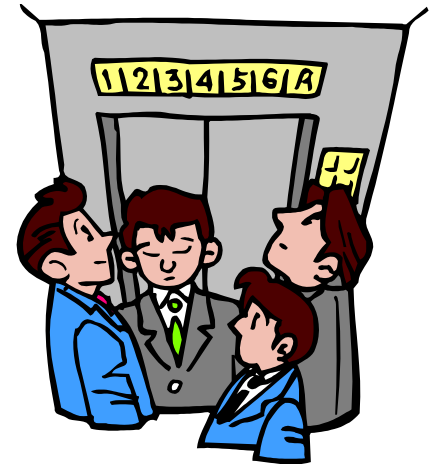




CS 431

The Semester in Elevator Speak

Carl Lagoze - Cornell University
May 3, 2006



Libraries as a model

- Elevator Speak
 - Tim Berners-Lee didn't invent information. Libraries have a centuries long tradition of information organization. We need to learn from that tradition but rethink it in the networked environment.
- The research
 - Coordination of physical and digital information
 - Machine learning from organized corpora
 - Balancing human and machine effort

Metadata

- Elevator Speak
 - Metadata is both a plague and a cure. In many cases it is necessary, but too much thinking about it relies on human input. Non-expert humans just don't do it well
- The research
 - Automatic generation of metadata from document context
 - Automatic generation of metadata for non-textual resources from related text

Tools and Standards

- Elevator Speak
 - The entire XML stack provides a suite of tools and standards that enrich our ability to process semi-structured data. However, considerable work remains to make this suite as efficient and robust as established relational technology
- Research Areas
 - Bridging the gap between fully structured and unstructured data
 - Overcoming the complexity problem

Semantic Web

- Elevator Speak
 - Despite the almost overwhelming hype, the work coming out of the semantic web initiative provides an important foundation for modeling and manipulating distributed semi-structured information.
- Research Areas
 - Efficient storage and querying of highly-scaled semantic graphs
 - Populating the semantic web with minimal human effort

Web-Scale Information Discovery

- Elevator Speak
 - The use of link structure and document context has dramatically advanced our ability to find and rank information at a massive scale
- Research Areas
 - Customization of search results based on user profiles, role, geographic location, etc.
 - Incorporating the deep web
 - Introducing the dimension of time in web analysis

Preservation

- Elevator Speak
 - Despite years of research in preservation of digital content it remains a difficult, expensive, and unresolved problem
- Research Areas
 - Understanding “just enough” preservation
 - Economic models of preservation

Scholarly Publishing

- Elevator Speak
 - We are in the midst of massive changes. It is not yet clear who are the losers and winners or how the technical/social/economic solutions will shake out.
- Research Areas
 - P2P and scholarly publishing
 - Integrating social networks and bibliographic networks
 - Economic models

Digital Rights Management

- Elevator Speak
 - Another issue, like scholarly publishing, that is on the front lines of the battle between the old (physical) and new (digital) worlds. Who “wins” has a much to do with politics and money as it does with technology
- Research Areas
 - Fair use and DRM
 - Web-scale DRM infrastructure/authentication
 - Business models for a digital society

The Big Elevator Speak

- As "*code*" infiltrates our social, political, cultural, and economic lives it's not just good old computer science any more. We can work to create the most optimal algorithms and engineer the best systems. But, their effect on our lives requires an awareness of social context, human behavior, and ethics.