

Fast and simple semantic class assignment for biomedical text

K. Bretonnel Cohen

Computational Bioscience Program
U. Colorado School of Medicine
and

Department of Linguistics

U. of Colorado at Boulder

kevin.cohen@gmail.com

Tom Christiansen

tchrist@perl.com

William A. Baumgartner Jr.

Computational Bioscience Program

U. Colorado School of Medicine

william.baumgartner@ucdenver.edu

Karin Verspoor

Computational Bioscience Program

U. Colorado School of Medicine

karin.verspoor@ucdenver.edu

Lawrence E. Hunter

Computational Bioscience Program

U. Colorado School of Medicine

larry.hunter@ucdenver.edu

Abstract

A simple and accurate method for assigning broad semantic classes to text strings is presented. The method is to map text strings to terms in ontologies based on a pipeline of exact matches, normalized strings, headword matching, and stemming headwords. The results of three experiments evaluating the technique are given. Five semantic classes are evaluated against the CRAFT corpus of full-text journal articles. Twenty semantic classes are evaluated against the corresponding full ontologies, i.e. by reflexive matching. One semantic class is evaluated against a structured test suite. Precision, recall, and F-measure on the corpus when evaluating against only the ontologies in the corpus is micro-averaged 67.06/78.49/72.32 and macro-averaged 69.84/83.12/75.31. Accuracy on the corpus when evaluating against all twenty semantic classes ranges from 77.12% to 95.73%. Reflexive matching is generally successful, but reveals a small number of errors in the implementation. Evaluation with the structured test suite reveals a number of characteristics of the performance of the approach.

1 Introduction

Broad semantic class assignment is useful for a number of language processing tasks, including coreference resolution (Hobbs, 1978), document classification (Caporaso et al., 2005), and information extraction (Baumgartner Jr. et al., 2008). A limited number of semantic classes have been studied extensively, such as assigning text strings to the

category *gene* or *protein* (Yeh et al., 2005; Smith et al., 2008), or the *PERSON*, *ORGANIZATION*, and *LOCATION* categories introduced in the Message Understanding Conferences (Chinchor, 1998). A larger number of semantic classes have received smaller amounts of attention, e.g. the classes in the GENIA ontology (Kim et al., 2004), various event types derived from the Gene Ontology (Kim et al., 2009), and diseases (Leaman and Gonzalez, 2008). However, many semantic types have not been studied at all. In addition, where ontologies are concerned, although there has been work on finding mentions or evidence of specific terms in text (Blaschke et al., 2005; Stoica and Hearst, 2006; Davis et al., 2006; Shah et al., 2009), there has been no work specifically addressing assigning multiple very broad semantic classes with potential overlap. In particular, this paper examines the problem of taking a set of ontologies and a text string (typically, but not necessarily, a noun phrase) as input and determining which ontology defines the semantic class that that text string refers to. We make an equivalence here between the notion of belonging to the domain of an ontology and belonging to a specific semantic class. For example, if a string in text refers to something in the domain of the Gene Ontology, we take it as belonging to a Gene Ontology semantic class (using the name of the ontology only for convenience); if a string in text refers to something belonging to the domain of the Sequence Ontology, we take it as belonging to a Sequence Ontology semantic class. We focus especially on rapid, simple methods for making such a determination.

The problem is most closely related to multi-class

classification, where in the case of this study we are including an unusually large number of categories, with possible overlap between them. A text string might refer to something that legitimately belongs to the domain of more than one ontology. For example, it might belong to the semantic classes of both the Gene Ontology and the Gene Regulation Ontology; regulation is an important and frequent concept in the Gene Ontology. This fact has consequences for defining the notion of a false positive class assignment; we return to this issue in the *Results* section.

2 Methods

2.1 Target semantic classes

The following ontologies were used to define semantic classes:

- Gene Ontology
- Sequence Ontology
- Foundational Model of Anatomy
- NCBI Taxonomy
- Chemical Entities of Biological Interest
- Phenotypic Quality
- BRENDA Tissue/Enzyme Source
- Cell Type Ontology
- Gene Regulation Ontology
- Homology Ontology
- Human Disease Ontology
- Human Phenotype Ontology
- Mammalian Phenotype Ontology
- Molecule Role Ontology
- Mouse Adult Gross Anatomy Ontology
- Mouse Pathology Ontology
- Protein Modification Ontology
- Protein-Protein Interaction Ontology

- Sample Processing and Separation Techniques Ontology
- Suggested Ontology for Pharmacogenomics

2.2 Methodology for assigning semantic class

We applied four simple techniques for attempting to match a text string to an ontology. They are arranged in order of decreasing stringency. That is, each subsequent method has looser requirements for a match. This both allows us to evaluate the contribution of each component more easily and, at run time, allows the user to set a stringency level, if the default is not desired.

2.2.1 Exact match

The first and most stringent technique is exact match. (This is essentially the only technique used by the NCBO (National Center for Biomedical Ontology) Annotator (Jonquet et al., 2009), although it can also do substring matching.) We normalize terms in the ontology and text strings in the input for case and look for a match.

2.2.2 Stripping

All non-alphanumeric characters, including whitespace, are deleted from the terms in the ontology and from text strings in the input (e.g. *cadmium-binding* and *cadmium binding* both become *cadmiumbinding*) and look for a match.

2.2.3 Head nouns

This method involves a lightweight linguistic analysis. We traversed each ontology and determined the head noun (see method below) of each term and synonym in the ontology. We then prepared a dictionary mapping from head nouns to lists of ontologies in which those head nouns were found.

Head nouns were determined by two simple heuristics (cf. (Collins, 1999)). For terms fitting the pattern *X of..* (where *of* represents any preposition) the term *X* was taken as the head noun. For all other terms, the rightmost word was taken as the head noun. These two heuristics were applied in sequence when applicable, so that for example *positive regulation of growth* (GO:0045927) becomes *positive regulation* by application of the first heuristic and *regulation* by application of the second heuristic. In the case of some ontologies, very limited pre-

processing was necessary—for example, it was necessary to delete double quotes that appeared around synonyms, and in some ontologies we had to delete strings like *[EXACT SYNONYM]* from some terms before extracting the head noun.

2.2.4 Stemming head nouns

In this technique, the headwords obtained by the previous step were stemmed with the Porter stemmer.

2.3 Corpus and other materials

We made use of three sources in our evaluation. One is the CRAFT (Colorado Richly Annotated Full Text) corpus (Verspoor et al., 2009; Cohen et al., 2010a). This is a collection of 97 full-text journal articles, comprising about 597,000 words, each of which has been used as evidence for at least one annotation by the Mouse Genome Informatics group. It has been annotated with a number of ontologies and database identifiers, including:

- Gene Ontology
- Sequence Ontology
- Cell Type Ontology
- NCBI Taxonomy
- Chemical Entities of Biological Interest (ChEBI)

In total, there are over 119,783 annotations. (For the breakdown across semantic categories, see Table 1.) All of these annotations were done by biological scientists and have been double-annotated with inter-annotator agreement in the nineties for most categories.

The second source is the full sets of terms from the twenty ontologies listed in the Introduction. All of the twenty ontologies that we used were obtained from the OBO portal. Version numbers are omitted here due to space limitations, but are available from the authors on request.

The third source is a structured test suite based on the Gene Ontology (Cohen et al., 2010b). Structured test suites are developed to test the performance of a system on specific categories of input types.

This test set was especially designed to test difficult cases that do not correspond to exact matches of Gene Ontology terms, as well as the full range of types of terms. The test suite includes 300 concepts from GO, as well as a number of transformations of their terms, such as *cells migrated* derived from the term *cell migration* and *migration of cells* derived from *cell migration*, classified according to a number of linguistic attributes, such as length, whether or not punctuation is included in the term, whether or not it includes function (stop) words, etc. This test suite determines at least one semantic category that should be returned for each term. Unlike using the entire ontologies, this evaluation method made detailed error analysis possible. This test suite has been used by other groups for broad characterizations of successes and failures of concept recognizers, and to tune the parameters of concept recognition systems.

2.4 Evaluation

We did three separate evaluations. In one, we compared the output of our system against manually-generated gold-standard annotations in the CRAFT corpus (op. cit.). This was possible only for the ontologies that have been annotated in CRAFT, which are listed above.

In the second evaluation, we used the entire ontologies themselves as inputs. In this method, all responses should be the same—for example, every term from the Gene Ontology should be classified as belonging to the GO semantic class.

In the third, we utilized the structured test suite described above.

2.4.1 Baselines

Two baselines are possible, but neither is optimal. The first would be to use MetaMap (Aronson, 2001), the industry standard for semantic category assignment. (Note that MetaMap assigns specific categories, not broad ones.) However, MetaMap outputs only semantic classes that are elements of the UMLS, which of the ontologies that we looked at, includes only the Gene Ontology. The other is the NCBO Annotator. The NCBO Annotator detects only exact matches (or substring matches) to ontology terms, so it is not clear that it is a strong enough baseline to allow for a stringent analysis of our ap-

proach.

3 Results

We present our results in three sections:

- For the CRAFT corpus
- For the ontologies themselves
- For the Gene Ontology test suite

3.1 Corpus results

Table 1 (see next page) shows the results on the CRAFT corpus if only the five ontologies that were actually annotated in CRAFT are used as inputs. The results are given for stemmed heads. Performance on the four techniques that make up the approach is cumulative, and results for stemmed heads reflects the application of all four techniques. In this case, where we evaluate against the corpus, it is possible to determine false positives, so we can give precision, recall, and F-measures for each semantic class, as well as for the corpus as a whole. Micro-averaged results were 67.06 precision, 78.49 recall, and 72.32 F-measure. Macro-averaged results were 69.84 precision, 83.12 recall, and 75.31 F-measure.

Table 2 (see next page) shows the results for the CRAFT corpus when all twenty ontologies are matched against the corpus data, including the many ontologies that are not annotated in the data. We give results for just the five annotated ontologies below. Rather than calculating precision, recall, and F-measure, we calculate only accuracy. This is because when classes other than the gold standard class is returned, we have no way of knowing if they are incorrect without manually examining them—that is, we have no way to identify false positives. If the set of classes returned included the gold standard class, a correct answer was counted. If the classifier returned zero or more classes and none of them was the gold standard, an incorrect answer was counted. Results are given separately for each of the four techniques. This allows us to evaluate the contribution of each technique to the overall results; the value in each column is cumulative, so the value for *Stemmed head* includes the contribution of all four of the techniques that make up the general approach. Accuracies of 77.12% to 95.73% were achieved, depending on the ontology. We see that

the linguistic technique of locating the head noun makes a contribution to all categories, but makes an especially strong contribution to the Gene Ontology and Cell Type Ontology classes. Stemming of head-words is also effective for all five categories. We see that exact match is effective only for those semantic classes for which terminology is relatively fixed, i.e. the NCBI taxonomy and chemical names. In some of the others, matching natural language text is very difficult by any technique. For example, of the 8,665 Sequence Ontology false negatives in the data reflected in the P/R/F values in Table 1, a full 2,050 are due to the single character +, which does not appear in any of the twenty ontologies that we examined and that was marked by the annotators as a Sequence Ontology term, *wild_type* (SO:0000817).

3.2 Ontology results

As the second form of evaluation, we used the terms from the ontologies themselves as the inputs to which we attempted to assign a semantic class. In this case, no annotation is required, and it is straightforwardly the case that each term in a given ontology should be assigned the semantic class of that ontology. We used only the head noun technique. We did *not* use the exact match or stripping heuristics, since they are guaranteed to return the correct answer, nor did we use stemming. Thus, this section of the evaluation gives us a good indication of the performance of the head noun approach.

As might be expected, almost all twenty ontologies returned results in the 97-100% correct rate. However, we noted much lower performance in two ontologies, the Sequence Ontology and the Molecule Role Ontology. This lower performance reflects a number of preprocessing errors or omissions. The fact that we were able to detect these low-performing ontologies indicates that our evaluation technique in this experiment—trying to match terms from an ontology against that ontology itself—is a robust evaluation technique and should be used in similar studies.

3.2.1 Structured test suite results

The third approach to evaluation involved use of the structured test suite. The structured test suite revealed a number of trends in the performance of the system.

Ontology	Annotations	Precision	Recall	F-measure
Gene Ontology	39,626	66.31	73.06	69.52
Sequence Ontology	40,692	63.00	72.21	67.29
Cell Type Ontology	8,383	53.58	87.27	66.40
NCBI Taxonomy	11,775	96.24	92.51	94.34
ChEBI	19,307	70.07	90.53	79.00
Total (micro-averaged)	119,783	67.06	78.49	72.32
Total (macro-averaged)		69.84	83.12	75.31

Table 1: Results on the CRAFT corpus when only the CRAFT ontologies are used as input. Results are for stemmed heads. Precision, recall, and F-measure are given for each semantic category in the corpus. *Totals* are micro-averaged (over all tokens) and macro-averaged (over all categories), respectively. P/R/F are cumulative, so that the results for stemmed heads reflect the application of all four techniques.

Ontology	Exact	Stripped	Head noun	Stemmed head
Gene Ontology	24.26	24.68	59.18	77.12
Sequence Ontology	44.28	47.63	56.63	73.33
Cell Type Ontology	25.26	25.80	70.09	88.38
NCBI Taxonomy	84.67	84.71	90.97	95.73
ChEBI	86.93	87.44	92.43	95.49

Table 2: Results on the CRAFT corpus when all twenty ontologies are used as input. Accuracy is given for each technique. Accuracy is cumulative, so that accuracy in the final column reflects the application of all four techniques.

- The headword technique works very well for recognizing syntactic variants. For example, if the GO term *induction of apoptosis* is written as *apoptosis induction*, the headword technique allows it to be picked up.
- The headword technique works in situations where text has been inserted into a term. For example, if the GO term *ensheathment of neurons* appears as *ensheathment of some neurons*, the headword technique will allow it to be picked up. If the GO term *regulation of growth* shows up as *regulation of vascular growth*, the headword technique will allow it to be picked up.
- The headword stemming technique allows us to pick up many verb phrases, which is important for event detection and event coreference. For example, if the GO term *cell migration* appears in text as *cells migrate*, the technique will detect it. The test suite also showed that failures to recognize verb phrases still occur when the morphological relationship between the nominal term and the verb are irregular, as for example between the GO term *growth* and the verb *grows*.
- The technique’s ability to handle coordination is very dependent on the type of coordination. For example, simple coordination (e.g. *cell migration and proliferation*) is handled well, but complex coordination (e.g. *cell migration, proliferation and adhesion*) is handled poorly.
- Stemming is necessary for recognition of plurals, regardless of the length of the term in words.
- The approach currently fails on irregular plurals, due to failure of the Porter stemmer to handle plurals like *nuclei* and *nucleoli* well.
- The approach handles classification of terms that others have characterized as “ungrammatical,” such as *transposition*, *DNA-mediated* (GO:0006313). This is important, because exact matches will always fail on these terms.

4 Discussion

4.1 Related work

We are not aware of similar work that tries to assign a large set of broad semantic categories to individual text strings. There is a body of work on selecting a single ontology for a domain or text. (Martínez-Romero et al., 2010) proposes a method for selecting an ontology given a list of terms, all of which must appear in the ontology. (Jonquet et al., 2009) describes an ontology recommender that first annotates terms in a text with the Open Biomedical Annotator service, then uses the sum of the scores of the individual annotations to recommend a single ontology for the domain as a whole.

4.2 Possible alternate approaches

Three possible alternative approaches exist, all of which would have as their goal the returning of a single best semantic class for every input. However, for the use cases that we have identified—coreference resolution, document classification, information extraction, and curator assistance—we are more interested in wide coverage of a broad range of semantic classes, so these approaches are not evaluated here. However, we describe them for completeness and for the use of researchers who might be interested in pursuing single-class assignment.

4.2.1 Frequent words

One alternative approach would be to use simple word frequencies. For example, for each ontology, one could determine the N most frequent words, filtering out stop words. At run time, check the words in each noun phrase in the text against the lists of frequent words. For every word from the text that appeared in the list of frequent words from some ontology, assign a score to each ontology in which it was found, weighting it according to its position in the list of frequent words. In theory, this could accommodate for the non-uniqueness of word-to-ontology mappings, i.e. the fact that a single word might appear in the lists for multiple ontologies. However, we found the technique to perform very poorly for differentiating between ontologies and do not recommend it.

4.2.2 Measuring informativeness

If the system is desired to return only one single semantic class per text string, then one approach would be to determine the informativeness of each word in each ontology. That is, we want to find the maximal probability of an ontology given a word from that ontology. This approach is very difficult to normalize for the wide variability in size of the many ontologies that we wanted to be able to deal with.

4.2.3 Combining scores

Finally, one could conceivably combine scores for matches obtained by the different strategies, weighting them according to their stringency, i.e. exact match receiving a higher weight than head noun match, which in turn would receive a higher weight than stemmed head noun match. This weighting might also include informativeness, as described above.

4.3 Why the linguistic method works

As pointed out above, the lightweight linguistic method makes a large contribution to the performance of the approach for some ontologies, particularly those for which the exact match and stripping techniques do not perform well. It works for two reasons, one related to the approach itself and one related to the nature of the OBO ontologies. From a methodological perspective, the approach is effective because headwords are a good reflection of the semantic content of the noun phrase and they are relatively easy to access via simple heuristics. Of course simple heuristics will fail, as we can observe most obviously in the cases where we failed to identify members of the ontologies in the second evaluation step. However, overall the approach works well enough to constitute a viable tool for coreference systems and other applications that benefit from the ability to assign broad semantic classes to text strings.

The approach is also able to succeed because of the nature of the OBO ontologies. OBO ontologies are meant to be orthogonal (Smith et al., 2007). A distributional analysis of the distribution of terms and words between the ontologies (data not shown here, although some of it is discussed below), as well as the false positives found in the corpus study, sug-

gests that orthogonality between the OBO ontologies is by no means complete. However, it holds often enough for the headword method to be effective.

4.4 Additional error analysis

In the section on the results for the structured test suite, we give a number of observations on contributions to errors, primarily related either to the characteristics of individual words or to particular syntactic instantiations of terms. Here, we discuss some aspects of the distribution of lexical items and of the corpus that contributed to errors.

- The ten most common headwords appear in from 6-16 of the twenty ontologies. However, they typically appear in one ontology at a frequency many orders of magnitude greater than their frequency in the other ontologies. Taking this frequency data into account for just these ten headwords would likely decrease false positives quite significantly.
- More than 50% of Gene Ontology terms share one of only ten headwords. Many of our Gene Ontology false negatives on the corpus are because the annotated text string does not contain a word such as *process* or *complex* that is the head word of the canonical term.

4.5 Future work

The heuristics that we implemented for extracting headwords from OBO terms were very simple, in keeping with our initial goal of developing an easy, fast method for semantic class assignment. However, it is clear that we could achieve substantial performance improvements from improving the heuristics. We may pursue this track, if it becomes clear that coreference performance would benefit from this when we incorporate the semantic classification approach into a coreference system.

On acceptance of the paper, we will make Perl and Java versions of the semantic class assigner publicly available on SourceForge.

4.6 Conclusion

The goal of this paper was to develop a simple approach to assigning text strings to an unprecedentedly large range of semantic classes, where mem-

bership in a semantic class is equated with belonging to the semantic domain of a specific ontology. The approach described in this paper is able to do that at a micro-averaged F-measure of 72.32 and macro-averaged F-measure of 75.31 as evaluated on a manually annotated corpus where false positives can be determined, and with an accuracy of 77.12-95.73% when only true positives and false negatives can be determined.

References

- A. Aronson. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. In *Proc AMIA 2001*, pages 17–21.
- William A. Baumgartner Jr., Zhiyong Lu, Helen L. Johnson, J. Gregory Caporaso, Jesse Paquette, Anna Lindemann, Elizabeth K. White, Olga Medvedeva, K. Bretonnel Cohen, and Lawrence Hunter. 2008. Concept recognition for extracting protein interaction relations from biomedical text. *Genome Biology*, 9.
- Christian Blaschke, Eduardo A. Leon, Martin Krallinger, and Alfonso Valencia. 2005. Evaluation of BioCreative assessment of task 2. *BMC Bioinformatics*, 6 Suppl 1.
- J. Gregory Caporaso, William A. Baumgartner Jr., K. Bretonnel Cohen, Helen L. Johnson, Jesse Paquette, and Lawrence Hunter. 2005. Concept recognition and the TREC Genomics tasks. In *The Fourteenth Text REtrieval Conference (TREC 2005) Proceedings*.
- Nancy A. Chinchor. 1998. Overview of MUC-7/MET-2.
- K. Bretonnel Cohen, Helen L. Johnson, Karin Verspoor, Christophe Roeder, and Lawrence E. Hunter. 2010a. The structural and content aspects of abstracts versus bodies of full text journal articles are different. *BMC Bioinformatics*, 11(492).
- K. Bretonnel Cohen, Christophe Roeder, William A. Baumgartner Jr., Lawrence Hunter, and Karin Verspoor. 2010b. Test suite design for biomedical ontology concept recognition systems. In *Proceedings of the Language Resources and Evaluation Conference*.
- Michael Collins. 1999. *Head-driven statistical models for natural language parsing*. Ph.D. thesis, University of Pennsylvania.
- N. Davis, H. Harkema, R. Gaizauskas, Y. K. Guo, M. Ghanem, T. Barnwell, Y. Guo, and J. Ratcliffe. 2006. Three approaches to GO-tagging biomedical abstracts. In *Proceedings of the Second International Symposium on Semantic Mining in Biomedicine*, pages 21–28, Jena, Germany.
- Jerry R. Hobbs. 1978. Resolving pronoun references. *Lingua*, 44:311–338.

- C. Jonquet, N.H. Shah, and M.A. Musen. 2009. Prototyping a biomedical ontology recommender service. In *Bio-Ontologies: Knowledge in Biology, ISMB/ECCB SIG*.
- Jin-Dong Kim, Tomoko Ohta, Yoshimasa Tsuruoka, Yuka Tateisi, and Nigel Collier. 2004. Introduction to the bio-entity recognition task at JNLPBA. In *Proceedings of the international joint workshop on natural language processing in biomedicine and its applications*, pages 70–75.
- Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP'09 shared task on event extraction. In *BioNLP 2009 Companion Volume: Shared Task on Entity Extraction*, pages 1–9.
- Robert Leaman and Graciela Gonzalez. 2008. BANNER: An executable survey of advances in biomedical named entity recognition. In *Pac Symp Biocomput.*
- Marcos Martínez-Romero, José Vázquez-Naya, Cristian R. Munteanu, Javier Pereira, and Alejandro Pazos. 2010. An approach for the automatic recommendation of ontologies using collaborative knowledge. In *KES 2010, Part II, LNAI 6277*, pages 74–81.
- Nigam H. Shah, Nipun Bhatia, Clement Jonquet, Daniel Rubin, Annie P. Chiang, and Mark A. Musen. 2009. Comparison of concept recognizers for building the Open Biomedical Annotator. *BMC Bioinformatics*, 10.
- Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J. Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J. Mungall, The OBI Consortium, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard H. Scheuermann, Nigam Shah, Patricia L. Whetzel, and Suzanna Lewis. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25:1251–1255.
- Larry Smith, Lorraine Tanabe, Rie Johnson nee Ando, Cheng-Ju Kuo, I-Fang Chung, Chun-Nan Hsu, Yu-Shi Lin, Roman Klinger, Christof Friedrich, Kuzman Ganchev, Manabu Torii, Hongfang Liu, Barry Haddow, Craig Struble, Richard Povinelli, Andreas Vlachos, William Baumgartner, Jr., Lawrence Hunter, Bob Carpenter, Richard Tzong-Han Tsai, Hong-Jie Dai, Feng Liu, Yifei Chen, Chengjie Sun, Sophia Katrenko, Pieter Adriaans, Christian Blaschke, Rafael Torres Perez, Mariana Neves, Preslav Nakov, Anna Divoli, Manuel Mana, Jacinto Mata-Vazquez, and W. John Wilbur. 2008. Overview of BioCreative II gene mention recognition. *Genome Biology*.
- E. Stoica and M. Hearst. 2006. Predicting gene functions from text using a cross-species approach. In *Proceedings of the 11th Pacific Symposium on Biocomputing*.
- Karin Verspoor, K. Bretonnel Cohen, and Lawrence Hunter. 2009. The textual characteristics of traditional and Open Access scientific journals are similar. *BMC Bioinformatics*, 10.
- A. Yeh, A. Morgan, M. Colosimo, and L. Hirschman. 2005. BioCreative task 1A: gene mention finding evaluation. *BMC Bioinformatics*, 6(Suppl. 1).