

Performance of Gene Name Recognition Tools on Patents

Maryam Habibi¹, David Luis Wiegandt¹, Florian Schmedding², and Ulf Leser¹

¹Knowledge Management in Bioinformatics, Humboldt-Universität zu Berlin, 12489 Berlin, Germany

²Averbis GmbH, 79106 Freiburg im Breisgau, Germany

¹{habibima,wigandtd,leser}@informatik.hu-berlin.de

² florian.schmedding@averbis.com

Abstract

The accurate identification of gene and protein names in patents is an essential step in many commercially highly relevant applications, such as patent retrieval, prior art search, or patent classification. Since patents exhibit a number of properties that make them quite different from scientific articles, it is questionable whether tools developed for the latter sort of texts will work equally well for the former. Answering this question is aggravated by the fact that only few annotated patent corpora exist which makes training hard. In this paper, we report on a comparative evaluation of four existing gene/protein named entity recognition and normalization tools trained on scientific articles regarding their performance on the two patent corpora. We analyze the tools with respect to different evaluation metrics to highlight their respective strengths and limitations. Our results reveal that the performances of these tools over patents are generally lower than for scientific articles. Exemplified by one of the four tools, we also show that training on annotated patents considerably improves performance on patent corpora. We conclude that more efforts must be taken to produce adequate training data for working with patents.

keywords: Patent Mining, Named Entity Recognition, Named Entity Normalization, Gene and Protein Entities, Performance Measurements.

1 Introduction

Patents are an important source of knowledge for biomedical research, yet receive compara-

bly little attention from the scientific community (Rodriguez-Esteban and Bundschus, 2016). While a multitude of annotated collections of scientific articles (mostly abstracts) for different classes of entities exist, only few such resources are available based on patents, although patents, like scientific abstracts, in principle are available to the public without fees. Patents are of paramount importance for many commercial activities in the field; accordingly, one may suspect that pharmaceutical and biotech companies should be highly interested in methods to automatically analyze patents, but, given the relative low number of publications on patent mining, they are obviously less interested in publishing their results or making their resources freely available (Roberts and Hayes, 2008).

There are multiple ways in which patents can be automatically analyzed by computer programs, such as patent classification (Iwayama et al., 2007), prior-art search (Harris et al., 2010), or search engines for patents (Lupu and Hanbury, 2013). In this work, we are interested in information extraction from patents; more specifically, we study the problem of recognizing (NER, named entity recognition) and normalizing (NEN, named entity normalization) names of genes and proteins (henceforth only called genes) in patents. Both gene NER and NEN are well researched problems in scientific articles (Leser and Hakenberg, 2005), but only few results exist for patents (Rodriguez-Esteban and Bundschus, 2016).

The most notable attempt to this problem was performed through the recent GPRO task (gene and protein related object task) part of the BioCreative V challenge (Krallinger et al., 2015). In this competition, teams had to extract mentions of genes and proteins from two manually annotated gold standard patent corpora consisting of patents' titles and abstracts. However, the

number of participating teams was rather limited; only four teams participated, submitting 16 runs. Furthermore, the evaluation was performed only at the mention level before normalization; thus the performance of gene name normalization was not assessed. The highest F-measure score at this challenge was 81.37 achieved by Leaman et al. (2015), which is a performance quite inferior to that achieved at gene NER tasks for scientific articles.

Given the lack of annotated patent corpora, it is tempting to try reusing models trained on other types of texts, especially on scientific abstracts as here a variety of gold standards are freely available. This approach was actually also taken by Leaman et al. (2015) which used an ensemble of different instantiations of the base tool GNormPlus (Wei et al., 2015) trained over different corpora. The question remains whether tools other than GNormPlus would be equally (or more or less) suited for such cross-text-type applications. To approach this question, we performed a comparative evaluation of four state-of-the-art gene NER/NEN tools on the two GPRO patent corpora, using their original models which were all trained on corpora made of scientific abstracts. Specifically, we compare GNAT (Solt et al., 2010), Gimli (Campos et al., 2013), GNormPlus (Wei et al., 2015), and GeneTuKit (Huang et al., 2011). We measure their execution time, describe their performance in terms of tagging and normalization quality and discuss their strengths and limitations using evaluations both at the mention level and at the document level. Furthermore, we report on the mention level performance of a high-recall ensemble, made of unifying the tagging outputs of GNormPlus and Gimli, and a high-precision ensemble, created by intersecting the tagging results of GNormPlus and GNAT. Eventually, we showcase the impact of the cross-text-type application by comparing the tagging performance of Gimli when trained on scientific abstracts with that when trained on patent abstracts.

Overall, our evaluation produces a diverse picture; GNAT seems to be the only tool fast enough to be applicable to truly large patent corpora and also achieves the best NER and NEN precision but a rather low recall; Gimli achieves a competitively high F-measure compared to GNormPlus, which attained the best F-measure, in cross-text-type evaluation but is very slow and does not per-

form entity normalization; and evaluation results improve considerably when systems are trained on documents of the same text type. We conclude that more efforts must be taken to produce adequate training data for working with patents.

2 Method

In this section, we present the existing gold standard patent corpora containing gene and protein annotations. We then describe a variety of gene/protein NER tools, and more specifically describe four freely available gene/protein NER systems studied here. Finally we explain the evaluation metrics used in our study.

2.1 Annotated Patent Corpora

We utilize two gold standard patent corpora containing annotations for genes and proteins. These two corpora are designed as training and development sets for GPRO task (gene and protein related object task) (Krallinger et al., 2015) at BioCreative V challenge. The two sets are noted GPRO_T and GPRO_D, each containing the title and the abstract of 7000 patents manually annotated using the same annotation guideline. Table 1 represents the details of these gold standard corpora including corpus size (the number of tokens separated by space), the number of patents, and the number of annotated entities.

Table 1: Details of the gold standard patent corpora containing the annotations for genes/proteins.

Corpus	Number of Patents	Number of Annotations
GPRO training set (GPRO_T) ≈ 662234 token (Krallinger et al., 2015)	7000 patents (title and abstracts)	4396 annotations
GPRO development set (GPRO_D) ≈ 648732 token (Krallinger et al., 2015)	7000 patents (title and abstracts)	3934 annotations

The annotated entities in both corpora can be assigned to one of “NESTED MENTIONS”, “IDENTIFIER”, “FULL NAME”, and “ABBREVIATION” classes. Moreover, these annotated entities are normalized to identifiers from gene databases such as UniProt, GenBank, HGNC, RefSeq, Ensembl, and so on. Since the tools, studied here, normalize detected entities into identifiers from EntrezGene database, we convert identifiers in the gold standard corpora into the ones

from EntrezGene database using *BioMart*¹ web service. We consider a subset of annotated entities that can be normalized to identifiers from EntrezGene database in our evaluation.

2.2 Gene NER Systems

Several NER tools for genes and proteins have been developed using machine learning approaches with conditional random fields (CRFs) such as GNORMPlus (Wei et al., 2015), AIIA Gene Mention Tagger (Hsu et al., 2008), GENIA Tagger (Kulick et al., 2004), Moara (Neves et al., 2010), tagtag (Cejuela et al., 2014), Gimli (Campos et al., 2013), GNAT (Solt et al., 2010), and GenTuKit (Huang et al., 2011). Among these various tools, we have chosen the ones with highest performance on scientific articles, GNORMPlus, GNAT, Gimli and GeneTuKit. GNAT and Gimli have been known as the baseline tools with the state-of-the-art performance on scientific articles. GeneTuKit has been selected as one of the high performance tools in gene normalization task at BioCreative III challenge (Lu et al., 2011), while GNORMPlus (Wei et al., 2015) has shown an improvement over the high performance tools at this challenge. Since these four tools are trained on at least one common corpus, their performances are comparable.

Table 2 summarizes the tokenization methods, and the training sets used by the mentioned NER tools. GNAT, Gimli and GNORMPlus train CRF models to recognize entities. GNAT and Gimli train CRF models using the BANNER implementation (Leaman and Gonzalez, 2008) while GNORMPlus uses the CRF++² implementation. The systems differ in their tokenization methods while all are using the BioCreative II GM corpus as part of their training sets to produce the models.

GeneTuKit as another NER system which we used here extracts and ranks entities based on a confidence score using an ensemble approach but it does not provide any information about the position of entities. GeneTuKit selects an entity if two methods out of three recognize it as a gene or a protein. The first method is a CRF model trained using BANNER similar to the three gene/protein NER tools mentioned above. The second method recognizes entities based on the ones listed in EntrezGene database. The third one is a CRF-

based method developed by training ABNER NER tool (Settles, 2005) using 32 full texts provided by BioCreative III challenge.

All the mentioned NER tools normalize detected entities into identifiers from EntrezGene database except Gimli which does not provide any normalization information for detected spans.

2.3 Evaluation Metrics

We first compare NER tools in terms of their execution time over full patent documents randomly chosen from *European Patent Office*³. Moreover, we compare the performance of these tools in terms of precision, recall, F-measure, true positive (TP), false positive (FP), and false negative (FN) counts. We perform exact matching to compute all the evaluation scores. The experiments are performed over the mentions and also the identifiers recognized by systems.

To calculate the performance values, both prediction and gold standard annotation files are converted into files in IOB format suggested by Klinger et al. (Klinger et al., 2008). In this format, every non-letter and non-digit character, and all number-letter changes are split and each token is represented by one of three chunk tags, B (begin), I (inside), O (outside). Then the evaluation scores are obtained using the commonly used conlleval⁴ script. We define the performance values at both the mention and the document levels in the following.

2.3.1 Mention Level Performance Values

The mention level scores are computed by considering the position of entities in each document. Precision measures the ratio of predicted gene mentions or identifiers assigned to the entities which are exactly matched with gene mentions or identifiers annotated in a gold standard corpus. Similarly, recall is measured as the ratio of gene mentions or identifiers in a gold standard corpus that appear at exactly the same location in prediction files. F-measure is the harmonic mean of the precision and the recall values. TP measures the number of gene mentions or identifiers in a gold standard corpus which are correctly recognized by a NER tool. FP is calculated by counting the number of spans or identifiers which are incorrectly recognized as gene mentions by NER tools.

¹See <http://www.biomart.org/>

²See <http://crfpp.googlecode.com/svn/trunk/doc/index.html>

³See <https://www.epo.org/>

⁴The tool is freely available at <http://www.cnts.ua.ac.be/conll2000/chunking>

Table 2: Details of the gene/protein NER tools in terms of their training sets, and tokenization method.

NER tool	NER training set	Tokenization method
GNAT (Solt et al., 2010)	-BioCreative II GM data	Tokenization at spaces, numbers, and punctuation marks
Gimli (Campos et al., 2013)	-BioCreative II GM data	Tokenization at every non-letter and non-digit characters
GeneTuKit (Huang et al., 2011)	-BioCreative II GM data -BioCreative III full text data	Semantic tokenization
GNormPlus (Wei et al., 2015)	-BioCreative II GM data -NLM Citation GIA collection	Tokenization at spaces, numbers and punctuation marks, and also transitions between uppercase and lowercase letters

FN is the number of gene spans or identifiers in a gold standard corpus which are not recognized by a NER tool.

Since both corpora are annotated using the same annotation guideline, we report the micro-average performance values at the mention level. These values are calculated by averaging over TP, FP, and FN counts computed for two patent corpora.

2.3.2 Document Level Performance Measurements

The scores at document level are calculated by disregarding the position of entities in documents. In addition, all duplicate occurrences of a mention or an identifier are ignored in our assessments. Precision at the document level is measured as the ratio of correctly predicted gene names or identifiers, among all the recognized entities or identifiers averaged over all documents. Recall is defined here as the ratio of correctly recognized gene names or identifiers divided by the total number of annotated gene entities or gene identifiers of a document which is averaged over all the documents in the corpus. F-measure is the harmonic mean of the precision and the recall values computed at the document level. Here, all the scores are calculated by averaging over all the patent texts in both corpora.

3 Results

We first compare NER tools in terms of their execution time over patent documents. Then, we assess the performance of each tool’s default model at both the mention and the document levels before and after normalization on patents.

3.1 Execution Time Analysis

We compare the execution time of these tools over 10 complete patents that have been randomly chosen from *European Patent Office* in domains of medicine, biochemistry, and biology.

The tools are run sequentially on two different machines. All the tools except Gimli are run over a machine (m_1) utilizing 2 Intel Core(TM) i5-3320M CPUs @ 2.6 GHz, 4GB RAM memory and Microsoft Windows operating system. We excluded Gimli because it does not normalize entities; therefore its execution time is not comparable with the others. Additionally, Gimli can be executed only on Mac iOS or Linux operating systems. Gimli requires a large amount of memory size, for which we exploit another machine (m_2) composed of 120 Intel Xeon CPUs @ 2.5 GHz, 1TB RAM and Linux operating system.

The execution time of each tool, run on one thread is reported in seconds and provided in Table 3. GNAT only requires 205 seconds to complete the task, and therefore it is the fastest NER tool rather than the others. GeneTuKit, is 10 times slower and requires 1970 seconds to finish the task. GNormPlus with execution time of 13415 seconds is a slower tool compared with both GNAT and GeneTuKit. The last one, Gimli, has the worst execution time (349999 seconds) while it utilizes a more powerful machine, m_2 . Consequently, the tools can be ranked, according to their execution time, from slow to fast as follows: Gimli < GNormPlus < GeneTuKit < GNAT.

To show a feeling of the execution times of these systems, we estimate them over 10 million patent documents using 8 parallel threads by extrapolating the above values. The estimated execution time for GNAT, GeneTuKit, and GNormPlus

Table 3: Execution time, in seconds, of four tools which are run on 10 full patent documents. The fastest tool, GNAT, is highlighted in bold font.

Machine	Gene/protein NER tool			
	GNAT	GeneTuKit	GNormPlus	Gimli
m_1	205	1970	13415	–
m_2	–	–	–	349999

are nearly 10 months, 8 years, 55 years, respectively which are too short compared to 1406 years for Gimli. The results indicate that the usage of these tools appears not to be a practical approach in reality except GNAT given large parallel systems.

3.2 Cross-Text-Type Comparison

We compare the performances of gene NER tools over patents at both the mention level and the document level (see Section 2.3) before and after normalization. The experiments are performed using the default models trained on non-patent articles.

3.2.1 Performance Values at Mention level

We calculate the performance values, i.e., precision, recall, and F-measure at the mention level for three gene/protein NER systems, GNAT, GNormPlus and Gimli before normalization as shown in Figure 1. The highest precision is obtained by GNAT and the highest recall is attained by Gimli over both corpora, while GNormPlus offers the highest F-measure value. We can infer that the superiority of each tool over patents depends on the application.

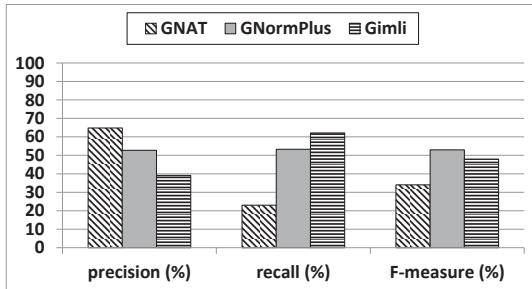


Figure 1: Evaluation scores at mention level in terms of precision, recall, and F-measure values over recognized spans by three gene/protein NER tools, GNAT, GNormPlus and Gimli.

The F-measure values obtained for GNAT, Gimli, and GNormPlus are 34.02%, 53.00%, and 48.00% respectively. These values are extremely

low, which implies that the models trained on scientific articles are not suitable enough to recognize gene and protein names from patent documents with complex writing structures. Moreover, we can rank the systems in terms of their F-measure values from low to high as follows: GNAT <Gimli<GNormPlus.

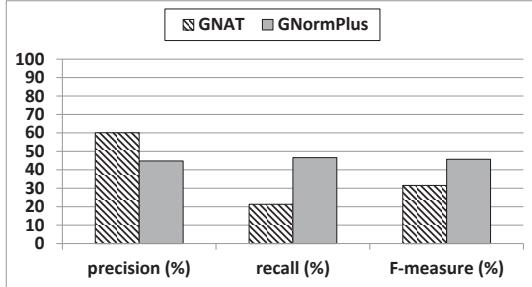


Figure 2: Evaluation scores at mention level with respect to precision, recall and F-measure values calculated for two gene NER tools after normalization.

In the following, we measure the performance values of gene/protein NER tools GNAT, and GNormPlus after normalization as depicted in Figure 2. We do not provide any values for Gimli, because it does not normalize recognized spans. We observe that their performance values are reduced after normalization. The precision values of both systems are reduced by around 6%. However, the decrease in the recall values of GNormPlus is higher than those of GNAT. The results also show that the highest precision on normalized entities is achieved by GNAT, whereas GNormPlus has the highest recall and F-measure values. We can rank two tools with respect to their F-measure values from low to high as follows: GNAT<GNormPlus. This ranking is completely in inverse order to the one obtained for their speed.

3.2.2 Performance Values at Document Level

We measure the performance values at the document level by ignoring the position of entities in documents as explained in Section 2.3.2. The results of four systems before normalization are provided in Figure 3. The results indicate that the performance values computed over the results of GNormPlus and Gimli are highly competitive and outperform those of GNAT and GeneTuKit. We can also rank the tools from low to high as follows: GeneTuKit<GNAT<Gimli<GNormPlus.

Similarly, we calculate document level perfor-

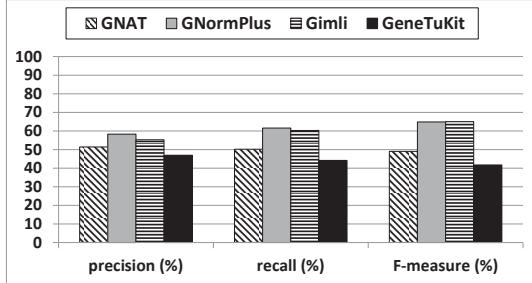


Figure 3: Evaluation scores at document level in terms of precision, recall, and F-measure using four gene NER tools before normalization.

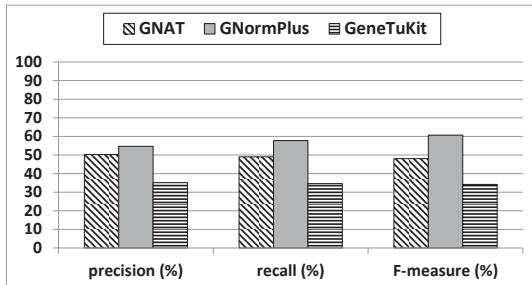


Figure 4: Evaluation scores at document level with respect to precision, recall, and F-measure using three gene NER tools after normalization.

mance values after normalization. The values for GNAT, GNormPlus and GeneTuKit are shown in Figure 4. We find no significant variation between the performance values of each tool before and after normalization at document level. Therefore, GNormPlus again outperforms the others, and the ranking obtained above for the systems remains unchanged after normalization.

4 Discussion

We have evaluated the performance of each individual NER tool in terms of their tagging and normalization quality on patents. Since we have observed big differences among their performance values, we have been motivated to measure the performance of ensembles built by unifying or intersecting the tagging outputs of pairs of systems. Additionally, we observed that the performances of default models trained on scientific articles are quite low on patents; therefore, we assess the impact of using patent training sets on the performance values measured on patent data. All the evaluations are performed at the mention level before normalization.

4.1 Ensemble Assessment

We compute the performance values obtained by intersecting or unifying the outputs of pairs of systems. Figures 5, and 6 respectively represent the micro-average precision, recall, and F-measure values obtained by intersecting and unifying the results of pairs. The highest precision achieved by intersecting the results of GNAT and Gimli which is at least 5% higher than the one obtained by one of the systems individually. Similarly, the highest recall value is attained by unifying the results of Gimli and GNormPlus which brought an improvement of around 10% over that of individual systems. Likewise, we observe an improvement of at least 2% on the F-measure value, obtained by unifying GNAT and GNormPlus outputs, compared to that of individual systems.

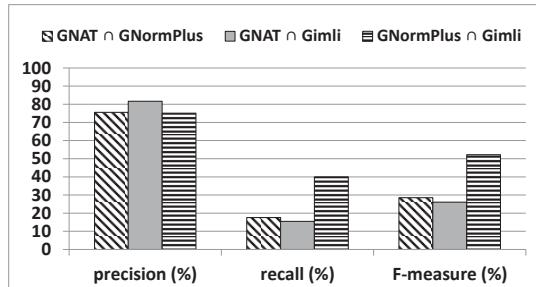


Figure 5: Evaluation scores calculated over the intersection of the pairs’ outputs at the mention level before normalization.

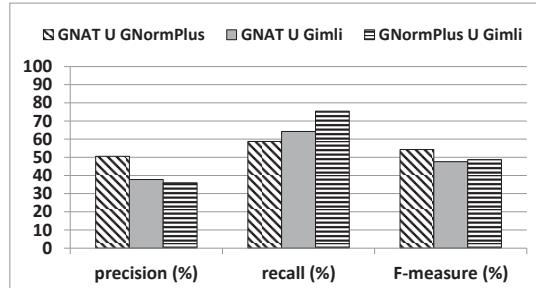


Figure 6: Evaluation scores calculated over the union of the pairs’ outputs at the mention level before normalization.

4.2 Retraining Using Patent Corpora

We retrain one of the four systems using patents to investigate whether exploiting patent training sets can enhance the tagging quality. As we did not found clear explanation about retraining procedure

in public API of GNormPlus as the best performing tool, we retrained Gimli as the second system with highest performance on patents.

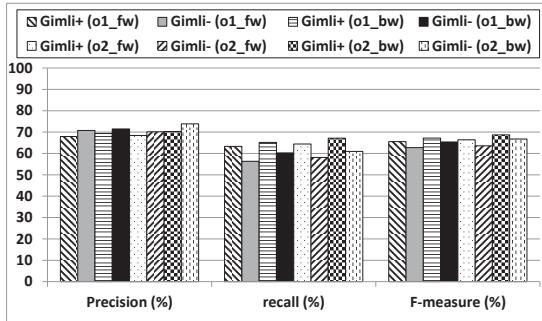


Figure 7: Evaluation scores in terms of precision, recall and F-measure are provided for the new models obtained by retraining Gimli using patent corpora. “+” and “-” denote GPRO_T and GPRO_D training sets respectively.

Gimli is retrained using both first-order and second-order CRF models with both forward (left to right) and backward (right to left) text parsing over patent corpora. 75% of each patent corpus, which is randomly selected, is used for training purpose, and the remaining samples are considered as the test set. Figure 7 represents the micro-average performance values of models trained using patent corpora.

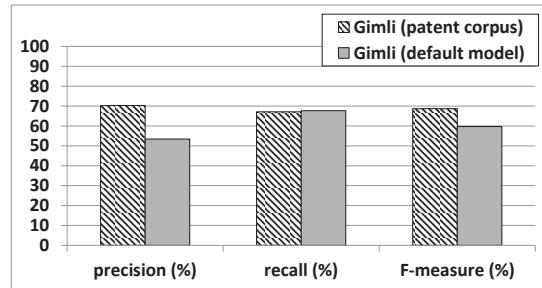


Figure 8: Evaluation scores in terms of precision, recall and F-measure are provided for the best model trained using patent corpora and the pre-trained model of Gimli.

We observe that, among different models obtained by retraining Gimli, the second order ones, using backward parsing, outperform the others. This finding is in agreement with the results of Gimli trained and tested on scientific journal articles (Campos et al., 2013).

We then compare the performance of the best model trained on patents, with the one trained on

scientific articles, in terms of their performance scores at the mention level. The precision, recall and F-measure values are provided in Figure 8. The results show that training Gimli using patent documents will improve precision by at least 15% and F-measure by around 10%, while the recall values are remained unchanged. Thus, utilizing patent texts for both training and test purposes will considerably improve tagging quality on patents.

5 Conclusion

In this paper, we measured the performance of several high performance gene/protein NER tools over available patent corpora. We compared their pre-trained models’ evaluated on patents at both the mention and the document levels before and after normalization. We observed that GNormPlus usually outperforms the others but it is limited by its long execution time over patents. However, we have shown that running GNAT over a huge number of patent documents will provide higher precision values within a reasonable execution time.

In addition to the comparison performed between the outputs of systems individually, we compared the performance of ensembles constructed by merging the results of pairs of systems. The results implied that using ensembles improves precision, recall and even F-measure scores. Finally we retrained Gimli using patent training sets, and observed a remarkable improvement in terms of precision and F-measure values on patents compared to those trained on scientific articles, which confirms the necessity of creating more annotated patent corpora.

Acknowledgments

The authors are grateful to the Federal Ministry for Economic Affairs and Energy (BMWi) for its financial support through the BioPatent project [KF2205219BZ4].

References

- David Campos, Sérgio Matos, and José Luís Oliveira. 2013. Gimli: open source and high-performance biomedical name recognition. *BMC bioinformatics*, 14(1).
- Juan Miguel Cejuela, Peter McQuilton, Laura Ponting, Steven J Marygold, Raymund Stefancsik, Gillian H Millburn, Burkhard Rost, FlyBase Consortium, et al. 2014. tagtog: interactive and text-mining-assisted

- annotation of gene mentions in PLOS full-text articles. *Database*, 2014:1–8.
- Christopher G Harris, Robert Arens, and Padmini Srinivasan. 2010. Comparison of IPC and USPC classification systems in patent prior art searches. In *Proceedings of the 3rd International Workshop on Patent Information Retrieval*, pages 27–32.
- Chun-Nan Hsu, Yu-Ming Chang, Cheng-Ju Kuo, Yu-Shi Lin, Han-Shen Huang, and I-Fang Chung. 2008. Integrating high dimensional bi-directional parsing models for gene mention tagging. *Bioinformatics*, 24(13):i286–i294.
- Minlie Huang, Jingchen Liu, and Xiaoyan Zhu. 2011. GeneTUKit: a software for document-level gene normalization. *Bioinformatics*, 27(7):1032–1033.
- Makoto Iwayama, Atsushi Fujii, and Noriko Kando. 2007. Overview of classification subtask at NTCIR-6 patent retrieval task. In *Proceedings of NTCIR-5 Workshop Meeting*.
- Roman Klinger, Corinna Kolářík, Juliane Fluck, Martin Hofmann-Apitius, and Christoph M Friedrich. 2008. Detection of IUPAC and IUPAC-like chemical names. *Bioinformatics*, 24(13):i268–i276.
- Martin Krallinger, Obdulia Rabal, Analia Lourenço, Martin Perez Perez, Gael Perez Rodriguez, Miguel Vazquez, F Leitner, J Oyarzabal, and A Valencia. 2015. Overview of the CHEMDNER patents task. In *Proceedings of the Fifth BioCreative Challenge Evaluation Workshop*, pages 63–75.
- Seth Kulick, Ann Bies, Mark Liberman, Mark Mandel, Ryan McDonald, Martha Palmer, Andrew Schein, Lyle Ungar, Scott Winters, and Pete White. 2004. Integrated annotation for biomedical information extraction. In *Proceedings of the Human Language Technology Conference and the Annual Meeting of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*, pages 61–68.
- Robert Leaman and Graciela Gonzalez. 2008. BANNER: an executable survey of advances in biomedical named entity recognition. In *Pacific Symposium on Biocomputing*, volume 13, pages 652–663.
- Robert Leaman, Chih-Hsuan Wei, Cherry Zou, and Zhiyong Lu. 2015. Mining patents with tmChem, GNORMPlus and an ensemble of open systems. In *Proceedings of the Fifth BioCreative Challenge Evaluation Workshop*, pages 140–146.
- Ulf Leser and Jörg Hakenberg. 2005. What makes a gene name? Named entity recognition in the biomedical literature. *Briefings in bioinformatics*, 6(4):357–369.
- Zhiyong Lu, Hung-Yu Kao, Chih-Hsuan Wei, Minlie Huang, Jingchen Liu, Cheng-Ju Kuo, Chun-Nan Hsu, Richard Tzong-Han Tsai, Hong-Jie Dai, Naoaki Okazaki, et al. 2011. The gene normalization task in BioCreative III. *BMC bioinformatics*, 12(8):1.
- Mihai Lupu and Allan Hanbury. 2013. Patent retrieval. *Foundations and Trends in Information Retrieval*, 7(1):1–97.
- Mariana L Neves, José Maria Carazo, and Alberto Pascual-Montano. 2010. Extracting and normalizing gene/protein mentions with the flexible and trainable moara java library. In *Linking Literature, Information, and Knowledge for Biology*, pages 71–80. Springer.
- Phoebe M Roberts and William S Hayes. 2008. Information needs and the role of text mining in drug development. In *Pacific Symposium on Biocomputing*, volume 13, pages 592–603.
- Raul Rodriguez-Esteban and Markus Bundschatz. 2016. Text mining patents for biomedical knowledge. *Drug discovery today*.
- Burr Settles. 2005. ABNER: an open source tool for automatically tagging genes, proteins and other entity names in text. *Bioinformatics*, 21(14):3191–3192.
- Illés Solt, Martin Gerner, Philippe Thomas, Goran Nejad, Casey M Bergman, Ulf Leser, and Jörg Hakenberg. 2010. Gene mention normalization in full texts using GNAT and LINNAEUS. In *Proceedings of the BioCreative III workshop (Bethesda, USA)*, pages 134–139.
- Chih-Hsuan Wei, Hung-Yu Kao, and Zhiyong Lu. 2015. GNORMPlus: An integrative approach for tagging genes, gene families, and protein domains. *BioMed research international*, 2015.