Network Working Group M. Smith, Editor
Request for Comments: DRAFT Netscape Communications Corp.
Obsoletes: RFC 2254 T. Howes
Expires: 9 February 2003 Loudcloud, Inc.
9 August 2002

LDAP: String Representation of Search Filters
<draft-ietf-ldapbis-filter-03.txt>

1. Status of this Memo

Discussion of this document should take place on the LDAP (v3) Revision (ldapbis) Working Group mailing list <ietf-ldapbis@openldap.org>.

2. Abstract

LDAP search filters are transmitted in the LDAP protocol using a binary representation that is appropriate for use on the network. This document defines a human-readable string representation of LDAP search filters that is appropriate for use in LDAP URLs and in other applications.

3. Table of Contents

4. Introduction

The Lightweight Directory Access Protocol (LDAP) [Protocol] defines a network representation of a search filter transmitted to an LDAP server. Some applications may find it useful to have a common way of representing these search filters in a human-readable form; LDAP URLs are an example of one such application. This document defines a human-readable string format for representing the full range of possible LDAP version 3 search filters, including extended match filters.

This document is an integral part of the LDAP Technical Specification [Roadmap].

This document replaces RFC 2254. Changes to RFC 2254 are summarized in Appendix A.

5. LDAP Search Filter Definition

An LDAPv3 search filter is defined in Section 4.5.1 of [Protocol] as follows:

```
     Filter ::= CHOICE {
             and                [0] SET SIZE (1..MAX) OF Filter,
             or                 [1] SET SIZE (1..MAX) OF Filter,
             not                [2] Filter,
             equalityMatch      [3] AttributeValueAssertion,
```

```
                substrings         [4] SubstringFilter,
                greaterOrEqual     [5] AttributeValueAssertion,
                lessOrEqual        [6] AttributeValueAssertion,
                present            [7] AttributeDescription,
                approxMatch        [8] AttributeValueAssertion,
                extensibleMatch    [9] MatchingRuleAssertion }

        SubstringFilter ::= SEQUENCE {
                type    AttributeDescription,
                -- at least one must be present,
                -- initial and final can occur at most once
                substrings    SEQUENCE OF CHOICE {
                        initial        [0] AssertionValue,
                        any            [1] AssertionValue,
                        final          [2] AssertionValue } }

        AttributeValueAssertion ::= SEQUENCE {
                attributeDesc   AttributeDescription,
                assertionValue  AssertionValue }

        MatchingRuleAssertion ::= SEQUENCE {
                matchingRule    [1] MatchingRuleId OPTIONAL,
                type            [2] AttributeDescription OPTIONAL,
                matchValue      [3] AssertionValue,
                dnAttributes    [4] BOOLEAN DEFAULT FALSE }

        AttributeDescription ::= LDAPString

        AttributeValue ::= OCTET STRING

        MatchingRuleId ::= LDAPString

        AssertionValue ::= OCTET STRING

        LDAPString ::= OCTET STRING
```

where the LDAPString above is limited to the UTF-8 encoding of the ISO 10646 character set [RFC2279]. The AttributeDescription is a string representation of the attribute description and is defined in [Protocol]. The AttributeValue and AssertionValue OCTET STRING have the form defined in [Syntaxes]. The Filter is encoded for transmission over a network using the Basic Encoding Rules defined in [ASN.1], with simplifications described in [Protocol].

6. String Search Filter Definition

The string representation of an LDAP search filter is defined by the following grammar, following the ABNF notation defined in [RFC2234].

The filter format uses a prefix notation.

```
      filter         = "(" filtercomp ")"
      filtercomp     = and / or / not / item
      and            = "&" filterlist
      or             = "|" filterlist
      not            = "!" filter
      filterlist     = 1*filter
      item           = simple / present / substring / extensible
      simple         = attr filtertype assertionvalue
      filtertype     = equal / approx / greater / less
      equal          = "="
      approx         = "~="
      greater        = ">="
      less           = "<="
      extensible     = attr [":dn"] [":" matchingrule] ":=" assertionvalue
                       / [":dn"] ":" matchingrule ":=" assertionvalue
                       / ":=" assertionvalue
      present        = attr "=*"
      substring      = attr "=" [initial] any [final]
      initial        = assertionvalue
      any            = "*" *(assertionvalue "*")
      final          = assertionvalue
      attr           = AttributeDescription
                         ; The <AttributeDescription> rule is defined in
                         ; Section 4.1.4 of [Protocol].
      matchingrule   = oid
                         ; The <oid> rule is defined in Section 2.1
                         ; of [Syntaxes] and is used to encode a
                         ; matching rule OBJECT IDENTIFIER.
      assertionvalue = valueencoding
                         ; The <valueencoding> rule is used to encode an
                         ; <AssertionValue> from Section 4.1.6 of [Protocol].
      valueencoding  = 0*(normal / escaped)
      normal         = %x01-27 / %x2b-5b / %x5d-7f
      escaped        = "\" hex hex
      hex            = %x30-39 / %x41-46 / %x61-66
```

Note that although both the <substring> and <present> productions in the grammar above can produce the "attr=*" construct, this construct is used only to denote a presence filter.

The <valueencoding> rule provides that the octets that represent the ASCII characters "*" (ASCII 0x2a), "(" (ASCII 0x28), ")" (ASCII 0x29), "\" (ASCII 0x5c), NUL (ASCII 0x00), and all octets greater than 0x7f are represented as a backslash "\" (ASCII 0x5c) followed by the two hexadecimal digits representing the value of the encoded

octet.

This simple escaping mechanism eliminates filter-parsing ambiguities and allows any filter that can be represented in LDAP to be represented as a NUL-terminated string. Other octets that are part of the <normal> set may be escaped using this mechanism, for example, non-printing ASCII characters.

For AssertionValues that contain UTF-8 character data, each octet of the character to be escaped is replaced by a backslash and two hex digits, which form a single octet in the code of the character.

For example, the filter checking whether the "cn" attribute contained a value with the character "*" anywhere in it would be represented as "(cn=*\2a*)".

7. Examples

This section gives a few examples of search filters written using this notation.

```
(cn=Babs Jensen)
(!(cn=Tim Howes))
(&(objectClass=Person)(|(sn=Jensen)(cn=Babs J*)))
(o=univ*of*mich*)
(seeAlso=)
```

The following examples illustrate the use of extensible matching.

```
(cn:1.2.3.4.5:=Fred Flintstone)
(cn:=Betty Rubble)
(sn:dn:2.4.6.8.10:=Barney Rubble)
(o:dn:=Ace Industry)
(:1.2.3:=Wilma Flintstone)
(:dn:2.4.6.8.10:=Dino)
(:=Fred Flintstone)
```

The first example shows use of the matching rule "1.2.3.4.5".

The second example demonstrates use of a MatchingRuleAssertion form without a matchingRule.

The third example illustrates the use of the ":dn" notation to indicate that matching rule "2.4.6.8.10" should be used when making comparisons, and that the attributes of an entry's distinguished name should be considered part of the entry when evaluating the match.

The fourth example denotes an equality match, except that DN components should be considered part of the entry when doing the match.

The fifth example is a filter that should be applied to any attribute supporting the matching rule given (since the attr has been omitted).

The sixth example is also a filter that should be applied to any attribute supporting the matching rule given. Attributes supporting the matching rule contained in the DN should also be considered.

The seventh and final example is a filter that should be applied to any attribute (since both the attr and matching rule have been omitted).

The following examples illustrate the use of the escaping mechanism.

```
(o=Parens R Us \28for all your parenthetical needs\29)
(cn=*\2A*)
(filename=C:\5cMyFile)
(bin=\00\00\00\04)
(sn=Lu\c4\8di\c4\87)
(1.3.6.1.4.1.1466.0;binary=\04\02\48\69)
```

The first example shows the use of the escaping mechanism to represent parenthesis characters. The second shows how to represent a "*" in an assertion value, preventing it from being interpreted as a substring indicator. The third illustrates the escaping of the backslash character.

The fourth example shows a filter searching for the four-byte value 0x00000004, illustrating the use of the escaping mechanism to represent arbitrary data, including NUL characters.

The fifth example illustrates the use of the escaping mechanism to represent various non-ASCII UTF-8 characters.

The sixth and final example demonstrates assertion of a BER encoded value.

8. Security Considerations

This memo describes a string representation of LDAP search filters. While the representation itself has no known security implications, LDAP search filters do. They are interpreted by LDAP servers to select entries from which data is retrieved. LDAP servers should take care to protect the data they maintain from unauthorized access.

Please refer to the Security Considerations sections of [Protocol] and [AuthMeth] for more information.

9. Normative References

[ASN.1] Specification of ASN.1 encoding rules: Basic, Canonical, and Distinguished Encoding Rules, ITU-T Recommendation X.690, 1994.

[AuthMeth] Harrison, R. (editor), "LDAP: Authentication Methods and Connection Level Security Mechanisms", draft-ietf-ldapbis-authmeth-xx.txt, a work in progress.

[Protocol] Sermersheim, J. (editor), "LDAP: The Protocol", draft-ietf-ldapbis-protocol-xx.txt, a work in progress.

[RFC2234] Crocker, D., Overell, P., "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, November 1997.

[RFC2279] Yergeau, F., "UTF-8, a transformation format of ISO 10646", RFC 2279, January 1998.

[Roadmap] Zeilenga, K. (editor), "LDAP: Technical Specification Road Map", draft-ietf-ldapbis-roadmap-xx.txt, a work in progress.

[Syntaxes] Dally, K. (editor), "LDAP: Syntaxes", draft-ietf-ldapbis-syntaxes-xx.txt, a work in progress.

10. Acknowledgments

This document replaces RFC 2254 by Tim Howes. Changes included in this revised specification are based upon discussions among the authors, discussions within the LDAP (v3) Revision Working Group (ldapbis), and discussions within other IETF Working Groups. The contributions of individuals in these working groups is gratefully acknowledged.

11. Authors' Address

Mark Smith, Editor
Netscape Communications Corp.
360 W. Caribbean Drive
Sunnyvale, CA 94089
USA
+1 650 937-3477
mcs@netscape.com

Tim Howes
Loudcloud, Inc.
599 N. Mathilda Ave.
Sunnyvale, CA 94086
USA
+1 408 744-7509
howes@loudcloud.com

12. Full Copyright Statement

13. Appendix A: Changes Since RFC 2254

13.1. Technical Changes

"String Search Filter Definition" section: replaced the "value" rule with a new "assertionvalue" rule within the "simple", "extensible", and "substring" ("initial", "any", and "final") rules. This matches a change made in [Syntaxes].
Revised the "attr", "matchingrule", and "assertionvalue" ABNF to more precisely reference productions from the [Protocol] and [Syntaxes] documents.

Introduced the "valueencoding" and associated "normal" and "escaped" rules to reduce the dependence on descriptive text.
Added a third option to the "extensible" production to allow creation of a MatchingRuleAssertion that only has a matchValue.

13.2. Editorial Changes

Changed document title to include "LDAP:" prefix.

IESG Note: removed note about lack of satisfactory mandatory authentication mechanisms.

Header and "Authors' Addresses" sections: added Mark Smith as the document editor and updated affiliation and contact information.

"Table of Contents" section: added.

Copyright: updated the year.

"Abstract" section: separated from introductory material.

"Introduction" section: new section; separated from the Abstract. Updated second paragraph to indicate that RFC 2254 is replaced by this document (instead of RFC 1960). Added reference to the [Roadmap] document.

"LDAP Search Filter Definition" section: made corrections to the LDAPv3 search filter ABNF so it matches that used in [Protocol].

"String Search Filter Definition" section: clarified the definition of 'value' (now 'assertionvalue') to take into account the fact that it is not precisely an AttributeAssertion from [Protocol] section 4.1.6 (special handling is required for some characters). Added a note that each octet of a character to be escaped is replaced by a backslash and two hex digits, which represent a single octet.

"Examples" section: added five additional examples: (seeAlso=), (cn:=Betty Rubble), (:1.2.3:=Wilma Flintstone), (:=Fred Flintstone), and (1.3.6.1.4.1.1466.0;binary=\04\02\48\69). Replaced one occurrence of "a value" with "an assertion value".

"Security Considerations" section: added references to [Protocol] and [AuthMeth].

"Normative References" section: renamed from "References" per new RFC guidelines. Changed from [1] style to [Protocol] style throughout the

document. Added entries for [AuthMeth] and [Roadmap] and updated UTF-8 reference to RFC 2279. Replaced RFC 822 reference with a reference to RFC 2234.

"Acknowledgments" section: added.

"Appendix A: Changes Since RFC 2254" section: added.

"Appendix B: Changes Since Previous Document Revision" section: added.

14. Appendix B: Changes Since Previous Document Revision

This appendix lists all changes relative to the last published revision, draft-ietf-ldapbis-filter-02.txt. Note that these changes are also included in Appendix A, but are included here for those who have already reviewed draft-ietf-ldapbis-filter-02.txt. This section will be removed before this document is published as an RFC.

14.1. Technical Changes

None.

14.2. Editorial Changes

"Abstract" section: separated from introductory material.

"Table of Contents" section: moved to correct location (after Abstract).

"Introduction" section: new section; separated from the Abstract.

"LDAP Search Filter Definition " section: updated section references to match current LDAPBis drafts. Made minor changes to the ASN.1 so it exactly matches that used in the Protocol document.

"Normative References" section: renamed from "References" per new RFC guidelines; changed author names to "Last, F." format for consistency.

"Authors' Address" section: updated Mark Smith's postal address.