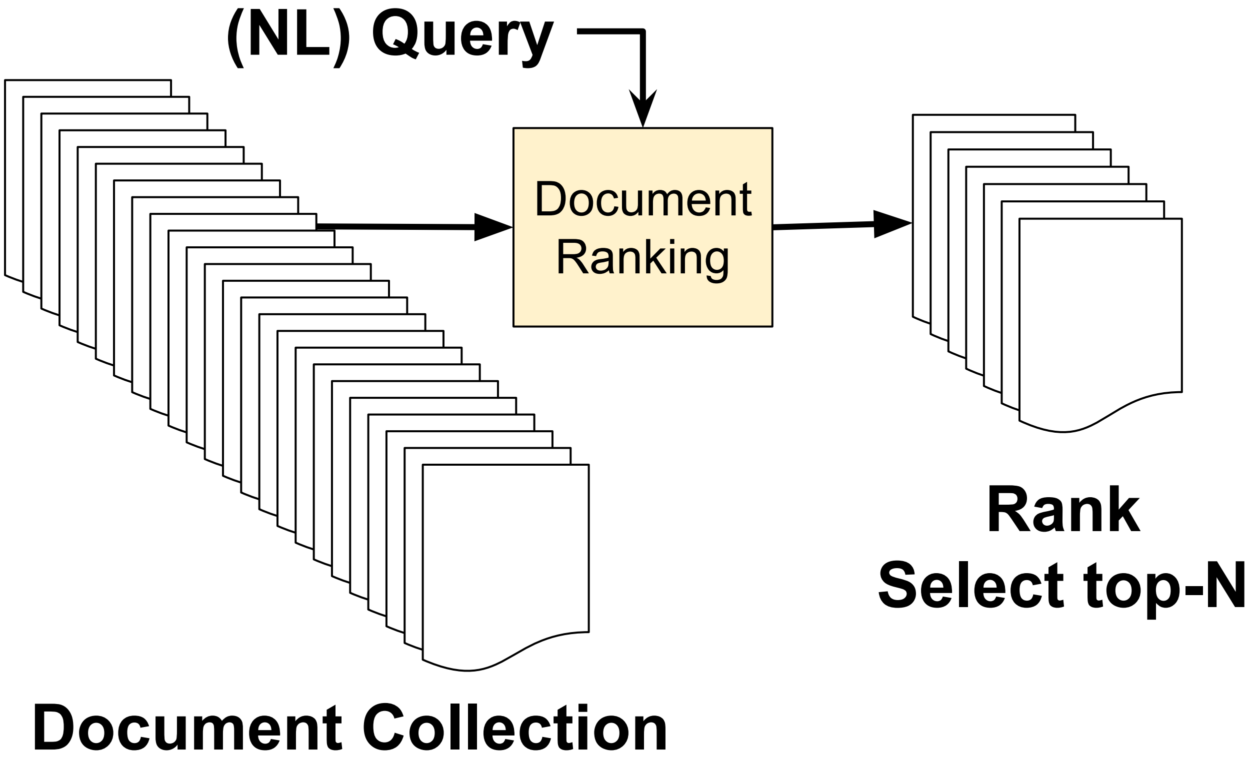
Deep Relevance Ranking Using Enhanced Document-Query Interactions

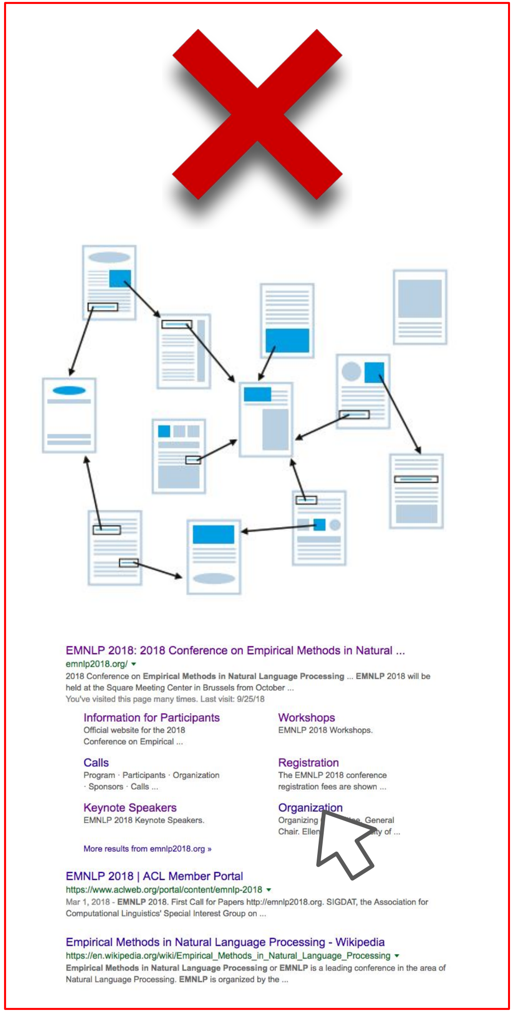
Ryan McDonald, George Brokos and Ion Androutsopoulos
https://github.com/nlpaueb

ΟΠΑ AUEB

Google AI

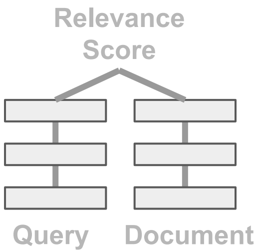
Ad-hoc Retrieval - Relevance Ranking



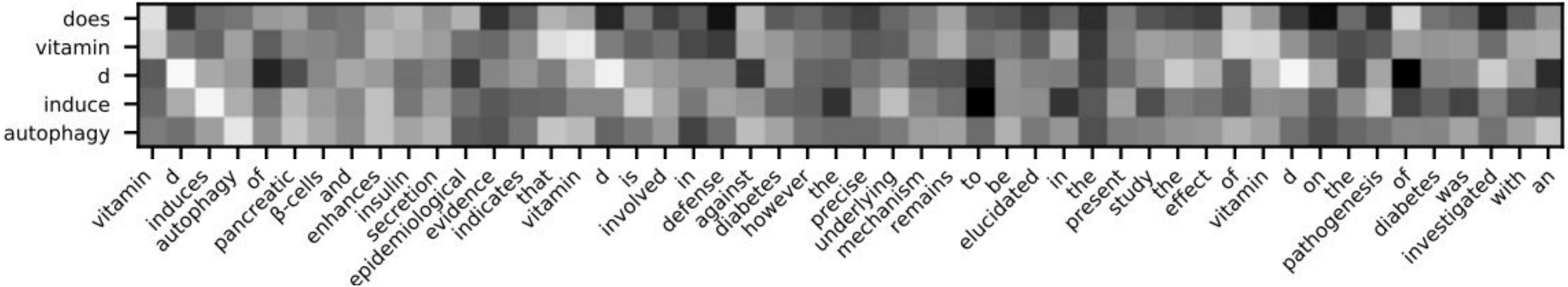


(Deep) Ad-hoc Retrieval / Relevance Ranking

- **Interaction-based**
    - DeepMatch (Lu and Li 2013)
    - ARC-II (Hu et al. 2014)
    - MatchPyramid (Pang et al. 2016)
    - **DRMM (Guo et al. 2016)**
    - **PACRR (Hui et al. 2017)**
    - DeepRank (Pang et al. 2017)
- Relevance-based

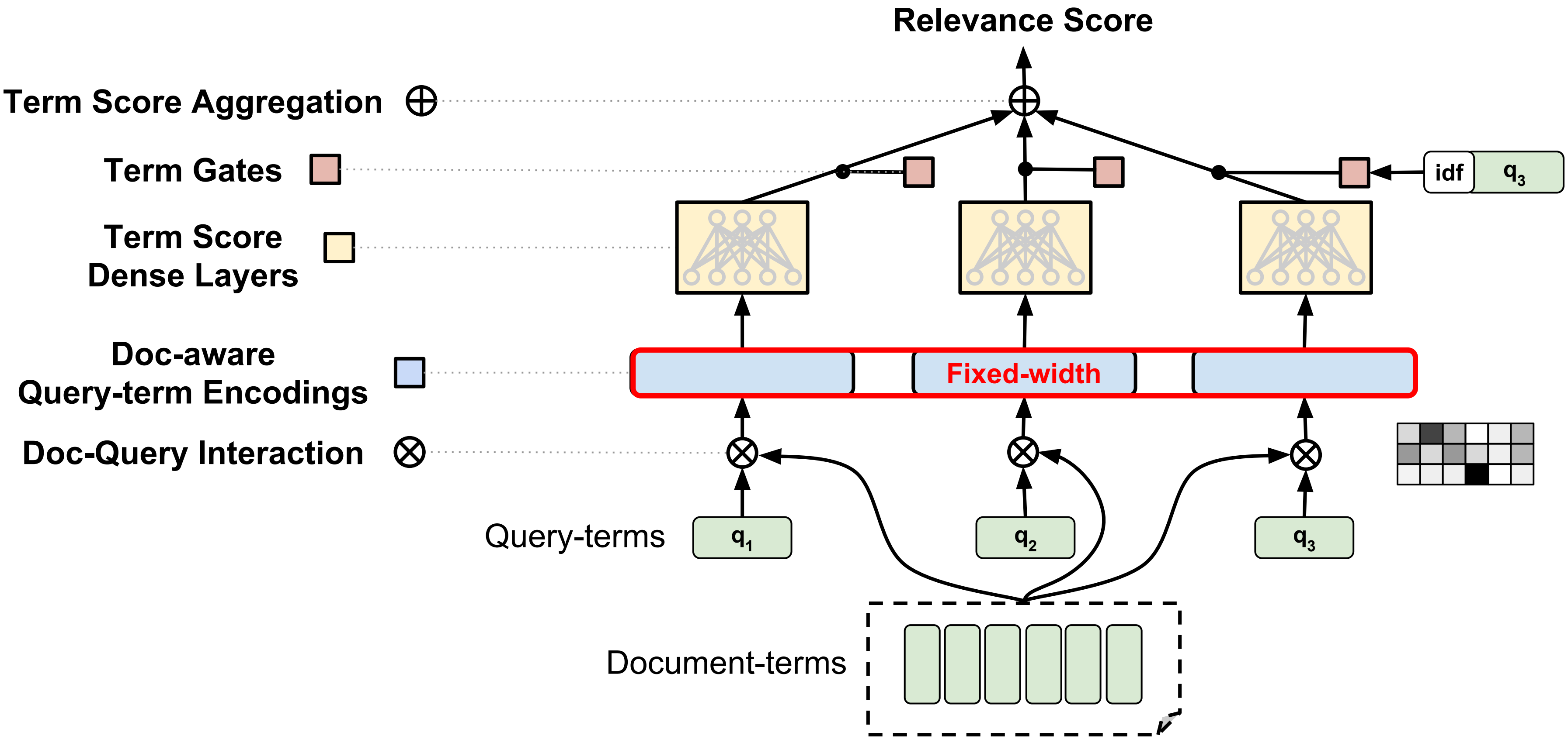


Query-Doc term similarity matrices



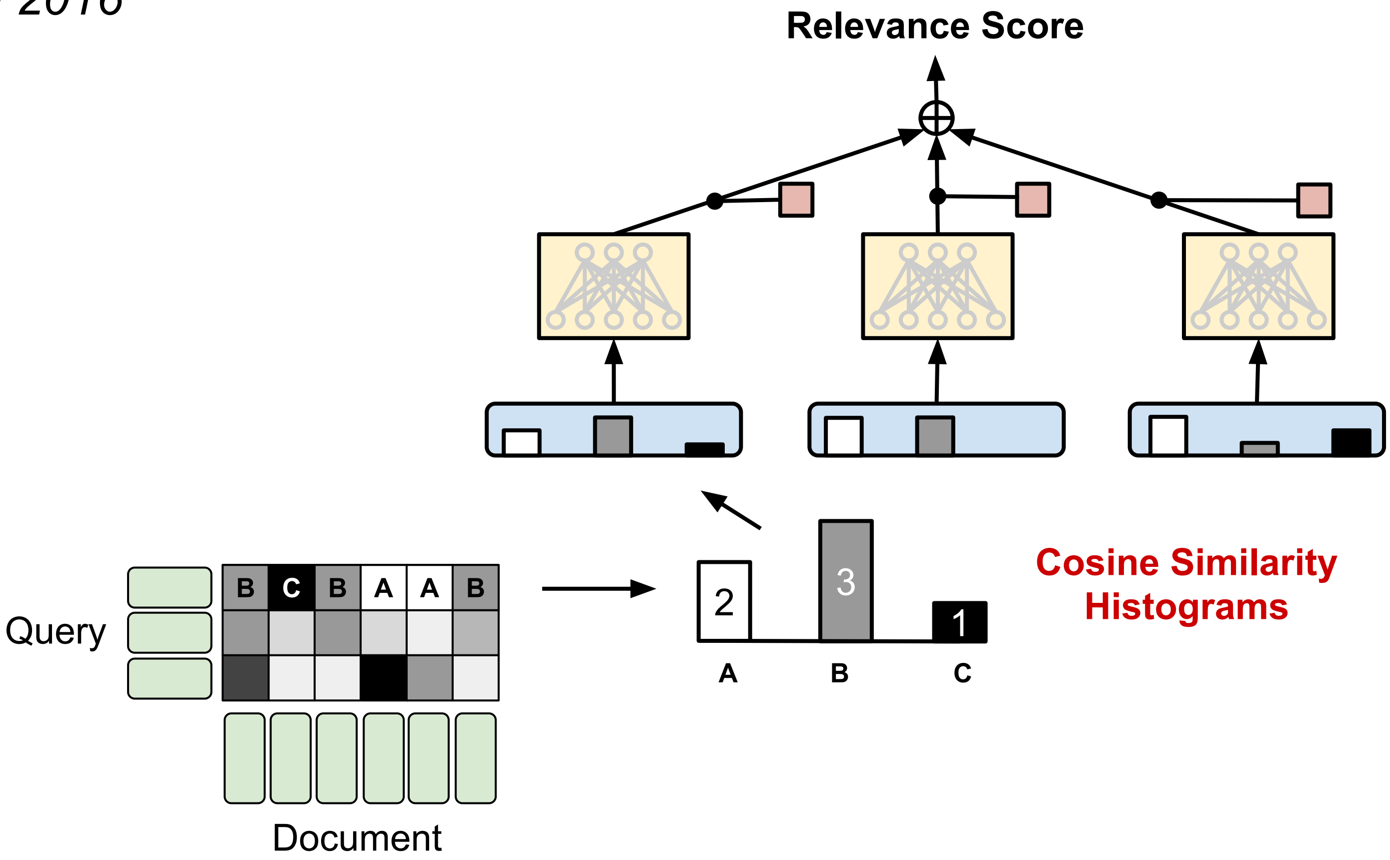
Deep Relevance Matching Model (DRMM)

Guo et al 2016

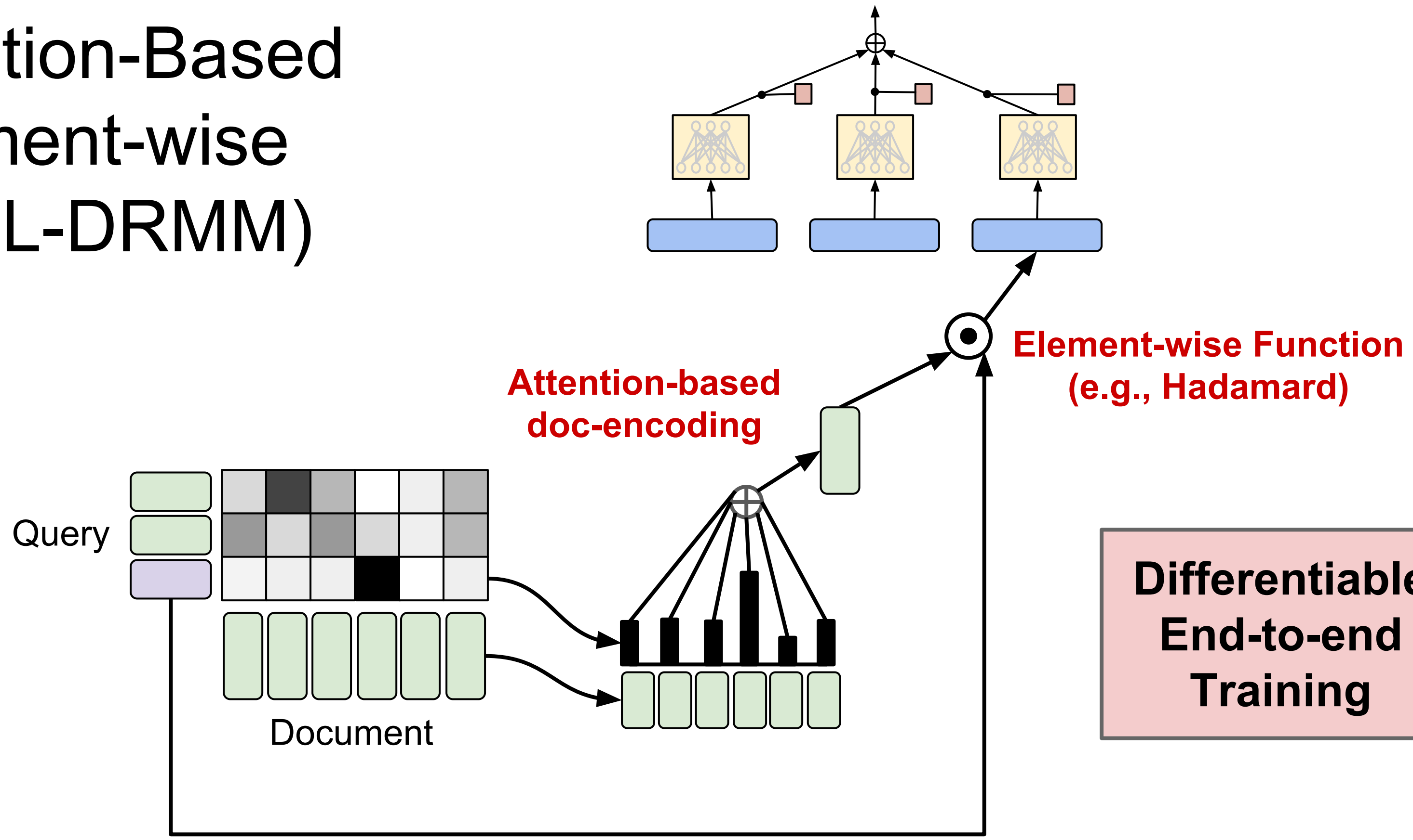


Deep Relevance Matching Model (DRMM)

Guo et al 2016

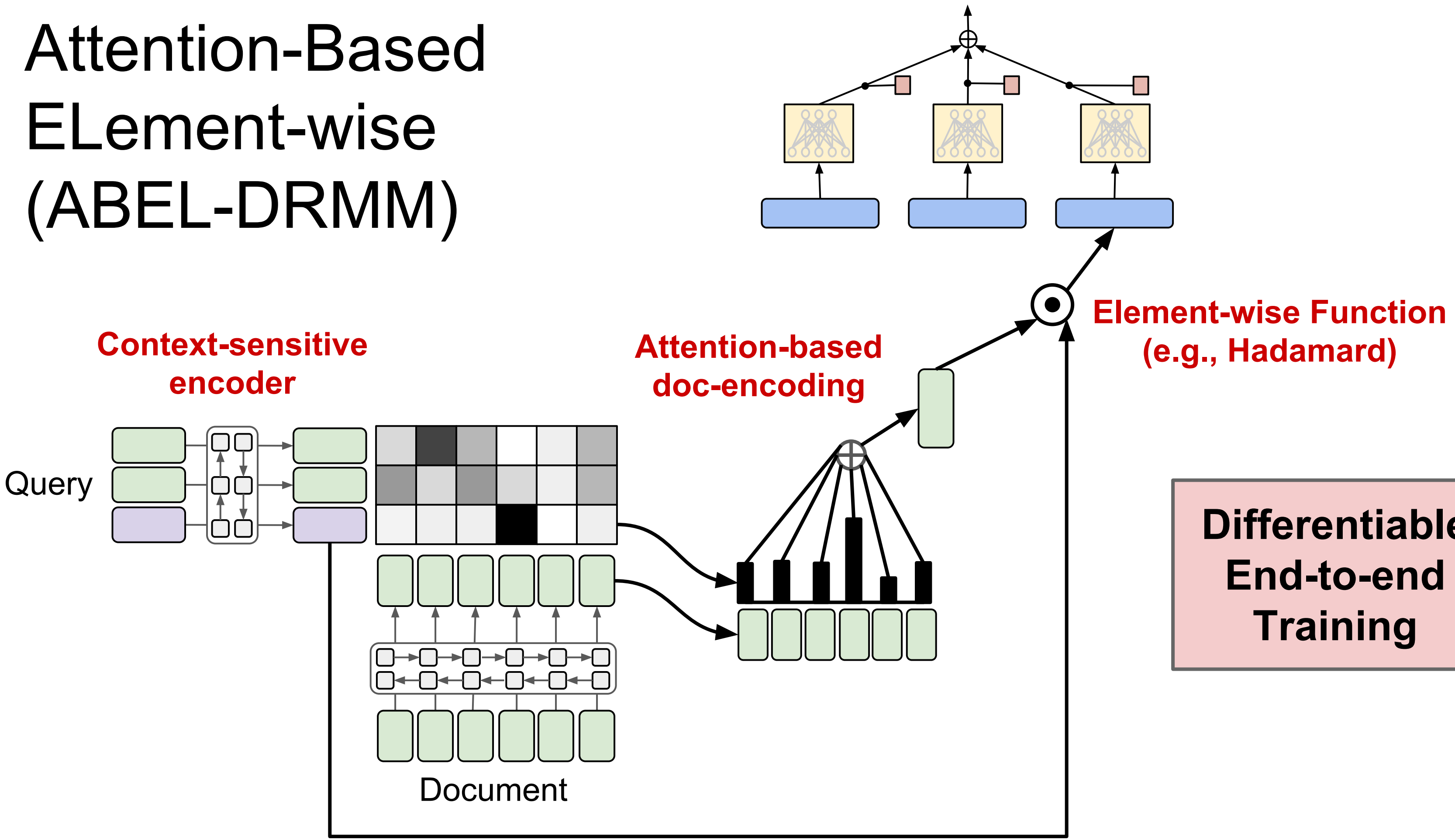


Attention-Based ELement-wise (ABEL-DRMM)



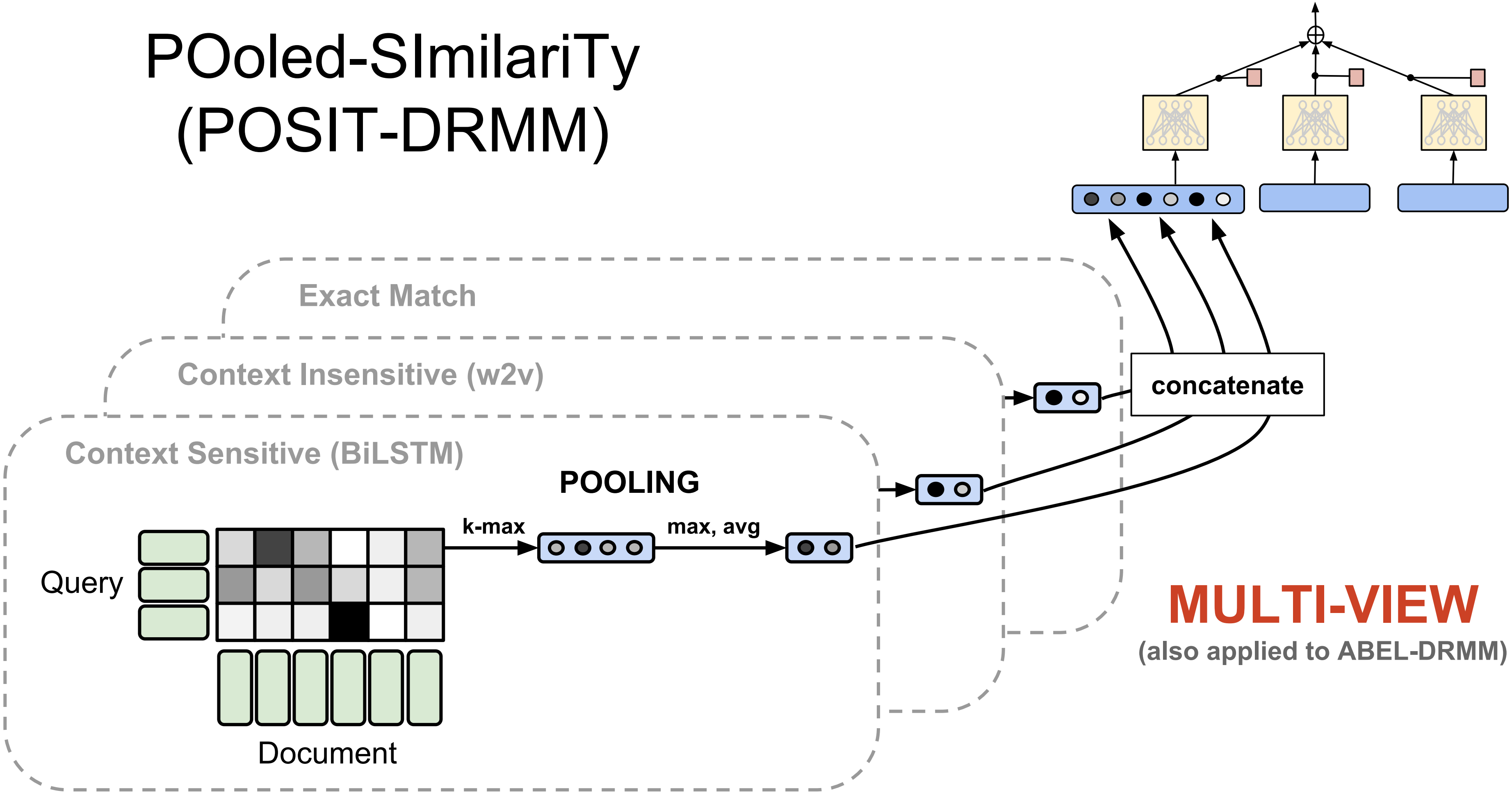
Differentiable End-to-end Training

Attention-Based ELement-wise (ABEL-DRMM)



Differentiable End-to-end Training

POoled-SImilariTy (POSIT-DRMM)



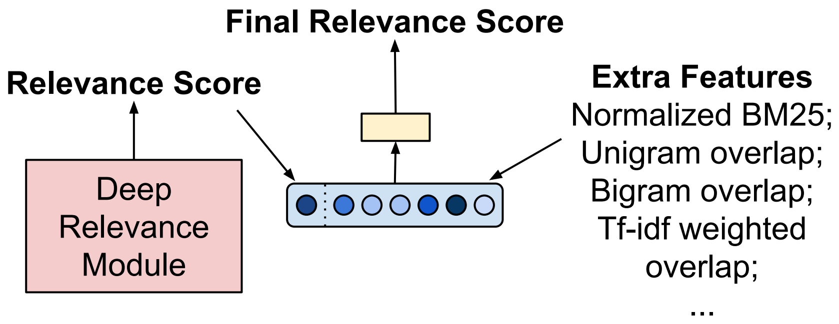
MULTI-VIEW

(also applied to ABEL-DRMM)

There is more going on … (read the paper)

- Re-ranked top N documents from traditional index using BM25 score

- All models use final linear layer
  - Severyn and Moschitti 2015
  - Mohan et al. 2017

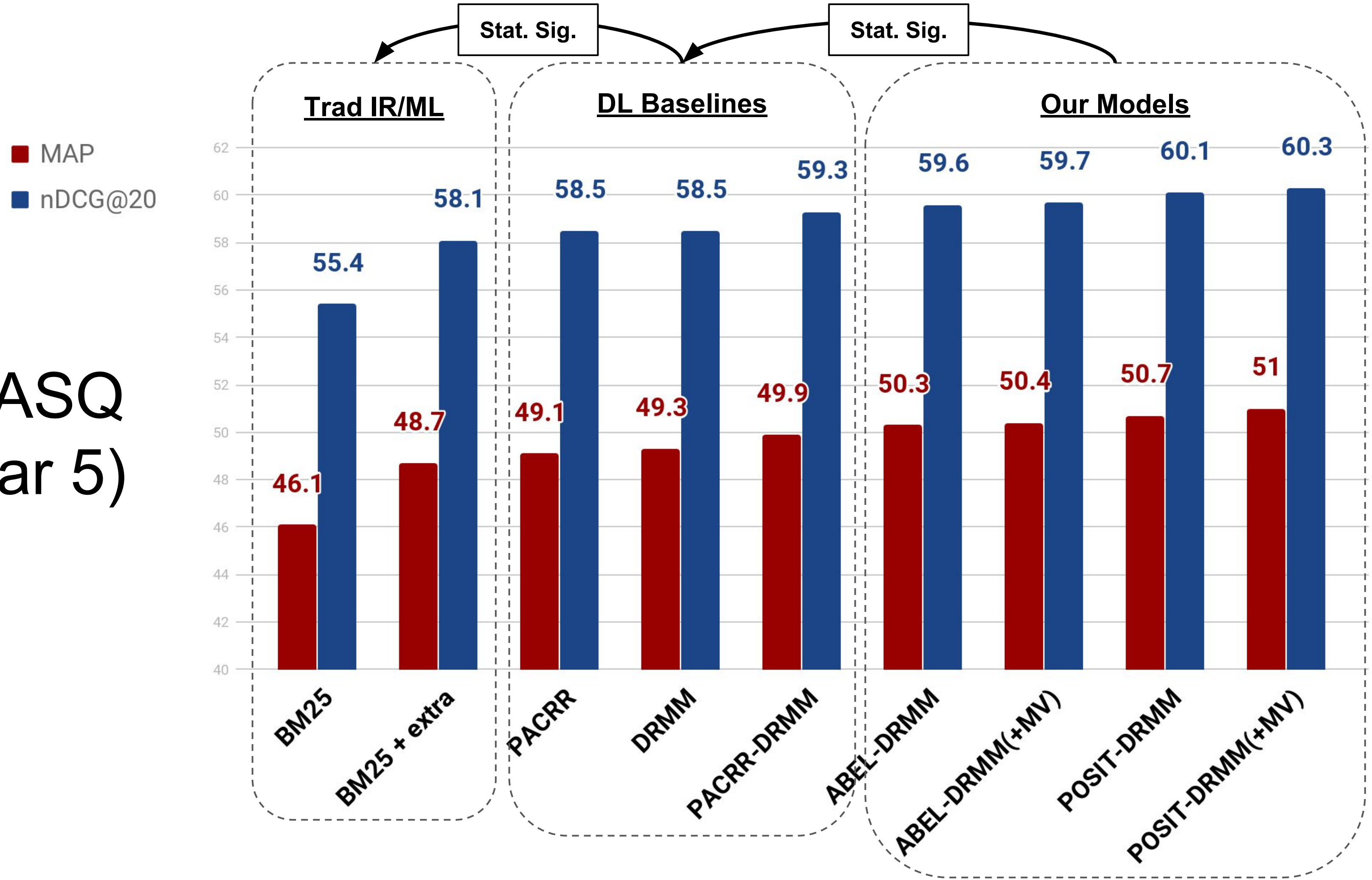


- Pairwise training with negative sampling

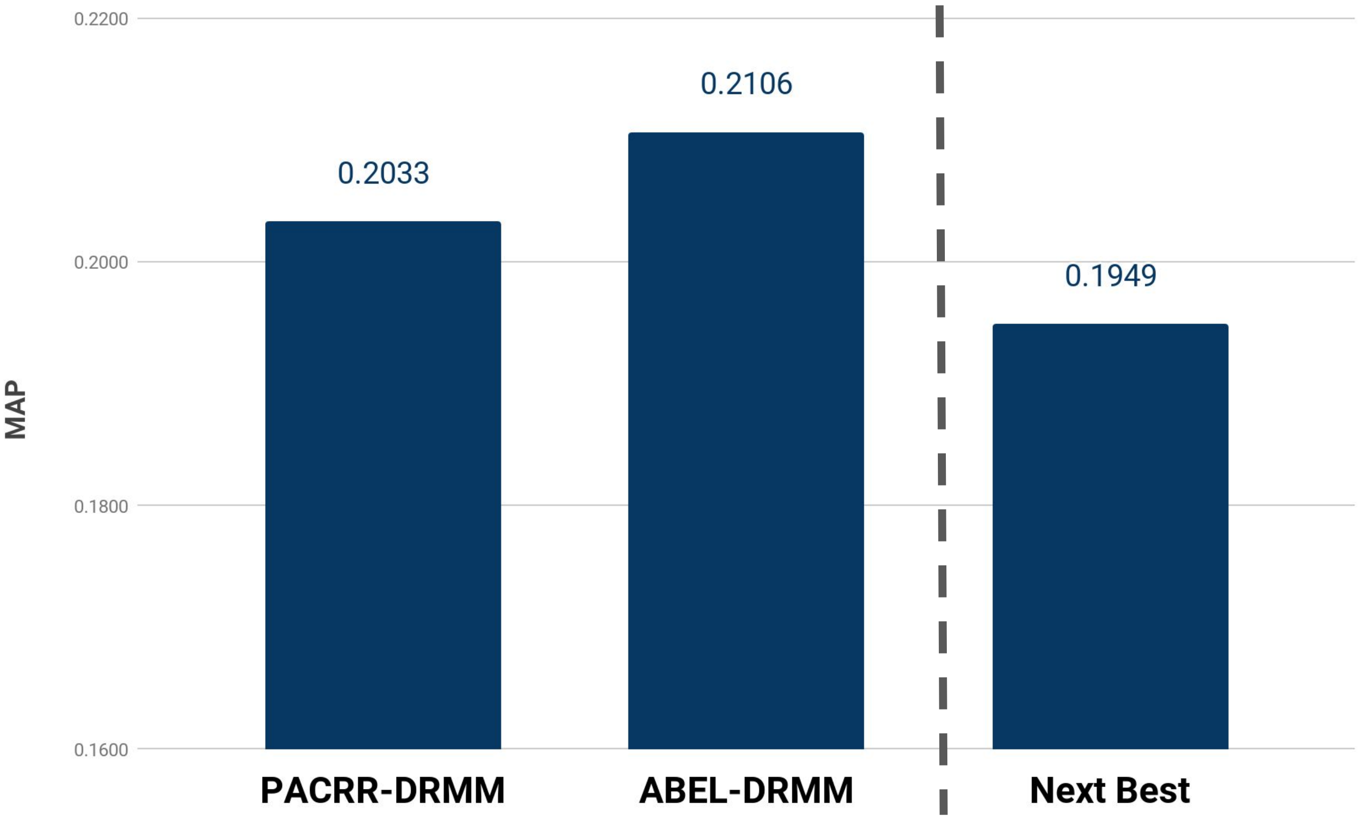
Experimental Datasets

	BioASQ Tsatsaronis et al. (2015)	TREC Robust Voorhees (2004)
Queries	2,251	250
Train / dev / test	Y1-4 / Y5b1 / Y5b2-5	5X-CV 60/20/20
Average query length	9.16	2.74
Collection size (documents)	17.7M	528K
Average document length	196.6	476.5
Vocabulary size	15.2M	1.4M
Average relevant documents	12.0	69.6

BioASQ (Year 5)

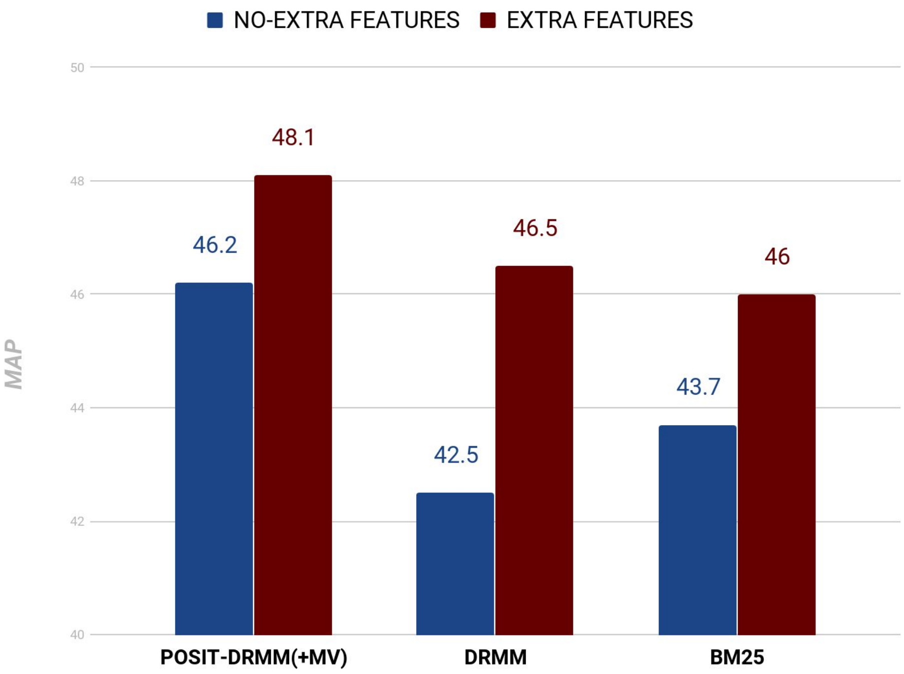


BioASQ Document Ranking (Year 6)

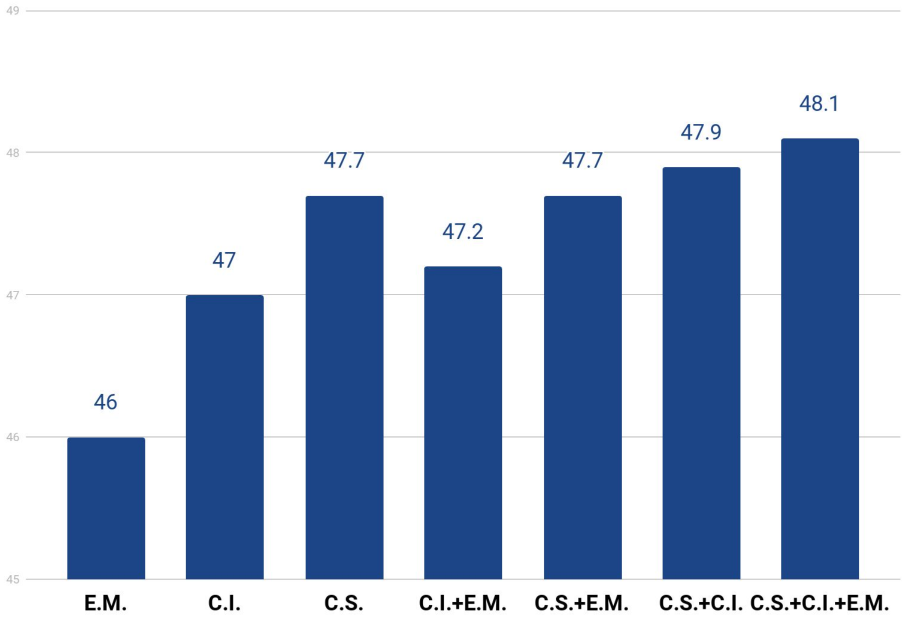


USTB: best system in years 3, 4, 5

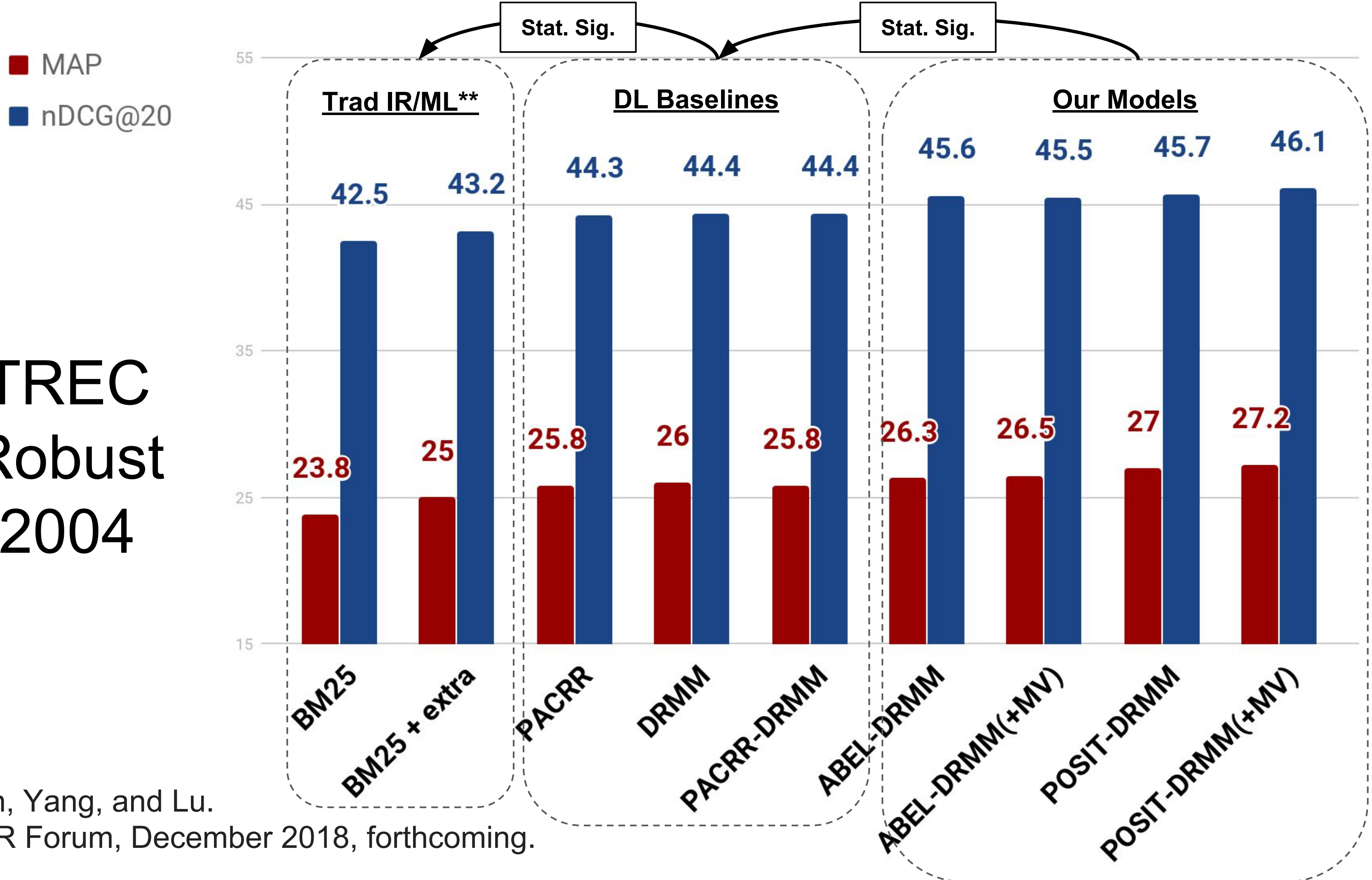
POSIT Ablation Study (BioASQ dev)



VIEW / MATCH-TYPE
C.I. = context-insensitive (word2vec)
C.S. = context-sensitive (BiLSTM)
E.M. = exact match



TREC Robust 2004



** Lin, Yang, and Lu.
SIGIR Forum, December 2018, forthcoming.

Summary

- Simple architectures can be effective for document ranking
  - Especially on NL query data sets
  - End-to-end training coupled with traditional IR signals
  - End-to-end architectures enables context-sensitive encoders

- POSIT-DRMM vs. ABEL-DRMM
  - Modeling best and top-k-average match improves accuracy

- Multi-view models promotes exact match in addition to vector match

Thanks!

https://github.com/nlpaueb/deep-relevance-ranking