



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

www.lib.monash.edu.au

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 **Report Date: 15 April 2007**

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

Service description: Access to materials and resources

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

Further Information:

Service: No. 2 – Information Resources: Collection Management

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

Key Performance Indicator	Result	Date reported/ to be reported	Period covered by statistics	Frequency to be reported	Measured by
95% of web links in the library catalogue are accurate	98.5% of web links in the library catalogue were accurate.	April, July, October and January	January – March 2007	Quarterly	% of accurate links in the library catalogue identified by monthly link checking program

Link checking programs – Voyager – catjob 9 or 10

Table 2-4. Cataloging batch jobs

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

This is from the Reporter Users Guide

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 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://voydb.lib.monash.edu.au/webadmin/ Google

Getting Started Latest Headlines

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](#)

http://voydb.lib.monash.edu.au/webadmin/cgi-bin/catjob.cgi

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: required

Job Scheduler

[Return to Main Page](#)

Email notification to: Simon.Huggard@lib.monash.edu.au

job: catjob -j 10



Catjob.log file

Fri Aug 17 10:19:11 2007 Job execution begun.

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...

Output file – catrprts.SYS.inp

For Cataloging batch jobs, the file that is created is always named `catrprts.xxxx.inp`.

catrprts.SYS.inp

```
09|97.2|...|http://...|URL|0|Invalid Response Code|Bibliographic|2200297|08/17/2007
09|97.2|...|http://...|URL|0|Invalid Response Code|Bibliographic|2199943|08/17/2007
09|97.2|...|http://...|URL|0|Invalid Response Code|Bibliographic|2200307|08/17/2007
09|97.2|...|http://...|URL|0|Invalid Response Code|Bibliographic|2200119|08/17/2007
09|97.2|...|http://...|URL|0|Invalid Response Code|Bibliographic|2201512|08/17/2007
09|97.2|...|http://...|URL|0|Invalid Response Code|Bibliographic|2201749|08/17/2007
09|97.2|142.206.72.67|http://142.206.72.67/r000_e.htm|URL|302|Moved Temporarily|Bibliographic|5974|08/17/2007
09|97.2|19thc-artworldwide.org|http://19thc-artworldwide.org|URL|0|Invalid Response Code|Bibliographic|2177891|08/17/2007
09|97.2|217.160.60.64|http://217.160.60.64/fb/link4.php?id=103|URL|0|Invalid Response Code|Bibliographic|1731442|08/17/2007
09|97.2|217.160.60.64|http://217.160.60.64/fb/link4.php?id=103|URL|0|Invalid Response Code|Holdings|2012540|08/17/2007
09|97.2|217.160.60.64|http://217.160.60.64/fb/link4.php?id=108|URL|0|Invalid Response Code|Bibliographic|1724384|08/17/2007
09|97.2|217.160.60.64|http://217.160.60.64/fb/link4.php?id=108|URL|0|Invalid Response Code|Holdings|2006127|08/17/2007
09|97.2|217.160.60.64|http://217.160.60.64/fb/link4.php?id=11|URL|0|Invalid Response Code|Bibliographic|1471800|08/17/2007
09|97.2|217.160.60.64|http://217.160.60.64/fb/link4.php?id=11|URL|0|Invalid Response Code|Holdings|2006950|08/17/2007
09|97.2|217.160.60.64|http://217.160.60.64/fb/link4.php?id=111|URL|0|Invalid Response Code|Bibliographic|1725481|08/17/2007
```



Entries in the log file

09|97.2||ezproxy.lib.monash.edu.au|http://ezproxy.lib.monash.edu.au/login?url=http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&res_id=xri:eebo&rft_val_fmt=&rft_id=xri:eebo:image:35198|URL| **0|Invalid Response Code**
|Holdings|2178856|08/17/2007

09|97.2||name.umdl.umich.edu|http://name.umdl.umich.edu/AAN7780|URL| **404|Not Found**
|Bibliographic|2105711|08/17/2007

Document Images - EEBO - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://eebo.chadwyck.com.ezproxy.lib.monash.edu.au/search/full_rec?SOURCE=pgimages.cfg&ACTION=By






Getting Started Latest Headlines

Display File University of Michigan Library Name R... Document Images - EEBO

EEBO EARLY ENGLISH BOOKS ONLINE

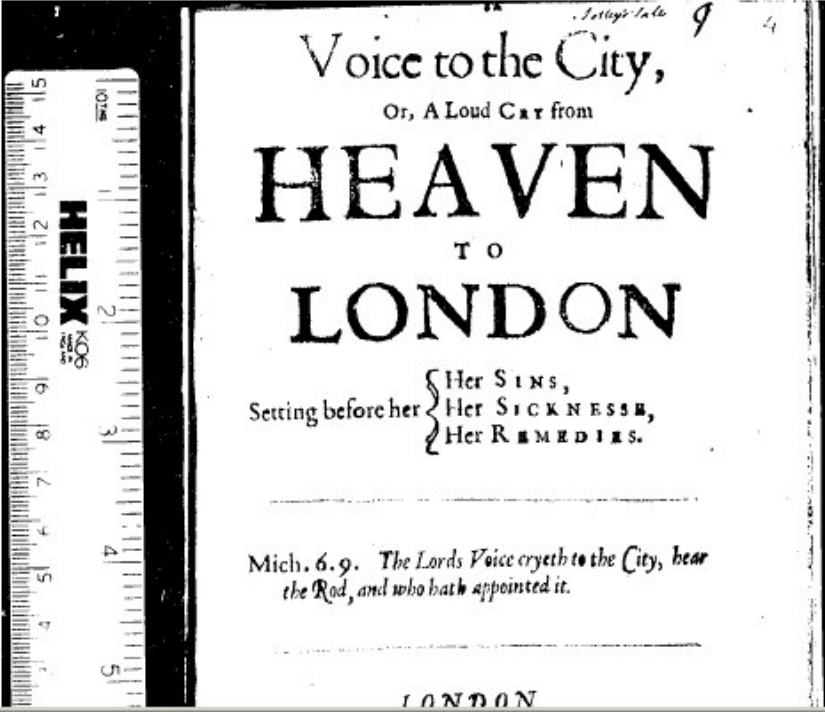
Document Images HOME MARKED LIST SEARCH HISTORY HELP ?

SEARCH: Basic | Advanced | Periodicals BROWSE: Authors | Thomason Tracts | Periodicals

View this document as:      [Help on downloading](#)

Add this record to your Marked List [Durable URL](#)

<< Previous illustration | Next illustration >> << Previous image | Next image >>



1649

Voice to the City,
Or, A Loud CRY from
HEAVEN
TO
LONDON

Setting before her { Her SINS,
Her SICKNESSE,
Her REMEDIES.

Mich. 6. 9. *The Lords Voice cryeth to the City, hear
the Rod, and who hath appointed it.*

LONDON

Done

Library

University of Michigan Library Name Resolver Service - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://quod.lib.umich.edu/cgi/b/bib/bibperm?q1=AAN7780

Getting Started Latest Headlines

Display File University of Michigan Library Na...

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.

1 of 1

Author: Cooper, James Fenimore, 1789-1851.
Title: Homeward bound; or, The chase. A tale of the sea. By J. Fenimore Cooper.
URL: <http://www.hti.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

1 of 1

Done

Printed Link Failure Report

856 Link Failure Report

08/17/2007

Date Range:

Link Host

name.umd.umich.edu

Link Type: URL

Link Text: *http://name.umd.umich.edu/AAN7780*

Error Code: 404

Error Text: *Not Found*

Record Type: *Holdings*

Record ID: 2178856

Date Updated: 08/17/2007

ezproxy.lib.monash.edu.au

Link Type: URL

Link Text: *http://ezproxy.lib.monash.edu.au/login?url=http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&res_id=xri*

Error Code: 0

Error Text: *Invalid Response Code*

Record Type: *Bibliographic*

Record ID: 2105711

Date Updated: 08/17/2007

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

Microsoft Visual FoxPro

File Edit Window Help

Voyager Elinks

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,

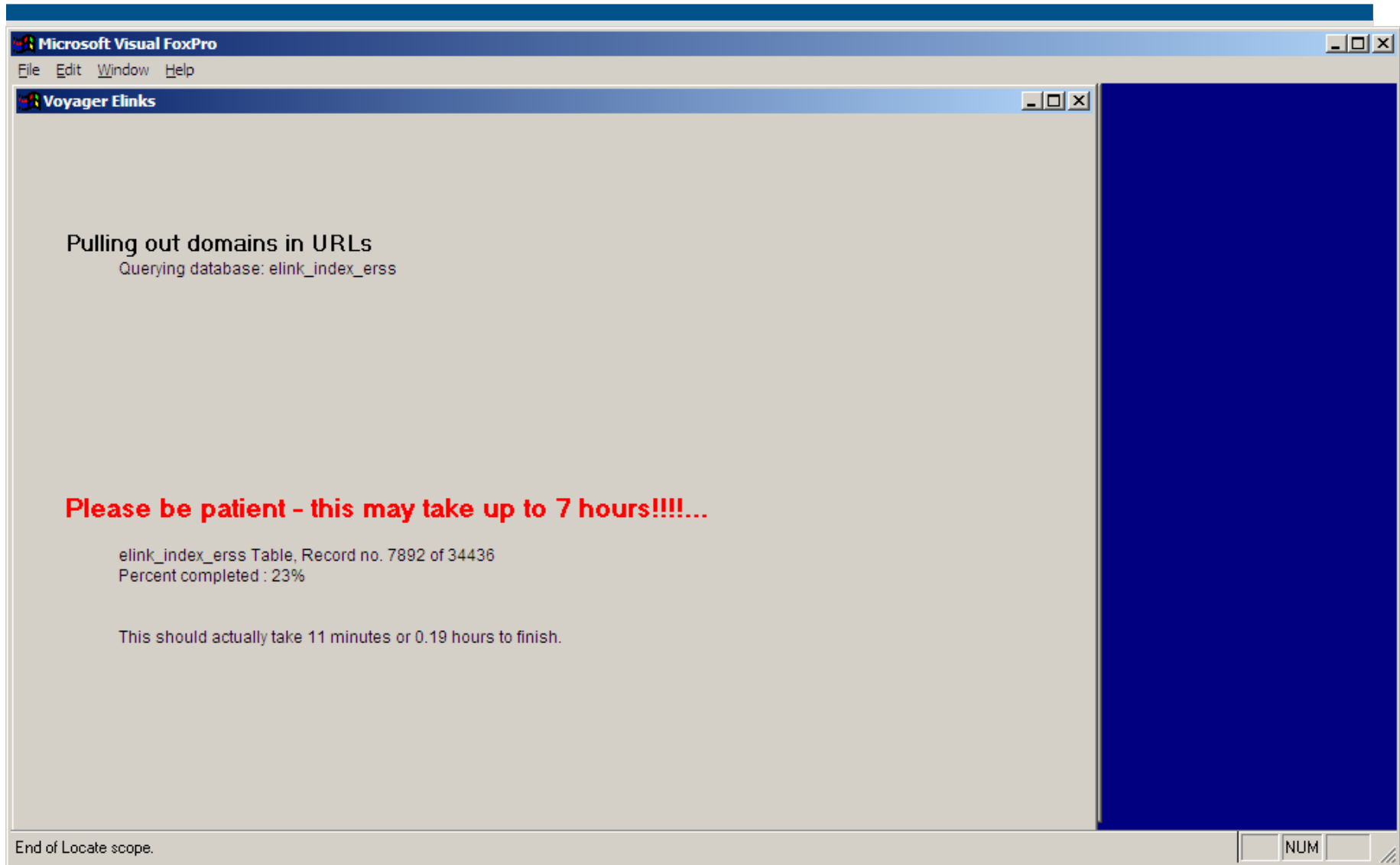
Program written in Microsoft Visual FoxPro 7 by Simon Hugg
21 August, 2001. Updated 21 August 2007 6:20pm

Running query... 51%

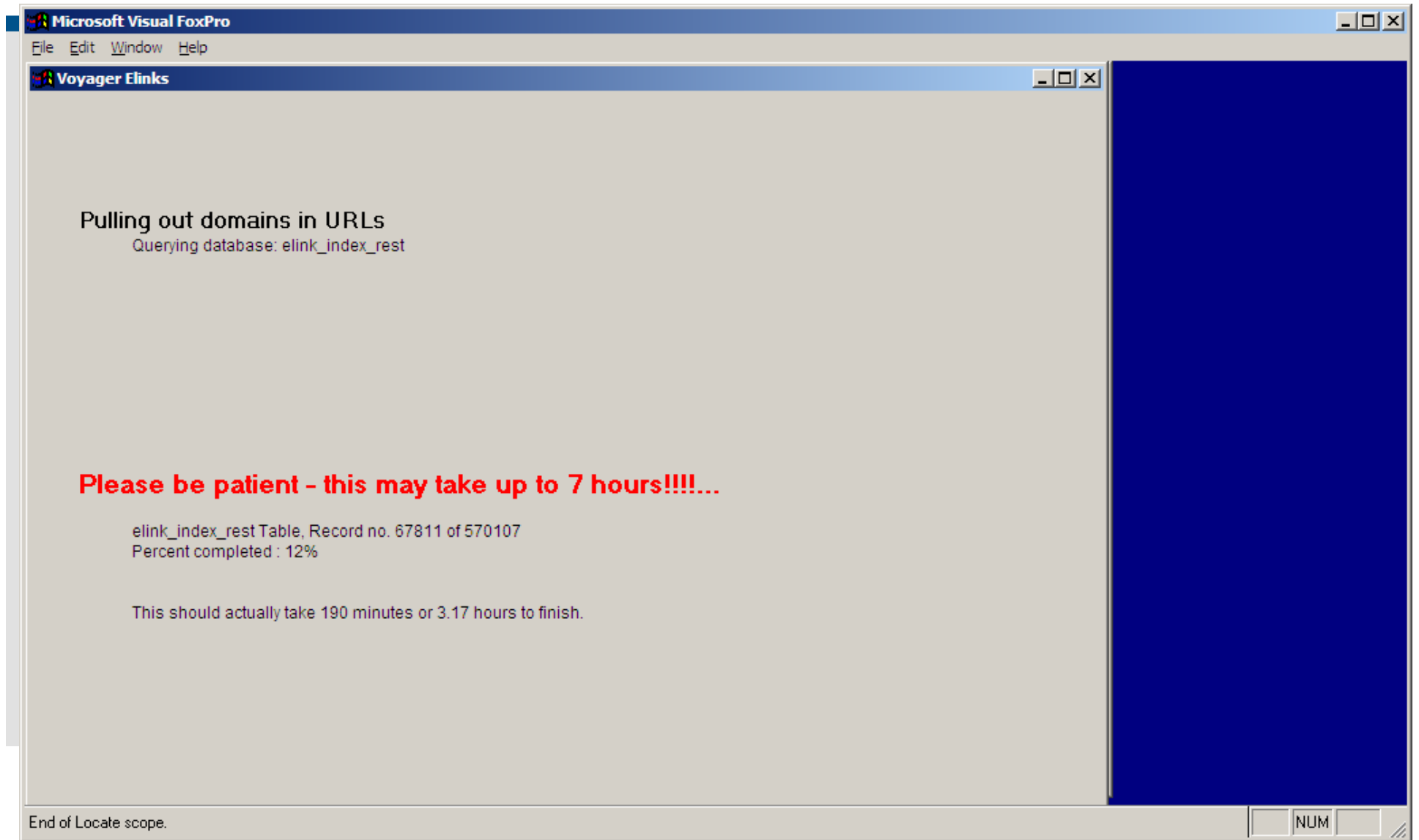
March - May 2007 - new code for SS records, TOC & other links and to fix some bugs
30 May 2005 - added code to remove MLO records
11-21 March 2005 - made changes to speed up the program.
24 March 2005 - added code to remove ERES locations
10 March 2005 - no longer gets suppressed records.
4 March 2004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

Record: 1/112127 Exclusive NUM







Elinks program in detail



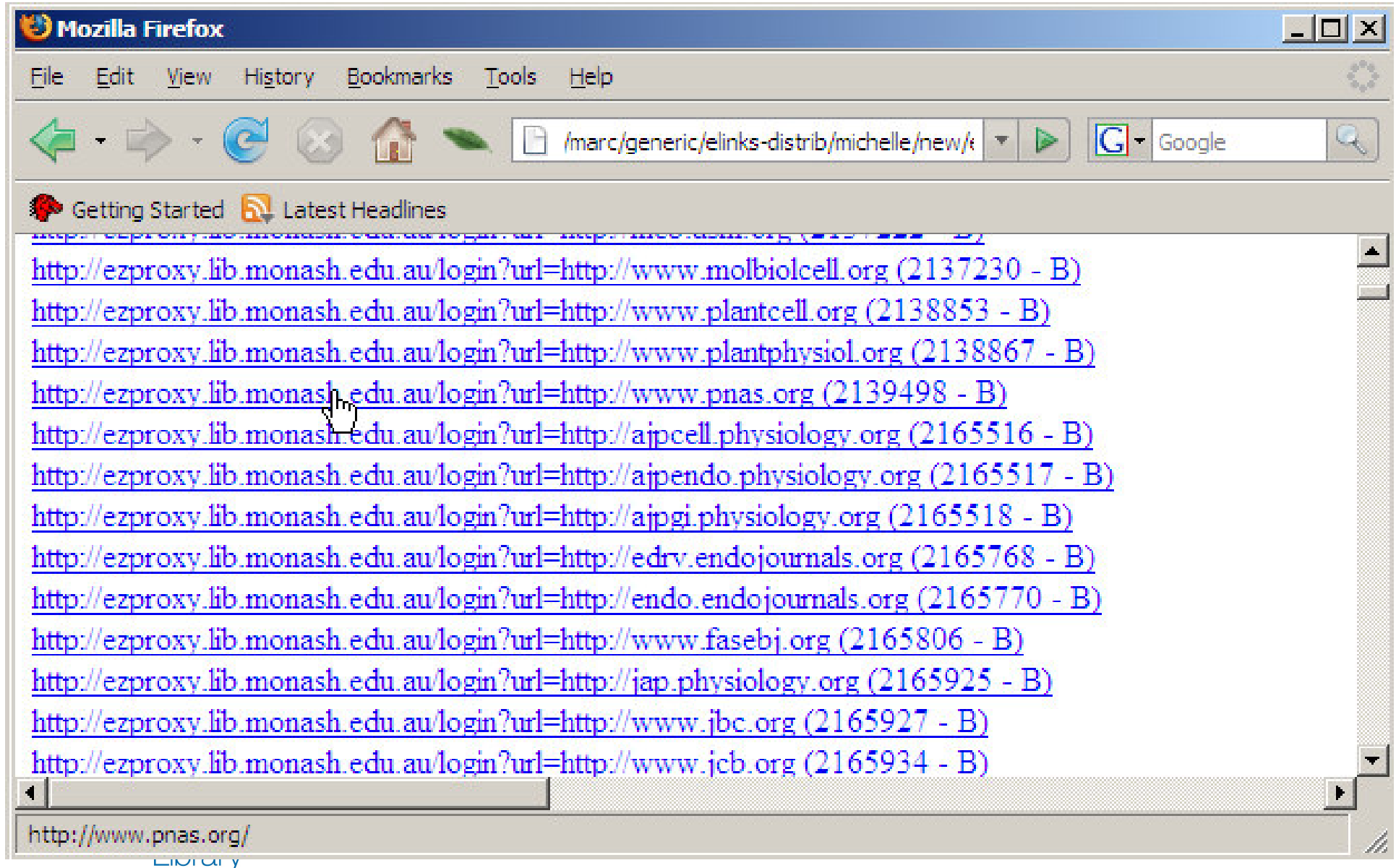
Elinks program



Elinks program – detailed description

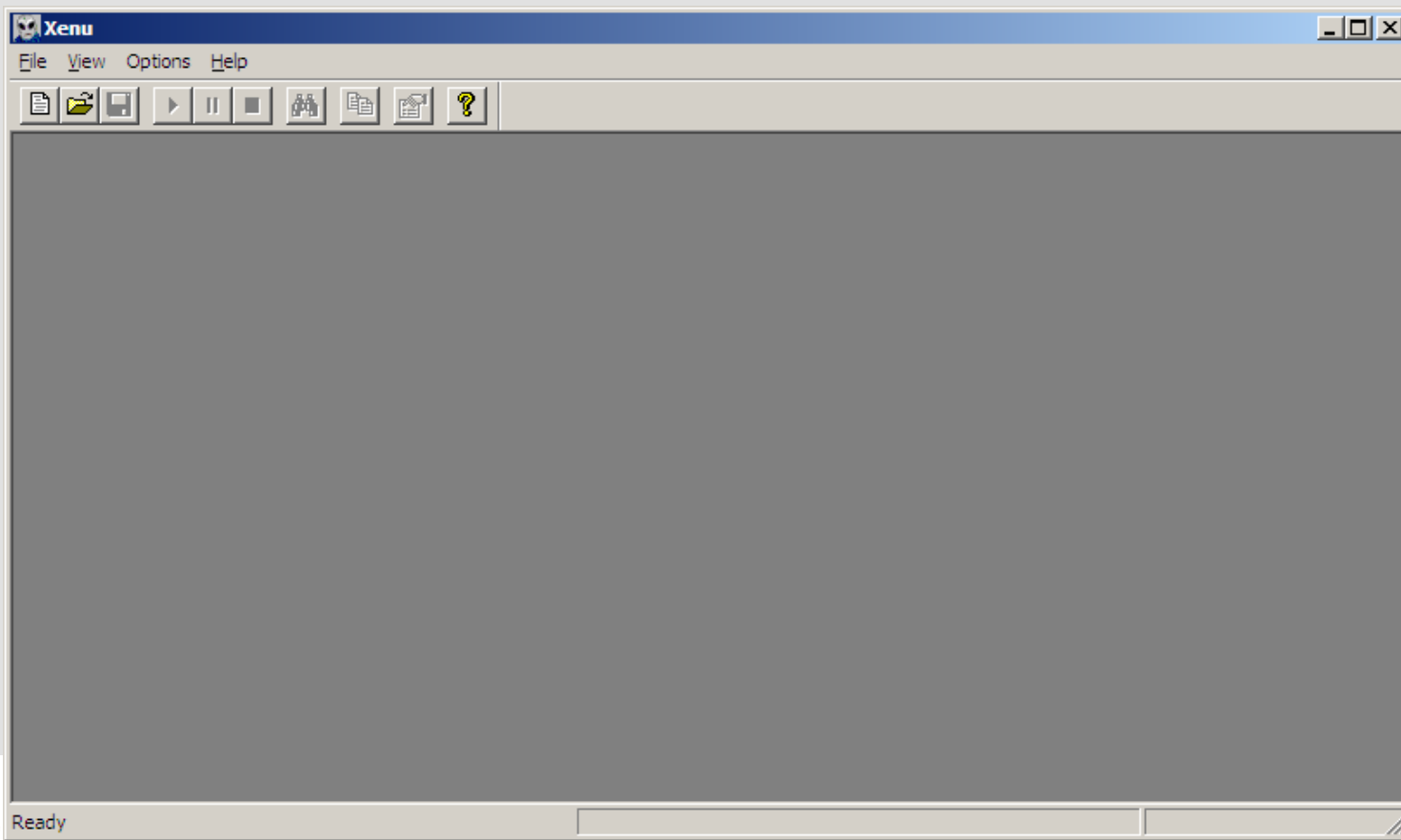
Name ▲	Size	Type	Date Modified
 elink_index_erss2007082116-02-43.html	117 KB	Firefox Document	2007-08-21 4:02 PM
 elink_index_erss2007082116-02-45.html	144 KB	Firefox Document	2007-08-21 4:02 PM
 elink_index_erss2007082116-02-48.html	159 KB	Firefox Document	2007-08-21 4:02 PM
 elink_index_erss2007082116-02-51.html	153 KB	Firefox Document	2007-08-21 4:02 PM
 elink_index_erss2007082116-02-54.html	34 KB	Firefox Document	2007-08-21 4:02 PM
 elink_index_erss-ezproxy2007082116-02-40.html	118 KB	Firefox Document	2007-08-21 4:02 PM

Elinks HTML files

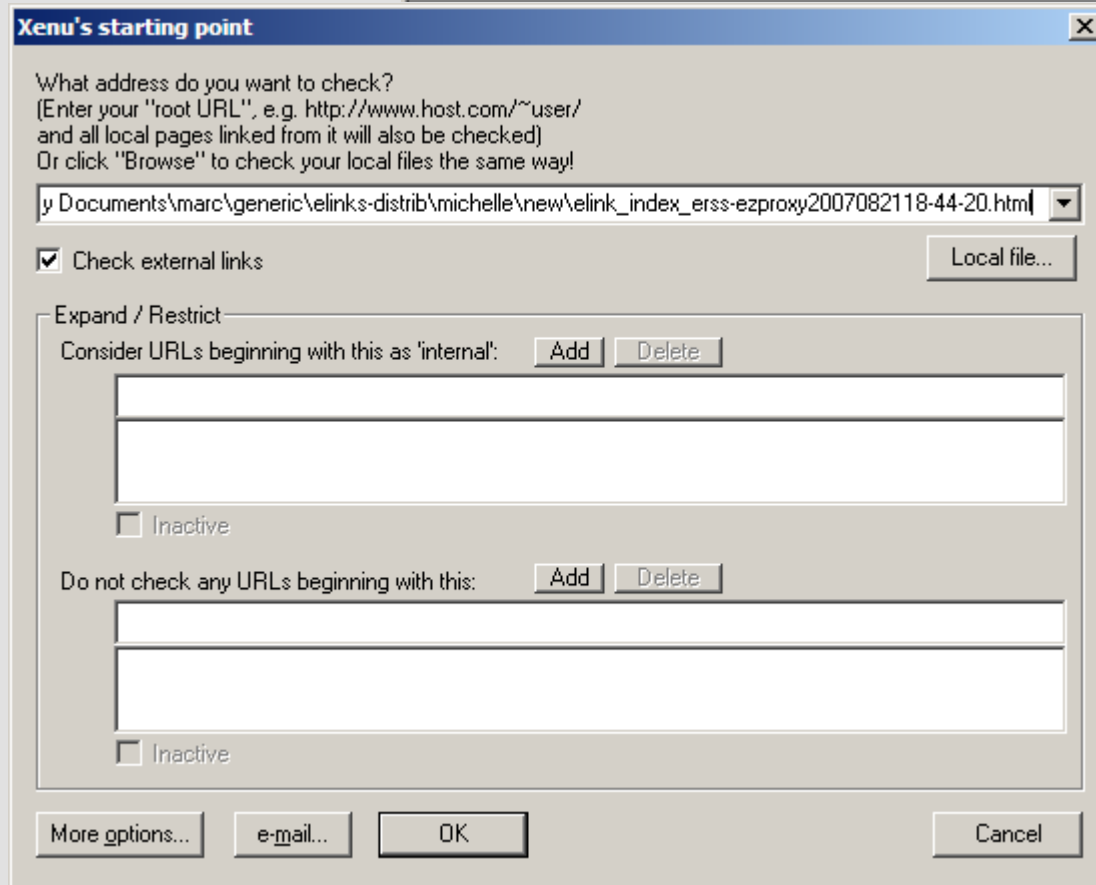
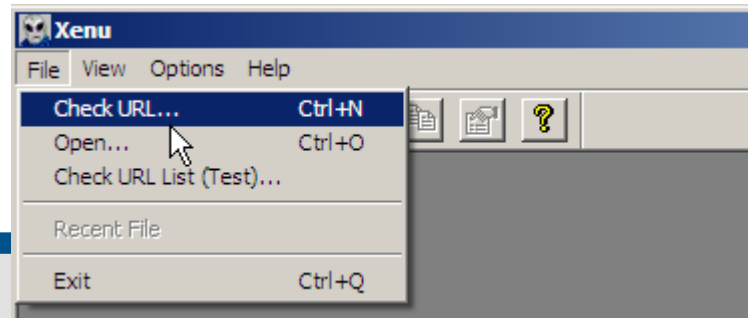


Link checking with Xenu

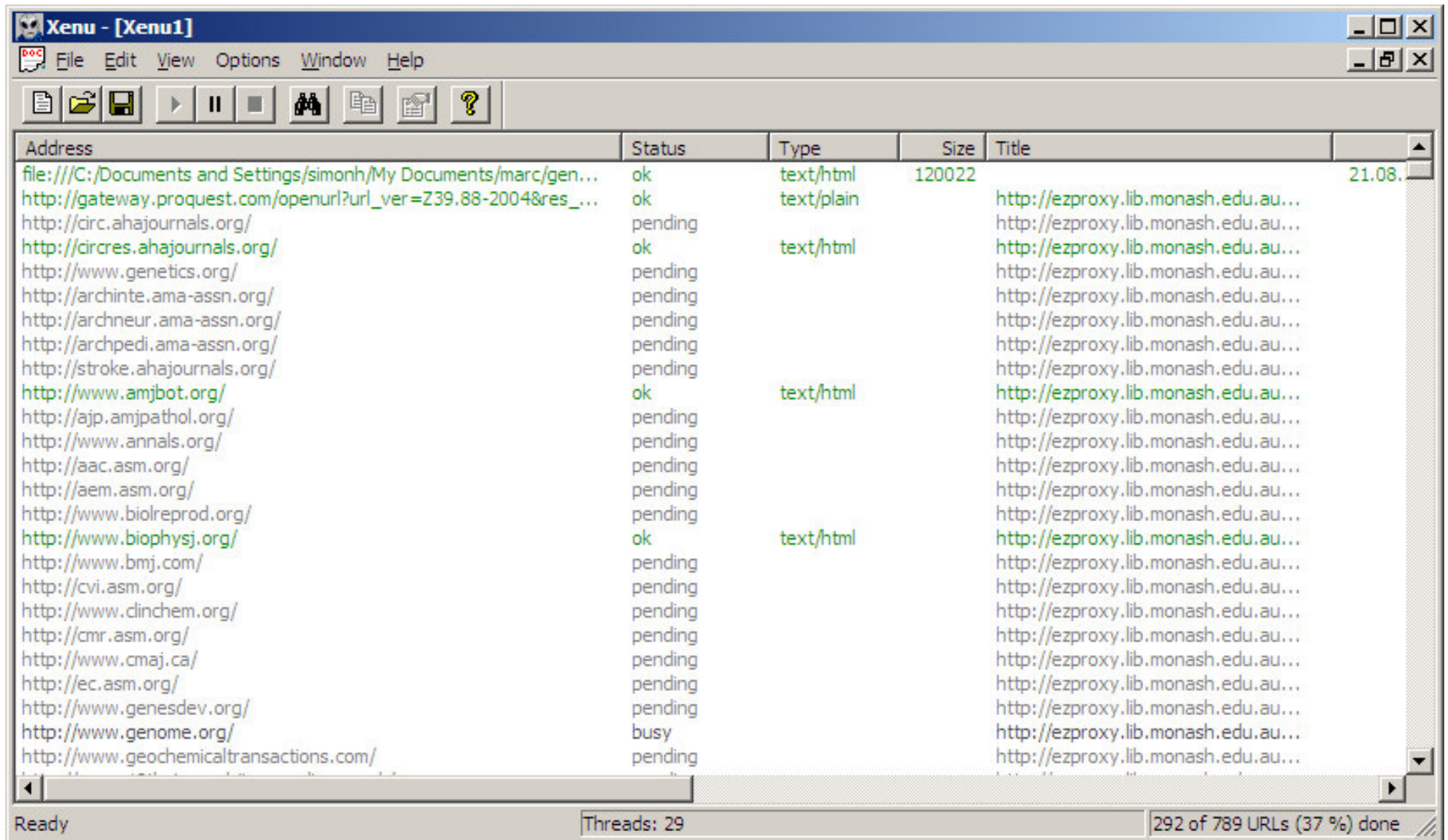
- **Xenu is freely available from:**
- <http://home.snafu.de/tilman/xenulink.html>



Xenu



Xenu

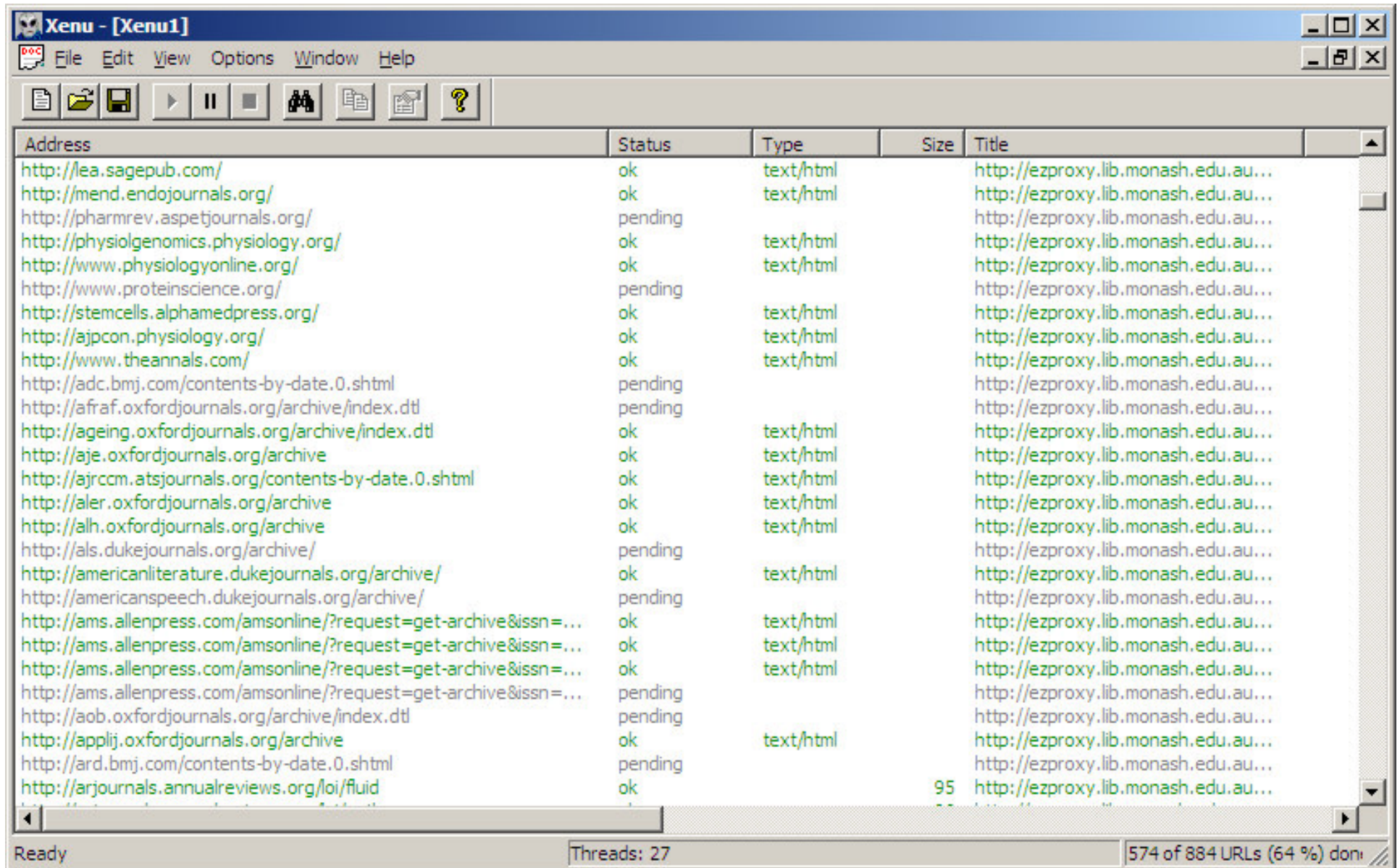


The screenshot shows the Xenu software interface. The window title is "Xenu - [Xenu1]". The menu bar includes "File", "Edit", "View", "Options", "Window", and "Help". The toolbar contains icons for file operations and navigation. The main area displays a table of download results:

Address	Status	Type	Size	Title
file:///C:/Documents and Settings/simonh/My Documents/marc/gen...	ok	text/html	120022	21.08.
http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&res_...	ok	text/plain		http://ezproxy.lib.monash.edu.au...
http://circ.ahajournals.org/	pending			http://ezproxy.lib.monash.edu.au...
http://circres.ahajournals.org/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://www.genetics.org/	pending			http://ezproxy.lib.monash.edu.au...
http://archinte.ama-assn.org/	pending			http://ezproxy.lib.monash.edu.au...
http://archneur.ama-assn.org/	pending			http://ezproxy.lib.monash.edu.au...
http://archpedi.ama-assn.org/	pending			http://ezproxy.lib.monash.edu.au...
http://stroke.ahajournals.org/	pending			http://ezproxy.lib.monash.edu.au...
http://www.amjbot.org/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ajp.amjpathol.org/	pending			http://ezproxy.lib.monash.edu.au...
http://www.annals.org/	pending			http://ezproxy.lib.monash.edu.au...
http://aac.asm.org/	pending			http://ezproxy.lib.monash.edu.au...
http://aem.asm.org/	pending			http://ezproxy.lib.monash.edu.au...
http://www.biolreprod.org/	pending			http://ezproxy.lib.monash.edu.au...
http://www.biophysj.org/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://www.bmj.com/	pending			http://ezproxy.lib.monash.edu.au...
http://cvi.asm.org/	pending			http://ezproxy.lib.monash.edu.au...
http://www.dinchem.org/	pending			http://ezproxy.lib.monash.edu.au...
http://cmr.asm.org/	pending			http://ezproxy.lib.monash.edu.au...
http://www.cmaj.ca/	pending			http://ezproxy.lib.monash.edu.au...
http://ec.asm.org/	pending			http://ezproxy.lib.monash.edu.au...
http://www.genesdev.org/	pending			http://ezproxy.lib.monash.edu.au...
http://www.genome.org/	busy			http://ezproxy.lib.monash.edu.au...
http://www.geochemicaltransactions.com/	pending			http://ezproxy.lib.monash.edu.au...

The status bar at the bottom shows "Ready", "Threads: 29", and "292 of 789 URLs (37 %) done".

Xenu



The screenshot shows the Xenu web crawler interface. The window title is "Xenu - [Xenu1]". The menu bar includes "File", "Edit", "View", "Options", "Window", and "Help". The toolbar contains icons for file operations and navigation. The main area displays a table of crawling progress.

Address	Status	Type	Size	Title
http://lea.sagepub.com/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://mend.endojournals.org/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://pharmrev.aspetjournals.org/	pending			http://ezproxy.lib.monash.edu.au...
http://physiolgenomics.physiology.org/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://www.physiologyonline.org/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://www.proteinscience.org/	pending			http://ezproxy.lib.monash.edu.au...
http://stemcells.alphamedpress.org/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ajpcon.physiology.org/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://www.theannals.com/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://adc.bmj.com/contents-by-date.0.shtml	pending			http://ezproxy.lib.monash.edu.au...
http://afraf.oxfordjournals.org/archive/index.dtl	pending			http://ezproxy.lib.monash.edu.au...
http://ageing.oxfordjournals.org/archive/index.dtl	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://aje.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ajrccm.atsjournals.org/contents-by-date.0.shtml	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://aler.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://alh.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://als.dukejournals.org/archive/	pending			http://ezproxy.lib.monash.edu.au...
http://americanliterature.dukejournals.org/archive/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://americanspeech.dukejournals.org/archive/	pending			http://ezproxy.lib.monash.edu.au...
http://ams.allenpress.com/amsonline/?request=get-archive&issn=...	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ams.allenpress.com/amsonline/?request=get-archive&issn=...	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ams.allenpress.com/amsonline/?request=get-archive&issn=...	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ams.allenpress.com/amsonline/?request=get-archive&issn=...	pending			http://ezproxy.lib.monash.edu.au...
http://aob.oxfordjournals.org/archive/index.dtl	pending			http://ezproxy.lib.monash.edu.au...
http://applij.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ard.bmj.com/contents-by-date.0.shtml	pending			http://ezproxy.lib.monash.edu.au...
http://arjournals.annualreviews.org/loi/fluid	ok		95	http://ezproxy.lib.monash.edu.au...

Ready Threads: 27 574 of 884 URLs (64 %) done

Xenu

The screenshot shows the Xenu software interface. The main window displays a list of URLs with columns for Address, Status, Type, Size, and Title. A dialog box titled "Xenu" is overlaid on the list, asking "Link sleuth finished. Do you want a report?" with "Yes" and "No" buttons. The status bar at the bottom shows "Ready", "Threads: 0", and "976 of 976 URLs (100 %) doi".

Address	Status	Type	Size	Title
http://adc.bmj.com/contents-by-date.0.shtml	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://afraf.oxfordjournals.org/archive/index.dtl	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ageing.oxfordjournals.org/archive/index.dtl	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://aje.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ajrccm.atsjournals.org/contents-by-date.0.shtml	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://aler.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://alh.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://als.dukejournals.org/archive/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://americanliterature.dukejournals.org/archive/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://americanspeech.dukejournals.org/archive/	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ams.allenpress.com/amsonline/?request=get-ar	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ams.allenpress.com/amsonline/?request=get-ar	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ams.allenpress.com/amsonline/?request=get-ar	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ams.allenpress.com/amsonline/?request=get-ar	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://aob.oxfordjournals.org/archive/index.dtl	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://aplij.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://ard.bmj.com/contents-by-date.0.shtml	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://arjournals.annualreviews.org/loi/fluid	ok		95	http://ezproxy.lib.monash.edu.au...
http://arjournals.annualreviews.org/loi/anthro	ok		96	http://ezproxy.lib.monash.edu.au...
http://arjournals.annualreviews.org/loi/astro	ok		95	http://ezproxy.lib.monash.edu.au...
http://arjournals.annualreviews.org/loi/biochem	ok		97	http://ezproxy.lib.monash.edu.au...
http://ascelibrary.aip.org/dbt/dbt.jsp?KEY=IJGNAI	ok			http://ezproxy.lib.monash.edu.au...
http://ascelibrary.aip.org/dbt/dbt.jsp?KEY=JCCOF2	ok			http://ezproxy.lib.monash.edu.au...
http://ascelibrary.aip.org/dbt/dbt.jsp?KEY=JCEMD4	ok			http://ezproxy.lib.monash.edu.au...
http://ascelibrary.aip.org/dbt/dbt.jsp?KEY=JOEEDU	ok			http://ezproxy.lib.monash.edu.au...
http://beheco.oxfordjournals.org/archive	ok	text/html		http://ezproxy.lib.monash.edu.au...
http://bioinformatics.oxfordjournals.org/archive/index.dtl	ok	text/html		http://ezproxy.lib.monash.edu.au...

Xenu's broken link report for August 22, 2007 at 18:11:26 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

file:///C:/DOCUME~1/simonh/LOCALS~1/Temp/TGH167.htm

Getting Started Latest Headlines

Xenu's broken link report

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

Root URL: file:///C:/Documents and Settings/simonh/My Documents/marc/generic/elinks-distrib/michelle/new/elink_index_erss-ezproxy2007082118-44-20.html

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

<http://ascelibrary.aip.org/error/cookies.jsp?url=http%3a//ascelibrary.aip.org/dbt/dbt.jsp%3fKEY=JCCOF2>
error code: 404 (not found), linked from page(s):
<http://ascelibrary.aip.org/dbt/dbt.jsp?KEY=JCCOF2>

Done



Xenu's broken link report for August 22, 2007 at 18:11:26 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

file:///C:/DOCUME~1/simonh/LOCALS~1/Temp/TGH167.htm

Getting Started Latest Headlines

Broken links, ordered by link:

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

<http://ascelibrary.aip.org/error/cookies.jsp?url=http%3a//ascelibrary.aip.org/dbt/dbt.jsp%3fKEY=JCCOF2>
error code: 404 (not found), linked from page(s):
<http://ascelibrary.aip.org/dbt/dbt.jsp?KEY=JCCOF2>

<http://ascelibrary.aip.org/error/cookies.jsp?url=http%3a//ascelibrary.aip.org/dbt/dbt.jsp%3fKEY=JCEMD4>
error code: 404 (not found), linked from page(s):
<http://ascelibrary.aip.org/dbt/dbt.jsp?KEY=JCEMD4>

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

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

<http://proquest.umi.com/pqdweb?RQT=318&pmid=36036>
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 erss-ezproxy2007082118-44-2>

<http://proquest.umi.com/pqdweb?RQT=318&pmid=38591>
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 erss-ezproxy2007082118-44-2>

<http://proquest.umi.com/pqdweb?RQT=318&pmid=46160>
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 erss-ezproxy2007082118-44-2>

<http://proquest.umi.com/pqdweb?RQT=318&pmid=55291>
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 erss-ezproxy2007082118-44-2>

Done



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
```

```
http://ezproxy.lib.monash.edu.au/login?url=http://www.sciencedirect.com/science/journal/00220728 (2124874 - B)  
http://ezproxy.lib.monash.edu.au/login?url=http://www.sciencedirect.com/science/journal/02604779 (2124914 - B)  
http://ezproxy.lib.monash.edu.au/login?url=http://www.sciencedirect.com/science/journal/00275107 (2124918 - B)  
http://ezproxy.lib.monash.edu.au/login?url=http://www.sciencedirect.com/science/journal/09692126 (2124971 - B)  
http://ezproxy.lib.monash.edu.au/login?url=http://www.scientific.net/1012-0386 (2222109 - B)
```

Find: Match case

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**



Past exams database

Quick search

Unit code: Search

- use ? to truncate eg *acc1?*
- separate unit codes with spaces eg *acc2351 afx2351*

[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.

Need help? Library frequently asked questions and online inquiries: [current students/staff](#) | [public users](#), [online chat](#), or phone +61 3 9905 5054

Something to say? Use our online inquiry service to send us your feedback and suggestions: [current students/staff](#) | [public users](#)

[Library home](#)

[my.monash](#)


© 2006 [Monash University](#) - [Caution](#) - [Privacy](#)

Record - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://exams.lib.monash.edu.au/cgi-bin/Pwebrecon.cgi?v1=2&ti=1,2&SEQ=20070823175526

Getting Started Latest Headlines

 **Past exams database**

[Back to results](#) [New search](#) [Search history](#) [Clear history](#) [Help](#)

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

[Previous](#) [Next](#)

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](#)

[Previous](#) [Next](#)

Done

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 examspdf 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)

<http://examspdf.lib.monash.edu.au/pdf/cl1996/04084163.pdf>

<http://examspdf.lib.monash.edu.au/pdf/cl1995/04083421.pdf>

<http://examspdf.lib.monash.edu.au/pdf/cl1994/04078156.pdf>

<http://examspdf.lib.monash.edu.au/pdf/cl1993/04074633.pdf>

<http://examspdf.lib.monash.edu.au/pdf/cl1997/04087949.pdf>

<http://examspdf.lib.monash.edu.au/pdf/cl1996/04084163.pdf>

<http://examspdf.lib.monash.edu.au/pdf/cl1995/04083421.pdf>

<http://examspdf.lib.monash.edu.au/pdf/cl1994/04078156.pdf>



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

do-exams-checks.pl

```
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

<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088255.pdf>:

12:00:34 ERROR 404: Not Found.

<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088256.pdf>:

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 verbosity, 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 results.txt
```

```
grep examspdf exams-broken.txt  
>>results.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.

<http://exampdf.lib.monash.edu.au/pdf/cl1997/04088255.pdf>

<http://exampdf.lib.monash.edu.au/pdf/cl1997/04088256.pdf>

<http://exampdf.lib.monash.edu.au/pdf/cl1997/04088258.pdf>

<http://exampdf.lib.monash.edu.au/pdf/cl1997/04088259.pdf>



examscheck.sh – emailing of results

```
cat results.txt | mailx -s 'Exam link checker  
results' Simon.Huggard@lib.monash.  
edu.au,michelle.rusiniak@lib.monash.ed  
u.au,nghi.duong@lib.monash.edu.au
```


Exam link checker results - Thunderbird

File Edit View Go Message Tools Help

Get Mail Write Address Book Reply Reply All Forward Tag Delete Junk Print Back Forward

Subject: Exam link checker results
From: Voyager Owner <voyager@alexander.lib.monash.edu.au>
Date: 2007-08-01 12:21 PM
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.

<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088255.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088256.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088258.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088259.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1999/04091955.pdf>
<http://examspdf.lib.monash.edu.au/pdf/gl1998/04080073.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cf1999/04091309.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088304.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cf1998/04090640.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cf1998/04090637.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cf1998/04090640.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cf1998/04090637.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088305.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088321.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088324.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088326.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088327.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088328.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088329.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088330.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088264.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088333.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088335.pdf>
<http://examspdf.lib.monash.edu.au/pdf/cl1997/04088336.pdf>

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**