Data cleansing makes online shopping a convenient, pleasant experience

Background

E-commerce sales are rising rapidly. The U.S. Commerce Department reports that last year alone, online shoppers spent $165.4 billion, up 14.8% from 2009. The Commerce Department's latest statistics also show that e-commerce is growing much faster than brick-and-mortar retail sales. In 2010, 4.2% of total retail spending was online, compared to 3.9% in 2009.

This trend is expected to continue as more mobile apps are developed to make online shopping more convenient and portable. Adobe reported this month that 62% of consumers owning web-connected mobile devices are buying goods through these devices. By the end of 2011, U.S. mobile commerce sales are expected to reach $5.3 billion, an increase of 83% from last year, according to Barclays Capital.

These statistics are good news for Earley & Associates' Fortune 100 client. This retailer's website is consistently ranked among the most visited, and last year, the company was rated among the top retailers in the country.

The retailer wants to make shopping a convenient, pleasant experience online. The key in achieving this is to help its customers navigate thousands of items to find the products they need and want quickly and easily. From October 2010 through February 2011, Earley & Associates teamed with this retailer to help the company achieve this goal.

The Challenge

A well-structured taxonomy and consistency in product descriptions and values are critical in ensuring a rewarding shopping experience online. "We have to figure out a way to keep a clean taxonomy—valid, structured, and relevant— but not sacrifice the quality of the site with it," said a Product Taxonomy team member. "We need to be the ambassadors of information science—why we're doing it this way, and how proper data structures can be useful."

Every day, the retailer receives product descriptions from its suppliers for the tens of thousands of products it sells. These descriptions include the attributes of the product (its color, dimensions, targeted age, sustainability claims, etc.) and the attribute values (green, length 28 inches, ages 6-12 months, whether the product is recyclable, etc.).

Managing the quality of product data can be a daunting task. This data is often inconsistent across suppliers. For example, one supplier may describe

> *"We have to figure out a way to keep a clean taxonomy—valid, structured, and relevant—but not sacrifice the quality of the site with it. We need to be the ambassadors of information science—why we're doing it this way, and how proper data structures can be useful."*
>
> ~ Product Taxonomy team member

Benefits of the Earley & Associates solution:

- Taxonomy assessment guides future growth
- Streamlined product data that's easy to maintain
- Documentation for knowledge retention and training future employees

the color of a dress as "green," while another may abbreviate the color as "grn." Sometimes product attributes and the relevant attribute values for that product are missing. Issues with product data quality result in lost sales.

Large retailers face these challenges in managing the quality of their product data:

- Maintaining the existing product database, or repository.
- Handling new and updated product data from its suppliers.
- Leveraging tools and systems to maintain the product data efficiently and effectively.

Therefore, deciding how to manage such a vast repository of information is a bigger, overarching challenge. The retailer chose to integrate two systems and hire a consulting team with expertise in taxonomies and information management, Earley & Associates (E&A), to help them get started.

The Solution

Before the E&A team arrived, the retailer decided to integrate two systems: its current custom, in-house product data repository and Oracle PDQ. To achieve the initial data clean-up, the product data would be extracted from the repository, data clean-up completed in PDQ, and the data updates uploaded to the repository. The updates would overwrite the existing data in the repository. When new

products are added or suppliers change products, the process for cleaning up this data would be the same.

In September 2010, Oracle consultants arrived at the retailer's headquarters to install PDQ and configure the system. The E&A team arrived one month later. Oracle trained the team on how to use PDQ to cleanse the retailer's data.

Next, the E&A team conducted an assessment of the retailer's business processes used in maintaining the company's product taxonomy, the product attributes, and the attribute values. The team also evaluated the existing taxonomy's structure and made recommendations on streamlining it so that products would be easier to find on the retailer's website. Then the E&A team worked with the retailer's Product Taxonomy team to implement these recommendations.

Once the business processes were in place and the taxonomy structure and rules agreed upon, the E&A team began cleansing the product data in PDQ. Team members applied rules to organize the product categories, to ensure consistency in product descriptions, and to set limits on product dimensions, such as furniture and clothing sizes. This work placed the data in a state of *maintainable maturity*, meaning that the retailer's Product Taxonomy team could update and maintain the data easily and effectively after the integration project was completed.

At the end of the project, the E&A team conducted knowledge transfer by training and mentoring the retailer's Product Taxonomy team members. The E&A team also provided documentation that covered all of the methods and best practices that it established for cleansing and maintaining the data.

The Results

The retailer's Product Taxonomy team is pleased with the results. The E&A team helped the retailer seamlessly re-engineer an existing internal business group into one that now supports new business tasks, processes, and skills.

Also, the retailer now has data that can be maintained easily and efficiently, along with documentation that guides them through the process. Training new team members will be easier with the documentation the E&A team provided. The recommendations the E&A team made in the taxonomy assessment will help the retailer in its strategic business decisions to expand its taxonomy.

The integration project also lays the foundation for the retailer to develop new business processes, such as inputting suppliers' product data directly into PDQ for verifying its quality and cleansing. This future direction will replace some manually intensive processes to move supplier data into the retailer systems.

References

Deatsch, Katie. "62% of consumers with web-connected mobile devices are buying goods via mobile." *Internet Retailer*, February 10, 2011. Retrieved February 22, 2011, from http://www.internetretailer.com/2011/02/10/62-consumers-have-purchased-goods-mobile-device.

Enright, Allison. "E-commerce sales rise 14.8% in 2010." *Internet Retailer*, February 17, 2011. Retrieved February 22, 2011, from http://www.internetretailer.com/2011/02/17/e-commerce-sales-rise-148-2010.