

Chapter 4.14

Applying Information Gathering Techniques in Business-to-Consumer and Web Scenarios

David Camacho

Universidad Autónoma de Madrid, Spain

ABSTRACT

The last decade has shown the e-business community and computer science researchers that there can be serious problems and pitfalls when e-companies are created. One of the problems is related to the necessity for the management of knowledge (data, information, or other electronic resources) from different companies. This chapter will focus on two important research fields that are currently working to solve this problem—Information Gathering (IG) techniques and Web-enabled Agent technologies. IG techniques are related to the problem of retrieval, extraction and integration of data from different (usually heterogeneous) sources into new forms. Agent and Multi-Agent technologies have been successfully applied in domains such as the Web. This chapter will show, using a specific IG Multi-Agent system called MAPWeb, how information gathering

techniques have been successfully combined with agent technologies to build new Web agent-based systems. These systems can be migrated into Business-to-Consumer (B2C) scenarios using several technologies related to the Semantic Web, such as SOAP, UDDI or Web services.

INTRODUCTION

The aim of this chapter is to show how several Computational Intelligence (CI) techniques have evolved to allow the implementation of intelligent Business-to-Consumer (B2C) applications using new technologies related to the Semantic Web. The evolution of these applications is made possible by several CI techniques (such as Machine Learning or Planning) that allow users to represent, integrate, and manage information and resources available in electronic repositories. It is also made

possible by the evolution of the World Wide Web and its associated technologies (such as XML or RDF). In addition, Web services are emerging as a major technology for achieving automated interactions between distributed and heterogeneous applications. Various technologies constitute the backbone of this achievement, for example WSDL, UDDI, and SOAP. These technologies aim at supporting the definition of services, their advertisement, and their binding for triggering purposes. The advantages of Web services have already been demonstrated and highlight their capacity to be combined into high-level business processes. These business processes are composites that consist of several component services. It is argued that composition via service interconnection allows more sophisticated services and applications to be hierarchically constructed from primitive ones (Curbera, Duftler, Khalaf, Nagy, Mukhi, & Weerawarana, 2002; McIlraith, Son, & Zeng, 2002; WWW Consortium, 2002).

Computational intelligence techniques allow for the implementation of robust, adaptable, and reliable applications, and these characteristics are needed for a successful deployment of B2C applications. CI techniques are therefore important in promoting and developing B2C solutions on the Web. Many B2C applications have now been deployed as Web applications with considerable repercussions for e-business. These new kinds of Web sites allow consumers to rent a car, book a hotel, schedule travel, buy music, books, etc. It is obvious that new Web-based B2C applications are flourishing in the commercial sphere of the Internet.

The Semantic Web (Berners-Lee, Hendler, & Lassila, 2001) with its tools and related technologies, (including Ontology, WSDL, UDDI, SOAP) and Web services are likely to fall short of realizing the complete automation often envisioned by e-business practitioners. The evolution from the current Web to the Semantic Web has created new business possibilities that go beyond what traditional Information Retrieval (Baeza-Yates

& Ribeiro-Neto, 1999; Jones & Willett, 1997), or Information Searching (Chen, Chung, Ramsey, & Yang, 2001; Howe & Dreilinger, 1997; Lieberman, 1995) techniques provide. Some of those possibilities arise because new problems can be solved using the available information from several electronic sources. These problems need to use multiple information sources to obtain a solution or solutions. The large number of available electronic sources and the opportunity to find any type of information makes it possible to think about building systems that can retrieve, reason and finally reuse the stored information to obtain new solutions, much like a person does when consulting those same Web sources. These types of systems need to use techniques that allow not only the retrieval of a set of documents, but also to integrate the knowledge stored in those documents using automated reasoning. To build this kind of systems it is necessary to deal with several problems such as:

- How to represent the problems to be solved or the knowledge sources that store useful information for them;
- How to implement the necessary mechanisms to retrieve the information;
- When useful information is found in different knowledge sources, how to integrate these into a common solution;
- Finally, if several electronically distributed sources are considered, how to build a robust and flexible system that allows for the addition of new sources or the modification of existing ones.

Previous techniques can be defined as Information Gathering (IG) (Fan & Gauch, 1999). IG tries to integrate a set of different information sources with the aim of querying them as if they were a single information source. IG systems are able to extract knowledge from documents by analyzing their contents looking for a particular structure. So a document is seen by these systems

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/applying-information-gathering-techniques-business/24355

Related Content

Supervised Machine Learning Methods for Cyber Threat Detection Using Genetic Algorithm

Daniel K. Gasu, Winfred Yaokumahand Justice Kwame Appati (2023). *AI and Its Convergence With Communication Technologies* (pp. 19-42).

www.irma-international.org/chapter/supervised-machine-learning-methods-for-cyber-threat-detection-using-genetic-algorithm/328930

Ambient Interface Design (AID) for the Ergonomically Challenged

Rosaleen Hegarty, Tom Lunney, Kevin Curranand Maurice Mulvenna (2010). *International Journal of Ambient Computing and Intelligence* (pp. 57-64).

www.irma-international.org/article/ambient-interface-design-aid-ergonomically/43863

Learning Fuzzy Network Using Sequence Bound Global Particle Swarm Optimizer

Satchidananda Dehuriand Sung-Bae Cho (2012). *International Journal of Fuzzy System Applications* (pp. 54-70).

www.irma-international.org/article/learning-fuzzy-network-using-sequence/63355

Segmentation of Spine Tumour Using K-Means and Active Contour and Feature Extraction Using GLCM

Malathi M., Sujatha Kesavanand Praveen K. (2021). *AI Innovation in Medical Imaging Diagnostics* (pp. 194-207).

www.irma-international.org/chapter/segmentation-of-spine-tumour-using-k-means-and-active-contour-and-feature-extraction-using-glcm/271754

Cardiac Fitness Status Among Male Paddy Cultivators: A Study in the Context of Emerging Increase in Ambient Temperature

Ayan Chatterjeeand Shankarashis Mukherjee (2022). *Handbook of Research on Lifestyle Sustainability and Management Solutions Using AI, Big Data Analytics, and Visualization* (pp. 160-167).

www.irma-international.org/chapter/cardiac-fitness-status-among-male-paddy-cultivators/298373