



اكتشاف المعلومات من البيانات الكبيرة على الإنترنت بواسطة تقنية الوكيل الدلالي

إعداد
رانية علي ناصر بافكير

بحث مقدم لنيل درجة الماجستير في علوم الحاسبات

إشراف
أ.د. كمال منصور جمبي

كلية الحاسبات وتقنية المعلومات
جامعة الملك عبدالعزيز
جدة - المملكة العربية السعودية
شعبان ١٤٣٨ هـ - مايو ٢٠١٧ م

اكتشاف المعلومات من البيانات الكبيرة على الإنترنت بواسطة تقنية الوكيل الدلالي

المستخلص

نظم المعلومات التقليدية غير قادرة على التعامل بكفاءة مع جميع مصادر البيانات الضخمة وسياقات المعلومات المتعددة التي تعتمد على الإنترنت كمنصة أساسية لها . حيث تواجه هذه النظم التقليدية العديد من المشاكل في استرجاع وتوحيد وتخزين ومعالجة واستخدام المعلومات الناتجة عن مصادر المعلومات المختلفة و غير المتجانسة و التي هي أساس العديد من المنظمات في عملية اتخاذ القرار . وبالرغم من أنه قد تم تطوير العديد من نظم المعلومات و الهياكل لإستخراج المعلومات ذات الصلة من شبكة الإنترنت، إلا أن الكثير منها يعتمد فقط على الهياكل النحوية في عملية استخراج المعلومات و التي تؤدي إلى نتائج غير مرضية . في حين أن البعض الآخر يستخدم تكنولوجيا الإنترنت الدلالي التقليدية، والتي تفقد المعنى الدلالي بسبب بعض التحديات والقيود المفروضة التي يجب التغلب عليها .

لذلك في هذا البحث , سوف نقوم بإنشاء مقترح لإكتشاف المعلومات من البيانات الكبيرة على الإنترنت بواسطة تقنية الوكيل الدلالي من أجل تحسين النتائج وذلك من خلال إيجاد قواعد جديدة تساهم في تقليل الوقت المستغرق لعملية الاكتشاف مع المحافظة على معاني المعلومات المستخرجة.

في هذه الأطروحة، قمنا بتطوير نهج الوكيل الدلالي الموحد القائم على علم الأنماط و(ontology) لتعزيز عملية اكتشاف المعرفة من البيانات الكبيرة على شبكة الإنترنت. حيث يستخرج النموذج المعلومات من مصادر البيانات المختلفة المنظمة وغير المنظمة ويعمل على تخزينها في قواعد بيانات موحدة مما يسهل عملية البحث والتحليل واعداد التقارير. كما يساعد على استخراج نتائج أكثر كفاءة وفعالية من خلال فهم طلب المستخدمين والمعنى السياقي للمصطلحات باستخدام ontology.

ولتقييم هذا العمل المقترح، أجريت العديد من التجارب لتقييم الأداء، ودقة النظام باستخدام مقاييس الأداء المختلفة. ولقد أظهرت النتائج أن استخدام تقنية ontology حسن من نسبة كفاءة الأداء ودقة النتائج بشكل أفضل من طريقة اكتشاف المعلومات التي تستند إلى البحث عن الكلمات الرئيسية.



Semantic Agent Based Knowledge Discovery from Big Data on the Web

By
Raniah Ali Nasser Bafaqeer

**A Thesis Submitted in Partial Fulfillment of the
Requirements for the Master Degree in Computer Science**

Supervised By
Dr. Kamal Jambi

**FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY
KING ABDULAZIZ UNIVERSITY
JEDDAH – SAUDI ARABIA
Shaban 1438 H – May 2017 G**

Semantic Agent Based Knowledge Discovery from Big Data on the Web

ABSTRACT

Nowadays, the emergence of Big Data has evolved from the fast rapid growth of the web resource on the internet. The size of the web becomes too massive, heterogeneous, and updated day by day, impacted by the fast development of networking, data collection capacity and data storage.

Data is the new oil, and like oil, data needs to be refined before its get value. Using Big Data analytics, organizations can extract usable information out of enormous, complex, interconnected and varied data sets leading to valuable insights. The traditional information systems are unable to cope efficiently with all new data sources and multiple contexts of information. Challenges are encountered in retrieving, standardizing, storing, processing and usage of information generated by various heterogeneous sources that are the basis for enabling systems for decision support organizations. The key to solving these problems lie in the establishment of a unified data model that either develops new algorithms that cope with heterogeneous data between distributed sites or new tools for preprocessing data to make them structured to fit existing algorithms.

We developed a unified semantic agent approach to enhance the knowledge discovery process from Big Data on the Web using ontology. The model extracts information from heterogeneous data sources, then defines a unified structure to store them that make it easily searchable, maximizing discovery and reporting for users. It gave the user more efficient and effective results by understanding the users' intent and the contextual meaning of terms as they appear in the searchable data space using ontology.

We conducted a series of experimental test. The test yielded the following results: using Ontology technique, have higher result's precision and recall than using the keyword based information discovery techniques because of the semantics of the keywords.

