BOOTStrep (FP6 - 028099) Project Synopsis * www.bootstrep.eu







BOOTStrep (Bootstrapping Of Ontologies and Terminologies STrategic REsearch Project) is pulling together already existing biological fact databases as well as various terminological repositories and is implementing a text analysis system which continuously increases its coverage by analysing biological documents.

Impact

Biological knowledge, up until now, is scattered in heterogeneous database formats and locked in unstructured natural language documents. The intended integration of biological knowledge in a homogeneous conceptual framework will ease access to this fragmented knowledge and substantially increase its usability for R & D purposes, e.g., in the European bio-tech and pharmaceutical industry.

BOOTStrep's main innovations

Knowledge integration and reuse in the biology domain are the main goals of the BOOTStrep project. The resources and text mining tools developed within the project are expected to boost the performance in various bio application tasks. In particular, BOOTStrep aims at

- exploiting already existing terminological resources (thesauri, classification systems, etc.) and combining them within a common, standardized conceptual representation framework,. Based on this domain-specific background knowledge advanced natural language technologies are employed for the analysis of biological documents in order to fill conceptual gaps in these resources by automatically acquiring new terms, concepts and relations,
- creating, incrementally maintaining and continuously updating a repository
 of biological facts based on employing a comprehensive bio-lexicon and a
 standards-based formal bio-ontology for text analysis. Facts are extracted
 from biological documents in a fully automatic way, they are subsequently
 filtered and validated for novelty, redundancy, contradiction, etc.,
- developing resources and resource-building NLP tools for text-based knowledge harvesting in order to support information extraction and text mining in the biology domain,
- allowing multilingual public access to continuously updated and validated biological fact repositories.

More details

- Project website: www.bootstrep.eu
- Press releases:
 - o http://idw-online.de/pages/de/news144153 (German)
 - http://idw-online.de/pages/de/news153353 (German)
 - http://supreme.coling.unijena.de/BootStrep/pub/Extern/PressPage/20060323_BOOTStrep_kick -off_press_release_English.pdf (English)

BOOTStrep's upcoming work

The project officially started in April 2006. Within the first year, the work will be separated into tool- and resource-oriented tasks. On the one hand, BOOTStrep will provide a software infrastructure needed for text analysis proper and knowledge capture from texts. On the other hand, an inventory of terms will be generated from a variety of existing resources, which may serve as links between facts from the unstructured literature and corresponding entries in the structured biomedical databases. Furthermore, the representational model for the terminological lexicon (Bio-Lexicon) and the ontological repository (Bio-Ontology) will have to be designed and populated.

Administrative details

BOOTStrep (FP6 - 028099) is a Specific Targeted Research Project (STREP) of the European Union's 6th Framework Programme, Thematic Priority 2 (Information Society Technologies) within the fourth call of the programme. It addresses the strategic objective "Semantic-based Knowledge and Content Systems".

BOOTStrep started on April 1, 2006 and will end on March 31, 2009.

The overall budget is 3.6 million euro.

Six partners from four European countries (Germany, U.K., Italy, France) and one Asian partner from Singapore are involved in the project.

List of participants

- Friedrich-Schiller Universität Jena (FSU), Germany (Project coordinator)
- European Molecular Biology Institute / European Biology Informatics (EMBL-EBI), UK
- Consiglio Nazionale delle Ricerche Istituto di Linguistica Computazionale (CNR-ILC), Italy
- University of Manchester (UoM), UK
- Université de Rennes (UR1), France
- Institute for Infocomm Research (I2R), Singapore
- Universitätsklinikum Freiburg (UKLFR), Germany

Events in connection with BOOTStrep

• BOOTStrep Kick-Off, April 12-13, 2006, Jena

Following the 2nd International Symposium on Semantic Mining in Biomedicine (SMBM 2006), the opening event of the European Union (EU) funded project BOOTStrep (Bootstrapping Of Ontologies and Terminologies STrategic REsearch Project) takes place April 12-13, 2006 at Friedrich-Schiller-University Jena.

Scientists from England, Italy, France, Singapore and Germany are meeting for the official opening event under the lead of Prof. Hahn, the chair of computational linguistics at Friedrich-Schiller-University for two years now. Primary goal of the kick-off event is, besides the intensification of the personal contacts, the development of a common view on the project, and the agreement on the first working steps.

2nd International Symposium on Semantic Mining in Biomedicine, April 9-12, 2006, Jena

The 2nd International Symposium on Semantic Mining in Biomedicine (SMBM) is organised by the EU Network of Excellence Semantic Mining and the Jena University Language & Information Engineeing (JULIE) Lab. For additional information visit:

http://supreme.coling.uni-jena.de/content/blogcategory/32/103.

• Workshop on Text Mining, Ontologies and NLP in Biomedicine, March 20-21, 2006, Manchester

Several distinct but closely related technological fields such as Natural language Processing and Text Mining, Intelligent Information Retrieval, Ontology Engineering and Knowledge Management, Knowledge Discovery, etc. have shown remarkable progress in recent years, which will bring revolution into the ways of conducting science. In particular, due to its rapid expansion and development as well as its data/knowledge intensive nature, life science and biomedical fields have been and will be at the centre of such methodological revolution. We invited researchers in Information Technology who had been successfully applying their research results to life science and those in biomedical fields who have been developing new Information-based approaches in biomedical fields. We are pleased to say that the aim of the workshop, i.e. bringing together researchers from different areas allowing exchange of ideas was successfully achieved and that we enjoyed alive and constructive discussion at the workshop. For additional information visit: http://www-tsujii.is.s.u-tokyo.ac.jp/jw-tmnlpo/.

Contact persons

Prof. Dr. Udo Hahn (Project Leader)

Institut für Germanistische Sprachwissenschaft der Universität Jena

Fürstengraben 30, 07743 Jena, GERMANY

Tel.: +49.3641.944.320 Fax: +49.3641.944.321

E-Mail: udo.hahn[at]uni-jena.de

Lars Zapf (Project Manager)

Institut für Germanistische Sprachwissenschaft der Universität Jena

Fürstengraben 30, 07743 Jena, GERMANY

Tel.: +49.3641.944.303 Fax: +49.3641.944.321

E-Mail: lars.zapf[at]uni-jena.de

BOOTStrep Project Synopsis last updated: 08//05/2006

BOOTStrep (FP6 - 028099) Project Synopsis * www.bootstrep.eu