Link Checking in Voyager

Simon Huggard, Systems Manager, Monash University Library
IGeLU, Brno. 4 September 2007
Outline

• About myself and Monash University
• Why do link checking?
• Voyager programs
• My programs – elinks and exams
• Future development
• Questions/discussion
About Monash University
About Monash University & the library

• MONASH UNIVERSITY
  – 6 campuses in Victoria, Australia, 1 in Malaysia and 1 in South Africa.
  – 53,000 students
  – 10 Faculties
  – 5000 academic and general staff
Monash University Library

**LIBRARY**
- 8 branch libraries
- 1.9 million books
- 60,000 journals
- 111,000 e-books & e-journals
- 600 electronic databases
- 265 library staff
- Systems Area – 16 staff - web, MULO, voyager/unix, help desk, it support.
# Service Level Agreement

**Support Service:** MONASH UNIVERSITY LIBRARY

**Service:** No. 1 – Information Resources: Access and Delivery

**Service description:** Access to materials and resources

<table>
<thead>
<tr>
<th>Key Performance Indicator</th>
<th>Result</th>
<th>Date reported/ to be reported</th>
<th>Period covered by statistics</th>
<th>Frequency to be reported</th>
<th>Measured by</th>
</tr>
</thead>
<tbody>
<tr>
<td>The library catalogue available during core service hours Monday to Friday 8.30am to 5.30pm with the exception of scheduled downtimes. There were 10 hours of outages during core hours which were unplanned.</td>
<td>System available 98.2% of core service hours.</td>
<td>April, July, October and January</td>
<td>January – March 2007</td>
<td>Quarterly</td>
<td>% of hours the library catalogue available during core service hours</td>
</tr>
</tbody>
</table>

**Further Information:**

**Service:** No. 2 – Information Resources: Collection Management

**Service description:** Development, selection, acquisition, cataloguing, maintenance and overall management of library resources

<table>
<thead>
<tr>
<th>Key Performance Indicator</th>
<th>Result</th>
<th>Date reported/ to be reported</th>
<th>Period covered by statistics</th>
<th>Frequency to be reported</th>
<th>Measured by</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% of web links in the library catalogue are accurate</td>
<td>98.5% of web links in the library catalogue were accurate.</td>
<td>April, July, October and January</td>
<td>January – March 2007</td>
<td>Quarterly</td>
<td>% of accurate links in the library catalogue identified by monthly link checking program</td>
</tr>
</tbody>
</table>
Link checking programs – Voyager – catjob 9 or 10

<table>
<thead>
<tr>
<th>Cataloging Batch Job Number</th>
<th>Cataloging Batch Job Name</th>
<th>Report Name</th>
<th>Record Type Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catjob 6</td>
<td>'See' References With Linked Bib Records Report</td>
<td>See Refs with linked bib records</td>
<td>Type 06</td>
</tr>
<tr>
<td>Catjob 7</td>
<td>'See' References Authorized In Another Authority Record Report</td>
<td>See Refs authorized in another record</td>
<td>Type 07</td>
</tr>
<tr>
<td>Catjob 8</td>
<td>'See Also' Reference Without Corresponding Authority Record Report</td>
<td>See Also Refs without an Authority record</td>
<td>Type 08</td>
</tr>
<tr>
<td>Catjob 9</td>
<td>HTTP Verification - Within Date Range</td>
<td>856 Link Failure report</td>
<td>Type 09</td>
</tr>
<tr>
<td>Catjob 10</td>
<td>HTTP Verification - All New Links</td>
<td>856 Link Failure report</td>
<td>Type 09</td>
</tr>
</tbody>
</table>
Voyager link checking program

- Catjob 10 only validates NEW links
- Catjob 9 will revalidate ALL links within a date range

NOTE:
Catjob 10 *does not* revalidate links that have previously been validated, whereas catjob 9 *does* revalidate links. As a result, you should run catjob 10 first to validate all new links then run catjob 9 to revalidate links that were last checked within the specified date range.

- These jobs can be run from command line on the server, or from WebAdmin.
Voyager Server Utilities
Version 6.1

Acquisitions
- Reports and Processes

Catalog
- Bulk Import MARC Records
- Bulk Export MARC Records
- Cataloging Reports

Circulation
- Reports and Notices
- Bursar Transfer

OPAC
- Reports and Processes

System Admin
- Log Files
- Report Files
Cataloging Reports and Processes

Select the report you want to run and press the submit button. Certain reports require a starting and ending date, and some require a location code. When you select one of these reports, you will be prompted for the appropriate information.

When the job is completed, you will receive notification via email that the job is complete as well as any additional information which was provided by the job. You will be able to retrieve the data generated using the Voyager reporter program (follow instructions in Accessories manual).

Also, check for the log file under "Voyager's Report Files" and look for "catjob log".

10. HTTP Verify - All

Email Address: Simon.Huggard@lib.mcl required

Submit CatJob

Job Scheduler

Return to Main Page

Email notification to: Simon.Huggard@lib.monash.edu.au
job: catjob -j 10

www.lib.monash.edu.au
Catjob.log file

Fri Aug 17 10:19:12 2007 Connection to Voyager Database successful...
Fri Aug 17 10:19:12 2007 HTTP Verify - All
Fri Aug 17 10:19:12 2007 Opening report file...
Fri Aug 17 10:19:12 2007 Building HTTP Bad Link Report...
For Cataloging batch jobs, the file that is created is always named catrprts.xxxx.inp.

`catrprts.SYS.inp`
## Entries in the log file

<table>
<thead>
<tr>
<th>Time</th>
<th>IP</th>
<th>URL</th>
<th>Error Code</th>
<th>Error Description</th>
<th>Holdings</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:57.2</td>
<td>192.0.2.192</td>
<td><a href="http://name.umdl.umich.edu/AAN7780">http://name.umdl.umich.edu/AAN7780</a></td>
<td>404</td>
<td>Not Found</td>
<td>2105711</td>
<td>08/17/2007</td>
</tr>
</tbody>
</table>
Voice to the City,
Or, A Loud Cry from HEAVEN TO LONDON

Her Sins, Setting before her, Her Sickness, Her Remedies.

Mich. 6.9. The Lord's Voice cryeth to the City, bear the Rod, and who hath appointed it.

LONDON
Your URL links to the following volume(s).
Bibliographic information is provided to confirm the link.
This information is provided by the
University of Michigan Library Name Resolver Service.

Author: Cooper, James Fenimore, 1789-1851.
Title: Homeward bound; or, The chase. A tale of the sea By J. Fenimore Cooper.
URL: http://www.lib.umich.edu/cgi/t/text/text-idx?c=moa;idno=AAN7780.0001.001
Extent: vi, [7]-471 p. 18 cm.
Place of publication: New York.
Publisher: Hurd & Houghton.
Date of publication: 1871.
Availability: These pages may be freely searched and displayed. Permission must be received for subsequent distribution in print or electronically. Please go to http://www.umdl.umich.edu/ for more information.
Title: Making of America (MOA)
In Collection(s): Making of America Books
Type: text
Format: xml
Language: ENG
# Printed Link Failure Report

## 856 Link Failure Report

**08/17/2007**

<table>
<thead>
<tr>
<th>Link Host</th>
<th>Date Updated: 08/17/2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>name.umdl.umich.edu</td>
<td></td>
</tr>
<tr>
<td><strong>Link Type:</strong> URL</td>
<td></td>
</tr>
<tr>
<td><strong>Error Code:</strong> 404</td>
<td></td>
</tr>
<tr>
<td><strong>Link Text:</strong> <a href="http://name.umdl.umich.edu/AAN7760">http://name.umdl.umich.edu/AAN7760</a></td>
<td>Not Found</td>
</tr>
<tr>
<td><strong>Record Type:</strong> Holdings</td>
<td></td>
</tr>
<tr>
<td><strong>Record ID:</strong> 2178856</td>
<td></td>
</tr>
<tr>
<td>ezproxy.lib.monash.edu.au</td>
<td></td>
</tr>
<tr>
<td><strong>Link Type:</strong> URL</td>
<td></td>
</tr>
<tr>
<td><strong>Error Code:</strong> 0</td>
<td></td>
</tr>
<tr>
<td><strong>Record Type:</strong> Bibliographic</td>
<td></td>
</tr>
<tr>
<td><strong>Record ID:</strong> 2105711</td>
<td></td>
</tr>
</tbody>
</table>

---

End of 856 Link Failure Report
Link checking around the world

- Most libraries export all the links from voyager
- Create an HTML file of links
- Link checking programs: many out there - Xenu - most popular
Other libraries

-------- Original Message --------
Subject: [VOYAGER-L]: Link Checker Program
Date: Tue, 15 May 2007 11:00:02 -0400
From: Jason Zou <qzou@lakeheadu.ca>
To: voyager-l <voyager-l@voyager.ship.edu>

Hi everyone,

At the EndUser 2007, I had a poster session about a link checker program I developed. I have updated the program based on some comments and suggestions gathered at the conference.

You can download the program from the following URL:
   http://erl.lakeheadu.ca/staff/doku.php?id=public:elink_checker

Hope you may find this helpful.

Jason

--
Jason (Qing) Zou
Systems Librarian
Lakehead University Library http://library.lakeheadu.ca
(807)343-8251, qzou@lakeheadu.ca
Link Checking at Monash

- **Elinks program** – to extract all links and massage them into meaningful smaller files
- **Exams link checker** – extracts all links and does checking online
Elink_index table

- Has just over 500,000 links in it
- Includes duplicates from MFHD and BIB records
- We don’t want to check all links
- Some databases don’t like you checking all links at once (and block your access)
- We don’t want to count link checking via our ezproxy server
Elinks program

- Copies and sorts links from the elink_index table
- Deduplication of links
- Checking of 4 links per database domain
Elinks program

- Pulls out TOC and book review URLs
- Pulls out IMDB (movie info) links
- Sorts into separate HTML files of 1000 links each
- Removes ezproxy from the URL
Elinks program

Voyager elinks program - Version for Michelle Rusiniak

Retrieving data from the E-links table.

Please wait..

The program excludes ERES, ERES-W and ERES-D locations as well as MLO records.
It gets all records, including those without MFHDs.

Our e-links table has over 490,000 records in it at present.

Program written in Microsoft Visual FoxPro 7 by Simon Huggett
21 August 2001. Updated 21 August 2007 6:20pm

Running query...

March - May 2007 - new code for SS records, TOC & other links and to fix some bugs
30 May 2005 - added code to remove MLO records
11-21 March 2005 - made changes to speed up the program.
24 March 2005 - added code to remove ERES locations
10 March 2005 - no longer gets suppressed records.
4 March 2004 - only gets 5 links from same database
Elinks program in detail

Pulling out domains in URLs
Querying database: elink_index_erss

Please be patient - this may take up to 7 hours!!!!...
elink_index_erss Table. Record no. 7892 of 34430
Percent completed: 23%

This should actually take 11 minutes or 0.19 hours to finish.
Elinks program

Pulling out domains in URLs
Querying database: elink_index_rest

Please be patient - this may take up to 7 hours!!!!...

elink_index_rest Table, Record no. 57811 of 570107
Percent completed : 12%

This should actually take 190 minutes or 3.17 hours to finish.
Elinks program – detailed description
Elinks HTML files
Link checking with Xenu

- Xenu is freely available from:
- [http://home.snafu.de/tilman/xenulink.html](http://home.snafu.de/tilman/xenulink.html)
Xenu

[Image of Xenu software interface showing a list of web addresses and their status, type, size, and title.]
<table>
<thead>
<tr>
<th>Address</th>
<th>Status</th>
<th>Type</th>
<th>Size</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://lea.sagepub.com/">http://lea.sagepub.com/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://mend.endojournals.org/">http://mend.endojournals.org/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://pharmrev.aspetjournals.org/">http://pharmrev.aspetjournals.org/</a></td>
<td>pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://physiolgenomics.physiology.org/">http://physiolgenomics.physiology.org/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.physiologyonline.org/">http://www.physiologyonline.org/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.proteinscience.org/">http://www.proteinscience.org/</a></td>
<td>pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://stemcells.alphamedpress.org/">http://stemcells.alphamedpress.org/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ajcon.physiology.org/">http://ajcon.physiology.org/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.theannals.com/">http://www.theannals.com/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://adc.bmj.com/contents-by-date.0.shtml">http://adc.bmj.com/contents-by-date.0.shtml</a></td>
<td>pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://afram.oxfordjournals.org/archive/index.dtl">http://afram.oxfordjournals.org/archive/index.dtl</a></td>
<td>pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ageing.oxfordjournals.org/archive/index.dtl">http://ageing.oxfordjournals.org/archive/index.dtl</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://aje.oxfordjournals.org/archive">http://aje.oxfordjournals.org/archive</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ajccm.atsjournals.org/contents-by-date.0.shtml">http://ajccm.atsjournals.org/contents-by-date.0.shtml</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://aler.oxfordjournals.org/archive">http://aler.oxfordjournals.org/archive</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://alh.oxfordjournals.org/archive">http://alh.oxfordjournals.org/archive</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://als.dukejournals.org/archive/">http://als.dukejournals.org/archive/</a></td>
<td>pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://americanliterature.dukejournals.org/archive/">http://americanliterature.dukejournals.org/archive/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://americanspeech.dukejournals.org/archive/">http://americanspeech.dukejournals.org/archive/</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=">http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=</a>...</td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=">http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=</a>...</td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=">http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=</a>...</td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=">http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=</a>...</td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=">http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=</a>...</td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=">http://ams.allenpress.com/amsonline/?request=get-archive&amp;issn=</a>...</td>
<td>pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://aob.oxfordjournals.org/archive/index.dtl">http://aob.oxfordjournals.org/archive/index.dtl</a></td>
<td>pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://aplj.oxfordjournals.org/archive">http://aplj.oxfordjournals.org/archive</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ard.bmj.com/contents-by-date.0.shtml">http://ard.bmj.com/contents-by-date.0.shtml</a></td>
<td>pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://arjournals.annualreviews.org/lo/fluid">http://arjournals.annualreviews.org/lo/fluid</a></td>
<td>ok</td>
<td>text/html</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ready Threads: 27 574 of 884 URLs (64 %) done
Xenu
Xenu's broken link report

Created on August 22, 2007 at 18:11:26


Table of contents

- Broken links, ordered by link
- Broken links, ordered by page
- List of redirected URLs
- List of valid URLs you can submit to a search engine
- Site Map of HTML pages with a Title
- Broken page-local links
- Orphan files
- Statistics for managers

[the skeptic's Dictionary: the truth is in here]

Broken links, ordered by link:

http://accelibrary.aip.org/error/cookies.jsp?url=http%3a//accelibrary.aip.org/dbt/dbt.jsp%3fKEY=1JGNAl
error code: 404 (not found), linked from page(s):
   http://accelibrary.aip.org/dbt/dbt.jsp?KEY=1JGNAL

http://accelibrary.aip.org/error/cookies.jsp?url=http%3a//accelibrary.aip.org/dbt/dbt.jsp%3fKEY=3CCOF2
error code: 404 (not found), linked from page(s):
   http://accelibrary.aip.org/dbt/dbt.jsp?KEY=3CCOF2

Done
Broken links, ordered by link:

http://ascelibrary.aip.org/error/cookies.jsp?url=http%3a%2f%2fascelibrary.aip.org%2fdbt%2fjsp%3fKEY=JONAI
error code: 404 (not found), linked from page(s):
http://ascelibrary.aip.org/dbt/jsp?KEY=JONAI

http://ascelibrary.aip.org/error/cookies.jsp?url=http%3a%2f%2fascelibrary.aip.org%2fdbt%2fjsp%3fKEY=JCOOF2
error code: 404 (not found), linked from page(s):
http://ascelibrary.aip.org/dbt/jsp?KEY=JCOOF2

http://ascelibrary.aip.org/error/cookies.jsp?url=http%3a%2f%2fascelibrary.aip.org%2fdbt%2fjsp%3fKEY=JCMD4
error code: 404 (not found), linked from page(s):
http://ascelibrary.aip.org/dbt/jsp?KEY=JCMD4

http://ascelibrary.aip.org/error/cookies.jsp?url=http%3a%2f%2fascelibrary.aip.org%2fdbt%2fjsp%3fKEY=JOFEDU
error code: 404 (not found), linked from page(s):
http://ascelibrary.aip.org/dbt/jsp?KEY=JOFEDU

http://eupub.oxfordjournals.org/archive/
error code: 12017 (cancelled / timeout), linked from page(s):
http://eupub.oxfordjournals.org/archive

http://proquest.um.com/pgweb?RQT=318&pmid=58036
error code: 403 (forbidden request), linked from page(s):
file:///C:/Documents and Settings/simonh/My Documents/marc/generic/elinks-distrib/michelle/new/elink index ersa-ezproxy1007082118-44-

http://proquest.um.com/pgweb?RQT=318&pmid=38861
error code: 403 (forbidden request), linked from page(s):
file:///C:/Documents and Settings/simonh/My Documents/marc/generic/elinks-distrib/michelle/new/elink index ersa-ezproxy1007082118-44-

http://proquest.um.com/pgweb?RQT=318&pmid=46190
error code: 403 (forbidden request), linked from page(s):
file:///C:/Documents and Settings/simonh/My Documents/marc/generic/elinks-distrib/michelle/new/elink index ersa-ezproxy1007082118-44-

http://proquest.um.com/pgweb?RQT=318&pmid=56291
error code: 403 (forbidden request), linked from page(s):
file:///C:/Documents and Settings/simonh/My Documents/marc/generic/elinks-distrib/michelle/new/elink index ersa-ezproxy1007082118-44-
Xenu

http://www.sciencedirect.com/science/journal/02604779
error code: 400 (no object data), linked from page(s):
file:///C:/Documents and Settings/simonh/My Documents/marc/ge:

Exams database

- Uses voyager
- Records are in MARC format
- All BIB records, with no MFHDs
- Only 5,700 records in the database
- There are over 17,500 links
Quick search

Unit code: [ ] Search

- use ? to truncate eg acc1
- separate unit codes with spaces eg acc2351 ax2361

more search options | more help

Read more about the Past exams database

News

Database problems
The exams database is experiencing problems. We are working on fixing these. Apologies for any inconvenience.

Latest additions:
Semester 1, 2007 exams are now available.

Downtime:
Routine backup occurs every day between 2.30am-3.00am. During this time the past exams database may be unavailable.
Now searching: Monash University Past Exams Database
Search request: Search = (semester 1) in Keyword Anywhere AND(2007) in Keyword Anywhere
Search results: displaying 2 of 11 records

BTF3091 Stock Exchange and Derivatives Law

Unit: BTF3091 Stock Exchange and Derivatives Law
Department/Centre: Business Law and Taxation
Faculty: Faculty of Business and Economics
Campus: Clayton, Caulfield
Exam Papers: Semester 1, 2007
            Semester 1, 2006
            Semester 1, 2005
            Semester 1, 2004
            Semester 1, 2003
            Semester 2, 2000
Exams database link checker

- Shell and perl programs
- Runs on the server on 1st day of each month
- Emails broken links (only) directly to cataloguers
- No maintenance or running of the program manually
- Takes about 2 hours to run (so if I used it for our 500,000 links in our main catalogue – it would take 200 hours or about 10 days to run!)
Exams link checker

00 11 1 * * /m1/voyager/examtestdb/local/examscheck.sh

#!/bin/ksh
# Using this to do link checking of PDF files from the examtestdb database
# Simon Huggard 29 May 2007

cd /m1/voyager/examtestdb/local
exams-link-checker.pl
do-exams-checks.pl
cp header.txt results.txt
grep exampdf exams-broken.txt >>results.txt
cat results.txt | mailx -s 'Exam link checker results' Simon.Huggard@lib.monash.edu.au,michelle.rusiniak@lib.monash.edu.au,nghi.duong@lib.monash.edu.au
exams-link-checker.pl

- SQL query for all links in elink_index
- Creates "exampdfs.txt" (list of all 17,506 of them)

do-exams-checks.pl

It creates one file:
examscheck.txt

This file will list all of the status codes that the program finds when checking each URL.

The program first reads in exampdfs.txt and puts each line in to an array.

It then checks each URL as it goes through each one sequentially – and writes the output to examscheck.txt
open (FHANDLEIN, "filename") || die "Couldnt open file for reading: $!";

    # while can read another line from file
while (<FHANDLEIN>) {
    $pdffile = $_;
    `/usr/local/bin/wget -nv --append-output /m1/voyager/examtestdb/local/examscheck.txt -- user=libstats --password=xxxxxx --spider $pdffile`;

} # end while read file

close(FHANDLEIN);
Examscheck.txt

200 OK
200 OK
200 OK
200 OK
12:00:34 ERROR 404: Not Found.
12:00:34 ERROR 404: Not Found.
200 OK
200 OK
Examscheck.txt

• We use the following options with wget:
• -nv, --no-verbose (turn off verboseness, without being quiet - So we only get the URL if there’s an error.)
• -- spider (this means just crawl the site but don’t download the actual file)
Back to examscheck.sh

Examscheck.sh now just does a couple of things more:

```
cp header.txt.txt results.txt

grep examspdf exams-broken.txt >> results.txt
```
This is an output from the exams link checker program called examscheck.sh which is running on the Voyager database server. This program is automatically setup to run on the first day of each month.

Once the output is finished, it emails the results to Michelle Rusiniak, Nghi Duong and Simon Huggard.

The output below is a list of all the BROKEN links in the exams database.

examscheck.sh – emailing of results

cat results.txt | mailx -s 'Exam link checker results' Simon.Huggard@lib.monash.edu.au,michelle.rusiniak@lib.monash.edu.au,nghi.duong@lib.monash.edu.au
Subject: Exam link checker results
From: Voyager Owner <voyager@alexander.lib.monash.edu.au>
Date: 2007-08-01 12:21PM
To: Simon.Huggard@lib.monash.edu.au, michelle.rusiniak@lib.monash.edu.au, nghi.duong@lib.monash.edu.au

This is an output from the exams link checker program called examscheck.sh which is running on the Voyager database server.
This program is automatically setup to run on the first day of each month.
Once the output is finished, it emails the results to Michelle Rusiniak, Nghi Duong and Simon Huggard.

The output below is a list of all the BROKEN links in the exams database.

Summary

• Link checking is boring
• Automated programs help a lot (the less intervention the better)
• Future plan is to copy the examscheck scripts and perl programs to do what the elinks program does
Thank you – any questions?

• All programs and documentation are at:
• http://users.monash.edu.au/~simonh/IGeLU/

• Email: Simon.Huggard@lib.monash.edu.au