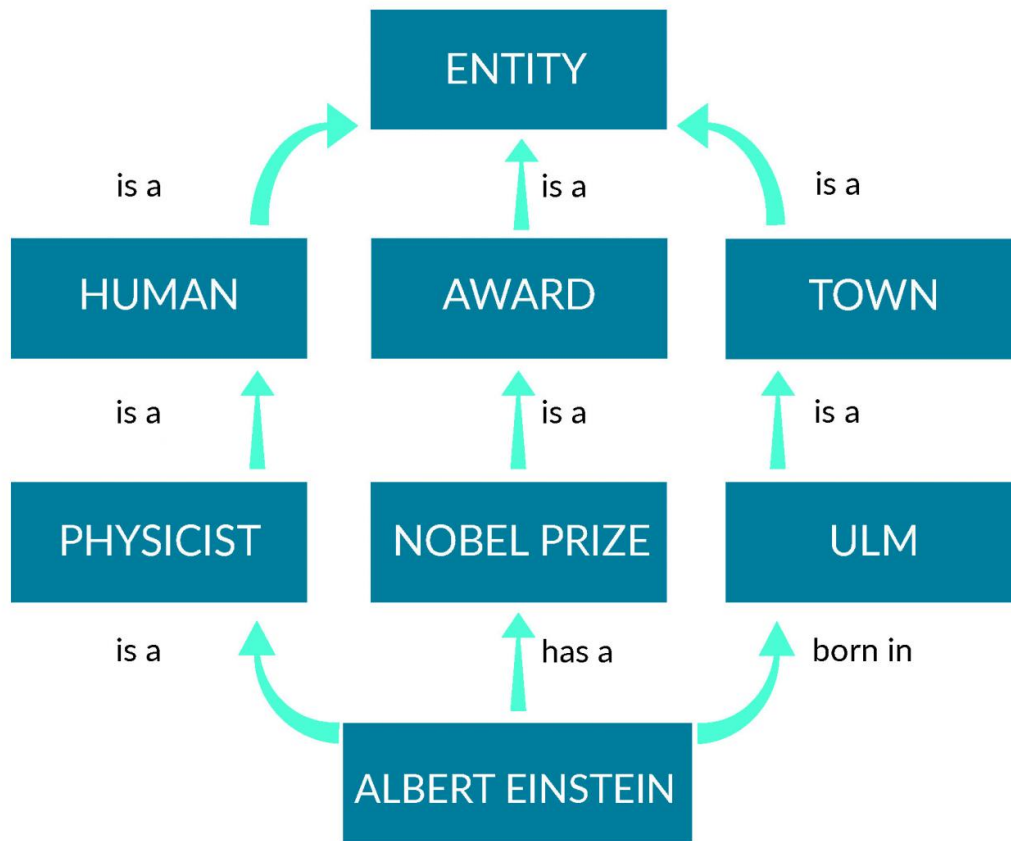# Source code of award-winning knowledge base is now available for everyone

September 1 2017

Through automatically generated entities and their relationships, computers can answer complex search queries. Credit: MPI

Almost every word has more than one meaning. Modern search engines solve this problem using knowledge bases. Yago was one of the first knowledge bases, developed by scientists at the Max Planck Institute for Informatics in Saarbrücken and the Télécom ParisTech in Paris. Last week, the researchers received an award for their work on Yago from the most important scientific journal in the field of artificial intelligence. Today, they are releasing Yago's source code.

"If you, for example, do a internet search for the German term 'Allianz', this is merely a collection of letters for the search engine," explains Professor Gerhard Weikum, Scientific Director at the Max Planck Institute for Computer Science in Saarbrücken. "With the aid of a [knowledge](#) base, this chain of letters can be mapped to possible meanings, for example, 'Allianz SE', the German financial and insurance services company, or the 'Rebel Alliance' from Star Wars films." These days, we cannot imagine not having this background knowledge in search engines, and it is only by the use of knowledge bases that Google for example can also show stock market prices, logos, and the managing director of Allianz SE in addition to the Web search results.

The knowledge bases began as academic research projects. "Especially Yago and, only a little later, DBpedia pioneered this field," Weikum says. The Yago project was the topic of the doctoral dissertation in 2007 by Fabian Suchanek at the Max Planck Institute for Informatics in Saarbrücken. More and more researchers took part in the project. Today, Yago is a collaboration of the Max Planck Institute, the Télécom ParisTech University, where Suchanek now holds a professorship, and the Max Planck spin-off, Ambiverse. Yago contains the knowledge of Wikipedia and other sources in a form that is readable for the computer. For example, the system knows that Allianz SE is headquartered in Munich, but also that the Rebel Alliance is fighting in the Star Wars

universe.

Since many applications in various areas of industry are using artificial intelligence to be more efficient and, most importantly, easier to operate, Yago is seeing wide use. Applications can, with the help of Yago, search in several languages as well as classify facts both spatially and temporally. Therefore, search queries like this are possible: "Tell me all the scientists who lived in the 20th century, were awarded a Nobel Prize, and were born in the greater Stuttgart area." An application example is provided by Primal.com, a Canadian startup, which uses Yago to help businesses better understand the interests of their customers and make content and product recommendations that serve their individual needs. The most high-profile use of Yago in the last ten years was when IBM used it in the Watson artificial intelligence system, which won the Jeopardy! quiz show in 2011.

Ambiverse itself provides another application. The spin-off of the Max Planck Institute analyzed the Panama Papers using Yago a few months ago. In only a few hours, Ambiverse succeeded in getting new insights about the owners of the Panama accounts, a task that would otherwise have required extensive manual effort.

Such an analysis was only possible because Yago categorizes all persons into a semantic structure. Up to now, computers have been able to store vast amounts of data, but they have been able neither to classify nor to understand that data. With Yago's structure, however, the computer has the ability to distinguish, for example, between Gerd Müller, the '74 soccer world champion, and Gerd Müller, the German minister of development. Johannes Hoffart, Managing Director of Ambiverse, explains: "Yago assigns people to contexts, so it is easy to determine whether more athletes or more politicians have accounts in Panama", Hoffart points out. Such structures used to be created manually, which, however, is very complex - both in terms of production and verification.

With the procedure researchers have developed for Yago, they cleverly get around that tedious task. Yago systematically digs for information in Wikipedia's knowledge pool. Not only is the information as to whether the people are athletes or politicians brought into a machine-readable form, but also the relationships between them. Thus, the relation "has headquarters in" connects Allianz SE with the city of Munich. Every Wikipedia article becomes an entity in the Yago knowledge base. In this way, in Yago the researchers are able to provide nearly 17 million entities and 150 million relationships between them.

Last week, the researchers behind Yago were awarded the Prominent Paper Award which recognizes outstanding papers published in the Artificial Intelligence Journal (AIJ) that are exceptional in their significance and impact over the past 5 years. The AIJ is the most important journal in the field of artificial intelligence. The prize itself was awarded at this year's IJCAI in Melbourne, the most significant scientific conference in the field of AI. Today, they are releasing the source code for their knowledge base on the platform GitHub under the open source license GNU GPL v3. This software license ensures that everyone is allowed to run, study, modify, and share the protected program code. "The developer community is getting a high-quality knowledge base," explains Prof. Fabian Suchanek, founder of the project. "We are not only hoping for further applications for Yago, but also look forward to contributions from the developer community."

The source code can now be downloaded from the following link: github.com/yago-naga/yago3

  More information: YAGO project page: www.mpi-inf.mpg.de/de/departme … arch/yago-naga/yago/

Citation: Source code of award-winning knowledge base is now available for everyone (2017, September 1) retrieved 27 April 2024 from https://phys.org/news/2017-09-source-code-award-winning-knowledge-base.html