

Consulting project case study: XML data export from InDesign pages to a database

A publisher of legal texts needed a way to extract product name, description and price information from hundreds of formatted page layout files as structured XML files so that the information could be imported into their product database system. Quark was unable to provide them with a stable solution that worked well, so they began a company-wide shift to InDesign. After training their staff during the transition to InDesign, they asked me to help them automate the XML export process.

I began by writing an InDesign script that prompts them for a folder, and then all the QuarkXPress files in that folder are batch converted to InDesign format using the Q2ID plug-in from Markzware. The simplified and reduced the amount of staff time needed to convert each Quark document to InDesign.

Each InDesign file contained name, description and price information for a single product in three different formats. Each of these formats needed to be exported to a separate XML file. Paragraph and character styles had been used consistently throughout all the files to designate name, description, price, bold, italic, etc.

I wrote an InDesign script that does the following to each InDesign file in a designated folder:

1. Creates XML tags for each piece of information that match the tags expected by the client's product database system
2. Maps InDesign paragraph styles to the XML tags
3. Locates the frames that contain the three different formats of information that will be exported and assigns XML tags to these frames
4. Exports three XML files, one for each tagged frame, to a folder specified by the user.

Unique challenge of this project:

- Developing an algorithm to locate the frames containing the data to be exported. Each page contained many other frames that were to be ignored, and the frames with the data were not in the same position on each page.

To discuss *your* data publishing project, contact Keith Gilbert at kgilbert@gilbertconsulting.com or 651-633-7148.

