

# Catégorisation et Désambiguïsation des Intérêts des Individus dans le Web Social

Coriane Nana Jipmo\*, Gianluca Quercini\*  
Nacéra Bennacer\*

\*Laboratoire de Recherche en Informatique (LRI)  
CentraleSupélec, Univ. Paris-Saclay  
Gif-sur-Yvette, France

{coriane.nanajipmo, gianluca.quercini, nacera.bennacer}@lri.fr

**Résumé.** Cet article présente une approche pour la catégorisation et la désambiguïsation des intérêts que les individus renseignent sur les réseaux sociaux en utilisant Wikipédia.

## 1 Introduction

Dans cet article, nous présentons une étude préliminaire sur le problème de caractérisation et de catégorisation des intérêts que les individus renseignent sur les réseaux sociaux, comme *reading*, *jogging*, *java*, etc. Ces intérêts étant exprimés en langage naturel, nous sommes confrontés au problème de la désambiguïsation dans un contexte limité, compte tenu du peu des informations accessibles dans les profils des individus. Les approches visant la désambiguïsation des *tags* dans les *folksonomies* sont confrontées aux même problème, bien qu'elles peuvent s'appuyer sur les ressources faisant l'objet des tags pour avoir un contexte plus riche [Garcia-Silva et al. (2012)]. Nous explorons une approche permettant de désambiguïser un intérêt d'un individu par la détermination d'un article Wikipédia qui contient la description de celui-ci. La désambiguïsation d'un intérêt se fera en utilisant les autres intérêts renseignés par l'individu comme contexte. Les résultats que nous avons obtenus sur 392 intérêts issus de 50 profils utilisateurs du réseau social *LiveJournal* sont encourageants.

## 2 Désambiguïsation des intérêts

Soit  $\mathcal{I}_u = \{I_1, I_2, \dots, I_n\}$  l'ensemble des intérêts qu'un individu  $\alpha$  renseigne sur son profil, exprimés sous la forme d'une chaîne de caractères en Anglais (*Computer*, *Music*, ...). Notre objectif est d'associer à chaque intérêt  $I_j$ ,  $j = 1, \dots, n$ , un article Wikipédia décrivant  $I_j$ ; à cet effet, pour chaque intérêt  $I_j$  notre approche sélectionne la page Wikipédia  $P$  ayant par titre  $I_j$ , si elle existe. Deux cas peuvent se présenter : (i)  $P$  est un article décrivant  $I_j$ , ou (ii)  $P$  est une page de désambiguïsation. Dans le premier cas, le mot décrivant l'intérêt  $I_j$  n'est pas ambigu (par ex., *Music*), au point que la majorité des utilisateurs Wikipédia ont trouvé un accord quant à son *interprétation* (ou, signification) par défaut (*Music* désigne une forme d'art) et lui ont associé un article. La page  $P$  sera alors choisie comme la seule interprétation de l'intérêt  $I_j$ .