Question Answering

Complex Questions: Summarizing Multiple Documents

Definition questions

Q: What is *water spinach?*

A: Water spinach (ipomoea aquatica) is a semi-aquatic leafy green plant with long hollow stems and spear- or heart-shaped leaves, widely grown throughout Asia as a leaf vegetable. The leaves and stems are often eaten stir-fried flavored with salt or in soups. Other common names include *morning glory vegetable, kangkong* (Malay), *rau muong (Viet.), ong choi (Cant.), and kong xin cai* (Mand.). It is not related to spinach, but is closely related to sweet potato and convolvulus.

Medical questions

Demner-Fushman and Lin (2007)

Q: In children with an acute febrile illness, what is the efficacy of single medication therapy with acetaminophen or ibuprofen in reducing fever?

A: Ibuprofen provided greater temperature decrement and longer duration of antipyresis than acetaminophen when the two drugs were administered in approximately equal doses. (*PubMedID: 1621668, Evidence Strength: A*)

Other complex questions

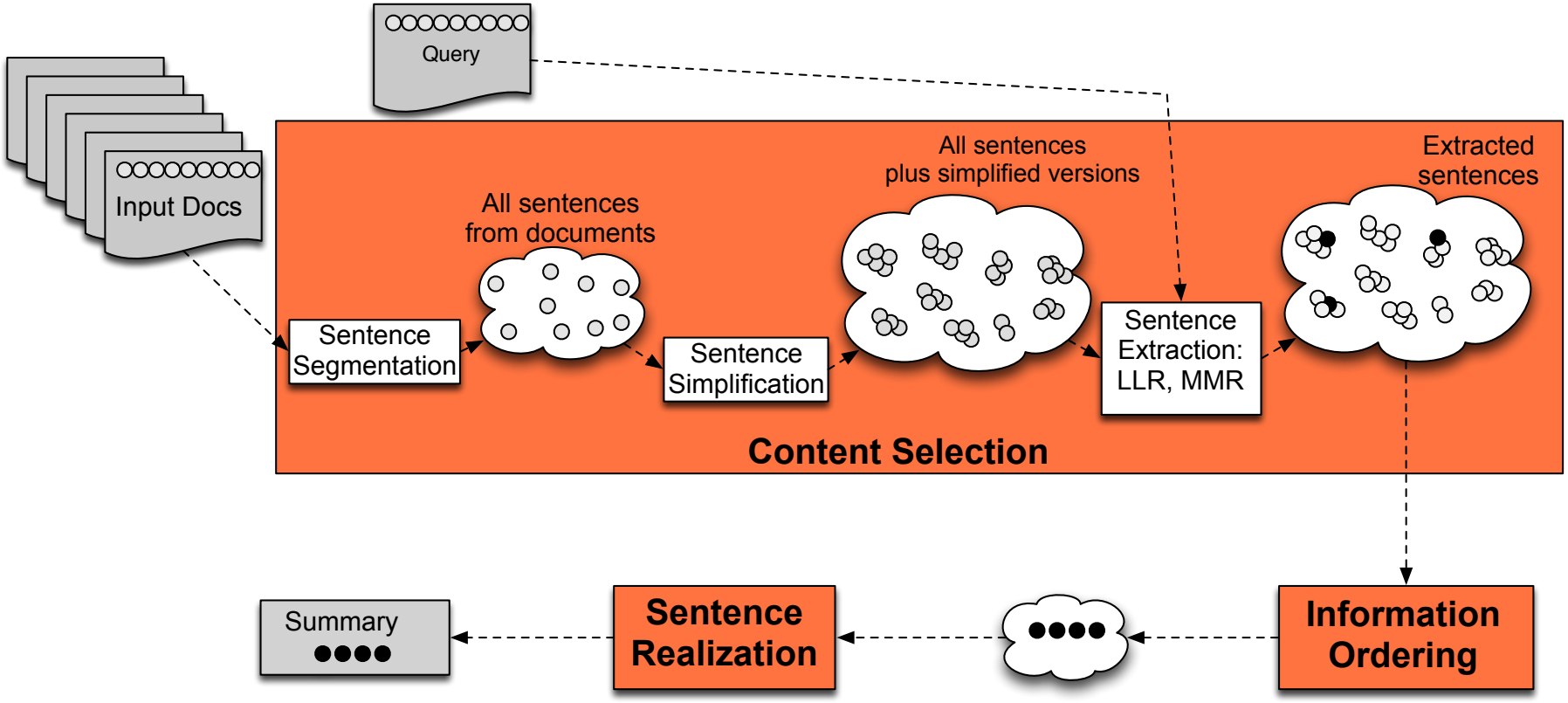
Modified from the DUC 2005 competition (Hoa Trang Dang 2005)

1. How is compost made and used for gardening (including different types of compost, their uses, origins and benefits)?
2. What causes train wrecks and what can be done to prevent them?
3. Where have poachers endangered wildlife, what wildlife has been endangered and what steps have been taken to prevent poaching?
4. What has been the human toll in death or injury of tropical storms in recent years?

Answering harder questions: Query-focused multi-document summarization

- The (bottom-up) snippet method
  - Find a set of relevant documents
  - Extract informative sentences from the documents
  - Order and modify the sentences into an answer
- The (top-down) information extraction method
  - build specific answerers for different question types:
    - definition questions
    - biography questions
    - certain medical questions

Query-Focused Multi-Document Summarization



Simplifying sentences

Zajic et al. (2007), Conroy et al. (2006), Vanderwende et al. (2007)

Simplest method: parse sentences, use rules to decide which modifiers to prune (more recently a wide variety of machine-learning methods)

appositives	Rajam, ~~28, an artist who was living at the time in Philadelphia~~, found the inspiration in the back of city magazines.
attribution clauses	Rebels agreed to talks with government officials, ~~international observers said Tuesday.~~
PPs without named entities	The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases [~~PP to a sustainable number~~]]
initial adverbials	"~~For example~~", "~~On the other hand~~", "~~As a matter of fact~~", "~~At this point~~"

Maximal Marginal Relevance (MMR)

Jaime Carbonell and Jade Goldstein, The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries, SIGIR-98

- An iterative method for content selection from multiple documents
- Iteratively (greedily) choose the best sentence to insert in the summary/answer so far:
  - Relevant: Maximally relevant to the user's query
    - high cosine similarity to the query
  - Novel: Minimally redundant with the summary/answer so far
    - low cosine similarity to the summary

$$\hat{s}_{MMR} = \max_{s \in D} \lambda sim(s,Q) \; - \; (1\text{-}\lambda) \max_{s \in S} sim(s,S)$$

- Stop when desired length

LLR+MMR: Choosing informative yet non-redundant sentences

- One of many ways to combine the intuitions of LLR and MMR:

1. Score each sentence based on LLR (including query words)
2. Include the sentence with highest score in the summary.
3. Iteratively add into the summary high-scoring sentences that are not redundant with summary so far.

Information Ordering

- Chronological ordering:
  - Order sentences by the date of the document (for summarizing news)..

  (Barzilay, Elhadad, and McKeown 2002)
- Coherence:
  - Choose orderings that make neighboring sentences similar (by cosine).
  - Choose orderings in which neighboring sentences discuss the same entity (Barzilay and Lapata 2007)
- Topical ordering
  - Learn the ordering of topics in the source documents

Domain-specific answering: The Information Extraction method

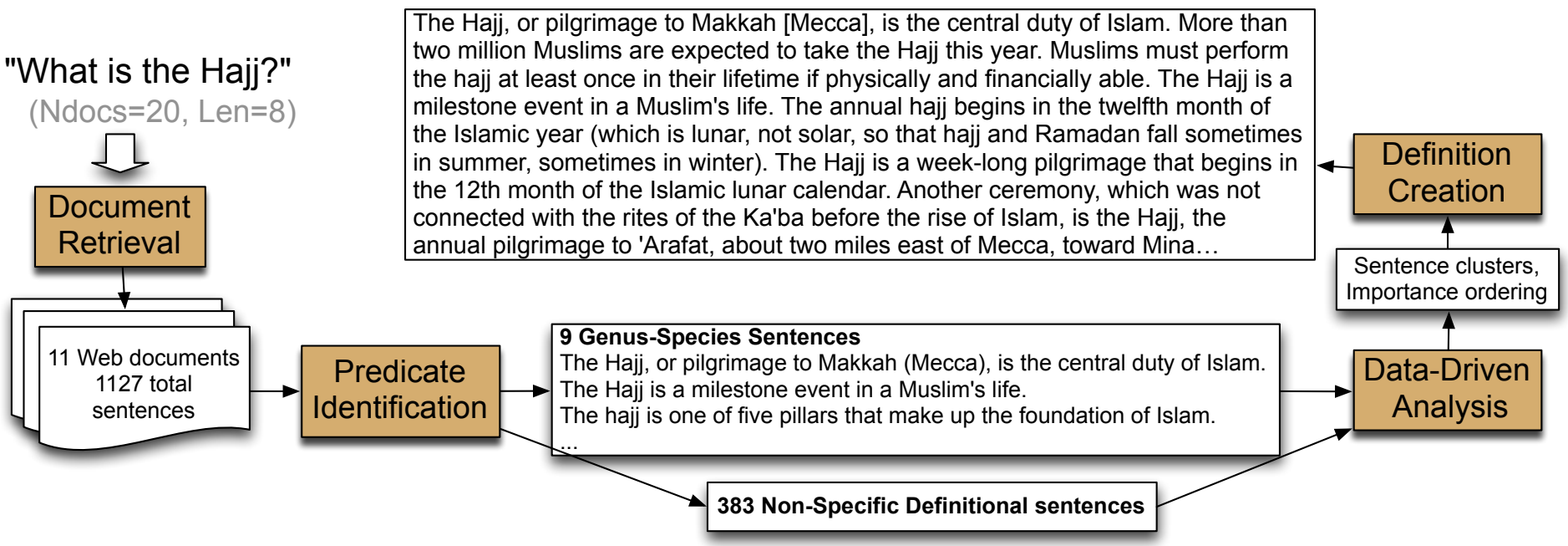
- a good **biography** of a person contains:
  - a person's **birth/death, fame factor, education, nationality** and so on
- a good **definition** contains:
  - **genus** or **hypernym**
    - *The Hajj is a type of ritual*
- a **medical answer about a drug's use** contains*:*
  - **the problem** (the medical condition),
  - **the intervention** (the drug or procedure), and
  - the **outcome** (the result of the study)**.**

Information that should be in the answer for 3 kinds of questions

Definition	
genus	The Hajj is a type of ritual
species	the annual hajj begins in the twelfth month of the Islamic year
synonym	The Hajj, or Pilgrimage to Mecca, is the central duty of Islam
subtype	Qiran, Tamattu', and Ifrad are three different types of Hajj
Biography	
dates	was assassinated on April 4, 1968
nationality	was born in Atlanta, Georgia
education	entered Boston University as a doctoral student
Drug efficacy	
population	37 otherwise healthy children aged 2 to 12 years
problem	acute, intercurrent, febrile illness
intervention	acetaminophen (10 mg/kg)
outcome	ibuprofen provided greater temperature decrement and longer duration of antipyresis than acetaminophen when the two drugs were administered in approximately equal doses

Architecture for complex question answering: definition questions

S. Blair-Goldensohn, K. McKeown and A. Schlaikjer. 2004. Answering Definition Questions: A Hybrid Approach.



Question Answering

Answering Questions by Summarizing Multiple Documents