

Un wiki sémantique basé sur le CIDOC CRM

Vincent Ribaud, Patrick Le Boeuf

► **To cite this version:**

Vincent Ribaud, Patrick Le Boeuf. Un wiki sémantique basé sur le CIDOC CRM. Sixième colloque Veille Stratégique Scientifique et Technologique (VSST 2010), Oct 2010, Toulouse, France. pp.1-12, 2010. <hal-00630579>

HAL Id: hal-00630579

<http://hal.univ-brest.fr/hal-00630579>

Submitted on 10 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UN WIKI SEMANTIQUE BASE SUR LE CIDOC CRM

[Vincent Ribaud](#) (*), [Patrick Le Boeuf](#) (**)

ribaud@univ-brest.fr, patrick.le-boeuf@bnf.fr

(*) [Université de Bretagne Occidentale](#), Département d'Informatique, 20 Avenue le Gorgeu C.S. 93837, 29283 Brest Cedex 3, France

(**) [Bibliothèque nationale de France](#), Service prospective et services documentaires, Quai François Mauriac, 75706 Paris cedex 13, France

Mots clefs :

Gestion des connaissances, modélisation des connaissances, wiki sémantique, CIDOC CRM

Keywords:

Knowledge management, knowledge modeling, semantic wiki, CIDOC CRM

Palabras clave :

Administración del conocimiento, formalización del conocimiento, wiki semántica, CIDOC CRM

Résumé

Cet article présente l'état actuel d'une expérience d'utilisation d'une ontologie de domaine (le CIDOC CRM, relative à la documentation muséographique) comme langage de représentation des connaissances. Les utilisateurs sont des étudiant-e-s de première année de licence STEM, qui pour moitié se destinent à l'informatique, et aux mathématiques ou à la physique pour l'autre moitié. Cette expérience se compose de séances de cours et de travaux pratiques, destinés à appuyer un projet d'écriture collective d'une histoire (ou un jeu) dans un hypertexte appuyé sur les technologies de ce qu'on appelle aujourd'hui le Web 2.0. Par groupe de [1], 2, 3 ou 4, les utilisateurs finaux modélisent, structurent, rédigent et mettent en forme l'histoire (ou le jeu) qu'il-elle-s ont librement choisi de raconter. Leur hypertexte est mis en œuvre dans un wiki sémantique hébergé sur Internet et le public visé est d'abord l'ensemble des étudiant-e-s et leurs réseaux d'« ami-e-s ». La plus grande partie du CIDOC CRM est utilisée comme noyau ontologique auquel les étudiant-e-s accrochent les classes et les propriétés du domaine représenté.

En plus des enjeux liés à l'édition collaborative, certaines difficultés de l'expérience sont liées à la rigueur requise dans la structuration et l'édition des connaissances. L'hypothèse est faite que le CIDOC CRM fournit une base de départ stable et extensible, et qu'une fois payé le prix d'entrée au CRM, celui-ci permet de guider assez naturellement l'utilisateur final dans la lecture comme dans l'écriture de l'histoire.

Cette expérience avait été conçue notamment dans le but d'attirer des élèves (notamment du sexe féminin) vers les études d'informatique. Très peu d'étudiant-e-s remettent en cause leur choix initial et donc l'approche « marketing » visée par ce cours n'obtient pas le succès escompté. Par contre, le domaine de la gestion des connaissances rencontre un réel succès, ce qui pourrait indiquer que, toutes proportions gardées, certains principes et artefacts de ce cours pourraient contribuer à la diffusion des ontologies et du Web sémantique en général, et du CIDOC CRM en particulier.

1 Introduction

Le terme “Web Sémantique” fait référence à la vision du W3C d’un web de données. Le but ultime du web de données est de permettre aux machines de faire du travail plus utile à l’homme et de développer des systèmes qui permettent des interactions de confiance à travers le réseau [1]. Pour un utilisateur final, le web sémantique devrait apporter, en plus de l’accès aux données, les relations entre les données. Comme le web de l’hypertexte, le web de données est construit avec des documents sur le Web. Cependant, dans le web de l’hypertexte, les liens sont des ancres vers des documents hypertextes écrits en HTML. Alors que, dans le web de données, les relations sont des liens entre n’importe quelles choses arbitraires (des objets ou des concepts) décrites en RDF et qu’on peut identifier avec des URIs. Les technologies du web sémantique doivent permettre aux utilisateurs finaux de créer des « entrepôts » de données sur le Web, d’effectuer des recherches, d’établir des inférences à l’aide de vocabulaires (ontologies), d’écrire des règles d’intégration et de raisonnements. L’environnement de travail « sémantique » d’un utilisateur doit jouer le rôle du « memex » imaginé par Vannevar Bush : “*a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.* [2, p. 43]”

La vision de Bush est que l’esprit humain (y compris sa mémoire) fonctionne de manière associative et que ses capacités de compréhension et de cognition peuvent être augmentées en ajustant l’information, et les associations entre informations, plutôt que de les enfermer dans des structures hiérarchiques rigides. “*Wholly new forms of encyclopedias will appear, ready-made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified.* [2, p. 45]”

Un des principaux apports du Web sémantique est de permettre à des communautés d’utilisateurs de partager (et d’inférer) des données porteuses de sens (des informations) grâce aux vocabulaires et ontologies. Cet article s’intéresse à la manière dont de petits groupes d’utilisateurs peuvent construire collectivement et partager des informations structurées à l’aide d’une ontologie évolutive. Un axiome de départ est qu’un wiki [sémantique] fournit une plate-forme collaborative très facile d’accès à des utilisateurs finaux. Le fait que le wiki soit sémantique permet de donner une signification aux différents liens qui relient les pages du wiki. La thèse principale défendue dans cet article est que le wiki sémantique doit être « garni » d’une ontologie (plus exactement d’un noyau ontologique extensible par les utilisateurs) et que parmi les ontologies de haut niveau disponibles telles que OpenCyc (<http://www.opencyc.org/>), WordNet (<http://wordnet.princeton.edu/>), le CIDOC CRM - modèle conceptuel de référence (Conceptual Reference Model, CRM) du Comité International pour la Documentation (CIDOC) de l’ICOM¹ - fournit un noyau de départ adapté et de taille raisonnable (86 classes et 137 propriétés). Le CIDOC CRM a été publié en 2006 par l’ISO en tant que norme internationale ISO 21127:2006 [3].

2 Positionnement du travail

Dans [4], Chan et Chao présentent un travail de recherche conduit parmi 68 petites et moyennes entreprises (PME) qui ont mis en œuvre des initiatives de gestion des connaissances. Les PME sont de taille un peu plus grosse que celles des groupes visés dans cet article (jusqu’à 5-6 personnes) mais la plupart des conclusions de cet article s’appliquent aussi à des groupes plus petits. Leurs conclusions sont que la gestion effective de connaissance est influencée par deux types de capacité, les infrastructures et les processus, qui doivent être déployés dans les PME.

¹ International Council of Museums, Conseil International des Musées, <http://icom.museum/>

Du point de vue des infrastructures, un gestionnaire de contenu (Content Management System, CMS) est généralement utilisé comme support principal de la gestion de la documentation. Plusieurs auteurs critiquent cependant la rigidité du contrôle éditorial requis par un CMS [5] ainsi que la nécessité d'un équilibre entre structure/contrainte et flexibilité [6]. Certains auteurs défendent l'utilisation de wikis et de RDF (Resource Description Framework) pour résoudre ces problèmes [7]. Concernant l'infrastructure, nous avons fait deux hypothèses. Une première hypothèse de travail est qu'un wiki (sémantique) est une plate-forme à la fois facile d'accès et favorisant le travail en groupe. Une deuxième hypothèse de travail est que les premiers pas de modélisation de connaissances d'un utilisateur final doit être fait « par l'exemple » et dans un environnement familier. La diffusion et l'outillage de Wikipédia (le logiciel libre MediaWiki <http://www.mediawiki.org/wiki/MediaWiki/fr>) fournit un bon point de départ. L'extension sémantique Semantic MediaWiki [8] - SMW - (http://semantic-mediawiki.org/wiki/Semantic_MediaWiki) permet la structuration des connaissances d'un wiki à un moindre coût qu'avec un « vrai » système de gestion de connaissance (*Knowledge Management System, KMS*). Nous avons donc retenu la plate-forme MediaWiki+SMW comme environnement s'approchant le plus de cette seconde hypothèse.

Rech et al. ont identifié plusieurs challenges relatifs aux processus de gestion de connaissances : enregistrement, réutilisation, recherche et partage [9]. Nous avons établi la simplicité d'utilisation comme première priorité. En effet, nous préférons utiliser les fonctionnalités (relativement) simples d'un wiki plutôt que de s'appuyer sur un système dédié (et probablement complexe) de gestion de connaissances. En accord avec [6], nous pensons que *“the strategy of building larger-scale editorial structures on the thinnest of technological foundations seems to also mean that the work of organizing and managing the content in the system can remain an editorial process, rather than a technical administration task.”* Lorsqu'on examine l'avis que la communauté émet sur ce qu'on peut appeler *Ontology-based Knowledge Management System (OKMS)* comme dans [10] ou [11] : *“... enterprise-wide OKMS is still in the early stage. Despite much research on ontology representation, engineering, and reasoning, features such as scalability, persistency, reliability, and transactions - standardized and widely adopted in classical database-driven information system - are typically not available in ontology-based system.”* Or, un wiki sémantique offre la plupart des fonctionnalités mentionnées quelquefois très simplifiées mais la plupart du temps suffisante pour la gestion des connaissances de type encyclopédique.

3 Le modèle conceptuel de référence du CIDOC

Le contenu des sections 3.1, 3.2 et 3.3 est un extrait simplifié de [12].

3.1 Objectifs et portée

Le CRM (Conceptual Reference Model) est un modèle qui vise à expliciter sans ambiguïté le sens des informations relatives à un « objet de musée » dans l'acception la plus large du terme (œuvre d'art, vestige archéologique tel qu'un clou ou un four à pain, corps momifié, fleur séchée ou animal empaillé, mais aussi quelque chose qu'il est matériellement impossible de conserver à l'intérieur d'un musée mais qu'une institution du patrimoine culturel peut gérer, comme un monument ou un site tel qu'une grotte préhistorique), et ce quelle que soit la structure des données qui véhiculent ces informations. Le CRM est à la fois un modèle orienté objet (reposant sur le mécanisme des « héritages » de classes et de propriétés) et un modèle de réseau sémantique mais c'est essentiellement, dans le cadre présenté dans cet article, une ontologie de domaine qui vise à définir formellement et à expliciter un ensemble de concepts sous-jacents, non nécessairement exprimés. La mention « de domaine » indique une restriction, une limitation à cette entreprise ontologique : il s'agit en fait tout simplement d'un effort visant à définir, non pas tout ce qui existe au monde (tâche virtuellement impossible et infinie), mais « seulement » tout ce qui présente un caractère de pertinence par rapport à la documentation muséographique.

3.2 Origine et perspectives

Le CRM a été élaboré par le Documentation Standards Working Group du CIDOC (Comité international pour la documentation), lequel est lui-même une émanation du Conseil international des Musées (ICOM). Il succède en fait à un modèle de données de type entité relation qui, au fil du temps, était devenu ingérable.

Le CRM a été soumis à l'ISO en 2000 et est devenu la norme internationale ISO 21127 en 2006 [3]. Le fait que le CRM soit devenu une norme ISO ne signifie nullement qu'il soit « obligatoire ». Cette validation contribue simplement à asseoir la position du CRM, à lui donner une plus grande « légitimité » et à le rendre disponible bien au delà du seul cercle de la documentation muséographique.

3.3 Contenu

Au cœur même du CRM, on trouve la notion d'événement, de phénomène temporel. En effet, si intuitivement on pourrait penser que ce qui est primordial c'est la description d'un objet matériel tel qu'il se présente aujourd'hui, il ne faut pas oublier qu'il s'agit avant tout de resituer cet objet dans un contexte historique. Ce qui importe, c'est à la fois l'environnement de cet objet au cours du temps (au moment de sa création, et lors de chacune de ses éventuelles modifications) et l'énoncé de tout ce qui a pu lui arriver.

Nous nous intéressons donc avant tout à l'*historique* de tout ce qui est arrivé aux objets que nous décrivons, c'est pourquoi le cœur même du modèle est constitué des *Entités Temporelles*, qui expriment soit le fait qu'un objet est « dans tel état » (classe *État de Conservation*), soit le fait qu'il « s'est produit quelque chose » (un *Événement*, une *Activité*, un *Déplacement*, une *Production*, un *Début d'Existence* ou une *Fin d'Existence*, etc.) ; à partir de cette notion centrale, nous nous posons un certain nombre de questions fondamentales :

Quand est-ce arrivé ? nous répondons à cette question au moyen de la classe *Durée* ;

Où est-ce arrivé ? cette question implique l'existence de la classe *Lieu* (la notion de lieu étant comprise comme pouvant se définir par rapport à une constellation déterminée de matière, indépendamment de l'emplacement de cette constellation : comme un poinçon sur une bague, qui reste toujours au même endroit par rapport à la bague, alors que la bague elle-même se déplace à travers le monde au gré des déplacements de la personne qui la porte) ;

Qui a fait ça ? question qui met en jeu la notion d'*Agent* (laquelle implique une notion d'intentionnalité : un singe n'est pas considéré comme une instance d'*Agent*, même s'il détruit un tableau) ;

À quoi est-ce arrivé ? (donc, quel est l'objet décrit ?), question qui débouche à la fois sur la notion de contenant et sur celle de contenu immatériel, donc sur les notions de *Chose Matérielle* (qui peut être naturel ou avoir été fabriqué de main d'homme) et d'*Objet Conceptuel* (qui nécessairement ne peut qu'avoir été « fabriqué », c'est-à-dire être un produit de l'activité humaine) ;

À ce stade, on éprouve le besoin de *classifier* toutes ces informations : l'activité de classification met en jeu la notion de *Type* (les types peuvent s'appliquer absolument à tout élément du système)

Enfin, pour pouvoir parler aisément de toutes ces choses, il faut bien évidemment leur donner un nom, ce qui met en jeu la notion d'*Appellation*.

3.4 Exemple

La figure 1 présente une copie d'écran d'une page typique sur un Ampèremètre.

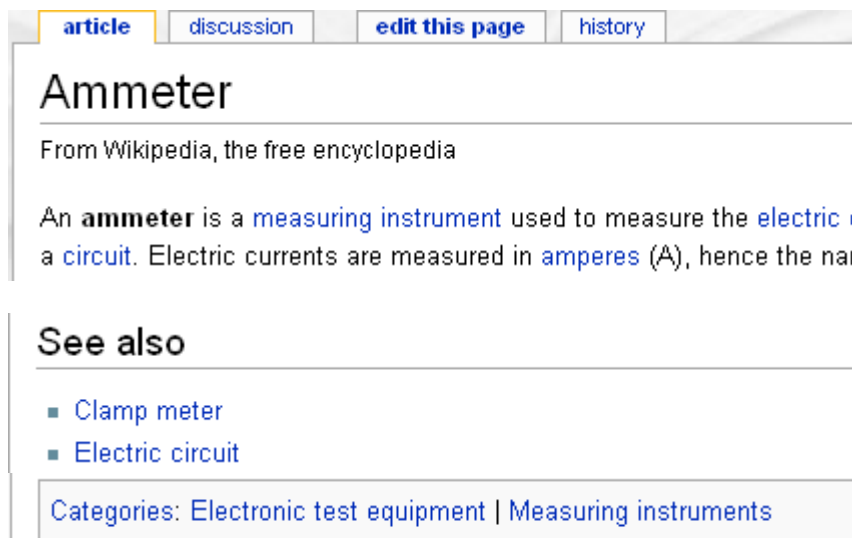


Figure 1 - Extrait d'une page de « Wikipédia » sur un ampèremètre.

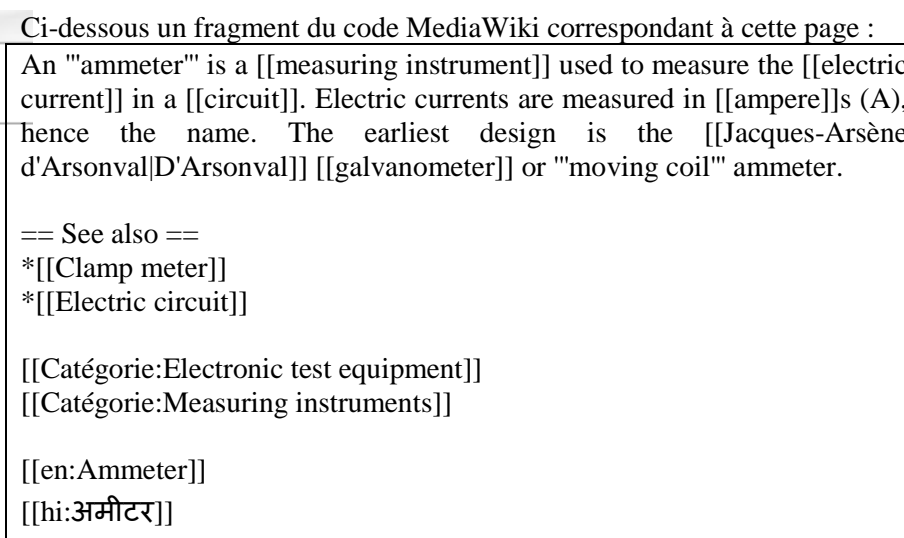


Figure 2 - Fragment de code MediaWiki d'une page de « Wikipédia » sur un ampèremètre.

Il existe une sémantique sous-jacente à la syntaxe des constructions utilisées dans les pages de code MediaWiki. Quelles sont les constructions de MediaWiki utilisées ? Que savons-nous et que comprenons-nous à la lecture du code de cette page ?

Cette page décrit « quelque chose » et l'URL de cette page sert d'identifiant pour ce « quelque chose » (qu'on appellerait une ressource au sens de RDF). Cette page contient des informations (valeurs) typées (syntaxiquement et sémantiquement). C'est donc un réservoir de triples RDF dont « quelque chose » est le sujet.

Cette page est rangée dans une ou plusieurs catégories, qu'on interprétera soit comme des types (avec la propriété *P2 has type*), soit comme des classes.

Cette page est liée à d'autres pages ; une des principaux apports de SMW est de permettre à l'utilisateur final de typer les liens, c'est-à-dire d'instancier des propriétés directes dont la(es) classe(s) d'appartenance de la page sont les domaines ou des propriétés inverses dont la(es) même(s) classe(s) sont les images. Dans la plupart des cas, le typage des propriétés sera à la charge de l'utilisateur, mais il existe des contextes où on peut inférer la propriété (comme la section See also, qui lie la page avec d'autres pages ; si ces pages sont des *E70 Thing*, ce lien peut être typé par *P130 shows features of*).

Les liens inter-langage (comme [[en:Ammeter]] [[hi:अमीटर]]) permettent de lier l'identifiant de la page avec d'autres identifiants (en théorie de la "même" page traduite dans d'autres langages).

Une des constructions les plus importantes de ce fragment de code est l'indication que cette page (relative à l'instance dont on parle) appartient aux catégories « Instrument de mesure électrique » et « Instrument électronique ». Une personne familière du domaine s'attend alors à disposer de valeurs pour les propriétés sous-jacentes à la catégorie (classe) « Instrument de mesure électrique » comme le type de courant (alternatif ou continu), la caractéristique mesurée (intensité, différence de potentiel, résistance ...), l'unité de mesure (Ampère, Volt, Ohm ...). Dans un wiki sémantique, ces valeurs (et les propriétés qu'elles instancient)

sont formellement définies par l'utilisateur, analysées lors de la publication de la page, converties et stockées sous forme de triples RDF, et retrouvées par l'emploi de requêtes sémantiques.

Tout wiki sémantique repose sur une organisation des classes et des propriétés - le modèle de données - qu'on appelle schéma (en XML ou RDFS), ontologie (en OWL). C'est le choix d'un schéma commun qui permet à la communauté d'utilisateurs de partager des significations, nous avons choisi d'utiliser un noyau ontologique basé sur une version réduite du CIDOC CRM [13].

Un des choix les plus importants fait par le catalogueur est de classer la page, soit dans une classe existante du noyau ontologique (par exemple *E24 Physical Man-Made Thing*), soit dans une classe spécifique au domaine (par exemple Scientific Instrument).

Si on prend la première option, les propriétés dont *E24 Physical Man-Made Thing* et ses sur-classes sont le domaine deviennent disponibles pour indexer ce fragment de code et un exemple typique de « sémantisation » est donné ci-dessous.

```
An "ammeter" is a [[measuring instrument]] used to measure the [[electric current]] in [[P39B.P40 observed dimension::ampere]]s (A), hence the name. The earliest design is the [[P14 carried out::Jacques-Arsène d'Arsonval|D'Arsonval]] [[P130 shows feature of::galvanometer]] or "moving coil" ammeter.

[[Category:E24 Physical Man-Made Thing]]
```

Figure 3 - Fragment de code « sémantique » correspondant au fragment de la figure 1.

La dernière instruction indique que ce fragment sémantique est associé à une instance de *E24 Physical Man-Made Thing* (une des classes de l'ontologie).

Lors de la publication de la page auquel ce fragment appartient, toutes les constructions utilisant la syntaxe `[[Px property name::value]]`, telle que `[[P130 shows feature of::galvanometer]]` sont converties en triples RDF (< sujet, prédicat, objet >) dont la page est le sujet, la propriété le prédicat et la valeur l'objet.

4 Utilisation finale de wikis sémantiques basés sur le CIDOC CRM

4.1 Cas d'étude

Cet article présente l'état actuel d'une expérience d'utilisation du CIDOC CRM comme langage de représentation des connaissances. Les utilisateurs sont des étudiant-e-s de première année de licence STEM (*science, technology, engineering, and mathematics*), qui pour moitié se destinent à l'informatique, et aux mathématiques ou à la physique pour l'autre moitié. Classiquement, cette expérience pédagogique se compose de séances de cours et de travaux pratiques, destinés à appuyer un projet d'écriture collective d'une histoire (ou un jeu) dans un hypertexte appuyé sur les technologies de ce qu'on appelle aujourd'hui le Web 2.0. Par groupe de 2, 3 ou 4, les utilisateurs finaux modélisent, structurent, rédigent et mettent en forme l'histoire (ou le jeu) qu'il-elle-s ont librement choisi de raconter. Leur hypertexte est mis en œuvre dans un wiki sémantique hébergé sur Internet et le public visé est d'abord l'ensemble des étudiant-e-s et leurs réseaux d'« ami-e-s ». La plus grande partie du CIDOC CRM est utilisée comme noyau ontologique auquel les étudiant-e-s accrochent les classes et les propriétés du domaine représenté.

En plus des enjeux liés à l'édition collaborative, certaines difficultés de l'expérience sont liées à la rigueur requise dans la structuration et l'édition des connaissances. L'hypothèse est faite que le CIDOC CRM fournit une base de départ stable et extensible, et qu'une fois payé le prix d'entrée au CRM, celui-ci permet de guider assez naturellement l'utilisateur final dans la lecture comme dans l'écriture de l'histoire.

Cet article présente les enjeux et les résultats des wikis sémantiques réalisés cette année.

4.2 Contenu du cours

Le cours comporte 15 heures de cours avec pour moitié des séances d'exercices, et 12 heures de travaux pratiques. Le cours adresse les points suivants :

- * Modèle E-A de base : entité-association binaire-attribut
- * Modèle E-A étendu : cardinalités, spécialisation/généralisation
- * Modèle conceptuel de référence - le CIDOC CRM : But et forme ; Présentation du contenu ; Comment utiliser le CIDOC CRM
- * Principes et présentation technique de MediaWiki
- * Analyse d'une histoire avec le CIDOC
- * Requêtes sémantiques : définition et usage
- * Introduction à FRBRoo : Œuvre, Expression, Manifestation
- * Modélisation centrée sur les événements

4.3 Critères d'évaluation du niveau de capacité

Pour des (presque) novices en informatique, on peut classer les objectifs du cours en 4 niveaux. Cette classification a été établie a posteriori pour évaluer ce qui a été compris. En ce sens, cette classification s'approche de la notion de «niveau de maturité (*capability level*) » comme on peut trouver dans les normes d'ingénierie du logiciel tel que le standard ISO 15504 [14] ou la norme CMMI (<http://www.sei.cmu.edu/cmmi/>).

L'ISO 15504 a une infrastructure de mesure comprenant plusieurs niveaux de maturité, chaque niveau étant associé à différents attributs [14, Part 2]. Le degré de réalisation d'un attribut est mesuré sur une échelle ordinale [14, Part 2, pp. 10-11]. Nous avons associé les attributs à des objectifs et utilisons la même échelle ordinale que dans la 15504 :

N Not achieved - Il n'y a pas ou très peu d'évidence que l'objectif a été atteint dans le wiki sémantique.

P Partially achieved - Il y a quelques évidences d'une approche vers et quelques résultats liés à l'objectif dans le wiki sémantique.

L Largely achieved - Il y a une évidence significative d'une approche vers et de nets résultats liés à l'objectif dans le wiki sémantique. Certaines faiblesses relatives à l'objectif peuvent cependant exister dans le wiki sémantique.

F Fully achieved - Il y a une nette évidence d'une approche complète et systématique vers et des résultats complets de l'objectif dans le wiki sémantique. Aucune faiblesse relative à l'objectif n'existe dans le wiki sémantique.

La suite de cette section présente les objectifs regroupés par niveau. Afin d'illustrer certains objectifs, quelques exemples sont pris à partir de la « sémantisation » d'une page Web, en l'occurrence celle du colloque VSST'2010. Sur la page d'accueil, on y trouve les informations suivantes :

Ce sixième colloque VSST et les ateliers qui s'y rattachent sont organisés conjointement par U.P.C./I.R.I.T./S.F.B.A.

LIEU : Université Paul Sabatier (Toulouse - France) et DATE : 25-29 Octobre 2010

PROGRAMME PROVISOIRE

2 liens : Inscription en ligne pour le colloque VSST'2010 et Pour mémoire: index des publications de VSST'2007 (Marrakech)

SECRETARIAT :

Colloque VSST'2010, UPS/IRIT/VSST, 118, route de Narbonne, 31062 Toulouse Cedex 9, 00 (33) 5.61.55.67.81 / 00 (33) 6.12.57.19.49, vsst'2010@irit.fr

4.3.1 Niveau 1: reproduction

A ce niveau, l'étudiant-e est capable de reproduire des constructions du wiki ou d'autres wikis : pages, sélection de propriétés et affectation de valeurs, téléversement de fichiers, mise en page simple. Il-elle est alors capable d'enrichir les données du wiki mais rarement sa structure. Les objectifs sont les suivants :

- Appréhender les principes de base et la syntaxe de MediaWiki et SMW
- Appréhender la différence entre catégorie (classe) et page (instance)
- Appréhender la différence entre attribut et association

La syntaxe de MediaWiki facilite l'édition de pages et l'insertion de liens hypertexte est aisée. On peut imaginer que le point d'entrée d'un site (sémantique) pour le colloque VSST'2010 serait une page Colloque VSST qui contiendrait un lien vers chaque édition et notamment la page principale de l'édition 2010. Quel statut donner à l'objet Colloque VSST : une instance d'une catégorie d'événements récurrents ou une catégorie à part entière qui serait une spécialisation de la catégorie Événement. Dans le premier cas, l'instance « Colloque VSST » serait liées aux différentes éditions avec une relation tout-partie, comme la propriété *P9 consiste en (fait partie de)*. Dans le deuxième cas, les différentes éditions seraient des instances de la nouvelle catégorie *Colloque VSST*. Quelle est la nature de la relation entre l'édition 2010 et l'Université Paul Sabatier : attribut ou association ? Dans le premier cas, l'Université n'est qu'un élément d'information de l'édition 2010 - une valeur de l'attribut Lieu. Dans le second cas, l'Université est une instance à part entière qui sera décrite dans sa propre page, et sera associée avec l'édition de 2010 par exemple avec la propriété *P7 a eu lieu dans (a été témoin de)*.

4.3.2 Niveau 2: personnalisation

A ce niveau, l'étudiant-e est capable de créer de nouvelles constructions (catégories, propriétés, modèles). Il-elle peut faire évoluer l'apparence du wiki en personnalisant les menus, le « skin » ... Il-elle est donc capable d'enrichir la structure du wiki autant que son contenu. Les objectifs sont les suivants :

- Appréhender l'impact d'une cardinalité sur l'interface et le contenu
- Appréhender ce que sont le domaine et l'image (en anglais, *range*) d'une propriété et leurs conséquences sur son utilisation
- Définir et mettre en œuvre des requêtes sémantiques

Le comité scientifique de l'édition 2010 a un président. Y en a-t-il toujours un-e et un-e seul-e ? Une cardinalité maximum à 1 permet de ne réserver qu'une seule place dans un formulaire de saisie et de consultation des éditions du colloque. Dans le cas contraire, il faut une interface capable d'afficher une liste. On peut créer une propriété Président de comité dont le domaine serait la catégorie *Colloque VSST* et l'image (*range*) la catégorie *E21 Personne* ; cette propriété ne peut s'employer que pour les éditions de VSST - pas pour d'autres colloques - et la présidence ne peut pas être tenue par une personne morale. Le domaine pourrait être une catégorie plus générale *Colloque* et l'image la catégorie plus générale *E39 Agent* (superclasse de groupe, collectivité, personne).

4.3.3 Niveau 3: transformation

A ce niveau, l'étudiant-e est capable de faire évoluer la structure du wiki, comme déplacer une catégorie, modifier la sémantique d'une propriété ou mettre à jour un modèle (pour prendre une analogie, dans le niveau précédent, on agit sur la décoration intérieure alors qu'au niveau transformation, on travaille sur le gros œuvre). Il-elle est confronté-e aux « vrais » problèmes de modélisation où on est amené à prendre des décisions et à appliquer des heuristiques. Il-elle est donc capable de modéliser son domaine à partir du CIDOC CRM et de l'implanter dans un wiki sémantique. Les objectifs sont les suivants :

- Appréhender la mise en œuvre de relations n-aires par des relations binaires
- Appréhender la modélisation centrée sur les événements
- Définir et utiliser les attributs inverses (notamment avec des requêtes sémantiques)

Par définition, un colloque est une rencontre de différentes personnes en un lieu et temps donné, qui peut donc être vu comme une relation n-aire. Le CIDOC CRM établit une distinction fondamentale entre entités persistantes (censés être indépendantes du temps tels les agents ou les objets) et entités temporelles (par définition dépendantes du temps mais la plupart du temps situées aussi dans l'espace), et un colloque fait naturellement partie de la deuxième catégorie. Ceci transforme de facto le colloque d'une relation n-aire à une instance de classe dotée d'un certain nombre de relations binaires (lieu, date, secrétariat ...). La transformation peut être moins triviale lorsque les relations n-aires à transformer n'invoquent pas le temps. Ainsi, « sémantiser » avec le CIDOC CRM un site « encyclopédique » de type Wikipédia peut s'avérer difficile, voire impossible à moins de recréer un fil historique pour chacun des articles de l' « encyclopédie ».

4.3.4 Niveau 4: réflexion

A ce niveau, l'étudiant-e est capable de raisonner à un niveau méta et fait la différence (parfois intuitivement) entre les niveaux d'abstraction des connaissances et de leur implantation dans le wiki. Il-elle est alors capable de voir les limites de l'implantation d'une ontologie de domaine dans un wiki sémantique, et a atteint un bon niveau de compréhension sur le Web sémantique. Les objectifs sont les suivants :

- Appréhender le carré ontologique²
- Appréhender la problématique des « propriétés de propriétés » et de leur mise en œuvre
- Appréhender les problèmes d'interopérabilité comme le multilinguisme

La notion de sous-propriété existe dans le CIDOC CRM - un auteur communiquant dans une session du colloque serait relié par la propriété *P14B a exécuté* à sa session, ce qui le relierait automatiquement à cette session avec la sur-propriété *P11B a participé* à. Cependant, on ne peut prévoir tous les cas d'extensibilité et l'emploi de « propriétés de propriétés » est souvent plus lisible. Ainsi les différents rôles des différentes personnes agissantes dans une session, respectivement président-e, présentateur-trice, auteur-trice, pourront spécialiser l'usage de la propriété *P14 réalisée par (a exécuté)* grâce à la propriété de propriété (*P14.1 in the role of: E55 Type*) où le rôle de chaque agent pourra être précisé.

Le multilinguisme des wikis (sémantiques ou non) pose en premier lieu le besoin de disposer d'identificateurs d'instances stables, afin de pouvoir référencer mutuellement deux (ou plusieurs) pages comme des traductions mutuelles. Or la stabilité n'apparaît qu'au fil du temps, lorsque la connaissance représentée dans ces wikis est arrivée suffisamment à maturité. Le multilinguisme et l'interopérabilité appellent à des solutions ad hoc liées au domaine d'application.

5 Evaluation des wikis sémantiques

5.1 Résultats

Les projets se sont déroulés de février à mai 2010. Le bilan ci-dessous a été établi après la remise du projet. Le tableau 1 présente les statistiques de chaque wiki : la 1^{ère} colonne donne l'URL du site (en français), la 2^{ème} le nombre de contributeur-trice-s par projet, la 3^{ème} la date d'évaluation, de la 4^{ème} à la 6^{ème} les objectifs du niveau 1, de la 7^{ème} à la 9^{ème} les objectifs du niveau 2, de la 10^{ème} à la 12^{ème} les objectifs du niveau 3, de la 13^{ème} à la 15^{ème} les objectifs du niveau 4, et la 16^{ème} et dernière la note globale (sur 20) attribuée au wiki sémantique.

² The Ontological Square is a four-categorical scheme obtained by crossing two formal distinctions which underpin conceptual modelling languages and top-level ontologies alike: that between types (universals) and tokens (particulars) on the one hand, and that between characters (features) and their bearers (substrates) on the other hand. [15]

Les groupes en italique sont constitués d'un seul élève.

URL du site	Nb. élèves	% de filles	Date évaluation	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3	4.1	4.2	4.3	Note
<i>finalfantasy13.referata.com</i>	1	0%	24-mai-10	F	F	F	F	F	F	F	F	F	P	N	N	16
<i>lesroismaudits.referata.com</i>	1	0%	06-mai-10	L	P	P	P	N	N	N	N	N	N	N	N	10
<i>lost.referata.com</i>	1	0%	26-mai-10	F	F	F	F	F	F	F	F	L	L	L	P	16
masseffect.referata.com	2	0%	26-mai-10	F	F	F	F	F	F	F	F	F	P	N	P	15
michaeljackson.referata.com	2	0%	26-mai-10	F	F	F	F	F	F	F	F	F	P	N	P	16
<i>oasis.referata.com</i>	1	0%	27-mai-10	F	F	F	F	F	L	L	L	N	N	N	N	13
stargatesatlantis.referata.com	2	0%	27-mai-10	F	F	F	F	F	F	L	L	N	N	N	N	14
stargatesg-1.referata.com	2	0%	27-mai-10	F	F	F	L	L	F	P	P	N	N	N	N	13
starwars.referata.com	2	0%	27-mai-10	F	F	F	F	F	F	L	L	F	N	N	N	15
<i>warcraft3.referata.com</i>	1	0%	07-mai-10	F	F	F	P	P	N	N	N	N	N	N	N	13
ncis.referata.com	3	33%	27-mai-10	F	F	F	F	F	P	L	L	N	N	N	N	13
alien2.referata.com	3	33%	18-mai-10	F	F	F	F	F	F	L	L	N	N	N	N	14
skins.referata.com	2	50%	27-mai-10	F	F	F	F	F	F	L	P	N	N	N	N	14
charlieetlachocolaterie.referata.com	3	67%	23-mai-10	F	F	F	F	F	F	F	F	P	P	N	P	16
clamp.referata.com	2	100%	23-mai-10	F	F	F	F	F	F	F	F	L	P	N	P	16
lol.referata.com	3	100%	26-mai-10	F	F	L	L	L	L	P	L	L	N	N	N	14
tintin.referata.com	4	100%	27-mai-10	F	F	F	F	F	F	L	L	N	N	N	N	15

Tableau 1 : Evaluation des niveaux de capacité sur 17 wikis sémantiques

5.2 Discussion

Une évaluation intermédiaire réalisée début avril a permis de détecter deux problèmes :

1. une confusion entre le contenu intellectuel d'une « Œuvre (incluant le contenu conceptuel qui sous-tend l'ensemble de ses versions linguistiques) et les réalisations matérielles de l'« Œuvre » (un objet spécifique par exemple composé de papier et d'une reliure si c'est un livre, ainsi que l'exemplaire identifié par un ISBN) ;
2. une difficulté (relative quelquefois au sujet du wiki, par exemple un jeu vidéo) d'avoir une approche de modélisation centrée sur les événements (les entités temporelles ou « perdurants ») alors qu'une encyclopédie a une approche centrée sur les entités persistantes (les « durables »).

L'IFLA (<http://www.ifla.org/>), Fédération internationale des associations de bibliothécaires et d'institutions est la principale organisation internationale représentant les intérêts des bibliothèques, des services d'information et de leurs usagers. L'IFLA a établi les fonctionnalités requises des notices Bibliographiques appelées Functional Requirements for Bibliographic Records (FRBR), c.à.d. l'essence même de ce sur quoi la notice bibliographique est censée renseigner, et l'essence même de ce que nous attendons de la notice en termes d'adéquation aux besoins des utilisateurs [16]. Les FRBR ont un modèle très clair à 4 niveaux des différents aspects d'une Œuvre (du concept à l'exemplaire) et ces aspects ont été introduits à la fin du cours pour tenter de résoudre le problème n° 1, mais ce modèle des FRBR n'a pas été compris par les élèves (ou bien a été introduit trop tard). Nous n'avons pas trouvé de cadre général s'adaptant à tous les types d'histoire racontées dans les wikis et le premier auteur a du fournir une réponse individuelle aux groupes « en panne » sur le problème n° 2. Dans la plupart des cas, le problème a été résolu grâce à l'assistance fournie.

Concernant la composition en genre des groupes, on peut remarquer empiriquement que - 1 - les groupes réduits à 1 élève sont exclusivement masculins, et - 2 - les groupes qui sont composées partiellement ou totalement d'élèves féminins ont une moyenne (14,5/20) supérieure à celle (14/20) des groupes composés exclusivement d'élèves masculins. Ce résultat empirique semble confirmé par une étude comme celle de [17] qui énonce qu'un cours basé sur les technologies du Web 2.0 et qui utilise des techniques de travail en groupe est plus susceptible d'attirer les élèves féminines que des cours de programmation classique. Nous avons observé que les groupes à majorité féminine s'encourageaient et s'aidaient plus que ceux à prédominance masculine, avec une répartition du travail naturelle et sans heurt. Dans les groupes masculins, nous avons observé plus de rivalités sur le leadership, conduisant à une moindre productivité.

Il existe plusieurs formalismes susceptibles d'appuyer la modélisation des connaissances: modèle entité-association, cartes topiques (topic map) ou réseau sémantique (semantic network). Bien que les deux derniers formalismes soient généralement considérés comme de bons outils de représentation de connaissances - et les réseaux sémantiques très proches des graphes RDF générés par un wiki sémantique - nous avons privilégié le modèle entité-association binaire-attribut étendu avec l'usage de la généralisation/spécialisation, à cause de son usage plus répandu dans la branche professionnelle informatique.

6 Conclusion

De nombreux travaux [17, 18, 19, 20] indiquent que la recherche d'informations sur le Web est en train d'évoluer d'un modèle QIRO (query in, results out) essentiellement basés sur l'indexation de mots-clés et le « ranking » pour sélectionner les résultats vers des modèles sémantiques incluant des recherches plus sophistiquées (statistiques ou linguistiques), des techniques de catégorisation et d'extraction de contenus, de nouveaux paradigmes de visualisation des résultats. La proposition faite dans cet article d'utiliser un wiki sémantique « garni » d'une ontologie de domaine extensible s'inscrit donc dans cette évolution.

Cette expérience avait été conçue pour montrer à des étudiant-e-s de STEM que l'informatique ne se résume pas à la programmation ou au bricolage de leur PC, notamment dans le but d'attirer des élèves (notamment du sexe féminin) vers les études d'informatique. L'évaluation empirique de l'expérience semble indiquer que les étudiant-e-s de STEM entrent en première année d'université avec un choix d'orientation précis et que même si une vision plus objective de l'informatique est obtenue à la fin de ce cours, très peu d'étudiant-e-s remettent en cause leur choix initial et donc que l'approche « marketing » visée par ce cours n'obtient pas le succès escompté. Par contre, le domaine de la gestion des connaissances rencontre un réel succès, ce qui pourrait indiquer que, toutes proportions gardées, certains principes et artefacts de ce cours pourraient contribuer à la diffusion des ontologies et du Web sémantique en général, et du CIDOC CRM en particulier.

7 Bibliographie

- [1] Semantic Web Official Homepage. <http://www.w3.org/standards/semanticweb/> (last accessed 2010 September, 13th)
- [2] Bush, V. As we may think. *interactions* 3, 2 (Mar. 1996), 35-46. 1996. [DOI](#)
- [3] International Standard Organization. ISO 21127:2006 Information and documentation -- A reference ontology for the interchange of cultural heritage information. ISO, 2006.
- [4] Chan, I. and Chao, C. Knowledge management in small and medium-sized enterprises. *Commun. ACM* 51, 4 (Apr. 2008), 83-88. 2008. [DOI](#)
- [5] García Alonso, J. M., Berrocal Olmeda, J. J. and Murillo Rodríguez J. M. Documentation Center: Simplifying the Documentation of Software Projects. Proc. Wiki4SE Workshop 4th International Symposium on Wikis, Porto, 2008.
- [6] Maxwell, J. W. Using wiki as a multi-mode publishing platform. In *Proceedings of the 25th Annual ACM international Conference on Design of Communication* (El Paso, Texas, USA, October 22 - 24, 2007). SIGDOC '07. ACM, New York, NY, 196-200. 2007. [DOI](#)
- [7] Rauschmayer, A. Next-Generation Wikis: What Users Expect; How RDF Helps. Third Semantic Wiki Workshop at ESWC, Redaktion Sun SITE, Aachen, 2009, poster.
- [8] Krötzsch, M., Vrandečić, D., Völkel, M., Haller, H., and Studer, R. 2007. Semantic Wikipedia. *Web Semant.* 5, 4 (Dec. 2007), 251-261. [DOI](#)
- [9] Rech, J., Bogner, C., and Haas, V. 2007. Using Wikis to Tackle Reuse in Software Projects. *IEEE Softw.* 24, 6 (Nov. 2007), 99-104. [DOI](#)
- [10] Maedche, A., Motik, B., Stojanovic, L., Studer, R., and Volz, R. Ontologies for enterprise knowledge management. *Intelligent Systems*, 18, 2 (Mar-Apr 2003), 26-33, [DOI](#)
- [11] Razmerita, L., 2007. Ontology-Based User Modeling for Knowledge Management Systems. Chapter In *Ontologies A Handbook of Principles, Concepts and Applications in Information Systems*, Integrated Series in Information Systems, Vol. 14, Springer-Verlag, pp 635-664, 2007, [DOI](#)
- [12] Le Boeuf, P. Le modèle CRM pour la documentation muséographique: s'attacher au sens pour n'être pas piégé par la forme. Journées d'étude de l'ADBS. 2003.
- [13] Crofts N., Doerr M., Gill T., Stead S., Stiff M. (editors). Definition of the CIDOC Conceptual Reference Model, http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.2.pdf (last accessed 2010 September, 13th)
- [14] ISO/IEC 15504:2004. Information technology -- Process assessment. Geneva: ISO, 2004.
- [15] 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.
- [16] Doerr, M., Le Boeuf, P. Modelling Intellectual Processes: The FRBR - CRM Harmonization. CIDOC Conference on September 11, Gothenburg, Sweden, 2006.
- [17] Roussinov, D., Fan, W., and Robles-Flores, J. Beyond keywords: Automated question answering on the web. *Commun. ACM* 51, 9 (Sep. 2008), 60-65. 2008. [DOI](#)
- [18] Hepp, M., Siorpaes, K., and Bachlechner, D. Harvesting Wiki Consensus: Using Wikipedia Entries as Vocabulary for Knowledge Management. *Internet Computing, IEEE*, vol.11, no.5, pp.54-65, Sept.-Oct. 2007. [DOI](#)
- [19] Kasneci, G., Ramanath, M., Suchanek, F., and Weikum, G. The YAGO-NAGA approach to knowledge discovery. *SIGMOD Rec.* 37, 4 (Mar. 2009), 41-47. 2009. [DOI](#)
- [20] Rao, R. From IR to Search, and Beyond. *Queue* 2, 3 (May. 2004), 66-73. 2004. [DOI](#)