Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study

Deborah L. McGuinness[1], Honglei Zeng[1], Paulo Pinheiro da Silva[2],
Li Ding[3], Dhyanesh Narayanan[1], Mayukh Bhaowal[1]

[1]Knowledge Systems, AI Lab, Department of Computer Science
Stanford University, California, USA
{dlm, hlzeng, dhyanesh, mayukhb}@ksl.stanford.edu
[2]Department of Computer Science
University of Texas at El Paso, Texas, USA
paulo@utep.edu
[3]Department of Computer Science
University of Maryland, Baltimore County, Maryland, USA
dingli1@umbc.edu

ABSTRACT

As collaborative repositories grow in popularity and use, issues concerning the quality and trustworthiness of information grow. Some current popular repositories contain contributions from a wide variety of users, many of which will be unknown to a potential end user. Additionally the content may change rapidly and information that was previously contributed by a known user may be updated by an unknown user. End users are now faced with more challenges as they evaluate how much they may want to rely on information that was generated and updated in this manner. A trust management layer has become an important requirement for the continued growth and acceptance of collaboratively developed and maintained information resources. In this paper, we will describe our initial investigations into designing and implementing an extensible trust management layer for collaborative and/or aggregated repositories of information. We leverage our work on the Inference Web explanation infrastructure and exploit and expand the Proof Markup Language to handle a simple notion of trust. Our work is designed to support representation, computation, and visualization of trust information. We have grounded our work in the setting of Wikipedia. In this paper, we present our vision, expose motivations, relate work to date on trust representation, and present a trust computation algorithm with experimental results. We also discuss some issues encountered in our work that we found interesting.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.3.5 [**Online Information Services**]: [Data Sharing, Web-based services]; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods

General Terms

Design, Languages, Management

Keywords

Trust, Wikipedia, Inference Web, Proof Markup Language, Open Editing

1. INTRODUCTION

One emerging pattern for building large information repositories is to encourage many people to collaborate in a distributed manner to create and maintain a repository of shared content. The notion of open editing has grown in popularity along with the notion of a Wiki, which in its simplest form allows users to freely create and edit web pages[1]. Wikipedia [1] is one popular Wiki that is a freely available online encyclopedia. Its size and diversity is one aspect of it that makes it an interesting motivating use case for our work. It has more than 900,000 registered authors[2] and three million articles. It has become perceived as a valuable resource and many people cite it as a credible information source. While recent studies (e.g. [2]) show that the science articles in Wikipedia are generally trustworthy, there have been some reports of claimed inaccuracies appearing in Wikipedia. For example, there was a widely reported situation where a journalist and a former official in the Kennedy administration, stated that Wikipedia contained an inaccurate biography article about him in 2005 [3]. The media coverage led to discussions about trustworthiness of content sources that have fairly liberal editing policies and also led to changes in Wikipedia's editing policy of anonymous authors.

One of the strengths of a collaborative information repository is that it may benefit from contributions of a wide diversity of users. Of course some of these users will have expertise levels that are untested and unknown to some end users. Additionally content in these repositories may change rapidly. Thus, trust management has become a critical component of such a system design. Without some form of trust management, these kinds of collaborative information repositories will have difficulty defending any particular level of authoritativeness and correctness. Additionally, without some notion of accountability in addition to the trust, these systems will only be able to provide end users with information but not with information about where the information came from and how trustworthy that source might be. The popular large implementations such as Wikipedia are currently addressing some of these issues, although currently not to the level that they will need to in the long run if they are to achieve their true potential.

Our work focuses on designing and building an extensible trust framework. We are investigating representation needs for the encoding of trust, methods for computing trust, and visualization of

---

[1]http://wiki.org/wiki.cgi?WhatIsWiki
[2]http://en.wikipedia.org/wiki/Special:Statistics

information that is informed by trust encodings. In our previous work on Inference Web, we have been designing and implementing an infrastructure for explaining answers from intelligent applications. One information source for these applications may be a collaboratively generated information repository such as Wikipedia. Our work on explaining answers focused us on where information came from and how it was manipulated to generate an answer. This work has also led us to investigate forms of trust encodings for information.

As we began to look more closely at aggregated information sources and collaborative, evolving information sources such as Wikipedia, we have found even more requirements for trust formulation. It is worth noting that an open (or mostly unrestricted) editing environment is quite different from some other social networks (e.g., eBay and Epinions) that have addressed trust. These social networks may be viewed as focusing on interactions between users while generating growing content but not typically generating changing content. For example, a transaction on eBay or a review on Epinions is typically created once and then remains unchanged. On the other hand, the content of collaborative information repositories like Wikis may be quite dynamic as it may be continually reviewed, shared, and updated by many different users. Trust formulation and requirements for rapidly changing repositories thus may be quite different from (mostly) monotonically growing repositories even though both may be perceived as trust problems.

Some social networks that have trust approaches that rely on explicit assertion of trust in a user resulting from feedback from transactions or ratings. Trust in Wikipedia has not been addressed explicitly in this manner. We began exploring the view that trust may be viewed as an implicit feature of the environment and we began looking for ways to make trust levels explicit and inspectable.

Significant research has been done on trust in various contexts (e.g., [4],[5]); however, most of the work assumes homogeneous context. Encryption and authentication (e.g., [6]) help secure trustworthiness in terms of the integrity and authenticity of information through pre-defined representation and functions. Distributed trust management (e.g., [7]) offers a flexible policy framework for judging if a person is trustworthy enough to perform an action through a common policy ontology and corresponding policy inference engine. Reputation systems (e.g., [8], [9]), and trust networks (based on social networks or P2P network) (e.g., [10],[11]) help compute trustworthiness of a person or an entity; again, using a pre-defined trust ontology and a common computation method.

The Web offers easy access to information from various sources and computational services at different locations. Thus, distributed web environments provide diverse and heterogeneous settings for trust researchers. For repositories of information like Wikipedia, trustworthiness information concerning an article or an author could be computed and published by many sources with varying degrees of reliability. When an end user is evaluating how to use (portions of) a Wikipedia article, it may be useful to view an aggregation of the trust information available concerning the article. The end user may thus want to effectively combine trust information from multiple sources using different representation schemes potentially using personalized trust computation methods. Unfortunately, research focused on enabling this scenario is sparse. Our investigations have been driven by our desire to work on distributed, heterogeneous, collaborative environments such as the web in general and collaborative, evolving information repositories in particular. Our goal is to provide an open, interoperable, and extensible framework that can provide a solution framework to the problems of trust we mentioned above.

In the way of background, Inference Web (IW) [12] enables Se-

mantic Web applications to generate portable proofs that contain information required to explain answers. One of challenge for users of any explanation system is evaluating trustworthiness of answers. Presentations of knowledge provenance, sources used and information manipulation steps performed to produce an answer help. It is also important to know how trustworthy any particular piece of information is, how trusted the author is etc. We thus have been motivated to add a trust representation extension to the Proof Markup Language. We will report here on our extension and describe how we are and plan to use it in our case study using Wikipedia.

We view Wikipedia as an example of a collaborative, evolving information repository that has variety in quality and coverage of its subject matter. We were inspired to look at Wikipedia as a case study for our trust extension work for the following reasons: (i) it is a large and growing collaborative repository yet is contained. It can be viewed as large enough to provide challenges of scale and trust. (ii) it stores much rich provenance information in comparison to typical collaborative information repository. (iii) it is in need of a trust solution.

Additionally, we believe that trust relationships can be computed from information contained and maintained by Wikipedia. Further, we believe that a solution infrastructure appropriate for Wikipedia may be widely reusable in other online system settings.

The rest of our paper is structured as follows. In section 2, we provide a vision of how we will use trust values once available to present trust information to users. We do this by describing a customizable trust view of information. In section 3, we show a citation-based approach, the link-ratio algorithm for computing trust. In section 4, we present some experimental results using the link-ratio algorithm in Wikipedia. In section 5, we discuss the implications of citation trust in Wikipedia and related work. We conclude our paper with a discussion of related work and future work.

Contributions presented in this paper to trust formulation in open collaborative, evolving settings include: an extension to the Proof Markup Language that creates a proof interlingua capable of encoding trust, a citation based trust algorithm (Link-ratio trust) designed to demonstrate our computational component and explore some characteristics of trust in Wikipedia; and a customizable visualization component for presenting Wikipedia content in a manner that has been informed by trust information.

2. TRUST TAB

In order to extend Wikipedia with a trust management component, we propose a new "trust" tab associated with each Wikipedia article. This trust tab will appear in addition to the conventional tabs of Wikipedia, i.e., "article", "edit", "history" and "discussion". The motivation is to render Wiki articles in ways that users can visually compare and identify text fragments of an article that are more (or less) credible than other fragments. The trust tab is supposed to be a primary tool for helping users to decide how much they should trust a particular article fragment. The rendering of each text fragment is to be based on degrees of trust. These degrees of trust may be between individual authors or they may be aggregated and thus may be viewed as a community trust level associated with an author of each fragment of the document.

Our present endeavor is to calculate and display trust information based on information already available in the Wikipedia and without the use of any external information sources, e.g., Wikipedia users. In the future, we will extend this approach to include feedback from external sources so as to inform the trust calculations with a wider set of input.

The trust tab is an addition to the conventional article tab in the sense that, when compared to the article tab, it adds a colored back-
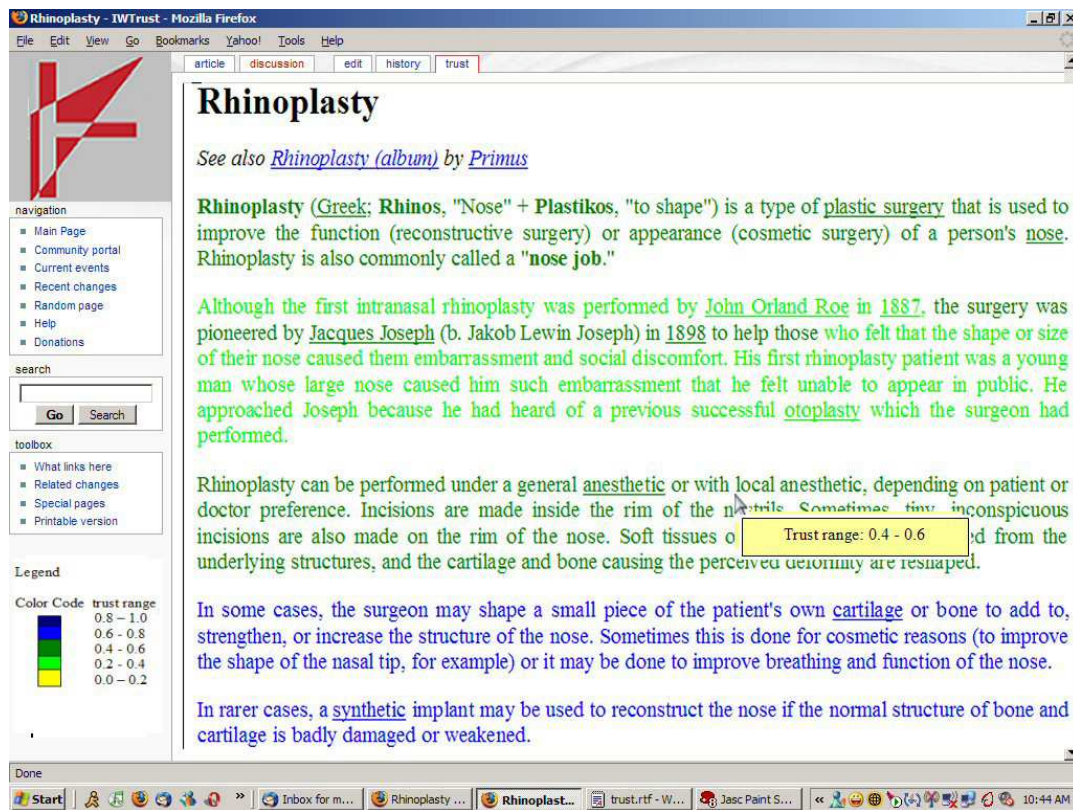
Figure 1: A Trust Tab Example in Wikipedia.

ground to text fragments in the article as shown in Figure 1. The new background color conforms to a color scheme which makes the presentation and its inherent meaning in terms of trust obvious and comprehensive.

According to the color code legend in the Figure 1, the degrees of aggregated trust of the fragments in the Rhinoplasty article range from 0.2 to 0.8 in a scale [0,1] where 0.0 is the total absence of trust and 1.0 is the total presence of trust. The exact meaning of this scale of trust is irrelevant for the trust tab that aims to provide a visual mechanism to compare the parts of the page that are more or less credible. The relative differential between the trust values is information that is useful to the end user. For instance, the trust tab says that the last fragment composed of the two last paragraphs of the page has a higher degree of trust than any other fragment in the page. Moreover, the second paragraph has the lowest degree of trust although the fragment "the surgery (...) in 1898 to help those" inside the paragraph have been added by a more credible author. [3]

The implementation of the trust tab has raised several issues related to Wikipedia. In the rest of this section, we briefly describe an approach to implement the trust tab. We will also present some experimental results of our effort to compute aggregated degrees of trust for the authors of article fragments as required for rendering useful trust tabs when no personalized trust relations are used.

2.1 Fragment Identification

The trust tab relies on the fact that Wikipedia articles can be seg-

mented into a sequence of text fragments where each fragment has a single author. We assume that several fragments in the article can have a single author. In order to compute a trust level for each fragment, the trust tab needs: (i) to identify each individual fragment in the article; (ii) to identify the author (and time stamp) of each fragment; and (iii) to compute a degree of trust for each author.

The Wikipedia database schema does not store individual fragments although it archives complete revisions of articles. Thus, one approach to fragment identification is to compare successive article revisions, e.g., using diff, and identify changes. Note, the granularity of the difference measure used is something we are exploring. By performing successive comparisons, the trust tab retrieves the individual fragments of an article as required in (i). Simultaneously, it identifies the time stamps and authors for the fragments as required in (ii). Trust computation associated with authors is discussed below in Section 3.

2.2 Provenance Annotation

Even though manual monitoring on Wikipedia has been enhanced recently, there may always be some users who will want information about degrees of trust in particular authors. Additionally, some malicious authors or programs may attempt to insert inappropriate or unwanted content in collaborative open systems like Wikipedia. As these systems grow, any level of manual monitoring will not be adequate since it will not be able to scale with the content size. Automatic methods are required to augment administrator's abilities to monitor updates and to help manage their workloads. Automated tools built upon the trust values may substantially improve the trustworthiness of Wikipedia: for example, as mentioned above, a trust tab implementation may provide users with trust information about

---

[3]The actual trust values used to render this page are just for expository purposes and are not intended to reflect that actual trust levels for this page; the figure is manually generated for demonstration purposes.

the articles they are viewing and help them to decide how much they should trust the articles.

Our trust tab approach depends on a mechanism for storing trust relations between authors as well as aggregated degrees of trust inferred from the Wikipedia content. This new stored content however, may not be enough to capture some important trust aspects of the system since Wikipedia is managed in a centralized manner. For instance, we still need to face two important issues in representing and obtaining knowledge provenance: (iv) how to capture provenance information not originally written by a user, e.g. a user may copy and paste some content from the Web to an Wiki article; and (v) how to make trust computation components independent of data storage.

For (iv), we need a more comprehensive vocabulary for annotating the provenance information. We are using the provenance part of Proof Markup Language (PML) [13] to fulfill this job. Beside person, PML also identifies many other types of information sources including website, organization, team, publication, and ontology. Upon updating a Wikipedia article, the editor may provide additional justification for his/her modifications. For example, when an editor adds one definition to an article, he/she may also specify that the definition is obtained from an online article and even specify the location of the related span of text.

For (v), we need explicit representation of provenance information. This is especially helpful when integrating multiple knowledge repositories which are managed independently. Our solution is to use the RDF/XML serialization of PML. To implement this idea, our design adds another "provenance" tab and exposes PML provenance information in RDF/XML format to agents (or web services) which are capable of computing trust using provenance information.

```
<iw:NodeSet
    rdf:about="http://en.wikipedia.org/wiki/Stanford">
 <iw:hasConclusion>"Article Fragment"</iw:hasConclusion>
 <iw:hasLanguage>en</iw:hasLanguage>
 <iw:isConsequentOf>
   <iw:InferenceStep>
     <iw:hasRule rdf:resource=
"http://iw.stanford.edu/registry/DPR/Told.owl#Told"/>
     <iw:hasSourceUsage>
       <iw:SourceUsage>
         <iw:hasAuthor >Harry</iw:hasAuthor>
         <iw:hasTimestamp>20051109</iw:hasTimestamp >
         <iw:hasParentID>2425693</iw:hasParentID >
       </iw:SourceUsage>
     </iw:hasSourceUsage>
   </iw:InferenceStep>
 </iw:isConsequentOf>
</iw:NodeSet> <iw:TrustRelation>
 <iw:hasTrustingParty rdf:resource=
"http://iw.stanford.edu/registry/ORG/Wikipedia.owl"/>
 <iw:hasTrustedParty>Harry</iw:hasTrustedParty>
 <iw:hasTrustValue>0.434</iw:hasTrustValue>
</iw:TrustRelation>
```

Figure 2: PML provenance annotation

The next step is to encode the trust information in PML. Figure 2 shows an example of such an encoding. In this example, Harry is the author of a fragment in the Stanford page and the Wikipedia community has an aggregated degree of trust of 0.434 in Harry. The use of a float for hasTrustValue is a simplification of the PML capabilities for representing trust values. More sophisticated, realistic approaches are discussed in [14]. PML encodings can then be used by automated programs for other presentations of trust information, or for use in more complex reasoning and question answering applications that may want to use trust input for filtering,

thresholding, etc.

2.3 Provenance Visualization

The trust tab applies conventional rendering techniques used by the article tab for rendering so that the typical style of articles is preserved in the trust tab. In addition to the use of these techniques, the trust tab also compares the content of the article with the PML encoding of the article. The trust tab views the PML encoding to be metadata for the page in the article tab. By comparing the page content with its PML encoding, the trust tab identifies fragments and the fragment authors. It also retrieves a pre-computed aggregated degree of trust for each author as stored in the newly created storage for trust in the Wikipedia database. From these degrees of trust and a color schema, the trust tab eventually identifies and sets the appropriate background color for each fragment.

3. CITATION-BASED TRUST

3.1 Trust issues in Wikipedia

In our work, we begin by considering how citation-based measures may be used to determine trust values. In some settings, an end user may be more inclined to rely on the content in a news story from a reputable newspaper, such as the New York Times, over the content that is published on a personal Blog, especially if the end user has no knowledge of the Blog or its author.

One way of computing trust of an author is to take an aggregated value from trust rankings of all of the articles written by the author. In order to share and visualize such trust information, we formalize trust as a numerical value between 0 and 1 and we view it as a measure of trustworthiness. In our setting, a value of 1 represents complete trust and value 0 represents unknown trustworthiness. Note, this differs from some approaches where a value 0 is interpreted as complete distrust. Although we have chosen a rather simplistic trust model in this work, we are also evaluating other, more sophisticated trust models that we may use to enhance our current model.

In this work, citation-based algorithms are a family of algorithms that derive trust based on citation relationships among entities. We refer to such derived trust as citation-based trust, or simply citation trust. We ground our work in Wikipedia and use it as a sandbox for evaluating citation trust.

One distinguishing characteristic of Wikipedia articles in comparison to general web documents is that Wikipedia articles are meant to be encyclopedia entries. We will refer to the title of a Wikipedia article (e.g. "Gauss's law"), as an *encyclopedia index term*. We note that encyclopedia index terms may occur, with or without citation, in other articles in Wikipedia. Since Wikipedia is an encyclopedia, one might expect that occurrences of encyclopedia index terms in other articles would refer back to the encyclopedia index term article, and in fact if a term appears but does so without citation, it might be viewed as a negative indicator of the quality of the index term entry. We will explore this notion and compute the number of non-citation occurrences of encyclopedia index terms. Two other useful measures of note in collaborative content settings are the number of citations a term (or article) receives and the citation trust of articles in which it is cited.

Consider the scenario where an article (i.e. its encyclopedia term) has many non-citation occurrences but few actual citations. One interpretation of this scenario is that the article may not be perceived to be worthy of a high trust value since few authors choose to cite the article when they mention the term [4]. In contrast, non-

---

[4] We will come back to this point in the discussion since another interpretation of a non-citation is simply ignorance of the article.

citation occurrences of a word or phrase on a typical web page may not mean anything about any associated trust levels since typical web page authors do not necessarily link every phrase that one would typically find in an encyclopedia to a web page describing the phrase.

In our work, we have begun explorations into citation ratios as a potential input to trust algorithms. In this paper, we will report on our investigations concerning link ratios. We define the Link-ratio of an article (i.e., the page with title x) as the ratio between the number of citations and the number of non-citation occurrences of the encyclopedia term x.

We provide the following motivation for exploring Link-ratio:

- Link-ratio is a trust measure unique to collaborative repositories of encyclopedic content. The fact that it is a ratio rather than a raw count of non-citation occurrences helps to minimize the impact of the difference between the numbers of occurrences of common vs. uncommon terms.

- Link-ratio is in the same family as the well respected PageRank [15], citation-based algorithm, which has been successfully used in many web settings. PageRank has also been studied in the context of Wikipedia. We will cite and discuss the results of this related research from other researchers ([16]).

- Unlike other social networks such as eBay, Wikipedia has no explicit trust assertions among authors and articles. Trust algorithms based on the transitivity property of trust cannot be directly applied without an initial set of trust values. Obtaining trust values manually for a content repository the size of Wikipedia is a large task. The Link-ratio approach may be used as one way to obtain initial trust values.

3.2 A Simple Wikipedia Model

Wikipedia may be (partially) characterized by the abstract model in Figure 3. Intuitively, Wikipedia consists of a set of articles (i.e. articles $d_1, d_2, ..., d_m$ in Figure 3). Each article (d_i) consists of a set of article fragments $(f_{i1}, f_{i2}, ..., f_{in_i})$, each of which is written by an author (a_j). An author may write more than one fragment in the same article. In addition, a fragment could link to other articles as citations. There are three types of links in Figure 3: author-fragment authorship links (solid lines from a_i to f_{jk}), fragment-article citation links (dotted lines from f_{ij} to d_k), and article-fragment membership links.
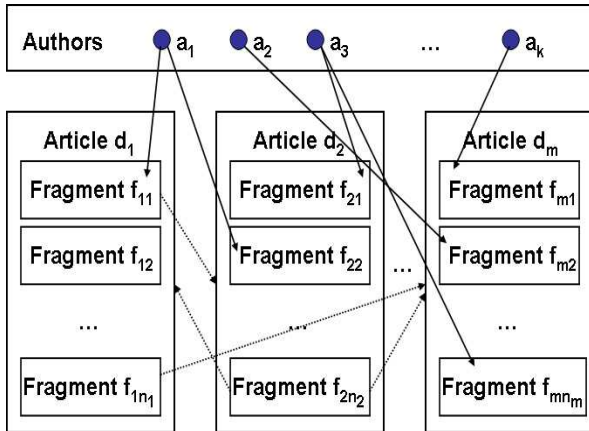

Figure 3: An Abstract Model of Wikipedia.

Our goal is to infer trustworthiness of authors, fragments and articles based on the above link structures. We also assume most Wikipedia authors have the genuine intention of providing accurate content.

In the following sections, we will show two citation-based trust algorithms, the Link-ratio algorithm and the PageRank algorithm. We will explain the link-ratio algorithm in detail but only briefly mention the well-known PageRank algorithm.

3.3 Link-ratio Algorithm

We first compute article-level trust in Wikipedia based on its rich citation structure. Assume d is an article, then $[[d]]$ refers to the hyperlink citation to this article d. For example, the article *Grape* refers to the article *Wine* by stating that "... used for making $[[wine]]$". When an article is linked to from another one, a certain trust is implied[5]. In this example, the author of *Grape* expresses his trust towards the article *Wine* by creating a citation to it. He believes that a user may benefit from further information on the wine topic by accessing the information contained in the article *Wine*.

In the link-ratio algorithm, we are interested in non-citation occurrences of an encyclopedia term. Thus, the algorithm looks for articles that contain a term d but do not link to article d. For example, in the article *Beer*, it is said that "Unfiltered beers may be stored much like wine for further conditioning ..." Both *Grape* and *Beer* mention the term "wine", but only *Grape* links to the article *Wine*. There may be many reasonable explanations for the omission of the wine citation in *Beer*: *Beer* may have been created before *Wine* was created; the author of *Beer* may be unaware that *Wine* exists; the *Beer* author may be in a hurry and may be limiting citations; the *Beer* author may not believe that the readers of this page need extra information on *wine*; or the author believes *Wine* is untrustworthy. Without further information, we are not able to determine the exact cause of a missing citation; therefore, we assume missing citations decreases the trustworthiness of an article that was not cited. Simultaneously, if one is keeping measures of how "known" a page it, the missing citation decreases this measure.

We define **Trust_doc**(d) to be the trust value of an article d. Based on the citation trust we discussed above, the more frequent $[[d]]$ occurs, the higher **Trust_doc**(d) is; the more non-citation occurrences of d are, the lower the trust value is.

$$\textbf{Trust\_doc}(d) = \frac{occurrences([[d]])}{occurrences([[d]]) + occurrences(d)} \quad (1)$$

Occurrences([[d]]) denotes the number of citations to an article d and *occurrences(d)* is the non-citation occurrences of term d. The citation trust is thereby defined to be the ratio between the occurrences of the citations to article d and the total occurrences of term d as a citation and a non-citation.

Wikipedia articles are often under constant revision. We refer to the change that an author commits in one edit session as *atomic change*. The latest version of an article can be simply viewed as the original article followed by a sequence of atomic changes. We define **Documents**(a) as the set of articles that author a has ever created and changed. We can calculate the *aggregated trust value* of an author a, **Trust_author**(a), based on the trustworthiness of **Documents**(a). Intuitively, the trust value of an author is an aggregated value of the trust values of all the articles he has contributed to. In Equation (2), we adopt the simple arithmetic mean, but other weighting functions are possible (e.g. weighted mean).

---

[5]This assumes that the link from the original text does not contain negative anchor text or description such as "examples of bad pages include $[[d]]$".

| **Documents**(a) | is the size of **Documents**(a), i.e., the number of articles that author a has contributed to.

$$\textbf{Trust\_author}(a) = \frac{\sum_{d \in \textbf{Documents}(a)} \textbf{Trust\_doc}(d)}{\mid \textbf{Documents}(a) \mid} \qquad (2)$$

One of our primary goals is to help users understand how much they should rely on information in articles. Since articles are composed of fragments, this also means that we want to help users compare trustworthiness of article fragments in the same article, each of which may be written by different authors. Since we have established author trust in Equation (2), we use a simple notion that assumes fragment trust is the same as the trust value of its author. If f is a fragment of an article and **Author**(f) denotes the author of this fragment, then we can define the trust of this fragment $Trust\_frag(f)$ as follows.

$$\textbf{Trust\_frag}(f) = \textbf{Trust\_author}(\textbf{Author}(f)) \qquad (3)$$

The notion of fragment trust being identical to author trust is a bit too simplisitic. Fragment trust may also depend on context. For example, Equation (3) would produce the same results for two article fragments from the same author, despite the possibility the author of is an expert on the topic of one fragment and is not an expert on the topic of another fragment.

Fortunately, Wikipedia classifies articles into different categories; for example, the Mathematics category is meant to hold articles about mathematics. If we define $c_1, c_2, ..., c_t$ to be the categories in Wikipedia, such that each of c_i is a collection of articles relating to the same topic, we can rewrite Equation (2) and Equation (3) to be topic-dependent.

$$\textbf{Trust\_author}(a, c_i) = \frac{\sum_{d \in \textbf{Documents}(a) \bigwedge d \in c_i} \textbf{Trust\_doc}(d)}{\mid \textbf{Documents}(a, c_i) \mid}$$
$$(4)$$

The trust of an author a on topic c_i ($Trust\_author(a, c_i)$) is the ratio between the average trust values of his contributed articles on topic c_i.

$$\textbf{Trust\_frag}(s) = \textbf{Trust\_author}(\textbf{Author}(s, c_i)) \qquad (5)$$

The trust of a fragment is now modified to be the trust of its author on the topic c_i, which the article of the fragment belongs to. Topic-specific trust may be viewed as a coarse approximation to context-based trust.

3.4 PageRank

We briefly mention the well known PageRank algorithm in this section as another example of citation-based approaches. PageRank is an algorithm for ranking web pages used by Google and other retrieval engines. Web pages that have high PageRank values are typically more highly regarded and trusted and many end users prefer to have them returned first.

According to [15], PageRank of a web page A is defined to be

$$\textbf{PR}(A) = (1 - d) + d\left(\frac{PR(t_1)}{C(t_1)} + ... + \frac{PR(t_n)}{C(t_n)}\right) \qquad (6)$$

In the Equation (6), $t_1, t_2, ..., t_n$ are pages linking to page A and $C(t_i)$ is the number of outgoing links that a page T_i has. d is a damping factor, empirically set to 0.85.

When calculating the PageRank of articles in Wikipedia, one can take two possible approaches:

a. Consider the presence of Wikipedia (as a collection of web pages) on the Web. This approach would take account into considerations the links between Wikipedia articles as well as the links from external websites to Wikipedia articles.

b. Consider Wikipedia as a set of interlinked articles in isolation and calculate the PageRank. This approach would account only for links that exist within Wikipedia. One could view it as an "internal PageRank" that is exclusive to the articles and associated citation structure in Wikipedia.

We are more interested in the second approach, because we intend to study the relative trustworthiness of articles within the Wikipedia collection. Consequently, allowing PageRank from external links to flow into this computation might not yield the desired results. Note that accounting for links from external pages would definitely help to account for added value to a Wikipedia article from the perspective of the entire Internet.

PageRank has been computed and studied in Wikipedia [16]. In section 5, we will cite and discuss the results, putting it in the context of citation trust and relating it to the Link-ratio algorithm and the general citation-based approach.

4. EXPERIMENTS

The main data set used in our experiments was the dump of the Wikipedia database taken in December, 2005. We computed the trustworthiness of Wikipedia articles using the link-ratio algorithm in Equation (1). In order to determine the citation trust of a given article, all the other articles in Wikipedia were parsed searching for the reference of the article under consideration, whether it was a plain occurrence or a linked reference.

The first experiment was to compute the link-ratio values of featured articles, normal articles, and clean-up articles in Wikipedia. Featured articles are expected to be the best articles in Wikipedia; they were reviewed for accuracy, completeness, and style by experts in the same fields. On the contrary, clean-up articles are those articles below the quality standard of Wikipedia and are viewed by editors as being in need of major revisions. Clean-up articles are typically manually marked by Wikipedia administrators or other authors. Normal articles are articles that are neither featured articles nor clean-up articles. Intuitively, featured articles are most trustworthy, clean-up articles are least trustworthy, and normal articles are somewhere in between.

We randomly chose 50 featured articles, 50 normal articles and 50 clean-up articles from the Geography category. Table 1 shows the average link-ratio values of each type of articles.

Table 1: Average link-ratio values of 50 articles in the Geography category

Type of the articles	Average Link-ratio value
Featured articles	0.34
Normal articles	0.26
Clean-up articles	0.21

As we may expect, featured articles have the highest link-ratio values while clean-up articles have the lowest value. The differences between normal articles and clean-up articles are rather small, possibly because normal articles have a wide range of trustworthiness and quality. In practice, we have viewed articles with a link-ratio over 0.30 as trustworthy, and articles with a value less than 0.15 as having unknown trustworthiness. For example, the article *Cleveland, Ohio* has a link-ratio 0.53, which means that over 50% of the times that the string "Cleveland, Ohio" occurs in documents, that string is linked to the article *Cleveland, Ohio*.

Our results are limited by the size of the article samples and their categorization. One source of rated articles was the class of featured articles. Unfortunately, currently, only 0.1% of Wikipedia articles are featured articles. In particular, there are less than 80 featured articles in the Geography category, which was our chosen topic area for evaluation. Since we are interested in topic-specific trust, lack of featured articles (and clean-up articles to a lesser extent) poses one challenge in evaluating the effectiveness of citation-based approach and other approaches, because there are no other explicit trust assertions in Wikipedia.

Our second observation is that the link-ratio value depends on not only the trustworthiness of an article but also on how "linkable" the encyclopedic index term is. For example, if one writes an article and it has the word "Love" in it, it is unlikely that the author will consider the linking the occurrence of the term "Love" to the article love. The author probably expects that readers of the new article know what the definition of love is and there is no need to link it to the encyclopedia entry. On the contrary, if one uses a scientific term such as "Gauss's law", it is likely that the author will consider linking to the encyclopedia article *gauss's law*, as the author may assume a typical reader may want more information concerning the topic. Thus the link-ratio result can be dependent on how common the term is as well as how likely it is to require supplemental information that is obtainable from an encyclopedic web page entry. In another example, names of famous people will have higher link-ratio values than those of general things like wine or coal. Table 2 shows increasing link-ratio values for terms that are less common and more specialized.

Table 2: Link-ratio values of common and less common cyclopedia terms

Type	Article	Value
General terms	English	0.003
	Love	0.004
	Beer	0.05
	Wine	0.06
General scientific terms	Broadcasting	0.02
	Electronics	0.07
Specialized scientific terms	Maxwell's equations	0.44
	Gauss's law	0.47
Names of famous people	John F. Kennedy	0.41
	Winston Churchill	0.59

Our third observation is that co-references of a term also plays an important role in determining the link-ratio value. For example, "Massachusetts Institute of Technology" has a much higher link-ratio value than its acronym "MIT", as shown in the Table 3. If an author writes the entire name as in the title, he likely does so as he specifically wants to link it to that article. After all, "Massachusetts Institute of Technology" is a more precise encoding than "MIT".

Table 3: Link-ratio values of Universities and their acronyms

Article	Link-ratio value
Massachusetts Institute of Technology	0.52
MIT	0.001
California Institute of Technology	0.69
Caltech	0.01
Carnegie Mellon University	0.65
CMU	0.002
University of California, Los Angeles	0.40
UCLA	0.15

5. DISCUSSION AND RELATED WORK

In general, our experiments support our intuition that the link ratio approach computes high trust values for specialized articles that are trustworthy. For example, we may conclude that the article *Lake Burley Griffin* is probably more trustworthy than the article *Lingaraj temple* since both terms are specialized geography names, and the former has a link-ratio 0.57 while the latter has only 0.1. This comparison of link ratio values was done between terms of the same type. Nevertheless, it is not informative to compare the link ratio value of *Lake Burley Griffin* article to the link-ratio value for the article on *Love*. When the link-ratio of an article is low, we can not determine whether it is because the article is untrustworthy or if it is low for another reason, such as would be the case for a common term like "love". Therefore, we interpret low link-ratio values as being of unknown trustworthiness, because we may not have sufficient information to determine its trustworthiness, not that we believe the article is untrustworthy. There are other considerations as well such as how new a page is - if the page has just been created, then there may be many non-citation occurrences of the phrase simply because the entry did not exist previously. This is an issue that could be handled with a kind of time stamp filtering though.

We do not expect link-ratio to be an accurate trust measure in isolation. It should either work with other trust measures, or be one component in a solution that utilizes multiple trust computation measures. In section 2, we proposed using PML for building trust layer solution. Our extension to PML for representing trust is intended to be used for encoding aggregated trust values that may have been computed using multiple approaches.

PageRank is a good candidate for an additional trust computation method since it has been useful in similar settings and it is also based on citation structures. [16] calculated the (internal) PageRank on a subset of Wikipedia articles. Specifically, approximately 109K articles from the normal entries of the Wikipedia English database were considered for their experiment. [16] uses the PageRank implementation available in the Java Universal Network/Graph Framework (JUNG) [17] open-source library. They noted that a large number of the highly ranked entries are the names of countries or years. The top 5 articles with their associated PageRank values are presented below:

Article	PageRank value	Link-ratio value
United States	0.003748	0.13
United Kingdom	0.001840	0.19
France	0.001663	0.19
2004	0.001584	0.06
Centuries	0.001264	0.12

The PageRank score may be viewed as a reflection of the relative popularity of an article in a collection of articles, as inferred from the link-structure within that collection. Obviously, there is no strong correlation between the PageRank scores and the link-ratio values, because PageRank is determined by the number of citations and the citation trust of cited articles, while link-ratio is determined by the number of citations and the number of non-citation occurrences. Nevertheless, it is useful to combine two approaches to find more evidence supporting accurate trust evaluation. For example, if both methods are used to calculate high trust values for the same article, we have more evidence that the article is trustworthy. Further, using the inference web approach, we can provide information concerning the trust value and how it was computed.

Wikipedia is different from the Web because Wikipedia articles are restricted to be encyclopedia entries. For example, the article "love" in Wikipedia may be viewed as a description of the definition of love, the scientific models and different point view of

love as opposed to any of the top 10 pages returned from a search for "love" using Google. Those pages are mostly websites about matching and dating services or love poetry resources. Citation-based algorithms may yield different results in a more general web setting. Popular (and potentially trustworthy) general web pages may be viewed as more interesting to link to than dry encyclopedic pages so they will return higher page rank scores and possibly higher link-ratio scores as well. We are continuing investigations into complementary methods and also on defining the conditions under which methods are more effective.

Our analysis is somewhat limited by the computational cost of the calculation of Wikipedia trustworthiness measures currently under investigation. For each article, we need to navigate all other articles for counting citations and non-citation occurrences. However, automated trust computing is essential in improving the trustworthiness of Wikipedia. In practice, incremental calculation of citation trust is desired because articles in Wikipedia are under constant revisions.

The trustworthiness of a Wikipedia article may be measured in different ways, for example, trust as a measure of accuracy of the article. Lih [18] studied the impact of press citation on the quality of a Wikipedia article in terms of number of editors and number of changes. Stvilia et al. [19] conducted a comprehensive qualitative analysis on various aspects of the information quality of Wikipedia article. While qualitative approaches are important, we are more interested in deriving quantitative metrics which can be automatically computed from Wikipedia database.

Link structure analysis on the Web has been extensively studied in the last of several years, e.g. [20] [21]. Social network and p2p network trust are also relevant to our work, e.g. [8] [10] [11] [22] [23]. Social networks usually have explicit trust assertions among the entities, such as user ratings of a movie, or to a transaction. However, Wikipedia lacks such explicit trust assertions. This is one of the reasons we began with the study of citation-based approaches, in which trust is implicit. Nevertheless, a hybrid model of trust propagation and a citation-based approach may be a more effective hybrid solution.

We are also interested in the representation of trust in large-scale and heterogeneous sources. Our markup representation for explanation information was designed to interoperate between applications needing to share answers and justifications. Similarly, our extension to this markup representation was designed to encode trust and to share that trust information between applications. This approach makes it possible to aggregate different trust values as calculated by different trust approaches. McGuinness and Pinheiro da Silva [12] present Inference Web, a framework for storing, exchanging, combining, abstracting, annotating, comparing and rendering proofs and proof fragments provided by reasoners embedded in Semantic Web applications and facilities. We are currently extending our Inference Web toolkit, including the IWTrust component, to include more support for encoding and sharing trust information.

6. CONCLUSION AND FUTURE WORK

Trust is a central issue when dealing with systems and environments that use information coming from multiple, unknown sources. In this paper, we have presented a vision of how one can use trust information to help users view and filter information in collaborative and evolving information repositories such as Wikipedia. Our tools enable users to develop their own opinion concerning how much and under what circumstances, they should trust information. We have extended PML to provide an interoperable and extensible encoding useful for capturing trust information including trust re-

lations between users. We have also designed a citation-based trust metric motivated by some characteristics of Wikipedia. We implemented the approach and presented some experimental results using Wikipedia data indicating that neither the Link-Ratio algorithm nor the PageRank algorithm proved to be effective enough alone for computing trustworthiness of assertions in an aggregated knowledge repository such as Wikipedia. Motivated by this observation, we have begun exploring new directions for computing trust in collaborative environments, using citation based trust as one building block. We intend to leverage the PML trust extension that we have proposed in this paper to work in combination with new trust algorithms.

While we implemented a single trust measure that was purely computational, we plan to continue our work along a number of dimensions. First, we believe that trust measures should include computational components yet we also want to allow stated trust values between entities (among users, between users and other sources, etc.) We are expanding our design to include stated trust values in addition to computed values. We are also expanding our design to include learning trust values by user instruction.

We have also begun investigations into more sophisticated models of trust. We extended PML with a very simple notion of trust and we are currently using a simple single value. We are exploring more complex measures of trust and we are working on formal descriptions so that different applications may use well defined definitions and values for trust and share those encodings among themselves. This would enable trust to be treated as a first-class entity and offer better flexibility in expressing complex trust relationships and multiple attributes that could codify trust.

The citation-based trust measure is intended to work as one component in a solution that utilizes multiple computational trust measures. We are exploring another approach based on the hypothesis that revision history may be a useful component in a hybrid approach for computing a measure of trustworthiness of articles. For example, one may assume that an article may become more trustworthy if it revised by a trustworthy author, and similarly, it may become less trustworthy if revised by an author who is known to be less trustworthy. Given the rich and accessible revision information in Wikipedia[6], we are working on a hybrid model that utilizes both citation-based trust and revision history-based trust. Preliminary experiments indicate that this hybrid approach using these two metrics performs far better than when a single model is used.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] : Wikipedia. (http://www.wikipedia.com)

[2] Giles, J.: Internet encyclopaedias go head to head. In: Nature 438, 900-901 (15 Dec 2005) News. (2005)

[3] : John seigenthaler sr. wikipedia biography controversy. (http://en.wikipedia.org/wiki/John_Seigenthaler_Sr._Wikipedia_biography_controversy)

[4] Castelfranchi, C., Tan, Y., eds.: Trust and Deception in Virtual Societies. Kluwer Academic Publishers (2001)

---

[6]Wikipedia authors have made approximately 41 million revisions, an average of 12 versions per article, over the last four years.

[5] Grandison, T., Sloman, M.: A survey of trust in internet application. IEEE Communications Surveys Tutorials (Fourth Quarter) **3**(4) (2000)

[6] Maurer, U.: Modelling a public-key infrastructure. In: ESORICS: European Symposium on Research in Computer Security, LNCS, Springer-Verlag (1996)

[7] Blaze, M., Feigenbaum, J., Lacy, J.: Decentralized trust management. In: Proceedings of the 1996 IEEE Symposium on Security and Privacy. (1996) 164–173

[8] Damiani, E., di Vimercati, S., Paraboschi, S., Samarati, P., Violante, F.: A reputation-based approach for choosing reliable resources in peer-to-peer networks. (2002) In 9th ACM Conf. on Computer and Communications Security.

[9] Mui, L.: Computational Models of Trust and Reputation: Agents, Evolutionary Games, and Social Networks. PhD thesis, MIT (2002)

[10] Kamvar, S.D., Schlosser, M.T., Garcia-Molina, H.: The eigentrust algorithm for reputation management in p2p networks. In: Proceedings of the 12th international conference on World Wide Web. (2003)

[11] Guha, R., Kumar, R., Raghavan, P., Tomkins, A.: Propagation of trust and distrust. In: Proceedings of the 13th international conference on World Wide Web, ACM Press (2004) 403–412

[12] McGuinness, D.L., Pinheiro da Silva, P.: Explaining answers from the semantic web: The inference web approach. In: Journal of Web Semantics. Volume 1. (2004) 397–413

[13] Pinheiro da Silva, P., McGuinness, D.L., Fikes, R.: A proof markup language for semantic web services. (In: Information Systems. (To appear))

[14] Cock, M.D., Pinheiro da Silva, P.: A many valued representation and propagation of trust and distrust. In: In Proceedings of International Workshop on Fuzzy Logic and Applications (WILF2005). (2005)

[15] Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University (1998)

[16] : Pagerank report on the wikipedia. (http://www.searchmorph.com/wp/2005/01/26/pagerank-report-on-the-wikipedia)

[17] : Java universal network/graph framework (jung). (http://jung.sourceforge.net/)

[18] Lih, A.: Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. In: Proceedings of the 5th International Symposium on Online Journalism. (2004)

[19] Stvilia, B., Twidale, M.B., Gasser, L., Smith, L.C.: Information quality discussion in wikipedia. In: Proceedings of the 2005 International Conference on Knowledge Management. (2005) 101–113

[20] Haveliwala, T.H.: Topic-sensitive pagerank. In: Proceedings of the Eleventh International World Wide Web Conference. (2002)

[21] Tomlin, J.A.: A new paradigm for ranking pages on the world wide web. In: Proceedings of the Twelveth International World Wide Web Conference. (2003)

[22] Xiong, L., Liu, L.: A reputation-based trust model for peer-to-peer ecommerce communities. (2003) Proceedings of the 4th ACM conference on Electronic commerce.

[23] Wang, Y., Vassileva, J.: Trust and reputation model in peer-to-peer networks. In: P2P'03. (2003)