

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

Annie Foret (équipe LIS, département DKM)

IRISA, Université Rennes 1, France

## Objectifs

**TermLis** est un contexte d'information logique (**Lis**) construit à partir de ressources terminologiques pour une utilisation flexible et extensible 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, avec des indications sur la qualité des données.
- Un tel contexte peut être enrichi par d'autres informations (de natures diverses), en le reliant à d'autres applications (par des actions associées selon des arguments dans le contexte).

Nous illustrons, à travers une réalisation concrète sur des données de **FranceTerme**, les avantages d'une telle approche pour des données terminologiques.

## Introduction

Cette réalisation vise 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, or les systèmes d'information logique sont conçus pour permettre une navigation flexible dans des données organisées en contexte logique. Des travaux se situent dans un cadre similaire [1, 6, 4, 2], pour des données et buts de natures différentes.

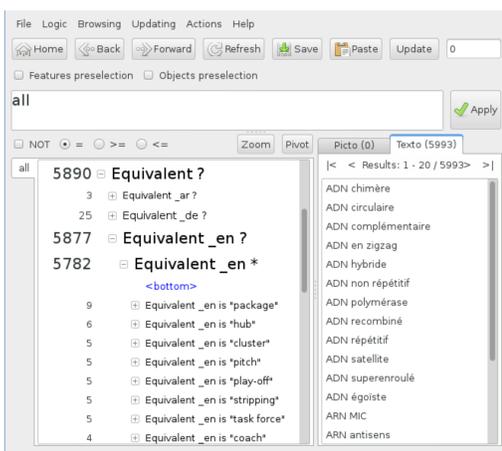


Figure 1: TermLis sans sélection ("all")

- Une vue en contexte logique comme **TermLis** 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** (seul outil de gestion de contexte logique disponible.)

## 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 ;
- avec un caractère attesté et règlementaire.

Des extensions ultérieures peuvent être envisagées : nouvelles données organisées selon un schéma analogue.

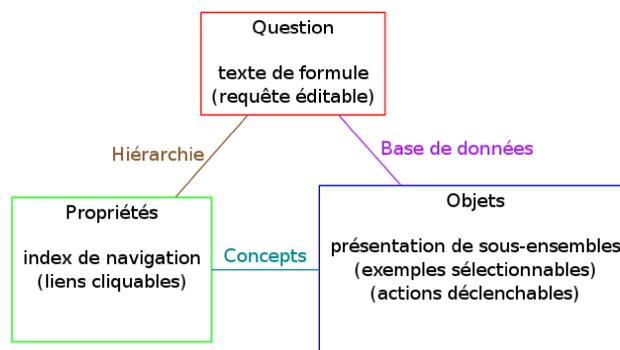


Figure 2: LIS et LCA

## Contexte logique

- Un *contexte logique* est défini par un ensemble fini  $\mathcal{O}$  d'objets  $o_i$  (de label  $l_i$ ) et pour chaque  $o_i$ , un ensemble fini de descriptions logiques  $d(o_i)$  pour un langage logique  $L$ .
- Un *système de gestion de contexte logique* permet de charger et d'exploiter un tel contexte, d'interroger par des requêtes logiques (explicites dans  $L$ , ou interactives) ; la réponse est alors un sous-contexte d'objets satisfaisant cette requête.

Nous avons utilisé **Camelis** (version 1)

<http://www.irisa.fr/LIS/ferre/camelis/> basé sur l'analyse de concept logique (LCA) [3].

LCA étend l'analyse de concept formel (FCA [5]) :

- 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 ceux qui vérifient  $int(c)$ .
- Ces concepts forment un treillis auquel correspond un *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 et de sources différents.

## Construction de TermLis.

Le contexte logique a été obtenu en réalisant un transducteur, appliqué au fichier source XML pour FranceTerme.

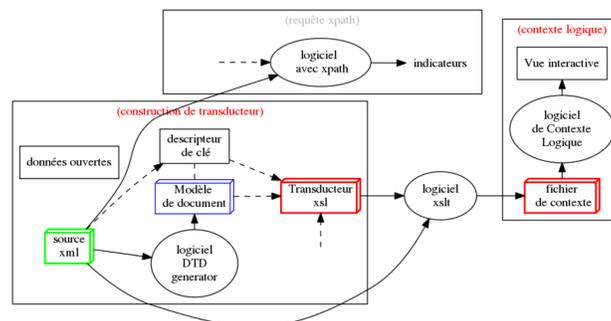


Figure 3: Transformation

Cette réalisation, s'est appuyée sur :

- un modèle du document XML
  - un xsd fourni (sans garantie) ;
  - une DTD générée automatiquement ;
- un modèle de contexte logique visé
  - un choix de types d'objets ;
  - une sélection de propriétés ;
- un lien entre les deux modèles,
  - la spécification d'une clé dans le source par une expression de chemin XML (ici `//Article[@id]` pour les codes d'Article).
  - et des expressions de contrôle (XPath)

## Utilisation du Contexte TermLis.



Figure 4: vue d'un sous-contexte

L'arbre de navigation à gauche permet des scénarios variés selon les facettes ouvertes et sélectionnées successivement

- avec divers types de données (chaines, dates etc.) ;
- combinaisons logiques ;
- variations, faux-amis éléments (ex: `<Attention>`) mis en évidence des défauts (redondances)

La *cohérence* est assurée entre les 3 fenêtres :

- texte de requête (formule logique) en haut ;
- présentation des objets (labels) 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 sont introduites par :

- des *axiomes* logiques pour des recherches avec inférences
- des *règles* ajoutant des propriétés ;
- des *actions* liant à d'autres ressources linguistiques ou traitements (analyseur,...).

Un découpage *modulaire* permet d'intégrer tout ou une sélection de ces possibilités.

## Quelques références

- P. Cellier, S. Ferré, M. Ducassé, and T. Charnois. Partial orders and logical concept analysis to explore patterns extracted by data mining. In *Conceptual Structures for Discovering Knowledge - 19th International Conference on Conceptual Structures, ICCS 2011*.
- I. Falk, D. Bernhard, and C. Gérard. From non word to new word: Automatically identifying neologisms in french newspapers. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*.
- S. Ferré and O. Ridoux. Introduction to logical information systems. *Inf. Process. Manage.*, 40(3):383–419, 2004.
- A. Foret and S. Ferré. On categorial grammars as logical information systems. In *Formal Concept Analysis, 8th International Conference, ICFA 2010*.
- B. Ganter and R. Wille. *Formal concept analysis - mathematical foundations*. Springer, 1999.
- S. Quiniou, P. Cellier, T. Charnois, and D. Legallois. What about sequential data mining techniques to identify linguistic patterns for stylistics? In *Computational Linguistics and Intelligent Text Processing - 13th International Conference, CICLing 2012*.

FranceTerme : <http://www.culture.fr/franceterme>

Logiciels Lis : <http://www.irisa.fr/LIS/software-fr>



- Web: <http://www.irisa.fr>
- Email: [foret@irisa.fr](mailto:foret@irisa.fr)