



TECHNICAL TRAINING KNOWLEDGE ARCHITECTURE AN004

FORREST SHENG BAO, PI – UNIV.AKRON
SHENGYONG WANG, CO-PI – UNIV.AKRON

FAA TECHNICAL MONITOR
GRETA BALLENTINE

PROJECT OVERVIEW



- Mainly literature review based tasks.
 - We will test some methods on the materials provided by FAA.
 - Thorough literature review on machine learning, natural language processing, information retrieve.
- Six-month plan in initial proposal (before knowing detailed expectation from FAA):
 - Task A: by 11/20/2016, understand FAA's needs
 - Task B: by 12/20/2016, Information extraction literature review
 - Task C: by 2/1/2017, Knowledge discovery literature review
 - Task D: by 3/1/2017, knowledge representation and reasoning literature review

PROGRESS UPDATE



- We are behind the schedule.
 - An extension is preferred.
- We have presented our first set of slides to our technical monitor in late January 2017 and are waiting for the feedback.
- Materials that need to be searched on is expected from FAA in a near future.
- This is related to another project of us.
- To be continued.

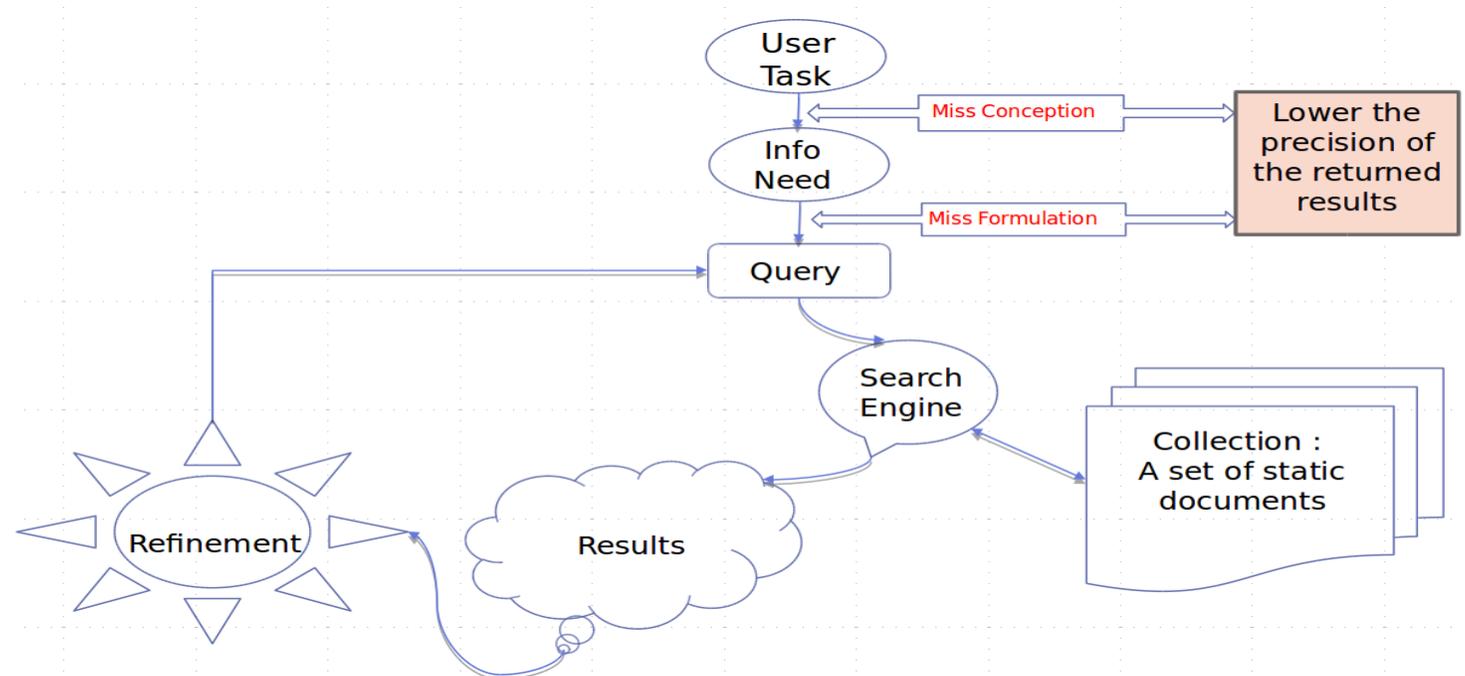
Information Retrieval

- Information retrieval is finding a material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within a large collections (usually stored in computers).

- **Cases of usage:**

- E-mail search
- Web search
- Corporate knowledge basis
- Legal information retrieval

Process flow



Google Experience

- Google turned its web search over to Artificial Intelligence using the RankBrain algorithm in 2015.
 - Google changes its search algorithm almost *500-600* times a year.
 - Google is using at least 200 ranking factors in their algorithms including social signals and trust. Google's algorithm is more complex and volatile than ever.
 - *RankBrain* uses artificial intelligence to embed vast amounts of written language into mathematical entities, called vectors, that the computer can understand.
 - If Rank Brain sees a word or phrase it isn't familiar with, the machine can make a guess as to what words or phrases might have a similar meaning and filter the result accordingly, making it more effective at handling never-before-seen search queries.
 - RankBrain has become the third-most important signal contributing to the result of a search query.

Semantic and Keyword Search

- The main difference between keyword and semantic based search engine is the time consumption. User takes time to sort the retrieved result in keyword engine whereas in case of semantic based search engine there is no sorting of search results.
- Keyword search uses the Conventional Information Retrieval technology which is based on the occurrence of words in documents. Therefore it's difficult to get a relevant results.
- Semantic search can deal with polysemy and synonymous words while the keyword search can't.
- Precision and recall evaluation values for the keyword search are low.

Query Expansion

- Query Expansion:

The issue of synonymy has an impact on the recall of most information retrieval systems.

- For example, aircraft and plane, thermodynamics and heat have the same meaning.

Methods tackling this problem: Expanding or reformulating query terms independent of the query and results returned from it, so that changes in the query wording will cause the new query to match other semantically similar terms.