Intellectual Property Rights Notice for Open Specifications Documentation

- **Technical Documentation.** Microsoft publishes Open Specifications documentation for protocols, file formats, languages, standards as well as overviews of the interaction among each of these technologies.

- **Copyrights.** This documentation is covered by Microsoft copyrights. Regardless of any other terms that are contained in the terms of use for the Microsoft website that hosts this documentation, you may make copies of it in order to develop implementations of the technologies described in the Open Specifications and may distribute portions of it in your implementations using these technologies or your documentation as necessary to properly document the implementation. You may also distribute in your implementation, with or without modification, any schema, IDL's, or code samples that are included in the documentation. This permission also applies to any documents that are referenced in the Open Specifications.

- **No Trade Secrets.** Microsoft does not claim any trade secret rights in this documentation.

- **Patents.** Microsoft has patents that may cover your implementations of the technologies described in the Open Specifications. Neither this notice nor Microsoft's delivery of the documentation grants any licenses under those or any other Microsoft patents. However, a given Open Specification may be covered by Microsoft Open Specification Promise or the Community Promise. If you would prefer a written license, or if the technologies described in the Open Specifications are not covered by the Open Specifications Promise or Community Promise, as applicable, patent licenses are available by contacting iplg@microsoft.com.

- **Trademarks.** The names of companies and products contained in this documentation may be covered by trademarks or similar intellectual property rights. This notice does not grant any licenses under those rights.

- **Fictitious Names.** The example companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted in this documentation are fictitious. No association with any real company, organization, product, domain name, email address, logo, person, place, or event is intended or should be inferred.

**Reservation of Rights.** All other rights are reserved, and this notice does not grant any rights other than specifically described above, whether by implication, estoppel, or otherwise.

**Tools.** The Open Specifications do not require the use of Microsoft programming tools or programming environments in order for you to develop an implementation. If you have access to Microsoft programming tools and environments you are free to take advantage of them. Certain Open Specifications are intended for use in conjunction with publicly available standard specifications and network programming art, and assumes that the reader either is familiar with the aforementioned material or has immediate access to it.
## Revision Summary

<table>
<thead>
<tr>
<th>Date</th>
<th>Revision History</th>
<th>Revision Class</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>02/19/2010</td>
<td>1.0</td>
<td>Major</td>
<td>Initial Availability</td>
</tr>
<tr>
<td>03/31/2010</td>
<td>1.01</td>
<td>Editorial</td>
<td>Revised and edited the technical content</td>
</tr>
<tr>
<td>04/30/2010</td>
<td>1.02</td>
<td>Editorial</td>
<td>Revised and edited the technical content</td>
</tr>
<tr>
<td>06/07/2010</td>
<td>1.03</td>
<td>Editorial</td>
<td>Revised and edited the technical content</td>
</tr>
<tr>
<td>06/29/2010</td>
<td>1.04</td>
<td>Editorial</td>
<td>Changed language and formatting in the technical content.</td>
</tr>
<tr>
<td>07/23/2010</td>
<td>1.04</td>
<td>No change</td>
<td>No changes to the meaning, language, or formatting of the technical content.</td>
</tr>
<tr>
<td>09/27/2010</td>
<td>1.04</td>
<td>No change</td>
<td>No changes to the meaning, language, or formatting of the technical content.</td>
</tr>
<tr>
<td>11/15/2010</td>
<td>1.04</td>
<td>No change</td>
<td>No changes to the meaning, language, or formatting of the technical content.</td>
</tr>
<tr>
<td>12/17/2010</td>
<td>1.04</td>
<td>No change</td>
<td>No changes to the meaning, language, or formatting of the technical content.</td>
</tr>
<tr>
<td>03/18/2011</td>
<td>1.04</td>
<td>No change</td>
<td>No changes to the meaning, language, or formatting of the technical content.</td>
</tr>
<tr>
<td>06/10/2011</td>
<td>1.04</td>
<td>No change</td>
<td>No changes to the meaning, language, or formatting of the technical content.</td>
</tr>
<tr>
<td>01/20/2012</td>
<td>1.5</td>
<td>Minor</td>
<td>Clarified the meaning of the technical content.</td>
</tr>
<tr>
<td>04/11/2012</td>
<td>1.5</td>
<td>No change</td>
<td>No changes to the meaning, language, or formatting of the technical content.</td>
</tr>
<tr>
<td>07/16/2012</td>
<td>1.5</td>
<td>No change</td>
<td>No changes to the meaning, language, or formatting of the technical content.</td>
</tr>
</tbody>
</table>
Table of Contents

1 Introduction .................................................................................................................. 6
   1.1 Glossary ................................................................................................................. 6
   1.2 References ............................................................................................................. 6
      1.2.1 Normative References ...................................................................................... 6
      1.2.2 Informative References ................................................................................... 7
   1.3 Structure Overview (Synopsis) ............................................................................... 7
   1.4 Relationship to Protocols and Other Structures ..................................................... 7
   1.5 Applicability Statement ......................................................................................... 7
   1.6 Versioning and Localization ................................................................................... 7
   1.7 Vendor-Extensible Fields ....................................................................................... 7

2 Structures ...................................................................................................................... 8
   2.1 Common File Structures ....................................................................................... 8
   2.2 Common File Naming Conventions ..................................................................... 9
   2.3 Search Clickthrough Files .................................................................................... 10
      2.3.1 <date>.clicks.txt ........................................................................................... 10
      2.3.2 <date>.queries.txt ......................................................................................... 10
      2.3.3 <date>.urls.txt ............................................................................................. 10
   2.4 Analysis Files ........................................................................................................ 10
      2.4.1 allfeeduris.<sf> files ....................................................................................... 10
      2.4.2 cid_by_cid_with_counts_and_query.<sf> ......................................................... 11
      2.4.3 cid_by_cid_with_counts_and_query_mergereduc.<sf> ................................... 11
      2.4.4 global_querycnt_by_query.<sf> .................................................................... 11
      2.4.5 local_querycnt_by_cid.<sf> ......................................................................... 11
      2.4.6 local_querycnt_by_cid_merge.<sf> ................................................................. 11
      2.4.7 local_querycnt_by_query.<sf> ..................................................................... 11
      2.4.8 local_querycnt_by_query_reduce.<sf> .............................................................. 12
      2.4.9 local_querycnt_by_url.<sf> ......................................................................... 12
      2.4.10 local_querycnt_by_url_merge.<sf> ................................................................. 12
      2.4.11 local_querycnt_by_urlid.<sf> ..................................................................... 12
      2.4.12 local_querycnt_by_urlid_reduce.<sf> .............................................................. 13
      2.4.13 semi_local_querycnt_by_queryid.<sf> ............................................................ 13
      2.4.14 semi_local_querycnt_by_urlid.<sf> ................................................................. 13
      2.4.15 semi_local_querycnt_by_urlid_map.<sf> ......................................................... 13
      2.4.16 semi_local_querycnt_pre_token.<sf> ............................................................... 13
      2.4.17 uris_by_contentid_ts.<sf> ............................................................................ 13
      2.4.18 uris_by_member.<sf> ...................................................................................... 14
      2.4.19 uris_by_member_reduce.<sf> ....................................................................... 14
      2.4.20 urls_by_urlhash_with_queries.<sf> ............................................................... 14
      2.4.21 urls_by_urlhash_with_queries_sort.<sf> ....................................................... 14
      2.4.22 urls_on_urlhash_with_queries.<sf> ............................................................... 14
      2.4.23 <col>_feeduris.<sf> ....................................................................................... 15
      2.4.24 <col>_feeduris_expanded.<sf> ...................................................................... 15
      2.4.25 <col>_feeduris_expanded_resplit.<sf> ............................................................ 15
      2.4.26 <gen>.queries_by_queryid.<sf> .................................................................. 15
      2.4.27 <gen>.queries_by_queryid_all.<sf> ................................................................. 15
      2.4.28 <gen>.queryinfo.<sf> .................................................................................. 15
      2.4.29 <gen>.urls_by_urlid.<sf> .............................................................................. 16
      2.4.30 <gen>.urls_by_urlid_all.<sf> ....................................................................... 16
      2.4.31 <gen>.<col>.unique_uris_by_uri.<sf> ............................................................... 16

[MS-FSSPRDF] — v20120630
SPRel Data File Formats

Copyright © 2012 Microsoft Corporation.

Release: July 16, 2012
3 Structure Examples ........................................................................................................................................ 21
3.1 Search Clickthrough Files .................................................................................................................. 21
3.1.1 <date>.clicks.txt files ..................................................................................................................... 21
3.1.2 <date>.queries.txt files .................................................................................................................. 21
3.1.3 <date>.urls.txt files ........................................................................................................................ 21
3.2 Analysis Files ....................................................................................................................................... 22
3.2.1 allfeeduris.<sf> files ........................................................................................................................ 22
3.2.2 cid_by_cid_with_counts_and_query.<sf> ......................................................................................... 22
3.2.3 cid_by_cid_with_counts_and_query_merge.<sf> ........................................................................... 22
3.2.4 global_querycnt_by_query.<sf> ....................................................................................................... 22
3.2.5 local_querycnt_by_cid.<sf> ............................................................................................................. 23
3.2.6 local_querycnt_by_cid_merge.<sf> .................................................................................................. 23
3.2.7 local_querycnt_by_query.<sf> ......................................................................................................... 23
3.2.8 local_querycnt_by_query_reduce.<sf> ............................................................................................ 23
3.2.9 local_querycnt_by_url.<sf> ............................................................................................................ 24
3.2.10 local_querycnt_by_url_merge.<sf> ............................................................................................... 24
3.2.11 local_querycnt_by_urlid.<sf> ........................................................................................................ 24
3.2.12 local_querycnt_by_urlid_reduce.<sf> ........................................................................................... 24
3.2.13 semi_local_querycnt_by_queryid.<sf> ........................................................................................... 24
3.2.14 semi_local_querycnt_by_urlid.<sf> ............................................................................................... 25
3.2.15 semi_local_querycnt_by_urlid_map.<sf> .................................................................................... 25
3.2.16 semi_local_querycnt_pre_token.<sf> ............................................................................................ 25
3.2.17 uris_by_contentid_ts.<sf> .............................................................................................................. 26
3.2.18 uris_by_member.<sf> ...................................................................................................................... 26
3.2.19 uris_by_member_reduce.<sf> .......................................................................................................... 26
3.2.20 urls_by_urlhash_with_queries.<sf> .............................................................................................. 26
3.2.21 urls_by_urlhash_with_queries_sort.<sf> ..................................................................................... 27
3.2.22 urls_on_urlhash_with_queries.<sf> .............................................................................................. 27
3.2.23 <col>_feeduris.<sf> ....................................................................................................................... 28
3.2.24 <col>_feeduris_expanded.<sf> ....................................................................................................... 28
3.2.25 <col>_feeduris_expanded_resplit.<sf> .......................................................................................... 28

Copyright © 2012 Microsoft Corporation.

Release: July 16, 2012
3.2.26  <gen>.queries_by_queryid.<sf> ................................................................. 28
3.2.27  <gen>.queries_by_queryid_all.<sf> ............................................................ 29
3.2.28  <gen>.queryinfo.<sf> .............................................................................. 29
3.2.29  <gen>.urls_by_urlid.<sf> ........................................................................ 30
3.2.30  <gen>.urls_by_urlid_all.<sf> .................................................................. 30
3.2.31  <gen>.<col>.unique_uris_by_url.<sf> ......................................................... 30
3.2.32  <gen>.<col>.uris_by_contentid.<sf> ......................................................... 31
3.2.33  <gen>.<col>.uris_by_contentid_ts.<sf> ..................................................... 31
3.2.34  <gen>.<col>.<host>_contentids_by_contentid_new.<sf> ........................... 31
3.2.35  <gen>.<col>.<host>_contentids_by_contentid_new_resplit.<sf>..<sf> ............ 31
3.2.36  <gen>.<col>.<host>_uris.0 ........................................................................ 32
3.2.37  <date>.clicks_by_urlid_and_queryid.<sf> ............................................... 32
3.2.38  <date>.clicks_by_urlid_and_queryid_sort.<sf><sf> .................................... 32
3.2.39  <date>.clicks_on_queryid.0 .................................................................. 32
3.2.40  <date>.local_querycnt_by_queryid.<sf> ................................................... 33
3.2.41  <date>.local_querycnt_by_queryid_reduce.<sf><sf> .................................... 33
3.2.42  <date>.queries_by_queryid.<sf> ............................................................... 33
3.2.43  <date>.queries_by_queryid_sort.0.<sf> ...................................................... 33
3.2.44  <date>.queries_on_queryid.0 ................................................................ 34
3.2.45  <date>.urls_by_urlid.<sf> ..................................................................... 34
3.2.46  <date>.urls_by_urlid_sort.0.<sf> ............................................................... 34
3.2.47  <date>.urls_on_urlid.0 ........................................................................... 34
3.3 Database Files............................................................................... 35
3.3.1  <gen>.sharepoint.rel.<part_num>.bin ......................................................... 35
3.3.2  <gen>.sharepoint.rel.<part_num>.idx ......................................................... 36
3.3.3  <gen>.sharepoint.rel.<part_num>.idx.ofs ................................................ 37

4 Security Considerations...................................................................... 38
5 Appendix A: Product Behavior.......................................................... 39
6 Change Tracking............................................................................... 40
7 Index ............................................................................................. 41
1 Introduction

This document specifies the SPRel Data File Formats used for client click input analysis. This analysis improves search relevancy by analyzing client feedback.

Sections 1.7 and 2 of this specification are normative and can contain the terms MAY, SHOULD, MUST, MUST NOT, and SHOULD NOT as defined in RFC 2119. All other sections and examples in this specification are informative.

1.1 Glossary

The following terms are defined in [MS-GLOS]:

Augmented Backus-Naur Form (ABNF)
Coordinated Universal Time (UTC)
little-endian
MD5 hash

The following terms are defined in [MS-OFCGLOS]:

absolute URL
base64 encoding
content collection
dictionary
document identifier
equivalence class
file
search clickthrough

The following terms are specific to this document:

MAY, SHOULD, MUST, SHOULD NOT, MUST NOT: These terms (in all caps) are used as described in [RFC2119]. All statements of optional behavior use either MAY, SHOULD, or SHOULD NOT.

1.2 References

References to Microsoft Open Specifications documentation do not include a publishing year because links are to the latest version of the technical documents, which are updated frequently. References to other documents include a publishing year when one is available.

1.2.1 Normative References

We conduct frequent surveys of the normative references to assure their continued availability. If you have any issue with finding a normative reference, please contact dochelp@microsoft.com. We will assist you in finding the relevant information. Please check the archive site, http://msdn2.microsoft.com/en-us/library/E4BD6494-06AD-4aed-9823-445E921C9624, as an additional source.

[MS-FSWASDR] Microsoft Corporation, "WebAnalyzer/SPRel Data Receiving Protocol Specification".


[MS-LCID] Microsoft Corporation, "Windows Language Code Identifier (LCID) Reference".
1.2.2 Informative References
[MS-FSFSDF] Microsoft Corporation, "FAST Distributed Make Worker Protocol Specification".
[MS-OFCGLOS] Microsoft Corporation, "Microsoft Office Master Glossary".

1.3 Structure Overview (Synopsis)
These files contain information for search clickthrough log analysis. Search clickthrough log analysis computes search ranks for documents. These logs are produced on a daily basis. A log contains information about how often and when a search query is processed. It also provides information about document retrieval, clicks, for a specific search query. This information is used as input to compute a search rank for the documents.

1.4 Relationship to Protocols and Other Structures
The file formats in this document are used by the protocol described in [MS-FSFSDF].

1.5 Applicability Statement
None.

1.6 Versioning and Localization
None.

1.7 Vendor-Extensible Fields
None.
2 Structures

This chapter specifies the directory structure, file names and internal data format for the files used for search clickthrough log analysis.

2.1 Common File Structures

A file MUST either be empty or contain a set of rows. Each row consists of one or more columns and is terminated by a newline. A newline is either a carriage return in combination with a line feed or only a line feed. Columns MUST be separated by a white space delimiter. If the column does not contain binary data, the columns MUST be ASCII encoded.

The common structure for a file with no binary data corresponds to the following rules written in Augmented Backus-Naur Form (ABNF), as specified in [RFC5234].

FILE = *LINE  
LINE = ROW NEWLINE  
BASE64 = 1*(ALPHA / DIGIT / "=" / "+" / "/")  
TOKEN = 1*(%x21-%ff)  
CID = 1*TOKEN  
CLICKEDURLRANK = 1*DIGIT  
COLLECTION = 1*TOKEN  
DATE = 4DIGIT "." 2DIGIT "." 2DIGIT  
GCOUNT = 1*DIGIT  
URLHASH = 1*DIGIT  
NUMURLS = 1*DIGIT  
LCID = 1*DIGIT  
LCOUNT = 1*DIGIT  
MEMBER = 1*(%x21-%ff)  
OP = "add" / "del"  
QUERIES = 1*BASE64  
QUERIESINFO = 1*BASE64  
QUERY = *(TOKEN SP) TOKEN  
QUERYID = 1*DIGIT  
TIMESTAMP = 1*DIGIT  
URL = 1*(%x21-%ff)  
URLID = 1*DIGIT  
PLACE = 1*DIGIT  
ZERO = "0"  
NEWLINE = (CRLF / LF)

Some of the ABNF rules are specified in more detail in the following table.

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CID</td>
<td>Specifies a document identifier(3).</td>
</tr>
<tr>
<td>CLICKEDURLRANK</td>
<td>Specifies the rank of a document in the search result.</td>
</tr>
<tr>
<td>COLLECTION</td>
<td>Specifies a content collection name.</td>
</tr>
<tr>
<td>DATE</td>
<td>Specifies a processing date. This date MUST be formatted as yyyy.mm.dd.</td>
</tr>
<tr>
<td>GCOUNT</td>
<td>Specifies the number of times a search query was processed.</td>
</tr>
<tr>
<td>Column</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>URLHASH</td>
<td>Specifies a 128-bit MD5 hash value of a URL. The hash value is represented as a numerical value.</td>
</tr>
<tr>
<td>NUMURLS</td>
<td>Specifies the total number of URLs associated with a specific search query.</td>
</tr>
<tr>
<td>LCID</td>
<td>Specifies a language code identifier, as specified in [MS-LCID].</td>
</tr>
<tr>
<td>LCOUNT</td>
<td>Specifies the number of times a URL was clicked for a particular search query.</td>
</tr>
<tr>
<td>MEMBER</td>
<td>Specifies a URL or document identifier that is in the equivalence class of the document. The MEMBER column MUST contain the value 0xc782, as specified section 2.2.1.3.2.1.4 in [MS-FSWASDR], when the OP column contains a DEL operation.</td>
</tr>
<tr>
<td>OP</td>
<td>Specifies a content submission operation. Values are:</td>
</tr>
<tr>
<td></td>
<td>- ADD: A new document was added to the index.</td>
</tr>
<tr>
<td></td>
<td>- DEL: A document was deleted from the index.</td>
</tr>
</tbody>
</table>
| QUERIES      | Specifies all search queries associated with a particular document identifier. This column contains a dictionary encoded with base64 encoding, as specified in [MS-FSWCU]. This dictionary MUST contain two keys: contentid and queries.  
|              | The contentid key specifies a document identifier (CID). The value of the "queries" key MUST contain an array, as specified in [MS-FSWCU], of query entries. Each query entry MUST be a five entry tuple, as specified in [MS-FSWCU]. Each tuple MUST contain the following column values; LCOUNT, GCOUNT, PLACE, NUMURLS and QUERY. |
| QUERIESINFO  | Specifies search queries and collections associated with a particular document identifier. This column contains a dictionary encoded with base64 encoding as specified for the QUERIES column. However, this dictionary MUST have one additional key, the collections key.  
|              | The collections key MUST contain an array of all the collections, as specified in [MS-FSWCU], where this document identifier is registered.         |
| QUERY        | Specifies an ASCII encoded search query string.                                                                                             |
| QUERYID      | Specifies a unique identifier for a search query.                                                                                            |
| TIMESTAMP    | Specifies a processing timestamp. This is a numerical value, it specifies the time in seconds after 00:00:00 1970-01-01 UTC.                      |
| URL          | Specifies an absolute URL.                                                                                                                   |
| URLID        | Specifies an unique identifier for a URL.                                                                                                |
| PLACE        | Specifies a numerical rank. This field ranks URLs for a specific search query.                                                              |
| ZERO         | This column is reserved, and MUST contain the ASCII character "0".                                                                           |

2.2 Common File Naming Conventions

Some files do not have a fixed file name. Parts of these file names have one or more variables, specified as follows.

- **date:** Specifies the date when this file was processed. The date format MUST be yyyymmmdd.
• **gen**: Specifies the generation of this file. This value begins at 0 and increases for each new version of the file.

• **sf**: Specifies the split factor number of a file. The split factor specifies the number of files into which to split the data. The value MUST be a number between 0 and up to the maximum split factor.

• **part_num**: Specifies a dataset partition number. This number is retrieved from the configuration file of the process.

• **col**: Specifies a collection name used by search clickthrough log analysis process.

• **host**: Specifies the worker hostname. The hostname is specified in the configuration file of the process.

### 2.3 Search Clickthrough Files

The files contain daily information from the search clickthrough logs. These files are used as input for search clickthrough analysis log process. For more information about ABNF parameters, see section 2.1.

#### 2.3.1 `<date>.clicks.txt`

File of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = QUERYID SP URLID SP LCID SP CLICKEDURLRANK
```

For more information about ABNF parameters, see section 2.1.

#### 2.3.2 `<date>.queries.txt`

Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = QUERYID SP DATE SP QUERY
```

For more information about ABNF parameters, see section 2.1.

#### 2.3.3 `<date>.urls.txt`

Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URLID SP DATE SP URL
```

For more information about ABNF parameters, see section 2.1.

### 2.4 Analysis Files

The analysis files consist of intermediate computation files of the search clickthrough log analysis process. For more information about ABNF parameters, see section 2.1.

#### 2.4.1 allfeeduris.<sf> files

Files of this type MUST contain rows that are specified by the following ABNF rule.
ROW = CID SP COLLECTION

For more information about ABNF parameters, see section 2.1.

2.4.2  cid_by_cid_with_counts_and_query.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = CID SP LCOUNT SP GCOUNT SP PLACE SP NUMURLS SP QUERY

For more information about ABNF parameters, see section 2.1.

2.4.3  cid_by_cid_with_counts_and_query_mergereduce.<sf>.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = CID SP LCOUNT SP GCOUNT SP PLACE SP NUMURLS SP QUERY

For more information about ABNF parameters, see section 2.1.

2.4.4  global_querycnt_by_query.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = GCOUNT SP QUERY

Rows MUST be sorted by the QUERY column.

For more information about ABNF parameters, see section 2.1.

2.4.5  local_querycnt_by_cid.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = CID SP LCOUNT SP QUERY

Rows MUST be sorted by the CID column.

For more information about ABNF parameters, see section 2.1.

2.4.6  local_querycnt_by_cid_merge.<sf>.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = CID SP LCOUNT SP QUERY

Rows MUST be sorted by the CID column.

For more information about ABNF parameters, see section 2.1.
2.4.7 **local_querycnt_by_query.<sf>**

Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = CID SP LCOUNT SP QUERY
```

Rows MUST be sorted by the **QUERY** column.

For more information about ABNF parameters, see section 2.1.

2.4.8 **local_querycnt_by_query_reduce.<sf>.<sf>**

Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = CID SP LCOUNT SP QUERY
```

Rows MUST be sorted by the **QUERY** column.

For more information about ABNF parameters, see section 2.1.

2.4.9 **local_querycnt_by_url.<sf>**

Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URL SP LCOUNT SP QUERY
```

Rows MUST be sorted by the **URL** column.

For more information about ABNF parameters, see section 2.1.

2.4.10 **local_querycnt_by_url_merge.<sf>.<sf>**

Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URL SP LCOUNT SP QUERY
```

Rows MUST be sorted by the **URL** column.

For more information about ABNF parameters, see section 2.1.

2.4.11 **local_querycnt_by_urlid.<sf>**

Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URLID SP LCOUNT SP QUERY
```

Rows MUST be sorted by the **URLID** column.

For more information about ABNF parameters, see section 2.1.
2.4.12  **local_querycnt_by_urlid_reduce.<sf>.<sf>**
Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URLID SP LCOUNT SP QUERY
```

Rows MUST be sorted by the `URLID` column.
For more information about ABNF parameters, see section 2.1.

2.4.13  **semi_local_querycnt_by_queryid.<sf>**
Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URLID SP QUERYID SP LCID SP LCOUNT
```

Rows MUST be sorted by the `QUERYID` column.
For more information about ABNF parameters, see section 2.1.

2.4.14  **semi_local_querycnt_by_urlid.<sf>**
Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URLID SP LCOUNT SP QUERY
```

Rows MUST be sorted by the `URLID` column.
For more information about ABNF parameters, see section 2.1.

2.4.15  **semi_local_querycnt_by_urlid_map.<sf>.<sf>**
Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URLID SP LCOUNT SP QUERY
```

Rows MUST be sorted by the `URLID` column.
For more information about ABNF parameters, see section 2.1.

2.4.16  **semi_local_querycnt_pre_token.<sf>**
Files of this type MUST contain rows that are specified by the following ABNF rule.

```
ROW = URLID SP LCOUNT SP LCID SP QUERY
```

For more information about ABNF parameters, see section 2.1.

2.4.17  **uris_by_contentid_ts.<sf>**
Files of this type MUST contain rows that are specified by the following ABNF rule.
ROWS MUST be sorted by the CID column.
For more information about ABNF parameters, see section 2.1.

2.4.18 uris_by_member.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = CID SP MEMBER

Rows MUST be sorted by the MEMBER column.
For more information about ABNF parameters, see section 2.1.

2.4.19 uris_by_member_reduce.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = CID SP MEMBER

Rows MUST be sorted by the MEMBER column.
For more information about ABNF parameters, see section 2.1.

2.4.20 urls_by_urlhash_with_queries.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = URLHASH SP QUERIES

For more information about ABNF parameters, see section 2.1.

2.4.21 urls_by_urlhash_with_queries_sort.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = URLHASH SP QUERIES

For more information about ABNF parameters, see section 2.1.

2.4.22 urls_on_urlhash_with_queries.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = URLHASH SP QUERIES

For more information about ABNF parameters, see section 2.1.
2.4.23  <col>_feeduris.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URL SP COLLECTION

For more information about ABNF parameters, see section 2.1.

2.4.24  <col>_feeduris_expand.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URL SP COLLECTION

For more information about ABNF parameters, see section 2.1.

2.4.25  <col>_feeduris_expand_resplit.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URL SP COLLECTION

For more information about ABNF parameters, see section 2.1.

2.4.26  <gen>.queries_by_queryid.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = QUERYID SP DATE SP QUERY

Rows MUST be sorted by the QUERYID column.
For more information about ABNF parameters, see section 2.1.

2.4.27  <gen>.queries_by_queryid_all.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = QUERYID SP DATE SP QUERY

Rows MUST be sorted by the QUERYID column.
For more information about ABNF parameters, see section 2.1.

2.4.28  <gen>.queryinfo.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = ZERO SP QUERIESINFO
For more information about ABNF parameters, see section 2.1.

2.4.29 <gen>.urls_by_urlid.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URLID SP DATE SP URL

For more information about ABNF parameters, see section 2.1.

2.4.30 <gen>.urls_by_urlid_all.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URLID SP DATE SP URL

For more information about ABNF parameters, see section 2.1.

2.4.31 <gen>.<col>.unique_uris_by_uri.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = CID

Rows MUST be sorted by the CID column.
For more information about ABNF parameters, see section 2.1.

2.4.32 <gen>.<col>.uris_by_contentid.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = CID SP TIMESTAMP SP OP SP MEMBER

Rows MUST be sorted by the CID column.
For more information about ABNF parameters, see section 2.1.

2.4.33 <gen>.<col>.uris_by_contentid_ts.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = CID SP TIMESTAMP SP OP SP MEMBER

Rows MUST be sorted by the CID column.
For more information about ABNF parameters, see section 2.1.

2.4.34 <gen>.<col>.<host>_contentids_by_contentid_new.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.
ROW = CID SP TIMESTAMP SP OP SP MEMBER

Rows MUST be sorted by the CID column.
For more information about ABNF parameters, see section 2.1.

2.4.35  <gen>.<col>.<host>_contentids_by_contentid_new_resplit.<sf>.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = CID SP TIMESTAMP SP OP SP MEMBER

Rows MUST be sorted by the CID column.
For more information about ABNF parameters, see section 2.1.

2.4.36  <gen>.<col>.<host>_uris.0
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = CID SP TIMESTAMP SP OP SP MEMBER

For more information about ABNF parameters, see section 2.1.

2.4.37  <date>.clicks_by_urlid_and_queryid.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = QUERYID SP URLID SP LCID SP CLICKEDURLRANK

For more information about ABNF parameters, see section 2.1.

2.4.38  <date>.clicks_by_urlid_and_queryid_sort.<sf>.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = QUERYID SP URLID SP LCID SP CLICKEDURLRANK

For more information about ABNF parameters, see section 2.1.

2.4.39  <date>.clicks_on_queryid.0
Files of this type MUST contain rows that are specified by the following ABNF rule.

ROW = QUERYID SP URLID SP LCID SP CLICKEDURLRANK

For more information about ABNF parameters, see section 2.1.
2.4.40  <date>.local_querycnt_by_queryid.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URLID SP QUERYID SP LCID SP LCOUNT

Rows MUST be sorted by the QUERYID column.
For more information about ABNF parameters, see section 2.1.

2.4.41  <date>.local_querycnt_by_queryid_reduce.<sf>.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URLID SP QUERYID SP LCID SP LCOUNT

Rows MUST be sorted by the QUERYID column.
For more information about ABNF parameters, see section 2.1.

2.4.42  <date>.queries_by_queryid.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = QUERYID SP DATE SP QUERY

Rows MUST be sorted by the QUERYID column.
For more information about ABNF parameters, see section 2.1.

2.4.43  <date>.queries_by_queryid_sort.0.<sf>

Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = QUERYID SP DATE SP QUERY

Rows MUST be sorted by the QUERYID column.
For more information about ABNF parameters, see section 2.1.

2.4.44  <date>.queries_on_queryid.0

Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = QUERYID SP DATE SP QUERY

Rows MUST be sorted by the QUERYID column.
For more information about ABNF parameters, see section 2.1.
2.4.45  <date>.urls_by_urlid.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URLID SP DATE SP URL

For more information about ABNF parameters, see section 2.1.

2.4.46  <date>.urls_by_urlid_sort.0.<sf>
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URLID SP DATE SP URL

For more information about ABNF parameters, see section 2.1.

2.4.47  <date>.urls_on_urlid.0
Files of this type MUST contain rows that are specified by the following ABNF rule.

    ROW = URLID SP DATE SP URL

For more information about ABNF parameters, see section 2.1.

2.5  Database Files

The database files consist of the resulting output from the analysis process.

2.5.1  <gen>.sharepoint.rel.<part_num>.bin
These files specify a set of records. The record size MUST be a multiple of 32, specified as a 32-bit signed integer in little-endian order before each record. If the record size is not a multiple of 32, the record MUST be padded with zeros.

The first record is a header record. The size of this record MUST be set to 124. This record MUST contain a marshaled representation of the following string, as specified in [MS-FSWCU].

    "{'offset_step': 32, 'len_field_type': 'I', 'serializer': 'pyfastmarshal', 'compression_type': 'gzip'}"

A record entry consists of a dictionary, as specified in [MS-FSWCU]. This dictionary MUST contain two keys: `contentid` and `queries`.

The `contentid` key MUST contain a document identifier (3) (CID).

The `queries` key MUST contain an array, as specified in [MS-FSWCU], with a set of query entries. Each query entry, as specified in [MS-FSWCU], is a five-entry tuple that MUST contain the columns `LCOUNT`, `GCOUNT`, `PLACE`, `NUMURLS`, and `QUERY`.

A record string MUST be serialized as specified in [MS-FSWCU]. All records MUST be compressed using the zlib format, as specified in [RFC1950]. For each compressed record, the compression
method and flags header MUST be removed. This means the protocol removes the first 2 bytes "78 9C" from every compressed record.

2.5.2  \(<gen>.sharepoint.rel.<part_num>.idx\)

This file specifies a set of hash values of the record entries, excluding the header record, in the \(<gen>.sharepoint.rel.<part_num>.bin\) file. Each hash value MUST be computed with the 32 most significant bits of a 128-bit MD5 hash. The 4-byte hash value MUST be specified in little-endian order. The 128-bit MD5 hash is calculated from the \(\text{contentid}\) field of the record entries in the \(<gen>sharepoint.rel.<part_num>.bin\) file. The hash values MUST be in the same order as the record entries.

2.5.3  \(<gen>.sharepoint.rel.<part_num>.idx.ofs\)

This file contains offsets to record entries in the \(<gen>.sharepoint.rel.<part_num>.bin\) file. Each offset is relative to the first record after the header record in the \(<gen>.sharepoint.rel.<part_num>.bin\) file. An offset is a 4-byte integer that is calculated by dividing the byte offset by 32. Each offset entry MUST be specified in little-endian order, and MUST be in the same order as the record entries.

2.6 Empty Files

All files that have an ".end" file extension MUST be empty. In addition, the following files MUST be empty: \(\text{cidcollapser\_done, done, sppartialupdate\_done and copied_<gen>.sharepoint.rel.<part_num>}\). These files are used for tracking internal states.
3 Structure Examples

The following are examples of information gathered by the analysis process.

3.1 Search Clickthrough Files

3.1.1 <date>.clicks.txt files

The following is an example of a <date>.clicks.txt file.

```
1017 2017 1 1
1018 2018 1 1
1019 2019 1 1
1020 2020 1 1
1021 2021 1 1
1022 2022 1 1
1023 2023 1 1
1024 2024 1 1
1025 2025 1 1
1026 2026 1 1
```

3.1.2 <date>.queries.txt files

The following is an example of a <date>.queries.txt file.

```
1000 2009.03.16 query1000 scope:"all sites"
1001 2009.03.17 query1001 scope:"all sites"
1002 2009.03.18 query1002 scope:"all sites"
1003 2009.03.19 query1003 scope:"all sites"
1004 2009.03.20 query1004 scope:"all sites"
1005 2009.03.21 query1005 scope:"all sites"
1006 2009.03.22 query1006 scope:"all sites"
1007 2009.03.23 query1007 scope:"all sites"
1008 2009.03.24 query1008 scope:"all sites"
```

3.1.3 <date>.urls.txt files

The following is an example of a <date>.urls.txt file.

```
2005 2009.03.21 http://www.alpineskihouse.com/2005
```
3.2 Analysis Files

3.2.1 allfeeduris.<sf> files

The following is an example of an allfeeduris.<sf> file.

\[
\begin{align*}
&\text{ssic://2143300394 sp} \\
&\text{ssic://2143300398 sp} \\
&\text{ssic://2143300399 sp} \\
&\text{ssic://2145300400 sp} \\
&\text{ssic://2145300402 sp} \\
&\text{ssic://2145300405 sp} \\
&\text{ssic://2145300407 sp} \\
&\text{ssic://2146300404 sp}
\end{align*}
\]

3.2.2 cid_by_cid_with_counts_and_query.<sf>

The following is an example of a cid_by_cid_with_counts_and_query.<sf> file.

\[
\begin{align*}
&\text{ssic://2143300394 22 22 1 1 query1022 scope all sites} \\
&\text{ssic://2143300398 26 26 1 1 query1026 scope all sites} \\
&\text{ssic://2143300399 27 27 1 1 query1027 scope all sites} \\
&\text{ssic://2145300400 6 6 1 1 query1006 scope all sites} \\
&\text{ssic://2145300402 4 4 1 1 query1004 scope all sites} \\
&\text{ssic://2145300405 3 3 1 1 query1003 scope all sites} \\
&\text{ssic://2145300407 1 1 1 1 query1001 scope all sites} \\
&\text{ssic://2146300404 15 15 1 1 query1015 scope all sites}
\end{align*}
\]

3.2.3 cid_by_cid_with_counts_and_query_mergereduce.<sf>.<sf>

The following is an example of a cid_by_cid_with_counts_and_query_mergereduce.<sf>.<sf> file.

\[
\begin{align*}
&\text{ssic://2145300405 3 3 1 1 query1003 scope all sites} \\
&\text{ssic://2145300400 6 6 1 1 query1006 scope all sites} \\
&\text{ssic://2146300404 15 15 1 1 query1015 scope all sites} \\
&\text{ssic://2143300394 22 22 1 1 query1022 scope all sites} \\
&\text{ssic://2143300398 26 26 1 1 query1026 scope all sites}
\end{align*}
\]

3.2.4 global_querycnt_by_query.<sf>

The following is an example of a global_querycnt_by_query.<sf> file.

\[
\begin{align*}
&2 \text{ query1002 scope all sites} \\
&3 \text{ query1003 scope all sites} \\
&5 \text{ query1005 scope all sites} \\
&6 \text{ query1006 scope all sites} \\
&9 \text{ query1009 scope all sites} \\
&13 \text{ query1013 scope all sites} \\
&14 \text{ query1014 scope all sites} \\
&15 \text{ query1015 scope all sites} \\
&19 \text{ query1019 scope all sites} \\
&22 \text{ query1022 scope all sites} \\
&25 \text{ query1025 scope all sites}
\end{align*}
\]
3.2.5 local\_querycnt\_by\_cid.<sf>

The following is an example of a local\_querycnt\_by\_cid.<sf> file.

```
ssic://2143300394 22 query1022 scope all sites
ssic://2143300398 26 query1026 scope all sites
ssic://2143300399 27 query1027 scope all sites
ssic://2145300400 6 query1006 scope all sites
ssic://2145300402 4 query1004 scope all sites
ssic://2145300405 3 query1003 scope all sites
ssic://2145300407 1 query1001 scope all sites
ssic://2146300404 15 query1015 scope all sites
```

3.2.6 local\_querycnt\_by\_cid\_merge.<sf>.<sf>

The following is an example of a local\_querycnt\_by\_cid\_merge.<sf>.<sf> file.

```
ssic://2145300405 3 query1003 scope all sites
ssic://2143300394 22 query1022 scope all sites
ssic://2143300398 26 query1026 scope all sites
```

3.2.7 local\_querycnt\_by\_query.<sf>

The following is an example of a local\_querycnt\_by\_query.<sf> file.

```
ssic://2145300404 2 query1002 scope all sites
ssic://2145300405 3 query1003 scope all sites
ssic://2145300403 5 query1005 scope all sites
ssic://2145300400 6 query1006 scope all sites
ssic://2145300415 9 query1009 scope all sites
ssic://2146300402 13 query1013 scope all sites
ssic://2146300405 14 query1014 scope all sites
ssic://2146300404 15 query1015 scope all sites
ssic://2146300408 19 query1019 scope all sites
ssic://2143300394 22 query1022 scope all sites
ssic://2143300397 25 query1025 scope all sites
ssic://2143300398 26 query1026 scope all sites
```

3.2.8 local\_querycnt\_by\_query\_reduce.<sf>.<sf>

The following is an example of a local\_querycnt\_by\_query\_reduce.<sf>.<sf> file.

```
ssic://2143300394 22 query1022 scope all sites
ssic://2143300398 26 query1026 scope all sites
ssic://2145300400 6 query1006 scope all sites
ssic://2145300405 3 query1003 scope all sites
ssic://2146300404 15 query1015 scope all sites
```
3.2.9  local_querycnt_by_url.<sf>

The following is an example of a local_querycnt_by_url.<sf> file.

http://www.alpineskihouse.com/2003 3 query1003 scope all sites
http://www.alpineskihouse.com/2007 7 query1007 scope all sites
http://www.alpineskihouse.com/2011 11 query1011 scope all sites
http://www.alpineskihouse.com/2014 14 query1014 scope all sites
http://www.alpineskihouse.com/2016 16 query1016 scope all sites
http://www.alpineskihouse.com/2021 21 query1021 scope all sites
http://www.alpineskihouse.com/2022 22 query1022 scope all sites
http://www.alpineskihouse.com/2026 26 query1026 scope all sites

3.2.10  local_querycnt_by_url_merge.<sf>.<sf>

The following is an example of a local_querycnt_by_url_merge.<sf>.<sf> file.

http://www.alpineskihouse.com/2003 3 query1003 scope all sites
http://www.alpineskihouse.com/2021 21 query1021 scope all sites

3.2.11  local_querycnt_by_urlid.<sf>

The following is example of a local_querycnt_by_urlid.<sf> file.

2001 1 query1001 scope all sites
2003 3 query1003 scope all sites
2006 6 query1006 scope all sites
2009 9 query1009 scope all sites
2012 12 query1012 scope all sites
2013 13 query1013 scope all sites
2017 17 query1017 scope all sites
2021 21 query1021 scope all sites
2024 24 query1024 scope all sites
2025 25 query1025 scope all sites
2027 27 query1027 scope all sites

3.2.12  local_querycnt_by_urlid_reduce.<sf>.<sf>

The following is an example of a local_querycnt_by_urlid_reduce.<sf>.<sf> file.

http://www.alpineskihouse.com/2007 7 query1007 scope all sites
http://www.alpineskihouse.com/2014 14 query1014 scope all sites
http://www.alpineskihouse.com/2026 26 query1026 scope all sites

3.2.13  semi_local_querycnt_by_queryid.<sf>

The following is an example of a semi_local_querycnt_by_queryid.<sf> file.

2002 1002 1 1
2002 1002 1 1
2004 1004 1 1
2004 1004 1 1
3.2.14  semi_local_querycnt_by_urlid.<sf>

The following is an example of a semi_local_querycnt_by_urlid.<sf> file.

2001 1 query1001 scope all sites
2003 1 query1003 scope all sites
2003 1 query1003 scope all sites
2003 1 query1003 scope all sites
2006 1 query1006 scope all sites
2006 1 query1006 scope all sites
2006 1 query1006 scope all sites
2006 1 query1006 scope all sites
2006 1 query1006 scope all sites
2006 1 query1006 scope all sites
2006 1 query1006 scope all sites

3.2.15  semi_local_querycnt_by_urlid_map.<sf>,<sf>

The following is an example of a semi_local_querycnt_by_urlid_map.<sf>,<sf> file.

2003 1 query1003 scope all sites
2003 1 query1003 scope all sites
2003 1 query1003 scope all sites
2012 1 query1012 scope all sites
2012 1 query1012 scope all sites
2012 1 query1012 scope all sites
2012 1 query1012 scope all sites
2012 1 query1012 scope all sites
2012 1 query1012 scope all sites
2012 1 query1012 scope all sites

3.2.16  semi_local_querycnt_pre_token.<sf>

The following is an example of a semi_local_querycnt_pre_token.<sf> file.

2002 1 1 query1002 scope:"all sites"
2002 1 1 query1002 scope:"all sites"
2004 1 1 query1004 scope:"all sites"
2004 1 1 query1004 scope:"all sites"
2004 1 1 query1004 scope:"all sites"
2004 1 1 query1004 scope:"all sites"
2006 1 1 query1006 scope:"all sites"
2006 1 1 query1006 scope:"all sites"
3.2.17 uris_by_contentid_ts.<sf>

The following is an example of an uris_by_contentid_ts.<sf> file.

```
ssic://2143300394 1239888256 ADD http://www.alpineskihouse.com/2022
ssic://2143300398 1239888256 ADD http://www.alpineskihouse.com/2026
ssic://2143300399 1239888256 ADD http://www.alpineskihouse.com/2027
```

3.2.18 uris_by_member.<sf>

The following is an example of a uris_by_member.<sf> file.

```
ssic://2143300393 http://www.alpineskihouse.com/2021
```

3.2.19 uris_by_member_reduce.<sf>

The following is an example of a uris_by_member_reduce.<sf> file.

```
ssic://2143300394 http://www.alpineskihouse.com/2022
ssic://2143300398 http://www.alpineskihouse.com/2026
```

3.2.20 urls_by_urlhash_with_queries.<sf>

The following is an example of a urls_by_urlhash_with_queries.<sf> file.

```
28768632944601236036478028165026923624
e3MJAAAAY29udGVudGlkc3QAAABzc2ljOi8vMjE0MzMwMDM4NXMHAAAAcXVlcnkxMDI5IHNjb3BlIGFsbCBzaXRlc3Q=
36707531868016354637501849906369623
e3MJAAAAY29udGVudGlkc3QAAABzc2ljOi8vMjE0NTMwMDQwOHMIAAAcXVlcnkxMDE4IHNjb3BlIGFsbCBzaXRlc3Q=
4960786479638935637760521368364923015
e3MJAAAAY29udGVudGlkc3QAAABzc2ljOi8vMjE0M0NjMjMwMDQwN3MIAAAcXVlcnkzMDExIHNjb3BlIGFsbCBzaXRlc3Q=
31280583718211359997925733612579346
```

Using the QUERIES column from the first row, results in the following raw data.
Decoding this information using base64 encoding results in the following raw data.

```
0000000: 7b 73 09 00 00 00 63 6f 6e 74 65 6e 74 69 64 73
0000010: 11 00 00 00 73 73 69 63 3a 2f 2f 32 31 34 33
0000020: 30 30 33 38 35 73 07 00 00 00 71 75 65 72 69 65
0000030: 73 5b 01 00 00 00 28 05 00 00 00 73 02 00 00 00
0000040: 32 39 73 02 00 00 00 32 39 73 01 00 00 00 31 73
0000050: 01 00 00 00 31 73 19 00 00 00 71 75 67
0000060: 3e3MJAAAAY29udGVudGlkcxEAAABzc2ljOi8vMjE0MzMwMDM5NHMHAAAAcXVlcnkxMDI2IHNjb3BlIGFsbCBzaXRlczA=
```

Deserializing the preceding information, as described in [MS-FSWCU], results in the following string.

```
{"contentid": 'ssic://2143300385', 'queries': [('29', '29', '1', '1', 'query1029 scope all sites')]}  
```

### 3.2.21 urls_by_urlhash_with_queries_sort.<sf>

The following is an example of a `urls_by_urlhash_with_queries_sort.<sf>` file.

```
26223550463997515671362586971738326722
```

### 3.2.22 urls_on_urlhash_with_queries.<sf>

The following is an example of a `urls_on_urlhash_with_queries.<sf>` file.

```
26223550463997515671362586971738326722
```

[MS-FSSPRDF] — v20120630
SPRel Data File Formats

Copyright © 2012 Microsoft Corporation.

Release: July 16, 2012
3.2.23 `<col>_feeduris.<sf>`

The following is an example of a `<col>_feeduris.<sf>` file.

```
ssic://2143300394 sp
ssic://2143300398 sp
ssic://2143300399 sp
ssic://2145300400 sp
ssic://2145300402 sp
ssic://2145300405 sp
ssic://2146300404 sp
```

3.2.24 `<col>_feeduris_expand.<sf>`

The following is an example of a `<col>_feeduris_expand.<sf>` file.

```
ssic://2143300394 sp
ssic://2145300405 sp
ssic://2145300407 sp
```

3.2.25 `<col>_feeduris_expand_resplit.<sf>`

The following is an example of a `<col>_feeduris_expand_resplit.<sf>` file.

```
ssic://2143300394 sp
ssic://2145300405 sp
ssic://2145300407 sp
```

3.2.26 `<gen>.queries_by_queryid.<sf>`

The following is an example of a `<gen>.queries_by_queryid.<sf>` file.

```
1002 2009.03.17 query1002 scope:"all sites"
1004 2009.03.19 query1004 scope:"all sites"
1006 2009.03.21 query1006 scope:"all sites"
1011 2009.03.26 query1011 scope:"all sites"
1018 2009.04.02 query1018 scope:"all sites"
1020 2009.04.04 query1020 scope:"all sites"
1022 2009.04.06 query1022 scope:all sites"
1023 2009.04.07 query1023 scope:"all sites"
1025 2009.04.09 query1025 scope:"all sites"
```
3.2.27 `<gen>.queries_by_queryid_all.<sf>`

The following is an example of a `<gen>.queries_by_queryid_all.<sf>` file.

1002 2009.03.17 query1002 scope:"all sites"
1004 2009.03.19 query1004 scope:"all sites"
1006 2009.03.21 query1006 scope:"all sites"
1011 2009.03.26 query1011 scope:"all sites"
1018 2009.04.02 query1018 scope:"all sites"
1020 2009.04.04 query1020 scope:"all sites"
1022 2009.04.06 query1022 scope:"all sites"
1023 2009.04.07 query1023 scope:"all sites"
1025 2009.04.09 query1025 scope:"all sites"
1027 2009.04.11 query1027 scope:"all sites"

3.2.28 `<gen>.queryinfo.<sf>`

The following is an example of a `<gen>.queryinfo.<sf>` file.

```
0
```

Using the `QUERIESINFO` column from the first row, results in the following raw data.

```
0000000: 65 33 4d 4a 41 41 41 41 59 32 39 75 64 47 56 75 ... e3MJAAAAY29udGVudGlc... 0000010: 64 47 6c 6b 63 78 45 41 41 41 42 7a 63 32 6c 6a ... dGlkcxEAAABzc2lj 0000020: 4f 69 38 76 4d 6a 45 30 4d 7a 4d 77 4d 44 4d ... Oi8vMjE0MzMwMDM5 0000030: 4e 48 4d 4l 41 41 41 41 59 32 39 73 62 47 56 6a ... NHMLAAAAY29ubGVj 0000040: 64 47 6c 76 62 6e 4n 62 41 51 41 41 41 48 46 7a ... dGlvbNhBQAAAAHM 0000050: 41 41 41 41 4d 6a 4a 7a 41 51 41 41 41 44 46 31 ... AAAAc3BzBAAAAAHF 0000060: 5a 58 4j 7b 7e 4b 61 41 41 41 43 67 46 36 4d ... ZXJpZXNwbAAACAgF 0000070: 41 41 41 41 63 63 42 7a 41 51 41 41 41 44 48 46 ... AAAAcwIAAAAyMnMC 0000080: 41 41 41 41 4d 6a 4d 74 41 51 41 41 41 44 46 7a ... AAAAMjzQAAAAAFz 0000090: 41 51 41 41 41 44 46 47 7a 47 51 41 41 41 48 46 ... AAAAaDfzQAAAAAF 00000a0: 5a 58 4j 7b 7e 4a 4a 7a 5a 58 3e 62 41 51 41 41 ... ZXJpZXNwbAAACAg 00000b0: 41 41 41 41 4d 6a 43 62 7a 41 51 41 41 41 44 46 ... AAAAcwIAAAAyMnMC 00000c0: 0a
```

Decoding this information using base64 encoding results in the following raw data.
Deserializing the preceding data, as described in [MS-FSWCU], results in the following string for the QUESTIESINFO column.

```
{
    'contentid': 'ssic://2143300394',
    'collections': ['sp'],
    'queries': [('22', '22', '1', '1', 'query1022 scope all sites')]
}
```

### 3.2.29 <gen>.urls_by_urlid.<sf>

The following is an example of a <gen>.urls_by_urlid.<sf> file.

```
2012 2009.03.27 http://www.alpineskihouse.com/2012
2013 2009.03.28 http://www.alpineskihouse.com/2013
2017 2009.04.01 http://www.alpineskihouse.com/2017
2021 2009.04.05 http://www.alpineskihouse.com/2021
```

### 3.2.30 <gen>.urls_by_urlid_all.<sf>

The following is an example of a <gen>.urls_by_urlid_all.<sf> file.

```
2012 2009.03.27 http://www.alpineskihouse.com/2012
2013 2009.03.28 http://www.alpineskihouse.com/2013
```

### 3.2.31 <gen>.<col>.unique_uris_by_uri.<sf>

The following is an example of a <gen>.<col>.unique_uris_by_uri.<sf> file.

```
ssic://2143300394
ssic://2143300398
ssic://2143300399
ssic://2145300400
ssic://2145300402
ssic://2145300405
ssic://2145300406
```
3.2.32  <gen>.<col>.uris_by_contentid.<sf>

The following is an example of a <gen>.<col>.uris_by_contentid.<sf> file.

```
ssic://2143300394 1239888256 ADD http://www.alpineskihouse.com/2022
ssic://2143300398 1239888256 ADD http://www.alpineskihouse.com/2026
ssic://2143300399 1239888256 ADD http://www.alpineskihouse.com/2027
```

3.2.33  <gen>.<col>.uris_by_contentid_ts.<sf>

The following is an example of a <gen>.<col>.uris_by_contentid_ts.<sf> file.

```
ssic://1134254170 1239888257 DEL ǂ
ssic://2143300394 1239888256 ADD http://www.alpineskihouse.com/2022
ssic://2143300398 1239888256 ADD http://www.alpineskihouse.com/2026
ssic://2143300399 1239888256 ADD http://www.alpineskihouse.com/2027
```

3.2.34  <gen>.<col>.<host>_contentids_by_contentid_new.<sf>

The following is an example of a <gen>.<col>.<host>_contentids_by_contentid_new.<sf> file.

```
ssic://1134254170 1239888257 DEL ǂ
ssic://2143300394 1239888256 ADD http://www.alpineskihouse.com/2022
ssic://2143300398 1239888256 ADD http://www.alpineskihouse.com/2026
ssic://2143300399 1239888256 ADD http://www.alpineskihouse.com/2027
```

3.2.35  <gen>.<col>.<host>_contentids_by_contentid_new_resplit.<sf>.<sf>

The following is an example of a <gen>.<col>.<host>_contentids_by_contentid_new_resplit.<sf>.<sf> file.

```
ssic://2143300394 1239888256 ADD http://www.alpineskihouse.com/2022
ssic://2143300398 1239888256 ADD http://www.alpineskihouse.com/2026
```
3.2.36  `<gen>.<col>.<host>_uris.0`

The following is an example of a `<gen>.<col>.<host>_uris.0` file.

```
ssic://2143300399 1239888256 ADD http://www.alpineskihouse.com/2027
ssic://1134254170 1239888257 DEL |

ssic://2143300394 1239888256 ADD http://www.alpineskihouse.com/2022
ssic://2143300395 1239888256 ADD http://www.alpineskihouse.com/2023
ssic://2143300396 1239888256 ADD http://www.alpineskihouse.com/2024
ssic://2143300397 1239888256 ADD http://www.alpineskihouse.com/2025
ssic://2143300398 1239888256 ADD http://www.alpineskihouse.com/2026
ssic://2143300399 1239888256 ADD http://www.alpineskihouse.com/2027
ssic://2143300384 1239888256 ADD http://www.alpineskihouse.com/2028
ssic://2143300385 1239888256 ADD http://www.alpineskihouse.com/2029
ssic://2144300395 1239888256 ADD http://www.alpineskihouse.com/2030
ssic://1134254170 1239888256 ADD http://invalid.com
ssic://1134254170 1239888257 DEL |
```

3.2.37  `<date>.clicks_by_urlid_and_queryid.<sf>`

The following is an example of a `<date>.clicks_by_urlid_and_queryid.<sf>` file.

```
1001 2001 1 1
1003 2003 1 1
1006 2006 1 1
1009 2009 1 1
1012 2012 1 1
1013 2013 1 1
1017 2017 1 1
1021 2021 1 1
```

3.2.38  `<date>.clicks_by_urlid_and_queryid_sort.<sf>.<sf>`

The following is an example of a `<date>.clicks_by_urlid_and_queryid_sort.<sf>.<sf>` file.

```
1001 2001 1 1
1003 2003 1 1
1006 2006 1 1
1009 2009 1 1
1012 2012 1 1
1013 2013 1 1
1017 2017 1 1
1021 2021 1 1
```

3.2.39  `<date>.clicks_on_queryid.0`

The following is an example of a `<date>.clicks_on_queryid.<sf>` file.

```
1001 2001 1 1
1002 2002 1 1
1003 2003 1 1
```
3.2.40  <date>.local_querycnt_by_queryid.<sf>

The following is an example of a <date>.local_querycnt_by_queryid.<sf> file.

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Day</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>1002</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2004</td>
<td>1004</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2006</td>
<td>1006</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2011</td>
<td>1011</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2018</td>
<td>1018</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2020</td>
<td>1020</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2022</td>
<td>1022</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2023</td>
<td>1023</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2.41  <date>.local_querycnt_by_queryid_reduce.<sf>.<sf>

The following is an example of a <date>.local_querycnt_by_queryid_reduce.<sf>.<sf> file.

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Day</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>1006</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2025</td>
<td>1025</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2027</td>
<td>1027</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2.42  <date>.queries_by_queryid.<sf>

The following is an example of a <date>.queries_by_queryid.<sf> file.

<table>
<thead>
<tr>
<th>ID</th>
<th>Date</th>
<th>Query</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>2009.03.16</td>
<td>query1000</td>
<td>all sites</td>
</tr>
<tr>
<td>1001</td>
<td>2009.03.17</td>
<td>query1001</td>
<td>all sites</td>
</tr>
<tr>
<td>1002</td>
<td>2009.03.18</td>
<td>query1002</td>
<td>all sites</td>
</tr>
<tr>
<td>1003</td>
<td>2009.03.19</td>
<td>query1003</td>
<td>all sites</td>
</tr>
<tr>
<td>1004</td>
<td>2009.03.20</td>
<td>query1004</td>
<td>all sites</td>
</tr>
<tr>
<td>1005</td>
<td>2009.03.21</td>
<td>query1005</td>
<td>all sites</td>
</tr>
<tr>
<td>1006</td>
<td>2009.03.22</td>
<td>query1006</td>
<td>all sites</td>
</tr>
</tbody>
</table>

3.2.43  <date>.queries_by_queryid_sort.0.<sf>

The following is an example of a <date>.queries_by_queryid_sort.0.<sf> file.

<table>
<thead>
<tr>
<th>ID</th>
<th>Date</th>
<th>Query</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>2009.03.16</td>
<td>query1000</td>
<td>all sites</td>
</tr>
<tr>
<td>1001</td>
<td>2009.03.17</td>
<td>query1001</td>
<td>all sites</td>
</tr>
<tr>
<td>1002</td>
<td>2009.03.18</td>
<td>query1002</td>
<td>all sites</td>
</tr>
<tr>
<td>1003</td>
<td>2009.03.19</td>
<td>query1003</td>
<td>all sites</td>
</tr>
<tr>
<td>1004</td>
<td>2009.03.20</td>
<td>query1004</td>
<td>all sites</td>
</tr>
<tr>
<td>1005</td>
<td>2009.03.21</td>
<td>query1005</td>
<td>all sites</td>
</tr>
<tr>
<td>1006</td>
<td>2009.03.22</td>
<td>query1006</td>
<td>all sites</td>
</tr>
</tbody>
</table>
3.2.44  <date>.queries_on_queryid.0

The following is an example of a <date>.queries_on_queryid.0 file.

```
1000 2009.03.16 query1000 scope:"all sites"
1001 2009.03.17 query1001 scope:"all sites"
1002 2009.03.18 query1002 scope:"all sites"
1003 2009.03.19 query1003 scope:"all sites"
1004 2009.03.20 query1004 scope:"all sites"
1005 2009.03.21 query1005 scope:"all sites"
1006 2009.03.22 query1006 scope:"all sites"
```

3.2.45  <date>.urls_by_urlid.<sf>

The following is an example of a <date>.urls_by_urlid.<sf> file.

```
2005 2009.03.21 http://www.alpineskihouse.com/2005
```

3.2.46  <date>.urls_by_urlid_sort.0.<sf>

The following is an example of a <date>.urls_by_urlid_sort.0.<sf> file.

```
2005 2009.03.21 http://www.alpineskihouse.com/2005
```

3.2.47  <date>.urls_on_urlid.0

The following is an example of a <date>.urls_on_urlid.0 file.

```
2005 2009.03.21 http://www.alpineskihouse.com/2005
```
### 3.3 Database Files

#### 3.3.1 `<gen>.sharepoint.rel.<part_num>.bin`

The following is raw data of a `<gen>.sharepoint.rel.<part_num>.bin` file.

```
0000000: 7c 00 00 00 7b 73 0b 00 00 00 6f 66 66 73 65 74  |...{s....offset
0000010: 5f 73 74 65 70 69 20 00 00 00 73 0e 00 00 00 6c  _stepi ...s....l
0000020: 65 6e 5f 66 69 65 6c 64 5f 74 79 70 65 73 01 00  en_field_types..
0000030: 00 00 49 73 0a 00 00 00 73 65 72 69 61 6c 69 7a  IS...
0000040: 65 72 73 0d 00 00 00 7y 66 61 73 74 6d 61 72 73  ers...pyfastmar
0000050: 73 68 61 6c 73 10 00 00 00 63 6f 6d 70 72 65 73  shals....compress
0000060: 73 69 6f 6e 5f 74 79 70 65 73 04 00 00 00 67 7a  sion_types....gz
0000070: 69 70 30 00 00 00 00 00 00 00 00 00 00 00 00 00  ip0.
0000080: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd  |......d``H..+I.
0000090: 2b c9 4l 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c  +.L).....3......
00000a0: 4d 4c 8d 0d 0c 4l 0c 8a d9 81 a2 85 a5 a9 45  ML...L.........E
00000b0: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 b4 80  ....@..+H...0.
00000c0: 99 a9 c5 73 74 65 70 69 20 00 00 00 73 0e 00 00 00 _stepi ...s....l
00000d0: 73 65 72 69 61 6c 69 7a 65 72 73 0d 00 00 00 7y  ser...pyfastmar
00000e0: 66 61 73 74 6d 61 72 73 73 68 61 6c 73 10 00 00  fastmar...shals
00000f0: 69 70 30 00 00 00 00 00 00 00 00 00 00 00 00 00  ip0.
0000100: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd  |......d``H..+I.
0000110: 2b c9 4l 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c  +.L).....3......
0000120: 4d 4c 8d 0d 0c 4l 0c cc 8a d9 81 a2 85 a5 a9 45  ML...L.........E
0000130: 99 a9 c5 73 74 65 70 69 20 00 00 00 73 0e 00 00 00 _stepi ...s....l
0000140: 73 65 72 69 61 6c 69 7a 65 72 73 0d 00 00 00 7y  ser...pyfastmar
0000150: 66 61 73 74 6d 61 72 73 73 68 61 6c 73 10 00 00  fastmar...shals
0000160: 69 70 30 00 00 00 00 00 00 00 00 00 00 00 00 00  ip0.
0000170: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd  |......d``H..+I.
0000180: 2b c9 4l 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c  +.L).....3......
0000190: 4d 4l 8d 0d 0c 4l 0c 8a d9 81 a2 85 a5 a9 45  ML...L.........E
00001a0: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 b4 80  ....@..+H...0.
00001b0: 99 a9 c5 73 74 65 70 69 20 00 00 00 73 0e 00 00 00 _stepi ...s....l
00001c0: 73 65 72 69 61 6c 69 7a 65 72 73 0d 00 00 00 7y  ser...pyfastmar
00001d0: 66 61 73 74 6d 61 72 73 73 68 61 6c 73 10 00 00  fastmar...shals
00001e0: 69 70 30 00 00 00 00 00 00 00 00 00 00 00 00 00  ip0.
00001f0: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd  |......d``H..+I.
0000200: 2b c9 4l 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c  +.L).....3......
0000210: 4d 4l 8d 0d 0c 4l 0c cc 8a d9 81 a2 85 a5 a9 45  ML...L.........E
0000220: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 b4 80  ....@..+H...0.
0000230: 99 a9 c5 73 74 65 70 69 20 00 00 00 73 0e 00 00 00 _stepi ...s....l
0000240: 73 65 72 69 61 6c 69 7a 65 72 73 0d 00 00 00 7y  ser...pyfastmar
0000250: 66 61 73 74 6d 61 72 73 73 68 61 6c 73 10 00 00  fastmar...shals
0000260: 69 70 30 00 00 00 00 00 00 00 00 00 00 00 00 00  ip0.

```
```
The first 4 bytes specify the header size, which is 124 bytes.

Deserializing the following 124 bytes as described in [MS-FSWCU] results in the following file header:

```
{"offset_step": 32, 'len_field_type': 'I', 'serializer': 'pyfastmarshal', 'compression_type': 'gzip'}
```

The next 4 bytes specify the size of the next compressed record, which is 124 bytes. Adding the zlib header as described in [RFC1950], in front of the following 124 bytes results in the following raw data:

```
0000000: 78 9c ab 2e e6 64 60 60 48 ce cf 2b 49 cd 2b c9 4L).....3......M.
0000010: 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c 4d cc L).....3.......Q
0000020: 8c 0d 0c 4c 0c 2c 8b d9 81 a2 85 a5 a9 45 99 a9 8c,.,L.,.....E..
0000030: c5 d1 8c 40 b6 06 2b 48 19 13 90 30 b4 80 51 20 ...@..+H...0..Q
0000040: 71 43 28 29 09 55 5f 69 68 60 68 a1 50 9c 9c 5f qC()_.U_ih`h.P.._
0000050: 90 aa 90 98 93 a3 50 9c 59 92 5a 6l 00 00 e5 2d ......P.Y.Zl...
0000060: 1a 29 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ..)..............
0000070: 00 00 00 00 00 00 00 00 00 00 00 00 00 00        ..............
```

Decompressing this information using the zlib format as described in [RFC1950], results in the following.

```
0000000: 7b 73 09 00 00 00 63 6f 6e 74 65 6e 74 69 64 73  contentids
0000010: 11 00 00 00 73 73 69 63 3a 2f 2f 32 31 34 36 33  ssc://21463
0000020: 30 30 34 07 31 38 73 07 00 00 00 31 38 73 01 00 00 00 18scope all sites
0000030: 31 38 73 02 00 00 00 31 38 73 01 00 00 00 31 73 18scope all sites
0000040: 01 00 00 00 31 73 19 00 00 00 71 75 65 72 79 31 1scope all sites
0000050: 30 31 38 20 73 63 6f 70 65 20 61 6c 6c 20 73 69 018 scope all sites
0000060: 74 6e 65 73 30 18 scope all sites0
```

Deserializing the data as described in [MS-FSWCU], results in the following record.

```
{"contentid": 'ssc://2146300409', 'queries': [('18', '18', '1', '1', 'query1018 scope all sites')]}
```

### 3.3.2  `<gen>.sharepoint.rel.<part_num>.idx`

The following is raw data from the a `<gen>.sharepoint.rel.<part_num>.idx` file.

```
0000000: 8a 72 5a 02 fc 33 11 04 1d a3 a4 15 0d 9d d 9d 1b  rZ..3............
0000010: 92 20 52 25 f8 4d be 2c 3e 2c 3e 2d 32 bf 6f 36  .R%..M...2..o6
```

[MS-FSSPRDF] — v20120630
SPRel Data File Formats

Copyright © 2012 Microsoft Corporation.

Release: July 16, 2012
The 128-bit MD5 hash value of the first record in section 3.3.1, "ssic://2146300409", is 0x025a728a0086c0491a829bb32176ea12. The 32 most significant bits of this hash value corresponds to the first 4 bytes in this file. The bytes are described in little-endian order.

### 3.3.3 <gen>.sharepoint.rel.<part_num>.idx.ofs

The following is raw data from the <gen>.sharepoint.rel. <part_num>.idx.ofs file.

The first 4 bytes, 0x000000, contain the offset to the first record entry in the <gen>.sharepoint.rel. <part_num>.bin file. The size of the file header, 124 bytes, and the record size integer, 4 bytes, are added to this offset. Thus, the first record begins at 128 bytes from the beginning of the file described in section 3.3.1. The following 4 bytes, 0x04000000, contain an offset of 4. This value is multiplied by 32, which results in an offset of 128. In addition, the header is added to this offset. Thus, the second record starts at 256 bytes from the start of the file.
4 Security Considerations

None.
5 Appendix A: Product Behavior

The information in this specification is applicable to the following Microsoft products or supplemental software. References to product versions include released service packs:

- Microsoft® FAST™ Search Server 2010

Exceptions, if any, are noted below. If a service pack or Quick Fix Engineering (QFE) number appears with the product version, behavior changed in that service pack or QFE. The new behavior also applies to subsequent service packs of the product unless otherwise specified. If a product edition appears with the product version, behavior is different in that product edition.

Unless otherwise specified, any statement of optional behavior in this specification that is prescribed using the terms SHOULD or SHOULD NOT implies product behavior in accordance with the SHOULD or SHOULD NOT prescription. Unless otherwise specified, the term MAY implies that the product does not follow the prescription.
6 Change Tracking

No table of changes is available. The document is either new or has had no changes since its last release.
7 Index
<col>_feeduris.<sf> 15
  example 28
<col>_feeduris_expand.<sf> 15
  example 28
<col>_feeduris_expand_resplit.<sf> 15
  example 28
<date>_clicks.txt 10
  example 21
<date>_clicks_by_urlid_and_queryid.<sf> 17
  example 32
<date>_clicks_by_urlid_and_queryid_sort.<sf>..<sf>
  > 17
  example 32
<date>_local_querycnt_by_queryid.<sf> 18
  example 33
<date>_local_querycnt_by_queryid_reduce.<sf>..<sf>
  > 17
  example 32
<date>_queries.txt 10
  example 21
<date>_queries_by_queryid.<sf> 18
  example 33
<date>_queries_by_queryid_sort.0..<sf>
  > 18
  example 33
<date>_queries_on_queryid.0 18
  example 34
<date>_urls.txt 10
  example 21
<date>_urls_by_urlid.<sf> 19
  example 34
<date>_urls_by_urlid_sort.0.<sf>
  > 19
  example 34
<date>_urls_on_urlid.0 19
  example 34
<gen>.<col>.<host>_contentids_by_contentid_new.<sf>
  > 16
  example 31
<gen>.<col>.<host>_contentids_by_contentid_new_resplit.<sf>..<sf>
  > 17
  example 31
<gen>.<col>.<host>_uris.0 17
  example 32
<gen>.<col>.unique_uris_by_url.<sf> 16
  example 30
<gen>.<col>.urls_by_contentid.<sf> 16
  example 31
<gen>.<col>.urls_by_contentid_ts..<sf>
  > 16
  example 31
<gen>.queries_by_queryid.<sf> 15
  example 28
<gen>.queries_by_queryid_all..<sf>
  > 15
  example 29
<gen>.queryinfo.<sf> 15
  example 29
<gen>.sharepoint.rel.<part_num>.bin 19
  example 35
<gen>.sharepoint.rel.<part_num>.idx 20
  example 36
<gen>.sharepoint.rel.<part_num>.idx.ofs 20
  example 37
<gen>.urls_by_urlid.<sf> 16
  example 30
<gen>.urls_by_urlid_all..<sf>
  > 16
  example 30
A
allfeeduris.<sf> 10
  example 22
Analysis files 10
<col>_feeduris.<sf> 15
<col>_feeduris_expand.<sf> 15
<col>_feeduris_expand_resplit.<sf> 15
<date>_clicks_by_urlid_and_queryid.<sf> 17
<date>_clicks_by_urlid_and_queryid_sort.<sf>..<sf>
  > 17
<date>_local_querycnt_by_queryid.<sf> 18
<date>_local_querycnt_by_queryid_reduce.<sf>..<sf>
  > 17
<date>_queries_by_queryid.<sf> 18
<date>_queries_by_queryid_sort.0..<sf>
  > 18
<date>_queries_on_queryid.0 18
<date>_urls_by_urlid..<sf>
  > 19
<date>_urls_by_urlid_sort.0..<sf>
  > 19
<date>_urls_on_urlid.0 19
<gen>.<col>.<host>_contentids_by_contentid_new.<sf>
  > 16
<gen>.<col>.<host>_contentids_by_contentid_new_resplit.<sf>..<sf>
  > 17
<gen>.<col>.<host> uris.0 17
<gen>.<col>.urls_by_urlid..<sf>
  > 16
<gen>.<col>.urls_by_urlid_all..<sf>
  > 16
allfeeduris.<sf> 10
  example 22
  example 36
Details
- Common file structures
- Common file naming conventions

Applicability 7

Change tracking 40

Analysis files
- Common data types and fields
- Common file structures
- Common file naming conventions

Data types and fields

- Common 8
- Database files 19

Database files
- Global querycnt by queryid.<sf> file 11
- Local querycnt by cid.<sf> file 11
- Local querycnt by cid_merge.<sf>.<sf> file 11
- Local querycnt by queryid.<sf> file 12
- Local querycnt by queryid_reduce.<sf>.<sf> file 11

Empty files 20

- Semi local querycnt by queryid.<sf> file 13
- Semi local querycnt by urlid.<sf> file 13
- Semi local querycnt by urlid_map.<sf>.<sf> file 13
- Semi local querycnt_pre_token.<sf> file 13

Examples (section 3 21, section 3 21)

- Analysis file

- Empty files 20

Copyright © 2012 Microsoft Corporation.

Release: July 16, 2012
F

<gen><col><host> contentids_by_contentid_new.<sf> 31
<gen><col><host> contentids_by_contentid_new_respli.t.<sf> 31
<gen><col><host> uris_0.<sf> 32
<gen><col> unique uris_by_uri.<sf> 30
<gen><col> uris_by_contentid.<sf> 31
<gen><col> queries_by_queryid.<sf> 28
<gen> queries_by_queryid.all.<sf> 29
<gen> queryinfo.<sf> 29
<gen> urls_by_urlid.<sf> 30
allfeeduris.<sf> 22
cid_by_cid_with_counts_and_query.<sf> 22
cid_by_cid_with_counts_and_query_mergeredu.ce.<sf> 22
global_querycnt_by_query.<sf> 22
local_querycnt_by_cid.<sf> 23
local_querycnt_by_cid_merge.<sf> 23
local_querycnt_by_query.<sf> 23
local_querycnt_by_query.reduce.<sf> 23
local_querycnt_by_url.<sf> 24
local_querycnt_by_url_merge.<sf> 24
local_querycnt_by_urlid.<sf> 24
local_querycnt_by_urlid_reduce.<sf> 24
semi_local_querycnt_by_queryid.<sf> 24
semi_local_querycnt_by_urlid.<sf> 25
semi_local_querycnt_by_urlid_map.<sf> 25
semi_local_querycnt_pre_token.<sf> 25
uris_by_contentid.<sf> 26
uris_by_member.<sf> 26
uris_by_member_reduce.<sf> 26
urls_by_urlhash_with_queries.<sf> 26
urls_by_urlhash_with_queries_sort.<sf> 27
urls_on_urlhash_with_queries.<sf> 27
Database file
<gen>sharepoint.rel.<part_num>.bin 35
<gen>sharepoint.rel.<part_num>.idx 36
<gen>sharepoint.rel.<part_num>.idx.ofs 37
search clickthrough file
<date> clicks.txt 21
<date> queries.txt 21
<date> urls.txt 21

Files vendor-extensible 7

col feeduris.<sf> 15
col feeduris expand.<sf> 15
col feeduris expand_resplilet.<sf> 15
<date> clicks.txt 10
<date> clicks_by_urlid_and_queryid.<sf> 17
<date> clicks_by_urlid_and_queryid_sort.<sf> 17
<date> clicks_on_queryid.0 17
<date> local_querycnt_by_queryid.<sf> 18
<date> local_querycnt_by_queryid_reduce.<sf> 18
<date> queries.txt 10
<date> queries_by_queryid.<sf> 18
<date> queries_by_queryid_sort.0.<sf> 18
<date> queries_on_queryid.0 18
<date> urls.txt 10
<date> urls_by_urlid.<sf> 19
<date> urls_by_urlid_sort.0.<sf> 19
<date> urls_on_urlid.0 19
<gen><col><host> contentids_by_contentid_new.<sf> 16
<gen><col><host> contentids_by_contentid_new_resplilet.<sf> 17
<gen><col><host> uris_0 17
<gen><col> unique uris_by_uri.<sf> 16
<gen><col> uris_by_contentid.<sf> 16
<gen><col> uris_by_contentid_ts.<sf> 16
<gen> queries_by_queryid.<sf> 15
<gen> queries_by_queryid_all.<sf> 15
<gen> queryinfo.<sf> 15
<gen> sharepoint.rel.<part_num>.bin 19
<gen> sharepoint.rel.<part_num>.idx 20
<gen> sharepoint.rel.<part_num>.idx.ofs 20
<date> urls_by_urlid.<sf> 16
<date> urls_by_urlid_all.<sf> 16
allfeeduris.<sf> 10
analysis 10
cid_by_cid_with_counts_and_query.<sf> 11
cid_by_cid_with_counts_and_query_mergereduc.e.<sf> 11
common naming conventions 9
database 19
empty 20
global_querycnt_by_query.<sf> 11
local_querycnt_by_cid.<sf> 11
local_querycnt_by_cid_merge.<sf> 11
local_querycnt_by_query.<sf> 12
local_querycnt_by_query.reduce.<sf> 12
local_querycnt_by_url.<sf> 12
local_querycnt_by_url_merge.<sf> 12
local_querycnt_by_urlid.<sf> 12
local_querycnt_by_urlid_reduce.<sf> 12
search clickthrough 10
semi_local_querycnt_by_queryid.<sf> 13
semi_local_querycnt_by_urlid.<sf> 13
semi_local_querycnt_by_urlid_map.<sf> 13
semi_local_querycnt_pre_token.<sf> 13
uris_by_contentid.<sf> 13
uris_by_member.<sf> 14
uris_by_member_reduce.<sf> 14
urls_by_urlhash_with_queries.<sf> 14
urls_by_urlhash_with_queries_sort.<sf> 14
urls_on_urlhash_with_queries.<sf> 14

G

global_querycnt_by_query.<sf> 11
example 22
Glossary 6

I

Implementer - security considerations 38
Informative references 7

Introduction 6

L

local_querycnt_by_cid.<sf> 11
    example 23
local_querycnt_by_cid_merge.<sf>.<sf> (section 2.4.6 11, section 2.4.7 12)
    example 23
local_querycnt_by_query.<sf>  
    example 23
local_querycnt_by_query_reduce.<sf>.<sf> 12
    example 23
local_querycnt_by_url.<sf> 12
    example 24
local_querycnt_by_url_merge.<sf>.<sf> 12
    example 24
local_querycnt_by_urlid.<sf> 12
    example 24
local_querycnt_by_urlid_reduce.<sf>.<sf> 13
    example 24
Localization 7

N

Normative references 6

O

Overview (synopsis) 7

P

Product behavior 39

R

References 6
    informative 7
    normative 6
Relationship to protocols and other structures 7

S

Search clickthrough files 10
   <date>.clicks.txt 10
   <date>.queries.txt 10
   <date>.urls.txt 10
Security - implementer considerations 38
semi_local_querycnt_by_queryid.<sf> 13
    example 24
semi_local_querycnt_by_urlid.<sf> 13
    example 25
semi_local_querycnt_by_urlid_map.<sf>.<sf> 13
    example 25
semi_local_querycnt_pre_token.<sf> 13
    example 25
Structures
    common file 8
    overview (section 2 8, section 2 8)