This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

American National Standard

ANSI/INCITS/ISO/IEC TR 19075-1:2011[2015]

(ISO/IEC TR 19075-1:2011, IDT)

Information technology - Database languages - SQL Technical Reports - Part 1: XQuery Regular Expression Support in SQL

Developed by

Where IT all begins

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

Adopted by INCITS (InterNational Committee for Information Technology Standards) as an American National Standard.

Date of ANSI Approval: 2/1/2015

Published by American National Standards Institute,
25 West 43rd Street, New York, New York 10036

Copyright 2015 by Information Technology Industry Council
(ITI). All rights reserved.

These materials are subject to copyright claims of International Standardization Organization (ISO), International Electrotechnical Commission (IEC), American National Standards Institute (ANSI), and Information Technology Industry Council (ITI). Not for resale. No part of this publication may be reproduced in any form, including an electronic retrieval system, without the prior written permission of ITI. All requests pertaining to this standard should be submitted to ITI, 1101 K Street NW, Suite 610, Washington DC 20005.
Printed in the United States of America

© ITIC 2015 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

TECHNICAL **ISO/IEC**

19075-1

First edition
2011-07-15

Information technology — Database languages — SQL Technical Reports —

Part 1: XQuery Regular Expression Support in SQL

Technologies de l'information — Langages de base de données — Rapport techniques SQL —

Partie 1: Support d'expressions régulières de XQuery en SQL

Reference number
ISO/IEC TR 19075-1:2011(E)

© ISO/IEC 2011

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2011

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

© ISO/IEC 2011 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

Contents

Page

Foreword v
Introduction vi
1 Scope 1
2 XQuery regular expressions 3
2.1 Matching a specific character 3
2.2 Metacharacters and escape sequences 4
2.3 Dot 5
2.4 Anchors 6
2.5 Line terminators 6
2.6 Bracket expressions 7
2.6.1 Listing characters 7
2.6.2 Matching a range 8
2.6.3 Negation 8
2.6.4 Character class subtraction 8
2.7 Alternation 8
2.8 Quantifiers 9
2.9 Locating a match 10
2.10 Capture and back-reference 11
2.11 Precedence 12
2.12 Modes 13
3 Operators using regular expressions 15
3.1 LIKE_REGEX 15
3.2 OCCURRENCES_REGEX 16
3.3 POSITION_REGEX 17
3.4 SUBSTRING_REGEX 19
3.5 TRANSLATE_REGEX 20
Bibliography 23
Index 25

©ISO/IEC 2011 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

(Blank page)

 ©ISO/IEC 2011 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide to publish a Technical Report. A Technical Report is entirely informative in nature and shall be subject to review every five years in the same manner as an International Standard.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 19075-1 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 32, *Data management and interchange*.

ISO/IEC TR 19075 consists of the following parts, under the general title *Information technology — Database languages — SQL Technical Reports*:

— *Part 1: XQuery Regular Expression Support in SQL*

©ISO/IEC 2011 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

Introduction

The organization of this part of ISO/IEC TR 19075 is as follows:

1) Clause 1, “Scope”, specifies the scope of this part of ISO/IEC TR 19075.

2) Clause 2, “XQuery regular expressions”, explains how XQuery regular expressions are formed.

3) Clause 3, “Operators using regular expressions”, explains how the SQL operators use regular expressions.

 ©ISO/IEC 2011 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

Information technology — Database languages — SQL Technical Reports —

Part 1:
XQuery Regular Expression Support in SQL

1 Scope

This Technical Report describes the regular expression support in SQL adopted from the regular expression syntax of [XQuery F&O], which is derived from Perl. This Technical Report discusses five operators using this regular expression syntax:

— LIKE_REGEX predicate, to determine the existence of a match to a regular expression.

— OCCURRENCES_REGEX numeric function, to determine the number of matches to a regular expression.

— POSITION_REGEX function, to determine the position of a match.

— SUBSTRING_REGEX function, to extract a substring matching a regular expression.

— TRANSLATE_REGEX function, to perform replacements using a regular expression.

©ISO/IEC 2011 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

(Blank page)

 ©ISO/IEC 2011 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.

2 XQuery regular expressions

XQuery regular expression syntax is specified in [XQuery F&O], section 7.6.1, “Regular expression syntax”. This paper references the XQuery specification, with two small modifications (required since character strings in an RDBMS are not necessarily normalized according to XML conventions). The following subsections provide an overview of this syntax.

The XQuery regular expression syntax is itself a modification of another regular expression syntax found in [XML Schema: Datatypes].

This section presents an overview of the capabilities of XQuery regular expression syntax. In the process, this section will illustrate some of the SQL operators. The SQL operators themselves are presented in the next section.

The following discussion does not cover every aspect of XQuery regular expressions; for this, [XQuery F&O] is the reference (though hardly a tutorial; a variety of popular works contain detailed treatments of regular expressions).

2.1 Matching a specific character

Perhaps the most elementary pattern matching requirement is the ability to match a single character or string. For most characters, this is done by simply writing the character in the regular expression. For example, suppose you want to know if a string S contains the letters “xyz”. This could be done with the following predicate:

```
S LIKE_REGEX 'xyz'
```

Note that the SQL LIKE predicate would require an exact match for “xyz”. However, the convention with regular expressions is that S need only contain a substring that is “xyz”. For example, all of the following values of S would yield *True* for the predicate above:

```
xyz
abcxyz123
1 xyz 2 xyz 3 xyz
```

Note that in the last example, there are actually three occurrences of the regular expression “xyz” within the tested value. The user may wish to know the number of occurrences of a match. This can be done with OCCURRENCES_REGEX. For example:

```
OCCURRENCES_REGEX ('xyz' IN '1 xyz 2 xyz 3 xyz') = 3
```

The user might also wish to know the position of a specific match. This can be done using POSITION_REGEX. For example, to learn the starting character position of the second occurrence,

```
POSITION_REGEX ( 'xyz' IN '1 xyz 2 xyz 3 xyz' OCCURRENCE 2 ) = 9
```

It is also possible to ask for the character position of the first character after the match. For example:

©ISO/IEC 2011 – All rights reserved

This is a preview of "ANSI/INCITS/ISO/IEC ...". Click here to purchase the full version from the ANSI store.


```
POSITION_REGEX ( AFTER 'xyz' IN '1 xyz 2 xyz 3 xyz' OCCURRENCE 2 ) = 12
```

If AFTER is used and the last character of the subject string is consumed, then the result is the length of the string plus 1 (one):

```
POSITION_REGEX ( AFTER 'xyz' IN 'xyz' ) = 4
```

2.2 Metacharacters and escape sequences

As mentioned, most characters can be matched by simply writing the character in the regular expression. However, certain characters are reserved as *metacharacters*. The complete list of metacharacters is:

```
. \ ? * + { } ( ) | [ ] ^ $
```

The use of each of these metacharacters will be explained later. If you want to match a metacharacter, then you need to use an *escape sequence*, consisting of a backslash ("`/`") followed by the metacharacter. For example, to test whether a string contains a dollar sign, you could write

```
S LIKE_REGEX '\$'
```

In particular, the escape sequence representing a backslash is two consecutive backslashes. There are various other defined escape sequences, matching either a single character, or any of a group of characters. The *single character escape sequence*s are:

\n	newline (U+000A)
\r	return (U+000D)
\t	tab (U+0009)
\-	minus sign ('-')

The so-called *category escape*s are exemplified by "`\p{L}`" or "`\p{Lu}`". A category escape begins with "`\p{`" followed by one uppercase letter, optionally a lowercase letter, and then the closing brace. In these example, "`\p{L}`" matches any letter (as defined by Unicode) and "`\p{Lu}`" matches any uppercase letter. Some interesting category escapes are listed below:

\p{L}	Any letter.
\p{Lu}	Any uppercase letter.
\p{Ll}	Any lowercase letter.
\p{Nd}	Any decimal digit.
\p{P}	Any punctuation mark.
\p{Z}	Any separator (space, line, paragraph, *etc*.).

The complete list of category escapes is found in [XML Schema: Datatypes], section F.1.1, "Character class escapes".

There are also *complementary category escape*s, which are exemplified by "`\P{L}`" or "`\P{Lu}`". A complementary category escape matches any character that would not be matched by the corresponding category

 ©ISO/IEC 2011 – All rights reserved