

Network Working Group	M. Hausenblas
Internet-Draft	DERI, NUI Galway
Updates: 4180 (if approved)	E. Wilde
Intended status: Standards Track	UC Berkeley
Expires: October 28, 2011	April 26, 2011

URI Fragment Identifiers for the text/csv Media Type
draft-hausenblas-csv-fragment-00

Abstract

This memo defines URI fragment identifiers for text/csv MIME entities. These fragment identifiers make it possible to refer to parts of a text/csv MIME entity, identified by cell, row, column, or slice.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on October 28, 2011.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- *1. [Introduction](#)
- *1.1. [What is text/csv?](#)
- *1.2. [Why text/csv Fragment Identifiers?](#)

- *1.2.1. [Motivation](#)
- *1.2.2. [Use Cases](#)
- *1.3. [Incremental Deployment](#)
- *1.4. [Notation Used in this Memo](#)
- *2. [Fragment Identification Methods](#)
 - *2.1. [Header](#)
 - *2.2. [Row-based selection](#)
 - *2.3. [Column-based selection](#)
 - *2.4. [Slice-based selection](#)
- *3. [Fragment Identification Syntax](#)
- *4. [Fragment Identifier Processing](#)
 - *4.1. [Syntax Errors in Fragment Identifiers](#)
- *5. [IANA Considerations](#)
- *6. [Security Considerations](#)
- *7. [Change Log](#)
- *8. [References](#)
 - *8.1. [Normative References](#)
 - *8.2. [Non-Normative References](#)
- *Appendix A. [Acknowledgements](#)
- *[Authors' Addresses](#)

1. Introduction

This memo updates the text/csv media type defined in RFC 4180 [\[RFC4180\]](#) by defining URI fragment identifiers for text/csv MIME entities. This section gives an introduction to the general concepts of text/csv MIME entities and URI fragment identifiers, and discusses the need for fragment identifiers for text/csv and deployment issues. [Section 2](#) discusses the principles and methods on which this memo is based. [Section 3](#) defines the syntax, and [Section 4](#) discusses processing of text/csv fragment identifiers.

1.1. What is text/csv?

Internet Media Types (often referred to as "MIME types") as defined in RFC 2045 [[RFC2045](#)] and RFC 2046 [[RFC2046](#)] are used to identify different types and sub-types of media. The text/csv media type is defined in RFC 4180 [[RFC4180](#)], using US-ASCII [[ASCII](#)] as the default character encoding (other character encodings can be used as well).

1.2. Why text/csv Fragment Identifiers?

URIs are the identification mechanism for resources on the Web. The URI syntax specified in RFC 3986 [[RFC3986](#)] optionally includes a so-called "fragment identifier", separated by a number sign ('#'). The fragment identifier consists of additional reference information to be interpreted by the user agent after the retrieval action has been successfully completed. The semantics of a fragment identifier is a property of the data resulting from a retrieval action, regardless of the type of URI used in the reference. Therefore, the format and interpretation of fragment identifiers is dependent on the media type of the retrieval result.

1.2.1. Motivation

Similar to the motivation in RFC 5147 [[RFC5147](#)], referring to specific parts of a resource can be very useful, because it enables users and applications to create more specific references. Users can create references to the part they really are interested in or want to talk about, rather than always pointing to a complete resource. Even though it is suggested that fragment identification methods are specified in a media type's MIME registration (see [[TypeReg](#)]), many media types do not have fragment identification methods associated with them.

Fragment identifiers are only useful if supported by the client, because they are only interpreted by the client. Therefore, a new fragment identification method will require some time to be adopted by clients, and older clients will not support it. However, because the URI still works even if the fragment identifier is not supported (the resource is retrieved, but the fragment identifier is not interpreted), rapid adoption is not highly critical to ensure the success of a new fragment identification method.

1.2.2. Use Cases

Fragment identifiers for text/csv as defined in this memo make it possible to refer to specific parts of a text/csv MIME entity. Use cases include, but are not limited to, discovery (what column headings or how many rows are available), selecting a part for visual rendering, stream processing, making assertions about a certain value (provenance, confidence, etc.), or data integration.

1.3. Incremental Deployment

As long as text/csv fragment identifiers are not supported universally, it is important to consider the implications of incremental deployment. Clients (for example, Web browsers) not supporting the text/csv fragment identifier described in this memo will work with URI references to text/csv MIME entities, but they will fail to understand the identification of the sub-resource specified by the fragment identifier, and thus will behave as if the complete resource was referenced. This is a reasonable fallback behavior, and in general users should take into account the possibility that a program interpreting a given URI will fail to interpret the fragment identifier part. Since fragment identifier evaluation is local to the client (and happens after retrieving the MIME entity), there is no reliable way for a server to determine whether a requesting client is using a URI containing a fragment identifier.

1.4. Notation Used in this Memo

The capitalized key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [[RFC2119](#)].

2. Fragment Identification Methods

This memo specifies fragment identification using following methods: header, row, column and slice. As of RFC 4180 [[RFC4180](#)] the header line is optional and hence the application of the method is dependent on the actual format of the text/csv MIME entity.

Throughout the sections below the following table in CSV is used:

```
date,temperature,place
2011-01-01,1,Galway
2011-01-02,-1,Galway
2011-01-03,0,Galway
2011-01-01,6,Berkeley
2011-01-02,8,Berkeley
2011-01-03,5,Berkeley
```

2.1. Header

For discovery purposes, the "head" scheme is used, returning the first row. If the "header" parameter per RFC 4180 [[RFC4180](#)] is available and its value is "present" the client can reliably determine that it is a header.

`http://example.com/data.csv#head`

Applied to the reference table, the above CSV fragment would select the header row, yielding:

```
date,temperature,place
```

2.2. Row-based selection

To select a specific record, the "row" scheme followed by a single number is used (the first record has the index 0). If the fragment is given in the form row:*, then no record is selected but the overall number of records is returned.

```
http://example.com/data.csv#row:2
```

The above CSV fragment yields: while the following computes the number of records (which equals 6, in the reference table)

```
2011-01-03,0,Galway
```

The following computes the number of records (which equals 6, in the reference table):

```
http://example.com/data.csv#row:*
```

2.3. Column-based selection

To select values from a certain column, the "col" scheme, followed either by a single number or the value of a header field is used.

```
http://example.com/data.csv#col:temperature
```

The above CSV fragment addresses a column by name, yielding:

```
1, -1,0,6,8,5
```

A column can also be addressed by position as shown in the next example:

```
http://example.com/data.csv#col:2
```

The above CSV fragment selects the third column:

```
Galway,Galway,Galway,Berkeley,Berkeley,Berkeley
```

[2.4. Slice-based selection](#)

To select a part of table, called a slice in the following, the "where" scheme is used. The allowed values are a comma-separated list of header fields with corresponding field values in the table.

```
http://example.com/data.csv#where:date=2011-01-01
```

The above CSV fragment selects a slice, yielding another CSV table as follows:

```
temperature,place
1,Galway
6,Berkeley
```

[3. Fragment Identification Syntax](#)

The syntax for the text/csv fragment identifiers is as follows. The following syntax definition uses ABNF as defined in RFC 4234 [\[RFC4234\]](#), including the rules DIGIT and HEXDIG. The mime-charset rule is defined in RFC 2978 [\[RFC2978\]](#).

NOTE: In the descriptions that follow, specified text values MUST be used exactly as given, using exactly the indicated lower-case letters. In this respect, the ABNF usage differs from [\[RFC4234\]](#).

```
csv-fragment = headersel / wheresel / colsel / rowsel
headersel = "head"
rowsel = "row:" rowspec
colsel = "col:" colspec
wheresel = "where:" kvpairs
kvpairs = 1*( col "=" val 0*1(",") )
col = 1*TEXTDATA
val = 1*TEXTDATA
colspec = column
rowspec = "*" / rownum
column = 1*TEXTDATA / 1*DIGIT
rownum = 1*DIGIT
TEXTDATA = %x23-2B / %x2D-3C / %x3E-7E
DIGIT = %x30-39
```

[4. Fragment Identifier Processing](#)

Applications implementing support for the mechanism described in this memo MUST behave as described in the following sections.

4.1. Syntax Errors in Fragment Identifiers

If a fragment identifier contains a syntax error (i.e., does not conform to the syntax specified in [Section 3](#)), then it MUST be ignored by clients. Clients MUST NOT make any attempt to correct or guess fragment identifiers. Syntax errors MAY be reported by clients.

5. IANA Considerations

Note to RFC Editor: Please change this section to read as follows after the IANA action has been completed: "IANA has added a reference to this specification in the Text/Plain Media Type registration."

IANA is requested to update the registration of the MIME Media type text/csv at <http://www.iana.org/assignments/media-types/text/> with the fragment identifier defined in this memo by adding a reference to this memo (with the appropriate RFC number once it is known).

6. Security Considerations

The fact that software implementing fragment identifiers for plain text and software not implementing them differs in behavior, and the fact that different software may show documents or fragments to users in different ways, can lead to misunderstandings on the part of users. Such misunderstandings might be exploited in a way similar to spoofing or phishing.

In particular, care has to be taken if fragment identifiers are used together with a mechanism that allows to show only the part of a document identified by a fragment. One scenario may be the use of a fragment identifier to hide small-print legal text. Another scenario may be the inclusion of site-key-like material, which may give the user the impression of using the real site rather than a fake site. Other scenarios may also be possible. Possible countermeasures may include but are not limited to displaying the included content within clearly visible boundaries and limiting inclusion to material from the same security realm or from realms that give explicit permission to be included in another realm.

Please note that the above issues all apply to the client side; fragment identifiers are not used when resolving an URI to retrieve the representation of a resource, but are only applied on the client side. Implementers and users of fragment identifiers for CSV text should also be aware of the security considerations in RFC 3986 [\[RFC3986\]](#) and RFC 3987 [\[RFC3987\]](#).

7. Change Log

Note to RFC Editor: Please remove this section before publication.

8. References

8.1. Normative References

- [1] [Freed, N. and N. S. Borenstein, "Multipurpose Internet Mail Extensions \(MIME\) Part One: Format of Internet Message Bodies"](#), RFC 2045, November 1996.
- [2] [Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions \(MIME\) Part Two: Media Types"](#), RFC 2046, November 1996.
- [3] [Shafranovich, Y., "Common Format and MIME Type for Comma-Separated Values \(CSV\) Files"](#), RFC 4180, October 2005.
- [4] Bradner, S., ["Key words for use in RFCs to Indicate Requirement Levels"](#), RFC 2119, March 1997.
- [5] Freed, N. and J. Postel, "IANA Charset Registration Procedures", BCP 19, October 2000.
- [6] [Berners-Lee, T., Fielding, R. T. and L. Masinter, "Uniform Resource Identifier \(URI\): Generic Syntax"](#), RFC 3986, January 2005.
- [7] Duerst, M. and M. Suignard, ["Internationalized Resource Identifiers \(IRI\)"](#), RFC 3987, January 2005.
- [8] Crocker, D. H. and P. Overell, ["Augmented BNF for Syntax Specifications: ABNF"](#), RFC 4234, October 2005.

8.2. Non-Normative References

- [1] ANSI X3.4-1986, ["Coded Character Set - 7-Bit American National Standard Code for Information Interchange"](#), STD 63, RFC 3629, 1992.
- [2] [Wilde, E. and M. J. Duerst, "URI Fragment Identifiers for the text/plain Media Type"](#), RFC 5147, April 2008.
- [3] Freed, N. and J. Klensin, ["Media Type Specifications and Registration Procedures"](#), RFC 4288, December 2005.

Appendix A. Acknowledgements

Thanks for comments and suggestions provided by ...

Authors' Addresses

Michael Hausenblas Hausenblas DERI, NUI Galway IDA Business Park
Galway, Ireland Phone: +353-91-495730 EMail:
michael.hausenblas@deri.org URI: <http://sw-app.org/about.html>

Erik Wilde Wilde UC Berkeley School of Information, 311 South Hall
Berkeley, CA 94720-4600, U.S.A. Phone: +1-510-6432253 EMail:
dret@berkeley.edu URI: <http://dret.net/netdret/>