SortSite
one-click site testing

SortSite 5
User Manual

one-click site testing

Overview

[Pro] indicates features only available in SortSite Professional Edition.

Introduction to SortSite

SortSite is a one-click web 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 then 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.

Web Server

Request
Web Pages
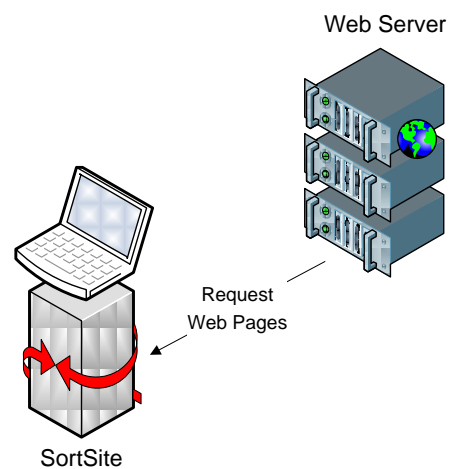
SortSite

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'

For more information see Advanced Options: Blocks below.

SortSite
one-click site testing

Checkpoints

Each page is checked for quality issues using over 700 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).

Errors

Check for over 40 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 (e.g. disclaimers or analytics tags)

Spell Checker

Check for spelling mistakes in English and French.

Accessibility

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

- W3C WCAG 1 Levels A, AA, AAA
- W3C WCAG 2 Levels A, AA, AAA

- US Federal Section 508

Checkpoints include:

- Missing and placeholder ALT tags
- Links only accessible via Flash movies
- Poor text contrast
- Animated images and sounds that cannot be paused
- Pages without structural markup
- Elements that behave unexpectedly when the keyboard is used for navigation

Browser Compatibility

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

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

Check for over 150 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

SortSite

one-click site testing

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)

Privacy [Pro]

Check for legal problems including:

- Use of cookies or tracking images without privacy statements

Search Optimization [Pro]

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
- HTML 5.0
- 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 [Pro]

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

This document is for informational purposes only. PowerMapper Software makes no warranties, express or implied in this document.

Copyright 1996-2015 PowerMapper Software

SortSite
one-click site testing

Scanning Sites, Folders and Pages

To scan a site:

1) Navigate to the site using the built-in web browser
2) Click the Check button

The software will scan each page in turn and display a report once it's done.
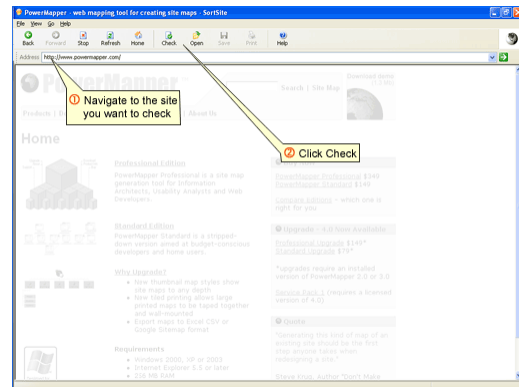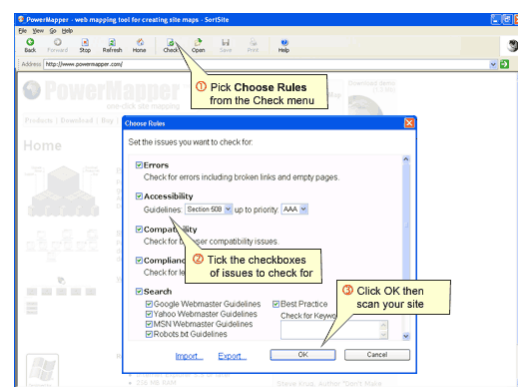
Scanning Options

Clicking the down arrow to the right of the Check button displays a menu that lets you:

- Check Entire Site – the default action for the check button
- Check Folder – scan pages in the same folder (starting the scan at www.example.com/folder/page.htm scans pages with addresses starting www.example.com/folder)
- Check Page – check a single page only (links are not checked)
- Check Page and Links – check a single page and outgoing links
- Check… – allows you to scan sites, folders and pages as well as specify additional pages to scan (such as landing pages)
- Recheck Current Page – use when the Source View is displayed to re-run checks on the page (useful when retesting pages after fixing issues)
- Choose Rules – chooses rules to run and other rule options

Choosing Rules

To choose which rules to run:

1) Pick the Choose Rules option from the Check menu.
2) Untick the checkboxes for any rules you want to omit from the scan.
3) Click OK.

This document is for informational purposes only. PowerMapper Software makes no warranties, express or implied in this document.

Copyright 1996-2015 PowerMapper Software

Sharing Rule Settings

The Choose Rules window allows you to import and export rules, allowing you to share common settings with other team members. This allows a team lead to define a corporate standard, then distribute to other people on the team.

Mapping Password Protected Pages

To map a password protected (authenticated) site:

1. Make sure Automatic Login is checked in the Options window (click the Options command on the View menu and go to the Links tab)
2. Login to the site in the usual way (either via a login page or the popup password dialog).
3. Once you've logged in navigate back to your home page and scan your site.

Note: If you have a logout link inside the secure area, then the scanner may be logged out if it follows the logout link. To prevent this happening go to Block Links (click the Options command on the View menu and go to the Links tab) and add something like the following:

```
*logout*
```

This stops the scanner following links with "logout" in the URL. You will need to adjust this to match the logout links (if any) used on your site.

Using Reports

Reports are broken down into a number of tabs and sub-tabs:

- Summary
  - Dashboard   overview of site issues
  - Site Map   shows distribution of issues across site
  - What was Tested   summary of what was tested
- Issues
  - Errors   technical errors such as broken links
  - Accessibility   WCAG and Section 508 accessibility
  - Compatibility   browser compatibility issues
  - Compliance [Pro]   legal compliance problems
  - Search [Pro]   search engine guidelines & best practice
  - Standards   W3 validation and technical standards
  - Usability [Pro]   usability issues
- Pages
  - Inventory   inventory of pages, images, scripts etc.
  - Issues by Page   issues grouped by page

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.
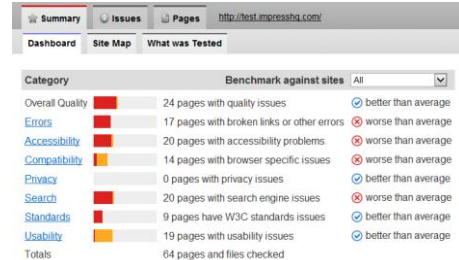
Site Map

The bars shows error counts for each page, allowing you to locate low-quality areas of the site. Clicking on a bar shows the Source View for the associated page.

SortSite
one-click site testing
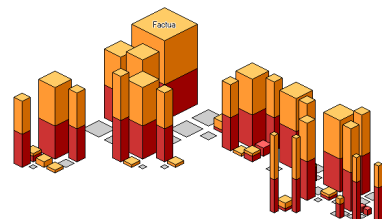
Issues by Category

The issues tabs show a list of issues for a single category (e.g. accessibility) with a summary of standards compliance status.

Clicking the ⚥ expansion button next to each issue shows URLs of affected pages. Clicking a URL displays the Source View for the URL.

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.

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.

Note: Line numbers refer to the line in the HTML page served by the web server, which may not correspond to line numbers used in server-side scripts like PHP, ASP.NET and JSP.

Inventory

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

Clicking the ⚥ expansion button next to each file type shows a list of URLs of that type.

Clicking on URLs in the inventory displays the Source View for each URL.

This document is for informational purposes only. PowerMapper Software makes no warranties, express or implied in this document.

Copyright 1996-2015 PowerMapper Software

Issues by Page

List of issues broken down by page. Each issue shows a list of affected pages. Clicking the ⊗ expansion button next to each URL shows issues for that page. Clicking a URL displays the Source View for the URL.

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.

Exporting Reports [Pro]

Reports can be exported using the Export command on the File menu. Supported formats include:

- HTML
- Word RTF
- Excel XML

SortSite

one-click site testing

Advanced Options

Application options are changed using the Options command on the View menu.

Reports

Allows you to specify settings control how many issues are reported.

Maximum pages shown per issue
The number of pages an issue is reported on in detail. This reduces noise from issues in page templates that affect every page of a site.

Maximum line numbers shown per issue
The number of lines per page an issue is reported on. This reduces noise from issues like unescaped '&' characters that can occur on nearly every line of every page of a site.

Maximum link limit for running rules
The maximum number of links per page that allow rules to run. Pages with a very large number of links can cause the rules engine to consume large amounts of memory.

Checking links to other sites
Either unlimited, or the The number of outbound links checked per external site. Checking lots of links to social book-marking sites can result in your book-marking account being suspended temporarily.
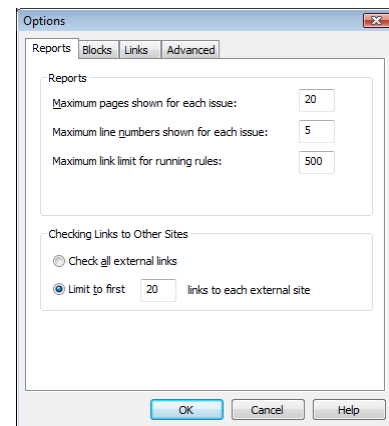
Blocks

Allows you to prevent pages being visited during the scan.

Obey Robots.txt
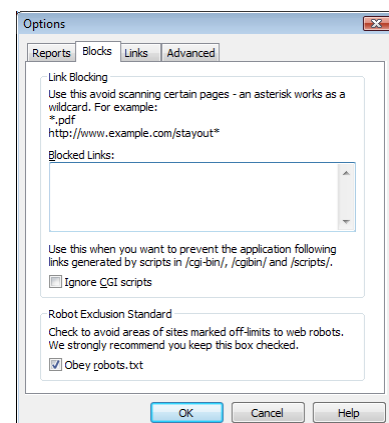When checked the web crawler will obey this protocol.

Note: *We strongly recommend you keep this checked.*

Ignore CGI scripts
Check to ignore links generated by scripts in the /cgi-bin/, /cgibin/ and /scripts/ directories.

Blocked Links
Allows you to specify URL patterns for pages to block. The syntax is the same as used in robots.txt. Examples:

`http://www.domain.com/copyright.htm`	Blocks a single page
`http://www.domain.com/stayout/*`	Blocks all pages in the stayout folder
`https:*`	Blocks all HTTPS links
`*.doc`	Blocks links to all pages with a .doc extension
`*print_friendly.htm`	Blocks links to URLs ending in print_friendly.htm
`*action=edit*`	Blocks links containing action=edit

Links

Allows you to control which links are followed during the scan.

Java Parameters
Lets you choose whether links in <PARAM> tags are followed. If checked any parameters containing "url", "href" or "link" in their names are treated as URLs.

JavaScript and Flash Links
Lets you choose whether links in JavaScript and Flash are followed. For example,

```
function GoTo()
{
    // the URL "bigframe1.htm" is
    // extracted from this JavaScript code
    parent.bigframe.location.href = "bigframe1.htm";
}
```

Follow links to related domains
When checked the scanner will follow links to child domains (e.g. if you scan domain.com then links to store.domain.com and www.domain.com will be followed) and sibling domains at the same level (e.g. if you scan www.domain.com then links to secure.domain.com and www2.domain.com will be followed).

Automatic Login
Check this box if you need to scan password protected sites.

Follow links to additional domains
Links to any additional domains listed in this box will be followed.
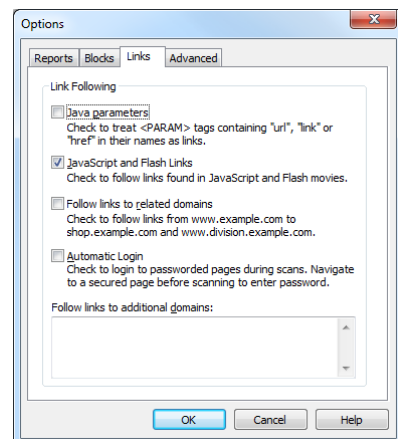
SortSite
one-click site testing

Advanced

Server Load
Lets you set a delay between loading pages to avoid placing undue load on a server.

Page Timeout
Controls the maximum time spent loading a page - the default setting is 1 minute.

Exploration Depth
Controls how deep the application will go into a web site. The default setting is your entire site, but you can restrict this to the top level pages of your site.

User Agent
Controls the user agent string sent to the web server (which is often recorded in web server logs). The default setting identifies the request as coming from a desktop browser. You can also choose a SortSite user agent string, or supply a custom user agent string. You sometimes need to supply a mobile browser user agent string to test some mobile sites (because they redirect desktop browsers onto the main desktop site).

Factory Reset
This resets all settings back to the initial factory settings.
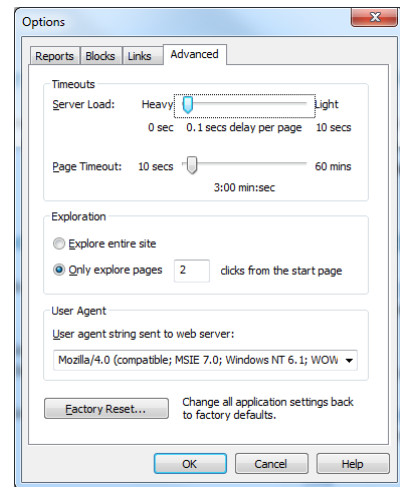
Recording Form Input

SortSite Professional can record form actions, for replay during a scan. For example, you can record text typed into your site's search box, and have this search replayed during subsequent scans, so that search results pages can be checked.

To record a form action:

1) Navigate to the page containing the form you want to record. For example, www.example.com/search
2) Select Record Form Input from the Check menu to start recording.
3) Type your test values into your form and submit it. This will display a new page with the results of the form submission (e.g. a page of search results). This new page will now be scanned each time the site is scanned.
4) Select Record Form Input again to stop recording. The recorded form action will be re-submitted each time www.example.com/search is scanned, and the results of form submission checked for problems.

Technical details: the Record Form Input command intercepts form GETS and POSTs and records (in encrypted form) the form fields in the post. Each time the page containing the form is visited during future scans, the form fields are re-submitted and the HTML returned by the GET or POST checked for problems.

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:

Example.schk
Example
  Map.ACC.htm
  Map.htm
  Report
  Res