

# The Changing Face of Publishing

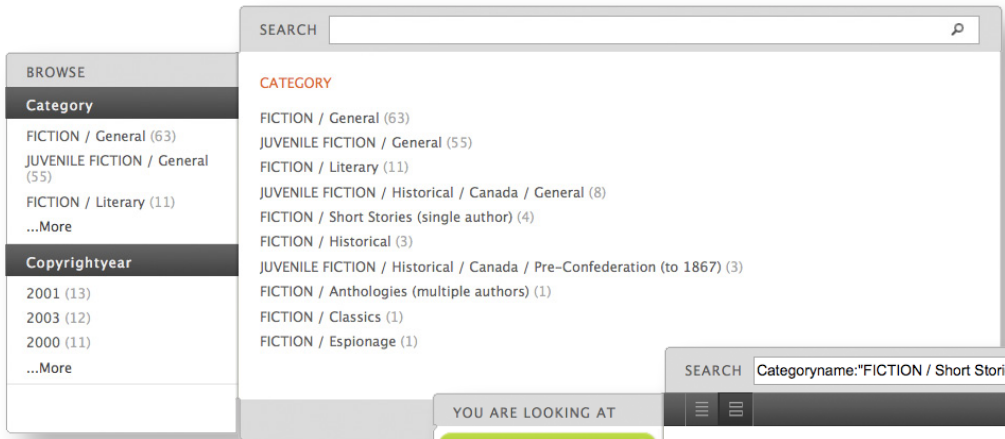
**W**hen it comes to change, few industries are undergoing the rapid, wholesale transformation that is unfolding in publishing right before our eyes. Traditional processes on the supply side are shifting, the demand for new formats is ever-evolving and, on a technical level, publishing takes on the characteristics of “data-processing.”



Bennett Cerf, the Publisher Archetype  
Database ‘Geek’, the New Face of Publishing

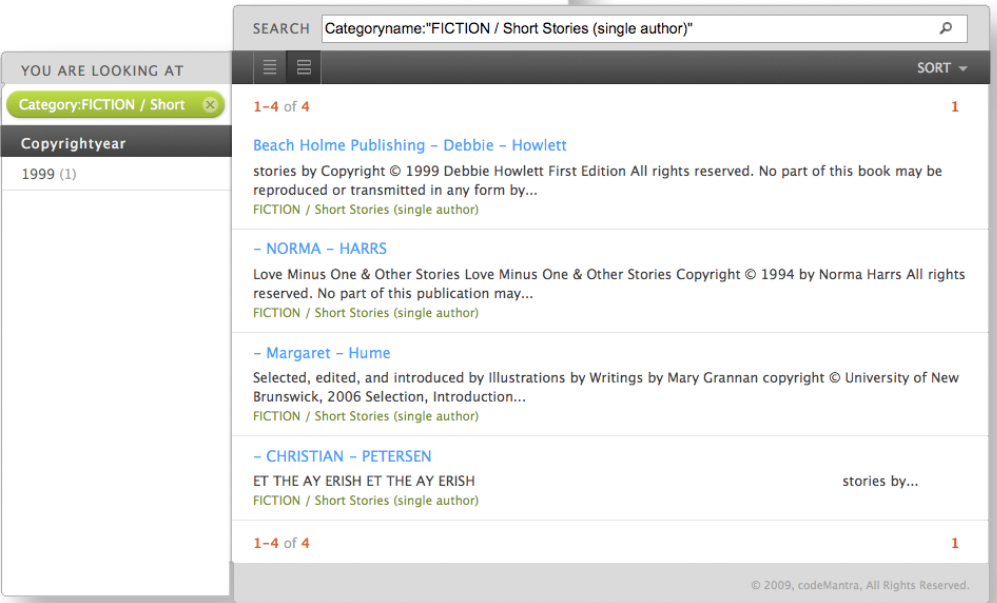
While the symbiotic relationship between a writer and editor continues to provide the fission that creates a great work, to a database engineer the ‘work,’ and the words that make it such, are mere objects to be organized and expressed in a structured format. Retrieving those objects with precise query and search accuracy is an entirely different skill set than the one required to assemble them for their power to inform, illuminate or entertain. Few publishers would see themselves as database engineers, but with the rapid evolution of digital markets and delivery systems, database technologies and the ‘geeks’ that manage them are emerging as key contributors to a successful publishing process.

As a publisher’s technology partner and provider of BPO services and software, codeMantra has reached out to Mark Logic, the leader in XML server technology, to help meet the growing demand for database publishing capabilities. Under the terms of its SaaS (Software as a Service) licensing agreement, codeMantra can now offer its collectionPoint 2.0 (cP 2.0) customers access to MarkLogic’s powerful XML Server. With this add-on application to cP 2.0, all or specified portions of a publisher’s content can be normalized to a standard XML schema. Unstructured content (such as PDFs or MS Office documents) can be readily converted to XML and indexed by MarkLogic XML Server. Like search engines, MarkLogic Server indexes words within documents. Unlike search engines, it also indexes XML elements, their structure, and the text within them. This means MarkLogic Server can quickly evaluate text queries, structural queries, and queries that combine both text and structural constraints – for example, search for diagram captions that mention “engine” in articles whose title contains “Airbus.”



**Figure 1** – MarkLogic enables users to browse content by category or copyright year, or perform detailed searches within the content itself.

**Figure 2** – Selecting Category name reveals the titles in that category, the author name, and a brief introduction into the text of the book. The introduction could be substituted with a brief description of the title that can be gathered from the title metadata.



The example above, demonstrates how Mark Logic enables users to browse, categorize and catalog content in several ways in order to make information more discoverable.

With MarkLogic Server, a publisher can provide rich navigation to external and internal users allowing them to systematically browse content or refine search results based on categories or tags. MarkLogic Server’s faceted navigation capabilities can be based on the XML structure of the content, entities, metadata, pre-existing taxonomies or taxonomies created with MarkLogic Server. Robust and refined search and retrieval across a publisher’s entire database of content is made possible using industry standard XQuery tools and technologies.

Furthermore, codeMantra’s application development team can build the functionality for a subscription database, allowing customers password protected access to all or select subsets of a publisher’s collection. Similarly, codeMantra engineers can develop tools and applications within cP 2.0 for the dynamic composition and delivery of content.

Working in tandem with Mark Logic, collectionPoint 2.0 hosts all versions and formats of a publisher’s digital assets. As always, cP 2.0 continues to function as the publisher’s digital warehouse, archiving and distribution platform of up-to-date and approved products, while MarkLogic Server provides a structured database of all publishing content along with powerful tools for data mining, search, retrieval, research and custom publishing. Neither the publisher’s internal users nor customers accessing Mark Logic enabled content will require proficiency in XML.