Overview of the Pathway Curation (PC) task of BioNLP Shared Task 2013

Tomoko Ohta [1], **Sampo Pyysalo** [1], **Rafal Rak** [1], **Andrew Rowley**[1], **Hong-Woo Chun**[2],
Sung-Jae Jung [2,3], **Chang-Hoo Jeong** [2] **Sung-Pil Choi** [2,3], **Jun'ichi Tsujii** [4], **Sophia Ananiadou** [1]

[1]National Centre for Text Mining and School of Computer Science, University of Manchester
[2]Software Research Center, Korea Institute of Science and Technology Information (KISTI)
[3]Department of Applied Information Science, University of Science and Technology (UST)
[4]Microsoft Research Asia, Beijing, China

Abstract

We present the Pathway Curation (PC) task, a main event extraction task of the BioNLP shared task (ST) 2013. The PC task concerns the automatic extraction of biomolecular reactions from text. The task setting, representation and semantics are defined with respect to pathway model standards and ontologies (SBML, BioPAX, SBO) and documents selected by relevance to specific model reactions. Two BioNLP ST 2013 participants successfully completed the PC task. The highest achieved F-score, 52.8%, indicates that event extraction is a promising approach to supporting pathway curation efforts. The PC task continues as an open challenge with data, resources and tools available from `http://2013.bionlp-st.org/`

1 Introduction

Following developments in molecular biology, biological phenomena are increasingly understood on the molecular level, as the products of complex systems of molecular reactions. Pathway models formalizing biomolecules and their reactions in machine readable representations are a key way of sharing and communicating human understanding of these phenomena and of developing computational models of biological systems (Kitano, 2002). Many pathway models integrate knowledge from hundreds or thousands of scientific publications, and their curation requires substantial manual effort. To support this effort, we have developed PathText (Kemper et al., 2010) which provides a seamless environment integrating a pathway visualizer, text mining systems and annotation tools. Furthermore, automatic processing of the domain literature could thus potentially play
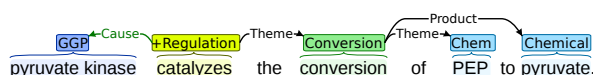

Figure 1: Event representation for a conversion reaction.

an important role in the support of pathway curation.

Information extraction targeting biomolecular reactions has been a major focus of efforts in biomedical natural language processing, with several tasks, resources, and tools addressing in particular protein-protein interactions (Krallinger et al., 2007; Pyysalo et al., 2008; Tikk et al., 2010). However, most such efforts have employed simple representations, such as entity pairs, that are not sufficient for capturing molecular reactions to the level of detail required to support the curation of pathway models. Additionally, previous efforts have not directly involved the semantics (e.g. reaction type definitions) of such models. Perhaps in part due to these reasons, natural language processing and information extraction methods have not been widely embraced by biomedical pathway curation communities (Ohta et al., 2011c; Ohta et al., 2011a).

We believe that the extraction of structured event representations (Figure 1) pursued in the BioNLP Shared Tasks offers many opportunities to make significant contributions to support the development, evaluation and maintenance of biomolecular pathways. The Pathway Curation (PC) task, a main task of the BioNLP Shared Task 2013, is proposed as a step toward realizing these opportunities. The PC task aims to evaluate the applicability of event extraction systems to pathway curation and to encourage the further development of methods for related tasks. The design of the task aims to address current issues in information extraction for pathway curation by explicitly basing its representation and extraction targets on ma-
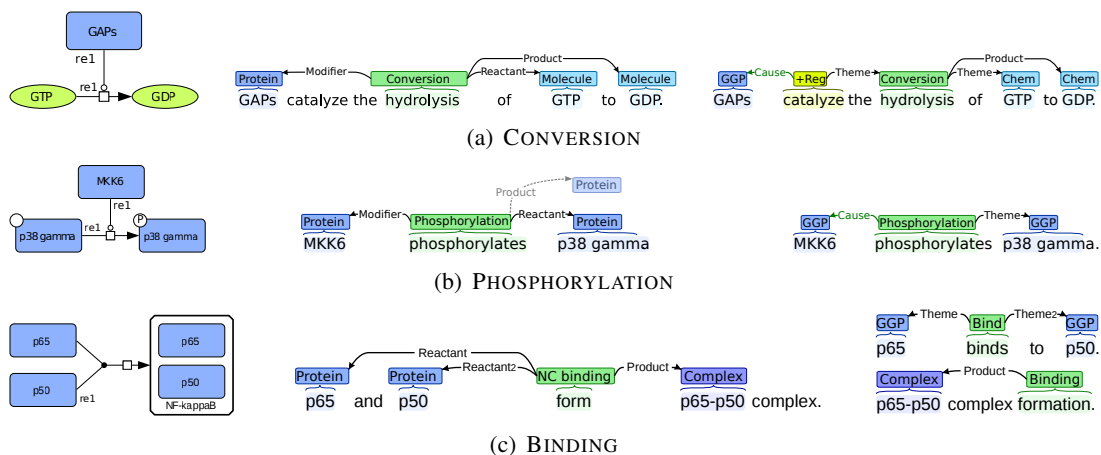
Figure 2: Illustration of pathway reaction (left), matching representation as an idealized text-bound event structure (middle) and applied event representation for statements actually appearing in text (right).

jor standards developed in the biomolecular pathway curation community, such as SBML (Hucka et al., 2003) and BioPAX (Mi et al., 2011), and ontologies such as the Systems Biology Ontology[1] (SBO) (Courtot et al., 2011). Further, The corpus texts are selected on the basis of relevance to a selection of pathway models from PANTHER Pathway DB[2] (Mi and Thomas, 2009) and BioModels[3] (Li et al., 2010) repositories. The PC task setting and its document selection protocol aim to account for both signalling and metabolic pathways, the latter of which has received comparatively little attention in recent domain IE efforts (Li et al., 2013).

2 Task setting

The PC task is formulated as an event extraction task (Ananiadou et al., 2010) following the general representation and task setting first introduced in the BioNLP ST 2009 (Kim et al., 2011). The primary aim is the extraction of event structures, or events, each of which can involve any number of physical entities or other events in specific roles.

The event representation is sufficiently expressive to allow the definition of event structures that closely parallel the definition of reactions in pathway representations such as SBML and BioPAX. These pathway representations differentiate between three primary groups of reaction participants: reactants ("inputs"), products ("outputs"), and modifiers, where the specific roles of modifiers can be further identified to differentiate e.g.

reaction catalysts from inhibitors. Correspondingly, the PC task applies the *Theme* role defined in previous BioNLP ST tasks to capture reactants, introduces a new *Product* role for products, and applies the previously defined *Cause* role and regulatory events to capture modifiers (Figure 2; see also Section 2.3).

It is important to note that while the event representation allows a one-to-one mapping to reactions in principle, an annotation scheme cannot guarantee that actual statements in text map to fully specified reactions: in free-form text, authors frequently omit mention of some entities taking part in reactions, perhaps most typically to avoid redundancies such as in "p38γ is phosphorylated into phospho-p38γ" (Figure 2b). Representations extracted from explicit statements in text will thus in some cases omit aspects of the corresponding complete reactions in pathway models.

Systems addressing the PC task are expected to extract events of specific types given 1) free-form text and 2) gold standard annotation for mentions of physical entities in that text. The task annotations also include equivalence relations and event modifications, a secondary extraction target. The annotation types are detailed below.

2.1 Entities

The entity annotation marks mentions of physical entities using start and end offsets in text (contiguous span) and a type selected from a fixed set. The following four entity types are marked in the PC task: SIMPLE CHEMICAL, annotated with reference to the Chemical Entities of Biological Interest (ChEBI) resource (Degtyarenko et al., 2008);

Entity type	Scope	Reference	Ontology ID
SIMPLE CHEMICAL	simple, non-repetitive chemical entities	ChEBI	SBO:0000247
GENE OR GENE PRODUCT	genes, RNA and proteins	gene/protein DBs	SBO:0000246
COMPLEX	entities of non-covalently linked components	complex DBs	SBO:0000253
CELLULAR COMPONENT	parts of cell and extracellular environment	GO-CC	SBO:0000290

Table 1: Entity types, definitions, and reference resources.

Event type	Core arguments	Additional arguments	Ontology ID
CONVERSION	*Theme*:Molecule, *Product*:Molecule		SBO:0000182
PHOSPHORYLATION	*Theme*:Molecule, *Cause*:Molecule	*Site*:SIMPLE CHEMICAL	SBO:0000216
DEPHOSPHORYLATION	*Theme*:Molecule, *Cause*:Molecule	*Site*:SIMPLE CHEMICAL	SBO:0000330
(Other modifications, such as ACETYLATION, defined similarly.)			
LOCALIZATION	*Theme*:Molecule	*At/From/ToLoc*:CELL. COMP.	GO:0051179
TRANSPORT	*Theme*:Molecule	*From/ToLoc*:CELL. COMP.	SBO:0000185
GENE EXPRESSION	*Theme*:GENE OR GENE PRODUCT		GO:0010467
TRANSCRIPTION	*Theme*:GENE OR GENE PRODUCT		SBO:0000183
TRANSLATION	*Theme*:GENE OR GENE PRODUCT		SBO:0000184
DEGRADATION	*Theme*:Molecule		SBO:0000179
BINDING	*Theme*:Molecule, *Product*:COMPLEX		SBO:0000177
DISSOCIATION	*Theme*:COMPLEX, *Product*:Molecule		SBO:0000180
REGULATION	*Theme*:ANY, *Cause*:ANY		GO:0065007
POSITIVE REGULATION	*Theme*:ANY, *Cause*:ANY		GO:0048518, GO:0044093
ACTIVATION	*Theme*:Molecule, *Cause*:ANY		SBO:0000412
NEGATIVE REGULATION	*Theme*:ANY, *Cause*:ANY		GO:0048519, GO:0044092
INACTIVATION	*Theme*:Molecule, *Cause*:ANY		SBO:0000412
PATHWAY	*Participant*:Molecule		SBO:0000375

Table 2: Event types and arguments. "Molecule" refers to an entity annotation of any of the types SIMPLE CHEMICAL, GENE OR GENE PRODUCT, or COMPLEX, and "ANY" refers to an annotation of any type, either entity or event. The indentation corresponds to ontological relationships between the event types: for example, PHOSPHORYLATION `is-a` CONVERSION and TRANSCRIPTION `part-of` GENE EXPRESSION.

GENE OR GENE PRODUCT, annotated with reference to gene and protein databases such as UniProt (Consortium, 2011), Entrez Gene (Maglott et al., 2005) and PFam (Finn et al., 2010); COMPLEX, annotated with reference to database resources covering complexes; and CELLULAR COMPONENT, annotated following the scope of the Gene Ontology `cellular component` subontology (Ashburner et al., 2000) (Table 1). For discussion of the relation between these types and the representations applied in pathway models, we refer to Ohta et al. (2011c).

In terms of mention types in text, the annotation for SIMPLE CHEMICAL, GENE OR GENE PRODUCT and COMPLEX covers entity name mentions only, while the annotation for CELLULAR COMPONENT covers entity name mentions, nominal mentions, and adjectival references (e.g. "*mitochondrial*").

2.2 Relations

The PC task defines one relation type, *Equiv* (equivalence), which can hold between entity
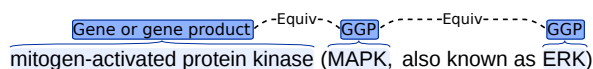

Figure 3: Example *Equiv* annotation.

mentions of the same type and specifies that they refer to the same real-world entity (Figure 3). These relations are only applied to determine if two events match during evaluation, where entities connected by an *Equiv* relation are considered interchangeable. Gold standard *Equiv* relations are applied also for test data, and systems participating in the task are not expected to extract these relations.

2.3 Events

The event annotation marks references to reactions, processes and comparable associations in scope of the annotation using the event representation. For the definition and scope of the event annotation, we rely primarily on the Systems Biology Ontology (SBO), drawing some general types not in scope of this ontology from the Gene Ontology (GO). Table 2 presents the event types anno-

Pathway	Repository	ID	Publication
mTOR	BioModels	MODEL1012220002	(Caron et al., 2010)
mTORC1 upstream regulators	BioModels	MODEL1012220003	(Caron et al., 2010)
TLR	BioModels	MODEL2463683119	(Oda and Kitano, 2006)
Yeast Cell Cycle	BioModels	MODEL1011020000	(Kaizu et al., 2010)
Rb	BioModels	MODEL4132046015	(Calzone et al., 2008)
EGFR	BioModels	MODEL2463576061	(Oda et al., 2005)
Human Metabolic Network	BioModels	MODEL6399676120	(Duarte et al., 2007)
NF-kappaB pathway	-	-	(Oda et al., 2008)
p38 MAPK	PANTHER DB	P05918	-
p53	PANTHER DB	P00059	-
p53 feedback loop pathway	PANTHER DB	P04392	-
Wnt signaling pathway	PANTHER DB	P00057	-

Table 3: Pathway models used to select documents for the task, with pathway repository model identifiers and publications presenting each model (when applicable).

tated in the PC task and their arguments. We refer again to Ohta et al. (2011c) for detailed discussion of the relation between these types and other representations applied in pathway models.

The role in which each event argument (entity or other event) participates in an event is specified as one of the following:

Theme entity/event that undergoes the effects of the event. For example, the entity that is transcribed in a TRANSCRIPTION event or transported in a TRANSPORT event.

Cause entity/event that is causally active in the event. Marks, for example, "P_1" in "P_1 inhibits P_2 expression".

AtLoc,FromLoc,ToLoc : location in which the *Theme* entity of a LOCALIZATION event is localized (*At*) in LOCALIZATION events not involving movement or is transported (or moves) from/to (*From/To*) in LOCALIZATION and TRANSPORT events involving movement.

Site site on the *Theme* entity that is modified in the event. Can be specified for modification events such as PHOSPHORYLATION.

Participant general role type identifying an entity that participates in some underspecified way in a high-level process. Only applied for the PATHWAY type.

2.4 Event modifications

In addition to events, the PC task defines a secondary extraction target, event modifications. Two modification types are defined: NEGATION and SPECULATION. Both are binary flags that modify events, the former marking an event as being explicitly stated as not occurring (e.g. "P is

not phosphorylated") and the latter as being stated in a speculative context ("P may be phosphorylated."). Both are defined in terms of annotation scope and semantics identically as in the BioNLP ST'09 (Kim et al., 2009).

2.5 Evaluation

The PC task evaluation applies the standard evaluation criteria established in the BioNLP ST 2009. These criteria relax exact matching between gold and predicted events in two aspects: approximate trigger boundary matching, and approximate recursive event matching. The former allows predicted event triggers to differ from gold triggers by one word, and the latter requires recursively referred events to only match in their core arguments (see Table 2). We refer to Kim et al. (2011) for a detailed definition of these criteria.

3 Corpus

This section presents the PC task corpus and its annotation process.

3.1 Document selection

To assure that the documents annotated for the PC task corpus are relevant to pathway reactions, we applied two complementary approaches, both selecting documents on the basis of relevance to a specific pathway reaction. First, we selected from the BioModels repository those pathway models with the largest numbers of manually created annotations referencing a specific PubMed document identifier. For each of these models, we extracted literature references, selected a random subset, downloaded the documents, and manually filtered to select abstracts that explicitly discuss relevant molecular reactions. Second, as only a small subset of models include explicit references to the

literature providing evidence for specific pathway reactions, we applied an alternative strategy where reactions from a selection of PANTHER DB models were entered into the PathText system (Kemper et al., 2010),[4] which is capable of suggesting documents relevant to given reactions based on an SBML model. We then selected a random set of reactions to query the system, and manually evaluated the highest-ranking documents to identify those whose abstracts explicitly discuss the selected reaction. We refer to Miwa et al. (2013a) for a detailed description of this approach. Table 3 presents the pathway models on which the document selection was based.

3.2 Annotation process

The base entity annotation for the PC corpus was created automatically using state-of-the-art entity mention taggers for each of the targeted entity types. For SIMPLE CHEMICAL tagging, the OS-CAR4 system (Jessop et al., 2011) trained on the chemical named entity recognition corpus of Corbett and Copestake (2008) was applied. For GENE OR GENE PRODUCT mention detection, the NERsuite[5] system trained on the BioCreative 2 Gene Mention task (Wilbur et al., 2007) corpus was used. NERsuite was also applied for CELLULAR COMPONENT mention detection, for this task trained on the Anatomical Entity Mention (AnEM) corpus (Ohta et al., 2012). Finally, COMPLEX annotations were created using a combination of a dictionary and heuristics making use of the GENE OR GENE PRODUCT annotation (for mentions such as "*cyclin E/CDK2 complex*"). To support the curation process, these tools were integrated into the NaCTeM text-analysis workflow system Argo (Rak et al., 2012).

Based on the evaluations of each of these tools in the studies presenting them, we expected initial automatic tagging performance to be in the range 80-90% in both precision and recall. Following initial automatic annotation, the entity mention annotation was manually revised to improve quality and consistency. As the entity annotation is not itself a target of extraction in the shared task, we did not separately evaluate the consistency of the revised entity mention annotation.

To assure that the quality and consistency of the event annotation are as high as possible, ini-

Item	Train	Devel	Test	Total
Documents	260	90	175	525
Words	53811	18579	35966	108356
Entities	7855	2734	5312	15901
Events	5992	2129	4004	12125
Modifications	317	80	174	571

Table 4: PC corpus statistics

tial event annotation was created entirely manually, without automatic support. This annotation effort was carried out using the BRAT annotation tool (Stenetorp et al., 2012) by a group of biologists in collaboration between NaCTeM and KISTI. Following initial annotator training and refinement of guidelines based on the event type definitions provided by the reference ontologies, the primary event annotation was created by three biologists. To evaluate and maintain annotation consistency, a random 20% of documents were annotated redundantly by all annotators, and these overlapping annotations were periodically evaluated and differences in annotation were discussed between the annotators and annotation coordinators. Following initial annotation, a round of semi-automatic consistency checks were applied using BRAT. Evaluation of the redundantly annotated documents using the primary task evaluation criteria gave an inter-annotator agreement of 61.0% in F-score. For the final corpus, the redundantly annotated documents were evaluated separately by an annotation coordinator to select the best of each set.[6]

The overall statistics of the corpus are summarized in Table 4. We note that the among the previous BioNLP ST corpora, only the GENIA (GE) task corpus has a larger number of annotated events than the PC corpus.

4 Results

4.1 Participation

Two groups submitted final results to the PC task, one from the National Centre for Text Mining (NaCTeM) and one from the University of Turku BioNLP group (TEES-2.1) (Table 5). Both participants applied their well-established, state-of-the-art event extraction systems, EventMine[7] (Miwa et al., 2012) (NaCTeM) and the Turku

---

			NLP		Events				Other resources	
Rank	Team	Org	Word	Parse	Trig.	Arg.	Group.	Modif.	Corpora	Other
1	NaCTeM	1NLP	Snowball	Enju, GDep	SVM	SVM	SVM	SVM	(see text)	triggers
2	TEES-2.1	1BI	Porter	McCCJ + SD	SVM	SVM	SVM	SVM	GE	hedge words

Table 5: Participants and summary of system descriptions. Abbreviations: BI=Bioinformatician, NLP=Natural Language Processing researcher, McCCJ=McClosky-Charniak-Johnson parser, Charniak=Charniak parser, SD=Stanford Dependency conversion, GE=GE task corpus.

Team	recall	prec.	F-score
NaCTeM	52.23	53.48	52.84
TEES-2.1	47.15	55.78	51.10

Table 6: Primary evaluation results

Event Extraction System[8] (Björne et al., 2011) (TEES). The two systems share the same overall architecture, a one-best pipeline with SVM-based stages for event trigger detection, trigger-argument relation detection, argument grouping into event structures, and modification prediction. The feature representations of both systems draw on substructures of dependency-like representations of sentence syntax, derived from full parses of input sentences. TEES applies the Charniak and Johnson (2005) parser with the McClosky (2009) biomedical model, converting the phrase-structure parses into dependencies using the Stanford tools (de Marneffe et al., 2006). By contrast, EventMine uses a combination of the predicate-argument structure analyses created by the deep parser Enju (Miyao and Tsujii, 2008) and the output of the the GDep best-first shift-reduce dependency parser (Sagae and Tsujii, 2007). All three parsers have models trained in part on the biomedical domain GENIA treebank (Tateisi et al., 2005).

Interestingly, both systems make use of the GE task data, but the application of EventMine extends on this considerably by applying a stacked model (Miwa et al., 2013b) with predictions also from models trained on the BioNLP ST 2011 EPI and ID tasks (Pyysalo et al., 2012) as well as from four corpora introduced outside of the shared tasks by Thompson et al. (2011), Pyysalo et al. (2011), Ohta et al. (2011b) and Ohta et al. (2011c).

4.2 Evaluation results

Table 6 summarizes the primary evaluation results. The two systems demonstrate broadly similar performance in terms of F-scores, with NaCTeM achieving an 1.7% point higher overall result.

However, the systems show quite different performance in terms of the precision/recall balance: while the NaCTeM system has little difference between precision and recall, TEES-2.1 shows a clear preference for precision, with 8.6% lower recall than precision.

Results are shown separately for each event type in Table 7. The results largely mirror the overall performance, with the NaCTeM system showing better performance for 13 out of the 21 event types present in the test data and more balanced precision and recall than TEES-2.1, which emphasizes precision over recall for almost all event types. Although the results do not include evaluation of EventMine with a reduced set of stacked models in training, the modest difference in performance suggests that comprehensive use of previously released event resources in EventMine did not confer a decisive advantage, perhaps in part due to differences in the event definitions between the PC task and previous resources.

Overall, the two systems appear quite similar not only in architecture but also performance, with the clearest systematic difference observed being the different emphases on precision vs. recall. As both systems are based on machine learning methods with real-valued outputs, it would be relatively straightforward to use prediction confidences to analyse performance over the entire precision-recall curve instead of a single fixed point. Such analysis could provide further insight into the relative strengths and weaknesses of these two systems.

5 Discussion

Although participation in this initial run of the PC task was somewhat limited, the two participating systems have been applied to a large variety of event extraction tasks over the last years and have shown consistently competitive performance with the state of the art (Björne and Salakoski, 2011; Miwa et al., 2012). It is thus reasonable to assume that the higher performance achieved by the

---
[8] http://jbjorne.github.io/TEES/

Event	NaCTeM			TEES-2.1		
	recall	prec.	F-score	recall	prec.	F-score
CONVERSION	34.33	35.48	34.90	35.82	42.86	**39.02**
PHOSPHORYLATION	62.46	55.94	59.02	53.40	66.00	**59.03**
DEPHOSPHORYLATION	45.00	56.25	**50.00**	35.00	77.78	48.28
ACETYLATION	69.57	72.73	71.11	82.61	76.00	**79.17**
DEACETYLATION	33.33	33.33	**33.33**	0.00	0.00	0.00
METHYLATION	42.86	60.00	50.00	57.14	80.00	**66.67**
DEMETHYLATION	100.00	100.00	100.00	100.00	100.00	100.00
UBIQUITINATION	52.94	64.29	58.06	58.82	76.92	**66.67**
DEUBIQUITINATION	100.00	100.00	100.00	100.00	100.00	100.00
LOCALIZATION	42.25	61.22	**50.00**	43.66	54.39	48.44
TRANSPORT	65.52	61.29	**63.33**	56.55	59.85	58.16
GENE EXPRESSION	90.65	83.15	**86.74**	84.55	79.39	81.89
TRANSCRIPTION	71.15	82.22	**76.29**	57.69	73.17	64.52
TRANSLATION	0.00	0.00	0.00	50.00	100.00	**66.67**
Simple-total	66.42	64.80	**65.60**	60.40	67.87	63.92
DEGRADATION	78.57	89.19	**83.54**	78.57	78.57	78.57
ACTIVATION	78.54	70.96	**74.56**	72.06	72.06	72.06
INACTIVATION	44.62	55.77	**49.57**	38.46	45.45	41.67
BINDING	64.96	47.30	**54.74**	53.96	53.96	53.96
DISSOCIATION	38.46	46.88	**42.25**	35.90	45.16	40.00
PATHWAY	84.91	75.50	**79.93**	70.94	75.50	73.15
General-total	69.07	62.69	**65.72**	61.16	65.74	63.37
REGULATION	33.33	33.97	33.65	29.73	39.51	**33.93**
POSITIVE REGULATION	35.49	42.81	38.81	34.51	45.45	**39.23**
NEGATIVE REGULATION	45.75	50.64	**48.07**	41.02	47.37	43.97
Regulation-total	37.73	42.79	**40.10**	35.17	44.76	39.39
Sub-total	53.47	53.96	**53.72**	48.23	56.22	51.92
NEGATION	24.52	35.87	29.13	25.16	41.30	**31.27**
SPECULATION	15.79	22.22	**18.46**	0.00	0.00	0.00
Modification-total	23.56	34.65	28.05	22.41	40.00	**28.73**
Total	52.23	53.48	**52.84**	47.15	55.78	51.10

Table 7: Primary evaluation results by event type.

task participants, a balanced F-score of 52.8%, is a good estimate of the performance level that can be attained for this task by present event extraction technology.

The results achieved by the two systems are broadly comparable to the best results achieved by any system in similar previously introduced event extraction tasks (Kim et al., 2012; Pyysalo et al., 2012). Given the novelty of the task domain and reference resource and the broad selection of documents, we find the results highly encouraging regarding the applicability of event extraction technology to supporting the development, evaluation, and maintenance of pathway models.

6 Conclusions

This paper presented the Pathway Curation (PC) task, a main event extraction task of the BioNLP ST 2013. The task was organized in collaboration between groups with an interest in pathway curation with the aim of evaluating and advancing the state of the art in event extraction toward methods for developing, evaluating and maintaining formal pathway models in representations such as SBML and BioPAX. We introduced an event extraction task setting with reference to pathway model standards and the Systems Biology Ontology, selected a set of 525 publication abstracts relevant to specific model reactions, and created fully manual

event annotation marking over 12,000 event structures in the corpus.

Two participants in the BioNLP ST 2013 submitted final predictions to the PC task, applying established, state-of-the-art event extraction systems, EventMine and the Turku Event Extraction System. Both systems achieved F-scores over 50%, with the EventMine system achieving the best overall result of 52.8%. This level of performance is broadly comparable with results achieved in comparable previously proposed tasks, indicating that current event extraction technology is applicable to the projected pathway curation support tasks.

To allow the further development and evaluation of event extraction methods for the task, the PC task continues as an open challenge to all interested participants, with the annotated corpus data, supporting resources, and evaluation tools available under open licenses from the task homepage, `http://2013.bionlp-st.org/`

Acknowledgments

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, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, et al. 2000. Gene ontology: tool for the unification of biology. *Nature genetics*, 25:25–29.

Jari Björne and Tapio Salakoski. 2011. Generalizing biomedical event extraction. In *Proceedings of the BioNLP Shared Task 2011 Workshop*, pages 183–191.

Jari Björne, Juho Heimonen, Filip Ginter, Antti Airola, Tapio Pahikkala, and Tapio Salakoski. 2011. Extracting contextualized complex biological events with rich graph-based feature sets. *Computational Intelligence*, 27(4):541–557.

Laurence Calzone, Amélie Gelay, Andrei Zinovyev, François Radvanyl, and Emmanuel Barillot. 2008. A comprehensive modular map of molecular interactions in rb/e2f pathway. *Molecular systems biology*, 4(1).

Etienne Caron, Samik Ghosh, Yukiko Matsuoka, Dariel Ashton-Beaucage, Marc Therrien, Sébastien Lemieux, Claude Perreault, Philippe P Roux, and Hiroaki Kitano. 2010. A comprehensive map of the mtor signaling network. *Molecular systems biology*, 6(1).

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

The UniProt Consortium. 2011. Ongoing and future developments at the universal protein resource. *Nucleic Acids Research*, 39(suppl 1):D214–D219.

Peter Corbett and Ann Copestake. 2008. Cascaded classifiers for confidence-based chemical named entity recognition. *BMC Bioinformatics*, 9(Suppl 11):S4.

Mélanie Courtot, Nick Juty, Christian Knüpfer, Dagmar Waltemath, Anna Zhukova, Andreas Dräger, Michel Dumontier, Andrew Finney, Martin Golebiewski, Janna Hastings, et al. 2011. Controlled vocabularies and semantics in systems biology. *Molecular systems biology*, 7(1).

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.

Kirill Degtyarenko, Paula De Matos, Marcus Ennis, Janna Hastings, Martin Zbinden, Alan Mcnaught, Rafael Alcántara, Michael Darsow, Mickaël Guedj, and Michael Ashburner. 2008. Chebi: a database and ontology for chemical entities of biological interest. *Nucleic acids research*, 36(suppl 1):D344–D350.

Natalie C Duarte, Scott A Becker, Neema Jamshidi, Ines Thiele, Monica L Mo, Thuy D Vo, Rohith Srivas, and Bernhard Ø Palsson. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, 104(6):1777–1782.

Robert D. Finn, Jaina Mistry, John Tate, Penny Coggill, Andreas Heger, Joanne E. Pollington, O. Luke Gavin, Prasad Gunasekaran, et al. 2010. The Pfam protein families database. *Nucleic Acids Research*, 38(suppl 1):D211–D222.

Michael Hucka, Andrew Finney, Herbert M Sauro, Hamid Bolouri, John C Doyle, Hiroaki Kitano, Adam P Arkin, Benjamin J Bornstein, et al. 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531.

David M. Jessop, Sam 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.

Kazunari Kaizu, Samik Ghosh, Yukiko Matsuoka, Hisao Moriya, Yuki Shimizu-Yoshida, and Hiroaki Kitano. 2010. A comprehensive molecular interaction map of the budding yeast cell cycle. *Molecular systems biology*, 6(1).

Brian Kemper, Takuya Matsuzaki, Yukiko Matsuoka, Yoshimasa Tsuruoka, Hiroaki Kitano, Sophia Ananiadou, and Jun'ichi Tsujii. 2010. Pathtext: a text mining integrator for biological pathway visualizations. *Bioinformatics*, 26(12):i374–i381.

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 BioNLP'09*.

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Junichi Tsujii. 2011. Extracting bio-molecular events from literature – the bionlp'09 shared task. *Computational Intelligence*, 27(4):513–540.

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.

Hiroaki Kitano. 2002. Systems biology: a brief overview. *Science*, 295(5560):1662–1664.

Martin Krallinger, Florian Leitner, and Alfonso Valencia. 2007. Assessment of the Second BioCreative PPI task: Automatic Extraction of Protein-Protein Interactions. In L. Hirschman, M. Krallinger, and A. Valencia, editors, *Proceedings of BioCreative II*, pages 29–39.

Chen Li, Marco Donizelli, Nicolas Rodriguez, Harish Dharuri, Lukas Endler, Vijayalakshmi Chelliah, Lu Li, Enuo He, et al. 2010. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 4:92.

Chen Li, Maria Liakata, and Dietrich Rebholz-Schuhmann. 2013. Biological network extraction from scientific literature: state of the art and challenges. *Briefings in bioinformatics*.

Donna Maglott, Jim Ostell, Kim D. Pruitt, and Tatiana Tatusova. 2005. Entrez gene: gene-centered information at ncbi. *Nucleic Acids Research*, 33(suppl 1):D54.

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

Huaiyu Mi and Paul Thomas. 2009. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. In *Protein Networks and Pathway Analysis*, pages 123–140. Springer.

Huaiyu Mi, Anushya Muruganujan, Emek Demir, Yukiko Matsuoka, Akira Funahashi, Hiroaki Kitano, and Paul D Thomas. 2011. Biopax support in celldesigner. *Bioinformatics*, 27(24):3437–3438.

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.

Makoto Miwa, Tomoko Ohta, Rafal Rak, Andrew Rowley, Douglas B. Kell, Sampo Pyysalo, and Sophia Ananiadou. 2013a. A method for integrating and ranking the evidence for biochemical pathways by mining reactions from text. *Bioinformatics*. in press.

Makoto Miwa, Sampo Pyysalo, Tomoko Ohta, and Sophia Ananiadou. 2013b. Wide coverage biomedical event extraction using multiple partially overlapping corpora. *BMC bioinformatics*, 14(1):175.

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

Kanae Oda and Hiroaki Kitano. 2006. A comprehensive map of the toll-like receptor signaling network. *Molecular Systems Biology*, 2(1).

Kanae Oda, Yukiko Matsuoka, Akira Funahashi, and Hiroaki Kitano. 2005. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular systems biology*, 1(1).

Kanae Oda, Jin-Dong Kim, Tomoko Ohta, Daisuke Okanohara, Takuya Matsuzaki, Yuka Tateisi, and Jun'ichi Tsujii. 2008. New challenges for text mining: mapping between text and manually curated pathways. *BMC bioinformatics*, 9(Suppl 3):S5.

Tomoko Ohta, Sampo Pyysalo, Sophia Ananiadou, and Junichi Tsujii. 2011a. Pathway curation support as an information extraction task. *Proceedings of LBM'11*.

Tomoko Ohta, Sampo Pyysalo, Makoto Miwa, and Jun'ichi Tsujii. 2011b. Event extraction for dna methylation. *Journal of Biomedical Semantics*, 2(Suppl 5):S2.

Tomoko Ohta, Sampo Pyysalo, and Jun'ichi Tsujii. 2011c. From pathways to biomolecular events: opportunities and challenges. In *Proceedings of BioNLP'11*, pages 105–113.

Tomoko Ohta, Sampo Pyysalo, Jun'ichi Tsujii, and Sophia Ananiadou. 2012. Open-domain anatomical entity mention detection. In *Proceedings of DSSD'12*, pages 27–36.

Sampo Pyysalo, Antti Airola, Juho Heimonen, Jari Björne, Filip Ginter, and Tapio Salakoski. 2008. Comparative analysis of five protein-protein interaction corpora. *BMC Bioinformatics*, 9(Suppl 3):S6.

Sampo Pyysalo, Tomoko Ohta, Makoto Miwa, and Jun'ichi Tsujii. 2011. Towards exhaustive event extraction for protein modifications. In *Proceedings of BioNLP'11*, pages 114–123.

Sampo Pyysalo, Tomoko Ohta, Rafal Rak, Dan Sullivan, Chunhong Mao, Chunxia Wang, Bruno Sobral, Jun'ichi Tsujii, and Sophia Ananiadou. 2012. Overview of the id, epi and rel tasks of bionlp shared task 2011. *BMC bioinformatics*, 13(Suppl 11):S2.

Rafal Rak, Andrew Rowley, William Black, and Sophia Ananiadou. 2012. Argo: an integrative, interactive, text mining-based workbench supporting curation. *Database: The Journal of Biological Databases and Curation*, 2012.

Kenji 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.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of EACL'12*, pages 102–107.

Yuka Tateisi, Akane Yakushiji, Tomoko Ohta, and Junichi Tsujii. 2005. Syntax annotation for the genia corpus. In *Proceedings of IJCNLP*, volume 5, pages 222–227.

Paul Thompson, Raheel Nawaz, John McNaught, and Sophia Ananiadou. 2011. Enriching a biomedical event corpus with meta-knowledge annotation. *BMC Bioinformatics*, 12(1):393.

Domonkos Tikk, Philippe Thomas, Peter Palaga, Jörg Hakenberg, and Ulf Leser. 2010. A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature. *PLoS Comput Biol*, 6(7):e1000837, 07.

John Wilbur, Lawrence Smith, and Lorraine Tanabe. 2007. BioCreative 2. Gene Mention Task. In L. Hirschman, M. Krallinger, and A. Valencia, editors, *Proceedings of BioCreative II*, pages 7–16.