

Document processing

- Preprocessing: detailed analysis of all texts in the corpus may be done a priori
 - e.g., annotate + index terms with semantic tags
- Retrieval: initial set of candidate answer-bearing documents are selected from a large collection
 - boolean retrieval methods may be used profitably
 - passage retrieval may be more appropriate
- Analysis:
 - part of speech (POS) tagging
 - named entity identification (NEI): recognizes multi-word strings as names of companies/persons, locations/addresses, quantities, etc.
 - may use *authority lists* (table lookup), e.g., WordNet, CIA World Factbook
 - shallow/deep syntactic analysis: obtains information about syntactic relations, semantic roles

Answer type identification \longleftrightarrow NEI