Universidade do Minho
Escola de Engenharia

David José Farinha Cuevas

Intelligent systems for energetic sustainability

Setembro de 2012

Universidade do Minho
Escola de Engenharia
Departamento de Informática

David José Farinha Cuevas

Intelligent systems for energetic sustainability

Dissertação de Mestrado
Mestrado em Engenharia Informática

Trabalho realizado sob orientação de
Cesar Analide

Setembro de 2012

Acknoledgements / Agradecimentos

O culminar de um trabalho desta dimensão é sempre uma excelente oportunidade para homenagear e agradecer a todas as pessoas que, de uma forma ou de outra, contribuíram para que a vida profissional, a aprendizagem, a camaradagem e a troca de experiências sejams empre rodeadas da melhor convivência e de menores dificuldades.

Iniciando, agradeço muito sinceramente ao Professor Cesar Analide, por me ter aceite neste projecto de braços abertos, fornecendo toda a força necessária para iniciar um projecto desta dimensão, enquanto a minha falhava.

Ao Fábio Silva, pelo grande companheirismo neste projecto, cuja supervisão e conselhos assim permitiram o desenrolar deste trabalho quando as dúvidas persistiam com mais frequência.

Ao Luciano Alves e Tiago Sá, pela grande amizade e camaradasgem, sempre presentes, que me fez conseguir ultrapassar todas as dificuldades escolares que senti no início do ano.

À minha mãe, por todo o esforço que tem feito para me manter nesta instituição, assim como nos estudos, de forma a nunca perder o sentido da vida e dos objectivos profissionais, e por todo o afecto que me proporciona para continuar a lutar.

Aos meus avós, Zeca e Júlia, um grande beijinho de agradecimento por sempre me quererem bem e pela nossa ligação, pelos conselhos e pela força necessária para ir em frente!

Às minhas grandes tias Tó, Gina, Virgínia e Antónia, aos meus enormes tios João, Nelo, Mário e Eduardo, às minhas primas, Maria João, Sara e Sílvia, aos meus primos, Paco, Litos, Tiago (Rocha), Marito (Rocha), Tiago (Alves), Dudu e Fred, um muito obrigado por serem a familia espetacular que são, e por me fazem sentir apoiado seja em que circunstância for.

À Dalila, Víctor e Catarina, porque, dada a distância, sei que vocês também puxaram pelo meu sucesso e pelo meu bem estar, e agradeço também, a vossa excelente companhia e afecto.

Aos meus colegas e professores do ISLAB, agradeço por todo o apoio que me deram, durante e após a minha recomposição.

À Andreia, pelo carinho, amor, amizade, força, e encorajamento que me fizeram tomar um rumo quando apenas via o futuro à deriva. Foste imprescindível, e não há palavras para agradecer o quanto fizeste por mim este ano.

Por útlimo, agradeço a muitos outros amigos, que deixo no anonimato por não querer correr o risco de esquecer algum.

A todos, um muito, mas muito sincero obrigado!

Para o meu pai

(Sei que brilharás sempre para mim...)

Resumo

Hoje em dia, a sociedade ainda possui uma grande tendência para um consumismo desleixado e irracional, especialmente nas nossas casas, levando assim a consumos desnecessários de energia, e consequentemente, ao dano do nossos sistema ambiente.

É pretendido, com este trabalho, desenvolver um sistema inteligente de gestão de consumos energéticos e baseado na aprendizagem das acções das pessoas que interagem com um determinado espaço. Este sistema irá actuar sobre uma rede de sensores, de onde é extraída toda a informação necessária para a aprendizaegm e gestão dos recursos. Com isto, é possível saber, de uma forma aproximada, que ações é que os utilizadores tomam, como por exemplo, a que horas do dia é que um dado utilizador entra num dado espaço. Através dos sensores ambientais, como por exemplo, um sensor de luminosidade, poderemos também determinar quais as condições necessárias para saber também quando é que um dado utilizador vai desligar ou ligar um interruptor ou até ligar um ar condicionado. Desta maneira, é possível conseguir adaptar o espaço que está a ser monitorizado, de modo a ser possível obter o melhor consumo energético e com isso tornar esse mesmo espaço um ambiente sustentável.

Esta solução deverá também informar o utilizador sobre que electrodomésticos (ou dispositivos que estão a ser monitorizados) são os mais consumidores (também deverá ter a percepção de qual é o que gasta menos) de forma a este poder regular o seu consumo numa fase inicial e posteriormente permitir ao sistema aprender.

Abstract

Nowadays, society stills possesses a great tendency to a slouch and irrational consumerism, specially at homes, leading the people to unnecessary power consumes and consequently, harm our natural environment and even leading also to expensive power bills.

It is intended, with this work, to develop an energetic resources management intelligent system, based on people behaviours that interact with a given physical space. This system must work alongside with a sensors network, from where important information can be collected, in order so that the system can learn. It is through this network that we can collect the information to be learnt (human behaviours and environmental values). This way, we can know, in an approximated way, which user's actions have more influence in the electric consume and in the natural environment. Through environmental sensors, we can, for instance, to know the hour of the day which a given user have more tendency to spend more electric power by using an air conditioner appliance. This way, it is possible to adapt and evolve the system to better fit the power consume and turn the physical space a sustainable environment.

This solution should not only warn the user about which electric appliances (or monitored devices) are the most power over-consuming but also warn him which is the less-consuming one, allowing the user to better regulate their consumes in an initial phase and let the system learn.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Ambient Intelligence (AmI) is a recent information technology paradigm, in which people become a part of a digital environment that is aware of their presence and context, sensing, adapting and responding to their needs [15].

These systems are characterized by their ubiquity, transparency and intelligence, where a given user must adopt his natural role, interacting normally with the environment. This user should not in any way, configure or program any device to achieve some kind of functionality. The intelligence must be transparent to the user, blending into the infrastructures (rooms, halls, laboratories and similar infrastructures) and then automatically the system will learn everyday behaviours based on the user's common tasks. There are some difficulties when we start to implement similar projects like this one, since:

- People are naturally non-deterministic and every person is highly individual, and, this way, are very unpredictable;

- The use of embedded devices (such as computers with some weak processors and memories) brings the lack of computation power and battery energy [15], needed to process large amount of data;

- Most of the times, spaces aren't populated by only one person, and the identification of each one of them becomes hard. Individual identification, control and track are important in order to better serve each person, but this implementation is also hard to do [20].

Since around 40% of our homes main source of energy is the electricity, and there has been registered a continuous increasing of the consumption of this source. This increment is directly associated to an increase on thermal comfort, and a growing number of electrical appliances in households.

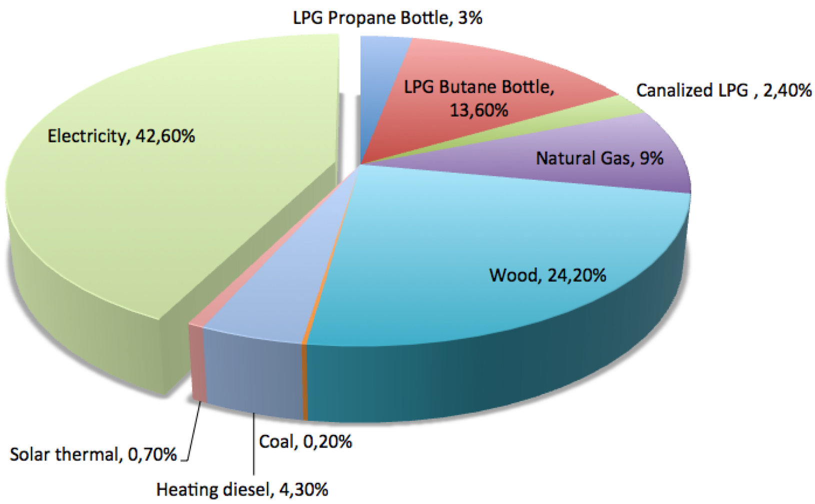

Figure 1.1: Distribution of energy consume in Portuguese homes, for each type of energy source - Portugal, 2010 [12]

More surveys have been made, and the statistics said that about 40% of the electricity spent in our home, are consumed in the kitchen. This number is sort of obvious, since that it is in the kitchen that are located mostly of the electric appliances, like microwave ovens, stove, fridge, and so on [12].

Given the great amount of electricity consumed in our homes, sometimes owners have some difficulties in predict their electric consumption (since they don't possess forecast abilities) in a given time interval. They don't know, most of the times, if their actual consumption profile is the most suitable for

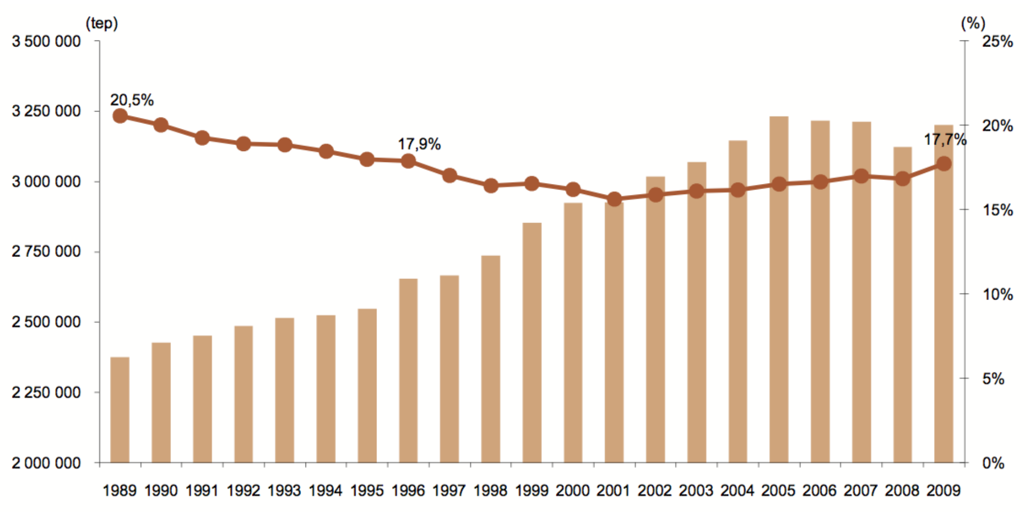

Figure 1.2: Consumption evolution on the domestic sector (tep) and weight (%) of the consumption of the domestic sector on the energy final consumption, 1989-2009 [12]

them and for the environment. This problem will also be approached in order to help users manage their control in a interval of time defined by them.

Since the great keyword of this work is: "Sustainability", it makes logic and is important to explain what's the meaning of this word. "Sustainability" can be defined as the development that meets the needs of the present without compromising the ability of future generations to meet their own needs [21].

Another great keyword that is directly associated to "Sustainability" keyword is the "Consumption" keyword. The definition proposed by the 1994 Oslo Symposium on Sustainable Consumption defines it as "the use of services and related products which respond to basic needs and bring a better quality of life while minimizing the use of natural resources and toxic materials as well as emissions of waste and pollutants over the life cycle of the service or product so as not to jeopardize the needs of future generations." [1]

The production and consumption system is the aggregate of all economic activity that guides the provision of goods and services that move through our lives and sustain us. It emerges from our needs and values, which determine what we produce and consume, and how much. This system is unsustainable on both social and ecological fronts, as the way we currently extract resources, transform matter and energy into products and services, and create waste is undermining lives and livelihoods and our planet's ability to sustain life [22].

Our society has become more consumerist than sustainable, and with that evolution (in an non-positive way), people doesn't care about the Environment and people's health and Environmental issues is getting worse and worse. Most of the energy supplies that our society use (to power their machines) is astonishingly centred on oil-derivable fuel and electric power (As shown, for example, in Figure 1.3). Besides this last energy source is a cleaner one, its use is not made from an efficient way, and there are few systems that helps people understanding their consumes and understand how can their acts and behaviour change and save the planet.

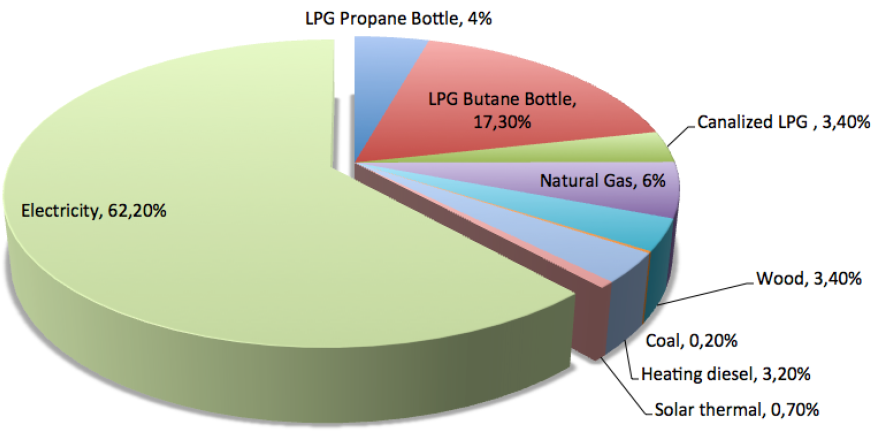

Figure 1.3: Average distribution of the spent energy at home, by type of source - Portugal, 2010 [12]

As it can be seen in Figure 1.3, the most spent energy at our homes, is the electricity, with a huge majority facing the others (About 60% of the total consume in a whole house is electric power). This means that electricity is the most used power in each home, and this means that this is the source that is more propitious to mismanagement, from older people, to the younger ones.

It is a motivation for this work, the use of the ISLab space. ISLab is located on Universidade do Minho (UM), more precisely at the Departamento de Informática (DI). This laboratory is equipped with an air conditioning appliance, working tables and seats for about 15 people. It has plenty of windows, turning the room very clear, and that windows have also their respective blinds in order to regulate the luminosity that enters in the room. The ISLab is usually crowded, becoming one potential good space to test our system. This is the motivation for this work is to teach and help people where they can save their power consumption and so, make our environments more Sustainable.

1.2 Main goals

The main objective of this work points to the development of a software capable to automatically manage and adapt energetic resources use depending on the people that inhabits a given house or a specific room in one building or space.

This software must study people's behaviours (in a non-intrusive way), inform people with the current plan of consumption, display important information about the devices in the controlled environment (on/off consoles, light lamps, open doors, wireless routers, computers, electric plugs, radiators, and others) and even forecast one report so that users can be aware of how much do they spend if they keep their actual behaviours. This system must be the most transparent as possible (in order to not influence people's behaviours) and help people understand which consequences will their behaviours provoke (in an environmental, energetic and economic way).

Another objective of this work is also to help people to practice the best "green" economy in these difficult times (world-wide crisis) that we are crossing nowadays. This dissertation has some development phases, that have to be decided, and they are shown in the list below, better organizing this dissertation. Splitting this work by steps, we have:

1. Collection of information through sensors such as luminosity, temperature, person presence, open doors, and others;
2. Implementation of a system based on the reviewed bibliography and on the exposed problem;
3. Test of a solution in a real physical environment, such as a laboratory room;
4. Processing of the application over obtained results in order to find some relevant conclusions;
5. Writing of Master's dissertation.

The first step item refers to the gathering of sensors values. This collection will form a values database that can be accessed by the application, to study and create profiles, corresponding to users behaviours. There are going to be settled sensors in a given study environment, such as laboratory or room, and data will be collected as many as possible, in order to create that information standard.

This information will be processed by a tool named WEKA [1]. WEKA has a collection of algorithms for data mining tasks, whose algorithms can either be applied directly to a given dataset or called from our Java code. Since WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization [3], it makes the perfect program to better understand all collected data than eye-sight interpretation.

1.3 Document structure

This dissertation is composed by a total of six chapters being the first one, the present chapter (Introduction 1).

[1] WEKA website: `http://www.cs.waikato.ac.nz/ml/weka`

This actual chapter is divided into 3 sections. In the first subchapter (Motivation 1.1), it is supposed to present the reader with some information about some motivation for this work, the second subchapter (Main goals 1.2) explains the main objectives this work will try to achieve and the last section (Document Structure 1.3) only means to explain the structure of this document and what will be told in every chapter and subchapter.

The second chapter (State of the art 2) contains the overview of the work done until now, in the same area than this work, and it is divided into 3 sections. The first section (Previous Work 5.1) resumes some of the first work that was developed in the beginning of AmI, the second section (More recent work... 2.2) refers to some work more recent than the ones presented in the previous section, and, for last, the last section (Technologies nowadays 2.3) has the meaning to present some technologies that are being developed nowadays and that could be possibly useful to some work of this kind.

The next (third) chapter (Contextualization 3) has the objective to explain some concepts that are going to be used and approached on this work, informing him so that he understands easily the nature of this work. This chapter is divided into 5 sections, and 6 subsections. The first section (Sensing 3.1) explains this new word and his meaning. The second section (Multi-Agent System 4.4) presents some concepts of Multi-Agent systems, their advantages and functionalities. The third section (Environment 3.3) explains briefly how can a Environment be useful to a AmI. The fourth section (Machine Learning 3.4), details the theory around the Machine Learning techniques, algorithms, and for what are used. The last section (Ubiquity 3.5) will explain this concept, and also explain this concept in the ISLab context.

The fourth chapter (Development 4) has the objective to explain the different phases of this work, the taken decisions, hardware choices and present some interfaces resultant from the developed system. This chapter is divided into 6 sections and 13 subsections. The first section (Data types 4.1) explains, the planning of the work, in some variables matter, in order to decide how to orient the development. The second section (Hardware 4.2) means to explain what happened along the development, in the matter of hardware, their problems and also solutions. The next, third chapter (Raw data 4.3) will explain some decisions taken in order so that data sets can be built and then start to process data and extract knowledge. The fourth section (Multi-Agent System 4.4) will explain the development of the multi-agent system and also present a general overview of the multi-agent of this work. The fifth section (Testing WEKA 4.5) has the objective to show how algorithms were chosen, tests made over the collected datasets and also the description of the Data Mining processes taken. The last and the sixth section (The application 4.6) will show some choices and thoughts about the GUI developed, changes and the lately developed platform.

The fifth chapter (Results 5) is divided into some 3 sections. The first section (Previous work 5.1) has the meaning of resume the work done, the second section (System Performance 5.2) will explain the improvements of the system in order to achieve better performance, in memory matter, and in computing performances, and the last and third section (Web Integration **??**sec:webIntegration)) informs the user that the beginning of a web platform was initiated.

The sixth and last chapter (Conclusions 6) has the purpose to settle some final considerations about the work done. This chapter is divided into 3 sections, and the first one (Synthethis 6.1) will resume and classify the work done, the second section (Relevant work 6.2) will expose some contributions that this work provided and, lastly, the third section (Future work 6.3) will suggest some future contributions in order to complete this work, at other levels.

Chapter 2

State of the art

Psychologists have already shown that with proper feedback, real time-information at salient times, and goal setting abilities, households can have up to 10% energy savings with small changes in their behaviours [18] [6] [25]. Yet most current methods of tracking resource consumption, such as via utility bills, remain inadequate [24]. Commercially, consumer devices that determine the electrical consumption of appliances—such as the Kill-A-Watt or Watts Up—are available today but these devices only track individual appliances, not total household energy use [1] [2].

Nowadays, people just want to satisfy their needs by playing video games, watching television or even by relaxing comfortably on heated bedrooms (using electric heaters or diesel burning systems). However, everyone has simple and characteristic behaviours while playing, relaxing on the living room, studying on the bedroom, or just by walking on a hall that may have consequences for the electric consumption. For example, in a hall, a child may leave all the lights on, just by slouch.

Another example is when someone turns on an electric heater or diesel burning heaters in the middle of the afternoon, to have his room or kitchen warm when he comes from work. Those behaviours, intentional, or irra-

[1] Kill-A-Watt main website: `http://www.p3international.com/products/special/p4400/p4400-ce.html`

[2] Watts up main website: `https://www.wattsupmeters.com/secure/index.php`

tional, may lead to an excessive and unnecessary consume of energetic or fuel resources (such as diesel or gas), translated to excessive loss of money and, thinking in a world-wide scale, the whole natural environment will be harmed due to the amount of pollution that can be generated [27].

These problems can be solved and avoided through the use of current technologies, like the ones that this work needs. There is a growing interest in designing smart environments that reason about residents, and this interest reflects on works as in [10] and [13],

2.1 Previous work...

There has been developed some works in the AmI area, some years ago, and this section will present some of that work.

One of the oldest approaches started in the late 90's, when the Massachussetts Institute of Technology (MIT) had started Ambience Intelligence researches known as the "Intelligent Room" [9]. At the University of Texas at Arlington (UTA), a multi-agent architecture approach has been proposed for controlling smart homes in the Managing an intelligent Versatile Home (MavHome) project [11]. This project is focused on finding and modelling sequential routines of user activities using prediction algorithms. Based on the discovered routines, control policies can be formulated to assist user activities in a proactive manner [5]

Besides these systems above described, there is another curious project: the Essex's iDorm, which is a intelligent system that monitors a dormitory and has the goal of learning the behaviour of its users. The background agent discretely controls the system according their users needs. These autonomous entities have usually a network connection associated, facilitating, this way, the communication and the cooperation between several other agents, forming a multi-agent system. These agents can also be adapted to mobile robots, giving them a new navigation perspective of the monitored environment. The space (in this case study, a bedroom) is equipped with a set of sensors that allows the system to recognize every action or behaviour of the users, even if they are lying on the bed or sitting on the desk's chair This system uses a electronic key about the size of a penny to distinguish different users differently on the managed space. [15].

2.2 More recent work...

The evolution of the technology has been crucial to the development of the AmI in the matter of, for example, some years ago, computers were not only too complex to work and use, but expensive as well. It was also a very precious and shared resourced, making it difficult to advance in technologies.

Nowadays, we can find computational power in mobile phones, refrigerators, Global Positioning System (GPS) navigation devices and even in washing machines, allowing us to work above those resources and develop solutions for our everyday environments. But even the possession of these kind of technologies, was not enough to flourish AmI in the matter that this kind of technology had to be accepted from the population, blending, incrementally, in our society. In health area, the improving of this area, is allowing to monitor patients more closely to their home, within their communities.

The idea of AmI is not new, but we can now seriously think about it as a reality and a discipline with a unique set of contributions. What we saw in science fiction movies, few years ago, like automatic sliding doors, or voice-recognition computers, is now a reality and more of these technologies are becoming even more present in our society. An intelligent environment is as transparent as more invisible are its components at human-sight. In few years, we could see that computers and computational devices had become smaller and smaller, and now we have already some devices that are almost undetectable, blending in the environment imperceptible to human eye [20]. As devices became even smaller, their portability became also greater, allowing us to transport them wherever we want, and the communication between them builds a network that may reach a whole building, emerging this way, the ubiquitous characteristic, or the characteristic of being present everywhere (omnipresence).

An Intelligent Environment must be able to see, listen, speak and even learn. These capacities (related to human intelligence) allow the system to be sensitive, adaptive and ubiquitous. Use of sensors is vital to the success of AmI being the link between the computational power and physical applications. The state of the environment is then collected, in order to also determine actions to take. Sensors size allow them to be integrated into almost any AmI system [20]. Wireless technology has also become more popular nowadays, turning our systems more mobile but more energy-battery consuming. These are the advantages and disadvantages of this technology.

There has been some application previously pursued by Ambient Intelligent researchers, such as Smart homes, Health monitoring and assistance, Hospitals, Transportation, Emergency Services and Education. As we can see, Ambient Intelligence can become a very interesting area of study. Another approach made on the University of South California, aims to the optimization of the energy consumed on a building, based on its physical infrastructure and also on environmental conditions, as well as on time dependent operational factors [16]

2.3 Technologies nowadays

There is recent projects and products that are accessible to our homes that help us to have more control over our places. Nest learning thermostat is one example of it. This device works as one traditional home thermostat but with some extra functionality. One person can turn its ring to set the wanted ambience temperature, but the device will start a mode that allows its system to learn about users preferences. This device can also control automatically the home temperature at night and know when homeowners get up in the morning and set their favourite temperature, all automatically. This device saves energy when the house is empty and never stop the learning process [3]. Sony has produced a power outlet that can measure and control manually the amount of power that a given appliance can use (by manually, it means turn on or off the equipment) and this outlet can also identify which user is using the appliance through RFID cards or tags. This is in some way useful to better how many Kilowatts a given user spend until the date, but in the other hand, it not learns anything about the users of the system, and so, there are no way to help users to better understand their behaviours, and consequently, teaching them where they can save more energy from their home appliances.

More related and recent work, and based on multi-agent systems, is SAVES project. This project was developed on the University of Southern California, and in this work, there was implemented a multi-agent system, in order to monitor and track power consumes, culminating on a system that can decrease power consumes [17].

[3] Nest website: `http://www.nest.com/`

Chapter 3

Contextualization

For this work, some concepts must be adopted to better understand the development of this required system. This chapter has the purpose to explain to the reader some concepts that will help him to better understand the work done in this dissertation. There are going to be described some concepts such as: sensing, Multi-Agent system (its advantages, functionalities and how it is composed), which kind of environment is required to be an object of study, its characteristics and what kind of information can be extracted from it. There are also going to be described the concept of Machine Learning and how important is this technology to develop and automated and auto-evolving system.

3.1 Sensing

Since there exists several kind of sensors, every single of them have some characteristics that are more suitable for some tasks than for others. For instance, touch sensors are more suitable to detect if someone or something touched the sensor, when compared to a force sensor. Since this kind of sensor can be attached on chairs and floors, it is possible to detect if something or somebody touched that surface. AmI is an emerging discipline that brings intelligence to our everyday environments and makes them sensitive to us

[20].

In order to achieve a better performance of this system, and since not every sensors retrieve values from the same nature, the monitored environment should be equipped with numerous kinds of sensors, in order to collect a richer set of information, improving this way, all the conclusions that are going to be achieved [4]. These environments should also be equipped with actuators allowing the adaptation or modifications of the environment accordingly to the calculated conclusions. Since there exists numerous types of sensors, there is a need to group them accordingly their functionalities and their objectives. These categories can be divided in [4]:

- Object sensors: Sensors attached to some objects in order to collect information about that object, in particular. These kind of objects can be, for instance, some electric home appliances, furniture or even switches or doors.

- Context sensors: Sensors capable of collect ambient information, in a continuous way, from some environment. These information can come from many ways, such as: luminosity, temperature, noise, and others.

- Motion sensors: These sensors, when set up on strategic spots, allow a system to collect people's movements in some space. Through these values, it is possible to conclude, for instance, if a person moves from a room to a kitchen or even to know where is "David", when the monitored space is a building's floor.

These group of systems have an huge scope when we are speaking in application areas, such as: Energetic sustainability, Security support, Hospital monitoring and even the comfort that we can have in our homes. Nowadays, there are numerous kinds of sensors, that can became part of our systems, such as:

- Electromagnetic sensors: sensors able to detect electromagnetic fields;

  - Compass: Localization device that points to the earth's magnetic pole. This is the ideal kind of device when it is needed the calculation of localization and trajectories.

- Ultrasonic: Sensors capable of detect sonic waves reflected on walls or objects, giving us the distance between the sensor and the object/target.

- Infra-red: sensors capable of identify values of luminosity through light reflection , such as obtaining distances (in a short range);

- Touch: Ability to recognize if there was a minimal touch of something or someone on the sensors surface;

- Vision: Sensors much probably equipped with a photographic or video camera, obtaining images in order to recognize patterns and gestures, for instance;

- Laser: Very similar to infra-red sensor, but has the particularity of obtaining information more precisely due to the different method of working, that is, the higher concentration of the light beam.

- Radar: Long range detection of moving targets (Static targets detection probably don't need radars for the effect).

- Stereoscopic vision: Vision genuinely natural, that consist in the ability of merging two images (naturally collected by our eyes) and with that resulting image, obtaining information about depth, distance, position and size of an target object, such as a 3D vision. This is the ideal kind of device for 3D restitution of scenarios through photos.

- GPS: Actually it is the most precise system that allows to obtain localization values from a given target;

- Accelerometers and velocimeters:

  - Accelerometers: A sensor that calculates the acceleration that a object can take. This allows what changes this object suffers, such as: travels, falls, turns and others.
  - Velocimeters: Sensor that calculates the instantaneous speed of a body or object. This sensor, combined with other kind of sensor (such as a GPS) and if we know the initial location of this object/body, we can obtain, approximately, the action radius of the body.

- Gyroscopes: Devices based on the principles of inertia used to provide data about directions, and also used to obtain data about balance (as used in robotics).

Given the extensive diversity of actual kinds of sensors, and the poor information that a single one of them can offer, the need to combine some of these kinds in order so it can be possible to extract a richer set of information and consequently better validate actions that can be taken posteriorly is needed.

Sensorial fusion appears as a term that defines the improvement of quality and quantity of knowledge extracted from the environment, leading also to some uncertain knowledge. From "uncertain knowledge", it is supposed to understand that a great quantity of information is not synonym of quality. For instance, if one man possesses one watch, has a better perception of time than someone that possesses two watches ("One man with one watch has the certain which time are, but the same man, with two watches may not have the exact perception of the time").

Sensorial Fusion can increase this way knowledge, but, as shown before, can also increase the number of uncertainties. In sum, a great quantity of information can lead to inconsistencies, leading this way to a incorrect notion of the managed environment. These "errors" (inconsistencies and conflicts) may come, for instance, from a same type sensors, which the main difference is the different calibration between them, making them to percept the same environment in different ways [2]. However, in order to fix some inconsistencies, there can be applied some certain techniques (such as sensorial fusion algorithms), increasing the efficiency of a given system. The resolution of these "errors" brings a better quality perception to what's happening in the environment, being possible to get better conclusions and consequently better custom service planning [3].

Said that, it is possible now, through a built dataset there can be now transmitted these values , all actions will be transmitted to actuators, which can be incorporated on objects (doors, windows, air conditioning, among others) in order to regulate (or evolve) the environment and satisfice automatically all users needs.

3.2 Multi-Agent System

A Multi-Agent System is considered as one development approach where a system can be developed in a modular way. This means that a system can be splitted into several modules, and these modules have the advantage of being adaptable and portable, increasing functionalities on systems fast and easy. For instance, an Agent A that belongs to a system X can be remotely called from a system Y, and this system Y will have access to all functionalities that the Agent A implements. A general Multi-Agent System is implemented as shown in the Figure 3.1.

These systems are composed of three main levels: a "Data Gathering" level, an "Acting" level and a "Reasoning Context" level. In the "Data Gathering" level, it is supposed that the system is connected to the Environment through sensors, and it is from those sensors that "Data" is gathered, or collected.

The "Acting" Level is the section from where the system changes the Environment, in a reverse way, when compared to the "Data Gathering" level, since the system has to be connected to some actuators and then, when these actuators receive actions delegated from the "Reasoning Context" Level, they

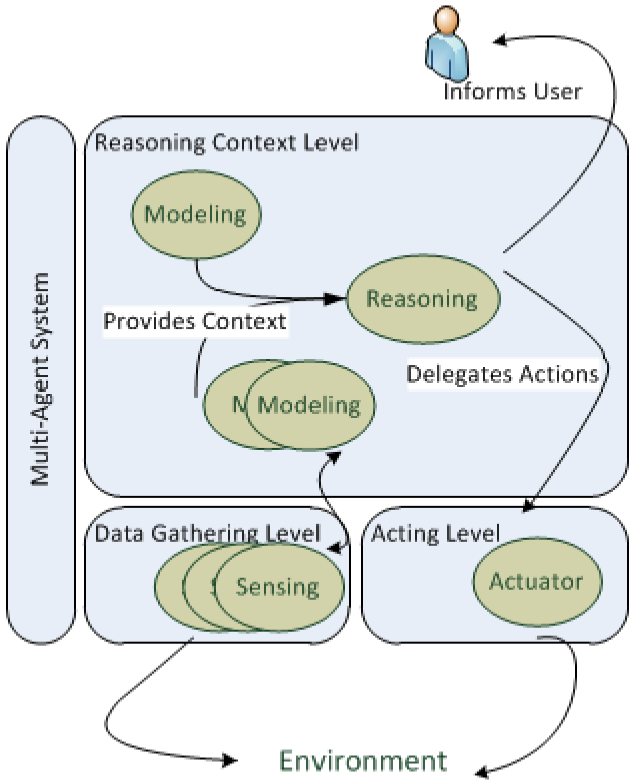

Figure 3.1: A Multi-Agent System general diagram

change the Environment, since they are also connected to the Environment the same way that Sensors.

It is supposed that, from data gathered from sensors, that the upper-layer "Reasoning Context Level" uses that data and generates some Models, and with that models, generate knowledge. This reasoning is one module that, when the models are generated, they are supposed to process and also generate some actions that must be taken so that the system can achieve its objectives. These actions are then delegated to the actuators, as described before.

The Reasoning module also informs user with important information or information that this user wants to know.

3.2.1 Multi-Agent Systems Tools and Frameworks

Nowadays there are some solid frameworks and tools so that multi-agents based applications can be more easily built.

One of the commonly used is the Java Agent DEvelopment framework (JADE) java framework. JADE is a framework that is developed in java and it allows the creation of agent-based applications. This framework uses the Foundation for Intelligent Physical Agents (FIPA) standards and implements successfully the Agent Management System (AMS), the Directory Facilitator (DF), and the Agent Communication Channel (ACC). For the communication between agents, this framework used an implemented language, denominated FIPA-Agent Communication Language (FIPA-ACL) [7].

Another tool that can be used to process some data mining is the Rapidminer software, being also one of the most used tools for knowledge extraction [1].

The intention of this subsection is not to provide a complete list of tools and platforms but rather a list of examples from which multi-agent DM systems may be developed.

3.3 Environment

There are several potential existing base-environments that are interesting in the matter of the information that can be extracted from them.

Some variables that can be defined from these environments, like temperature, noise, luminosity, electric power and much more. However, for this work, temperature, luminosity and RFID values (equivalent to each person entrance and exit from that environment) variables.

[1] Accordingly to the poll made on May, 2009 `http://www.kdnuggets.com/polls/2009/data-mining-tools-used.htm`

3.4 Machine Learning

There are several algorithms that can be applied on data mining processes. Each one of them has particular ways of work with data and different conclusions (rules, tables and trees, p.e) can be generated. This allows a developer to choose which type of conclusions wants to work with.

3.4.1 Data mining

Data mining is one of the techniques that is used by to achieve Machine learning.

Data mining is a set of processes that use raw information (or values), and generate knowledge, or usable information. Since the values collected come from different types of sensors, these raw values must be transformed through a battery of processes (algorithms) so that this set of values can be used in an heterogeneous way (this means that there is no further different values like Celsius, for Temperature, Lux, for Luminosity, for instance.

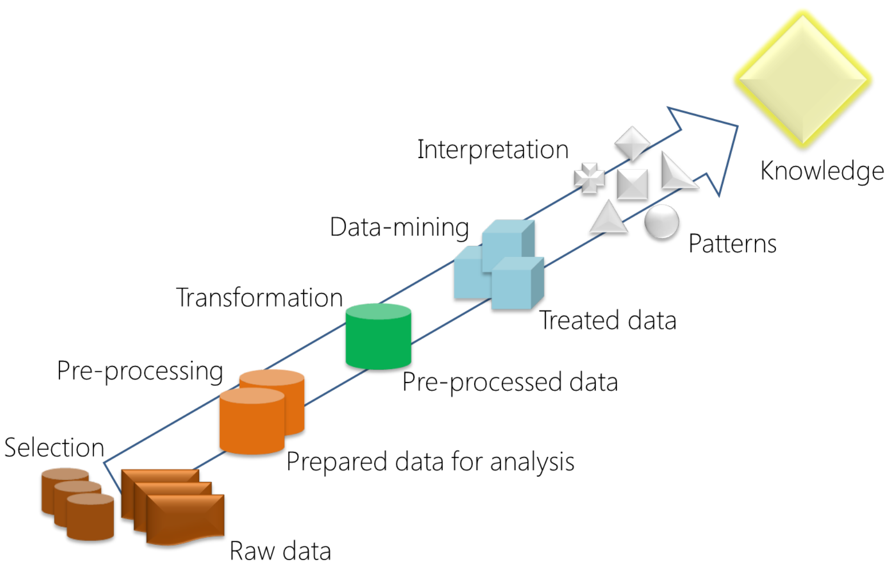

Figure 3.2: Knowledge extraction processes [3] .

1. **Selection** The phase of choosing the data that are going to be used in the process of knowledge extraction, since this data comes form heterogeneous sources and no always are useful to the resolution of a given problem;

2. **Pre-processing** Phase of cleaning the data, removing noise errors and blank fields and also represents the normalization of the data. This phase represents nearly 80 % of the knowledge extraction's effort, since information has to be correctly treated in order so that other phases can correctly process the data.

3. **Transformation** In this phase, data will be classified by subject, chronologically, aggregating or making clusters between them;

4. **Data mining** Some knowledge extraction algorithms may be applied in this phase in order to find the best set of techniques and methods to resolve the problem;

5. **Interpretation** This is the last phase of the process of knowledge extraction, where the found patterns by the resolution of the problem are used to the decision-making. This phase represents also the analysis of returning to the previous phase and applies a more suitable algorithm.

3.4.2 Clustering

Alongside with the classification algorithms, there are also another class of data mining algorithms that can be used. It's the Segmentation algorithms group. These algorithms are also interesting to use when a given user wants to group sets of data. These algorithms could also be incorporated to the application developed in this work, being another way to analyse the dataset.

Those generated clusters allow a user to make groups of data, and then determine patters from the dataset. There can be also set the number of clusters that the algorithm must generate in order to the business or problem needs. There are advantages when using clustering algorithms since these use every attribute to analyse the data, unlike the classification algorithms, that use only a subset of the attributes are used in the model.

The disadvantage of this kind of algorithms is the previous knowledge of how many groups a user wants to create. This task may be hard to plan, since there are datasets with thousands (or even millions) of records and the notion of group can be difficult to achieve. However, for small / medium datasets, these kind of algorithms can be very useful since they can quickly process the data and so generate more conclusions [2]

There are, in WEKA, also various clustering algorithms, such as SimpleKMeans or FarthestFirst. However, these algorithms weren't tested since there was an exhaustive testing of the Classification algorithms. These algorithms could be also bring another perspective to the analysis of this work in the matter of finding important groups of information (or values / instances) in the dataset. This way, we can say that the application of these algorithms on this solution is inconclusive in the matter that there are no results at all.

Clustering algorithms starts to find an important value or values of data, and then tries to group similar instances of data, forming, this way, groups of information. It is possible with this entities compare to other sets of information an quickly assimilate relations between them.

The use of these algorithms may be applied on various areas since from marketing, seismic foresight and insurers agencies. It is important to discover new groups of clients in order to maximize sells, identify common geographical sectors and even discover the most risky clients to make contracts [3]

One simple and intuitive example of this kind of algorithms is the comparison between products in a supermarket. There has been tested this algorithm over the database of purchases by the clients, and there has been found groups of clients that, if they bought baby dippers, they also bought beer. Besides the strange association between this products, this kind of rich information allow the supermarket manager to order their employees to change

[2]IBM data mining topics: `http://www.ibm.com/developerworks/opensource/library/os-weka2/index.html`

the place of the beers, next to the baby dippers. This simple action improved the sales of beers significantly, bringing more profit to the commerce.

3.4.3 Association

Association algorithms have the purpose to process a data set and then generate knowledge through association rules. These rules associate parameters between each other and find relationships between then.

This way, it is possible to calculate, or to know which parameter does influence another given parameter, and adjustments can be made, in order to better achieve the mais objectives of the problem.

3.4.4 Classification

A classification method can be described as the mapping of objects or instances on classes. Classification has the purpose to organize and categorize data into distinct classes, and it is possible also to make some predictions through this process. In this process, it is possible to predict, given a model, a new class of objects, and then map new data.

Classification is used to foresee discrete values. If the values that are about to be treated are continuous, then we are talking about prediction, but classification is considered as the prediction of discrete values. [1].

Building a classifier

The building of a classifier is divided into three phases:

- Construction of a model (Learning);
- Evaluation of the model (Correction or accuracy);
- Use of the model (Classification or prediction);

In the "construction of a model" phase (Learning) there must be denominated a class identifier, which is a predefined class, or attribute, and the set of all the objects is denominated of training data set. The model can assume representations based on:

- Classification rules (*if-else* rules);
- Decision trees;
- Mathematic formulae.

The "evaluation of the model" phase, corresponds to a model's evaluation, based on some tests. It is supposed, in this phase, to estimate the model correction rate, comparing the classes of the test dataset with the results of the classification made by the model, and that comparing results on a correction rate (percentage of the correctly classified test cases).

In the "Use of the model" phase, the model has to be put on the field and be used in order to classify unknown objects (objects that don't belong to the test objects). In this phase, it is supposed that the model receive new objects and then try to predict the value of a given attribute.

Classification methods

There are several methods of classification, among which there can be found:

- Decision trees;
- Artifitial Neural Networks;
- Bayesian classification;
- Closest neighbour (k-Nearest Neighbour);
- Case based reasoning;
- Genetic algorithms;
- and others.

Decision Trees

The decision trees classification method consists on the implementation of a test to one attribute, represented by a node. The branches that connect the nodes represents one value to the tested attribute and to every single leaf (leaf, in this case, means the last node that the tree has, when looked from up to down) is given a classification. The top node is the only entry point to the tree, and it represents all the objects in the tree.

These decision trees are generated, normally, from a top-down way, and its made by two phases:

1. Building;

2. Pruning.

In a first stage ("Building"), every objects make part of the tree and the examples are divided, recursively, in order to the selected attribute. in the second and last stage of the tree generation ("Pruning"), the main goal is to remove all the unnecessary branches, in order to increase the correction and precision of the classification.

There are some decision trees algorithms such as:

- ID3, ID4, ID6;

- C4.5, C5 (these are used very commonly);

- J48 (and this one as well);

- CHAID, CART.

These Decision Trees are algorithms that are specifically suitable for some problems with large dimensions.

3.5 Ubiquity

Ubiquity can be described as the ability of being omnipresence, this is, the ability of being everywhere, at the same time. Observation can perhaps be considered one of the most accurate forms of monitoring and track people and objects, but this tracking and monitoring can be only accurate when the system that tracks or monitors, possesses technology that allows to collect data in different spaces, in an disguised way so that the user does not feels observed or influenced in every action he takes (this isn't applied to objects) [19].

3.5.1 Pervasiveness

In AmI developments, it is important that pervasiveness can only be achieved when hardware is highly blended in the environment so that the user doesn't feel that he is observed and its actions are influenced.

In this work, pervasiveness couldn't be completely achieved since the RFID tag reader had a very short detection range and every ISLab user had to pass their RFID tag very close to the reader. This way, the system couldn't record efficiently every entrance and exit from the ISLab due to the users forgetfulness when they, for example, came out from the ISLab and forgot to pass again they RFID tag to register their exit. If this happened in a Friday, the system would "think" that that person was in the ISLab through the weekend.

Chapter 4

Development

This chapter has the objective of explain what has been done in this work and what modifications did this work suffered and the motive that these decisions were taken.

4.1 Data types

In order to achieve and complete the main objectives of this work, this system must be able to determine and control the energetic sustainability of a given environment, we started to find some suitable variables. To do so, in this section, there are going to be settled some definitions about the information that is going to be treated in this work.

Looking to the hardware available in the ISLab, there was not available any kind of sensors able to connect to some home appliances and measure how much that home appliance would consume. Given this, this work will have to be oriented to some other kind of values, allowing later, to be easily adapted to other kind of data (power values), and then fully complete the purpose of this work.

4.1.1 Temperature, luminosity and person tracking

Given the lack of power collecting data sensors, there must be adopted new kinds of values, and they are: Temperature, Luminosity, and person tracking (by person tracking, it means a system to understand where people, or a given person are/is), being this last one of the most important value to collect. With these variables, the system will try to predict, how many people are in the room, given the luminosity and temperature of the environment. In order to collect these kind of values, it is supposed to choose one suitable environment that allows to collect this data, and the ISLab room was chosen.

The explanation for the selection of the ISLab was the fact of temperature and luminosity variables are both adjustable in this space. Luminosity can be adjusted through windows (and their blinds), and through light lamps and temperature can also be regulated through the activation (or deactivation) of the air conditioning. These two variables will be collected in a cyclic way (interval of 30 seconds between each collection), unlike the person tracking variable, that is collected occasionally / sporadically.

By person location, we mean the actual status of a person in the ISLab, i.e, to know if who is inside, or outside of the laboratory, in a given moment.

Every entrance and exit of a identified person is persisted (on a database), and since there are some RFID hardware available, these values are going to be registered in this format. Each person must possess a RFID code, and every entrance (as well for every exit) on the ISLab, will represent one record on the database. So, when the system starts, it is supposed that the laboratory is empty, and when somebody arrives, it is supposed to register one record (meaning the entrance). When the same person is about to leave, there will be registered another record, representing the exit of that person.

Alongside with the unique identification code, it will be persisted also a timestamp (A timestamp is a value, in milliseconds, from which can be extracted all date information, such as Hour, Day, or even the day of the week).

The tracking of each person is one of the key-aspects of this work. When a system knows where a user is (geographically), there can be studied, or generated patterns (or routines) that better describe each character's actions, and with this, determine what's the influence that every person has in the managed environment. These three variables were chosen, principally, due to the phidgets equipment that was available in the ISLab, since there were touch, luminosity, temperature, RFID and touch sensors.

4.2 Hardware

4.2.1 Nebusens N-Core

In the early days, this work was about to be implemented alongside with Nebusens N-Core [1]. Nebusens [2] is a Spanish hardware enterprise that produces N-Core as another solution to create wireless sensor networks, based on the IEEE 802.15.4/ZigBee pattern, being one of the main characteristics the easy scalability and modularity of new components, allowing the easy and fast growing of a wireless network.

The available kit was composed by one Sirius A device (Figure 4.1a), and about five Sirius B (Figure 4.1b) and D devices (Figure 4.1c) and some magnetic sensors and some Light-Emitting Diode (LED). The Sirius A is the main device, that connects to a computer and from where data will flow, possibly generating a bottleneck in this point. Sirius B and D should not also act as data repeaters, such as routers, but sensors can also be attached through some connectors that are located on these devices, and then place these devices somewhere in order to collect data via wireless.

[1]Nebusens N-core website `http://www.nebusens.com/index.php/en/products/n-core-en`

[2]Nebusens website: `http://www.nebusens.com/index.php/en/`

(a) (b) (c)

Figure 4.1: N-Core Sirius A (a), B (b) and D (c)

In Figure 4.2, it is presented how this hardware could be set up. In a practical test, Sirius B and C should be connected to some sensors, and these, in turn, connected to some several home appliances such as TVs, Air Conditioning or even electric sockets, in order to study when they are powered on, and for how long. Sirius B could be used to person tracking, since it can be used as a tag, where a given user could keep it in the pocket.

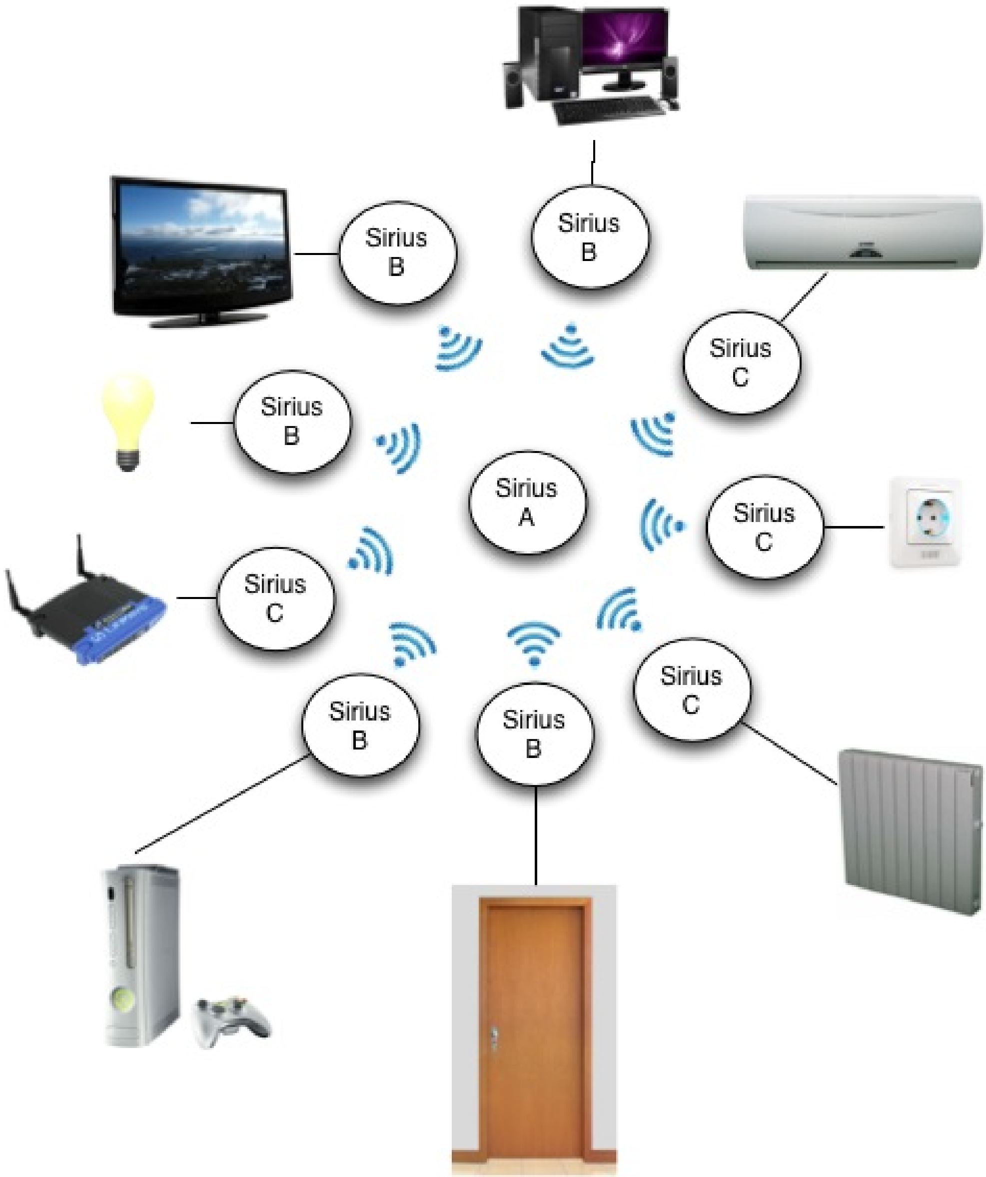

Figure 4.2: A network based on N-Core devices connected to some home appliances and doors

This set up, could not, however, be built successfully since the kit wasn't equipped with any kind of instruction manual or some how-to, and not even in Nebusens website, in order to better understand how to communicate between all the Sirius devices and with some computer. This way, we had to change the hardware responsible to collect data, to a new one that would be easier to connect to a pc and communicate better with sensors.

4.2.2 Phidgets

Phidgets hardware were the second chance to try to implement this work. Phidgets are small devices that possesses highly connectivity to computers (through Universal Serial Bus (USB) ports) and their sensors are also easily attachable to Phidget boards. In the laboratory, it was used one from two Phidgets starter kits, and it was composed by:

- Precision Light sensor ref.(1127)
- Temperature sensor ref.(1124)
- PhidgetRFID tag reader ref.(1023)
- Interface 8/8/8 kit ref.(1018)
- Touch sensor ref.(1110)
- RFID tags ref.(3007_0 and 3009_0)

Since this kit wasn't equipped with some electric power data collecting sensors, there were chosen some sensors and boards presented in the list above so that some information from different natures can be collected and try to study people's behaviours and then try to learn their behaviours and estimate some consumes. For example, it is supposed that some home appliances (such as lights) can be turned on or off when the luminosity of the outside is very low. If the fact of turning the lights on is associated to the fact of the luminosity of the day being too low, then probably, in the next two, three times, the system will learn if this is a common behaviour (turning the lights on when the luminosity is low) and then, try to turn the lights on automatically.

For this work, Precision light sensor (Figure 4.3) will be used collect the luminosity values in the ISLab and temperature sensor (Figure 4.3) will collect the temperature values along the time. It was decided that these values, since their environmental nature, would be collected cyclically, in order to detect some unusual variations, and, when crossed with other kind sensors, determine some patterns. The time period of 30 seconds between each temperature and luminosity value collection was set.

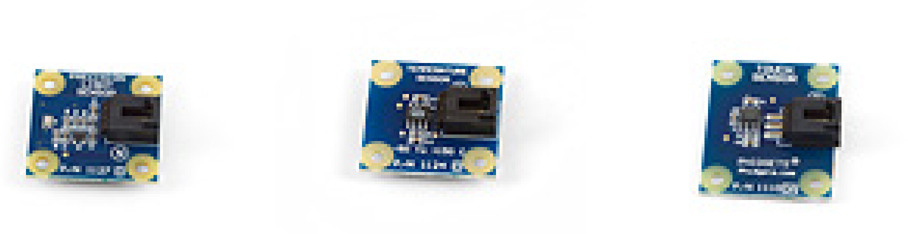

Figure 4.3: Phidgets precise temperature sensor (On the left), and Phidgets precise luminosity sensor (On the middle). The phidgets sensor represented on the right is the sensor touch

The RFID tag reader (Figure 4.4) will be responsible for collecting information about some RFID tags (as shown in Figure 4.4) and when they were detected. Each RFID tag will be distributed for highest number of the persons that usually frequents the laboratory since each one of them possesses and unique identification code that allows to recognize each person individually.

When somebody passes the tag in front of the reader, that unique code is detected and will be persisted on database. These passages will represent, for the system, one of two states: or the person is entering the ISLab, or leaving it. When the application begins, it makes sense that the system assumes room is empty, and then it will start to detect every movement that every person does.

Figure 4.4: Phidgets RFID tag reader (To the left), RFID black tag (On the center) and RFID keychain black tag

The touch sensor (as in Figure 4.3) will have a different function. In order to detect some variations in the luminosity values, and since the ISLab possesses some windows equipped with some blinds that can be opened and closed, it is supposed that, every time these are opened and closed, the luminosity values will suffer a great variation. In order to detect if somebody closed or opened the window, it will also be set a touch sensor so that, when somebody moves in the blinds, he/she has to press slightly on the touch sensor. The system will recognize actions the same way that RFID tag reader works, that is, when the system starts, all blinds are opened, and the next touch detected on the sensors will represent the close action of the blind.

All these sensors described above are connected to the phidgets logical board (the same as the one represented on Figure 4.5). This board has capacity to connect up to eight digital sensors (as the ones that were described above), and also allows the connection of sixteen analog sensors. This board is connected to a computer through its USB connector, as well to the exception of the RFID tag reader (Figure 4.4), allowing the collection of data through this interface.

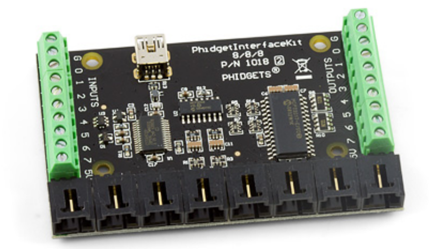

Figure 4.5: Phidgets board 8-8-8

The implementation of the system itself is going to be on Java programming language, and it has been designed into three four distinct layers (FIgure 4.6).

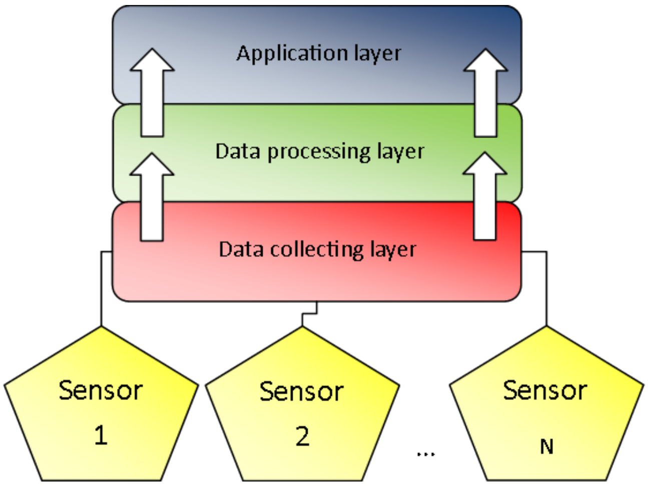

Figure 4.6: A general schema application architecture

The Figure 4.6 shows, in a general way, the layers that will be implemented. Starting from below, the sensor network is represented by "Sensor 1", "Sensor 2" and "Sensor N", being "N" an unknown number of sensors. This network is going to be connected to the "Data collecting layer", where all the information is going to be collected, and persisted into database. In both "Data collecting layer" and "Data Processing layer", there are going to be implemented JADE agents, and these agents will communicate to the most upper layer, "Application layer", where the GUI receives all the info sent from JADE agents and changes so that the user can see the system report and better manage their environment [26]. The data starts to flow from the lower layer to the layer immediately above, until the "Application layer".

4.3 Raw data

First of all, after the assembly of all the phidgets to the main board (Figure 4.5), they were fixed to a wall (Figure 4.7), in order to better collect luminosity values, with no interference (such the passage of a person in the front of the luminosity sensor). In Figure 4.7, the set up can be seen, next to a door, so that the tag reader can be easily accessible, and so that everybody didn't forget to pass the corresponding tag and then register each entrance / exit.

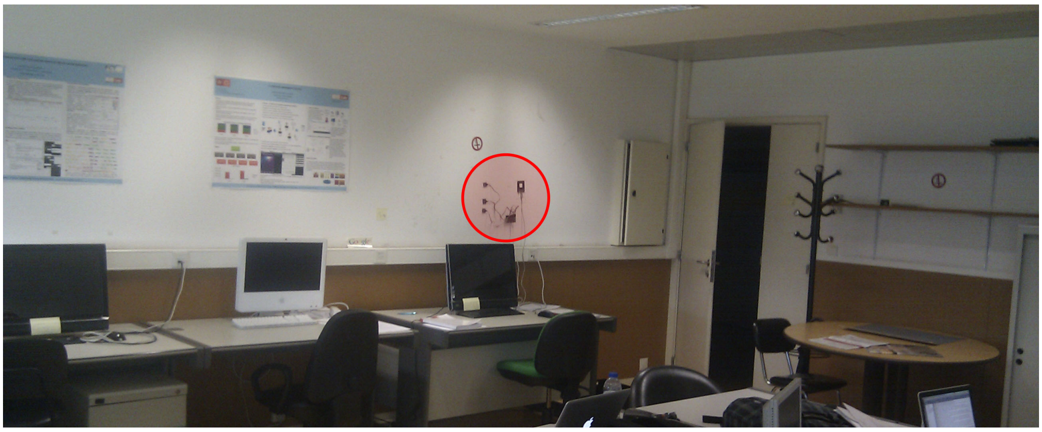

Figure 4.7: The set up with sensors fixed on the ISLab wall

4.3.1 Collection of the raw data

This set up was working for about one month, collecting data continuously. This data was persisted into a database generated through the schema presented in Figure 7.1, at the Attachments (Chapter 7). This is a very simple schema, since the relations between the different kind of data weren't too complexes.

In this schema, it is possible to recognize 5 tables:

- luminosityValues;
- temperatureValues;
- rfidValues;
- time;
- touchValues;

Noticing the name of the tables, it is easy to understand what each kind of values did them stored, unlike the "time" table. "Time" table has the purpose of working as the system clock, gathering all timestamps (explained above) of all temperature, RFID, luminosity and touch record. This is the most important table on the database since it is where the system will collect data about when did the collection of data happened in order to calculate important statistics later.

4.3.2 Datasets

The WEKA Application Programming Interface (API) allows the connection to a database, and also to query it, in order to retrieve some instances (from instances, it is pretended to understand that are a set of values, where are present some required values) from it. This API queries the database and retrieve result sets in the shape of datasets, and these datasets are made of instances with 5 attributes (timestamp, Luminosity value, temperature value, hour of the day and number of persons)

Table 4.1: Instance

timestamp	luxValue	tempValue	hourOfTheDay	numberOfPersons
timestamp	lux_Value	temp_Value	hour_day_value	number_persons_value

4.4 MAS

After being more familiar with the concepts of WEKA API's datasets and instances, it was implemented the MAS, and there was implemented four agents, as shown in Figure 4.8.

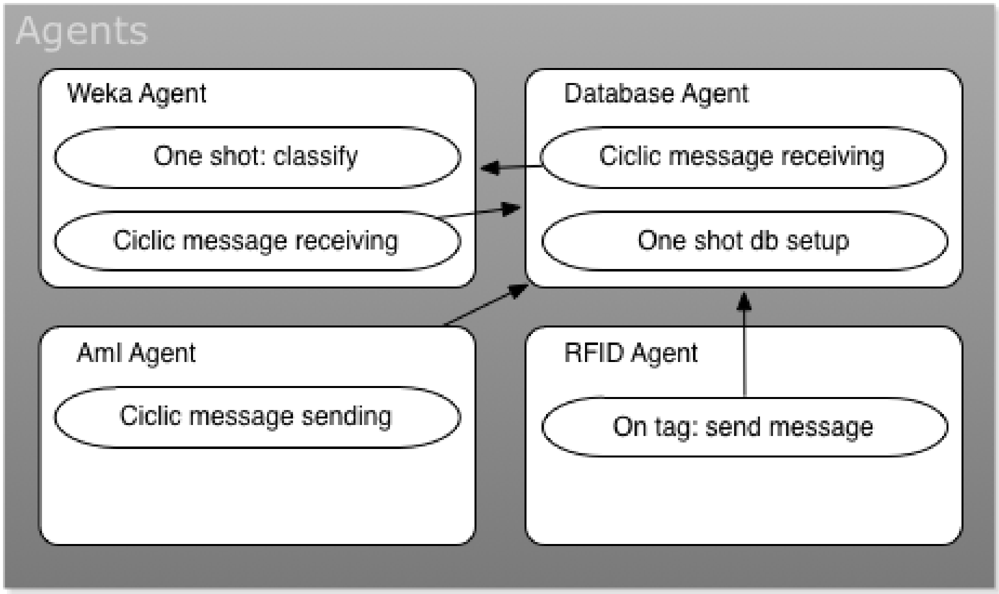

Figure 4.8: The multi agent system implemented in this work

- RFID Agent.
- AmI Agent;
- Database agent;

- WEKA agent;

The RFID agent function is to collect and send each tag detected by the RFID reader and send it to the database agent so that this record can be persisted. This agent implements also a One Shot behaviour, that set up events so that the system can trigger more sending messages to the database agent, when a new RFID tag is detected.

The AmI agent function, similarly to the RFID agent, is to collect temperature, luminosity and the Window State values through the temperature, luminosity and touch sensors. After this, this agent prepares a message with those collected values and sent it to the database agent.

The database agent is responsible to persist all the information collected through the sensors and sent from RFID and AmI agents into a database. This agent is also responsible to control the GUI and its components, in order to better display all the information to the user. This agent also possesses two behaviours: an One-Shot behaviour, responsible to establish connection to the database, and the other behaviour is a cyclic one, that receives all the messages sent to the database agent, checks which (AmI or RFID) agent sent that message, and then process (persists) upon this sender. If this message is sent from WEKA agent, then database agent is supposed to refresh every information the GUI is showing.

Weka agent is the agent responsible to execute all the knowledge extraction processes that data mining requires, as described in Figure 3.2. This agent has implemented two behaviours. One of those is a one-shot behaviour that accesses to the database in order to collect all the collected raw values and builds the training dataset.

The second behaviour (a cyclic one) is executed continuously and is responsible to receive and process all the messages received. In this behaviour, a message is received, and then, it is extracted from this message, the function that is necessary to execute.

There are two options to be executed, or to classify, or to predict one dataset instance. If the instance is going to be classified, there will be applied a classifying algorithm and then associated to other similar classified instances. If the instance is going to be predicted, it is possible to determine a value of the variable that we want to predict, and then try to estimate the user consumes.

In this work, it is possible to predict, with the instances values of Temperature, Hour and Luminosity, how many persons will be at the given interval of time.

4.5 Testing WEKA

There are several algorithms that can be applied on data mining processes. Each one of them has particular ways of work with data and different conclusions (rules, tables and trees, p.e) can be generated (As explained in 3.4). This allows a developer to choose which type of conclusions wants to work with.

For this work, WEKA [8] data mining tool was chosen since it is an intuitive and easy way to work with datasets. Pre-processing a dataset, choosing algorithms, see their description, apply the algorithms over a dataset and even analyse their results are some of the features that this tool allows us to do.

The third point of the methodology steps refers to the test of a prototype, and the its installation on a physical space, in order to test it and collect some data to work with. It is supposed that this prototype collect information about people walking in and out of that space (such as a room or laboratory), and then persist that data into a dataset.

The fourth point refers to the running of the application over a robust dataset, so that it can work and process the knowledge extraction phases as referred before, and with those results, build constructive and important conclusions.

4.5.1 WEKA Explorer

WEKA tool possesses an GUI where a given user can use the "Explorer" functionality from it. This explorer allows an user to connect to a MySQL database and process data. After the selection of the data that the user wants to work with, he can do lots of operations, from Pre-processing until Association or Classification methods.

In this work, this explorer tool was used to determine the efficiency of all algorithms, to facilitate the selection of the algorithms later so that can be more easily chosen later.

4.5.2 Pre-processing

This subsection explains some steps taken since the collection of the data from the database, the application of filters and changes done to the data before the processing of the algorithms and the later knowledge analysis.

First of all, since some values were collected in the data-collecting phase, as numeric values. Since the RIpple-DOwn Rule (Ridor) algorithm demands that all attributes must be in Nominal form, then a filter was implemented in order to change the nature of these variables.

Another pre-processing technique that was applied on the data set was the remove technique. Remove technique aims to the removal of noise and by noise, is considered all the instances or data that doesn't bring or misrepresent valuable knowledge. So, there was applied another filter in order to remove noise present on the dataset and just only after the end of the removal, the algorithm was applied over the dataset and then continued the data mining phases

4.5.3 Classification

Classification method was chosen as one of the techniques that can be applied over the dataset so that prediction can be achieved and processed.

After the data collection phase, there were tested all the classification algorithms that WEKA Explorer provides us, over the dataset built with sensors values. Tables 4.2 4.3 4.4 4.5 4.6 4.7 show the percentage of correctly classified instances from each classification algorithm.

Table 4.2: Bayes-classifying algorithms

Algorithm	% of correctly classified instances
AODE	78
AODEsr	79
BayesNet	65
HNB	78
NaiveBayes	65
NaiveBayesSimple	65
NaiveBayesUpdatable	65
WAODE	79

Table 4.3: Lazy-classifying algorithms

Algorithm	% of correctly classified instances
IB1	77
IBk	81
kStar	75
LBR	81
LWL	65

There are some algorithms that, due to lack of memory heap or very slow processing time, couldn't be completely executed, so, these are useless to this work and they are presented in the Table 4.8.

Table 4.4: Meta-classifying algorithms

Algorithm	% of correctly classified instances
AdaBoost	52
AdaBoostM1	52
AttributeSelectedClassifier	81
Bagging	81
ClassificationViaClustering	48
CVParameterSelection	52
Dagging	74
Decorate	81
DecisionTable	82
DTNB	81
END	81
FilteredClassifier	81
Grading	52
LogitBoost	64
MultiBoostAB	52
MultiClassClassifier	67
MultiScheme	52
RacedIncrementalLogiBoost	64
RandomComittee	81
RandomSubspace	75
RotationForest	81
Stacking	52
StackingC	52
Vote	52

In these kind of projects, it is important to achieve a great percentage of correctly classified instances that these algorithms results. This allows a

Table 4.5: Misc-classifying algorithms

Algorithm	% of correctly classified instances
HyperPipes	54
VFI	42

Table 4.6: Rules-classifying algorithms

Algorithm	% of correctly classified instances
ConjunctiveRule	52
DecisionTable	81
DTNB	81
JRip	71
NNge	77
OneR	59
PART	81
Ridor	80
ZeroR	52

given system to correctly adapt the environment to what a user really needs, without fake or wrong predictions. If the system predictions are most of the time wrong, the actuators won't respond to the user needs, turning that system difficult to live with.

In the end of the tests phase, it was selected one of the most efficient test-algorithm (better classified) to better respond to the users needs.

4.5.4 The chosen algorithm

Since the percentage of correctly classified instances was one important choice factor, the Ridor algorithm was selected since it obtained, among other

Table 4.7: Trees-classifying algorithms

Algorithm	% of correctly classified instances
BFTree	81
DecisionStump	52
ID3	81
J48	81
LADTree	61
NBTree	81
RandomForest	81
RandomTree	81
REPTree	81
SimpleCart	81
UserClassifier	52
VFI	42

algorithms, a very high score, as shown on the Table 4.6. In this Table, it can be seen that about 80 % of all instances of the dataset are correctly classified and, this way, it is one possible starting algorithm.

This algorithm belongs to the *weka.classifiers.rules* package algorithms, and this means that the information is represented by rules. These rules format is:

Default rule: "Instance_processed = value (number_of_total_instances / number_of_correctly_classified_instances) [=>] [Exception Rule]

Exception rule: "field_attribute" = (interval_of_the_previous_attribute) => (number_of_total_attribute_instances / number_of_correctly_classified_attrribute_instances) [number_of_total_instances / number_of_correctly_classified_instances]

Table 4.8: Inconclusive algorithms

Algorithm's name	Algorithm's type
PART	Meta
LibLinear	Functions
LibSVM	Functions
Logistic	Functions
MultiLayerPerception	Functions
RBFNetwork	Functions
SimpleLogistic	Functions
SMO	Functions
SerializedClassified	Misc
CitationKNN	Mi
MINND	Mi
MISMO	Mi
MIWrapper	Mi
SimpleMI	Mi
ClassificationViaRegression	Meta
CostSensitiveClassifier	Meta
Cost	Meta
GridSearch	Meta
MetaCost	Meta
OrdinalClassClassifier	Meta
Prism	Rules
FT	Trees
LMT	Trees

The output of this algorithm is designed in a tree-like aspect. This algorithm builds the base of the tree by calculating the default rule, and the remaining nodes (child nodes) are result of the calculation of exception rules, and every node is connected by branches. Each default rule excepts the father node rule, in a cascading like exception tree. In the end, this algorithm is supposed to generate the best rules that classify instances [14].

4.5.5 Problems

The Ridor algorithm 4.5.4, when applied through the WEKA GUI explorer, the dataset was processed smoothly and with no problems of lack of memory, or similar problems. However, when WEKA API was implemented on this system, the system could not open a database connection and extract all the 90000 records and then classify this dataset, while the connection was open. This problem was raised and the caused appeared to be hardware limitations, since the machine that where the system was being developed only possessed 3GB of Random Access Memory (RAM), and that processing seemed to require a bigger RAM storage.

In order to overcome this limitation, another algorithm was selected, and then tested to classify the dataset. The C4.5 classification algorithm was selected, by looking to the 4.7, and confirmed that this algorithm has also a high percentage of correctly classified instances (about 80%). This algorithm has a different nature that Ridor, since it is a tree-based algorithm [23], and in the phase of testing on the system application, it didn't require very specific hardware requirements. The WEKA API implements this algorithm in Java (which is named J48) and its was adapted to work alongside with the system of this work).

4.6 The application

This section will explain every development that the GUI suffered since the beginning of the work.

4.6.1 The GUI - Listing

The GUI need to be easy to understand and has to present important information to the user. So, in an initial phase, we thought on designing a simple list to show the persons that are in the ISLab or outside. This list was then implemented, showing the names of the persons that possess the RFID tag (Figure 4.4) and attends the ISLab and, in front of each name, it was decided that to represent when the person is outside the laboratory, the red color, and, if he is inside, the green colour. These colors were chosen since they are very intuitive and easy to learn.

Figure 4.9: Screen shot taken in the first stage of the application

As we can see in the Figure 4.9, the application is starting with a red list, meaning that everybody is outside the ISLab, and then the application starts to register each entrance, by changing the red colour to green.

After some time, it was decided that this application should have another panel, where the system manager could watch the general consumes overview, and predictions. And with that change, it was needed to separate the list presented on Figure 4.9 onto a new panel. The result of this is depicted on Figure 4.10.

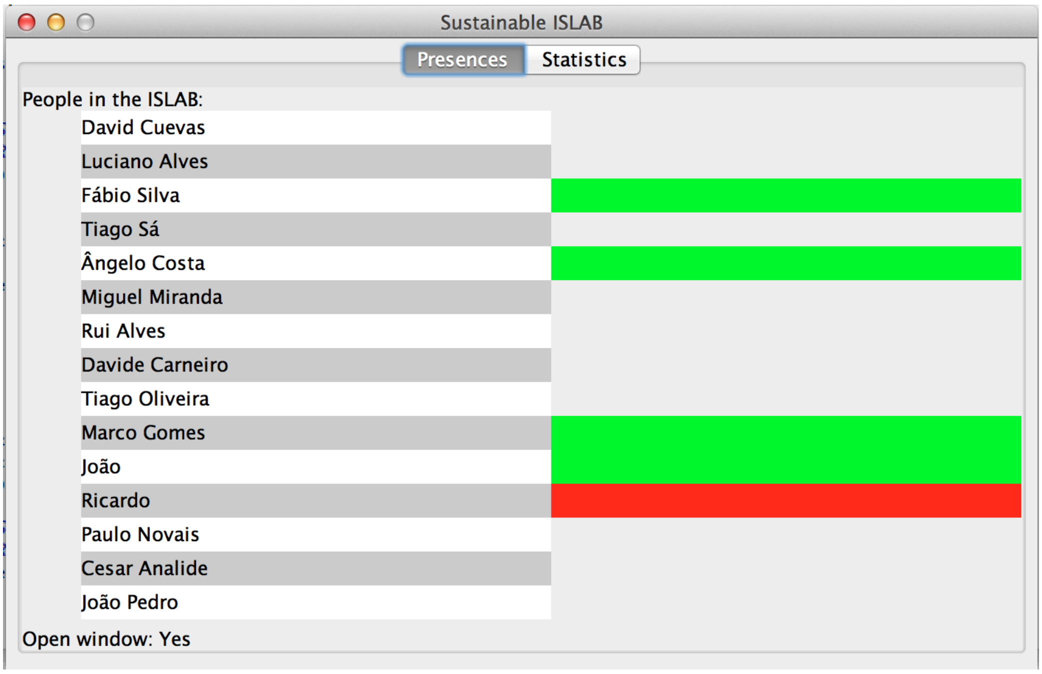

Figure 4.10: Screen shot taken from the first tab of the application

With this, it was also implemented some improvements at a graphical level, that can be observed clearly on the Figure 4.10.

4.6.2 The GUI - Reports

The next step of implementing this application was the report tab. The report tab is supposed to show to the user all the relevant information, such as: the most active person (by active, we mean the person who spent the most time in the ISLab), as well the less active person, the most consuming appliance (that is, the device that consumed more power since the application started), and even the predicted number of persons that will be in one hour in the room. This last value is the value that the C4.5 algorithm returns, giving the prediction power to the system. The Figure 4.11 shows the screenshot that correspond to the report with the parameters above descibed.

In the last figure, all the report parameters are in a processing phase, waiting for the start of the implemented agents and the classification of the

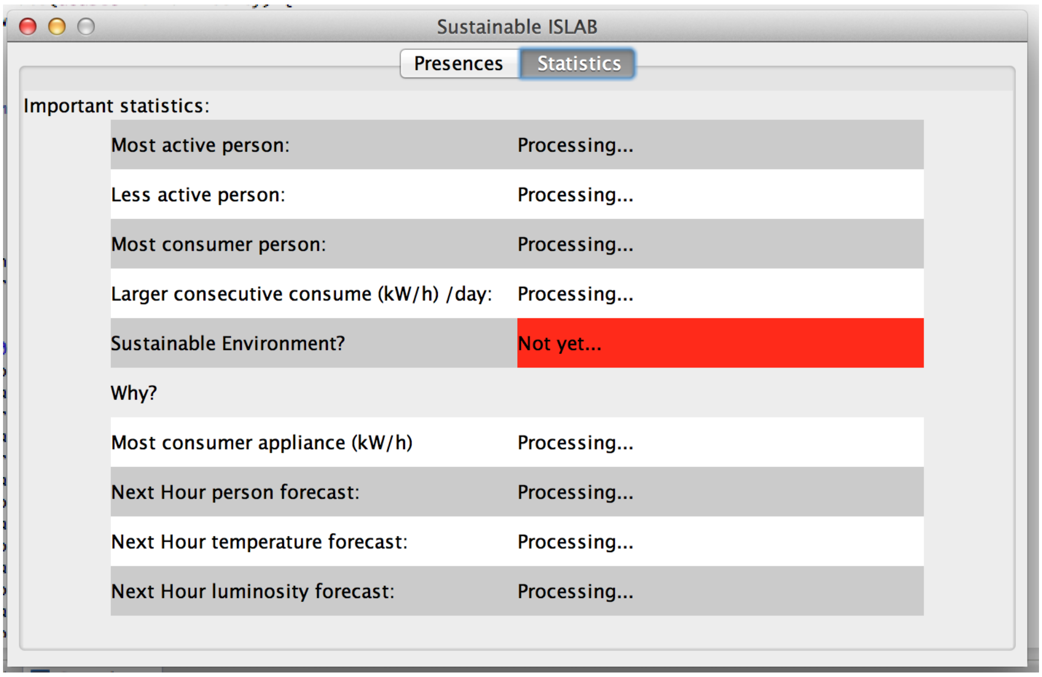

Figure 4.11: Screen shot taken from the second tab, on the very start of the application

dataset since the system needs to use the classifiers in order to predict new instances of values.

After the initial processing, some forecasts and predictions were made, and then presented on the report, so that the manager can start to see some results and also start to make some decisions. It can be seen that the "Most consumer appliance" is not yet calculated, since there was any hardware available to measure power consumes. "Sustainable environment?" and "Most consumer appliance" cells were maintained since another kind of forecasts was made.

As explained above, despite of the lack of the consume information on the system, temperature, as luminosity and person forecast values are already available since the classifier that the system uses is already classifying new instances, and predicted the values on the figure.

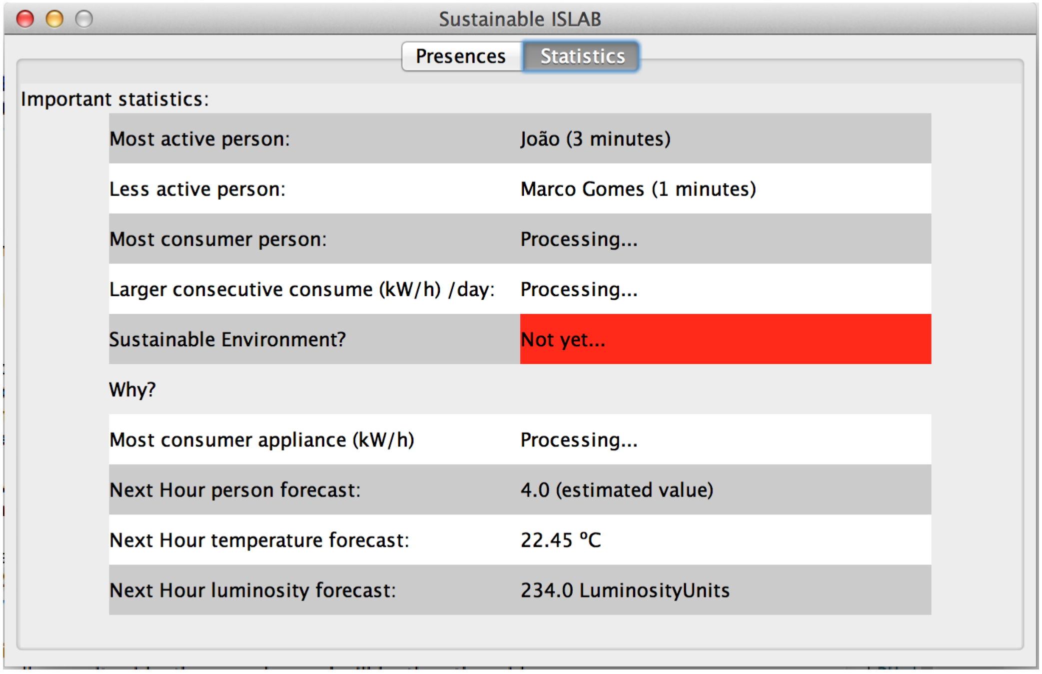

Figure 4.12: Screen shot from the report, with some parameters already calculated

4.6.3 The Web platform

Since this system consisted on the development of a desktop application, it does not bring any kind of advantage if someone wants to access to his home reports remotely. This way, a web platform was suggested and started to be developed in order to respond to this need, and allowing to the house owner to control every sensors agents, and can also obtain information that are available in the report, only through a web browser.

In the Figure 4.13, there can be seen the web site, developed in HyperText Markup Language (HTML), Cascading Style Sheets (CSS) and Java Server Pages (JSP), and it is from this web site that the manager can have access to the system's report. This interface, despite of the still rudimentary layout, it provides already all the informations that allows to the manager to take actions and also change previous actions in order to achieve a sustainable environment at his home.

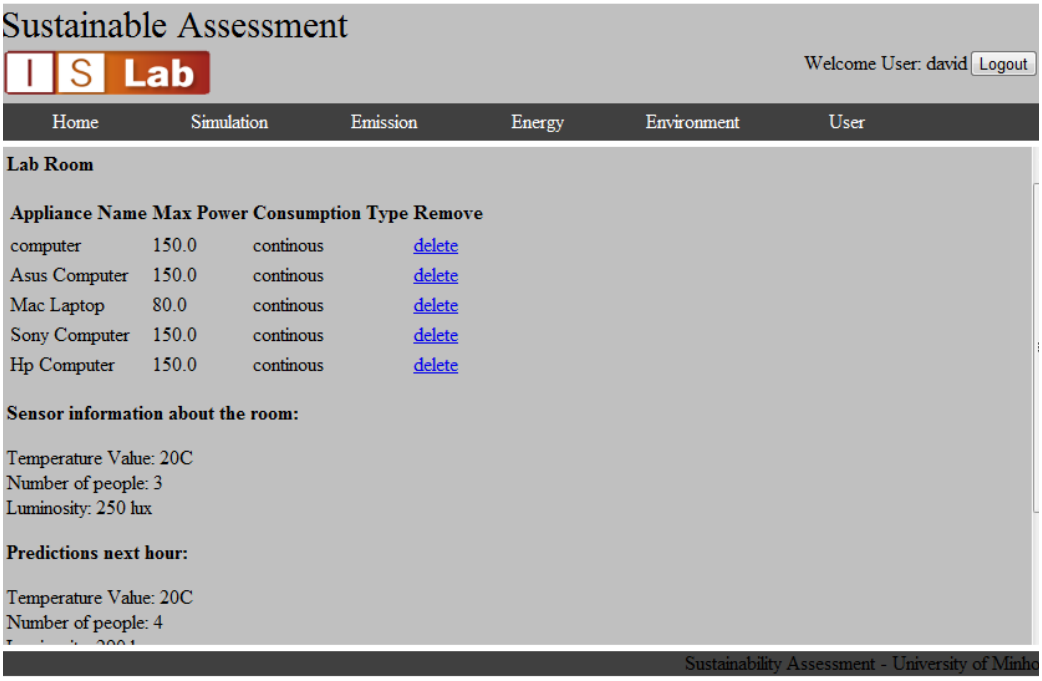

Figure 4.13: The web application screenshot

Chapter 5

Results

After the work is done, there must be made some considerations about the some modifications and final results that were obtained, and that's the goal of this chapter.

5.1 Previous work

The system was able to predict, based on the data set collected on the data collecting phase, most of the times the variable that the manager wants to know, and since the reports explains very clearly what devices are the most consuming ones (even there were just simulated values, based on the time that everyone was on the laboratory), and which ones were the most sustainable.

5.2 System Performance

About system performance, there was a significant improvement comparing the "young" version of this system, and the last version. When comparing the extreme processing power that was required when Ridor algorithm was selected, there could be observed a significant improvement , after the change of algorithm to the C4.5, specially on the speed of processing. This means that this system doesn't need to be deployed on very powerful and expensive hardware configurations, since the requirements are very accessible.

5.3 Web Integration

The web interface brought a new level of management since it helps system managers to better manage the energy resources that their homes consume, and even control devices.

Chapter 6

Conclusion

This chapter has the purpose to describe what are the conclusions achieved by this work and what nees to be done in the future.

6.1 Synthethis

In the end, this work is able to predict in an evolutive way, values of temperature, luminosity and even the number of persons that will be on a specific hour of the day. The system can also report the most active and unactive person that attends the ISLab and even estimate which appliance is the most consumer (these last values are estimated).

There was implemented also a web interface, that allows a user to control the system agents remotely and have access also to the reports that are generated locally on the main computer.

6.2 Relevant work

During this research several parallel activities and tasks were performed that have contributed to the evolution of this work. The list of such activities is described below:

1. Silva, F., Cuevas, D., Analide, C., Neves, J. and Marques, J. "*Sen-*

sorization and Intelligent Systems in Energetic Sustainable Environments", at IARIA Conference. Despite the acceptance of this article, the team decided that the same is not going to be published on the conference book [1].

2. Silva, F., Cuevas, D., Analide, C., Neves, J. and Marques, J. "*Sensorization and Intelligent Systems in Energetic Sustainable Environments*" (short paper), at the IDC conference [2]. This article is going to be published at the "Intelligent Distributed Computing VI - Proceedings of the 6th International Symposium on Intelligent Distributed Computing - IDC 2012" book.

6.3 Future work

For future work, and connecting a power-collecting Phidgets sensor, it is very simple to start collecting these values and learn consumption patterns to generate more reliable reports. The idea of creating an social network based on power consumption is always a good starting point to achieve a mutual helping / challenging neighbours to save energy and turn their homes even more sustainable.

[1]IARIA conference hyperlink: `http://www.iaria.org/conferences2012/CfPAMBIENT12.html`

[2]IDC2012 conference hyperlink `http://idc2012.deis.unical.it/index.html`

Bibliography

[1] Cesar Analide, Paulo Novais, and José Neves. Classificação para extracção de conhecimento, 2010.

[2] Cesar Analide, Paulo Novais, and José Neves. Fusão sensorial, 2010.

[3] Cesar Analide, Paulo Novais, and José Neves. Preparação de dados para extracção de conhecimentos, 2010.

[4] Asier Artizia, Juan Carlos Augusto, and Alberto Izaguirre. Spatial and temporal aspects for pattern representation and discovery in intelligent environments. 2008.

[5] J.C. Augusto and C.D. Nugent. *Designing smart homes: the role of artificial intelligence*. Springer, 2006.

[6] A. Baum, J.E. Singer, and J.L. Singer. *Advances in Environmental Psychology: Volume 3: Energy Conservation, Psychological Perspectives*. Advances in Environmental Psychology. Taylor & Francis, 1982.

[7] Fabio Luigi Bellifemine, Giovanni Caire, and Dominic Greenwood. *Developing Multi-Agent Systems with JADE*. Wiley, 2007.

[8] Remco R. Bouckaert, Eibe Frank, Mark A. Hall, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten, and Soeren Sonnenburg. Weka—experiences with a java open-source project. 2010.

[9] R. A. Brooks. The intelligent room project. Washington, DC, USA, 1997. IEEE Computer Society.

[10] Diane Cook and Sajal Das. *Smart Environments: Technology, Protocols and Applications (Wiley Series on Parallel and Distributed Computing)*. Wiley-Interscience, 2004.

[11] Diane J. Cook, Juan C. Augusto, and Vikramaditya R. Jakkula. Ambient intelligence:technologies, applications, and opportunities. *Elsevier*, page 22, April 2009.

[12] Instituto Nacional de Estatísticas (Statistics Portugal). *Inquérito ao Consumo de Energia no Sector Doméstico*, volume 19. 2010.

[13] Faiyaz Doctor, Hani Hagras, and Victor Callaghan. A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 35(1):55–65, 2005.

[14] Brian R. Gaines and Paul Compton. Induction of ripple-down rules applied to modeling large databases. *J. Intell. Inf. Syst.*, 5(3):211–228, 1995.

[15] H. Hagras, V. Callaghan, M. Colley, G. Clarke, A. Pounds-Cornish, and H. Duman. Creating an ambient-intelligence environment using embedded agents. *Intelligent Systems, IEEE*, 19(6):12–20, November 2004.

[16] K. Klein, G. Kavulya, F. Jazizadeh, J. Kwak, B. Becerik-Gerber, P. Varakantham, and M. Tambe. Towards optimization of building energy and occupant comfort using multi-agent simulation. *The 28th International Symposium on Automation and Robotics in Construction*, 2011.

[17] Jun-young Kwak, Pradeep Varakantham, Rajiv Maheswaran, Milind Tambe, Farrokh Jazizadeh, Geoffrey Kavulya, Laura Klein, Burcin Becerik-Gerber, Timothy Hayes, and Wendy Wood. *SAVES: a sustainable multiagent application to conserve building energy considering oc-*

cupants. AAMAS '12. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2012.

[18] L. T. McCalley and Cees J. H. Midden. Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of Economic Psychology*, 23(5):589–603, 2002.

[19] Stuart Moran and Keiichi Nakata. Ubiquitous monitoring and user behaviour: A preliminary model. *J. Ambient Intell. Smart Environ.*, 2(1):67–80, January 2010.

[20] H. Nakashima, H. Aghajan, and J.C. Augusto. *Handbook of Ambient Intelligence and Smart Environments*. Springer, 2009.

[21] United nations. *Report of the World Commission on Environment and Development: Our Common Future: Chapter 2, towards sustainable development.* 1987.

[22] E. Prinet. Sustainable consumption & production. *One earth*, 2011.

[23] Ross Quinlan. *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[24] Darby S. *Making it obvious: designing feedback into energy consumption*, pages 685–696. Berlin, Springer-Verlag, January 2001.

[25] C. Seligman, L.J. Becker, and J.M. Darley. Encouraging residential energy conservation through feedback. *Journal of Economic Psychology*, 3(5):589–603, 1981.

[26] Kevin I. K. Wang, Waleed H. Abdulla, and Zoran Salcic. Fast track article: Ambient intelligence platform using multi-agent system and mobile ubiquitous hardware. *Pervasive Mob. Comput.*, 5:558–573, October 2009.

[27] G. Wood and M. Newborough. Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. *Energy and Buildings*, 35(8):821–841, September 2003.

Chapter 7

Attachments

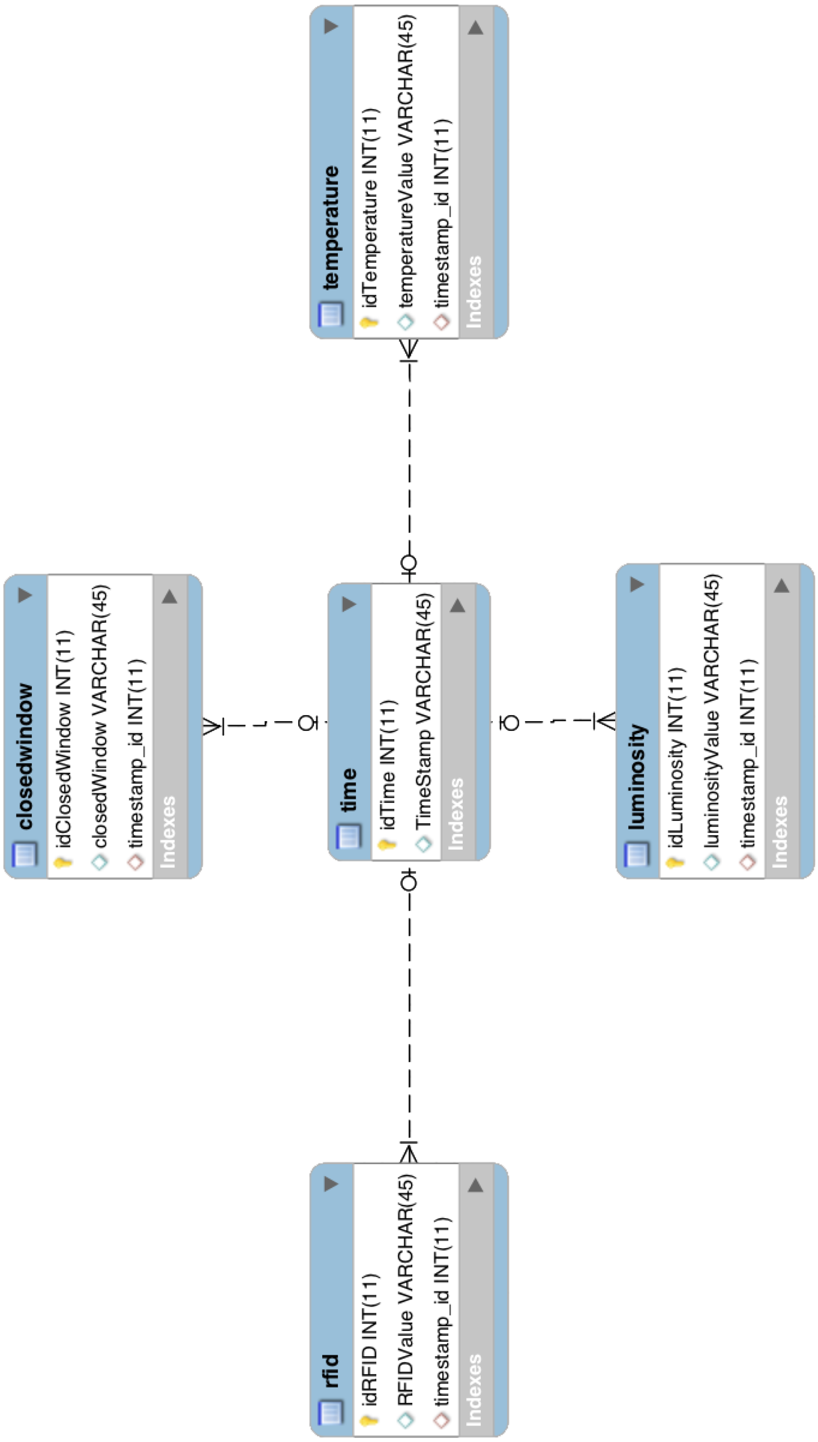

Figure 7.1: The system database schema