Hunting for Inconsistencies in Multilingual DBpedia with QAKiS

Elena Cabrio, Julien Cojan, Serena Villata, and Fabien Gandon

INRIA Sophia Antipolis, France
{firstname.lastname}@inria.fr

Abstract. QAKiS, a system for open domain Question Answering over linked data, allows to query DBpedia multilingual chapters with natural language questions. But since such chapters can contain different information w.r.t. the English version (e.g., more specificity on certain topics, or fill information gaps), *i)* different results can be obtained for the same query, and *ii)* the combination of these query results may lead to inconsistent information about the same topic. To reconcile information obtained by distributed SPARQL endpoints, an argumentation-based module is integrated into QAKiS to reason over inconsistent information sets, and to provide a unique and motivated answer to the user.

1 Introduction

In the Web of Data, the combination of the information items concerning a single real-world object coming from different data sources, e.g., the results of a single SPARQL query on different endpoints, may lead to an inconsistent results set, mining the overall quality of the data itself. In particular, this problem arises while querying DBpedia multilingual chapters, since different information can be provided for the same query (e.g. answers can be either identical, contradictory, or one can subsume the other). To reconcile information provided by multilingual DBpedia chapters to obtain a consistent results set, we embed an argumentation module in QAKiS, that *i)* detects the semantic relations linking each piece of information to the others returned by the different SPARQL endpoints, and *ii)* adopts *abstract bipolar argumentation theory* to reason over the inconsistencies among the answers, and to return a consistent (sub)set of them to the user.

An abstract bipolar argumentation framework (BAF) [2] represents a negative relation between elements called *arguments* through a binary *attack* relation, and a positive relation among arguments through a binary *support* relation. Argumentation semantics then allow to reason about the arguments and their relations to detect the set of *accepted* arguments, i.e., those considered as believable by an external evaluator with full knowledge of the BAF. However, such kind of crisp evaluation of the arguments is not suitable for the real life scenarios where a numerical value is required. This is why we adopt and extend the fuzzy labeling algorithm proposed in [3] to consider also the support relation in addition to the attack one.

The overall argumentation framework together with the acceptability degree of each argument is used to motivate to the user the answer the system returns.

2 Extending QAKiS to reason over inconsistent answers

QAKiS (Question Answering wiKiFramework-based System)[1] [1] addresses the task of QA over structured knowledge-bases (e.g., DBpedia), where the relevant information is expressed also in unstructured forms (e.g., Wikipedia pages). It implements a relation-based match for question interpretation, to convert the user question into a query language (e.g., SPARQL). More specifically, it makes use of relational patterns (automatically extracted from Wikipedia), that capture different ways to express a certain relation in a given language. QAKiS is composed of four main modules (Fig. 1): *i)* the **query generator** takes the user question as input, generates the typed questions, and the SPARQL queries from the retrieved patterns; *ii)* a **Named Entity (NE) Recognizer**; *iii)* the **pattern matcher** takes as input a typed question, and retrieves the patterns matching it with the highest similarity; and *iv)* the **SPARQL package** handles the queries to DBpedia. QAKiS targets questions containing a NE related to the answer through one ontological property, i.e., questions match a single pattern.

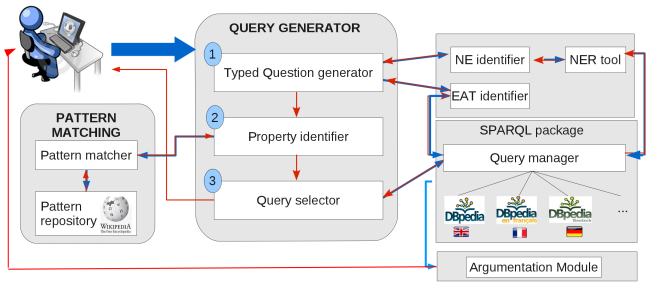

Fig. 1: QAKiS workflow

Given the answers retrieved by DBpedia multilingual endpoints for a SPARQL query, the argumentation module assigns a support or attack relation between the arguments (see Fig 2): *i)* **identity [assigned relation: support]**: if two endpoints provide identical answers (Fig. 2a-b where both French and English DBpedia SPARQL endpoints provide *Italy* as answer to *Where is the Colosseum located?*; *sameAs* links are used to recognize the translation of the same word in multilingual DBpedia). Arguments are merged into a unique one becoming highly acceptable as shared by several sources[2]; *i)* **subsumption [assigned relation: support]**, when one of the answers is more specific than the other, both in terms of spacial relation (Fig. 2d) and hyperonymy (Fig. 2c where *Gibson* is_a *Guitar*)[3]; *iii)* **conflict [assigned relation: attack]**, if the answers are different,

[1] `http://qakis.org/qakisArgumentation`

[2] The starting confidence score of this argument is calculated as the *arctangent* of the confidence scores of the endpoints providing such answer (max value = 1).

[3] External sources of semantic knowledge are exploited, e.g., GeoNames, YAGO.

and there is no subsumption (Fig. 2e-f where the locations of the Colosseum by Italian and English DBpedia are contradictory). When each endpoint provides a *list* of values as answer (e.g., DBpedia non-functional properties, Fig. 2g), QAKiS does not consider arguments of the same list as conflictual.

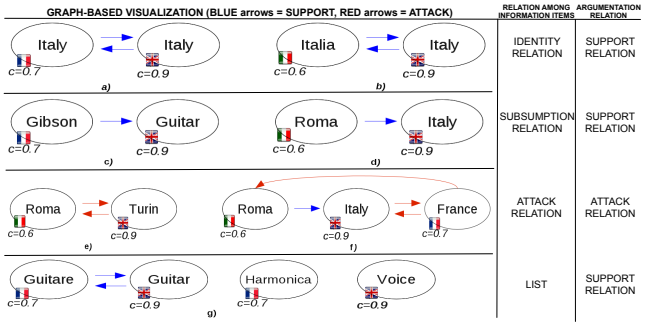

Fig. 2: Semantic relations and their mapping in argumentation

We assign an apriori confidence score to the endpoints according to their dimensions and solidity in terms of maintenance (other methods are under investigation). Starting from the obtained set of arguments and relations, the module calculates the arguments' acceptability degree (i.e., the arguments that will be proposed to the user as more reliable). We propose a bipolar fuzzy labeling algorithm where $\mathcal{A}$ is a fuzzy set of trustful arguments, and $\mathcal{A}(A) = \max_{s \in \text{src}(A)} \tau_s$ is the membership degree of argument A in $\mathcal{A}$ given by the trust degree of the most reliable source offering argument A, where τ_s is the degree to which source $s \in \text{src}(A)$ is evaluated as reliable. A bipolar fuzzy labeling is a total function $\alpha : \mathcal{A} \to [0, 1]$. We say that α is a bipolar fuzzy labeling *iff*, for all arguments A, $\alpha(A) = \texttt{avg}\{\min\{\mathcal{A}(A), 1 - \max_{B:B \overrightarrow{attack}\, A} \alpha(B)\}; \max_{C:C \overrightarrow{support}\, A} \alpha(C)\}$. $\alpha(A) = 0$ means that A is outright unacceptable, $\alpha(A) = 1$ means A is fully acceptable. All cases in-between provide the degree of the acceptability of the arguments which are considered accepted at the end, if they overcome a certain threshold. The result of the fuzzy labeling is the arguments confidence score.

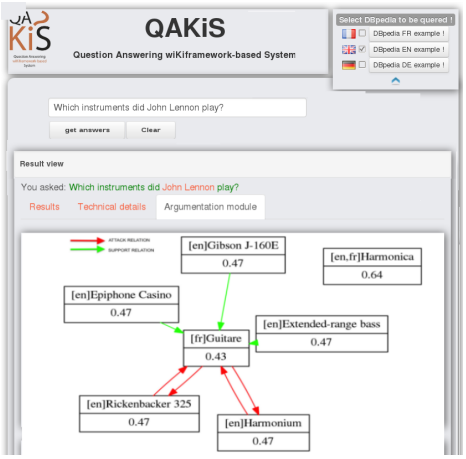

Fig. 3: QAKiS demo interface.

Figure 3 shows the QAKiS demo interface. The user can select the DBpedia chapter he wants to query besides English, i.e. French or German DBpedia (top right corner) [1]. Then the user can either write a question or select among a list of examples. QAKiS outputs *i)* the user question, *ii)* the generated typed question, *iii)* the pattern matched, *iv)* the generated SPARQL query, *v)* the answer, and *vi)* the graph of the answers by the different endpoints and their relations, together with their confidence score.

Since QAKiS currently targets only questions containing a NE related to the answer through one ontological property, we extracted from QALD-2[4] data the questions corresponding to such criterion, i.e. 58 questions, and we run them over QAKiS (querying English, German and French DBpedia endpoints). Since QALD-2 questions are created for English DBpedia, only in 25/58 cases there are at least two endpoints that provide an answer. We carried out two sets of experiments. In Experiment 1 (input: the answers obtained from the different DBpedia endpoints, manually creating the SPARQL query), performances of the argumentation module in identifying the arguments from the endpoints are F-meas. 0.97, in relation assignment are F-meas. 0.72. Errors in arguments identification are due to missing *SameAs* links in DBpedia: the algorithm does not merge translations of the same answer, and it considers them as different. Wrong relation assignments are mainly due to missing attacks among arguments.

Since QAKiS performances are about ∼50%, the results of Experiment 2 (submitting natural language questions to QAKiS) are obtained accordingly, F-meas 0.72 for argument identification and F-meas 0.55 for relation assignment (the argumentation module is biased by QAKiS mistakes). The average computation cost of the argumentation module is ∼5s for 1-answer, and ∼125s for n-answers questions. The complexity is quadratic, at least one SPARQL query is sent for each couple of answers. We are working on the algorithm optimization.

3 Future perspectives

Extensions are planned in several directions: *i)* to let the user assign the confidence degree to the information sources embedding this feature in the QAKiS interface; *ii)* extend the set of ontologies we consider to detect further relations (positive and negative) among the information items; *iii)* perform a user evaluation campaign to verify which kind of visualization is better usable.

References

1. Cabrio, E., Cojan, J., Gandon, F., , Hallili, A.: Querying multilingual DBpedia with QAKiS. In: Procs of ESWC 2013. Demo papers. (2013)
2. Cayrol, C., Lagasquie-Schiex, M.C.: Bipolarity in argumentation graphs: Towards a better understanding. In: Procs of SUM 2011, pp. 137–148. LNCS, v. 6929 (2011)
3. da Costa Pereira, C., Tettamanzi, A., Villata, S.: Changing one's mind: Erase or rewind? In: Procs of IJCAI2011. pp. 164–171. IJCAI/AAAI (2011)

[4] Question Answering Linked Data challenge: `http://bit.ly/QALD2`