

A stylized, light-colored illustration of a plant with several leaves and a cluster of small, round buds or flowers, positioned on the left side of the slide against a dark brown background.

SQL言語の開発と日本の 貢献

東京外国語大学

アジア・アフリカ言語文化研究所

芝野耕司

概要

- Ted Coddが提唱した関係モデルの実装は、当初、IBMで研究開発が行われ、その中で、データベース言語として、SQL言語が開発された。
- その後、SQL言語の開発の場は、ISOに移され、SQL87以来、最新のSQL2013まで大きく機能強化が行われてきた。
- SQL言語の開発で日本は、マルチメディア拡張であるSQL/MMの提案を含め、主要な貢献を行ってきた。
- ここでは、30年超におよぶ開発の中で日本は何を考え、貢献を行ってきたかを振り返る。

関係モデルの提案とその実装

- Ted Coddによる関係モデルの提案
 - Codd, E.F. (1970). “A Relational Model of Data for Large Shared Data Banks”. *Communications of the ACM* 13 (6): 377–387. doi:10.1145/362384.362685.
- PRTV (*Peterlee Relational Test Vehicle*)
 - 当時イギリスのPeterleeにあったIBM UKサイエンティフィックセンターで開発された最初の関係データベース管理システム
 - Todd, Stephen (1976). "The Peterlee Relational Test Vehicle - A System Overview". *IBM Systems Journal* 15 (4): 285–308. doi:10.1147/sj.154.0285.
 - 関係代数に基づくISBL (Information Systems Base Language)による関係モデルのより厳密な実装, アクセスパスの最適化, ユーザ定義関数, UDFを用いた地理情報システムアプリケーション
 - その後, PRTVの中心的研究者のToddがSan Jose研究所に移動し, 関係データベースの研究はSan Joseに移動

SQLの誕生

- ToddがPRTVを持ってきたことからPRTVのDBMSカーネルをもとにSystem-Rの研究がSan Joseで開始
- Don Chamberlinによって開発されたSEQUELをSystem-Rの問合せ言語に採用
 - Chamberlin, Donald D; Boyce, Raymond F (1974). "SEQUEL: A Structured English Query Language". *Proceedings of the 1974 ACM SIGFIDET Workshop on Data Description, Access and Control (Association for Computing Machinery)*: 249–64. Retrieved 2007-06-09.
 - COBOLなどと同様の英語構文に似た構文を採用
 - SEQUELがイギリスで通信機器の登録商標として登録されていたため、SQL (Structured Query Language)に改称
- System-RのカーネルをJim Grayが新規に書き直す
 - J. Gray, "Notes on Database Operating Systems", *Operating Systems, an Advanced Course*, Bayer et. al. eds., Lecture notes in Computer Science 60, Springer-Verlag, 1978, pp. 393-481.
 - 同一DBMSカーネル上にSQL, IMSなど複数のDBインタフェースを実装することを目指す

閑話休題

- System-R関連プロジェクトとして、各国の研究所で自然言語問合せの研究が行われた
 - San Jose, Rendezvous
 - E.F. Codd, "Seven Steps to Rendezvous with the Casual User," *Proc. IFIP Working Conf. Data Base Management*, 1974, pp. 179–200.
 - 日本IBM東京サイエンティフィックセンター, ヤチマタ
 - 鷹尾洋一, 間下浩之, 諸橋正幸, 藤崎哲之助, 渋谷政昭, ヤチマタにおける名詞句データ・モデル, *情報処理学会研究報告データベースシステム (DBS)*, 1976(33(1976-DBS-032)), 1-9 (1977-01-13)
- HTML, XML ← SGML ← GML (IBM Script/VSのマーク付言語)
 - Charles Goldfarb、Edward Mosher、Raymond Lorieによって開発されGMLと命名。LorieはSystem-Rプロジェクトの主要メンバー
 - SQLの設計者Chamberlinも1980年代は文書処理WSの研究に
 - ol, ul, dl, pなどのHTMLのタグ名はGML由来

データベースの標準化

- CODASYL Database
 - COBOLの標準化を行ったCODASYL (Conference/Committee on Data Systems Languages)が1969年にData Base Task Group (DBTG)がネットワークデータベースモデルのための言語仕様を発表。標準化を目指す。
 - IBM委員よりMinority Reportとして、仕様の問題点が指摘され、標準化が延期となる。
- ANSI SPARCでデータベース言語の標準化を検討
 - CCAのBrodieを委員長とする委員会でデータベース言語の標準化を行うことを勧告。ただし、具体構文ではなく、抽象構文としてインタフェースの規定を行うことが勧告された。
 - M.L. Brodie, J. W. Shmidt eds, Final report of the ANSI/X3/SPARC DBS-SG relational database task group, ACM SIGMOD Record Volume 12 Issue 4, July 1982, Pages 1 - 62
 - 1978年にANSI X3H2を設け、NDL(Network Database Language)及びRDL (Relational Database Language)の標準化を開始。
 - ANSIではI (International)プロジェクトとし、他の言語のようにANSI主導の開発ではなく、ISOの開発プロジェクトとして標準化を開始。
 - モジュール言語を抽象構文仕様として採用。

SQL87

- 最初のSQL規格
 - 開発当初の名称RDLをイギリス提案でSQLに変更
 - 抽象構文であるモジュール言語に加えて埋込み構文の標準化
- ANSIでは規格票の出版ではなく、規格の承認日時でANSI SQL86と呼ぶが、ANSIとISOは同一テキストで1987年初めに出版。1987年正月に最終校正を行い、数十件の編集上の問題をISOエディタのPhil Shawに送付し、これも反映して、国際規格を出版。
- この仕様は。もともと当時IBM社内で複数のSQL開発が行われており、社内標準として開発していたものをISO/ANSIに持ち込んだものの発展形。
 - IBM社内のSQL実装
 - メインフレームDOS/VSE, VM/CMS環境SQL/DS, MVS環境DB2, AS400環境SQL400, OS2EE環境の少なくとも4つの主要な製品が存在
- SQL87ではM. StonebrakerのINGRESの問合せ言語であるQUELの制限に対応するため部分実装のレベル1を設けた。

SQL89

- 関係モデルの提唱者であるCoddの批判の最大の問題である整合性制約サポートを追加する拡張
- 日本語機能拡張
 - 日本では、IBM, 日立, 富士通, 日本電気, 東芝, 三菱, 沖電気などの国内各社及びOracle, Sybase, Informixなどと調整し, SQLでの日本語サポート仕様を開発し, SQL92へ国際提案を行うとともに, ISOに了承を取り, JIS SQL89でいち早く, 日本語機能拡張を行い, JIS X 3005-1990で先取り標準化を行った。

SQL92

- SQL言語への批判及び拡張要求への対応のため、大幅な機能拡張を行った。
 - 各国語機能を含むデータ型拡張, 動的SQL, 言語の直交性の改善, 整合性機能拡張, 外部結合, 一時表, 動的SQL, DDL拡張など
 - 日本からは, 各国語機能(日本語サポート), 定義域の追加などの主要機能追加を提案
- マルチパート規格へ
 - SQL/CLI ODBC APIの標準化
 - SQL/PSM 計算完備で格納可能な手続きを含む言語拡張
 - 日本から拡張を提案。

SQLのパート

1. Part 1: Framework (SQL/Framework) 枠組
2. Part 2: Foundation (SQL/Foundation) 基本機能
3. Part 3: Call-Level Interface (SQL/CLI) 呼出しレベルインタフェース
4. Part 4: Persistent Stored Modules (SQL/PSM) 永続格納モジュール
5. Part 9: Management of External Data (SQL/MED) 外部データ管理
6. Part 10: Object Language Bindings (SQL/OLB) オブジェクト言語結合
7. Part 11: Information and Definition Schemas (SQL/Schemata) スキーマ
8. Part 13: SQL Routines and Types Using the Java™ Programming Language (SQL/JRT) Javaで用いるSQLルーチン及び型
9. Part 14: XML-Related Specifications (SQL/XML) XML関連仕様

SQL99

- オブジェクトリレーショナル・データベース機能拡張, 再帰問合せ, OLAPサポート, トリガ, 役割, トランザクション制御機能拡張など
- 配列, 大規模オブジェクト, ユーザ定義型などを含むデータ型
- 日本からは, 副表, ビュー更新規則, 正規表現をサポートするSIMILAR 述語などの主要機能追加を提案

• SQL2003

- ウィンドウ関数, 順序型など

• SQL2008

SQL/MM

- 当時データベース言語グループのラポータ(主査)であったUSのLen Gallagherと共同でプロジェクト提案を行い, SQL担当のデータベース言語グループ(DBL RG)とは別のISO/IEC JTC1/SC21 (OSI)/WG3 SQL/MM RGを設置。その後, SC21の廃止により新しく組織されたSC32 (Data Management and Interchange)配下のWGに名称変更。ISO/IEC JTC1/SC32/WG4 (SQL/MM)
- SQL/MMのパート ISO/IEC 13249 Information Technology --- Database languages --- SQL multimedia and application packages ---
 1. ISO/IEC 13249-1 Part 1: Framework 枠組
 2. ISO/IEC 13249-2 Part 2: Full-Text フルテキスト (FT_, Full-Text)
 3. ISO/IEC 13249-3 --- Part 3: Spatial 地理情報 (ST_, Spatio-Temporal)
 4. ISO/IEC 13249-5 --- Part 5: Still Image 静止画像 (SI_, Still Image)
 5. ISO/IEC 13249-6 --- Part 6: Data Mining データマイニング (DM_, Data Mining)
 6. ISO/IEC TS 13249-7 packages --- Part 7: History 履歴 (HS_, History)
- 最新版
 - ISO/IEC 13249-3: 2016 Information technology -- Database languages -- SQL multimedia and application packages -- Part 3: Spatial 1328 pp.
(SC32/WG4) 発行: 2016-01-15

SQL2011での主な改正点(規格票解説より引用)

- 旧規格, すなわち, JIS X 3005-2:2010 (ISO/IEC 9075-2:2008)からの主な改正点を, 次に示す。
 - モジュール言語AdaでのVARCHAR及びNUMERICデータ型[機能B211]の追加。
 - ルーチン言語AdaでのVARCHAR及びNUMERICデータ型[機能B221]の追加。
 - SQL呼出し関数のDATE型のSQLパラメタに対するTIMESTAMP型の引数の指定を可能にするためのDATE型優先順序リストへのTIMESTAMP型の追加[機能F054]。
 - 実表の全ての行を削除するTRUNCATE TABLE文での, 識別列の値を継続するか又は再開始するかを選択機能[機能F202]の追加(識別列の値の再開始機能の追加を含む。)
 - 条件によって行をINSERTするかUPDATEするかを制御できるMERGE文での, 条件によって行をDELETEする機能[機能F314]の追加。
 - 列にNOT NULL制約を追加する機能及び列のNOT NULL制約を削除する機能[機能F383]の追加。
 - IDENTITY特性句を削除することによって識別列を識別列でない列に変更する機能[機能F384]の追加。
 - 他, 30個

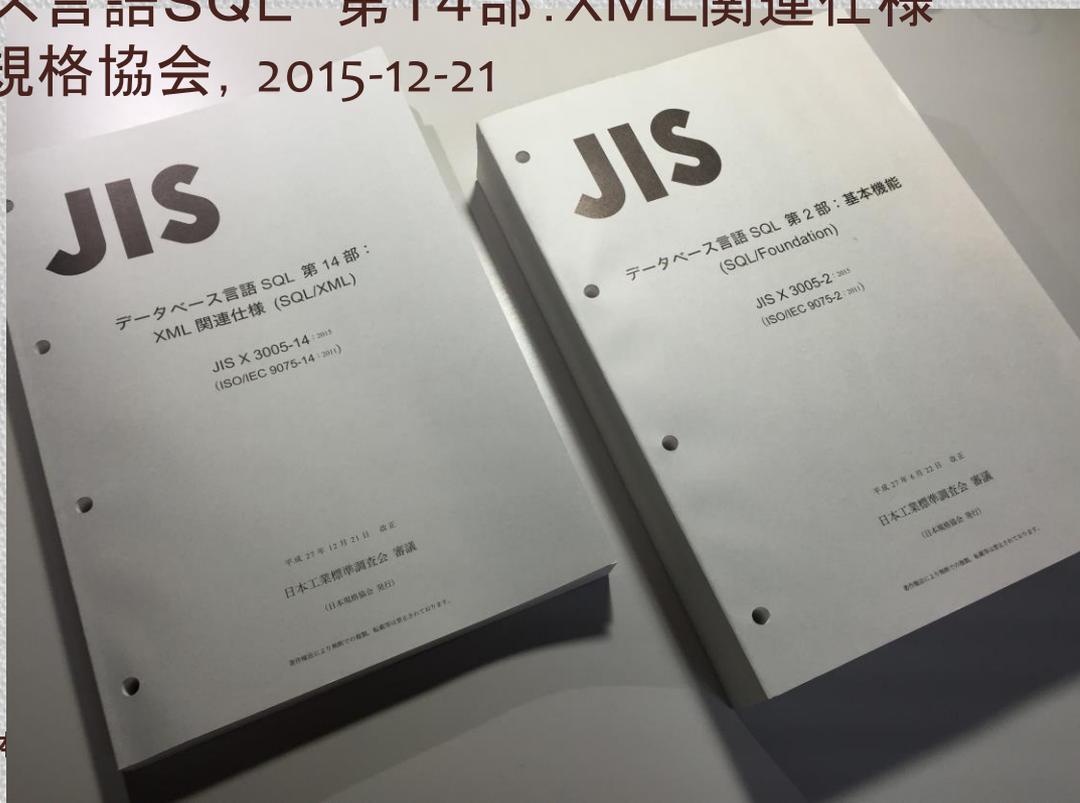
JISとISとの対応

- JIS X 3005:1987(制定) **ISO/IEC 9075:1987 全訳**
- JIS X 3005:1990(改正) **ISO/IEC 9075:1989 全訳**
- JIS X 3005:1995(改正) **ISO/IEC 9075:1992 全訳**
- JIS X 3005-3:1996(制定) ISO/IEC 9075-3:1996
- JIS X 3005:1997(追補1)(改正) ISO/IEC 9075:1992 Technical Corrigendum 1:1996
- JIS X 3005-4:1998(制定) ISO/IEC 9075-4:1996
- JIS X 3005-1:2002(制定), JIS X 3005-2:2002(制定), JIS X 3005-3:2002(改正), 及びJIS X 3005-4:2002(改正) **ISO/IEC 9075:1999**
- JIS X 3005-10:2003(制定) ISO/IEC 9075-10: 2000 (SQL/OLB)
- JIS X 3005-9:2003(制定) ISO/IEC 9075-9:2001 (SQL/MED)
- JIS X 3005-1:2003(追補1)(改正) 及びJIS X 3005-2:2003(追補1)(改正) ISO/IEC 9075: 2001 Amendment 1
- JIS X 3005-14:2006(制定) 及びJIS X 3005-13:2007(制定) JIS X 3005-14:2006(制定) 及びJIS X 3005-13:2007(制定) **ISO/IEC 9075: 2003**
- JIS X 3005-1:2010(改正)(以下, 旧規格という。) 及びJIS X 3005-2:2010(改正) **ISO/IEC 9075: 2008 全訳**
- JIS X 3005-1:2014(改正), JIS X 3005-2:2015(改正) 及びJIS X 3005-14:2015(改正) **ISO/IEC 9075-1:2011 全訳**

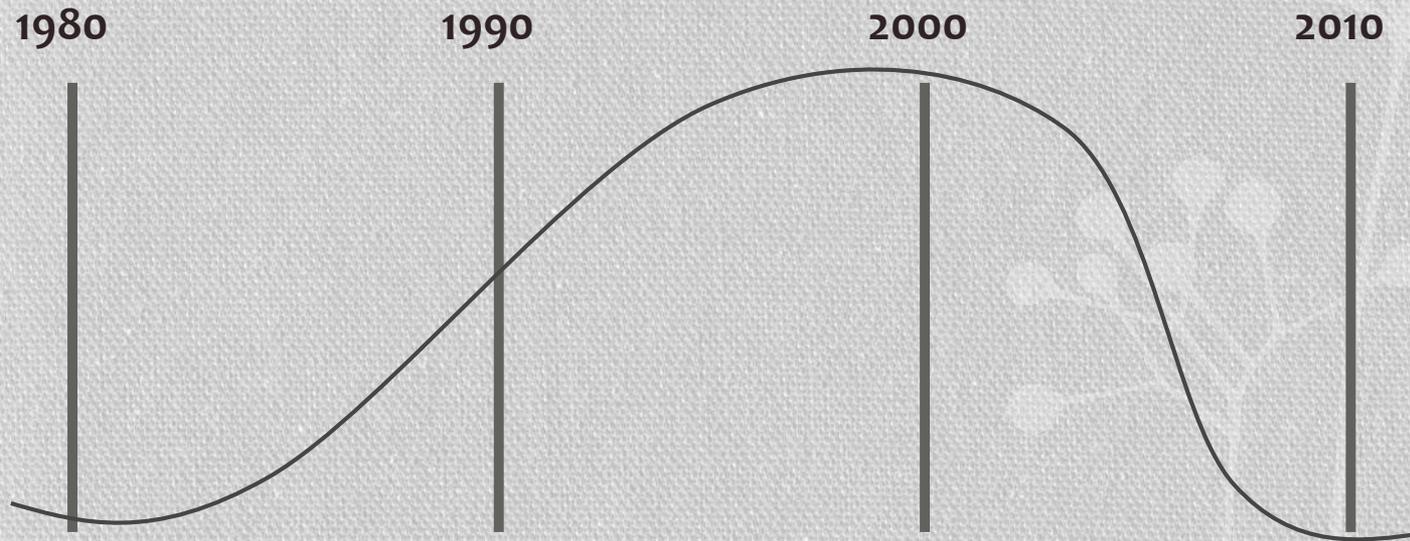
最新のJIS/IS SQL

- JIS X 3005-1: 2014データベース言語SQL 第1部:枠組(SQL/Framework), 60p, 日本規格協会, 2014-08-20
- JIS X 3005-2: 2015 データベース言語SQL第2部:基本機能(SQL/Foundation), 1112p, 日本規格協会, 2015-06-22
- JIS X 3005-2: 2015データベース言語SQL 第14部:XML関連仕様(SQL/XML), 334p, 日本規格協会, 2015-12-21

- 対応国際規格
- ISO/IEC 9075:2011
- (2016年改正予定)



SQLのライフサイクル



- 導入期 ~1987 研究開発から先行製品の発表
- 成長期 ~1995 標準化を契機に多くのSQL製品が出現
- 成熟期 ~2005 M&A, 撤退により商用製品が大幅に淘汰される一方, 主要なOSS製品が出現
- 衰退期

国際で受け入れられていない最近の日本提案

- SQL/Security
 - SQL関連のヴァルネラビリティ対策機能拡張
 - SQLインジェクションなどの典型的なSQL関連セキュリティ問題への対応
 - 既存のSQLセキュリティ機能, 具体的には権限, 役割などに返りレコード数の上限を設定できるようにするなどの機能拡張
- Hash型拡張及び永続格納変数
 - これまでに拡張してきた配列型, マルチセット型に加え, Hash型を追加する。
 - また, 表だけではなく, 変数についてもPSMと同様のスキーマオブジェクトとして永続格納可能とする拡張
 - 昨年, 日立の小寺氏が200ページ弱の提案を準備
 - SQLとNoSQLの機能統合
 - 10月のSC32/WG3中間会議で受け入れられなかった。

日本提案が受け入れられない理由

- USの委員数の大幅な減少。USの開発力の低下から大きな機能拡張を受け入れる可能性が大幅に低下
- 日本においても、委員及びアクティブに参加する企業の大幅な減少。
- OSSによってソフトウェア規格の必要性は減少したとも考えられるが、言語規格などの標準化の必要性は依然として存在する。

現在の研究課題

- JIS・IS関連規格開発
 - SQL, SQL/MMの開発
 - 文字コード関連規格の整備
 - 日本語処理関連のJIS整備
- COE, 特色GP, 現代GP, 教育GP及び大学院GPの遂行
- 大学院生の指導
- → 応用言語学, 教育工学, コーパス言語学, 語用論

現在進行中の研究プロジェクト

- 日本語e-LearningシステムJPLANGプロジェクト
 - ゼロから大学教育を日本語で受けることができるようになるまでの日本語教育をe-Learningで可能とするシステム及びコンテンツ整備
 - 1000時間分の日本語教材の公開
 - 2万人以上の登録ユーザからのアクセスを含む2000万アクセスのサービス
 - Ruby on Railsを基本とする独自開発LMS
- 国際標準に基づく新しい日本語教材の開発
 - CEFR (Common European Framework of Reference for Languages)で規定される場面・機能シラバスに基づくCan-doベースの教材を実際の言語運用に基づくだ〜規模コーパスから作成することを目的とする。
 - 97JIS(漢字コード)の時に収集した日本の全教科書に加え, 5年ほど前に全教科書を購入し, 教科書コーパスを構築
 - 東京キー局の字幕付きテレビ番組を2012年12月以降全部録画し, 字幕コーパスを構築
 - 場面・機能ごとの例文の取得
 - Formulaic Linguisticsにもとづき, 日本語の決まり文句, 定形表現, 慣用句の抽出のためのMap/Reduceを用いたビッグデータ解析