May 2014

A Computational Linguistic Approach towards Understanding Wikipedia's Article for Deletion (AfD) Discussions

Wanting Mao
The University of Western Ontario

Supervisor
Dr. Robert E. Mercer
The University of Western Ontario

Joint Supervisor
Dr. Lu Xiao
The University of Western Ontario

Graduate Program in Computer Science

A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

A COMPUTATIONAL LINGUISTIC APPROACH TOWARDS UNDERSTANDING WIKIPEDIA'S ARTICLE FOR DELETION (AFD) DISCUSSIONS

(Thesis format: Monograph)

by

Wanting Mao

Graduate Program in Computer Science

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

With the thriving of online deliberation, Wikipedia's Article for Deletion (AfD) discussion has drawn a number of researchers' attention in the past decade. In this thesis we aim to solve two main problems: 1) how to help new users effectively participate in the discussion; and 2) how to make it efficient for administrators to make decision based on the discussion. To solve the first problem, we obtain a knowledge repository for new users by recognizing imperatives. We propose a method to detect imperatives based on syntactic analysis of the texts. And the result shows a good precision and reasonable recall. To solve the second problem, we propose a decision making support system that provides administrators with an reorganized overview of a discussion. We first divide the arguments in the discussion into several groups based on similarity; then further divide each group into subgroups based on sentiment (positive, neutral and negative). In order to classify sentiment polarity, we propose a recursive algorithm based on the dependency structure of the text. Comparing with the state of the art sentiment analysis tool by Stanford, our algorithm shows a promising result of 3-categories classification without requiring a large training dataset.

Keywords: natural language processing, speech act, sentiment analysis, knowledge management, decision making

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

Wikipedia, a free Internet encyclopedia, has become influential worldwide since 2001. A large group of volunteers collaboratively participate in building the encyclopedia, including the process of creation, editing, deletion, etc. To ensure the quality of the encyclopedia, deletion of articles happens all the time.

There are three main methods in the process of deletion. An article might qualify for "speedy deletion" if it is clearly inappropriate and meets certain specific criteria, and "proposed deletion" may be applied when an article is considered as uncontroversially non-encyclopedic. If it is a controversial article, a discussion called "Article for Deletion"(AfD) is held to determine whether this article should be deleted.

This thesis focuses on AfD since it is the most deliberative among the three main deletion methods. It is open to any user to join in the discussion and make a comment. One example of discussion in AfD is given in Figure 1.1. In each discussion, the user who nominates the AfD gives his argument for deletion of the article. Then other users can read and vote "Keep", "Delete", "Merge", etc. After their votes, usually they should justify their votes by elaborating their arguments. Although their vote seems to be helpful, it cannot be the determining factor in deciding whether the article should be deleted. The administrator has to review the discussion and make a decision according to the arguments in the deletion discussion.

One problem is that newcomers join in the discussion continuously. And it is not surprising that they make the same mistakes as other newcomers. That is, certain mis-

The following discussion is an archived debate of the proposed deletion of the article below. ***Please do not modify it.*** *Subsequent comments should be made on the appropriate discussion page (such as the article's talk page or in a deletion review). No further edits should be made to this page.*

The result was **speedy delete** per A7 by Jimfbleak. (non-admin closure) ★☆ DUCKISJAMMMY☆★ 11:05, 16 January 2013 (UTC)

Interpersonal wellness [edit]

Interpersonal wellness (edit | talk | history | links | watch | logs | views) – (View AfD · Stats)
(*Find sources:* "Interpersonal wellness" – books · scholar · JSTOR · free images)

Fails Notability guideline Veggies (*talk*) 22:58, 15 January 2013 (UTC)

- **Delete** I agree with Veggies. This is a neologism invented by the article's author. I PROD'ed the article simultaneously with Veggie's AFD nomination; I think the deletion is non-controversial enough to not need an AFD. WikiDan61 ChatMe! ReadMe!! 23:01, 15 January 2013 (UTC)
- **Delete** I concur. Not much that hasn't already been said here. Bagheera (talk) 23:33, 15 January 2013 (UTC)
- **Comment:** I did a Google Books search to determine if "interpersonal wellness" is used in any contexts independent of that described in the article. It appears that the concept has been discussed in a few WP:RS, especially in the *The Praeger Handbook Of Education And Psychology, Volume 1, pg. 342.*. The article as currently written clearly is not up to Wikipedia standards (e.g. WP:OR), but a completely re-written article with appropriate sources may have some WP:POTENTIAL. --Mike Agricola (talk) 00:11, 16 January 2013 (UTC)
- **Delete**. It's not clear that "interpersonal wellness" is anything more than a vague neologism; I don't see a compelling source that establishes this as a technical term with a set meaning. Worse, I found a source using this term from 1994, long before the origin claimed in the entry, and I can find no reliable source that mentions Joyce Odidison in connection with this term (for example, she is not mentioned in the Praeger Handbook reference given above). Hairhorn (talk) 04:18, 16 January 2013 (UTC)

The above discussion is preserved as an archive of the debate. ***Please do not modify it.*** *Subsequent comments should be made on the appropriate discussion page (such as the article's talk page or in a deletion review). No further edits should be made to this page.*

Figure 1.1: An example of discussion in AfD

takes are easily made by novices during the discussion. If we can find out the common mistakes from the past discussions and provide them to newcomers, they may understand policies and guidelines better, and hopefully avoid these common mistakes. However, it is challenging because of the large amount of information.

Another problem involves the administrator making the decision. Everyday, 50 to 100 new discussions appear and a long discussion can be pages long. The process of reviewing the textual discussion requires much time and effort, thus how to reduce the workload for an administrator has drawn our attention.

In this thesis, we will analyze existing problems in the process of deletion discussion from two perspectives: knowledge management and decision making. In terms of knowledge management, we will analyze the discussions and propose feasible methods to help educate new users and prevent them from making certain mistakes. As for decision making, we will propose a decision making support system that makes it easier for

administrators to review the discussions so that they can make rational decision more efficiently. In this thesis, we try to solve these problems from a linguistic perspective by using natural language processing (NLP) techniques.

The rest of this thesis is organized as follows. In Chapter 2, we will discuss related research that has been done on Wikipedia AfD and some natural language processing techniques and tools. Knowledge management involving helping new users will be discussed in Chapter 3. And how to help administrators with decision making will be presented in Chapter 4. Finally, Chapter 5 will conclude our work in this thesis and will discuss future work.

Chapter 2

Literature Review

2.1 Deletion Discussion in Wikipedia

There is a lot of research that has been done on Wikipedia's Articles for Deletion (AfD). In [29], Schneider et al. studied decision factors in deletion discussions in Wikipedia. They first identified the factors that impact the decision about whether to delete a certain article. Then, they analyzed the importance of these factors. There are four factors determining a decision for deletion: notability, sources, maintenance and bias. Notability is the most decisive factor, while bias could never close a debate by itself. Although sources and notability are distinct, they are closely related, since all notability is supported by sources. Maintenance often leads to a deletion discussion. These factors can be used in both 'keep' and 'delete' arguments, as listed in Table 2.1.

In their study, they point out that sometimes when we fail to gain sufficient discussion or the article has been changed during the discussion period, we are in a position of no consensus. There are conflicts around consensus values. Novices seem confused and very emotional, so there is little constructive engagement between new comers and experts. It also shows that the Wikipedia policy itself comes under attack in the discussion. Since Wikipedia policies are complex, it is difficult for novices to understand them.

Schneider et al. investigated the difference in arguments from novices and experienced users[30]. Some common problematic arguments such as personal preference and requesting a favor often happens in novices' arguments since they often bring their per-

factors	Example (used to justify 'keep')	Example (used to justify 'delete')
Notability	A quick search shows that the term is clearly notable.	The article is a disaster and the person is of indeterminable notability.
Sources	I believe this is a real thing, and this is the term used by reliable sources.	I could find no significant coverage of his work in reliable sources.
Maintenance	Yes it needs improvement and I have made a start on referencing the article.	No need for this as well as it's pretty much a duplicate.
Bias	It is by no means spam (it does not promote the products).	I thought for sure this was going to be a well-meaning college student trying to turn a term paper into an encyclopedia article.

Table 2.1: Four main factors used in arguments for keeping and deleting an article

sonal emotions to the discussion and may be confused about the policy of deletion. For example, they may make a personal attack such as: "your comment was thick headed." They also found that 'no consensus' discussions appeared in these two cases: lack of sufficient discussion, and involvement with novice nominator. One severe problem novices have is that they fail to provide significant and reliable sources to justify their opinions. On the contrary, experts are more familiar with policies and guidelines. They use policies effectively to justify their arguments. However, it is found that some experts also make strategic errors during discussion and they try to challenge the policies themselves.

Xiao and Askin have done a study on the factors that influence online deliberation[40]. They examined the types of rationales in Wikipedia AfD discussions. They identified nine types of rationales in the discussions with the three major types being notability, credibility, and policy. They also found that non-unanimous discussions draw more users to join in the discussion.

2.2 Natural Language Processing

2.2.1 Speech Act

A Speech Act is a performative utterance in communication. It was originally proposed by J. L. Austin in the 1950s[1]. While many sentences hold a truth-value (or proposition) as part of the utterance, there are some sentences that are considered as neither true nor false. Austin proposed that all sentences have speech acts. A speaker might be performing any of three acts when speaking: a locutionary act, an illocutionary act, and a perlocutionary act. A locutionary act is the act of uttering words, phrases or clauses that conveys literal meaning by means of a lexicon, syntax and phonology. An illocutionary act is the act of expressing the speaker's intention. A perlocutionary act is the act performed by or resulting from saying something; it is the consequence of, or the change brought about by the utterance.

For example, my saying to you "Follow the guideline." (a locutionary act) counts as commanding you to follow the guideline (an illocutionary act), and if you obey my command I have thereby succeeded in persuading you to follow the guideline (a perlocutionary act).

Illocutionary act

Among the three acts, the illocutionary act is the most central part in speech act theory. The American philosopher-linguist John Searle classified illocutionary acts into five groups [31].

Representatives: stating or describing, saying what the speaker believes to be true.

- e.g.: I have never seen the man before.

Directives: trying to get the hearer to do something.

- e.g.: Open the window!

Commissives: committing the speaker himself to some future course of action.

- e.g.: I promise to come.

Expressives: expressing feelings or attitude towards an existing state.

- e.g.: I'm sorry for the mess I have made.

Declarations: bringing about immediate changes by saying something.

- e.g.: I now declare that the meeting begins.

Directives

In this thesis, we focus on directives since the issuer of the directive expresses his/her desire to make another participant in the discussion to do something by using a directive. It is a kind of communication between participants in the deletion discussion. And they often lead to some action. For example, once one participant says, "Could you provide the source to me?", another participant responds with some sources. Here we can see how effectively the directive works through the discussion.

The form of directives varies according to the context. S. Ervin-Tripp's work [11], lists six types of directives (Table 2.2).

Need statements	I need a match.
Imperatives	Give me a match.
Imbedded imperatives	Could you give me a match?
Permission directives	May I have a match?
Question directives	Got a match?
Hints	The matches are all gone.

Table 2.2: Six types of directives

When a "need" statement occurs between people of different ranks, the speaker usually is the superior of the hearer (e.g., a doctor to a nurse). Other cases of "need" statements occurs between family members. For example, a four year old boy says to his mother, "I need a toy car, mommy".

Imperatives express a command, that is, a request that asks someone to do or not to do something. In most cases, the predicate in imperatives is an action verb and the subject is second-person (you). The subject is typically eliminated. As well, there can be words of politeness and adverbial modifiers of the verb:

- Please do this sort of check in the future.
- Just avoid those sorts of comments and perhaps strike the one above.

Imperatives can also be used to express a suggestion, an invitation, a wish, an apology, etc.:

- Let's have dinner together. (suggestion)
- Come in and have a seat. (invitation)
- Have a good vacation! (wish)
- Pardon me (apology)

Cohortatives (first person plural imperatives) are normally used in suggestions as we can see from the above example.

Embedded imperatives are developed from imperatives with a kind of formal addition. Examples are:

- Can you open the window?
- Would you mind opening the window?

In Sinclair et al.'s work [32], they proposed a modal directive rule:

An interrogative clause can be interpreted as a command if: a) it contains a modal verb such as can, will, could; b) the subject is also an addressee; c) the predicate describes an action which can be done at the time of utterance.

Some sentences are ambiguous such as "can you swim?". If this utterance occurs in the classroom, it would be interpreted as a general question with an 'yes' or 'no' answer. Whereas if it occurs by a swimming pool, then it would be interpreted as a command to jump into the pool and swim. Context substantially determines whether these sentences are commands depending on the feasibility of the demanding action.

Permission directives have the following structure: modal + beneficiary + verb + ? Examples are:

- May I have the salt?
- Can I talk to Mary?

Question directives and hints are implicit. In these two cases, speakers usually don't express their intentions directly. Thus, it is easy for listeners to ignore them.

Han proposed the logical form of imperatives [14], which contains directive force and unrealized interpretation. It can be represented as $directive(irrealis(p))$. Assume the hearer has a plan set. The utterance expressed by the speaker to direct the hearer to add a plan p to the plan set is an imperative.

One type of utterance looks like an imperative but actually it is not. It is called an imperative-like construction. E.g.: "Miss this train, and you will be late." The speaker does not intend to make the hearer miss the train. Instead, "miss this train" is used as a conditional function in this example. The difference between imperatives and imperative-like constructions is that imperatives contain both directive and irrealis feature, whereas imperative-like constructions lack direct (illocutionary) force. In Han's work, he also pointed out that imperative-like constructions allow only second person subjects. Nevertheless, imperatives also allow first and third person subjects (e.g., let's do it; let her go).

2.2.2 Rhetorical Structure Theory

Rhetorical Structure Theory (RST) is a theory for describing the organization of natural text and identifying the relationship between text spans. It was originally proposed by Mann and Thompson in 1987 [21]. Ever since then, RST has caught extensive attention in diverse fields, especially in the linguistics area, including text generation, discourse analysis and cross-linguistic studies. However, it is not easy to automatically generate the rhetorical structure of texts due to ambiguity and complexity. This is evidenced in the following two segments:

A. I love to collect classic automobiles. My favorite car is my 1899 Duryea.

B. I love to collect classic automobiles. My favorite car is my 2010 Toyota.

Obviously, segment A makes sense. It presents the fact that the author loves his 1899 Duryea, preceded by the fact that he loves classic automobiles. Nevertheless, readers

may not understand the author's intention in segment B. If we add "however" between the two sentences in segment B as below:

- I love to collect classic automobiles. However, my favorite car is my 2010 Toyota.

It would be clear to readers that although the author loves to collect classic automobiles, his favorite car is his 2010 Toyota which is apparently not a classic car. For both segments, there exist some relation between the two sentences. In segment A, it is easy to observe the implicit relation—Elaboration between the two sentences, while in segment B, the Contrast relation is relatively hard to detect. In the original RST paper [21], Mann and Thompson defined a set of 23 relations, including Elaboration and Contrast.

Hierarchy in Rhetorical Structure

The first step to analyze the text is dividing it into primary units. Units are called spans as well. They can be clauses, sentences, etc. Then one span will be connected to another by adding a particular relation between them. Thus we will have a new span composed of the two primary units. By doing it recursively, a tree structure with one top-level relation will be formed. To illustrate how it works, here is an example. We have a paragraph, which has been divided into three units as follows:

(1) I love to collect classic automobiles.
(2) My favorite car is my 1899 Duryea.
(3) However, I prefer to drive my 2010 Toyota.

After dividing it into 3 units, we connect (1) and (2) with the relation elaboration. After that, we treat (1)(2) as a new text span and we connect it with (3) and then assign a Contrast relation between them. Thus, a complete tree structure has been built as shown in Figure 2.1.

Multiplicity

People used to regard a construct as unambiguous, however, it often happens that a text can be analyzed in several ways due to the way that RST is defined. Giving the same

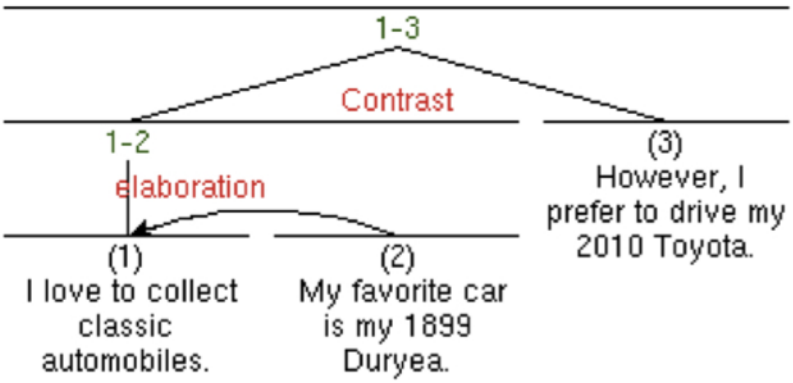


Figure 2.1: Hierarchy in a rhetorical structure tree

text to several analysts, they often come up with different analyses. It is even possible that one analyst gives two or more analyses for the same text. Five kinds of multiplicity are listed below:

- Boundary judgements
- Text structure ambiguity
- Simultaneous analyses
- Difference between analysts
- Analytical error

Tasks and solutions

As we have discussed, the multiplicity of possibilities in RST brings difficulty in analyzing text. Given that problem, it is even more difficult for computers to analyze text automatically. There are three main tasks in RST: 1. Determining the elementary units; 2. Building discourse structure; 3. Defining the relations that hold between parts of text. To address these tasks, many researchers have devoted themselves to them. Marcu proposed a surface-based algorithm to identify the discourse markers and elementary units in one sentence at a time[22]. Taboada studied how often a rhetorical relation is signaled by a discourse marker on both task-oriented dialogues and newspaper articles [37]. Soricut and Marcu proposed a sentence level discourse parsing system concerning both discourse segmentation and discourse tree building [35]. Later Sporleder and Lascarides

used machine learning methods to predict high-level discourse structure [36]. Bach et al. proposed the current state-of-the-art discourse parser–unlabeled discourse parsing system in the RST framework [2].

2.2.3 Relationships between text fragments

Text fragments are usually bound to each other in a well-structured article or online deliberation such as a discussion in AfD. Recognizing these relationships between texts (e.g., sentence to sentence, paragraph to paragraph) would help to better understand the context. In this section, we will first give an introduction to textual entailment and then review the related work in similarity between texts.

Textual entailment

Textual entailment (TE) represents a directed relationship between two text fragments. The two fragments are called text (T) and hypothesis (H). If the reader can infer H by reading T, we would say T entails H ($T \Rightarrow H$). The directional relation cannot be reversed, that is, when T entails H, the reverse relation $H \Rightarrow T$ does not always hold. More specifically, three different relations in textual entailment are listed below:

Positive entailment:

T: If you had checked Google Scholar, you would see that the top result has 13,311 citations.

H: Very high cites in GS.

Negative entailment (contradiction):

T: Redirecting the page to the lead actors' future projects section will be cool.

H: I don't think it is wise to redirect to the original film.

Non-entailment

T: Sources say that this film is under production; what happens if it is cancelled?

H: "Cancelled" is a whole different issue, and would likely prevent any article recreation.

Cabrio and Villata combined textual entailment and argumentation theory to generate a framework for supporting online debates[5]. A common problem for online debates is that a participant who wants to participate in the middle of a debate may have difficulties in reviewing the past discussion and identifying the accepted arguments. In their work, they aimed to help the participants better understand the ongoing debates. By using textual entailment, we can detect supporting arguments (positive entailment) and attacking arguments (negative entailment). Given a set of arguments and the attacks among them, an argumentation framework (AF) can be used to detect the arguments being accepted. In particular, an argument is accepted if all the attacks of it are rejected or it is not attacked. An argument is rejected if at least one argument that attacks it is accepted. An example of a simple argument framework is given in Figure 2.2. Plain arrows stand for attacking and dashed arrows stand for supporting. Double bordered items (A1, A2, A3) are the accepted arguments since the only argument A3 attacking A1 is rejected by A4 and A4 and A2 are not rejected (attacked).

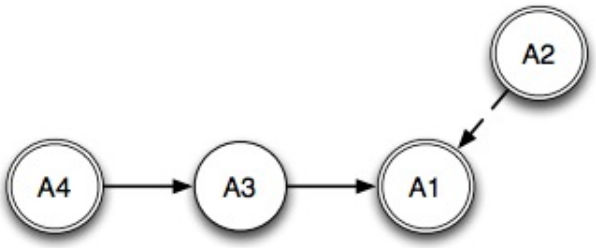


Figure 2.2: An example of an argument framework (AF) [5]

The experiment in Cabrio and Villata's work was conducted by using Debatepedia (an encyclopedia of debates). They randomly selected 200 T-H pairs, 100 for training and 100 for testing. Each data set consists of 55 entailment and 45 contradiction pairs. To recognize the TE relation in each T-H pair, they used EDITS (Edit Distance Textual Entailment Suite) [18].

They first evaluate the performance of EDITS by assigning the entailment relations to each T-H pair. As shown in Table 2.3, EDITS provides an accuracy of 0.69 on the training set and 0.67 on the test set. The baseline for comparison is based on a word overlap algorithm which gives an accuracy of 0.61 and 0.62 on the training and the test

		Train			Test		
	rel	Pr.	Rec.	Acc.	Pr.	Rec.	Acc.
EDITS	yes	0.71	0.73	**0.69**	0.69	0.72	**0.67**
	no	0.66	0.64		0.64	0.6	
WordOverl.	yes	0.64	0.65	**0.61**	0.64	0.67	**0.62**
	no	0.56	0.55		0.58	0.55	

Table 2.3: TE system performances on the Debatepedia data set (precision, recall and accuracy) [5]

set, respectively.

They also assessed how EDITS performs on argument acceptability. Comparing the accepted arguments in the correct argumentation frameworks with the ones in the frameworks generated basing on EDITS, they obtained a precision of 0.74, a recall of 0.76 and an accuracy of 0.75. Although EDITS makes mistakes recognizing textual entailment relations, the result of their combined approach is still promising.

Text similarity

Text similarity can be interpreted as similarity between sentences, paragraphs, documents, etc. It has been used in various aspects in NLP such as information retrieval, text classification, and automatic evaluation. The most fundamental part is word similarity. We consider words to be similar in the following conditions:

- Synonyms
- Antonyms
- Similar concept (red, green)
- Similar context (doctor, hospital)
- Hyponym/hypernym relation (dog, pet)

WordNet, a word-to-word similarity library, was developed by Pedersen et al. [24] and has been widely used to compute the similarity at a coarser granularity (e.g., sentence-to-sentence similarity). Various methods to deal with text similarity have been proposed over the past decades.

The most fundamental method to assess the similarity between texts is based on lexical overlap. It is not just as simple as calculating the common words in two text fragments, but a number of parameters are utilized to compute the similarity score. Specifically, first we would remove punctuation and stopwords from the texts; next we may want to keep only some specific words or keep all words; then we may use a weighting scheme. All these parameters are adjustable according to the texts we are analyzing.

A greedy method is proposed by Mihalcea et al. [23] They used the equation below to calculate the similarity score between two text fragments (i.e., T1 and T2). For each word in T1 (T2), the maximum similarity score to any word in T2 (T1) is used. The WordNet similarity we have mentioned previously can be used for assigning similarity score between every pair of words in the two texts.

$$sim(T1,T2) = \frac{1}{2}(\frac{\sum_{w\in T1} max\{sim(w,T2) * idf(w)\}}{\sum_{w\in T1} idf(w)} + \frac{\sum_{w\in T2} max\{sim(w,T1) * idf(w)\}}{\sum_{w\in T2} idf(w)}) \tag{2.1}$$

Accordingly, Rus and Lintean proposed an optimal method to compute text similarity based on word-to-word similarity[26]. It is similar to an optimal assignment problem. Given a weighted complete bipartite graph ($G = X \bigcup Y; X \times Y$), with weight $w(xy)$ on edge xy, we need to find a matching from X to Y with a maximum total weight. Their results showed that the optimal method outperformed the greedy method in terms of accuracy and kappa statistics.

Latent dirichlet allocation (LDA) is a probabilistic generative model proposed by Blei et al. [4] The basic idea of LDA is that documents are represented as distributions of underlying topics and topics are represented as distributions of words. Rus et al. came up with a semantic similarity measure based on LDA [28]. The word-to-word similarity measure is defined as follows ($\varphi_t(w)$ represents the probability of word w in topic t):

$$LDA_w2w(w,v) = \sum_{t=1}^{T} \varphi_t(w)\varphi_t(v) \tag{2.2}$$

Since documents are represented as distributions of topics, text-to-text similarity should be computed based on the similarity of these distributions (for all the details see [28]).

First Word	Second Word	Third Word (Not Extracted)
JJ	NN or NNS	anything
RB, RBR, or RBS	JJ	not NN nor NNS
JJ	JJ	not NN nor NNS
RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

Table 2.4: Patterns of tags for extracting two-word phrases from reviews [38].

2.2.4 Sentiment analysis

The rise of sentiment analysis has paralleled a similar interest in social media and e-commerce. As e-commerce grows rapidly, there can be thousands of reviews for some popular products. It becomes difficult for a potential customer to read them and make a decision regarding which to buy. Also, it's difficult for a manufacturer to keep track of and manage customer opinions.

Basically, sentiment analysis is meant to determine the polarity of a certain text, which can be positive, negative and neutral. Researchers and companies have been extensively studying sentiment analysis over the last decade. Most of the early work is aimed at analyzing the polarity of customer reviews (e.g., product reviews, restaurant reviews). Apart from business, sentiment analysis has also been applied in other domains such as politics and sociology. By analyzing the large amount of information from social networks like Facebook and Twitter, politicians may have an overview of the public's opinions.

An early work by Turney classifies reviews as recommended or not recommended [38]. First, phrases are extracted using the patterns of part of speech shown in Table 2.4. Then, the semantic orientation (SO) of the phrases is estimated. Pointwise Mutual Information (PMI) was used in calculating the semantic orientation of a phrase. If the average of the semantic orientation of the phrases in a review is positive, then the review is classified as recommended; if negative, then not recommended. This method was tested on reviews of automobiles, banks, movies and travel destinations. The results indicate that the accuracy on reviews of movies is obviously lower than the others.

To illustrate why movie reviews are difficult to classify, a sample phrase from movie

Movie:	The Matrix
Author's Rating:	recommended (5 stars)
Average SO:	-0.219 (not recommended)
Sample Phrase:	more evil [RBR JJ]
SO of Sample Phrase:	-4.384
Context of Sample Phrase:	The slow, methodical way he spoke. I loved it! It made him seem more arrogant and even **more evil**.

Table 2.5: A sample phrase from movie review [38]

review is given in Table 2.5. The phrase "more evil" has a strong negative orientation. It describes a successful character in this movie, but it will not make it a bad one. A good movie often contains bad roles or unpleasant scenes. Thus the two factors involved in a movie review make it difficult to classify: the elements of the movie including roles and events; the other one is the whole movie such as quality and style.

Kim and Hovy presented a system detecting opinion holders and the sentiment of the opinion [17]. The system architecture is shown in Figure 2.3. First it extracts sentences with both topic phrases and holder candidates. A named entity tagger is used for identifying holder candidates. Next, it defines the sentiment region of the opinion. Four ways to delimit the region are used in their work:

- Window1: full sentence
- Window2: words between Holder and Topic
- Window3: window2 $\pm$ 2 words
- Window4: window 2 to the end of sentence

Then the system calculates the polarity of the sentiment words. Finally it combines the sentiments of these words to determine the polarity of the given text. Three approaches are used for sentiment synthesis: a) **Model 0:** $\prod$(signs in region), signs are positive(+1) or negative(-1); b) **Model 1:** harmonic mean of the sentiment strengths; c) **Model 2:** geometric mean of the sentiment strengths.

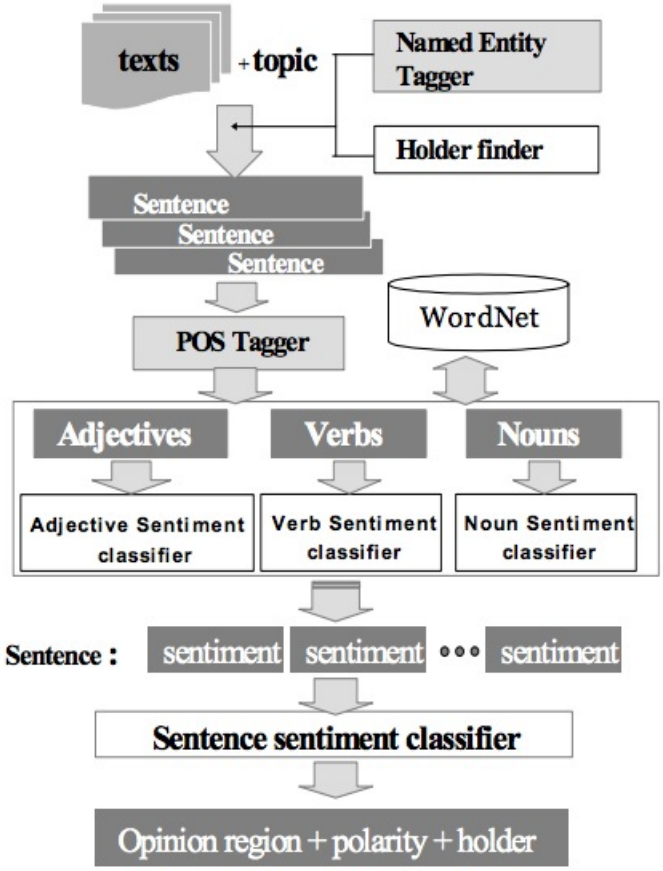


Figure 2.3: System architecture [17]

Figure 2.4 and Figure 2.5 shows the results with manually annotated holder and automatic detected holder. The best performance with manually annotated holder reaches an accuracy of 0.81 and 0.67 with automatic detected holder. It also shows that the region window 4 outperforms the other regions.

Hu and Liu proposed feature-based opinion summarization to help potential consumers make decisions [15]. The architecture is shown in Figure 2.6. First product features are extracted from customer reviews. Then for each feature, the opinion sentence and their sentiment orientations (positive or negative) are identified. Finally, a summary is generated.

Product features can be either implicit or explicit in customer reviews. Considering the following reviews about a digital camera:

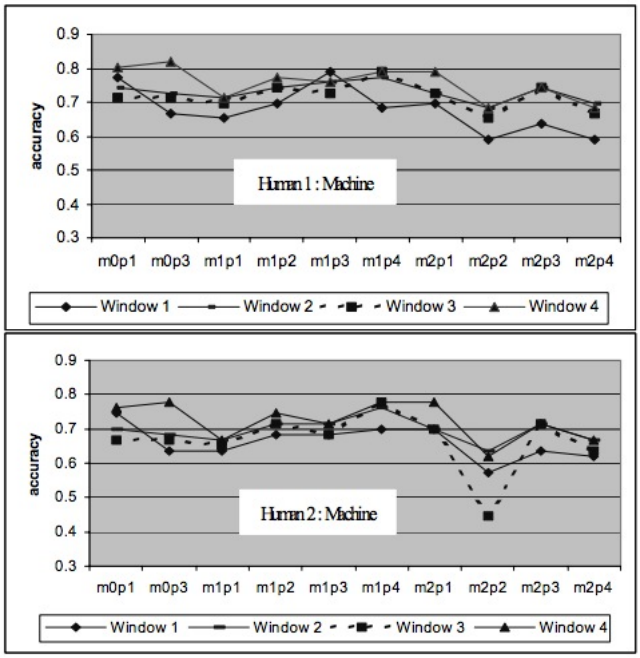


Figure 2.4: Results with manually annotated holder. [17]

- The pictures are very clear.
- While light, it will not easily fit in pockets.

The first sentence indicates that the user is pleased with the picture taken with the camera. Obviously "picture" is one of the features a camera has. However, some implicit features are relatively difficult to detect such as the second sentence. The user is actually unsatisfied with the size of the camera without the word "size". Their work only focused on explicit features.

To determine the semantic orientation of opinion words, they proposed a simple method by using the adjective synonym set and antonym set in WordNet. In WordNet, each adjective has a cluster of synonyms and a cluster of antonyms. Generally speaking, if an adjective is positive (negative), then its synonyms are very likely to be positive (negative), and its antonyms tend to be negative (positive). Based on this idea, we can predict the semantic orientation of an adjective. Starting with a set of seed adjectives (whose semantic orientation has been annotated), we can grow this set by adding new adjectives with predicted orientations.

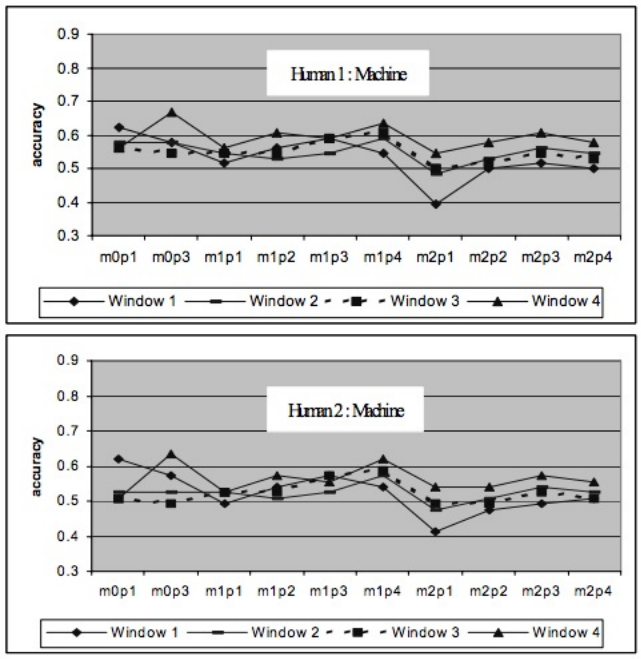


Figure 2.5: Results with automatic detected holder. [17]

Once we have a set of opinion words, we can predict the orientations of opinion sentences. It is worth noting that negation words such as “no”, “not”, “never” may change the orientation of the sentence. In their work, they negate the orientation of an opinion word if a negation word is found within a distance of 5 words from the opinion word. For example, the orientation of the sentence “the pictures are not clear” can be correctly negated. However, this method is too simple in terms of negation. In this thesis, we will elaborate the methods to deal with negation.

Socher et al. proposed a Recursive Neural Tensor Network (RNTN) model to deal with semantic compositionality [33]. They introduced the Stanford Sentiment Treebank containing 11,855 sentences from movie reviews in the form of parse trees. It also annotates the 215,154 fine grained phrases with sentiment labels (very negative, negative, neutral, positive, very positive). An example of RNTN predicting 5 sentiment classes on every level of a parse tree is shown in Figure 2.7.

They compared the accuracy obtained by RNTN with a standard recursive neural network (RNN), matrix-vector RNN (MV-RNN), Naïve Bayes (NB) and Support Vector Machine (SVM). Table 2.6 shows the results. RNTN performs better than the others, es-

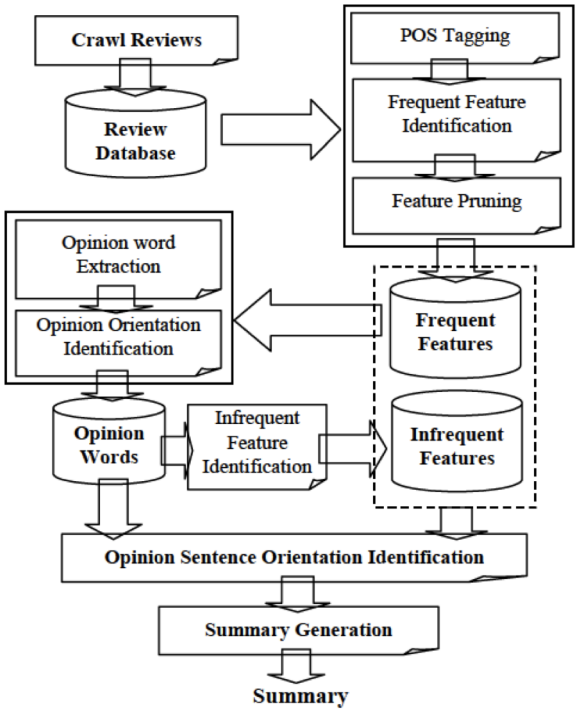


Figure 2.6: Architecture of feature-based opinion summarization [15].

pecially the methods using only bag of words (NB and SVM). It indicates the importance of using a parse tree during sentiment analysis. Unsurprisingly, detecting sentiment at the phrase level is easier than at the sentence level. And the longer the sentence is, the harder the analysis is.

Apart from customer reviews, sentiment analysis has been applied extensively in online deliberation and social media. Li and Wu used text mining combined with sentiment analysis to detect and forecast online forums hotspot [20]. To calculate the sentiment

Model	Fine-grained		Positive/Negative	
	All	Root	All	Root
NB	67.2	41.0	82.6	81.8
SVM	64.3	40.7	84.6	79.4
RNN	79.0	43.2	86.1	82.4
MV-RNN	78.7	44.4	86.8	82.9
RNTN	80.7	45.7	87.6	85.4

Table 2.6: Accuracy for fine grained (5 classes) and binary predictions at sentence level (root) and all nodes [33].

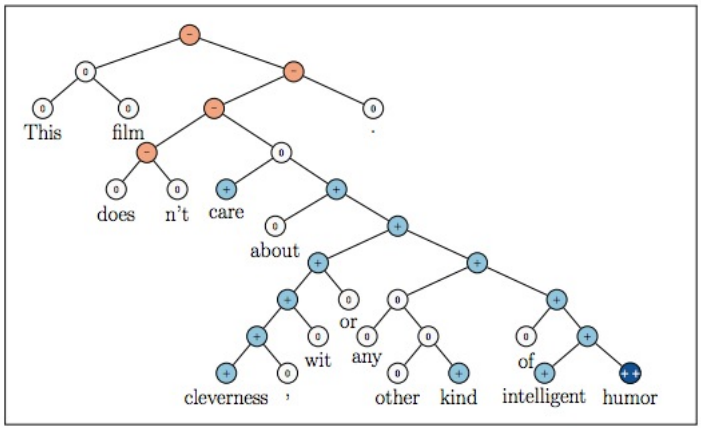


Figure 2.7: An example of Recursive Neural Tensor Network predicting 5 sentiment classes on a parse tree [33].

value of an article, they proposed a simple algorithm. They interpreted the article as a sequence of key words and the basic idea of the algorithm is to add up the sentiment values of each key word in the article. In order to assign a value to each key word, they used a Chinese dictionary with sentiment labels given by HowNet [1] and derived several word lists including a positive word list, a negative word list, a privative list (e.g., non, less) and 5 modifier lists with different sentiment intensities. A key word's sentiment value is determined by its prior polarity and the privatives and the modifiers near it. A drawback of this algorithm is that the distance of the privative from the key word is indecisive and a privative far from the key word can also affect its polarity.

Balahur presented a method for sentiment analysis on Twitter data by using supervised learning. [3] In the preprocessing of tweets, they dealt with the repeated punctuation signs, emoticons, upper and lower case, slangs and repeated letters in a word. In particular, if a word is matched in sentiment lexicons, it will be replaced with its sentiment label (positive, negative, high positive and high negative). Similarly, they replaced the modifiers that negate, intensify or diminish the sentiment word with labels of "negator", "intensifier" or "diminisher". Then they used Support Vector Machines Sequential Minimal Optimization (SVM SMO) to classify three different data sets. The best result is obtained when using unigram and bigram features that appear at least twice with the

[1] http://www.keenage.com/html/e_index.html

replaced labels of sentiment words and modifiers.

Sentiment analysis has also been used in online community management [34]. Sood et al. employed a machine learning method to detect negative content such as personal insults and profanity. They proposed a multistep classifier by combining valence analysis and a Support Vector Machine (SVM) to detect insults and to classify the insult object. The multistep classifier reached 0.5038 F-measure and an accuracy of 0.9082. It is worth noting that they pointed out that a sentiment analysis system performs well on text which is from a domain similar to the training corpus, but likely poorly on that from an irrelevant domain.

2.2.5 NLP Tools

This thesis aims to solve the problem for discussion in Wikipedia AfD from two aspects: to help participant better participate in the discussion and to help administrators make decisions efficiently. Specifically, we analyze the syntactic structure of sentences in the discussion texts, the dependency relations in these sentences, the similarity between sentences, etc. Thus, we have extensively used existing NLP tools to analyze our data. In this section, we will introduce the tools that we used: the BLLIP parser, the Stanford Parser, EDITS, and SEMILAR.

The BLLIP parser is a reranking syntactic parser built by Charniak and Johnson [7] [8]. It takes each sentence (one per line) as input and outputs a Penn Treebank-style phrase structure tree. The default model was trained on the Wall Street Journal (WSJ) Penn Treebank (PTB). The BLLIP parser is also known as the Charniak-Johnson parser, the Charniak parser, and the Brown reranking parser. In this thesis, we use BLLIP parser because it makes fewer error comparing to other parsers in terms of syntactic analysis.

The Stanford parser generates the Stanford typed dependency representation and phrase structure trees [10]. The dependency representation provides users with a simple description of the grammatical relationships of the words in a sentence. Each of these relations is a binary relation between a governor and a dependent. For example, the

dependency representation of the sentence "I agree with you" is:
nsubj(agree-2, I-1)
root(ROOT-0, agree-2)
prep(agree-2, with-3)
pobj(with-3, you-4)
From "nsubj(agree-2, I-1)", we can understand that 'I' is the subject of 'agree'. In this thesis, we used this parser to generate the dependency representation for our data.

EDITS is an open-source package for recognizing textual entailment [18]. It creates an Entailment Engine for training a model and then the model is used to predict the entailment relations between two text fragments. The algorithm is based on the distance between a T-H (Text-Hypothesis) pair and the cost of edit operations including insertion, substitution and deletion to transform T into H. Lexical and semantic similarity is applied in the calculation of distance.

SEMILAR is a semantic similarity toolkit [27]. It includes implementations of various semantic similarity algorithms proposed over the last decade, ranging from word-to-word similarity to document-to-document similarity. Some methods in this toolkit are listed below:

- Lexical overlap
- Greedy method based on word-to-word similarity
- Optimal method based on word-to-word similarity
- Weighted Latent Semantic Analysis (LSA)
- Similarity measure based on Latent Dirichlet Allocation (LDA)
- Quadratic Assignment Problem (QAP)

Chapter 3

Knowledge Management

Knowledge management (KM) is the process of capturing, developing, sharing, and effectively using organizational knowledge[9]. It has been widely used in various applications including business management, information system, human resource management, social media, etc. KM focuses on integrating, organizing and sharing knowledge to make organizational improvements on the basis of existing knowledge. The genesis of this knowledge happens through collecting information sources, extracting useful information from these sources, then processing the information in different ways. For example, to write a report, one needs to read a wide range of material or collect information sources in other forms; and find out what can be used in the report; next add one's thoughts to the report; finally the report is assembled and finished. What one has learned during the process becomes one's own knowledge.

Wikipedia is considered to be one of the most successful knowledge management systems. A large group of volunteers collaboratively participate in building and maintaining the encyclopedia, which includes the processes of creation, editing, deletion, etc. To ensure the quality of the encyclopedia, deletion of articles happens continually. If an article is controversial, an online discussion called "Article for Deletion"(AfD) will be held to determine whether the article should be deleted. It is open to any user to participate in the discussion and make a comment. There can be 50 to 100 discussions per day and some discussions can be lengthy. What can be learned from the discussions and how the overwhelming amount of information can be dealt with are two key problems in terms

of knowledge management that we consider in this thesis. Our contribution is to use natural language techniques to deal with these two problems.

In this chapter, we will first explain the motivation for applying speech act theory to analyze the large amount of data from the AfD discussions. Next, how to identify speech acts will be presented. In particular, we will introduce some methods to detect imperatives, a kind of speech act, by using natural language processing techniques. Then, we will evaluate the results obtained using our methods. By the end of this chapter, we will discuss what we can learn from the knowledge repository produced by our methods.

3.1 Motivation

Due to the large amount of data in Wikipedia: Article for Deletion discussions, it would seem impossible for new users to read through all of the previous discussions and grasp the important information. In this thesis, we aim to provide help to new users by identifying the information that is potentially useful for them.

The main question here is what information do we consider important. It is found that certain mistakes are easily made by novices during the discussions. If we can discover the common mistakes (e.g., being emotional when making a comment) from the past discussions and provide them to new comers, they may understand policies and guidelines better, and hopefully avoid these common mistakes. Looking through the discussion, it is found that some sentences are of instructive significance. For example, "Please refrain from making personal attacks". This sentence is an imperative that warns users not to make personal attacks. Personal attack is a serious problem and impacts the quality of the discussion in a negative way. New users should avoid this kind of problem. If we can collect the sentences with instructive significance, we will have a knowledge base of instructions proposed by users. By analyzing it, we will be able to develop a set of instructions for educating new users. If we include the example above in these instructions, new users will be educated not to be aggressive and make personal attacks. Generally speaking, the problems that have been mentioned frequently in the previous discussions are worth noticing, because we do not want new users to make the mistakes

that have appeared repeatedly.

3.2 Identifying Speech Acts

As we have mentioned in Chapter 2, a speech act is a performative utterance in communication. It consists of a locutionary act, an illocutionary act, and a perlocutionary act. Among the three acts, the illocutionary act is the most central part of a speech act. An illocutionary act is the act of expressing the speaker's intention. Directives are the only type of utterance used to make the hearer do something among the five illocutionary act categories. Thus, the directive has drawn our attention in terms of collecting useful information (i.e., sentences with instructive significance). Some directive sentences from AfD discussions are listed below:

- Add the information, and please give us some information so we can judge these sources.
- Let's avoid compounding the BLP issues caused by the existence of this article, in violation of notability and BLP policies, by having it snow-deleted post-haste.
- You must first discuss the matter there, and you need to be specific.
- Perhaps time would be better spent adding more and improving the article rather than just arguing here.
- Instead of complaining, how about finding such content and improving the article?

As we can see from the above examples, some users directly suggest or command other users to do something such as the the first one. The addressee of the suggestion can also be the user himself such as the second example. The third one is obviously commanding someone to discuss the matter first and to be specific. The first three examples are imperatives, which express the suggestion or command directly. And this kind of directive is relatively easy to detect. However, sometimes people present a command in an indirect way often to be polite, as illustrated by the last two examples. Since the form of this kind of utterance varies, it is difficult to define a rule for recognizing it by computer. In

this thesis, we only detect direct imperatives and leave indirect imperative recognition for future work.

3.3 Detecting Imperatives

In English, a typical imperative is expressed by using the base form of a verb, normally without a subject. To detect this kind of imperative, we need to analyze the grammatical structure of sentences. Consider the sentence below:

- Please refrain from making personal attacks.

The Penn Treebank-style phrase structure of this sentence is:
(ROOT (S (INTJ (VB please)) (VP (VB refrain) (PP (IN from) (S (VP (VBG making) (NP (JJ personal) (NNS attacks)))))) (. .)))

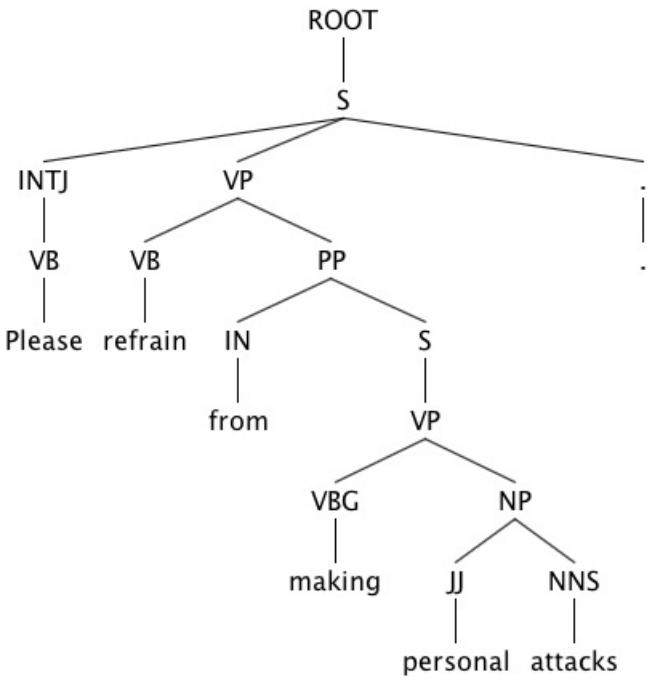


Figure 3.1: Graphical representation of Penn Treebank-style phrase structure tree for the sentence: *Please refrain from making personal attacks.*

Figure 3.1 gives a graphical representation of the structure tree. It shows not only the syntax of the sentence but also the part-of-speech (POS) tags. We also need to know the

dependency relations in this sentence. The result of the Stanford dependency parser for this sentence is:

discourse(refrain-2, please-1)

root(ROOT-0, refrain-2)

prep(refrain-2, from-3)

pcomp(from-3, making-4)

amod(attacks-6, personal-5)

dobj(making-4, attacks-6)

To make it easier for readers to understand, we present the dependency structure tree in a graph (Figure 3.2). "Refrain" is the root of this sentence and there is no subject. To

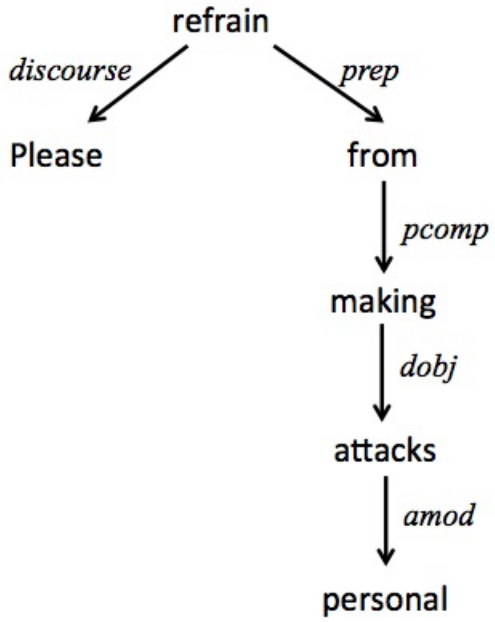


Figure 3.2: Graphical representation of dependency structure tree for the sentence: *Please refrain from making personal attacks.*.

generate a rule for detecting this type of sentences, we need to find out how it differs from other sentences in terms of phrase structure and dependency structure. Then we can make computer automatically detect and extract the useful sentences from the overwhelming information. For comparison, we use a declarative sentence.

- She told him not to make personal attacks.

The phrase structure is:
(ROOT (S (NP (PRP She)) (VP (VBD told) (S (NP (PRP him)) (RB not) (VP (TO to) (VP (VB make) (NP (JJ personal) (NNS attacks)))))) (. .)))
The dependency relations are:
nsubj(told-2, She-1)
root(ROOT-0, told-2)
nsubj(make-6, him-3)
neg(make-6, not-4)
aux(make-6, to-5)
xcomp(told-2, make-6)
amod(attacks-8, personal-7)
dobj(make-6, attacks-8)

In this sentence, "told" is the root, and we can also find a subject "she" in the dependency relation nsubj(told-2, She-1). Note that "told" is an inflected verb, whose base form is "tell". We can observe the difference between verbs in base form and those being inflected from their part-of-speech (POS). Verbs in base from are tagged as VB, while verbs in past tense are tagged as VBD.

According to our observation, a typical imperative contains a verb in base form without any subject. Therefore, the basic rule for imperative recognition is to find those sentences with a verb (in its base form) as the root in the phrase structure and this particular verb has no subject child in the dependency structure.

Another form of imperative we have mentioned in the previous section is like the sentence: "You must first discuss the matter there, and you need to be specific". In our thesis we apply a simple rule. That is, we use the form that a personal pronoun or noun (e.g., you, we, username) followed by a modal verb (e.g., should, must, need) to recognize the speech act. It can be a command of asking someone to do something or a prohibition against some acts. For instance, the sentence "you can't vote twice" starts with a personal pronoun "you" and a modal verb with negation "can't" follows immediately. Specifically, the sentences with the following forms tend to be imperatives though our observation:

- you should / must / need to / have to / can not ...
- we must ...

This kind of imperative can be easily detected by keyword searching.

3.4 Evaluation

In this section, we will evaluate the performance of our methods to detect imperatives. We first ask two human annotators to extract all imperatives and we then calculate their agreement. Then the two annotators have a discussion about their disagreements and finally reach an agreement on all imperatives, which becomes our gold standard. Finally, we compare the result produced by our methods with this gold standard.

The two annotators are undergraduate students at The University of Western Ontario, one is from biology and the other one is from linguistics. They were asked to extract imperative sentences from our data. Before annotating, we gave them several examples to annotate and discussed whether their judgements were correct. In this way, they would have a better understanding about how to annotate imperatives. Then they were asked to extract separately imperative sentences from one day's AfD discussion. However, when we compared their annotation, we found a lot of disagreement. To calculate their agreement, we used Cohen's kappa coefficient [6]. The equation for kappa value is:

$$kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \tag{3.1}$$

where $Pr(a)$ is the relative observed agreement and $Pr(e)$ is the hypothetical probability of chance agreement.

The value of kappa is 0.436, which indicates a moderate agreement according to the magnitude guidelines in [19]. The agreement is much lower than our expectation, so we discussed the annotation task with the two annotators. We found that they had different standards on judging whether a sentence is an imperative or not. One tended to be strict and the other one more lenient. For example:

- This should probably be a merge recommendation.

This is a typical sentence extracted by one of the annotators. From this example, we can tell that the speaker suggests merging, but it seems to be the speaker expressing an opinion rather than giving a directive speech act. It can be recognized as an indirect directive but hardly as an imperative.

Then we asked the two annotators to annotate another day's data. This time they were asked to be strict. After calculation, the value of kappa is 0.576. Since there are over 1,000 sentences in one day's data and imperatives are only a small portion of the data, we hypothesized that the annotator might miss some when extracting them. Thus we asked each annotator to read the imperatives extracted only by the other annotator and to give their judgement for these potential imperatives. After that, we calculated the kappa coefficient again and it increased to 0.883 which is considered to be almost perfect. However, we still had disagreement, so we asked them to discuss the remaining disagreed upon sentences and to reach agreement. The sentences agreed by both becomes our gold standard.

Among the disagreement, we found that two typical type of sentences:

- If you know of any, list them here or add them to the article and I'll vote keep.
- The article needs to be cleaned up though.

One is conditional sentence. Although the illocutionary force is effective under certain condition, the main clause is still regarded as imperative. Thus, in our gold standard, we agree that this type of sentence is imperative. Another one is ambiguous in deciding whether it is imperative or not. It usually has the meaning of "something needs to be done". Apparently the speaker hopes something to be done by someone, but no hearer is specified. Therefore, we decide to recognize this type of sentence as non-imperative.

Finally, we compared the imperatives extracted by our methods with the gold standard. To generate the phrase structure tree and dependency tree, we use BLLIP parser and Stanford parser. The result is shown in Table 3.1. The computation of precision, recall and F-measure are shown below:

$$precision = \frac{|\{relavant\ document\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|} \tag{3.2}$$

$$recall = \frac{|\{relavant\ document\} \cap \{retrieved\ documents\}|}{|\{relavant\ documents\}|} \tag{3.3}$$

$$F = 2 \times \frac{precision \times recall}{precision + recall} \tag{3.4}$$

Our methods produce a high precision of 0.8447 and good recall of 0.7337. Table 3.2 illustrates some examples being detected incorrectly and those that our methods fail to detect. A majority of sentences in the category "incorrectly detect" are those with an implicit subject "I". They look like imperatives, which use the base form of a verb, normally without a subject. However, they are statements in which the speaker has eliminated the subject "I". As for the sentences that our methods fail to detect, one reason is the incorrect analysis by the parsing tool. Another reason is that the sentence with the form of subject + modal verb, but the subject is a specific noun (person or organization) instead of a pronoun. These imperatives are more difficult to detect due to their flexibility.

Precision	Recall	F-measure
0.8447	0.7337	0.7874

Table 3.1: Precision, Recall and F-measure of the detection of imperatives produced by our methods

3.5 Discussion

To obtain a knowledge repository that can be used to educate new users, we need to extract a large amount of useful information from previous Article for Deletion discussions. In our thesis the information that we are interested in is delivered by imperative sentences. We have demonstrated that our natural language processing methods can effectively recognize imperatives. Thus, we have chosen one week of discussions in each month in the year 2013 and we have applied our methods to this corpus to obtain a knowledge repository.

Wikipedia policy plays a key role in deletion discussions. It provides users with standards and rules to follow and resolves conflicts during discussion. Citing a policy in

Incorrectly detect	**Fail to detect**
Have no idea what this is even supposed to prove and is certainly not significant coverage about him.	But please note that I did ask for these links very early in the piece , and you have only just provided them.
Agree with most of the rest of this.	All this needs to be addressed.
If that's all then we must Delete, which is why I ask if there are Urdu or other language sources, or other spellings of the name?	The nominator should please refrain from further nominations in this Gettysburg deletion spree as it seems clear that he is not following deletion policy.

Table 3.2: Examples being incorrectly detected as imperative and those that our methods fail to detect.

an argument makes the argumentative point more powerful and persuasive. In this thesis we will only analyze the extracted imperatives that contain explicit policy references and leave the ones without policy references for future work.

There are hundreds of Wikipedia policies, which make it impossible for new users to read them all before they make a comment in a discussion. If we can provide them with a list of policies to which they might need to pay more attention, it may prove helpful to them to understand how to propose a better point of argumentation to justify their opinion. Since an imperative aims to suggest or command someone to do something, it can also be considered as a guide for new users. What has been mentioned frequently in the previous discussions is worth noticing. Table 3.3 illustrates a few policies that have been mentioned frequently in our newly constructed knowledge repository and some examples of their appearance in imperatives. "WP:GNG"[1] is a general notability guideline; "WP:RS"[2] requires Wikipedia articles to be based on reliable sources; "WP:BEFORE"[3] is a guideline for nominating articles for deletion; "WP:NOTINHERITED"[4] indicates that notability is not usually inherited; "WP:BIO"[5] is a notability guideline for biographies.

[1]http://en.wikipedia.org/wiki/WP:GNG
[2]http://en.wikipedia.org/wiki/WP:RS
[3]http://en.wikipedia.org/wiki/WP:BEFORE
[4]http://en.wikipedia.org/wiki/WP:NOTINHERITED
[5]http://en.wikipedia.org/wiki/WP:BIO

From the examples, we can see that users justify themselves or refute other arguments by using Wikipedia policies. They ask others to read some certain policies, which means these policies are likely to be neglected by them.

Given the list of policies being mentioned frequently in our knowledge repository, new users can participate in the deletion discussion more effectively by reviewing it first. They can not only use the policy to support their own argumentation, but also to avoid mistakes that violate the policy.

Policy	Examples
WP:GNG	Please carefully study the General Notability Guideline (WP:GNG), which expects "significant coverage in reliable sources that are independent of the subject". With respect to your first question, See WP:GNG bullet point 3 which advises editors that the nature of the sources need to be considered when evaluating notability. Please check WP:GNG for general notability guideline and WP:BIO, so far, Brady Haran doesn't meet the sufficient criteria for having an article within WP.
WP:RS	Please read WP:RS to see which sources may be considered reliable. Please see WP:RS for what constitutes a reliable source for Wikipedia articles. 42of8, please read WP:RS, WP:V, and WP:GNG.
WP:BEFORE	AFD isn't the place for sourcing issues, see WP:BEFORE. While you're looking, please see WP:BEFORE, particularly Read and understand these policies and guidelines. Have a read of WP:BEFORE.
WP:NOTINHERITED	please see WP:NOTINHERITED, subject isn't notable themselves just because they are an acquaintance to multiple heads of state. And the point I forgot to make, the wikipedia guidelines are very clear that notability is NOT inherited - saying you think that notability is inherited in contradiction of the guidelines won't help your case either, see WP:NOTINHERITED.
WP:BIO	Read under WP:BIO, WP:BASIC : "If the depth of coverage in any given source is not substantial, then multiple independent sources may be combined to demonstrate notability". Please note that WP:BIO states: "Failure to meet these criteria is not conclusive proof that a subject should not be included".

Table 3.3: The policies that appear frequently in imperatives and some examples of their appearance

Chapter 4

Decision Making

Decision making can be regarded as a problem solving process ending with a solution from among several alternative choices. It can involve the analysis of a large amount of information regarding a certain problem. In order to make a rational decision when confronted with information overload, the main issue we need to deal with is the quantity of information. In AfD, although users give their votes (delete, keep, redirect, etc.), the votes are not the key factor in determining the final decision. Administrators need to read through the discussions and make decisions based on these discussions. However, reviewing the discussion is time-consuming. Thus we need to find an effective way to help administrators make decisions. In this thesis, we propose a decision making support system by using natural language processing (NLP) techniques including identifying relations between texts and identifying sentiment polarity of sentences. The architecture of the system is shown in Figure 4.1. For an AfD discussion, the first thing to do is to eliminate redundancy. Then, the system classifies similar arguments into groups. For each group, it further classifies the arguments into 3 groups: positive, neutral and negative sentiment.

In this chapter, we will first explain the motivation for using NLP techniques to solve decision making problems. This is done in Section 4.1. Next, we will introduce some methods for identifying relations between texts in Section 4.2. Then in Section 4.3, how to identify sentiment polarity will be explained. Finally, we will present a prototype of the decision making support system in Section 4.4.

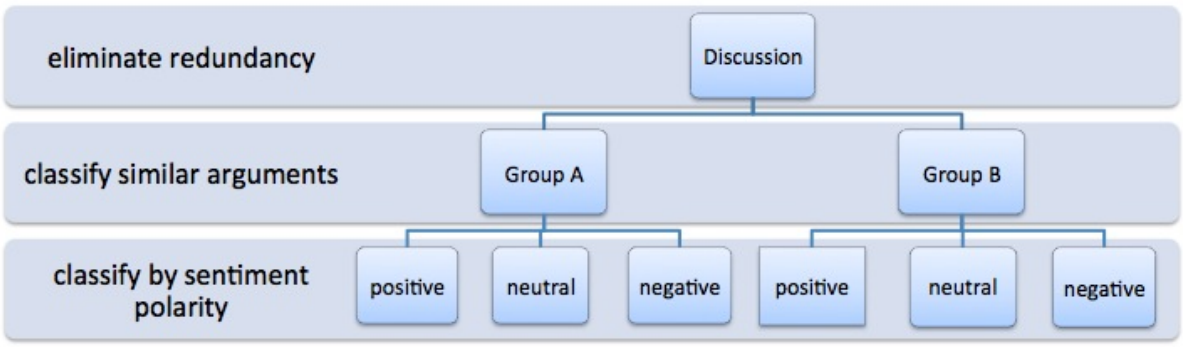


Figure 4.1: Architecture of decision making support system.

4.1 Motivation

For the purpose of making rational decisions effectively, we need to make it easier for the administrators to read the discussions. By observing some discussions in AfD, we find that some information is redundant. That is, some users repeat the arguments that have already been proposed by other users. For example, two users' comments *"Could be redirected to OpenXMA, the content of which isn't all that different from this article"* and *"Redirected to OpenXMA as suggested"* are considered redundant. The redundant information itself does not add a new perspective to the final decision making. It would be a waste of time to read these redundant arguments. In this thesis, we aim to eliminate the redundancy, but we want to record how many times that the same argument is mentioned by different users. Although they are redundant, they are important and useful to some extent in decision making, since users tend to use them several times to justify their opinions.

In one user's deliberation, there can be several arguments involving distinct factors (notability, source, maintenance, bias, etc.). Classifying the arguments that talk about the same factor into a single group would make it easier for an administrator to balance the different opinions regarding a certain factor. Consider the two arguments below:

- A: Nothing to show that this is any more notable than any of the millions of other intersections in the world.

- B: This is a major intersection in a provincial capital that appears to likely be

named after the June 5, 1963 demonstrations in Iran, which suggests that it's a pretty big deal in context.

The two arguments are proposed by two users. Both of them are talking about the notability of the intersection, while A thinks this intersection is not notable, but B holds the contradictory opinion that it is notable and elaborates his reason for notability.

In this thesis, we first use relations between texts to eliminate redundancy and classify arguments into groups. Then we further classify the arguments as pros, cons and neutral based on sentiment analysis. To make the process of decision making more efficient, we design a system that administrators could use to review the arguments by groups and the pros and cons in each group are explicitly classified.

4.2 Identifying relations between texts

In this section, we will introduce two types of relations between texts including textual entailment and text-to-text similarity. In terms of eliminating redundancy, either of them can be used. For classifying arguments into groups, we have decided to use text-to-text similarity.

4.2.1 Textual Entailment

Textual entailment (TE) represents a relation with a direction between two text fragments (i.e., text (T) and hypothesis (H)). If we can infer H by reading T, we would say T entails H ($T \Rightarrow H$). In other words, the information we can get from H is also contained in T. In this perspective, H is redundant. For example:

- T: If you had checked Google Scholar, you would see that the top result has 13,311 citations.
- H: Very high cites in GS.

Suppose we have the above sentences in a deletion discussion. If we remove H, no information is lost, since T also indicates "very high cites in GS".

4.2.2 Text-to-text Similarity

Text similarity can be interpreted as similarity between sentences, paragraphs, documents, etc. In this thesis, we use sentence-to-sentence similarity. Consider the following pair of sentences in a discussion about the article of a primary school:

- Non-notable elementary school.
- It is not a notable primary school.

The two sentences are almost the same. It would be safe to delete either one of them in order to eliminate the redundancy. In this thesis, we only keep one sentence when we find several highly similar sentences.

Another situation is that although two sentences are talking about the same thing (i.e., they are similar), they are presented in different perspectives, which can be opposite positions. Consider the sentences below:

- The topic is notable, being covered in numerous sources.
- There don't seem to be any reputable sources at all in this article.
- I could find no significant coverage of his work in reliable sources.

All of these sentences involve source. We can say they are similar or related, but we would not consider any of them redundant. The first holds a pro view towards source, while the second and third one holds a con. In our work, we want to recognize these similar (related) sentences and classify them into a group. As for how to distinguish pros and cons, we will solve it by using sentiment analysis techniques discussed in the next section.

4.2.3 Experiments and Evaluation

To test different methods for identifying relations between texts, we extract 80 pairs of sentences from deletion discussions in AfD (see Appendix A). We manually annotate them as being similar or not, and being entailed or not. Then we use EDITS, an open-source package for recognizing textual entailment [18], to predict the entailment relations

between each pair of sentences. We want to use this tool to eliminate redundancy, but to recognize and keep contradictory pairs. However, the results show that EDITS incorrectly recognizes the pairs which are similar but contradictory as entailment, such as the following pair of sentences:

- However, the article fails WP:GNG and WP:NFOOTBALL.
- Article satisfies both WP:GNG and WP:NFOOTBALL.

We can say the two sentences are similar/related, but there is no entailment relation between them. Keeping the contradictory sentences is very important in the process of decision making. Thus, using EDITS to eliminate redundancy seems not to be feasible.

Another option is to use sentence-to-sentence similarity. In our experiment, we use SEMILAR, a semantic similarity toolkit [27]. We have tested three approaches provided in SEMILAR: optimum method based on WordNet, similarity based on Latent Semantic Analysis (LSA) and similarity based on Latent Dirichlet Analysis (LDA). SEMILAR assigns a similarity score to each pair of sentences in the range from 0 to 1 (see Appendix B). To evaluate the accuracy of the three approaches, we find a threshold to divide the result into two groups (i.e., similar and not similar). We compute the accuracy for 101 thresholds ranging from 0.00 to 1.00 with an interval of 0.01 and find the highest accuracy. For the optimum method based on WordNet, when we set threshold at 0.13, the best accuracy of 76.25% is obtained. For the approach based on LSA, when we set the threshold at 0.21, we reach an accuracy of 76.25%. And for LDA, the accuracy is 75% when the threshold is located between 0.13 and 0.21. In this thesis, we choose to use the optimum method based WordNet in the decision making support system, since it costs much less time (less than 1/2) than the other two and it performs well.

4.3 Identify sentiment polarity in texts

Sentiment analysis on customer reviews (e.g., product reviews, movie reviews, and restaurant reviews) has been studied extensively. However, how to classify arguments (e.g., deletion discussions in AfD) by their sentiment polarity has rarely been proposed. It

is relatively difficult to analyze the sentiment of arguments since they tend to be less explicit compared to customer reviews. Additionally, customer reviews express opinions involving a particular object, event, or place, while AfD arguments tend to be more comprehensive and divergent. In particular, one AfD involves a variety of aspects ranging from event to place, people and organization. Accordingly, the arguments in deletion discussions in AfD represent broad knowledge instead of a focussed topic.

The granularity of sentiment analysis ranges from word to sentence, paragraph and document. In terms of the word level, several sentiment word lists have been developed recently, such as SentiWordnet [12] and MPQA Subjectivity Lexicon[25]. Most of the coarser granularity sentiment analyses are based on these sentiment word lists. For example, the MPQA Subjectivity Lexicon was used in Wilson et al.'s work to recognize contextual polarity at the phrase level [39]. Each entry in this word list contains word token, part-of-speech (POS), prior polarity, etc. In this thesis, we also use the MPQA Subjectivity Lexicon to identify sentiment polarity at the sentence level.

A number of methods in sentiment analysis have been proposed using the bag of words representation of text. These methods may perform well at the paragraph and document level because these levels depend a lot simply on the number of sentiment words in the text. However, at the sentence level, just counting sentiment words falls short of an appropriate technique to determine the sentiment polarity of the sentence. It is not wise to ignore the order of words and the syntactic structure, which the bag of words representation of text does, since the polarity of a word can be negated or changed by other words such as its modifier or the grammatical subject of the sentence. In Section 4.2.1, we will elaborate the different types of negations that exist. An approach to recognizing the sentiment polarity at the sentence level will be presented in Section 4.2.2. And we will evaluate this approach in Section 4.2.3.

4.3.1 Negation

Negation[1] plays a significant role in sentiment analysis. This is one key factor that makes the analysis so difficult. In the context of sentiment, negation occurs in a variety of forms in text. It can be the word *not* next to a sentiment word (e.g., "not notable"), or far from the sentiment word such as the negation in the subject (e.g., "no one thinks that it is notable"). We need to identify not only negation words but also the scope of the negation. Consider the sentences listed below:

1. I *agree* that the place is *notable*.
2. I *don't agree* that the place is *notable*.
3. I *disagree* that the place is *notable*.
4. *Neither* one of us *agrees* that the place is *notable*.
5. The place is of *indeterminable notability*.
6. It is a *violation* of *notability*.

Sentence 1 is obviously a positive argument. The prior polarity of the words *agree* and *notable* are both positive. Sentences 2 and 3 hold the opposite polarity of sentence 1 by using the negation word *not* and *disagree*. *Not agree* has the same meaning as *disagree* and their negation scope is the dependent (subordinate) clause *the place is not notable*, which is positive. After negation, the overall sentence is negative. Sentence 4 expresses negative opinion by negating the subject. In sentence 5, *indeterminable* is a negative word, while *notability* is a positive word. The phrase *indeterminable notability* becomes negative. Similarly, in sentence 6, *violation* is a negative word and the phrase *violation of notability* transforms the positive object of the preposition into a negative phrase due to the effect of the preposition. If we just use a shallow analyzer which simply adds the polarity of all the words in a sentence (consider positive as 1, negative as -1, neutral as 0), then we will not obtain satisfactory results. We will now analyze different types of

[1]The term "negation" is usually used in the context of the meaning of a piece of text. Throughout this section, we use it in a more focussed way, to mean the reversal of sentiment polarity suggested by a piece of text.

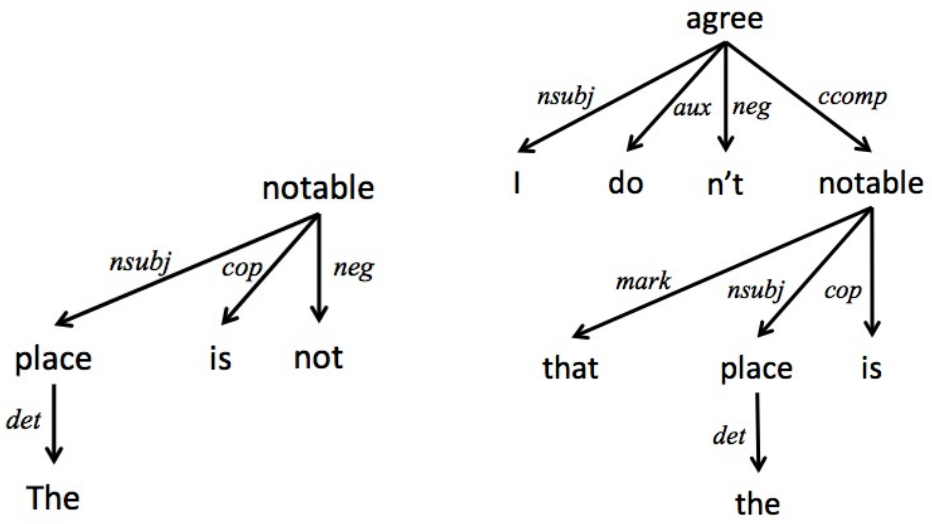
negations in detail and explain how to recognize various types of negation in syntactic and dependency structures.

Local negation

Local negation is the easiest type of negation to recognize. A *not* usually modifies the sentiment word. The dependency structure tree for the following sentences is shown in Figure 4.2.

- The place is not notable.
- I don't agree that the place is notable.

We can see that *not* is the child of *notable* and *agree* in the dependency trees of the two sentences. And the relation between *not* and the sentiment word is "neg (negation modifier)". The positive adjective *notable* with its modifier *not* becomes negative as *non-notable*. Likewise, *not* modifying the positive verb *agree* makes it negative as *disagree*.



(a) (b)

Figure 4.2: Graphical representation of dependency structure trees for the sentences: (a) The place is not notable. (b) I don't agree that the place is notable.

Predicate negation

Another type of negation is found in the predicate by using verbs with negative polarity. For example:

- I disagree that the place is notable.

In this sentence, the verb *disagree* is negative and it negates the dependent clause *the place is notable*. As shown in Figure 4.3, the clause is the child of the verb in the dependency structure tree. Apart from negating dependent clause, a negative verb also negates its object (e.g., it violates notability).

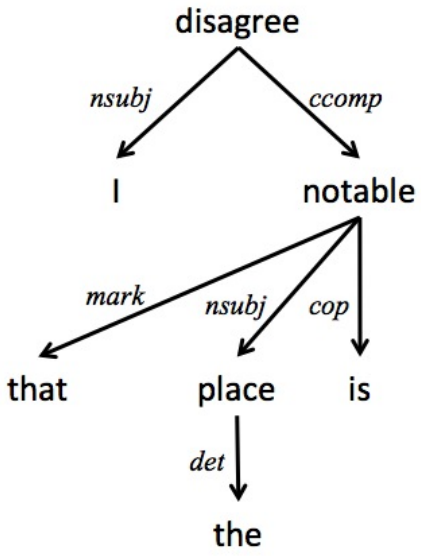


Figure 4.3: Graphical representation of dependency structure tree for the sentence: I disagree that the place is notable.

Subject negation

A negative subject leads to the negation of its predicate. For example:

- Neither one of us agrees that the place is notable.

The subject *neither one of us* is negative and it negates the positive predicate *agrees*. The dependency relation between the subject (child) and the predicate (parent) is "nsubj

(nominal subject)" as shown in Figure 4.4. The parent in the "nsubj" relation is not always a verb. When the verb is a copular verb, the parent is the complement of the copular verb such as an adjective or a noun, for example, *none of these places is notable*. In this sentence, the subject *none of these places* reverses its parent *notable*'s polarity.

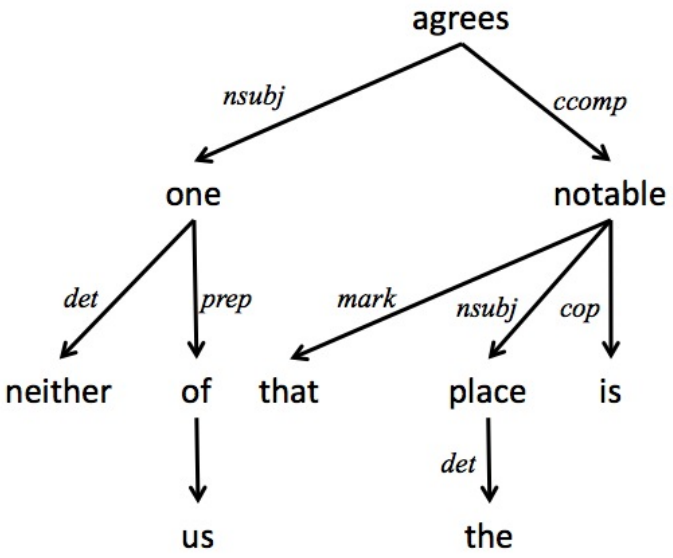


Figure 4.4: Graphical representation of dependency structure tree for the sentence: Neither one of us agrees that the place is notable.

Preposition negation

The preposition *of* plays an important role in negation. Usually the polarity of the object following the preposition *of* can be changed by the word modified by the preposition. An example is:

- It is a violation of notability.

It can be seen from Figure 4.5 that the preposition *of* is the child of *violation* in the relation "prep (prepositional modifier)" and the parent of *notability* in the relation "pobj(object of a preposition)". Obviously the negative word *violation* negates the positive word *notability* through the preposition *of*.

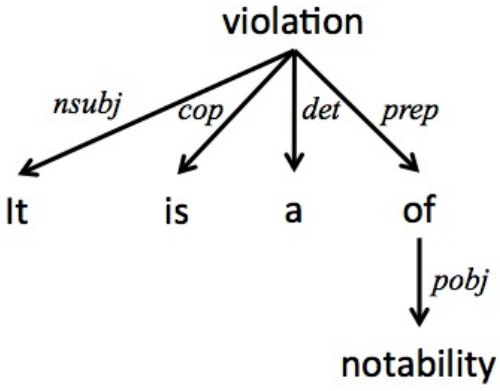


Figure 4.5: Graphical representation of dependency structure tree for the sentence: It is a violation of notability.

Modifier negation

Some sentiment words' polarities can be negated by their modifiers. An example is *indeterminable notability*. *Notability* is a positive word, while it is modified by a negative word *indeterminable*; as a consequence, the phrase becomes negative. A negative modifier might also negate a negative word, such as *little damage*, *never fail*. However, a word is not always negated by a negative modifier (e.g., *terribly allergic*). It may remain as its prior polarity. It is also worth noticing that global context affects the phrase polarity. In this thesis, we are analyzing the deletion discussion in AfD, in which the phrase *original research* is considered to be negative. Although it seems to be positive most of time, it actually violates the Wikipedia policy for being non-encyclopedic. A Wikipedia article is not supposed to be original research. Thus the polarity of *original research* will be annotated as negative. There is no simple way to determine how the modifier affects the word being modified. That is, we cannot predict the polarity of a phrase just by the polarity of each word.

In order to determine the polarity of a phrase involving a sentiment word, we use machine learning methods. The phrases found in the thesis data are composed of two words in the following combination:

- Noun modified by adjective

- Noun modified by noun
- Adjective modified by adverb
- Adverb modified by adverb
- Verb modified by adverb

At least one of the words in a phrase has to be sentiment word. We use 6 attributes to describe a two-word phrase:

- First word token
- Second word token
- First word polarity
- Second word polarity
- First word POS (part-of-speech)
- Second word POS

Based on these six attributes, we want to predict the polarity of the phrase. For example, the phrase *indeterminable notability* is described as indeterminable, notability, negative, positive, adjective, noun. Its polarity is labeled as negative by a human. In this thesis, we annotate the polarity of a number of instances (phrases) by hand. By using this labeled corpus, supervised machine learning methods can be applied to build a model to predict the unlabeled phrase polarity more accurately. The corpus and the outcomes of some machine learning methods will be discussed in Section 4.3.3.

4.3.2 Methods to Determine Polarity

After analyzing a number of sentences, we propose a recursive algorithm. Rather than just calculating the sum of the polarities of the words in a sentence, this algorithm is based on the dependency structure tree of the sentence. In our algorithm, we assign a polarity score to each node in the dependency structure tree according to its position

and its related nodes. We have discussed the different types of negation in the previous section. We integrate them in our algorithm. The flow chart of calculating the polarity score for a node is shown in Figure 4.6. Take a node as input, and the polarity score for the node as output. A node contains the following information:

- its position in the sentence (sequence number in the sentence)
- word token
- part-of-speech (POS)
- prior polarity
- parent node
- child nodes
- dependency relations between itself and its children

The POS of a node is obtained by examining the phrase structure tree of the sentence. The prior polarity of a node is the out-of-context polarity determined by the MPQA Subjectivity Lexicon [25]. We use the word token and the POS to match the clues in the Lexicon and assign the prior polarity of the matching clue to the node. The parent node, child nodes and dependency relations between the node itself and its children are determined by the dependency structure tree. For a sentence, its polarity score is the polarity score of its root node.

In the recursive algorithm for calculating the polarity score of a node, we first check if it is a leaf node, i.e. it has no children. It it is a leaf node, we will assign its prior polarity, which is obtained from the lexicon (0 as neutral, -1 as negative, 1 as positive), as its polarity score. Otherwise we check if it has a modifier. If it does, we will use the machine learned model to determine the polarity of the phrase (node with the modifier) and assign this polarity to the node. If it doesn't have a modifier, the node polarity remains as its prior polarity. Next, check if there is a negation, which can be detected by looking up the dependency relations between the node and its children. Once the negation is found, the node polarity is negated. Then, for the node whose POS is 'verb' and its polarity

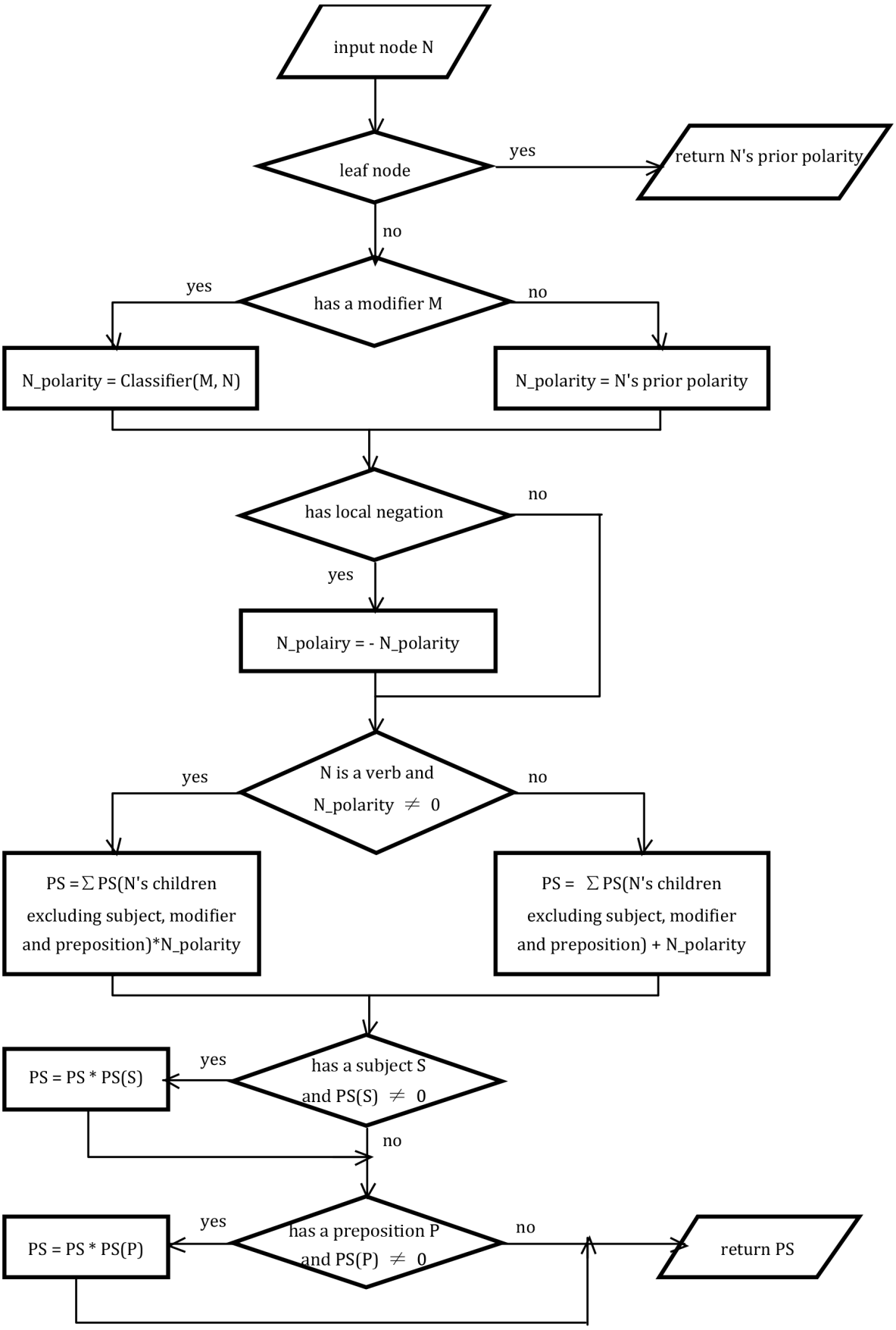
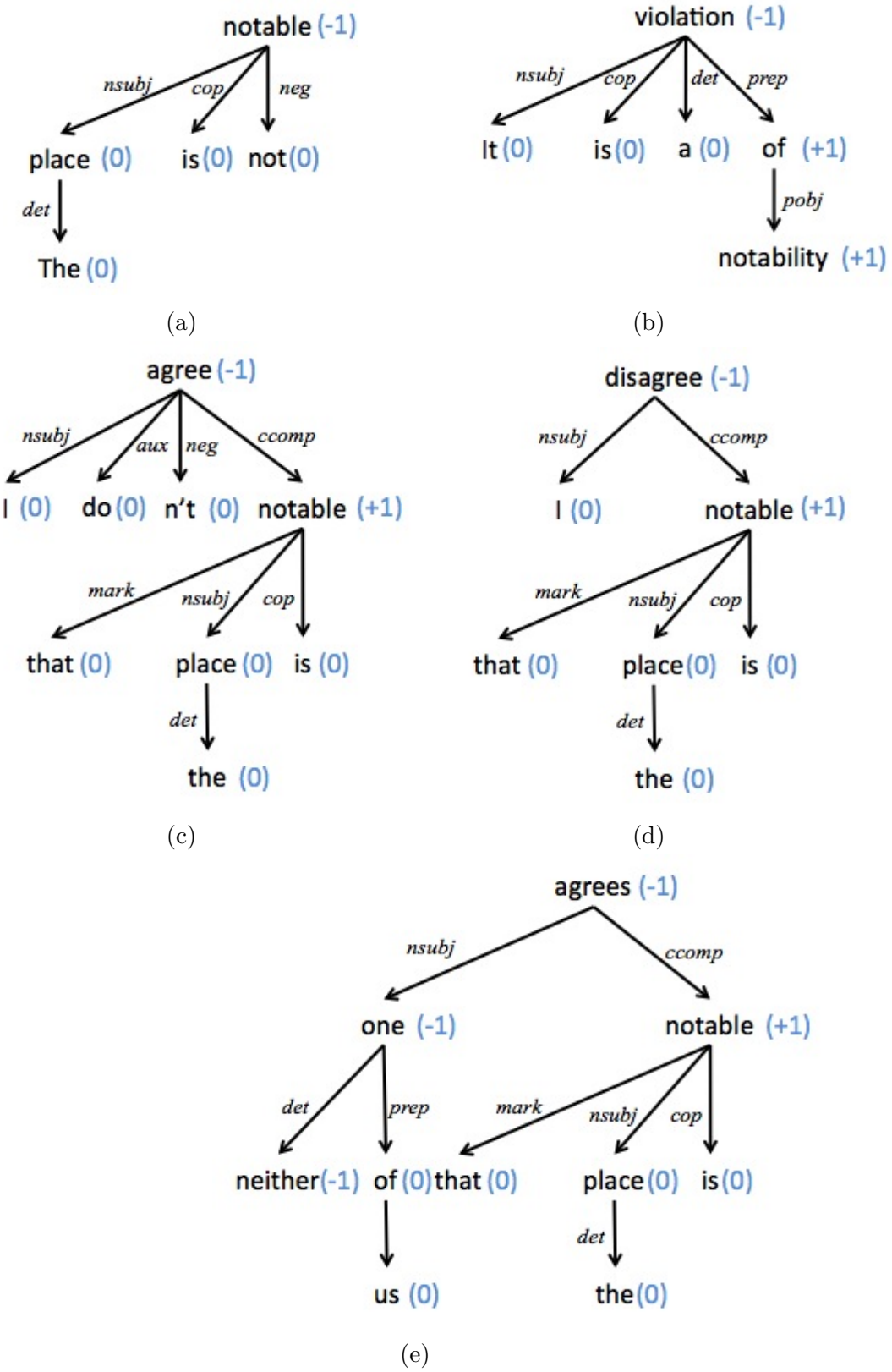


Figure 4.6: Flow chart of calculating the polarity score for a node.

is not 0 (neutral), we multiply the node polarity by the sum of the polarity scores of the child nodes excluding those that are subject, modifier and preposition since they are involved in negation and need to be considered separately. If the verb node's polarity is 0 (neutral), then we only perform the sum without multiplication. For any node whose POS is not 'verb', we calculate the sum of its child nodes' polarities excluding those that are subject, modifier and preposition and add the node polarity to it. The result is the temporary polarity score. The next step is to check for subject and preposition relations, both of which can be detected in the dependency relations. If there is a subject, we multiply the temporary polarity score by the polarity score of the subject node. And similarly, when a preposition occurs, the temporary polarity score is multiplied by the polarity score of the preposition node. Then the result of the computation is the final polarity score of the input node. The polarity score ranges from negative infinity to positive infinity; while it usually ranges from -3 to +3 for most sentences.

If the polarity score of the root node is less than or equal to -1 (greater than or equal $+1$), then the polarity of the sentence is negative (positive). Otherwise, the sentence, whose root node's polarity score is 0, is neutral.

Figure 4.7 illustrates the polarity score for each node for the six sentences we have discussed in Section 4.3.1. Subfigures (a) and (c) show how local negation changes the polarity in a sentence. (b) is an example of preposition negation showing that multiplication is more effective than adding the polarity of each word (in this case simply summing would give 0 (neutral) for this phrase). Predicate negation is shown in (d). For (e), we observe two positive words *agrees* and *notable* and one negative word *neither*. If we were to use a simple bag of words method, we would get a resulting positive polarity. However, the negative word *neither*, being part of the subject, plays a dominant role in this sentence. In our algorithm, to calculate the root node polarity score, since it is a verb, we first multiply its prior polarity +1 with the polarity score of the node *notable*, which is +1. Then we multiply the result +1 with the polarity score of the subject node, which is -1. And we get the result -1 as the polarity score of the root node, which also indicates the overall polarity of the sentence is negative.



(a) (b) (c) (d) (e)

Figure 4.7: Polarity score on every node in dependency structure for the sentences in Section 4.3.1

4.3.3 Experimentation and Evaluation

In this thesis, the sentiment wordlist we use is the MPQA Subjectivity Lexicon [25], which consists of over 8,000 subjectivity clues. Each clue has several features including type, word token, part-of-speech (POS), prior polarity. For each word in our analysis, we use the word token and its POS which can be obtained with the use of a syntactic parser to match the subjectivity clue in the MPQA lexicon. If it has a match, then we assign the prior polarity of the matching clue to the word. Although the lexicon is large enough, we have still needed to add a few clues according to the context of our data such as *notability* and *source*. In this section, we will first discuss the experiment and outcomes of the phrase polarity prediction using three supervised machine learning methods. Then, the performance of our recursive algorithm will be presented.

Phrase polarity prediction

For the phrase polarity prediction experiment, we extracted 795 two-word phrases from deletion discussions in AfD and annotated their polarity manually. They all follow the combination being discussed under Modifier negation in Section 4.3.1 and at least one of the two words is a sentiment word. 280 phrases are labeled as neutral, 320 are positive and 195 are negative. Since the data is unbalanced, we use the SMOTE filter in WEKA [13] to balance the data. After balancing, we have 961 balanced instances.

The machine learning methods for phrase polarity prediction in our experiment are Naïve Bayes, k-nearest neighbor (KNN) and decision tree; because they are basic machine learning methods that have been used in many researches. We use 10-fold cross validation to evaluate them. The results are shown in Table 4.1. The accuracy produced by KNN is the highest among the three methods. We assigned different values to k, and it attained the best performance when k=1.

	Naïve Bayes	K-nearest neighbor	Decision Tree
Accuracy (%)	81.58	84.08	78.46

Table 4.1: Accuracy of phrase polarity prediction by Naïve Bayes, K-nearest neighbor and Decision Tree

If we investigate the confusion matrix of classification by KNN (as shown in Table 4.2), only 9 positive instances are classified as negative and 10 negative instances as positive. It indicates that KNN seldom makes mistake in distinguishing positive with negative instances, which is good for our analysis. However, confusion between neutral instances and positive ones are much more difficult for the classification. We leave this problem for future work. The performance of KNN in phrase polarity prediction is quite promising, so we will utilize KNN in our recursive algorithm to predict sentiment polarity at the sentence level .

		Predicted class		
		neutral	positive	negative
Actual class	neutral	255	41	24
	positive	49	262	9
	negative	20	10	291

Table 4.2: Confusion matrix of phrase polarity classification by K-nearest neighbor

Sentence polarity prediction

To evaluate the performance of sentence-level sentiment polarity prediction by our method, we randomly selected 236 sentences from deletion discussions in AfD. 83 sentences are annotated as positive, 102 as negative and 51 as neutral. We compare the performance of our recursive algorithm with the Stanford sentiment analysis method [33]. In our algorithm, we use a machine learning method to determine the polarity of a phrase with a modifier. We have previously discussed the performance of different machine learning approaches. We use the one with the highest accuracy, which is KNN, in our algorithm. In order to demonstrate whether it improves the system, we test our algorithm in both settings: with and without KNN classification in the step of determining the phrase polarity. The accuracy of 3-class sentiment polarity prediction (positive, negative and neutral) is shown in Table 4.3. The confusion matrix shown in Table 4.4 and recall, precision and F-measure for each category are shown in Table 4.5.

The Stanford sentiment analysis model only achieves an accuracy of 48.73%, while our recursive algorithm with machine learning reaches the highest accuracy of 60.17%.

	Stanford sentiment analysis	recursive algorithm without machine learning method	recursive algorithm with machine learning method
Accuracy (%)	48.73	58.47	60.17

Table 4.3: Accuracy of sentence polarity prediction by Stanford sentiment analysis and our recursive algorithm with and without machine learning.

		Predicted class		
		positive	neutral	negative
Actual class	positive	57	18	8
	neutral	13	28	10
	negative	17	26	59

Table 4.4: Confusion matrix of sentence polarity prediction by recursive algorithm with machine learning method.

And unsurprisingly, the accuracy of our algorithm without machine learning (58.47%) is slightly lower than that with machine learning, which verifies our hypothesis that using the machine learning approach to determine the phrase polarity can improve the overall performance.

The most likely reason for the poor performance by the Stanford sentiment analysis model on this task is the difference between the corpus used to train their model and the corpus used in our evaluation. Their training corpus consists of movie reviews. We suspect that most people would express strong sentiment using certain sentiment words. On the other hand, our task is to analyze deletion discussions in AfD, in which sentiment can be very subtle and implicit. Additionally, the same sentiment word may express a distinct polarity in different contexts. For example *original research* is negative in our corpus, while you might not recognize it as negative in movie reviews. Another example is *horrific*, which can be considered as positive in a horror movie review. However you would hardly classify it as positive in our corpus. As a consequence, using the Stanford sentiment analysis model, which is trained on a quite different corpus, to predict the sentiment polarity of the sentences in our corpus would lead to a low accuracy. We have included it here only to provide a baseline, since, to our knowledge, no other baseline exists for our task.

	Recall	Precision
Positive	0.6867	0.7125
Neutral	0.5490	0.3889
Negative	0.5784	0.7662

Table 4.5: Category-based analysis of sentence polarity prediction by recursive algorithm with machine learning method.

The performance of the Stanford sentiment analysis model on movie reviews is shown in Table 2.6. It has a reported accuracy of 45.7% and 85.4% for sentence-level sentiment prediction in terms of 5-class (very negative, negative, neutral, positive, very positive) and 2-class (negative, positive) predictions, respectively. Although they don't have the result for 3-class prediction as we do, we can infer from their result that an accuracy around 60% produced by our approach (3-class prediction) is reasonable and promising.

4.4 Prototype of a decision making support system

The purpose of a decision making support system is to make it easier for an administrator to review the deletion discussion. Thus, we have designed a prototype, which turns an unstructured discussion into a user-friendly well-structured overview of the discussion. We provide several options for an administrator to choose what to show including representative arguments and those with Wikipedia policies. The prototype is illustrated in Figure 4.8. There are six main step in this prototype:

1. First we need to input a discussion. The user interface (UI) design of this step is shown in Figure 4.9. All arguments in a discussion include the vote (keep, delete, merge, etc.), deliberation and the user name of the person making the argument. Once we have an input, a discussion is ready for analyzing. When we click "Analyze", the system will start analyzing this discussion.

2. The second step is eliminating redundancy in the discussion. All of our analyses are at the sentence level. If a sentence consists of two or more independent clauses, then we treat each clause as a sentence. After that, we compute the similarity

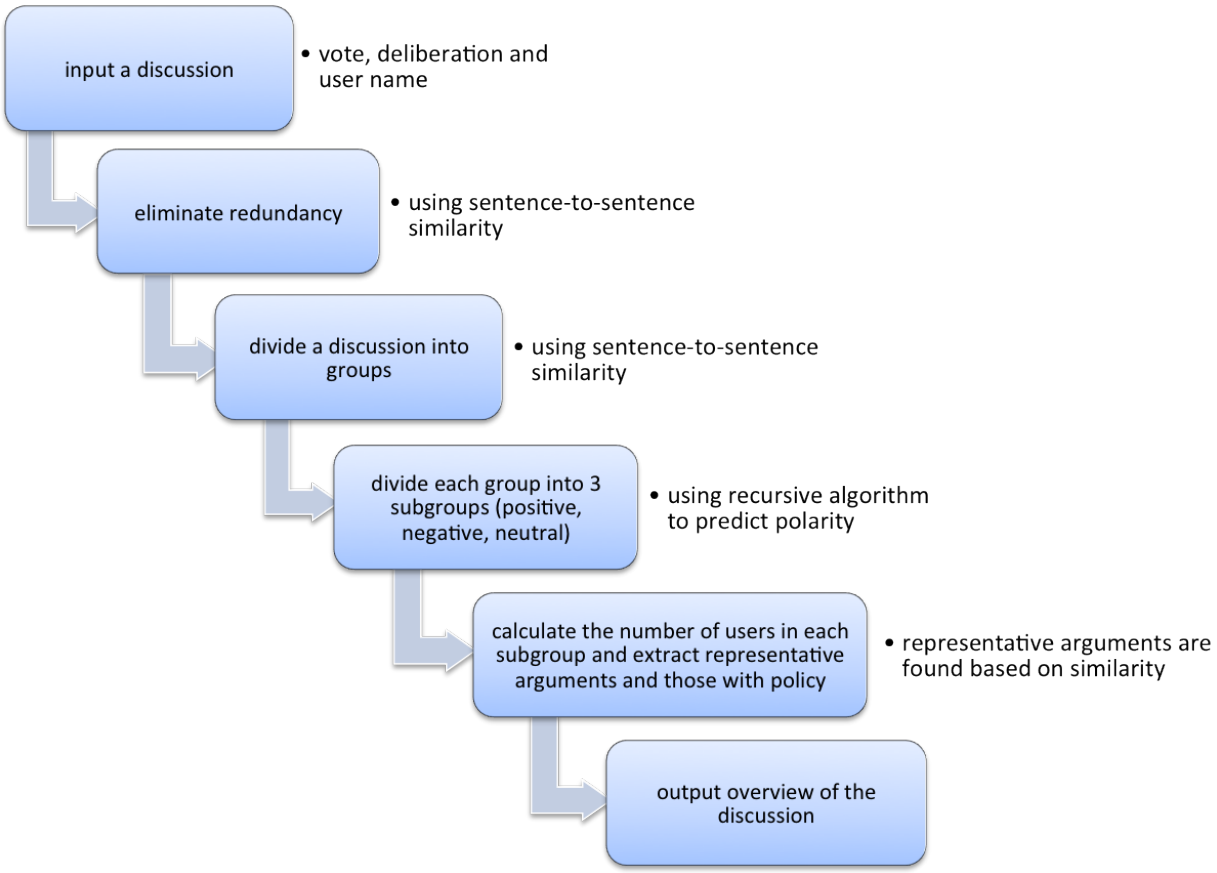


Figure 4.8: Prototype of decision making support system.

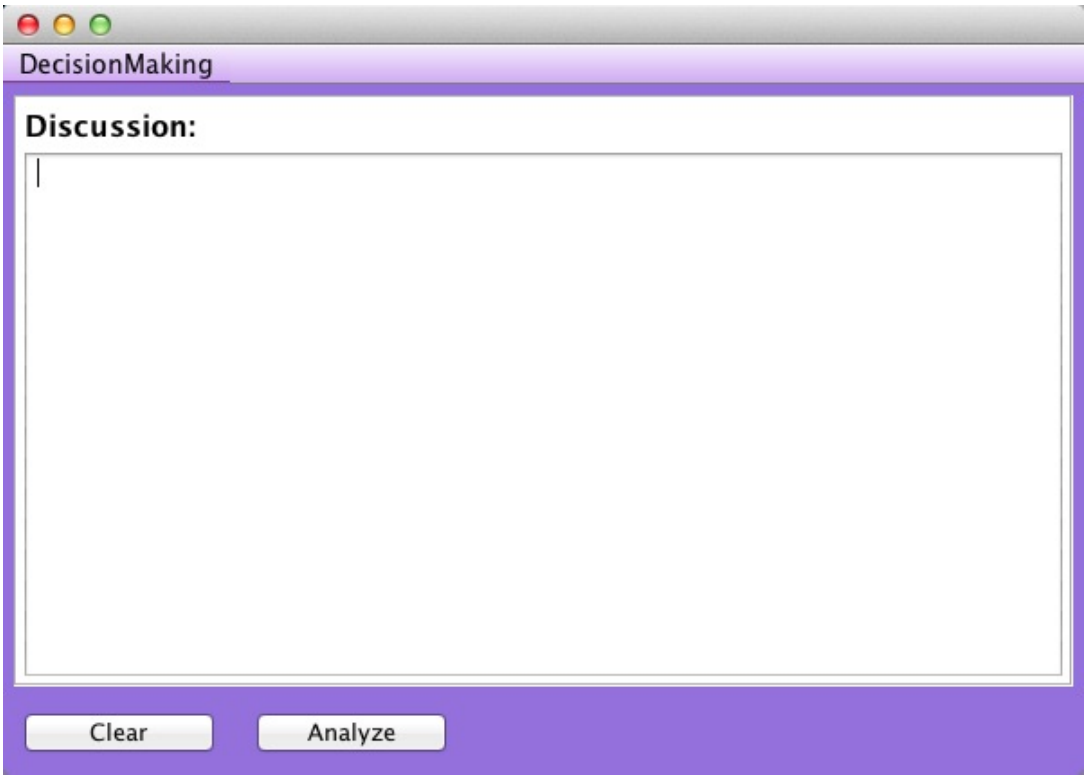


Figure 4.9: UI design of input for decision making support system.

scores between pairs of sentences. If the similarity score is 1, which means they are highly similar or the same, then we remove one of the two sentences. However, we want to record such redundancy from different users so that we can keep track of it in the step calculating the number of users.

3. Once we have the discussion without redundancy, we can divide the discussion into groups based on the similarity between sentences. Since we have computed the similarity score between every pair of sentences in the discussion, we have a similarity matrix. And we can transform the similarity matrix to a dissimilarity matrix by transforming x to 1/x. Then our problem is to cluster a dissimilarity matrix. Thus we can use a clustering method like hierarchical clustering [16] to cluster the sentences into groups. The advantage of hierarchical clustering is that its cluster partitioning can be determined by similarity score; whereas most clustering methods require the number of clusters that you want. Since a long discussion may involves more aspects (i.e., more clusters), clustering based on the fixed number is inappropriate here. Using hierarchical clustering ensure that all the sentences in the same group are similar to each other and dissimilar with ones in other groups. As a consequence, the sentences in the same group are related to a common theme.

4. Now, we have several groups of similar sentences/arguments. We need to further divide each group into 3 subgroups (positive, negative, neutral) based on sentiment analysis, because among the similar sentences, some hold positive opinion, some negative and some neutral. In order to classify these sentences into subgroups by sentiment polarity, we use the polarity prediction method proposed in Section 4.3 to determine the polarity of each sentence.

5. After subgroup division is done, we calculate the number of users who hold positive, negative and neutral opinions separately in each subgroup. And we extract one argument as a representative for each subgroup since sometimes there are too many sentences in one subgroup and it can be time-consuming to read them all. The representative is the sentence that has the highest sum of similarity score with other sentences in the same subgroup. Additionally, we extract all arguments involving

policy, as policy is very important in deletion discussions.

6. Finally, the UI design of the output is shown in Figure 4.10. The system first lists the number of "keep" and "delete" votes. And the overview of the discussion is provided. In particular, arguments with common themes are in the same row, and positive, neutral and negative opinions are shown in different columns. By default, the system only shows one representative argument in each cell and the number of users who support it. In addition, the administrator can choose to show policy-related arguments and all arguments in each column (i.e., positive, neutral and negative).

4.5 Discussion

We evaluated our sentiment analysis algorithm using the corpus from the Wikipedia Article for Deletion (AfD) forum. The comparison of our approach and the Stanford sentiment analysis in analyzing the corpus shows that our approach has a good performance when compared with the state-of-the art Stanford sentiment analysis tool. Like many sentiment analysis tools, the Stanford sentiment analysis is trained on a corpus of movie reviews. Our study shows that the accuracy in sentiment analysis is over 10% higher with our algorithm, as compared to when we used the Stanford tool to analyze the Wikipedia AfD deliberations. This suggests that to achieve a satisfactory performance in sentence-level sentiment analysis of online deliberation content we may need to use a training set that is closer to argumentation data. Our algorithm does not require a large training dataset but achieves a promising performance, which contributes to the research activities in this area.

We presented a prototype of a decision making support system. By using the system, an administrator only needs to copy the discussion from the AfD forum and paste it in the text field as an input to the system. Then after clicking the "Analyze" button, the system provides the administrator with a well-structured overview of the discussion. Having the overview of the discussion, the administrator should quickly grasp the key

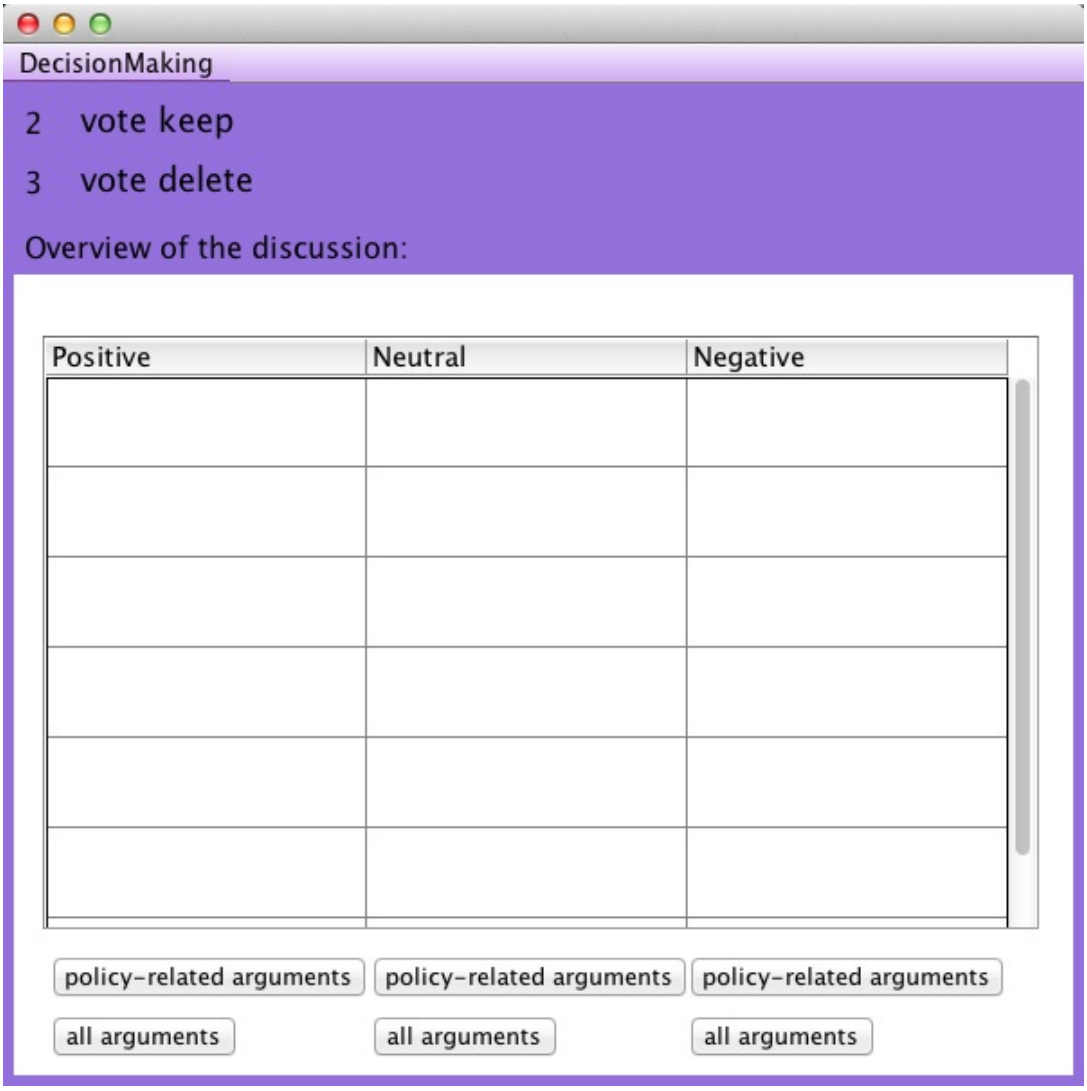


Figure 4.10: UI design of output for decision making support system.

points the discussion involves, and distinct opinions about each point.

The processing time of the system is determined by the length of the discussion. For a middle size discussion which contains 20 to 30 sentences, it takes around 2 minutes.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

In this thesis, we have proposed some approaches to solving online deliberation problems from two perspective: to help users participate in the discussion (knowledge management) and to help administrators review the discussion to help them make their final decisions (decision making). We have focused on article for deletion (AfD) discussions in Wikipedia which is a typical type of online deliberation. To make the processing of deletion discussions more efficient, we have used natural language processing methods. By applying speech act theory to recognize useful information, we have been able to obtain a knowledge repository. By analyzing relationships between text fragments and sentiment in text fragments, a decision making support system has been developed.

In terms of knowledge management, the main question here is what information do we consider important. Generally speaking, the mistakes that have been made frequently in the previous AfD discussions are worth noticing, because we do not want new participants to make the mistakes that have appeared repeatedly in the previous discussion. In particular, we are interested in how to identify one type of speech act, the directive, when it is formed as an imperative sentence.

First, we have analyzed a typical type of imperative, which is formed by using a verb in its base form, normally without a subject. Specifically, we have analyzed the Penn Treebank-style phrase structure and the dependency structure of sentences which

is generated by the Stanford Parser. The basic rule of imperative recognition is to find the sentences with a verb (in its base form) as the root in the phrase structure and this particular verb has no subject child in the dependency structure. Then, we have included the form that a personal pronoun or noun (e.g., you, we, username) followed by a modal verb (e.g., should, must, need) to recognize another subgroup of imperatives.

In our experiments, we asked two human annotators to extract all imperatives and then calculate their agreement. They were trained on annotating our data until they reached an agreement of 0.883 kappa coefficient. All remaining disagreements were eliminated after they discussed the conflicts and reached consensus.

To evaluate the performance of our method, we have compared the result given by our method with the result agreed by the two annotators. Our method produces a high precision of 0.8447 and good recall of 0.7337. Thus, we can say that our approach can effectively recognize imperatives.

To obtain a knowledge repository that can be used to educate new participants, we have curated one week's discussions in each month in the year 2013 and have applied our methods on this corpus to obtain a knowledge repository. Since Wikipedia policy plays a key role in deletion discussions, we have extracted a few policies that are mentioned frequently in the knowledge repository. Given the list of policies being mentioned frequently in our knowledge repository, new participants can participate in the deletion discussion more effectively by reviewing them first.

Deliberation is a type of informal logical communication whose purpose is to rationalize the process of reaching a decision. To reach the decision, people often need to weigh different opinions and rationales expressed in the deliberation. AfD deliberations are numerous and can be lengthy. One foreseen issue in this context is the possible daunting task of reading through all the deliberation content and identifying and evaluating diverse key points and related rationales.

This study is interested in addressing this issue through a computational linguistics approach. We have developed an approach that combines a text-to-text similarity technique with a sentence-level sentiment analysis method. The deliberation content is first divided into groups based on the similarity of texts, then within each group we use

a recursive algorithm to examine the sentiment polarity of each sentence according to the identified similar topic to further classify the sentences into three groups: *positive*, *neutral*, and *negative*.

We have proposed a recursive algorithm to predict the sentiment polarity on sentence level. Instead of just calculating the sum of the polarity of each word in a sentence, we have taken dependency structure into account, since the polarity of a word can be negated or changed by other words such as its modifier or the subject of the sentence. Then we have presented several types of negation including local negation, predicate negation, subject negation, preposition negation and modifier negation. And we have also explained how to detect them in the dependency structure tree.

To evaluate our method, we have compared it with the Stanford sentiment analysis, the state-of-art tool. The result shows that the accuracy of our method is over 10% higher than the Stanford sentiment analysis (which has been trained on a different sentiment corpus). Our method does not require a large training dataset but achieves a quite promising performance.

Finally, we have developed a prototype of the decision making support system. The input is a discussion, and the output is a well-organized overview of the discussion. By using this decision making support system, the administrator should have a clear overview of a discussion and weigh each aspect to make a wise decision.

5.2 Future work

There still are improvements that can be done based on our work. One is a more comprehensive knowledge repository analysis in Chapter 3. In this thesis, we have only developed a list of policies that are frequently mentioned in the knowledge repository. However, there is a large amount of information other than the list of policies that could be extracted from this repository. Some examples from our repository are:

- Please refrain from making personal attacks.
- Remember, notability can't be inherited.

- Please consider providing further rationale, as it is possible that the closer of this discussion may otherwise not provide much weight to your vote, because it's somewhat ambiguous.

These arguments may not involve mentioning specific policies, but they are important and helpful in educating new users and even experienced users. Thus, future work would include the analysis of these arguments in our repository. If we could summarize the repository, that would be beneficial for the Wikipedia community and thereby improve the quality of AfD discussion. Summarization is currently a topic of significant interest in the computational linguistics community.

All of our analyses in this thesis are based at the sentence level. However, the rhetorical structure at a coarser granularity has not been taken into account since it is not easy to automatically generate the rhetorical structure of the texts due to ambiguity and complexity. The current tools did not reach a satisfactory level in our context (i.e., the deletion discussion in AfD). Thus, we need to explore a way to generate rhetorical structure in our context in the future.

We have proposed a prototype of decision making support system and we need to evaluate how it effects the decisions making process and how it helps the administrator in the future. And more factors can be added in the system such as different weights given to new and experienced members.

Although our approaches aim to solve problems in AfD in Wikipedia, it is possible to extend our work to a broader context. And more problems in online deliberation need to be investigated apart from the two perspectives in this thesis. For example, how expertise level of the participants affects the deliberation and the final decision. All in all, we hope this thesis provides inspiration for future research in this area.

Bibliography

[1] John Langshaw Austin. *How to do things with words*, volume 1955. Oxford university press, 1975.

[2] Ngo Xuan Bach, Nguyen Le Minh, and Akira Shimazu. Udrst: A novel system for unlabeled discourse parsing in the rst framework. In *Advances in Natural Language Processing*, pages 250–261. Springer, 2012.

[3] Alexandra Balahur. Sentiment analysis in social media texts. *WASSA 2013*, page 120, 2013.

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[5] Elena Cabrio and Serena Villata. Natural language arguments: A combined approach. In *ECAI*, volume 242, pages 205–210, 2012.

[6] Jean Carletta. Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22(2):249–254, 1996.

[7] Eugene Charniak. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, pages 132–139. Association for Computational Linguistics, 2000.

[8] Eugene Charniak and Mark Johnson. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 173–180. Association for Computational Linguistics, 2005.

[9] Thomas H Davenport. Saving it's soul: Human-centered information management. *Harvard business review*, 72(2):119–31, 1994.

[10] Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454, 2006.

[11] Susan Ervin-Tripp. Is sybil there? the structure of some american english directives. *Language in society*, 5(01):25–66, 1976.

[12] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422, 2006.

[13] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[14] Chung-hye Han. *The structure and interpretation of imperatives: mood and force in Universal Grammar*. Psychology Press, 2000.

[15] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[16] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.

[17] Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics, 2004.

[18] Milen Kouylekov and Matteo Negri. An open-source package for recognizing textual entailment. In *Proceedings of the ACL 2010 System Demonstrations*, pages 42–47. Association for Computational Linguistics, 2010.

[19] J Richard Landis, Gary G Koch, et al. The measurement of observer agreement for categorical data. *biometrics*, 33(1):159–174, 1977.

[20] Nan Li and Desheng Dash Wu. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48(2):354–368, 2010.

[21] William C Mann and Sandra A Thompson. Rhetorical structure theory: A theory of text organization. Technical report, DTIC Document, 1987.

[22] Daniel Marcu. A surface-based approach to identifying discourse markers and elementary textual units in unrestricted texts. In *Proceedings of the COLING-ACL 1998 Workshop on Discourse Relations and Discourse Markers*, pages 1–7, 1998.

[23] Rada Mihalcea, Courtney Corley, and Carlo Strapparava. Corpus-based and knowledge-based measures of text semantic similarity. In *AAAI*, volume 6, pages 775–780, 2006.

[24] Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. Wordnet:: Similarity: measuring the relatedness of concepts. In *Demonstration Papers at HLT-NAACL 2004*, pages 38–41. Association for Computational Linguistics, 2004.

[25] Ellen Riloff and Janyce Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 105–112. Association for Computational Linguistics, 2003.

[26] Vasile Rus and Mihai Lintean. A comparison of greedy and optimal assessment of natural language student input using word-to-word similarity metrics. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 157–162. Association for Computational Linguistics, 2012.

[27] Vasile Rus, Mihai Lintean, Rajendra Banjade, Nobal Niraula, and Dan Stefanescu. Semilar: The semantic similarity toolkit. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 163–168. Citeseer, 2013.

[28] Vasile Rus, Nobal Niraula, and Rajendra Banjade. Similarity measures based on latent dirichlet allocation. In *Computational Linguistics and Intelligent Text Processing*, pages 459–470. Springer, 2013.

[29] Jodi Schneider, Alexandre Passant, and Stefan Decker. Deletion discussions in wikipedia: Decision factors and outcomes. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, page 17. ACM, 2012.

[30] Jodi Schneider, Krystian Samp, Alexandre Passant, and Stefan Decker. Arguments about deletion: how experience improves the acceptability of arguments in ad-hoc online task groups. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1069–1080. ACM, 2013.

[31] John R Searle. A classification of illocutionary acts. *Language in society*, 5(01):1–23, 1976.

[32] J.M.H. Sinclair and M. Coulthard. *Towards an analysis of discourse: the English used by teachers and pupils*. Oxford University Press, 1975.

[33] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1631–1642, 2013.

[34] Sara Owsley Sood, Elizabeth F Churchill, and Judd Antin. Automatic identification of personal insults on social news sites. *Journal of the American Society for Information Science and Technology*, 63(2):270–285, 2012.

[35] Radu Soricut and Daniel Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 149–156. Association for Computational Linguistics, 2003.

[36] Caroline Sporleder and Alex Lascarides. Combining hierarchical clustering and machine learning to predict high-level discourse structure. In *Proceedings of the 20th international conference on Computational Linguistics*, page 43. Association for Computational Linguistics, 2004.

[37] Maite Taboada. Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, 38(4):567–592, 2006.

[38] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

[39] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005.

[40] Lu Xiao and Nicole Askin. What influences online deliberation? a wikipedia study. *Journal of the Association for Information Science and Technology*, 2014.

Appendix A

Entailment in Sentences

Here we list the 80 T-H (Text-Hypothesis) pairs with their actual entailment relations and predicted ones by EDITS.

pair"1" entailment="NO" predicted = "NO"
T: And the suggestion that the battle is not notable is utterly false as this engagement is covered in detail in numerous sources.
H: Neither one of us believes that the battle needs to be covered in this much detail.

pair"2" entailment="YES" predicted = "YES"
T: The article is a disaster and the person is of indeterminable notability.
H: No indication of notability.

pair"3" entailment="YES" predicted = "NO"
T: The articles mention him in passing but are not about him.
H: article fails to explain why this person is notable.

pair"4" entailment="YES" predicted = "YES"
T: Chief executive of a first-level subdivision of a sovereign state is inherently notable.
H: This is a chief executive-related list that is inherently notable as reason from Yk Yk Yk

pair“5” entailment=“YES” predicted = “NO”
T: The film has a strong and not-trivial coverage in Google Books, including books like L’Avventurosa storia del cinema italiano raccontata dai suoi protagonisti, 1960-1969, 007 All’Italiana, Mondo exotica: sounds, visions, obsessions of the cocktail generation, Dizionario dei film italiani stracult, Spionaggio, avventura, eroi moderni.
H: It is widely covered in texts and repertories about Italian genre films.

pair“6” entailment=“NO” predicted = “YES”
T: The suggestion that chemicals such as acetic acid or acetone are an original invention here is preposterous.
H: I was obviously not suggesting that ”acetic acid” itself has been an original invention on Wikipedia!

pair“7” entailment=“NO” predicted = “NO”
T: Sources say that this film is under production; what happens if it is cancelled?
H: ”Cancelled” is a whole different issue, and would likely prevent any article recreation.

pair“8” entailment=“YES” predicted = “YES”
T: Non-notable primary school.
H: Non-notable school.

pair“9” entailment=“NO” predicted = “NO”
T: The only coverage I could find was trivial, passing mentions.
H: Google search turned up nothing promising.

pair“10” entailment=“NO” predicted = “YES”
T: We don’t need articles on specific football line-upspartnerships, that’s just overkill.
H: Our colonial cousins seem very keen on articles on specific line-upspartnerships.

pair“11” entailment=“YES” predicted = “YES”

T: Though some people might take the page as not important and irrelevant to have its own article in Wikipedia, I think the article should not be deleted as because it provides the readers information, regarding who are this top people from the hip hop genre who had greatly shaped the hip hop culture and the hip hop scene yearly.
H: The article has significant or has demonstrated effects in the hip hop culture and of the hip hop entertainment industry.

pair"12" entailment="YES" predicted = "YES"
T: Could be redirected to OpenXMA, the content of which isn't all that different from this article.
H: Redirect to OpenXMA, as suggested.

pair"13" entailment="NO" predicted = "NO"
T: Redirecting the page to the lead actors future projects section will be cool.
H: I don't think it is wise to redirect to the original film.

pair"14" entailment="YES" predicted = "NO"
T: The linked sources do not confirm that there's really anything worth redirecting.
H: I don't think it is wise to redirect to the original film.

pair" pair"15" entailment="NO" predicted = "NO"
T: Clearly too early, as she has only appeared in a few small $25k ITF tournaments, where she always lost in first round.
H: Information in the article is also factually inaccurate, as she is wrongly listed with 2 ITF titles in the infobox.

pair"16" entailment="NO" predicted = "NO"
T: Any news website that has a wide coverage (especially it's in internet and used by Google News as well) will surely make an impact in today's culture.
H: I see many Wordpress blogs in Google News searches so that proves nothing.

pair"17" entailment="NO" predicted = "NO"
T: It greatly summarizes the people's accomplishments grouped into categories: arts, humanities, and business (rationale for the categories are found in the article).
H: Being noted in different articles across the web, will certainly increase the notability of the list and of the blog itself.

pair"18" entailment="YES" predicted = "YES"
T: Being a part of the Google news feed does not show notability.
H: Being in a Google News search doesn't show notability.

pair"19" entailment="NO" predicted = "YES"
T: The notability of the blog itself seems to be debatable and, more pertinantly, the annual list itself doesn't satisfy me as having received significant coverage from multiple reliable sources.
H: But anyways, sources stated above are notable.

pair"20" entailment="NO" predicted = "NO"
T: The article has significant or has demonstrated effects in the hip hop culture and of the hip hop entertainment industry.
H: But anyways, sources stated above are notable.

pair"21" entailment="YES" predicted = "YES"
T: The notability of the blog itself seems to be debatable and, more pertinantly, the annual list itself doesn't satisfy me as having received significant coverage from multiple reliable sources.
H: List compiled by a non-notable blog.

pair"22" entailment="NO" predicted = "NO"
T: There is no justification to splinter the description of Longstreet's second day attacks

into multiple articles.
H: The topic is notable, being covered in numerous sources.

pair"23" entailment="YES" predicted = "NO"
T: There is no justification to splinter the description of Longstreet's second day attacks into multiple articles.
H: I agree with the nominator's statement – there's no need for an attack by a single division to have a separate article.

pair"24" entailment="YES" predicted = "NO"
T: I don't see why we can't cover the basics of the assault on the Second Day page and refer anyone looking for more details to the appropriate source.
H: There is absolutely no reason why we need this level of detailed coverage of the battle.

pair"25" entailment="NO" predicted = "NO"
T: I think the issue here is not if it can be referenced but if this is notable enough to have a seperete article from the Gettysburg article, which I believe is what Hal Jespersen is saying as well.
H: The topic is notable, being covered in numerous sources.

pair"26" entailment="YES" predicted = "NO"
T: As the merge discussion hasn't yet closed and the merge isn't finalised, I think sending this article to AfD is premature, to say the least.
H: This article should never have been nominated for AFD.

pair"27" entailment="NO" predicted = "NO"
T: There don't seem to be any reputable sources at all in this article.
H: Reliable sources found.

pair"28" entailment="YES" predicted = "NO"

T: Most of Cruise's roles aren't notable enough for an article to themselves (and some of them are very well acted) and there's no reason to think that this would be any different.
H: The character is not notible.

pair"29" entailment="NO" predicted = "NO"
T: Subject has been mentioned in passing in books and news but not with significant depth to meet WP:GNG.
H: Founder of two significant companies is sufficient.

pair"30" entailment="YES" predicted = "YES"
T: The Daily Advertiser has at least 50 news articles on the Sturt Mall, locals sources are relevant and as point out by LauraHale, it has some non-local coverage!
H: There are more sources that could be used from The Daily Advertiser.

pair"31" entailment="YES" predicted = "YES"
T: the fact it exists is not in doubt and it will certainly have been described in published books about card games such as this one.
H: It clearly exists.

pair"32" entailment="YES" predicted = "NO"
T: Absolutely fails notability guidelines, not even close.
H: non-notable, unsalvageable.

pair"33" entailment="NO" predicted = "NO"
T: I think it is probably an encyclopedic subject, if structured correctly.
H: So I'm inclined to say that it's probably something we want to consider keeping around.

pair"34" entailment="NO" predicted = "YES"
T: That is an POV not supported by policy, why shouldn't all malls be notable, what if they have had coverage elsewhere?

H: Not all malls should be notable.

pair“35” entailment=“YES” predicted = “YES”

T: It is true that not all malls are notable.

H: Not all malls should be notable.

pair“36” entailment=“NO” predicted = “NO”

T: Sources like this and this and this this this, this this this also help convince me it is notable.

H: I was unable to find coverage in reliable sources that would create one.

pair“37” entailment=“YES” predicted = “NO”

T: Sources like this and this and this this this, this this this also help convince me it is notable.

H: Locals sources are relevant and as point out by LauraHale

pair“38” entailment=“YES” predicted = “YES”

T: Very significant figure in the UK, has been covered in the media many times, and to cap it all has been knighted, which easily meets WP:BIO: ”The person has received a well-known and significant award or honor”.

H: Very notable figure in the UK in his own right.

pair“39” entailment=“YES” predicted = “YES”

T: Very significant figure in the UK, has been covered in the media many times, and to cap it all has been knighted, which easily meets WP:BIO: ”The person has received a well-known and significant award or honor”.

H: Very frequently in the media on numerous issues.

pair“40” entailment=“YES” predicted = “YES”

T: He is already sufficiently mentioned there and, as someone noted above, the serving

Chairman does not have his own article.
H: He's already mentioned there.

pair"41" entailment="NO" predicted = "NO"
T: I'm surprised this article is this short, because there's loads of stuff that could be written about him from reliable sources.
H: Without giving examples of this coverage, your !vote looks a bit like just vouching for it.

pair"42" entailment="NO" predicted = "NO"
T: The fact that current chair of the MCB doesn't have a wiki article doesn't necessarily prove anything more than no-one having got round to writing the article yet.
H: He is already sufficiently mentioned there and, as someone noted above, the serving Chairman does not have his own article.

pair"43" entailment="NO" predicted = "NO"
T: Very significant figure in the UK, has been covered in the media many times, and to cap it all has been knighted.
H: If a knighthood isn't a well-known and significant award or honour then I don't know what is.

pair"44" entailment="NO" predicted = "NO"
T: There may be no sources in googleland, but the ones now in the article satisfy the notability standard.
H: Despite some misgivings, I do assume good faith for this deletion nomination.

pair"45" entailment="NO" predicted = "YES"
T: However, the article fails WP:GNG and WP:NFOOTBALL.
H: Article satisfied both WP:GNG and WP:NFOOTBALL.

pair“46” entailment=“NO” predicted = “YES”

T: Hoax article, the real Rizky Syawaludin is 16-year-old (as mentioned in the source) and there is no other Rizky Syawaludin that has scored 55 goals in Indonesia.

H: Rizky Syawaludin is fake.

pair“47” entailment=“YES” predicted = “NO”

T: Since the exact same content is already present in that article - there is nothing to merge.

H: no need for this as well as it’s pretty much a duplicate.

pair“48” entailment=“NO” predicted = “NO”

T: Since the exact same content is already present in that article - there is nothing to merge.

H: Some information could be merged.

pair“49” entailment=“YES” predicted = “YES”

T: No sources in the article or on the talk page, searched the first 20 pages of results from the search ”Risk Devolution” + ”Warcraft”, nothing resembling a reliable source covering this game mode in any detail.

H: A search returns no significant coverage by reliable sources.

pair“50” entailment=“NO” predicted = “NO”

T: This is a major intersection in a provincial capital that appears to likely be named after the June 5, 1963 demonstrations in Iran, which suggests that it’s a pretty big deal in context.

H: Nothing to show that this is any more notable than any of the millions of other intersections in the world.

pair“51” entailment=“NO” predicted = “NO”

T: I don’t see a compelling source that establishes this as a technical term with a set

meaning.
H: I did a Google Books search to determine if "interpersonal wellness" is used in any contexts independent of that described in the article.

pair"52" entailment="NO" predicted = "NO"
T: Not to be discouraging, but any artist who only sells 100 or less copies of their music (per the article) is still at approximately garage-band level.
H: I cannot find coverage to indicate notability is satisfied.

pair"53" entailment="NO" predicted = "NO"
T: I'm terribly allergic to all these Wikipedia lists and so much blatantly promotional product placement, as it were.
H: Nike gets enough advertising and articles like this add nothing to Wikipedia.

pair"54" entailment="NO" predicted = "NO"
T: And we have the information and hence we can make article is not a valid reason.
H: From my part and view, this is enough for us to have information of him on our open wikipedia.

pair"55" entailment="YES" predicted = "NO"
T: Not all dance competition winners are notable, regardless sometimes how much attention they receive.
H: Fails notablity criterias, only known for winniing a TV show competition.

pair"56" entailment="NO" predicted = "NO"
T: Searching Google News and Google News Archives for the same terms also yielded no evidence of notability.
H: There is no significant coverage in independent reliable sources to indicate that the subject meets general inclusion criteria, or that specific for creative people.

pair"57" entailment="YES" predicted = "NO"
T: Without independent (and in this case more important reliably published) sources that cover the subject in non-trivial detail, he does not pass WP:GNG.
H: Delete per failure to meet WP:GNG.

pair"58" entailment="YES" predicted = "YES"
T: I looked all the way to page 20 without finding any reliable, third-party sources.
H: No coverage in independent reliable sources.

pair"59" entailment="NO" predicted = "NO"
T: With multiple secondary sources already extant in the article, the topic of the article is notable and the article should be kept.
H: Accepted from AfC because Drawbridge is extensively covered in several reputable national news sources, for example Forbes and the BBC.

pair"60" entailment="NO" predicted = "NO"
T: There are tons of references available to verify each of those facts, and we do tend to have articles about American major league sports owners.
H: Multimillionaire developer and former owner of two professional sports teams(Oakland Athletics and Seattle Seahawks) is likely to have the quantity and quality of secondary coverage needed to satisfy WP:BIO.

pair"61" entailment="NO" predicted = "NO"
T: The article is in piss poor shape, but as Edison says, there's plenty of coverage out there.
H: Recent cleanup work and sourcing (added since nom) demonstrates notability and solves BLP issue.

pair"62" entailment="YES" predicted = "NO"
T: if you had checked Google Scholar, you would see that the top result has 13,311 cita-

tions.
H: Very high cites in GS.

pair"63" entailment="NO" predicted = "NO"
T: The problem with this article is that it is a mere stub, and tells us nothing of what he did as Commissioner.
H: This is admittedly borderline, but I think commissioner is an important enough position for its holders to be inherently notable.

pair"64" entailment="YES" predicted = "NO"
T: GBooks and GScholar show multiple sources to confirm the significance of the Goodyear Silents sports teams in deaf culture.
H: There are plenty of references.

pair"65" entailment="NO" predicted = "NO"
T: I couldn't find reliable secondary references for notability, but YouTube videos and entries at tv.com show the topic as verifiable.
H: quite frankly, I can't really think of any characters from the show that would meet notability

pair"66" entailment="YES" predicted = "YES"
T: Critical reviews in gaming sites have traditionally been sufficient to meet WP:N and there has been several AFD's where a video game article has been kept due to reviews from notable gaming sites.
H: The reviews make the videogame notable.

pair"67" entailment="NO" predicted = "NO"
T: I have added various references to the article.
H: The sources brought forward by Mcewan would be sufficient on their own.

pair"68" entailment="NO" predicted = "NO"

T: It may be a verified secondary school but the Wikipedia requirement for notability is to have reliable sources.

H: As an Afrikaans-language school it is not realistic to expect Google hits in English.

pair"69" entailment="YES" predicted = "YES"

T: Experience shows that with enough local research high schools invariably can be made to meet WP:ORG.

H: Almost all secondary schools can be found to be notable if enough research is done.

pair"70" entailment="NO" predicted = "NO"

T: Considering that this is an Afrikaans-language school unlikely to have extensive coverage in English, those of us who work primarily in English may have difficulty finding sources on this school.

H: I've picked up around the web indicate that this is an Afrikaans-language school (thus, there may not be a lot of content about it in English) and a boarding school, largely enrolling rural children from a large area.

pair"71" entailment="NO" predicted = "NO"

T: A valid stub article waiting to be expanded.

H: Nomination does not state a valid basis for deletion.

pair"72" entailment="NO" predicted = "NO"

T: The fact the article is a tiny stub is unimportant as regards notability.

H: Keep as stub unless someone demonstrates that this isn't a real place.

pair"73" entailment="NO" predicted = "NO"

T: Some unreleased songs from artists may not be notable, but her's clearly are.

H: Entries here do not need to be notable recordings, they only need to be a part of Spears's career which is verifiable.

pair"74" entailment="YES" predicted = "YES"

T: I can say that there are enough secondary sources for this to pass GNG muster.

H: Adequate secondary sources.

pair"75" entailment="YES" predicted = "NO"

T: it's unsourced and also looks like a hoax.

H: No source given.

pair"76" entailment="NO" predicted = "NO"

T: The creator's user page reveals that they are highly imaginative.

H: I'm not quite willing to call this a hoax, as there do appear to be Tamil language sources found when searching in that language.

pair"77" entailment="NO" predicted = "NO"

T: The overall topic has received significant coverage in reliable sources and passes Wikipedia's General notability guideline.

H: Plus, this article are still young and need more attention from other editors to expand it.

pair"78" entailment="NO" predicted = "NO"

T: Although there is still an issue with an over-abundance of individual articles on every Nortel product, a list article on their overall product line is entirely reasonable.

H: Nortel and its products are notable given their prominent role in telecommunications infrastructure.

pair"79" entailment="YES" predicted = "YES"

T: social media, forums and blogs are not considered reliable sources and a lack of reliable sources means this subject fails WP:GNG.

H: no reliable secondary sources.

pair"80" entailment="NO" predicted = "NO"

T: I'm not familiar enough yet with policy on inherent notability or otherwise of pretenders to extinct titles to decide keep or delete.

H: The reason he is notable is because he is head of the Imperial House of France, in that capacity you will find he is always referred to as a Prince in sources.

Appendix B

Similarity in Sentences

Here we list the 80 pairs of sentences with their actual similarity and predicted similarity score by LSA and LDA in SEMILAR.

pair"1" similarity="YES" LSA = 0.4 LDA = 0.4
T: And the suggestion that the battle is not notable is utterly false as this engagement is covered in detail in numerous sources.
H: Neither one of us believes that the battle needs to be covered in this much detail.

pair"2" similarity="YES" LSA = 0.29 LDA =0.29
T: The article is a disaster and the person is of indeterminable notability.
H: No indication of notability.

pair"3" similarity="YES" LSA = 0.33 LDA = 0.25
T: The articles mention him in passing but are not about him.
H: article fails to explain why this person is notable.

pair"4" similarity="YES" LSA = 0.38 LDA = 0.38
T: Chief executive of a first-level subdivision of a sovereign state is inherently notable.
H: This is a chief executive-related list that is inherently notable as reason from Yk Yk Yk

pair“5” similarity=“YES” LSA = 0.05 LDA = 0.05

T: The film has a strong and not-trivial coverage in Google Books, including books like L'Avventurosa storia del cinema italiano raccontata dai suoi protagonisti, 1960-1969, 007 All'Italiana, Mondo exotica: sounds, visions, obsessions of the cocktail generation, Dizionario dei film italiani stracult, Spionaggio, avventura, eroi moderni.

H: It is widely covered in texts and repertories about Italian genre films.

pair“6” similarity=“YES” LSA = 0.67 LDA = 0.67

T: The suggestion that chemicals such as acetic acid or acetone are an original invention here is preposterous.

H: I was obviously not suggesting that ”acetic acid” itself has been an original invention on Wikipedia!

pair“7” similarity=“NO” LSA = 0.2 LDA = 0.2

T: Sources say that this film is under production; what happens if it is cancelled?

H: ”Cancelled” is a whole different issue, and would likely prevent any article recreation.

pair“8” similarity=“YES” LSA = ‘0.8 LDA = 0.8

T: Non-notable primary school.

H: Non-notable school.

pair“9” similarity=“NO” LSA = 0 LDA = 0

T: The only coverage I could find was trivial, passing mentions.

H: Google search turned up nothing promising.

pair“10” similarity=“YES” LSA = 0.46 LDA = 0.46

T: We don't need articles on specific football line-upspartnerships, that's just overkill.

H: Our colonial cousins seem very keen on articles on specific line-upspartnerships.

pair“11” similarity=“YES” LSA = 0.36 LDA = 0.32

T: Though some people might take the page as not important and irrelevant to have its own article in Wikipedia, I think the article should not be deleted as because it provides the readers information, regarding who are this top people from the hip hop genre who had greatly shaped the hip hop culture and the hip hop scene yearly.
H: The article has significant or has demonstrated effects in the hip hop culture and of the hip hop entertainment industry.

pair"12" similarity="YES" LSA = 0.5 LDA = 0.5
T: Could be redirected to OpenXMA, the content of which isn't all that different from this article.
H: Redirect to OpenXMA, as suggested.

pair"13" similarity="YES" LSA = 0.15 LDA = 0.15
T: Redirecting the page to the lead actors future projects section will be cool.
H: I don't think it is wise to redirect to the original film.

pair"14" similarity="YES" LSA = 0.24 LDA = 0.18
T: The linked sources do not confirm that there's really anything worth redirecting.
H: I don't think it is wise to redirect to the original film.

pair" pair"15" similarity="NO" LSA = 0.16 LDA = 0.13
T: Clearly too early, as she has only appeared in a few small $25k ITF tournaments, where she always lost in first round.
H: Information in the article is also factually inaccurate, as she is wrongly listed with 2 ITF titles in the infobox.

pair"16" similarity="YES" LSA = 0.21 LDA = 0.21
T: Any news website that has a wide coverage (especially it's in internet and used by Google News as well) will surely make an impact in today's culture.
H: I see many Wordpress blogs in Google News searches so that proves nothing.

pair"17" similarity="NO" LSA = 0.1 LDA = 0.1

T: It greatly summarizes the people's accomplishments grouped into categories: arts, humanities, and business (rationale for the categories are found in the article).

H: Being noted in different articles across the web, will certainly increase the notability of the list and of the blog itself.

pair"18" similarity="YES" LSA = 0.67 LDA = 0.67

T: Being a part of the Google news feed does not show notability.

H: Being in a Google News search doesn't show notability.

pair"19" similarity="YES" LSA = 0.22 LDA = 0.22

T: The notability of the blog itself seems to be debatable and, more pertinantly, the annual list itself doesn't satisfy me as having received significant coverage from multiple reliable sources.

H: But anyways, sources stated above are notable.

pair"20" similarity="NO" LSA = 0 LDA = 0

T: The article has significant or has demonstrated effects in the hip hop culture and of the hip hop entertainment industry.

H: But anyways, sources stated above are notable.

pair"21" similarity="YES" LSA = 0.22 LDA = 0.22

T: The notability of the blog itself seems to be debatable and, more pertinantly, the annual list itself doesn't satisfy me as having received significant coverage from multiple reliable sources.

H: List compiled by a non-notable blog.

pair"22" similarity="NO" LSA = 0 LDA = 0

T: There is no justification to splinter the description of Longstreet's second day attacks

into multiple articles.
H: The topic is notable, being covered in numerous sources.

pair"23" similarity="YES" LSA = 0.32 LDA = 0.32
T: There is no justification to splinter the description of Longstreet's second day attacks into multiple articles.
H: I agree with the nominator's statement – there's no need for an attack by a single division to have a separate article.

pair"24" similarity="YES" LSA = 0.11 LDA = 0.11
T: I don't see why we can't cover the basics of the assault on the Second Day page and refer anyone looking for more details to the appropriate source.
H: There is absolutely no reason why we need this level of detailed coverage of the battle.

pair"25" similarity="NO" LSA = 0.13 LDA = 0.13
T: I think the issue here is not if it can be referenced but if this is notable enough to have a seperete article from the Gettysburg article, which I believe is what Hal Jespersen is saying as well.
H: The topic is notable, being covered in numerous sources.

pair"26" similarity="YES" LSA = 0.29 LDA = 0.29
T: As the merge discussion hasn't yet closed and the merge isn't finalised, I think sending this article to AfD is premature, to say the least.
H: This article should never have been nominated for AFD.

pair"27" similarity="YES" LSA = 0.29 LDA = 0.29
T: There don't seem to be any reputable sources at all in this article.
H: Reliable sources found.

pair"28" similarity="YES" LSA = 0 LDA = 0

T: Most of Cruise's roles aren't notable enough for an article to themselves (and some of them are very well acted) and there's no reason to think that this would be any different.
H: The character is not notible.

pair"29" similarity="NO" LSA = 0.14 LDA = 0.14
T: Subject has been mentioned in passing in books and news but not with significant depth to meet WP:GNG.
H: Founder of two significant companies is sufficient.

pair"30" similarity="YES" LSA = 0.38 LDA = 0.38
T: The Daily Advertiser has at least 50 news articles on the Sturt Mall, locals sources are relevant and as point out by LauraHale, it has some non-local coverage!
H: There are more sources that could be used from The Daily Advertiser.

pair"31" similarity="YES" LSA = 0.25 LDA = 0.25
T: the fact it exists is not in doubt and it will certainly have been described in published books about card games such as this one.
H: It clearly exists.

pair"32" similarity="YES" LSA = 0 LDA = 0
T: Absolutely fails notability guidelines, not even close.
H: non-notable, unsalvageable.

pair"33" similarity="NO" LSA = 0.18 LDA = 0.18
T: I think it is probably an encyclopedic subject, if structured correctly.
H: So I'm inclined to say that it's probably something we want to consider keeping around.

pair"34" similarity="YES" LSA = 0.4 LDA = 0.4
T: That is an POV not supported by policy, why shouldn't all malls be notable, what if they have had coverage elsewhere?

H: Not all malls should be notable.

pair“35” similarity=“YES” LSA = 0.8 LDA = 0.8

T: It is true that not all malls are notable.

H: Not all malls should be notable.

pair“36” similarity=“YES” LSA = 0.22 LDA = 0.22

T: Sources like this and this and this this this, this this this also help convince me it is notable.

H: I was unable to find coverage in reliable sources that would create one.

pair“37” similarity=“YES” LSA = 0.25 LDA = 0.25

T: Sources like this and this and this this this, this this this also help convince me it is notable.

H: Locals sources are relevant and as point out by LauraHale

pair“38” similarity=“YES” LSA = 0.18 LDA = 0.18

T: Very significant figure in the UK, has been covered in the media many times, and to cap it all has been knighted, which easily meets WP:BIO: ”The person has received a well-known and significant award or honor”.

H: Very notable figure in the UK in his own right.

pair“39” similarity=“YES” LSA = 0.09 LDA = 0.09

T: Very significant figure in the UK, has been covered in the media many times, and to cap it all has been knighted, which easily meets WP:BIO: ”The person has received a well-known and significant award or honor”.

H: Very frequently in the media on numerous issues.

pair“40” similarity=“YES” LSA = 0.22 LDA = 0.22

T: He is already sufficiently mentioned there and, as someone noted above, the serving

Chairman does not have his own article.

H: He's already mentioned there.

pair"41" similarity="NO" LSA = 0 LDA = 0

T: I'm surprised this article is this short, because there's loads of stuff that could be written about him from reliable sources.

H: Without giving examples of this coverage, your !vote looks a bit like just vouching for it.

pair"42" similarity="NO" LSA = 0.1 LDA = 0.1

T: The fact that current chair of the MCB doesn't have a wiki article doesn't necessarily prove anything more than no-one having got round to writing the article yet.

H: He is already sufficiently mentioned there and, as someone noted above, the serving Chairman does not have his own article.

pair"43" similarity="YES" LSA = 0.13 LDA = 0.13

T: Very significant figure in the UK, has been covered in the media many times, and to cap it all has been knighted.

H: If a knighthood isn't a well-known and significant award or honour then I don't know what is.

pair"44" similarity="NO" LSA = 0 LDA = 0

T: There may be no sources in googleland, but the ones now in the article satisfy the notability standard.

H: Despite some misgivings, I do assume good faith for this deletion nomination.

pair"45" similarity="YES" LSA = 0.75 LDA = 0.75

T: However, the article fails WP:GNG and WP:NFOOTBALL.

H: Article satisfied both WP:GNG and WP:NFOOTBALL.

pair“46” similarity=“YES” LSA = 0.24 LDA = 0.24

T: Hoax article, the real Rizky Syawaludin is 16-year-old (as mentioned in the source) and there is no other Rizky Syawaludin that has scored 55 goals in Indonesia.

H: Rizky Syawaludin is fake.

pair“47” similarity=“YES” LSA = 0 LDA = 0

T: Since the exact same content is already present in that article - there is nothing to merge.

H: no need for this as well as it’s pretty much a duplicate.

pair“48” similarity=“YES” LSA = 0.33 LDA = 0.33

T: Since the exact same content is already present in that article - there is nothing to merge.

H: Some information could be merged.

pair“49” similarity=“YES” LSA = 0.23 LDA = 0.23

T: No sources in the article or on the talk page, searched the first 20 pages of results from the search ”Risk Devolution” + ”Warcraft”, nothing resembling a reliable source covering this game mode in any detail.

H: A search returns no significant coverage by reliable sources.

pair“50” similarity=“YES” LSA = 0.1 LDA = 0.1

T: This is a major intersection in a provincial capital that appears to likely be named after the June 5, 1963 demonstrations in Iran, which suggests that it’s a pretty big deal in context.

H: Nothing to show that this is any more notable than any of the millions of other intersections in the world.

pair“51” similarity=“NO” LSA = 0 LDA = 0

T: I don’t see a compelling source that establishes this as a technical term with a set

meaning.
H: I did a Google Books search to determine if "interpersonal wellness" is used in any contexts independent of that described in the article.

pair"52" similarity="NO" LSA = 0 LDA = 0
T: Not to be discouraging, but any artist who only sells 100 or less copies of their music (per the article) is still at approximately garage-band level.
H: I cannot find coverage to indicate notability is satisfied.

pair"53" similarity="YES" LSA = 0.14 LDA = 0.14
T: I'm terribly allergic to all these Wikipedia lists and so much blatantly promotional product placement, as it were.
H: Nike gets enough advertising and articles like this add nothing to Wikipedia.

pair"54" similarity="YES" LSA = 0.25 LDA = 0.25
T: And we have the information and hence we can make article is not a valid reason.
H: From my part and view, this is enough for us to have information of him on our open wikipedia.

pair"55" similarity="YES" LSA = 0.29 LDA = 0.29
T: Not all dance competition winners are notable, regardless sometimes how much attention they receive.
H: Fails notablity criterias, only known for winniing a TV show competition.

pair"56" similarity="NO" LSA = 0.03 LDA = 0
T: Searching Google News and Google News Archives for the same terms also yielded no evidence of notability.
H: There is no significant coverage in independent reliable sources to indicate that the subject meets general inclusion criteria, or that specific for creative people.

pair"57" similarity="YES" LSA = 0,25 LDA = 0.25

T: Without independent (and in this case more important reliably published) sources that cover the subject in non-trivial detail, he does not pass WP:GNG.

H: Delete per failure to meet WP:GNG.

pair"58" similarity="YES" LSA = 0.36 LDA = 0.36

T: I looked all the way to page 20 without finding any reliable, third-party sources.

H: No coverage in independent reliable sources.

pair"59" similarity="NO" LSA = 0.09 LDA = 0.09

T: With multiple secondary sources already extant in the article, the topic of the article is notable and the article should be kept.

H: Accepted from AfC because Drawbridge is extensively covered in several reputable national news sources, for example Forbes and the BBC.

pair"60" similarity="NO" LSA = 0.14 LDA = 0.14

T: There are tons of references available to verify each of those facts, and we do tend to have articles about American major league sports owners.

H: Multimillionaire developer and former owner of two professional sports teams(Oakland Athletics and Seattle Seahawks) is likely to have the quantity and quality of secondary coverage needed to satisfy WP:BIO.

pair"61" similarity="NO" LSA = 0 LDA = 0

T: The article is in piss poor shape, but as Edison says, there's plenty of coverage out there.

H: Recent cleanup work and sourcing (added since nom) demonstrates notability and solves BLP issue.

pair"62" similarity="YES" LSA = 0 LDA = 0

T: if you had checked Google Scholar, you would see that the top result has 13,311 cita-

tions.

H: Very high cites in GS.

pair“63” similarity=“NO” LSA = 0.17 LDA = 0,17

T: The problem with this article is that it is a mere stub, and tells us nothing of what he did as Commissioner.

H: This is admittedly borderline, but I think commissioner is an important enough position for its holders to be inherently notable.

pair“64” similarity=“YES” LSA = 0 LDA = 0

T: GBooks and GScholar show multiple sources to confirm the significance of the Goodyear Silents sports teams in deaf culture.

H: There are plenty of references.

pair“65” similarity=“NO” LSA = 0.24 LDA = 0.24

T: I couldn’t find reliable secondary references for notability, but YouTube videos and entries at tv.com show the topic as verifiable.

H: quite frankly, I can’t really think of any characters from the show that would meet notability

pair“66” similarity=“YES” LSA = 0.18 LDA = 0.18

T: Critical reviews in gaming sites have traditionally been sufficient to meet WP:N and there has been several AFD’s where a video game article has been kept due to reviews from notable gaming sites.

H: The reviews make the videogame notable.

pair“67” similarity=“NO” LSA = 0 LDA = 0

T: I have added various references to the article.

H: The sources brought forward by Mcewan would be sufficient on their own.

pair"68" similarity="YES" LSA = 0.13 LDA = 0.13

T: It may be a verified secondary school but the Wikipedia requirement for notability is to have reliable sources.

H: As an Afrikaans-language school it is not realistic to expect Google hits in English.

pair"69" similarity="YES" LSA = 0.31 LDA = 0.31

T: Experience shows that with enough local research high schools invariably can be made to meet WP:ORG.

H: Almost all secondary schools can be found to be notable if enough research is done.

pair"70" similarity="NO" LSA = 0.30 LDA = 0.30

T: Considering that this is an Afrikaans-language school unlikely to have extensive coverage in English, those of us who work primarily in English may have difficulty finding sources on this school.

H: I've picked up around the web indicate that this is an Afrikaans-language school (thus, there may not be a lot of content about it in English) and a boarding school, largely enrolling rural children from a large area.

pair"71" similarity="NO" LSA = 0.22 LDA = 0.22

T: A valid stub article waiting to be expanded.

H: Nomination does not state a valid basis for deletion.

pair"72" similarity="NO" LSA = 0.18 LDA = 0.18

T: The fact the article is a tiny stub is unimportant as regards notability.

H: Keep as stub unless someone demonstrates that this isn't a real place.

pair"73" similarity="YES" LSA = 0.33 LDA = 0.33

T: Some unreleased songs from artists may not be notable, but her's clearly are.

H: Entries here do not need to be notable recordings, they only need to be a part of Spears's career which is verifiable.

pair"74" similarity="YES" LSA = 0.5 LDA = 0.5

T: I can say that there are enough secondary sources for this to pass GNG muster.

H: Adequate secondary sources.

pair"75" similarity="YES" LSA = 0 LDA = 0

T: it's unsourced and also looks like a hoax.

H: No source given.

pair"76" similarity="NO" LSA = 0 LDA = 0

T: The creator's user page reveals that they are highly imaginative.

H: I'm not quite willing to call this a hoax, as there do appear to be Tamil language sources found when searching in that language.

pair"77" similarity="NO" LSA = 0 LDA = 0

T: The overall topic has received significant coverage in reliable sources and passes Wikipedia's General notability guideline.

H: Plus, this article are still young and need more attention from other editors to expand it.

pair"78" similarity="NO" LSA = 0.2 LDA = 0.2

T: Although there is still an issue with an over-abundance of individual articles on every Nortel product, a list article on their overall product line is entirely reasonable.

H: Nortel and its products are notable given their prominent role in telecommunications infrastructure.

pair"79" similarity="YES" LSA = 0.22 LDA = 0.22

T: social media, forums and blogs are not considered reliable sources and a lack of reliable sources means this subject fails WP:GNG.

H: no reliable secondary sources.

pair"80" similarity="NO" LSA = 0.10 LDA = 0.10

T: I'm not familiar enough yet with policy on inherent notability or otherwise of pretenders to extinct titles to decide keep or delete.

H: The reason he is notable is because he is head of the Imperial House of France, in that capacity you will find he is always referred to as a Prince in sources.

Curriculum Vitae

Name:	Wanting Mao
Post-Secondary Education and Degrees:	University of Western Ontario London, ON 2012 - present M.Sc candidate Xidian University Xi'an, China 2008 - 2012 B.Eng
Related Work Experience:	Teaching Assistant and Research Assistant The University of Western Ontario 2012 - 2013