Next Generation Information Retrieval: What Next After Google?

Dongwon Lee
IST / Penn State
April 22, 2005

“information about lingering brothers circus at pittsburgh 2005”
“information about ringling brothers circus at pittsburgh 2005”

“Researcher who is working on database and information retrieval at Penn State”
“What is the actress Helen Hunter doing these days?”

Can and Can’t

- What Google Can:
  - Large-scale Super-fast Document Retrieval
  - Syntactic match (ie, keyword comparison)

- What Google Can’t:
  - Natural Language Processing
  - Information Retrieval
  - No way for Knowledge Retrieval

Feel free to replace “Google” by other search engines like “Yahoo” or “A9”, etc.
Document Retrieval vs. Information Retrieval

Information Retrieval vs. Knowledge Retrieval

Figure from Eric Brill, "The NLP Silver Bullet for Search", CIDR 2005
Knowledge Retrieval

Semantic Web to the Rescue

- "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation…"

- Building Blocks
  - RDF (Resource Description Framework): asserts facts on web resources
  - RDF Schema: defines vocabularies
  - OWL (Web Ontology Language): defines relationships among vocabularies

- Will it succeed ???