LREC 2008 Workshop

Building and evaluating resources for biomedical text mining

26 May 2008
Marrakech, Morocco

There has been tremendous work in biomedical text mining over the last decade. The size and coverage of the available literature and demands for text mining applications in the domains of biology and biomedicine are constantly increasing. These domains have become one of the driving application areas for the natural language processing community, resulting in a series of workshops and conferences that have reported on the progress in the field. Most of the work has focused on solving specific problems, often using task-tailored and private data sets. This data is rarely reused, in particular outside the efforts of the providers. This has changed during the last years, as many research groups have made available resources that have been built either purposely or as by-products of research or evaluation efforts. A number of projects, initiatives and organisations have been dedicated to building and providing biomedical text mining resources (e.g. GENIA, PennBioIE, TREC Genomics track, BioCreative, Yapex, LLL05, BOOTStrep, JNLPBA, KDD data, Medstract, BioText, etc.). Although several resources have been provided for and from the community to support both training and evaluation of text mining applications, there have been few efforts to provide community-wide discussions on design, availability and interoperability of resources for bio-text mining.

This workshop will focus on assessment of the current state-of-the-art in building and evaluating resources for biomedical text mining, in particular on guidelines and annotation schemes, formats and availability. A particular focus will be on open access resources. The workshop will address building and designing lexical and knowledge repositories (controlled vocabularies, terminologies, ontologies, factual databases) and annotated corpora, as well as their evaluation and usability.

We invite papers reporting on biomedical resources specifically used to facilitate biomedical text mining and the process of designing, building, updating, delivering, evaluating and disseminating them. A focus of the workshop is on lexical and knowledge repositories and annotated corpora. A further focus is on design guidelines, standards for building resources, storage and exchange format, interoperability of resources and last, on exploring new directions for their dissemination.

Topics of interest include but are not limited to:

- Building biomedical resources: controlled vocabularies, terminologies, ontologies, corpora
- Guidelines and annotation schemas, challenges, interoperability
- Building task-specific resources
- Reengineering existing biomedical or general language resources
- Augmentation of resources with biomedical features
- Update and evolution of resources
- Lightly annotated and noisy resources
- Tools for exploration of resources
- Data exchange formats
- Standards for building resources

- Documenting and disseminating resources
- Evaluation of resources

Submissions:

Submissions must describe original, completed or in progress, and unpublished work. Your paper (up to 8 pages) should be formatted according to the stylesheets provided by LREC 2008. Please upload your electronic submissions in PDF format using EasyChair at http://www.easychair.org/conferences/?conf=berbmtm08.

Each submission will be reviewed by three program committee members. Paper review will be double blind, so papers should not include authors' names and affiliations. Self-references are to be avoided — instead of "As we showed in Smith et al. 1999...", use "As Smith et al. 1999 showed....".

Accepted papers will be published in the workshop proceedings, and may be presented either as a poster or an oral presentation. Selected expanded papers will be published in a special issue of the Language Resources and Evaluation (LRE) journal.

Important dates:

- February 24, 2008   Paper submissions due (before 23:00 GMT)
- March 28, 2008   Notification of acceptance
- April 11, 2008   Camera-ready papers due
- May 26, 2008   Workshop

Venue:

The workshop, sponsored by NaCTeM, will be held in conjunction with the LREC 2008 conference in Palais des Congres Mansour Eddahbi, in Marrakech, Morocco.

Organisers:

- Sophia Ananiadou, National Centre for Text Mining, University of Manchester, UK
- Monica Monachini, Istituto di Linguistica Computazionale, Pisa, Italy
- Goran Nenadic, University of Manchester, UK
- Jian Su, Institute for Infocomm Research, Singapore

Program committee members:

- Olivier Bodenreider, NLM, USA
- Paul Buitelaar, DFKI, Germany
- Nicoletta Calzolari, CNR, Italy
- Kevin B. Cohen, MITRE, USA
- Nigel Collier, National Institute for Informatics, Japan
- Walter Daelemans, University of Antwerp, Belgium
- Beatrice Daille, University of Nantes, France
- Udo Hahn, Jena University, Germany
- Marti Hearst, Berkeley, USA

- Martin Krallinger, Protein Design group, Spain
- Ewan Klein, Edinburgh University, UK
- Mark Liberman, CIS, UPenn, USA
- Hong Fang Liu, Georgetown University Medical Center, USA
- John McNaught, University of Manchester, UK
- Simonetta Montemagni, CNR, Italy
- Adeline Nazarenko, LIPN, Paris 13, France
- Claire Nedellec, CNRS, Framce
- John Pestian, Computational Medicine Center, Cincinnati Children's, USA
- Dietrich Rebholz-Schuhmann, EMBL-EBI, UK
- Patrick Ruch, University Hospital of Geneva and Swiss Federal Institute of Technology
- Guergana Savova, Mayo Clinic, USA
- Hagit Shatkay, Queen's University, USA
- Stefan Schulz, Freiburg University Hospital, Germany
- Jun-ichi Tsujii, University of Tokyo, Japan and University of Manchester, UK
- Yoshimasa Tsuruoka, University of Manchester, UK
- Karin Verspoor, Los Alamos National Labs, USA
- Pierre Zweigenbaum, LIMSI-CNRS, France

Workshop contact person:

Sophia.Ananiadou@manchester.ac.uk, National Centre for Text Mining, Computer Science, University of Manchester