

Estonian National Library runs ABBYY Recognition Server for gothic text OCR

Thousands of unique hardcopy materials, some dated early 19th century, are accessed in searchable formats via web

In recent years Estonia has been confidently establishing itself as a strong IT power. It was the first country in the world to run governmental e-Elections. Its tax paying system is largely internet-based so that over 90% of Estonians submit their tax declarations online. Also, there is e-Banking, where 98% of all transactions are carried out online. As for the public sector, Estonia has e-Schools and e-Police. Overall, Estonians like to think of their country as an e-Country, and they have all grounds to prove it.

All of the above is a strong indicator of the nation's commitment to deploying IT technologies for optimizing and enhancing processes in every aspect of life.

One of the beneficiaries of Estonia's e-movement is Estonian National Library (NLE). Established in 1918, it offers thousands of materials for public access. The Library launched electronic information services in 1993. Since then, the amount of e-services on its website has increased manifold, so the primary task for NLE during past few years has been to create and increase online access to the services to meet growing requirements of the public. For instance, in 2006 NLE joined the European Commission's Books2eBooks project which involves developing a new online service of ordering digital books. Under this project, an e-book is produced on request and is delivered to customer in PDF with a full-text search enabled.



About Nekstom OÜ

Established in 1994, it is one of ABBYY's oldest resellers. Nekstom OÜ is also one of the pioneers in Estonia in localizing software. For more information please visit www.nekstom.ee



About Estonian National Library

The biggest library in Estonia, founded in 1918. It offers thousands of materials for public access. For more information, please visit www.nlib.ee

Estonian National Library and thousands of its unique assets



A tremendous project was launched by NLE in 2006. The Library began to register and store Estonian publications in the digital archive called DIGAR. Printed materials – newspapers, magazines and books, accumulated during 80+ years since the Library opened, make up the largest part of its assets. Due to wear and tear, the Library stopped lending newspapers to public. Because of their irreplaceability, it was also important to retain the materials in their original form. Microfilm came in helpful for a while, but became obsolete due to its inefficiency and inconvenience: no text search, no indexing, and it gets scratched.

The question arose: what to do with thousands of old printed materials – 600,000 pages of newspapers and 35,000 pages of magazines, with publication dates as far back as 1821? How to preserve them and yet let the public read them? Given the delicate condition of some materials, their uniqueness and their large volumes, an OCR solution had to be found that was both trustworthy and powerful. However, the above-mentioned conditions weren't all.

21st century ABBYY technology and 19th century gothic typography

A very significant fact needs to be mentioned here: 19th century Estonian alphabet. Not every OCR program has the power to recognize gothic text. ABBYY Recognition Server has. The acclaimed OCR technology was proposed by ABBYY's Estonian partner, Nekstom OÜ. Recognition Server's main strengths are its fully-automated performance, unprecedented recognition accuracy, unattended server-based processing, unmatched scalability and flexible integration tools. On top of that, it saves processed files in a format most suitable for archiving or preservation, such as PDF, PDF/A, TIFF or RTF.

Estonian alphabet of 200 years ago had its particularities, such as unique letters. When first books were printed in Estonian language, there were no typography facilities located in Estonia. Books were printed in Sweden, Finland, Germany or other places that used whatever typography settings they could at that time, and this is why the same letter could look different and some different letters looked very similar. For example, it was very difficult to tell difference between gothic Ä or Ö. To address this, especially for NLE, Recognition Server was delivered with customized gothic OCR, based on ABBYY FineReader XIX Engine OCR. Initially, after testing ABBYY FineReader XIX on old texts in Estonian, it became evident that there was an additional need for enhancing the OCR for Fraktur letters in Estonian language. Considerable work was done for finding out and specifying in detail the missing Estonian characters. For that purpose, different publications, printed in different printing houses from 1800-1940, were compared. The examples were scanned and delivered to Nekstom OÜ. The enhancement of ABBYY FineReader XIX OCR Engine with additional Estonian characters gave a new quality and better recognition level for old Estonian texts in books, journals and newspapers.

With ABBYY Recognition Server up and running, this is what the process looks like at NLE now:

- Paper-based materials are scanned:
 - On bigger stations, Hybrid Device OK 300 Hybrid Colour scanner is set up. This scanner allows to do microfilms and image files at once.
 - Konica Minolta PS-7000 scanners are used on smaller stations.
- OCR and verification are performed by ABBYY Recognition Server and ABBYY FineReader
- A searchable PDF is created and archived into DIGAR
- DIGAR is open-accessed via NLE's website, <http://digar.nlib.ee>

The customer feedback is fantastic: *"Before the project we had only image archive, but now, thanks to ABBYY, it is turned into a flexible and convenient tool, and the materials are easily accessible to historians, students or anyone who has a wish or need."*

© 2008 ABBYY.
ABBYY, ABBYY Logo and FormReader are either registered trademarks or trademarks of ABBYY Software Ltd.
All other trade marks are property of their respective owners.
ABBYY LLC.: 129301, Moscow, Russia, P.O. Box #54, tel.: +7 (495) 783 3700, fax: +7 (495) 783 2663, office@abbyy.com; www.abbyy.com.
Part# 3157e