Knowledge Engineering for NLP

April 3, 2006

Test suites

Overview

- General announcements
- Q&A from lab 1
- Evaluation in computational linguistics
- Uses of test suites
- Evaluating precision grammars
- Levels of adequacy
- Test suite software ([incr tsdb()] demo)
- If time, Lab 2 preview

Be sure to ask questions!!

- About: lab instructions, phenomena in your language, the Matrix, the LKB, HPSG
- Who's using the FAQ?
- Who spent $>$ 10 minutes figuring out an LKB error?
- What do you want to know about Lab 1?

Evaluation and computational linguistics

- Why is evaluation so prominent in computational linguistics?
- Why is it not so prominent in other subfields of linguistics?
- What about CS?

Uses of testsuites

- How far do I have left to go?
  - Internal metric
  - Objective comparison of different systems

    → How do you evaluate precision grammars?
- Where have I been?
  - Regression testing
  - Documentation

Kinds of test suites

- Hand constructed
  - Controlled vocabulary
  - Positive and negative examples
  - Controlled ambiguity
- Corpus based
  - More open vocabulary
  - Greater ambiguity
  - Haphazard ungrammatical examples
  - Application-focused
- Which kind for which use?

Evaluating precision grammars

- Coverage against a corpus
  - Which corpus?
  - Challenges of lexical acquisition
- Coverage of phenomena
  - How does one choose phenomena?
  - What did TSNLP do?
- Comparison across languages

Levels of adequacy (1/2)

- Grammaticality
- “Right” structure
- “Right” dependencies
- Exact match semantics

Levels of adequacy (2/2)

- Only legitimate parses (how do you tell?)
- Some set of parses including the preferred one
- Preferred parse only/first
- Preferred parse within first N

Test suite software

- What does the LKB batch parse utility do?
- What else would you like it to do?
- → [incr tsdb()] demo

Overview

- General announcements
- LKB Formalism con’t
- Evaluation in computational linguistics
- Uses of test suites
- Evaluating precision grammars
- Levels of adequacy
- Test suite software ([incr tsdb()] demo)
- Lab 2 preview...