RADCON Monitoring Module
Specification and Design

Prepared By

Samhaa El-Beltagy

February, 2004

1 Introduction

Carrying out monitoring activities with respect to web sites or information systems can offer invaluable data to system developers as well as to management staff as to how these systems are being used and how they can be enhanced. It is also an important evaluation indicator as it can show whether the system is actually being used for the purpose it was built for, or not. It is also often the case, that after carrying out monitoring activities, the system developers might find out that their system is being used in ways that they had not anticipated when designing the system.

Both the VERCON and RADCON information systems were designed with specific goals in mind to serve a specific set of users or stake holders. It is thus very important to have an advanced and well thought out monitoring utility in place for use with the both those systems in order to both access their use and identify areas of demand. It is the aim of this document to specify the requirements of monitoring activities that need to be implemented for the VERCON/RADCON systems as well as a design for a system that can carry these activities. This document is divided as follows: section2, provides a brief description of the problem, and the general requirements that the monitoring system will aim to satisfy, section 3 provides a list of reports that should be generated by the system while section 4 addressed design and implementation issues related to reports outlined in section 3. Using this document, a system developer should be able to implement the VERCON/RADCON monitoring system.

2 Problem Statement and General Requirements

Most Web usage analysis tools rely only on log files for carrying out their tasks. While this is sufficient for the general purpose of evaluating traffic on some specific web site, it was found to be inadequate for monitoring a system such as VERCON or RADCON. For example, this kind of data on its own can not be used to “understand” who uses the system or to correlate viewed pages to users so as to know which type of users are interested in what; using log file information on its own reports stating that 63% of some system’s users are researchers, and that the most popular web page in a site among researchers is that dealing with expert system, can simply not be done. Also, web pages have no defined “semantics” meaning that someone other than the site developer will be unable to make sense of the reports being offered. To clarify, a report stating the page *main.h*tm has been visited 890 times in June is less meaningful than a report stating the *VERCON’s Front page* has been visited 890 in June. In systems such as VERCON and RADCON, many subsystems exist (VERCON) or will exist (RADCON). Each of these subsystems is made up of a set of pages or scripts, which achieve its overall functionality. These pages may or may not have names that tie them to the subsystem in which they are being used. Using traditional monitoring utilities, a typical report will relate various requested statistics to these pages. While a system developer or a site manager might be capable of relating statistics generated for these individual pages to a specific sub-system, a project manager will be lost as he/she won’t be able to figure out which pages correspond to which system. To overcome these limitations, the monitoring system needs to be augmented with metadata about individual pages.

To address the first problem use can be made of the VERCON/RADCON user registration policy. This policy requires that users of the VERCON and RADCON systems register in order to be able to use the systems. Every time a user accesses the

system they need to login. By collecting data about the user when s/he first registers and then tracking his/her usage of the system using the log-in information, we can collect data as to the activities they carry out when visiting the site and co-relate this with pages that they view in ways that can be useful to the further development of the site.

Given the above requirements, the monitoring system to be developed needs to incorporate two tools: the first should be one for adding meta-data to various web pages so as provide them with meaningful descriptors that can be understood by anyone while the second should be a web based report generating tool that can be used by any party who has appropriate access privileges in order to gain access to reports generated by the monitoring system. While the next section briefly outlines reports that should be generated by the second tool, section 4 specifies in details the operations of both tools.

3 Required Monitoring Reports

In the development of a monitoring utility, it is important to identify the kind of reports that this utility will be capable of generating. In the following next two subsections, these reports are identified, and their functionality briefly outlined. While section 3.1 deals mainly with reports or indicators that are specific to the VERCON and RADCON systems, section 3.2 lists the more general type of reports that are traditionally available in any Web usage monitoring utility.

3.1 Specific VERCON/RADCON Indicators

1. Breakdown of system users by occupation (Researcher, Extension Worker, Government Official, System developer, Student, etc) represented as a percentage of all users under consideration. This may be generated for:
   - all system users
   - users who registered in some given month
   - users who registered over a given date range (inputs = start date and end date)

2. Breakdown of system users by affiliation (Research Centers, Universities, NGOs, Private Sector, Government, etc) represented as a percentage of all users under consideration. This may be generated for:
   - all system users
   - users who registered in some given month
   - users who registered over a given date range (inputs = start date and end date)

3. Breakdown of system users by their gender (male, female) represented as a percentage of all users under consideration. This may be generated for:
   - all system users
   - users who registered in some given month
   - users who registered over a given date range (inputs = start date and end date)

4. Breakdown of system users by their governorate (Cairo, Giza, Kafer El-Shiek, etc) represented as a percentage of all users under consideration. This may be generated for
   - all system users
   - users who registered in some given month
   - users who registered over a given date range (inputs = start date and end date)

5. Forum specific statistics:
    - Total number of active participants per forum
    - Number of new messages posted per forum

The above information should be displayed for:

- a given month or
- a given date range (inputs = start date and end date)

6. Co-relation between viewed pages and various users
    - Top ***n*** pages viewed by users of some given occupation
    - Top ***n*** pages viewed by users of some given affiliation
    - Top ***n*** pages viewed by users of some given gender
    - Top ***n*** pages viewed by users of some given governate

The above information should be displayed for:

- a given month or
- a given date range (inputs = start date and end date)

3.2 General Indicators

7. Number of hits for each page:
    - For a given day
    - For a given range (inputs = start date and end date)
    - For a given month
    - For a given year

    To be selected from a drop down menu.
    To be represented in textual as well graphical format.

8. Maximum load Time

This is defined by the number of overall hits encountered by the server in any given hour.

- For a given day
- For a given range within a given day (inputs= start time and end time)
- For a given range (inputs = start date and end date)
- For a given month

9. Number of visitors over the entire site, determined by the total number of unique IPs that access individual pages, and presented for:
    - a given day
    - a given range (inputs = start date and end date)
    - a given month

10. Breakdown of access by country. This will also be presented for:
    - a given day
    - a given range (inputs = start date and end date)
    - a given month

11. A list of accessing IPs/host names for any given page to be presented for
    - given day
    - a given range (inputs = start date and end date)
    - a given month

This list will provide statistics as to the number of times a given IP has accessed this individual. To be selected from a drop down menu.

12. Some usage statistics that include:
    - Top n sites that access the server the most
    - Day on which maximum number of hits were made in a given month

4 Design & Implementation Issues

As stated before, the development of a monitoring utility for the VERCON/RADCON monitoring system, entails the implementation of two tools: one to assign to them meaningful names/descriptors that can be understood by anyone, and another to generate reports. Both reports can be used by any party who has appropriate access privileges. The implementation of these tools will result in changes in the current VERCON database, and will introduce a component that will have to be installed in various newly developed RADCON web pages. In the following subsections, each of these is described in details, and any changes that need to be made to the currently developed system or database are outlined.

4.1 The Meta-data assignment tool

The purpose of this tool is simply to allow system developer to provide meaningful names that can be used to refer to pages that they develop. This way, when a report is generated using the reporting tool, data can be displayed using these meaningful names instead of the more ambiguous page names.

4.1.1 Required Database Changes

Two tables should be added to the system's database. The first of these will simply be used to map between a page name as it is represented in the file system, and a meaningful, but brief name that describes the page. For example, the file named "VerconProject/ExDoc5.asp" should be associated with the name: "Search In Extension Documents" while the file "wheat/main.asp" should be associated with "The wheat Expert system". The second table should simply contain a list of user Ids that refer to users who can access this system (Assuming that a user of this system is already a registered system user).

4.1.2 The User Interface

The first screen that should appear is a login screen. There should be two separate log-in screens for the VERCON system and the RADCON systems. Once the user has successfully logged into the system, they should be offered a screen to allow them to select a file and assign a meaningful name to that file. Figure 1 is provided as a guideline as to what this screen should look like. Through this screen a user should be also be able to edit the values of previously assigned names. Another option that should be offered through this tool, is the option to simply list all the names of the previously assigned file names which in Figure 1 is represented as a hyperlink at the bottom of the page. Following this hyperlink should result in the appearance of a page similar to that shown in Figure 2 and following a hyperlink in the screen shown in Figure 2 should result in the user being able to edit the name/descriptor displayed.

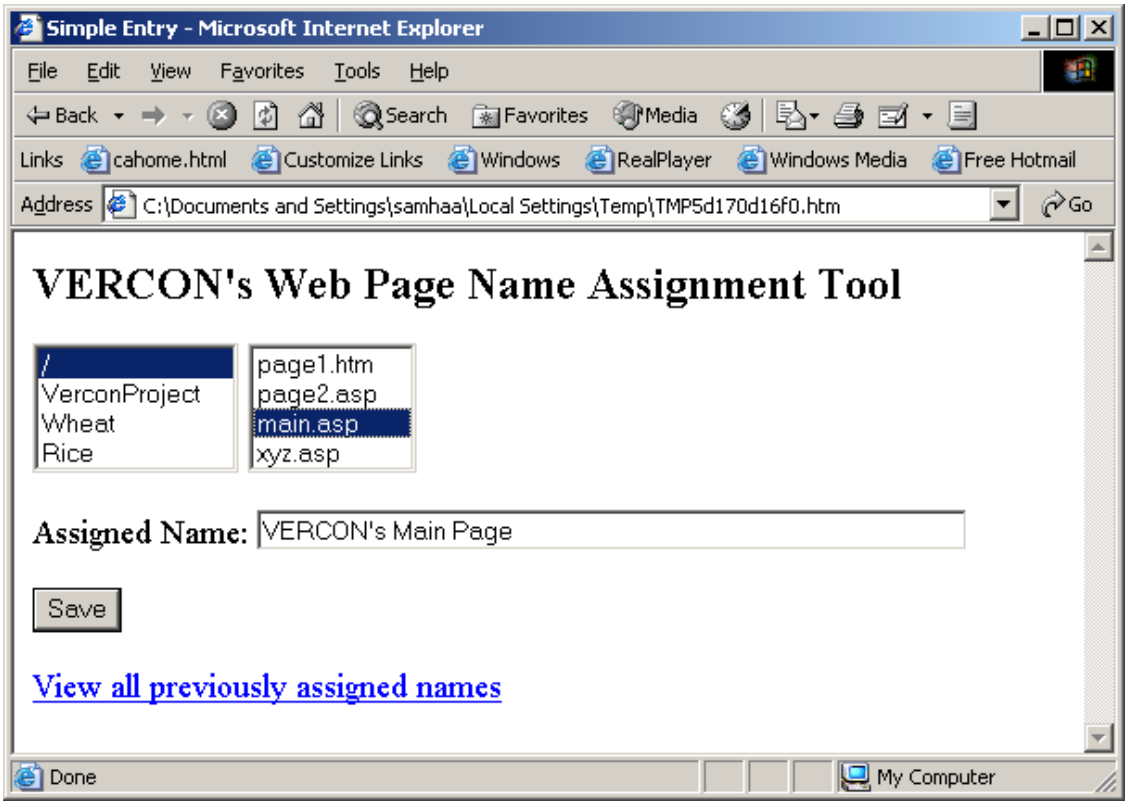

Figure 1: Screen for selecting and assigning a meaningful name to a file

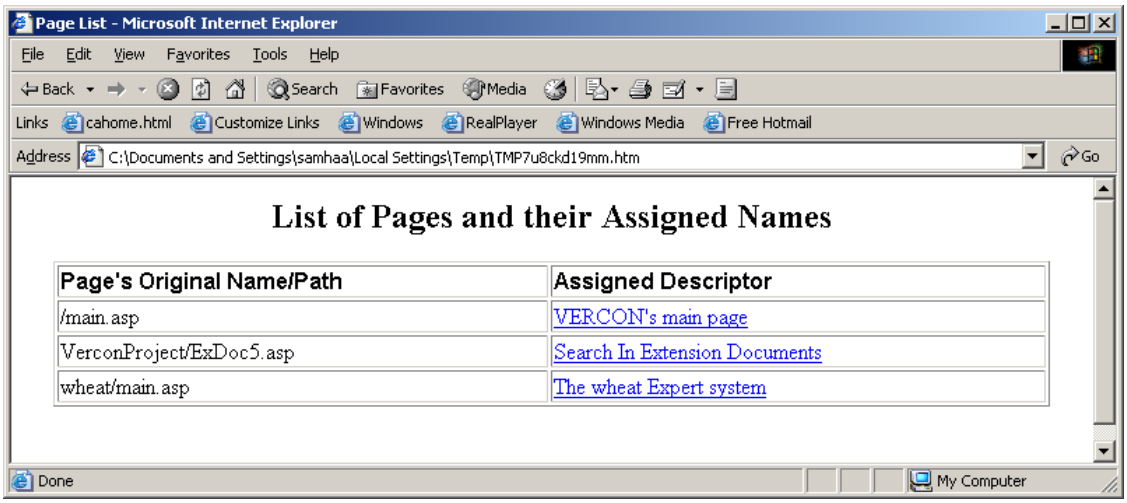

Figure 2: Screen for viewing previously assigned descriptors

4.2 The Report Generation tool

The report generating tool will be used to generate the reports discussed in section 3. While traditionally, a monitoring report generation tool will rely primarily on log files stored on the server machine to generate various access statistics, this tool will have to employ other means for carrying out its task. This is because, as discussed before, the

role of this tool goes beyond generating access statistics. Specifically, this tool will have to rely on a component that keeps track of user log-ins and access patterns and record those in the system's database. The details of this component, required changes to the database, as well as the overall interface of this tool, are described in the following subsections.

4.2.1 User Log-In Tracking Component

The aim of this component is to keep track on users and their access patterns. Every time a page is accessed, the following information should be saved in the database:

- Page-ID: ID of page being requested
- Date: date on which a request was made
- Time: (time at which a request was made)
- req-ip: IP of requesting client
- req-host: IP of requesting client, this will be derived from the IP whenever possible
- userID: ID of the user who logged into the page
- SessionID: The session ID for the logged in user
- User-Agent: requesting software/agent