

Truthfulness Verification System

Tathagata Dasgupta, ABM Musa
Email: {tdasgu2, amusa2}@uic.edu
URL: <http://code.google.com/p/tverifier>

1 Progress Summary

Mar 9 to Mar 29

- Formulated the initial steps for feature extraction for each doubt unit using Wikipedia. See page [10] of section 5.3.2

Feb 23 to Mar 9

- Formulated the procedure of truthfulness verification in a more structured way
- As a result of the procedure formulation, the sections Step-by-step Algorithm for Truthfulness and Verification Examples of Truthfulness Verification are written in a complete revised way compared to the proposal
- Apart from getting out feet wet with the MQL api, the daunting task is to develop generalized framework that is flexible enough to fit the wide variety of sentences into the rigid structure of Freebase. The learning curve is steep, given the richness and massive size of documentation. Though we are groping in dark currently, we are hopeful that the pieces of the puzzle will soon fit in.
- Looked into applicability of POS tagger in the context of our problem and found it useful for constructing freebase queries
- Got some initial promising results by exploring different options such as Wikipedia, answer.com beyond Freebase ontologies.
- Worked on porting of existing T-verifier code to linux. The original T-verifier code is for windows only and not generic enough at this moment. We also contacted with the student author about the problems we are facing to port the code to linux. We hope to resolve it soon.

2 Introduction

Web has become the most prevalent source of information now a days. However, many information on the web is untruthful. Also because of the widespread reach of web, sometimes web is used

to propagate untruthful facts for social and political reasons. Hence with the increasingly use of web as information source, verification of information truthfulness became an important facet to consider.

The popular search engines extract information from web based on keywords and metadata without considering the truthfulness of the facts. Also, there have been not much research in this area. To our best knowledge, the most recent work on this area is T-verifier [5], which uses results from search engines to verify truthfulness of statements. T-verifier performs very well on the test-dataset. However, T-verifier has some problems with the overall approach to verify the truthfulness of statements and there is room for improvement. In this work, we will extend the T-verifier system so that it's weaknesses can be resolved to get a more robust truthfulness verification system.

3 Current System

Current system for truthfulness verification is called T-verifier [5], which uses two phase methods for truthfulness verification of statements. Each of these two phases rely heavily on search results returned by popular search engines. T-verifier takes the doubtful statements (DS) as input from the user along with the doubtful unit (DU). Phrase after removing DU from DS is called topic unit (TU).

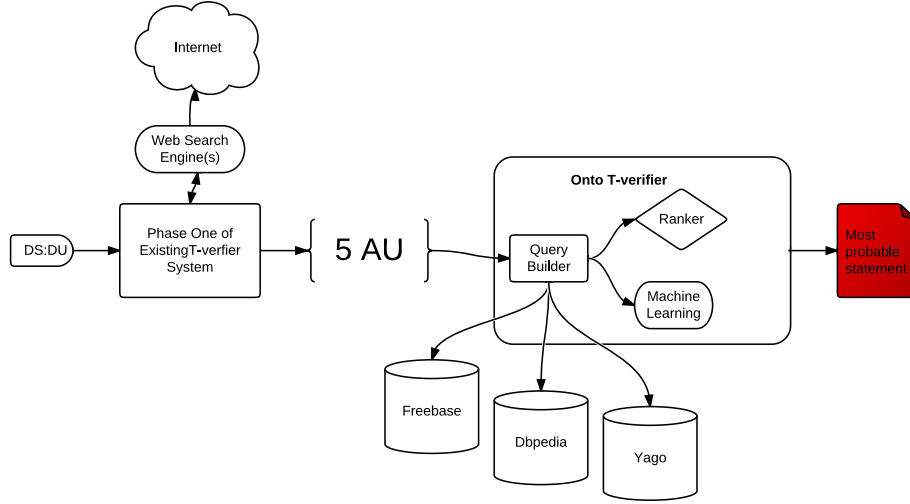
At the first phase, T-verifier generates alternative statements by supplying TU to search engine and collecting the relevant alternate DU -s. However, from basic web search may result in lot of alternate DU -s that are not semantically or logically relevant to the original DS . Hence, T-verifier uses combination of seven features to rank alternate DU -s. These features primarily exploit the facts that relevant alternative units frequently co-occur, people often mention both misconception and truthful conception together, data-type matching, and sense-closeness. T-verifier chooses top 5 alternative statements based on top 5 alternate DU -s obtained in this phase.

At the second phase, top 5 alternative statements from phase 1 is supplied to the search engine again. Then the returned searched result is ranked by multiple rankers such as Alternative Unit Ranker, Hits ranker, Text-feature Ranker, Domain Authority Ranker etc. Then all those ranks are merged to form an overall ranking among the alternate statements and top statement in this final merged ranking is considered as truthful statement.

4 Problems with Current System

Although results from the tested dataset achieves good performance (90% accuracy), failing of T-verifier for some statements shows that it has some inherent problems and there is room for improvement.

First, T-verifier assumes that truthful statements will be more propagated in the web compared to the untruthful statement. However, this may be not true in general because of intended and planned propaganda for establishing some untruthful statements. This kind of propagandas are even becoming more common now a days due to widespread reach of Internet. T-verifier also showed that "*Hillary Clinton is the President of United States*" has more hits than "*Hillary Clinton is the Secretary of State*". Although T-verifier was able to find the correct statement in this case using



(a) System overview

Figure 1: System overview

multiple ranks together, in general the untruthful statement can be prevalent in the web compared to truthful statements.

Second, T-verifier do not use the reputation of the information source. T-verifier uses only search results returned by the search engine irrespective of the origin and believability of the information origin. Hence there is room for improvement here to give more weight to the information obtained from trustworthy sources such as Wikipedia.

5 Proposed System

5.1 Description of extraction algorithm

For disambiguation among the alternative statements, Wikipedia is generally used as an authoritative source. On the other hand DBpedia, Freebase have organized the massive amount of data in a searchable fashion e.g. DBpedia uses SPARQL endpoint, Freebase uses MQL api. Open source implementation of Python wrappers exist for both the interfaces exist and appear to be mature enough for our needs.

5.1.1 Freebase

Freebase has information about approximately 20 million Topics, each one having a unique Id, which can help distinguish multiple entities which have similar names, such as Henry Ford the industrialist vs Henry Ford the footballer. Most of the topics are associated with one or more types[1] (such as people, places, books, films, etc) and may have additional properties like "date of birth" for a person or latitude and longitude for a location. Freebase not only contains data from the Wikipedia but also other sources; users can submit data to the Freebase datastore and expand it in richness. We tinkered with the api[2] and it appeared to be the most viable starting point for the project.

Listing 1: Minimal code to Freebase

```
import freebase
import pprint

query = [{
    "a:starring": [{
        "actor": "Meg Ryan"
    }],
    "b:starring": [{
        "actor": "Tom Hanks"
    }],
    "type": "/film/film",
    "*": [],
}]

pp = pprint.PrettyPrinter(indent=4)
result = freebase.mqlread(query)

print "Movie names & their various forms"

for i in result:
    pp.pprint(i["key"])
```

Listing 2: Cleaned Output

```
Movie names & their various forms
[ '158982',
  'You$0027ve_Got_Mail ',
  '18171032',
  ...,
  'E-m$0040il_f$00FCr_Dich ',
```

```

    'youve-got-mail ' ]
[   '176489 ',
    'Joe_Versus_the_Volcano ',
    'Joe_Vs$002E_The_Volcano ',
    'Brain_Cloud ',
    ...
    '2327353 ',
    'joe-versus-the-volcano ' ]
[   '226198 ',
    'Sleepless_in_Seattle ',
    'Sleepless_In_Seattle ',
    ....
    '169146 ',
    '106482 ',
    'Insonnia_d$0027amore ',
    '62812 ',
    'Schlaflos_in_Seattle ',
    'sleepless-in-seattle ' \ ]

```

The above results show how the three movies starring Tom Hanks and Meg Ryan. When we query Google with Tom Hanks and Meg Ryan, the top result is a page from Answers.com where a user has asked which are the movies where the two actors appear together, and the answer lists these three movies namely - “Joe versus the Volcano”, “ Sleepless in Seatle” and “You’ve got Mail”. A quick lookup of the Wikipedia and IMDB pages also confirm the same.

5.1.2 Dbpedia

DBpedia is a similar project to Freebase, but it focuses mainly on the content available from Wikipedia. It scores in being precisely importing the data from the info boxes in Wikipedia pages, but at this stage it does not seem to be offering anything additional over Freebase [4]. We are yet to explore its programmatic interface [3].

5.1.3 Yago

YAGO is a semantic knowledge base with over 900,000 entities (like persons, organizations, cities, etc.) and uses Wikipedia and Wordnet as its main source of information. We are yet to explore the programmatic interfaces it provides and how we can use it for the project.

5.2 Truthfulness verification

5.2.1 Building queries from the data supplied by the user

Formulating a proper Freebase query is for our specific purpose is a different process than the standard way of querying a search engine that does full text search on text documents. We start by introducing the various abstraction levels associated with the freebase data.

- A *type* is a conceptual container of related *properties* commonly needed to describe a certain aspect of a *topic*.
- A *topic* can be assigned one or more types (the default type being /common/topic)
- As *properties* are grouped into *types*, *types* are grouped into *domains*.
- *Domains*, *types*, and *properties* are given IDs in a *namespace/key* hierarchy.
- Common well-known topics are given IDs in the /en namespace, which are human-readable English strings.
- *Topics* are uniquely identified within Freebase by *GUIDs*.
- *Properties* are *multi-value* by default, and multi-value properties and single-value properties can be queried in the same way.

In order to transform a sentence to a freebase query we have to one to one map a content word from the sentence(*TU* plus *DU*) to the above mentioned abstraction. In other words, the process involves identifying the contextual meaning of the content words. Although this is pretty intuitive when we do it manually, trying to achieving this programmatically is one the challenging aspect of the project. One way of doing it is using a part of speech taggers, along with chunk extraction and named entity recognition. (details in next section)

The key point of distinction is that this result set is the set of records from a hierarchical database. We can not stuff in every word from the topic units into a query to freebase, as MQL(metaweb query language) is Query By Example language, and has a rigid structure which is not immediately obvious given a sentence in natural language. We incrementally build a query Q starting from with one word w_i from the word list L extracted from the TU . Let the results associated with q_{w_i} be (R_{w_i}) . Initially $Q = \{q_1 = w_i\}$

The following are the possible cases

- No results - In this case we get the synset from Wordnet S_{w_i} and repeat the search with each word in the synset w_j^s .
- If there is no match with the word or its synset, we reject w_i from the query Q and move on to the next word in L and repeat the process.
- If R is not empty, we retain the w_i or w_j^s in Q . Then we take the each of the remaining words w_j from L and S_{w_j} and search in R_{w_i} . If w_j or a synonym of it w_j^s is found to occur in the result, we augment Q with w_j (or w_j^s). So Q now becomes $Q = \{q_1 = w_i, q_2 = w_j\}$ or $\{q_1 = w_i, q_2 = w_j^s\}$
- With the new Q we again query Freebase and repeat the above steps.

- We terminate when all the words (and in their synsets) in L have been substituted. This allows us to form the most appropriate query Q from the TU . Note though this is essentially a breadth first search of the graph, we would not be traversing very deep (though the branching factor can be pretty high) because of the small number of content words in TU and their synsets.
- If the result returned by this query Q contains the DU , we can say with a good degree of confidence that the statement is true.

5.3 Examples of Truthfulness Verification

5.3.1 Use of Freebase

For all the 50 sentences mentioned in the original paper we tried the default POS tagger that comes with the Natural Language toolkit along with NE chunker, Binary NE Chunker and the IEER NE Chunker. None of the yeilded good results. So we used the Illinois named entity extractor from UIUC, which gave comparatively better results primarily because its database is built from various sources like the Wikipedia, Brown Hierarchical Word Clusters etc.

- Correctly tagged 19
- Partly correct 14
- Wrong/no identification 12

Consider one of the following sentences:

"Tom Hanks was the lead actress in the movie Sleepless in Seattle"

Tom/NNP Hanks/NNP was/VBD the/DT lead/NN actress/NN in/IN the/DT movie/NN Sleepless/NNP in/IN Seattle/NNP Phrases and Named Entities

PERSON: Tom/NNP PERSON: Hanks/NNP GPE: Seattle/NNP

Content words for this sentence are "Tom Hanks", "lead", "actress", "movie", "Sleepless in Seattle".

While "actress" is a domain in freebase, it does not contain anything yet. So we look into the synset of the word "actress" from wordnet, which includes *"female actor"*.

Now for the noun phrase *"Sleepless in Seattle"*, we can generate "id" for the query as "sleepless in seattle". But this id will be associated with many properties. To select the relevant property we can use synset obtained from the actress. And this synset has *actor*, which is one of the property for id *"Sleepless in Seattle"*. Hence the Freebase query can be following:

```
[{
  "id" : "/en/sleepless_in_seattle"
  "/film/film/starring" : [{ "actor" : null }]
}]
```

Now the result of this query returns following output:

```
"code":          "/api/status/ok",
"result": [{
  "/film/film/starring": [
    {
      "actor": "Tom Hanks"
    },
    {
      "actor": "Meg Ryan"
    },
    {
      "actor": "Bill Pullman"
    },
    {
      "actor": "Rosie O'Donnell"
    },
    {
      "actor": "Rob Reiner"
    },
    {
      "actor": "Victor Garber"
    },
    {
      "actor": "Gaby Hoffmann"
    },
    {
      "actor": "Carey Lowell"
    },
    {
      "actor": "David Hyde Pierce"
    },
    {
      "actor": "Ross Malinger"
    },
    {
      "actor": "Frances Conroy"
    },
    {
      "actor": "Rita Wilson"
    }
  ]
}]
```



```

    }
  ],
  "id": "/en/sleepless_in_Seattle"
}],
"status": "200 OK",
"transaction_id": "cache;cache03.p01.sjc1:8101;2011-03-09T05:09:44Z;0032"
}

```

One of the actors in this result set is "Tom Hanks" that matches with our content word "Tom Hanks" in the given sentence. Now as we know earlier that actress means female actor, we can use the keyword female to find the fact that female is type of gender and Freebase id of "tom hanks" has a property gender associated with it. So we can formulate following query.

```

[ {
  "id" : "/en/tom_hanks"
  "/people/person/gender" : {}
} ]

```

The result of this query is following:

```

{
  "code": "/api/status/ok",
  "result": [ {
    "/people/person/gender": {
      "id": "/en/male",
      "name": "Male",
      .
      .
      .
    }
  }
}

```

Here the gender is Male, which contradicts with our gender female. Hence we can decide that this statement is false.

For finding the true statement i.e. the actress we can use all actors obtained in the first query result and form second query with their names and output the truthful sentence if we get female as the gender.

5.3.2 Use of Wikipedia

Although structured databases such as Freebase, DBPedia extract the information from Wikipedia and format them into a structured and query-friendly way, it does not contain all the relevant information all the time. For example, consider the following sentence:

Les Paul invented the electric guitar

The wikipedia page "http://en.wikipedia.org/wiki/Electric_guitar" contains the inventor of the electric guitar in one of the sentences in the body. But this information is not available in the Freebase.

We can use a very simple and elegant method to find the relevant information from Wikipedia by using simple web-scraping and pattern matching using regular expression. The process is described below.

After getting the wikipedia page we can convert the html file into text file and then we can separate all the sentences in different lines. The sentence is the part of the whole text that exist between two full-stops. Another important observation is that the relevant information will be within the same sentence in the question-answer type sentences almost all the time. So we can match the content words to the sentences extracted from Wikipedia page and verify the truthfulness of the sentence.

For the example sentence above, after processing the Wikipedia page and matching by inventor we get following two sentences:

- *"Commercial production began in late summer of 1932 by Electro-Patent-Instrument Company Los_Angeles, a partnership of Adolph_Rickenbacker, PaulBarth and George_Beauchamp, the inventor"*
- *In a morecontemporary style, Little Willie Joe, the inventor of the Unitar, had a rhythmand_blues instrumental hit in the 1950s with "Twitchy", recorded with the ReneHall Orchestra*

When we do a regular expression search of the *DU* "Les Paul" in these two sentences, we do not get any match. Hence we can conclude that the statement is not true.

Contrary to our prior report, Wikipedia does have a very rich API, which the wiki software, Mediawiki provides. Of the various features that this API provides there are two searching mechanisms.

5.3.3 Opensearch Protocol

The first one is Opensearch protocol, which gets pages whose name case-insensitively match a given string. When default limit(10) is reached, results are ordered by number of incoming links. For example if we search for Forrest Gump, we will get:

Listing 3: Result using Opensearch

```
[
  * "Forrest Gump"
  [
    o "Forrest Gump"
    o "Forrest Gump (character)"
    o "Forrest Gump (novel)"
    o "Forrest Gump (soundtrack)"
    o "Forrest Gump Original Motion Picture Score"
    o "Forrest Gump (disambiguation)"
  ]
]
```

]]

This to a certain extent provides some semantic information attached alongwith the term.

5.3.4 Fulltext search

This provides a richer resultset, which syntactically resembles as the results returned for query on a web search engine. However, the interpretation of the returned results and the methods to analyse it has to significantly different. A real world example might make this distinction more clear. Searching the web is more like asking which books in the library has information on a particular topic, but full text search on wikipedia is more like pulling out a single volume of one's favourite encyclopedia and searching the index pages to find if there is an article on this topic. So each article has a definite focus and it is closely defined by its title. Majority of the wikipedia pages adhere to some structure, which might reveal more than the text that is in there for the article. For example the interwiki links, the See also links etc. point towards related topics. Before exploring those areas, we first focus on the various aspects that are immediately available from the api:

1. srinfo What metadata to return. Type: one of totalhits, suggestion
2. srlimit How many total pages to return. Type: limit
3. srnamespace The namespace(s) to enumerate. Type: namespace
4. sroffset Use this value to continue paging (return by query). Type: integer
5. srprop What properties to return:
 - (a) size - Adds the size of the page in bytes
 - (b) wordcount - Adds the word count of the page
 - (c) timestamp - Adds the timestamp of when the page was last edited
 - (d) score - Adds the score (if any) from the search engine
 - (e) snippet - Adds a parsed snippet of the page
 - (f) titlesnippet - Adds a parsed snippet of the page title
 - (g) redirectsnippet - Adds a parsed snippet of the redirect
 - (h) redirecttitle - Adds a parsed snippet of the redirect title
 - (i) sectionsnippet - Adds a parsed snippet of the matching section
 - (j) sectiontitle - Adds a parsed snippet of the matching section title
 - (k) hasrelated - Indicates whether a related search is available
6. srredirects Include redirect pages in the search. Type: bool
7. srsearch (required) Search for all page titles (or content) that has this value. Type: string

8. srwhat Search inside the text. Searching titles is disabled in Wikipedia.

5.4 Problem transformation - Recognizing Text Entailment

While a thorough experimentation needs to be done in how to utilize all these parameters to do a better extraction of the underlying semantics, we realize limiting the knowledge base to Wikipedia, reduces the problem to an instance of Recognizing Text Entailment (RTE). Over the last few years amazing progress has been made in this track, and the state of the art algorithms are quite complicated to be implemented within this short time frame. We start off with a simple idea for RTE, and list some experimentation that we plan to do ahead.

The main problem of RTE is that given a Hypothesis(H) and Text(T), the algorithm has to detect if T entails H.

T:The sale was made to pay Yukos' US\$ 27.5 billion tax bill, Yuganskneftegaz was originally sold for US\$ 9.4 billion to a little known company Baikalfinansgroup which was later bought by the Russian state-owned oil company Rosneft .

H:Baikalfinansgroup was sold to Rosneft.

The above example is taken from Recognizing Textual Entailment (RTE) 3 Challenge Corpora. The correct answer in this case, (i.e. yes T entails H) is easy for humans to comprehend, but obviously the greatest challenging task to do programmatically.

The basis of the transformation is that each *DS* now becomes a Hypothesis(H) and the wikipedia article the Text(T).

5.5 With or without the doubt unit

The original approach left out the doubt unit in order to not bias the returned results. However, in this case it might actually be beneficial. We can also go the other way, that is using the *DU* to search article and then check in that article if there is sufficient intersection with the remainder of the doubtful statement. While this may work well for names and places, this would not work out for dates enforcing us to have separate rules attached to different data types. We have only tried without the DU right now.

5.5.1 Without DU

We remove the doubt unit (DU) and stop words from the topic unit and use the remaining key words to do a full text search. The results returned by Lucene, search engine that powers these searches are very accurate. But our task is much more complex, and it requires even finer analysis of the result.

Of the various properties that a returned result may be decorated with, the two most important ones are:

- **titlesnippet** - This is probably the most important category of all. If the doubt unit is located here then it would mean that there is a wikipedia page for the doubt unit, and the text contains keywords from the doubtful statement. This is the strongest indication that that the statement might be true.
- **snippet** - This is the SRR in this case, that is formed by the concatenation of parts from three sentences that surround matched query keywords in the article content. The important thing to consider over here are the distaces between two such sentence parts and the number of matches each part includes. For a long page, there may be located at two very different areas in the page and be about very different contexts. In order to handle this we can take a weighted distance measure that takes the number of words between two matching sentence parts $0.33 * \text{number of matches in a sentence part} / (\text{length of the sentence part} * \text{length of document})$

Listing 4: Feature extraction initial steps

```

snippet_split_weights = {}
snippet_splits = split snippets by '<b>...</b>'
for snippet_split in snippet_splits:
    snippet_split_len = length(snippet_split)
    snippet_split_match_count = count number of matches in snippet_split
    snippet_split_weights[snippet_splits.index(snippet_split)]
        = snippet_split_match_count / snippet_split_len
snippet_weight = 0
for snippet_split_weight in snippet_split_weights.values():
    snippet_weight += snippet_split_weight
snippet_weight = 0.33 * snippet_weight

#TITLE
title_match_count = count all matches in result['titlesnippet']
title_len = length(result['title'])
title_weight = title_match_count / title_len

#TODO need to determine appropriate wieghts
overall_result_weight = (title_weight + snippet_weight) / 2

if regex_tokenize:
    tokenizer = RegexpTokenizer('([A-Z]\.)+|\w+|\$[\d\.]+' )

    snippet_tokens = set(tokenizer.tokenize(snippet))
    title_tokens = set(tokenizer.tokenize(result['title']))
    doubt_unit_tokens = set(tokenizer.tokenize(doubt_unit))
else:
```

```

        snippet_tokens = set(lower case the snippet and tokenize by space)
        title_tokens = set(lower case the title and tokenize by space)
        doubt_unit_tokens = set(lower case the doubt_unit and tokenize by space)

found = False

#If the match is found in the snippet
if snippet_tokens.intersection(doubt_unit_tokens):
    print "S",';',title_weight, ';', snippet_weight, ';', overall_result_weight
    found = True

#if the match is found in the Title
if title_tokens.intersection(doubt_unit_tokens):
    print "T",';',title_weight, ';', snippet_weight, ';', overall_result_weight
    found = True

#If no match is found
if title_tokens.intersection(doubt_unit_tokens):
    print "N",';',title_weight, ';', snippet_weight, ';', overall_result_weight
    found = True

```

As for the rest of the properties like redirecttitles, redirectsnippets, sectiontitle, sectionsnippets did not return values that seem to be of immediate help. However, they made find their use in generating alternative units. We have not yet experimented with them yet.

Currently the code that we have written only checks for presence of the doubt unit in each of the above mentioned property, for the fifty sentences that were there in the original paper.

The following are the immediate concerns

- Better matching: We tried two ways to match the DU-s in the returned results, one a conservative approach that would try to tokenize so that abbreviations like "U.S.A" and monetary amounts like "\$23.00" are kept as tokens. In the more relaxed approach the tokens were all lowercased and split by spaces. This obviously is the most crucial part of the entire process and it appears we should have separate strategy based on the data type of the doubt unit.
- Building the classifier: We got the data from TREC 9, which is in a question answer format and have finished converting 100 of the 711 sentences all to statement form with correct answers. Once we have some more sentences labeled we will use it to train a classifier from the features extracted above.

5.5.2 Data Description

The attached file shows the data collected for each sentence from wikipedia. We take the top 10 results returned by searching wikipeda, and calculate the values mentioned above. "S" denotes that

DU appears in the snippet, "T", shows that it appears in the Title, and "N" means the *DU* was not found anywhere. The other values are separated by ';' are title weight, snippet weight, overall result weight.

5.5.3 Use of answer.com

We found that answer.com also contains the correct answers to the for the doubt sentences in almost of all the cases. The precise Q&A format of closely matching the sentences excites us about the possibilities of using the above mentioned strategies for this site. A lot of domain specific sites (like stackexchange.com, quora.com) have become very popular since 2009, which deviate from the traditional forums in the sense that the user generated content is voted, summarised and exposed to a programmatic interface.

•

6 Conclusion

In this paper, we described current system for truthfulness verification of statements found in the web called T-Verifier and described the possible improvements to the current system. We will primarily use authentic information sources such as Wikipedia and Ontologies extracted from Wikipedia such as FreeBase, DBPedia, and YOGA to found the truthfulness of doubtful statements.

References

- [1] <http://blog.freebase.com/2007/03/19/the-type-system-in-freebase/>.
- [2] <http://code.google.com/p/freebase-python/>.
- [3] <http://sourceforge.net/projects/sparql-wrapper/files/sparql-wrapper-python/1.4.2/>.
- [4] <http://wiki.freebase.com/wiki/dbpedia>.
- [5] Clement Yu Xian Li, Weiyi Meng. T-verifier: Verifying truthfulness of fact statements. In *ICDE*, 2011.