The Role of Roles in Classifying Annotated Biomedical Text

Son Doan, Ai Kawazoe, Nigel Collier
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan
{doan,zoeai,collier}@nii.ac.jp

Abstract

This paper investigates the roles of named entities (NE's) in annotated biomedical text classification. In the annotation schema of BioCaster, a text mining system for public health protection, important concepts that reflect information about infectious diseases were conceptually analyzed with a formal ontological methodology. Concepts were classified as Types, while others were identified as being Roles. Types are specified as NE classes and Roles are integrated into NEs as attributes. We focus on the Roles of NEs by extracting and using them in different ways as features in the classifier. Experimental results show that: 1) Roles for each NE greatly helped improve performance of the system, 2) combining information about NE classes with their Roles contribute significantly to the improvement of performance. We discuss in detail the effect of each Role on the accuracy of text classification.

1 Introduction

Today, the Internet is a powerful tool for discovering novel information via news feed providers. This is becoming increasingly important for the public health domain because it can help to detect emerging and re-emerging diseases. In infectious disease surveillance systems such as the Global Public Health Intelligence Network (GPHIN) system (Public Health Agency of Canada, 2004) and ProMed-Mail (International Society for Infectious Diseases, 2001), the detection and tracking of outbreaks using the Internet has been proven to be a key source of information for public health workers, clinicians, and researchers interested in communicable diseases. The basis for such systems is the monitoring of a large number of news articles simultaneously. The classification of news articles into disease-related or none disease-related classes is the first stage in any automated approach to this task. In practice though there are a large number of news articles whose main subject is related to diseases but which should not necessarily be notified to users together with a relatively small number of high priority articles that experts should be actively alerted to. Alerting criteria broadly include news related to newly emerging diseases, the spread of diseases across international borders, the deliberate release of a human or engineered pathogen, etc. The use of only raw text in the classification process inevitably fails to resolve many subtle ambiguities, for example semantic class ambiguities in polysemous words like "virus", "fever", "outbreak", and "control" which all exhibit a variety of senses depending on context. These different senses appear with relatively high frequency in the news especially in headlines. A further challenge is that diseases can be denoted by many variant forms. Therefore we consider that the use of advanced natural language processing (NLP) techniques like named entity recognition (NER) and anaphora resolution are needed in order to achieve high classification accuracy.

Text classification is defined as the task of assigning documents into one or more predefined cat-

BioNLP 2007: Biological, translational, and clinical language processing, pages 17–24,
Prague, June 2007.

egories. As shown by (Cohen and Hersh, 2005), an accurate text classification system can be especially valuable to database curators. A document in the biomedical domain can be annotated using NER techniques with enriched semantic information in the form of NEs such as the disease, pathogen, location, and time. NER and term identification in general have been recognized as an important research topic both in the NLP and biomedical communities (Krauthammer and Nenadic, 2004). However, an investigation into the contribution of NEs on the performance of annotated biomedical text classification has remained an open question until now. There are two main reasons for this: Firstly there are a small number of open annotation schema for biomedical text, and secondly there is no benchmark annotated data for testing.

The BioCaster project (Collier, 2006) is working towards the detection and tracking of disease outbreaks from Internet news articles. Although there are several schema for biomedical text (Wilbur et al., 2006), little work has been done on developing one specifically for public health related text. BioCaster therefore provides an annotation schema that can fill this gap. Our schema, which is based on discussions with biologists, computational linguists and public health experts, helps identify entities related to infectious diseases which are then used to build up a detailed picture of events in later stages of text mining. One significant aspect of the schema is that it is based on conceptual analysis with a formal ontological methodology. As discussed in (Kawazoe et al., 2006), by applying meta-properties (Guarino and Welty, 2000a; Guarino and Welty, 2000b), our "markable" concepts are classified into "Type" and "Role". Information about Role concepts is integrated into the schema as attributes on NEs. This work takes the investigation one step forward by showing empirical evidence for the usefulness of Role concepts in a practical application.

In this paper, we focus on the task of text classification, proceeding under the simplifying assumption that given enough annotated training data for NEs and their Roles both can be automatically tagged with high accuracy. In recent years there have been many studies on text classification using general methods (Sebastiani, 2002; Yang and Liu, 1999) semi-structured texts (Kudo and Matsumoto, 2004), and XML classification (Zaki and Aggarwal, 2003). Other research has investigated the contribution of semantic information in the form of synonyms, syntax, etc. in text representation (Bloehdorn and Hotho, 2004; Hotho et al., 2003; Frürnkranz et al., 1998). Feature selection (Scott and Matwin, 1999) has also been studied. The contribution of this paper is to provide an analysis and evaluation on the Roles of NEs in annotated text classification.

The rest of this paper is organized as follows: in Section 2, we outline the BioCaster schema for the annotation of terms in biomedical text; Section 3 presents a description of the BioCaster gold standard corpus; Section 4 provides details of the method and experimental results of classification on the gold standard corpus. Finally we draw some conclusions in Section 5.

2 BioCaster Schema for Annotation of Terms in Biomedical Text

The BioCaster annotation schema is a component of the BioCaster text mining project. We have identified several important concepts that reflect information about infectious diseases, and created guidelines for annotating them as target entity classes in texts. Based on the conceptual analysis using meta-properties (rigidity, identity, and dependency) developed by Guarino and Welty (2000a; 2000b), categories of important concepts were classified as Types, i.e., properties which are rigid[1] and supply identity conditions, while others were identified as being Roles, properties which are anti-rigid[2] and dependent. The 18 categories of Type concepts are specified as NE classes which we denote here in upper case. These include PERSON, LOCATION, ORGANIZATION, TIME, DISEASE, CONDITION (status of patient such as "hospitalized" or "in stable condition"), OUTBREAK (event of group infection), VIRUS, ANATOMY (body part), PRODUCT (biological product such as "vaccine"), NONHUMAN (animals), DNA, RNA, PROTEIN, CONTROL (control measures to contain the disease), BACTERIA, CHEMICAL and SYMPTOM. The three Role concepts we explore are *case* (dis-

[1]A property is *rigid* if every instance of that property necessarily has the property, i.e. in every possible world.

[2]A property is *anti-rigid* if no instance of that property necessarily has the property.

eased person), *transmission* (source of infection) and *therapeutic* (therapeutic agent). These are integrated into the annotation schema as XML attributes which are associated with some XML elements denoting Type concepts. PERSON takes a *case* attribute, NONHUMAN and ANATOMY take *transmission*, PRODUCT takes *transmission* and *therapeutic* and CHEMICAL takes *therapeutic*. For PERSON we added another attribute *number* (number of people). Each attribute has only one value, the value of *number* is *one* or *many*, and the value of *case*, *transmission*, *therapeutic* is *true* or *false*. This is summarized in Table 1. In the rest of this paper, we call *case*, *transmission*, and *therapeutic* "Role attributes" (or "Role" for short) and *number* a "Quality attributes" (or "Quality" for short).

A NE in a biomedical text is annotated following the BioCaster annotation schema in XML format as follows,

```
<NAME cl="Named Entity"
attribute1="value1" attribute2="value2"
...  </NAME>,
```

where `"Named Entity"` is one of the names for the 18 BioCaster NEs and `attribute1, attribute2, ...` are the names of the NE's Role/Quality attributes, `"value1", "value2", ...` are values corresponding to Role/Quality attributes. Further details of the annotation guidelines are discussed in (Kawazoe et al., 2006).

3 BioCaster Gold Standard Data Corpus

The BioCaster gold standard corpus was collected from Internet news and manually annotated by two doctoral students. The annotation of a news article proceeded as follows. Firstly, NEs are annotated following the BioCaster schema and guidelines. Secondly, each annotated article is manually assigned into one of four relevancy categories: *alert*, *publish*, *check*, and *reject*. The assignment is based on guidelines that we made following discussions with epidemiologists and a survey of World Health Organization (WHO) reports (World Health Organization, 2004). These categories are currently being used operationally by the GPHIN system which is used by the WHO and other public health agencies. Where there were major differences of opinion in NE annotation or relevancy assignment between the two annotators, we consulted a public health expert in order to decide the most appropriate assignment. Finally we had a total of 500 articles that were fully annotated. While this is small compared to other data sets in text classification, we consider that it is large enough to obtain a preliminary indication about the usefulness of Role attributes.

The following is an example of an annotated article in the BioCaster gold standard corpus.

Example.

```
<DOC id="000125" language="en-us"
source="WHO" domain="health"
subdomain="disease"
date_published="2005-03-17"
relevancy="alert"> <NAME cl="DISEASE">
Acute fever </NAME> and <NAME
cl="DISEASE"> rash syndrome </NAME> in
<NAME cl="LOCATION">Nigeria</NAME> <NAME
cl="TIME"> 17 March 2005 </NAME><NAME
cl="ORGANIZATION"> WHO</NAME> has received
reports of <NAME cl="PERSON" case="true"
number="many"> 1118 cases </NAME>
including <NAME cl="PERSON" case="true"
number="many">76 deaths</NAME>case
fatality rate, 6.8% reported in 12
Local Government Areas (LGAs) of <NAME
cl="LOCATION">damawa </NAME> state, <NAME
cl="LOCATION"> Nigeria</NAME> as of <NAME
cl="TIME">28 February 2005</NAME>.  The
cases have been clinically diagnosed
as <NAME cl="DISEASE"> measles </NAME>
but no laboratory diagnosis has been
made to date.  Other states, including
<NAME cl="LOCATION">Gombe</NAME>,
<NAME cl="LOCATION">Jigawa</NAME>,<NAME
cl="LOCATION">Kaduna</NAME>, <NAME
cl="LOCATION">Kano</NAME>, and <NAME
cl="LOCATION">Kebbi</NAME> have all
reported <NAME cl="OUTBREAK"> outbreaks
</NAME> of <NAME cl="DISEASE"> measles
</NAME>...  </DOC>
```

We grouped the 500 articles into 2 categories: *reject* and *relevant*. The *reject* category corresponds simply to articles with label *reject* while the *relevant* category includes articles with labels *alert*, *publish*, and *check*. We conflated the *alert*, *publish* and *check* categories because we hypothesized that distinguishing between non-reject (relevant) categories

Named entity	Role/Quality attributes	Named entity	Role/Quality attributes
PERSON	case, number	ANATOMY	transmission
ORGANIZATION	none	SYMPTOM	none
LOCATION	none	CONTROL	none
TIME	none	CHEMICAL	therapeutic
DISEASE	none	BACTERIA	none
CONDITION	none	PRODUCT	transmission, therapeutic
NONHUMAN	transmission	DNA	none
VIRUS	none	RNA	none
OUTBREAK	none	PROTEIN	none

Table 1: Lists of Named entity classes and their Role/Quality attributes in BioCaster annotation schema.

would require higher level semantic knowledge such as pathogen infectivity and previous occurrence history which is the job of the text mining system and the end user. Finally we had a total of 269 news articles belong to the *reject* category and 231 news articles belong to the *relevant* category. The statistical information about NEs is shown in Table 2. In the table, "+" stands for the frequency of NEs in the *relevant* category and "-" stands for the frequency of NEs in the *reject* category.

4 Experiments

4.1 Method

We used the BioCaster gold standard corpus to investigate the effect of NE classes and their Role attributes on performance of classification. In order to avoid unnecessary data, we removed the first line containing DOC tag of all article in the corpus. The validation is as follows. We randomly divided the data set into 10 parts. Each of the first 9 parts has 23 articles belonging to the *relevant* category and 27 articles belonging to the *reject* category; the 10th part has 24 articles belonging to the *relevant* and 26 articles belonging to the *reject* categories. Then, we implemented 10-fold cross validation: 9 parts for training and 1 part for testing sets. For the training set we extracted NEs classes and their Roles as features to build a classifier. The remaining part was used for testing.

The classifier we use in this paper is the standard Näive Bayes classifier (Mitchell, 1997). In the preprocessing we did not use a stop list and no word stemming. The experiments were implemented in Linux OS, using the Bow toolkit (McCallum, 1996).

The details of extracting NEs and their Roles from annotated texts are the followings. For the sake of convenience, we divided features into 3 groups: Features for each NE, features for NEs with Role/Quality, and features for combined NEs with Role/Quality.

1. Features for each NE: Each NE is extracted and used with raw text as features. We denoted NE1 as features extracted from named entity NE1. For example, DISEASE1 means features are raw text and DISEASE class, VIRUS1 means features are raw text and VIRUS class. An example of features for PERSON1 is shown in Table 3.

2. Features for NEs with Role/Quality: We investigated the effect of NEs with Roles/Qualities, i.e., *case*, *number*, *therapeutic*, and *transmission*. Features are chosen as follows.

   - PERSON+case+number: Raw text and PERSON class with both Role *case* and Quality *number* are used as features.
   - PERSON+case: Raw text and PERSON class with Role *case* are used as features.
   - PERSON+number: Raw text and PERSON class and Quality *number* are used as features.
   - NONHUMAN+trans: Raw text and NONHUMAN class and Role *transmission* are used as features.
   - ANATOMY+trans: Raw text and ANATOMY class and Role *transmission* are used as features.

NE class	Frequency	Total	NE class	Frequency	Total
PERSON	+3291/-4978	8269	ANATOMY	+263/-224	487
ORGANIZATION	+1405/-3460	4865	SYMPTOM	+293/-105	398
LOCATION	+2432/-2409	4841	CONTROL	+282/-87	369
TIME	+1159/-1518	2677	CHEMICAL	+108/-185	293
DISEASE	+1164/-456	1620	BACTERIA	+136/-103	239
CONDITION	+689/-206	895	PRODUCT	+124/-74	198
NONHUMAN	+393/-344	737	DNA	+8/-55	63
VIRUS	+428/-127	555	RNA	+0/-55	55
OUTBREAK	+460/-75	535	PROTEIN	+5/-32	37

Table 2: The frequency of NE classes in the BioCaster gold standard corpus, "+" denotes the frequency in the *relevant* category and "-" denotes the frequency in the *reject* category.

Example of annotated text	`<NAME cl="ORGANIZATION"> WHO</NAME> has received reports of <NAME cl="PERSON" case="true" number="many"> 1118 cases </NAME>`
Text only	"WHO", "has", "received", "reports", "of", "1118", "cases"
PERSON1	"WHO", "has", "received", "reports", "of", "1118", "cases", "PERSON"
PERSON+case+number	"WHO", "has", "received", "reports", "of", "1118", "cases", "PERSON", "case", "number"
PERSON+case	"WHO", "has", "received", "reports", "of", "1118", "cases", "PERSON", "case"
PERSON+number	"WHO", "has", "received", "reports", "of", "1118", "cases", "PERSON", "number"

Table 3: An example of using different features for PERSON class as training data.

- PRODUCT+trans+thera: Raw text and PRODUCT class and both Roles *transmission* and *therapeutic* are used as features.
- PRODUCT+trans: Raw text and PRODUCT class and Role *transmission* are used as features.
- PRODUCT+thera: Raw text and PRODUCT class and Role *therapeutic* are used as features.
- CHEMICAL+thera: Raw text and CHEMICAL class and Role *therapeutic* are used as features.

3. Features for combined NEs with Roles. We investigate features for disease-related NEs which include DISEASE, VIRUS, BACTERIA, SYMPTOM, CONDITION, CONTROL, DNA, PROTEIN, RNA, OUTBREAK, PRODUCT, ANATOMY, NONHUMAN, CHEMICAL and features for all NEs with their Roles, i.e., *therapeutic* and *transmission*. We investigated 5 different features as follows:

    - Text only: Only raw text is used as features.
    - Text+DiseaseNEs: Raw text and all 14 NEs disease-related classes are used as features.
    - Text+DiseaseNEs+Roles: Raw text and all 14 NEs disease-related classes with Roles are used as features. We note that there are two Roles *therapeutic* and *transmission* in this case.
    - Text+AllNEs: Raw text and all NE classes are used as features.
    - Text+AllNEs+Roles: Raw text and all NE classes with Roles are used as features. In this case we have all 3 Roles *case*, *therapeutic* and *transmission*.

An example of using different features for PER-

	YES is correct	*NO* is correct
Assigned *YES*	a	b
Assigned *NO*	c	d

Table 4: A contingency table.

SON class is shown in Table 3.

4.2 Results and Discussions

The details of experimental results are shown in the following sections. We use two performance measures, standard Precision/Recall and accuracy. They are calculated based on the two-way contingency table in Table 4. In the table, a counts the assigned and correct cases, b counts the assigned and incorrect cases, c counts the not assigned but incorrect cases, and d counts the not assigned and correct cases (Yang, 1999). Then,

$$\text{Precision} = \frac{a}{a+b}, \text{ and Recall} = \frac{a}{a+c}.$$

Accuracy is defined as accuracy=$(a + d)/(a + b + c + d)$.

4.2.1 Effectiveness of Each NE Class

In order to investigate the effect of NEs on performance, we consider the baseline as the method using text only. In experiment the baseline achieved a performance of 74.40% accuracy and 64.35% Precision, 100% Recall. We can see that Recall always achieves 100% in all cases. This may be due to the small size of data. However it is interesting that we can observe the change of Precision measure - an important measure in our case. Hereafter we discuss accuracy and Precision only.

The effectiveness of each NE class is shown in Table 5. The results show that each NE does not have the same effect. Compared to the baseline, nearly half the total NEs (7/18) help improve performance while the others do not have a significant affect.

Looking at the distribution of NE frequency in Table 2, it seems that the higher the frequency of the NE class, the better the performance it provides. For example, PERSON achieved the best of all (76.80% accuracy, 66.57% Precision compared to 74.40% accuracy and 64.35% Precision when using raw text). However this trend is not always followed, for example, the TIME class tends to reduce performance when compared to raw text. This is natural as there is no obvious correlation between time and relevancy. From the result tables we can conclude that the effectiveness of each NE on the performance of classification in our corpus is decreased in the following order.

PERSON > LOCATION > ORGANIZATION > DISEASE > CONDITION = VIRUS = OUTBREAK > NONHUMAN = ANATOMY = SYMPTOM = CONTROL = BACTERIA = PRODUCT = PROTEIN > CHEMICAL = DNA = RNA > TIME

In particular, 7 NEs, i.e., PERSON, LOCATION, ORGANIZATION, DISEASE, CONDITION, VIRUS, OUTBREAK improve performance, while TIME significantly reduces it. Two NEs DNA and RNA that have low frequency weakly reduce performance.

4.2.2 Effectiveness of Roles on Classification

In this Section we investigate the effect of each Role on performance. The experimental results are shown in Table 6. We can easily observe that Roles in NEs improved both the accuracy and Precision significantly.

We first consider the Role *case*. This Role is associated to PERSON which has highest frequency in the corpus. Role *case* helped improve the accuracy from 76.8% to 80.60%, and Precision from 66.57% to 74.43% for PERSON. This is significant when we compare to the baseline with 74.4% accuracy and 64.35% Precision. We note that PERSON has another attribute, the Quality *number*. Role *case* helps PERSON with Quality number improve the accuracy from 78.00% to 81.80% and Precision from 67.74% to 71.74%. Moreover, we can obviously draw the relative comparison about effectiveness between Role *case* and Quality *number* from these results, it yields that *case* > *number*.

We proceed to investigate the effect of Roles *therapeutic* and *transmission*. Obviously we see that their effects on performance are positive. Specifically, *transmission* help NONHUMAN improve the accuracy from 74.40% to 74.60%, *therapeutic* helps CHEMICAL improve the accuracy from 74.20% to 74.40%. They both have not effects on some minor NE classes like ANATOMY and PRODUCT. If we had more training data with more of these minor NE classes we hope to see a positive effect from

Named entity	Accuracy	Pre/Rec	Named entity	Accuracy	Pre/Rec
PERSON1	76.80	66.57/100	**ANATOMY1**	74.40	64.35/100
ORGANIZATION1	75.40	65.25/100	SYMPTOM1	74.40	64.35/100
LOCATION1	75.60	65.44/100	CONTROL1	74.40	64.35/100
TIME1	73.00	63.11/100	**CHEMICAL1**	74.20	64.17/100
DISEASE1	75.00	64.89/100	BACTERIA1	74.40	64.35/100
CONDITION1	74.60	64.53/100	**PRODUCT1**	74.40	64.35/100
NONHUMAN1	74.40	64.35/100	DNA1	74.20	64.17/100
VIRUS1	74.60	64.53/100	RNA1	74.20	64.17/100
OUTBREAK1	74.60	64.53/100	PROTEIN1	74.40	64.35/100

Table 5: Performance of each NE class in which features of NEs in bold text have Role attributes.

FEATURES	Accuracy	Pre/Rec
Baseline	74.40	64.35/100
PERSON1	76.80	66.57/100
PERSON+number	78.00	67.74/100
PERSON+case	**80.60**	**74.43/100**
PERSON+case+number	**81.80**	**71.74/100**
NONHUMAN1	74.40	64.35/100
NONHUMAN+trans	**74.60**	**64.53/100**
ANATOMY1	74.40	64.35/100
ANATOMY+trans	**74.40**	**64.35/100**
PRODUCT1	74.40	64.35/100
PRODUCT+trans	**74.40**	**64.35/100**
PRODUCT+therapeutic	**74.40**	**64.35/100**
PRODUCT+trans+thera	**74.40**	**64.35/100**
CHEMICAL1	74.20	64.17/100
CHEMICAL+therapeutic	**74.40**	**64.35/100**

Table 6: Performance of Role attributes with their NEs.

Roles on them. Interestingly, while NEs associated to Roles do not improve the accuracy like NONHUMAN and CHEMICAL, their Roles helped improve the accuracy. Based on the improvements of *transmission* and *therapeutic* in Table 6, we can draw their effectiveness are the same on their NEs, that is *therapeutic* = *transmission*.

When we compare the effect of all Roles on performance, we can see that the improvements of Role *case* and also Quality *number* are much higher than the improvements of Roles *therapeutic* and *transmission*. We think this is because the frequency of PERSON (NE associated to Role *case* and Quality *number*) is higher than the frequency of NEs which are associated to Roles *therapeutic* and *transmission* in the corpus. Then, we can have the effect of Roles/Qualities is in the order *case* > *number* > *therapeutic* = *transmission*.

FEATURES	Accuracy	Pre/Rec
Baseline	74.40	64.35/100
Text+DiseaseNEs	75.80	65.63/100
Text+DiseaseNEs+Roles	**76.20**	**66.00/100**
Text+AllNEs	79.40	69.16/100
Text+AllNEs+Roles	**84.40**	**74.76/100**

Table 7: The performance of combined NEs with their Roles.

4.2.3 Effectiveness of Combined NEs with Roles

We continue to investigate the effectiveness of Roles for combined NEs. The experimental results are given in Table 7. We note that there are two Roles *therapeutic* and *transmission* in disease-related NE classes, and all 3 Roles *case*, *therapeutic* and *transmission* in all NE classes.

We can easily see that Roles improved performance of text classification significantly. In details, for disease-related NE classes, Roles *therapeutic* and *transmission* helped to improve the accuracy from 74.40% to 76.20%, and Precision from 64.35% to 66.% compared to the baseline. For all NE classes, all 3 Roles *case*, *therapeutic*, and *transmission* help to improve the accuracy from 74.40% to 84.40% and Precision from 64.35% to 74.76%. We conclude that all 3 Roles achieved the best results in performance.

5 Conclusion

This paper has focused on the contribution of Roles in biomedical annotated text classification. The experimental results indicated that:

1. Roles of each NE greatly help improve performance of the system.
2. The effect of Role/Quality attributes on classification was decreased in the order as follows: *case* $>$ *number* $>$ *therapeutic* $=$ *transmission*.
3. Combined NE classes with Roles contribute significantly to the improvement of performance.

Acknowledgments

The authors wish to thank Mika Shigematsu and Kiyosu Taniguchi at the National Institute of Infectious Diseases for useful discussions. This work was supported by Grants-in-Aid from the Japan Society for the Promotion of Science (grant no. 18049071).

References

S. Bloehdorn and A. Hotho. 2004. Boosting for text classification with semantic features. In *Proc. of the Workshop on Mining for and from the Semantic Web at the 10th ACM SIGKDD 2004*, pages 70–87.

A.M. Cohen and W.R. Hersh. 2005. A survey of current work in biomedical text mining. *Briefing in BioInformatics*, 6(3):57–71.

N. Collier. 2006. BioCaster text mining project. `http://biocaster.nii.ac.jp`.

J. Frürnkranz, T. Mitchell, and E. Riloff. 1998. A case study in using linguistic phrases for text categorization on the WWW. In *Working Notes of the AAAI/ICML Workshop on Learning for Text Categorization*, pages 5–13.

N. Guarino and C. Welty. 2000a. A formal ontology of properties. In *Proceedings of the 2000 Conference on Knowledge Engineering and Knowledge Management (EKAW-2000)*, pages 97–112.

N. Guarino and C. Welty. 2000b. Ontological analysis of taxonomic relations. In *Proceedings of the International Conference on Conceptual Modeling*, pages 210–224.

A. Hotho, S. Staab, and G. Stumme. 2003. WordNet improves text document clustering. In *Proc. of the SIGIR 2003 Semantic Web Workshop, 2003*.

International Society for Infectious Diseases. 2001. Promed mail. `http://www.promedmail.org`.

A. Kawazoe, L. Jin, M. Shigematsu, R. Barrero, K. Taniguchi, and N. Collier. 2006. The development of a schema for the annotation of terms in the BioCaster disease detection/tracking system. In *Proceedings of the International Workshop on Biomedical Ontology in Action (KR-MED 2006)*, pages 77–85.

M. Krauthammer and G. Nenadic. 2004. Term identification in the biomedical literature. *Journal of Biomedical Informatics*, 37(6):512–526.

T. Kudo and Y. Matsumoto. 2004. A boosting algorithm for classification of semi-structured text. In *Proceedings of the 2004 Conference on Empirical Methods in NLP*, pages 301–308.

A.K. McCallum. 1996. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. `http://www.cs.cmu.edu/~mccallum/bow`.

T.M. Mitchell. 1997. *Machine Learning*. McGraw-Hill.

Public Health Agency of Canada. 2004. Global Public Heath Intelligence Network (GPHIN). `http://www.gphin.org`.

S. Scott and S. Matwin. 1999. Feature engineering for text classification. In *Proc. of International Conference on Machine Learning 1999*, pages 379–388.

F. Sebastiani. 2002. Machine learning in automated text categorization. *ACM computing survey*, 34(1):1–47.

W. J. Wilbur, A. Rzhetsky, and H. Shatkay. 2006. New directions in biomedical text annotation: definition, guidelines and corpus construction. *BMC Bioinformatics*, 7(356):1471–2105.

World Health Organization. 2004. ICD10, International Statistical Classification of Diseases and Related Health Problems, Tenth Revision.

Y. Yang and X. Liu. 1999. A re-examination of text categorization methods. In *Proc. of 22th ACM Int'l. Conf. on Research and Development in Information Retrieval*, pages 42–49.

Y. Yang. 1999. An evaluation of statistical approaches to text categorization. *Information Retrieval Journal*, 1:69–90.

M.J. Zaki and C.C. Aggarwal. 2003. XRules: an effective structural classifier for XML data. In *Proceedings of the ninth ACM SIGKDD International Conference, 2003*, pages 316–325.