APEX COVANTAGE

How JSTOR Successfully Digitized Thousands of Pages of Bi-Directional Hebrew & Latin Content

Challenge

JSTOR maintains a digital library of hundreds of thousands of archival and current scholarly publications for individuals and institutions to access online. When JSTOR looked to add over half a million pages of Hebrew journal content dating back to 1922, it presented several major technical challenges:

1. **Hebrew reads right-to-left and includes a completely different set of characters that English-based OCR software is not designed to process**

2. **Hebrew characters with diacritics (dots and dashes in, above, and under characters) oftentimes come out jumbled, incorrect, and out of order post-scanning**

3. **Incorporating metadata in both Hebrew and English required integration of two different bi-directional alphabets**

JSTOR turned to Apex, a partner of over 20 years with the proven ability to adapt to changing demands and requirements.

IZAAC

Apex integrated our IZAAC conversion software with a third-party software solution to recognize right-to-left text and diacritic markings and accurately convert them.

Solution

Apex responded by developing a meticulous 40-step workflow that resulted in a highly-accurate digital journal complete with metadata meeting JSTOR's detailed requirements. Apex integrated our IZAAC conversion software with a third-party software solution to recognize right-to-left text and diacritic markings and accurately convert them. Hebrew language experts based in Israel were recruited to distinguish subtle variations in text unable to be resolved by software alone, further increasing accuracy. IZAAC also enabled the creation and zoning of metadata and non-metadata elements in both Hebrew and English. Work was seamlessly executed across an international team based in the United States, India, and Israel.
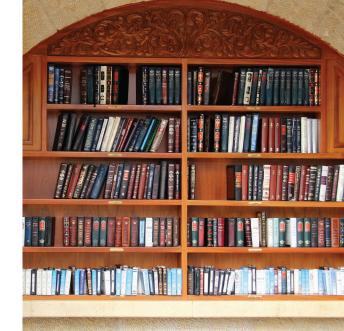
Results

Once the software was ready, JSTOR was able to complete its Hebrew journal project and make its collection of historic and current journals available online for scholarly pursuit. Partnering with Apex led to remarkable results:

- **500,000 journal pages were successfully converted**

- **All JSTOR quality standards were met or exceeded**

- **XML metadata was created for each article, issue and journal**

- **Reference XML metadata, OCR files, illustration files, and page image files were provided for each journal**

Apex continues to work with JSTOR, delivering over 1.5 million pages of content per year in XML, OCR text, and image formats.

500,000 journal pages were successfully converted

All JSTOR quality standards were met or exceeded

Once the software was ready, JSTOR was able to complete its Hebrew journal project and make its collection of historic and current journals available online for scholarly pursuit. Partnering with Apex led to remarkable results.

About Apex CoVantage

Apex is a leading supplier of data conversion, editorial, and content-enhancement services to publishers worldwide. By balancing leading technology and broad industry expertise, Apex delivers exceptional content and media solutions to its customers. Apex has been managing large-scale digitization projects for libraries, publishers, and media companies around the world for nearly 30 years.

+1.703.709.3000

content@apexcovantage.com

@apex_content

linkedin.com/company/apex-covantage

facebook.com/ApexCoVantage

apexcovantage.com

Apex CoVantage
200 Presidents Plaza
198 Van Buren Street
Herndon, VA USA 20170-5338

Learn more at **apexcovantage.com**