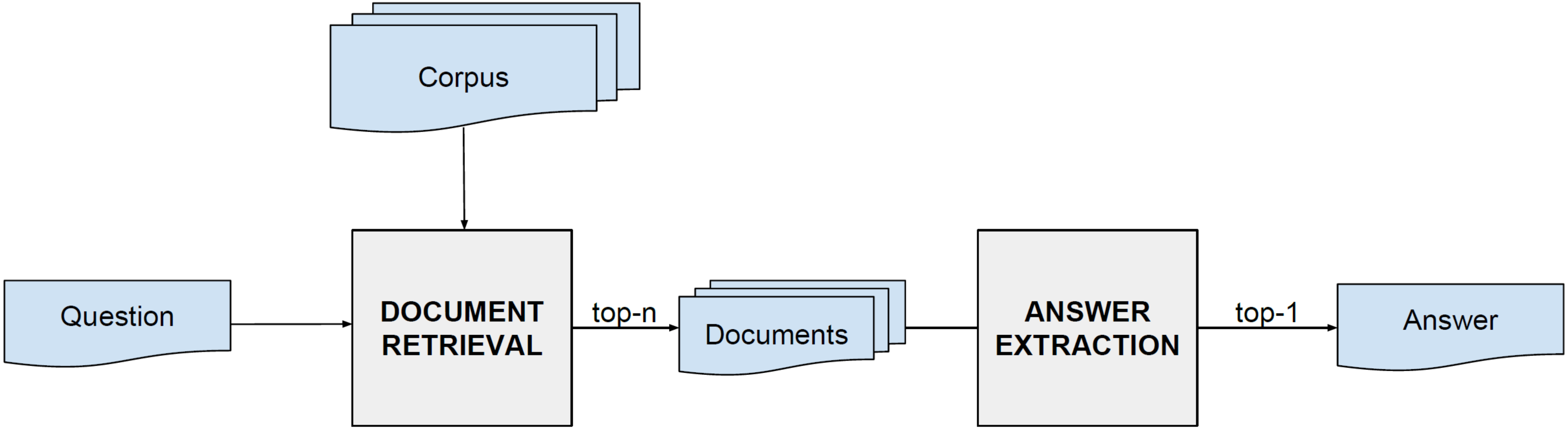
ETH zürich

Adaptive Document Retrieval for Deep Question Answering

Bernhard Kratzwald and Stefan Feuerriegel
ETH Zürich

Empirical Methods in Natural Language Processing (EMNLP) 2018

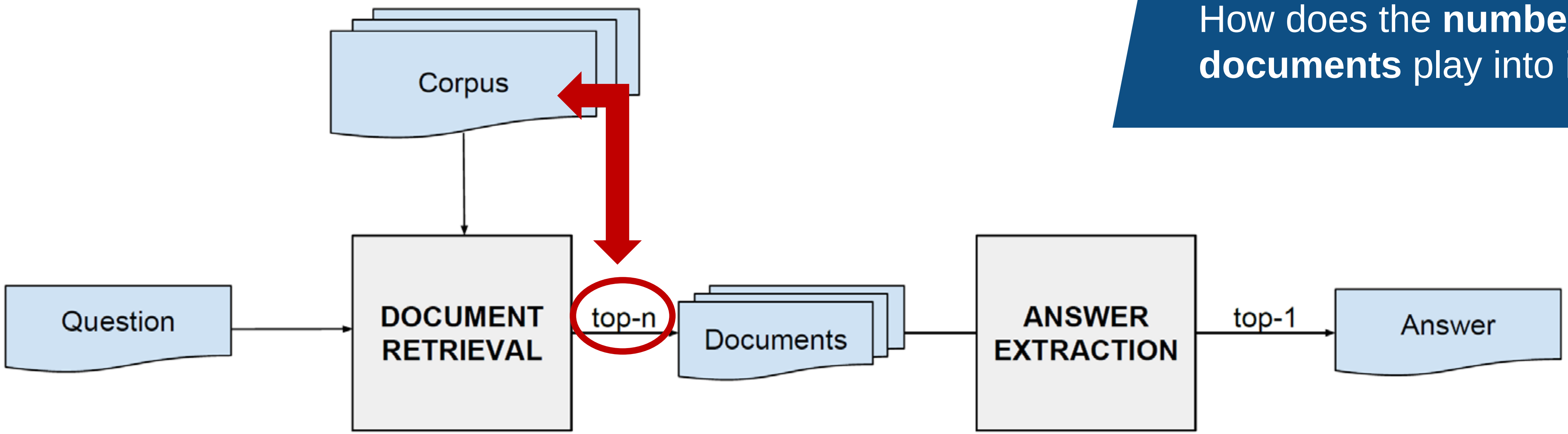
Content based QA systems with neural answer extraction



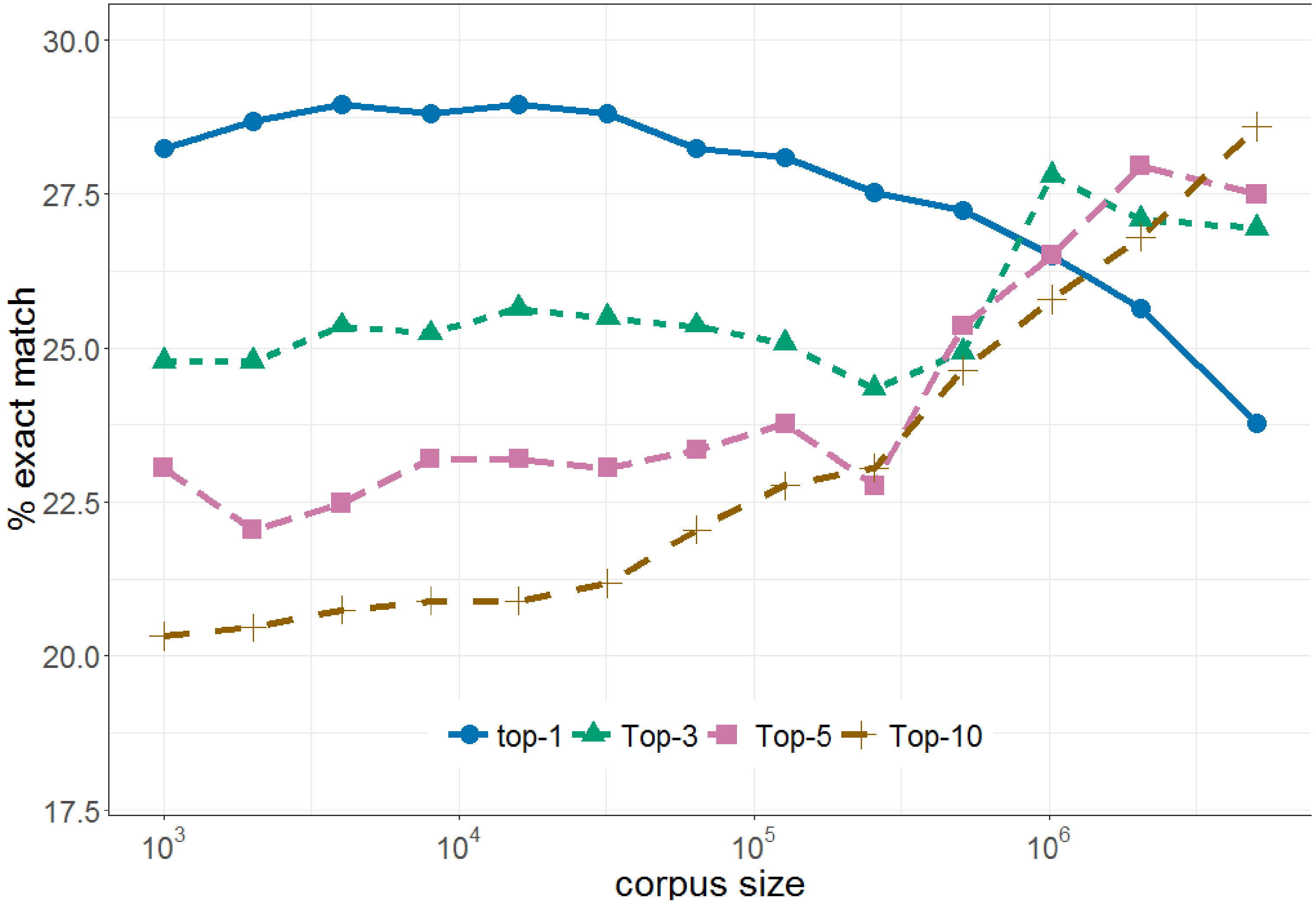
How does top-*n* document retrieval influence the performance of neural QA systems?

How does the **corpus size effect the over-all performance** of our QA system?

How does the **number of selected documents** play into it?



Varying top-*n* retrieval under a dynamic corpus

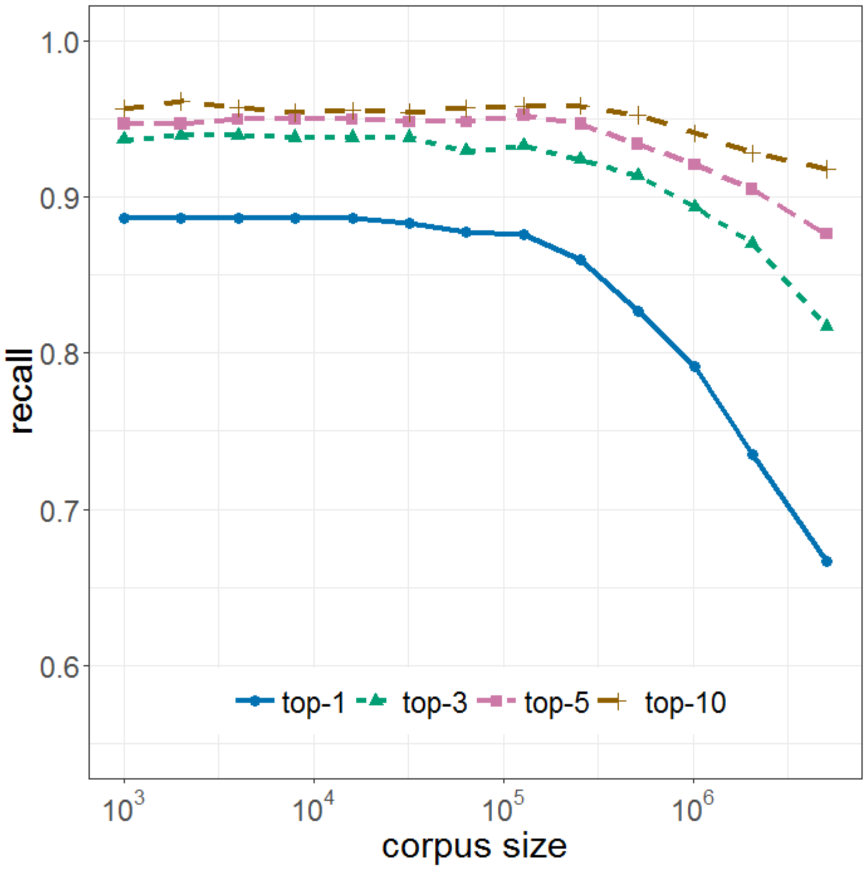


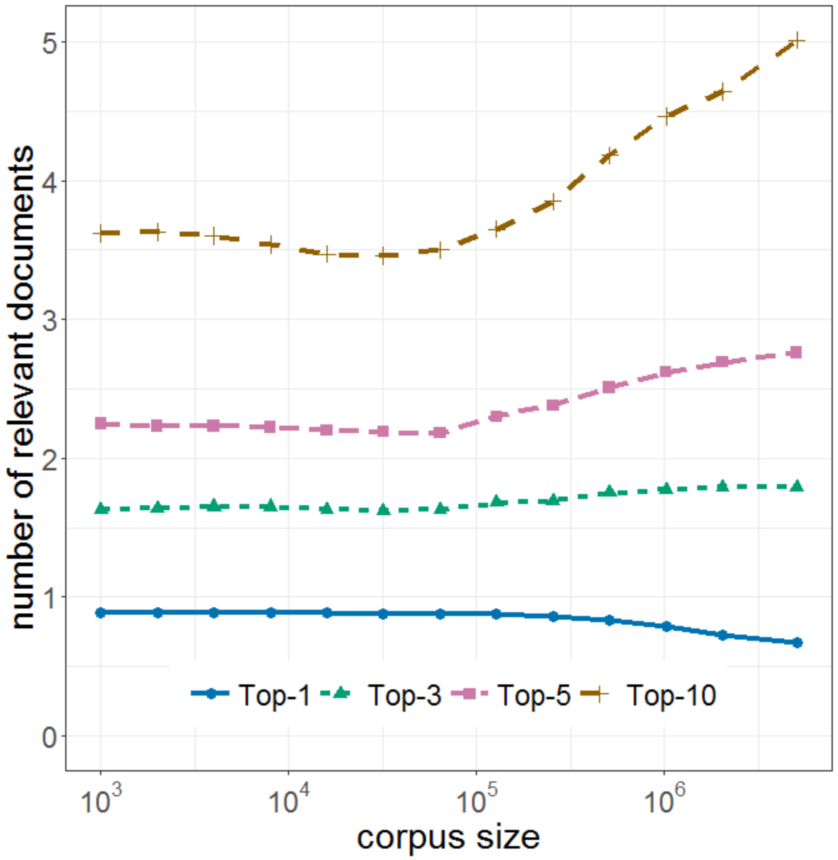
Findings

- For small corpora, a top-1 system outperforms any other configuration
- For bigger corpora, a top-*n* system preforms better than any top-1

Experiments based on DrQA System:
Chen, Danqi, et al. "Reading Wikipedia to Answer Open-Domain Questions." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (2017)

Document retrieval under a dynamic corpus



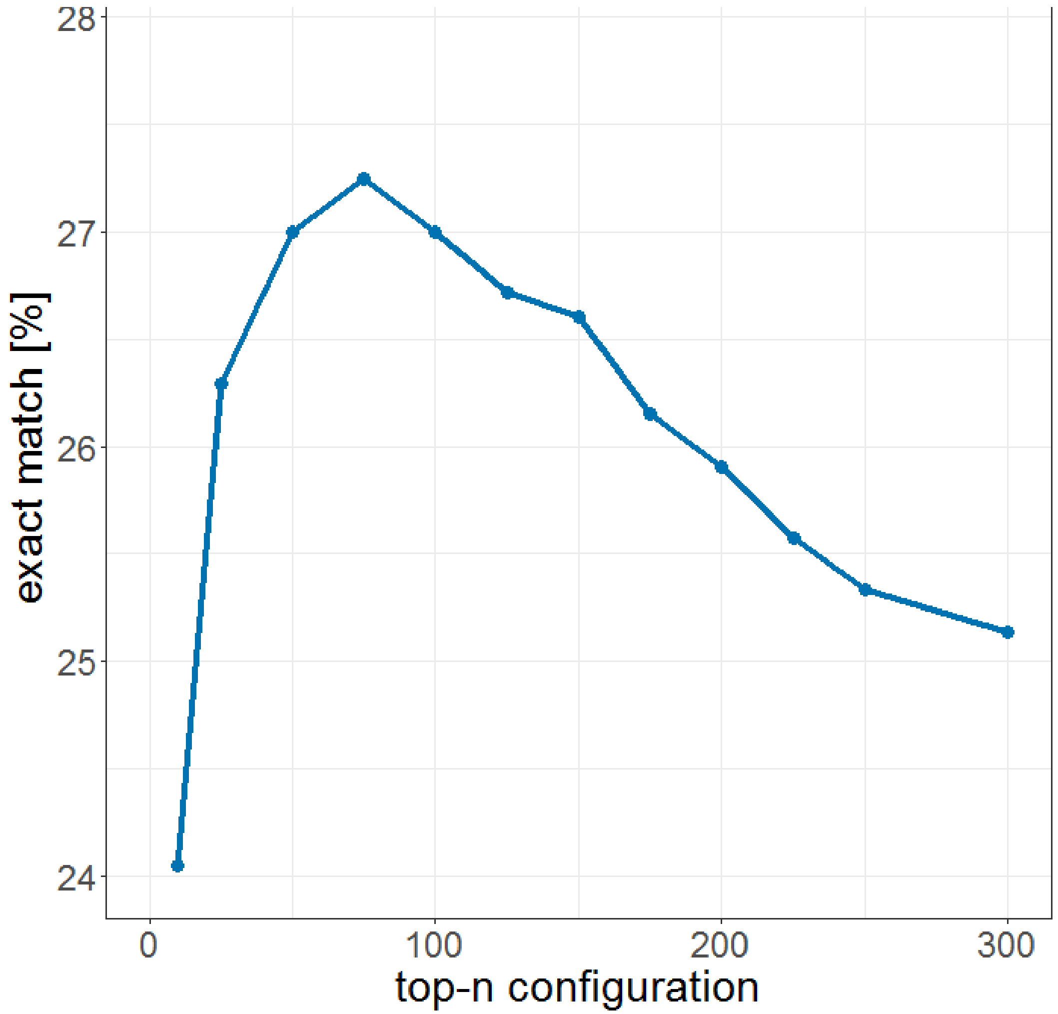


Findings

- The **recall@1** is very volatile when the corpus grows
- **Top-*n*** retrieval results in a higher density of information

Experiments based on DrQA System:
Chen, Danqi, et al. "Reading Wikipedia to Answer Open-Domain Questions." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*

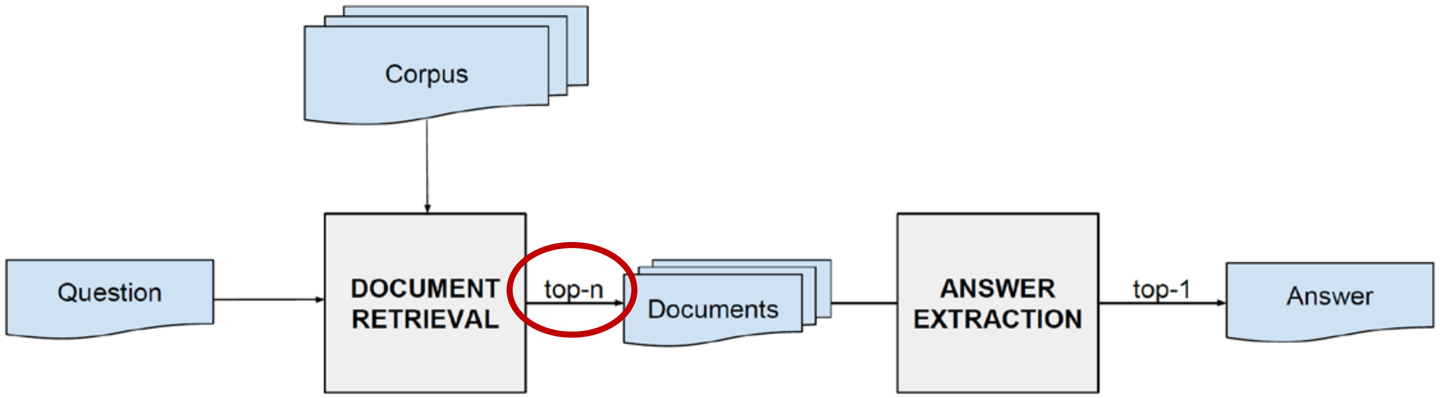
Varying top-*n* retrieval under a static corpus



Findings

- In the beginning, **selecting more paragraphs** results in higher performance
- Selecting **too many paragraphs** results in a performance loss

The noise information trade-off in neural question answering



Noise-information trade-off

- Pick a **query dependent number of documents** to select
- Select enough documents to answer the question
  → but not too many to select a wrong candidate answer

A threshold baseline for adaptive retrieval

- Number of documents is determined independently for every query

$$n_i = \max_k \sum_{j=1}^{k} s_i^{(j)} < \theta$$

$s_i^{(j)}$ score of document *j*, in query *i*

- **The Idea:** The more confident we are, the less documents we want to select

A ordinal regression model for adaptive retrieval

- We model the **number of top-*n* documents** by

$$y_i = f([s_i^{(1)}, \ldots, s_i^{(\tau)}]) = \lceil s_i^T \beta \rceil$$

- **The Idea:** Learn the **cut-off point** by minimizing

$$\mathcal{L} = \| \lceil X\beta \rceil - y \|_1 + \lambda \, \|\beta\|_2$$

- For prediction, we use an **additional off-set**

$$\hat{n}_i = \lceil {s'}_i^T \hat{\beta} \rceil + b$$

Results on the full Wikipedia corpus

	SQuAD	TREC	WebQuestions	WikiMovies
DrQA (Chen et al., 2017)†	29.3	27.5	18.5	36.6
Threshold-based ($\theta = 0.75$)	**29.8**	28.7	19.2	38.6
Ordinal regression ($b = 1$)	29.7	28.1	19.4	38.0
Ordinal regression ($b = 3$)	29.6	**29.3**	**19.6**	38.4
R^3 (Wang et al., 2018)	29.1	28.4	17.1	**38.8**

†: Numbers vary slightly from those reported in the original paper, as the public repository was optimized for runtime performance.

Experiment

- [DrQA] system #1
- Corpus: Full Wikipedia

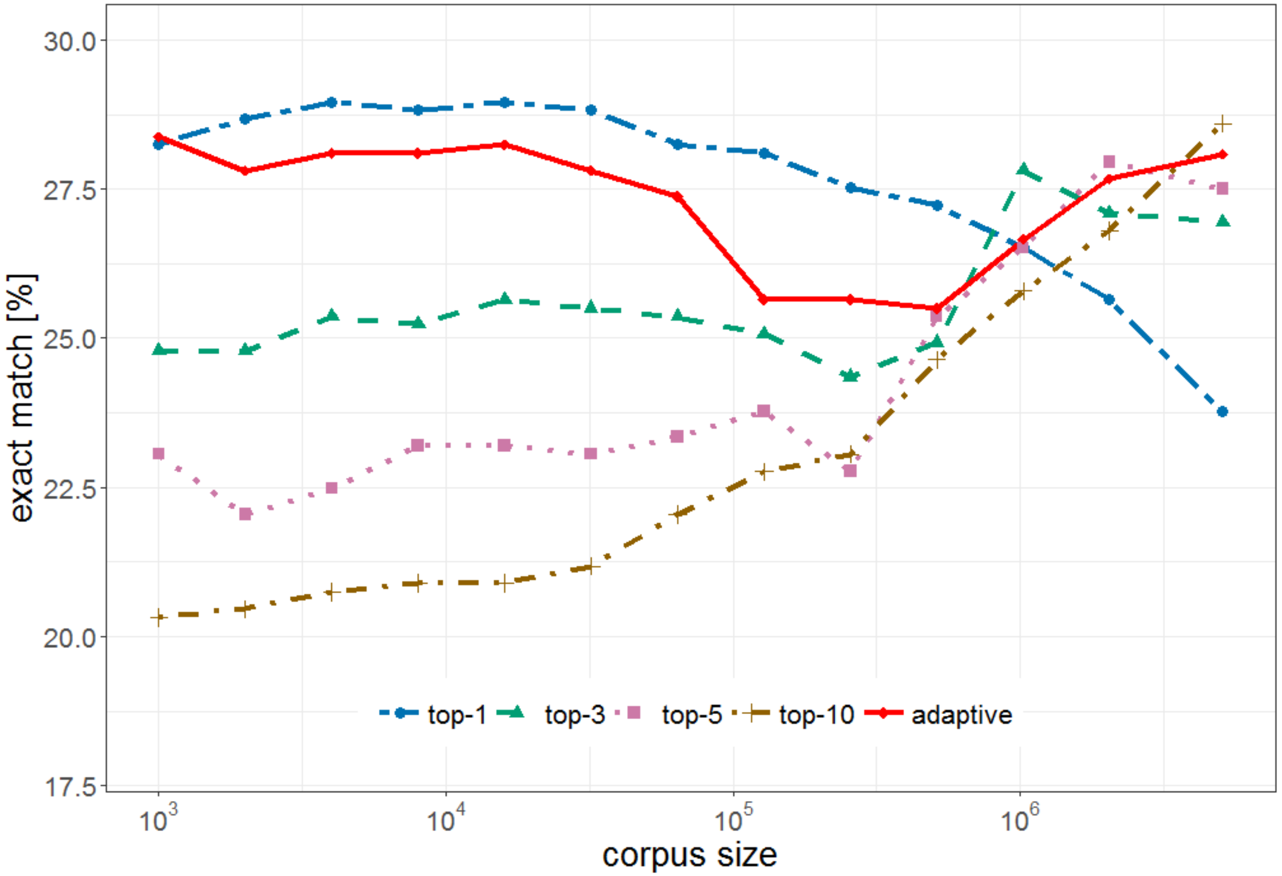
Experiment

- Paragraph system #2
- Corpus: Full Wikipedia

	SQuAD	TREC	WebQuestions	WikiMovies
Top-50 System	27.0	23.5	15.1	24.4
Top-80 System	27.2	25.9	14.9	26.0
Threshold-based ($\theta = 0.75, \tau = 100$)	27.2	**27.1**	15.4	26.3
Ordinal regression ($b = 3, \tau = 250$)	**27.3**	**27.1**	**16.7**	**26.5**

Chen, Danqi, et al. "Reading Wikipedia to Answer Open-Domain Questions." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (2017)
Wang, Shuohang, et al. "R3: Reinforced Reader-Ranker for Open-Domain Question Answering.“ *Conference on Artificial Intelligence AAAI* (2018).

Results under varying corporas



The adaptive approach remains the most **robust**

It's the 'safest' bet in terms of regret
→ i.e. when the corpus grows/shrinks over time

Experiments based on DrQA System:
Chen, Danqi, et al. "Reading Wikipedia to Answer Open-Domain Questions." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (2017)

Thank you very much!

Questions?

Feel free to reach out to us: bkratzwald@ethz.ch

Slides and more background:
https://github.com/bernhard2202/adaptive-ir-for-qa