

Natural Language Interfaces for Data Warehouses

Nicolas Kuchmann-Beauger* and Marie-Aude Aufaure**

*SAP Research – 157/157 rue Anatole France – 92309 Levallois-Perret Cedex
nicolas.kuchmann-beauger@sap.com,

**École Centrale Paris – Grande Voie des Vignes – 92290 Châtenay-Malabry
marie-aude.aufaure@ecp.fr

Abstract. Data warehouses process a huge mass of data, that are often modeled in a way humans hardly understand. In order to make the search paradigm more accessible to end users, some efforts have been made in the field of Business Intelligence. However, expressing information need as a structured query is still an artificial task. This explains why “natural” approaches are preferred nowadays. In this paper, we present a Question Answering (Q&A) system for structured data in a context of BI. Some benefits of our proposal are described through an iPhone/iPad application and through an HTML prototype.

1 Introduction

In recent years, a new class of search systems, such as Wolfram|Alpha¹, became popular: question answering systems that operate on well curated knowledge bases and thus deliver highly reliable answers to user questions. Industrial solutions have been designed by the BI community to support users in formulating queries. However, the query models are often oversimplistic compared to the complex questions that users have in mind. Query interfaces in commercial BI tools allow users to combine dimensions and measures to build queries.

New device capabilities (increased smartphones and tablets memory and storage) as well as the generalization of Internet access to mobile devices bring new challenges. In particular, recent success of speech-to-text technologies like Siri² has made natural search interfaces popular. This observation, and the fact that users are more comfortable to use such query interfaces compared to very structured ones have been pointed out by Hearst (2011). As a result, there is a strong requirement from non-expert employees (e.g., sales representatives) to access their data on their mobile devices and to formulate queries without any knowledge on the underlying query language. In addition, mobile devices bring also new information about users (e.g., geographic location). Such information are processed by context-aware algorithms.

As a proposition to overcome the shortcomings of existing BI tools, and taking into account new trends in terms of context sharing and mobility in information retrieval, we present in this paper an answering system for data in warehouses. The remaining of the paper is structured as follows: Section 2 presents hierarchies in multidimensional analysis, and Section 3 introduces

1. designed by Mathematica, see <http://www.wolfram.com/mathematica/>.

2. For more information about Siri: <http://www.apple.com/iphone/features/#siri>