

AI tool summarizes lengthy papers in a sentence

November 24 2020, by Peter Grad



Credit: Pixabay/CC0 Public Domain

Scholars have a nifty way of alerting colleagues to lengthy treatises that they find simply not worth their time to read.

They tag such documents "tl;dr"—too long, didn't read.

It's kind of a 21st century spin on the 420-year-old notion Shakespeare's Polonius relayed to the king and queen in "Hamlet": "Brevity," he suggested, "is the soul of wit."

The Allen Institute for Artificial Intelligence in Seattle has taken both sentiments to heart and this week unveiled a system that offers extreme condensation of lengthy computer-science reports to slash the time it take to review such literature.

Semantic Scholar is a [research tool](#) powered by AI and used for scientific research. With its new summarization feature, it surveys massive numbers of [scientific research](#) papers and reduces them to one-sentence summaries. More than 7 million users a month have been accessing Semantic Scholar.

Currently, there are 10 million computer-science papers in Semantic Scholar's database. According to Dan Weld, who oversees the database, papers from other disciplines will gradually be added.

The system offers a great advantage to researchers who up to now have had to rely on scanning numerous titles and often lengthy abstracts, an especially trying task on mobile devices. Following early tests, reaction has been positive. "People seem to really like it," Weld said.

There have been a variety of Natural Language Processing programs developed over the years to summarize documents. They generally use one of two approaches: the extractive approach focuses on selecting representative text and using it verbatim in the summary. For instance, Paper Digest, developed in 2018, appears to extract key sentences rather than rewriting findings in its own words.

The other approach is abstractive; it uses natural language generation algorithms to create summaries with original wording. Improvements in AI natural language generation in recent years have made this approach the favored one among programmers.

Semantic Scholar is notable for achieving the greatest compression rate of all summarizing tools. With scientific papers averaging 5,000 words, Semantic Scholar's summaries are around 21 words. That averages to summaries 1/238th the size of the reports. The closest Semantic Scholar competitor compresses documents to only 1/36th of the report size.

According to Jevin West, an information scientist at the University of Washington in Seattle who tested the new program, "I predict that this kind of tool will become a standard feature of scholarly search in the near future. Actually, given the need, I am amazed it has taken this long to see it in practice."

He noted that it is not yet perfect, "but it's definitely a step in the right direction," he said.

The Allen Institute team is making their code available for free. They also have set up a demonstration site open to all. scitldr.apps.allenai.org/

Currently, only papers written in English are being accepted. But the program's authors hope to include documents in other languages eventually.

More information: github.com/allenai/scitldr

www.semanticscholar.org/

Citation: AI tool summarizes lengthy papers in a sentence (2020, November 24) retrieved 3 June 2024 from <https://techxplore.com/news/2020-11-ai-tool-lengthy-papers-sentence.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.