
Different users-Different support?

Support systems for the evaluation of credibility and the role of
familiarity

BACHELOR THESIS

Lotta Schulze

s0210382

University of Twente

Department of Cognitive Psychology and Ergonomics

16-12-2011

1st supervisor: Teun Lucassen, MSc.

2nd supervisor: Prof. Dr. Jan Maarten Schraagen

Abstract

Open access encyclopedias such as Wikipedia have gained an increasing importance for the information search in the past ten years. Both, the amount of information provided and number of users have increased with the growing importance of the World Wide Web (WWW). Every user has the opportunity to upload information. This huge amount of unsupervised information leads to problems with the evaluation of its credibility. Regular users have to distinguish between information concerning its credibility. This distinction asks skills from regular users they mostly do not have developed completely. Support systems can help users evaluating the credibility of information. The purpose of this study was to examine the relation of familiarity of users with a topic, its influence on the users' method of evaluating credibility and two different support systems. One support system incorporated surface features as an indicator for credibility. The other support system used semantic comparisons with other websites as an indicator for credibility.

Fourty academic psychology students were divided in two groups. One group got familiar topics of the presented articles and the other group got unfamiliar articles. Both groups evaluated five Wikipedia articles. The articles were presented with either positive or negative advice of one of the support systems and for the last article participants chose one support system. A questionnaire which asked trust evaluations about the article and the presented support system as well as the influence of the advice was completed by the participants after each article. Participants were asked to give motivations for their answers.

The unfamiliar group showed significantly more trust and influence of the support system which incorporated surface features as an indicator for credibility than by a support system which used semantic comparisons. The familiar group showed no significant differences between the two support systems.

Users who differ in their familiarity with a topic incorporate different elements of information to evaluate credibility. Support systems that incorporate the same features as the users in their evaluation of credibility were trusted more and had more influence.

Samenvatting

Online encyclopedieën zoals Wikipedia zijn ontwikkeld waar elke gebruiker de mogelijkheid heeft informatie te uploaden. Deze enorme hoeveelheid ongecontroleerde informatie leidt tot problemen met betrekking tot de evaluatie van haar betrouwbaarheid. Regelmatige gebruikers moeten onderscheiden tussen betrouwbare en onbetrouwbare informatie. Daarvoor hebben gebruikers vaardigheden nodig die ze meestal nog niet volledig ontwikkeld hebben. Ondersteuningssystemen kunnen helpen om informatie op haar betrouwbaarheid te evalueren. Doel van het onderzoek was de samenhang tussen bekendheid van gebruikers met een onderwerp, de invloed ervan op de wijze van gebruikers om betrouwbaarheid te evalueren en uiteindelijk de gevolgen daarvan voor twee verschillende ondersteuningssystemen te onderzoeken. Het eerste ondersteuningssysteem gebruikte oppervlakkige kenmerken van informatie voor een betrouwbaarheidsevaluatie. Het andere ondersteuningssysteem gebruikte inhoudelijke vergelijkingen met andere websites om de betrouwbaarheid van informatie te evalueren.

Veertig academische psychologie studenten werden verdeeld over twee groepen. Deze groepen verschilden in de onderwerpen van de aangeboden artikelen. Een groep kreeg artikelen met bekende onderwerpen en de andere groep artikelen over onbekende onderwerpen. Beide groepen hebben vijf Wikipedia artikelen geëvalueerd. De artikelen werden gepresenteerd met een positief of negatief advies van een van de ondersteuningssystemen. Voor de laatste artikel iedereen koos een van de ondersteuningssystemen. Na elk artikel werd een vragenlijst ingevuld om het vertrouwen in het artikel en het ondersteuningssysteem te bepalen. Bovendien werd naar de invloed van het ondersteuningssysteem en naar motivaties voor hun antwoorden gevraagd. De groep die onbekend was met de onderwerpen in de artikelen toonde significant meer vertrouwen in en invloed door het ondersteuningssysteem dat oppervlakkige eigenschappen als een indicator voor betrouwbaarheid gebruikt dan in een ondersteuningssysteem dat inhoudelijke vergelijkingen gebruikt. De groep die bekend was met de onderwerpen in de artikelen toonde geen significante verschillen tussen de twee ondersteuningssystemen.

Gebruikers die in hun bekendheid verschillen met het onderwerp gebruiken verschillende elementen van informatie om de betrouwbaarheid te evalueren. Ondersteuningssystemen die dezelfde kenmerken als de gebruikers opnemen in hun evaluatie van de betrouwbaarheid van informatie werden als betrouwbaarder ervaren en hadden meer invloed.

Content

1. Introduction	5
1.1. Problem identification	6
1.2. Existing support systems	7
1.3. Definitions of trust and credibility	9
1.4. Definition of familiarity	9
1.5. Theoretical framework	9
2. Method.....	12
2.1. Participants	12
2.2. Design and procedure.....	12
2.3. Independent variables.....	13
2.3.1. Familiarity	13
2.3.2. Support systems.....	14
2.3.3. Advice of the support system	15
2.4. Dependent variables	16
2.4.1. Manipulation check of familiarity.....	16
2.4.2. Trust in the article	16
2.4.3. Trust in the support system	16
2.4.4. Influence of the support system	16
2.4.5. Preference of a support system.....	17
2.4.6. Motivations of participants.....	17
2.5. Data analysis	17
3. Results	18
3.1. Manipulation check of familiarity.....	18
3.2. Preference of a support system.....	18
3.3. Trust in the support system	20
3.4. Influence of the support system	22

4. Discussion	26
4.1. Unfamiliar group	26
4.2. Familiar group	27
4.3. Limitations and future research.....	28
4.4. Conclusion.....	30
Addendum.....	31
References	33
Appendix A	36

1. Introduction

The World Wide Web (WWW) as an international exchange platform of information was developed in the late 80s and published in 1991. Since then, the WWW is growing strongly. Nowadays almost everyone uses the different services of information-gathering and -exchange. In particular, the direct and easy access to all kinds of information makes the WWW very attractive to the people all over the industrialized world. In 2010, it has been shown that 67.6% of the population of the European Union use the Internet. All over the world an increase in the WWW-use of 444% from 2000 to 2010 was reported (Internetworldstat, 2010).

The WWW offers various ways of information-gathering and -exchange. Examples are social networks, blogs or online encyclopedias. This study researches the information credibility of online encyclopedias. A core element of online encyclopedias is the open access, which means that everybody has the opportunity to upload information to be included in online encyclopedias. Online encyclopedias are frequently used to gather information. The most popular online encyclopedia today is Wikipedia which was developed in 2001. Since then the amount of articles and number of users is constantly growing (Lih, 2004). Wikipedia consists nowadays of more than 19 million articles in more than 280 languages¹. Wikipedia is a user-based website which is of open access for every user. Everyone has the opportunity to publish or modify information without an examination of its accuracy. While most articles on Wikipedia present correct information, some articles are published which present wrong information or a limited perspective on a certain topic. On the one hand, the absence of supervision offers the opportunity for fast and direct modification of information. On the other hand, this freedom leads to a lot of unsupervised information in the WWW and can be easily abused (Voss, 2005). Unsupervised information can be credible or less credible. This evaluation requires skills from users which they mostly do not have developed thoroughly (Brand-Gruwel, Wopereis & Vermetten, 2005). It is important to help users to distinguish between credible and less credible information.

Appropriate support systems which give an advice on the credibility of the information are one opportunity. These support systems distinguish credible and less credible information based on different features.

¹ Wikipedia. (2011). Retrieved 22 July 2011, from <http://en.wikipedia.org/wiki/Wikipedia>

This study investigates which features of information are incorporated by two types of users in their credibility evaluation and its consequences for different support systems. The users differ in their familiarity with various topics.

Familiarity is expected to be an important characteristic of users and their information-gathering process in the WWW because it influences which elements of information are incorporated as an indicator for credibility. Users with prior knowledge on a topic can consider the accuracy of presented information through comparisons to their own knowledge. They incorporate semantic elements of information. Users without knowledge on a topic are not able to make such comparisons. They circumvent this limitation through incorporating surface features such as references or text structure (Lucassen & Schraagen, 2011a).

In conclusion, familiarity influences the users' method of evaluating information and which elements of information are incorporated as an indicator for credibility. Therefore, this study manipulates familiarity to examine its consequences for a support system that includes semantic comparisons as an indicator for credibility and a support system that incorporates surface features as an indicator for credibility. The first support system will be designed towards users who have some knowledge on the topic, the other aims at users without prior knowledge. This study attempts to answer the following research question:

How does familiarity influence the preference, trust and influence of a support system?

The aim of this study is to investigate if users who differed in their familiarity with the topics show a difference in the preference, trust and influence of two different support systems.

With the enlightening of the relation between familiarity and the preference, trust and influence of different support systems, appropriate support systems can be introduced to make information-gathering via the WWW much safer and easier for regular users.

1.1. Problem identification

In the last two decades, our society has transformed in an information-dependent society. In this society, everyone has access to a huge amount of information on the WWW. This information can concern any circumstance, ranging from health topics, to job-related themes and trivial information. A study in 2009 has shown that most Americans are searching via the WWW about the symptoms of the swine flu pandemic (Allen, 2009). The transformation of the society as well as of the WWW evolved very quickly. As described earlier, this

transformation asks appropriate skills to work adequately with the information presented which a lot of users have not developed. Lazonder & Rouet (2008) have shown that students frequently use the web to gather information but their skills to organize and evaluate the presented information with respect to credibility and quality were not proficient.

These problematic consequences of the change in the information-gathering process via the WWW have strengthened the scientific interest in the WWW and its interaction with regular users. Several studies have aimed at an improvement of the interaction between users and the WWW, especially the importance of cognitive skills to search and distinguish information appropriately (Metzger, Flanagan & Zwarun, 2003; Lazonder & Rouet, 2008). Human factors are taken into account for a better interaction and use of the advantages the WWW offers (Chen & Macredie, 2010).

Today, research focuses on the direct interaction between regular users and credible information on the WWW (Metzger, 2007). The need to develop support systems which help users evaluating the credibility of information is increasingly recognized. It is important to consider the user characteristics in the development of support systems (Lucassen & Schraagen, 2010).

1.2. Existing support systems

Different support systems exist which should help Wikipedia editors or end-users to evaluate the credibility of information. These following support systems are either designed for Wikipedia editors, end-users or both. A strict distinction is difficult because editors are often also end-users and vice versa. A short summary of prominent support systems for Wikipedia follows.

The *WikiScanner* traces IP-addresses from authors who are editing or changing articles anonymously. Through this process anonymous authors can be identified who are editing self-serving articles. These authors' sections are deleted and their IP-addresses are banned. The *WikiScanner* was developed to detect vandalism in Wikipedia. As a result, quality of Wikipedia articles can be enhanced (Potthast, Stein & Gerling, 2008). This system is designed for Wikipedia editors rather than end-users.

WikiDashboard shows important patterns in the history of articles to improve the transparency of Wikipedia for regular users. A dashboard at every site from Wikipedia shows

patterns of re-editions of the article, the different authors who work on the article and their activity in the last week (Suh, Chi, Kittur & Pendleton, 2008)

WikipediaViz is a program that visualizes five important features of an article according to its quality and credibility. These five features include the number of words of the article, the number of contributors and their length of contribution, the number and lengths of edits, the number of references and internal links and, last, the length and activity of the discussion (Chevalier, Huot & Fekete, 2010).

WikiTrust colours the background of each word of the articles according to its credibility. The background of credible words is white whereas the background of non-credible words is dark orange. The credibility of words depends on its “survival duration” which is a number to indicate the time the words are added to an article and remain unchanged. Newly added words are assigned to the credibility of their author. Author’s credibility is based on his average credibility of edits. *WikiTrust* is based on the assumption that errors are quickly detected and enhanced. Therefore, older words which are not changed are more credible (Adler, Chatterjee, de Alfaro, Faella, Pye, & Raman, 2008).

Lucassen & Schraagen (2011b) have shown that users are influenced by *Wikitrust* but they did not evaluate the system as having added value. Furthermore, its usefulness was questioned because the coloured background led to reading difficulties and uncertainty what to do with the information from *WikiTrust*. A study in 2006 has shown that a “history-based trust model” could predict the quality of articles (Zeng, Alhossaini, Ding, Fikes & McGuinness, 2006).

In conclusion, the described support systems incorporate features of Wikipedia, such as its open character for everyone which lead to several authors of an article. The history of the article shows the diversity of authors as well as the duration-time of different sections. In the development of the existing support systems, human factors are not taken into account as far as the usefulness and handling of the information for regular users is not concerned.

One limitation of these different support systems is that they are mainly based on the way Wikipedia works. Another approach for developing useful support systems may be to focus on regular users, their characteristics and their evaluation of credibility. This study introduces two different support systems that are based on the users’ method to evaluate the credibility of information, independent from the way Wikipedia works. Users’ characteristics can play an important role in the evaluating process of support systems and their usefulness. This study

examines the familiarity of users with a topic and its consequences for the two presented support systems.

1.3. Definitions of trust and credibility

The definitions of trust and credibility are not consistent in literature and different authors use different definitions. For this study, credibility is defined as the “believability” of information (Metzger, 2007). Trust refers to two dimensions: the believability of information and based on this, the intention to use the information. On the one hand, trust includes the risk of trusting a wrong source. On the other hand, trust includes the possibility of gaining knowledge (Simon, 2010).

1.4. Definition of familiarity

The definition of familiarity is also not consistent in literature and a lot of authors incorporate the term familiarity and expert as similar as the term domain knowledge but these three concepts have different concerns. Familiarity refers to the prior knowledge someone has about different topics or different domains (“novices”). Familiarity can become domain knowledge through focusing on a specific domain of prior knowledge. Domain knowledge includes the facts that someone knows about a specific topic within a domain (“apprentices”). The development of becoming an expert in a specific domain demands years of structural analysis of a specific domain (Alexander, 1992; Feltovich, Pritula & Ericsson, 2006). In conclusion, these three concepts have a hierarchical structure. This study focuses on the familiarity with different topics in different domains.

1.5. Theoretical framework

Various models describe the cognitive way of people making evaluations of the credibility of information. One prominent model was suggested by Metzger (2007) based on the dual-process model of Chaiken (Chaiken, 1980). Both models incorporate characteristics of users as a crucial factor for the decision which elements of information are used to make an evaluation.

The dual-process theory from Chaiken distinguishes two routes of making an evaluation: The heuristic route and the systematic route. The heuristic way of evaluation is based on superficial cues of the information. Simple, effortless heuristics are used to make an evaluation. The systematic way of evaluation refers to an effortful way of thinking based on profound features of the information. Different features of information are incorporated in the

evaluation process (Todorov, Chaiken & Henderson, 2002). Metzger has added two important characteristics of users which influence the information-processing: the motivation to take an effortful way to take a decision and the ability of users to evaluate credibility correctly. Motivation depends on the importance of information whereas ability depends on the skills users have developed to handle information from the web (Metzger, 2007).

These models show that users with different levels of motivation and ability incorporate different elements of information as an indicator of credibility. Lucassen en Schraagen (2011a) incorporated three other crucial factors of users which greatly influence how users deal with information: Familiarity with a topic, information skills and source experience. Their 3S model emphasizes in which way different levels of these factors influence the evaluation process of credibility. Familiar users incorporate semantic features of information which they can compare to prior knowledge. Semantic features refer for example to the accuracy, completeness, scope and neutrality of presented information (Lucassen & Schraagen, 2011a). Users unfamiliar with the subject matter are not able to evaluate semantic features through making comparisons with their own knowledge. To deal with that, they are focusing on surface features of information. Surface features refer for example to the length of the information, references, pictures and the writing style (Lucassen & Schraagen, 2011a). Unfamiliar as well as familiar users can be influenced by earlier experiences they have done with the source of the information. This can bias the evaluation process of users because they do not incorporate information features to evaluate credibility. They evaluate the information dependent on its source.

Research has shown that academic students have good information skills and know the importance of information features such as references (Lucassen & Schraagen, 2010). This study uses academic students as respondents. Thus, a good level of information skills can be expected but the familiarity with the topic varies. Furthermore, this study uses only one source of information to control the influence of experiences with various sources.

In conclusion, users who differ in their familiarity with a topic incorporate different features of information to evaluate its credibility. Hypothetically, this influences not only the way information is evaluated but also the way users deal with advice of different support systems. Familiar users incorporate semantic features of information to evaluate credibility. We assume that familiar users prefer a support system that uses semantic features as indicators for credibility of information because this system compares the presented information with other

information from different websites. Thus, this system incorporates similar semantic elements of information as familiar users in the evaluation process of credibility. Through this agreement of used features in credibility evaluation, familiar users have more trust in such a support system and are more influenced by its advice.

This leads to the following hypothesis:

Hypothesis 1: Users who are familiar with the topic a) prefer b) have more trust in c) are more influenced by a support system that incorporates semantic comparisons as an indicator for the credibility of information than by a support system which uses surface features as an indicator for the credibility of information.

According to the 3S model, users who are unfamiliar with a topic use their information skills to evaluate the credibility of information (Lucassen & Schraagen, 2011a). We assume that unfamiliar users prefer a support system that uses surface features as an indicator for credibility because the used features of this support system match with the features users incorporate on their own. Through this agreement of features used in credibility evaluation, unfamiliar users have more trust in this support system and are more influenced by its advice.

This leads to the following hypothesis:

Hypothesis 2: Users who are unfamiliar with the topic a) prefer b) have more trust in c) are more influenced by a support system that incorporates surface features as an indicator for the credibility of information than by a support system that incorporates semantic comparisons as an indicator for the credibility of information.

In conclusion, we assume that users prefer, trust and are more influenced by a support system which matches with their own method of credibility evaluation.

2. Method

2.1. Participants

This study was conducted with 40 academic psychology students. They were enrolled in average 2.94 ($SD = 1.66$) years at a Dutch university. Their average age was 22.3 ($SD = 2.01$). 28 of them were German and 12 were Dutch. All participants had a high proficiency in English as well as in the Dutch language. On the one hand, students were used as participants because they were easily recruited ('convenience sample'). On the other hand, students used Wikipedia frequently (Head & Eisenberg, 2010). Therefore, all participants are familiar with Wikipedia and the way the online encyclopedia works. Informed consent was obtained from each of the participants prior to the experiment.

2.2. Design and procedure

The experiment was implemented using Lime Survey, a tool with which questionnaires were developed. The participants were assigned to one of two groups: familiar group ($N=20$) or unfamiliar group ($N=20$). The two groups differed in the topics of the articles which are included in the questionnaire. One group received articles with probably familiar topics whereas the other group received articles with probably unfamiliar topics. Both groups followed the same experimental procedure. Every participant evaluated in total five articles which were screenshots from original Wikipedia articles. All articles presented were from "Start-Class" quality. The Wikipedia Editorial Team Assessment has ranked articles from "Stub" articles which were very low in quality to "FA" articles which were very high in quality. 721,344 articles from the English version are ranked as Start-class articles². Start-class articles are not completely developed and require further revision. This led to ambiguity in the credibility of the articles because they had trustworthy characteristics (e.g. scientific references) as well as untrustworthy characteristics (e.g. poor writing style). This resulted in a difficulty to evaluate the credibility of the article on its own.

The questionnaires of the experiment were prepared in Dutch. The articles from Wikipedia were in English because the quality of articles from the English version was ranked through the Wikipedia Editorial Team Assessment. The experiment started with a questionnaire about demographics and their experiences with Wikipedia. A list of the topics of the five articles

² Wikipedia. (2011). Retrieved 22 July 2011, from <http://en.wikipedia.org/wiki/Wikipedia:1.0/A>

was presented. To check the manipulation of familiarity through the different topics of presented articles all participants were asked to which extent they were familiar with the topics of the articles on a 7-point Likert scale. Prior to the presentation of the articles, the two support systems used in this experiment were explained. In addition, this explanation of the support systems was given to the participants on paper. This paper could be used during the experiment.

All participants evaluated five articles according to its credibility. The first two articles were presented with the DisputeFinder support system. The next two articles were presented with the WikiCheck support system (the two support systems are explained in a following paragraph). One of the articles in both conditions was evaluated negatively through the support system and the other positively. For the last article, participants had to choose a support system. After each article, a questionnaire followed in which the participants were asked their trust in the article, their trust in the support system and their influence of the support system. The answers were given on 7-point Likert scales. Open questions asked motivations for their answers.

2.3. Independent variables

2.3.1. Familiarity

Familiarity is a characteristic of the articles which was manipulated through the choice of topics of the presented articles. The topics of the articles are presented in Table 1. The topics of the articles were chosen from various topics to be a representative sample and did not include country-specific topics. Country-specific topics include for example the Dutch royal family or the German political system. Through the nationality of the participants, they differed in their knowledge about their own country. This different knowledge could bias the manipulation of familiarity. All articles were received from the English Wikipedia on May, 2nd 2011. Articles used in this experiment were selected based on their “Start-Class” quality. Within “Start-class” articles, topics were chosen that were either study-related (e.g. mental model), age-related (e.g. Tamagotchi, Online community and Austin Powers in Goldmember) or general definitions (e.g. Abstinence). Unfamiliar “Start-Class” articles were selected within very specific domain topics which include no topics of general knowledge.

Table 1 Topics of the presented articles

Articles of the familiar group	Articles of the unfamiliar group
Abstinence	Chinoiserie
Austin Powers in Goldmember	Bob Black
Online community	Ruben's tube
Mental model	Richard Evan Schultes
Tamagotchi	Esoteric Cosmology

2.3.2. Support systems

Two different support systems were presented. One of the systems aimed at the familiar group and one of the systems aimed at the unfamiliar group.

The first support tool, named WikiCheck, worked with a complicated computation which took numerous surface features of the articles into account. These features included for example the history of the articles, the references or the writing style. This support system was based on the study by Lucassen & Schraagen who studied the use of three different support tools (Lucassen & Schraagen, in preparation). One of these support systems incorporates the same elements of information as the WikiCheck support system. Results of this study showed that the trust and influence of the complicated system (as the WikiCheck system) was higher than the trust and influence of a heuristic support system. Therefore, this system was used in this study in comparison with another support system.

The second support-tool, named DisputeFinder, compared the presented information from the article with information from other websites about the topic. This support tool was based on the idea of Ennals, Byler, Agosta & Rosario (2010). They proposed that the use of different websites could simplify the gathering of credible information. The DisputeFinder searched on other websites if information from the Wikipedia article was disputed. Through searching specific patterns as “falsely claimed that...” or “the misconception that...” a statistical value was calculated to give an advice about the credibility of Wikipedia articles.

The participants got an explanation of the two support systems. These explanations were as following:

“DisputeFinder: This system searches on other websites if the information from Wikipedia is disputed. It searches for patterns as “falsely claimed that...” or “the misconception that...”. Through calculating the found “disputed claims” positive or negative advice about the credibility of the article is given” (translated from a Dutch description, see Appendix A).

“WikiCheck: This system uses an adaptive neural network that incorporates different aspects of the article (for example: the authors of the article, the history and the references) to evaluate the credibility of the article. Based on this algorithm a positive or negative advice is given” (translated from a Dutch description, see Appendix A)

The two support tools were simulated and did not really exist. They were presented to the participants as if they worked.

2.3.3. Advice of the support system

The advice of the support systems was either positive or negative. The support tools were automatic and gave either an orange exclamation mark with an explanation as negative advice or a green hook with an explanation as positive advice about the credibility of the article. Figures 1 a-d show examples of such advices.

Figure 1a Information box containing positive advice of the DiputeFinder



Figure 1b Information box containing negative advice of the DisputeFinder



Figure 1c Information box containing positive advice of the WikiCheck

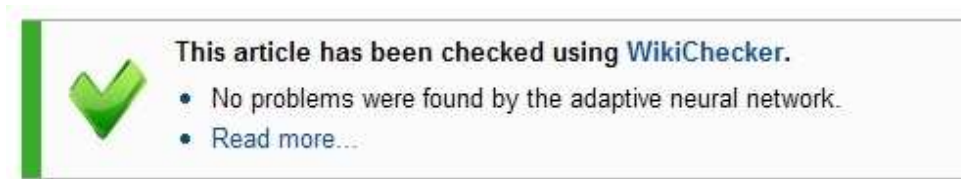
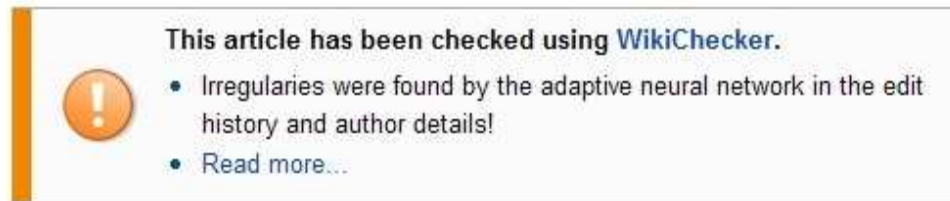


Figure 1d Information box containing negative advice of the WikiCheck



2.4. Dependent variables

2.4.1. Manipulation check of familiarity

The dependent variable of familiarity examined the manipulation of familiarity between the two groups. The participants assessed their familiarity with the topics of the articles through answering to which extent they were familiar with the presented topics on a 7-point Likert scale.

2.4.2. Trust in the article

The participant's trust in the article was measured after each article by a 7-point Likert scale. The mean scores of each participant were added up to measure their overall trust in the articles. The mean scores were analyzed for each group.

2.4.3. Trust in the support system

The trust in the presented support system was measured after each article by a 7-point Likert scale. The mean trust scores were analyzed for each group.

2.4.4. Influence of the support system

Participants rated their influence of the advice of a support system on a 7-point Likert scale. Furthermore, their trust in the article when one support system showed either positive or negative advice is analyzed. The mean scores for each group were used for further analysis.

2.4.5. Preference of a support system

The choice which support system was used for the last article was measured on a dichotomous variable with two options. The first option was the WikiCheck support system and the second option was the DisputeFinder support system.

2.4.6. Motivations of participants

The participants were asked to give motivations for their preference of a support system and their trust in the article as well as in the support system. These motivations were classified and the main categories were analyzed through Chi Square-tests. Motivations which did not fit were categorized as “Others”. For each group, 25% of the questionnaires were analyzed through two raters to calculate their agreement and improve the reliability of the classification. For the unfamiliar group, the calculated kappa was 0.88 and for the familiar group 0.91 was calculated. This indicated a good agreement. The classifications of the motivations were presented in table 3, 5 and 8.

2.5. Data analysis

The variables were mostly measured by 7-point Likert scales. These scales measured presumably at an ordinal level. Therefore, the variables were tested through non-parametric tests. Data analysis is done with statistical software (SPSS PASW Statistics 18).

3. Results

3.1. Manipulation check of familiarity

The experiment manipulated the familiarity of the participants through the topics of the presented articles. To check the manipulation of this variable, a Mann-Whitney test for the two groups was used. The familiar group ($M = 3.0$, $SD = 0.94$) showed a significantly higher level of familiarity than the unfamiliar group ($M = 1.19$, $SD = 0.35$); $Z = -3.38$, $p = 0.00$.

The following sections describe the different aspects of the hypotheses, namely preference of a support system, trust in a support system and influence of a support system.

3.2. Preference of a support system

Chi-Square-tests were used to analyze the preference of either support system. The unfamiliar group showed a trend towards a significant preference of the WikiCheck support system ($\chi^2(1) = 3.2$, $p = 0.07$). No significant preference was found for the familiar group ($\chi^2(1) = 0.20$, $p = 0.66$). Table 2 presents the number of preferences of each support system for the unfamiliar and the familiar group.

Table 2 Preference of a Support System

Support system	Unfamiliar group	Familiar group
WikiCheck	14	9
DisputeFinder	6	11

The open motivations showed that 90% of the unfamiliar group and 80% of the familiar group motivated their preference through the used criteria of the support systems. Table 3 presents the percentages and number of motivations.

For the unfamiliar group, a significant difference between the positive evaluation of the used criteria of the DisputeFinder and the positive evaluation of the used criteria of the WikiCheck was found ($\chi^2(1) = 4.77$, $p = 0.03$). A positive evaluation of the used criteria of the WikiCheck was mentioned more (65% versus 20%). 10% of the unfamiliar group motivated their preference with the dependence of WikiCheck on Wikipedia: “Wikicheck is probably influenced by Wikipedia.”

No significant difference in the familiar group was found between the positive evaluation of the used criteria of the DisputeFinder or the WikiCheck ($\chi^2(1) = 0.34, p = 0.56$). Examples of motivations are: “I think it is a good idea that the DisputeFinder compares the information with other websites” or “I think the WikiCheck gives better advice, because it takes different details into account”. Negative evaluations of the used criteria of the support systems were also not significant ($\chi^2(1) = 1.00, p = 0.32$). 10% of the familiar group motivated their preference to which extent the support systems are influenced by Wikipedia, for example: “WikiCheck is a program from Wikipedia, so I have chosen the DisputeFinder because this is an independent program”. The other 10% would prefer no support system.

Table 3 Percentages and number of motivations of both groups for their preference of a support system

	Percentages	<i>N</i>
Unfamiliar group		
Positive criteria of WikiCheck	65%	13
More details used	50%	10
Objectivity	10%	2
algorithm	5%	1
Positive criteria of DisputeFinder	20%	4
Semantic comparison	10%	2
Traceability	10%	2
Negative criteria of WikiCheck	0%	0
Negative criteria of DisputeFinder	5%	1
Reliability of other websites	5%	1
Dependence on Wikipedia	10%	2
Familiar group		
Positive criteria of WikiCheck	25%	5
More details used	25%	5
Positive criteria of DisputeFinder	35%	7
Semantic comparison	20%	4
Traceability	15%	3
Negative criteria of WikiCheck	5%	1
Unreliable criteria	5%	1
Negative criteria of DisputeFinder	15%	3
Reliability of other websites	10%	2
searches only disputed claims	5%	1
No support system used	10%	2
Dependence on Wikipedia	10%	2

The found results did not confirm hypothesis 1a which stated that the familiar group preferred a support system that incorporated semantic comparisons as an indicator for credibility over a surface-based support system. Hypothesis 2a that the unfamiliar group preferred a support system that used surface features as an indicator for credibility over a semantic-based support system was rejected.

3.3. Trust in the support system

Wilcoxon signed-rank tests showed that the unfamiliar group trusted the WikiCheck support system more than the DisputeFinder support system ($Z = -2.08, p = 0.02$). Table 4 presents the means and standard deviations of both groups. The familiar group showed no significant difference in trust in either the DisputeFinder or the WikiCheck support system ($Z = -0.24, p = 0.4$).

Table 4 Means and standard deviations of both groups for the trust in the support system

	<i>Means</i>	<i>SD</i>
<i>Unfamiliar group</i>		
Trust in Wikicheck	5.15*	0.99
Trust in DisputeFinder	4.55*	1.21
<i>Familiar group</i>		
Trust in Wikicheck	4.2	1.33
Trust in DisputeFinder	4.35	1.18

* $p < 0,05$ (significant difference)

Open questions about their motivations to trust a support system demonstrated that the used criteria of the support systems were important for both groups. Understandability was another factor noted. Table 5 presents the percentages and number of motivations of both groups.

The unfamiliar group evaluated the criteria of the WikiCheck significantly more positively than the used criteria of the DisputeFinder ($\chi^2(1) = 6.74, p = 0.01$). The used criteria of the DisputeFinder were evaluated significantly more negatively than the used criteria of the WikiCheck ($\chi^2(1) = 9.00, p = 0.00$). One participant noted: "If I understand it right, the DisputeFinder searches everywhere on the web to find conflicting information. I think that this is always for every article the case." 19% criticized the understandability of the support

systems and 8% evaluated their trust in the support system through the conformity of the advice with their own impression of the article.

The familiar group showed neither a significant difference in the positive evaluation of the used criteria of the two support systems ($\chi^2(1) = 0.14, p = 0.71$) nor a significant difference in the negative evaluation of the used criteria of the support systems ($\chi^2(1) = 3.00, p = 0.09$). Examples are “The DisputeFinder seems to be credible because it searches on the whole web” or “The adaptive neural network of WikiCheck is quite large, it investigates references etc which is quite important for the credibility.” 28% criticized the understandability of the two support systems and 11% evaluated their trust in the support system through the comparisons of the advice with their own knowledge.

Table 5 Percentages and number of motivations of both groups for their trust in the support system

	Percentages	<i>N</i>
Unfamiliar group		
Positive criteria of WikiCheck	27%	27
More details used	10%	10
Focus on references	7%	7
Focus on characteristics of the article	4%	4
Used algorithm	2%	2
Internal comparison	4%	4
Positive criteria of DisputeFinder	11%	11
Comparison with independent websites	8%	8
Presentation of advice	3%	3
Negative criteria of WikiCheck	5%	5
Dependence on Wikipedia	1%	1
Unreliable criteria	4%	4
Negative criteria of DisputeFinder	20%	20
Reliability of other websites	8%	8
Difficulty for controversial topics with contradicting statements	12%	12
Understandability	19%	19
Other websites used for the advice are unknown	7%	7
Reasons for the advice are not comprehensible	9%	9
Number of contradictions unknown	3%	3
Conformity with own impression	8%	8
Others	10%	10

Familiar group		
Positive criteria of WikiCheck	15%	15
More details used	10%	10
Characteristics of the article used	5%	5
Positive criteria of DisputeFinder	13%	13
Comparisons with the whole web	6%	6
Looks for contradicting statements	6%	6
Scientific presentation of advice	1%	1
Negative criteria of WikiCheck	9%	9
Unreliable criteria used	5%	5
Too much attention for references	4%	4
Negative criteria of DisputeFinder	18%	18
Trivial websites used	3%	3
Reliability of other websites	10%	10
Difficulty for controversial topics with contradicting statements	5%	5
Understandability	28%	28
Unobvious method of support systems	18%	18
Type of irregularities unknown	10%	10
Comparisons of the advice with own knowledge	11%	11
Others	6%	6

These results confirmed aspect b of hypothesis 2: The unfamiliar group had more trust in a support system that incorporated surface features as an indicator for credibility than in a semantic-based support system. Hypothesis 1b which stated that the familiar group had more trust in a support system that incorporated semantic comparisons as an indicator for credibility than in a surface-based support system was rejected.

3.4. Influence of the support system

Wilcoxon signed-rank tests showed that the unfamiliar group had significantly less trust in the article when the WikiCheck support system showed negative advice in comparison to positive advice of the WikiCheck support system ($Z = -3.66, p = 0.00$). Table 6 presents the means and standard deviations of both groups. A significant difference was also found when the DisputeFinder support system showed positive advice compared with negative advice of the DisputeFinder. The unfamiliar group had significantly less trust in the article when the Dispute Finder showed negative advice in comparison with positive advice of the DisputeFinder ($Z = -2.13, p = 0.02$). Furthermore, the unfamiliar group had significantly more trust in the article when the WikiCheck support system showed positive advice in comparison

to positive advice of the DisputeFinder ($Z = -2.15, p = 0.02$). The unfamiliar group had significantly less trust in the article when the WikiCheck support system showed negative advice in comparison to negative advice of the DisputeFinder ($Z = 1.79, p = 0.04$).

The familiar group did not show more trust in the article when the DisputeFinder showed positive advice compared with negative advice from the DisputeFinder ($Z = -1.44; p = 0.08$). The familiar group showed more trust in the article when the WikiCheck support system showed positive advice then when the WikiCheck support system showed negative advice ($Z = -2.1, p = 0.02$). Furthermore, the familiar group did not show significantly more trust in the article by positive advice of the WikiCheck support system in comparison to positive advice of the DisputeFinder support system ($Z = -0.59, p = 0.28$). A significant influence on the trust in the article was also not found when the WikiCheck support system showed negative advice compared with negative advice of the DisputeFinder ($Z = -0.49, p = 0.31$).

Table 6 Means and standard deviations of both groups for the trust in the article

	<i>Means (trust in article)</i>	<i>SD (trust in article)</i>
<i>Unfamiliar group</i>		
Positive advice of WikiCheck	5.75*	0.72
Negative advice of WikiCheck	3.85*	1.27
Positive advice of DisputeFinder	5.25*	1.02
Negative advice of DisputeFinder	4.55*	1.02
<i>Familiar group</i>		
Positive advice of WikiCheck	5.45* ₁	1.0
Negative advice of WikiCheck	4.65* ₂	1.43
Positive advice of DisputeFinder	5.15	1.5
Negative advice of DisputeFinder	4.45	1.1

* $p < 0,05$ (significant differences) *₁ $p < 0,05$ (significant difference only in comparison to negative advice of WikiCheck) *₂ $p < 0,05$ (significant difference only in comparison to positive advice of Wikicheck)

These results were strengthened by the self-report of the participants to which extent they were influenced by the support systems. Table 7 presents the means and standard deviations of both groups.

Wilcoxon signed-rank tests showed that the unfamiliar group scored significantly higher on the influence of the WikiCheck support system than on the influence of the DisputeFinder ($Z = -2.53, p = 0.01$).

The familiar group showed no significant differences in their self-reported influence of the two support systems ($Z = -0.35, p = 0.37$).

Table 7 Means and standard deviations of both groups for their self-reported influence by the support systems

	<i>Mean</i>	<i>SD</i>
<i>Unfamiliar group</i>		
Self-reported influence WikiCheck	4.92*	0.88
Self-reported influence DisputeFinder	4.08*	1.13
<i>Familiar group</i>		
Self-reported influence WikiCheck	3.33	1.63
Self-reported influence DisputeFinder	3.5	1.44

* $p < 0,05$ (significant differences)

Open motivations for their trust in the articles demonstrated that the unfamiliar group noted surface features (such as one participant said: "It is accurately referenced to other articles of statistics.") significantly more than semantic features ($\chi^2(1) = 60.24, p = 0.00$). 30% based their trust in the article on the advice of the support system. Table 6 presents the percentages and number of motivations of both groups.

The familiar group showed no significant difference in the notation of surface and semantic features ($\chi^2(1) = 0.89, p = 0.35$). Examples of motivations are: "I base my answer on the correlation between the content of the article and my prior knowledge" or "The way the article is written and structured and the number of references." 17% noted the advice of the

support system as an indicator for their trust in the article and 5% motivated their trust in the article with their first impression or feeling.

Table 8 Percentages and number of motivations of both groups for their trust in the article

	Percentages	<i>N</i>
Unfamiliar group		
Surface features	66%	66
References	51%	51
Textstructure	7%	7
Writing style	4%	4
Text length	4%	4
Semantic features	2%	2
Accuracy	1%	1
Neutrality	1%	1
Advice of support system	30%	30
Advice of WikiCheck	24%	24
Advice of DisputeFinder	6%	6
Others	2%	2
Familiar group		
Surface features	40%	40
References	34%	34
Writing style	4 %	4
Text structure	2 %	2
Semantic features	32%	32
Completeness	14%	14
Neutrality	12%	12
Accuracy	6%	6
First impression/feeling	5%	5
Advice of support system	17%	17
Advice of WikiCheck	11%	11
Advice of DisputeFinder	6%	6
Others	6%	6

Aspect c of hypothesis 2 that the unfamiliar group was more influenced by the advice of a support system that incorporated surface features as an indicator for credibility than by a semantic-based support system was confirmed. Aspect c of hypothesis 1 which stated that the familiar group was more influenced by a support system which incorporated semantic comparisons than by a surface- based support system was rejected.

4. Discussion

The aim of this study was to investigate if users who differed in their familiarity with the presented topics showed a difference in the preference, trust and influence of two different support systems. As described earlier, the two support systems used the users' method of evaluating credibility of information. The 3S model predicted that familiar and unfamiliar users incorporated different elements of information to evaluate credibility: semantic features, surface features and source experience (Lucassen & Schraagen, 2011a). We assumed that users preferred the support system which used the same method of evaluating credibility as them. Furthermore, they had more trust in such a support system and were more influenced by it. Through the hypotheses these assumptions were examined. We found that users who differ in their familiarity with a topic show a different preference, trust and influence of the presented support systems. Furthermore, they incorporated different elements of information to evaluate credibility.

4.1. Unfamiliar group

The results of this study support the hypotheses of the unfamiliar group that they have more trust in a surface-based support system. Furthermore, they are more influenced by it than by a support system which uses semantic comparisons. The hypothesis that the unfamiliar group prefers a support system which incorporates surface features as an indicator for credibility is rejected, however, the corresponding preference shows a trend toward significance. The advices of both support systems influence the unfamiliar group because they have lower trust in the article when one support system showed negative advice. This negative influence is for the advice of DisputeFinder only significant in comparison with positive advice from the DisputeFinder. In comparison with the advice of the WikiCheck support system, the advices of the WikiCheck support system have significantly more influence.

An explanation for their preference and trust in a surface-based support system is given through the motivations of the participants. The unfamiliar group evaluates the used criteria of a surface-based support system more positively than the criteria of a support system which incorporates semantic comparisons as an indicator for credibility. The understandability of the advice of both support systems is criticized because users do not know precisely where contradicting statements or irregularities are found. They miss more information about the advice. Furthermore, two participants interpret the name of the support system because they

described WikiCheck as a program which is affiliated with Wikipedia. The DisputeFinder is described as an independent program.

The motivations noted for trust in the article show that the unfamiliar group uses significantly more surface features than semantic features to evaluate an article. Furthermore, the influence of a support system can be explained through the notation of the advice of a support system as a reason for their trust judgement. 1/3 of the participants use the advice of a support system in their credibility evaluation. As predicted, the unfamiliar group uses surface features as references or writing style to judge the credibility of an article.

This study confirms that unfamiliar users incorporate surface features of information to evaluate credibility. They have more trust in a surface-based support system and are more influenced by it. Although the preference is not significant, it shows a trend towards the preference of the support system that incorporates surface features of information. An assumption of these finding is that users trust and are more influenced by the support system which fits with their own method of evaluating credibility. If users explicitly notice this agreement and the method the presented support systems use has to be studied in following studies.

4.2. Familiar group

The results of this study reject the hypotheses of the familiar group. The familiar group shows no significant preference for or trust in a support system that incorporates semantic comparisons as an indicator for credibility. Furthermore, the familiar group shows significantly no more influence of the support system that incorporates semantic comparisons over a surface-based support system. The results indicate that the familiar group have more trust in the article when the WikiCheck support system show positive advice but only in comparison with negative advice of the WikiCheck support system. No significant difference in their trust in the article is found for positive or negative advice of the DisputeFinder. In comparison with the advice of each other, neither the advice of the WikiCheck support system nor the advice of the DisputeFinder support system has significantly more influence.

One possible explanation concerns the judgement of criteria quality of the support systems. Motivations for preference and trust of a support system show that the familiar group evaluates the used criteria of a support system which incorporates semantic comparisons not more positively than the used criteria of the surface-based support system. As described by

the unfamiliar group, two participants of the familiar group also interpret the names of the support systems. WikiCheck is influenced by Wikipedia whereas the DisputeFinder works as an independent program. Furthermore, familiar participants note that they preferred no support system.

Another explanation for the rejection of the hypotheses is demonstrated through the noted motivations for trust in the article. These motivations show that the familiar group incorporates semantic features as well as surface-features almost equally. Only 1/5 of the familiar group notes the advice of the support systems in their credibility evaluation.

This study rejects that familiar users incorporate mainly semantic features of information in their credibility evaluation. The findings show that familiar users use semantic features as well as surface features, especially references to an equal extent. This can explain why our hypotheses that familiar users prefer, trust and are more influenced by a support system that incorporates semantic features over a surface-based support system cannot be confirmed. This support system does not fit with the users' method of evaluating credibility because it only incorporates semantic features. Both support systems do not fit. The familiar group shows no significant preference, trust and influence for any support system. For further research, it is important to study what users think about the method the presented support systems use for evaluating credibility and if they explicitly notice that one support system uses only semantic features and the other only surface features.

4.3. Limitations and future research

This study has a few limitations which have to be considered.

The first limitation considers the presentation of the two support systems. For every participant the DisputeFinder was presented first and the WikiCheck follows. This procedure can lead to a sequence effect which can bias the findings. It is important to change the sequences of the presented support systems to minimize the influence of the first support system presented on the following support system. We recommend repeating this experiment with a balanced sequence of the two support systems. Furthermore, in a follow-up study it is important to check explicitly the manipulation respectively the explanation of the two support systems and if the participants develop different perspectives of the presented support systems. It is necessary to study if users consciously notice that one support system incorporate only surface features and the other support system only semantic features. This is

not explicitly asked in this study. In addition, it has to be studied if users know consciously their method of evaluating credibility and its' match with the method of evaluating credibility of the support systems.

Second, the sample consists only of academic psychology students. The demographics of the participants could influence the results. They have a high education level and are familiar with psychological research methods through their study. Furthermore, students have good information skills and are aware of the importance of references. They incorporate references mainly in their credibility judgements (Lucassen & Schraagen, 2010). An overestimation of references has to be considered. On the one hand, using only students as participants can bias the findings. On the other hand, students are the group that mainly uses Wikipedia and are familiar with the way the website works (Head & Eisenberg, 2010). Furthermore, a greater sample size can enhance the reliability of the results. Trends towards significance can be verified. We propose to repeat this experiment with participants with other demographical features to consider the influence of academic education. The WWW is a daily used medium at school. Therefore, school children can be recruited for such an experiment because they are familiar with Wikipedia. In addition, their information skills (e.g. the importance of surface features as references) are not proficiently developed as in academic students. It would be interesting if the findings of this study can be strengthened if the overall information skills of the participants are lower.

A third limitation considers the names of the two support systems. The WikiCheck support system can be interpreted as a program which is affiliated with Wikipedia whereas the DisputeFinder works as an independent program. Participants noted that they reject the WikiCheck support system because of its dependence on Wikipedia. These misconceptions can lead to wrong evaluations.

A last limitation considers the research setting that can influence the results. The experiment was done in a laboratory which can result in very low motivation of participants to evaluate systematically. As described earlier, the evaluation of credibility depends on two factors: the ability and the motivation to evaluate (Metzger, 2007). Participants in this laboratory setting are not affected by the results of their evaluation which resulted in low motivation. Likely, the users' evaluation was heuristically. This can be different from real life situations. If students use information from Wikipedia for a coursework, their motivation to find credible information is higher than in a research setting because this can have negative consequences.

This difference in motivation can influence the evaluation of support systems and which of them are used in real-life circumstances. We propose to repeat this experiment in a real-life setting. The experiment can be involved in a course-work about familiar and unfamiliar topics.

More research is needed to strengthen the results obtained and get more insight in the interaction of familiarity of users, its influence on the users' method of evaluating credibility and different types of support systems. Further research should aim at the examination of different support systems that use the users' method of evaluating credibility. Furthermore, this study did not focus on the manipulation of the explanation of the two support systems. An interesting research question for following research can be if the participants consciously notice that one support system uses only semantic features and the other only surface features. Through this manipulation, the explanation that users prefer, trust and are more influenced by a support system which uses the same method of evaluating credibility can be strengthened.

4.4. Conclusion

The results demonstrate that unfamiliar users show a demand for appropriate support systems especially support systems that incorporate surface features as an indicator for the credibility of information. Based on these findings, we recommend developing such support systems using surface features which help users evaluating credibility of unfamiliar information.

Furthermore, the results demonstrate that the method which familiar users use to evaluate credibility does not fit with any of the presented support systems. They show no significant preference, trust or influence of one of the support systems. We recommend repeating a similar study design with a support system that incorporates surface as well as semantic features as an indicator of credibility. This support system fits with their method of evaluating credibility. For following research it is important to consider the manipulation of the explanation of the presented support systems explicitly.

Knowing why users trust and use support systems leads to a better understanding why present support systems are not completely appropriate and can help to develop support systems based on the users' methods of evaluating credibility. The approach to incorporate users' methods of evaluating credibility offers new opportunities for support systems based on users' characteristics and demands. Following research is necessary to enlighten the relation between users' method of credibility evaluation and appropriate support systems.

Addendum

In July 2011, a Wikipedia Recommender System (WRS) was introduced in Wikipedia which helps users to judge the credibility and quality of Wikipedia articles. The WRS was developed by Jensen, Pilkauskas & Lefèvre (2011). This support system works with users' evaluations and recommendations of the articles. The credibility of articles is calculated based on the feedback given through other Wikipedia users on different scales, for example the accuracy of the article, the completeness or the writing style. Another important factor which influences the calculated "trust value" is the expertise of the users. The users' level of expertise in different domains is established through their activity within a domain and their recommendations. All Wikipedia articles are classified in one of 15 categories. The activity of users within Wikipedia gives an indication of their expertise in specific categories of the classification system. Recommendations and evaluations of users who are identified as experts in one category are valued more than recommendations of users who are no experts in this category. In conclusion, the Wikipedia recommender System is an expert-rating system which gives regular users an indication of the credibility of the articles.

This support system differs from other implemented support systems such as for example *WikiTrust* or *WikiDashboard* because it is based on users' feedback of different criteria. The criteria included semantic as well as surface features.

The study in this paper introduces two support systems which incorporated either semantic or surface features. Furthermore, these two support systems are automated and not based on user ratings. The results of this study demonstrated that support systems which incorporate similar elements of information as Wikipedia users in their evaluation process of credibility have more influence and are trusted more. Participants in this study are either familiar or unfamiliar with the topic which leads to a different method of evaluating credibility. Familiar users incorporate semantic and surface elements of information and unfamiliar users incorporate surface elements. One support system fits with the elements of the unfamiliar group through incorporating surface elements (WikiCheck). The other support system only incorporates semantic features which did not fit with the elements the familiar group uses in credibility evaluation (DisputeFinder).

The WRS offers a new opportunity to strengthen the results of this study because this system matches with the method of evaluating credibility of familiar users. We recommend to repeat this study with the WRS as one support system which aims at the familiar group and the

WikiCheck support system which aims at the unfamiliar group. A limitation is that the WRS is a user-based support system whereas the WikiCheck is an automated support system. This can bias the findings. Another approach can be not only to exchange the DisputeFinder support system but also the WikiCheck support system. The development of a user-based support system which incorporates only surface features in its credibility evaluation is one opportunity.

In conclusion, it is important not only to implement support systems but to investigate the usefulness and handling of any support systems through empirical research with regular users. Without such research, support systems are implemented but not used by regular users.

References

- Adler, B. T., Chatterjee, K., de Alfaro, L., Faella, M., Pye, I., & Raman, V. (2008). Assigning trust to wikipedia content. *Technical Report UCSC-SOE-08-07*, School of Engineering, University of California, Santa Cruz, CA, USA.
- Alexander, P. A. (1992). Domain knowledge: evolving themes and emerging concerns. *Educational Psychologist*, 27(1), 33-51.
- Allen, J. T. (2009). Swine Flu' where americans are turning for flu facts, Pew Research Center, Washington, DC, retrieved 20th march 2011, from <http://pewresearch.org/pubs/1217/internet-swine-flu-source-top-websites-search>.
- Brand-Gruwel, S., Wopereis, I. & Vermetten, Y. (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior*, 21(3), 487-508.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 1980(39), 752-766.
- Chen, S. Y. & Macredie, R. (2010). Web-based interaction: a review of three important human factors. *International Journal of Information Management*, 30(5), 379-387.
- Chevalier, F., Huot, S., & Fekete, J. D. (2010). WikipediaViz: Conveying article quality for casual wikipedia readers. PacificVis '10: IEEE Pacific Visualization Symposium, pages 215-222.
- Ennals, R., Byler, D., Agosta, J. M., and Rosario, B. (2010). What is disputed on the web? In *Proceedings of the 4th workshop on Information credibility, WICOW '10*, pages 67-74, New York, NY, USA. ACM.
- Head, A.J. & Eisenberg, M.B. (2010). How today's college students use wikipedia for course-related research. *First Monday* 15(3).
- Jensen, C.D., Pilkauskas, P. & Lefèvre, T. (2011). Classification of recommender expertise in the wikipedia recommender system. *Journal of Information Processing*, 19, 345-363.

- Lazonder, A. W. & Rouet, J. F. (2008). Information problem solving instruction: some cognitive and metacognitive issues. *Computers in Human Behavior*, 24(3), 753-765.
- Lih, A.(2004). Wikipedia as participatory journalism: reliable sources? Metrics for evaluating collaborative media as a news resource. In *Paper for the 5th International Symposium on Online Journalism*, Austin, USA.
- Lucassen, T. & Schraagen, J. M. (2011a). Factual accuracy and trust in information: the role of expertise. *J. Am. Soc. Inf. Sci.*, 62(7), 1232-1242.
- Lucassen, T. & Schraagen, J. M. (2011b). Evaluating WikiTrust: a trust support tool for wikipedia. *First Monday*, 16(5).
- Lucassen, T. & Schraagen, J. M. (2010). Trust in wikipedia: how users trust information from an unknown source. In *Proceedings of the 4th workshop on Information credibility*, WICOW '10, pages 19-26, New York, NY, USA. ACM.
- Lucassen,T. & Schraagen, J. M. (in preparation). Evaluation of credibility support tools for Wikipedia.
- Metzger, M. J. (2007). Making sense of credibility on the web: models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13), 2078-2091.
- Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education*, 41(3), 271–290.
- Potthast, M., Stein, B., and Gerling, R. (2008). Automatic vandalism detection in Wikipedia. In Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., and White, R. *Advances in Information Retrieval*, volume 4956 of *Lecture Notes in Computer Science*, chapter 75, pages 663-668. Springer Berlin / Heidelberg, Berlin, Heidelberg.
- Simon, J. (2010). The entanglement of trust and knowledge on the web. *Ethics and Information Technology*, 12(4), 343-355.
- Suh, B., Chi, E. H., Kittur, A., & Pendleton, B. A. (2008). Lifting the veil: improving accountability and social transparency in wikipedia with WikiDashboard. In *CHI*, Florence, Italy.

Todorov,A., Chaiken,S. & Henderson,M. D. (2002). The heuristic-systematic model of information-processing. In J. P. Dillard& M. Pfau (2002), *The Persuasion Handbook: Developments in Theory and Practice* (pp.195-213). Thousand Oaks,CA: Sage Publications,Inc.

Internetworldstat. (2010). Usage and population statistics.Retrieved 19th march 2011, from <http://www.internetworldstats.com/stats.htm>.

Voss, J. (2005). Measuring Wikipedia. In *International Conference of the International Society for Scientometrics and Informetrics* : 10th,Stockholm (Sweden),24-28 July 2005.

Zeng, H., Alhossaini, M. A., Ding, L., Fikes, R., & McGuinness (2006). Computing trust from revision history. In *Proceedings of the 2006 International Conference on Privacy, Security and Trust*.

Appendix A Explanation of support systems

Dispute Finder

Bij de eerste twee artikelen kun je gebruik maken van het “DisputeFinder” system. Dit systeem zoekt op andere websites of informatie uit dit wikipedia-artikel wordt tegengesproken. Dit wordt gedaan door te zoeken naar patronen zoals „ falsely claimed that...” en “the misconception that...”.

Op basis van het aantal gevonden “disputed claims” wordt er een positief of negatief advies gegeven over de betrouwbaarheid van het artikel. Je vindt dit advies bovenaan, bij de titel.

WikiCheck

Bij de laatste twee artikelen kun je gebruik maken van het “WikiCheck” system. Dit system maakt gebruik van een adaptief neural network die op basis van veel eigenschappen van het artikel (onder andere diverse details van de auteurs, de bewerkingsgeschiedenis, en de referentielijst) een betrouwbaarheidsoordeel geeft.

Op basis van het algoritme wordt er een positief of negatief advies gegeven over de betrouwbaarheid van het artikel. Je vindt dit advies bovenaan, bij de titel.