What Computers Should Know, Shouldn't Know, and Shouldn't Believe

Gerhard Weikum
Max Planck Institute for Informatics
Saarland Informatics Campus, Germany
weikum@mpi-inf.mpg.de

Automatically constructed knowledge bases (KB's) are a powerful asset for search, analytics, recommendations and data integration, with intensive use at big industrial stakeholders. Examples are the knowledge graphs for search engines (e.g., Google, Bing, Baidu) and social networks (e.g., Facebook), as well as domain-specific KB's (e.g., Bloomberg, Walmart). These achievements are rooted in academic research and community projects. The largest general-purpose KB's with publicly accessible contents are BabelNet, DBpedia, Wikidata, and Yago. They contain millions of entities, organized in hundreds to hundred thousands of semantic classes, and billions of relational facts on entities. These and other knowledge and data resources are interlinked at the entity level, forming the Web of Linked Open Data.

What Computers Should Know

Despite their wealth of facts, none of the major KB's can ever be complete. KB's have been constructed and are maintained with focus on encyclopedic knowledge about prominent and business-relevant entities, and often with strong reliance on Wikipedia. This way of knowledge acquisition misses out on a number of important dimensions, posing open challenges for next-generation KB's.

Temporal and Emerging Knowledge: Change is the only constant in knowledge. Attribute values of entities (e.g., city populations) and relationships between entities (e.g., the CEO of a company or a person's spouse) change over time. New entities of interest are created all the time and need to be added to the KB (e.g., new songs, sports matches, babies of celebrities). Existing entities may be irrelevant for a KB at some point, but become prominent at a later point.

So KB's must be continuously updated. This requires keeping *versions of facts*, along with their *temporal validity* scopes. Some of the major KB's have rigorously followed this principle (e.g., [5]). However, as we aim to capture long-tail and emerging entities as well, advanced methods are needed for capturing the relevant time points and time spans (see, e.g., [7, 22]).

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3038914

Salient Facts: KB's have been built in an opportunistic manner, mostly relying on Wikipedia. If Wikipedia does not have the information or if a fact is stated only in sophisticated form in the article's text, all the KB's miss out on it. What is notable about the Nick Cave album "Abbatoir Blues"? Many KB's contain this album, listing its individual songs. No KB points out, though, that the song "Let the Bells Ring" is about Johnny Cash and that the song "O Children" is used in one of the Harry Potter movies. Part of the problem is the poor coverage of predicate types: KB's are missing predicates like *songIsAbout*. There is a research opportunity here: reasoning about KB incompleteness [4] and capturing the truly notable facts.

Commonsense Knowledge: Automatically constructed KB's have mostly focused on harvesting encyclopedic fact knowledge. However, for semantic search and other intelligent applications (e.g., conversational bots in social media), machines need a broader understanding of the world: properties of everyday objects, human activities, plausibility invariants and more.

This calls for the goal of distilling commonsense knowledge from Internet sources: properties of objects like size, color, shape, parts or substance of which an object is made of, etc., and knowledge on which objects are used for which activities as well as when and where certain activities typically happen. For example, a rock concert involves musicians, instruments – almost always including drums and guitars, speakers, a microphone for the singer; the typical location is a stage, and so on. Projects on acquiring commonsense include ConceptNet [17] and WebChild [19, 21]. An exciting direction is organizing knowledge on *human activities* [2, 20, 23]. However, it is still a long way to go for computers to learn what every child knows.

Socio-Cultural Knowledge: Another dimension where KB's have a huge gap is the socio-cultural context. Consider statements on people making discoveries or inventions. On first glance, one would expect that these are objective and universally agreed upon. On second thought, however, it becomes clear that it depends on the background and viewpoint of users. For example, people in the US would say that the computer was invented by Eckert and Mauchley, whereas a German would give the credit to Konrad Zuse and a British may point out Alan Turing (or perhaps Charles Babbage). This depends not just on geographical context: teenagers may widely think of Steve Jobs as the (re-) inventor of the (mobile) computer. For commonsense knowledge, it is even more critical to capture socio-cultural contexts.

What Computers Shouldn't Know

Some of the knowledge graphs of the big Internet players are reported to contain several hundred millions of individual entities, with billions of facts. This suggests that knowledge is captured not only about prominent people, places and products, but also about common users of online services. For example, the KB of a search engine or social network may contain information about a user's purchases, traveling, clicks, etc. By linking the information with general world knowledge, the provider can perform generalizable inference. For example, a user who often visits Thai restaurants probably likes spicy food and would appreciate recommendations for Szechuan restaurants.

All this happens in the interest of the user and with her consent. However, it opens the door towards comprehensive data gathering about a user's online behavior and, indirectly, her life. With extensive tracking by third-party companies, this may result in detailed user profiling over extended time periods. Such long-term profiling in turn incurs risks of intruding on a user's privacy (see, e.g., [15] for a study on slowly evolving loss of privacy). This is not necessarily in the form of de-anonymizing the user, but could result from generating personalized ads that the user may not want to be displayed while her co-workers can see them (e.g., about medical topics). More severe forms of discrimination may arise, such as data-driven learning-based ratings of users and algorithmic decision-making on job applications, visa approval or denial, health insurance issues, etc.

Traditional privacy mechanisms are geared for anonymizing single datasets and do not address these kinds of privacy and discrimination risks. In fact, one can argue that there is no way of completely preventing privacy breaches, as users do want to post information and opinions and share them with online communities. What is crucially needed instead, is a way of making users aware of such risks and provide them with informative guidance on their online actions [13]. We are working on principles for a *privacy risk advisor* tool that assists individual users in this spirit [1].

What Computers Shouldn't Believe

Knowledge gathering has inherent uncertainty, as information extraction cannot guarantee hundred percent correctness. The more knowledge bases tap into contents about long-tail entities, individual users and sophisticated relationships, the higher the degree of uncertainty and risk of capturing flawed information. The same argument applies to acquiring commonsense and socio-cultural knowledge. This raises the challenge of assessing the credibility of information and the trustworthiness of underlying sources.

Coping with doubtful statements for KB curation has been investigated under the theme of *fact checking*, sometimes called *truth discovery*; see [8] for a survey. A typical use case is to debunk an alleged fact on Obama's birthplace being Nairobi in Kenya, and identifying the true birthplace Honolulu on Hawaii. This can be done based on Web evidence for and statistics about alternative variants, typically using joint inference on fact credibility and source trustworthiness (see, e.g., [10, 11, 12]). However, the problem is much broader and more critical, as wrong or even fabricated claims often appear in natural language text, such as news articles or social media posts. This requires a deeper analysis of textual contents, with reasoning about the origin and temporal dissemination of claims (see, e.g., [3, 6, 9, 14, 16, 18]).

Both traditional media (e.g., newspapers or digital news feeds) and social media (e.g., micro-blogs or online communities) exhibit an increasing fraction of misleading or even manipulative contents, ranging from completely faked over "post-factual" (with pseudo-truth based on how the world "feels" to someone) all the way to wild claims. Since prominent politicians, business leaders and other stakeholders use these channels to influence the public, the media landscape is more and more becoming a twilight zone and battleground. Inferring what is factual truth and what is "post-truth", "alternative truth" or plain fake in this space, and explaining this to users, is a huge challenge and research opportunity.

Acknowledgement

Our research on privacy and trust is partly supported by the ERC Synergy Grant 610150 (imPACT).

References

[1] J. Biega et al.: R-Susceptibility: an IR-Centric Approach to Assessing Privacy Risks for Users in Online Communities. SIGIR'16

[2] C.X. Chu et al.: Distilling Task Knowledge from How-To Communities. WWW'17

[3] W. Ferreira, A. Vlachos: Emergent: a Novel Data-Set for Stance Classification. HLT-NAACL'16

[4] L. Galarraga et al.: Predicting Completeness in Knowledge Bases. WSDM'17

[5] J. Hoffart et al.: Yago2: a Spatially and Temporally Enhanced Knowledge Base from Wikipedia. Art.Int. 194, 2013

[6] S. Kumar et al.: Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. WWW'16

[7] E. Kuzey et al.: As Time Goes By: Comprehensive Tagging of Textual Phrases with Temporal Scopes. WWW'16

[8] Y. Li et al.: A Survey on Truth Discovery. SIGKDD Explorations 17(2), 2015

[9] J. Ma et al.: Detecting Rumors from Microblogs with Recurrent Neural Networks. IJCAI'16

[10] S. Mukherjee et al.: People on Drugs: Credibility of User Statements in Health Communities. KDD'14

[11] N. Nakashole, T. Mitchell: Language-Aware Truth Assessment of Fact Candidates. ACL'14

[12] J. Pasternack, D. Roth: Latent Credibility Analysis. WWW'13

[13] S.T. Peddinti et al.: Understanding Sensitivity by Analyzing Anonymity. IEEE Security & Privacy 13(2), 2015

[14] K. Popat et al.: Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media. WWW'17

[15] M.-A. Rizoiu et al.: Evolution of Privacy Loss in Wikipedia. WSDM'16

[16] M. Samadi et al.: ClaimEval: Integrated and Flexible Framework for Claim Evaluation Using Credibility of Sources. AAAI'16

[17] R.Speer, C.Havasi: Representing General Relational Knowledge in ConceptNet 5. LREC'12

[18] B. Tabibian et al.: Distilling Information Reliability and Source Trustworthiness from Digital Traces. WWW'17

[19] N. Tandon et al.: WebChild: Harvesting and Organizing Commonsense Knowledge from the Web. WSDM'14

[20] N. Tandon et al.: Knowlywood: Mining Activity Knowledge From Hollywood Narratives. CIKM'15

[21] N. Tandon et al.: Commonsense in Parts: Mining Part-Whole Relations from the Web and Image Tags. AAAI'16

[22] Y. Wang et al.: Coupling Label Propagation and Constraints for Temporal Fact Extraction. ACL'12

[23] Z. Yang, E. Nyberg: Leveraging Procedural Knowledge for Task-oriented Search. SIGIR'15