A systematic approach to evaluate enterprise ontologies using testing techniques of software usability

Alessandro Viola Pizzoleto¹, and Hilda Carvalho de Oliveira²

¹Department of Computer Science Federal University of São Carlos - UFSCar São Carlos, SP, Brazil
²Dep. of Statistics, Applied Mathematics and Computer Science
Univ Estadual Paulista - UNESP Rio Claro, SP, Brazil
alessandropizzoleto@gmail.com, hildarc@unesp.br

Abstract—this paper introduces a systematic for evaluating the enterprise ontologies using software usability testing techniques. The objective is to support developers of enterprise ontologies to verify if the ontology that has been developed is easy and simple to use and navigate as well as if it eases the content addressed understanding. Inconsistencies in the content can also be found. The systematic adapts software usability testing techniques with the users’ participation in an organized and objective way. A case study is presented for the evaluation of an enterprise ontology considering the processes of the software quality model MPS-SW. The results contributed for improvements in the ontology, because it was possible to correct inconsistencies and problems in the organization and relationships of the content.

Keywords: Enterprise ontology; Usability; Software processes; Ontology evaluation; MR-MPS-SW

1. Introduction

Researches about ontologies have been significant over the years, including many types and their applications. In general, ontologies represent knowledge between different systems and/or components, improving the collaborative use of such content. They are semantic models to capture and represent aspects of a real domain. This work was directed to a type of ontology known as Enterprise Ontology, describing concepts and relationships that exist in an enterprise domain.

An Enterprise Ontology is a concept that is defined for the Enterprise Project [1], from the inclusion of new concepts from project TOVE [1]. This project aims at adapt the methods of ontology modeling for business and change management, providing a common vocabulary in the organization. It facilitates the development of systems that handle the organization’s knowledge, promoting integration between tools that manipulate knowledge related to ontology, through the databases sharing created from its ontological structure.

According to Uschold e King [1], in order to build an ontology enterprise there are four stages to follow: (1) identification of the proposal of the ontology to determine the formality level of the ontology description; (2) construction of ontology, capturing and coding knowledge, including integration with other ontologies; (3) ontology evaluation throughout the process; (4) formal documentation (definition of constants, predicates and axioms), reviewing the stages of identification the scope and formalization.

According to Blomqvist [2], an enterprise ontology can be built following five stages: (1) requirements analysis, considering scope and use cases; (2) iterative construction, with middle-out approach to covering the requirements specifications; (3) implementation with appropriate tool; (4) evaluation of the clarity, consistency and usability and (5) maintenance. For the author, the construction of an enterprise ontology can be manual or automatic.

In this context, this paper covers step (3) of [1] and step (4) of [2], proposing a systematic to evaluate the clarity and usability of an ontology with human participation.

For that, the concept of Rubin’s evaluation tests [3] was considered, which could be applied to any product development cycle. According to the author, evaluation tests verify that the conceptual models have been implemented properly if the user can develop real tasks, identifying the specific deficiencies of usability. In this type of test, the user navigates between the screens following a specific sequence. The data collected by the evaluator must be analyzed, in order to generate improvements in the system specifications [4].

Thus, Section 2 presents a short literature review about evaluation of ontologies. Section 3 presents the systematic proposed to evaluate enterprise ontologies, composed to six steps for evaluation. A case study is presented in Section 4, considering a business ontology about the levels G and F from the software quality model MPS-SW. The conclusions about using the systematic approach are presented in Section 5.

2. Literature reviews

In the last decade, many research related to the evaluation of ontologies have been found in the literature. Some authors proposed a systematic of generic evaluation [5] [6] [8], evaluation focusing on reuse [9] and others defined criteria and evaluation metrics [10] [8], standards-based methods
have arisen recently [11]. However, the same approach of usability tests proposed in this paper was not found.

Felix, Taofiki e Adetokunbo [12] proposed a spectral analysis of the graph generated by the ontology information. The authors changed the relationship of classes in a tree structure where nodes and edges are used to represent the concepts (classes, objects and entities) and the relationship between them. The structure and the structural dimension demonstrate the complexity and help rapid recovery concepts. This proposal requires an ontologies’ documentation, to simplify the tests and which results prove the veracity of knowledge contained in the ontology.

Poveda-Vilalaso [13] presents the "OOPS!" tool, with the purpose of detecting flaws in ontologies. This tool identifies failures caused by inexperienced developers in building ontologies. "OOPS!" may be used in any domain ontologies implemented in any language. This tool can sort the flaws in three levels of severity: critical, important and less relevant [13]. The tool is useful for validation process of the ontology. However, it analyzes and identifies only predefined faults in a pre-existing configuration. Currently, the tool has 50 preconfigured failures.

Ma et al. [14] propose an approach that examines the ontology in search of inconsistencies by using measurement of axioms and relevant subsets. For each type of inconsistency, a weight for the calculation of the impact value of each type is assigned. Some algorithms for the calculation of the impact value of each inconsistency type are proposed. The aim is to identify and classify the axioms and subassemblies more susceptible to give rise to inconsistencies in the ontology, as well as the concepts that must be modified [15].

Brank, Grobelnik e Mladenić [15] analyzed some types of systematic assessment and defined four evaluation approaches, as well as five levels of expertise about them, as shown in Table I. The differences of the levels adopted by the researchers were treated.

Levels from Table I are explained below, (a) to (d). (a) Lexical, vocabulary, concepts and data: the assessment verify the concepts included in the ontology and the vocabulary used to represent and identify these concepts. Documents related to the problem domain processed with the ontology are used to assist the evaluation; (b) Hierarchy and taxonomy: the relationships between the concepts are validated, especially if the relationship is one-to-many; (c) Other semantic relations: all relations between the concepts that are not "is-a" are evaluated separately; (d) Context, application: whereas an ontology can be part of a set of ontologies or part of a software, the tests should be run on each usage environment; (e) Syntactic level: manually created ontologies are validated, whereas the language used for description of the ontology and its documentation; (f) Structure, architecture, design: predefined criteria are required to analyze the ontology structure. This evaluation is usually made manually.

The approaches to evaluate ontologies presented in Table I are presented below from (1) to (4). (1) Golden standard: compares the syntax used in the ontology definition with the syntax of the language with which it was written. (2) Application-based: this approach considers that there is an integrated application to ontology. Thus, an ontology can be classified as efficient or not, depending on the part that the application uses. The ontology is evaluated by the results of the application, requesting documentation with the possible results. (3) Data-Driven: compares the information contained in the ontology with existing data (usually a set of textual documents) in ontology domain. (4) Humans: the evaluation is performed by humans, evaluating whether the ontology introduces a predefined set of criteria, standards and requirements related to your domain.

Note that the approach (4) Humans is the only one that can cover all levels of evaluation of an ontology. The systematic proposal in Section III follows this direction, to allow the evaluation of a business ontology at all levels, from (a) to (f). That is because the evaluator has to decide which tests levels to cover and which profiles the participants will have.

3. Systematic proposal

According to Rubin [3], usability corresponds to the quality degree of the interface interaction with users. The quality is linked to the principles: a) ease learning; b) ease of task memorization in case of intermittent use; c) user productivity on tasks execution; d) prevention, targeting the reduction of errors from users; e) user satisfaction. Tests for evaluation of the degree of usability of a product are usually used with the participation of potential users.

Thus, this section presents the systematic proposal for running usability tests with users to evaluate enterprise ontologies. The systematic consists of six steps and presented in subsections from A to F.

Note that the ontology evaluation is generally performed through a software tool such as Protégé, for instance. In this case, it is important to be clear to the evaluator and the participants that the tool is not being evaluated. If the participant fails to meet the tool used, the evaluator should grant some minutes to present the main tool features
to the end. This time is not considered in the tests. It is important not to lose focus in the definition of the tests: evaluate the structure, navigability, completeness and depth of knowledge contained in the ontology and as the user identifies information.

3.1 Definition of participants’ profiles

Efficient tests require that all levels mentioned in Table I are evaluated. Thus, participants with good knowledge about the domain of ontology must be selected. However, it is important that participants with different knowledge levels and experience are also selected. It is recommended to classify the degree of knowledge such as beginner, intermediate, and expert. Roles can be associated with each class of knowledge degree. The amount of profiles and participants in each knowledge degree depends upon the depth you want to achieve with the test run. However, it is not recommended a greater number than fifteen participants nor less than five, for reasons of results and sustainability.

3.2 Preparation of documents to guide the application of usability testing

Seven documents should be made to guide the application of the tests in enterprises ontologies:

(1) Test plan: contains information regarding the purpose of the test, the statement of the problem to be tested, the methodology to guide the test run, the tasks and responsibilities of the evaluator, the metrics of evaluation and preparation of the environment for the participants;

(2) Roadmap of the evaluator: contains information to guide the evaluator or evaluation team, during the tests running:-the objectives; -the environment and the equipment; -the evaluation rules; -ontology functionality; -protocols and procedure; -forms used in the test;

(3) Orientation script for the participant: contains information that must be submitted for each participant before starting the tests: the test objectives and how it will be conducted. The intention is to help the participant on the difficulties encountered in operating the software tool of ontologies used, difficulties in understanding the information contained in the ontology, etc.:

(4) Task list: it is composed by different tasks to be performed by the participants, considering increasing levels of difficulty. A certain degree of difficulty must be associated with each task. The tasks must be defined with the participant’s intention to cover all points of enterprise ontology are verified. It is recommended that the tasks be defined based on the ten heuristics of Nielsen for software systems [16]. This is an unusual use form of these heuristics. Nevertheless, they deal with important aspects to be evaluated in a software and that can be evaluated on business ontologies. The task list should be split into two documents: one to be delivered to the participant that has only the tasks; another for the evaluator, with the tasks and information on implementation and results to be obtained, to serve as a guide to the monitoring of the tasks implementation;

(5) Form for data collection: document used by the assessor for explanations relating to the execution of each task by the participant, including their attitudes during the test. This document can be implemented through a text field under each task from the task list; in this case, a copy of the task list with the fields for notes should be used for each participant;

(6) Evaluation questionnaire: contains questions to be answered by each participant at the end of the test. The goal is to evaluate the perception of the participants in relation to the difficulties they found, their expectations and suggestions for improvements in the ontology. Examples of information required from participant: level of ease to find information; difficulties found during navigation by ontology; positive and negative aspects regarding the use of ontology, as to the information contained in it and about the organization of information; ontology development failures;

(7) Terms of image use: document that must be signed by the participant to allow the use of his image, in order to confirm the test. The participant does not need to show his face.

3.3 Contact with participants and implementation of usability testing

Whereas the evaluator defined the participants profiles in step 1 (see subsection A), candidates must be contacted. This is a process that must provide for refusal of some candidates due to the lack of time. However, the evaluator should emphasize that the test should spend approximately one hour only and the schedule can be adjusted.

The evaluator can act in two ways for carrying out the tests: go to the location where is the end or invite the participant to come to a location prepared specifically for the test. The necessary infrastructure must be provided for both. However, in the second situation, the laboratory environment should be pleasing to the end.

It is very important to inform the participant that he is not being evaluated, but the ontology. He must be aware that the ontology must satisfy their needs otherwise it should be adjusted. The participant must also receive a short guidance on how to use the ontology presentation system.

Specific software tools for the execution and usability testing analyze are recommended. These tools allow recording all the participant’s actions to performing the tasks. Some tools allow the evaluator to make notes about participant’s observations in real time. It is important that the tool support the evaluator during the tabulation process the results.

It is interesting that the evaluator present a possible usage scenario for each task, as well as the participant understands the purpose of that task execution. The intention is that the participant imagines a real situation of your day-to-day life, motivating him to the task.
It should be noted that all participants must perform the same tests and the conduct of the evaluator must be the same in all tests. All supporting documents for the tests should have been adequately provided.

3.4 Analysis of usability testing

After the fourth stage of the tests, the data recorded by the support tool and documented notes must be analyzed. Overall, some examples of facts that can be verified: - Number of times the mouse was used for a participant to complete a task; - Sequence of actions for the task (may be different from that provided by the evaluator sequence); - Participant’s behavior through their facial expressions and body during the execution of the task; - Important observations made by participants during the test; - Difficulties found by the participant to complete the task; - Unfinished tasks and the reasons for the participant not finish them.

It is important to consider the time used by each participant to perform each task. This indicator helps the developer to make adjustments in the ontology in order to reduce the time spent and the paths to facilitate obtaining the desired results.

The evaluator should compare the features of ontology that the participant used to complete the task with the resources provided. If the resources used were different, it is appropriate to examine whether the results were also satisfactory for the participant.

When there is an occurrence of several errors in the execution of a task, this can be a sign that the ontology is not self-explanatory, requiring adjustments. Generally, the occurrence of many errors is related to the lack of knowledge about the concepts represented in the ontology - which impairs its use.

Usability testing is also used to verify the level of acceptance of the ontology by the participants. The evaluation questionnaire filled out by participants support this type of evaluation. The observations and evaluator’s notes help to identify gaps and necessary improvements that were not observed in other analyzes.

3.5 Preparation of the recommendations

Following the analysis step, it is important to develop a list of the ontology strengths as well as a guide of recommendations for improvement of weaknesses. This guide may contain inferred guidelines between the group of participants and the evaluator, together or separately. These guidelines must be clear and objective to ease the implementation.

The ten Nielsen heuristics can help to organize the recommendations. While this form of heuristics use is not usual, it helps to support the view for user needs to access to the ontology information [16]. All the participants signed the "terms of use".

4. Case study

This section introduces the systematic enforcement proposal in section III, whereas a business ontology on the levels G and F software quality model MPS-SW [17]. The version of the ontology considered for this work was the v. 1.1 not yet tested with users. The systematic application is presented in subsections A to E, where each section corresponds to a step, with the exception of the steps 3 and 4, which are in the same subsection.

![Part of the ontology in OWL](image)

The MPS-SW model has well defined rules for people involved in the process, with the following responsibilities: (1) execution of the process; (2) monitoring of the evaluation process; (3) auditor ship (if the process was executed properly and meet the objectives); (4) validation of the processes (if the case meets the criteria of the company’s internal policy).

The MPS-SW model is part of the Brazilian Software process improvement (MPS.BR) based on CMMI (Capability Maturity Model Integration) and ISO/IEC 12207 and 15504, 20000. This model includes internationally recognized practices to the implementation and evaluation of processes associated with the development of software. The idea is to meet the business needs of the software industry. The model defines seven levels of processes: G to A (highest level). G level processes refer to: Project Management and Requirements management. F level processes are: Measurement, Acquisition, Configuration Management e Quality Assurance. Each level adds processes to lower levels. The whole process is composed of attributes, and results of these attributes (RAP) well defined, which must be documented. The effective implementation of the processes must meet the expected results, which are evidenced by a work product or a significant change in the process state. The certification can be granted from the G level, including.
The proposed ontology by Pizzoleto [17] aims at offering an alternative way to organize the levels G and F of the MPS-SW model. The intention is to support the understanding of several guides for implementation and evaluation model. The ontology adds terms and explanations of the Project Management Body of Knowledge (PMBOK) and indicators according to the perspectives of the Balanced Scorecard (BSC) model among other indicators of processes. It adds also information provided by experts [17]. Figure 1 presents a part of ontology in OWL (Web Ontology Language).

4.1 Participant profile

The definition of the test participants’ profiles considered different knowledge levels about the MPS-SW model in theory and practice. Knowledge about the implementation and evaluation model was also considered. The Knowledge levels were defined: (1) People with low model knowledge (Beginners); (2) Project and/or quality managers, with some experience in the model; (3) Consultant and/or evaluator of model (high experience in the model). Table 2 shows the definitions for the quantity and functions of the participants as to the experience.

The "Profile Questionnaire" was composed of questions that approached the items proposed on the systematic: educational background, work experience and experience with the MPS-SW model. This information helped in data analysis.

4.2 Documents for the testing application

All seven documents proposed in step 2 were systematically developed. They have been key so that questions during the tests were objectively answered. These documents had the objective to maintain uniformity of treatment for all participants.

The “Task List” was developed with the amount of fourteen tasks, with three difficulty levels: low, medium and high. The sequence of execution was: three tasks with low difficulty, four tasks with medium difficulty level and seven tasks with a high difficulty level. The tasks addressed the following subjects: - Ontology classification; - Composition of expected results (Planning and schedule project, Document requirements); - Information dealt strategic moments of the MPS-SW model (Milestone review, Project estimation). It must be observed that the definition of tasks considered aspects approached in the ten heuristics of Nielsen. Table II presents a table used for this purpose. Notes were also made in the “Form of data collection”.

4.3 Contact with participants and test execution

Nine participants accepted the invitation to schedule the test, four beginner participants in the study of MPS-SW model, two participants with the Project Manager role, one as Quality Manager and two as evaluators/consultants of MPS-SW. It is important to notice that the amount planned has been amended in practice according to the participants’ availability. Another important element for the invitation acceptance by these people was to inform the estimated time for the tests. Some people could not participate due to schedule and time problems. Next, the defined date was important to remember them by e-mail and phone call.

Two participants were from different cities of the evaluator, distant of about 500 km. However, for all participants, the location selected to do the tests was their own workplace, considering geographical distance problems, traveling time and daily activities of participants. For this, a structure of a mobile laboratory was mounted: two laptops with screens of 15.6" and with the software Morae TechSmith Recorder and Observer (version 3.3.2) in each one. The computers were interconnected through network with crossover connection. Thus, the Morae TechSmith Observer software was used for notes and comments on what the participant was doing.

On the user’s computer was installed Protégé system and version 1.1 of ontology, considered alpha version by the author [17]. The concern with the Protégé system interface influence in the tests was mitigated, teaching previously the participant to use the interface before the execution of the tasks (about five minutes for this). It must be observed that Garcia [18] mentions that even users without experience in ontology editors can perform tasks without too much difficulty (no drastic errors).

The tests were conducted for approximately 45 minutes in the presence only of the participant and the evaluator on the place. åJFÅJGåJG illustrates two participants filling the "Profile Questionnaire". All the tests were performed according to the recommendations in a uniform manner and without the need for reviewer interference with the execution of tasks.
of the task. The evaluator made a scenario to the participant before the execution of each task.

### 4.4 Tests Analysis

All collected data with Morae Observer system during the tests were imported into Morae TechSmith Manager software (Figure 3). This system provides several types of analyzes, including the examples mentioned earlier in the subsection 3.4. All events that occurred during the execution of the tests can be analyzed with the support of graphs generated by the system. During the tests, videos with the movement made on the screen and the participant’s picture on the corner of the screen were recorded. Sound recordings/voice were also made. This recorded material can be accessed, allowing insertion of comments.

All participants were able to complete all fourteen tasks. Some participants asked for help because they were unable to identify some terms used in the ontology. This happened most often with "beginners" participants due to limited experience with the MPS-SW model. The other participants made a few questions. Fig. 4 shows the average help requests from all participants for each of the fourteen tasks (x-axis: from 1 to 14). Considering graphs as in Figure 4 and the notes taken by the evaluator, it was possible to verify that the great majority of requests for assistance was made when the participant was in a state of navigation and not knew where to go. While the participant was performing the task and went through the tree superclasses and subclasses, the navigation mechanism became clearer as well as the associated information. Thus, requests for assistance have been decreasing with the continuity of tasks.

### 4.5 Preparation of recommendations

The beta version of the ontology developed was adequate to meet the original purpose. However, some recommendations for changes were presented by the participants and by the evaluator. These recommendations can be summarized in the following requirements: (1) creation of new connection properties between superclasses and subclasses; (2) adjustments on the links between subclasses because chaining failures were identified; (3) identification of parallel flows to facilitate users to access the intended purpose; (4) improvements in the layout of classes; (5) simplification of terms used in the ontology.

All five recommendations are effected with the definition of a set of thirteen tasks presented in Table III. Each task has been performed through a set of activities. Thus, a new version of the ontology was made available for use (v.1.2). This beta version is available in a repository of free ontologies: http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library.
5. Conclusions

This paper showed a new way of evaluating enterprise ontologies using software usability evaluation techniques with the participation of users. The approach was presented in a systematic way, through six well-defined stages. A case study was presented using a business ontology on the levels G and F of the MPS-SW software quality model. A diagnosis of problems and inconsistencies was generated. The recommendations originated from the analysis and participants comments were implemented, introducing several improvements in the ontology. A beta version of the ontology has been made available in a public repository.

Another important point of the systematic proposed is the human-ontology interaction, because the participant’s point of view is analyzed in the test and their satisfaction is measured. Furthermore, the methodology proposed to evaluate business ontologies can be used in other types of ontologies, if appropriate.

References