Text Summarisation

Types:

- single document vs. multi document
- text vs. speech
- monologue vs. dialogue
- generic vs. query-specific

Applications:
- summarising news articles on a single topic from different sources
- providing summaries of meetings (speech summarisation)
- automatic generation of abstracts for scientific papers
- shortening search engine output
- shortening text to fit an output device (generating subtitles, sentence compression)
Text Summarisation

Types:
- single document vs. multi document
- text vs. speech
- monologue vs. dialogue
- generic vs. query-specific

Applications:
- summarising news articles on a single topic from different sources
- providing summaries of meetings (speech summarisation)
- automatic generation of abstracts for scientific papers
- shortening search engine output
- shortening text to fit an output device (generating subtitles, sentence compression)
What makes a good summary?
What makes a good summary?

- original document is compressed/shortened
- summary is faithful to original
- summary contains most important information (according to user needs)
- summary is well-formed (coherence, grammatical sentences, anaphora etc.)
Extractive Summarisation:
Identify most important sentences/clauses in input document(s) and concatenate with minor corrections (anaphora resolution etc.). No generation step involved.

Abstractive Summarisation:
Identify most important information in input document(s) and paraphrase. Generation step required (paraphrasing).
Text Summarisation

Subtasks:

- content selection
- content organisation and linguistic realisation
Content selection based on:
Content selection based on:

- RST-info (select Nuclei only; Marcu 1997, 2000)
- lexical chains (to determine important sentences; Barzilay & Elhadad 1997)
- positional properties (first sentence of document/paragraph etc.)
- cue phrases (*to sum up*)
- tf-idf to find central concepts (and subsequental central sentences)
- degree of lexical connectiveness between passages
Sentence Ordering:

- same order as in input document
- possibly “smooth” output summary:
  - correct/normalise anaphora, temporal expressions etc.
  - aggregate (i.e., join two sentences)
  - add connectives (e.g. *and*)
Content Selection

- information extraction (template filling)

Linguistic Realisation

- text generation from template database
Document 1, BBC News

Barack Obama’s campaign to become the Democratic candidate for US president received a major lift with the coveted backing of Senator Edward Kennedy. The seal of approval from the patriarch of one of America’s . . .

Document 2, The Independent

Barack Obama took his re-energised campaign for the Democratic nomination across the Deep South yesterday after crushing Hillary Clinton in a racially polarised primary in South Carolina. Adding to the renewed momentum was an emotional endorsement from the daughter of John F Kennedy . . . That endorsement will be echoed by her uncle today as Senator Edward Kennedy announces his backing for the man who has emerged battle-hardened.

Document 3, The Times

Barack Obama yesterday paraded a trio of Kennedy clan endorsements, evidence that large sections of the Democratic establishment are now severing ties with the dominant force in the party for 16 years: Hillary and Bill Clinton. . . .
Document 1, BBC News
Barack Obama’s campaign to become the Democratic candidate for US president received a major lift with the coveted backing of Senator Edward Kennedy. The seal of approval from the patriarch of one of America’s . . .

Document 2, The Independent
Barack Obama took his re-energised campaign for the Democratic nomination across the Deep South yesterday after crushing Hillary Clinton in a racially polarised primary in South Carolina. Adding to the renewed momentum was an emotional endorsement from the daughter of John F Kennedy . . . That endorsement will be echoed by her uncle today as Senator Edward Kennedy announces his backing for the man who has emerged battle-hardened.

Document 3, The Times
Barack Obama yesterday paraded a trio of Kennedy clan endorsements, evidence that large sections of the Democratic establishment are now severing ties with the dominant force in the party for 16 years: Hillary and Bill Clinton. . . .
Challenges

- identify documents on the same topic (document clustering)
- recognising and coping with redundancy (sentence alignment and/or clustering)
- identifying important differences between documents (e.g. contradictory information, bias in source documents)
- temporal ordering more difficult (time may have elapsed between publication of two documents)
- temporal expressions have to be normalised (e.g. today)
- less easy to ensure coherence (source document may have different foci etc.)
Challenges

- identify documents on the same topic (document clustering)
- recognising and coping with redundancy (sentence alignment and/or clustering)
- identifying important differences between documents (e.g. contradictory information, bias in source documents)
- temporal ordering more difficult (time may have elapsed between publication of two documents)
- temporal expressions have to be normalised (e.g. today)
- less easy to ensure coherence (source document may have different foci etc.)
Method 1: Assessment by Human Judges

- compression rate ok?
- important information there?
- faithful to original?
- readable, coherent?
Method 2: Comparison with Gold Standard Human Generated Summaries

- Human comparison of the two summaries
- Precision, recall (for extractive, single document summaries)
- N-gram statistics, e.g., ROUGE (recall-oriented measure) (cf. BLEU for MT, precision-oriented)

Problem: no unique gold standard summary
Method 2: Comparison with Gold Standard Human Generated Summaries

- human comparison of the two summaries
- precision, recall (for extractive, single document summaries)
- n-gram statistics, e.g., ROUGE (recall-oriented measure) (cf. BLEU for MT, precision-oriented)

**ROUGE: Recall-Oriented Understudy for Gisting Evaluation**

\[
ROUGE_2 = \frac{\sum_{S \in \text{RefSummaries}} \sum_{\text{bigram} \in S} \text{Count}_{match}(\text{bigram})}{\sum_{S \in \text{RefSummaries}} \sum_{\text{bigram} \in S} \text{Count}(\text{bigram})}
\]
Method 3: Intrinsic/Application-based

- automatic evaluation of well-formedness/coherence (essay scoring methods)
- suitability for a task
- interpretability by a human