شبكة المعلومات الجامعية

بسم الله الرحمن الرحيم

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الالكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها

علي هذه الأفلام قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأفلام بعيدا عن الغبار

في درجة حرارة من ١٥-٢٥ مئوية ورطوبة نسبية من ٢٠-٤٠%

To be Kept away from Dust in Dry Cool place of 15-25- c and relative humidity 20-40%

بعض الوثائق الاصلية تالفة

بالرسالة صفحات لم

ترد بالاصل

Faculty of Computers and Information
Cairo University

Extracting Ontology from Different Sources

By
Mohammed Yehia Hassan Dahab

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Doctor of Philosophy
Computer Science

Supervised by

Prof. Dr. Ahmed Rafea
Prof. Dr. Khaled Shaalan
Dr. Hesham Hassan

2007

A thesis submitted to the faculty of Computers and Information, Cairo University, in partial fulfillment of requirements for the degree of Doctor of Philosophy of Computer Science in the department of Computer Science.

I certify that this work has not been accepted in substance for any acadmic degree and is not being concurrently submitted in candidature for any other degree.

Any portion of this thesis for which I am indebted to other sources are mentioned and explicit references are given.

Student: Mohammed Yehia Hassan Dahab

Approval Sheet

Extracting Ontology from Different Sources

By
Mohammed Yehia Hassan Dahab

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy
Computer Science

Faculty of Computer and Information
Cairo University

This thesis is for the Ph.D. Degree in Computer Science, Department of Computer Science, Faculty of Computer and Information, Cairo University, has been approved by:

Name	Signature
Prof.Dr. Ibrahim Farag	
Prof.Dr. Ahmad Rafea	
Prof.Dr. Ayman Al-Dosokey	
Dr. Hesham Hassan	

ACKNOWLEDGEMENTS

In general writing a thesis and doing the research that comes with it is a solitary business. Fortunately therefore it is good to have a committee that is willing to read about and to listen to your descriptions of the work you have done so far In this respect I would especially like to thank Prof. Ahmed Rafea and Dr. Hesham Hassan for their feedback which they provided me. Both of them gave me a lot of direction by making me realize which parts of my research were potentially interesting.

Other people, who were in my committee, as colleagues or friends, gave me feedback also and I would like to thank them for that. Dr. Bassam Bayaa probably would deserve a separate paragraph in order to thank him for his help.

Finally I want to express my gratitude to my wife and my family for supporting me to accomplish this work.

ABSTRACT

Ontologies have become a popular research topic and have been investigated by several artificial intelligence research communities, including knowledge engineering and management, natural language processing and knowledge representation. Ontologies meet a major demand in these fields: they establish shared and common understanding of a domain to facilitate communication among different software applications and users.

Ontology construction, in large, complex application domains, can be lengthy, costly and controversial. Most of existing ontologies construction tools support construction of ontological relations (e.g. taxonomy, equivalence, etc.) but they do not support construction of domain relations, non-taxonomic conceptual relationships, (e.g. causes, caused by, has-member, contain, etc.). The aim of this thesis is to investigate the automation of the ontology building process from structured and unstructured sources. This thesis also aims at proposing approaches related to merging, integrating and maintaining extracted or existing domain ontologies.

To achieve the thesis objectives we surveyed the most important ontology representation languages and tools and developed tools for extracting ontologies from text, knowledge bases, thesauri and semantic lexicons. One of these tools has been developed to extract taxonomy of plant diseases and insects that affect plant parts using a knowledge base, a thesaurus for agricultural terms (AGROVOC) and a semantic lexicon (WordNet). A second tool has been developed to extract relations between symptoms on plant parts and their causes namely diseases and insects, through analyzing available text on the web and scientific papers.

The methodologies and tools have been applied and tested for extracting different ontologies and integrating them automatically to form one general ontology. Results showed the success of the proposed methodologies and tools.

Contents