

SIABO

Semantic Information Access through Biomedical Ontologies

Troels Andreasen, Henrik Bulskov, Tine Lassen, **Sine Zambach**
CBIT, Roskilde University, Denmark

Per Anker Jensen, **Bodil Nistrup Madsen**, Hanne Erdman Thomsen
ISV, Copenhagen Business School, Denmark

Jørgen Fischer Nilsson, Bartłomiej Antoni Szymczak
DTU Informatics, Technical University of Denmark

Industrial partner

- The healthcare company Novo Nordisk A/S
- Novo Nordisk needs access to the contents of the increasing amount of documentation
- Only a semantics-based approach to information management is adequate to that task

Aim of the SIABO project

- Provide an approach to representing, organizing, and accessing the conceptual content of biomedical texts using a formal ontology

The meaning content of each document is described as a set of arbitrarily complex conceptual feature structures facilitating detailed comparison of the content of documents.

Outline of our presentation

- Ontological semantics
- Generative Ontology
- Domain Ontologies
- Extraction of concepts from text
- Querying information and knowledge

- Contact e-mail: siabo@ruc.dk

Ontological semantics

- Text chunks with identical meaning but different linguistic forms are to be mapped to the same node in an ontology
- Facilitates identification of paraphrases, concept relations and measurement of distances between key concepts in texts
- Introduces the notion of generative ontologies, i.e., infinite ontologies providing increasingly specialized concepts

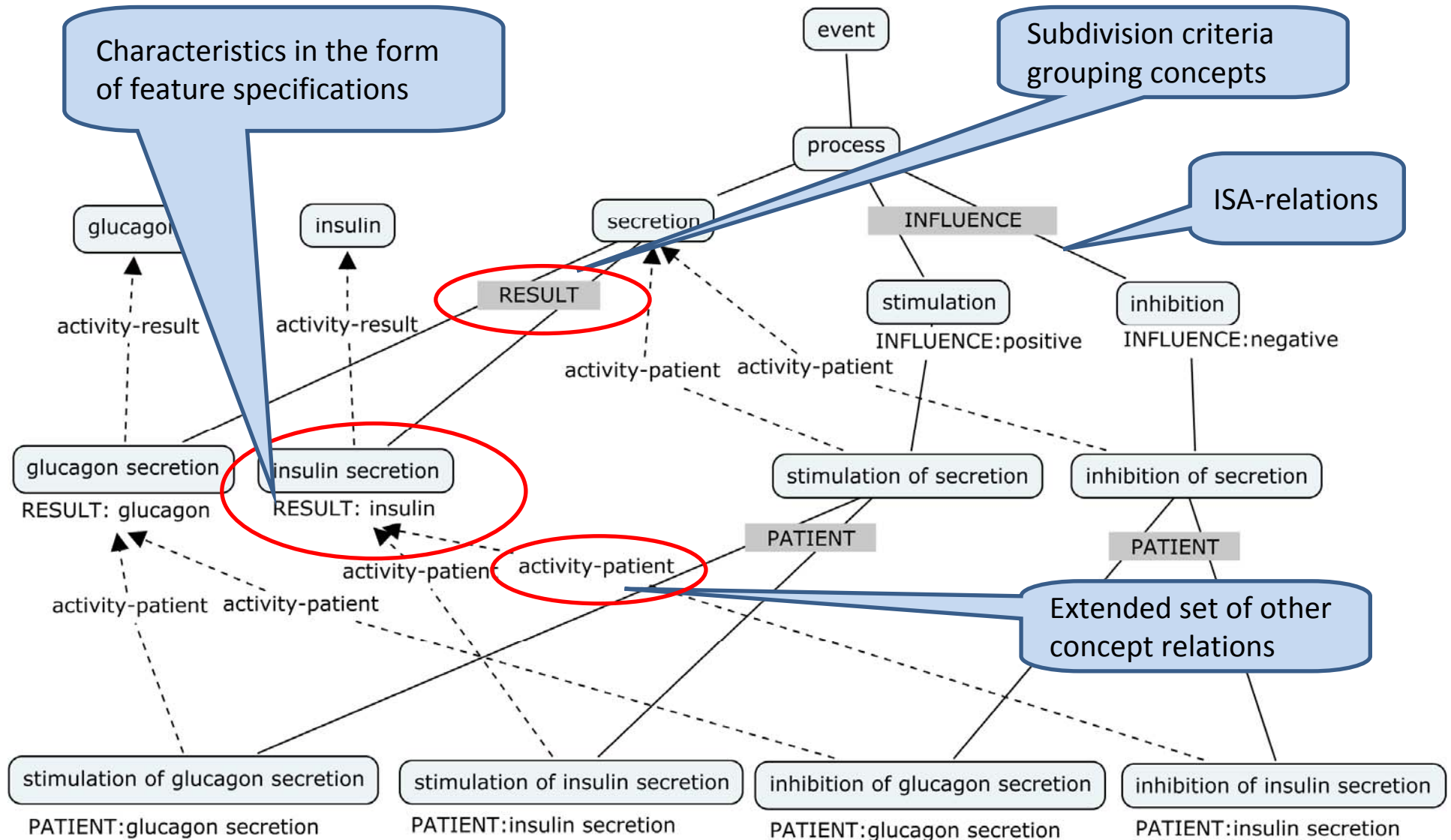
Focus

- Ontological engineering of biomedical ontologies applying lattices and relation-algebras
- Clear affinities to contemporary research in the Semantic Web area, description logic as well as XML approaches

The need for domain ontologies

- A domain ontology is a validated fragment of the generative ontology
- Domain ontologies supplement and refine already existing ontologies for the domain, such as UMLS
- UMLS is not specific enough as regards concepts and concept relations, and in many cases the existing resources are imprecise
- Validated domain ontologies are based on
 - principles of terminological ontologies
 - a text corpus
 - knowledge of domain experts

Extract of the domain ontology of insulin production



Extracting concepts from text

- Purpose is semantic information retrieval by:
 1. Identifying semantic knowledge and map it into the generative ontology
 2. Semantic annotation/indexing for both search corpus and query

Two different approaches are presented:

- Pattern based approach
- Synthesis approach

Pattern based approach

- Accessing semantic roles through lexical resources: NOMLEX, VerbNet and WordNet
- Automatic extraction of semantic roles using a syntactic analysis
- Example:
 - np-1 *inhibits* np+1 - > inhibition[AGT:np-1, PNT:np+1]

Ontograbber: Synthesis approach

- The generative ontology provides potentially infinitely many admissible concepts
- That is, feature structures which are constrained by so-called ontological affinities
- The Ontograbber matches these recursive structures against phrases in the text
- Assisted by linguistic grammar rules
- Example:
 - Substance *inhibits* Process -> inhibition[AGT: Substance, PTN:Process]

Querying Information and knowledge

- The prototype utilizes the semantically indexed corpus
- Analyses the query
- Retrieves documents that match the semantics of the query
- Example
 - Query: Insulin blocks glucose secretion
 - Analysis: inhibition[AGT:insulin,PNT:secretion[RES:glucose]]
 - Matches in text:
 - (...) inhibition of glucose release by insulin (...)*
 - (...) insulin inhibits glucose secretion (...)*

www.SIABO.dk

