Recognizing Textual Entailment

Carlos Areces Claire Gardent

{areces,gardent}@loria.fr

Langue et Dialogue, LORIA
Nancy, France

7th February 2007

Outline

- Natural Language Processing (NLP) at LORIA
- Textual entailment recognition
  - Associating text and meaning
  - Acquiring knowledge
  - Reasoning
- The Erasmus Mundus LCT Master
- Other Topics of Interest

NLP at LORIA

- **Langue et Dialogue, Calligramme:** Computational Linguistics, Computational Semantics
- **Parole:** Speech, Statistical Computational Linguistics
- **Orpailleur:** Semantic web, Ontologies

Computational Semantics

Associating Text and Meaning

- Text ⇒ Meaning (Semantic Construction)
- Meaning ⇒ Text (Surface Realisation)

NLP and Reasoning

- Textual Entailment Recognition (Reasoning and Analysis)
- Generating natural sounding text (Reasoning and Generation)

Recognising Textual Entailment

T: The debate lasted many hours, but after much bitter discussion the Labour Party decided to back Tony Blair's policy on Iraq.

H: The Labour Party backed Tony Blair's Iraq policy.

Does T textually entails H ?

The Bricks

What do we need to recognize textual entailment ?

- Semantic Construction: Text $\Rightarrow$ Meaning Representation
- Knowledge: Lexical semantics, Ontologies
- Reasoners: Theorem provers NLP, Model builders, specialised reasoners
- Evaluation Data: Test suites, Annotated corpora

Semantic Construction

- Based on a Tree Adjoining Grammar
  - produced semi-automatically by compilation from a factorised grammar description
    http://sourcesup.cru.fr/xmg
  - used also in generation (surface realisation) mode
    http://trac.loria.fr/~geni
- Tabular TAG parser
  - http://dyalog.gforge.inria.fr
- Semantic construction module in Prolog
  - http://trac.loria.fr/~semconst

Going Further

- Scaling up
  - Extending the grammar and the lexicons
  - Improving the grammar: error mining
  - Improving parsing efficiency: parsing with factorised trees
- Testing the semantic representations used
  - Disambiguation (statistical or symbolic)
  - Using generation: mesure (and reduce) overgeneration
  - Using (automatically generated) test suites
- Creating and testing semantic representations with different levels of granularity (Predicate/Argument structures, Description Logic, FOL, DRS, etc.)

Knowledge Acquisition

To reason about the meaning of texts, we need Knowledge about

- Word meanings: e.g., definitional, synonymic, hyperonymic, antonymic, etc.
  - Extract this information from general and synonym dictionnaries
  - VerbNet, FrameNet for French
- Concepts: i.e., ontologies and relations between concepts and words

Inference for NLP

- Once we have designed grammars with a semantic dimension we can obtain semantic representations for texts.
- Once we have adquired knowledge we have the the background information in terms of which the semantics should be interpreted.
- But how do we put them together? How do we 'query' this information?
- Intuitively: Inference is the tool that let us explore the combination of semantic representation and background knowledge.

Inference for NLP: Levels of Representation

- We take a very broad definition of inference:
  - Statistical Inference (e.g., a probabilistic model)
  - Decidable Inference in weak languages (e.g., satisfiability of Propositional Logic)
  - Decidable Inference in expressive languages (e.g., knowledgebase consistency in Description Logics)
  - Undecidable, but Noninteractive Inference (e.g., theorem proving for First Order Logic)
  - Undecidable and Interactive Inference (e.g., interactive theorem proving in Higher Order Logics)
- Which do we use for each NLP task?

Zooming In: Ontologies and Description Logics

- Description Logics (DLs) are formal languages specially tailores for knowledge representation (e.g., ontologies)
- Members of the group work on theoretical and practical aspects of inference in DLs and similar languages.
- Our interest is twofold:
  - Explore how to make the best use of DL inference tools in NLP applications (representations, complex inference task, etc.)
  - Extend the expressive power of these tools without loosing their good computational behaviour (new logica operators, integration with other reasoning tools, etc.).

Erasmus Mundus LCT Masters

- New EM Language and Communication Technologies Masters approved by the Erasmsu Mundus Consortium.
- Partners: Nancy 2, University of Saarbrücken, Charles University of Prague, Groningen University, University of Bolzano, University of Malta.
- Set Up: One year of study in two Universities from the consortium to obtain a double degree.
- Starts: Academic year 2007/2008.
- 150 applications within 15 days of registration opening.


European Masters Program
Language & Communication Technologies
Education and Culture
Erasmus Mundus
European Masters Program in Language and Communication Technologies (LCT)
The Erasmus Mundus European Masters Program in Language and Communication Technologies (LCT) is designed to meet the demands of industry and research in a rapidly growing area. It offers education and training opportunities for the next generations of leaders in research and innovation. It provides students with profound knowledge and insight into the various disciplines that contribute to the methods of language and communication technologies and it strengthens their ability to work according to scientific methods. Moreover, the students acquire practice-oriented knowledge by choosing appropriate combinations of modules in Language Technology, Computational and Theoretical Linguistics, and Computer Science.
It involves studying one year at one of the Universities of the consortium, and completing the second year with a stay in one of the partner Universities. After this, the student will obtain two already approved Master of Science/Arts degrees with legal value in the countries of issue.
Duration: 2 years
Language of instruction: English
Summary of Study Programme: The course consists of Compulsory Modules and Advanced Modules in Language Technology and Computer Science, complemented by a Project, and a Masters thesis, for a total of 120 ECTS credits.
Last modified: 2006-12-22
Search
Contents
Home
LCT Intro
LCT Consortium
Programme Outline
Application Procedure
Financial Information
Teaching Body
Strategic Advisory Board
Endorsed by
Contact Information

Erasmus Mundus Masters

- Main Aim: Make European masters an alternative worldwide.
- Aprox. 20 full grants for non European students provided by the EU.
- Support for exchange of European students and lecturers between Erasmus Mundus Consortia and non European Universities.
- Encourages interaction and integration between the partners of an approved Erasmus Mundus consortium.

Summing Up

- Semantic Construction
  - Lexicon Acquisition; Grammar Debugging; Parsing and Disambiguation; ...
- Knowledge Adquisition
  - {Word, Verb, Frame}-Net for French; Mapping between Concepts and Words; ...
- Inference for NLP
  - Which level of representation? How computationally expensive? ...
- EM Language and Communication Technology Masters
  - Student and scholars exchange; Curricula integration; Double degree (as a first step towards a join degree?); ...
- Other Topics of Interest
  - Generation; Dialogue; Other NLP Teams at LORIA (Calligramme, Orpailleur, Parole)...