# **Semantic Search**

Semantic Search Requirements for a Large Industrial Supply Organization

Application must understand the intent and provide personalized search experience.

## **Customer**

Customer is a leading value-add industrial distributor offering products, services and solutions that enable its customers to achieve higher levels of growth, productivity, and profitability. Customer is one of the nation's largest direct marketers and distributors of a broad range of metalworking and maintenance, repair and operations (MRO) products to customers throughout North America.

Customer deals with 1.5 Million cataloged items, 3000 suppliers and 6 million non-cataloged items.

## **Background/Brief Description**

Major sales channel for the customer is the B2B ECommerce Website which allows their customers to browse and search the digitalized version of their Sales catalog and place orders online. The catalog constitutes more than 1.5 Million products across 5000 taxonomies which can have Level1-Level4 categories.

End users of the web site constitutes of customers from industrial domains ranging from freightliners, aerospace, automotive, manufacturing etc.

Current Requirement is to Create Semantic search solution to enhance the search experience.

## **Problems Faced**

* Customer context is not taken for search
* Many searches result in empty pages or no-item results page (like categories)
* Personalized recommendation is outsourced and expensive per click solution

## **Solution Expected**

* Create Semantic search solution to enhance the search experience.
* Create a search solution that combines the features of Semantic search and thesaurus
* Create a search application that’s able to understand and process content written in natural language.

## **Users**

* Procurement Managers, Shop Floor Associates, Technical experts working in Industrial Distribution Industries like Metal working, MRO.

## **Acceptance Criteria**

* Create Semantic search solution to enhance the search experience. The search solution must be able to understand the intent / meaning of the search query based on the demography of the customer as well as the available contexts like previous purchase history of the customer, his industry back ground, tools which he is interested in, search patterns of other customers in similar industry etc. For example, a simple search query ‘Battery’ could mean automotive OR rechargeable OR power tool batteries depending upon the context of the search.
* Create a search solution that combines the features of Semantic search and thesaurus, to be able to understand synonyms during a search operation. For example, the following pairs of words mean the same: Drinking fountains /Bubblers, Rubbish bin/Trash bin/Garbage bin. The search engine must be able to recognize the synonyms and bring up the results accordingly.
* Create a search application that’s able to understand and process content written in natural language. The application must be able to understand, tag/categorize such information and provide a great search experience for the user

Good to have: NLP, Voice, image and context-based search

* Search Solution should be able to visibly improve accuracy of search results based on keywords, user back ground and history. It need not be specific to industrial distribution domain, but the team should be able to demonstrate the efficacy over a sample data set.
* Search solution should be able to demonstrate thesaurus like capabilities over a set of keywords without hardcoding the rules
* Search Solution should be able to demonstrate Voice & Image based search (Not mandatory)

## **In-Scope And Out of Scope**

* State the benefits that can be achieved by the solution that is being provided (ROI, Time & Money Vs Returns so that it can be derived whether the solution is reasonable to implement)
* Non Catalog search is out of scope (Like search over Marketing Contents, Documents etc)

## **Assumptions**

* Solution comes with reasonable security in place protecting both customer and item data. Solution should not involve exposing entire Item data as a scrapable Data dump in Cloud or other areas (including website) so that competitors can download and reuse. Customer specific data shouldn’t be exposed to another customer.

## **References**

* [www.mscdirect.com](http://www.mscdirect.com)
* <https://www.mscdirect.com/browse/?searchterm=batteries&hdrsrh=true&rd=k> This leads to a category page
* <https://www.mscdirect.com/browse/?searchterm=Adapters&hdrsrh=true&rd=k> This leads to a Cross Category result page
* <https://www.mscdirect.com/browse/tn/?searchterm=bubblers&hdrsrh=true> This leads to search restults for drinking fountain, but the thesaurus is a hard coded rule.

## **Technology**

* Any technologies not involving AWS & associated feature set can be used.
* The Search Solution is hosted on Oracle Endeca front ended by java and JSP.
* Preferred stacks DB2, MongoDB, Elasticsearch, Java, J2EE (No Spring) etc.
* Also would be open to other technologies except AWS and associated features. Preferred not to use complex stacks which involves lot of maintenance.