LINK ON THE BLINK? USING CHECKERS TO FIND SITE WRECKERS
--Fred Heidenreich, Information Services Librarian--Paul Bracke, Head of Systems and Networking--Arizona Health Sciences Library--University of Arizona--Tucson, Arizona--

What do link checkers do?
Link checkers examine the validity of links within a document, or an entire web site. They automatically request pages from a web server, and judge whether or not a link is broken based on the resulting HTTP response code. A few sample codes:
- 200=OK
- 301=Moved Permanently
- 307=Temporary Redirect
- 401=Authorization Required
- 403=Access Forbidden
- 404=Page Not Found

What kinds of link checkers are available?
Link checkers generally fall into one of three categories:
- Stand-alone applications (example: Xenu)
- Web service (example: Hotpulse)
- Integrated into a web authoring tool (example: FrontPage, Dreamweaver)

For a full list of available checkers, see:
- Google Directory: Link Management:
  http://directory.google.com/Top/Computers/Software/Internet/Website_Management/Link_Management/
- Hardin MD: Link Checkers:
  http://www.ib.usce.edu/hardin/ml/linkcheckers.html
- Library Web Manager's Reference Center:
  http://sunsite.berkeley.edu/Web4Lib/RefCenter/hardin/ml/linkcheckers.html

What kind of errors do link checkers find and how well do they work?
- Broken or lapsed HTML links
- A page has been temporarily moved
- A syntactic error is present in the link (e.g., a trailing slash is omitted from the URL)
- There are network problems
- The server is temporarily down

Research has shown that educational WWW hyperlinks in the sciences are highly ephemeral. Broken links also referred to as “link rot” is thought to be the second most serious problem on the Web - following slow connections - see table illustrating the loss of distributed e-resources.

<table>
<thead>
<tr>
<th>How Rapidly Do Science Education Hyperlinks Become Extinct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research has shown that educational WWW hyperlinks in the sciences are highly ephemeral. Broken links also referred to as “link rot” is thought to be the second most serious problem on the Web - following slow connections - see table illustrating the loss of distributed e-resources.</td>
</tr>
</tbody>
</table>

What kind of errors do link checkers find and how well do they work?
- Broken or lapsed HTML links
- A page has been temporarily moved
- A syntactic error is present in the link (e.g., a trailing slash is omitted from the URL)
- There are network problems
- The server is temporarily down

Some link checkers scan websites only for 404 (“Page not Found”) and 403 (“Access Forbidden”) errors, and often not check links in Javascript, frames, or Flash.

Dynamic sites have special requirements for link checking.

How do link checkers find and how well do they work?
- Broken or lapsed HTML links
- A page has been temporarily moved
- A syntactic error is present in the link (e.g., a trailing slash is omitted from the URL)
- There are network problems
- The server is temporarily down

Some link checkers scan websites only for 404 (“Page not Found”) and 403 (“Access Forbidden”) errors, and often not check links in Javascript, frames, or Flash.

Dynamic sites have special requirements for link checking.

I have some broken links, what next?
- See if the site has just moved and update link
- Identify alternate sites providing the same kinds of information
- Or delete

For a full list of available checkers, see:
- Google Directory: Link Management:
  http://directory.google.com/Top/Computers/Software/Internet/Website_Management/Link_Management/
- Hardin MD: Link Checkers:
  http://www.ib.usce.edu/hardin/ml/linkcheckers.html
- Library Web Manager’s Reference Center:
  http://sunsite.berkeley.edu/Web4Lib/RefCenter/hardin/ml/linkcheckers.html