Question Answering (QA)

MAI – HLT

MERITXELL GONZÀLEZ

Outline

- Introduction
- A bit of History of QA
- General QA systems
- Factoid QA
- Beyond Factoid QA
- Domain Restricted QA
- QA over Linked Open Data
- Open domain QA

Introduction

- **QA** can be defined as the task of given a user information need, expressed as a NL question (or Controlled NL) provide to the user the correct answer to the question

- … As opposed to a set of documents where the answer can be found (IR systems).

Introduction

- Many NLP sub-tasks are involved in QA and many approaches can be followed for approaching these tasks.
- [Weston at al, 2015] (Fabebook AI Research) have proposed a framework and a set of synthetic tasks with the aim of helping to develop learning algorithms for text understanding and reasoning and applying them to QA.
  - The goal is to categorize different kinds of questions into skill sets, which become their tasks.
  - The tasks are publicly available at http://fb.ai/babi [paper]
  - Source code to generate the tasks is available at https://github.com/facebook/bAbI-tasks.

Introduction

1. Single Supporting Fact

   Mary went to the bathroom. Where is Mary?

2. Two Supporting Facts

   John is in the playground. John picked up the football. Where is the football?

3. Three Supporting Facts

   John picked up the apple. John went to the office. John went to the kitchen. Where was the apple before the kitchen?

4. Two Argument Relations

   The office is north of the bedroom. The bedroom is north of the bathroom. What is the bedroom north of?

Introduction

5. Three Argument Relations

   Mary gave the cake to Fred. Fred gave the cake to Bill. Jeff was given the milk by Bill. Who did Fred give the cake to?

6. Yes/No Questions

   John moved to the playground. Is John in the playground?

7. Counting

   Daniel picked up the football. Daniel dropped the football. Daniel got the milk. Daniel took the apple. How many objects is Daniel holding?

8. Lists/Sets

   Daniel picks up the football. Daniel drops the newspaper. Daniel picks up the milk. What is Daniel holding?

Introduction

9. Simple Negation

   Fred is no longer in the office. Sandra is in the garden. Is Fred in the office?

10. Indefinite Knowledge

    John is either in the classroom or the playground. Is John in the classroom?

11. Basic Coreference

    Daniel was in the kitchen. Then he went to the studio. Where is Daniel?

12. Conjunction

    Mary and Jeff went to the kitchen. Then Jeff went to the park. Where is Mary?

Introduction

13. Compound Coreference

Daniel and Sandra journeyed to the office. Then they went to the garden.
Where is Daniel?

14. Time Reasoning

In the afternoon Julie went to the park. Yesterday Julie was at school.

Julie went to the cinema this evening. Where did Julie go after the park?

15. Basic Deduction

Sheep are afraid of wolves. Cats are afraid of dogs. Mice are afraid of cats. Gertrude is a sheep. What is Gertrude afraid of?

16. Basic Induction

Lily is a swan. Lily is white. Greg is a swan. What color is Greg?

Introduction

17. Positional Reasoning

    The triangle is to the right of the blue square. The red square is on top of the blue square. The red sphere is to the right of the blue square. Is the red square to the left of the triangle?

18. Size Reasoning

    The football fits in the suitcase. The suitcase fits in the cupboard. The box is smaller than the football. Will the box fit in the suitcase?

19. Path Finding

    The kitchen is north of the hallway. The bathroom is west of the bedroom. The den is east of the hallway. The office is south of the bedroom. How do you go from den to kitchen?

20. Agent's Motivations

    John is hungry. John goes to the kitchen. John grabbed the apple there.

    Daniel is hungry. Where does Daniel go?

Introduction

- QA systems can be viewed as a natural extension of IR systems.
  - In IR systems user information needs are expressed through a query, usually consisting on a set of keywords
  - The query is then used for retrieving from a knowledge base
    - a collection of documents
    - the whole Web
    - a domain restricted collection
    - a corporative textual database
  - The output of IR consists of a, sometimes ranked, set of documents retrieved from the knowledge base.

Introduction

- If the query is well formulated, a good IR system should retrieve in its best ranked documents those satisfying the user information needs.
- In QA systems, the query consists of a NL question and the output is the exact answer to the query.

Introduction

- Instead of allowing full NL expressivity, some NL applications prefer to work with controlled NL (CNL)
  - basically those heavily based on human-computer interaction, as QA, dialog systems, etc.

- Using CNL results on more robust interfacing, reduces the ambiguity inherent to NL and improves the parsing performance of the systems.
  - There exist generic CNL but also domain-restricted CNL.

Introduction

- Currently, CNL systems are applied only to English
- The term Controlled English is frequently used for referring to these systems
- CNL implies several tasks:
  - Defining a grammar for a controlled language
  - Linguistic Engineering for building, testing, and maintaining the grammars
  - Tuning the grammars to new domains
  - Parsing

Outline

- Introduction
- A bit of History of QA
- General QA systems
- Factoid QA
- Beyond Factoid QA
- Domain Restricted QA
- QA over Linked Open Data
- Open domain QA

A bit of History of QA

- The origin of QA can be found in the eighties of last century, with the development of many **NLI to computer applications**, especially NLI to databases.

- The term QA was started to be used in the framework of QA tracks within TREC challenges (starting with TREC-8 in 1999).

- Modern QA systems acquired a multilingual dimension in the framework of CLEF (from 2003) challenges, and currently included into the framework of TAC (from 2008).

A bit of History of QA

- Some QA systems accessible through the Web are START (MIT), AnswerBus, Webclopedia (ISI), Ask (before AskJeeves), LCC, PowerAnswer, **IBM's Watson, or Wolfram Alpha.**

- Conventional IR systems use basically statistical approaches, QA systems are, increasingly, evolving towards more complex questions, NLP techniques, both for processing the question and for extracting the answer.

A bit of History of QA

Related disciplines are

- IR
- Information routing, filtering, harvesting
- Answer Finding: able to discover in a collection of question/answer pairs (as FAQ), the questions closest to the original one for retrieving the corresponding answer.
- Paraphrase detection
- Textual Entailment
- Information Integration
- Knowledge Base Population

A bit of History of QA

- Initially QA was limited to Factoid Questions (Factoid QA), where the questions consisted on asking for a fact.
- Answering to these questions was not specially challenging because an assertive formulation of the answer is likely to be found in the collection.
  - For instance, for the question "Where was Barack Obama born", the assertion "Obama was born in Hawaii" probably occurs in the collection.

One of the lines of research in QA was to increase the complexity and expressivity of the questions.

Outline

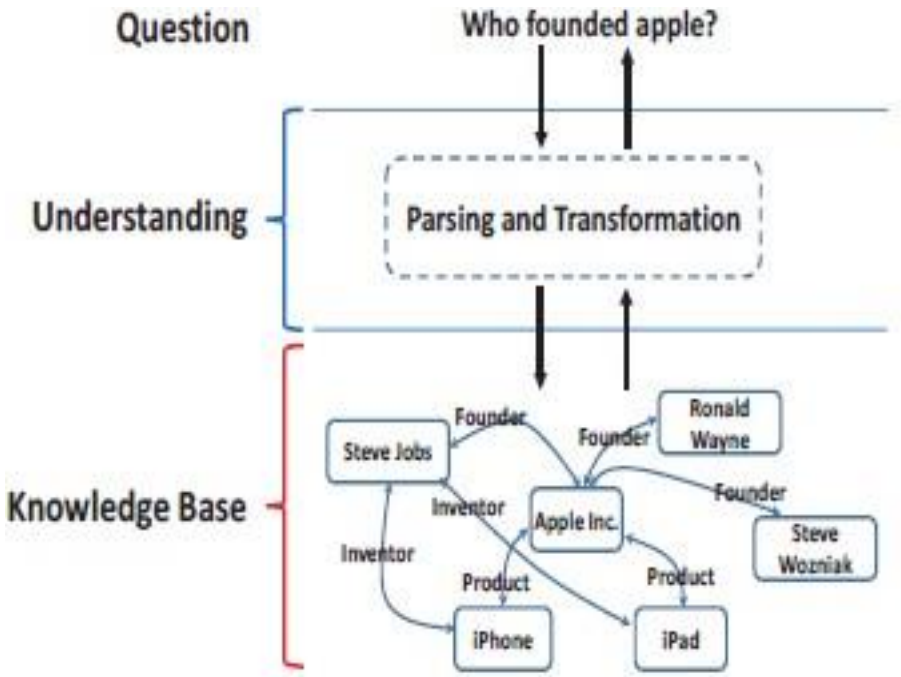
- Introduction
- A bit of History of QA
- General QA systems
- Factoid QA
- Beyond Factoid QA
- Domain Restricted QA
- QA over Linked Open Data
- Open domain QA

General QA systems

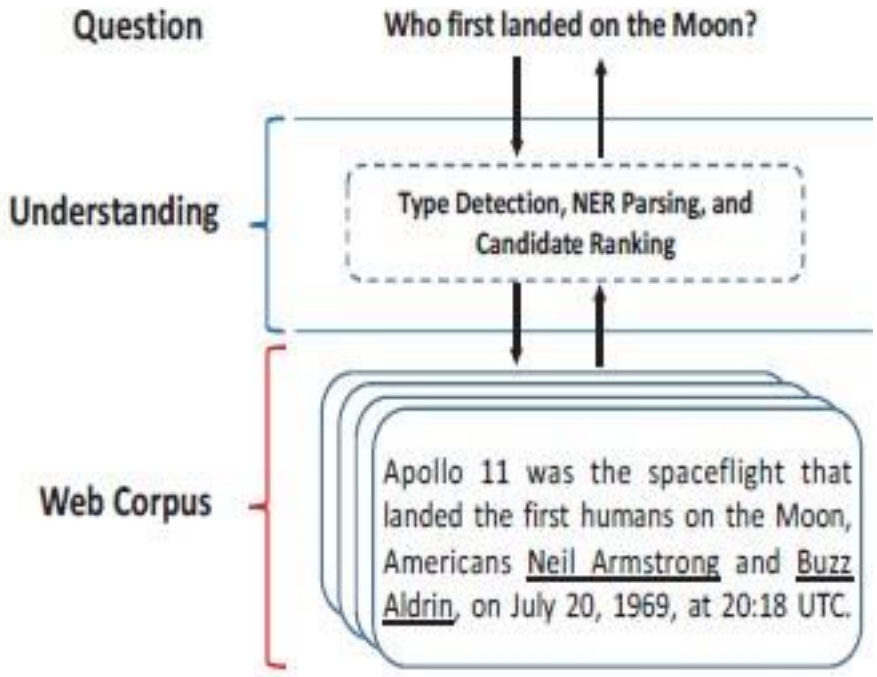
- Initially Conventional QA systems use to be structured into four modules (or steps):
  - Question Processing
  - IR of relevant documents
  - IR of relevant passages (fragments)
  - Answer Extraction.
- So, the language technologies involved have to cover these 4 tasks:
  - Linguistic analysis of questions using general or specific parsers (and grammars).

    This includes the definition of an appropriate tagset for classifying the questions (Question Type, QT)!
  - IR engines (general or specific for the task)
  - IE techniques for extracting the answer

General QA systems



(a) KB-based QA Systems



(b) Web-based QA Systems

Outline

- Introduction
- A bit of History of QA
- General QA systems
- Factoid QA
- Beyond Factoid QA
- Domain Restricted QA
- QA over Linked Open Data
- Open domain QA

Factoid QA

For “simple” and “factoid” we mean that the question asks for a fact, the formulation of the question is simple (there are no additional constraints) and the answer can be retrieved from a single document with no additional processing.

Most of the available QA systems satisfy this definition.

Factoid QA

Some examples of Factual Questions from TREC 8 Contest
How much folic acid should an expectant mother get daily?
Who invented the paper clip?
What university was Woodrow Wilson president of?
Where is Rider College located?
Name a film in which Jude Law acted.
Where do lobsters like to live?
Who was Picasso?

Beyond Factoid QA

- One of the lines of research in QA was to **increase the complexity and expressivity of the questions.**

- Some of these lines are the following:
  - **Why QA**, is the task of retrieving answers for a given Why Question, as "Why are tsunamis generated?".
  - An in depth analysis of causality, causal relations, causal inference is needed.

Beyond Factoid QA

- **Definitional QA** (DQA), where the answer probably has to be synthesized from partial pieces of information extracted from several documents.
  - Definitions can refer to people, “Who was Unamuno”, organizations, “what is the FAO”, or terms, “what is synonymy”.
  - As many DQA systems synthesize definitions from multiple sources NL Generation techniques, Slot Filling, Entity Linking and IE are widely used.

Beyond Factoid QA

- **List QA**, where the answer consists of a list.
- In some cases the whole list can be found in a document but frequently the members of the list have to be collected from different documents: “French presidents after World War II”
- Linked questions: “Who was Picasso? When and where was he born and dead?”

Beyond Factoid QA

- **Dialog** based QA:

USR: Time span of Georges Bush presidency

SYS: There are two USA presidents named Georges Bush, which do you refer to?

- The form of dialog usually implies harder forms of anaphora and the use of more NLP techniques.
  - Anaphora resolution
  - Dialog management
  - Dialog grammars
  - NL Generation

Beyond Factoid QA

- **Time and space** constrained questions: “Who was the second USA republican president after the Vietnam war?”.

  - For answering this question a QA system should probably split the complex question into a set of related simpler questions:

“When did the Vietnam war ends?”, giving Date_1, “USA presidents after Date_1”, giving Person_1, Person_2, Person_3,

Beyond Factoid QA

- **Opinion QA**. Starting in TAC challenges and following the current trends in Opinion Mining and Sentiment Analysis, Opinion QA is currently subject of very active research.
  - Consider the question "Why do people enjoy Starbucks coffee?"
  - Correctly answering this question implies not only locating a candidate answer but also check whether it is an opinion and whether its polarity is positive.
  - Other examples are "What Quevedo thought about Góngora?", "Arguments pro and against arms control in USA".

Beyond Factoid QA

- Additional technologies are needed for dealing with this task:
  - The Question processing step is more complex because not only the QT and EAT (expected answer type) have to be recognized but also the constraints related to opinion mining, sentiment analysis, and polarity detection.

- In the answer extraction step some issues arise:
  - i) classifying a detected assertion as informative, opinion, etc.
  - ii) extracting its polarity,
  - iii) detecting the subject of the opinion, etc.

Frequently the opinion information has to be extracted from social nets. Processing this kind of documents needs specialized linguistic processors.

Outline

- Introduction
- A bit of History of QA
- General QA systems
- Factoid QA
- Beyond Factoid QA
- Domain Restricted QA
- QA over Linked Open Data
- Open domain QA

DRQA

- **Domain restricted QA, DRQA**. Where both questions and search space are restricted to a given domain.
- Many domains have been faced, geographic, tourism, economics, etc. Perhaps the domain object of the most applications is the **medical**.
- DRQA are applied to specific tasks and use domain specific lexicons, terminologies, KBs, ontologies, ….
- Search spaces are smaller and so approaches based on the redundancy of answers (as votation techniques) are useless.
- User's requirements use to be high, and system performance is more precision than recall oriented (no answer is better than a wrong answer).
- Questions and documents are challenging and frequently contain acronyms, non-textual content (tables, itemized lists, etc.), domain specific jargon, etc.

DRQA

Some examples from BioAsk 2018		
Is Hirschsprung disease a mendelian or a multifactorial disorder?	Summary	Disease
List signaling molecules (ligands) that interact with the receptor EGFR?	List	Biology
Is the protein Papilin secreted?	Yesno	Biology
Are long non coding RNAs spliced?	Yesno	Biology
Is RANKL secreted from the cells?	Yesno	Biology

DRQA

- Clinical question answering (Clinical QA).
  - [Demner-Fushman et al, 2009]

  - Questions occurring in clinical situations could pertain to:
    - Information on particular patients
    - Data on health and sickness within the local population
    - Medical knowledge
    - Local information on doctors available for referral
    - Information on local social influences and expectation
    - Information on scientific, political, legal, social, management, and ethical changes affecting both how medicine is practiced and how doctors interact with individual patients

Outline

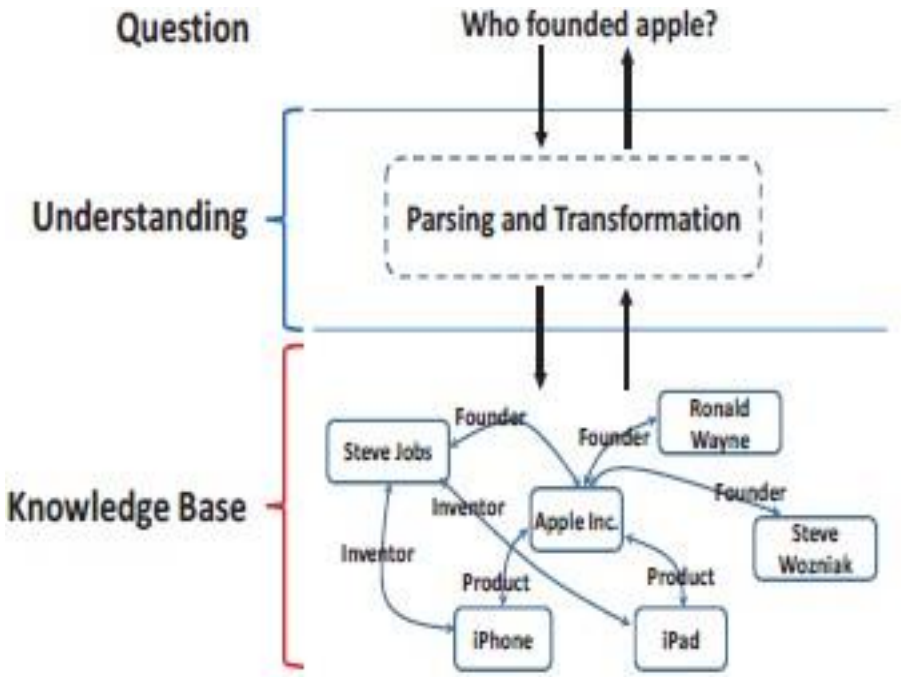
- Introduction
- A bit of History of QA
- General QA systems
- Factoid QA
- Beyond Factoid QA
- Domain Restricted QA
- QA over Ontologies & Linked Open Data
- Open domain QA

Q&A over Ontologies

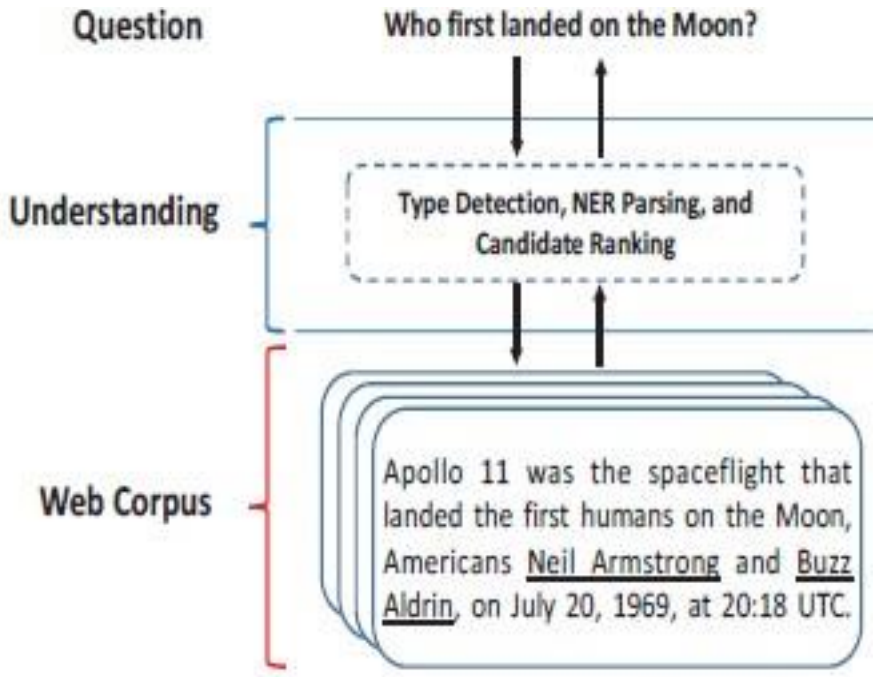
- **QA over domain ontologies**, Ontology-based QA (ObQA).

- In this case the answers are looked up not in free text documents but in ontologies taking profit not only of the linguistic (terminological) data included into the ontology but also over their relations, properties, and inferential capabilities.

General QA systems



(a) KB-based QA Systems



(b) Web-based QA Systems

Q&A over Linked Data

- In the framework of the **Semantic Web** there has been recently a huge growth of available **open and closed domain resources**.
- Many of these resources are included into the **Linked Open Data (LOD) initiative.**
- Relevant resources in LOD are FreeBase and DBPedia as open domain LOD, and BioPortal (medical and genomic) or LinkedGeoData (geographic) as closed domain LOD.
- QA systems using LOD as search space are referred as **QALD**

Q&A over LOD

- Open-domain Question Answering
- Answer question on any topic
  - query a KB with natural language
  - Semantic Representation = KB entities + relations

Q&A over LOD: challenges

- Difficulties for dealing with RDF datasets
- Mapping questions into SPARQL is not an easy task. Several problems arise and have to be faced:
  - Different namespaces coexist in DBpedia, some of them belonging to DBpedia itself, and others corresponding to links from DBpedia to other ontologies, as Yago.
  - For instance, looking for the generic term 'Mountain' we find 217 categories in Yago namespace (e.g. http://dbpedia.org/class/yago/Mountain109359803), 10 DBpedia properties (e.g. http://dbpedia.org/ontology/highestMountain), and 6 DBpedia ontology categories (e.g. http://dbpedia.org/ontology/Mountain)

Q&A over LOD: challenges

- Lack of coherence in the nomenclature used for naming DBpedia entries (classes, properties and instances).
- Use of lower/upper case, singular/plural forms, abbreviations, order of simple components ofthe multi-word expressions, inclusion of parenthesis, underscores, and other orthographic marks is rather arbitrary or at least difficult to interpret.
- The following properties (among many others) were found in DBpedia when looking for number of members:
    - http://dbpedia.org/property/memberNo
    - http://dbpedia.org/property/members
    - http://dbpedia.org/property/member
    - http://dbpedia.org/property/numMembers
    - http://dbpedia.org/property/membersNumbers
    - http://dbpedia.org/property/noOfMembers

Q&A over LOD

- The habitual clash when mapping terms of the NL expression of the question into terms of the ontology is obviously present. The habitual problems of polysemy (a term of the question can be mapped to many terms of the ontology : classes, properties, and instances) and synonymy (an ontology term can be referred by different question terms) frequently occur.

- The directionality of the relations in the ontology is not always clear. For instance, it is not clear whether http://dbpedia.org/property/mayor/ links a city to a person or a person to a city.

- Depending on the Question Type (QT), the EAT and the complexity of the question, partially reflected in the constraints provided by the Question Processing module, resolving the mapping can be more or less difficult.

Outline

- Introduction
- A bit of History of QA
- General QA systems
- Factoid QA
- Beyond Factoid QA
- Domain Restricted QA
- QA over Linked Open Data
- Open domain QA

Open domain Q&A

New and big datasets: allows the use of Deep Learning Techniques

dataset	size	description
quasarT	43000	Open-domain trivia questions.
quasarS	37000	Fill-in-the-gap queries constructed from definitions of software entity tags on Stack Overflow.
SQuAD1.1	98169	Stanford Question Answering Dataset (SQuAD) is a reading comprehension data set, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable.
SQuAD2.0	142192	Combines the 100,000 questions in SQuAD1.1 with over 50,000 new, unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering.

Open domain Q&A

dataset	size	description
TREC (Curated)	2180	TREC questions with answers added.
WikiMovies	181679	This includes only the QA part of the Movie Dialog data set, but using three different settings of knowledge: using a traditional knowledge base (KB), using Wikipedia as the source of knowledge, or using IE (information extraction) over Wikipedia. This allows to test the ability of models to directly read documents to answer questions, and to compare this to traditional KBs in the same setting.
TriviaQA	109767	reading comprehension data set containing over 650K question-answer-evidence triples. TriviaQA includes 95K question-answer pairs authored by trivia enthusiasts and independently gathered evidence documents, six per question on average, that provide high quality distant supervision for answering the questions.

Open domain Q&A

dataset	size	description
Quora	404302	Allows to train and test models of semantic equivalence, based on actual Quora data. The dataset consists of over 400,000 lines of potential question duplicate pairs. Each line contains IDs for each question in the pair, the full text for each question, and a binary value that indicates whether the line truly contains a duplicate pair.
WikiQA	29261	Set of question and sentence pairs, collected and annotated for research on open-domain question answering. In addition, the WikiQA data set also includes questions for which there are no correct sentences, enabling researchers to work on answer triggering, a critical component in any QA system.

Open domain Q&A

Neural Models

- A. Fader, L. Zettlemoyer & O. Etzioni. ACL, 2013:
  - Paraphrasing
- A. Fader, L. Zettlemoyer & O. Etzioni. KDD, 2014
  - Embedding of both questions and LOD triples
- Y. Kim.
  - Convolutional neural networks for sentence classification.
- C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, 2015
  - combined architecture of CNN and LSTM
- D. Chen, A. Fisch, J. Weston, and A. Bordes, 2017
- W.-t. Yih, M.-W. Chang, C. Meek, and A. Pastusiak, 2013
  - Alignments
- C. Xiong, V. Zhong, and R. Socher, 2016.
  - Dynamic coattention networks