The Group Natural and Formal Languages

Projects and Research Directions

Frank Drewes
with input and help by “the group”

CS Research Day
The Information Age is a Language Age

Twitter
scientific journals
legal documents
product reviews
Wikipedia
Facebook
Spam
Commercials

Too much for manual processing!
Too valuable for being disregarded!
Natural and Formal Languages

- semiotics
- context
- ambiguity
- mixture
- text
- change
- mixture
- vagueness
- speech
- syntax
- semantics
- pragmatics
- transduction
- automaton
- precision
- algorithm
- learning
- complexity
- relevance
- cleaning
- summarizing
- translation
- understanding
- interfaces
- classification
- learning
We focus on methods that exploit structure and have a sound theoretical basis.
Some projects...
• Natural language analysis usually starts with a syntactic analysis.
  • There are many different grammar formalisms for doing this.
  • Efficient parsing is central, but not entirely understood.

What are the parameters that enable efficient syntactic analysis? Is there a trade-off between efficiency and descriptional power?
we the children Hans the house let help paint
ACC DAT
"we let the children help Hans paint the house"
Text Classification

Financing: VR project grant
    UmU Young Researcher Award
    FOI & Industrial Doctoral School
    \{ all J. Björklund

People: J. Björklund, N. Zechner, N.N.

The amount of text makes automatic classification a central issue.

• Do I have an unknown aunt in Nigeria or is it spam?
  • How is our product viewed in social media?
  • What is this text about?
  • Who wrote this text?

Can structure help to answer theses questions?
Niklas Zechner <zechner@cs.umu.se> 20 november 2012 15:08
Till: Martin Berglund
Kopia: Johanna Björklund, Suna Bensch, Frank Drewes, Henrik Björklund
Re: EU application - progress beyond zero

It looks good to me. If I was a component-based text analysis tool, I would notice the following:

- it is written in a very formal style, and looks a lot like an EU application

- parser has some difficulties with phrases like "With all of the above work packages not only will the individually interesting and relevant research result"

- we can identify the author with 90% certainty, from phrases like "to a great part" and "a fair bit"
Spacebook: Spoken Guidance for Pedestrians

Financing: FP7 STREP project (M. Minock)

People: M. Minock, J. Mollevik
(L. Edblom, S. Hegner, M. Karlsson, L. Närva, M. Åsander)

Partners: 6 universities and one company

Why should I wish to stare at my iPhone when I visit another city?

- I want to be told how to find my way.
  - I want to ask “How old is this bridge?”
  - I want reasonable answers.

Speech-driven hands and eyes free support for pedestrians navigating and exploring the city.
This video shows trials of the SpaceBook Long Thin System. The main aim was to demonstrate the integration of modules from different partners. This system was evaluated in Edinburgh, Scotland.

http://www.youtube.com/watch?v=qQZnwrOyeTE
Multi-Dimensional Language Analysis

Financing: VINNOVA VINNMER grant (S. Bensch)
People: S. Bensch, H. Björklund, F. Drewes

- Natural language does not only have the syntactic dimension.
  - Others are phonetics, semantics, pragmatics, ...
    - Modern linguistic theories consider them as interlinked.
  - We want to develop formalisms that capture this.
    - This could enable better analyses.

We call our formalism **Millstream systems**.
Mary loves Peter.
Further Areas of Interest

- Shuffled languages
  - Extended formalisms for regular languages
    - Efficient evaluation of XML queries
      - Extensions of context-free graph grammars

...and, of course, Berglund’s conjecture!
For those interested...

We will organize 2 international conferences here in Umeå:

Thank very for listening much you!

[Swiss English]