b152593

A HYBRID APPROACH FOR NATURAL LANGUAGE DATABASE QUERY TRANSLATION

Pornpimon Teekayuphun

A Thesis Submitted in Partial
Fulfillment of the Requirements for the Degree of
Master of Science (Applied Information System)
School of Applied Statistics
National Institute of Development Administration

2006

A HYBRID APPROACH FOR NATURAL LANGUAGE DATABASE QUERY TRANSLATION

Pornpimon Teekayuphun

School of Applied Statistics

The Examining Committee Approved This Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science (Applied Information System)

Assistant Professor ..... Committee Chairman

(Ohm Sornil, Ph.D.)

Associate Professor ..... Committee

(Pipat Hiranvanichakorn, Ph.D.)

Associate Professor ..... Committee

(Surapong Auwatanamongkol, Ph.D.)

Associate Professor ..... Acting Dean

(Surapong Auwatanamongkol, Ph.D.)

Date 11 May 2007

ABSTRACT

Title of Thesis	A Hybrid Approach for Natural Language Database Query Translation
Author	Ms. Pornpimon Teekeyuphun
Degree	Master of Science (Applied Information System)
Year	2006

Natural language interface to database allows users to access information stored in a database by supplying requests expressed in some natural language to the database system. In this paper, an approach for translating queries into Structured Query Language is presented in the context of an injury surveillance database, where queries and data are expressed in Thai language.

The process consists of three major phases: keyword formation, keyword sense disambiguation, and Structured Query Language query formulation. The system attempts to extract keywords, groups of words or phrases which have specific meaning in a database context, by grouping contiguous words within the user input. A technique from data compression, Prediction by Partial Matching, is applied to extract keywords from the user input. However, the keywords may refer two or more semantic objects in the database domain. The ambiguities of keywords are resolved by using a random walk with restart algorithm. After extracting all keywords from the user input, the algorithm is applied to generate Structured Query Language statement to retrieve the information in the database.

Each phase and the overall process are evaluated in details. The results show that the overall query translation achieves an accuracy of 87%.

ACKNOWLEDGEMENTS

The author would like to express sincere thanks to my major advisor, Dr.Ohm Sornil, for his valuable suggestion, continuous support, encouragement and guidance in making this thesis a successful one. Besides my advisors, I would like to thank the rest of my thesis committee, Dr. Pipat Hiranvanichakorn and Dr. Surapong Auwatanamongkol, for their thoughtful comments and suggestions.

I also wish to extend thanks and appreciation to Dr.Witaya Chadbunchachai and all staffs in Secretariat Office of Trauma Center, Khon Kaen Hospital for their kindness and cooperation in supplying the data for this thesis.

Pornpimon Teekayuphun

March 2007