

Knowledge Security

An Empirical Use of IT – Child Abuse Monitor System Model

Tiago Pereira¹ and Henrique Santos²

¹*Algoritmi Centre, University of Minho, Azurém, Guimarães, Portugal*

²*Information Systems Department, University of Minho, Azurém, Guimarães, Portugal*

Keywords: Health Care Knowledge Sensitivity, Health Care Decision Support System, Ontology, Health Care Knowledge Security, Knowledge Management, Topic Models, Information Retrieval, Text Mining.

Abstract: The Information Security, nowadays, faces new threats such as the process of massive information in which are applied artificial intelligence techniques with the goal of predicting and classify our actions. Thus, knowledge about our behaviour, likes, dislikes, among others, leads us to consider that Knowledge Security appears has the natural evolution of Information Security. On the other side of the same coin we have new possibilities to monitor health, the wellbeing and abnormal symptoms, reactions to treatments, alert for insulin insufficiency, pacemaker malfunction, among others. Child abuse cases, it is a subject of most importance in our society, although, these cases are, from suspicion to signalization, difficult to identify since strong evidences are needed. Typically, health care services deal with these cases in an earlier stage with evidences based on the emergency diagnosis, but, yet, not sufficient and with lack of information, thus, further analysis is needed from experts' teams. The main goal of these teams is to protect the child from the possibility of occurrence of more abuses. We have developed a prototype that automatically could predict and alert to situations that could be needed to use the measure of the protection of the child, using digitalised child abuse processes, knowledge management and artificial intelligence techniques with 83% of true positives. In this research, we addressed both sides of the coin, Knowledge Security and the benefits of the Knowledge Discovery defining, in our opinion, the fourth generation of Knowledge Management - Value Creation and Knowledge.

1 INTRODUCTION

Knowledge Security it is seen here as a natural evolution from information security. Currently, we face new paradigms concerning the context of "BigData", the huge amount of data that is processed, the machine learning algorithms and other artificial intelligence which are implemented, in such way that is possible to predict our present and future actions, in other words, knowledge about our behaviour, preferences, among others. Thus, the protection of the right to privacy, promotion of confidentiality, integrity and availability - fundamental properties of InfoSec (ISO, 2013, Oladimeji et al., 2011). The context of knowledge Security demands a new approach. To promote, protect and preserve knowledge about us is necessary to concern about knowledge management and information security management that indicates that we should define what we need to protect. This

decision, it is related to the value of that knowledge represents to us. Concerning the organizations, the situation is similar. If we want to implement security treatments, we should define the security object to protect. In this research, the health care issues were a mean to demonstrate the need of knowledge security, also to realize that in what kind of form Knowledge Sensibility allow the characterization of critical knowledge. Based on this premise, we have formulated the research question: Can we automatically classify health care information as critical concerning laws and regulations, terms and knowledge sensibility in order to preserve it? Using the design science research methodology, we have developed a prototype that can distinguish child abuse documents from others based on health care knowledge (symptoms, effects, among others) allowing us to identify a case that should have further analysis by child protection committee and by its critical value if it should or not be

implemented the child protection by prevention. The Childhood protection is a subject with high value for the society, but, the Child Abuse cases are difficult to identify (JN, 2013). The process from suspicious to accusation is very difficult to achieve. It must configure very strong evidences. Typically, Health Care services deal with these cases from the beginning where there are evidences based on the diagnosis, but they aren't enough to promote the accusation. Besides that, this subject is highly sensitive because there are legal aspects to deal with such as: the patient privacy, paternity issues, medical confidentiality, among others.

2 HEALTHCARE CRITICAL KNOWLEGE

The development of standards of health care software is a big step to the interoperability between information systems in this area. There are, at least, six entities that have developed standards in this field: The American Society for Testing and Materials with ASTM-E31, The American National Standards Institute with ANSI-HL7, The European Committee for Standardization with CEN-TC251, The International Organization for Standardization with ISO-TC215, The Association of Electrical Equipment and Medical Imaging Manufactures with NEMA-DICOM and IEEE with multiple standards (IRMA, 2013). With the evolution of the health care information systems the access to patient information using Electronic Health Record Systems (EHRS) is facilitated, e. g. Urgency treatment data, health monitoring data, among others. According to the health domain analysis report from the technical committee from HL7 about security and privacy of health care information, particularly, in exchange of information between information systems and according to HL7 Security and Privacy Ontology was possible for us to identify critical knowledge concepts in health care domain (WG, 2013, HL7, 2010): Substances abuse, Sexual abuse and domestic violence, Genetic disease, Sexual transmitted disease, Sickle Cell, Sexuality and Reproductive, HIV/AIDS, Psychiatry and Taboo (HL7, 2010), see figure 1.

From this, we explore the subject General Abuses with the focus on the child abuse and based on regulations and legal documentation we have constructed an ontology that maps the concepts: symptoms, behaviour and other evidences of child abuse (Saúde, 2008). For the different phases and

objectives of this research we have used multiple research techniques: survey literature techniques, content analysis and proof of concept in Design Research context.

“Design Science research is a research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artefacts, thereby contributing new knowledge to the body of scientific evidence. The designed artefacts are both useful and fundamental in understanding that problem” (Hevner and Chatterjee, 2010).

The design science research has its roots in the sciences of the artificial. Artificial as something that is created by humans that doesn't exists in Nature. Design Research is fundamentally a problem-solving paradigm. It consists in seeking innovation through ideas, practices, technical abilities and products obtained from a set of routines such as: analysis, design, implementation, and use of information systems concerning the effectiveness and efficiency achievement on organizations. The outputs forms of Design research could be: constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems).

3 CHILD ABUSE CRITICAL KNOWLEDGE MONITORING SYSTEM

The model of the system, see figure 1 and 2, is defined by four components: the knowledge capture component that allows to extract topics and the descriptors from documents, the critical knowledge ontology component that is used to filter the critical topics and descriptors addressing the context of ontology; the critical knowledge repository component that stores all files produced in all process of the system; and the alert and log component that we need the use of Artificial Intelligent algorithms and techniques that allow to predict actions. Each component is based on a variety of information systems fields (Pereira and Santos, 2012, Pereira and Santos, 2013a).

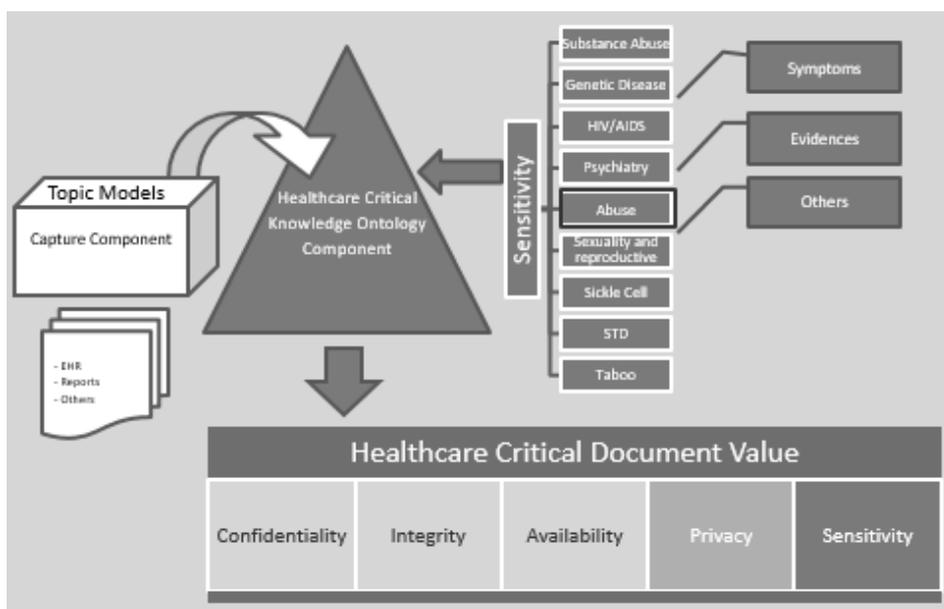


Figure 1: Child Abuse Critical Knowledge Monitoring System.

3.1 The Knowledge Capture Component (KC)

The knowledge capture component requisites are: extracting tokens from documents in a variety of formats, such as text and audio. Additionally, the implemented system supports other formats: video (extracting sound and text within the video), webpages, among others; and transform the tokens extracted in such format that could be searchable concerning the privacy and confidentiality, integrity and availability of documents. In order to do it, we have implemented a topic model approach using two methods, Latent Dirichlet Allocation - LDA (Hofmann, 2001, Steyvers and Griffiths, 2007) and Pachinko Allocation Model - PAM (Mimno et al., 2007, Pereira and Santos, 2015) and not that kind of techniques that, normally, extract names, locations, among others. The PAM has been chosen because it can establish relations between topics and topic descriptors. The use of the topic model approach is fundamental because allows to driven topics from documents and ignore (because of its lower occurrence within a document) personal data (names, contacts and addresses) complying with the privacy and information security properties. To use topic models from the extracted tokens from the documents we needed to filter (Efraim Turban, 2010) them, essentially, tokens less than four (configurable) characters and trivial discourse tokens such as “and”, “or”, punctuation, among others. As an output of this component we get a searchable set

of descriptors clustered by topics that co-occur in the document.

Table 1: Knowledge Science fields used in Knowledge Capture Component.

KC Component – Knowledge fields
TextMining
Information Retrieval

3.2 The Critical Knowledge Ontology Component (CKO)

The critical knowledge ontology component requisites are: allowing the editing of the critical ontology; and matching the ontology with the output of the knowledge capture component. This component uses Portégé Editor and Portégé API, see acknowledgments, for the matching procedure with the topic descriptors.

Table 2: Knowledge Science Fields used in Critical Knowledge Ontology Component.

CKO Component – Knowledge fields
Health Care / Child Abuse - Sensitivity
Knowledge Engineering
Knowledge Management

3.3 The Critical Knowledge Repository Component

The critical knowledge repository component

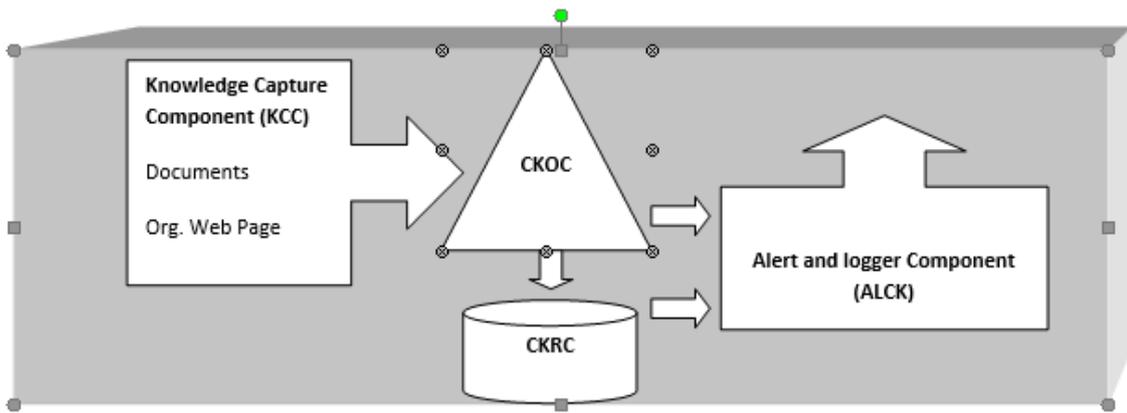


Figure 2: Critical Knowledge Monitor System Model.

requisites are allowing the storage of the outputs from the KCC and CKOC. This component uses a document management tool with the control version capability. The control versions could be useful in future implementations and we could analyse multiple diagnosis of the same patient in an historical perspective.

Table 3: Knowledge Science fields used in Critical Knowledge Repository Component.

CKR Component – Knowledge fields
Document Repository Management
Information Security – File Encryption

3.4 The Alert and Log Component

The alert and log component requisites are: alerting the user by email of the probability of the document containing evidences about a child abuse; and register the evidences identified by the system of each document for further analysis. The system should select what cases should trigger an alert or register only. In order to do this, we'll use artificial intelligence classification algorithms very well tested and implemented to assess the evidences and give value to the sensibility of the document in this context.

Table 4: Knowledge Science fields used in Alert and Log Component.

AL Component – Knowledge fields
Artificial Intelligence – Classification algorithms
Information and Communication Technology

4 RESEARCH RESULTS

With the precious help of the Young and Child

protection Committee of Fafe, we could have accessed to 16 anonymized processes of child abuses. The goal of this committee is to decide whenever they should apply the protection measure of the child, based, mostly, on the accusation report, medical reports, teachers and school reports and diligences. Ten of the processes refer to cases in the final stage, so we used them to train the Child Abuse Critical Knowledge Monitoring System. The other six processes were in early stages and we've used to validate the accuracy of the system for the detection of a child abuse case to the prediction of the measure to adopt. All of the six cases were detected as child abuse cases. Two of them, the system considered that they don't need child protection with a certainty of 99% - Group A. Another two, the system considered that should have a child protection measure with certainty of 99% - Group B. Finally, the last two the system considered that should have the measure with certainty of 70% and 89% - Group C. From the perspective of the committee of the group A, independently of the system considerations, they don't need the child protection measure. For the group B, the committee already have implemented the measure. Finally, for the group C, the committee considered that they shouldn't have the measure of child protection. We've communicated this results to the committee and one and a half month, later, they reply to us. In the case of 89% of certainty of the system, the committee, clearly, decided to apply the child protection measure. For the 70% case of certainty they apply the measure but, because it has been changed family/home context of the child that lead to that conclusion.

4.1 Results Validation

The system was validated in three stages:

Laboratorial context, Real context as stated in the previous paragraph. Since we developed the prototype with the knowledge engineering and management body of knowledge we have validated the system with the matrix of quality dimensions of knowledge management and knowledge management system (Pereira and Santos, 2013b).

4.1.1 Laboratory Context

The objectives of laboratory context were validating the system requisites and systemic function of all components. In this, were used documents in other areas beside child abuse to validate false positives. And documents of the theme child abuses to validate the true positives.

Table 5: Child Abuse process characteristics - Final Stage.

Process	Words (OCR)	Pages	Protection Measure	Court Decision
CPCJ1	2543	13	No	Não
CPCJ2	1373	5	No	No
CPCJ3	3725	10	Yes	No
CPCJ4	1352	4	Yes	No
CPCJ5	604	2	No	No
CPCJ6	918	3	No	No
CPCJ7	1664	4	Yes	Yes
CPCJ8	1433	4	No	No
CPCJ9	1397	4	Yes	No
CPCJ10	781	3	No	No
Total	15790	52	10	

4.1.2 Real Context

In the validation on real context the processes of child abuses that we had access have these characteristics, see tables 5 and 6:

Table 6: Child Abuse process characteristics - Earlier Stages.

Process	Words (OCR)	Pages	Protection measure	Court Decision
CPCJ11	128	1	Undefined	No
CPCJ12	932	4	Undefined	No
CPCJ13	694	2	Undefined	No
CPCJ14	340	2	Yes	No
CPCJ15	249	2	Undefined	No
CPCJ16	842	3	Yes	No
Total	15790	52	6	

4.1.2 The Matrix of Quality Dimensions of Knowledge Management and Knowledge Management Systems

The matrix of quality dimensions of knowledge management and knowledge management systems allows us to look to our management knowledge system on three different dimensions: System, see tables 7, 8 and 9, User, see table 10 and Organization and perform an auto critic analysis. Since this is a prototype and not a final product with broad implementation, we cannot evaluate the organization dimension that addresses subjects, such as: Organizational support, Perceived usefulness of knowledge Sharing/ Intent to use/ Perceived benefit, Knowledge/System/Net benefits, Knowledge / Information quality and service quality.

Table 7: System Quality Dimension.

System Quality	
Stability	The lab experiment allows to see that the documents output is the same in different times of analysis and it kept on real documents.
Response time	The system has the capacity of processing 273 words per second and 1,11 seconds per page.
User-friendly and ease of use	Since this is a prototype this aspect is less treated in this research.
Knowledge Classification	The system is capable of relevant knowledge recognition in documents of the target subject.
Technical resources	According to knowledge management this prototype is capable of capture knowledge from a variety of document types, identifies knowledge addressing an ontology and applies algorithms of classification to validate the identified knowledge.

Table 8: Ontology Quality.

Ontology Quality	
Accuracy and Correctness	The ontology was based in child abuses regulation and procedures of medical staff (Saúde, 2008).
Authority	Besides it was based on existing HL7 ontology, the child abuse ontology it's our authorship.
Clarity	All the ontology concepts are identified and supported by regulation.
Interpretability	The concepts mapped on this ontology are from Healthcare area, since the users are from healthcare to the interpretability facilitated.
Completeness/Coverage	All the concepts of child abuse mentioned in regulation were mapped on this ontology.
Consistency	There's no ambiguity in the concepts although, the ontology is multiple language (English and Portuguese).
History	The ontology is based on an HL7 ontology and is suitable to be accoupled to it.
Infrastructure	The ontology infrastructure is scanned, automatically, by the prototype so its definition is correct.
Knowledge Sharing	The ontology is OWL compliant so it can be used and updated by others.
Lawfulness	The ontology is based on regulation and it preserves the privacy by technical implementation.
Metadata Evolution	The ontology is OWL compliant so it can be extended with other ontologies.
Minimality	The concepts of the ontology are reduced to its atomicity.
Purpose	The ontology achieves completely the objectives.
Relevance	This ontology its fundamental to the functionality of the prototype.
Richness	The ontology addresses the concepts and its relation so it's more valuable than dictionaries or bags of words.
Security	The ontology editor allows us to control the access to the ontology, although, in this research it's not addressed.
Strategy	This ontology is in a standard format so it could be updated as the theme evolves.
Traceability	The ontology editor allows us to control the versions of the ontology and restore them, although, in this research it's not addressed.

Table 9: Knowledge Retainer Quality.

Knowledge Retainer Quality	
Accuracy	The repository component is based on opensource software, MongoDB. It's a repository oriented for documents that's is appropriated to distributed systems with replication technics and unique identifier of documents.
Authority	Since it's opensourced there is no problem about authorship.
Expertise	This is a very specific software to management documents with scalability. This was the main issue that lead us to choose it.
Consistency	The repository by redundancy assures high consistency to documents access.
Credibility	This software has the maturity of 10 years since it was developed in 2007 and largely used.
Degree of detail	For each document, it's possible to include Metadata allowing us to add its characteristics and notations.
History	The MongoDB has implemented a control version that allows us to access previous versions
Reuse	The prototype allows us to process again documents by option.
Relevance	This tool is essential for the performance of this prototype and possible adaption to bigdata systems.
Sharing usefulness	Since this tools is prepared to distributed systems context it allows multiple users with high performance.
Degree of context	Since it's a document repository allows to store and retrieve all the documents produced in the analysis process of the system.
Accessibility	The characteristics of this tool defines an high availability to documents.
Degree of socialisation	Since this prototype deals with restricted access documents – critical knowledge this variable doesn't apply to it.
Security	All the security mechanisms, such as: access control, SSL, among others are possible to implement on this system.
Willingness to share	Since this prototype deals with restricted access documents – critical knowledge this variable doesn't apply to it.

Table 10: User Dimension.

User Dimension	
Job performance	This prototype could increase the efficiency on the identification of child abuse cases and promote faster the child protection measure.
Productivity, Easefulness tasks and system	The results achieved with this system allows us to interpret that directly the productivity could be increased and we could alert child abuses cases that need to be further analysed.
Knowledge that meets the needs	This prototype is very specific to healthcare area, but it's possible to ply it to any area of knowledge and any language support.
Content	The prototype describes the relevant evidences presented in child abuse process.
Accuracy	With the available results and without a broad application the prototype presents us good quality in child abuses cases identification.
Format	The prototype was developed in a programming language that could be used in all operating systems.
Ease of use and timeless	Since this is a prototype the easiness of use was not the principle concerned to us. Although, we can consider that the time of the all process has good performance.

5 CONCLUSIONS

Beside the small sample that we have access to, by the kind collaboration of young and child protection Committee (YCPC), we could achieve these results: from sixteen child abuses processes documents, all of them were identified by the system as pertinent to child abuses context. Ten of the processes had complete all phases of the YCPC analysis and they were used in training of the system in the model definition and the other six that were in the preliminary stages of the YCPC analysis, were used as a test of the model. As a final result, the system was capable of identify two processes as critical in order to apply the child protection and the same processes had that measure applied by YCPC. Two others were not considered by the system and by the YCPC as critical to use the child protection measure and the last two processes were considered critical by the system but not yet been classified by YCPC.

Based on these results, it's possible to conclude that with the definition of the sensible topics in health care, with the use of laws and regulations and health care terminology promote the identification of the sensible concepts to develop an ontology about child abuse cases critical knowledge that objectively address the research question - Can we automatically classify health care information as critical concerning laws and regulations, terms and knowledge sensibility in order to preserve it? Thru knowledge sensibility it was possible to get the critical documents and the prediction of the use of the child protection measure that is the main goal of the organization - YCPC. The ontology of the health care critical knowledge, the critical knowledge monitor system (prototype) and the knowledge sensitivity applied to knowledge security are the contributes of this research.

Since this prototype is generic, as long as we apply to it other ontologies and objectives, we could address other aspects of Health Care Sensitivity, such as: Adults Abuses, Substance Abuse, Genetic Diseases, HIV/AIDS, Psychiatry, Sexuality and Reproductive, Sickle Cells, Sexual Transmitted Diseases and Taboo, identified by HL7 initiative as Health Care Sensibility areas. Technically speaking, the prototype it is prepared to apply to it "BigData" systems, since the repository is a Document Based already used in this kind of systems and the Capture component and Ontology Component could be the MAP and The Alert and log Component could be the Reduce part of the MapReduce process of "BigData" or even, we could plugin it to Electronic Health Records System and monitor these Health Care aspects.

ACKNOWLEDGEMENTS

This work is financed by FEDER funds through the Competitive Factors Operational Program – COMPETE and Portuguese national funds through FCT – Fundação para a ciência e tecnologia in project FCOMP-01-0124-FEDER-022674.

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

REFERENCES

- Efraim Turban, R. S. a. D. D. (2010) *Decision Support and Business Intelligence Systems*. 9 th edn.: Prentice Hall Press Upper Saddle River, NJ, USA.
- Hevner, A. and Chatterjee, S. (2010) 'Design Science Research in Information Systems', *Design Research in Information Systems Integrated Series in Information Systems*: Springer US, pp. 9-22.
- HL7, S. W. g. o. 2010. Composite Security and Privacy - Domain analysis Model Report. May 2010 Ballot ed.: HL7.
- Hofmann, T. (2001) 'Unsupervised learning by probabilistic latent semantic analysis', *Machine Learning*, 42(1), pp. 177-196.
- IRMA (2013) *User-Driven Healthcare: Concepts, Methodologies, Tools, and Applications*. IGI Global.
- ISO 2013. ISO/IEC 27001:2013 Information technology - Security techniques - Information security management systems - Requirements. ISO.
- JN (2013) 'Maus tratos a crianças cada vez mais perversos e difíceis de identificar', *Jornal de Notícias*. Available at: http://www.jn.pt/PaginaInicial/Sociedade/interior.aspx?content_id=3529230&page=2 (Accessed: 23-03-2014).
- Mimno, D., Li, W. and McCallum, A. 'Mixtures of hierarchical topics with pachinko allocation'. 633-640.
- Oladimeji, E. A., Chung, L., Jung, H. T. and Kim, J. 'Managing Security and Privacy in Ubiquitous eHealth Information Interchange'. *ICUIMC '11*. 2011. New York, NY, USA: ACM, 26:1-26:10.
- Pereira, T. and Santos, H. (2013) 'Health Care Critical Knowledge Monitor System Model:Health Care Critical Knowledge Ontology Component'. *Developing a healthier environment under worldwide economical constraints. SHEWC2013 - XIII Safety, Health and Environment World Congress*, Porto, Portugal: COPEC - Science and Education Research Council, 002.
- Pereira, T. and Santos, H. (2015) 'Child Abuse Monitor System Model: A Health Care Critical Knowledge Monitor System', in Giaffreda, R., Vieriu, R.-L., Pasher, E., Bendersky, G., Jara, A.J., Rodrigues, J.J.P.C., Dekel, E. & Mandler, B. (eds.) *Internet of Things. User-Centric IoT Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*: Springer International Publishing, pp. 255-261.
- Pereira, T. R. and Santos, H. (2012) 'Critical Knowledge Monitor System Model: Healthcare Context'. *European Conference on Knowledge Management*, Cartagena, Spain, 5-6 September: Academic Conferences International.
- Pereira, T. R. and Santos, H. (2013b) 'The Matrix of Quality Dimensions of Knowledge Management: Knowledge Management Assessment Models Review', *Knowledge Management: An International Journal*, 12(1), pp. 33-41.
- Saúde, M. d. 2008. Despacho n.º 31292/2008. 236. Diário da república: INCM.
- Steyvers, M. and Griffiths, T. (2007) 'Probabilistic topic models', *Handbook of latent semantic analysis*, 427(7), pp. 424-440.
- WG, H. S. O. 2013. HL7 Version 3 Standard: Security and Privacy Ontology, Release 1. Version 3 ed.: HL7.