

解説



データベース関連技術の標準化

3. マルチメディアデータベース SQL/MMの標準化[†]

芝野 耕 司^{††}

1. はじめに

マルチメディアデータベースおよびアプリケーションパッケージの標準化であるSQL multimedia and application packages (SQL/MM) プロジェクトは、1993年に正式に国際標準化機構 (ISO) のプロジェクトとして承認され、現在、規格仕様の開発を行っている¹⁾。

このプロジェクトの提案にあたっては、日本規格協会情報技術標準化研究センターフルテキストマルチメディアデータベース調査研究委員会でプロジェクト概要およびプロジェクト提案の検討を行い、アメリカ、カナダなどの主要国と共同で、プロジェクト提案を行った。

この規格は、SQL規格の関連規格として、マルチメディアを中心とした、より応用分野に隣接した規格仕様開発を目指し、開発を行っている。

この規格では、SQL 3のオブジェクト指向拡張を用いて、形式的な仕様定義を行っている。

実際の仕様開発では、すでに大きなアプリケーション分野を持ち、大きな市場と開発会社が存在する分野から順に規格仕様の開発を行っている。そのため、現在は、全文データベース、地理情報データベースおよび科学技術データベースの3つの分野を中心とした仕様の開発を行っている。そして、新しい分野として、インターネットで注目を集めているWWW (World Wide Web) におけるマルチメディア機能の開発を計画している。

プロジェクトとしては、マルチメディアすべてをサポートする仕様を10パート以上に分かれる複数パート規格として、開発することを予定しているが、当初開発としては、ニーズがはっきりと

している。次の4つのパートの開発から行うこととした¹⁾。

- Part 1. フレームワーク
- Part 2. フルテキスト
- Part 3. 地理情報
- Part 4. 汎用機能

これら以外に想定しているパートには、静止画像、動画像、音声、多次元情報などのパートがある。

Part 1.フレームワーク²⁾では、全体構造および仕様記述法を規定し、SQL 3のSQL/MM形式記述への利用法、特に、SQL 3のオブジェクト指向拡張およびユーザ定義構文機能の利用法を規定する。

Part 2.フルテキスト³⁾の主な適用分野は、従来からある全文検索分野に加えて、ATA/AIA (Air Transport Association/Air Space Industry Association) での航空機のSGMLをベースとした電子化マニュアル、CALSなどを主要な適用分野とし、全文データベースのための仕様を規定している。

Part 3. 地理情報⁴⁾は、石油および資源探査分野での地理/空間情報のデータベース機能を中心に開発し、地理情報システムについての国際規格を開発する新しい技術委員会である、新TC (TC 211) と協力して開発。この分野でのAPI機能も提供することを目指している。

また、この規格の開発にあたってアメリカは、これまでデータベース言語を担当してきたANSI X 3 H 2と地理情報システムを担当するANSI X 3 L 1とが共同で対応している。このことによって、データベースベンダだけではなく、地理情報システムの主要なベンダがすべてこの規格の開発に参加している。

Part 4. 汎用機能⁵⁾では、算術、統計関数など

[†] Standardization of Multimedia Database-SQL/MM
by Kohji SHIBANO (Tokyo International University).
^{††} 東京国際大学

の機能を提供する。すなわち、このパートは、これまで科学技術データベースまたは統計データベースとして検討されてきた機能の標準化にあたる。

これらのパートに加えて、インターネットでのマルチメディアサービス、特に、WWWでのマルチメディアサーバとして必要とされるマルチメディアデータベース機能の開発を早急に進めることも計画している。

現在開発を進めているパートのほとんどは、1996年7月にCD (Committee Draft), 1997年初頭にDIS (Draft International Standard), 1997年中にIS (International Standard) にすることを予定している。同時に、データベースベンダおよびアプリケーションベンダでは先取的な実装を行おうとしている。

まず、SQLマルチメディアの基本アーキテクチャについて述べ、次いで、それぞれのパートについて述べよう。

2. SQLマルチメディアの基本アーキテクチャ

SQL Multimedia and Application Packages (SQL/MM) は、SQL規格のコンパニオン規格として、開発を進めている。SQL言語とSQL/MM言語との関係は、プログラム言語における言語の基本構成要素とライブラリとの関係にほぼ相当する。

すなわち、SQLマルチメディアでは、図-1に示すように、SQL言語を用いたSQLのライブラリとして、アプリケーションに隣接するクラスライブラリまたはオブジェクトパッケージの開発を

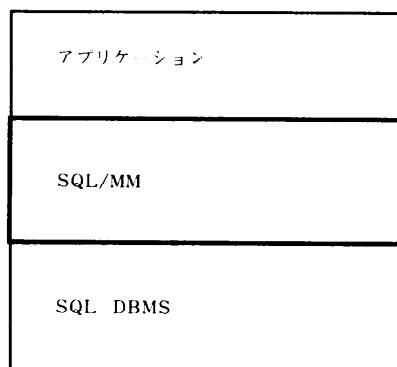


図-1 SQL/MMの基本アーキテクチャ

行っている。一般のマルチメディアを中心とするアプリケーションは、SQL/MMを利用し、開発を行うことを想定している。

こうしたアーキテクチャを想定することによって、従来型のデータを格納する関係データベースと新しいマルチメディアデータを格納するSQL/MMとを統合的に利用可能とすることを目指している。

また、実際の仕様開発を担当する主要なメンバーは、SQLと共通であり、完全にSQL仕様と互換性を保った開発を行っている。実際に、SQL/MMの仕様は、SQLを形式定義言語として利用し、その多くを記述している。この点では、SQL/MMは、SQL仕様の最も重要な、最も初期のユーザの役割を果たしている。

同時に、巨大オブジェクト (Binary Large Object; BLOB および Character Large Object; CLOB) の要求やSQL/MMの言語仕様をSQL言語とシームレスに取り扱うことを可能にするためのSQL側の機能であるユーザ定義演算子など、SQL/MMサポートのために、多くのSQLの機能開発も並行して行ってきた。

一方では、SQL3の機能およびシステムが大規模なものとなることから、SQL3を完全には前提とせず、SQL/MMの機能を直接実現し、SQL-89レベルでの実装でも使えるようにするための配慮も重要であり、SQL/MMへの適合性においても、この配慮をしている。

次にSQL3で記述しているSQL/MM仕様の記述例を示す。この仕様記述は、Part4での数値関数の定義例である。

Part4における数値関数 (arc cosine) の定義例

```
CREATE FUNCTION ArcCsc(x REAL)
  RETURNS REAL
  LANGUAGE SQL
  BEGIN
    DECLARE i COMPLEX DEFAULT
      Complex(0, 1);
    DECLARE z COMPLEX DEFAULT
      Complex(x, 0);
    DECLARE InvalidInput EXCEPTION
      FOR SQLSTATE 'G4F18';
```

```

IF Abs(x)>1
  THEN SIGNAL InvalidInput;
END IF;
RETURN RealPart(-i
  * Log(Invert(z)+
    Sqrt(1-Invert(z * z)));
END

```

3. Part 2 フルテキスト

Part 2 フルテキストでは、全文データベース検索のためのデータベース言語拡張を規定する。

このパートは、このプロジェクトの発端となったパートであり、SGML などによってマーク付けされた構造を持つ全文情報に対しても、データベース機能を提供することを目指している。

このパートは、ATA/AIA で開発していた SGML データベースサポートと日本規格協会が開発していたフルテキストデータベース機能との 2 つの仕様を中心に開発が始まった。

これらの具体的な仕様で先立って、CALIS フェーズ 2 で要求されていた SGML でマーク付けされた構造を持った全文データベース機能を実現することが目標である。

個別の機能としては、大きくは、従来の全文情報検索システムで実現されてきた機能を実現することおよび構造を持った全文情報に対する検索機能を与えることを中心として言語機能の開発を行った。ここでの検索機能は、SQL 1 からある LIKE 述語、SQL 3 の日本で開発した正規表現を含む文字列照合検索を実現する SIMILAR 述語に加えて、単語間の隣接関係およびシソーラスを用いた検索を実現するものである。

具体的には、隣接演算を行う FT_Proxi、発音に関連した表記の揺れを取り扱う FT_Soundex、シソーラスに関連する広義の語に展開する FT_BroaderTerm、狭義の語に展開する FT_NarrowerTerm、同義語展開を行う FT_Synonym、類義語展開を行う FT_PREFERREDTerm、関連語展開を行う FT_RelatedTerm、最上位語展開を行う FT_TopTerm がある。

この全文情報検索に対応する検索機能では、シソーラスを活用する検索機能を取り入れているが、このシソーラス自体も SQL の表として取り

扱うことを可能にすることによって、拡張性を確保している。

これらの検索指定は、SQL の述語と同様に拡張構文で定義され、ユーザからは、SQL 言語の直接的な拡張の形で、定義することとした。

実際に検索を指定する構文の一部を次にあげる。

```

<search primary> ::= <text literal>
  | <asterisk> <text literal>
  | <text literal> <asterisk>
  | <text literal> <plus sign>
  | <text function invocation>
<text function invocation> ::=
  <Proximity function invocation>
  | <Is_About function invocation>
  | <expansion function invocation>
<expansion function invocation> ::=
  <Soundex_Exp function invocation>
  | <Broader_Term function invocation>
  | <Narrower_Term function invocation>
  | <Synonym function invocation>
  | <Preferred_Term function invocation>
  | <Related_Term function invocation>
  | <Top_Term function invocation>
<Proximity function invocation> ::=
  PROXIMITY <left paren>
  <token list 1>
  <comma> <distance> <comma>
  <unit> <comma>
  <order> <comma> <token list 2>
  <right paren>

```

また、シソーラスは、次のように定義している。次に語の階層関係を定義する TERM_HIERARCHY、語の同意関係を定義する TERM_SYNONYM を含む、シソーラス定義の一部をあげる。

```

CREATE SCHEMA SQLMM THESAURUS
  AUTHORIZATION SQLMM_THESAURUS

```

```

CREATE TABLE TERM_DICTIONARY
(
  TERMID INTEGER NOT NULL
  DEFAULT 0,
  EXPR CHARACTER
  VARYING(Maximun_Term_Length),
  THNAME_DIC CHARACTER
  VARYING(Thes_Name_Length),
  PRIMARY KEY(TERMID)
)
CREATE TABLE TERM_HIERARCHY
(
  TERMID INTEGER NOT NULL,
  NARROWER_TERMID INTEGER,
  THNAME_HRR CHARACTER
  VARYING(Thes_Name_Length),
  PRIMARY KEY
  (TERMID, NARROWER_TERMID,
  THNAME_HRR),
  FOREIGN KEY(NARROWER_TERMID)
  REFERENCES TERM_DICTIONARY,
  FOREIGN KEY(TERMID,
  THNAME_HRR)
  REFERENCES TERM_DICTIONARY
)
CREATE TABLE TERM_SYNONYM
(
  TERMID INTEGER NOT NULL,
  SYNONYM_TERMID INTEGER NOT
  NULL,
  PREFERRED_TERMID INTEGER,
  THNAME_SYN CHARACTER
  VARYING(Thes_Name_Length),
  PRIMARY KEY(TERMID,
  SYNONYM_TERMID, THNAME_SYN),
  FOREIGN KEY(SYNONYM_TERMID,
  THNAME_SYN)
  REFERENCES TERM_DICTIONARY,
  FOREIGN KEY(PREFERRED_TERMID,
  THNAME_SYN)
  REFERENCES TERM_DICTIONARY,
  FOREIGN KEY(TERMID,
  THNAME_SYN)
  REFERENCES TERM_DICTIONARY)

```

一方、構造を持った全文に対する検索では、SGMLだけではなく、それ以外の構造を持った全文情報を一般的に取り扱いのための抽象データ型を定義し、これを用いた検索を可能としている。

4. Part 3 地理情報

地理情報のパートは、最も期待されているパートである。すなわち、このパートの開発には、主要なデータベースベンダおよび地理情報システムベンダが参加し、強力にこの仕様をサポートしているのみならず、世界で最も大きなユーザ団体からも強いサポートがある。このユーザ団体とは、石油会社のソフトウェア標準を決定している POSC (Petrorium Oil Software Consorcium) である。

こうした強い要請から、少なくともこのパートに関しては、1997年中にも国際規格の制定が要請されている。

具体的な仕様としては、位相的な演算子を前提に、空間データをサポートする抽象データ型、幾何抽象データ型、メタデータ抽象データ型からなる。

空間データをサポートする抽象データ型に関連して、地理情報で重要ないくつかの定義域を定義している。これらの定義域としては、地図投影法や幾何的な基本性質を表す定義域がある。そして、基本的な2次元および3次元の座標系を空間情報記述のための抽象データ型として規定している。

幾何抽象データ型には、上位の型として、幾何オブジェクトなどの抽象データ型を定義する。図-2に幾何オブジェクトの型階層を示す。点、線、領域などの基本空間オブジェクトについても、抽象データ型として定義する。こうした幾何抽象データ型に関連する特性値を抽象データ型として定義している。

また、それぞれの抽象データ型には、その抽象データ型を取り扱うための関数を定義している。次にこの関数の定義例を示す。

例 曲線 ST_Curve の定義の一部

```

CREATE TYPE ST_Curve
  UNDER ST_Line

```

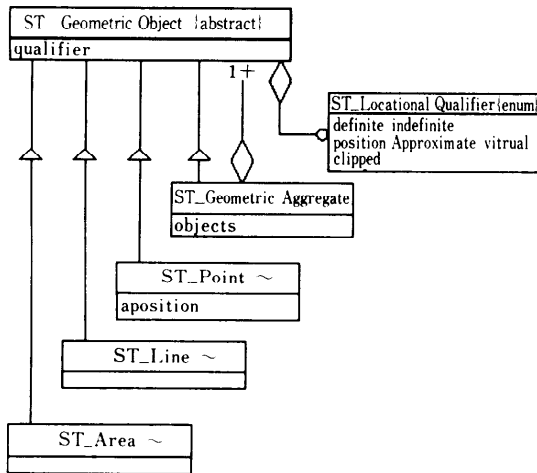


図-2 ST_GeometricObject 型階層図

```

(ST_Points LIST(ST_Point)
  NOT NULL,
  ST_CurveType CHARACTER(1),

```

```

PUBLIC FUNCTION ST_Curve(
  lpts LIST(ST_Point),
  ctype CHARACTER(1))
  RETURNS ST_Curve
  LANGUAGE SQL
  BEGIN
    DECLARE rtn ST_Curve;

    SET rtn=ST_Curve();
    SET rtn..ST_Points=lpts;
    SET rtn..ST_CurveType=ctype;
    RETURN rtn;
  END,

```

このように、このパートでは、最初の規格として、地理情報データベースの取扱いのために最小限必要とする機能に絞って、仕様開発を行っている。

空間情報と時間情報とを統合的に取り扱うことや、資源探査での利用など、より具体的な応用に近い仕様に関しても、開発を行っているが、これらの仕様に関しては、最初の規格ではなく、将来の改正に回すこととしている。

5. Part 4 汎用機能

仕様開発の最初の時点では、この汎用機能は、もともとは、フレームワークに含め、すべてのパートで必要とする汎用的な機能を規定することを目指した。しかし、その後、多岐にわたる開発の結果として、科学技術分野で必要とする多くの数学分野での関数を提供することとなることから、別のパートとして、フレームワークから分離することとした。

また、このパートでは、数学の異なったさまざまな分野の関数を含むことから、他のパートとは異なって、基本的には、各章単位での実装を想定し、この規格の適合性に関しても、各章単位での適合性とする事とした。

各章は、次の抽象データ型の機能仕様からなる。

- 角度
- 配列
- 複素数
- ベクトル
- 数値関数
- ブール集合
- 参照型モデル

角度抽象データ型は、極座標系における演算を定義する。

配列抽象データ型では、配列を取り扱う。この配列の取扱いには、sparseな配列の取扱いおよび1次元と多次元の配列の取扱いを含む。

複素数抽象データ型は、複素数に対する四則演算および実数部、虚数部の取り出しなどの演算を含む。

ベクトル抽象データ型は、線形代数で取り扱われる実数および複素数の有限次元のベクトル空間を取り扱う。ベクトルの四則演算、共役ベクトル、ノルムなどの演算を含む。

また、2次元、3次元のユークリッド空間の取扱いも含む。

数値関数は、基本数値関数、べき乗などの組合せ関数、指数関数、三角関数など一般の数値関数を含む。

ブール集合抽象データ型は、ブール集合に対する基本演算を定義する。

また、これらの抽象データ型のそれぞれには、

データベース検索中で広く利用される比較述語を用いることが可能になるように、IsEqual 関数が定義されている。

これらの抽象データ型を含む、このパートでは、型テンプレートの機能など、SQL 3の仕様確定にあたって将来先送りとしたSQL 4の多くの機能を利用している抽象データ型定義が多くあり、今後の大幅な見直しが必要である。

6. おわりに

現在のところ、SQL/MMでは、フルテキスト、地理情報および汎用機能（科学・数学）の3つの分野での応用よりのデータベース機能拡張クラスライブラリを開発してきた。

このうちのフルテキストおよび地理情報関連に関しては、すでに一部のデータベースベンダで製品としての実装が進んでいる。特に、地理情報分野に関しては、前述のように、データベースベンダだけではなく、主要な地理情報システムベンダが参加し、早期の実装および製品化を予定している。

この点では、SQL/MM規格は、近い将来多くの製品に採用される規格になる可能性がきわめて高い。

この開発では、実際の市場での要求の多い項目を中心とし、具体的には、それぞれの応用分野ごとに特定市場向けの製品が存在するものについて仕様開発を行ってきた。

こうした規格仕様開発の方法論自体は、決して間違ったものではない、と思う。しかし、マルチメディアデータベース機能として見た場合に、現在、中心的に取り組んでいるフルテキスト分野と地理情報分野とだけではどこがマルチメディアであるのか、というそしりを受けかねない。このことは、SQLマルチメディアプロジェクトの最初からの課題であった。

最近では、インターネットでのWWWの急速な普及によって、WWWのバックボーンサーバとしてデータベース管理システムを利用するケースが増えてきており、主要なデータベース管理システムでは、このWebサーバ機能を持つものが増えてきている。SQLマルチメディアでもでき

る限り早急にこうした課題に取り組むことが必要である。

こうした新しくかつ急速に広がってきている要求にも答える必要があり、現在、静止画像を含む、WWWで用いられるマルチメディアをサポートし、Webのバックボーンデータベースとしての機能を実現する仕様の標準化を進める予定である。

また、この新しいパートの開発にあたっては、最近の国際標準化機構が推進している市場要求にタイムリーに取り組むことを可能にする手段を採用し、1~2年程度で、Webマルチメディアサポートに関しても、マルチメディアデータベースの標準化の一貫として取り組むことを検討している。

参考文献

- 1) ISO, SQL/MM Project Plan, ISO/IEC JTC 1/SC 21/WG 3 SQL/MM MCI-02(1996).
- 2) ISO, SQL/MM Part 1: Framework, ISO/IEC JTC 1/SC 21/WG 3 SQL/MM MCI-03(1996).
- 3) ISO, SQL/MM Part 2: Full-Text, ISO/IEC JTC 1/SC 21/WG 3 SQL/MM MCI-04(1996).
- 4) ISO, SQL/MM Part 3: Spatial, ISO/IEC JTC 1/SC 21/WG 3 SQL/MM MCI-05(1996).
- 5) ISO, SQL/MM Part 4: General Purpose Facilities, ISO/IEC JTC 1/SC 21/WG 3 SQL/MM MCI-06(1996).

(平成8年4月18日受付)



芝野 耕司 (正会員)

1979年京都大学大学院工学研究科数理工学専攻博士課程前期課程修了、三井情報開発(株)、日本アイ・ビー・エム(株)東京基礎研

究所を経て、現在、東京国際大学商学部経営情報学科教授。ISO/IEC SC 21/WG 3 SQL/MM RG 主査、情報規格調査会 SQL SG 主査、SQL/MM SG 主査、JIS X 3005 SQL 原案委員会主査、JIS X 4051 日本語組版原案委員会主査、JIS X 4061 日本語照合順番原案委員会主査、情報規格調査会 SC 2 専門委員会委員長、JISコード委員会委員長など。工業標準調査会、産業構造審議会臨時委員。主な研究テーマ：データベース管理システム、電子文書処理、フルテキストマルチメディアデータベース、日本語処理、文字コード。