



HAL
open science

Réseaux sémantiques étendus pour la gestion d'informations et de connaissances dans le domaine de la Toxicologie Nucléaire Environnementale

Joël Quinqueton, Reena Shetty, Pierre-Michel Riccio

► To cite this version:

Joël Quinqueton, Reena Shetty, Pierre-Michel Riccio. Réseaux sémantiques étendus pour la gestion d'informations et de connaissances dans le domaine de la Toxicologie Nucléaire Environnementale. RIC: Rencontres Intelligence Collective, May 2006, Nîmes, France. 2006. lirmm-00121929

HAL Id: lirmm-00121929

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00121929>

Submitted on 4 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Réseaux sémantiques étendus pour la gestion d'informations et de connaissances dans le domaine de la Toxicologie Nucléaire Environnementale

Reena T.N. SHETTY1, Pierre-Michel RICCIO2 et Joël QUINQUETON3

1- Doctorante Ecole des mines de Paris, 2- Maître de Conférences Ecole des mines d'Alès, 3- Professeur Université Montpellier

CENTRE DE RECHERCHE



Objectifs

Le programme **ToxNuc-E** rassemble plus de **600 chercheurs** du CEA, du CNRS, de l'INRA et de l'INSERM sur le thème de la Toxicologie Nucléaire Environnementale. Dans ce contexte, notre travail consiste à imaginer et mettre en œuvre de nouveaux outils, essentiellement logiciels, pour accompagner ces chercheurs répartis en **15 projets scientifiques** : dans la gestion d'informations et de connaissances, et dans la quête d'informations.

Mots clés : modèles de données, gestion d'information, gestion de connaissances, web sémantique, clustering, algorithmes d'extraction d'informations, ontologies, indexation de contenus.

Problématique

L'association des compétences et du savoir-faire de **spécialistes de différentes disciplines** (chimie, biologie, physique, informatique, ...) est une caractéristique du programme ToxNuc-E. Ces spécialistes possèdent un langage propre, travaillent à partir de **concepts spécifiques**, ce qui nécessite la construction d'un **référentiel commun**. Celui-ci est tout à la fois le socle et le résultat de ce programme de recherche. Pour créer ce référentiel, modèle d'informations et de connaissances, il n'existe pas à ce jour d'autre solution que de créer une **ontologie** de façon manuelle : ce qui compte tenu de la complexité du domaine est presque impossible. Nous proposons une alternative, les **réseaux sémantiques étendus**, construits de façon automatique à partir : d'une part de petits **réseaux sémantiques** élaborés par un groupe de chercheurs et, d'autre part de gros **réseaux de proximités** calculés mathématiquement à partir d'un ensemble de documents sélectionnés.

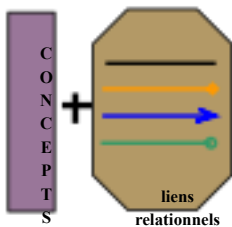


Fig. 1 : Réseau sémantique du projet Arabidopsis - RS

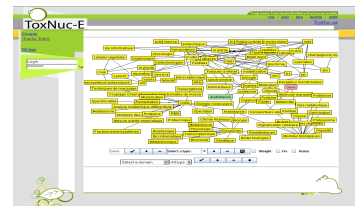
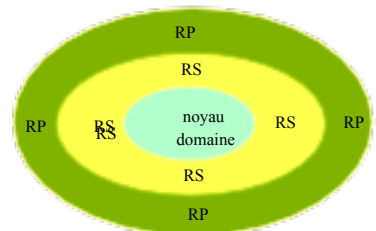


Fig. 3 : Réseau sémantique étendu du projet Arabidopsis - RSE

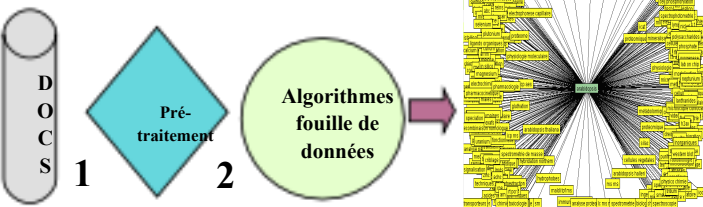


Fig. 2 : Réseau de proximité du programme ToxNuc-E - RP

Architecture

Le réseau sémantique **RS** construit à l'aide d'un éditeur de graphe est automatiquement combiné avec le réseau de proximité **RP** pour former le réseau sémantique étendu **RSE**.

Cette architecture permet de construire facilement de nombreux réseaux sémantiques étendus : par exemple un pour chacun des 15 projets du programme ToxNuc-E.

Un algorithme de classification permet alors de traiter de nouveaux documents, interne ou externe, et d'indiquer pour chacun d'entre eux les thématiques dominantes.

Avancement des travaux

Les réseaux de proximités sont actuellement construits par combinaison de trois méthodes : ACP, K-means et mots associés qui rendent compte de la proximité spatiale des termes dans le contenu documentaire. Nous envisageons prochainement d'introduire d'autres méthodes comme les SOMs pour améliorer la qualité des résultats.

Notre objectif est aussi de spécialiser les réseaux sémantiques étendus (par utilisateur) par intégration des **modèles utilisateurs** (profils et usages en situation).

Une première série d'applications, issues directement de ces travaux sera déployée sur ToxNuc-E en fin d'année 2006.