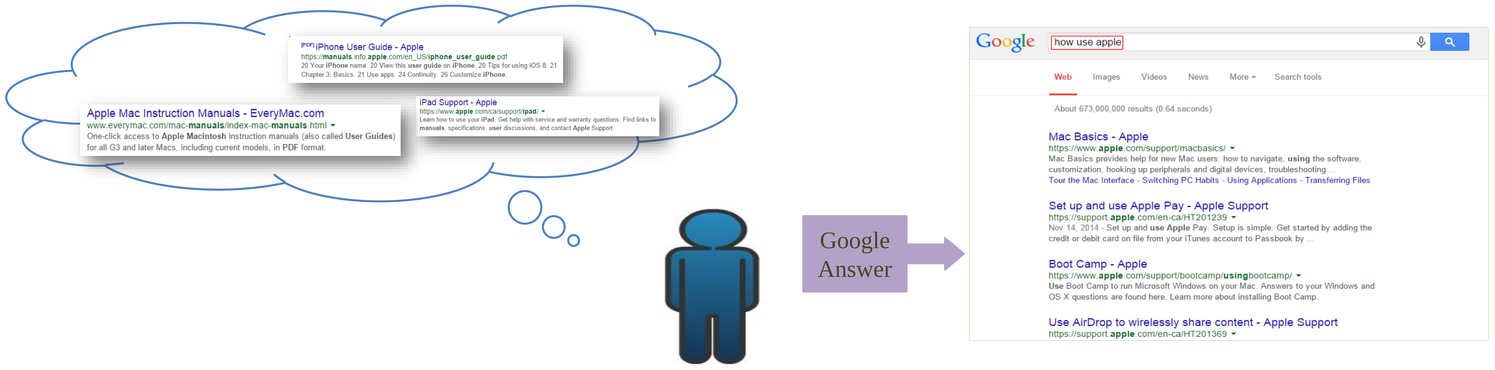
Indexing Infrastructure for Semantic Search

Fatemeh Lashkari and Ebrahim Bagheri

Introduction

Semantic search considers the **intent** and **contextual meaning** of terms in the corpus and query in order to improve the accuracy of the search.

In semantic search, a **combination** of **words**, **formal knowledge** (e.g. Fredericton is a location), and **semantic annotations** (e.g. Fredericton is in Canada) is indexed.

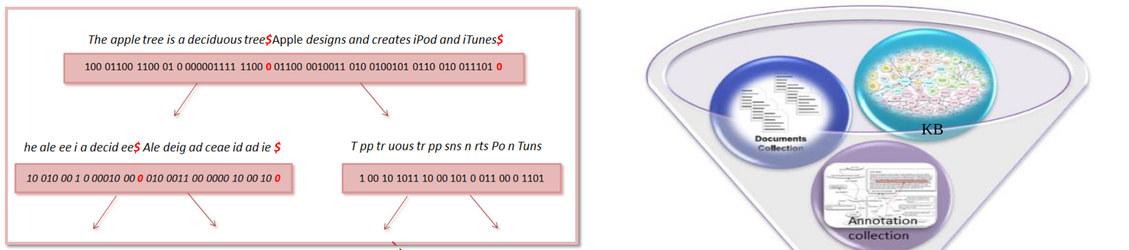


Approach

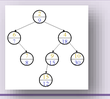
❖ Semantic Index

Our semantic index, named ***Devine index***, consists of **keyword**, **entity** and **type indices. T**hese three indices are designed based on **wavelet tree** and **inverted index** structures.

Type index stores **is-a** relations between concepts. Entity index maps each entity to its statistical information based on semantic annotations.



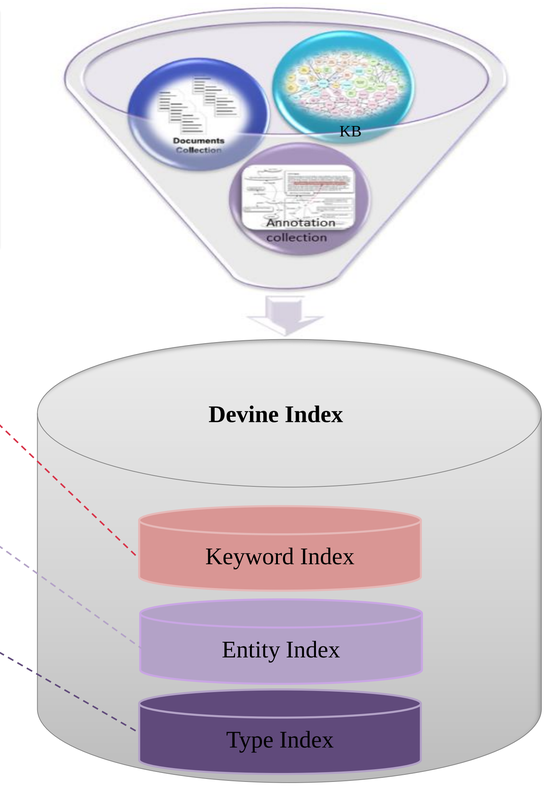
Wavelet Tree

Apple	
Apple Inc.	

Inverted Index
Posting list is a ***Treap***

Fruit	{apple, orange, ..}
Company	{Apple Inc., Samsung,….}

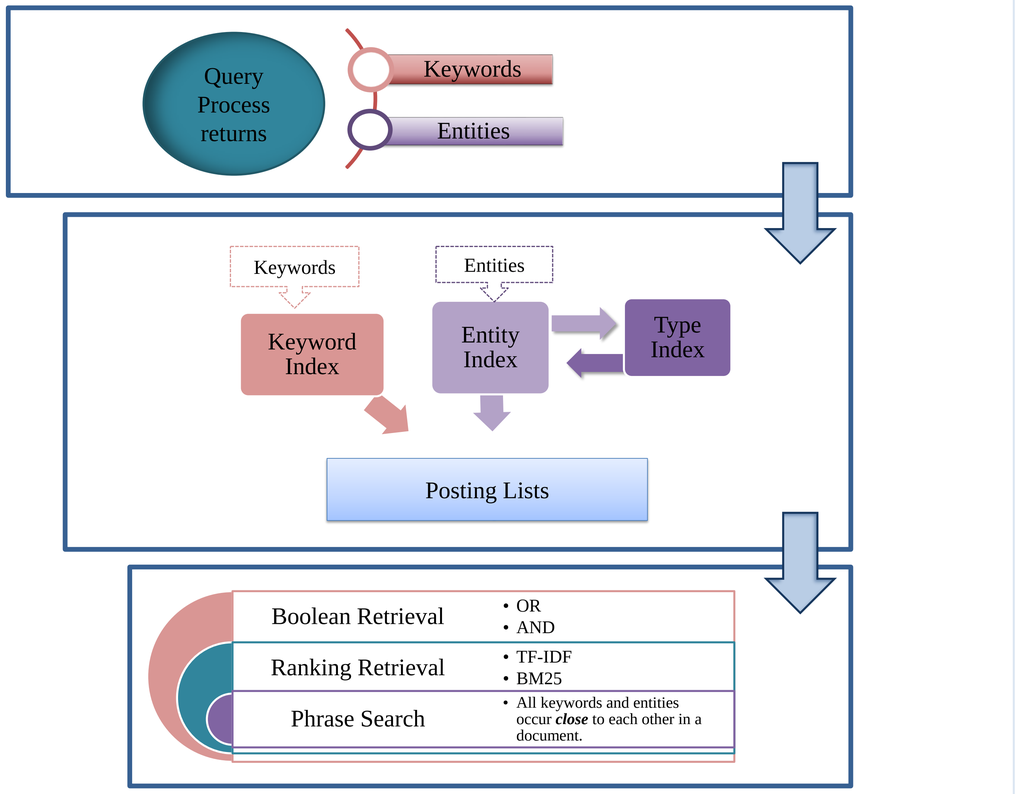
Inverted Index
Posting list is a list of ***entities***



❖ Query Retrieval

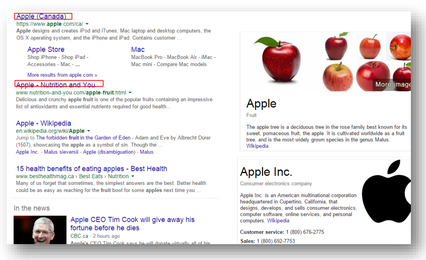
Devine index supports **Boolean retrieva**l, **Ranking retrieval, Ranking AND**, and **Phrase search** efficiently with respect to space usage.

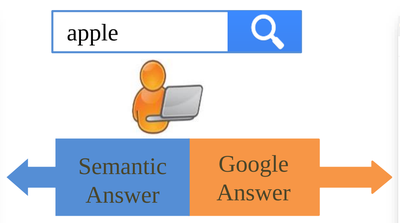
Documents are retrieved on the basis of relevance to ontological concepts, along with keywords.

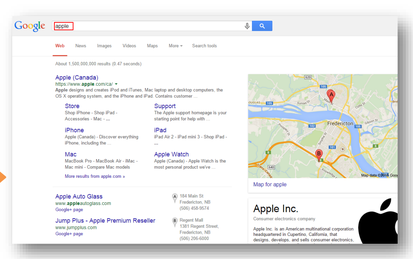


Conclusion

- ❖ Devine index can play the role of an indexer for **Document oriented semantic search** and **entity search**.
- ❖ Devine index is evaluated based on **space usage** and **query response time**.
- ❖ Devine index is an **in-memory** index structure.







Faculty of COMPUTER SCIENCE
UNB Fredericton · New Brunswick · Canada
RYERSON Laboratory for Systems, Software and Semantics
Intelligent and Adaptive Systems
2015 Faculty of Computer Science Research Expo. April 10, Wu Conference Center, University of New Brunswick