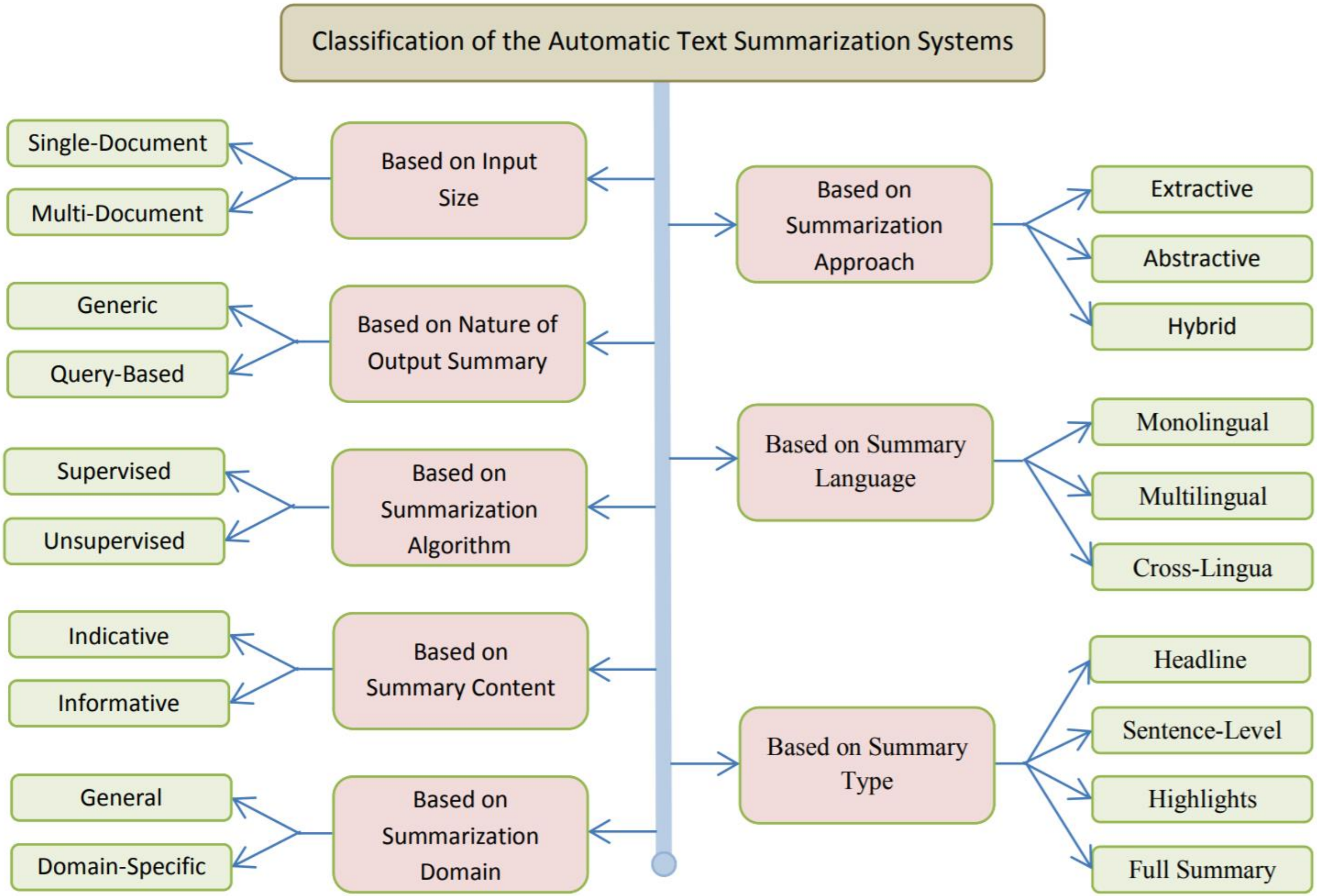
Text summarization methods

Prof. Luca Cagliero
Dipartimento di Automatica e Informatica
Politecnico di Torino

Lecture goal

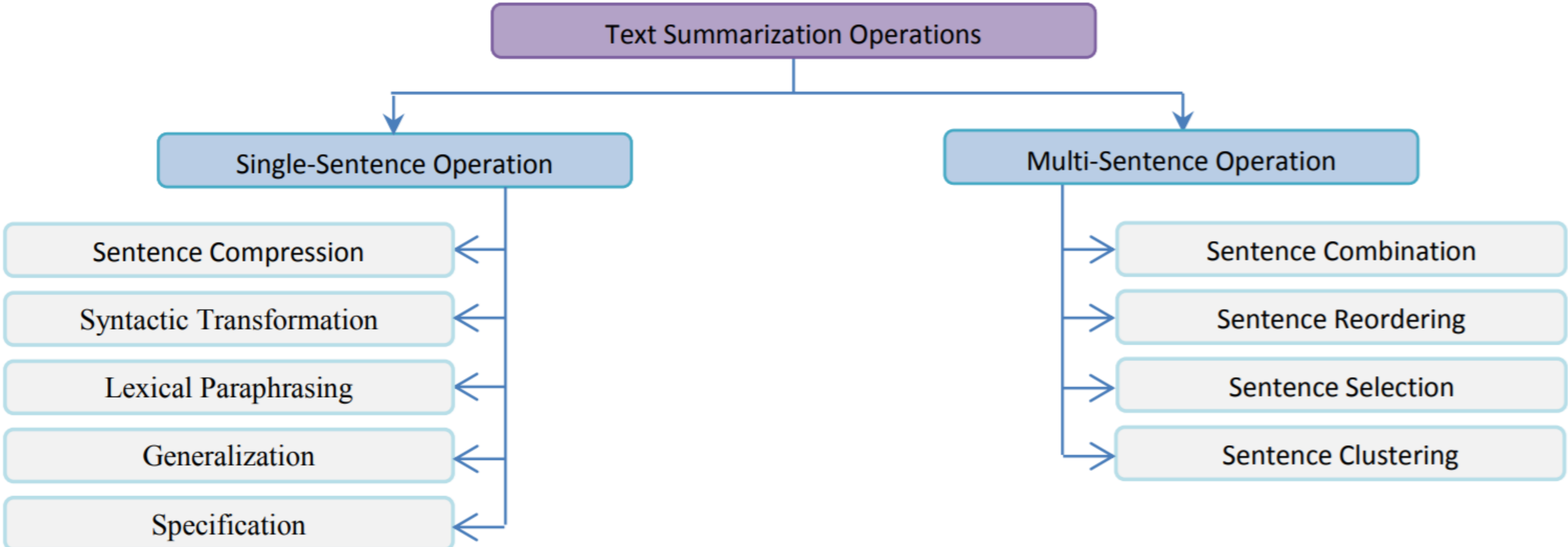
- Classifications of text summarization methods
- Overview of the main strategies
  - Clustering-based
  - Graph-based
  - Itemset-based
  - Optimization-based

Classification of text summarization methods



Automatic text summarization: A comprehensive survey. Wafaa S.El-Kassasa, Cherif R.Salamaab, Ahmed A.Rafeab, Hoda K.Mohameda. ESWA 2021

Key operations



Automatic text summarization: A comprehensive survey. Wafaa S.El-Kassasa, Cherif R.Salamaab, Ahmed A.Rafeab, Hoda K.Mohameda. ESWA 2021

Key operations

- Sentence Compression/Reduction
  - removal of unimportant parts (e.g. phrases) to shorten the original sentence.
- Syntactic Transformation
  - transformation of a sentence by changing its syntactic structure (e.g. the position of the subject in a sentence may be moved from the end to the front). This operation may be used in both sentence compression and sentence combination operations.
- Lexical Paraphrasing
  - replacement of phrases by their paraphrases.
- Generalization
  - replacement of phrases or clauses by more general descriptions.
- Specification
  - replacement of phrases or clauses by more specific descriptions.

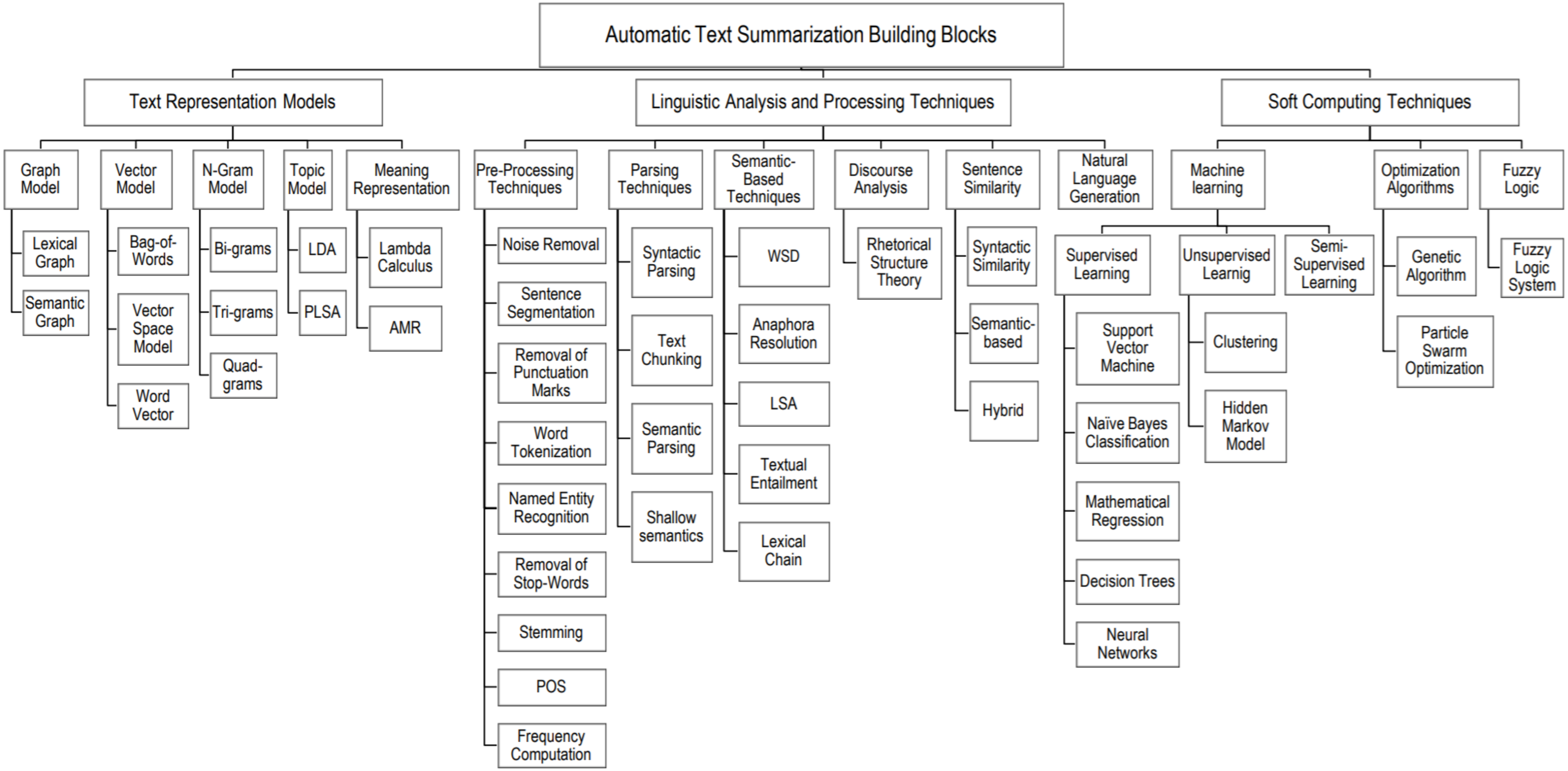
Automatic text summarization: A comprehensive survey. Wafaa S.El-Kassasa, Cherif R.Salamaab, Ahmed A.Rafeab, Hoda K.Mohameda. ESWA 2021

Key operations

- Sentence Combination/Fusion
  - merge of two or more original sentences into a single summary sentence
- Sentence Reordering
  - change of the order of summary sentences
    - E.g., an ending sentence in an original text is placed at the beginning of the summary
- Sentence Selection
  - selection of one sentence from two or more similar sentences
- Sentence Clustering
  - grouping of the sentences into different clusters
  - This operation is very useful in multi-document summarization
    - e.g., identify the subject and cluster the sentences by subject

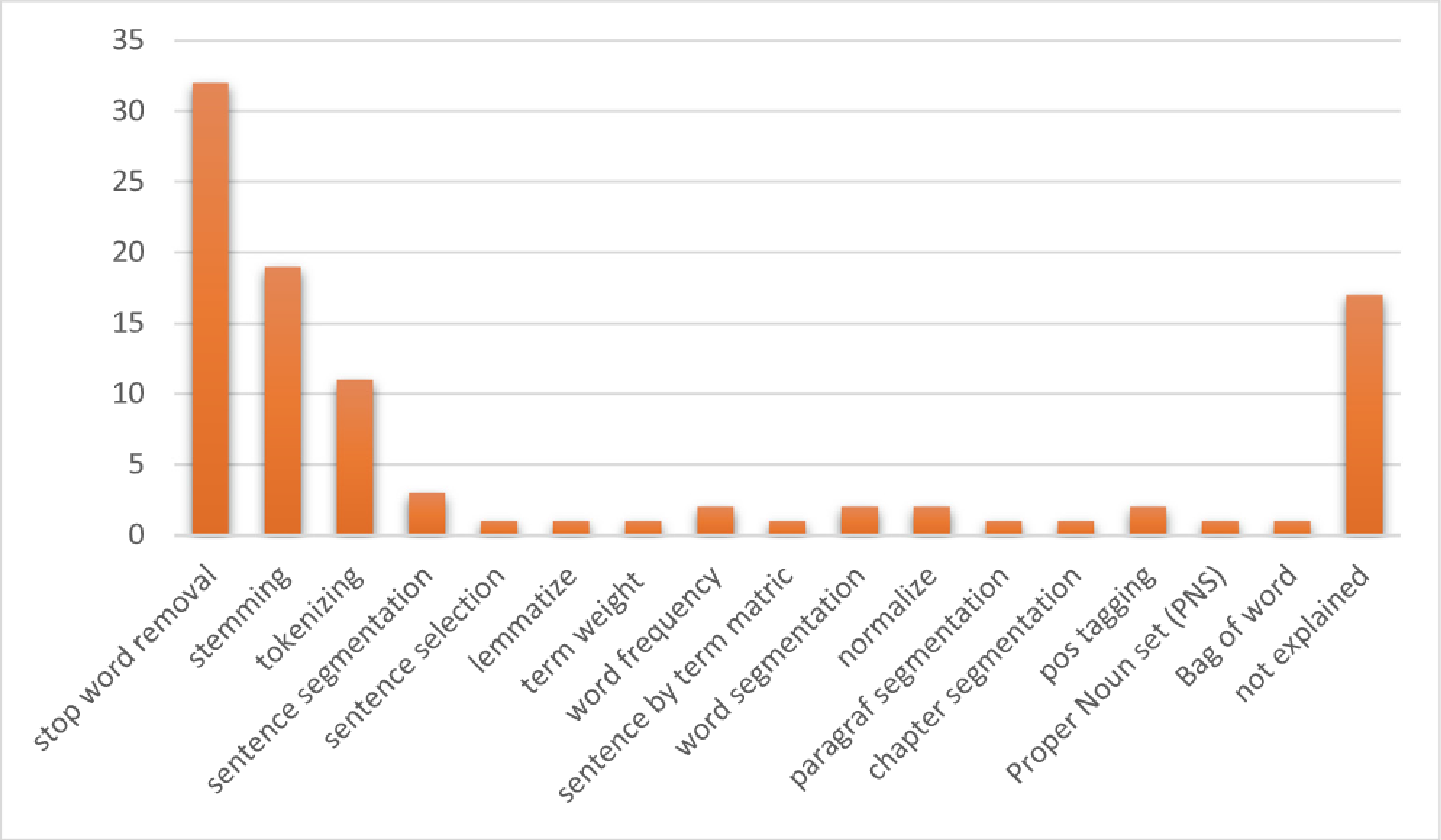
Automatic text summarization: A comprehensive survey. Wafaa S.El-Kassasa, Cherif R.Salamaab, Ahmed A.Rafeab, Hoda K.Mohameda. ESWA 2021

Building blocks



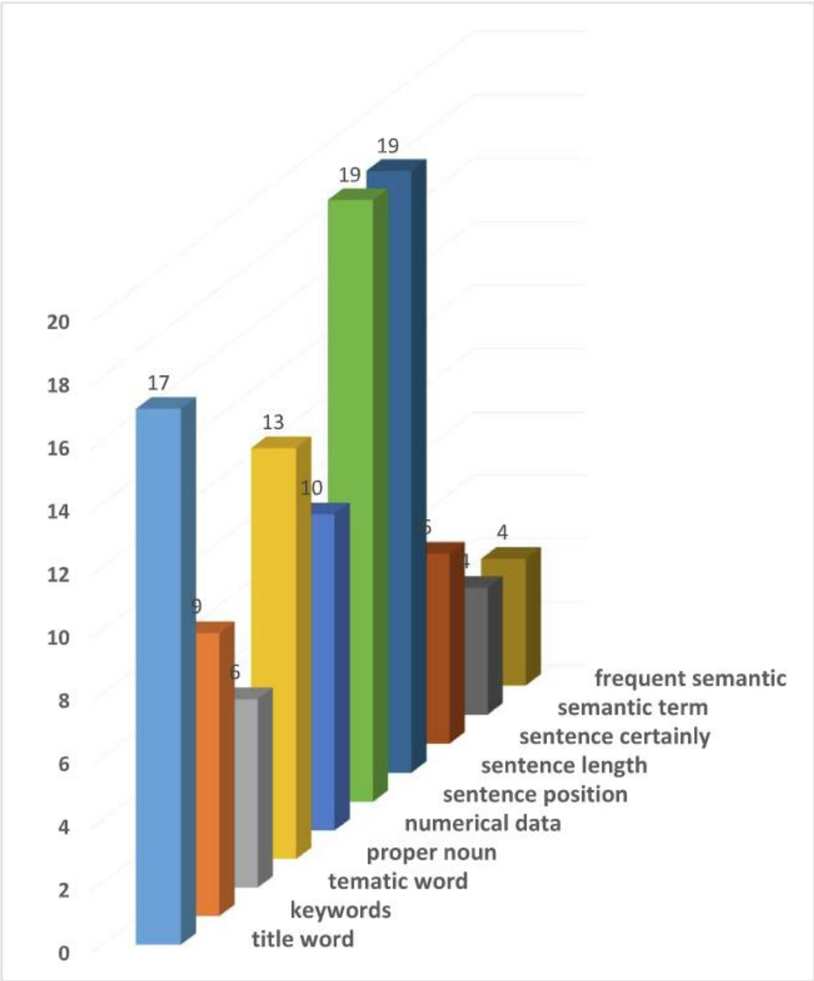
Automatic text summarization: A comprehensive survey. Wafaa S.El-Kassasa, Cherif R.Salamaab, Ahmed A.Rafeab, Hoda K.Mohameda. ESWA 2021

Text preprocessing



. P. Widyassari, S. Rustad, G. F. Shidik et al., Review of automatic text summarization techniques & methods, Journal of King SaudUniversity – Computer and Information Sciences,https://doi.org/10.1016/j.jksuci.2020.05.006

Features



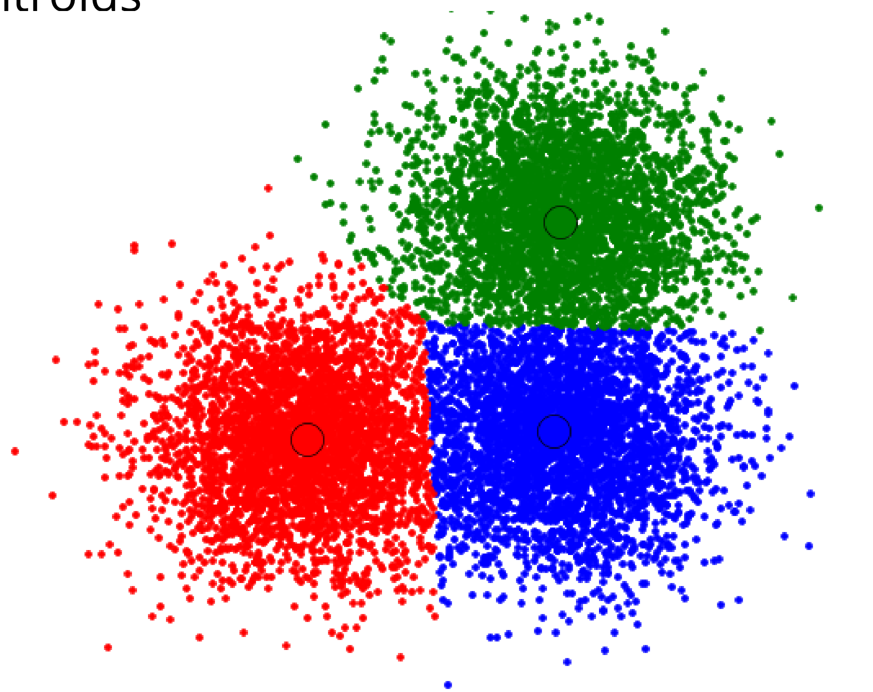
- Title word
  - similarity of a word to the title
  - the greater the similarity of words to the title means the more likely the word is included in the summary.
- Keyword
  - a word that has a large judgment or main word that often appears in a sentence and words that are important words in a sentence.
  - LSA, SVD, TF-IDF
- Sentence position score
  - Variable that shows the position of the sentence, N is a variable that shows the total sentence in the document, and i is a variable that shows the sentence.

$$SP_i = 1 - \frac{i-1}{N}$$

. P. Widyassari, S. Rustad, G. F. Shidik et al., Review of automatic text summarization techniques & methods, Journal of King SaudUniversity – Computer and Information Sciences,https://doi.org/10.1016/j.jksuci.2020.05.006

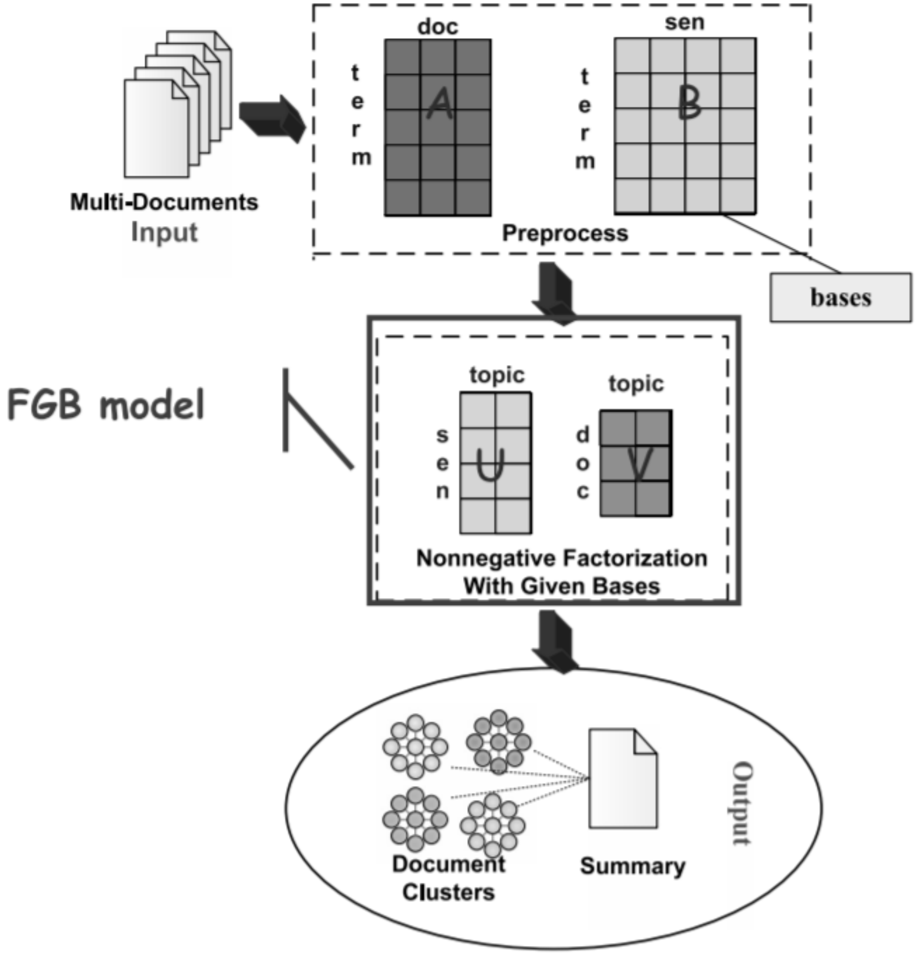
Clustering-based approaches

- Group homogenous documents/sentences/words using clustering algorithms
- Pick cluster representatives
  - E.g., medoids, centroids



Dingding Wang, Shenghuo Zhu, Tao Li, Yun Chi, and Yihong Gong. 2011. Integrating Document Clustering and Multidocument Summarization. *ACM Trans. Knowl. Discov. Data* 5, 3, Article 14 (August 2011), 26 pages. DOI=http://dx.doi.org/10.1145/1993077.1993078

Example of clustering-based approach

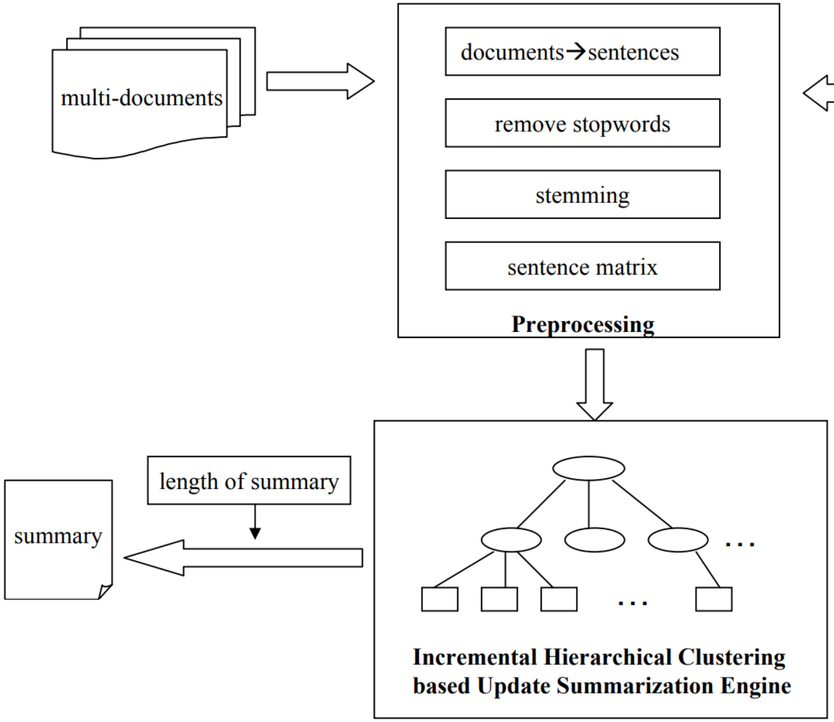


- Get the document-term matrix and the sentence-term matrix
- Perform non-negative factorization on the document-term matrix using the sentence-term matrix as the basis
- Obtain the document-topic matrix and sentence-topic matrices
- Generate the document clusters
- The corresponding summaries are generated simultaneously

Dingding Wang, Shenghuo Zhu, Tao Li, Yun Chi, and Yihong Gong. 2011. Integrating Document Clustering and Multidocument Summarization. ACM Trans. Knowl. Discov. Data 5, 3, Article 14 (August 2011), 26 pages. DOI:https://doi.org/10.1145/1993077.1993078

Example of clustering-based approach

- Document update summarization based on incremental hierarchical clustering
  - Select the most representative sentences to summarize each node of the hierarchy and its subtrees (each leaf node is a sentence)
  - Once a new sentence arrives, the sentence hierarchy is updated accordingly



Dingding Wang and Tao Li. 2010. Document update summarization using incremental hierarchical clustering. In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10). ACM, New York, NY, USA, 279-288.

Clustering-based approaches: pros and cons

Pros:

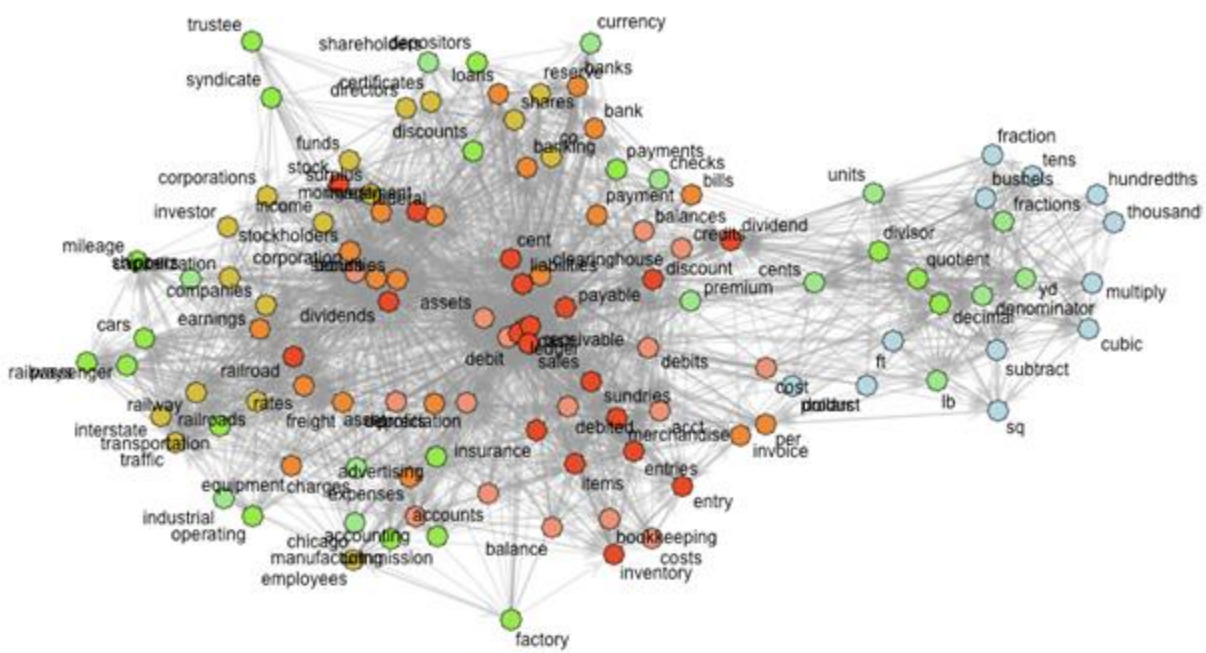
- Language-agnostic
- Incremental
  - Via hierarchical clustering
- Fairly robust to noise
  - density-based clustering

Cons:

- Limited effectiveness on complex document collections compared to other techniques
  - E.g., itemset-based models, LSA-based techniques

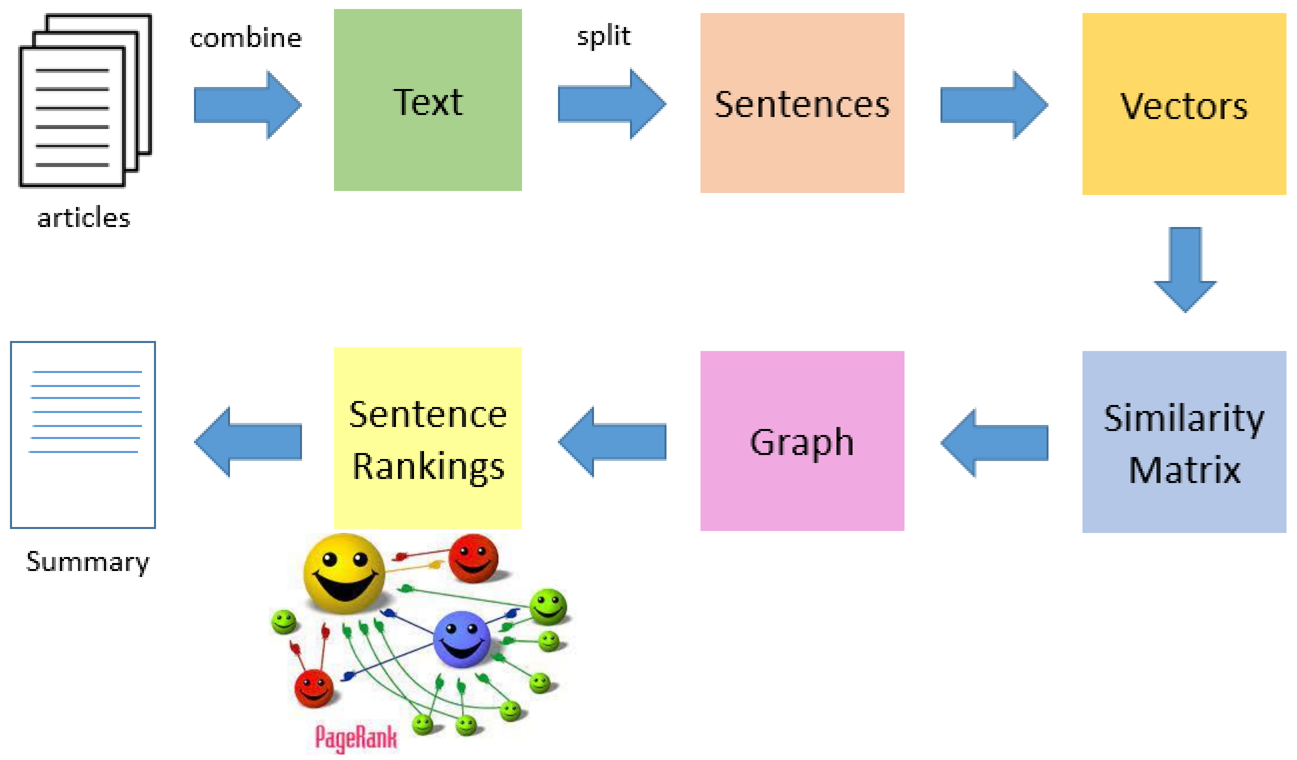
Graph-based approaches

- Document content is modelled as a graph
  - Terms or sentences are graph nodes
  - Edges are weighted by co-occurrence-based measures
- Graph are ranked using popular algorithms (e.g., PageRank, HITS) to shortlist the most relevant content



Graph-based approach: TextRank

- Unsupervised graph-based approach based on the concept of graph centrality



TextRank: Bringing Order into Texts. Rada Mihalcea and Paul Tarau

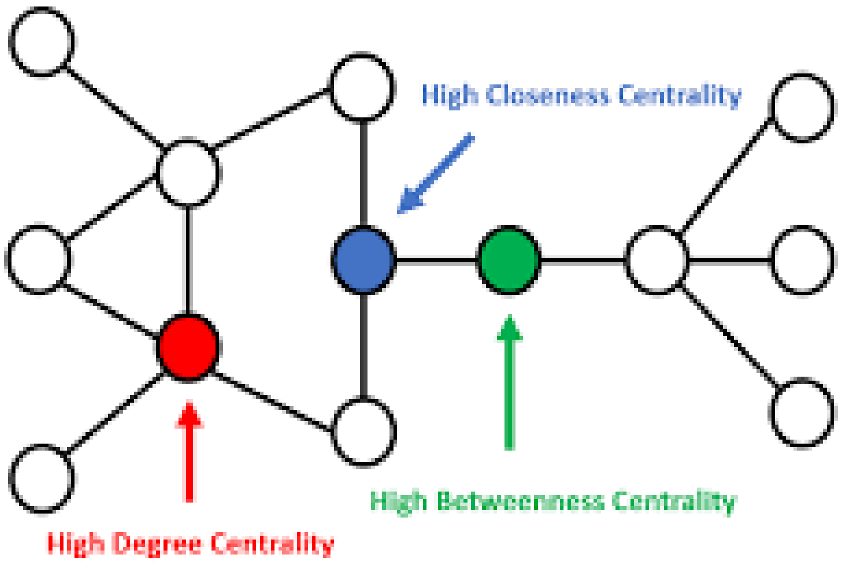
Graph-based approach: TextRank

- Main steps
  1. Identify text units that best define the task at hand and add them as vertices in the graph
     - E.g., sentences, keyphrases
  2. Identify relations that connect such text units, and use these relations to draw edges between vertices in the graph
     - Edges can be directed or undirected, weighted or unweighted
     - To compute keyword similarity use a co-occurrence relation, controlled by the distance between word occurrences
       - two vertices are connected if their corresponding lexical units co-occur within a window of maximum word
  3. Iterate the graph-based ranking algorithm until convergence
  4. Sort vertices based on their final score
     - Use the values attached to each vertex for ranking/selection decisions

TextRank: Bringing Order into Texts. Rada Mihalcea and Paul Tarau

Graph-based approach: LexRank

- Unsupervised graph-based approach based on the concept of graph centrality



G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approach: LexRank

- Key elements
  - Many sentences are expected to be somehow similar to each other since they are all about the same topic.
  - If one sentence is very similar to many others, it will likely be a sentence of great importance.
  - Sentences *recommend* other similar sentences to the reader.
    - The importance of this sentence also stems from the importance of the sentences "recommending" it.
  - To get ranked highly and placed in a summary, a sentence must be similar to many sentences that are, in turn, also similar to many other sentences.

G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approaches: LexRank

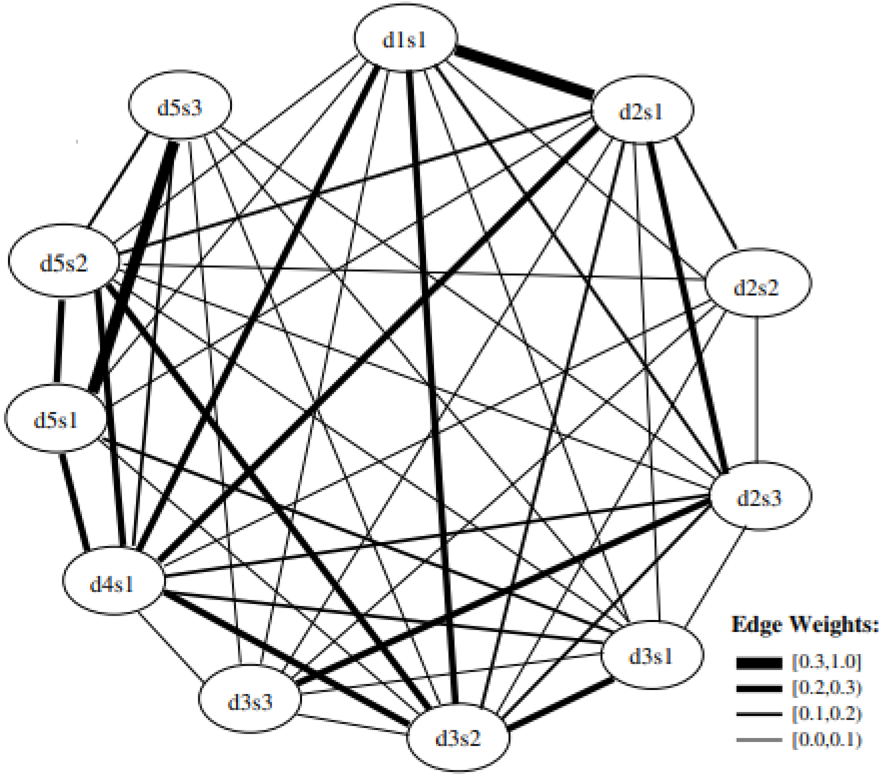
- Step 1
  - compute the cosine similarity matrix between the tf-idf sentence vectors

	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.45	0.02	0.17	0.03	0.22	0.03	0.28	0.06	0.06	0.00
2	0.45	1.00	0.16	0.27	0.03	0.19	0.03	0.21	0.03	0.15	0.00
3	0.02	0.16	1.00	0.03	0.00	0.01	0.03	0.04	0.00	0.01	0.00
4	0.17	0.27	0.03	1.00	0.01	0.16	0.28	0.17	0.00	0.09	0.01
5	0.03	0.03	0.00	0.01	1.00	0.29	0.05	0.15	0.20	0.04	0.18
6	0.22	0.19	0.01	0.16	0.29	1.00	0.05	0.29	0.04	0.20	0.03
7	0.03	0.03	0.03	0.28	0.05	0.05	1.00	0.06	0.00	0.00	0.01
8	0.28	0.21	0.04	0.17	0.15	0.29	0.06	1.00	0.25	0.20	0.17
9	0.06	0.03	0.00	0.00	0.20	0.04	0.00	0.25	1.00	0.26	0.38
10	0.06	0.15	0.01	0.09	0.04	0.20	0.00	0.20	0.26	1.00	0.12
11	0.00	0.00	0.00	0.01	0.18	0.03	0.01	0.17	0.38	0.12	1.00

G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approach: LexRank

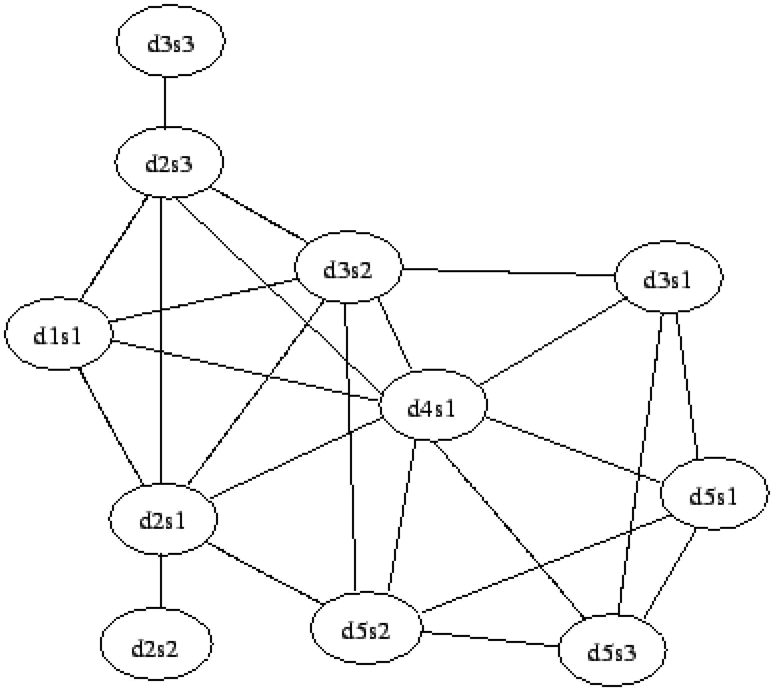
- Step 2
  - Prune the edges corresponding to low similarity values



G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approaches: LexRank

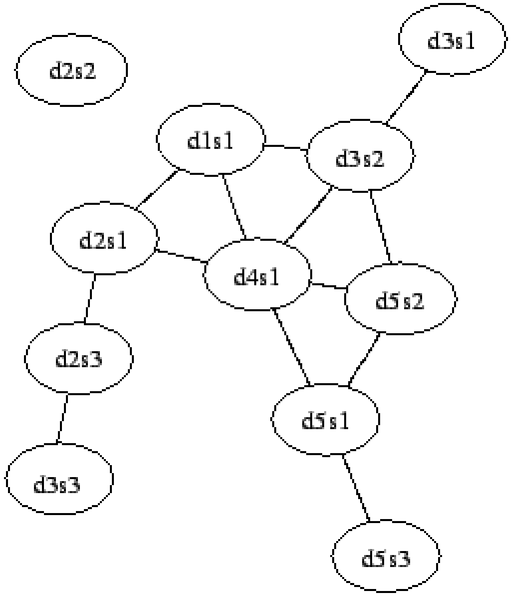
- Step 3
  - Generate the unweighted similarity graph (thr=0.1)



G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approaches: LexRank

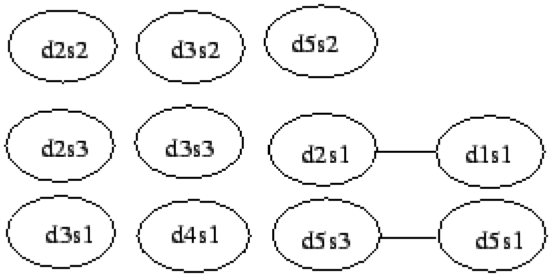
- Step 3
  - Generate the unweighted similarity graph (thr=0.2)



G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approaches: LexRank

- Step 3
  - Generate the unweighted similarity graph (thr=0.3)



G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approaches: LexRank

- Step 4
  - Compute the degree centrality score of each sentence
    - The degree centrality of a sentence is the degree of the corresponding node in the similarity graph
    - Each edge is a vote to determine the overall sentence centrality

ID	Degree (0.1)	Degree (0.2)	Degree (0.3)
d1s1	5	4	2
d2s1	7	4	2
d2s2	2	1	1
d2s3	6	3	1
d3s1	5	2	1
d3s2	7	5	1
d3s3	2	2	1
d4s1	9	6	1
d5s1	5	4	2
d5s2	6	4	1
d5s3	5	2	2

G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approaches: LexRank

- Step 5
  - Apply PageRank to the graph built on top of the adjacency matrix

Centrality of node u

$$p(u) = \sum_{v \in adj[u]} \frac{p(v)}{deg(v)}$$

Where Adj[u] is the set of nodes that are adjacent to u Deg(v) is the degree of node v

The adjacency matrix B is defined by

$$\mathbf{B}(i,j) = \frac{\mathbf{A}(i,j)}{\sum_k \mathbf{A}(i,k)}$$

G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approaches: LexRank

- Step 5(b)
  - Variant of the centrality of node u when the similarity graph is weighted:

Centrality of node u

$$p(u) = \frac{d}{N} + (1-d) \sum_{v \in adj[u]} \frac{\text{idf-modified-cosine}(u,v)}{\sum_{z \in adj[v]} \text{idf-modified-cosine}(z,v)} p(v)$$

Where the strength of the similarity links is obtained using the cosine similarity between the sentence vectors, weights are normalized by the row sums and a damping factor *d* is added for the convergence of the method

Notice that there exist variants using BM25, BERT similarity, etc.

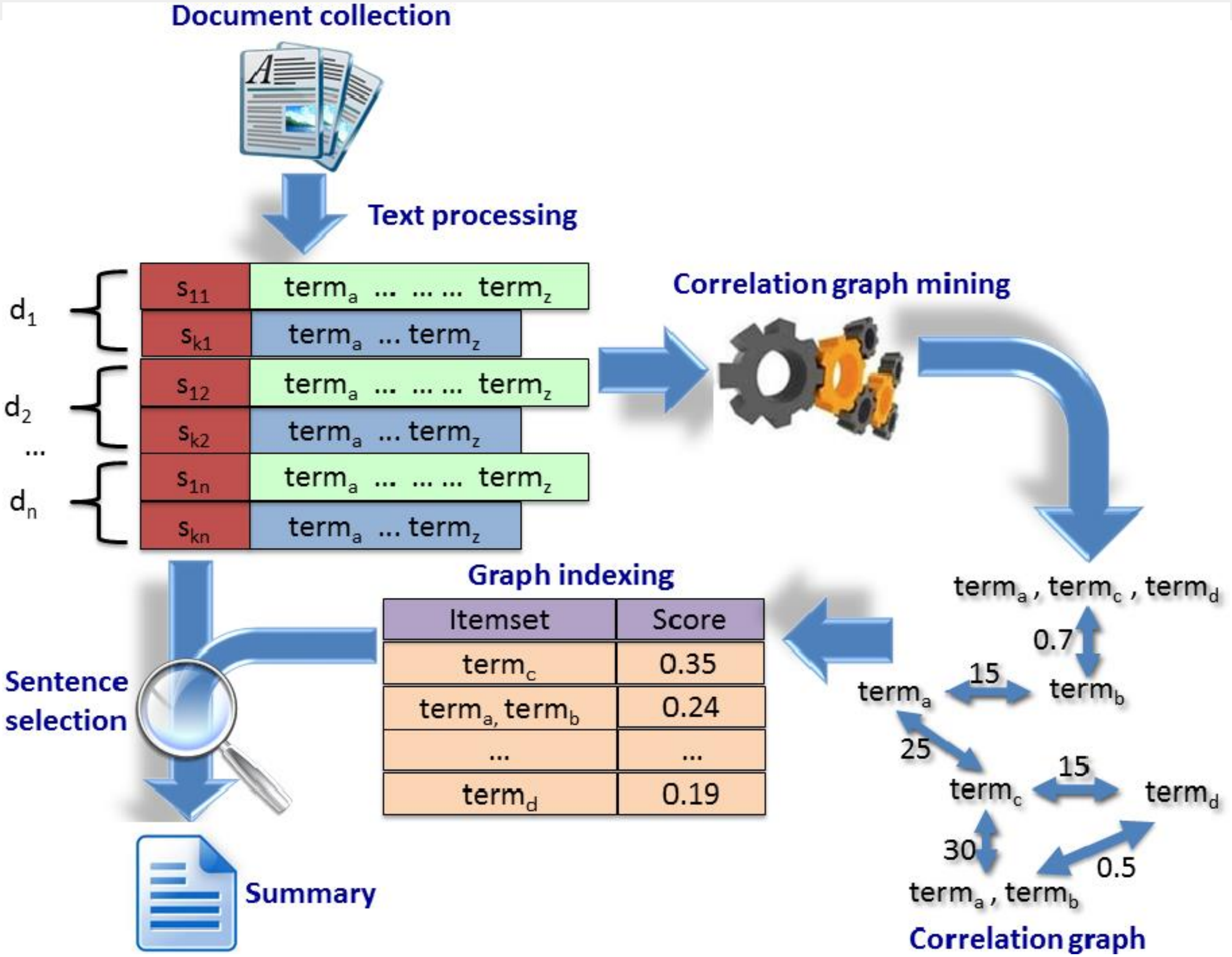
G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approach: GraphSum

- Unsupervised graph-based approach relying on both association rule mining and graph centrality scoring of Bag-Of-Words
- Intuition
  - Graph nodes are itemsets consisting of Bag-Of-Words
    - Rather than single terms or entire sentences
  - Graph edges are association rules and are weighted by the confidence level of the extracted association
    - Only frequent and positively correlated rules are considered
  - Graph ranking is based on HITS

G. Erkan and D. Radev. Lexpagerank: Prestige in multi-document text summarization. In Proceedings of EMNLP 2004, 2004

Graph-based approach: GraphSum



Elena Baralis, Luca Cagliero, Naeem A. Mahoto, Alessandro Fiori.
GraphSum: Discovering correlations among multiple terms for graph-based summarization. Inf. Sci. 249: 96-109 (2013)

Graph-based approaches: pros and cons

Pros:

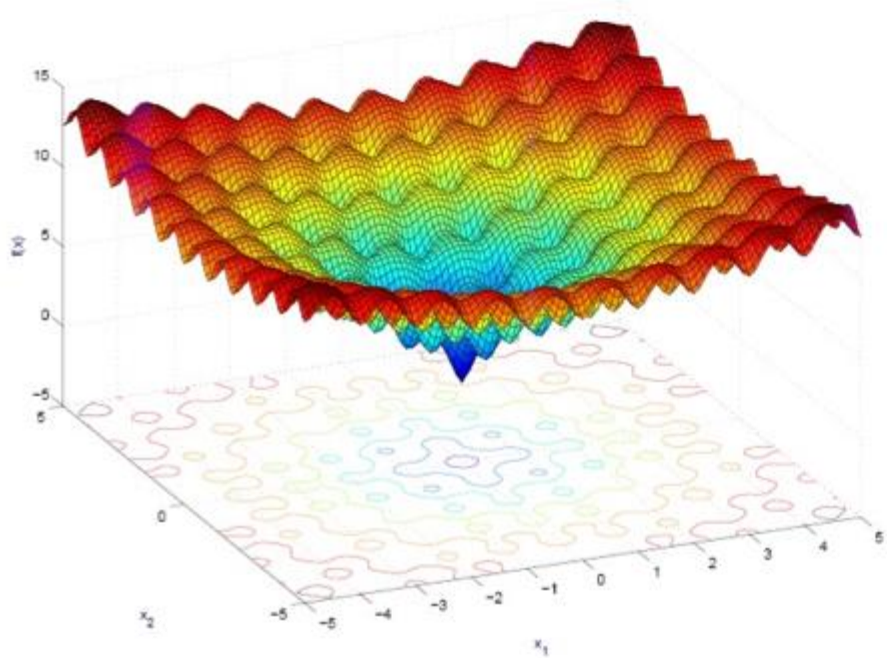
- High effectiveness
- Easy-to-use

Cons:

- High computational complexity
- Not incremental

Optimization-based approaches

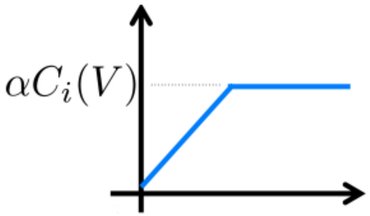
- Shortlist document sentences by exploring the search space and optimizing a given objective function
  - Exploit complementary techniques/models such as Integer Linear Programming and Submodular Functions



Optimization-based approaches: SubModular

- The summarization task is modeled as a Knapsack problem
  - Given a set of objects $V = \{v_1, \ldots, v_n\}$ and a function $F : 2^V \rightarrow \mathbb{R}$ that returns real values for any subset $S \subseteq V$
  - Find the subset of bounded size $|S| <= K$, i.e., $\arg\max_{S \subseteq V} F(S)$

$$F(S) = \underbrace{R(S)}_{\substack{\uparrow \\ \text{Relevance}}} + \underbrace{\lambda D(S)}_{\substack{\uparrow \\ \text{Diversity}}}$$



$$R(S) = \sum_i \min\{C_i(S), \alpha C_i(V)\}$$

$$C_i(S) = \sum_{j \in S} \omega_{i,j}$$

$C_i(S)$: How well is sentence i "covered" by S

$\omega_{i,j}$: Similarity between i and j

Lin and Bilmes. A Class of Submodular Functions for Document Summarization. ACL 2011

Optimization-based approaches: SubModular

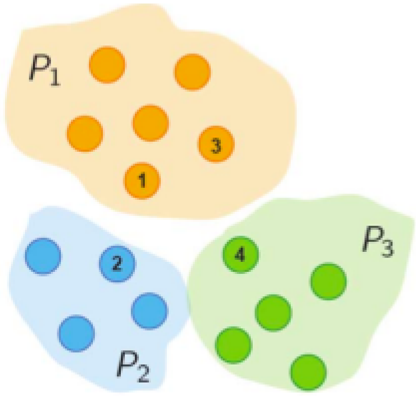
- Diversity of a summary

$$D(S) = \sum_{i=1}^{K} \sqrt{\sum_{j \in P_i \cap S} r_j}$$

$$r_j = \frac{1}{N} \sum_{i} \omega_{i,j}$$

r_j: Relevance of sentence j to doc.

$\omega_{i,j}$: Similarity between i and j



Clustering of sentences in document

Lin and Bilmes. A Class of Submodular Functions for Document Summarization. ACL 2011

Optimization-based approaches: COSUM

- Combine clustering with optimization for multi-document summarization
  - Cluster sentences into homogeneous groups using K-Means
  - Select salient sentences from the clusters using optimization techniques
    - Objective function: harmonic mean of the objective functions enforcing the coverage and diversity of the selected sentences in the summary

COSUM: Text summarization based on clustering and optimization. Rasim M. Alguliyev,Ramiz M. Aliguliyev,Nijat R. Isazade,Asad Abdi,Norisma Idris. Expert Sytstems. Wiley Online Library. https://doi.org/10.1111/exsy.12340

Optimization-based approaches: Maximal Marginal Relevance

- Summarization framework that balances summary salience and redundancy
- Score of sentence s_j

$$m_j^t = \lambda\, S(s_j, D) - (1 - \lambda)\, \max_{s^* \in CurSum} R(s_j, s^*)$$

Where λ (between 0 and 1) is the weight balancing saliency and redundancy

Optimization-based approaches: pros and cons

Pros:

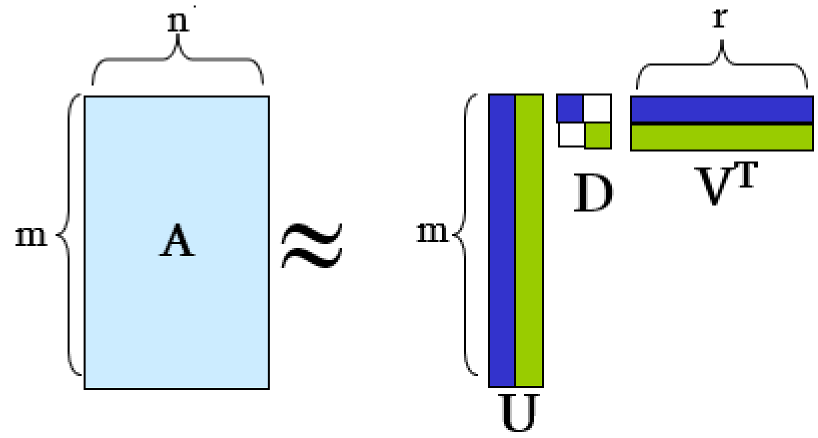
- Good performance
- Possibility to combine ML and optimization

Cons:

- Limited scalability
- Limited explainability

LSA-based summarization

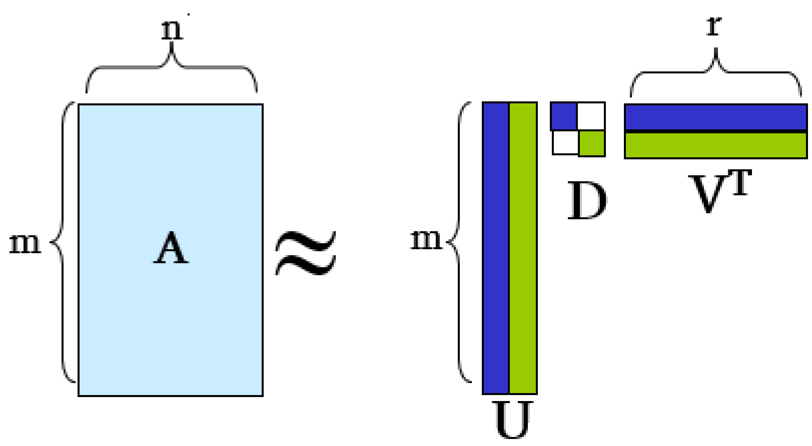
- Transform a sentence-by-term matrix into a sentence-by-latent topic matrix using the SVD decomposition



LSA-based Multi-Document Summarization. J. Steinberger.

LSA-based summarization

- Pick the sentences that best cover the latent topics in the approximated matrix A'
  - Eigenvectors are filtered to keep only the most significant features
    - Heuristic approach
  - The approximated matrix A' is iteratively updated to avoid redundancy



LSA-based Multi-Document Summarization. J. Steinberger. 2011

LSA-based summarization: pros and cons

Pros:

- Simplicity
- Scalability
- Language independence

Cons:

- It considers only word-level relations

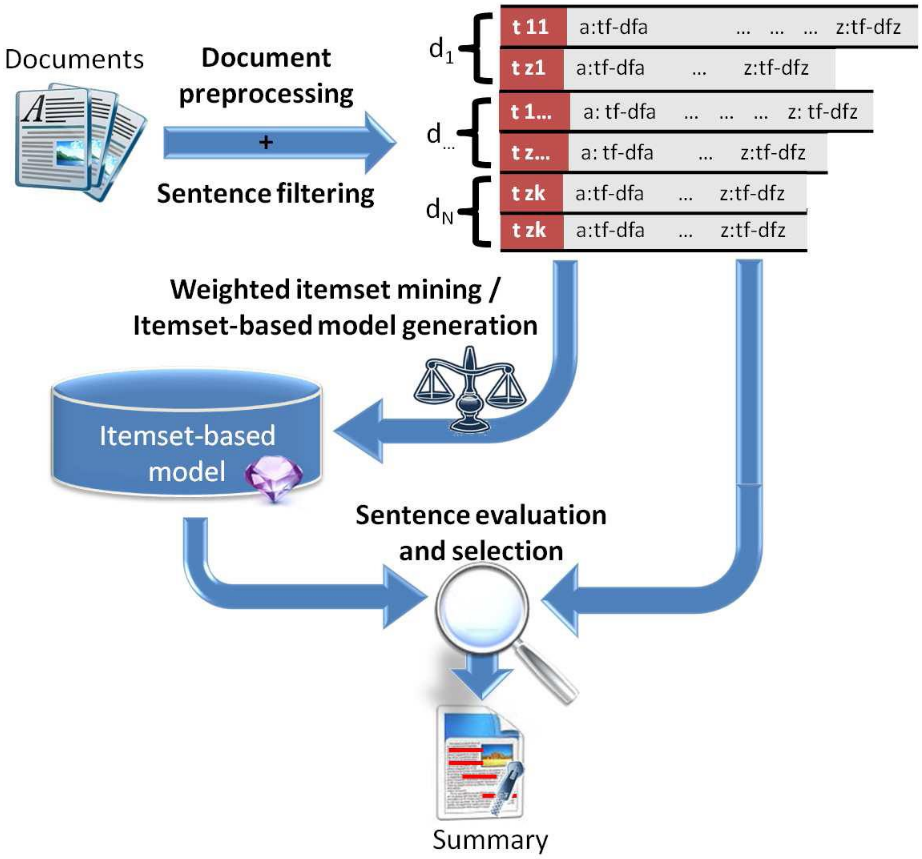
Itemset-based summarization

- Perform sentence selection on top of an itemset-based model
  - Itemsets represent co-occurrences among multiple document terms
    - Go beyond word-level associations
  - Adaptative summary size
    - Depending on the itemset distribution

LSA-based Multi-Document Summarization. J. Steinberger.

Itemset-based summarization

- Key steps



Elena Baralis, Luca Cagliero, Alessandro Fiori, Paolo Garza.
MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets. ACM Trans. Inf. Syst. 34(1): 5:1-5:35 (2015)

Itemset-based summarization

- An itemset is a set of terms of arbitrary length
- An itemset occurs in a given transaction if all its items (terms) are contained in the corresponding sentence
  - E.g., itemset {treat, diseas} occurs in sentence
- Item occurrences in the transactions are weighted

Elena Baralis, Luca Cagliero, Alessandro Fiori, Paolo Garza.
MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets. ACM Trans. Inf. Syst. 34(1): 5:1-5:35 (2015)

Itemset-based summarization

- Item occurrences in the transactions are weighted
- Itemset occurrences are weighted
  - The occurrence weight of an itemset in a transaction is the least item weight
    - E.g., if items *treat* and *diseas* in sentence 1 have weights 0.5 and 0.7, respectively, then {*treat, diseas*} has occurrence weight equal to 0.5
    - Most conservative option

Elena Baralis, Luca Cagliero, Alessandro Fiori, Paolo Garza. MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets. ACM Trans. Inf. Syst. 34(1): 5:1-5:35 (2015)

 |

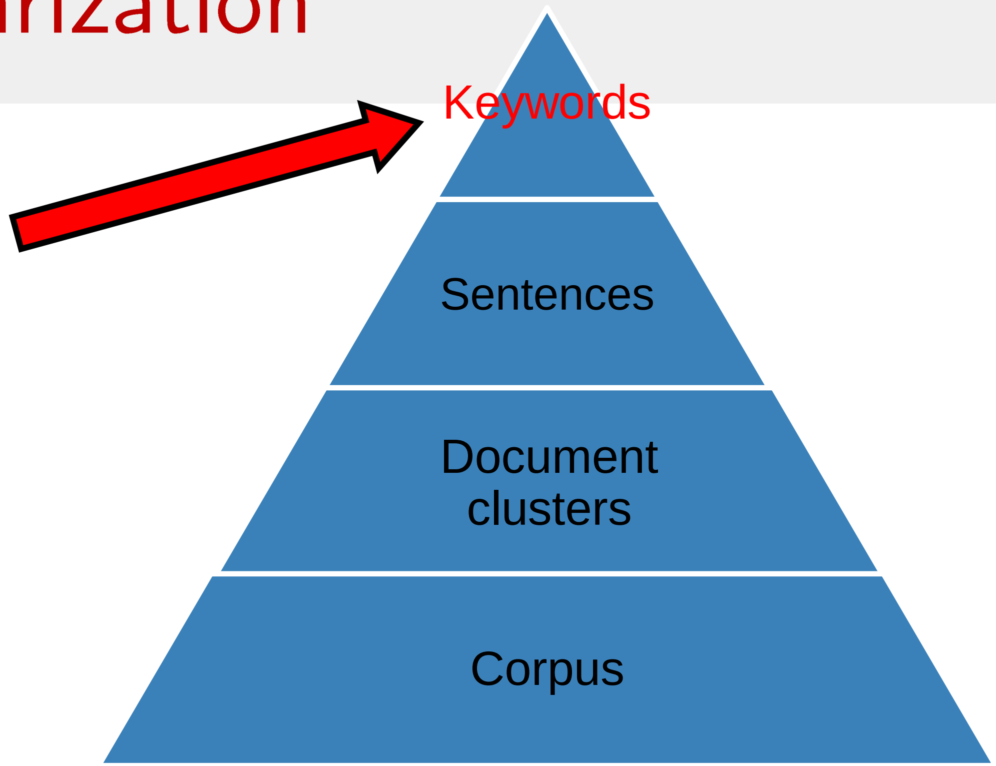
Itemset-based summarization

- Extract all the possible combinations of terms whose weighted frequency of occurrence in the analyzed data is above a given threshold
  - The weighted frequency of occurrence (weighted support) of an itemset is the average occurrence weight in the analyzed dataset
    - E.g., if itemset {treat, diseas} occurs in sentence 1 with weight 0.5 and in sentence 2 with weight 0.7 the weighted support of {treat, diseas} is 0.6

Elena Baralis, Luca Cagliero, Alessandro Fiori, Paolo Garza.
MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets. ACM Trans. Inf. Syst. 34(1): 5:1-5:35 (2015)

Itemset-based summarization

Gener Diseas (0.778)
Diseas Disorder (0.774)
Stem Cell (0.671)
…
Stem Cell Esophag Tissue (0.215)
...

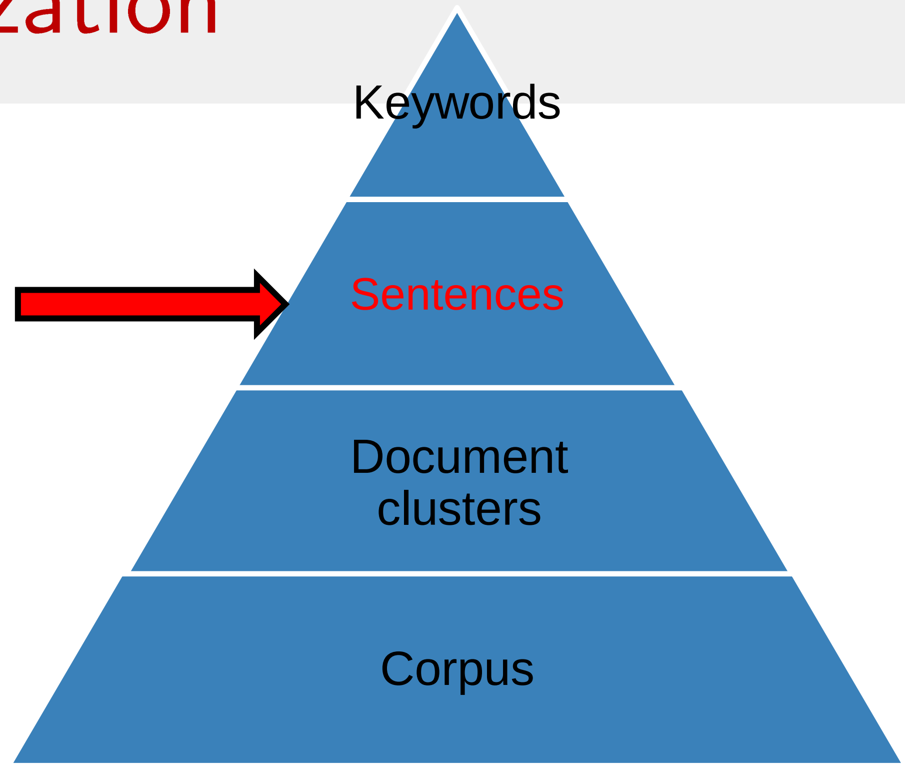


- Combinations of two or more terms
  - Longer combinations provide more specific context
  - Weighted support value (in brackets) indicate relative importance in the document collection

Elena Baralis, Luca Cagliero, Alessandro Fiori, Paolo Garza.
MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets. ACM Trans. Inf. Syst. 34(1): 5:1-5:35 (2015)

Itemset-based summarization

This treatment for a disease …
~~*This is another treatment …*~~
This work covers different topics …



- Select the sentences that best correspond to the mined itemsets
  - Sentences covered by many frequent itemsets are most likely to represent informative content

Elena Baralis, Luca Cagliero, Alessandro Fiori, Paolo Garza.
MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets. ACM Trans. Inf. Syst. 34(1): 5:1-5:35 (2015)

 |

Itemset-based summarization

- Sentence selection
  - Optimal solution
    - Select the minimal set of sentences that maximizes itemset coverage
    - Minimum redundancy
      - Minimal number of sentences
    - Maximal information
      - Maximal number of itemsets per sentence
- Greedy heuristics
  - Iterate over the set of sentences
  - At each iteration, pick the sentence that is covered by the maximal number of itemsets
  - Stop when all the mined itemsets cover a sentence

Elena Baralis, Luca Cagliero, Alessandro Fiori, Paolo Garza.
MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets. ACM Trans. Inf. Syst. 34(1): 5:1-5:35 (2015)

Itemset-based summarization

- Sentence ordering matters
  - The first sentence selected by the greedy heuristics is potentially the most informative one
  - It should appear first in the summary
- No fixed number of selected sentences
  - It depends on the data distribution
  - Low ranked sentences can be postpruned if need be

Elena Baralis, Luca Cagliero, Alessandro Fiori, Paolo Garza.
MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets. ACM Trans. Inf. Syst. 34(1): 5:1-5:35 (2015)

Itemset-based approaches: pros and cons

Pros:

- High performance
- Model explainability
- Language independence

Cons:

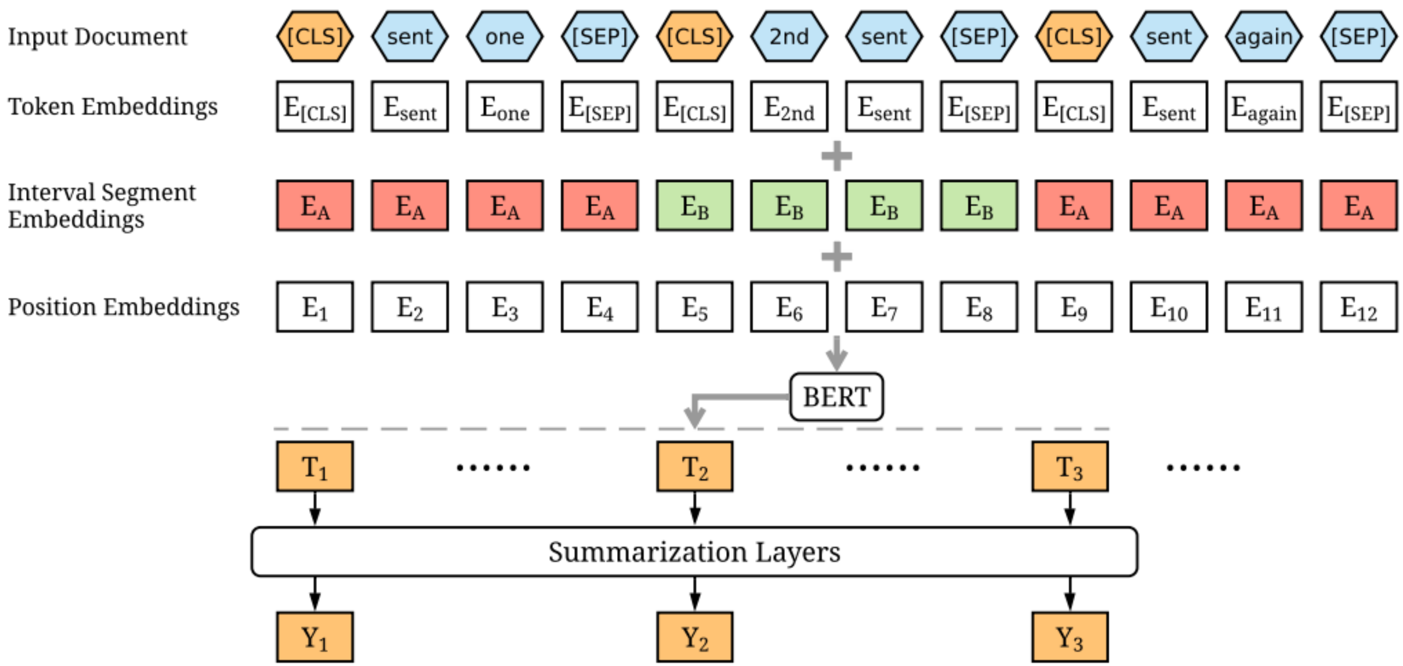
- High computational complexity
- Limited scalability

Neural summarization

- Estimate sentence relevance using Deep Learning models
  - Self-supervised
    - Learn vector text representations (e.g., BERT)
    - Compute semantic similarity using cosine similarity
    - Apply traditional methods to the encoded vectors
      - E.g., Apply TextRank to BERT encodings: Lab practice
  - Supervised
    - Train a sentence classification model using humanly generated summaries
    - Apply the model to new documents to extract the summary

Neural summarization: BERTSum

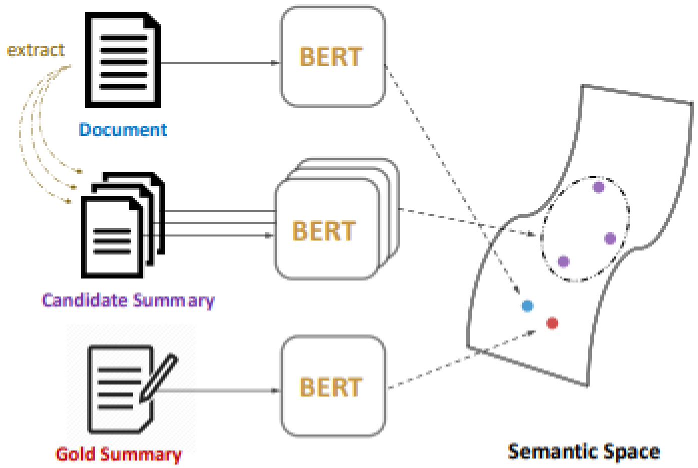
- The extractive summarization task is modeled as a **sentence classification task**
  - **True** if the sentence belongs to the target summary, **False** otherwise
- The interval segment embedding is used to separate the input sentences



Yang Liu, Mirella Lapata: Text Summarization with Pretrained Encoders. EMNLP/IJCNLP (1) 2019: 3728-3738

Neural summarization: MatchSum

- Intuition
  - A good summary should be more semantically similar as a whole to the source document than the unqualified summaries
  - Sentences are prepruned to limit the number of possible combinations
    - BERTSum
  - Candidate summaries are combinations of the shortlisted sentences



Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, Xuanjing Huang (2020). Extractive Summarization as Text Matching. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 6197–6208. July 5 - 10, 2020. Association for Computational Linguistics

Neural summarization: MatchSum

- The extractive summarization task is modeled as a **semantic text matching problem**
  - It matches the contextual representations of the document with gold summary and candidate sentences extracted from the document
  - Better candidate summaries should be semantically closer to the document and the gold summary should be the closest

Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, Xuanjing Huang (2020). Extractive Summarization as Text Matching. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 6197–6208. July 5 - 10, 2020. Association for Computational Linguistics

Neural approaches: pros and cons

Pros:

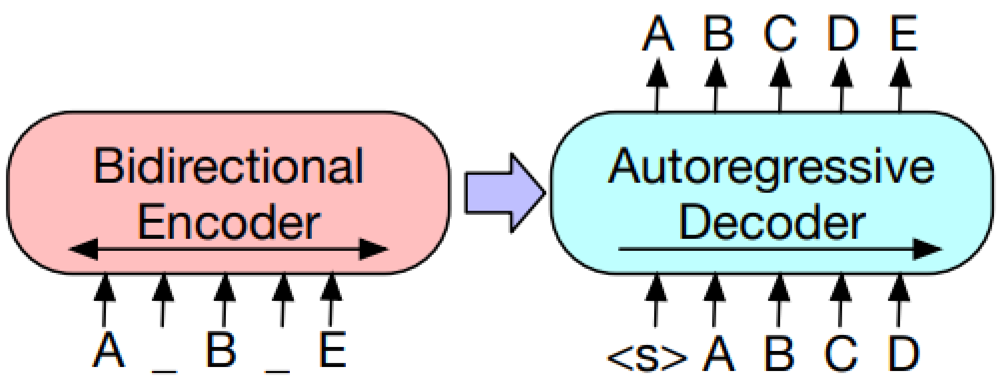
- High performance
- Semantic-aware

Cons:

- Need for large-scale datasets
  - Annotated for supervised learning
- Require expensive HW resources
- Low explainability

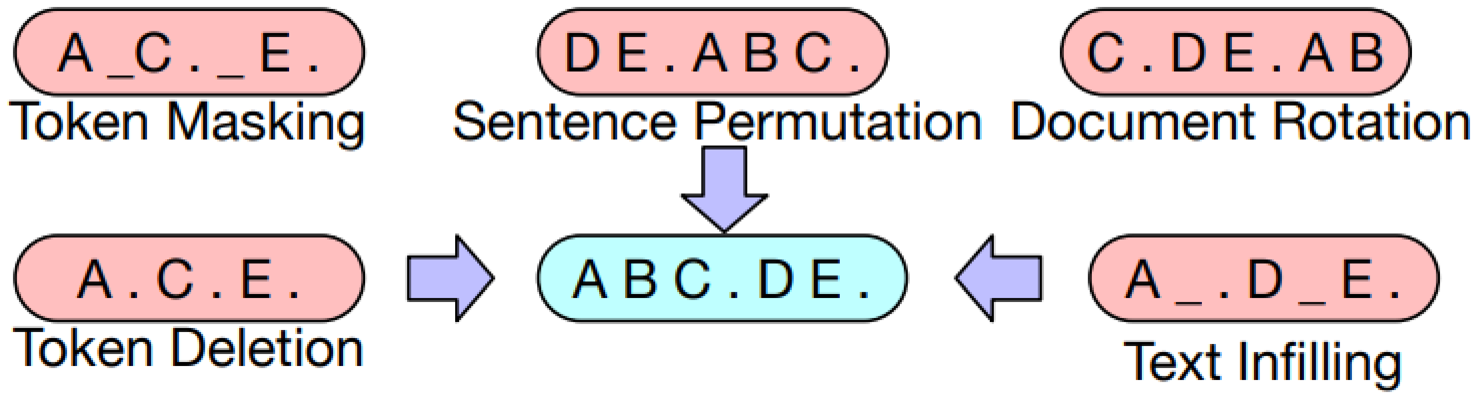
Abstractive neural summarization: BART

- Supervised method for Natural Language Generation, Translation, and Comprehension by **Facebook Inc.**
- Based on pretraining and fine-tuning like BERT with a new **noising strategy**
  - The corrupted document (left) is encoded with a bidirectional model first
  - Then the likelihood of the original document (right) is calculated with an autoregressive decoder.
- For fine-tuning, an uncorrupted document is provided as input to both the encoder and decoder
  - Use the representations from the final hidden state of the decoder



BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer. ACL 2019

BART transformation for noising



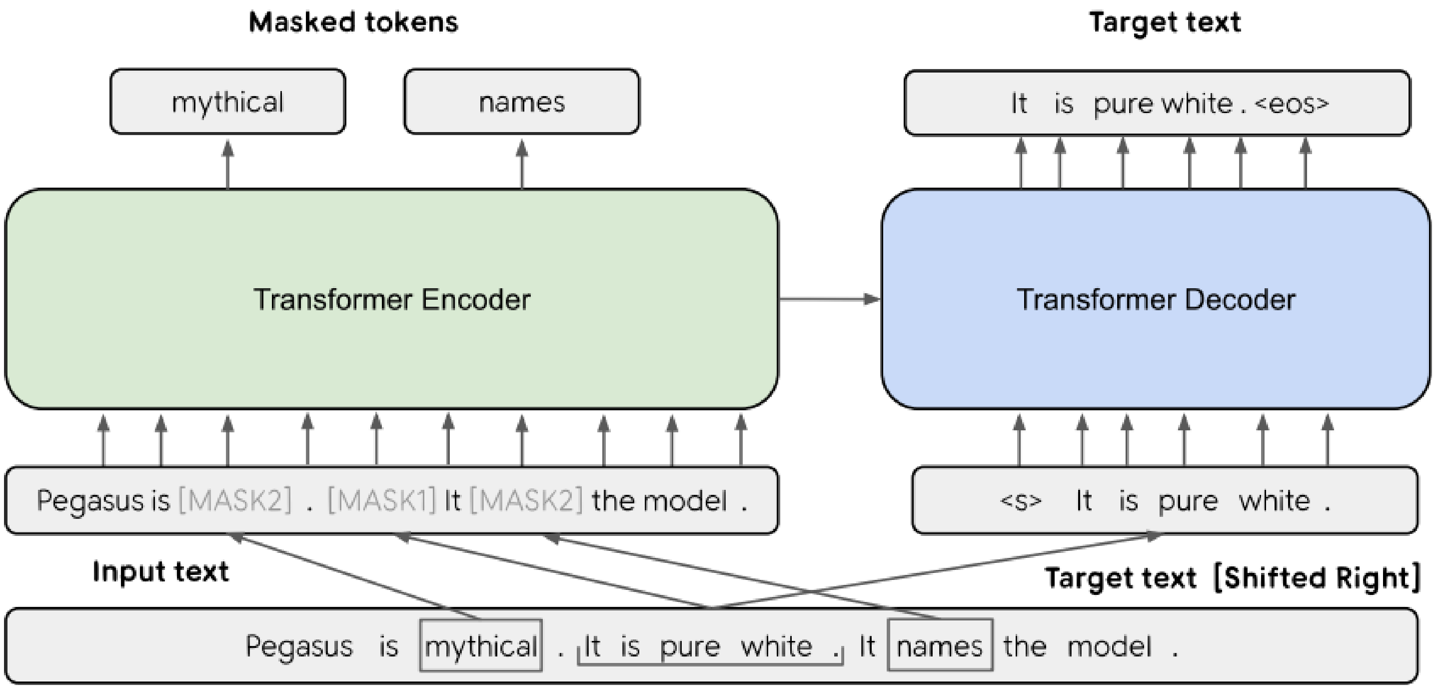
BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer. ACL 2019

Abstractive neural summarization: PEGASUS

- Pre-training large Transformer-based encoder-decoder model on massive text corpora by **Google Inc.**
- New self-supervised pretraining objective
  - **Apply Gap Sentence Generation** and **Masked Language Modeling** simultaneously
- Three sentences
  - One sentence is masked with [MASK1] and used as target generation text (GSG)
  - The other two sentences remain in the input, but some tokens are randomly masked by [MASK2] (MLM)

PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. Jingqing Zhang, Yao Zhao, Mohammad Saleh, Peter J. Liu. ACL 2020

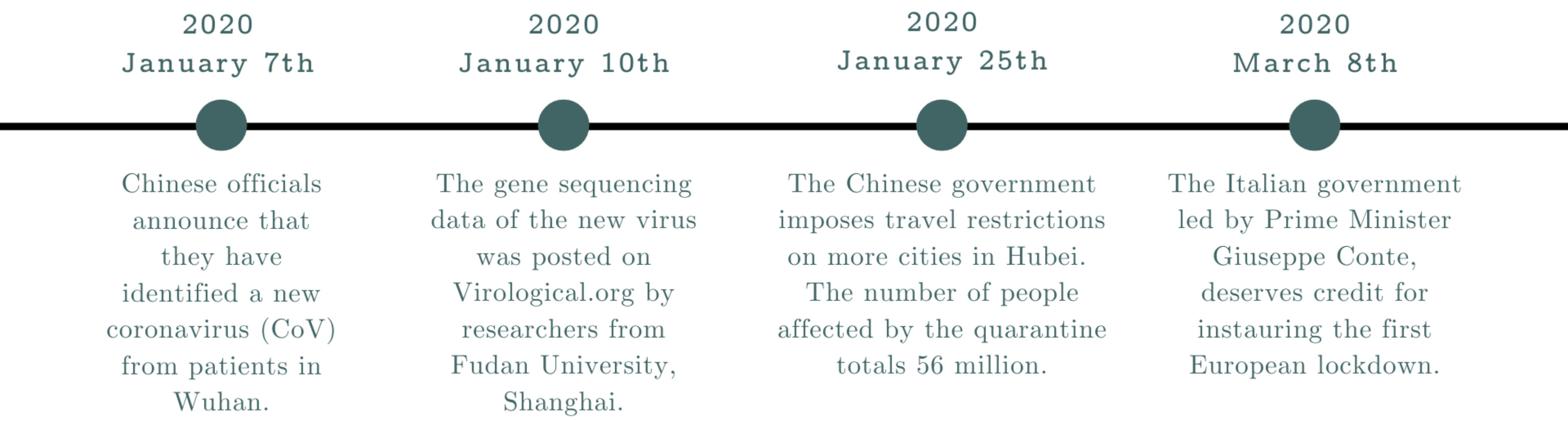
Abstractive neural summarization: PEGASUS



PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. Jingqing Zhang, Yao Zhao, Mohammad Saleh, Peter J. Liu. ACL 2020

TimeLine summarization

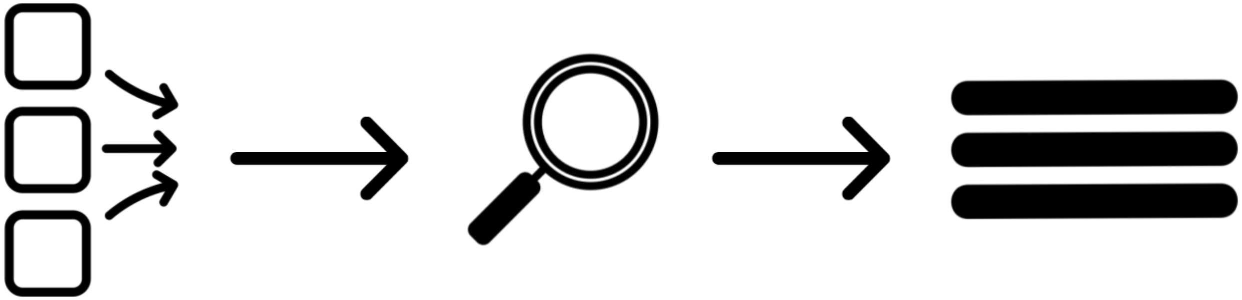
- A timeline is a chronological arrangement of a given topic
    - It includes the relevant dates and key insights for each of them



Source: https://covidreference.com/timeline

TimeLine summarization

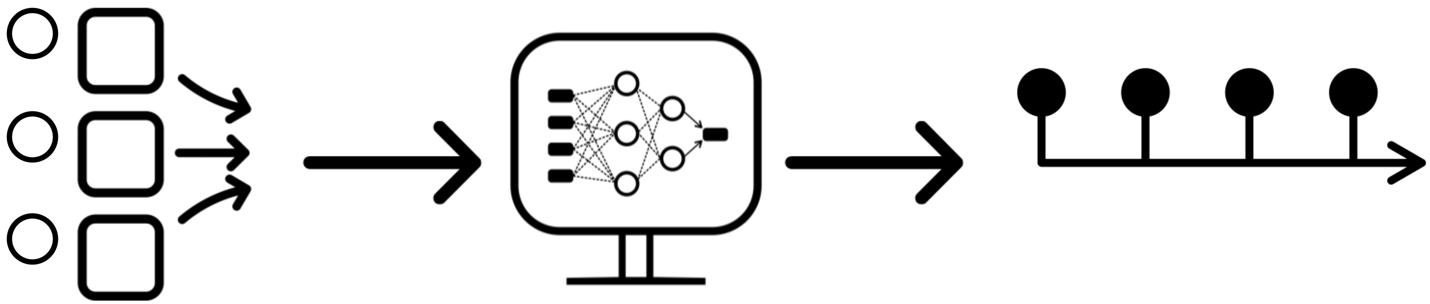
- How does TLS work?
  - Given one or more documents, the system generate a shorten version conveying the most relevant information



* Credit: Image components by David Christensen from the Noun Project.

TimeLine summarization

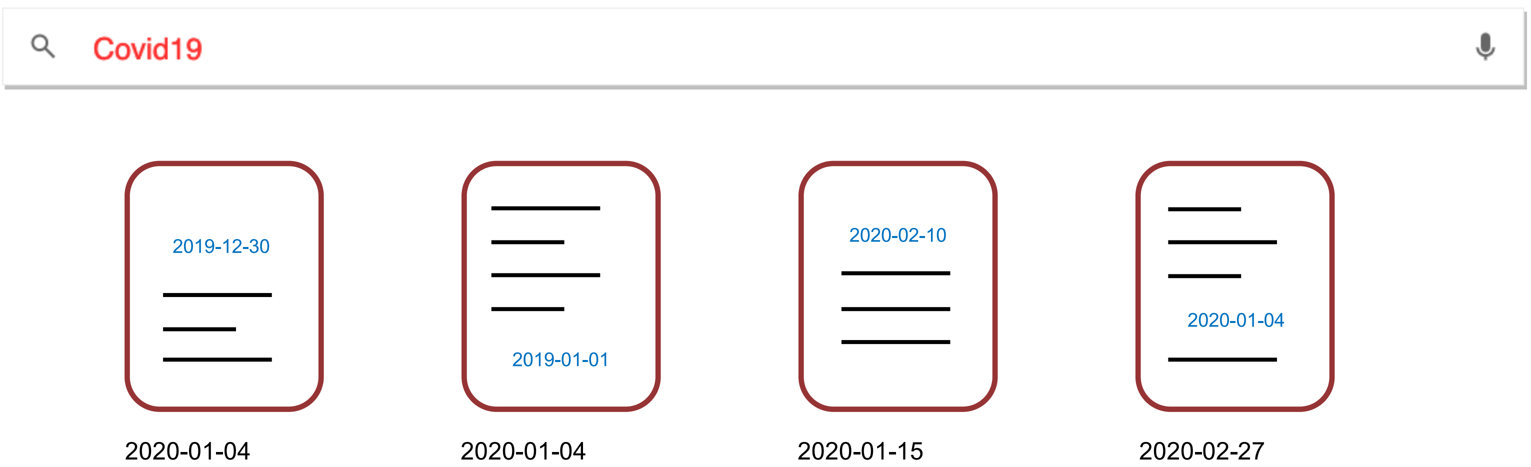
- How does TLS work?
  1. Retrieve key dates related to a specific event
  2. Summarize and track event evolution



* Credit: Image components by David Christensen from the Noun Project.

TimeLine summarization

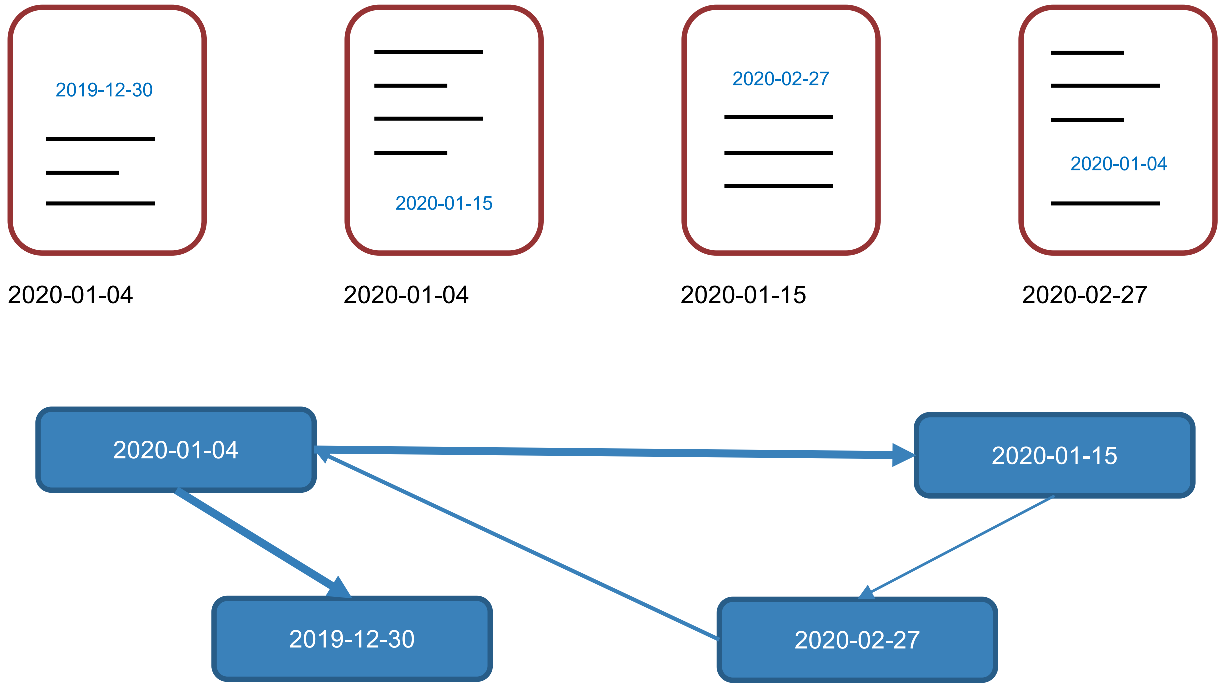
1. Retrieve topic-related corpus
2. Input data are tagged with temporal references



Strötgen, J., & Gertz, M. (2015, September). *A baseline temporal tagger for all languages*. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 541-547).

TimeLine summarization

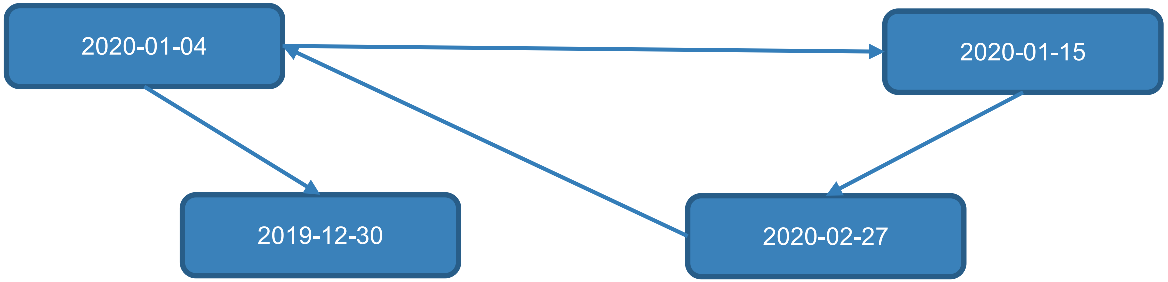
3. Graph-based date reference model



Strötgen, J., & Gertz, M. (2015, September). *A baseline temporal tagger for all languages*. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 541-547).

TimeLine summarization

3. Graph ranking

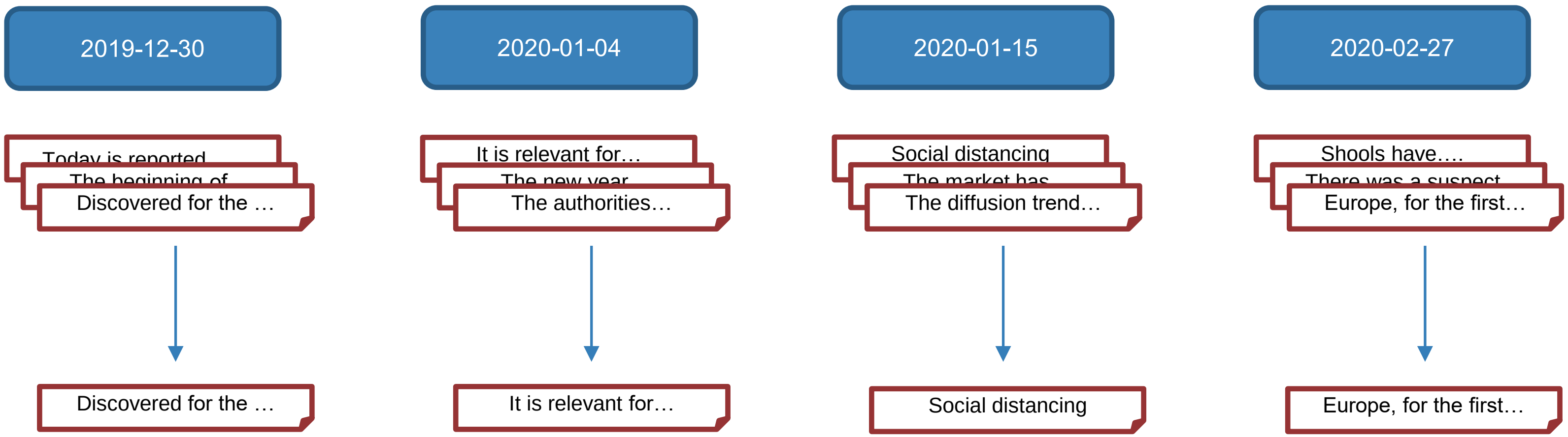


Node centrality reflects the date-level importance in the tracked event

Ghalandari, D. G., & Ifrim, G. (2020, July). Examining the State-of-the-Art in News Timeline Summarization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 1322-1334).

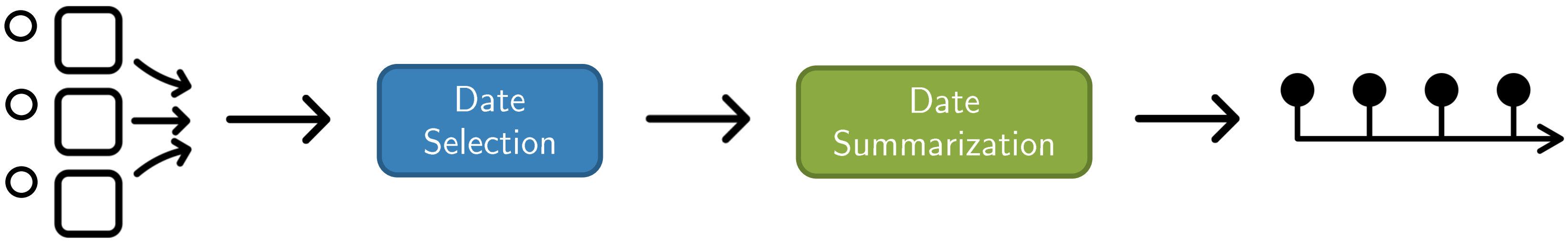
TimeLine Summarization pipeline

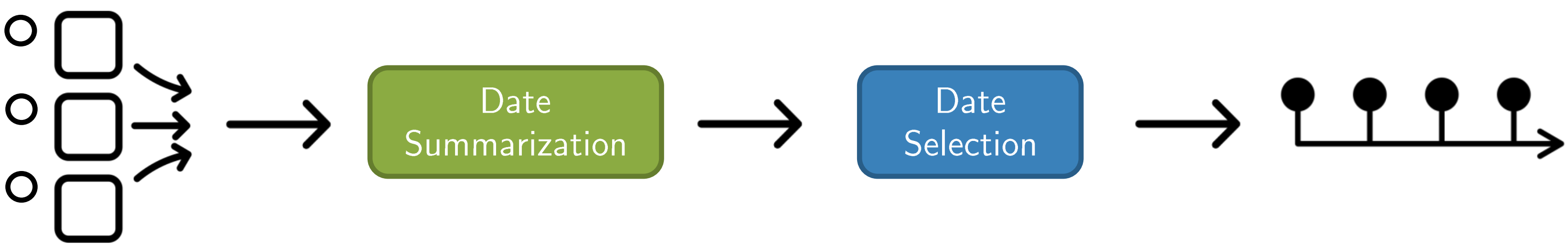
4. Date Summarization



a) Date-specific news content also includes the referencing text.
b) A selection of date-specific content is extracted separately for each selected date

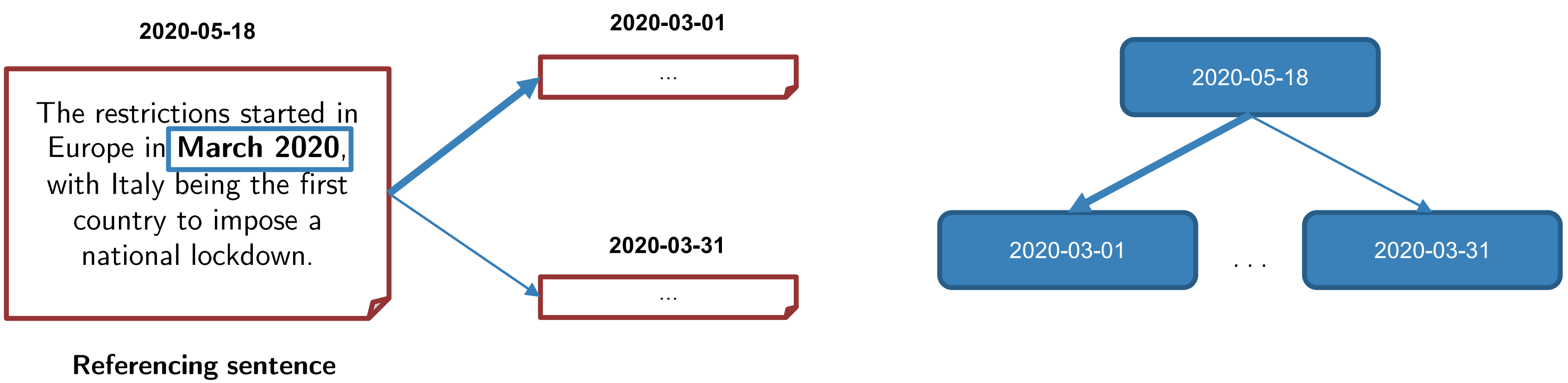
Summarize Dates First





Moreno La Quatra, Luca Cagliero, Elena Baralis, Alberto Messina, Maurizio Montagnuolo. Summarize Dates First: A Paradigm Shift in Timeline Summarization. SIGIR 2021: 418-427

Summarize Dates First: high-level date reference



Moreno La Quatra, Luca Cagliero, Elena Baralis, Alberto Messina, Maurizio Montagnuolo. Summarize Dates First: A Paradigm Shift in Timeline Summarization. SIGIR 2021: 418-427

Summarize Dates First: Incremental timeline updating

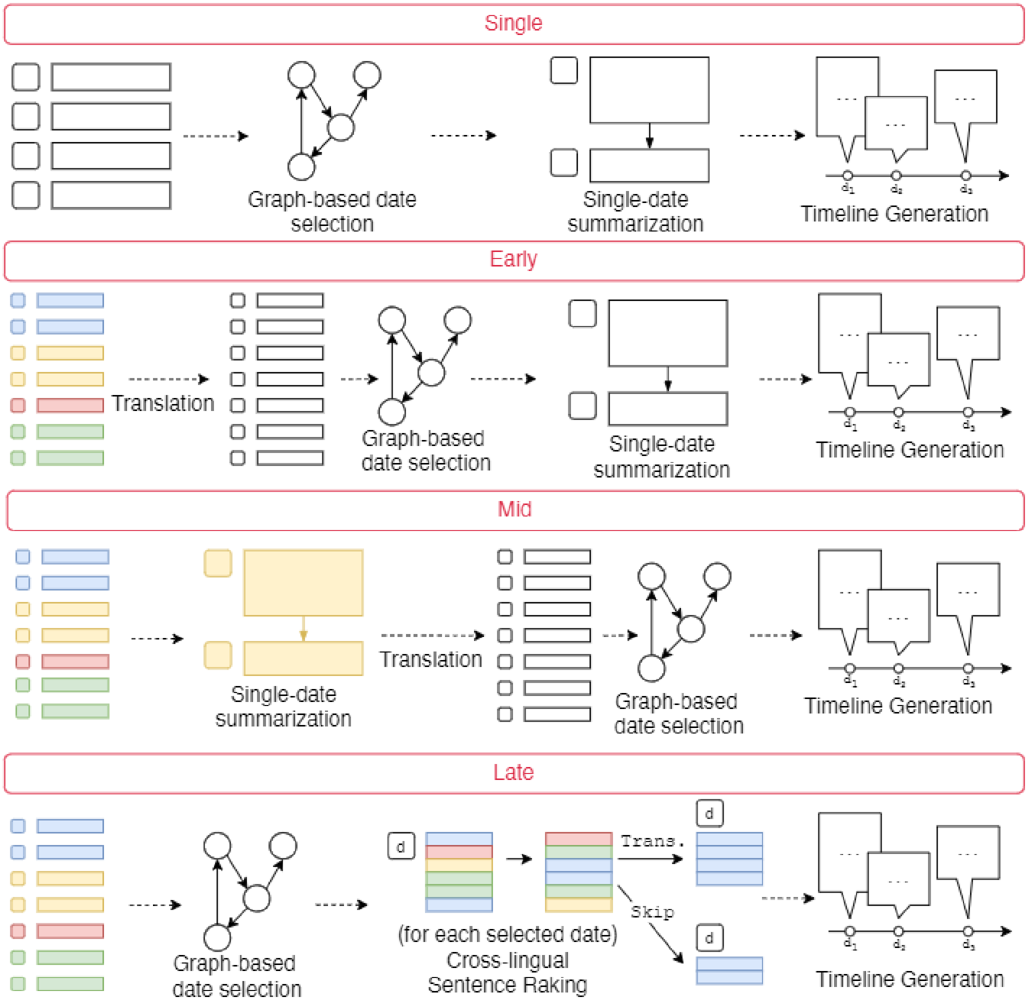
- Traditional pipelines do not allow incremental updates.
- Summarization is the most computationally expensive step.

In SDF TLS:

- Frequent updates are more efficient.
- Summaries updates only for a small subset of dates.

Moreno La Quatra, Luca Cagliero, Elena Baralis, Alberto Messina, Maurizio Montagnuolo. Summarize Dates First: A Paradigm Shift in Timeline Summarization. SIGIR 2021: 418-427

Towards Cross-Lingual TimeLine Summarization



Cross-lingual timeline summarization. Luca Cagliero, Moreno La Quatra, Paolo Garza and Elena Baralis. IEEE AIKE 2021

Acknowledgements and copyright license

- Acknowledgements
  - I would like to thank Dr. Moreno La Quatra, who collaborated to the writing and revision of the teaching content

- Affiliation
  - The author and his staff are currently members of the Database and Data Mining Group at Dipartimento di Automatica e Informatica (Politecnico di Torino) and of the SmartData interdepartmental centre
    - https://dbdmg.polito.it
    - https://smartdata.polito.it

Thank you!

