Smarter searching in archives using newly developed interface

September 3 2012

Large quantities of data are flowing into archives each day: Newspapers and books are being digitised, whereas video material is being supplied directly in digital format. Search engine technology is therefore growing in importance. All of this digitised material provides a wealth of information for researchers in the humanities and social sciences, but can they also find what they are looking for amongst these so-called 'big data'?

According to Marc Bron, PhD student at the Intelligent Systems Lab Amsterdam (ISLA) at the University of Amsterdam, that depends on various factors. For certain material, researchers know that it is in the archive and which search terms they should use to retrieve it. However, in the majority of cases researchers come to the archive with a research question and they must first search for suitable material and explore the content of the archive.

Finding relevant material

One important difficulty in finding relevant material lies in the formulation of the search question that can be entered into the search engine. The search terms used by researchers can differ from the terminology archivists use to describe the material, even though they both mean more or less the same thing. For example, a researcher might enter the term 'migrant', whereas an archivist has used the term 'foreigner'. The second problem arises if material is found. Researchers cannot establish whether or not they have collected all of the relevant material or if other interesting things can still be found that they are not

yet aware of.

Explorative interface provides a solution

In order to tackle these problems, Bron has developed an explorative interface together with colleagues at ISLA, the Centre for Television in Transition of Utrecht University and the Netherlands Institute for Sound and Vision. This interface is called MeRDES, an acronym for Media Researchers' Data Exploration Suite. It can be used to compare the outcomes of different search queries in rich archives, such as those of the Netherlands Institute for Sound and Vision.

Researchers can visualise the number of programmes that are relevant for each of the search queries in order to form an impression about how much information is available for different aspects of a subject. For example, using this approach the growing use of the term 'migrant' in archive material can be compared with the use of the term 'foreigner'. The amount of material available for a subject and how this compares to other subjects can exert a considerable influence on the approach used for the research and the questions that can ultimately be answered.

Marc Bron and postdoc Jasmijn Van Gorp (Utrecht University) tested the interface by carrying out a user study with 40 media scientists. Bron presented the outcomes of their research at the International conference of the Special Interest Group on Information Retrieval (SIGIR) that was held from 12 to 16 August in Portland (Oregon, United States). A demo of the interface is available at: zookma.science.uva.nl/merdesdemo.

Provided by Netherlands Organisation for Scientific Research (NWO)

Citation: Smarter searching in archives using newly developed interface (2012, September 3) retrieved 25 April 2024 from

https://phys.org/news/2012-09-smarter-archives-newly-interface.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.