CS-AD 220 – Spring 2016

Natural Language Processing

Session 28: 12-May-16

Prof. Nizar Habash

Slides from Jurafsky and Martin with modifications

Final Exam

- **Final is May 16, 1pm to 4pm in CR-002**
- **Open Book (J+M)**
- **No Internet**
- **Cheating →0% on the full exam**
  - **What is "cheating"?**

Assignment #4
Open Challenge Results

	True Case (Official)		Lower Case		
	BLEU	**1/2/3/4 grams**	**BLEU**	**1/2/3/4 grams**	**BP**
Anonymous					
MissingFreshman					
ThreeStooges					
Quanfucius					
Google	45.39	78.1/54.1/37.9/26.5	47.62	80.4/56.5/40.1/28.2	1.000

Assignment #4
Open Challenge Results

	True Case (Official)		Lower Case		
	BLEU	**1/2/3/4 grams**	**BLEU**	**1/2/3/4 grams**	**BP**
Anonymous	27.42	64.0/35.9/20.6/12.4	28.99	66.3/38.0/22.0/13.2	0.991
MissingFreshman	23.61	60.6/32.1/17.3/9.7	24.98	62.6/33.8/18.5/10.4	0.988
ThreeStooges	18.97	55.8/26.4/13.5/7.4	31.30	69.8/41.0/24.8/15.3	0.970
Quanfucius	2.62	22.9/5.4/1.7/0.4	2.68	23.7/5.6/1.7/0.4	0.868
Quanfucius2	25.29	65.6/36.1/21.0/12.5	26.47	67.7/37.7/22.1/13.3	0.900
Google	45.39	78.1/54.1/37.9/26.5	47.62	80.4/56.5/40.1/28.2	1.000

- First Place (+5%): Anonymous (Talha, Joe)
- Second Place (+4%): Missing A Freshman (Jiwon, Batu, Paula)
- Third Place (+3%): The Three Stooges (Maverick, Shantanu, Will)
- Fourth Place (+2%): Quanfucius (Quan, Kenny, Frederik)

Text Summarization

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

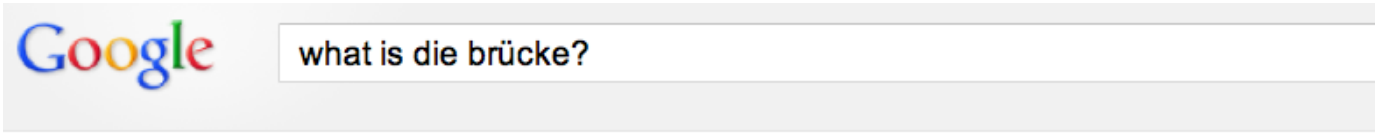
- **Single-document summarization**
  - Given a single document, produce
    - abstract
    - outline
    - headline
- **Multiple-document summarization**
  - Given a group of documents, produce a gist of the content:
    - a series of news stories on the same event
    - a set of web pages about some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

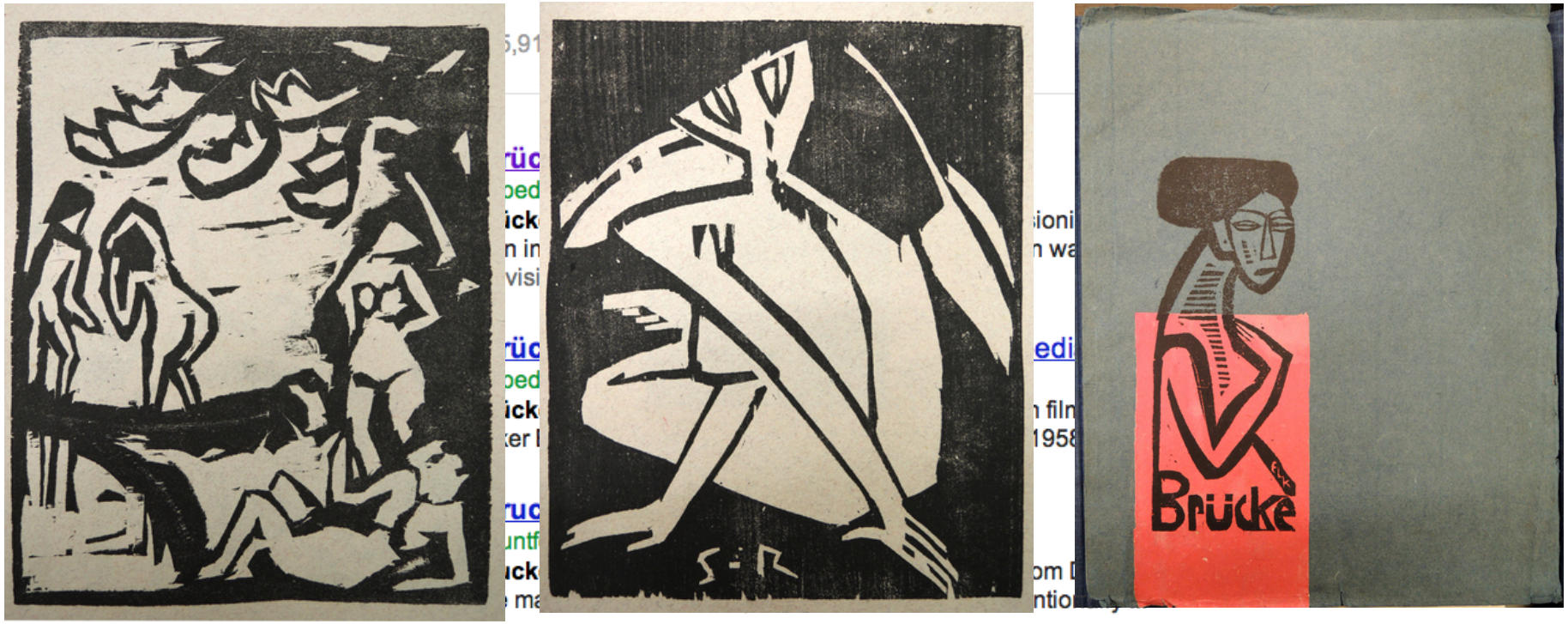
Query-focused Summarization & Generic Summarization

- Generic summarization:
  - Summarize the content of a document
- Query-focused summarization:
  - 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





Summarization for Question Answering: Multiple documents

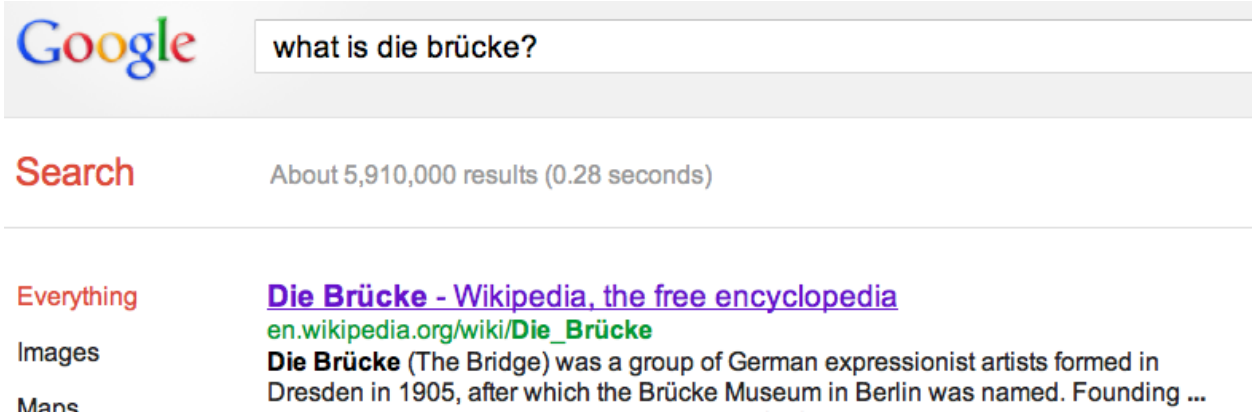
Create answers to complex questions summarizing multiple documents.

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

Extractive summarization & Abstractive summarization

- Extractive summarization:
  - 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



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

Snippets: query-focused summaries

Was cast-metal movable type invented in korea?

About 591,000 results (0.14 seconds)

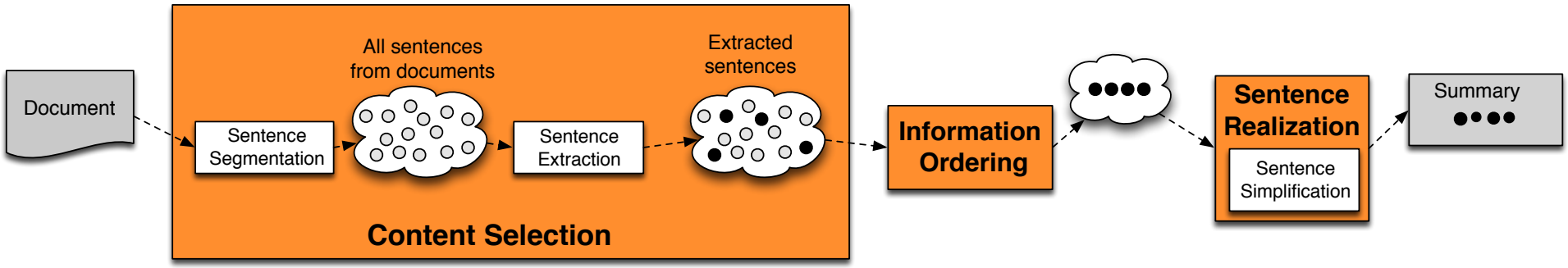
Movable type - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Movable_type**
Jump to **Metal movable type**: Transition from wood type to **metal** type occurred in 1234 **...** The following description of the **Korean** font **casting ...** In the early fifteenth century, however, the **Koreans invented** a form of **movable type** 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&region=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 ...**

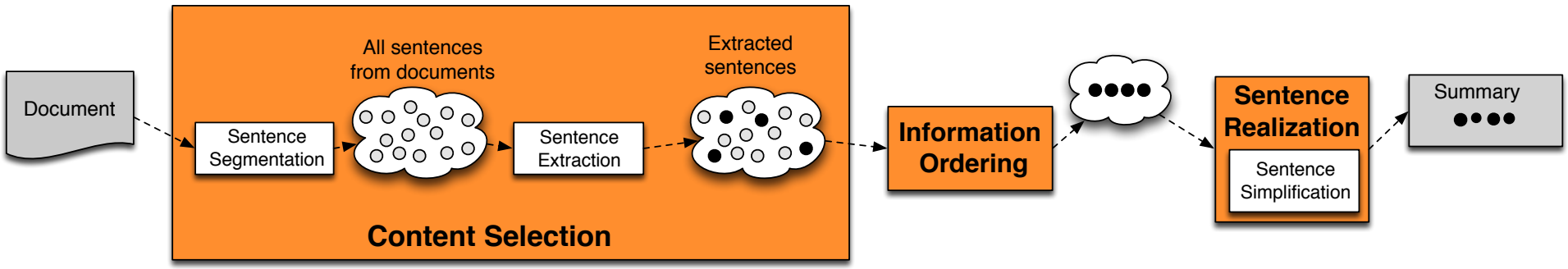
Summarization: Three Stages

1. 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

1. 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.

- 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

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

(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**

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?

$$ROUGE2 = \frac{\sum_{s \in \{\text{RefSummaries}\}} \sum_{\text{bigrams } i \in S} \min(count(i,X), count(i,S))}{\sum_{s \in \{\text{RefSummaries}\}} \sum_{\text{bigrams } i \in S} count(i,S)}$$

A ROUGE example:
Q: “What is water spinach?”

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

Human 2: Water spinach is a 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.

- ROUGE2 = $\frac{3 + 3 + 6}{10 + 9 + 9}$ = 12/28 = .43

Final Exam Review

- Open Book!
- Syntax, POS tagging
- Machine translation
- Lexical semantics, similarity measures
- Question answering, information retrieval, summarization
- NLP system building, evaluation

Thanks everyone for a wonderful semester!

Have an enjoyable and productive summer!