INTERNATIONAL STANDARD

ISO/IEC 19075-5

First edition

Information technology — Guidance for the use of database language SQL —

Part 5: Row pattern recognition

PROOF/ÉPREUVE

Reference number
ISO/IEC 19075-5:2021(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Tables

Table

Page

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents), or the IEC list of patent declarations received (see patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 32, *Data management and interchange*.

This first edition of ISO/IEC 19075-5 cancels and replaces ISO/IEC TR 19075-5:2016.

This document is intended to be used in conjunction with the following editions of the parts of the ISO/IEC 9075 series:

— ISO/IEC 9075-1, sixth edition or later;

— ISO/IEC 9075-2, sixth edition or later;

— ISO/IEC 9075-3, sixth edition or later;

— ISO/IEC 9075-4, seventh edition or later;

— ISO/IEC 9075-9, fifth edition or later;

— ISO/IEC 9075-10, fifth edition or later;

— ISO/IEC 9075-11, fifth edition or later;

— ISO/IEC 9075-13, fifth edition or later;

— ISO/IEC 9075-14, sixth edition or later;

— ISO/IEC 9075-15, second edition or later;

— ISO/IEC 9075-16, first edition or later.

A list of all parts in the ISO/IEC 19075 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/-national-committees.

Introduction

This document discusses the syntax and semantics for recognizing patterns in rows of a table, as defined in ISO/IEC 9075-2.

The organization of this document is as follows:

1) Clause 1, "Scope", specifies the scope of this document.

2) Clause 2, "Normative references", identifies additional standards that, through reference in this document, constitute provisions of this document.

3) Clause 3, "Terms and definitions", defines the terms and definitions used in this document.

4) Clause 4, "Row pattern recognition: FROM clause", discusses Feature R010, "Row pattern recognition: FROM clause".

5) Clause 5, "Expressions in MEASURES and DEFINE", discusses scalar expression syntax in row pattern matching.

6) Clause 6, "Row pattern recognition: WINDOW clause", discusses Feature R020, "Row pattern recognition: WINDOW clause". Clause 6, "Row pattern recognition: WINDOW clause", does not duplicate material already presented in Clause 4, "Row pattern recognition: FROM clause" and Clause 5, "Expressions in MEASURES and DEFINE", which should be read even if the reader is only interested in Feature R020, "Row pattern recognition: WINDOW clause".

7) Clause 7, "Pattern matching rules", discusses the formal rules of pattern matching.

Information technology — Guidance for the use of database language SQL —

Part 5:
Row pattern recognition

1 Scope

This document discusses the syntax and semantics for recognizing patterns in rows of a table, as defined in ISO/IEC 9075-2, commonly called "SQL/RPR".

SQL/RPR defines two features regarding row pattern recognition:

— Feature R010, "Row pattern recognition: FROM clause"

— Feature R020, "Row pattern recognition: WINDOW clause"

These two features have considerable syntax and semantics in common, the principle difference being whether the syntax is placed in the FROM clause or in the WINDOW clause.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 9075-1, *Information technology — Database languages — SQL — Part 1: Framework (SQL/Framework)*

ISO/IEC 9075-2, *Information technology — Database languages — SQL — Part 2: Foundation (SQL/Foundation)*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 9075-1 apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

4 Row pattern recognition: FROM clause

4.1 Context of row pattern recognition

The requirements for the material discussed in this document shall be as specified in ISO/IEC 9075-1 and ISO/IEC 9075-2.

4.2 Introduction to the FROM clause in row pattern recognition

Feature R010, "Row pattern recognition: FROM clause" of SQL/RPR enhances the capability of the FROM clause with a MATCH_RECOGNIZE clause to specify a row pattern. The syntax and semantics of a row pattern is discussed through examples presented throughout this Clause of this document.

There are two principal variants of the MATCH_RECOGNIZE clause:

1) ONE ROW PER MATCH, which returns a single summary row for each match of the pattern (the default).
2) ALL ROWS PER MATCH, which returns one row for each row of each match. There are three suboptions, to control whether to also return empty matches or unmatched rows.

4.3 Example of ONE ROW PER MATCH

The following example illustrates MATCH_RECOGNIZE with the ONE ROW PER MATCH option. Let Ticker (Symbol, Tradeday, Price) be a table with three columns representing historical stock prices. Symbol is a character column, Tradeday is a date column, and Price is a numeric column.

NOTE 1 — All examples in this document use mixed-case identifiers for the names of tables, columns, etc., whereas SQL key words are shown in uppercase. Unquoted identifiers are actually equivalent to uppercase, so the column headings of sample results will be shown with the identifiers converted to uppercase.

It is desired to partition the data by Symbol, sort it into increasing Tradeday order, and then detect maximal "V" patterns in Price: a strictly falling price, followed by a strictly increasing price. For each match to a V pattern, it is desired to report the starting price, the price at the bottom of the V, the ending price, and the average price across the entire pattern.

The following query may be used to solve this pattern matching problem:

```
SELECT M.Symbol, /* ticker symbol */
       M.Matchno, /* sequential match number */
       M.Startp, /* starting price */
       M.Bottomp, /* bottom price */
       M.Endp, /* ending price */
       M.Avgp /* average price */
FROM Ticker
       MATCH_RECOGNIZE (
         PARTITION BY Symbol
         ORDER BY Tradeday
         MEASURES MATCH_NUMBER() AS Matchno,
```

```
            A.Price AS Startp,
            LAST (B.Price) AS Bottomp,
            LAST (C.Price) AS Endp,
            AVG (U.Price) AS Avgp
    ONE ROW PER MATCH
    AFTER MATCH SKIP PAST LAST ROW
    PATTERN (A B+ C+)
    SUBSET U = (A, B, C)
    DEFINE /* A defaults to True, matches any row */
           B AS B.Price < PREV (B.Price),
           C AS C.Price > PREV (C.Price)
) AS M
```

In the example above, the principal syntactic elements of MATCH_RECOGNIZE are presented on separate lines. In this example:

— Ticker is the name of the row pattern input table. In this example, the row pattern input table is a table or view. The row pattern input table may also be a derived table (in-line view).

— MATCH_RECOGNIZE introduces the syntax for row pattern recognition.

— PARTITION BY specifies how to partition the row pattern input table. The PARTITION BY clause is a list of columns of the row pattern input table. This clause is optional; if omitted, there are no row pattern partitioning columns, and the entire row pattern input table constitutes a single row pattern partition.

— ORDER BY specifies how to order the rows within row pattern partitions. The ORDER BY clause is a list of columns of the row pattern input table. This clause is optional; if omitted, the order of rows in row pattern partitions is completely non-deterministic. However, since non-deterministic ordering will defeat the purpose of most row pattern recognition, the ORDER BY clause will usually be specified.

— MEASURES specifies row pattern measure columns, whose values are calculated by evaluating expressions related to the match. The first row pattern measure column in this example uses the special nullary function MATCH_NUMBER(), whose value is the sequential number of a match within a row pattern partition. The third and fourth row pattern measure columns in this example use the LAST operation, which obtains the value of an expression in the last row that is mapped by a row pattern match to a row pattern variable. LAST is one of the row pattern navigation operations introduced by SQL/RPR, discussed in Subclause 5.6, “Row pattern navigation operations”.

  The result of the MATCH_RECOGNIZE clause is called the row pattern output table. When ONE ROW PER MATCH is specified, as in this example, the row pattern output table has one column for each row pattern partitioning column and one column for each row pattern measure column.

— ONE ROW PER MATCH specifies that the row pattern output table will have a single row for each match that is found in the row pattern input table.

— AFTER MATCH SKIP clause specifies where to resume looking for the next row pattern match after successfully finding a match. In this example, AFTER MATCH SKIP PAST LAST ROW specifies that pattern matching will resume after the last row of a successful match.

— PATTERN specifies the row pattern that is sought in the row pattern input table. A row pattern is a regular expression using primary row pattern variables. In this example, the row pattern has three primary row pattern variables (A, B, and C).

— SUBSET defines the union row pattern variable U as the union of the primary row pattern variables A, B, and C.

— DEFINE specifies the Boolean condition that defines a primary row pattern variable; a row shall satisfy the Boolean condition in order to be mapped to a particular primary row pattern variable. This example uses PREV, a row pattern navigation operation that evaluates an expression in the previous row. If a primary row pattern variable is not defined in the DEFINE clause, then the definition