

Méthodologie de construction d'ontologie

Semantic Web Primer, chap 7

Semantic Web for the working ontologist, chap 12

Guy Lapalme

Université de Montréal

lapalme@iro.umontreal.ca

Types de construction

- AAA Slogan
 - Anybody can say Anything about Any topic
 - d'où le besoin d'une certaine méthodologie
- Manuelle
- Réutilisation d'ontologies existantes
- Semi-automatique

Manuelle - I a

Déterminer la portée

Une ontologie

- pas un but en soi
- *modèle ou abstraction* d'un domaine
 - pas d'ontologie *correcte*
 - alternatives multiples
- fonction de l'utilisation prévue
 - domaine ?
 - quels types de questions ?
 - utilisateur et *mainteneur* ?
- considérer la réutilisation

Manuelle Ib

- préciser les questions auxquelles le modèle devra répondre
- prévoir l'intégration
 - de nouvelles informations
 - du modèle dans un cadre plus général
- imaginer comment se passera le processus d'inférence

Manuelle -2

Énumérer les termes

- Énumération des termes du domaine
 - noms : noms de classe
 - verbes : noms des propriétés
- Exemple des accords mets-vins
 - vin, cépage, établissement vinicole, rouge, blanc poisson, viande rouge
 - être situé à, avoir comme couleur, contenir comme sucre
- Considérer les schémas de BD, vocabulaires, thesaurus déjà présents dans l'entreprise

Manuelle -3

Définir la taxonomie

- Placer les termes les uns par rapport aux autres
 - top/down ou bottom/up (ou les deux)
- S'assurer que c'est bien une hiérarchie
 - si A est sous-classe de B,
toutes les instances de A sont instances de B
 - *Pinot Noir* est sous-classe de *Vin Rouge*

Manuelle - 4

Définir les propriétés

- souvent fait en même temps que la définition des classes
- associer les propriétés à la plus haute classe à laquelle elle s'applique
- définir les domaines et portées
- tension entre généralité des domaines et portées et spécificité pour détecter les inconsistances

Manuelle - 5

Définir les facettes

- Cardinalité
au moins/plus 1
- Domaine, portée, type
- Valeurs requises
- Caractéristiques relationnelles
symétrie, transitivité, inverse, fonctionnelle
- On peut maintenant vérifier la consistance de l'ontologie

Manuelle - 5

Définir les instances

- Le but d'une ontologie est d'organiser les instances
- Le nombre d'instances est beaucoup plus grand que le nombre de classes
- Souvent créées automatiquement
Base de données ou extraites d'un corpus
- Vérification d'anomalies

Modélisation pour la réutilisation

- Une ontologie sera *aussi* lue par des humains
- Trouver des noms évocateurs pour les classes
- Ajouter des `rdfs:label`, `rdfs:comment`, `rdfs:seeAlso`
- Conventions habituelles (ie celles du W3C)
 - ressources en CamelCase
 - nom d'une classe est au singulier et débute par une majuscule
 - nom d'une propriété débute par une minuscule

Garder trace des classes et des individus

- Les instances des uns sont les classes des autres
 - espèces d'animal vs les animaux
 - lignes d'affaires vs les entreprises
- une classe devrait pouvoir avoir des instances
- les valeurs précises des propriétés devraient s'appliquer aux instances

Tester la modélisation

- Peut-on répondre aux questions identifiées au début ?
- Comment se comporte l'ontologie en présence de sources multiples ?
- Quelles inférences peut-on tirer ?

Réutilisation d'ontologies

- Codification par des experts
 - Cancer Ontology du National Cancer Institute
 - Art and Architecture Thesaurus
 - Thesaurus of Geographic Names
- Vocabulaires intégrés
 - Unified Medical Language System
(750 000 concepts, 10M de liens)
- Upper level ontologies (gros bon sens...)
 - Cyc, Standard Upperlevel Ontology

Réutilisation d'ontologies

- Hiérarchies de sujets
 - Open directory project
- Ressources linguistiques
 - WordNet
- Librairies d'ontologies
 - DAML (DARPA Agent Markup Language)
 - Protégé
 - Swoogle

Problème d'acquisition de connaissances

- Création manuelle d'ontologie est longue, coûteuse, très spécialisée et parfois fastidieuse
- Méthodes d'apprentissage pour
 - acquisition et extraction des connaissances
 - maintenance et révision des connaissances

Tâches envisageables par apprentissage

- Extraction d'ontologies
- Extraction de données et métadonnées
- Fusion et mise en correspondance d'ontologies
- Maintenance d'ontologies par l'analyse des instances
- Observation des usagers

Techniques d'apprentissage

- Clustering
- Mise à jour incrémentale d'ontologie
- Support pour *l'ingénieur des connaissances*
- Amélioration d'ontologies pour la langue naturelle
- Apprentissage d'ontologie

Conclusion

- Création d'ontologies reste un *art*
- Pas encore de recette miracle
- Exemples de création
 - Ontology Development 101: A Guide to Creating Your First Ontology
 - Développement d'une ontologie 101