

Course: Data and Information Literacy

Lecture: 4 Answers to Self-Assessment Questions

Lecturer: Dr. Johnson Masinde

1. How do indexing techniques influence the efficiency and effectiveness of information retrieval systems in locating relevant documents based on user queries?

Indexing techniques play a crucial role in the efficiency and effectiveness of information retrieval systems by organizing and structuring large collections of documents for faster and more accurate retrieval. Efficient indexing reduces the time required to search for relevant documents by creating a structured representation of document contents, typically through keywords or terms. For example, inverted indexing—one of the most common techniques—maps each keyword to the documents in which it appears, enabling quick lookups. Effective indexing also enhances precision and recall by capturing important aspects of documents (e.g., term frequency, metadata) that help the system locate the most relevant documents based on a user's query. Without proper indexing, retrieval systems would have to scan through each document in full, slowing the process and reducing the system's ability to deliver relevant results efficiently.

2. What are some common methods used to implement query expansion in information retrieval systems?

- **Synonym Expansion:** The system adds synonyms or similar terms to the query. For example, if the user searches for "car," the system might add "automobile" to retrieve a broader set of results.
- **Relevance Feedback:** Based on an initial set of search results, the user identifies relevant documents, and the system expands the query using terms from those documents.
- **Thesaurus-Based Expansion:** The system uses a predefined thesaurus or ontology to find related terms. This approach ensures semantic expansion, often based on hierarchical relationships (e.g., broadening a query on "dog" to include "canine").

- **Automatic Term Generation:** Machine learning techniques, such as word embeddings, generate additional terms that are statistically related to the query.
- **Contextual Expansion:** The system adds terms based on the context in which the original query is being made, incorporating domain-specific or situational terms.

3. In what ways do ranking algorithms impact the user experience in information retrieval systems?

Ranking algorithms directly affect the user experience by determining the order in which search results are presented. Well-designed ranking algorithms prioritize relevance, ensuring that the most pertinent and useful documents appear at the top of the search results, thus improving user satisfaction. If the ranking is accurate, users can quickly find what they are looking for without needing to scroll through multiple pages of less relevant results. For example, algorithms like **TF-IDF (Term Frequency-Inverse Document Frequency)** and **BM25** rank documents based on term relevance, while more advanced algorithms like **PageRank** or **machine learning-based ranking** (e.g., Google's RankBrain) incorporate factors like user behavior, authority, and content freshness. Poor ranking can lead to user frustration if irrelevant or low-quality results dominate the top positions, negatively affecting trust in the system's ability to meet user needs. Therefore, ranking impacts both efficiency and satisfaction, influencing the overall usability and success of the system.