ORACLE ENTERPRISE DATA QUALITY
FOR ORACLE ENDECA

Oracle Enterprise Data Quality (EDQ) is a complete data quality platform including profiling, standardization, match, merge, governance and case management capabilities. It includes best-of-breed features for both customer (party) and (product item) data – making it both broader and deeper than most competing data quality systems. Oracle Endeca is a best-in-class search system that leverages structured information to enhance the search experience. EDQ can efficiently standardize structured and semistructured data and deliver it to Endeca to optimize the search experience.

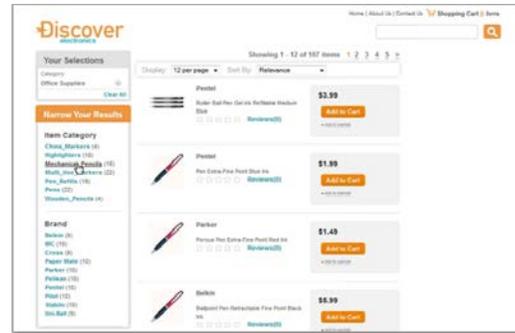## Oracle Endeca – Guided Navigation is Key to Effective Search

Before Endeca, search was largely based on matching characters in the query to characters in whatever information was being searched. This simple technique works well for small volumes of reasonably well standardized data, but Endeca took this to the next level by allowing the search results to be further refined by sub-selecting various attributes of the initial search – a technique they called 'Guided Navigation'. I.e. a search for 'motor' can be refined by power, size, type, etc. or a search for 'shirt' can be refined by color, size, pattern, material, etc.

This technique was revolutionary, and has now become an accepted pre-requisite for any world-class search, especially in retail. But it exposed a hidden problem – the product data to support this has to be much more standardized and much better organized to be navigated in this way.

Here are some insights from a survey of Endeca customers:

Insight #1 – Guided Navigation is key, but data is a problem – Most Endeca users (88%) view Guided Navigation as 'very important' to their use of the system. A majority (71%) of respondents say that having more dimensions would improve their search experience but more than three-quarters (76%) say that they are prevented from deploying more search dimensions due to the difficulty of structuring and standardizing their product data.

Insight #2 – Changing data structures make integration harder – The nature of product data is that it is always evolving, as are the search requirements. Endeca is quite capable of handling the changing product categories and dimensions, but the systems that feed Endeca need to be kept in-sync and that can lead to a high cost of maintenance in the integration area, which again limits the use of the kind of data that delivers an optimal search experience. 20% of customers have built a custom data integration application (usually at high expense), but few survey respondents are happy with the result.

KEY BENEFITS

- Enable the optimal Endeca search experience with optimized data
- Improve search effectiveness with more dimensions for navigation
- Improve search experience with fully standardized dimensions
- Reduce operational cost using automation to improve data
- Reduce integration costs through adaptive integration

KEY FEATURES

Oracle Enterprise Data Quality

- Broad data quality platform with best-of-breed capabilities for customer data and product data
- Integrated design promotes high usability for business users
- Modern architecture enables flexible deployment, including as callable web services

Oracle Enterprise Data Quality Connector for Endeca

- Includes full range of EDQ capabilities
- Pre-built templates to prepare data for use in Endeca
- Pre-built integration synchronizes metadata between systems to eliminate integration

Support for Endeca

- Populates dimension – for more ways to find an item
- Standardizes dimensions – for cleaner and more organized drop-down lists
- Standardizes ranges – to avoid manual mapping of items into ranges
- Category-based rules – allows different rule sets per category to eliminate 'over generalization' (applying a single set of rules across all categories)
- Allows setup of precedence rules

ORACLE

Navigation dimensions are used to refine Endeca searches

Product Data – Never 'Good Enough'

Product data is notorious for its lack of standards. How many attributes should you hold for a motor, or a capacitor, or a pair of shoes? What are the allowable values for those attributes, and what should you do if they are missing? There is no single answer to these questions, but they cannot be ignored or your products (items, SKUs, components, assemblies, etc.) will have different descriptions, and little standardized information with which to search and match them – and your Product Value Chain will be in chaos.

Most often in a Product Value Chain, your product information is supplied by someone else, a vendor, supplier, component engineer, salesman, warehouse manager or whomever else has been designated to enter the data. Unfortunately, because there are no universal product data standards, the data is created in quite different ways to meet many different needs beyond search, and it comes from many different sources. Consequently it is usually 'non-standard' to say the least. More commonly it is a mess and needs some major effort to make it consistent and 'fit for purpose' – and remember, this is not a 1-time effort as this data is constantly changing.

The need to enhance and standardize the data is especially acute if the product data is to be published on a website for browsing and search. This could be on a retailer's website, or it could be the website of a manufacturer or distributor, but the more data there is, the more critical it is that the data be 'searchable', which means it should be complete, consistent and well attributed.

Oracle Enterprise Data Quality for Product Data – a Unique Approach

Oracle Enterprise Data Quality is a broad data quality platform, but includes some capabilities specifically built to handle product data that are unavailable in other data quality systems. These capabilities are built on two core capabilities:

• **Semantic Recognition** – This is the ability to recognize information in context based on semantic information in the record. This allows relatively unstructured information to be recognized, extracted and standardized based on the context of the category. In the case of EDQ, the semantic model can learn from feedback captured during operation. This 'auto-learning' capability allows the creation and maintenance of a large amount of semantic rules in a very efficient way.

• **Category Specific Architecture** – It is rare that one has to deal with only one or two product categories. More typically product data comes in hundreds or thousands of categories (note that product categories are groupings of products that share similar vocabulary, attributes and rules. For example, laptop computers might be a category, while handbags might be another).

ORACLE®

As each category has different vocabulary, attributes and rules for parsing and extracting information, it is important that the system have a category architecture to keep the rules neatly organized. Without such an architecture, it would be impossible to build and maintain the necessary rules.

EDQ includes these patented fundamental capabilities and more, based on a long history of practical experience in dealing with the complexities of product data.

Oracle Enterprise Data Quality Connector for Endeca

The core capabilities of the EDQ Connector for Endeca are twofold:

- **Semantic Data Standardization** – Using semantic recognition, product records are classified, attributes (dimensions in Endeca) are extracted and standardized into valid values and/or ranges. Optionally, search keywords and standardized descriptions can also be generated. This is the data that enables Endeca to deliver a rich and engaging search experience.

- **Adaptive Integration** – The EDQ system manages the category based rules and schema for data standardization and is able to use this information (metadata) to load into Endeca and configure how it expects to receive data. In this way the complex and changing product data schemas are managed in one place but synchronized across both system thus eliminating a major integration and maintenance problem.
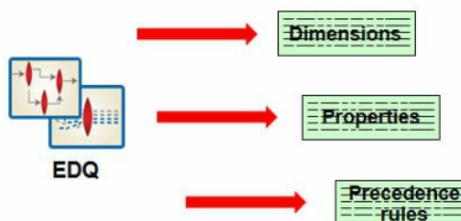
Enterprise Data Quality populates Dimension, Properties and Precedence rules in Endeca

These capabilities are designed for use by business users (who know the required data rules) rather than IT staff and taken together with the above capabilities allows a fast, efficient and highly responsive method of optimizing data for search. This allows Endeca to do what is does best – deliver a rich and engaging search experience.

Contact Us

For more information about Oracle Enterprise Data Quality products and the EDQ Connector for Endeca, visit oracle.com or call +1.800.ORACLE1 to speak to an Oracle representative.

Oracle is committed to developing practices and products that help protect the environment

ORACLE®