

Internet Draft
Document: [draft-ietf-imapext-regex-00.txt](#)
Expires: September 2000

R. Gellens
QUALCOMM
March 2000

IMAP Regular Expressions SEARCH 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/lid-abstracts.txt>>

The list of Internet-Draft Shadow Directories can be accessed at
<<http://www.ietf.org/shadow.html>>.

A version of this draft document is intended for submission to the RFC editor as a Proposed Standard for the Internet Community. Discussion and suggestions for improvement are requested.

Copyright Notice

Copyright (C) The Internet Society 2000. All Rights Reserved.

Table of Contents

1.	Abstract	2
2.	Conventions Used in this Document	2
3.	Comments	2
4.	Open Issues	2
5.	REGEX Modifier to SEARCH (and UID SEARCH) Criteria	3
6.	Formal Syntax Changes	3
7.	Regular Expression Details	3
8.	Examples	4
9.	References	4
10.	Security Considerations	5
11.	Acknowledgments	5
12.	Author's Address	5
13.	Full Copyright Statement	5

[1.](#) Abstract

This memo describes a regular-expression search facility for the [IMAP] protocol.

A server advertises support for this facility by the capability name REGEX.

[2.](#) Conventions Used in this Document

The 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](#) [[KEYWORDS](#)].

[3.](#) Comments

Public comments can be sent to the IETF IMAP Extensions mailing list, <ietf-impext@imc.org>. To subscribe, send a message to <ietf-impext-request@imc.org> with the word SUBSCRIBE as the body. Private comments should be sent to the author.

[4.](#) Open Issues

- Should the regular expression syntax be specified in a simple

form, as is done here, or should it be a reference to a published, more complex form, such as POSIX 1003.2?

- Should ^ and \$ be used for token, line, or a message part (such as a header field)?

Gellens

Expires September 2000

[Page 2]Internet

- Should REGEX come after the search key?
- The current Formal Syntax allows for silliness such as SEARCH REGEXP REGEXP CC "gork"

5. REGEX Modifier to SEARCH (and UID SEARCH) Criteria

This extension adds an additional optional REGEXP modifier to all SEARCH criteria which take strings. When this modifier is present, the search criteria string is interpreted as a regular expression, as described in [section 7](#).

If the string supplied with the search criteria does not contain a valid regular expression, the server MUST return a BAD response.

If the search criteria does not take a string, the server MUST return a BAD response.

6. Formal Syntax Changes

This section described the changes to the Formal Syntax of the IMAP protocol, using [\[ABNF\]](#):

```
search-key =/ REGEX SP search-key
            ; modifies existing IMAP search-key so
            ; that string values in search-key are treated
            ; as regular expressions for pattern matching
```

7. Regular Expression Details

The regular expression syntax described in this section is a subset of that used in many applications and systems. It is however very simple and does not include the logical operators AND and OR.

Searches using regular expressions are always substring matches except when the regular expression contains the characters '^' or

'\$'.

Character	Function
<any except those listed in this table>	Matches itself
.	Matches any character
a*	Matches zero or more 'a's
a+	Matches one or more 'a's
[ab]	Matches 'a' or 'b'
[a-c]	Matches 'a', 'b' or 'c'
[^ab]	Matches any character except 'a' or 'b'

Gellens

Expires September 2000

[Page 3]Internet

^	Matches beginning of a token
\$	Matches end of a token
\	Next character matches itself

Examples

String	Matches	Doesn't Match
hello	xhelloy	heello
h.llo	hello	helio
h.*o	hello	helloa
h[a-f]llo	hello	hgllo
^he.*	hello	ehello
.*lo\$	hello	helloo
hel+o	hello	helo

8. Examples

This example finds messages that have "MAKE***MONEY***FAST" in the subject, but not "MAKE!**MONEY**!FAST":

```
C: Z SEARCH REGEX SUBJECT "MAKE[ _\-\]*+MONEY[ _\-\]*+FAST"
S: * SEARCH 2 22 98 2048
S: Z OK SEARCH Completed
```

This example uses an invalid regular expression:

```
C: Y SEARCH REGEX TO ".["
S: Y BAD Invalid regular expression syntax
```

This example uses REGEX with a search criteria that does not take a string:

```
C: X SEARCH REGEX LARGER 10000
S: X BAD Cannot use regular expressions with the LARGER criteria
```

This example searches for messages without the \ANSWERED flag where the envelope from matches the regular expression "mump.*wump" and the text matches the substring "a[0-9]*" (note that the text criteria is not a regular expression):

```
C: W SEARCH UNANSWERED REGEX FROM "mump.*wump" TEXT "a[0-9]*"
S: * SEARCH 3
S: W OK SEARCH Completed
```

Gellens

Expires September 2000

[Page 4]Internet

9. References

[ABNF] Crocker, Overell, "Augmented BNF for Syntax Specifications: ABNF", [RFC 2234](#), Internet Mail Consortium, Demon Internet Ltd., November 1997. <<ftp://ftp.isi.edu/in-notes/rfc2234.txt>>

[IMAP4] Crispin, "Internet Message Access Protocol - Version 4rev1", [RFC 2060](#), University of Washington, December 1996.

[KEYWORDS] Bradner, "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), Harvard University, March 1997. <<ftp://ftp.isi.edu/in-notes/rfc2119.txt>>

10. Security Considerations

This extension does not alter the security semantics of IMAP.

11. Acknowledgments

The table in [section 7](#) is based on the one in [RFC 1835](#).

12. Author's Address

Randall Gellens
QUALCOMM Incorporated
5775 Morehouse Drive

+1 858 651 5115
randy@qualcomm.com

San Diego, CA 92121-2779
U.S.A.

13. Full Copyright Statement

Copyright (C) The Internet Society 2000. 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.

Gellens

Expires September 2000

[Page 5]Internet

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.

