

My research vision is to equip machines with knowledge about real-world entities, which is needed in AI applications such as recommendation systems and chatbots. My PhD's focus is on **enriching open-world Knowledge Bases (KBs) with informative negated statements**.

Web-scale KBs store almost *only positive* statements, and miss out on negative statements. Due to incompleteness of open-world KBs, absent statements are considered unknown, rather than false. Not being able to formally distinguish whether a statement is false or unknown poses challenges in a variety of applications, e.g., lack of robustness in QA systems due to uncertainty. My work makes the case for *selective materialization* of statements that do *not* hold. To obtain such knowledge, my research combined information retrieval, machine learning, and text mining methods. It discovers correct and informative negative statements in three different settings:

1. **Negation inference from encyclopedic KBs:** statistical inference method for compiling and ranking simple and complex negative statements in well-canonicalized KBs, based on expectations from positive statements of peer entities, identified using graph- and embeddings-based measures [9, 10, 7].
2. **Negation inference from commonsense KBs:** statistical inference method for compiling informative negative statements about every day concepts using large external taxonomies to identify candidate statements, which are scrutinized using semantic- and pre-trained-language-model-based scorings [11, 6].
3. **Negation inference from text:** extractive method to identify informative negative statements from search engines' query logs using a few handcrafted meta-patterns, which are instantiated with entity mentions to retrieve most frequent occurrences [7].

Sample discovered statements are shown in the table.

Setting	Negated Statement
Wikidata [12] (encyclopedic)	NOT(Stephen Hawking, award, Nobel Prize in Physics) NOT(Switzerland, member of, European Union)
Google query-log (text)	Denmark does not use euro. Amazon does not accept paypal.
Ascent++ [2] (commonsense)	NOT(Lawyer, work with, patient) NOT(Peanut, is, nut)

This research was funded by Max Planck Institute for Informatics (MPII) and the German Research Foundation (DFG). The outcome of this work includes **conference** [7, 11], **demo** [10, 6], **journal** [9], and **workshop** [8] papers. Our introductory paper [7] where we set up the problem and make the case for explicit materialization of negations received the **best paper award** at AKBC'20 as voted by conference attendees. Moreover, I presented a summary of this work as part of a **tutorial** on completeness of open-world KBs at premier conferences, namely VLDB'21, KR'21, ISWC'21, and WWW'22, as part of the sister-conference tracks at ISWC'20 and AKBC'22, and an invited talk at The University of Edinburgh. These appearances received **more than 200 attendees**.

Prior Research. During my master’s studies, I worked on **effective searching of KBs**. KBs are typically searched using triple-pattern queries, which often return too many or too few results, making it difficult for users to find relevant answers to their needs. I presented methods which extend both the searched KB and queries with keywords allowing users to form a wider range of queries, perform automatic query relaxation if no matches are found, and diversify top-k results. This work was funded by the American University of Beirut’s research board (URB), and resulted in **conference** [4] and **journal** [5] publications (the latter was also presented at ISWC’18).

Future Plans. My next goal is to explore interdisciplinary research by mixing my expertise on knowledge discovery with social computing, with focus on problems of importance for humans and society at large. In particular, I am interested the 3 following areas:

1. **Political sciences:** (*short term*) KBs are a good source for interconnections between political figures and their affiliations (parties, movements). I plan to utilize this in **identifying political leanings of social media users**. With access to users’ posts and a large KB, we create their political profiles, including politicians they admire and viewpoints on certain issues. Next, I intend to **study the change of political positions of individuals** by analyzing their political profiles in different time frames, e.g., a person becoming less or more conservative with time.
2. **Sociology:** (*short term*) One of the limitations of existing commonsense KBs is their lack of *cultural context*. Due to recent contributions on extracting cultural commonsense knowledge at scale [3, 1], I intend to extend my work on commonsense negations to **infer cultural no nos**, e.g., leaving a tip at a restaurant in Japan v. in the U.S.
3. **Arabic social media:** (*medium term*) As a native speaker, I plan to exploit the progress in Arabic NLP ¹ to **crawl and curate much needed datasets about the Arab world**, e.g., arabs’ reactions to the latest COVID-vaccines, e.g., tweets, news articles. (*long term*) More generally, I intend to **study Arab democracies**, e.g., Lebanon and Tunisia, in an effort to contribute to addressing interesting open questions such as: How do they differ from other democracies? What are the culture-specific factors in predicting an individual or a group’s political leanings? In retrospect, what are the public’s views on the Arab spring?

Results of this research agenda will be published at premier venues: knowledge discovery (WSDM, CIKM), semantic web (ISWC, ESWC), and social computing (ICWSM, CSCW).

Note: While my areas of interest are generally set, the specifics of this plan are tentative, as I am still exploring various research problems. Therefore, I am also open for research on other related topics, e.g., Demography.

¹<https://aclanthology.org/venues/wanlp/>

References

- [1] Anurag Acharya, Kartik Talamadupula, and Mark A Finlayson. Towards an atlas of cultural commonsense for machine reasoning. *arXiv preprint arXiv:2009.05664*, 2020.
- [2] Tuan-Phong Nguyen, Simon Razniewski, Julien Romero, and Gerhard Weikum. Refined commonsense knowledge from large-scale web contents. *TKDE*, 2022.
- [3] Tuan-Phong Nguyen, Simon Razniewski, Aparna Varde, and Gerhard Weikum. Extracting cultural commonsense knowledge at scale. *WWW*, 2023.
- [4] [Hiba Arnaout](#) and Shady Elbassuoni. Result diversity for rdf search. In *IC3K*, 2016.
- [5] [Hiba Arnaout](#) and Shady Elbassuoni. Effective searching of RDF knowledge graphs. *JWS*, 2018.
- [6] [Hiba Arnaout](#), Tuan-Phong Nguyen, Simon Razniewski, and Gerhard Weikum. Uncommonsense in action! informative negations for commonsense knowledge bases. In *WSDM*, 2023.
- [7] [Hiba Arnaout](#), Simon Razniewski, and Gerhard Weikum. Enriching knowledge bases with interesting negative statements. In *AKBC*, 2020.
- [8] [Hiba Arnaout](#), Simon Razniewski, Gerhard Weikum, and Jeff Z. Pan. Negative knowledge for open-world wikidata. In *WWW Companion*, 2021.
- [9] [Hiba Arnaout](#), Simon Razniewski, Gerhard Weikum, and Jeff Z. Pan. Negative statements considered useful. *JWS*, 2021.
- [10] [Hiba Arnaout](#), Simon Razniewski, Gerhard Weikum, and Jeff Z. Pan. Wikinegata: A knowledge base with interesting negative statements. *PVLDB*, 2021.
- [11] [Hiba Arnaout](#), Simon Razniewski, Gerhard Weikum, and Jeff Z. Pan. Uncommonsense: Informative negative knowledge about everyday concepts. In *CIKM*, 2022.
- [12] D. Vrandečić and M. Krötzsch. Wikidata: a free collaborative knowledge base. *CACM*, 2014.