Service Genre: Search

Introduction

The Introduction provides a brief, standalone overview of the Service Genre. It is for a non technical reader. It may duplicate other material in the Service Genre Description.

The search service genre provides an abstract service end point for discovery of data objects (e.g., LOM metadata XML documents, SCOs, content packages) from collections within a managed resource. The query process takes a query and retrieves information about the corresponding objects in the identified resource.

Service Genre Description

The Service Genre Description is the complete, formal documentation of the Service Genre.

Search is the process by which an application presents a *query* to a resource and the resource responds by returning information about objects in the collection that *match* the query or search criteria. A service interface for the resource provides a mechanism for the external agent (accessor) to send the query to the resource to get the set of query results. The returned set of results is generally made available to an external user or application. The returned results may be the objects that match the query, but are more typically information about the matching objects. This information may include object locators, which allows the search service to be used together with an obtain service that allows access to the full objects. How the search service represents queries and how it matches the query criteria are specified within service expressions that specialize this service genre.

This is a general description of a search service genre, independent of application end point, resource, query language, or underlying communications protocols and service models. The service genre does not include a mechanism to authenticate clients. The service genre can be used in conjunction with authorization methods to control the return and filtering of results. Such authorization can be included in a service expression extending this service genre, or as part of a service usage model that combines the authorization service genre with the search service genre. If the query results only include object identifiers and not object locators, the service genre can also be used in a service usage model with resolution processes (including appropriate copy provisions) to obtain an object when given an object identifier.

Service Genre Metadata

The Service Genre Metadata contains basic labeling, classification and a version history for the Service Genre.

Name

- Service Genre Name: search
- LSAL ID: hdl:1870/662CC57E60FD4EF092F7F0E8C1D6986
- ADL Name: search {collection registry}, search {CSDB}, search {competency resource}, search {HR data resource}, search {knowledge base}, search {metadata registry}, search {repository}, search {repository registry}, search {repository}, search {rights license repository}, search {task list}, search {training catalog repository}, search {TSDB}
- JADL IPA ID: hdl:JADL-IPA-NA/2D71C289AF31422989C74063405CAAFF (source for derivative)
- FRED Service Genre Name: search
- FRED ID: hdl:FREDNA/819843C836A649E4AFE2B82186AB5D60 (source for derivative)

Classification

Classification Facets:

- Service Genre Status: Unapproved
- Domains: Repository, Content Authoring, Training
- Domain Coverage: Multiple
- Deployment Status: Developmental
- Deployment Scale: Isolated

- Maturity: Immature
- Composition: Individual
- Purpose: Exemplar, Application

Technical Facets:

- State Behavior: Stateful
- Transactional Behavior: Non Transactional
- Batch Behavior: Individual
- Time Constraint Behavior: None
- Service End Point: Provider
- Authentication / Authorization: Not Auth'ed
- Exposure: Public

Version

- LSAL Version: 1.0.0 [hdl:1870/662CC57E60FD4EF092F7F0E8C1D6986]
- JADL IPA Version: 1.0.0 [hdl:JADL-IPA-NA/2D71C289AF31422989C74063405CAAFF]

Version History			
Version	**Date**	**Author**	**Description**
0.50	2007-09-04	DR	Initial D^{3}UI version hdl:JADL-IPA-NA/2D71C289AF31422989C74063405CAAFF
0.51	2007-09-25	DR	Editorial review, consistency.
0.70	2007-10-07	DR	Draft for review.
0.90	2007-10-14	DR	Final editorial.
1.0.0	2007-10-21	DR	Final D^{3}UI V 1.0.0.
1.0.0	2008-06-18	DR	LSAL V1.0.0. Derivative from D^{3}UI version.

Notation

The Notation element includes conventions used to describe the Service Genre.

The words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL in this document are to be interpreted as described in [RFC 2119].

Notational conventions follow those given in the LSAL ***Service Notation and Document Conventions***.

The identification and versioning scheme follows those given in the LSAL ***Service Component Identification Scheme***.

The service classification scheme follows those given in the LSAL ***Service Classification Scheme***.

The Service Genre Description follows those given in the LSAL ***Service Genre Description Guidelines***.

Description

The Description element is an informal, standalone, non technical narrative description of the Service Genre (problem, process, business-level capabilities and workflow).

The search service genre provides the mechanism to query a resource to find objects matching the query. It is an example of a request-response process where a single request returns a composite or multi-part response. It may be part of a stateful workflow. The search service genre may be used to send a query to a single or to multiple resources through the single search service end point; if these resources are hosted as distinct entities (e.g., as separate repositories), the function provided by the service genre is termed *federated search*.

The resource is assumed to be a collection of data objects, each of which is discoverable through a set of attributes (keyword, label, object text) collectively denoted *search terms*. These search terms are used in the matching or query process to determine the results set, i.e., the objects that match the query are determined according to a comparison between the objects and the search terms. The prime use of this service genre is to discover one or more of the objects in the collection that match the query.

This service genre focuses on discovery of information from repositories, registries and other similar data collections of discoverable content objects. Typically, the data retrieved will be information about the matching objects in the repository being searched (not the entire object). The search may specify what type of information is to be returned (e.g., identifier, metadata, summary, full object) for the objects in the results. Searching is used to discover content objects and present information about them to external applications and end users. The service genre provides a discovery interface for any content repository or registry that is part of the federation. Combined with a service such as the obtain service, it provides the mechanism to retrieve the discovered object.

The service genre may be specialized in service expressions to gather particular types of information about the objects in the results, to specify query language or to specify communications, messaging and transport protocols.

As defined, the search service genre is not access controlled; i.e., any client may attempt to contact a search service end point. There are no authentication controls. The service end point for the resource is responsible for determining what results it will return and from which clients it will accept requests.

Note, as with other service genres, there is no expectation that all service implementations will be interoperable. Different service expressions MAY take different approaches to defining interfaces and data models.

Usage Scenarios

The Usage Scenarios element is an informal, non technical description of how the Service Genre is used. An illustration of process or problem workflows, expressed using services, is included. An illustration of an application using the components of the Service Genre may be included (but not a description of the design of the application). No critical or essential information required to understand the Service Genre should be included.

In the simplest case, an end user (or a system acting on behalf of the end user) wishes to ascertain what objects are available from the resource that match the given search criteria. The requestor issues the query request, and receives a response in terms of metadata describing the matching objects. The requestor may process the metadata further, to identify particular objects of interest. This further processing may take the form of another search (operating on the resource and constraining the previous query; or it may be local search on the retrieved data). Alternatively, the results may be presented to the end user to enable browsing through the search results. The requestor may then use data from the retrieved metadata descriptions (typically an identifier or locator) to trigger a distinct service of obtaining or accessing the objects identified in by the search. A typical usage workflow is:

1. Make search request.
2. Retrieve set of metadata for matching objects.
3. (Optional) Identify objects of interest through metadata description.
4. (Optional) Obtain objects of interest through identifiers and locators included with metadata description.

This scenario may be enhanced as follows:

- The search returns the matching objects themselves, rather than metadata describing them. Step 4 is unnecessary.
- The metadata descriptions retrieved are filtered, so that only specific fields are returned. For instance, only the locators of the objects concerned are returned, so that the requestor may proceed directly to obtaining them. If enough information is still present in the filtered record, Step 3 is still possible.
- The complete set of matching records is not retrieved in the one request, due to processing or client constraints. The retrieval of matching records is spread out over several request–response pairs, amounting to repetitions of steps 2–4; the results set persists over a series of requests (possibly not all). Each response contains a resumption token, which is passed back to the resource to retrieve the next subset of matching

records: for example, the response containing the 100th though 149th matching result records (object descriptions) would contain a resumption token, which could be used to request the next 50 records (the 150th through 199th matching result records). The resumption token enables flow control.

- The requestor may not already know what information is available from the provider, including what attributes the resource may be queried by, what the possible formats for results are, what the possible query languages are, etc. Preparatory to making the search request, the requestor may query the resource for metadata about its search interface, in order to formulate its search request accurately.

The basic usage scenario for the search service genre is to embed a query interface in another tool, e.g., within the *Reload* editor, as demonstrated in the (D3UI) Content Authoring service usage model. The user enters a query which is passed to the search service. Search returns information about items that match the query. These are displayed to the user and the user selects the object to process. All of the interactions with the search service are hidden from the end user (the end user MAY not know which service end point was used) and the query, response, display, selection process is fully automated. This is essentially a simple GUI search interface to a collection managed by a resource that is embedded in another tool.

Applicability

The Applicability element details when the Service Genre is used or not used. It represents specific constraints and assumptions on the use of the Service Genre. It is more specific and normative than the informal Usage Scenarios. No critical or essential information required to understand the Service Genre should be included.

As defined, the search service genre is applicable for discovery of objects managed by any resource.

The service genre is applicable when results returned are for:

- the discovered objects
- metadata about objects that match the query
- identifiers or locator proxies for objects that match the query

The service genre is applicable to discovery from a single service end point. There may be one or more resources behind the service end point.

The service genre is applicable when results are returned in one set, or when returned in chunks over a stateful workflow.

Any processing of the search results is out of scope for this service genre.

This service genre does not define behavior when the resource requires authentication to permit search.

This service genre does not define behavior when the resource requires authorization or access controls to permit search.

This service genre does not define behavior when the search service end point attempts to apply a filter relying on information not available to the resource. For instance, if a filter relies on authorization policies or rules that are not communicated to the resource through the query interface, the behavior is not defined.

This service genre does not define behavior if communications need to be secure.

Functionality

The Functionality element details and illustrates the behaviors provided by the Service Genre, in terms of services, workflows, messages, resources, and data objects. It is not a technical description of the Service Genre, but it must provide sufficient information to develop the Requests & Behaviors of the Service Genre and to evaluate conformance of the Service Genre to the stated behaviors. It should not include implementation-specific information.

The search service genre supports two types of functions:

Query Functions. These functions are used to query the resource and obtain the results set.
Requests SHALL specify:

- A query defined in a query language [REQUIRED].

Requests MAY specify:

- The query language being used in the request [OPTIONAL].
  - A query language specification SHOULD refer to existing standards and SHOULD include version numbers or version information.
- The desired format or digital representation for the result, e.g., the schema of the objects in the results set [OPTIONAL].
  - Representations SHOULD refer to existing standards and SHOULD include version numbers or version information.
- Filters to be applied to the results set before it is returned to the client [OPTIONAL]. For example, the filter may specify: selection of parts of the results set, formatting the results set, sorting the results set, grouping similar results, ranking, merging.

Mechanisms MAY exist to provide flow control and persistence of results so that large results sets are returned in chunks or so that queries MAY request a part of a results set.

The query functionality MAY be split across multiple behaviors and requests. Some service expressions that specialize the service genre MAY manage control information separately from the query request while others MAY combine the control and query request into a single behavior. Separating these behaviors implies that a query request is stateful.

In federated search, the specification of which subset of resources to query is assumed to be a part of the query specification or the control data, not a separate function.

Search Description Functions. These functions are used to get descriptive information about the capabilities of the service end point. This information enables a client to successfully communicate with and obtain information about the service end point that is providing the interface to the resource. The information SHOULD be machine processible, so that a service can formulate a query based on search description information without human intervention. Description information gathered MAY include:

- Description of the resource at the service end point.
- Description (machine processible) of communications and transport protocols supported.
  - Protocol information SHOULD include version numbers.
- Description of information that can be included in a query request (e.g., search terms).
  - Schema and digital representation format information SHOULD include version numbers.

No other functionality is defined. The functionality that is defined MAY be extended. Major new or additional functionality SHOULD NOT be included; extended capabilities SHOULD be included in other service genres.

Authenticated query SHALL be defined as a separate service genre.

Requests & Behaviors

The Requests & Behaviors element details all of the behaviors exposed by the Service Genre. It lists functionality that can be used by applications or Service Implementations. The information must be sufficient to specialize the Service Genre to one or more Service Expressions.

The format and definitions for requests and responses SHALL be defined by the service expressions that specialize the service genre. Requests and behaviors SHALL meet the following conditions:

- At least one request providing a *Query Function* SHALL be defined.
  - The request SHALL process the defined query language.
  - The query language MAY be profiled.
  - The request SHOULD include mechanisms to specify or control the format or content of results sets.
  - The request SHOULD include mechanisms to specify the part of the results set to be returned.
  - Responses SHOULD include flow control.
- At least one request providing a *Search Description Function* SHOULD be defined.
  - The response SHALL return basic metadata about the target resource.
  - The response SHOULD include protocol and result format information.
- Responses SHALL include error indicators or other needed control information.
  - Indicators SHALL be available for the requests as a whole (e.g., malformed query).
  - Separate indicators SHALL be used to describe the availability of the results.

Use & Interactions

The Use & Interactions element details how the how the Requests & Behaviors are combined to provide the stated functionality of the Service Genre. This is a precise technical description of how the Service Genre provides its capabilities.

The model for a client to interact with a service implementation SHALL be defined by the service expressions that specialize the service genre.

Structure

The Structure element provides a conceptual model of how the Service Genre manipulates data and state to provide results in response to requests. An illustration of the structure should be included. The structure information is used to specialize the Service Genre to one or more Service Expressions, but is not needed to understand how to use or interact with the Service Genre.

The structure of the service genre SHALL be defined by the service expressions that specialize the service genre.

The service endpoint provides the interfaces for the requests and responses for query functionality and search description functionality. Internally the service implementation MAY include various features (indexing, flow control, a query persistence layer, etc.) needed to provide the defined functionality. The service genre interfaces with the resource holding the object collection and the search description data.

The overall flow of interactions with the service is illustrated in the diagram.

If the response to a search request is too large to fit in a single response, then mechanisms are used to retrieve the response from the resource over a series of requests, each request asking for another subset of the set of matching objects. The set of objects matching the query needs to persist on the resource over (a part of) the duration of the series of requests (*persistent results*). Moreover, the sequence of requests for subsets of the results set needs to be coordinated, so that the entire results set is ultimately retrieved (*flow control*). The data used to coordinate the series of requests is termed *control information*.

Search results may be passed through a *filter* before being passed back to the client. A filter specifies a data transformation that is applied to the response prior to being delivered to the client. The client receives the transformed version of the response, rather than the original response.

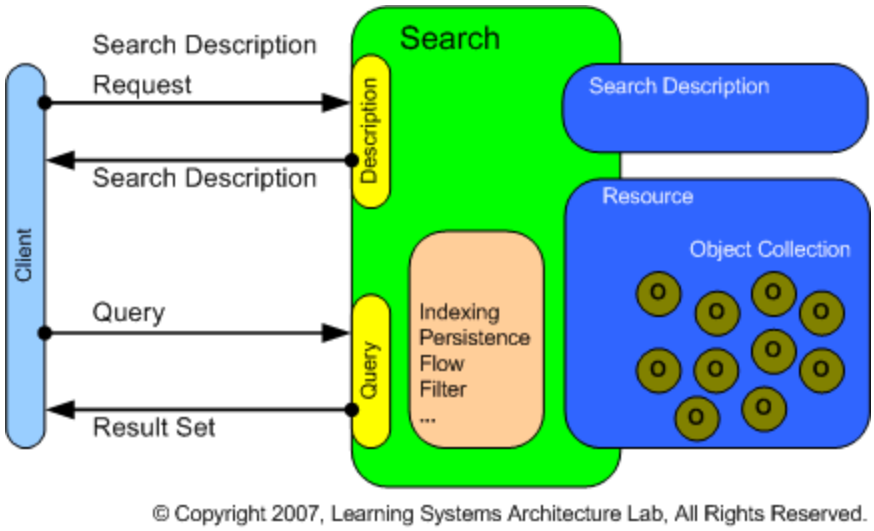

Figure 1: Search Service Genre [© Copyright 2007, Learning Systems Architecture Lab, All Rights Reserved.]

Design Decisions & Tradeoffs

The Design Decisions & Tradeoffs element documents overall choices, tradeoffs and their implications on the design of the Service Genre. It does not address the issues related to the internal details of a Service Implementation used to implement the Service Expression based on the Service Genre. No critical or essential information required to understand the Service Genre should be included.

Internal processes for search are not defined; the service genre only defines interfaces and observable behaviors, not internal models and structure. Processes SHALL be defined in the service expressions that specialize the service genre.

Search is designed to not require authentication. Anyone can search. The process of obtaining the corresponding object may require authentication or authorization controls. Authenticated or authorized search could be described in a separate service genre.

Implementation Guide & Dependencies

The Implementation Guide & Dependencies element describes issues of concern in specializing the Service Genre to one or more Service Expressions and their corresponding Service Implementations. Resolution of issues discussed is deferred to the actual Service Implementation design. No critical or essential information required to understand the Service Genre should be included.

The following design decisions apply to the service expressions that specialize the service genre.

Design:

- The service expression MAY include flow control for managing results. The service expression SHALL define if flow control is supported, and limits on request size, results sets size and results set persistence.
- The service expression MAY include the specification of the query communications protocol as part of its definition (e.g., as in SRW) or it MAY layer the functions on top of another defined communications protocol (e.g., using SQI as the communications and control protocol).
- The service expression SHOULD clearly and cleanly separate service search description functions from query functions.

Consistency:

- The service implementation SHOULD ensure that all discoverable objects and all defined data formats are discoverable and included in the results set, i.e., if an object is in the resource, it should be searchable; if a data format is specified, the results set should be expressible in the defined format.
- The service implementation SHOULD ensure that once an object is included in a persistent results set, the object description SHOULD remain available from the resource in the data format specified throughout the lifetime of the results set. If all objects included remain available throughout, the results set is referred to as consistent.
- The service implementation MAY NOT ensure consistency of results across stateful queries. For example, if a results set is large and it is returned in chunks across multiple steps, an object that is part of the results set may be deleted from the resource but this deletion may not be reflected in the results set.
- The timing of updates and transactions on the resource MAY impact search requests in a way that would omit objects from the results set.

Performance:

- A service implementation SHALL be capable of handling simultaneous requests from different clients.
- A service implementation SHOULD implement an indexing scheme or equivalent method to permit efficient discovery.
- Load balancing SHOULD be implemented for large resource collections or those that are searched frequently.

Security and Privacy Considerations:

- Service implementations may be subject to denial-of-service attacks.
- Care should be taken to maintain privacy of any personal data or other records that may disclose usage patterns.
- There are no authorization or authentication controls. Care should be taken to maintain data privacy.

Applicable Standards

The Applicable Standards element lists domain-specific standards applicable to the Service Genre as a whole. Standards are described in terms of name, version and citation link. Conformance requirements and extensions should be noted. Standards used to implement applications are excluded. No critical or essential information required to understand the Service Genre should be included.

None. No standards are directly applicable to the service genre as a whole.

The service expressions that specialize the service genre SHALL be defined in terms of standards:

- Service expressions SHALL specify applicable query language standards or query language standard profiles (e.g., CQL).
- Service expressions SHALL specify an applicable data model of what is searchable (i.e., what the allowed search terms are; e.g., Dublin Core).
- Service expressions SHALL specify applicable query communications interfaces (e.g., SRW/SRU, Z39-50).
- Service expressions SHALL specify applicable data encoding and representations for the returned objects (e.g., XML representation of LOM).
- Service expressions SHALL specify applicable communications, encoding and transport protocols.

Known Uses

The Known Uses element documents actual uses of the Service Genre in applications and systems, including how used, extensions, limits.

Actual: The service genre is specialized in the D^{3}UI development-to-deposit prototype using *Reload* to discover content from a federated metadata registry.

Actual: The service genre is specialized in the D[3]UI discovery-to-delivery prototype using the *ADL SRTE* to discover content from a federated metadata registry.

Potential: The service genre could be used in any content management or content creation process to discover content from a managed resource.

Potential: The service genre could be used in a service usage model for a repository federation. Search could be used to discover content from any resource in the repository federation (repository, registry).

Potential service expressions: Specializations of the service genre include:

- search resource – sru basic: Search a repository using SRU as defined (CQL, http).
- search resource – srw basic: Search a repository using SRW as defined (CQL, SOAP).
- search resource – srw ws: Search a repository using an SRW web service (CQL, WSDL over SOAP; based on WS-I).
- search resource – open search google ajax: Search a repository using Open Search as defined with the Google query language used in the Google AJAX API.
- search federation – sqi basic: Federated search using SQI.

Service Genre Dependencies

The Service Genre Dependencies lists other Service Genres that this Service Genre is dependent upon. Dependent Service Genres are identified by name and version.

None.

Related Service Usage Models

The Related Service Usage Models element documents and illustrates how the Service Genre is used in Service Usage Models. Related Service Usage Models are identified by name and version. No critical or essential information required to understand the Service Genre should be included.

(D[3]UI) content authoring: V1.0.0.
[D[3]UI content-authoring-sum-v100 / hdl:1870/3FD212C661F1464AA79DBB22D3DE00FF]
Search (genre) is part of the (D[3]UI) content authoring service usage model (genre based) and is used within an authoring and content deposit workflow to find content.

(D[3]UI) delivery: V1.0.0.
[D[3]UI delivery-sum-v100 / hdl:1870/01D0643B992B489ABEB4D89806918450]
Search (genre) is part of the (D[3]UI) delivery service usage model (genre based) and is used within a discovery-to-delivery workflow to find content.

(FRED) repository federation: V1.0.0.
[link to service usage model]
Search (genre) is a part of the repository federation service usage model (genre based) and is used to discover content from the repositories and collections that participate in the federation. The repository federation service usage model provides an integrated set of service genres used to populate and use the metadata registry that supports a repository federation. Functionality includes content management (creation and management of metadata objects within a repository federation), content discovery (discovery of content objects from a repository federation) and content delivery (retrieval of and access to content objects discovered through a repository federation).

(ADL-R) repository federation: V1.0.0.
[ADL-R repository-federation-sum-v100 / hdl:1870/E2FE4AD428A1468FA284E270245F72D7]
The (ADL-R) repository federation service usage model is a derivative subset of the (FRED) repository federation service usage model. For the purposes of this service genre they are interchangeable.

Related Service Patterns

The Related Service Patterns element documents and illustrates how the Service Genre is used in Service Patterns. Related Service Patterns are identified by name and version.

None.

References

The References element includes references and bibliographic citations to works needed to understand the Service Genre.

None.

Glossary & Terminology

The Glossary & Terminology element defines domain-specific terms used in documenting the Service Genre.

Terms in the LSAL ***Service Glossary*** are applicable to this Service Genre.

Working Notes / Things To Do

The Working Notes element documents open issues in the development of the Service Genre and is for internal project use only. It should be deleted before the Service Genre is submitted for publication.

None.