SortSite

one-click site testing

SortSite Enterprise
User Manual

SortSite

one-click site testing

Overview

This manual describes how to use the SortSite Enterprise web application.

Introduction to SortSite

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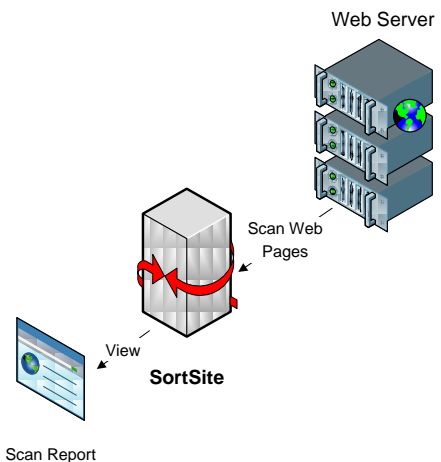
This process is called web crawling, and is used by search engine agents like Googlebot to find pages to index. Like other well-behaved web crawlers, SortSite obeys the Robots Exclusion Protocol (robots.txt) which describes areas of sites that are off limits to web crawlers.

In addition, you can add URL patterns to ignore:

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

For more information see the section "Blocking Pages" below.

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).

Each page is scanned for over 700 quality checkpoints.

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. See the Choosing Rules section on how to set the spelling language.

Accessibility

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

- W3C WCAG 2 Levels A, AA, AAA

- W3C WCAG 1 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 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

Check for privacy issues including:

- Use of cookies or tracking images without privacy 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
- 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

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

SortSite

one-click site testing

Scanning Sites, Folders and Pages

To scan a site:

1) Login to the web application (login link at top of every page on powermapper.com)
2) Type the address of the site you want to scan
3) Click the **Scan** button
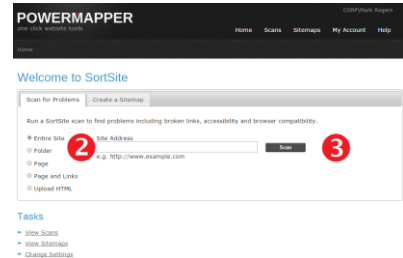
A scan request is sent to the SortSite Enterprise servers, which scan each page of the site in turn and display a report once it's done.

Scanning Options

The dropdown box to the left of the Scan button displays a menu that lets you choose the scope of the scan:

- Site – the entire site (apart from any blocked links)
- 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)
- Page – check a single page only (links are not checked)
- Page and Links – check a single page and broken links
- Upload HTML - upload an HTML file to scan

Choosing Rules

Once the a site has been scanned, you can change the checkpoints tested on each page:

1) Click the **Scans** link on the navigation bar at the top of the screen.
2) Click the **Edit Rules and Schedule** link next to the scan you want to modify.
3) Modify the rule settings then click the **Save** button at the bottom of the page.
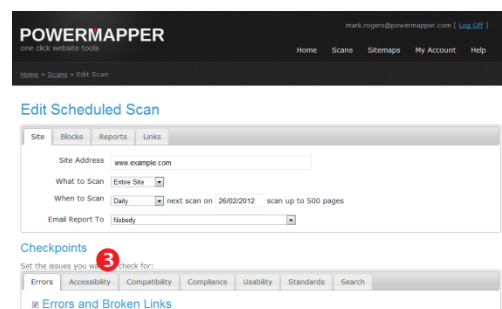
Note: changes saved here become the default settings for any new scans created (but don't affect existing scans, other than the one being edited).

Forms and Passwords

The Forms and Passwords tab on the Edit Scan page allows you to fill in a set of HTML form values to be replayed during scanning. Forms are often used to

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

provide login details, so you can also use this to login to sites that use forms authentication.

1) Choose the replay method use by the form being tested: GET or POST. You determine this by looking at the HTML code of the form being tested using either View Source or F12 Developer Tools in a browser. The following examples show how to choose the method:

   <form method="POST" action="login.aspx" /> <!—choose POST →
   <form method="GET" action="search.php" /> <!— choose GET →
   <form action=" search.php" /> <!— no method: choose GET →

2) Use the address (URL) of the page you viewed the source of in step 1 as the "from this page" setting.

3) Choose the action URL. The following examples show how to choose the action:

   <!— use http://google.com/search as the action URL →
   <form method="GET" action="http://google.com/search" />

   <!— use login.aspx as the action URL →
   <form method="POST" action="login.aspx" />

   <!— no action: leave the action URL blank →
   <form method="GET"/>

4) Choose the HTML field names to fill out before the form is submitted. You need to determine these by looking at the HTML code of the form page:

   <!— use "query" as the HTML field name →
   <form>
           <input type="text" name="query" />
           <input type="submit" name=" submit" value="Search" />
   </form>

   <!— use "txtUserName" and "txtPassword" as the HTML field names →
   <form>
           <input name="txtUserName" type="text" />
           <input name="txtPassword" type="password" />
           <button type="submit" name=" submit" >Login</button>
   </form>

Note: logins using HTTP basic authentication or Windows integrated authentication are not currently supported.

Scheduling and Emailing Reports

The Edit Scan page also allows you schedule regular scans (daily, weekly, monthly) and email reports to one of the registered users (see Managing Users below). The options are:

- Site Address – the URL of the site/folder/page being scanned
- What to Scan – site/folder/page (see Scanning Options above)
- When to Scan – when to run scheduled scans
- Email Report To – who to email scheduled scan results to

Report Visibility

You can control whether a scan reports and settings can be seen by other team members, or are only visible to you.

Blocking Pages

Adding entries to the Blocked Links box allows you to stop pages being visited during the scan.

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

Note: if the target site has a robots.txt file, then any pages blocked by the robots.txt file are always blocked (blocked links will be reported as an information message on the Errors tab of the report).

Setting Report Limits

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

Maximum pages shown per issue
The maximum number of pages each 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.

Checking links to other sites
Either unlimited, or the number of outbound links checked per external site. Useful if a site has social bookmarking links on every page (checking lots of these slows scans down).

Controlling Link Following

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

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).

Follow links to additional domains
Links to any additional domains listed in this box will be followed.

Deleting Scans

Old scan reports are automatically deleted after a fixed period to conserve disk space:

- Scheduled scans run monthly are deleted after 12 months.
- Scheduled scans run weekly are deleted after 12 weeks.
- Scheduled scans run daily are deleted after 12 days.
- Scans run via Quick Scan are deleted after 10 days.

If you want to manually delete a scan or scan history use the Delete Scan link. This gives the following options:

- Delete scan reports, except most recent one
  Deletes all but the most recent scan report for this url, but keeps scan schedule and settings
- Delete all scan reports
  Deletes all scan reports for this url, but keeps scan schedule and settings
- Delete url, scan settings and reports
  Deletes everything for this URL, and the URL no longer appears in the Scans list

SortSite

one-click site testing

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   legal compliance problems
  - Search   search engine guidelines & best practice
  - Standards   W3 validation and technical standards
  - Usability   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.
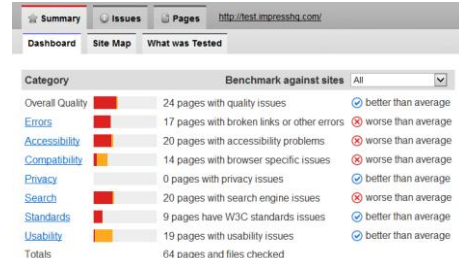
Site Map
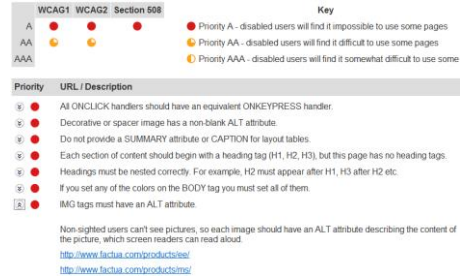
The bars shows error counts for each page, allowing you to locate low-quality areas of the site. Clicking on a bar shows the Source View for the associated page.

SortSite

one-click site testing
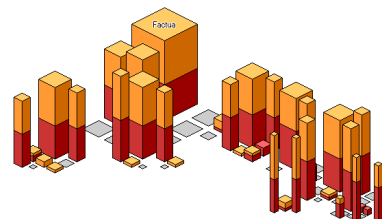
Issues by Category

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.
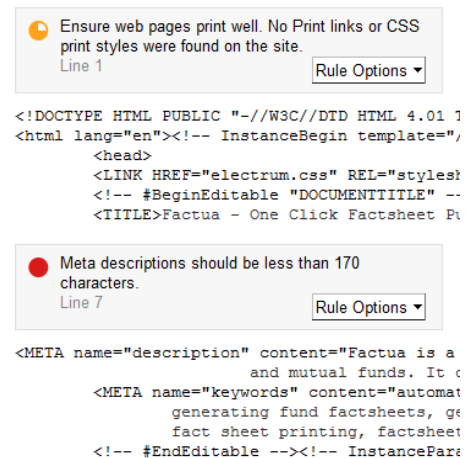
Source View

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.

SortSite
one-click site testing

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

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

- Full Report (Word RTF)
  RTF version of report that can be opened in any word processor

- Inventory Report (Excel CSV)
  List of pages / images / files found by scan

- Issue Report (Excel CSV)
  List of issues that can be opened by any spreadsheet

- Link Report (Excel CSV)
  List of broken links that can be opened by any spreadsheet

Account Management

Manage Billing

For credit card customers, shows the most recent invoice and provides a login link to the billing management system which provides:

- Cancellations, downgrades and upgrades
- Changing credit card details
- Charge history

Note: customers who have purchased through a reseller should contact their reseller for billing changes and queries.

Manage Users

Allows an administrator to manage users:

- Add new users (up to your plan maximum)
- Change email addresses and permissions
- Disable or delete users

Organization Settings

Allows an administrator to control the following settings:

- Organization name
- Installation serial number
- Whether robots.txt files are obeyed (see the section on "Blocking Pages" above)
- Email sender address (From address for password resets and emailed reports use)
- SMTP server name (the DNS name or IP address of your email server)
- SMTP server port (the port number of your email server)

My Settings

Allows the currently logged in user to set date and time formats displayed in the user interface.

Change Password

Allows the currently logged in user to change their password.

Password Reset

Available from the login page. Allows a user with an active account to securely reset their password by sending a reset link to their email address.