Insight Log Analyzer Tool:

Introduction:

The Insight log analyzer tool is intended to provide general statistics on use of the Luna Insight Software. It is a Perl script that parses logs generated by the Insight server and creates reports in HTML format.

Installation:

Files Included:

doc/ -- contains this document
examples/ -- contains sample commands, log files, a properties file and outputs.
perl/ -- contains the analyzer script and all required files

Where to install the log analyzer:

Since log analyzer is a processor intensive Perl script, it will probably make sense to execute the log analyzer on a different machine other than where your collection manager is installed.

When choosing where to install and run the log analyzer, be aware that it requires access to the following log files:

For Insight 5.0+ users:

- *the collection's web server logs or media manager logs (depending on where images are served from)*
- *the collection's access, query, and export logs* (to determine who logged in, searches run, and images exported) These are generally located in the same folder as your collection server
- *the browser's access and query logs* (to determine who logged in and searches run) These are generally located in the JSP Browser's logs folder.

For versions before 5.0:

- *the collection's web server logs*
- *the collection's access, query, and export logs*
- *the browser's access and query logs*

For more information on the log files above, refer to 'Log Files' section.

Installation:

1) The following changes to the BrowserInsight.conf are necessary to enable logging support:
   a. Session logging must be configured as follows:

```
SessionLogFile = Browser-Insight-Session-Log.txt
```

```
SessionLogMode = 1
```

b. The Search Query log should just be the filename (not including the full path to the file) for example:

```
SearchQueryLogFile = Browser-Insight-Query-Log.txt
```

2) Install ActivePerl 5.6.1 or higher for Windows on your machine (if it is not already installed). A copy of ActivePerl is included in the following folder: \CD3_Upgrade\utilities\media_server\IIS\, or it can be downloaded from: http://aspn.activestate.com/ASPN/Downloads/ActivePerl/
3) Copy the log analyzer folder on to a computer where you will be running log analyzer.
4) Create a folder on the same computer and copy all the required log files from the server to this location.
5) Install the following Perl modules. You can download the latest versions of these modules via the Perl Package Manager (PPM) or from CPAN (http://search.cpan.org). To start the Perl Package Manager, type PPM in a DOS prompt. Type the following to install the required modules:
   - install Time-modules
   - install Archive-Zip
   - install Net-IP
   - install CGI

   These should install the modules below:
   - Time::ParseDate (included in Time-modules)
   - Archive::Zip
   - Net::IP
   - CGI (generally included with all Perl distributions)

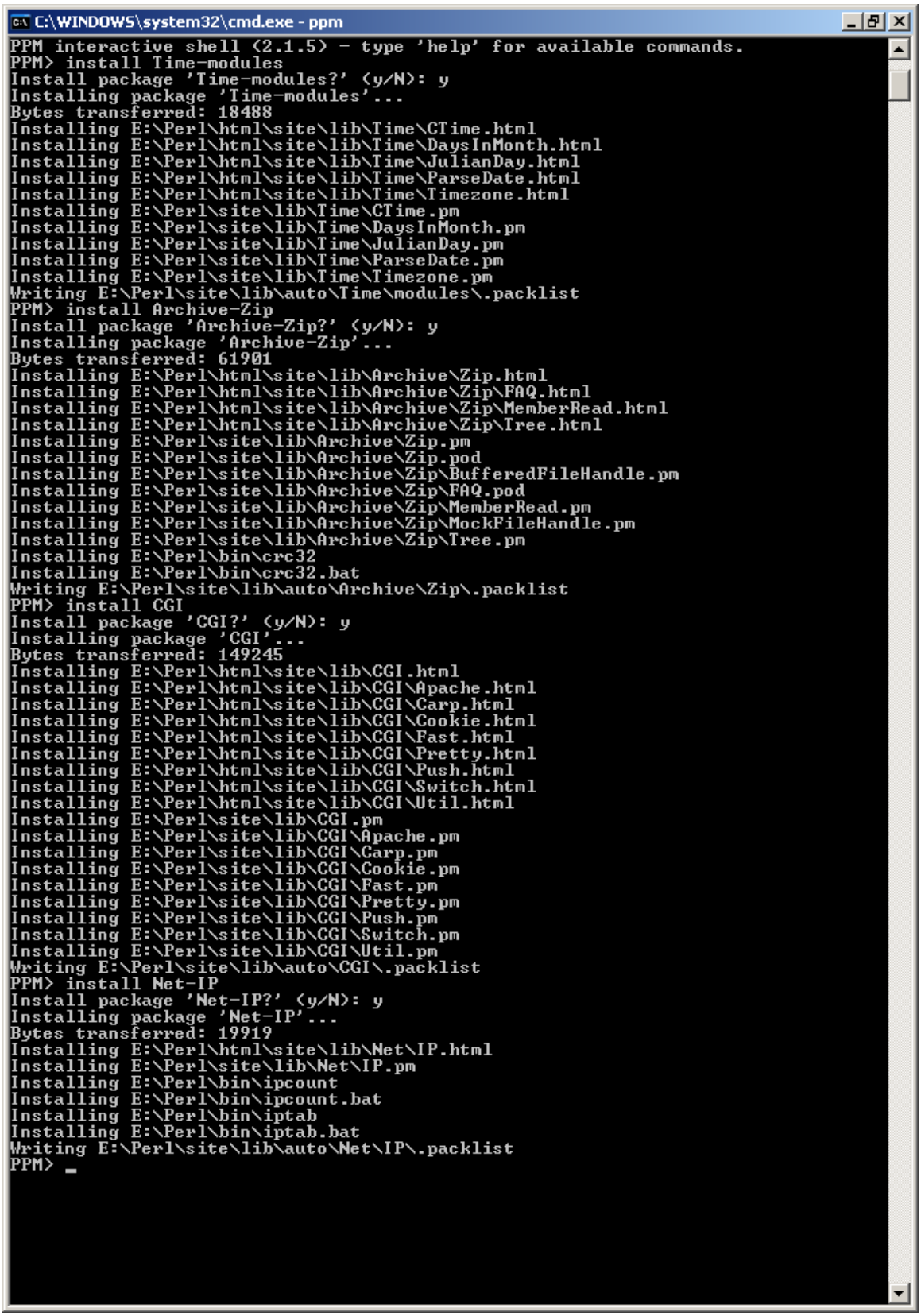


Figure 1: Screenshot of installing Perl modules using PPM

6) Edit first line of the analyzer.pl file to point to the perl executable.

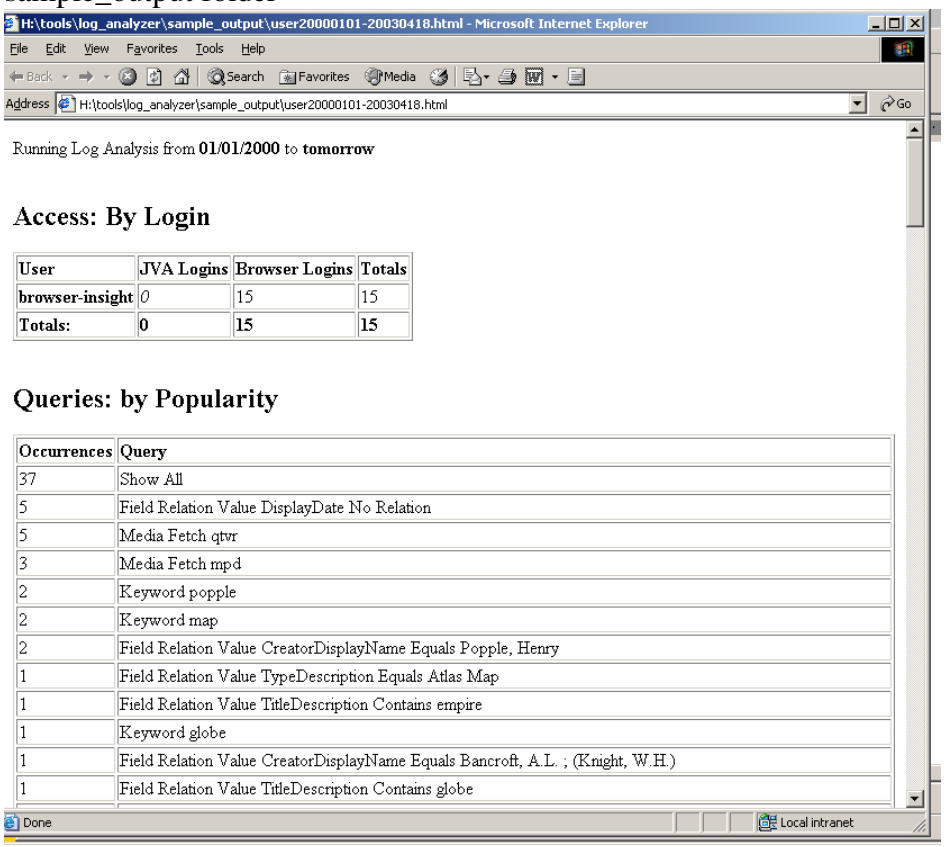
   #!c:\perl\bin\perl.exe --> #!/path/to/excecutable

Sample output:

Sample log data and sample log files are included in the installation package in examples directory. Finally, sample invocations of the log analyzer for common reports are stored in the examples\commands folder. See Figure 2 for a screenshot of a sample report.

Note: The sample .cmd files must be executed in the same directory as Analyzer.pl.

Figure 2: Screenshot of a sample report (other samples are available in the sample_output folder



Configuration and Use:

Valid Log Files:

Apache, IIS, and Media Security Server Log files:

IIS Log Files:
For IIS log files to be processed, they must have the following values in the W3C format. Date, Time, Client IP, Method, URI Stem, URI Query, Status

Apache Log Files:
For Apache log files to be processed, they must be in either common or combined format.

Media Manager Log Files:
These will work in their current format.

Log files:
The following log files are needed to properly produce a report:

- IIS (W3C format *see above*) or APACHE (common) Web server log or Media Manager log (Insight 5.0+) [required for ***Image Access*** reports]
- Insight JVA Access Log [required for ***Access:By Login*** reports]
  (in collection_manager/*three_letter_collection_name*/logs)
- Insight JVA Query Log [required for ***Queries*** reports]
  (in collection_manager/*three_letter_collection_name*/logs)
- Insight JVA Export Log [required for ***Export*** reports]
  (in collection_manager/*three_letter_collection_name*/logs)
- Insight Browser Access Log (Optional). Note: as the browser can support multiple collections one log file is created per collection and has the appropriate Collection ID appended to the beginning of the file (in order to easily determine which is the correct log file)
- Insight Browser Query Log (Optional) Note: as the browser can support multiple collections one log file is created per collection and has the appropriate Collection ID appended to the beginning of the file (in order to easily determine which is the correct log file)

Invocation:

perl analyzer.pl usertype=inst process=all headerfile=headerfile.html ...

Passing parameters to analyzer.pl:

Command line properties can be passed to analyzer.pl either via the command line or via a properties file specified on the command line. The properties file is a simple way to pass the same set of properties over multiple invocations (good for IP mapping).

Note on log directories:
The log directories passed to the log analyzer must contain the JVA and Browser log files for ONE COLLECTION ONLY, with the exception of the web server log. This is to ensure that log data from multiple collections are not combined.

Usertypes:
Insight supports two different concepts of users: Individual users and Institutional users. Individual users are defined by usernames in the user manager while institutional users are defined by a group name (the one paired with an access-key/code-key). The log analyzer by default will assume that it should treat users as institutional users, however this can be overridden by the usertype parameter.

Note: recording access of individual users by username requires a 4.0+ client.

Figure 2: Configuration parameters for the Log Analyzer

Parameter	Example	Repeat?	definition	Required
logdir	logdir=c:\logs	Y	Specifies one or many log directories where logfiles may be found for processing	YES
fromdate	fromdate=01/01/2000 or fromdate=last year or fromdate=last month or fromdate=last week or fromdate=yesterday	N	Specifies the first valid day for a log entry	YES
todate	todate=01/01/2000 or todate=last year or todate=last month or todate=last week or todate=today or todate=tomorrow	N	Specifies the last valid day for a log entry	YES
outdir	outdir=c:\inetpub\...	N	Specifies the output directory for reports	YES
cid	cid=1	N	Specifies the collection id for the collection to report on	YES
iid	iid=LUNA	N	Specifies the institution id	YES
vcid	vcid=NA	N	Specifies the virtual collection id (NA if this is not a virtual collection)	YES
propsfile	propsfile=analyzer.dat	N	Specifies the filename of the properties file (see figure 3 for an example)	NO
usertype	usertype=Institution or usertype=User	N	Specifies whether users are grouped by their groupname or by their username	NO
process	process=all or process=username	N	Specifies the user(s) to process. All specifies process all user, specifying a user limits processing to just that user	NO
headerfile	headerfile=header.html	N	Prepends the headerfile to the beginning of the log report (good for adding an image or title)	NO
footerfile	footerfile=footer.html	N	Appends the footerfile to the end of the log report	NO
fileprefix	fileprefix=luna	N	Specifies the prefix for a filename, i.e. if rumsey is specified, a filename might be rumsey2001-2004.html	NO

logsuffix	logsuffix=log	Y	Specifies one or many allowable suffixes for a log file (so the script will not try and parse invalid file types). Note: zip files are a valid format	NO
overideWebRegex	overideWebRegex=/luna/|client=luna	N	Allows an administrator to filter down the HTTP log further using a Regular Expression, in example, URL contains /luna/ or client=luna	NO
mingroupfilter	mingroupfilter=10	N	Specifies the minimum number of times a group is accessed before it is displayed in the group access log	NO
minresultfilter	minresultfilter=10	N	Specifies the minimum number of results for a search before that search is displayed in the search results log	NO
minpopfilter	minpopfilter=10	N	Specifies the minimum number of times a search is performed before it is displayed in the popularity log	NO
minaccessfilter	minaccessfilter=10	N	Specifies the minimum times a user must login before they are displayed in the access log	NO
minexportinstfilter	minexportinstfilter=10	N	Specifies the minimum number of times an institution exports images before it is displayed in the export log	NO
minexportfilefilter	minexportfilefilter=10	N	Specifies the minimum number of times an image is exported before it is displayed in the export log	NO
minwebinstfilter	minwebinstfilter=10	N	Specifies the minimum number of times an institution accesses images before it is displayed in the image access log	NO
minwebfilefilter	minwebfilefilter=10	N	Specifies the minimum number of times an image is accessed before it is displayed in the image access log	NO

map	map=131.229.219.0-131.229.219.255: TEST	Y	Specifies an IP address or range of IP addresses to map to a given user or group. Valid IPs include:
			
			`A range: 20.34.101.207 – 201.3.9.99`
			`An IP: 193.0.1.46`
				NO

Figure 3: Sample properties file (file.props):

```
# Note: lines starting with a # are comments

# locations of log directories
logdir=H:\tools\log_analyzer\logs\

# Note: it is highly recommended that log analyzer run on a different computer other than your server.
# The following parameters for logdir are for illustration only.
logdir=C:\WINDOWS\system32\Logfiles\
logdir=C:\Program Files\LunaImaging\5.5\collection_manager\ABC\logs
logdir=C:\Program Files\LunaImaging\5.5\media_manager\logs\

# allowable suffixes for log files
logsuffix=log,txt
logsuffix=zip

# Map IP addresses to a given user or group
map=131.229.219.0-131.229.219.255: TEST

# collection ID
cid=10000

# institution ID
iid=VRA

# virtual collection ID
vcid=NA
```