Language Technology

Within the Human Language Technology Group we develop efficient and resource lean Natural Language Processing (NLP) methods, resources and tools for information access and refinement using Language Technology for very large text sets, with a special interest in the medical domain. Our main focus is the Swedish language, but we also work with English and other languages. Research areas include information extraction, text summarisation, text generation, semantic modelling as well as question and message answering, information retrieval and healthcare informatics.

Developing efficient Natural Language Processing

Language technology is an interdisciplinary research field that draws its main bulk of methods and techniques from computer science and linguistics, but also from diverse fields such as statistics, mathematics, pedagogics and psychology. The applications can be found at the core of many information systems as well as systems for computer mediated communication and learning.

Automatic Text Summarisation
With the abundance of digitally stored information, it has now become near impossible to manually search. The information must by some means be filtered and extracted in order to avoid drowning in it.

Computer Assisted Language Learning (CALL)
is a broad field, and Language technology is playing an increasingly important part in this.

Cross-Lingual Information Retrieval
These techniques allow people to write queries in one language and retrieve relevant documents written in another language.

Email and Short Message Answering
While question answering systems answer single-sentence questions, email answering, or short message answering, delivers answers to messages that contain several sentences. Just like our e-mails. The task of email answering can be considered similar to the task of document classification.

Language Modelling and Lexical Semantics
Distributional patterns in corpora can be exploited to build mathematical models of language use that contain information about the relative meaning of linguistic units, typically words.

Natural Language Processing of Health Records
The information in medical records can be used to assist the clinician during her daily work when reading and writing records about the patient, for example to obtain an overview of the patient record by automatic text summarisation.

Semantic Information Extraction
For some information extraction needs, accurate, relevant and situation-specific information extraction is crucial. This involves, e.g., distinguishing factual information from speculative or negated information.

Web Mining
is the application of data and text mining techniques to Web related data with the aim to discover structural, usage and content related patterns.

Contacts

Hercules Dalianis, professor
Martin Duneld, Ph.D.
Ola Knutsson, Ph.D.
Erik Sneiders, Ph.D.
Sumithra Velupillai, Ph.D.

Please consult the weblink for contact information.

Ongoing projects

Automated translation of radiology reports into general Swedish

Nordic Center of Excellence-The Nordic Information for Action eScience Center: creating interoperability of nordic cancer biobanks using data- and text mining

Detect-HAI - Detection of Hospital Acquired Infections through language technology analysis of electronic patient records.

High-performance DAta mining for Drug Effect detection

Increasing automation of request management for public authorities by language technology

Interlock: Stockholm - San Diego - Inter-Language collaboration in clinical NLP

Department of Computer and Systems Science, DSV

www.dsv.su.se/language

Stockholm University