Overview of the Cancer Genetics (CG) task of BioNLP Shared Task 2013

Sampo Pyysalo  Tomoko Ohta  Sophia Ananiadou
National Centre for Text Mining and School of Computer Science, University of Manchester
sampo.pyysalo@gmail.com, tomoko.ohta@manchester.ac.uk,
sophia.ananiadou@manchester.ac.uk

Abstract

We present the design, preparation, results and analysis of the Cancer Genetics (CG) event extraction task, a main task of the BioNLP Shared Task (ST) 2013. The CG task is an information extraction task targeting the recognition of events in text, represented as structured n-ary associations of given physical entities. In addition to addressing the cancer domain, the CG task is differentiated from previous event extraction tasks in the BioNLP ST series in addressing a wide range of pathological processes and multiple levels of biological organization, ranging from the molecular through the cellular and organ levels up to whole organisms. Final test set submissions were accepted from six teams. The highest-performing system achieved an F-score of 55.4%. This level of performance is broadly comparable with the state of the art for established molecular-level extraction tasks, demonstrating that event extraction resources and methods generalize well to higher levels of biological organization and pathological processes as well as physical entities and processes at higher levels of biological organization, including e.g. mutation, cell proliferation, apoptosis, blood vessel development, and metastasis.

The CG task aims to advance the development of such event extraction methods and the capacity for automatic analysis of texts on cancer biology.

1 Introduction

Despite decades of focused research efforts, cancer remains one of the leading causes of death worldwide. It is now well understood that cancer is a broad class of diseases with a complex genetic basis, involving changes in multiple molecular pathways (Hanahan and Weinberg, 2000; Haber et al., 2011). The scientific literature on cancer is enormous, and our understanding of cancer is developing rapidly: a query of the PubMed literature database for cancer returns 2.7 million scientific article citations, with 140,000 citations from 2012. To build and maintain comprehensive, up-to-date knowledge bases on cancer genetics, automatic support for managing the literature is thus required.

The BioNLP Shared Task (ST) series has been instrumental in encouraging the development of methods and resources for the automatic extraction of bio-processes from text, but efforts within this framework have been almost exclusively focused on normal physiological processes and on molecular-level entities and events (Kim et al., 2011a; Kim et al., 2011b). To be relevant to cancer biology, event extraction technology must be generalized to be able to address also pathological processes as well as physical entities and processes at higher levels of biological organization, including e.g. mutation, cell proliferation, apoptosis, blood vessel development, and metastasis. The CG task aims to advance the development of such event extraction methods and the capacity for automatic analysis of texts on cancer biology.

The CG task introduces a novel corpus covering multiple subdomains of cancer biology, based in part on a previously introduced angiogenesis subdomain resource (Pyysalo et al., 2012a). To extend event extraction to upper levels of biological organization and pathological processes, the task defines a set of 18 entity and 40 event types based on domain ontologies such as the Common Anatomy Reference Ontology and Gene Ontology, more than doubling the number of event and event types from those considered in previous BioNLP ST extraction tasks.

This paper presents the design of the CG task, introduces the groups and systems taking part in the task, and presents evaluation results and analysis.
2 Task definition

The CG task goal is the automatic extraction of events (Ananiadou et al., 2010) from text. The applied representation and task setting extend on those first established in the BioNLP ST 2009 (Kim et al., 2011a). Each event has a type such as GROWTH or METASTASIS and is associated with a specific span of characters expressing the event, termed the event trigger. Events can take any number of arguments, each of which is identified as participating in the event in a specific role (e.g. Theme or Cause). Event arguments may be either (physical) entities or other events, allowing complex event structures that capture e.g. one event causing or preventing another. Finally, events may be marked by flags identifying extra-propositional aspects such as occurrence in a speculative or negative context. Examples of CG task extraction targets are shown in Figure 1.

The following sections present the categories of annotation and the specific annotated types involved in the CG task: entities, relations, events, and event modifications. To focus efforts on novel challenges, the CG task follows the general convention of the BioNLP ST series of only requiring participants to extract events and their modifications. For other categories of annotation, correct (gold standard) annotations are provided also for test data.

2.1 Entities

The entity types defined in the CG task are shown in Table 1. The molecular level entity types largely match the scope of types such as PROTEIN and CHEMICAL included in previous ST tasks (Kim et al., 2012; Pyysalo et al., 2012b). However, the CG types are more fine grained, and the types PROTEIN DOMAIN OR REGION and DNA DOMAIN OR REGION are used in favor of the non-specific type ENTITY, applied in a number of previous tasks for additional event arguments (see Section 2.3). The definitions of the anatomical entity types are drawn primarily from the Common Anatomy Reference Ontology (Haendel et al., 2008), a small, species-independent upper-level ontology based on the Foundational Model of Anatomy (Rosse and Mejino Jr, 2003). We refer to Ohta et al. (2012) for more detailed discussion of the anatomical entity type definitions.

2.2 Relations

The CG task does not target the extraction of any standalone relations. However, following the model of past BioNLP ST tasks, the CG corpus is annotated by Equiv (equivalence) relations, symmetric, transitive relations that identify two entity mentions as referring to the same entity (Figure 2). These relations primarily mark local aliases and are applied only in evaluation. When determining whether a predicted event matches a gold event,
Table 2: Event types and their arguments. Nesting corresponds to ontological structure (is-a/part-of). The affixes ?, *, and + denote zero or one, zero or more, and one or more, respectively. GGP abbreviates for GENE OR GENE PRODUCT. For brevity, additional argument types are not shown in table: Loc arguments take an anatomical entity type, and Site PROTEIN/DNA DOMAIN OR REGION.

differences in references to equivalent entities are ignored, so that e.g. an event referring to CML as its Theme instead of chronic myeloid leukemia would be considered to match the event shown in Figure 2.

2.3 Events

Table 2 summarizes the event types defined in the CG task. As in most previous BioNLP ST task settings, the event types are defined primarily with reference to the Gene Ontology (GO) (Ashburner et al., 2000). However, GO explicitly excludes from its scope pathological processes, which are critically important to the CG task. To capture pathological processes, we systematically expand the scope GO-based event types to include also analogous processes involving pathological entities. For example, statements such as “cancer growth” are annotated with GROWTH events by analogy to processes such as “organ growth”. Second, we introduce a number of event types explicitly accounting for pathological processes with no analogous normal physiological process, such as METASTASIS. Finally, many important effects are discussed in the literature through statements involving experimenter action such as transfect and treat (Figure 1). To capture such statements, we introduce the general PLANNED PROCESS type, defined with reference to the Ontology for Biomedical Investigations (Brinkman et al., 2010).

The event argument roles largely match those
established in previous BioNLP ST tasks (Kim et al., 2012; Pyysalo et al., 2012b): Theme identifies the arguments undergoing the primary effects of the event, Cause those that are responsible for its occurrence, and Participant those whose precise role is not stated. Site is used to identify specific parts of Theme entities affected (e.g. phosphorylated residues) and the Loc roles entities where the event takes place (AtLoc) and start and end points of movement (FromLoc and ToLoc).

2.4 Event modifications

The CG task follows many previous BioNLP ST tasks in including the event modification types Negation and Speculation in its extraction targets. These modifications apply to events, marking them as explicitly negated and speculatively stated, respectively (Kim et al., 2011a).

2.5 Evaluation

The CG task evaluation follows the criteria originally defined in the BioNLP ST’09, requiring events extracted by systems to otherwise match gold standard events exactly, but allowing trigger spans to differ from gold spans by single words (approximate span matching) and not requiring matching of additional arguments (see Table 2) for events referred from other events (approximate recursive matching). These criteria are discussed in detail by Kim et al. (2011a).

3 Corpus

3.1 Document selection

The corpus texts are the titles and abstracts of publications from the PubMed literature database, selected on the basis of relevance to cancer genetics, specifically with respect to major subdomains relating to established hallmarks of cancer (Hana- han and Weinberg, 2000). Of the 600 documents forming the CG task corpus, 250 were previously released as part of the MLEE corpus (Pyysalo et al., 2012a) involving the angiogenesis subdomain. The remaining 350 were selected by iter-

| Domain          | Documents | Query terms                                           |
|-----------------|-----------|-------------------------------------------------------|
| Carcinogenesis  | 150       | cell transformation, neoplastic AND (proteins OR genes) |
| Metastasis      | 100       | neoplasm metastasis AND (proteins OR genes)           |
| Apoptosis       | 50        | apoptosis AND (proteins OR genes)                     |
| Glucose metabolism | 50   | (glucose/metabolism OR glycolysis) AND neoplasms       |

Table 3: Queries for document selection. All query terms were restricted to MeSH Term matches only (e.g. "apoptosis"[MeSH Terms])

Table 4: Corpus statistics

| Item             | Train | Devel | Test  | Total |
|------------------|-------|-------|-------|-------|
| Documents        | 300   | 100   | 200   | 600   |
| Words            | 66082 | 21732 | 42064 | 129878|
| Entities         | 11034 | 3665  | 6984  | 21683 |
| Relations        | 466   | 176   | 275   | 917   |
| Events           | 8803  | 2915  | 5530  | 17248 |
| Modifications    | 670   | 214   | 442   | 1326  |

Attractively formulating PubMed queries consisting of MeSH terms relevant to subdomains such as apoptosis and metastasis (Table 3). Following initial query formulation, random sets of abstracts were selected from each domain and manually examined to select a final set of documents that specifically discuss both the target process and its molecular foundations.

3.2 Annotation process

The corpus annotation was created using the BRAT annotation tool (Stenetorp et al., 2012) by a single PhD biologist with extensive experience in event annotation (Tomoko Ohta). For the entity annotation, we created preliminary annotation using the following automatic named entity and entity mention taggers: BANNER (Leaman and Gonzales, 2008) trained on the GENETAG corpus (Tanabe et al., 2005) for Gene OR Gene Product entities, Oscar4 (Jessop et al., 2011) for Simple Chemical and Amino Acid entities, NERSuite\(^1\) trained on the AnEM corpus (Ohta et al., 2012) for anatomical entities, and LINNAEUS (Gerner et al., 2010) for Organism mentions. Processing was performed on a custom pipeline originally developed for the BioNLP ST’11 (Stenetorp et al., 2011). Following preliminary automatic annotation, all entity annotations were manually revised to create the final entity annotation.

By contrast to entity annotation, no automatic preprocessing was applied for event annotation to avoid any possibility of bias introduced by initial application of automatic methods. The event annotation extended the guidelines and manual

\(^1\)http://nersuite.nlplab.org
Table 5: Participating teams and references to system descriptions. Abbreviations: BI=Bioinformatician, NLP=Natural Language Processing researcher, CS=Computer Scientist, LI=Linguist, ML=Machine Learning researcher.

| Team     | Institution                                                      | Members                                      | Members                                                                 |
|----------|------------------------------------------------------------------|----------------------------------------------|-------------------------------------------------------------------------|
| TEES-2.1 | University of Turku                                            | 1 BI                                         | (Björne and Salakoski, 2013)                                            |
| NaCTeM   | National Centre for Text Mining                                  | 1 NLP                                        | (Miwa and Ananiadou, 2013)                                              |
| NCBI     | National Center for Biotechnology Information                   | 3 BI                                         | (Liu et al., 2013)                                                      |
| RelAgent | RelAgent Private Ltd.                                           | 1 LI, 1 CS                                   | (Ramanan and Nathan, 2013)                                              |
| UET-NII  | University of Engineering and Technology, Vietnam and National Institute of Informatics, Japan | 6 CS                                         | (Tran et al., 2013)                                                     |
| ISI      | Indian Statistical Institute                                    | 2 ML, 2 NLP                                  |                                                                         |

Table 6: Summary of system architectures. Abbreviations: CoreNLP=Stanford CoreNLP, Porter=Porter stemmer, BLem=BioLemmatizer, Snowball=Snowball stemmer, McCCJ=McClosky-Charniak-Johnson parser, Charniak=Charniak parser, SD=Stanford Dependency conversion

annotation process introduced by Pyysalo et al. (2012a). Following the initial annotation, a number of revision passes were made to further improve the consistency of the annotation using a variety of automatically supported methods.\(^2\)

3.3 Corpus statistics

Table 4 summarizes the corpus statistics for the training, development and test sets, representing 50%, 17%, and 33% of the documents, respectively. The CG task corpus is the largest of the BioNLP ST 2013 corpora by most measures, including the number of annotated events.

4 Participation

Final results to the CG task were successfully submitted by six teams, from six different academic groups and one company, representing a broad range of expertise ranging from biology to machine learning, natural language processing, and linguistics (Table 5).

The characteristics of the participating systems are summarized in Table 6. There is an interesting spread of extraction approaches, with two systems applying SVM-based pipeline architectures shown successful in previous BioNLP ST events, one applying a joint pattern matching approach, one a rule-based approach, and two systems parsing-based approaches to event extraction. Together, these systems represent all broad classes of approaches applied to event extraction in previous BioNLP ST events. Three of the six systems addressed also the event modification (negation and speculation) extraction aspects of the task.

Although all systems perform syntactic analysis of input texts, there is a fair amount of variety in the applied parsers, which include the parser of Charniak and Johnson (2005) with the biomedical domain model of McClosky (2009) and the Stanford Dependency conversion (de Marneffe et al., 2006) – the choice in many systems in BioNLP ST'11 – as well as Enju (Miyao and Tsujii, 2008), GDep (Sagae and Tsujii, 2007), Stanford CoreNLP\(^3\), and a custom parser by RelAgent (Ramanan and Nathan, 2013). Simple stemming algorithms such as that of Porter (1980) remain popular for word-level processing, with just the NCBI system using a dedicated biomedical domain lemmatizer (Liu et al., 2012).

The task setting explicitly allows the use of any external resources, including other corpora, and previously released event resources contain significant numbers of annotations that are relevant

\(^2\)There was no opportunity to train a second annotator in order to evaluate IAA specifically for the new CG corpus annotation. However, based on our previous evaluation using the same protocol (Pyysalo et al., 2012a), we expect the consistency of the final annotation to fall in the 70-80% F-score range (primary task evaluation criteria).

\(^3\)http://nlp.stanford.edu/software/corenlp.shtml
Table 7: Primary evaluation results
to the molecular level events annotated in the CG
task. Nevertheless, only the TEES and NCBI
teams made use of corpora other than the task
data, both using the GE corpus (Kim et al., 2012)
and NCBI using also the EPI corpus (Pyysalo et
al., 2012b). In addition to corpora annotated for
events, lexical resources derived from such cor-
pora, containing trigger and hedge expressions,
were applied by three teams.

We refer to the descriptions presented by each
of the participating teams (see Table 5) for further
detail on the systems and their implementations.

5 Results

The primary evaluation results are summarized in
Table 7. The highest performance is achieved by
the established machine learning-based TEES sys-
tem, with an F-score of 55%. Previous versions
of the same system achieved the highest perform-
ance in the BioNLP ST’09 (52% F-score) and
in four out of eight tasks in BioNLP ST’11 (53%
F-score for the comparable GE task) (Björne and
Salakoski, 2011). The performance of the system
ranked second, EventMine (Miwa et al., 2012),
is likewise broadly comparable to the results for
the same system on the GE task considered in
BioNLP ST’09 and ‘11. The NCBI submission
also extends a system that participated in the
ST’11 GE task, then achieving a somewhat lower
F-score of 41.13% (Liu et al., 2011). By con-
trast, the RelAgent, UET-NII and ISI submissions
involve systems that were not previously applied
in BioNLP ST events. Thus, in each case where
system performance for previously proposed event
extraction tasks is known, the results indicate that
the systems generalize to CG task extraction tar-
gets without loss in performance.

These parallels with results for previously intro-
duced tasks involving molecular-level events are
interesting, in particular considering that the CG
task involves more than twice the number of en-
tity and event types included in previously con-
sidered BioNLP ST tasks. The results suggest
not only that event extraction methods generalize
well to higher levels of biological organization,
but also that overall performance is not primar-
ily limited by the number of targeted types. It is
also notable that the complexity of the task set-
ing does not exclude rule-based systems such as
that of RelAgent, which scores within 10% points
of the highest-ranking system. While the parser-
based systems of UET-NII and ISI perform be-
low others here, it should be noted that related ap-
proaches have achieved competitive performance
in previous BioNLP ST tasks (McClosky et al.,
2011), suggesting that further development could
lead to improvements for systems based on these
architectures. As is characteristic for event extrac-
tion systems in general, all systems show notably
higher precision than recall, with the performance
of the UET-NII and ISI systems in particular pri-
marily limited by low recall.

The F-score results are shown separately for
each event type in Table 8. As suggested by the
overall results, the novel categories of events in-
volving anatomical and pathological entities are
not particularly challenging for most systems,
with results roughly mirroring performance for
molecular level events; the best results by event
category are 77% F-score for anatomical, 68%
for pathological, and 73% for molecular. Of
the newly introduced CG event categories, only
planned processes involving intentional human in-
tervention appear to represent difficulties, with the
best-performing system for PLANNED PROCESS
reaching only 41% F-score. Two previously es-
tablished categories of events remain challenging:
general events – best 53% F-score – including
BINDING (often taking multiple arguments) and
LOCALIZATION (frequent additional arguments),
and regulation category events, which often form
complex event structures by involving events as ar-
guments. Event modifications, addressed by three
of the six participating teams, show comparatively
low levels of extraction performance, with a best
result of 40% F-score for NEGATION and 30%
for SPECULATION. However, as in previous tasks
(Kim et al., 2011a), this is in part due to the com-
pound nature of the problem: for an event modifi-
cation attribute to be extracted correctly, the event
that it attaches to must also be correct.

Further details on system performance and anal-
yses are available on the shared task home page.
6 Discussion and conclusions

We have presented the Cancer Genetics (CG) task, an information extraction task introduced as a main task of the BioNLP Shared Task (ST) 2013. The task is motivated by the needs of maintaining up-to-date knowledge bases of the enormous and fast-growing literature on cancer genetics, and extends previously proposed BioNLP ST tasks in several aspects, including the inclusion of entities and events at levels of biological organization above the molecular and the explicit inclusion of pathological and planned processes among extraction targets. To address these extraction goals, we introduced a new corpus covering various subdomains of cancer genetics, annotated for 18 entity and 40 event types and marking over 17,000 manually annotated events in 600 publication abstracts.

Final submissions to the CG task were received from six groups, who applied a variety of approaches including machine learning-based clas-
sifier pipelines, parsing-based approaches, and pattern- and rule-based systems. The best-performing system achieved an F-score of 55.4%, a level of performance comparable to the state of the art in established molecular level event extraction tasks. The results indicate that event extraction methods generalize well across the novel aspects introduced in the CG task and that event extraction is applicable to the automatic processing of the cancer literature.

Following convention in the BioNLP Shared Task series, the Cancer Genetics task will continue as an open challenge available to all interested participants. The CG task corpus, supporting resources and evaluation tools are available from http://2013.bionlp-st.org/.

Acknowledgments

We wish to thank the BioNLP ST 2013 CG task participants and supporting resource providers for their invaluable contributions to making this task a success. This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) [BB/G53025X/1].

References

Sophia Ananiadou, Sampo Pyysalo, Jun’ichi Tsujii, and Douglas B. Kell. 2010. Event extraction for systems biology by text mining the literature. Trends in Biotechnology, 28(7):381–390.

Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, et al. 2000. Gene ontology: tool for the unification of biology. Nature genetics, 25(1):25–29.

Jari Björne and Tapio Salakoski. 2011. Generalizing biomedical event extraction. In Proceedings of BioNLP’11, pages 183–191.

Jari Björne and Tapio Salakoski. 2013. TEES 2.1: Automated annotation scheme learning in the bioNLP 2013 shared task. In Proceedings of BioNLP Shared Task 2013.

Ryan R Brinkman, Mélanie Courtot, Dirk Derom, Jennifer M Fostel, et al. 2010. Modeling biomedical experimental processes with OBI. J Biomed Semantics, 1(Suppl 1):S7.

Eugene Charniak and Mark Johnson. 2005. Coarse-to-Fine n-Best Parsing and MaxEnt Discriminative Reranking. In Proceedings of ACL’05, pages 173–180.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D Manning. 2006. Generating typed dependency parses from phrase structure parses. In Proceedings of LREC, volume 6, pages 449–454.

Martin Gerner, Goran Nenadic, and Casey M Bergman. 2010. Linnaeus: a species name identification system for biomedical literature. BMC bioinformatics, 11(1):85.

Daniel A Haber, Nathanael S Gray, and Jose Baselga. 2011. The evolving war on cancer. Cell, 145(1):19–24.

Melissa A Haendel, Fabian Neuhaus, David Osumi-Sutherland, Paula M Mabee, Jos LV Mejino Jr, Chris J Mungall, and Barry Smith. 2008. CARO—the common anatomy reference ontology. pages 327–349.

Douglas Hanahan and Robert A Weinberg. 2000. The hallmarks of cancer. Cell, 100(1):57–70.

David M Jessop, Sam E Adams, Egon L Willighagen, Lezan Hawizy, and Peter Murray-Rust. 2011. Oscar4: a flexible architecture for chemical text-mining. Journal of Cheminformatics, 3(1):1–12.

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun’ichi Tsujii. 2011a. Extracting bio-molecular events from literature - the BioNLP’09 shared task. Computational Intelligence, 27(4):513–540.

Jin-Dong Kim, Sampo Pyysalo, Tomoko Ohta, Robert Bossy, and Jun’ichi Tsujii. 2011b. Overview of BioNLP Shared Task 2011. In Proceedings of BioNLP’11.

Jin-Dong Kim, Ngan Nguyen, Yue Wang, Jun’ichi Tsujii, Toshihisa Takagi, and Akinori Yonezawa. 2012. The genia event and protein coreference tasks of the bionlp shared task 2011. BMC bioinformatics, 13(Suppl 11):S1.

Robert Leaman and Graciela Gonzalez. 2008. BANNER: an executable survey of advances in biomedical named entity recognition. Proceedings of the Pacific Symposium on Biocomputing (PSB’08), pages 652–663.

Haibin Liu, Ravi Kumar Komandur, and Karin Verspoor. 2011. From graphs to events: A subgraph matching approach for information extraction from biomedical text. In Proceedings of BioNLP’11, pages 164–172.

Haibin Liu, Tom Christiansen, William A Baumgartner Jr, Karin Verspoor, et al. 2012. Biolemmatizer: a lemmatization tool for morphological processing of biomedical text. Journal of biomedical semantics, 3(3).

Haibin Liu, Karin Verspoor, Donald Comeau, Andrew MacKinlay, and W John Wilbur. 2013. Generalizing an approximate subgraph matching-based system to extract events in molecular biology and cancer genetics. In Proceedings of BioNLP Shared Task 2013 Workshop.
David McClosky, Mihai Surdeanu, and Christopher D Manning. 2011. Event extraction as dependency parsing for bioNLP 2011. In Proceedings BioNLP’11, pages 41–45.

David McClosky. 2009. *Any Domain Parsing: Automatic Domain Adaptation for Natural Language Parsing*. Ph.D. thesis, Department of Computer Science, Brown University.

Makoto Miwa and Sophia Ananiadou. 2013. NaCTeM EventMine for bioNLP 2013 CG and PC tasks. In Proceedings of BioNLP Shared Task 2013 Workshop.

Makoto Miwa, Paul Thompson, and Sophia Ananiadou. 2012. Boosting automatic event extraction from the literature using domain adaptation and coreference resolution. *Bioinformatics*, 28(13):1759–1765.

Yusuke Miyao and Jun’ichi Tsujii. 2008. Feature forest models for probabilistic hpsg parsing. *Computational Linguistics*, 34(1):35–80.

Tomoko Ohta, Sampo Pyysalo, Jun’ichi Tsujii, and Sophia Ananiadou. 2012. Open-domain anatomical entity mention detection. In Proceedings of DSSD 2012, pages 27–36.

Martin F Porter. 1980. An algorithm for suffix stripping. *Program: electronic library and information systems*, 14(3):130–137.

Sampo Pyysalo, Tomoko Ohta, Makoto Miwa, Han-Chrel Cho, Jun’ichi Tsujii, and Sophia Ananiadou. 2012a. Event extraction across multiple levels of biological organization. *Bioinformatics*, 28(18):i575–i581.

Sampo Pyysalo, Tomoko Ohta, Rafal Rak, Dan Sullivan, Chunhong Mao, Chunxia Wang, Bruno Sobral, Jun’ichi Tsujii, and Sophia Ananiadou. 2012b. Overview of the ID, EPI and REL tasks of BioNLP Shared Task 2011. *BMC bioinformatics*, 13(Suppl 11):S2.

SV Ramanan and P. Senthil Nathan. 2013. Performance and limitations of the linguistically motivated cocoa/peaberry system in a broad biological domain. In Proceedings of BioNLP Shared Task 2013 Workshop.

Cornelius Rosse and José LV Mejino Jr. 2003. A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of biomedical informatics*, 36(6):478–500.

Kenny Sagae and Jun’ichi Tsujii. 2007. Dependency parsing and domain adaptation with lr models and parser ensembles. In Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007, pages 1044–1050.