Chapter XVII

Use of Log Analysis and Text Mining for Simple Knowledge Extraction:
Case Study of a Science Center on the Web

Leo Tan Wee Hin, Nanyang Technological University, Singapore

R. Subramaniam, Nanyang Technological University, Singapore

Daniel Tan Teck Meng, Singapore Science Centre, Singapore

Abstract

Log analysis of server data has been used to study the Web site of the Singapore Science Center, which is the largest Web site among all science centers in the world. This has yielded a wealth of data, which has been useful in assessing the effectiveness of the content hosted on the site. Additionally, the use of text-mining to structure an effective query interface for the Science Net database, which is an online repository of over 6,000 questions and answers on science and technology, is assessed. A commentary on the use of log analysis for virtual science centers is also presented.

Introduction

Data archived in the Web logs of servers represent a potentially rich source of information (Jones et al., 2000; Ren et al., 2002; Zhang, 1999). However, the raw data recorded on server logs are not of much use. It has to be first cleaned to remove redundant information such as, for example, sound, graphics, video and image files embedded on a Web page for which a hit has already been registered, and then stored in a data warehouse before using appropriate techniques to extract useful information (Joshi et al., 1999). Log analysis software is needed to unravel the data, and the information mined can help to evaluate the efficacy of a Web site through the identification of patterns and trends embedded in the data.

Published studies on the effectiveness of commercially available log analysis software are rather sparse in the primary reference literature, not surprisingly since the use of such software may be construed as endorsement of the product. Since such software represents, to a good extent, the state-of-the-art development in the field, its use can permit an evaluation of its effectiveness in real life settings. This can help to further bridge the chasm between theory and practice as well as provide a basis for further research. It was with this objective that the present study was undertaken.

More specifically, this study aims to use log analysis software to study the access patterns and trends from data archived in the server logs of the Web site of the Singapore Science Centre, which is the largest Web site among all science centers in the world. It is recognized that administrators and stakeholders would need reassurance that investments in a virtual science center are reaping dividends in the form of increased Web traffic. Log analysis software aims to contribute towards this. Sophisticated analysis of data archived in server logs is not the object of this study, and is therefore not addressed here. Only simple statistical analysis is used, as this is adequate for the needs of most science centers. An additional objective of this study is to use a commercially available text-mining software to construct a query interface for information retrieval by users of the Science Net database on this Web site.

Related Work and Rationale for Study

Science centers are institutions for the popularizing of science and technology among the masses (Danilov, 1982). The need for people to be cognizant of science and technology is a given in today's age of globalization and rapid scientific advances, and science centers do this in a way that makes people realize the impact of science and technology in their everyday lives. Commonly, this entails the use of science exhibitions and various promotional activities. Raising popular science literacy levels of the people through such initiatives empowers them to make informed decisions as well as become meaningful participants in science and technology driven nation-building efforts.

The Exploratorium in San Francisco pioneered the science center movement in 1968 (Oppenheimer, 1972; Delacote, 1998) and made a success of it for others to emulate the

Related Content

MACBETH: Development of a Training Game for the Mitigation of Cognitive Bias

Norah E. Dunbar, Scott N. Wilson, Bradley J. Adame, Javier Elizondo, Matthew L. Jensen, Claude H. Miller, Abigail Allums Kauffman, Toby Seltsam, Elena Bessarabova, Cindy Vincent, Sara K. Straub, Ryan Ralston, Christopher L. Dulawan, Dennis Ramirez, Kurt Squire, Joseph S. Valacichand Judee K. Burgoon (2013). *International Journal of Game-Based Learning (pp. 7-26).*

www.irma-international.org/article/macbeth/96976

Ontology-Based Competency Management for Corporate E-Learning

Fotis Draganidisand Gregoris Mentzas (2007). *Competencies in Organizational E-Learning: Concepts and Tools (pp. 319-332).*

www.irma-international.org/chapter/ontology-based-competency-management-corporate/6760

An Evaluation of the Added Value of Co-Design in the Development of an Educational Game for Road Safety

Anissa All, Jan Van Looyand Elena Patricia Nuñez Castellar (2013). *International Journal of Game-Based Learning (pp. 1-17).*

www.irma-international.org/article/evaluation-added-value-design-development/77312

E-Learning as Nation Building

Marco Adriaand Katy Campbell (2007). *Making the Transition to E-Learning: Strategies and Issues (pp. 1-16).*

www.irma-international.org/chapter/learning-nation-building/25610

Cognitive Modeling of Learning Using Big Data From a Science-Based Game Development Environment

Leonard Annetta, Richard Lamb, Denise M. Bresslerand David B. Vallett (2020). *International Journal of Game-Based Learning (pp. 22-39).*

www.irma-international.org/article/cognitive-modeling-of-learning-using-big-data-from-a-science-based-game-development-environment/262195