The *fagreferent* as a key factor in the information society

Diana Santos
SINTEF Telecom and Informatics /
Distributed resource center for the computational processing of Portuguese

Presentation plan

- Introduction
  - *Fagreference* and my job
  - Information society...
- Three real cases
  - Area overview
  - Undeserved recognition
  - Newspaper categories
- Three relevant references
- Summing up

Two main ideas in this talk

- It is necessary to know a subject in order to perform meaningful search
  - *so, what we need is people who can discriminate what is good and bad*
- A computer can HELP with a task, not replace a person who is good at it
  - *technology-seduced decision-makers fail to acknowledge the human capital importance*

My background

- Formal education (*sivilingeniør*) in electronic engineering and computers
- MSc and PhD in natural language processing
- 14 years of research and development in natural language processing, machine translation, web services, contrastive linguistics
- In charge of the *Distributed resource center for the computational processing of Portuguese*

The Computational Processing of Portuguese project

- Catalogue the area
- Identify its main problems
- Suggest courses of action
- Make the resources more widely available
- Create a Web portal for the area
- Create and/or improve the resources
- Evaluate them

http://www.portugues.mct.pt

Information society

consumers, providers, processors, agents interested in
- disseminating
- concealing
- forfeiting
- destroying

INFORMATION

Society

- Associated with people / living beings
- Different degrees of authority
- Well defined rules/roles

It is not basically a technological issue!

IS=Computer-mediated lives

- learn, teach, communicate, work
- produce, buy, play, sell, do research
- enjoy (travel, visit museums, gossip, sex, emotions, company)

The naive view

You can do everything better with a computer because it has access to an enormous network of resources **????**

Three counterexamples to the naive view

The only thing in common, is that they belong to my own experience

- Different strategies to map a *fag*: the weight of authority
- The inability to choose: how to select from non-classified information
- The difficulty to categorize: there are no waterproof general categories

Area map in language technology

- MLIS/EuroMAP initiative in Portugal
  very defficient coverage, strong bias
- the Computational Processing of Portuguese project's way
  only objective criteria were taken into account
- the Finnish approach
  consult the authorities for expert profiling

Portuguese scientists...

Thank you for being a Portuguese scientist. My son is in seventh grade at (...), USA. He is in the process of completing a project on Portugal. Joe needed a Portuguese scientist profiled in his report. Excuse our ignorance; however, we knew of no such person, until we searched the Internet.

CETEMPúblico: Knowledge organization

- Criteria for newspaper sections
  - place (International, National, Local…)
  - size and place (Short, First page, Last page)
  - origin (international, experts)
  - time (50 years ago; last minute)
  - authorship (experts, celebrities)
- Almost no subject matter
  - sports, politics, gossip, cooking, computers

Three relevant references

- The social life of information
- Hypersearching the Web
- Information extraction

Further reading...

Tunnel design

- *it is easy to overlook the resources that lie beyond the immediacy of the information tunnel, even when they are quite substantial*
- *his enthusiasm had screened out an enormous array of people, organizations, and institutions involved in this "direct" touch*
- *it is in these steps -- from sources to reporters to editors and news organizations -- that news is made.*

Brown, John Seely, and Duguid, Paul. "Special Issue with Excerpts from: The Social Life of Information" *First Monday* **5** (4) (April 3, 2000)

Authority and hubs

*In addition to **expert sites that have garnered many recommendations**, the Web is full of another type of page: **hubs** that link to those prestigious locations, tacitly radiating influence outward to them. (...) So even if we find it difficult to define "authorities" and "hubs" in isolation, we can state this much: a respected authority is a page that is referred to by many good hubs; a useful hub is a location that points to many valuable authorities.*

The CLEVER project
http://www.sciam.com/1999/0699issue/0699raghavan.html

A digression

- Trying to find the CLEVER article in Altavista: I knew it had the word *hubs* in it, and its topic was Web search

  word count: hubs: 528305;
  internet: 93790965;
  search: 101138013
- Trying to retrieve the Brown & Duguid article from the Web:

  stale address

Digression inside the digression

Trying to find "MDF" (supposedly a sort of standard for/related to Web-based teaching), in the first 100 hits provided by Altavista

- Main Distributing Frame
- Main distribution facility
- map definition files
- Meta Data Format
- Menu Definition File
- Multiple Domain Feature
- motion displacement field
- Medium Density Fibreboard (in the description of a car audio system)

Digression^3

Trying to find *control* in Linguistics Abstracts [control = grammatical category catering for the understood subject of e.g. an infinitive]

- voice control, attentional control, theme control, plus all cases where control is mentioned in passing, not the topic
- 30% in the first 100

Question answering

- TREC-9 question answering track
  - Real examples from Microsoft Encarta's logs (previously from FAQFinder logs)
  - Human judges

Ellen M. Voorhes & Dawn M. Tice, "The TREC-8 Question Answering Track", *Proc. Second International Conference on Language Resources and Evaluation, LREC 2000* (Athens, 31 May -2 June 2000), Vol III, pp.1501-8.

Reflections

- to select this/any collection of references, human discretion is necessary
- give meaning to / interpret experience into a form of an argument so that others can agree or disagree but aprehend the message
- to provide a coherent picture of a large amount of information, for the purposes of another human, you need a human expert

a fool and a computer do not solve any problem

Lessons learned from AI

- What computers are worse at is common sense
- Imagining that a computer would have access to unlimited information and flawless deductive capabilities...

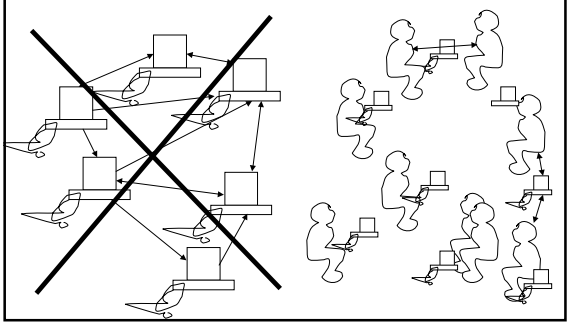
Why do church bells ring on Sundays?
Because the bellman pulls the cord

Concluding remarks

¶ Only knowledge allows gathering of knowledge

¶ Computers provide infrastructural help for humans, not a replacement for them

¶ Evaluation of computer systems performance is done against human judgements/performance

¶ You need people who can do their job well, so they can both help develop the tools and really use them to do the job better

This kind of information society