

the Net-based environments. This section discusses means and ways to mine data from Net-based environments.

All activity that takes place on a Web site or in a virtual environment is normally logged or stored by the program owner. However, the difference between the promise of accurate and meaningful information and the reality of what one finds in the thousands of lines of raw data produced by a Web server log has inspired a host of Web analysis applications and even more customized solutions. These applications produce a wide variety of individual and summary data, some of which may be useful to the e-researcher. If, for example, one is studying the use of a dedicated educational suite of Web-enabled software, such as First Class, WebCT, or BlackBoard, the program may itself be gathering and presenting data on user activity. This is almost always the easiest data to access. However, most of the data collected by these programs is designed for educational purposes, rather than research purposes and, hence, may not be optimized for e-research. As such, the e-researcher may need to turn to one of the more sophisticated commercial or freeware tools designed to assist researchers in identifying and measuring the activities that users engage in while visiting a site. Unfortunately, most of these tools are focused clearly on the e-commerce market. The tools available, while interesting and of potential value for some types of e-research, may be too focused on analysis of behavior that is directly related to current or future sales prospects. Concurrently the prices of some of these products reflect their commercial orientation.

Aberdeen Consulting coined the term *insight-to-effort* ratio in regard to Web analyzer software to highlight the amount of effort required of the e-researcher to extract meaningful insights from the behavior of users of the site. Complex analysis information may be very time consuming and require special programming and data extraction skills. At the lowest level, a researcher can use simple text analysis tools to examine and extract data from the Web logs themselves. However these logs are usually overly detailed and not formatted for ease of understanding or analysis. In all likelihood, work spent analyzing raw server logs produces a very low insight-to-effort ratio. Alternatively, an e-researcher may choose to lease or buy a high-end Web analyzer package and achieve insights at low effort, though probably at a high price. The e-researcher's task, then, is to select a set of tools that provides a high insight-to-effort ratio without exceeding our often limited research budgets.

Beginning e-researchers might wonder just what type of information a Web server routinely captures. The Australian Web service, VSBWEB, provides a real-time analysis of a variety of Web sites that it supports. Reviewing the reports at <http://vcsweb.com/logs/> provides a glimpse of the types of information available from the logs of a standard UNIX-based Web server and the output formatted using the open-source analysis program Analog (<http://www.analog.cx/>).

The basic features of Web analysis programs relevant to researchers include:

- Number of hits at specific pages.

- Amount of time between hits, thus indicating the time visitors spend at each page.
- Reviews of the path followed by subjects through particular educational sites.