Tracking and analyzing visits to the Simbios and SimTK websites

Jeanette P. Schmidt

Version 1.1

August, 2005

Summary

Tracking visitors to the SimTK.org and Simbios.stanford.edu sites is, among others, a request by the NIH science officers that will help them analyze the use of the sites. In addition, it may help the Simbios staff improve the usability of SimTK.org. This document largely addresses the needs for SimTK.org, a subset of the listed features would be sufficient for Simbios. Tracking should be put in place for SimTK 1.0 with all the functionality we need now and what we may need eventually – it should not be necessary to revisit this basic functionality.

1 Introduction

This document is organized as follows. The items that require tracking are listed in Section 2., Section 3 describes the granularity with which tracking should occur. Section 4 discusses requirements for the presentation of the tracking output. Section 5 discusses things outside our scope.

2 What to capture

We would like to track visits to SimTK.org on a daily basis and record the following statistics:

1) Number of unique visitors:
   A unique visitor is defined as an IP address who came to visit the site and who viewed at least one page. This should measure the number of different persons that visited the site. The requirement to view at least one page should distinguish “people visits” from visits by automated engines (worms, crawlers, etc). We may want to optionally record these separately but they should not be counted as “visitors”.
2) Number of total visits:
   This is less informative than unique visits, but gives some information on how many times a given visitor “returns”. It counts the number of visits to the site -- a new visit is defined as an incoming visitor not connected to the site in the last 60 minutes
3) Pages:
   Number of times a page is visited (summed for all visitors and visits))

4) Hits:
   Number of times a page, image or file is viewed or downloaded. This is different from pages as it also counts downloads of image or other data.

5) Ability to break down each of these measures by

   a) Geographic region:
      It is most important to distinguish the Stanford.edu domain from the rest. Since number of unique visitors will be one of our key measures, it is probably OK to not worry about not being able to identify Stanford staff logging in from home as part of Stanford. But as long as we are capturing geographic location, it would be nice to organize the geographic regions in addition. Depending on the complexity of implementation a breakdown by city & state, by state only or by time zone would be acceptable. Clearly the breakdown by city & state provides the most information, but it may be difficult to achieve.

   b) Operating systems & Browsers (this is optional and certainly not required for "PR" purposes but will come in handy down the line)
      Visits broken down by operating system and browsers.

   c) Registered and unregistered users

3 Granularity

We would like to generate statistics for the entire site, on a project per project basis, for each page and for each link (to a document or other site) as well as downloads from the subversion repository. In addition the number of downloads should be grouped for items that naturally form one unit. The only explicit requirement for the grouping consists of versions of an executable or document. For example if a program has 4 versions we would like to know how many times the last version of the program has been downloaded, as well as the total number of downloads for the program (regardless of version). Clearly if the author decides to change the name of the program or create a new project for a new version – the program would be considered a new program

4 Presentation

It is very important that all these statistics can be easily graphically presented. These should be extremely easy to obtain by any Simbios staff member. AWstats is a free program that has most (if not all) of the required properties, (see http://awstats.sourceforge.net/#WHAT). Awstats may require an additional plug-in (possibly not free) to capture the geographic location of the visitors as required in this document.

5 Not in scope

We will not track the activities of individual users, but use this tracking just for global information. We may should post somewhere on the website what information we gather and for what purpose and include that we will not track the activities of individual users.

6 References

AWstats: http://awstats.sourceforge.net

ClickTracks: http://www.clicktracks.com

Acknowledgments

This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 GM072970. Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics.