Do you always blindly believe what you are told?

How to explain which pieces of data contributed to your answer and where it came from

Contact persons: Katja Hose (khose@cs.aau.dk) and maybe Luis Galárraga (luis.galarraga@inria.fr)

Our lives depend more and more on the help of "machines", such as online search engines, personal assistants, assisted decision making, etc. It does not need much imagination or science fiction to understand that the more we get used to data-driven algorithmic systems making decisions for us or providing us with recommendations and information, the more we are relying on them in blind trust – without second-guessing whether they are actually acting in our best interest or not. Hence, the fact that these systems more or less resemble black boxes that offer no explanations, has given rise to concerns regarding fairness[1] and trust in such systems. So, the goal of this project is to help make the cores of these black boxes more transparent by providing explanations for their answers. In particular, we will focus on knowledge graphs[2] since they are used as integral components in many of these systems[3,4] to manage and organize information about concepts and entities of interest.

Theme 1: Data Provenance

A basic way of building trust in a system is to enable it to describe the data on which it relies, namely, where this data comes from, how it was pre-processed, integrated, etc. Put differently, we are interested in *data provenance*. For knowledge graphs, this kind of provenance can be expressed and encoded in different ways (PROV-O[5], reification[6], RDF-star[7], etc.). Although different alternatives are available, we still lack systems that efficiently make use of them. This becomes particularly prominent when we consider that knowledge graphs are rarely static but evolve over time, e.g., the president of the United States changes over time. Hence, aspects that projects in this theme could target are, for example, provenance-enhanced knowledge management, provenance-aware query processing in single-graph and multiple-graph settings, time-travel queries over evolving knowledge graphs, etc.

---

[1] Evaggelia Pitoura, Kostas Stefanidis, Georgia Koutrika: Fairness in Rankings and Recommendations: An Overview. CoRR abs/2104.05994 (2021)
[2] Aidan Hogan et al.: Knowledge Graphs. Morgan & Claypool Publishers, 2021, https://kgbook.org/
[3] https://en.wikipedia.org/wiki/Google_Knowledge_Graph
[4] https://cacm.acm.org/magazines/2019/8/238342-industry-scale-knowledge-graphs
[5] https://www.w3.org/TR/prov-o/
[6] https://www.w3.org/wiki/RdfReification
[7] https://w3c.github.io/rdf-star/cg-spec/editors_draft.html

Theme 2: Query Provenance

Whereas Theme 1 focuses on the origin of the data itself, another way to build trust in a system is to provide information about how the system uses the available data to come up with a certain answer for a given query: the so-called *how- or why-provenance*. In the context of knowledge graphs, the principle is that systems provide an expression describing which pieces of the graph were combined in which way to produce a particular answer to a given query[8]. Aspects that projects in this theme could target are, for example, how query provenance can be computed in the presence of multiple (potentially distributed) knowledge graphs or in evolving knowledge graphs, etc.

Theme 3: Interactive exploration of provenance

Even though a system might return information about how an answer was derived for a given query (query/how/why provenance, polynomials) and even though it might be possible to trace the origin and updates of a particular triple (data provenance, reification/RDF-star), it is still very difficult for a non-expert to access the information and explore it. Aspects that projects in this theme could target are, for example, how query and data provenance can be combined[9], how to explore this information efficiently and interactively, provide users with the opportunity to ask "why-not" (or "how-not" – depending on the provenance scheme) questions, i.e., why a particular answer was not part of the result, etc.

Relevant Literature

- Daniel Hernández, Luis Galárraga, Katja Hose: Computing How-Provenance for SPARQL Queries via Query Rewriting. Proc. VLDB Endow. 14(13): 3389-3401 (2021)
- Melanie Herschel, Ralf Diestelkämper, Houssem Ben Lahmar: A survey on provenance: What for? What form? What from? VLDB J. 26(6): 881-906 (2017)
- Daniel Hernández, Luis Galárraga, Katja Hose: Visualizing How-Provenance Explanations for SPARQL Queries. Submitted as a demo paper to TheWebConf 2023 (currently under review, students can build up on this work)
- See footnotes above

[8] Daniel Hernández, Luis Galárraga, Katja Hose: Computing How-Provenance for SPARQL Queries via Query Rewriting. Proc. VLDB Endow. 14(13): 3389-3401 (2021)

[9] It is probably sufficient to consider how prevenance (set of triples involved in deriving an answer) for this purpose instead of why provenance (polynomials describing how triples are combined to derive the result).