INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 2)
Available online at: www.ijariit.com

Intelligent model to identify the duplicate products in catalog management

M. Kavitha
15tucs114@skct.edu.in
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

Akilandeswari E.
akilaelangovan29@gmail.com
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

P. Tamije Selvy
p.tamijeselvy@skct.edu.in
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

ABSTRACT

Nowadays electronic catalog management provide more information about the products to the customers. Catalog management in the e-commerce system supports the definition, storage, retrieval, and management of product information in the e-commerce process. In which electronic catalog contains a duplicate of the product so it reduces the customer's experiences and it should not provide much more relevant information about the products in the search recommendations. This paper presents the method to solve the duplicates of the product and increase the search recommendations with the help of the Relational Database Management System (RDMS) and (DBMS). So, it creates a systematic and impressive catalog that helps the customers to get all the product descriptions and images they need and reduces the duplicates to make a buying decision.

Keywords— *Electronic commerce, Catalog, Merchant, Eliminate duplicate*

1. INTRODUCTION

A catalog management system provides the means for defining, storing, and retrieving E-catalogs while supporting multiple views and to transparently support the various processes of an e-commerce application and it is very useful for the merchant to upload the products without the duplicate products. It increases customer experiences to view the products in a detailed manner. There are many critical issues in the design and implementation of a catalog management system include modeling & design, standardization, translation, publishing, indexing, and version management.

Online business is the business exchanges with the assistance of the web from anyplace at any time. The exciting environment of electronic commerce is a broad range of interaction processes among the various market participants; ordering the product, transport and delivery, invoice and payment, etc. A typical e-commerce transaction consists of suppliers and consumers to gather the information and exploring potential market partners for goods and services, exchanging with the offers, movement of goods, payments, and other transaction activities.

Electronic catalogs hold most of the information required in the stream of processes and the computerized nature of e-catalogs allows for opportunities to automate and streamline many of the processes involved. Consequently, e-catalogs form the basis of an e-commerce transaction and catalog (or product information) management is a function that is required in almost all e-commerce systems. The catalog management in the e-commerce website is the system used to increase the customer experience and helps the customer to know more about the products in detail.

The merchant will tend to view the products in terms of the materials and processes associated with uploading the details of the product with the help of the schemas. And, the merchant would be more interested in the product's size and weight. The difference in views can be clearly described the variety of product classification schemes that are in use nowadays. So here are many critical issues in the design and implementation of a catalog management system include modeling and design, standardization, translation, publishing, indexing, and version management.

© 2019, www.IJARIIT.com All Rights Reserved

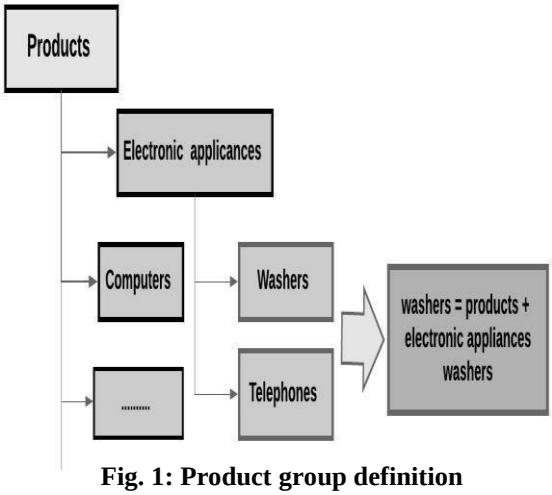

Fig. 1: Product group definition

2. DATABASE DESIGN

Normally, the catalog in e-commerce would need to be persisted in a database system to be useful. The challenge here is, it can't envision every one of the items it would have amid its lifetime. Even if it happens, it is a waste of storage space to create and populate the database tables for products that may come after years. This presents a challenge to design the data model as these items and products would need to be persisted in a database, typically using a relational database management system [3]. An E-catalog needs to store a wide variety of product categories, to cover all possible products of a business during its lifetime. Changing structure or recoding the E-catalog, as a software application, is expensive and may introduce new bugs. Thus, this should be avoided as much as possible. Different contexts or subject areas would need to store different types of products and items.

2.1 E-Commerce systems need a high availability E-catalog

Any software application accessing RDBMS or (DBMS) tables must have a priori information on its own (figure 1). Thus, the structure and data type stored in the relational database management system. Modifying one or more of these database tables, for example, to define a new product type, would usually require modification of code accessing these tables. Therefore, it is not possible to change the table structure or add new tables in the catalog database at application runtime. This must be done offline and the application code would need to be retrofitted and redeployed.

2.2 Product attributes data need to be stored in the atomic form:

To support accurate product searches and comparisons, these operations need to be done on atomic data values stored in columns. Storing all product information in one large text field would not enable extraction of this information when needed. For example, a customer wants to get a list of refrigerators between the sizes of 18c.f. and 22c.f, with best energy efficiency. This means the customers need to search on "Size" and "Energy consumption" attributes of the Refrigerator category. If this information were embedded in a description text field with other properties, it would be a difficult task to extract them for this search.

2.3 Organize the product types (categories) in a hierarchy of parent-child relationships (use the inheritance pattern):

Each parent would contain attributes that are common to all its children, i.e. a child would inherit the attributes of its parent. This schema is to represent the product or item categories in the catalog, to capture common attributes in the hierarchy. For example, a common attribute for all products in a store would be price, quantity, manufacturer, discounts offered, and product name. This schema would allow new categories to be defined with minimal effort, as they would inherit attributes from existing ones.

2.4 Allow the administrator system user to define a new category when needed:

Define the category name, its parent category (multiple categories for multi-inheritance), its attribute names and their types.

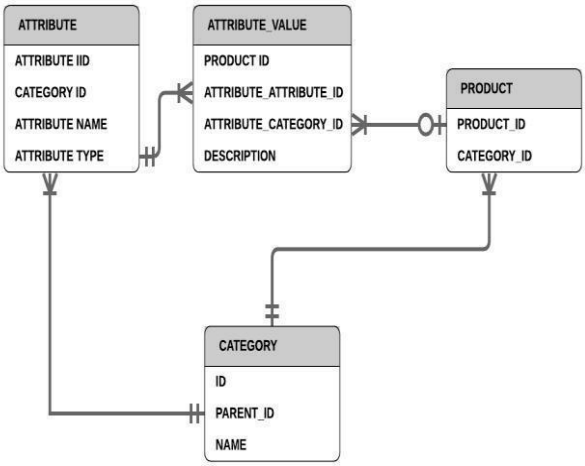

Fig. 2: Catalog data model

© 2019, www.IJARIIT.com All Rights Reserved

CATEGORY: The ID column stores unique id for this category and it is the primary key for the table.PARENT_ID contains the parent category id. This is also a foreign key referencing the ID column, as any parent category would be a category itself. This column would contain a NULL value for the category at the hierarchy top, as it has no parent. NAME contains the name of the category. ABSTRACT_IND is a flag to be set if this category should have no concrete products in it. This table is related to relation R1 to itself as a category may have multiple child categories.

ATTRIBUTE table: contains the attributes of a category. Each row represents an attribute. The ATTRIBUTE_ID is a unique id for the attribute, and it belongs to the category defined by CATEGORY_ID. ATT_NAME contains the name of the attribute. ATT_TYPE contains the attribute type. ATTRIBUTE_VALUE table: contains values of those attributes defined in the ATTRIBUTE table. This value is stored in the VALUE column. This table would store attribute values ATTRIBUTE_ATTRIBUTE_ID values associated with concrete products defined by PRODUCT_ID that belongs to a category identified by ATTRIBUTE_CATEGORY_ID.

PRODUCT table: This table associates the real instance of the product to a certain category. It contains the information about which product (PRODUCT_ID) is of which product type (CATEGORY_ID). Relation R/2 means that a category would have many attributes, and each attribute belongs to one category. Relation R/3 shows that an attribute may have many values stored, each value is associated with a different concrete product. Relation R/4 shows that a category may have zero, one or multiple products stored in the catalog of that category type. One variation of the implementation would be to define attributes common to all products as columns in the PRODUCT table, instead of defining and storing them as attributes in ATTRIBUTE and ATTRIBUT_VALUE tables.

3. PRODUCT CLASSIFICATION

The main benefits of using this catalog are to remove the duplicates of the product and maintained at the scale. Apart from this, millions of dollars are spent developing software that maintains information about products, buying the history of users for products, etc. But as the catalog size and number of suppliers keep growing the problem of maintaining the catalog accurately grows exponentially. The main misclassification of products can lead to poor customer experience (erroneous or irrelevant search results and recommendations). In this product classification, there are many challenges includes inaccurate tags impact product discovery, lack of real-time visibility into inventory and inability to analyze and react to demand. **CrowdANALYTIX** approach is used to optimize (eliminating the duplicates) the catalog management in the e-commerce system and reduces the speed, increases the performance as like (figure 2). The process involved the following key steps:

Main keys: Pre-processing, Transformation and Accuracy of the internal product catalog data as mentioned in (Figure 3).

Consistent data: Once the models were processed, ensure that the consistent and accurate data to maintaining the quality of the product.

There are three different ways to solve the product classifications in catalog. They are:

- Data entry crowdsourcing platforms
- Product APIs
- dataXTM

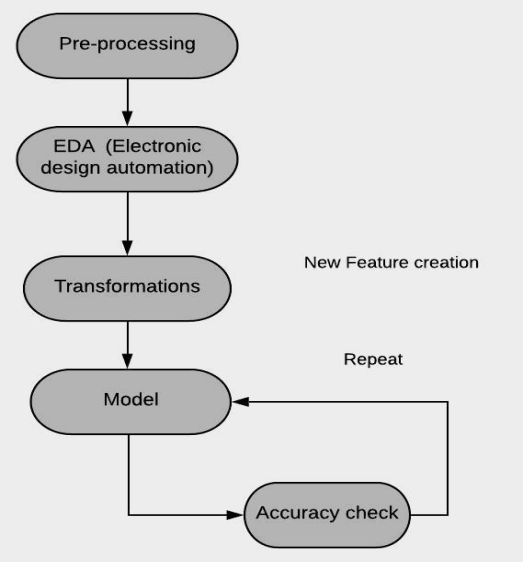

Fig. 3: Steps for predicting the accurate result

3.1 Approach 1

Data entry crowdsourcing platforms: Crowdsourcing platforms like Mechanical Turk provides access to a large number of products across the world. It gives a Quality of data (Since humans enter the values, data quality is reasonably good, however, human data entry errors do creep in and need to be watched out for). Its main advantage is, it gives a Speed of updates (Distributed Workforces can handle at most 40,000 SKUs in a month). This isn't always ideal since products on a catalog need to be on board every day and any delay in optimizing the catalog can result in lost sales due to the poor customer experience.

© 2019, www.IJARIIT.com All Rights Reserved

3.2 Approach 2

Product APIs: Platforms like Wiser, Index, Semantics3 and other web-based data platforms that claim to have pre-built data repositories that can be integrated with existing catalogs. Integration is very easy and it is cost efficient while using this approach.

3.3 Approach 3

Automated attribute prediction algorithms: Some sophisticated retailers build a bunch of NLP and ML algorithms to extract values of data attributes from the image, title and description of products. Since the image, title and description provided by suppliers are generally trustworthy, this method provides accurate attribute values if the models are built at a precision of 90% or more. Suitability for small retailers (There is a one-time effort in building models but post that the cost is very manageable)

3.4 Approach 4

dataXTM: uses ML algorithms to automatically update product catalog data. Thousands of models are built by a crowd of data scientists, one model corresponding to each product attribute. The models are then made accessible through an API for near real-time updates. It minimizing errors and reducing costs

4. IMPLEMENTATION

Now the results from the above product classification approach are going to use in the RDMS tables to retrieve and display in the web pages to the end users.

Create the table with the name ATTRIBUTE, ATTRIBUTE VALUE, PRODUCT, CATEGORY and having fields have discussed earlier in the database design module. Now the table along with the sample data looks like in Table 4.1, 4.2, 4.3, 4.4

Table 1: Attribute Table

ATTRIBUTE_ID	CATEGORY_ID	ATTRIBUTE_NAME	ATTRIBUTE_VALUE
1	101	BRAND	1
2	102	MODEL	2
3	103	COLOR	3

Table 2: Attribute value

PRODUCT_ID	ATTRIBUTE_ ATTRIBUTE _ID	ATTRIBUTE_CATEGORY_ID	VALUE
1	1	101	Samsung
	2		Galaxy J5
	3		Black
2	1	101	Samsung
	2		Galaxy J5
	3		White

Table 3: Product table

PRODUCT_ID	CATEGORY_ID
1	101
2	102
3	103

5. CONCLUSION

The proposed catalog in e-commerce presents a method of how to eliminate the duplicates in the products that are displayed to the customers and increases the search recommendations by using the **CrowdANALYTIX** approach in an efficient way. With the help of this catalog management lots of customers can buy their needs according to their wish and by reducing the duplicates in the products based upon the category reducing the performance issues.

6. REFERENCES

[1] Anant Jhingran, "Moving up the food chain: Supporting E-Commerce Applications on Databases", SIGMOD Record, vol. 29, no. 4.2.

[2] Kiryoong Kim, Dongkyu Kim, Jeuk Kim, Sangwook Park, Ighoon Lee, Sang-goo Lee, Jong-hoon Chun, "An Evaluation of Dynamic Electronic Catalog Models in Relational Database system. (RDBMS)

[3] Systems" in Managing E-Commerce and Mobile Computing technologies in IRM .908 publishing.

© 2019, www.IJARIIT.com All Rights Reserved

[4] Eduardo B. Fernandez, Yi Liu, and RouYi Pan, Patterns for Internet shops, in PLOP 2001.
[5] Dongkyu Kim, Sang-goo Lee, Jonghoon Chun, Sangwook Park, Jaeyoung Oh, Catalog Management in E-Commerce Systems, proc. of Computer Science and Technology (CST 2003).
[6] D. Kim, et al, Catalog integration for electronic commerce through category-hierarchy merging technique, Proc. 12th IEEE Workshop on Research Issues on Data Engineering, Engineering E-Commerce/E-Business Systems (RIDE-2EC.
[7] A. Jhingran, Moving up the food chain: Supporting e- commerce applications on databases, ***Technical Report, IBM Almaden Research Center***, 2000.
[8] S. Park, et al, MCI: Mass catalog indexing for searching large- scale catalog databases, Technical Report, CoreLogiX Inc., Seoul, Korea.
[9] A. Jhingran, Moving up the food chain: Supporting e-commerce applications on databases, Technical Report, IBM Almaden Research Center.
[10] T. Anders. Controlling of dynamic IT services based upon Service Level Agreements. Proceedings of the 2004 WTC/ICC symposium, Seoul, Korea.
[11] I. Vista. The Quality of Service Management Solutions Whitepaper.
[12] K. Kim, et al, An Experimental Evaluation of Dynamic Electronic Catalog Models In Relational Database Systems, Proc. of the Information Resources Management Association International Conf.
[13] Setrag Khoshafian, George P. Copeland, Thomas Jagodis, Haran Boral, and Patrick Valduriez. A query processing strategy for the decomposed storage model. In ICDE, pages 636-643. IEEE Computer Society.
[14] Min Wang, Yuan chi Chang, and Sriram Padmanabhan. Supporting efficient parametric search of e-commerce data: A loosely coupled solution. In EDBT, volume 2287, pages 409-426, Springer.

BIOGRAPHY

M. Kavitha
Student, Computer Science and Engineering
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

E. Akilandeswari
Student, Computer Science and Engineering
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

Dr. P. Tamije Selvy
Professor, Computer Science and Engineering
Sri Krishna College of Technology, Coimbatore, Tamil Nadu

© 2019, www.IJARIIT.com All Rights Reserved