Managing Semantic Content for the Web

Presentation by
Wesley W. Owen

Introduction

• Meaning of “Semantic”
  - of or relating to meaning in language
  Merriam-Webster's Dictionary 2002

• Meaning of “Semantic Web”
  - define meaning to web content
Overview of Presentation

- Introduction to SCORE
- System Architecture Overview
- Semantic Metadata Extraction
- Semantic Search Engine
- Performance and Scalability

Semantic Web (1/2)

- Scientific American article says 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.”
Semantic Web (2/2)

- Semantic Web could be as revolutionary as the web itself says researchers at Stanford University

SCORE Introduction (1/2)

- SCORE stands for:
  - Semantic Content Organization and Retrieval Engine
- Based on research from University of Georgia’s Large Scale Distributed Information Systems
SCORE Introduction (2/2)

- SCORE engine and support semantic applications such as:
  - Context-sensitive search
  - Browsing
  - Correlation
  - Normalization
  - Content analysis

How SCORE Works (overview)

- Semantic organization and use of metadata
- Semantic normalization
- Semantic search
- Semantic association
Semantic organization and use of metadata (overview)

Metadata has 2 forms
- Syntactic
- Semantic

Syntactic
- Noncontextual information about content such as language, format, etc.
- No insight of the documents meaning
Semantic

- Domain specific information about the content
- Example: finance domain: company name, ticker, industry, executives, etc.

Semantic normalization

- Tie like terms together
- Example: Yahoo’s founder known as both “Chief Yahoo” and “Yahoo Founder”.
Semantic search

• Able to search semantically
• Example: Movies Robert Redford directed, but not acted in.
• Example: palm
  - Company “Palm”
  - Technology: Palm
    • Operating system
    • Product (PDA: palm)

Semantic association

• Being able to associate search terms with other highly likely terms to provide relevant material.
• Example: A search on Intel would bring up Intel, reports about the semiconductor sector, etc.
System Architecture Overview (1/2)

- 2 components of ontology:
  - WorldModel: The definitional component
  - Knowledgebase: the assertional component

System Architecture Overview (2/2)

- Score System Architecture
  - WorldModel and Knowledgebase
  - Content Processing
  - Semantic applications
Semantic Metadata Extraction of Structured and Semistructured Text

• Four Key Capabilities
  - Extracting
  - Identifying
  - Enhancing
  - Maintaining

Four Key Capabilities (1/2)

• Extracting
  - Extracting metadata by scanning unstructured text as well as by exploiting the content structure

• Identifying
  - Identifying both domain-specific (semantic) and domain-independent (syntactic) metadata
Four Key Capabilities (2/2)

- **Enhancing**
  - Enhancing the extracted information using the Knowledgebase
- **Maintaining**
  - Maintaining the Knowledgebase using extraction technology, avoiding the problem of static, soon-obsolete dictionaries

Example of hierarchical classification

- “person” gets subclassed with: politician, artist, sportsperson
- “sportsperson” get subclassed with: coach, athlete, etc.
Metadata Extraction from Unstructured Text

- Ambiguities
  - Classification
    - Classifies documents in categories from human generated rules, or from training documents
  - Knowledgebase
    - Relates words in documents that are in the same entity classes

Semantic Search Engine

- User can provide search attributes for highly precise results
  - Example: Golf::player=Tiger
- Search Engine keeps metabase and knowledge base in main memory, providing very fast queries (10 ms)
- All query results are returned as XML
Performance and Scalability (1/2)

- Current version of SCORE running on:
  - Dual Pentium III 766 Mhz
  - 2 GB Ram
  - RedHat Linux
    - Apache
    - MySQL

Performance and Scalability (2/2)

- Main memory holds 4.5 million documents per server
- Can scale by adding more RAM to each server, or distribute over more servers.
Conclusion

• SCORE has shown the first steps as a proof on concept for the Semantic Web
  - Works well in well defined areas
  - Manages some ways of growing, but needs help and work in this area

References (1/3)

• Merriam-Webster's Dictionary 2002
References (2/3)


References (3/3)

Any Questions?