

# Semantic Based Search Engines

Janet Hui-wen Hsiao

## Outline

- ◆ Introduction
- ◆ Previous & Related Works
- ◆ Motivation
- ◆ Basic Model
- ◆ Implementation
- ◆ Demo

2001/3/24

Semantic Based Search Engine

2

## Introduction

- ◆ Search Engines:
  - Keyword based: not always reflect what we are looking for.
  - Link based: need the help of keyword based method.
  - Semantic Based: search information based on the semantic of the query sentence.

2001/3/24

Semantic Based Search Engine

3

## Previous & Related Works

- ◆ Apply Natural Language Processing techniques:
  - need not understand complex rules for composing advanced queries.
  - Some examples:
    - Ask Jeeves: one of the first search engines using NLP.
    - Albert: support many languages (English, French, Spanish, German, Portuguese, etc.).
    - Other relevant products about information retrieval and search: Autonomy, LexiQuest, e-Cyc ...

2001/3/24

Semantic Based Search Engine

4

## Previous & Related Works

- ◆ Ask Jeeves: <http://www.askjeeves.com/>
  - combine the strengths of natural-language parsing software, data mining process, knowledge-base creation and maintenance tools with the cognitive strength.
  - rate Web sites according to how users interact with their content and services .
  - Drawbacks:
    - Need people to create large semantic network.
    - Need more interaction to narrow down the domain. (Because it has to deal with every possible case.)

2001/3/24

Semantic Based Search Engine

5

## Previous & Related Works

- ◆ Albert: <http://www.albert.com>
  - Features:
    - Contextual understanding
    - Automatic learning capabilities
    - Search personalization with user profiles
    - Tolerance to misspelling and mistyping
  - Has the same problem as Ask Jeeves

2001/3/24

Semantic Based Search Engine

6

## Previous & Related Works

- ◆ **Cyc** (en-cyc-lopedia)
  - World's largest general-purpose knowledge base.
  - Feeding information into a computer is a very tedious task.
  - E-Cyc: bring the story into the Internet
    - Cyc knowledge Server allows Internet and intranet web sites to add common-sense intelligence to their environments.
    - Separates out the different meanings of ambiguous concepts.
    - Teach Me button: ask for feed back
  - Test site: <http://beta.hotbot.com/>

2001/3/24

Semantic Based Search Engine

7

## Previous & Related Works

- ◆ **Webmind** : Intended to be the first real AI
  - Construct a brain system: an unified framework accomplishes every major aspect of the mind studied in psychology and brain science.
  - Between neural network and semantic network.
  - Basic idea (from an introduction article):
    - TextNode represents particular information in a text, or some "higher level" node represents more general topics. (actors)
    - Spreading activation: when a node get attention, it spreads some to its neighbors.
    - Return nodes with highest activation value.

2001/3/24

Semantic Based Search Engine

8

## Previous & Related Works

- ◆ **Webmind** (cont)
  - Goal: is built to want to answer queries.
  - Overview: not only match keywords and numerical data, but to **conceptually understand it**.
    - Incorporating with user feedbacks.
    - Extract sentiment from documents and classify them according to whether they express positive, negative or neutral attitudes
    - Recommendation system: In addition to returning the usual exact-match data, Webmind Recommendation can provide near matches.

2001/3/24

Semantic Based Search Engine

9

## Motivation

- ◆ Problem: We need a BIG semantic network inside the semantic based search engine. However, it is hard to build a good one.
  - Some words have domain specific meanings. If we return all possibilities, it only produces more garbage. (interactive querying)
  - Some queries need domain knowledge to respond well.

2001/3/24

Semantic Based Search Engine

10

## Motivation

- ◆ Observation:
  - Most problems concerning with "semantic" are about "domain" → Can we build a better search engine if the domain is known?
- ◆ Goal:
  - A semantic search engine which can perform better in local network area (e.g. [www.cs.sfu.ca](http://www.cs.sfu.ca)) than other global search engine. (because the domain is known)

2001/3/24

Semantic Based Search Engine

11

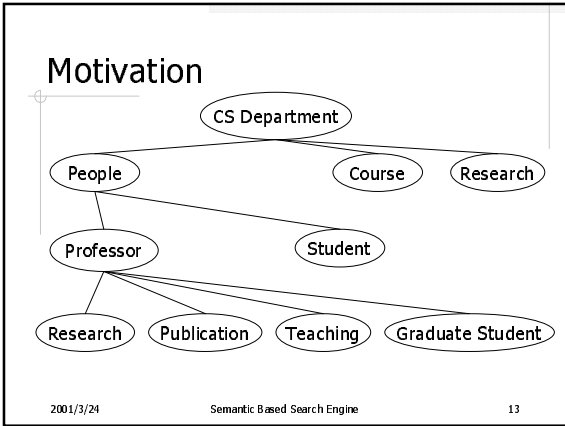
## Motivation

- ◆ Observations:
  - In a certain domain, most pages on the Internet have "templates".
  - Templates basically form semantic relationships.
  - Templates → XML, keep information in hierarchical templates.
  - If we encode semantic information inside an XML file, it can be queried like a database with XML query language.

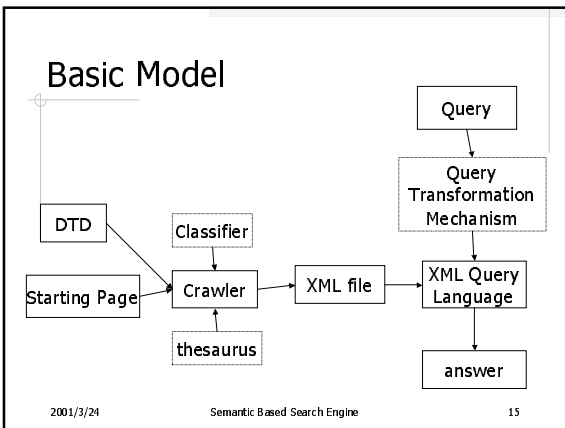
2001/3/24

Semantic Based Search Engine

12



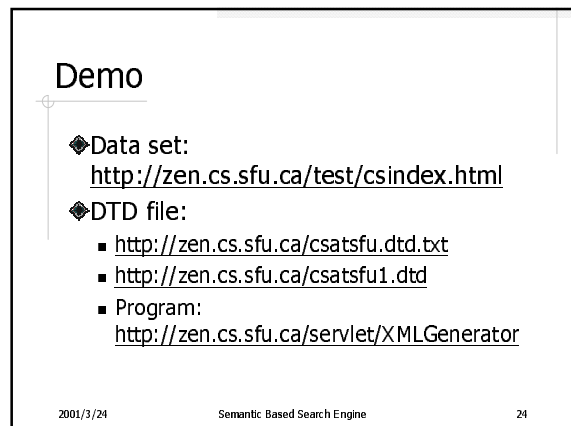
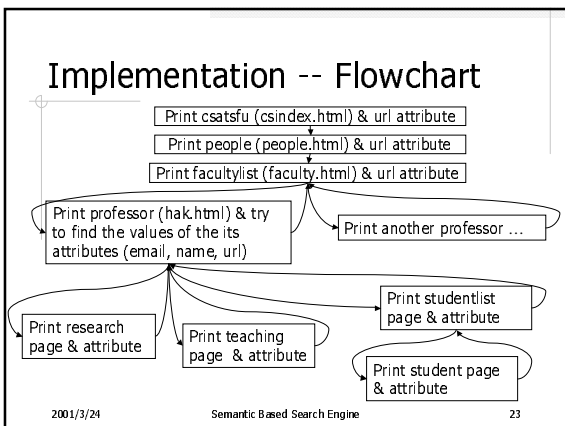
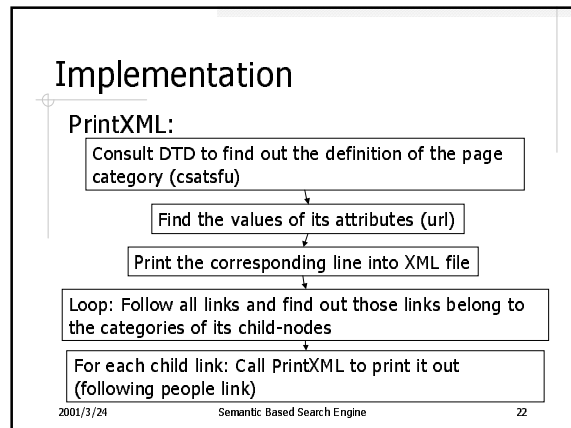
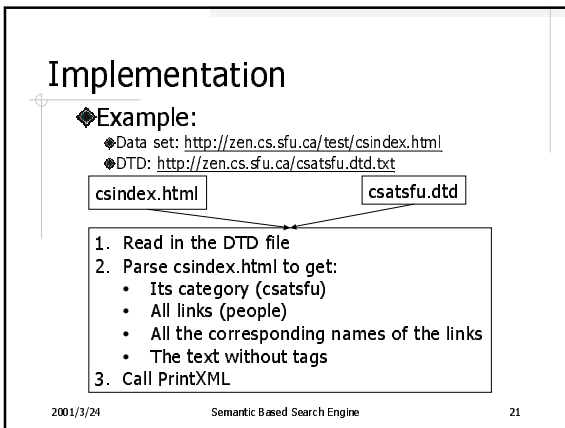
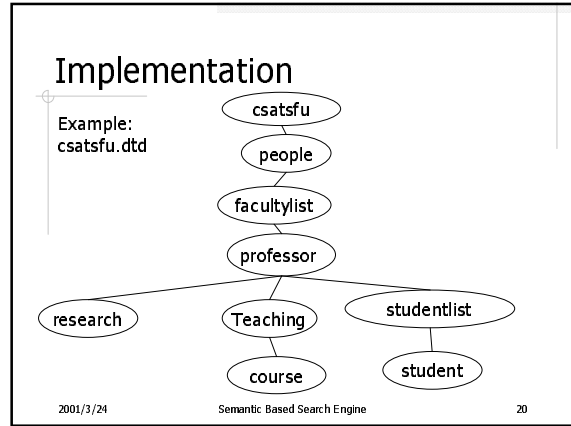
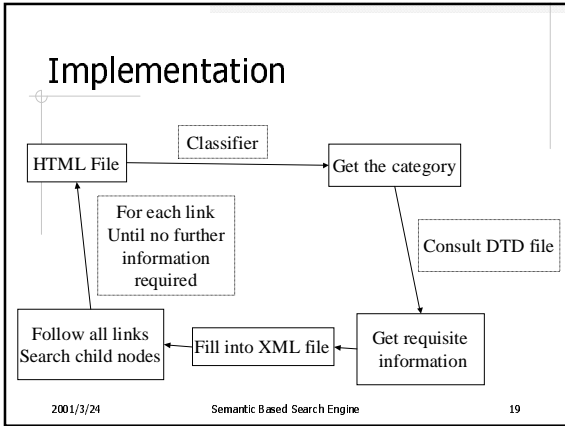
- ### Basic Model
- ◆ Inputs:
    - Web pages and structures for a certain domain
    - DTD file to define the data types in that domain
  - ◆ Outputs:
    - An XML file with useful information
  - ◆ Tools:
    - Web Classifier
    - Thesaurus: synonyms, proper nouns ...
- 2001/3/24 Semantic Based Search Engine 14



- ### Basic Model
- ◆ Example: inside [www.cs.sfu.ca](http://www.cs.sfu.ca) domain
    - DTD (professor template only): <http://zen.cs.sfu.ca/csatsfu.dtd.txt>
    - XML: <http://zen.cs.sfu.ca/csatsfu.xml>
- 2001/3/24 Semantic Based Search Engine 16

- ### Implementation
- ◆ Goal:
    - get information from the starting page, according to the DTD file, and then crawl down until all necessary information has been found. Then construct an XML file.
  - ◆ Wrapper:
    - Information has been “wrapped” with suitable wrappers by the designers of the web pages.
- 2001/3/24 Semantic Based Search Engine 17

- ### Implementation
- ◆ Some possible heuristic for extracting information (Professor Template):
    - Email: search “mailto:” link
    - Name: look up the proper name database.
    - Keyword search: research, teaching, graduate students ...
- 2001/3/24 Semantic Based Search Engine 18



## Further Research

- ◆ How to transform a keyword query into an XML query?

2001/3/24

Semantic Based Search Engine

25

## Links

- ◆ <http://www.ai.about.com/compute/ai/library/weekly/aa101300a.htm>
- ◆ <http://ai.about.com/compute/ai/library/weekly/aa021900a.htm>
- ◆ <http://www.askgeeves.com>
- ◆ <http://www.albert.com>
- ◆ <http://www.webmind.com>
- ◆ <http://www.e-cyc.com>

2001/3/24

Semantic Based Search Engine

26