

SortSite

one-click site testing

SortSite Technical Overview White Paper

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

Background

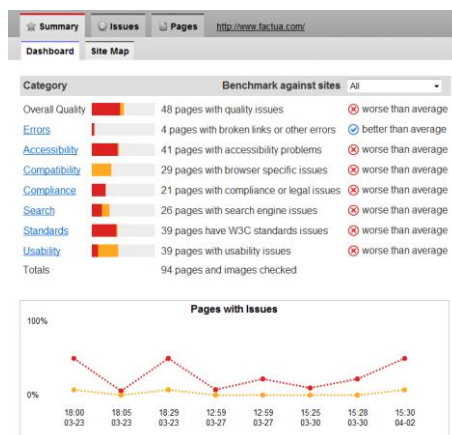
Finding quality issues on web sites is difficult and time-consuming. Most sites have evolved over time, and are changing constantly.

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.

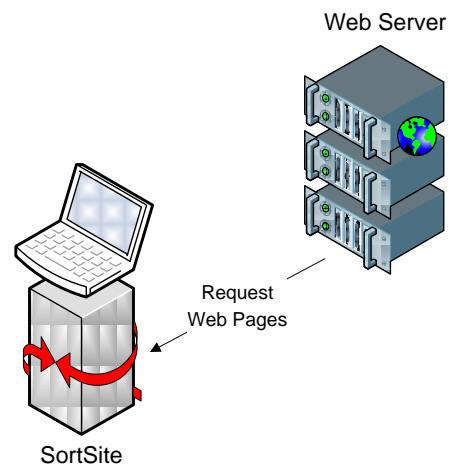


Example SortSite Report

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, ColdFusion and SharePoint.

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'

SortSite

one-click site testing

Core Technology

The engine at the core of SortSite, and its sister product 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 regular maintenance releases every 4-8 weeks. More complex feature suggestions are scheduled for the next major release.

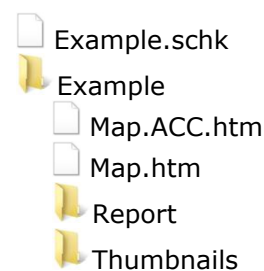
Backup and Restore

SortSite stores site structure and issues in a binary .schk file, with generated reports stored in an associated folder with the same

name as the .schk file (minus extension).

Files are only held open while reads and writes are in progress so saved data can be archived and restored using any backup solution.

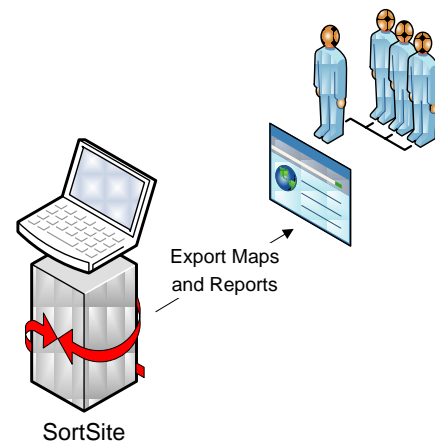
Saving a file called "Example" produces the following files and folders:



Product Features

Features Overview

- Dashboard showing issue counts, trends and benchmarks
- Drilldown issue reports
- Inventory of site content
- Site map showing error counts
- Standards based checkpoints for accessibility, browser compatibility, legal compliance, broken links, search engine guidelines, usability
- Integrated HTML / XHTML validator



Interoperability

The Export command allows map structure and metadata to be exported to a variety of output formats for sharing with co-workers or processing in applications like Excel and SQL Server.

The following export formats are supported:

Full Report	HTML Word RTF
Link Report - Excel Issue Report - Excel	Excel XML

Checkpoints

Each page is scanned for over 450 quality checkpoints.

Each page is checked for quality issues using over 450 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 1 Levels A, AA, AAA (74 checkpoints)
- W3C WCAG 2 Levels A, AA, AAA (106 checkpoints)
- US Federal Section 508 (34 checkpoints)

Checkpoints include:

- Missing ALT tags
- Links only accessible via Flash
- Poor text contrast
- Animated images and sounds that cannot be paused
- Elements that only work with a mouse

Spell Checking

Spell checking for:

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

Errors

Check for 32 major functional problems including:

- Broken links and anchors in HTML, CSS, Flash, 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
- Firefox
- Safari
- Opera
- Google Chrome

Check for over 100 issues including:

- Use of HTML tags and attributes not supported by all browsers
- Use of CSS rules 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

Legal Compliance

Check for legal problems including:

- Use of no-commercial-use code on commercial sites
- Use of cookies or tracking images without privacy statements
- Missing copyright statements

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 60 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
- 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

(HTML 5 will be supported when the standard is finalized)

Usability

Check against usability standards:

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

Check for over 50 user experience problems including:

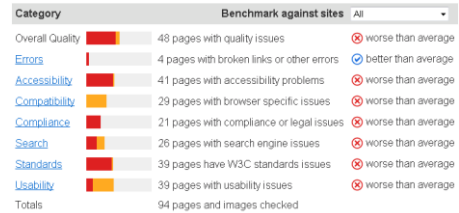
- Slow loading pages
- Use of Windows or Mac specific fonts
- Pages with no navigation options
- Pages requiring university level reading skills
- Displaying unsolicited popup windows
- Pages with default titles like "Untitled-1"
- Using underlined text which looks like a link
- Forms that require shifting 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.

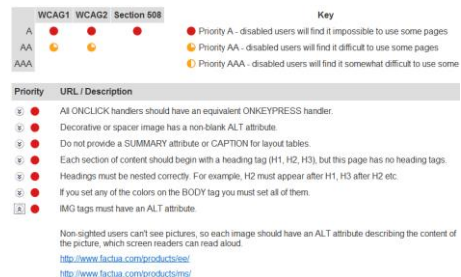
Issues are divided into categories: errors; accessibility; browser compatibility; legal compliance; search guidelines; standards and usability. Clicking on a category title drills down to the Issues by Category report.



Issues by Category

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

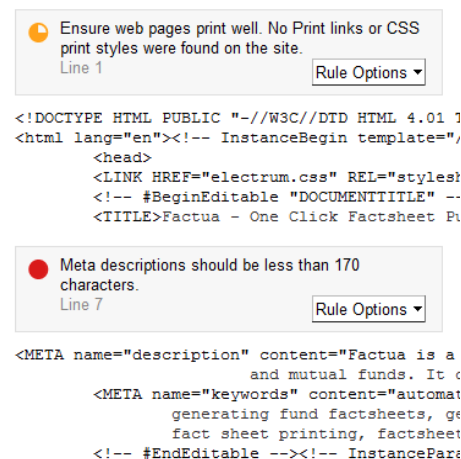
Each issue shows a list of affected pages. Clicking on the URL drills down to the View Source report.



View Source

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.

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.

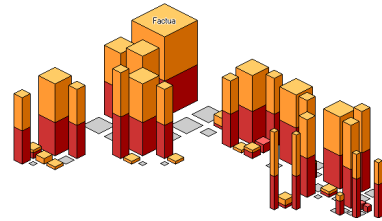


SortSite

one-click site testing

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.

Pages	URLs	Resource Type
HTML	45 pages	HTML pages
ASP	0 pages	Classic ASP dynamic pages
ASPX	0 pages	ASP.NET dynamic pages
PHP	0 pages	PHP dynamic pages
Ruby	0 pages	Ruby dynamic pages
Other	0 pages	Other dynamic pages

Images	URLs	Resource Type
GIF	33 images	GIF images
PNG	0 images	PNG images
JPEG	3 images	JPEG images
image/jpeg	http://www.factua.com/factsheets/images/euro.jpg	
image/jpeg	http://www.factua.com/factsheets/images/image001.jpg	
image/jpeg	http://www.factua.com/images/map.jpg	
Other	0 images	Other image types

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.

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 25% of the Fortune 100, and customers include Boeing, NASA, Bank of America, Dell and Microsoft.