BioNLP Shared Task 2011: Supporting Resources

Pontus Stenetorp[*†] **Goran Topić**[*] **Sampo Pyysalo**[*]
Tomoko Ohta[*] **Jin-Dong Kim**[‡] **and Jun'ichi Tsujii**[$]
[*]Tsujii Laboratory, Department of Computer Science, University of Tokyo, Tokyo, Japan
[†]Aizawa Laboratory, Department of Computer Science, University of Tokyo, Tokyo, Japan
[‡] Database Center for Life Science,
Research Organization of Information and Systems, Tokyo, Japan
[$]Microsoft Research Asia, Beijing, People's Republic of China
{pontus,goran,smp,okap}@is.s.u-tokyo.ac.jp
jdkim@dbcls.rois.ac.jp
jtsujii@microsoft.com

Abstract

This paper describes the supporting resources provided for the BioNLP Shared Task 2011. These resources were constructed with the goal to alleviate some of the burden of system development from the participants and allow them to focus on the novel aspects of constructing their event extraction systems. With the availability of these resources we also seek to enable the evaluation of the applicability of specific tools and representations towards improving the performance of event extraction systems. Additionally we supplied evaluation software and services and constructed a visualisation tool, *stav*, which visualises event extraction results and annotations. These resources helped the participants make sure that their final submissions and research efforts were on track during the development stages and evaluate their progress throughout the duration of the shared task. The visualisation software was also employed to show the differences between the gold annotations and those of the submitted results, allowing the participants to better understand the performance of their system. The resources, evaluation tools and visualisation tool are provided freely for research purposes and can be found at http://sites.google.com/site/bionlpst/

1 Introduction

For the BioNLP'09 Shared Task (Kim et al., 2009), the first in the ongoing series, the organisers provided the participants with automatically generated syntactic analyses for the sentences from the annotated data. For evaluation purposes, tools were made publicly available as both distributed software and online services. These resources were well received. A majority of the participants made use of one or more of the syntactic analyses, which have remained available after the shared task ended and have been employed in at least two independent efforts studying the contribution of different tools and forms of syntactic representation to the domain of information extraction (Miwa et al., 2010; Buyko and Hahn, 2010). The evaluation software for the BioNLP'09 Shared Task has also been widely adopted in subsequent studies (Miwa et al., 2010; Poon and Vanderwende, 2010; Björne et al., 2010).

The reception and research contribution from providing these resources encouraged us to continue providing similar resources for the BioNLP Shared Task 2011 (Kim et al., 2011a). Along with the parses we also encouraged the participants and external groups to process the data with any NLP (Natural Language Processing) tools of their choice and make the results available to the participants.

We provided continuous verification and evaluation of the participating systems using a suite of in-house evaluation tools. Lastly, we provided a tool for visualising the annotated data to enable the participants to better grasp the results of their experiments and to help gain a deeper understanding of the underlying concepts and the annotated data. This paper presents these supporting resources.

2 Data

This section introduces the data resources provided by the organisers, participants and external groups for the shared task.

Proceedings of BioNLP Shared Task 2011 Workshop, pages 112–120,
Portland, Oregon, USA, 24 June, 2011.

Task	Provider	Tool
CO	University of Utah	Reconcile
CO	University of Zürich	UZCRS
CO	University of Turku	TEES
REL	University of Turku	TEES

Table 1: Supporting task analyses provided, TEES is the Turku Event Extraction System and UZCRS is the University of Zürich Coreference Resolution System

2.1 Supporting task analyses

The shared task included three Supporting Tasks: Coreference (CO) (Nguyen et al., 2011), Entity relations (REL) (Pyysalo et al., 2011b) and Gene renaming (REN) (Jourde et al., 2011). In the shared task schedule, the supporting tasks were carried out before the main tasks (Kim et al., 2011b; Pyysalo et al., 2011a; Ohta et al., 2011; Bossy et al., 2011) in order to allow participants to make use of analyses from the systems participating in the Supporting Tasks for their main task event extraction systems.

Error analysis of BioNLP'09 shared task submissions indicated that coreference was the most frequent feature of events that could not be correctly extracted by any participating system. Further, events involving statements of non-trivial relations between participating entities were a frequent cause of extraction errors. Thus, the CO and REL tasks were explicitly designed to support parts of the main event extraction tasks where it had been suggested that they could improve the system performance.

Table 1 shows the supporting task analyses provided to the participants. For the main tasks, we are currently aware of one group (Emadzadeh et al., 2011) that made use of the REL task analyses in their system. However, while a number of systems involved coreference resolution in some form, we are not aware of any teams using the CO task analyses specifically, perhaps due in part to the tight schedule and the somewhat limited results of the CO task. These data will remain available to allow future research into the benefits of these resources for event extraction.

2.2 Syntactic analyses

For syntactic analyses we provided parses for all the task data in various formats from a wide range of parsers (see Table 2). With the exception of the Pro3Gres[1] parser (Schneider et al., 2007), the parsers were set up and run by the task organisers. The emphasis was put on availability for research purposes and variety of parsing models and frameworks to allow evaluation of their applicability for different tasks.

In part following up on the results of Miwa et al. (2010) and Buyko and Hahn (2010) regarding the impact on performance of event extraction systems depending on the dependency parse representation, we aimed to provide several dependency parse formats. Stanford Dependencies (SD) and Collapsed Stanford Dependencies (SDC), as described by de Marneffe et al. (2006), were generated by converting Penn Treebank (PTB)-style (Marcus et al., 1993) output using the Stanford CoreNLP Tools[2] into the two dependency formats. We also provided Conference on Computational Natural Language Learning style dependency parses (CoNLL-X) (Buchholz and Marsi, 2006) which were also converted from PTB-style output, but for this we used the conversion tool[3] from Johansson and Nugues (2007). While this conversion tool was not designed with converting the output from statistical parsers in mind (but rather to convert between treebanks), it has previously been applied successfully for this task (Miyao et al., 2008; Miwa et al., 2010).

The text from all documents provided were split into sentences using the Genia Sentence Splitter[4] (Sætre et al., 2007) and then postprocessed using a set of heuristics to correct frequently occurring errors. The sentences were then tokenised using a tokenisation script created by the organisers intended to replicate the tokenisation of the Genia Tree Bank (GTB) (Tateisi et al., 2005). This tokenised and sentence-split data was then used as input for all parsers.

We used two deep parsers that provide phrase structure analysis enriched with deep sentence struc-

[1] https://files.ifi.uzh.ch/cl/gschneid/parser/

[2] http://nlp.stanford.edu/software/corenlp.shtml

[3] http://nlp.cs.lth.se/software/treebank_converter/

[4] http://www-tsujii.is.s.u-tokyo.ac.jp/~y-matsu/geniass/

Name	Format(s)	Model	Availability	BioNLP'09
Berkeley	PTB, SD, SDC, CoNLL-X	News	Binary, Source	No
C&C	CCG, SD	Biomedical	Binary, Source	Yes
Enju	HPSG, PTB, SD, SDC, CoNLL-X	Biomedical	Binary	No
GDep	CoNLL-X	Biomedical	Binary, Source	Yes
McCCJ	PTB, SD, SDC, CoNLL-X	Biomedical	Source	Yes
Pro3Gres	Pro3Gres	Combination	–	No
Stanford	PTB, SD, SDC, CoNLL-X	Combination	Binary, Source	Yes

Table 2: Parsers, the formats for which their output was provided and which type of model that was used. The availability column signifies public availability (without making an explicit request) for research purposes

tures, for example predicate-argument structure for Head-Driven Phrase Structure Grammar (HPSG). First we used the C&C Combinatory Categorial Grammar (CCG) parser[5] (C&C) by Clark and Curran (2004) using the biomedical model described in Rimell and Clark (2009) which was trained on GTB. Unlike all other parsers for which we supplied SD and SDC dependency parses, the C&C output was converted from its native format using a separate conversion script provided by the C&C authors. Regrettably we were unable to provide CoNLL-X format output for this parser due to the lack of PTB-style output. The other deep parser used was the HPSG parser Enju[6] by Miyao and Tsujii (2008), also trained on GTB.

We also applied the frequently adopted Stanford Parser[7] (Klein and Manning, 2003) using a mixed model which includes data from the biomedical domain, and the Charniak Johnson re-ranking parser[8] (Charniak and Johnson, 2005) using the self-trained biomedical model from McClosky (2009) (McCCJ).

For the BioNLP'09 shared task it was observed that the Bikel parser[9] (Bikel, 2004), which used a non-biomedical model and can be argued that it uses the somewhat dated Collins' parsing model (Collins, 1996), did not contribute towards event extraction performance as strongly as other parses supplied for the same data. We therefore wanted to supply a parser that can compete with the ones above in a domain which is different from the biomedical domain to see whether conclusions could be drawn as to the importance of using a biomedical model. For this we used the Berkeley parser[10] (Petrov et al., 2006). Lastly we used a native dependency parser, the GENIA Dependency parser (GDep) by Sagae and Tsujii (2007).

At least one team (Choudhury et al., 2011) performed experiments on some of the provided lexical analyses and among the 14 submissions for the EPI and ID tasks, 13 submissions utilised tools for which resources were provided by the organisers of the shared task. We intend to follow up on whether or not the majority of the teams ran the tools themselves or used the provided analyses.

2.3 Other analyses

The call for analyses was open to all interested parties and all forms of analysis. In addition to the Supporting Task analyses (CO and REL) and syntactic analyses provided by various groups, the University of Antwerp CLiPS center (Morante et al., 2010) responded to the call providing negation/speculation analyses in the BioScope corpus format (Szarvas et al., 2008).

Although this resource was not utilised by the participants for the main task, possibly due to a lack of time, it is our hope that by keeping the data available it can lead to further development of the participating systems and analysis of BioScope and BioNLP ST-style hedging annotations.

3 Tools

This section presents the tools produced by the organisers for the purpose of the shared task.

[5] http://svn.ask.it.usyd.edu.au/trac/candc/

[6] http://www-tsujii.is.s.u-tokyo.ac.jp/enju/

[7] http://nlp.stanford.edu/software/lex-parser.shtml

[8] ftp://ftp.cs.brown.edu/pub/nlparser/

[9] http://www.cis.upenn.edu/~dbikel/software.html

[10] http://code.google.com/p/berkeleyparser/

```
1  10411007-E1       Regulation    <Exp>regulate[26-34] <Theme>TNF-alpha[79-88]    ↵
       ↳<Excerpt>[regulate] an enhancer activity in the third intron of [TNF-alpha]
2  10411007-E2       Gene_expression       <Exp>activity[282-290] <Theme>TNF-alpha[252-261]       ↵
       ↳<Excerpt>[TNF-alpha] gene displayed weak [activity]
3  10411007-E3       +Regulation   <Exp>when[291-295] <Theme>E2    <Excerpt>[when]
```

Figure 1: Text output from the BioNLP'09 Shared Event Viewer with line numbering and newline markings

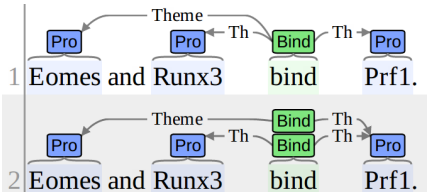


Figure 2: An illustration of collective (sentence 1) and distributive reading (sentence 2). "Theme" is abbreviated as "Th" and "Protein" as "Pro" when there is a lack of space

3.1 Visualisation

The annotation data in the format specified by the shared task is not intended to be human-readable – yet researchers need to be able to visualise the data in order to understand the results of their experiments. However, there is a scarcity of tools that can be used for this purpose. There are three available for event annotations in the BioNLP ST format that we are aware of.

One is the BioNLP'09 Shared Task Event Viewer[11], a simple text-based annotation viewer: it aggregates data from the annotations, and outputs it in a format (Figure 1) that is meant to be further processed by a utility such as `grep`.

Another is What's Wrong with My NLP[12], which visualises relation annotations (see Figure 3a) – but is unable to display some of the information contained in the Shared Task data. Notably, the distributive and collective readings of an event are not distinguished (Figure 2). It also displays all annotations on a single line, which makes reading and analysing longer sentences, let alone whole documents, somewhat difficult.

The last one is U-Compare[13] (Kano et al., 2009), which is a comprehensive suite of tools designed for managing NLP workflows, integrating many available services. However, the annotation visualisation component, illustrated in Figure 3b, is not optimised for displaying complex event structures. Each annotation is marked by underlining its text segment using a different colour per annotation type, and a role in an event is represented by a similarly coloured arc between the related underlined text segments. The implementation leaves some things to be desired: there is no detailed information added in the display unless the user explicitly requests it, and then it is displayed in a separate panel, away from the text it annotates. The text spacing makes no allowance for the annotations, with opaque lines crossing over it, with the effect of making both the annotations and the text hard to read if the annotations are above a certain degree of complexity.

As a result of the difficulties of these existing tools, in order to extract a piece of annotated text and rework it into a graph that could be embedded into a publication, users usually read off the annotations, then create a graph from scratch using vector drawing or image editing software.

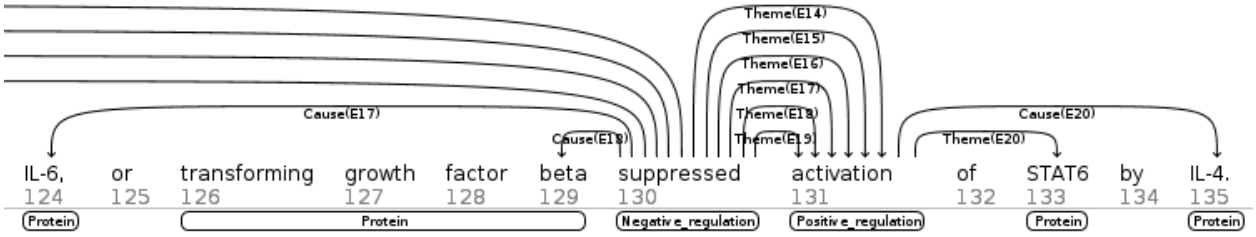
To address these issues, we created a visualisation tool named *stav* (*stav* Text Annotation Visualizer), that can read the data formatted according to the Shared Task specification and aims to present it to the user in a form that can be grasped at a glance. Events and entities are annotated immediately above the text, and the roles within an event by labelled arcs between them (Figure 3c). In a very complex graph, users can highlight the object or association of interest to follow it even more easily. Special features of annotations, such as negation or speculation, are shown by unique visual cues, and more in-depth, technical information that is usually not required can be requested by floating the mouse cursor over the annotation (as seen in Figure 5).

We took care to minimise arc crossovers, and to

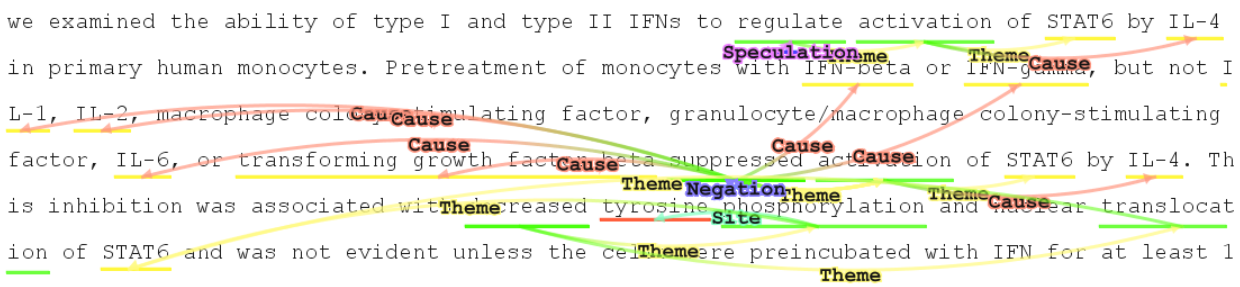
[11] http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/SharedTask/downloads.shtml

[12] http://code.google.com/p/whatswrong/

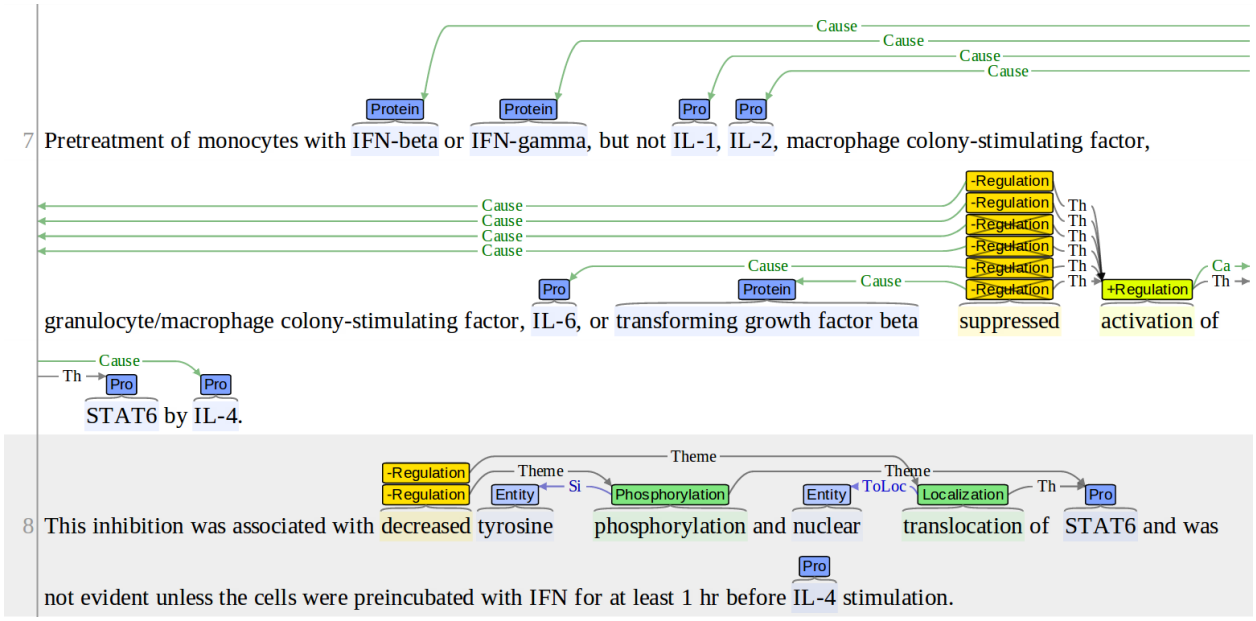
[13] http://u-compare.org/bionlp2009.html



(a) Visualisation using What's Wrong with My NLP



(b) Visualisation using U-Compare



(c) Visualisation using *stav*

Figure 3: Different visualisations of complex textual annotations of Dickensheets et al. (1999)

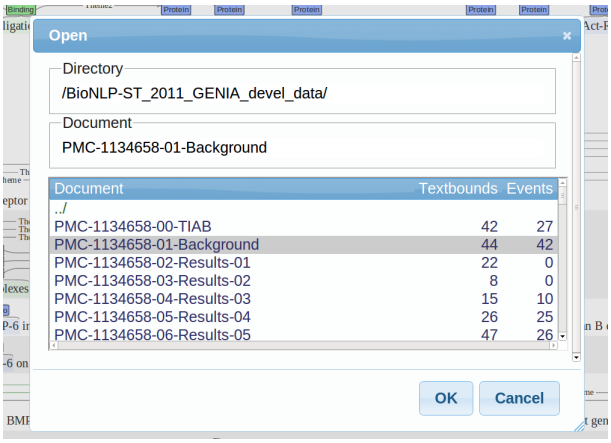


Figure 4: A screenshot of the *stav* file-browser

keep them away from the text itself, in order to maintain text readability. The text is spaced to accommodate the annotations between the rows. While this does end up using more screen real-estate, it keeps the text legible, and annotations adjacent to the text. The text is broken up into lines, and each sentence is also forced into a new line, and given a numerical identifier. The effect of this is that the text is laid out vertically, like an article would be, but with large spacing to accomodate the annotations. The arcs are similarly continued on successive lines, and can easily be traced – even in case of them spanning multiple lines, by the use of mouseover highlighting. To preserve the distributionality information of the annotation, any event annotations are duplicated for each event, as demonstrated in the example in Figure 2.

stav is not limited to the Shared Task datasets with appropriate configuration settings, it could also visualise other kinds of relational annotations such as: frame structures (Fillmore, 1976) and dependency parses (de Marneffe et al., 2006).

To achieve our objectives above, we use the Dynamic Scalable Vector Graphics (SVG) functionality (i.e. SVG manipulated by JavaScript) provided by most modern browsers to render the WYSIWYG (What You See Is What You Get) representation of the annotated document. An added benefit from this technique is that the installation process, if any, is very simple: although not all browsers are currently supported, the two that we specifically tested against are Safari[14] and Google Chrome[15]; the former comes preinstalled with the Mac OS X operating system, while the latter can be installed even by relatively non-technical users. The design is kept modular using a dispatcher pattern, in order to allow the inclusion of the visualiser tool into other JavaScript-based projects. The client-server architecture also allows centralisation of data, so that every user can inspect an uploaded dataset without the hassle of downloading and importing into a desktop application, simply by opening an URL which can uniquely identify a document, or even a single annotation. A screenshot of the *stav* file browser can be seen in Figure 4.

3.2 Evaluation Tools

The tasks of BioNLP-ST 2011 exhibit very high complexity, including multiple non-trivial subproblems that are partially, but not entirely, independent of each other. With such tasks, the evaluation of participating systems itself becomes a major challenge. Clearly defined evaluation criteria and their precise implementation is critical not only for the comparison of submissions, but also to help participants follow the status of their development and to identify the specific strengths and weaknesses of their approach.

A further challenge arising from the complexity of the tasks is the need to process the relatively intricate format in which annotations are represented, which in turn carries a risk of errors in submissions. To reduce the risk of submissions being rejected or the evaluation showing poor results due to formatting errors, tools for checking the validity of the file format and annotation semantics are indispensable.

For these reasons, we placed emphasis in the organisation of the BioNLP-ST'11 on making tools for format checking, validation and evaluation available to the participants already during the early stages of system development. The tools were made available in two ways: as downloads, and as online services. With downloaded tools, participants can perform format checking and evaluation at any time without online access, allowing more efficient optimisation processes. Each task in BioNLP-ST also

[14] http://www.apple.com/safari

[15] http://www.google.com/chrome

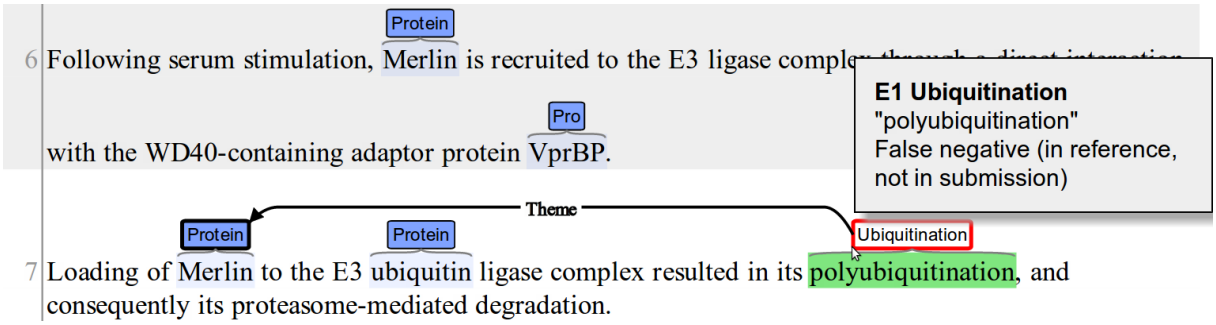


Figure 5: An example of a false negative illustrated by the evaluation tools in co-ordination with *stav*

maintained an online evaluation tool for the development set during the development period. The online evaluation is intended to provide an identical interface and criteria for submitted data as the final online submission system, allowing participants to be better prepared for the final submission. With online evaluation, the organisers could also monitor submissions to ensure that there were no problems in, for example, the evaluation software implementations.

The system logs of online evaluation systems show that the majority of the participants submitted at least one package with formatting errors, confirming the importance of tools for format checking. Further, most of the participants made use of the online development set evaluation at least once before their final submission.

To enhance the evaluation tools we drew upon the *stav* visualiser to provide a view of the submitted results. This was done by comparing the submitted results and the gold data to produce a visualisation where errors are highlighted, as illustrated in Figure 5. This experimental feature was available for the EPI and ID tasks and we believe that by doing so it enables participants to better understand the performance of their system and work on remedies for current shortcomings.

4 Discussion and Conclusions

Among the teams participating in the EPI and ID tasks, a great majority utilised tools for which resources were made available by the organisers. We hope that the continued availability of the parses will encourage further investigation into the applicability of these and similar tools and representations.

As for the analysis of the supporting analyses provided by external groups and the participants, we are so far aware of only limited use of these resources among the participants, but the resources will remain available and we are looking forward to see future work using them.

To enable reproducibility of our resources, we provide a publicly accessible repository containing the automated procedure and our processing scripts used to produce the released data. This repository also contains detailed instructions on the options and versions used for each parser and, if the software license permits it, includes the source code or binary that was used to produce the processed data. For the cases where the license restricts redistribution, instructions and links are provided on how to obtain the same version that was used. We propose that using a multitude of parses and formats can benefit not just the task of event extraction but other NLP tasks as well.

We have also made our evaluation tools and visualisation tool *stav* available along with instructions on how to run it and use it in coordination with the shared task resources. The responses from the participants in relation to the visualisation tool were very positive, and we see this as encouragement to advance the application of visualisation as a way to better reach a wider understanding and unification of the concept of events for biomedical event extraction.

All of the resources described in this paper are available at http://sites.google.com/site/bionlpst/.

Acknowledgements

We would like to thank Jari Björne of the University of Turku BioNLP group; Gerold Schneider, Fabio Rinaldi, Simon Clematide and Don Tuggener of the Univerity of Zurich Computational Linguistics group; Roser Morante of University of Antwerp CLiPS center; and Youngjun Kim of the University of Utah Natural Language Processing Research Group for their generosity with their time and expertise in providing us with supporting analyses.

This work was supported by Grant-in-Aid for Specially Promoted Research (MEXT, Japan) and the Royal Swedish Academy of Sciences.

References

Daniel M. Bikel. 2004. Intricacies of Collins' Parsing Model. *Computational Linguistics*, 30(4):479–511.

J. Björne, F. Ginter, S. Pyysalo, J. Tsujii, and T. Salakoski. 2010. Complex event extraction at PubMed scale. *Bioinformatics*, 26(12):i382.

Robert Bossy, Julien Jourde, Philippe Bessières, Marteen van de Guchte, and Claire Nédellec. 2011. BioNLP Shared Task 2011 - Bacteria Biotope. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

S. Buchholz and E. Marsi. 2006. CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, pages 149–164. Association for Computational Linguistics.

E. Buyko and U. Hahn. 2010. Evaluating the impact of alternative dependency graph encodings on solving event extraction tasks. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 982–992. Association for Computational Linguistics.

Eugene Charniak and Mark Johnson. 2005. Coarse-to-Fine n-Best Parsing and MaxEnt Discriminative Reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 173–180.

Pallavi Choudhury, Michael Gamon, Chris Quirk, and Lucy Vanderwende. 2011. MSR-NLP entry in BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

S. Clark and J.R. Curran. 2004. Parsing the WSJ using CCG and log-linear models. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 103. Association for Computational Linguistics.

Michael John Collins. 1996. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pages 184–191, Santa Cruz, California, USA, June. Association for Computational Linguistics.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating Typed Dependency Parses from Phrase Structure Parses. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, pages 449–454.

H.L. Dickensheets, C. Venkataraman, U. Schindler, and R.P. Donnelly. 1999. Interferons inhibit activation of STAT6 by interleukin 4 in human monocytes by inducing SOCS-1 gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 96(19):10800.

Ehsan Emadzadeh, Azadeh Nikfarjam, and Graciela Gonzalez. 2011. A generalizable and efficient machine learning approach for biological event extraction from text. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

Charles J. Fillmore. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32.

R. Johansson and P. Nugues. 2007. Extended constituent-to-dependency conversion for English. In *Proc. of the 16th Nordic Conference on Computational Linguistics (NODALIDA)*.

Julien Jourde, Alain-Pierre Manine, Philippe Veber, Karën Fort, Robert Bossy, Erick Alphonse, and Philippe Bessières. 2011. BioNLP Shared Task 2011 - Bacteria Gene Interactions and Renaming. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

Yoshinobu Kano, William Baumgartner, Luke McCrohon, Sophia Ananiadou, Kevin Cohen, Larry Hunter, and Jun'ichi Tsujii. 2009. U-Compare: share and compare text mining tools with UIMA. *Bioinformatics*, 25(15):1997–1998, May.

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP'09 Shared Task on Event Extraction. In *Proceedings of Natural Language Processing in Biomedicine (BioNLP) NAACL 2009 Workshop*, pages 1–9.

Jin-Dong Kim, Sampo Pyysalo, Tomoko Ohta, Robert Bossy, and Jun'ichi Tsujii. 2011a. Overview of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

Jin-Dong Kim, Yue Wang, Toshihisa Takagi, and Akinori Yonezawa. 2011b. Overview of the Genia Event task in BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

D. Klein and C.D. Manning. 2003. Fast exact inference with a factored model for natural language parsing. *Advances in neural information processing systems*, pages 3–10.

M.P Marcus, B. Santorini, and M.A Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Tree Bank. *Computational Linguistics*, pages 313–318.

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

M. Miwa, S. Pyysalo, T. Hara, and J. Tsujii. 2010. Evaluating Dependency Representation for Event Extraction. In *In the 23rd International Conference on Computational Linguistics (COLING 2010)*, pages 779–787.

Y. Miyao and J. Tsujii. 2008. Feature forest models for probabilistic HPSG parsing. *Computational Linguistics*, 34(1):35–80.

Yusuke Miyao, Rune Sætre, Kenji Sagae, Takuya Matsuzaki, and Jun'ichi Tsujii. 2008. Task-oriented evaluation of syntactic parsers and their representations. In *Proceedings of ACL-08: HLT*, pages 46–54, Columbus, Ohio, June. Association for Computational Linguistics.

R. Morante, V. Van Asch, and W. Daelemans. 2010. Memory-based resolution of in-sentence scopes of hedge cues. *CoNLL-2010: Shared Task*, page 40.

Ngan Nguyen, Jin-Dong Kim, and Jun'ichi Tsujii. 2011. Overview of the Protein Coreference task in BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

Tomoko Ohta, Sampo Pyysalo, and Jun'ichi Tsujii. 2011. Overview of the Epigenetics and Post-translational Modifications (EPI) task of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

S. Petrov, L. Barrett, R. Thibaux, and D. Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 433–440. Association for Computational Linguistics.

H. Poon and L. Vanderwende. 2010. Joint inference for knowledge extraction from biomedical literature. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 813–821. Association for Computational Linguistics.

Sampo Pyysalo, Tomoko Ohta, Rafal Rak, Dan Sullivan, Chunhong Mao, Chunxia Wang, Bruno Sobral, Jun'ichi Tsujii, and Sophia Ananiadou. 2011a. Overview of the Infectious Diseases (ID) task of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

Sampo Pyysalo, Tomoko Ohta, and Jun'ichi Tsujii. 2011b. Overview of the Entity Relations (REL) supporting task of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

Laura Rimell and Stephen Clark. 2009. Porting a lexicalized-grammar parser to the biomedical domain. *Journal of Biomedical Informatics*, 42(5):852 – 865. Biomedical Natural Language Processing.

R. Sætre, K. Yoshida, A. Yakushiji, Y. Miyao, Y. Matsubyashi, and T. Ohta. 2007. AKANE system: protein-protein interaction pairs in BioCreAtIvE2 challenge, PPI-IPS subtask. In *Proceedings of the Second BioCreative Challenge Workshop*, pages 209–212.

Kenji Sagae and Jun'ichi Tsujii. 2007. Dependency parsing and domain adaptation with LR models and parser ensembles. In *Proceedings of the CoNLL 2007 Shared Task*.

G. Schneider, M. Hess, and P. Merlo. 2007. Hybrid long-distance functional dependency parsing. *Unpublished PhD thesis, Institute of Computational Linguistics, University of Zurich*.

G. Szarvas, V. Vincze, R. Farkas, and J. Csirik. 2008. The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 38–45. Association for Computational Linguistics.

Y. Tateisi, A. Yakushiji, T. Ohta, and J. Tsujii. 2005. Syntax Annotation for the GENIA corpus. In *Proceedings of the IJCNLP*, pages 222–227.