# What is Disputed on the Web?

Rob Ennals
Intel Labs Berkeley
2150 Shattuck Ave
Berkeley, CA, USA
robert.ennals@intel.com

Dan Byler
School of Information
University of California at
Berkeley
Berkeley, CA, USA
daniel.byler@berkeley.edu

John Mark Agosta
Intel Labs Santa Clara
2200 Mission College Blvd
Santa Clara, CA, USA
john.m.agosta@intel.com

Barbara Rosario
Intel Labs Santa Clara
2200 Mission College Blvd
Santa Clara, CA, USA
barbara.rosario@intel.com

## ABSTRACT

We present a method for automatically acquiring of a corpus of disputed claims from the web. We consider a factual claim to be *disputed* if a page on the web suggests both that the claim is false and also that other people say it is true.

Our tool extracts disputed claims by searching the web for patterns such as “falsely claimed that $X$” and then using a statistical classifier to select text that appears to be making a disputed claim.

We argue that such a corpus of disputed claims is useful for a wide range of applications related to information credibility on the web, and we report what our current corpus reveals about what is being disputed on the web.

## Categories and Subject Descriptors

H.3.1 [**INFORMATION STORAGE AND RETRIEVAL**]: Content Analysis and Indexing; I.2.7 [**ARTIFICIAL INTELLIGENCE**]: Natural Language Processing

## General Terms

Design, Human Factors

## Keywords

dispute, web, credibility

## 1. INTRODUCTION

The web contains a vast number of pages written by a vast number of people. In many cases, these people disagree with each other, and the pages they write contain conflicting information. For users to extract reliable information from the web, it is important that they be able to determine the credibility of the information that they read. One way to determine whether information is credible is to check whether it is disputed by other sources.

In this paper we describe a method for automatically acquiring a corpus of disputed claims from the web. We use the word *claim* to mean a statement of opinion or fact contained in the text of a web page. We consider a claim to be *disputed* if a page exists on the web that suggests both that the claim is false, and also that there are others who say the claim is true. We are also interested in *who* disputes a particular claim, since a user is likely to be more interested in being told that a claim is disputed by a large number of sources that they believe are credible, rather than a small number of sources that they do not trust.

We identify disputed claims by searching for lexical patterns such as “the misconception that” (Section 3). For example, if a page contains the text “the misconception that *the moon is made of cheese*”, this suggests that the author believes that others are saying that the moon is made of cheese, and that the author themselves disputes this claim. In Figure 1 we list some of the patterns we use. In practice, many of the strings that match such a pattern are not well-formed disputed claims. We use a statistical classifier to select only those strings that appear to make a well-formed unambiguous claim.

We believe that this corpus of disputed claims will be useful for many purposes. We are currently using it as part of our Dispute Finder [11] project to automatically highlight phrases on web pages that are disputed by other sources. We are also exploring other applications of this corpus, including automatically detecting disputed claims in human speech, creating visualization tools that let users know what is disputed about a topic that interests them, and building statistical tools that allow people to look for patterns in Internet debate (Section 4). In Section 5 we show how our corpus can be used to reveal trends in dispute on the web.

We have made our corpus publicly available for other researchers to download at `http://confront.intel-research.net/`. It currently contains claims extracted from pages written on the days between October 1st 2009 and January 31st 2010, and pages written between January 1st and January 31st of every year from 2000 to 2010. This amounts to approximately 4.2 million strings that our classifier believes are making disputed claims. We plan to expand our corpus as we crawl more of the web and improve our algorithms.

We believe that ours is the first attempt to automatically acquire a corpus of disputed claims from the Web.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
*WICOW'10,* April 27, 2010, Raleigh, North Carolina, USA.
Copyright 2010 ACM 978-1-60558-940-4/10/04 ...$10.00.

## 2. BACKGROUND AND RELATED WORK

The problem of determining information credibility on the web is becoming increasingly important. In the past a user would typically obtain information from a relatively small set of sources such as books, TV channels, and radio stations. The user would have some idea of the reputation and biases of each of these sources, and publishing barriers and quality control would ensure that these sources only published information that met their standards. The web is very different. A user has access to a vast number of different sources, but knows little about the biases or reputation of most of them. Moreover, the Internet has little in the way of publishing barriers or quality control. If a user is to extract reliable information from the web then they need to either restrict themselves to a small set of trusted sources, or use some kind of mechanism to determine the credibility of the information that they access.

There are several ways that a user can determine whether to trust information that they find on a web page. The user can check the reputation of the person or organization that wrote the web page; they can observe whether the web page has features that they would expect a credible source to have; or they can check whether the information is consistent with information available from other sources.

A user can check the reputation of a source using a service such as SourceWatch.org, which publishes manually curated information about the reputation and known biases of various sources. Trustpilot.com produce a Firefox extension that warns a user when they are looking at a web page hosted by a company that they believe is not trustworthy. Alternatively, a user can simply use a search engine such as Google to look for information about the source's reputation. While these tools can be very useful, trustworthy sources sometimes publish unreliable information, and untrusted sources sometimes contain useful information. For example, a reliable source may have been misled by an unreliable source they were using themselves; or a small unrated blog may publish information that is useful, insightful, and accurate.

Researchers have identified a variety of metrics that can be used to automatically estimate the quality of a document based on looking at its content. For Wikipedia, Blumenstock [4] estimates the quality of an article by the word count and WikiTrust [1] identifies potentially unreliable sections of an article by analyzing its edit history. Custard and Sumner [7] use a combination of metrics to measure web site quality, including number of links and whether it contains videos. Fogg et al [13] have shown that users commonly evaluate the credibility of a web site based on factors such as the design look, the information structure of the site, and the tone of the writing; some of the factors identified by Fogg et al could potentially be measured automatically and used to guide a user. These metrics do a good job of detecting pages that resemble pages that contain unreliable information, but they do not protect against authors who write untrustworthy information in the style of a trustworthy document.

A third approach is to inform a user when a document contains information that other sources disagree with. Annotation tools such as ReframeIt.com, ShiftSpace.org, and SpinSpotter.com allow a user to manually annotate a web page that they disagree with, overlaying their own opinions on top of existing content. Cohere [32] allows a user to manually annotate a web page as making a claim that is part of an argumentation graph. These tools work well when a user has marked up the page the user is reading, but are limited by the requirement that a human needs to manually mark up each page.

In previous work, we built an extension to the Firefox web browser called Dispute Finder [11] that informs a user when a web page makes a claim that is in its database of known disputed claims. For example, if a user is reading a page that says "Elvis is alive" and Dispute Finder recognizes this as being an instance of a known disputed claim, then Dispute Finder will highlight the text as being disputed and direct the user to web pages that put forward alternative points of view. In its currently released version, Dispute Finder builds a corpus of disputed claims by allowing users to add disputed claims to its corpus manually, and by scraping a small set of web sites that manually curate such claims (currently Politifact.org and Snopes.com). This approach scales better than having users manually mark individual pages, but it is still difficult to make this approach scale to the huge number of claims on the web that are disputed.

A more scalable approach is to automatically detect when web pages disagree with each other. Tools like Statement Map [25] and WISDOM [20] use *contradiction detection* to inform a user when one document contains a statement that contradicts a statement on another web page.

There has been significant work on detecting contradictions. Condoravdi et al [6] argue that contradiction detection is one of the key tasks in language understanding. De Marneffe et al [9] present a taxonomy of the different ways that claims can contradict each other and describe a system that combines many techniques to detect different kinds of contradictions. AuContraire [30] uses TextRunner [12, 2] to infer subject-verb-object relationships, looks for cases where a verb maps the same subject to multiple objects, and uses semantic knowledge to determine whether this implies that there is a contradiction. Harabagiu et al [16] look for contradictions where one claim is a negated paraphrase of another. The RTE-3 Recognizing Textual Entailment challenge [14] included an optional contradiction detection task, allowing different groups building contradiction detection algorithms to compare their results.

Detecting contradictions has proven to be a hard problem [14]. Much of this difficulty comes from the fact that one typically needs deep semantic and contextual knowledge to determine whether two statements that look like they might contradict each other actually do. Two instances of the same phrase may have the same sense, but a different reference, or two different phrases may refer to the same thing. For example "George Bush is married to Barbara Bush" does not contradict "George Bush is married to Laura Bush" because there is more than one George Bush; "Alan Turing was born in England" does not contradict "Alan Turing was born in London" because London is in England; and "It is raining is San Francisco" does not contradict "It is not raining in San Francisco" if the statements were made at different times.

In our work, we attempt to bypass the difficulty of contradiction detection by looking for *disputes* rather than contradictions. Rather than looking for claims that contradict each other, we look for linguistic clues that writers believe that there is a contradiction between what they believe and what others believe; and that this contradiction is important. Looking for disputes rather than contradictions allows humans to do the hard work of identifying contradictions and deciding whether they are important. If a page says "Falsely claimed that Obama is a Muslim" then that tells us that the author believes that other web pages are claiming that "Obama is a Muslim" and that the author believes this claim is false. For our purposes, disputes are also more interesting than contradictions; we are interested in the social side of dispute; we want to know who disagrees with who, why they disagree, and what they think is important.

One disadvantage of looking for disputed claims rather than contradictions is that it limits us to only finding those claims that some

| **Frequency*<br>on the web** | **Pattern** | **Pattern<br>Precision** | **Classifier<br>Accuracy** |
|---|---|---|---|
| *Plausible to not believe it* | | | |
| 495,000,000 | believe that | 48% | 65% |
| 101,000,000 | claim that | 46% | 72% |
| *It is notable that others believe it* | | | |
| 40,100,000 | claiming that | 39% | 65% |
| 11,700,000 | who believe that | 69% | 73% |
| 8,400,000 | who think that | 49% | 66% |
| *Something false* | | | |
| 5,790,000 | the myth that | 62% | 73% |
| 3,260,000 | into believing that | 52% | 72% |
| 1,690,000 | the lie that | 52% | 76% |
| 1,410,000 | it is not true that | 64% | 78% |
| 1,220,000 | the delusion that | 54% | 76% |
| 1,140,000 | the misconception that | 67% | 81% |
| 676,000 | the mistaken belief that | 51% | 74% |

*Yahoo's approximate estimate

**Figure 1: Some of the patterns we use to find disputed claims**

human has explicitly said they disagree with. For a small corpus this would be a limiting factor; however our hypothesis is that the web is sufficiently large that, at least for areas of general interest, any sufficiently important claim that is false will be explicitly disputed by somebody.

Our primary motivation for automatically acquiring a large corpus of disputed claims is to use this corpus to enable tools like Dispute Finder to automatically inform users when they encounter information that this corpus says is disputed. However, as we discuss in Section 4, we believe such a corpus could be useful for many other purposes too.

Another closely related area is sentiment analysis/opinion mining [18, 28]. Opinion mining tries to determine what an author's opinion is about certain objects or certain features of certain objects. A focal application has been automatically summarizing of product reviews to produce an overview of the product features that are viewed favorably or negatively. Dispute mining could be thought of as opinion mining applied to beliefs and ideas, rather than features of objects. To put it another way, dispute mining can be thought of as a second order of opinion mining in which we are looking for opinions about opinions.

# 3. FINDING DISPUTED CLAIMS

We implemented a three-stage process to build our corpus of disputed claims, each described in the following Sections. We first create a set of lexical patterns such as "the misconception that" and "it is not the case that" (Section 3.1). We then search the Web for these patterns (Section 3.2). Finally, we filter the resulting strings to give only those that resemble unambiguous falsifiable claims (Section 3.3).

## 3.1 Patterns for Finding Disputes

We hand-crafted a set of 54 patterns that we hypothesize may indicate that a claim $S$ is disputed, such as "false claim that $S$" or or "it is not true that $S$". Figure 1 shows some of the patterns we use.

This is an instance of pattern matching, a method that has been used extensively for many NLP tasks. Hearst [17] searched for patterns such as "$X$ such as $Y$" to infer hyponyms; Riloff [29] searched for words such as "kidnapped" to find information about terrorist events; Caraballo [5] searched for patterns like "$X$ and $Y$" to find nouns of the same type; Girju et al [15] searched for patterns like "$X$'s $Y$" and "$X$ of the $Y$" to infer metronyms; and TextRunner [12] searches for more general patterns to extract logical relationships from the web. These types of methods differ on how the patterns are generated and on how the extraction of the relevant information is performed. Most often they construct the patterns automatically and often choose them on the basis of statistics.

In this work we started with an initial set of manually crafted search patterns. We then identified additional patterns by searching for the text of known disputed claims, observing what text commonly occurred as a prefix of a known disputed claim, and manually choosing those patterns that made sense to us. This technique is inspired by bootstrapping algorithms [19, 15, 33] that start with a small number of seed items that belong to a semantic class of interest and use these seed items to learn extraction patterns that can be used to find new members of the same semantic class. In our case, the semantic entities of interest are the disputed claims. Currently, we *manually* choose the patterns to retain, rather than having the algorithm pick the patterns automatically.

Figure 1 shows a subset of the total of 54 patterns that we used. For each pattern, the *frequency* is the number of pages that Yahoo estimates contain this pattern. The *precision* is an estimate of the proportion of text strings matched by the pattern that are unambiguous disputed claims. We discuss precision further in Section 3.3.

The linguistic patterns we use are instances of *modal phrases* that qualify the meaning of the subordinate clause that follows the word "that" in the pattern. Specifically we are interested in modals that qualify belief. The phrase offers a syntactic clue that the statement following it is not to be taken at "face value". Modal clues are the basis for some of the preliminary syntactic analysis we've attempted as a way to generalize claim-finding templates.

Modals serve a broad variety of purposes; we are interested particularly in **epistemic** modality. Epistemic phrases qualify the degree to which the claim is known [27]. Most linguistic scholarship has focused on *possibility* and *necessity* as epistemic qualifiers. In contrast, our interest is in qualifiers that call the claim into question, as a presumption, speculation, or a questionable opinion. This can often, but not always done by attributing the claim to some other person's belief, either implicitly or explicitly. The author in so doing raises an issue about the truth, or lack of truth of the claim.

Not all patterns have the same meaning. We grouped our patterns into three loose groups, based on manual inspection of the pages that use them (Figure 1):

- **False:** The author believes the claim is wrong
- **Notable that others believe it:** The author believes that it is notable that others believe it. For example "Fans are claiming that *Elvis is alive*". Further inspection is required to determine whether the author disagrees with the claim. For example, do they follow it with a word like "but" or "despite"?
- **Plausible to not believe it:** The author feels the need to say that someone believes it, implying that it is plausible that

|  | **extracted text** | **valid** | **details** |
|---|---|---|---|
| 1. | the false claim that *won't go away* | no | "wont go away" is not a statement |
| 2. | falsely claimed that *he didn't do it* | no | "he" and "it" are unbound referents |
| 3. | falsely claimed that *federal labor laws do not apply* | no | ambiguous: apply to what? |
| 4. | false claim that *Elvis is alive* despite all evidence | yes | but drop everything after "despite" |
| 5. | wrongly believe that *the moon is made of cheese* | yes | it is a statement about the world |

**Figure 2: We filter out text that does not look like a statement**

| **accuracy** | **count** | **percent** | **cumulative** |
|---|---|---|---|
| dead on | 16 | 32% | 32% |
| within 10 days | 9 | 18% | 50% |
| within 30 days | 12 | 24% | 74% |
| within 90 days | 3 | 6% | 80% |
| within 1 year | 6 | 12% | 92% |
| off by more than 1 year | 4 | 8% | 100% |

**Figure 3: Accuracy of simple date detection**

someone else might not believe it. This is subtly different from the previous group. For example "Fans believe that *Elvis is alive*" is less strong than "Fans are claiming that *Elvis is alive*". It is useful to exclude the cases where the pattern is prefixed by "I", since such sentences rarely contain useful disputed claims. For example "Some believe that Elvis is dead" implies much more skepticism than "I believe Elvis is dead".

Which pattern group one should use depends on what one wishes to use the resulting corpus of disputed claims for. If one wants to only have claims for which we can present web pages that argue that the claim is false, then only the first set of patterns should be used. If however one just wishes to see what things people think are worthy of having opinions expressed about, then the full set is useful.

## 3.2 Searching the Web for Claims

We use the Yahoo BOSS API [34] to search for occurrences of our patterns on the web. Using Yahoo BOSS allows us to search for patterns on a vast number of web pages without having to build our own search engine. We search Yahoo BOSS for a raw string such as "the misconception that", download every page that Yahoo lists in its search results, and then look for text that matches any of our known patterns.

Since we are using a general purpose search engine, we are limited to searching for simple text strings. This requires us to generate multiple patterns that use different synonyms of the same word; for example we have both "the lie that" and "the deceit that". We are also restricted to searching for patterns that consist entirely of either a prefix or a suffix. For example we could not search for "say that $S$ despite", because there is no way to require that "despite" and "say that" be in the same sentence.

In the longer term we may work round these issues by building our own search infrastructure, but Yahoo BOSS provides an excellent platform for prototyping the idea to see if it works.

It is useful to be able to update our corpus as new pages appear on the web without having to repeatedly download the same pages. The Yahoo BOSS API does not offer the ability to request only recent pages, or to specify a particular date range. We instead simulate this behavior by including a literal date string in the queries we pass to Yahoo BOSS. For example, if we want to find claims that were disputed on January 10th 2010, then rather than searching for `"it is not true that"`, we instead search for `"it is not true that" "January 10 2010"`.

The reasoning behind this technique is that many articles on the web include the date the article was posted in the text of the web page, particularly news articles and blog posts. This approach has false positives and false negatives. Some web pages include dates that are not the date the article was posted, some web pages do not include the date the article was posted, and many web pages include dates in a different format. Despite being very primitive, this method seems to work reasonably well. When we manually inspected 50 pages retrieved by our algorithm, we found that 92% were written on the correct year, and 74% were written in the correct month (Figure 3). We recognize that our current implementation suffers from a systematic bias towards sites that write their dates in US format (month first) rather than UK format (day first).

Knowing the rough date when someone believed or disputed something allows us to produce fuzzy time trends (Section 5.1). Moreover, it is important to know when people believed particular things, since things that were true in the past may not be true any more (e.g. Elvis was alive in the past, but isn't now).

## 3.3 Filtering Claims

Searching the web for patterns yields a collection of strings. From this collection we want to filter out any strings that are not making a falsifiable claim and any strings that make a claim that is ambiguous.

Figure 2 shows some examples of the kind of strings that can match our patterns and the reasons why we want to filter some of them out. A chunk of text might not be a statement (case 1), it might include unbound referents such as "it" or "they" (case 2), or it might otherwise be ambiguous what entities it is referring to (case 3).

The *pattern precision* column in Figure 1 shows the estimated precision for different patterns. This is the estimated percentage of the strings that match the pattern that are unambiguous falsifiable claims. We estimated the precision of each pattern by randomly selecting 100 strings that match the pattern and then manually marking strings that we subjectively decided were well formed and unambiguous.

Whether a claim is unambiguous is often subjective. For example, in the string "Obama is the president" it is possible that "Obama" is referring to someone other than "Barack Obama" and "the president" is referring to a position other than "President of the United States". When manually judging claims, we considered a claim to be unambiguous if an average person would be pretty sure what the claim meant without knowing its context.

We built a statistical classifier that attempts to automatically determine whether any given string matched by a pattern is an unambiguous disputed claim. This is a classical text classification task where the text is classified as either being an unambiguous falsifiable claim or an invalid one (binary classification). Naive Bayes classifiers [10] have been used extensively for language tasks (for

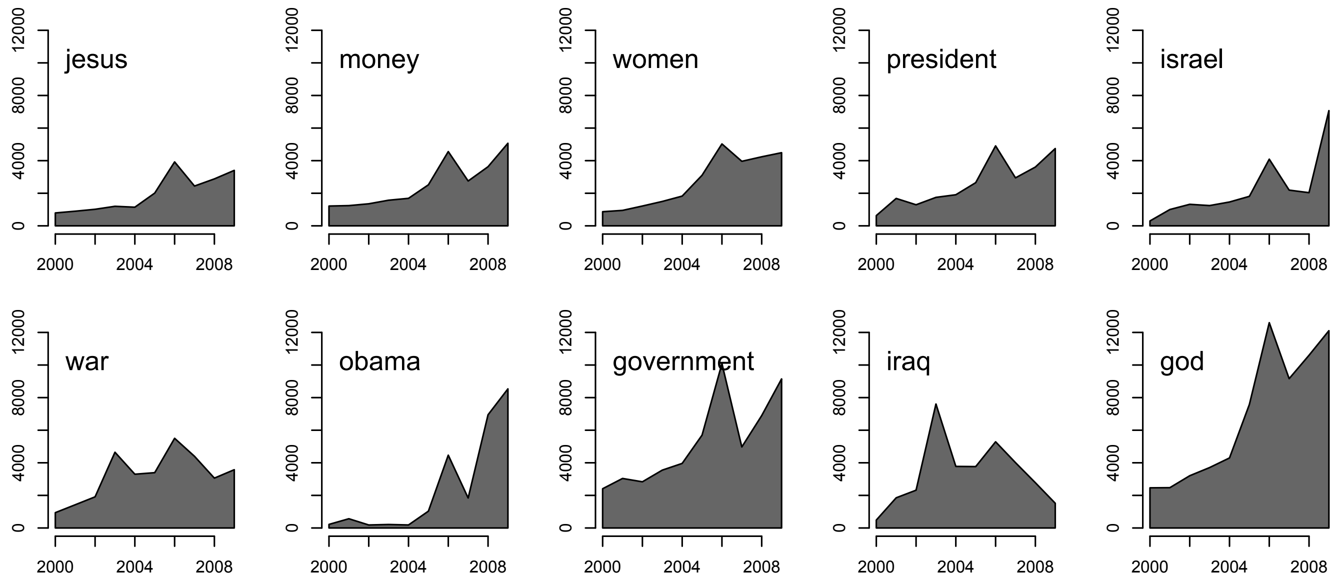

**Figure 4: Number of claims containing a particular noun made during January of any particular year**

example, to decide whether an email is spam or not [31] and to decide the topic of a document [23]). We used a simple Naive Bayes classifier whose features were the words that were present in the string, the parts of speech that were present in the string, and the identity and part-of-speech tag of the first and second words.

The *classifier accuracy* column in Figure 1 shows how accurately our classifier was able to determine whether any given string matched by a particular pattern is an unambiguous disputable claim. To provide a baseline for what these figures mean, consider that a random classifier would have 50% accuracy, a classifier that always said "yes" would have accuracy equal to the pattern precision (e.g. 62% for "the myth that"), and a classifier that always said "no" would have accuracy equal to 100 minus the pattern precision (e.g. 52% for "believe that").

We can observe that the patterns in the "something false" group have much higher precision than the "plausible to not believe it" group, but appear much less frequently on the web. Similarly, we can observe that our classifier performs much better on the "something false" patterns, with accuracy getting into the high seventies.

Manual inspection suggests that the classifier is fairly good at filtering out strings that are not claims (case 1) but does poorly at working out whether a claim is unambiguous. This is not surprising. We anticipate that a more sophisticated technique with real world knowledge would be needed in order to accurately determine whether a claim was unambiguous.

In some cases a disputed claim will be followed by information about where the claim was made, or why the claim is wrong. For example "he claimed that the moon was made of cheese *on his show*" or "he claimed that the moon is made of cheese *despite contrary evidence*". We chop off such suffixes by discarding any text that follows a word such as "despite", "however", "but", or "was", provided that word was proceeded by something that looks like a statement.

### 3.4 Context

We save the surrounding text, page title, and URL for each of the disputed claims we see made on the web. At present we do not do anything with this data, but we anticipate that it could be useful for many purposes; for example it may be possible to resolve the meaning of ambiguous words or entities based on the words nearby like Wikify [24] does. Since many web sites have content written by multiple writers with different degrees of credibility, it is also useful to be able to know which writer disputed a particular claim. Other useful contextual information includes what organization hosted the page, who they link to, what meme-like phrases they make near to the claim, and whether the page is written in an "expert" style.

## 4. USING DISPUTED CLAIMS

There are many potential uses for a corpus of disputed claims.

In previous work, we have built Dispute Finder [11], a web browser extension that informs a user if the page they are browsing appears to make claims that are disputed by other sources. In the future, we intend that Dispute Finder will consult our corpus of disputed claims to determine whether claims made on a web page are disputed.

In future work, we intend to apply a similar dispute-finding approach to many other sources of information. We are working on a tool that listens to conversations and informs a user if they hear or say anything that is disputed by a reliable source. We are also interested in enhancing standard tools such as search engines, news aggregators, news readers, and email clients to inform a user if the information they are consuming may be disputed.

Since information may only become disputed after you read it, we are also interested in building tools that track all the information a user consumes and alert the user if something they read in the past ceases to be true, is corrected, or becomes disputed. Such a tool could also track the topics a user is interested in and inform the user when something new is disputed about the topic.

A third category of applications could be broadly classified as research tools: tools that allow a user to explore what is disputed about a particular topic, or to more broadly examine the social dynamics of disputes on the web. As an increasing proportion of human knowledge is brought to the web, these tools will be increasingly useful for sociological and historical research.

# 5. ANALYZING THE DATA

We downloaded all text strings that matched each of our 54 patterns for all dates from October 1st 2009 to January 31st 2010, and for all dates from January 1st to January 31st for every year from 2000 to 2010. This gave us a total of 7 million text strings, of which 4.2 million were identified by our classifier as being unambiguous disputable claims. This does not mean that there are 4.2 million distinct beliefs that are disputed, since many of the claims in our corpus are repeating claims made elsewhere in the corpus.

## 5.1 Trends over time

Figure 4 shows the frequency with which the ten most common nouns were mentioned in particular years. For example the claim “Obama is a muslim” contains the nouns “Obama” and “Muslim”. We used the NLTK [3] part-of-speech tagger to identify the nouns found in each of our claims. We then analyzed the distribution of noun frequencies in claims made between January 1st and January 31st of each year from 2000 to 2010. Although the dates we searched for were in January, the claims we found are often from other points in the year, due to the noisy nature of our date search mechanism (Figure 3). We excluded uninteresting nouns such as “someone”, “anybody” and “something”, and plotted graphs for the ten most disputed nouns (Figure 4). All graphs are plotted on the same scale. The y axis is the number of claims that contained the noun, and the x axis is the year in which the claim was made. It is likely that a Named Entity Recognition algorithm [26] would produce better results, but even nouns prove to be quite interesting.

We manually inspected our claims to determine what was going on in the world that made particular nouns spike at particular times:

- **Obama:** Barack Obama is not a topic of dispute until 2004 when he spoke at the Democratic convention. His level of controversy then rose in 2008, in the run up to the November 2008 US presidential elections and then again after he became president. We believe that most of the mentions of Obama before 2004 are due to date classification errors.

- **God:** God is always a controversial topic. We theorize that the spike in 2006 may relate to the 2006 publication of the Richard Dawkins book “the God Delusion”[8].

- **Iraq:** The biggest spike seems to correspond to the initial invasion in 2003.

- **Israel:** Spikes seem to correspond to the 2009 Gaza conflict and the 2006 Lebanon conflict.

These graphs are similar to those produced by MemeTracker [22], which graphed the times when particular memes were discussed on the web. MemeTracker tracked frequently occurring phrases, rather than nouns used in disputed claims. Kittur et al [21] graphed the categories on Wikipedia that contained the most conflict.

| Noun | Word Cluster | Freq |
|---|---|---|
| **god** | *no significant clusters* | |
| **government** | health care | 2605 |
| | united states | 1710 |
| **iraq** | weapons mass destruction | 3874 |
| | united states | 2972 |
| | posed imminent threat | 2547 |
| | saddam hussein | 2088 |
| | damaging manner justify armed forces | 1814 |
| | national security interests | 1673 |
| | use nation | 1503 |
| | al qaeda | 1200 |
| **israel** | united states | 1211 |
| | middle east | 556 |
| | off map | 543 |
| | west bank | 426 |
| | second coming | 296 |
| **jesus** | *no significant clusters* | |
| **money** | root evil | 503 |
| **obama** | working best interests | 1077 |
| | health care | 1074 |
| | united states | 741 |
| | birth certificate | 434 |
| | nobel prize | 311 |
| **president** | united states | 1695 |
| | white house | 716 |
| | george w | 618 |
| | health care | 491 |
| | saddam hussein | 369 |
| | off map | 363 |
| | mahmoud ahmadinejad | 320 |
| | sarah palin | 277 |
| **war** | united states | 1323 |
| | mass destruction | 484 |
| | middle east | 421 |
| **women** | *no significant clusters* | |

**Figure 5: Largest word clusters found for nouns**

## 5.2 What is Disputed about a Noun?

Once we know what nouns are disputed, it is interesting to know what it is about a noun that is disputed. For example, what is it that people are disputing about “Obama” or “Iraq”? To address this question, we took the set of claims that contained each popular noun (e.g. “Obama”) and looked for sets of words that occurred together much more frequently than would be expected if claims were formed randomly. For example in our claims about “Obama”, we found the word cluster “health care” because most claims about Obama that contain the word “health” also contain the word “care” and vice-versa. We used an agglomerative clustering algorithm to group together words that co-occur in the same claims.

Figure 5 shows the most frequently occurring word clusters that we found for each noun. For each cluster, the count is the number of claims about that noun that contain all of the words in the cluster. Most of these will make sense to people who have been following news in the United States. For example, we see that disputed claims about Obama often talk about his Nobel Prize or his birth certificate, and we see discussion about President Ahmadinejad’s alleged statement that Israel should be wiped off the map (which the authors had not previously realized was disputed). Some of the word clusters are not particularly interesting; for example many claims talk about the United States, even when that is not an important part of what is disputed.

It would be interesting to see how these word clusters compare to word clusters for all sentences on the web about these nouns, including those that are not disputed. Unfortunately we do not currently have access to such data.

In the future, we plan to use ideas from textual entailment [14] to cluster claims together directly, and then automatically train a classifier that we can use to identify text that is making a disputed claim. This will allow us to better attack some of the problems we outlined in Section 4.

## 6. CONCLUSIONS

Our research has shown that the web can be used as a corpus to determine what claims people identify as being disputed. Although it is likely that any specific claim found on the web is not authoritative, the web as a whole aggregates the collection of views that can be be found online. Furthermore, because individuals express opinions freely online, a snapshot of the disputed claims online can provide valuable insights into the cultural zeitgeist of the time.

Creating a comprehensive corpus of disputed claims presents many challenges: how to create patterns, how to search for patterns, how to avoid malformed sentences, how to resolve or filter out ambiguous claims, how to determine if two claims are paraphrases, and how to learn interesting things from the data set. For each of these tasks, we have implemented enough to get useful results, but there is much scope to do things better.

There are many possible applications for a corpus of disputed claims. These include real-time dispute recognition, autonomous research agents that operate in the background, and sociocultural research.

Our data set currently includes approximately 4.2 million records. In the future we intend to expand this by several orders of magnitude.

## 7. REFERENCES

[1] B Thomas Adler, Krishnendu Chatterjee, Luca De Alfaro, Marco Faella, Ian Pye, and Vishwanath Raman. Assigning Trust to Wikipedia Content. In *WikiSym*, 2008.

[2] M. Banko, O. Etzioni, and T. Center. The tradeoffs between open and traditional relation extraction. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08)*, 2008.

[3] Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly, 2009.

[4] Joshua E Blumenstock. Size Matters: Word Count as a Measure of Quality on Wikipedia. *Computing*, pages 2007–2008, 2008.

[5] Sharon Caraballo. *Automatic construction of a hypernym-labeled noun hierarchy from text Sharon A. Caraballo*. Phd, Brown University, 2001.

[6] Cleo Condoravdi, Dick Crouch, Valeria de Paiva, Reinhard Stolle, and Daniel G. Bobrow. Entailment, intensionality and text understanding. In *Proceedings of the HLT-NAACL 2003 workshop on Text meaning -*, pages 38–45, Morristown, NJ, USA, 2003. Association for Computational Linguistics.

[7] Myra Custard and Tamara Sumner. Using Machine Learning to Support Quality Judgements. *D-Lib Magazine*, 11(10), 2005.

[8] Richard Dawkins. *The God Delusion*. Houghton Mifflin Harcourt, 2006.

[9] M.C. de Marneffe, A.N. Rafferty, and C.D. Manning. Finding contradictions in text. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08)*, 2008.

[10] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

[11] Rob Ennals, Beth Trushkowsky, John Mark Agosta, Tye Rattenbury, and Tad Hirsch. Highlighting Disputed Claims on the Web. In *Proceedings of the 19th international conference on World Wide Web - WWW 2010*, 2010.

[12] Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S Weld. Open information Extraction from the Web. *Communications of the ACM*, 51(12), 2008.

[13] B J Fogg, Persuasive Technology Lab, Cordura Hall, Cathy Soohoo, David R Danielson, Ellen R Tauber, Sliced Bread Design, Mountain View, Julianne Stanford, Leslie Marable, and Consumer Webwatch. How Do Users Evaluate the Credibility of Web Sites? A Study with Over 2,500 Participants. *World Wide Web Internet And Web Information Systems*, pages 1–15, 2003.

[14] D. Giampiccolo, B. Magnini, I. Dagan, and B. Dolan. The third pascal recognizing textual entailment challenge. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, number June, page 1âĂŞ9, 2007.

[15] Roxana Girju, Adriana Badulescu, and Dan Moldovan. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of HLT/NAACL*, volume 3, 2003.

[16] S. Harabagiu, A. Hickl, and F. Lacatusu. Negation, contrast and contradiction in text processing. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 755. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[17] Marti a. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Computational linguistics*, Morristown, NJ, USA, 1992. Association for Computational Linguistics.

[18] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, page 168, 2004.

[19] Rosie Jones, Andrew McCallum, Kamal Nigam, and Ellen Riloff. Bootstrapping for text learning tasks. In *IJCAI-99 Workshop on Text Mining: Foundations, Techniques and Applications*, 1999.

[20] Daisuke Kawahara, Sadao Kurohashi, and Kentaro Inui. Grasping Major Statements and Their Contradictions Toward Information Credibility Analysis of Web Contents. *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pages 393–397, 2008.

[21] Aniket Kittur, Ed H Chi, and Bongwon Suh. What's in Wikipedia? Mapping Topics and Conflict Using Socially Annotated Category Structure. In *CHI*, pages 1509–1512, 2009.

[22] Jure Lescovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the Dynamics of the News Cycle. 2009.

[23] A. McCallum and K. Nigam. A comparison of event models for naive bayes text classification, 1998.

[24] Rada Mihalcea and Andras Csomai. Wikify! Linking Documents to Encyclopedic Knowledge. In *CIKM*, pages 233–241, 2007.

[25] Koji Murakami, Eric Nichols, Suguru Matsuyoshi, Asuka Sumida, Shouko Masuda, Kentaro Inui, and Yuji Matsumoto. Statement Map: Assisting Information Credibility Analysis by Visualizing Arguments. In *Workshop on Information Credibility on the Web (WICOW)*, 2009.

[26] D. Nadeau and S. Sekine. A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1), 2007.

[27] Frank Robert Palmer. *Mood and Modality*. Cambridge Univesity Press, 2001.

[28] Bo Pang and Lillian Lee. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of the Association for Computational Linguistics*, 2004.

[29] E. Riloff. Automatically constructing a dictionary for information extraction tasks. In *Proceedings of the National Conference on Artificial Intelligence*, 1993.

[30] Alan Ritter, Doug Downey, Stephen Soderland, and Oren Etzioni. It's a Contradiction - No, it's Not: A Case Study using Functional Relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2008.

[31] Mehran Sahami, Susan Dumais, David Heckerman, and Eric Horvitz. A bayesian approach to filtering junk e-mail. In *Proceedings of the AAAI'98 Workshop on Learning for Text Categorization*, 1998.

[32] Simon Buckingham Shum. Cohere: Towards Web 2.0 Argumentation. In *Computational Models of Argument (COMMA)*, volume 44, 2008.

[33] R. Snow, D. Jurafsky, and A.Y. Ng. Learning syntactic patterns for automatic hypernym discovery. In *Advances in Neural Information Processing Systems*, volume 17. Citeseer, 2005.

[34] Yahoo BOSS. `http://developer.yahoo.com/search/boss/`.