Foundations of Language Science and Technology

Technological Foundations II

Stephan Busemann

German Research Center for Artificial Intelligence
(DFKI GmbH)

Overview

- Language Technologies vs Human Language Processing
- Evaluation Techniques
- Exploring the LT World (http://www.lt-world.org)
- Exercise

Performance Revisited

Competence / Performance

- Competence: skills and abilities needed to solve a problem. Can not be observed directly.
- Performance: behaviour in solving a problem. Can be observed.

Applied to Language

- People know the grammar of English. This is their **competence**.
- People often produce deviant, ungrammatical utterances (which may be understood by others). This is their **performance**.

LT systems

- No distinction between competence and performance.
- However a system's performance usually differs from human performance when given the same task.

Human and Machine Performance

Assume a two-party dialogue application. USR is a human customer in a travel agency. SYS is a consultation system for travel recommendations.

After some talk…

USR *I'd like one of the smaller hotels, with a pool. I'm a nonswimmer.*
SYS *You may wish to stay at the BelAir. They have both an indoor and a large outdoor pool.*
USR *Are these pools deep?*
SYS ?? … ??

- Out of domain talk may lead to disrupture
- System doesn't know the concept of a pool's depth. It doesn't have data about pool depth either. It can't reason about this situation. At most: *„I don't know what you mean by a pool being deep."*
- A human agent should be able to explain, infer and cooperate: *„I don't know how deep they are. But the hotel has wading pools, too. So you'll most certainly find a safe area in the water."*

Avoiding Errors

Humans try to anticipate and avoid errors by quickly choosing a „safer solution“.
Ex.: style used in foreign language text production

- Speaker should like to say: „sanitize the public image of LT“
- Speaker self-monitors his message: „improve the public picture of LT“
- Speaker realizes that „picture“ is the wrong word
- Speaker doesn‘t use metaphor at all: „contribute positive connotations to LT in public“

Language technologies don‘t usually have a dedicated mechanism for error anticipation and avoidance.

- LT errors from basic methods or component technologies show in the output
- No feedback architecture allowing inter-component interaction

What is Language Understanding?

Understanding „understanding“: Verifyiable Scenarios in which an intellectual effort can be demonstrated that involves reasonable action as a consequence of a linguistic stimulus

There are different ways to define „language understanding“

- Tourist satisfied with a trip recommended by a computer agent in the course of a NL dialogue
- Translating a text from one language into another
- User constructing an electric circuit upon NL advice by a computer
- Robot fetching something after being told to do so

What language understanding is not:

- Successful runs of a parser that maps text input onto a logical form output
- Phone routing systems (predefined interpretations)
- Airport flight information (predefined utterances)

Modeling Language Understanding is Always Partial

Linguistic coverage

- *I'd like to fly to Cuba.*
- *Are there still flights to Cuba?*
- *Can you please book me a flight to Cuba!*
- …

Conceptual (out of domain) coverage

- *With BA, food is better.*
- *I have fear of flying.*
- *Why not by car?*
- …

Social coverage (adolescence)

- No social learning
- No social experience
- No social integration

In constructing a model we necessarily exclude anything that is not modeled

Constructing models is not the right way to making computer performance more similar to human performance

Adding models of thought, behavior, social roles etc. will improve performance, but still remain deficient

Evaluation Techniques I

How can we assess whether our technology lives up to expectations?
How can we compare a technology with other technologies that do the same thing?

Glassbox evaluation (competence predicted by theory) vs.
Blackbox evaluation (performance of implemented system)

1. Introspection

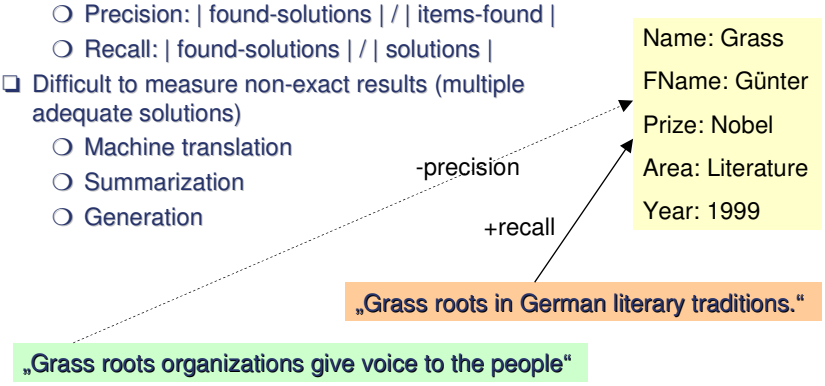
- Author of system sits back and checking what is plausible
- Self-evaluation
- No general validity of results

2. Group tests

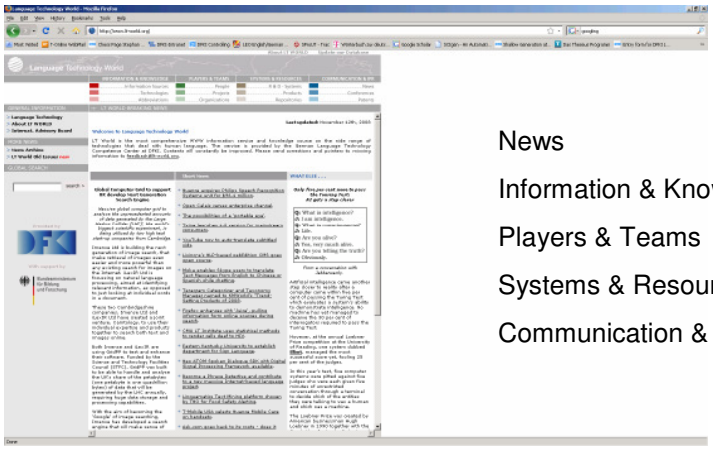
- A group of possible intended users (= hire a few motivated undergraduates) is testing the system
- Slow, costly, difficult to get reliably
- General validity questionable

3. Comparison against „gold standard“

- Corpus of representative texts
- Annotations of results to be tested
- Comparison with system results
- Measures
  - Precision: | found-solutions | / | items-found |
  - Recall: | found-solutions | / | solutions |
- Difficult to measure non-exact results (multiple adequate solutions)
  - Machine translation
  - Summarization
  - Generation



http://www.lt.world.org is the major Internet portal for language technologies.
First result when googling „language technology“. Maintained at DFKI.



News

Information & Knowledge

Players & Teams

Systems & Resources

Communication & IPR

Improving on the LT World

Choose a language technology in one of the subsections of the Technologies area.

1. Consider the information associated with it.
   - ❑ Is it still current?
   - ❑ Can you find newer relevant information on Google?
   - ❑ Do available link resources maintain relevant newer information?

2. Do you have other recommendations regarding missing / outdated technologies?

Write up your findings (one page only) and present them on Friday (5-10 mins).

If your results are used for an update of LT World, you will be duly acknowledged.