Enhancing Search: Events and Their Discourse Context

Sophia Ananiadou*, Paul Thompson, and Raheel Nawaz

National Centre for Text Mining, Manchester Institute of Biotechnology,
University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK
`{sophia.ananiadou,paul.thompson}@manchester.ac.uk,`
`raheel.nawaz@cs.man.ac.uk`

Abstract. Event-based search systems have become of increasing interest. This paper provides an overview of recent advances in event-based text mining, with an emphasis on biomedical text. We focus particularly on the enrichment of events with information relating to their interpretation according to surrounding textual and discourse contexts. We describe our annotation scheme used to capture this information at the event level, report on the corpora that have so far been enriched according to this scheme and provide details of our experiments to recognise this information automatically.

Keywords: event extraction, text mining, semantic search, discourse analysis.

1 Introduction

Data deluge makes finding relevant information increasingly difficult. Searching using keywords will usually return far more documents than are relevant to a query. A researcher interested in which proteins are *positively regulated* by IL-2 would typically expect the following sentence answering his query:

(1) p21ras proteins are activated by IL-2 in normal human T lymphocytes.

Using PubMed, a document containing (1) would be amongst the results retrieved using the search terms *IL-2* and *activate*. However, documents containing information directly relevant to the user's query may be hard to locate. Search engines view documents as "bags of words", omitting *relations* between search terms and do not incorporate variability in query terms e.g., acronyms, synonymous terms. Although users are interested in retrieving information about biological reactions that correspond to *positive regulations,* this can be expressed not only by the verb *activate* but other variations, e.g., *stimulate* or *affect*, or nominalisations such as *activation, activator, effect, stimulation*. Contextual interpretation is also important for a user, e.g., regulation may be negated: *p21ras proteins are not activated by IL-2 in normal human T lymphocytes.* Alternatively, there may be other types of information about the regulation specified in its textual context: *Our results suggest that p21ras proteins are strongly activated by IL-2 in normal human T lymphocytes.* The ability to specify

---

[*] Corresponding author.

A. Gelbukh (Ed.): CICLing 2013, Part II, LNCS 7817, pp. 318–334, 2013.

restrictions regarding interpretation or discourse function helps to further focus search results. For example, certain users may be interested specifically in negated interactions, whilst others may want to exclude them from their retrieved results. Other cases where interpretation can be important include matching hypotheses with experimental observations/evidence, or detecting contradictions that occur in the literature.

The above limitations of search engines can be alleviated through the integration of text mining methods [1–3] into customised search systems such as event-based search systems. Events are structured, semantic representations of pieces of knowledge contained within a text. In biomedicine, they include various biological processes, such as regulation, expression and transcription. Examples from newswire include terrorist attacks, company takeovers, personnel appointments, etc. In event-based search systems, searches take place over these structured events, not over unstructured text.

The sophistication of event-based search systems can be increased by automatic identification of contextual information, including information about discourse structure, such as causality relations [4], as well as level of certainty, negation, intensity of biological reaction, etc., and by allowing such types of information to be specified as restrictions on the types of events to be retrieved. We call these different types of information *meta-knowledge.*

In this paper, we firstly provide an overview of event-based text mining. We present our annotation scheme for enriched events with meta-knowledge information, corpora with event annotations, and describe how to train systems to recognise meta-knowledge information at the event level automatically.

2 Event-Based Text Mining

Recognising events in text usually involves separate identification and/or categorisation of several pieces of information in the text: triggers, the words around an event, and the event participants or arguments. Participants can include the instigator (or cause) of the event, the thing affected by the event (theme), etc. The information conveyed in (1) could thus be represented as a structured event as follows (based on [5]).

EVENT_TYPE: *positive_regulation*
TRIGGER: *activates*
CAUSE: *IL-2:PROTEIN*
THEME: *p21ras proteins:PROTEIN*
LOCATION: *normal human T lymphocytes:CELL*

The event representation above has been assigned an event type, drawn from an ontology of event types. Each participant is also assigned a named entity type. Participants can also be events themselves, i.e., events can be embedded within other events. Work on event extraction has not been limited to biomedical text, many earlier efforts were focused on newswire text.

The specific features of text, in terms of, e.g., the structure and language used, varies between domains. Event extraction systems thus must be adapted or reconfigured for different domains.

2.1 EventMine

EventMine [6] is a state-of the-art event extraction system, deployed for BioNLP Shared tasks on event extraction [7, 8]. It outperforms all systems in the BioNLP'09 ST subtask Task 2 and BioNLP'11 main tasks (GENIA, ID and EPI), achieving F-measures of 58.3%, 58.0%, 59.1%, 54.4%, other systems achieving F-measures of 57.4%, 56.0%, 57.6%, 53.3%, respectively. New features are constantly added, most recently by employing domain adaptation and coreference resolution [9]. It is a machine learning-based pipeline system with three detection modules for: (i) Event trigger/entity—assigns an appropriate trigger/entity category to each word that potentially constitutes the head word of an event participant; (ii) Event argument—finds semantic pair-wise relations among event participants; (iii) Multi-argument event—merges several pair-wise relations into complete event structures. It is designed to extract event structures from parser output. Any dependency parser could be substituted, but currently we use a combination of Enju [10] and GDep [11]. It extracts various token-related features (character n-grams, base form, parts-of-speech, etc). Contextual information is included in the feature set by taking into account dependency paths involving the focused word, n-grams of words surrounding the target word and its dependencies, and n-grams of words surrounding triggers and their identified arguments.

2.2 Semantic Event Searching: MEDIE

MEDIE[1] [12] facilitates event-based searching. A deep syntactic analyser tuned to the biomedical domain [13], an event expression recogniser and a named entity recogniser [14] provide its data. Queries take the form of <subject, verb, object> to specify an event, where *subject* and *object* refer to grammatical relations with the verb. In (1), the subject corresponds to the *Cause* participant, whilst the object corresponds to the *Theme*. One or more of the three "slots" in the query template can be left empty, in order to increase or decrease the specificity of the query: to find out which proteins are positively regulated by IL-2 we would specify: <*IL-2, activate, ?*>.

MEDIE addresses the issues of the simple keyword search engine, at least to a certain extent: (i) Only documents in which the specified grammatical relations hold between the search terms are retrieved, thus eliminating many of the spurious results retrieved by a traditional search engine; (ii) MEDIE detects named entities and event trigger terms, which are then linked with databases and ontologies. This allows automatic expansion of searches to include variants of search terms listed in these resources; (iii) Each sentence is automatically classified by title, objective, method, result or conclusion, and searches can specify which of these sentence types to consider when retrieving results [15]. For example, events in result sentences are likely to contain definite experimental results, whilst conclusion sentences will usually contain analyses or conclusions about experimental results.

Despite its advantages over a traditional search engine, MEDIE has limitations. It only allows two event participants (subject and object). Information on time, environmental conditions and manner is considered to be highly important to their correct interpretation [16]. MEDIE's search template is tied to the syntactic structure of the

---

[1] `http://www.nactem.ac.uk/medie/`

text. An approach in which users specify restrictions in terms of semantic rather than grammatical roles is more desirable. For instance, Cause and Theme semantic arguments do not consistently correspond to grammatical subject and object for all verbs. A semantic approach is even more desirable if additional participants (location, environmental conditions, etc.) may be specified as part of the search.

The meta-knowledge aspect of MEDIE involves classifying sentences as: *title, objective, method, result* or *conclusion*. Whilst useful, assignment of such information at the sentence level is often not sufficient when extracting information at the level of events. Sentence (2) helps to illustrate this:

> *(2) We conclude that the inhibition of the MAP kinase cascade with PD98059, a specific inhibitor of MAPK kinase 1, may prevent the inhibition of the alpha2 integrin subunit.*

In (2), two "top-level" events can be identified:

a) A somewhat tentative conclusion: *the inhibition of the MAP kinase cascade with PD98059 may prevent the inhibition of the alpha2 integrin subunit.*
b) A general fact: *PD98059 is a specific inhibitor of MAPK kinase 1.*

Sentence (2) is likely to be classified by MEDIE as a *conclusion*. However, the two events identified here have different interpretations: a) is fairly typical of the type of event that would normally be expected to occur in a *conclusion*, i.e., an analytical conclusion based on experimental results. However, events with other types of interpretations can also occur in such sentences as in b). Thus, to support event-based searching, it is preferable for any information relating to discourse structure and interpretation to be assigned at the level of the event, rather than at the sentence level.

2.3 Semantic Event Searching

MEDIE's search strategy is largely based on syntactic analysis of text. By allowing specification of search criteria via an intuitive semantic template that abstracts from the way events are specified in text, users without linguistic expertise can easily specify their exact search criteria. An ideal template would allow specification of the following types of search options: (i) Specification of event types (chosen from a fixed set) as an alternative to specific event trigger words or phrases. Ontologies of event types provide the user with control over the level of generality of the results returned by the query; (ii) Use of semantic role types rather than grammatical relations when specifying restrictions on event participants; (iii) A flexible way of specifying restrictions on the values of particular participants, in the form of either particular terms (e.g., *NF-kappa B*), NE classes (e.g., *PROTEIN*), or a combination; (iv) Specification of meta-knowledge about the event, e.g., should only facts be retrieved or are experimental analyses also acceptable. If so, are highly speculative analyses of interest, or only more definite analyses? The main challenges of producing a system that can extract events to match such a template are the following: (i) How each ontological event type manifests itself in the text, i.e., which words and phrases can be used as triggers; (ii) How the syntactic structure of the text maps in various ways to the semantic arguments of different types of events; and (iii) How meta-knowledge information about the event can be derived from the textual context of the event.

Resources such as the GENIA event ontology [5] provide an inventory of relevant biomedical event types. It is linked to the Gene Ontology [17], and allows us to obtain potential trigger words for some event type. Also, the BioLexicon [18], a large-scale terminological resource, can help with syntax-semantics mapping in event extraction: it provides an account of the syntactic and semantic behaviour of biomedical verbs.

In general, whilst external resources can help to improve the accuracy of event extraction systems, they are usually not sufficient to facilitate the recognition of a customised set of event types. A well-established method of adapting text mining systems to new domains is through training using annotated corpora [19-21]. To facilitate training of event extraction systems, corpora containing semantically annotated events, e.g., BioInfer [22], GENIA event corpus [5], MLEE[2] [23] and GREC [24] have been used. These corpora vary in several ways, including the richness of the ontologies used to categorise events and named entities, the numbers and types of the semantic arguments identified, the types of meta-knowledge information included as part of the annotation and the overall corpus size. By far the largest is the GENIA event corpus, consisting of 1000 abstracts, containing a total of 36,858 events.

3 Event Interpretation, Discourse Structure and Meta-knowledge

The BioNLP shared tasks distinguish speculated and non-speculated events. However, speculation can be expressed to varying degrees, and the ability to distinguish between these is useful for certain tasks, e.g., slight hedging indicates the authors are quite confident about the results of their analyses, but they may include a hedging device as a safeguard. In contrast, larger amounts of speculation can indicate that the event should be taken as a hypothesis.

Events with no explicit specification of speculation may have different interpretations. An event may be presented as the subject of an investigation, a known fact, an experimental observation or as an outcome of analysing experimental results. We may also distinguish events that represent knowledge cited from a previously published paper and events that constitute part of the new knowledge contribution in the current paper. Depending on the nature and criticality of the task being undertaken, some or all of the above distinctions may be important when searching for events in text.

A more detailed distinction between events is needed, according to their intended interpretation, based upon their textual and discourse context. To facilitate the automatic recognition of such information at the event level, we have designed an annotation scheme, tailored to enrich event-annotated corpora with meta-knowledge [25]. Whilst the current scheme version is tailored to annotating biomedical events, it is possible to identify domain-specific and domain-independent aspects such that, by extending a core set of concepts, it can be tailored to other domains. Following our description of the current biomedical annotation model, we describe our preliminary

---

[2] `http://www.nactem.ac.uk/MLEE/`

efforts to adapt our model to the social history domain. Much of this work has concerned either speculation/certainty level detection [26, 27], or assignment of information relating to the general information content or discourse function of the sentences, which has been carried out on abstracts [28, 29] and full papers [30, 31].

A smaller number of annotation schemes and systems has considered annotation of either multiple aspects of meta-knowledge, e.g., assigning both a general information category and if the sentence refers to new or previous work [32, 33] or both negation and speculation [34]. Uniquely amongst the above cited corpora, [34] also annotates the clue expressions (i.e. the negative and speculative keywords) on which the annotations are based, as does [35], which annotates several types of information relating to the interpretation of information in newspaper articles.

Few schemes explicitly annotate meta-knowledge clue expressions, yet these have been shown to be highly important for the recognition of meta-knowledge. For example, corpus-based studies of hedging (i.e., speculative statements) in biological texts [36, 37] found that 85% of hedges are conveyed through lexical means. Specific lexical markers can also denote other types of information pertinent to meta-knowledge identification, e.g., markers of certainty [38], as well as deductions or sensory (i.e., visual) evidence [36]. We have also shown that different types of meta-knowledge may be expressed through different words in the same sentence [39]. Thus, although meta-knowledge is not always conveyed through lexical clues (and conversely, presence of a particular lexical clue in the sentence does not guarantee the "expected" meta-knowledge interpretation), we consider the identification of meta-knowledge clue expressions as one of the keys to accurate meta-knowledge identification.

Other annotation schemes consider e.g., clauses [40] or sentence segments [41], to account for the fact that several types of information can be specified in one sentence.

Our multi-dimensional scheme for enriching events with meta-knowledge takes inspiration from other schemes, but, given that event structures are different from continuous spans of text, it has been tailored to encode exactly the types of information that can be readily identified for events. Indeed, it has been shown that the information encoded by our scheme can provide complementary information to that encoded by sentence and clause-based schemes [42].

3.1 Meta-knowledge Scheme for Biomedical Events

Our multi-dimensional meta-knowledge scheme maximally captures useful information specified about events in their textual context. Each dimension consists of a set of complete, mutually-exclusive categories: an event belongs to just one category in each dimension. Moreover, the interplay between the different dimension values can be used to derive further information (hyper-dimensions) on event interpretation. To minimise annotation burden, the number of possible categories within each dimension has been kept as small as possible, whilst respecting important distinctions in meta-knowledge we have observed during our corpus study. A brief overview of the dimensions of our scheme and their possible values is provided below.

Knowledge Type (KT): Captures the event's general information content. Each event is classified as: Investigation (enquiries, examinations), *Observation* (direct experimental observations), *Analysis* (inferences, interpretations, conjectures), *Method* (experimental methods), *Fact* (general facts, well-established knowledge) or *Other* (default: expresses incomplete information, or the KT is contextually unclear).

Certainty Level (CL): Encodes the confidence or certainty level ascribed to the event according to three levels: *L3* (default: no expression of uncertainty), *L2* (high confidence or slight speculation) or *L1* (low confidence or considerable speculation).

Polarity: Identifies negated events. We define negation as the absence or non-existence of an entity or a process. Possible values are *Positive* (default) and *Negative*.

Manner: Captures information about the rate, level, strength or intensity of the event: *High* (event occurs at a high rate or level of intensity), *Low* (event occurs at a low rate or level of intensity) or *Neutral* (default: no indication of rate/intensity).

Source: Encodes the source of the knowledge being expressed by the event as *Current* (default: the current study) or *Other* (any other source).

Hyper-Dimensions: Correspond to additional information that can be inferred by considering combinations of some of the explicitly annotated dimensions. We have identified two such hyper-dimensions each with *Yes* or *No* values: *New Knowledge* (inferred from *KT, Source* and *CL*) and Hypothesis (inferred from *KT* and *CL*).

3.2 Meta-knowledge Annotation of Biomedical Corpora

The scheme was firstly applied to the GENIA event corpus of 1,000 abstracts (36,858 events) to create the GENIA-MK corpus [43]. Whilst the scheme was designed via examination of biomedical abstracts, it is also important that meta-knowledge should also be readily identifiable for events in full papers, especially given the recent trend of extending event extraction techniques to apply to full papers [44]. This means that a classifier trained on abstracts is unlikely to give optimal performance if applied to full papers.

We are currently creating a corpus of full papers with meta-knowledge annotation. Our preliminary set of 4 papers (1,710 events) has already been annotated via the GENIA event annotation scheme [45]. Future work will involve the meta-knowledge enrichment of full papers that have been annotated with different types of events, such as those made available following BioNLP 2011 and 2013 shared tasks.

3.3 Analysis of Meta-knowledge Annotations in Biomedical Corpora

We have analysed our two meta-knowledge enriched corpora (full papers and abstracts), to discover and compare their different types of characteristics of events. Table 1 reports our analysis, giving the relative frequencies (RF) of events assigned meta-knowledge values in abstracts (A) and full papers (FP). To make clearer differences in the distribution of meta-knowledge values between these two text types, difference in relative frequencies is also shown, together with percentage change.

Table 1. Comparison of meta-knowledge values in full papers and abstracts

Dim.	Cat.	RF (FP)	RF (A)	Diff. in RF (FP – A)	% Change in RF
KT	Ana.	22.2%	17.8%	4.4%	24.8%
	Inv.	3.8%	5.3%	-1.5%	-39.0%
	Obs.	36.3%	34.7%	1.4%	4.1%
	Fact	4.1%	8.1%	-4.0%	-98.7%
	Meth.	5.8%	2.6%	3.2%	120.8%
	Oth.	27.8%	31.3%	-3.5%	-12.7%
CL	L1	2.3%	2.1%	0.2%	9.7%
	L2	9.5%	6.0%	3.5%	57.6%
	L3	88.2%	91.9%	-3.7%	-4.2%
Polarity	Neg.	3.6%	6.1%	-2.5%	-66.7%
	Pos.	96.4%	93.9%	2.5%	2.6%
Manner	High	3.9%	3.8%	0.1%	2.2%
	Low	0.8%	0.8%	0.0%	0.0%
	Neut.	95.2%	95.3%	-0.1%	-0.1%
Source	Cur.	80.0%	98.5%	-18.5%	-23.1%
	Oth.	20.0%	1.5%	18.5%	1248.6%
Hyper-Dimensions	N.K	28.6%	43.4%	-14.8%	-51.7%
	Hypo.	15.2%	13.4%	1.8%	13.4%

In Table 1, in most cases, the “rankings” of each value within a particular dimension remain constant between full papers and abstracts and the absolute differences between the relative frequencies are small. However, the percentage change in relative frequencies between abstracts and full papers reveals significant differences.

Most notable is the difference between the relative frequencies of events that are assigned *Source=Other* in full papers and abstracts—full papers contain 12.5 times more such events than abstracts. Furthermore, citations, the most common way to denote previous work, are often not allowed in abstracts: full papers normally mention related work extensively, most notably in Background and Discussion sections.

There are differences in values of the KT dimension. For example, *Method* events are more than twice as abundant (in terms of relative frequency) in full papers. Since the average size of abstracts in the GENIA event corpus is 9 to 10 sentences [5], the relative frequency of *Fact* events in abstracts is high (over 8%). In full papers, factual events only appear half as frequently. The reason is that the only type of section in full papers in which *Fact* events occur with any significant frequency is *Background* (over 7% of all events in this section type), where the current state of knowledge is also discussed in detail. In contrast, other sections in full papers are more concerned with experimental details. A similar argument explains why *Investigation* events are more frequent in abstracts: most abstracts describe the purpose of the study, but a smaller proportion of full papers is devoted to describing investigations. In contrast to the previous two categories mentioned, the proportion of *Analysis* events is ~25% higher in full papers: in contrast to *Fact* and *Investigation* events, *Analysis* events are found with high frequency in several sections of full papers. There is much less variation in the *Observation* category, suggesting that the clear reporting of experimental observations is equally important throughout both full papers and abstracts.

Authors are more cautious in detailing their results in the body of papers, to maintain credibility in case these are later disproved. “Scientists gain credibility by stating the strongest claims they can for their evidence, but they also need to insure against overstatement.” ([36] p. 257). Authors achieve this by using slight hedging (*L2*). Greater speculation (*L1*) is less common, as credibility is thus reduced. The fact that the proportion of slightly hedged *Analysis* events is particularly high in the *Results*, *Discussion* and *Conclusion* sections of full papers, rising as high as 51% in the *Discussion* sections, explains why *L2* events are over 57% more frequent in full papers.

Regarding *Polarity*, the relative frequency of negated events is significantly (67%) higher in abstracts. This is partly due to the fact that negative results are sometimes more significant than positive results [46], and are thus highlighted in the abstracts.

There is little difference in the relative frequencies of different values of *Manner* in both text types. For the hyper-dimensions, there is a higher proportion of hypotheses whilst for *New Knowledge*, there is a more significant difference. In abstracts, just under half of all events report new knowledge: unsurprising, given the previously specified main aims of abstracts. In contrast, there is much more room in full papers for describing and discussing previous work, and speculating about results.

3.4 Adapting the Meta-knowledge Scheme to a New Domain

We have investigated applying our scheme in the ISHER project[3] on social history, which aims to enhance search over digitised social history resources, through text mining-based rich semantic metadata extraction for collection indexing, clustering and classification, thus supporting semantic search. Semantic metadata include both named entities and events. As part of the training data, we use the Automatic Content Evaluation (ACE) 2005 corpus, which contains events, i.e., *Conflict* (Attack, Demonstrate) and *Justice* (e.g. Arrest-Jail, Sentence, Fine, etc.). We are enriching relevant events in the corpus with meta-knowledge annotation.

Three of the original meta-knowledge dimensions are useful for ACE, i.e., *Polarity, Source* and *CL*, as these dimensions and their values represent general characteristics of all text types. *Manner* is not relevant to the social history domain but *Knowledge Type (KT)* is, although a different set of values may be required for each different domain. The existing categories are very specific to academic papers, while an examination of events in the ACE corpus suggests that, although some categories may remain constant across different domains, other categories are domain specific. For example, although events describing facts and analyses of information can also be observed in the ACE corpus,

---

[3] `http://www.nactem.ac.uk/DID-ISHER/`

other types of information require different categories, e.g.: hypothetical events: *It could swell to as much as $500 billion if* ***we go to war in Iraq****, or* opinions: *Dan Snyder of Baden, Pennsylvania writes, "****Bush should torture the al Qaeda chief operations officer****."*

4 Automatic Detection of Meta-knowledge

We have extended EventMine to extract events and assign meta-knowledge to them. In EventMine-MK [47], meta-knowledge assignment is implemented as a separate module. We used two types of features: **event structure** concerned with the text surrounding both the event trigger and its arguments, both in terms of immediate context and dependency paths and **meta-knowledge clue features.** They include the position in the abstract of the sentence that contains the event, which is used since certain types of meta-knowledge (particularly events belonging to different values within the KT dimension) tend to appear in fixed places in abstracts (e.g., events with the KT type Fact or Observation often appear towards the beginning of an abstract). A citation feature refers to the presence of citations. Citations are extracted via a regular expression that matches parentheses or brackets surrounding numbers (e.g., [108]) or sequences ending in 4 digits (e.g., (..., 1998)). Clues for Other (Source dimension) often constitute citations, and thus are not covered by the clue dictionaries.

Through experimentation with different combinations of the above feature types, we found that the exact combinations of features that produce the best results vary according to the meta-knowledge dimension under consideration. However, since for each dimension, the difference between the performance of the best setting and the setting in which all features are enabled is less than 1%, we decided to enable all features for all dimensions, due to the extra computational and spatial costs that would be required to calculate and store a different set of features for each dimension.

Table 2. Results of applying EventMine-MK to the ST-MK corpus (**R**ecall, **P**recision, **F**-score)

Dimension	Average type	R/P/F	+GENIA (R/P/F)	Majority (R/P/F)
KT	Macro	56.5 / 60.9 / 57.3	56.2/ 59.7 / 57.3	16.7 / 6.8 / 9.6
	Micro	74.6 / 74.6 / 74.6	73.9/ 73.9 / 73.9	40.5 / 40.5 / 40.5
CL	Macro	57.0/49.3/52.3	66.8/87.1/69.2	33.3 / 32.2 / 32.8
	Micro	96.6 / 96.6 / 96.6	97.7 / 97.7 / 97.7	96.7 / 96.7 / 96.7
Polarity	Macro	84.5 / 77.2 / 80.3	82.5 / 79.8 / 81.0	50.0 / 47.9 / 48.9
	Micro	96.4 / 96.4 / 96.4	96.9 / 96.9 / 96.9	95.9 / 95.9 / 95.9
Manner	Macro	91.9 / 76.4 / 82.8	91.4 / 76.8 / 82.8	33.3 / 31.4 / 32.3
	Micro	96.2 / 96.2 / 96.2	96.3 / 96.3 / 96.3	94.1 / 94.1 / 94.1
Source	Macro	82.1 / 90.7 / 85.9	82.1/94.8/87.4	50.0 / 49.3 / 49.7
	Micro	99.3 / 99.3 / 99.3	99.4 / 99.4 / 99.4	98.6 / 98.6 / 98.6

EventMine-MK also makes use of a meta-knowledge prediction model, trained on the original GENIA-MK corpus, which has richer semantic information about events than the ST corpus (the corpus from the BioNLP'09 shared task, on which EventMine-MK was trained), and a much greater number of events. Thus, a model trained on the GENIA-MK corpus should predict meta-knowledge more accurately than one trained on the ST-MK corpus. Unfortunately, the differences in event types and distribution of meta-knowledge values in the ST corpus mean that direct application of the GENIA-MK trained model to predict meta-knowledge on the ST data will not produce optimal results. However, we did find that indirect use of the GENIA-MK model on the ST data (i.e., by adding additional features based on this model to the meta-knowledge model trained on the ST corpus), improved performance of meta-knowledge assignment. Table 2 reports results of applying EventMine-MK to the ST-MK corpus. For each dimension, micro and macro average scores are shown. For reference, a majority baseline is shown, i.e., the scores that would be achieved if each event was assigned the dimension value that appears most frequently in the corpus. Although performance is different for each dimension, it is in most cases superior to the baseline, in some cases by a significant margin. The +GENIA column shows the effects of adding features based on the model trained on the GENIA-MK corpus. For most dimensions, we see some improvement when these features are added. The effect is very noticeable when the macro average of the CL dimension is considered.

Table 3. Comparison of EventMine-MK with other systems on the task of negation and speculation detection

	Negation	**Speculation**	**Total**
EventMine-MK (+clues)	**29.96**/42.24 /**35.05**	21.63/36.59 / 27.19	**25.98**/39.79 / **31.43**
EventMine-MK	28.19/36.16 / 31.68	22.12/**41.82** / 28.93	25.29/38.33 / 30.47
[48]	22.03/49.02 / 30.40	19.23/38.46 / 25.64	20.69/43.69 / 28.08
[49]	18.06/46.59 / 26.03	**23.08**/40.00/ **29.27**	20.46/42.79 / 27.68
[50]	15.86/**50.74** / 24.17	14.98/50.75 / 23.13	16.83/**50.72** / 25.27

To evaluate and compare EventMine-MK, we applied it to the BioNLP'09 ST subtask (Task 3) of extracting events with associated negation and speculation information. Although this task does not deal with all aspects of meta-knowledge that can be predicted by our system, there are currently no other systems that can predict the values of other meta-knowledge dimensions at the event level, and so further comparisons cannot be undertaken. Two versions of EventMine-MK were trained, one on the ST-MK corpus, and one on the original ST corpus, which was annotated for negation and speculation, but not for negation and speculation clues. This latter corpus was the one used by the other systems compared for training, and so allows more direct comparison. Performance is reasonably low for all systems in Table 3, because the evaluation settings take into account event extraction performance as well as negation/speculation detection. We see that whether or not EventMine-MK is trained using meta-knowledge clues, it outperforms the top 3 systems that participated in the original task, in terms of both overall F-scores and F-scores for negation detection. We also see that meta-knowledge clue annotation helps improve performance, especially in detection of negated events. This provides strong evidence that our decision to annotate meta-knowledge clues was correct. For speculation, a small decrease in performance is observed when meta-knowledge clues are taken into account. However, this decrease reinforces the analysis by [44], that speculation annotations in the ST corpus do not conform to the standardised notion of speculation, i.e., in contrast to the events enriched with meta-knowledge annotation, events occurring with modal verbs (e.g., *may*) and epistemic adverbs (e.g., *probably*) are rarely annotated as speculative in the ST corpus. The model was also applied to the full-text subset of BioNLP-ST'11 GENIA corpus, to investigate differences in the distribution of meta-knowledge values in full papers and abstracts. Since the results obtained are somewhat different to our manual annotation efforts, this provides further evidence for our earlier hypothesis that different models need to be trained for abstracts and full papers.

4.1 Experiments with Individual Dimensions

EventMine-MK is designed to be robust and efficient, facilitating scalability to large scale event extraction. The overall efficiency of the framework used, together with spatial and computational costs, are all important considerations. The meta-knowledge module of EventMine-MK uses the same machine learning algorithm as other modules, i.e., SVMs, and also uses the same set of features for each meta-knowledge dimension. Although EventMine-MK produces very competitive results for negation and speculation detection, we decided to do some smaller-scale experiments to investigate whether the results for other meta-knowledge dimensions could be improved by using alternative feature sets and/or machine learning algorithms.

The meta-knowledge specific features of EventMine-MK take into account several more general observations about textual features that can affect meta-knowledge values. However, we decided to carry out a more in-depth analysis of individual dimensions to help suggest a customised set of features for each dimension, leading to improved prediction results. So far, we have carried out such analyses for two meta-knowledge dimensions, *Manner* and *Source*. Each study is characterised by detailed analysis of the contexts in which the different values of the given meta-knowledge dimension can occur, together with the different types of clues that can be used.

Both studies use a different set of core features, falling into six different categories, i.e., syntactic, semantic, lexical, lexicosemantic, dependency and constituency. The core features are more wide ranging than those used by the meta-knowledge prediction module of EventMine, particularly the use of sematic information about the bio-event (semantic types of events and participants, semantic roles assigned to participants, etc.), and the use of constituency features as well as dependency features.

Our detailed analysis of *Manner* cues [51] revealed that 8% of clues for *High* manner are of the form *n-fold*, in which *n* represents a number. Since *n* can vary, matching with clue lists is not the correct strategy here. In addition, the exact form of expression can vary, and in the GENIA-MK corpus, 13 different variants of this numerical expression have been annotated as *High* cues. Some examples include *2-fold*, 4-*6 fold*, *5-to 7-fold*, etc. Accordingly, we use customised regular expressions to find such clues, which are subsequently included amongst the lexical clues extracted. In addition, the expression of negation inverts the polarity of a manner cue. For example, the word *significant* acts as a *High* cue, but its negated form (*no/not significant*) is a *Low* cue. Therefore, one of the lexical features used determines whether a negation cue is in the textual context of the event.

For the *Source* dimension [52] customised features include the tense of the main verb in the sentence (since events with *Source=Other* are often reported using the past tense). Also, positional features are included, as over 80% of *Other* events were found to occur in the first half of abstracts. A further interesting result of our analysis is that there is a correlation between event complexity and *Source* value. By event complexity, we mean whether an event is simple (i.e, if all of its participants are entities) or complex (i.e., if one or more of its participants is itself an event). Our analysis revealed that in abstracts, an arbitrary complex event is 2.6 times more likely than an arbitrary simple event to have knowledge source value of *Other*, whilst 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.*

For both the *Source* and *Manner* dimensions, training was done using the GENIA-MK corpus. The classifiers assume events have been pre-annotated: they do not attempt to recognise events as well as meta-knowledge. In both cases, we used the Random Forest algorithm to carry out the training, which develops an ensemble/forest of Decision Trees from randomly sampled subspaces of the input features. We used this algorithm since it has been successfully applied to various text mining and bioinformatics tasks. In particular, our recent experiments on detecting negated events [53] revealed that the Random Forest algorithm outperforms several other algorithms, including SVMs.

The experiments use 10-fold cross validation, so results can be compared to those produced when training the SVM classifier of the EventMine-MK meta-knowledge module on the GENIA-MK corpus, as we also report 10-fold cross validation results.

For *Source,* the best result achieved by the SVM classifier was micro-averaged F-Score of 98.4%, which is the same as majority baseline for this dimension (since the vast majority of events are assigned *Source=Current*). In comparison, the Random Forest classifier was able to achieve an improvement over the majority baseline, with a micro-averaged F-score of 99.4%. For *Manner,* the micro-averaged F-score for the SVM classifier was 95.4%, which again is virtually the same as the majority baseline for this dimension. In contrast, the micro-averaged F-Score for the Random Forest classifier is almost 3% higher, i.e., 98.3%. In terms of macro-averages, the gap is greater, with the SVM classifier achieving a macro-averaged F-Score of 59.2% for *Manner*, compared to 83.9% for the Random Forest classifier.

Results show that the Random Forest classifier is better suited to meta-knowledge value prediction than an SVM classifier, when used with customised feature sets. Future work involves a detailed analysis of combinations of features and learning algorithms to arrive at an optimal solution for automatic meta-knowledge prediction.

5 Conclusion

We have examined various aspects of event-based text mining. We have looked at how various text mining techniques can enhance users' search experience and help them to locate the information they require in a much more focussed and efficient way. In particular, we have discussed how analysing the structure of documents can help to restrict search results, how query expansion techniques can help to increase the number of relevant documents returned by a search; and how automatic recognition of various types of information (i.e., meta-knowledge) about the discourse and textual contexts of events provide even greater potential to refine event-based search queries according to the specific tasks being undertaken by individual users.

Using the MEDIE event-based search engine as an example, we showed that, whilst searching for events according to syntactic structures constitutes a robust method that offers improvements over traditional search engines, a more semantically based approach to extracting and searching for events provides greater power and flexibility to the user. The emergence of corpora with semantic annotation, coupled with the challenges posed by the BioNLP shared tasks, has helped to drive the development of more semantically-oriented event extraction systems. In particular, we provided details of EventMine, our own state-of-the-art event extraction system.

Since most event extraction systems to date have not attempted the detailed recognition of meta-knowledge information, the latter part of the paper has provided a summary of the various types of research that we have carried out within this area, including: the motivation for carrying out meta-knowledge recognition at the event level, rather than higher-level text units; the design of the original meta-knowledge scheme, tailored to enriching biomedical events; the application of the scheme to corpora of both abstracts and full papers; a comparison of the differences in the annotation results between abstracts and full papers; an investigation into how the meta-knowledge scheme could be adapted to other domains; and finally, an explanation of our most recent work, which has involved a number of efforts to train systems to recognise meta-knowledge automatically. This latest work consisted of two strands. The first of these extended EventMine with a module that is able to assign meta-knowledge information along the five dimensions of the scheme to automatically extracted events. The module follows the same structure as other modules in the EventMine pipeline, and is intended to strike a balance between accuracy, efficiency and robustness. Our second strand of work involved smaller scale experiments to investigate ways of improving on the prediction of meta-knowledge values, by considering different feature sets, coupled with a different machine-learning algorithm. The encouraging results achieved by these experiments suggest that there may be ways to improve the extraction of meta-knowledge in future versions of EventMine-MK.

As future work, we will analyse the effects that different machine learning algorithms and sets of features have on the accurate prediction of meta-knowledge information. We also aim to integrate more sophisticated event extraction technology in our search engines including MEDIE, Europe PubMed Central (a search engine over an archive of 25 million abstracts and 2 million full texts in the life sciences), as well as search engines operating on texts in other domains, such as ISHER.

Acknowledgements. The work described in this paper has been funded by the Meta-Net4U project (ICT PSP Programme, Grant Agreement: No 270893) and the JISC-funded Integrated Social History Environment for Research (ISHER)-Digging into Social Unrest project, which is part of the Digging into Data Challenge.

References

1. Zweigenbaum, P., Demner-Fushman, D., Yu, H., Cohen, K.B.: Frontiers of Biomedical Text Mining: Current Progress. Brief Bioinform. 8, 358–375 (2007)
2. Ananiadou, S., Kell, D.B., Tsujii, J.: Text Mining and its Potential Applications in Systems Biology. Trends Biotechnol. 24, 571–579 (2006)

3. Ananiadou, S., Nenadic, G.: Automatic Terminology Management in Biomedicine. In: Ananiadou, S., McNaught, J. (eds.) Text Mining for Biology and Biomedicine, pp. 67–98. Artech House, London (2006)
4. Mihăilă, C., Ohta, T., Pyysalo, S., Ananiadou, S.: BioCause: Annotating and Analysing Causality in the Biomedical Domain. BMC Bioinformatics 14, 2 (2013)
5. Kim, J., Ohta, T., Tsujii, J.: Corpus Annotation for Mining Biomedical Events from Literature. BMC Bioinformatics 9, 10 (2008)
6. 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)
7. Pyysalo, S., Ohta, T., Rak, R., Sullivan, D., Mao, C., Wang, C., Sobral, B., Tsujii, J., Ananiadou, S.: Overview of the ID, EPI and REL Tasks of BioNLP Shared Task 2011. BMC Bioinformatics 13 (suppl. 11), S2 (2012)
8. Pyysalo, S., Ohta, T., Rak, R., Sullivan, D., Mao, C., Wang, C., Sobral, B., Tsujii, J., Ananiadou, S.: Overview of the Infectious Diseases (ID) Task of BioNLP Shared Task 2011. In: BioNLP Shared Task 2011 Workshop, pp. 26–35. Association for Computational Linguistics (2011)
9. Miwa, M., Thompson, P., Ananiadou, S.: Boosting Automatic Event Extraction from the Literature using Domain Adaptation and Coreference Resolution. Bioinformatics 28(13), 1759–1765 (2012)
10. Miyao, Y., Sagae, K., Saetre, R., Matsuzaki, T., Tsujii, J.: Evaluating Contributions of Natural Language Parsers to Protein-Protein Interaction Extraction. Bioinformatics 25, 394–400 (2009)
11. Sagae, K., Tsujii, J.I.: Dependency Parsing and Domain Adaptation with LR Models and Parser Ensembles. In: Proceedings of the CoNLL 2007 Shared Task Session of EMNLP-CoNLL 2007, pp. 1044–1050. Association for Computational Linguistics (2007)
12. 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. In: Proceedings of Coling/ACL, pp. 1017–1024. Association for Computational Linguistics (2006)
13. Hara, T., Miyao, Y., Tsujii, J.: Adapting a Probabilistic Disambiguation Model of an HPSG Parser to a New Domain. In: Dale, R., Wong, K.-F., Su, J., Kwong, O.Y. (eds.) IJCNLP 2005. LNCS (LNAI), vol. 3651, pp. 199–210. Springer, Heidelberg (2005)
14. Tsuruoka, Y., Tsujii, J.: Bidirectional Inference with the Easiest-First Strategy for Tagging Sequence Data. In: Proceedings of HLT/EMNLP 2005, pp. 467–474. Association for Computational Linguistics (2005)
15. Hirohata, K., Okazaki, N., Ananiadou, S., Ishizuka, M.: Identifying Sections in Scientific Abstracts using Conditional Random Fields. In: Proceedings of the 3rd International Joint Conference on Natural Language Processing, pp. 381–388. Association for Computational Linguistics (2008)
16. Tsai, R.T., Chou, W.C., Su, Y.S., Lin, Y.C., Sung, C.L., Dai, H.J., Yeh, I.T., Ku, W., Sung, T.Y., Hsu, W.L.: BIOSMILE: a Semantic Role Labeling System for Biomedical Verbs using a Maximum-Entropy Model with Automatically Generated Template Features. BMC Bioinformatics 8, 325 (2007)
17. 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., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene Ontology: Tool for the Unification of Biology. Nature Genetics 25, 25–29 (2000)

18. Thompson, P., McNaught, J., Montemagni, S., Calzolari, N., Del Gratta, R., Lee, V., Marchi, S., Monachini, M., Pezik, P., Quochi, V., Rupp, C.J., Sasaki, Y., Venturi, G., Rebholz-Schuhmann, D., Ananiadou, S.: The BioLexicon: a Large-Scale Terminological Resource for Biomedical Text Mining. BMC Bioinformatics 12, 397 (2011)
19. Kim, J.T., Moldovan, D.I.: Acquisition of Linguistic Patterns for Knowledge-Based Information Extraction. IEEE Transactions on Knowledge and Data Engineering 7, 713–724 (1995)
20. Soderland, S.: Learning Information Extraction Rules for Semi-structured and Free Text. Machine Learning 34, 233–272 (1999)
21. Califf, M.E., Mooney, R.J.: Bottom-Up Relational Learning of Pattern Matching Rules for Information Extraction. Journal of Machine Learning Research 4, 177–210 (2003)
22. 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)
23. Pyysalo, S., Ohta, T., Miwa, M., Cho, H.-C., Tsujii, J.I., Ananiadou, S.: Event Extraction across Multiple Levels of Biological Organization. Bioinformatics 28, i575–i581 (2012)
24. Thompson, P., Iqbal, S.A., McNaught, J., Ananiadou, S.: Construction of an Annotated Corpus to Support Biomedical Information Extraction. BMC Bioinformatics 10, 349 (2009)
25. Nawaz, R., Thompson, P., McNaught, J., Ananiadou, S.: Meta-Knowledge Annotation of Bio-Events. In: Proceedings of LREC 2010, pp. 2498–2507. ELRA (2010)
26. Light, M., Qiu, X.Y., Srinivasan, P.: The Language of Bioscience: Facts, Speculations, and Statements in between. In: Proceedings of the BioLink 2004 Workshop at HLT/NAACL, pp. 17–24. Association for Computational Linguistics (2004)
27. Medlock, B., Briscoe, T.: Weakly Supervised Learning for Hedge Classification in Scientific Literature. In: Proceedings of ACL, pp. 992–999. Association for Computational Linguistics (2007)
28. Ruch, P., Boyer, C., Chichester, C., Tbahriti, I., Geissbühler, A., Fabry, P., Gobeill, J., Pillet, V., Rebholz-Schuhmann, D., Lovis, C.: Using Argumentation to Extract Key Sentences from Biomedical Abstracts. Int. J. Med. Informatics 76, 195–200 (2007)
29. McKnight, L., Srinivasan, P.: Categorization of Sentence Types in Medical Abstracts. In: Procedings of AMIA Annual Symposium, pp. 440–444. AMIA (2003)
30. Mizuta, Y., Korhonen, A., Mullen, T., Collier, N.: Zone Analysis in Biology Articles as a Basis for Information Extraction. Int. J. Med. Informatics 75, 468–487 (2006)
31. Teufel, S., Carletta, J., Moens, M.: An Annotation Scheme for Discourse-Level Argumentation in Research Articles. In: Proceedings of EACL, pp. 110–117. Association for Computational Linguistics (1999)
32. Liakata, M., Teufel, S., Siddharthan, A., Batchelor, C.: Corpora for the Conceptualisation and Zoning of Scientific Papers. In: Proceedings of LREC 2010, pp. 2054–2061. ELRA (2010)
33. 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(7), 991–1000 (2012)
34. Vincze, V., Szarvas, G., Farkas, R., Mora, G., Csirik, J.: The BioScope Corpus: Biomedical Texts Annotated for Uncertainty, Negation and their Scopes. BMC Bioinformatics 9, S9 (2008)
35. Rubin, V., Liddy, E., Kando, N.: Certainty Identification in Texts: Categorization Model and Manual Tagging Results. In: Shanahan, J.G., Qu, Y., Wiebe, J. (eds.) Computing Attitude and Affect in Text: Theory and Applications, pp. 61–76. Springer, Heidelberg (2006)

36. Hyland, K.: Talking to the Academy: Forms of Hedging in Science Research Articles. Written Communication 13, 251–281 (1996)
37. Hyland, K.: Writing without Conviction? Hedging in Science Research Articles. Applied Linguistics 17, 433–454 (1996)
38. Rizomilioti, V.: Exploring Epistemic Modality in Academic Discourse Using Corpora. In: Arnó Macià, E., Soler Cervera, A., Rueda Ramos, C. (eds.) Information Technology in Languages for Specific Purposes, pp. 53–71. Springer, New York (2006)
39. Thompson, P., Venturi, G., McNaught, J., Montemagni, S., Ananiadou, S.: Categorising Modality in Biomedical Texts. In: Proceedings of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining, pp. 27–34. ELRA (2008)
40. de Waard, A., Pander Maat, H.: Categorizing Epistemic Segment Types in Biology Research Articles. In: Proceedings of the Workshop on Linguistic and Psycholinguistic Approaches to Text Structuring, LPTS 2009 (2009)
41. Wilbur, W.J., Rzhetsky, A., Shatkay, H.: New Directions in Biomedical Text Annotations: Definitions, Guidelines and Corpus Construction. BMC Bioinformatics 7, 356 (2006)
42. Liakata, M., Thompson, P., de Waard, A., Nawaz, R., Maat, H.P., Ananiadou, S.: A Three-Way Perspective on Scientific Discourse Annotation for Knowledge Extraction. In: Proceedings of the ACL Workshop on Detecting Structure in Scholarly Discourse (DSSD), pp. 37–46. Association for Computational Linguistics (2012)
43. Thompson, P., Nawaz, R., McNaught, J., Ananiadou, S.: Enriching a Biomedical Event Corpus with Meta-knowledge Annotation. BMC Bioinformatics 12, 393 (2011)
44. Cohen, K.B., Johnson, H.L., Verspoor, K., Roeder, C., Hunter, L.E.: The Structural and Content Aspects of Abstracts versus Bodies of Full Text Journal Articles are Different. BMC Bioinformatics 11, 492 (2010)
45. Nawaz, R., Thompson, P., Ananiadou, S.: Meta-Knowledge Annotation at the Event Level: Comparison between Abstracts and Full Papers. In: Proceedings of the Third LREC Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2012), pp. 24–21. ELRA (2012)
46. Knight, J.: Negative Results: Null and void. Nature 422, 554–555 (2003)
47. Miwa, M., Thompson, P., McNaught, J., Kell, D.B., Ananiadou, S.: Extracting Semantically Enriched Events from Biomedical Literature. BMC Bioinformatics 13, 108 (2012)
48. Bjorne, J., Salakoski, T.: Generalizing Biomedical Event Extraction. In: Proceedings of the BioNLP Shared Task 2011 Workshop, pp. 183–191. Association for Computational Linguistics (2011)
49. Kilicoglu, H., Bergler, S.: Adapting a General Semantic Interpretation Approach to Biological Event Extraction. In: Proceedings of BioNLP Shared Task 2011 Workshop, pp. 173–182. Association for Computational Linguistics (2011)
50. Kilicoglu, H., Bergler, S.: Syntactic Dependency Based Heuristics for Biological Event Extraction. In: Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task, pp. 119–127. Association for Computational Linguistics (2009)
51. Nawaz, R., Thompson, P., Ananiadou, S.: Identification of Manner in Bio-Events. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012), pp. 3505–3510. ELRA (2012)
52. Nawaz, R., Thompson, P., Ananiadou, S.: Something Old, Something New: Identifying Knowledge Source in Bio-Events. In: Proceedings of CICLing 2013 (2013)
53. Nawaz, R., Thompson, P., Ananiadou, S.: Negated Bio-events: Analysis and Identification. BMC Bioinformatics 14, 14 (2013)