

Programme de l'Atelier du Réseau MIA
“Intégration de sources/masses de données hétérogènes et ontologies”
Montpellier SupAgro
25-26 juin 2018

Lundi 25 juin

13h-13h30 Accueil

13h30-16h00 Exposés, discussions, démo

13h30-14h30 « Le module sensoriel de l'ontologie MIAM » Sylvie Desprès (Université Paris 13 – LIMICS)

14h30-15h « Ontology-Mediated Query Answering and Heterogeneous Data » Federico Ulliana (LIRMM)

15h-15h30 « Extraction et Inférence de Connaissances d'Assemblages Mécaniques Comportant une Représentation CAO 3D » Harold Vilmart (INRIA - IMAGINE)

15h30-18h30 Visite

20h Diner

Mardi 26 juin

9h-10h30 Exposés, discussions, démo

9h-9h30 « Florilège: une base de données intégrative de phénotypes microbiens grâce à l'ontologie Ontobiotope et aux traitements de text-mining » Estelle chaix (INRA)

Ce travail se concentre sur l'extraction d'informations pertinentes sur les micro-organismes des produits alimentaires, en mettant l'accent sur la flore positive. Des travaux antérieurs ont montré que les relations entre les microbes, leur lieu de vie (nourriture comme habitat) et leurs propriétés phénotypiques sont des informations intéressantes pour les biologistes (In Ovide 2017). L'information textuelle est sous-exploitée car elle est exprimée en langage naturel, et donc non structurée. L'objectif de ce travail est de fournir un accès unifié à l'information de diverses sources: articles scientifiques, mais aussi des champs de texte libre de bases de données. Nous avons conçu un corpus composé de références d'articles scientifiques provenant de la base de données bibliographiques PubMed et des champs de texte des catalogues des centres de ressources biologiques, par exemple Inra CIRM (Centre International de Ressources Microbiennes, <http://www.inra.fr/cirm>), DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen) et des bases de données génétiques majeures, par exemple GenBank, BioSample. Le processus de text-mining derrière l'extraction d'information a été mis en place par l'Inra/Bibliome à l'aide des outils Alvis. Le

traitement appliqué au corpus est le même quelle que soit la source de l'information textuelle : (i) la détection d'entités des parties pertinentes du texte (mots ou des groupes de mots) qui sont identifiées et assignées à un type, "taxon", "habitat" ou "phénotype" ; (ii) la normalisation qui assigne une catégorie de la ressource de connaissance pertinente aux entités identifiées, (taxonomie du NCBI ou l'ontologie Ontobiotope), (iii) et enfin, l'extraction de relations qui lie ces entités entre elles. Le résultat est stocké dans la base de données appelée Florilege.

9h30-10h « Extraction, génération et analyse d'alignements entre ontologies dans AgroPortal » Elcio Abrahao (LIRMM)

10h-10h30 « Construction d'un OAD pour le choix d'emballage alimentaire qui combine perception consommateurs et dires d'experts. » Bruno Yun (INRA)

10h30-11h Pause-café

11h-13h Exposés, discussions, démo

11h-11h30 « Combining Apples and Oranges: A Flexible Representation for Defeasible Logics and Repair Semantics » Abdelraouf Hecham (LIRMM)

We propose Statement Graphs (SG), a new logical formalism for defeasible reasoning and repair semantics based on argumentation. Using a flexible labelling function, SGs can capture the variants of defeasible reasoning (ambiguity blocking or propagating, with or without team defeat, and circular reasoning) and certain repair semantics (IAR, ICAR, and -possibly- ICR, AR). We also evaluate our approach with respect to human reasoning and propose a working first order defeasible reasoning tool that, compared to the state of the art, has richer expressivity at no added computational cost.

11h30-12h30 « Data Infrastructures: La gestion de données distribuées et hétérogènes à tous les niveaux » Mathieu d'Aquin (NUI Galway, Ireland)

12h30-13h Discussion et bilan