

SemanticMining

Semantic Interoperability and Data Mining in Biomedicine

The main concern of **SemanticMining** is semantic interoperability, which simply means that semantics is preserved in communication between information systems, a condition which should be natural but has proven to be very hard to achieve, especially so in the complex application area of health care.

Objectives of the project

Traditionally academic departments in the domain of biomedical informatics have their roots either in computer science, system engineering or in a medical or clinical context. **SemanticMining** is composed of partners from these scientific areas including public health care organisations, all bringing their experience and in-depths knowledge together into a common framework.

The general objective of **SemanticMining** as a Network of Excellence (NoE) is to bridge gaps in the European research infrastructure and to facilitate cross-fertilisation between scientific disciplines.

A main concern of **SemanticMining** is semantic interoperability in communication between health care information systems. A key component in support of interoperability is ontologies by which information is semantically well defined.

The long-term goal of **SemanticMining** will be the development of generic methods and tools supporting the critical tasks of the field: data mining, knowledge discovery, knowledge representation, abstraction and indexing of information, semanticbased information retrieval in a complex and highdimensional information space.

Project Description

Researchers in the network play an influential role in the process of harmonisation and further development of ontologies and terminology systems. Examples of areas of interaction are the Gene Ontology, the Foundational Model of Anatomy, SNO-MED CT and the CNPU coding system.

The research carried out also address the need for approaches in Europe which will bridge language barriers and facilitate access for non- English native persons to the large scientific corpus of texts written in English.

The research activities in **SemanticMining** are focused around the following areas:

- principles in ontology engineering
- evaluation of SNOMED CT
- terminology systems in laboratory medicine
- impact of ontologies on health statistics
- the construction of a multi-lingual medical dictionary
- information retrieval in bioinformatics
- the semantic-based electronic health record

“The NoE has active interaction with standardisation bodies such as CEN TC251 and HL7”

Scenario

An important application area is the development of the semantically well defined electronic health record (EHR). The challenge is to define interoperable means of specifying classes of data within the EHR with sufficient granularity and precision that clinical applications, decision support systems and other tools can create or retrieve data values or sets of patients precisely matching any given clinical criterion. Research is ongoing on the boundary problem between information models of a generic EHR and terminology models used for populating them, and prototypes of the openEHR Archetype model are being developed.

Expected Results & Impacts

It is well known that the health care system is facing a series of challenges concerning quality and costeffectiveness. The distribution of health care services in ways which allow the patient to take an active part in relevant decisions and the provision of evidence-based medicine at all levels in the system and the effective use and reuse of information are all key issues for the organisation of health care delivery in Europe.

The information and communication technology infrastructure should reflect a view of the health care system as a seamless system where information can flow under the necessary forms of regulation, across organisational and professional – and national – borders.

The need for cross-referencing between biological and clinical information provides another grand challenge. The vast amount of data available in bioinformatics databases together with the growing volume of electronically available clinical information call for automated (or at least semi-automated) methods for high-quality indexing, annotation, and cross-referencing through discovery of patterns and relationships. Thus there is a need for harmonisation and resources for the integration of data derived from divergent sources of the sort which ontology can provide.

Health and health care are not only important for each individual but also important indicators of the state of a society. Therefore statistics about health are an important part of the information system. Issues in focus are the scope of health and health care statistics, the tools used for coding and classification, as well as problems of quality and comparability of data. A basic research question is how the move from traditional classifications to reference terminologies may improve the quality of health statistics. While several coding systems are utilised in health care domains such as diagnoses, health problems, and interventions, the challenge is to allow aggregation according to different aspects and to assure high information quality on all levels of data abstraction.

Keywords:

biomedical informatics,
electronic health records,
semantic interoperability,
ontologies, data mining



Semantic Mining Semantic Interoperability and Data Mining in Biomedicine

Project co-ordinator:

Hans Åhlfeldt (Linköping University, SE)

Contact person:

Hans Åhlfeldt

Tel: +46 13 227574

Fax: +46 13 101902

Email: hans.ahlfeldt@imt.liu.se

Website: www.semanticmining.org

Partners:

- Albert-Ludwigs-Universitaet Freiburg (DE)
- Christian-Albrechts-Universitaet Zu Kiel (DE)
- Consiglio Nazionale Delle Ricerche (IT)
- Egeszseguyi Strategiai Kutatointezet (HU)
- European Dynamics Advanced Systems Of Telecommunications Informatics And Telematics S.A. (GR)
- European Molecular Biology Laboratory
- Friedrich-Schiller-Universitaet Jena (DE)
- Goeteborgs Universitet (SE)
- Institut National De La Santé Et De La Recherche Médicale (FR)
- Karolinska Institutet (SE)
- Kith As - Kompetansesenter For Informasjonsteknologi I Helsevesenet As (NO)
- Linkoepings Universitet (SE)
- Merrall-Ross International Limited
- National Research And Development Centre For Welfare And Health (FI)
- Socialstyrelsen (SE)
- Sundhedsstyrelsen (DK)
- The Open University (UK)
- The Victoria University Of Manchester (UK)
- Universitaet Des Saarlandes (DE)
- Universitaet Leipzig (DE)
- Universitaetsklinikum Freiburg (DE)
- Universite De Geneve (CH)
- University College London (UK)
- University Of Brighton (UK)
- Uppsala Universitet (SE)
- Vaestra Goetalands Laens Landsting (SE)

Timetable: from 01/04 – to 12/06

Total cost: € 6,384,000

EC funding: € 5,000,000

Instrument: NoE

Project Identifier:

IST-2002-507505