

Information Retrieval: aka “Google-lite”

CMSC 11500

Introduction to Computer Programming

November 27, 2002

Roadmap

- Information Retrieval (IR)
 - Goal: Match Information Need to Document Concept
 - Solution: Vector Space Model
 - Representation of Documents and Queries
 - Computing Similarity
- Implementation:
 - Indexing: Documents \rightarrow Vectors
 - Query Construction: Query \rightarrow Vector
 - Retrieval: Finding “Best” match: Query/Document

The Information Retrieval Task

- Goal:
 - Match the information need expressed by user
 - (the Query)
 - With concepts in documents
 - (the Document collection)
- Issues:
 - How do we represent documents and queries ?
 - How do we know if they're “similar”? Match?