BUILDING GENERALIZE QA SYSTEM, SLR

Muhammad Suleman
Department of Computer Science
FAST - NUCES, Karachi, Pakistan
Email: ssulemans@gmail.com

Mohammad Zoaib
Department of Computer Science
FAST - NUCES, Karachi, Pakistan
Email: k132290@nu.edu.pk

Hussain Shabbir
Department of Computer Science
FAST - NUCES, Karachi, Pakistan
Email: k132166@nu.edu.pk

Hatim Ali Asghar
Department of Computer Science
FAST - NUCES, Karachi, Pakistan
Email: k132260@nu.edu.pk

Humdiya Raza
Department of Computer Science
FAST - NUCES, Karachi, Pakistan
Email: k132086@nu.edu.pk

Tahir.Q.Syed
Department of Computer Science
FAST - NUCES, Karachi, Pakistan
Email: syed.tahir@nu.edu.pk

***Abstract*—Building Question Answering (QA) systems is a very important task in Information Retrieval (IR) / Natural Language Processing (NLP) domain. In IR / NLP domain there are many tasks which are similar which means solution of one task can be used to solve another task. These tasks include building QA systems, paraphrase detection, semantic similarity between sentences/words, semantic entailment, machine comprehension, slot filling and other like tasks. We found that many of these tasks are tackled in research using different techniques and different datasets. We also found that although standardized but most datasets are very small and they cannot solve generalize semantic assignment problem which lies at the heart of all these problems. Recently new dataset are published with larger size in the hope that they will be useful in building Generalize QA system. We do systematic literature review (SLR) of almost all the papers from 2010 to 2012 related to above mentioned problems. We try to find the correct direction for building a Generalize QA system that will be helpful in QA on any open domain corpus/dataset. We extract all the techniques / features / dataset / evaluation metrics / state of the art results published by different papers developing QA systems or doing like tasks. In the light of this statistics we answer some hypothesized question which we think are really helpful in building Generalize QA systems. For SLR we apply procedure as defined in [38].**

***Keywords*—QA systems, Deep Learning, Literature Review, Cognitive Systems, Information Retrieval.**

I. INTRODUCTION

In this paper, the investigation revolves around the development of the Generalized Question Answering (QA) System. For this purpose, a systematic literature review of X papers has been carried forward. The selection of relevant papers was made after applying a strict inclusion /exclusion principle. We have crafted a number of research questions and hypothesis. Answering all of these questions is necessary for building a generalized QA system. After systematic review of all the selected papers the answers of most of the questions and hypothesis were identified. However, the answers of some hypothesis need more research; therefore, it has been marked in the recommendations for the future work. In section 2 the definition of the criteria was pursued according to which papers were included in this investigation. In section 3 we present our research questions and hypothesis, the answers of the questions have been given in the light of statistics that has been gathered by the virtue of the literature review. In section 4 the List of Datasets has been tabulated accompanied by the List of QA systems along with datasets they are using and State of the art results these datasets are achieving. The section 5 highlighted the future directions for building a Generalize QA system. Finally section 6 encompasses of the comprehensive conclusion. In section 7 entails the references to all the papers we used for our SLR.

II. INCLUSION/EXCLUSION PRINCIPLE

Mostly we add papers after 2010 but include some heavily cited papers from earlier dates. Another criterion is we take papers with at least 10 citations, some time we break this rule by less cited but more relevant paper. We exclude all papers with non-supporting arguments as well as paper re-attempting same problem without any impact on research.

III. LIST OF QUESTIONS AND HYPOTHESIS

1) What are available IR/Cognitive systems?
2) What data sets are present and how much sample from Generalize QA they cover?
3) What are problem solved by deep learning?
4) What hybrid systems (Using both IR/Cognitive systems and deep learning)?
5) Future directions to solve Generalized QA problem?
6) Which technique is better, either cognitive system techniques or deep learning techniques?

We hypothesize that for almost all type of data deep learning architecture will work, but how to extract features for each deep learning architecture using IR/Cognitive system techniques needs careful design.

IV. LIST OF QA SYSTEMS

By Generalize QA system we mean a QA system which is capable of answering questions related to any open domain corpus. For example, corpus can include all Wikipedia pages or partial set of Wikipedia pages, collection of X (X is a positive integer) Washington post newspapers, a novel of Sherlock Holmes, a physics, history or geography textbook. QA system will be provided with corpus/dataset for training.

If after training QA system is able to answer all/most of the generalized questions related to any corpus correctly with high accuracy, we will classify this system as Generalize QA system. QA systems that are not capable of scaling for generalize question answering over open domain corpus are classified as Non-Generalize QA system. Below is a list of most of the datasets/corpus currently available.

V. TAXONOMY

In our literature review, we have read many research papers regarding cognitive system, deep learning, and hybrid approaches (which is using both cognitive approach as well as deep learning approach). We have figured out there are vast numbers of techniques defined in these approaches.

Firstly, discussing about the cognitive system research papers and features discussed in these research papers and model used in these research papers. Our first paper [5] uses integration of the spatio-temporal dimension, SPOTL, an extension of the original SPO-triple model to time and space technique, whereas second paper [06, 06] uses transformation function and semantic similarity technique using POS tagging and entity recognition approach. While some other papers like [11, 12, 13, 14, 15,16, 17, 18, 19, 20,21, 22, 23,85, 89,90, 91, 97] from these all mentioned papers most of the papers first uses translates natural language into a logical form, which is then evaluated to produce a vector and from those vector answer is being extracted while some other papers clearly uses semantic similarity technique using Frames semantic, co-reference, word embedding, and sliding window similarities. [81, 82] present a novel approach to answer validation based on the intuition that the amount of implicit knowledge which connects an answer to a question can be quantitatively estimated by exploiting the redundancy of Web information. [83] Purposes a structured inference system for this task, formulated as an Integer Linear Program (ILP) that answers natural language questions using a semi-structured knowledge base derived from text, including questions requiring multi-step inference and a combination of multiple facts. [84] This paper describes the design and implementation of the slot-lling system prepared by Stanfords natural language processing group for the 2011 Knowledge-Base Population track at the Text Analysis Conference.

There are many research papers regarding natural language question answering who are using natural language processing techniques these papers include [24, 25, 26, 27, 28,29,30, 31, 32, 33, 46, 47,70, 76, 78, 86,93] these all papers include a specific type of information retrieval techniques. These papers have built a Question Answering system which attempts to find out the correct answer to the question pose in natural language. As we know Question answering is multidisciplinary. From the technological perspective, these papers use natural or statistical language processing, information retrieval, and knowledge representation and reasoning as potential building blocks. Moreover, it includes text classification, information extraction and summarization techniques. In general, these questions answering system (QAS) has three components such as question classification, information retrieval, and answer extraction. These components play an essential role in QAS. Question classification play primary role in in most of the mentioned papers to categorize the question based upon on the type of its entity. One of the important methodology is Information retrieval which extract out applicable answer post by their intelligent question answering system. Finally, answer extraction module is rising topics in these Natural language research papers where these systems are often requiring ranking and validating a candidates answer. Some papers were on analyzing the semantics as the first step in NLP. [60] Proposes an algorithm labeling semantic classes and for leveraging them to extract is-a relationships using a top-down approach. [61, 95] We propose a new approach of obtaining expansion terms, based on selecting terms from past user queries that are associated with documents in the collection. [62, 65] proposes the use of hierarchical PitmanYor processes to model statistical dependencies between meaning representations of predicates and those of their arguments. [66] Proposes the technique of Alignment and bridging lexicalized syntactic denotation and having Test accuracy on web Questions i.e. 35.7, Accuracy. [67] This paper reviews six different information retrieval models: vector space model with cosine similarity, vector space model with weighted sum, latent semantic indexing, query likelihood model with Dirichlet smoothing, query likelihood model with linear smoothing, and query likelihood model with Dirchlet and LDA topic model smoothing. [68, 91] uses techniques like statistical parser, two-step pattern-based technique, novel technique, and SPARQL-like query language. [69] Combined ideas from imitation learning and agenda based parsing to train a semantic parser that search spartial logical forms in a more strategic order. [71] Paper addresses a missed opportunity to use crowdsourcing to understand the query itself. It proposes a novel hybrid human-machine approach that leverages the crowd to gain knowledge of query structure and entity relationships. [72] Uses unsupervised slot induction, semantic slot lling, semantic representation. [73] Presents an open-domain textual Question-Answering system that uses several feedback loops to enhance its performance. These feedback loops combine in a new way statistical results with syntactic, semantic, or pragmatic information derived from texts and lexical databases. [74, 75] papers rst presents the OpQA corpus of opinion questions and answers. Using the corpus, we compare the properties of fact and opinion questions and answers. Based on the disparate characteristics of opinion vs. fact answers, we argue that traditional fact-based QA approaches may have difculty in an MPQA setting without modication.[77] uses Semantic web, Question answering, Ontologies and Natural language techniques.[79] uses Algorithms used for finding top-k answers and approximations of top-k answers using proximity search.

Secondly, discussing about Deep learning architectures. The papers we have been through has [34] uses deep learning approach to match questions with answers by considering semantic encoding. [35] Proposes the discriminative algorithm for natural language parsing, based on a deep recurrent convolutional graph transformer network (GTN). [36] Presents a novel deep learning architecture which provides a semantic parsing system through the union of two neural models of language semantics. [37] Attempted to get an understanding of an extensive empirical evaluation of 19 different deep learning architectures. [38, 92] proposes a novel way of learning a neural network classifier for use in a greedy, transition-based dependency parser. [39] Proposed sequence to sequence framework of neural

Network. [40] The goal is to identify the subset that contains the answer using neural networks. [41] Proposes use of convolutional neural networks (CNN) trained on top of pre-trained word vectors for sentence-level classification tasks. [42] Proposes to learn morphological embeddings and propagate morphological information through the tree using a recursive composition procedure. [43, 98] details a compositional distributional framework based on a rich form of word embeddings that aims at facilitating the interactions between words in the context of a sentence. [44, 87] proposes an attention based neural matching model for ranking short answer text. [45] Presents a convolutional neural network architecture for re ranking pairs of short texts. [57, 99] also highlights the training of neural networks for question answering systems. [58, 88] paper finds out the feature types for ranking answers of QA systems. [63] Proposes a tri-modal deep belief network (tri-DBN) to extract a unified representation for the query, question, and answer. [64] Model consists of multiple agents and the communication between them is learned alongside their policy. [80, 100] uses, text classification, machine learning, support vector machine, kernel method.

TABLE I: Question Answering system using various datasets/corpus

S.NO	QA system	Dataset/Corpus
1	Match-Lstm[43],End-to-End Answer Chunk Extraction [285], Fine-Grained Gating [287],Dynamic Co-Attention Networks [288],	Squad
2	M Richardson et al[169],[42,62,65,70]	MCTest
3	Kushman et al. [37]	Algebra
4	Yang et al. [130]	WikiQA
5	Voorhees and Tice [290],[26,33,55,58,59,76,257,259,262,283]	TREC-QA
6	Hermann et al [71]	CNN/Daily Mail
7	[57,74]	Microsoft research Paraphrase identification task dataset
8	[62,101,109,129]	Dbpedia
9	[63,71]	BABi dataset
10	[96,97,128,162,257,262]	World Wide Web
11	[291]	Yahoo dataset
12	[179]	PARALAX dataset
13	IBM Watson[73]	Wikipedia Comprehension

TABLE II: Datasets/corpus with their purpose and sizes

S.NO	Dataset/Corpus	Purpose/Goal	Size
1	SQuad	Machine Comprehension	100k
2	MCTest	Machine Comprehension	2640
3	Algebra	Word Problems	524
4	WikiQA	Question Answering	3047
5	TREC-QA	Question Answering	1479
6	CNN/Daily Mail	Machine Comprehension	1.4 M
7	Microsoft research Paraphrase identification task dataset	Paraphrase Identification	train: 4, 076 test: 1,725
8	Dbpedia	Question Answering	3 GB
9	BABi dataset	Question Answering	¡1M
10	World Wide Web	Question Answering	
11	Yahoo dataset	Question Answering	2.5 GB
12	PARALAX dataset	Question Answering	1.8 million pairs of question and single relation
13	Wikipedia	Machine Comprehension	51 GB

TABLE III: State of the art achieving Deep learning techniques [1,2]

S.NO	Deep Learning Techniques	References
1	SAMS-RecNN	Cheng and Kartsaklis (2015)
2	Multi-Perspective CNN	He et al.(2015)
3	REL-TK	Filice et al.(2015)
4	Rao (2016) - PairwiseRank + Multi-Perspective CNN	Rao et al. (2016)
5	Yang (2016) - Attention-Based Neural Matching Model	Yang et al.(2016)
6	W and N (2015) - Three-Layer BLSTM+BM25	Wang and Nyberg (2015)
7	DT-RNN model	Mohit Iyyer (2014)
8	N gram of higher degrees Neural Network	A Severyn(2015)
9	Convolutional Neural Network for Modelling Sentences	N Kalchbrenner (2014)
10	Recurrent Neural Networks for Sequence Learning	ZC Lipton(2015)
11	Dependency Parser using Neural Networks	D Chen (2014)
12	Convolutional Neural Networks for Sentence Classification	Y Kim (2014)
13	Deep Neural Networks for Syntactic Parsing of Morphologically Rich Languages	J Legrand(2016)
14	Deep Neural Networks for Semantic Similarity Measurement	H He (2016)
15	Natural Language Grammatical Inference with Recurrent Neural Networks	S Lawrence(2000)

TABLE IV: State of the art achieving IR/Cognitive Systems techniques [1,2]

S.NO	IR/Cognitive System Techniques	References
1	LSA latent semantic space	Hassan (2011)
2	ESA explicit semantic space	Hassan (2011)
3	SSA salient semantic space	Hassan (2011)
4	ParaDetect	Zia and Wasif(2012)
5	TF-KLD	Ji and Eisenstein (2013)
6	Yao (2013)	Yao et al. (2013)
7	S and M (2013)	Severyn and Moschitti (2013)
8	Shnarch (2013) Backward	Shnarch (2013)
9	Yih (2013) LCLR	Yih et al. (2013)
10	W and I (2015)	Wang and Ittycheriah (2015)
11	L.D.C Model	Wang et al. (2016)
12	H and S (2010)	Heilman and Smith (2010)
13	W and M (2010)	Wang and Manning (2010)

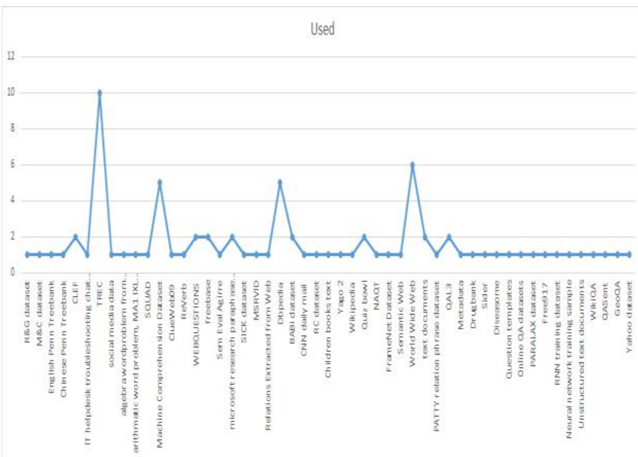

Fig. 1: No of datasets used by the researchers [49-59]

VI. Statistics

Most of the above-mentioned QA systems are Non-Generalize QA systems. TREC is oldest and most popular dataset for question answering but QA systems that target track are not capable of generalization. This is because of small size of dataset. A variety of QA systems are built using TREC dataset. The attainment of the State of the art results for TREC can be found from aclweb website [1]. WikiQA, CNN/Daily mail, MCTest has same problem of small size. Microsoft research paraphrase detection MSPR is a dataset used for paraphrase detection, a task similar to QA systems. State of the art results for many systems using MSPR dataset can be found here [2]. IBM Watson uses comprehension and facts from Wikipedia corpus. WATSON won several competitions in jeopardy. Unfortunately, WATSON is closed source due to which research community is unable to check its accuracy claims for many datasets/corpuses. In addition, it is also difficult to identify the effectiveness of WATSONs cognitive system. IBM Watson works on factoid question [73], in jeopardy competition it demonstrated good performance concerning generalize factoid question answering. Most of the new QA systems are designed using deep learning models. Most of the datasets used deep learning models. SQUAD is a new dataset containing 100,000+ rows of crowd sourced comprehension plus QA. Good thing about SQUAD is that, it is the only available dataset that can be proved to be a sound aid for Generalize QA system due to its large dataset size. State of the art performance for SQUAD dataset is achieved using deep learning models. It can be concluded whether SQUAD generalizes to give correct answer on any dataset after testing it on many datasets/corpuses. For a Generalize QA system there are many other domains to cater including algebra, word problem solver, end of chapter problem solver for probability, number theory, mechanics and thermodynamic problems and puzzle solvers. It is found that only 2 paper for solving word problems [36, 37]. We think that a dataset like SQUAD for word problems is of prime importance, such a dataset can be made using end of chapter problems in undergrad standard mathematical text books.

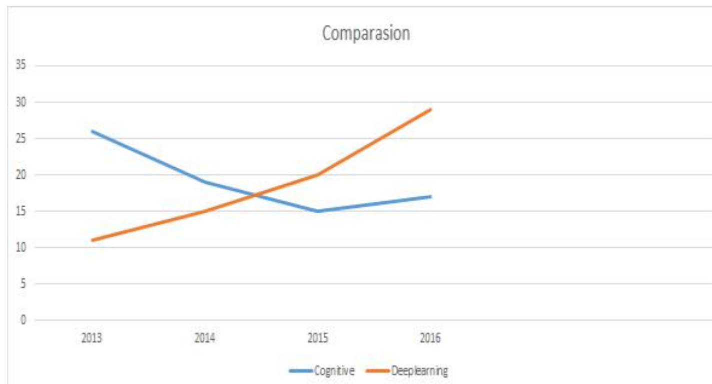

Fig. 2: A graph showing from year 2013 - 2016, how cognitive system curve is falling down as with the increase of deep learning. [101 - 116]

Our next question revolved around the identification of a better technique from the array of cognitive system techniques or deep learning techniques. Graph in Figure 2 clearly shows that deep learning and neural network based architectures are continuously dominating cognitive systems. But the real question is what will be proper architecture for utilizing deep learning.

Some papers use deep learning for each and every task [34,186], although this approach seems independent from human resources but it cannot be generalized due to enormous amount of training data. Most of the papers that demonstrated the state of art results using deep learning architecture or neural networks feed are using automatic features generated by IR or cognitive system techniques. Below table and chart shows features used and their proportion

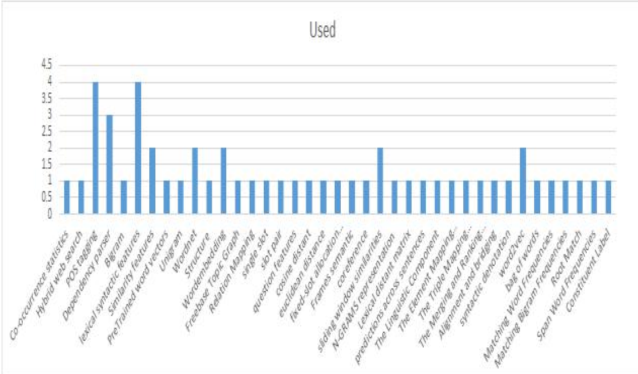

Fig. 3: No of features used by the researchers in their models.

Although these papers discuss reasoning for using one feature in place of other but a proper guideline of what features works best for which type of data is still far from being published.

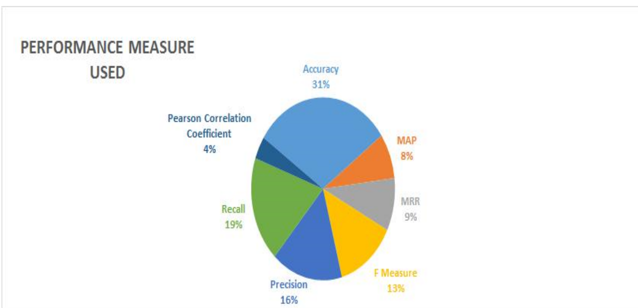

Fig. 4: Performance Measure used by the researchers to evaluate systems

VII. FUTURE WORK

We hypothesize that for almost all type of data deep learning architecture will work, but we need to figure out which data can be handled using which feature. As an example, we hypothesize that there can be a single deep learning architecture which after training on all pages in Wikipedia category history can answer most generalized set of questions related to history, we can use same deep learning architecture for geography category, but we cannot use same architecture for

psychology category. For psychology category we need another deep learning architecture, since history and geography are just informative while psychology adds concepts to information. We deliberately do not give example of maths or physics category because they are trivially different from history and geography category.

Both of the deep learning architecture differs in layers, features extracted from IR techniques and other hyperparameters. Similarly, there should be different deep learning architectures for comprehension in novels of Sherlock Holmes and codes written in law or medicine books. For a generalized QA system, all these architectures should be tightly coupled in a single system. IR/Cognitive system techniques can be used to decide which neural architecture can be used for which task. In addition, IR/Cognitive system techniques can be used for generating features for all the deep learning architectures. Due to scope of this paper we will not give answer of this hypothesis in this paper, instead we consider it as a future work.

VIII. CONCLUSION

In this paper, we investigated how to build a Generalized Question Answering (QA) System. IBM Watson[101] was considered as such a system but it was closed source due to which we cannot verify its claims. Most of the obstacles in Generalized QA systems were due to lack of large dataset. With the inclusion of SQUAD in dataset it looks that this problem is overcome. Development of QA systems is now dominated by deep learning architecture. Mostly features of deep learning architectures are extracted using IR/Cognitive system techniques. Still there is lacking in QA dataset which should include problems from algebra domain, word problems, end of chapter problems from probability, number theory, mechanics and thermodynamic textbooks and logic puzzles. To improve training performance of deep learning architecture for Generalize QA systems we give some insights and leave complete architecture and its implementation as future work.

REFERENCES

[1] https://aclweb.org/aclwiki/index.php?title=Question Answering (State of the art)
[2] https://aclweb.org/aclwiki/index.php?title=Para phrase Identification (State of the art)
[3] Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.
[4] Guangyou Zhou-Yubo Chen-Daojian Zeng-Jun Zhao - Proceedings of the 22nd ACM international conference on Conference on information and knowledge management - CIKM '13 2013
[5] Hoffart, J., Suchanek, F. M., Berberich, K., and Weikum, G. (2013). YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence, 194, 28-61.
[6] Hamani, M. S., and Maamri, R. (2013, August). Word Semantic Similarity Based on Document's Title. In 2013 24th International Workshop on Database and Expert Systems Applications (pp. 43-47). IEEE.
[7] Angeli et al.2014] Gabor Angeli, Sonal Gupta, Melvin Jose,Christopher D Manning, Christopher R e, Julie Tibshirani,Jean Y Wu, Sen Wu, and Ce Zhang. 2014. Stanfords 2014slot filling systems.TAC KBPVector Space Semantic Parsing: A Framework for Compositional Vector Space Models
[8] Hecht, B., Teevan, J., Morris, M.R. and Liebling, D. 2012. SearchBuddies: Bringing search engines into the conversation. Proc. of ICWSM 2012, 138-145.A Multi-Strategy and Multi-Source Approach to Question Answering
[9] Deepa Yogish, Prof. Manjunath TN Prof. Ravindra S Hegadi 2016 A Survey of Intelligent Question Answering System Using NLP and Information Retrieval Techniques International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 5, May 2016
[10] Weston, J., Bordes, A., Chopra, S., Rush, A. M., van Merrinboer, B., Joulin, A., and Mikolov, T. (2015). Towards ai-complete question answering: A set of prerequisite toy tasks. arXiv preprint arXiv:1502.05698.The SemSets Model for Ad-hoc Semantic List Search
[11] Krishnamurthy, Jayant, and Tom Mitchell. "Vector space semantic parsing: A framework for compositional vector space models." (2013): 1.
[12] Bordes, Antoine, et al. "Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing." AISTATS. Vol. 22. 2012.
[13] Pakray, Partha, and et al. "A Hybrid Question Answering System based on Information Retrieval and Answer Validation." CLEF (Notebook Papers/Labs/Workshop). 2011.

[14] Starzyk, Janusz A., and Marek Jaszuk. ”A question answer approach to building semantic memory.” International Conference on Artificial Intelligence and Soft Computing. Springer Berlin Heidelberg, 2012.
[15] Rodrigo, lvaro, et al. ”A Question Answering System based on Information Retrieval and Validation.” CLEF (Notebook Papers/LABs/Workshops). 2010.
[16] Kolomiyets, Oleksandr, and Marie-Francine Moens. ”A survey on question answering technology from an information retrieval perspective.” Information Sciences 181.24 (2011): 5412-5434.
[17] Mooney, Ray. ”Learning semantic parsers: An important but understudied problem.” Working notes of the AAAI spring symposium on language learning. 2004.
[18] Jansen, Peter, Mihai Surdeanu, and Peter Clark. ”Discourse Complements Lexical Semantics for Non-factoid Answer Reranking.” ACL (1). 2014.
[19] Gliozzo, Alfio, et al. ”Semantic technologies in ibm watsontm.” ACL 2013 (2013): 85.
[20] Rindflesch, Thomas C., and Alan R. Aronson. ”Semantic processing in information retrieval.” Proceedings of the Annual Symposium on Computer Application in Medical Care. American Medical Informatics Association, 1993.
[21] Maind, Ankush, Anil Deorankar, and Prashant Chatur. ”Measurement of semantic similarity between words: A survey.” International Journal of Computer Science, Engineering and Information Technology 2.6 (2012): 189-194.
[22] Fillmore, Charles J., and Collin F. Baker. ”Frame semantics for text understanding.” Proceedings of WordNet and Other Lexical Resources Workshop, NAACL. 2001.
[23] Java, Akshay, Tim Finin, and Sergei Nirenburg. ”Text understanding agents and the Semantic Web.” System Sciences, 2006. HICSS’06. Proceedings of the 39th Annual Hawaii International Conference on. Vol. 3. IEEE, 2006.
[24] Dong, Tiansi, et al. ”A natural language question answering system as a participant in human Q and A portals.” Twenty-Second International Joint Conference on Artificial Intelligence. 2011.
[25] Socher, Richard, et al. ”Parsing natural scenes and natural language with recursive neural networks.” Proceedings of the 28th international conference on machine learning (ICML-11). 2011.
[26] Hirschman, Lynette, and Robert Gaizauskas. ”Natural language question answering: the view from here.” natural language engineering 7.04 (2001): 275-300.
[27] Schuster, Sebastian, and Christopher D. Manning. ”Enhanced english universal dependencies: An improved representation for natural language understanding tasks.” Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). 2016.
[28] Lin, Qian, and Hongyu Xiong. ”Dynamic Memory Network on Natural Language Question-Answering.”
[29] Angeli, Gabor, and Christopher D. Manning. ”NaturalLI: Natural Logic Inference for Common Sense Reasoning.” EMNLP. 2014.
[30] Song, Dezhao, et al. ”TR Discover: A Natural Language Question Answering System for Interlinked Datasets.” International Semantic Web Conference (Posters and Demos). 2015.
[31] Shamsfard, Mehrnoush, and Ahmad Abdollahzadeh Barforoush. ”Learning ontologies from natural language texts.” International journal of human-computer studies 60.1 (2004): 17-63.
[32] Thompson, Cynthia A., Mary Elaine Califf, and Raymond J. Mooney. ”Active learning for natural language parsing and information extraction.” ICML. 1999.
[33] Socher, Richard, et al. ”Parsing natural scenes and natural language with recursive neural networks.” Proceedings of the 28th international conference on machine learning (ICML-11). 2011.
[34] Yu, Lei, et al. ”Deep learning for answer sentence selection.” arXiv preprint arXiv:1412.1632 (2014).
[35] Collobert, Ronan. ”Deep Learning for Efficient Discriminative Parsing.” AISTATS. Vol. 15. 2011.
[36] Grefenstette, Edward, et al. ”A deep architecture for semantic parsing.” arXiv preprint arXiv: 1404.7296 (2014).
[37] Laha, Anirban, and Vikas Raykar. ”An Empirical Evaluation of various Deep Learning Architectures for Bi-Sequence Classification Tasks.” arXiv preprint arXiv:1607.04853 (2016).
[38] Chen, Danqi, and Christopher D. Manning. ”A Fast and Accurate Dependency Parser using Neural Networks.” EMNLP. 2014.
[39] Vinyals, Oriol, and Quoc Le. ”A neural conversational model.” arXiv preprint arXiv: 1506.05869 (2015).
[40] Rao, Jinfeng, Hua He, and Jimmy Lin. ”Noise-Contrastive Estimation for Answer Selection with Deep Neural Networks.” Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016.
[41] Kim, Yoon. ”Convolutional neural networks for sentence classification.” arXiv preprint arXiv: 1408.5882 (2014).
[42] Legrand, Jol, and Ronan Collobert. ”Deep neural networks for syntactic parsing of morphologically rich languages.” Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. No. EPFL-CONF-219841. 2016.
[43] Cheng, Jianpeng, and Dimitri Kartsaklis. ”Syntax-aware multi-sense word embeddings for deep compositional models of meaning.” arXiv preprint arXiv: 1508.02354 (2015).
[44] Yang, Liu, et al. ”aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model.” Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016.
[45] Severyn, Aliaksei, and Alessandro Moschitti. ”Learning to rank short text pairs with convolutional deep neural networks.” Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015.
[46] Tellex, Stefanie, et al. ”Quantitative evaluation of passage retrieval algorithms for question answering.” Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, 2003.
[47] Cabrio, Elena, et al. ”Question answering based annotation for a corpus of spoken requests.” Proceedings of the workshop on the Semantic Representation of Spoken Language, Salamanca, Spain, November. 2007.
[48] [48] Sak, Haim, Andrew Senior, and Franoise Beaufays. ”Long short-term memory recurrent neural network architectures for large scale acoustic modeling.” Fifteenth Annual Conference of the International Speech Communication Association. 2014.
[49] Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.
[50] Kayes, I., Kourtellis, N., Quercia, D., Iamnitchi, A., and Bonchi, F. (2015, August). Cultures in community question answering. In Proceedings of the 26th ACM Conference on Hypertext and Social Media (pp. 175-184). ACM.
[51] Tahri, A., and Tibermacine, O. (2013). DBPEDIA BASED FACTOID QUESTION ANSWERING SYSTEM. International Journal of Web and Semantic Technology, 4(3), 23.
[52] Moussa, A. M., and Abdel-Kader, R. F. (2011). Qasyo: A question answering system for yago ontology. International Journal of Database Theory and Application, 4(2), 99.
[53] Lopez, V., Uren, V., Motta, E., and Pasin,
[54] M. (2007). AquaLog: An ontology-driven question answering system for organizational semantic intranets. Web Semantics: Science, Services and Agents on the World Wide Web, 5(2), 72-105.
[55] Litkowski, K. C. (1999). Question-Answering Using Semantic Relation Triples. In TREC.
[56] Cucerzan, S., and Agichtein, E. (2005). Factoid Question Answering over Unstructured and Structured Web Content. In TREC (Vol. 72, p. 90).
[57] H. Cui, R. X. Sun, K. Y. Li, M. Y. Kan, T. S. Chua. 2005. Question answering passage retrieval using dependency relations. In Proceedings of the ACM SIGIR, 400407. ACM Press.
[58] Tellex, S., Katz, B., Lin, J., Fern, A., and Marton, G. (2003). Quantitative evaluation of passage retrieval algorithms for question answering. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 4147.
[59] R. D. Burke, K. J. Hammond, V. A. Kulyukin, S. L. Lytinen, N. Tomuro, and S. Schoenberg. Question answering from frequently asked question files: Experiences with the faq finder system. Technical report, 1997.
[60] Pantel, Patrick, and Deepak Ravichandran. ”Automatically Labeling Semantic Classes.” HLT-NAACL. Vol. 4. 2004.
[61] Billerbeck, Bodo, et al. ”Query expansion using associated queries.” Proceedings of the twelfth international conference on Information and knowledge management. ACM, 2003.
[62] Titov, Ivan, and Alexandre Klementiev. ”A Bayesian model for unsupervised semantic parsing.” Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011.

[63] Jiang, Nan, et al. "Modeling Joint Representation with Tri-Modal Deep Belief Networks for Query and Question Matching." IEICE TRANSACTIONS on Information and Systems 99.4 (2016): 927-935.
[64] Sukhbaatar, Sainbayar, and Rob Fergus. "Learning multiagent communication with backpropagation." Advances in Neural Information Processing Systems. 2016.
[65] Jones, Mark P., Paul Hudak, and Sebastian Shaumyan. "Using Types to Parse Natural Language." Functional Programming. 1995.
[66] Berant, Jonathan, et al. "Semantic Parsing on Freebase from Question-Answer Pairs." EMNLP. Vol. 2. No. 5. 2013.
[67] Binkley, Dave, et al. TR 140512: on information retrieval models. Technical Report www. Cs. Loyola. Edu/binkley/papers/TR-140512. Pdf, Loyola University, 2014.
[68] Mousavi, Hamid, et al. "Mining semantic structures from syntactic structures in free text documents." Semantic Computing (ICSC), 2014 IEEE International Conference on. IEEE, 2014.
[69] Berant, Jonathan, and Percy Liang. "Imitation learning of agenda-based semantic parsers." Transactions of the Association for Computational Linguistics 3 (2015): 545-558.
[70] Soricut, Radu, and Eric Brill. "Automatic Question Answering: Beyond the Factoid." HLT-NAACL. 2004.
[71] Demartini, Gianluca, et al. "CrowdQ: Crowdsourced Query Understanding." CIDR. 2013.
[72] Chen, Yun-Nung, William Yang Wang, and Alexander I. Rudnicky. "Unsupervised induction and filling of semantic slots for spoken dialogue systems using frame-semantic parsing." Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on. IEEE, 2013.
[73] Harabagiu, Sanda, et al. "The role of lexico-semantic feedback in open-domain textual question-answering." Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2001.
[74] Jiang, Jay J., and David W. Conrath. "Semantic similarity based on corpus statistics and lexical taxonomy." arXiv preprint cmp-lg/9709008 (1997).
[75] Stoyanov, Veselin, Claire Cardie, and Janyce Wiebe. "Multi-perspective question answering using the OpQA corpus." Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005.
[76] Bouma, Gosse. "Linguistic knowledge and question answering." Proceedings of the Workshop KRAQ'06 on Knowledge and Reasoning for Language Processing. Association for Computational Linguistics, 2006.
[77] Lopez, Vanessa, et al. "AquaLog: An ontology-driven question answering system for organizational semantic intranets." Web Semantics: Science, Services and Agents on the World Wide Web 5.2 (2007): 72-105.
[78] Lin, Dekang, and Patrick Pantel. "Discovery of inference rules for question-answering." Natural Language Engineering 7.04 (2001): 343-360.
[79] Surdeanu, Mihai, Massimiliano Ciaramita, and Hugo Zaragoza. "Learning to Rank Answers on Large Online QA Collections." ACL. Vol. 8. 2008.
[80] Zhang, Dell, and Wee Sun Lee. "Question classification using support vector machines." Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, 2003.
[81] Cui, Hang, et al. "Question answering passage retrieval using dependency relations." Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2005.
[82] Magnini, Bernardo, et al. "Is it the right answer?: exploiting web redundancy for Answer Validation." Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002.
[83] Khashabi, Daniel, et al. "Question answering via integer programming over semi-structured knowledge." arXiv preprint arXiv:1604.06076 (2016).
[84] Surdeanu, Mihai, et al. "Stanford's Distantly-Supervised Slot-Filling System." TAC. 2011.
[85] Maind, Ankush, Anil Deorankar, and Prashant Chatur. "Measurement of semantic similarity between words: A survey." International Journal of Computer Science, Engineering and Information Technology 2.6 (2012): 189-194.
[86] Soricut, Radu, and Eric Brill. "Automatic Question Answering: Beyond the Factoid." HLT-NAACL. 2004.
[87] Yang, Liu, et al. "aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model." Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016.
[88] Surdeanu, Mihai, Massimiliano Ciaramita, and Hugo Zaragoza. "Learning to Rank Answers on Large Online QA Collections." ACL. Vol. 8. 2008.
[89] Jansen, Peter, Mihai Surdeanu, and Peter Clark. "Discourse Complements Lexical Semantics for Non-factoid Answer Reranking." ACL (1). 2014.
[90] Maind, Ankush, Anil Deorankar, and Prashant Chatur. "Measurement of semantic similarity between words: A survey." International Journal of Computer Science, Engineeringand Information Technology 2.6 (2012): 189-194.
[91] Krishnamurthy, Jayant, and Tom Mitchell. "Vector space semantic parsing: A framework for compositional vector space models." (2013): 1
[92] Chen, Danqi, and Christopher D. Manning. "A Fast and Accurate Dependency Parser using Neural Networks." EMNLP. 2014.
[93] Song, Dezhao, et al. "TR Discover: A Natural Language Question Answering System for Interlinked Datasets." International Semantic Web Conference (Posters and Demos). 2015
[94] Hamani, Mohamed Said, and Ramdane Maamri. "Word Semantic Similarity Based on Document's Title." Database and Expert Systems Applications (DEXA), 2013 24th International Workshop on. IEEE, 2013.
[95] Billerbeck, Bodo, et al. "Query expansion using associated queries." Proceedings of the twelfth international conference on Information and knowledge management. ACM, 2003.
[96] Mousavi, Hamid, et al. "Mining semantic structures from syntactic structures in free text documents." Semantic Computing (ICSC), 2014 IEEE International Conference on. IEEE, 2014.
[97] Java, Akshay, Tim Finin, and Sergei Nirenburg. "Text understanding agents and the Semantic Web." System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on. Vol. 3. IEEE, 2006.
[98] Cheng, Jianpeng, and Dimitri Kartsaklis. "Syntax-aware multi-sense word embeddings for deep compositional models of meaning." arXiv preprint arXiv: 1508.02354 (2015).
[99] Srivastava, Rupesh K., Klaus Greff, and Jrgen Schmidhuber. "Training very deep networks." Advances in neural information processing systems. 2015.
[100] Zhang, Dell, and Wee Sun Lee. "Question classification using support vector machines." Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, 2003
[101] Lally, A., Prager, J. M., McCord, M. C., Boguraev, B. K., Patwardhan, S., Fan, J., ... and Chu-Carroll, J. (2012). Question analysis: How Watson reads a clue. IBM Journal of Research and Development, 56(3.4), 2-1.
[102] Irsoy, O., and Cardie, C. (2014, October). Opinion Mining with Deep Recurrent Neural Networks. In EMNLP (pp. 720-728).
[103] Durrett, G., and Klein, D. (2015). Neural crf parsing. arXiv preprint arXiv:1507.03641.
[104] Neural responding machine for short-text conversation.
[105] Wang, C., Kalyanpur, A., Fan, J., Boguraev, B. K., and Gondek, D. C. (2012). Relation extraction and scoring in DeepQA. IBM Journal of Research and Development, 56(3.4), 9-1.
[106] Sneiders, E. (2002, June). Automated question answering using question templates that cover the conceptual model of the database. In International Conference on Application of Natural Language to Information Systems (pp. 235-239). Springer Berlin Heidelberg.
[107] Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. (2008, April). Predictors of answer quality in online QandA sites. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 865-874). ACM.
[108] Ye, Z., Jia, Z., Yang, Y., Huang, J., and Yin, H. (2015, October). Research on Open Domain Question Answering System. In National CCF Conference on Natural Language Processing and Chinese Computing (pp. 527-540). Springer International Publishing.
[109] Hirschman, L., and Gaizauskas, R. (2001). Natural language question answering: the view from here. natural language engineering, 7(04), 275-300.
[110] Noy, N. F. (2004). Semantic integration: a survey of ontology-based approaches. ACM Sigmod Record, 33(4), 65-70.

[111] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958.

[112] Meng, F., Lu, Z., Wang, M., Li, H., Jiang, W., and Liu, Q. (2015). Encoding source language with convolutional neural network for machine translation. arXiv preprint arXiv:1503.01838.

[113] Yih, Wen-tau and Chang, Ming-Wei and Meek, Christopher and Pastusiak, Andrzej. 2013. Question Answering Using Enhanced Lexical Semantic Models. In ACL 2013.

[114] Abney, S., Collins, M., and Singhal, A. (2000, April). Answer extraction. In Proceedings of the sixth conference on Applied natural language processing(pp. 296-301). Association for Computational Linguistics.

[115] Zhiguo Wang, Haitao Mi and Abraham Ittycheriah. 2016. Sentence Similarity Learning by Lexical Decomposition and Composition. In Coling 2016.

[116] Khashabi, D., Khot, T., Sabharwal, A., Clark, P., Etzioni, O., and Roth, D. (2016). Question Answering via Integer Programming over Semi-Structured Knowledge. arXiv preprint arXiv:1604.06076.