

Lecture 2.3

Text Summarization

Text Summarization

- **Goal:** produce a reduced version of a text that contains information that is important or relevant to understand the content.
- **Summarization Applications**
 - ▣ **outlines or abstracts** of any document, article, etc
 - ▣ **summaries** of email threads
 - ▣ **action items** from a meeting
 - ▣ **simplifying** text by compressing sentences
 - ▣ News **summarization**

'Types' of Summary?

- Indicative vs. informative
...used for quick categorization vs. content processing.
- Extract vs. abstract
...lists fragments of text vs. re-phrases content coherently.
- Generic vs. query-oriented
...provides author's view vs. reflects user's interest.
- Background vs. just-the-news
...assumes reader's prior knowledge is poor vs. up-to-date.
- Single-document vs. multi-document source
...based on one text document vs. fuses many text documents.

Aspects that Describe Summaries

□ Input

- *subject type*: domain
- *genre*: newspaper articles, editorials, letters, reports...
- *form*: regular text structure; free-form
- *source size*: single doc; multiple docs (few; many)

□ Purpose

- *situation*: embedded in larger system (MT, IR) or not?
- *audience*: focused or general
- *usage*: IR, sorting, skimming...

□ Output

- *completeness*: include all aspects, or focus on some?
- *format*: paragraph, table, etc.
- *style*: informative, indicative, aggregative, critical...

What to summarize?

□ **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:

- *Query based summarization tries to answer a question by summarizing a document that has the information to construct the answer*

Extractive summarization

□ Extractive summarization:

- ▮ create the summary from phrases or sentences in the source document(s)
- ▮ Selects some of the best describing sentences from the whole text and present to the user
- ▮ The number of sentences can be specified by number or proportion
- ▮ Eg:
 - take the first sentence
 - Take the first sentences of each paragraph
 - Take the sentence which contains high frequency words

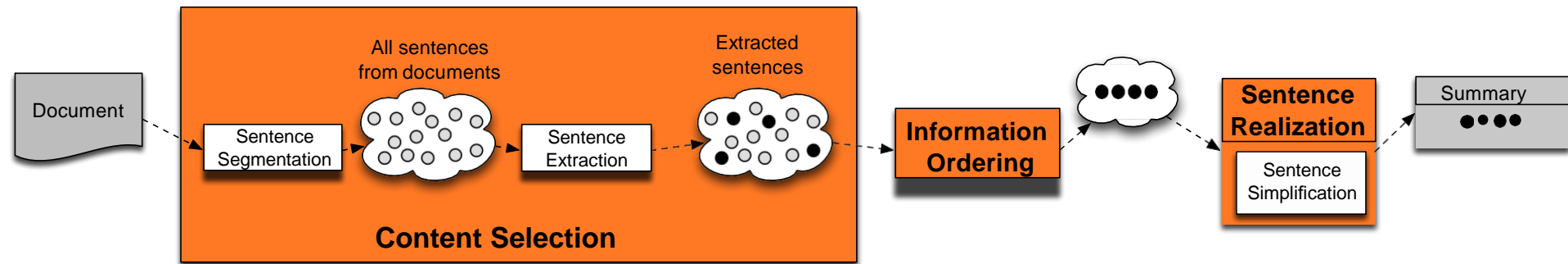
Abstractive summarization

□ Abstractive summarization:

- ❑ express the ideas in the source documents using (at least in part) different words
- ❑ More sophisticated summarization approach where the system is expected to compose its own summary of the whole text.

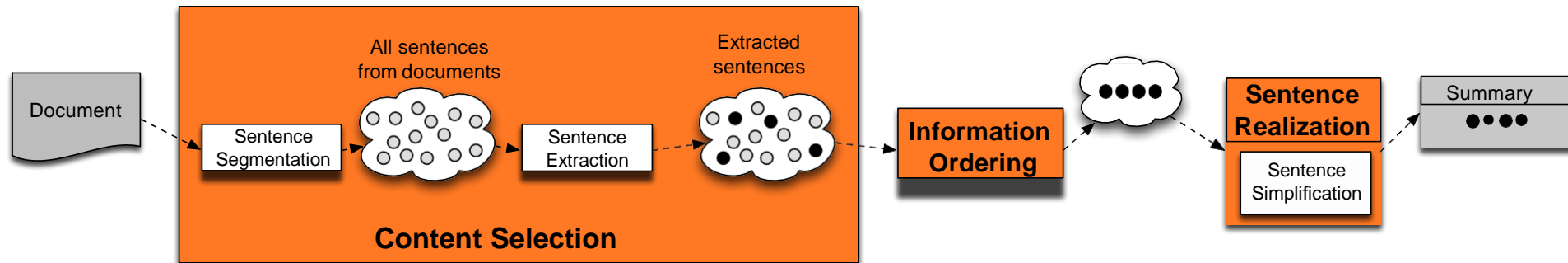
Summarization: Three Stages

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



The End