ABBYYY Helps Genealogy Society Increase Access to and Awareness of Historical Records

**FineReader reduces workload and speeds up project timelines for one Norwegian Organisation**

DIS-Hordaland, located in Bergen, Norway, is a section of the Genealogy Society of Norway (DIS-Norge) and specializes in providing databases for genealogists around the world focusing on people who lived in Hordaland. Their goal is not only to supply efficient databases for professional and private genealogists, but also to raise the awareness for genealogy amongst the Norwegian people.

**The challenge**

Free access to a database containing personal information has a long tradition in Norway. To populate these databases DIS-Norge volunteers have been manually transcribing parish records, censuses, and other public material for years. DIS-Hordaland was recently tasked with creating a database of persons in Bergen who were at least 21 years old at the date of the local elections in 1934. The database would need to contain information, such as name, address and date of birth of any citizen who met this criterion.

Because most of the documents containing the data were typewritten or printed, DIS-Hordaland determined that using text recognition (optical character recognition) software, to convert the documents to editable text would save the volunteer staff valuable time and resources. As Tore J. Mehl, Executive Manager of DIS-Hordaland explains, “what if these could be transcribed using modern tools such as an OCR program? We could then publish several sets of printed source documents with much less effort and in a shorter amount of time than doing it by hand.”

The organisation had a few requirements: first, the software should successfully digitise complex layouts, such as colour-scanned typewriter pages in A4-format with varying font types. It was also important that the program could recognise tables as this is the form in which much of the information appears in source documents. Finally, any software would also have to be able to produce editable spreadsheets that could be exported as CSV files and imported directly into a database.

![Original Document (Paper)](image1)

![OCR Results](image2)
The solution

DIS-Hordaland researched and tested a number of OCR-programs. After careful consideration and seeing it listed in many rankings of the best OCR programs, Tore J. Mehl, purchased a copy of ABBYY FineReader Professional Edition in order to test its features. Further investigation of the capabilities of the software resulted in DIS-Hordaland choosing ABBYY FineReader as a best-fit solution for their needs.

“The final reason that convinced us that ABBYY FineReader was the right product was its reliability in converting tables, something critical for us because of the common occurrence of tabular information in our sources. No other tool had the table processing feature as found in ABBYY,” affirms Tore J. Mehl. “Another great benefit was the software’s ease of use. We needed a tool that required very little training for our staff before they could get to work.”

The volunteers involved in the project went through a brief introduction course to learn how to use ABBYY FineReader and to understand other important aspects of the project. They were more than happy with the results: the advanced recognition capabilities of ABBYY FineReader provided full access to information locked on paper and images and virtually eliminated time-consuming retyping and reformatting. The process for the volunteers is quite simple: paper documents are scanned and fed into ABBYY FineReader. The images are converted to text which DIS-Hordaland chooses to have in Microsoft® Excel® format. The spreadsheet is then exported to a database, which enables it to appear on a web site, publically available for a world-wide audience of researchers and enthusiasts.

The outcome

With ABBYY FineReader DIS-Hordaland saved a considerable amount of time and resources. This was particularly valuable for the organisation as work for all projects is performed by volunteers outside of their ‘normal’ jobs. The conversion of 10 pages encompassing OCR-processing of typewritten text with ABBYY FineReader, additional proof-reading and final processing in Excel took about 1 hour 20 minutes. Under normal conditions, a similar process would have taken up to 4 hours of error-prone manual re-typing. DIS-Hordaland estimates that they saved roughly 400 hours of manual work in total for this project.

“The database could only be created so quickly and efficiently with the help of ABBYY FineReader. Instead of 9 to 12 months of manual work, with ABBYY FineReader we completed the project in only 3 months,” asserts Tore J. Mehl. “The published database has become a real success. In the first two months of the web site being live there have been more than 44000 searches in the database. This is an average of approximately 370 per day, and the interest in our database is only growing.”