

T2 - Ingénierie et web des données

- <http://devlog.Cnrs.fr/jdev2017>
- <http://devlog.Cnrs.fr/jdev2017/t2>

Objectifs

Solutions d'ingénierie nécessaires à la maîtrise des masses de données et à leur valorisation sur le web. De la modélisation à la conception des systèmes d'information internet. Des données structurées et semi-structurées au web sémantique.

Description

Nous vivons une révolution des données. La quantité produite de données augmente de manière exponentielle. Le web a donné naissance à notre monde interconnecté et rend accessible ces masses de données au plus grand nombre.

Comment intégrer ces sources de données dans les projets scientifiques et comment valoriser les nouveaux résultats de recherche produits?

Nous nous intéresserons aux différentes solutions du processus d'informatisation, du développement et de la maintenance des systèmes d'information.

De la modélisation du domaine avec UML en passant par le modèle entité-association, nous ferons le point sur l'apport du modèle du web sémantique.

Au-delà des moteurs relationnels et des moteurs NOSQL, nous verrons l'apport de SparQL.

Nous étudierons en quoi ce nouveau paradigme permet ou non d'explicitier, de représenter, de construire et de capitaliser les connaissances pour différentes communautés scientifiques et techniques. Nous serons attentifs à la rétro-ingénierie, à la capitalisation et à l'évolution de l'existant, à l'ingénierie des systèmes d'information internet.

Mots clés

- Modélisation des données
- web sémantique
- RDFS/RDF
- OWL
- Thésaurus/Ontologie
- SGBDR
- NOSQL
- SPARQL
- Recherche d'information

- Combinaison de ressource (mashup)
- Alignement de données
- ETL
- Graph mining
- Raisonneur
- Sécurité
- Qualité des données
- Visualisation des données
- Modéliser les traitements
- Prototypage rapide

Public

- communautés : réseaux DEVLOG, Calcul, RBDD, RENATIS, science des données, développement web

Planning

- Mardi 4/07 Matin: AP01
- Mercredi 5/07 Matin: Plénière (présentations ci-dessous)
- Mercredi 5/07 après-midi: A04, A05
- Mercredi 5/07 soir (17h30-19h00): GT06, GT01
- Jeudi 6/07 Matin: A01, A03, GT03/GT04, GT07
- Jeudi 6/07 Après-midi: A02, A06, GT02, GT05/GT08

Présentations

- 09h00-10h00 : Tutoriel sur le Linked Data - **Manuel Atencia (LIG & Inria Grenoble Rhône-Alpes)**
- 10h00-10h40 : Best Practices & Design Patterns pour l'ingénierie des données avec les langages du Web sémantique - **Olivier Corby (INRIA, I3S, Sophia Antipolis)**
- 10h40-11h00 : Pause
- 11h00-11h40 : Intégrer des Sources de Données Hétérogènes dans le Web de Données - **Franck Michel (I3S, CNRS, Université Côte d'Azur)**
- 11h40-12h05 : Retour d'expérience sur DOREMUS : données musicales en réutilisation (<http://www.doremus.org/>) - **Konstantin Todorov (LIRMM/Université de Montpellier)**
- 12h05-12h30 : Retour d'expérience sur Data Persee (<http://data.persee.fr/>) - **Viviane Boulétreau (ENS Lyon)**

Ateliers

La durée d'un atelier est de 3h.

Voir agenda :

- [T2.AP01](#) : Prise en main de SPARQL avec Wikidata - ma première requête - **Karima Rafes (BorderCloud)**
- [T2.A01](#) : Modéliser les ontologies : cas d'application d'une ontologie pour l'annotation de photos - **Manuel Atencia (LIG & Inria Grenoble Rhône-Alpes)**
- [T2.A02](#) : Comment faire une application avec des données liées: annotation de photo en utilisant les technos du web sémantique avec Jena et l'ontologie faite en T2.A01 - **Jérôme David (Enseignant-chercheur à l'Université Grenoble Alpes. Membre de l'équipe MoEx, LIG & Inria Grenoble Rhône-Alpes)**
- [T2.A03](#) : Alignement d'ontologies et interconnexion de données du web - **Kostantin Todorov (LIRMM/Université de Montpellier) & Franck Michel (I3S, CNRS, Université Côte d'Azur) & Manel Achichi (LIRMM/Université de Montpellier)**
- [T2.A04](#) : Atelier SparQL, niveau avancé - **Karima Rafes (BorderCloud)**
- [T2.A05](#) : STTL, un langage de transformation de graphes RDF basé sur SPARQL (analogue à XSLT) qui permet de concevoir des navigateurs hypertexte sur le Web de données - **Olivier Corby (INRIA, I3S, Sophia Antipolis)**
- [T2.A06](#) : Interrogation efficace de bases de données relationnelles avec SPARQL et Ontop - **Benjamin Cogrel (Post-doctorant, KRDB, Université Libre de Bozen-Bolzano)**

Groupes de travail

La durée d'un groupe de travail est de 1h30. Un groupe de travail permet de discuter, présenter, animer autour d'un sujet.

- [T2.GT01](#) : : Utilisation des vocabulaires contrôlés pour vérifier des données - **Baptiste Laporte (CESAB/FRB) & Marie-Claude Quidoz (CEFE/CNRS)**
- [T2.GT02](#) : Outils, compétences et savoir faire nécessaire à la fouille de graphes : quelques cas d'utilisation et les défis qui les accompagnent - **Romain David (IMBE/CNRS) & Luc HOGIE (CNRS/I3S/Université de Nice Sophia Antipolis/Inria) & Anna COHEN NABEIRO (FRB/Ecoscope) & Sophie PAMERLON (GBIF France)**
- [T2.GT03](#) : Quelles technologies du Web Sémantique choisir pour son système d'information ? - **Benjamin Cogrel (Post-doctorant, KRDB, Université Libre de Bozen-Bolzano)**
- [T2.GT04](#) - : Utilisation des technologies du Web Sémantique pour l'urbanisation et l'ouverture d'une infrastructure de recherche - **Adrien Desseigne (CNRS TGIR Hum-Num)**
- [T2.GT05](#) - [T2.GT08](#) : Extraction d'information du web : illustration aux réseaux de chercheurs - Crawling, Harvesting, Scraping sur le web de données pour la synthèse de réseaux scientifiques - **Sonia Guerin-Hamdi (CNRS/ISH) [T2.GT05](#)**

+ La plateforme « The NetWork of Researchers» pour moissonner, analyser, visualiser et valoriser les données de la production]] scientifiques sous forme de graphe - **A Phat LY (CNRS/LLL) & Yvan Stroppa (CNRS/Paul Painlevé) [T2.GT08](#)**

- [T2.GT06](#) : Quelles sont les compétences professionnelles à développer par les administrateurs de systèmes d'information dédiés à la gestion des connaissances (formation, certification, ...) ? - **Karima Rafes (BorderCloud)**
- [T2.GT07](#) : Référencement des données pour la recherche : retour d'expériences du Center for Data Science de l'université Paris-Saclay - **Karima Rafes (BorderCloud)**
- [T2.GT08](#) : Extraction d'information du web : illustration aux réseaux de chercheurs - La plateforme « The NetWork of Researchers » pour moissonner, analyser, visualiser et valoriser les données de la production scientifiques sous forme de graphe - **A Phat LY (CNRS/LLL) & Yvan Stroppa (CNRS/Paul Painlevé) Fusionné avec [T2.GT08](#)**

Voir aussi

- [T5.GT13](#) - les identifiants permanent, pourquoi, comment? La problématique de la citation des données (et des requêtes!) - RDA : Françoise Génova et/ou François-André (OMPs) ?

⇒ Devenir référent pour un schéma de données - comment on les nomme pour être pérenne. URI, DOI, ARK - Données d'autorités et leur gouvernance. Citations des données et des requêtes

- [T5.GT14](#) - Intéropérabilité des référentiels de données pour lier les ensembles de données variés - (Raphael Ritz, RDA) ⇒ Requêtage, accès aux données : fédérée/distribuée au niveau sparql, SQL (clause Service, merge local et données distant) (le mardi matin)