Handbook of Research on Natural Language Processing and Smart Service Systems

Rodolfo Abraham Pazos-Rangel
Tecnológico Nacional de México, Mexico & Instituto Tecnológico de Ciudad Madero, Mexico

Rogelio Florencia-Juarez
Universidad Autónoma de Ciudad Juárez, Mexico

Mario Andrés Paredes-Valverde
Tecnológico Nacional de México, Mexico & Instituto Tecnológico de Orizaba, Mexico

Gilberto Rivera
Universidad Autónoma de Ciudad Juárez, Mexico

IGI Global
PUBLISHER of TIMELY KNOWLEDGE

A volume in the Advances in Computational Intelligence and Robotics (ACIR) Book Series

Published in the United States of America by
IGI Global
Engineering Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Names: Pazos-Rangel, Rodolfo Abraham, 1951- editor.
Title: Handbook of research on natural language processing and smart service systems / Rodolfo Abraham Pazos-Rangel, Rogelio Florencia-Juarez, Mario Andrés Paredes-Valverde, Gilberto Rivera, editors.
Description: Hershey, PA : Engineering Science Reference, an imprint of IGI Global, [2020] | Includes bibliographical references and index. | Summary: "This book is a collection of innovative research on the integration and development of intelligent software tools and their various applications within professional environments"-- Provided by publisher.
Identifiers: LCCN 2019058351 (print) | LCCN 2019058352 (ebook) | ISBN 9781799847304 (hardcover) | ISBN 9781799847311 (ebook)
Subjects: LCSH: Natural language processing (Computer science) | Natural language generation (Computer science) | Computational linguistics.
Classification: LCC QA76.9.N38 H3645 2020 (print) | LCC QA76.9.N38 (ebook) | DDC 006.3/5--dc23
LC record available at https://lccn.loc.gov/2019058351
LC ebook record available at https://lccn.loc.gov/2019058352

This book is published in the IGI Global book series Advances in Computational Intelligence and Robotics (ACIR) (ISSN: 2327-0411; eISSN: 2327-042X)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

List of Contributors

Table of Contents

Chapter 4
Two New Challenging Resources to Evaluate Natural Language Interfaces to Databases Generated Based on Geobase and Geoquery

Juan Javier González-Barbosa
Tecnológico Nacional de México, Mexico & Instituto Tecnológico de Ciudad Madero, Mexico

Juan Frausto Solís
Tecnológico Nacional de México, Mexico & Instituto Tecnológico de Ciudad Madero, Mexico

Juan Paulo Sánchez-Hernández
Universidad Politécnica del Estado de Morelos, Mexico

Julia Patricia Sanchez-Solís
Universidad Autónoma de Ciudad Juárez, Mexico

ABSTRACT

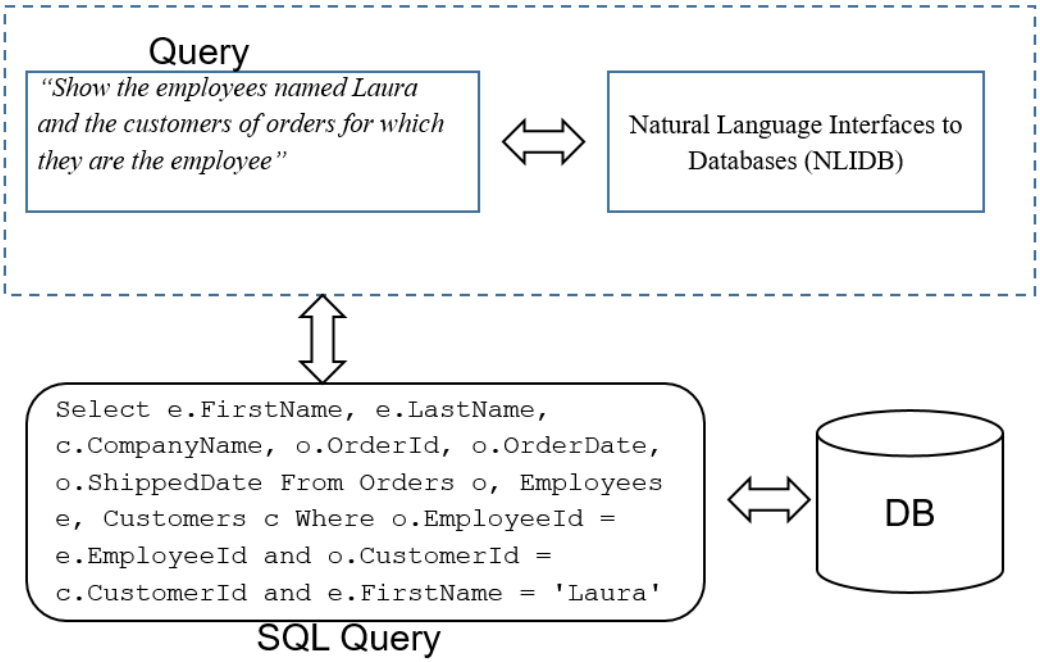
Databases and corpora are essential resources to evaluate the performance of Natural Language Interfaces to Databases (NLIDB). The Geobase database and the Geoquery corpus (Geoquery250 and Geoquery880) are among the most commonly used. In this chapter, the authors analyze both resources to offer two elaborate resources: 1) N-Geobase, which is a relational database, and 2) the corpus Geoquery270. The former follows the standard normalization procedure, then N-Geobase has a schema similar to enterprise databases. Geoquery270 consists of 270 queries selected from Geoquery880, preserving the same kind of natural language problems as Geoquery880, but with more challenging issues for an NLIDB than Geoquery250. To evaluate the new resources, they compared the performance of the NLIDB using Geoquery270 and Geoquery250. The results indicated that Geoquery270 was the harder corpus, while Geoquery250 is the easier one. Consequently, this chapter offers a broader range of resources to NLIDB designers.

DOI: 10.4018/978-1-7998-4730-4.ch004

INTRODUCTION

An NLIDB (Natural Language Interface to Database) is a system that can be used to access information in a database by typing sentences in a natural language such as English, French, or any other (Androutsopoulos, Ritchie, & Thanisch, 1995). For example, the sentence on the left side of **Figure 1** is the query "Show all the employees called Laura and the customers whose orders have been generated by these employees"; for this sentence, the NLIDB on the right side of this figure will generate an equivalent query in SQL (Structured Query Language) to extract the information demanded by the user.

Figure 1. A simple query processed with a Natural Language Interface to Database (NLIDB)



Notable success on modern techniques to translate queries in natural language to SQL has been developed. From the 1960s to the 1990s, the first applications of NLIDBS were published, examples follow:

- In 1961, BASEBALL was introduced for a database about baseball games (Green Jr, Wolf, Chomsky, & Laugh, 1961).
- In 1972, LUNAR was applied for lunar rocks and soil composition (Woods, Kaplan, & Nash-Webber, 1972).
- In the 1980s, CHAT-80 was implemented in Prolog, and its database integrated by facts (true sentences in Prolog) relating to 150 of the world's countries (Warren, 1981).
- In 1987, TEAM used a database that comprises geographic data (Grosz, Appelt, Martin, & Pereira, 1987).