

Evidentiality for Text Trustworthiness Detection

Qi Su^{1,2}, Chu-Ren Huang and Helen Kai-yun Chen

¹Department of Chinese & Bilingual Studies, The Hong Kong Polytechnic University

²Key Laboratory of Computational Linguistics, Peking University

sukia@pku.edu.cn, {helenkychen, churen.huang}@gmail.com

Abstract

Evidentiality is the linguistic representation of the nature of evidence for a statement. In other words, it is the linguistically encoded evidence for the trustworthiness of a statement. In this paper, we aim to explore how linguistically encoded information of evidentiality can contribute to the prediction of trustworthiness in natural language processing (NLP). We propose to incorporate evidentiality into a framework of machine learning based text classification. We first construct a taxonomy of evidentials. Then experiments involving collaborative question answering (CQA) are designed and implemented using this taxonomy. The experimental results confirm that evidentiality is an important clue for text trustworthiness detection. With the binarized vector setting, evidential based text representation model has considerably performed better than both the bag-of-word model and the content word based model. Most crucially, we show that the best trustworthiness detection result is achieved when evidentiality is incorporated in a linguistically sophisticated model where their meanings are interpreted in both semantic and pragmatic terms.

1 Introduction

With the exponential increase in web sites and documents, the amount of information is no longer a main concern for automatic knowledge acquisition. This trend raises, however, at least two new issues. The first is how to locate the information which exactly meets our needs among the vast web content. Efforts to address this issue can be exemplified by advanced re-

search in information retrieval, information extraction, etc. The second is how to judge the validity of the acquired information, that is, the trustworthiness of information. This issue has attracted considerable interest in some related research areas recently. Taking the specific information retrieval task, question answering (QA) as an example, a QA system attempts to retrieve the most appropriate answers to questions from web resources. To determine the trustworthiness of the extracted candidate answers, a common approach is to exploit the co-occurrence frequency of questions and candidate answers. That is, if a candidate answer co-occurs more frequently with the question than other candidates, the QA system may judge it as the best answer (Magnini, 2002). This approach presupposes and relies crucially on information redundancy. Although this heuristic method is simple and straightforward, it is not applicable to all cases. For the applications which don't involve much information redundancy, the heuristic could cease to be effective. The task of collaborative question answering (CQA) which we will address in this paper is just one of such examples. For a user posted question, there are usually only few answers provided. So, the heuristic is not useful in providing the best answer. In addition, since the spread of unsubstantiated rumors on the Internet is so pervasive, the high-frequency information on the Web sometimes may mislead the judgment of trustworthiness. In terms of the above consideration, it is essential to look for other approaches which allow directly modeling of the trustworthiness of a text.

Given that non-textual features (such as user's Web behavior) used in text trustworthiness detection are often manipulated by information providers, as well as no directly related textual features for the task has been proposed up to

date, we need a more felicitous model for detecting the trustworthiness of statements. Noting that *evidentiality* is often linguistically encoded and hence provides inherent information on trustworthiness for a statement, we propose to incorporate the linguistic model of evidentiality in our study. Specifically, we incorporate evidentiality into a machine learning based text classification framework, and attempt to verify the validity of evidentiality in trustworthiness prediction of text information in the context of collaborative question answering. The experimental results show that evidentials are important clues in predicting the trustworthiness of text. Since none of the task-specific heuristics has been incorporated, the current approach could also be easily adapted to fit other natural language processing applications.

The paper proceeds as follows. In section 2 we discuss related work on text trustworthiness detection. The section is divided into two parts: the current methodology and the textual features for analysis in the task. Section 3 introduces the linguistic researches on evidentiality and our taxonomy of evidentials based on the trustworthiness indication. Section 4 presents the experiment settings and results. Finally, in section 5 we discuss the experiment results and conclude the current research.

2 Related Work

The research of text trustworthiness is very helpful for many other natural language processing applications. For example, in their research on question answering, Banerjee and Han (2009) modulate answer grade by using a weighted combination of the original score and answer credibility evaluation. Also, Weerkamp and Rijke (2008) incorporate textual credibility indicators in the retrieval process to improve topical blog posts retrieval. Gyongyi et al (2004) propose a TrustRank algorithm for semi-automatically separating reputable, good Web pages from spams.

2.1 General Approaches for Text Trustworthiness Detection

In past research, the judgment for the trustworthiness or credibility of a given text content is usually tackled from two aspects: entity oriented and content oriented (Rubin and Liddy, 2005). The former approach takes into consideration the information providers' individual profiles, such as their identity, reputation, authority and

past web behavior; whereas the latter approach considers the actual content of texts. Metzger (2007) reviews several cognitive models of credibility assessment and points out that credibility is a multifaceted concept with two primary dimensions: expertise and trustworthiness. Following Metzger's framework, Rubin and Liddy (2005) compile a list of factors that users may take into account in assessing credibility of blog sites. This list could also be summarized as the above mentioned two-folds: the bloggers' profiles and the information posted in the entries.

Comparing these two aspects, most existing research on text trustworthiness focuses on the user oriented features. Lots of user oriented features have been proposed in the research of credibility detection. To score the user oriented features such as user's authority, a common approach is based on a graph-based ranking algorithm such as HITS and PageRank (Zhang et al, 2007; Bouguessa et al, 2008).

In the research of text trustworthiness detection, the overwhelmingly adaption of non-textual features such as entity profiles over text content based features reflect some researchers' belief that superficial textual features cannot meet the need of text credibility identification (Jeon et al, 2006). In this paper, we examine the lexical semantic feature of evidential and argue that evidentiality, as a linguistically instantiated representation of quality of information content, offers a robust processing model for text trustworthiness detection.

The detection of information trustworthiness also has promising application values. Google News¹ is just such an application that ranks search results according to the credibility of the news. Other online news aggregation service, such as NewTrust², also focuses on providing users with credible and high quality news and stories. The existed applications, however, rely on either the quality of web sites or user voting. So, it is anticipated that the improvement on the technology of text trustworthiness detection by incorporating lexical semantic cues such as evidentiality may shed light on these applications.

2.2 Textual Feature Based Text Trustworthiness Detection

Although non-textual features have been popular in text credibility detection, there has been a few research focusing on textual features so far. Gil

¹ <http://news.google.com/>

² <http://www.newstrust.net/>

and Artz (2006) argue that the degree of trust in an entity is only one ingredient in deciding whether or not to trust the information it provides. They further point out that entity-centered issues are made with respect to publicly available data and services, and thus will not be possible in many cases. In their research of topical blog posts retrieval, Weerkamp and Rijke (2008) also consider only textual credibility indicators since they mentioned that additional resources (such as bloggers' profiles) is hard to obtain for technical or legal reasons.

However, most research which utilizes textual features in text trustworthiness detection usually equates writing quality of document with its trustworthiness. Therefore, some secondary features which may not directly related to trustworthiness are proposed, including spelling errors, the lack of leading capitals, the large number of exclamation markers, personal pronouns and text length (Weerkamp and Rijke, 2008). There has not been attempted to directly evaluate inherent linguistic cues for trustworthiness of a statement.

3 On Evidentiality in Text

Evidentiality, as an explicit linguistic system to encode quality of information, offers obvious and straightforward evidence for text trustworthiness detection. Yet it has not attracted the attention which it deserves in most of the natural language processing studies. In this paper, we aim to explore how we can incorporate the linguistic model of evidentiality into a robust and efficient machine learning based text classification framework.

Aikhenvald (2003) observes that every language has some way of making reference to the source of information. Once the language is being used, it always imprinted with the subjective relationship from the speakers towards the information. Evidentiality is information providers' specifications for the information sources and their attitudes toward the information. As a common linguistic phenomenon to all the languages, it has attracted linguists' attention since the beginning of 20th century. In any language, evidentiality is a semantic category which could be expressed on both grammatical level (as in some American Indian language) and lexical level (as in English, Chinese and many other languages). The linguistic expressions of evidentiality are named as evidentials or evidential markers.

Mushin (2000) defines evidential as a marker which qualifies the reliability of information. It is an explicit expression of the speaker's attitudes toward the trustworthiness of information source. For instance,

- a). *It's probably raining.*
- b). *It must be raining.*
- c). *It sounds like it's raining.*
- d). *I think/guess/suppose it's raining.*
- e). *I can hear/see/feel/smell it raining.*

It is obvious that the information provided in the above examples is subjective. The information expresses the personal experience or attitudes, while at the same time reflects the speakers' estimation for the trustworthiness of the statement by information providers.

3.1 The Definition of Evidentiality

There are two dimensions of the linguistic definition for evidentiality. The term evidentiality is originally introduced by Jakobson (1957) as a label for the verbal category indicating the alleged source of information about the narrated events. In line with Jakobson's definition, the narrow definition of evidentiality proposed by other researchers focuses mainly on the specification of the information sources, that is, the evidence through which information is acquired (DeLancey, 2001). Comparing with the narrow definition, the board definition explains evidentiality in a much wider sense, and characterizes evidentiality as expressions of speaker's attitude toward information, typically expressed by modalities (Chafe, 1986; Mushin, 2000).

Ifantidou (2001) also holds that evidential has two main functions: 1) indicating the source of knowledge; 2) indicating the speaker's degree of certainty about the proposition expressed. He further divides them in details as follows.

a) Information can be acquired in various ways, including observation (e.g. *see*), hearsay (e.g. *hear, reportedly*), inference (e.g. *must, deduce*), memory (e.g. *recall*).

b) Evidentiality can indicate the speaker's degree of certainty, including certain propositional attitude (e.g. *think, guess*) and adverbials (e.g. *certainly, surely*), also epistemic models (e.g. *may, ought to*).

3.2 The Taxonomy of Evidentials

Evidentiality has its hierarchy which forms a continuum that marks from the highest to the least trustworthiness. Up to now, there are many hierarchical schemes proposed by researchers.

	Absolute	High	Moderate	Low
Attributive/modal adverb	<i>certainly, sure, of course, definitely, absolutely, undoubtedly</i>	<i>clearly, obviously, apparently, really, always</i>	<i>Seemingly, probably</i>	<i>maybe, personally, perhaps, possibly, presumably</i>
Lexical verb	<i>report, certain</i>	<i>believe, see</i>	<i>seem, think, sound</i>	<i>doubt, wish, wonder, infer, assume, forecast, fell, heard</i>
Auxiliary verb		<i>must</i>	<i>ought, should, would, could, can</i>	<i>may, might</i>
Epistemic adjective	<i>definite</i>		<i>possible, likely, unlikely, probable, positive, potential</i>	<i>not sure, doubtful</i>

Table 1. The Categorization and Inside Items of Evidentiality

Oswalt (1986) suggests a priority hierarchy of evidentials as:

Performative > Factual > Visual > Auditory > Inferential > Quotative

In this evidential hierarchy, *performative* carries the highest degree of trustworthiness since Oswalt considers that the speaker is speaking of the act he himself is performing. It is the most reliable source of evidence for the knowledge of that event.

Whereas Barners (1984) proposes the following hierarchy:

Visual > Non-visual > Apparent > Second-hand > Assumed

He points out that visual evidence takes precedence over the auditory evidence and is more reliable.

The above two hierarchies are based on the narrow definition of evidentiality mentioned above. There are also some hierarchies involving the board definition of evidentiality, such as Chafe (1986)'s categories of evidentiality.

In this paper, we adopt a broad definition of evidentiality and focus on a trustworthiness categorization. This categorization follows the model of four-dimensional certainty categorization by Rubin et al (2005). In this model, it is suggested that the division of the certainty level dimension into four categories - *Absolute, High, Moderate* and *Low*. With some revision, there are different items of evidential words and phrased that we extracted from the corpus. These items from each category to be adopted in our experiments are presented in Table 1.

4 Incorporating Evidentiality into Machine Learning for Trustworthiness Detection

In this section, we apply evidentiality in an actual implement of text trustworthiness detection.

It is based on a specific web application service, collaborative question answering (CQA), in which the trustworthiness of text content is very helpful for finding the best answers in the service.

With the development of Web2.0, the services of CQA in community media have largely attracted people's attention. Comparing with the general ad hoc information searching, question answering could help in finding the most accurate answers extracted from the vast web content. Whereas in the collaborative question answering, the CQA community media just provide a web space in which users can freely post their questions, and at the same time other users may answer these questions based on their knowledge and interests. Due to the advantage of interactivity, CQA usually could settle some questions which cannot be dealt with by ad hoc information retrieval. However, since the platform is open to anyone, the quality of the answers provided by users is hard to identify. People may present answers of various qualities due to the limitation of their knowledge, attitude and purpose of answering the questions. As a result, the issue of how to identify the most trustworthy answers from the user-provided content turns out to be the most challenging part to the system.

As mentioned previously, the trustworthiness of text content could be identified from two dimensions. The first one relies on the features related with information distributors. The second one relies on the content of a text. In current research we focus on textual features, especially the feature of evidentiality in texts. The feature will be incorporated into a machine learning based text classification framework in order to identify the best answers for CQA questions.

4.1 The Dataset

For the experiments, we use the snapshot of Yahoo! Answers dataset which is crawled by Emory University³. Since our experiments only involve text features, we use the answer parts from it without considering the question sets and user profiles. Such information could be incorporated to achieve a higher performance in the future.

With regard to the text classification problems, there is typically a substantial class distribution skew (Forman, 2003). For the Yahoo! Answers dataset, a question only has one best answer and accordingly all the other answers will be marked as non-best answers. Thus the class of best answer contains much fewer texts than the class of non-best answers. In our dataset (a proportion of the overall CQA dataset provided by Emory University), the number of best answers is 2,165, and the number of non-best answers is 17,654. The proportion of the size of the two answer sets is around 1:8.15, showing a significant skew. For a better comparison of experimental results, we use a balanced dataset which is generated from a normal distribution dataset.

A 10-fold validation is used for the evaluation, where the datasets of best and non-best answers are divided into 10 subsets of approximately equal size respectively. In the normally distributed dataset, we use one of the ten subsets as the test set, while the other nine are combined together to form the training set. In the balanced dataset, for each subset of the non-best answers, we only use the first k answers, in which k is the size of each subset of best answers. The training data and test data used in the machine learning process are shown in Table 2.

	Training /Test Set	Best answer	Non-best answer
normal distribution dataset	training	19,490	158,889
	test	2,165	17,654
balanced dataset	training	19,490	19,490
	test	2,165	2,165

Table 2. The Dataset Used for the Experiments

4.2 Experiment Settings

To conduct a machine learning based classification for best answers and non-best answers, we first need to construct the feature vectors. The representation of text is the core issue in the ma-

chine learning model for text classification. In text domains, feature selection plays an essential role to make the learning task efficient and more accurate. As the baseline comparison, we use the following feature vector settings.

- **Baseline1** represents using all the words in the text as features (when the frequency of the word in the dataset is bigger than a predefined threshold j).

- **Baseline2** represents using all the content words (here we include the four main categories of content words - nouns, verbs, adjectives and adverbs identified by a POS tagger) in the dataset as features.

We use both the above two baselines. The bag-of-words model of Baseline1 is a conventional method in text representation. However, since not all the words are linguistically significant, in Baseline2, we consider only the content words in the dataset, since content words convey the core meaning of a sentence.

For the evidentiality-based classification, we adopt the following feature vector settings.

- **Evidential** represents using all the evidentials in text as features.

- **Evidential'** represents using all the evidentials except for those in the category of *Moderate* as features.

- **Evid.cat4** represents using the four evidentiality categories of *Absolute*, *High*, *Moderate* and *Low* from Table 1.

- **Evid.cat2** represents using the two categories of *Absolute* and *High* as the positive evidential and *Moderate* and *Low* as the negative evidential.

- **Evid.cat2'** omits the evidential category of *Moderate*, and represents using the two categories of *Absolute* and *High* as the positive evidential and only the category of *Low* as the negative evidential feature.

Some researchers have proved that usually a Boolean indicator of whether the feature item occurred in the document is sufficient for classification (Forman, 2003). Although there are also some other feature weighting schemes such as term frequency (TF), document frequency (DF), etc, comparison of these different weighting schemes is not the object of the current research. So in this paper, we only consider Boolean weighting. In the Boolean text representation model, each feature represents the Boolean occurrence of a word, evidential, or evidential category according to the different feature settings. By the experimental settings, we want to verify the hypothesis that incorporating the knowledge

³ <http://ir.mathcs.emory.edu/shared>

of evidentiality into text representation can lead to improvement in classification performance.

In our experiment, we perform text preprocessing including word segmentation and part-of-speech (POS) tagging. The Stanford Log-linear Part-Of-Speech Tagger (<http://nlp.stanford.edu/software/tagger.shtml>) is used for POS tagging. We adopt support vector machine (SVM) as the machine learning model to classify best answers from non-best ones, and use the SVMlight package (<http://svmlight.joachims.org>) as the classifier with the default parameters and a linear kernel. For the evaluation, we use the metrics of precision (**Prec.** as in table 3), recall (**Rec.** as in table 3), accuracy (**Acc.** as in table 3) and **F1**: F₁-measure, the harmonic mean of the precision and recall.

4.3 Evaluation

Table 3 shows the experimental results using the balanced dataset with Boolean weighting. The focus of the experiment evaluation is on identifying the best answers, so the evaluation metrics are all for the best answers collection. From the table, we see increases of the two feature vector setting of evidentials over both baseline results. The highest improvement is 14.85%, achieved by the feature set of Evidential'. However, there is no increase found in the settings of using evidential categories. This means that although the category of evidentials in indicating text trustworthiness is obvious for human, it is not necessary a preferred feature for machine learning.

	Prec.	Rec.	Acc.	F1
Baseline1	45.62%	51.51%	45.15%	47.94%
Baseline2	59.58%	39.20%	56.30%	47.28%
Evidential	67.78%	44.18%	61.59%	53.49%
Evidential'	47.40%	90.12%	45.06%	62.13%
Evid.cat4	64.15%	25.85%	55.70%	36.85%
Evid.cat2	60.86%	28.21%	55.03%	38.55%
Evid.cat2'	40.35%	25.85%	43.81%	31.51%

Table 3. Experimental Results Using the Balanced Training/Test Dataset (with Boolean Weighting)

To eliminate the potential effect of term weighting scheme on performance trend among different text representation models, we also conduct experiments using TF weighting. By the

experiments, we aim to compare the relative performances of different feature vectors constructed with evidentials, and the results are demonstrated in Table 4.

	Prec.	Rec.	Acc.	F1
Evidential	66.78%	45.57%	61.45%	54.17%
Evidential'	59.66%	20.82%	53.37%	30.87%
Evid.cat4	50.00%	18.14%	50.00%	26.63%
Evid.cat2	55.91%	16.39%	51.73%	25.35%

Table 4. Experimental Results Using the Balanced Training/Test Dataset (with TF Weighting)

From the table, it can be observed that using evidentials as features shows better improvement in the performance than the category of evidentials as a feature. A similar performance has been summarized in Table 3.

Finally, but not the least, to better understand the effect of evidential category on the machine learning performance, we design additional experiments as follows.

- **Evid_cat1** stands for combining the four evidential categories into one, and uses only this one category of evidential as a feature. The approach of Boolean weighting is actually the same as a rule-based approach that classifies the test dataset according to whether evidential occurs or not.

BOOL	Prec.	Rec.	Acc.	F1
Evid_cat1	59.42%	61.59%	59.76%	60.49%

Table 5. Experimental Results Using the Balanced Training/Test Dataset (with Boolean Weighting; Only One Evidential Category)

Table 5 presents a set of interesting experimental result. In the result, all the four evaluation metrics show performance increases comparing to the baseline, and it even outperforms almost all the other results from both weighting schemes. Based on this result, it is suggested that evidentiality still may contribute to the task of text trustworthiness detection. Moreover, it can significantly reduce the dimensionality of feature space (e.g. for Baseline 1, the dimensionality of feature dimension is 218,328 in one of our cases; while for the experiment of Evidential, it reduced to only 51 as shown in Table

1). However, we should address the question of why not all types of evidential features demonstrate improvement of detection. We will further discuss the issue from a pragmatic viewpoint in the next section.

5 Conclusion and Discussion

In this paper, we propose to incorporate the linguistic knowledge of evidentiality in the NLP task of trustworthiness prediction. As evidentiality is an integral and inherent part of any statement and explicitly expresses information about the trustworthiness of this statement, it should provide the most robust and direct model for trustworthiness detection. We first set up the taxonomy of lexical evidentials. By incorporating evidentiality into a machine learning based text classification framework, we conduct experiments for a specific application, CQA. The evidentials in the dataset are extracted to form different text representation schemes. Our experimental results using evidentials show improvements up to 14.85% over the baselines. However, not all types of evidential features contributed to the improvement of detection. We also compared the effect of different types of evidential based feature representation schemes on the classification performance.

The way to model evidentiality for trustworthiness detection which we adopted in our initial experiment design actually could also be explained by Grice's Maxim of Quality: be truthful. As the Maxim of Quality requires one "not to say that for which one lacks adequate evidence", we hypothesize that evidential constructions mark the adequacy of evidence and should indicate reliable answers. However, the results from our experiments only partially supported this hypothesis. The results showed a satisfactory performance was achieved when all evidential markers were treated as negative evidence for reliability. This result could then be accounted by invoking another Gricean maxim: Quantity.

The Maxim of Quantity requires that "one makes his/her contribution as informative as is required, and at the same time does not make the contribution more informative than is required." As evidentiality is not grammaticalized in English, the use of evidentiality is not a required grammatical element. An answer marked by evidentials would violate Maxim of Quantity if it is correct. The Maxim of Quantity predicts that good answers are plain statements without evidential markers. On the frequent use of eviden-

tial markers for less reliable answers can be accounted for by speakers' attempt to follow both Maxims of Quality and Quantity. The evidential marks are used to compensate for the fact that speakers are not very confident about the answer, yet would like to adhere to the Maxim of Quality. In other words, evidentials are not likely to be used in reliable answers because of the Maxim of Quality, but it is likely used in less reliable answers because the speakers may try to provide proof of adequate evidence by a grammatical device instead of providing true answer.

Therefore, this model elaborated above takes into account not only the grammatical function of evidential constructions but also how this linguistic structure is used as a pragmatic/discourse device. In other words, this study suggests that modeling linguistic theory in NLP needs to take a more comprehensive approach than the simple modular approach where only one module (based on evidentiality) is used. Linguistic modeling needs to consider both how linguistic structure/knowledge is represented and processed, we also need to model how a particular linguistic device is used.

In the further works, we plan to continue developing and elaborate on a multi-modular linguistic model of evidentiality for knowledge acquisition. We will also explore the possibility of incorporating other features, both textual and non-textual, to further improve performances in the tasks of text trustworthiness detection.

References

- Agichtein E, Castillo C, and etc. 2008. Finding high-quality content in social media. In *Proceedings of WSDM2008*.
- Aikhenvald A and Dixon, ed. 2003. *Studies in evidentiality*. Amsterdam/Philadelphia: John Benjamins Publishing Company
- Banerjee P, Han H. 2009. Credibility: A Language Modeling Approach to Answer Validation, In *Proceedings of NAACL HLT 2009*, Boulder, Bolorado, US
- Barners J. 1984. Evidentials in the Tuyuca Verb. IN *International Journal of American Linguistics*, 50
- Bouguessa M, Dumoulin B, Wang S. 2008. Identifying Authoritative Actors in Question-Answering Forums - The Case of Yahoo! Answers, In *Proceedings of KDD'08*, Las Vegas, Nevada, USA
- Chafe W. 1986. Evidentiality: The Linguistic Coding of Epistemology, Evidentiality in English Conversation and Academic Writing. In Chafe and Nich-

- ols, (ed.). *Evidentiality: The Linguistic Coding of Epistemology*. Norwood, NJ: Ablex
- DeLancey S. 2001. The mirative and evidentiality. In *Journal of Pragmatic*, 33
- Forman G. 2003. An Extensive Empirical Study of Feature Selection Metrics for Text Classification, In *Journal of Machine Learning Research*, 3
- Gil Y, Artz D. 2006. Towards Content Trust of Web Resources, In *Proceedings of the 15th International World Wide Web Conference*, Edinburgh, Scotland
- Gyongyi Z, Molina H, Pedersen J. 2004. Combating Web Spam with TrustRank. In *Proceedings of the 30th VLDB Conference*, Toronto, Canada
- Ifantidou E. 2001. Evidentials and Relevance. *John Benjamins Publishing Company*.
- Jeon J, Croft W, Lee J and Park S. 2006. A Framework to Predict the Quality of Answers with Non-textual Features, In *Proceedings of SIGIR '06*, Seattle, Washington, USA
- Leopold E, Kindermann J. 2002. Text Categorization with Support Vector Machines. How to Represent Texts in Input Space?, In *Machine Learning*, 46, 423-444
- Oswalt R. 1986. The evidential system of Kashaya. IN Chafe W and Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex
- Rubin V, Liddy E, Kando N. 2005. Certainty Identification in Texts: Categorization Model and Manual Tagging Results, In Shanahan J and et al (Eds.), *Computing Attitude and Affect in Text: Theory and Applications (The Information Retrieval Series)*: Springer-Verlag New York, Inc.
- Rubin V, Liddy E. 2006. Assessing Credibility of Weblogs, In *Proceedings of the AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs (CAAW)*
- Magnini B, Negri M, Prevete R and Tanev H. 2002. Is It the Right Answer? Exploiting Web Redundancy for Answer Validation, In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, PA
- Metzger M. 2007. Evaluating Online Information and Recommendations for Future Research, *Journal of the American Society for Information Science and Technology*, 58(13)
- Mushin I. 2000. Evidentiality and Deixis in Retelling, In *Journal of Pragmatics*, 32
- Weerkamp W, Rijke M. 2008. Credibility Improves Topical Blog Post Retrieval. In *Proceedings of ACL08: HLT*
- Zhang J, Ackerman M, Adamic L. 2007. Expertise Networks in Online Communities: Structure and Algorithms. In *Proceedings of the 16th ACM International World Wide Web Conference (WWW'07)*