Collaborative Project

FROCKG - Fact Checking for Large Enterprise Knowledge Graphs

Project Number: E! 113314 Start Date of Project: 2020/01/01 Duration: 36 months

Deliverable 3.1: First version of the explanations algorithms including initial benchmarking results

Dissemination Level	Public
Due Date of Deliverable	Month 18, 2021/06/30
Actual Submission Date	Month 18, 2021/06/30
Work Package	WP3
Task	T3.1
Type	Report
Approval Status	Work in progress
Version	1.0
Number of Pages	14

Abstract:
This report presents the first prototype of the explanation generation component of the FROCKG platform. The report explains the functionality of the component. It also describes the automatic evaluation that has been setup and the results that have been achieved by this first prototype. A detailed analysis of the results concludes the report.

The information in this document reflects only the author's views and Eurostars is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.

Project by Eurostars.

History

Version	Date	Reason	Revised by
0.1	2021/05/06	final TOC	Michael Röder
0.2	2021/05/20	Drafted content	Michael Röder, Farshad Afshari
0.3	2021/06/09	Evaluation results added	Farshad Afshari
0.4	2021/06/14	First draft finished	Michael Röder, Farshad Afshari
1.0	2021/06/29	Final version	Michael Röder, Farshad Afshari

Author List

Organization	Name	Contact Information
Paderborn University	Michael Röder	michael.roeder@upb.de
Paderborn University	Farshad Afshari	farshad.afshari@uni-paderborn.de

FROCKG

Contents

Introduction

The goal of Fact Checking is to decide whether a given fact is true or false. This is typically done based on some reference knowledge. Most approaches try to identify evidence within this reference knowledge to support or refute the given fact. Work package 2 of the FROCKG project develops different Fact Checking services. One approach relies on textual reference knowledge, i.e., the given fact is searched within the reference corpus. Text snippets that give evidence for the fact are extracted and rated based on a machine learning model. After that, the service calculates a final truth value based on the extracted pieces of evidence. This service is complemented by a second, knowledge-graph-based approach. This approach relies on connections between the subject and object of the given fact within a reference knowledge graph. Figure 1 shows an example for these connections (dubbed paths) for the example fact "`Barack Obama, nationality, the United States of America`".

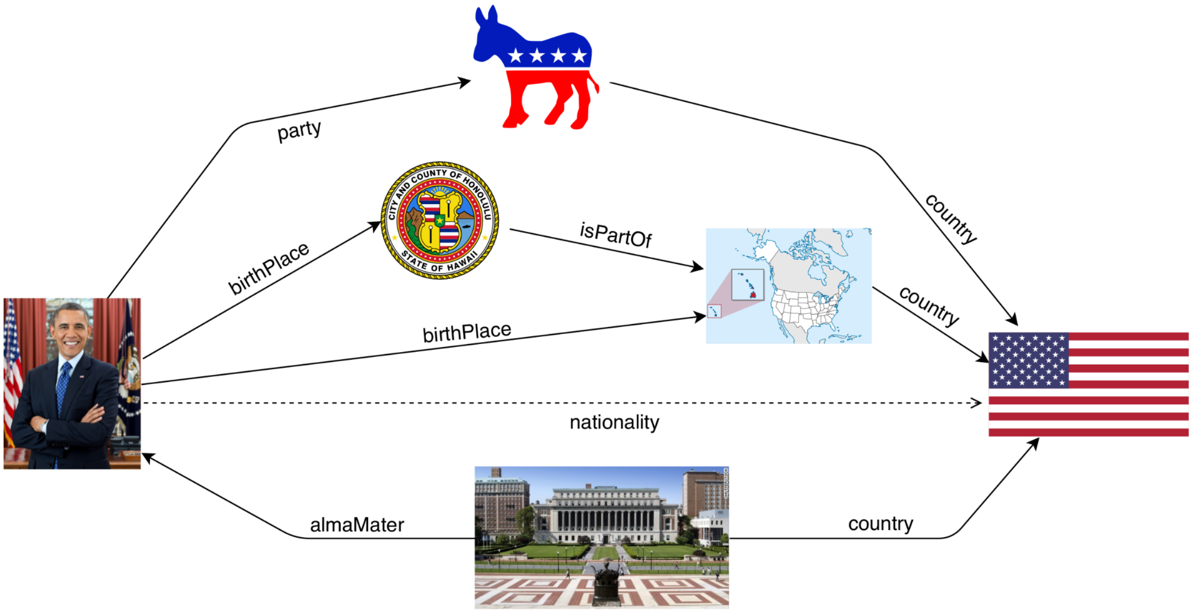

Figure 1: The fact that should be checked (the dotted line) and additional paths between the resources Barack Obama and the United States of America extracted from the DBpedia.

The Fact Cecking service scores these paths with respect to their statistical co-occurrence with facts that share the same property as the given fact. With respect to the example, the path comprising the two properties "`birthplace`" and "`country`" will be rated based on the number of times this path occurs in the reference knowledge graph and how often it connects a person and a country which are also connected with the "`nationality`" property.

Both Fact Checking services return a final truth score for the given fact and the single pieces of evidence they found (i.e., text snippets and paths, respectively). However, a user may not have the background knowledge to interpret the scores, the paths or the context of the extracted text snippets. The main goal of this work package is the generation of natural language explanations for the results of the Fact Checking services. These explanations should enable a user to understand the results of the FROCKG platform. To this end, we develop an explanation service within this work package. As defined in D1.2 [1], the input of this service are the pieces of evidence that the Fact Checking algorithms identified. Each piece of evidence comes with a score that expresses to which extent it supports the given fact. In addition, the final truth value that the services calculated for the given fact is provided as well. Depending on the type of evidence, different strategies have been implemented to incorporate them into an explanation.

This deliverable describes the first prototype that was implemented to achieve the aforementioned goal. The University Paderborn (UPB) leads the development. It should be mentioned that because of the pandemic situation the position of a researcher working for the FROCKG project at the UPB couldn't be staffed before mid of August 2020 while the project already started in January 2020. To ensure that this won't delay the project, the work package lead decided to focus on the main functionalities of the explanation component. This ensures that the prototype fulfills the requirements defined in D1.2 [1] and D1.3 [2], and is ready to be used by the other work package.

Evidence from Text

Text-based Fact Checking services rely on textual reference knowledge, i.e., a text corpus. An example of such an approach is FactCheck [3]. FactCheck uses a local corpus by utilizing Elasticsearch to index the documents of the corpus. When it is used to check a given fact, it transforms the fact into keyword queries. These queries are used to search for relevant documents in the reference corpus. The retrieved documents are further analysed and parts of the text in which the searched keywords occur are extracted. Each extracted piece of text is rated by a machine learning algorithm with respect to its quality and support for the given fact. In its final step, FactCheck relies on a trained classifier that uses all identified pieces of evidence to decide whether the given fact is true or false. The result is a truth value for the given fact as well as the extracted pieces of evidence, an identifier of the documents from where they have been extracted and their scores.

Tables 1 and 2 show some examples of extracted text snippets for the two given facts "`Albert Einstein, birthplace, Germany`" and "`Elizabeth FitzHugh, spouse, William Parr 1st Baron Parr of Kendal`".

ID	Feature	Value
1	Document	https://en.wikipedia.org/wiki/Culture_of_Switzerland
	Text snippet	Physicist Albert Einstein , born in Germany , moved to Switzerland in 1895 at the age of 16 and became a Swiss citizen in 1901 .
	Truth value	0.9701708046574612

2	Document	https://en.wikipedia.org/wiki/Bernhard_Caesar_Einstein
	Text snippet	Hans Albert was born on 10 July 1930 in Dortmund , Germany , where Hans Albert was involved in a bridge building project .
	Truth value	0.9998544980570143
3	Document	https://en.wikipedia.org/wiki/Hans_Albert_Einstein
	Text snippet	Einstein 's father , Albert , left Germany in 1933 to escape the virulently antisemitic Nazi threat
	Truth value	0.9087569184551185

Table 1: Example texts extracted for the example fact ”`Albert Einstein, birthplace, Germany`”.

ID	Feature	Value
1	Document	https://en.wikipedia.org/wiki/Alice_Montacute,_5th_Countess_of_Salisbury
	Text snippet	Their daughter , Elizabeth , married William Parr , 1st Baron Parr of Kendal
	Truth value	0.8407566318935811
2	Document	https://en.wikipedia.org/wiki/Alice_Neville
	Text snippet	Elizabeth FitzHugh , who married firstly William Parr , 1st Baron Parr of Kendal and secondly Sir Nicholas Vaux
	Truth value	0.8164785150220529
3	Document	https://en.wikipedia.org/wiki/Baron_FitzHugh

	Text snippet	They had five sons and six daughters , including Elizabeth FitzHugh , who married William Parr , 1st Baron Parr of Kendal and then Nicholas Vaux , 1st Baron Vaux of Harrowden
	Truth value	0.6559214724088098

Table 2: Example texts extracted for the example fact ”`Elizabeth FitzHugh, spouse, William Parr 1st Baron Parr of Kendal`”.

Evidence from KGs

A knowledge-graph-based Fact Checking approach makes use of structured information of a reference knowledge graph to identify pieces of evidence that support or refute the given fact. These pieces of evidence are typically paths that connect the given subject and object. COPAAL [4] is an open-source knowledge-graph-based Fact Checking algorithm. For a given fact, COPAAL searches for paths between the subject and object of the fact within the reference knowledge graph. These paths are scored whether they corroborate the existence of the fact using an NPMI-based heuristic.
The paths that are identified by COPAAL can be transformed into a natural language representation. We use the rule-based LD2NL approach for that.[1] Table 3 shows examples of facts and the verbalization of a single path that has been found by COPAAL for the given fact.

Fact	Path	Generated text
subject : Achilleas Gerokostopoulo predicate : death place object : Greece	[Achilleas_Gerokostopoulos, deathPlace, Patras], [Patras, country, Greece]	Achilleas Gerokostopoulos' death place is Patras. Patras' country is Greece.
subject : Polemon I of Pontus predicate : child object : Artaxias III	[Pythodorida_of_Pontus, spouse, Polemon_I_of_Pontus], [Pythodorida_of_Pontus, child, Artaxias_III]	Pythodorida of Pontus' spouse is Polemon I of Pontus. Pythodorida of Pontus' child is Artaxias III.
subject : Battle of Williamsport	[Battle_of_Williamsport, place,	Battle of Williamsport relates an entity to the populated

[1] https://github.com/dice-group/ld2nl

predicate : place object : Maryland	Washington_County,_Maryland], [Washington_County,_Maryland, state, Maryland]	place in which it is located. Washington County, Maryland. Washington County, Maryland's state is Maryland.
subject : Alexander of Battenberg predicate :spouse object : Johanna Loisinger	[Alexander_of_Battenberg, spouse, Johanna_Loisinger]	Alexander of Battenberg's spouse is Johanna Loisinger.

Table 3: Example facts, a single path that has been found by COPAAL, and the verbalization of these paths.

Merging Pieces of Evidence

The FROCKG platform offers the addition of several Fact Checking algorithms. The results of these algorithms will be merged into a single result. For explaining the result of the Fact Checking process to a user, the platform has to be able to merge the pieces of evidence that originate from these different sources. To this end, we implemented a component that merges the pieces of evidence into a single explanation.[2]

We have chosen a generic approach that allows the addition of further services, if necessary. The steps of our approach are as follows:

1. Sort the pieces of evidence per service based on their scores.
2. Select the best *n* elements per service.
3. For each service with a non-empty list of evidence, generate an introductory sentence and add the verbalized pieces of evidence.

In the first step, the pieces of evidence of a Fact Checking service are sorted with respect to their importance. This step guarantees that the explanation will rely on the most important pieces of evidence.

The second step selects *n* elements per service. This is necessary since a Fact Checking service may return a large amount of pieces of evidence and the users of the FROCKG platform won't have the time to read through all of them. The number of elements that are chosen is configurable.

The third step ensures that a coherent text is generated. This increases the quality of the explanation and gives a context for the single pieces of evidence, which makes it easier for a user to understand them. Table 4 lists explanations generated for some example facts that have been checked using FactCheck [3] and COPAAL [4]. It should be noted that the pieces of text in quotes are extracted from documents and are not further adapted by our approach.

Fact	Explanation

[2] The open-source implementation of the component can be found at https://github.com/dice-group/FROCKG

subject : Tay Zonday predicate : birthPlace object : Minneapolis	We found the following evidence in our reference knowledge base: Tay Zonday's birth place is Minneapolis. Tay Zonday's birth place is Minnesota. Minnesota's largest city is Minneapolis. We found the following evidence in our reference corpus: “Zonday was born Adam Nyerere Bahner in Minneapolis , Minnesota .” “While a PhD student in Minneapolis , Minnesota , Zonday began performing at open mic nights in 2006 .”
subject : Subramaniam Siva predicate : writer object : Pori (film)	We found the following evidence in our reference knowledge base: Pori (film)'s director is Subramaniam Siva. We found the following evidence in our reference corpus: “Pori (English : Sparks) is a 2007 Indian Tamil-language action comedy film directed by Subramaniam Siva , who rendered his debut Thiruda Thirudi , a 2004 runaway blockbuster .”
subject : Michael Rummenigge predicate : affiliation object : Borussia Dortmund	We did not find any evidence in our reference knowledge base. We found the following evidence in our reference corpus: “His brother Michael Rummenigge was also a noteworthy footballer . His played as forward for Bayern Munich and Borussia Dortmund from 1982 -- 88 and 1988 -- 94 , respectively .”
subject : Artem Laguta predicate : nationality object : Russia	We found the following evidence in our reference knowledge base: Artem Laguta's nationality is Russia. We found the following evidence in our reference corpus: “Laguta have two speedway licences : Russian (MFR) and Latvian (LaMSF) .” “Laguta is current Individual Latvian Champion. He is a three-time Russian champion.” “Grigory Laguta (born April 9 , 1984) is a Russian-born Latvian motorcycle speedway rider .”

Table 4: Example facts and the explanations generated for them based on the top 3 results of COPAAL and FactCheck, respectively.

Evaluation

As described in D1.3, we created a benchmarking dataset for the generation of explanations. This allows us to evaluate the fluency of the explanations based on common metrics. For our evaluation, we use the evaluation platform BENG [5]. It measures the following metrics:

- BLEU [6]: The Bilingual Evaluation Understudy Score is a metric for evaluating a generated sentence to a reference sentence. A perfect match results in a score of 1.0, whereas a perfect mismatch results in a score of 0.0.
- BLEU NLTK [7]: The original BLEU was designed for the document-level, BLEU NLTK uses a smoothing technique that is implemented in the NIST official BLEU toolkit. It assigns a geometric sequence starting from 1/2 to the n-grams with 0 matches.
- METEOR [8]: The Metric for Evaluation of Translation with Explicit ORdering is used for the evaluation of machine translation output. The metric is based on the harmonic mean of unigram precision and recall, with recall weighted higher than precision.
- chrF++ [9]: This measure is based on character n-gram precision and recall enhanced with word n-grams. The tool calculates the F-score averaged on all character and word n-grams. The default character n-gram order is 6 while the word n-gram order is 2.
- TER [10]: The translation error rate measures the number of edits needed to change the generated text into the given explanation.

Dataset

The evaluation dataset relies on the YAGO 3.0 knowledge base. It comprises 5628167 triples. We randomly selected 1000 triples and executed the fact checking algorithms for them.[3] After that, we selected facts that were correctly identified as true by our fact checking system. For 41 of these facts, we manually created explanations. To this end, one Fact Checking expert created the explanation. A second expert checked these manually created explanations and marked those that he identified as not good. These marked explanations were then discussed and fixed by the two experts.[4]

Results

Table 5 shows the results of the experiment.[5] It can be seen that the values are in a moderate area. This means that the explanations are close to the human created explanations but can be further improved.

Measures	Value

[3] The results of the checks can be found in D2.1
[4] The created dataset can be found at https://hobbitdata.informatik.uni-leipzig.de/projects/FROCKG
[5] The results are available online in the BENG platform at https://beng.dice-research.org/gerbil/experiment?id=202106100004

BLEU	43.23
BLEU NLTK	0.44
METEOR	0.42
chrF++	0.70
TER	1.00

Table 5: Values measured for the different metrics.

Table 6 shows several examples. It should be noted that the quoted pieces of text stem from the reference corpus and have not been generated by the explanation component. Hence, their quality is not taken into account during the following discussion.

The first example shows that the generated explanation is close to the human created sentences in some of the cases. However, the other three examples show that humans tend to write shorter explanations while the machine-generated explanation comes with several verbose expressions that could be omitted.

The second example shows two points that can be improved in the future. First, the fact the user asked for was directly found in the knowledge base. So there is no further explanation needed than stating that. However, the generated explanation would list more evidence. Second, no evidence was found by the text-based Fact Checking component. Stating that, might not be necessary and is completely ignored by the expert's explanation.

The third example shows that facts from the knowledge base are given as evidence although they are very similar to each other. This repetition of pieces of evidence that are very similar to each other. increases the length of the explanation but does not add new information.

The fourth example shows a similar situation to the second example. For this fact, the knowledge-graph-based Fact Checking approach could not find any evidence. However, this is a detail that the user might not be directly interested in and our expert's explanation does not include it.

Machine Explanation	Human Explanation
We found the following evidence in our reference knowledge base: Artem Laguta's nationality is Russia. We found the following evidence in our reference corpus: “Laguta have two speedway licences : Russian (MFR) and Latvian (LaMSF) .” “Laguta is current Individual Latvian Champion. He is a three-time Russian champion.”	In my knowledge base, I found the following fact that backs up this claim: Artem Laguta's nationality is Russia. In addition, I found the following pieces of text that could serve as evidence: "Laguta have two speedway licences : Russian (MFR) and Latvian (LaMSF) ." "Laguta is current Individual Latvian Champion. He is a three-time Russian champion."

"Grigory Laguta (born April 9 , 1984) is a Russian-born Latvian motorcycle speedway rider ."	"Grigory Laguta (born April 9 , 1984) is a Russian-born Latvian motorcycle speedway rider ."
We found the following evidence in our reference knowledge base: Isao Yamagata's birth place is London. We did not find any evidence in our reference corpus	In my knowledge base, I found the following fact that backs up this claim: Isao Yamagata's birth place is London.
We found the following evidence in our reference knowledge base: Garba Lawal's team is C.D. Santa Clara. Garba Lawal's club is C.D. Santa Clara. We did not find any evidence in our reference corpus	In my knowledge base, I found the following fact that backs up this claim: Garba Lawal's team is C.D. Santa Clara.
We did not find any evidence in our reference knowledge base. We found the following evidence in our reference corpus: "His brother Michael Rummenigge was also a noteworthy footballer . His played as forward for Bayern Munich and Borussia Dortmund from 1982 -- 88 and 1988 -- 94 , respectively ."	I found the following pieces of text that could serve as evidence: "His brother Michael Rummenigge was also a noteworthy footballer . His played as forward for Bayern Munich and Borussia Dortmund from 1982 -- 88 and 1988 -- 94 , respectively ."

Table 6: Example of machine explanations and human explanations.

Summary

With the work described in this report, WP3 fulfilled all necessary steps to achieve the project's second milestone. The evaluation results presented above are promising. We showed that the explanation generation component is able to generate explanations that are already close to a human explanation. However, we also identified some parts that can be further improved. Future work within WP3 will concentrate on these improvements.

We also provided a dataset for the explanation of fact checking results. To the best of our knowledge, this is the first dataset of its kind. It will be used to constantly measure the improvement of the generated explanations during the rest of WP3.

References

[1] Deliverable 1.2: Component architecture and interfaces. FROCKG project. (2020)

[2] Deliverable 1.3: Requirements Specification. FROCKG project. (2020)

[3] Zafar Habeeb Syed, Michael Röder, and Axel-Cyrille Ngonga Ngomo: “FactCheck: Validating RDF Triples using Textual Evidence”. In Proceedings of the International Conference on Information and Knowledge Management (CIKM), 2018.

[4] Zafar Habeeb Syed, Michael Röder, and Axel-Cyrille Ngonga Ngomo: “Unsupervised Discovery of Corroborative Paths for Fact Validation”. The Semantic Web – ISWC 2019 , page 630–646. Cham, Springer International Publishing, (2019)

[5] Thiago Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van der Lee, Simon Mille, Diego Moussallem, and Anastasia Shimorina: “The 2020 Bilingual, Bi-Directional WebNLG+ Shared Task Overview and Evaluation Results (WebNLG+ 2020)”. Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+), Dec 2020, Dublin/Virtual, Ireland.

[6] Kishore Papineni, Salim Roukos, Todd Ward, and WeiJing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics

[7] Boxing Chen and Colin Cherry. 2014. A systematic comparison of smoothing techniques for sentencelevel BLEU. In Proceedings of the Ninth Workshop on Statistical Machine Translation, pages 362–367, Baltimore, Maryland, USA. Association for Computational Linguistics.

[8] Alon Lavie and Abhaya Agarwal. 2007. METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In Proceedings of the Second Workshop on Statistical Machine Translation, pages 228–231, Prague, Czech Republic. Association for Computational Linguistics.

[9] Maja Popovic. 2017. ´ chrF++: words helping character n-grams. In Proceedings of the Second Conference on Machine Translation, pages 612–618, Copenhagen, Denmark. Association for Computational Linguistics.

[10] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In In Proceedings of Association for Machine Translation in the Americas, pages 223–231.