

ORAL PRESENTATION

Open Access

Entities, relations, events: representing biomolecular semantics

Sampo Pyysalo

From Workshop on Advances in Bio Text Mining
Ghent, Belgium. 10-11 May 2010

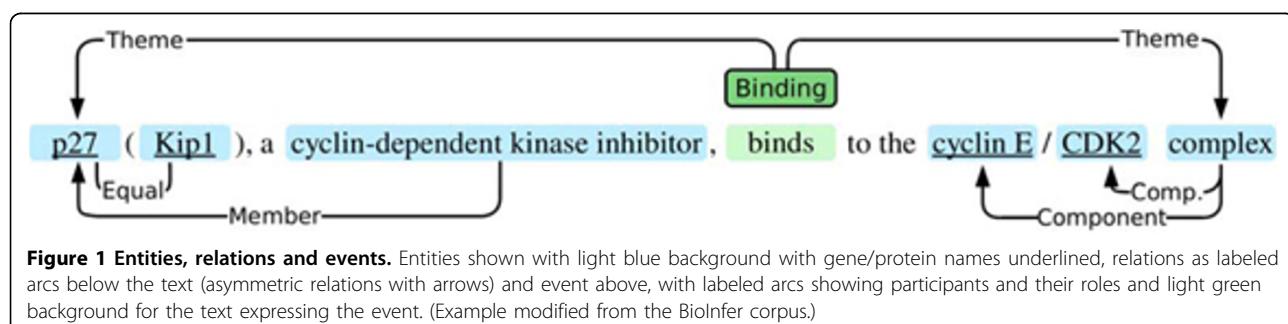
Biomedical information extraction efforts have until recently primarily focused on the detection of mentions of named entities (NEs) (e.g. genes and proteins) and the recognition of simple associations of these entities, predominantly modeled as pairwise relations. While applicable to many key tasks such as the recognition of protein-protein interactions, the limitations of the relation representation are becoming increasingly apparent in the pursuit of advanced extraction and text mining targets such as Gene Ontology annotations and metabolic and signaling pathways.

A number of recent studies have proposed more expressive alternatives to the relation representation, along with annotated resources such as the BioInfer (<http://www.it.utu.fi/BioInfer>) and GENIA Event (<http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/>) corpora. A major step toward practical systems capable of extracting such representations was taken in the BioNLP 2009 Shared Task on Event Extraction [1]. Providing annotation for gene/protein NEs as a starting point, the task centered on the extraction of an event representation that can

capture the associations of arbitrary numbers of participants in specified roles (e.g. Theme and Cause). The representation further connects events to specific statements in text and treats them as primary objects of annotation, allowing events to act as participants in other events and to be specified as being negated or stated speculatively.

Mentions of entity names (e.g. p53) serve as the basis for event extraction as they provide a connection to specific real-world entities. However, this choice implies some approximations in representation: statements involving, for example, complex of c-Rel and p50 are modeled as events with the NEs (c-Rel and p50) as participants. Marking either a non-specific term such as complex or the entire phrase as a participant can capture more context, but also opens a new question for automatic processing: what do events involving such entities imply for the NEs that connect the representation to reality?

Pairwise relations specifying how NEs are associated with terms in their context provide one possible answer.



Correspondence: smp@is.s.u-tokyo.ac.jp
Department of Computer Science, University of Tokyo, Japan

A small set of basic relation types with well-defined semantics such as object-component (for e.g. complex-subunit associations) and collection-member (for family-protein) can characterize many NE-term associations and provide specific meaning to general terms [2]. Re-introducing pairwise relations in this role suggests a detailed representation where both NEs and general terms are marked as entities, relations connect the two, and events model statements of change involving the entities, with specific NEs and terms originally stated as participants (Figure 1).

Whether the detail afforded by such a model is of sufficient practical value to outweigh the challenges in its automatic extraction remains an interesting question for future study.

Published: 20 September 2010

References

1. Kim JD, Ohta T, Pyysalo S, Kano Y, Tsujii J: Overview of BioNLP'09 Shared Task on Event Extraction. *Proceedings of the BioNLP 2009 Shared Task, Boulder, Colorado* 2009, 1-9 [<http://www.aclweb.org/anthology/W09-1401>].
2. Pyysalo S, Ohta T, Kim JD, Tsujii J: Static Relations: a Piece in the Biomedical Information Extraction Puzzle. *Proceedings of Natural Language Processing in Biomedicine (BioNLP) NAACL 2009 Workshop* Boulder, Colorado: Association for Computational Linguistics 2009, 1-9.

doi:10.1186/1471-2105-11-S5-O6

Cite this article as: Pyysalo: Entities, relations, events: representing biomolecular semantics. *BMC Bioinformatics* 2010 11(Suppl 5):O6.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

