Statement Map: Assisting Information Credibility Analysis by Visualizing Arguments

Koji Murakami† Eric Nichols† Suguru Matsuyoshi† Asuka Sumida†¶
Shouko Masuda†‡ Kentaro Inui† Yuji Matsumoto†
†Nara Institute of Science and Technology
8615, Takayama-cho, Ikoma, Nara 640-0192 JAPAN
‡Osaka Prefecture University
1-1, Gakuen, Naka-ku, Sakai, Osaka 599-8531 JAPAN
¶Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292 Japan
{kmurakami,eric-n,matuyosi,asuka-s,shouko,inui,matsu}@is.naist.jp

ABSTRACT

In this paper we introduce Statement Map, a project designed to help users navigate the vast amounts of information on the internet and come to informed opinions on topics of interest. It does this by mining the Web for a variety of viewpoints and presenting them to users together with supporting evidence in a way that makes it clear how the viewpoints are related. In this paper, we discuss the need to address issues of information credibility on the internet, outline the development of Statement Map generators for Japanese and English, discuss the technical issues that are being addressed, and report on the construction of the resources necessary to meet the project's goals.

Categories and Subject Descriptors

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

General Terms

Algorithms, Languages

Keywords

Statement Map, Recognizing Semantic Relations

1. MOTIVATION

1.1 The Internet as a Source of Information

The importance of the internet as a source of information cannot be disputed. A recent poll [26] by the Pew Research Center found that among Americans the internet has overtaken newspapers as a news outlet and rivaled television for those surveyed under the age of thirty. Another poll showed that over 80% of Japanese internet users check news online on a daily basis [43].

Recent research reports that people are turning to the internet for information on important decisions like health care, medical information, and large purchases [17]; yet these users often lack the knowledge necessary to evaluate the credibility of online information [23]. This is particularly worrisome when one considers the dearth of bad information present on the Web; reports of users losing money in internet auction scams or personal information to phishers are commonplace. In an age where publication is as simple as uploading a file, and anyone with a computer can reach a large audience, the old adage not to believe everything you read is more relevant than ever.

1.2 The Anti-Vaccination Movement: A Cautionary Tale

The anti-vaccination movement (hereafter "the anti-vax movement") is a good example of the danger of disinformation. In 1998, a group of researchers in the UK lead by Dr. Andrew Wakefield published a study implying a causal connection between Measles, Mumps, and Rubella (MMR) vaccinations and the development of autism in children [41]. Though further scrutiny of these initial results disproved the autism-vaccination link - culminating in the withdrawal of endorsements by 10 of the study's 12 authors - the damage had already been done.

The mainstream media picked up on the study, amplifying fears about the safety of vaccinations in an already nervous public. An anti-vaccination movement soon formed, popularized by celebrity activists. Online communities[1] developed, insulating their members against the medical evidence to the contrary. Vaccination rates plummeted despite the best efforts of public health organizations [8].

The result of the spread of the anti-vax movements was that in 2008, for the first time in over a decade, there was a resurgence in the number of reported cases of measles in both the United States [2] and Europe. The situation in the UK was serious enough to be elevated to an endemic [7]. Measles, which in the 1990s was considered a cured disease, was making a comeback.

[1] `http://www.ageofautism.com`

WICOW'09, April 20, 2009, Madrid, Spain.

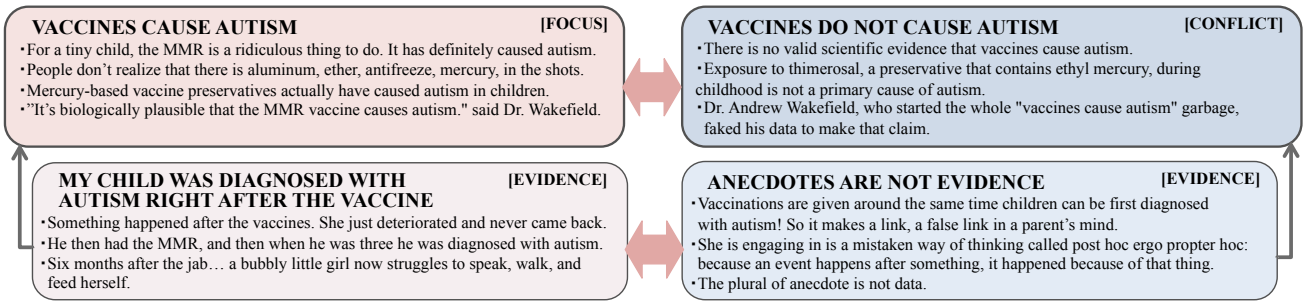

Figure 1: An example STATEMENT MAP for the query "Do vaccines cause autism?"

2. SUPPORTING CREDIBILITY EVALUATION OF ONLINE INFORMATION

The case of the anti-vax movement causing a resurgence in measles is a sad one, but it could have been prevented. Wakefield et al.'s study [41] was repeated numerous times in an attempt to verify the connection between MMR vaccinations and autism, and the results were overwhelmingly against such a causative connection[2]. But this information did not get to the very people most concerned about the safety of vaccinations. Part of the blame belongs with the mainstream media which was more interested in entertaining conspiracy theories than presenting the wealth of evidence disproving a vaccination-autism link, but the underlying problem that people did not know how to find trustworthy evidence to the contrary is illustrative of the importance of evaluating the credibility of information.

Clearly the problem of evaluating information credibility is important, but how can we help users decide what information to trust? One approach taken by several projects is to educate users about how to identify good information online. Services like `http://www.snopes.com` and `http://www.factcheck.org` debunk urban myths and provide fact checking to commonly made political claims. The Quackometer[3] uses language models to identify pseudo-scientific language in webpages. Sense About Science[4] campaigns to educate users about the importance of the scientific method and peer review. Credibility Commons[5] provides tools to help users automatically evaluate the credibility of webpages. Finally, a number of professionals in fields ranging from science and medicine to history and economics share their expert opinions with the public through blogs.

The above projects are all invaluable, but users are often not aware of them, and there may not always be a dedicated resource for a given user's topic of interest. More needs to be done to connect users with the good information out there on the Web. In order to come to an informed opinion, users need to be presented with all of the viewpoints on a topic and the justification or supporting evidence for each one.

Gaining a comprehensive understanding of all possible viewpoints for a topic can be difficult given the large amount of information present on the internet. The bias of search engines to rank pages by popularity makes it more likely that users may not be exposed to all sources of information. Lankes recognizes this problem [17] saying:

> "Consequently, more popular pages are selected and are displayed higher in the search results. Because few people go beyond the first few pages of the search output, however, 'even if a page is of high quality, the page may be completely ignored by Web users simply because its current popularity is very low.' This kind of a system sets up a sort of 'popularity equals credibility' heuristic that could be dangerous or at least disadvantageous to students' learning."

One response to this problem of popularity bias is shown in Reference Extract[6]. Building on research that showed "library websites were seen as more credible than those of museums, governments, and commercial and commercial services,"[7] the project is creating a search engine where ranking gives preference to sources trusted by librarians. This is essentially a *source-based approach* to evaluating credibility.

3. STATEMENT MAP

3.1 Mapping Arguments

The STATEMENT MAP Project instead adopts a *content-based approach* to assisting internet users with evaluating the credibility of online information. Its goal is to present the user with a comprehensive survey of opinions on a topic and show how they relate to each other. We do this by organizing them into groups of agreeing and conflicting opinions and displaying the logical support for each group.

Consider the case of an anxious new parent who is worried about whether vaccines are really safe for his or her child. The top ten results for a simple Google search on "autism" currently[8] contain links to the WikiPedia page on autism, the webpages of several governmental health organizations, including the National Institute of Health, and several charities for autism, but it also includes links to several anti-vax webpages. Telling the difference between these similar-looking but importantly different sources of information can be difficult.

Figure 1 shows the results a similar query - "Do vaccines cause autism?" - would produce with STATEMENT MAP. The

[2]An updating list of studies can be found at `http://en.wikipedia.org/wiki/MMR_vaccine_controversy`
[3]`http://www.quackometer.net/`
[4]`http://www.senseaboutscience.org.uk`
[5]`http://credibilitycommons.org`
[6]`http://referencextract.org`
[7]`http://referencextract.org/?page_id=3&page=3`
[8]As of 2009-02-20

group in the upper-left is labeled FOCUS, and it contains statements that are closest to the user's query. In this case these are opinions that support a causal link between vaccines and autism. An example is the claim "Mercury-based vaccine preservatives actually have caused autism."

The group in the upper-right is labeled CONFLICT, and it contains statements that are in opposition to the statements of focus. This includes the counter-claim "There is no valid scientific evidence that vaccines cause autism."

The thick, red, bi-directional arrows connecting the FOCUS and CONFLICT groups help that opposition in opinion stand out to the user. It is clear that these are strongly opposing opinions. The groups labeled EVIDENCE at the bottom of the figure contain supporting evidence for the FOCUS statements and CONFLICT statements. They are linked by thin, gray, mono-directional arrows.

When the concerned parent in our example looks at this STATEMENT MAP, he or she will see that some opinions support the query "Do vaccines cause autism?" while other opinions do not, but it will also show what support there is for each of these viewpoints.

Ultimately it will be up to him or her to weight the anecdotal evidence of the anti-vaxxers against the medical evidence and logical arguments of the scientific community, but by providing all of the information to the user in a way that makes it easy to see the support or lack thereof for each viewpoint, the STATEMENT MAP Project helps the user come to an informed conclusion.

3.2 Statement Map Generation as NLP Tasks

To generate a STATEMENT MAP, we need to recognize the following 3 kinds of semantic relations between statements on the Web:

- **AGREEMENT** to cluster similar statements
- **CONFLICT** to capture differences of opinion
- **EVIDENCE** to support other statements

In our task, we define a relation AGREEMENT, which becomes true if and only if the relation ENTAILMENT is true between the two statement in both directions at the same time. Identifying logical relations such as AGREEMENT, ENTAILMENT or CONFLICT between statements is the focus of Recognizing Textual Entailment (RTE), the task of deciding whether the meaning of one text is entailed from another text. A major task of the RTE Challenge is identifying ENTAILMENT or CONTRADICTION between Text (t) and Hypothesis (h). Over the last several years, corpora have been constructed for this task, annotated with several thousand (t,h) pairs. In the corpora, each pair was tagged with its related tasks (Information Extraction, Question Answering, Information Retrieval and Summarization). The RTE Challenge has successfully employed a variety of techniques in order to recognize instances of textual entailment, including methods based on: measuring the degree of lexical overlap between bag of words [9, 40], the alignment of graphs created from syntactic or semantic dependencies [20, 18], statistical classifiers which leverage a wide range of features [11], or reference rule generation [38]. These approaches have been successful in recognizing entailment pairs in the corpus, but more robust models of recognizing logical relations are still desirable.

While RTE focuses on only two logical relations, ENTAILMENT and CONTRADICTION, other relations such as EVIDENCE are exempted. There is another task of recognizing relations between sentences, CST (Cross-Document Structure Theory) developed by Radev et. al [30]. CST is an expanded rhetorical structure analysis based on RST [45], and attempts to describe the relations that exist between two or more sentences from different source documents that are related to the same topic, as well as those that come from a single source document. To support this research, the CSTBank corpus [29] was constructed. CSTBank is composed of clusters in which topically related articles are collected. There are 18 kinds of relations in this corpus, not limited to just EQUIVALENCE or CONTRADICTION, but also including JUDGEMENT, ELABORATION, and REFINEMENT. Etoh et al. [6] constructed a Japanese Cross-document Relation Corpus and redefined 14 kinds of semantic relations to fit their corpus.

Relation Type	Relation Label
Logical Relations	Agreement
	Coflict
	Entailment
Attitudinal Relations	Agreeing Opinion
	Conflicting Opinion
	Agreeing Evaluative Polarity
	Conflicting Evaluative Polarity

Table 1: A typology of semantic relations in STATEMENT MAP generation

Zhang et. al [46] attempted to classify CST relations between sentence pairs extracted from topically related documents. However, they used a vector space model and tried multi-class classification. The results were not satisfactory. This observation may indicate that the recognition methods for each relation should be developed separately. Miyabe et al. [24] attempted to recognize relations that were defined in a Japanese cross-document relation corpus [6]. But their target relations included only EQUIVALENCE and TRANSITION; the other relations have not been targeted. Recognizing EVIDENCE is indispensable for STATEMENT MAP. We need to develop robust methods for its identification.

The goal of RTE is to recognize logical and factual relations between sentences in a pair, while CST targets objective expressions because newspaper articles related to the same topic are utilized. Facts, which can be extracted from newspaper articles, have been used in conventional NLP research, such as Information Extraction or Factoid Question Answering. However, there are a lot of opinions on the Web, and it is important to fully survey the opinions related to a user's topic of interest to generate a STATEMENT MAP. The task specifications of both RTE and CST do not cover opinions and their relations as illustrated below.

(1) a. There is absolutely no connection between vaccines and autism.
    b. I do believe that there is a correlation between vaccinations and autism.

Subjective statements, such as opinions, have recently been the focus of various NLP research, such as review analysis, opinion extraction, opinion QA, and sentiment analysis. In the corpus conducted by the MPQA Project (Multi-Perspective Question Answering) [44], individual expressions corresponding to explicit mentions of private states, speech events, and expressive subjective elements are tagged.

The recognition of AGREEMENT, CONFLICT, and EVIDENCE may be sufficient to generate STATEMENT MAPS if

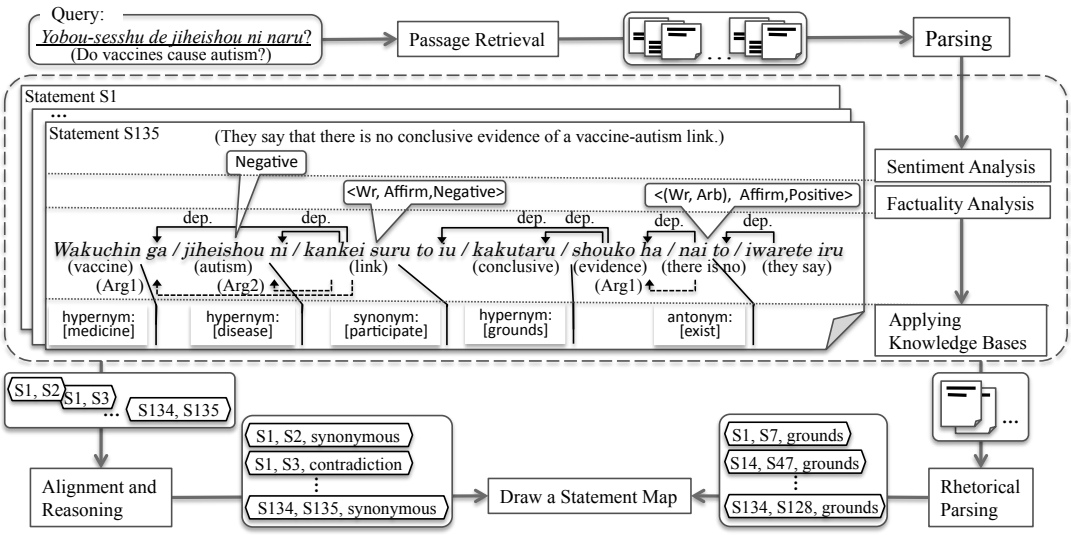

Figure 2: Overview of the Japanese STATEMENT MAP generation system

it were limited to factual statements, however, we also need to determine the source of an opinion and their attitude in order to recognize semantic relations when dealing with opinions. We need to recognize not only logical relations, such as ENTAILMENT and CONFLICT, but also several expanded relations in order to handle opinions. Table 1 shows the semantic relations to capture for STATEMENT MAP generation. To recognize attitudinal relations, we need to combine RTE methodologies, attribution analysis for capturing the source of an opinion, and sentiment analysis to recognize the semantic relations.

4. TECHNICAL ISSUES

4.1 Necessary Technology

As we mentioned in the previous section, we want to identify several kinds of semantic relations between statements, including AGREEMENT, CONFLICT and EVIDENCE, for generating a statement map for a given query sentence.

At first, recognition of AGREEMENT requires framework to recognize entailment and computing similarity between two statements. The recognition of CONFLICT requires a framework for recognizing contradictions as well as one for identifying negative expressions. Furthermore, in order to identify the person that is expressing a private opinion in a statement, his/her attitude, and his/her evaluation to the event in question, it is necessary to perform attribution (or source), modality, and sentiment analysis, respectively.

Recognition of EVIDENCE requires partial rhetorical parsing that identifies conclusion-evidence pairs in a given document. Because the conclusion-evidence pairs found in a single document are more credible than ones that are extracted from multiple documents, we separate the EVIDENCE recognition process from the recognition processes for AGREEMENT and CONFLICT.

4.2 Overview of Statement Map Generation

We are developing systems that generate STATEMENT MAP in Japanese and English. Figure 2 shows an overview of our system for generating STATEMENT MAP in Japanese. This system receives a query sentence and generates a STATEMENT MAP related to it. The algorithm for STATEMENT MAP generation is as follows:

1. Retrieve documents related to a query from the Web
2. Split strings in the documents into sentences
3. Perform tokenization and POS tagging with ChaSen[9]
4. Conduct dependency parsing with CaboCha[10]
5. Carry out predicate-argument structure analysis and anaphora resolution with Syncha[11]
6. Carry out the following two processes concurrently:
   (A) Rhetorical Parsing for each document
   (B) Recognizing semantic relations between each pair of statements
7. Draw a STATEMENT MAP using with the obtained pairs of statements and relations

In order to construct a system for process (A), we plan to make a corpus annotated with EVIDENCE statements and employ machine learning techniques to train EVIDENCE detectors, as proposed in [15, 42, 13].

The above process (B) consists of the following sub-processes:

(a) Sentiment analysis
(b) Factuality analysis that is composed of source identification, and modality and polarity analyses
(c) Applying rules in knowledge bases to statements
(d) Alignment and reasoning

In subsections 4.3, 4.4 and 4.5, we describe processes (a), (b), and (c) respectively in more detail. In subsection 4.6, we propose the construction of a corpus which can be used as training data for process (d). In recent years, various approaches for recognizing textual entailment have been proposed, including methods of aligning syntactic or semantic dependency graphs [10, 19, 4] and frameworks for alignment using supervised machine learning[3, 18]. These approaches have encouraging results and should be directly applicable to our alignment and reasoning task.

[9] http://chasen-legacy.sourceforge.jp/
[10] http://chasen.org/~taku/software/cabocha/
[11] http://sourceforge.jp/projects/syncha/

Resource	# of synset	# of words
WordNet	7,408	29,349
Wikipedia	113,401	307,851
(Types)	101,946	328,534

Table 2: Collected synsets of entities in Japanese.

We describe our system for generating STATEMENT MAP in English in subsection 4.7, but it uses the same basic algorithm as for Japanese.

4.3 Sentiment Analysis

In order to detect strings with implicit sentiment and expressive subjective elements in a given statement, we perform sentiment analysis.

Expressions of emotion, evaluation and reputation, each of which has a sentiment orientation (i.e. positive or negative), have been collected in existing sentiment lexicons such as SentiWordNet [5] for EnglishÂÿ and Kobayashi's sentiment lexicon [16] as well as Higashiyama's sentiment lexicon [12] for Japanese. We extracted and manually checked 5,500 predicates and 13,312 compound nouns from Web documents using the methods in [16] and [12], respectively.

Our sentiment lexicon includes "*zenkai* (complete recovery)" and "*seiseki ga agaru* (raise one's grade)" that represent positive sentiment, and "*byoki* (disease)" and "*koshosuru* (breakdown)" that represent negative sentiment. Our sentiment analyzer detects strings with implicit sentiment and expressive subjective elements using this lexicon, and marks them with their sentiment orientations.

4.4 Factuality Analysis

For each event mention, we want to identify the modal status of the event entity referred to in the event mention. Namely, we want to know whether the event actually took place, will take place, or is just hypothetical.

We have created a new markup scheme for annotating event mentions with factuality information [14, 33]. We annotate each event mention in a given text with a triplet <Source, Modality, Polarity>.

The Source slot represents the person or entity that is expressing the private state in a given text. This attribute conforms to the framework of *nested sources* proposed by Wiebe et al. [44].

The Modality slot specifies the author's mental or communicative attitude toward the event in question. We have divided Japanese modalities into 24 classes based on the following five aspects:

1. whether the statement is a fact, an opinion, a question, or just hypothetical,
2. whether the event has a conditional statement,
3. when the event happened or will happen,
4. whether the source is the agent of the event, and
5. the degree of certainty.

The classes of modality that we defined include:

Affirmation, Inference, Counterfactual, Conditional Affirmative, Deserving, Schedule, Intention, Wish, Request, Hypothesis, etc.

The Polarity slot denotes the polarity status of the event in question. This slot can take one of the values "Positive", "Negative", or "Unknown" depending on the context.

Resource	# of pairs
Web documents	627,791
WordNet	876,861
Category-subcategory in the Wikipedia	642,695
Articles in the Wikipedia	1,492,233
(Types)	3,534,357

Table 3: Total Japanese entity hyponyms collected

With the above markup scheme, a factuality analyzer would produce the triplet <(writer, John), Intend, Positive> for the underlined event mention in "*John wa gakko ni iku tsumori da* (John will go to school)".

We have created a corpus from blog posts and other Web documents. It has 9,111 event mentions manually annotated with <Modality, Polarity> pairs. We plan to annotate these event mentions with Source information as well, and are using this corpus to train a factuality analyzer [14].

4.5 Two Knowledge Bases for Recognizing Logical Relations between Statements

Recognizing logical relations between statements requires a lange amount of knowledge about relations between words of various categores: such as nouns, verbs, adjectives, adverbs and so on. In this subsection, we describe two knowledge bases that can provide inference rules: a knowledge base of relations between entities, and another knowledge base of relations between events.

4.5.1 A Knowledge Base of Relations between Entities

Our knowledge base for entities in Japanese consists of the following two components:

a. 101,946 synsets composed of 328,534 nouns (including compound nouns), and
b. 3,534,357 hyponym relations

We collected synsets of Japanese entities from the following two resources:

1. Synsets from the Japanese WordNet [1]
2. Redirect pages in Wikipedia[12]

We extracted the synsets from the Japanese WordNet that contain more than one word. A redirect page in the Wikipedia includes the page title and its alias expressions, which, together, make up a synset of entities. So, we collected synsets from Wikipedia as well. Table 2 shows the number of synsets collected in our knowledge base.

We collected Japanese hyponyms using the following four methods:

1. Applying syntactic patterns with and without parentheses to Web documents [35]
2. Collecting hyponymy relations from the Japanese WordNet
3. Extracting <category, subcategory> pairs from Wikipedia
4. Extracting candidates from Japanese articles in Wikipedia using an SVM model [36, 37].

Table 3 shows the number of collected hyponymys in our knowledge base.

[12] `http://ja.wikipedia.org/`. We used the Wikipedia archive from 2008/06/07.

4.5.2 A Knowledge Base of Relations between Events

Our knowledge base for events [21], which are represented as predicate-argument structures (PASs) in Japanese, consists of the following two components:

a. a thesaurus that includes 9,582 PASs, and
b. 52,722 binary relations between PASs

The set of binary relations has nine types of logical relations: near synonym, antonym, hypernym, inseparable event, presupposition, effect, goal, co-occurrence, and means, where the latter seven relations suggest some kind of entailment.

4.6 Constrcting a Japanese Corpus for System Evaluation

There are not currently any corpora that focus on semantic relations between both facts and opinions, and there are many challenges in constructing such a corpus. In this section, we describe the specification of the corpus we are constructing and our method of collecting samples from Web.

4.6.1 Characteristics of the Corpus

We focus on only attitudinal and logical relations to construct the corpus because we believe that examples of EVIDENCE should be collected by another paradigm using discourse structures. The structure of an entry in the corpus is represented by the 4-tuple <statement_1, statement_2, entailment_flag, semantic_relation>. The statements in an entry should be collected from different Web documents because the all statements focused on in our task come from real sentences on the Web. Sentences on the Web generally consists of more than one statement or has complex structure. It is difficult to recognize only one semantic relation between a pair of sentences. We attempt to obtain reasonable constituent text segments as statements, and each pair of statements is labeled with one of the semantic relations shown in Table 1 or with "no relation." When a sentence includes several semantic segments, more than one statement can be extracted. So a statement can reflect writer's affirmation in the original sentence. If extracted statement lacks some semantic information, such as pronoun or arguments, human annotators manually add this information. If ENTAILMENT can be recognized between an obtained statement and the original sentence, we tag the pair as in the following example:

(2) a. There is no link between the MMR vaccine and autism, according to the largest ever published study about this controversial issue.
b. There is no link between the MMR vaccine and autism.

This annotation allows our corpus to be used in RTE-like tasks as well.

4.6.2 Corpus Construction Procedure

As a first step, the following procedure is carried out to collect Web documents related to the same topic:

1. Retrieve documents relevant to a specific user query with the search engine TSUBAKI [34]
2. Detect major sub-topic words based on term or document frequency in the document set
3. Extract real sentences including sub-topic words, and eliminate advertisements and extremely similar sentences, using heurisitics to treat them as spam

For example, "regulation", "research" or "possibility" were selected as sub-topics for the query "clone technology".

To prepare a pair of two statements, we consider several options. One option is that we obtain statements from all sentences at first, and then exhaustively pair up extracted statements. However, preparing statements from sentences is too labor-intensive for human annotators, and an arbitrary pair of statements generally does not have the semantic relations we want. We have to select the way to prepare statement pairs that minimizes annotator costs while still obtaining the semantically relevant statements. To do this we pair up sentences which share semantically relevant statements first, and then we extract statements from each sentence in the pair. The relevant statements in a pair of sentences can take the form of identical words, paraphrases, or other similar expressions. We calculate lexical similarity between two sentences to determine whether they can be treated as a pair or not. The calculation method is based on Bag of Words technique, and is closed to BoLI [32]. Our preliminary experiments show that unigrams of KANJI and KATAKANA expressions, single nouns, compound nouns, verbs and adjectives worked well as features.

4.7 Statement Map Generation in English

While the majority of the technical discussion in this paper has focused on issues encountered and resources for handling Japanese, the techniques discussed are also applicable to English. As we showed in Section 3, the information that we show to the user and the semantic relations that need to be identified are not language-dependent.

The algorithm used to generate STATEMENT MAPS is also applicable to English as long as suitable tagger and parsers are used. We are currently carrying out experiments using MXPOST [31] for POS tagging, the MST Parser [22] for dependency parsing, ASSERT [28] for Semantic Role Labeling, and WordNet for recognizing logical relations.

In addition, there are already rich, lexical resources that correspond to many of the resources being constructed for Japanese. WordNet and OpenCyc are broad-coverage, freely-available ontologies that can be used to annotate English dependency graphs with semantic relations. VerbNet, FrameNet and other resources can be used to provide predicate argument structures, much like SynCha does for Japanese.

In order to expand the project to handle English, however, we need appropriate training data to help us build tools capable of detecting statements of focus in web data and the logical relations of interest to users. We do so by exploiting a previously-untapped data source: *scientific blogs*.

4.7.1 Scientific Blogs as a Corpus

Let us return to the example of the anti-vax movement to show the potential of scientific blogs as a corpus. In October of 2008, *Us Magazine* published an interview with celebrity and anti-vax activist Jenny McCarthy claiming she had cured her autistic son by changing his diet [39]. The interview, which offered no evidence to support this claim, angered Phil Plait, a professional astronomer and blogger for the scientific news source *Discover Magazine*.

The author of *Bad Astronomy*, whose other pursuits include debunking the claims of moon landing skeptics and presiding over the James Randi Educational Foundation[13], is not a medical doctor, but as a scientist he has a healthy respect for the scientific process: the verification of testable hypotheses through reporducable experiments. So he wrote

[13] `http://www.randi.org`

an entry at his blog, *Bad Astronomy* [27], critical of the *Us Magazine* piece.

Bad Astronomy's author pointed out that medical doctors have not verified the claimed recovery of Jenny McCarthy's son, and explained the logical fallacy of *post hoc ergo propter hoc*[14] present in both her claims of a vaccination-autism link and her son's cure through a change in diet. He reminded his readers that failure to get their children vaccinated helps spread infectious diseases like measles, and ended with a plea not to believe the anti-vaxxers' groundless claims.

Soon, other members of the scientific blogging community had noticed the *Bad Astronomy* post, and weighted in with their own opinions. One blogger pointed readers to `http://www.stopjenny.com`stopjenny.com, a website dedicated to refuting the arguments of the anti-vax movement and its spokeswoman. An entire discussion about the credibility problems in the *Us Magazine* article was sparked by the blog post at *Bad Astronomy*.

This kind of linked discussion on the same topic, participated in by many members of the blogging community is what makes the construction of our corpus possible. The authors of scientific blogs share a common goal of celebrating good science while tearing down bad science. They seek out examples of bad science (and bad science reporting) in the mainstream media and on the internet and refute them point-by-point, explaining the logical fallacies and other common pitfalls. When bad science appears, it is often surrounded by controversy: global warming denialism, safety concerns regarding the Large Hadron Collider, and alternative medicine are often-addressed topics by science bloggers. Furthermore, the blogs posts are written for a general audience in an informal, easy-to-understand manner, instead of the terse, jargon-laden prose common to scientific publications.

We construct our corpus by forming *discussions* – collections of posts from different blogs discussing and organized around a single topic or article. The structure of the blogs and the interlinking nature of the blogging community facilitate this task. Tags in each blog post make it easy to identify the topic of discussion. Blog posts contain a link to the *source of interest* – the original mainstream media news article, event, or other blog post that inspired authors to respond with their own opinions. Once *discussion*s are formed, we identify *statements of focus* – opinions, facts or justification pertinent to the topic of discussion – and annotate the logical relations between them. We describe methods for automatically expanding the corpus data and identifying candidates for annotation in [25].

5. CONCLUSION AND FUTURE WORKS

In this paper we introduced Statement Map, a project designed to help users navigate the vast amounts of information on the internet and come to informed opinions on topics of interest, and outlined the development of Statement Map generators for Japanese and English. While this project is still in its early stages, we have a clear grasp of the problems that need to be solved and are working to further develop both the Japanese and English resources necessary to make Statement Map a reality.

[14]"after therefore because of:" mistaking precedence for causality

Acknowledgments

This work is supported by the National Institute of Information and Communications Technology Japan. Francisco Dalla Rosa Soares and Hiram Calvo also provided invaluable feedback. Finally, we would like to thank all of the bloggers, especially Ben Goldacre of `http://www.badscience.net` for the data.

6. REFERENCES

[1] F. Bond, H. Isahara, K. Kanzaki, and K. Uchimoto. Boot-strapping a wordnet using multiple existing wordnets. In *Proc. the 6th International Language Resources and Evaluation (LREC'08)*, 2008.

[2] CDC. Update: Measles outbreaks continue in U.S. *Website for Centers for Disease Control and Prevention*, 2008. Available at: `http://www.cdc.gov/Features/MeaslesUpdate/`.

[3] N. Chambers, D. Cer, T. Grenager, D. Hall, C. Kiddon, B. MacCartney, M.-C. de Marneffe, D. Ramage, E. Yeh, and C. D. Manning. Learning alignments and leveraging natural logic. In *ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 165–170, 2007.

[4] M.-C. de Marneffe, T. Grenager, B. MacCartney, D. Cer, D. Romage, C. kiddon, and C. D. Manning. Aligning semantic graphs for textual inference and machine reading. In *Proc. of AAAI Spring Symposium Series: Machine Reading*, 2007.

[5] A. Esuli and F. Sebastiani. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC-06)*, pages 417–422, 2006.

[6] J. Etoh and M. Okumura. Cross-document relationship between sentences corpus. In *Proc. of the 14th Annual Meeting of the Association for Natural Language Processing*, pages 482–485, 2005. (in Japanese).

[7] Eurosurvelliance. Measels once again endemic in the United Kingdom. *Eurosurvelliance*, 13(27), 2008. Available at: `http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18919`.

[8] Finding Dulcenia. European health officials cope with measles outbreaks, lower vaccination rates. *Finding Dulcenia: Librarian of the Internet*, 2009. Available at: `http://www.findingdulcinea.com/news/health/2009/jan/European-Health-Officials-Cope-With-Measles-Outbreaks-Lower-Vaccination-Rates.html`.

[9] O. Glickman, I. Dagan, and M. Koppel. Web based textual entailment. In *Proc. of the First PASCAL Recognizing Textual Entailment Workshop*, 2005.

[10] A. Haghighi, A. Ng, and C. Manning. Robust textual inference via graph matching. In *Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 387–394, 2005.

[11] A. Hickl, J. Williams, J. Bensley, K. R. B. Rink, and Y. Shi. Recognizing textual entailment with lcc's groundhog system. In *Proc. of the Second PASCAL Challenges Workshop*, 2005.

[12] M. Higashiyama. *Acquiring Noun Polarity Knowledge Using Selectional Preferences*. MASTER Thesis, 2008. (in Japanese).

[13] R. Iida, K. Inui, and Y. Matsumoto. The task definition of evidence-conclusion relation extraction and its preliminary empirical evaluation. In *Proc. of the 15th Annual Meeting of the Association for Natural Language Processing*, 2009. (in Japanese).

[14] K. Inui, S. Abe, H. Morita, M. Eguchi, A. Sumida, C. Sao, K. Hara, K. Murakami, and S. Matsuyoshi. Experience mining: Building a large-scale database of personal experiences and opinions from web documents. In *Proc. of*

the 2008 IEEE/WIC/ACM International Conference on Web Intelligence, pages 314–321, 2008.

[15] T. Inui, K. Inui, and Y. Matsumoto. Acquiring causal knowledge from text using the connective marker tame. *ACM Transactions on Asian Language Information Processing*, 4(4):435–474, 2005.

[16] N. Kobayashi, K. Inui, and Y. Matsumoto. Opinion mining from web documents: Extraction and structurization. *Journal of the Japanese Society for Artificial Intelligence*, 22(2):227–238, 2007.

[17] D. R. Lankes. Trusting the internet: New approaches to credibility tools. In M. J. Metzger and Andrew, editors, *Digital Media, Youth, and Credibility*, pages 101–122. MIT Press, 2008.

[18] B. MacCartney, M. Galley, and C. D. Manning. A phrase-based alignment model for natural language inference. In *Proc. of 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP-08)*, pages 802–811, 2008.

[19] B. MacCartney, T. Grenager, M.-C. de Marneffe, D. Cer, and C. D. Manning. Learning to recognize features of valid textual entailments. In *Proc. of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2006)*, 2006.

[20] E. Marsi and E. Krahmer. Classification of semantic relations by humans and machines. In *In ACL-05 Workshop on Empirical Modeling of Semantic Equivalence and Entailment*, pages 1–6, 2005.

[21] S. Matsuyoshi, K. Murakami, Y. Matsumoto, , and K. Inui. A database of relations between predicate argument structures for recognizing textual entailment and contradiction. In *Proc. of the 2nd International Symposium on Universal Communication (ISUC2008)*, pages 366–373, 2008.

[22] R. McDonald, K. Crammer, and F. Pereira. Online large-margin training of dependency parsers. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 91–98, Morristown, NJ, USA, 2005. Association for Computational Linguistics.

[23] M. J. Metzger. Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13):2078–2091, 2007.

[24] Y. Miyabe, H. Takamura, and M. Okumura. Identifying cross-document relations between sentences. In *In Proc. of the 3rd International Joint Conference on Natural Language Processing (IJCNLP-08)*, pages 141–148, 2008.

[25] E. Nichols, K. Murakami, K. Inui, and Y. Matsumoto. Constructing a scientific blog corpus for information credibility analysis. In *Proc. of the 15th Annual Meeting of the Association for Natural Language Processing*, 2009.

[26] Pew Research. Internet overtakes newspapers as news outlet. *Website for the Pew Research Center for the People & the Press*, 2008. Available at: `http://people-press.org/report/479/internet-overtakes-newspapers-as-news-source`.

[27] P. Plait. âĂębut how do we recover from Jenny McCarthy? *Bad Astronomy*, 2008. Available at: `http://blogs.discovermagazine.com/bad astronomy/2008/10/20/but-how-do-we-recover-from-jenny-mccarthy/`.

[28] S. Pradhan, W. Ward, K. Hacioglu, J. Martin, and D. Jurafsky. Shallow semantic parsing using support vector machines. In *Proceedings of the Human Language Technology Conference/North American chapter of the Association of Computational Linguistics (HLT/NAACL)*, Boston, MA, 2004.

[29] D. Radev, J. Otterbacher, and Z. Zhang. CSTBank: Cross-document Structure Theory Bank. http://tangra.si.umich.edu/clair/CSTBank, 2003.

[30] D. R. Radev. Common theory of information fusion from multiple text sources step one: Cross-document structure. In *Proc. the 1st SIGdial workshop on Discourse and dialogue*, pages 74–83, 2000.

[31] A. Ratnaparkhi. A maximum entropy model for part-of-speech tagging. In E. Brill and K. Church, editors, *Proceedings of the Empirical Methods in Natural Language Processing*, pages 133–142, 1996.

[32] D. Roth and M. Sammons. Semantic and logical inference model for textual entailment. In *Proc. of the Third PASCAL Recognizing Textual Entailment Workshop*, 2007.

[33] C. Sao, M. Eguchi, S. Matsuyoshi, and K. Inui. An annotation scheme for capturing modality and polarity of events in japanese text. In *Proc. of the 15th Annual Meeting of the Association for Natural Language Processing*, 2009. (in Japanese).

[34] K. Shinzato, T. Shibata, D. Kawahara, C. Hashimoto, and S. Kurohashi. Tsubaki: An open search engine infrastructure for developing new information access methodology. In *Proc. the 3rd International Joint Conference on Natural Language Processing (IJCNLP2008)*, pages 189–196, 2008.

[35] A. Sumida, K. Torisawa, and K. Shinzato. Concept-instance relation extraction from simple noun sequences using a search engine on a web repository. In *Proc. the Web Content Mining with Human Language Technologies workshop on the 5th International Semantic Web*, 2006.

[36] A. Sumida, N. Yoshinaga, and K. Torisawa. Boosting precision and recall of hyponymy relation acquisition from hierarchical layouts in wikipedia. In *Proc. the 6th International Language Resources and Evaluation (LREC'08)*, 2008.

[37] A. Sumida, N. Yoshinaga, K. Torisawa, and K. Mannari. Acquiring a large number of hyponymy relations from the wikipedia. In *Proc. of the 14th Annual Meeting of the Association for Natural Language Processing*, pages 769–772, 2008. (in Japanese).

[38] I. Szpektor, E. Shnarch, and I. Dagan. Instance-based evaluation of entailment rule acquisition. In *Proc. of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 456–463, 2007.

[39] Us Magazine. Jenny McCarthy: My son no longer has autism. *Website for Us Magazine*, 2008. Available at: `http://www.usmagazine.com/news/jenny-mccarthy-my-son-is-no-longer-autistic/`.

[40] J. V. and M. de Rijke. Recognizing textual entailment using lexical similarity. In *Proc. of the First PASCAL Challenges Workshop*, 2005.

[41] A. J. Wakefield, S. H. Murch, A. Anthony, J. Linnell, D. M. Casson, M. Malik, M. Berelowitz, A. P. Dhillon, M. A. Thomson, P. Harvey, A. Valentine, S. E. Davies, and J. A. Walker-Smith. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*, 351(9103), 1998.

[42] B. Wellner and J. Pustejovsky. Automatically identifying the arguments of discourse connectives. In *Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 92–101, 2007.

[43] What Japan Thinks. Checking internet news in Japan. *Website for What Japan Thinks*, 2008. Available at: `http://whatjapanthinks.com/2008/07/13/checking-internet-news-in-japan/#more-1225`.

[44] J. Wiebe, T. Wilson, and C. Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210, 2005.

[45] M. William and S. Thompson. Rhetorical structure theory: towards a functional theory of text organization. *Text*, 8(3):243–281, 1988.

[46] Z. Zhang and D. Radev. Combining labeled and unlabeled data for learning cross-document structural relationships. In *Proc. the Proceedings of IJC-NLP*, 2004.