Automatic Summarization

Lecture Information Retrieval, DS

8th of March, 2019

Maartje ter Hoeve m.a.terhoeve@uva.nl @maartjeterhoeve

Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today's focus: Single Document Summarization (text & multimodal)

Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today's focus: Single Document Summarization (text & multimodal)

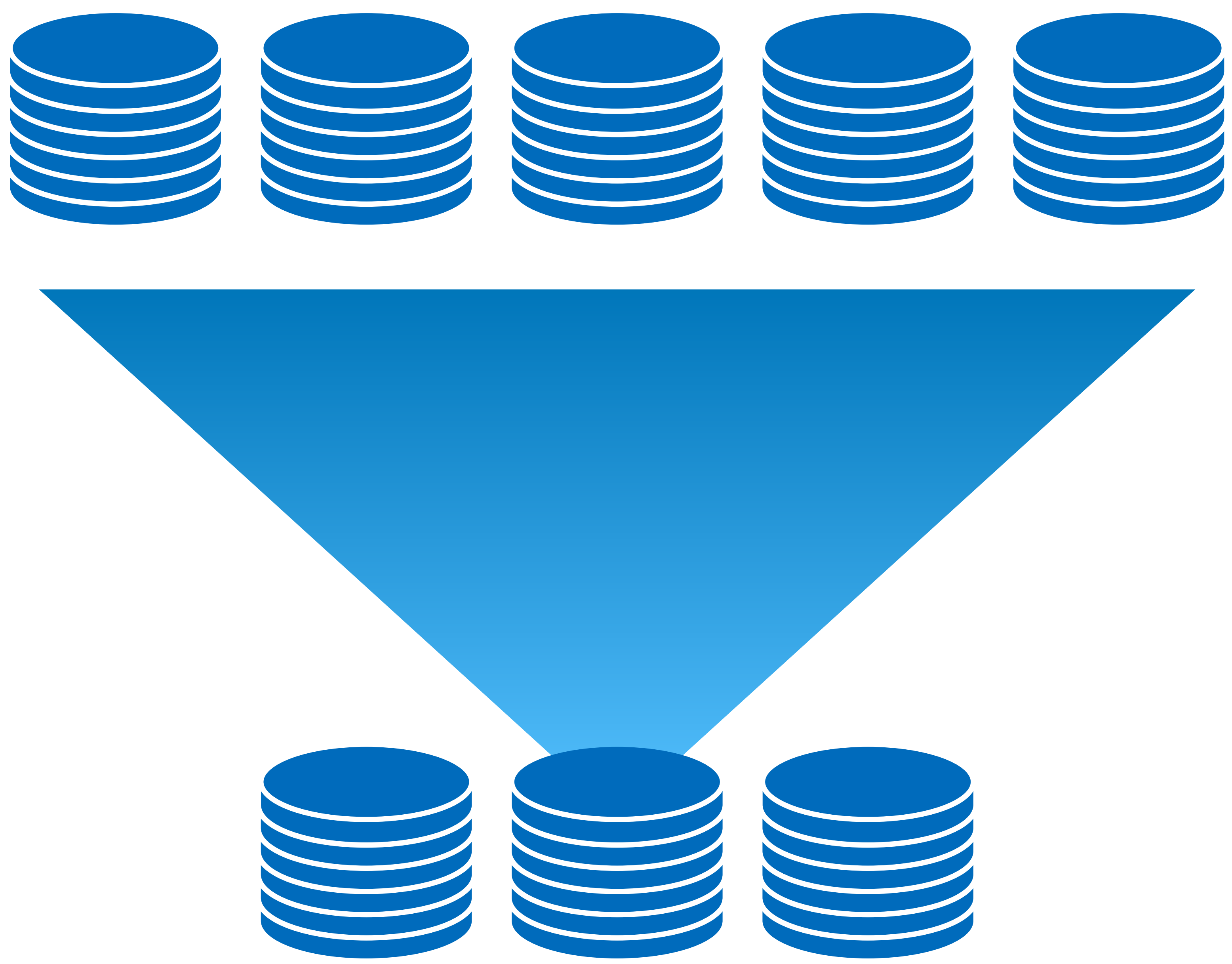
Disclaimer! The field is way too big to cover everything in one lecture, so we will just touch upon a selection of different work!

Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today's focus: Single Document Summarization (text & multimodal)

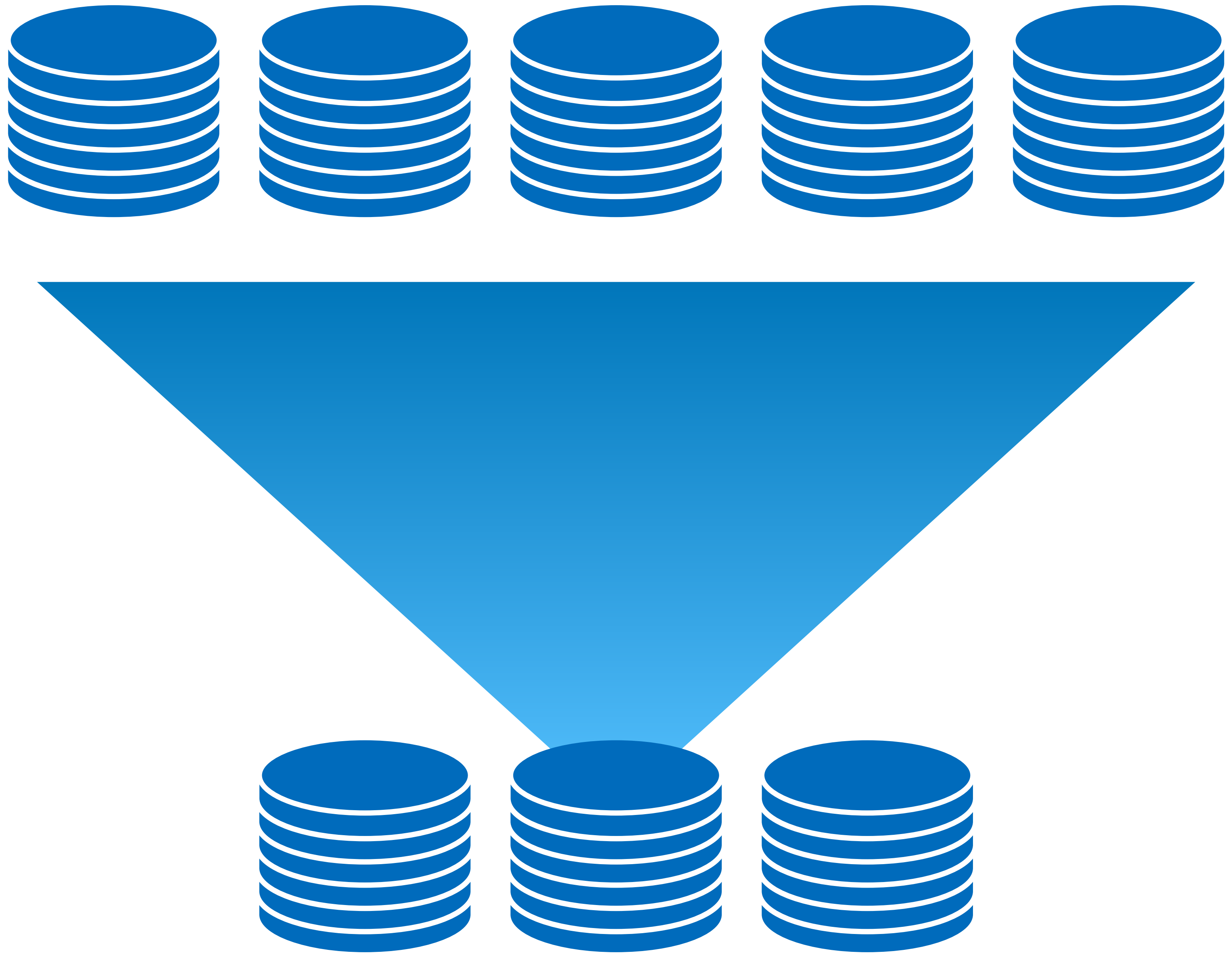
What is summarization?

What is summarization?



Task: Find the important bits in all the data and return these

What is summarization?



Task: Find the important bits in all the data and return these

Why would you be interested in this?

Why would you be interested in this?

Academically

Can we learn a smaller representation that still captures our input?

Practical standpoint

Countless examples where amount of information available is too much to manually digest.

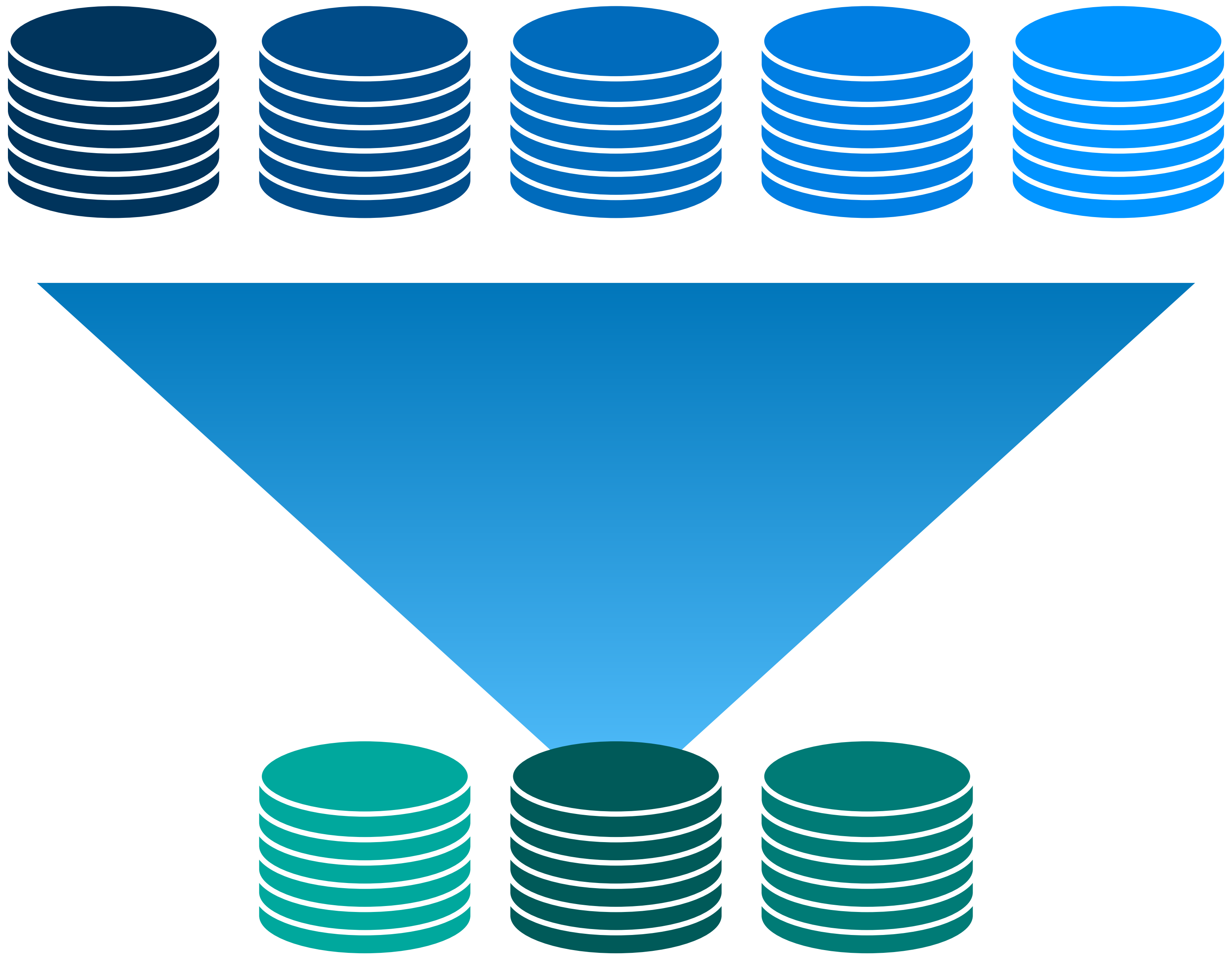
Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today's focus: Single Document Summarization (text & multimodal)

What can summaries look like?

Extractive summarization

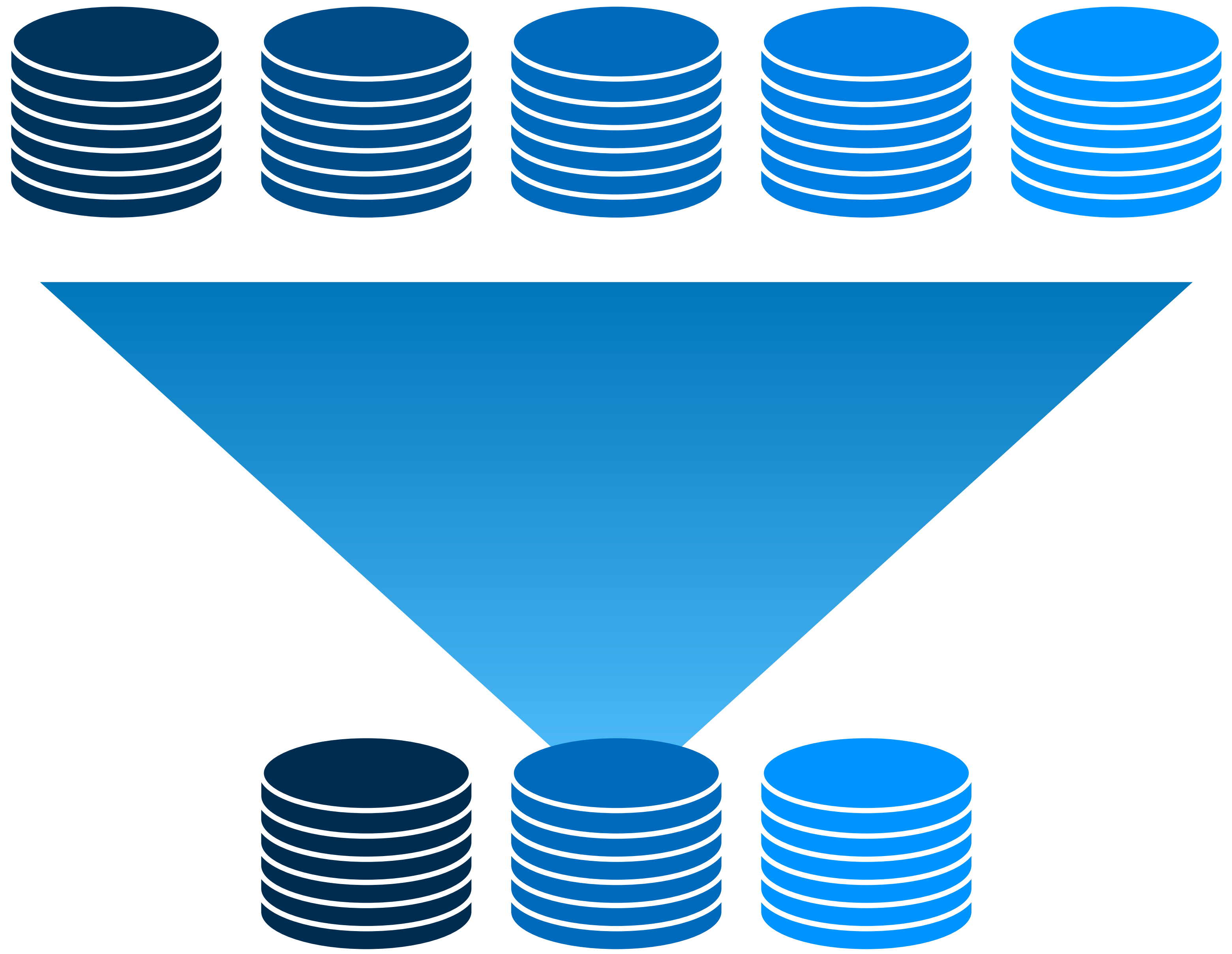
Abstractive summarization

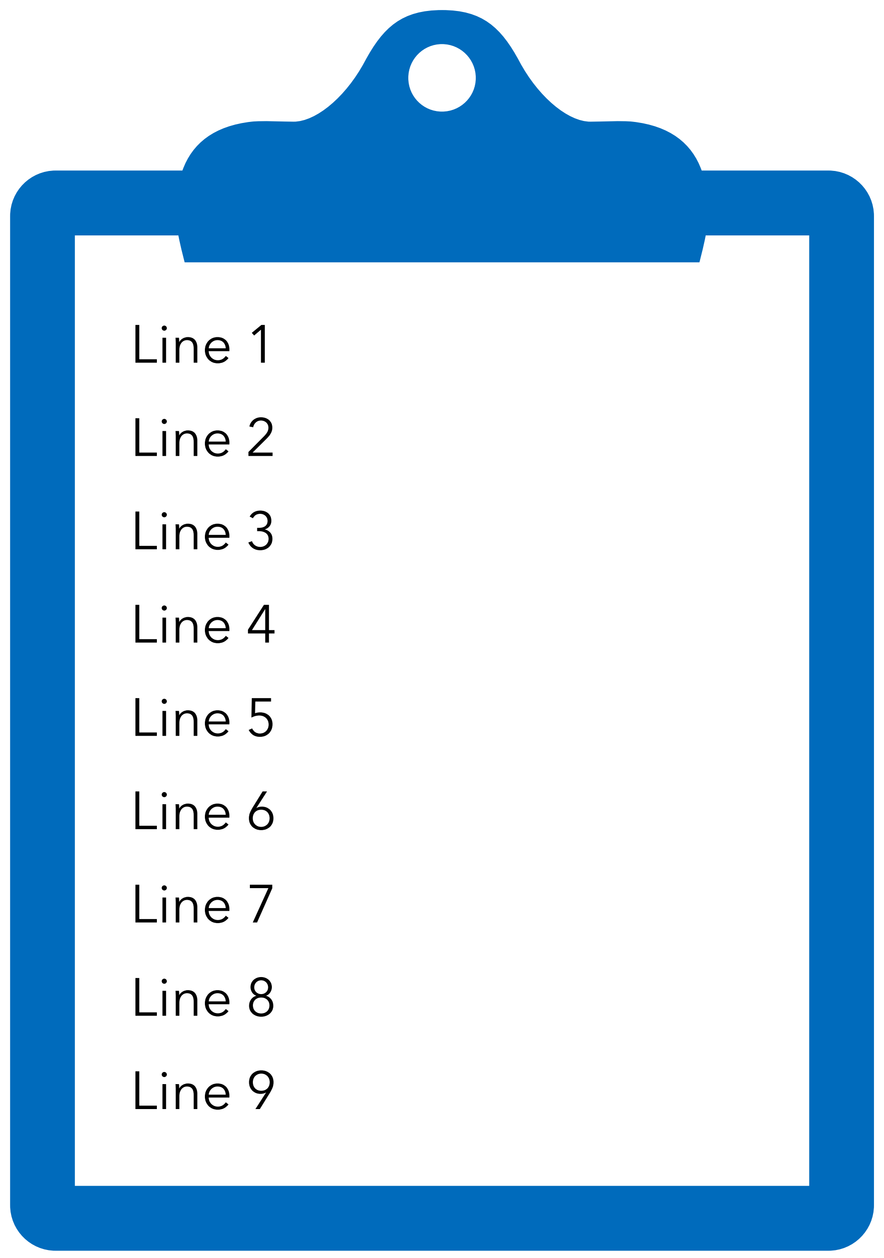


UNIVERSITY OF AMSTERDAM
Maartje ter Hoeve
m.a.terhoeve@uva.nl
@maartjeterhoeve

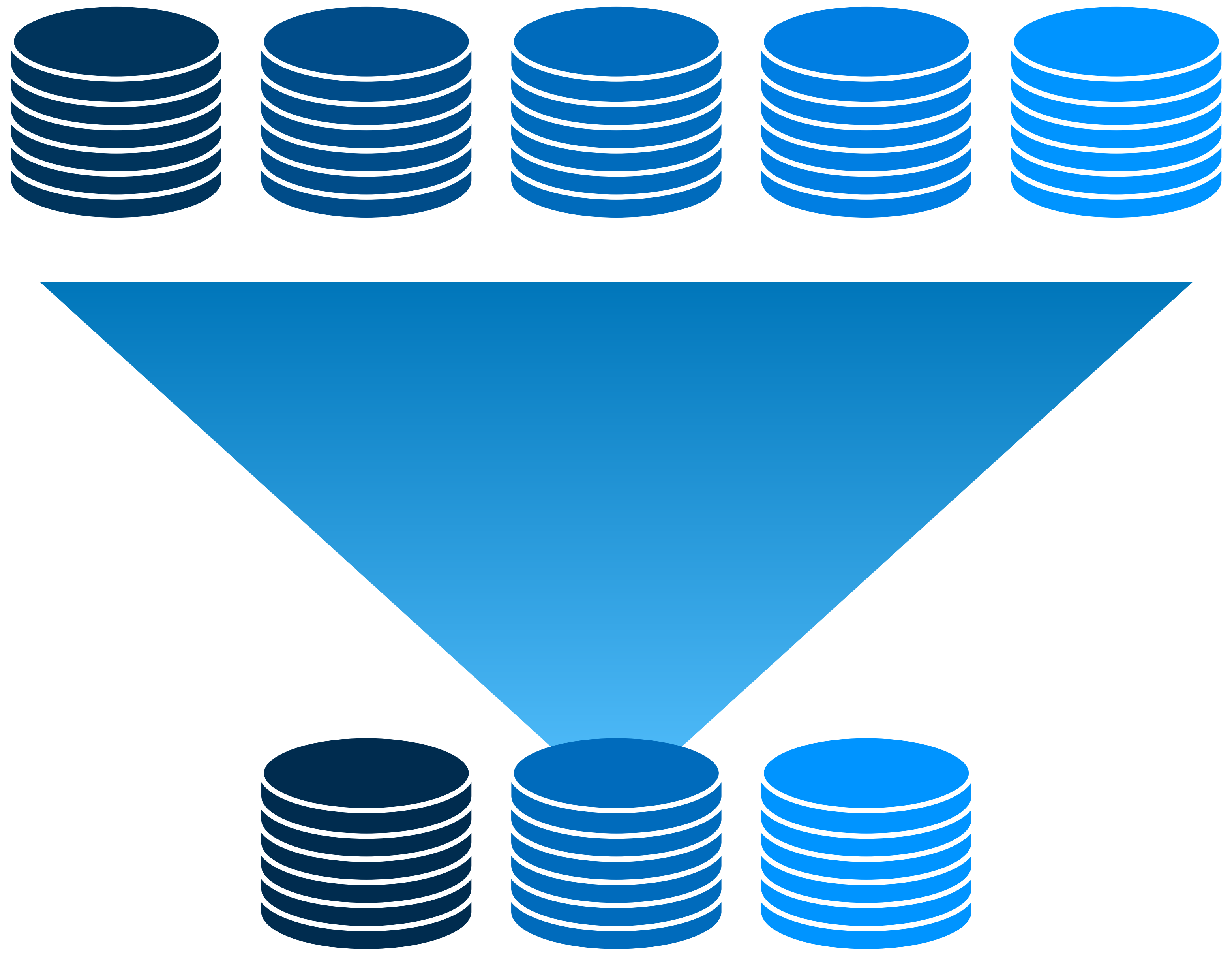
Extractive Summarization

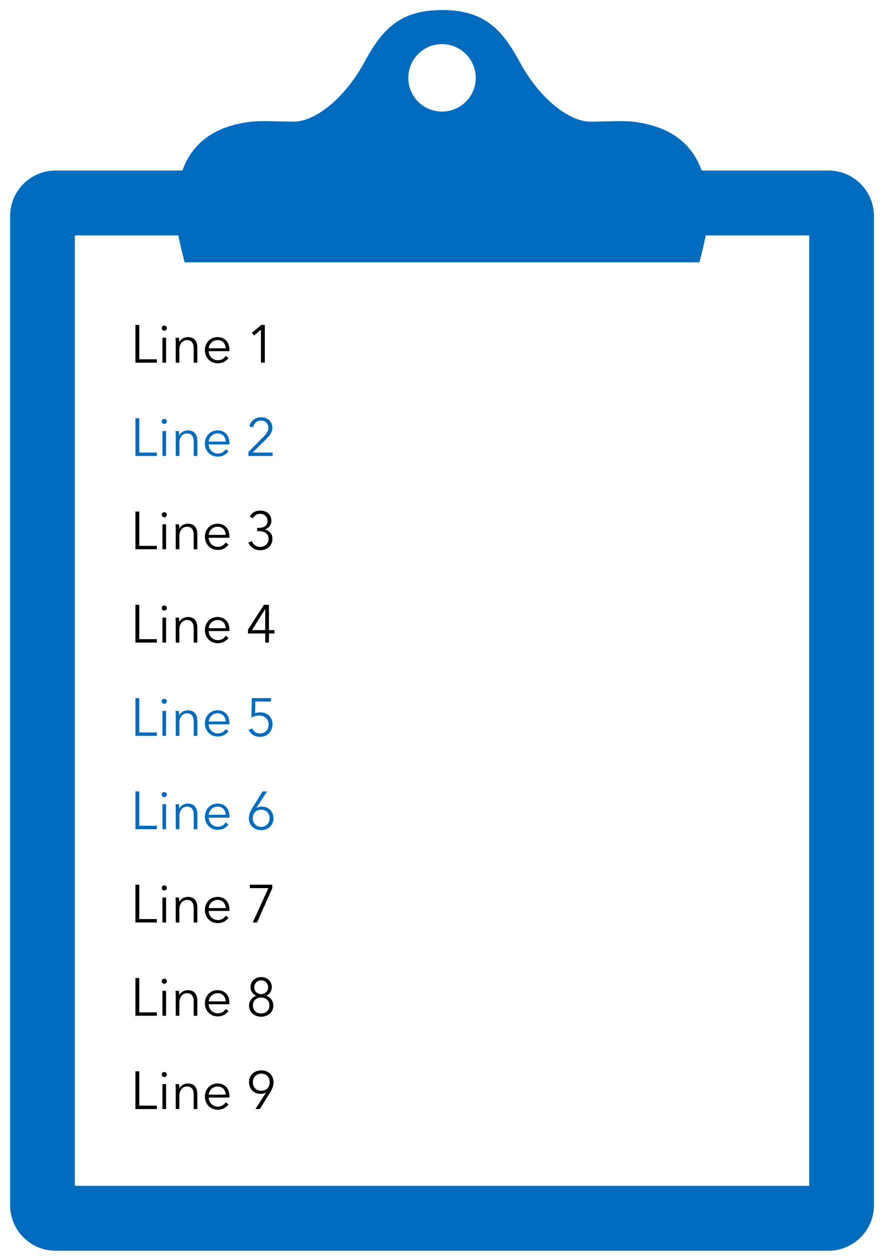
Extractive Summarization



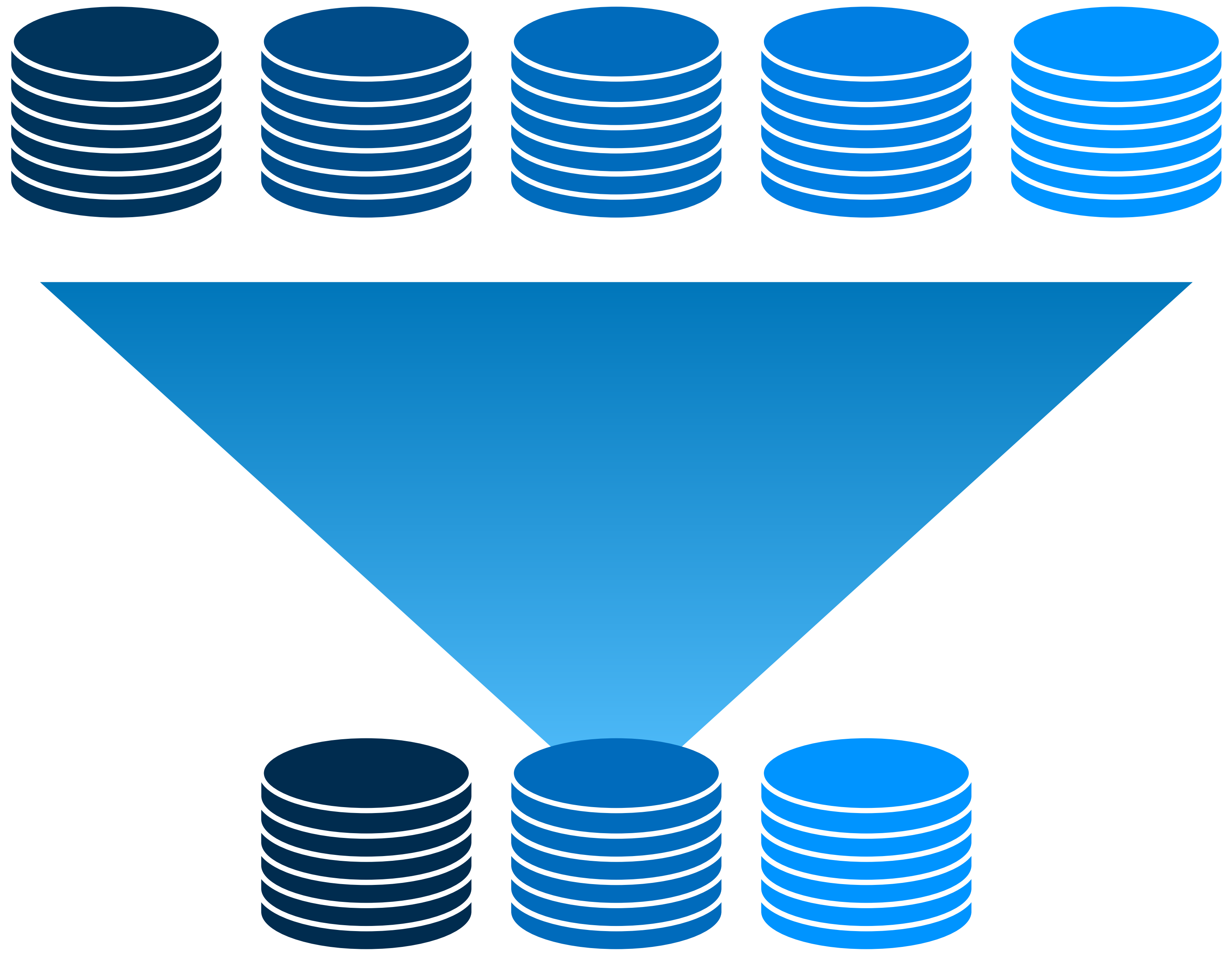


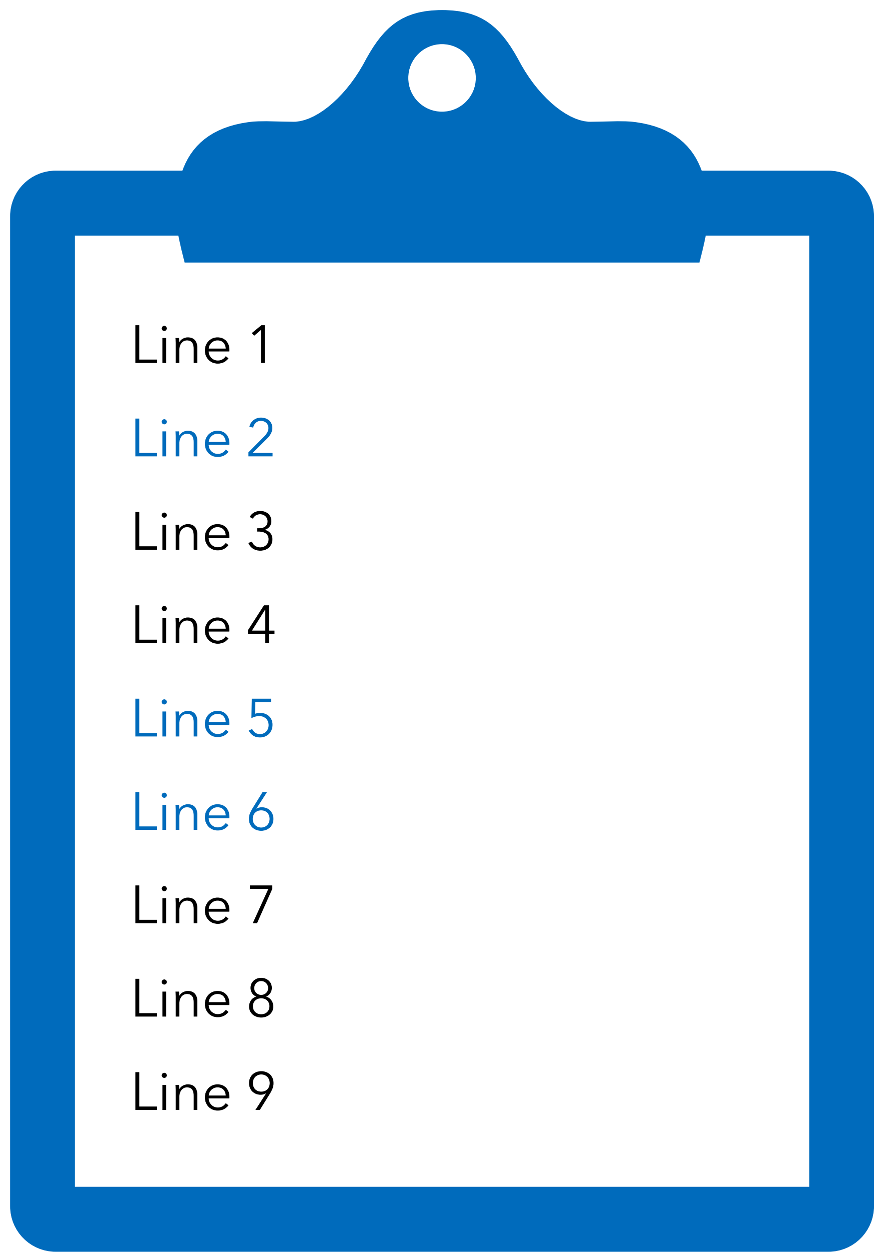
Extractive Summarization

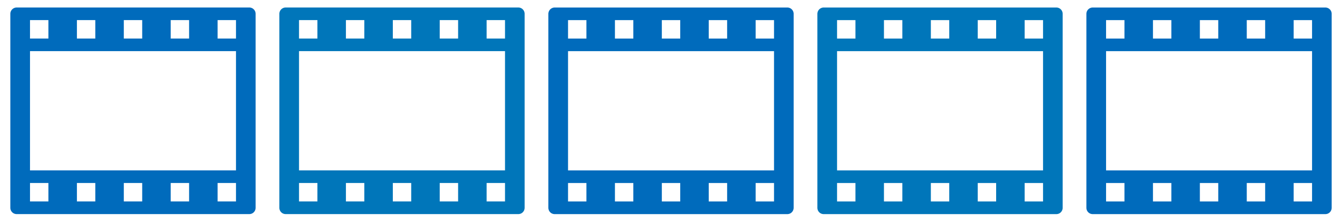




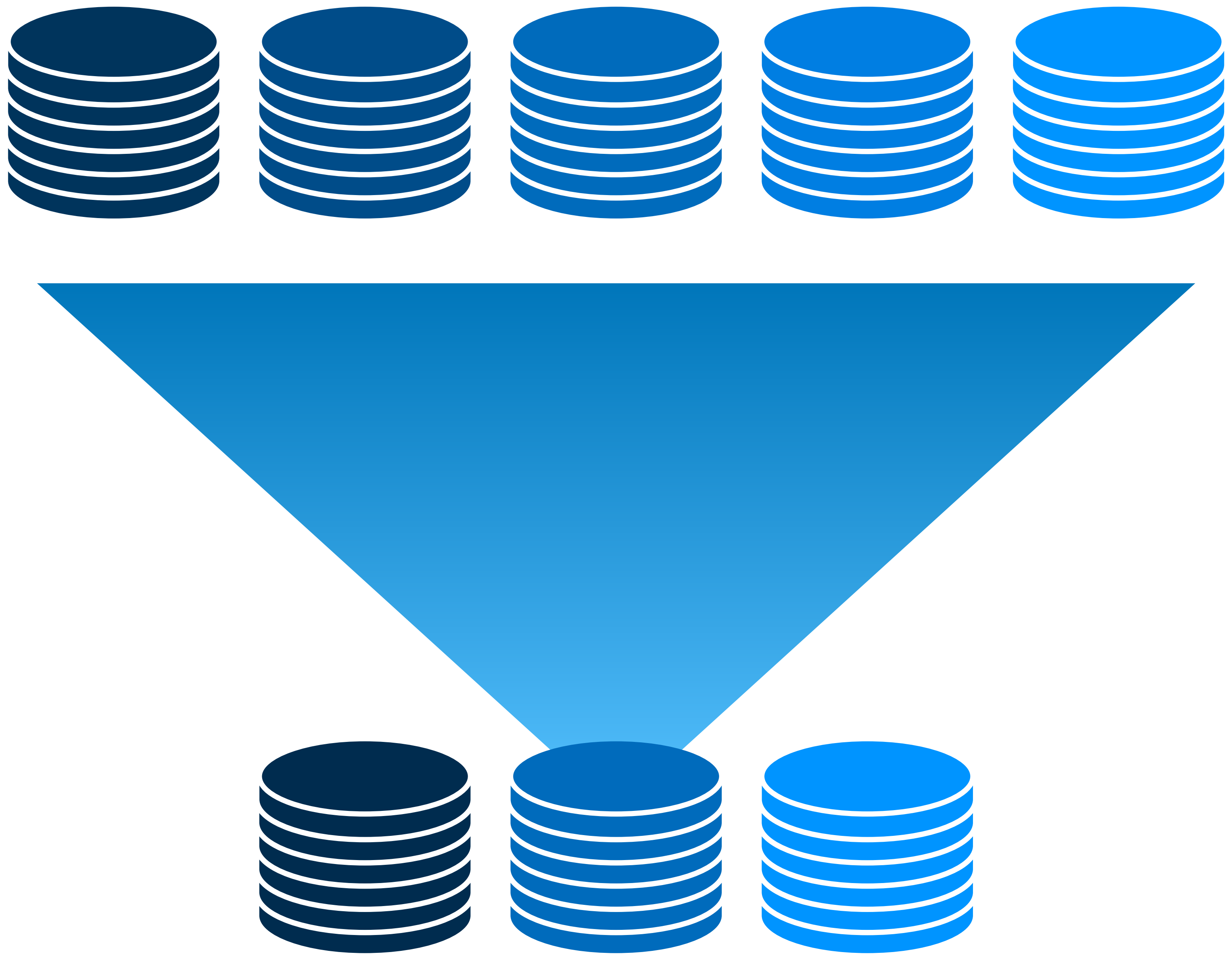
Extractive Summarization

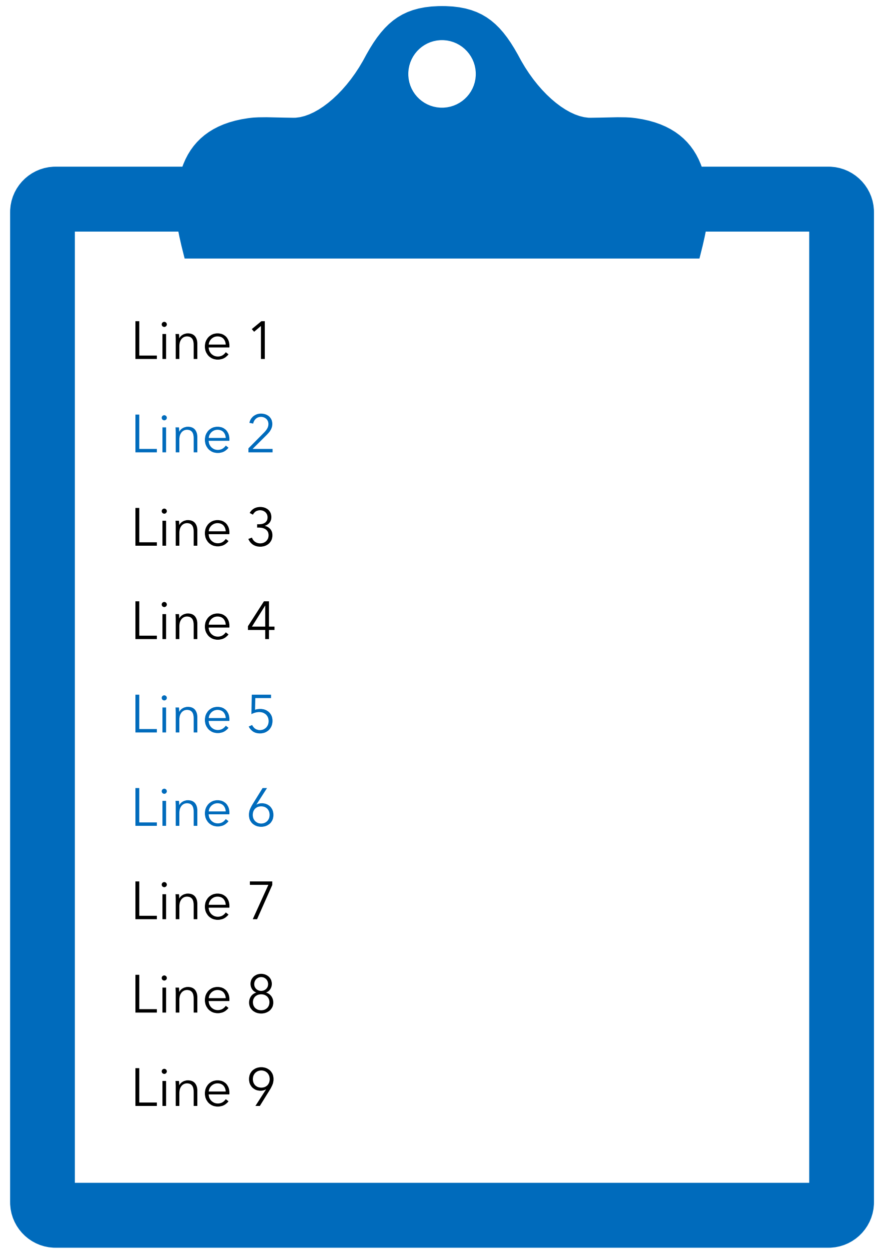


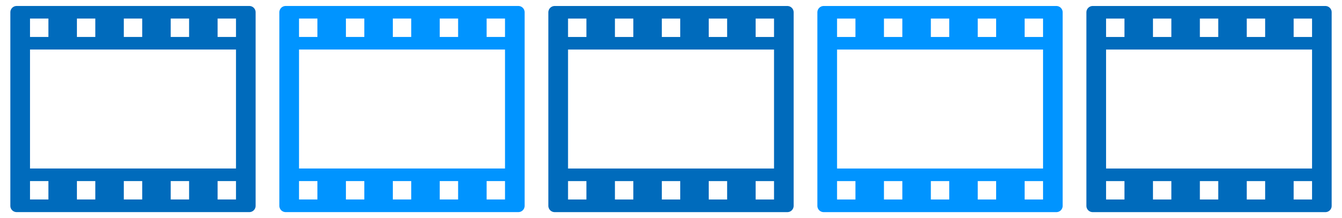




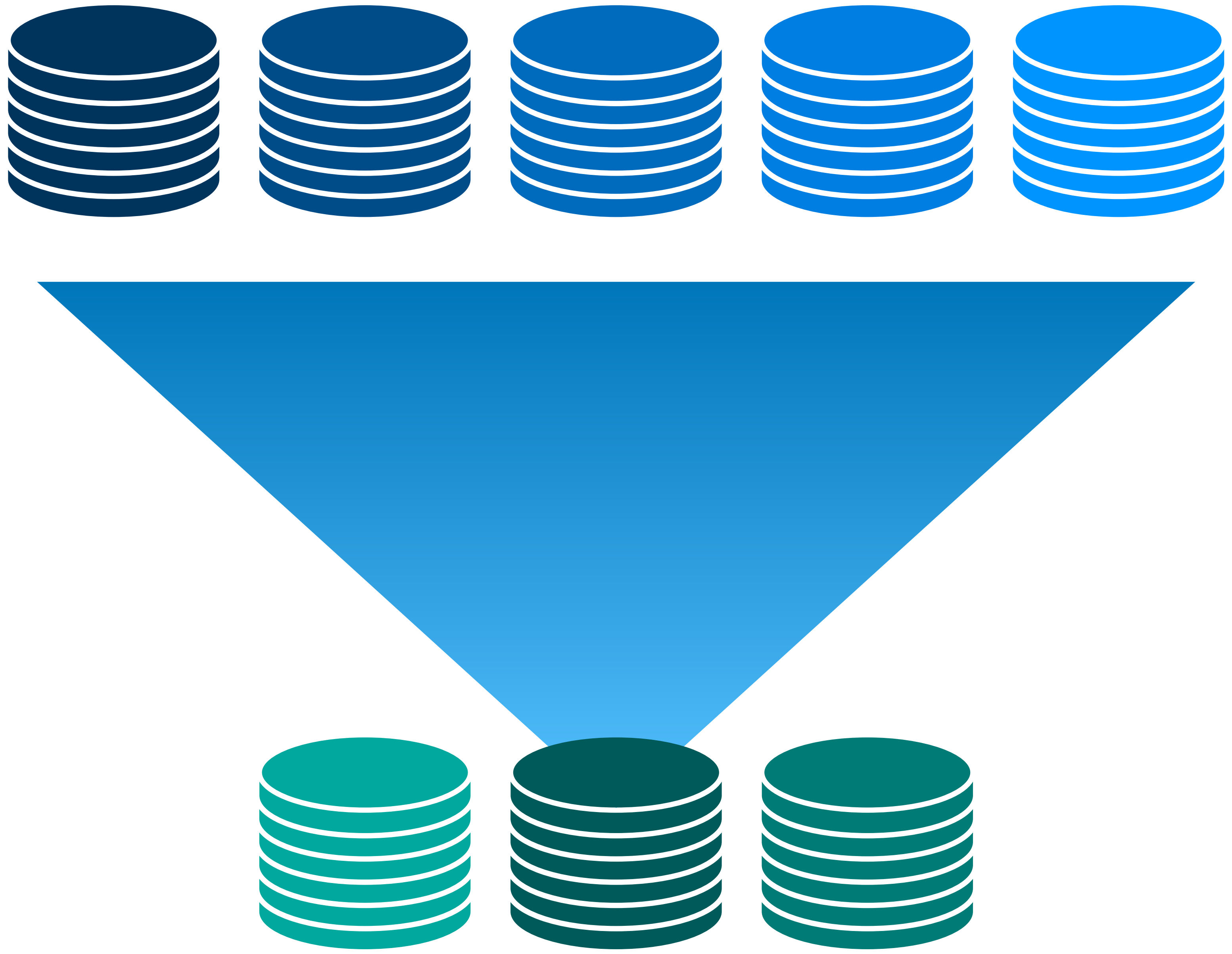
Extractive Summarization

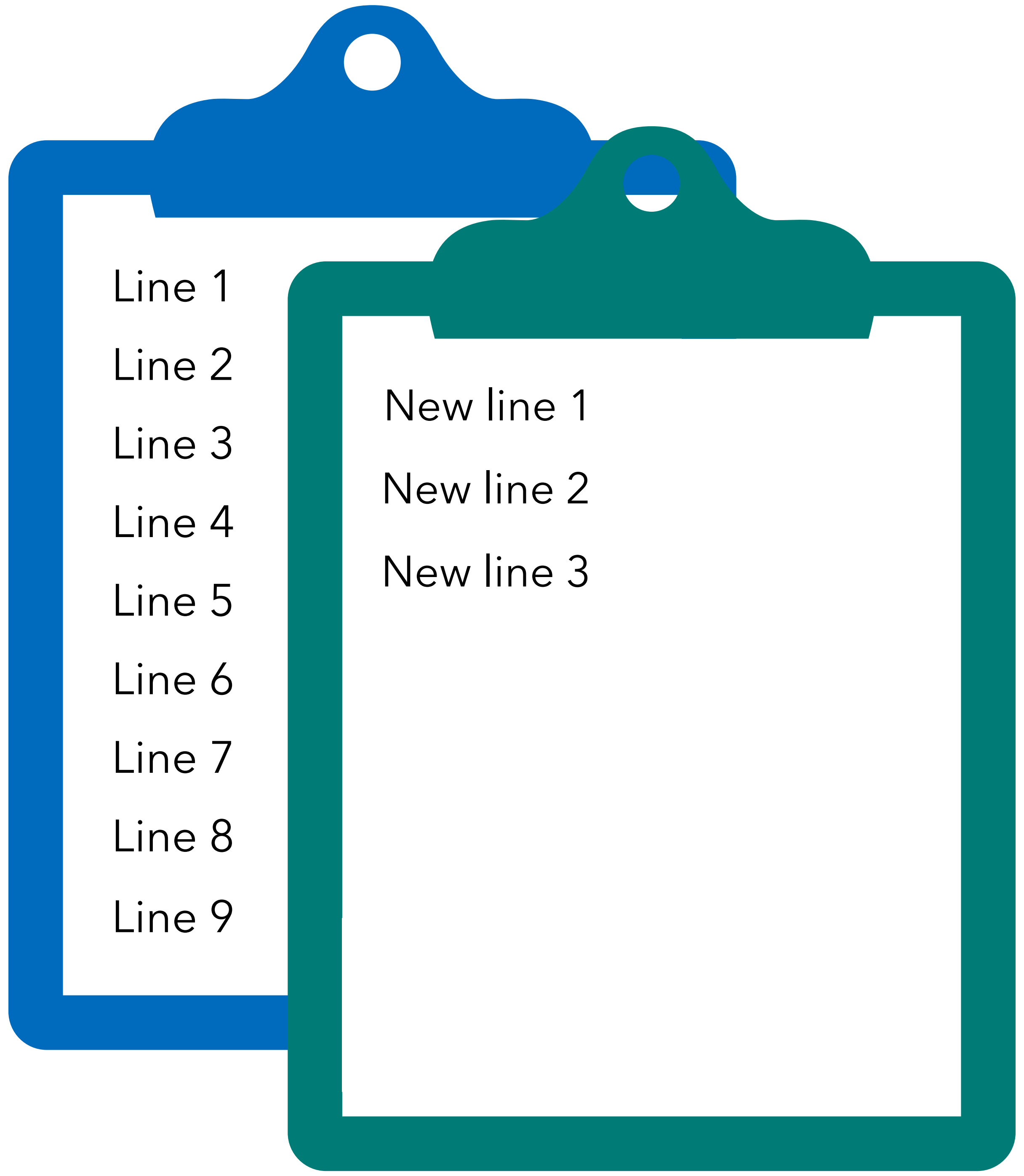


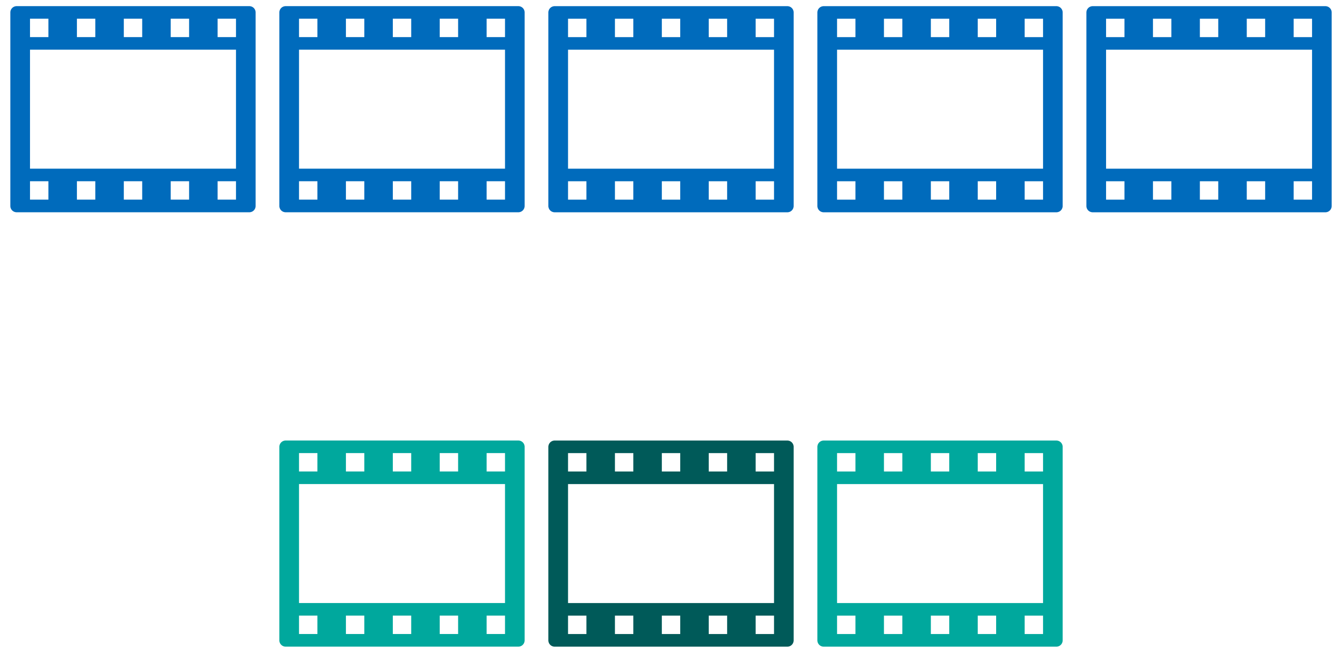




Abstractive Summarization







Focus of this lecture

Single Document Summarization

- Text
- Multimodal

Maartje ter Hoeve m.a.terhoeve@uva.nl @maartjeterhoeve

Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today's focus: Single Document Summarization (text & multimodal)

Evaluation - what makes a summary 'good'?

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

ROUGE: Recall-Oriented Understudy for Gisting Evaluation

ROUGE is a metric to evaluate textual summaries.

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

Predicted summary

ROUGE computes the quality of a summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

Reference summary

ROUGE is a metric to compute the quality of a textual summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

Predicted summary

ROUGE computes the quality of a summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

Reference summary

ROUGE is a metric to compute the quality of a textual summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

Predicted summary

ROUGE computes the quality of a summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

Unigrams: 22

Bigrams: 21

Reference summary

ROUGE is a metric to compute the quality of a textual summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

Unigrams: 27

Bigrams: 26

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

Predicted summary

ROUGE computes the quality of a summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

Reference summary

ROUGE is a metric to compute the quality of a textual summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

Overlapping Unigrams: 21

Overlapping Bigrams: 18

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

$$ROUGE-N_{Recall} = \frac{\textbf{\# overlapping n-grams}}{\textbf{\# n-grams in reference summary}}$$

$$ROUGE-N_{Precision} = \frac{\textbf{\# overlapping n-grams}}{\textbf{\# n-grams in predicted summary}}$$

$$ROUGE-N_F = \frac{(1+\beta^2) \times ROUGE-N_{Recall} \times ROUGE-N_{Precision}}{ROUGE-N_{Recall} + \beta^2 \times ROUGE-N_{Precision}}$$

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

$$ROUGE-1_{Recall} = \frac{\textbf{\# overlapping unigrams}}{\textbf{\# unigrams in reference summary}} = \frac{21}{27} = 0.78$$

$$ROUGE-1_{Precision} = \frac{\textbf{\# overlapping unigrams}}{\textbf{\# unigrams in predicted summary}} = \frac{21}{22} = 0.95$$

$$ROUGE-1_{F} = \frac{(1+\beta^2) \times ROUGE-1_{Recall} \times ROUGE-1_{Precision}}{ROUGE-1_{Recall} + \beta^2 \times ROUGE-1_{Precision}} = 2 \times \frac{0.78 \times 0.95}{0.78 + 0.95} = 0.86$$

We use $\beta = 1$

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

$$ROUGE - 1_{Recall} = \frac{\textbf{\# overlapping unigrams}}{\textbf{\# unigrams in reference summary}} = \frac{21}{27} = 0.78$$

$$ROUGE - 1_{Precision} = \frac{\textbf{\# overlapping unigrams}}{\textbf{\# unigrams in predicted summary}} = \frac{21}{22} = 0.95$$

$$ROUGE - 1_{F} = \frac{(1+\beta^2) \times ROUGE - 1_{Recall} \times ROUGE - 1_{Precision}}{ROUGE - 1_{Recall} + \beta^2 \times ROUGE - 1_{Precision}} = 2 \times \frac{0.78 \times 0.95}{0.78 + 0.95} = 0.86$$

We use $\beta = 1$

Same computation for Rouge-2,

but use the bigram counts

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

ROUGE-L

Do the same, but for longest common subsequence.

(Use the union of LCS for multiple sentences.)

Multiple references

Use argmax of all ROUGE-N scores.

ROUGE

Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, edited by Stan Szpakowicz Marie-Francine Moens, 74–81. Barcelona, Spain: Association for Computational Linguistics, 2004.

Correlation with Human Judgement

Okay?

People are not totally convinced and often perform a human evaluation as well.

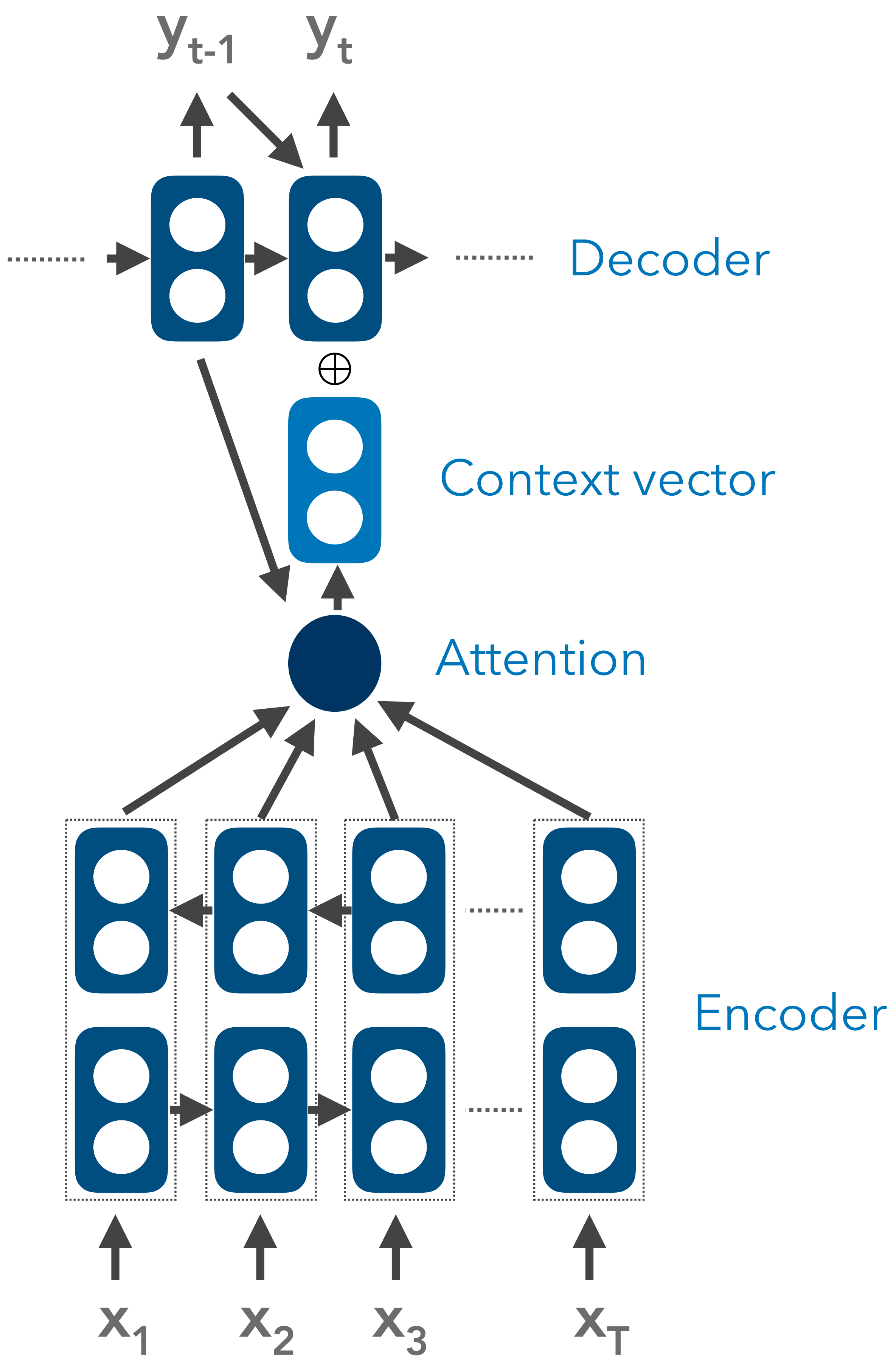
Content

Single Document Summarization - Text

Extractive summarization

- Historical methods
- Binary labeling strategy
- Reinforcement learning approach
- …

Abstractive summarization



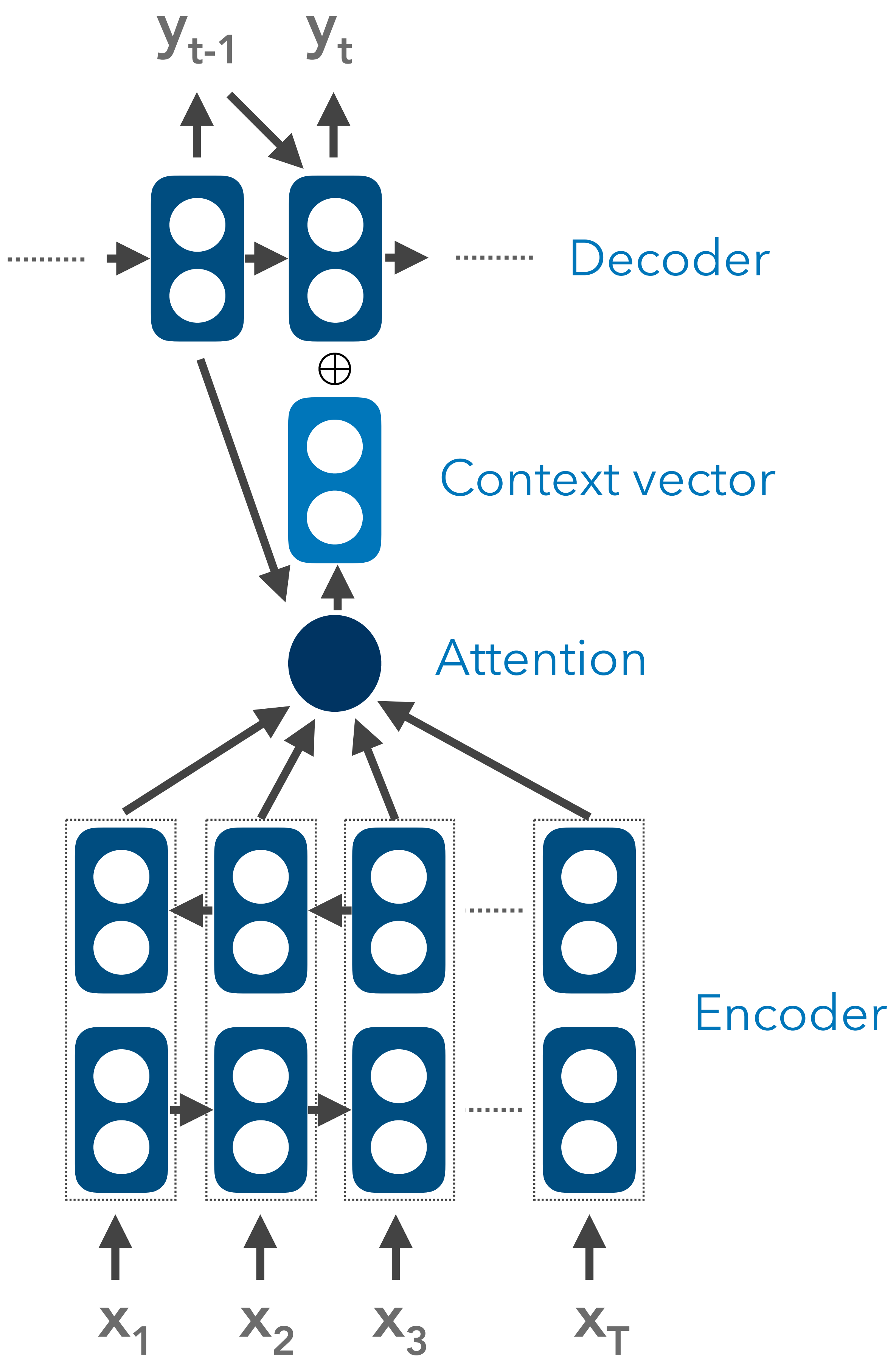
Single Document Summarization - Text

Extractive summarization

- Historical methods
- Binary labeling str y
- Rei g approach
- …

Disclaimer! The field is way too big to cover everything in one lecture, so we will just touch upon a selection of different work!

Abstractive summarization



Extractive Single Document Summarization - Text (1)

Erkan, Gunes, and Dragomir R. Radev. "LexRank: Graph-Based Lexical Centrality as Salience in Text Summarization." *Journal of Artificial Intelligence Research* 22 (December 1, 2004): 457–79. https://doi.org/10.1613/jair.1523.

Historical methods - **LexRank** as example

Make a graph from the document. Each sentence is a node.

Compute similarity between sentences. These scores are the edges.

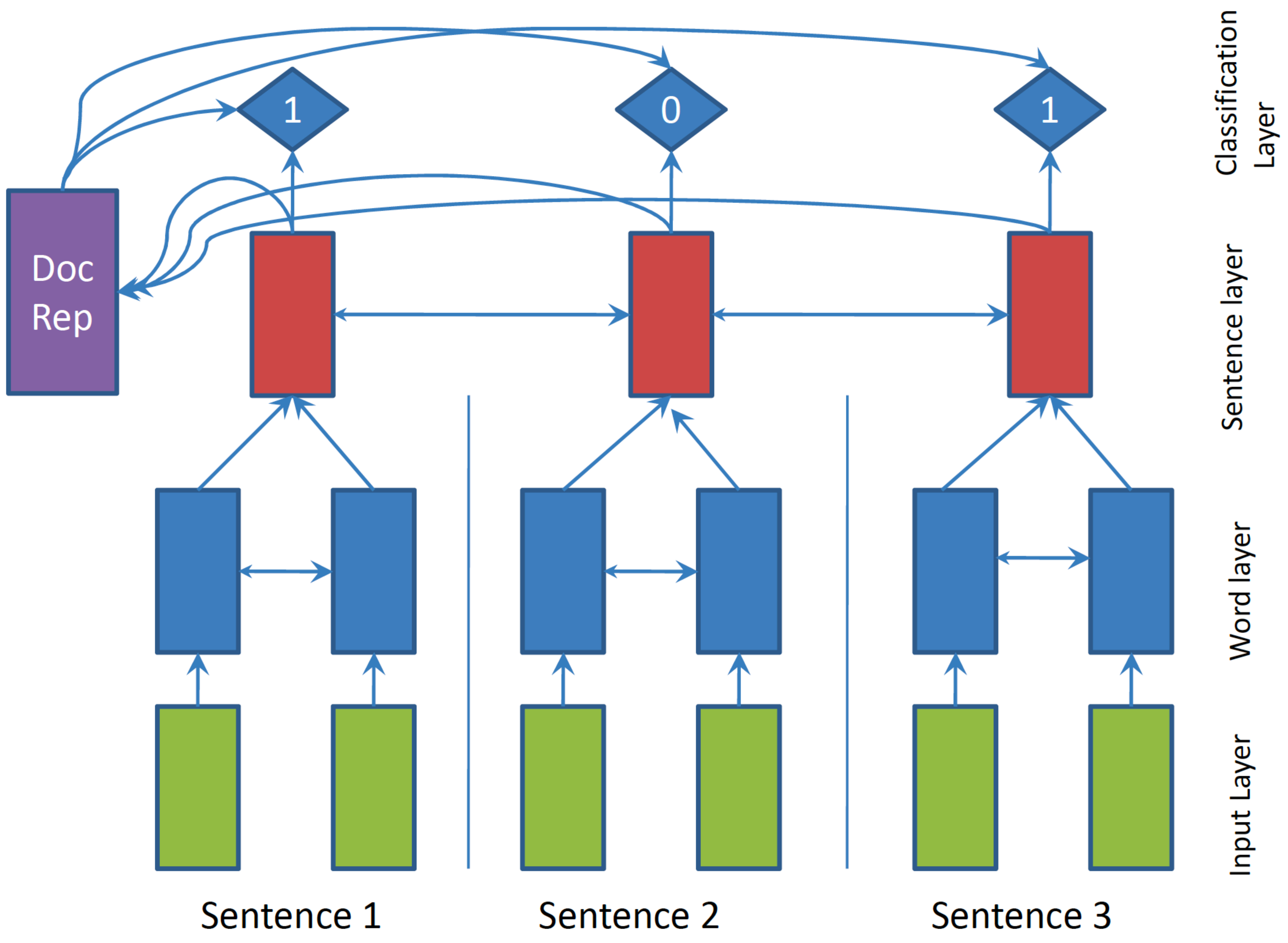
Use PageRank to rank these sentences.

Select the highest ranking sentences as the summary.

Unsupervised!

Extractive Single Document Summarization - Text (2)

Nallapati, Ramesh, Feifei Zhai, and Bowen Zhou. "SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents." *ArXiv:1611.04230 [Cs]*, November 13, 2016. http://arxiv.org/abs/1611.04230.



Binary labeling strategy

Classify each sentence as either 'belongs to summary' or as 'does not belong to summary'.

Extractive Single Document Summarization - Text (3)

Dong, Yue, Yikang Shen, Eric Crawford, Herke van Hoof, and Jackie Chi Kit Cheung. "BanditSum: Extractive Summarization as a Contextual Bandit." In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3739–3748. Brussels, Belgium: Association for Computational Linguistics, 2018. http://www.aclweb.org/anthology/D18-1409.

Reinforcement Learning Strategy

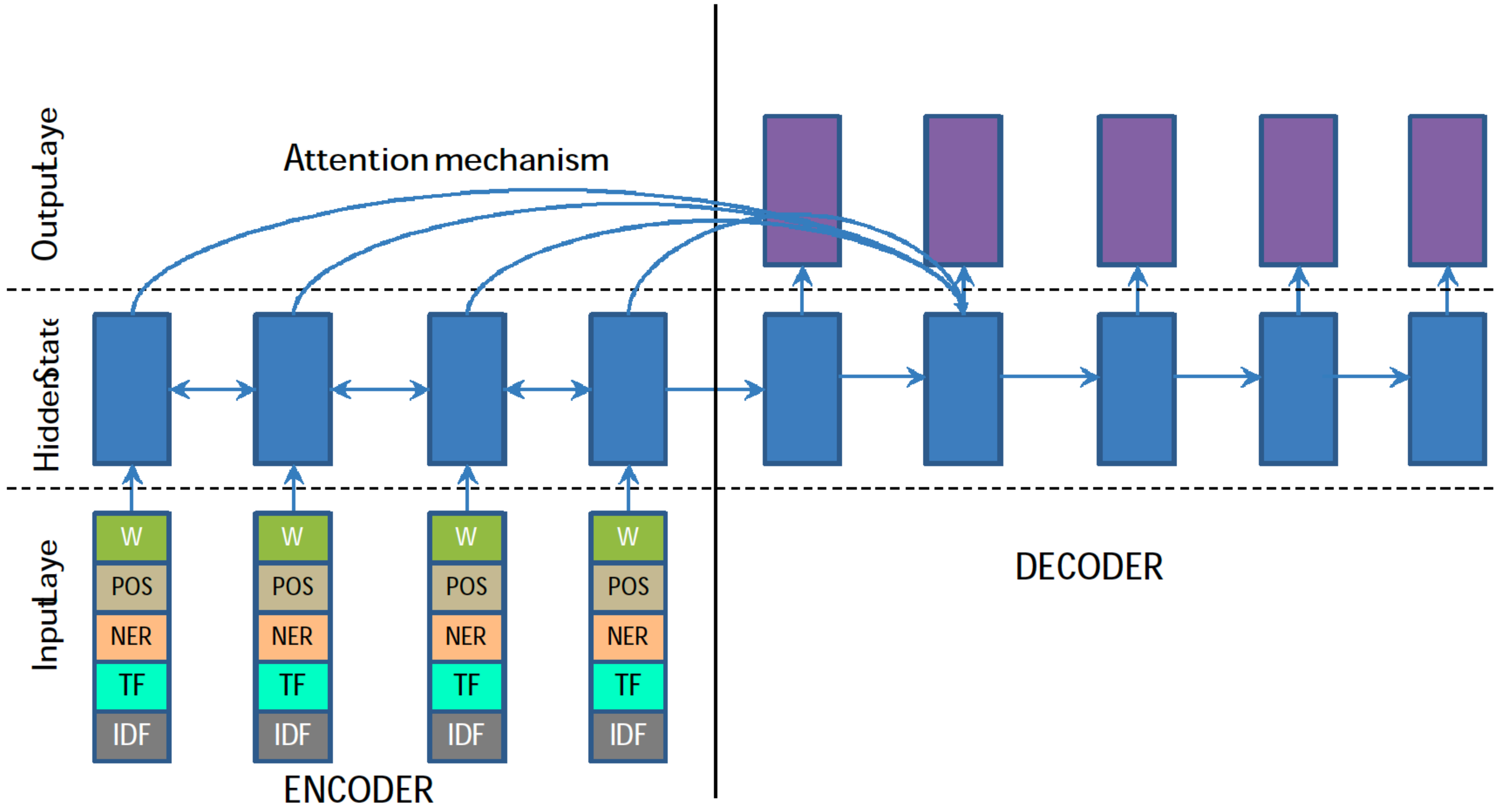
Each document is a context, combinations of sentences in the document are actions.

Learn a policy to perform the best action - get the best summary.

This method optimises directly for ROUGE!

Abstractive Single Document Summarization - Text (1)

Nallapati, Ramesh, Bowen Zhou, Cicero Nogueira dos santos, Caglar Gulcehre, and Bing Xiang. "Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond." *ArXiv:1602.06023 [Cs]*, February 18, 2016. http://arxiv.org/abs/1602.06023.

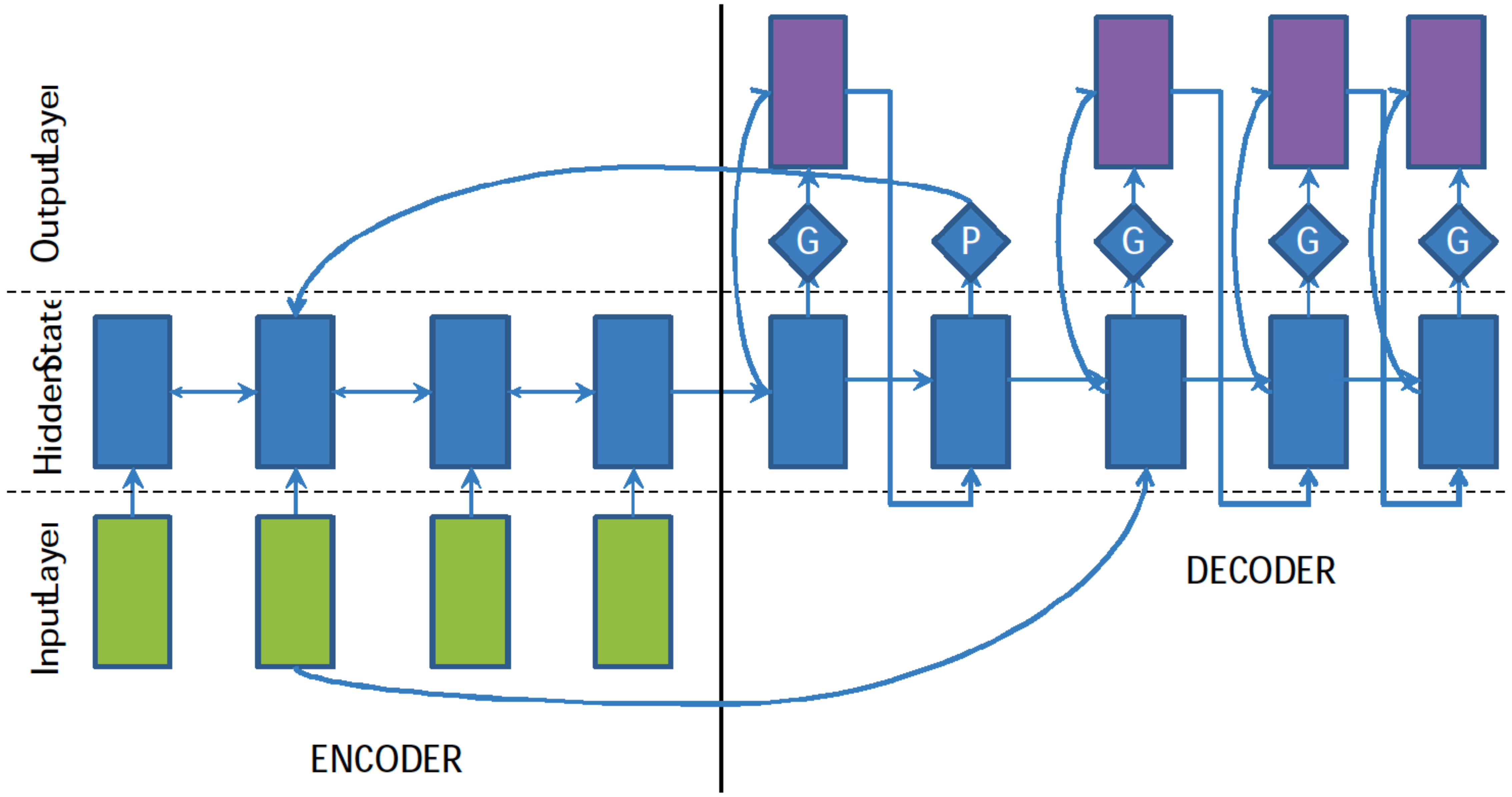


Main idea / contributions

- Introduce encoder / decoder structure for summarisation.
- Add tags to the encoder.

Abstractive Single Document Summarization - Text (1)

Nallapati, Ramesh, Bowen Zhou, Cicero Nogueira dos santos, Caglar Gulcehre, and Bing Xiang. "Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond." *ArXiv:1602.06023 [Cs]*, February 18, 2016. http://arxiv.org/abs/1602.06023.

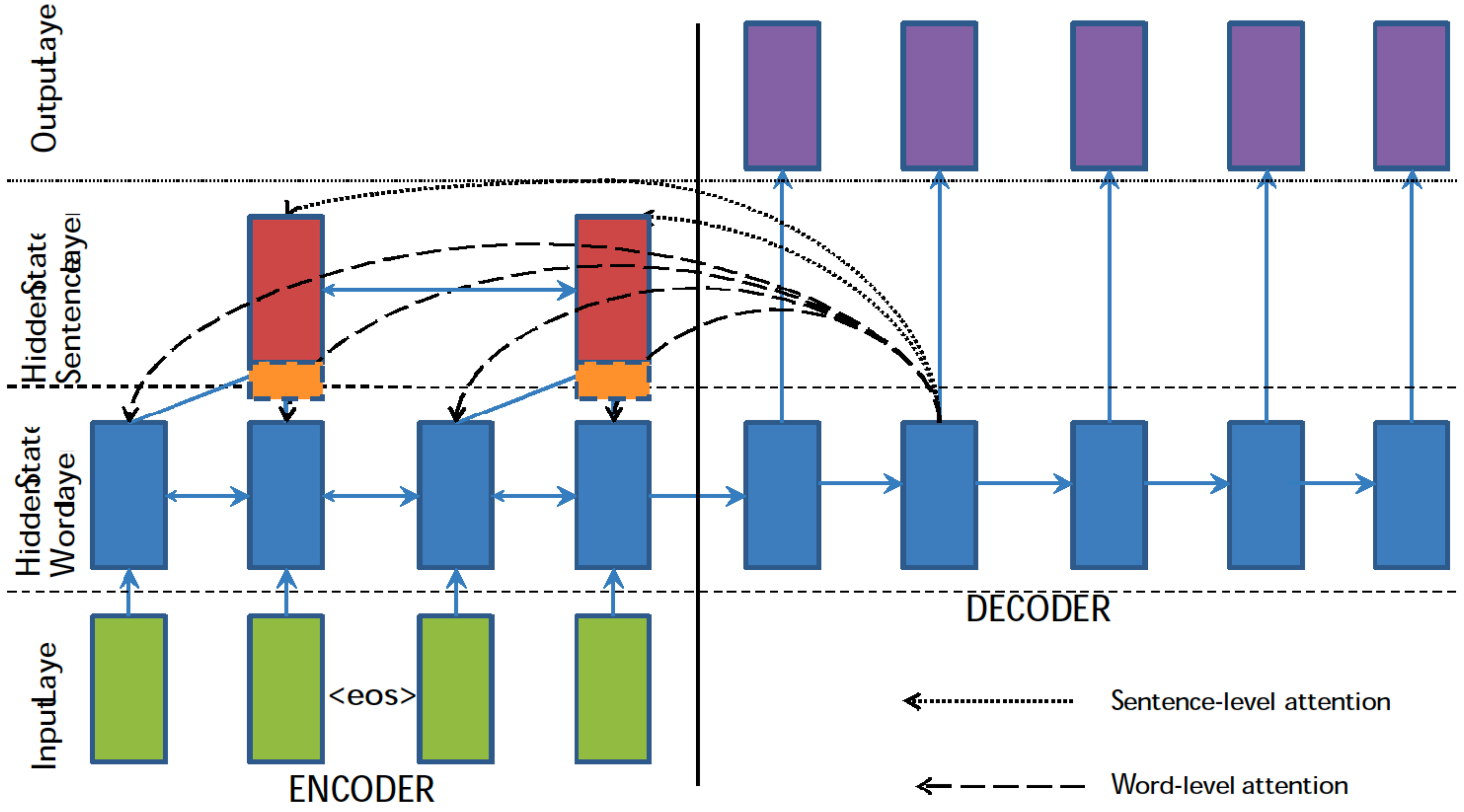


Main idea / contributions

- Introduce encoder / decoder structure for summarisation.
- Add tags to the encoder.
- Introduce pointing probability, for unknown words.

Abstractive Single Document Summarization - Text (1)

Nallapati, Ramesh, Bowen Zhou, Cicero Nogueira dos santos, Caglar Gulcehre, and Bing Xiang. "Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond." *ArXiv:1602.06023 [Cs]*, February 18, 2016. http://arxiv.org/abs/1602.06023.

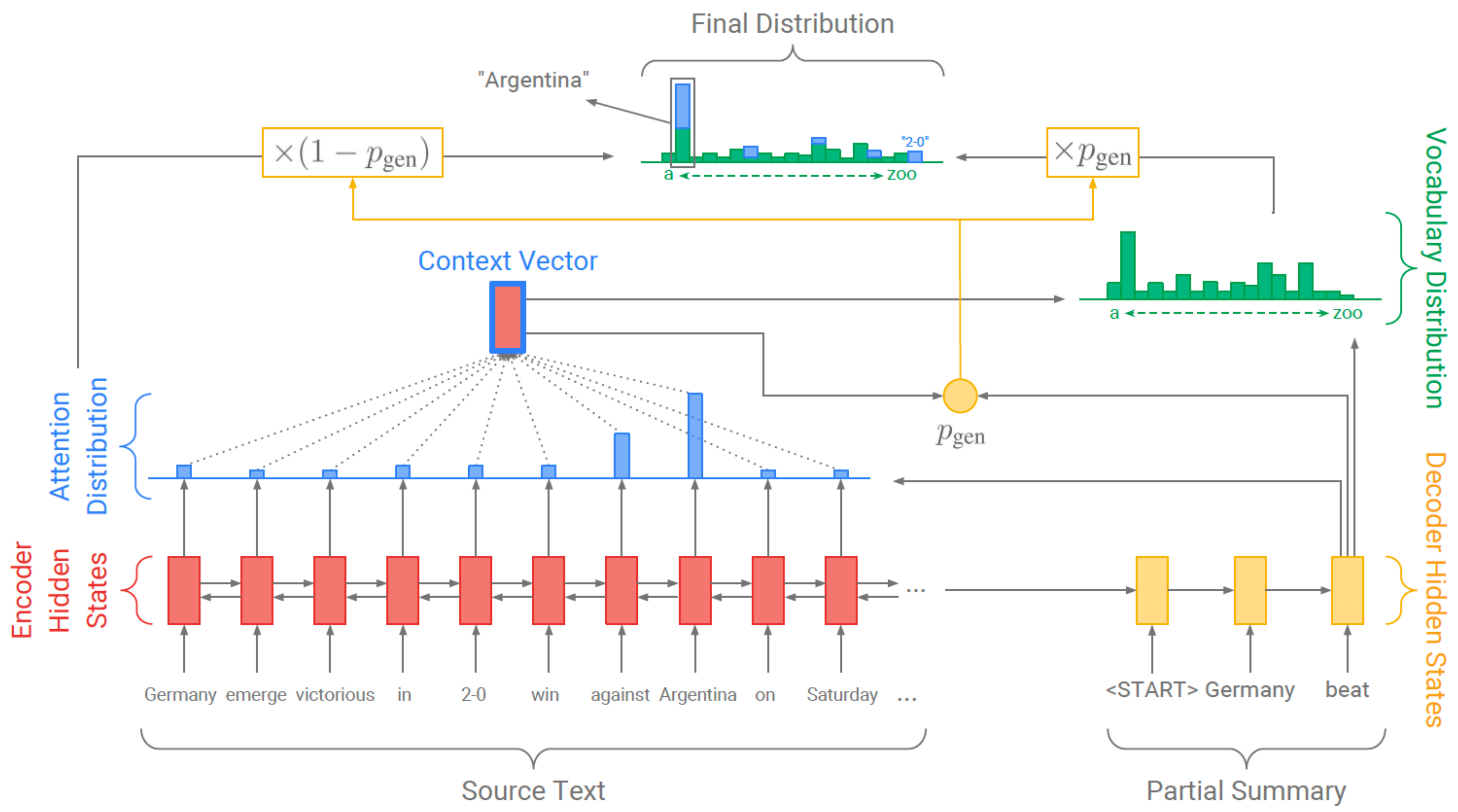


Main idea / contributions

- Introduce encoder / decoder structure for summarisation.
- Add tags to the encoder.
- Introduce pointing probability, for unknown words.
- Introduce hierarchical attention mechanism.

Abstractive Single Document Summarization - Text (2)

See, Abigail, Peter J. Liu, and Christopher D. Manning. "Get To The Point: Summarization with Pointer-Generator Networks." *ArXiv:1704.04368 [Cs]*, April 14, 2017. http://arxiv.org/abs/1704.04368.



Main idea / contributions

- Make use of pointer network, that can copy words from the input **at any time** - less unknown words.
- Introduce coverage mechanism that keeps track of previously paid attention.

Abstractive Single Document Summarization - Text

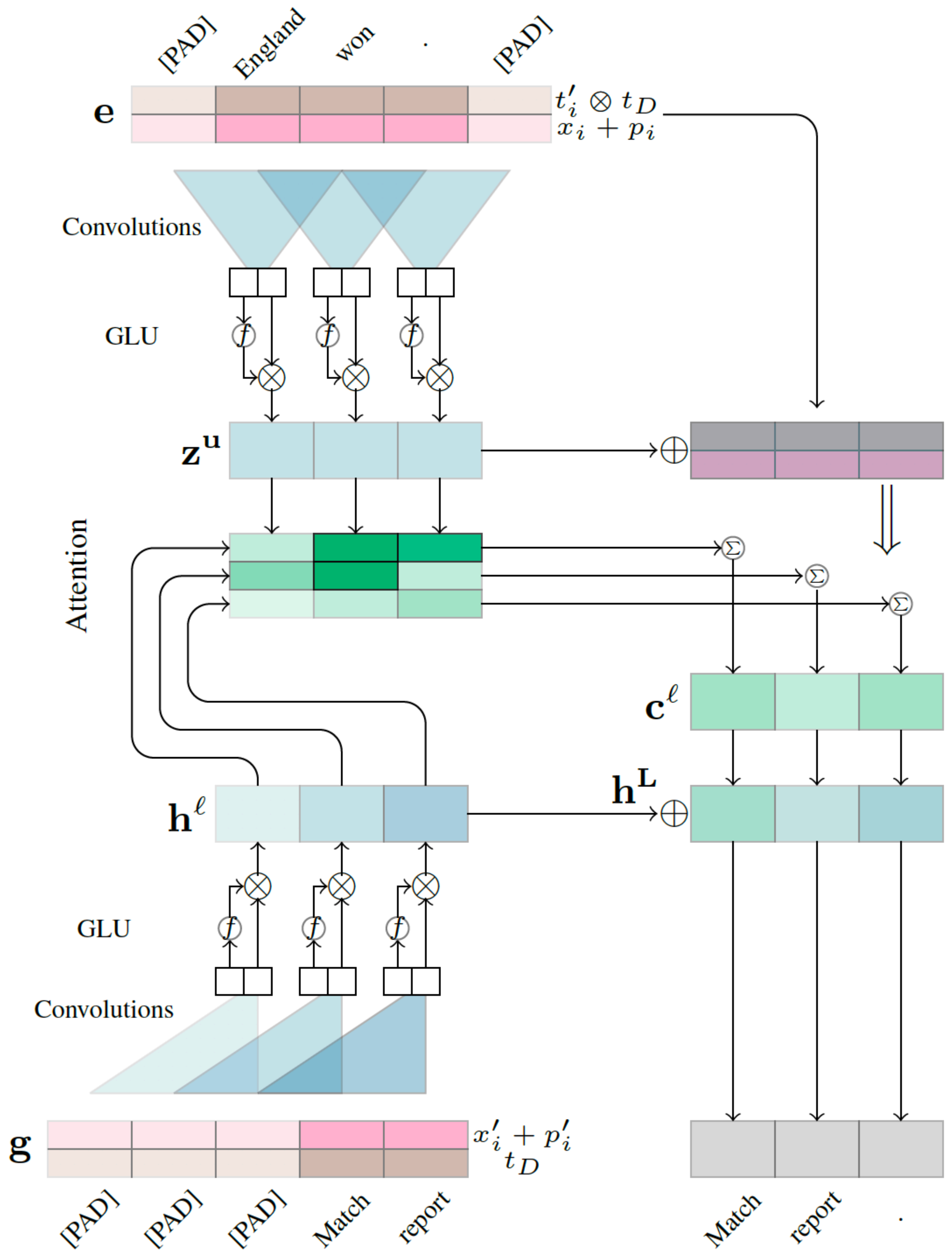
Challenges

Questionable grammaticality.

Not very abstractive - Next method addresses this.

Abstractive Single Document Summarization - Text (3)

Narayan, Shashi, Shay B. Cohen, and Mirella Lapata. "Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization." In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 1797–1807. Brussels, Belgium: Association for Computational Linguistics, 2018. http://www.aclweb.org/anthology/D18-1206.



Main idea / contributions

- Use convolutions, to capture the hierarchical structure of the text.
- Add topic information, to force the summary into a certain direction.
- They introduce a new dataset!

Single Document Summarization - Multimodal

Multimodal Summarization with Multimodal Output

Zhu, Junnan, Haoran Li, Tianshang Liu, Yu Zhou, Jiajun Zhang, and Chengqing Zong. "MSMO: Multimodal Summarization with Multimodal Output." In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4154–4164. Brussels, Belgium: Association for Computational Linguistics, 2018. http://www.aclweb.org/anthology/D18-1448.

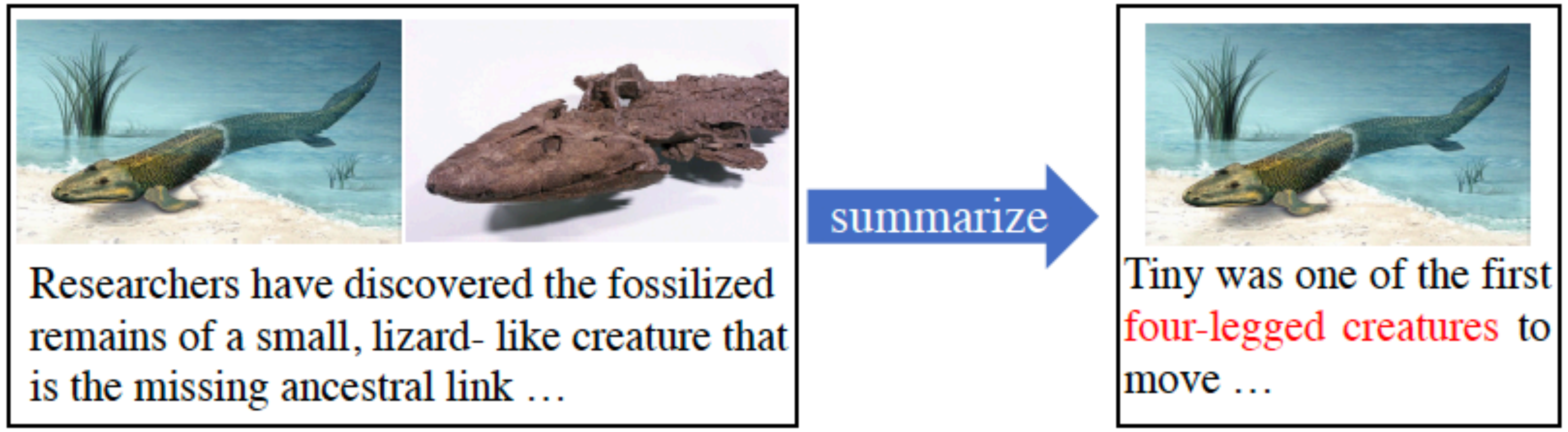


Figure 1: The illustration of our proposed task – Multimodal Summarization with Multimodal Output (MSMO). The image can help better understand the text in the red font.

Multimodal Summarization with Multimodal Output

Zhu, Junnan, Haoran Li, Tianshang Liu, Yu Zhou, Jiajun Zhang, and Chengqing Zong. "MSMO: Multimodal Summarization with Multimodal Output." In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4154–4164. Brussels, Belgium: Association for Computational Linguistics, 2018. http://www.aclweb.org/anthology/D18-1448.

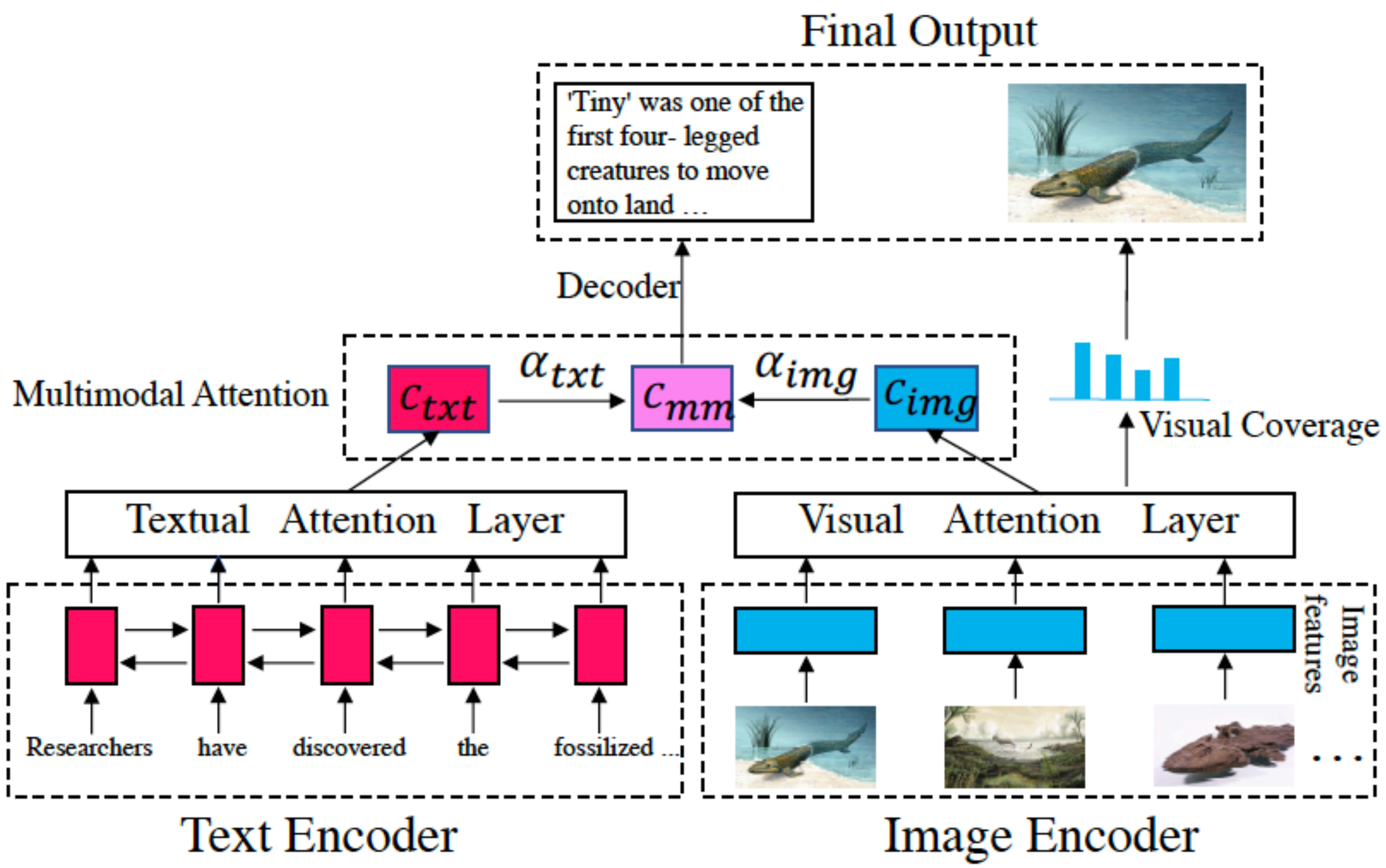


Figure 2: The framework of our model.

Summary

- Automatic summarization has the goal to automatically find the important bits in all the data and return these.

Summary

- Automatic summarization has the goal to automatically find the important bits in all the data and return these.
- We can summarise different types of data. In this lecture we have focussed on text and we have briefly touched upon multimodal summarization.

Summary

- Automatic summarization has the goal to automatically find the important bits in all the data and return these.
- We can summarise different types of data. In this lecture we have focussed on text and we have briefly touched upon multimodal summarization.
- Summarization techniques can be divided into extractive methods and abstractive methods.

Summary

- Automatic summarization has the goal to automatically find the important bits in all the data and return these.
- We can summarise different types of data. In this lecture we have focussed on text and we have briefly touched upon multimodal summarization.
- Summarization techniques can be divided into extractive methods and abstractive methods.
- ROUGE and Human Evaluation are used to evaluate the produced summaries.

Questions?

Slides are available at maartjeth.github.io/#talks