ATAPY Media Service Operations: Helping Preserve Swedish Cultural Heritage

ATAPY's track record in Scandinavian countries includes such projects as digitization of a large collection of books for the Royal Danish Library, creating an archive of XVIII-century Northern European prints for a UNESCO educational project and magazine backlog conversion for a Danish musical publishing house. As a result of this work, ATAPY was recently entrusted with several new Scandinavian projects primarily focused in Sweden.

Building an archive of Old Swedish plays for Riksteatern Sweden

The Swedish National Touring Theatre (Riksteatern Sweden) made decided to convert its collection of Old Swedish plays into a digital format.

According to the project requirements, texts were to be converted to Microsoft® Word with the original page design preserved as much as possible. Due to the material's age and layout specifics, the job required an expert knowledge of OCR technology in addition to a heavy manual formatting effort.

In this project, ATAPY processed more than 12,000 pages in Old Swedish. Double verification was used in almost 50% of the material to ensure high recognition accuracy and excellent text searchability. All the digitized material is now available online on one of Rikrteatern's web sites in Microsoft® Word and PDF formats.

Riksteatern is the name of the popular "National Touring Theater"/"National Theater Company" in Sweden. Established in 1933 with the goal to promote and produce quality theater throughout Sweden, Riksteatern is now the largest touring theater company in Sweden. It is financed and owned by 240 local Swedish economic associations.

Digitization of Old Prints Collection for Gothenburg University

ATAPY converted a series of old Swedish printed sources dating back to XVIII-XIX centuries into text format for the Gothenburg University Library.

This ongoing project, comprised of several phases, currently exceeds 75,000 pages and approx. 65% of the material was subject to full verification. Material yielding better OCR results, underwent partial verification (uncertainly recognized symbols only). This approach made it possible to provide considerable cost savings. The next step, ATAPY conducted a manual markup of files for subsequent conversion to XML.
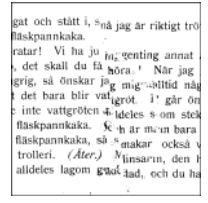
The University of Gothenburg is a major university in Northern Europe (approximately 37,000 students). The University's 40 Departments cover most scientific disciplines, making the University one of Sweden's most diversified higher education institutions.

In both projects, ATAPY faced challenges typical for working with old books:

- Low quality of the original page images (old weathered paper, pale print, etc.);
- Uneven lines, "jumping" print, varied spacing between letters, words, lines, etc.;
- Old Swedish words and grammar (ABBYY FineReader and sometimes even FineReader XIX dictionaries failed).

ATAPY overcame these challenges by using ABBYY FineReader XIX (a specialized package for processing prints in Old European languages and typefaces), smart segmentation of the material and applying qualified manual services when necessary. ATAPY's stratedy is to automate work whenever possible to minimize customer cost without sacrificing quality.

Creation of an electronic archive of Selma Lagerlöf's works for National Library of Sweden

Selma Lagerlöf (1858-1940) is one of Sweden's most prominent authors, winner of the Nobel Prize in Literature and a Swedish Academy Member. She left a literary legasy of more than 2,500 pages. In 2010, the National Library of Sweden launched a project to make this material available online. One of ATAPY's former customers recommended the company as an excellent service provider with affordable prices and hands-on experience with sources in Scandinavian languages.

The project involved the following phases:

- OCR of scanned images;
- Full verification of OCR results;
- XML markup of basic layout elements: titles, page numbers, separator elements, etc.

ATAPY processed the material with a limited deadline by using three media service operators. That year the Library was able to publish selected portions of its Lagerlöf Collection online tö commemorate the 150th anniversary of Selma Lagerlöf's birth.

The National Library of Sweden is a state agency with offices in Stockholm. The Library has collected virtually everything printed in Sweden or in Swedish since 1661. Currently the Library coordinates services and programs for all research libraries in Sweden and administers LIBRIS, the Swedish national library catalog system.

ATAPY Software
Engineernaya Street, 16
630090 Novosibirsk, Russia
Tel. +7 383 36 39 699
www.atapy.com ■ office@atapy.com