



TermLis : un contexte d'information logique pour des ressources terminologiques

Annie Foret

► To cite this version:

Annie Foret. TermLis : un contexte d'information logique pour des ressources terminologiques . Traitement Automatique des Langues Naturelles, Jun 2015, Caen, France. Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles, 2015, Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles. <http://www.atala.org/taln_archives/TALN/TALN-2015>. <hal-01252802>

HAL Id: hal-01252802

<https://hal.inria.fr/hal-01252802>

Submitted on 8 Jan 2016

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.

TermLis : un contexte d'information logique pour des ressources terminologiques.

Annie Foret
IRISA, Université Rennes 1, France
foret@irisa.fr

Résumé. Nous présentons `TermLis` un contexte d'information logique construit à partir de ressources terminologiques disponibles en xml (FranceTerme), pour une utilisation flexible avec un logiciel de contexte logique (CAMELIS). Une vue en contexte logique permet d'explorer des informations de manière flexible, sans rédaction de requête a priori, et d'obtenir aussi des indications sur la qualité des données. Un tel contexte peut être enrichi par d'autres informations (de natures diverses), mais aussi en le reliant à d'autres applications (par des actions associées selon des arguments fournis par le contexte). Nous montrons comment utiliser `TermLis` et nous illustrons, à travers cette réalisation concrète sur des données de FranceTerme, les avantages d'une telle approche pour des données terminologiques.

Abstract.

TermLis : a logical information context for terminological resources.

We present `TermLis` a logical information context constructed from terminological resources available in XML (FranceTerme), for a flexible use with a logical context system (CAMELIS). A logical view of a context allows to explore information in a flexible way, without writing explicit queries, it may also provide insights on the quality of the data. Such a context can be enriched by other information (of diverse natures), it can also be linked with other applications (according to arguments supplied by the context). We show how to use `TermLis` and we illustrate, through this concrete realization from FranceTerme data, the advantages of such an approach with terminological data.

Mots-clés : Applications multilingues, Classification, Extraction d'information, Fouilles de données textuelles, Recherche d'information, Ressources du langage, Données Ouvertes, Qualité des données, Données légales.

Keywords : Multilingual applications, Classification, Information extraction, Textual data mining, Information retrieval, Linguistic resources, Open Data, Information Quality, Legal Information.

Introduction. Cette réalisation se situe dans le cadre de travaux visant à rendre exploitables des données linguistiques par l'approche des systèmes d'information logiques : de telles données ne sont pas toujours simples d'utilisation sans aide, par ailleurs les systèmes d'information logique sont conçus pour permettre une navigation flexible dans des données organisées comme un contexte logique. D'autres travaux se situent dans un cadre similaire (Cellier *et al.*, 2011; Quiniou *et al.*, 2012; Foret & Ferré, 2010; Falk *et al.*, 2014) mais les données y sont de natures différentes, et les buts visés aussi.

Cette démonstration illustre comment une vue en contexte logique permet d'explorer des données terminologiques de manière flexible, avec des facettes sémantiques, des possibilités d'inférences logiques et sans rédaction de requête a priori. Nous traitons le cas des données FranceTerme avec l'outil de gestion de contexte `Camelis` (le seul outil de gestion de contexte logique disponible à notre connaissance).

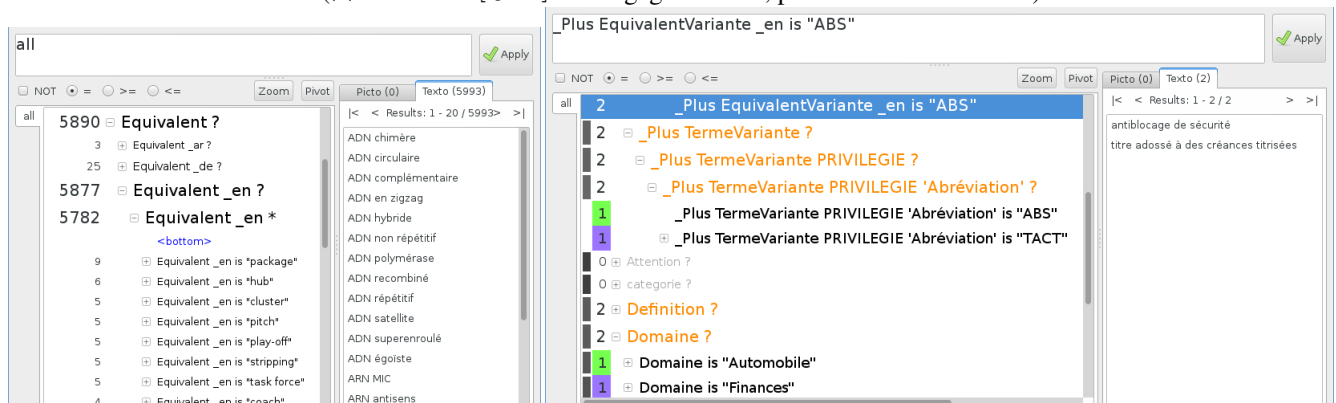
Ressource linguistique. La ressource utilisée présente une richesse de structure : aspects multilingues, des définitions, des relations synonymes etc., une variabilité en fonction d'un domaine/sous-domaine ou en fonction de critères linguistiques (plusieurs variantes d'anglais par exemple), l'absence possible ou la répétition possible de certains types d'informations ; son caractère règlementaire (avec des informations source et date de publication) nous intéresse aussi. Des extensions ultérieures peuvent être envisagées : pour de nouvelles données que l'on organisera selon un schéma analogue.

Contexte logique. Un contexte logique est défini par un ensemble fini d'objets \mathcal{O} , chaque objet o_i ayant un label l_i (a priori plusieurs objets peuvent avoir le même label), et pour chaque objet o_i , un ensemble fini de descriptions logiques $d(o_i)$ chaque expression étant une formule bien formée pour un langage logique donné L . Un *système de gestion de contexte logique* permet de charger et d'exploiter un tel contexte, en permettant l'interrogation d'un contexte par des requêtes logiques (explicites ou interactives), la réponse est alors un sous-contexte d'objets satisfaisant cette requête. Nous avons utilisé `Camelis` (version 1, accessible à <http://www.irisa.fr/LIS/ferre/camelis/>) ce logiciel est basé sur l'analyse de concept logique (LCA) définie en (Ferré & Ridoux, 2004), une extension de l'analyse de concept formel

(FCA, voir (Ganter & Wille, 1999)) : un *concept logique*, noté c , est un couple formé d'une extension $ext(c)$ (un ensemble d'objets) et d'une intension $int(c)$ (une formule) tel que les éléments de $ext(c)$ sont exactement ceux qui vérifient $int(c)$; ces concepts forment un treillis auquel correspond l'*arbre de navigation logique* et incrémental dans la fenêtre gauche du logiciel. Le logiciel Camelis est aussi prévu pour gérer des ensembles d'objets de types différents.

Contexte TermLis. Le contexte logique a été obtenu en réalisant un transducteur, appliqué au fichier source XML pour FranceTerme. Cette réalisation, dont nous illustrons l'usage avec Camelis (cf figures), s'est appuyée sur :

- un modèle du document XML (une DTD générée automatiquement) ;
- un modèle de contexte logique visé ; avec notamment : un choix de types d'objets ; une sélection de propriétés ;
- un lien entre les deux modèles, comprenant en particulier la spécification d'une clé dans le source (par une expression de chemin XML (`//Article[@id]` en langage XPATH, pour les codes d'Article).



L'arbre de navigation à gauche permet des scénarios variés selon les facettes ouvertes et sélectionnées successivement (all correspond au contexte initial) : recherches simples avec divers types de données (chaines, dates etc.) ; combinaisons logiques ; mise en évidence de variations/faux-amis, d'un élément (comme <Attention>), voire des défauts (redondances).

La *cohérence* est assurée entre les 3 fenêtres (requête/formule logique en haut ; objets à droite ; index de propriétés/arbre de navigation à gauche). Un autre avantage de l'exploration avec Camelis est *sa navigation sûre* évitant les résultats vides.

Des capacités complémentaires pour le contexte sont introduites par : des axiomes logiques pour des recherches avec inférences ; des règles ajoutant des propriétés ; des actions liant à d'autres ressources (CNRTL) ou traitements (analyseur,...). Nous illustrons aussi ces points, un découpage modulaire permettant d'intégrer tout ou une sélection de ces compléments.

Références

- CELLIER P., FERRÉ S., DUCASSÉ M. & CHARNOIS T. (2011). Partial orders and logical concept analysis to explore patterns extracted by data mining. In S. ANDREWS, S. POLOVINA, R. HILL & B. AKHGAR, Eds., *Conceptual Structures for Discovering Knowledge - 19th International Conference on Conceptual Structures, ICCS 2011, Derby, UK, July 25-29, 2011. Proceedings*, volume 6828 of *Lecture Notes in Computer Science*, p. 77–90 : Springer.
- FALK I., BERNHARD D. & GÉRARD C. (2014). From non word to new word : Automatically identifying neologisms in french newspapers. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, H. LOFTSSON, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Eds., *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26-31, 2014.*, p. 4337–4344 : European Language Resources Association (ELRA).
- FERRÉ S. & RIDOUX O. (2004). Introduction to logical information systems. *Inf. Process. Manage.*, **40**(3), 383–419.
- FORET A. & FERRÉ S. (2010). On categorial grammars as logical information systems. In L. KWUIDA & B. SERTKAYA, Eds., *Formal Concept Analysis, 8th International Conference, ICFCA 2010, Agadir, Morocco, March 15-18, 2010. Proceedings*, volume 5986 of *Lecture Notes in Computer Science*, p. 225–240 : Springer.
- GANTER B. & WILLE R. (1999). *Formal concept analysis - mathematical foundations*. Springer.
- QUINIOU S., CELLIER P., CHARNOIS T. & LEGALLOIS D. (2012). What about sequential data mining techniques to identify linguistic patterns for stylistics ? In A. F. GELBUKH, Ed., *Computational Linguistics and Intelligent Text Processing - 13th International Conference, CICLing 2012, New Delhi, India, March 11-17, 2012, Proceedings, Part I*, volume 7181 of *Lecture Notes in Computer Science*, p. 166–177 : Springer.