Links with Answers: Query Answering for Customer Support⋆

Veronika Thost[1,2], Achille Fokoue[2], Vittorio Castelli[2], and Salim Roukos[2]

[1] MIT-IBM Watson AI Lab
[2] IBM Research
veronika.thost@ibm.com, {achille,vittorio,roukos}@us.ibm.com

1 Introduction

The increasing complexity of today's systems has triggered a tremendous demand for customer support [2]. Customers often ask about standard procedures or common issues that are well documented *somewhere* so that their questions could be answered efficiently in an automated fashion. But current solutions often fail in addressing their problems – we assume that every reader has enough experience with unhelpful chat bots, automated phone systems, and standardized e-mail replies in order to see that there is room for improvement.

One main challenge for automated solutions is to understand (enough of) the customer question to either link it to the answer, if it is known by the system, or forward it to special, e.g., human support. Natural language understanding is not easy in general and an entire area of research by itself. However, in the context of a concrete system, the most relevant vocabulary can be narrowed down to the one of the documents containing the answers and its specific semantics are known. For example, the user types and relevant tasks and tools are usually fixed and thus could be described semantically and applied for answering the queries.

In this paper, we focus on a set of natural language questions asking for IT support from within IBM. Each question is associated with an URL, the link to a document containing the answer. These documents are available in HTML but, alas, not annotated in the form of Semantic Web (SW) pages; we assume that this is a scenario which is common, also outside of IBM. In this initial investigation, we show that ontologies and reasoning using SW technologies can help in resolving problems with underspecification, ambiguity, and variety in the questions, and in finding answers which are missed by standard information retrieval (IR) approaches. We compare a standard TF-IDF-based approach to a semantic solution based on a custom ontology, the state-of-the-art rule learning system AMIE [3], and SWRL reasoning, using the learned rules. Although the benefits of including semantic knowledge become evident, we point out several issues with and gaps in the current tool set; challenges for future development of semantic technologies. In this sense, we propose a semantic baseline to motivate research in an area of rising importance where, in our opinion, SW technologies could provide fundamental benefits.

2   The ITS Dataset

We focus on the ITS dataset (IT Support dataset), a set of natural language questions asking for IT support from within IBM, where each question is associated with an URL, the link to a document containing the answer. Alas, we were not able to disclose the data so far, but we are planning to do so.

Altogether, there are 966 queries, all in English, with a min/max/avg length of 1/150/24 words. Figure 1 gives a more detailed overview on the lengths. The following two examples show that some questions are very short, mentioning only the basic problem, while others are rather complex, containing more detailed problem descriptions:

> *Password is not working please send recovery key to change the password*
>
> *My laptop has been having some issues. 1) Slows down and freezes. 2) sometimes Does not shut down because [...] 3) it has screen refresh issues and when I am working in excel some numbers can not be read*

There are 276 different links (URLs). Figure 1 shows an aggregated view of the number of example questions associated with the same URL; for example, there are 20 different URLs for which we have 3 query examples. Note that, for 175 links, we have only a single question as example. Initially, the answers to the queries were determined by human agents who answered the questions by consulting the corresponding documents. However, while our data contains only the original user questions, it is likely that there was a discussion process where the agent asked for additional information in order to capture the problem of the user. For that reason, some of the original query and answer pairings cannot be rated to be correct in a more general context. For instance, one query is associated to a user guide in Chinese although it does not mention anything related to Chinese or China at all; hence, the English version of the user guide
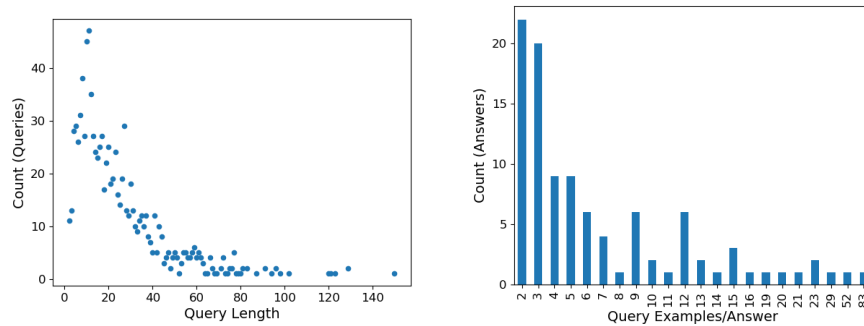

Fig. 1: Overview of query length and of number of query examples/answer; for readability, the latter does not show the 175 answers with only one query example.

should rather be considered as the answer. The question and answer pairs have been validated, 20% of the data by more than one annotator, to estimate the confidence of that validation. Since the inter-annotator agreement was only 84%, we afterwards checked the data by ourselves, changed some answers, and removed pairs we found ambiguous. There are several observations we can make about the dataset. In general, it is too small to apply learning as a standalone solution. In particular, there are several links for which we have only few query examples. Also, many queries are underspecified descriptions of the actual problem. For instance, the first example query above mentions a password but not its kind – users generally have dozens of passwords at IBM; on the other hand, the mentioning of a recovery key might be a hint to the password of a PC.

3 A Baseline for Ontology-Based Query Answering

We propose a semantic baseline based on a (very small) custom ontology, the rule learning system AMIE, and SWRL reasoning, using the learned rules.
Representing the Queries as Facts. Each query is preprocessed by repairing common typos, filtering out stopwords, and stemming the words. Then, a set of facts is created for each by considering the remaining words as constants and using a single predicate HasKey. For the above example query, we have the following facts (amongst others), including one for the answer:

$$\text{HasKey(q0,laptop), HasKey(q0,issue), HasKey(q0,slow), HasKey(q0,down)}$$

The Ontology. Our ontology is a small, manually built set of rdf:type statements relevant for the ITS queries. The main goal is generalization during training. For instance, we have statements to integrate different device types:

$$\text{Type(ipad,device), Type(laptop,device), Type(computer,device)}$$

We have not found any publicly available ontology for the IT support domain.
Rule Learning. As outlined above, we use AMIE for learning rules such as:

$$\text{HasKey(X,Y), Type(Y,device),HasKey(X,slow)} \rightarrow \text{HasAns(X,www.hw.com)}$$

SWRL Reasoning. Given a set of queries as facts (w/o answer facts), we then can use SWRL reasoning to find possible answers.

4 Experiments

We compared our semantic baseline (SB) to a standard IR approach (IRB) using TF-IDF and Linear Support Vector classification, with a train/test split of 4/1, and the test set containing only queries for whose answer the training set contains at least 2 examples. We compared the top 1/5/10 answers per query of IRB to those of SB, for which we considered a random subset of the same size or less, if it did not yield that many answers. Note that SB may return

several (unrated) answers since several rules may apply. Best results are: precision@1/5/10 is 0.50/0.15/0.08 (IRB) and 0.49/0.35/0.33 (SB); recall@1/5/10 is 0.50/0.73/0.79 (IRB) and 0.39/0.55/0.55 (SB). However, the approaches are quite different in nature, so that a comparison based just on these numbers is misleading. Random checks show, e.g., that the answers of SB are closer to the correct ones than the additional ones (i.e., the ones not top-rated) of IRB. The poster gives more details about all experiments we ran and also describes performance effects of the AMIE parameter selection, such as maximal rule length. Observe also that, to learn any rules at all for a link, both approaches need some examples, so they cannot be considered as stand-alone solutions. Nevertheless, we believe that SW methods are useful to augment IR approaches.

5   Conclusions

Our initial investigation of applying state-of-the-art SW technology for IT support question answering revealed several open challenges, amongst others:

- There is little work on ontologies for customer support [1, 4].
- Semantic parsing of natural language – into a better representation than our simple keyword facts – is an active area of research but has not been considered much in the SW context (e.g., its integration with ontologies etc.)
- State-of-the-art rule learning systems only consider Datalog, though some with negation, and struggle with larger rule bodies. More complex rules could model the queries more faithfully.
- The links, or rather the linked documentation, could similarly be described in an ontology. Such a hierarchy would also allow for suggesting more general answers in case there is doubt about the exact answer.
- There is basically no public example data.
- The documents could be used as additional information for learning, and especially help to link to answers with no or only few example queries.

Overall, the performance of our approach is mostly limited by the simple query representation and by the rule learning system. On the other hand, it turned out that in most cases an ontology which is a simple hierarchy is sufficient to resolve ambiguity and meaning. There are larger efforts at IBM to build an IT support ontology and a question answering dataset. In this paper, we share our insights in first work with the data since customer support is a use case of rising importance, to motivate discussion and research in the SW community.

References

1. eClassOWL. http://www.heppnetz.de/projects/eclassowl/
2. https://www.inc.com/peter-roesler/american-express-study-shows-rising-consumer-expectations-for-good-customer-service.html, accessed: 2019-04-04
3. Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F.M.: Fast rule mining in ontological knowledge bases with AMIE+. VLDB J. **24**(6), 707–730 (2015)
4. Quan, T.T., Nguyen, T.D.: Ontology evolution for customer services. In: Proceedings of the Knowledge Representation Ontology Workshop. pp. 61–69 (2008)