

Link rot

Link rot (also called **link death**, **link breaking**, or **reference rot**) is the phenomenon of hyperlinks tending over time to cease to point to their originally targeted file, web page, or server due to that resource being relocated to a new address or becoming permanently unavailable. A link that no longer points to its target, often called a *broken* or *dead* link, is a specific form of dangling pointer.

The rate of link rot is a subject of study and research due to its significance to the internet's ability to preserve information. Estimates of that rate vary dramatically between studies.

Contents

Prevalence

Causes

Prevention and detection

See also

Further reading

Notes & references

External links

Prevalence

A number of studies have examined the prevalence of link rot within the World Wide Web, in academic literature that uses URLs to cite web content, and within digital libraries.

A 2003 study found that on the Web, about one link out of every 200 broke each week,^[1] suggesting a half-life of 138 weeks. This rate was largely confirmed by a 2016–2017 study of links in Yahoo! Directory (which had stopped updating in 2014 after 21 years of development) that found the half-life of the directory's links to be two years.^[2]

A 2004 study showed that subsets of Web links (such as those targeting specific file types or those hosted by academic institution) could have dramatically different half-lives.^[3] The URLs selected for publication appear to have greater longevity than the average URL. A 2015 study by Weblock analyzed more than 180,000 links from references in the full-text corpora of three major open access publishers and found a half-life of about 14 years,^[4] generally confirming a 2005 study that found that half of the URLs cited in D-Lib Magazine articles were active 10 years after publication.^[5] Other studies have found higher rates of link rot in academic literature but typically suggest a half-life of four years or greater.^{[6][7]} A 2013 study in BMC Bioinformatics analyzed nearly 15,000 links in abstracts from Thomson Reuters's Web of Science citation index and found that the median lifespan of web pages was 9.3 years, and just 62% were archived.^[8]

A 2002 study suggested that link rot within digital libraries is considerably slower than on the web, finding that about 3% of the objects were no longer accessible after one year^[9] (equating to a half-life of nearly 23 years).

Causes

Link rot can result from several occurrences. A target web page may be removed. The server that hosts the target page could fail, be removed from service, or relocate to a new domain name. A domain name's registration may lapse or be transferred to another party. Some causes will result in the link failing to find any target and returning an error such as HTTP 404. Other causes will cause a link to target content other than what was intended by the link's author.

Other reasons for broken links include:

- the restructuring of websites that causes changes in URLs (e.g. `domain.net/pine_tree` might be moved to `domain.net/tree/pine`)
- relocation of formerly free content to behind a paywall
- a change in server architecture that results in code such as PHP functioning differently
- dynamic page content such as search results that changes by design
- the presence of user-specific information (such as a login name) within the link
- deliberate blocking by content filters or firewalls
- the removal of gTLDs^[10]

Prevention and detection

Strategies for preventing link rot can focus on placing content where its likelihood of persisting is higher, authoring links that are less likely to be broken, taking steps to preserve existing links, or repairing links whose targets have been relocated or removed.

The creation of URLs that will not change with time is the fundamental method of preventing link rot. Preventive planning has been championed by Tim Berners-Lee and other web pioneers.^[11]

Strategies pertaining to the authorship of links include:

- linking to primary rather than secondary sources and prioritizing stable sites
- avoiding links that point to resources on researchers' personal pages^[5]
- using clean URLs^[12] or otherwise employing URL normalization or URL canonicalization
- using permalinks and persistent identifiers such as ARKs, DOIs, Handle System references, and PURLs
- avoiding linking to documents other than web pages^[12]
- avoiding deep linking
- linking to web archives such as the Internet Archive,^[13] WebCite,^[14] Archive.is, Perma.cc,^[15] or Amber^[16]

Strategies pertaining to the protection of existing links include:

- using redirection mechanisms such as HTTP 301 to automatically refer browsers and crawlers to relocated content
- using content management systems which can automatically update links when content within the same site is relocated or automatically replace links with canonical URLs^[17]
- integrating search resources into HTTP 404 pages^[18]

The detection of broken links may be done manually or automatically. Automated methods include plug-ins for content management systems as well as standalone broken-link checkers such as like Xenu's Link Sleuth. Automatic checking may not detect links that return a soft 404 or links that return a 200 OK response but point to content that has changed.^[19]

See also

- Software rot
- Digital preservation
- Deletionism and inclusionism in Wikipedia

Further reading

- Markwell, John; Brooks, David W. (2002). "Broken Links: The Ephemeral Nature of Educational WWW Hyperlinks". *Journal of Science Education and Technology*. **11** (2): 105–108. doi:10.1023/A:1014627511641 (<https://doi.org/10.1023%2FA%3A1014627511641>).
- Gomes, Daniel; Silva, Mário J. (2006). "Modelling Information Persistence on the Web" (<https://web.archive.org/web/20110716190802/http://xldb.di.fc.ul.pt/daniel/docs/papers/gomes06urlPersistence.pdf>) (PDF). *Proceedings of the 6th International Conference on Web Engineering*. ICWE'06. Archived from the original (<http://xldb.di.fc.ul.pt/daniel/docs/papers/gomes06urlPersistence.pdf>) (PDF) on 2011-07-16. Retrieved 14 September 2010.
- Dellavalle, Robert P.; Hester, Eric J.; Heilig, Lauren F.; Drake, Amanda L.; Kuntzman, Jeff W.; Graber, Marla; Schilling, Lisa M. (2003). "Going, Going, Gone: Lost Internet References" (<http://www.researchgate.net/publication/9030760>). *Science*. **302** (5646): 787–788. doi:10.1126/science.1088234 (<https://doi.org/10.1126%2Fscience.1088234>). PMID 14593153 (<https://pubmed.ncbi.nlm.nih.gov/14593153>).
- Koehler, Wallace (1999). "An Analysis of Web Page and Web Site Constancy and Permanence". *Journal of the American Society for Information Science*. **50** (2): 162–180. doi:10.1002/(SICI)1097-4571(1999)50:2<162::AID-ASI7>3.0.CO;2-B (<https://doi.org/10.1002%2F%28SICI%291097-4571%281999%2950%3A2%3C162%3A%3AAID-ASI7%3E3.0.CO%3B2-B>).
- Sellitto, Carmine (2005). "The impact of impermanent Web-located citations: A study of 123 scholarly conference publications" (http://vuir.vu.edu.au/336/1/The_impact_of_impermanent_Web.pdf) (PDF). *Journal of the American Society for Information Science and Technology*. **56** (7): 695–703. CiteSeerX 10.1.1.473.2732 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.473.2732>). doi:10.1002/asi.20159 (<https://doi.org/10.1002%2Fasi.20159>).

Notes & references

Notes

References

1. Fetterly, Dennis; Manasse, Mark; Najork, Marc; Wiener, Janet (2003). "A large-scale study of the evolution of web pages" (<http://www2003.org/cdrom/papers/refereed/p097/P97%20sources/p97-fetterly.html>). *Proceedings of the 12th international conference on World Wide Web*. Archived (<https://web.archive.org/web/20110709175020/http://www2003.org/cdrom/papers/refereed/p097/P97%20sources/p97-fetterly.html>) from the original on 9 July 2011. Retrieved 14 September 2010.

2. van der Graaf, Hans. "The half-life of a link is two year" (<http://blog.zomdir.com/2017/10/the-half-life-of-link-is-two-year.html>). *ZOMDir's blog*. Archived (<https://web.archive.org/web/20171017041901/http://blog.zomdir.com/2017/10/the-half-life-of-link-is-two-year.html>) from the original on 2017-10-17. Retrieved 2019-01-31.
3. Koehler, Wallace (2004). "A longitudinal study of web pages continued: a consideration of document persistence" (<http://www.informationr.net/ir/9-2/paper174.html>). *Information Research*. **9** (2). Archived (<https://web.archive.org/web/20170911062629/http://www.informationr.net/ir/9-2/paper174.html>) from the original on 2017-09-11. Retrieved 2019-01-31.
4. "All-Time Weblock Report" (<https://web.archive.org/web/20160304081204/https://weblock.io/report?id=all>). August 2015. Archived from the original (<https://weblock.io/report?id=all>) on 4 March 2016. Retrieved 12 January 2016.
5. McCown, Frank; Chan, Sheffan; Nelson, Michael L.; Bollen, Johan (2005). "The Availability and Persistence of Web References in D-Lib Magazine" (<https://web.archive.org/web/20120717000118/http://www.iwaw.net/05/papers/iwaw05-mccown1.pdf>) (PDF). *Proceedings of the 5th International Web Archiving Workshop and Digital Preservation (IWAW'05)*. Archived from the original (<http://www.iwaw.net/05/papers/iwaw05-mccown1.pdf>) (PDF) on 2012-07-17. Retrieved 2005-10-12.
6. Spinellis, Diomidis (2003). "The Decay and Failures of Web References" (<http://www.spinellis.gr/pubs/jrnl/2003-CACM-URLcite/html/urlcite.html>). *Communications of the ACM*. **46** (1): 71–77. CiteSeerX [10.1.1.12.9599](https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.9599) (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.9599>). doi:[10.1145/602421.602422](https://doi.org/10.1145/602421.602422) (<https://doi.org/10.1145/602421.602422>). Archived (<https://web.archive.org/web/20200723030709/https://www.spinellis.gr/pubs/jrnl/2003-CACM-URLcite/html/urlcite.html>) from the original on 2020-07-23. Retrieved 2007-09-29.
7. Steve Lawrence; David M. Pennock; Gary William Flake; et al. (March 2001). "Persistence of Web References in Scientific Research". *Computer*. **34** (3): 26–31. CiteSeerX [10.1.1.97.9695](https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.9695) (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.9695>). doi:[10.1109/2.901164](https://doi.org/10.1109/2.901164) (<https://doi.org/10.1109/2.901164>). ISSN [0018-9162](https://www.worldcat.org/issn/0018-9162) (<https://www.worldcat.org/issn/0018-9162>). Wikidata [Q21012586](https://www.wikidata.org/wiki/Q21012586).
8. Hennessey, Jason; Xijin Ge, Steven (2013). "A Cross Disciplinary Study of Link Decay and the Effectiveness of Mitigation Techniques" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3851533>). *BMC Bioinformatics*. **14**: S5. doi:[10.1186/1471-2105-14-S14-S5](https://doi.org/10.1186/1471-2105-14-S14-S5) (<https://doi.org/10.1186/1471-2105-14-S14-S5>). PMC [3851533](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3851533) (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3851533>). PMID [24266891](https://pubmed.ncbi.nlm.nih.gov/24266891) (<https://pubmed.ncbi.nlm.nih.gov/24266891>).
9. Nelson, Michael L.; Allen, B. Danette (2002). "Object Persistence and Availability in Digital Libraries" (https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1008&context=computer_science_fac_pubs). *D-Lib Magazine*. **8** (1). doi:[10.1045/january2002-nelson](https://doi.org/10.1045/january2002-nelson) (<https://doi.org/10.1045/january2002-nelson>). Archived (https://web.archive.org/web/20200719044311/https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1008&context=computer_science_fac_pubs) from the original on 2020-07-19. Retrieved 2019-09-24.
10. "The death of a TLD" (<https://blog.benjojo.co.uk/post/the-death-of-a-tld>). *blog.benjojo.co.uk*. Archived (<https://web.archive.org/web/20180726220205/https://blog.benjojo.co.uk/post/the-death-of-a-tld>) from the original on 2018-07-26. Retrieved 2018-07-27.
11. Berners-Lee, Tim (1998). "Cool URIs Don't Change" (<https://www.w3.org/Provider/Style/URI>). Archived (<https://web.archive.org/web/20000302064802/http://www.w3.org/Provider/Style/URI>) from the original on 2000-03-02. Retrieved 2019-01-31.
12. Kille, Leighton Walter (8 November 2014). "The Growing Problem of Internet "Link Rot" and Best Practices for Media and Online Publishers" (<http://journalistsresource.org/studies/society/internet/website-linking-best-practices-media-online-publishers>). Journalist's Resource, Harvard Kennedy School. Archived (<https://web.archive.org/web/20150112034707/http://journalistsresource.org/studies/society/internet/website-linking-best-practices-media-online-publishers>) from the original on 12 January 2015. Retrieved 16 January 2015.

13. "Internet Archive: Digital Library of Free Books, Movies, Music & Wayback Machine" (<https://archive.org/>). 2001-03-10. Archived (<https://web.archive.org/web/19970126045828/http://www.archive.org/>) from the original on 26 January 1997. Retrieved 7 October 2013.
14. Eysenbach, Gunther; Trudel, Mathieu (2005). "Going, going, still there: Using the WebCite service to permanently archive cited web pages" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550686>). *Journal of Medical Internet Research*. 7 (5): e60. doi:10.2196/jmir.7.5.e60 (<https://doi.org/10.2196%2Fjmir.7.5.e60>). PMC 1550686 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550686>). PMID 16403724 (<https://pubmed.ncbi.nlm.nih.gov/16403724>).
15. Zittrain, Jonathan; Albert, Kendra; Lessig, Lawrence (12 June 2014). "Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations" (https://cdn.harvardlawreview.org/wp-content/uploads/2014/03/forvol127_zittrain.pdf) (PDF). *Legal Information Management*. 14 (2): 88–99. doi:10.1017/S1472669614000255 (<https://doi.org/10.1017%2FS1472669614000255>). Archived (https://web.archive.org/web/20201101012831/https://cdn.harvardlawreview.org/wp-content/uploads/2014/03/forvol127_zittrain.pdf) (PDF) from the original on 1 November 2020. Retrieved 10 June 2020.
16. "Harvard University's Berkman Center Releases Amber, a "Mutual Aid" Tool for Bloggers & Website Owners to Help Keep the Web Available | Berkman Center" (<https://cyber.law.harvard.edu/node/99276>). *cyber.law.harvard.edu*. Archived (<https://web.archive.org/web/20160202042259/https://cyber.law.harvard.edu/node/99276>) from the original on 2016-02-02. Retrieved 2016-01-28.
17. Rønn-Jensen, Jesper (2007-10-05). "Software Eliminates User Errors And Linkrot" (<http://justaddwater.dk/2007/10/05/blog-software-eliminates-user-errors-and-linkrot/>). Justaddwater.dk. Archived (<https://web.archive.org/web/20071011033526/http://justaddwater.dk/2007/10/05/blog-software-eliminates-user-errors-and-linkrot/>) from the original on 11 October 2007. Retrieved 5 October 2007.
18. Mueller, John (2007-12-14). "FYI on Google Toolbar's Latest Features" (<http://googlewebmastercentral.blogspot.com/2007/12/fyi-on-google-toolbars-latest-features.html>). Google Webmaster Central Blog. Archived (<https://web.archive.org/web/20080913132848/http://googlewebmastercentral.blogspot.com/2007/12/fyi-on-google-toolbars-latest-features.html>) from the original on 13 September 2008. Retrieved 9 July 2008.
19. Bar-Yossef, Ziv; Broder, Andrei Z.; Kumar, Ravi; Tomkins, Andrew (2004). "Sic transit gloria telae: towards an understanding of the Web's decay". *Proceedings of the 13th international conference on World Wide Web – WWW '04*. pp. 328–337. CiteSeerX 10.1.1.1.9406 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1.9406>). doi:10.1145/988672.988716 (<https://doi.org/10.1145%2F988672.988716>). ISBN 978-1581138443.

External links

- Future-Proofing Your URIs (<https://web.archive.org/web/20061109071631/http://www.wrox.com/WileyCDA/Section/id-301495.html>)
 - Jakob Nielsen, "Fighting Linkrot" (<http://www.useit.com/alertbox/980614.html>), *Jakob Nielsen's Alertbox*, June 14, 1998.
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Link_rot&oldid=1015926088"

This page was last edited on 4 April 2021, at 11:04 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.