



HAL
open science

TyDI: un outil d'assistance à la construction de terminologie structurée

Sophie Aubin, Wiktoria Golik, Frederic Papazian, Claire Nédellec

► To cite this version:

Sophie Aubin, Wiktoria Golik, Frederic Papazian, Claire Nédellec. TyDI: un outil d'assistance à la construction de terminologie structurée. 8. International conference on terminology and artificial intelligence: TIA 2009, Nov 2009, Toulouse, France. hal-02756848

HAL Id: hal-02756848

<https://hal.inrae.fr/hal-02756848>

Submitted on 3 Jun 2020

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.

TyDI : un outil d'assistance à la construction de terminologie structurée

Sophie Aubin, Wiktoria Golik, Frédéric Papazian, Claire Nédellec

MIG (Mathématique, Informatique et Génome), INRA
{prénom.nom}@jouy.inra.fr

Mots-clés : Construction de terminologie, Interface homme-machine, Termes candidats, Relations sémantiques entre termes.

1 Présentation

TyDI est destiné à assister la construction de terminologie. Il permet en particulier de caractériser des termes pertinents pour une application et de les structurer grâce à des liens de synonymie et d'hyperonymie. Ses deux utilisations principales sont la construction de terminologie à partir de candidats termes extraits automatiquement de corpus de texte en langage naturel et la sélection de termes pertinents à partir d'une terminologie existante. Les informations linguistiques associées aux termes font des terminologies obtenues des ressources exploitables pour l'indexation manuelle ou automatique en texte plein ou pour la construction d'ontologie.

Avec TyDI, la mesure de la pertinence et la représentativité d'un terme pour une application donnée repose sur deux types d'analyse, celle de la pertinence de termes sémantiquement proches et celle de l'usage du terme à travers son contexte dans les textes. La proximité de forme et de structure syntaxique des termes rend partiellement compte de leur proximité sémantique, c'est cette piste qui est suivie dans TyDI, par opposition à la similarité de contexte telle qu'exploitée en sémantique distributionnelle. TyDI propose un large éventail de fonctionnalités destinées à permettre à l'utilisateur de visualiser des sous-ensembles de termes morphologiquement proches avec leurs informations linguistiques associées.

Ces informations sont principalement l'origine du terme, sa forme de surface, son lemme, sa structure syntaxique, sa tête, ses expansions, ses sous-termes et leur nombre, son nombre de mots, son nombre d'occurrences par corpus, les termes variants, les synonymes et hyper/hyponymes, telles que provenant de terminologies existantes, ou calculées par les outils en amont de l'utilisation de TyDI. Ce sont des extracteurs de termes tels que YaTeA, et des outils de calculs de variants morphosyntaxiques tels que FastR. Ces informations sont facultatives, TyDI pouvant traiter de simples listes de termes. Le panneau principal (*TermGrid*) en jaune dans la Fig. 1 affiche la liste des termes tels que sélectionnés par le filtre défini dans la grille au-dessus. Outre les informations décrites ci-dessus, la mention de l'utilisateur courant (*Fred* dans Fig. 1) est affichée avec la mesure de pertinence attribuée au terme (V / D ici). Les valeurs assignées par d'autres utilisateurs pour le même projet peuvent être affichées et filtrées. L'utilisateur peut également associer un

commentaire. Les valeurs des filtres des champs *Form*, *Lemma*, *Syntactic category* et *Expansion* sont des combinaisons de caractères et de métacaractères (* et ?). Divers critères de sélection et de tri par colonne sont également possibles à l'aide d'expressions régulières. Le panneau gris *OccurrenceContextWindow* visualise les occurrences du terme sélectionné en contexte dans le corpus choisi. Une recherche à l'aide d'un moteur externe tel que *Google* ou celui de *Wikipédia* peut être déclenchée. Le panneau rose permet de visualiser et d'éditer les relations sémantiques des termes sélectionnés dans la grille par simple glisser-déposer. Ce sont les relations d'hyper/hyponymie (*illegal fish* ici), d'antonymie et de synonymie pour lesquels l'utilisateur précise les cas d'acronymie, de variation typographique (erreur d'OCR, de segmentation) et de traduction. Un terme représentatif doit être désigné pour chaque classe sémantique définie par ses synonymes (*illegal cod* ici). Une autre fenêtre permet de visualiser sous forme de liste et sous forme de graphe, de filtrer et de qualifier les relations entre termes variants.

Le panneau bleu affiche les projets de l'utilisateur courant. L'architecture de TyDI est une architecture de Web service avec un client lourd. Les projets sont gérés par une base de données ce qui permet un accès multi-utilisateur et la sauvegarde temps réel.

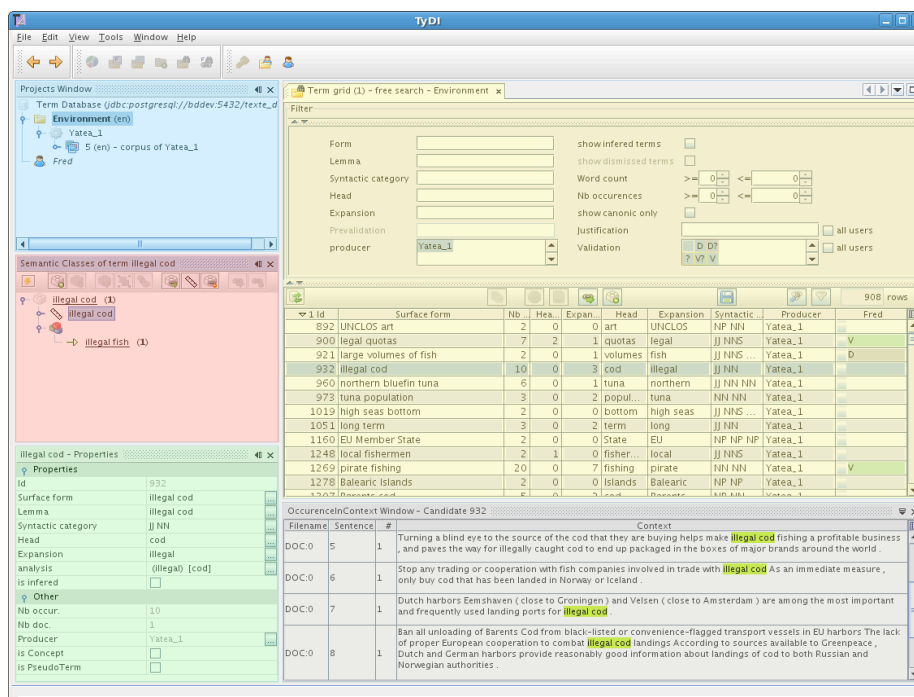


Fig. 1 – Fenêtre principale de TyDI

Remerciement : Ce travail est partiellement financé par Oséo à travers le projet Quero. Les auteurs remercient particulièrement Sandra Makuntima et Bernard Teyssendier pour leur contribution à l'évaluation de TyDI par l'utilisateur.