

A Google for handwriting

April 28 2015, by Josefin Svensson



Per Cullhed, Head of the Conservation Unit at Uppsala University library.
Credit: Mikael Wallerstedt

To be able to use computers to analyse and search handwritten texts would revolutionise research in the humanities. And the technology to digitise printed books and make them searchable already exists.

Uppsala University library has recently launched a digital platform—Alvin—where digitised works from cultural heritage

collections are now being collected into a single database. With just a few clicks it will be possible to search collections, opening up new possibilities for researchers and other interested parties.

'The works are searchable, for example via Google, which means you can go back over historical materials and find new angles. The texts don't need to be consulted on site either, which provides greater accessibility', says Per Cullhed, development strategist at Uppsala University library.

When the university library digitises printed books from heritage collections, it uses software that converts the pages to digital text, known as Optical Character Recognition (OCR). The software interprets the printed information and makes it searchable. With handwriting, HTR technology - handwritten text recognition - is used instead. It is the development of this technology which is creating something of a race among researchers worldwide.

'You want to be the first to find a program that works. If someone today had an algorithm to carry out large-scale digital searches of things like the collection of manuscripts in the Vatican Library, it would be worth a fortune. Whilst the market value is enormous, so is the scale of the task', says Anders Brun, project manager at the Department of Information Technology.

In the interdisciplinary research project 'From Quill to Bytes', Anders Brun and his colleagues are trying to develop a method that makes it possible to analyse and search large amounts of handwritten texts. The project involves basic research, which in the longer term should result in finished software.

'We usually call it a Google for handwriting; a way of quickly finding what you are looking for even though the amount of information is enormous', he says.

The project started in January 2013 and will run for about five years. Financing consists mainly of a framework grant from the Research Council of 13.7 million SEK.

Frederick Wahlberg, PhD student at the Department of Information Technology, is currently working on medieval manuscripts in Old Swedish in collaboration with Mats Dahllöf, researcher in linguistics and philology, and Lars Mårtensson, associate professor at the Department of Scandinavian Languages. Later in the project, they will be looking at the more recent Waller Collection, which is in the university library.

'The texts are very difficult to read and it is imperative to collaborate across disciplinary boundaries if we are to succeed in this', says Fredrik Wahlberg.

The core of the work is all about text decoding, achieving a method via which the computer tries to interpret the digital image of the text. The researchers are trying to avoid text interpretation because handwritten text can look very different depending on who was holding the pen. Instead, they want to teach the computer to interpret the material.

'Using expert knowledge, we try to give the computer the right answer for a small portion of the material and then automate this', says Fredrik Wahlberg.

The experts' knowledge of what is interesting and how various writers differ help them move forward in their work.

'The computer can help us, but it can't solve all our problems. There still needs to be expert knowledge in order to interpret the material and make corrections', says Anders Brun.

However, for researchers in the humanities, the opportunity to make

manuscripts searchable on a large scale would revolutionise their work and create all kinds of new possibilities.

'This kind of software is a bit of a Holy Grail for researchers who want to break new digital ground in areas such as history, religious studies and linguistics. It would mean so much to research', says Anders Brun.

Provided by Uppsala University

Citation: A Google for handwriting (2015, April 28) retrieved 7 February 2023 from <https://phys.org/news/2015-04-google.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.