Recap

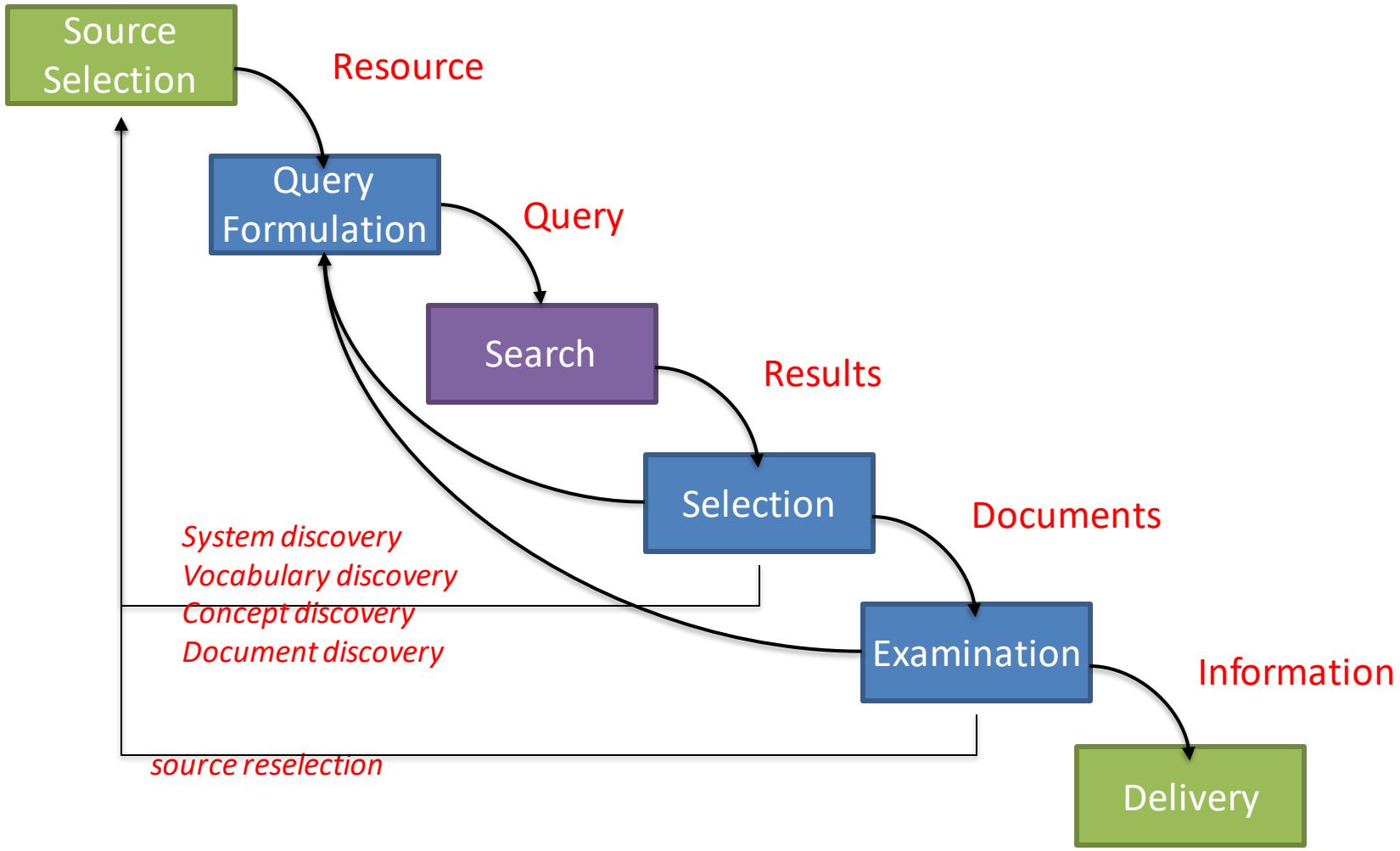
- Unstructured content is increasing so is commercialization (market cap) of such content
- IR is a science of finding material (documents) of unstructured nature that satisfies some information need within large collections
- Search engines are the Key to the knowledge kingdom
- Role of paid search
- Search wars – conventional and new set of players
- History/Evolution of Search engines

Evolution of Search Engines

2

- Crawling and Indexing
- Topic directories
- Clustering and Classification
- Hyperlink analysis
- Resource discovery and vertical portals
- Semantic Web
- ???

information retrieval cycle



The central problem in search

Searcher

Author

Concepts

Concepts

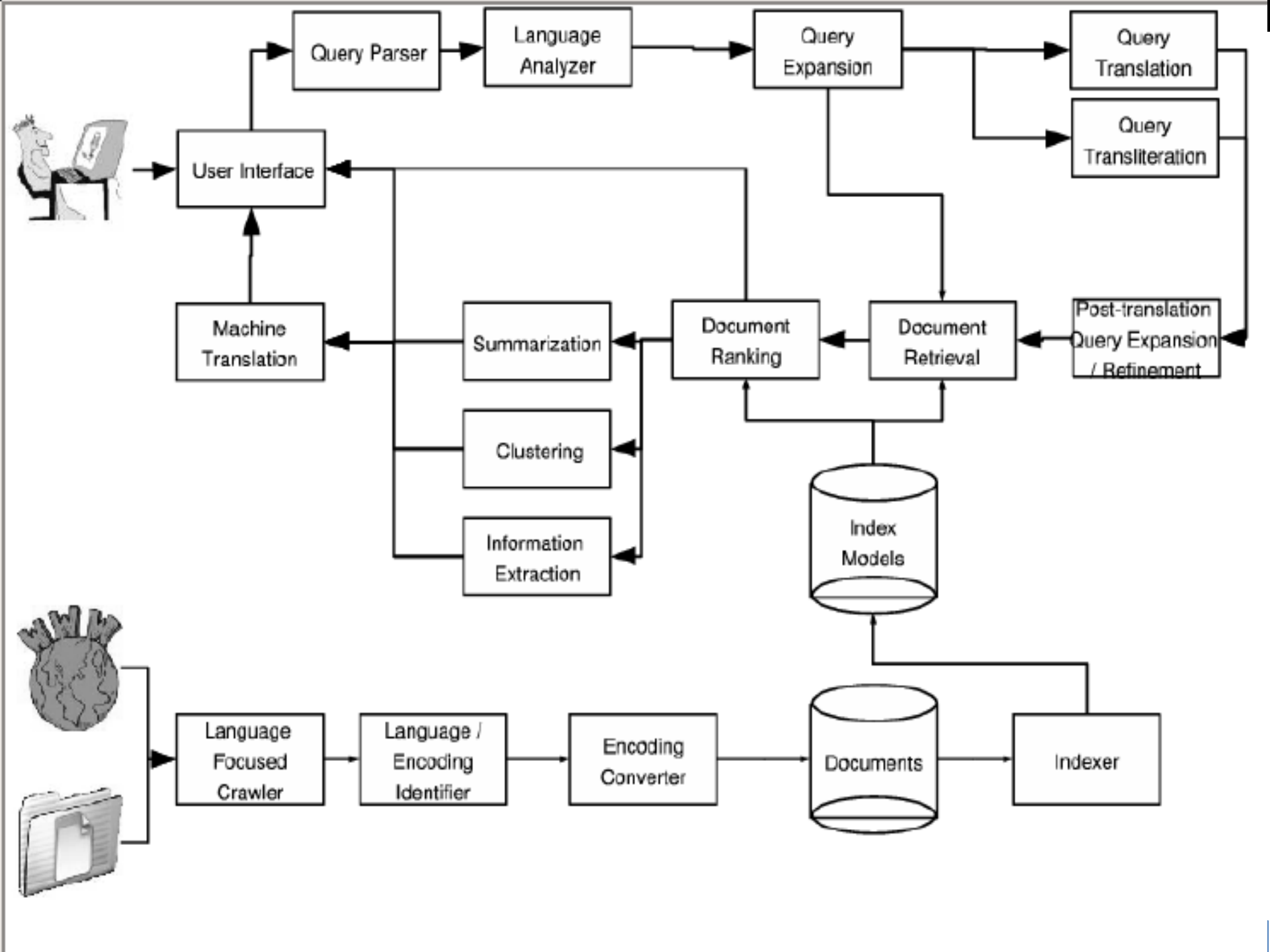
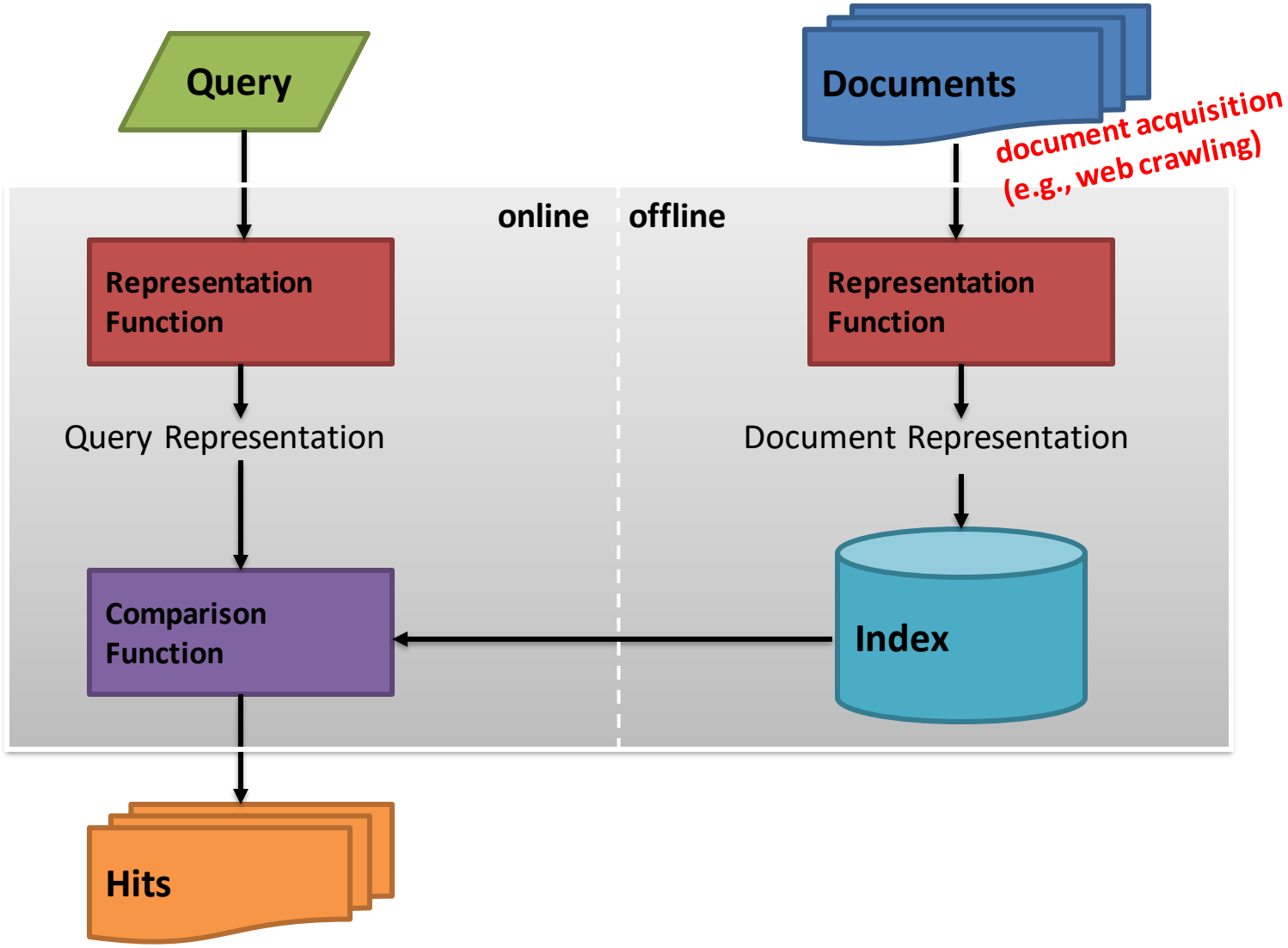
Query Terms

"tragic love story"

Document Terms

"fateful star-crossed romance"

Do these represent the same concepts?


Query Parser
Language Analyzer
Query Expansion
Query Translation
Query Transliteration
User Interface
Machine Translation
Summarization
Document Ranking
Document Retrieval
Post-translation Query Expansion / Refinement
Clustering
Information Extraction
Index Models
Language Focused Crawler
Language / Encoding Identifier
Encoding Converter
Documents
Indexer

abstract ir architecture