

ONTOLOGY TOOL FOR KNOWLEDGE ACQUISITION IN A VIRTUALISED ICT INFRASTRUCTURE

Florin POSTOLACHE¹

¹"Mircea cel Batran" Naval Academy, florin.postolache@anmb.ro

Abstract: Nowadays, physical (hardware) environments virtualization has expanded rapidly, becoming an absolute necessity for IT infrastructure reconfiguration. The system virtualization and cloud computing are highly debated topics in the last 10 years which lead towards numerous strategies, sometimes quite different. IT Virtualization increases the levels of resources utilization. Due to load balancing of the available physical resources, by virtualization we increase their level of usage, without a proportional growth of costs, as in a normal extension of infrastructure. The research necessity is imposed by the rapid evolution of the information technology, which leads to finding new ways of organizing the hardware infrastructure, applications, as well as new ways of solving the operations, regardless their complexity. A big part of the infrastructure can become partially or totally virtualized and thus, the processes gain a dynamic and distributed character, while the static and hierarchical structures become more and more adaptable and flexible. Thus, for knowledge acquisition and structuring, we consider necessary the development of a peculiar ontology involving a common set of constraints and a conceptual environment, having as a main goal the relations between the information classes and modeling of knowledge. Having established the goal, i.e. defining an ontology of the IT infrastructure virtualization field, we will suggest a methodology necessary to classifying the virtualized systems, offering to the specialists' community a new model structured through the medium of the included layers, which allows capturing the relations between entities / services included in a layer, as well as the relations between the components of different layers. The virtualized IT environment, structured according to the proposed approach also allows us to deepen knowledge and understand the virtualization domain, its components and their relationships.

Keywords: Ontology, virtualised IT infrastructure, knowledge acquisition

Introduction

In a virtualised IT systems, knowledge acquisition is a systematic process, firstly involving the presentation of processes, procedures and steps that occur throughout a proposed acquisition of knowledge.

However, a good knowledge of all system peculiarities represent a decisive factor on the successful acquisition of knowledge for fault diagnosis, where knowledge acquisition regarding the fault reaction of the target system is required (faults / symptoms and manifestations, fault's granularity and relations between them in various contexts running).

Thus, a common set of constraints and a joint conceptual framework for relations between the classes of information and modeling knowledge are decisive in the ontology development, the adopted methodology being through the layers that make up the model, the adopted methodology being through the layers that make up the model since is possible to know the relationships between entities / services that are part of a layer as well as the relationships between the components of different layers.

The aim of this paper is to highlight some challenges regarding the method of

representation and structuring the layers of the proposed model within the ontology and how they approach IT field targeted.

The proposed approach

We outline the proposed ontology around the concept of stack consisting of layers (Figure 1). We recall here that the layers grouped within the abstracted infrastructure have been made according to the environment in which they are placed and because of virtualisation, each layer further comprise one or more services with equivalent levels of abstraction.

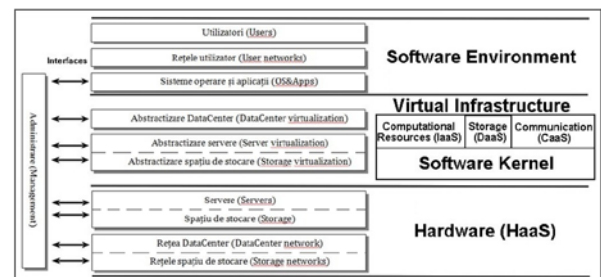


Figure 1. Virtualised IT infrastructure

Ontology role is to capture knowledge in a particular area of interest and also, how entities can be grouped, ranked, or divided according to their similarities and differences. Because some

languages provide various facilities, we use the latest languages for ontology development, especially Tropes, a software designed for Semantic Search Engine, Text Analysis & Semantics developed by Semantic-Knowledge¹.

Ontology Research Tool

Designed for Information Science and Scientific studies, using semantic classifications, and reliable analysis techniques, Tropes - a Natural Language Processing and Semantic Classification software - guarantees pertinence and quality in Text Analysis.

We considered, in the ontology development, the current state of present development tools and standards that allow for different ontologies to be imported and placed to work together easily and also, the stand-alone systems with certain specific characteristics. This remove inconveniences related to ontology designs which gives a general application, so they can be used widely and be easily shared.

Analysis of written or spoken texts requires answers from certain questions concerning certain objectives. Moreover, having the skill and opportunity to reuse and re-combine ontologies easier if they have certain characteristics and meet certain principles, a written text analysis method go beyond in the development or extension of the ontology because text is reduced as far as possible to their essentials.

Extraction of Relevant Information, Qualitative Analysis and Categorization, Chronological Analysis and finally the Text Contraction (Summarization) represents the important steps which immediately detect contexts, isolate themes and identify principal actors in order to place entities in context and rapidly compare it with other and where the principal episodes can be isolated and the discussion blocks visualized.

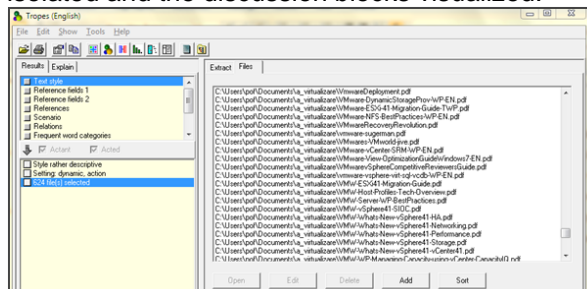


Figure 2. Loading documents

Based on intended analysis strategy, to set up ontologies (personalized classifications), modify or restructure the equivalent classes, customize information retrieval and define an analysis grid - aiming at automatic generation of an analysis

report, or statistics in order to export results - require a specific classifications adapted to analysis objectives.

Tropes software GUI, an intelligent manager based on Semantic Networks and Natural Language Text Analysis technologies, highlights all hypothetical interpretations through proposed hypertext graphs and automatically sorted keyword lists.

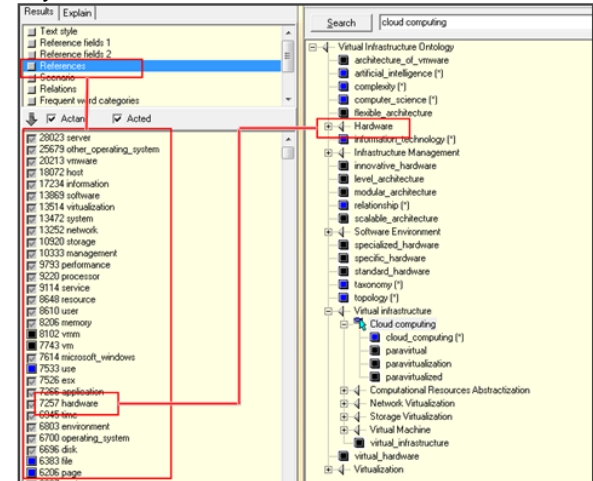


Figure 3. Scenario design

In a context analysis, a set of techniques applied for a given text highlights the main actors, the relationship, hierarchy and the mutual relations between actors (Figure 3).

The Results

The proposed ontology relates to hierarchical structuring of knowledge through the category's layers according to the essential qualities (relevant and / or cognitive) and also by developing knowledge structures that allow to different IT systems to understand and use each other's content.

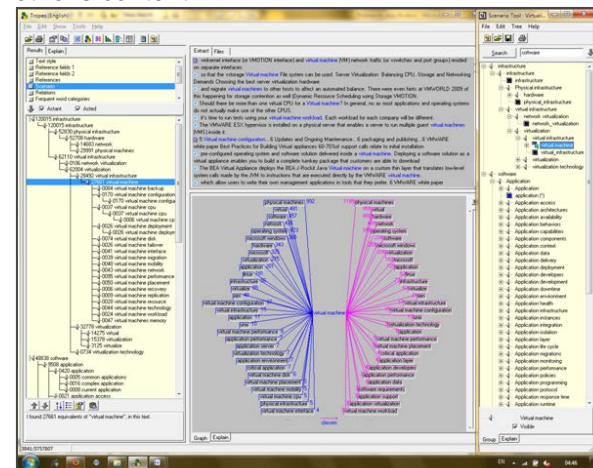


Figure 4. Ontology graphical interface

For the field of interest, ontology creation are based on the following main components:

¹ <http://www.semantic-knowledge.com/>

- **Classes:** things in a field such as physical entities, persons, documents, organizations, activities, ideas, locations, etc.
- **Attributes:** properties of classes such as color, size, age, storage capacity, number of components, difficulty, etc.
- **Values:** the specific properties of a thing that distinguishes it from other things, such as black, white, easy, heavy, long, difficult, etc.
- **Relationships:** how classes are associated with each other.

Within the undertaken research, the structure is determined by the manner and purpose in which knowledge must be represented both to users and to other IT systems. We take into account two basic ways to knowledge representation: one based on relationships, the other based on attributes and values. The first can be viewed as a network of concepts one related to another, each link representing a connection. This is the basic format of a scheme (concept map). The other, based on attributes and values, is viewed as a frame (frame). Ontology scenario includes a number of Semantic Groups that can be arranged in a hierarchy up to 9 levels, i.e. groups of words and / or equivalent classes.

Terminology extraction

A rapid development of Scenarios is performed using terminology extraction that highlights expressions and combinations thereof from context, suggesting also a list of references to complete the Scenario.

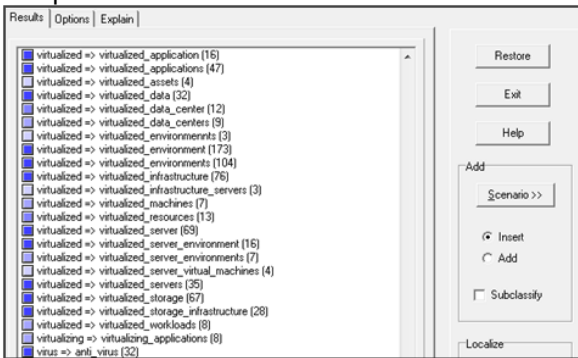


Figure 5. Terminology extraction

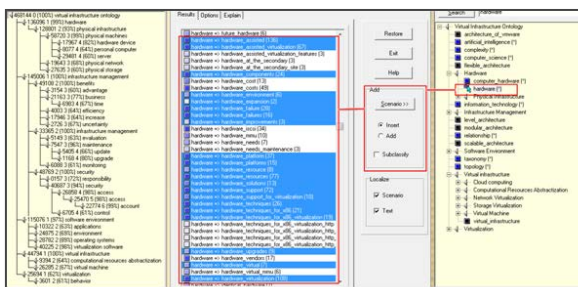


Figure 6. Expressions selection for Scenario

Thus, for the proposed approach, the extraction terminology tool (Figure 5) identify the most significant expressions and combinations thereof, thus obtaining a more precise classification concomitantly with a rapid enrichment of the scenario proposed by grouping acronyms and phrases associated.

The terms or expressions resulting from extraction are preceded, depending on how much are used in the context, by a color code, which varies from darker for the most used to a shade open for the insignificant terms or expressions.

Classes and Relations representation

For relations representation are available four types of graphs: [Area], [Star], [Distribution] or [Episodes] and [Actors].

The reference graphs (area and star) show the Relations between the Equivalent classes. They are oriented: the classes on the left of the central class are its predecessors, those on the right its successors. Distribution graph displays the distribution of equivalent class or relations between equivalent classes. The episodes graph displays the episodes in chronological order.

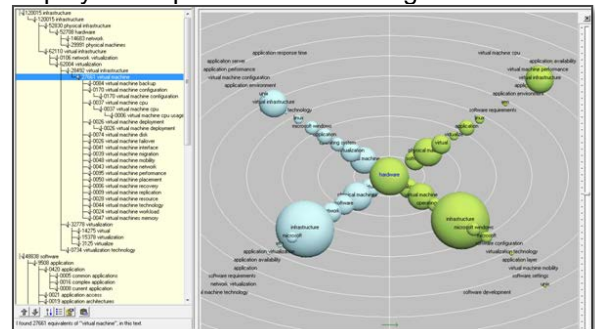


Figure 7. Area graph

The Area graph (Figure 7) represents the equivalent classes as spheres, their surface is proportional to the relevance of the entity (the number of words it contains). When two classes are close together, they share many Relations, and when they are far from one another, they share few Relations. The number of Relations connecting the central class and the other classes is proportional to the distance between them.

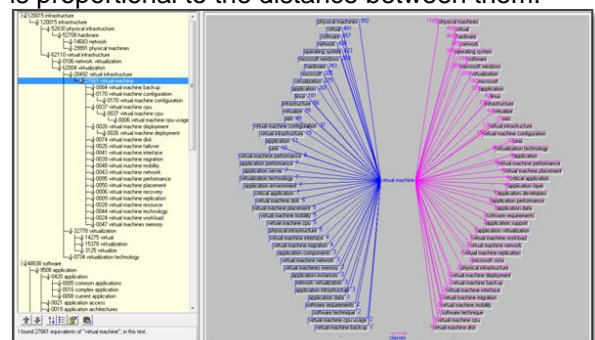


Figure 8. Star graph

For this purpose we considered a systematic approach. In other words, we must have the right methodologies in order to follow and use the frameworks. These helped us to focus us on the knowledge that we need and how to get them in the most efficient way.

Whenever was possible we reuse the knowledge from previous projects or from generic models of knowledge such as generic taxonomies (practice and science of classification). Due to this fact, we do not started with an empty knowledge base, thus using, combining and adapting the existing "skeletons" 've developed a new structure that has provided us guidance on sorting the captured knowledge. For this purpose we used a specific software which relieved the effort, giving us a set of tools and techniques that we can use when we acquire, analyze and shape knowledge.

Contributions to knowledge acquisition refer to:

1. An appropriate methodology concerning the classification of virtualized ICT infrastructure around the concept of grouping layers that compose the model (Figure 1). The proposed ontology allowing to capture the relationships between layer entities / services and also the relationships between components of different layers, making an easier introduction to understanding the fields, its components and their relationships.

2. The structuring of IT virtualization field regarding the acquisition of knowledge.

Structuring the interested field obtained the advantage that each agent (human or intelligent) specialized in a particular layer will identify and insert on the correct layer the specific case, allowing a better location, a precisely level of discrimination components by default a small granularity, in such cases the diagnosis being made in full details, that enabling an intervention in troubleshooting.

3. Devising an ontology for our field of interest thereby contributing to:

- Sharing a common vocabulary and a better understanding of the domain.
- Reuse of domain knowledge for further development.
- Formulating the assumptions in an explicit way regarding the field of interest.
- Separation of operational knowledge from domain knowledge.
- A deeper analysis of the domain of interest.
- Usage of knowledge base to share knowledge with other computer systems or as part from developing an intelligent computational system.

BIBLIOGRAPHY

- [1] Y. Kodratoff, Comparing Machine Learning and Knowledge Discovery in DataBase, July 1999.
- [2] Y. Kodratoff, Knowledge Discovery in Texts: a definition and applications, in Foundation of Intelligent Systems, Ras & Skowron (Eds.) LNAI1609, Springer, 1999.
- [3] N. Langlie, Educational Technology Leaders: Competencies For A Conceptual Age, Dissertation Presented in Partial Fulfillment Of the Requirements for the Degree Doctor of Philosophy, Capella University, 2008.
- [4] J. Lyons (*), Semantics, Cambridge, Cambridge University Press, 1993.
- [5] J. Vander Putten & A. L. Nolen, Qualitative Data Analysis: Comparing Results From Constant Comparative and Computer Software Methods, Conference of the Association for the Study of Higher Education, November 5-8, 2008 in Jacksonville
- [6] Florin POSTOLACHE, ACHIZIȚIA ȘI STRUCTURAREA CUNOȘTINȚELOR ÎN SISTEMELE IT VIRTUALIZATE, Editura Zigotto, 2012, ISBN 978-606-669-005-8;
- [7] <http://www.semantic-knowledge.com/index.htm>