



HAL
open science

Le projet : le wiktionnaire sémantique des sciences humaines et sociales (WS/SHS) pour les espaces franco-maghrébin et francophone

Hammou Fadili

► To cite this version:

Hammou Fadili. Le projet : le wiktionnaire sémantique des sciences humaines et sociales (WS/SHS) pour les espaces franco-maghrébin et francophone. *Faits de langue et Société*, 2020, 6. hal-03093003

HAL Id: hal-03093003

<https://cnam.hal.science/hal-03093003v1>

Submitted on 1 Feb 2021

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.

LE PROJET : LE WIKTIONNAIRE SEMANTIQUE DES SCIENCES HUMAINES ET SOCIALES (WS-SHS) POUR LES ESPACES FRANCO-MAGHREBIN ET FRANCOPHONE

*The project: The Semantic Wiktionary of Human and
Social Sciences (WS-SHS) for Franco-Maghrebian and
Francophone Spaces*

Par / By

Hammou Fadili

Résumé

Cet article présente une contribution à la construction d'un Wiktionnaire des sciences sociales et humaines (SHS). Cette dernière est basée sur une extension de la plateforme « Mediawiki sémantique » existante afin qu'elle puisse prendre en compte aussi bien les aspects multiculturels et multidisciplinaires des SHS mais aussi pour une représentation de ses entrées selon le standard ISO 1951 relatif aux dictionnaires. Sa construction devrait permettre aux chercheurs des deux rives de la Méditerranée d'échanger et de partager des connaissances dans le domaine des sciences sociales et humaines et cela quels que soient leurs lieux géographiques de travail et/ou de résidence. Sa réalisation a été effectuée suivant des normes et protocoles bien spécifiques, en vue de son intégration dans le Web des données ouvertes (Linked Open Data ou LOD). Après le rappel du contexte du projet et les choix technologiques effectués, nous détaillerons les caractéristiques fonctionnelles et conceptuelles, suivies d'une brève description de l'application développée avant la conclusion.

Mots clés : Wiktionnaire, Données liées, Web des données, Web sémantique, Web 2.0, Web 3.0, RDF, SPARQL, Wiki sémantique, LOD

Abstract

This article presents a contribution to the construction of a Social and Human Sciences Wiktionary (SHS). The latter is based on an extension of the existing "Semantic Mediawiki" platform so that it can take into account

both the multicultural and multidisciplinary aspects of SHS but also for a representation of its entries according to the ISO 1951 standard relating to dictionaries. Its construction should allow researchers from both shores of the Mediterranean to exchange and share knowledge in the field of social and human sciences, regardless of their geographical place of work and / or residence. Its realization was carried out according to very specific standards and protocols, with a view to its integration into the Web of open data (Linked Open Data or LOD). After recalling the context of the project and the technological choices made, we will detail the functional and conceptual characteristics, followed by a brief description of the application developed before the conclusion.

Keywords: Wiktionary, linked open data (LOD), Web of data, Semantic Web, Web 2.0, Web 3.0, RDF, SPARQL, Semantic Wiki

Introduction

Les sciences sociales et humaines jouent un rôle primordial dans la compréhension et l'interprétation du contexte économique, culturel et social dans lesquels vivent et agissent les populations. L'évolution de la recherche dans ce domaine passe inévitablement par l'échange et le partage des connaissances entre les chercheurs. A la demande de la DRIC (Direction des relations internationales et de la coopération du ministère français de l'éducation nationale), un séminaire avait réuni en janvier 2002 à la Maison des Sciences de l'Homme (MSH) trente chercheurs maghrébins et français. L'objectif était de réfléchir ensemble sur la situation actuelle de la coopération en sciences sociales et humaines entre la France et les pays du Maghreb, et sur les moyens de la développer. C'est dans ce contexte qu'a été défini et lancé en 2005 le programme FSP-Maghreb.

Le comité de pilotage de ce programma a confié à la Fondation MSH la responsabilité de définir et coordonner, en liaison avec les équipes des 23 groupes de recherche ayant participé au FSP, un projet de construction d'un dictionnaire et d'un contenu multilingue, multidisciplinaire et multiculturel en collaboration, dont le laboratoire CEDRIC du Conservatoire des Arts et Métiers est l'un des principaux partenaires. Une fois réalisé, la plateforme du dictionnaire permettra de favoriser les échanges entre chercheurs maghrébins et leurs partenaires français en mettant en commun un ensemble de savoirs automatiquement exploitables, en Français, Arabe, Amazigh et Anglais, sur les deux cultures et les deux sociétés.

Dire que les contenus doivent être automatiquement exploitables, partageables, etc., cela veut dire qu'ils doivent être mis à disposition dans un contexte de données structurées, sémantiquement annotées et liées. Il s'agit, essentiellement dans notre cas, de définir un schéma de description bien adapté et de faire évoluer les fonctionnalités actuellement offertes par les technologies du Web de données qui est une implémentation pragmatique des évolutions du Web 2.0 (réseaux sociaux) et du Web 3.0 (Web sémantique), afin de prendre en charge les problématiques du domaine des sciences humaines et sociales tels que les aspects multilingues, multidisciplinaires et multiculturels des données. Ceci par l'adaptation et l'utilisation de la technologie « WIKI sémantique » et plus particulièrement de la plateforme Mediawiki sémantique de la fondation Wikimédia ; afin prendre en charge les problématiques liées à la normalisation et à l'intégration du dit dictionnaire dans le Web de données ouvertes (LOD).

L'objectif de cet article est de présenter les technologies utilisées et les travaux réalisés pour la mise en œuvre de l'application du Wiktionnaire sémantique des sciences humaines et sociales (WS-SHS). La section 1 rappelle le contexte et les motivations du présent travail. La section 2 décrit les choix technologiques effectués. La section 3 présente quelques éléments de conception et l'architecture du modèle retenus. La section 3 montre une vue d'ensemble du système développé. La dernière section conclut l'article et rappelle quelques perspectives et travaux futurs.

1. Contexte & motivations

Le programme du FSP Maghreb vise à appuyer la recherche en SHS au Maghreb et à développer les échanges entre les chercheurs et institutions de recherche maghrébines et leurs partenaires français dans une logique de réseaux et d'ouverture à l'international, notamment dans un cadre euro-méditerranéen. Il vise également à mettre en commun, par une meilleure maîtrise partagée de l'arabe, de l'amazigh et du français, et par une politique d'incitation à la traduction, un ensemble de savoirs sur les deux cultures et les deux sociétés, et à en assurer la diffusion internationale la plus large possible. Il vise ainsi à développer le débat d'idées entre les deux rives de la Méditerranée, et à conforter une communauté scientifique franco-maghrébine engagée dans un dialogue continu.

Trois objectifs principaux ont été retenus pour y parvenir :

- Mettre en place des réseaux de chercheurs des deux rives de la Méditerranée dans le cadre de programmes de recherche conjointe (réseaux intra-maghrébines et Maghreb-France) ;
- Aider au renforcement et au développement de communautés de jeunes chercheurs en Sciences humaines et sociales au Maghreb, mieux intégrées au sein de la communauté scientifique internationale ;
- Développer l'accès à l'information scientifique récente et l'échange de connaissances.

C'est dans ce contexte, que 23 projets thématiques ont été engagés, en plus d'un projet transversal commun à tous les projets, pour la mise en place d'une plateforme de communication, de création et d'échanges entre les chercheurs dans le domaine des SHS. D'un point de vue informatique, cette exigence peut se traduire, par le besoin de mettre en place une « sorte » d'ontologie multilingue, multiculturel et multidisciplinaire du domaine des sciences sociales et humaines (SHS) ; c'est-à-dire d'un espace et d'une base de connaissances du même domaine partagés et reconnus par la communauté des chercheurs Français et Maghrébines.

Cette demande est justifiée par le fait :

- Qu'il n'existe aucun dictionnaire des sciences sociales et humaines franco-maghrébin reflétant l'état de la coopération scientifique et culturelle entre la France et le Maghreb dont les lexiques & concepts pourraient évoluer de manière indépendante
- Qu'il y a un grand déséquilibre, d'un point de vue de la disponibilité des ressources numériques, entre le Français et les langues du Maghreb
- Les dictionnaires peu nombreux sont incomplets
- Les traducteurs existants (également peu nombreux) simplifient la traduction en faisant du « mot-à-mot », i.e. les traductions ne prennent pas en compte tous les aspects liés aux contextes des définitions des concepts
- Les corpus potentiels pouvant constituer des sources de données sont également rares
- Les travaux sur les nouvelles technologies et la normalisation des données des langues du Maghreb sont encore à leurs débuts

- Etc.

Pour y répondre, nous avons fait le choix, d'une solution basée sur un Wiktionnaire sémantique des SHS (WS-SHS), qui est l'objet du présent article.

1.1 Choix technologiques

La solution proposée pour répondre aux exigences exprimées dans le programme, repose sur des choix technologiques particuliers, que nous proposons de décrire dans ce qui suit.

1.1.1 Web de données (Web of data) & Données liées ouvertes (Linked Open Data ou LOD)

1.1.1.1 Le Web de données

Une donnée, non interprétée, est un élément fondamental décrivant un fait brut, devient une information une fois interprétée. Jusqu'à présent, l'interprétation automatique des données est très limitée, car, d'une part, les recherches dans ce domaine sont encore à leur début, et d'autre part, cela nécessite des technologies sophistiquées difficiles à mettre en place. D'où la nécessité de mettre en place des solutions innovantes, simplifiant les technologies existantes comme le Web sémantique, le TAL, etc., pour les rendre réalisables et rendre aussi leurs données exploitables. C'est ce contexte qui a favorisé l'émergence du Web des données (Web of Data).

Le Web de données est bien sûr une notion récente, qui est entrain de se populariser et se généraliser au niveau de la publication et de la diffusion des données « structurées » ou « fortement annotées sémantiquement » sur le Web. Il est constitué d'ensembles liés de données liées ouvertes (Linked Open Data ou LOD), suivant des standards et protocoles normalisés ; facilitant leur interprétation, leur exploitation et leur partage. Il est considéré par certains comme une implémentation du Web sémantique et, par d'autres comme une évolution pragmatique du Web sémantique, très vaste et très difficile à maîtriser.

« Le Web des données (Linked Data, en anglais) est une initiative du W3C (Consortium World Wide Web) visant à favoriser la publication de données structurées sur le Web, non pas sous la forme de silos de données isolés les uns des autres, mais en les reliant entre elles pour constituer un réseau global d'informations. » Source Wikipédia.

1.1.1.2 Données liées ouvertes

Une donnée ouverte (Open Data) est une donnée mise à disposition dans des formats ouverts et normalisés permettant son exploitation sous des licences ouvertes, libres et gratuites garantissant sa réutilisation dans un contexte international.

Le concept de données liées (Linked Data) est un ensemble de données « structurées » ou « fortement annotées », liées suivant des standards normalisés.

La combinaison des deux notions constitue ce qu'on appelle la notion de données liées ouvertes (Linked Open Data) qui a pour but l'exploitation du Web par les machines et par les êtres humains, faisant abstraction aux spécificités technologiques de toutes les sources de données liées. Cela consiste aussi à

répondre à la nécessité de disposer de données légalement "ouvertes", c'est à dire librement (ré-) utilisables par l'utilisateur et pour n'importe quel but.

C'est un moyen puissant pour connecter les utilisateurs à la connaissance et les utilisateurs entre eux **chacun dans sa langue** et dans des environnements hétérogènes et conviviaux.

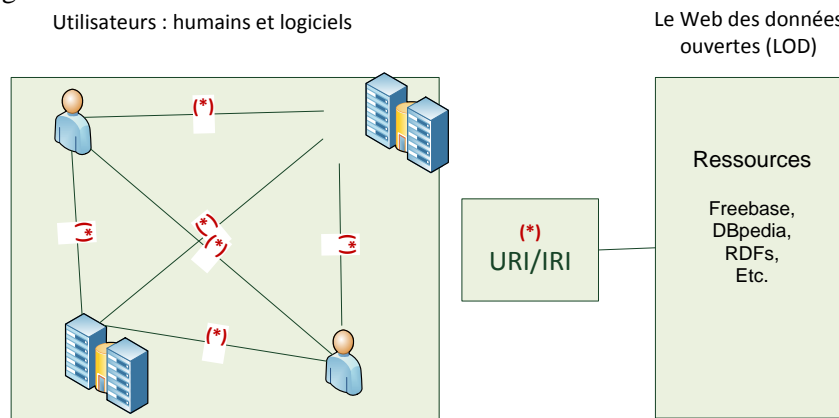


Figure 1 : Les utilisateurs du LOD

1.1.1.3 Mise à disposition des données dans le Web de données

Le but du Web de données est de permettre une évolution du Web actuel (Web 2.0) où les utilisateurs interagissent avec des sources de données fixes, vers un « Web de données », où les utilisateurs et les machines interagissent potentiellement avec le Web entier pouvant être considéré, à terme, comme une seule base de données universelle, composée des sources de données liées et compatibles. A ce stade de l'évolution de cette nouvelle notion du Web moderne, elle est surtout largement utilisée par les grandes institutions et les gouvernements, pour la mise à disposition de leurs données publiques, dans le cadre de la bonne gouvernance et transparence.

Pour produire et lier les données au Web de données, plusieurs normes, protocoles, standards, etc., sur lesquels doit reposer l'intégration des données, ont été mis en place. Tim Berners-Lee, dans ses travaux, les a résumés à travers les quatre principes suivants :

1. On doit utiliser les URIs comme noms des ressources.
2. On doit utiliser les URIs HTTP accessibles aux utilisateurs.
3. On doit utiliser les normes et standards RDF et SPARQL pour la description et l'interrogation des données.
4. Et puis, on doit créer des liens vers d'autres URIs afin de permettre de découvrir de nouvelles ressources.

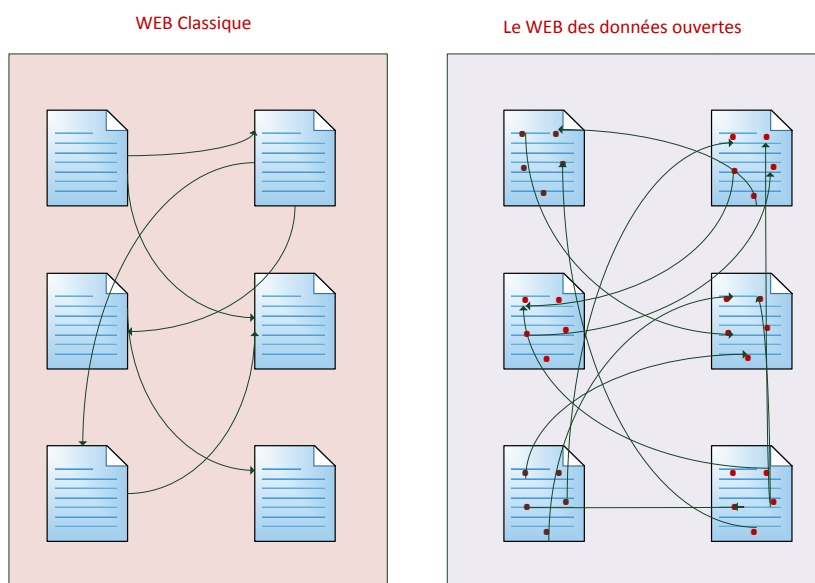


Figure 2 : Evolution du Web vers le LOD

Contrairement au Web classique permettant, seulement la création de liens entre les pages, le Web des données ouvertes permet, en plus, la création de liens entre les données à l'intérieur des pages. Pour mesurer la qualité de la publication des données, Tim Berners-Lee a également défini la classification en nombre d'(*) suivante.

Les conditions associées à chaque ligne sont, à cumuler en plus, de celles des lignes qui précédent.	
*	Publication des données sous n'importe quel format, mais avec une licence libre (données ouvertes).
**	Publication des données structurées (RDF, Excel, XML, CSV, etc.).
***	Utilisation des formats non-propriétaires (exclure Excel et tous les formats propriétaires).
****	Utilisation des standards du W3C pour identifier les ressources (URI, RDF, SPARQ, etc.).
*****	Création de lien vers d'autres ressources (définition du contexte des données).

Figure 3 : Qualité des données dans le LOD

1.1.2 Wiki sémantique, une implémentation de type « Web sémantique »

1.1.2.1 *Wiki*

En 1995, Ward Cunningham a inventé les wikis pour faciliter l'écriture collaborative de contenus.

Un Wiki est en général une application ouverte, universelle et multilingue en ligne. C'est une application de type Système de Gestion de Contenu (CMS) de sites Web permettant à des personnes autorisées, d'éditer et de publier facilement et rapidement des contenus en ligne, de les faire évoluer via des processus de travail collaboratif par mutualisation de compétences, une gestion complète des versions et des historiques des contenus et enfin une gestion des notifications permettant à des personnes intéressées par des thèmes particuliers d'être alertées à chaque fois que des créations, modifications ou suppressions de contenus par rapport à leurs thématiques favorites se reproduisent.

La technologie Wiki est basée sur des solutions légères moins coûteuses, accessibles à tous, où l'utilisation ne nécessite pas de formation particulière ; ce qui peut inciter à la création et encourager la production de nouveaux contenus. Ces solutions peuvent être perçues comme outils de communication de groupes, capables de tisser des liens entre communautés en créant des « réseaux sociaux » autour de centres d'intérêts communs. Ces « sociétés virtuelles » peuvent être un levier pour la création de groupes, augmenter la production de la « littérature grise » et ainsi créer une « intelligence collective ». Les Wikis peuvent être aussi considérés comme des outils permettant de sauvegarder la mémoire d'un groupe, d'un projet, d'une institution, et d'en assurer une meilleure diffusion du fait qu'ils sont très populaires et largement consultés. Les Wiki ont été adaptés et utilisés dans plusieurs domaines : espaces de rédaction, espaces pédagogiques, édition/publication électronique, bibliothèques numériques, Blogs, veille technologique, etc.

1.1.2.2 *Mediawiki*

MediaWiki est une plateforme Open source, développé à l'origine pour Wikipédia et aujourd'hui utilisé par plusieurs autres projets de l'association à but non lucratif Wikimedia Foundation ainsi que par beaucoup d'autres institutions privées ou publiques. Wikipédia l'un des projets du Wikimedia Foundation créé en 2001, est devenu le site web le plus visité écrit avec un wiki, développé avec Mediawiki. Néanmoins, le Mediawiki classique et équivalents souffrent de plusieurs limitations, liées surtout à l'interprétation et à l'exploitation de leurs contenus par des processus de traitements automatiques. Les contenus de type plein-texte nécessitant des Techniques du Traitement Automatique des Langues (TAL) difficile à mettre en place et dont les résultats de recherche sont encore peu satisfaisants.

1.1.2.3 *Mediawiki sémantique*

Pour remédier à ces limitations, et permettre la création et la gestion de contenus compréhensibles et utilisables au même temps par les machines et par les humains, plusieurs solutions basées sur la notion du Web sémantiques ont été

mis en place, dont, dans le cas de Wikimedia Foundation, avec l'évolution de Mediawiki vers le Mediawiki sémantique ou Semantic mediawiki (SMW). Ce dernier a l'avantage de combiner les technologies Wiki et du Web sémantique. Cette extension permet de structurer, rechercher, organiser, naviguer, évaluer et partager des contenus, en les rendant exploitables, aussi bien par les humains que par les machines ; à travers l'utilisation, entre autres, du langage de balises wiki « Wiki-ML » pour les annotations sémantiques, ainsi que beaucoup d'autres extensions annexes.

1.1.2.4 *Semantic forms & pageSchema*

Semantic forms est une extension du Mediawiki sémantique fournissant des mécanismes de saisie, basés sur des formulaires sémantiques, permettant ainsi la facilité d'édition des données et l'évitement des erreurs de frappe pour la saisie des annotations sémantiques. Cette extension comporte plusieurs utilitaires et outils, comme par exemple, l'introduction de fonctions pour l'auto-complétion des textes, le contrôle du vocabulaire, ... La génération de tels formulaires se fait par le biais d'une autre extension appelée « pageSchema » à partir d'un schéma XML facile à éditer et paramétrer via un méta-formulaire de définition du schéma.

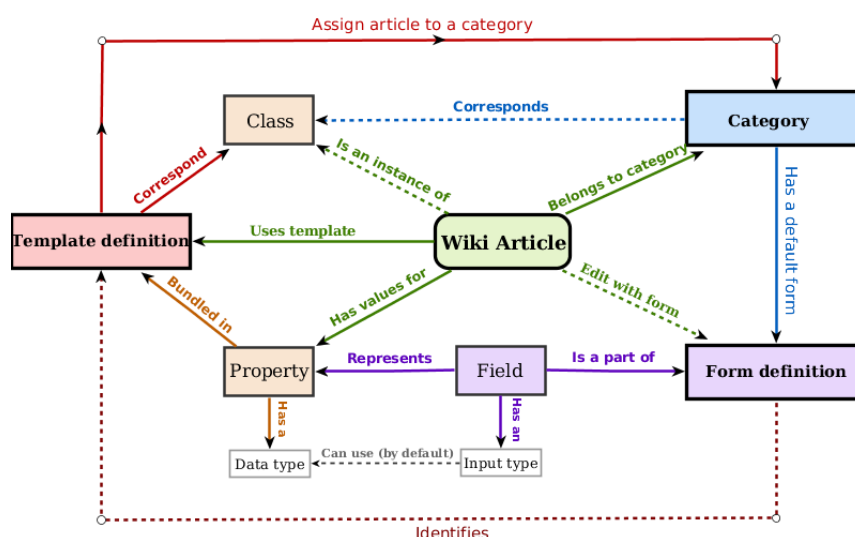


Figure 4 : Meta-modèle « Semantic forms » (Source : EduTechWiki)

C'est une solution idéale pour les besoins du projet, car, cela permet la mise en place d'un Wiktionnaire basé sur un schéma dynamique et extensible « à la demande », en fonction des besoins. C'était une réelle exigence au départ du projet. Etant donné, comme décrit précédemment, que nous devons respecter la norme relative aux dictionnaires ISO 1951, très complexe à mettre en place dans sa totalité, notre choix consistait à prendre en compte un sous-ensemble pertinent du schéma, qu'on peut étendre, « à la demande », en fonction des besoins et disposer ainsi d'une implémentation « POTENTIELLEMENT COMPLETE » de la totalité

de la norme ISO des dictionnaires. Cette souplesse a été décisive pour les choix technologiques retenus.

1.2 Modélisation

1.2.1 Aspects fonctionnels

L'analyse des besoins nous a permis de déduire, un premier modèle fonctionnel de l'application. Le Wiktionnaire des SHS doit :

- permettre une rédaction distribuée selon les localisations géographiques (France et Maghreb) des différentes équipes participant au programme. C'est l'une des plus importantes fonctionnalités du système.
- être extensible et évolutif, sachant que de nouveaux besoins peuvent apparaître lors de la conception, l'implémentation et l'exploitation du dictionnaire.
- être extensible de sorte à englober les langues : anglais, français, arabe et amazigh dans un premier temps puis une extension à toutes les langues du bassin méditerranéen.
- supporter différents alphabets et graphies des langues retenues.
- intégrer les différentes disciplines scientifiques du programme FSP, et offrir des passerelles entre celles-ci. Parmi ces disciplines scientifiques, on peut citer l'anthropologie, la sociologie, l'urbanisme, les sciences de l'environnement, etc.
- être compatible avec les normes dictionnaires existantes afin de faciliter les échanges.
- intégrer les fonctions d'import/export pour la compilation des contenus.
- permettre l'alimentation de la base dictionnaire dans un format structuré suivant la définition d'une entrée et de sa traduction, et vice versa afin de faciliter l'alimentation, les échanges et la génération des contenus
- être doté d'interfaces et de techniques lui assurant une grande ergonomie qui facilitera son utilisation par des chercheurs.
- prendre en charge l'aspect sémantique des définitions des entrées et de leurs traductions afin de permettre des renvois intelligents d'une langue à une autre, ou encore d'un parler à un autre.
- être libre d'accès en consultation, mais réservé en gestion.

1.2.2 Aspects conceptuels

La conception du Wiktionnaire en SHS repose sur des correspondances entre les éléments de départ dans leurs contextes pour la langue source et les éléments d'arrivée dans leurs contextes pour la/les langue(s) cible(s) selon un sous-ensemble du schéma de la norme ISO1951.

On doit prendre en compte le fait qu'une entrée A_k dans une langue source peut avoir plusieurs sens et donc plusieurs traductions $B_1, \dots, B_j, \dots, B_m$ dans la langue cible. Cette même entrée A_k peut être définie avec plusieurs éléments $A_1, \dots, A_i, \dots, A_n$ du schéma du dictionnaire (synonyme, antonyme, étymologie, hyperonyme, hyponyme, etc.) qui peuvent être à leurs tours des entrées dans la même langue source et par conséquent, peuvent avoir plusieurs sens dans cette même langue source et plusieurs traductions dans la langue cible (voir figure 5). Notons que

selon le sens de la traduction une langue source peut devenir cible et réciproquement.

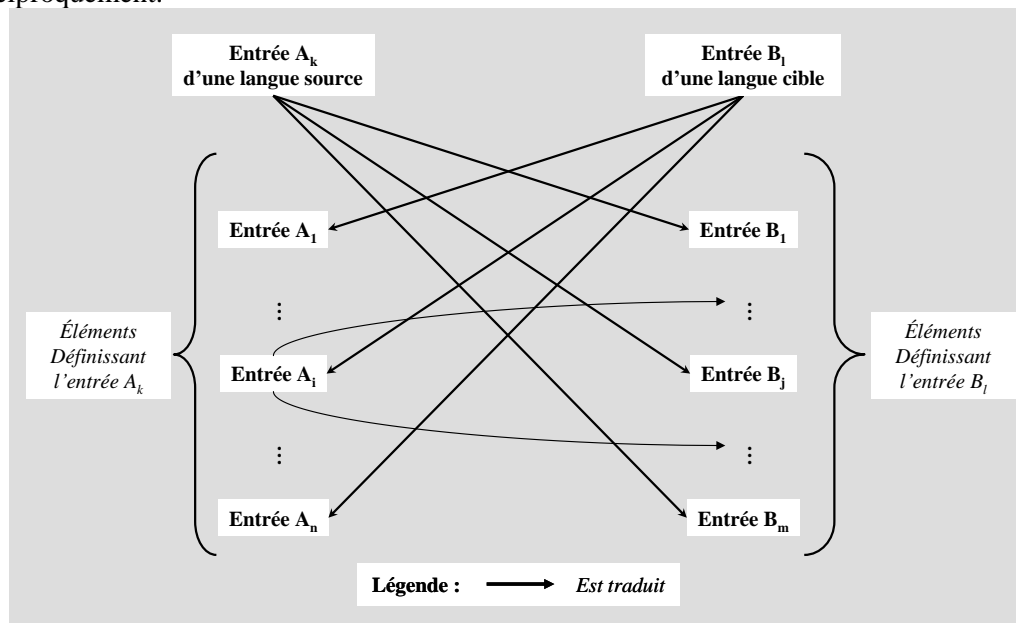


Figure 5 : Extrait du schéma du dictionnaire des SHS

Ce schéma, récapitulant ce qui a été décrit précédemment, montre la complexité des renvois entre la langue source et la langue cible. Elle nécessite la définition d'un schéma spécifique au Wiktionnaire des SHS dans chaque langue ainsi que la mise en place d'un système de gestion des correspondances d'une manière automatique. Pour ce faire, nous pouvons procéder tout d'abord par une première simplification du problème de départ, qui consiste à associer à une entrée source (mot, locution, etc.) un ou plusieurs sens (définitions) qui renvoient à une ou plusieurs entrées cibles ; puis revenir du terme traduit, pris cette fois-ci comme entrée source. Ce procédé doit être pris en charge par la mise en place d'un système de guidage d'aide à la génération et la définition des correspondances entre les entrées, leurs définitions et leurs traductions dans les différentes langues et que l'utilisateur peut compléter à partir des suggestions du système suivant les critères suivants :

- La définition d'une entrée se fait par la description d'une fiche suivant un format structuré déterminé par le schéma du Wiktionnaire sémantique des SHS et composé de : contexte, définitions, relations sémantiques, traductions, indications grammaticales, parlars, etc.
- la signification attribuée à une entrée dépend du contexte de définition. Ce dernier est décrit par un ensemble fini et connu de paramètres contextuels des aspects : temporels, géographiques, disciplinaires, culturels, linguistiques, etc.

- Les relations entre les termes, se fait par le biais de relations sémantiques telles que : la synonymie, l'antonymie, l'hyperonymie, l'hyponymie, l'isonymie, la conversion, etc.
- l'ensemble des éléments servant à décrire une entrée fait partie de la norme ISO 1951 [ISO 1951].

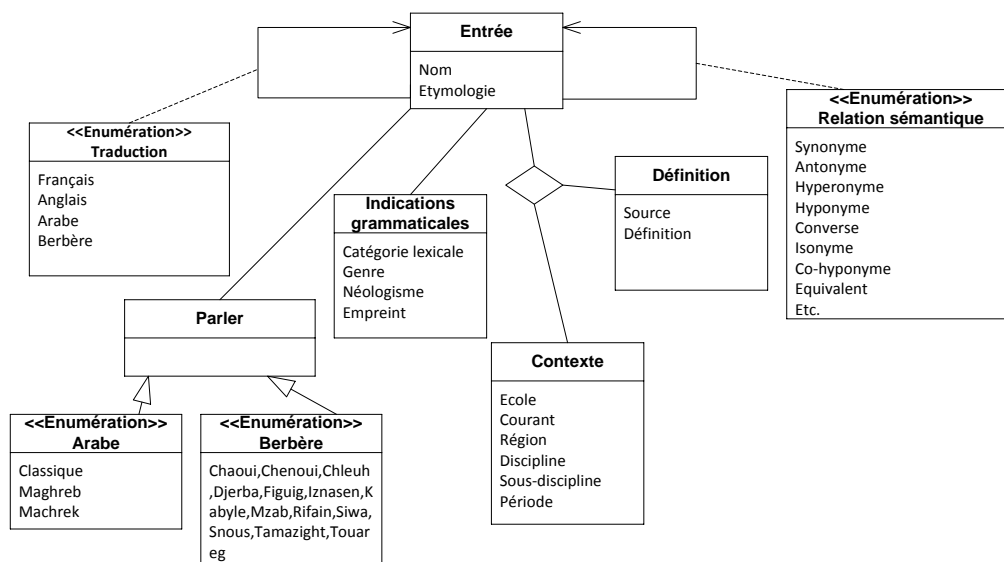


Figure 6 : modèle du schéma du Wiktionnaire

Pour la mise en place du dictionnaire, nous nous sommes basés sur un sous-ensemble de la norme ISO mentionnée précédemment pour retenir un schéma de base, à partir duquel, on a défini ou traduit des versions pour chacune des langues. Concrètement, nous avons créé deux versions pour chaque formulaire, une en Français et une traduite dans la langue qu'il représente : Français, Arabe, Amazigh ou Anglais. Ceci, permettra, à chacun de renseigner le Wiktionnaire dans sa langue et au non arabophones et aux non amazighophones de le renseigner en utilisant les versions françaises de tous les formulaires.

1.3 Vue d'ensemble du système développé

Le paragraphe suivant décrit une vue d'ensemble de la plateforme mise en place, à travers 3 sous-paragraphe dédiés respectivement, aux fonctionnalités principales offertes par le système, à nos contributions pour sa mise en place et à un aperçu de l'application à travers quelques copies d'écrans.

1.3.1 Principales fonctionnalités

L'analyse des besoins nous a permis de déduire un certain ensemble de fonctionnalités importantes que nous devons retenir et implémenter dans la plateforme. A rappeler que l'application été conçue et développée dans le but d'être compatible et intégrable dans le web de données et dont les données doivent être gérées dans un esprit de partage et coopération. Ci-après une liste non

exhaustive des principales fonctionnalités et services implémentés et mis à disposition par la plateforme.

Edition et gestion :

- Gestion ergonomique, collaborative et orientée communautés
- Sécurisation des accès et des contenus
- Saisie et annotation des données via des formulaires
- Gestion automatique de liens
- Génération automatique de listes
- Complétion automatique de la saisie

Structuration :

- Catégorisation et hiérarchisation des données
- Instanciation des catégories et création des articles
- Annotations et création des propriétés et relations des articles
- Génération automatique « à la demande » des versions multilingues du schéma

schéma

- Extension automatique « à la demande » du schéma à d'autres éléments de la norme

Interrogation/navigation :

- Interrogation sémantique les contenus
- Navigation par ontologie
- Recherche à facettes
- Export/import RDF
- Exposition d'un point de terminaison SPARQL
- Intégration dans le Web de données (triplets RDF)
- Navigation riche entre les langues

Diffusion et enrichissement automatique du dictionnaire :

- Meilleure diffusion : l'intégration du dictionnaire dans le Web de données, permet bien sûr une meilleure diffusion ; les sources de données qui lui sont liées sont dans la plupart des cas très connues et populaires, comme par exemple : le dbpedia.

- Enrichissement automatique : nous avons créé pour cela des requêtes paramétrées SPARQL permettant, suivant le contexte, de générer des contenus bien spécifiques, soit pour créer de nouvelles pages, soit pour enrichir d'une manière automatique la définition d'un concept SHS du dictionnaire. Le Wiktionnaire peut se servir potentiellement de la totalité du Web de données compatibles pour enrichir ses contenus.

1.3.2 Contribution et améliorations

L'implémentation de l'application a été réalisée à partir de la version de base de la plateforme Mediawiki sémantique et de ses extensions. Nous avons procédé à des développements, améliorations et adaptations des différents outils, que l'on peut résumer comme suit :

- Développement et adaptation du module associé au méta-formulaire basé sur « pageSchema », pour la génération automatique des formulaires du module

« semantic forms », et la prise en charge de toutes les langues dont l'arabe. Les versions arabes de ces modules n'existent pas

- Développement et adaptation du module « semantic forms » pour la génération de modèles d'affichage multilingues dont le cas de l'arabe
 - Personnalisation des feuilles de styles en fonction de la langue affichée
 - Définition d'un modèle pertinent du schéma du dictionnaire
 - Basé sur des catégories, propriétés, relations, ...
 - Définition des versions spécifiques du schéma de base, et ses versions légèrement différentes pour l'Arabe et l'Amazigh
 - Définition et génération de tous les formulaires

- Adaptation d'un schéma de base pour les langues retenues pour le dictionnaire :
 - Anglais, français, arabe et amazigh

- Contexte
 - Ecole(s)/ courant(s)
 - Région(s)
 - Discipline(s)
 - Sous-discipline(s)
 - Période(s)
- Etymologie
- Indications grammaticales
 - Catégorie(s) lexicale(s)
 - Genre
 - Néologisme (dans le cas de l'arabe et de l'amazigh)
- Parler(s) (dans le cas de l'arabe et de l'amazigh)
- Définition(s)
 - Source
 - Définition
- Relation(s) sémantique(s)
 - Synonyme(s)
 - Antonyme(s)
 - Hyperonyme(s)
 - Hyponyme(s)
 - Converse(s)
 - lsonyme(s)
 - Co-hyponyme(s)
- Traduction(s) (à partir de n'importe quelle langue, assurer les traductions vers : le français, l'anglais, l'arabe et l'amazigh)

Extrait du schéma : FR

- Source
- Contexte
 - Discipline
 - Période
 - Région
- Etymologie
- Définitions
- Relations sémantiques
 - Synonymes
 - Antonymes
 - Hyperonymes
 - Hyponymes
 - Converse
 - Etc.
- Traductions
 - Traductions arabes
 - Traductions berbères
- Texte libre

Extrait du schéma : AR

- المصدر
- سياق
 - موضوع
 - فترة
 - منطقة
- أصل
- تعريف
- العلاقة الدلالية
 - المرادفات
 - متناقض
 - عام
 - محدد
 - عكس
 - Etc.
- الترجمات
 - الترجمة الفرنسية
 - الترجمة الامازيغية
- Texte libre

Extrait du schéma : AM

- zzey
- Amnad
 - tugrint
 - tallit
 - tamnat
- Azur
- Tabadut/Tibatutin
- Issayen Itasnamekten
 - Akniw
 - Ameglawal
 - Tawacult
 - Mmuzzeg
 - Amgal
 - Etc.
- Tisuklitin
 - Tasuqilt tafransit
 - Tasuqilt taerabt
- Texte libre

Figure 7 : Adaptation d'un extrait du schéma du dictionnaire

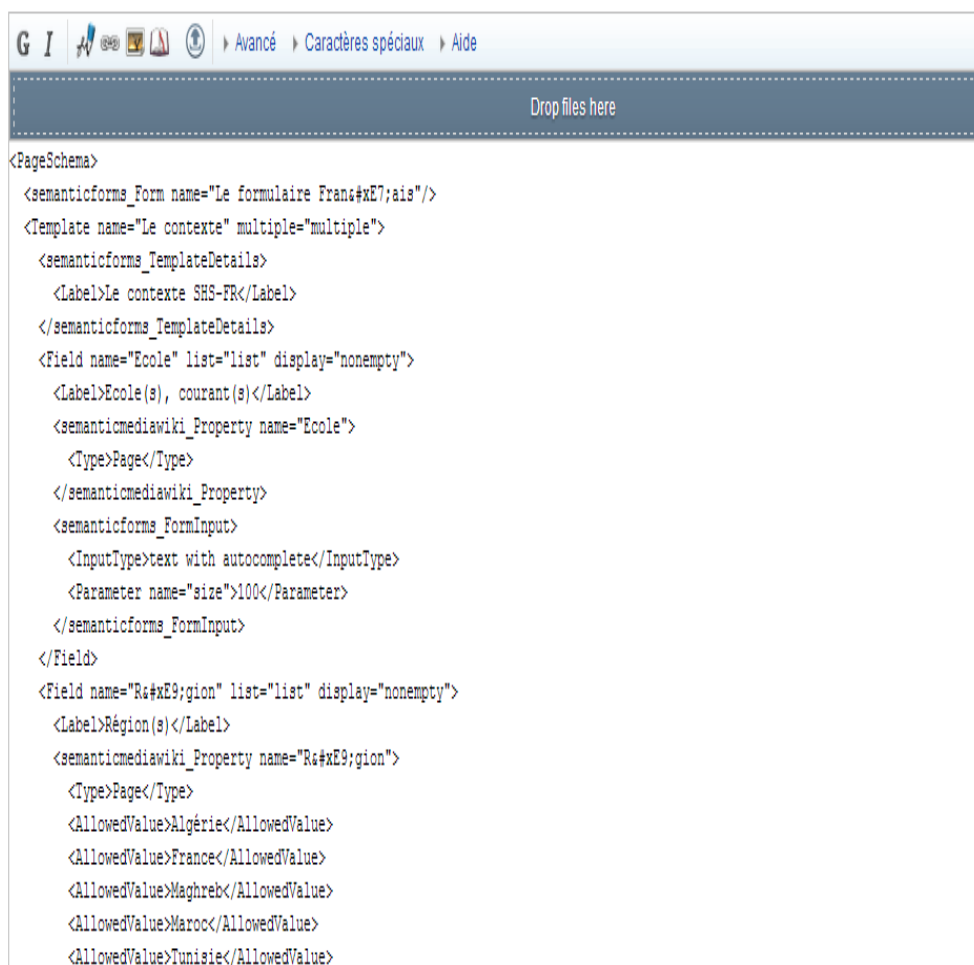
1.3.3 L'application

L'application est issue d'une adaptation de la technologie du MédiaWiki sémantique et de ses extensions pour prendre en charge la totalité des besoins de gestion du sous-ensemble du schéma du Wiktionnaire retenu dans les différentes langues. Ce dernier doit être facilement et rapidement extensible « à la demande », et doit couvrir, potentiellement, la totalité de la norme ISO1951. Nous avons pris en compte les spécificités de chaque langue, en particulier de l'arabe et son intégration dans le modèle de saisie orienté formulaire, comme par exemple l'écriture de la droite vers la gauche. La solution, a l'avantage de nous offrir la

possibilité de réaliser, dans un premier temps un Wiktionnaire anglais-français-arabe-amazigh, facilement extensible à d'autres langues pratiquées dans le bassin méditerranéen, contrainte majeure d'implémentation spécifiée dans les besoins initiaux du projet.

Figure 8 : Formulaire de gestion des schémas

Le méta-formulaire de création du schéma du Wiktionnaire permettant la génération automatique des formulaires de saisie ou d'annotation sémantique.



```
<PageSchema>
<semanticforms_Form name="Le formulaire Frans#xE7;ais"/>
<Template name="Le contexte" multiple="multiple">
  <semanticforms_TemplateDetails>
    <Label>Le contexte SHS-FR</Label>
  </semanticforms_TemplateDetails>
  <Field name="Ecole" list="list" display="nonempty">
    <Label>Ecole(s), courant(s)</Label>
    <semanticmediawiki_Property name="Ecole">
      <Type>Page</Type>
    </semanticmediawiki_Property>
    <semanticforms_FormInput>
      <InputType>text with autocomplete</InputType>
      <Parameter name="size">100</Parameter>
    </semanticforms_FormInput>
  </Field>
  <Field name="Rs#xE9;gion" list="list" display="nonempty">
    <Label>Région(s)</Label>
    <semanticmediawiki_Property name="Rs#xE9;gion">
      <Type>Page</Type>
      <AllowedValue>Algérie</AllowedValue>
      <AllowedValue>France</AllowedValue>
      <AllowedValue>Maghreb</AllowedValue>
      <AllowedValue>Maroc</AllowedValue>
      <AllowedValue>Tunisie</AllowedValue>
    </semanticmediawiki_Property>
  </Field>
</Template>
</PageSchema>
```

Figure 9 : Version XML générée du schéma

Version XML du schéma généré par le méta-formulaire de gestion de schémas précédent.

Modifier Le formulaire Français : Entrepreneur(1)

La(es) source(s) SHS-FR

La(es) source(s):
Auteur1

Ajouter une autre

Le contexte SHS-FR

La(es) région(s):
 France Maroc Algérie Tunisie Maghreb

La(es) discipline(s):
 Histoire sociologie droit

La(es) période(s):
 17eme 18eme 19eme 20eme 21eme 15ème 16ème

Ajouter une autre

Ethymologie

L'éthymologie:
Eth1, Eth2.

Ajouter une autre

Définition FR

Définition 1:
On parle également de l'économie lato sensu comme de la situation économique d'un pays ou d'une zone, c'est-à-dire de sa position conjoncturelle (par rapport aux cycles économiques) ou structurelle. Dans ce sens, l'économie est donc un quasi-synonyme à la fois de système et de régime. Enfin, de manière générale, en français, on parle d'économie comme synonyme de réduction de dépense ou d'épargne. L'économie peut en effet être le résultat d'une organisation interne plus efficiente : on parle alors d'économie interne. La baisse du coût moyen due à l'augmentation de la dimension de l'entreprise constitue une économie d'échelle ou économie

Définition 2:
On parle également de l'économie lato sensu comme de la situation économique d'un pays ou d'une zone, c'est-à-dire de sa position conjoncturelle (par rapport aux cycles économiques) ou structurelle. Dans ce sens, l'économie est donc un quasi-synonyme à la fois de système et de régime. Enfin, de manière générale, en français, on parle d'économie comme synonyme de réduction de dépense ou d'épargne. L'économie peut en effet être le résultat d'une organisation interne plus efficiente : on parle alors d'économie interne. La baisse du coût moyen due à l'augmentation de la dimension de l'entreprise constitue une économie d'échelle ou économie

Figure 10 : Formulaire d'édition.

Formulaire d'édition implémentant le schéma du Wiktionnaire retenu à un moment donné pour une langue donnée.

Entrepreneur(1)	
La(es) source(s)	
La(es) source(s)	Auteur1
Le contexte	
La(es) région(s)	France, Maroc
La(es) discipline(s)	Histoire, sociologie
La(es) période(s)	17eme, 18eme, 19eme
La(es) Ethymologie(s)	
L'éthymologie	Ethy1, Ethy2
La(es) Définition(s) FR	
Définition 1	<p>On parle également de l'économie lato sensu comme de la situation économique d'un pays ou d'une zone, c'est-à-dire de sa position conjoncturelle (par rapport aux cycles économiques) ou structurelle. Dans ce sens, l'économie est donc un quasi-synonyme à la fois de système et de régime. Enfin, de manière générale, en français, on parle d'économie comme synonyme de réduction de dépense ou d'épargne. L'économie peut en effet être le résultat d'une organisation interne plus efficiente : on parle alors d'économie interne. La baisse du coût moyen due à l'augmentation de la dimension de l'entreprise constitue une économie d'échelle ou économie de dimension. L'économie peut résulter d'un phénomène extérieur au pouvoir de décision de l'agent : on parle alors d'économie externe ou externalités qui peuvent être soit positives, si elles apportent un plus aux agents économiques, ou négatives, dans le cas contraire.</p> <p>On parle également de l'économie lato sensu comme de la situation économique d'un pays ou d'une zone, c'est-à-dire de sa position conjoncturelle (par rapport aux cycles économiques) ou structurelle. Dans ce sens, l'économie est donc un quasi-synonyme à la fois de système et de régime. Enfin, de manière générale, en français, on parle d'économie comme synonyme de réduction de dépense ou d'épargne. L'économie peut en effet être le résultat d'une organisation interne plus efficiente : on parle alors d'économie interne. La baisse du coût moyen due à l'augmentation de la dimension de l'entreprise constitue une économie d'échelle ou économie de dimension. L'économie peut résulter d'un phénomène extérieur au pouvoir de décision de l'agent : on parle alors d'économie externe ou externalités qui peuvent être soit positives, si elles apportent un plus aux agents économiques, ou négatives, dans le cas contraire.</p>
Définition 2	<p>On parle également de l'économie lato sensu comme de la situation économique d'un pays ou d'une zone, c'est-à-dire de sa position conjoncturelle (par rapport aux cycles économiques) ou structurelle. Dans ce sens, l'économie est donc un quasi-synonyme à la fois de système et de régime. Enfin, de manière générale, en français, on parle d'économie comme synonyme de réduction de dépense ou d'épargne. L'économie peut en effet être le résultat d'une organisation interne plus efficiente : on parle alors d'économie interne. La baisse du coût moyen due à l'augmentation de la dimension de l'entreprise constitue une économie d'échelle ou économie de dimension. L'économie peut résulter d'un phénomène extérieur au pouvoir de décision de l'agent : on parle alors d'économie externe ou externalités qui peuvent être soit positives, si elles apportent un plus aux agents économiques, ou négatives, dans le cas contraire.</p>

Figure 11 : Interface de navigation

Interface de lecture des entrées associées à des articles annotés suivant les modèles de description associés à chaque langue.

Point de terminaison SPARQL

Enter SPARQL query:

```
@PREFIX w : <Special:URIResolver/> .
SELECT *
WHERE { ?s ?p ?o }
LIMIT 25
```

Query by Equivalent URIs
 Output Equivalent URIs
 Output format: **HTML**

(RDF/XML requires creating triples using CONSTRUCT)

Result:

s	p	o
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#DatatypeProperty
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://www.w3.org/2000/01/rdf-schema#label	Discipline(s)
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://semantico-mediawiki.org/swiki/1.09page	http://france-maghreb.msh-paris.fr/wiktionnaire/index.php/Attrib%20Dis
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://www.w3.org/2000/01/rdf-schema#isDefinedBy	http://france-maghreb.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://semantico-mediawiki.org/swiki/1.09wiki:namespace	102
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://semantico-mediawiki.org/swiki/1.09wiki:Page:Creator:Date	2013-07-23T10:17:37Z
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	2496496 9299
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://semantico-mediawiki.org/swiki/1.09wiki:Page:LastEditor	http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://semantico-mediawiki.org/swiki/1.09wiki:Page:Modification:Date	2013-08-17T16:00:38Z
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	2496553 17266
http://www.msh-paris.fr/wiktionnaire/index.php/Spc%3%ABscial%20UR%20Resoluer%20Attrib%203%20Discipline%2028a-29	http://semantico-mediawiki.org/swiki/1.09special:Property:_PIVAL	!Histoire

Figure 12 : Point de terminaison et interrogation SPARQL

1.4 Conclusion et travaux futurs

Nous avons présenté dans cet article une approche de construction d'un dictionnaire multilingue, multidisciplinaire et multiculturel franco-maghrébin pour les sciences sociales et humaines. Ce projet a été initié par la Fondation Maison des Sciences de l'Homme.

L'architecture mise en place repose sur une conception générique représentant le choix du schéma et des langues retenus à un moment donné, mais facilement extensible à d'autres schémas et généralisable à d'autres langues. La plateforme développée est une extension de la plateforme Mediawiki et plus particulièrement de la plateforme Mediawiki sémantique, garantissant une utilisation ergonomique : saisie, annotations, consultation, navigation et recherche sémantiques. Elle est en phase de production après plusieurs mois de développements-tests-améliorations et une validation par un comité d'experts/chercheurs du domaine.

Comme perspective de recherche à ce travail, on prévoit, l'intégration du Wiktionnaire dans le Web de données, puis l'utilisation des données, structurées, sémantiquement annotées et liées (intégrées dans le Web de données) pour enrichir un travail de recherche parallèle, concernant une nouvelle approche de détection automatique et contextuelle du sens dans le texte : Basé sur les relations lexico-sémantiques et la notion de contexte.

Hammou Fadili

**Laboratoire CEDRIC du Conservatoire National des Arts et Métiers
de Paris
(Pôle Recherche & Prospective, Programme Maghreb) de la FMSH, Paris,
France**

REFERENCES WEBOGRAPHIQUES

<http://www.w3.org/DesignIssues/LinkedData.html>
<http://linkeddata.org/>
<http://linguistics.okfn.org/2011/05/20/the-open-linguistics-working-group/>
<http://ldl2012.lod2.eu/>
<http://inkdroid.org/lod-graph/>
<http://www.mediawiki.org/wiki/Special:ExtensionDistributor>
<https://semantic-mediawiki.org/>

ANNEXE

1. Le projet

Le projet Wiktionnaire Sémantique des Sciences Humaines et Sociales (WS-SHS) a été défini par le comité scientifique du FSP Maghreb pour répondre aux besoins des chercheurs du Maghreb et de l'espace francophone, qui en avait confié la réalisation à la FMSH, gestionnaire du FSP, dont il devait constituer le principal prolongement.

1.1. Coordination

Maurice Aymard (Historien, FMSH/EHESS), Hammou Fadili (FMSH/CNAM, Chercheur habilité à diriger les recherches (HDR) en informatique) et Patrice Pognan (Linguiste, professeur émérite, INALCO).

1.2. Institutions associées au Maghreb

IRCAM (Rabat), CRASC (Oran), CREAD (Alger), l'Ecole Normale Supérieure en Informatique (Alger), Universités de Tizi-Ouzou et de Fès)

1.3. Equipes françaises associées

LDIF (Paris 13, Mejrî Salah), STIH (Paris 4, Jean-Pierre Desclées), PLIDAM (Inalco, Patrice Pognan).

1.4. Principaux chercheurs associés

Meftaha Ameer (IRCAM), Miloud Taifi (Fès), Mohand Salhi Akli (Tizi-Ouzou), Malika Chakiri (Paris 5), Moktar Meroufel (Mascara/Oran), Fatma Gutbi (Khartoum).

Ce réseau scientifique, en cours d'élargissement, repose sur l'implication des chercheurs et d'équipes engagés dans des recherches correspondant aux objectifs du Wiktionnaire, et susceptibles d'y intégrer leurs résultats. La FMSH assure la mise au point et le fonctionnement de la plateforme informatique, et lui apporte son soutien logistique.

1.5. Objectif

Construire un ensemble de contenus définissant, dans le champ des SHS des concepts multidisciplinaires, multilingues et multiculturels, structurés, normalisés et annotés par des enrichissements sémantiques. Il vise à favoriser les échanges entre les chercheurs du Maghreb et leurs partenaires francophones et à leur permettre à la fois de « parler le même langage », mais aussi de comprendre les difficultés inhérentes à toute opération de traduction pour des disciplines dont le vocabulaire est emprunté pour une partie aux langues naturelles, sur lesquelles s'appuient en priorité les études sur le terrain, et pour une autre partie à une ou plusieurs langues scientifiques à ambition internationale, élaborées et codifiées à différents moments de l'histoire et dans différents contextes politiques et culturels. Très souvent, chaque terme peut renvoyer à plusieurs dans d'autres langues : d'où la nécessité d'enrichir le dictionnaire par un choix représentatif d'exemples empruntés à des auteurs de référence pour chaque langue dans chaque discipline concernée.

1.6. Situation actuelle

Les résultats du projet Wiktionnaire sémantique des SHS sont en phase d'exploitation et de production. Un réseau de chercheurs français et maghrébins actif a été constitué autour d'une plateforme informatique, mise en place pour permettre une gestion complète des données en son sein ; et respectant les normes dictionnairiques dans le contexte du Linked Open Data (LOD). La base contient actuellement plus de 1300 entrées dans les trois langues : français, arabe et berbère ; dont, par exemple, les traductions des concepts de Bourdieu, Mauss, Berque et de Durkheim. Des chercheurs du réseau ont récemment traduit le dictionnaire d'Oswald Ducrot, il sera également intégré au Wiktionnaire.