

Séminaire rdatadev : 16/12/2016

Retour séminaire ISS2016 Informations Spatialisées Sémantisées (23- 25 Novembre 2016)

<http://devlog.cnrs.fr/iss2016>

V. Chaffard (LTHE)

Le web sémantique appliquée à des données à référence spatiale

Web sémantique: cf réseau SARI séminaire SARI Web Sémantique du nov 2015 + TP juin 2016

https://wiki.grenoble-inp.fr/sari/lib/exe/fetch.php?media=theme:web:linked_data_ws_irstea_2016.pdf

Web de données / Web sémantique :

- publier sur le Web des données accessibles par des machines (Web de données)
- + lier entre elles ces ressources en les structurant entre elles par rapport à 1 hiérarchie de classe et en donnant une sémantique aux relations utilisées (Web sémantique)

Objectif : -> Lier des données ensemble accroît leur valeur car va générer des nouvelles connaissances en suscitant des usages auxquels la personne qui les a produites n'a pas pensé

enjeux: décloisonnement des différentes sources de données hétérogènes et souvent séparées en silos de données

Information spatialisée et sémantique. Quelles convergences?

Historique et état de l'art lus au travers des modèles conceptuel, logique et physique des données et des traitements

Nathalie Abadie (IGN) . Pascal Auda (OMP)

Historique des SIG:

- > SIG autonome station de travail (ArcInfo ESRI 1982, GRASS 1983, QGIS 2002, ...)
- > SIG distribué: BDD spatial (PostGIS2001, spatiaLite 2008, ...) + OGC WMS/WFS (Mapserver 1994 , GeoServer, geoNetwork 2001,)

SIG distribué (architecture client / serveur, BDD centralisée) s'appuie sur des **standard OGC**:

- **Simple Feature Access** : modèle de classe pour décrire les données vecteur (Geometry <- Point / Curve / Surface / GeometryCollection) + méthodes (intersect, within, contains, ...)
- Implémentation: **Simple Feature Acces spatial SQL**: tables prédéfinies pour stocker la géométrie/topologie (GEOMETRY_COLUMNS, SPATIAL_REF_SYS, ...) + méthodes géométriques (ST_Point, ST_Polygon, ...) /topologiques (ST_intersect, ST_Within, ...) . Implémenté dans SGBD spatial (Ex: PostGIS)

-> SIG dispersé avec du Web sémantique spatialisée : ontologie spatio temporelle + langage de requete GeoSPARQL

Ontologie: décrit de façon formelle la sémantique de sources d'informations, ie la sémantique de termes de vocabulaires et des relations entre ces termes. Favorise ainsi leur interopérabilité sémantique.

GeoSPARQL: → Extension du langage SPARQL pour interroger des données géographiques sémantisées s'appuie sur ces standards

conclusion: Les modèles standards des SIG sont adoptés progressivement pour représenter et manipuler des objets géographiques sur le Web de données.

Etat de l'art du web sémantique et des données liées (linked data)

Philippe Genou et Danièle Ziebelin (LIG-Steamer)

WEB DE DONNEES: publier sur le Web des données accessibles par des machines

1/ Nommer les ressources par des URI HTTP (Uniform Resource Identifier) déréférencables. Ces URI doivent délivrer une représentation de la ressource sous différentes manières grace à de la négociation de contenu (entete HTTP Ex: Accept: text/html ou Accept: application/xml)

**2/ Décrire la ressource selon le standard RDF (Resource Description Framework).
Recommandation du W3C**

Pour déclarer des faits connus sur ces entités sous la forme d'un triplet:

Sujet (URI) / Prédicat ou relation (URI) / Objet (URI/Literal)

Ex: Georges Brassens / Was born in / Sète

Sujet URI :http://dbpedia.org/resource/Georges_Brassens / Relation URI:
<http://dbpedia.org/ontology/birthPlace> / Valeur URI: <http://dbpedia.org/resource/Sete>

Plusieurs formats de sérialisation: XML/RDF, N-Triples, **Turtle** (le + lisible pour 1 humain),
...

3/ Déployer des triple store et Requeter ces données : SPARQL

langage SPARQL : SQL pour le Web de données

W3C standard

langage fait pour requeter des triplets sujet/Predicat/objet via HTTP GET/POST/ou SOAP)
résultat en XML ou JSON

GeoSPARQL: → Extension du langage SPARQL pour interroger des données géographiques (standard OGC)

Un vocabulaire pour représenter des données géographiques conforme au modèle OGC

Simple Features Access:

- classes « Simple Feature Access »
- types de données gmlLiteral et wktLiteral pour sérialiser les géométries en GML (Geography Markup Language. XML schéma) ou WKT (Well Known Text. Ex: POINT(6 10), LINESTRING(3 4,10 50,20 25), POLYGON((1 1,5 1,5 5,1 5,1 1)))
- propriétés topologiques pour expliciter des relations spatiales entre features ou géométries

Des fonctions spatiales pour étendre SPARQL et permettre de calculer :

- des distances
- des buffers, des enveloppes convexes, des intersections, des contours,...
- des prédicats topologiques (Ex: Intersection entre polygones. matrice d'intersection DE9IM)

vocabularies/ontologies:

- geo: <http://www.opengis.net/ont/geosparql#>
- sf: <http://www.opengis.net/ont/sf#>
- gml: <http://www.opengis.net/ont/gml#>

Pour découvrir ces ontologies: <https://prefix.cc/>

Ex: Sparql EndPoint LinkedGeoData:

effort pour apporter dimension spatiale aux web de données/ web sémantique . s'appuie sur info d'OpenstreetMap

<http://linkedgeodata.org/OnlineAccess/SparqlEndpoints>

Requete: trouver tous les resto qui sont à moins de 1 km du théâtre de la mer à Sète:

Prefix rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>

Prefix ogc: <<http://www.opengis.net/ont/geosparql#>>

Prefix geom: <<http://geovocab.org/geometry#>>

Prefix lgdo: <<http://linkedgeodata.org/ontology/>>

Select ?s, ?l From<<http://linkedgeodata.org>>

Where

{

?s a lgdo:Restaurant;

rdfs:label?l ;

geom:geometry[

ogc:asWKT?g

].

Filter(bif:st_intersects(?g, bif:st_point(3.692764, 43.393794), 1)) .

}

WEB SEMANTIQUE:

4/ lier entre elles ces ressources en :

4.1/ les structurant entre elles par rapport à 1 hiérarchie de classe

W3C standard:

- RDFs (RDFschéma) : Modèle de classe entre concept . définit des classes, leurs propriétés et organise leur hierarchie (héritage multiple possible) (Class, Property, subclassOf, subPropertyOf,)

- OWL (Web Ontology Langage) : étend RDFs avec une description formelle bcp plus riche.

Ex: cardinalité, union, intersection, ... entre classes

Ex: owl:equivalentClass, owl: disjointWith, owl:sameAs

Le sujet <http://dbpedia.org/resource/Sete> a un lien owl:sameAs sur l'objet Sete dans la base de données Geonames (nom de lieux) <http://sws.geonames.org/2974733/>

En remontant dans le parent (arrondissement de Montpellier), on va retrouver un lien owl:sameAs sur l'INSEE.

<owl:sameAs rdf:resource="<http://id.insee.fr/geo/arrondissement/343>"/>

4.2/ en donnant une sémantique aux relations utilisées

W3C standard: SKOS (Simple Knowledge Organisation System)

permet de décrire entre concept des relations associatives (skos:related), hierarchiques (skos:narrower/ skos:broader), termes équivalents (skos:closeMatch),

L'information spatialisée sémantisée

Projet ANR Datalift (2010-2014): PUBLICATION ET INTERCONNEXION DES DONNÉES GÉOGRAPHIQUES,

Publication du IGN GEOFLA® (département) selon les bonnes pratiques du Web des données, à l'aide de la plateforme Datalift

Fayçal Hamdi (CNAM) / Nathalie Abadie (IGN)

<http://datalift.org>

Permet de transformer des données SIG shape en RDF et les publier sur le Web dans un triple store

Academic partners: INRIA, LIRMM Montpellier (Laboratory of Informatics, Robotics, and Microelectronics)

Institutional partners: IGN, INSEE

Industry partners: ATOS, Mondeca

- Sélection des schémas, vocabulaires, ontologies permettant de décrire les données: réutiliser les vocabulaires existants

pour explorer les vocabulaires: <http://lov.okfn.org/dataset/lov/>

démarche expliquée par Nathalie Abadie

- choix des vocabulaires pour les géométries: ETENDRE GEOSPARQL ET NEOGEO (propriété, restriction) pour fournir des géométries structurées (liste ordonnées de points, ...)
- identification d'un syst. de coord de référence (par ex: on doit pouvoir identifier tous les SCR français)

- Conversion des données en RDF: Push-button SHP to RDF conversion (using GeoTools)
- Interconnexion : lier les données entre elles en faisant le lien entre d'autres ontologies (ex: liens d'équivalence entre les données sameAs). Avec GeoNames par ex ou INSEE

- Publication sur le Web

Démonstration de l'application DataLift. Disponible sous Git. tourne sous java 1.7.

IGN Sparql Endpoint:

<http://data.ign.fr/>

Ex: de requete Récupérer tous les départements de moins N habitants

Conclusion: ces technos liés au mouvement Opendata. Web sémantique (Linked RDF) = 5ième étoile sur l'échelle mesurant à quel point les données sont ouvertes sur le Web
From open data to linked open data (Tim burners Lee):

- 1ere étoile: data on the Web (whatever format)
- 2 ième étoile: structured data in proprietary format (e.g. Excell instead of image scan)
- 3ième étoile: structured data in non proprietary format (e.g. csv instead of Excell)
- 4ième étoile: linked data in RDF standard (RDF, SPARQL)
- 5ième étoile: linked open data . Link to other data (RDFs, SKOS, OWL)

Systeme d'information

IDS GeoOrchestra: l'OGC et ses usages, architecture et feuille de route de Geo-orchestra et de l'OGC

Rodéric Béra et Hervé Squividant (INRA, Agrocampus ouest, Quimper)

<http://slides.com/stlejer-morvrini/>

Open Street Map, Sémantique Web et IDS

Rodéric Béra (INRA, Agrocampus ouest, Quimper)

convergence entre IDS et OSM:

- possibilité d'importer une base OSM dans postgis
- flux WMS pret à l'emploi des fonds OSM

- outils OSM vers IDS: OSM2geoOrchestra :addons pour rentrer base OSM dans Geo-orchestra: intérêt pour des projets au Sud ! (ex. de projet Sénégal; Burkina. Extraction de données taggués dispensaire d'OSM)

See also:

<http://ontology.irstea.fr/>

<https://data.irstea.fr>