The Digital **Apprentice**

Two years ago, countless stories appeared in the media about a computer winning *Jeopardy*, a tricky quiz show on American television. This raises a number of questions: What can computers know? How do they use this knowledge for language comprehension and for dialog with human beings? And what can be done when machines collect facts about users that aren't in the users' best interests?

TEXT **GERHARD WEIKUM**

Do computers have the potential to be the intellectual equal of humans, or even to be superior to them? Scientists engaged in research in informatics and its sub-discipline, artificial intelligence, have been seeking an answer to this question ever since Alan Turing proposed a test more than fifty years ago: Can a computer that is communicating with a human partner via a text interface behave in such a way that the person, even after some considerable time, would be unable to tell whether they are communicating with a human being or a machine?

Human knowledge is almost completely digitized and systematically organized.

Today, human knowledge – in books, essays, news reports and other texts – is almost completely digitized and systematically organized. The most famous example of a digital knowledge corpus is the online encyclopedia *Wikipedia*. But computers are initially unable to understand *Wikipedia* because the texts are written for people.

All that has changed in recent years. Extensive, machine-readable knowledge bases, such as Google's *Knowledge Graph*, enable computers to comprehend texts in a way that goes beyond mere identification of the words in a query, for instance. Computer algorithms create semantic relations between the terms and, in doing so, enable a semantic search. They can even then correctly answer questions containing ambiguous terms. And their understanding of semantics means that computers can also understand the meaning of texts written for people, like the articles in *Wikipedia*.

The knowledge bases that facilitate this deeper understanding of language were produced largely automatically, and are constantly being updated and expanded. The *Knowledge Graph* knows more than twenty million people, places, movies, pharmaceuticals, sports events, and much more besides, along with more than a billion facts about these entities and how they relate to one another. Google uses this formidable knowledge to better understand search queries, to improve its search result rankings, to provide better recommendations for users of YouTube and

Photo: Ikon Images/Fi Online

other web portals, and to make intelligent suggestions of restaurants, concerts and other places to go.

Three projects in particular have developed the methods for automatically constructing such comprehensive knowledge bases. These are *DBpedia* at the Freie Universität Berlin and the University of Leipzig; *Freebase*, which was bought by Google and now forms the heart of the *Knowledge Graph*; and *Yago*, which we at the Max Planck Institute for Informatics have been developing since 2005.

An important first dimension of digital knowledge consists in collecting so-called entities, naming them unambiguously and assigning them to semantic classes, such as people, places, organizations or events. It is primarily *Yago* that does this on a grand scale by employing smart algorithms to link category names from *Wikipedia* with the manually created thesaurus *WordNet*. The resulting knowledge base contains almost 10 million entities and more than 300,000 fine-grained and hierarchically organized classes, such as politicians, musicians, bass players, rock ballads, heavy metal songs, benefit concerts and open-air operas.

The second dimension of a knowledge base comprises facts about entities. These include not only such attributes as the height of a soccer goalkeeper and the number of international matches he has played in, but also relations between entities, such as the goalkeeper's birthplace, the teams he has played for, his wife's name, a country's capital or the board members of a certain company.

Language is often ambiguous. Names and phrases can be interpreted in different ways.

Finally, the third dimension is made up of rules that express general relations – regardless of specific entities. These include rules like the fact that each person has precisely one birthplace and that a country's capital city must be located in that country. However, rules like these can entail uncertainties – they don't necessarily apply 100 percent of the time. A person is likely to live in the same town as their spouse, or in the town where they work.

Machines need general knowledge of this kind in order to logically link several facts together. If you don't have any point of reference, for instance, for where Angela Merkel lives, but you do know that her husband works at Humboldt University in Berlin, the computer is able to conclude that Germany's Chancellor lives in Berlin.

Language is often ambiguous. This may be due to sentence structure, but much more frequently it is because names and phrases can be interpreted in different ways. Consider the following sentence: "Page played *Kashmir* on his Gibson." Is this about Larry Page, the founder of Google, meeting actor and director Mel Gibson on the edge of the Himalayas? That obviously makes no sense! People can see that immediately on the basis of their empirical knowledge; a machine, however, must analyze the sentence systematically and algorithmically. In fact, what we have here is a statement about Led Zeppelin guitarist Jimmy Page playing the song *Kashmir* on a Gibson Les Paul guitar.

In order to really understand a sentence, the machine also has to recognize and semantically interpret the relations between the entities involved. The verb "play", for instance, can relate to games, sports, music, tricks and much more besides. The probability of "play" being used in the sense of *MusicianPerformsSong* is actually very high if the ambiguous names "Page" and "Kashmir" denote a musician and a piece of music.

By the same token, a sentence employing "play" in the above-mentioned context of *MusicianPerformsSong* is highly likely to mention a musician and a song. These mutual dependencies in the interpretation of verb and noun phrases are resolved with the help of optimization algorithms.

Digital knowledge combined with ample statistics and clever algorithms therefore gives machines an astonishingly strong language comprehension ability. And of course this doesn't stop at single sentences in statement form, but also encompasses questions, en-

tire paragraphs, lengthy essays and scientific publications, as well as man-machine dialogue.

A difficult example of a sentence in question form is: "Who did scores for westerns?" Here, the analysis must work out that "scores" relates to movie soundtracks, that "westerns" refers to western movies, and that the slang formulation "did" should be interpreted as denoting the *ComposedMusic* relation. With this linguistic understanding, the computer can supply an answer straight out of its knowledge base – perhaps Ennio Morricone, who wrote the soundtrack to the movie *Once Upon a Time in the West*.

A computer's knowledge and language processing capabilities are still subject to immense limitations. Frequently, it all hinges on the abundance of the underlying statistics or the extent of training in learning techniques. And then there are languages like Mandarin, which are very difficult to parse and display a much more complex degree of ambiguity than is present in English or German. In some languages, such as Bambara and Urdu, there is no sizable corpus of digital texts and thus no comprehensive statistics are available.

However, taking the advances of the past decade and extrapolating them, we may possibly be able to expect the kind of performance that could come very close to passing the above-mentioned Turing test, perhaps as early as 2020. We could give the computer a school biology textbook "to read" – and the machine would then be able to answer questions at the level of an oral school-leaving exam. Or imagine a game in which several online users prepare meals with a virtual version of British chef Jamie Oliver. In order for Jamie to be able to react to the mistakes his apprentices make while preparing tiramisu, the computer must analyze their conversations and gestures, their facial expressions and visual impressions, and combine them with its culinary knowledge.

Thirty years ago, the field of artificial intelligence conducted a now-derided attempt to implement automatic expert systems for medical diagnosis. Though the endeavor failed at the time, it is now coming within reach in varied form. Imagine a doctor meeting with a patient to discuss the patient's symptoms and the results of their first lab tests. The computer listens in, adopting the role of advisory assistant.

With its encyclopedic knowledge of the subject at hand, the digital assistant can supply crucial hints as to potential diagnoses that can be ruled out, or recommend additional tests that can specifically discriminate various hypotheses. The computer can also join in the conversation and address questions to the doctor or the patient. In this future scenario, the machine plays a very significant role, but leaves the decisions and the responsibility to the human expert.

We may potentially fall prey to user tracking, advertising and other effects we didn't necessarily ask for.

Digital knowledge and intelligent language comprehension don't stop at news, celebrities and general knowledge. They are also methodological building blocks that can be employed to collect knowledge about all of us and our likes and preferences, which can be used to make intelligent recommendations and facilitate man-machine interaction. The sources from which this is drawn are our many and varied interactions with the Internet – be it through our membership in social networks or through our smartphone and everything we do with it.

This can potentially cause us to fall prey to user tracking, advertising and other effects we didn't necessarily ask for. In year one after the NSA scandal, it's obvious how much the privacy of every one of us can be affected by that. Digital background knowledge plays a key role, as the following fictional scenario illustrates.

Zoe, a young woman from Namibia who is studying in Europe, posts photos and other material on her page in a social network. She also recommends to her friends there movies and music, including the indie

Photo: Ikon Images/Fr Online

rock singer Nive Nielsen from Greenland. Zoe is known in the network by her real name and her brief public profile.

Zoe has thyroid problems, takes the drug Synthroid and suffers from side effects. She finds an online forum on health matters, joins under a pseudonym and participates in discussions. Lastly, Zoe also uses search engines to research alternative medications, such as Levothroid, but also to search for movies about Apartheid or information about her favorite singer, Nive Nielsen. The search engines recognize Zoe only as an anonymous user, but an Internet observer from the tracking and targeting industry can collect her search and click history over an extended period of time.

> The Privacy Advisor has very personal knowledge about Zoe.

This may look like a harmless scenario, but there's a great deal more to it. An algorithm with background knowledge could make connections between Zoe's three identities in the digital world. The attacker could use a knowledge base to work out that Synthroid and Levothroid are drugs to treat the same condition, namely an underactive thyroid. With the help of other hints, it could then conclude that the person in the health forum and in the search history is one and the same.

Moreover, there is an extremely low statistical probability of two different young women from Africa being interested in the same Greenlandic singer and other non-mainstream topics. The attacker can thus link the search history with Zoe's identity in the social network. And so it follows that Zoe must be the same person who discussed her thyroid problems in the health forum. This opens the door to unwanted junk mail, possible issues with her health insurance, and other consequences that may turn out to be more than merely unpleasant.

What we outlined here is an automated attack on Zoe's privacy. It thrives on precisely the same knowledge and language technology on the part of computers that we previously viewed as a help and a blessing for mankind. Counteracting it systematically and permanently could itself be based on digital knowledge and language technology: a personal software tool called Privacy Advisor. Such a tool would continuously observe what Zoe does on the Internet, would be aware of the activities she engages in and the things she likes. And it would permanently analyze the risk inherent in the extent to which Zoe reveals critical things about herself that a powerful attacker could exploit. If and when the tool raises an alarm, it would need to explain the situation to Zoe and suggest what she should do instead to mitigate the risk.

The Privacy Advisor is a concept that is, in fact, based to a large extent on machine knowledge and language comprehension. But it has an advantage over potential attackers: not only does it have world knowledge and general knowledge, it also has very personal knowledge about Zoe. For the tool to be trusted by Zoe, it needs to be designed as open-source software and to have been verified by many different programmers. It draws its effectiveness from being configured around Zoe herself, and from the personal knowledge base it references.

Michael Backes (Saarland University), Peter Druschel and Rupak Majumdar (Max Planck Institute for Software Systems) and the author are working on realizing this vision within the imPACT project, which is funded by an ERC Synergy Grant. The project aims to achieve a scientific understanding of all relevant dimensions of the social marketplace the Internet has become, along with its potential tensions: besides privacy itself, user accountability, service compliance and trust in information and knowledge are fundamental pillars of a future Internet.

This article illustrates the extent to which a computer is capable of acquiring knowledge and language – intellectual skills that seem to be reserved for humans. We have seen that today's machines automatically collect and organize digital knowledge on a large scale and use it for an ever-improving linguistic understanding. The following hypotheses are presented in the interest of stimulating further contemplation and discussion:

Photo: Ikon Images/Fi Online

Machines will be vastly superior to man in many application scenarios in the not-too-distant future, such as in answering knowledge-intensive questions or in automatically abstracting long texts or entire corpora and preparing them for analysis. Machines will also be capable of passing high school level exams. As such, machines will come very close to passing the Turing test. One could consider this the simulation of intelligent behavior on the basis of knowledge, statistics and algorithms. When it comes to the effect elicited in applications, it is irrelevant whether the intelligence we're dealing with is "artificial" or "real."

In situations that call for intuition and cognitive flexibility, machines will never really be superior to people, but they may prove to be indispensable assistants. One example of such a situation would be in helping with medical diagnoses. Here, a computer could function as an almost full-fledged participant in the discussion for both doctor and patient. But there will always be situations where a machine is unable to imitate us: humor, irony, flirting and other emotions will surely remain the domain of humans for some considerable time to come.

We can teach computers to warn us if we are revealing too much of ourselves on the Internet.

Given that computers are increasingly analyzing the meaning of texts in social media and making connections between words and phrases, a whole raft of new applications is opening up to them – but these aren't always in the best interests of users: one of the things that semantic understanding does is enable machines to analyze us humans more comprehensively. However, we don't need to sit back and take it: ultimately, we can teach computers to use their knowledge of semantics and relations to warn us if we are giving away too much information on the Internet that algorithms could link to create detailed personality profiles. ◀

THE AUTHOR

Prof. Dr. Gerhard Weikum, born in 1957, studied informatics and earned his Ph.D. at TU Darmstadt. He worked at that university as an Assistant Professor in 1987. Other post-doc positions he held were at MCC in Austin, Texas, and at ETH Zurich, where he also held a professorship between 1990 and 1994. He later went to Saarland University in Saarbrücken. He has been a Director and Scientific Member at the Max Planck Institute for Informatics since 2003.

Photo: Private collection