

End-user story telling with a CIDOC CRM- based semantic wiki

1. Introduction

This paper presents the current state of an experience intended to use the CIDOC CRM as a knowledge representation language. STEM freshmen freely constitute groups with 2 to 4 members and choose a theme; groups have to model, to structure, to write and to present a story within a web-hosted semantic wiki. The main part of the CIDOC CRM is used as an ontological core where students are hanging up classes and properties of the domain related to the story. The hypothesis is made that the CIDOC CRM is providing a stable and extensible basis; and that, once the entry ticket has been paid, the CRM guides the end-user in a quite natural manner for reading, and writing as well, the story.

2. Work hypotheses

Summing up the choices we made (and benefits expected), we can say that - 1 - we selected a semantic wiki among different Knowledge Management Systems (KMS) [Maedche02], mainly for its ease of learning and use. Then, - 2 - we picked up MediaWiki and its semantic extension SMW (Semantic MediaWiki [Krötzsch07]), mainly because of the familiarity due to the widespread use of Wikipedia and the quantity of systems using MediaWiki (and SMW); Finally - 3- , among suitable ontologies, we chose the CIDOC CRM [Crofts and al. 2010], mainly for its broad scope, its free diffusion and its associated resources such as documentation and presentation. We apologize that none of these choices have been dictated with epistemological or experimental reasons, but merely with the « Keep It Simple » principle, that, obviously, is reducing the scientific scope of this paper.

3. Semantic wikis and CIDOC CRM

wiki is basically a set of pages interconnected with links and a semantic wiki allows us to give different meanings to links and to value a set of properties associated with each page (typically called metadata). There is an underlying semantic to the statements syntax used in MediaWiki pages. One of the main indications is given according to categories that a page is claiming to belong. Thus, a domain-familiar user is expecting to find some values related to properties that may be used for these categories. In a semantic wiki, these values (and the properties that the values are instantiating) are formally defined by the end-user, parsed when the page is published, processed and stored as RDF triples, and retrieved during semantic searches.

It is the choice of a common schema (ontology) that allow a user community to share meaning, and we chose to use an ontological core based on a slightly extended version of the « reduced CRM-compatible form » found in the Definition of the CIDOC Conceptual Reference Model [Crofts and al. 2010]. We added a dozen of classes (and most of associated domain properties), mainly to add E21 Person and E40 Legal Body, E31 Document, E38 Image and to achieve a kind of “ontological closure”.

4. Objectives and contents of the course

This course has been given four times over two years: two times in a Master of Information Technology; and two times to “freshman” students during the STEM Bachelor first year (STEM: science, technology, engineering, and mathematics). For the second edition of the course, we noticeably increased the number of hours and we significantly reduced the content. The current version has 15 hours of lessons (including the half-time with exercises); and 12 hours of labs. Tutoring is given without limits on students requests through e-mails exchange and tutor’s intervention into the students’ wikis.

The appraisal of the course’s first edition in Master leads us to hide as much as possible the quagmire of Web 2.0 technologies. The appraisal of the course’s first edition to freshman students leads us to impose a CIDOC CRM reduced version as an ontological core and to support knowledge representation modelling with an Entity-Relationship approach rather a semantic network approach.

Measurements of the second edition for fresh[wo]men are quantitative (thanks to wiki statistics) and qualitative (based on an assessment of an objectives level for each wiki). After a brief presentation of the course content and a classification of objectives, the remaining of this paper is an attempt to interpret quantitative and qualitative results in order to identify the failed spots of the course and to plan possible improvements.

5. A classification of objectives and difficulties

Course objectives can be classified in 4 levels. This classification is much more close to the concept of « capability level » of ISO 15504 standard rather than a classification such as Bloom’s taxonomy [Bloom56]. The 15504 standard has a capability dimension, based upon a measurement framework comprising six process capability levels and their associated process attributes [15504, Part 2, p. v]. Within the 15504, the extent of achievement of a process attribute is measured using an ordinal scale of measurement [15504, Part 2, pp. 10-11]. We use the same scale of measurement of an objective regarding its application in the semantic wiki: N Not achieved, P Partially achieved, L Largely achieved, F Fully achieved. The ordinal points defined above shall be understood in terms of a percentage scale representing extent of achievement.

Global and detailed objectives of each level are presented below.

1. Level 1: reproduction

The student is able to enrich wiki data (content) but rarely its structure. Objectives are:

1. To apprehend the base principles and the syntax of MediaWiki and SMW.
2. To apprehend the difference between category (class) and page (instance).
3. To apprehend the difference between attribute and relationship.

2. Level 2: customization

The student is able to enrich wiki content and structure as well. Objectives are:

ABSTRACT

1. To apprehend impact of a property quantification on user interfaces and wiki structures
2. To apprehend what is the domain and the range of a property and consequences on use.
3. To define and implement semantic searches.

Level 3: transformation

The student is able to model the story domain with the CIDOC CRM and to implement it in a semantic wiki. Objectives are:

1. To apprehend the implementation of an n-ary relationship with binary relationship.
2. To apprehend what is an event-centred modelling.
3. To define and implement inverse attributes (especially through semantic searches)

4. Level 4: reflection

The student reached a true capability level on the Semantic Web. Objectives are:

1. To apprehend the ontological square².
2. To apprehend the problematic of « property of property » and its implementation.
3. To apprehend interoperability issues as multilingualism.

6. Wikis assessment

5. Intermediary appraisal

An appraisal has been established 2 weeks before the deadline of the course. Quantitative facts are given in table 1. Columns meaning is: 2 - effective number of students in the group; 3 - percentage of female students in the group; 4 - number of classes added to the CRM core; 5 - number of used properties added to the CRM core; 6 - number of valued properties (CRM and domain-specific); 7 - number of uploaded files; 8 - number of wiki visits. URLs of wikis is given in column 1 but unfortunately for English readers, wikis are written in French.

Wiki URL	Students nb.	Women %.	Class nb.	Template nb.	Used properties	Valued properties	Up-loaded files	Number of visits
allen2.referata.com	3	33%	3	3	13	200	9	679
charlieetlachocolaterie.referata.com	3	67%	8	13	29	547	19	1618
clamp.referata.com	2	100%	5	5	30	530	33	1385
finalfantasy13.referata.com	1	0%	2	5	11	211	22	470
lesfrescosoen.referata.com	1	0%	0	0	1	3	0	55
lesrolsmaudits.referata.com	1	0%	0	0	1	26	0	127
lol.referata.com	4	75%	3	3	9	185	19	578
lost.referata.com	1	0%	1	1	4	32	0	142
masseffect.referata.com	2	0%	6	2	14	182	1	551
michaeljackson.referata.com	2	0%	6	5	30	378	10	798
ncis.referata.com	3	33%	4	3	30	51	0	575
oasis.referata.com	1	0%	5	4	14	125	4	156
skins.referata.com	2	50%	2	0	24	61	14	785
stargatesatlantis.referata.com	2	0%	3	3	18	222	8	550
stargatesg-1.referata.com	2	0%	3	3	8	41	1	252
starwars.referata.com	2	0%	3	5	15	143	0	436
tintin.referata.com	4	100%	4	4	32	279	18	1137
warcraft3.referata.com	1	0%	1	-	1	-	102	2568

Table 1. Quantitative statistics

ABSTRACT

Some groups reduced to one student (initially or during the course) present data in italics. About the qualitative appraisal using levels and objectives presented in section 5, 18 semantic wikis (for 37 active students) were assessed as follows:

- 1 wiki are in a state of neglect ;
- 1 wiki do not reach Largely or Fully the level 1 ;
- 6 wikis reach Largely or Fully the level 1, but not the level 2;
- 5 wikis reach Largely or Fully the level 2, but not the level 3;
- 5 wikis reach Largely or Fully the level 3, but not the level 4;

We may also notice that between 2 and 5 students reach individually the level 4. Wikis assessment let us detect two main problems:

PB1: confusion between immaterial work (corresponding to the concept of F2 Expression in FRBRoo) and (physical) realisation of the (corresponding to the concept of F5 Item in FRBRoo)

PB2: a difficulty (related to the domain subject in some wikis, e.g. video games) to have an event-centred modelling (hence, a modelling with a persistent entity-centred point of view).

In order to resolve the first point PB1, we added - in a hurry - a lecture to introduce a simplified view of FRBRoo and we gave strong directives to update wikis in order to differentiate F2 Expression aspects from F5 Item aspects. Solving the second point PB2 is under process, but we did not find a universal solution (suitable to all kind of stories: artists' group or single artist, film, TV series, video games, and cartoons). Hence, it requires an individual coaching and dedicated directives has been given to students in order to put back some wikis on the rails.

6. Final assessment (quantitative and qualitative)

Final appraisal will be established in 2 weeks. As the intermediary appraisal, it will be quantitative (related to wiki statistics) and qualitative (related to the level reached). We will also attempt to see how far both problems above were (partially or totally) solved.

7. By-product results

Looking at the gender composition of groups who reached a satisfactory level, we empirically remarked that groups partially or totally composed of female students are assessed at a significant higher level than male-only students group. A second-order result of this work - to be compared with other similar results, such as [Hardy 08] - may be that a course that relies on Web 2.0 technologies and uses group work is much more attracting female students than classical programming courses.

It could be also noted that students are at ease while interacting with tutors through e-mails and wikis, probably because they belongs to the generation Y or Net generation³.

8. Conclusion

This experience has been performed in order to show STEM students that computer science is not only related to potter about a PC or geek programming, partly to attract female

ABSTRACT

students to computing studies. Day life applications such as Internet, electronic commerce, multimedia, social knowledge building, have been taken as a start point in order to lead to computing topics. Looking at students' choice for the second university year seems to indicate that this « marketing » did not produced expected results. However, the part of the course presented in this paper reached a real students' interest and success, that may contribute to the broadcast of ontologies and Semantic Web in general, and the CIDOC CRM widespread in particular.

9. References

Maedche, A.; Motik, B.; Stojanovic, L.; Studer, R.; Volz, R.; 2003, Ontologies for enterprise knowledge management, Intelligent Systems, IEEE , vol.18, no.2, pp. 26- 33, Mar-Apr 2003 doi: 10.1109/MIS.2003.1193654

Krötzsch, M. ; Vrandečić, D.; Völkel, M.; Haller, H. and Studer, R.; “Semantic Wikipedia”, Journal of Web Semantics 5/2007, Elsevier, Amsterdam, 2007, pp. 251–261.

Crofts, N.; Doerr, M.; Gill, T.; Stead, S.; Stiff, M. (editors), “Definition of the CIDOC Conceptual Reference Model”, January 2010, http://cidoc.ics.forth.gr/docs/cidoc_crm_version_5.0.2.pdf (last accessed, April, 39th, 2010)

ISO/IEC 15504:2004, “Information technology -- Process assessment”. Geneva: ISO, 2004.

Benjamin S. Bloom & David R. Krathwohl, The classification of educational goals, by a committee of college and university examiners, Longmans, New York, 1956

Schneider, L. 2008. The Ontological Square and its Logic. In Proceeding of the 2008 Conference on Formal ontology in information Systems: Proceedings of the Fifth international Conference (FOIS 2008) C. Eschenbach and M. Grüninger, Eds. Frontiers in Artificial Intelligence and Applications, vol. 183. IOS Press, Amsterdam, The Netherlands, 36-48.

Hardy, N. 2008. Women in computer science: harnessing the power of web 2.0 to draw women to computer science fields. In *Proceedings of the 9th ACM SIGITE Conference on information Technology Education* (Cincinnati, OH, USA, October 16 - 18, 2008). SIGITE '08. ACM, New York, NY, 59-60. DOI= 10.1145/1414558.1414576

¹ Subject to Patrick Le Bœuf's agreement

² The Ontological Square is a four-categorical scheme that is obtained by crossing two formal distinctions [...] that between types (or universals) and tokens (or particulars) on the one hand, and that between characters (or features) and their bearers (or substrates) on the other hand [Schneider 2008].

³ http://www.businessweek.com/managing/content/mar2008/ca20080313_241443.htm

<p>Vincent Ribaud (with the review of Patrick Le Bœuf¹)</p>
