

# A Mining Algorithm for Authentication of Data

Mr.Ashwin kumar H.N<sup>1</sup>, Dr. Manjunath Rao.L<sup>2</sup>  
Research Scholar, CMJ University, Meghalaya, India

## Abstract

TRUTH FINDER is an approach that utilizes the interdependency between web site trustworthiness and fact confidence to find trustable web sites and true facts. The input of Truth Finder is a large number of facts about properties of a certain type of objects. The facts are provided by many web sites. The aim of Truth Finder is to identify the true fact among them. The world-wide web has become the most important information source for most of us. Unfortunately, there is no guarantee for the correctness of information on the web. Moreover, different web sites often provide conflicting information on a subject, such as different specifications for the same product. In this application we propose a solution to a new problem called Veracity, i.e., "conformity to truth", which studies how to find true facts from a large amount of conflicting information on many subjects that is provided by various web sites. We design a general framework for the Veracity problem, and invent an algorithm called Truth Finder, which utilizes the relationships between web sites and their information, i.e., a web site is trustworthy if it provides many pieces of true information, and a piece of information is likely to be true if it is provided by many trustworthy web sites.

IJSER

## Introduction:

The world-wide web has become a necessary part of our lives, and might have become the most important information source for most people. Everyday people retrieve all kinds of information from the web. For example, when shopping online, people find product specifications from web sites like Amazon.com or ShopZilla.com. When looking for interesting DVDs, they get information and read movie reviews on web sites such as NetFlix.com or IMDB.com. "Is the world-wide web always trustable?" Unfortunately, the answer is "no".

There is no guarantee for the correctness of information on the web. Even worse, different web sites often provide conflicting information, we use the word "fact" to represent something that is claimed as a fact by some web site, and such a fact can be either true or false. There are often conflicting facts on the web, such as different sets of authors for a book. There are also many web sites, some of which are more trustworthy. Because of the interdependency between facts and websites, we choose an iterative computational method. At each iteration, the probabilities of facts being true and the trust-worthiness of web sites are inferred from each other.

The trustworthiness of a web site does not depend on how many facts it provides, but on the accuracy of those facts. Nor can we compute the probability of a fact being true by adding up the trustworthiness of web sites providing it. These lead to non-linearity in computation. Second and more importantly, different facts influence each other. For example, if a web site says a book is written by "Jessamyn Wendell", and another says "Jessamyn BurnsWendell", then these two web sites actually support each other although they provide slightly different facts. For this conflict, first, we formulate the Veracity problem about

how to discover true facts from conflicting information. Second, we propose a framework to solve this problem, by defining the trustworthiness of web sites, confidence of facts, and influences between facts. Finally, we propose an algorithm called Truth Finder for identifying true facts using iterative methods.

## Existing System:

User visits different web sites, which often provide conflicting information on various subjects on which we browse, such as different specifications for the same product. There is no guarantee for the correctness of information on the web. Which leads to confusions among the users? Thus leads to a new problem called Veracity, i.e., conformity to truth, is not guaranteed. Further, the drawbacks are 1) The popularity of web pages does not necessarily lead to accuracy of information. 2) Even the most popular website may contain many errors. 3) Whereas some comparatively not-so-popular websites may provide more accurate information. Truth Validation and Veracity Analysis with Information Networks. Jiawei Han Data Mining Group, Computer Science . University of Illinois at Urbana-Champaign. Outlet: Truth Finder: Truth Validation by Information Network Analysis. Beyond Truth Finder: Multiple Versions of Truth and Evolution of Truth. Enhancing Truth Validation by Info Net Analysis: The RankClus & NetClus Methodology

## Problem Definition

In this paper, we propose a new problem called the Veracity problem, which is formulated as follows: Given a large amount of conflicting information about many objects, which is provided by multiple websites, how can we discover the true fact about each object? We use the word "fact" to represent something that is claimed as a fact

by some website and such a fact can be either true or false.

Online Store	Authors
Powell's books	Holtzblatt, Karen
Barnes & Noble	Karen Holtzblatt, Jessamyn Wendell, Shelley Wood
A1 Books	Karen Holtzblatt, Jessamyn Burns Wendell, Shelley Wood
Cornwall books	Holtzblatt-Karen, Wendell-Jessamyn Burns, Wood
Mellon's books	Wendell, Jessamyn
Lakeside books	Wendell, Jessamynholtzblatt, Karenwood, Shelley
Blackwell online	Wendell, Jessamyn, Holtzblatt, Karen, Wood, Shelley

In this paper, we only study the facts that are either properties of objects or relationships between two objects we also require that the facts can be parsed from the web pages. There are often conflicting facts on the Web, such as different sets of authors for a book. There are also many websites, some of which are more trustworthy than others. A fact is likely to be true if it is provided by trustworthy websites. A website is trustworthy if most facts it provides are true. At each iteration, the probabilities of facts being

true and the trustworthiness of websites are inferred from each other.

### Proposed System

TRUTHFINDER is an approach that utilizes the interdependency between web site trustworthiness and fact confidence to find trustable web sites and true facts. TRUTHFINDER achieves high accuracy at finding true facts and at the same time identifies websites that provide more accurate information. There is a guarantee for the correctness of information on the web which based on the users browsing count. Conformity to truth is guaranteed. TRUTHFINDER avoids confusion among users in search of truth. Thus our experiments show that TRUTHFINDER achieves very high accuracy in discovering true facts.

### Motivation:

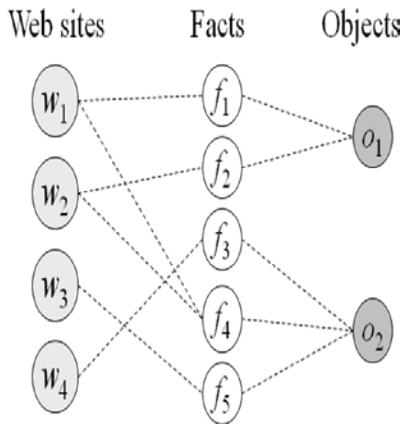
Why truth validation and veracity analysis? Information sharing; Sharing trustable, quality information; Information security: Protecting trustable information and its sources. Identifying which information providers are suspicious ones: frequently providing false information. The trustworthiness problem of the web (according to a survey): 54% of Internet users trust news web sites most of time. 26% for web sites that sell products. 12% for blogs. Truth Finder: Truth discovery on the Web by link analysis. Among multiple conflict results, can we automatically identify which one is likely the true fact? Veracity (conformity to truth): Given a large amount of conflicting information about many objects, provided by multiple web sites (or other information providers), how to discover the true fact about each object?

### Conflicting Information on the Web

Different websites often provide conflicting info. On a subject, e.g., Authors of "Rapid Contextual Design"

Table of conflicting information

Mapping It to Information Networks Each object may have a set of



conflicting facts  
 E.g., different author names for a book And each web site provides some facts How to find the true fact for each object?

Figure Inter dependency between websites and facts

Basic Heuristics for Problem Solving: There is usually only one true fact for a property of an object. This true fact appears to be the same or similar on different web sites. E.g., "Jennifer Widom" vs. "J. Widom". The false facts on different web sites are less likely to be the same or similar. False facts are often introduced by random factors. A web site that provides mostly true facts for many objects will likely provide true facts for other objects

Mutual Consolidation between Confidence of Facts and Trustworthiness of Providers: Confidence of facts ↔ Trustworthiness of web sites; A fact has high confidence if it is provided by (many) trustworthy web sites; A web site is trustworthy if it provides many facts with high confidence. The Truth Finder mechanism, an overview: Initially, each web site is equally trustworthy. Based on the above four heuristics, infer fact confidence from web site

trustworthiness, and then backwards. Repeat until achieving stable state Inference on Trust worthness

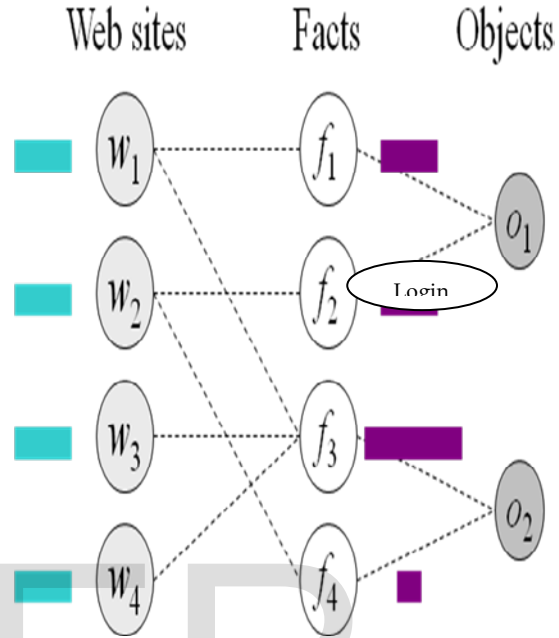


Figure Inference of web site trustworthiness & fact confidence

True facts and trustable web sites will become apparent after some iteration

Implementation:

In computer science, implementation is the practical application of a methodology or algorithm to fulfill a desired purpose. For example, one might create a computer program that sorts a list of numbers in ascending order. To do so, one would implement a known method of sorting. In software architectural design, the top-level structure and organization of the system is described and various components are identified (how the system is decomposed and organized into components). It must describe the interfaces between these components. Software implementation design describes each component of the system.

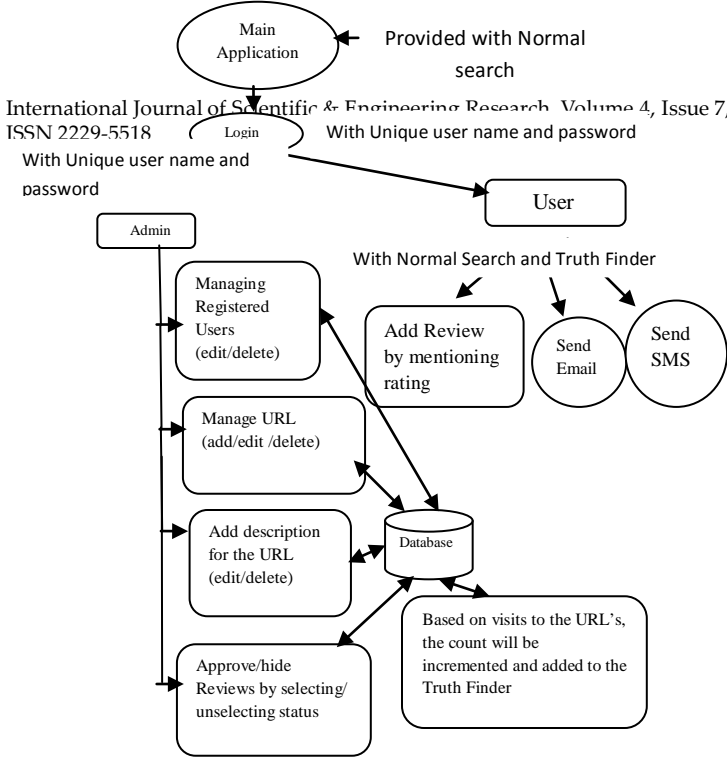


Figure Data Flow






**Algorithm Stage**

- Step 1: Login page:
- Step 2: Admin Login page:
- Step 3: Admin login control
- Step 4: Arguments
- Step 5: admin login page
- Step 6: AdminManageRegisteredUsers.aspx
- Step 7: Approve status:
- Step 8: User registration page:
- Step 9: User Manage Website Content Page:

**Results:**

Outcome	Description
	In order to search the contents using truth finder search the user has to sign in

	using username and password.
	Here the new user registration will taken place and it will be stored in the user table.
	Snapshot of normal search; Result screen for the normal search method.
	Snapshot of truth finder search; Truth finder search displays the contents along with different view counts.
	Snapshot of registered users; The registered user's details will be maintained here by the admin. He can perform the edit/delete

	operation.		reviews to the users by setting the status property to true.
	Snapshot of URL's; The URL's and the visitor counts are maintained here by the admin. New URL's can be added here.		Snapshot of sending SMS; Searched contents can be sending through sms to other users.
	Snapshot of URL contents; The contents regarding to particular keyword along with the websites and hits are present here. The admin will add the new contents.		Snapshot of sending email; Searched contents can also be sending through email.
	Snapshot of reviews; The reviews about a particular website given by the user are stored here. The admin can display this	<p><b>Conclusion:</b></p> <p>We cannot compute the trustworthiness of a website by adding up the weights of its facts as in, nor can we compute the probability of a fact being true by adding up the trustworthiness of websites providing it instead, we have to resort to probabilistic computation. Second and more importantly, different facts influence each other. For example, if a website says that a book is written by "Jessamyn Wendell" and another says "Jessamyn Burns Wendell," then these two websites actually support each other although they provide slightly different facts. We</p>	

incorporate such influences between facts into our computational model. In summary, we make three major distributions in this paper. First, we formulate the Veracity problem about how to discover true facts from conflicting information. Second, we propose a framework to solve this. TRUTHFINDER achieves high correctness at finding true facts and at the same time identifies websites that provide more accurate information provided by many trustworthy web sites. TRUTHFINDER achieves high accuracy at finding true facts and at the same time identifies websites that provide more accurate information. Conformity to truth is guaranteed. The project benefits the user to make efficient and effective search, but the quality of information on the web has always been a major concern for internet users. Hence there can be studies on what factors of data quality are important for users and machine learning approaches for distinguishing high-quality and low-quality web pages, where the quality is defined by human preferences. It can also show that information quality measures can help improve the effectiveness of web search. Further, The work can reduce the uncertainties in the discovered patterns that are used to solve the problem. This can be done by discovering only the useful patterns. Clustering based on techniques can be used to determine the useful patterns.

#### References

- [1] R.B. Bradford, "Technical factors in the creation of large full text databases," Department of Energy InfoTech Conference, Oak Ridge, Tennessee, May 1991
- [2] S. Tsujimoto, H. Asada, "Major components of a complete text reading system", Proc. Of IEEE, July, 1992, Vol. 80, No. 7, pp. 1133-1149.

[3] M. Bokser, "Omnidocument technologies," Proceedings of the IEEE, 80(7):1066-1078, 1992

[4] G. Nagy, S. Seth, and K. Einspahr, "Decoding substitution ciphers by means of word matching with application to OCR," IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9(5):710-715, 1987

[5] Baird, H.S., and K. Thompson, "Reading Chess", IEEE Trans. PAMI, 12(6):552-559, June 1990

IJSER