

# Healthcare Critical Knowledge Monitor System Model: Healthcare Knowledge Capture Component Specification

**Tiago C Pereira\* and Henrique Santos**

Department of Information System, Portugal

ISSN: 2576-8816



**\*Corresponding author:** Tiago C Pereira,  
Department of Information System,  
Portugal

**Submission:**  October 14, 2019

**Published:**  November 08, 2019

Volume 8 - Issue 2

**How to cite this article:** Tiago C Pereira\*  
and Henrique Santos. Healthcare  
Critical Knowledge Monitor System  
Model: Healthcare Knowledge Capture  
Component Specification. Res Med Eng Sci.  
8(2).RMES.000685.2019.  
DOI: [10.31031/RMES.2019.08.000685](https://doi.org/10.31031/RMES.2019.08.000685)

**Copyright@** Tiago C Pereira, This  
article is distributed under the terms of  
the Creative Commons Attribution 4.0  
International License, which permits  
unrestricted use and redistribution  
provided that the original author and  
source are credited.

## Abstract

The constant evolution of Information and Communication Technologies brings new opportunities with multiple forms of communication, therefore new ways of sharing knowledge. In healthcare practice, being updated it's extremely important in order to provide a better service. The growth of multiple sources of knowledge, mostly supported by technology, gives more opportunity to achieve it. Another way to be updated is sharing knowledge with partners and communities, which is becoming part of the healthcare organizations culture (i.e. Electronic Health Record systems). Healthcare organizations manage with personal information concerning to patients such as clinical treatment, clinical history, drug administration, diseases casuistic, among others; and from many sources, such as patient's feedback, knowledge from suppliers, knowledge from Internet sources, knowledge from decision support systems and inference knowledge (e.g. knowledge from data mining techniques) that are supported by computer-based systems and therefore, demands cautious when are ethical and legal aspects involved. The Critical Knowledge Monitor System Model, proposed, allows knowledge sharing in a controlled ambient and could be a part of the answer to this paradigm that healthcare organizations face. To implement the Critical Knowledge Monitor System model we'll need, to capture knowledge in multiple forms in healthcare organizations, to accomplish it knowledge engineering techniques such as text mining techniques, Information retrieval, ontology construction, among others, should be applied. Since not all knowledge manage by healthcare organizations could be considered critical (or much critical), it's necessary to use specific clinical constructs, such as sensitivity and, we believe, combining it with information security principles CIA and Privacy we could assess documents and classify them as critical to the healthcare organization.

**Keywords:** Healthcare critical knowledge; Knowledge capture; Topic models; Ontology; Knowledge engineering and knowledge security

## Introduction

The recent efforts in healthcare information systems on centralizing health data to bring interoperability between healthcare organizations, such as electronic health record systems had been a breakthrough in Healthcare Organizations. In healthcare practice, being updated it's extremely important in order to provide a better service. The growth of multiple sources of knowledge, mostly supported by technology, gives more opportunity to achieve it. Another way to be updated is sharing knowledge with partners and communities, which is becoming part of the healthcare organizations culture. Sittig et al. [1] refers that there are parallel systems of communications, given the example of the people that carry out discussion by email and others that carry out conversations on wikis [1] and, eventually, other Web2.0 technologies. Another aspect is the use of the existing knowledge wide sources: patient's feedback; knowledge from suppliers; knowledge from Internet sources; knowledge from decision support systems; and inference knowledge (e.g. Knowledge from Data Mining techniques). But, data, information and clinical knowledge managed by healthcare organizations could be considered as critical, consequently it demands cautious when we use computer-based systems to store and retrieve it. Malin et al. [2] explain the risks associated to biomedical databanks and open-access translational research information systems that facilitates knowledge sharing but could allow re-identification of clinic research participants. Sinnot [3] refers that some unencrypted clinical data is shared by CDs through the post and by email as the worst-case scenarios. These aspects could affect ethical and legal issues, such as patient's privacy. Beside the concerns to personal information, healthcare organizations also manage with knowledge about clinical procedures, treatment history, combination of drugs, disease casuistic, clinical

research under development or even knowledge protected by law, which could be considered as critical knowledge. In this research, ontology engineering will be used to support healthcare critical knowledge. The healthcare critical knowledge ontology should be tailored to the healthcare organization in focus to comply with multiple factors, such as: organizational culture, terminology used, health department specifications, among others. Ontologies had been widely used in recent efforts of Knowledge Management Systems (KMS). The ontology will be driven, mostly, from a focus group that will be performed in a public health organization, using Confidentiality, Integrity, availability and Privacy (CIAP) [4] as dimensions and sensitivity of healthcare knowledge as stated in standards such as HL7. Then, we intent to capture knowledge from documents (text, sound, video, images and html) and throughout the use of topic models approach identify the core concepts of the documents-Healthcare Knowledge Capture component. Finally, confront these concepts with the critical knowledge ontology to accurate the sensitivity of the documents and set the value for that document. We proposed this Critical Knowledge Monitor System Model, as part of a solution to the challenges and opportunities [1,5,6] that healthcare organizations face addressing the research question "Can we automatically classify healthcare knowledge as critical concerning organization culture, terminology and sensitivity in order to preserve it?". In section 2, we define what we consider as healthcare critical knowledge, Section 3, we described text mining and how we are using it, Section 4, we describe the knowledge engineering context that is used in the system, Section 5, we described the critical knowledge system model, and particularly the knowledge capture component specification, Section 6, we explain how we are using topic models to extract the core concepts and relations from documents, Section 7, we described how we intent to evaluate the system and all its components, Section 8, we present the preliminary results of the Healthcare Knowledge Capture component, finally, in Section 9, we present conclusions and the next steps that we intent to do.

### Healthcare critical knowledge

There are several perspectives to define knowledge [7]. In the humanistic perspective the knowledge is a conception developed in human's mind. Other authors relate knowledge with two concepts, data and information, defining a hierarchical structure between them. In a concise manner, data is the simplest form, followed by information and, at the top of the hierarchy, we find knowledge. Tuomi apud [7] states that we cannot produce data without the necessary knowledge. More recently, Turban [8] defines data as facts, values and measures or statistical; information as organized or processed data, accurate and available on time; and knowledge as contextualized and relevant information, focused in action. Alavi, based on Polanyi [7], defines knowledge as Tacit or Explicit. Tacit knowledge is the knowledge possessed by humans, not formalized through schemes or written documents and, normally, it is only possible to transfer to another person through learning in work context. Explicit knowledge is the knowledge formalized in books, schemes, reports and other documents. This view is the most common in the knowledge management field. Healthcare

organizations manage with personal information concerning to patients such as clinical treatment, clinical history, drug administration, diseases casuistic, among others; and from many sources, such as patient's feedback, knowledge from suppliers, knowledge from Internet sources, knowledge from decision support systems and inference knowledge (e.g. knowledge from data mining techniques) that are supported by computer-based systems and therefore, demands cautious when are ethical and legal aspects involved. Thus, healthcare knowledge could be considered as critical which justify the use of knowledge management in order to preserve it. Since not all healthcare knowledge could be managed, it's necessary to use specific clinical constructs to classify it. This lead us to assess the sensitivity of healthcare knowledge. By Healthcare Critical Knowledge we consider knowledge that has highly sensitive. According to HL7 [9], sensitivity could be about substance abuse, HIV/AIDS, psychiatry, sexual and domestic violence, vip patient or taboo. We believe that, with the use of the key aspects of Information Security-CIAP and relate them with knowledge sensitivity we can give a value to the clinical knowledge, thus, we could focus in what knowledge we should preserve.

### Text mining

Text Mining (TM) concerns to a set of techniques to extract knowledge from text-based documents or Corpus and therefore apply learning algorithms and Data Mining tasks to them. Some of these techniques came from informational retrieval (IR). We could say that the main difference between DM and TM is the source of Knowledge. There are pre-defined models to extract knowledge from Text Documents, such as location, organization name, addresses, among others; inclusively, we can build an ontology based in the text document knowledge extraction. This is interesting and useful for this research to compare a branch of pre-defined ontology with the driven ontology. Namely, the use of IR and machine learning techniques to obtain knowledge concepts and its relations such as [10,11]: topic models, document similarity, among others.

### Knowledge engineering

Xiaohu [12] had compiled what he considers as principles of Knowledge Engineering (KE) and some of them are valid to KM [12]: Different types of knowledge need different approaches; Different types of experts and expertise need different methods; Different ways to represent knowledge are determinant to acquire, validate and reuse of knowledge; Different uses of knowledge need a Knowledge acquisition oriented to the organization's goals; Efficiency of the knowledge acquisition process needs structured methods; Finally KE is the process of eliciting knowledge to a specific purpose of KMS. Organizing and codifying knowledge is a core function of a KE. The essential here is to preserve knowledge relations with other concepts and get a location where relevant knowledge is. Making a review of the literature, we can find some techniques to organize knowledge [13-20]: Concept Maps, Thesaurus, "Yellow Pages", Ontologies, among others. In this paper we only mention Ontologies because that is the structure that we choose from the KE techniques. The reason of this choice

is that in KMS the use of Ontology is becoming more usual and it's the technique structure that better covers the concepts and its relations. In this paper we only mention Ontologies because that is the structure that we choose from the KE techniques. The reason of this choice is that in KMS the use of Ontology is becoming more usual and it's the technique structure that better covers the concepts and its relations.

**Ontologies**

Ontology [14,15,17] is the definition of the terms and concepts used to describe and represent a specific domain of interest or area of knowledge. Ontology is mostly used to represent knowledge in Knowledge Engineering context. Kapoor and Sharma define the phases of ontology engineering process as [14]: Ontology Scope, Ontology Capture, Ontology Encoding, Ontology Integration, Ontology Evaluation and Ontology Documentation. The methodologies in Ontology Construction are supported by tools. These tools could enable automatic construction [17], manual construction or hybrid methods, which allows to have some automatic or manual phases of ontology construction. Most tools for ontology building, such as editors, mergers and extractors use files to store the ontology although protégé (version 3.4) uses files and Database Management Systems (DBMS). The language used by these tools could be determinant in the type of Ontology Storing because it is based in XML and Resource Description Framework (RDF), such as OWL- Ontology Web Language, that is a standard of W3C. Recently, in 2008, a query language to RDF was defined as W3C standard, named SPARQL. In this research, a hybrid method will be used. The manual part of the ontology construction will be driven from a focus group approach, the automatic part will be the management of this ontology. The language will be OWL. As a standard it could allow more interoperability with other systems. After surveying the existing ontology tools we choose Portégé as the more appropriated for this research.

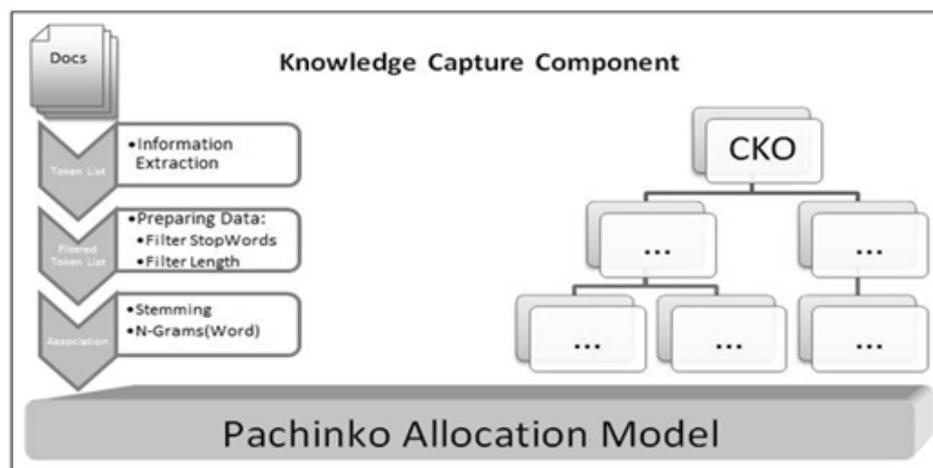
**Critical knowledge monitor system model**

The Critical Knowledge Monitor System (CKMS) [21] should be capable of support healthcare organizational critical knowledge

ontology and repository, monitor the use, access and alterations of the critical knowledge within the organization and by the external entities that the organization has knowledge transfer. In order to achieve this, the system must monitor the stored documents and probably communications based on computing of the organization and alert when critical knowledge is in use. To preserve critical knowledge, we need to develop a knowledge-base or repository that will support the maintenance (use, access and alterations) of the critical knowledge. To develop this component, we need to survey knowledge-based technologies and choose one in a justified manner. There are some requirements that we already know, such as: access control, support for multi-type of content, open source, multiple languages, among others. To achieve prevention of breakthrough of critical knowledge this system model must monitor, in a contextualize form, the organizational critical knowledge transferred in internal and external systems. As internal we include all systems of organization that potentially could use critical knowledge, such as: web pages, document repositories, ERP, CRM, SCM, among others. As external systems we focus on mail servers, and document exchange, or in another words, the systems that provide communication with external entities. These systems deal with different types of content. To capture content to compare with the critical knowledge ontology we need multiple technologies used in many IST areas, such as: Text Mining and Web Mining, Information Retrieval and Knowledge Engineering. We expect that the system could give us the value of the document concerning healthcare critical knowledge. The system is composed by four components: Knowledge Capture Component, Critical Knowledge Ontology component, Critical Knowledge Ontology Repository and Alert and Log Critical Knowledge.

**Knowledge capture component**

The Critical Knowledge Monitor System model needs to capture knowledge from multiple sources (text, sound, video and html). We will use Text mining, Web mining and Information Retrieval techniques in this component. In this section we describe the necessary steps to obtain concepts and some relations from documents, see Figure 1.



**Figure 1:** Knowledge capture component of CKMS Model.

### Extracting text and preparing data

In this step the principle demands are multiple languages (i.e. English and Portuguese) support and multiple sources of documents (i.e. presentations, documents, html pages, sounds, images and video). Using Information Extraction (IE) techniques we could extract a list of tokens from a document. This form is not suitable to apply topic models directly, so preparing data is essential to achieve it. To do so, we will use filters to the tokens to enhance the process and to accurate it, such as: filter Stop Words allow us to clean the tokens like “and, or, punctuation,...”; filter Word Length allow us to clean the small tokens less than i.e. 5 characters. Next, and because clinic terms are very peculiar, we need to aggregate terms that are formed by more than one word, normally called n-gram (Words). In this system number of words could be parameterized.

### Applying topic models

Once we have the ontology built, we will use it to classify documents as critical. In order to achieve it a topic model approach will be applied to identify the topics on documents from the same organization where the ontology will be driven. In a non-supervised form, PAM will be applied to documents, and then using the healthcare critical ontology as a source of concepts/topics and its relevance to healthcare critical knowledge we could classify the topics/documents as critical. We expect to get the relevance of criticality from the perspective of CIAP and set the value of the document to the healthcare organization. To better understanding of this approach we described the two most recent techniques of topic modeling.

**Latent dirichlet allocation (LDA):** LDA is based in hidden variable model that are structure distributions were observable data interact with hidden random variables. Observed data are the words of each document and hidden variables represent the latent topical structure, using a posterior distribution of hidden variables when using observed documents determinates, a hidden topical decomposition of a collection of documents [22]. There are variants of this technique that try to get a hierarchical structure of topics. LDA correlates words in a document but doesn't correlate topics among documents. This issue is addressed by another technique called PAM.

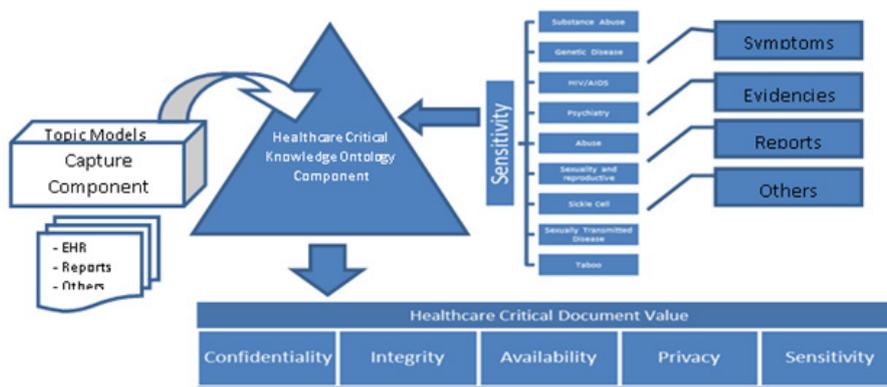
**Pachinko allocation model:** PAM uses directed acyclic graph (DAG) structure to represent and learn topics correlation. It assumes that topics are distributions over the words, but also over other topics. Which leaf node of DAG structure is associated with a word in vocabulary, and each non-leaf node corresponds to a topic, having a distribution over its children (words), at this point there isn't much difference from LDA. Beside this, the topics could have also topics as children, representing the correlation between them.

### Evaluating the system

This model will be evaluated by a prototype (laboratory experiment) - the knowledge capture component is part of it - concerning existing models of KM and KMS Success assessment. Surveying literature we found four models that evaluate the success of these systems, three of which are based in the Delone and McLean Information Systems Success Model, largely used in Information Systems. The other model is based on data warehouses quality, is the most technical and the only one that concerns to ontology quality assessment which is important to KMS based on ontologies, since it is the structure used to categorize, search and retrieve knowledge [23,24].

### Preliminary Results

The Knowledge Capture component is already implemented in Java. We already implement the extraction step and topic models, too. In applying topic models, we implement and tested both techniques LDA and PAM. As stated before, PAM could be more useful because it gives the correlation between topics and it's more approximated to a branch of ontology. We could say that we're using Topic Models in non-supervised way. We tested both languages in focus (English and Portuguese). We also tested various documents (.doc, .pdf, .ppt and html). Using documents about cardiology we found topics strictly directed to the theme. We have partially constructed the Healthcare Critical Ontology and we have already tested the system throughout Capture to Ontology Matching, see Figure 2, and achieve only the concepts and relations from the document that are considered critical to the healthcare organization. So far, the performance of the Knowledge Capture component is good staying under the 30 seconds per document. We considered that the requisites for this component had been achieved.



**Figure 2:** Knowledge capture components and criticle knowledge ontology component interconnection.

## Conclusion and Future Work

The CKMS Model could be applied to existing healthcare systems such as Electronic Health Record Systems allowing the identification, log and alert of knowledge that is critical, based on its sensitivity and CIAP considerations. The capture component, particularly, could be applied to other contexts since it allows capturing from many sources transforming text into concept and relations to match with ontologies from different domains. The next step will be the conclusion of the construction of the healthcare critical knowledge ontology using HL7-Security Work Group - security and privacy ontology (Draft) and beside Physical Abuse that we already have mapped, defining the remain characteristics (symptoms, evidences, reports and others) of Substance Abuse, Psychiatry, Genetic Disease, HIV/AIDS, Sickle Cell, Sexuality and reproduction, Sexually transmitted Diseases and Taboo concerning healthcare sensitivity throughout a focus group to be performed in a public healthcare organization. We believe that the target of this focus group would be medical and administrative personnel. Medical personnel since the CKMS could allow medical confidentiality preservation and administrative personnel because of the knowledge and information exchange between healthcare organizations.

## Acknowledgment

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.

## References

- Sittig DF, Wright A, Simonaitis L, Carpenter JD, Allen GO, et al. (2010) The state of the art in clinical knowledge management: An inventory of tools and techniques. *International Journal of Medical Informatics* 79(1): 44-57.
- Malin B, Karp D, Scheuermann RH (2010) Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. *J Investig Med* 58(1): 11-8.
- Sinnott RO, Stell AJ (2009) Data privacy by design: digital infrastructures for clinical collaborations-Enlighten, USA.
- ISO, "ISO 7498-2: (1989) Information processing systems-open systems interconnection-basic reference model-part 2: security architecture.
- Bali RK, Dwivedi A, Naguib R (2005) *Clinical knowledge management: opportunities and challenges*: IDEA Group Publishing, USA.
- Prokosch HU, Ganslandt T (2009) Perspectives for medical informatics. Reusing the electronic medical record for clinical research. *Methods Inf Med* 48(1): 38-44.
- Alavi M, Leidner DE (2001) Review: knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly* 25(1): 107-136.
- (2010) Efraim turban, decision support and business intelligence systems. In: (9<sup>th</sup> edn): Prentice Hall Press Upper Saddle River, NJ, USA.
- SWgo HL7 (2010) Composite security and privacy-domain analysis model report, HL7.
- W Li, Callum A (2006) Pachinko allocation: DAG-structured mixture models of topic correlations," in Proceedings of the 23<sup>rd</sup> international conference on Machine learning, Pittsburgh, Pennsylvania, USA, pp. 577-584.
- Blei DM, Lafferty JD (2006) Dynamic topic models, in Proceedings of the 23<sup>rd</sup> international conference on machine learning, Pittsburgh, Pennsylvania, USA, pp. 113-120.
- Xiaohui Y (2009) To facilitate knowledge management using basic principles of knowledge engineering, pp 94-97.
- Hellström T (2004) Mapping knowledge and intellectual capital in academic environments: A focus group study," Emerald Group Publishing Limited, UK.
- Kapoor B, Sharma S (2010) A comparative study ontology building tools for semantic web applications. *International journal of Web & Semantic Technology* 1(3) 1-13.
- Wang G (2010) Methodology research of ontology building in semantic web. *Computer and Information Science* 3(4).
- Guoqiang SJ, Qiang W, Qi Liu (2010) REA-based enterprise business domain ontology construction. *Journal of Software* 5(5): 552-529.
- Park J, Cho W, Rho S (2010) Evaluating ontology extraction tools using a comprehensive evaluation framework. *Data Knowl Eng* 69(10): 1043-1061.
- Li Z, Sun Y (2009) Ontology based knowledge management in intelligent tutoring systems, pp. 1-4.
- Irani Z, Sharif AM, Peter ED (2009) Mapping knowledge management and organizational learning in support of organizational memory. *International Journal of Production Economics* 122(1): 200-215.
- Cicortas A, Jordan VS, Fortis AE (2012) Considerations on construction ontologies, USA.
- Pereira TR, Santos H (2012) Critical knowledge monitor system model: Heath care Context, USA.
- Blei DM, Lafferty JD (2009) Topic models text mining: classification, clustering, and applications, pp. 71-94.
- Stojanovic N, Stojanovic L (2002) Usage-oriented evolution of ontology-based knowledge management systems," on the move to meaningful internet systems. OTM Confederated International Conferences "On the Move to Meaningful Internet Systems, pp. 1186-1204.
- Pereira TR, Santos H (2013) The matrix of quality dimensions of knowledge management: Knowledge management assessment models review," knowledge management. *An International Journal* 12(1): 33-41.

For possible submissions Click below:

[Submit Article](#)