

## **Sieve -- Regular Expression Extension**

### Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#). Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>

To view the list Internet-Draft Shadow Directories, see <http://www.ietf.org/shadow.html>.

Distribution of this memo is unlimited.

### Copyright Notice

Copyright (C) The Internet Society 2001. All Rights Reserved.

### Abstract

In some cases, it is desirable to have a string matching mechanism which is more powerful than a simple exact match, a substring match or a glob-style wildcard match. The regular expression matching mechanism defined in this draft should allow users to isolate just about any string or address in a message header or envelope.

## Table of Contents

Status of this Memo . . . . .	<a href="#">1</a>
Copyright Notice . . . . .	<a href="#">1</a>
Abstract . . . . .	<a href="#">1</a>
<a href="#">0.</a> Meta-information on this draft . . . . .	<a href="#">3</a>
<a href="#">0.1.</a> Discussion . . . . .	<a href="#">3</a>
<a href="#">0.2.</a> Noted Changes . . . . .	<a href="#">3</a>
<a href="#">0.2.1</a> since -03 . . . . .	<a href="#">3</a>
<a href="#">0.2.2</a> since -02 . . . . .	<a href="#">3</a>
<a href="#">0.2.3</a> since -01 . . . . .	<a href="#">3</a>
<a href="#">0.2.4</a> since -00 . . . . .	<a href="#">3</a>
<a href="#">1.</a> Introduction . . . . .	<a href="#">4</a>
<a href="#">2.</a> Capability Identifier . . . . .	<a href="#">4</a>
<a href="#">3.</a> Regex Match Type . . . . .	<a href="#">4</a>
<a href="#">4.</a> Security Considerations . . . . .	<a href="#">6</a>
<a href="#">5.</a> Acknowledgments . . . . .	<a href="#">6</a>
<a href="#">6.</a> Author's Address . . . . .	<a href="#">6</a>
<a href="#">Appendix A.</a> References . . . . .	<a href="#">7</a>
<a href="#">Appendix B.</a> Full Copyright Statement . . . . .	<a href="#">7</a>

Expires March 4, 2002

Murchison

[Page 2]

## **0. Meta-information on this draft**

This information is intended to facilitate discussion. It will be removed when this document leaves the Internet-Draft stage.

### **0.1. Discussion**

This draft is intended to be an extension to the Sieve mail filtering language, available from the Internet-Drafts repository as [<ftp://ftp.ietf.org/internet-drafts/draft-showalter-sieve-12.txt>](ftp://ftp.ietf.org/internet-drafts/draft-showalter-sieve-12.txt) (where 12 is the version number, which is actually currently 12).

This draft and the Sieve language itself are being discussed on the MTA Filters mailing list at [<ietf-mta-filters@imc.org>](mailto:ietf-mta-filters@imc.org). Subscription requests can be sent to [<ietf-mta-filters-request@imc.org>](mailto:ietf-mta-filters-request@imc.org) (send an email message with the word "subscribe" in the body). More information on the mailing list along with a WWW archive of back messages is available at [<http://www.imc.org/ietf-mta-filters/>](http://www.imc.org/ietf-mta-filters/).

### **0.2. Noted Changes**

#### **0.2.1 since -03**

Added security concerns with poor implementations.

#### **0.2.2 since -02**

Editorial changes (re-submission).

#### **0.2.3 since -01**

Added augmented MATCH-TYPE syntax element.

Editorial changes.

#### **0.2.4 since -00**

Added POSIX.2 ERE summary.

Added examples and editorial changes.

Expires March 4, 2002

Murchison

[Page 3]

## **1. Introduction**

This is an extension to the Sieve language defined by [\[SIEVE\]](#) for comparing strings to regular expressions.

Conventions for notations are as in [\[SIEVE\]](#) [section 1.1](#), including use of [\[KEYWORDS\]](#).

## **2. Capability Identifier**

The capability string associated with the extension defined in this document is "regex".

## **3. Regex Match Type**

Commands that support matching may take the optional tagged argument ":regex" to specify that a regular expression match should be performed. The ":regex" match type is subject to the same rules and restrictions as the standard match types defined in [\[SIEVE\]](#). For convenience, the "MATCH-TYPE" syntax element defined in [\[SIEVE\]](#) is augmented here as follows:

```
MATCH-TYPE  =/  ":regex"
```

Example:

```
require "regex";

# Try to catch unsolicited email.
if anyof (
  # if a message is not to me (with optional +detail),
  not address :regex ["to", "cc", "bcc"]
    "me(\\+.*)?@company.com",

  # or the subject contains 2 or more dollar signs,
  header :regex "subject" "\\$\\$+",

  # or the subject is all uppercase (no lowercase)
  header :regex :comparator "i;octet" "subject"
    "^[^:lower:]*$" ) {

  discard;      # junk it
}
```

The ":regex" match type is compatible with both the "i;octet" and "i;ascii-casemap" comparators and may be used with them.

Expires March 4, 2002

Murchison

[Page 4]

Implementations MUST support extended regular expressions (EREs) as defined by [POSIX.2]. Any regular expression not defined by [POSIX.2], including [POSIX.2] basic regular expressions, word boundaries and backreferences are not supported by this extension.

The following table provides a brief summary of the regular expressions that MUST be supported. This table is presented here only as a guideline. [POSIX.2] should be used as the definitive reference.

Expression	Pattern
Items to match a single character	
.	Match any single character except newline.
[ ]	Bracket expression. Match any one of the enclosed characters. A hyphen (-) indicates a range of consecutive characters.
[^ ]	Negated bracket expression. Match any one character NOT in the enclosed list. A hyphen (-) indicates a range of consecutive characters.
\\	Escape the following special character (match the literal character). Undefined for other characters. NOTE: Unlike [POSIX.2], a double-backslash is required as per section 2.4.2 of [SIEVE].
Items to be used within a bracket expression (localization)	
[ : : ]	Character class (alnum, alpha, blank, cntrl, digit, graph, lower, print, punct, space, upper, xdigit).
[ = = ]	Character equivalents.
[ . . ]	Collating sequence.
Quantifiers - Items to count the preceding regular expression	
?	Match zero or one instances.
*	Match zero or more instances.
+	Match one or more instances.
{n,m}	Match any number of instances between n and m (inclusive). {n} matches exactly n instances. {n,} matches n or more instances.



Expires March 4, 2002

Murchison

[Page 5]

<u>Expression</u>	<u>Pattern</u>
Anchoring - Items to match positions	
^	Match the beginning of the line or string.
\$	Match the end of the line or string.
Other constructs	
	Alternation. Match either of the separated regular expressions.
( )	Group the enclosed regular expression(s).

#### 4. Security Considerations

Security considerations are discussed in [SIEVE]. It is believed that this extension doesn't introduce any additional security concerns.

However, a poor implementation COULD introduce security problems ranging from degradation of performance to denial of service. If an implementation uses a third-party regular expression library, that library should be checked for potentially problematic regular expressions, such as "(.\*)\*".

## 5. Acknowledgments

Thanks to Tim Showalter, Alexey Melnikov, Tony Hansen and Phil Pennock for their help with this document.

## 6. Author's Address

Kenneth Murchison  
Oceana Matrix Ltd.  
21 Princeton Place  
Orchard Park, NY 14127

Phone: (716) 662-8973  
EMail: ken@oceana.com

Expires March 4, 2002

Murchison

[Page 6]

## **Appendix A. References**

[KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", Harvard University, [RFC 2119](#), March, 1997.

[SIEVE] Showalter, T., "Sieve: A Mail Filtering Language", Mirapoint, Inc., [RFC 3028](#), January 2001.

[POSIX.2], "Portable Operating System Interface (POSIX). Part 2, Shell and utilities", National Institute of Standards and Technology (U.S.).

## **Appendix B. Full Copyright Statement**

Copyright (C) The Internet Society 2001. All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Expires March 4, 2002

Murchison

[Page 7]