



Text Understanding for Computers

T.M. Kenter

Text Understanding for Computers

A long-standing challenge for computers communicating with humans is to pass the Turing test, i.e., to communicate in such a way that it is impossible for humans to determine whether they are talking to a computer or another human being. The field of *natural language understanding* — which studies automatic means of capturing the semantics of textual content — plays a central part in this long-term goal of artificial intelligence research. Natural language understanding can itself be understood at different levels. In this thesis, we make contributions to automatic understanding of text at the level of words, short texts, and full documents.

Understanding texts at the word level, means understanding how words relate to each other semantically. For example, do two words or phrases mean approximately the same thing? Does a particular word still mean the same thing it used to, say, 50 years ago? Or, as is the question central to the first part of this thesis, can we automatically detect which words people used in different periods in time to refer to a particular concept?

When word-level semantics are understood to a sufficient degree, an attempt can be made at capturing the meaning of short pieces of text, such as sentences. The question we ask ourselves in the second part of this thesis is: can we automatically determine, from the word-level up, if two sentences have a similar meaning?

Finally, in the third part of this thesis, document-level text understanding is the focus of our interest. In particular, we study multiple approaches to the reading comprehension task, where a computer reads a document and answers questions about it.

Today, text understanding is at an unprecedented level. It plays a role in many tasks, among which are dialogue systems, or conversational agents. For the first time in history people can have more or less meaningful conversations with digital assistants, like the Google Assistant, Amazon's Alexa, Apple's Siri and Microsoft's Cortana. Still, the conversational systems we have today are miles away from omniscient systems like the Star Trek computer or any other hyper-intelligent AI envisioned in science fiction, and they will probably not pass the Turing test any time soon. However, they can perform mundane tasks like setting alarms, making appointments in a personal calendar, switching on the lights, playing television shows, or reporting what the weather is going to be like, all simply by being asked to do so in natural language. As new functionalities are being added, time will tell which ones will catch on, a process itself influencing how we will interact with computers. Do we stick to them being personal assistants that can do simple tasks for us? Or do we end up

having full-fledged conversations with them, like the one the bots on the cover of this book seem to be engaged in?