Summarization in Question Answering





Text Summarization

Goal: produce an abridged version of a text that contains information that is important or relevant to a user.

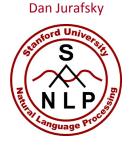
Summarization Applications

- outlines or abstracts of any document, article, etc.
- **summaries** of email threads
- action items from a meeting
- simplifying text by compressing sentences



What to summarize? Single vs. multiple documents

- Single-document summarization
 - Given a single document, produce
 - abstract
 - outline
 - headline
- Multiple-document summarization
 - Given a group of documents, produce a gist of the content:
 - a series of news stories on the same event
 - a set of web pages about some topic or question



Query-focused Summarization & Generic Summarization

- Generic summarization:
 - Summarize the content of a document
- Query-focused summarization: 類似第20章提到的Complex Question
 - summarize a document with respect to an information need expressed in a user query.
 - a kind of complex question answering:
 - Answer a question by summarizing a document that has the information to construct the answer

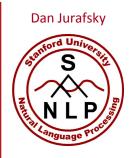




Summarization for Question Answering: Snippets

- Create snippets summarizing a web page for a query
 - Google: 156 characters (about 26 words) plus title and link

Canala		
Google	what is die brücke?	
Search	About 5,910,000 results (0.28 seconds)	
	這些預覽文字就是snippets,算是這個page的摘要	
Everything	Die Brücke - Wikipedia, the free encyclopedia	
Images	en.wikipedia.org/wiki/Die_Brücke	
	Die Brücke (The Bridge) was a group of German expressionist artists formed in Dresden in 1905, after which the Brücke Museum in Berlin was named. Founding	
Maps	You've visited this page 5 times. Last visit: 4/16/12	
Videos		
News	Die Brücke (film) - Wikipedia, the free encyclopedia	
News	en.wikipedia.org/wiki/ Die_Brücke_ (film)	
Shopping	Die Brücke (English: The Bridge) is a 1959 West German film directed by Austrian filmmaker Bernhard Wicki. It is based on the eponymous 1958 novel by	
Applications		
More	Die Brucke - Die Brucke Art	
MOLG	www.huntfor.com/arthistory/c20th/diebrucke.htm	
	Die Brucke was the association of artist expressionists from Dresden, Germany Die	
Con Francisco	Brucke made use of a technique that was controlled, intentionally	



Summarization for Question Answering: Multiple documents

Create answers to complex questions summarizing multiple documents.

- Instead of giving a snippet for each document
- Create a cohesive answer that combines information from each document



Extractive summarization & Abstractive summarization

- Extractive summarization: 從原文中擷取關鍵字,變成摘要
 - create the summary from phrases or sentences in the source document(s)
- Abstractive summarization:
 - express the ideas in the source documents using (at least in part) different words



Simple baseline: take the first sentence

第一行通常是Topic Sentence

也就是整段文字的摘要

Google

可以當作是Summarization的Baseline what is die brücke?

Search

About 5,910,000 results (0.28 seconds)

Everything

Die Brücke - Wikipedia, the free encyclopedia

Images

en.wikipedia.org/wiki/Die_Brücke

Mans

Die Brücke (The Bridge) was a group of German expressionist artists formed in Dresden in 1905, after which the Brücke Museum in Berlin was named. Founding ...

Die Brücke

From Wikipedia, the free encyclopedia

For other uses, see Die Brücke (disambiguation).

Die Brücke (The Bridge) was a group of German expressionist artists formed in Dresden in 1905, after which the Brücke Museum in Berlin was named. Founding members were Fritz Bleyl, Erich Heckel, Ernst Ludwig Kirchner and Karl Schmidt-Rottluff. Later members were Emil Nolde, Max Pechstein and Otto Mueller. The seminal group had a major impact on the evolution of modern art in the 20th century and the creation of expressionism.^[1]

Die Brücke is sometimes compared to the Fauves. Both movements shared interests in primitivist art. Both

Summarization in Question Answering

Generating Snippets and other Single-Document Answers



Snippets: query-focused summaries

Was cast-metal movable type invented in korea?

About 591,000 results (0.14 seconds)

Movable type - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Movable_type

Jump to <u>Metal movable type</u>: Transition from wood type to <u>metal</u> type occurred in 1234 ... The following description of the **Korean** font <u>casting</u> ... In the early fifteenth century, however, the **Koreans invented** a form of <u>movable type</u> that has ...

History of printing in East Asia - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/History_of_printing_in_East_Asia

The following description of the **Korean** font **casting** process was recorded by the ... While **metal movable type** printing was **invented in Korea** and the oldest ...

Korea, 1000–1400 A.D. | Heilbrunn Timeline of Art History | The ... www.metmuseum.org/toah/ht/?period=07®ion=eak

The **invention** and use of **cast-metal movable type** in **Korea** in the early thirteenth century predates by two centuries Gutenberg's **invention** of metal **movable type** ...

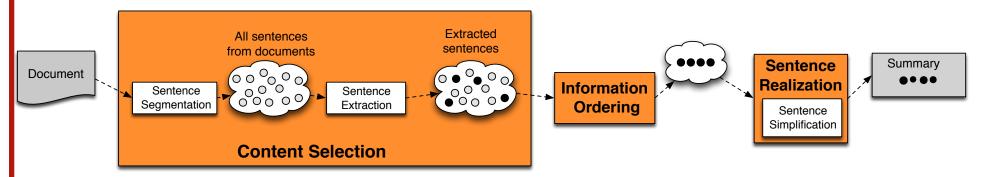




Summarization: Three Stages

基本上大多數的工作都是在第一階段 Content Selection

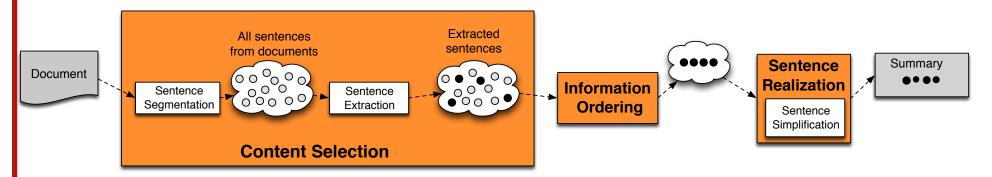
- content selection: choose sentences to extract from the document
- 2. information ordering: choose an order to place them in the summary
- 3. sentence realization: clean up the sentences





Basic Summarization Algorithm

- content selection: choose sentences to extract from the document
- 2. information ordering: just use document order
- 3. sentence realization: keep original sentences





Unsupervised content selection

H. P. Luhn. 1958. The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development. 2:2, 159-165.

把出現「關鍵字」的句子擷取出來,關鍵字可以用TF-IDF去判斷,越高就代表越可能是關鍵字

- Intuition dating back to Luhn (1958):
 - Choose sentences that have salient or informative words
- Two approaches to defining salient words
 - 1. tf-idf: weigh each word w_i in document j by tf-idf $weight(w_i) = tf_{ij} \times idf_i$
 - 2. topic signature: choose a smaller set of salient words
 - mutual information
 - log-likelihood ratio (LLR) Dunning (1993), Lin and Hovy (2000)

$$weight(w_i) = \begin{cases} 1 & \text{if } -2\log\lambda(w_i) > 10\\ 0 & \text{otherwise} \end{cases}$$



Topic signature-based content selection with queries

Conroy, Schlesinger, and O'Leary 2006

- choose words that are informative either
 - by log-likelihood ratio (LLR)
 - or by appearing in the query

把LLR高的,以及有出現在問題Query的字的weight設成1

$$weight(w_i) = \begin{cases} 1 & \text{if } -2\log\lambda(w_i) > 10 \\ 1 & \text{if } w_i \in question \\ 0 & \text{otherwise} \end{cases}$$
 (could learn more complex weights)

Weigh a sentence (or window) by weight of its words:

$$weight(s) = \frac{1}{|S|} \sum_{w \in S} weight(w)$$



Supervised content selection

- Given:
 - a labeled training set of good summaries for each document
- Align:
 - the sentences in the document with sentences in the summary
- Extract features
 - position (first sentence?)
 - length of sentence
 - word informativeness, cue phrases
 - cohesion
- Train
 - a binary classifier (put sentence in summary? yes or no)

- Problems:
 - hard to get labeled training data
 - alignment difficult
 - performance not better than unsupervised algorithms
- So in practice:
 - Unsupervised content selection is more common

訓練一個分類器

用來決定每一句話是否應該放到Summary

但問題是很難拿到Label

因為原文的Summary不會總是和原文用一樣的句子

Generating Snippets and other Single-Document Answers

Evaluating Summaries: ROUGE



ROUGE (Recall Oriented Understudy for Gisting Evaluation)

Lin and Hovy 2003

- Intrinsic metric for automatically evaluating summaries
 - Based on BLEU (a metric used for machine translation)
 - Not as good as human evaluation ("Did this answer the user's question?")
 - But much more convenient
- Given a document D, and an automatic summary X:
 - 1. Have N humans produce a set of reference summaries of D
 - 2. Run system, giving automatic summary X
 - 3. What percentage of the bigrams from the reference summaries appear in X?

$$ROUGE - 2 = \frac{\sum_{s \in \{\text{RefSummaries}\} \text{ bigrams } i \in S} \min(count(i, X), count(i, S))}{\sum_{s \in \{\text{RefSummaries}\} \text{ bigrams } i \in S} count(i, S)}$$



A ROUGE example:

Q: "What is water spinach?"

3

問題是人工的Summary用字可能會改寫原文的字 所以在設計Human Summary時,要注意盡量使用原文的字

Human 1: Water spinach is a green leafy vegetable grown in the tropics.

Human 2: Water spinach is a semi-aquatic tropical plant grown as a vegetable.

Human 3: Water spinach is a commonly eaten leaf vegetable of Asia.

• System answer: Water spinach is a leaf vegetable commonly eaten in tropical areas of Asia.

$$\frac{3+3+6}{10+9+9}$$

Σ總Bi-gram數

$$= 12/28 = .43$$

Evaluating Summaries: ROUGE

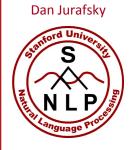
Complex Questions:
Summarizing
Multiple Documents



Definition questions

Q: What is water spinach? 假設沒辦法從 I R 取得這段文章,有辦法「產生」嗎?

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.



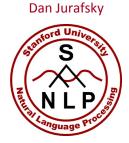
Medical questions

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?



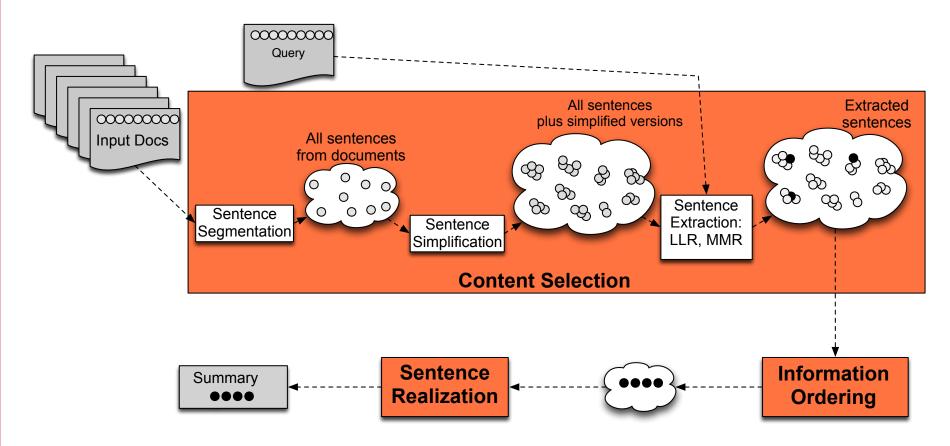


Answering harder questions: Query-focused multi-document summarization

- The (bottom-up) snippet method 找出很多相關文章 從這些文章擷取句子
 - Find a set of relevant documents 重組這些句子
 - 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 Bottom-up





Simplifying sentences

Zajic et al. (2007), Conroy et al. (2006), Vanderwende et al. (2007) 把長句改寫成意義相近的短句

用手寫Rules去做

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

appositives 刪除同位語	Rajam, 28, an artist who was living at the time in Philadelphia, found the inspiration in the back of city magazines.
attribution clauses 刪除不必要的屬性	Rebels agreed to talks with government officials, international observers said Tuesday.
PPs without named entities 刪除沒有NER的子句	The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases [PP to a sustainable number]]
initial adverbials 刪掉轉折句等	"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

上一頁會從很多文件找出需要的句子,也把這些句子改短了,下一步就是要做Ranking找出最適合放進Summary的句子

- 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

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

Stop when desired length

兩個原則挑選最適合的句子

- 1. 和User Query最相關的
- 2. 和目前Summary最不重複的

相關?重複?

用Cosine Similarity



LLR+MMR: Choosing informative yet non-redundant sentences

- One of many ways to combine the intuitions of LLR and MMR:
- Score each sentence based on LLR (including query words)
- 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

句子選好了, 要怎麼排序?

- Chronological ordering: 照日期排
 - Order sentences by the date of the document (for summarizing news)..
 (Barzilay, Elhadad, and McKeown 2002)
- Coherence:

照相近語意排

- 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

照擷取句子的原文的順序排

• 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).

要定義每一個領域可能會需要的Entity種類這樣QA系統才知道要怎麼抓關鍵字



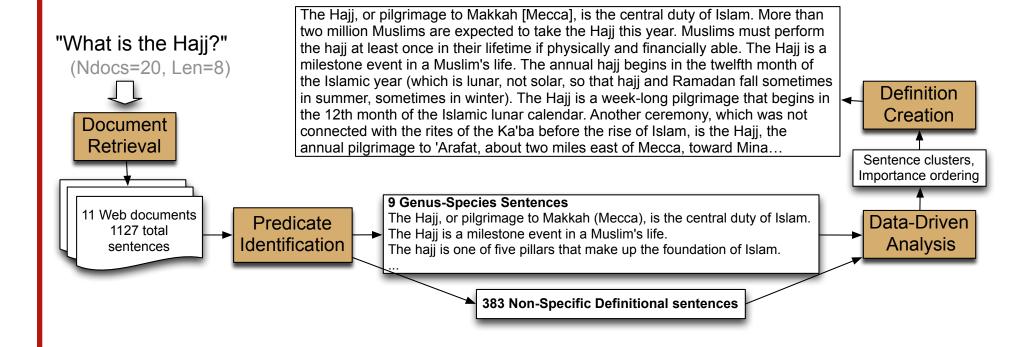
Information that should be in the answer for 3 kinds of questions

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.



Answering Questions by Summarizing Multiple Documents