IJCLA VOL. 4, NO. 1, JAN-JUN 2013, PP. 129–144
RECEIVED 07/12/12 ACCEPTED 11/01/13 FINAL 14/03/13

Something Old, Something New: Identifying Knowledge Source in Bio-events

RAHEEL NAWAZ, PAUL THOMPSON, AND SOPHIA ANANIADOU

University of Manchester, UK

ABSTRACT

Locating new experimental knowledge in biomedical texts is important for several tasks undertaken by biologists. Although several systems can distinguish between new and existing knowledge, this generally happens at the text zone level. In contrast to text zones, bio-events constitute structured representations of biomedical knowledge. They bridge text with domain knowledge and can be used to develop sophisticated semantic search systems. Typically, event extraction systems locate and classify events and their arguments, but ignore interpretative information (meta-knowledge) from their textual context. Since several events (often nested) can occur in a sentence, determining which event(s) are affected by which textual clues can be complex. We have analysed knowledge source annotation in two bio-event corpora: GENIA-MK (abstracts) and FP-MK (full papers), and have developed a system to classify bio-events automatically according to their knowledge source. Our system performs with an accuracy of over 99% on both abstracts and full papers.

KEYWORDS: *knowledge source, new knowledge, meta-knowledge, event, bio-event, machine learning*

1 Introduction

In recent years, several annotation schemes, e.g., [1-4] have been developed to identify and classify textual zones (i.e., continuous spans of text, such as sentences and clauses) in scientific papers, according to their rhetorical status or general information content. In most cases, these corpora have subsequently been used as a basis for training systems to recognise this information automatically, e.g., [5-7]. Common to all of these systems is the ability to identify information about *knowledge source*. That is, whether the text zone refers to new work being described in the paper, or to work that has already been described elsewhere. Such systems can be instrumental in helping users to search for text zones that contain new experimental knowledge. The identification of such information is important for several tasks in which biologists have to search and review the literature. One such example is the maintenance of models of biological processes, such as pathways [8]. As new reactions or new evidence for reactions become available in the literature, these should be added to the corresponding pathway(s). Another area where this information is useful is in the curation of biomedical databases. One of the tasks involved in keeping such databases up-to date is to search for new evidence for a particular interaction (e.g., gene regulation) within the literature [9].

In the types of task outlined above, the biologist is likely to be looking for specific types of biological processes or reactions, and specific types of information about them, e.g., what caused the reaction to occur, where the reaction took place, etc. Although the text zone classification systems cannot help with this kind of task, another type of system, i.e., an event extraction system, can be extremely useful. Event extraction systems are usually developed through training on manually annotated bio-event corpora, e.g., GENIA [10], BioInfer [11] and GREC [12]. These corpora identify named entities, such as genes and proteins, as well as the bio-events in which these entities participate. Systems are then trained to extract bio-event structures automatically from texts. The recent BioNLP Shared Tasks on event extraction in 2009 [13] and 2011 [14] have helped to stimulate considerable advances in event extraction research.

Event extraction facilitates the development of sophisticated semantic-based search systems, e.g., [15], which allow researchers to perform structured searches over events extracted from a large body of text [16]. Although search constraints can typically be specified in terms of event type (i.e., the process or reaction of interest) and/or the types of named

entities participating in the event, the ability to specify knowledge source as a constraint is not available. Bio-events are typically contained within a single sentence, and the existing text zone identification systems would normally be able to determine knowledge source at the sentence level. However, events are not the same as text zones. Whilst text zones constitute continuous spans of text, events usually consist of several discontinuous text spans, which correspond to different elements of the event, e.g., participants, location, etc. [17]. There are also (usually) several events contained within a single sentence. This means that just because a sentence or clause may be identifiable as having a particular knowledge source, it does not follow that all events contained within that text zone will have the same knowledge source; each event may have its own interpretation, and determining which events are affected by particular textual clues can be complex. For example, consider the following sentence:

Previous studies have shown that inhibition of the MAP kinase cascade with PD98059, a specific inhibitor of MAPK kinase 1, may prevent the rapid expression of the alpha2 integrin subunit.

This sentence contains not only a speculative analysis from an *Other* source, i.e., *Inhibition of the MAP kinase may prevent the expression of the alpha2 integrin subunit*, but also a general fact, i.e., *PD98059 is a specific inhibitor of MAPK kinase 1*. The main verb in the sentence (i.e., *prevent*) describes the information that has been reported in previous studies. In a sentence-based annotation scheme, this is likely to be the only information that is encoded. However, this means that the general fact is disregarded. Some annotation schemes have attempted to overcome the fact that sentences may contain multiple types of information by annotating meta-knowledge below the sentence level, i.e., clauses [18, 19] or segments [20]. In the case of the latter scheme, a new segment is created whenever there is a change in the meta-knowledge being expressed.

In the sentence above, however, it is not possible to split the sentence into continuous segments, since the general fact is embedded within the speculative analysis. In an event-based view of the sentence, this does not matter, since events consist of structures with different "slots", each of which is filled by a different text span, drawn from anywhere within the sentence. In this way, we say that the speculative analysis is triggered by the verb *prevent*, and has the participants *Inhibition of the MAP kinase* and *the rapid expression of the alpha2 integ-*

rin subunit. Similarly, the general fact can be encoded as a separate event. Only the speculative analysis event is referring to work being carried out as part of a particular study. The general fact event is considered to be established knowledge, and so it would not be correct to attribute this event to a particular previous study.

In order to allow further information to be encoded in event extraction systems, [21] proposed a multidimensional event-based meta-knowledge annotation scheme that includes knowledge source as a dimension of event interpretation. Other dimensions included in the scheme are: knowledge type, certainty level (allowing, amongst other things, speculative analyses to be encoded), polarity, and manner. This scheme has been manually applied to a number of different corpora. Firstly, the GENIA event corpus, comprising 1000 MEDLINE abstracts, was enriched to create the GENIA-MK corpus [22]. Secondly, a corpus of 4 full papers with event annotations has been enriched to create the FP-MK corpus [23]. A third, on-going effort is the application of the scheme to a corpus of stem cell research papers [24].

This paper describes our work on analysis and automated identification of knowledge source information about bio-events, using the GENIA-MK (abstracts) and FP-MK (full papers) corpora for training and testing. In both corpora, each event is ascribed one of two knowledge source values, i.e., *Current*, for events relating to work described in the current paper (default value), or *Other*, for events relating to work originally described elsewhere. Although the analysis carried out in [23] reveals that there are significant differences in the distributions of the different knowledge source values in abstracts and full papers, and that the textual means of denoting *Other* events also varies between abstracts and full papers, our system is able to perform to an almost identical level of accuracy on both text types, i.e., 99.6% and 99.4%, for abstracts and full papers, respectively.

2 Background

2.1 *Bio-event*

In its most general form, a **textual event** can be described as an action, relation, process or state [25]. More specifically, an event is a structured semantic representation of a certain piece of information contained within the text. Events are usually anchored to particular text fragments that are central to the description of the event, e.g., *event-*

trigger, *event-participants* and *event-location*, etc. A **bio-event** is a textual event specialised for the biomedical domain, in that it constitutes a dynamic bio-relation involving one or more participants [10]. These participants can be bio-entities or (other) bio-events, and are each assigned a semantic role like *theme* and *cause*, etc. Bio-events and bio-entities are also typically assigned semantic types/classes from particular taxonomies/ontologies. Consider the sentence S1: "*It has previously been reported [12] that LTB4 augments c-jun mRNA*". This sentence contains a single bio-event of type *positive_regulation*, which is anchored to the verb *augments*. Figure 1 shows a typical structured representation of this bio-event. The event has two participants: *c-jun mRNA* and *LTB4*, which have both been assigned their respective semantic types and roles within the event.

TRIGGER:	***augmented***
TYPE:	**positive_regulation**
THEME:	***c-jun mRNA*** : RNA_molecule
CAUSE:	***LTB4*** : organic_molecule

Fig. 1. Typical representation of the bio-event contained in sentence S1

2.2 *Knowledge Source*

As mentioned above, information about knowledge source is an integral part of a number of schemes for annotating text zones and their functions. The argumentative zoning (AZ) scheme, first introduced in [1], distinguishes sentences that mention OWN work presented in the current paper and OTHER specific work presented in another paper. Later extensions based on this scheme [2, 26] recognized that different types of information about OWN work can usefully be distinguished, such as OWN_METHD (methods) and OWN_RES (results) or OWN_CONC (conclusions). Multi-dimensional schemes allow several pieces of information to be associated with a given text span, and thus provide more flexibility regarding the types of information that can be encoded. Several such schemes encode information about knowledge source as a separate dimension, e.g., the scheme of [6] includes a novelty attribute (*New* or *Old*) that is distinct from their knowledge type attribute (*Background, Method, Conclusion*, etc.) The scheme of [3] identified five dimensions of information that could reliably be identi-

fied about text fragments (mostly clauses or sentences). Their *evidence* dimension includes information about the source of knowledge expressed in the text fragment. It has four possible values, which have similarities with some of the evidence codes used during the annotation for the Gene Ontology [27]. These values are: *E0*: no indication of evidence; *E1*: mention of evidence with no explicit reference; *E2*: explicit reference is made to other papers to support the assertion; *E3*: experimental evidence is provided directly in the text.

In the event-based meta-knowledge scheme of Nawaz et al. [21], information about the knowledge source of the event is encoded using the *Source* dimension, which has two possible values. The **Other** value is assigned when the event can be attributed to a previous study. This value is normally determined through the presence of explicit clues, e.g., *previously*, *recent studies*, etc., or cited papers, in the vicinity of the event. The **Current** value is assigned when the event makes an assertion that can be attributed to the current study. This is the default category, and is assigned in the absence of explicit lexical or contextual clues, although explicit clues such as *the present study* may be encountered. As an example, the bio-event in sentence S1 (section 1.1) has been attributed to another study through the use of an in-text citation. Therefore, it will be assigned the knowledge source value of *Other*.

2.3 *Annotation of Knowledge Source in GENIA-MK and FP-MK Corpora*

The GENIA-MK corpus consists of 1000 MEDLINE abstracts, containing 36,858 events, each of which has been annotated according to the meta-knowledge scheme described in [23]. In this corpus, slightly fewer than 2% of all events are assigned a *Source* value of *Other*. This is not surprising: abstracts are meant to provide a summary of the work carried out in a given paper and, given the very limited space, there is little opportunity to discuss previous work. Indeed, the use of citations is often prohibited in abstracts.

The FP-MK corpus consists of 4 full papers, in which 1,710 events have been annotated according to the same meta-knowledge scheme. In contrast to the GENIA-MK corpus, nearly 20% of all events in the FP-MK corpus belong to the *Other* category. The analysis provided in [23] examines the distribution of *Source* annotations in the various different sections in full papers. The study reports that by far the highest concentration of *Other* events is in the *Background* sections of the papers, where over 40% of the events are attributed to other sources. This is

expected, since it is normally in the *Background* section where one encounters the highest concentration of descriptions of previous work. The *Discussion* sections of the papers also have a high concentration (over 25%) of *Other* events, since it is common to compare and contrast the outcomes of the current work with those of previous related studies as part of the discussion. The frequency of *Other* events in the remaining sections is considerably lower. For example, in the *Results* sections of the papers, less than 7% of events are annotated as *Other.*

3 Analysis of *Other* Events

3.1 *Clue Frequency*

Table 1 shows the most commonly annotated clue expressions for *Source=Other* in the GENIA-MK (abstracts) and FP-MK (full papers) corpora respectively. For abstracts, several clue expressions contain the adverbs *previously* or *recently*, or their adjectival equivalents. The phrases *have been* and *has been* have also been annotated as clues with reasonably high frequency, the reason being that the use of the passive voice with the present perfect tense (e.g. *has been studied)* is a common means to indicate that an event has previously been completed (e.g., in a previous study), but yet has relevance to the current study.

Table 1. Most frequently annotated *Other* clues in GENIA-MK and FP-MK corpora

GENIA-MK (abstracts)			FP-MK (full papers)		
Cue	Freq	%	Clue	Freq	%
previously	118	21.7%	Citation	267	78.3%
has been	89	16.3%	has been	41	12.0%
recently	67	12.3%	previously	6	1.8%
have been	39	7.2%	recently	6	1.8%
previous	38	7.0%	latter example	4	1.2%
recent	32	5.9%	studies have shown	4	1.2%
earlier	6	1.1%	we and others	4	1.2%

In contrast to abstracts, the vast majority of clue expressions in full papers correspond to citations. However, similarly to abstracts, the use

of the present perfect tense is also quite common. Other explicit markers (such as *previously* and *recently*) constitute less than 10% of the clue expressions.

3.2 *Clue Ambiguity*

The presence of an *Other* clue in a sentence is not in itself sufficient evidence for assigning the knowledge source value of *Other* to all events in the sentence. While a sentence contains, on average, 4 bio-events, the majority of *Other* clues affect only one event in the sentence, i.e., the knowledge source value for the remaining events in the sentence is *Current*. Therefore, it is highly important that the syntactic/semantic structure of the sentence is considered, in order to determine which, if any, of the events are being affected by the clue. For example, the existence/type of dependency/constituency relations between the event participants and any *Other* clue(s) present in the sentence can be considered.

Furthermore, some of the *Other* clues (e.g., the tense of the sentence) are inherently ambiguous, and only indicate an *Other* event in certain contexts. For example, the clue expression *has/have been* is a significant clue for *Other* events – it accounts for over 23% of all *Other* events in abstracts and 12% of all *Other* events in full papers. However, an analysis of events from the sentences containing the phrase *has/have been* in the GENIA-MK corpus reveals that only 8% of these events are of type *Other*. This proportion is even lower (7%) for full papers.

3.3 *Event Complexity*

We examined the distribution of events assigned the value *Source=Other* amongst **simple** and **complex** events. By simple event, we mean an event whose participants are all entities, whilst a complex event is one with at least one participant which is itself an event. In abstracts, 67% of *Other* events are complex. Conversely, 2.26% of complex events are of type *Other*, while only 0.88% of simple events are of type *Other*. This means that an arbitrary complex event is 2.6 times more likely than an arbitrary simple event to have knowledge source value of *Other*.

In full papers, an even greater proportion of *Other* events (i.e., 72%) is complex. A total of 3.32% of complex events are of type *Other*, while only 0.73% of simple events belong to this type. Therefore, in

full papers, an arbitrary complex event is 4.5 times more likely than an arbitrary simple event to have knowledge source value of *Other*.

3.4 *Relative Position within Text*

In abstracts, 74% of *Other* events appear in the 2nd, 3rd or 4th sentence. Furthermore, over 80% of the *Other* events appear in the first half of the abstract.

In full papers, the section to which the sentence containing the event belongs is more significant than the relative position of the sentence within the paper or even within a section. For example, over 60% of all *Other* events found in full papers occur within the *Background* section.

4 Classifier Design

Based on the analysis of *Other* events, we engineered 7 feature sets. We used the Enju parser [28] to obtain the lexical and syntactic information required to construct these features. A brief explanation of each feature set is as follows:

- *Syntactic features* include the tense of the sentence (since *Other* events will normally be reported using the past tense), the POS tag of the *event*-trigger, and the POS tag(s) of *Other* clue(s) found in the sentence.
- *Semantic features* include the type of the bio-event and the type and role of each participant.
- *Lexical features.* Since the presence of lexical clues is usually key to determining *Other* events, these features include whether an *Other* clue is present in the sentence, and the clue itself. The clue list was compiled by combining the clue lists extracted from the GENIA-MK and FP-MK corpora, together with regular expressions to identify citations, which are also often important for the identification of *Other* events.
- *Lexico-semantic features.* Since the presence of an *Other* clue in a sentence does not usually affect all events within the sentence, these features help to determine the likelihood that a particular lexical clue for *Other* affects a given event. The features include the proximity (surface distance) between the *Other* clue and various event compo-

nents (*event-trigger*, *event-participants* and *event-location*), whether the *Other* clue precedes or follows the *event-trigger*, etc.

– *Dependency (lexico-syntactic) features.* Proximity of *Other* clues to event components is not always sufficient to determine which events they affect. In more complex sentences, it can be important to consider syntactic structure, since the *Other* clue may not occur close to the event components, but still be structurally related. For this reason, these features are based around the presence of direct and indirect dependency relations between the *Other* clue present in the sentence and the *event-trigger*, and the length of these dependency paths.

– *Constituency (lexico-syntactic) features.* This is a further class of structural features. They are based around the *command* [29] and *scope* relations, which are derived from the constituency parse tree. The command features consider the existence of S-, VP- and NP-command relations between the *Other* clue and the *event-trigger*. The scope features consider whether the *event-trigger* falls under the syntactic scope of the *Other* clue.

– *Positional features.* As mentioned above, *Other* events are far more numerous in certain sections of full papers, while within abstracts, earlier sentences are most likely to contain such events. Therefore, we include amongst our features the section in which the sentence containing the event appears (for abstracts all events have the same value and this feature becomes redundant), and the relative position of the sentence containing the event, both within the entire text and within the section.

We used the Random Forest [30] algorithm, which develops an ensemble/forest of Decision Trees from randomly sampled subspaces of the input features. Once the forest has been created, new instances are classified by first obtaining individual classifications from each tree and then using a majority vote to attain the final classification. We used the WEKA [31] implementation of the Random Forest algorithm, which is based on [30]. Our optimization settings included: (1) setting the number of trees in the forest to 10, (2) setting the number of features used to build individual trees to log(N+1), where N is the total number of features, (3) setting no restrictions on the depth of individual trees.

5 Results and Discussion

We conducted a series of experiments using different clue lists and feature set combinations. All results were 10-fold cross validated. The best results for abstracts and full papers are shown in Table 2. In both cases, the best results were achieved by using the 7 most frequent clues (Table 1) and all feature sets.

Table 2. Best results for GENIA-MK and FP-MK

Category	GENIA-MK (abstracts)			FP-MK (full papers)		
	P	R	F	P	R	F
Current	99.6%	99.8%	99.7%	99.5%	99.2%	99.3%
Other	83.3%	70.8%	75.6%	81.3%	70.1%	75.3%
Overall	99.4%	99.4%	99.4%	95.9%	93.4%	94.6%

5.1 *Abstracts*

In abstracts, only 2% of all events are of type *Other*; therefore, the baseline accuracy (through majority-class allocation) is 98%. Our system achieves an overall accuracy of 99.6%, which is considerably higher than this baseline. Recall for the *Other* category is significantly lower than the precision (over 10%). This is mainly due to the difficulty in identifying and disambiguating *Other* clues. The overall system precision and recall are both 99.4%.

5.2 *Full Papers*

The proportion of *Other* events in full papers is almost 10 times greater than in abstracts, with just under 20% of all events belonging to the *Other* category. The baseline classification accuracy for full papers is thus 80%. Therefore, statistically, identification of knowledge source in full papers is a harder task than in abstracts. However, our system achieves a very high overall accuracy of 99.4%. The main difference between the *Other* events in abstracts and full papers is the occurrence of explicit citations as clues. Since our system also includes citation related features, it is able to perform equally well on both corpora.

Similarly to the results for abstracts, precision for full papers is significantly higher than recall. Again, this is mainly due to the difficulty in identifying/disambiguating *Other* clues. This is also reflected in overall system performance as well, where precision is 2.5% higher than recall.

5.3 *Discussion*

Our results are the first that concern the detection of knowledge source at the event level. However, some comparisons can be drawn with similar previous work at the clause, sentence, and zone level. The text zone classification system of [5] achieved a precision/recall of 51%/30% for their OTHER category and a precision/recall of 85%/86% for the OWN category. [32] achieved an overall F-score of 70% for automatic zone classification, including BACKGROUND and OWN zones. The clause classification system reported by [7] performed with F-scores of 89%, 57%, 94% and 91% for the E0, E1, E2, and E3 classes respectively. [6], whose classification is performed at the sentence level, achieved an F-score of 64% for their BACKGROUND class; however, they did not try to identify the novelty attributes separately. Although we identify knowledge source at the event level, which is more challenging than similar tasks at the clause/sentence/zone level, our results are significantly higher. This is partly because we have cast the problem as a binary classification rather than a multi-category classification.

In our system, the most common reason for misclassification was the inability of the system to identify *Other* clues. This accounted for over 52% of the misclassified events. A significant proportion (32%) of misclassified events belonged to sentences with complex syntactic structures, e.g., where the *event-trigger* and the *Other* clue belonged to different clauses. These misclassifications can be partly attributed to parsing limitations, especially in terms of identifying complex dependency relations.

6 Conclusion

The isolation of new experimental knowledge in large volumes of text is important for several tasks undertaken by biologists. Although the ability to search for events of interest can significantly reduce the biologist's workload in finding relevant information, even more time could

be saved by facilitating further refinement of the search results to include only events pertaining to reliable new experimental knowledge. This goal can be achieved through the automatic recognition of event meta-knowledge. One of the most crucial aspects of identifying new experimental knowledge is to determine the knowledge source of the event.

In this paper, we have analysed the event-level knowledge source annotations in the GENIA-MK corpus (abstracts) and the FP-MK corpus (full papers). This analysis was used to inform the process of designing a system to recognise knowledge source automatically. We have shown that the knowledge source of events can be recognised to a high degree of accuracy. In abstracts, the overall accuracy is 99.6% and the overall F-score is 99.4%. The baseline accuracy for abstracts is already extremely high (98%), given that there are few events in abstracts that refer to previous work. However, a more significant result is that the performance of the classifier on full papers is almost as high as for abstracts, even though the baseline accuracy for full papers (80%) is considerably lower than for abstracts. On full papers, the classifier performs with an overall accuracy of 99.4% and achieves an overall F-score of 94.6%. These results provide encouraging evidence that the knowledge source of biomedical events can be predicted very reliably, regardless of text type. We plan to use our system to assist in the (semi-)automatic annotation of other corpora containing bio-event or relation annotation, e.g., [11, 12, 33]. This will pave the way for a more advanced system, able to recognise source information for a wider range of event and relation types. By integrating our classification system with event extraction systems, such as [34], we will be able to develop more sophisticated systems that can extract events with associated source information fully automatically. Events are also relevant to other domains. For example, the ACE 2005 evaluation involved the recognition of events in the general language domain, including events relating to conflict, business and justice. We are in the process of adapting our meta-knowledge scheme to this domain, which will allow systems to be trained to recognise knowledge source for events in alternative domains.

ACKNOWLEDGEMENTS. The work described in this paper has been funded by the MetaNet4U project (ICT PSP Programme, Grant Agreement: No 270893).

References

1. Teufel, S., Carletta, J., Moens, M.: An annotation scheme for discourse-level argumentation in research articles. Proceedings of EACL 110–117 (1999)
2. Mizuta, Y., Korhonen, A., Mullen, T., Collier, N.: Zone analysis in biology articles as a basis for information extraction. International Journal of Medical Informatics **75**:468–487 (2006)
3. Wilbur, W.J., Rzhetsky, A., Shatkay, H.: New directions in biomedical text annotations: definitions, guidelines and corpus construction. BMC Bioinformatics **7**:356 (2006)
4. Liakata, M., Teufel, S., Siddharthan, A., Batchelor, C.: Corpora for the conceptualisation and zoning of scientific papers. Proceedings of LREC 2010, 2054-2061 (2010)
5. Teufel, S.: Argumentative Zoning. Univ. of Edinburgh, Edinburgh (1999)
6. Liakata, M., Saha, S., Dobnik, S., Batchelor, C., Rebholz-Schuhmann, D.: Automatic recognition of conceptualisation zones in scientific articles and two life science applications. Bioinformatics **28** (2012)
7. Shatkay, H., Pan, F., Rzhetsky, A., Wilbur, W.J.: Multi-dimensional classification of biomedical text: toward automated, practical provision of high-utility text to diverse users. Bioinformatics **24**:2086–2093 (2008)
8. Oda, K., Kim, J.-D., Ohta, T., Okanohara, D., Matsuzaki, T., Tateisi, Y., Tsujii, J.i.: New challenges for text mining: mapping between text and manually curated pathways. BMC Bioinformatics **9:** S5 (2008)
9. Yeh, A.S., Hirschman, L., Morgan, A.A.: Evaluation of text data mining for database curation: lessons learned from the KDD Challenge Cup. Bioinformatics **19**: i331-i339 (2003)
10. Kim, J.-D., Ohta, T., Tsujii, J.: Corpus annotation for mining biomedical events from literature. BMC Bioinformatics **9** (2008)
11. Pyysalo, S., Ginter, F., Heimonen, J., Bjorne, J., Boberg, J., Jarvinen, J., Salakoski, T.: BioInfer: a corpus for information extraction in the biomedical domain. BMC Bioinformatics **8**: 50 (2007)
12. Thompson, P., Iqbal, S., McNaught, J., Ananiadou, S.: Construction of an annotated corpus to support biomedical information extraction. BMC Bioinformatics **10**:349 (2009)
13. Kim, J.D., Pyysalo, S., Ohta, T., Bossy, R., Nguyen, N., Tsujii, J.: Overview of BioNLP Shared Task 2011. Proceedings of BioNLP Shared Task 2011 Workshop, 1–6 (2011)
14. Kim, J.-D., Pyysalo, S., Nedellec, C., Ananiadou, S., Tsujii, J. (eds.): Selected Articles from the BioNLP Shared Task 2011, Vol. 13. BMC Bioinformatics (2012)

15. Miyao, Y., Ohta, T., Masuda, K., Tsuruoka, Y., Yoshida, K., Ninomiya, T., Tsujii, J.: Semantic Retrieval for the Accurate Identification of Relational Concepts in Massive Textbases. Proceedings of ACL, 1017–1024 (2006)
16. Ananiadou, S., Pyysalo, S., Tsujii, J., Kell, D.B.: Event extraction for systems biology by text mining the literature. Trends Biotechnol **28**: 381–390 (2010)
17. Kim, J.D., Ohta, T., Tsujii, J.: Corpus annotation for mining biomedical events from literature. BMC Bioinformatics **9**: 10 (2008)
18. Waard, A.d., Shum, B., Carusi, A., Park, J., Samwald, M., Sándor, Á.: Hypotheses, Evidence and Relationships: The HypER Approach for Representing Scientific Knowledge Claims. ISWC 2009, the 8th International Semantic Web Conference, Washington, DC., USA (2009)
19. Rubin, V.L.: Stating with certainty or stating with doubt: Intercoder reliability results for manual annotation of epistemically modalized statements. Human Language Technologies Conference: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, 141–144 (2007)
20. Wilbur, W.J., Rzhetsky, A., Shatkay, H.: New directions in biomedical text annotation: definitions, guidelines and corpus construction. BMC Bioinformatics **7**: 356 (2006)
21. Nawaz, R., Thompson, P., McNaught, J., Ananiadou, S.: Meta-Knowledge Annotation of Bio-Events. Proceedings of LREC 2010, 2498–2507 (2010)
22. Thompson, P., Nawaz, R., McNaught, J., Ananiadou, S.: Enriching a biomedical event corpus with meta-knowledge annotation. BMC Bioinformatics **12**: 393 (2011)
23. Nawaz, R., Thompson, P., Ananiadou, S.: Meta-Knowledge Annotation at the Event Level: Comparison between Abstracts and Full Papers. Proceedings of the Third LREC Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2012), 24-21 (2012)
24. Neves, M., Damaschun, A., Kurtz, A., Leser, U.: Annotating and evaluating text for stem cell research. Proceedings of the Third LREC Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2012), 16-23 (2012)
25. Sauri, R., Pustejovsky, J.: FactBank: A Corpus Annotated with Event Factuality. Language Resources and Evaluation **43**, 227–268 (2009)
26. Teufel, S., Siddharthan, A., Batchelor, C.: Towards discipline-independent argumentative zoning: Evidence from chemistry and computational linguistics. Proceedings of EMNLP 2009, 1493–1502 (2009)
27. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richard-

son, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene Ontology: tool for the unification of biology. Nature Genetics **25**: 25–29 (2000)

28. Miyao, Y., Tsujii, J.: Feature Forest Models for Probabilistic HPSG Parsing. Computational Linguistics **34**: 35–80 (2008)
29. Langacker, R.: On Pronominalization and the Chain of Command. In: Reibel, D., Schane, S. (eds.): Modern Studies in English. Prentice-Hall, Englewood Cliffs, NJ., 160–186 (1969)
30. Breiman, L.: Random forests. Machine Learning **45**: 5–32 (2001)
31. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explorations **11**: 10–18 (2009)
32. Mullen, T., Mizuta, Y., Collier, N.: A baseline feature set for learning rhetorical zones using full articles in the biomedical domain. ACM SIGKDD Explorations **7**: 52-58 (2005)
33. Bunescu, R., Ge, R., Kate, R.J., Marcotte, E.M., Mooney, R.J., Ramani, A.K., Wong, Y.W.: Comparative experiments on learning information extractors for proteins and their interactions. Artif Intell Med. **33**: 139–155 (2005)
34. Miwa, M., Saetre, R., Kim, J.D., Tsujii, J.: Event extraction with complex event classification using rich features. J Bioinform Comput Biol **8**: 131–146 (2010)

Raheel Nawaz
National Centre for Text Mining,
Manchester Interdisciplinary Biocentre, University of Manchester,
131 Princess Street, Manchester, M1 7DN, UK
E-mail: <raheel.nawaz@cs.man.ac.uk>

Paul Thompson
National Centre for Text Mining,
Manchester Interdisciplinary Biocentre, University of Manchester,
131 Princess Street, Manchester, M1 7DN, UK
E-mail: <paul.thompson@manchester.ac.uk>

Sophia Ananiadou
National Centre for Text Mining,
Manchester Interdisciplinary Biocentre, University of Manchester,
131 Princess Street, Manchester, M1 7DN, UK
E-mail: <sophia.ananiadou@manchester.ac.uk>