The presented thesis, entitled "*Application of semantic technologies for the creation of high sensitivity medical differential diagnosis intelligent systems*", has been developed with the aim of demonstrating that the application of semantic technologies allows the creation of high sensitivity medical differential diagnosis systems.

In order to provide a better understanding of this summary, this chapter will be divided in several sections which will represent the main chapters that have been presented before.

1.1. STATE OF THE ART

The state of the art of this thesis has been divided in two main parts. In the first part, a wide analysis of the research and the systems related to the main topic of the thesis was presented, focusing on these systems, which have been developed with the aim of providing differential diagnosis systems with a general purpose. Among these main researches we can enhance some works such as MYCIN (Shortliffe & Buchanan, 1975), DXPlain (Hupp et al., 1986), Isabel (Graber & Mathew, 2008) and DiagnosisPro (Aronson, 1997).

In the second part of the state of the art a brief analysis of the main technologies which has been used to verify the thesis hypothesis was presented. Between these technologies we should mention the Semantic Technologies as part of Artificial Intelligence (Berners-Lee, 2006), Rule-based systems, logic-based inference, and decision support systems technologies.

1.2. HYPOTHESIS

As discussed above, the main research done could be summarized as a research based on demonstrating the use of semantic technologies for developing medical differential diagnosis systems.

The research has been specifically focused on verifying if these technologies represent a good technological support for developing high sensitivity diagnosis systems. However, apart from this main hypothesis, other research hypotheses were defined in order to provide an accurate research definition.

Basically, three hypotheses were defined:

1. The use of semantic technologies allows the creation of high sensitivity medical differential diagnosis with accuracy and efficiency. The concept high sensitivity makes reference to the ability of a system to provide results even when the number of inputs is too low.
2. It is possible to solve the multilevel diagnosis applying these technologies. Multilevel diagnosis makes reference to the ability of determining a disease even when the diagnosis criterion of the disease is formed by other diseases.
3. It is possible to create a process which calculates diagnosis alternatives in the approach in which the signs or symptoms of a disease were induced by drug consumption, and for hence, its presence is not a finding of the pathology to diagnose.

1.3. KNOWLEDGE REPRESENTATION

The representation of the knowledge was one of the main problems which were set out at the beginning of this thesis. An extended analysis of the situation was done in order to achieve the best solution. The solution which was finally chosen was the use of diagnostic criteria as representation model. In the analysis performed to check the different terminologies or ontologies which nowadays exist in healthcare domain we have included SNOMED-CT (Rogers & Bodenreider, 2009), Open GALEN (Amaral et al., 2000) and OBO-Foundry (Smith et al., 2007) among others.

After a painstaking analysis of these systems and models, a new ontological model was developed in order to achieve the necessities to verify the hypotheses. The model was developed using SNOMED-CT as a terminology model. A new ontology architecture model was developed using root ontology as meta-ontology to define the relations which exist between the diseases and the findings (signs, symptoms and other diseases).

Another important factor in the knowledge representation process is the information source. In this chapter, where the information comes from is explained. In this case, given that we are working with very sensitive information (medical information, and more concretely, diagnosis information, it should come from reliable sources which allow the validity of the data to be guaranteed), the information source used was medical books which include titles such as Harrisons (Harrison, 2009). A deep research was done with these books as medical reference to extract the information related to the disease, in order to generate the necessary knowledge which will be used in the inference process. It is important to highlight that the use of automatic process shouldn't be followed because the reliability of the information provided by automatic process is very low. The medical information should be contrasted by experts.

The knowledge representation chapter also includes information about how the ontology population process was done. This process includes the generation of the instances which the diseases and findings represent, the relations between these instances and the generation of the rules.

1.4. ODDIN, ADONIS AND SEDELO

Chapters 5 and 6 show information about the first version of the systems developed to verify the hypotheses of the thesis.

ODDIN (García-Crespo et al., 2010), was the first developed thesis version. This system was developed using a legacy representation knowledge model to the one presented in the representation knowledge chapter. In ODDIN, a first version of the representation model was developed using description logics. The preliminary results of this development show that the appliance of description logics to this field provides very poor results regarding the temporal efficiency of the system. However, also a first version using Jena Rules was developed, providing much better results than the previous approach. Nevertheless, the main lack of ODDIN system was in the fact that the system was not able to deal with multilevel diagnosis problem. The system was able to perform simple and complex diagnosis, but it was not ready to perform multilevel diagnosis.

One important factor that was also developed in the ODDIN system is the ability to generate multiple alternative diagnosis results depending on the drugs which the patient is taking; this factor will be used for the verification of hypotheses H3.

ADONIS (Rodríguez-González, 2009b; Rodríguez-González et al., 2011) and SEDELO (Rodríguez-González et al., 2011a) systems suppose an improvement to the previous ones (ODDIN). The systems were developed based on the description logics of ODDIN. A new version of these description logics was developed in order to allow these systems to perform multilevel diagnosis. ADONIS system solves the multilevel diagnosis problem partially, adding the problematic of making normal inference impossible in some contexts.

In SEDELO, a new improvement to the description logics developed in ADONIS was done in order to solve the problems introduced in this system. However, the efficiency problem arose once again, making the use of the system in real environments impossible.

1.5. MDSS-OPM

MDSS-OPM is the final version of the systems developed in this thesis. The main difference between this system and the previous ones is the ability to perform multilevel diagnosis with efficiency. Basically, the new model of rules which allows the performance of the multilevel diagnosis is presented in this chapter. The rule language used, as in ODDIN, is Jena Rules language. The reason to use this language is based on the capacity of this language to implement a negation process, which is not possible to do with other semantic rule languages like SWRL, for example. This chapter also includes an interesting approach to understand the inference process which is behind the Jena rules engine and the new ontology design based (Rodríguez-González et al., 2012).

Another important aspect which is presented in this chapter is a probabilistic model based on epidemiological data. This model is presented as a proof-of-concept model for calculating the probability of diagnosing a disease based on epidemiological data.

1.6. EVALUATION

The evaluation chapter shows the main aspects of the evaluation process followed in this thesis.

In the evaluation of a Diagnosis Decision Support System (DDSS) several approaches could be followed. In this chapter, an analysis of the main tools which are normally used for the evaluation of this kind of tools is done. The evaluation of this type of systems is normally done following three main approaches: evaluation of diagnosis accuracy, evaluation of temporal and spatial complexity and evaluation of the techniques employed.

The analysis of the state of the art reveals that there is an important lack of standard methodologies to measure these approaches. Within the context of this thesis, the most important factors to measure are accuracy and temporal complexity. For this reason, given this lack, a standard methodology for the measurement of the accuracy and temporal complexity of the approach developed has been created.

The methodology created is based on the work done by Kaplan (2001). Kaplan states that the measurement of DDSS accuracy should be done following one of these three approaches: 1) comparison of the system with other systems; 2) comparison of the system with a gold standard; or 3) comparison of the system with experts.

Given that the comparison with other systems is not applicable (different knowledge bases and different evaluation approaches) and there is not current gold standard to compare the results, the option which will be followed is the comparison of the system with experts.

For this, a methodology based on the creation of clinical cases to be solved by both experts and the system was developed. Once the results are provided by experts (assessors) and the system, an arbitration process is made to confirm the results provided (using other experts (referees) for this purpose). Once the arbitration process is finished, binary classifier metrics (Precision and Recall (Sensitivity) (Makhoul et al., 1999), Accuracy and Specifity - PRAS) are used to measure the accuracy of the system. Apart from these basic metrics, a Matthews Correlation Coefficient (MCC (Matthews, 1975; Baldi et al., 2000)) is used. MCC metric provides balanced results of PRAS metrics, allowing us to have a single result of the comparison between system and assessors.

This methodology includes the analysis of temporal efficiency of the system and the assessors and the correlation which exists among the referees that participate in the arbitration process.

Finally, in this section, an analysis of the techniques used in the development of the thesis compared with other techniques was done.

1.7. VERIFICATION

The verification of the hypothesis set previously is done by using the results of the evaluation. The first hypothesis, which states that "*The application of semantic technologies allows the creation of high sensitivity medical differential diagnosis with accuracy and efficiency*", was verified using the results provided by the evaluation.

Concretely, the verification was based on the premise that we know that the developed systems for the verification of the hypotheses used semantic technologies. Based on this, we only need to verify that the developed system is a high sensitivity system (high recall rate) which is accurate (high MCC rate) and efficient. Once we check the results of the evaluation, we can see that the recall rate of the system is 97.92 against 70.38% of the assessors. In the case of MCC rate we obtain a 93.59% for the system and a 61.40% for the assessors. In both cases, a statistical analysis using a T-Student confirms that there are significant differences between system and assessors, confirming this part of the hypothesis from a statistical point of view. In the case of the efficiency, both the tables showed in this part of the evaluation and the T-Student performed again show significant differences between system and assessors.

In the case of the second hypothesis, which states that "*It is possible to solve the multilevel diagnosis applying these technologies*", it was verified by using the results provided in Verification section. In fact, this hypothesis could be automatically verified when hypothesis H1 was verified, given that the system was developed to perform multilevel diagnosis. However, to provide a better understanding of this hypothesis and how it was verified, the verification chapter shows more information about this verification.

Finally, the third hypothesis, which states that "*It is possible to create a process which calculates diagnosis alternatives in the approach in which the signs or symptoms of a disease* *were induced by drug consumption, and for hence, its presence it is not a finding of the pathology to diagnose*", was verified by a qualitative evaluation using referees.

1.8. CONCLUSIONS

The research of this thesis was focused on demonstrating the application of semantic technologies for the creation of high sensitivity diagnosis decision support systems.

In this development, the requirements of accuracy and efficiency to the developed systems are included.

Knowledge representation is one of the topics which this thesis has dealt with, focusing on medical knowledge representation. However, this knowledge, as seen before, has several ways to be represented and used, using terminologies, taxonomies and similar elements. However, given that the concrete medical knowledge which this thesis makes use of is the one that depends on the differential diagnosis domain, a redefinition of the current knowledge representation elements was necessary.

Apart from the knowledge representation used in the diagnosis process, the generation of a set of inference elements with the aim of verifying the hypotheses set out was also necessary. The first solution which was applied, allows the verification of these hypotheses partially, however, it generates a set of problems refereed to the efficiency. For this reason, using rule-based systems in Jena Rules format instead of description logics was agreed. The reason to use Jena Rules was that this rule language allows emulating the negation concept, necessary to establish a close world assumption (CWA) which allows the generation of diagnosis which fulfill with the requirements of high sensitivity, accuracy and efficiency. These rules were also applied to generate the solution which allows performing multilevel diagnosis.

The implementations of the systems created with the aim of verifying the hypotheses also need an evaluation process to confirm this verification. For this reason, an evaluation methodology was created to evaluate the accuracy and efficiency of the systems created. In this context, binary classifier metrics where used, including Precision, Recall (Sensitivity), Accuracy and Specifity, and adding a final summary metric: Matthews Correlation Coefficient (MCC).

In this document, in the evaluation and verification chapters only summary results were showed. However, the evaluation performed reveals very interesting results like the behavior of the assessors in the diagnosis of each of the diseases which are included in the knowledge base or the reason behind the high sensitivity results among several others.

To sum up in only one sentence, the final conclusion from this research is that the application of semantic technologies allows the creation of accurate and efficient high sensitivity diagnosis systems, which perform multilevel diagnosis using ontologies as knowledge representation model.

REFERENCES

Amaral, M., Roberts, A. & Rector, A. (2000). NLP techniques associated with the OpenGALEN ontology for semi-automatic textual extraction of medical knowledge: abstracting and mapping equivalent linguistic and logical constructs. *Proceeding of AMIA Annual Symposium*, 76-80.

Aronson, A. (1997). DiagnosisPro: The Ultimate Differential Diagnosis Assistant. *Journal Of the American Medical Association* **277(5)**, 426.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. F. & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* **16**, 412-424.

Berners-Lee, T. (2006). Artificial Intelligence and the Semantic Web. *W3C Talk*.

García Crespo, A., Rodríguez Gonzalez, A., Mencke, M., Gómez Berbís, J. & Colomo Palacios, R. (2010). ODDIN: Ontology-driven Differential Diagnosis based on Logical Inference and Probabilistic Refinements. *Expert Systems with Applications*, **37(3)**, 2621-2628.

Graber, M. & Mathew, A. (2008). Performance of a Web-Based Clinical Diagnosis Support System for Internists. *Journal of General Internal Medicine* **23(1)**, 37-40.

Harrison, T. & Anthony, S.Fauci, A., (ed.). (2009). *Harrison: Principios de medicina interna*. McGraw-Hill.

Hupp, J., Cimino, J., Hoffer, E., Lowe, H. & Barnett, G. (1986). DXPlain: a computer-based diagnostic knowledge base. *Proc. of Fifth World Congress on Medical Informatics*.

Kaplan, B. (2001). Evaluating informatics applications—clinical decision support systems literature review. *International Journal of Medical informatics* **64(1)**, 15-37.

Makhoul, J., Kubala, F., Schwartz, R. & Weischedel, R. (1999). Performance measures for information extraction. *Proceedings of DARPA Broadcast News Workshop*.

Matthews, B. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **405**, 442-451.

Rodríguez-Gonzalez, A., Garcia Crespo, A. & Colomo Palacios, R., Labra-Gayo, J.E., Gomez-Berbis, J.M., Alor-Hernandez, G. (2011). Automated Diagnosis through Ontologies and Logical Descriptions: the ADONIS approach, *International Journal of Decision Support System Technology (IJDSST)*, **3(2)**, 21-39

Rodríguez-González, A., Labra, J., Alor-Hernandez, G., Gomez, J. & Posada-Gomez, R. (2009b). ADONIS: Automated diagnosis system based on sound and precise logical descriptions. *Computer-Based Medical Systems, 2009. CBMS 2009*, 1-8.

Rodríguez-Gonzalez, A., Labra-Gayo, J., Colomo-Palacios, R., Mayer, M., Gomez-Berbis, J. & García-Crespo, A. (2011a). SeDeLo: Using Semantics and Description Logics to support aided clinical diagnosis. *Journal of Medical Systems*.

Rodríguez-González, A., Hernandez-Chan, G., Colomo-Palacios, R., Gomez-Berbis, J.M., García-Crespo, A., Alor-Hernandez, G. (2012). Towards an Ontology to support semantics enabled Diagnostic Decision Support Systems. Special issue on Semantic Web and Healthcare. Current Bioinformatics. (Impact Factor 2010: 0.976) (to appear)

Rogers, J. & Bodenreider, O. (2009). SNOMED CT: Browsing the Browsers.

Shortliffe, E. & Buchanan, B. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences* **23**, 351-379.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L., Eilbeck, K., Ireland, A., Mungall, C., OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S., Scheuermann, R., Shah, N., Whetzel, P. & Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* **25**, 251-1255.