One Framework to Run them All: A Unified Framework for Reading Comprehension Style Question Answering Systems

Preksha Nema
(CS15D201)
Advisors:
Mitesh M. Khapra
Balaraman Ravindran

List of Papers :

- B. Dhingra, H. Liu, Z. Yang, W. W. Cohen, and R. Salakhutdinov. **Gated-attention readers for text comprehension**. In Proceedings of the55th Annual Meeting of the Association for Computational Linguistics,ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers, pages 1832–1846, 2017. doi: 10.18653/v1/P17-1168.
- M. J. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi. **Bidirectional attention flow for machine comprehension**.CoRR, abs/1611.01603, 2016.
- C. Xiong, V. Zhong, and R. Socher. **Dynamic coattention networks forquestion answering**. CoRR, abs/1611.01604, 2016.
- W. Wang, N. Yang, F. Wei, B. Chang, and M. Zhou. **Gated self-matching networks for reading comprehension and question answering**. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1:Long Papers, pages 189–198, 2017.
- K. M. Hermann, T. Kocisk ́y, E. Grefenstette, L. Espeholt, W. Kay,M. Suleyman,and P. Blunsom. **Teaching machines to read and comprehend**. In Advances in Neural Information Processing Systems 28:Annual Conference on Neural Information Processing Systems 2015,December 7-12,2015,Montreal,Quebec,Canada, pages 1693–1701, 2015.

Reading Comprehension Style QA

PASSAGE

The role of teacher is often formal and ongoing, carried out at a school or other place of formal education. In many countries, a person who wishes to become a teacher must first obtain specified professional qualifications or credentials from a university or college. These professional qualifications may include the study of pedagogy, the science of teaching. Teachers, like other professionals, may have to continue their education after they qualify, a process known as continuing professional development. Teachers may use a **lesson plan** to facilitate student learning, providing a course of study which is called **the curriculum.**

Query: ***What is a course of study called?***

Answer: ***The curriculum***

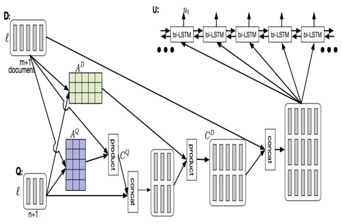
Query: **What can a teacher use to help students learn?**

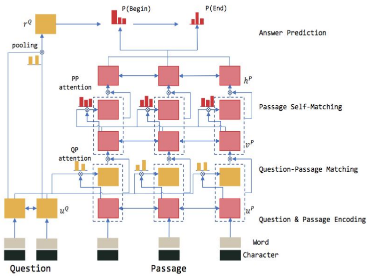
Answer: ***lesson plan***

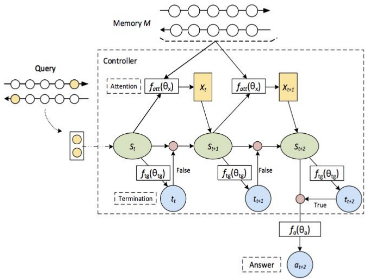
Current Datasets

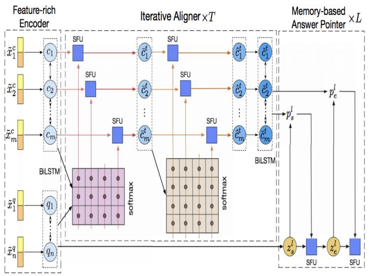
Datasets	Question Source	Formulation	Size
CNN/Dailymail[9]	summary + Cloze	fill in single entity	1.4M
CBT[10]	Cloze	fill in single entity	688K
SQuAD[26]	crowdsourced	spans in passages	100K
RACE[17]	Standardized tests	Multiple choice	80K
MS-Marco[23]	query logs	Human generated	+100K
Trivia-QA[14]	query logs	answer span	95K
MCTest[27]	crowdsourced	multiple choice	2640
WDW[24]	Cloze	fill in single entity	+185K

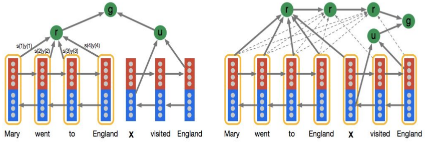
Wave of Neural RCQA models:





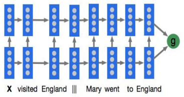




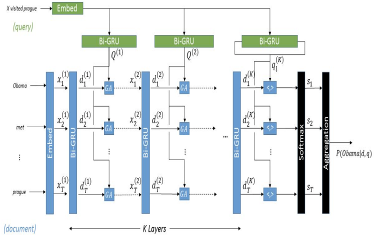


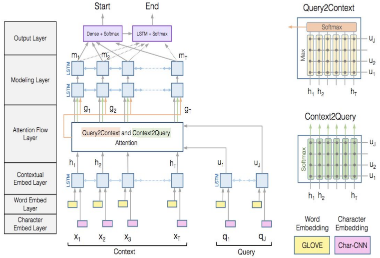
(a) Attentive Reader.

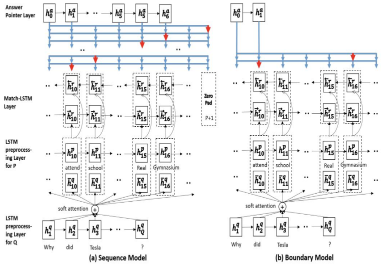
(b) Impatient Reader.



(c) A two layer Deep LSTM Reader with the question encoded before the document.





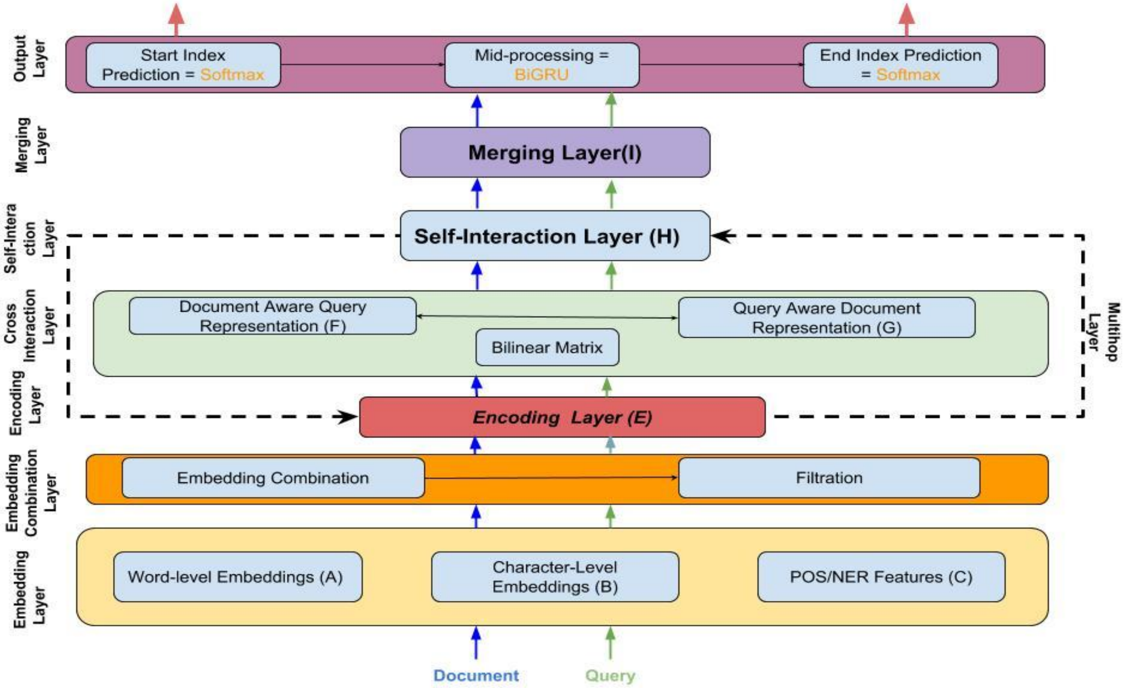


The Problem:

- Too many models..
- Each work fine-tunes a different module to beat the state-of-the-art performance

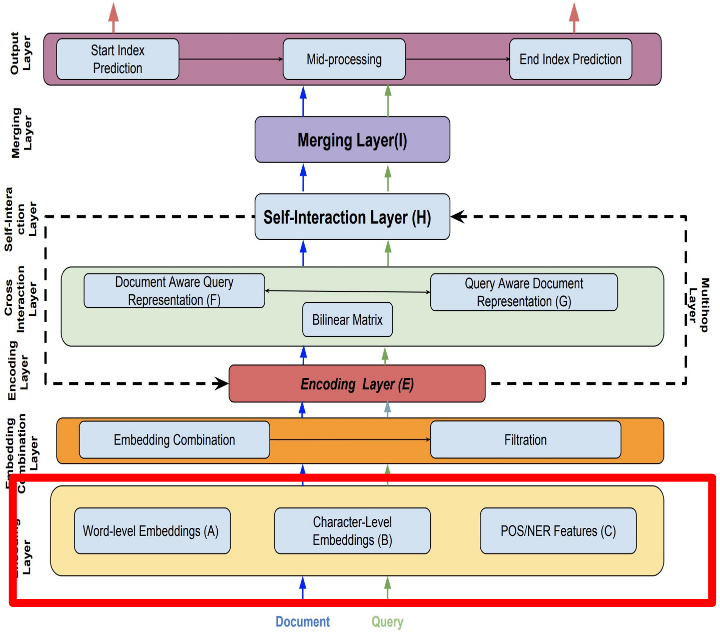
Need a common platform to analyze all the works in an apples-to-apples fashion

Proposed RCQA Framework



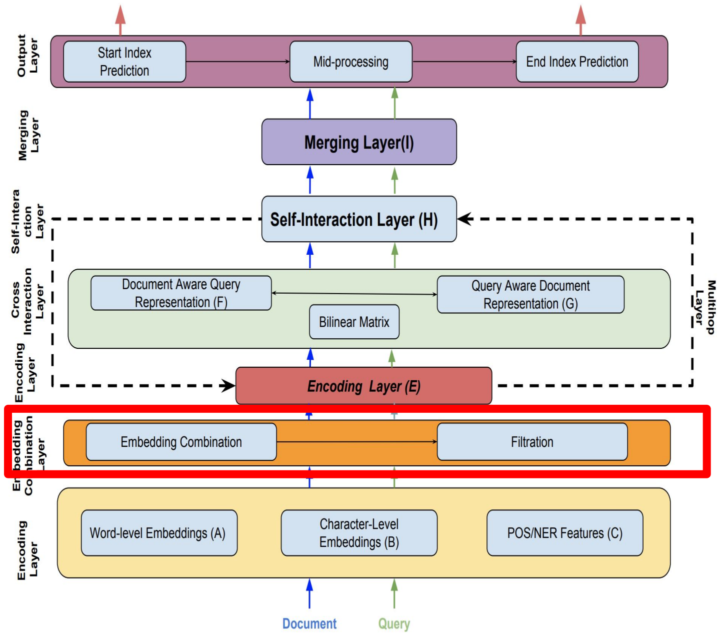
Embedding Layer

- **Objective:** Each word in query and passage could be represented as a *d*-dimensional vector
- **Modules:**
  - Word Embedding Module
  - Character Embedding Module
  - POS/NE/TF-IDF Features Module
- *Existing Works using this layer:*
  - **ALL**



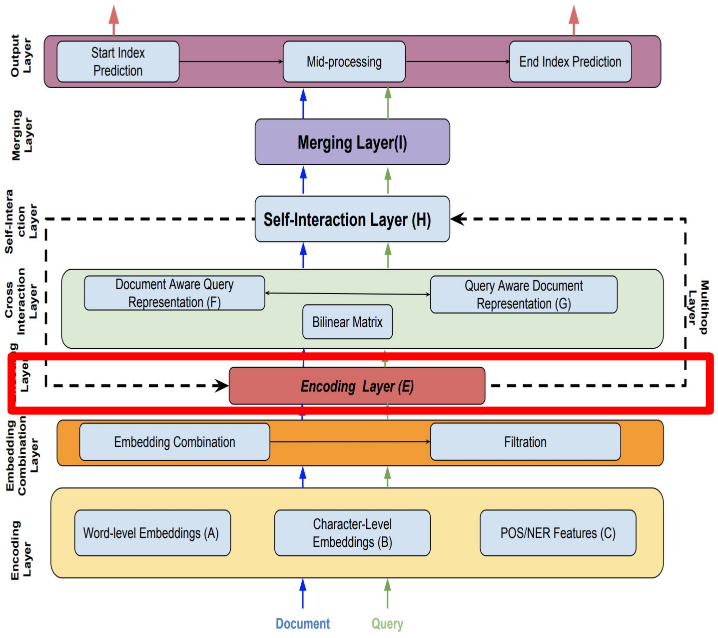
Embedding Aggregation Layer

- **Objective:** If more than one module is used in the previous layer, this layer combines the different vector representations of the same word
- **Methods for Aggregation**
  - Concatenation
  - Algebraic Sum
  - Scalar Gate
  - Vector Gate
- *Existing Works using this layer:*
  - [6], [37], [29], [8], [32]



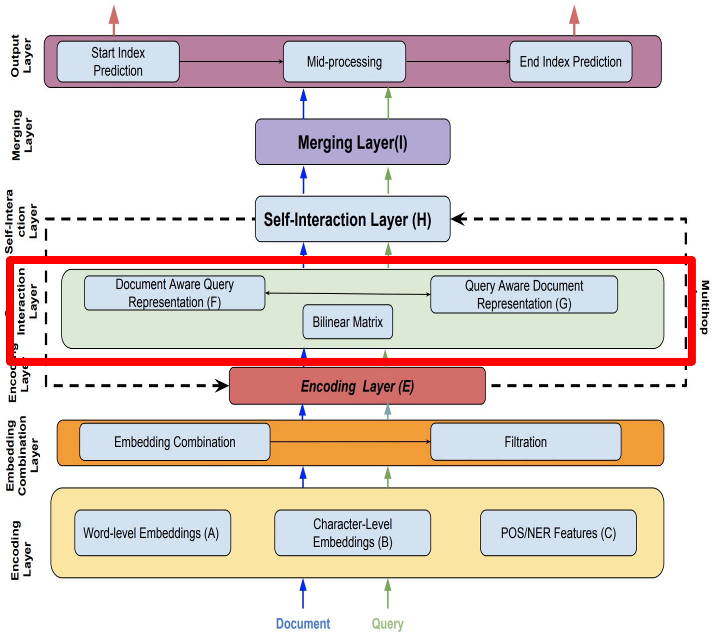
Encoding Layer

- **Objective:** Contextual information for each of the word in the document and query is encoded using RNNs:
- **Methods for Encoding:**
  - Bidirectional LSTMs
  - Stacked Bidirectional LSTMs
  - Deep LSTMs with skip connections across layers
- *Exisiting Works Using this layer*
  - **ALL**



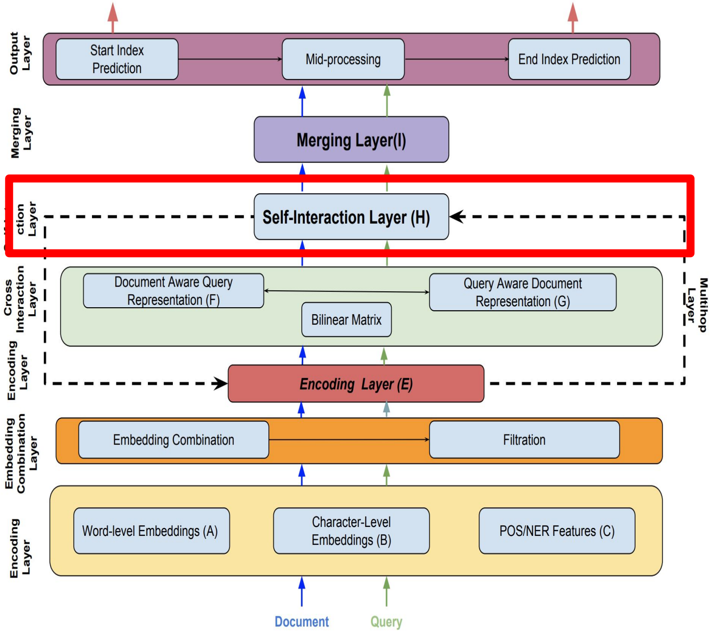
Cross Interaction Layer

- **Objective:** Captures the interaction between document and query, either at word level or sentence-level
- **Modules**:
    - Document Aware Query Representation: Read document in light of the query
    - Query Aware Document Representation: Read query in light of the document
    - Bilinear Matrix: Captures word-word interaction between query and document
- *ExistingWorks Using this layer*
    - **ALL**



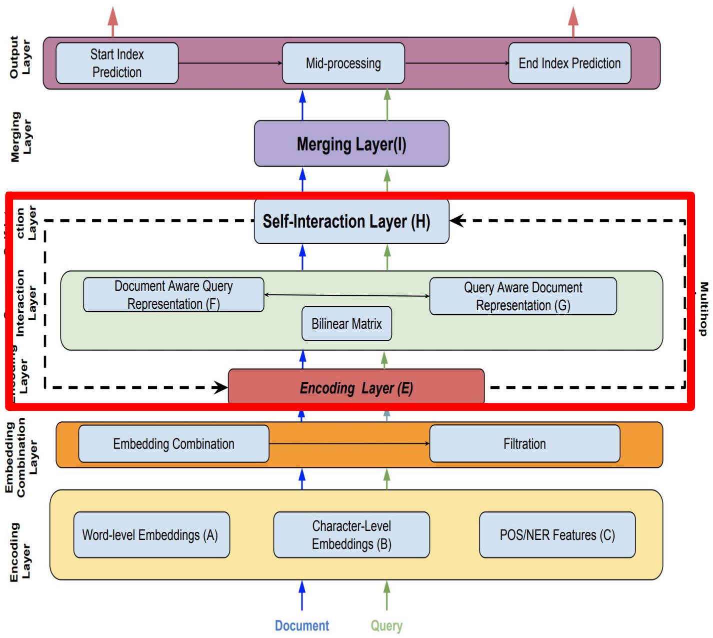
Self Interaction Layer

- **Motivation:** To infer some of the answer, it is necessary to infer from more than one sentence in a document (might not be continuous)
- **Objective**: Capture long term dependency by computing word-word interaction of every document word pair
- **Module**:
  - Bilinear Matrix
- *Existing Works Using this layer*
  - [8],[29],[12],[35]



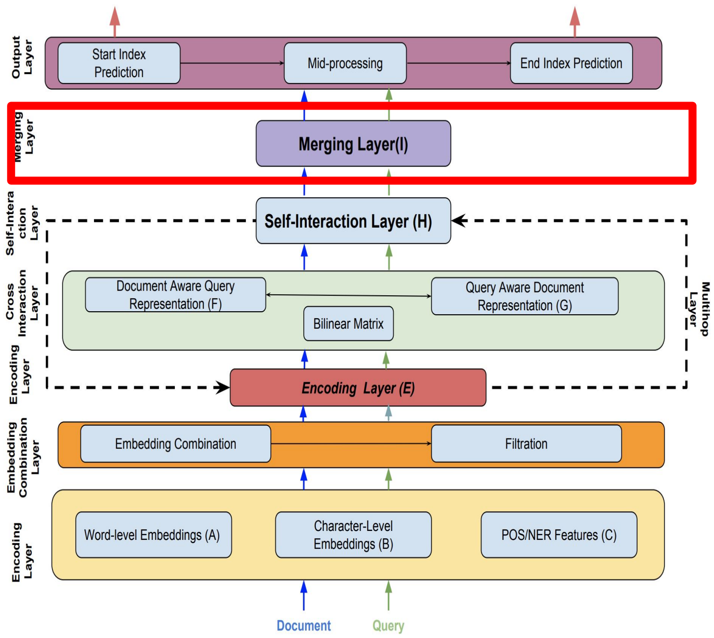
Multihop Layer

- **Motivation:** Documents need to be re-read a couple of times to be confident about the answer
- **Objective**: Decides on how many times document needs to be re-read and ensures information flow is consistent from self-interaction layer to encoding layer
- **Module**:
  - Bilinear Matrix
- *Existing Works Using this layer*
  - [30], [6], [31]



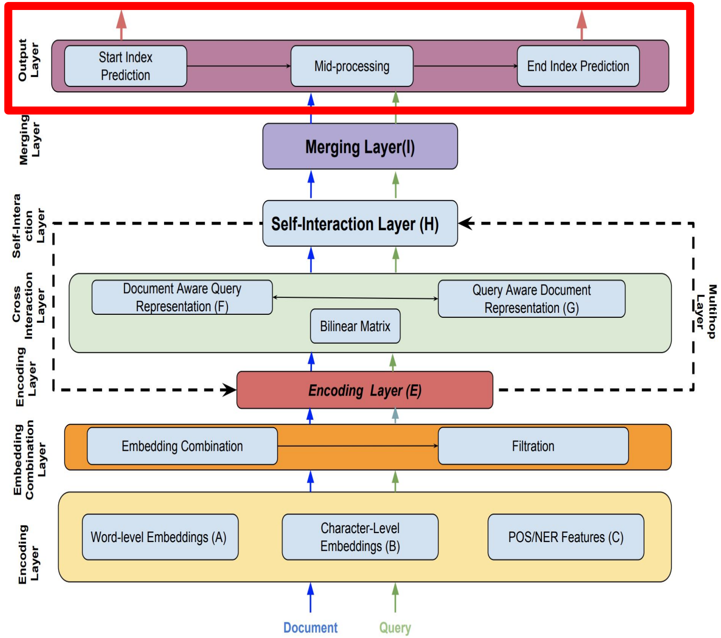
Merging Layer

- **Objective**: Merge the representations obtained using cross-interaction and self-interaction layers
- **Module**:
  - Bilinear Matrix
- *Existing Works Using this layer*
  - **ALL**



Ouput Layer

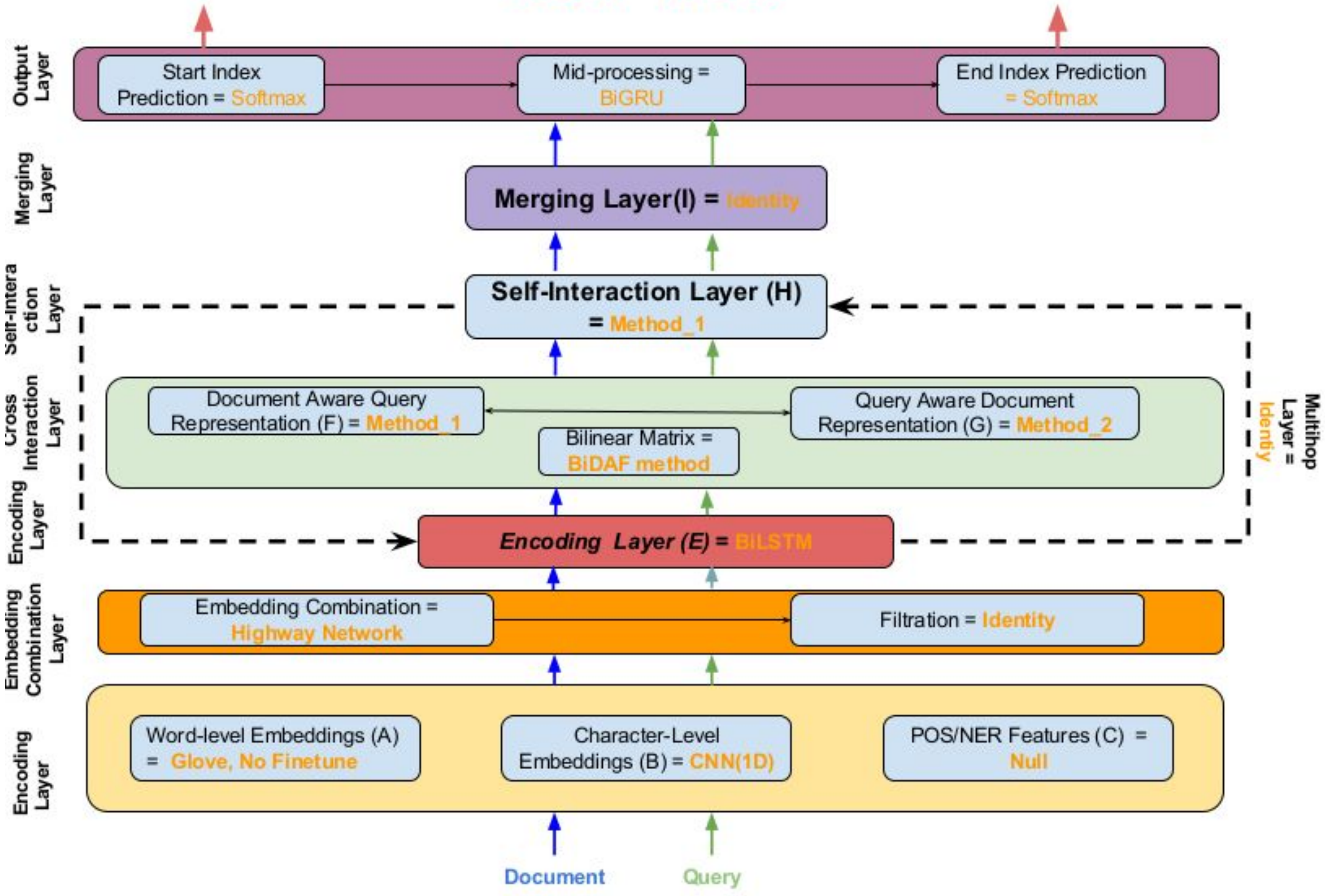
- **Objective**: Uses the representation obtained from merging layer to predict the correct span or generate the answer in natural language
- *Existing Works Using this layer*
  - **ALL**



Conclusion

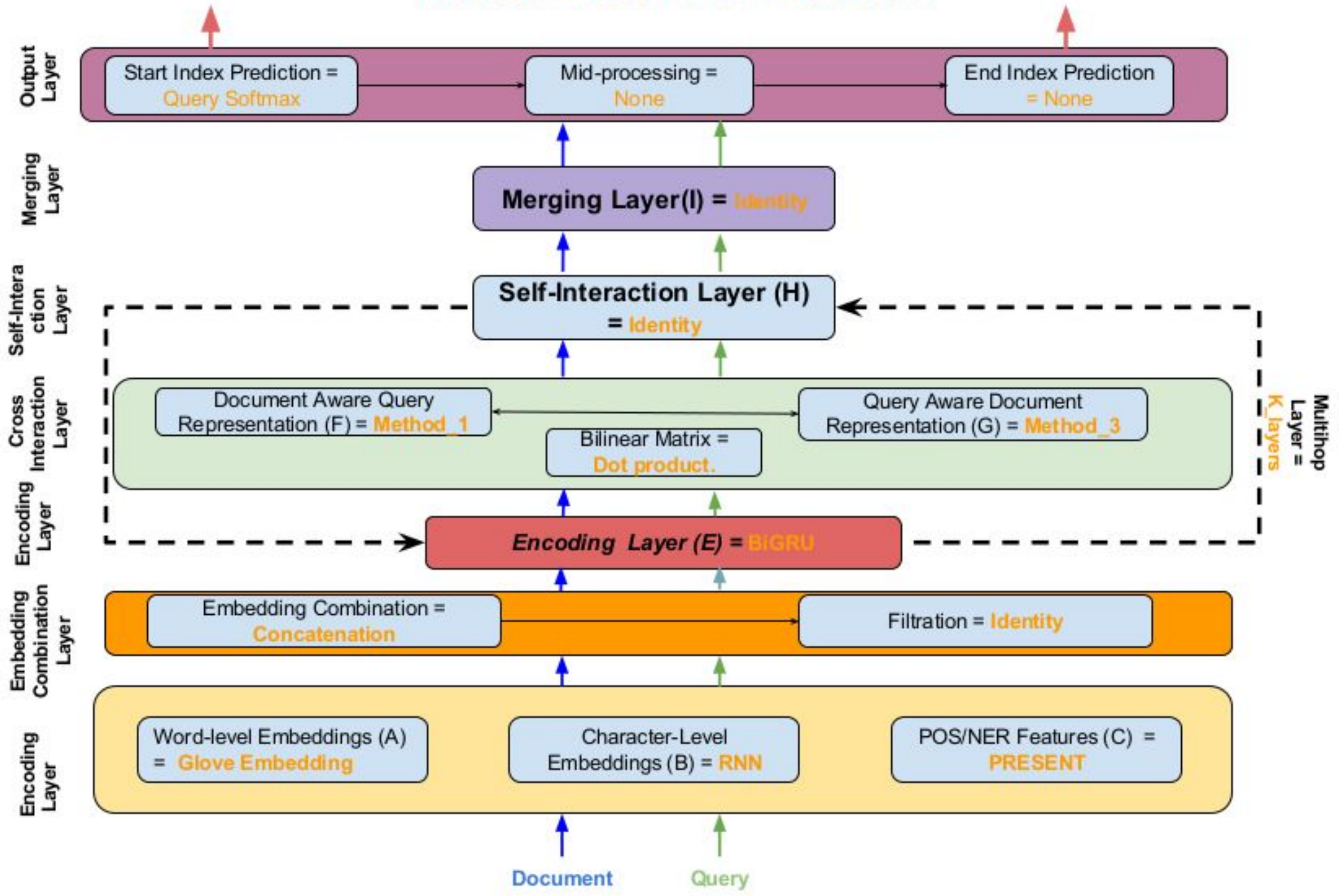
- This unified framework will help us understand, the plethora of concepts introduced in RCQA models, better
- This framework would help in comparing the efficacy of different modules/models

BiDAF Model

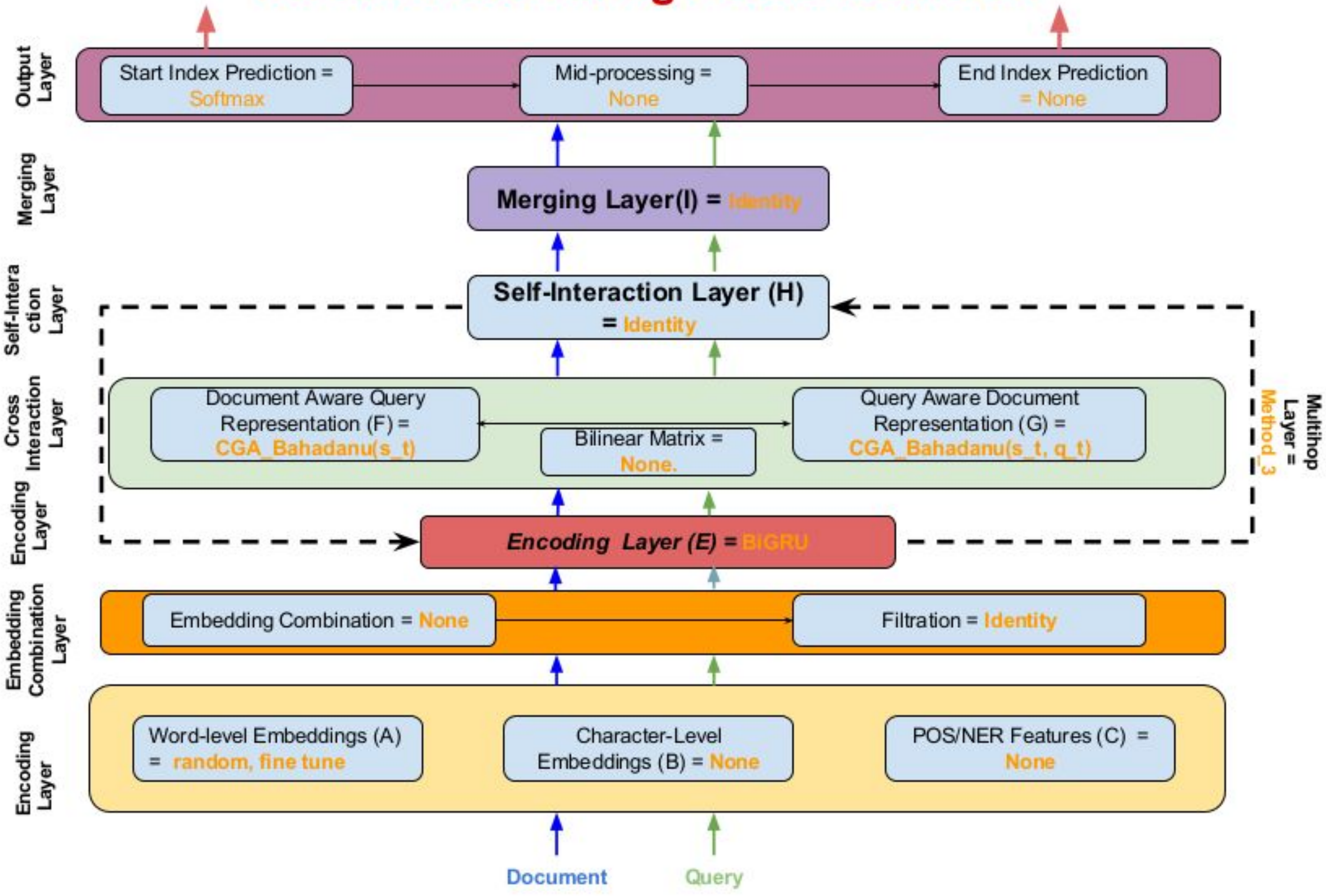


Output Layer
Start Index Prediction = Softmax
Mid-processing = BiGRU
End Index Prediction = Softmax
Merging Layer
Merging Layer(I) = Identity
Self-Interaction Layer
Self-Interaction Layer (H) = Method_1
Cross Interaction Layer
Document Aware Query Representation (F) = Method_1
Query Aware Document Representation (G) = Method_2
Bilinear Matrix = BiDAF method
Multihop Layer = Identiy
Encoding Layer
Encoding Layer (E) = BiLSTM
Embedding Combination Layer
Embedding Combination = Highway Network
Filtration = Identity
Encoding Layer
Word-level Embeddings (A) = Glove, No Finetune
Character-Level Embeddings (B) = CNN(1D)
POS/NER Features (C) = Null
Document
Query

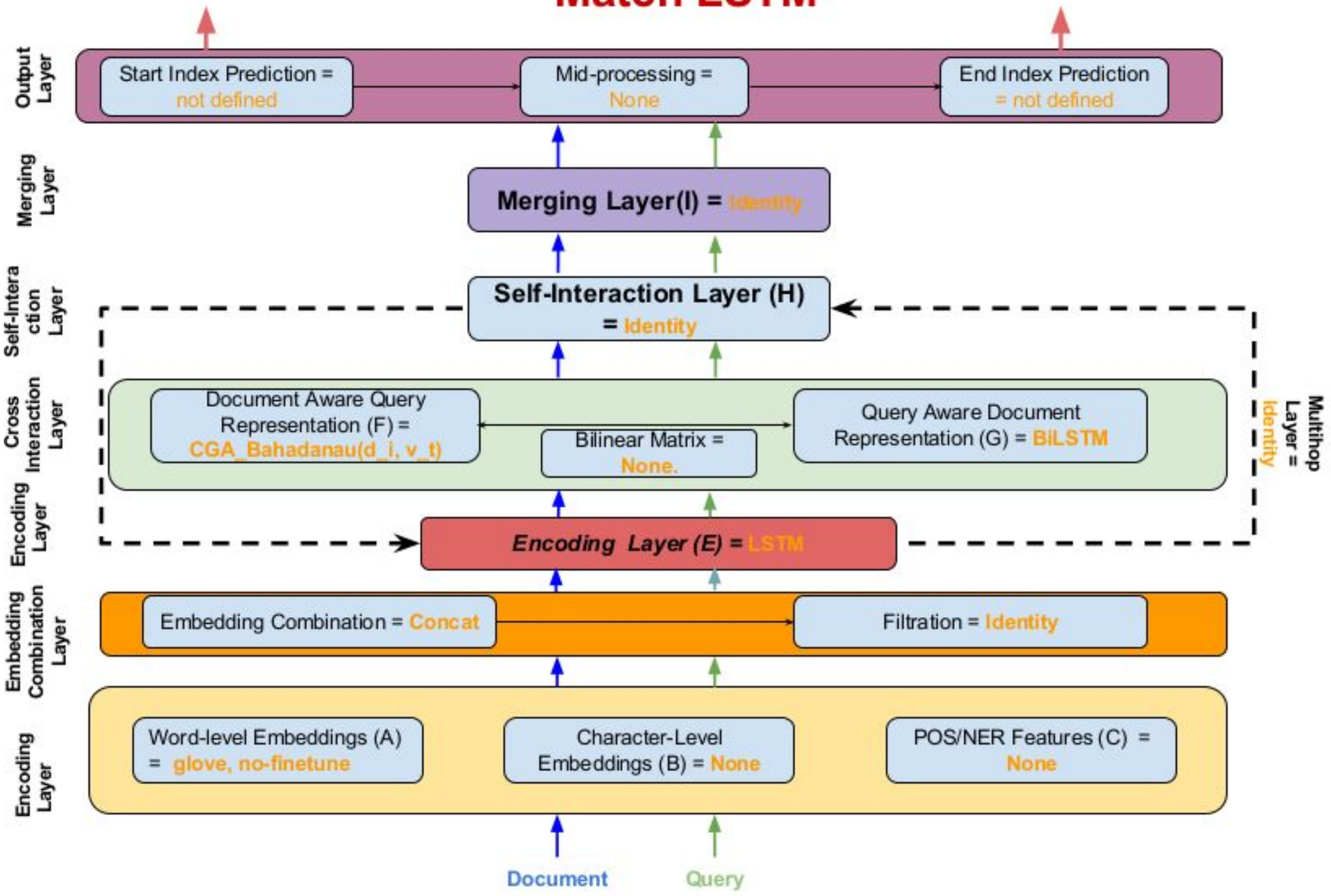
Gated Attention Reader



Iterative Alternating Neural Attention

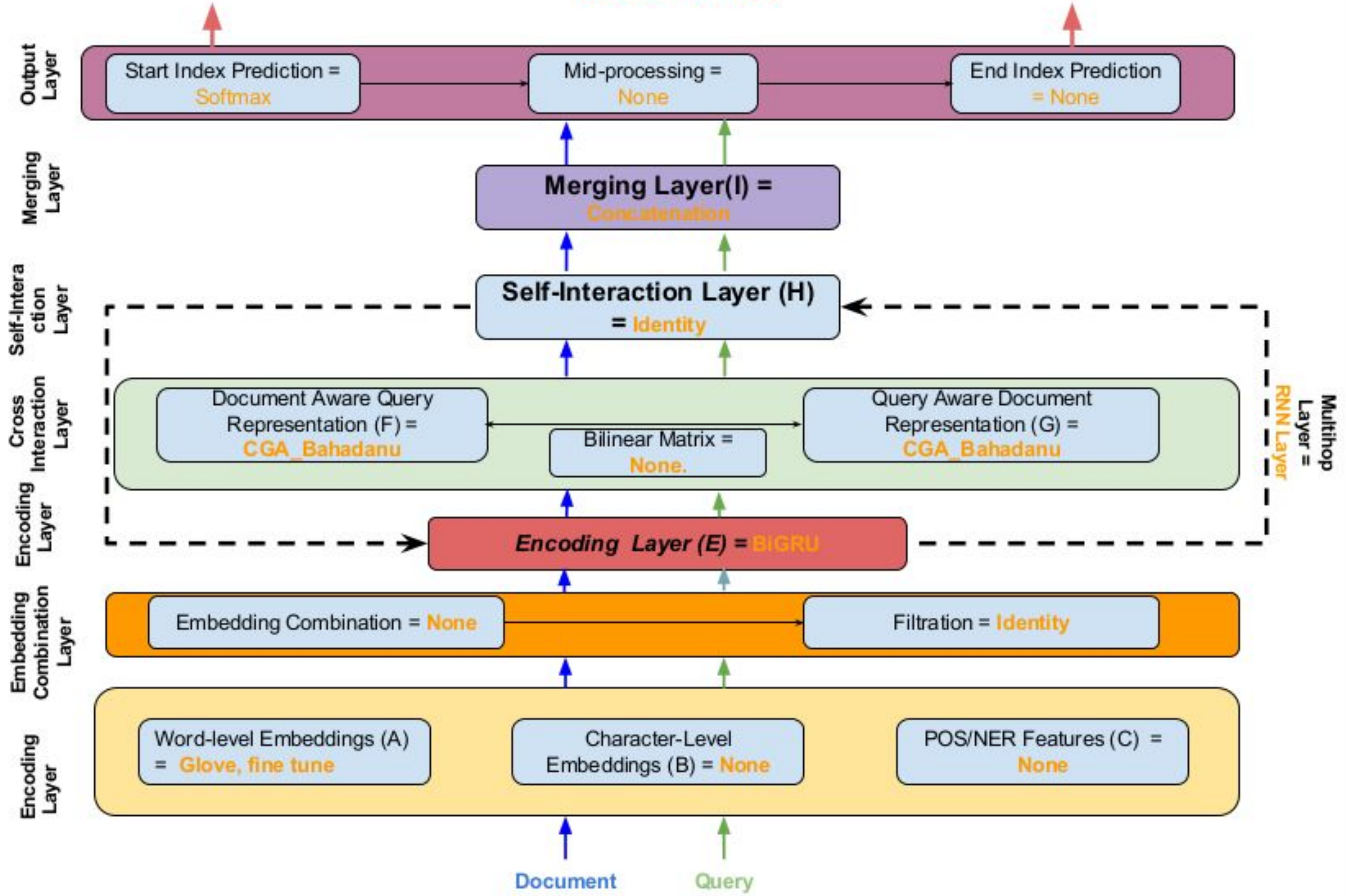


Match LSTM



Output Layer
Start Index Prediction = not defined
Mid-processing = None
End Index Prediction = not defined
Merging Layer
Merging Layer(I) = Identity
Self-Interaction Layer
Self-Interaction Layer (H) = Identity
Cross Interaction Layer
Document Aware Query Representation (F) = CGA_Bahadanau(d_i, v_t)
Bilinear Matrix = None.
Query Aware Document Representation (G) = BiLSTM
Multihop Layer = Identity
Encoding Layer
Encoding Layer (E) = LSTM
Embedding Combination Layer
Embedding Combination = Concat
Filtration = Identity
Encoding Layer
Word-level Embeddings (A) = glove, no-finetune
Character-Level Embeddings (B) = None
POS/NER Features (C) = None
Document
Query

Reasonet



Output Layer
Start Index Prediction = Softmax
Mid-processing = None
End Index Prediction = None
Merging Layer
Merging Layer(I) = Concatenation
Self-Interaction Layer
Self-Interaction Layer (H) = Identity
Cross Interaction Layer
Document Aware Query Representation (F) = CGA_Bahadanu
Bilinear Matrix = None.
Query Aware Document Representation (G) = CGA_Bahadanu
Multihop Layer = RNN Layer
Encoding Layer
Encoding Layer (E) = BiGRU
Embedding Combination Layer
Embedding Combination = None
Filtration = Identity
Encoding Layer
Word-level Embeddings (A) = Glove, fine tune
Character-Level Embeddings (B) = None
POS/NER Features (C) = None
Document
Query

Questions ?

References:

[1] Learning Answer-Entailing Structures for Machine Comprehension ,July 2015

[2] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate.CoRR, abs/1409.0473, 2014

[3] D. Chen, J. Bolton, and C. D. Manning. A thorough examination of the cnn/daily mail reading comprehension task.CoRR, abs/1606.02858, 2016

[4] K. Cho, B. van Merrienboer, C ̧ . G ̈ul ̧cehre, F. Bougares, H. Schwenk, andY. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation.CoRR, abs/1406.1078, 2014.

[5] Y. Cui, Z. Chen, S. Wei, S. Wang, T. Liu, and G. Hu. Attention-over-attention neural networks for reading comprehension. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1:Long Papers, pages 593–602, 2017. doi: 10.18653/v1/P17-1055.

[6] B. Dhingra, H. Liu, Z. Yang, W. W. Cohen, and R. Salakhutdinov.Gated-attention readers for text comprehension. InProceedings of the55th Annual Meeting of the Association for Computational Linguistics,ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers, pages 1832–1846, 2017. doi: 10.18653/v1/P17-1168.

References:

[7] Y. Gong and S. R. Bowman. Ruminating reader: Reasoning with gated multi-hop attention.CoRR, abs/1704.07415, 2017.

[8] N. L. C. Group and M. R. Asia. R-net: Machine reading comprehension with self-matching networks.ACL, abs/1606.02245, 2017.

[9] K. M. Hermann, T. Kociský, E. Grefenstette, L. Espeholt, W. Kay,M. Suleyman,and P. Blunsom.Teaching machines to read and comprehend. In Advances in Neural Information Processing Systems 28:Annual Conference on Neural Information Processing Systems 2015,December 7-12,2015,Montreal,Quebec,Canada, pages 1693–1701, 2015.

[10] F. Hill, A. Bordes, S. Chopra, and J. Weston. The goldilocks principle:Reading children's books with explicit memory representations.CoRR,abs/1511.02301, 2015. [11] L. Hirschman, M. Light, E. Breck, and J. D. Burger. Deep read: A reading comprehension system. In Proceedings of the 37th Annual Meeting of ththe 19 Association for Computational Linguistics on Computational Linguistics,ACL '99, pages 325–332, Stroudsburg, PA, USA, 1999. Association for Computational Linguistics.

[12] M. Hu, Y. Peng, and X. Qiu. Mnemonic reader for machine comprehension.CoRR, abs/1705.02798, 2017.

[13] H. Huang, C. Zhu, Y. Shen, and W. Chen.Fusionnet: Fusing viafully-aware attention with application to machine comprehension.CoRR,abs/1711.07341, 2017.

References:

[14] M. Joshi, E. Choi, D. S. Weld, and L. Zettlemoyer. Triviaqa: A large scaled supervised challenge dataset for reading comprehension. InACL(1), pages 1601–1611. Association for Computational Linguistics, 2017.

[15] R. Kadlec, M. Schmid, O. Bajgar, and J. Kleindienst. Text understanding with the attention sum reader network. In ACL (1). The Association for Computer Linguistics, 2016.

[16] R. Kadlec, M. Schmid, O. Bajgar, and J. Kleindienst. Text understanding with the attention sum reader network. InProceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers, 2016.

[17] G. Lai, Q. Xie, H. Liu, Y. Yang, and E. H. Hovy. RACE: large-scale reading comprehension dataset from examinations. InProceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017, pages 785–794, 2017.

[18] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit.In Association for Computational Linguistics (ACL) System Demonstrations, pages 55–60, 2014.

[19] Y. Miyamoto and K. Cho. Gated word-character recurrent language model.In EMNLP, pages 1992–1997. The Association for Computational Linguistics, 2016.

References:

[20] K. Narasimhan and R. Barzilay. Machine comprehension with discourse relations. pages 1253–1262, 01 2015.

[21] H. T. Ng, L. H. Teo, J. Lai, and J. L. P. Kwan. A machine learning approach to answering questions for reading comprehension tests. InIn Proceedings of EMNLP/VLC-2000 at ACL-2000, 2000.

[22] T. Nguyen, M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder,and L. Deng. MS MARCO: A human generated machine reading comprehension dataset. InProceedings of the Workshop on Cognitive Computation: Integrating neural and symbolic approaches 2016 co-located with the 30th Annual Conference on Neural Information Processing Systems NIPS 2016), Barcelona, Spain, December 9, 2016., 2016.

[23] T. Nguyen, M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, andL. Deng. Ms marco: A human generated machine reading comprehension dataset. November 2016.

[24] T. Onishi, H. Wang, M. Bansal, K. Gimpel, and D. A. McAllester. Who did what: A large-scale person-centered cloze dataset. InProceedings ofthe 2016 Conference on Empirical Methods in Natural Language Processing,EMNLP 2016, Austin, Texas, USA, November 1-4, 2016, pages 2230–2235,2016.

[25] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors forward representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, 2014.

References:

[26] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang. Squad: 100,000+ questions for machine comprehension of text. In EMNLP, pages 2383–2392. The Association for Computational Linguistics, 2016.

[27] M. Richardson, C. J. C. Burges, and E. Renshaw. Mctest: A challenge dataset for the open-domain machine comprehension of text.

[28] E. Riloff and M. Thelen. A rule-based question answering system for reading comprehension tests. In Proceedings of the 2000 ANLP/NAACL Workshop on Reading Comprehension Tests As Evaluation for Computer-based Language Understanding Systems

[29] M. J. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi. Bidirectional attention flow for machine comprehension.CoRR, abs/1611.01603, 2016.

[30] Y. Shen, P. Huang, J. Gao, and W. Chen. Reasonet: Learning to stop reading in machine comprehension. In KDD, pages 1047–1055. ACM, 2017.

[31] A. Sordoni, P. Bachman, and Y. Bengio. Iterative alternating neural attention for machine reading.CoRR, abs/1606.02245, 2016.

[32] C. Tan, F. Wei, N. Yang, W. Lv, and M. Zhou. S-net: From answer extraction to answer generation for machine reading comprehension.CoRR,abs/1706.04815, 2017.

References:

[33] A. Trischler, T. Wang, X. Yuan, J. Harris, A. Sordoni, P. Bachman,and K. Suleman. Newsqa: A machine comprehension dataset.CoRR,abs/1611.09830, 2016.

[34] A. Trischler, Z. Ye, X. Yuan, P. Bachman, A. Sordoni, and K. Suleman. Natural language comprehension with the epireader. In EMNLP, pages 128–137. The Association for Computational Linguistics, 2016.21

[35] W. Wang, N. Yang, F. Wei, B. Chang, and M. Zhou. Gated self-matching networks for reading comprehension and question answering. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1:Long Papers, pages 189–198, 2017.

[36] Z. Wang, H. Mi, W. Hamza, and R. Florian. Multi-perspective context matching for machine comprehension. CoRR, abs/1612.04211, 2016.

[37] C. Xiong, V. Zhong, and R. Socher. Dynamic coattention networks forquestion answering.CoRR, abs/1611.01604, 2016.