Text Summarization

Florina Piroi

Content

- Introduction
- Processing & baseline algorithms
- Evaluation

- Very active research area

Introduction

Motivation for Automatic Text Summarization

- Lots and lots of documents, mass of information, unstructured text
    - ➢ "too much information kills information"
- No standards in how to process them

Summaries written by:

- document author
- Professional summarizer
- Third party

Writing a summary is difficult – cognitive process!

select, comprehend (and reformulate), stay coherent

consistency (same person, different summaries at different times)

Challenges in automating it

Motivation for Automatic Text Summarization

- Lots and lots of documents, mass of information, unstructured text
    - ➢ “too much information kills information”
- No standards in how to process them

Automatic Text Summarization

- Condense the text
- Maintain relevant information
- Limited understanding of the text

Why (Automatically) Summarize?

- Reduce reading time
- Selection process is easier
- Effective indexing (by librarians)
- Automatic summarization is less biased
- Useful in Question-Answering systems
- Enable service providers to process more docs

What is a Text Summarization?

- 1979 (ANSI) : [An abstract] is an abbreviated, accurate representation of the contents of a document, preferably prepared by its author(s) for publication with it. Such abstracts are useful in access publications and machine-readable databases.
- 1980 (van Dijk): The primary function of abstracts is to indicate and predict the structure and content of the text.
- 1973 (Cleveland): An abstract summarizes the essential contents of a particular knowledge record, and it is a true surrogate of the document.

Produced by people!

What is a *Automatic* Text Summarization?

- OED : The creation of a shortened version of a text by a computer program. The product of this procedure still contains the most important points of the original text.

- 2001 (Spärk-Jones, Sakai): A summary is a reductive transformation of a source text into a summary text by extraction or generation.

-> compression, loss of information

- Which information to include in a summary?
  - Important & representative content (how do you measure that?)

What is a *Automatic* Text Summarization?

- OED : The creation of a shortened version of a text by a computer program. The product of this procedure still contains the most important points of the original text.

- 2005 (Hovy): An automatic summary is a text generated by a software, that is coherent and contains a significant amount of relevant information from the source text. Its compression rate τ is less than a third of the length of the original document.

$$\boxed{\tau = \frac{|Summary|}{|Source|}}$$

10%, 15% - 30%

What is a *Automatic* Text Summarization?

- OED : The creation of a shortened version of a text by a computer program. The product of this procedure still contains the most important points of the original text.

- 2005 (Hovy): An automatic summary is a text generated by a software, that is coherent and contains a significant amount of ***relevant information*** from the source text. Its compression rate τ is less than a third of the length of the original document.

$$\boxed{\tau = \frac{|Summary|}{|Source|}}$$

10%, 15% - 30%

Generating a Summary?

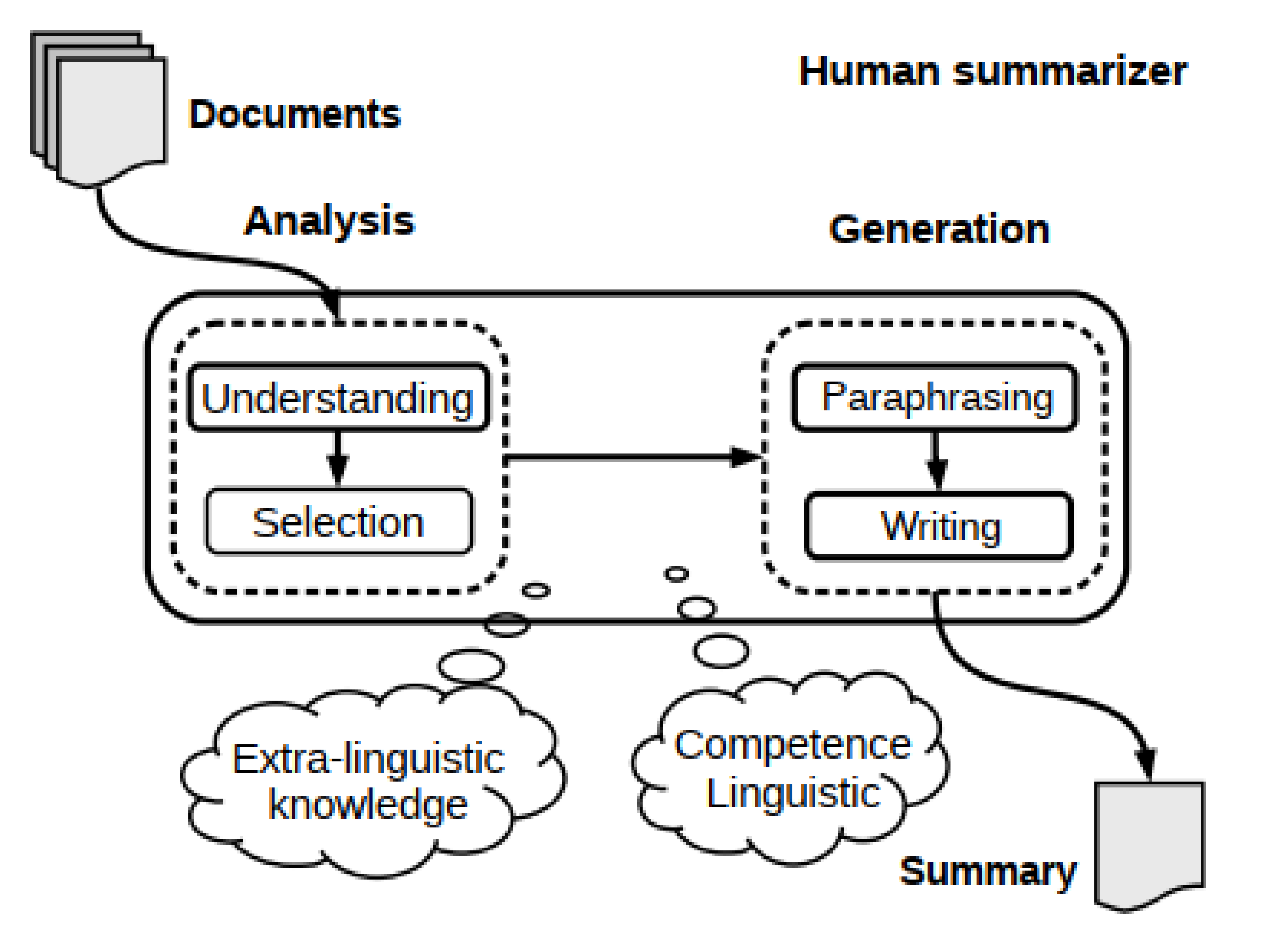
- “Art of Abstracting” – Edward Cremmins (1996)

- Highly structured and coherent

- Introduction (1-2 general sentences, 2-3 specific)
- Aims (1 sentence)
- Methods (1-2 sentences)
- Results (3-4 sentences)
- Conclusions (1-2 sentences)

(Summarizer need not be an expert, 8-12 minutes, no deep understanding)

- Should we replicate this?
- Probably not …
- … but exploitable properties!



Summaries: (High-level) Objectives

- Direct objective (why):
  - Give overview of the document
  - Update the user
  - Eliminate language barriers
  - Information Retrieval
- Indirect objective (what/how):
  - Classify, index
  - Extract keywords
  - …

Summaries … by their function

- Indicative summary
  - Give info about the topics discussed
  - Resembles a table of contents
- Informative summary
  - Reflect the content of the source text
  - Explaining the arguments
  - Shortened version of the document

Summaries … by number of documents

- Single-document
- Multi-document
  - Not always heterogeneous
  - Usually on the same topic

Summaries … by genre

- News summary (news articles)
- Specialised (on a specific domain – law, science, technology)
- Literary (narrative summaries, etc.)
- Encyclopaedic (summary of encyclopaedic docs)
- Social networks (of blogs, very short docs, etc.)

Summaries … by type

- Extract

  Fragments from the original document
- Abstract

  Reformulating/paraphrasing
- Sentence compression

  Same number of sentences as in the original doc, but shorter

Summaries … by type of summarizer

- Author summary
  Written by the author
- Expert summary
  Person other than the author, not an expert summarizer
- Professional summary
  Person does not specify in the field, but knows how to summarize

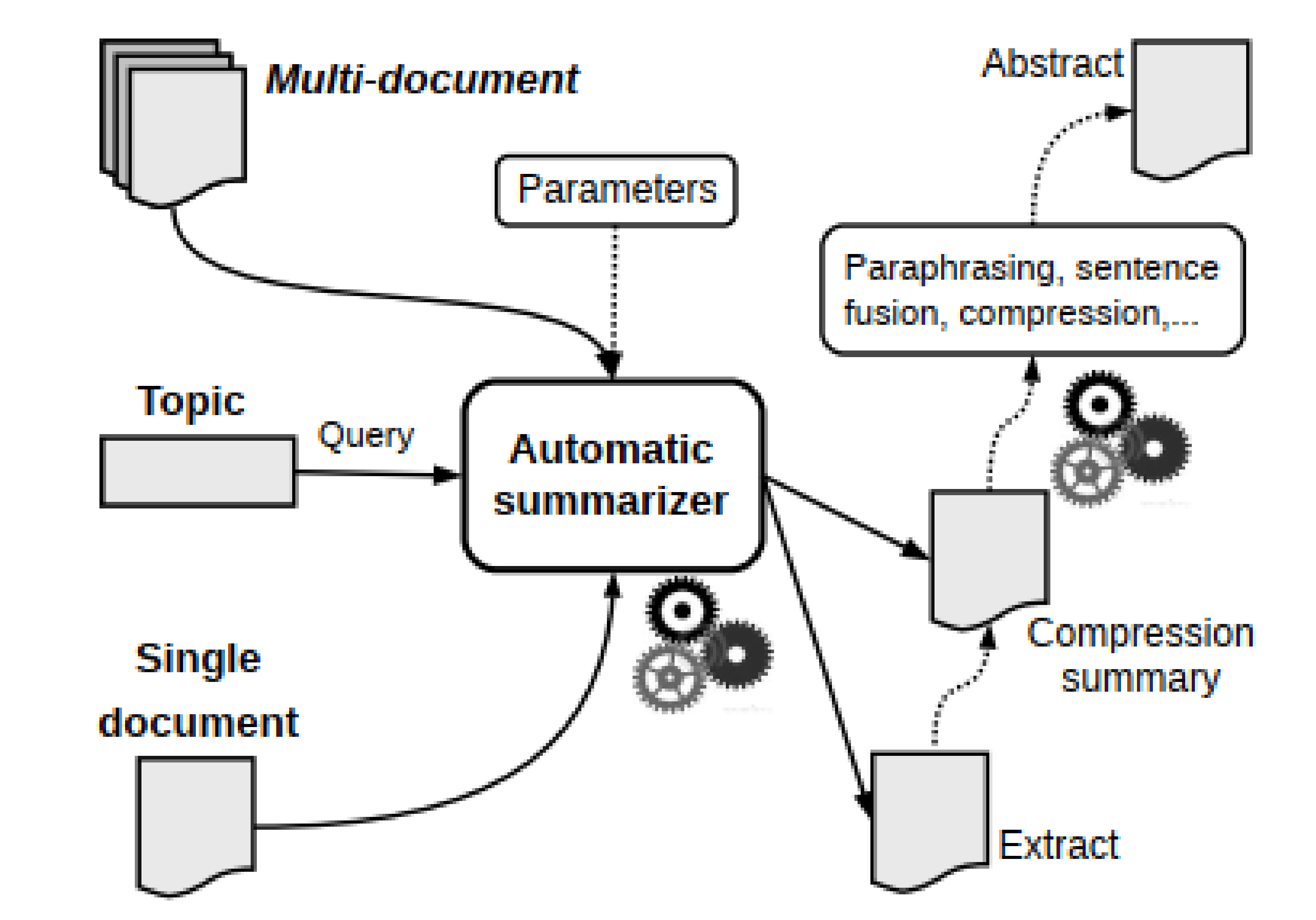
Summaries … by context

- Generic summary
  Ignores information need
- Query guided summary
  Guided by some user information need/query
- Update summary
  User familiar with topic, show only new, important information

Summaries … by target audience

- Without a profile
  - Independent of the user characteristics – only based on document
- User profile
  - Targeted at users interested in particular domains (phys, chem, etc)

Generating automatic text summaries



Applications

- Better IR and IE performance
- Q&A
- Newswire generation
- RSS feed summarization
- E-mail and e-mail thread summarization
- Meeting summaries
- Automatic generation and extraction of titles
- Opinion summarization
- Domain-specific summarization (medicine, law, etc).
- …

Processing & Baseline Summarization Algorithms

Extraction vs. Abstraction vs. Compression

- Summarization by
  - Extraction
  - Abstraction
  - Sentence compression

Pre-processing: Before the Summary

- Objectives
    - Word normalization
    - Lexicon reduction

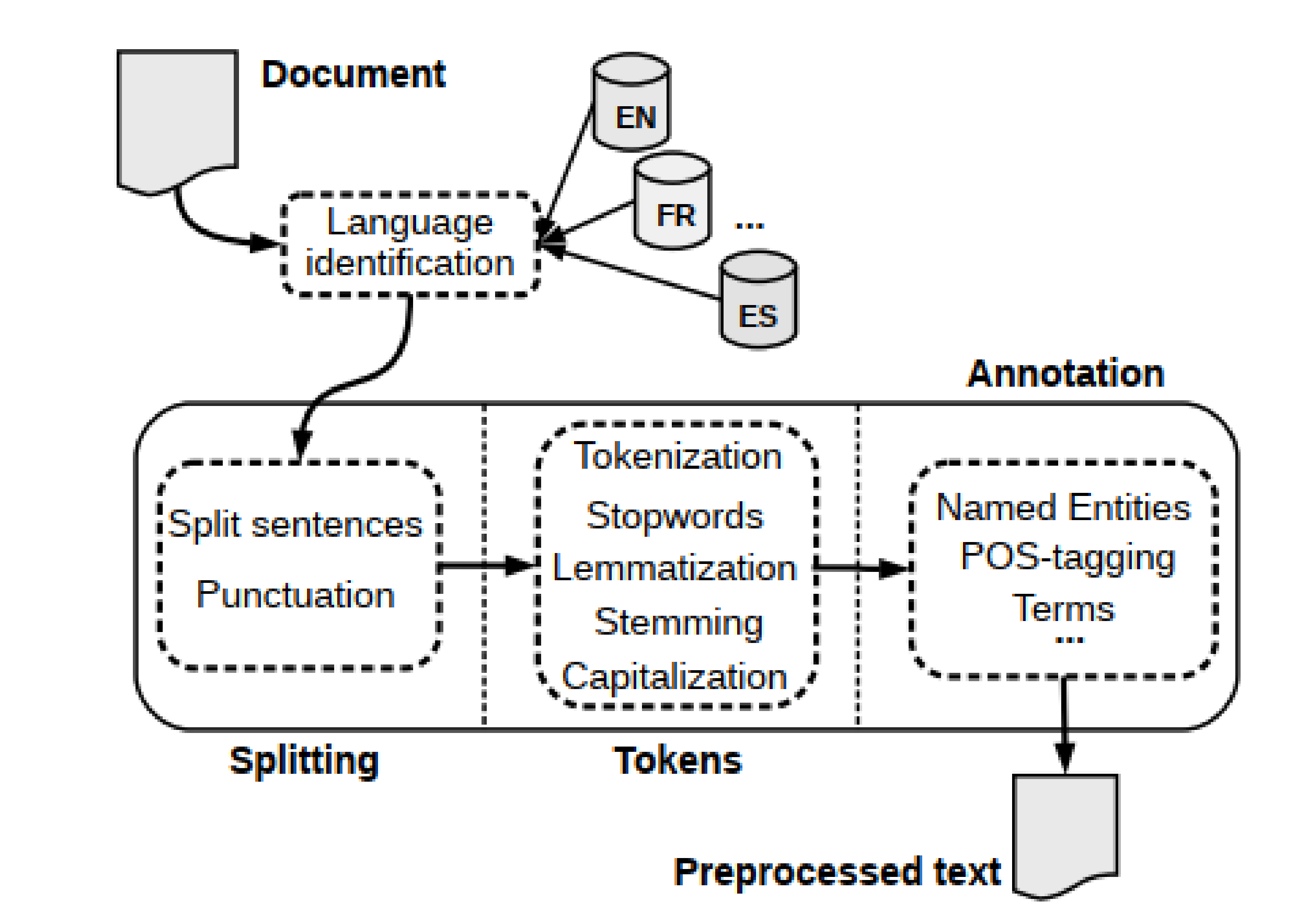
- Essential to providing clean, adequate representations of the source documents.

Pre-processing: Before the Summary

- (Standard) stages:
  - Split into segments / paragraphs / sentences / etc.
  - Split segments into words / tokens
  - Normalization: lemmatization, stemming, etc.
  - Filtering (stop word removal)
- Optional:
  - Annotations (e.g. POS)
  - Named Entity recognition (NE)
  - Term extraction (keywords)
  - Weighting (tf-idf, embedding vectors, etc.)

Pre-processing: Before the Summary

- (Standard) stages:
  - Split into segments / pa
  - Split segments into wor
  - Normalization: lemmati
  - Filtering (stop word rem
- Optional:
  - Annotations (e.g. POS)
  - Named Entity recognitic
  - Term extraction (keywo
  - Weighting (tf-idf, embed



Extraction vs. Abstraction vs. Compression

- Summarization by
  - Extraction
  - Abstraction
  - Sentence compression

- Summary: reduced representation, keeps essential content of a text
- Abstract: reformulating sentences (“good student”)
- Extract: picking out (extracting) pieces of the original text (“bad student”).

- Misuses leading to confusion & ambiguity: Zusammenfassung, Résumé

Extraction vs. Abstraction vs. Compression

- Select units of texts based on informativity
- Assemble them in a useful way.
- Aim: give an overview of the original text’s content.
- Length: determined by compression rate.

    - (2000) 81.5% sentences borrowed verbatim (manually generated summaries)
    - (2004) 70% sentences in summaries appear in the original text

Extraction vs. Abstraction vs. Compression

- (Radev et al 2002) Algorithms are:
  - Surface-level
    - No linguistic processing
    - Word weighting
    - Used for docs with a fixed structure
    - Segment position
    - Words / Key sequences
  - Intermediate-level
  - Deep parsing techniques

Extraction vs. Abstraction vs. Compression

- (Radev et al 2002) Algorithms are:
  - Surface-level
  - Intermediate-level
    - Use linguistic information
    - Not as complex as deep parsing algorithms
    - Use lexical chains – the strong chains go into the summary.
    - Anaphoric references
  - Deep parsing techniques

Extraction vs. Abstraction vs. Compression

- (Radev et al 2002) Algorithms are:
  - Surface-level
  - Intermediate-level
  - Deep parsing techniques
    - Use in-depth linguistic techniques
    - Exploit discursive structure theory (rhetorical structure theory)
    - Divide text into discursive (rhetorical) units
    - Apply weights to these units
    - Highest weighed units are selected for the summary

Extraction vs. Abstraction vs. Compression

- Final objective in Summarization
- Text understanding
- Generate correct & coherent text
- Can combine IE, IR and Natural Language Generation techniques

Extraction vs. Abstraction vs. Compression

- Sentence compression idea:
  - Eliminate non-essential information in a sentence
  - Preserve correct grammar
  - All sentences kept
  - No solution to it, yet.

- Applications: generate titles, subtitles for media, display info on mobile devices

- Compression rate up to 33% according to some authors.

Extraction vs. Abstraction vs. Compression

- Sentence compression idea:
  - Eliminate non-essential information in a sentence
  - Preserve correct grammar
  - All sentences kept
  - No solution to it, yet.

- Multi-sentence fusion (multi-sentence compression):
  - more documents
  - Redundant sentences indicator for importance
  - Sentences from different docs fused and compressed
  - Form of paraphrasing

Limits of Extraction

- Cohesion
- Coherence

(hint on how to evaluate the quality of a summary)

Limits of Extraction

- Cohesion

  Cohesion between the sentences of a text is related to the absence of anaphora and unresolved temporal references in the generated summary.

  1) I like cherries
  2) I am allergic to them.

- Resolving anaphora in text is difficult

Limits of Extraction

- Coherence

  The coherence of the summary is the absence of contradictions and/or redundancy in the sequence of sentences in a document.

Source: The aircraft manufacturing industry is still in crisis in 2011. In 2010 Airbus faced many problems due to financial difficulties and was forced to halt production of its new long-haul carrier A380. Transatlantic, its Boeing rival, has finally started production of its B770. The new B770 has received a large number of pre-orders from Asian airline companies. This is good news for the manufacturer who may end up owning a large share of the aircraft industry.

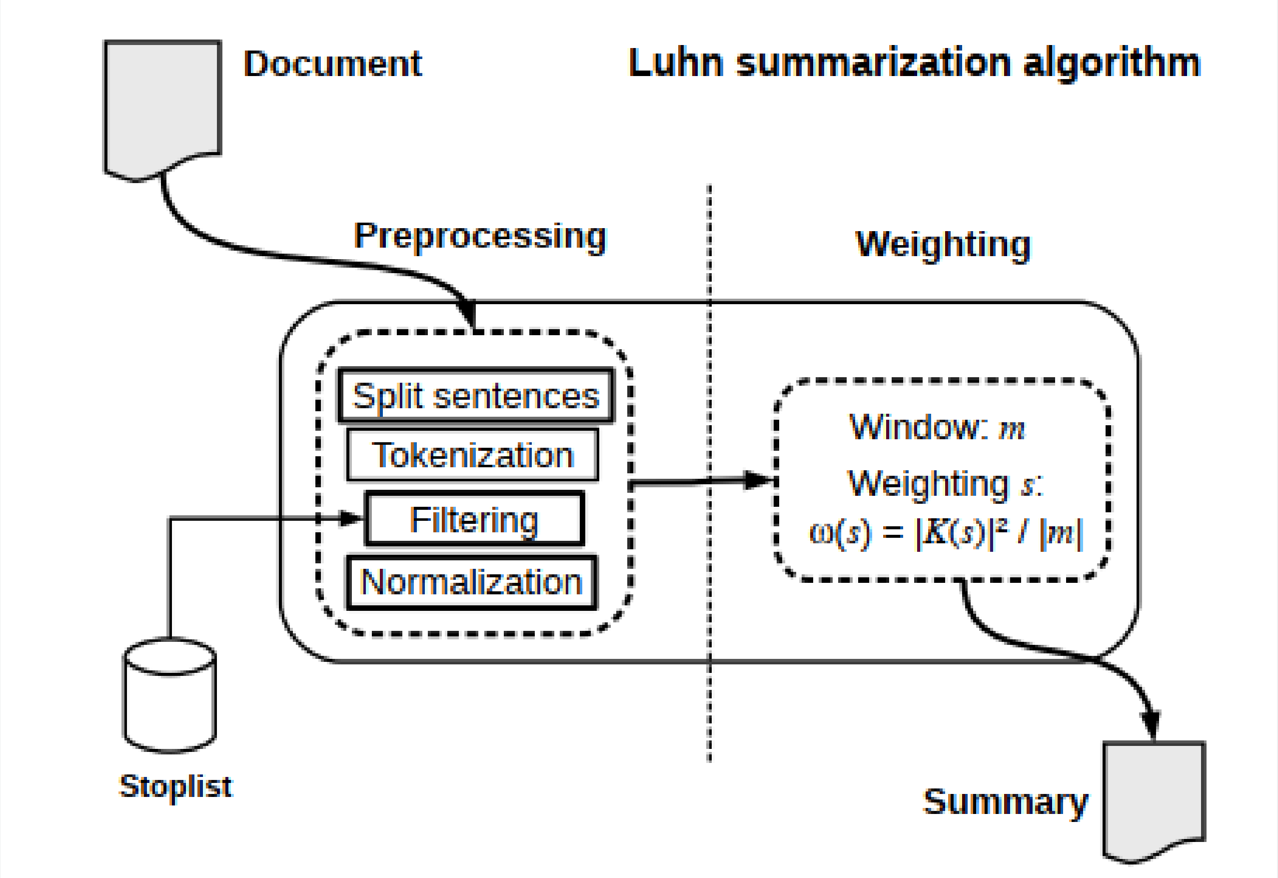
Summary: In 2010 Airbus faced many problems due to financial difficulties and was forced to halt production of its new long-haul carrier A380. This is good news for the manufacturer who may end up owning a large share of the aircraft industry.

Algorithms: Luhn's Automatic Creation of Literature Abstracts

- Hans Peter Luhn (1896 – 1964)
  - Pioneer in automatic text summarization
  - 1958 (while at IBM) developed a simple algorithm to extract abstracts from scientific articles
  - Was worried about the information overload.
  - Recognized his work was "miles away" from text understanding & linguistics

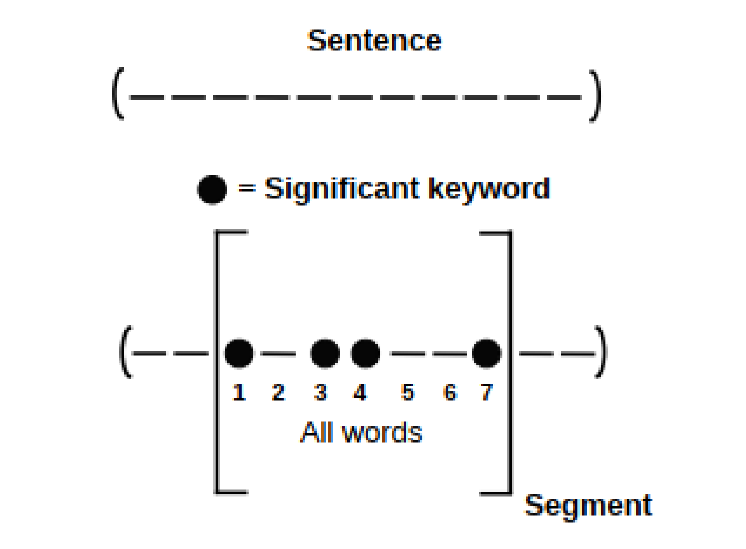
Algorithms: Luhn's Automatic Creation of Literature Abstracts

- Two phases:
  - Preprocess source document
  - Sentence weighing
  - (Fortran & Lisp)

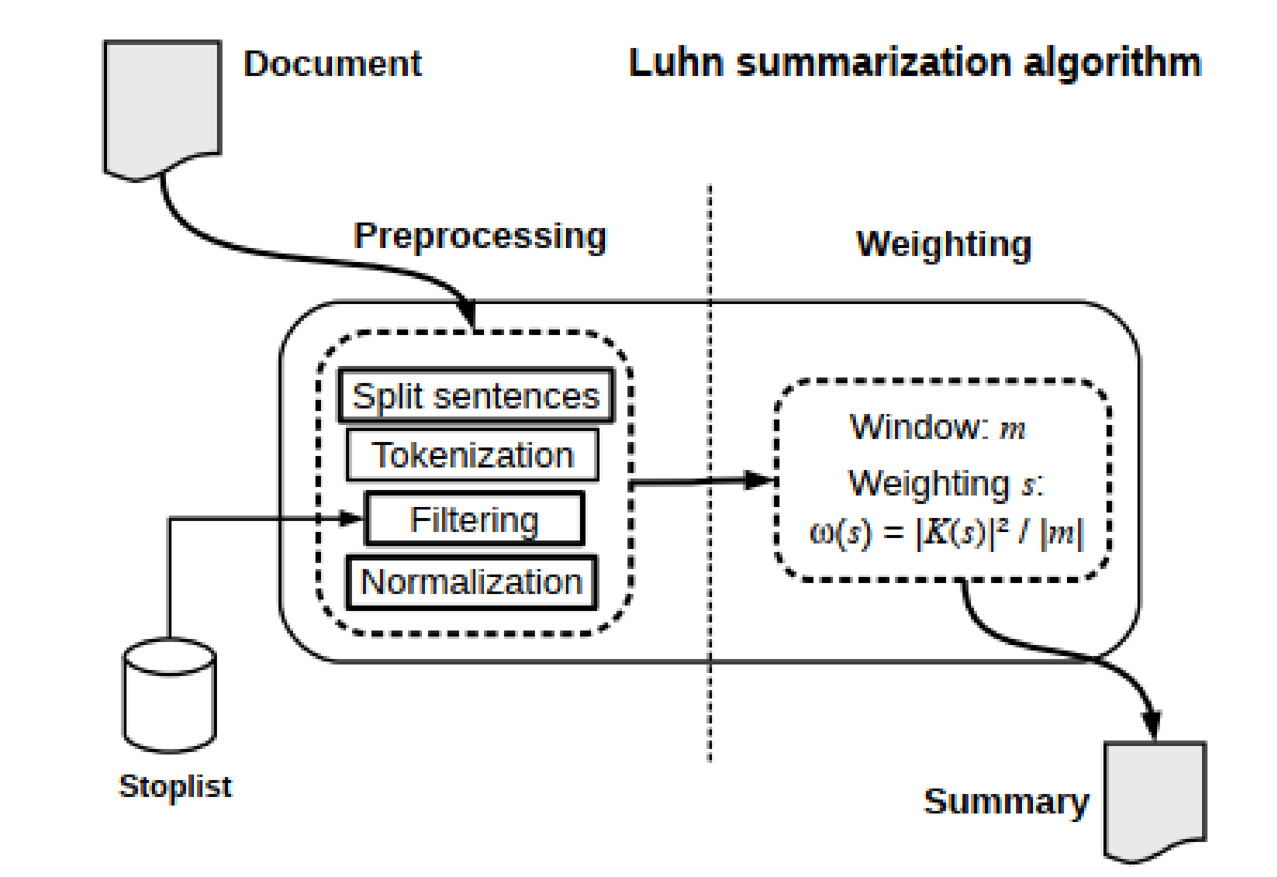


Algorithms: Luhn's Automatic Creation of Literature Abstracts

- Two phases:
  - Preprocess source document
  - Sentence weighing
  - (Fortran & Lisp)



  - Sentence weight: 16 / 7



Informatics

Algorithms: Graph-based Approaches

- Graph models to represent textual information:
  - Semantic disambiguation, POS tagging, IE, opinion analysis, etc.
  - Vertices (nodes) are semantic/cognitive units (words)
  - Edges are relations between units.

- Graphs -> *very sparse* binary matrix representation
- Effectively used in e-mail thread summarization, multisentence compression, tweet summarization, keyword detection, etc.

Algorithms: Graph-based Approaches

- PageRank
  - Computer representation for sparse matrices
  - Vertex centrality (from SNA)

- (central to IR – webpage ranking)
- Popularity of page *i* computed wrt. links starting/ending from/on *i*

$$p_R(i) = \frac{1-d}{N} + d\sum_{j=1}^{N} \frac{p_R(j)}{C(j)}$$

Algorithms: Graph-based Approaches

- LexRank & TextRank
  - Widely used
  - Based on the PageRank algorithm
  - Apply the graph to the concept of sentence weighting
  - Identify the most “prestigious” sentence in the graph!

- Document as graph
- Nodes are text units (sentences)
- Edges are similarity values between text units.
- “Central” sentences make it to the summary

Algorithms: Graph-based Approaches

- LexRank & TextRank
  - Document as graph
  - Nodes are text units (sentences)
  - Edges are similarity values between text units.
  - “Central” sentences make it to the summary
- Two algorithm stages:
  - Graph construction (sentences + similarity scores)
  - Link analysis (sentence “recommendation”)

Algorithms: Graph-based Approaches

- Graph construction:
  - Nodes: take whole sentences (simplest)
  - Edges: morphological similarity (~recommendation)
    - Cosine, word overlap, etc.
- $G = (V, E), E \subseteq V \times V$
- V_i, $s_j \in In(s_i)$
- $\sum_{s_k \in Out(s_j)}$

Algorithms: Graph-based Approaches

- Sentence weighting
  - $G = (V, E), E \subseteq V \times V$
  - $V_i,\ s_j \in In(s_i)$
  - $\sum_{s_k \in Out(s_j)}$

- LexRank:
  - (multidoc)

$$\omega(s_i) = \frac{d}{N} + (1-d) \sum_{s_j \in In(s_i)} \frac{\mathrm{sim}(s_i, s_j)}{\sum_{s_k \in \mathrm{Out}(s_j)} \mathrm{sim}(s_k, s_j)} \omega(s_j)$$

http://lexrank.herokuapp.com/index.html

Algorithms: Graph-based Approaches

- Sentence weighting
  - $G = (V, E), E \subseteq V \times V$
  - $V_i,\ s_j \in In(s_i)$
  - $\sum_{s_k \in Out(s_j)}$

- TextRank:
  - (single doc)

$$\omega(s_i) = (1-d) + d \sum_{s_j \in In(s_i)} \frac{w_{j,i}}{\sum_{s_k \in \mathrm{Out}(s_j)} w_{jk}} \omega(s_j)$$

$$w_{j,i} = \frac{\sum_{w \in i,j} C_w^i + C_w^j}{\log|s_i| + \log|s_j|}$$

TU WIEN Informatics

http://lexrank.herokuapp.com/index.html

Algorithms: Machine Learning Approaches

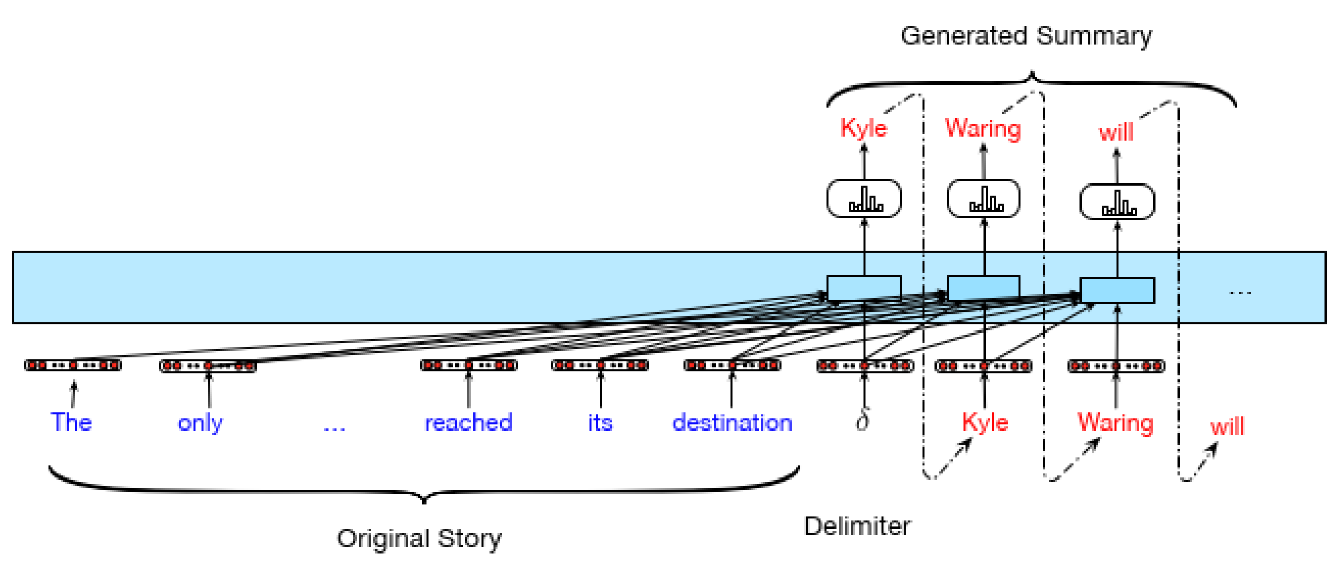
- Position of keywords -> weight of sentences -> selection into summary
- Use other parameters (features), too
- How to decide their contribution to the selection of a sentence?

- ML to estimate parameter values from their occurrences on a previously *tagged* corpus.
  - {article, summary}
  - Train a Bayes classifier to estimate sentence probability

  - Features: sentence length cut-off, fixed-phrases, position of sentence in paragraph, lexicon feature, upper case words,
  - (Kupiec et. al. 95 – small data set, 188 pairs)

Algorithms: Machine Learning Approaches

- Deep Learning methods
- Transformer-based autoregressive models
- Training data: {article, separator, summary}
  - (a_1, …, a_m, SEP, s_1, …, s_n)
  - Handled as long sequences



Algorithms: Machine Learning Approaches

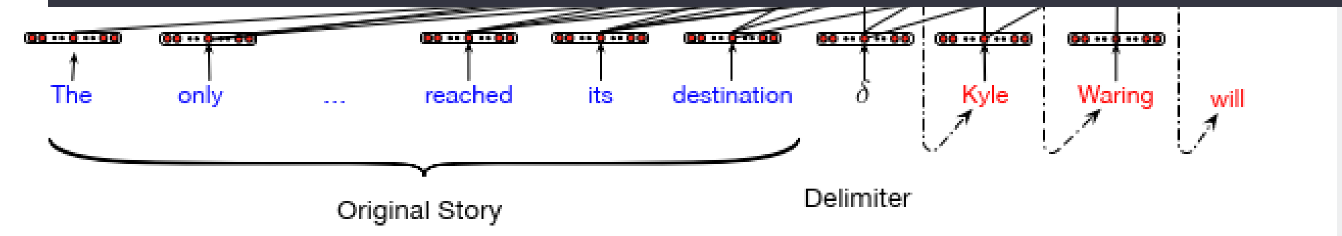
- Deep Learning methods
- Transformer-based autoregr
- Training data: {article, separa
  - (a_1, ..., a_m, SEP, s_1, ..., s_
  - Handled as long sequences

You

summarize this: The aircraft manufacturing industry is still in crisis in 2011. In 2010 Airbus faced many problems due to financial difficulties and was forced to halt production of its new long-haul carrier A380. Transatlantic, its Boeing rival, has finally started production of its B770. The new B770 has received a large number of pre-orders from Asian airline companies. This is good news for the manufacturer who may end up owning a large share of the aircraft industry.

ChatGPT

In 2011, Airbus faced financial issues, pausing A380 production, while Boeing's B770, receiving significant pre-orders from Asian airlines, is poised to dominate the aircraft industry.



Evaluation

Evaluation

- Input (text corpus) → Text Summarizer → Output (text summary)

- The little prince visited a planet inhabited by a drunkard. He talked to the drunkard and asked him why he drank. The response he received left him perplexed.

- The little prince visited a planet inhabited by a drunkard. After several bizarre answers from the drunkard, the little prince understood that the drunkard drank to forget that he was ashamed of drinking. The visit was very short and the little prince went away perplexed.

No perfect summary

Evaluation

- Intrinsic
  - measures the system's performance on its own
  - Against human reference or gold standard reference
  - Baseline summaries:
    - Randomly extract *n* sentences
    - Use lead sentences, first *n* sentences
    - Use lead-end sentences, first *n* and the last *m*
  - Task specific baselines are also used.
- Extrinsic
  - how summaries are good enough to accomplish the purpose of some other specific task (IR, report generation, etc)

Semi-automatic Summary Evaluation

- Levels of granularity
  - Sentence
  - Word

Semi-automatic Summary Evaluation

- Sentence level evaluation:
  - Have a set of "perfect" summaries (*reference* summaries)
  - Precision & Recall for content overlap

- Precision: how many sentences match the ones in the reference summary?
- Recall: how many sentences were "forgotten" by a system
- F: harmonically combines P and R

$$P = \frac{|\mathrm{Sum}_{\mathrm{ref}} \cap \mathrm{Sum}_{\mathrm{can}}|}{|\mathrm{Sum}_{\mathrm{can}}|} \quad R = \frac{|\mathrm{Sum}_{\mathrm{ref}} \cap \mathrm{Sum}_{\mathrm{can}}|}{|\mathrm{Sum}_{\mathrm{ref}}|} \quad F = 2 \cdot \frac{P \cdot R}{P + R}$$

Semi-automatic Summary Evaluation

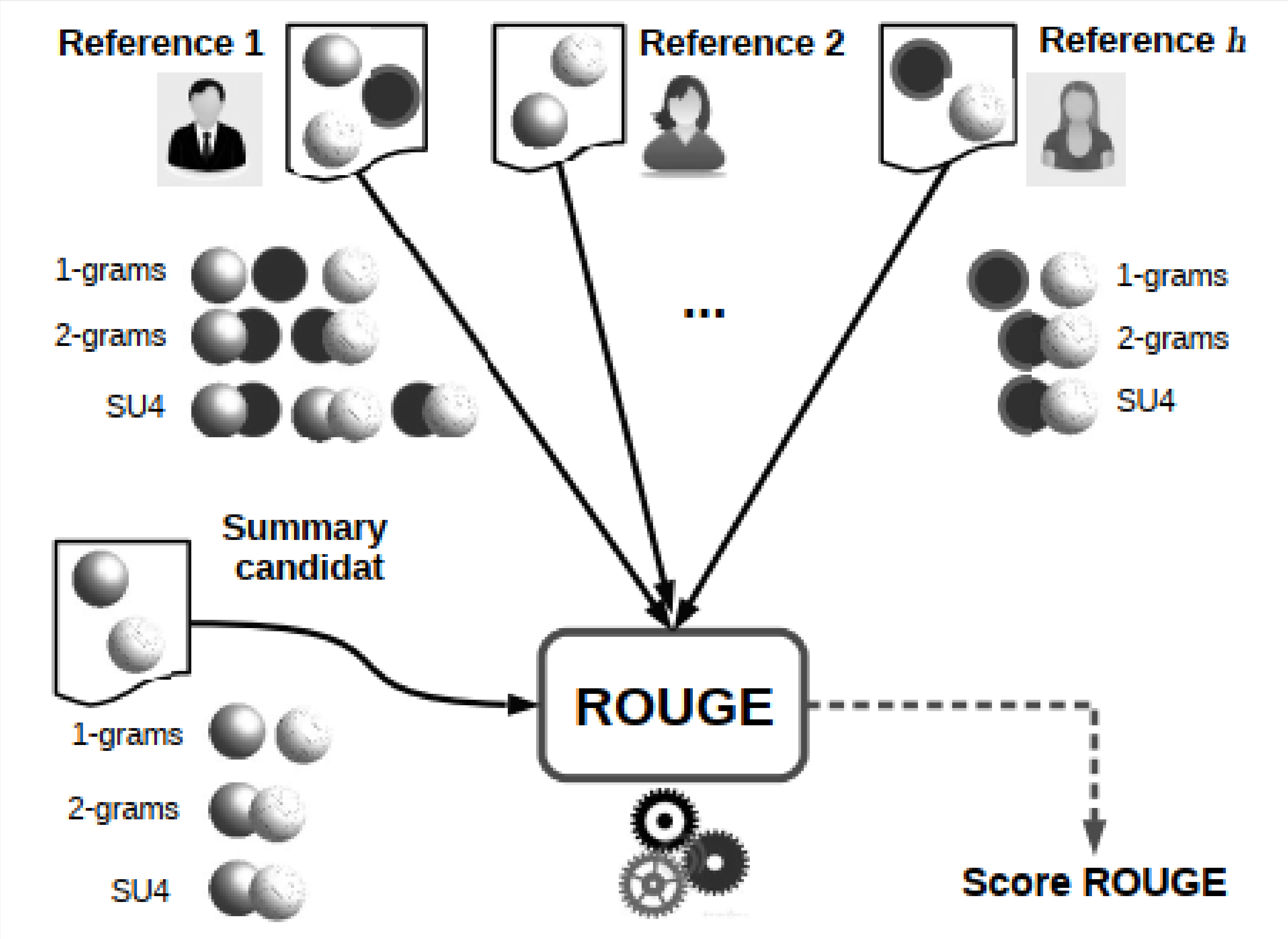
- Word level evaluation:
  - ROUGE (Recall-Oriented Understudy for Gisting Evaluation)
  - PYRAMID
  - BASIC ELEMENTS (BE)

- Measure similarity to a reference summary.

- ROUGE (2004): look at difference in word distribution between the candidate and the reference summary.
- N-grams!

ROUGE

- Inspired by BLEU



ROUGE

- Number of n-grams to be used in the Rouge-n metric:

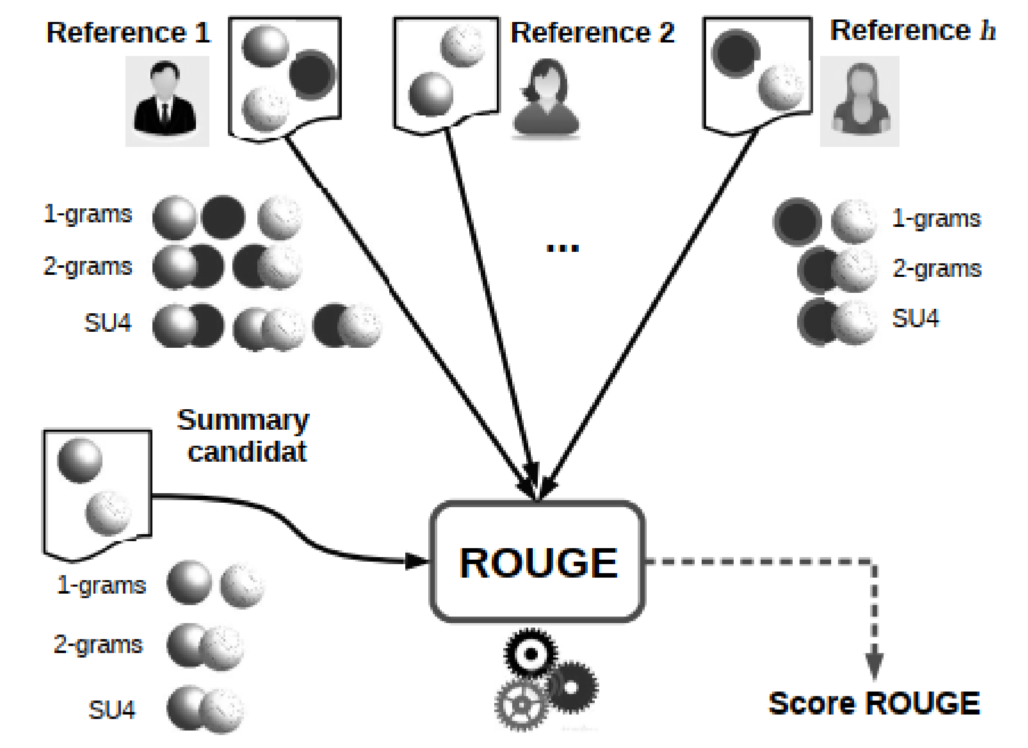
Sentence	*Intellectuals solve problems; geniuses prevent them.*	Count
ROUGE-1	Intellectuals ▷ solve ▷ problems ▷ geniuses ▷ prevent ▷ them	6
ROUGE-2	Intellectuals solve ▷ solve problems ▷ problems geniuses ▷ geniuses prevent ▷ prevent them ▷	5
ROUGE-SU4	Intellectuals solve ▷ Intellectuals problems ▷ Intellectuals geniuses ▷ solve problems ▷ solve geniuses ▷ solve prevent ▷ problems geniuses ▷ problems prevent ▷ problems them ▷ geniuses prevent ▷ geniuses them ▷ prevent them	12

ROUGE

- Rouge-2, Rouge-SU4 mostly used

- Rouge-SUγ ~ ROUGE-2 using skip units (SU) of size ≤ γ
- Rouge-SU4 – max window length is 4.

$$\text{ROUGE-}n = \frac{\sum_{n\text{-grams}} \in \{\mathrm{Sum}_{\mathrm{can}} \cap \mathrm{Sum}_{\mathrm{ref}}\}}{\sum_{n\text{-grams}} \in \mathrm{Sum}_{\mathrm{ref}}}$$



ROUGE-SUγ

- γ = 4 (mostly used)
- γ = 4 == ROUGE-2
- Bigram counts in a window of size *n*:

$$\mathrm{Count}(k,n) = C\binom{n}{k} - \sum_{0}^{k-\gamma}(k-\gamma)\ ;\ \gamma \geq 1$$

ROUGE-n

- Exhaustive character
- Limited by content representation

- “blue convertible car” vs. “blue automobile”
- “The airplane A380” vs. “The Airbus 380 aeroplane”

- Possible to cheat for high ROUGE scores.
  - Sjöbergh 2007 study
  - Markov automata, bigram frequencies only.

ROUGE-n

- Exhaustive ch
- Limited by co

- “blue convert
- “The airplane

- Possible to ch
  - Sjöbergh 2
  - Markov au

of the hurricane andrew had been injured and the storm caused by the gulf of mexico and louisiana and at least 150 000 people died on the us insurers expect to the florida and there are likely to be concentrated among other insurers have been badly damaged a result of damage caused in the state and to dollars 20bn of new orleans then to pay out on monday and new iberia associated tornadoes devastated inhabitants of miami

⋮

industry is the costliest disaster in florida as a quarter of hurricane hugo which hit the industry analysts cautioned that the bahamas on tuesday night or shut down

Live Summarization Systems

https://www.tools4noobs.com/summarize/

http://lexrank.herokuapp.com/index.html

http://proceedings.mlr.press/v119/zhang20ae.html

Evaluation:

S. Shahbaz, D. Schwabe, and M. Potthast. Summary Workbench: Unifying Application and Evaluation of Text Summarization Models, (EMNLP 2022, arXiv)

Content

- Introduction
- Processing & baseline algorithms
- Evaluation

Not covered: Abstracting text, Domain-specific summarization, query-guided summarization…

Luhn, H.P.: The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development. 2, 159–165 (1958). https://doi.org/10.1147/rd.22.0159.

Kupiec, J., Pedersen, J., Chen, F.: A Trainable Document Summarizer. Presented at the (1995).

Torres-Moreno, "Automatic Text Summarization", 2014, Wiley

Radev, D.R., Hovy, E., McKeown, K.: Introduction to the Special Issue on Summarization. Computational Linguistics. 28, 399–408 (2002).

Radev, Dragomir R., Timothy Allison, Sasha Blair-Goldensohn, John Blitzer, Arda Celebi, Stanko Dimitrov, Elliott Drabek, et al. "MEAD - A Platform for Multidocument Multilingual Text Summarization," 2004. https://doi.org/10.7916/D8MG7XZT.

Lloret, E.: TEXT SUMMARIZATION : AN OVERVIEW.