Digitization of Eighteenth Century Historical Works

Background
The client wanted to digitize (convert as images and XML data) more than 150,000 printed books of the 18th century available as microfilms, and publish the digital editions for enabling easy search, retrieval, display, and print.

Solution
This project involved creating a digital edition of more than 150,000 printed books having over 46 million pages within a time span of one year for web availability. To achieve this, HTC leveraged its Imaging and Document Management framework to deliver over 3.5 million pages per month. This framework enabled HTC to perform image capture, Optical Character Recognition (OCR), capture optional items (index fields, table of contents, etc.), and generate XML output of the text according to the client’s specifications.

Benefits
- Easy search and retrieval of the digitized collection
- Enabled a large number of users to easily access the digital books anytime and anywhere
- Ability to reuse existing resources limited by its format such as the materials stored on microfilm

Customer Speak
The quality of images and XML data that you have delivered has helped us create a highly sophisticated digital library. This has revolutionized the way research is done using old English books.”
- Director – Technology