A Corpus for Hybrid Question Answering Systems

Brigitte Grau, Anne-Laure Ligozat

A Corpus for Hybrid Question Answering Systems

Brigitte Grau
LIMSI, CNRS, ENSIIE, University Paris-Saclay
Orsay, France
brigitte.grau@limsi.fr

Anne-Laure Ligozat
LIMSI, CNRS, ENSIIE, University Paris-Saclay
Orsay, France
anne-laure.ligozat@limsi.fr

ABSTRACT

Question answering has been the focus of a lot of researches and evaluation campaigns, either for text-based systems (TREC and CLEF evaluation campaigns for example), or for knowledge-based systems (QALD, BioASQ). Few systems have effectively combined both types of resources and methods in order to exploit the fruitfulness of merging the two kinds of information repositories. The only evaluation QA track that focuses on hybrid QA is QALD since 2014. As it is a recent task, few annotated data are available (around 150 questions). In this paper, we present a question answering dataset that was constructed to develop and evaluate hybrid question answering systems. In order to create this corpus, we collected several textual corpora and augmented them with entities and relations of a knowledge base by retrieving paths in the knowledge base which allow to answer the questions. The resulting corpus contains 4300 question-answer pairs and 1600 have a true link with DBpedia.

KEYWORDS

Hybrid Question Answering, Corpus

ACM Reference Format:
Brigitte Grau and Anne-Laure Ligozat. 2018. A Corpus for Hybrid Question Answering Systems. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3184558.3191540

1 INTRODUCTION

Question answering (QA) systems provide a user-friendly tool for seeking different kinds of resources, as they allow the user to enter questions written in natural language. Such systems have to extract the answers from relevant documents or to query a database or even to exploit both kinds of resources.

Question answering has been the focus of a lot of researches and evaluation campaigns, either for text-based systems (TREC and CLEF evaluation campaigns for example), or for knowledge-based systems (QALD, BioASQ). Few systems [6, 24, 26] have effectively combined both types of resources and methods in order to exploit the fruitfulness of merging the two kinds of information repositories. For example, if presented with the question *At which college did the only American actor that received the César Award study?*, a system will find the actor more probably in texts and his college in a knowledge base. Textual resources contain a great amount of information, but require complex natural language processing (NLP) tools for extracting answers. On the contrary, knowledge bases contain structured information, which makes it possible to query them directly once the question has been translated into the appropriate query language. Yet, if knowledge bases are much more reliable, they remain incomplete and do not contain as much information as texts. Moreover, KB are not dedicated to store contextual information or information about all the entities. Only the most famous entities have entries in a KB. Hybrid QA systems aim at exploiting both types of information.

The only QA evaluation track that focuses on hybrid QA is QALD, since 2014 [18]. As it is a recent task, few annotated data are available for learning. Moreover, in this track, all answers are KB entities. Other question answering datasets exist, but they are built for developing systems dedicated to search in one resource only: Trec, CLEF and SQuAD datasets for textual QA, QALD and WebQuestions datasets for knowledge base QA. Thus they are not suitable for training or evaluating a hybrid QA system. Textual datasets do not provide the answer URIs, when they exist, which are required to evaluate the results of a knowledge base search. Concerning KB datasets, QALD dataset contains too few examples, and WebQuestions contains mostly simple questions, i.e. questions that can be solved by a single triple and do not require a hybrid approach.

In this paper, we present a question answering dataset that was constructed to develop and evaluate hybrid question answering systems. It contains both question and answer pairs in textual form, and references to the KB. The textual mentions of entities have a reference to their entity in the knowledge base, and useful relations of the KB are added to the pairs enabling to align text and structured representations. The questions must be complex enough for needing the resort to a hybrid solution.

In order to create this corpus, we collected several textual corpora and augmented them with entities and relations of a knowledge base by retrieving the paths in the knowledge base which allow to answer the questions. The resulting corpus contains 4300 question-answer pairs where 1600 have a true link with DBpedia and can be used for learning and testing hybrid QA systems as well as improving KB systems on complex questions[1].

2 RELATED WORK

2.1 Question answering systems

Most QA systems are dedicated to search for an answer either in text or in a knowledge base, but not both. Textual QA systems rely mainly on methods able to recognize the similarity between a question and a sentence, or more generally a textual entailment relation between them. Their goal is mainly to model the recognition of lexical and syntactic overlaps that take into account lexical

[1]The corpus can be found at https://zenodo.org/record/1186300#.Wpbj-eYiE5s

WWW '18 Companion, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3191540

variations. Methods range from feature based learning methods [9, 23, 28, 29] to neural network methods, for example [11][2], that show better performances when dealing with lexical similarity.

The extraction of the exact answer involves criteria based on the determination of the expected answer type and its matching in the candidate sentences, based on named entity recognition and syntactic features. It also ranges from feature based learning approaches [10] to neural network approaches [5][3]

One of the main challenges when querying a KB given a NL question concerns the alignment of a question with the KB triples, which needs to overcome lexical gap and to adapt the question parsing to the KB schema in order to determine which phrases are entity or relation mentions (see Diefenbach et al. [8] for a recent survey). The query can be generated by representing the question based on semantic graphs [2, 30] or patterns [17] and transforming this representation into a query. Yao and Van Durme adopt a less sequential approach and extract a subgraph of the KB around the entities recognized in the question. Deep neural network methods are similar to those applied on text and compare a question representation learned from word embeddings to triple representations. The triple representation are learned from the KB triples [4] or from their label [13].

Some hybrid QA approaches were developed. Besides former methods that use in parallel both resources for searching an answer [6, 7, 12] , [6] also use them in a complementary strategy for verifying a candidate answer type. Some recent works develop a collaborative strategy. Yahya et al. developed query extension and relaxation techniques to search for information in the text contexts associated to triples. Park et al., on the contrary, first search for information in texts annotated with KB entities, and use SPARQL queries if the text strategy is not successful. In [25], a hybrid search is really performed with the decomposition of the questions into subparts that are searched in both kinds of resources and provided answers are aggregated for the final answer selection.

The exploration of hybrid approaches needs to enlarge the available datasets and to provide them with information that allow to learn and evaluate the alignment of text and KB data. It will also benefit the mono-source QA systems.

2.2 QA datasets

2.2.1 Text QA corpora. QA evaluation campaigns have been organized at TREC between 1999 and 2007. TREC questions are built from logs of search engines and answers have to be extracted from texts, that are mainly newswires. Datasets of question-answer pairs are freely available, but without their supporting passages, i.e. passages that support and justify the answer given. Around 500 factual questions[4] were released each year.

CLEF QA campaigns took place between 2003 and 2009 with an important evolution over the time concerning the type of questions and the documents to search. Documents are newswires and articles from Wikipedia. As CLEF propose evaluations for the European languages, the questions are conceived so that they can be answered from texts in each studied language [5]. Each year, 200 questions were provided and their textual answers were released.

Questions in these corpora concern news that were present in newspapers. Thus, some questions are closely linked to the actuality at the time of the newspapers. Others are timeless and concerns encyclopedic information.

Apart from evaluation campaigns, some corpora have been distributed. The Microsoft Research Question-Answering Corpus[6] is a corpus of 400 factual questions on Encarta 98.

The Stanford Question Answering Dataset (SQuAD)[7] [16] a corpus of more than 100 000 question-answer pairs concerning more than 500 articles of Wikipedia that were collected by crowdsourcing. This corpus is now largely used for developing textual QA systems based on deep learning.

2.2.2 Knowledge base corpora. QALD evaluation campaign exists since 2011, whose objective is to evaluate the performances of QA systems over a knowledge base. The datasets are made of about 200 textual questions every year. They must be answered using DBpedia and are provided with a SPARQL query. The answers are DBpedia URIs. One limitation of this corpus is the low number of questions, which makes it hard to use in learning methods. Another limitation is the bias due to the fact that this corpus was built specifically to be answerable on a knowledge base.

WebQuestions [3] is a question-answer dataset on Freebase. The textual questions have been build using Google Suggest API queries and Amazon Mechanical Turk to filter them. Answers are the Freebase URIs. It is constituted of 3778 examples for training and 2032 for test. The questions are rather simple in their form, as most of them can be represented as a single triple of the knowledge base and do not seem complex enough to evaluate a hybrid question answering system.

2.2.3 Hybrid corpora. QALD has created a task for hybrid QA since 2014. The objective of the task is to answer questions by using both triples from DBpedia and the abstract of each relevant Wikipedia article. Until now, it has provided around 150 question-answer pairs that can be used to evaluate a question answering hybrid system, but the number of questions remains low for learning.

3 REQUIREMENTS OF THE NEW DATASET

In a first step, we explore which kinds of phenomena occurring in QA could benefit a hybrid approach, so that it will allow us to define the content of our new corpus.

We consider a knowledge base which is made of triplets and stores binary relations (subject, *predicate*, object) about instances, such as DBpedia or Freebase, and a text corpus. We also define a hybrid search as follows: finding the answer would require to search for information in the two sources and aggregate them.

The possibility to find answers using either text or a knowledge base, or both, is related to the question content, – are they about an entity, an event, a concept – and the type of answer that is expected – a definition, a factual information, an entity, an explanation, etc.

[2]More references are given at http://aclweb.org/aclwiki/index.php?title=Question_Answering_State_of_the_art

[3]Results of several Deep Neural Network models using the SQUAD Dataset can be found on the leaderboard https://rajpurkar.github.io/SQuAD-explorer/

[4]http://trec.nist.gov/data/qamain.html

[5]http://nlp.uned.es/clef-qa/repository/qa.php

[6]https://www.microsoft.com/en-us/download/details.aspx?id=52318

[7]https://rajpurkar.github.io/SQuAD-explorer/

–. As soon as the question mentions an entity or the answer is a known entity, a KB search can be fired in addition or in place of a textual search. However, we wanted to examine more closely which cases would need a hybrid resolution. A preliminary corpus study [8] lead us to define the following cases:

- The answer is a property value of an entity given in the question (direct relationship): *Who is Bill Clinton's daughter's husband?*. The answer can be searched directly in both sources, or be solved by hybrid search because it requires finding three entities and two relations which could not be both present in a KB: (Bill Clinton, *daughter*, Chelsea Clinton) and (Chelsea Clinton, *married*, Marc Mervinsky) ;
- The answer is about an event, either about a role or the name of the event: *Who is the assassin of Martin Luther King?*. In general the answer will come from texts, especially if the event involves unknown entities as events are often not represented in KB;
- A combination of the two preceding cases, for example a direct relation with an entity having a role in an event (composition of relations): *Where was the assassin of Martin Luther King born?*. A hybrid search should be done.
- The answer is either an instance or a concept : *What animal lays blue eggs ?*. The answer will certainly come from text, with for example "The Collonca are without tail and lay blue eggs", and the verification that Collonca are animals can be operated on one or the other source.
- The relation with the answer is contextual; it can be an opinion or related to an event, for example as in *Which country bought petroleum to Irak during the embargo?* (about the *embargo* event). The answer can only be found in texts;
- A definition: *What is an atom?*. The answer is in texts ;
- A result coming from an operator of aggregation (comparison, sorting, counting): *Give the ten bigger French companies*. The answers can come from texts, as long as the searched information is explicit, but they are easier to deduce from a knowledge base.

In conclusion, if we want a corpus to be helpful for hybrid research, we must ensure that: 1) at least one entity is present in the question or the answer; 2) there exists at least one relation in the KB relevant for answering the question; 3) the question often contains additional information so that its whole meaning cannot be represented by a unique relation.

4 CONSTRUCTION OF THE DATASET

In order to build hybrid systems, it is better to have a dataset with a large number of questions, that are long enough so that they would more probably require a hybrid reasoning. It is thus required that the questions be linked to a reference in a knowledge base.

Our intent is to augment an existing corpus, and not to build a new one from scratch. To obtain such a corpus, two approaches are possible. The first one consists in using a corpus of questions for knowledge base QA systems and adding texts related to the pairs. The available datasets of that type contain either too short questions (WebQuestions) or too few questions (QALD). The second approach consists in using a textual QA dataset and complementing it with information that come from the knowledge base. We chose this second approach because the corpus of questions on texts that are available (TREC and CLEF) are often about information present in knowledge bases (DBpedia). Part of these questions are also complex and long enough to make it possible to build a hybrid system.

4.1 Corpus sources

Name	Number of questions
Clef 2004 : MultiEight	700
Clef 2005 : MultiNine	200
Trec 1999 -> 2007	3400
total	4300

Table 1: Corpus sources

We selected factual questions in CLEF and TREC datasets. We kept the datasets of CLEF 2004 and 2005 (cf [14] and [19]) that concern newspapers from 1994 and 1995. They enclose several kinds of manually written and translated questions: factual questions, "how" questions and definition questions. The factual questions types are: TIME, MEASURE, PERSON, ORGANISATION, LOCATION, OBJECT, MANNER, OTHER. Definition questions are limited to persons and organizations.

The TREC datasets (cf [20], [1], [21] and [22]) are based on similar documents. The questions were written manually in Trec 1999 and conceived from logs for the campaigns of 2000 to 2004. They cover factual and definition questions.

4.2 Augmentation of the corpus

In order to determine if questions are hybrid, we can try to determine whether it is possible to solve them only with the knowledge base. A first step consists in determining 1) if an entity of the knowledge base answers the question; 2) if an entity of the knowledge base is the main entity (the focus) in the question; 3) if a path of the KB links the two entities.

Several cases can then be possible:

- the answer can be identified as a knowledge base entity, but no path can be found with an entity of the question: the knowledge base and the texts could be combined to find the answer;
- an entity can be identified in the question, but the entity of the answer cannot be identified: then the entities could help the resolution by using the knowledge base, but the resolution must be done with text;
- a path can be found: then the question might be solved completely by the knowledge base.

These cases are all potentially hybrid: the third case can be solved using only the knowledge base but in practice information from texts can help to find an answer.

Thus the next step is to augment each question-answer pair with data from DBpedia: answer URI, question entity URIs, paths between these two entities (named KB path in the following).

[8] on 9 227 questions for text QA that were annotated by a named entity recognizer, about 59 % of questions mention an entity

In order to distribute a more self-contained corpus, we also augment the pairs with a related Wikipedia paragraph if found. To do that, we use the code released by [5].

We will now define more closely a KB path. First, a path that leads to the answer, called *answer path*, is a finite sequence $((e_n, r_n, e_{n+1}))_n$ with:

- $n \in [0, N[$, N is the length of the path,
- e_n is either an entity of the KB e_n^K represented as an URI, or an entity of a text e_n^T represented as a sequence of words,
- r_n is either a relation of the knowledge base r_n^K represented as an URI, or a relation of a text r_n^T represented as a sequence of words,
- e_0 is the focus entity in the question while e_N is the answer.

The answer path allows to answer the question but note that an answer path does not necessarily contain all the information given in the question; it is not a representation of the meaning of the question, neither a full justification of the answer and does not correspond to a query as in QALD corpora. The supplementary information in the question will help to find it, using text or KB.

A hybrid answer path contains both kinds of entity: at least one entity from a knowledge base and one from a text.

Given these definitions, a hybrid pair (q, a), with q a question and a an answer, which potentially requires a hybrid resolution is a pair that is associated to a hybrid answer path. Thus, a hybrid resolution could consists in finding a part of the answer path in KB, i.e. a KB path, and complete missing information using texts or in using textual information outside the answer path for selecting the answer path. We will give some examples below.

A KB path links a question entity to the answer entity, so it has the same extremities as the answer path. For evaluating its relevance, we will do it in reference to the answer path.

In order to annotate pairs with the KB information, firstly answers and questions are associated with URIs of the entities they mention by using DBpedia Spotlight. We then compute the KB paths at one and two steps between the pairs of entities by querying DBpedia.

- A one-step KB path (e_q, p,e_r) is made of one entity of the question, e_q, the answer entity, e_r, and a predicate, p between the two. For example in the question *Who is Shimon Peres?* whose answer is *Israeli Foreign Affairs Minister*, the entities e_q *res:Shimon_Peres* and e_r *res:Ministry_of_Foreign_Affairs_(Israel)* are identified respectively in the question and the answer and the path with the predicate *dbp:office* is found and allows to answer the question.
- A two-step KB path is made of 2 entities e_q and e_r, 2 predicates p_1 and p_2 and an intermediary entity e_i : ((e_q,p_1,e_i), (e_i,p_2,e_r)). A 2-step path means that the question is complex enough and needs some reasoning. We limit the paths to a length of 2 for a computational reason and because we think that, in general, paths longer than 2 steps are not relevant and are not part of the answer path.

In order to examine the relevance of these annotations, we validated them manually. We first validated the annotation of the answers with the found URIs. Then, for every validated entities, we examined the KB paths. We do not explicit the answer path, as it would have to be done manually. Thus, when we decide the validity of a KB path as a possible sub-path of an answer path, we have to decide if some words of the question can be associated with some predicates found in the KB path, i.e. if they could be a mention of the relation.

The kinds of possible annotation of one-step or two-step KB paths are:

- Correct: the KB path is an answer path. Several cases can occur:
  - the KB path is an answer path, i.e. the predicates correspond to question words, and it is fully justified, i.e. contains all the information given in the question. For example (CERN, headquarters, Canton_of_Geneva), (Canton_of_Geneva, capital, Geneva) correctly answers the question *Where is the CERN located?* with Geneva as answer
  - the KB path is an answer path and does not cover all the information given in the question. For example, (James_Bond, portrayer, Pierce_Brosnan) correctly answer *Who plays James Bond in the latest film of the 007 series?* with Pierce Brosnan as answer [9], however the fact that it is the last one is not in the path.
- partial path: the KB path is a sub-path of the answer path: the predicate is not precise enough and requires to be completed by other information. For example, (Kim_Il-sung, allegiance, North_Korea) partially answers *Who was the president of North Korea before 1994?* with Kim Il-sung as answer. The information that Kim Il-sung is a president is more precise than allegiance.
- related path: the predicates found are related to the question words but do not correspond to a sub-path of the answer path.
  - the relation matches some information given in the question, but does not belong to the answer path. For example (Java, designer, Sun_Microsystems) is relevant to the question *What will Microsoft license from SUN?* with answer Java, as it makes the connection between Java and SUN, but it does not answer the question about Microsoft license.
  - as the preceding case, but with a too vague relation or a relation topically related to the question, i.e. the relation is not completely out of topic.
- incorrect path: the KB path is not in relation with the answer path, i.e. all the other cases. For example (Mississippi, flower, Magnolia) does not answer *What is the nickname of the state of Mississippi?* with answer Magnolia, as flower cannot be matched with some question words. This relation could be useful in some reasoning, for example if we know that often a nickname of a state is its associated flower, however we can expect that we will find the direct relation in texts. Thus incorrect path means that the question is better solved with textual information.

A lot of two-step paths have been found but only a small part is correct. In order to shorten the annotation time, we only examine

[9] Some questions implicitly involve a time stamp, as with latest, and the answer may not be correct anymore.

two-steps paths of questions that do not have a validated one-step path. Moreover, some irrelevant relations have been removed: subject, align, direction, seeAlso which are not semantic relations and often yield false paths.

5 ANALYSIS OF THE ANNOTATED CORPUS

We computed some statistics on the corpus after validation that are given in Table 2.

Kind of question	Number of questions
2-step path	290
1-step path	269
the answer is a URI	1699
total	4300

Table 2: Corpus statistics

There are around 1700 questions that have a connection with DBpedia and text. Among them, around 560 are annotated with a useful KB path for their resolution.

Our corpus can be used for hybrid QA but also for KB QA in case of correct KB paths. As 1-length answer paths are usually found in corpus for KB QA, we wanted to know whether the 2-path questions bring new interesting cases that are not usually part of the existing KB corpora.

We found three categories of such questions:

(1) the answer requires to do some inferences;
(2) the answer requires some geographical reasoning;
(3) the resolution needs a complex adaptation to the schema of the knowledge base.

5.1 Inferences

Some questions can only be answered by human-made inference obtained by relation composition according to the meaning of the relations involved.

- Question : Name the children of Sani Abacha
- Answer : Mohammed Abacha
- Path :
  - (Sani_Abacha, *spouse*, Maryam_Abacha)
  - (Maryam_Abacha, *child*, Mohammed_Abacha)

This kind of question is usually answered by a single triple: the entity of the person, the relation *child* and the entity of its child. Our DBpedia version does not contain this information for Sani Abacha.

However the path from *Sani Abacha* to *Mohammed Abacha* that goes through the spouse relation of *Sani Abacha* has been found. Answering that question with the KB means inferring that the children of his wife are also his children.

- Question : What is Jane Goodall known for?
- Answer : London-born primatologist
- Path :
  - (Dian_Fossey, *influences*, Jane_Goodall)
  - (Dian_Fossey, *fields*, Primatology)

The field of Jane Goodal is not directly available in DBpedia, but the fact that she was influenced by Dian Fossey who is a primatologist makes it possible to infer she is a primatologist too.

In these cases, the question words do not explicitly refer to the two relations, and requires a complex treatment for finding them.

5.2 Geographical relations

For some questions that expect a type of location as answer, some geographical inference may be needed to find a precise location.

- Question : Where is the Leaning Tower?
- Answers : Pisa
- Path :
  - (Leaning_Tower_of_Pisa, *province*, Province_of_Pisa)
  - (Province_of_Pisa, *seat*, Pisa)

To link Pisa to the *Leaning_Tower_of_Pisa*, the path goes through *Province_of_Pisa* which has a relation to *Pisa*. Province of Pisa could be an accepted answer, but being able to find the Pisa entity is more precise. This example does not picture a general rule for geographical inference (the reasoning for finding the precise location of the Leaning Tower will not apply in other cases)

5.3 Adaptation to a complex knowledge base schema

One of the main challenges in KB QA is to match question words to the relevant relations. This is particularly difficult when this matching involves a one to many or a many to one correspondence.

- Question : What is the name of the chairman of the Federal Reserve Board?
- Answers : Alan Greenspan
- Path :
  - (Federal_Reserve_Board_of_Governors, *leaderTitle*, Chair_of_the_Federal_Reserve)
  - (Alan_Greenspan, *title*, Chair_of_the_Federal_Reserve)

The answer path goes through *Chair_of_the_Federal_Reserve* to find *Alan Greenspan*. Chairman is involved in an entity name, Chair_of_the_Federal_Reserve and in relation names, leaderTitle and title, which makes it hard to find.

- Question : What party does Edouard Balladur represent?
- Answers : conservative
- Path :
  - (Edouard_Balladur, *party*, Union_for_a_Popular_Movement)
  - (Union_for_a_Popular_Movement, *ideology*, Conservatism)

The expected answer for this question is *conservative*. In fact, it is not the name of the party but its ideology. A direct triple between *Édouard_Balladur* and *Conservatism* cannot be found but it is possible to find a path between these 2 entities by going through the *Union_for_a_Popular_Movement* entity. According to the granularity of the expected answer, the word party has to be linked to two relations.

6 CONCLUSION

QA systems are generally dedicated to search in a single kind of source, a knowledge base or texts. However, using the two kinds of

resources would lead to build more powerful systems. A challenge is then to study which kinds of phenomena occur, which kind of cooperation can be useful and how to leverage QA methods. For these purposes, we enrich automatically a textual corpus conceived for textual QA with KB annotations (KB entities and a path between them of length 1 or 2) in order to complement textual triples (question, answer, passage) with relevant KB material. While paths made of one relation are often correct, 2-length paths need to be curated. After annotation, the corpus contains around 1700 questions that have a connection with DBpedia and text. Among them, around 560 are annotated with a KB path useful for their resolution. We also showed that searching for a KB solution to questions that have been asked in a text retrieval context leads to propose new kinds of questions that require complex reasoning. For future work, we envisage to explore the automatic curating of the corpus, in order to define a distant supervision process that does not generate too much noise and allows for building larger corpora.

ACKNOWLEDGMENTS

The work is partly supported by the ANR project GoAsQ (ANR-15-CE23-0022) and the FUI project Pulsar.

REFERENCES

[1] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. 1999. *Modern information retrieval*. Vol. 463. ACM press New York.

[2] Romain Beaumont, Brigitte Grau, and Anne-Laure Ligozat. 2015. SemGraphQA@QALD5: LIMSI participation at QALD5@CLEF. In *Working Notes of CLEF 2015*.

[3] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic parsing on freebase from question-answer pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 1533–1544.

[4] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*. 2787–2795.

[5] Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. 2017. Reading Wikipedia to Answer Open-Domain Questions. In *Proceedings of ACL 2017*. 1870–1879.

[6] Jennifer Chu-Carroll, James Fan, BK Boguraev, David Carmel, Dafna Sheinwald, and Chris Welty. 2012. Finding needles in the haystack: Search and candidate generation. *IBM Journal of Research and Development* 56, 3.4 (2012), 6–1.

[7] Silviu Cucerzan and Eugene Agichtein. 2005. Factoid Question Answering over Unstructured and Structured Web Content.. In *TREC Proceedings*, Vol. 72. 90.

[8] Dennis Diefenbach, Vanessa Lopez, Kamal Singh, and Pierre Maret. 2017. Core techniques of question answering systems over knowledge bases: a survey. *Knowledge and Information Systems* (2017), 1–41.

[9] Martin Gleize and Brigitte Grau. 2015. A Unified Kernel Approach for Learning Typed Sentence Rewritings. In *Proceedings of ACL-IJCNLP*.

[10] Arnaud Grappy, Brigitte Grau, Mathieu-Henri Falco, Anne-Laure Ligozat, Isabelle Robba, and Anne Vilnat. 2011. Selecting answers to questions from Web documents by a robust validation process. In *Proceedings of Web Intelligence Conference (WI)*.

[11] Hua He and Jimmy Lin. 2016. Pairwise Word Interaction Modeling with Deep Neural Networks for Semantic Similarity Measurement. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, 937–948. http://www.aclweb.org/anthology/N16-1108

[12] Wesley Hildebrandt, Boris Katz, and Jimmy J Lin. 2004. Answering Definition Questions Using Multiple Knowledge Sources.. In *HLT-NAACL*. 49–56.

[13] Denis Lukovnikov, Asja Fischer, Jens Lehmann, and Sören Auer. 2017. Neural network-based question answering over knowledge graphs on word and character level. In *Proceedings of the 26th international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1211–1220.

[14] Bernardo Magnini, Alessandro Vallin, Christelle Ayache, Gregor Erbach, Anselmo Peñas, Maarten De Rijke, Paulo Rocha, Kiril Simov, and Richard Sutcliffe. 2004. Overview of the CLEF 2004 multilingual question answering track. In *Workshop of the Cross-Language Evaluation Forum for European Languages*. Springer, 371–391.

[15] Seonyeong Park, Soonchoul Kwon, Byungsoo Kim, and Gary Geunbae Lee. 2015. ISOFT at QALD-5: Hybrid question answering system over linked data and text data. In *Working Notes of CLEF 2015*.

[16] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250* (2016).

[17] Christina Unger, Lorenz Bühmann, Jens Lehmann, Axel-Cyrille Ngonga Ngomo, Daniel Gerber, and Philipp Cimiano. 2012. Template-based question answering over RDF data. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 639–648.

[18] Christina Unger, Corina Forascu, Vanessa Lopez, ACN Ngomo, E Cabrio, Philipp Cimiano, and Sebastian Walter. 2014. Question Answering over Linked Data (QALD-4). CLEF Conference.

[19] Alessandro Vallin, Bernardo Magnini, Danilo Giampiccolo, Lili Aunimo, Christelle Ayache, Petya Osenova, Anselmo Peñas, Maarten De Rijke, Bogdan Sacaleanu, Diana Santos, et al. 2005. Overview of the CLEF 2005 multilingual question answering track. In *Workshop of the Cross-Language Evaluation Forum for European Languages*. Springer, 307–331.

[20] Ellen M Voorhees. 2001. Overview of TREC 2001 Question Answering Track.. In *TREC Conference*.

[21] Ellen M Voorhees. 2003. Overview of the TREC 2003 Question Answering Track.. In *TREC Conference*, Vol. 2003. 54–68.

[22] Ellen M Voorhees. 2004. Overview of the TREC 2004 Question Answering Track. (2004).

[23] Mengqiu Wang and Christopher D Manning. 2010. Probabilistic tree-edit models with structured latent variables for textual entailment and question answering. In *COLING*.

[24] Kun Xu, Yansong Feng, Songfang Huang, and Dongyan Zhao. 2016. Hybrid Question Answering over Knowledge Base and Free Text. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee, Osaka, Japan, 2397–2407. http://aclweb.org/anthology/C16-1226

[25] Kun Xu, Siva Reddy, Yansong Feng, Songfang Huang, and Dongyan Zhao. 2016. Question Answering on Freebase via Relation Extraction and Textual Evidence. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 2326–2336. http://www.aclweb.org/anthology/P16-1220

[26] Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, and Gerhard Weikum. 2013. Robust question answering over the web of linked data. In *Proceedings of CIKM*.

[27] Xuchen Yao and Benjamin Van Durme. 2014. Information extraction over structured data: Question answering with freebase. In *Proceedings of ACL*.

[28] Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch, and Peter Clark. 2013. A Lightweight and High Performance Monolingual Word Aligner.. In *Proceedings of ACL*. 702–707.

[29] Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch, and Peter Clark. 2013. Semi-Markov Phrase-Based Monolingual Alignment.. In *Proceedings of EMNLP*. 590–600.

[30] Lei Zou, Ruizhe Huang, Haixun Wang, Jeffer Xu Yu, Wenqiang He, and Dongyan Zhao. 2014. Natural language question answering over RDF: a graph data driven approach. In *Proceedings of SIGMOD*.