

# Information Technology Trends

**Elizabeth D. Liddy**

Professor & Director  
Center for Natural Language Processing  
School of Information Studies  
Syracuse University

May 9, 2007

# Difficulty of the Assigned Task!!

- Its hard to *WOW* you!
  - So much advance coverage and speculation in the press
  - Movies & TV make the future seem present

- ▶ Series 60 home page
- ▶ Introduction / Reviews
- ▶ Technical specifications
- ▶ **S60 software catalog**
- ▶ Series 60 Downloads
- ▶ FAQ & Tips
- ▶ Series 60 Gallery
- ▶ S60 Discussion Forums
- ▶ **S60 software shop**

- What's hot**
- **20 newest programs**
  - **20 top rated programs**

- Categories**
- **S60 software categories**
  - Miscellaneous Utilities (370)
  - Games and Entertainment (686)
  - Graphics & Multimedia (177)
  - Security (46)
  - Internet (78)
  - Emulation (21)
  - On-line services (7)
  - Programming, Development (24)
  - Finance & Business (34)
  - Compression (6)
  - Health & Medicine (23)
  - Corporate (1)
  - PC Programs (16)
  - Calculators & Math tools (22)
  - Food & Drinks (13)
  - Document Editors/Readers (26)
  - Learning & Dictionaries (34)
  - Time (23)
  - Sport (5)
  - Database Management (6)
  - Literature (2)
  - UI personalization (9)
  - Travel, Sport & Hobby (13)
  - Other (4)
  - Religion (21)
  - Software Bundles (2)

### SounderCover by Simea

[Buy It USD](#) [Buy It EUR](#) [Try It](#)

**Google Maps for mobile** - Get maps, directions & traffic info on your phone. Download today! [www.google.com/gmm](http://www.google.com/gmm)

[Ads by Google](#)

[Advertise on this site](#)

Did you ever wish you could hide your location when talking on the [phone](#)? Ever wanted to give the impression you were somewhere else?

SounderCover gives you the ability to add a background sound to any incoming or outgoing call, giving the impression that you really are in the environment where the background sound is normally heard.

Did you wake up late for work and you want your boss to think you're caught in traffic? Select the Traffic Jam background and give him a call from your bedroom :). He will hear your voice on top of traffic jam [sounds](#).

**SoundCover**

**Background tune:**

- circusparade.amr**
- dentist.amr
- menatwork.amr
- phoning-15s.amr
- phoning-30s.amr

Select

Pretend you're at the dentist, in the park, on the street, caught in a thunderstorm, near heavy machinery or at a circus parade. The possibilities are endless! You can even use your own prerecorded sounds or sounds downloaded from [the Internet](#).

# Difficulty of the Assigned Task!!

- Its hard to WOW you!
  - So much advance coverage and speculation in the press
  - Movies & TV make the future seem present
- 6 trends worth watching
- What are the underlying technology advances that will enable these applications?
  - Particular emphasis on language-based software, search & their applications

# Over-Arching Trends

- **Miniaturization**
  - Devices, data compression, nano-technologies
  - *ttyl, lol, bff, l8r, b4*
- **Personalization**
  - “*you-ness*”
- **Localization**
  - Contextualized everything
- **Pervasiveness**
  - ICT-enabled ‘any time, any place’
- **Affective Computing**
  - System-awareness of emotional state
- **Community-Based Information Provision**
  - Reliance on the Wisdom of the Masses

# Personalization

- Approach → Problem = global relevance
  - Are we really all the same?
- Or, by knowing some things about you, can a search engine give you better results?
  - Similar to Recommender Systems on Amazon, etc.
- Broadening this – provide “personalized” results based not only on who you are, but who you know
  - Friends, colleagues and anyone in your social network will influence the type of results you see
- Community-based profiles for improving your personalized search
  - Many web search queries originate from “identifiable communities” of searchers with similar needs & preferences
    - System must recognize 2 queries as expressing same need
    - Use community’s click-through history to re-rank hits
    - Empirical tests shows this improves results for tight communities

# Localization

- System captures behavior & builds probabilistic model of your needs / preferences
  - By day, by time, by location
- Immersive, location-aware, self-tracking technology that allows you to receive instant information targeted to the area that you're searching from
  - Finding the right auto mechanic closest to your job
  - A Thai restaurant in the city you're visiting
- Remember Microsoft's Implicit Query?
  - System would keep an eye on what you're doing and ask if you want to launch a search based on the context of your task
  - If composing an email about upcoming trip to NYC, Implicit Query would give you the options of searching for NYC hotels, airfares, events & restaurants

# Affective Computing

- Computers capable of reading, interpreting, and acting on people's emotions and moods
  - Using Machine Learning, vocal and visible recognition, and sensors capable of detecting signals indicating emotion
  - Requires ability to recognize connotative as well as denotative meaning
    - *Is this email apology really sincere?*
- Possible applications:
  1. Interprets an incoming message's linguistic clues and classifies it as *nice*, *nasty*, or *neutral*, and suggests a fitting emotional response
  2. Identifies searcher's emotional state and re-ranks search results accordingly
  3. Advises email writer of how a message will be interpreted by others and, before they send it, inquires if they intend it

# Community-Based Information

- Wisdom of the Crowds
  - *Why the Many Are Smarter Than the Few*
  - American Idol – popularity vs. expert view
  - Leveraging human capital on the Web
    - Social tagging – ESP game for labeling of Google images
  - Amapedia – a community of users who share information about the best products they own and love
- Wikipedia
  - 7.2 million articles in 251 languages
    - 1.7 million of which are in the English edition.
    - Ranks among the top 15 most-visited websites worldwide
    - Favors consensus over credentials
      - Criticized for reliability & accuracy
      - Susceptibility to vandalism
  - Need protective practices

# MINDS

- Recent, government-convened set of workshops to establish an agenda of paradigm-shifting research in the Human Language Technologies

**M – Machine Translation**

**I – Information Retrieval**

**N – Natural Language Processing**

**D – Data Resources**

**S – Speech Understanding**

- Outcome is a report based on input from leading researchers in each HLT area, motivated ONLY by what each research area thinks is important to ITS goals

# Machine Translation

- Real-time, fully-automatic, felicitous ability for text in any single language to be translated into any of the other languages in the world
  - 6,809 languages (est.)
- If you try Google's BabelFish, you know that open domain MT is a ways off
- Currently, PDA-like devices being tested by US forces in Iraq to provide Arabic  $\leftrightarrow$  English translations
  - Limited vocabulary based on situation and what might be said
  - Word-by-word translation only works in a closed domain
  - Need to utilize higher levels of language processing that take intent, context, and world knowledge into account

# Information Retrieval

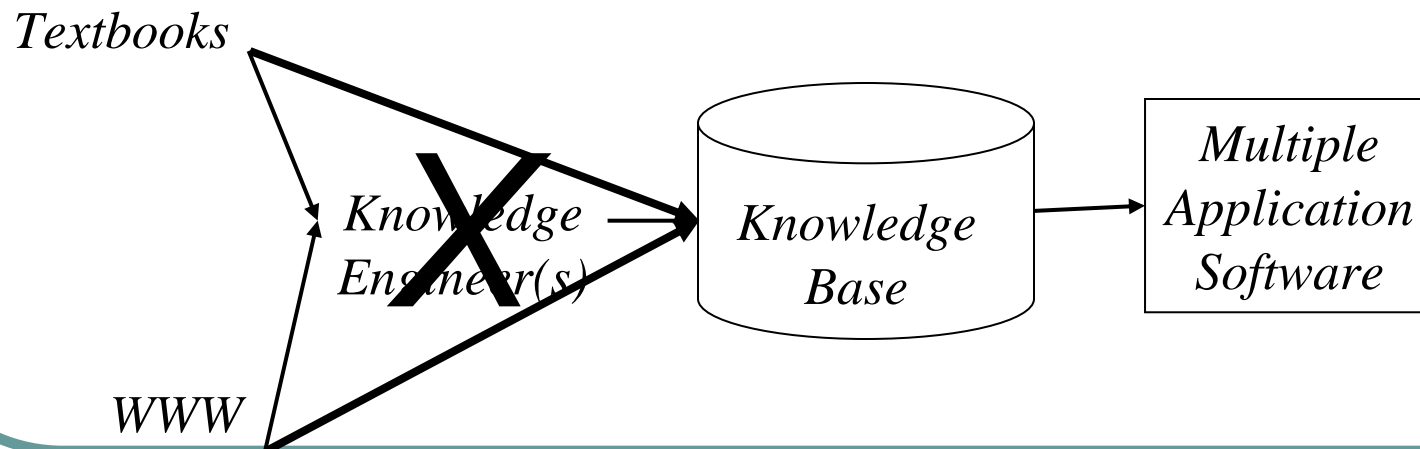
- A search engine that understands your *real* information need as stated in a full natural language question
  - The same as you would express it if you were asking another person
- Response will be an **answer**, not just a list of URLs of potentially relevant documents
  - Provided with supporting evidence, summarized from across multiple sources
- Future of search reflects notions discussed in *“The Long Tail”*
  - Web-based search exhibits a demand curve much like an online store's, with huge appeal of the top items, trailing off quickly for less popular ones
  - Challenge is to devise search engine that:
    1. Continues to provide excellent access to popular items
    2. Gets folks into the long tail on the open web to appropriate docs

# Natural Language Processing

- Ability of a system to understand language in both its obviousness and its subtleties, the same as humans do
- Much success at understanding smallest units of language
  - Reducing to root – *hide, hiding, hid*
  - Part-of-speech tagging
  - Syntax-based parsing
  - Word-based semantics
- Next frontiers are understanding the implicit meanings intended
  - Plan / goal of speaker
  - Connotative & emotive aspects of meaning
- Promise of success using Sublanguage Grammar methodology where the peculiarities of language use in a particular domain for a particular purpose can be learned

# Machine Reading

- **Challenge:** Although most of the world's knowledge is available in text resources,
  - Software today cannot improve its effectiveness on a task through reading and learning from those texts
- **Today:** Software experts & knowledge engineers meticulously, manually improve system performance by adding knowledge



# Machine Reading

- **Future:** Robust NLP + Machine Learning offers the potential to bridge gap from text → knowledge, but need to be able to learn:
  - Encyclopedic lexical knowledge
  - Domain & genre structure
  - Mapping between language and knowledge representation
- Applications
  - Read a user manual and answer complex help questions by owner
  - Answer Help Desk questions regarding software use
  - An intelligent tutor that reads a textbook and interacts with and aids the student, based on what it has learned from its reading

# Machine Reading Evaluation

- Instruct system to answer questions
  - Before reading a text for the first time
- Have system read a book / chapter / report in a subject area,
  - Biology textbook or software user manual
- Ask the system the same set of questions
- Measure the delta in question-answering capability from before reading to after
- Or, instruct system to ask questions regarding text just read
- Or, instruct system to summarize what it has read

# Socially-Aware Language Understanding

- **Challenge:** Incorporate social-context understanding in a system's interpretation of language
  - Requires system to accomplish deeper levels of interpretation
    - Discourse & Pragmatics
  - Beyond literal meaning – connotative as well as denotative
    - Politeness, sarcasm, humor, etc
- **Future:** Personalized NLP – Conversational systems that self-adapt to the person & the context
  - Ability of agent & person or 2 agents to jointly construct meaning
    - Each having own experience and expertise, but ability to take other's perspective into account to understand
    - Use of subtle features that highlight human-like linguistic intuitions to better understand and communicate

# Data Resources

- We'll skip this one
- Focus on multiple research communities agreeing to representation scheme, methodology for annotating sample texts with all the levels of meaning of interest to be learned
- **Future:** A range of creative ways to acquire annotated training data:
  - Leveraging human capital on the Web
    - Social tagging – ESP tagging / Open Minds
  - Active learning as a methodology
  - Performance improves as depth & breadth of annotation builds

# Speech Understanding

- Everyday audio, not just single-speaker, prepared talks
- Ability to quickly adapt to all voices, accents, background noises, multiple speakers, overlapping voices, mobile input devices (e.g. cell phones)
- If speech to text can be accomplished, then all of the advances in NLP and its applications can be applied to spoken text as well
- Dragon – version Professional 9 by Nuance
  - Up to 99% accurate
  - \$1,000

# Hybrid Future HLT Application

- Scenario - International meeting with many speakers and multiple languages
- **Speech Understanding** and **Machine Translation** will produce accurate transcripts of everything said in multiple languages
- **Natural Language Processing** would generate appropriately written meeting minutes, press releases, summaries, etc
- **Information Retrieval** would enable users to ask questions about motions, outcomes, who voted, etc

# New iSchool Movement

- Schools focused on research & education across **information**, **technology**, and **people**
  - Commitment to learning and understanding of the role of information in human endeavors
- Begun by Ray VonDran, Dean, iSchool, Syracuse
- Now a consortium of 19 schools

UC – Berkeley	Michigan
UC – Irvine	North Carolina
UC - Los Angeles	Penn State
Drexel	Pittsburgh
Florida State	Rutgers
Georgia Tech	Syracuse
Illinois @ Urbana-Champaign	UT @ Austin
Indiana	Toronto
Maryland	Washington
- Highly sought iSchool graduates - highest starting salary of any school at SU

# Summary

- ACM-SIGIR
  - <http://www.acm.org/sigs/sigir/>
- ACL - HLT
  - <http://www.aclweb.org/>
- Search Engine Watch
  - <http://searchenginewatch.com/>
- Annual Search Engine Meeting
  - <http://www.infonortics.com/searchengines/>
- MINDS Report
  - <http://www.itl.nist.gov/iaui/894.02/minds.html>