



# Diplo LinkChecker

## Installation Guide

Dan Booth  
03/02/2012

<http://www.diplo.co.uk/>

# Table of Contents

---

Introduction.....	3
Features.....	3
How It Works .....	3
Revision History .....	4
Installing using a package.....	5
Manual installation.....	6
Configuring .....	7

# Introduction

---

Diplo LinkChecker is a simple HTTP based link checker for Umbraco that adds a link checking facility to the Content section of your Umbraco installation. It has been tested against version 4.7 and up and is built against the NET 4.0 framework.

## Features

- LinkChecker will check both internal and external links
- LinkChecker checks the entire page contents - not just Rich Text areas
- LinkChecker will also check all assets, such as the paths to images, included CSS files, JavaScript etc.
- Links within each page are checked asynchronously for speed
- Once a link has been checked it won't be checked again if it occurs elsewhere in the site
- Pages are batched into manageable chunks (by default 5) so that users are not left waiting for ages for the entire site to be checked
- Quick link to edit any page with broken links in it
- Advanced options can be configured via an included XML .config file

## How It Works

LinkChecker works in the following manner:

- It generates a list of every page in your Umbraco site
- It then breaks these down into small batches (by default 5 pages)
- It then makes an HTTP request to each page in the batch in turn
- It then extracts all the links, images and included assets from the page and makes an HTTP HEAD request to each one in turn
- LinkChecker then examines the status code returned by the server to determine whether the link is OK (200), Not Found (404 etc.) or requires a Warning (other status codes)

## Screenshot

Get Started | Last Edits | Change Password | **Link Checker**

### Link Checker

Include in report links that are:  
 OK  Warning  Error  
 Show all pages checked (even if they have no links)

Finished checking 18 links in Pages 1 to 5. Found 3 links with errors or warnings. So far checked a total of 18 links in 5 pages.  
 Press the 'Continue Link Check' button above to check the next set of 5 pages.

Checked Page	Edit Link	Created	Last Edited	Parent Section	Links	Status
<a href="#">Home</a>	<a href="#">Edit</a>	31/01/2012 by <a href="#">admin</a>	31/01/2012 by admin	<a href="#">Home</a>	13	200 OK

The following links in 'Home' have been checked:

Status	Message	Checked Link	Link Text	Type	Internal?	
OK	200 OK	<a href="/css/all.css">/css/all.css</a>		link	Yes	<a href="#">Help</a>
OK	200 OK	<a href="#">/</a>		link	Yes	<a href="#">Help</a>
Error	404 Not Found	<a href="/styles/broken.css">/styles/broken.css</a>		link	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/about-us.aspx">/about-us.aspx</a>	About Us	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/login.aspx">/login.aspx</a>	Login	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/another-page.aspx">/another-page.aspx</a>	Another Page	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/test-page.aspx">/test-page.aspx</a>	Test Page	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/yet-another-page.aspx">/yet-another-page.aspx</a>	Yet Another Page	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/news.aspx">/news.aspx</a>	News	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/news/base-package-released.aspx">/news/base-package-released.aspx</a>	Base Package Released	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/news/nokia-bets-on-windows-7-phone.aspx">/news/nokia-bets-on-windows-7-phone.aspx</a>	Nokia Bets on Windows 7 Phone	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/news/life-sentences-to-be-mandatory.asp...">/news/life-sentences-to-be-mandatory.asp...</a>	Life Sentences to be Mandatory	hyperlink	Yes	<a href="#">Help</a>
OK	200 OK	<a href="/news.aspx?altTemplate=RssFeed">/news.aspx?altTemplate=RssFeed</a>	Subscribe to RSS Feed	hyperlink	Yes	<a href="#">Help</a>

Checked Page	Edit Link	Created	Last Edited	Parent Section	Links	Status
<a href="#">About Us</a>	<a href="#">Edit</a>	31/01/2012 by <a href="#">admin</a>	03/02/2012 by admin	<a href="#">Home</a>	1	200 OK

The following links in 'About Us' have been checked:

Status	Message	Checked Link	Link Text	Type	Internal?	
OK	200 OK	<a href="#">/</a>		link	Yes	<a href="#">Help</a>

## Links

Download Link: <http://our.umbraco.org/projects/backoffice-extensions/diplo-linkchecker>

Blog Post: <http://www.diplo.co.uk/blog/2012/2/3/diplo-linkchecker-for-umbraco.aspx>

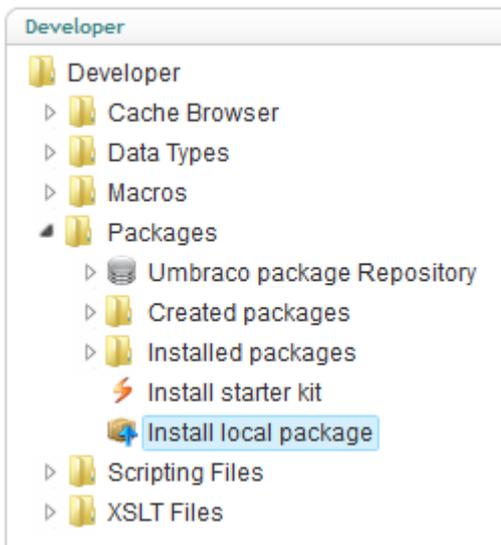
## Revision History

- Version 1, 3rd February 2011 - Initial release

# Installing using a package

---

1. Download the package to your local PC
2. Go to the Developer section of Umbraco and select Install Local Package from the Packages tree
3. Browse to the package file you downloaded and follow the instructions
4. Once installed go to the Content section of your website and you should see a new Link Checker tab
5. Get checkin' those naughty links!



# Manual installation

---

If for some reason the package installation fails or due to permissions or other reasons is not an option on your system, please follow this tedious manual installation process.

The manual process consists of several files:

- Copy the two DLL files to your Umbraco /bin folder
- Copy DiploLinkChecker.config to your /config folder
- Create a folder called Diplo within your /usercontrols folder
- Copy LinkChecker.ascx into the above Diplo folder
- Edit Dashboard.config in your /config folder to add a reference to the LinkChecker.ascx control:

```
<section alias="DiploLinkChecker">
  <areas>
    <area>content</area>
  </areas>
  <tab caption="Link Checker">
    <control>/usercontrols/Diplo/LinkChecker.ascx</control>
  </tab>
</section>
```

# Configuring

---

The LinkChecker doesn't need any configuring – it should work “out-of-the-box”.

However, you can tweak some of the options for the LinkChecker by editing `DiploLinkChecker.config` which is installed into your `/config` folder. By default this looks like:

```
<LinkChecker>

  <!-- Number of Umbraco Pages to Check At Once (larger numbers will slow things
  down) -->
  <pageSize>5</pageSize>

  <!-- What part of the page to check - can be EntireDocument, Body or Head -->
  <documentSection>EntireDocument</documentSection>

  <!-- Can be HEAD, GET or POST (HEAD is quickest but some servers don't respect
  it (eg. Wikipedia). Use GET for those.) -->
  <checkLinksRequestMethod>HEAD</checkLinksRequestMethod>

  <!-- Whether to follow redirects or not -->
  <allowRedirects>True</allowRedirects>

  <!-- Timeout in milliseconds before deciding if a link is not found.
  Lower numbers will speed things up but may cause false-negatives with slow
  servers. -->
  <timeOutMiliseconds>20000</timeOutMiliseconds>

</LinkChecker>
```

The options you can change are:

<b>pageSize</b>	Integer	The number of pages to check at once. The larger you make this the longer the link checker will be running for. Make it too large and you may encounter time-out problems.
<b>documentSection</b>	EntireDocument, Body, Head	Select whether you want to check links in the entire page or just the body section or just the head.
<b>checkLinksRequestMethod</b>	HEAD, GET, POST	Select the method by which the link checker makes an HTTP request. Usually HEAD is quicker (as it doesn't require downloading the entire page) but some servers, such as Wikipedia, don't accept this method. In those cases use GET. See <a href="http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html">http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html</a>
<b>allowRedirects</b>	True, False	Whether the link checker follows links that are redirects.
<b>timeOutMiliseconds</b>	Integer	The amount of time the link checker waits before determining that the server is not

		responding and therefore there is a problem with the link. Making this smaller might improve speed, but can cause "false positives" with slow web-servers.
--	--	--