

### **1. What is Information Retrieval?**

Information Retrieval (IR) is the process of obtaining relevant information from a large repository, typically in response to a user query. IR systems are designed to search, retrieve, and rank documents (or other types of information) based on their relevance to user-provided queries. It's what powers search engines like Google. When you search "best pizza places near me" on Google, the search engine retrieves documents (web pages, reviews, maps) that are likely to answer your question. As the amount of digital content grows, effective information retrieval becomes increasingly essential for users to find what they need quickly and accurately.

### **2. What are some of the challenges in Information Retrieval?**

- Query ambiguity
- Synonymy and polysemy
- Scalability with big data
- Ranking relevance
- Multilingual retrieval
- User intent understanding

Users may input vague or ambiguous queries, and words can have multiple meanings. Systems must handle diverse languages, massive data, and return the most relevant results. The word "Java" could refer to a programming language, coffee, or even a drink. Figuring out what the user wants is challenging. Overcoming these challenges is key to providing accurate and fast search results, especially in critical domains like legal, medical, or academic search.

### **3. Name some of the research work that can be done under the area of Information Retrieval?**

- Semantic search improvement
- Personalization in IR
- Voice-based IR systems
- Cross-lingual IR
- Neural IR models using deep learning
- IR in low-resource languages

These research areas aim to improve how IR systems understand user intent, context, and language, especially with AI and machine learning advancements. Using transformer models (like BERT) to improve Google's search result accuracy is a recent research success in IR.

Continuous research helps evolve IR systems to handle modern user needs, like searching via voice assistants or in multiple languages.

#### **4. Can we address both the issues of effectiveness and efficiency of an IR system at the same time? Why?**

It's possible, but it involves trade-offs. Effectiveness refers to how well the system retrieves relevant documents. Efficiency refers to how *quickly* it retrieves them. Improving one often affects the other. Highly effective methods (e.g., deep learning) may be computationally expensive, reducing efficiency. Google uses a combination of fast keyword-based methods for initial filtering (efficient) and then applies more complex neural re-ranking on the top results. With hybrid approaches—like multi-stage ranking systems—we can balance both goals. Advances in hardware and optimized algorithms also help.