Chapter 3.5
The Weather Tool:
An Agent-Based Approach to Information Integration

Gunjan Kalra
Quantum Leap Innovations, USA

ABSTRACT

This chapter discusses the process of providing information in its most accurate, complete form to its users and the difficulties faced by the users of the current information systems. The chapter describes the impact of prevalent technologies such as the multi-agent systems and the Semantic Web in the area of information supply via an example implementation and a model use case. The chapter offers a potentially more efficient and robust approach to information integration and supply process. The chapter intends to highlight the complexities inherent in the process of information supply and the role of emerging information technologies in solving these challenges.

INTRODUCTION

The progress in distributed computing has played a significant role in facilitating replacement of large, monolithic databases with smaller and conceptually self-contained databases. This has resulted in efficient yet "ever-expanding" clusters of heterogeneous sources of data (*data sources*) such as the traditional databases, Web sites, e-mails, and operational applications. The need to view and analyze data from various data sources unified under one application context has become more prevalent and has gained prominence among data analysts, both in the information technology industry and academia. For example, large-scale enterprises have a growing demand for enterprise integration products, while data mining researchers need a unified view of data from different sources to select and test their algorithms. The task of providing integrated data management within one umbrella is particularly challenging. In this chapter, we describe the need and evolution of the Weather Tool system, a weather data provisioning and integration system accessible via a standard application programming interface (API). The chapter begins with an introduction to the evolution of sources of data and to the problem of weather data integration. Some of the previous

and existing data integration technologies and products are then summarized. The chapter will also address some of the new technologies that the data/information integration systems benefit greatly from, including the agent technology and the Semantic Web. We then provide an overview of the Weather Tool architecture (and an example application) that enables integration of several data sources to provide a unified data environment. The architecture discussion describes how *data processing services* are provisioned in a dynamically-changing environment. The advantages and pitfalls of such a system are elaborated, followed by a discussion of potential improvements in the architecture. The chapter briefly discusses the core ideas of Quantum Leap Innovation, Inc. (QLI) intelligent data management (IDM) framework, the generalized, enhanced version of the Weather Tool effort. This framework is aimed at performing beyond the Weather Tool's single repository approach to data integration by incorporating several strategies for on-demand, data access and analysis remotely, thereby avoiding creation of redundant intermediary repositories. Finally, the chapter highlights the need, evolution, and importance of data/information integration efforts.

EVOLUTION FROM LARGE DATABASES TO LARGE DATA ENVIRONMENT

Historically, due to high set-up and maintenance costs for databases, extremely large volumes of data were stored in one database; this strategy was considered efficient and had the advantage of easy access to all data. In reaction to the scale and complexity of such large databases, data analysts have often created and exploited selective views of subsets of the large databases. Since the advent of cheaper, flexible, and more efficient data-storage solutions, the large databases have made way for individual databases that store smaller sets of relevant data. These databases have varying data

access mechanisms and protocols that are usually dependent upon non-standard, product specific technology. Some examples of commonly-used relational databases are MySQL, Oracle, Postgres, and DB2.

Data-access mechanisms for the databases vary in complexity of implementation and user-friendliness. The most common, structurally straightforward but tricky to use data-access mechanism is Structured Query Language (SQL) that allows formation of data queries via a set of special commands and keywords. Relatively more expressive and easy-to-use approaches involve the graphical user interface (GUI) applications developed using technologies such as Visual Basic, Java, and so forth. However, these involve long implementation times. More recent tools that have been developed for easier implementation, use, and data access are the Web sites that enable access to remote databases via Web pages. The Web sites/portals that are easily customized to create data-access forms have become an important source of information in today's world. There are examples of Web sites functioning not only as data-access mechanisms but also as sources of intermediate transient data that can become crucial input to relevant applications. Examples include unprocessed weather data, hourly stock market data, and event announcements. Transient data inherently do not require elaborate storage but often becomes invaluable for certain categories of consumers. For instance, unprocessed weather data can help in assessing weather conditions and road conditions in the absence of validated weather data. Event announcements, though not useful once the event is over, can be of great use in estimating population densities and movements in cities and can also help in highlighting the factors that affect the cities' economies.

The popular adoption of Web sites as front-ends to relational databases has resulted in the emergence of many heterogeneous data sources, a large data environment with easy accessibility via the Internet. Other than the inherent efficiency and

Related Content

Big Data Opportunities: Lung Image Segmentation for a Coronary Artery Diseases Monitoring System
 Karthikeyan, Alex Khangand K. Krishnaveni (2023). *AI and IoT-Based Technologies for Precision Medicine (pp. 314-323).*
www.irma-international.org/chapter/big-data-opportunities/332842

A Web Services Implementation of a User-Centered Knowledge Management Platform for the Construction Industry
Yacine Rezguiand Farid Meziane (2005). *International Journal of Intelligent Information Technologies (pp. 1-19).*
www.irma-international.org/article/web-services-implementation-user-centered/2390

A Roadmap on Updates
Fernando Zacarías Flores, Dionicio Zacarías Flores, Rosalba Cuapa Cantoand Luis Miguel Guzmán Muñoz (2009). *Encyclopedia of Artificial Intelligence (pp. 1370-1375).*
www.irma-international.org/chapter/roadmap-updates/10418

Virtual Insights, Real Solutions: The Promise of Augmented Reality in Medicine
Jaspreet Kaur (2024). *Approaches to Human-Centered AI in Healthcare (pp. 20-41).*
www.irma-international.org/chapter/virtual-insights-real-solutions/342006

Heterogeneous Influence Maximization Through Community Detection in Social Networks
Jaya Krishna Raguruand Devi Prasad Sharma (2021). *International Journal of Ambient Computing and Intelligence (pp. 118-131).*
www.irma-international.org/article/heterogeneous-influence-maximization-through-community-detection-in-social-networks/289629