IT452 Advanced Web and Internet

Set11
Search Engines & SEO

Outline

• How do search engines work?
  – Basic operation
  – What makes a good one?
  – What makes it difficult?
• Web Design with search engines in mind
Search Engines – Basic Operation

• Crawler

• Indexer

• Query Engine

Crawler

• How does it find the pages?

• Does it crawl everything?

• How fast does it crawl?
Indexer

- Parse document
- Remember
  - Whole text
  - Words
  - Phrases
  - Link text
- Builds an “inverted index”

Query Engine

- Process text query from user
- Return ranked set of hopefully relevant pages
- Ranking factors
  - 1. Query-specific
  - 2. Page-specific
  - 3.
PageRank

• Original basis of Google – still important
• Two interpretations:
  – Random walk
  – Pages voting

• Does it depends on the query?

SEO

• Goal

• What does it consider?

• Types
SEO 0.1

- Early search engines heavily dependent on meta tags
  - What to do?
    - White hat:
    - Black hat:
  - Key issue: easy to___________________________

SEO 1.0

- Modern search engines depend heavily on links
  - What to do?
    - White hat:
    - Black hat:
Good principles

- Clear hierarchy
- Links to all pages (static), not as images
- Useful content
- Links from relevant sites
- Good title / alt / meta
- Limit dynamically generated pages (or # args)
- No broken links, < 100 links
- Use robots.txt – exclude internal search results
- Fresh content

Bad principles

- Stuff with lots of irrelevant content
- Show different version of content to crawler
- Link schemes, farms
- Hidden text and links
- Pages designed just for search engines, not users
- Automated querying
- Deception in general