

This is a preprint of the article: Michalis Mountantonakis, Yannis Tzitzikas, “ChatGPT Responses Validation through Knowledge Graphs”, ERCIM News No. 134 (July 2023)

ChatGPT Responses Validation through Knowledge Graphs

by Michalis Mountantonakis and Yannis Tzitzikas (FORTH-ICS and University of Crete)

The novel artificial intelligence ChatGPT chatbot offers detailed responses across many domains of knowledge; however, quite often it returns erroneous facts even for popular persons, events and places. To tackle this problem, we present GPT•LODS, a novel prototype that annotates and validates ChatGPT responses by leveraging one or more RDF knowledge graphs.

There is a recent trend for using the novel artificial intelligence ChatGPT chatbot, which is an innovative application of large language models (LLMs) that provides detailed and articulate responses across many domains of knowledge. However, in many cases, it returns plausible-sounding but incorrect or inaccurate responses, it does not provide justifications, and its current version has “limited knowledge of world and events after 2021”. On the other hand, there is a high proliferation of knowledge graphs (KGs) that are modeled using the Resource Description Framework (RDF) model over any real domain. These KGs offer high-quality structured data by recording their provenance, whereas most of the popular RDF KGs are updated at least periodically. Therefore, the key notion is how to enable the combination of ChatGPT and RDF KGs, for making it feasible to enrich and validate any ChatGPT response, and this is quite challenging since it requires access to numerous RDF KGs, sources and resources in general.

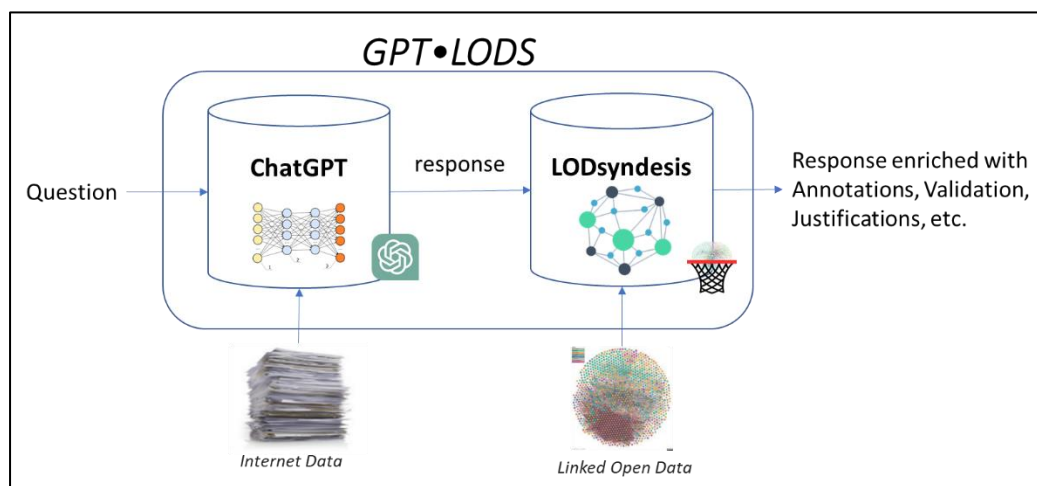


Figure 1: The key notion of combining ChatGPT with RDF knowledge graphs.

The Information Systems Laboratory of the Institute of Computer Science of FORTH designs and develops innovative algorithms and tools for enabling the combination of ChatGPT and RDF KGs. We call the corresponding services “GPT•LODS” [1], [L1] (the name of GPT•LODS stems from the mathematical notation for function composition). The key idea, illustrated in Figure 1, is to send a

question to ChatGPT, which has been trained by using data from web sources (such as Wikipedia, books and news articles), and then to enrich its response by using hundreds of RDF KGs through LODsyndesis [2], [L2], which aggregates data from hundreds of RDF KGs from several domains, containing in total more than 2 billion triples. The current version of *GPT•LODS* (accessible online in [L1], also offering tutorial videos), provides two different types of services: (i) an annotation and enrichment service for enabling the identification, linking and enrichment of the entities of a ChatGPT response, and (ii) a fact-checking service for validating the facts of a ChatGPT response, and accompanying them with provenance information. Below we describe these services through an example where we ask ChatGPT “Who was the scorer of the UEFA Euro 2004 Final”, as shown in Figure 2.

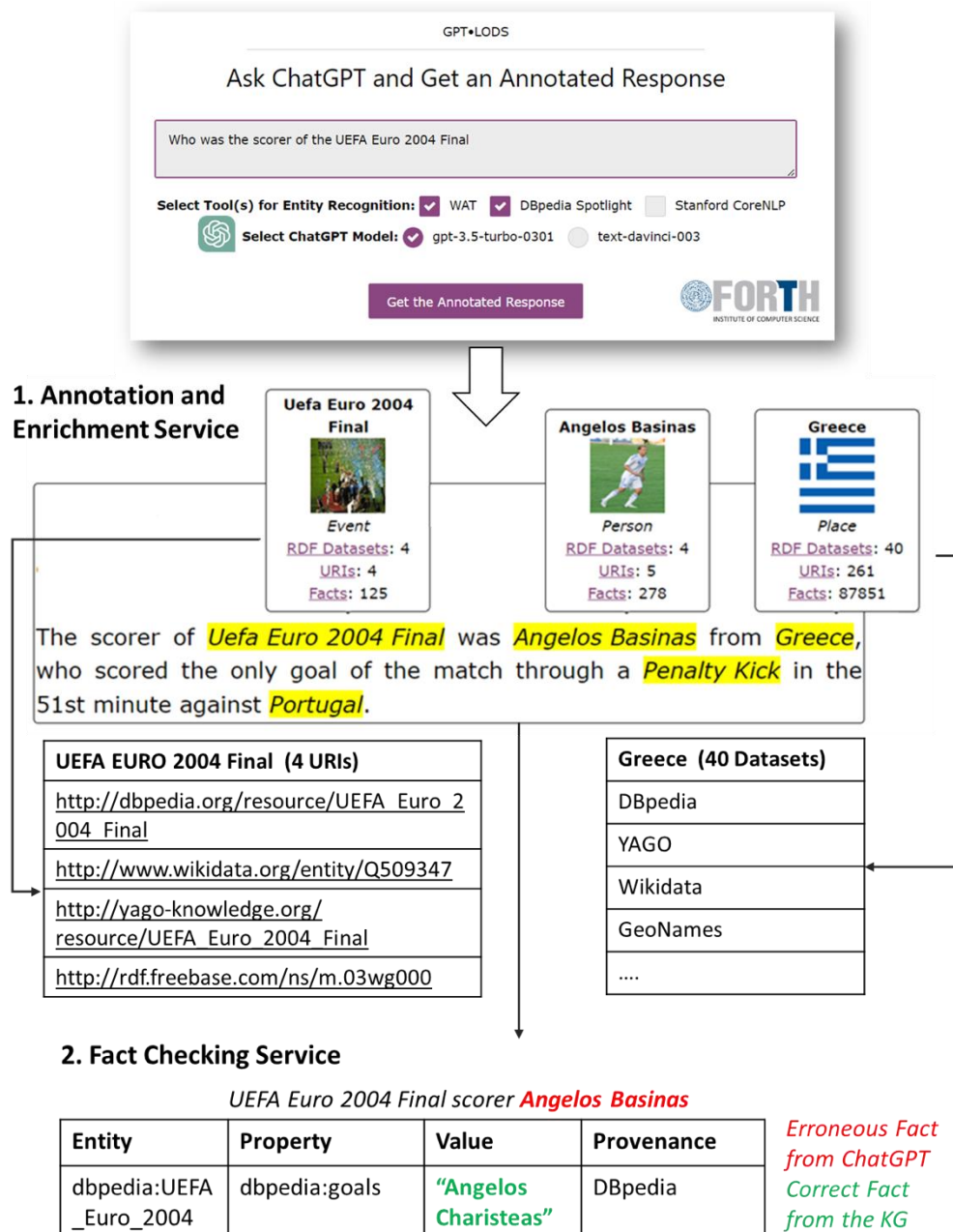


Figure 2: The services of *GPT•LODS*.

The annotation and enrichment service retrieves a ChatGPT response and offers real-time annotation, linking and enrichment of its entities based on hundreds of RDF KGs, by using natural language processing tools (specifically named entity recognition and linking tools). In the example of Figure

2, the system managed to identify and link the entities of the response, in particular, it found more information (links, images, entity type, datasets and facts) for each of the entities of the response, for example, for the entity “Greece” it found 261 URIs and 87 thousand facts from 40 RDF KGs. By clicking on one of those links, one can browse all this information (see the lower part of Figure 2), for example, see on the left side all the URIs for the entity “UEFA Euro 2004 Final” and on the right side the RDF datasets (or KGs) including information about “Greece”.

However, the key problem of the ChatGPT response of Figure 2 is that it contains some erroneous facts: the scorer of the UEFA Euro 2004 Final was “Angelos Charisteas” with a header and not “Angelos Basinas” through a penalty kick. To tackle this problem, the fact-checking service first collects the facts from the ChatGPT response (in RDF format), and then through dedicated algorithms based on semantic web techniques, word embeddings and sentence similarity metrics, finds the most similar fact(s) in the RDF KGs indexed by LODsyndesis. The objective is both to confirm the correct ChatGPT facts and to find the correct answer for erroneous ChatGPT facts from existing RDF KGs. In the lower side of Figure 2, we can see that the fact-checking service managed to find the correct answer for the scorer of “UEFA Euro 2004 Final”, that is, “Angelos Charisteas”, and provided also the right provenance for this information (i.e. the DBpedia KG).

In the future we plan to extend GPT•LODS for supporting fact validation from more types of sources (including web pages), to evaluate the quality of the validation service and to provide a REST API, for enabling the exploitation of these services in various applications.

Links:

[L1] <https://demos.isl.ics.forth.gr/GPToLODS>

[L2] <https://demos.isl.ics.forth.gr/lodsyndesis/>

References:

[1] M. Mountantonakis, and Y. Tzitzikas, “Using multiple RDF knowledge graphs for enriching ChatGPT responses,” In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2023)*, Demo Track. 2023.

[2] M. Mountantonakis, and Y. Tzitzikas, 2020. “Content-based union and complement metrics for dataset search over RDF knowledge graphs,” *Journal of Data and Information Quality (JDIQ)*, vol. 12, no. 2, pp.1–31, 2020.

Please contact:

Michalis Mountantonakis, FORTH-ICS and University of Crete

mountant@ics.forth.gr

Yannis Tzitzikas, FORTH-ICS and University of Crete

tzitzik@ics.forth.gr, +30 2810 391621,