

Federated Search

Supporting the Federation of Multiple Heterogeneous Sources Into a Single Unified View

Daston's Service-Oriented Architecture (SOA) developers have designed and fielded a Federated Search framework that can be used to support the federation of multiple heterogeneous sources into a single unified view with minimal effort. This paper highlights the solution's architecture, technology employed, and its implementation at the Defense Information Systems Agency (DISA).

Overview

Daston has developed a Federated Search framework that can be used to support the federation of search results from multiple heterogeneous sources into a single unified view with minimal effort. The search framework is comprised of a template services API for accessing sources, as well as a set of basic search services and search user interface (UI).

Through implementation of the API, the framework is able handle data from virtually any type of data source whether it is a Google index, Lucene/Regain index, database, or even another service. The framework provides some built-in service implementations and additional adapters can be implemented following the template services API.

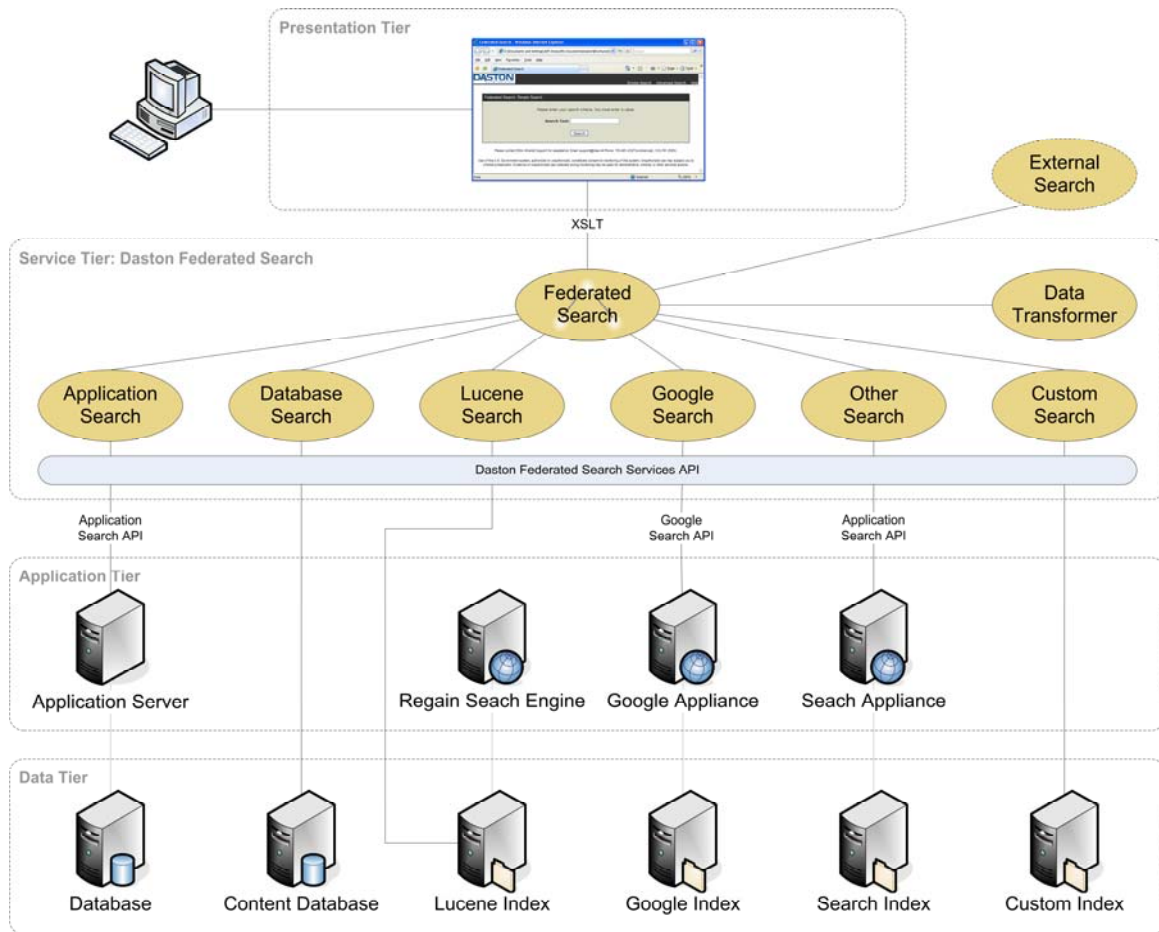
An organization only needs to implement the template services for their sources and then integrate the results into their presentation tier or the search UI.

Solution Architecture

The architecture for the Federated Search services includes following distributed components:

- **Federated Services Search API**
This is the core search API that defines the standards and provides core search service functionality including the service composition logic. It defines the interfaces for the basic and composed services.
- **Federated Search Services**
The implementation of the basic and composed search service interfaces. This component holds the service implementations that are specific to data sources, the federated search composition service, and any required data transformation services.
- **Search UI**
This spring-based front end application invokes the federated search and displays the search results using XSLT.

Daston's Federated Search Framework Architecture



The built-in service inventory provides the following service templates; the actual service implementation of an organization will extend one of these services:

- **Basic Search Service** - A basic search service template that represents the base of all search services. It provides the basic search API, along with the implementation of some common search capabilities. To implement an actual search service for a given data source, the service needs to extend this Basic Search Service.
- **Basic Federated Search Service** - A basic composition service to federate multiple individual search services. While the individual search services need to implement certain API of the Basic Search service, the

actual implementation of an instance of the Federated Search service simply needs to extend the Basic Federated Search service; all of the required functionalities including the federation logic for federated search, are already provided in this service.

- **Search Client API** - Sample client code to show how to invoke search services.
- **Security API** - The search API provides a basic security API for querying data sources for user access-based data. Each service instance can be implemented to pass on the logged in users credential to the corresponding data source for log in purposes.

- Authentication Integration Point - Daston's search framework provides a template for passing the user information similar to that of the NCES Federated Search service. At DISA, this information is used to perform authentication against the user directory (LDAP) via a Java Authentication and Authorization Service (JAAS) module. The JAAS module is responsible for querying the LDAP to ensure that users have the access control rights (permissions) required to do the actions performed.

Technology

The framework was designed following the principles of Service Orientation with a focus on agnostic services so they can be reused for composition purposes within the service inventory. It was implemented following various SOA design patterns and utilizes web services standards and XML technology for maximum flexibility and interoperability. This allows it to be easily deployed either as a stand-alone application or integrated into an existing SOA implementation.

The WS-Standards utilized in the framework are:

- | | |
|-----------------------|-------------------|
| ▪ SOAP | ▪ WSDL |
| ▪ WS-Addressing | ▪ WS-ServiceGroup |
| ▪ WS-MetadataExchange | ▪ XML Namespaces |
| ▪ XML Schema | ▪ XPath |
| ▪ XQuery | ▪ XSLT |

Future enhancements will include implementing

- | | |
|------------------|--------------------|
| ▪ WS-Routing | ▪ WS-Messaging |
| ▪ WS-Transaction | ▪ WS-BPEL |
| ▪ SAML | ▪ WS-Orchestration |

Each search service has an XML configuration file and a standardized contract published via a WSDL. The frameworks' search algorithm is implemented to perform multiple concurrent requests to multiple data sources simultaneously with predefined timeout values for the most optimal performance experience.

The various parameters of a search request are defined in the xml configuration file and can be changed to alter the behavior accordingly. The federated search results are displayed to the end users using XSLT to "translate" the search results into displayable HTML.

The framework was implemented in Java and built on the Apache Axis2 (<http://ws.apache.org/axis2>) web service stack. A WSO2 (<http://wso2.com>) Enterprise Services Bus is also integrated to provides WS-Standard services and a proxy web service to allow for a distributed deployment. Depending on the deployment platform, it is possible to add adapters that run on other web service containers.



DISA Implementation

Daston utilized this framework to implement a federated search solution at the Defense Information Systems Agency (DISA). The primary goal of this federated search application is to help DISA users find information that are stored across multiple organizations within and outside the agency, and bring them all together into one single view seamlessly to the end users. DISA believes this will help its users find pertinent contents for the upcoming BRAC move both for the existing and new employees.

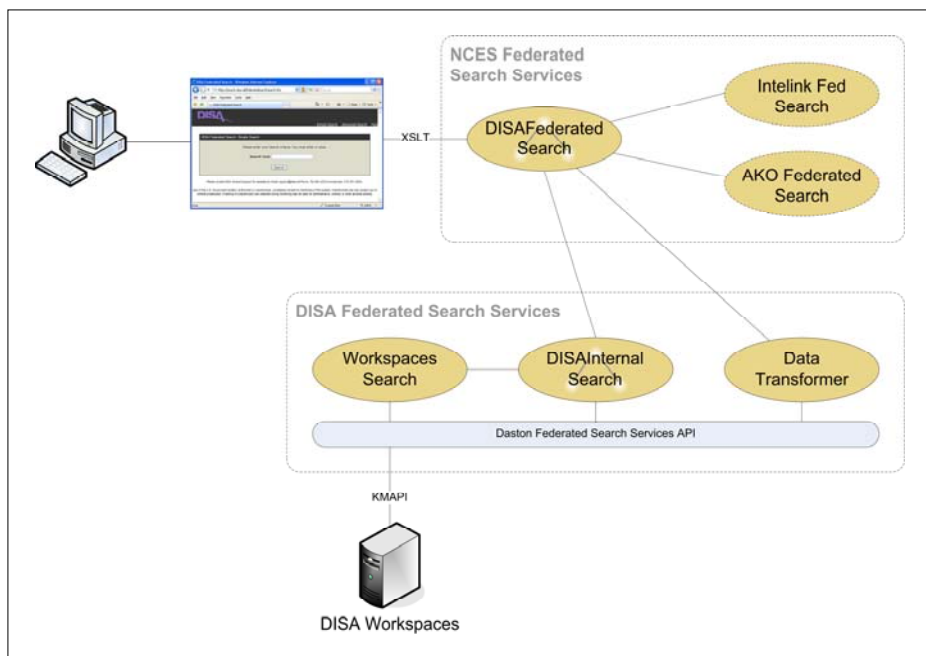
The initial implementation focused on three data sources: DISA Workspaces, Intelink and the Army Knowledge On-Line. The framework also allows DISA Internal content to be exposed to the enterprise through the implementation of the NCES Federated Search Services and inclusion as a source for the NCES Enterprise Search service.

The authentication mechanism functions as follows:

- The DISA Federated Search application requires a user to be logged in to the Workspaces i.e. to the AKO/DKO SSO infrastructure. The UI code retrieves the edipi of the user performing the search upon login and keeps it in session for the search requests performed by the user.

- The UI passes the edipi value as part of the search request to the DISA Federated Search service.
- The search service passes the edipi value to the JAAS module of the Workspaces. Every single access to the Workspaces is protected by this JAAS module.
- The JAAS module authenticates against the user directory (LDAP).

The solution, which is shown in the diagram below, has established a basis for integrating all internal and external DISA content into a single search interface.



DISA's Federated Search Implementation

The DISA Federated Search service is the main service which invokes both external services (Intelink Fed Search and AKO Federated Search) as well as the internal DISA Internal Search services, to combine search data related to DISA into a single data set.

Daston Experience, Innovation and Results

Daston provides the technology infrastructure and the implementation expertise to help your organization set up a foundation for a Federated Search capability. Daston can also help implement additional services to provide a fully functional search application that will federate various data sources into one single view for your end users. Daston has the trained SOA professionals to help integrate these services into your existing SOA infrastructure or establish a new one.

For more information on how we can help your organization implement a Federated Search capability, please contact Michael Pait, Senior Vice President of Business Development, 703.314.8287; michael.pait@daston.com