

ISO/IEC JTC 1/SC 32 N 0447

Date: 2000-02-15

REPLACES: --

<p style="text-align: center;">ISO/IEC JTC 1/SC 32</p> <p style="text-align: center;">Data Management and Interchange</p> <p style="text-align: center;">Secretariat: United States of America (ANSI) Administered by Pacific Northwest National Laboratory on behalf of ANSI</p>

DOCUMENT TYPE	Other document (Open)
TITLE	Proposal for a new SQL/MM subproject - -SQL/MM Part 6 - Data Mining
SOURCE	SC 32/WG 5
PROJECT NUMBER	1.32.04.01.06.00 [64.06.00.00.00]
STATUS	Proposal for project split approved by SC 32/WG 4 and JTC1/SC32. This is passed to JTC 1 for approval
REFERENCES	
ACTION ID.	ACT
REQUESTED ACTION	JTC 1 approval
DUE DATE	
Number of Pages	5
LANGUAGE USED	English
DISTRIBUTION	P & L Members SC Chair WG Conveners and Secretaries

Douglas Mann, Secretariat, ISO/IEC JTC 1/SC 32

Pacific Northwest National Laboratory *, 901 D Street, SW., Suite 900, Washington, DC, 20024-2115, United States of America

Telephone: +1 703 575 2114; Facsimile: +1 703 681 9180; E-mail: MannD@battelle.org

available from the JTC 1/SC 32 WebSite <http://bwonotes5.wdc.pnl.gov/SC32/JTC1SC32.nsf>

*Pacific Northwest National Laboratory (PNL) administers the ISO/IEC JTC 1/SC 32 Secretariat on behalf of ANSI

DIN NI32/194-99
ISO SC32 WG4 N11
SQL/MM SAF-020
Date: November 4, 1999

ISO

International Organisation for Standardization

ISO/IEC JTC1 SC32
Data Management and Interchange
Secretariat: USA (ANSI)

Title: Proposal for a new SQL/MM subproject -
SQL/MM Part 6 - Data Mining

Source: Germany

Project: ISO/IEC 1.21.64

Status: Proposal for consideration by DIN NI-32 and possible submission to
ISO/IEC JTC1/SC32

References:

- [1] SQL/MM LGW-002, SQL Multimedia and Application Packages (SQL/MM) Project Plan, Paul Cotton, April 1997.
- [2] ISO/IEC JTC1/SC32 N0188 (ISO/IEC FCD 13249-1), Information Technology - Database Languages - SQL Multimedia and Application Packages - Part 1: Framework, October 1998.
- [3] ISO/IEC JTC1/SC32 N0??? (ISO/IEC FDIS 13249-2:199x), Information Technology - Database Languages - SQL Multimedia and Application Packages - Part 2: Full-Text, 1999.
- [4] ISO/IEC FDIS 13249-3:1999, Information Technology - Database Languages - SQL Multimedia and Application Packages - Part 3: Spatial, 1999.
- [5] ISO/IEC 9075-1:1999, Information Technology - Database Languages - SQL - Part 1: Framework.
- [6] ISO/IEC 9075-2:1999, Information Technology - Database Languages - SQL - Part 2: Foundation.
- [7] ISO/IEC 9075-3:1999, Information Technology - Database Languages - SQL - Part 3: Call-Level Interface (SQL/CLI).
- [8] ISO/IEC 9075-4:1999, Information Technology - Database Languages - SQL - Part 4: Persistent Stored Modules (SQL/PSM).
- [9] ISO/IEC 9075-5:1999, Information Technology - Database Languages - SQL - Part 5: Host Language Bindings (SQL/Bindings).
- [10] ISO/IEC JTC1/SC32 WG4: SAF-001, Minutes of the SQL/MM WG4 Meeting and Part 1 - Framework FCD Editing Meeting, May 22-27, 1999, Matsue Japan, Paul Scarponcini, June 1999.
- [11] The Data Mining Group: <http://www.dmg.org>.
- [12] PMML 1.0 - Predictive Model Markup Language, http://www.dmg.org/public/techreports/pmml-1_0.html.
- [13] Microsoft PressPass - Microsoft Announces Vendor Initiative To Create Open Data Mining Specification, <http://www.microsoft.com/PressPass/press/1999/May99/DataMiningpr.htm>
- [14] ISO/IEC JTC1/SC21 N7727, "Criteria for NP's and Program Extensions", April 1993.

Discussion

With the increasing importance of data mining in data processing and especially in decision support applications and data warehouse environments it is essential that a standard interface be defined to permit the execution of data mining functions and the deployment of data mining models through relational database systems.

To this end, Germany seeks to extend the SC32 Project 1.21.64 (SQL/MM [1]) to add a new Subproject for a separate part of SQL/MM that will provide the necessary extensions for data mining in SQL/MM implementations.

Needs

Database Language SQL (ISO/IEC 9075:1989) has become a very important standard in many application areas. The SQL 1992 enhancement (ISO/IEC 9075:1992) adds new features and facilities to make SQL a comprehensive language for definition, manipulation, protection, and administration of table-structured data. A second substantial enhancement (ISO/IEC JTC1 SC32 WG3 project 03.04), called SQL99, has recently been standardized. The SQL99 specification [5]-[9] is a forward-looking SQL enhancement that intends to provide a computationally complete language for defining and managing persistent objects.

An important feature in the SQL99 specification is known as object oriented extensions of the SQL language including user-defined types (UDTs), including methods, object references, subtypes and inheritance, polymorphism, and integration with existing facilities. The base data types in SQL99 include: fixed-length and variable-length character strings, fixed-length and variable-length bit strings, fixed and floating point numerics, dates, times, time stamps, intervals, Booleans, and binary and character large objects.

Since SQL is used extensively in many different application areas, the SQL/MM project was created to define robust packages of specific user-defined types and routines for use in the following areas: Full-Text [3] and Spatial [4]. The project also includes a Framework.

At its meeting in Matsue (Japan) in May 1999, SC32 WG4 felt that "an opportunity exists for OLAP to be developed partly in DBL and partly in MM" [10]. The present proposal is an exact match of that position.

Database users are interested in a standard for data mining in order to be able to arbitrarily combine databases and the data mining based applications. From this perspective, data mining is just a sophisticated tool to extract information or to aggregate the original data. This is pretty similar to the functionalities which are provided by SQL today. Hence, standardizing data mining through SQL/MM would be nothing but a natural extension of today's "simple" storage and retrieval mechanisms.

Existing Practise

Today there are no standards for handling data mining objects. Microsoft has started an initiative to establish a de facto standard by defining data mining extensions for the OLE DB [13] specification. The so-called Data Mining Group (DMG [11]) currently defines a standard for the representation of data mining models in XML (PMML [12]). But there are no standards, de jure

or de facto, for the support of data mining in relational database management systems. There are a number of data mining products implemented on top of various commercial and research DBMS engines but none of these has reached a "critical" mass in the marketplace.

With the standardized definition of Character Large Objects (CLOB's) in SQL99 it is possible to handle the various data mining model formats and functions through a Data Mining data type family.

Expected Stability

Specifications in this standard will be derived from classical definitions in different application areas, or from newly emerging standards in other technical committees, and their interface will be defined using standard SQL definition mechanisms. Packages will be designed to be extensible so that new structures and/or operations can be added in an upward-compatible manner. These specifications will be stable as long as the SQL standard itself is stable.

Program of Work

Specify a new Part of the emerging SQL/MM standard, e.g. "SQL/MM - Part 6: Data Mining" to include the user-defined types and routines for a family of data mining techniques.

Justification for Program Extension

Reference [10] identifies SC32 requirements for approval of program extensions (subdivisions and minor enhancements) of existing SC32 projects. Germany supports the extension of the existing SQL/MM Project (1.21.24) as follows:

- a) The rationale for the SQL/MM Data Mining extension of the SQL/MM project is given in the "Needs" and "Existing Practice" sections above. This proposed work is an extension to the authorized work in the SQL/MM project plan "for robust packages of user-defined type definitions which will have broad appeal across different application areas and could benefit from specification in a single International Standard" [2].
- b) Mr. Friedemann Schwenkreis, a member of DIN NI32, is willing to be the subproject editor for SQL/MM Part 6: Data Mining.