VERA: A Platform for Veracity Estimation over Web Data

Mouhamadou Lamine Ba, Laure Berti-Equille, Hossam M. Hammady
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
{mlba,lberti,hhammady}@qf.org.qa

ABSTRACT

Social networks and the Web in general are characterized by multiple information sources often claiming conflicting data values. Data veracity is hard to estimate, especially when there is no prior knowledge about the sources or the claims and in time-dependent scenarios where initially very few observers can report first information. Despite the wide set of recently proposed truth discovery approaches, "no-one-fits-all" solution emerges for estimating the veracity of on-line information in open contexts. However, analyzing the space of information conflicts and sources' agreements might be relevant, as well as ensembling multiple truth discovery methods. This demonstration presents VERA, a Web-based platform that supports information extraction from Web textual data and micro-texts from Twitter and estimates data veracity. Given a user query, VERA systematically extracts entities and relations from Web content, structures them as claims relevant to the query, expands the query for gathering more evidences and conflicting/corroborating information. Finally, VERA combines multiple truth discovery algorithms through active learning and returns the veracity label and score of each data value and the trustworthiness scores of the sources. VERA will be demonstrated through several real-world scenarios to show its potential value for on-line time-dependent fact-checking.

1. INTRODUCTION

With the recent development of computational journalism [3, 7], on-line fact-checkers such as FactCheck[1], Snopes[2], PolitiFact[3], TruthorFiction[4] or OpenSecrets[5], and ClaimBuster[6] have lately gained unprecedented attention as their goal is to verify on-line information for public opinion and automate Web-scale fact-checking. But estimating the veracity of data still remains a challenging problem: extracting structured information from large, heterogeneous corpora of textual and multimedia documents and integrating these multi-source data are difficult tasks. Web data and micro-texts from social media can be noisy, outdated, incorrect, conflicting, and thus unreliable, often due to information extraction errors, disagreements, biased observations, disparate or low quality of the sources.

[1] www.factcheck.org
[2] www.snopes.com
[3] www.politifact.com
[4] www.truthorfiction.com
[5] www.opensecrets.org
[6] idir-server2.uta.edu/claimbuster

Many truth discovery methods have been proposed to deal with data veracity estimation – see [1] for a survey. They are mostly applied to structured data and compute iteratively the accuracy of the sources claiming some data as a function of the veracity scores of their data and the veracity scores are computed as a function of the accuracy of their sources. Recent approaches have been developed to discover true values extracted from textual content in a large corpus of Web sources using various information extractors [5, 15]. These solutions extend previous probabilistic models based on iterative vote counting and integrate the extraction systems' error in truth discovery computation.

Nevertheless, most approaches operate on a static set of structured claims from a fixed corpus of information sources. They usually do not expand dynamically the search space to gather additional evidences and controversial or corroborating claims. Moreover, several studies have proven that a "one-fits-all" solution does not seem to be achievable for a wide range of truth discovery scenarios [11] and we argue that ensembling truth discovery methods can significantly improve the quality performance of current results.

In this demo, we present VERA, a Web-based platform that supports the full pipeline of truth discovery from Web unstructured corpus and tweets: ranging from information extraction from raw texts and micro-texts and data fusion to truth discovery and visualization. VERA offers several advantages over previous work as it includes:

- Extraction and fusion of multi-source information to answer a factual query defined by the user;
- Dynamic query expansion to enlarge the set of sources and claims and collect additional, timely evidences;
- Combination of multiple truth discovery algorithms using active learning-based ensembling in order to effectively discover true values from conflicting ones;
- Explanation of the truth discovery results;
- Visualization artifacts to better understand the information space with disagreeing vs. agreeing sources and corroborating vs. conflicting claims.

To the best of our knowledge, this work is the first attempt to demonstrate truth discovery in action from Web information and Twitter social network, overcoming limitations of single truth discovery methods with active ensembling to estimate data veracity.

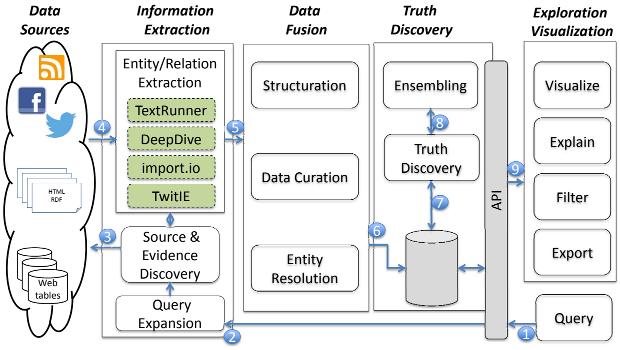


Figure 1: VERA Platform Architecture

VERA platform, RESTful API and additional material including data sets are available at: `dafna.qcri.org`.

2. VERA PLATFORM

VERA platform architecture and its workflow (from 1 to 9) are depicted in Figure 1. Our system is composed of four layers: (1) Information Extraction layer is in charge of querying the corpus of Web documents and tweets, discovering new resources and expanding the user query (e.g., the query "number of people killed in November 2015 Paris attacks" is expanded by related keywords queries such as "how many dead in Paris terrorist attacks", "Paris victims", and "Paris shootings casualties"); this layer applies various text processing techniques to extract relevant information items; (2) Data Fusion layer is in charge of transforming extracted information into structured claims and applies various data formatting and curation techniques including deduplication and entity resolution; it groups together the structured claims referring to the same real-world event or entity property; (3) Truth Discovery layer is responsible for executing and ensembling various truth discovery methods; it determines which claims are true or false by computing their veracity scores and estimating the trustworthiness scores of the sources. This layer can integrate users' a priori knowledge (when available), e.g., about the reliability of certain sources or the hardness of certain claims; (4) Result Exploration and Visualization layer provides three types of result: (i) veracity score and label for each claim and trustworthiness score for each source; (ii) explanations of the truth discovery results; (iii) and various visualization artifacts for exploring source polarity and claim controversy.

2.1 Information Extraction and Data Fusion

Information Extraction. Information extraction in VERA is modular and various information extractors can be added depending on the initial set of resources and the application domain of the query. Currently, VERA uses four extractors: TextRunner [13], TweetIE [2], DeepDive [16], and `import.io`[7]. TextRunner is used as an open information extraction system from a predefined set of Web corpus including *Google*, *ClueWeb*, *News*, *Nell*, and *Wikipedia* corpus. Each particular corpus aggregates data from various Web sources and domain-specific Websites.

When a user query is submitted to VERA, it is reformulated and submitted to user-selected extractors to query the sources of the corpora. Then, each extractor extracts a set of candidate answers using natural language processing techniques and other external resources: for example, in the case of TextRunner, Freebase ontology[8] is leveraged and the user query is transformed into a triplet (e_1, r, e_2) and sent to TextRunner. The argument r specifies a possible relationship between the two entities e_1 and e_2. Partial knowledge about the real-world can be captured when one argument is unknown and TextRunner finds out every possible candidate values. VERA then transforms the set of candidate answers and completes the output of TextRunner with its respective identified sources.

To expand the Web corpus, one can specify DeepDive extractors or use external APIs such as `import.io`. In the case of DeepDive, each predefined extractor takes as input a collection of textual documents and a set of labeled examples of relations (using DBpedia knowledge base, for example). DeepDive extractor instance first extracts entities and candidates relationships by leveraging the outputs of natural language processing over the training examples. Then, it uses statistical learning with user-defined inference rules and training examples of the relations to extract.

As for TwitIE, it is applied, when chosen, to the set of tweets collected on-demand regarding a given user query. It is used for natural language and micro-text processing, named entity recognition, and relation extraction from tweets. Additional scripts have been developed for filtering non-textual contents and tweets that are irrelevant to the user query.

Data Fusion. Once information is extracted by the extractor instances of TextRunner, DeepDive, `import.io`, and TwitIE, it is transformed into a structured claim such as: (claimID, sourceID, Object:Property, claimedValue, timestamp, evidence). Each claim has an identifier, a source identifier, a value for a particular property of the queried object, a timestamp and the content from which the claim has been extracted as a textual evidence. Entity resolution is achieved to group the claims referring to the same real-world entity and property into the same cluster. The clusters of quadruplets constitute the input data of the Truth Discovery layer. An illustrative example of cluster for the property `NumberOfVictims` for the object `November_2015_France_Terrorist_Attack` is given in Table 1.

Storage. Queries, claims, labels and scores are stored in a PostgreSQL database instance through Amazon S3 for result exploration tasks or truth discovery processes.

2.2 Truth Discovery

VERA supports an adaptive truth discovery approach based on ensembling and active learning to compute the optimal labeling and scoring results from a set of truth discovery algorithms. Our approach outperforms individual truth discovery technique on any given data set. It actively leverages the user's knowledge when available for finding the true claims and update the trustworthiness score of the sources. When user's knowledge or training data are not available, VERA still provides meaningful results using ensembling of methods with minimizing the disagreement between methods.

Competing Classifiers. In our context, the truth discovery algorithms are considered as binary classifiers whose goal is to label each conflicting value as a true or false answer to the user query.

VERA integrates and combines twelve state-of-the art truth discovery algorithms that can be classified in three categories as follows: (1) Agreement-based methods including TruthFinder [14], Cosine, 2-Estimates and 3-Estimates [6], AccuNoDep [4]; (2) MAP Estimation-based methods including MLE [12], LTM [17], SimpleLCA and GuessLCA [9]; and (3) Bayesian Inference-based methods including Depen, Accu, and AccuSim [4].

[7] `https://import.io`

[8] `https://www.freebase.com/`

Claim	Source	Extracted Value	Textual Evidence	Timestamp
C_5	cnn.com	At least 128	Paris massacre: At least 128 killed in gunfire and blasts, French officials say	Nov 27, 2015
C_4	theguardian	120	Paris attacks kill more than 120 people – as it happened	Nov 26, 2015
C_3	news.sky.com	130	Number Of Paris Attacks Victims Rises To 130	Nov 20, 2015
C_2	bbc.com	130	Tributes have been paid to the 130 people who lost their lives in the Paris terror attacks.	Nov 16, 2015
C_1	@TBurgesWatson	35	BREAKING.This is what we know: 35 dead, 100 hostages taken at a concert venue. Various drive by shootings. Explosions at a #Paris stadium.	Nov 13, 2015

Table 1: Example of conflicting answers for the query "How many victims in Paris Attacks"

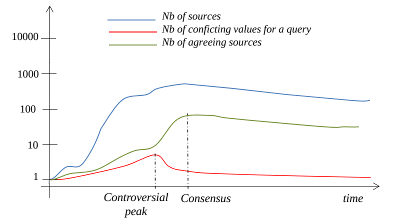


Figure 2: Illustration of Time-dependent Consensus

Ensembling. Ensembling is a semi-supervised learning approach combining various competing models that has been demonstrated to be very effective in many disciplines [10]. VERA ensembling method discovers the optimal set (ensemble) of classifiers for the truth discovery classification problem by actively learning from an oracle, e.g., the user or a reference model over a sample of data. Ensembling enables to perform classification consistently well across various data sets without having to determine *a priori* a suitable classifier type. VERA exploits ensembling for combining truth discovery methods in the two following cases: (1) When the user can provide either *a priori* or *a posteriori* truth labels for few claims as instances of the ground truth (under a limited budget and with expected guarantees); (2) When no ground truth training data are available.

In the first case, VERA actively learns from the user's labels over the training data, finds the best ensemble of classifiers and returns the veracity labels and scores for the rest of the data.

In the second case, VERA selects the candidate ensembling which satisfies our time-dependent consensus model. This model captures three intuitive ideas: (i) Initially, very few sources with diverse authoritativeness degrees may observe an event and report information (e.g., in case of disaster or bombing). As long as the information is not confirmed (or denied) by a sufficient number of other independent sources, unknown or non-reputable sources should not be penalized and the authoritativeness of the sources should not influence the veracity estimation; (ii) The number of conflicting information claimed by multiple sources has a decreasing variability in time; and (iii) The majority of sources cannot be trusted until a certain time-point where a consensus on the fact (i.e., the true value) is reached as illustrated in Figure 2.

Truth discovery is hard in practical scenarios because there is often no prior ground truth guiding the selection of an algorithm, in particular when the context is dynamic. More importantly, a large set of labeled data (or training data) is generally out-of-reach in the context of Web and social media data. As a consequence, it remains usually hard to evaluate the precision/recall of existing algorithms on real-world data.

However, users may have background knowledge about some real-world facts or about the reliability of particular sources. Such knowledge can be very valuable not only *a priori* for guiding VERA truth discovery computation, but also *a posteriori* for evaluating and backtracking the errors.

In our demonstration scenarios, we will show the two operational modes of VERA which combines truth discovery and open information extraction with ensemble-based active learning for estimating data veracity with and without prior knowledge of the user, also learning from posterior knowledge in order to improve the next truth discovery computations.

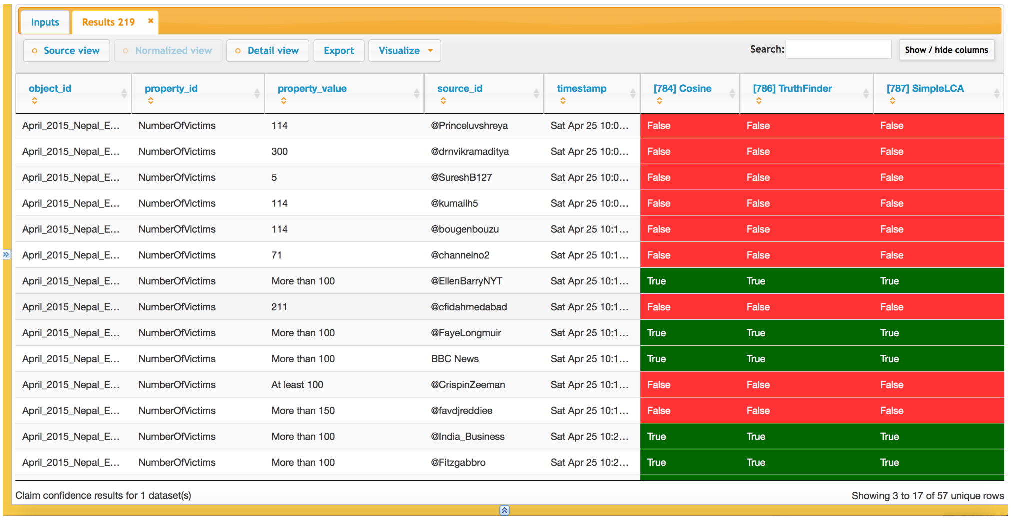
2.3 Result Exploration

VERA features a layer for result visualization and explanation basically consisting of a set of Web user interfaces (panels and option widgets) and decision algorithms which respectively enable viewing and browsing the truth discovery results to obtain more deeper insights and understand how the estimation of the veracity of each claim has been computed by the system. This layer also provides the ability to input truth labels for a limited subset of claims. Explanation is accomplished in VERA through APIs whereas result visualization renders the output of the truth discovery process to ease user exploration and interaction with the system, as we will detail hereafter.

Visualization. VERA currently supports three artifacts to visualize and browse the results of query answering. When a user query is submitted to VERA, candidate claims are extracted and truth discovery computation is performed. A list-based artifact then presents the candidate answers to the user. For each object property related to the user query, the candidate answers are ranked in a decreasing order of veracity scores, i.e., the answer with the highest veracity score (the most likely to be true) is listed first, then the answer with the second highest veracity score is given, and so on. Each line of the result list contains a claimed answer, its veracity score and label (`True` or `False`) returned by the Truth Discovery layer, with the option widget to view the set of sources supporting it (`view_sources`), and an illustrative excerpt from the content of the Web document or tweet from which the answer has been extracted. Candidate claims are presented to the user with an additional widget option `user_input_label` that enables the user to eventually propose a labelling or add background knowledge. For further result exploration, VERA has set two other visualization artifacts. Indeed, when the user chooses the option widget `view_sources`, VERA presents the complete list of sources which support the corresponding claim, their trustworthiness scores computed by the Truth Discovery layer, and their associated corpus. Finally, the user can access to the explanation window to better understand the estimation of the veracity score of a particular answer by clicking on it.

Explanation. VERA relies on DAFNA API [**?**] to generate explanations for the results of the truth discovery process. It is accessible via its endpoint **runs**. Given a claim identifier, the system can explain how the returned score of veracity has been computed using the ensemble of truth discovery algorithms.

To this end, DAFNA builds an explanation decision tree representing the different choices made by the ensemble applied to the candidate answers. Every decision tree is built from the number of supporting vs. disagreeing sources, their trustworthy levels, and the set of conflicting values. VERA leverages the explanation functionality of DAFNA to provide insights of the results when requested by the user.



object_id	property_id	property_value	source_id	timestamp	[784] Cosine	[786] TruthFinder	[787] SimpleLCA
April_2015_Nepal_E...	NumberOfVictims	114	@Princeluvshreya	Sat Apr 25 10:0...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	300	@drnvikramaditya	Sat Apr 25 10:0...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	5	@SureshB127	Sat Apr 25 10:0...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	114	@kumailh5	Sat Apr 25 10:0...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	114	@bougenbouzu	Sat Apr 25 10:1...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	71	@channelno2	Sat Apr 25 10:1...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	More than 100	@EllenBarryNYT	Sat Apr 25 10:1...	True	True	True
April_2015_Nepal_E...	NumberOfVictims	211	@cfidahmedabad	Sat Apr 25 10:1...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	More than 100	@FayeLongmuir	Sat Apr 25 10:1...	True	True	True
April_2015_Nepal_E...	NumberOfVictims	More than 100	BBC News	Sat Apr 25 10:1...	True	True	True
April_2015_Nepal_E...	NumberOfVictims	At least 100	@CrispinZeeman	Sat Apr 25 10:1...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	More than 150	@favdjreddiee	Sat Apr 25 10:1...	False	False	False
April_2015_Nepal_E...	NumberOfVictims	More than 100	@India_Business	Sat Apr 25 10:2...	True	True	True
April_2015_Nepal_E...	NumberOfVictims	More than 100	@Fitzgabbro	Sat Apr 25 10:2...	True	True	True

Figure 3: VERA: Behind the Scene

3. DEMONSTRATION SCENARIOS

During the demo we will show how VERA estimates the veracity of multi-source information from Web sources and tweets. The audience will see two main truth discovery scenarios that can not be accomplished using conventional search engines or existing truth discovery methods and we will show how they can be handled using VERA.

Real-Time Fact-Checking for Crisis Situations. In crisis situations, time is critical when an emergency response must be issued as soon as possible. Often the only information the public receives about the situation or the disaster is through the media (usually by authoritative sources) only once it is verified and but also immediately through social media as volunteered information that still needs to be checked. In this demonstration scenario, VERA is combined to AIDR [8] to estimate the veracity of claims extracted from the content of tweets. Figure 3 presents VERA behind the scene and shows the results of multiple truth discovery methods. These methods have been applied to the claims extracted by TwitIE and structured by VERA from a collection of tweets related to Nepal earth quake on April 25, 2015 in the time window 10:06AM–10:23AM as presented in the figure; the tweets were classified through AIDR in the category "injured or dead people". As the time goes by, more information corroborate the true number of casualties. VERA source view also provides the trustworthiness scoring and ranking of the sources (twitter ids and Web sources) evolving in time.

Rumors. Nowadays, rumors about facts related to persons (e.g., celebrities) or hot events are ubiquitous on the Web. Some rumors are purposely propagated for misinformation or propaganda, (e.g., Barack Obama is born in Kenya) and others are tied to a certain context which requires to have more information as soon as possible to confirm or deny them (e.g., the rumor of the bombing of "Les Halles shopping center" during Paris attacks in November 2015). Such kinds of rumors often spread out very quickly in social media due to the lack of effective means to detect them. This scenario will show how VERA operates on rumors by leveraging the sources' trustworthiness and time-dependent consensus in truth discovery computation.

4. REFERENCES

[1] L. Berti-Equille and J. Borge-Holthoefer. *Veracity of Big Data*. Morgan & Claypool, 2015.
[2] K. Bontcheva, L. Derczynski, A. Funk, M. A. Greenwood, D. Maynard, and N. Aswani. TwitIE: An open-source information extraction pipeline for microblog text. In *Proc. of the Intl. Conf. on Recent Advances in Natural Language Processing*, 2013.
[3] S. Cohen, J. T. Hamilton, and F. Turner. Computational journalism. *CACM*, 54(10):66–71, 2011.
[4] X. L. Dong, L. Berti-Equille, and D. Srivastava. Integrating conflicting data: The role of source dependence. *PVLDB*, 2(1):550–561, 2009.
[5] X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *KDD'14*, pages 601–610, 2014.
[6] A. Galland, S. Abiteboul, A. Marian, and P. Senellart. Corroborating Information from Disagreeing Views. In *WSDM'10*, pages 131–140, 2010.
[7] N. Hassan, C. Li, and M. Tremayne. Detecting check-worthy factual claims in presidential debates. In *CIKM'15*, pages 1835–1838, 2015.
[8] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg. AIDR: Artificial intelligence for disaster response. In *WWW'14*, 2014.
[9] J. Pasternack and D. Roth. Latent Credibility Analysis. In *WWW'13*, pages 1009–1020, 2013.
[10] B. Settles. *Active Learning*. Morgan & Claypool, 2012.
[11] D. A. Waguih and L. Berti-Equille. Truth discovery algorithms: An experimental evaluation. *CoRR*, 1409.6428, 2014.
[12] D. Wang, L. M. Kaplan, H. K. Le, and T. F. Abdelzaher. On Truth Discovery in Social Sensing: a Maximum Likelihood Estimation Approach. In *IPSN*, pages 233–244, 2012.
[13] A. Yates, M. Cafarella, M. Banko, O. Etzioni, M. Broadhead, and S. Soderland. TextRunner: Open Information Extraction on the Web. In *NAACL*, pages 25–26, 2007.
[14] X. Yin, J. Han, and P. S. Yu. Truth Discovery with Multiple Conflicting Information Providers on the Web. *TKDE*, 20(6):796–808, 2008.
[15] D. Yu, H. Huang, T. Cassidy, H. Ji, C. Wang, S. Zhi, J. Han, C. R. Voss, and M. Magdon-Ismail. The wisdom of minority: Unsupervised slot filling validation based on multi-dimensional truth-finding. In *COLING 2014*, pages 1567–1578, 2014.
[16] C. Zhang. *DeepDive: A Data Management System for Automatic Knowledge Base Construction*. PhD thesis, University of Wisconsin-Madison, USA, 2015.
[17] B. Zhao, B. I. P. Rubinstein, J. Gemmell, and J. Han. A Bayesian Approach to Discovering Truth from Conflicting Sources for Data Integration. *PVLDB*, 5(6):550–561, 2012.