
Utilisation d'ontologies dans une application médicale décisionnelle

Thibaud Raby*,,*** — Frédéric Ravaut*****

**France Télécom / Orange Labs
BIZZ / VMC / Données et santé
28, chemin du vieux chêne – BP 98, 38243 Meylan Cedex, France
thibaud.raby@orange-ftgroup.com*

***Laboratoire LISSI
Université de Paris Est Créteil
120 rue Paul Armangot – 94400 Vitry-sur-Seine*

****Laboratoire d'Analyse et Contrôle des Systèmes Complexes
ECE Paris École d'Ingénieur
37 Quai de Grenelle – 75015 Paris
ravaut@ece.fr*

RÉSUMÉ. Dans un contexte de formation aux pratiques avancées infirmières, nous décrivons dans cet article un outil de formation qui exploite en temps réel une modélisation de connaissances de santé à l'aide d'ontologies. Nous nous attachons dans un premier temps à montrer comment ces ontologies peuvent s'intégrer dans la réalisation d'un outil de formation de type serious game. Enfin, nous présentons un outil décisionnel permettant d'établir un diagnostic à l'aide d'une ontologie.

MOTS-CLÉS : ontologie, serious game, Web sémantique, modélisation des connaissances, pratiques avancées de santé, formation, application médicale.

1. Contexte : un besoin de formation en pratiques avancées

Depuis 2006, en France, le ministère de la santé et des groupes de professionnels de santé dont l'Union Nationale Des Associations d'Infirmiers/Infirmières De Bloc Opératoire Diplômé(e)s d'État (UNAIBODE), travaillent sur la réingénierie des diplômes paramédicaux. Cette restructuration des diplômes, et plus précisément de la formation, va permettre aux infirmiers de s'intégrer dans le parcours universitaire type LMD (Licence-Master-Doctorat).

En Novembre 2009, l'École des Hautes Études en Santé Publique (EHESP) et l'Université de Marseille Méditerranée proposent une telle formation via le premier Master de pratiques avancées en soins infirmiers pour répondre aux évolutions de la profession infirmière conséquentes à la mise en place de la loi HPST (Hôpital, Patient, Santé et Territoires). [WEB 1]

Dans son domaine d'activité spécifique, l'infirmière de pratique avancée aborde les situations de soins des patients et de leur entourage sur une perspective infirmière :

- elle dispense des soins infirmiers experts dans trois domaines spécifiques : oncologie, gériatrie et coordination des parcours complexes de soins.

- elle est en outre habilitée à dispenser des soins dans le domaine du diagnostic et de la prescription selon des modalités de supervision médicale (protocole, prescription anticipée...) fixées par la réglementation encadrant son exercice professionnel.

- elle inscrit son action en complémentarité avec celle des autres professionnels de santé et du secteur social composant l'équipe interdisciplinaire. [WEB 2]

Les personnels infirmiers en pratiques avancées interviennent dans des lieux de soins variés qui vont de la chambre d'hôpital ou de maison de retraite à la chambre de patient à domicile.

Ils font face à des situations impliquant une autonomie accrue par rapport aux situations classiques des personnels soignants : prises de décisions, organisation du suivi des patients à distance. Dans ce cadre, ces personnels seront amenés à utiliser les nouvelles technologies pour améliorer le confort du patient, son information et son suivi à distance.

Ceci nécessite un important effort de formation.

Pour parfaire cette formation, et en partenariat avec l'EHESP, il est envisagé de fournir aux étudiants infirmiers une application de type jeu sérieux (*serious game*) pour laquelle un personnel médical se retrouve en situation de devoir prendre une décision qui engage l'état d'un patient virtuel dont les fonctions vitales seront « monitorées » en temps réel.

Dans ce contexte, concevoir un outil de monitoring intelligent permettrait d'impliquer les étudiants à la surveillance des données physiologiques d'un patient virtuel et, à terme, leur permettre de s'exercer sur les pratiques avancées de santé.

Le premier socle nécessaire à un tel outil de monitoring intelligent se construit autour de la modélisation des connaissances de santé nécessaires à l'étudiant en pratiques avancées dans un but de formation. On entend par intelligent la capacité envisagée de l'outil à interagir avec les données pertinentes de son environnement ainsi qu'avec les informations fournies par des intervenants médicaux en temps réel.

Sensibilisé au besoin de modéliser ses connaissances et son vocabulaire de façon formelle et explicite, l'univers médical expérimente l'utilisation d'ontologies pour modéliser des connaissances de santé depuis un certain temps à l'image des ressources ontologiques de l'*Unified*

Medical Language System (UMLS) [PIS 98]. Cet article se focalise sur les travaux préliminaires de construction d'une ontologie à même de répondre à de futures exigences en vue de son intégration dans un outil de formation de type *serious game* ; puis sur la création d'un outil décisionnel de type application Web permettant d'expérimenter et d'envisager son utilisation.

Le chapitre 2 décrit les raisons qui nous ont conduit à utiliser les ontologies comme modèle de représentation de l'information. Les chapitres 3 et 4 s'attardent sur nos choix techniques. Enfin, le chapitre 5 conclue l'article.

2. Les ontologies informatiques, technologie du Web sémantique

Plusieurs façons de modéliser la connaissance en santé nous étaient possibles. Les deux plus répandues étant la méthode hiérarchique (par exemple le langage XML) ou la méthode des bases de données relationnelles (par exemple MySQL, MS SQL).

Une troisième façon de modéliser de l'information étant les graphes de données.

Dans l'utilisation des deux méthodes classiques de représentation de l'information, il est admis comme concept que certains éléments de données ont plus d'importance, de prépondérance que les autres. Par exemple, un document XML est construit en nœuds d'informations, chaque nœud possédant des nœuds d'informations parents. À la racine du document, ou plus haut niveau de hiérarchie, se trouve le nœud *root* qui ne possède pas de parents.

Dans une représentation en graphe de données, il n'y a pas de concept de hiérarchie ou de prépondérance d'un élément, un graphe consistant en plusieurs ressources reliées entre elles, sans ressource ayant une importance intrinsèque sur une autre. Cette méthode de représentation a été considérée comme base de travail pour le développement du Web sémantique.

Les sous-sections à venir présentent le langage de description des ontologies ainsi que les avantages à utiliser les ontologies dans le contexte présenté au chapitre 1.

2.1. La représentation d'un graphe de données à l'aide du RDF

Pour représenter des connaissances de type graphe de données on utilise le langage *Resource Description Framework (RDF)* [WEB 3] sous la forme de déclarations (ou *statemenst*) RDF.

Une déclaration RDF est représentée par ce qu'on appelle un triplet. Un triplet est constitué d'un sujet, d'un prédicat (aussi appelé propriété) et d'un objet.

Pour modéliser une connaissance telle que « le pull est rouge », on peut utiliser le triplet suivant :

- sujet : le pull
- prédicat : est de couleur
- objet : la couleur rouge

Nous détaillerons le langage RDF plus spécifiquement au chapitre 3.

Les ontologies sont construites sur des documents RDF et utilisent ainsi une représentation de l'information en graphes de données, un des fondements de base du Web sémantique.

2.2. L'intérêt de l'utilisation des ontologies

L'intérêt théorique de l'utilisation d'ontologies, spécifications formelles explicites de termes d'un domaine et de relations entre elles [GRU 93], est multiple :

- partager une compréhension commune sur la structure d'informations liées à l'apprentissage entre les utilisateurs et les machines (par exemple entre softwares) [KAN 06].
- permettre la réutilisation de domaines de connaissances existants.

À ces intérêts théoriques nous pouvons ajouter ou en préciser certains dans un contexte de modélisation de connaissance médicale et de son intégration dans un outil de formation, tels que :

- la nécessité d'établir un vocabulaire commun sur des notions de santé admises par tous les acteurs du monde de la santé.
- l'évolution constante des normes de santé implique la possibilité de mettre à jour de façon aisée notre base de connaissances.
- la nécessité de séparer le contenu médical du futur outil de formation *serious game* pour faciliter la mise à jour des données en garantissant l'intégrité des deux systèmes alors indépendants.
- la possibilité d'inclure au sein de l'ontologie des notions d'intelligence artificielle comme dans un moteur de jeu classique. Ceci à l'aide des diverses possibilités offertes par l'extension *Web Ontology Language* (OWL) [SIR 07] du langage RDF utilisé pour décrire notre ontologie.
- la convention de représentation des données dans une ontologie permet de faciliter l'agrégation de données multi-sources (capteurs, personnels de santé, textes) et multi-formats. On peut ainsi unifier la représentation de l'information pour définir ensuite des raisonnements adaptés au contexte d'exécution d'un outil de formation.

Bien sûr, certaines de nos exigences peuvent être respectées en utilisant une modélisation des connaissances de type base de données par exemple, mais l'ontologie en OWL intègre nativement des fonctionnalités telles que la création de classes, d'instances (ou individus) et l'ajout de propriétés entre ses différents concepts (cf. chapitre 3). Intégrer de telles fonctionnalités dans une base de données relationnelle semble demander un effort d'intégration et d'adaptation de la technologie plus conséquent. Par ailleurs, la modélisation de données offerte par le langage OWL se rapproche d'une organisation de type programmation orientée objet utilisée naturellement dans la conception de jeux vidéos [SAL 04], nous avons donc jugé utile de faciliter la future intégration de notre base de connaissance médicale dans un moteur de jeu par le biais d'une ontologie.

2.3. Les informations de santé à modéliser : de la maladie à la procédure de soin en passant par les symptômes et les données physiologiques

Dans un cheminement de pensée décisionnel classique utilisé par les personnels de santé, l'étude des symptômes et les observations et analyses qui en découlent conduisent au diagnostic qui s'accompagne d'une procédure de soin adaptée. Ce cheminement se rapproche du modèle SOAP (*Subjective, Objective, Assessment, Plan*) utilisé par certains médecins pour décrire dans une fiche informative l'état de santé complet d'un patient [WEE 69].

Dans un tel modèle :

- les nœuds S décrivent les symptômes et signes cliniques du patient.
- les nœuds O décrivent les observations et analyses du médecin.
- les nœuds A correspondent aux maladies et problèmes de santé du patient.
- les nœuds P correspondent aux procédures de soins ou plans de santé mis en place pour résoudre les problèmes de santé du patient.

A la réutilisation d'une ontologie existante dans le domaine de la santé, il a été préféré la création d'une ontologie à partir d'un arbre décisionnel représentant au mieux le cheminement logique depuis un ou plusieurs symptômes jusqu'aux diagnostics et procédures de soin. Cette démarche permet, à travers un modèle informatique, la représentation d'une prise de décision formalisée faisant partie intégrante d'une formation étudiante à des pratiques de santé, l'établissement d'un diagnostic dans ce cas. La possibilité de personnaliser les propriétés contenues dans cette ontologie nous autorise également l'inclusion de simples notions de logique en amont du développement d'un moteur d'intelligence artificielle propre au jeu vidéo de formation, ce qui nous conforte dans le choix d'une ontologie pour modéliser les connaissances santé.

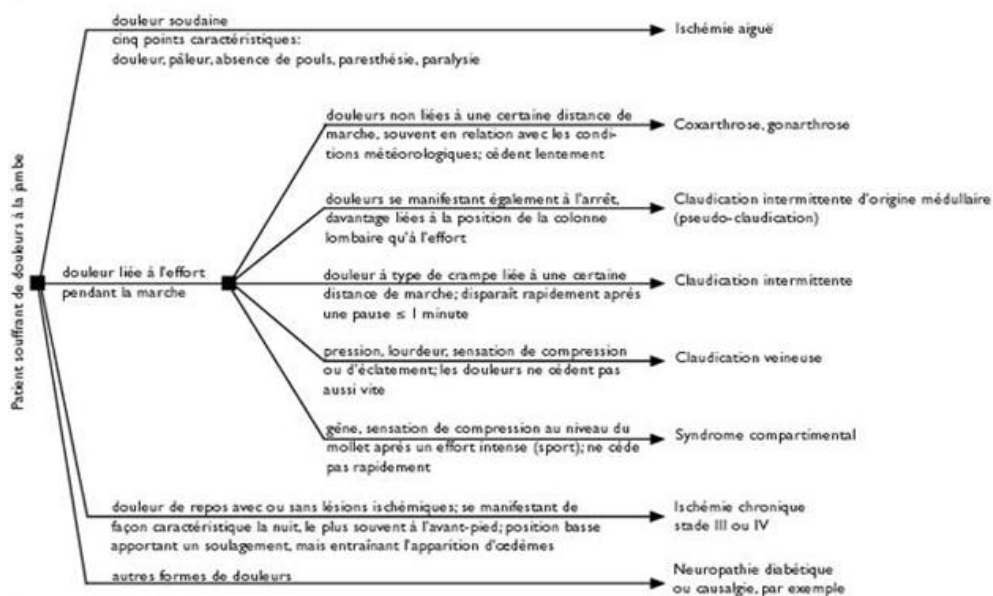


Figure 1. Arbre logique du cheminement diagnostique dans le cas d'une douleur à la jambe.

Pour les besoins de l'exercice, seuls ont été conservés les symptômes et maladies associées (soit les parties S et A du modèle SOAP). Ce choix a été réalisé en vue de l'automatisation dans un outil applicatif du cheminement logique d'une pensée diagnostique qui, partant d'un symptôme sélectionné, aboutit au diagnostic d'une maladie grâce à l'ontologie.

3. La construction de l'ontologie à l'aide du langage OWL

Le langage OWL (*Web Ontology Language*) est une technologie du Web sémantique [BER 01] et une application du langage RDF, recommandé par le W3C [WEB 4] pour décrire les ressources du Web sémantique. Les ontologies OWL sont encapsulées dans des déclarations RDF et utilisent les identificateurs RDF (*rdf:id*) et les propriétés RDF associées (*rdf:datatype*).

Les éléments contenus dans les *namespace* OWL (*owl:*) identifient les composantes logiques de l'ontologie. Ces composantes logiques, ou notions de logique, se présentent sous la forme de relations binaires entre concepts (les classes de l'ontologie) et entre instances desdits concepts (les individus) à l'aide de propriétés. Ces propriétés sont multiples, des *Object* et *Datatype Properties* aux propriétés fonctionnelles, inverses, transitives...

La notion de hiérarchie entre les concepts de l'ontologie, à l'image de la hiérarchie des classes en programmation orientée objet, permet d'ajouter une composante d'intelligence artificielle supplémentaire nativement à l'ontologie.

Chaque individu possède un identifiant unique en tant que ressource du Web sémantique. Il est possible d'associer à l'individu des labels ou des commentaires, par exemple une définition, par le biais du *rdfs:comment*. Il est également possible de spécifier qu'un individu est en tout point identique à un autre (cela se révèle utile lorsque plusieurs orthographes existent pour un même mot, ou si l'on veut désigner le même individu dans deux langues différentes).

Pour définir les relations entre concepts ou individus, OWL permet la définition de propriétés (ou relations) dont les deux principales [HOR 09] sont :

- les *Object Properties* qui relient deux individus entre eux
- les *Datatype Properties* qui lient un individu à une valeur ou à une portée de valeurs numériques ou une chaîne de caractères par exemple.

```
<owl:NamedIndividual rdf:about="http://localhost/diagnostic#Gene_au_niveau_du_mollet.">  
  <rdf:type rdf:resource="http://localhost/diagnostic#Douleur_durant_un_effort"/>  
  <hasPathologie rdf:resource="http://localhost/diagnostic#Syndrome_compartimental."/>  
</owl:NamedIndividual>
```

Figure 2. L'individu « Gène au niveau du mollet » membre de la classe « Douleur durant un effort » est associé à la pathologie « Syndrome compartimental » par la propriété « hasPathologie ».

3.1. Protégé et le module d'extension OWL

Protégé [WEB 5] est un environnement de développement d'ontologies open-source dont les fonctionnalités sont les suivantes : édition des classes, des propriétés et des individus de l'ontologie. Protégé intègre le module d'extension (ou plug-in) OWL et plusieurs standards du Web sémantique depuis 2003, et offre l'avantage d'être supporté par une forte communauté de développeurs, d'universitaires ou d'utilisateurs du monde professionnel (plus de 170 000

utilisateurs au début 2011) et de proposer une interface utilisateur intuitive. D'où son choix pour modéliser notre première ontologie.

Le plug-in OWL de Protégé permet l'inclusion des composants du langage OWL au sein de l'ontologie.

Ce plug-in fournit une interface utilisateur avec plusieurs onglets :

- l'onglet *OWL Classes* affiche la hiérarchie des classes de l'ontologie et permet au développeur de créer et d'éditer des classes.
- l'onglet *Properties* est utilisé pour créer et éditer les propriétés de l'ontologie.
- l'onglet *Individuals* est utilisé pour créer et éditer les individus de l'ontologie.
- un onglet *OWL Viz* qui permet de visualiser l'ontologie sous forme de graphe.

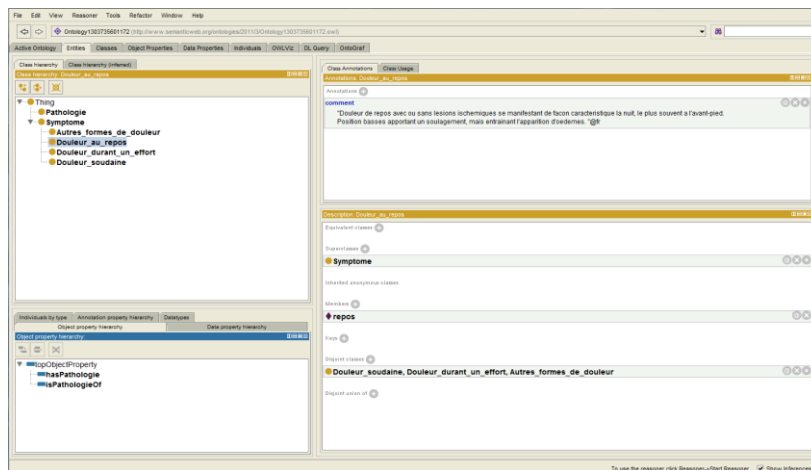


Figure 3. Interface de Protégé avec le plug-in OWL intégré.

3.1.1. Différences de modélisation entre Protégé-OWL et la programmation orientée objet

À la différence de la programmation orientée objet (POO) dans laquelle un individu (ou instance) est directement associé à sa classe mère par une relation d'appartenance (ou héritage) lors de son instanciation, Protégé-OWL ne fonctionne pas de la même façon pour lier individus et classes d'appartenance.

Protégé-OWL utilise un système de restriction sur les classes. Les restrictions sont des propriétés particulières des classes qui expriment de façon explicite les conditions nécessaires pour qu'un individu devienne membre d'une classe.

Ce concept propre à Protégé-OWL pourrait dérouter dans un premier temps les programmeurs issus du monde de la POO.

Protégé-OWL dispose ainsi d'un moteur d'inférence qui permet d'associer de façon automatique les individus à leurs classes une fois les restrictions de classes explicitées.

Bien qu'utilisant des termes similaires, classes et instances sous Protégé-OWL différent du mode de fonctionnement des classes et instances en POO.

La POO sera principalement centrée sur les méthodes des classes, le programmeur réalisant les décisions conceptuelles en s'appuyant sur les propriétés opérationnelles d'une classe, alors que le concepteur d'ontologie s'appuiera sur les propriétés structurelles des classes.

3.2. Interroger l'ontologie : utilisation du protocole de requête SPARQL

Une fois l'ontologie construite et sauvegardée dans un fichier OWL, il faut utiliser un protocole de requête spécifique au langage RDF construit sur des graphes de données (à l'image du protocole SQL pour interroger une base de données).

Le SPARQL (prononcer « *sparkle* », acronyme de *SPARQL Protocol And RDF Query Language*), recommandé depuis le 15 Janvier 2008 par le W3C [w3 08], définit la syntaxe et la sémantique nécessaire à l'expression de requêtes sur une base de connaissances RDF. SPARQL est adapté à la structure spécifique des graphes de données RDF, et s'appuie sur les triplets qui les constituent. SPARQL permet d'exprimer des requêtes de type interrogatives ou constructives.

```
$query = ' PREFIX ontologie: <http://localhost/diagnostic/>
SELECT ?x
WHERE { ?x rdfs:subClassOf ontologie:Symptome } ; '
```

Figure 4. Requête SPARQL retournant l'ensemble des sous classes de type « *Symptôme* ».

4. Une application pour exploiter l'ontologie

Pour exploiter une ontologie et effectuer des requêtes SPARQL plusieurs choix étaient possibles, ceux-ci allant du moteur de recherche sémantique (*SPARQLer* [WEB 6]), aux applications de type *stand-alone* (*CORESE* [WEB 7]), ainsi qu'à la possibilité de créer des applications plus complexes en C# ou Java (*Jena Semantic Framework* [WEB 8]).

Notre choix s'est porté sur un *framework PHP* nommé *ARC2* [WEB 9].

Ce choix nous semblait être celui apportant le plus de souplesse dans la conception d'une application Web selon nos critères : créer dans sa totalité un outil de diagnostic automatisé pour visualiser les résultats de nos requêtes et également réaliser cette expérience dans un environnement Web en vue de futures problématiques liées à l'intervention de capteurs médicaux connectés sur réseau. En gardant à l'esprit la nécessaire intégration de mécanismes propres aux jeux vidéo, nous souhaitions que la chaîne de prise de décision soit automatisée pour cette première approche sur l'utilisation d'ontologies. C'est pourquoi cette première ontologie a été construite en s'inspirant d'un cas médical décrit sous la forme d'un arbre décisionnel (Fig. 1.).

L'application Web se présente sous la forme d'un formulaire dans lequel l'utilisateur final peut sélectionner un symptôme. Pour obtenir un diagnostic lié au choix du symptôme, l'application se

charge ensuite d'envoyer une requête sur l'ontologie pour obtenir la pathologie associée au symptôme.

The figure consists of two side-by-side screenshots of a web application interface titled 'Outil d'aide à la prise de décision' and 'Utilisation d'ontologies comme base de connaissances'. The interface features the Orange logo in the top right corner.

Left Screenshot (Form):

- Field: 'Entrer un nom ou identifiant pour le patient :'
- Section: 'Sélectionnez le symptôme dont souffre le patient :'
- Options:
 - Autres formes de douleur
 - Douleur au repos
 - Douleur durant un effort
 - Douleur soudaine
- Text: 'Commentaires sur le symptôme : Douleur de repos avec ou sans lésions ischémiques se manifestant de façon caractéristique la nuit, le plus souvent à l'avant-pied. Position basses apportant un soulagement, mais entraînant l'apparition d'œdèmes.'
- Section: 'Préciser le symptôme :'
- Dropdown menu: 'Douleur de type crampe liée à une certaine distance de marche' (selected)
- Other options in dropdown: 'Douleur de type crampe liée à une certaine distance de marche', 'Gêne au niveau du mollet', 'Sensation de compression'
- Buttons: 'Lancer le diagnostic', 'Effacer les données'

Right Screenshot (Result):

- Text: 'L'ontologie a établi le diagnostic suivant : Le patient souffre de la pathologie **Syndrome compartimental**.'
- Section: 'Récapitulatif des données transmises :'
- Table:

Nom du patient	Symptôme sélectionné	Précisions éventuelles
john doe	Douleur au repos	Gêne au niveau du mollet
- Button: 'Lancer un nouveau diagnostic'

Figure 5. Écran de sélection des symptômes à travers un formulaire généré par l'ontologie puis affichage du diagnostic.

5. Conclusion et travaux futurs

Cette première approche nous a permis d'envisager l'utilisation des ontologies en tant que modèle de représentation des connaissances médicales. Les notions dédiées à la formation d'infirmières en pratiques avancées ont des contraintes d'actualisation fréquentes et nécessitent d'être exprimées dans un vocabulaire commun interopérable entre utilisateurs et machines. Ce pourquoi les ontologies sont très adaptées.

Cependant, l'outil présenté dans cet article est un outil de diagnostic et non de formation.

Nous envisageons par la suite de créer une base de connaissances en partenariat avec l'EHESP qui fournira :

- les connaissances physiologiques sous-jacentes à la prise de décision dans l'exercice infirmier en pratiques avancées,
- un contenu pertinent de formation.

A l'aide de mécanismes propres aux moteurs de jeux vidéo, combinés avec l'utilisation d'ontologies, nous prévoyons d'utiliser notre base de connaissance pour faire évoluer en temps réel les scénarii et scripts du jeu de formation de type *serious game*.

Les symptômes d'un patient virtuel entraîneront la mise en pratique des connaissances médicales et infirmières. Celle la même entraînera l'apparition de symptômes différents, en respect des connaissances physiologiques, dans un univers de jeu libre où chacune des actions de l'apprenant sera prise en compte et influera sur l'environnement de jeu à l'aide de la base de connaissances ontologique.

Cette tentative dans un environnement web nous a apporté les résultats attendus en termes de rapidité d'exécution, élément indispensable au bon déroulement d'un jeu vidéo temps réel.

La création d'ontologies insiste également sur l'intégration, en tant que fondement sémantique, de relations entre ses propres concepts, ce qui demeure un atout considérable pour une intégration dans un moteur d'intelligence artificielle de jeu plus avancé.

Nous garderons à l'esprit que le sujet présenté dans cet article est à la croisée de deux univers, l'informatique et la santé, ce qui constitue la principale difficulté du travail.

Bibliographie

- [BER 01] BERNERS-LEE T., HENDLER J., LASSILA O., « The Semantic Web. », Scientific American Magazine, 17 Mai 2001.
- [GRU 93] GRUBER T.R., « A translation approach to portable ontology specification. », Knowledge Acquisition 5, 1993, p. 199-220.
- [HOR 09] HORRIDGE M., DRUMMOND N., JUPP S., MOULTON G., STEVENS R., « A practical guide to building OWL ontologies using Protégé 4 and CO-ODE tools Edition 1.2. », the university of Manchester, 13 Mars 2009.
- [KAN 06] KANELLOPOULOS D., KOTSIANTIS S., PINTELAS P., « Ontology-based learning applications: a development methodology. », 24th IASTED Software Engineering, Février 2006.
- [PIS 98] PISANELLI DM., GANGEMI A., STEVE G., « An ontological analysis of the UMLS Methathesaurus. », AMIA Symp., 1998, p.810-814.
- [SAL 04] SALEN K., ZIMMERMAN E., « Rules of Play: Game Design Fundamentals. », MIT Press, 2004.
- [SIR 07] SIRIN E., PARSIA B., GRAU B.C., KALYANPUR A., KATZ Y., « A practical OWL-DL reasoner. », Journal of Web Semantics 5(2), 2007.
- [WEE 69] WEED L.L., « Medical records, medical education, and patient care. The problem-oriented record as a basic tool. », OH: Case Western Reserve University, 1969.
- [W3 08] W3.ORG, « W3C Semantic Web Activity News - SPARQL is a Recommendation », Janvier 2008.
- [WEB 1] <http://www.sante.gouv.fr/la-loi-hopital-patients-sante-et-territoires.html>
- [WEB 2] <http://www.ehesp.fr/wp-content/uploads/2009/11/dossier-master-sci.pdf>
- [WEB 3] <http://www.w3.org/TR/PR-rdf-syntax/>
- [WEB 4] <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- [WEB 5] http://protege.stanford.edu/doc/users_guide/index.html
- [WEB 6] <http://www.sparql.org/query.html>
- [WEB 7] <http://www-sop.inria.fr/edelweiss/software/corese/>
- [WEB 8] <http://jena.sourceforge.net/>
- [WEB 9] <https://github.com/semsol/arc2/wiki>