

DIC9305 Logique, informatique et sciences cognitives

Connaissance, sémantique et inférence

Roger Villemaire

Département d'informatique
UQAM

le 15 février 2024



© 2021-2024 Roger Villemaire, villemaire.roger@uqam.ca
Creative Commons Paternité - Pas d'Utilisation Commerciale - Pas de Modification 3.0 non transcrit.

Plan

- 1 Données et Connaissances
- 2 Ontologies
- 3 Web Sémantique
- 4 Normes du W3C
- 5 Logique de Description

Données

- D'énormes quantités de données sont aujourd'hui stockées, gérées et intégrées sur des supports informatiques,
- et ceci dans tous les domaines : affaires, sciences de la vie, édition, éducation, information, etc.
- On voit donc le développement interdisciplinaire d'une *Science des Données* pour profiter au maximum de cette énorme quantité de connaissances à l'aide d'outils et de méthodes informatiques.

Données ouvertes

- Bien qu'on souligne depuis longtemps que la *société de l'information (ou de la connaissance)* a succédé à la société industrielle et que l'informatique joue un rôle central dans les entreprises et dans nos vies, ce qui émerge aujourd'hui est le désir d'intégrer le plus largement possible des sources d'information des plus hétérogènes.
- Il s'agit donc de passer d'un
 - modèle *fermé* où les données sont structurées en vue d'une application bien précise, à un
 - modèle *ouvert* où les données seront accessibles à une application quelconque.

Partage

- Mais pour partager les données et rendre l'information accessible, interopérable, et réutilisable, il est nécessaire d'établir
 - des formats normalisés,
 - des primitives à la sémantique claire,
 - et des méthodes algorithmiques efficaces pour exploiter l'information ainsi représentée.
- Pour atteindre ces objectifs l'approche ontologique est aujourd'hui dominante.

Ontologie

- En philosophie, l'ontologie est l'étude de *l'être* et réfère à la nature de la réalité et des choses (être vivants, objets, etc.) ainsi qu'à la nature de leurs relations.
- En informatique, *une* ontologie est un vocabulaire contrôlé qui permet de décrire les propriétés fondamentales des entités d'un domaine, ainsi que leurs relations.

Ontologie (informatique)

- Il s'agit d'organiser l'information pour en faciliter le partage et l'usage à l'aide d'outils informatiques.
- Il est donc nécessaire d'établir et d'utiliser des formats normalisés pour les données.
- On veut aussi favoriser l'interopérabilité et la réutilisation de l'information.
- Il est donc primordial d'éviter le développement de terminologies différentes, qui pourraient s'avérer difficiles à mettre en correspondance.

Modélisation

- Une ontologie est donc une *modélisation formelle* qui permet
 - de décrire précisément les concepts et relations d'un domaine,
 - d'en inférer des conséquences,
 - et, plus généralement, de raisonner pour aussi déterminer les incohérences et limitations du modèle ainsi construit.
- C'est donc tout naturellement que la logique, en tant que science du raisonnement, y joue un rôle de premier plan.

- Timothy John Berners-Lee a établi au début des années 1990, alors qu'il travaillait au CERN (Conseil européen pour la recherche nucléaire), les fondations du WWW (World Wide Web),
 - un réseau (web) de documents (hyper)reliés (*linked*).
- Ceci a permis la prodigieuse dissémination d'information, dont nous profitons tous les jours.
- Il a aussi fondé, en 1994, le W3C (World Wide Web Consortium) qui gère les normes qui assurent l'interopérabilité des outils du Web.

Web Sémantique

- C'est aussi au milieu des années 1990 que Berners-Lee attire l'attention sur le fait que partager des documents n'est pas suffisant pour un véritable partage de l'information et son utilisation par la machine. Il lance donc le mouvement du *Web Sémantique* pour créer
 - un réseau d'information (hyper)reliées (*web of information*).
- Il ne s'agit donc plus seulement de disséminer des documents, mais de partager de l'information dont la sémantique (sens) peut être exploitée par la machine.

Objectifs

- Les objectifs du Web Sémantique sont de permettre
 - de lier des informations de sources disparates, par ex. hôtel, restaurant, places de stationnement,
 - de partager et intégrer de l'information dans un même domaine, par ex. bibliothèques, données biologiques, etc.

URI

- Le Web Sémantique nécessite tout d'abord une façon de déterminer l'identité d'une ressource.
- Ceci s'effectue à l'aide d'un *Uniform Resource Identifier* (URI) qui assure une identification globale, à travers tout le Web.

URI et URL

- Le format d'une URI est normalisé par l'Internet Engineering Task Force (IETF) dans le RFC 3986 (Request for Comments)

`scheme:[//[user:password@]host[:port]][/]path[?query][#fragment]`

- Par exemple l'URI de l'UQAM sur DBpedia est `http://dbpedia.org/resource/UQAM`
- Une URI a la même forme qu'une URL (*Uniform Resource Locator*) et si elle peut être déréférencé par un navigateur Web, elle mènera normalement à une page explicative.
 - C'est une méthode importante de documentation.
 - De plus, le format de la donnée peut être souvent choisi par une négociation HTML, ce qui permet aussi de répondre aux besoins algorithmiques.

CURIE

- Pour simplifier la présentation des URI ont utilise en général des CURIE (*Compact URI*) :
 - définition du préfixe, par exemple db pour `http://dbpedia.org/resource/`
 - la CURIE [db:UQAM] représente alors `http://dbpedia.org/resource/UQAM`

- Pour le *Resource Description Framework* (RDF), la pierre angulaire du Web Sémantique, ce qui compte pour ce cours est que :
 - l'information est représentée par des *triplets* formés d'un *sujet*, d'un *prédicat*, et d'un *objet*.
 - par exemple (*Roger*, *suit*, *DIC9305*).
 - Il s'agit donc d'une représentation par un graphe, dont les sommets sont des entités (sujets et objets), et les arêtes les couples des relations (prédicats).



Graphe de Connaissance

- En relation avec la façon dont la connaissance est représentée dans le cadre du Web Sémantique on parle donc
 - de *triplestore* ou de *RDF store*, mais aussi
 - de Graphe de connaissance (*Knowledge Graph*).
- De plus, l'information arrive au fur et à mesure (*monde ouvert*) et on découvre donc la connaissance progressivement, lorsqu'elle se présente.

Applications

- Les technologies du Web Sémantique sont appliquées dans :
 - portails web alimentés par des triplets RDF (*RDF-backed web portals*,
 - DBpedia, un effort collectif pour extraire l'information des projets de la fondation Wikimedia,
 - Wikidata, qui est plutôt un effort collectif de partage d'information,
 - Schema.org une initiative conjointe, démarrée par Google, Microsoft, Yahoo, et Yandex pour structurer les données de pages Web et d'autres applications de l'Internet,
 - le partage de données dans des formats comme RDF par des gouvernements, des villes etc.,
 - Open Graph Protocol de Facebook pour intégrer des informations extérieures dans les réseaux sociaux.

Exemple

- L'Open Graph protocol permet l'ajout à des pages HTML d'informations en format RDFa.
- Ceci est entre autre utilisé par Facebook, Twitter, LinkedIn, Pinterest, Mattermost pour représenter correctement (*rich preview*) un lien partagé.

```
<meta property="og:locale" content="fr_CA" />
<meta property="og:type" content="website" />
<meta property="og:title" content="UQAM | Université
  du Québec à Montréal | Accueil" />
<meta property="og:description" content="L'Université
  du Québec à Montréal (UQAM) offre plus de 300 programmes d'étu
<meta property="og:url" content="https://uqam.ca/" />
<meta property="og:image" content="https://ressources-medias.
  uqam.ca/media/uploads/20200306151917/campuscentral.jpg" />
<meta property="og:image:width" content="1080" />
<meta property="og:image:height" content="763" />
```

- Décrire l'information n'est pas suffisant pour en inférer les conséquences.
- Il est nécessaire aussi d'ajouter les contraintes, comme celles entre les prédicats, indispensables à leur signification.
 - Par exemple, si la fourchette est à *gauche de* l'assiette, l'assiette est à *droite de* la fourchette !
- Ceci est le rôle du Web Ontology Language (OWL) qui est une représentation, dans les technologies du Web Sémantique, de la *Logique de Description*.

Logique de Description

- La logique de description permet de représenter
 - des *classes* regroupant des *individus* (ou instances),
 - des *rôles*, soit des relations binaires entre individus,
 - ainsi que des contraintes entre ces notions.
- De plus, l'inférence de connaissances implicites est en général réalisable par des algorithmes qui permettent donc un raisonnement automatique.

Concepts

- Un *concept* dénote un ensemble d'*individus* :
 - *Etudiant*, représente le concept “être un étudiant”,
 - *Prof*, représente le concept “être un professeur”,
 - *Cours*, représente le concept “être un cours”,

- Un *rôle* dénote une *relation binaire* entre individus :
 - *suit*, représente le rôle “... suit ...”,
 - *donne*, représente le rôle “... donne ...”,
 - *a_echoue*, représente le rôle “... a échoué ...”.

- L'outil Protégé permet de définir des ontologies OWL.
- À l'aide de *raisonneurs* comme `Hermit`, on peut y
 - vérifier si un concept est contradictoire,
 - calculer la hiérarchie *inférée*,
 - et en fait faire toute tâche expressible en logique de description.