

# [MS-FSSPRDF]: SPRel Data Files Format Specification

---

## 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's Open Specification Promise (available here: <http://www.microsoft.com/interop/osp>) or the Community Promise (available here: <http://www.microsoft.com/interop/cp/default.msp>). 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](mailto: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

Date	Revision History	Revision Class	Comments
02/19/2010	1.0	Major	Initial Availability
03/31/2010	1.01	Editorial	Revised and edited the technical content
04/30/2010	1.02	Editorial	Revised and edited the technical content
06/07/2010	1.03	Editorial	Revised and edited the technical content
06/29/2010	1.04	Editorial	Changed language and formatting in the technical content.
07/23/2010	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
09/27/2010	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
11/15/2010	1.04	No change	No changes to the meaning, language, or formatting of the technical content.
12/17/2010	1.04	No change	No changes to the meaning, language, or formatting of the technical content.

# Table of Contents

<b>1 Introduction</b>	<b>6</b>
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
<b>2 Structures</b>	<b>8</b>
2.1 Common File Structures	8
2.2 Common File Naming Conventions	10
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	11
2.4 Analysis Files	11
2.4.1 allfeeduris.<sf> files	11
2.4.2 cid_by_cid_with_counts_and_query.<sf>	11
2.4.3 cid_by_cid_with_counts_and_query_mergereduce.<sf>.<sf>	11
2.4.4 global_querycnt_by_query.<sf>	11
2.4.5 local_querycnt_by_cid.<sf>	12
2.4.6 local_querycnt_by_cid_merge.<sf>.<sf>	12
2.4.7 local_querycnt_by_query.<sf>	12
2.4.8 local_querycnt_by_query_reduce.<sf>.<sf>	12
2.4.9 local_querycnt_by_url.<sf>	12
2.4.10 local_querycnt_by_url_merge.<sf>.<sf>	13
2.4.11 local_querycnt_by_urlid.<sf>	13
2.4.12 local_querycnt_by_urlid_reduce.<sf>.<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>.<sf>	14
2.4.16 semi_local_querycnt_pre_token.<sf>	14
2.4.17 uris_by_contentid_ts.<sf>	14
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>	15
2.4.22 urls_on_urlhash_with_queries.<sf>	15
2.4.23 <col>_feeduris.<sf>	15
2.4.24 <col>_feeduris_expand.<sf>	15
2.4.25 <col>_feeduris_expand_resplit.<sf>	15
2.4.26 <gen>.queries_by_queryid.<sf>	15
2.4.27 <gen>.queries_by_queryid_all.<sf>	16
2.4.28 <gen>.queryinfo.<sf>	16
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

2.4.32	<gen>.<col>.uris_by_contentid.<sf>	17
2.4.33	<gen>.<col>.uris_by_contentid_ts.<sf>	17
2.4.34	<gen>.<col>.<host>_contentids_by_contentid_new.<sf>	17
2.4.35	<gen>.<col>.<host>_contentids_by_contentid_new_resplit.<sf>.<sf>	17
2.4.36	<gen>.<col>.<host>_uris.0	17
2.4.37	<date>.clicks_by_urlid_and_queryid.<sf>	17
2.4.38	<date>.clicks_by_urlid_and_queryid_sort.<sf>.<sf>	18
2.4.39	<date>.clicks_on_queryid.0	18
2.4.40	<date>.local_querycnt_by_queryid.<sf>	18
2.4.41	<date>.local_querycnt_by_queryid_reduce.<sf>.<sf>	18
2.4.42	<date>.queries_by_queryid.<sf>	18
2.4.43	<date>.queries_by_queryid_sort.0.<sf>	19
2.4.44	<date>.queries_on_queryid.0	19
2.4.45	<date>.urls_by_urlid.<sf>	19
2.4.46	<date>.urls_by_urlid_sort.0.<sf>	19
2.4.47	<date>.urls_on_urlid.0	19
2.5	Database Files	19
2.5.1	<gen>.sharepoint.rel.<part_num>.bin	20
2.5.2	<gen>.sharepoint.rel.<part_num>.idx	20
2.5.3	<gen>.sharepoint.rel.<part_num>.idx ofs	20
2.6	Empty Files	20
<b>3</b>	<b>Structure Examples</b>	<b>21</b>
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_mergereducer.<sf>.<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>.<sf>	23
3.2.7	local_querycnt_by_query.<sf>	23
3.2.8	local_querycnt_by_query_reduce.<sf>.<sf>	23
3.2.9	local_querycnt_by_url.<sf>	24
3.2.10	local_querycnt_by_url_merge.<sf>.<sf>	24
3.2.11	local_querycnt_by_urlid.<sf>	24
3.2.12	local_querycnt_by_urlid_reduce.<sf>.<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>.<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_expand.<sf>	28
3.2.25	<col>_feeduris_expand_resplit.<sf>	28

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_uri.<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
<b>4</b>	<b>Security Considerations.....</b>	<b>38</b>
<b>5</b>	<b>Appendix A: Product Behavior.....</b>	<b>39</b>
<b>6</b>	<b>Change Tracking.....</b>	<b>40</b>
<b>7</b>	<b>Index.....</b>	<b>41</b>

# 1 Introduction

This document specifies the file names and file formats used for client click input analysis. This analysis improves search relevancy by analyzing client feedback.

## 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**  
**Base 64 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

### 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](mailto: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](#)", November 2009.

[MS-FSWCU] Microsoft Corporation, "[WebAnalyzer/Crawler Utility Structure Specification](#)", November 2009.

[MS-LCID] Microsoft Corporation, "[Windows Language Code Identifier \(LCID\) Reference](#)", July 2007.

[RFC1950] Deutsch, P., and Gailly, J-L., "ZLIB Compressed Data Format Specification version 3.3", RFC 1950, May 1996, <http://www.ietf.org/rfc/rfc1950.txt>

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, <http://www.ietf.org/rfc/rfc2119.txt>

[RFC5234] Crocker, D., Ed., and Overell, P., "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, January 2008, <http://www.ietf.org/rfc/rfc5234.txt>

### 1.2.2 Informative References

[MS-FSFDMMW] Microsoft Corporation, "[FAST Distributed Make Worker Protocol Specification](#)", February 2010.

[MS-GLOS] Microsoft Corporation, "[Windows Protocols Master Glossary](#)", March 2007.

[MS-OFGLGLOS] Microsoft Corporation, "[Microsoft Office Master Glossary](#)", June 2008.

### 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 [\[MS-FSFDMMW\]](#).

### 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

; Section [2.3](#) and [2.4](#) specify ABNF rules for each type of ROW.

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.

Column	Description
<b>CID</b>	Specifies a <b>document identifier(3)</b> .
<b>CLICKEDURLRANK</b>	Specifies the rank of a document in the search result.
<b>COLLECTION</b>	Specifies a <b>content collection</b> name.
<b>DATE</b>	Specifies a processing date. This date MUST be formatted as yyyy.mm.dd.
<b>GCOUNT</b>	Specifies the number of times a search query was processed.
<b>URLHASH</b>	Specifies a 128-bit <b>MD5 hash</b> value of a URL. The hash value is represented as a numerical value.
<b>NUMURLS</b>	Specifies the total number of URLs associated with a specific search query.
<b>LCID</b>	Specifies a language code identifier, as specified in <a href="#">[MS-LCID]</a> .
<b>LCOUNT</b>	Specifies the number of times a URL was clicked for a particular search query.
<b>MEMBER</b>	Specifies a URL or document identifier that is in the <b>equivalence class</b> of the document. The <b>MEMBER</b> column MUST contain the value 0xc782, as specified section 2.2.1.3.2.1.4 in <a href="#">[MS-FSWASDR]</a> , when the <b>OP</b> column contains a DEL operation.
<b>OP</b>	Specifies a content submission operation. Values are: <ul style="list-style-type: none"><li>- ADD: A new document was added to the index.</li><li>- DEL: A document was deleted from the index</li></ul>
<b>QUERIES</b>	Specifies all search queries associated with a particular document identifier. This column contains a <b>base 64 encoded dictionary</b> , as specified in <a href="#">[MS-FSWCU]</a> . This dictionary MUST contain two keys: <b>contentid</b> and <b>queries</b> . The <b>contentid</b> key specifies a document identifier (CID). The value of the "queries" key MUST contain an array, as specified in <a href="#">[MS-FSWCU]</a> , of query entries. Each query entry MUST be a five entry tuple, as specified in <a href="#">[MS-FSWCU]</a> . Each tuple MUST contain the following column values; LCOUNT, GCOUNT, PLACE, NUMURLS and QUERY.
<b>QUERIESINFO</b>	Specifies search queries and collections associated with a particular document identifier. This column contains a base 64 encoded dictionary as specified for the <b>QUERIES</b> column. However, this dictionary MUST have one additional key, the <b>collections</b> key. The <b>collections</b> key MUST contain an array of all the collections, as specified in <a href="#">[MS-FSWCU]</a> , where this document identifier is registered.
<b>QUERY</b>	Specifies an ASCII encoded search query string.

Column	Description
<b>QUERYID</b>	Specifies a unique identifier for a search query.
<b>TIMESTAMP</b>	Specifies a processing timestamp. This is a numerical value, it specifies the time in seconds after 00:00:00 1970-01-01 <b>UTC</b> .
<b>URL</b>	Specifies an <b>absolute URL</b> .
<b>URLID</b>	Specifies an unique identifier for a URL.
<b>PLACE</b>	Specifies a numerical rank. This field ranks URLs for a specific search query.
<b>ZERO</b>	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 `yyyymmdd`.
- `<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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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.

```
ROW = CID SP TIMESTAMP SP OP SP MEMBER
```

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

For more information on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 on 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 marshalled, 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 two first 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 **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 a ".end" file extension MUST be empty. In addition, the following files MUST be empty: 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.

```
2000 2009.03.16 http://www.alpineskihouse.com/2000
2001 2009.03.17 http://www.alpineskihouse.com/2001
2002 2009.03.18 http://www.alpineskihouse.com/2002
2003 2009.03.19 http://www.alpineskihouse.com/2003
2004 2009.03.20 http://www.alpineskihouse.com/2004
2005 2009.03.21 http://www.alpineskihouse.com/2005
2006 2009.03.22 http://www.alpineskihouse.com/2006
2007 2009.03.23 http://www.alpineskihouse.com/2007
2008 2009.03.24 http://www.alpineskihouse.com/2008
```

## 3.2 Analysis Files

### 3.2.1 allfeeduris.<sf> files

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

```
ssic://2143300394 sp
ssic://2143300398 sp
ssic://2143300399 sp
ssic://2145300400 sp
ssic://2145300402 sp
ssic://2145300405 sp
ssic://2145300407 sp
ssic://2146300404 sp
```

### 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.

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

### 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.

```
ssic://2145300405 3 3 1 1 query1003 scope all sites
ssic://2145300400 6 6 1 1 query1006 scope all sites
ssic://2146300404 15 15 1 1 query1015 scope all sites
ssic://2143300394 22 22 1 1 query1022 scope all sites
ssic://2143300398 26 26 1 1 query1026 scope all sites
```

### 3.2.4 global\_querycnt\_by\_query.<sf>

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

```
2 query1002 scope all sites
3 query1003 scope all sites
5 query1005 scope all sites
6 query1006 scope all sites
9 query1009 scope all sites
13 query1013 scope all sites
14 query1014 scope all sites
15 query1015 scope all sites
19 query1019 scope all sites
22 query1022 scope all sites
25 query1025 scope all sites
```

```
26 query1026 scope all sites
```

### 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
```



```
2004 1004 1 1
2004 1004 1 1
2006 1006 1 1
2006 1006 1 1
2006 1006 1 1
2006 1006 1 1
2006 1006 1 1
2006 1006 1 1
2006 1006 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
```

### 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"
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
ssic://2145300400 1239888256 ADD http://www.alpineskihouse.com/2006
ssic://2145300402 1239888256 ADD http://www.alpineskihouse.com/2004
ssic://2145300405 1239888256 ADD http://www.alpineskihouse.com/2003
```

### 3.2.18 uris\_by\_member.<sf>

The following is an example of a uris\_by\_member.<sf> file.

```
ssic://2145300405 http://www.alpineskihouse.com/2003
ssic://2145300401 http://www.alpineskihouse.com/2007
ssic://2146300400 http://www.alpineskihouse.com/2011
ssic://2146300405 http://www.alpineskihouse.com/2014
ssic://2146300407 http://www.alpineskihouse.com/2016
ssic://2143300393 http://www.alpineskihouse.com/2021
ssic://2143300394 http://www.alpineskihouse.com/2022
ssic://2143300398 http://www.alpineskihouse.com/2026
```

### 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
ssic://2145300405 http://www.alpineskihouse.com/2003
```

### 3.2.20 urls\_by\_urlhash\_with\_queries.<sf>

The following is an example of a urls\_by\_urlhash\_with\_queries.<sf> file.

```
28768632944601236036478028165026923624
e3MJAAAAY29udGVudG1kcxEAAABzc2ljOi8vMjE0MzMDM4NDNMHAAAAcXVlcml1c1sBAAAAKAUAAABzAgAAADI5cwIAA
AAyOXMBAAAAMXMBAAAAMXMAAAAACXVlcnkxMDI5IHNjb3BlIGFsbCBzaXRlc2A=
36707531866801635426375018499063369623
e3MJAAAAY29udGVudG1kcxEAAABzc2ljOi8vMjE0NjMwMDQwN3MHAAAAcXVlcml1c1sBAAAAKAUAAABzAgAAADE2cwIAA
AAxNmBAAAAMXMBAAAAMXMAAAAACXVlcnkxMDE2IHNjb3BlIGFsbCBzaXRlc2A=
49607864796389356377605213683634923015
e3MJAAAAY29udGVudG1kcxEAAABzc2ljOi8vMjE0NjMwMDQwOHMHAAAAcXVlcml1c1sBAAAAKAUAAABzAgAAADE5cwIAA
AAxOXMBAAAAMXMBAAAAMXMAAAAACXVlcnkxMDE5IHNjb3BlIGFsbCBzaXRlc2A=
3128085837182113593997925733612579346
e3MJAAAAY29udGVudG1kcxEAAABzc2ljOi8vMjE0NjMwMDQwOXMHAAAAcXVlcml1c1sBAAAAKAUAAABzAgAAADE4cwIAA
AAxOHMBAAAAMXMBAAAAMXMAAAAACXVlcnkxMDE4IHNjb3BlIGFsbCBzaXRlc2A=
5406235469349875302439361261800735185
e3MJAAAAY29udGVudG1kcxEAAABzc2ljOi8vMjE0NTMwMDQwMXMHAAAAcXVlcml1c1sBAAAAKAUAAABzAQAAADdzAQAAA
DdzAQAAADfzAQAAADfzGQAAAHF1ZXJ5MTAwNyBzY29wZSBhbGwgc210ZXMw
```

Using the **QUERIES** 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 e3MJAAAAAY29udGVu
0000010: 64 47 6c 6b 63 78 45 41 41 41 42 7a 63 32 6c 6a dG1kcxEAAABzc21j
0000020: 4f 69 38 76 4d 6a 45 30 4d 7a 4d 77 4d 44 4d 34 Oi8vMjE0MzMwMDM4
0000030: 4e 58 4d 48 41 41 41 41 63 58 56 6c 63 6d 6c 6c NXMHAAAAcXVlcmll
0000040: 63 31 73 42 41 41 41 41 4b 41 55 41 41 41 42 7a c1sBAAAAKAUAAABz
0000050: 41 67 41 41 41 44 49 35 63 77 49 41 41 41 41 79 AgAAADI5cwIAAAAY
0000060: 4f 58 4d 42 41 41 41 41 4d 58 4d 42 41 41 41 41 OXMBAAAAAMXMBAAAA
0000070: 4d 58 4d 5a 41 41 41 41 63 58 56 6c 63 6e 6b 78 MXMZAAAAcXVlcnkx
0000080: 4d 44 49 35 49 48 4e 6a 62 33 42 6c 49 47 46 73 MDI5IHNjb3BlIGFs
0000090: 62 43 42 7a 61 58 52 6c 63 7a 41 3d bCBzaXRlcZA=

```

Decoding this information using base 64 encoding results in the following raw data.

```

0000000: 7b 73 09 00 00 00 63 6f 6e 74 65 6e 74 69 64 73 {s....contentids
0000010: 11 00 00 00 73 73 69 63 3a 2f 2f 32 31 34 33 33 ...ssic://21433
0000020: 30 30 33 38 35 73 07 00 00 00 71 75 65 72 69 65 00385s....querie
0000030: 73 5b 01 00 00 00 28 05 00 00 00 73 02 00 00 00 s[....(....s....
0000040: 32 39 73 02 00 00 00 32 39 73 01 00 00 00 31 73 29s....29s....1s
0000050: 01 00 00 00 31 73 19 00 00 00 71 75 65 72 79 31 ....1s....query1
0000060: 30 32 39 20 73 63 6f 70 65 20 61 6c 6c 20 73 69 029 scope all si
0000070: 74 65 73 30 tes0

```

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.

```

262235504639975156713625869717383236722
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0MzMwMDM5NHMHAAAAcXVlcmllc1sBAAAAKAUAAABzAgAAADIYcwIAA
AAyMnMBAAAAAMXMBAAAAAMXMAAAAacXVlcnkxMDIyIHNjb3BlIGFsbcBzaXRlcZA=
11414261698623317584290355546833026730
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0MzMwMDM5OHMHAAAAcXVlcmllc1sBAAAAKAUAAABzAgAAADI2cwIAA
AAyNnMBAAAAAMXMBAAAAAMXMAAAAacXVlcnkxMDI2IHNjb3BlIGFsbcBzaXRlcZA=
214165211352470348401251767931789719422
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0MzMwMDM5OXMHAAAAcXVlcmllc1sBAAAAKAUAAABzAgAAADI3cwIAA
AAyN3MBAAAAAMXMBAAAAAMXMAAAAacXVlcnkxMDI3IHNjb3BlIGFsbcBzaXRlcZA=
162766349373572592373858349095155889970
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0NTMwMDQwMHMHAAAAcXVlcmllc1sBAAAAKAUAAABzAQAAADZzAQAAA
DZzAQAAADfzAQAAADfzQAAAHF1ZXJ5MTAwNiBzY29wZSBhbGwgc210ZXMw
197388348979834279235465721912469386177

```

### 3.2.22 urls\_on\_urlhash\_with\_queries.<sf>

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

```

262235504639975156713625869717383236722
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0MzMwMDM5NHMHAAAAcXVlcmllc1sBAAAAKAUAAABzAgAAADIYcwIAA
AAyMnMBAAAAAMXMBAAAAAMXMAAAAacXVlcnkxMDIyIHNjb3BlIGFsbcBzaXRlcZA=

```

```
114142616986233175842903555546833026730
e3MJAAAAY29udGVudG1kcxEAAABzc2ljOi8vMjE0MzMDM5OHMHAACXVlcml1c1sBAAAKAUAAABzAgAAADI2cwIAA
AAyNnMBAAAAMXMBAAAAMXMAAAAcXVlcnkxMDI2IHJb3BlIGFsbCBzaXRlc2A=
214165211352470348401251767931789719422
e3MJAAAAY29udGVudG1kcxEAAABzc2ljOi8vMjE0MzMDM5OXMHAACXVlcml1c1sBAAAKAUAAABzAgAAADI3cwIAA
AAyN3MBAAAAMXMBAAAAMXMAAAAcXVlcnkxMDI3IHJb3BlIGFsbCBzaXRlc2A=
162766349373572592373858349095155889970
e3MJAAAAY29udGVudG1kcxEAAABzc2ljOi8vMjE0MzMDQwMHMHAACXVlcml1c1sBAAAKAUAAABzAQAAADZzAQAAA
DZzAQAAADZzAQAAADZzAQAAAHF1ZXJ5MTAwNiBzY29wZSBhbGwgc2l0ZXMw
```

### 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://2145300407 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"
```

```
1027 2009.04.11 query1027 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
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0MzMwMDM5OHMLAAAAAY29sbGVjdGlvbnNbaQAAAHMCAAAAc3BzBwAAAHF1ZXJpZXNbaQAAACgFAAAAcwIAAAAYMnMCAAAAMjJzAQAAADfzAQAAADfzGQAAAHF1ZXJ5MTAyMiBzY29wZSBhbGwgc210ZXMw
0
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0MzMwMDM5OHMLAAAAAY29sbGVjdGlvbnNbaQAAAHMCAAAAc3BzBwAAAHF1ZXJpZXNbaQAAACgFAAAAcwIAAAAYNnMCAAAAMjZzAQAAADfzAQAAADfzGQAAAHF1ZXJ5MTAyNiBzY29wZSBhbGwgc210ZXMw
0
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0MzMwMDM5OXMLAAAAAY29sbGVjdGlvbnNbaQAAAHMCAAAAc3BzBwAAAHF1ZXJpZXNbaQAAACgFAAAAcwIAAAAYN3MCAAAAMjdzAQAAADfzAQAAADfzGQAAAHF1ZXJ5MTAyNyBzY29wZSBhbGwgc210ZXMw
0
e3MJAAAAAY29udGVudG1kcxEAAABzc21jOi8vMjE0NTMwMDQwMHMLAAAAAY29sbGVjdGlvbnNbaQAAAHMCAAAAc3BzBwAAAHF1ZXJpZXNbaQAAACgFAAAAcwEAAA2cwEAAA2cwEAAAxcwEAAAxcxkAAABxdWVyeTEwMDY
```

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 e3MJAAAAAY29udGVu
0000010: 64 47 6c 6b 63 78 45 41 41 41 42 7a 63 32 6c 6a dG1kcxEAAABzc21j
0000020: 4f 69 38 76 4d 6a 45 30 4d 7a 4d 77 4d 44 4d 35 Oi8vMjE0MzMwMDM5
0000030: 4e 48 4d 4c 41 41 41 41 59 32 39 73 62 47 56 6a NHMLAAAAAY29sbGVj
0000040: 64 47 6c 76 62 6e 4e 62 41 51 41 41 41 48 4d 43 dGlvbnNbaQAAAHMC
0000050: 41 41 41 41 63 33 42 7a 42 77 41 41 41 48 46 31 AAAAc3BzBwAAAHF1
0000060: 5a 58 4a 70 5a 58 4e 62 41 51 41 41 41 43 67 46 ZXJpZXNbaQAAACgF
0000070: 41 41 41 41 63 77 49 41 41 41 41 79 4d 6e 4d 43 AAAAcwIAAAAYMnMC
0000080: 41 41 41 41 4d 6a 4a 7a 41 51 41 41 41 44 46 7a AAAAMjJzAQAAADfz
0000090: 41 51 41 41 41 44 46 7a 47 51 41 41 41 48 46 31 AQAAADfzGQAAAHF1
00000a0: 5a 58 4a 35 4d 54 41 79 4d 69 42 7a 59 32 39 77 ZXJ5MTAyMiBzY29w
00000b0: 5a 53 42 68 62 47 77 67 63 32 6c 30 5a 58 4d 77 ZSBhbGwgc210ZXMw
00000c0: 0a .
```

Decoding this information using base 64 encoding results in the following raw data.

```

0000000: 7b 73 09 00 00 00 63 6f 6e 74 65 6e 74 69 64 73 {s....contentids
0000010: 11 00 00 00 73 73 69 63 3a 2f 2f 32 31 34 33 33 ....ssic://21433
0000020: 30 30 33 39 34 73 0b 00 00 00 63 6f 6c 6c 65 63 00394s....collec
0000030: 74 69 6f 6e 73 5b 01 00 00 00 73 02 00 00 00 73 tions[....s....s
0000040: 70 73 07 00 00 00 71 75 65 72 69 65 73 5b 01 00 ps....queries[.
0000050: 00 00 28 05 00 00 00 73 02 00 00 00 32 32 73 02 ..(....s....22s.
0000060: 00 00 00 32 32 73 01 00 00 00 31 73 01 00 00 00 ...22s....1s....
0000070: 31 73 19 00 00 00 71 75 65 72 79 31 30 32 32 20 1s....query1022
0000080: 73 63 6f 70 65 20 61 6c 6c 20 73 69 74 65 73 30 scope all sites0

```

Deserializing the preceding data, as described in [\[MS-FSWCU\]](#), results in the following string for the **QUERIESINFO** 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.

```

2001 2009.03.16 http://www.alpineskihouse.com/2001
2003 2009.03.18 http://www.alpineskihouse.com/2003
2006 2009.03.21 http://www.alpineskihouse.com/2006
2009 2009.03.24 http://www.alpineskihouse.com/2009
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.

```

2001 2009.03.16 http://www.alpineskihouse.com/2001
2003 2009.03.18 http://www.alpineskihouse.com/2003
2006 2009.03.21 http://www.alpineskihouse.com/2006
2009 2009.03.24 http://www.alpineskihouse.com/2009
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
ssic://2145300400 1239888256 ADD http://www.alpineskihouse.com/2006
ssic://2145300402 1239888256 ADD http://www.alpineskihouse.com/2004
ssic://2145300405 1239888256 ADD http://www.alpineskihouse.com/2003
```

### 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 †
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
ssic://2145300400 1239888256 ADD http://www.alpineskihouse.com/2006
ssic://2145300402 1239888256 ADD http://www.alpineskihouse.com/2004
ssic://2145300405 1239888256 ADD http://www.alpineskihouse.com/2003
ssic://2145300406 1239888256 ADD http://www.alpineskihouse.com/2000
ssic://2145300407 1239888256 ADD http://www.alpineskihouse.com/2001
ssic://2146300404 1239888256 ADD http://www.alpineskihouse.com/2015
```

### 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 †
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
ssic://2145300400 1239888256 ADD http://www.alpineskihouse.com/2006
ssic://2145300402 1239888256 ADD http://www.alpineskihouse.com/2004
ssic://2145300405 1239888256 ADD http://www.alpineskihouse.com/2003
ssic://2145300406 1239888256 ADD http://www.alpineskihouse.com/2000
ssic://2145300407 1239888256 ADD http://www.alpineskihouse.com/2001
ssic://2146300404 1239888256 ADD http://www.alpineskihouse.com/2015
```

### 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://2145300406 1239888256 ADD http://www.alpineskihouse.com/2000
ssic://2145300407 1239888256 ADD http://www.alpineskihouse.com/2001
ssic://2145300405 1239888256 ADD http://www.alpineskihouse.com/2003
ssic://2145300402 1239888256 ADD http://www.alpineskihouse.com/2004
ssic://2145300400 1239888256 ADD http://www.alpineskihouse.com/2006
ssic://2146300404 1239888256 ADD http://www.alpineskihouse.com/2015
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
ssic://1134254170 1239888257 DEL ‡
```

### 3.2.36 <gen>.<col>.<host>\_uris.0

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

```
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
```



```
1004 2004 1 1
1005 2005 1 1
1006 2006 1 1
1007 2007 1 1
1008 2008 1 1
1009 2009 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.

```
2002 1002 1 1
2004 1004 1 1
2006 1006 1 1
2011 1011 1 1
2018 1018 1 1
2020 1020 1 1
2022 1022 1 1
2023 1023 1 1
```

### 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.

```
2006 1006 1 1
2025 1025 1 1
2027 1027 1 1
```

### 3.2.42 <date>.queries\_by\_queryid.<sf>

The following is an example of a <date>.queries\_by \_queryid .<sf> 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"
```

### 3.2.43 <date>.queries\_by \_queryid\_sort.0.<sf>

The following is an example of a <date>.queries\_by \_queryid\_sort.0.<sf> 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"
```

### 3.2.44 <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.

```
2000 2009.03.16 http://www.alpineskihouse.com/2000
2001 2009.03.17 http://www.alpineskihouse.com/2001
2002 2009.03.18 http://www.alpineskihouse.com/2002
2003 2009.03.19 http://www.alpineskihouse.com/2003
2004 2009.03.20 http://www.alpineskihouse.com/2004
2005 2009.03.21 http://www.alpineskihouse.com/2005
2006 2009.03.22 http://www.alpineskihouse.com/2006
```

### 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.

```
2000 2009.03.16 http://www.alpineskihouse.com/2000
2001 2009.03.17 http://www.alpineskihouse.com/2001
2002 2009.03.18 http://www.alpineskihouse.com/2002
2003 2009.03.19 http://www.alpineskihouse.com/2003
2004 2009.03.20 http://www.alpineskihouse.com/2004
2005 2009.03.21 http://www.alpineskihouse.com/2005
2006 2009.03.22 http://www.alpineskihouse.com/2006
```

### 3.2.47 <date>.urls\_on\_urlid.0

The following is an example of a <date>.urls\_on\_urlid.0 file.

```
2000 2009.03.16 http://www.alpineskihouse.com/2000
2001 2009.03.17 http://www.alpineskihouse.com/2001
2002 2009.03.18 http://www.alpineskihouse.com/2002
2003 2009.03.19 http://www.alpineskihouse.com/2003
2004 2009.03.20 http://www.alpineskihouse.com/2004
2005 2009.03.21 http://www.alpineskihouse.com/2005
2006 2009.03.22 http://www.alpineskihouse.com/2006
```

### 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....serializ
0000040: 65 72 73 0d 00 00 00 70 79 66 61 73 74 6d 61 72 ers....pyfastmar
0000050: 73 68 61 6c 73 10 00 00 00 63 6f 6d 70 72 65 73 shals....compres
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 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c +.L).....3.....
00000a0: 4d cc 8c 0d 0c 4c 0c 2c 8b d9 81 a2 85 a5 a9 45 M....L.,.....E
00000b0: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 b4 80 .....@..+H...0..
00000c0: 51 20 71 43 28 29 09 55 5f 69 68 60 68 a1 50 9c Q qC().U_ih`h.P.
00000d0: 9c 5f 90 aa 90 98 93 a3 50 9c 59 92 5a 6c 00 00 ._.P.Y.Zl..
00000e0: e5 2d 1a 29 00 00 00 00 00 00 00 00 00 00 00 00 .-.).....
00000f0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
0000100: 5c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd \.....d`H..+I.
0000110: 2b c9 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c +.L).....3.....
0000120: 4d 4c 8d 0d 0c 4c 0c 0c 8b d9 81 a2 85 a5 a9 45 ML...L.....E
0000130: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 88 65 8e 44 1a .....@..+H...e.D.
0000140: 42 49 49 a8 ea 4a 43 03 03 73 85 e2 e4 fc 82 54 BII..JC..s.....T
0000150: 85 c4 9c 1c 85 e2 cc 92 d4 62 03 00 b2 40 19 b8 .....b...@..
0000160: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd |.....d`H..+I.
0000170: 2b c9 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c +.L).....3.....
0000180: 4d 8c 8d 0d 0c 8c 2d 4c 8b d9 81 a2 85 a5 a9 45 M.....-L.....E
0000190: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 b2 84 .....@..+H...0..
00001a0: 51 20 71 43 28 29 09 55 5f 69 68 60 64 a9 50 9c Q qC().U_ih`d.P.
00001b0: 9c 5f 90 aa 90 98 93 a3 50 9c 59 92 5a 6c 00 00 ._.P.Y.Zl..
00001c0: e6 06 1a 2f 00 00 00 00 00 00 00 00 00 00 00 00 .../.....
00001d0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
00001e0: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd |.....d`H..+I.
00001f0: 2b c9 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c +.L).....3.....
0000200: 4d cc 8c 0d 0c 4c 0c cc 8b d9 81 a2 85 a5 a9 45 M....L.....E
0000210: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 34 83 .....@..+H...04.
0000220: 51 20 71 43 28 29 09 55 5f 69 68 60 68 a6 50 9c Q qC().U_ih`h.P.
0000230: 9c 5f 90 aa 90 98 93 a3 50 9c 59 92 5a 6c 00 00 ._.P.Y.Zl..
0000240: e3 ab 1a 21 00 00 00 00 00 00 00 00 00 00 00 00 ...!.....
0000250: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
0000260: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd |.....d`H..+I.
0000270: 2b c9 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c +.L).....3.....
0000280: 4d cc 8c 0d 0c 4c 0c 2c 8a d9 81 a2 85 a5 a9 45 M....L.,.....E
0000290: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 b4 84 .....@..+H...0..
00002a0: 51 20 71 43 28 29 09 55 5f 69 68 60 68 a9 50 9c Q qC().U_ih`h.P.
00002b0: 9c 5f 90 aa 90 98 93 a3 50 9c 59 92 5a 6c 00 00 ._.P.Y.Zl..
00002c0: e5 4e 1a 2b 00 00 00 00 00 00 00 00 00 00 00 00 .N.+.....
00002d0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
00002e0: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd |.....d`H..+I.
00002f0: 2b c9 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c +.L).....3.....
0000300: 4d cc 8c 0d 0c 4c 0c 8c 8a d9 81 a2 85 a5 a9 45 M....L.....E
0000310: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 34 86 .....@..+H...04.
0000320: 51 20 71 43 28 29 09 55 5f 69 68 60 68 ac 50 9c Q qC().U_ih`h.P.
0000330: 9c 5f 90 aa 90 98 93 a3 50 9c 59 92 5a 6c 00 00 ._.P.Y.Zl..
```

```

0000340: e0 c8 1a 13 00 00 00 00 00 00 00 00 00 00 00 .....
0000350: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
0000360: 7c 00 00 00 ab 2e e6 64 60 60 48 ce cf 2b 49 cd |.....d`H..+I.
0000370: 2b c9 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c +.L).....3.....
0000380: 4d 8c 8d 0d 0c 8c 2d cd 8a d9 81 a2 85 a5 a9 45 M.....-.....E
0000390: 99 a9 c5 d1 8c 40 b6 06 2b 48 19 13 90 30 32 81 .....@..+H...02.
00003a0: 51 20 71 43 28 29 09 55 5f 69 68 60 64 a2 50 9c Q qC().U_ih`d.P.

```

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 x....d`H..+I.+
0000010: 4c 29 16 04 f2 8a 8b 33 93 ad f4 f5 8d 0c 4d cc L).....3.....M.
0000020: 8c 0d 0c 4c 0c 2c 8b d9 81 a2 85 a5 a9 45 99 a9 ...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 6c 00 00 e5 2d .....P.Y.Zl...-
0000060: 1a 29 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 {s....contentids
0000010: 11 00 00 00 73 73 69 63 3a 2f 2f 32 31 34 36 33 ....ssic://21463
0000020: 30 30 34 30 39 73 07 00 00 00 71 75 65 72 69 65 00409s....querie
0000030: 73 5b 01 00 00 00 28 05 00 00 00 73 02 00 00 00 s[....(....s....
0000040: 31 38 73 02 00 00 00 31 38 73 01 00 00 00 31 73 18s....18s....1s
0000050: 01 00 00 00 31 73 19 00 00 00 71 75 65 72 79 31 ....1s....query1
0000060: 30 31 38 20 73 63 6f 70 65 20 61 6c 6c 20 73 69 018 scope all si
0000070: 74 65 73 30 tes0

```

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

```

{'contentid': 'ssic://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 0d9 d 9d 1b .rZ..3.....
0000010: 92 20 52 25 f8 4d be 2c 3e d3 fd 32 b9 bf 6f 36 . R%.M.,>..2..o6

```

```

0000020: 5a c8 98 38 ca ef 08 43 82 11 df 55 95 94 b2 62 Z..8...C...U...b
0000030: de 1f 8d 64 ad a3 fb 77 9b a8 73 7a c1 9d 2d 7f ...d...w...sz...-
0000040: fa 9c 7f 94 27 b8 1e a1 a3 7f 57 a5 1d ae e3 a8 ....'.....W.....
0000050: 78 81 0c b6 fe 2b 78 b6 ac b5 1b b8 9b c1 30 b9 x....+x.....0.
0000060: 15 52 55 bd 99 b8 48 c5 cb c4 9d ca 87 08 42 e7 .RU...H.....B.
0000070: 4f ba 17 eb 55 e9 dc ef O...U...

```

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.

```

0000000: 00 00 00 00 04 00 00 00 07 00 00 00 0b 00 00 00 .....
0000010: 0f 00 00 00 13 00 00 00 17 00 00 00 1b 00 00 00 .....
0000020: 1e 00 00 00 21 00 00 00 25 00 00 00 29 00 00 00 ....!...%...)...
0000030: 2d 00 00 00 31 00 00 00 35 00 00 00 38 00 00 00 -...1...5...8...
0000040: 3b 00 00 00 3e 00 00 00 42 00 00 00 45 00 00 00 ;...>...B...E...
0000050: 49 00 00 00 4d 00 00 00 51 00 00 00 55 00 00 00 I...M...Q...U...
0000060: 59 00 00 00 5c 00 00 00 60 00 00 00 64 00 00 00 Y...\...`...d...
0000070: 68 00 00 00 6c 00 00 00 h...l...

```

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, MUST be 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>](#)  
  [≥](#) 18  
  [example](#) 32  
[<date>.clicks on queryid.0](#) 18  
  [example](#) 32  
[<date>.local querycnt by queryid.<sf>](#) 18  
  [example](#) 33  
[<date>.local querycnt by queryid reduce.<sf>.<sf>](#)  
  [f>](#) 18  
  [example](#) 33  
[<date>.queries.txt](#) 10  
  [example](#) 21  
[<date>.queries by queryid.<sf>](#) 18  
  [example](#) 33  
[<date>.queries by queryid sort.0.<sf>](#) 19  
  [example](#) 33  
[<date>.queries on queryid.0](#) 19  
  [example](#) 34  
[<date>.urls.txt](#) 11  
  [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>](#) 17  
  [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 uri.<sf>](#) 16  
  [example](#) 30  
[<gen>.<col>.uris by contentid.<sf>](#) 17  
  [example](#) 31  
[<gen>.<col>.uris by contentid ts.<sf>](#) 17  
  [example](#) 31  
[<gen>.queries by queryid.<sf>](#) 15  
  [example](#) 28  
[<gen>.queries by queryid all.<sf>](#) 16  
  [example](#) 29  
[<gen>.queryinfo.<sf>](#) 16  
  [example](#) 29  
[<gen>.sharepoint.rel.<part\\_num>.bin](#) 20  
  [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>](#) 11  
  [example](#) 22  
[Analysis files](#) 11  
  [<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>](#)  
  [sf>](#) 18  
  [<date>.clicks on queryid.0](#) 18  
  [<date>.local querycnt by queryid.<sf>](#) 18  
  [<date>.local querycnt by queryid reduce.<sf>.<sf>](#)  
  [sf>](#) 18  
  [<date>.queries by queryid.<sf>](#) 18  
  [<date>.queries by queryid sort.0.<sf>](#) 19  
  [<date>.queries on queryid.0](#) 19  
  [<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>](#) 17  
  [<gen>.<col>.<host> contentids by contentid new resplit.<sf>.<sf>](#) 17  
  [<gen>.<col>.<host> uris.0](#) 17  
  [<gen>.<col>.unique uris by uri.<sf>](#) 16  
  [<gen>.<col>.uris by contentid.<sf>](#) 17  
  [<gen>.<col>.uris by contentid ts.<sf>](#) 17  
  [<gen>.queries by queryid.<sf>](#) 15  
  [<gen>.queries by queryid all.<sf>](#) 16  
  [<gen>.queryinfo.<sf>](#) 16  
  [<gen>.urls by urlid.<sf>](#) 16  
  [<gen>.urls by urlid all.<sf>](#) 16  
[allfeeduris.<sf>](#) 11  
[cid by cid with counts and query.<sf>](#) 11  
[cid by cid with counts and query mergereduc e.<sf>.<sf>](#) 11  
[global querycnt by query.<sf>](#) 11  
[local querycnt by cid.<sf>](#) 12  
[local querycnt by cid merge.<sf>.<sf>](#) 12  
[local querycnt by query.<sf>](#) 12  
[local querycnt by query reduce.<sf>.<sf>](#) 12  
[local querycnt by url.<sf>](#) 12  
[local querycnt by url merge.<sf>.<sf>](#) 13  
[local querycnt by urlid.<sf>](#) 13  
[local querycnt by urlid reduce.<sf>.<sf>](#) 13  
[semi local querycnt by queryid.<sf>](#) 13  
[semi local querycnt by urlid.<sf>](#) 13  
[semi local querycnt by urlid map.<sf>.<sf>](#) 14

[semi\\_local\\_querycnt\\_pre\\_token.<sf>](#) 14  
[uris\\_by\\_contentid\\_ts.<sf>](#) 14  
[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>](#) 15  
[urls\\_on\\_urlhash\\_with\\_queries.<sf>](#) 15  
[Applicability](#) 7

## C

[Change tracking](#) 40  
[cid\\_by\\_cid\\_with\\_counts\\_and\\_query.<sf>](#) 11  
[example](#) 22  
[cid\\_by\\_cid\\_with\\_counts\\_and\\_query\\_mergereduc.<sf>.<sf>](#) 11  
[example](#) 22  
Common data types and fields ([section 2 8](#), [section 2 8](#))  
[Common file naming conventions](#) 10  
[Common file structures](#) 8

## D

Data types and fields  
[common](#) 8  
[Data types and fields - common](#) 8  
[Database files](#) 19  
[<gen>.sharepoint.rel.<part\\_num>.bin](#) 20  
[<gen>.sharepoint.rel.<part\\_num>.idx](#) 20  
[<gen>.sharepoint.rel.<part\\_num>.idx ofs](#) 20  
Details  
[<col> feeduris.<sf>](#) file 15  
[<col> feeduris\\_expand.<sf>](#) file 15  
[<col> feeduris\\_expand\\_resplit.<sf>](#) file 15  
[<date>.clicks.txt](#) file 10  
[<date>.clicks\\_by\\_urlid\\_and\\_queryid.<sf>](#) file 17  
[<date>.clicks\\_by\\_urlid\\_and\\_queryid\\_sort.<sf>.<sf>](#) file 18  
[<date>.clicks\\_on\\_queryid.0](#) file 18  
[<date>.local\\_querycnt\\_by\\_queryid.<sf>](#) 18  
[<date>.local\\_querycnt\\_by\\_queryid\\_reduce.<sf>.<sf>](#) 18  
[<date>.queries.txt](#) file 10  
[<date>.queries\\_by\\_queryid.<sf>](#) 18  
[<date>.queries\\_by\\_queryid\\_sort.0.<sf>](#) 19  
[<date>.queries\\_on\\_queryid.0](#) 19  
[<date>.urls.txt](#) file 11  
[<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>](#) file 17  
[<gen>.<col>.<host> contentids\\_by\\_contentid\\_new\\_resplit.<sf>.<sf>](#) file 17  
[<gen>.<col>.<host> uris.0](#) file 17  
[<gen>.<col>.unique\\_uris\\_by\\_uri.<sf>](#) file 16  
[<gen>.<col>.uris\\_by\\_contentid.<sf>](#) file 17  
[<gen>.<col>.uris\\_by\\_contentid\\_ts.<sf>](#) file 17  
[<gen>.queries\\_by\\_queryid.<sf>](#) file 15  
[<gen>.queries\\_by\\_queryid\\_all.<sf>](#) file 16  
[<gen>.queryinfo.<sf>](#) file 16

[<gen>.sharepoint.rel.<part\\_num>.bin](#) 20  
[<gen>.sharepoint.rel.<part\\_num>.idx](#) 20  
[<gen>.sharepoint.rel.<part\\_num>.idx ofs](#) 20  
[<gen>.uris\\_by\\_urlid.<sf>](#) file 16  
[<gen>.uris\\_by\\_urlid\\_all.<sf>](#) file 16  
[allfeeduris.<sf>](#) file 11  
[analysis files](#) 11  
[cid\\_by\\_cid\\_with\\_counts\\_and\\_query.<sf>](#) file 11  
[cid\\_by\\_cid\\_with\\_counts\\_and\\_query\\_mergereduc\\_e.<sf>.<sf>](#) file 11  
common data types and fields ([section 2 8](#), [section 2 8](#))  
[common file naming conventions](#) 10  
[common file structures](#) 8  
[database files](#) 19  
[empty files](#) 20  
[global\\_querycnt\\_by\\_query.<sf>](#) file 11  
[local\\_querycnt\\_by\\_cid.<sf>](#) file 12  
[local\\_querycnt\\_by\\_cid\\_merge.<sf>.<sf>](#) file 12  
[local\\_querycnt\\_by\\_query.<sf>](#) file 12  
[local\\_querycnt\\_by\\_query\\_reduce.<sf>.<sf>](#) file 12  
[local\\_querycnt\\_by\\_url.<sf>](#) file 12  
[local\\_querycnt\\_by\\_url\\_merge.<sf>.<sf>](#) file 13  
[local\\_querycnt\\_by\\_urlid.<sf>](#) file 13  
[local\\_querycnt\\_by\\_urlid\\_reduce.<sf>.<sf>](#) file 13  
[search clickthrough files](#) 10  
[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 14  
[semi\\_local\\_querycnt\\_pre\\_token.<sf>](#) file 14  
[uris\\_by\\_contentid\\_ts.<sf>](#) file 14  
[uris\\_by\\_member.<sf>](#) file 14  
[uris\\_by\\_member\\_reduce.<sf>](#) file 14  
[urls\\_by\\_urlhash\\_with\\_queries.<sf>](#) file 14  
[urls\\_by\\_urlhash\\_with\\_queries\\_sort.<sf>](#) file 15  
[urls\\_on\\_urlhash\\_with\\_queries.<sf>](#) file 15

## E

[Empty files](#) 20  
Examples ([section 3 21](#), [section 3 21](#))  
analysis file  
[<col> feeduris.<sf>](#) 28  
[<col> feeduris\\_expand.<sf>](#) 28  
[<col> feeduris\\_expand\\_resplit.<sf>](#) 28  
[<date>.clicks\\_by\\_urlid\\_and\\_queryid.<sf>](#) 32  
[<date>.clicks\\_by\\_urlid\\_and\\_queryid\\_sort.<sf>.<sf>](#) 32  
[<date>.clicks\\_on\\_queryid.0](#) 32  
[<date>.local\\_querycnt\\_by\\_queryid.<sf>](#) 33  
[<date>.local\\_querycnt\\_by\\_queryid\\_reduce.<sf>.<sf>](#) 33  
[<date>.queries\\_by\\_queryid.<sf>](#) 33  
[<date>.queries\\_by\\_queryid\\_sort.0.<sf>](#) 33  
[<date>.queries\\_on\\_queryid.0](#) 34  
[<date>.urls\\_by\\_urlid.<sf>](#) 34  
[<date>.urls\\_by\\_urlid\\_sort.0.<sf>](#) 34  
[<date>.urls\\_on\\_urlid.0](#) 34

<a href="#">&lt;gen&gt;.&lt;col&gt;.&lt;host&gt; contentids by contentid new.&lt;sf&gt;</a>	31	<a href="#">&lt;date&gt;.queries by queryid.&lt;sf&gt;</a>	18
<a href="#">&lt;gen&gt;.&lt;col&gt;.&lt;host&gt; contentids by contentid new resplit.&lt;sf&gt;.&lt;sf&gt;</a>	31	<a href="#">&lt;date&gt;.queries by queryid sort.0.&lt;sf&gt;</a>	19
<a href="#">&lt;gen&gt;.&lt;col&gt;.&lt;host&gt; uris.0</a>	32	<a href="#">&lt;date&gt;.queries on queryid.0</a>	19
<a href="#">&lt;gen&gt;.&lt;col&gt;.unique uris by uri.&lt;sf&gt;</a>	30	<a href="#">&lt;date&gt;.urls.txt</a>	11
<a href="#">&lt;gen&gt;.&lt;col&gt;.uris by contentid.&lt;sf&gt;</a>	31	<a href="#">&lt;date&gt;.urls by urlid.&lt;sf&gt;</a>	19
<a href="#">&lt;gen&gt;.&lt;col&gt;.uris by contentid ts.&lt;sf&gt;</a>	31	<a href="#">&lt;date&gt;.urls by urlid sort.0.&lt;sf&gt;</a>	19
<a href="#">&lt;gen&gt;.queries by queryid.&lt;sf&gt;</a>	28	<a href="#">&lt;date&gt;.urls on urlid.0</a>	19
<a href="#">&lt;gen&gt;.queries by queryid all.&lt;sf&gt;</a>	29	<a href="#">&lt;gen&gt;.&lt;col&gt;.&lt;host&gt; contentids by contentid new.&lt;sf&gt;</a>	17
<a href="#">&lt;gen&gt;.queryinfo.&lt;sf&gt;</a>	29	<a href="#">&lt;gen&gt;.&lt;col&gt;.&lt;host&gt; contentids by contentid new resplit.&lt;sf&gt;.&lt;sf&gt;</a>	17
<a href="#">&lt;gen&gt;.urls by urlid.&lt;sf&gt;</a>	30	<a href="#">&lt;gen&gt;.&lt;col&gt;.&lt;host&gt; uris.0</a>	17
<a href="#">&lt;gen&gt;.urls by urlid all.&lt;sf&gt;</a>	30	<a href="#">&lt;gen&gt;.&lt;col&gt;.unique uris by uri.&lt;sf&gt;</a>	16
<a href="#">allfeeduris.&lt;sf&gt;</a>	22	<a href="#">&lt;gen&gt;.&lt;col&gt;.uris by contentid.&lt;sf&gt;</a>	17
<a href="#">cid by cid with counts and query.&lt;sf&gt;</a>	22	<a href="#">&lt;gen&gt;.&lt;col&gt;.uris by contentid ts.&lt;sf&gt;</a>	17
<a href="#">cid by cid with counts and query mergededu ce.&lt;sf&gt;.&lt;sf&gt;</a>	22	<a href="#">&lt;gen&gt;.queries by queryid.&lt;sf&gt;</a>	15
<a href="#">global querycnt by query.&lt;sf&gt;</a>	22	<a href="#">&lt;gen&gt;.queries by queryid all.&lt;sf&gt;</a>	16
<a href="#">local querycnt by cid.&lt;sf&gt;</a>	23	<a href="#">&lt;gen&gt;.queryinfo.&lt;sf&gt;</a>	16
<a href="#">local querycnt by cid merge.&lt;sf&gt;.&lt;sf&gt;</a>	23	<a href="#">&lt;gen&gt;.sharepoint.rel.&lt;part num&gt;.bin</a>	20
<a href="#">local querycnt by query.&lt;sf&gt;</a>	23	<a href="#">&lt;gen&gt;.sharepoint.rel.&lt;part num&gt;.idx</a>	20
<a href="#">local querycnt by query reduce.&lt;sf&gt;.&lt;sf&gt;</a>	23	<a href="#">&lt;gen&gt;.sharepoint.rel.&lt;part num&gt;.idx ofs</a>	20
<a href="#">local querycnt by url.&lt;sf&gt;</a>	24	<a href="#">&lt;gen&gt;.urls by urlid.&lt;sf&gt;</a>	16
<a href="#">local querycnt by url merge.&lt;sf&gt;.&lt;sf&gt;</a>	24	<a href="#">&lt;gen&gt;.urls by urlid all.&lt;sf&gt;</a>	16
<a href="#">local querycnt by urlid.&lt;sf&gt;</a>	24	<a href="#">allfeeduris.&lt;sf&gt;</a>	11
<a href="#">local querycnt by urlid reduce.&lt;sf&gt;.&lt;sf&gt;</a>	24	<a href="#">analysis</a>	11
<a href="#">semi local querycnt by queryid.&lt;sf&gt;</a>	24	<a href="#">cid by cid with counts and query.&lt;sf&gt;</a>	11
<a href="#">semi local querycnt by urlid.&lt;sf&gt;</a>	25	<a href="#">cid by cid with counts and query mergededuc e.&lt;sf&gt;.&lt;sf&gt;</a>	11
<a href="#">semi local querycnt by urlid map.&lt;sf&gt;.&lt;sf&gt;</a>	25	<a href="#">common naming conventions</a>	10
<a href="#">semi local querycnt pre token.&lt;sf&gt;</a>	25	<a href="#">database</a>	19
<a href="#">uris by contentid ts.&lt;sf&gt;</a>	26	<a href="#">empty</a>	20
<a href="#">uris by member.&lt;sf&gt;</a>	26	<a href="#">global querycnt by query.&lt;sf&gt;</a>	11
<a href="#">uris by member reduce.&lt;sf&gt;</a>	26	<a href="#">local querycnt by cid.&lt;sf&gt;</a>	12
<a href="#">urls by urlhash with queries.&lt;sf&gt;</a>	26	<a href="#">local querycnt by cid merge.&lt;sf&gt;.&lt;sf&gt;</a>	12
<a href="#">urls by urlhash with queries sort.&lt;sf&gt;</a>	27	<a href="#">local querycnt by query.&lt;sf&gt;</a>	12
<a href="#">urls on urlhash with queries.&lt;sf&gt;</a>	27	<a href="#">local querycnt by query reduce.&lt;sf&gt;.&lt;sf&gt;</a>	12
Database file		<a href="#">local querycnt by url.&lt;sf&gt;</a>	12
<a href="#">&lt;gen&gt;.sharepoint.rel.&lt;part num&gt;.bin</a>	35	<a href="#">local querycnt by url merge.&lt;sf&gt;.&lt;sf&gt;</a>	13
<a href="#">&lt;gen&gt;.sharepoint.rel.&lt;part num&gt;.idx</a>	36	<a href="#">local querycnt by urlid.&lt;sf&gt;</a>	13
<a href="#">&lt;gen&gt;.sharepoint.rel.&lt;part num&gt;.idx ofs</a>	37	<a href="#">local querycnt by urlid reduce.&lt;sf&gt;.&lt;sf&gt;</a>	13
search clickthrough file		<a href="#">search clickthrough</a>	10
<a href="#">&lt;date&gt;.clicks.txt</a>	21	<a href="#">semi local querycnt by queryid.&lt;sf&gt;</a>	13
<a href="#">&lt;date&gt;.queries.txt</a>	21	<a href="#">semi local querycnt by urlid.&lt;sf&gt;</a>	13
<a href="#">&lt;date&gt;.urls.txt</a>	21	<a href="#">semi local querycnt by urlid map.&lt;sf&gt;.&lt;sf&gt;</a>	14
<b>F</b>		<a href="#">semi local querycnt pre token.&lt;sf&gt;</a>	14
<a href="#">Fields - vendor-extensible</a>	7	<a href="#">uris by contentid ts.&lt;sf&gt;</a>	14
Files		<a href="#">uris by member.&lt;sf&gt;</a>	14
<a href="#">&lt;col&gt; feeduris.&lt;sf&gt;</a>	15	<a href="#">uris by member reduce.&lt;sf&gt;</a>	14
<a href="#">&lt;col&gt; feeduris expand.&lt;sf&gt;</a>	15	<a href="#">uris by urlhash with queries.&lt;sf&gt;</a>	14
<a href="#">&lt;col&gt; feeduris expand resplit.&lt;sf&gt;</a>	15	<a href="#">uris by urlhash with queries sort.&lt;sf&gt;</a>	15
<a href="#">&lt;date&gt;.clicks.txt</a>	10	<a href="#">urls on urlhash with queries.&lt;sf&gt;</a>	15
<a href="#">&lt;date&gt;.clicks by urlid and queryid.&lt;sf&gt;</a>	17	<b>G</b>	
<a href="#">&lt;date&gt;.clicks by urlid and queryid sort.&lt;sf&gt;.&lt;sf&gt;</a>	18	<a href="#">global querycnt by query.&lt;sf&gt;</a>	11
<a href="#">&lt;date&gt;.clicks on queryid.0</a>	18	<a href="#">example</a>	22
<a href="#">&lt;date&gt;.local querycnt by queryid.&lt;sf&gt;</a>	18	<a href="#">Glossary</a>	6
<a href="#">&lt;date&gt;.local querycnt by queryid reduce.&lt;sf&gt;.&lt;sf&gt;</a>	18	<b>I</b>	
<a href="#">&lt;date&gt;.queries.txt</a>	10	<a href="#">Implementer - security considerations</a>	38

[Informative references](#) 7  
[Introduction](#) 6

## L

[local\\_querycnt\\_by\\_cid.<sf>](#) 12  
  [example](#) 23  
[local\\_querycnt\\_by\\_cid\\_merge.<sf>.<sf>](#) ([section 2.4.6](#) 12, [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>](#) 13  
  [example](#) 24  
[local\\_querycnt\\_by\\_urlid.<sf>](#) 13  
  [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  
  [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](#) 11  
[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>](#) 14  
  [example](#) 25  
[semi\\_local\\_querycnt\\_pre\\_token.<sf>](#) 14  
  [example](#) 25  
Structures  
  [common file](#) 8  
  overview ([section 2](#) 8, [section 2](#) 8)

## T

[Tracking changes](#) 40

## U

[uris\\_by\\_contentid\\_ts.<sf>](#) 14  
  [example](#) 26  
[uris\\_by\\_member.<sf>](#) 14  
  [example](#) 26  
[uris\\_by\\_member\\_reduce.<sf>](#) 14  
  [example](#) 26  
[uris\\_by\\_urlhash\\_with\\_queries.<sf>](#) 14  
  [example](#) 26  
[uris\\_by\\_urlhash\\_with\\_queries\\_sort.<sf>](#) 15  
  [example](#) 27  
[uris\\_on\\_urlhash\\_with\\_queries.<sf>](#) 15  
  [example](#) 27

## V

[Vendor-extensible fields](#) 7  
[Versioning](#) 7