

NISO RP-2005-02

NISO Metasearch Initiative

Search and Retrieval Results Set Metadata

A Recommended Practice of the National Information Standards Organization

Standards Committee BC / Task Group 3

**Version 1.0
September 13, 2005**

Published by the National Information Standards Organization
Bethesda, MD



Metasearch Search and Retrieval Results Set Metadata

Introduction

A sub group of the Metasearch Initiative Search and Retrieval Task Group was assigned the responsibility of identifying a core set of metadata that provide information about a result set at both the aggregate level (see Aggregate tab in this workbook) and the individual record (see Individual tab in this workbook). These data elements are intended to be used by content providers to provide better quality of information returned through a variety of methods. It may also be used to ensure that the needs of metasearch products are met by a given protocol.

The group used the following assumptions in developing the recommended metadata elements:

- 1 The elements in this document apply to metadata related to the results of a search. Specifically excluded is information about the query interpretation (subquery: recommendation) and elements specific to a response (e.g. RetMax and RetStart).
- 2 Result set metadata is nested. So many of the data elements can be reused in the context of a specific database. Elements such as "results by database: count" can be handled through "record count" in a nested structure.
- 3 It is recognized that there may be value to providing result set level defaults for particular elements (e.g. schema id); however this document does not address that particular issues. That is left to a particular protocol to define those elements.

Metasearch Initiative Search and Retrieval Results Set Metadata

Aggregate Results

Element	Description	Notes
Branding	Provides branding information from the content provider. May be an image, text, or pointer to branding information.	
Clustering element	Contains the two sub-elements: field and value.	Reconciling across clustering mechanisms from different providers may be problematic
clustering field	The field used for clustering.	
clustering value	The value of the cluster term (common value) e.g. English.	
Clustering ID	An id that points to the result set representing records in the cluster. Particularly useful when an element/field is not appropriate.	
Database name	The name of the database providing the result set.	May be different than what the user selected. E.g., the search used a database name that represented an aggregated set of databases and the results are return by individual database.
Diagnostic Message	Informative message returned with the result set.	
List of terms	Expanded query; what actually got searched.	
Posting count by	Counts related to terms.	
Ranking algorithm	Algorithm used to rank the result set.	
Record count	Number of total records in the result set.	
Resources Used	What did it cost to generate the result set (the cost so far) or execution time.	
Results status	Report on the success/failure of the query operation.	
Sort Order	Sort order of the records in the result set.	

Metasearch Initiative Search and Retrieval Results Set Metadata

Individual Record

Element	Description	Notes
Application URI	Application to be used (e.g. takes user to native interface)	
Character set	Character set of the record	
Constraint	A human readable string expressing constraints on the use of the record. (May be an element out of the descriptive record)	May be of use to provider. If available it is likely to be important for the metasearch engine to display
Cost Info	A human readable string describing cost information to access the record.	
Date creation	Creation data of the administrative metadata	Can be used for alerts services
Database name	The name of the database that the record belongs to for display purposes.	
Date of Last Modification	Last time the record was modified in the database.	
Date of Last Review	Last time the record was reviewed or validated.	
Full Document URL	Fetches document out of the database to be used in a different context.	
Language of Record	Language metadata record.	
ID local	A control or record id that is unique to the database.	Of potential value if provider supports searching on the value or for use with secondary applications (e.g. OpenURL).
Modifiable by	A single agent which is able to modify the record.	
Position in result set	The order of the record in the result set.	
Processing instructions	Instructions used for processing results.	
Rank	Relevancy ranking for ranked retrieval.* See also Score, below.	If the result set is relevance ranked, the rank and position in result set elements are redundant.
Record Created By	The agent which created the record.	
Record Modified By	The agent which modified the record at the given time stamp.	
Schema instance	Instance of schema used.	Important to metasearch provider to process the document. Should be present if not at result set level
Schemas available	Listing of schemas that the record is available in.	
Score	Relevancy score for ranked retrieval.* See also Rank, above.	Important to metasearch engines as a way to combine result sets. This allows the providers to give information about the importance of the record.
Size	The size of the record for the schema	The schema specific leads to additional questions in the area if this is handling structured metadata.
User message	Human readable message	
Within	A single local identifier for the record within the database, equivalent to the ID attribute in XML.	

** Note: Rank and score pertain to relevancy ranking for ranked retrieval. The server may assign a score and/or a rank to a result set record. The score or rank applies to the record relative to other records in the result set. A score is an integer between zero and 100. The rank of a record is an integer from 1 to N, where N is the result set size. (A higher score means more relevant, while a lower rank means more relevant.) It is assumed that if record A has a better (higher) score than record B, then it will also have a better (lower) rank, however score and rank differ in the following respect: no two records in the result set have the same rank, and for every integer between 1 and N (where N is the result set size) there is exactly one record with that rank; on the other hand, more than one result set record may share the same score, and there need not be a record for every possible score. The reason for defining both (Score and Rank) is that some ranked retrieval systems score records while others rank records. The format of the content is left to the protocol. [Paraphrased from Z39.50 Utility Attribute Set (<http://www.loc.gov/z3950/agency/attrarch/archive/util-d3.html>). Although it talks about access points, they have the same semantics as for retrieval elements.]*