Software provides a clear overview in long documents

July 25 2014

Veröffentlichung nur unter Nennung des Copyrights
Any publication must include a copyright notice

Copyright: Bayerische Staatsbibliothek / H.-R. Schulz

In the future, a software will help users better analyze long texts such as the documents for calls for bids, which are often more than one thousand pages long. Experts at Siemens' global research unit Corporate Technology have developed a search function that enables users to simultaneously look for key words and sections of text in all of the documents of a call for bids, for example, without having to actually open any of the files. This makes the search very fast so that it

only takes a few milliseconds before users can read the search results in the documents. The picture shows an employee of the bavarian state library working in its digitalization center at a scanning robot. Credit: Bayerische Staatsbibliothek / H.-R. Schulz

In the future, a software will help users better analyze long texts such as the documents for calls for bids, which are often more than one thousand pages long. Experts at Siemens' global research unit Corporate Technology have developed a search function that enables users to simultaneously look for key words and sections of text in all of the documents of a call for bids, for example, without having to actually open any of the files. This makes the search very fast so that it only takes a few milliseconds before users can read the search results in the documents. The experts also developed a component that checks to see how requirements have changed compared to previous versions of a specific text. As reported in the current issue of "Pictures of the Future" magazine, the ultimate goal is to create a semantic software that recognizes interrelationships in order to find relevant information.

Corporate Technology originally developed the software as part of a feasibility study regarding the digitization of all land registers in Germany. A system was required that could record automated information regarding owners, property sizes, outstanding mortgages, and other matters from the land registers of the past 50 years (around 500 million pages of PDF files). The software had to be able to extract the required information with the help of the respective document structure. The software also had to be able to handle scans of poorly copied typewritten pages or repeatedly corrected documents.

To develop the software for calls for bids in industry, the researchers at CT are cooperating closely with colleagues from the corresponding

Siemens businesses. The researchers are using this as a basis for developing characteristic search algorithms that enable users to find all of the information that a document contains about certain topics such as safety or pollution control.

Because calls for bids are repeatedly adjusted during a project, the software then identifies and displays any changes compared to previous versions of the document. In the third step, the software looks for analogies to previous, similar calls for bids so that users can see how certain requirements were evaluated in the past. The automatic semantic evaluation of large documents for a bid saves time, prevents mistakes, and makes it easier for users to integrate and analyze changes that were made at short notice.

Provided by Siemens

Citation: Software provides a clear overview in long documents (2014, July 25) retrieved 16 April 2024 from https://phys.org/news/2014-07-software-overview-documents.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.