MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ DJILLALI LIABES SIDI BEL ABBES

FACULTÉ DES SCIENCES DE L'INGÉNIEUR

RESUMÉ DE MÉMOIRE DE MAGISTER

Nom & Prénom(s)	Messaoudi Kaddour
E-mail (obligatoire)	meskadinf@hotmail.com
Spécialité	Informatique Option : Systèmes d'Information Web (SIW)
Titre	Matching de contenu des résumés des textes
Date de soutenance	18 Mai 2014
Nom, prénom(s) et grade de l'encadreur	MCA. Lehireche Ahmed

Résumé :
L'avènement de l'informatique et l'accroissement du nombre de documents électroniques stockéssur les divers supports électroniques et sur le Web ; Enparticulier les données textuelleont permis ledéveloppement d'outils d'analyse et de traitement automatique des textes, notamment la classificationautomatique. Classer des textes est devenu indispensable pour assister les utilisateurs de ces collections dedocuments. La classification permet d'explorer etde répertorier toutes ces immenses banques de donnéestextuelles.
Notre travail est essentiellement la classification des documents basée sur une technique de similarité du contenu.
En premier, nous avons utilisé les techniques standards du Text Mining ; En semi supervisé ne considérant que deux classes. Cette première étape est assujettis a un comportement structurelle .i.e. la technique du TEXT MINING est une technique syntaxique.
Nous avons conçu une applicationqui utilise les techniques de similarité et s'appuie principalement sur deux types d'algorithmes, Cosinus et Manhattan pour calculer le score de similarité entre tous les résumés. Ensuite nous avons classé également ces résumés selon la valeur de leurs scores en deux classes : classe des résumés similaires et non similaires.
Comme deuxième travail, Nous avons étudié le problème du « NOT » sur deux corpus qui ne sont différencié que par la négation. Évidement la technique du TEXT MINING ne détecte pas la négation qui est un problème de CONTENU.
Nous proposons une approche pour résoudre le Problème de la négation et une implémentation.

Mots clés :
Classification, similarité syntaxique, Catégorisation, fouille de texte, similarité sémantique, la négation.

Abstract
The advent of computers and the increasing number of electronic documents stored on various electronic media and the Web, in particular textual data has enabled the development of analysis and automatic text processing tools, including Automatic classification. Classifying texts has become essential to assist users of these collections of documents. The classification allows exploring and identifying all these huge banks of textual data.
Our work is essentially the classification of documents based on similarity techniques of the "content".
First, we used standard techniques of Text Mining; a semi supervised one, considering only two classes. This first step is subject to a structural behavior .i.e. the TEXT MINING technique is syntactic.
We designed an application that uses the techniques of similarity and relies primarily on two types of algorithms, Cosine and Manhattan to calculate the score of similarity between all the texts abstracts. Then we also rated them by mean of the value of their scores in two classes: Class of similar and dissimilar abstracts.
As a second work, we studied the problem of the "NOT" using only two corpora that are differentiated only by negation. It is evident that the technique TEXT MINING does not detect the negation which is a problem of CONTENT.
We propose an approach to solve the problem of negation and an implementation.

Keywords :

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ DJILLALI LIABES SIDI BEL ABBES

FACULTÉ DES SCIENCES DE L'INGÉNIEUR

RESUMÉ DE MÉMOIRE DE MAGISTER

Classification, similarité syntaxique, Catégorisation, fouille de texte, similarité sémantique, la négation.