Exploring the Role of Named Entities for Uncertainty Recognition in Event Detection

Masnizah Mohd and Kiyoaki Shirai
Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan

Keywords: Uncertainty, Named Entities, User, Event Detection.

Abstract: Ambiguous information contributes to the uncertainty issue. Type of information such as using named entities has been proved to provide significant information to the user compared to the 'bag-of-words' in identifying an event. So what else could contribute to the uncertainty in an event detection? We propose to answer this question by analysing the distribution of named entities across topics, and explore the potential of named entities in a user experiment. We construct an event detection task with 20 users and use news dataset from Topic Detection and Tracking (TDT) corpus, under the Sports and Politics categories. We analyse the results from five uncertainty dimensions: too little information, too much information, complex information, ambiguous information and conflicting information. These dimensions are categorise as two factors; amount and type of information. There was no statistical significance difference in the amount of information given with the number of successful event detected. However, with little information and high named entities has contributes in reducing uncertainty. In addition, the amount of information and information quality are mutually independent. Our results suggest that uncertainty vary substantially between the amount of information and type of information in event detection.

1 INTRODUCTION

"Uncertainty" in event detection and tracking task can be interpreted as lack of information and inability to interpret or determine an event due to the little or too much information. This task rely on named entities as one of the important features used to detect an event which occurs in a topic.

The objective of this experiment is to explore the potential of named entities (NEs) for uncertainty in event detection. Therefore we take two approaches. The former, proposed by Mohd and Mabrook (2014) in the context of Topic Detection and Tracking (TDT) systems, proved that named entities was useful in improving Tracking task (including sub Profiling activity) performance meanwhile *bag of words* (BOW) improved user performance in Detection task. Therefore type of information either NEs or BOW were considered. However the potential role of NEs in reducing or increasing uncertainty in TDT has never been explored.

In the second approach, we go beyond 'type of information' by also considering the 'amount of information or stories' provided. Previously, Hurley et al., (2011) evaluated uncertainty in online news focusing on cancer topic from 5 dimensions; too little information, too much information, ambiguous information, complex information and conflicting information. Therefore in this experiment we considered the 5 dimensions of uncertainty introduced by Hurley et al., (2011) with the type of information. We want to investigate the potential role of NEs or BOW for uncertainty recognition in event detection.

These two approaches are important to validate research questions:

- Is there any relationship between the 'type of information' and 'amount of information' for uncertainty recognition in event detection?
- What is the 'type of information' that are considered as complex, ambiguous and conflicting to the user?

We conducted a set of user experiments that concern various kinds of entities (e.g. Person, Location, Organization, Date, Time, Money, and Percent) across topics (Politic, Sports). Our experiment include 1000 evaluations of news stories.

This paper is organized as follows. Section 2 describes related work and positions our approach.

Mohd, M. and Shirai, K..
Exploring the Role of Named Entities for Uncertainty Recognition in Event Detection.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 3: KMIS*, pages 335-341
ISBN: 978-989-758-158-8

Section 3 discuss the methodology applied to construct the user experiment. Finally in Section 4, we report the findings. We end the paper with conclusions and thoughts for future work.

2 RELATED WORK

Uncertainty is one of the challenges in information seeking and retrieval (Chowdhury et al., 2011). Many attempts have been done in developing uncertainty model by investigating human information behaviour in information seeking and retrieval process (Ingwersen, 1992). There are few work that proposed natural language processing technique such as from syntactical and semantic approach to reduce uncertainty (Goodman, 2008; Topka, 2013).

Several linguistic research aim at modelling the use of modality, but very few concentrate on uncertainty, for instance the Certainty Categorization Model proposed by Rubin (2006). This model was based on four dimension; Level, Perspective, Focus and Time to characterize uncertainty. For level dimension, they considered the words such as 'might buy' and 'will come' to be classified as Absolute level or Low level. Meanwhile in Perspective level, they analysed on how the sentences are reported from writer's point of view. Focus dimension differentiated between Abstract and Factual information. Finally for Time dimension, they analysed the sentences based on past, present and future time. Then Goujon (2009) enhanced the Certainty Categorization Model proposed. The enhanced model includes the identification of the local source, which was important to the end user in validating the reliability of the reported discourse. It also takes into account the reality and unreality of an information which was specified in the source text, rather than the Focus dimension. Thus the enhanced dimensions consist of five; Level, Perspective, Time, Reality and Source Name to characterize uncertainty.

There are also few work in measuring uncertainty in message. Mishel (1988) has introduced forms of uncertainty (ambiguity, complexity, volume of information and unpredictability) and Babrow (1998) dimensions of uncertainty were combined to form five forms of dimensions of uncertainty in messages. Instances of uncertainty related content within a message are such as message characteristic (specific words, phrases or sentences). Then Hurley (2011) enhanced the dimension of uncertainty into five dimensions: too little information (volume), too much information (volume), complex information, ambiguous information and conflicting information. These five forms of uncertainty in messages was easily identified in news article and been implemented in cancer news article.

In the context of TDT research, researchers have attempted to build better document models, developing similarity metrics or better document representations (Chen and Ku, 2002). This has led to a series of research efforts that concentrate on improving document representation by applying Named Entity Recognition (Chen and Ku, 2002). Mohd and Mabrook (2014) investigated the potential of named entities in TDT tasks and they discovered that NEs has improved both tasks. However there is no work has evaluate the role of NEs for uncertainty recognition in event detection task. This is the first work that explored the five dimensions of uncertainty in TDT.

3 METHOD

There are two approaches in this work. First we analysed the distribution of named entities (NEs) across topics (Section 3.2) and secondly we conducted a user experiment (Section 3.3 - 3.4) to explore the potential of named entities for uncertainty recognition in event detection task.

3.1 Dataset

We used 300 news documents from Topic Detection and Tracking (TDT) corpus. There are 2 categories (Politics and Sports) with 10 topics and 50 events occurred as shown in Table 1. On average, there are 5 events and 30 documents/story per topic. In TDT, a topic consist of several events and an event consist of several stories or documents.

Table 1: Topics and events for Politics and Sports categories.

Topic: [P1] Current Conflict with Iraq (20015)
Event • Current Conflict with Iraq • Iraq announces it will block inspections • Iraq prevents inspection team from entering • Reaction to blocked inspection team • Inspection team withdrawn • Hussein may stop cooperating with inspections
Topic: [P2] Clinton-Jiang Debate (20096)
Event • Plans, preparations for Clinton's trip to China • Clinton leaves for China • Clinton's activities in China • Freedom of worship for Chinese citizens • Reaction to Clinton's trip

Table 1: Topics and events for Politics and Sports categories. (cont.)

Topic: [P3] Gingrich Resigns (30024)
Event • Reaction to elections, Gingrich faces challenge to speakership • Largent, Livingston to challenge GOP leaders • Gingrich announces he will resign • Reaction to, reflection on Gingrich resignation • Candidates emerge for speakership
Topic: [P4] US Mid-term Elections (30050)
Event • Clinton campaigns for Democrats • Impeachment hearings begin • Effect of impeachment hearings on campaigns • Budget negotiations • Effect of budget on campaigns • Impact of other issues on campaigns
Topic: [P5] Clinton's Gaza Trip (30053)
Event • Clinton visits Middle East • Police fire on West Bank demonstrators • White House praises PLO revocation • Clinton comments on impeachment hearings • Clinton meets with Middle East leaders • Netanyahu will not hand over more land
Topic: [S1] 1998 Winter Olympics (20013)
Event • Preparation for Olympics • Olympic games open • Olympic contests, results
Topic: [S2] Super Bowl '98 (20033)
Event • Preparations, predictions for Super Bowl • Broncos win Super Bowl • Post-game celebrations, riots
Topic: [S3] NBA finals (20087)
Event • Basketball regional finals • Finals • Bulls win championship, Chicago celebrates
Topic: [S4] Yankees vs. Padres in World Series (31026)
Event • Padres win NLCS • Yankees win ALCS • Game 1 of World Series • Joe DiMaggio, Darryl Strawberry illnesses • Game 2 of World Series • Game 3 of World Series
Topic: [S5] Joe DiMaggio Illness (31036)
Event • DiMaggio in hospital for pneumonia • Debate, discussion over heart attack and lung cancer • Doctors confirm DiMaggio had lung cancer • Reflection on DiMaggio • DiMaggio develops infection, improves, then coma • DiMaggio improves • DiMaggio tells doctors to stop updating press

3.2 Named Entity Recognition

We used ANNIE (A Nearly-New Information Extraction System) that has been developed using GATE (Cunningham, 2002). It is an example of a lexical resource and rule-based approach to IE. It was used to identify regions of text corresponding to the seven MUC-7 named entity types (Person, Location, Organization, Date, Time, Money, and Percent).

The ANNIE system consists of seven processing resources organized into an application pipeline. These include a tokenizer, a gazetteer, a sentence splitter, a POS (parts of speech) tagger, and a named entity transducer. Each of them is associated with a language resource containing data or rules, i.e. tokenizer rules, gazetteer lookup lists, sentence segmentation rules, a POS lexicon, and NE transduction rules. The resources that are rule-based use JAPE (Java Annotations Pattern Engine) grammar rules to match patterns using regular expressions over annotations in order to create new annotations. JAPE rules can match against annotations, annotation features, token attributes, lookup types, and/or parts of speech and can take any java-based action in response to a matched pattern. Based on ANNIE's capability, therefore we were not building a NER system and instead using the existing system to recognise named entities in a document. We used it for its accurate entity, pronoun and nominal co-references extraction.

3.3 Procedure

We conducted a user experiment with 20 users from October, 2014 to January, 2015 at the School of Information Science, Japan Advanced Institute of Science and Technology (JAIST). The users were postgraduate students and the average age of the users was 20–30 years. Users were asked about their topic familiarity and topic interest before they started with the task. 1000 tasks were performed (20 users, 10 topics, 5 events) in this experiment. Five topics on Politic (P1-P5) and five topics on Sports (S1-S5) were conducted in two sessions in the Event Detection task (Section 3.4). After completion of the tasks, users were interviewed about their opinion or experience in performing the task. They were given 15 minutes to identify an event for each topic. The time assigned to each task was sufficient based on the feedback received from the pilot test conducted. The users were offered a short break (5–15 minutes) after the first session. A Latin square (Sparck-Jones, 1997) was used to construct the experimental design (refer Table 2). This allowed us to evaluate the same topic using

different amount of information. This was important to justify whether the type of information helped the users to detect an event even though they were given a low amount of information.

Table 2: Experimental design.

Users	Session 1					Session 2				
	Topic (Politic)					Topic (Sports)				
	P1	P2	P3	P4	P5	S1	S2	S3	S4	S5
1-10	L	L	L	L	L	H	H	H	H	H
11-20	H	H	H	H	H	L	L	L	L	L

L=Too little information/stories (Low)
H=Too much information/stories (High)

Low information means users received 40% or less stories while High information means they received 70% or more stories. 12 stories will be selected from the 30 stories using random sampling method to unsure equal probability of selection for each article in an event under the Low information category.

We also gave attention to the information quality. The ability to detect an event might be associated with the amount and quality of information given. User might claimed that they received poor information, hence we would like to avoid any issue due to the quality of information provided during the experiment. Thus these 2 aspects; amount and information quality are mutually independent. The stories in TDT corpus consist of quality information as it reflect the way it was annotated (TDT Annotation Manual). An integral and key part of the corpus is the annotation of the corpus in terms of the events discussed in the stories as shown in Table 3.

Table 3: Example of story for topic 'Current Conflict with Iraq'.

Topic	Event	Story
Current Conflict with Iraq	Iraq announces it will block inspections	Iraq says it will block one of the U.N. inspection teams being led by an American until the team is recomposed with fewer Americans. The team carried out its duties today without interference.
	Iraq prevents inspection team from entering	A new crisis may be developing in Baghdad. The Iraqi government blocked an American-led team of U.N. weapons inspectors from doing its work today. The White House says it's coordinating a response at the U.N. Security Council.

We defined successful event based on the keywords or the important terms in the event detected given by user. We classified event into three categories:

a. None: where users did not provide any event or they did not complete the task;
b. Successful: where users provide the right keyword or the important terms in the event detected
c. Unsuccessful: where users failed to provide the right keyword or the important terms in the event detected

We also conflated different keywords referring to the same context and meaning as shown in Table 4.

Table 4: Example of event detected by users and their category.

Predefined event	Event detected by users	Category
Gingrich announces he will resign	Gingrich resignation	Successful
	Newt Gingrich leave job	Successful
	Newt Gingrich dissatisfaction	Unsuccessful
	Gerald Ford resign	Unsuccessful

3.4 Event Detection Task

The event detection task was designed based on five uncertainty dimension. A user's session consisted of the following stages, carried out in a single block of time. In this task, the users had to detect an event for a given topic. The procedure for performing the task was as follows.

a. Users were welcomed and asked to read the introduction to the experiment provided on an information sheet. This set of instructions was developed to ensure that each user received precisely the same information. Users could retain the information sheet after the user experiment.
b. The users were given a short overview of what the experiment would entail. We also explained our role in this experiment – i.e. to provide users with support and remind users of the time taken in performing the task.
c. Users were asked to complete an entry questionnaire to provide us with background information.
d. Event Detection task
   - Users were asked to perform the task by identifying what was the event by following the experimental design (as shown in Table 2). Users were given 15 minutes to identify an event, and could stop early if they were unable to find any more relevant information.
   - Then there was one sub-activity in this task: profiling. Profiling required the user to provide the important keyword that was

considered as; ambiguous, complex and conflicting information (refer Table 5).

Table 5: Definition of ambiguous, complex and conflicting information.

Category	Definition
Ambiguous	Keyword which is not clear and have several possible meanings or interpretations in detecting an event.
Complex	Keyword which is difficult to understand in detecting an event.
Conflicting	Keyword which is contradict or different in detecting an event.

e. Users performed the post-evaluation interview. In this session, we asked user about their experiences in performing the task.

4 RESULTS

Findings revealed that there was no statistical significance difference between topics and topic familiarity (Mann-Whitney Test, p=0.496). The users were not familiar with the topics given in the Event Detection task (mean=2.04 sd=1.05). There were also no statistically significant difference between the users and their topic interest (Mann-Whitney Test, p=0.844). Their topic interest was average (mean=3.29 sd=1.11). This is a good indication of the experiment since the users are not affected by external factors such as their topic familiarity and topic interest.

4.1 Named Entity Distribution Across Topics

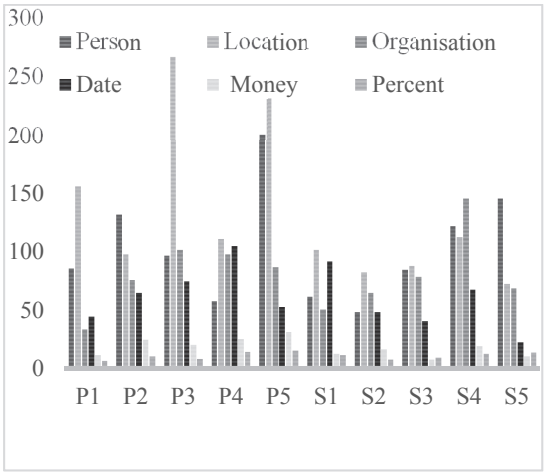

Figure 1: Named entities across topics.

Figure 1 summarizes the distribution of NE across topics. Topic P5 (Clinton's Gaza Trip) has the highest percentage of Person (19.4%), Money (14.3%) and Percent (13.0%) NEs. Meanwhile topic S4 (Yankees vs. Padres in World Series) has the highest Organisation (18.1%) and Date (17%) NEs. Topic P3 (Gingrich Resigns) has the highest Location NEs (20.2%). The distribution of NEs are affected by the topics and events occurred. One possibility is the nature of the topic that has caused certain NEs to appear frequently.

4.2 User Evaluation

In this section we discussed the evaluation on the amount and type of information by analysing the rate of successful event detected (discussed in Section 3.3) in conjunction with the amount of information/stories and named entities distribution across topics.

4.2.1 Amount of Information

The entire event detection task was successful, with 93.9% of the task being successful and 6.1% being unsuccessful.

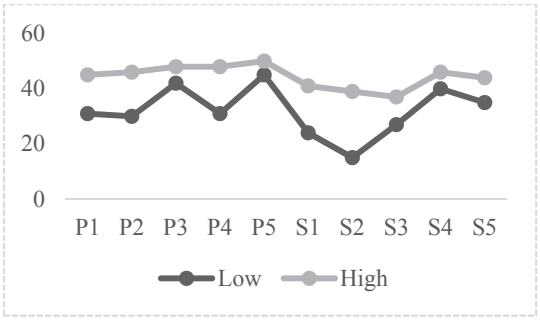

Figure 2: Successful event detection rate for high and low information/stories across topics.

We associated the successful rate in identifying an event with the low uncertainty rate. Users were able to successfully identify an event if they were certain with the stories or information received. Figure 2 shows the number of successful event detected for high and low stories across topics. There was no statistical significance difference in the number of successful event detected (Mann-Whitney Test, $p>0.05$). However there was a statistical significance difference in the number of successful event detected (Mann-Whitney Test, $p<0.05$) in conjunction with the low number of stories across topics. Users were able to successfully detect an event even they were provided with low number of stories. This occurred

in topics P3 (Gingrich Resigns), P5 (Clinton's Gaza Trip) and S4 (Yankees vs. Padres in World Series).

One of the possibility was the high distribution of named entities for these topics as shown in Table 6. Topic P3 has the highest Location NEs (20.2%). Topic P5 has the highest Person (19.4%), Money (14.3%) and Percent (13.0%) NEs. Meanwhile topic S4 has the highest Organisation (18.1%) and Date (17%) NEs.

Table 6: Named entities distribution.

Topic	Person	Location	Organisation	Date	Money	Percent
P1	8.3	11.8	4.2	7.3	6.5	6.1
P2	12.7	7.4	9.4	10.6	13.5	9.6
P3	9.3	20.2	12.6	12.2	11.4	8.8
P4	5.6	8.4	12.1	11.0	14.1	12.9
P5	19.4	17.5	10.8	8.6	14.3	13.0
S1	6.0	7.7	6.3	14.9	8.0	10.4
S2	4.7	6.3	8.1	8.0	11.2	7.0
S3	8.2	6.7	9.8	6.7	4.3	8.7
S4	11.8	8.5	18.1	17.0	10.8	11.3
S5	14.1	5.5	8.6	3.7	5.9	12.2

Users managed to detect an event when they were provided with high number of stories compared to when they were provided with low number of stories. However there was an exception if the low number of stories have high number of NEs. This indicated that named entities could reduce the uncertainty in event detection although users were provided with low number of stories.

4.2.2 Type of Information

Users issued an average of 22 keywords of NEs per topic and an average of 4 keywords of NEs per event. Meanwhile they also provided double the amount for BOW; an average of 47 keywords of BOW per topic and an average of 9 keywords of BOW per event. The number of keywords that were labelled as ambiguous, complex and conflict in event detection. It is important to analyse the distribution of these keywords to identify what type of information (NEs, BOW) and in which condition (low, high) will significantly contribute to uncertainty as shown in Table 7.

There was no significant difference in type of information across topics for complex and conflict dimension (Mann-Whitney Test, $p>0.05$). Users tends to provide almost the same average amount of keywords between NEs (mean=14) and BOW (mean=16) per topic.

However there was a statistical significance difference for type of information (Mann-Whitney Test, $p<0.05$) in conjunction with the low information/stories across topics for ambiguous dimension. User list out less NEs (mean=3.01, sd=4.05) as an ambiguous information compared to BOW (mean=9.42, sd=6.45) when they were provided with low information/stories in topics P3 (Gingrich Resigns), P5 (Clinton's Gaza Trip) and S4 (Yankees vs. Padres in World Series). One of the reason probably the high number of NEs occurred in these topics has help user in event detection task.

Table 7: Type of information distribution across different settings.

		Frequency (%)		
Topic	Amount of info./ stories	Ambiguous	Complex	Conflict
P1	Low	34.7 (65.3)	54.3 (45.7)	44.1 (55.9)
	High	45.7 (54.3)	50.2 (49.8)	48.2 (51.8)
P2	Low	28.1 (71.9)	44.7 (55.3)	54.4 (45.6)
	High	42.4 (57.6)	54.1 (45.9)	46.7 (53.3)
P3	Low	15.6 (84.4)	47.9 (52.1)	55.8 (44.2)
	High	51.2 (48.8)	50.5 (49.5)	55.2 (44.8)
P4	Low	23.7 (76.3)	47.8 (52.2)	44.9 (55.1)
	High	47.1 (52.9)	45.5 (54.5)	54.7 (45.3)
P5	Low	12.6 (87.4)	54.5 (45.5)	47.7 (52.3)
	High	40.5 (59.5)	47.4 (52.3)	50.7 (49.3)
S1	Low	29.6 (70.4)	54.8 (45.2)	47.5 (52.5)
	High	50.3 (49.7)	47.8 (52.2)	45.9 (54.1)
S2	Low	31.7 (68.3)	44.2 (55.8)	47.2 (52.8)
	High	53.3 (46.7)	55.3 (44.7)	44.4 (55.6)
S3	Low	25.5 (74.5)	49.9 (50.1)	54.2 (45.8)
	High	48.9 (51.1)	47.1 (52.9)	45.8 (54.2)
S4	Low	19.8 (80.2)	55.1 (44.9)	47.6 (52.4)
	High	52.1 (47.9)	45.6 (54.4)	50.4 (49.6)
S5	Low	33.5 (66.5)	54.6 (45.4)	54.8 (45.2)
	High	46.2 (53.8)	50.0 (50.0)	47.2 (52.8)

* Figure in bracket referring to the frequency of BOW (%)

This indicates that user perceived BOW as the ambiguous information when they were given with low number of stories to detect an event.

4.2.3 Post Evaluation Interview

During the post-evaluation interview, 85% of the users agreed that the BOW were more descriptive thus making the event detection task difficult. Meanwhile 92% of the users agreed that named entities has produced interesting and specific information which has helped them to become focus in identifying an event even they were provided with low number of stories.

98% of the users agreed that ambiguous, complex and conflicting information has nothing to do with their understanding of the meaning of a term. 95% of the user claimed that concentrating on the named entities appeared in a stories has help them to successfully perform the task.

5 CONCLUSIONS

User are able to detect an event even when they were provided with low information or stories. Low percentage of NEs was labelled as 'ambiguous' by user during the event detection task. Thus NEs reduce ambiguity and uncertainty in event detection, compared to *bag of words* which is more descriptive. This is one of the justification that NEs increased the user confidence in understanding the flow of stories by providing user with high quality forms of information. Associating NEs occurred in an event could be one of user strategy to increase their understanding of a topic. Therefore another future direction lies is by analysing named entity recognition and linking to reduce uncertainty.

REFERENCES

TDT Corpus. https://catalog.ldc.upenn.edu/LDC98T25.

TDT Annotation Manual. https://catalog.ldc.upenn.edu/docs/LDC2006T19/TDT2004V1.2.pdf.

Mohd, M and Mabrook, O., 2014. Investigating the Combination of Bag of Words and Named Entities Approach in Tracking and Detection Tasks among Journalists. *Journal of Information Science Theory and Practice*. 2(4), pp 31-38.

Hurley, R. J. 2011. Uncertain about cancer? so is online news. *Communication Currents*, 6 (5). http://www.natcom.org/CommCurrentsArticle.aspx?id=1703.

Chowdhury, S., Gibb, F., & Landoni, M. 2011. Uncertainty in information seeking and retrieval: A study in an academic environment. *Inf. Process. Manage*. 47, 2 (March 2011), pp. 157-175. DOI=http://dx.doi.org/10.1016/j.ipm.2010.09.006.

Ingwersen, P. 1992. Information Retrieval Interaction, Taylor Graham, London.

Goodman, N. D., Vikash, K., Mansinghka, D. R., Bonawitz, K., and Tenenbaum, J.B. 2008. Church: A language for generative models. In *Proceedings of the Twenty-Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence*, Corvallis, Oregon, pp. 220-229.

Topka, L. V., 2013. Situation of uncertainty: pragmatic, semantic, and syntactic aspects of investigation. *European Scientific Journal*. 9 (Sept. 2013), pp. 60-69.

Rubin, V. L., Liddy, E. D., and Kando, N. 2006. Certainty identification in texts: categorization model and manual tagging results. *Springer, Dordrecht*, The Netherlands, vol. 20, pp. 61-76.

Goujon, B. 2009. Uncertainty detection for information extraction. In *Proceedings of the International Conference RANLP 2009*, Borovets, Bulgaria, pp. 118-122.

Mishel, M. H. 1988. Uncertainty in illness. *Image J Nurs Sch*, vol. 20, pp. 225–232.

Babrow, A. S., Kasch, C. R., and Ford, L. A. 1998. The many meanings of uncertainty in illness: towards a systematic accounting. Heatlh Communication, vol.10, pp. 1-23.

Chen H. and Ku L. W. 2002. An NLP and IR approach to topic detection. In: Allan J (ed.) Topic detection and tracking: Event-based information organization. Norwell, MA: *Kluwer Academic Publishers*, pp. 243–264.

Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. 2002. GATE: A framework and graphical development environment for robust NLP tools and applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, July 2002, Philadelphia, PA, pp. 168–175.

Sparck-Jones, K., and Willet, P. 1997. *Readings in Information Retrieval*. San Francisco, CA: Morgan Kaufmann.