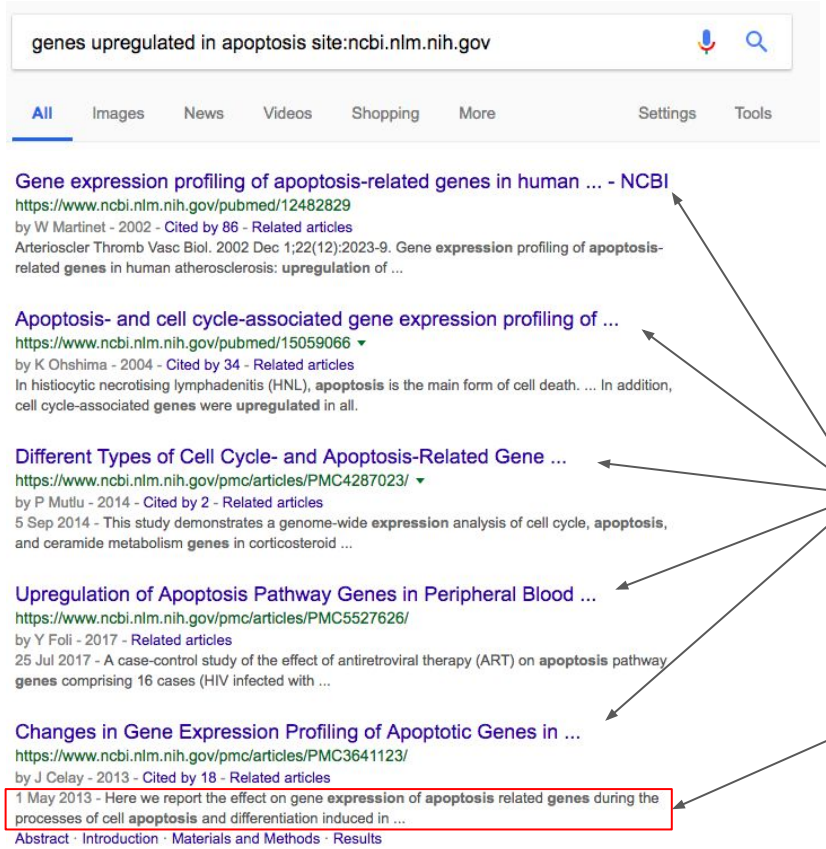
AUEB at BioASQ 6: Document and Snippet Retrieval

George Brokos, Polyvios Liosis, Ryan McDonald,
Dimitris Pappas and Ion Androutsopoulos
https://github.com/nlpaueb

Document and Snippet Ranking



BioASQ Task 6B Phase A

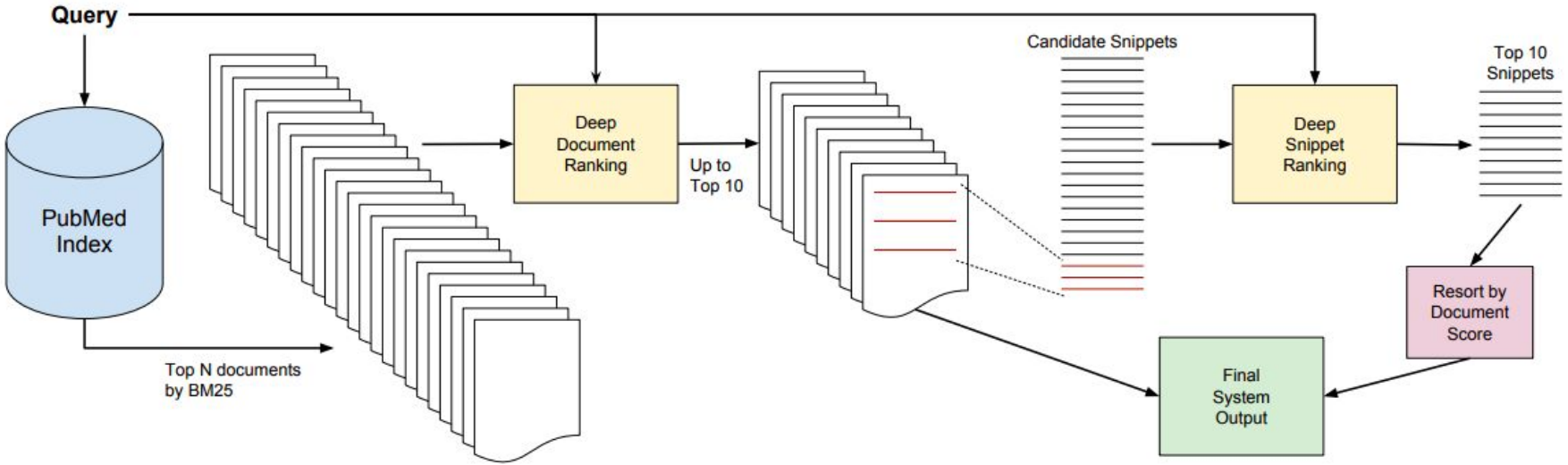
For each question

1. Select and rank 10 documents
2. Select and rank 10 snippets from these docs
   a. No restriction on length
   b. A document can yield multiple snippets
   c. A document can yield no snippets

RELEVANT DOCUMENTS

SUPPORTING SNIPPETS

Overall System Architecture

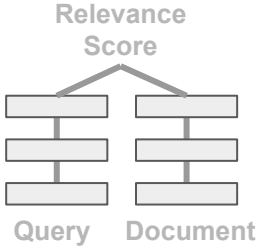


Document Ranking

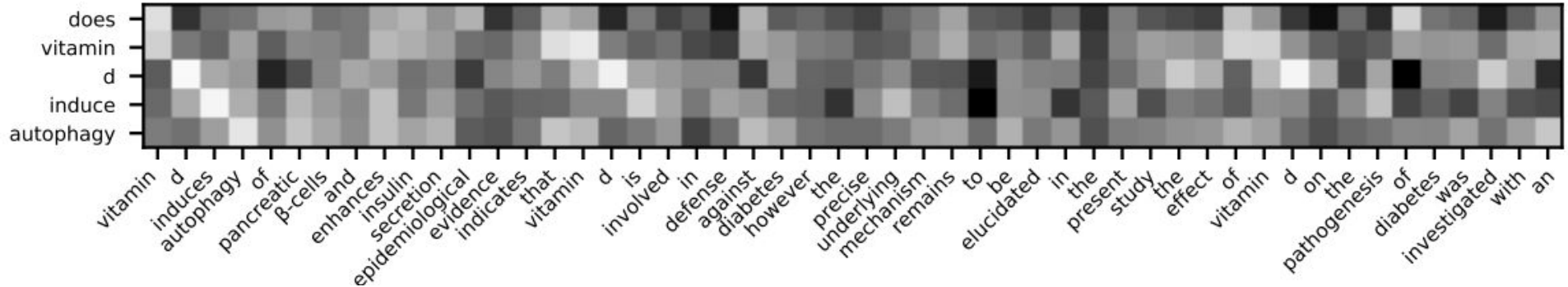
Deep Relevance Ranking Using Enhanced Document-Query Interactions
McDonald, Brokos and Androutsopoulos
EMNLP 2018
Saturday November 3rd at 9:36
Session: 5C: IR / Text Mining

(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

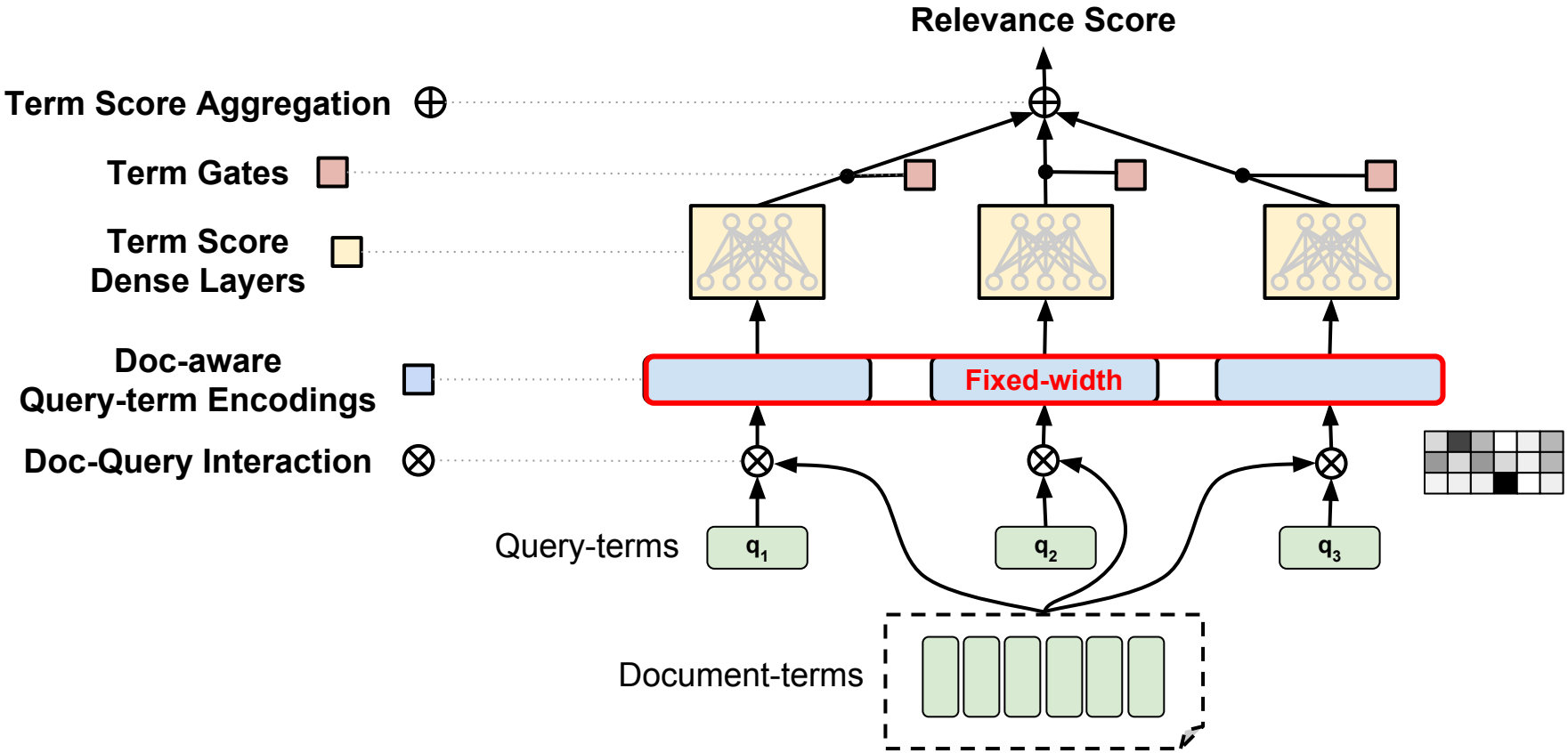


Query-Doc term similarity matrices



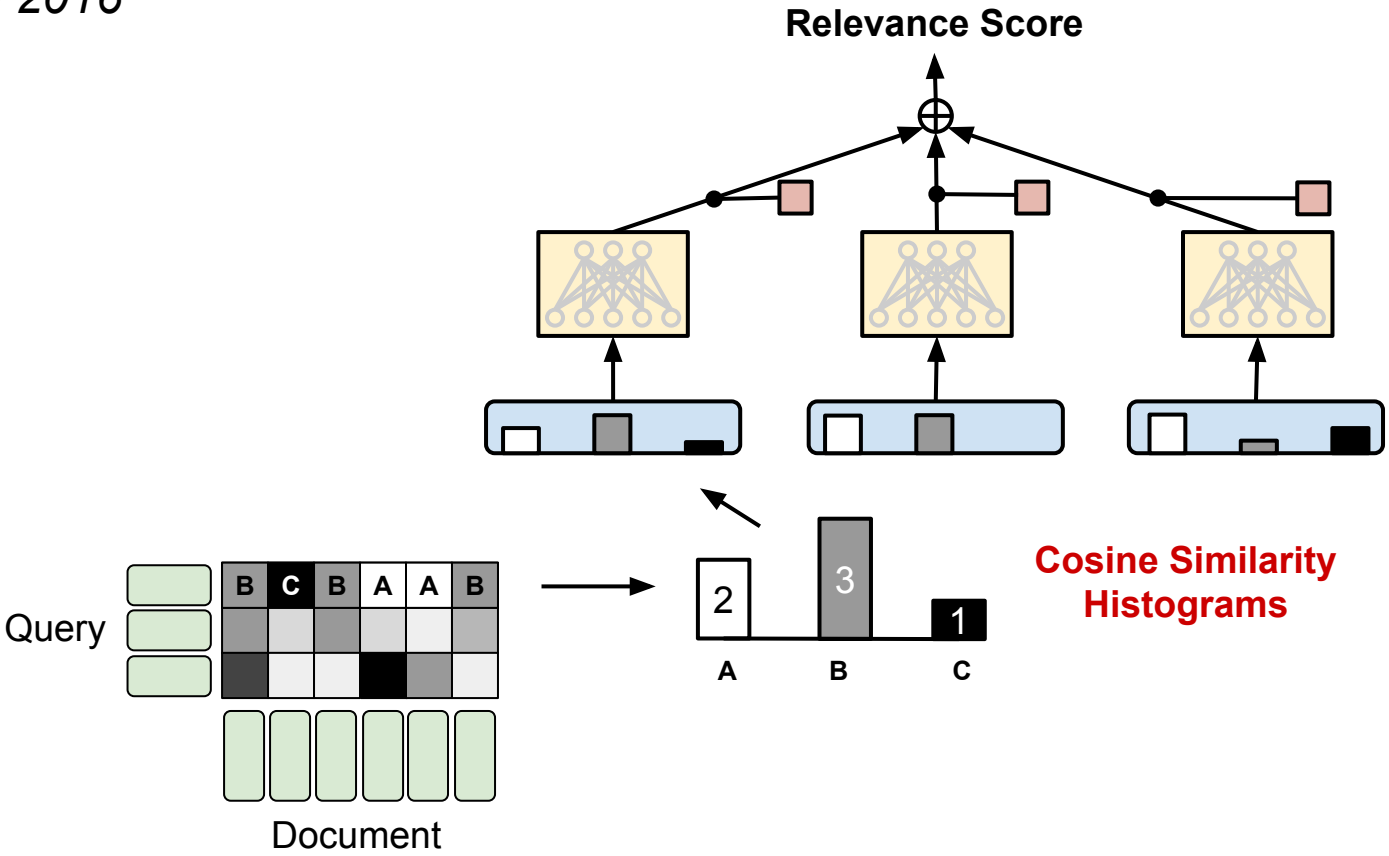
Deep Relevance Matching Model (DRMM)

Guo et al 2016

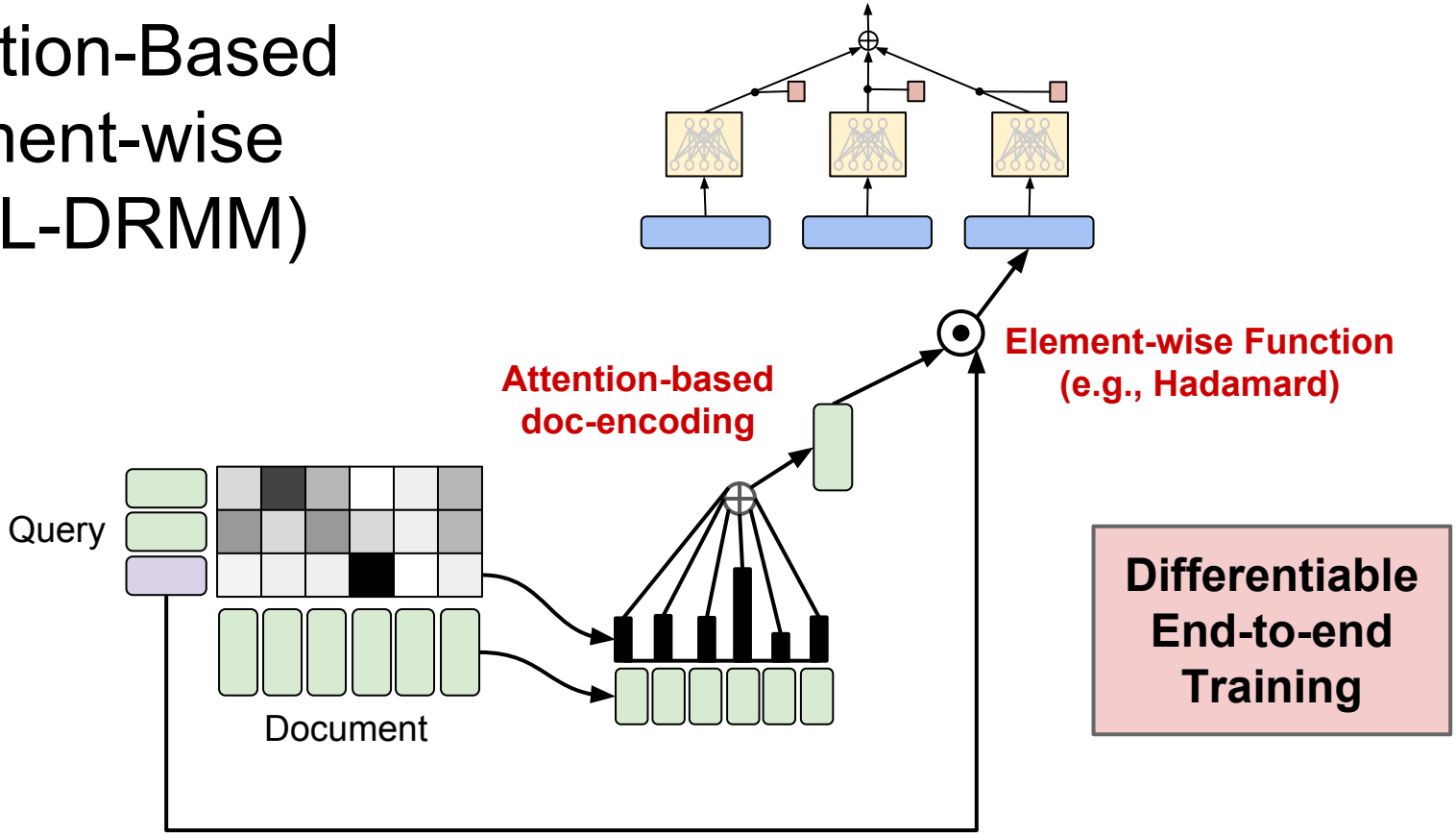


Deep Relevance Matching Model (DRMM)

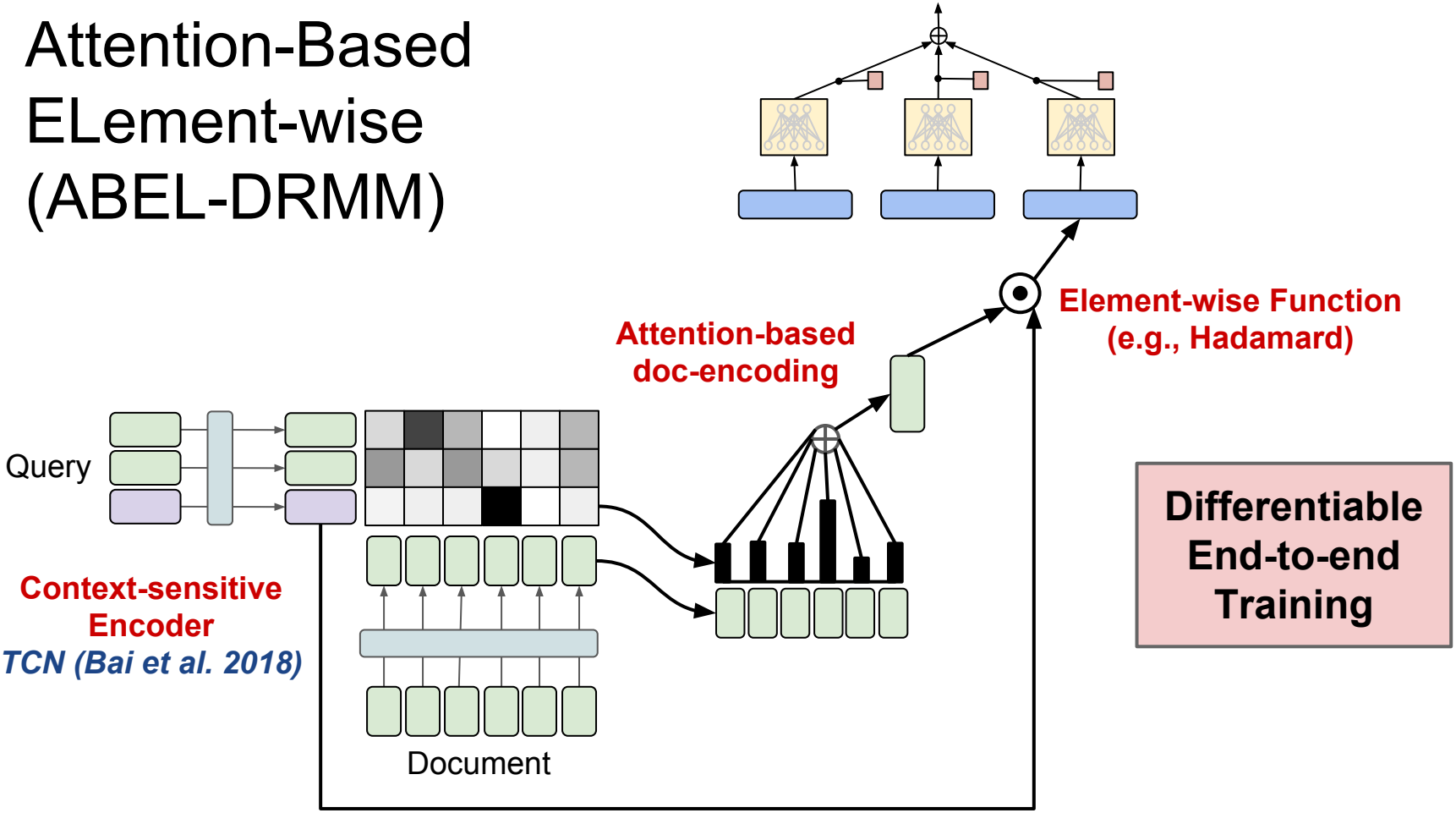
Guo et al 2016



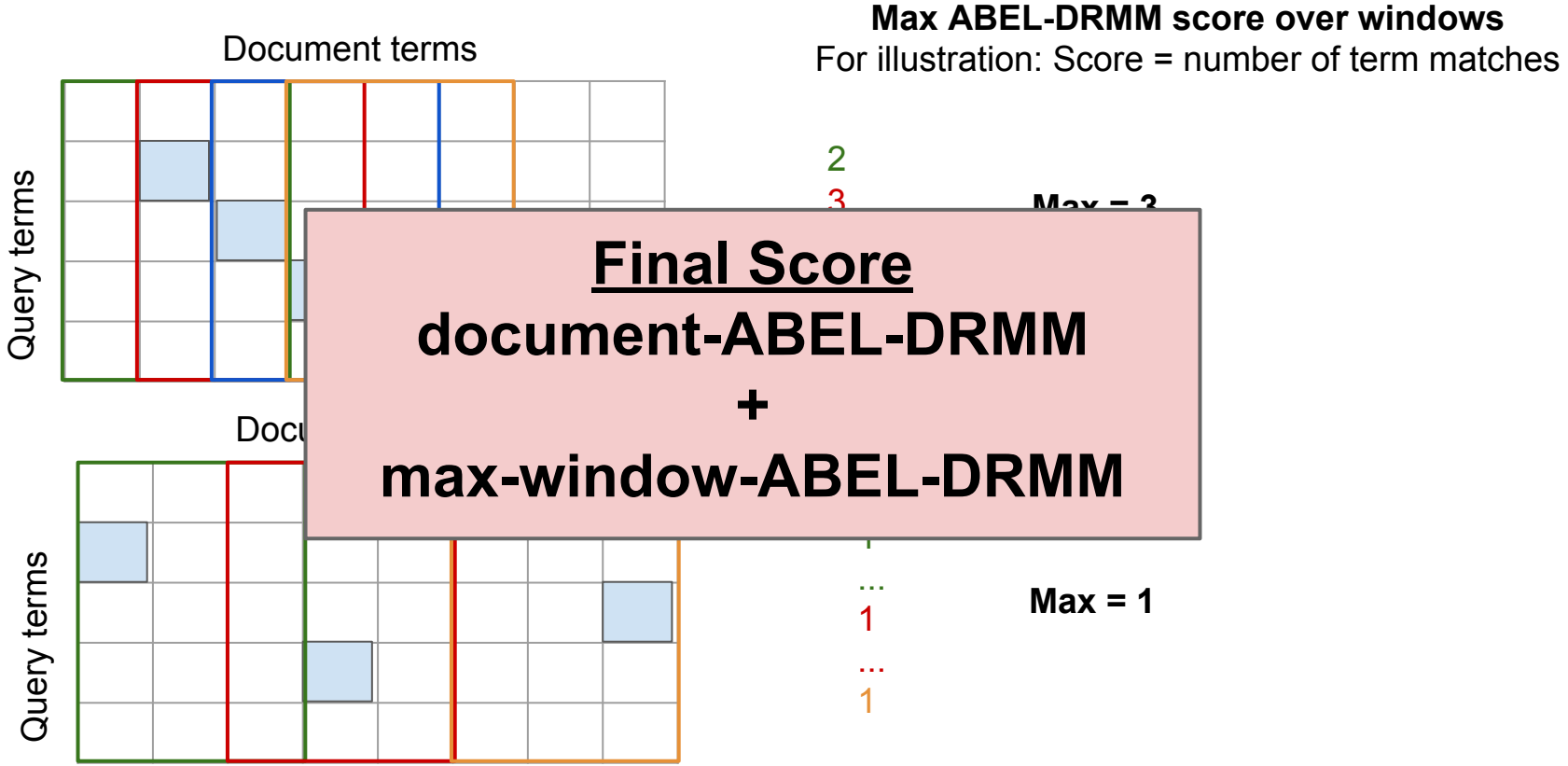
Attention-Based ELement-wise (ABEL-DRMM)

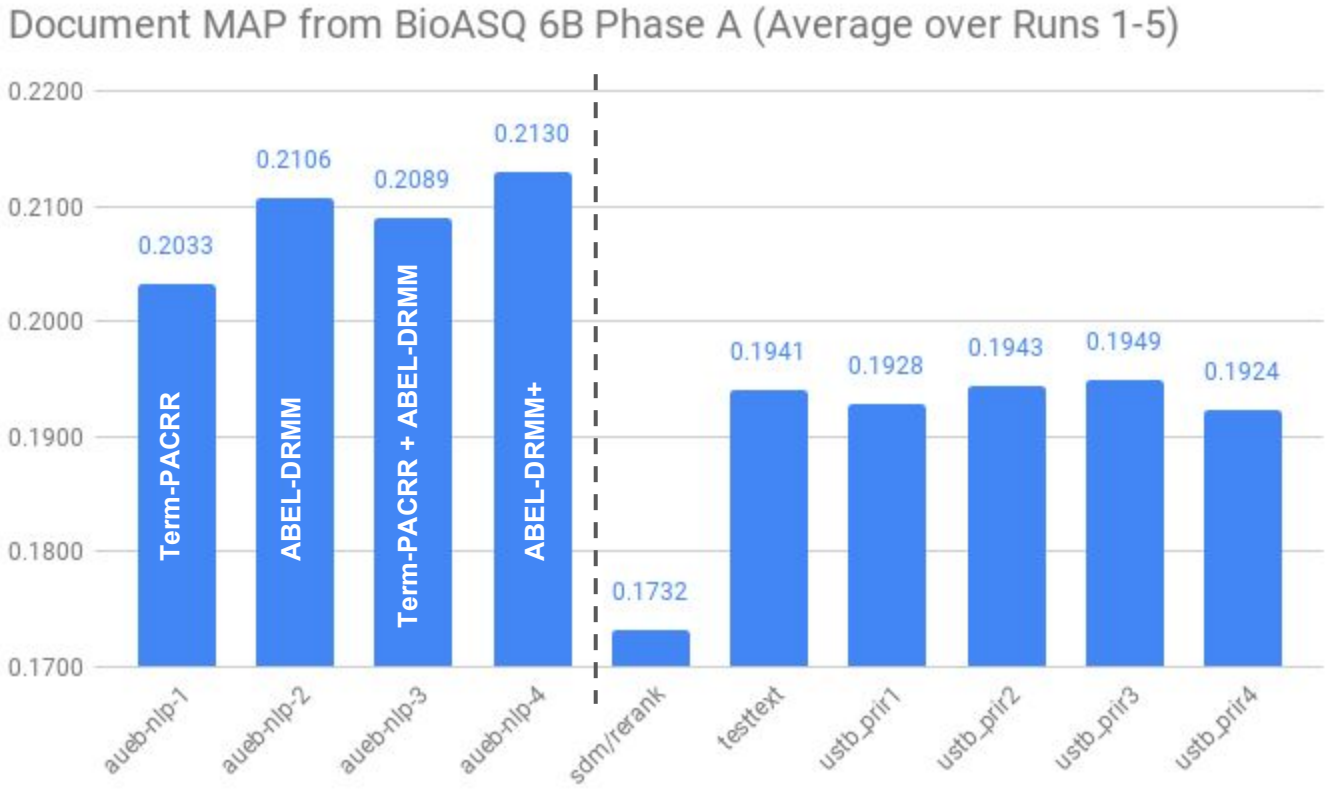


Attention-Based ELement-wise (ABEL-DRMM)



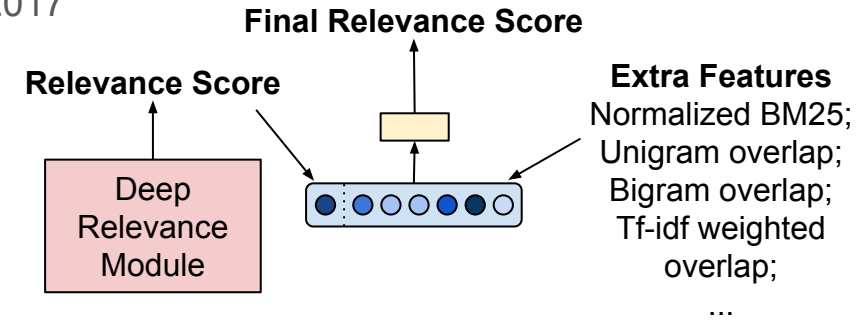
ABEL-DRMM+: Add density




Document MAP from BioASQ 6B Phase A (Average over Runs 1-5)
0.2200
0.2100
0.2000
0.1900
0.1800
0.1700
0.2033
0.2106
0.2089
0.2130
0.1732
0.1941
0.1928
0.1943
0.1949
0.1924
Term-PACRR
ABEL-DRMM
Term-PACRR + ABEL-DRMM
ABEL-DRMM+
aueb-nlp-1
aueb-nlp-2
aueb-nlp-3
aueb-nlp-4
sdm/rerank
testtext
ustb_prir1
ustb_prir2
ustb_prir3
ustb_prir4

Devil's in the details … (read the paper)

- Re-ranked top 100 documents from pubmed index using BM25 score
- System 1-3 are combinations of multiple runs (8)
- All models use final linear layer over
  - Severyn and Moschitti 2015, Mohan et al. 2017

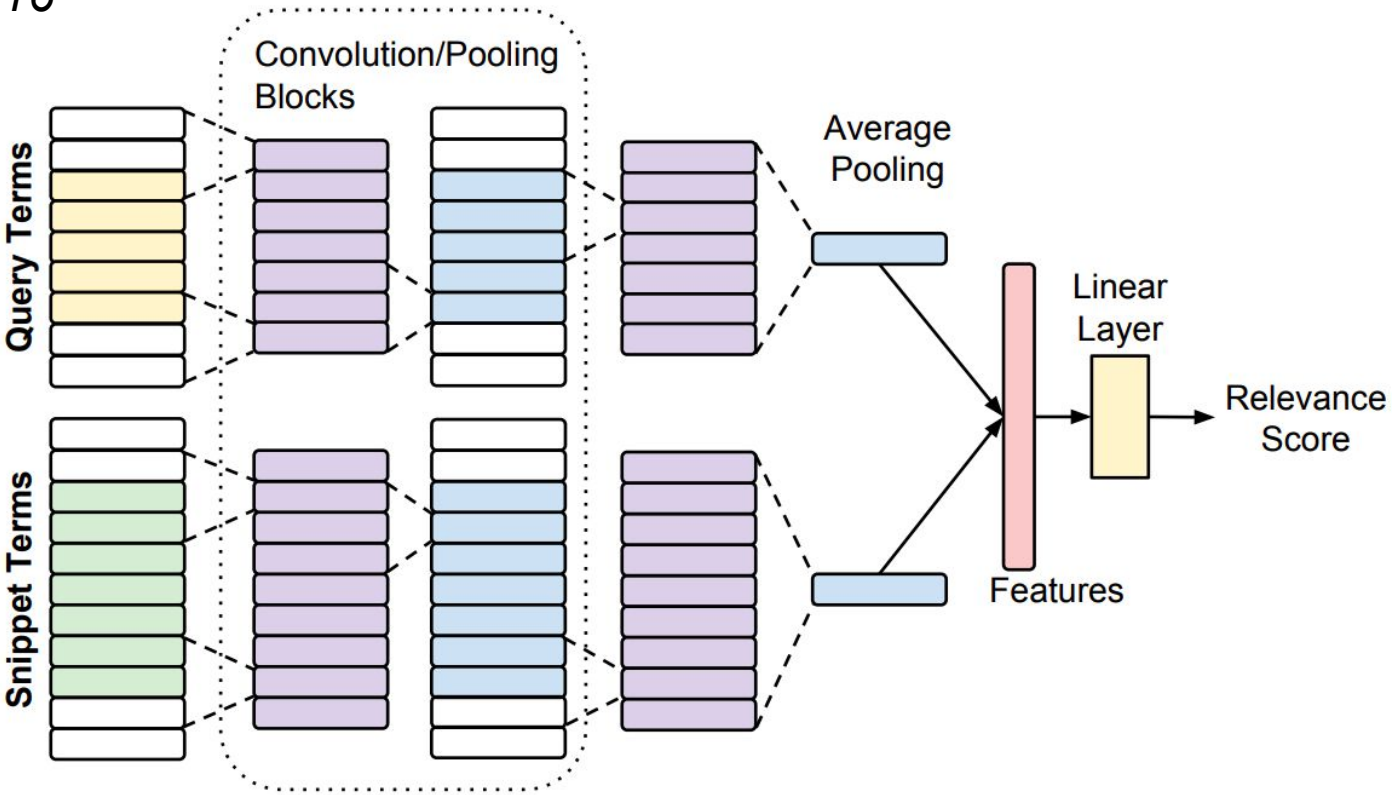


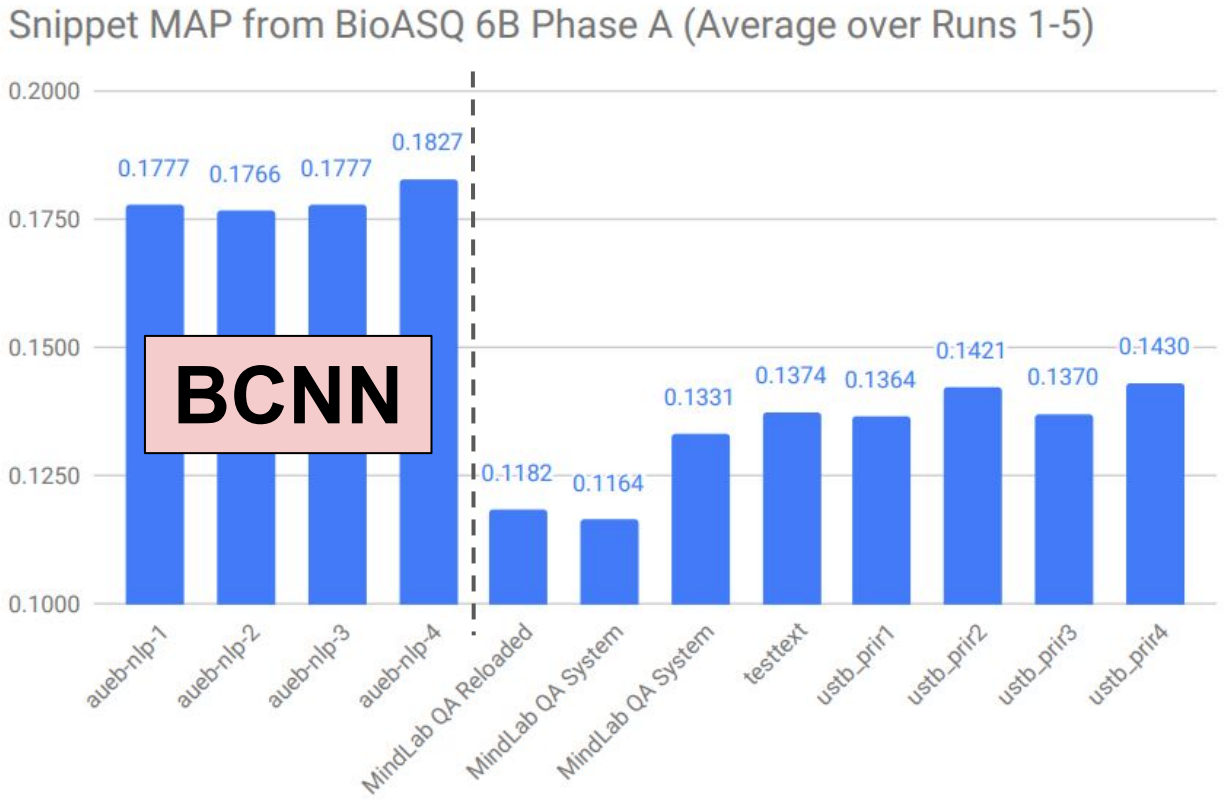
- Years 1-5 used to train models
- Pairwise training with neg. sampling

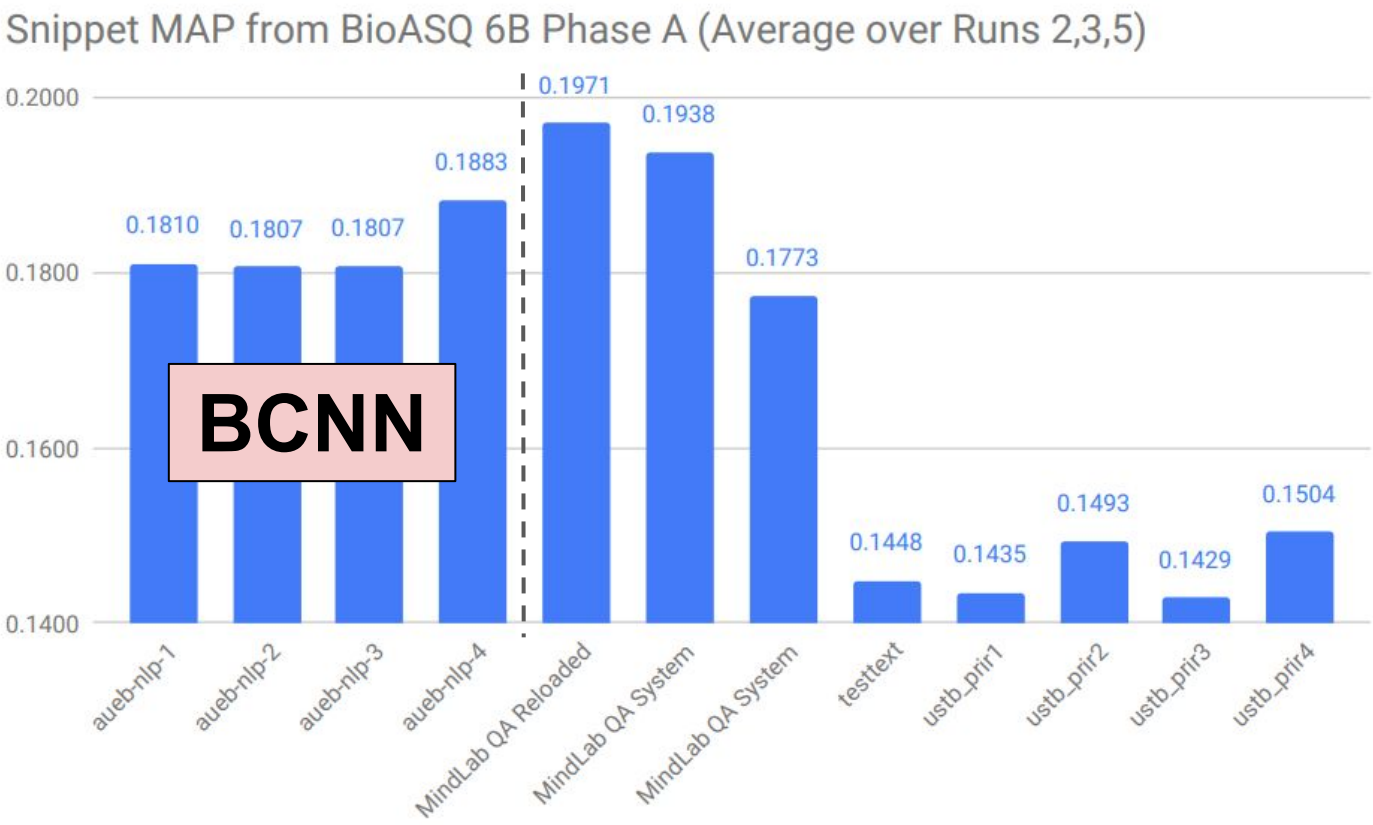
Snippet Ranking

Basic CNN Text Matching

Yin et al. 2016



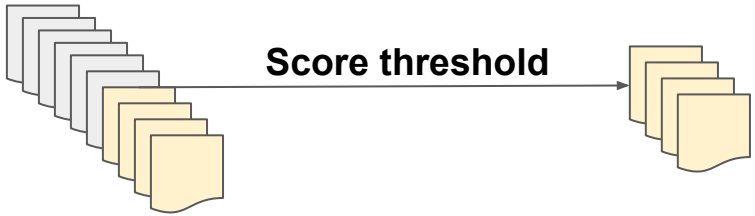

Snippet MAP from BioASQ 6B Phase A (Average over Runs 1-5)
0.2000
0.1750
0.1500
0.1250
0.1000
BCNN
0.1777
0.1766
0.1777
0.1827
0.1182
0.1164
0.1331
0.1374
0.1364
0.1421
0.1370
0.1430
aueb-nlp-1
aueb-nlp-2
aueb-nlp-3
aueb-nlp-4
MindLab QA Reloaded
MindLab QA System
MindLab QA System
testtext
ustb_prir1
ustb_prir2
ustb_prir3
ustb_prir4

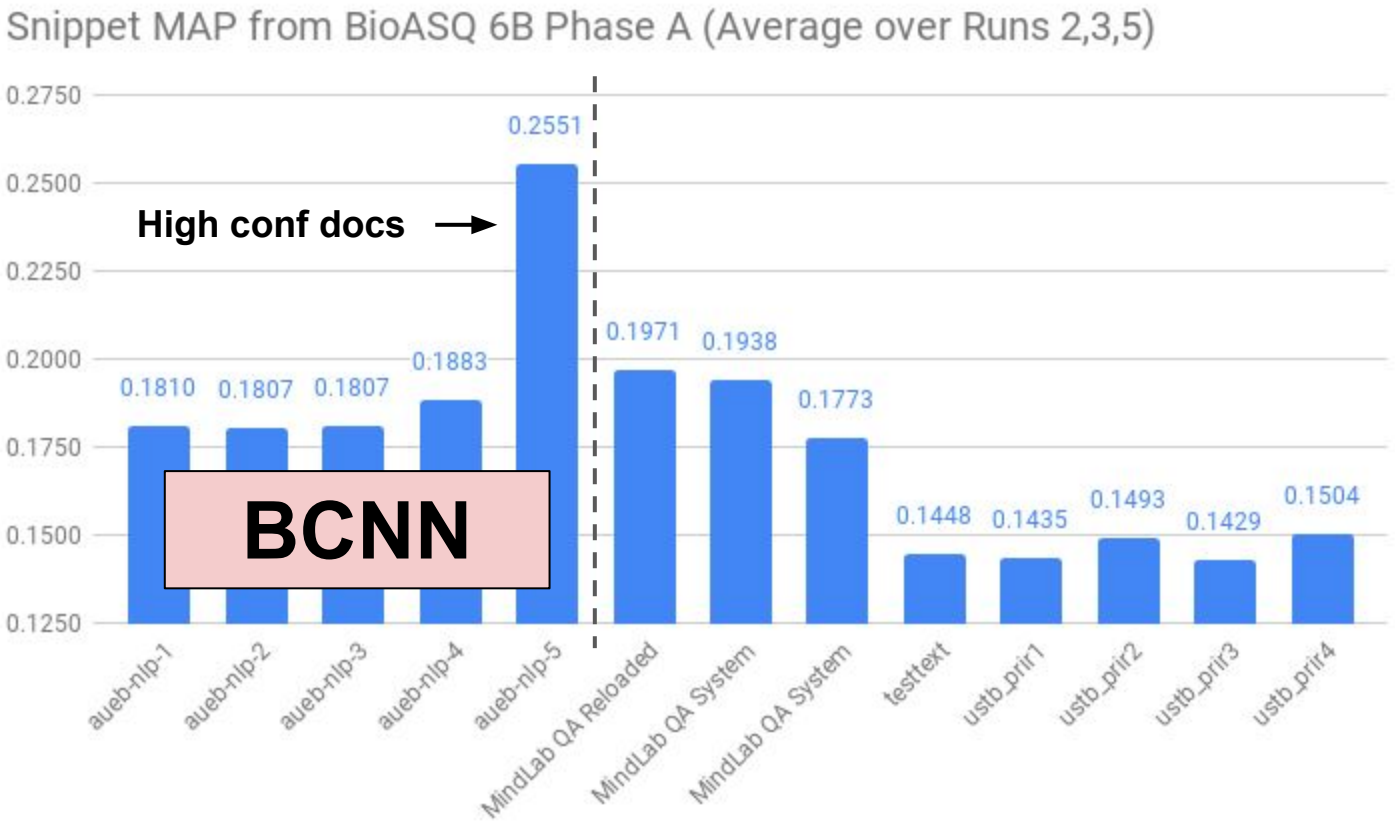

Snippet MAP from BioASQ 6B Phase A (Average over Runs 2,3,5)
0.2000
0.1800
0.1600
0.1400
0.1810
0.1807
0.1807
0.1883
0.1971
0.1938
0.1773
0.1448
0.1435
0.1493
0.1429
0.1504
BCNN
aueb-nlp-1
aueb-nlp-2
aueb-nlp-3
aueb-nlp-4
MindLab QA Reloaded
MindLab QA System
MindLab QA System
testtext
ustb_prir1
ustb_prir2
ustb_prir3
ustb_prir4

Training-testing mismatch

- Snippet extractor trained only on gold documents
- Learns to extract good snippets from good documents
- Not calibrated for non-relevant documents seen at test time

- Heuristic -- only extract snippets from high scoring docs
  - Often returns less than 10 documents




Snippet MAP from BioASQ 6B Phase A (Average over Runs 2,3,5)
0.2750
0.2500
0.2250
0.2000
0.1750
0.1500
0.1250
High conf docs
BCNN
0.1810
0.1807
0.1807
0.1883
0.2551
0.1971
0.1938
0.1773
0.1448
0.1435
0.1493
0.1429
0.1504
aueb-nlp-1
aueb-nlp-2
aueb-nlp-3
aueb-nlp-4
aueb-nlp-5
MindLab QA Reloaded
MindLab QA System
MindLab QA System
testtext
ustb_prir1
ustb_prir2
ustb_prir3
ustb_prir4

Conclusions

- Deep IR models are effective for ranking docs and snippets from Biomed data
  - Propose in paper: End-to-end training with extensions to DRMM
  - Need to incorporate deep IR scores with traditional features

- Snippet retrieval: incorporate document score
  - Heuristics go a long way
  - More principled methods likely to improve more

Thanks!

https://github.com/nlpaueb/aueb-bioasq6