	17
TO → AC → US	31	6.6	12
TO → QU → US	23	4.9	9
TO → AU → US	39	8.3	16

user's information need state changes from Q3 to Q4, the user perceives relevance according to whether the information can solve the problem in the situation, in which he/she makes situation relevance judgment (judge the usefulness of information).

Considering the above two findings, the empirical study proposed the research model to be verified as shown in **Figure 1**. The model expresses two types of relations to be verified. First, the relationship between clues and corresponding RC need to be verified. Clues are information attributes or characteristics that reflect the connotations of RC (for example, as shown in **Table 3**, DT, DK, DD, DTS are the clues that reflect the connotation of topicality). Second, the relationships among scientific data of RC need to be verified. Based on the results of the exploratory study (as shown in **Table 5**), it is assumed that topicality, as a pre-requisite RC, has positive effects on data quality, authority, accessibility and usefulness judgment (H1–H4), while data quality, accessibility and authority judgment have positive effects on the final judgment of data usefulness (H5–H7). The specific hypotheses are described in H1–H7:

H1–H4: DT, DK, DD, and DTS reflect the connotation of topicality which has a positive effect on data quality, authority, accessibility and usefulness judgment

H5: DQI, DPPM, DSRO, and DVV reflect the connotation of data quality which has a positive effect on data usefulness judgment

H6: DP, DODP, and DSP reflect the connotation of data authority which has a positive effect on data usefulness judgment

H7: DAC, DSL, DS, DSD and reflect the connotation of data accessibility which has a positive effect on data usefulness judgment

5.2 Data collection

5.2.1 Subjects

The subjects of the empirical study also came from the Fifth "Sharing Cup". The subjects were presented with the same competition task. In the empirical study, 564 subjects participated in the questionnaire survey, and 544 valid questionnaires were finally used (see Section 5.4.1 for detailed demographic information).

5.2.2 Data collection process

Based on the results of the exploratory research, the corresponding questionnaire was designed in the empirical research (see Appendix B). The subjects scored each measurement variable according to its importance using a six-level scale – the importance increases continuously from zero (never pay attention) to five (very important).

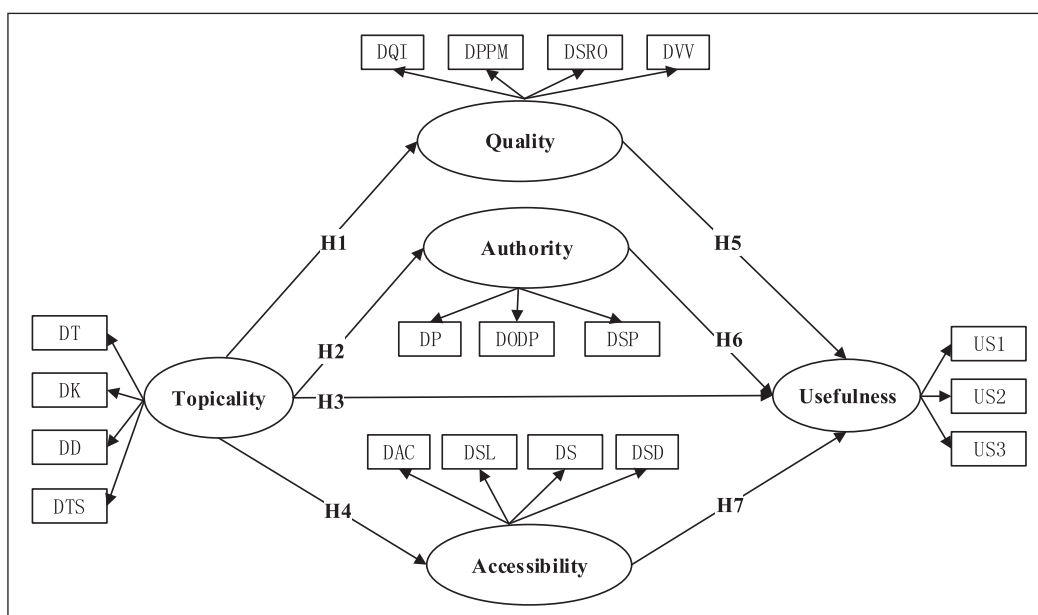


Figure 1: Research model.

5.3 Data analysis process

A strict psychological measurement method, structural equation model (SEM), was used in this study. Anderson and Gerbing (1988) proposed this method to develop and verify theoretical assumptions. As an effective psychometric analysis method, SEM has been widely used in behavioural science, marketing, education and other fields (Garson, 2016). The analysis process of SEM was divided into two steps: measurement model and structural model analysis. The measurement model was used to verify the structural stability between the measurement index and the latent variable. For example, whether DT can be a measurement index of topicality needs to be verified. The structural model was used to verify the stability of the relationship between latent variables. For example, whether data quality judgment has impact on the data usefulness judgment needs to be verified.

There are two types of SEM, covariance-based (CB-SEM) and variance-based partial least squares (PLS-SEM). CB-SEM follows a maximum likelihood (ML) estimation procedure and aims at reproducing the covariance matrix without focusing on explained variance (Hair et al., 2011). Whereas PLS-SEM uses a regression-based partial least squares estimation method with the goal of explaining the latent constructs' variance by minimizing the error terms (Hair et al., 2014). The two methods are complementary with each other. The most important reason to select CB-SEM or PLS-SEM is the research goal or research context. Hair et al. (2011: 144) recommended:

- “If the goal is predicting key target constructs or identifying key ‘driver’ constructs, select PLS-SEM.
- If the goal is theory testing, theory confirmation, or comparison of alternative theories, select CB-SEM.
- If the research is exploratory or an extension of an existing structural theory, select PLS-SEM.”

Accordingly, this study aims at verifying the RC using structure: topicality as the driver construct, quality/authority/accessibility as intermediary constructs, and usefulness as the target construct. And this study is also an exploratory study that first adopts PLS-SEM in RC using structure. Therefore, we finally chose PLS-SEM and employed the SmartPLS¹⁰ as analysis tools.

5.4 Results and findings

5.4.1 Demographic information

This study received 544 valid questionnaires (excluding 20 invalid questionnaires), with the recovery rate of 96%. The gender ration of the subjects was balanced (M = 49.5%, F = 50.5%), the majority of subjects were postgraduate students (postgraduate = 95.6%, other = 4.4%), and the age range was mainly 18–30 (18–30 = 91.4%, other = 8.6%). In the aspect of user's familiarity with scientific data, 84% of the subjects participated in at least one data-related research project, and for 92% of the subjects, their scientific data retrieval time accounted for more than 20% of the time of IR.

5.4.2 Measurement model

The measurement model verifies the structural validity of the construction. Structural validity tests the internal consistency, convergence validity and discrimination validity of construction. In this study, SmartPLS3 was used to evaluate the structural validity of the measurement model (Ringle, Wende & Becker, 2015). Cronbach's alpha (α) and composite reliability (C.R) are important indicators to measure internal consistency. In confirmatory research, the threshold of C.R, Cronbach's alpha (α) and standardized loading (SL) are required to be greater than 0.7 (Hair, Ringle & Sarstedt, 2011). Convergence validity is verified by average variance extracted (AVE). AVE should be greater than 0.5 in the confirmatory study. As shown in **Table 6**, C.R., SL, α and AVE all meet the above requirements.

The discrimination validity is verified by Fornel-Larcker-Criterion (Fornell & Larcker, 1981). **Table 7** shows that if the top value (square root of AVE) in each column is greater than other values in that column, the discrimination validity is positive. As shown in **Tables 6** and **7**, the measurement model in this study meets all requirements. The clues (data attributes) are effective as measurement index of corresponding RC (as shown in **Table 6**, all are significant at the level of $p < 0.001$).

5.4.3 Structural Model

The structural model tests the research hypotheses to interpret the prediction ability of the model. As shown in **Figure 2** of the structural equation model verified in this study, the path coefficients are all normalized

¹⁰ <https://www.smartpls.com/>.

Table 6: Reflective measurements.

RC	Clues	Mean	SD	SL	AVE	C.R	α
Topicality					0.545	0.826	0.719
	DT	4.412	1.362	0.657***			
	DK	4.756	1.190	0.816***			
	DD	4.579	1.281	0.784***			
	DTS	4.524	1.309	0.687***			
Quality					0.534	0.820	0.708
	DQI	4.634	1.322	0.811***			
	DPPM	4.211	1.354	0.720***			
	DSRO	4.022	1.411	0.691***			
	DVV	4.110	1.459	0.694***			
Authority					0.670	0.859	0.752
	DP	3.761	1.355	0.750***			
	DODP	3.671	1.409	0.860***			
	DSP	3.998	1.389	0.842***			
Accessibility					0.546	0.827	0.720
	DAC	4.278	1.281	0.765***			
	DSL	3.991	1.429	0.769***			
	DS	3.404	1.346	0.622***			
	DSD	4.881	1.308	0.788***			
Usefulness					0.591	0.812	0.650
	US1	4.233	1.332	0.692***			
	US2	4.237	1.384	0.814***			
	US3	3.803	1.373	0.796***			

Note: *** Significant at 0.001 (two-tailed); SL = standardized loading; C.R = composite reliability; α = Cronbach's alpha; AVE = average variance extracted.

Table 7: Fornell-Larcker-Criterion.

	Latent Variable Correlations(LVC)					Discriminant Validity met? (Square root of AVE>LVC?)
	AC	AU	QU	TO	US	
AC	0.739					Yes
AU	0.626	0.819				Yes
QU	0.684	0.596	0.731			Yes
TO	0.634	0.505	0.653	0.738		Yes
US	0.672	0.562	0.689	0.610	0.769	Yes

Note: The top value in each column is the value of square root of AVE, which replaces self-correlation value of 1.

coefficients. The validity of latent variable relation is tested by bootstrap resampling technique (5000 bootstrap samples; no sign changes), which provides p-values and CLs to evaluate the significance of paths (Nevitt & Hancock, 2001). The results show that H1, H2, H3, H4, H5 and H7 are significant at the level of $p < 0.001$, and H6 is significant at the level of $p < 0.05$. The research hypotheses were all valid.

The interpretation and prediction capabilities of the PLS-SEM model were verified by the following indicators: R^2 and composite-based standardized root mean square residual (SRMR). R^2 is an important indicator to explain the predictive ability of the model, and the bigger the value of the R^2 is, the stronger of the model's

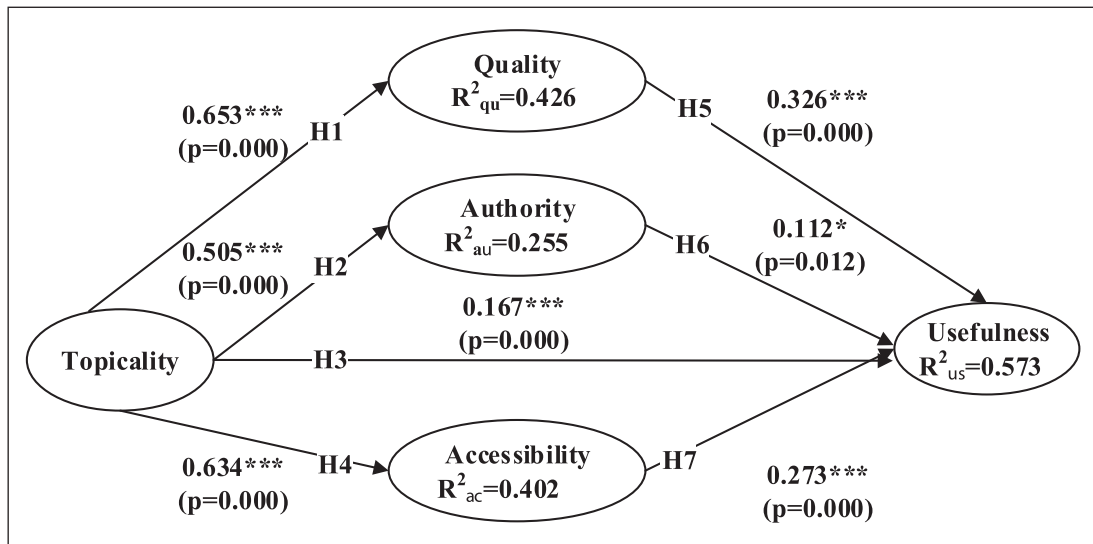


Figure 2: RC use structure model of scientific data users.

Note: Hypothesis testing result with SmartPLS3; SRMR = 0.088; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

prediction for the variance explanation of the endogenous variable will be. The R-square values of 0.25, 0.40 and 0.75 respectively indicate the weak, medium and strong level of the prediction ability of the model (Latan & Ramli, 2013; Hair et al., 2011). As show in **Figure 2**, the results of the coefficients are all great than the minimum threshold of 0.25. Furthermore, three of them are greater than 0.4, which means the model has a medium level of the prediction ability. In addition, from a review of similar studies (e.g., Lew and Sinkovics, 2013; Sarkar et al., 2001; Ralf and Siegfried, 2015) some authors used PLS-SEM. We concluded that the cut-off criteria of R^2 of the model in this study were acceptable.

SRMR is an index to evaluate the overall fitting degree of the model. It measures the discrepancy between the observed correlation matrix and the model-implied correlation matrix. The smaller the SRMR value is, the better the model will fit. In CB-SEM, the model has a good fit when SRMR is less than 0.08 (Hu & Bentler, 1998, 1999). Whereas in terms of research goal and context (see details of the difference between CB-SEM and PLS-SEM in Section 5.3), the recommended minimum threshold of SRMR recommended might be 0.1 in PLS-SEM (Henseler et al., 2014; Cangur & Ercan, 2015; Garson, 2016).

Therefore, as shown in **Figure 2**, R^2 of latent variables in this study are all greater than 0.25. Meanwhile, SRMR = 0.088 is less than the lenient threshold of 0.1 in PLS-SEM. Because this research belongs to exploratory research, the values of R^2 and SRMR show that the model has moderate ability for interpretation and prediction.

5.4.4 The model of RC use

The model verified and revealed the basic structure of RC use, which was characterized by the following three aspects. First, topicality (cause variable) is taken as the starting point of the user's relevance judgment of scientific data, affecting other levels of relevance judgment (as shown in **Figure 2**, H1–H4 are all significant at the level of $p < 0.001$). Second, quality, authority, and accessibility (intermediate variables) judgment are important processes of the user's relevance judgment of scientific data, which ultimately affect users' judgment on the usefulness of scientific data (as shown in **Figure 2**, H5, H7 are significant at the level of $p < 0.001$, and H6 is significant at the level of $p < 0.05$). Third, usefulness (result variable) expresses the user's comprehensive perception of the utility of scientific data in solving problems as the result of user relevance judgment. Based on the scientific data RC use path, user's behaviour patterns of relevance judgment of scientific data were discussed (see Section 6.2 for specific discussion).

6 Discussion and Implications

6.1 The relevance criteria of scientific data

This study identified 5 RC (topicality, accessibility, quality, authority and usefulness) and 18 corresponding clues of scientific data as shown in **Table 3**. Dozens of RC are used in the user relevance judgment for documents and images (Barry, 1994; Barry & Schamber, 1998; Choi, 2002). It is difficult to consider every RJ in

the practice of IR. While the number of RC used by scientific data users is relatively small, and there is a path structure for RC usage as shown in **Figure 2**. This study made new definitions to fit the context of scientific data research as shown in **Table 4**, though the concept of these 5 criteria are not proposed for the first time in this study. More importantly, the RC usage path reflects the patterns of scientific data user's relevance judgment behaviours.

6.2 Summary of relevance judgment patterns based on the use of relevance criteria

6.2.1 The pattern of "data topicality judgment as the first step or starting point"

The two-phase study verified that topicality plays a fundamental role and functions as a prerequisite in user relevance judgment on scientific data. **Table 3** showed that topicality is the most frequently used criterion with 325 coding nodes. **Table 5** illustrated that all 7 RC usage paths take topicality as the starting point and the PLS-SEM also verify that topicality has a positive effect on data quality, authority, usefulness and accessibility (as shown in **Figure 2**, H1–H4 are significant at the level of $p < 0.001$).

The prerequisite function of topicality has also been confirmed in user relevance judgement on texts/documents, images, and audio as information carriers (Barry, 1994; Wang and Soergel, 1998; Choi, 2002; Crystal and Greenberg, 2006; Laplante, 2010). However, as shown in **Table 3**, clues (data attributes) that stimulate scientific data users to make topicality judgment included not only general textual information such as data title, data description, data keywords, but also data time scope information, which are the clues often used to judge the novelty or recency of documents (Xu and Chen, 2006). This difference originates from the essential feature of scientific data, which is expected to be "evidences" rather than "novel viewpoints or new discoveries" from documents.

6.2.2 The pattern of "data reliability judgment as the necessary process"

As shown in **Figure 2**, H5 and H7 are significant at the level of $p < 0.001$, and H6 is significant at the level of $p < 0.05$. The results revealed that scientific data users pay "special" attention to the RC of data quality, authority and accessibility in scientific data relevance judgment. It can be summarised as the pattern of "data reliability judgment as the necessary process". That is embodied in the following aspects. Firstly, quality and authority represent scientific data user's judgment on the validity of data as "evidences", because data without quality and authority are useless in solving practical problems. Secondly, accessibility is used as a conditional criterion for users to judge the "evidences" of data, because users cannot make sufficient judgment without the whole data or adequate information.

6.2.3 The pattern of "data utility judgment as final purpose"

The results of the two-phase study verified that usefulness is the target and result variable for scientific data relevance judgment. Scientific data user's judgment of topicality, quality, authority and accessibility all have positive effects on usefulness judgment (as shown in **Figure 2**, H3, H5 and H7 are significant at the level of $p < 0.001$, H6 is significant at the level of $p < 0.05$, $R_{us}^2 = 0.573$). The research results verified that scientific data relevance judgment is a typical situational relevance judgment, which takes a pragmatic and measurable perspective and is operated as the utility/usefulness of the information objects to the user's situational task at hand (Borlund, 2003; Cosjin & Ingwersen, 2000; Xu & Chen, 2006).

The nature of fact-bearing of scientific data leads to the scientific data relevance judgment as a result of users' judgment of the utility of data as "evidences" to solve practical problems. This is also one of the most essential differences between data and document user. This difference also suggests that it is necessary to develop proprietary scientific data retrieval systems and algorithms.

6.3 Limitations and future directions

Before drawing any implications, some limitations should be mentioned. First, this research took academic search as the research situation, the results may not explain the non-academic search situation well. Second, the study took student groups as samples, and the results of the study should be carefully extended to other user groups. Third, as it is the first study that adopts PLS-SEM in RC use structure, this research is a typical exploratory one with the moderate ability of the research model. Therefore, the results should be interpreted with caution. The research will be developed in the future in a dual path: the first direction is high level of generalization, which means the test of more situations and groups of users; the other way is to apply its findings to design and develop an interactive and cognition-friendly retrieval system specific to scientific data.

6.4 Implications

Except for enlarging the understanding of how human make relevance judgment on scientific data (or in a more general sense, information), the research seeks to upgrade or even trigger off a sort of data-specific and cognition-friendly retrieval technique based on the understanding of the relevance judgment pattern of scientific data users. As to current findings of the research, it can contribute to achieving the ultimate goal at least in the following aspects.

6.4.1 Implications for metadata schema design

The essential features and attributes of scientific data and the RC employed by users suggest a more cognitive data description schema for informed decision of how to select a dataset. Traditionally, people describe datasets with various metadata schema. Comparing the user relevance judgement, including criteria and its underlying attributes of datasets, as well as the usage pattern of criteria, it is clear that traditional metadata schema cannot provide sufficient information for relevance decision. Given the importance role of representation and description of information to IR, it is imperative for researchers to provide suitable dataset description schema for a cognition-friendly data retrieval system.

6.4.2 Implications for user relevance algorithm design

The combination use structure of RC reveals the defect of a system-oriented relevance algorithm which can only partially capture topicality. Meanwhile, it also calls for the user-oriented relevance algorithm that comprehensively considers the paths and strength of different RC on relevance judgment. Previous researchers have done some exploratory studies in multi-criteria decision. Xu and Chen (2006) proposed multiple criteria (topicality/reliability/understandability/novelty/scope) use model by using the algorithm of multiple regression. However, Xu and Chen's model only considered the influence strength of RC but did not consider the path among RC. Célia (2010, 2012) developed a multiple-criteria (aboutness/coverage/appropriateness/reliability) relevance evaluation model that considers the prerequisite role of aboutness by using the algorithm of priority aggregation. Specific to the findings of three RC use patterns in scientific data user's relevance judgment, it will be a challenge and direction for future research to comprehensively understand these patterns in the form of algorithms.

6.4.3 Implications for interactive data retrieval systems

The usage patterns and their underlying cognitive mechanism throw light on an interactive mechanism of data retrieval systems. The results show that user relevance judgment on scientific data does not just depend on one CR, nor make the final decision at beginning stage. It is an interactive process, in which user make differential levels of relevance judgment before reaching the final decision. However, traditional term-matching technologies, which just partially captures user's information need in the aspect of topicality rather than considering usage patterns of other RC. It calls for an interactive scientific data retrieval system with more cognition-friendly.

7. Conclusion

The research carries out a two-phase study to explore how users judge the relevance of scientific datasets. Five RC and three patterns of the RC usage are identified in the context of data retrieval within an academic situation relating to data use. The findings will contribute to deepening the understanding of user relevance judgment, and will give suggestions and instructions for designing a novel interactive, cognition-friendly and hence more effective data-specific retrieval system.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Interview outline. DOI: <https://doi.org/10.5334/dsj-2020-009.s1>
- **Appendix B.** Major items of the questionnaire. DOI: <https://doi.org/10.5334/dsj-2020-009.s2>

Ethics and Consent

The study was approved by the Logistics Department for Civilian Ethics Committee of Agricultural Information Institute, Chinese Academy of Agricultural Sciences.

All subjects who participated in the experiment were provided with and signed an informed consent form. All relevant ethical safeguards have been met with regard to subject protection.

Acknowledgements

This work was supported by grants from National Scientific and Technical Infrastructure Center of China project titled “The Research on Quality Assessment of Scientific Data (2018DDJ1ZZ16)”, and Agricultural Science, Technology Innovation Project of CAAS (Project No.CAAS-ASTIP-2016-AII), and the project of Fundamental Research Funds for CAAS, “The Construction of Platform and Data Mining for Crop Breeding (Y2016XK08)” and “The Establishment of Data Collection Standard and System Development for Agricultural Experimental Stations (Y2019PT27)”.

The authors also acknowledge the data collection support from “National Science & Technology Infrastructure of China” (<http://www.escience.net.cn>).

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Jianping Liu contributed to experimental design, data analysis, manuscript writing and revision. Jian Wang and Guomin Zhou contributed to guiding the experimental design and the revision of the manuscript. Mo Wang and Lei Shi contributed to the revision of the manuscript.

Author Information

Jianping Liu is a research associate at the Agricultural Information Institute, Chinese Academy of Agricultural Sciences. He is a PhD student in information retrieval technology in Chinese Academy of Agricultural Sciences. His research has been published in *Library & Information Science Research* (2019) and *The 3rd International Conference on Computer Science and Application Engineering*(2019), and *Man-Machine-Environment system engineering: proceedings of the 19th international conference on MMESE*(2019).

Jian Wang is a professor in information sciences at the Cognitive Computing Research Laboratory of Agricultural Information Institute, Chinese Academy of Agricultural Sciences. He holds a PhD in geoinformatics from the Chinese Academy of Sciences. His research has been published in *Sensor Letters* (2010) and *Library & Information Science Research* (2019).

Guomin Zhou is a chancellor's professor in information sciences, deputy director of Department of Science and Technology Management, Chinese Academy of Agricultural Sciences. He holds a PhD in information sciences from the Chinese Academy of Sciences. He is the executive editor of the *Journal of Agricultural Big Data in China*. His research has been published in *Intelligent Automation and Soft Computing* (2010) and *Library & Information Science Research* (2019).

Mo Wang is an assistant professor at Cognitive Computing Research Laboratory of Agricultural Information Institute, Chinese Academy of Agricultural Sciences. He holds a PhD in Geographical Information Systems from the Chinese Academy of Sciences.

Lei Shi is a professor in information resources management at the National Science and Technology Infrastructure Center in China.

References

- Anderson, JC** and **Gerbing, DW**. 1988. Structural Equation Modeling in Practice: A Review of the Two-Step Approach. *Psychological Bulletin*, 103(3): 411–423. DOI: <https://doi.org/10.1037/0033-2909.103.3.411>
- Barry, CL**. 1994. User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, 45(3): 149–159. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199404\)45:3<149::AID-ASI5>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-4571(199404)45:3<149::AID-ASI5>3.0.CO;2-J)
- Barry, CL** and **Schamber, L**. 1998. Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, 34: 219–236. DOI: [https://doi.org/10.1016/S0306-4573\(97\)00078-2](https://doi.org/10.1016/S0306-4573(97)00078-2)
- Bateman, JA**. 1998. Modeling Changes in End-user Relevance Criteria: An Information Seeking Study. Unpublished thesis (PhD), University of North Texas.
- Borgman, C**. 2013. *Scholarship in the networked world: Big data, little data, no data*. Oxford: Oliver Smithies Lecture, Balliol College.
- Borlund, P**. 2003. The concept of relevance in IR. *Journal of the American Society for Information Science & Technology*, 54(10): 913–925. DOI: <https://doi.org/10.1002/asi.10286>

- Cangur, S** and **Ercan, I**. 2015. Comparison of Model Fit Indices Used in Structural Equation Modeling Under Multivariate Normality. *Journal of Modern Applied Statistical Methods*, 14(1). DOI: <https://doi.org/10.22237/jmasm/1430453580>
- Choi, Y** and **Rasmussen, EM**. 2002. User's relevance criteria in image retrieval in American history. *Information Processing & Management*, 38(5): 695–726. DOI: [https://doi.org/10.1016/S0306-4573\(01\)00059-0](https://doi.org/10.1016/S0306-4573(01)00059-0)
- Cosijn, E** and **Ingwersen, P**. 2000. Dimensions of relevance. *Information Processing & Management*, 36(4): 533–550. DOI: [https://doi.org/10.1016/S0306-4573\(99\)00072-2](https://doi.org/10.1016/S0306-4573(99)00072-2)
- Crystal, A** and **Greenberg, J**. 2006. Relevance criteria identified by health information users during Web searches. *Journal of the American Society for Information Science and Technology*, 57(10): 1368–1382. DOI: <https://doi.org/10.1002/asi.20436>
- da Costa Pereira, C**, **Dragoni, M** and **Pasi, G**. 2010. User Evaluation of Multidimensional Relevance Assessment. *IIR 2010 – Proceedings of the First Italian Information Retrieval Workshop*, Padua, Italy, January 27–28, 2010.
- da Costa Pereira, C**, **Dragoni, M** and **Pasi, G**. 2012. Multidimensional relevance: prioritized aggregation in a personalized information retrieval setting. *Information Processing & Management*, 48(2): 340–357. DOI: <https://doi.org/10.1016/j.ipm.2011.07.001>
- Downe-Wamboldt, B**. 1992. Content analysis: method, applications, and issues. *Issues in Health Care of Women*, 13(3): 313–321. DOI: <https://doi.org/10.1080/07399339209516006>
- Fornell, C** and **Larcker, DF**. 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1): 39–50. DOI: <https://doi.org/10.1177/002224378101800104>
- Garson, GD**. 2016. *Partial least squares: Regression and structural equation models*. Asheboro, NC: Statistical Associates Publishers.
- Gregory, K**, **Groth, P**, **Cousijn, H**, **Scharnhorst, A** and **Wyatt, S**. 2017. Searching Data: A Review of Observational Data Retrieval Practices in Selected Disciplines. *arXiv: Digital Libraries*.
- Greisdorf, HF**. 2000. Relevance thresholds: A conjunctive/disjunctive model of end-user cognition as an evaluative process. Unpublished thesis (PhD), Denton, TX: University of North Texas.
- Hair, JF**, **Hult, GTM**, **Ringle, CM** and **Sarstedt, M**. 2014. A primer on partial least squares structural equation modeling (PLS-SEM). Thousand Oaks: Sage.
- Hair, JF**, **Ringle, CM** and **Sarstedt, M**. 2011. PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory & Practice*, 19(2): 139–152. DOI: <https://doi.org/10.2753/MTP1069-6679190202>
- Harter, SP**. 1992. Psychological relevance and information science. *Journal of the Association for Information Science & Technology*, 43(9): 602–615. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199210\)43:9<602::AID-ASI3>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-4571(199210)43:9<602::AID-ASI3>3.0.CO;2-Q)
- Henseler, J**, **Dijkstra, TK**, **Sarstedt, M**, **Ringle, CM**, **Diamantopoulos, A**, **Straub, D**, **Ketchen, D**, **Hair, JF**, **Hult, GTM** and **Calantone, RJ**. 2014. Common Beliefs and Reality About Partial Least Squares: Comments on Rönkkö & Evermann (2013). *Social Science Electronic Publishing*, 17(2): 182–209. DOI: <https://doi.org/10.1177/1094428114526928>
- Hu, L** and **Bentler, PM**. 1998. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4): 424–453. DOI: <https://doi.org/10.1037/1082-989X.3.4.424>
- Hu, L** and **Bentler, PM**. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1): 1–55. DOI: <https://doi.org/10.1080/10705519909540118>
- Ingwersen, P** and **Järvelin, K**. 2011. *The Turn: Integration of Information Seeking and Retrieval in Context*. Berlin, Germany: Springer Publishing Company, Incorporated.
- Inskip, C**, **Macfarlane, A** and **Rafferty, P**. 2010. *Creative professional users' musical relevance criteria*. Sage Publications, Inc. DOI: <https://doi.org/10.1177/0165551510374006>
- Laplante, A**. 2010. Users' Relevance Criteria in Music Retrieval in Everyday Life: An Exploratory Study. In: *International Society for Music Information Retrieval Conference, Ismir 2010*, Utrecht, Netherlands, August, 2010, 601–606.
- Latan, H** and **Ramli, NA**. 2013. *The Results of Partial Least Squares-Structural Equation Modelling Analyses (PLS-SEM)*. Social Science Electronic Publishing. DOI: <https://doi.org/10.2139/ssrn.2364191>
- Lew, YK** and **Sinkovics, RR**. 2013. Crossing borders and industry sectors: behavioral governance in strategic alliances and product innovation for competitive advantage. *Long Range Planning*, 46(s1–2): 13–38. DOI: <https://doi.org/10.1016/j.lrp.2012.09.006>

- Liu, JP, Wang, J and Zhou, GM.** 2019. Understanding relevance judgment in the view of perceived value. *Library & Information Science Research*, 41(4). DOI: <https://doi.org/10.1016/j.lisr.2019.100982>
- Markkula, M and Sormunen, E.** 2000. End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive. *Information Retrieval*, 1(4): 259–285. DOI: <https://doi.org/10.1023/A:1009995816485>
- Nevitt, J and Hancock, GR.** 2001. Performance of Bootstrapping Approaches to Model Test Statistics and Parameter Standard Error Estimation in Structural Equation Modeling. *Structural Equation Modeling A Multidisciplinary Journal*, 8(3): 353–377. DOI: https://doi.org/10.1207/S15328007SEM0803_2
- Perreault, WD and Leigh, LE.** 1989. Reliability of nominal data based on qualitative judgments. *Journal of Marketing Research*, 26(2): 135–148. DOI: <https://doi.org/10.1177/002224378902600201>
- Ralf, W and Siegfried, PG.** 2015. The impact of dynamic capabilities on operational marketing and technological capabilities: investigating the role of environmental turbulence. *Journal of the Academy of Marketing Science*, 43(2): 181–199. DOI: <https://doi.org/10.1007/s11747-014-0380-y>
- Ringle, CM, Wende, S and Becker, JM.** 2015. Smart PLS 3. Bönningstedt: SmartPLS. Retrieved from <http://www.smartpls.com>.
- Sanderson, M and Croft, WB.** 2012. The History of Information Retrieval Research. *Proceedings of the IEEE*, 100(6): 1444–1451. DOI: <https://doi.org/10.1109/JPROC.2012.2189916>
- Saracevic, T.** 1975. RELEVANCE: A Review of and a Framework for the Thinking on the Notion in Information Science. *Journal of the Association for Information Science & Technology*, 26: 1915–1933. DOI: <https://doi.org/10.1002/asi.20682>
- Saracevic, T.** 2016. The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really? *Synthesis Lectures on Information Concepts Retrieval & Services*, 8(3): 1–109. DOI: <https://doi.org/10.2200/S00723ED1V01Y201607ICR050>
- Savolainen, R and Kari, J.** 2013. User-defined relevance criteria in web searching. *Journal of Documentation*, 62(6): 685–707. DOI: <https://doi.org/10.1108/00220410610714921>
- Schamber, L.** 1990. Users' Criteria for Evaluation in a Multimedia Environment. *Proceedings of the American Society for Information Science & Technology*, 28(6): 126–133.
- Taylor, RS.** 1968. Question-negotiation and information seeking in libraries. *College & Research Libraries*, 29. DOI: https://doi.org/10.5860/crl_29_03_178
- Tombros, A, Ruthven, I and Jose, JM.** 2003. Searchers' criteria for assessing web pages. In: *International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, Toronto, Canada, 28 July–01 Aug 2003. 385–386.
- Wang, PL and Soergel, D.** 1998. A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science*, 49(2): 115–133. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199802\)49:2<115::AID-ASI3>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1097-4571(199802)49:2<115::AID-ASI3>3.0.CO;2-T)
- Xu, Y and Chen, Z.** 2006. Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science & Technology*, 57(7): 961–973. DOI: <https://doi.org/10.1002/asi.20361>

How to cite this article: Liu, J, Wang, J, Zhou, G, Wang, M and Shi, L. 2020. How Do People Make Relevance Judgment of Scientific Data? *Data Science Journal*, 19: 9, pp. 1–15. DOI: <https://doi.org/10.5334/dsj-2020-009>

Submitted: 09 May 2019

Accepted: 15 January 2020

Published: 09 March 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 