Elsevier Scopus Search Prototype

March 17, 2014

Version 1.0

**Content**

1. Batch/Bulk loading. Snapshot of Scopus content (Core records, author records, and affiliation records) from June 2013. The content will be in the current XOCS storage format (not syndication format). This implies that certain content that is currently available to FAST for indexing will not be present in the PoC and this may impact functionality that can be tested. This is most pronounced in the author records since associated institution record information will not have been merged into the author records. The Content will be provided in S3 buckets with appropriate IAM credentials for retrieval. Signed URLS for the ‘keys’ to the assets stored in the S3 buckets are provided in the appendix. The snapshot consists of:
   1. 50,773,230 core records stored in the abstract-xml S3 bucket.
   2. 49,705,979 author records stored in the author-xml S3 bucket
   3. 8,163,349 affiliation records stored in the affiliation-xml S3 bucket.
2. In addition to the raw content, the loading system must compute current corpus-wide statistical measurements and make them searchable in the index. One of the prime requirements for the Scopus product is the need for a tight synchronization of derived document statistic counts between the search index and content repository as modifications to the corpus are made. The proposed solution must ensure this synchronization.
   1. Current publication count for each author.
   2. Current publication count for each affiliation.
   3. Current cited by count for each core record.
3. In addition to the statistical measurements, the loading system must also maintain the list of citing identifiers for each core record. This information must be kept synchronized and updated into each core record.
4. The designed system must make any document changes and associated statistical counts visible within the Scopus product no more than 24 hours after receipt by the XOCS Fabrication system. In other words, it’s acceptable to ‘flip’ the index once per day.
5. Test updates during the PoC using repeat chunks of documents from the original load. The PoC team has been monitoring production update volumes over a 3-month window to determine the typical daily update load for each content type. There will be a set of document identifiers provided that will be used to update each content type during testing. Since we don’t want to change our ‘static’ search index, we will treat all changes to the search index as updates (replacing the content with the same content). We will not add (or delete) content when applying updates. The daily update volumes are:
   1. core records (400,000)
   2. author records (400,000)
   3. affiliation records (250,000)
6. Ensure that at least a 100% and 300% average update volume for each content type can be fully processed within the 24 hour indexing window. Signed URLS for the ‘keys’ to the assets (core, author, affiliation) stored in the S3 buckets that should be used for updates are provided in the appendix.
7. For author records, the system must accommodate a once-per week update of 2 million additional records (that must also be processed within the 24 hour indexing window). This is to accommodate the weekly merge of affiliation changes into the author records. A signed URL for the ‘keys’ to the author assets stored in the S3 bucket that should be used for updates are provided in the appendix.
8. Test impacts during the PoC for the publication count and cited by count changes. The PoC team has been monitoring production update volumes over a 3-month window to determine the usual daily impact for each count.
   1. Core cited by counts (500,000)
   2. Author publication counts (1,000,000)
   3. Affiliation publication counts (25,000)
9. Ensure that at least a 100% and 300% average publication count and cited by count can be fully processed within the 24 hour indexing window. Signed URLS for the ‘keys’ to the assets (core, author, affiliation) stored in the S3 buckets that should be used for publication count and cited by count updates are provided in the appendix. A default value of 10 can be used for both the publication count and cited by count value.
10. Identify architecture and cost to re-index entire corpus.
    1. Re-index of entire corpus (for all content types) must happen within 24 hours.
    2. Must include cited by count and publication count information.
    3. Must include citing document information.
11. Handle Elsevier XML content including namespaces, DTD versions, etc.
12. Indexing is based on the Scopus 13.3 Index Profile.

* 1. Match common index profiles for Scopus (string data, numeric data, date data, facet fields, sortable fields, returnable fields, etc.)
  2. Fields necessary for exact phrase (punctuation sensitive queries) should be defined.
  3. Fields necessary for boundary matching queries (starts-with or equals) are only required for the author initials and author surname fields and only in the ‘author’ cluster.
  4. Sample Java code will be provided to help with mapping fields identified in the to the appropriate xpath expressions.
  5. Some adjustments have been made to the Scopus 13.3 Index Profile where fields/data may not be readily available. In those circumstances, the sample Java code will highlight the deviations.
  6. A signed URL for the Scopus 13.3 Index Profile is provided in the appendix.

**Query Sets for Load Testing**

1. Query sets were harvested from production Scopus traffic from November 19 through December 3, 2013. There is a separate query set for each content type (cores, authors, and affiliations). Signed URLS for the query sets stored in the S3 buckets are provided in the appendix. The snapshot consists of:
   1. 91,533 affiliation record queries
   2. 289,655 author record queries
   3. 2,179,581 core records queries

The above query sets have been further divided into the following categories based on patterns observed in the query sets. This will help identify/isolate performance characteristics for each of these subsets.

Affiliation Record Queries

* 90,504 afid queries (55,788 unique)
* 1,029 User/Other queries (921 unique)
* Navigatiors (1032)

Author Record Queries

* 176,510 auid queries (112,091 unique)
* 52,221 auidafid queries (44,239 unique)
* 60,924 user/other Queries (47,338 unique)
* Navigators (41,932)

Core Record Queries

* 138,308 doi queries (115,213 unique)
* 246,687 eid queries (230,073 unique)
* 97,661 sdeid queries (65,493 unique)
* 169,473 refeid queries (144,780 unique)
* 381,369 auid queries (290,295 unqiue)
* 15,428 afid queries (6,930 unique)
* 16,268 issnisbn queries (5,602 unique)
* 33,441 collecid queries (31,607 unique)
* 53,639 refeidneid queries (52,227 unique)
* 26,738 srcid queries (19,173 unique)
* 566,691 scopout queries (523,392 unique)
* 433,878 user/other (354,057 unique)
* Navigators (226,799)

1. All queries will be defined in Elsevier XQueryX (an xml representation of the query). Fields referenced in the XQueryX should map to the fields defined in the Scopus 13.3 Index Profile. Some exceptions will occur with some facets and sortable fields, as all information may not be available. In those situations, alternative fields will be suggested. The XQueryX will define the following:
   1. User query
   2. Filter query
   3. Fields to be returned
   4. Navigators (facets) to be returned
   5. Sort fields
   6. Number of results to return
   7. Starting offset to return results
   8. Whether to highlight (only core record queries)
2. Load Testing queries have been simplified to only include basic AND/OR/NOT. Nested/scoped queries (searching for a specific author or affiliation) and proximity queries have been removed. Boundary matching and punctuation sensitive queries will still exist in the Load Test query set.
3. Stop words can be removed from a query provided they are not part of a phrase. A signed URL for he Scopus stopword list is provided in the appendix
4. For core queries identified as being ordered by score (relevancy) the following field weighting should be used when the fields are explicitly contained in the query or are part of a composite field specified in the query. Both the punctuation sensitive version of the field and regular version of the field should be boosted.
   1. Field ‘abs’ weight of ‘8’. Defined in composite fields ‘all’, ‘allmed’, and ‘allsmall’.
   2. Field ‘affil’ weight of ‘2’. Defined in composite field ‘all’.
   3. Field ‘auth’ weight of ‘8’. Defined in composite fields ‘all’ and ‘allmed’.
   4. Field ‘chemname’ weight of ‘2’. Defined in composite field ‘all’.
   5. Field ‘coden’ weight of ‘2’. Defined in composite field ‘all’.
   6. Field ‘collab’ weight of ‘2’. Defined in composite field ‘all’.
   7. Field ‘confall’ weight of ‘2’. Defined in composite field ‘all’.
   8. Field ‘corres’ weight of ‘2’. Defined in composite field ‘all’.
   9. Field ‘doi’ weight of ‘2’. Defined in composite field ‘all’.
   10. Field ‘ed’ weight of ‘2’. Defined in composite field ‘all’.
   11. Field ‘isbn’ weight of ‘2’. Defined in composite field ‘all’.
   12. Field ‘issn’ weight of ‘2’. Defined in composite field ‘all’.
   13. Field ‘itemtitle’ weight ‘12’. Defined in composite fields ‘all’, ‘allmed’, and ‘allsmall’.
   14. Field ‘keywords’ weight ’8’. Defined in composite fields ‘all’, ‘allmed’, and ‘allsmall’.
   15. Field ‘lang’ weight of ‘2’. Defined in composite field ‘all’.
   16. Field ‘pub’ weight of ‘2’. Defined in composite field ‘all’.
   17. Field ‘srctitle’ weight of ‘2’. Defined in composite field ‘all’.
5. The system must support the following qps against each content type and performance criteria. Queries against a content type have been further divided into ‘navigator’ queries vs. ‘no navigator’ queries as they will likely have different performance requirements. Detailed below is the breakdown of the queries (and percentages of the overall query mix) for each type.

This performance data was gathered from March 2014. It does include the network latency from the FAST Search Service in Dayton to FAST engine in Boston.

**Volume of Navigator vs No Navigator Searches**

|  |  |  |
| --- | --- | --- |
| Cluster | Navigator | No Navigator |
| Scopus | 10% | 90% |
| Institution | 1% | 99% |
| Author | 15% | 85% |

**Max & Avg QPS**

|  |  |  |
| --- | --- | --- |
| Cluster | MAX QPS | Average QPS |
| Scopus | 91 | 65 |
| Institution | 8 | 4 |
| Author | 15 | 9 |

**Performance By Cluster (Milliseconds)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Mean | Median | P95th | Max |
| Scopus Navigators | 452 | 259 | 1022 | 65058 |
| Scopus No Navigators | 397 | 171 | 633 | 179189 |
| Author Navigators | 242 | 101 | 1013 | 29091 |
| Author No Navigators | 131 | 46 | 174 | 28168 |
| Institution Navigators | 129 | 103 | 243 | 979 |
| Institution No Navigators | 95 | 70 | 180 | 3643 |

The data below was based on the collected set of queries used to drive the Hothouse testing.

|  |  |  |
| --- | --- | --- |
| Content Type | Query | Percentage of Mix |
| Affiliation | Infra (afid) | 99% |
| Affiliation | User/other | 1% |
| Affiliation | Navigators | 1% |
| Affiliation | No Navigators | 99% |
| Author | Infra (auid) | 61% |
| Author | Infra (auidafid) | 18% |
| Author | User/other | 21% |
| Author | Navigators | 15% |
| Author | No Navigators | 85% |
| Core | Infra (doi) | 6% |
| Core | Infra (eid) | 11% |
| Core | Infra (sdeid) | 4% |
| Core | Infra (refeid) | 8% |
| Core | Infra (auid) | 17% |
| Core | Infra (afid) | 1% |
| Core | Infra (issnisbn) | 1% |
| Core | Infra (collecid) | 2% |
| Core | Infra (refeidneid) | 3% |
| Core | Infra (srcid) | 1% |
| Core | Infra (scopout) | 26% |
| Core | User/other | 20% |
| Core | Navigators | 10% |
| Core | No Navigators | 90% |

1. The working assumption is the search index will not be updated in place so there is no need to execute the query load test concurrently with the loading of content.

**Search Features**

1. Nested/scoped queries. This feature provides the ability to search for a specific author, specific affiliation, etc. There are different fields (per content type) that require content scoping. Some fields require additional sub scoping within a scoped query. For example, within a scoped reference query, it is also possible to search for a specific author (within that reference) and even a within a specific author within a reference, to search for a specific given name or surname.
   1. Cores
      1. Auth

Subscopes :

* + - * authlast;
      * authfirstini;
      * authemail
    1. Affil

Subscopes:

* affilcity;
* affilctry;
* affilorg
  + 1. Fundall

Subscopes:

* fundno;
* fundacr;
* fundsponsor
  + 1. Ref

Subscopes :

* + - * refeid;
      * refauth;
        + (sub-subscope authlast; authfirstini)
      * reftitle;
      * langreftitle;
      * refsrctitle;
      * refpubyr;
      * refpubyrtxt;
      * refpg;
      * refpgfirst;
      * refartnum;
      * refscp;
      * refauthid
  1. Author
     1. Affil

Subscopes:

* affilname;
* affilcity;
* affilctry

1. Content/Search normalization (e.g. Stemming, synonyms, etc.)
   1. Equivalence is single direction.
   2. Equivalence for UK/US spelling forms and Greek letters vs. their names
   3. UTF-8 Unicode characters.
   4. Results must display the original (non-equivalenced) text.
   5. A signed URL Scopus synonym list stored in an S3 bucket is provided in the appendix .
   6. Lemmatization is not required. Basic stemming is sufficient. SOLR should use KStem and MarkLogic should use the ‘basic’ stemmer.
2. Complex boolean search support
   1. AND is the default connector.
   2. Use white space and punctuation (non-letter, non-number) for default tokenization.
   3. Support AND/OR/NOT. (in that order of priority)
3. Support Proximity queries
   1. Ordered
   2. Unordered
4. Special field types
   1. Support numeric fields and comparison operators
   2. Support Date content fields with comparison operators.
   3. Support ‘exact phrase’ or ‘punctuation sensitive’ fields specified in the Scopus 13.3 Index Profile.
5. Sorting. Queries (as specified in the XQueryX) can specify ordering of the results based on the following Scopus 13.3 Index Profile fields (or relevance). Both ascending/descending as well as the ability to specify multiple fields must be supported. While the Scopus 13.3 Index Profile might specify additional fields as needing to support sorting, the following fields were actually used in the examined queries.
   1. Cores
      1. Auth
      2. Datesort
      3. Fastloaddate
      4. Itemtitle
      5. Numcitedby
      6. Pubyr
      7. Relevance
   2. Authors
      1. Active
      2. Afdispcity
      3. Afdispctry
      4. Affilsortname
      5. Authid
      6. Count
      7. Preffirst
      8. Preffirstsort
      9. Prefini
      10. Preflast
      11. Relevance
   3. Affiliations
      1. Affilcity
      2. Affilctry
      3. Afid
      4. Count
      5. Sortname
6. Facets. While the Scopus 13.3 Index Profile might specify additional fields for facets, the following fields were actually used in the examined queries. Facet counts must be accurate (across the entire result set) and not estimates. Facet values can be ordered by count as well as alphabetically (those that start with a specific letter or digit). A maximum of 160 values for a facet can be returned. Facet values can also be used to limit/exclude search results.
   1. Cores
      1. Scoaucite
      2. Scoaffilctry
      3. Scoafid
      4. Scoauthid
      5. Scoexactkeyword
      6. Scoexactsrctitle
      7. Scolang
      8. Scoprefnameauid
      9. Scopubyr
      10. Scosrctype
      11. Scostatustype
      12. Scosubjabbr
      13. Scosubtype
   2. Author
      1. Autactivenav
      2. Autaffilcitynav
      3. Autaffilctrynav
      4. Autafnameidnav
      5. Autsrctitle
      6. Autsubjclus
   3. Affiliation
      1. Instaffilcitynav
      2. Instaffilctrynav
7. Return Fields. While the Scopus 13.3 Index Profile might specify additional fields as being returned, the following fields were actually used in the examined queries. For specified fields in the Scopus 13.3 Index Profile, XML markup will need to be maintained (and returned).
   1. Cores
      1. Abs
      2. Absavail
      3. Affil
      4. Affilcity
      5. Affilctry
      6. Afid
      7. Artnum
      8. Auth
      9. Authemail
      10. Authgrpid
      11. Authid
      12. Authkeywords
      13. Authsuff
      14. Coden
      15. Collec
      16. Collecid
      17. Dateloaded
      18. Datesort
      19. Db
      20. Dbdocid
      21. Doi
      22. Dummycode
      23. Dummylink
      24. Eid
      25. Eissn
      26. Fastloaddate
      27. Groupid
      28. Idxterms
      29. Intid
      30. Isbn
      31. Issn
      32. Issnp
      33. Issue
      34. Itemtitle
      35. Loadnum
      36. Numcitedby
      37. Pg
      38. Pgfirst
      39. Pglast
      40. Pii
      41. Pmid
      42. Pub
      43. Pubdatetxt
      44. Pubyr
      45. Refcount
      46. Refeid
      47. Restrictedaccess
      48. Sdeid
      49. Srcid
      50. Srctitle
      51. Srctitleabbr
      52. Srctype
      53. Statustype
      54. Subjabbr
      55. Subtype
      56. Vol
   2. Authors
      1. Active
      2. Afdispcity
      3. Afdispctry
      4. Afdispname
      5. Affilcity
      6. Affilctry
      7. Affilname
      8. Afhistdispname
      9. Afid
      10. Afnameid
      11. Authfirst
      12. Authid
      13. Authlast
      14. Count
      15. Eid
      16. Namevarfirst
      17. Namevarini
      18. Namevarlast
      19. Preffirst
      20. Prefini
      21. Preflast
      22. Pubrangefirst
      23. Pubrangelast
      24. Srctitle
      25. Subjabbr
      26. Suppress
   3. Affiliations
      1. Affilcity
      2. Affilctry
      3. Affilname
      4. Afid
      5. Certscore
      6. Count
      7. Eid
      8. Namevar
      9. Parafid
      10. Prefname
      11. Prefparname
8. Amount of control over search algorithm/results (e.g. Boosting fields, multiple rank algorithms, boosting at query time, ability to modify scoring algorithms, etc.). As mentioned previously in the performance testing section, leverage the specified boosts for core queries. No additional boosting will be required for author or affiliation queries.
9. Highlighting results list. Identifying location of where hits were found in the result list.
   1. Highlighting only needs to be applied when specified in the XQueryX.
   2. Highlighting is only required for the abs field in the returned hit list. (Only core records).
10. SpellCheck with customized dictionary is not required for the PoC.
11. Multi-language - fold case and accents, diacritical marks.
    1. Term Equivalency only applies to English language.
    2. No language customizations have been applied.
    3. Default search engine document and language capabilities are applied.
12. Phrase Searching. Both strict and loose phrase searching needs to be supported.
    1. Loose and exact phrase modes are available only for fields identified as needing it in the Scopus 13.3 Index Profile. Mode selection (loose/exact) if specified in the XQueryX is ignored for other fields.
    2. Loose phrases ignore punctuation, consider stemmed words and term equivalents when applied, and support wildcards. Exact phrases require exact matches including punctuation, and neither stemmed, term equivalents nor wildcards are available. All phrase searches are case-insensitive and ignore accents/diacritics.
13. Wildcarding.
    1. Both single, multi-character.
    2. Multiple wildcards per query/term
    3. Prefix/infix/postfix
    4. Honored in ”loose” phrase searching.
14. Boundary Matching – starts with, ends with, equals. Case insensitive. While indicated this is needed across the various content types (and various fields), the examined queries only show this is used in the ‘author’ content type and only in the fields for author given name and surname. Subsequently, the PoC only needs to support boundary matching for author given name (and surname) within the ‘author’ content type.
15. Formal relevance testing will not be required.
16. More Like This functionality. Within the ‘core’ cluster, a ‘more like this’ functionality is provided for references, keywords, and authors. For references and authors, a query consisting of the refid and authid values are used. For keywords, a query consisting of the keywords is used. Examples of these queries are contained in the load query set.
17. A list of queries to validate all functionality and counts (for results and facets) will be provided and executed manually. This will minimally include nested/scope queries and proximity queries. Due to stemming algorithm differences, it is unlikely counts will be exact when comparisons are made across search engines. A signed URL for the functionality and count queries is provided in the appendix.
18. Query length needs to support up to 64K at a minimum.
19. Client API exposed /ease of integration with existing product platforms via REST endpoint.
20. Search debug/tuning capabilities (why a search is returning what it does).
21. A signed URL for he Scopus stopword list is provided in the appendix. Stopwords should be applied to specified fields identified in the in the Scopus 13.3 Index Profile (with the exception of the punctuation sensitive versions of the fields).

**Benchmarking**

1. Based on harvested production queries for cores, authors, and affiliations.
2. The queries provided in XQueryX should be transformed into the native search syntax and stored in a scalable repository (such as Amazon DynamoDB). For each content type, the queries should be stored in a separate table with the ‘key’ identifying the query. The key consists of the following 3 components:

* Content type (core, auth, or affil) for the query.
* Each content type can have the following subtypes
  + affil (afid, user)
  + auth (auid,auidafid,user)
  + core (doi,eid,sdeid,refeid,auid,afid,issnisbn,collecid,refeidneid,srcid,user)
* Offset within the respective XQueryX file for the content type and subtype.

So, the 25th query in the ‘affiliation’ content type and subtype ‘afid’ set would have a key of ‘affil\_afid\_25’. The 33rd key for the ‘core’ content type and subtype of ‘doi’ set would have a key of ‘core\_doi\_33’.

1. Each load test will run 3 times to cover variability found in AWS deployment.
2. The Load Test framework will issue REST requests to a Node endpoint. This request will contain the ‘key’ for the query to execute. The Node endpoint will retrieve the query from DynamoDB and POST the query to the search engine endpoint. Both the Node endpoint and the search engine will reside within the same AWS region (and AWS zone).
3. It is permissible to run some queries to “prime the caches” prior to beginning the performance testing runs.

Appendix



**S3 buckets**

abstract-xml (abstract/core records)

affiliation-xml (affiliation records)

author-xml (author records)

**Signed URLs for the keys in the S3 buckets**

abstract-xml keys

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/abskeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399855460&Signature=Yi1O4gwSEJSfQvEbEuhh3A2I%2BE8%3D>

affiliation-xml keys

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/affilkeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399855561&Signature=PTeMwRYOTYYWKkjFViBhBXG4eDE%3D>

author-xml keys

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/authkeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399855624&Signature=7gvoVIMVt8Eo2ZBoj9h14ME45qk%3D>

**Signed URLs for the XQueryX query sets**

affiliation (afid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimInstitutionAfid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400194114&Signature=XLzpEUXLwMhsa1jS29wrL9fAHJQ%3D>

affiliation (user)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimInstitutionUser.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400194179&Signature=XqVwTfKJ6SAPIalJYuP%2BaVZaRU0%3D>

author (auid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimAuthorAuid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400194242&Signature=AGOybA6tdH0D9U0VAoM0%2BJ/fl1g%3D>

author (auidafid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimAuthorAuidafid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400449189&Signature=dYqc6jKS3zOp537c8YZXaRwx9Ws%3D>

author (user)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimAuthorUser.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400194319&Signature=XRlETN2sdsFfRUXM%2BnMtCoQugzY%3D>

core (doi)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainDoi.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207021&Signature=CLL2Wr%2BUQLgjgU3gYYXjxyzXAuQ%3D>

core (eid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainEid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207053&Signature=cwq/z6%2B2J3fqFUaKIqjw%2Bge2vqw%3D>

core (sdeid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainSdeid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207076&Signature=uRXHQmCYvu874G1kNmzLMExKWsQ%3D>

core (refeid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainRefeid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207102&Signature=VBr2hJ5awp6iYF5DMsdc0QF4ZeU%3D>

core (auid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainAuid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207130&Signature=kqmr6ZCZZimDH6GQmvh0k8VrQ5c%3D>

core (afid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainAfid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207191&Signature=GGb1Hg7QEYhKP1wOPvTx9E4UKwI%3D>

core (issnisbn)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainIssnisbn.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207218&Signature=KfCIacxwOHAtEamWmb9JzyeQU7k%3D>

core (collecid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainCollecid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207245&Signature=TBTUQVjltvpu6MwXyRqhD0Uj1u4%3D>

core (refeidneid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainRefeidneid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207281&Signature=ecxyugkJGF5gCaoXwmWneqdvm9A%3D>

core (srcid)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainSrcid.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207304&Signature=bmtyInTzjKnSR/QNlDuCxcCcHcQ%3D>

core (scopout)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainScopout.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1430637777&Signature=F3nsAV9xcB0DLK4zBkNKO%2BfJRuE%3D>

core (user)

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/dec052013/slimMainUser.txt.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400207329&Signature=qpJlKuw9anV8t8N8K8W2bnMNCdI%3D>

**Signed URLs for the Load test keys**

affiliation (afid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/affiliationAfidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400194914&Signature=wMPnCFi0sJuQXd6uDZ7/WLGuLV8%3D>

affiliation (user)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/affiliationUserLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400195039&Signature=tbmBNEKgGfX6VplfH8Ksi/Sn8bo%3D>

author (auid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/authorAuidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400195770&Signature=V%2BySWq/fwZJULbOseYGzVrPjO/E%3D>

author (auidafid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/authorAuidafidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400449695&Signature=2ihsKUtoAkOdr1jzFh6myPxECuE%3D>

author (user)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/authorUserLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400195818&Signature=qmGV9yFvRx0zmMLLh6wcxZjvE6A%3D>

core (doi)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreDoiLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210540&Signature=YztHTMioDkCL2XRGTFfRzvg9b70%3D>

core (eid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreEidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210569&Signature=96OuyAM3v104GNXjSUun%2BtcYMfc%3D>

core (sdeid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreSdeidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210590&Signature=vhNl4u6IHt7yGgJNAeNHP4Mqol0%3D>

core (refeid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreRefeidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210629&Signature=2fO%2BF0aU1xR0aggg/O0JqTLH0mc%3D>

core (auid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreAuidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210654&Signature=%2Bn3xcsuTg8roxK1jPl%2BSUxNcofA%3D>

core (afid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreAfidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210675&Signature=rQJXuLt7mqWA9p8NIS2WJ%2Bxni1c%3D>

core (issnisbn)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreIssnisbnLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210699&Signature=rTYFrUR4H3SqXJMS2dm8d9IG%2B00%3D>

core (collecid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreCollecidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210723&Signature=OejFqUDvNqNELUzSVWfq5a1dT48%3D>

core (refeidneid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreRefeidneidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210741&Signature=yh6ja2iXsRkg/eiZz0tafhPEX7k%3D>

core (srcid)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreSrcidLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210761&Signature=3FCyx0cxdgjFqVzwUNroGNYpkKI%3D>

core (scopout)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreScopoutLoadQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1430552406&Signature=raMyArMuWkq39RZBCseQ27Rg0bo%3D>

core (user)

<https://els-ats.s3.amazonaws.com/scopusSolr/node/loadTestKeySets/coreUserQueryKeys.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400210780&Signature=XtT1Sg2VnBF/NCIVlvDgxirBz7w%3D>

**XSLT Stylesheets (for generating queries in SOLR syntax)**

core

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/stylesheets/CoreXqueryX2Solr.xsl?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399859156&Signature=dzfv0urbcN5Evk2OyVNAVEH3cgs%3D>

affiliation

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/stylesheets/AffiliationXqueryX2Solr.xsl?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399859279&Signature=br/dvZNVafpSzV%2B/5kSk0jMB5as%3D>

author

<https://els-ats.s3.amazonaws.com/scopusSolr/querysets/stylesheets/AuthorXqueryX2Solr.xsl?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399860164&Signature=Q3DHjy1%2BUO/QohOF9Vt2wrG3IYw%3D>

**Signed URLs for stopwords and synonyms**

stopwords

<https://els-ats.s3.amazonaws.com/scopusSolr/solrConfig/common/stopwords.txt?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399861559&Signature=3GTpeVw7dFRk2ypOkJeqSQxPlBw%3D>

synonyms

<https://els-ats.s3.amazonaws.com/scopusSolr/solrConfig/common/synonyms.txt?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399861527&Signature=j9TLxjjSGxmehD9tB%2BQIYWMAcEE%3D>

**Signed URL for Scopus 13.3 Index Profile**

<https://els-ats.s3.amazonaws.com/scopusSolr/Scopus%20SC13-3%20V52%20Draft.xls?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399860505&Signature=dY04KSvh6l7A672XASblv1RjVZ4%3D>

**Signed URL for Scopus Feature/Count Queries**

<https://els-ats.s3.amazonaws.com/scopusSolr/ScopusHotHouseFeatureCountQueries.xls?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400551001&Signature=pu82ZJFyFTPsMi2Xsk6HIFg9vdo%3D>

**Signed URLs for content update keys**

core (100%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/abskeys1x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399866083&Signature=Qoj0pTLEOevzji9XMlAnaRoc4YM%3D>

affiliation (100%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/affilkeys1x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399866059&Signature=QifUmgsdAJv/EddIkd7nZBz%2BM6g%3D>

author (100%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/authkeys1x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399866030&Signature=iXRQhaDwrGtvYFbhce0rXH7eXp8%3D>

core (300%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/abskeys3x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399866209&Signature=2NQm0%2BYgwxXBAG71RSlpS2Gk5a4%3D>

affiliation (300%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/affilkeys3x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399866177&Signature=HGQzo9USXj88vUiWBEHgMy5mpDY%3D>

author (300%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/authkeys3x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1399866154&Signature=g0KqrG7NsUuO0Eo0o0ldjjC1rl4%3D>

Author 2 Million Update

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/authupdkeys2m.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1431091733&Signature=cwsDSxyZBtrb8lHZsoyZdpqkikk%3D>

**Signed URLs for publication and cited by update keys**

author publication count (100%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/authpckeys1x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1431091789&Signature=ytihc0TJJ9a7zMxFPz%2BAW3NJK4c%3D>

affiliation publication count (100%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/affilpckeys1x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1431091935&Signature=75TIHrIB%2BQPgeTkXvAu/LDHBUTU%3D>

core cited by count (100%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/abskeyscbc1x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400041962&Signature=9jGPeAQi3CF8Zae1EXSml5h0H24%3D>

author publication count (300%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/authpckeys3x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1431091981&Signature=oyyNcbvktXusLP8%2BxungA/85S/8%3D>

affiliation publication count (300%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/affilpckeys3x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1431092037&Signature=XCpjl5bbtSkT3HcWD7lm7DVwwzE%3D>

core cited by count (300%)

<https://els-ats.s3.amazonaws.com/scopusSolr/keys/abskeyscbc3x.gz?AWSAccessKeyId=AKIAIQ2VDFJYKESDOTUQ&Expires=1400042000&Signature=xdpxoglXi8OM4qE5/NrK%2BBGkOdw%3D>