

Paper 331-2010

Integrating the Power of SAS® with the Ease of Search across the Enterprise

Craig Rubendall, SAS Institute Inc., Cary, NC

Chandra Prakash and Sudheer Prem, SAS Institute Inc., Pune, India

ABSTRACT

Enterprise search engines provide easy access to enterprise resources through search. Search engines unify various information sources into useful information based on relevance that empowers a company's employees, customers, and partners to take quick actions to meet evolving business needs.

Surfacing SAS content through enterprise search engines enables you to integrate the power of SAS with the ease of search across the enterprise.

We discuss the various approaches of surfacing SAS contents through the search facilities like Google Search Appliance, Microsoft Search Server (used by SharePoint), Teragram® Search Engine, and the search toolbar facility in browsers.

INTRODUCTION

Just as search engines provide an easy way to access information that is on the Web, enterprise search facilities provide an easy way to access the information within an enterprise. You can integrate SAS business intelligence (BI) content with the enterprise search facilities to increase the reach of BI content to a larger audience within the enterprise.

We describe the following approaches of integrating SAS content with enterprise search facilities:

1. crawl and index
2. feed integration
3. custom search provider

We provide an introduction to the SAS product Search Interface to SAS Content. We touch upon the security aspect of the various approaches. We also explore integrating the Search Interface to SAS Content product with the browser's search toolbar facility.

You can integrate SAS content such as dashboards, reports, stored processes, and information maps with the search engines.

Figure 1 shows a sample result of searching SAS contents through the Teragram Search Engine front-end.

We wrote a SAS Stored Process to generate Key Performance Indicators (KPIs), and then registered this stored process as a search provider with the Teragram Search Engine front-end. SAS contents are pushed to the Teragram index as feeds with categories included. Teragram applies categorization on the search results and provides faceted search.

SAS - Teragram Integrated Search
sas

Faceted Search

Related Information Maps

[sales_summary](#) (4)

[YearlySalesandProfitsMap](#) (3)

[PromoSales_PromoSummary](#) (3)

Related Chart Types

[Bar](#) (6)

[Plot](#) (5)

[Waterfall](#) (3)

[Pie](#) (2)

Related Metadata Types

[Report](#) (14)

[StoredProcess](#) (4)

Related Owners

[sasadm](#) (6)

[sastrust](#) (1)

Related Keywords

[NEWFILE](#) (2)

[PROC_GCHART](#) (2)

[GIF](#) (2)

Live KPIs created using SAS stored process

Q1

Target vs Actual Sales

Q2

Target vs Actual Sales

Q3

Target vs Actual Sales

Q4

Target vs Actual Sales

Search results coming from search engine's index in which the SAS contents are loaded as feeds

Found 21 matches

- [Yearly Sales by Customer Type.srx](#)
This Report shows the yearly Sales made by each customer.
Data Items: AnnualAmount, Customer Type Code, Customer Active Code, Averagepurchaseamount, Quarterlyamount etc.
09/17/2009 - <http://fdc1box.na.sas.com:8080/SASSearchService/Controller?forward=ReportProxy&openType=ReportI>
- [Line Item Sales by Country by Channel.srx](#)
This report shows the net and total line item sales amounts by country and by order channel
Data Items: Order Channel, Net Line Item Sales Amount, Total Line Item Sale Amount, Country Code, City Name.
09/17/2009 - <http://fdc1box.na.sas.com:8080/SASSearchService/Controller?forward=ReportProxy&openType=Reps>
- [Progressive Bar Chart showing Sales Progress.srx](#)
Progressive chart showing Year wise sales progress across different regions.
Data Items: Year Region Sales

Figure 1. SAS Content Searched through the Teragram Search Client

CRAWL AND INDEX APPROACH

“Crawl and Index” is the traditional approach used by search engines to make content searchable. The crawler scans available content sources as scheduled. There are multiple strategies to crawl the content sources. These strategies help to crawl optimally in a specified frequency. Once the content source is crawled, it will be indexed in the search engine’s repository. When the user searches for information, the search engine searches in this index and displays the results.

Documents in various formats are supported in this approach – for example, HTML pages, MS Office documents, and PDF documents.

You can leverage the crawler facility of search engines to support searching SAS content. This approach works particularly well for reports that do not change frequently, that take a long time to produce, or that need to be archived. Such reports cannot (or should not) be produced at the moment the viewer wants to see them, so they must be cached. The search engine crawls these locations, caches them in its index, and provides them as search results. The downside to this method is that because the report is a cached version, the search result might not be completely up-to-date. It is only as current as the last time the job was run and crawled. The deciding factor is how dynamic the back-end data is and how current the viewer expects the reports to be.

SURFACING SAS CONTENT TO SEARCH ENGINES VIA THE CRAWLER

The SAS Output Delivery System (ODS) makes new report formatting options available in SAS. These format options can be HTML, PDF, RTF, MS Excel, and so on. ODS separates your output into component parts so that the output can be sent to any ODS destination that you specify.

You can use ODS to produce various types of reports and publish them to a storage location. This storage location is registered with the search engine’s crawler. The search engine crawls the report files as per schedule. SAS results will be displayed along with other search results.

Figure 2 shows the process flow:

1. A batch process creates SAS reports using ODS in a variety of different formats. These formats can be images, MS office documents (for example, Word, Excel, PowerPoint, and so on), PDF, and so on.
2. These reports are copied to a storage location that is exposed to the search engine. This location can be inside the document root of a Web server or a network file share.
3. The search engine crawls these reports and updates its index.
4. The user searches for information by entering the query in a search client (browser, and so on). The search client submits the query via an HTTP GET request.
5. The search engine queries the search index for all possible results (based on the search engine's configuration).
6. The search engine applies preferred results, rules, and formatting based on the setup and query. The results are returned to the search client to be displayed to the user.

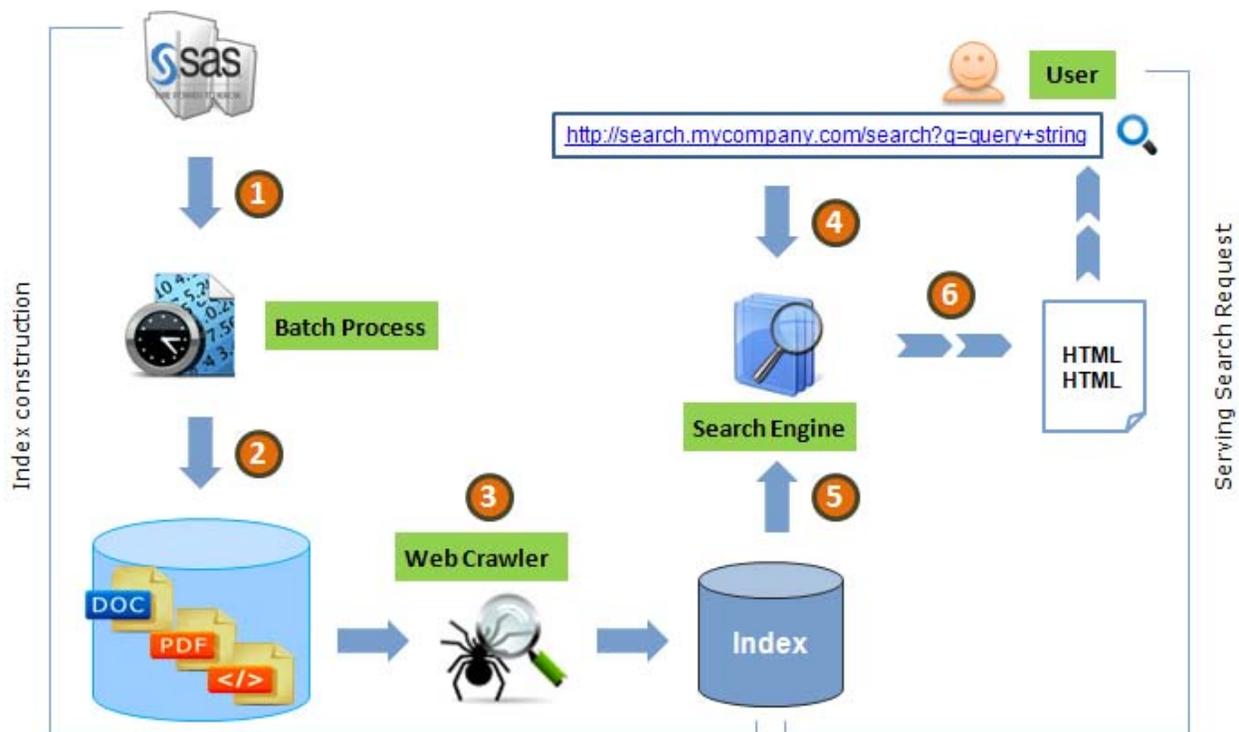


Figure 2. Surfacing SAS Content to Search Engines via a Crawler

FEED INTEGRATION APPROACH

Search engines allow feeding the documents as feeds for indexing. This approach is useful for content that requires special handling, content that might not be found through links on crawled Web pages, or content that is most useful when it is crawled at a specific time.

You can use feeds to push data and metadata into the index of the search engines. To push feeds to multiple search engines, you need to create feeds in different formats in the schema specified by each search engine.

SURFACING SAS CONTENT TO SEARCH ENGINES AS FEEDS

You can schedule batch programs to generate feeds to update the index of search engines. Figure 3 shows how the process works:

1. A timer or trigger-based application reads the SAS content and converts the content to XML in the schema specified by the search engine.

2. This feed XML is pushed to the search engine for indexing. The search engine updates its index with this feed.
3. The user searches for information by entering the query in a search client (browser, and so on). The search client submits the query via an HTTP GET request.
4. The search engine queries the search index for all possible results (based on the search engine's configuration).
5. The search engine applies preferred results, rules, and formatting based on the setup and query. The results are returned to the search client to be displayed to the user.

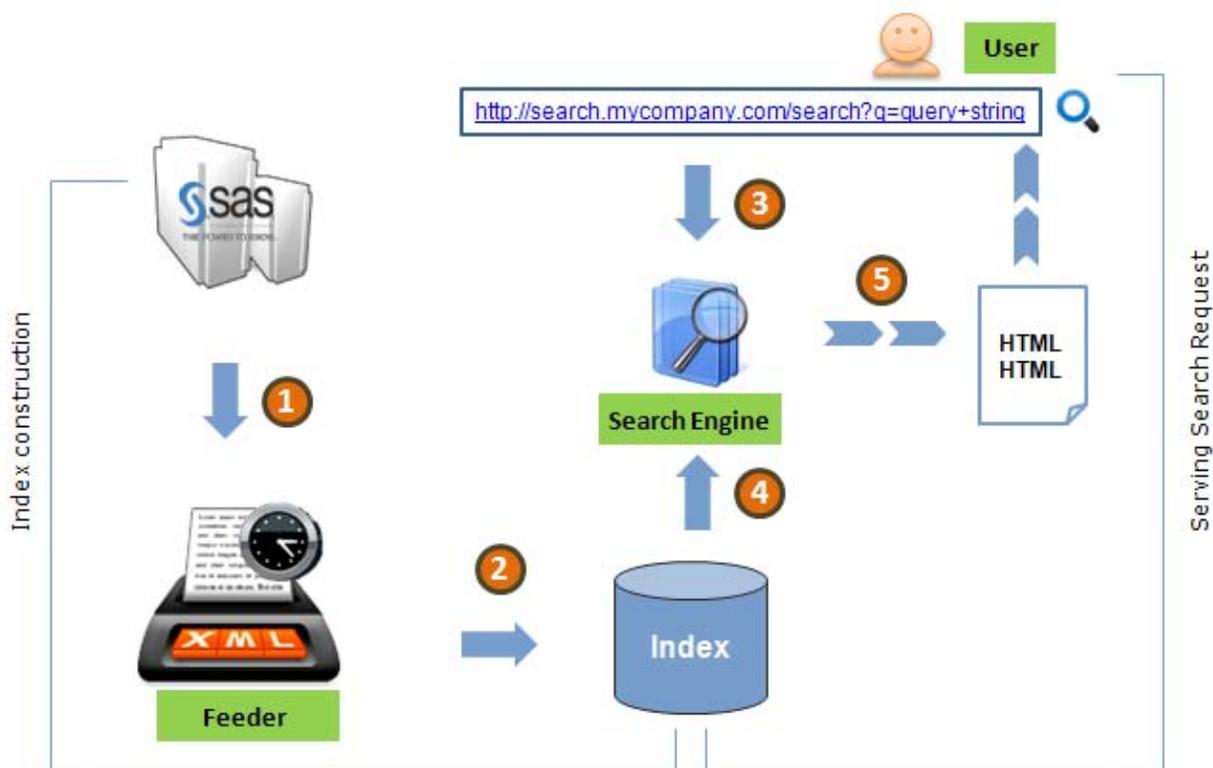


Figure 3. Feeding SAS Content into Search Engines

You can create a feed containing report information from the SAS Metadata Server® and the SAS Content Server®, with the URL for opening these reports in a respective viewer (SAS Web Report Studio®, Stored Process Web application, and so on). This feed will be pushed to the search engine's index. In this approach, since only the metadata is indexed and not the actual report content, the user can see the most updated report content by clicking the URL in the search result. If metadata is modified, the index can be updated with the modified feed.

CUSTOM SEARCH PROVIDER APPROACH

Enterprise search engines provide ways to aggregate search results from various search providers. When a search is performed, the search engine sends the search query to the matching search providers and displays the search results received from these search providers along with the search results from its own index. Google Search Appliance allows third-party search providers to be registered as a OneBox module. Microsoft Enterprise Search Solutions allow third-party search providers to be registered as a federated location.

You can leverage this feature to surface SAS content to enterprise search engines using a SAS Stored Process and Search Interface to SAS Content as search providers. Using this approach enables you to perform real-time search.

INTEGRATING A SAS STORED PROCESS AS A SEARCH PROVIDER

A stored process is a SAS program that is stored on a server and that can be executed by requesting applications. Stored processes can be used for Web reporting, analytics, building Web applications, delivering packages to clients or to the middle tier, and publishing results to channels or repositories. Stored processes can also access any SAS

data source or external file and create new data sets, files, or other data targets supported by SAS. A stored process can be executed through the SAS Stored Process Web Application and have the results returned to a Web browser.

You can use a stored process to provide content as search result in three ways:

1. A search provider that is registered with the search engine - The stored process accepts the search query and returns the search results to the search engine in the format expected by the search engine. Search engines can limit the number of results to be displayed.
2. A link in the search result (for example, a "More Results" link) that invokes a stored process that sends all results in HTML format.
3. A link in the search result that invokes a stored process that creates a report/graph/KPI to show to the user.

The request can come from either the search engine or the browser. The result returned should conform to the standards of the search engine.

To integrate a SAS Stored Process with search engines, do the following:

1. Create ODS templates to format the search result to XML/HTML format conforming to the search engine's response format schema and send the XML/HTML result back to the search engine.
2. Write a stored process for searching the search expression/query in the SAS content (SAS data sets, cubes, and so on) and use the ODS template to format the contents.
3. Designate the server on which the stored process runs and register the stored process in the SAS Metadata Server through SAS Management Console so that the stored process can be invoked through the SAS Stored Process Web application.
4. Register the URL of the SAS Stored Process Web application with the stored process as a parameter as a search provider in the search engine.

INTEGRATING SEARCH INTERFACE TO SAS CONTENT AS A SEARCH PROVIDER

Search Interface to SAS Content is a part of the SAS 9.2 BI platform. The interface provides a way to search the platform through various search engines. Search Interface to SAS Content is a Web application that enables searching SAS content (such as reports and information maps) in the SAS Metadata Server.

Search Interface to SAS Content can integrate with the enterprise search facilities as a search provider, such as:

- Google Search Appliance, as a OneBox module
- Microsoft Enterprise Search Solutions as a federated location
- Teragram

To integrate with these search engines, follow the detailed steps in the respective integration document available on the product's download site.

In addition to searching SAS content, Search Interface to SAS Content leverages metadata relationships to provide additional navigation (for example, similar reports). It supports launching the search results in the appropriate application (SAS Information Delivery Portal®, SAS Web Report Studio, SAS Stored Process Web application, and so on). It is integrated with the SAS Logon Manager for authentication.

INTEGRATING SEARCH INTERFACE TO SAS CONTENT WITH A BROWSER SEARCH TOOLBAR

The latest browsers support integrating OpenSearch clients as search providers. Search Interface to SAS Content supports OpenSearch. You can leverage this facility to integrate it with a browser's search toolbar. This approach does not require any enterprise search facility.

We have integrated Search interface to SAS Content with Browser Search toolbar as a search provider for the following browsers:

- Microsoft Internet Explorer 7.0 or later
- Mozilla Firefox 2.0 or later

After installing Search Interface to SAS Content in a supported Web server, follow the steps under the section "**Integrating Search Interface to SAS Content in the browser search toolbar as a Search Provider**" in the "Readme" document available on the product's download site.

SECURED SEARCH

Security is always a consideration when surfacing SAS content via search. To provide authorized SAS content as search results to the search engines, the user has to be authenticated by the SAS Intelligence Platform.

The Crawl and Index approach or feed loading approach stores the SAS content in the search engine's index, and SAS authorization cannot be applied to this index. Usually, document-level authorization is not supported by the search engine. When you follow one of these two approaches and want to secure SAS documents, you must either leverage the security feature of the search engine or add only public SAS content to the search engine.

Search Interface to SAS Content supports public search and secured search. When you select public search, the results for the Web anonymous user will be returned. When the secured search is selected, SAS Logon Manager authenticates the user and provides search results that the user is authorized to see.

When integrating Search Interface to SAS Content as a search provider to a search engine, various authentication protocols can be selected. They can be *No authentication*, *Basic Authentication*, *authentication using third-party authentication providers*, and *single sign-on (SSO)*.

Refer to the **SAS 9.2 Intelligence Platform Security Administration Guide** provided in the **REFERENCES** section to configure authentication for the SAS Intelligence Platform.

Refer to search engine-specific documentation for the supported authentication protocol for a search provider.

CONCLUSION

The SAS BI platform is very robust and it provides various options to integrate with search engines. You can exploit these options to expand the reach of SAS content to the larger audience of an enterprise through various search engines.

If there is an enterprise search facility, then the Crawl and Index approach provides an easy way to surface batch content. It has limitations such as the inability to crawl proprietary information, providing outdated results, and the inability to integrate results from various sources. Using the feed approach enables you to overcome some of the limitations.

The SAS Stored Process integration approach can be used to provide real-time results from various sources such as the metadata server, SAS data sets, and custom programs. This requires custom programming and development.

The Search Interface to SAS Content product can be used to provide a comprehensive search experience with little or no custom development. The advantages of this product are seamless integration with the SAS middle-tier platform, the ability to integrate with Web browsers through their search toolbar, integration with multiple search facilities and formats, integration with various solutions for viewing the search results, SSO with these solutions, and so on.

REFERENCES

Google (2009), "Google OneBox for Enterprise Developer's Guide", Available at <http://code.google.com/apis/searchappliance/documentation/62/oneboxguide.html>

Microsoft Corporation (2008) "MSDN Federated Search Overview [Search Server 2008]" Available at <http://msdn.microsoft.com/en-us/library/bb931080.aspx>

SAS Institute Inc. (2009), "Building a Web Application with SAS Stored Processes", Available at <http://support.sas.com/documentation/cdl/en/stpug/61271/HTML/default/a003152855.htm>

SAS Institute Inc. (2009), "SAS 9.2 Intelligence Platform Security Administration Guide", Available at <http://support.sas.com/documentation/cdl/en/bisecag/61133/HTML/default/titlepage.htm>

SAS Institute Inc. (2009), "Search Interface to SAS Content download site", Available at <http://www.sas.com/apps/demosdownloads/setupcat.jsp?cat=Base+SAS+Software>

RECOMMENDED READING

SAS Institute Inc. (2009), "SAS ODS (Output Delivery System)", Available at
<<http://support.sas.com/rnd/base/ods/index.html>>

CONTACT INFORMATION

Craig Rubendall
SAS Institute Inc.
100 SAS Campus Dr.
Room R2261
CARY, NC 27513

Chandra Prakash
SAS Research and Development (India)
Level 2A & Level 3, Cybercity Tower 5, Margarpatta City
Pune, Maharashtra, India – 411013

Sudheer Prem
SAS Research and Development (India)
Level 2A & Level 3, Cybercity Tower 5, Margarpatta City
Pune, Maharashtra, India – 411013

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.