Computational Linguistics

CS579: Fall Semester 2012

Department of Computer Science
Korea Advanced Institute of Science and Technology

Jong C. Park

Course Objective (1/2)

- Computational linguistics is an interdisciplinary field, involving linguistics, computer science, mathematics, logic, cognitive science, and cognitive psychology.
- This course explores the theoretical aspects of computational linguistics, focusing on its subfield known as computational semantics.

Course Objective (2/2)

- We examine issues of representation and inference, so that they will be used to model human competence in language use.
- Through the course, the students will be able to understand the core theoretical aspects of computational linguistics and use them to strengthen the formal base of any applications that involve the processing of human language.

Administrative Details (1/2)

- Instructor
  - Prof. Jong C. Park (박종철)
  - email: park@cs.kaist.ac.kr
  - Office: Room 2406, CS Bldg.
  - Phone: x3541
- Teaching Assistants
  - Yong-Jae Lee, Jaehyun Han
  - Email: cs579@nlp.kaist.ac.kr
  - Phone: x7741

Administrative Details (2/2)

- Homepage
  - http://nlp.kaist.ac.kr/~cs579
- Time and Place
  - Tuesdays and Thursdays
  - 16:00 ~ 17:30
  - Room 2112, CS Bldg.

Course Materials (1/2)

- Primary
  - Patrick Blackburn and Johan Bos, Representation and Inference for Natural Language: A First Course in Computational Semantics, CSLI Studies in Computational Linguistics, CSLI Publications, 2005.
- Secondary
  - Patrick Blackburn, Johan Bos, and Kristina Striegnitz, Learn Prolog Now!, College Publications, King's College, 2006. http://www.learnprolognow.org

Course Materials (1/2)

- Others
  - Steven Bird, Ewan Klein, and Edward Loper, Natural Language Processing with Python, O'Reilly, 2009.
  - Daniel Jurafsky and James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2nd edition, Pearson:Prentice Hall, 2009.

Weekly Schedule (1/2)

- Week 1: Introduction
- Week 2: Prolog
- Week 3: Prolog
- Week 4: First Order Logic
- Week 5: First Order Logic; Lambda Calculus
- Week 6: Lambda Calculus
- Week 7: Midterm Exam (10/23)

Weekly Schedule (2/2)

- Week 8: Underspecified Representations
- Week 9: Underspecified Representations
- Week 10: Propositional Inference
- Week 11: Propositional Inference
- Week 12: First Order Inference
- Week 13: First Order Inference
- Week 14: Putting It All Together
- Week 15: Final Exam (12/18)

Evaluation

- Midterm Exam: 25%
- Final Exam: 25%
- Attendance: 20%
- Homework/Class Participation: 30%

INTRODUCTION

Introduction
Implementations
Interactions
Theoretical Limitations

Introduction

- Formal semantics
  - Giving model-theoretic interpretations to fragments of natural language
- Computational semantics
  - Semantic construction
    - Every boxer loves Mia.
    - ∀x(BOXER(x) -> LOVE(x,MIA))
  - Inference
    - Every boxer loves Mia. Butch is a boxer. Butch does not love Mia.

Introduction

- Theorem proving
- Consistency checking
- Model checking
- Model building

Implementations (1/2)

- Curt: Clever Use of Reasoning Tools
  - Baby Curt
    - no inference tools
  - Rugrat Curt
    - consistency checking
  - Clever Curt
    - theorem proving and model building

Implementations (2/2)

- Sensitive Curt
  - informativity checking
- Scrupulous Curt
  - equivalence reasoning
- Knowledgeable Curt
  - background knowledge
- Helpful Curt
  - query information

Interactions

- Baby Curt

> Vincent loves Mia.
Curt: OK.
> Every woman knows a boxer.
Curt: OK.
> Mia smokes.
Curt: OK.
> Mia does not smoke.
Curt: OK.

Interactions

- Rugrat Curt

  > Mia smokes.

  Curt: OK.

  > Mia does not smoke.

  Message (consistency checking): proof found.

  Curt: No! I do not believe that!

Interactions

- Rugrat Curt (continued)

  > Vincent is a man.

  Curt: OK.

  > Mia likes every man.

  Curt: OK.

  > Mia does not like Vincent.

  Curt: OK.

Interactions

- Clever Curt

  > Mia dances.

  Message (consistency checking): mace found a result.

  Curt: OK.

  > Jody dances.

  Message (consistency checking): mace found a result.

  Curt: OK.

Interactions

- Clever Curt (continued)

  > Mia is not Jody.

  Message (consistency checking): mace found a result.

  Curt: OK.

  > Vincent knows every boxer.

  Message (consistency checking): mace found a result.

  Curt: OK.

Interactions

- Clever Curt (continued)

  > Butch is a boxer.

  Message (consistency checking): mace found a result.

  Curt: OK.

  > Vincent knows Butch.

  Message (consistency checking): mace found a result.

  Curt: OK.

Interactions

- Clever Curt (continued)

  > Mia smokes.

  Message (consistency checking): mace found a result.

  Curt: OK.

  > Mia smokes.

  Message (consistency checking): mace found a result.

  Curt: OK.

  > Mia smokes.

  Message (consistency checking): mace found a result.

  Curt: OK.

Interactions

- Sensitive Curt

  > Mia smokes.

  Message (consistency checking): mace found a result.

  Curt: OK.

  > Mia smokes.

  Message (consistency checking): mace found a result.

  Message (informativity checking): otter found a result.

  Curt: Well, that is obvious!

Interactions

- Scrupulous Curt

  > A boxer loves a woman

  Message (consistency checking): mace found a result

  Message (consistency checking): mace found a result.

  Message (informativity checking): mace found a result

  Message (informativity checking): otter found a result

  Curt: OK.

  > summary

  Message (eliminating equivalent readings): there are two readings:

Interactions

- Knowledgeable Curt

> Jody is a woman.

Message (informativity checking): mace found a result.

Message (consistency checking): mace found a result.

Curt: OK.

> Jody is a plant.

Message (consistency checking): otter found a result.

Curt: No! I do not believe that!

Interactions

- Helpful Curt

> Vincent knows every woman.

Curt: OK.

> Mia is a woman.

Curt: OK.

> Who knows Mia?

Curt: This question makes sense!

Curt: vincent

> Who is a plant?

Curt: I have no idea.

Theoretical Limitations

- The satisfiability problem is shown to be polynomial, NP-complete, EXPTIME-complete, NEXPTIME-complete and undecidable for a series of fragments of English (syllogism, relative clauses, non-copula verbs, restricted anaphora, unrestricted anaphora, respectively).

(Ian Pratt-Hartmann, Fragments of Language, Kluwer Academic Publishers, 2003. link)