Contact
CaleaDorobanţilor, no. 15-17, room 2017, Bucharest, Romania
Web: http://dbjournal.ro
E-mail: editordbjournal@gmail.com; editor@dbjournal.ro

CONTENTS

Designing a Currier Optimized Route Mobile Application

Luiza Cecilia MAZARE, Simona Vasilica OPREA
Department of Economic Informatics and Cybernetics
The Bucharest University of Economic Studies, Romania
luiza.mazare@gmail.com, simona.oprea@csie.ase.ro

Nowadays, most activities have had encountered changes due to the technology progress process. There is an obvious preference for the use of applications that help people with the everyday tasks, such as: getting a cab, booking a ticket at the cinema or even waiting for the online shopping to arrive. In this paper, we propose a mobile application for currier activities with optimization route that improves the performance.

Keyword: *currier optimization route, mobile application design*

1 Introduction

Online shopping was a new yet weird concept in the beginning, but after some time it became the first choice for most of the people. The idea of online shopping was applied for all sorts of products, such as: furniture, clothes, books, cosmetics, flowers and much more items that we use every day. The work schedule most people have nowadays doesn't leave so much time left for going to stores so online shopping is a very good alternative. This process can be done through a mobile application or through a web application. In the end, the result is the same through both of these options: the items are home delivered without any effort from the buyer and being less time consuming.

Regarding online shopping, we also need to mention a very important actor: the currier companies. These companies take the products or the packages from the sender and deliver them to the receiver. This entire process helps the way that the desired product gets from the warehouse to the receiver in almost no time, even maybe in the very same day. The arrival of the order is influenced by the location where the destination point is and sometimes by the shopper options.

Given these aspects, the currier companies started to grow more and more. The payment for the ordered items can be done either online or to the courier. Also, an important aspect is that currier companies deliver packages from one person to another.

The main benefit of this entire process is that the delivery is way less time consuming, giving the opportunity to place orders with emergency status (this means being delivered very fast). The competition between currier companies is at the level of delivery speed. Therefore, it is intended to minimize the time the customer has to wait in order to receive the package.

The entire process described above can be improved by a mobile application. Most people nowadays own a smartphone or even a tablet, and most of the activities that we could only do from standing at the computer or at the laptop can also be accessed from the phone. Mobile applications are widespread because they offer, through a user-friendly interface, benefits within certain domains. An application for a currier company helps the customer to take a new step towards the easy delivery of packages.

2 Similar applications

At the moment, the currier companies are on the rise, being the preferred method in the case of people who want to move a particular object from one area to another. Every company that works in this activity field has a web page or mobile application, through which is connected with the potential customers, providing relevant information about the services they offer.

A very popular application is the one from FAN Courier. It provides the ability to monitor order status using AWB Tracking technology, generating an estimated order price, dynamics map viewing, having a user-friendly simple interface, but very comprehensive.

A similar application is the AfterShip Package Tracker. This application main purpose is keeping track of the location where the order placed by the client is located. This application is an important support for other popular application used in this domain. The features they use consist in: notifications sent by the application, data synchronization, no ads and it can be used free. The process of using this application consists in typing the order number, then the application will identify the package. It also has the ability to scan the barcode.

A well-known currier company is FedEx. It has an application that implements order tracking, notifications sent to the client through the app, and the ability of using it after the creating of an account. An interesting functionality is the implementation of digital signature. The application also has the ability of scanning the barcode, estimate the price and the time the package will reach its destination.

The 17TRACK application also aims to monitor your orders. It provides support for international currier companies. Its main functionalities are: code scanning, changing the language in which the application can be used, etc. The application works on mobile devices and also on tablets. Once the status of an order has changed, the client is notified. A very important feature is that the application synchronizes its data in cloud.

Another currier application is CourierManager that can be used in correlation with the web page. Its functionalities are related to the physical track of couriers through a map, generating information about their previous experience. The application is connected to a portable printer and it can monitor the amount of money that has been paid. Google Navigator and Google Maps applications are used to generate routes. It also allows the display of order's content.

3 Designing the application

In order to access the application, the client needs to create an account. Through it, the user can access all the functionalities offered by the computer system. Account creation is done by filling in a form. Subsequently, when placing an order, it is necessary to fill in a form. Once the order has been sent, it can be reviewed and its status can be tracked. Also, the saving of an order is marked by receiving a confirmation email. The order is picked up by the courier. His account is created by the administrator and grants him access to all the functionalities within the application. The courier takes the order to the warehouse to be sorted and delivered to the warehouse situated in the delivery sector. From that point the order will be taken over by another courier who is commissioned to deliver the package to the consignee. The application's functionalities are presented in Figure 1.

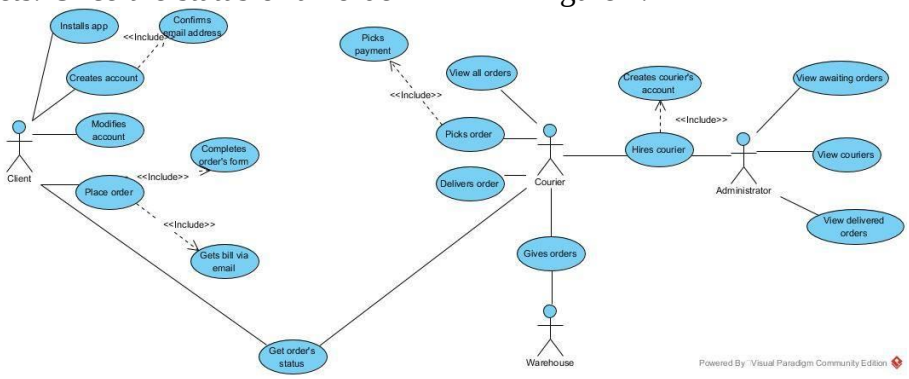


Fig. 1. Use case diagram

The most important cases in the application are: placing and delivering an order. Figures 2 and 3 illustrates the two use cases. The customer needs an account. Then the order is registered and appears as an awaiting order for the couriers in a particular sector. The client receives a confirmation email.

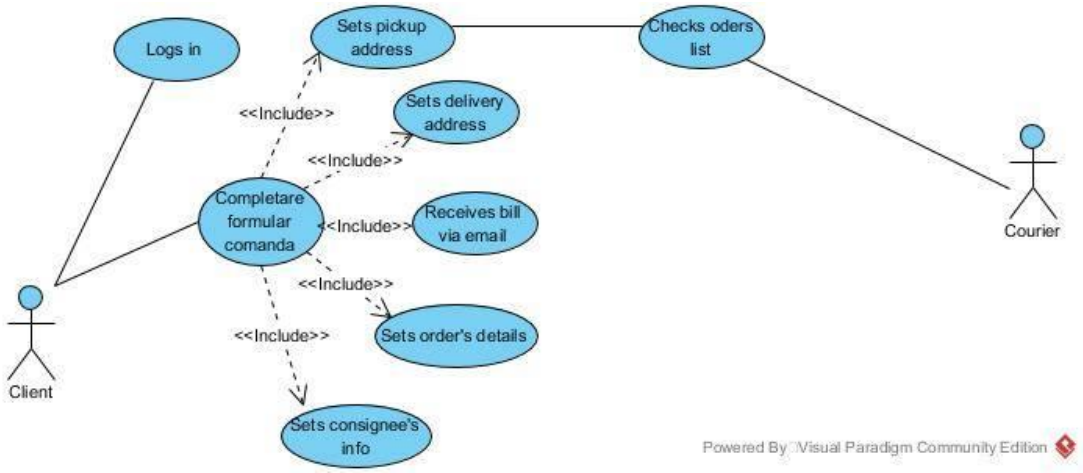


Fig. 2. Placing an order

In Figure 3 is illustrated the case of picking an order by courier and delivering it to the warehouse. The courier has to be logged in the application. The courier needs to go to the picking point and take the package along with the payment.

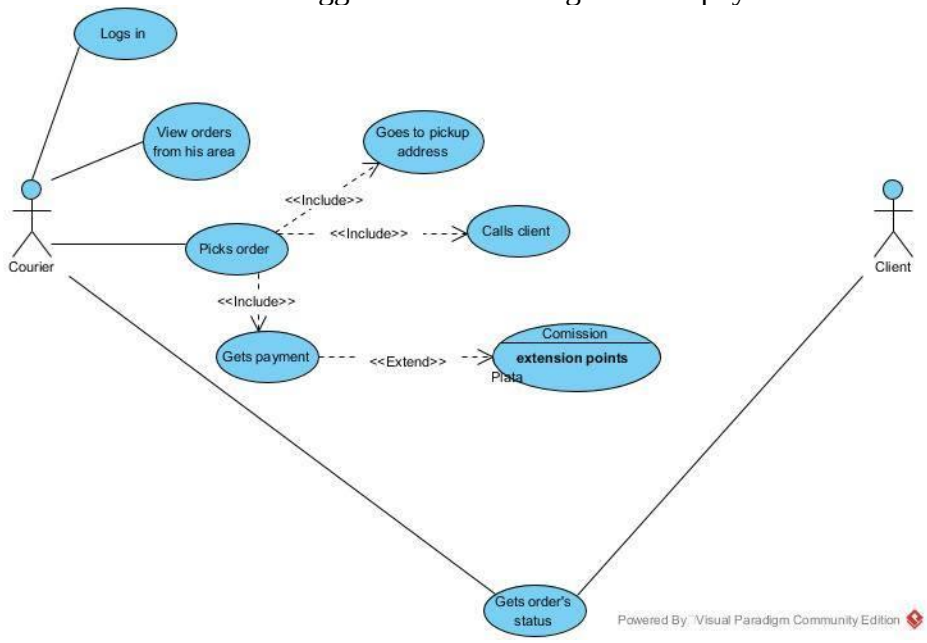


Fig. 3. Picking up an order

The creation of an account only applies to clients and couriers, administrators having their accounts created previously. The user opens the application, accesses the client menu and completes the registration form. If the fields from the form are not validated then the form cannot be submitted. Once the account has been successfully created, the email address needs to be confirmed. After this step, the account can be used without any other restrictions. If the email address is not confirmed, the account cannot be accessed. The process of creating an account for the employee starts when the courier is employed. The administrator creates the account for the courier. In Figure 4, the activity diagram is shown.

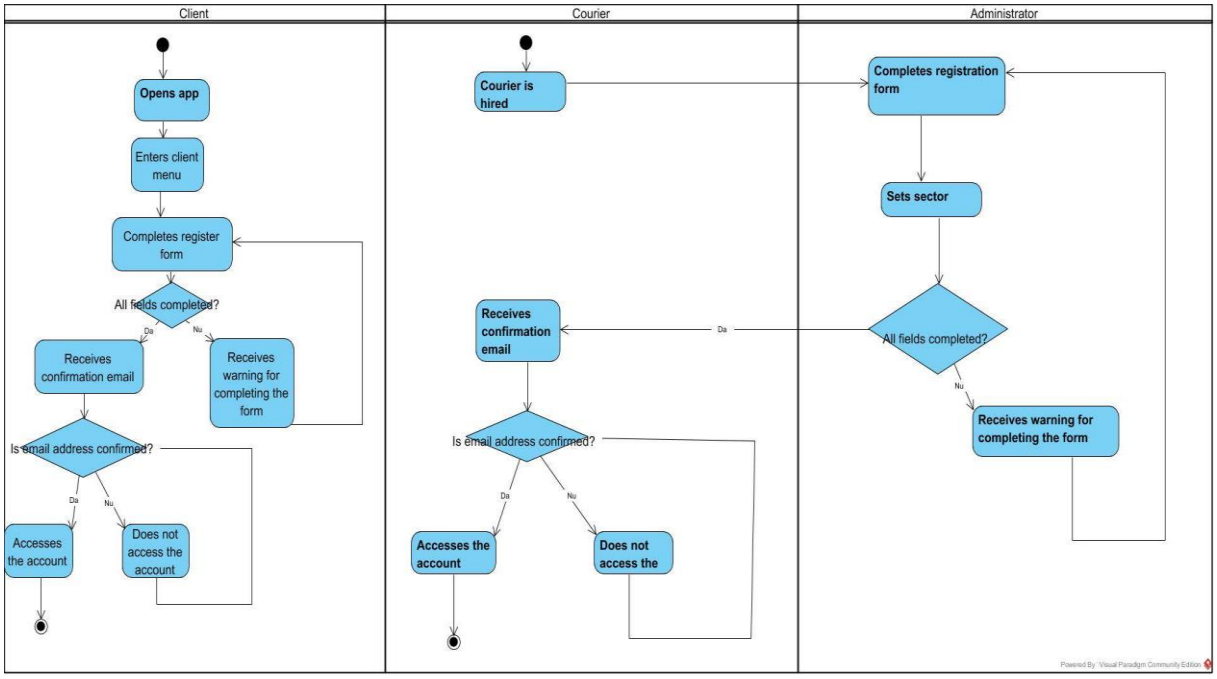


Fig. 4. Activity diagram for the creation of the account

Figure 5 shows the process of picking an order. This involves client-side authentication.

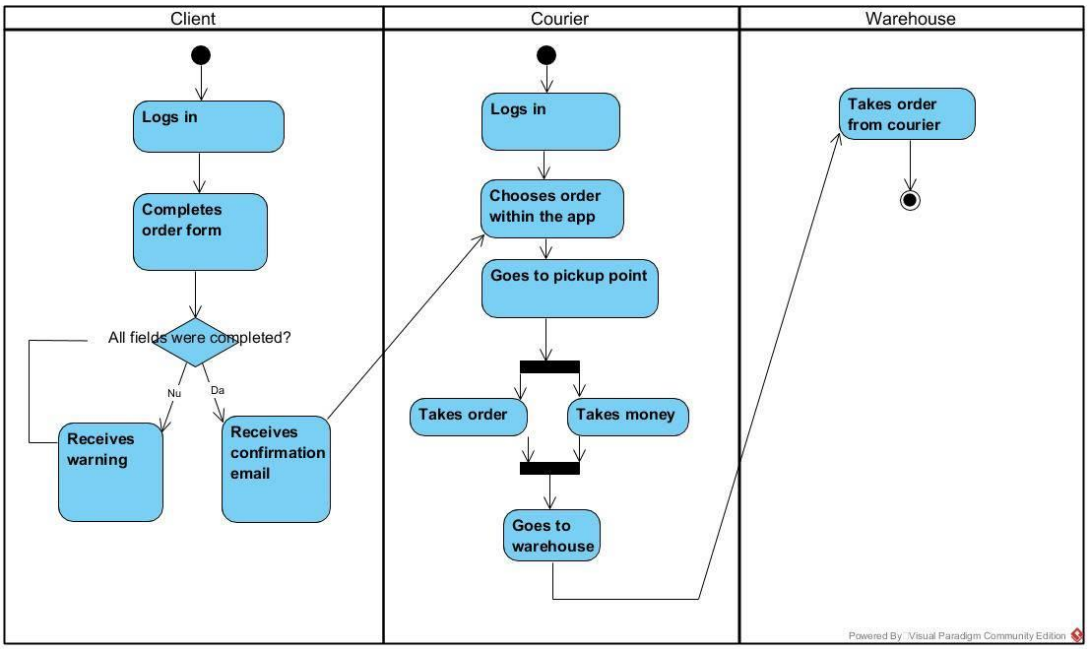


Fig. 5. Picking up an order

Following internal sorting within the warehouse, the order is picked up by the courier. It goes to the destination, contacts the recipient and takes over the command in the possession of the courier (as in Figure 6).

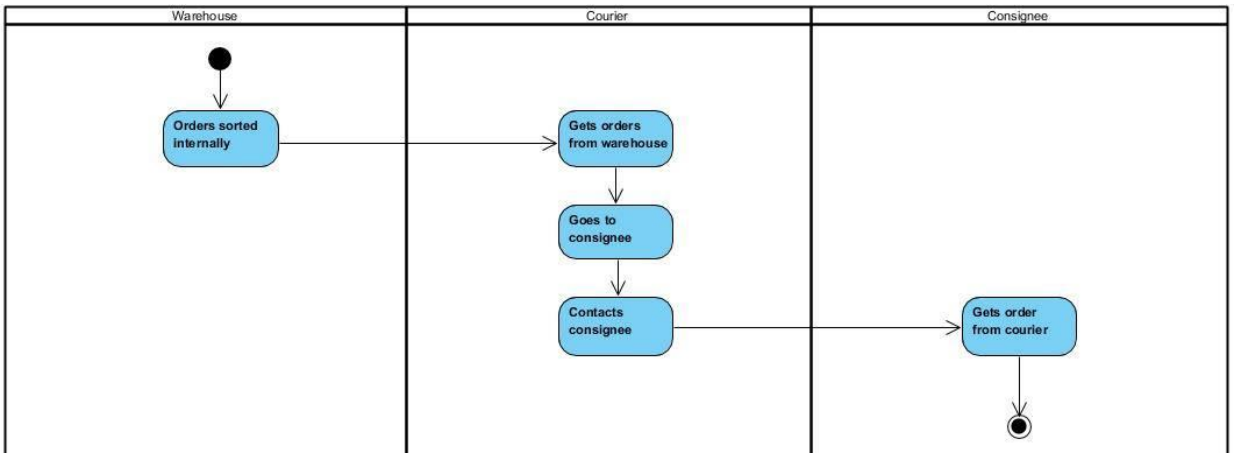


Fig. 6. Delivering an order

The BPMN process diagram illustrated in Figure 7 represents the way the application works in the client-specific interface. This process is executable and basically shows the main thread of actions a client is naturally pursuing when using the application.

Fig. 7. Process diagram

The diagram illustrated in Figure 8 is a BPMN collaboration diagram. It deals with the process of placement of an order by the client and picking up by the courier. It highlights the exchanges of information between the two entities and the way they communicate with each other during this process.

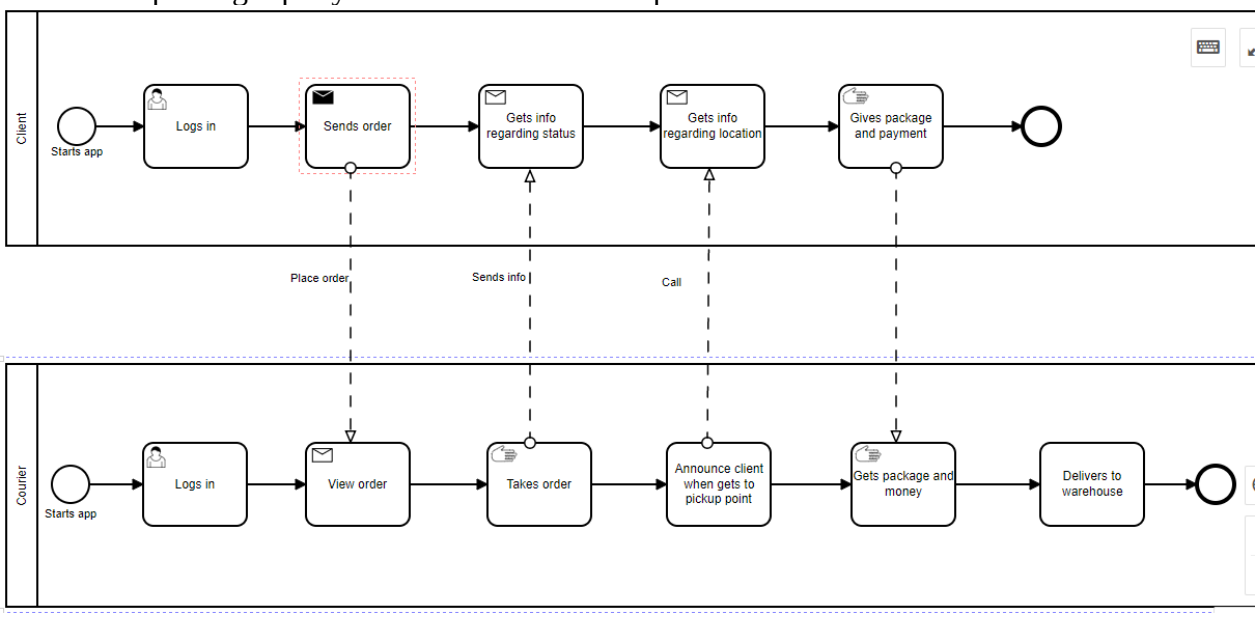


Fig. 8. Collaboration BPMN diagram for placing an order

An important process within the application is the delivery of an order. To illustrate it, we use a BPMN collaboration chart (as in Figure 9).

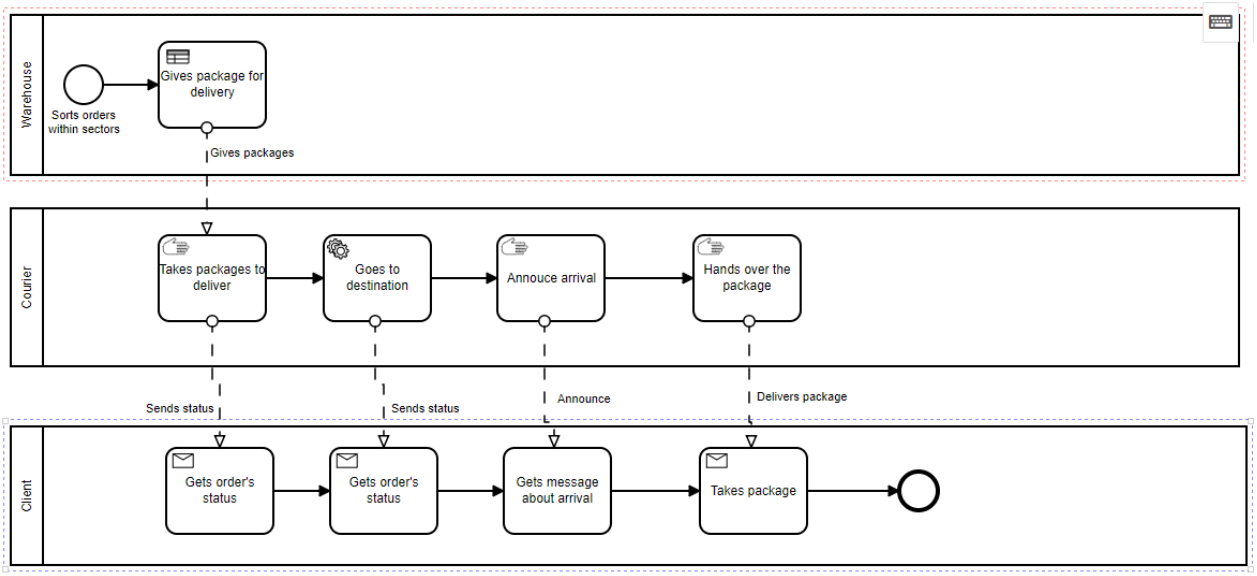


Fig. 9. Collaboration diagram for delivering an order

The Database schema is represented in Figure 10.

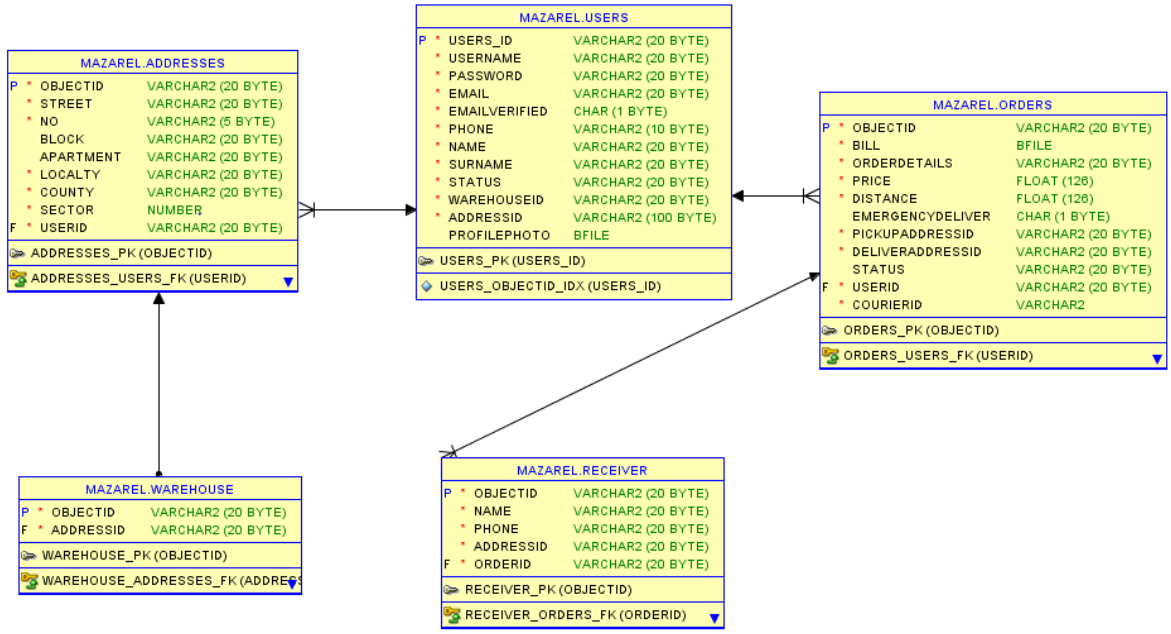


Fig. 10. Database schema

4 Software technologies

Parse server is an open version that supports any type of infrastructure running on Node.js. The Parse server has some essential features, such as:

- Uses a MongoDB database and does not depend on a database hosted in Parse;
- Provides the ability to migrate any existing application to its own infrastructure;
- The built-in application based on a Parse server can be tested and developed locally using Node.
-The server does not depend on a backend hosted in Parse [2].

The databases used in the Parse server are MongoDB and Postgres. For security of the application, it is not necessary to use a client key. The server itself includes client key, REST API, .NET and JavaScript. Each application has a unique id through which the application security process is performed. Also, the read-only key is available to the developer. By enabling it, the server will perform all reading operations, but will not be able to perform

any write operation. This type of key is used in the dashboard.

MongoDB. It is a NoSQL database. NoSQL concept includes a huge variety of technologies that have emerged due to demand for the development of modern applications [3]. Application developers process very large volumes of rapidly changing data such as: structured, unstructured, semi-structured, polymorphic.

Android Studio. The structure of a project in Android Studio includes Android modules, bookstores and Google apps. The language used in Android Studio is Java. The most important files in a project made in Android Studio are: Manifest file, Java files and res files. The Manifest file expresses permissions that allow the application to access various pre-installed applications on that device. Java files represent the place where all the functions and functionality of the application are implemented, representing the behavior of the application during its use. Files that do not contain code are stored in res files. These are XML files that help add visual elements to the application, such as buttons, colors, images, etc. [4]. Android Studio can be integrated with Google Services, a way to make it easier to deploy features like Google Maps, Google Direction.

Google play services. Google Play Services can use the latest and most popular APIs without the need for support on your device. When versions of Google Play are undergoing improvements, they are automatically distributed and delivered via Android SDK Manager [5].

Google Maps. The direction arrow is intended to find the route from a starting point to an arrival point. Through it, one can get directions by various means of transport, such as walking, driving, public transport or pedal. At the same time, one can specify destinations, stop points in the route as text, longitudinally or latitude coordinate, or location IDs [6]. Through this API, optimal routes are returned, and the first factor behind the optimization of a route is the time spent on the route. Also, in the process of calculating the optimal route, there are other important factors that account for, such as the distance and the number of curves.

The travelling salesman problem. Generating an optimized route for delivering and taking orders, it is an important step in completing the menu for the courier interface. To implement this functionality, we use the Google Maps API. By querying the database, we took over all the addresses to which a courier has to arrive within a day and created a link. A JSON file is found when accessing the address. The link should follow the format provided by Google Maps. Within the JSON file, the addresses are stored in an optimized order. Also, the directions the driver must follow to reach the destination points are included. With Intent objects, the link must be sent to the Google Maps app. Then, it was necessary to create another address that respects a different format to be compatible with the map application. So, we obtained the addresses from JSON file first, then stored them inside a list we used to create the new link. This method has ensured that the addresses will remain in the same optimized order. The goal is to find the shortest route between various destinations where the salesman must arrive using the algorithm for the travelling salesman problem [7].

5 Application interface depiction

The four major functionalities of the application are: placing an order, tracking the status of the placed order, order history, and how to package parcels as in Figure 11.

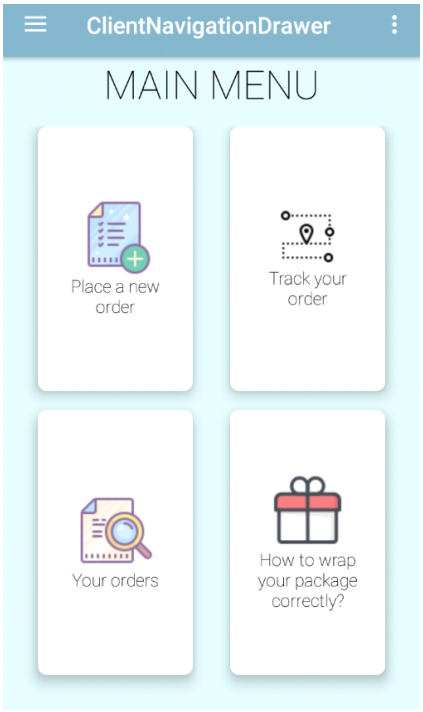


Fig. 11. Client interface menu

The order placement form gives the client the possibility to enter a new address or use one of the saved ones so far. Once the user sends commands from or to a specific address, it is automatically saved in the database. It must then enter information about the recipient, such as: name, first name, telephone number. It also requires entering information about the subject you are sending, such as the number of pounds or envelopes, weight, length, width, height. These fields contain validations, and it is not possible to send very large or heavy objects. The price is calculated dynamically, based on the weight and size you entered. The user may opt for an emergency delivery, but it will cause an extra change. It is not acceptable to send it to the server if the data are not complete or valid and the required fields are left blank. Errors are generated on the fields where the entered data did not meet the required requirements. Once the form is sent to the server, the customer receives the invoice by email. This is in pdf format and contains the important information in the order, plus the unique QR code of the order. The invoice must be printed and glued on the package to be scanned later by the courier.

The basic functionality in the courier interface is the way the route he has to take is optimized. This route must be optimized in terms of both distance and time. The route is made up of all the addresses he needs to reach, and if additional addresses are added, the route is reordered to include the new added point. Figures 12 and 13 illustrate two routes, one optimized, and the other non-optimized, in which the addresses are given in a random order. The first criteria to be considered in the optimization process is the distance traveled, observing that in the optimized route, the distance is 28 kilometers. In the non-optimized route, the distance traveled is 38 kilometers, a major difference in the resources that the company should allocate to the same transport. Also, the time difference is approximately one hour between the two routes. It can also be seen that addressing is done in a different order within the two routes.

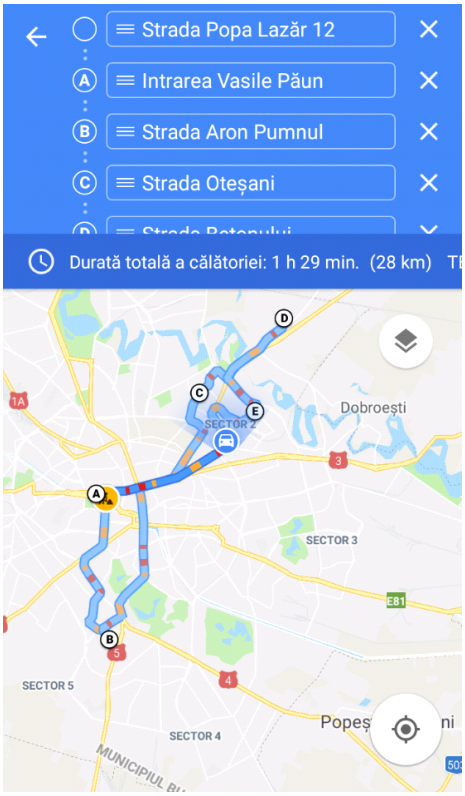


Fig. 12. Optimized route

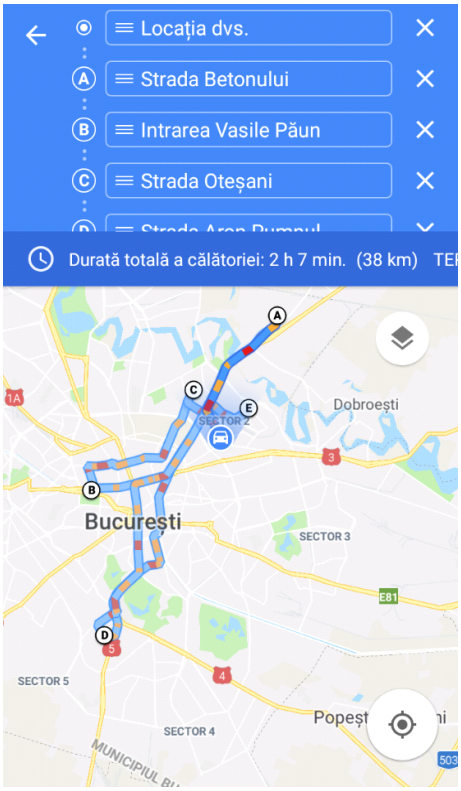


Fig. 13. Non-optimized route

The interface for administrators is completely different from the ones described above, both in terms of design and functionalities. The menu is based on three windows that illustrate the stage of the entire business activity. The first window lists all company employees, along with contact details. This enables them to be contacted if it is necessary. In the second window are all commands that have not been retrieved at that time. In the third window are all commands that have been delivered to the recipient. This includes the reference of all orders and the way they are delivered.

The application also generates graphs that help explain the company's performance from a financial and productivity point of view. The first graph refers to the ratings obtained from the feedback process. These are illustrated as in Figure 14. In this way, conclusions can be drawn on how customers perceive the services offered by the company.

The second graph illustrates the number of orders delivered during the last week (Figure 15). This way one can see how much the company's productivity increased or decreased, as the case may be, from one day to the next, or the level of demand for services offered by the company.

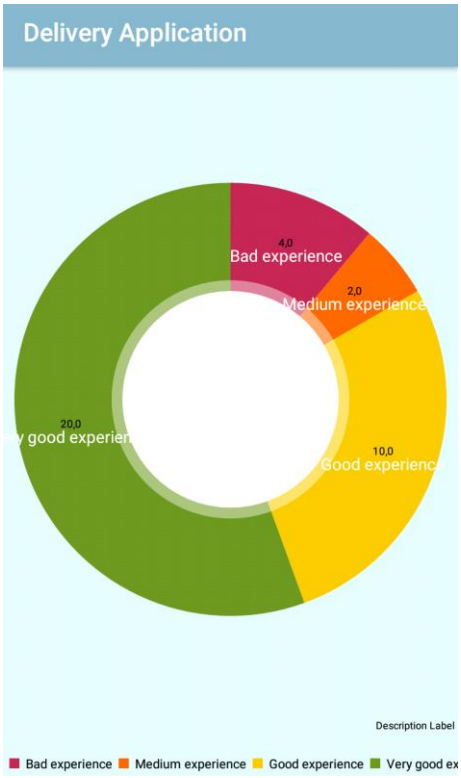


Fig. 14. Feedback chart

Fig. 15. Last week deliveries

6 Conclusions and future works

In this paper, we propose a mobile application for currier activities with optimization route using algorithm for the travelling salesman problem that improves the performance. By the optimized routs, both distances and transporting time were reduced.

References

[1] Anca Andreescu-*Dezvoltarea sistemelor software pentru managementul afacerilor*, București, Editura ASE, 2010.

[2] Parse Server guide, 2010. Available online at: http://docs.parseplatform.org, accesed June 2017;

[3] MongoDB guide, 2012. Available online at: https://www.mongodb.com, accesed June 2017;

[4] Android Studio guide, 2012. Available online at: https://developer.android.com, accesed June 2017;

[5] Erik Hellman-*Android Programming: Pushing the Limits*, Editura Wiley, 2015.

[6] Google Maps Platform guide, 2012. Available online at: https://developers.google.com, accesed June 2017;

[7] Hector Ortega-Arranz, Diego R. Llanos, Arturo Gonzales-Escriano-*The Shortest-Path Problem: Analysis and Comparison of Methods*, Editura Morgan&Claypool, 2014.

Luiza-Cecilia MAZĂRE (b. January 6, 1996) is a third year student at the University of Economic Studies from Bucharest, Faculty of Cybernetics, Statistics and Economic Studies. She has graduated Theoretic High-School "Grigore Moisil" from Tulcea in 2015 with specialization in Mathematics-Informatics, advanced English. She is familiar with Java, C# and Microsoft Dynamics 365.

Simona-Vasilica OPREA (b. July 14, 1978) is an Assistant at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from the Bucharest University of Economic Studies of Bucharest. She has graduated the Polytechnic University in 2001, holds a Master Diploma in Infrastructure Management Program, Yokohama National University, Japan in 2007, a PhD diploma in Electrical Engineering in 2009 and a PhD diploma in Economic Informatics in 2017. She is the author of over 50 articles.

Informatics solution for Smart Garden based on Sensors

Diana-Alexandra RADU, Simona-Vasilica OPREA
Department of Economic Informatics and Cybernetics
The Bucharest University of Economic Studies, Romania
radu_diana96@yahoo.com, simona.oprea@csie.ase.ro

This paper aims to develop an application dedicated to the growth and maintenance of plants, whether leguminous or decorative plants, which involves the traditional cultivation method that uses the soil. It consists in two main components: a mobile application that is be able to run on the Android operating system, and the physical component, which is referred to, in what follows, by the term "SmartGarden".

Keyword: *sensors, Arduino, smart garden, monitoring*

1 Introduction

It is well-known that fruit and vegetable producers use intensive crop farming, in order to obtain large crops in a short period of time. This method is used both to increase profits and to meet market demand, taking into account that the global population is constantly on the rise. These agricultural practices come at a price. In recent years, there has been a decline in the quality of fruits and vegetables available on the market, both in Romania and all around the world. Moreover, vegetables found in stores outside the season can be toxic because of the chemicals they are being treated with, to increase their resistance in time. They also lack taste and nutrients because they either are hybrid breeds, specially created to be cultivated in greenhouses, or are not grown under optimum conditions.

Considering these aspects, the best and most reliable alternative would be the cultivation of your own vegetables or fruits. However, people today no longer have the time or resources to grow their own vegetables, or are not quite concerned about this. This is why any technology innovation that occurs in this area is very important, as it could radically change the way plants are grown today. Also, developing devices that facilitate the growth and maintenance of plants could lead more people to choose to cultivate their own plants.

Two of the main ways to cultivate plants are: the traditional method, in which the soil is used as a rooting medium and a source of getting nutrients, and the hydroponic culture. Hydroponics means that plants grow without soil, relying on mineral nutrient solutions in a water solvent, from which the roots of the plants directly uptake the necessary nutrients. Most specialists believe that hypotonic products are not organic because they do not use soil as a growth substrate.

The physical part of the informatics solution proposed in this paper is the one that collects information from the environment and transmits it to the database, in order to be further processed. SmartGarden consists in several sensors and other hardware components linked to a motherboard. The main features are the build-in LED lamp and irrigation system. The mobile application displays data coming from sensors and provides remote control for SmartGarden. It also allows users to adjust the settings to meet their needs, and to manage a list of plants and devices.

2 Similar applications

The idea of an automated garden has been exploited so far, reaching similar solutions to the one proposed in this paper. The main difference is that most products which are already on the market use hydroponic growing.

The first solution that I am going to present is Plantui6 [1]. It consists of a water tank, an LED lamp and spaces designed for plant development. The first difference is, as we mentioned above, the hydroponic cultivation method. A second disparity is that Plantui6 uses white light, unlike SmartGarden that uses combined red and blue LEDs for greater efficiency. Also, this solution does not come with an application through which the values in the environment can be monitored. At the same time, it does not show any remote control method. An advantage of this solution is that after 16 hours of operation the device will stop for 8 hours called "sleep time" to save energy.

Another solution that resembles the application we propose is AeroGarden [2]. It consists of a water tank, a blue and red light bulb, and seed planting compartments. AeroGarden Wi-Fi is the mobile application for the solution, which promises to monitor the amount of water, light and nutrients. Also, the built-in LED lamp can be remotely controlled via the mobile application.

CLICK & GROW [3] is another solution for plant growth. Unlike the products mentioned above, it is the only one that uses soil instead of water. The soil used is called 'smart soil' and promises an optimal pH, oxygen and nutrient level. The system has the same components as the above-mentioned devices. It also does not come with a mobile application or a LED lamp.

The elements of innovation presented by the proposed application are the functionalities of the mobile application. From the descriptions of the solutions analyzed before, it is noticeable that there were not used sensors to capture information about several factors that could impact on plant development, such as air humidity, temperature, light, soil humidity. The application, in addition to providing real-time information, provides the option of seeing the changes in the form of a graph for a more eloquent representation that facilitates understanding. Another strength of the proposed solution is the "automatic mode", which allows the device to turn on certain components (lamp or irrigation system) without needing the intervention of a user. Another possible advantage of our solution might be that it is intended for growing plants in the soil and not in the water, considering that most of these products are made for hydroponic crops.

3 Application design

In Figure 1 the actions that the user can take in the mobile application are presented. There is a simple association relationship between the user and all use cases. As we can see, the include relationship was used to represent that most actions (add plant, add device, check status of parameters, modify account data, choose operating mode) that can only be performed after the authentication has been done in advance. Also, the "Check status of parameters" case extends three other cases: "Turn on irrigation system", "Turn on lamp", "View History". This extension highlights the fact that these extended cases take place optionally or under certain conditions. Thus, in the manual mode of operation of the system, if after checking the state of the parameters, the user wishes to take one or more of the extended actions, they will take place, otherwise not. On the other hand, in the automatic mode, the system will execute the extended cases: "Turn on irrigation system" and "Turn on lamp" when the soil moisture and brightness parameters will drop below the lower limit.

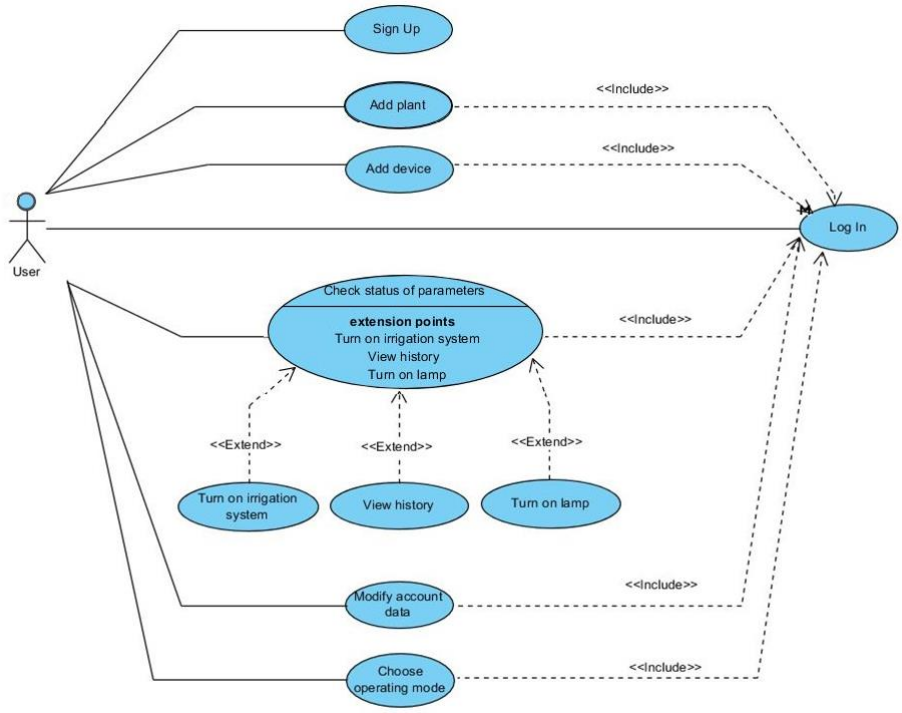


Fig. 1. Use case diagram

Activity diagram for "Turn on lamp". In the "Turn on lamp" activity diagram, we can see the steps taken by both user and system in order to control the lamp remotely (as in Figure 2). The logged-in user can turn on the lamp by pressing the button on the home page of the application. Once the button is pressed, a request is sent to the database to change the status of the lamp to 0, corresponding to the state "Off", or 1, corresponding to the "On" state. The motherboard reads the values in the database and acts as a result of the change the used made.

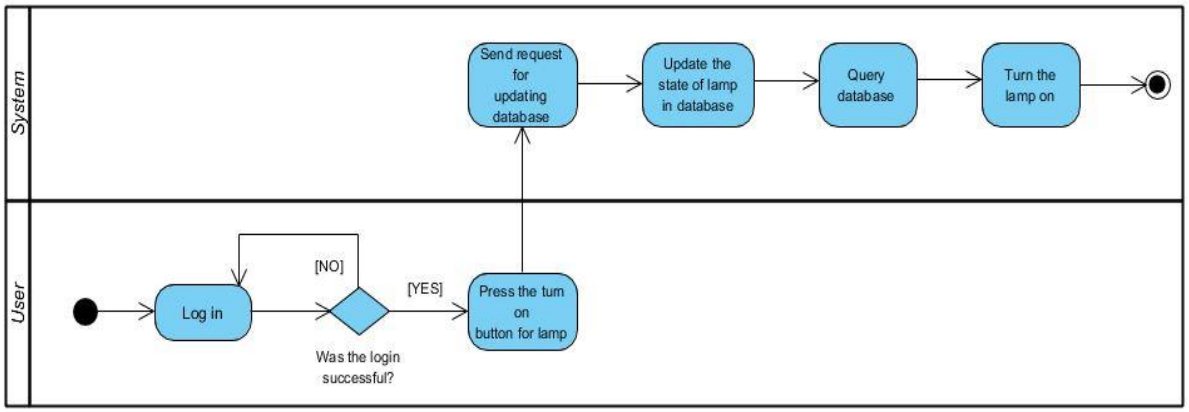


Fig. 2. Activity diagram for "Turn on lamp"

The *class diagram* is designed to graphically describe the static part of the system by presenting classes and relationships between (as in Figure 3). The classes used in the project are: Accounts, Devices, Devices_models, Plants, Default_plants, Categories, Sensor_data.

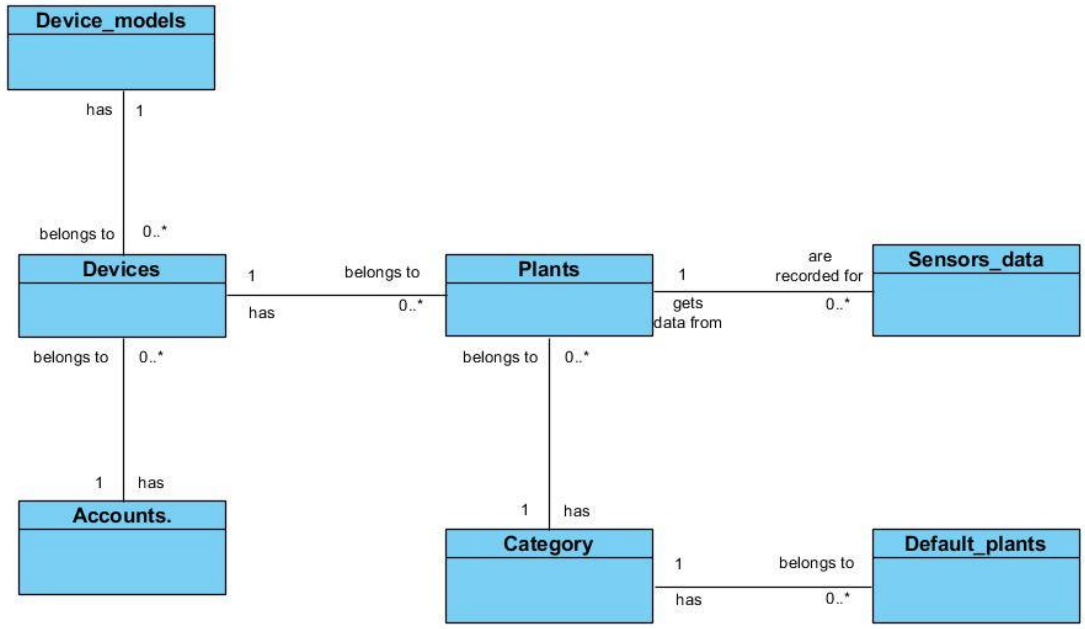


Fig. 3. Class diagram

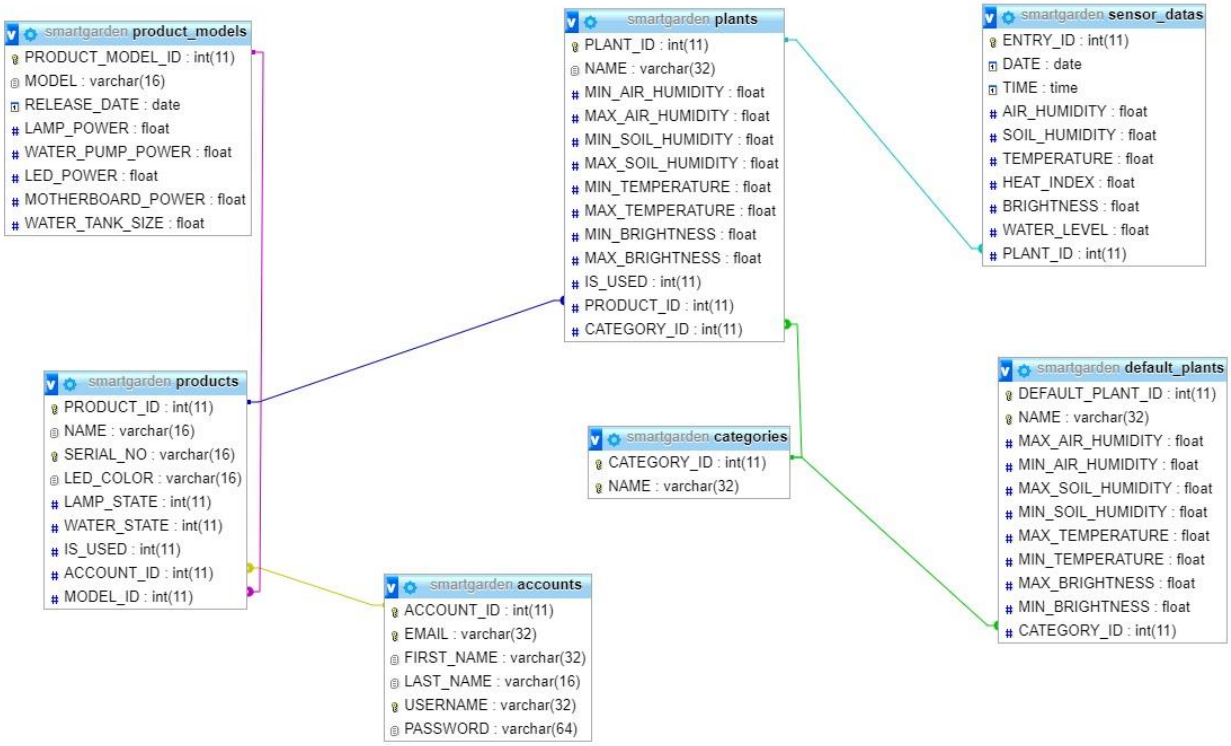


Fig. 4. Database logical schema

Relationships between entities can be expressed as follows: a device model belongs to one or more devices, a device has only one model; an account has one or more devices added, but a device belongs to a single account; a device, over time, can grow one or more plants, and one plant is grown in single device; a category contains one or more plants, and one plant belongs to a single category; a category contains one or more default plants, and a default plant belongs to a single category; for one plant are recorded one or more

inputs from the sensors, and one input from the sensors is recorded for a plant.

4 Software Technologies

As specified in the introduction, there are two components: the mobile application and the physical component, SmartGarden. The technologies used in the project are the following:

- Android platform and Java language
- Arduino software and C ++ language
- MySQL
- LAMP server

The mobile application is built through the Android Studio development environment using the Java programming language and is compatible with the Android operating system. It can run on devices that have at least Android version 4.4 KitKat, with a minimum API of 19.

Arduino software is a development environment used in the automation process to write code for the motherboard, using C/C++. To simplify code writing, the following libraries: ESP8266, WiFiClient.h, ESP8266WebServer.h, SimpleDHT.h were used. The first three libraries were used for the WiFi module and the last one for the temperature and humidity sensors.

MySQL is used to store data gathered from SmartGarden. Since records are being generated every 7 seconds, a database management system with a large storage capacity is needed. This is the means of communication between SmartGarden and the mobile application. It also manages user accounts, storing either personal data and plant listings or user configurations about how the application works.

The LAMP server is used to establish the communication between SmartGarden and the mobile application. SmartGarden queries or inserts data into the database via PHP pages running on the server. Both queries and insertions from SmartGarden are done using the POST method. Using the same method, the application queries the database and takes data coming from SmartGarden. There is another method that can execute the task, called GET. The difference between GET and POST methods consists in the security they provide. The GET method is less secure because the values are explicitly written in the link, remaining in the browser's history, and can be intercepted by unauthorized people. The POST method has a higher degree of security because the values are encoded and transmitted via HTTP headers. However, the level of security also depends on the HTTP protocol, so by using Secure HTTP (HTTPS) the level of data protection is higher.

To provide remote control, by using the mobile application, the status of the lamp / irrigation system in the database changes to 1 or 0 when the ”turn on” / ”turn off” button is pressed. SmartGarden reads the information about the new state of the device, and reacts by turning them on or off.

5 Application interface depiction

SmartGarden contains the following hardware components:

The *motherboard* is the main component of the SmartGarden project. It works as a control center where all the information transmitted by the sensors is collected and it is further made available to the mobile application through a database. The board not only gathers data, but also controls the mechanical part of the project, such as: a mini water pump used for soil irrigation, but also an LED lamp to provide the necessary amount of light (as in Figure 5).

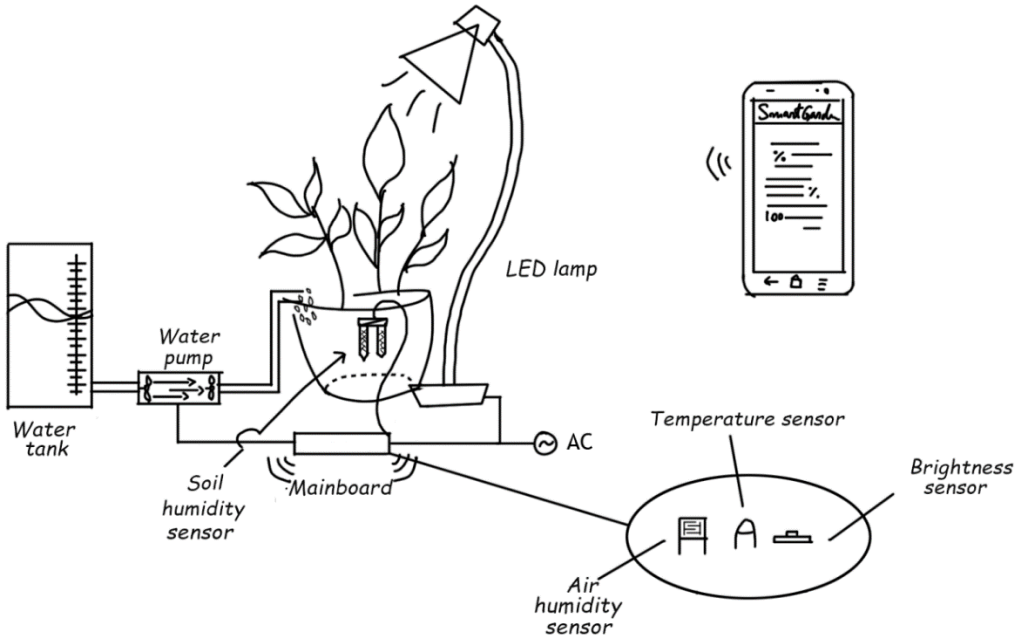


Fig. 5. Schema of components

The mainboard is like a computer, as it contains common elements such as a processor, internal memory, RAM, but also a very important element, a WiFi module capable of receiving or transmitting packets of data on a local network or on the Internet. At the computer level, the mainboard acts both as a server and as a client. On the server side, it provides a web page where you can set the name and password of the router to which it will connect in order to access the Internet, and on the client side it connects to the database.

Sensors. In order to get quicker, yet more accurate information, a range of efficient sensors has been chosen. Next, the sensors used in the project implementation and their characteristics are presented:

a) The temperature sensor - detects the temperature in the air and transmits it to the motherboard in digital format.
- The measured temperature range is -40 ° C to + 80 ° C;
- The accuracy of the measurement is ± 0.5 ° C.

b) Air humidity sensor - detects relative humidity (RH) in air and transmits it to the motherboard in digital format.
- The measured humidity range is between 0% RH and 100% RH;
- The accuracy of the measurement is ± 5% RH.

c) The soil humidity sensor - detects soil moisture and transmits it to the mainboard in analog format.
- The measured humidity range is between 0% and 100%;
- The accuracy of the measurement is ± 1%.

d) Ambient light sensor - records the level of ambient light and transmits it to the mainboard in analog format.
- The accuracy of the measurement is similar to the human eye;
- The detection angle of the sun rays is ± 60 ° and best responds around the value of 570nm of the light wavelength, which corresponds to natural light or the one produced by LEDs. The light produced by incandescent bulbs is detected very poorly because it produces another spectrum of light. This is the desired effect because incandescent light does not help plants grow.

e) Water level sensor - detects the water level in the tank and transmits it to the motherboard in analog format.

The *LED lamp* will extend the duration of exposure of plant to light. There are certain wavelengths in the visible light spectrum that are ideal for plant growth, the red and blue ones. Different proportions of red and blue light affect different types of chlorophyll differently. Depending on the stage of development, the optimal wavelengths may vary. For example, in the blooming phase, a red light spectrum is most suitable to plants. LEDs are the best choice for this solution because they do not produce overheating and have low energy consumption. Moreover, they produce the same wavelengths as the sun's rays [4].

The *irrigation system* ensures an optimal level of soil moisture. It has the following components: a water pump, two hoses and a tank with a capacity of 0.8L. By using the appropriate sensor, we will know when the soil humidity has dropped below the lower limit, so that the system needs to be turned on. This avoids both water deprivation of the plant and the use of too much water.

Multiplexer. Because the motherboard has only one analog input, it was necessary to introduce a 16-channel multiplexer in order to extend the number of analog inputs. It receives from the motherboard the channel (in binary format) from which we want to intercept the analog value and read any input on the indicated channel. The information is then returned to the board.

Relay module. In order to turn on and off the lamp and the water pump, it has been used a 4-relay module. It works as a switch, but instead of the mechanical action of pressing a button, a digital signal of 0 or 1 is transmitted to the module, which in turn closes or opens the electrical circuit. The relay module is controlled by the main board through two pins, one for the water pump and the other for the lamp (Figure 6).

Fig. 6. SmartGarden – water and light sources

Controller for LED strip. The LED strip is used in the SmartGarden's exterior design. In order to obtain different color shades, it is necessary to build a controller. This is a set of three transistors that, along with the signal received from the motherboard, control the intensity and color of SmartGarden's lights (Figure 7).

Fig. 7. SmartGarden – in operation

The LED strip consumes a higher current than the one that the motherboard can support, so an external power supply is needed.

The Android application will provide the following features:

Register of a new user. When creating a new account, the user must provide some personal data (name, surname, email, username, and password) that is filled into

a form of the application. These data will be stored in the database. Taking into account that the table containing users will contain sensitive content such as personal info and passwords, the information is encrypted using the bcrypt algorithm, because it is considered to be the most secure and efficient.

Authentication. After the account is created, the user will provide the username and password in order to login and be able to use the application. To verify the correctness of the data entered, it will be checked in the database whether the username and password combination belongs to a user. If they are correct, authentication will be made successfully, otherwise, the application will display an appropriate message to let the user know that one of the information they provided is incorrect.

Display the values from sensors. The sensors integrated in Smart Garden will record the values of environmental parameter every 7 seconds. These values will be stored in the database, in a table dedicated to sensors data only. The mobile application will retrieve this data and display it both in the main menu (Figure 8) for a quick check and on separate pages dedicated to each sensor. Individual pages will also allow you to view a history of recorded values. Thus, we can see the evolution of certain parameters that influence plant development, throughout the day. The graph and report option will also be available for easier viewing and understanding of the changes that have taken place (as in Figure 9 and Figure 10).

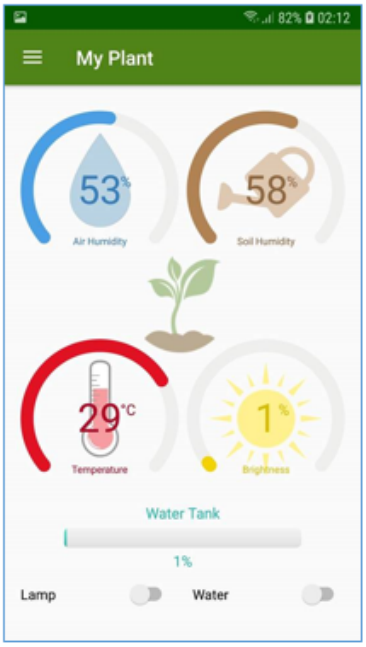


Fig. 8. Home page app

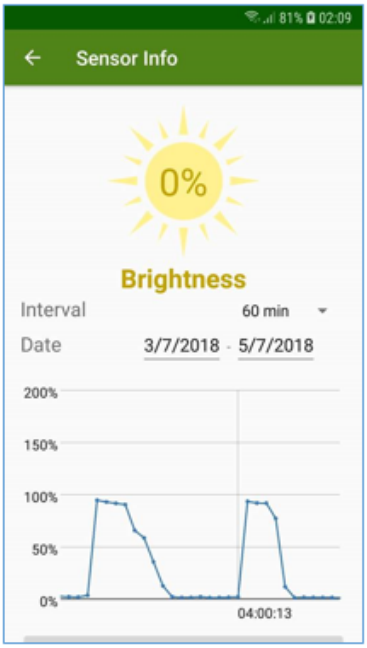


Fig. 9. Graph for the evolution of brightness

Fig. 10. Report for the evolution of brightness

Adding a plant. The option to add a plant can be accessed via the side menu of the mobile application. Here will be provided information about the plant such as: name, optimum parameter values (air humidity, soil humidity, temperature and ambient light intensity). If the user does not know the optimal values for good development of the plant, he can choose to add a default plant provided by the application. Subsequently, the user will be able to modify the entered data provided previously or remove a plant from the list, if desired.

User notification. One of the most important features of the application is the notifications sent to the user when the values of a parameter exceeded the optimal interval limits. It will also signal when the water level in the tank has reached a low

level. Depending on this, the user can make certain decisions in order to fix the situation.

Remote control. Smart Garden has a plant watering system and a special LED lamp for plant growing. It provides two modes of use: automatic and manual. If we choose the automatic mode, SmartGarden will no longer need our intervention to take care of the plant. It will take into account the measured values of the sensors and the optimal values and will automatically turn on the irrigation system or LED lamp when needed. If we choose the manual mode, the application will send a notification to the user, indicating that the irrigation system should be turned on because the soil humidity has dropped below the lower limit. Also, if the exposure time to sun light is too short, it can be substituted by using the LED lamp, which can also be turned on or off remotely, without the need for physical presence.

6 Conclusions and future works

We consider that the resulting application has achieved the goal set in the beginning, namely to automate the plant monitoring and maintenance process.

The application could be improved by the following aspects:

- improving both the mobile application design and the physical component aspect;
- more detailed analyzes on the evolution of parameters values over time;
- a system that holds the flower pot in place or a system that notifies the user when the pot have been removed, in order to avoid the situation in which the irrigation system is turned on in the absence of the pot, which could be damaging for the hardware components;
- plant growth monitoring;
- predictions;
- control of temperature and air humidity;
- notifications via email or SMS.

References

[1] Plantui6 Smart Garden, Available online: www.plantui.com
[2] AeroGarden Miracle Gro, Available online: www.aerogarden.com
[3] CLICK &GROW, Available online: eu.clickandgrow.com
[4] B. Chiang, J. Puckett (2010) – „Growing Indoor Plants Using LED Lighting”, www.kessil.com

Diana-Alexandra RADU (b. June 10, 1996) is a senior year student at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from the Bucharest University of Economic Studies of Bucharest.

Simona-Vasilica OPREA (b. July 14, 1978) is an Assistant at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from the Bucharest University of Economic Studies of Bucharest. She has graduated the Polytechnic University in 2001, holds a Master Diploma in Infrastructure Management Program, Yokohama National University, Japan in 2007, a PhD diploma in Electrical Engineering in 2009 and a PhD diploma in Economic Informatics in 2017. She is the author of over 50 articles.

Automated Trading Software - Design and Integration in Business Intelligence Systems

Cristian PĂUNA
Economic Informatics Doctoral School - Academy of Economic Studies
cristian.pauna@ie.ase.ro

After the introduction of the electronic execution systems in all main stock exchanges in the world, the role of the automated trading software in the business intelligence systems of any financial or investment company became significant. Designing of reliable trading software to build and send automated orders based on quantitative mathematical models applied in the historical and real-time price data is a challenge for nowadays. Algorithmic trading and high-frequency trading engines become today a relevant part of any trading system and their specific characteristics related with the fast execution trading process and capital management involves specific measures to be used. Smart integration of the trading software in the business intelligence systems is also a sensitive theme for any financial and investment activity, a plenty of functional, control and execution issues being subjects of researches for future improvements. This paper wants to gather together more particular aspects on this subject, based on the experience of last years, opening the way for future topics.
Keywords: *automated trading software (ATS), business intelligence systems (BIS), business process management (BPM), algorithmic trading (AT), high-frequency trading (HFT).*

1 Introduction

Trading shares in an organized stock market is already a stand-alone activity for hundreds of years. Since the transfer of an equity from a seller to a buyer for the money is a sustained process, it underwent a whole series of significant transformations.

After some organized markets appeared between the 15th and 16th century in Belgium and Netherlands, "the United East India Company was chartered in 1602" in Amsterdam and "a new bourse was begun in 1608" [1]. It was founded the first company with shares that can be bought and sold by the shareholders and this new trading shares activity does not affect the main activity of the company. "Altogether, 1,143 investors subscribed to the initial capital of the of the Company's Amsterdam chamber" [2]. "The Dutch pioneering several financial innovations that helped lay the foundations of modern financial system" [3]. It was only a step to the continuous trading of shares in the incipient stock market. A new era has begun, more stock exchanges was developed worldwide.

It is well known the image of a trading floor in a stock exchange, where the communications where made in open outcry language until recently. Even "until 2009 trades on the floor of the New York Stock Exchange always involved a face-to-face interaction" [4]. In this system all bids and offers were made out of the open market, giving all participants the chance to compete with the best price. All estimations were based on the last closed price, on the whole price history using the latest fundamental news, but no estimation was possible to be made based on the real-time price movement as we are used today.

In the last decades, introducing the electronic systems in the stock exchange activity was a significant step. Electronic trading (ET) is the method that use information technology to bring together buyers and sellers in a virtual market place using an electronic trading platform and a network that links all participants. The huge difference is now that the real-time price is available, which makes possible a more accurate prediction. The second important difference is that all orders are executed electronically, in a very fast

execution process, fact that affect the volatility and the speed of the price movement.

ET was born practically in 1971 when National Association of Securities Dealers Automated Quotations (NASDAQ) was founded. "NASDAQ was the world's first electronic stock market" [5]. After that, in 1987 Globex trading system was founded by CME Group, "conceived in 1987 and launched fully in 1992" [5]. It allowed access for ET of treasuries bonds, commodities and foreign exchange. A little bit later, the rival ET system CBOT (Chicago Board of Trade) was implemented, "an electronic trading platform that allowed for trading to take place alongside that took place in the CBOT pits" [5].

The fast IT development was a key factor in the ET domain. In the years of 2010 and 2011 massive investments were made in technology in all main stock exchanges in the world. This process determined the majority or the classical trading floors to be changed in ET systems, the classical brokers were removed from the trading chain and all trading orders were made and processed electronically. The impact was tremendous in reducing the costs of transactions, gather the liquidity and increasing the transparency of the price movement.

In this new context, the activity of any financial or investment company is driven by a BIS which is developed and optimized individually and related with the objectives of each institution. ET allowed a BPM to be automatized. In this process, using an ATS has become today a necessity for any trading and investment systems, to permit a quick adaptability to the market conditions that are changing practically every second.

Using technology to analyze real-time price has allowed the development of a new type of trading based on mathematical and statistical algorithms: algorithmic trading (AT). Computer software generate automatically buy and sell orders without any human intervention to trade individual financial instruments. High-frequency trading (HFT) term "has gained some significant attention due to the flash crash in the U.S. on May 6, 2010" [6]. It is a specialized form of AT characterized by high turnover and especially by high order-to-trade involved ratios, generating a huge number of trades. We will see below the place occupied by AT and HTF through the ATS in the BIS of a financial or investment company.

2 Business Intelligence Systems using Automated Trading Software

It is well known that BIS comprises the strategies and informational technologies for data analysis. Once applied IT in the trading activity, the BIS that includes ET turned into a special architecture, as a result of the main role played by the ATS.

The real-time data analysis for prediction and risk management in the ET systems place ATS to be the main engine in the BIS of a financial or investment company. Because the profit is obtained directly from the actions of the ATS, the real-time data mining processes in ET have become the priority research subjects in the latest years. The low-latency data flux together with the fast speed data mining processes allows a very fast build for the buy and sell orders. ATS is linked directly with the liquidity and capital infrastructure to obtain the profit.

Even the real-time data are continuously stored in the data warehouse of the BIS together with the historical data, the real-time flux is used in parallel by the ATS to ensure the rapidity of the trading decisions and actions. After the data mining process and order execution, all resulting data are reversed in the data flow to be stored in the data warehouse. After that the BIS is aligned again with the usual configuration, all BIS results being obtained with the data already stored. However some reports are automatically generated by the ATS.

The particularity of the BIS with ATS is the special management of the real-time data flux to ensure a low-latency

functionality. The real-time data flux includes the price time series provided by data sources linked directly with the stock exchanges, real-time fundamental data and news provided by specialized external sources and the real-time capital and liquidity data provided by external sources linked directly with the bank and brokerage system.

The key of success using BIS with ATS in a financial or investment company, is how the real-time data flux is organized and used, how the real-time data mining process is made together with the risk and capital management and how fast can be sent and executed the trading orders in the brokerage system. In Figure 1. is presented the data floe of BIS with ET:

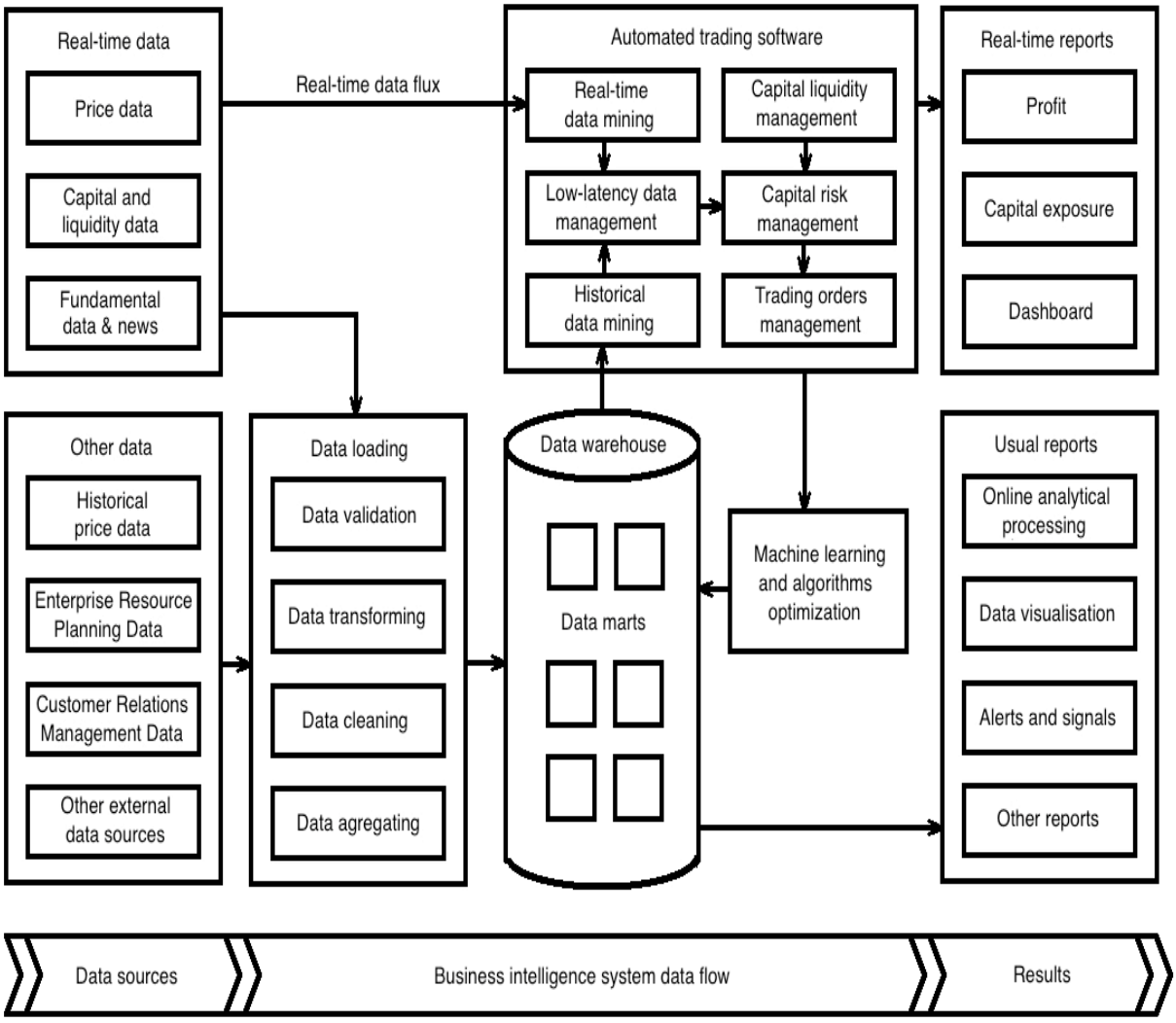


Fig. 1. Data flow of the Business Intelligence System using Automated Trading Software

It is well known today that "researchers believe the Cloud is a big part of the future of business intelligence" [7]. In the new time of the analytic data management of BIS, to solve the most problems regarding the complexity, costs and inflexibility, the cloud computing is a solution for majority systems. It is found that due to the specificity of information managed in BIS with ATS, there is a significant latency in the process of adoption of the cloud technologies in the companies using ET. There is no documented argumentation for this assertion except the personal experience in this field in the last ten years. Due to the specificity of the processed

data, due to the fact that the data mining procedures and the obtained results are strictly confidential and because of the speed and low-latency requested in the BIS with ATS, the cloud technology adoption is systematic delayed in many BIS in the financial and investment field. The classical "in house secured" BIS solution is still preferred.

2.1 Algorithmic trading

Algorithmic trading (AT) term refers to "any form of trading using sophisticated algorithms (programmed systems) to automate all or some part of the trade" [8]. Triggering from the notion of algorithm, as to be a finite set of precise instructions performed in a prescribed sequence to achieve a goal, especially a mathematical rule or procedure used to compute a desired result, the algorithmic trading activity ca be defined as to be the process of using computers programmed to follow a defined set of instructions for placing trades in order to generate profit at a speed and frequency that is impossible for a human trader. The main characteristics of the AT is the complexity and large computing volume ensuring in the same time a very high speed for the process of data mining, price prediction and building and sending the trading orders, which is also the intended purpose. Considering the different aspects of the processes involved in algorithmic trading, we can have some several categories of systems in this field:

- depending on the level of automation: semiautomated and automated trading systems;
- depending on the algorithm type used: statistical, neural or arbitrage trading systems;
- depending on the frequency of orders: long term, medium term and short term systems.

The long term trading systems are usual named as investments systems, being characterized by a small number of trades made to be profitable on long term (several weeks, even months). The short term trading systems are usual named and assimilated as to be high-frequency trading systems (HFT). These systems implied a very high number of orders, rapid order execution and cancelation, very short holding periods (seconds, minutes, hours), low-latency data and speed data-mining required and a specific focus on a high liquid equities [6].

2.2 Data mining in financial trading

In AT data is processed to find a profitable solution to buy or to sell an equity in a specified moment of time. The basic principle is very simple. Computers do everything using well-established algorithms. However, developing a stable and profitable system is a very laborious job, paying a very particular attention to the used algorithms.

Since statistics became widespread in the context of computer emergence, we can see several types of analytic tools: "descriptive analytics focus on reports of what has happened", "predictive analytics extend statistical and/or artificial intelligence to provide forecasting capability", "diagnostic analytics can apply analysis to sensor input to direct control systems automatically" and "prescriptive analytics applies quantitative models to optimize systems, or at least to identify improved systems" [9]. Considering the specifics of each methodology, "data mining includes descriptive and predictive modelling", "operations research includes all four" categories [9].

In financial trading we use descriptive modelling to analyse the past evolution of the price movement of an equity. This process is sometimes called "historical data-mining" (HDM) and the results help to understand the general trend and movement circumstances of a price. The HDM process use all available historical data in the warehouse (Figure 1.) and it is a significant part of the data management procedures in any trading system.

Predictive analysis which is involving different forecasting models in practically

present in majority fields of the human activity. In financial trading the predictive analytics is practically the main engine of each trading system. Using the real-time price data and the historical data, the predictive models try to determine the probability for a price increasing or decreasing. Because this process is using the real-time data flux, sometimes is called as "real-time data mining" (RTDM). The process is using the real time data without any pre-cleaning or validation filter and must provide a fast and reliable response to the low-latency data management module of the trading system (Figure 1.). In this process the algorithms used are very important but we know also that the "speed is nature of the scripting or programming language used" [10]. This means a very significant role in the process of design and development of a trading system the optimization of all used resources.

In the RTDM are included different types of mathematical algorithms for prediction. Some of them are using typical statistical functions together with different mathematical functions and price transformations. Classical algorithms based on moving averages or regressions are well known. However for a better efficiency, very complex algorithms are designed in this section and this is the key of success in any trading or investment company. Large amount of resources and effort are spent here, a better algorithm being what any investor want.

Other algorithms in RTDM are based on neural networks methods. On this chapter the algorithms try to use a continuous optimization to find the best way to the optimal solution. Usual this type of data-mining algorithms are combined with the statistical functions.

The third category in RTDM is the arbitrage type. These methods are based only on the real-time data price and are used in all trading procedures regarding equities with non-centralized price. Today is well known the huge interest in the arbitrage algorithms adapted for all cryptocurrencies, big difference between the sell and the buy price of a cryptocurrency in two different exchanges being a huge opportunity for profit. Important investment are made latterly in this field to develop algorithms and to reap all these opportunities.

3. Automated trading software

An ATS is a software which is receiving the real-time and historical price data of an equity, generates the signals for buying and selling of the equity based on well-determined algorithms, sets the volume of trading based on the capital liquidity and the capital a defined risk level, builds the trading orders and send them to the brokerage account without any human intervention.

The purposed objective of the ATS is to generate profit into the investor's account. Once established the algorithms and the functional parameters, the ATS will run continuous and automatically, being the main part of any automated trading system. The design of an ATS must take care about some specific requirements: a). low-latency data processing; b). fast trading orders processing; c). profitable trading strategies with measurable capital exposure level; d). automated capital risk management.

The low-latency data processing and fast data orders processing are required to ensure that the trading orders arrive in the brokerage account system before the price is changed. This is a defining requirement for ATS because an order with an obsolete price is ignored and will not be executed. Consequently the processing speed is the first design direction for an ATS.

The second design direction is the measurable risk involved by the ATS. Not all trading strategies are suitable, some of the strategies can not measure priori the risk involved. In an ATS can be included any profitable strategy with a measurable capital exposure level. With this parameter the capital risk management will be possible to be automatically made, in this

way the trading strategy will be integrated with the risk management procedures.

Receiving the real-time and historical data, ATS applies different data-mining methods to extract information and to calculate probabilities for an increasing or decreasing of the price based on different trading strategies. Depending on the time frame used for each strategy, there are several categories between high-frequency trading and investment strategies: very-short term, short term, medium term and long term strategies.

Once a buy or a sell opportunity was found, a trading signal is generated. The trading signal is a record that includes the action (buy/sell), the price and the code of the equity. This signal can not be yet executed because the volume of the transaction is still missing.

In the capital and risk management module, using the information about the available capital liquidity (C), based on the capital exposure level for each (i=1, 2, ... n) used trading strategy (ξ_i) we can calculate the volume of the current transaction (V_i). The total capital exposure level (R) of the ATS is the maximum limit of the exposure level:

$$V_i = C * \xi_i \text{ with } R \geq \xi_1 + \xi_2 + ... + \xi_n \quad (1)$$

In this way the total capital exposure of any ATS can be controlled to the maximal value (R). With an automated stop loss at that maximal value, the risk can be limited and controlled.

4. Conclusions

In the context of ET, the ATS occupies a central place in the BIS of any financial or investment company. ATS has become today a necessity in this environment, to permit a quick adaptability to the market conditions that are changing very fast. The computational and data processing speed is the first design requirements of any ATS.

Based on more trading strategies, applied on very-short, short, medium or long term, the ATS require a real-time price and capital liquidity data flux together with the historical price data. The data mining process, where the trading strategies are included, is organized in two individual modules: the real-time data mining module and the historical data mining module, depending on what data type are mined. The results are processed by a low-latency data management module which incorporates the trading signals. Based on the information provided by the capital and risk management module, the trading signals are transformed into trading orders and are sent into the brokerage account system. The speed for building and sending orders to the capital account is very important because the trading orders must arrive in the brokerage system before the price is changing, otherwise the obsolete price orders will be ignored and not executed.

The second design direction for an ATS is the measurable risk involved by the trading system. It is required by the automated capital and risk management process. This requirement is obtained by incorporating into the ATS only trading strategies with measurable capital exposure level. Knowing the capital exposure involved by each trading strategy, the total risk exposure involved by the ATS can be in real-time calculated, and therefore can be limited as needed. Together with the real-time capital liquidity informations it is possible to build an automated capital management process that will manage in real time the capital and the risk involved.

Because the profit in any financial or investment company is made directly by the actions of the ATS, the design and integration in BIS has gained a special attention. Thinking that we have only about ten-twelve years of ATS in the context of a history of hundreds of years of financial trading in the stock exchanges, we can say that we are still on the beginning in this domain, large research and improvements are still to come in the trading automation.

References

[1] V. Barbour, *Capitalism in Amsterdam in the Seventeenth Century*, University of Michigan Press, 1963, ISBN 978-0404613136, pp. 17.
[2] L. Petram, *The World's First Stock Exchange*, Columbia University Press 2014, ISBN 978-0231163781, pp. 9.
[3] Wikipedia - Electronic trading - Internet:https://en.wikipedia.org/wiki/Stock_market, January 29, 2018 [January 31, 2018].
[4] Wikipedia - Open outcry - Internet: https://en.wikipedia.org/wiki/Open_outcry, January 9, 2018 [January 28, 2018].
[5] Wikipedia - Electronic trading - Internet:https://en.wikipedia.org/wiki/Electronic_trading, December 1, 2017 [January 28, 2018].
[6] P. Gomber, B. Arndt, M. Lutat, T. Uhle, *High-Frequency Trading*, Business & Information Systems Engineering Journal, April 2013, pp. 6-15.
[7] S. Ouf, M. Nasr, *The Cloud Computing: The Future of BI in the Cloud*, International Journal of Computer Theory and Engineering, Vol. 3, No. 6, 2011, pp. 750.
[8] P. Treleaven, M. Galas, V. Lalchand, *Algorithmic Trading Review*, Communication of the ACM, 2013, vol 56, no. 11, pp. 78.
[9] D.L. Olson, D. Wu, *Predictive Data Mining Models*, Springer Science Business Media Singapore, 2017, ISBN 978-981-10-2543-3, pp. 5-6.
[10] J.O. Katz, D.L. McCormick, *The encyclopedia of trading strategies*, McGraw-Hill, New York, USA, 2000 ISBN 978-007-0580990-2, pp. 22.

Cristian PĂUNA graduated the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 1999 and also he is a graduate of the Faculty of Aircrafts of the Polytechnic University of Bucharest in 1995. He got the title of Master of Science in Special Aerospace Engineering in 1996. In the last decades he had a sustained activity in software development industry, especially applied in the financial trading domain. Based on several original mathematical algorithms he is the author of more automated trading software. At present he is the Principal Software Developer of Algo Trading Service Ltd. and he is involved as PhD student in Economic Informatics Doctoral School of the Academy of Economic Studies.

Using SVM in Classification

Ștefan PREDA
Economic Informatics Doctoral School
The Bucharest University of Economic Studies, Romania
spreda2003@yahoo.com

Machine learning technology is very often used in present days. Exist many machine learning models, the theory has evolved in last year's due exponential use of those technologies. A challenge is to apply those advanced technologies in practice to solve different problems or to improve systems.

1 Short theory about machine learning

Machine learning (ML) is a software that use previous data, or information, or experience it learn from the past and predict or make decisions about future data, making generalization of the existing data. In simple mode, ML predict something named output based on something named input. Input information is the known experience.

If output and input are labeled pairs, we have supervised learning [1]. Example of supervised learning is determining generated power of photovoltaic system based on weather parameters like solar irradiation level, temperature, wind speed, etc.

In case when input data is not labeled learning, process is an unsupervised learning. We can have an example in this case also from the same area, let's presume we need to classify the type of defects for a photovoltaic system based on some description of the status of components or elements that show component status, it is classification problem for determining specific groups. In [1] is mentioned also semi-supervised learning which use both approach of supervising and
unsupervised learning, we have input data labeled and unlabeled. An example in this case can be the same as previous example, but which use also for output defect classification also apart from status parameters also labeled parameters that have specific numeric values.

Input variables are named usual explanatory variables or regressors, manipulated variables [1], [2]. Output variables are named response variables or explained variable, dependent variables, measured variables. In our case response variable is generated power, explanatory variables can be: wind speed, temperature, ceiling, real feel temperature shade, real feel temperature, relative humidity, cloud cover, dew point, pressure, apparent temperature, wet bulb temperature, visibility, wind chill temperature, etc.

Sample data collection which is used for learning is named training set, collection which is used to test if estimated data is correct is named test set [1], [3]. Machine learning can do many tasks, for example:

- Classification, example ML determine if the share price will rise;
- Regression, example ML will predict generated power of a wind turbine based on wind speed.
- Clustering, putting observations in groups, for example putting mobile users in specific categories based on consuming pattern.
- Dimensionality reduction is the task when input dataset contain many parameters, this correlated with huge data set make data analyze difficult, to easy this, using dimensionality reduction, from entire set off parameters will be chosen most significant one.

2 About Support Vector Machine

2.1. SVM Classification

Support Vector Machine (SVM) is a ML that can be used for classification or regression. In [4] classification case methods start from input variables that can be assimilated like vectors {x1, x2, ..., xN}. Any of those vectors have features, for example for vector xj those are noted generic {aj1, aj2, …, ajm}. Like in real world we categorize things based on features here we try to find if based on the m features mentioned those vectors are in one class or the other(s). In simple way for two classes noted C1 and C2, the problem is to find if vectors xi pertain to C1 or C2. The easy way to separate data is using lines or hyperplanes. If our input variables xj can be separated by lines or hyperplanes, then those are linearly separable. After reading all SVM theory some time ago, I think linear way was first considered because it is easier to be mathematically explained.

If we have two input vectors, then the suggestive way to represent input data is graphically like below

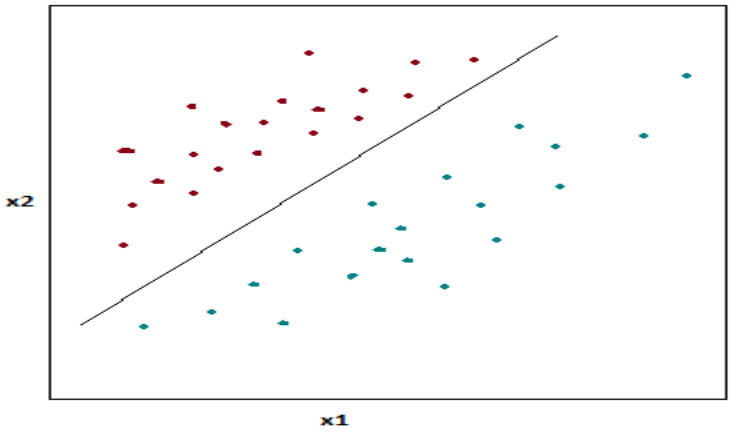


Fig. 1. Data separated by line in two-dimensional case

If we have 3 input vectors figure 1 will be three dimensional, and similar for n input vectors we will have "n" dimensional representation. In [5] the use of lines and hyperplanes in SVM classification is underlined more by using term linear classifiers. Basically, delimiting groups using lines or hyperplanes is straightforward and it can be computed mathematically.

We can observe that if data is separable or classified like in figure 1 between two groups there is not only a line, it is an area, which is named margin, like in Figure 2.

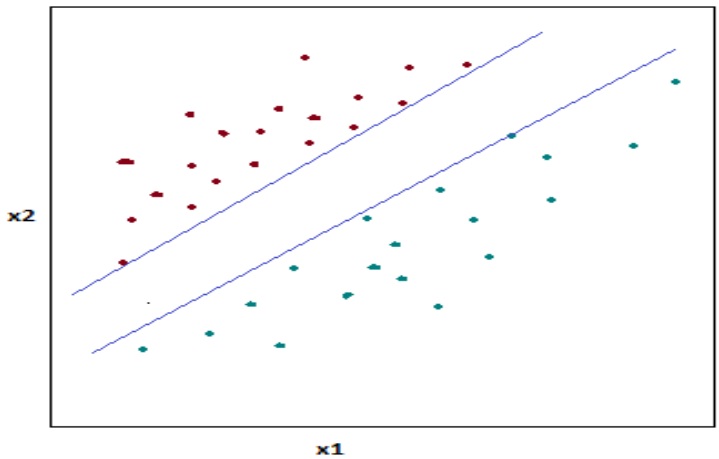


Fig. 2. Margin

Classification is better if the margin is wide.
The vectors that lies on edge lines or hyperplanes are named support vectors, from here the method name.
Generalizing for n vectors (n experiences, n input data) that is used in training stage of ML, the goal of SVM is to find a hyperplane that classify vectors in two classes, as there are many such solutions (or hyperplanes) the best one is that which assure the maximum size for margin.
To optimize this SVM use a function

$$S(x) = wT^* x + b;$$

where x – is an input vector
wT – is waiting vector, it is orthogonal to hyperplane and control hyperplane direction
b – is a scalar, a bias, it controls hyperplane position
For a specific input vector, SVM get that:

- It pertains to class c1 if $S(x) = wT^* x + b > 0$;
- It pertains to class c2 if $S(x) = wT^* x + b < 0$.

In real world the delimitation between objects is not always a line or a hyperplane, it can be other figure like nonlinear curves [5], [6], an for example for two-dimensional case:

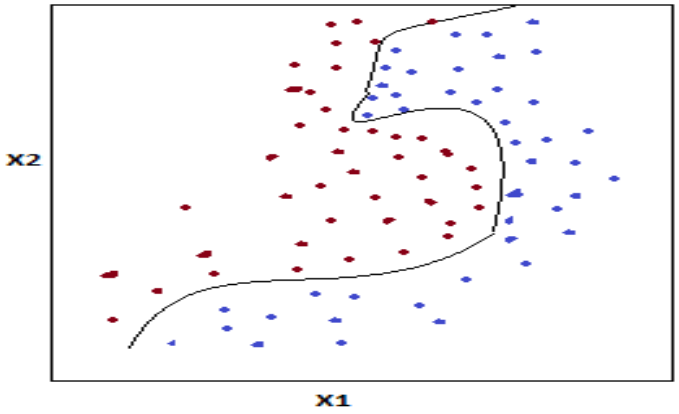


Fig. 3. Non-linear classification

For classification in this case SVM need a nonlinear approach, for this SVM use a kernel function, this map data into other space where classification can be done using hyperplanes [6]
There are many kernel functions, most used linear, polynomial, radial basis and sigmoid [6], [7], [8].

2.2 SVM regression

SVM can be used not only for classification, it can be used also for regression. A good and simple explanation for SVM principle used in regression is in [9]. Below figure describe it:

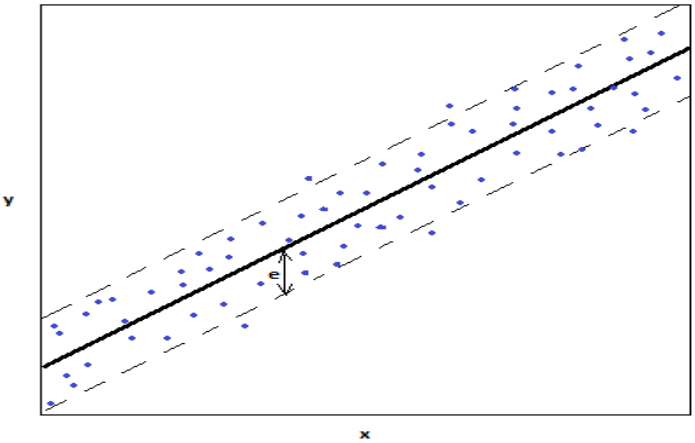


Fig. 4. Using SVM for regression

If in SVM classification principle is to have vectors or input data delimited by line (hyperplane) and to have a margin as big as possible, in SVM regression principle is revers, means to have vectors as much as near to the line (hyperplane), or to have vectors in a margin as narrow as possible. Margin size in SVM regression algorithm is controlled by a hyperparameter “e” as specified in figure 4.
SVM regression similar like classification can be linear or nonlinear. Also, nonlinear SVM regression have similar kernel functions: linear, polynomial, radial basis and sigmoid [10]

3 Example Using R for SVM classification in photovoltaic systems.

3.1 SVM with linear kernel

We will use data from a logger that is connected to a MPPT solar controller which is used in photovoltaic system. Target is to classify “BatteryStatus" of the battery bank based on "BatteryVoltageV", "BatterySOC" parameters all collected from logger which record data in csv or xlsx format. I will give some summary explanations in special for code that is related to SVM
First for loading necessary R libraries we use

```
library("e1071")
library("xlsx")
library("dplyr")
```

R SVM related functions are in "e1071" package
to load xlsx data logger we use:

```
setwd("C:/date")
print(getwd())
datalog2.frame <- read.xlsx("dateS2.xlsx",
sheetIndex = 1)
```

For training we will use 400 records from 476 records that are in data frame. From all data frame columns, we will use the following colomns: "BatteryVoltageV", "BatterySOC", "BatteryStatus". To fulfil those later we will use variables:

```
strain <- sample(476, 400)
head(strain, 5)
coloane <- c("BatteryVoltageV",
"BatterySOC", "BatteryStatus")
```

Data frame named datalog2.frame contain all parameters recorded by logger, because we will use in analyze 3 parameters we will extract those from data frame and create new data frame named datalog3.frame.

```
datalog3.frame <-
data.frame(datalog2.frame$BatteryVoltage
V, datalog2.frame$BatterySOC,
datalog2.frame$BatteryStatus)
names(datalog3.frame)[1:3] =
c("BatteryVoltageV", "BatterySOC",
"BatteryStatus")
head(datalog3.frame, 1)
```

For training we will we will create a new data frame named datalog3_train.frame it contain 400 records.

```
datalog3_train.frame <-
datalog3.frame[strain, coloane]
```

For test we will use the remaining 76 records from 476

```
stest <- datalog3.frame[-strain, coloane]
head(stest, 2)
```

To create SVM model we will use [12], [13], [14], [15]

```
mysvmlearn <- svm(BatteryStatus ~ ., data
= datalog3_train.frame, kernel = "linear",
cost = 0.1, scale = FALSE)
```

To plot classification, we will use [12], [13], [14], [15]:

```
plot(mysvmlearn, datalog3_train.frame[,
coloane])
```

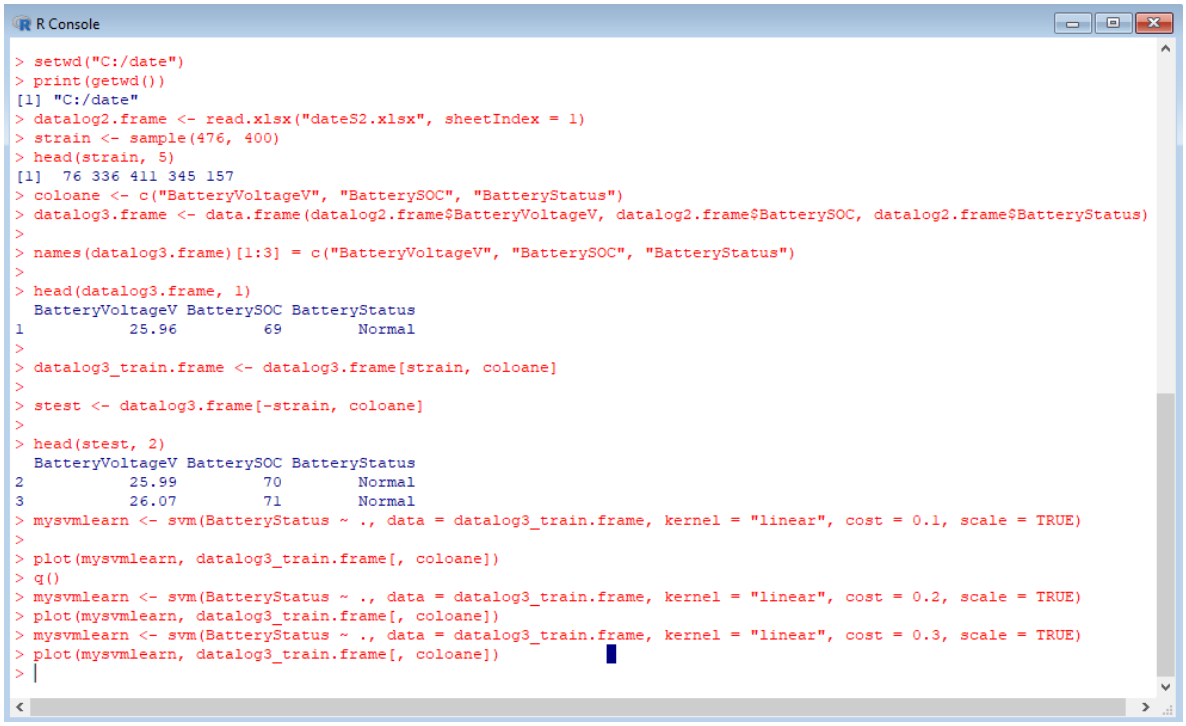


Fig. 5. Run SVM for classification

Based on [11], [12] when build model and use SVM function in R an important parameter is the cost. Cost parameter control margins and training errors. If cost is small this can generate big margins and make SVM susceptible to errors, revers if cost is big errors are less and margins are small.

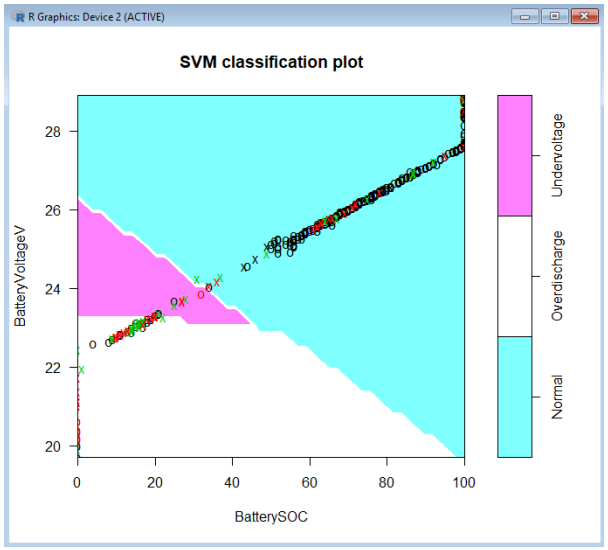


Fig. 6. Plot for cost=0.3

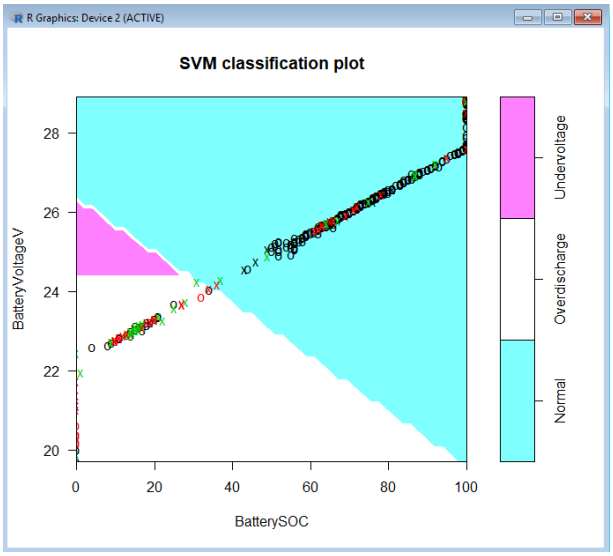


Fig. 7. Plot for cost=0.2

```
datatuned <- tune(svm, BatteryStatus ~ .,
data = datalog3_train.frame, kernel =
"linear", ranges = list(cost = c(0.001, 0.01,
0.1, 0.2, 0.3, 0.5, 1, 5, 10, 50, 100)))
summary(datatuned)
```

This will show:

```
> summary(datatuned)
Parameter tuning of ‘svm’:
- sampling method: 10-fold cross
validation
- best parameters:
 cost
 0.01
```

```
- best performance: 0.1128846
- Detailed performance results:
   cost     error dispersion
1  1e-03 0.2205128 0.06417731
2  1e-02 0.1128846 0.05589104
3  1e-01 0.1453205 0.04362853
4  2e-01 0.1528846 0.04157534
5  3e-01 0.1503846 0.04084294
6  5e-01 0.1478846 0.03814453
7  1e+00 0.1478846 0.03814453
8  5e+00 0.1528846 0.04321340
9  1e+01 0.1453205 0.04200667
10 5e+01 0.1428205 0.04404488
11 1e+02 0.1353205 0.04434103
```

We see the best performance is considered for cost=0.01, it has smallest error. However, we see also that for bigger cost like 50, 100 error decrease and also dispersion decrease, it means that tune function should be run repeatedly for other ranges that include this trend.

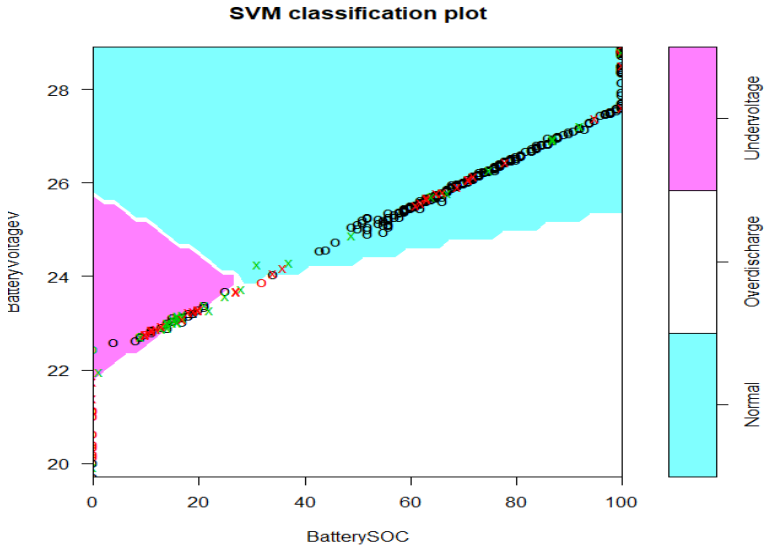


Fig. 8. Plot for cost=100

It is interesting to observe that for cost=100 not only that error and dispersion decrease but also classification is more appropriate by reality, means it show normal area to be area where BateryVoltageV is higher, which is good.

To predict we will use with predict function generated model SVM data frame, named mysvmlearn, and test data:

```
 predict.frame <- predict(mysvmlearn,
stest, type = "class")
```

To display standard confusion matrix, we will use

```
> table(stest[, 1], predict.frame)
  predict.frame
      Normal Overdischarge Undervoltage
  22.24      0           0           1
  22.74      0           0           1
  22.85      0           0           1
  22.87      0           0           1
  22.9       0           0           1
  22.91      0           0           1
  23         0           1           1
  23.04      0           0           1
  23.05      0           0           1
  23.12      0           0           1
  23.17      0           0           1
  23.25      0           0           1
  23.26      0           0           1
  23.73      0           1           0
  …….lines omitted ……..
  28.33      1           0           0
  28.34      1           0           0
  28.51      1           0           0
  28.85      3           0           0
>
```

The interpretation of confusion matrix in this case is smoothly.
For example, first line
22.24 0 0 1
There is 1 record that is classified to be Undervoltage, this seems ok like interpretation

For line
28.34 1 0 0
It means there is 1 record that is classified to be Normal, this also seems ok.

3.2. SVM with polynomial kernel

We will use SVM with polynomial kernel, in the same time with model build we will run tune function to get best parameters for model. Commands are:

```
polinomtune = tune.svm(BatteryStatus ~ .,
data=datalog3_train.frame,
kernel="polynomial", degree=c(3,4,5),
coef0=c(0.1,0.5,1,2,3,4))
> summary(polinomtune)
```

Response is:

```
Parameter tuning of 'svm':
- sampling method: 10-fold cross
validation
- best parameters:
 degree coef0
      3   0.5
- best performance: 0.09525641

- Detailed performance results:
   degree coef0      error dispersion
1       3   0.1 0.09782051 0.03455756
2       4   0.1 0.10544872 0.04612342
3       5   0.1 0.10044872 0.04647723
4       3   0.5 0.09525641 0.03296725
5       4   0.5 0.09782051 0.04183038
6       5   0.5 0.09794872 0.04862114
7       3   1.0 0.10794872 0.04495231
8       4   1.0 0.09532051 0.03901788
9       5   1.0 0.10294872 0.04562073
10      3   2.0 0.10282051 0.04009153
11      4   2.0 0.10532051 0.03892650
12      5   2.0 0.10544872 0.04612342
13      3   3.0 0.10532051 0.03709964
14      4   3.0 0.10532051 0.03892650
15      5   3.0 0.10288462 0.04208615
16      3   4.0 0.10532051 0.03709964
17      4   4.0 0.10538462 0.04097352
18      5   4.0 0.10294872 0.04246732
```

The best parameters for model are degree=3 and coef0=0.5
Running model for degree=3 and coef0=0.5 and representing graphic we have :

```
dataframePolinom=svm(BatteryStatus ~ .,
data=datalog3_train.frame,
kernel="polynomial", degree=3,
coef0=0.5)
plot(dataframePolinom,
datalog3_train.frame[, coloane])
```

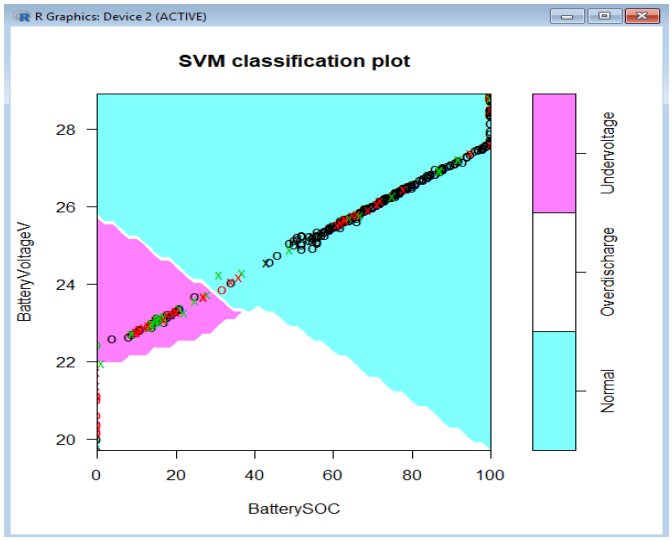


Fig. 9. SVM run with polynomial kernel

4 Conclusions

SVM was used in current example for a photovoltaic system to classify data related to battery banc, results obtained are verified in practice, we saw for example in the interpretation of confusion matrix for SVM with polynomial kernel that results match battery voltage and status from real life.

Also, SVM classification either linear or polynomial kernel, need tuning to detect best parameters that assure for model a good accuracy. In practice it can be necessary even to run tune function with many range of parameters such as to fine a combination of parameters that a re suitable for use case.

SVM classification results offer good graphical results that can be used in practice, graphs can be used to monitor systems in time. For example, from figure 7 can be created for data sets for very week and those can be compared, in this way we have information if something changes in system related to this component. Referring to this in [16] is specified that monitoring system by comparing graphs this is one of the best recommendation

References

[1] Raúl Garreta; Guillermo Moncecchi; Trent Hauck; Gavin Hackeling, scikit-learn : Machine Learning Simplified, Packt Publishing, Pub. Date: November 10, 2017, Web ISBN-13: 978-1-78883-152-9

[2] Henrik Brink Joseph W. Richards Mark Fetherolf, Real-World Machine Learning, Manning Publications, September 20, 2016, Print ISBN-13: 978-1-61729-192-0

[3] Scott V. Burger, Introduction to Machine Learning with R, O'Reilly Media, Inc., March 21, 2018, Print ISBN: 978-1-4919-7644-9

[4] Machine Learning, Mohssen Mohammed; Muhammad Badruddin Khan; Eihab Mohammed Bashier, CRC Press, August 19, 2016, Print ISBN-13: 978-1-4987-0538-7

[5] Jason Bell, Machine Learning: Hands-On for Developers and Technical Professionals, John Wiley & Sons, November 3, 2014, Print ISBN-13: 978-1-118-88906-0

[6] Himani Bhavsar, Mahesh H. Panchal, A Review on Support Vector Machine for Data Classification, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 1, Issue 10, December 2012

[7] Arti Patle, Deepak Singh Chouhan, SVM kernel functions for classification, Advances in Technology and Engineering (ICATE), 2013 International Conference, IEEE Xplore: 06 June 2013

[8] Hussein A Kazem, Jabar H. Yousif, Miqdam T Chaichan, Modelling of Daily Solar Energy System Prediction using Support Vector Machine for Oman, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 20 (2016) pp. 10166-10172 © Research India Publications.

[9] Aurélien Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow, O'Reilly Media, Inc., March 30, 2017

[10] Giuseppe Ciaburro, Regression Analysis with R, Publisher: Packt Publishing, January 31, 2018, Print ISBN-13: 978-1-78862-730-6

[11] Kuntal Ganguly, R Data Analysis Cookbook - Second Edition, Packt Publishing, September 20, 2017, Web ISBN-13: 978-1-78712-531-5

[12] Viswa Viswanathan; Shanthi Viswanathan; Atmajitsinh Gohil; Chiu Yu-Wei, R: Recipes for Analysis,

Visualization and Machine Learning, Packt Publishing, November 24, 2016, Print ISBN-13: 978-1-78728-959-8

[13] Cory Lesmeister, Mastering Machine Learning with R, Packt Publishing, October 28, 2015, Print ISBN-13: 978-1-78398-452-7

[14] Raghav Bali; Dipanjan Sarkar; Brett Lantz; Cory Lesmeister, Unleash Machine Learning Techniques, Packt Publishing, October 24, 2016, Web ISBN-13: 978-1-78712-828-6

[15] Cory Lesmeister, Mastering Machine Learning with R - Second Edition, Packt Publishing, April 24, 2017, Print ISBN-13: 978-1-78728-747-1

[16] IEA International Energy Agency, Photovoltaic and Solar Forecasting: State of the Art, Report IEA PVPS T14-01:2013.

Stefan PREDA graduated the Faculty of Economic Cybernetics, Statistics and Informatics, with a bachelor degree in Economic Informatics in 2013. In 2015 he got his master degree from the same faculty of the Bucharest University of Economic Studies, specialization in Databases Support for Business. Currently he is working on Oracle Corporation, like Senior Principal Technical Support Engineer, software analyst, in Fusion Middleware, EMEA Identity Management team.

Semantic databases: a smart perspective for today innovation in e-business

Georgiana STĂNESCU (NICOLAIE)
PHD Student at Doctoral School in Economic Informatics
Bucharest University of Economic Studies, Romania
georgianastanescu17@gmail.com

The purpose of this article is to demonstrate that the e-business domain needs from today, have helped in the apparition of a semantic database, which is more reliable for business purpose. It is already a fact that we live in a digital era where society needs a more usable and accessible web environment. Looking back, we can see that the emergence of the World Wide Web in 1989 had a constructive effect on society. As the world is on a continuous change and the volume of data which became available on-line is increasing over time, there has been a need to introduce semantic technologies, which will make the web domain more efficient. The Internet aspect and semantic technology have helped to publish online all the information contained by an e-business, making it accessible from all over the world without limitation barriers.

Keywords: *semantic database, web, semantic technology, e-business*

1 Introduction

If we look in the past, we can see that the emergence of the World Wide Web has had a constructive effect on our society. Numerous facilities, activities, which until recently were considered traditional, are now available in the web with the help of the current semantic technology. Today, we can automate and facilitate the exchange of information between people without much effort.

Before the appearance of semantic technologies or even the World Wide Web, the e-business concept was already a part of our lives. Everyone was making business as a part of everyday leaving, but in the traditional way.

The appearance of the internet and semantic technology, helped in publishing online all the information an e-business have and make it accessible from all over the world, breaking the limitation barrier.

We can easily see that the semantic web is like an addition to today's World Wide Web and helps people to distribute content without any boundaries from websites or applications. From all the numerously mentions of semantic web we can highlight the following descriptions: 'the Semantic Web is an extension of the existing World Wide Web. It provides a standardized way of expressing the relationships between web pages, to allow machines to understand the meaning of hyperlinked information.'[1]

The semantic web was the one which initiate the creation of semantic technologies, along with all its languages and the related applications. In our days, the idea of semantic technology has greatly spread, being successfully adopted, both nationally and globally.

Along with all mentioned semantic qualities, we must find a way to sustain the relation between information in a business. The constant growth of web environment demands a necessity for new services and used methods that satisfy the user modern needs.

2 The components of semantic technologies

The semantic technologies are based on more than 10 years of researches and are like an extension for web service, along with all its improvements. With the help of semantic technologies, the static web is

evolving to a much friendly user perspective.

The word 'semantics', derived from Greek, is the meaning and signification of some notions. The field of semantics is very complex and abstract, linking knowledge and implementation. Thus, when it comes to semantic technologies, we refer to a suite of languages, accumulated in a single work area that facilitates the integration of business in the web.

The first mention of semantic technologies was in 2001, in an article by Tim Berners-Lee, who stated that ‘semantically the web is not a separate web, but an extension of the current one, where the information is well defined and allows for better cooperation between people and computers. ‘[2]

As far as we know, semantic technologies offer us 'a new approach to managing information and processes, the fundamental principle of which is the creation and use of semantic meta-data'. [3]

What we should know about semantic technologies is that they simplify the relations and the communication between the programs. The semantic technologies help in connecting all the data, has a better integration in the systems, the implementation is much easier and understandable.

According to an article [4] from Economic Informatics magazine, no. 3 (35)/ 2005, written by Sabin Corneliu Buraga, the 'architecture of semantic web is a functional one, because the constitution of the specification is based on its incremental some languages, starting from the lower level (ie the metadata) and reaching higher levels (eg, logic level).'

From the following figure we can see that the most useful and important semantic technologies in the web environment are eXtensible Markup Language (XML) and Resource Description Framework (RDF), but also there can be mentioned the followings: SPARQL(Simple Protocol and RDF Query Language), OWL (Web Ontology Language), RDFa (Resource Description Framework attributes), JSON-LD (JavaScript Object Notation for Linked Data), SKOS (Simple Knowledge Organization Syatem), RDFS (Resource Description Framework), GRDDL (Gleaning Resource Descriptions from Dialects of Languages), HTML (Hyper Text Markup Language), CSS (Cascading Style Sheets), HTTP (Hypertext Transfer Protocol).[5]

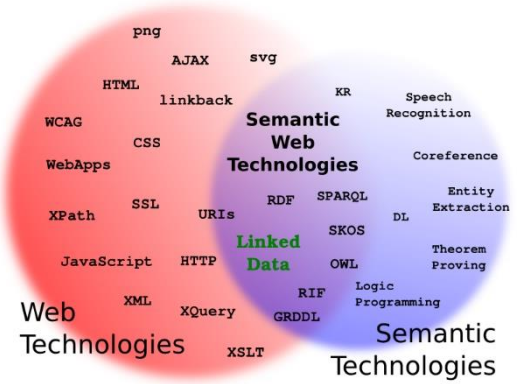


Figure 1. The components of semantic technologies[5]

All of the mentioned languages are a part of semantic technologies and which facilitates the relationship between data and sharing across applications.

The architecture of the web is based on combination of those and the main facility is that the content of a page is easily updated or changed without the necessity of changing the main schema.

They have the role to improve existing web, not by making a change, but by keeping the initial content and adding useful information in the web structure.

With the use of semantic technology, the old way of organizing information and important data is simpler. The web environment has evolved from static way of listing information, to a more interactive and user-friendly way. Also, according to Tim Berners Lee, 'the web was designed as an information space, with the goal that it should be useful not only for human-human

communication, but also that machines would be able to participate and help.'[6]
We cannot mention the technologies used by the semantic web without reference to the pyramid created in 2000 by Tim Berners-Lee.

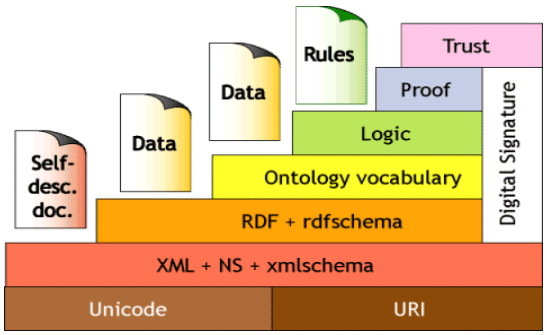


Figure 2. The pyramid of semantic technologies [7]

It details the protocols and challenges behind the emergence of the semantic web, as we can see in the following figure.
At the syntactic level, XML (eXtensible Markup Language) is used. It allows structured web content, accessibility to information is easier due to the way data is stored. XML is very widespread in the web domain and is often used to transfer data over the web.
The simplest example of using XML is the presence on the web of a service provider who offers their services through the internet and wants to be contacted by potential customers.
The next element in the pyramid is the Resource Description Framework (RDF), which appeared in the continuation of XML, to represent metadata in the World Wide Web. Data from the semantic web uses RDF to store the data of a web page modification, content and author details.
RDF, used together with XML, has the role of improving the existing web, not by modifying it, but by adding useful information to the structure of the already existing web, being some of the most used semantic technologies.

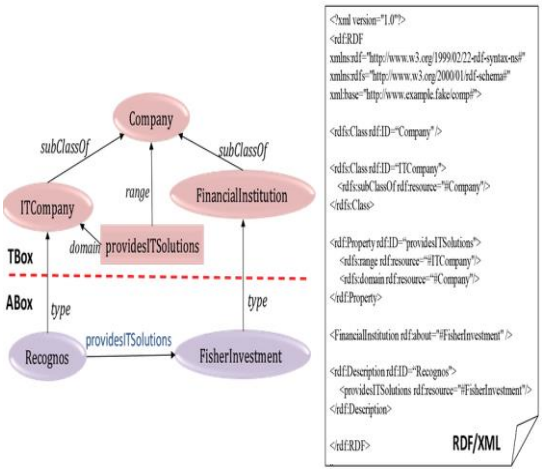


Figura 3. RDF syntax example [8]

The other languages listed: OWL, RDFa, RDFS, etc., are specially designed to help develop complex and rich information sites as well as the relationships between data.
The components of the listed semantic technologies are used to implement and design applications that serve an e-business.
In the article 'Considerations on the Use of Semantic Web Technologies in the Context of e-Business Applications' [4] written by Dr. Sabin-Corneliu Buraga, he defines the use of semantic technology components in application development and data processing of an e-business.
He classifies the use of used language on levels, all from metadata and advancing to the logical level, offering numerous possibilities for the requirements of an application, as follows:
- metadata is based on the RDF language 'which provides the general framework for expressing simple semantic statements' [4];
- schemes, 'provides the possibility of specifying simple ontologies to define a hierarchical description of concepts and properties' [4];
- the logical level includes complex languages, which help to make ontologies for the semantic web.[4]

3 Semantic technologies: where are we now?

The demands of society have led to technological developments, which have helped create a more accessible and usable web environment. It is already a fact that we

live in a digital age where society needs a more usable and accessible web environment.

Today, there is a need to introduce semantic technologies in all the domains, starting from IT, banking, business, etc., which will make the web domain much more efficient. This will simplifie the relationship between existing software products.

Online users rely on the available offers found in a simple search on the internet for the products or services they want. Semantic technologies provide users with a search process based on product or service attributes that best fit their needs, providing relevant information in a timely manner. Selling online products is more efficient with the implementation of semantic technologies.

Tim Berners-Lee claimed that the semantic web is 'a new form of web content that is significant for computers and unleashes a revolution of new possibilities.'[10]

From the following table we can see that semantic technologies are based on more than ten years of research and evolution. The new web 3.0 is viewed as an extension for the current web service, along with all the improvements it assigns.

Table 1. Evolution of web

Used technology	**Evolution of semantic technologies**
1980-1989 - Windows and PC era - Some basic scripts and web technologies - FTP	- Beginning of computer era
1990-2000 - World Wide Web - Servers - SQL databases - HTML - HTTP	Web 1.0: - Read only web - Non-interactive - Limited web access - Focus on content - Networking
2000-2010 - PHP - JAVA - XML - FLASH - RDF - Media sharing - Social websites - Multimedia - Cloud computing	Web 2.0: - Read and write web - Social software - User interactive content - Sharing knowledge - Interactive sites - Focus on user needs
2010 – Today - Semantic databases - Semantic search - RDF/RDFS - OWL	Web 3.0: - 'Intelligent web' - Semantic technologies - Media technologies - Virtual reality - Big data - More dynamic - Information control - User oriented platforms

If we look closer in the evolution of web table, we can see that everything starting from the bases of web 1.0 and continuing with the evolution of the web 2.0 is evolving step by step. The current web 3.0 is viewed as an extension of the other two, with an impressive evolution and adaptability among businesses.
Web 3.0 attributes numerous improvements to the virtual environment by using semantic technologies, evolving in a more user-friendly way. From day to day, information available online increases as a number, which leads to many unstructured data and a large amount of information in web content.

3 Semantic databases management in e-business

The structured data, which we are already used to, will be much accessible with the used of relational databases. Using the semantics, the business information stored in a database can be used in many ways, besides simple search queries or storing information.
HyperText Markup Language (HTML) is used to build a website to ensure information creation and structuring. This language has been used since the emergence of the web domain, but it is static and provides only a way to present the information. It is very widespread, being used in most web pages due to browser compatibility and reduced integration complexity. 'Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data'[11]
One of the big challenge from our days is the need to share and manage the volume of data stored in an e-business. For this, the database should have a semantic approach and make use of SPARQL(Simple Protocol and RDF Query Language), OWL (Web Ontology Language), OIL (Ontology Inference Layer) and RDF (Resource Description Framework) atributes.
For recording data, the databases are using RDF, which has a graph-based model and for describing the relations attributes and creating a set of values, the OWL language is used.
Today the basic architecture for semantic web is easy to understand, as we can see from the following figure, 'the concepts of ontology and of rules are the basis of the data integration'.[12]

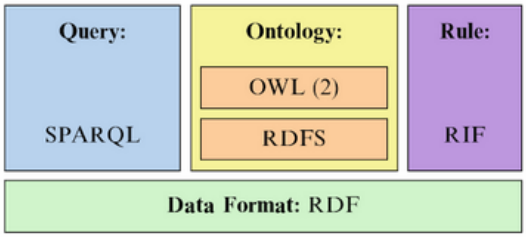


Figure 4. Basic semantic web architecture [12]

The difference between the usual databases and the semantic ones can be seen from the use of semantic query language SPARQL. According to [13], 'semantics is a data model that focuses on relationships, which adds contextual meaning around the data so it can be better understood, searched and shared. By using semantics, leading organizations are integrating their data faster and easier and building smarter applications with richer analytic capabilities.'[13]
To understand better the use of semantic databases in an e-business domain, I have created the following structure, which is an overview of all the important factors which are making use of semantics. The schema describes references and also the relationships between them. A database should always correlate values with the specific references.

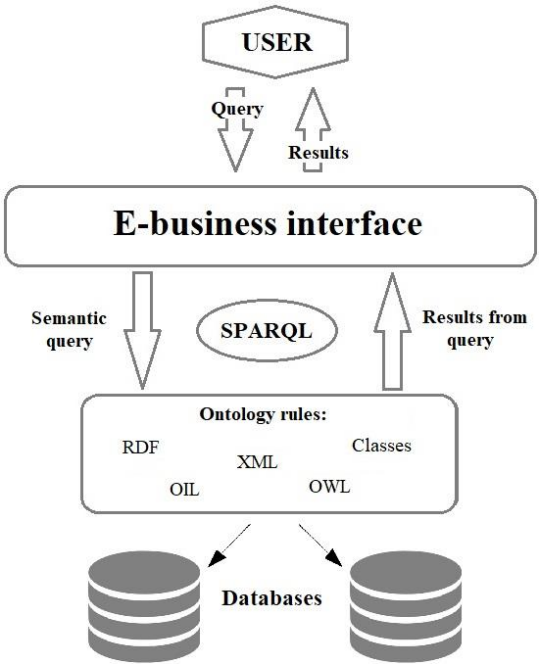


Fig 5. Semantic databases overview

A user will make a simple query, related to his needs, in the interface and expects relevant results. Between the e-business interface and the databases, there must be a SQL query, but as we use the semantic databases, we must use also a semantic search query language, which is SPARQL.

According to article Database-to-Ontology Mapping Generation for Semantic Interoperability, 'when an SPARQL query is received by a data provider service, it is translated to an SQL query using the mappings between the database and the local ontology. The SQL query is executed in the database and its result is encapsulated as an SPARQL response and returned to the query web service. The query web service then collects the responses returned from data provider services and recomposes them in one coherent response which will be sent to the visualization web service.'[14]

If a database is based on semantic relational data and is making use of ontologies, this kind of query can extract more information from a database, than the one based only on simple SQL language.

Also, 'when a query is submitted to the system, it is analyzed by this service and decomposed into a set of modular queries
s. Then using the mapping directory in the knowledge base, the query web
service redirects the single queries to the suitable data provider services.'[14]

The old way of organizing data and information becomes more advanced with the use of semantic technologies

We can say that 'by exposing the semantics of the data, machines can then utilize the information in more interesting ways than just storing it or displaying it.' [11]

4 The advantages of using semantic databases

Is well known that semantic database technologies deliver a lot of benefits, that businesses of all types should be making use of.

At high level, semantic databases offer five main benefits[15]:

- they work with your existing relational databases;
- they align with Web technologies;
- their underlying technology speeds integration of multiple databases;
- they're based on data structures that are flexible by design;
- they can help enterprises big data challenges.

A business model is constantly changing due to the dynamic environment of the business market. When the market changes and the business model has to adapt to maintain the viability of the company and, of course, the level of profit.

Using semantic technologies in existing business models, customers can access the information they need for products or services more quickly, purchase products, pay online, etc.

By integrating semantic databases into the chosen business model, companies can have a broader view of the products or services they offer.

'By preserving semantic structure, we can query the database at different levels of semantic meaning, from very specific to

very general. When we query the database, we don't just get back a list of records – we get back fully "rehydrated" semantic types'. [11]

5 Example of business which are making use of semantic technologies

Semantic technologies are scalable, leading to application in all areas of an e-business. Analyzing the current state of use of semantic technologies in e-business, one can notice that large software companies such as IBM, Oracle, Microsoft have already begun to introduce semantic technology into their platforms.

The article 'Advanced Technologies in e-Tourism' [16] highlights the application of semantic technologies in e-tourism. This branch of e-business is the ideal environment for application of semantic technologies, due to the dynamic information that must be contained: maps, tourist routes, accommodation and transport information, etc.

By integrating web technologies, a better search of information, better promotion of tourist packages and services is made. Typically, tourists search for vast information without a well-defined context, so semantic technologies filter information as useful as possible, providing relevant information. The application of semantic technologies in e-tourism is constantly developing due to the diverse level of information.

Another model of use of semantic technologies would be in the field of e-learning. The way of learning has evolved considerably, anyone can access the data needed to learn from anywhere and from any mobile device. In the article 'Evolution of eLearning based on web 3.0 and semantic technologies' [17], it is emphasized that the semantic web revolutionized the learning environment, access to education being quick and diversified, depending on the needs of users. The e-learning field has evolved with the help of semantic technologies from the static to interactive stage, now including various services, such as downloading course documents, uploading projects, interactive tutorials, evaluating online knowledge, etc.

Another example of using semantic technologies is the supercomputer created by IBM, called Watson, 'combining artificial intelligence and sophisticated analysis software for optimal performance as a machine for answering questions.' [18]

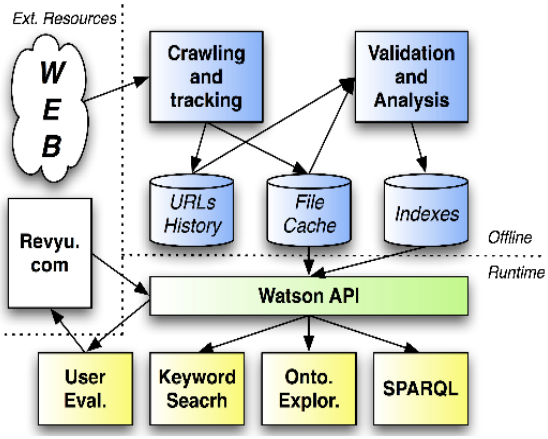


Figure 6 – Watson architecture[18]

With the help of semantic technologies, in combination with today's large data volume, Watson has access to an impressive amount of information and develops methods to 'structure search based on ontological relationships.' [18]

There are lots of systems that use semantic technologies as a way of searching, but the Watson concept is different. We can say that, 'Watson is the only tool that provides the required level of service, for applications to dynamically exploit web semantic data.' [19]

The Watson architecture helps to gather semantic content available on the web, analyze it in detail, and find ways to easily access the data found.

A more recent integration of semantic technologies would be in cloud computing, according to the article 'Semantic Technologies for Enterprise Cloud Managemen' [20]. The cloud service appeared with Web 3.0 and is the way that the storage space has moved into the web. It

works like an internet computer and provides the ability to access information stored anytime and anywhere. Large software companies have entered the business market with data storage capabilities on virtual servers.

6 Conclusions

In my opinion, the e-business domain is about the strategy of a business and the improving the efficiency with the use of semantic technologies. Implementing, adding or changing something in e-business is easy with semantic technologies. All the e-business components combined with a web interface, results a very strong and sustainable company, improving the hole business process.

The semantic technologies can shape the virtual space and make it more affordable and reachable for e -business domain. The world is evolving along with the semantic technologies and the e-business domain gets all the advantages and benefits from implementing all the technology in the needed tools.

The semantic technologies offer better interoperability of information and better integration, once it is used in e-business domain.

The internet aspect and semantic technology have helped to publish online all the information of an e-business, making it accessible from all over the world without limitation barriers.

More and more companies adopt global semantic technology, to capitalize on the opportunities it offers. At the local level, e-business offers small business development opportunities and semantic technologies are being implemented.

The importance of this article is amplified by the use of semantic technologies in the databases used in an e-business.

Thanks to semantic technologies, we can see that the e-business field takes full advantage of all the advantages offered by the web.

Once a business is online, it is much easier to promote its interests and brings a wealth of benefits to the company.

Semantic databases are successfully adopted and represent an important part of business models, participating sustainably in their development and integration.

By using the databases described above, we obtain an appropriate basis for tools that enable understanding, sharing and communicating, changing, measuring, and simulating business models.

References

[1] *Semantic web*. Available at: http://semanticweb.org/wiki/Main_Page.html.

[2] Cody Benson, *Introduction to the Semantic Web Vision and Technologies*. Available at: http://www.semanticfocus.com/blog/entry/title/introduction-to-the-semantic-web-vision-and-technologies-part-1-overview/, 2007.

[3] John Davies, Rudi Studer, Paul Warren, *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. Accesible at: https://books.google.co.uk/books?hl=ro&lr=&id=ePCrymWkcl0C&oi=fnd&pg=PR1&dq=semantic+web+technologies&ots=wvs6oA054W&sig=-ebY7DPZcS3URb9JePmrOE0jTc0&redir_esc=y#v=onepage&q=semantic%20web%20technologies&f=false; 2006.

[4] Sabin-Corneliu BURAGA, *Considerations Regarding the Use of Semantic Web Technologies in the Context of E-business Applications*. Article published in articol publicat in Economic Informatics Magazine, nr. 3(35)/2005.

[5] *Want a semantic web/ linked data job?* Available at: http://afterglowlee.blogspot.co.uk/2011/07/want-semantic-web-linked-data-job.html

[6] Greg Ross, *An Introduction to Team Berners-Lee's Semantic Web*. Available at: http://www.techrepublic.com/article/an-

introduction-to-tim-berners-lees-semantic-web/, 2005

[7] *From Data Federation Pyramid to the Semantic Web 'Birthday Cake'*. Available at: http://www.mkbergman.com/231/from-data-federation-pyramid-to-the-semantic-web-birthday-cake/

[8] Alina Dia Miron, *Semantic web scurta introducere*. Available at: http://www.todaysoftmag.ro/article/82/semantic-web-scurta-introducere

[9] Tim Berners-Lee, *The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*. Article published in Scientific American, May 2001

[10] Mark Clifton, *Semantic Database: Concept, Architecture and Implementation*, Available at: https://www.codeproject.com/Articles/832959/Semantic-Database-Concept-Architecture-and-Impleme, 24 October 2014

[11] Sven Groppe, *Data Management and Query Processing in Semantic Web Databases*, Available at: https://books.google.co.uk/books?hl=en&lr=&id=HdVZ2MhzP_4C&oi=fnd&pg=PR3&dq=semantic+databases&ots=KXmqlEMtun&sig=wFex7oniKC98qMvKC248wJwjgBM#v=onepage&q=semantic%20databases&f=false, 2011

[12] *Gain context for your data with semantics.* Available at: https://info.marklogic.com/semantics

[13] Raji Ghawi, Nadine Cullot, *Database-to-Ontology Mapping Generation for Semantic Interoperability*. Available at: file:///C:/backup/virtual/de%20pe%20desktop/docs/InterDB07-Ghawi.pdf, September 23-28, 2007

[14] *Semantic Databases: Destiny Or Distraction?* Available at: https://www.informationweek.com/database/semantic-databases-destiny-or-distraction/d/d-id/1107587, 29 November 2012

[15] Pavel Cech, Vladimír Bureš, *Advanced Technologies in e-Tourism*. Available at: http://www.wseas.us/e-library/conferences/2009/genova/ACS/ACS-12.pdf

[16] Fayed Ghaleb1, Sameh Daoud1, Ahmad Hasna2, Jihad M. ALJa'am2, Samir A. El-Seoud3 si Hosam El-Sofany. *E-Learning Model Based On Semantic Web Technology*. 15 May 2006

[17] *IBM Watson supercomputer*. Available at: http://whatis.techtarget.com/definition/IBM-Watson-supercomputer

[18] Jérôme Euzenat, INRIA Grenoble Rhône-Alpes. *Watson, more than a Semantic Web search engine*. Available at: http://www.semantic-web-journal.net/sites/default/files/swj96_1.pdf

[19] Peter Haase, Tobias MathaB, Michael Schmidt, Andreas Eberhart, Ulrich Walther. *Semantic Technologies for Enterprise Cloud Management.* Available at: http://www2.informatik.uni-freiburg.de/~mschmidt/docs/iswc10.pdf

Georgiana STĂNESCU (Nicolaie) graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Bucharest University of Economic Studies (ASE) in 2010. After that she obtain a Master Degree in Multimedia and Audiovisual Production from Faculty of Journalism and Mass Communication, in the Master of multimedia and audiovisual production, Bucharest (Romania). In present she is a PhD student in Economic Informatics at Economic Informatics Doctoral School, Bucharest University of Economic Studies (ASE), specialization: semantic technologies applied in e-business.

Data Model for Electricity Consumption Management

Iuliana ȘIMONCA (BOTHA) [1], Simona Vasilica OPREA[1],
Adina UȚĂ[1], Ion LUNGU[1], Osman Bulent TOR[2]
[1]The Bucharest university of Economic Studies, Romania
[2]EPRA Engineering Procurement Research Analysis, Turkey
iuliana.botha@ie.ase.ro; simona.oprea@csie.ase.ro; adina.uta@ie.ase.ro;
ion.lungu@ie.ase.ro; osman.tor@epra.com.tr

The power management system targets three major sectors: energy production, market trading and energy consumption. Energy consumption management is based on optimization models and methods that are implemented with actual system data. In order to identify these methods we have analyzed a series of related papers on optimizing energy consumption.
As a result, in this paper, we present the improvements that we made on the data model for electricity consumption approach from SMARTRADE project.

Keyword: *electricity consumption, renewable energy, optimization methods*

1 Introduction

The management of electricity consumption implies making the right decisions for energy consumption optimization in the context of demand-management mechanisms. All consumption optimization models and methods use industry-specific data. In this context, an important role in the management of electricity consumption has its way of organizing the data. It is very important that the data model contain accurate, consistent, relevant data, real-time links and fast access to real-time decisions.

2. Related works on energy consumption optimization

Energy consumption management is based on optimization models and methods that are implemented with actual system data. In order to identify these methods we have analyzed a series of related papers on optimizing energy consumption.
In the paper [1], the authors present an analysis of models and methods for energy consumption optimization for residential users in the context of demand-management mechanisms, which are used to reduce peak demand or to avoid emergency cases.
In the paper [2] the authors make a review of various concerns about optimization methods and highlight the fact that since the creation of a renewable energy structure entails very high costs in terms of design and long-term planning of energy systems, it is very important to be able to select the best alternative between different renewable energy systems.
An important issue is the planning of renewable energy systems at Community level, which consists in formulating local energy consumption policies, developing energy resource and energy allocation patterns, economic development and energy structure as well as analyzing the interactions between economic cost and safety in power supply.
In order to obtain solutions to ensure the allocation of energy resources / services and the expansion of the desired capacities, the minimum system cost, maximum system reliability and maximized energy security, some authors have opted for the application of interval linear programming, chance-constrained programming, and mixed integer-linear programming.
The paper [3] presents a non-linear programming model developed for a typical factory, which presents the optimal values and times for the electrical energy

consumption based on doing activities and operations in factory.

In [4] it is developed a distributed optimization method by which they calculate optimal consumption levels in an iterative manner and the intermediate results obtained are used by smart meters to calculate probability distributions of start times for tasks. Using these distributions, smart meters can quickly plan their tasks.

The article [5] analyzes the minimization of power consumption in wireless systems where nodes operate on battery and [6] makes a review on optimization modeling of energy, as Deterministic Optimization Modeling, Inexact Optimization Modeling, Model-Based Decision Support Tools.

The studies [7], [8] and [10] propose multi-objective optimization methods for developing energy systems using primary energy consumption and initial cost. [9] proposes a long-term dynamic multi-objective planning model for distribution network expansion along with distributed energy options which, using an immune genetic algorithm-based (I-GA) algorithm, optimizes costs and emissions by determining the optimal schemes of sizing, placement and dynamics of investments on distributed generation units and network reinforcements over the planning period.

The paper [11] analyzes and synthesizes energy-optimization methods using either cloud-based video surveillance systems or video surveillance systems in general. The paper highlights the limitations of the application of the methods in the cloud video surveillance system due to the architecture and the energy consumption characteristics of the system.

In the dissertation thesis [12], the author develops a method to minimize the energy consumption of production equipment using a mathematical model for multi-objective problems on single or parallel machines, to minimize energy consumption in production planning. The method uses a genetic algorithm, greedy randomized adaptive search procedure and meta-heuristic hybrid approaches to obtain sets of reasonable approximate solutions in reasonable time used in decision-making on production planning.

3. Electricity management system architecture

The power management system targets three major sectors: energy production, market trading and energy consumption.

The informatics prototype will be developed in a modular manner in order to support a flexible and scalable implementation. Thus, the proposed models (Figure 1) are:

- Electricity Consumption (EL) that aims to provide a demand response solution with load forecast (LF), load profiles (LP) and load optimization (LO) models;
- Electricity Generation (GEN) that aims to increase the actual forecasting accuracy through generation forecast (GF) model and also to provide a better generation management through generation optimization (GO) model;
- Market Trading (MT) will implement market trading models (MTM) and simulations (MS) to provide a decision support system for market trading activities;
- Analytics & BI (ABI) aims to provide a friendly, interactive and flexible interface for analysing the Key Performance Indicators (KPI) by accessing a web portal with reports, charts and dashboards.

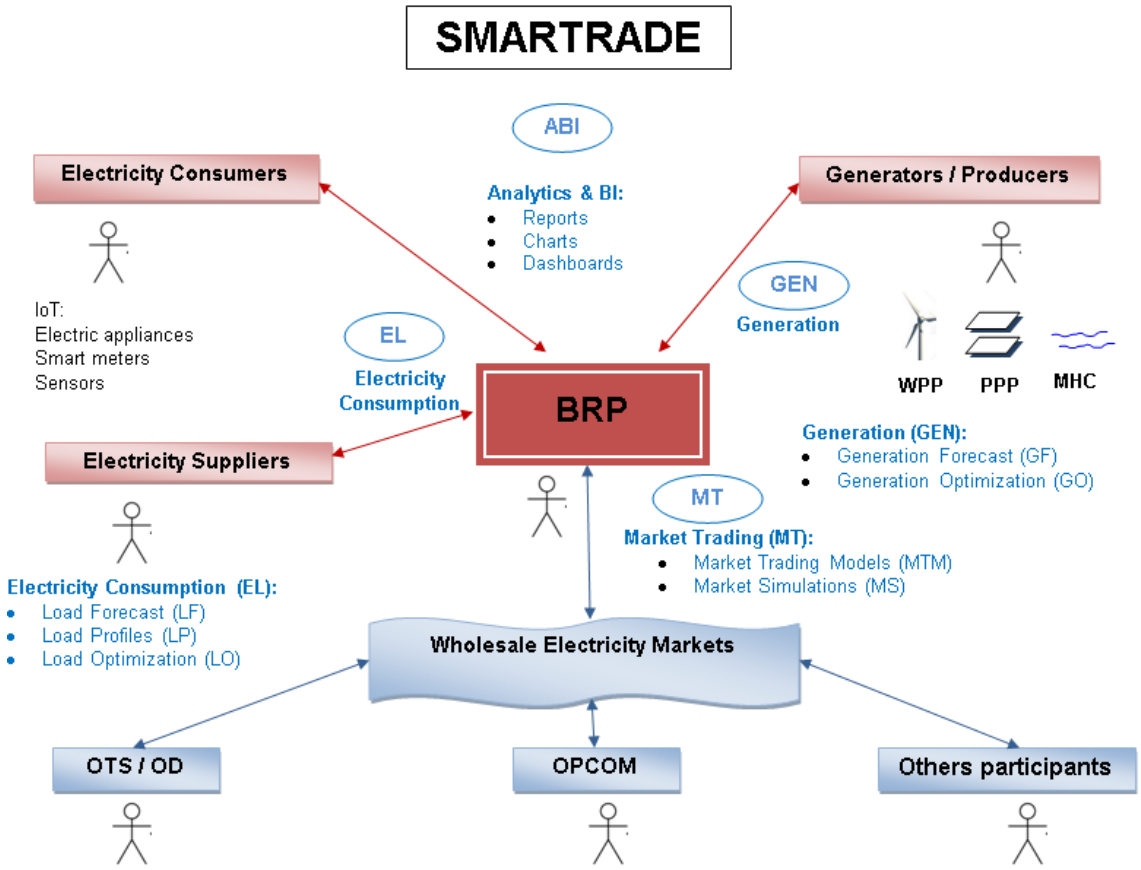


Fig. 1. Electricity management system

We designed the systems' architecture based on three levels: data, models and interfaces. The complete architecture is represented in Figure 2

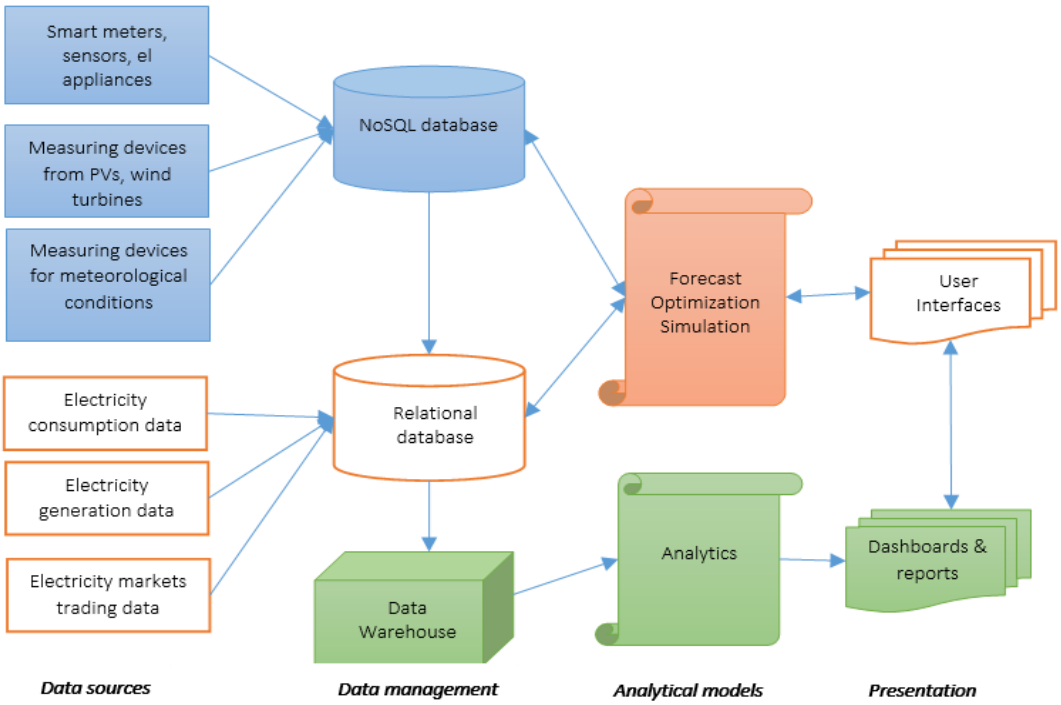


Fig. 2. SMARTRADE proposed architecture

4. Identifying data and their relationships in the electricity consumption management system

The data model involves identifying data collections, establishing relationships, identifying attributes, and integrity constraints.

In order to develop the database schema for each module, we design the activity diagrams (Figures 3-6) and also revised the data models to support the forecasting and optimization models.

The main activities identified in the electricity consumption model are as follows:

- CO: The main actors, consumer and supplier, will negotiate consumption and generation on the place of consumption. The consumer accepts tariffs and it is signed the contract. Using the configuration from the place of consumption, the supplier install / configure Meters & Sensors;
- LP: Based on the consumer data, the supplier will select the profiles and the algorithm used;
- LF: Using weather data and historical consumption the supplier will configure the forecast parameters and will select and apply the algorithm. Based on the comparison of the predicted values with actual values, the supplier will send real consumption to the BRP (Balance Responsible Party);
- LO: The supplier configures optimization parameters, which will lead to the optimization results that need to be viewed or adjusted. As a result, the consumer will access and implement the optimized consumption program in order to minimize the costs.

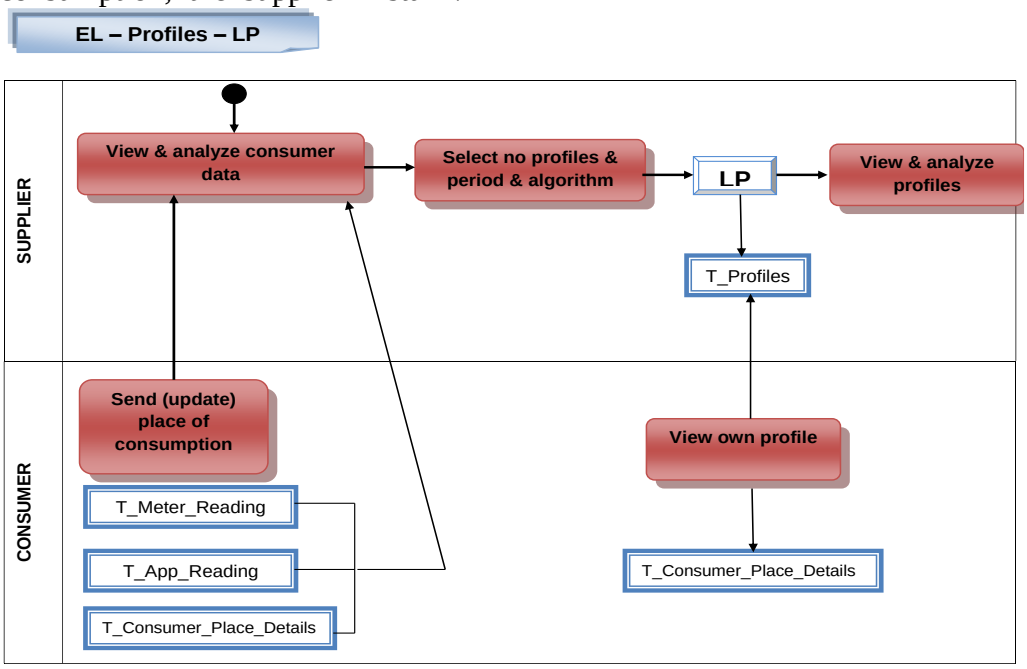


Fig. 3. Activity diagram for Load Profiles (LP)

The activities flow for the **Load Profiles** is detailed in the above figure and consists in the following activities:

1. View & analyze consumer data
2. Select number of profiles & period & algorithm
3. View & analyze profiles
4. Send (update) place of consumption
5. View own profile

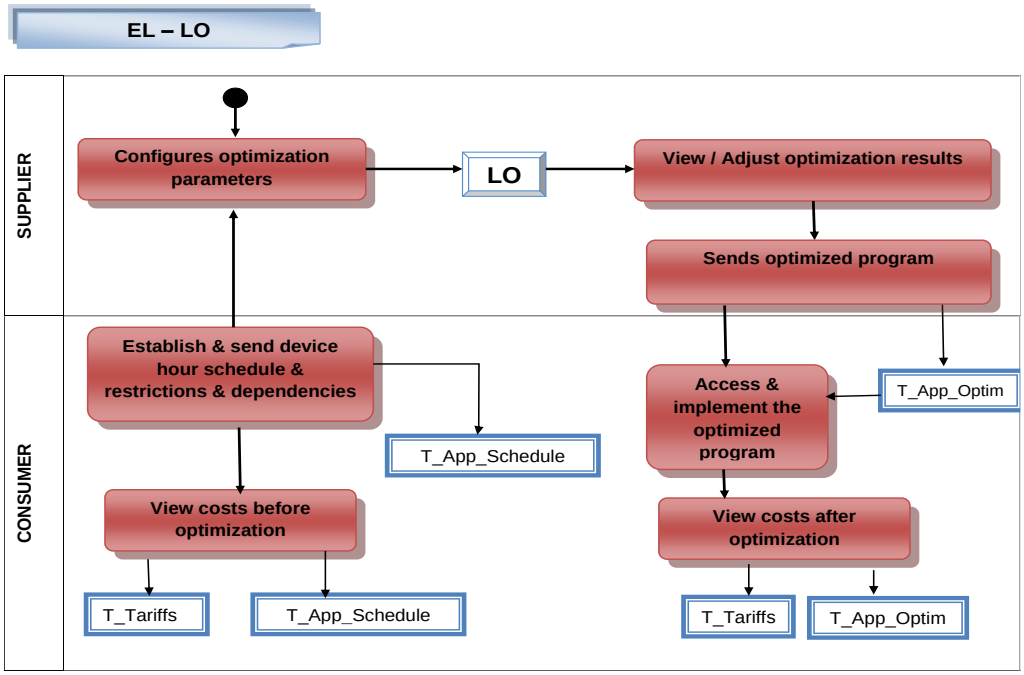


Fig. 4. Activity diagram for Load Optimization (LO)

The activities flow for the **Load Optimization** is detailed in the above figure and consists in the following activities:

1. Configures optimization parameters
2. View / Adjust optimization results
3. Sends optimized program
4. Establish & send device hour schedule & restrictions & dependencies
5. View costs before optimization
6. Access & implement the optimized program
7. View costs after optimization

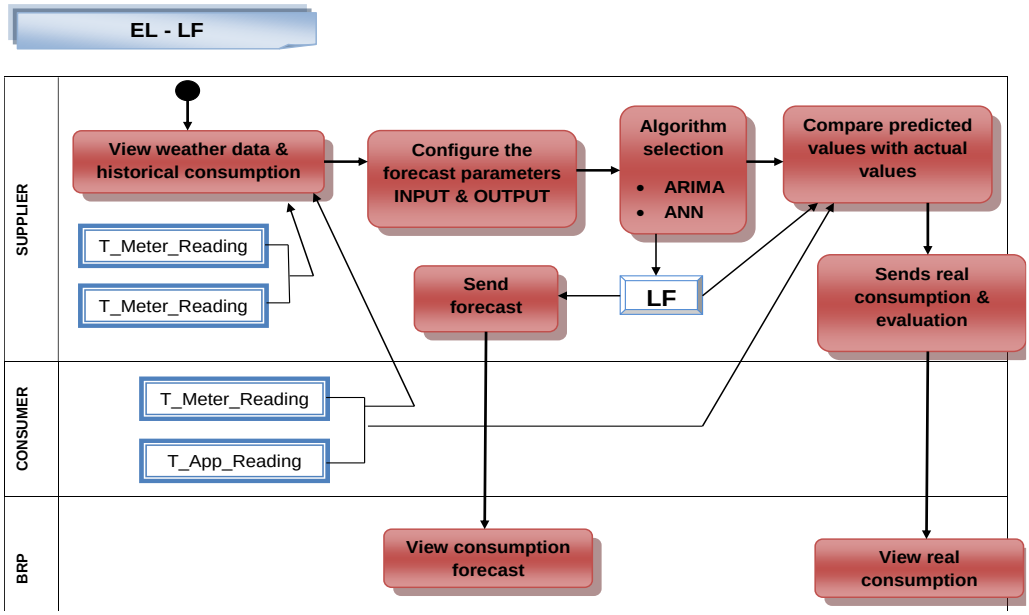


Fig. 5. Activity diagram for Load Forecast (LF)

The activities flow for the **Load Forecast** is detailed in the above figure and consists in the following activities:

1. View weather data & historical consumption
2. Configure the forecast parameters INPUT & OUTPUT

3. Algorithm selection (ARIMA. ANN)
4. Compare predicted values with realized
5. Sends real consumption & evaluation
6. Send forecast
7. View forecast consumption
8. View real consumption

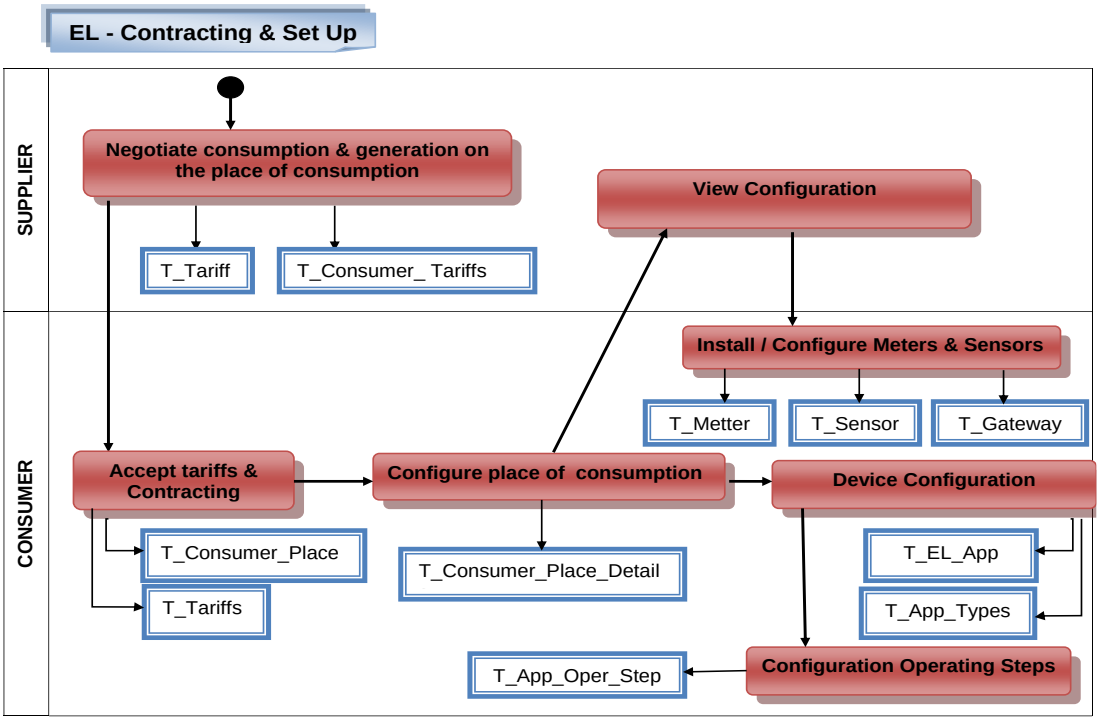


Fig. 6. Activity diagram for Contracting (CO)

The activities flow for the **Contracting & Set Up** is detailed in the above figure and consists in the following activities:

1. Negotiate consumption & generation on the place of consumption
2. Accept tariffs & Contracting
3. Configure place of consumption
4. View Configuration
5. Install / Configure Meters & Sensors
6. Device Configuration
7. Configuration Operating Steps

5. Database schema for Electricity Consumption module

First stage of the data model was detailed in [13] and contains tables initially identified following the system analysis.

In the current stage, we have refined the database, adding new tables or improving the existing ones.

The new tables included in the database are the following:

- T_WEATHER_STATIONS
- T_WEATHER_READINGS
- T_APPLIANCE_OPTIM
- T_APPLIANCE_SCHEDULE
- T_APPLIANCE_OPER_STEP
- T_SENSORS
- T_GATEWAY

Improved EL database schema is represented in the Figure 7.

Fig. 7. Refined database schema

6. Conclusions and future works

In this paper, we have analyzed a series of related papers on optimizing energy consumption and we have proposed some improvements for the electricity consumption data model.

7. Acknowledgment

This paper presents the scientific results of the project "Intelligent system for trading on wholesale electricity market" (SMARTRADE), co-financed by the European Regional Development Fund (ERDF), through the Competitiveness Operational Programme (COP) 2014-2020, priority axis 1 – Research, technological development and innovation (RD&I) to support economic competitiveness and business development, Action 1.1.4 - Attracting high-level personnel from abroad in order to enhance the RD capacity, contract ID P_37_418, no. 62/05.09.2016, beneficiary: The Bucharest University of Economic Studies.

References

[1] A.Barbato, A.Capone, "Optimization Models and Methods for Demand-Side Management of Residential Users: A Survey", OPEN ACCESS energies, ISSN 1996-1073

[2] R. Banos, F. Manzano-Agugliaro, ,,F.G. Montoya, C.Gil, A. Alcayde, J. Gómez, Optimization methods applied to renewable and sustainable energy: A review, , Renewable and Sustainable Energy Reviews 15 (2011) 1753–1766, Renewable and Sustainable Energy Reviews 15 (2011) 1753–1766

[3] O.Bakhtiar, D.Moazzami, K.Rezaie, A non-linear programming model for optimization of the electrical energy consumption in typical factory, Applied Mathematics and Computation 187(2):944-950, April 2007, DOI 10.1016/j.amc.2006.09.138,

[4] S.He, M.Wallace, C.Wilson, A.Liebman, Fast Electrical Demand Optimization Under Real-Time Pricing, The AAAI-17 Workshop on Artificial Intelligence for Smart Grids and Smart Buildings WS-17-05

[5] S.Cui, A.J.Goldsmith, A.Bahai, Energy-constrained Modulation Optimization, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, SEPTEMBER, 2005

[6] Y.Zeng, Y.Cai, G.Huang, J.Dai, A Review on Optimization Modeling of Energy Systems Planning and GHG Emission Mitigation under Uncertainty, Energies 2011, 4, 1624-1656; doi:10.3390/en4101624, ISSN 1996-1073

[7] Y.Liu, H.Dong, N.Lohse, S.Petrovic, A multi-objective genetic algorithm for optimisation of energy consumption and shop floor production performance, International Journal of Production Economics, Volume 179, September 2016, Pages 259-272

[8] Kong, D-S; Jang, Y-S; Huh, J-H, A Multi-objective Optimization Method for Energy System Design Considering Initial Cost and Primary Energy Consumption", Korean Journal of Air-Conditioning and Refrigeration Engineering, Volume 26, Issue 8, 2014, pp.357-365, DOI : 10.6110/KJACR.2014.26.8.35

[9] Soroudi A, Ehsan M, Zareipour H. A practical eco-environmental distribution network planning model including fuel cells and non-renewable distributed energy resources. Renewable Energy 2011;36(1):179–88.

[10] Mario E. Castro Gama, Quan Pan, Salman A.M. Fadl-Elmola, Andreja Jonoski, Multivariate Optimization to Decrease total Energy Consumption in the Water Supply System of Abbiategrasso (Milan, Italy), Environmental Engineering and Management Journal, September 2015, Vol.14, No. 9, 2019-2029

[11] Y.Xiong; Y.Zhang; X.Chen; M.Wu, Survey of energy consumption optimization methods for cloud video surveillance system, Control

Conference (CCC), 2013 32nd Chinese ISBN: 978-9-8815-6383-5,
[12] G.Mouzon, Operational Methods and Models for Minimization of Energy Consumption in a Manufacturing Environment" May 2008, Wichita State University, https://soar.wichita.edu/bitstream/handle/10057/1954/d08006.pdf;sequence=1
[13] S-V.Oprea, O.BulentTor, Data model for Demand Side Management, Database Systems Journal, vol. VIII, no.2/2017, ISSN 2069 - 3230

Iuliana ȘIMONCA (BOTHA) is Associate Professor at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from the Bucharest University of Economic Studies. She has graduated the Faculty of Cybernetics, Statistics and Economic Informatics in 2006 and she holds a PhD diploma in Economic Informatics from 2012. She is author/co-author of 10 books and 44 papers published in journals or conference proceedings. She participated as team member in 8 research projects. Her scientific fields of interest include: Database Systems, Data Warehouses, Business Intelligence, Design of Economic Information Systems

Simona-Vasilica OPREA is an Assistant at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from the Bucharest University of Economic Studies. She has graduated the Polytechnic University in 2001, holds a Master Diploma in Infrastructure Management Program, Yokohama National University, Japan in 2007, a PhD diploma in Electrical Engineering in 2009 and a PhD diploma in Economic Informatics in 2017. She is the author of over 50 articles.

Adina Ileana UȚĂ is University Professor in Economic Informatics Department, Academy of Economic Studies of Bucharest. She published over 30 articles in journals and magazines in computer science, informatics and business management fields, over 50 papers presented at national and international conferences, symposiums and workshops and she was member or director in over 40 research projects. In 2000 she finished the doctoral stage, the title of her PhD thesis being: The automation of analysis and forecast activities related to final energy consumption. Her interest domains related to computer science are: software development methodologies and development systems.

Ion LUNGU graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 1974. He got the title of doctor in economy in the specialty economic informatics in 1983. He has been directing graduates who study towards getting a doctor's degree since 1999. At present he is a professor in the department of the faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies of Bucharest. He had documentary activity and specialization with the Eindhoven Technical University of Holland, the Economic University of Athens and Economic University of Milan. His domains of work are: informatics systems and databases.

Osman Bülent TÖR received the B.S., M.S. and Ph.D. degrees from Middle East Technical University (METU), Turkey, in 1998, 2001 and 2008, respectively. He has almost 20 years working experience in the electrical power sector. At this moment, he is working as Managing Partner and Director for EPRA. Prior to this position, he was leading the Power Systems Analysis and Planning Department of TÜBİTAK-Ankara, as a chief power system R&D engineer for 12 years, before he left TÜBİTAK and joined to EPRA in 2013. He has been lecturing as a part-time instructor at the METU - EE 578:Power System Planning, EE 573: Stability and Dynamics of Power Systems. He is the author of several journals and conference papers in the power system area.

Adoption of Internet of Things Technologies in Enterprises

Floarea NĂSTASE[1], Andrei MIHALACHE[1],
Paul MARINESCU[2], Ionuţ MINCIUNĂ[2]
[1]Faculty of Cybernetics, Statistics and Economic Informatics,
[2]Faculty of Management,
University of Economic Studies, Bucharest, Romania
nastasef@ase.ro, andrei@mdata.ro,
paulmarinescu1@gmail.com, ionut.minciuna@gmail.com

The Internet of Things (IoT) pervades any device that has or that can be upgraded with an Internet connection capability. IoT has become a key concept linking uniquely identifiable things to their virtual representations over the Internet. Currently, this approach has spread widely throughout most of areas, enterprises or groups of people, on the thriving express lane provided by IPv6 protocol. This newer version of IP has more than enough addresses, about 3.4×10^{38} addresses to serve all IP networking needs for the foreseeable future when more than twenty-four billion smart things will be connected by 2020. As we move from www (static pages web) to web2 (social networking web) to web3 (ubiquitous computing web), the need for data-on-demand using sophisticated intuitive queries increases significantly.
Keywords: *Internet of Things, Wireless Sensor Networks, ubiquitous sensing, open innovation, pervasive information, smart things, smart environments, business model innovation*

1 Introduction

The *Internet of Things* (IoT) architecture arouses the premises of an organic approach that extends the actual extreme frontier in Internet communications.

As identified by Atzori [1], *Internet of Things* can be realized in three paradigms – internet-oriented (middleware), things oriented (sensors) and semantic-oriented (knowledge).

Although this type of delineation is required due to the interdisciplinary nature of the subject, the usefulness of IoT can be unleashed only in an application domain where the three paradigms intersect.

According to the experts, this next revolution in digital technology is a global, immersive, pervasive, ambient networked computing environment built through the continued proliferation of smart sensors, cameras, software, databases, and massive data centres in a world-spanning information fabric known as the Internet of Things.

At present, we illustrated in **Fig. 1.** that IoT creates a connectivity [2] between six elements:

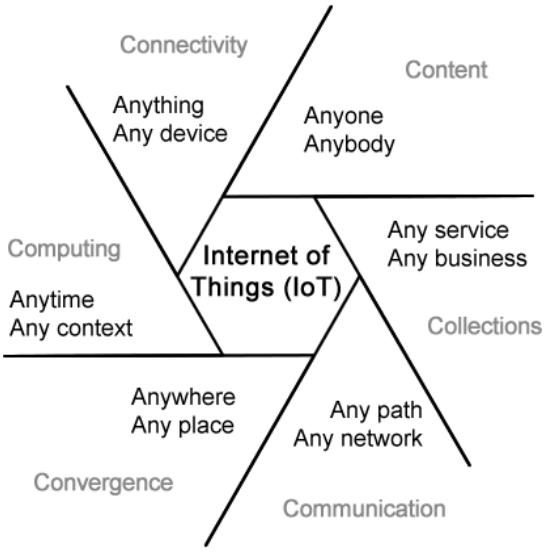


Fig. 1. Connectivity of IoT

- *Anything*, any device that we already use (desktops, notebooks, tablets, smartphones), devices using connectivity in new ways (household appliances like smart TVs or smart sensors and actuators), devices

previously inactive, but now intelligent (smart bulbs, watches or vehicles) or completely new devices (virtual reality headsets and other gadgets);

- *Anytime*, any context enables the real-time computing, monitoring, control and continuous optimization of even a large number of devices, easily, from a single centralized control point; moreover, the computing capability of the device enables all kinds of logic to be written and executed: *if*, *then*, *and*, *or*, *else*, *nor*, and so on depending on self or other devices;
- *Anywhere*, any place the remote devices are situated, they all point towards their manufacturer/owner who is updated at certain intervals or real-time data or parameters settings that he is designed to provide;
- *Any network*, any path the IoT device can connect to a gateway using any communication method like GPRS/2G/3G/4G, Bluetooth, ZigBee, Serial/USB, Ethernet, VPN, Wi-Fi access point, MQTT etc.; because the vast majority of IoT end devices are engineered to operate independently of network connectivity, *individual* data messages are completely *uncritical*;
- *Any service*, any business process can be exposed and called to retrieve information or to activate a certain action; moreover, IoT is more than a business tool for managing business processes more efficiently and more effectively – it also enables a more convenient way of life [3];
- *Anybody*, anyone who is interested or has access to the device or service.

'Things' are active participants in business, information and social processes where they are enabled to interact and communicate among themselves and with the environment by exchanging data and information sensed about the environment, while reacting autonomously to the real/physical world events and influencing it by running processes that trigger actions and create services with or without direct human intervention.

The popularity of different paradigms varies with time. The web search popularity, as measured by the Google search trends during the last 14 years for the terms *Internet of Things*, *Wireless Sensor Networks* and *Ubiquitous Computing* are shown in **Fig. 2.** [4].

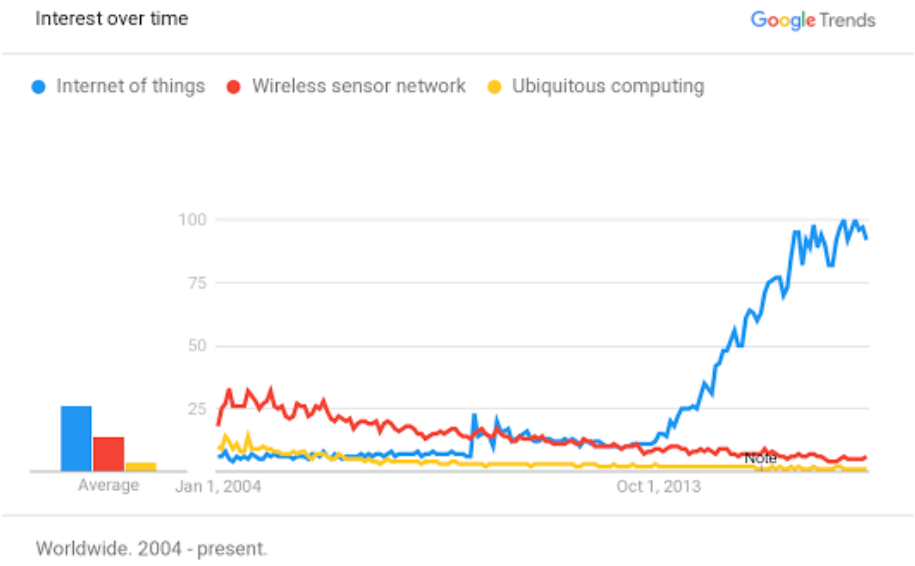


Fig. 2. Google search trends since 2004:
■ Internet of Things, ■ Wireless Sensor Networks, ■ Ubiquitous Computing.

As it can be seen, since IoT has come into existence, search volume is consistently increasing with the falling trend for Wireless Sensor Networks. Between August 2009 and August 2013, the concurrence terms Internet of Things and Wireless Sensor Networks is tight, but ascending trend of IoT continues for the next years as other enabling technologies converge to form a genuine Internet of Things.

Internet of Things has started gaining popularity and overpassed Wireless Sensor Networks and Ubiquitous Computing since August 2013. In fact, this reflects the social acceptability of the technology as consumers look for more data about various topics of interest.

IoT involves the integration of various technologies and domain knowledge across industries. Without cross-industry collaboration, the development of IoT applications will be limited in its scope and reach [5]. Therefore, breaking through the limitations and stalled situation requires an open collaborative approach whereby people together could innovate through the power of open innovation, promoting self-development and seeding the growth of IoT implementation

One of the most popular industry models for IoT implementation today is to combine *cloud computing* and *big data* analytics technology to develop application-specific solutions.

2 Communication requirements

Anything can be an IoT device if it can send or receive data from the cloud and is designed to process a unique task using cloud recommendations.

IoT grants "smart objects" to be active participants in business, information and social processes where they are enabled to interact and communicate among themselves and with the environment by exchanging data and information "sensed" from the environment, while reacting autonomously to the "real world" events and influencing it by running processes that trigger actions and create services with or without direct human intervention.

Ubiquitous services enable interaction with these "smart things" using standard interfaces that provide the necessary link via the Internet, to query and change their state and retrieve any pervasive information associated with them, taking into account security and privacy issues.

The architecture of the original Internet was created long before communicating with billions of very simple devices such as sensors and appliances was ever envisioned.

The present explosion of these much simpler devices creates tremendous challenges for the current networking paradigm in terms of the number of devices, unprecedented demands for low-cost connectivity, and the impossibility of managing remote and diverse equipment.

From a high-level perspective, there are three IoT components [6] which enables seamless ubicomp:

1. Hardware - made up of sensors, actuators, and embedded communication hardware
2. Middleware – on-demand storage and computing tools for data analytics and
3. Presentation - novel easy to understand visualization and interpretation tools which can be widely accessed on different platforms and which can be designed for different applications [7].

A radical evolution of the current Internet into a Network of interconnected objects that not only harvests information from the environment (sensing) and interacts with the physical world (actuation/ command/control) but also uses existing Internet standards to provide services for information transfer, analytics, applications and communications.

Fueled by the prevalence of devices enabled by open wireless technology such as Bluetooth, radio frequency identification (RFID), Wi-Fi and telephonic data services as well as embedded sensor and actuator nodes, IoT has stepped out of its infancy and is on the verge of transforming the

current static Internet into a fully integrated Future Internet [8]. Internet revolution led to the interconnection between people at an unprecedented scale and pace. The next revolution will be the interconnection between objects to create a smart environment. Only in 2011, the number of interconnected devices on the planet overtook the actual number of people.

Smart connectivity with existing networks and context-aware computation using network resources is an indispensable part of IoT. With the growing presence of Wi-Fi and 4G-LTE wireless Internet access, the evolution toward ubiquitous information and communication networks is already evident. However, for the Internet of Things vision to successfully emerge, the computing criterion will need to go beyond traditional mobile computing scenarios that use smartphones and evolve into connecting everyday existing objects and embedding intelligence into our environment. For technology to disappear from the consciousness of the user, the Internet of Things demands:

1. a shared understanding of the situation of its users and their appliances,
2. software architectures and pervasive communication networks to process and convey the contextual information to where it is relevant, and
3. the analytics tools in the Internet of Things that aim for autonomous and smart behavior.

With these three fundamental grounds in place, smart connectivity and context-aware computation can be accomplished.

The first direct consequence of the IoT is the generation of huge quantities of data, where every physical or virtual object connected to the IoT may have a digital twin in the cloud, which could be generating regular updates.

IoT developments show that we will have about twenty-four billion connected devices by the year 2020 [9], which will average out to nine devices per person on earth and to many more per person in digital societies.

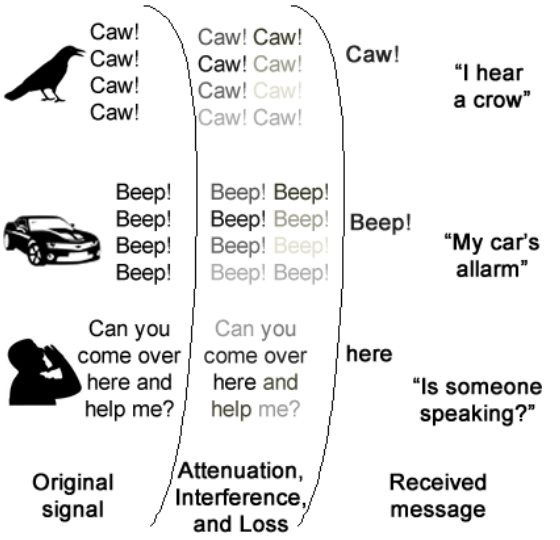


Fig. 3. The results of a lossy connection at an endpoint

Although these challenges are becoming evident now, they will pose a greater, more severe problem as this revolution accelerates.

This paper aims to define a new paradigm for the Internet of Things to be adopted both by individuals and organizations; but first, the need of an IoT architecture much more organic approach compared with traditional networking because it represents an extreme frontier in communications.

The scope and diversity of the devices to be connected are huge, and the connections to the edges of the network where these devices will be arrayed will be "low fidelity": low-speed, lossy (where attenuation and interference may cause lost but generally insignificant data), and intermittent, as shown in **Fig. 3**. At the same time, much of the communication will be machine-to-machine and in tiny snatches of data, which is completely the opposite of networks such as the traditional Internet.

3 Appearance of ubiquitous computing

In the late 1980s researchers tried to create human-to-human interface through technology, but their effort resulted in the creation of the *ubiquitous computing* discipline, whose objective is to embed technology into the background of everyday life. Currently, we are in the post-PC era where smartphones and other handheld devices are changing our environment by making it more interactive as well as informative.

Mark Weiser, the forefather of Ubiquitous Computing (*ubicomp*), defined a smart environment [10] as – the physical world that is richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives, and connected through a continuous network.

The appearance of the Internet has marked a foremost milestone towards achieving ubicomp‘s vision which enables individual devices to communicate with any other device in the world. The inter-networking reveals the potential of a seemingly endless amount of distributed computing resources and storage owned by various owners.

In contrast to Weiser’s calm computing approach, Rogers proposes a human-centric ubicomp which makes use of human creativity in exploiting the environment and extending their capabilities [11]. He proposes a domain-specific ubicomp solution when he says: “In terms of who should benefit, it is useful to think of how ubicomp technologies can be developed not for the Sal’s of the world, but for particular domains that can be set up and customized by an individual firm or organization, such as for agriculture production, environmental restoration or retailing.”

Caceres and Friday [12] discuss the progress, opportunities, and challenges during the 20-year anniversary of ubicomp. They suggest the building blocks of ubicomp and the characteristics of the system to adapt to the changing world. More importantly, they identify two critical technologies for growing the ubicomp infrastructure - *Cloud Computing* and the *Internet of Things*.

The advancements and convergence of micro-electro-mechanical systems (MEMS) technology, wireless communications, and digital electronics have resulted in the development of miniature devices having the ability to sense, compute and communicate wirelessly in short distances. These miniature devices called nodes interconnect to form a *Wireless Sensor Networks* (WSN) and find wide application in environmental monitoring, infrastructure monitoring, traffic monitoring, retail, etc. [13]. This has the ability to provide ubiquitous sensing capability which is critical in realizing the overall vision of ubicomp as outlined by Weiser [10].

In order to build a complete IoT vision, an efficient, secure, scalable and market-oriented computing and storage resourcing are essential. Cloud computing [14] is the most recent paradigm to emerge which promises reliable services delivered through next-generation data centers that are based on virtualized storage technologies. This platform acts as a receiver of data from the ubiquitous sensors; as a computer to analyze and interpret the data; as well as providing the user with easy to understand web-based visualization. The ubiquitous sensing and processing works in the background, *hidden* from the user, and provides him pervasive information.

This new integrated Sensor-Actuator-Internet framework is the core technology around which a smart environment will be shaped: information generated will be shared across diverse platforms and applications, to develop a common operating picture of an open innovation smart environment, where control of certain unrestricted ‘Things’ are made possible. As we shifted from www (static pages web) to web2 (social networking web) to web3 (ubiquitous computing web), the need for data-on-demand using

complex intuitive queries increases. To take full advantage of the available Internet technology, there is a need to deploy large-scale, platform-independent, wireless sensor network infrastructure that includes data management and processing, actuation and analytics. *Cloud computing* promises high reliability, scalability, and autonomy to provide ubiquitous access, dynamic resource discovery and composability required for the next-generation Internet of Things applications. Consumers will be able to choose the service level by changing the Quality of Service parameters.

4 Smart environments and 'things' uniquely identification

According to Forrester Research [15], a smart environment uses information and communications technologies to make the critical infrastructure components and services of a city's administration, education, healthcare, public safety, real estate, transportation and utilities more aware, interactive and efficient.

If we give this definition a more user-centric approach and do not restrict it to any standard communication protocol, it will allow long-lasting applications to be developed and deployed using the available state-of-the-art protocols at any given point in time.

Open standards are required to use and extend its functionality. It will be a huge network, considering that every object has its virtual representation. Open innovation clusters have a great contribution to the definition and development of the present and future versions of the Internet of Things.

Therefore, in IoT scalability is required. The Internet of Things will need to be flexible enough to adapt to changing requirements and technological developments. Its development can be accelerated through the availability of open source software that allows anyone to implement and test new functionalities. Another opportunity to experiment and test new functionalities are living lab initiatives, where service providers and users participate in a collaborative environment.

The Internet of Things requires a more holistic architecture that includes layering of standards, separation of data models and interfaces, provision of extension mechanisms, specification of data models and interfaces, initially in a neutral abstract manner (e.g., using UML), then with provision of specific transport bindings (e.g., web services) and schema bindings (e.g., XML).

This holistic architecture is possible using open innovation to build a cloud-based open community environment in which participants with different expertise and vertical industry knowledge can share resources and leverage the different development tools from each other. Innovators could come together in the community and exchange innovative ideas and creative knowledge to effectively find appropriate products or solutions that fill the gaps of each other's shortcomings [5].

A future Internet of Things has to integrate stakeholders who will be affected by the Internet of Things, such as citizens, small and medium enterprises, governmental institutions and policymakers, to meet and match key societal and economic needs. Applications that recognize and improve the fundamental qualities of life for users, businesses, society and the environment are needed [8].

Finally, it needs a sustainable infrastructure to provide a basis for the necessary investments.

Still, one of the greatest challenges of IoT, critical for its success is the unique identification of billions of devices but also to control remote devices through the Internet. The few most critical features of creating a unique address are: uniqueness, reliability, persistence and scalability.

Every element that is already connected and those that are going to be connected must be identified by their unique identification, location, and functionalities.

The current IPv4 may support to an extent where a group of cohabiting sensor devices can be identified geographically, but not individually. The Internet Mobility attributes in the IPV6 may alleviate some of the device identification problems; however, the heterogeneous nature of wireless nodes, variable data types, concurrent operations and the confluence of data from devices exacerbates the problem further [16].

Persistent network channel data traffic ubiquitously and. Although the TCP/IP takes care of this mechanism by routing in a more reliable and efficient way, from source to destination, the IoT faces a bottleneck at the interface between the gateway and wireless sensor devices. Furthermore, the scalability of the device address of the existing network must be sustainable. The addition of networks and devices must not hamper the performance of the network, the functioning of the devices, the reliability of the data over the network or the effective use of the devices from the user interface.

To solve these issues, the Uniform Resource Name (URN) system is considered fundamental for the development of IoT. URN creates replicas of the resources that can be accessed through the URL. With large amounts of spatial data being gathered, it is often quite important to take advantage of the benefits of metadata for transferring the information from a database to the user via the Internet [17]. IPv6 also gives a very good option to access the resources uniquely and remotely. Another critical development in addressing is the development of a lightweight IPv6 that will enable addressing home appliances uniquely.

Wireless sensor networks (considering them as building blocks of IoT), which run on a different stack compared to the Internet, cannot possess IPv6 stack to address individually and hence a subnet with a gateway having a URN will be required. With this in mind, we then need a layer for addressing sensor devices by the relevant gateway. At the subnet level, the URN for the sensor devices could be the unique IDs rather than human-friendly names as in the www, and a lookup table at the gateway to address this device. Further, at the node level, each sensor will have a URN (as numbers) for sensors to be addressed by the gateway [18]. The entire network now forms a web of connectivity from users (high-level) to sensors (low-level) that is addressable (through URN), accessible (through URL) and controllable (through URC).

5 Data storage and analytics

One of the most important outcomes is the creation of an unprecedented amount of data. Storage, ownership, and expiry of the data become critical issues. The internet consumes up to 5% of the total energy generated today and with these types of demands, it is sure to go up even further.

Machine-to-machine communications require minimal packaging and presentation overhead. For example, a moisture sensor in a farmer’s field may have only a single value to send of volumetric water content. It can be communicated in a few characters of data, perhaps with the addition of a location/identification tag. This value might change slowly throughout the day, but the frequency of meaningful updates will be low. Similar terse communication forms can be imagined for millions of other types of IoT sensors and devices. Many of these IoT devices may be simplex or nearly simplex in data flows, simply broadcasting a state or reading over and over while switched on without even the capacity to “listen” for a reply.

This raises another aspect of the typical IoT message: it’s individually unimportant. For simple sensors and state machines, the variations in conditions over time may be small. Thus, any individual transmission from the majority of IoT devices is likely completely uncritical. These messages are being collected and interpreted elsewhere

in the network, and a gap in data will simply be ignored or extrapolated (see **Fig. 4**).

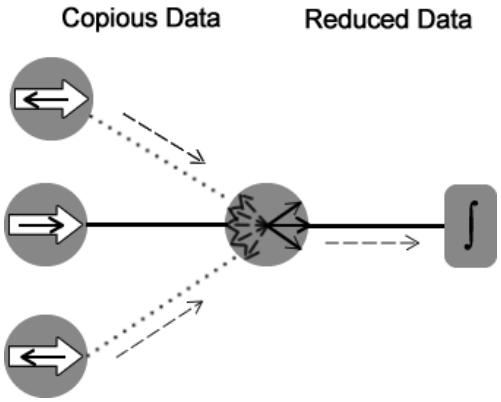


Fig. 4. Multiple identical messages may be received; some are discarded

Even more complex devices, such as a remotely monitored diesel generator, should generate little more traffic, again in terse formats unintelligible to humans, but gathered and interpreted by other devices in the IoT. Overall, the meaningful amount of data generated from each IoT device is vanishingly small—nearly exactly the opposite of the trends seen in the traditional Internet.

Today's traditional Internet is extremely reliable, even if labeled "best effort." Overprovisioning of bandwidth (for normal situations) and backbone routing diversity have created an expectation of high service levels among Internet users. "Cloud" architectures and the structure of modern business organizations are built on this expectation of Internet quality and reliability.

But at the extreme edges of the network that will make up the vast statistical majority of the IoT, connections may often be intermittent and inconsistent in quality.

Traditional protocols such as TCP/IP are designed to deal with lossy and inconsistent connections by resending data. Even though the data flowing to or from any individual IoT device may be exceedingly small, it will grow quite large in aggregate IoT traffic. The inefficiencies of resending vast quantities of *mostly individually unimportant* data are clearly an unnecessary redundancy.

Data must be stored and used intelligently for smart monitoring and actuation. It is important to develop artificial intelligence algorithms which could be centralized or distributed based on the need. Novel fusion algorithms need to be developed to make sense of the data collected. State-of-the-art non-linear, temporal machine learning methods based on evolutionary algorithms, genetic algorithms, neural networks, and other artificial intelligence techniques are necessary to achieve automated decision making [19].

Visualization is critical for an IoT application as this allows the interaction of the user with the environment. With recent advances in touchscreen technologies, use of smart tablets and phones has become very intuitive. For an average person to fully benefit from the IoT revolution, attractive and easy to understand visualization has to be created. This will also enable policymakers to convert data into knowledge, which is critical in fast decision making. Extraction of meaningful information from raw data is non-trivial. This encompasses both event detection and visualization of the associated raw and modeled data, with information represented according to the needs of the end-user

6 Enterprises

We refer to the 'Network of Things' within a work environment as an enterprise-based application. Information collected from such networks are used only by the owners and the data may be released selectively.

Environmental monitoring is the first common application which is implemented to keep track of the number of occupants and manage the utilities within the building (e.g., HVAC – heating, ventilation, and air conditioning, or lighting).

Sensors have always been an integral part of the factory setup for security, automation, climate control, etc. This will eventually be replaced by a wireless

system giving the flexibility to make changes to the setup whenever required. This is nothing but an IoT subnet dedicated to factory maintenance.

One of the major IoT application areas that are already drawing attention is Smart Environment IoT [20]. There are several testbeds being implemented and many more planned in the coming years. The smart environment includes subsystems as shown in **Table 1.** and the characteristics from a technological perspective are listed briefly. It should be noted that each of the subdomains covers many focus groups and the data will be shared [8].

Table 1. Smart environment application domains

	Smart home/ office	**Smart retail**	**Smart city**	**Smart agriculture/ forest**	**Smart water**	**Smart transport-tation**
Network size	Small	Small	Medium	Medium/large	Large	Large
Users	Very few, family members	Few, community level	Many, policy makers, general public	Few, landowners, Policy makers	Few, government	Large, general public
Energy	Rechargeable battery	Rechargeable battery	Rechargeable battery, Energy harvesting	Energy harvesting	Energy harvesting	Rechargeable battery, energy harvesting
Internet connectivity	Wifi, 3G, 4G LTE backbone	Wifi, 3G, 4G LTE backbone	Wifi, 3G, 4G LTE backbone	Wi-Fi, satellite communica-tion	Satellite communica-tion, microwave links	Wi-Fi, satellite communica-tion
Data management	Local server	Local server	Shared server	Local server, shared server	Shared server	Shared server
IoT devices	RFID, WSN	RFID, WSN	RFID, WSN	WSN	Single sensors	RFID, WSN, single sensors
Bandwidth requirement	Small	Small	Large	Medium	Medium	Medium/ large

The applications or use-cases within the urban environment that can benefit from the realization of a smart city WSN capability are shown in **Table 2.**

These applications are grouped according to their impact areas.

Table 2. Potential IoT applications identified by different focus groups of the city of Bucharest

Citizens	
Healthcare	Triage, patient monitoring, personnel monitoring, disease spread modeling and containment—real-time health status and predictive information to assist practitioners in the field, or policy decisions in pandemic scenarios
Emergency services, defense	Remote personnel monitoring (health, location); resource management and distribution, response planning; sensors built into building infrastructure to guide first responders in emergencies or disaster scenarios
Crowd monitoring	Crowd flow monitoring for emergency management; efficient use of public and retail spaces; workflow in commercial environments
Transport	
Traffic management	Intelligent transportation through real-time traffic information and path optimization
Infrastructure	Sensors built into infrastructure to monitor structural fatigue and other maintenance;

monitoring	accident monitoring for incident management and emergency response coordination
Services	
Water	Water quality, leakage, usage, distribution, waste management
Building management	Temperature, humidity control, activity monitoring for energy usage management, Ventilation and Air Conditioning (HVAC)
Environment	Air pollution, noise monitoring, waterways, industry monitoring

This includes the effect on citizens considering health and wellbeing issues; transport in light of its impact on mobility, productivity, pollution; and services in terms of critical community services managed and provided by local government to city inhabitants.

7 Conclusions

The Internet of Things is essentially about how data is retrieved from sensors and chips to drive various business processes, automate them, enable new applications and service customers in entirely new approaches.

The evolution of the next-generation mobile systems will depend on the creativity of the users in designing new applications. IoT is an ideal emerging technology to influence this domain by providing new evolving data and the required computational resources for creating innovative applications.

8 Acknowledgment

We are grateful to all of the members of the Department of Informatics and Economic Cybernetics and Department of Management at The University of Economic Studies with whom we have the pleasure to work during this and other related projects. Each of the members of these departments provides extensive personal and professional guidance and shares a great deal about both scientific research and life in general.

References

[1] L. Atzori, A. Iera, G. Morabito, The Internet of Things: A survey, *Computer Networks*, Volume 54, pp. 2787–2805, 2010.

[2] O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, A. Bassi, I. S. Jubert, M. Mazura, M. Harrison, M. Eisenhauer, P. Doody, "Internet of things strategic research roadmap.", *Internet of Things-Global Technological and Societal Trends, 1*, ISBN 978-8-792-32967-7, pp. 9-52, 2011.

[3] D. Uckelmann, M. Harrison, F. Michahelles, "An Architectural Approach Towards the Future Internet of Things", *Architecting the Internet of Things*, ISBN: 978-3-642-19156-5, pp. 1-22, Springer-Verlag Berlin Heidelberg, 2011.

[4] Google Trends, *Google*. https://trends.google.com/trends/explore?date=all&q=%2Fm%2F02vnd10,%2Fm%2F043x2v,%2Fm%2F07v58

[5] A. Perng, "IoT Development Enters the Open Innovation Craze" - White Paper, NEXCOM, 2016.

[6] S. Tilak, N. Abu-Ghazaleh, W. Heinzelman, A taxonomy of wireless micro-sensor network models, Acm Mobile Computing and Communications Review. 6, pp. 28–36, 2002.

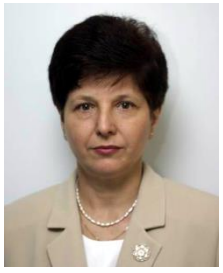
[7] M. Tory, T. Moller, Rethinking Visualization: A High-Level Taxonomy, Information Visualiza-tion, INFOVIS 2004. Symposium on. IEEE, pp. 151–158, 2004.

[8] J Gubbi, R Buyya, S Marusic, M Palaniswami "Internet of Things (IoT): A vision, architectural elements, and future directions", *Elsevier. Future Generation Computer Systems*. Volume 29, Issue 7, pp. 1645-1660, 2013.

[9] Routledge, The Internet of Things: From RFID to the Next-Generation Pervasive Networked Systems (Hardback) – Routledge, 2008.

[10] Weiser M., Gold R., The origins of ubiquitous computing research at

PARC in the late 1980s, *IBM Systems Journal*. 1999.
[11] Y. Rogers, "Moving on from weiser's vision of calm computing: Engaging ubicomp experiences", UbiComp, *Ubiquitous Computing*, 2006.
[12] R. Caceres, A. Friday, "Ubicomp Systems at 20: Progress, Opportunities, and Challenges", *IEEE Pervasive Computing* vol. 11, pp. 14–21, 2012.
[13] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless Sensor Networks: A Survey", *Computing Networks*, pp. 38 393–422, 2002.
[14] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", *Future Generation Computer Systems*, vol. 25, pp. 599–616, 2009.
[15] J. Belissent, "Getting clever about smart cities: new opportunities require new business models", *Forrester Research*, 2010.
[16] M. Zorzi, A. Gluhak, S. Lange, A. Bassi, From today's Intranet of Things to a future Internet of Things: a wireless- and mobility-related view, IEEE Wireless Communications 17, pp. 43–51, 2010.
[17] N. Honle, U.P. Kappeler, D. Nicklas, T. Schwarz, M. Grossmann, Benefits of integrating meta data into a context model, pp. 25–29, 2005.
[18] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, T. Razafindralambo, A survey on facilities for experimental Internet of Things research, IEEE Communications Magazine 49, pp. 58–67, 2011.
[19] A. Ghosh, S.K. Das, Coverage and connectivity issues in wireless sensor networks: a survey, *Pervasive and Mobile Computing* 4, pp. 303–334, 2008.
[20] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, T. Razafindralambo, A survey on facilities for experimental Internet of Things research, "IEEE Communications Magazine" vol. 49, pp. 58–67, 2011.



Floarea NĂSTASE graduated the Faculty of The Faculty of Automatic Control and Computer Science of the University POLITEHNICĂ of Bucharest in 1975. She obtained the title of doctor of engineering degree from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 1995. She has been directing graduates who study towards getting a doctor's degree since 2000. At present, she is a professor in the department of the faculty of Cybernetics, Statistics and Economic Informatics of the University of Economic Studies of Bucharest.

Andrei MIHALACHE has a background in computer science and is interested in database related issues and enterprise modeling techniques. Currently, he is a PhD candidate in the field of Economic Informatics at the University of Economic Studies. His research interests include: cloud computing, internet of things, big data, enterprise data model, adaptable information systems, business process innovation, IT project management, software engineering, software improvement and web collaborative technologies.

Paul MARINESCU graduated the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies of Bucharest in 1976 licensed in the field of Economic Informatics. In 1997, he obtained the title of doctor in economics at the Bucharest National Institute of Economic Research "Costin C. Kirițescu". He is also a consultant, a trainer for entrepreneurs and evaluator for international conferences. At present, he is a professor in the department of the faculty of Management of the University of Economic Studies of Bucharest.

Ionuț MINCIUNĂ is a PhD Candidate at the Faculty of Management, University of Economic Studies in Bucharest, Romania. He is also a business consultant, an academic assistant and a trainer for entrepreneurs. His research interests include: crowd-based models, open-innovation, external innovation, entrepreneurship, outsourcing, project management, portfolio management, technology acquisition, web collaborative technologies and business models.

What can Romanian Supermarkets do in order to improve the Customer Experience? Recommendation Systems in Creating Personalized Offers

Cristina STOICESCU, Cristina STANCIU-MÂNDRULEANU
University of Economic Studies, Bucharest, Romania
kris_stoicesku@yahoo.com, cristina.stanciu@csie.ase.ro

Modern food retail stores have started to penetrate the Romanian market over 20 years ago and since then they have gained more than half of the total market share. More than 10 international brands have opened over 2500 stores creating a competitive environment where each company needs to develop new business strategies in order to achieve positive financial results. In order to gain new clients supermarkets are opening new stores every year, but in order to keep their actual clients they need to improve the customer experience. Introducing fidelity cards, loyalty points, instant awards, tasting sessions or moving in the online area are some of the actions taken by supermarkets in Romania, but none of them have yet developed a recommendation system that can propose personalized offers. This article describes a hybrid algorithmic model based on data mining and collaborative filtering techniques. The goal of the model is to recommend date specific products, similar products and cross-shopping cart products that fulfill the custom needs of each consumer. So at the end of the shopping session each client will receive an offer that includes specific products and customized discounts to his buying behavior. As already sustained by other markets, personalized offers will improve the customer experience and will increase the sales of supermarkets.
Keywords: *consumer behavior, food retailers, recommendation systems, customer experience*

1 Introduction

In the last decades, consumer behavior has changed drastically, becoming a focus point for all companies around the world. Consumers of the 21st century are more informed, the buying process being completed before a first interaction with sales, they are socially networked, more in-store purchases being abandoned due to negative online sentiment and less loyal, more customers being willing to try a new brand to get better customer service. The main causes for this phenomenon are the demographic, social and economic changes and the fast evolution of technology, which reshaped the consumer behavior. As a consequence, companies started to develop their own models in order to quantify consumers needs which have become more diverse and more custom, and to better predict their uncertain and changing behavior.

In Romania, the food retail market is becoming saturated, new stores being opened annually in order to satisfy the extended demands. In this context, in order to gain more market share, the food retail stores have realized the importance of gaining the loyalty of their customers.

The main goal of this paper is to describe a model that can anticipate the needs of consumers and generate personalized offers, which will increase the sales of the company. The model is a hybrid algorithm based on data mining and collaborative filtering techniques that in the end will create a personalized recommendation for each client. The offer includes three types of products: products that are specific to the client shopping cart, products bought by other clients with a similar shopping cart and specific day products. In this way, it is encouraged the purchase of more products from the same category, and the diversification of the current shopping cart. By increasing the number of the products sold, the financial results of the food retail companies will be improved.

Section 2 of the paper reviews some of the research literature concerning the evolution

of consumers behaviour and the actual trends that shapes the food retail market in Romania. Section 3 presents the context needed for implementing the model that generates personalized recommendation, while section 4 describes in detail the main steps of model. Section 5 contains the conclusions and future research.

2 Literature survey

Consumer behavior is an emergent phenomenon that evolved along with human development. Thus, during prehistoric the human behavior occurred in a very limited way, people being grouped in small families with the only concern of surviving. According to Miller (2009), much later, people began to develop social skills that eventually led to the emergence of money, social status, wealth and ultimately to shaping consumer behavior. The study of consumer behavior is an interdisciplinary science that integrates information from psychology, sociology, social psychology, anthropology, artificial intelligence and economics, which currently has started to gain attention of all companies who want to survive in a competitive and evolving environment, including food retailers. [1]

Burke (2014) performed several studies regarding consumer behavior in supermarkets and he noted that 85% of the products purchased by a consumer don't change over time, economic agents having stable preferences for most of the bought products. [2] According to this, in order to increase its sales, a food retailer needs to primarily focus on quantity and not on variety. However, only a combination of both methods will determine in the end improved financial results. In Romania, supermarkets and hypermarkets are using bonus points and loyalty points programs to gain more market share, offering only general discounts or discount that are dedicated to an entire customer segment. Therefore, Romanian food retailers are focusing exclusively on the price strategy and not on loyalty programs based on personalized recommendations, programs already adopted and tested in other countries like USA or UK.

According to a study by Precima (2017) traditional marketing techniques are still dominating the food retail markets around the world, although companies are not very pleased with their results, 22% generating zero benefits for the companies questioned in 2017. Traditional techniques aim to identify the customer segments that will positively react to an already existing general promotion, while current trends suggest the necessity to create personalized recommendations based on the needs and behavior of each consumer. [3]

During LEAD Marketing Conference, McVie (2017) spoke about the importance of personalized recommendations in food retail stores and how to set discounts according to their impact on business performance. There is a directly proportional relationship between discounts, sales volume and profit, its intensity becoming negative once the discounts reach a 40% threshold, the optimal discount level being between 10% and 20% when all financial indicators are recording maximum rates. Discounts up to 20% influence loyal customers and generate high sales volume, while discounts over 30% primarily affect customers who follow exclusively offers and generate small earnings. Discounts over 40% are generating loss. [4]

In conclusion, personalized discounts should fall within a range of 5% to 20%, focusing primarily on loyal customers. In this regard, Romanian food retailers should introduce personalized offers to gain customer loyalty and competitive advantages, by integrating recommendation systems in their business strategy. According to Isinkaye et al. (2015), recommendation systems assume data collection, data processing and adaptation of data filtering techniques, finally generating a recommendation, a feedback loop being present between these stages. Recommender systems may use

different approaches like collaborative filtering, content-based filtering or hybrid techniques. [5]

According to Leskovec et al. (2014) content-based filtering techniques have the advantage of being independent of other consumer choices, and if a consumer's preferences change the recommendation system can be adapted in a very short time. The main disadvantage is that products must have similar content because the data are collected for the same types of variables and in the absence of data the results can be disruptive. [6] According to Herlocker et al. (2004), collaborative filtering techniques are mainly used when heterogeneous products with different characteristics need to be selected. Collaborative filtering techniques start from the assumption that users with similar shopping carts have common preferences, so recommendations will be made on the basis of similarities between consumers, not products. [7]

In 2006, Netflix launched a contest to improve its recommendation system named Cinematch. At that time, the recommendation system relied on content-based filtering techniques, using as input the ranking results and reviews of users. Recently, in order to generate personalized recommendations for each consumer, user profile analysis techniques based on historical activity have been also integrated into the system (Gomez-Uribe and Hunt, 2016). [8]

The large volume of information has also become a problem for social networks like Facebook, selecting the right information that might caught consumers attention becoming a real challenge. Baatarjav et al. (2008) proposed a recommendation system for group socialization on Facebook by integrating hierarchical cluster analysis techniques and decision trees. According to their results, the group recommendation system has an accuracy of 73%, using as main data the consumers profile. [9]

Recommendation systems are migrating from e-commerce to brick-and-mortar industries, becoming a solution for companies that sell a large range products and want to obtain a stable market share and gain customer loyalty.

3 Hypotheses and Prerequisites

The goal of the model is to generate personalized offers that will satisfy the specific needs of each consumer. In order to define and describe the model, the following points were taken into consideration:

- Who will receive the personalized offer – new or already existing customers;
- How to select and filter the recommended products – how many products are included in the recommendation and how are they prioritized within the recommendation;
- How to customize the additional information – the discounts, the validity or the newsletters;
- How to distribute the personalized offer – electronic or printed and when should it be delivered.

3.1. Customer selection

The model is based on a prescriptive system which analysis the historical data of consumers purchases in order to generate the personalized recommendations. Therefore, the first prerequisite is the presence of an instrument used to collect information regarding consumer behavior. In this sense, loyalty cards is the best solution. This marketing strategy already exists on the Romanian food retail market, but is used only for isolated situations.

Consumers that hold a loyalty card will be considered already existing customers for which a history of purchase can be analysed and recommendations can be generated, while consumers who do not have such a card will be considered new customers, although they might possibly do regular shopping. The offers generated by the model will be fully personalized and

will not include general offers. In order to eliminate the risk that a general offer would be more attractive than the personalized one, negatively affecting the recommendation system, the company may cumulate the discounts or may introduce the general offers as variable in the model increasing significantly its complexity. According to the business strategy, the company can also decide to create personalized offers for new customers. This might happen if the company is launching on a new market, if it wants to build a stable database or seeks to gain a larger market share.

Food retail market in Romania is becoming saturate and the already existing stores have stable positions on the market so their main focus should be on gaining customer loyalty. The model will generate full personalized offers only for consumers who have a fidelity card, while general offers will primarily appeal to attracting new customers. The customers will be able to cumulate the discounts.

3.2. Product selection

The model splits the products into three categories, date specific products, specific shopping cart products and products different from the consumers shopping cart. For each category a different selection and filtering logic will be defined. The opportunity cost increases along with the number of products a consumer can choose from, which can make the buying decision very difficult. So, the total number of products included in the personalized offer must be limited, a larger discount for fewer products having a higher impact than a lower discount for multiple products.

The model will recommend every time 6 products for each consumer, the number of products specific to the shopping cart being directly proportional to the total number of products, thus having the highest weight. Only one product date specific and one product different from the shopping cart will be included in the offer. Products will be prioritized according to the category they belong to, so the first products will be those date specific and different from consumer shopping cart, while the other products will be those specific to the shopping cart.

3.3. Additional information selection

Discounts on birthdays or name days is one of the traditional marketing techniques used by food retailers. The model will generate a general discount of 15% on the customer's birthday and 5% on his name day. The discount will be applied only once for a full shopping cart within three days, including the celebration day. This discount will be included in the offer as a header separate section before all other discounts and will be announced in advance so it may be included in several successive recommendations. This discount will not affect the total number of products included in the offer which is a default configuration. The discount will be automatically granted once the customer is identified in the system and will be cumulated with all the other offers. If birthday and name day are in the same period, only the birthday discount will be generated, as its value is significantly higher.

The value of discounts for the other products is influenced by customer loyalty, being directly proportional to the frequency of purchases and purchasing power of the consumer. The discount can be customized for each product, influencing the complexity of the model, or can be applied as a general discount for the entire personalized offer. The model will determine a general discount for all products included in the recommendation.

The discounts will be applied on a time frame established by default or it can be also customized according to the consumer behavior. In Romanian, shopping in supermarkets is a weekly activity for the majority of consumers, so the model will generate for the recommendation a one week validity period that will start the day after the current shopping session.

Depending on the company's strategy, the offer can be used only once in that time frame or whenever the customer comes to shopping during that time.
In order to make recommendations even more attractive, they will also include a personalized Newsletter, based on nutritional information, diets, or recipes for a food product included in the offer or based on up-to-date information for a non-food product included in the offer. This will also be a separate section at the end of the offer or on the back of it, and the information will not be repetitive.
For customers who have a low shopping frequency, the quantity of the products sold needs to be increased, so selecting one of the specific shopping cart products included in the offer would be the best solution for generating the Newsletter. For customers who have a high shopping frequency, they aim is to increase the diversity of their shopping carts, so the Newsletter will be generated for products different from the shopping cart or even for random products that have not been purchased or recommended before.

3.4. Distribution channel

The personalized recommendations will be delivered once a day only for the first shopping session in order not to stimulate splitting the products into several purchases. Usually offers have the greatest impact when consumer are still focused on the shopping session, so the recommendations will be distributed at the end of the shopping session in order to increase the view rate. The consumer will be able to receive a printed offer along with the shopping bill or by e-mail. The consumer will be able to choose the distribution mode when requesting the fidelity card and will then be able to change this option according to his needs. No matter how the recommendation is distributed, the consumer must present it at the end of the shopping session in order to apply the discounts.

4 Input variables and main steps of the model

The shopping frequency is represented by the number of shopping sessions made in a recent time unit. The time unit can be predefined by the company or it can be represented by the entire period a consumer has become a loyal client by purchasing a loyalty card. As consumer behavior may change over time, affecting also the frequency of purchases, the time unit is represented by the last six months. The shopping frequency is expressed in days and is calculated using the following formula:

$$SF = \frac{TU}{SS}$$

SF – shopping frequency
TU – time unit
SS – number of shopping sessions in the defined time unit

Customers with a shopping frequency less than 7 days will be considered loyal customers, while customers with a shopping frequency greater than one month will be considered casual customers.
There are many social, psychological and economic factors that influence consumer behavior, such as gender, age, level of training, socio-professional category, residence environment, the main factor being the income. For most consumers the products price is a decisive component in the purchasing process, products from the same category being sold at different prices depending on their quality or supplier. The purchasing power of a consumer will be determined by the price category of the purchased products. The model attaches to each product a price category by calculating the ratio between the price of the product and the average price of all products in the same category. Thus, each product belongs to one of the following categories:

- Category 1 – budget product
- Category 2 – mass product
- Category 3 – premium product
- Category 4 – luxury product

The purchasing power is calculated by selecting all products purchased in the predefined time unit and applying the following weighted arithmetic mean:

$$PP = \begin{cases} \left[\frac{1*x_1+2*x_2+3*x_3+4*x_4}{x_1+x_2+x_3+x_4}\right], \\ \left\{\frac{1*x_1+2*x_2+3*x_3+4*x_4}{x_1+x_2+x_3+x_4}\right\} \leq 0{,}5 \\ \left[\frac{1*x_1+2*x_2+3*x_3+4*x_4}{x_1+x_2+x_3+x_4}\right]+1, \\ \left\{\frac{1*x_1+2*x_2+3*x_3+4*x_4}{x_1+x_2+x_3+x_4}\right\} > 0{,}5 \end{cases}$$

x_i – number of products from Category i
PP – purchasing power
The purchasing power can be an essential factor in dividing customers into clusters and thus reducing the complexity of the customer based filtering algorithm. Also, the purchasing power of the customer is used to accurately identify the exact product that will be recommended to a customer, always selecting products whose price category does not exceed the customer's purchasing power.
The discount calculated by the model will be between 5% and 15%, being directly proportional to the shopping frequency. Therefore, the discount will be granted on the basis of the following distribution function:

$$D = \begin{cases} 5\%, & FC > 14\ zile \\ 10\% * \frac{7}{SF}, & 5 < FC < 14 \\ 15\%, & FC < 5\ zile \end{cases}$$

D – personalized discount
SF – shopping frequency

A product should not be included in two consecutive offers, thus a *recommendation range* is defined, expressed in number of days.

$$R = SF * c$$

R – recommendation range
SF – shopping frequency
c – constant, defined by the company

The constant c is influenced by the diversity of the products and the company's business strategy. The model will consider c = 4, so a product will appear in no more than one recommendation per month.

The selection and filtering algorithm consists of three parts, one for each product category included in the recommendation. The model selects one date specific product, one product different from the shopping cart and N-2 shopping cart specific products. Taking into account the impact of opportunity cost, the number of products included in the final personalized recommendation is N = 6. The algorithm is executed only for one consumer at a time. The main steps of the model are summarized in figure 1.

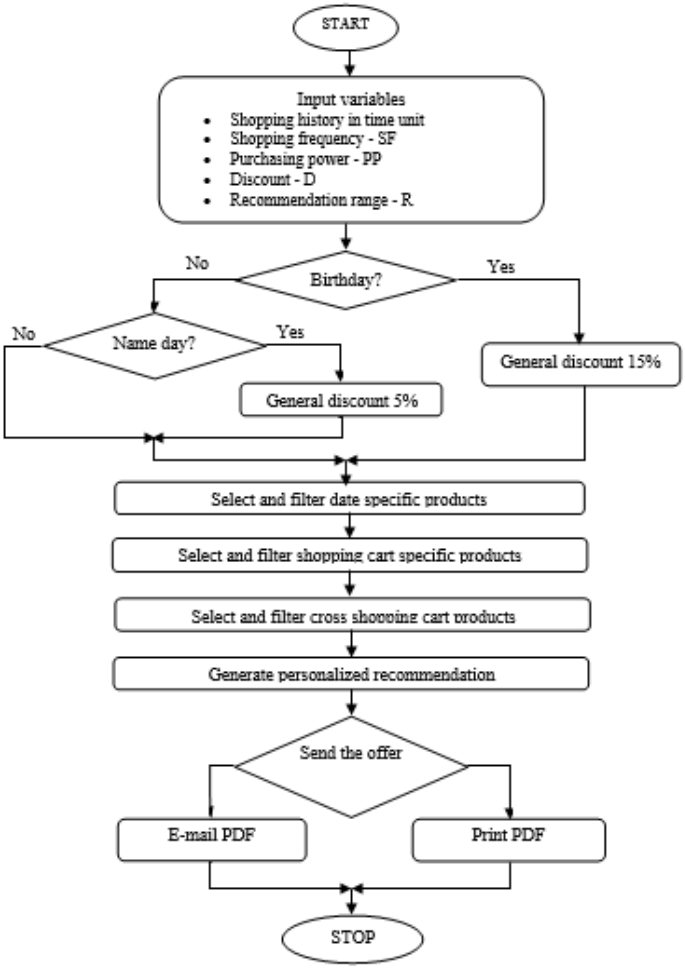


Fig. 1. The Logical schema of the recommendation model

4.1. Select and filter date specific products

STEP 1: check if the current date is included in a special period, like Christmas, Easter or St. Valentines. A custom table in the database maintains the special periods.

STEP 2: select all date specific products. Each product will be marked if it is specific to a period and each product will not be associated with more than one special period.

STEP 3: select all products previously purchased by the customer, regardless of the predefined time unit. This category of products is sold with an annual frequency so the time unit restriction is eliminated.

STEP 4: identify all date specific products purchased by the customer over time. If there is more than one product, the model will select the one purchased most often. If such a product could not be identified, then the date specific product is randomly selected.

4.2. Select and filter shopping cart specific products

Shopping cart specific products are also divided into three categories, identical products, similar products and complementary products, the algorithm having the same main logic for each subcategory.

The identical products are the products that were previously purchased by the customer. They are identified based on the purchasing history in the predefined time unit. For this category the price is not a filtering condition.

STEP 1: select all products previously purchased in the predefined time unit and remove one time purchased products which are considered spontaneous products.

STEP 2: remove the products that are no longer sold by the food retail store.

STEP 3: check if the products can be included in the recommendation according to the consumption period and the recommendation range.

A product can't be included in two consecutive recommendations and it can't be recommended before it is purchased and consumed. If the date of the last product recommendation summed up with the recommendation range exceeds the current date, or if the date of the last purchase cumulated with the consumption period exceeds the current date, then the product is removed from the selection.

The similar products are the products included in the same category as those previously purchased by the consumer. They could be identified by using an item based filtering techniques, but due to the diversity of the products the model will be slow and inefficient. So, the similar products are identified by CPV code. The Common Procurement Vocabulary (2008) establishes a single classification system for public procurement aimed at standardising the references used by contracting authorities and entities to describe procurement contracts. CPV codes are based on a branched structure of up to nine digits.

STEP 1: select all products similar to those already purchased in the predefined time unit. Select those products for which the first six digits of the CPV code is the same and the price category is the same as the consumer's purchasing power.

STEP 2: remove the products that are previously purchased because their selection has already been made using the identical algorithm.

STEP 3: remove the products that are no longer sold by the food retail store.

STEP 4: check if the products can be included in the recommendation. If the date of the last product recommendation summed up with the recommendation range exceeds the current date, the product is removed from the selection.

The complementary products are those products that are sold separately, but can be consumed together with other products, being in a direct association relationship.

STEP 1: select all products that are complementary to those already purchased in the predefined time unit, taking into account the price category. Select those products for which the price category is the same as the consumer's purchasing power.

STEP 2: remove the products that were previously purchased or the similar products because their selection has already been made using the above described algorithms.

STEP 3: remove the products that are no longer sold by the food retail store.

STEP 4: check if the products can be included in the recommendation. If the date of the last product recommendation summed up with the recommendation range exceeds the current date, the product is removed from the selection.

The model selects X1 identical products, X2 similar products and X3 complementary products, all being products specific to the consumer's shopping cart. A product will be uniquely found in one of three categories. The model selects randomly N-2 products to be included in the recommendation.

4.3. Select and filter cross shopping cart products

In order to select the products that are different from the shopping cart a customer based filtering technique is used by the model. To reduce the complexity of the model the customers are split in consumer clusters using as main rule of differentiation the purchasing power. The algorithm is applied only for customers from the same cluster.

STEP 1: select all customers who are part of the same cluster as the consumer for whom the model is running.

STEP 2: select all items purchased by each customer in the predefined the time unit. Determine how many times each product was purchased. Identify the common products purchased, taking into account only the products from the current shopping cart.

STEP 3: identify the customers who have purchased the most common products and select the different product that was purchased most often and is not already included in the offer.

STEP 4: remove the products that are no longer sold by the food retail store and check if the products can be included in the recommendation.

For example, the current bill contains products A, B, C, D, E. There are 4 consumers belonging to the same cluster and having the same purchasing power as the consumer for which the recommendation is generated. The consumer based filtering technique is illustrated in table 1.

Table 1. Consumer based filtering technique implemented in the model

Consumers	**C_1**	**C_2**	**C_3**	**C_4**
Products purchased in the predefined time unit	A_2 G_7 F_6 D_3	B_3 D_7 M_5 O_3 N_1	D_1 F_6 G_2 N_9	A_4 C_6 H_2 N_9
Common products	2 (A_2,D_3)	2 (B_3,D_7)	1 (D_1)	2 (A_4,C_6)
Degree of similarity	5	10	-	10
Cross shopping cart products	N_{10} M_5 O_3 H_2			

The model selects product N because it has been purchased most often by the consumers with the highest degree of

similarity. The degree of similarity between two consumers is calculated as the number of common products purchased by the two consumers. If product N is specific to a period, is specific to the shopping cart, it is no longer sold or if it has been included in a recent offer, then another different product that meets the filtering conditions is recommended.

5 Conclusions and future research

Due to the exponential growth of population and the rapid evolution of technology, consumers needs have become in the last centuries very diverse. Given this radical change of the economic environment where the bargaining power has shifted from sellers to buyers, companies need to adapt their business and marketing strategies in order to survive in a competitive environment.

The food retail market in Romania is becoming saturated, so the stores are realizing that keeping their actual clients is less expensive than attracting new ones and that gaining consumers loyalty have a big impact on their financial results. So, more and more companies are starting to use the newest technologies and software to increase customer satisfaction. This strategy can also be adopted by food retail stores.

The model described in this paper is based on the principles of recommendation systems originally developed for e-commerce companies. The hybrid algorithmic model based on data mining and collaborative filtering techniques generates real time personalized offers. The model can be used by the large food retail stores in Romania, as it analyses a wide range of products that have different characteristics. The complexity of the model may increase exponentially depending on the number of variables and techniques integrated, the ultimate goal being the recommendation of a limited number of products per consumer. The products are selected and filtered according to the consumer's specific buying behavior, and the final offer is personalized to the smallest details.

In the future, the model will be implemented in the SAP ERP program and tested on a real-life data base. The generated recommendations will be given to the trial participants in order to quantify the results of the model and to make the necessary adjustments.

References

[1] G. Miller, *Spent: Sex, Evolution and Consumer Behavior*, Penguin Books, 2009.

[2] R. R. Burke, *Innovative Uses of Video Data Analysis in Marketing: Measuring Shopper Attention, Engagement, and Purchase*, AMA-ECMI-EMAC Marketing & Innovation Symposium, Erasmus University, Rotterdam, 2014.

[3] Precima, *How important is price and promotion personalization to grocers and shoppers?*, article retrieved March 15, 2018, from https://www.precima.com/content/how-important-is-price-and-promotion-personalization-to-grocers-and-shoppers

[4] G. McVie, *Are personalized offers the remedy for ineffective trade promotions?*, LEAD Virtual Marketing Conference, 2017.

[5] F. O. Isinkaye, Y. O. Folajimi, B. A. Ojokoh, *Recommendation systems: Principles, methods and evaluation*, Egyptian Informatics Journal, 16(3), 261-273, 2015.

[6] J. Leskovec, A. Rajaraman, J. D. Ullman, *Mining of Massive Datasets*, 2nd Edition, Cambridge University Press, 2014.

[7] J. L. Erlocker, J. A. Konstan, L. G. Terveen, J. T. Rield, *Evaluating collaborative filtering recommender systems*, ACM Trans Inform Syst, 22(1), 5-53, 2004.

[8] C. A. Gomez-Uribe, N. Hunt, *The Netflix Recommender System: Algorithms, Business Value, and Innovation*, Transactions on

Management Information Systems, 6(4), no.13, 2016.
[9] E. A. Baatarjav, S. Phithakkitnukoon, R. Dantu, *Group Recommendation System for Facebook,* OTM: On the Move to Meaningful Internet Systems, 211-219, 2008.

Cristina STOICESCU graduated the Faculty of Economic Cybernetics, Statistics and Informatics, with a bachelor degree in Economic Cybernetics in 2012. In 2014 she got her masters degree from the same faculty of the Bucharest University of Economic Studies, specialization in Cybernetics and Quantitative Economics. Currently she is taking her PhD degree in Economic Cybernetics and Statistics at Bucharest University of Economic Studies, coordinated by Professor Crişan Albu. Her PhD thesis is "Experimental Economics in shaping consumer behavior".

Cristina STANCIU-MÂNDRULEANU obtained her bachelor's degree as an economist after graduating the Faculty of Economics Sciences, at the University of Craiova in 2001. In 2009, she completed her masters degree at the Bucharest University of Economic Studies - Faculty of Master Administration. Currently, she is taking her PhD degree in Economic Cybernetics and Statistics at Bucharest University of Economic Studies, coordinated by Professor Crișan Albu. Her PhD thesis is "Means and methods of analyzing the behavior of a company with application on the medical oxygen market".

Solutions for Big Data Processing and Analytics in Context of Smart Homes

Adela BÂRA[1], Bogdan TUDORICĂ[2], Razvan Cristian MARALES[3]
[1]The Bucharest University of Economic Studies
[2]Petroleum - Gas University of Ploieşti
[3] METRO SYSTEMS Romania
bara.adela@ie.ase.ro, tudorica_bogdan@yahoo.com, razvan.marales@csie.ase.ro

Abstract. *The paper analysis Big Data storage and processing solutions applicable in the smart grid context in case of large volume of data coming from sensors (smart meters, IoT appliances). A flexible architecture for data management is proposed that consists in three layers: relational database (RB) tier, big data (DG) tier and data warehouse (DW) tier. A proof of concept implementation is also provided, using Elasticsearch and Kibana.*
Keywords: Big Data, Smart Home, IoT, Smart Metering, Data Management

1 Introduction

In the recent years, due to the development needs of certain big companies activating in the digital industry, the big data processing techniques were developed very fast. As presented in [1], the main types of big data processing techniques are:

- batch processing, which was initially accomplished by using Apache Hadoop as a big data analysis platform for big data. Apache Hadoop is based on the MapReduce concept, which was initiated by Google's batch processing programming model. The main functionality of Apache Hadoop is successfully done by dividing a very large data set into multiple smaller data sets, which are then processed in parallel; then, the reducer job get the results from all these smaller data sets. As the main disadvantages of using Apache Hadoop and MapReduce are represented by the fact that MapReduce cannot be used for real time sensor data or streaming data processing, and that the set of machine learning algorithms provided by Apache Hadoop is not enough to meet requirements for smart grid big data analysis. Considering this facts, the authors do not recommend Apache Hadoop as a good choice for big data analysis on smart grid systems. Another big data processing solution, as presented in [2], is Apache Spark. Apache Spark is defined as a general purpose cluster computing platform, which delivers flexibility, scalability and speed to meet the challenges of big data, even for the case of using big data in smart grid analysis. Apache Spark has many advantages: not only has the ability for batch processing, but it also is capable of doing iterative and streaming processing, it has more efficient machine learning algorithms and enhanced linear algebra libraries.
- stream processing, which involves processing each new instance of data, rather than waiting for the next batch, and then re-process everything – this way the data processing technique avoids unnecessary repetition of re-processing the data. This model of data processing enables Apache Spark to perform analytics on data with dynamic behaviour – this feature is very important in the situation when data comes continuously, in real time, and from different data sources.

- iterative processing, represents another category of big data processing, and is mainly used to solve problems that cannot be addressed with batch or stream processing. One of the main characteristics is that it can process all variety of data types frequently – in general, the model is time consuming due to their continuous operations of writing and reading which represents each iteration of the system. Once again, Apache Spark is the current leader for iterative processing because it has the power to process and hold data in memory across the cluster, and the data is written back only after the completion of the iterative process.

The concept of Smart Home involves an efficient management of demand response regarding consumers' behaviour, integrating different residential generation sources and electric appliances through smart meters, providing user friendly applications and real time billing systems. In this context, ICT solutions should be developed to support a real time integration of heterogeneous sources gathered from smart meters, intelligent appliances and micro-generation sources. Papers [3-9] provide an overview of smart home and demand response concept and functionalities: responsible for an efficient connection and exploitation of generation and consumption sources, providing automatic and real-time management of the grid, optimizing the level of reliability and improving the electricity providers' services which lead to energy savings and lower costs. An important aspect in developing ICT solutions for demand response is the design of a scalable architecture for data management. At the current moment, a wealth of data storage technologies is available, based on different ideas and with appropriate advantages and disadvantages:

- Direct storage in a data file. A multitude of competing formats are currently available for storing data in text or similar files, such as Comma Separated Values (CSV), eXtensible Markup Language (XML), JavaScript Object Notation (JSON) or Yet Another Markup Language (YAML);
- Storage in a classic relational database accompanied by a complete and user-controlled database management system (DBMS). Examples of such databases are: Oracle Database, MySQL, MS SQL Server, PostgreSQL, IBM DB2;
- Storage in a document store type NoSQL database characterized by schema-free organization of data. Records should not have a uniform structure and may have different columns. Data types of individual column values may be different for each record. Columns can have multiple values (arrays). Records can have a nested structure [10]. Examples of such databases are: MongoDB, Elasticsearch, CouchBase, CouchDB;
- Storage in a key-value store NoSQL database that can store pairs of keys and values, and serves to recover a value when the appropriate key is known. These databases are not normally suitable for complex applications but, in certain circumstances, provide specific benefits [10]. Examples of such databases are: Redis, Aerospike or Oracle NoSQL;
- Storage in a wide-column store NoSQL database that involves the storage of data in records with the ability to contain a very large number of columns

that dynamically vary. The column names are also variable, the record keys are not fixed, and, since a record can have billions of columns, databases of this type can be seen as two-dimensional key-value matrices [10-11]. Examples of such databases are: Cassandra, HBase, Microsoft Azure Table Storage and Google Cloud Bigtable;

- Storage in a time series database optimized to manage time series data, each record being associated with a time stamp. Time series data can be produced by sensors, smart meters or RFIDs - IoT equipment, or can represent the values recorded in a high-speed stock exchange system. These databases are designed to efficiently collect, store and query different time series data. While these data can also be stored in other types of databases - from key-value stores to relational systems, specific challenges sometimes require specialized systems [11]. Examples of such databases are: InfluxDB, Kdb+, RRDtool, Graphite și OpenTSDB.

2 Big data solutions for Smart Home

2.1. Big Data processing solution

During the development of the SMARTRADE project, the success of an efficient data integration is related to the correct choice of the technology used to store and manage data. The nature of the project involves the use of a number of different data storage technologies, due to the stratified nature of the target application. It is assumed that the final prototype will contain at least the following levels of operation, virtually forming a stack of applications:

- Level 1 - Storage of data at the level of the IoT (Internet of Things) type systems that substantiate the acquisition and / or local centralization of data;
- Level 2 - The primary storage of data accumulated from the IoT systems and other sources, prior to any cleaning, aggregation, and analysis of data;
- Level 3 - Secondary data storage, used after data cleaning has been applied. Aggregation and analysis operations may be applied on the secondary storage data, whether we are talking about supervised analysis (common queries, but also classical statistical analysis), or unsupervised analysis (data mining operations);
- Level 4 - Tertiary storage of processed data / processing results, which is to be the integrated storage in a final consumption estimation application, resulting in competitive offers for the wholesale electricity market.

Each of the above four levels has distinct features that require the use of a particular storage technology among those listed in the previous section.

Storage at the level 1 of the application stack

The storage of data in IoT devices implies significant restrictions, both in terms of available storage space and usable processing capacity (doubled narrowed in turn by the limited performance of the processors used and the need for low energy consumption, which is encountered in many IoT devices). In [12] IoT devices are classified into four categories according to their complexity level:

- Extremely simple devices whose communications are limited to signalling their state in case of an event. Both the memory and the processing capacity of these devices are reduced to the minimum required for operation, and most of them do not

have proper storage memory. Some smart-metering devices from older generations could fall into this category;

- Slightly more complex IoT devices. The memory and processing capability of these devices is reduced, but it allows processing and storage not directly linked to the main function of the device. They also involve the transmission of the measured data in a single sense, but various forms of aggregation and representation of data are also possible. Many of the new generation smart metering devices belong to this category. For this IoT device type, some data storage is available at least for logging. It is also important to note that many such devices are proprietary devices and one can't interact directly with the software installed on them or with storage media and can't choose the data storage technology used. As a solution to this situation, the use of modular equipment and / or equipment fully or partially based on open technologies is recommended. In this category of devices, if it is possible to control the storage, it is most likely that only data files can be used [13] because either the address space is too small or the memory is not large enough to implement a query mechanism;
- Interactive devices capable of bi-directional communication. Additionally, such devices offer programmability and many of the capabilities of personal computing systems (multiple storage systems, multiple inputs and outputs, both analogue and digital). Use of data files and embedded relational or NoSQL databases is recommended for this category of devices [13];
- Intelligent devices with high processing and storage capabilities raised to the level they even allow running artificial intelligence applications. On this type of devices, it is possible to run non-embedded, relational or NoSQL databases [14].

Storage at the level 2 of the application stack

The method chosen for the storage of data accumulated from IoT devices depends specifically on the number of IoT devices from which data is collected and the data traffic density for each device (the frequency with which data is transmitted multiplied by the size of each transmission). Under a certain number of IoT devices from which data is collected and under a certain density of data traffic per device, it is perfectly possible to use any type of commonly used database, being preferable, in this context, to use relational databases. The motivation for this choice is given by the superior data query facilities offered by these products.

When a critical level is reached, which is mainly dependent, as highlighted above, on the number of devices from which data is collected and the density of data traffic per device (a number of other factors need to be taken into account, such as data communication system capabilities and temporal overlay of data transmissions from multiple IoT devices), the use of relational databases becomes undesirable, for reasons of limited performance.

The limitations given by the use of relational databases are quite well known. For example, MS SQL Server supports theoretically 32767 competing connections over a cluster of servers, but in real-world applications, their number is severely limited by the computing system used for the database server and the data communications system, reaching, for example, 50-200 competing connections per server. A similar situation is encountered in MySQL where the maximum recommended number of concurrent connections per server instance in version 5.5 is 500 (with an average of 100 competing connections) for a transactional usage scenario. The same situation can be seen in Oracle Database 12c - a maximum of 2000 concurrent

connections to a database instance. On the other hand, the initial arguments that led to the success of NoSQL solutions on the market, the volume of data circulated and its velocity (to which, subsequently, variety and veracity were added to form the 4 V's of NoSQL databases), are indeed visible even at a superficial glance. For this reason, for any data collection system that involves large amounts of data, a NoSQL database is the recommended solution. One has to determine only the NoSQL database subtype to be used, depending on the proposed application architecture - the three main types of NoSQL databases are available (document stores, key-value stores, and wide columns stores), but also some less used types can be taken into account. Among the less commonly used database types, for smart-metering applications, the databases specialized in the storage of time series are very important.

Storage at the level 3 of the application stack

For the secondary data storage, the same considerations, from the point of view of volume and velocity of data, that imposed the solution at the second level of the application stack, should be considered. Even though the number of data sources is significantly reduced, the data volume is probably the same order of magnitude as the one at the level 2 of the application stack.

Storage at the level 4 of the application stack

If a separate application is preferred, as a consumer of the results of the analyses performed on the data generated by the IoT devices, such an application is likely to be only a small application, limited both as complexity and as the volume of used data - it will not have as inputs the actual data which is generated, collected and analysed at the other levels, but the power generation and consumption profiles resulting from these data, in conjunction with other types of information (e.g. prices from electricity producers and transporters, etc.). For such an application, the need to use a NoSQL storage solution is no longer stringent, although it can be considered for various practical reasons. The implementation for a complete big data solution involves the combination of multiple data storage technologies, for the various operating levels specified above. None of the existing data storage technologies meet the requirements of all these levels and it is necessary to select a separate solution for each of them. Particular attention should be paid in this context, not only to the data solutions used per se, but also to the modalities found for the transfer of data between these solutions.

2.2 Big Data Analytics

In the paper [15] are presented the most important tools, adopted for forecasting big data, and the following techniques are mentioned: Factor models, Bayesian models and Neural Networks. The paper [16] also illustrates the importance of big data, by the fact that Big Data and predictive analysis goes hand in hand in the modern age, with companies focusing on obtaining real time forecasts using the increasingly available data. The importance of big data analytics derives the need of developing the existing data mining techniques to meet the requirements, as stated in paper [17]. The author states that there is a need for the adoption of powerful tools, such as Data Mining techniques, which can aid in modeling the complex relationships that are inherent in Big Data. To support the need for more powerful tools, the paper [18] states that the process of forecasting big data is very hard, and identifies one important reason why the process gets so complicated, as the traditional forecasting tools cannot handle the size, speed and complexity inherent in Big Data. These characteristics of big data are than completed in the paper [19], where the author also identifies another important factor of the problems arising in big data

when forecasting big data, as the lack of a structure in these data sets and the size.
The main challenges when using big data for forecasting were identified in the paper [15], as follows:

- the skills required for forecasting with Big Data and the availability of personnel – as presented in the paper [20], developing these skills can be a big challenge. The authors recommend that in order to overcome this issue, Higher Educational Institutes should upgrade the syllabuses to incorporate the skills necessary for understanding, analyzing, evaluating and forecasting with Big Data;
- the noise that is likely distorting the signal and the accuracy of the forecast is another extremely important challenge in Big Data forecasting as identified by [21];
- new techniques should be developed for filtering the noise in Big Data to enable accurate and meaningful forecasts;
- hardware and software requirements for forecasting big data are another great challenge of big data, as is also stated in the paper [19].

Paper [22] presents statistical significance as a major threat in big data forecasting accuracy, as is very possible to make false discoveries from Big Data. Supporting this idea, in the paper [23] is observed that given the quantity of data to be processed, and forecasted, is a very complex task to differentiate between randomness and statistically significant outcomes;
Architecture of algorithms and the Data Mining techniques have been designed to handle data of much smaller sizes, as opposed to the size of Big Data is also considered a high challenge and in the paper [24] is contained a detailed evaluation of challenges, associated with the application of Data Mining techniques to Big Data;
In paper [15], the authors also consider Big Data itself a challenge for big data forecasting, because of its main characteristics: evolves and changes in real time; includes unstructured data; highly complex in structure.

2.3 Configuration of the Big Data Management solution

The proposed data management solution will be developed as a scalable and customized framework using open solutions based on cloud computing. The data framework will contain several tiers that can be customized. These tiers will allow complete customization, using different technologies for interconnecting smart meters and sensors (*SM/IoT tier*); fast processing and real-time analyses through big data (*BG-Tier*); high-reliable, low-latency, secure and performant data transmission and management with relational databases (*RB-Tier*); historical and advanced analytics through data warehouse (*DW-Tier*). These tiers are illustrated in Figure 1.

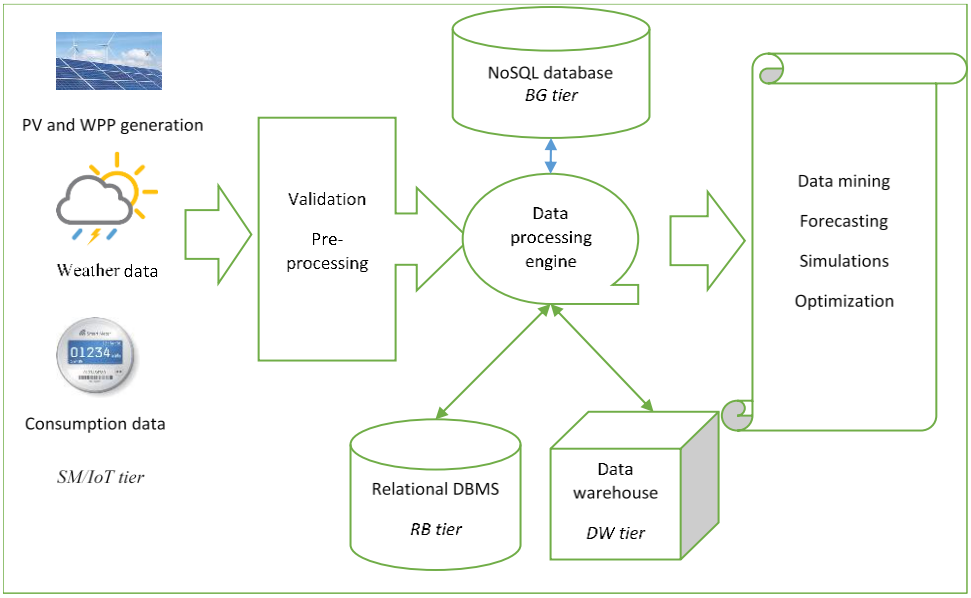


Fig. 1. Data management solution for SMARTRADE project

3. Proof of concept implementation

For the proof of concept we implemented a scalable system using different technologies. The system is able to retrieve data about electricity consumption from residential consumers and send it to the data processing engine, where the data is validated and stored in a persistent database. Based on the final data, the system can perform various calculations that will support the purpose of forecasting and optimisation.

For simulations we used a TP LINK HS110 smart plug connected to electric appliances, which is able to provide information such as the current consumption, power or voltage in real-time. The smart plug is communicating through WiFi with a Raspberry PI 3 mini-controller. The mini-controller is running on a Raspbian Linux distribution having a NodeJs application inside that is querying data about energy consumption from the smart plug at a fixed interval and distributes it outside the house to a messaging queue. Each time the NodeJs application is making a query and receives a valid response, it is seen in the system as an event. The event is basically created in the house and it propagates in the system until it is stored in the database.

We decided to use the messaging queue and the publish/subscribe pattern to ensure the speed and scalability of our application. From the message queue, the information is consumed by a Java application that is deployed in a public cloud. The application has the role of pre-processing the information, check the consistency and validate it. In the end, if the data is valid, it gets stored in an Elasticseach database on a specific index.

To have a cleaner view on the data in the system, we integrated Kibana in our database. Kibana is a Elasticsearch visualization tool that extends the database capabilities by providing new features such as UI, charts and graphs, real-time monitoring and alerts based on queries. To understand the data in real-time or during a certain period of time better, we implemented Kibana visualizers. As an example, Figure 2 presents two charts regarding energy consumption - the average (left-side) and the sum (right-side) of energy consumption monitored in one day. Using the visualizers, we were able to find the consumption patterns of the consumer. In Figure 2 two important peaks during the day are spotted - one in the morning and the other one in the evening.

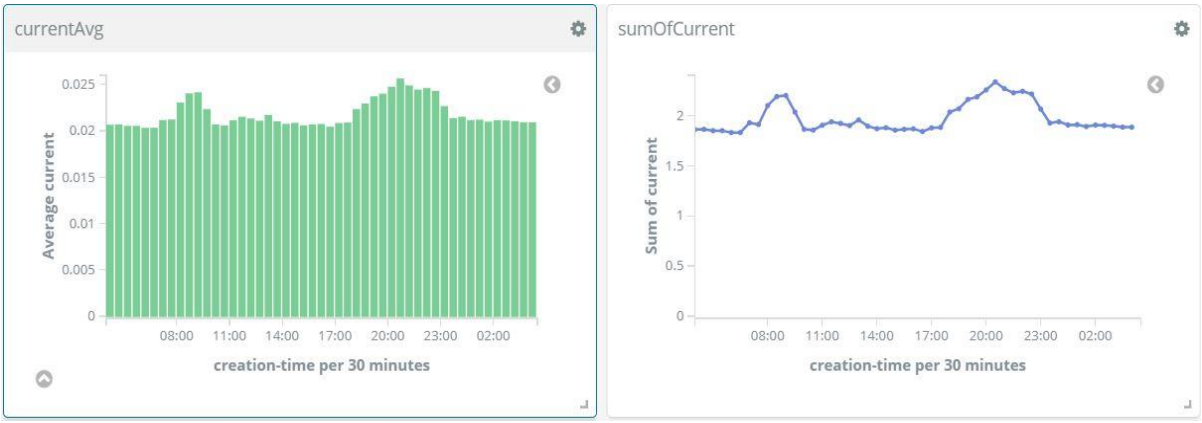


Fig. 2. Kibana visualization of one day energy consumption

In the proof of concept we implemented two tiers described in the paper, the SM/IoT tier and DB tier. In the future we are planning to introduce also the other two tiers in order to develop a complete scalable prototype.

4. Conclusions

The integration of IoT electric appliances in current development of the Smart Home concept require a scalable data management solution. The paper presented an overview of the Big Data solutions that are suitable for demand response management and proposed a scalable architecture implemented in the SMARTRADE project that consist in several tiers for data storage and processing. Two of these tiers are partially implemented during the project and some results were presented in the paper as a proof of concept.

Acknowledgement

This paper present scientific results of the project "Intelligent system for trading on wholesale electricity market" (SMARTRADE), co-financed by the European Regional Development Fund (ERDF), through the Competitiveness Operational Programme (COP) 2014-2020, priority axis 1 – Research, technological development and innovation (RD&I) to support economic competitiveness and business development, Action 1.1.4 - Attracting high-level personnel from abroad in order to enhance the RD capacity, contract ID P_37_418, no. 62/05.09.2016, beneficiary The Bucharest University of Economic Studies.

References

[1] R. Shyam, G. Bharathi, K. Sachin, P. Prabaharan, K. Soman Pa - "Apache Spark a Big Data Analytics Platform for Smart Grid", Procedia Technology, 21, 171-178, 2015

[2] M. Zaharia, C. Mosharaf, J. Michael, S. Scott, I. Stoica - "Spark: cluster computing with working sets", Proceedings of the 2nd USENIX conference on Hot topics in cloud computing, 10-10, 2010

[3] W. Wang, Z. Lu, "Cyber security in the smart grid: survey and challenges", Comput. Netw., 57, 13, 44-71, 2013

[4] M. McGranaghan, D. H. L. Schmitt, F. Cleveland, E. Lambert, "Enabling the integrated grid: leveraging data to integrate distributed resources and customers", IEEE Power Energy Mag., 14, 83-93, 2016

[5] V. Agarwal, L. H. Tsoukalas, "Smart grids: importance of power quality", Proceedings of first international conference on energy-efficient computing and networking, 13-15, April 2010, Berlin, 136-143, 2010

[6] S. M. Amin, "Smart grid: overview, issues and opportunities. Advances and challenges in sensing, modelling, simulation, optimization and control", Eur. J. Control, 17, 547-567, 2011

[7] Y. Yan, Y. Qian, H. Sharif, D. Tipper, "A survey on smart grid communication infrastructures: motivations, requirements and challenges", IEEE Commun. Surv.

Tutor., 15, 5-20, 2013
[8] SV. Oprea, B. Tudorică, A. Belciu, I. Botha "Internet of Things, Challenges for Demand Side Management", Informatică Economică, vol. 21, no.4/2017
[9] A. Florea, V. Diaconita, I. Dorobat, Business process modeling for sales processes automation, 15th EBES International Conference, 8-10 January 2015, Lisbon, Portugal, 2015
[10] S. Edlich, List of NoSQL databases". NoSQL - Your Ultimate Guide to the Non-Relational universe. Available http://nosql-database.org/ [December, 2017]
[11] DE-Engines.Com, Ranking of Relational DBMS. Available: https://db-engines.com/en/ranking/relational+dbms [December, 2017]
[12] B. Jones, The 4 Device Types in the Internet of Things, from a Data Perspective. Available: http://www.datasciencecentral.com/profiles/blogs/the-4-device-types-in-the-internet-of-things-from-a-data [December, 2017]
[13] B. Cole, XML - the embedded industry's not so secret weapon? Available: https://www.embedded.com/electronics-blogs/cole-bin/4217831/XML-the-embedded-industrys-not-so-secret-weapon [March 2018]
[14] G. Carvalho, Raspberry as a dedicated server? The result is amazing. Available: https://www.copahost.com/blog/is-it-possible-to-run-a-web-server-in-a-raspberry-pi-3-as-a-dedicated-server/ [March 2018]
[15] H. Hassani, E. Sirima Silva - "Forecasting with Big Data: A Review", Ann. Data. Sci., 2(1), 5-19, DOI 10.1007/s40745-015-0029-9, 2015
[16] D.Bernstein - "Big data's greatest power: predictive analysis" (2013) http://www.equest.com/cartoons/cartoons-2013/big-datas-greatest-power-predictive-analytics/ [March 2018]
[17] H. R. Varian - "Big data: new tricks for econometrics", J. Econ. Perspect, 28, 2, 3-28, 2014
[18] S. Madden - "From databases to big data", IEEE Internet Comput., 16, 3, 4-6, 2012
[19] D. Arribas-Bel - "Accidental, open and everywhere: emerging data sources for the understanding of cities", Appl. Geogr., 49, 45-53, 2014
[20] L. Einav, J. D. Levin - "The data revolution and economic analysis", Working Paper No. 19035, National Bureau of Economic Research, 2013
[21] N. Silver - "The signal and the noise: the art and science of prediction", Penguin Books, Westmins, 2012
[22] S. Lohr - "The age of big data" (2013) http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html [March 2018]
[23] B. Efron - "Large-scale inference: empirical Bayes methods for estimation, testing and prediction", Cambridge University Press, Cambridge, 2010
[24] H. Hassani, G. Saporta, E. S. Silva - "Data mining and official statistics: the past, the present and the future", Big Data, 2, 1, BD1-BD10, 2014

Adela BÂRA (b. October 11, 1978) is Professor at the Economic Informatics Department at the Faculty of Cybernetics, Statistics and Economic Informatics from the Academy of Economic Studies of Bucharest. She has graduated the Faculty of Economic Cybernetics in 2002, holds a PhD diploma in Economics from 2007. She is the author of 15 books in the domain of economic informatics, over 50 published scientific papers and articles (among which over 20 articles are indexed in international databases, ISI proceedings, SCOPUS and 16 of them are WoS indexed). She participated as team member in 5 research projects and has gained as project manager 4 research grants, financed from national research

programs. Domains of competence: Database systems, Big data, Business Intelligence, Decision Support Systems, Data Mining & Analitics, Artificial Neural Networks.

Bogdan TUDORICĂ (b. July 10, 1976) is Lecturer at the Petroleum-Gas University of Ploiesti, Faculty of Economic Sciences. He has graduated the Faculty of Letters and Sciences of Petroleum-Gas University of Ploiesti in 1998 and holds a PhD diploma in Economic Informatics from 2015. He is author or co-author of 2 books in the domain of economic informatics, over 20 published scientific papers and articles (among which over 17 articles are indexed in international databases or ISI proceedings). He participated as team member in 3 research projects and 10 development projects. Domains of competence: Database systems, Big Data, Computer networks, Programming, Data Security.

Razvan Marales, Java Developer. METRO SYSTEMS Romania with interests in IoT, Cloud and Microservices. Currently, he is PhD student at the Bucharest Unoversity of Economic Studies, in the fied of Economic Informatics.