Futures of IR Systems

W. Frisch
Institute of Government, European Studies and Comparative Social Science
University Vienna

The Seven Ages of Information Retrieval

Vannevar Bush's 1945 article set a goal of fast access to the contents of the world's libraries which looks like it will be achieved by 2010, sixty-five years later.

Bush's Prediction

IR Childhood (1945-1955)

- Ideas conceived
  - Information explosion after World War II
  - Possibility of information processing machine
- Memex
  - The hardware seems mostly out of date.
    - user inserting 5000 pages per day into a personal repository and it taking hundreds of years to fill it up.
  - the software goals have not been achieved.

The Schoolboy (1960s)

- Many many experiments
- Use of Precision and Recall
- Use of relevance feedback

Adulthood (1970s)

- The invention of
  - word processing systems
  - time-sharing systems
- The beginning of information industry
  - OCLC
  - DIALOG
  - BRS

Maturity (1980s)
Mid-Life Crisis (1990s)
- Internet put IR to the test.
- Better understanding of the limit of IR.
- Large scale evaluations
- Digital Libraries projects

Predictions
- Fulfillment (2000s)
- Retirement (2010)

IR research

Top Ten Research Issues
10. Relevance Feedback.
9. Information Extraction.
7. Effective Retrieval.

Examples of Advanced IR Systems
- Intelligent retrieval
- Natural language dialog
- Text mining
- Knowledge-based indexing
<table>
<thead>
<tr>
<th>Intelligent Retrieval</th>
<th>Natural Dialog with the Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Expert systems</td>
<td>- Use natural language</td>
</tr>
<tr>
<td>- Case-based reasoning</td>
<td>- Let the computer keep a profile of your life-long</td>
</tr>
<tr>
<td>- Natural language understanding</td>
<td>information seeking patterns</td>
</tr>
<tr>
<td>- Automatic abstracting/summarizing</td>
<td>- Understand information received</td>
</tr>
<tr>
<td></td>
<td>- understand user’s information needs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Text mining</th>
<th>IR &amp; Text Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Information Summation</td>
<td>- IR helps users find information that is already</td>
</tr>
<tr>
<td>- Automatic summary</td>
<td>known and that exists in a document.</td>
</tr>
<tr>
<td>- Automatic translation</td>
<td>- TM attempts to examine a collection of documents</td>
</tr>
<tr>
<td>- Automatic clustering and</td>
<td>and discover information not contained in any</td>
</tr>
<tr>
<td>classification</td>
<td>individual document in the collection.</td>
</tr>
<tr>
<td>- Knowledge Discovery</td>
<td>-</td>
</tr>
<tr>
<td>- Associative query expansion</td>
<td>-</td>
</tr>
<tr>
<td>- Use of domain knowledge</td>
<td>-</td>
</tr>
<tr>
<td>- explore document association</td>
<td>-</td>
</tr>
<tr>
<td>- Generate visualization maps</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neural Network Computing</th>
<th>Knowledge-based Indexing</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Learn how the human brain works</td>
<td>- Automatic indexing + concept indexing.</td>
</tr>
<tr>
<td>- Create massive associative networks where each node has limited process capability (intelligence)</td>
<td>- Index not only documents but also ideas.</td>
</tr>
<tr>
<td>- Make the network learns by its own processing patterns</td>
<td>-</td>
</tr>
</tbody>
</table>
From Internet to Semantic Web
- Internet
  - Digital files are connected together
- WWW
  - Web pages are connected
- Digital Library
  - Information collections are connected
- Knowledge Net
  - Ideas are connected

Personal Information Agent?
- Find me anything I ask it to find.
- Help me clarify what I really need.
- Let me know what information is available
- Help me convert information into knowledge

Human-Computer Collaboration
- The nature of information seeking requires the dialog between the user and the computer to be deeper than what HCI describes.
  - Interaction -- to act upon one another
  - Collaboration -- work together purposely in an intellectual endeavor
  - Let the computer amplify and augment human's intellect.

Search Engines in the News
- Searching the Web gets easier with engines that try to read your mind
  - Now Google saves lives! Someone wondering if they were having a heart attack did a search on Google, found a page explaining the symptoms, then got to a hospital for help. Without Google, “I'd be dead today,” he's quoted as saying. A look at various search engines, search technology and tips.

Search Engines in the News (Commentary)
- Desperately Seeking Search Technology
  - BusinessWeek Online, September 24, 2001 by Robert D. Hof
  - After e-mail, people spend the most time online searching for stuff. Market researcher Jupiter Media Metrix Inc. found that 80% of online users will abandon a site if the search function doesn't work well. Says Martha M. Frey, an analyst at market researcher Patricia Seybold Group: “You could make a case that the main reason e-commerce is unprofitable is that the power of search has been overlooked.”

Building Web Sites With Depth
- Web Techniques, February 2001 by Jakob Nielsen and Marie Tahir
- Search is one of the most common, and one of the least successful ways that users look for things on the Web. Search is often as bad as the worst salesperson or customer service representative.
Google vs. SearchKing
- An ad marketer took Google to court over changes the popular search engine made to its page-ranking system.
  - When Google changed its PageRank, SearchKing lost business
  - Should the power of Google be regulated?
    - Like public utility or highway?

Google vs. Searchking
- Google vs Searchking: United States District Court in Oklahoma gave Google First Amendment protection in SearchKing's suit against Google for harm to its business from changes in Google's page ranking algorithms. SearchKing is a link farm for its customers. Last year its position in Google results dropped dramatically as Google adjusted its ranking rules. Google argued that PageRanks are opinions and protected by free speech. It won that point, but the Court did say that Google manually lowered SearchKing's PageRank. --Jan. 27.2003

Google vs. Evil
- Daniel Brandt, who runs google-watch.org, attacked PageRank, accusing the company of being unfair and undemocratic. Brandt urged the FTC to investigate Google and regulate it as a public utility - as a company that, in effect, controls access to the Internet's natural resources.
- Google vs. Evil

Finally ....
- Recommended Readings on Futures of Search Engine Technologies