

RESUMÉ DE MÉMOIRE DE MAGISTER

Nom & Prénom(s)	BENNABI Rim Sakina
E-mail (obligatoire)	rym_avenir@yahoo.fr
Spécialité	Informatique.
Titre	Fouille de Texte sur Grille
Date de soutenance	16/12/2013
Nom, prénom(s) et grade de l'encadreur	Dr. ELBERRICHI Zakaria MCA

Résumé :

Le besoin continu d'acquisition de nouvelles connaissances a poussé les différents organismes (scientifiques, gouvernementaux, éducatifs, commerciaux,...) vers les solutions de fouille de données. Ces données sont souvent sous une forme textuelle et nécessitent donc un prétraitement spécifique avant d'être traitées sur machine.

La fouille de texte, ou l'application du data mining sur les documents textuels permet d'extraire des informations utiles et non triviales à partir de collections de documents textuels souvent complexes et volumineuses. Or, les applications de fouille de texte ont rapidement évolué rendant dans certains cas, les systèmes et les infrastructures conventionnels incapables de répondre aux exigences de telles applications. Une solution à ce problème serait de recourir aux calcul distribué et plus précisément aux « grilles de calcul » afin de bénéficier d'une grande puissance de calcul à travers l'intégration et le partage de plusieurs ressources computationnelles et de stockage se trouvant sur des machines localement reliées ou géographiquement distribuées.

La grille permet de réaliser de nouvelles applications de fouille de texte, et fournit une infrastructure qui répond aux besoins - en termes de scalabilité et de distribution - des tâches d'extraction de connaissance à partir de texte.

Mots clés :

Fouille de texte, grille de calcul, prétraitement, fouille de données, KPPV

Abstract

The continuing need for new knowledge has led various organizations (scientific, governmental, educational , commercial, ...) to the data mining solutions. These data are often in textual form and therefore require special pre-treatment before being processed on machine.

Text mining, or the application of data mining on textual documents can extract useful and non-trivial information from complex and voluminous textual collections. However, text mining has evolved rapidly making in some cases, systems and conventional infrastructure unable to meet the requirements of such applications. One solution to this problem would be to use distributed computing and more precisely " grid computing " to benefit from high computing power through integration and sharing of several computational and storage resources located on machines connected locally or geographically distributed .

The grid enables new applications of text mining, and provides an infrastructure that meets the needs - in terms of scalability and distribution - of knowledge extraction tasks from text.

Keywords :

text mining, grid computing, preprocessing, data mining, KNN

©©©©

استمرار الحاجة إلى معرفة جديدة، أدت المنظمات المختلفة (العلمية و الحكومية و التعليمية و التجارية ، ...) إلى حلول تنقيب البيانات. هذه البيانات هي في كثير من الأحيان في شكل نصي، وتتطلب المعالجة المسبقة قبل أن يتم معالجتها آليا .
تنقيب النصوص، أو تطبيق تنقيب البيانات على الوثائق النصية، يمكن من استخراج معلومات مفيدة و ذات أهمية من المجموعات النصية المعقدة و الضخمة. ومع ذلك، فقد تطور تنقيب النصوص بسرعة جاعلا في بعض الحالات الأنظمة و البنية التحتية التقليدية غير قادرة على تلبية متطلبات هذه التطبيقات . الحل لهذه المشكلة يكمن في استخدام الحوسبة الموزعة و على نحو أدق " الحوسبة الشبكية " للاستفادة من قوة الحوسبة العالية من خلال التكامل و تقاسم المصادر الحاسوبية و التخزينية العدة المتواجدة على آلات متصلة محليا أو موزعة جغرافيا.
الشبكة تنتج تطبيقات جديدة من تنقيب النصوص، و تمنح البنية التحتية التي تلبى الاحتياجات - من حيث قابلية التوسع و التوزيع - من مهام استخراج المعرفة من النص.

كلمات مفتاحية

تنقيب النصوص، الحوسبة الشبكية، المعالجة المسبقة، تنقيب البيانات، ك-أقرب جار .