

Chapter XXXII

Mining User Activity Data in Higher Education Open Systems

Owen G. McGrath

University of California at Berkeley, USA

ABSTRACT

Higher education IT project managers have always relied on user activity data as logged in one form or another. Summarized counts of users and performance trends serve as essential sources of information for those who need to analyze problems, monitor security, improve software, perform capacity planning, etc. With the reach of the Internet extending into all aspects of higher education research and teaching, however, new questions have arisen as to how, where, and when user activity gets captured and analyzed. Tracking and understanding remote users and their round-the-clock activities is a major technical and analytical challenge within today's cyber-infrastructure. As open content publishing and open source development projects thrive in higher education there are some side effects on usage analysis. This chapter examines how data mining solutions – particularly Web usage mining methods – are being taken up in three open systems project management contexts: digital libraries, online museums, and course management systems. In describing the issues and challenges that motivate data mining applications in these three contexts, the chapter provides an overview of how data mining integrates within project management processes. The chapter also touches on ways in which data mining can be augmented by the complementary practice of data visualization.

INTRODUCTION

Before the advent of the Web and of large cross-institutional open source and open content projects, the job of tracking and reporting use was typically accomplished with commercial or home-grown utility software. An academic technology group at a university could confine its efforts to meeting the reporting needs of the local institution by using the features of in-house systems (McGrath, 2005). With today's global access to open content via open source applications on the Web, however, crucial questions arise as to how usage information can be captured and managed, how to ensure that the analysis of this information gets integrated into key management processes, and how to address the sheer scale and pace of usage as they outstrip conventional methods.

Higher education project managers still need to gather and analyze user activity information within contemporary open content or open source Web frameworks, but find themselves increasingly faced with unpredictable access patterns from unknown numbers of unseen users coming from who knows where. This chapter describes a range of higher education efforts – in digital libraries, online museums, and course management systems -- to develop practical approaches towards building and managing distributed capture and analysis systems for large Web-based production deployments. Moving beyond pure research, these projects meet a wide range of real tracking, monitoring, and reporting needs by applying data mining techniques in practice.

BACKGROUND

For decades, automatically generated logs have been a vital source of information within academic computing. From timeshare minicomputers to lab-based workstations, logs have provided data at the heart of acquisition, recharge, licensing, security, usability, and capacity planning processes. In the

Internet era, Web log data becomes an even more essential and often sole source of information for all of these same reasons together with the added urgency of supporting round-the-clock, remote access. Tracking users' interactions with digital library content, online museum collections, or e-learning material is crucial. Tools for making sense of usage within these systems depend heavily on the underlying Web technology. The monitoring features available to administrators of these systems rely heavily on Web application server logs as a record of the visitor's access. Further complicating the usage analysis problem is the trend within higher education towards open content, open source application frameworks, distance learning and a more general embrace of open online research collaboration (Lynch, 2007). Academic institutions are also increasingly involved in collaborative efforts to develop open source alternatives to commercial applications such as repositories, portals, and collaboration environments (Olsen, 2003). This shift in the locus of software development away from commercial companies and into loosely organized consortia of higher education institutions, however, ends up with noticeably different processes and results.

Behind open source projects such as Dspace, Fedora, uPortal, Sakai and Moodle are evolving organizations still experimenting with management structures that might allow them to coordinate distributed software development better (Wheeler, 2007). Many of the same technical principles - simple services, loose-coupling, lightweight data standards - that make these open source projects realizable also present challenges to building out full-featured Web analytic tools like those found in commercial systems. In particular, the lightweight distributed design can make it more difficult to develop comprehensive features in areas such as user event logging and usage analysis. To facilitate the development of new tools by loose-knit communities of developers, these consortia aim at minimal overhead and lean requirements in order for applications

12 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/mining-user-activity-data-higher/21653

Related Content

Organizational Commitment in the IS Workplace: An Empirical Investigation of its Antecedents and Implications

Qiang Tu, Bhanu Raghunathan and T. S. Raghunathan (2002). *Advanced Topics in Information Resources Management, Volume 1* (pp. 352-374).

www.irma-international.org/chapter/organizational-commitment-workplace/4594

ERP Selection using an AHP-based Decision Support System

Maria Manuela Cruz-Cunha, Joaquim P. Silva, Joaquim José Gonçalves, José António Fernandes and Paulo Silva Ávila (2016). *Information Resources Management Journal* (pp. 65-81).

www.irma-international.org/article/erp-selection-using-an-ahp-based-decision-support-system/164900

The Information Plan for Celerity Enterprises, Inc.: A Teaching Case

Laurie Schatzberg (2000). *Annals of Cases on Information Technology: Applications and Management in Organizations* (pp. 187-213).

www.irma-international.org/article/information-plan-celerity-enterprises-inc/44635

Collaborative Virtual Environments

Thrasyvoulos Tsiatsos and Andreas Konstantinidis (2009). *Encyclopedia of Information Science and Technology, Second Edition* (pp. 583-588).

www.irma-international.org/chapter/collaborative-virtual-environments/13633

A Method to Quantify Corpus Similarity and its Application to Quantifying the Degree of Literality in a Document

Etienne Denoual (2008). *Information Communication Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 751-763).

www.irma-international.org/chapter/method-quantify-corpus-similarity-its/22698