Knowledge Engineering for NLP

January 12, 2009

Test suites

Overview

- General announcements
- Evaluation in computational linguistics
- Uses of test suites
- Evaluating precision grammars
- Levels of adequacy
- Test suite software ([incr tsdb()] demo)
- Look at instructions for Lab 2

Ask more 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?

Questions from Lab 1 (1/5)

- How can you identify only part of a feature structure?
- What is the significance of number of edges in a batch parse in regard to differences between revisions of a grammar?
- I’m curious as to how I can come up with a coherent multiple inheritance sche me for my types without constructing all of my types first.
- How are we going to implement things like the SHAC?

Questions from Lab 1 (2/5)

- Why did we pull the number (e.g. 3sing) into the type rather than using something like arg-st?
- Why is ‘label’ defined twice in types.tdl?
- Why does creating a single Head-Complement Rule change the number of edges for SOME sentences?
- What’s the significance of the node-label type name, i.e. ‘np-label.’ Where else does it appear and must it be unique?
- Using Trollet, why do the type hierarchy windows freeze if I don’t close them before reloading the grammar?

Questions from Lab 1 (3/5)

- Is there a listing of the lkb commands used in emacs anywhere?
- Can I get the Windows version?
- Is it possible to get the syntax highlighting in emacs to work properly with .tdl files.
- Why is it acceptable to treat the COMPS value as *ne-list* when it is declared as type *list*. In order to change the Head-Complement Rule, I used FIRST and REST, but these features are declared for *ne-list* not *list*.

Questions from Lab 1 (4/5)

- It seems that infixes and circumixes are not handled. Is this true? I suppose a circumix could be treated as a prefix and suffix combined, but since the rule requires you to chose one or the other, it would have to be broken into two rules. Is there a way to enforce t hat both rules are applied since having only part of it would lead to an ill-formed word if only one were applied?
- This lab didn’t ask for it so I didn’t include it, but I was wondering if we bother ma king phrase supertypes other than labels? I found myself doing phrase & head noun a lot.

Questions from Lab 1 (5/5)

- I was kind of stressed out about possibly having to create trans, intrans, ditrans, and then past, 3sing pres and non3sing pres for each type.

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
- Evaluation in computational linguistics
- Uses of test suites
- Evaluating precision grammars
- Levels of adequacy
- Test suite software ([incr tsdb()] demo)
- Look at instructions for Lab 2