Improve Customer Experience Based on Recommendation and Detection of a Pattern Change in Eating Habits

W. Velásquez*, *Member, IEEE*, Andres Munoz-Arcentales†, Andreina Montoya‡, M. Chalen§, *Member, IEEE*
Universidad Politécnica de Madrid, Avenida Complutense 30, 28040, Spain
Email: (wa.velasquez,andreina.montoyap‡)@alumnos.upm.es*
Escuela Superior Politécnica del Litoral, ESPOL, FIEC, P.O. Box 09-01-5863 Guayaquil, Ecuador
Email: (wavelasq, joanmuno†, mchalen§)@espol.edu.ec*

***Abstract*—This paper states a new approach for improving the consumer experience through the use of gaph database in order to extract the consumer eating habits and the pattern change from a large amount of anonimyzed data. For achieving that, a web system was developed for making recipes recommendations using recommender techniques, inference, graph database, multimedia visualization and SMS payment, for offering to the user an improve customer experience.**

***Keywords*-Recommender system; eating habits; graph databases; on-line payments; neo4j; inference; anonymized data.**

I. Introduction

Every day, people is producing many data at the social media on the internet that many companies uses to generate personalized marketing to persuade customers to buy their products. So, persons receive a lot of buying offers and recommendations from their social media friends but this data does not bring a fixed knowledge [1].

A recommender is a system that is responsible for providing users personalized and differentiated information about certain products and/or services that may be of interest based on the profile or searches previously consulted. As we know, these systems already dominate the market, especially in social networks, e.g., Facebook[1] and Twitter[2] using recommendation algorithms to know the affinities between users who may or may are in a kinship [2]. In this context, some applications have been developed using social networks, such as the proposal Kazienko et al. [3] where it is described a multidimensional social network into the social recommender system or the work of Xin Liu and Karl Aberer [4] with a contextual recommendation system in a social network. In addition, recommendation systems have penetrated internet shopping interactions through catalogs like Amazon [5], Ali-express, Alibaba, etc., based on previous customer searches or purchases of articles or services [6].

This project pretends to be more than a user's data recommendation system, it offers a service to help customers to get food recipes based on their purchases and presents information about how many calories they are consuming daily or per dish prepared to lead a healthy diet.

The following parts of this paper are divided into: Section II,'A description of the state of the art of technology used'. Section III, 'Presents the simulation scheme and configuration'. Section IV, 'Analysis of the different databases and their behaviors', and Section V, 'Conclusion of the results and data analyzed'.

[1] www.facebook.com
[2] www.twitter.com

II. State of the Art

In this section, a brief review of the concepts of recommender systems and a graph database is presented to provide a little background of this kind of technologies, which serve as part of the analysis for choosing the correct method to build a platform for improving the customer experience based on eating habits.

A. Recommender System

The recommender is a subclass of an information filtering system to provide items suggestions to be used by the user [7] [8]. Recommender systems have become popular in recent years and many big companies have adopted these techniques to increase their sales and improve customer experiences using multimedia platforms [9].

Nowadays, the design of recommender systems considers two main commonly used approaches, one is collaborative filtering that predict what users will like based on their similarities with other users to make recommendations using a large amount of data on users' behaviors, activities or preferences [9]. The second is content-based filtering, this method combines the description of an item and the profile of a user preferences to provide a suggestion of selecting an item with similar characteristics to another item that the user liked or linked in the past [10].

In recent years, researches have demonstrated that a combination of collaborative filtering and content-based filtering produce better results to make recommendations in some cases. An example of this hybrid recommender approach is Netflix that brings recommendations to the user based on searching habits of similar users and the highly-rated films made by himself [11].

B. Databases

There is a great variety of databases applied to different applications that can be classified in: Relational Database (RDB) and Non-Relational Databases (NoSQL) [12].

- *RDB*: It is a set of tables that contain specific data. The table contains one or more categories of data in columns. Each row in a table contains a data instance of an object. Relationships can be established between the tables to resemble the real world [13].
- *NoSQL*: These types of databases present a structured storage, depending on the way in which they store the data, they can be in: documents, columns, key/value, multi-value and oriented to graphs [14].

Currently, most of the recommendation systems use a NoSQL database because of the properties they have in managing the data, such as: Facebook and Twitter [15] using Cassandra. For this development, considering the concept and graph database utilization, it had been used neo4j [16].

Graph databases have many interesting features like reliability, high availability, fast response, high level of replication as other similarities with No-Sql database [17]. These characteristics make that this type of databases is been widely used in different type of sectors as bank, networking, aviation, etc. but the real advantage of the graph database use is the relationship established between the element and power of inference that this structure provides. Because of these features this database is used in fraud detection, real time recommendation, graph based search, networking, manage and operate, real time traffic, etc.

One of the most known case of use this type of database was the case of ”Panama Papers” where the International Consortium of Investigative Journalists (ICIJ) used neo4j database for storing and processing the leaked information, provided by all the journalist network [18], it provided a posterior use in the same sector as is mentioned by Gustavo C. G. van Erven, Maristela Holanda and Rommel N. Carvalho in Detecting Evidence of Fraud in the Brazilian Government Using Graph Databases [19]. On the other hand, Walmart used a graph database for improving the customer experience in all his more than 11000 stores. It Replaced complex batch process real-time online recommendations, building a simple, real-time recommendation system with low-latency queries, providing better and faster recommendations by combining historical and session data[20].

All that experience related to the use of graph databases for making a real time recommendation and infer information, made us to use this type of technology as a core tool for improving the customer experience based on the recommendation and detection of a pattern change in eating habits.

III. Data Extraction and Representation

This section describes the data provided and state the techniques used to process it in order to get the information needed for building the recommender system.

A. Experimental Data

The data in the shape of anonymised purchase tickets were provided by Carrefour Company [21], we use the set of tickets as a core for building the recommender system. The tickets consist of a 1 GB json file, this file contains a sample of more than 580,000 tickets for two different stores that date from Jan 2016 to May 2016 containing more than 50,000 different products, over 60,000 clients. The file contains the fields described in Table I.

Table I
Fields that contains customer’s purchase ticket

Field	Description	Type
ID	Number id for that individual ticket.	NumberInt
MALL	Store where the ticket was printed. It has two values, 1 and 2.	NumberInt
DATE	Date and time the ticket was printed.	ISODate
CLIENT	Some tickets will have a Customer ID. Many tickets will share a Customer ID.	NumberLong
ITEMS	List of items contained in the printed ticket. The list contains a dictionary with a product description (desc), the amount charged (net_am), and the number of units bought (n_unit).	[{ ”net_am” : NumberInt, ”n_unit” : NumberInt, ”desc”,: String, }]

B. Graph Structure

The information provided by the purchase tickets allows us to construct a graph representation of the data, and store it in a neo4j database. For taking advantage of the information provided, we establish strategic relationships between the different nodes of the graph in order to determine the consumer habits using a simple cypher query’s, using this way of retrieving data, we get a fast response being this a essential feature for providing a real time recommendations. The nodes and relationship of the graph constructed are presented in the Table II and a graphical representation of this database is showed in the Fig 1.

Moreover, as we state in section II, a recommender system use two techniques, collaborating filtering and content based filtering, or a combination of both. This graph representation of the data uses the concepts of these two techniques in a simpler manner. For example, if a want to recommend a recipe for a user, we can determine it through the relation established between a product, and user even if that product is part of a determined recipe, this type of information is easily retrieved using a simple cypher query.

Table II
PROPERTIES OF NODES AND RELATIONSHIPS IN THE NEO4J DATABASE

Element	Name	Data
Node	Store	id; name; address; longitude; latitude
	Person	id; name; surname
	Ticket	id; name; total; date
	Item	id; name; price; weight; calories
	Recipe	id; name
Relationship	Contains	quantity
	Creates	id
	Gets	date
	Needs	quantity
	Stores	quantity

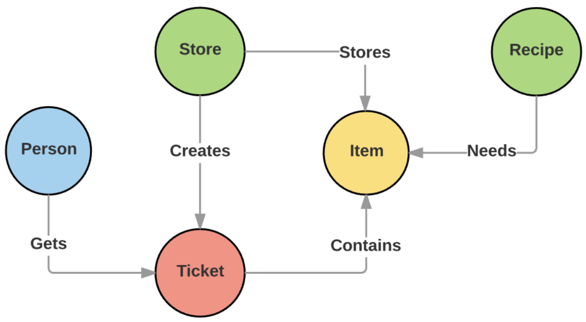

Figure 1. Interaction between nodes and relationships

IV. SYSTEM DESIGN

The main goal of this work is to build a recommender system that provides recipe recommendations based based on purchase that the user made in a group of stores. In this section we describe all components of the system previously mentioned using the information of the purchased tickets described in section III.

The system is composed by the system backend, in which we have the database with the information of all the items purchased by each user, this database is connected with a recommendation algorithm developed in python, which is in charge of executing all the queries to the database for getting the recipes based on his consumer habits, and send the recommendations to the web application. Finally the recipes are shown to the user through the GUI of the application, also, the information with the location of the store in which he/she can find each product is shown. The Figure 2 presents all the interaction that the user and the system component have between them.

A. Customer

The customer must have an identifier or user that allows him/her to access the system and check the recipes recommended by the application based on the purchase tickets.

B. Web Application

The web application is developed using HTML5, JavaScript and CSS3. In addition, for the server, it is used as a Python programming language for the interaction between the client and the neo4j database.

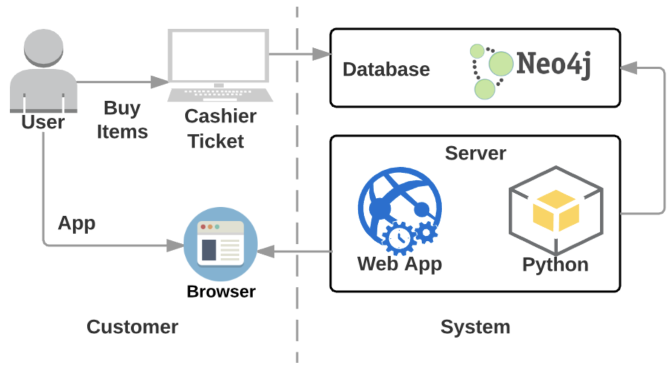

Figure 2. System Components

C. On-line Payment

The user has the option to make the purchases on-line for the supermarket to take them to his home and an order module of articles through SMS using TELESTAX [22] has been implemented, as shown in Figure 3:

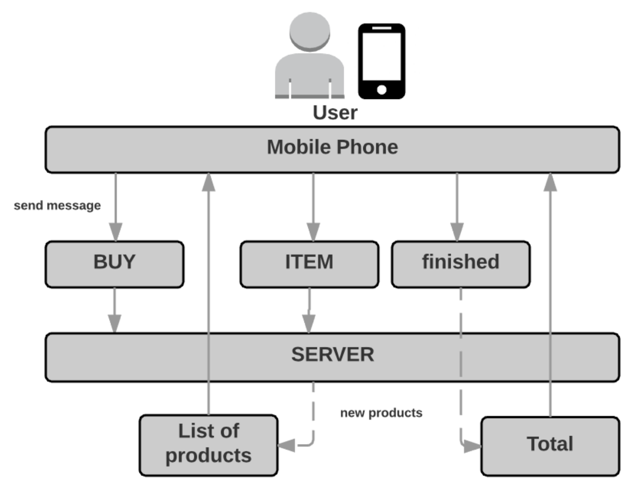

Figure 3. Interaction between the user and the system (Delivery)

The user sends a SMS with the word ”BUY”, the server sends a list of products to the user, the user responds with items that he/she wants to buy (one or many). When he/she has finished with the whole list, the user sends a SMS with the word ”YES” and the server sends the total pay. One of the motivation for implement this type of transactions is to introduce the use of SMS payment associated to electronic money accounts that many counties are promoting in the recent years.

V. Results

As a result of this work all the components described in the previous sections were implemented and test in the TADHACK Global Competition 2016, wining the first place in the Global category Carrefour Challenge [23]. In the following subsections a more detailed presentation of the developed system is shown.

A. Application

This section describes the results obtained in the application design, as shown in Figure 4, when the user logins, he/she has the option to review the recipes that are recommended to cook based on the purchase tickets. To provide this, we follow the next procedure. First obtain the products most bought by the user, using the following cypher query:

```
Match (a:Person)-[r:GETS]->(b:Ticket)-
[r1:CONTAINS]->(c:Item) Where a.id_person
='"+_user+"' Return SUM(r1.quantity)
```

Then, we choose the recipe that contains the largest quantity of products purchased.

```
Match (a:Recipe)-[r:NEEDS]->(b:Item) Where
b.id_item="_item" Return DISTINCT(a)
```

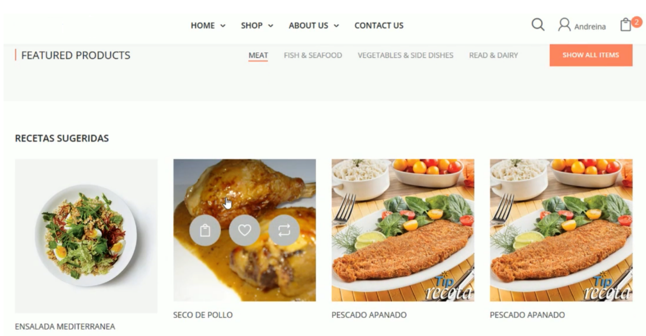

Figure 4. Recipes for a user profile

In figure 5, selecting the "seco de pollo" recipe (Ecuadorian food), the products or items that the recipe needs to be prepared with a brief description of the preparation procedure and his location in the store of the supermarket company are shown on the left side. This is achieved using the follow cypher query:

```
Match (a:Recipe)-[r:NEEDS]->(b:Item)
Where a.id_recipe="+_id+" return
[r.quantity, b.name, b.calories, b.cost]
```

In addition, an item search option was implemented using Sketch up for providing to the user a interactive visualization system for locate a determined product inside of a selected store[24] (a 3D design tool provided by Google), as shown in Figure 6.

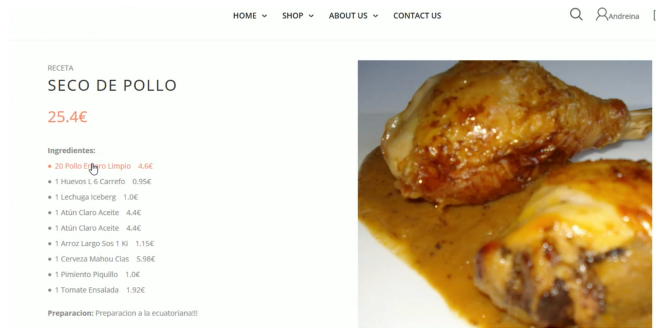

Figure 5. "Seco de pollo" recipe: Products and method of preparation

Figure 6. 3D virtual way to search for a product with Sketch Up

B. Change Patterns

The system allows to detect certain patterns in the clients' food habits, using variables such as: calories, meat and vegetables. In Table III the user maintains his eating habits and Table IV shows a change in his habits. The system recommends recipes based on these changes, even if it exceeds the parameter (calories) indicates an improvement in their habits. All of this information is easily to determine due to the inference power that provide the relationships established in the graph representation of the stored data as we showed in section III.

Table III
Person who maintains the same eating habits

Item	january (Unit / Month)	Febraury (Unit / Month)
Long Rice	2	5
Chicken breasts	8	10
Pineapple juice	3	8
Beer	12	14

VI. Conclusions

This implementation start a new approach for making recommendations using the inference process and the base concept of recommender systems. Also adopt the use of a graph database as a more lightweight way to manage and use the data in a large scale systems. The system presented in this article was tested by people related to the supermarket

Table IV
PERSON WHO CHANGES HIS/HER EATING HABITS

Item	January (Unit / Month)	April (Unit / Month)
Long Rice	2	2
Chicken breasts	8	4
Bacalao Fish	0	4
Beer	12	0

industry, giving a positive feedback even in the technical process for building the system and the improvement in the user experience.

Despite that the graph database presents wide advantages in the use and the storage of the information with many success cases as the Walmart case presented in the section II, his use in the sector of the e-commerce in not widely spread. But with this new grasp of the use of this technology, many sectors have shown interest in adopting new techniques that allow to the supermarket company to offer better services.

ACKNOWLEDGMENT

Thanks to the Escuela Superior Politécnica del Litoral (ESPOL) and to the Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) for financing this research works.

REFERENCES

[1] F. Ricci, L. Rokach, and B. Shapira, *Introduction to recommender systems handbook*. Springer, 2011.

[2] J. He and W. W. Chu, *A social network-based recommender system (SNRS)*. Springer, 2010.

[3] P. Kazienko, K. Musial, and T. Kajdanowicz, "Multidimensional social network in the social recommender system," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 4, pp. 746–759, 2011.

[4] X. Liu and K. Aberer, "Soco: a social network aided context-aware recommender system," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 781–802.

[5] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, vol. 7, no. 1, pp. 76–80, 2003.

[6] W. Velasquez, *Recomendador sensible a contexto de elementos educativos*. Editorial Académica Española, 2016.

[7] T. Mahmood and F. Ricci, "Improving recommender systems with adaptive conversational strategies," in *Proceedings of the 20th ACM conference on Hypertext and hypermedia*. ACM, 2009, pp. 73–82.

[8] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, 1997.

[9] H.-H. Chen, I. Ororbia, G. Alexander, and C. L. Giles, "Expertseer: a keyphrase based expert recommender for digital libraries," *arXiv preprint arXiv:1511.02058*, 2015.

[10] C. C. Aggarwal, *Recommender systems*. Springer, 2016.

[11] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE transactions on knowledge and data engineering*, vol. 17, no. 6, pp. 734–749, 2005.

[12] W. A. V. Vargas, "Bases de datos orientadas a grafos y su enfoque en el mundo real."

[13] C. Date, *The relational database dictionary*. Apress, 2008.

[14] E. Redmond and J. R. Wilson, *Seven databases in seven weeks: a guide to modern databases and the NoSQL movement*. Pragmatic Bookshelf, 2012.

[15] T. Github, "Cassandra Database Github," https://github.com/cassandra-rb/cassandra, 2014, [Online].

[16] A. Vukotic, N. Watt, T. Abedrabbo, D. Fox, and J. Partner, *Neo4j in action*. Manning, 2015.

[17] J. J. Miller, "Graph database applications and concepts with neo4j," in *Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA*, vol. 2324, 2013, p. 36.

[18] M. Cabra. (2016) How the ICIJ Used Neo4j to Unravel the Panama Papers data and research unit, icij. [Online]. Available: https://neo4j.com/blog/icij-neo4j-unravel-panama-papers/

[19] G. C. van Erven, M. Holanda, and R. N. Carvalho, "Detecting evidence of fraud in the brazilian government using graph databases," in *World Conference on Information Systems and Technologies*. Springer, 2017, pp. 464–473.

[20] K. Nixon. (2016) How Walmart Uses Neo4j for Retail Competitive Advantage dproduct team. [Online]. Available: https://neo4j.com/blog/walmart-neo4j-competitive-advantage/

[21] C. C. Github, "Challenge Carrefour - TadHack 2016 Github," https://github.com/ging/carrefour_basket_data_challenge, 2016, [Online].

[22] Telestax, "Real Time Communication - Telestax," https://telestax.com, 2017, [Online].

[23] A. QUAYLE. (2016) TADHack 2016 Winners tadhack. [Online]. Available: http://blog.tadhack.com/2016/10/16/tadhack-2016-winners/

[24] A. Chopra, *Google SketchUp for Dummies*. John Wiley & Sons, 2007.