

# TER

Construction d'une ontologie à partir d'un corpus de textes avec l'ACF

February 4, 2017

## 1 vocabulaire

- ontologie : ensemble structuré des termes et concepts représentant le sens d'un champ d'informations, que ce soit par les métadonnées d'un espace de noms, ou les éléments d'un domaine de connaissances.
- concepts formels : ensemble disjoints objets/propriétés.

<https://hal.archives-ouvertes.fr/inria-00167678/document>

## 2 résumé

### 2.1 création d'une méthodologie semi-automatique

de construction d'ontologie à partir de corpus de textes sur un domaine spécifique.

### 2.2 on utilise :

- la classification d'objets d'après les propriétés partagées en utilisant l'analyse de concepts formels (ACF).pour la construction d'un treillis de concepts. Ce treillis va servir à construire un noyau d'ontologie.
- enrichir cette ontologie avec des relations transversales en utilisant une nouvelle méthode : l'analyse relationnelle de concepts (ARC). Chaque concept de l'ontologie résultante est défini puis représenté en Logique de Descriptions (LDs)

## 3 Introduction

### 3.1 système d'acquisition des connaissances

est important pour n'importe quel domaine spécifique. Il permet aux experts de raisonner sur les connaissances du domaine et de les partager. Néanmoins, il faut savoir que ces systèmes souffrent tous de ce qu'on appelle "goulot d'étranglement dans l'acquisition de connaissances" (Cimiano et al., 2005), c'est-à-dire la difficulté d'actualiser le modèle du domaine en question.

## **3.2 l'ontologie est construite en plusieurs étapes**

### **3.2.1 construction du noyau de l'ontologie**

Ce noyau représente la hiérarchie des concepts en s'appuyant sur la méthode de Cimiano (Cimiano et al., 2005) qui utilise l'Analyse de Concepts Formels (ACF) (Ganter, 1999), méthode formelle qui regroupe un ensemble d'objets d'après les propriétés qu'ils partagent et, de façon duale, regroupe un ensemble de propriétés d'après les objets qui les possèdent. De plus, les objets ne sont pas seulement définis par leurs propriétés, ils sont aussi définis par les relations qu'ils entretiennent avec d'autres objets.

### **3.2.2 prise en compte des relations transversales entre objets**

approche fondée sur une extension de l'ACF, l'Analyse Relationnelle de Concept (ARC), qui classe un ensemble d'objets célestes par rapport aux relations qu'ils entretiennent avec d'autres objets. Cette extension d'ontologie nous oblige à utiliser un langage en LDs plus expressif qui permet de prendre en compte les restrictions de relations

### **3.2.3 processus de fouille de textes**

### **3.2.4 hiérarchie de concepts construite avec l'ACF**

### **3.2.5 enrichissement du noyau d'ontologie avec l'ARC**

## **4 Méthodologie**

### **4.1 utilisation de la méthontology**

La "Methontology" construit une ontologie à partir des termes extraits des ressources (les ressources ne sont pas spécifiées) et a pour objectif de donner une définition à chaque concept et à chaque relation de l'ontologie dans un langage de LDs.

### **4.2 les étapes sont :**

#### **4.2.1 extraction de termes**

#### **4.2.2 construction du noyau de l'ontologie**

#### **4.2.3 extraction des relations transversales**

#### **4.2.4 regroupement des 2 modules pour obtenir l'ontologie complète**

## **5 Fouille de texte**

### **5.1 Objectif**

construire une hiérarchie de concepts

## **5.2 méthode de Malaisé**

utiliser des patrons définitoires pour extraire les définitions de chaque terme à partir du corpus de textes. Cette méthode donne de très bons résultats quand elle est utilisée pour un corpus de type dictionnaire(terme : définition ou terme1 est un terme2), mais pas pour un corpus de textes tel que celui de l'astronomie, car il est difficile de trouver une définition à chaque objet.

## **5.3 méthode de Harris**

l'étude des régularités syntaxiques dans un corpus de sous-langage (ou langage spécialisé) permet d'identifier des schémas syntaxiques formés de combinaisons de classes qui reflètent les connaissances du domaine traité(Faure & Nedellec, 1998)

# **6 Construction du noyau de l'ontologie**

2 types de travaux pour la construction d'une hiérarchie de concepts à partir de corpus de textes :

- les co-occurrences des termes dans les textes
- nature symbolique

## **6.1 passage du treillis à l'ontologie et étiquetage par des experts**

## **6.2 représentation des concepts en logique de description**

# **7 Extraction des relations transversales**

## **7.1 analyse formelle de concepts**

L'idée est de construire un contexte formel pour chaque relation transversale extraite du texte

## **7.2 représentation des concepts en logique de descriptions**