The Effects of Source Credibility Ratings in a Cultural Heritage Information Aggregator

Alia Amin
Semantic Media Interfaces,
CWI, the Netherlands
Alia.Amin@cwi.nl

Junte Zhang
Archives and Information
Studies, Faculty of
Humanities, University of
Amsterdam,the Netherlands
J.Zhang@uva.nl

Henriette Cramer
Human-Computer Studies
Laboratory, University of
Amsterdam, the Netherlands
H.S.M.Cramer@uva.nl

Lynda Hardman
Semantic Media Interfaces,
CWI, the Netherlands
Human-Computer Studies
Laboratory, University of
Amsterdam, the Netherlands
Lynda.Hardman@cwi.nl

Vanessa Evers
Human-Computer Studies
Laboratory, University of
Amsterdam, the Netherlands
V.Evers@uva.nl

ABSTRACT

State of the art web search applications allow the user to aggregate information from many sources. Because of this, users are confronted with having to assess the reliability of information from different sources. This paper reports on an empirical user study on the effect of displaying credibility ratings of multiple cultural heritage sources (e.g. museum websites, art blogs) on users' search performance and selection. The study investigated whether source credibility has an influence on users' search performance when they are confronted with only a few information sources or where there are many. The results of our online interactive study (n=122) show that by presenting the source credibility information explicitly, people's confidence in their selection of information significantly increases, even though it does not necessarily make search more time efficient. Additionally, we highlight credibility issues that are applicable beyond the cultural heritage domain, such as issues related to credibility measures and choice of visualization.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: Models and Principles—*User/Machine Systems*; H.3.7 [**Digital Libraries**]: System issues, User issues; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Design, Experimentation, Human Factors.

1. INTRODUCTION

Consider a student who is looking for information about Dutch painters in the 17th century who made portraits in a certain artstyle. This information may be online, but bits and pieces may exist on different HTML web pages, Wikipedia articles, weblogs, etc. The student has to choose which pages to look at from a large number presented in the search result page. Before making this decision, s/he may examine where the information comes from and who wrote it before looking at the article itself.

The situation above occurs frequently to all of us in slightly different settings. Many web search aggregators enable us to find information from different sources simultaneously [22]. These systems enable us to quickly retrieve information from multiple sources but the decision as to which information source to go for is something that is left to the user. When interacting with such systems, we constantly need to assess the credibility of the information sources, the authors and the content. In this decision process, credibility plays an important role.

Even though the number of information aggregators is expanding, there are only few studies that report on how the added complexity of having to deal with multiple information sources influences the users' ability to make decisions and select the appropriate information. The aim of this study is to investigate how displaying the credibility ratings of cultural heritage sources affects the user's confidence and time to search for information.

The outline of this paper is as follow: related work and problem statement are given in the next sections, the experimental setup is laid out in Section 4, the results are presented in Section 5 and finally we conclude this paper with discussions and conclusions in Section 6.

2. RELATED WORK

2.1 Credibility and the Web

We adopt the simple notion of credibility as *believabilty* [13]. One of the earliest research about credibility was conducted

by [16]. That paper specifically focused on *source credibility*, i.e. credibility of a source. More recently, it is becoming an important issue for research on the World Wide Web and information access as well. As new information systems are emerging that combine information from multiple sources, the effects of source credibility for improving information access comes back as a research issue. In [12], distinctions are made between several types of credibilities, like Web credibility, site credibility, sponsor credibility, news credibility, etc. Credibility research is complex, and researchers do not necessarily or completely agree with its dimensions as a concept. However, a comprehensive and extensive literature overview about credibility is presented in [24], where credibility is related to a multidisciplinary framework, and related to other concepts (which are not the same as credibility), such as 'quality', 'authority', 'trust', and 'persuasion'. There are also different types of credibility by attaching it to potentially interlinked objects of assessment, like source credibility, media credibility, and message credibility. In this paper we mainly focus credibility on the object of the source.

There has been ample research conducted about credibility, and in a plethora of domains. For example, the credibility in the online encyclopaedia Wikipedia [4, 20], and website design for achieving credible websites [27]. Credibility in academic information and in the online encyclopaedia Wikipedia [21, 4], health and medicines [19, 9, 29, 6, 23, 11], media and news providers [17, 26, 28, 5, 3], website design for achieving credible websites [27, 15, 12]. These research reports different aspects of credibility. For example, in [11] the objectives were to describe techniques for retrieval and appraisal used by consumers when they search for health information on the Internet. They identified factors to determine the website credibility, where authority of a source, email, credentials and qualifications can be applied to the source credibility. The credibility of the source appeared to be a common determinant in the criteria of all participants who looked online for information about medicines [23]. Source credibility is also a very important issue in the cultural heritage domain, especially for cultural heritage experts [1] and historians [8]. Both studies have reported that it is important for these experts to be able to assess the credibility of the source before using their information.

2.2 Aggregated Search

As defined in [22], Aggregated Search deals with the task of searching and assembling information from a variety of sources, and placing that information in a single interface. Examples of generic Information Retrieval (IR) systems are Alpha Yahoo[1], Google Universal Search[2] and Naver[3], a Korean search portal. There are also domain specific IR systems, such as Google Scholar[4], a search engine that harvests information from publishers, preprint repositories, universities and scholarly organizations, or WorldCat[5], a union catalog of more than 10.000 libraries. In the cultural heritage domain, ECulture Multimedian[25] and CultureSampo[10] allow to search for information from various museums or similarly the README system [30] for historical archives.

[1] http://au.alpha.yahoo.com/
[2] http://www.google.com/intl/en/press/pressrel/universalsearch_20070516.html
[3] http://www.naver.com/
[4] http://scholar.google.com/
[5] http://www.worldcat.org/

For these systems, several issues remain open questions, such as what is the best way to present information that would allow the user to interact with the system in the most natural way [22]. This include the question of what kind of information is essential for the presentation and how to best present it.

2.3 Transparency

In the context of the evaluation of IR systems, it has been pointed out that a central question for the design of interactive systems is the amount of knowledge that is needed about a system [18]. On the one hand, interfaces can hide the inner workings of a system as much as possible and put the focus on the user's task. However, on the other hand, some knowledge and control may be necessary to enhance interaction with (different components of) the system. This deals at its core with the issue of *transparency*. Conversely, transparency could also influence the trust and acceptance of systems.

As pointed out in [7], there has not been a clear-cut consensus in numerous different studies which empirically tested the effects of system transparency. There were mixed findings: transparency could enhance or even worsen user interaction with systems. There is no general guideline and it highly depends on the implementation and the application. For example, the study that was conducted in [7] tested the effects of transparency with an adaptive recommender system. It was found that transparency increased the acceptance of recommendations and makes a system more understandable, which correlates with the perceived competence of users (and thus enhances the interaction with the system).

2.4 Credibility Measures

In a computer credibility research [27], credibility is interpreted as believability. The authors point out that credibility has 2 dimensions: The *trustworthiness* dimension of credibility captures the perceived goodness, morality of the source or 'well-intentioned', 'truthful', and 'unbiased'. The *expertise* dimension of credibility captures the perceived knowledge and skill of the source, or 'knowledgeable', 'experienced', and 'competent'. They continue by pointing out that credible people are believable people; credible information is believable information. Credibility is a perceived quality, reflected by phrases, such as 'trust the information', 'accept the advice', 'believe the output'.

Most research on credibility take quantitative empirical social science experiments as their approach. In this paper we adopt a similar approach.

3. PROBLEM STATEMENT

When presented with multiple sources users need to make several decisions about the information search results. This added complexity is likely to influence users' search performance. Our main research question is therefore:

- Does source credibility improve information access to information aggregated from multiple sources?

We elaborate our research question through three research hypotheses. First, when the users are confronted with multiple sources, they need to establish the credibility of the source and its information. We expect that the users will feel less confident in their selection of information from the numerous aggregated results:

H1 *The confidence users have in the reliability of the information they selected will decrease when the number of sources increases.*

Results from [7] indicate that information transparency can positively influence the acceptance of recommendations coming from a single source. This may also be the case for information from multiple sources provided by an information aggregator. By presenting the origin of the information and the credibility ratings, a user may be able to assess information quicker and be more assured about their decisions. Our conjecture is that it will result in higher confidence levels and less time needed to search. This leads to the following hypotheses:

H2 *Displaying the 'source credibility' will give users greater confidence in the information they select.*

H3 *Displaying the 'source credibility' will produce a shorter search time when compared to when it is not displayed.*

4. EXPERIMENTAL SETUP

To test our hypotheses, we conducted a two phase studies. The first study was a pre-study designed to elicit source credibility scores for the cultural heritage sources, such as museums and art websites, that will be used for the Study 2. The second study aimed to investigate how displaying source credibility influences the user search efficiency.

4.1 Study 1: Measuring Cultural Heritage Source Credibility

For the purpose of eliciting credibility scores of 12 art sources from 4 categories (see Table 1) we developed an interactive online survey (see Figure 1) which automatically generated a random set of 6 art sources, out of a total of 12, for the participant to assess on credibility.

4.1.1 Procedure

When evaluating credibility of an online information, both the receptor's attribute [2] and the source's attribute [27, 12, 15] should be taken into consideration. Based on this literature, we define the credibility measures for the cultural heritage domain:

Receptor's attributes knowledge (of arts and culture), and reliance (degree of importance of arts and culture).

Source attributes Trustworthiness, Completeness, Goodwill, and Expertise.

Participants were asked demographic questions reflecting the receptor's attribute. To measure the source attributes, participants are given information about the different sources taken from the sources' public website. The type of chosen information were taken from[11, 21] and adapted for the cultural heritage domain. This information is: *logo*, *name*, *contact information*, *mission statement*, *history*, *repository*, *source type*, and *organization*. For each source, participants were asked to assess the 4 source attributes by using a 5-point scale (see Fig. 1).

Trustworthiness (T) *"I believe this source will give cultural information that is neutral and it has good intentions."*

Completeness (C) *"I believe this source is able to give me everything I need to know about an artwork."*

Goodwill (G) *"I believe this source genuinely cares about arts and culture."*

Expertise (E) *"This source appears to be a leader in its area of speciality."*

Additionally, participants were also asked whether they were familiar with or have prior knowledge about each source by using 5-point Likert scale.

Source knowledge *"I already have a lot of knowledge about this source."*

4.1.2 Participants

In total, 57 participants were recruited from Dutch online discussion boards over a period of 3 weeks. The gender is almost evenly distributed (male: 47.4%, female: 52.6%), they were highly educated (61.4% University education), were very experienced with the Internet (M = 4.39, SD = 0.80), had some arts and culture knowledge (M = 3.14, SD = 0.93), and had lived their entire life in the Netherlands (M = 25.49, SD = 9.93).

4.1.3 Cultural heritage source credibility scores

To examine the reliability of our 4 source attributes value, we calculated the Cronbach's α and found the value of 0.70, which is sufficient to determine the credibility of a source.

Figure 1: Example screenshot used in Study 1 where participants judge the credibiltiy of a cultural heritage source

The credibility scores of the 12 sources are depicted in Table 1. The scores of source attributes were averaged into a single numeric value. The museums were perceived as most credible by the participants, followed by arts websites. The general websites, which do not exclusively cover arts and culture, score lowest. Interestingly, *Groove Art Online* was considered almost as credible as the museum sources. Overall, the sources were assessed as what we expected beforehand. We also checked for the relationship between source knowledge and the perceived credibility score. We found weak

Source	Credibility score M (SD)	γ	Source Type
Joods Historisch Museum	4.09 (.61)	6.7	
Rijksmuseum A'dam	4.34 (.59)	7.5	T1
Museum Volkenkunde	4.05 (.55)	6.6	
absolutearts.com	3.75 (.55)	5.6	
ArtLex Art Dictionary	3.25 (.77)	4.2	T2
Grove Art Online	3.96 (.69)	6.3	
art.blogging.la	3.20 (.66)	4.1	
Artblog.net	3.13 (.74)	3.9	T3
ArtsJournal	3.29 (.66)	4.3	
About.com	2.94 (.82)	3.5	
Infoplease	2.93 (.72)	3.4	T4
Wikipedia	3.05 (.86)	3.7	

Table 1: Sources' Credibility Score, where T1: Museums, T2: Arts Websites, T3: Art Blogs, T4: General Websites. γ represents the marked up credibility score as calculated by equation (1) and (2).

relationship between source knowledge and source credibility. Only *Art.blogging.la* was significant (2-tailed) (r=.39, n=28, p<.041) and no correlation for the other sources. We conclude that for the sources that were used in this study, participants gave the credibility scores based only on the type of information given (e.g. source history, mission statement) and independent of their knowledge with the source.

We use these credibility scores as basis for our follow up study. To increase the difference between the sources, we mapped the credibility score by using a quadratic function (see eq.1). This results in larger differentials between the scores which will make the visualization clearer for Study 2[6]. The score Ψ was calculated for each score by taking the quadratic value of credibility scores and translated to a 0 to 10 scale (0: lowest credibility, 10: highest credibility).

$$\Psi_i = \left(\frac{T_i + C_i + G_i + E_i}{4} \right)^2 \cdot \frac{1}{2.5} \quad (1)$$

$$\gamma = \frac{1}{K_{total}} \cdot \sum_{i=1}^{N} (\Psi_i \cdot K_i) \quad (2)$$

where the rating T is the *Trustworthiness*, rating C is *Completeness*, rating G is *Goodwill*, rating E is *Expertise* for the i-th participant, and $\Psi \in [0, 10]$. Finally, we take the users source knowledge into account for the mark up credibility score of each source γ by taking Ψ from each i-th participant, and use the *Source Knowledge* K_i as propagation factor, where $K_i \in [1, 5]$ and K_{total} is the sum of all K-s (see Table 1 for mark up credibility score γ for all sources).

4.2 Study 2: Cultural Heritage Source Credibility Effects

In Study 2 we investigated the effects of displaying credibility scores of cultural heritage sources on users confidence and search time. We use the mark up credibility scores γ from Study 1 in a bar visualization for different cultural heritage information sources.

---

[6]We acknowledge this mapping could be done with other functions, such as logarithmic, which we will explore in future research.

4.2.1 Procedure

An experiment was carried out to assess the effects of multiple information sources and source credibility on search performance and user confidence. A 2x2 experiment was carried out with number of sources (few = 4 sources, many = 12 sources) and source credibility (ratings or no ratings) as between-subject variables (see Table 2). An interactive online survey was developed in which participants were randomly and automatically assigned to one of the four conditions (FN, FR, MN, or MR).

	Few Sources (4 sources)	Many Sources (12 sources)
No credibility rating	*FN*	*MN*
With credibility rating	*FR*	*MR*

Table 2: Study 2 setup with 4 conditions: the number of cultural heritage information sources (few or many) vs. availability of the source credibility score ratings (displayed or not displayed).

Each participant was given 3 tasks. In each task, a participant needs to select the culture of origin and the description of artwork that seemed most accurate to them. The possible answers shown come from different cultural heritage information sources. In the FN and MN conditions, the participant only sees the name of the information source. In the FR and MR condition, the participant is also given extra information on the credibility ratings of each information source (see example in Fig. 2 from Study 1). In total, participants conducted 3 tasks. The tasks order is randomized across all participants. After each task, participants were asked to rate the confidence of their selection in a 5-point scale.

4.2.2 Variables

The descriptions available for the artwork were all designed to be concise, equally realistic and plausible for all four conditions. Number of sources was manipulated by either presenting information from 4 or 12 sources. As a greater number of choices also implies a longer time needed to complete a choice task [8], this effect had to be controlled. The answers for the many sources conditions were short sentence, for the few sources conditions answers were usually longer (around 3 sentences).

Source credibility transparency was manipulated by displaying (or hiding) the source credibility for each chunk of information from a different source. Each task consisted of low and high credibility sources. In the no source credibility conditions (FN and MN) the orange rating bar and the credibility indication were omitted (see Figure 2). We presented low and high credibility ratings (not controlled), because we wanted to simulate a more realistic use-case and interactive scenario.

We wanted to focus on and manipulate only the influence of credibility of source and not the credibility of content. Thus, in every task, the information content provided for the artwork was fictional but equally plausible (as assessed by multiple art experts). We have also chosen unknown artworks to make sure that the participants could not rely on prior knowledge about the artwork to answer the questions in the task.

The dependent variables measured were *efficiency* (the time it took a participant to finish the search task), and

Figure 2: Example screenshots used in Study 2. *Top*: few sources-with credibility ratings condition (FR). *Bottom*: few sources-no credibility ratings condition (FN).

participant's *confidence* in the accuracy of the information selected. *Search accuracy* (giving the right answer) was not measured as the information given to the participants were fictional but equally plausible.

We also checked whether participants rely on the content of the information or the source (Q1 and Q2). Additionally, for the rating conditions (FR and MR), we asked an extra question on the visualization (Q3). The answer selected by the participant was expressed on a 5-point scale.

Q1 Reliance on source *"I only choose the answer from the source I trust most."*

Q2 Reliance on content *"I only choose the answer that seems most correct to me."*

Q3 Reliance on visualization *"I only choose the answer based on the visualization."*

4.2.3 Participants

Participants ($N = 122$) for Study 2 were recruited in the Netherlands through a mailing to students and research staff at Dutch research institutes in a period of 2 weeks. The valid response ratios were 81.6% (FN, $N = 30$), 71.4% (MN, $N = 31$), 81.6% (FR, $N = 30$), and 75.0% (MR, $N = 31$).

The participants demographics were similar to those in the first study. The gender was equally divided, the average age was 29 years (SD=10.15), they were highly educated (68.0% University level), were very experienced with the Internet (M=4.50, SD=.71), had modest knowledge of Cultural Heritage (M=2.71, SD=.92), but they gave a high importance

to Cultural Heritage (M=3.64, SD=1.00) and most had lived their entire lives in the Netherlands (M=25.07, SD=12.35).

5. RESULTS

In the analysis, Kruskal-Wallis (K-W) test was performed to check for the effects across the four groups[7]. Afterwards, the Mann-Whitney (M-W) U test was used as a post-hoc test to check for effects on the dependent variables (efficiency, confidence, reliance of content and reliance on source). We conducted the M-W U test for both the few sources conditions (FN and FR) and the many sources conditions (MN and MR), and for with credibility ratings (FN and MN) and no credibility ratings conditions (FR and MR).

5.1 Effects of Displaying Credibility Ratings

A significant effect was found on confidence in the accuracy using the K-W test ($\chi^2(3)$=11.16, p(2- tailed)=.011). Scores on confidence were highest for both with credibility ratings/many sources - MR (M=3.90) and the with credibility ratings/few sources - FR (M=4.19) conditions. Therefore we accept Hypothesis 2: displaying the 'source credibility' ratings do give users greater confidence in the information they select.

There was also a significant effect on efficiency using the K-W test ($\chi^2(3)$=17.80, p(2-tailed)=.000). The efficiency was surprisingly the highest for the no credibility rating/few sources condition (FN), where the average rank was 47.16

---

[7]Nonparametric statistics is used as the data did not meet parametric assumptions

(M=214.26 s) and the non-transparent/many sources condition with an average rank of 69.63 (M=458.03 s). This means we cannot support Hypothesis 3.

The M-W U test was used to check for more effects. In the few sources condition, the effect of transparency on reliance on content was significant (U = 341.500, z = -2.1, p(2-tailed)=.038). When the source credibility rating is displayed, the average rank was 27.02 (Mdn=4.00), whereas when it was not displayed, there was an average rank of 35.98 (Mdn=4.00). In the many sources conditions, we did not find a significant effect. We did not find other significant effects.

5.2 Effects of Number of Sources

As we haven taken into account Hick's law [14] to control for the amount of information, we can compare the mean values of the few and many sources conditions. We found that the participants need significantly more time per source to complete the tasks as the number of sources is tripled. The efficiency was highest for the no credibility rating/few sources condition (FN), where the average rank was 47.16 (M=214.26) and highest for the with credibility rating/few sources condition (FR) with an average rank of 50.29 (M=227.07). Our a priori assumption is confirmed. Moreover, the confidence drops, but not significantly, either when there is no credibility ratings (U=405.500, z=-.870, p(2-tailed)=.384), or with credibility ratings (U=393.500, z=-1.038, p(2-tailed)=.299), so we cannot support Hypothesis 1.

With the M-W U test, we found that the reliance on the content was significant when there was no credibility ratings (U=301.00, z=-2.495, p(2-tailed)=.013). Participants who had information from 4 sources had an average rank of 36.29 (Mdn=4.00) and participants who had 12 sources to assess had an average rank of 25.53 (Mdn=4.00), and hence relied less on the given content. We did not find more significant effects.

5.3 Effects Between Source and Content

Participants answered several statements about what their decisions were based on: the answer (content) or the source. For the with credibility ratings conditions (FR and MR), an extra option was added: the visualization. The bar chart in Figure 5.2 shows the means from a 5-scale Likert value from 2 statements.

It shows that our participants agree more with statement 2 than statement 1 in all the conditions. We check for statistical significant (2-tailed) correlation between these 2 variables for all 4 conditions with Spearman's rho. We found a large negative correlation for condition 1 (r = -.60, n = 31, p < .0001), no correlation for condition 2 (r = -.01, n = 30, p < .95), a significant medium negative correlation in condition 3 (r = -.49, n = 31, p < .006), and a medium but not significant relationship in condition 4 (r = -.34, n = 30, p < .06). The results show that there is a significant strong relationship between source and answer for choosing information when people are confronted with the few sources conditions.

The bar chart in Figure 5.2 depicts what our participants choose for **Q3 Reliance on visualization**, which resulted in nominal data. Participants are asked to prioritize between the 2 factors that we focus on, and additionally the 3rd factor "the source credibility" is introduced in the *with*
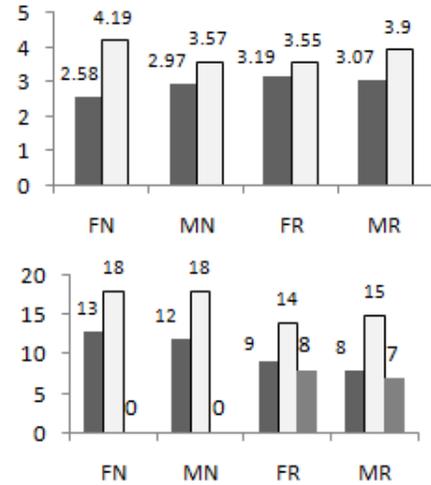
Figure 3: *Top:* **Mean decision score per condition based on Q1 - Reliance on the source (black bar) or Q2 - Reliance on the Content (white bar) Score Min:1 Max:5**
Bottom: **The number of people per condition that based decision on either: source (black bar), content (white bar) or visualization (grey bar).**

source credibility conditions. It shows that a majority of participants explicitly gave a higher priority to the *answers* in all conditions, which indicates that a large part of the participants preferred to use common sense, instead of blindly relying on the visualization of the source credibility. It also shows that participants were almost equally divided when choosing for either the source or the visualization of its credibility. However, we did not find this to be significant.

According to [24], previous research has shown that credible sources are seen as likely to produce credible content, and credible contents are likely seen as to have originated from credible sources. The results of our study also show this clear tendency, where participants tend to select the highest ratings, however, this was also not significant.

5.4 Qualitative Feedback

In our studies, we also solicited user feedback on the experience and thoughts about the study they participated in. Making the source credibility ratings available significantly improved the confidence, however, some people do not rely blindly on it or understood it fully:

- *"The reliability meter seduces people to herd behavior."* (-Participant 26)

- *"For the first answer I trusted the Wikipedia and Amsterdam museum site, but since the answers for question 2 and further didn't make sense for these sites, I started looking only at the answers and not the sources."* (-Participant 31)

- *"I would explain the confidence bar more thoroughly at the start of the experiment. I understood it eventually, but it took some time."* (-Participant 72)

Some participants commented that the tasks given for both studies are difficult, especially the case with the many

sources and no visualization condition. Our results show that this low confidence is significantly dependent on the cultural heritage knowledge of a participant, which is lacking for most of our novice participants.

- *"...average respondent gets their culture from NGC or Discovery Channel for no longer than average 10 minutes a week."* (-Participant 12)

- *"Whoa. I don't know anything about Japanese or Chinese art. I might just as well done a random selection..."* (-Participant 108)

6. DISCUSSION AND CONCLUSION

This study contributes to the scant research on interaction with applications that aggregate information from multiple sources. We investigated how displaying source credibility ratings influences people's behavior when accessing information. Our findings also suggest that performance is negatively influenced when users have to select information from many sources as compared with few sources, as people relied less on the given information, were less confident, and needed significantly more time as the number of sources increased. We also found that presenting credibility ratings do boost the confidence level for novice users. By making the source credibility ratings available, users can select information more confidently, even though people do not blindly rely on it.

In our study, it took participants more time to choose an answer in the conditions where the credibility ratings of the source was displayed. We expected the participants to lean heavily on the credibility indication to select (what seemed to them) the most reliable information, thus reducing search time. This is, however, not the case. The credibility indications may add to the user's cognitive load, or even that participants enjoy this visualization and choose to examine it longer.

Our research has prompted many follow up questions for future work. There is still a lot to be investigated on the topic of credibility ratings. For example, source credibility measures across different domains; In our experiment, we have anchored credibility rating to the 4 credibility measures for the Cultural Heritage domain. It would be interesting if there were other measures that are domain specific for example for the health, law, news, or financial domains.

Moreover, we question whether these credibility measures are the same for novice as well as expert users. We know from previous study [1] that experts have their own standards on which information sources that they trust and use.

Another question is how we can elicit credibility assessments of sources in a cheap way, given the plethora online sources in numerous domains. Examples of credibility scores assessments are: by user votes, by dedicated credibility assessors, by most visited, or by most linked/referred, etc. Should there be different assessments for different domains?

Furthermore, in our study, we manipulated the actual information to make them all equally plausible, in the real world low credible information may come from a high credible source, and vice versa. It will be interesting to investigate the choice behavior when a user is confronted with contradictions.

Finally, visualization of credibility ratings; There are different ways to visualize credibility, such as presenting the scores as numbers, bars, stars or other visualizations. Alternatively, it is also possible to visualize a stamp of approval for information sources that passed certain criteria. In our research, we only considered credibility visualization in a list result type presentation. Different types of result presentation, such as map view in `grokker.com`, may require other credibility visualization techniques. This raise the question whether the type of credibility visualization matters at all and if so, how does it influences users' search behavior. As online search systems act more as intelligent information aggregators, source credibility issues will be more prominent and will demand more attention from the Web community.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] A. Amin, J. van Ossenbruggen, L. Hardman, and A. van Nispen. Understanding cultural heritage experts' information seeking needs. In Proc. *JCDL '08*, pages 39–47, New York, NY, USA, 2008. ACM.

[2] D. K. Berlo, J. B. Lemert, and R. J. Mertz. Dimensions for Evaluating the Acceptability of Message Sources. *Public Opin Q*, 33(4):563–576, 1970.

[3] W. P. Cassidy. Online news credibility: An examination of the perceptions of newspaper journalists. *Journal of Computer-Mediated Communication*, 12(2):478–498, 2007.

[4] T. Chesney. An empirical examination of wikipedia's credibility. *First Monday*, 11(11), Nov. 2006.

[5] J. H. Choi, J. H. Watt, and M. Lynch. Perceptions of news credibility about the war in iraq: Why war opponents perceived the internet as the most credible medium. *Journal of Computer-Mediated Communication*, 12(1):209–229, 2006.

[6] M. Craigie, B. Loader, R. Burrows, and S. Muncer. Reliability of health information on the internet: An examination of experts' ratings. *J Med Internet Res*, 4(1):2, Jan 2002.

[7] H. Cramer, V. Evers, S. Ramlal, M. Someren, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18(5):455–496, 2008.

[8] W. Duff, B. Craig, and J. Cherry. Historians' use of archival sources: Promises and pitfalls of the digital age. *Public Historian*, 26(2):7–22, 2004.

[9] M. S. Eastin. Credibility assessments of online health information: The effects of source expertise and knowledge of content. *Journal of Computer-Mediated Communication*, 6(4):0–0, 2001.

[10] E. Hyvonen, et al. Culturesampo – finnish cultural heritage collections on the semantic web 2.0. In *Proc. of the 1st International Symposium on Digital humanities for Japanese Arts and Cultures (DH-JAC-2009), Ritsumeikan, Kyoto, Japan*, March 2009.

[11] G. Eysenbach and C. Kohler. How do consumers search for and appraise health information on the

world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ*, 324(7337):573–577, 2002.

[12] A. J. Flanagin and M. J. Metzger. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media Society*, 9(2):319–342, 2007.

[13] B. J. Fogg and H. Tseng. The elements of computer credibility. In Proc. *CHI '99*, pp. 80–87, ACM.

[14] W. E. Hick. On the rate of gain of information. *The Quarterly Journal of Experimental Psychology*, 4(1):11–26, 1952.

[15] T. Hong. The influence of structural and message features on web site credibility. *J. Am. Soc. Inf. Sci. Technol.*, 57(1):114–127, 2006.

[16] C. Hovland and W. Weiss. The influence of source credibility on communication effectiveness. *The Public Opinion Quarterly*, 15(4):635–650, 1951.

[17] S. Kiousis. Public trust or mistrust? perceptions of media credibility in the information age. *Mass Communication & Society*, 4(4):381–403, 2001.

[18] J. Koenemann and N. J. Belkin. A case for interaction: a study of interactive information retrieval behavior and effectiveness. Proc. of the *SIGCHI conference on Human factors in computing systems*, pages 205–212, New York, NY, USA, 1996. ACM.

[19] D. A. B. Lindberg and B. L. Humphreys. Medicine and Health on the Internet: The Good, the Bad, and the Ugly. *JAMA*, 280(15):1303–1304, 1998.

[20] R. Lopes and L. Carriço. On the credibility of wikipedia: an accessibility perspective. In Proc. *WICOW '08*, pp. 27–34, 2008. ACM.

[21] M. J. Metzger, A. J. Flanagin, and L. Zwarun. College student web use, perceptions of information credibility, and verification behavior. *Comput. Educ.*, 41(3):271–290, 2003.

[22] V. Murdock and M. Lalmas. Workshop on aggregated search. *SIGIR Forum*, 42(2):80–83, 2008. ACM.

[23] G. Peterson, P. Aslani, and A. K. Williams. How do consumers search for and appraise information on medicines on the internet? a qualitative study using focus groups. *J Med Internet Res*, 5(4):33, Dec 2003.

[24] S. Y. Rieh and D. R. Danielson. Credibility: A multidisciplinary framework. In *Annual Review of Information Science and Technology*, pp. 307–364, 2007.

[25] G. Schreiber, A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. van Kersen, M. de Niet, B. Omelayenko, J. van Ossenbruggen, R. Siebes, J. Taekema, J. Wielemaker, and B. Wielinga. Multimedian e-culture demonstrator. In Proc. *ISWC 2006*, vol. 4273 of *LNCS*, pages 951–958. Springer, 2006.

[26] S. S. Sundar and C. Nass. Conceptualizing sources in online news. *Journal of Communication*, 51(1):52–72, 2001.

[27] S. Tseng and B. J. Fogg. Credibility and computing technology. *Commun. ACM*, 42(5):39–44, 1999.

[28] Y. Tsfati and J. N. Cappella. Why do people watch news they do not trust? the need for cognition as a moderator in the association between news media skepticism and exposure. *Media Psychology*, 7(3):251–271, 2005.

[29] B. J. Walther, Z. Wang, and T. Loh. The effect of top-level domains and advertisements on health web site credibility. *J Med Internet Res*, 6(3), Sep 2004.

[30] J. Zhang, K. N. Fachry, and J. Kamps. Access to archival finding aids: Context matters. In Proc. *ECDL '08*, pages 455–457, Berlin, Heidelberg, 2008. Springer-Verlag.