

Digitization of copyright protected newspapers in Norway

Roger Jøsevoid - National Library of Norway

The background

In 2004 The National Library revised its digitization strategy. This was followed up with a decision in 2006 to digitize the whole collection. Because newspapers are a much sought-after source material, a plan for a newspaper digitization program was launched in 2007. Aftenposten, the largest and most influential Norwegian newspaper, was contacted for a possible cooperation on digitization of their historical archives and open this for use by a wider audience than those, who visit The National Library. The result was a partnership-agreement in three parts.

- Based on a 50/50 cost sharing agreement between The National Library and the Aftenposten the entire historical newspaper archive should be digitized. Aftenposten and The National Library will subsequently have their own digital archival copy.
- Aftenposten would for the future deposit the printing file in PDF format in addition to the traditional printed newspapers.
- The digitized historical newspaper archive will be made text searchable through OCR and should be freely accessible in every Norwegian library. These are all the newspapers older than a week.

The National Library's choice of Aftenposten to be the first newspaper was to make a pilot-agreement with the largest and most complex newspaper. It was meant to set a standard that other newspapers would easily follow.

In 2009 the digitization of the historic newspaper archive was completed and Aftenposten and The National Library were left with a complete historical newspapers archive each. For Aftenposten the motivation for cooperation was twofold. They wanted to streamline a labour intensive internal archive service. Additionally they wanted to offer their subscribers access to the historical archive as an additional subscription service.

For National Library the motivation was in line with the mandate of making cultural heritage material as simple as possible accessible to largest possible group of users. In this case through all Norwegian libraries. In Norway there is a public library in every municipality, many of them with affiliates. In addition comes 53 university and college libraries and 140 special libraries. As legal deposit institution The National Library has the most complete Norwegian newspaper collection and it is therefore natural to assume the task of building a complete digital historical newspaper archive.

The digitization

The main strategy is to digitize from newspaper microfilm copies. Only where the microfilm does not exist or do not have a good enough quality the original print-version will be digitized. As the program to digitize the complete historical newspaper legacy will happen over time and will vary in annual volume in line with agreements with newspapers, The National Library decided to out-source the digitization rather than setting up an in-house digitization operation. This also includes OCR of the digitized text, proofreading and tagging. Digitization of print editions is an in-house operation.

The functionality

Since the digitized newspapers are OCR processed it is possible to do single word search and search for phrases. As The National Library also has the role of being The Norwegian Language Bank the agreements with the papers states that the full text of the digitized newspapers will be included as text corpora thus covered by services available through The Language Bank. As of today The National

Library has launched its own N-gram search corresponding the Google N-gram search, and new tools for text mining is under development.

The legal framework

The agreements signed with the papers assume that the newspapers are responsible for ensuring that adequate rights clearance for providing access to the archives in Norwegian libraries are secured.

The size of the digital archive

Today the historic newspaper archive contains almost 400 different newspaper titles (not corrected for some newspapers having changed names throughout history). In total there are around 970,000 digitized newspapers with 22 million newspaper pages. Efforts to make agreements have taken place in a period where printed newspapers are challenged as news media and advertising media and where the traditional printed newspaper is put to the test by digital media. Despite the positive reception of requests for cooperation some newspapers for these reasons have not been able to use money on digitization of its historical archives. Nevertheless many of the major national and regional newspapers have signed agreements.

The usage

End of year 2015 it is possible to access the historic newspaper collection at 470 Norwegian libraries. Counting of use of the service is based on logging of IP addresses accessing the collection. As libraries typically have one single IP address the number of personal users is thus unknown. The most precise indication of use at the moment is the number of pages viewed on average per month. This has increased from 25,000 in Average in 2012 to 225,000 on average in 2015. A user survey conducted in 2013 shows that while 80% of those visiting nb.no reports access to digitized Norwegian books as the main usage (full reading access to all books published in Norway until the year 2001), the corresponding figure for newspapers is only 11%. However, the historical newspapers collection is the main service among those accessing nb.no from their local library.

In line with the increasing volume of titles and pages, sorting capabilities and more generally the all-over user interface becomes increasingly important. In September 2015, a project to renew and improve the National Library's digital library services was launched. In this project, all media accessed through the service are subjected to testing and re-design in order to increase usability and accessibility.