

SortSite Technical Overview White Paper

SortSite saves weeks of manual testing by scanning web sites for quality issues.

Background

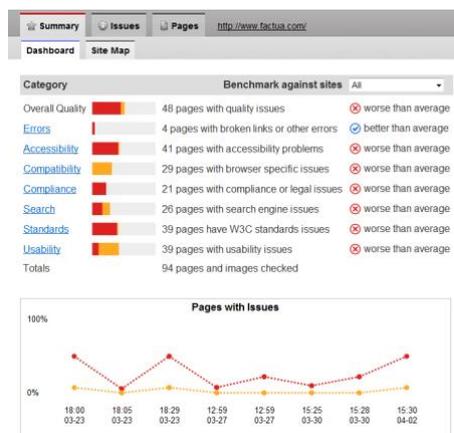
Finding quality issues on web sites is difficult and time-consuming. Most sites have evolved over time, and are changing constantly.

For large sites manual testing is very time-consuming and error prone, so organizations are increasingly turning to automated tools.

Introduction to SortSite

SortSite is a one-click site testing tool that runs hundreds of quality checkpoints on each page of a web site.

It produces easy to read reports that drill down from a high level dashboard to the exact location of each problem.

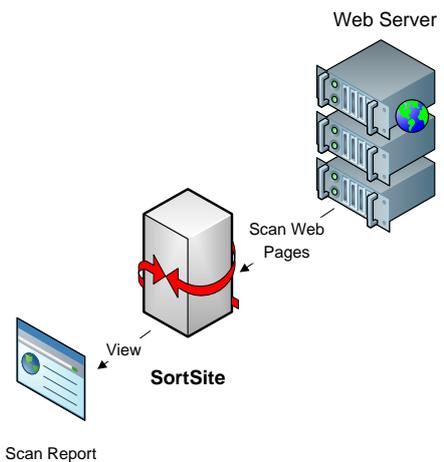


Example SortSite Report

How SortSite Works

SortSite scans each page of a site by running checkpoints, and then examines the page for links to un-scanned pages. It scans the newly found pages and repeats this

process until all pages have been scanned.



Pages are requested from the web server in the same way that a web browser requests them. This means the product works with password protected pages as well as server technologies like ASP.NET, JSP and ColdFusion.

This process is called web crawling, and is used by search engine agents like Googlebot to find pages to index. Like other well-behaved web crawlers, SortSite obeys the Robots Exclusion Protocol (robots.txt) which describes areas of sites that are off limits to web crawlers.

In addition, you can add URL patterns to ignore:

- *.pdf – to ignore all PDF files
- *logout* - to ignore all URLs containing 'logout'

Core Technology

The engine at the core of SortSite, and its sister product

SortSite

one-click site testing

PowerMapper, has been in production use for over a decade.

Continuous development based on feedback from a large customer base over these timescales means the product is very robust. Low risk product feature suggestions are incorporated into quarterly maintenance releases. More complex feature suggestions are scheduled for the next major release.

Product Editions

SortSite is available as traditional desktop software, or a web-based

pay-as-you-go subscription. The product editions are:

- SortSite Desktop Standard Windows and macOS
- SortSite Desktop Professional Windows and macOS
- SortSite OnDemand Web-based subscription (SaaS)

Product Features

Features Overview

- Automated scanning of entire websites, website folders and individual pages
- Scheduled scans
- Very easy to use
- Over 1300 comprehensive checks for:
 - Broken links, server configuration errors and spell checking (English and French)
 - Accessibility (WCAG 2.0 AAA, WCAG 2.1 AAA, and Section 508)
 - Browser Compatibility (Chrome, Safari, Firefox, Edge, Internet Explorer, Opera and iPhone/iPad)
 - Search Optimization (Google and Bing Webmaster Guidelines, Robots.txt, Search Best Practices)
 - Web Standards (HTML, XHTML and CSS validation)
 - Usability (Usability.gov Guidelines, W3C Best Practices)
- User-friendly reports with:
 - Dashboard showing issue counts, trends and benchmarks
 - Drilldown issue reports
 - Source View showing web page code annotated with issues and how-to-fix help
 - Inventory of site content
 - Site map showing error counts

Checkpoints

Each page is scanned for over 1300 quality checkpoints.

Each page is checked for quality issues using over 1300 checkpoints. Checkpoints are grouped into guidelines (e.g. Google Webmaster Guidelines or Section 508 Accessibility Guidelines) which can be enabled or disabled as a group. Individual checkpoints can also be enabled or disabled. This is often done when issues are triggered by code that can't be changed (e.g. third party components).

Accessibility

Check that disabled users can access your site using national and international accessibility standards:

- W3C WCAG 2.0 Levels A, AA, AAA (over 340 rules covering 39 checkpoints)
- W3C WCAG 2.1 Levels A, AA, AAA (over 350 rules covering 45 checkpoints)
- US Federal Section 508 (over 330 rules covering A and AA checkpoints)

Checkpoints include:

- Missing ALT attributes
- Poor text contrast
- Animated images and sounds that cannot be paused
- Controls that can only be actioned with a mouse

Spell Checking

Spell checking for:

- English - US
- English - UK
- English - Canada
- English - Australia
- French

Errors

Check for over 40 major functional problems including:

- Broken links and anchors in HTML, CSS, PDF and Office documents
- Missing or corrupt images

- Server configuration issues like inconsistent MIME types and character set encodings
- Script errors
- Domains about to expire
- SSL certificate problems
- Placeholder text like "Lorem Ipsum" and "TODO"
- Empty pages
- Page code and content larger than a specified size (page weight)
- Pages containing specified text (e.g. inappropriate language, competitor brand names)
- Pages missing required text or code (e.g. disclaimers or analytics tags)

Browser Compatibility

No major browser provides a full implementation of W3C standards. Check your HTML, CSS, JavaScript and images work with major browsers:

- Internet Explorer
- Edge
- Firefox
- Safari
- Opera
- Google Chrome

Check for over 250 issues including:

- Use of HTML elements and attributes not supported by all browsers
- Use of CSS features not supported by all browsers (Internet Explorer in particular has patchy CSS support)

- JavaScript and DOM properties and methods that trigger bugs or aren't implemented (e.g. in Firefox window.opener.closed returns random values after a window is closed)
- Images that don't display in some browsers (e.g. PNG files that are 4097 or 4098 bytes in length don't display in IE6 and BMP files only display in Windows browsers)
- Website features that work in IE6, but don't work in modern browsers

Search Optimization

Check against search engine guidelines and best practices:

- Google Search Guidelines
- Yahoo Search Guidelines
- Bing Search Guidelines
- Best Practice Keyword Optimization

Check for over 40 issues including:

- Duplicate and missing page titles
- Dynamic pages with too many URL parameters to index
- Pages with hidden text which can trigger ranking penalties
- Presence of specified keywords in critical areas like page titles and headings

Web Standards

Validate code using the same integrated HTML validation engine as the W3C validation service:

- HTML, XHTML, CSS and SVG validation
- Find markup deprecated by W3C

Supported W3C standards include:

- HTML 2.0

- HTML 3.2
- HTML 4.0
- HTML 4.01
- HTML 5
- ISO HTML 15445:2000
- MathML 2.0
- MathML 3.0
- XHTML 1.0
- XHTML 1.1
- SVG 1.0
- SVG 1.1
- SVG Basic 1.1
- SVG Tiny 1.1
- SMIL 1.0
- SMIL 2.0
- SMIL 2.1
- SMIL 3.0

Usability

Check against usability standards:

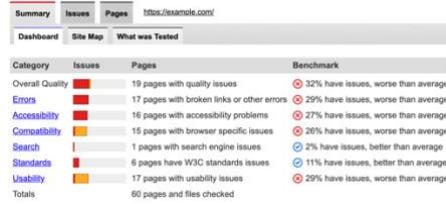
- Usability.gov guidelines
- W3C best practices
- Readability guidelines

Check for over 35 user experience problems including:

- Slow loading pages
- Use of Windows or Mac specific fonts
- Pages with no navigation options
- Pages requiring university level reading skills
- Displaying unsolicited popup windows
- Pages with default titles like "Untitled-1"
- Using underlined text which looks like a link
- Forms that require shifting from keyboard to mouse to complete
- Long URLs which wrap onto two lines when emailed
- Images that cause the page to jump about while loading

Reports

Dashboard



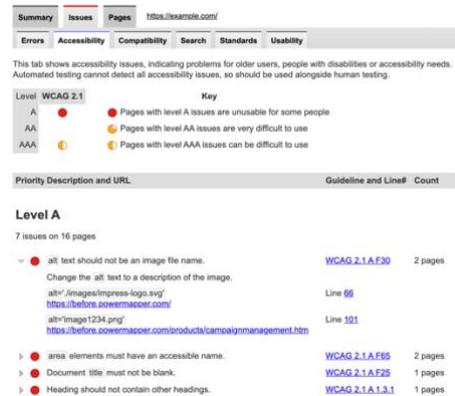
Overview of site issues; issue count over time; and benchmarks against similar sites.

Issues are divided into categories: errors; accessibility; browser compatibility; legal compliance; search guidelines; standards and usability. Clicking on a category title drills down to the Issues by Category report.

Issues by Category

This report shows a list of issues for a single category (e.g. accessibility) with a summary of standards compliance status.

Each issue shows a list of affected pages. Clicking on the URL drills down to the View Source report.



SortSite

one-click site testing

Source View

This view displays the source code of the file (HTML, CSS or JavaScript) with issues and errors shown alongside the line of code that caused them, and help describing how to fix the problem.

This view also shows a summary of page attributes: MIME type; last modified date; download time; size; list of links in and out; list of images, script files and CSS files.

● Ensure web pages print well. No Print links or CSS print styles were found on the site.
Line 1 Rule Options ▾

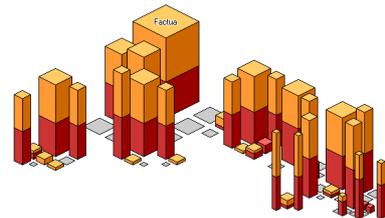
```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 1
<html lang="en"><!-- InstanceBegin template="
<head>
<LINK HREF="electrum.css" REL="styles!
<!-- #BeginEditable "DOCUMENTITITLE" --
<TITLE>Factua - One Click Factsheet Pr
```

● Meta descriptions should be less than 170 characters.
Line 7 Rule Options ▾

```
<META name="description" content="Factua is a
and mutual funds. It c
<META name="keywords" content="automat
generating fund factsheets, ge
fact sheet printing, factsheet
<!-- #EndEditable --><!-- InstancePar
```

Site Map

The bars shows error counts for each page, allowing you to locate low-quality areas of the site.



Inventory

List of pages and content on site broken down by type and technology.

Clicking on the URLs displays the view source report.

| Summary | Issues | Pages | https://example.com/ |
|---|-----------|------------------------|----------------------|
| Inventory Issues by Page | | | |
| This tab shows an inventory of all pages, images and resources. | | | |
| Type | URL | Description | |
| Documents | | | |
| > HTML | 17 pages | HTML pages | |
| > PDF | 1 files | Adobe Acrobat PDFs | |
| > Other | 0 files | Other documents | |
| Images | | | |
| > GIF | 20 images | GIF images | |
| > PNG | 5 images | PNG images | |
| > JPEG | 0 images | JPEG images | |
| > SVG | 1 images | SVG images | |
| > Other | 0 images | Other image types | |
| Misc | | | |
| > CSS | 5 files | Cascading Style Sheets | |
| > JavaScript | 0 files | JavaScript files | |
| > Fonts | 0 files | Web fonts | |

Issues by Page

List of issues broken down by page. Each issue shows a list of affected pages. Clicking on the URL drills down to the View Source report.

| Issues | URL / Description |
|--------|---|
| 30 | http://www.factua.com/ |
| 26 | http://www.factua.com/products/index.htm |
| 28 | http://www.factua.com/products/ee/ |
| 29 | http://www.factua.com/products/ms/ |
| 3 | http://www.factua.com/products/datasheet.htm |
| 30 | http://www.factua.com/products/demo.htm |
| 2 | http://www.factua.com/factsheets/rock.pdf |
| | ● Use the LANG attribute to identify the language of the page. |
| | ● PDFs must be tagged to be accessible by screen readers. |

About PowerMapper Software

PowerMapper Software is a UK-based software company specializing in web site analysis products. The company's first product, PowerMapper, was released in 1997.

PowerMapper Software products are used by over 30% of the Fortune 100, and customers include Boeing, NASA, Bank of America, Dell and Microsoft.