chist-era

Credibility Models for Multimedia Streams

Alexandru L. Ginsca*, Mihai Lupu**, Adrian Popescu*

*CEA, LIST, LVIC France

** Vienna University of Technology, ISIS, IMP

Contacts: lupu@ifs.tuwien.ac.at, adrian.popescu@cea.fr

MUCKE Project, Deliverable 3.1

28/01/2014

Contents

Abstract

This report presents a detailed analysis of existing credibility models from different areas of the Web. An extensive survey of existing work shows that there is a very rich body of work pertaining to different aspects and interpretations of credibility, particularly for different types of textual content (web sites, blogs, tweets etc.). The study focuses on automatic prediction and on integration of credibility aspects in a retrieval process. In the multimedia domain we have found significantly less material. The few attempts that we have identified focus on video and audio content. In this sense, the focus of our project, MUCKE, on image content complements and completes the studies on credibility observed in the survey.

The second part of this report sketches the credibility model which will be implemented in MUCKE. The model will identify and combine context and content features using appropriate machine learning techniques. The ultimate goal is to provide an efficient estimation of the credibility of the content shared by multimedia social network users. To achive this goal, the current document provides the necessary background information and the proposed model distilled from it.

1 INTRODUCTION

The MUCKE 3.1 Deliverable analyzes past and current work on different aspects of credibility and proposes a model for handling multimedia information streams shared on social media. In doing so it covers the transition from credibility in the Web 1.0 static environment to the dynamic Web 2.0 environment. This report is a summary focusing on MUCKE-specific issues. The full report is pending publication.

1.1 MOTIVATION

According to a 2011 Pew Research Center [1], about 50% of computer literate individuals, with at least a college degree take most of the national and international news from the Web. That is: more than television, newspapers, radio or magazines. It is therefore easy for the reader to relate the need for credibility on the web. Even more, in addition to the social group to which these readers belong, it is also a fact that a majority of the population in the developed world has access to and continuously uses the Web to seek information.

Initial works on Web credibility include research on understanding users' mental models when assessing credibility and on the development and evaluation of interventions to help people better judge credibility online. The field of captology [2] studies how technology can be designed to persuade end-users. Much prior work in the area of credibility approaches the topic from a captology perspective, with a goal of understanding how people evaluate credibility so as to help designers create websites that will appear more credible. Examples include Schneiderman's guidelines for designing trust online [3] and Ivory and Hearst's tool for high quality site design [4]. This line of prior research has shown that users consider many different pieces of information to help them evaluate the credibility of Web pages. Tseng and Fogg [5] categorize this information into four types of credibility:

- *Presumed credibility* is based on general assumptions in the users' mind (e.g.the trustworthiness of domain identifiers).
- *Surface credibility* is derived from inspection of a site, is often based on a first impression that a user has of a site, and is often influenced by how professional the site's design appears.
- *Earned credibility* refers to trust established over time, and is often influenced by a site's ease of use and its ability to consistently provide trustworthy information.
- *Reputed credibility* refers to third party opinions of the site, such as any certificates or awards the site has won.

These four types of credibility refer to a human-based assessment of it. While works that tackle automatic credibility prediction for textual content already exist (and we shall cover them in the following sections), this topic is virtually uncovered for multimedia information streams. In MUCKE,

we focus on this last type of information and we have identified three main concepts to which credibility can be applied: **data**, **users** and **systems**. The first two are explicitly mentioned in the MUCKE project proposal while the latter was identified as important during the bibliographical research on credibility. The credibility of data, users and systems is of course tightly interlinked. Our purpose is to start credibility estimation from single data pieces, aggregate these individual pieces into estimations of user credibility and finally exploit user credibility in an image retrieval system in order to demonstrate its practical applicability.

Concerning data and user credibility, this report covers the latest trends in Web credibility research and detaches itself from other surveys, such as the work of Lazar et al. [6], who examine the research literature in the area of web credibility until the year 2007. They examine the general credibility of web sites, online communication, such as e-mail and instant messaging and discuss the implications for multiple populations (users, web developers, browser designers, and librarians), whereas we focus on the latest works on credibility in social media and, from a technical perspective, we are mainly interested in automatic methods used for credibility predictions. Other surveys related to the notion of credibility include the works on trust, Golbeck [7] who focuses on trust on Web, including webpages, websites, Semantic Web data but also services including peer-to-peer environments and Web services and Cofta [8], who presents a literature review on trust in information and communication technology (ICT) structures and expertise, represented by the work of Balog et al. [9], who cover the subject of expertise retrieval in the field of information retrieval.

1.2 CREDIBILITY OF AN INFORMATION ACCESS SYSTEM

A recent EuroStat report[1] shows that within the European Union (28 countries), in 2011, 71% of all individuals had used a search engine to find information. Certainly, these percentages are likely to drop in regions under development, but Internet penetration is on the rise even in the most remote places [10]. In fact, it is likely that Internet adoption will outpace e-literacy[11], and at least some users will have the feeling of trying to quench their thirst for information from a fire hydrant. In the case of the Web, this fire hydrant ejects an amorphous mix of useful, useless and malignant information. Web search engines play an undisputed vital role in this process. In addition to topical relevance, they also use simple and efficient metrics to estimate the importance of a web page (e.g. PageRank, HITS algorithms). There are several issues with the status quo:

1. PageRank-like algorithms are substituting a hard problem (credibility) by an easier problem (popularity) ;

2. there is the assumption that the search engine is an impartial information indexer with the users' best interests at heart. Even if that were the case for all search engines, the web routes search results through a variety of intermediary nodes, most of the time without encryption.

[1] `http://bit.ly/167xo82` Visited: July 2013, Most recent data: 2011

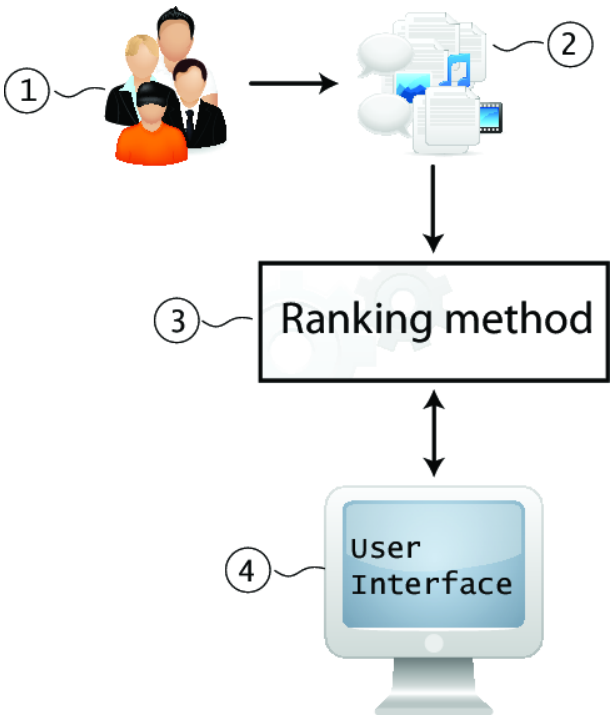

FIGURE 1: A TYPICAL INFORMATION RETRIEVAL SYSTEM.

The negation of the assumption, as the existence of third party intermediaries put into question the credibility we can assign to a search result list;

3. for the purposes of assessing credibility, the solutions to both of the above issues feed into a recursive credibility question unless the user can develop an understanding of the results provided.

Concerning system credibility, this survey will address primarily the first two problems, and only partially the third (particularly because it includes a vast research area in Human-Computer Interaction). In doing so, we strive to avoid the philosophical, psychological and sociological issues of trust and trustworthiness—major components of credibility—and focus on technical aspects.

1.3 DEFINITIONS

Before proceeding, we should provide a definition of the two elements under discussion here: credibility and IR.

Credibility. A majority of researchers identify two components of credibility, namely trustworthiness and expertise [12], but we also identified quality and reliability as important aspects. In general, trustworthiness is unbiased, truthful, well intentioned, while expertise is knowledgeable, experienced, or competent. In addition, we will discuss quality, which is often seen as an intrinsic characteristic of content shared on the Web, and reliability, which refers to the extent to which something can be regarded as dependent and consistent.

Each of the four components in Figure 1 and the above list has their own role to play in the general assessment of credibility. For each of them, the two components of credibility, expertise and

trustworthiness, have different meanings. In the following three sections we detail this for each of them and summarize the existing work.

IR System. Figure 1 shows a highly schematized version of a retrieval system. The IR Engine itself may be considered to be only the *Ranking method*, which in this case includes the indexing, similarity scoring and any other components the retrieval system might have (e.g. relevance feedback). But there are other important components, particularly for the consideration of credibility:

1. Significant amount of information online is directly attributed to a person, be it the editor or author of an article, or the owner of a blog or twitter feed.

2. The data itself, generated by the above-mentioned user and to be indexed and made retrievable by the system

3. The retrieval system itself

4. The interface used to present results to the end-users and to accept their feedback.

In Section **??**, we detail each of the four concepts linked to credibility and provide arguments for this relation and in Section **??**, we group the works on credibility based on their final purpose. In Section **??**, we cover literature that is placed in the latest direction of web credibility, the development of methods for web page credibility prediction and website feature exploration in this context. We also pay special attention to the medical domain and present a brief review of blog credibility. In Section **??** we present the latest works on credibility in social networks, with a focus on Twitter and Community Question Answering platforms, while in Section **??**, we cover an emerging line of research, credibility in the multimedia domain. In Section **??**, we discuss credibility in the context of information retrieval systems. In Section **??**, we present a list of existing resources that can be used for the assessment of credibility and, given that public ground truth datasets for assessing multimedia data credibility do not exist, we briefly describe three proposals for creating such datasets. Before concluding, in Section 2 we present a user credibility model which combines the mining of user context (i.e. social network features, social activity etc.) and user content (i.e. analysis of textual and image content shared online) in order to predict her/his credibility.

2 MUCKE CREDIBILITY MODEL FOR MULTIMEDIA STREAMS

The credibility model used in MUCKE distinguishes four components: *expertise*, *trustworthiness*, *quality*, and *reliability*, with the following definitions:

expertise denotes the knowledge ability of the object to provide truthful information;

trustworthiness denotes the estimated intent of the object to provide truthful information;

TABLE 1: PERCEIVED IMPORTANCE OR APPLICABILITY OF EACH CREDIBILITY COMPONENT FOR EACH IR SYSTEM COMPONENT

	expertise	trustworthiness	quality	reliability
User (creator)	●●●	●●●	●○○	●○○
Data object	●○○	●○○	●●●	○○○
IR engine	●●○	●●○	●○○	●●●
HCI	○○○	●●○	●●●	●●●

quality denotes the ability of the object to convey the truthfulness of the information provided;

reliability denotes the consistency of the object's ability to provide information.

In typical information retrieval or extraction systems, each of the four components may be applied to each of the four levels where the issue of credibility may appear. However, the importance of each component at each level is different and differences are illustrated in Table 1.

For instance, while it is possible to talk about reliability of the data creator, in the sense we have defined above, it makes more sense to talk about reliability of the engine and of the HCI component, if we are to have a general estimation of credibility. Similarly, while we may consider the expertise of a document (i.e. the coverage and detail in which it describes the topic), it makes little sense to talk about the expertise of the human-computer interface.

In doing a credibility assessment, while it would be desirable to consider the system as a whole, in practice we may chose to focus on a particular component and a particular aspect of credibility. For instance, in the case of multimedia streams, consisting of images and tags associated with them, we may focus on the user creating the tags. As we mentioned, in MUCKE, we focus on image collections annotated by identified users and work over a collection of Flickr photos. The credibility of each user can be derived by aggregating contextual and content features which characterize her/his contributions to the multimedia sharing platforms. Simply put, given a set of user features, the purpose is to attribute a numeric credibility score to the contributions of each user which accounts for the quality of her/his contributions in the multimedia social network. The extensive review of credibility prediction literature convinced us that credibility estimation for multimedia streams should be cast as a machine learning problem. As we mentioned, initial tests on the DIV400 collection indicate that a combination of contextual and content features is necessary. It became clear than no single feature provides a good estimate of credibility when taken in isolation. However, results also showed that they are often complementary and that an aggregation of weak signals given by separate features boosts the quality of results. These tests also allowed us to calibrate the size of the test collections that need to be manually annotated for accurate credibility estimation. A description of the three collections that we intend to build in order to compute and assess credibility estimations is provided in Subsection **??**.

One important question that will be investigated is whether credibility should be considered as a global user feature or as a domain-related characteristic. For instance, if we consider the case of an image retrieval application, is it more efficient to attribute a unique credibility or to adapt this credibility score to each query issued? The hypothesis that we want to test is if the quality of user annotation varies from one topic to another. For instance, a user which is a bird expert or enthusiast might tag her/his photos of birds more accurately than another user who takes photos of birds only occasionally. Global scores are simpler to compute because credibility scores can be computed offline and reused for any query issued in the system. Topic adapted scores need to be be computed online since an adaptation of the features to the context of the query. A similar feature set will be used in both cases but at least a part of these features need to be reweighted for domain adapted credibility. Going back to the example of a query with a bird name, it might be interesting to give more weight to textual or visual features which relate to this conceptual domain.

In the following subsections we briefly introduce contextual and content features that will be tested and then list a series of machine learning methods that we consider fitted for our task. The list of features is non-exhaustive and a more complete version along with an evaluation of their usefulness for credibility estimation will be provided in Deliverables 3.2 and 3.3 of the project.

2.1 CONTEXTUAL FEATURES FOR CREDIBILITY

Following existing works about credibility prediction in social networks described in Section **??**, we will test a number of contextual features which might prove useful for credibility estimations. These include:

- *social features* - first, the position of the user in the Flickr network will be computed using PageRank-like algorithms. Second, each user's participation in Flickr groups will be assessed.
- *popularity features* - these features give feeback about the echo that a user's contributions have in the community and include the average number of views and the average number of comments per photo.
- *vocabulary features* - these features give a statistical characterization of a user's tags and include: average number of tags per photo, size of the vocabulary used, percentage of bulk uploads, the percentage of unique tags.
- *temporal features* - features which account for the temporal activity of a Flickr user: user account creation date, number of days when photos were uploaded, regularity of uploads.

2.2 CONTENT FEATURES FOR CREDIBILITY

Multimedia mining methods will be devised in order to assess the correspondence between tags associated to a photo and its content. Given the uncontrolled character of Flickr tags, wide coverage

resources (in terms of concepts and languages are needed to assess credibility. We have identified and preprocessed the following datasets:

- *Wikipedia* - the popular online encyclopedia is used to extract data-driven similarity scores between concepts and to create translation dictionaries.
- *WordNet*[2] - the hierarchical lexical database is classically used to compute hierarchical similarity scores between concepts.
- *ImageNet*[3] - this image collection illustrates over 20,000 WordNet concepts with 14 million Web images. The relevance of ImageNet photographs for the concept that they illustrate was manually validated and this dataset is a very valuable resource for creating visual concept models.
- *MUCKE collection* - the image dataset collected in MUCKE (around 90 million images for 60,000 Wikipedia concepts) has a wider coverage compared to ImageNet but no manual validation of the included photos. It will be used to investigate to what extent the use of a noisy image collection is efficient in visual processing.

We describe textual and visual processing methods that will be tested to assess credibility in the following subsections.

2.2.1 TEXTUAL PROCESSING

Textual processing has two purposes: align content from different resources and create actual features for credibility. The resources cited above are only partly aligned and it is necessary to create a mapping between Wikipedia, WordNet on the one side and Flickr on the other. For instance, visual resources are available only for English while Flickr tags are often times multilingual. In order to assess the credibility of non-English taggers, tag translation is necessary. Although imperfect, it is preferable to not being able to process tags at all. Wikipedia was already used to create translation dictionaries for a large number of languages and will also be exploited in MUCKE. Tag ambiguity, i.e. knowing to which sense of a word a given Flickr tag refers to, is also important in credibility estimation. For instance, if a user tags with Jaguar, the photo can refer to a feline or to a car and, depending upon the actual sense, different textual/visual processing should be applied. A textual disambiguation method based on Wikipedia and applicable to a large number of languages is currently studied as part of Task 2 of MUCKE.

The tags used by a Flickr user can be exploited in order to build different features for credibility estimation. Below we present a tentative list of textual features which will be used:

- *tag specificity* - a hypothesis is made that there is a correlation between the level of specificity of tags and their quality. For instance, a user who tags with Sun Conure is possible more

[2]http://wordnet.princeton.edu/

[3]http://www.image-net.org/

knowledgeable about birds than another user who tags with parrot or bird. If this behavior is consistent across the user's Flickr annotations, it might indicate that the user who uses specific tags elicits a higher level of expertise. Specificity can be derived either from hierarchical resources, such as WordNet, by looking at the depth of tags in the hierarchy or from Flickr itself by computing the commonness of tags and assuming that rarer tags are more specific.

- *tag consistency* - shared representations of tags can be created by leveraging contributions of a large number of users and be used to assess to what point an individual user's contributions match with these social representations.

- *feedback mining* - opinion mining techniques can be exploited in order to evaluate the comments left by other Flickr users about a given user's photo. An adaptation of techniques is needed since the Flickr comments vocabulary contains a badges and words which are specific to it.

2.2.2 VISUAL PROCESSING

Here our main purpose is to assess to what point there is a link between a user's tags and the actual content of the images. Simply put, given a photo tagged with Sun Conure and sky, we want to find out if each of these concepts is actually depicted in the image or not. Inspired by work done in multimedia and computer vision communities, we started working on creating a large number of visual concept models. ImageNet was already processed and linear SVM models were created for over 17,000 concepts. The results of the ImageNet Large Scale Visual Recognition Challenge [4] have shown that the use of convolutional neural networks (CNNs) results in a significant performance improvement compared to the use of SVMs. We plan to use CNNs in the future but, given the relatively low maturity of existing tools, their efficient usage is not straightforward. Although large, the array of concepts which were already modeled ensures a partial coverage of tags used in Flickr and different improvement directions will be explored:

- *tag translation* - as we mentioned, ImageNet exists only in English and this has two negative consequences. First, tags in other languages are not taken into account since they are not modeled in ImageNet. Second, there are cases of inter-lingual ambiguity. For instance, mare in Italian has sea as main translation in English whereas mare in English stands for a female horse. The translation dictionaries created with Wikipedia will be used to evaluate the usefulness of automatic tag translation.

- *tag disambiguation* - there are two cases of ambiguity that we will process. The easiest one is when a tag has a corresponding ImageNet representation among the different senses represented in the resource. For instance, a photo of a dog (as animal) can be accurately process since this sense of the word is covered in ImageNet. A simple manner to disambiguate is

[4] http://www.image-net.org/challenges/LSVRC/2013/

such cases is to consider the sense which is visually closest to the test photo. Ambiguity is much more complicated to handle when the targeted sense of the word is not illustrated in ImageNet. Inversely, a photo of Jaguar (as car) will be compared to the ImageNet representation of jaguar (as animal) since there is no synset for Jaguar (as car) in WordNet. A special case is that of given names, such as John or Bill, which also have other meanings in English and are frequently used to tag photos of people in Flickr. Face detection techniques could be used here to process given names properly.

- *collection extension* - the MUCKE collection is not manually validated and thus noisy but it will be used in order to supplement ImageNet with additional concepts in order to have an improved coverage of the Flickr vocabulary. This should fill in important gaps, notably concerning named entities which are frequent in Flickr but not illustrated in ImageNet. Such entities refer mostly artefact names (i.e. Renault, BMW, Airbus).

2.3 MACHINE LEARNING METHODS FOR CREDIBILITY

Based on the nature of the MUCKE datasets described in Section **??**, we identify two machine learning problems. The first one is a regression scenario in which we set the goal to predict a user's credibility score viewed as a continuous variable. For this task, we use the direct credibility assessment collection. The second one falls under the learning to rank paradigm, which is well suited to be used for directly predicting rankings using user credibility features. We evaluate these models on the test collections built for the generic IR scenario and the domain-adapted IR scenario. For the latter, some learning to rank algorithms also have the advantage of jointly modelling the query and the list of results. This allows us to include domain credibility features in a more direct manner.

2.3.1 PREDICTING USER CREDIBILITY SCORES

We test on the direct credibility collection presented in Section **??** several supervised learning algorithms, such as linear regression, support vector regression [13] and gradient boosting machines [14]. Due to the nature of the collection, containing few instances and a small set of features, the choice of a particular learning algorithm does not have a strong impact on the prediction accuracy but plays a greater role towards feature analysis. We view the learning process as a means of validating our hypothesis for user credibility indicators and to evaluate the feature engineering efforts.

2.3.2 PREDICTING CREDIBILITY INDUCED USER RANKINGS

Learning to rank algorithms are typically grouped into three categories:

- *pointwise approaches* - in the case of these approaches, the ranking problem is reduced to a classification or, such is our case, a regression problem. The ranking in the training data is not kept as a whole and each item in the ranking becomes an instance for a supervised learning

algorithm. The label given to each instance is chosen so that it can be used to reproduce the initial ranking. Examples of such algorithms are OC SVM and McRank.

- *pairwise approaches* - in these types of approaches, the problem is also transformed into a classification or regression problem but this time an instance is a pair of items from the initial ranking used for training. The label is chosen so that it expresses a preference for an item over the other one in a pair. For instance, the label may be a binary variable indicating whether the first item in the pair should be ranked higher than the second. Popular algorithms from this category include RankSVM, LambdaMART and RankBoost.

- *listwise approaches* - these approaches exploit the rankings as a and use ranking lists in both learning and prediction. For training, the labelled items associated with a query are viewed as a single instance. One specific advantage of these methods is the possibility to chose an evaluation measure for rankings (e.g. NDCG, MAP) to be directly optimized during training. For this category, some noticeable algorithms are: AdaRank, PermuRank and ListNet.

For an initial round of experiments, we focus on pairwise and listwise approaches and more exactly on those using boosting techniques (e.g. LambdaMART, AdaRank, RankBoost). Besides these, we also test RankSVM. Our choice of algorithms is based on the outcome of the Yahoo! Learning To Rank Challenge, in which the winners used and ensemble of LambdaMART rankers [15] and most of the top ranked teams used boosting algorithms [16].

2.3.3 FEATURE SELECTION AND ANALYSIS

The machine learning experiments carry the purpose not only of predicting a user credibility score or ranking but also to highlight those features or groups of features that have a stronger influence in the prediction. This allows us to draw some conclusions on which types of features are good indicators for credibility (contextual features vs.content features, textual features vs. visual features etc.) Another benefit from such an analysis is that it will allow us to channel future feature engineering efforts. Besides looking at the direct correlation with the credibility score, estimates for the the relative importance of features are obtained using different learning methods. One straightforward way to do this is to compare the quality of predictions, either the credibility score or a ranking, when using the same algorithm with different subsets of features. Another method for obtaining feature importance estimates is to exploit the nature of those algorithms that produce feature weightings as a by-product of the predictions process. For example, the weights can be the coefficients of a linear model, the relative rank of a feature used as a decision node in a decision tree model or the average of these ranks in models that use ensembles of trees.

3 CONCLUSIONS

This literature suvey demonstrates the difficulty in defining credibility, particularly in such a way as to make it amenable to automatic assessment and estimation. We have shown that credibility can be understood in many ways, depending on the context in which is used. It can be applied to individual pieces of information, to users or to entire systems and it can be defined subjectively, as a quality perceived by the system users or, in particular scenarios, more objectively. In MUCKE, we will follow this last path and, following the experience of benchmarking exercises in IR, attempt a functional definition of credibility in the context of multimedia information retrieval and extraction systems.

While building on recent tentatives to automatically predict credibility of textual information, the current work departs from existing approaches via the utilization of multimedia mining techniques and the exploration of context features which are fitted in this context. A first step in this process is the creation of new resources to provide the experimental testing grounds not only for our credibility estimation efforts, but also those of the research community at large.

References

[1] Pew Research Center. Internet Gains on Television as Public's Main News Source. Technical report, The Pew Research Center for the People and the Press, 2011.

[2] Bernardine MC Atkinson. Captology: A critical review. In *Persuasive Technology*, pages 171–182. Springer, 2006.

[3] Ben Shneiderman. Designing trust into online experiences. *Communications of the ACM*, 43(12):57–59, 2000.

[4] Melody Y Ivory and Marti A Hearst. Statistical profiles of highly-rated web sites. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 367–374. ACM, 2002.

[5] Shawn Tseng and BJ Fogg. Credibility and computing technology. *Communications of the ACM*, 42(5):39–44, 1999.

[6] Jonathan Lazar, Gabriele Meiselwitz, and Jinjuan Feng. *Understanding web credibility: a synthesis of the research literature*, volume 1. Now Publishers Inc, 2007.

[7] Jennifer Golbeck. Trust on the world wide web: a survey. *Found. Trends Web Sci.*, 1(2):131–197, January 2006.

[8] P. Cofta. The trustworthy and trusted web. *Foundations and Trends in Web Science*, 2(4), 2011.

[9] Krisztian Balog. Expertise retrieval. *Foundations and Trends® in Information Retrieval*, 6(2-3):127–256, 2012.

[10] David Talbot. African Entrepreneurs Deflate Google's Internet Balloon Idea. *MIT Technology Review*, 2013.

[11] E. Wyatt. Most of u.s. is wired, but millions aren't plugged in. *The New York Times*, August 18 2013.

[12] B. J. Fogg and Hsiang Tseng. The elements of computer credibility. In *Proc. of SIGCHI*, 1999.

[13] Debasish Basak, Srimanta Pal, and Dipak Chandra Patranabis. Support vector regression. *Neural Information Processing-Letters and Reviews*, 11(10):203–224, 2007.

[14] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.

[15] Christopher JC Burges, Krysta Marie Svore, Paul N Bennett, Andrzej Pastusiak, and Qiang Wu. Learning to rank using an ensemble of lambda-gradient models. *Journal of Machine Learning Research-Proceedings Track*, 14:25–35, 2011.

[16] Olivier Chapelle and Yi Chang. Yahoo! learning to rank challenge overview. *Journal of Machine Learning Research-Proceedings Track*, 14:1–24, 2011.