تصميم معدل صفحات باستخدام تقنية الويب الدلالية

رسالة أعدت لحيل درجة الماجستير في هندسة التحكم الآلي والحواسب

إعداد المهندس
ماهر سلامة

بتفرج

مشترث أساسي
د. م. نبيل عبس (أستاذ مساعد)

المهندس منصور صويص (مدرس)

1434هـ - 2013م
تصميم مدخل صفحات باستخدام تكنولوجيا الويب الدلالية

م. ماهر حلمي1 د. بلال عباس2 د. منصور صبص3

ملخص البحث

تم في هذا البحث تصميم وبناء نظام حاسوبي لاستخراج المعلومات من مواقع الويب وتضمينها بصيغة Resource Description Framework (RDF) منصة وصف المصدر، إذ تتضمن هذه الصيغة مجموعة من التلاليات متفاوتة عددها وتوزعها اعتمادًا على مجال عمل النظام والمعلومات المراد استخراجها من الصور، لغرض تطبيق تلك التلاليات من الممثل النصي غير المهبل إلى الممثل المهبل، مما يسهل التعامل مع وإجراء العمليات عليه.

المعلومات التي سيتم استخراجها من موقع الويب تتضمن في الآتي:

- التلاليات: مثل أسماء الأشخاص والأماكن والشركات التي تم ذكرها مع مراعاة أن تكون هذه العملية موصوفة بشكل يتيح لنا توصية مجال عملنا دون تحويل التطبيق.
- وصف النص: أي المجال الذي يتعلق به النص (أعمال، تفتيتات،...).
- تحديد درجة الثباية بين صفحات الويب المختلفة.
- معلومات مستردة من بيئة المستند نفسه: مثل اسم الكاتب، الكلمات المفتاحية.

الكلمات المفتاحية: الويب الدلالي، ثلاثيات RDF، أنظمة استخراج المعلومات، المكتبات، اللغة الموسعة XML
DESIGNING AND IMPLEMENTING Web Pages Parser Using Semantic Web Technology

E. Maher Salama¹  Dr.Eng. Nabil Abbas²  Dr.Eng. Monsef Soes³

Summary of Research

The purpose of the Information Extraction (IE) systems structure is that it depends on the processing of the natural language. In addition to the use of the data structure itself considering that the system input is Web pages written in the HTML language, while the output is the semantics which have been extracted from these pages as RDF triples using the semantic Web. The final result is stored either as normal text or in the data storage AllegroGraph. The information extracted from the Web page can be summarized as:

- Entities, such as, the names of people, places, and companies which have been mentioned. It should be taken into consideration that this operation is described in a way which allows a kind of scalability without any change of the application.
- The text description, i.e. the domain which relates to the text (news, techniques.....).
- Similar pages and the degree of the similarity between them.
- Extracted information from the structure of the page itself, such as, author’s name and key words.

Key Words: Semantic Web, RDF Triples, Information Extraction Systems, Ontology, XML

¹ Master Student – Auto Control and Computers Dept.-Faculty of Mech. and Elec. Eng.-Albaath Univ.
² Dr.E. in Auto Control and Computers Dept.-Faculty of Mech. and Elec. Eng.-Albaath Univ.
³ Dr.E. in software eng. Dept.-Faculty of IT Eng.-Albaath Univ.