

[illegible]



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 ...](#)

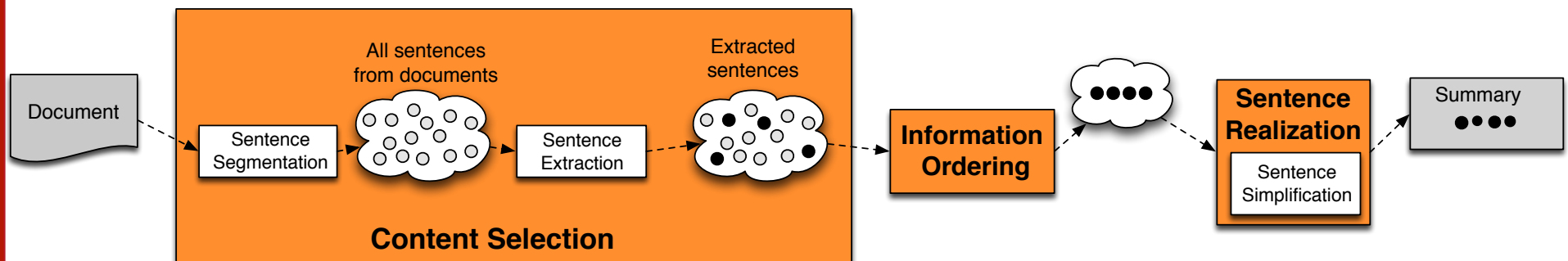
www.metmuseum.org/toah/ht/?period=07®ion=eak

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

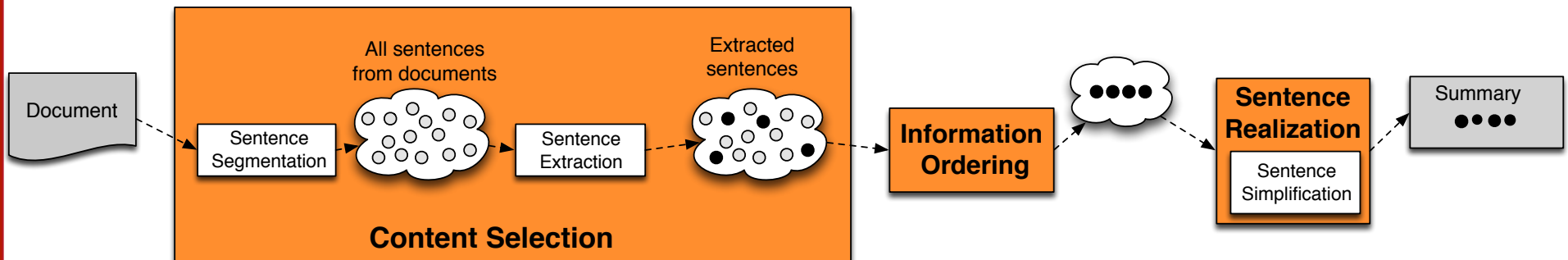
1. **content selection**: choose sentences to extract from the document
2. **information ordering**: choose an order to place them in the summary
3. **sentence realization**: clean up the sentences





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

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

- 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**

[illegible]

Generating Snippets and other Single- Document Answers