

Data links quick fix

February 13 2014

Software that can fix 90 percent of broken links in the web of data, assuming the resources are still on the site's server, has been developed by researchers in Iran. The details are reported this month in the *International Journal Web Engineering and Technology*.

Everyone knows the frustration of following a link to an interesting web site only to discover the target page is no longer there and to be presented with an error page. However, more frustrating and with wider implications for science, healthcare, industry and other areas is when machines communicate and expect to find specific resources that turn out to be missing or dislocated from their identifier. This can cause problems when a computer is processing large amounts of data in a financial or scientific analysis, for instance. If the resource is still on the servers, then it should be retrievable given a sufficiently effective [algorithm](#) that can recreate the missing links.

Computing engineers Mohammad Pourzaferani and Mohammad Ali Nematbakhsh of the University of Isfahan explain that previous efforts to address the issue of broken links in the web of data have focused on the destination point. This approach has two inherent limitations. First, it homes in on a single point of failure whereas there might be wider issues across a database. Secondly, it relies on knowledge of the destination data source.

The team has now introduced a method for fixing broken links which is based on the source point of links and a way to discover the new address of the digital entity that has become detached. Their method creates a

superior and an inferior dataset which lets them create an exclusive data graph that can be monitored over time in order to identify changes and trap missing links as resources become detached.

"The proposed algorithm uses the fact that entities preserve their structure event after movement to another location. Therefore, the algorithm creates an exclusive graph structure for each entity," explains Pourzaferani. This graph consists of two types of entity called 'Superior' and 'Inferior'. Which are entities point to the detached entity and point by it, respectively. When the broken link is detected the algorithm starts its task to find the new location for detached entity or the best similar candidate for it. To this end, the crawler controller module searches for the superiors of each entity in the inferior dataset, and vice versa. After some steps the search space is narrowed and the best candidate is chosen."

The researchers tested the algorithm on two snapshots of DBpedia within which are contained almost 300,000 person entities. Their algorithm identified almost 5,000 entities that changed between the first and second snapshot recorded some time later. The algorithm demonstrated its prowess in relocating 9 out of 10 of the broken links.

More information: "Repairing broken RDF links in the web of data" in *Int. J. Web Engineering and Technology*, 2013, 8, 395-411

Provided by Inderscience Publishers

Citation: Data links quick fix (2014, February 13) retrieved 11 May 2024 from <https://phys.org/news/2014-02-links-quick.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.