

Chapter 12

Semantic search technology

Search engines

Human directed search

- Keywords, concepts and references
- Counts occurrences of keywords

Automated search

- Based on Web crawlers, spiders, bots, or agents
- Individual search engines
- Metasearchers

Search engines

Ranking and relevancy

- Location and frequency of keywords
- Polysemy and synonymy

Three criteria for useful searches

- Recall
- Precision
- Meaningful ranking

Search engines

Google search algorithm

- PageRank
- Uses a vast link structure to rank pages

$$PR(A) = (1 - d) + d\left(\frac{PR(T_1)}{C(T_1)} + \dots + d\left(\frac{PR(T_n)}{C(T_n)}\right)\right)$$

PR(A) is PageRank of page A

PR(T1) is PageRank of page T1

C(T1) is the number of outgoing links from page T1

d is a damping factor in the range $0 < d < 1$, usually set to 0.85

Search engines

Semantic search

Searching techniques

- Incompleteness
- Uninformed: Depth first search, Breadth first search
- Informed: Best-first search, A*

Web search agents

Searches the web itself using documents connected through links

Four main phases: Initialization, perception, action and effect

Moves from initial phase to a loop of perception, action and effect.

Semantic methods

Latent semantic index searches looks at patterns of words within a set of documents to determine if they are related to each other semantically

Stemming: A process of removing words which do not carry meaning

Local weighting: Words that appear multiple times in a document is given a higher weight than words that appear once.

Web of documents and words: connecting all documents to all words.

Tap

Extracts knowledge bases from unstructured and semistructured bodies of text

Swoogle

Crawler based indexing and information retrieval system.

Swoogle architecture has four major components: SWD discovery, metadata creation, data analysis and interface.

OntologyRank based on google's PageRank.

Chapter 13

Semantic patterns and adaptive software

Patterns in software design

Becoming a master chess player:

- First, learn the rules: names of pieces, legal moves, chess board, etc.
- Second, learn the basic principles: piece value, center squares, etc.
- Third, study the games of the masters, and learn and apply them.

Same goes for becoming a master of software design!

Design patterns

Pattern frame:

Everything surrounding the pattern: architecture, design, implementation.
Database, application, deployment, infrastructure.

A bit of history:

Originated with Christopher Alexander during the late 1970s.

Pattern movement stayed quiet for a while until it reappeared at a conference.

Since then, it's importance and popularity has grown.

Patterns defined

“A pattern describes a recurring problem that occurs in a given context and based upon a set of guiding forces recommends a solution.”

Whenever a problem occurs, the same problem will probably occur again some time. A pattern will simplify the process of solving such problems.

The solution isn't necessarily a set of instructions, but more high-level abstractions and architectural guidelines.

Semantic patterns

May be useful as a communication tool between developers and designers.

The goal is to find characteristics of semantic models that remain constant, which is hard, because different languages have different models.

Useful patterns for semantics could include upper ontologies and domain ontologies, closure axioms and open world reasoning, n-ary relations, and classes as values.

There is a need for a technique of describing new semantic primitives at a higher level of abstraction.

Semantic patterns contd.

Building the basic idea of semantic patterns on the Web requires technical representation and the establishment of semantic pattern libraries.

SemPL (**S**emantic **P**atterns **L**anguage) - a model for efficient coding of semantics.

Self-organizing and adaptive software

Self-organizing software: A network's ability to organize and configure itself according to environmental conditions.

Adaptive software: The ability of applications to learn and adapt to the changing conditions in the network.

Genetic algorithms: A model for machine learning in which a population of randomly created units goes through a selection process of evolution.

Learning algorithm: A process that takes a data set from a database as input and after performing an algorithmic operation returns an output statement representing learning