**Summary:**

* Two possibilities for factoid QA: on 1) structured DB (e.g., Wikidata) [1,3,4,5], 2) un- or semi-structured texts (e.g., Wikipedia) [7].
* 1) To evaluate factoid QA systems over Knowledge-Graph databases (QA-KG), benchmark datasets are usually generated by exporting facts (KG content) and having annotators create questions with corresponding answers [1,4]. Examples: SimpleQuestions dataset [3,4].
* 2) To factoid QA systems over un- or semi-structured texts, annotators create questions on the texts with corresponding answers pointing to specific passages in the text [7].
* Often Neural Networks are used for answer prediction [4,5].
* MathQA evaluation should be oriented on QAPD [1], Platypus [3], and SimpleQuestions dataset [4].
* Need to create a novel benchmark on seeding list from AnnoMathTeX paper (persisted in Mathmlben).

---

**[1] QAPD: an ontology-based question answering system in the physics domain**

We generate a set of questions according to the ontology domain. For this purpose, 10 human experts were asked to query our representation of the physics domain (Electricity and Electromagnetism). The ontology, together with a list of real entities and attributes extracted from our data source are shown to them in order to generate questions for any kind of data they were interested in. The questions include real data instances for the concepts defined in the ontology (i.e., “what is the unit of resistance?”, “is ohm the unit of the resistance?” and “what is transformer?”). 2. This step aims to identify and categorize the entity instances, and then replace them by their corresponding ontology concept. To do this, we identify all named entities in this set of questions. For example, the question “what is the unit of voltage?” is annotated “what + [Be/Have] + [Relation] + [Concept/Attribute]”. Once this process was finished, we delete all duplicated questions. 3. Eventually, the set of questions is classified into clusters or categories according to their semantic equivalences. Two questions are semantically identical when they illustrate same information and talk about similar idea or provide the same ontological concepts. For example, the questions: “what is the unit of voltage?” or “what is the unit of potential difference?” belong to the same semantic category. Each cluster demonstrates how to deal with its elements, what inference process is required and what kind of answer can be expected. Once the questions are categorized, we associate a SQL statement to each cluster. This statement will enable to obtain the answer for any of the questions in the cluster. The SQL statement includes the generic concepts (e.g., Electric\_element). Therefore, before applying the SQL commend to database, these general concepts must be replaced by the original data instances determined in the user’s query. It is worth noting, different questions may be represented by the same cluster, since, in natural language there can be various ways of asking for a specific ontology concept. For example, the questions: “what is the unit of voltage?” or “what is the unit of potential difference?” requested the unit of voltage using two ways (“ voltage”, “potential difference”). Thus, we identified and stored the various ways of mentioning each ontology concept in a question. This knowledge can be useful for the ISM approach in order to determine paraphrases regarding the concepts that appear in the users’ question.

**[2] Characterizing Searches for Mathematical Concepts**

…

**[3] Platypus – A Multilingual Question Answering Platform for Wikidata**

<https://hal.archives-ouvertes.fr/hal-01730479/document>

In this paper we present Platypus, a natural language question answering system. Our objective is to provide the research community with a production-ready multilingual question answering platform that targets Wikidata, the largest general-purpose knowledge base on the Semantic Web. Our platform can answer complex queries in several languages, using hybrid grammatical and template based techniques.

Evaluated on SimpleQuestions dataset

<https://github.com/davidgolub/SimpleQA/tree/master/datasets/SimpleQuestions>

**SimpleQuestions Dataset**

[4] [Large-scale Simple Question Answering with Memory Networks](https://paperswithcode.com/paper/large-scale-simple-question-answering-with)

<https://arxiv.org/pdf/1506.02075v1.pdf>

We collected SimpleQuestions in two phases. The first phase consisted of shortlisting the set of facts from Freebase to be annotated with questions. We used FB2M as background KB and removed all facts with undefined relationship type, i.e., containing the word freebase. We also removed all facts for which the (subject, relation-ship) pair had more than a threshold number of ob-jects. This filtering step is crucial to remove facts. which would result in trivial uninformative ques-tions, such as,Name a person who is an actor?.The threshold was set to 10.In the second phase, these selected facts weresampled and delivered to human annotators togenerate questions from them. For the sampling,each fact was associated with a probability whichdefined as a function of its relationship frequencyin the KB: to favor variability, facts with relation-ship appearing more frequently were given lowerprobabilities. For each sampled facts, annotatorswere shown the facts along with hyperlinks tofreebase.comto provide some context whileframing the question. Given this information, an-notators were asked to phrase a question involvingthe subject and the relationship of the fact, withthe answer being the object. The annotators wereexplicitly instructed to phrase the question differ-ently as much as possible, if they encounter multi-ple facts with similar relationship. They were alsogiven the option of skipping facts if they wish todo so. This was very important to avoid the anno-tators to write a boiler plate questions when theyhad no background knowledge about some facts.

[5] Simple Question Answering by Attentive Convolutional Neural Network

<https://arxiv.org/pdf/1606.03391.pdf>

Factoid QA datasets from search-engine logs

[6] <https://stackoverflow.com/questions/30585760/where-can-i-find-a-corpus-of-search-engine-queries>

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

[8] <https://www.kaggle.com/c/yandex-personalized-web-search-challenge>

[9] An empirical analysis of question queries in a largescale Persian search engine log

<https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1951&context=dubaipapers>

[10] <https://trends.google.com/trends/?geo=US>