A decision support system for risk analysis and diagnosis of hereditary cancer

César Calatrava*, Mauro J. Oruezabal†, Jose A. Olivas*, Francisco P. Romero* and Jesus Serrano-Guerrero*
*High School of Computer Engineering
University of Castilla La Mancha, Ciudad Real, Spain
†University Hospital Rey Juan Carlos, Madrid, Spain

***Abstract*—**The complexity associated with the process of the capture of medical decisions resides in the managing of considerable information deriving sometimes in a series of clinical mistakes. Due to this, the clinical decision support systems are receiving a major importance nowadays in the context of medicine. The family identification with a high risk of suffering hereditary cancer is an arduous labor. The aim of the present work is the improvement of the process of identification of suspicious families that can suffer cancer of genetic character with the development of a decision support system for risk analysis and diagnosis of hereditary cancer.

***Keywords*—**Clinical decision support system, cancer, genogram, genetic factors, probandus

I. Introduction

In recent years, the presence of the computer science has increased considerably in the medical area, up to the point that we can find from tools in charge of the documentary medical management up to clinical decision support system for the decision-making, facilitating of notable form the work of the sanitary personnel.

The main discipline responsible for this unity is Knowledge Engineering [1], whose main purpose is the acquisition, representation and conceptualization of knowledge. All this is orientated to emulate various intelligent human capabilities.

Inside the field of the Engineering of the Knowledge Decision Support Systems (DSS) stand out [2] [3], which try to simulate the knowledge of a person expert in a domain of specific knowledge. These systems arise in order to overcome the limitations of other IT systems popularized previously or to give response and coverage to situations not solved up to the date.

The complexity associated with the process of capture of medical decisions resides in the managing of considerable information deriving sometimes in a series of clinical mistakes, it is shown in some reports [4], which states that these might be the eighth reason of death in the industrialized countries. Due to this, the decision support systems are receiving a major importance nowadays in the context of medicine, receiving the name of Clinical Decision Support Systems (CDSS) [5].

We can define a CDSS as a knowledge system designed to help health professionals to take preventive clinical decisions or diagnostic based on a series of data of the patients. The implication of the patients in the capture of decisions is increasing in the last years, for what the design of the CDSS is orientated increasingly to the integration of his preferences.

The health, nowadays, is a topic of great scope in the society, and the cancer one of his more clear exponents, due to the scanty percentage of patients who manage to overcome successfully the disease. The term cancer [6] includes a numerous group of diseases characterized by the proliferation of abnormal cells, which divide and grow without control being able to invade other fabrics of the body.

A high percentage of the patients that suffers some type of cancer does it in sporadic form, namely, there does not exist any genetic risk [7] of suffering the disease. Nevertheless, there exists a small percentage of patients (5%-10%) that present a tumor hereditary syndrome and that is due to genetic factors that the individual carries from his birth. Accordingly, the fact of inheriting a genetic susceptibility to developing the disease does not want to say that it will end up by developing it, but the risk to suffering it is significantly superior to the rest of the population.

The family identification with a high risk of suffering hereditary cancer is an arduous labor. From the clinical point of view, the identification of a series of characteristics that must put in alert the medical personnel should be important. Some of these characteristics can be the high incident of cancer in the family, the appearance of the cancer at an early age, the occurrence of the same type of cancer, among others.

The aim of the present work is the improvement in the process of identification of suspicious families that can suffer cancer of genetic character, first, with the automatic construction of the genogram [8] (genealogical clinical tree) of the patient based on his health record, and secondly with the evaluation of this genogram, later, to recommend a series of actions to continue or to give advices according to the diagnosis obtained. This will be achieved by the development of a decision support system for risk analysis and diagnosis of hereditary cancer.

The rest of the document has been organized in the following form: section II gathers the previous investigation carried out in the area of interest, paying special attention to the genogram and the standard *Continuity of Care Record* (CCR) as a model of health record, section III describes in detail the parts that shape the system as well as its functioning, in the next section, part IV, the design of an experiment on the system functioning is shown. Finally section V contains the final conclusions and future works of the system exposed in the present work.

II. Related Work

In the process of identification of families suspicious of suffering hereditary cancer there arises the need to compile a series of information both of the patient and of the members of his family to carry out the above mentioned process. This need is covered with a genealogical clinical tree called genogram. A genogram gathers graphically (in the shape of a genealogical tree) the basic information of, at least, three generations of a family. In this case, for the system that exposes in the present document the necessary information of each of the familiar members, included the *probandus*[1] it is the following one: age, sex, presence of tumor or not, type of tumor, in case of existing, age of diagnosis, age of death and reasons of the death.

Diverse tools exist for the construction of genograms, among which we find *GenoPro* [2], this genealogical tool offers us a practical solution at the moment of creating familiar trees and genograms. In addition, it allows the storage of additional information as images and other resources.

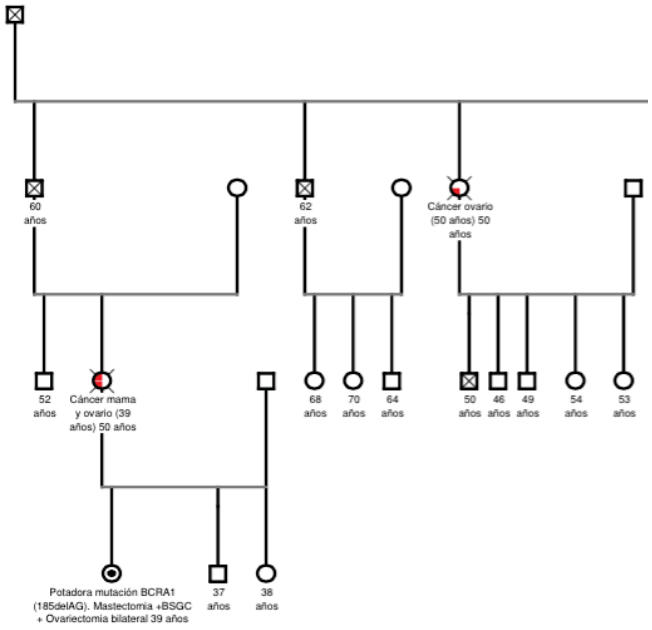

Figure 1. Example of genogram in *GenoPro*

However, GenoPro, which allows us the construction of genogramas to carry out the identification and evaluation of the suspicious families of genetic cancer is not the tool most adapted for this, principally due to the fact that it does not allow the automatic generation of the genograms based on the health record of the patient. On the contrary, it only allows the manual construction and this can lead in some occasions to mistakes. As it was commented previously in the introduction, one of the principal innovations that introduces this system is the automatic construction of the genogram based on the health record of the patient.

[1]an individual affected with a disorder who is the first subject in a study (as of a genetic character in a family lineage)

[2]http://www.genopro.com/es/, (Last visit - April 22, 2015)

Therefore, it is important to have a robust and flexible model of health record to facilitate the task of construction. The model of health record chosen has been the standard *Continuity of Care Record (CCR)* [9], which is a specification of the medical records developed by *AST International* and a set of institutions and companies related to the sanitary sector.

The principal aim of this standard is the improvement in the record of the medical information of the patients. Instead of sheltering a complete record of the patient, the CCR contains the most relevant and important information of the patient allowing this way the exchange between the sanitary personnel. The CCR is designed so that a sanitary expert can create it easily across one electronic health record.

In order to cope with the model of a health record needed by the system, it has been necessary to do an extension of the standard CCR, on the basis of the foregoing, the following objects have been added to the model CCR: HealthStatusActor, Tumor and FamilyTumorHistory.

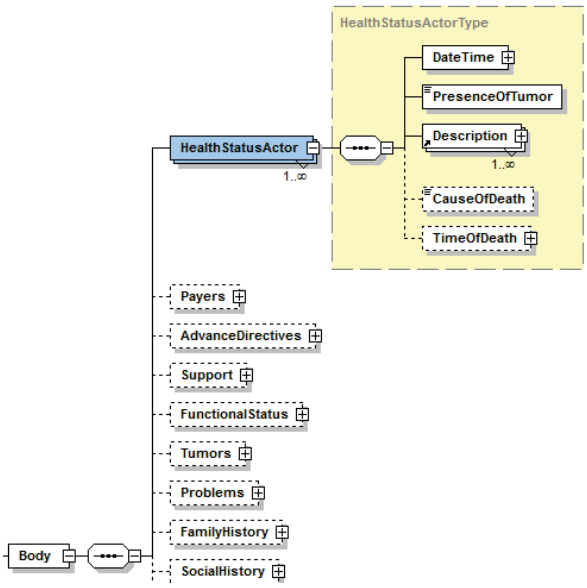

Figure 2. Extension of the *HealthStatusActor* object in CCR

- HealthStatusActor: this object has been added to the model to be able to face the requirement of the expert of indicating the age of the actor (*probandus* or family) and the presence or not of a tumor. For *probandus* will be referenced from *Body* and for the rest of relatives from *FamilyMember*.
- Tumor: the motive of his incorporation has been being able to gather in a precise and correct form the information of the tumor necessary for the correct construction of the genogram. This object is also going to contain two fields for the tumors with typical phenotype and the benign tumors with predisposition to turn in malignant.
- FamilyTumorHistory: object added to *FamilyHistory* to handle the familiar precedents affected by cancer that existed in the family and the healthy ones.

III. Proposal

The principal aim of the system is the improvement in the process of identification of the families that present a high risk of suffering hereditary cancer, as well as of their characteristics or those of a concrete relative for later to value and to recommend the actions to continue according to the evaluation carried out.

As for the structure of the system, we differentiate three parts, on the one hand, the module for the automatic generation of the genogram of a patient based on his health record, and for other one, the part that deals with the analysis and evaluation of this genogram to check the risk the *probandus* or some concrete relative for syndrome of predisposition to the cancer (SPC). Finally, but not less important, a manager of clinical histories for the management of the same ones completes the structure of the CDSS.

Next, each of these modules will be exposed in a detailed form.

With the module of generation of genograms what is looked is to construct a tool capable of generating the complete genogram of a patient of automatic form based on his clinical history. The principal advantage of this tool is that on having treated with health histories that follow a medical standard as the CCR, it allows the possibility of generating genograms of diverse nature, not only orientated to the oncology area as in this case. Other advantages are the considerable profit of time opposite to the manual construction as well as the decrease of the risk to committing mistakes in its creation.

```
▼<Code>
   <Value>394856008</Value>
   <CodingSystem>SNOMEDCT</CodingSystem>
   <Version>2014</Version>
  </Code>
 </ActorRole>
▼<HealthStatusActor>
   <PresenceOfTumor>No</PresenceOfTumor>
```

Figure 3. Fragment of health record in *CCR*

Departing from a health history any one, figure 3, and after a process of transformation which we can divide in two phases, the first, in charge of the extraction of the elements to representing of the clinical record, and the second, which will carry out the graphical representation of these elements, we will obtain in this way the complete genogram of a patient similar to that of the figure 4. The principal advantage that presents representation is that every type of cancer has an own symbol associated. For this reason, the doctor only with a brief glimpse will be able to identify the types of cancer that suffers a family. For example, the bilateral breast cancer is represented with a circle of red color.

As shown in figure 4, of the healthy nodes their current age is registered and of the nodes affected by some type of cancer the age of the diagnosis and the current age is shown. The cases that have deceased are marked by a cross and their age of death is reflected in the genogram, if the reason of the death was due to some tumor will be reflected in the genograma, in opposite case it is not.

On the other hand, the module in charge of the analysis and evaluation of the genogram is formed by two submodules, one will take the previous evaluation of the genogram as a principal function with the aim to generate a preliminary report that contains a series of useful indicators for the doctor. The another submodule has most of the weight of the CDSS, it will evaluate and analyze the genogram of exhaustive form to determine if the family presents a high risk of suffering hereditary cancer.

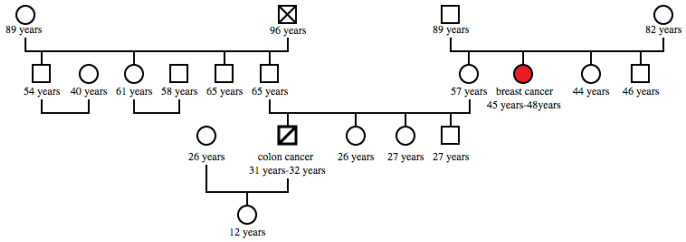

Figure 4. Genogram model obtained after the generation

The purpose of realizing a previous analysis of the genogram is orientated to the use of the system on the part of doctors of primary care, that is to say, doctors not expert in the field of the oncology and who do not present the sufficient knowledge on this speciality. Therefore, this analysis comes to cover the above mentioned lack of knowledge with the generation of a report, with this first analysis the following indicators will be obtained:

- Classification of the genogram in informative or not informative: the genograms can be classified in informative or not informative according to the characteristics that they present. A genogram is informative for the following cases:
  - Hereditary Breast Ovarian Cancer Syndrome (HBOC): existence of at least two women over 65 years in the branch where there are the cases that the syndrome defines.
  - Other syndromes: Existence of at least three persons (man or woman) over 65 years,

  the persons who are exposed above can have or not cancer. The rest of genograms that do not fulfill these criteria will be classified like not informative.

- Associate cases and his relation with syndromes of predisposition to the cancer: correspond to the nodes affected by cancer and that due to the criteria that they present can be classified in someone of the syndromes of predisposition to the cancer as the syndrome of Li-Fraumeni or Peutz-Jeghers.

- Sporadic cases: identification of sporadic cases in the genogram, nodes affected by cancer, but the origin of this cancer does not have any relation with genetic factors, with the syndromes of predisposition to the cancer. His identification is very important in order that the doctor does not confuse them with nodes affected by hereditary cancer.

- Index case and cases not indexed: search and identification inside the genogram of the case index and the cases not indexed. In a genogram the index node is that one that either presents the strangest tumor, for example, a triple negative breast cancer only affects 10% of the population, or it is the youngest node affected by some type of tumor. The rest of nodes of the genogram correspond to cases not indexed.
- Phenocopies: they correspond to the cases of sporadic cancer in a family with a syndrome of hereditary cancer. Seemingly (for the phenotype) they are similar cases affected by syndrome of genetic cancer, but do not have the mutation (different genotype). Therefore, they receive the name of phenocopy the cases that have a similar phenotype, but a different genotype.

In the figure 5, we can see the knowledge map of this module.

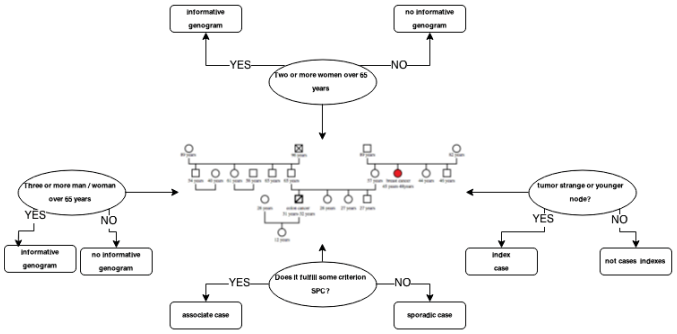

Figure 5. Knowledge map

On the other hand, the submodule in charge of the exhaustive analysis of the genograma is much more complex than the previous one that we have seen, this is due to the dimension of the evaluation that it has to make. It will realize a meticulous analysis of the genograma to verify the characteristics of each one of the nodes that shape the genogram.

It will be contemplated the diagnosis of three types of cancer such as breast cancer, ovary and colon, in addition to various syndromes related to the genetic tumors. On having finished the evaluation, we will obtain a report with the analysis realized on the genogram similar to that of the figure 7 of the next section.

Next, the criteria extracted in the meetings with the expert from one of the tumors included in the system, the cancer of breast and ovary, are exposed:

- *High-risk families*[3]
  - A case of cancer of breast to minor or equal age of 40 years.
  - Cancer of breast and cancer of ovary in the same patient, to any age.
  - Two or more cases of cancer of breast, one of them diagnosed to a minor or equal age of 50 years, or bilateral.
  - A case of cancer of breast diagnosed to a minor or equal age of 50 years or bilateral, and a cancer of ovary in a relative of the first or second degree.
  - Three cases of cancer of breast or ovary (at least one of ovary), in relatives of the first or second degree.
  - Two cases of cancer of ovary in relatives of the first[4] or second[5] degree.
  - A case of cancer of breast in male, and another case of cancer of breast (male or woman) or ovary in a relative of the first or second degree.
- *Families of moderate risk*[6]:
  - Two cancers of breast in relatives of the first degree, diagnosed between 51 and 60 years.
  - A cancer of breast in a relative of the first degree and other one in a relative of the second degree, if the sum of the ages to the diagnosis is minor or equal 118 years.

To conclude, with the development of the module of clinical management one has sought to implement a tool capable of automating to the maximum everything the process of management of the medical reports of the patients. Those clinic records manager takes as basic functions the storage and recovery of the records, beside having a small clinical seeker.

The clinical histories are loaded in a documentary warehouse in order to facilitate the access and the recovery of the same ones at any time determined. The user of the only system must worry about the selection of the history that it wants to load, so the clinic records manager will take charge storing it in his corresponding place. It is important to indicate that the above mentioned manager offers us the possibility of creating collections, the collections allow to group different document that share a series of similar characteristics. For example, if in the system we have diverse of clinical histories corresponding to a series of relatives we might create a collection for this family and hereby be able to guard his medical records of form more organized under the same ”group”.

Normally, as soon as the clinical histories are stored his modification or update will not be necessary. Therefore, the activity of recovery is focused on the most part the consultation of the clinical histories. The doctor will be able to accede to them for the consultation of any aspect of his interest.

Finally, the clinical seeker will be of great usefulness at the moment of realizing searches on one or several clinical histories with the aim to obtain a series of statistics or useful indicators. This seeker will cover different cases of consultation since it can be this, so the doctor will get to know the total number of affected women by cancer of breast minors 30-year-old, between others.

[3]The families of high risk are candidatas to: 1) consultation of genetic advice; 2) analysis of the genes BRCA1 and BRCA2, and 3) measures of follow-up.

[4]Any individual who is separated by a meiosis from one from the members of his family (it is to say, father / mother, brother / sister, son / daughter)

[5]Any individual who is separated by two meiosis from one of the members from his family; relative with whom an individual shares the fourth part of his genes (it is to say, grandparents, grandsons, uncle, aunt, nephew, niece)

[6]The families of moderate risk can benefit from a consultation of genetic advice, and in advisable measures of follow-up of the organs reveille beyond applied in the general population

IV. EXPERIMENT

Once the functioning of CDSS has been described accurately and detailed. We will carry out the design of an experiment where will develop each of the involved steps in the process, from the generation of the genogram up for his evaluation and analysis.

As it was mentioned before in section III, the responsible module for the automatic generation of the genogram will take the health record of the patient as entry, stored in the clinical repository, for further processing. A genogram similar to that of the figure 4 will be obtained. This can be found in the previous section.

The next step in the process of analysis is the evaluation of the genogram for the generation of a preliminary report it contains a set of indicators as indicated previously. Next, figure 6, we show the result of this assessment on a particular genogram.

```
Análisis previo del genograma 23as54ju9032sd2890

                REPORTE PRELIMINAR

- CLASIFICACIÓN DEL GENOGRAMA: Genograma no informativo.

- CASOS ASOCIADOS: Nodo I.1, Nodo II.1, Nodo II.4 y Nodo III.2

- CASOS ESPORÁDICOS: No existen evidencias de ello debido a la edad de aparición del
cáncer en el Nodo I.1.

- CASO ÍNDICE: Nodo III.3 debido a criterio nodo más joven en desarrollar cáncer.

- CASOS NO ÍNDICE: Resto nodos que conforman el genograma.

- FENOCOPIAS: No existen evidencias de ello.
```

Figure 6. Preliminary report after the evaluation of the genogram

With this report the doctor would know quickly and easily the general characteristics of the genogram, likewise, it can use to corroborate or refute his diagnosis on the analyzed genogram.

In this case, we are facing a not informative genogram. It possesses four associate cases. Sporadic cases do not exist, i.e., cases that do not guard relation with any syndrome of predisposition to the cancer. The index node is the *probandus* of the genogram.

```
Análisis final del genograma 23as54ju9032sd2890

DIAGNÓSTICO FINAL: Familia afectada por el Síndrome de Cáncer Mama y Ovario Hereditario
(CMOH)

El probando (III.3), es una mujer de 46 años que tuvo un cáncer de mama a los 35.
Han aparecido otros tres casos de cáncer de mama en mujeres jóvenes (35, 40 y 42 años),
dos de ellos en la misma mujer (II.1). Hay además un cáncer de ovario en una mujer de 40
años.
La enfermedad podría venir por la línea paterna de la primera generación ya que el
individuo I.1 tuvo un cáncer de próstata a los 62 años.
En el CMOH es frecuente observar cáncer de próstata a veces con una edad de aparición más
joven que en los casos esporádicos.
```

Figure 7. Final report after exhaustive evaluation of the genogram

In the figure 7, we can observe the obtained result after the accomplishment of the exhaustive analysis of the genogram. This report is composed of two distinct parts like are diagnosis and analysis or explanation of the same. In the shown example, the family is affected by the Hereditary Breast Ovarian Cancer Syndrome (HBOC) and this diagnosis is argued in the lower part through a series of reasons and taken facts in the analysis that has been realized on the genogram. After this, the doctor will decide the measures to take in relation to family (follow-up more meticulous of the nodes with risk, genetic testing for *probandus*).

In order to provide to the system of different cases of test for his improvement and refinement, it focused mainly on the automatic construction of the genogram and further evaluation, has been carried out the development of an automatic generator of health record based on a Bayes classifier, i.e., a probabilistic classifier based on Baye's theorem. To calculate the probability that a node has of getting cancer or no. Moreover, has been applied 1:

$$\frac{p(type\_cancer) * p(g\_age\_p)}{p(type\_cancer) + p(g\_age\_0) + ... + p(g\_age\_n)} \quad (1)$$

where $p(type\_cancer)$ is the probability which has a node of suffering a certain type of cancer or not, $p(g\_age\_p)$ the probability of suffering this cancer in function to the group of age to which it belongs and $p(g\_age\_n)$ it is the probability that the different groups of age have.

From an objective population, a series of clinical data and the probabilities of incident and mortality for different types of cancer, family health records will be generated to realize an exhaustive training of the module in charge of the automatic construction of the genogram.

Likewise, an analysis has been carried out on the obtained genograms by applying *eye tracking* techniques [11].

The issues, to be evaluated, would be different. On the one hand, we have the comparative evaluation between the expert knowledge acquired of the analysis of the genogram and the royal procedure applied. This question, which identify the rest of the document as (RQ1), it would help additional at the moment of to guarantee or refine the procedure followed by the expert not only at the time of capture steps, if not at the moment of acquiring each and every of the stages that the expert follows for the resolution of the problem.

Furthermore, the evaluation of the presentation of the genogram (RQ2) will be very interesting, with the aim of showing a genogram an expert and/or the system in the most appropriate way, which would result in improved system efficiency and performance.

The experiment involves exposing a genograms battery for evaluation of all of them by the expert. As result there will be obtained the confirmation or refutation of a possible diagnosis of cancer for each of the patients under study.

In the first question RQ1, the analysis of the results would come given by a quantitative value of comparison between the original proceeding raised by the acquisition of knowledge and the actual applied procedure by the expert in the analysis of the genogram, by obtaining the recorded data by the *eye tracking* devices.

For the question RQ2, the analysis would come given by a study of the order of the steps followed by the participant of the study with regard to a genogram, of such a form, that these steps gathered by means of the technologies of eye tracking, possibly leading to a pattern as the order and the observed zone of the sample, in order to improve the presentation of genograms. To finish, the results will rest on a subjective valuation it brings over of the perception of the participant of the study with relation to the presentation of the genograms.

V. Conclusions and Future Work

In this article "A decision support system for risk analysis and diagnosis of hereditary cancer" has been presented. The proposed system pretend improve the identification process of suspicious families that suffer hereditary cancer, the process is arduous and is very important from the clinical point of view the identification of a series of characteristic aspects that present these families his for immediate diagnostic.

The main contribution of the system is the identification and diagnosis of several types of cancer such as breast cancer, ovary cancer and colon, in addition to different syndromes related Li-Fraumeni, Peutz Jeghers, among others. Finally, the automatic generation of the genogram is a great innovation, this fact will lead to a reduction of errors which could be committed in the manual construction of the genogram as well as a considerable decrease in the time of confection of the same.

In the future, we plan to realize various experiments with the aim of improving of incremental form the functioning of the system, thus, an increasingly reliable and robust system will be built. Future research will focus on system expansion in the diagnosis part, that is to say, it will have to evaluate and analyze more types of hereditary cancer and syndromes of predisposition.

References

[1] Montes, C., Gómez, A., Juristo, N. and Pazos, J., *Ingeniería del Conocimiento*. Ceura, 1997.

[2] Scott, M., *Management Decision Systems: Computer-Based Support for Decision Making*. Harvard University Press, 1971.

[3] Keen, P., *Decision Support Systems: A research perspective*. *Massachusetts Institute of Technology*, March 1980.

[4] Corrigan, J., Kohn, L. and Donaldson, M., *To err is human: Building a safer health system*. Committee on Quality of Health Care in America, Institute of Medicine, 2000.

[5] Greenes, R., *Clinical Decision Support: The Road Ahead*. Academic Press, 2011.

[6] Benajmin, R. and Robins, M., *Cancer Biology*, 3rd ed. Harlow, England: Pearson Education Limited, 2006.

[7] Gascón, A., Cruz, J., de Lara, A., Osorio, A., Perea, J., Robledo, M., Bando, H., Benítez. J. and Urioste, M., *Planteamientos básicos del cáncer hereditario; principales síndromes*. ADEMAS Comunicación Gráfica, s.l., 2011.

[8] Kennedy, V., *Genograms*. MAI Review, 2010.

[9] ASTM INTERNATIONAL, *Standar Specifitacion for Continuity of Care Record*. http://www.astm.org/Standards/E2369.htm, (Last visit - April 22, 2015).

[10] Goldstein, M. and Woff, D., *Bayes Linear Statistics: Theory and Methods*. Wiley & Sons Ltd, 2007.

[11] Hassan, Y. and Herrero, V., *Eye-Tracking en Interacción Persona-Ordenador*, 2007.