

バーチャル天文台 JVO プロトタイプシステムの開発

Development of a Prototype System of the Japanese Virtual Observatory (JVO)

田中 昌宏^{*1} 白崎 裕治^{*2} 本田 敏志^{*2}
大石 雅寿^{*1} 水本 好彦^{*2} 安田 直樹^{*3}
増永 良文^{*4}

Masahiro TANAKA Yuji SHIRASAKI
Satoshi HONDA Masatoshi OHISHI
Yoshihiko MIZUMOTO Naoki YASUDA
Yoshifumi MASUNAGA

天文学の分野では、観測装置の発達とともに蓄積される観測データ量も膨大になってきている。これらの多波長データを容易に「再観測」し、これまで困難であった研究を可能にするための「装置」がバーチャル天文台(VO)であり、国立天文台では JVO という計画を推進している。VO では世界各地に分散配置された天文データベース間および解析機能との連携を目指している。我々は SQL に対して VO 用に拡張した検索言語を定義し、その検索言語によって分散データサーバ間で連携した検索と解析をおこなう JVO プロトタイプシステムを実装した。本論文はその概要について述べる。

Advance in astronomical instruments has enhanced the amount of observational data. Virtual Observatory (VO) is a tool for "re-observing" those multi-wavelength data and enabling researches that have been difficult. The National Astronomical Observatory of Japan is conducting the Japanese Virtual Observatory (JVO) project. VO aims at federation of astronomical databases and analysis tools distributed around the world. We defined a query language for VO (JVO) by extending the SQL, and implemented JVO prototype systems that perform searches and analyses for federated data servers by using the JVOQL.

1. はじめに

すばる望遠鏡を始めとする世界中の望遠鏡や宇宙望遠鏡など最新の装置による観測および大規模なサーベイ観測により、良質で大量のデータが大量に生み出される時代になった。そうしたデータは一次処理後アーカイブ化され、一定期間経たのち一般に公開される。これは天文学研究においてデ

ータアーカイブの重要性が高まってきているからである。特に電波からガンマ線にわたる広い波長範囲の観測データを使った統計的な研究には単一の望遠鏡による観測データだけでは不可能であり、データアーカイブの活用が必須となる。しかし実際にそのような研究を実行するにはハードルがある。というのは、天文学データをアーカイブして公開する機関は望遠鏡を開発・運用する天文台である場合が多く、公開するデータの形式や方法も天文台ごとに統一されておらず、世界各地に分散配置されたデータアーカイブから研究に必要な数々のデータを取得し、データを変換・加工して解析可能な状態にするまでには、非常に大きな労力が必要となるからである。そこで、究極の天文学データアーカイブ検索システムとして、世界中に分散配置されたアーカイブサーバが連携することにより、世界中のすべての観測データの中から必要なデータを簡単に検索し解析できるシステムが検討され始めている。このようなシステムは、計算機に蓄えられた観測データに基づいて仮想的な宇宙を構築し、それを自在に「観測」するための「天文台」であるということから「Virtual Observatory (VO)」と呼ばれている。VO における大きな課題は、分散配置された計算機資源を連携利用することである。このためグリッドなど、情報処理の分野で最近発展している分散計算機リソースを有効利用する技術の導入が検討されている。

VO 計画は各国で立ち上がっており、国立天文台においても Japanese Virtual Observatory (JVO) [1][2]が進められている。JVO では、独自の VO システムの開発の他、各国の VO が相互連携するために設立された International Virtual Observatory Alliance (IVOA) [3]にも加盟し、VO に関する国際標準仕様の策定などをおこなっている。

VO は、リレーショナルデータベース(RDB)をそのまま集めれば構築できるわけではなく、天文学の要求に応じた機能拡張が必要となる。その中で重要なものは、天球座標、クロスマッチ、および画像・スペクトルデータの扱いである。SQL を拡張することにより、これらを自然に扱うことができる VO 用検索言語 JVOQL を定義した。そして JVOQL で記述された検索命令によって実際に動作するプロトタイプシステムを開発した。本稿ではこれらの概要について述べる。

2. 天文学用検索言語 JVOQL

はじめに VO への検索命令を記述する言語の検討について述べる。SQL に対して以下のような仕様拡張をおこなうことにより、JVO 用検索言語 JVOQL を定義した。

2.1 天文データリソースの命名形式

VO では複数の天文台に分散配置されたデータベースを统一的に扱うことを目指している。その場合、世界中の何百何千とある天文データベースの中から必要なリソースを一意に指定することが必要になる。IVOA では VO におけるデータリソースに対する一意的な命名法が提案されている[4]。JVO におけるテーブル命名方法には、IVOA での提案と互換性を持つドット記法を採用した。テーブルの表記方法は次のようになる。

[[AUTHORITY.]DATABASE.]TABLE [[AS] ALIAS]
テーブル名の先頭にはデータベースを管理する機関に与えられる名前が付き、ドットの後にデータベース名が来る。このデータベース名は機関内で重複しないように命名する。最後にテーブル名が来る。データベース名より上はテーブル名が一意に指定できる条件下では省略可能とする。

*1 正会員 国立天文台 {masahiro.tanaka, masatoshi.ohishi}@nao.ac.jp

*2 非会員 国立天文台 {yuji.shirasaki, honda.satoshi, mizumoto.y}@nao.ac.jp

*3 非会員 東京大学宇宙線研究所
yasuda@icrr.u-tokyo.ac.jp

*4 正会員 お茶の水女子大学理学部情報科学科
masunaga@is.ocha.ac.jp

2.2 天体カタログの検索

天体カタログは、天体の位置や明るさといったデータをテーブルにしたものであり、SQL によって比較的自然的に扱うことができる。しかし SQL にはないが天文学や VO では重要な機能を表現するため、以下のような拡張をおこなった。

2.2.1 天球座標での領域指定

天体の位置や領域の指定には、天球を球面座標で表した天球座標系が用いられる。天球座標系には、赤道座標系や銀河座標系など数種類ある。赤道座標系は歳差により年月とともに移動するため、いつの座標かを示す元期の指定も必要である。そこで天球上の位置の指定には以下の記述方法を考えた。

POINT(経度, 緯度 [, 座標系])

経度・緯度には度を単位とする値か、時・分・秒、度・分・秒形式 (12h34m56.7s, -76d54m32.1s または 12:34:56.7, -76:54:32.1) を指定する。座標系には、赤道座標系の場合は元期によって B1950 または J2000 を指定する。領域指定文は、この位置指定文を用いて次のいずれかのように記述する。

BOX(位置指定文, 幅, 高さ [, 回転角])

CIRCLE(位置指定文, 半径)

POLYGON(位置指定文のリスト)

BOX は領域の形状を四角形で表し、CIRCLE は円形で表す。それぞれ第1引数に中心座標として前述の位置指定文を指定する。第2引数以降で領域の大きさを指定する。領域の大きさには単位として度(deg)、分角(arcmin)、秒角(arcsec)のうちいずれかを付けることができる。省略時の単位は度である。BOX の場合、回転角を指定できる。回転角は四角形の縦方向が座標系の北極へ向かう場合に0度とし、東側への回転を正とする。POLYGON は引数として位置指定文のリストをとり、それらを結んでできる領域を表す。これらの領域指定文を WHERE 句に検索条件として記述できる。

2.2.2 クロスマッチ(Xmatch)検索

クロスマッチ検索とは、異なる天体カタログの中から同じ天体を探すことであり、多波長観測データを使った研究には必須である。クロスマッチは RDB において JOIN をとることと似ている。JOIN と異なるのは、同じ天体かどうかをまず座標位置の近さで判断することである。しかし天体の位置には不定性がある。天体の位置の精度は、望遠鏡の口径や観測波長などで決まる空間分解能や、観測時の空の状態などの条件などによって決まる。したがって同じ天体かどうかの判断は、天体間の角距離がある上限以下という条件でおこなう。クロスマッチは既存の RDB にはないが、天文では必須となるため、次のように XMATCH 条件文を定義した。

XMATCH(テーブル1, [!] テーブル2) < 精度
[NEAREST|BRIGHTEST|ALL]

「精度」により二つのテーブルにある天体間の角距離の上限を指定する。ここでマッチする天体は複数になりうるため、それに続く省略可能なキーワードにより、マッチする天体候補を絞りこむ方法を指定できる。NEAREST を指定した場合は精度の範囲内で位置が一致する天体のうち、最も近くの天体を選ぶ。BRIGHTEST を指定した場合は最も明るい天体を選ぶ。ALL が指定された場合はマッチする天体すべてを返す。テーブル2の前に記号!がある場合は、排他的クロスマッチを行う。すなわち、テーブル1の天体から「精度」により指定される範囲内にテーブル2に天体が存在しないという条件の天体を選ぶ。これは、例えば可視光では暗くて見えないが、赤外では明るく光っているような特殊な天体を選択する場合に有効である。

2.2.3 カラムの分類名

一般にデータベースのカラム名はテーブルごとに異なり、このカラム名を知らなければ SQL を書くことはできない。しかし天文データベースでは、赤経、赤緯(赤道座標系における経度と緯度)や等級など、ある程度決まった種類のカラムが使われる。こうしたカラムを分類する目的で考案されたのが、Unified Column Description (UCD)[5]である。UCD はカラムの意味を示す IVOA における標準的な名前である。例えば赤経は POS_EQ_RA_MAIN、B バンド等級は PHOT_MAG_B と表される。JVOQL では、UCD をカラム名の代わりに使用できることとした。

2.3 画像・スペクトル検索

天体カタログとは異なり、画像やスペクトルに対しては SQL によるテーブル操作とは異なる扱いが必要である。画像データは2次元のグリッドで表われ、各々のピクセルには主に放射強度、すなわち明るさのデータが記録される。天文における画像の特徴に、天球面から平面への投影になることが挙げられる。天文学では、画像の他に、光や電波などの信号を分光して得られたスペクトルデータを扱うことも多い。各天体について測定したスペクトルは、波長あるいは周波数方向の1次元のグリッドデータとなる。画像の各点についてスペクトルを測定した場合は3次元グリッドデータとなり、これをデータキューブと呼ぶ。画像やスペクトルの検索はデータキューブへの範囲指定となる。こうした画像・スペクトルデータを統一的に扱うための記述方法を考察した。まず、画像を四角形で切り出す場合は次のように記述する。

```
SELECT 画像データ.BOX(位置指定文, 幅, 高さ)
FROM 画像データ
```

領域が他の形であれば、BOX(...)の代わりに2.1節で述べた領域指定文を指定することもできる。また、スペクトル切り出しの場合は領域指定文の代わりに SPECTRUM と記述する。さらにカタログ検索条件を次のように同時に記述すれば、検索条件にマッチした天体の画像・スペクトルデータを切り出す、という意味になる。

```
SELECT c.BOX(POINT(c.ra, c.dec), 10 arcsec,
              10 arcsec)
FROM catalog c
WHERE BOX(POINT(270,60), 1 deg, 1 deg)
```

この他に、個々のデータベースを特定せず、波長などをキーにして不特定なカタログ・画像を検索する方法についても考察した。ここではその詳細は割愛するが、その実現方法については今後の課題である。

3. JVO プロトタイプの開発

3.1 システム概要

前章で述べた検索言語の機能を実現することが、JVO プロトタイプの目的の1つである。すべての機能を実現することは困難であるから、まずユースケースを設定し、その機能を実現するシステムの構築を目標とした。プロトタイプ構築後、評価を行い、採用した技術の見直しや改良について検討した。プロトタイプはこれまでに第1版と第2版を開発した。

図1にJVOプロトタイプのシステム構成を示す。利用者はまず Web ブラウザで JVO ポータルサーバにアクセスし、JVOQL で記述した検索命令を実行制御コントローラに送る。コントローラは命令を解析し、レジストリを参照して要求されたサービスを提供するサーバを引き当て、検索命令の実行手順を作成する。この手順に従い、グリッドツールキット

(3.3.4 節を参照) 経由で各サーバに実行要求を送る。

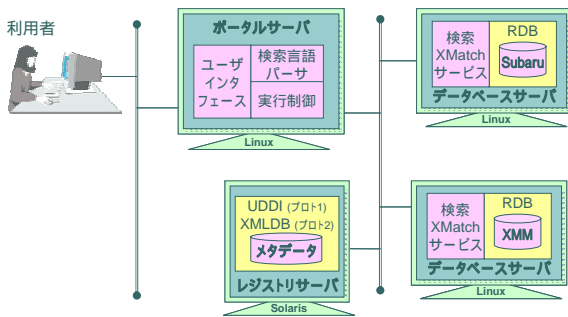


図1 JVO プロトタイプのシステム構成

Fig.1 System Configuration of JVO prototype

3.2 ユースケース

JVO の利用例(ユースケース)として、すばる望遠鏡の主焦点カメラ Suprime-Cam (2k × 4k ピクセル × 10 CCDs)によって撮影された Subaru Deep Field (SDF) という領域のデータ利用を考えた。ここで用いたデータの観測波長は *i'* バンド (683-854nm) と *z'* バンド (834-993nm) の 2 つである。それぞれのバンドごとに、画像データおよびその画像から検出された天体のカタログデータを用意した。分散データベース連携の実証のため、これらはそれぞれ異なるサーバにストアし、検索サービス・クロスマッチサービス・画像切り出しサービスを構築した。

ターゲットとなるクエリは、「*i'* バンドと *z'* バンドでクロスマッチをとり、両方のカタログに含まれる天体のリストおよびその画像を返す」というものである。これを JVOQL で記述すると、以下ようになる。

```
select
  i.POS_EQ_RA_MAIN as ra1,
  i.POS_EQ_DEC_MAIN as dec1,
  i.PHOT_SDSS_I,
  z.POS_EQ_RA_MAIN as ra2,
  z.POS_EQ_DEC_MAIN as dec2,
  z.PHOT_SDSS_Z,
  i.BOX(POINT(ra1,dec1), 0.005, 0.005),
  z.BOX(POINT(ra2,dec2), 0.005, 0.005)
from
  naoj.sdf.i i,
  naoj.sdf.z z
where
  XMATCH(i, z) < 10 arcsec NEAREST and
  BOX(POINT(201., 27.4), 0.10, 0.10) and
  i.PHOT_SDSS_I - z.PHOT_SDSS_Z > 0.0 and
  i.PHOT_SDSS_I < 20.0 and
  z.PHOT_SDSS_Z < 20.0
```

ここで WHERE 句にある XMATCH という条件はテーブル *i* と *z* とでクロスマッチを取ることを表わし、BOX(...) という条件は赤経 201 度、赤緯 27.4 度を中心とする 0.1 度四方の領域以内の天体を検索することを表わす。SELECT 句の i.BOX(...), z.BOX(...) というカラムは、WHERE 句にマッチした天体の位置を中心に、それぞれ *i'* バンドと *z'* バンドの画像を 0.005 度四方で切り出すことを表わす。JVO システムはこのクエリを受け取ると以下のように動作する。(1) JVOQL をパースして個々のサーバへのクエリに分解。(2) 個々のクエリを 2 つのサーバに送り、それぞれ条件にマッチする天体数を取得。(3) マッチした天体数が少ない方

のサーバで検索を実行。結果をもう一方のサーバへ送る。(4) 送られたデータとサーバ内のデータでクロスマッチを実行。(5) クロスマッチの結果を基に、画像を切り出す。(6) 検索結果を元にテーブルデータおよび画像を表示する。このように実検索の前に個数検索をおこなうのは、転送量をできるだけ少なくするためである。

3.3 実装

3.3.1 転送データ形式

検索結果やデータサーバ間でやりとりするデータ形式は、将来他の VO との接続を容易にするため、標準的な形式を用いた。天体カタログデータの場合は VOTable[6]、画像・スペクトルデータの場合は FITS (Flexible Image Transfer System)[7] である。VOTable は、XML でテーブルデータを記述する形式であり、現在の VO における国際標準である。VOTable に記録されるデータは、大雑把に言えば、テーブル形式のデータ本体、およびテーブルのカラムについての付加情報(カラムメタデータ)から成る。一方 FITS は天文学で主に画像を記録するために以前から一般的に使われているデータフォーマットである。FITS もヘッダ部とデータ部からなり、ヘッダ部には、観測時刻、観測装置、観測波長、座標系、単位などの情報を記録できる。

3.3.2 データサーバ

個々のデータサーバには天体カタログ検索サービス、クロスマッチ検索サービス、および画像検索サービスを配置した。

天体カタログ検索サービスの構築には RDB を利用した。このプロトタイプ開発の目的の 1 つは、異種分散データベース同士で連携した検索を実証することである。そこで Solaris マシンと Linux マシンという異機種種のハードウェアを組み合わせたテスト環境を用い、DBMS には Oracle と PostgreSQL の 2 種類を使用した。RDB にはない機能、すなわち領域指定構文を解釈すること、検索結果にカラムメタデータを付加して VOTable で出力する機能などを独自に実装した。また、天球座標を効率的に検索するため、天球座標の領域を 1 次元のインデックスで表わす HTM (Hierarchical Triangular Mesh)[8]を採用した。

クロスマッチ検索は 2 つのテーブル間でおこなう。片方のテーブルは VOTable、もう片方のテーブルは RDB 中のテーブルである。このサービスは、前述の検索サービスを拡張することにより構築した。

さらに 2.3 節で述べた画像検索のため、入力の VOTable に含まれる天体の座標を中心とした FITS 画像を、指定のサイズで切り出す、というサービスを実装した。

3.3.3 実行コントローラ

実行コントローラは、JVOQL で記述された検索条件をパースし、個々のリモートデータサーバへの検索条件へと分解し、実行手順を作成する、という機能を持つ。また、これは動的に変化する実行手順にも対応できるようにした。というのは、クロスマッチ検索は 3.2 節で述べたように、クロスマッチ以外の検索条件に合う天体の個数が少ない方のサーバを先に検索するという、動的な実行手順となるからである。そこで、実行手順の途中でそれまでの実行結果に基づき次回以降のサービスの実行手順を作成するという機能を取り入れた。さらに 1 つのサーバへの実行が失敗したときは別の似た機能を持つサーバへ再試行するという機能により、システムの堅牢性を確保した。

3.3.4 Globus Toolkit の利用

世界中に分散した天文データベース連携のため、ネットワ

ーク接続された計算機間の遠隔実行・データ転送が必要になる。ここにグリッド技術の応用が検討されている。JVO プロトタイプの構築には、標準的なグリッドのミドルウェアとして、Globus Alliance[9]で開発された Globus Toolkit を利用し、VO への応用の適否を検証した。

プロトタイプ 1 の遠隔実行の実装には、Globus Toolkit 2 に付属の globus-job-run というコマンドを用いた。しかしこの方法では 1 回の遠隔実行に 1 分以上かかり、1 回の検索には複数回の遠隔実行が伴うことから全体で 10 分以上もかかった。この理由は、globus-job-run が長時間ジョブの発行という利用方法を想定しており、リアルタイム処理には向かないためである。そこで、プロトタイプ第 2 版では Globus Toolkit 3 で新たに導入された「グリッドサービス」を遠隔実行に用いた。その結果、遠隔実行のオーバーヘッドは 30 ミリ秒程度になり、1 回の検索にかかる時間は数秒ですむようになった。こうしてグリッドによる実用的な分散データベース連携を実証した。

3.3.5 レジストリ

分散計算機のどこにどのような天文データが存在するかわかるように、天文データのメタデータを保管し、検索できるようにしたものをここではレジストリと呼ぶ。プロトタイプ第 1 版では、天体カタログの検索には Web サービスに用いられる UDDI を利用した。しかし UDDI はサービス検索が目的であるため、観測波長・領域といった天文データに特有な情報の検索はできなかった。そこでプロトタイプ第 2 版では、天体カタログのメタデータをレジストリに登録し、レジストリに対して検索することとした。メタデータの仕様は、ここでも他の VO との連携を視野に入れ、国際的な VO における標準化案[10] (XML で記述) と互換性を持つように設計した。このデータを XML データベース(製品「カレアレア」を利用)に投入し、XPath で検索できるようにした。

3.4 結果表示

検索に成功すると、結果はポータルサーバに送られ、利用者は VOTable の内容を Web 上でブラウズできる(図 2)。切り出した画像やスペクトルの早見画像も表示でき、FITS データのダウンロードも可能である。さらに重力レンズ探しといった解析ツールとも連携している。

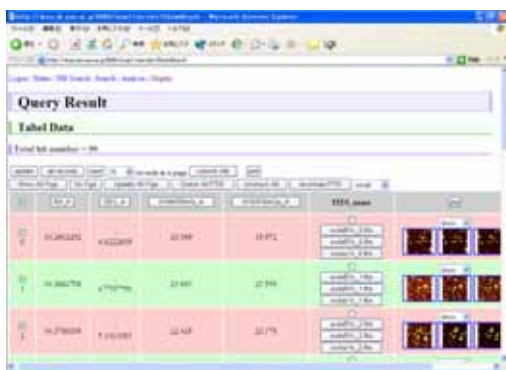


図 2 結果表示画面

Fig.2 Result Display

4. まとめ

バーチャル天文台 JVO 計画では、SQL を拡張した VO 用検索言語を定義し、その言語で記述された検索命令を分散デ

ータサーバ環境で実行する JVO プロトタイプシステムを実装した。このプロトタイプにより、JVO システムの枠組みを確立できた。プロトタイプの評価を踏まえて、今後 JVO 運用システムの開発を計画している。

[謝辞]

本研究の JVO プロトタイプの開発においては、富士通(株)および(株)セックの方々のご支援を頂いた。本研究は科学費補助金特定領域研究「IT の深化の基盤を築く情報学研究」(課題番号 15017289) の補助を受けておこなわれた。

[文献]

- [1] 大石雅寿: “Japanese Virtual Observatory の構築” 天文月報, Vol.95, No.12, pp.566-575 (2002)
- [2] JVO <http://jvo.nao.ac.jp/>
- [3] International Virtual Observatory Alliance <http://www.ivoa.net/>
- [4] IVOA Identifiers <http://www.ivoa.net/Documents/PR/Identifiers/Identifiers.html>
- [5] Unified Content Descriptors <http://vizier.u-strasbg.fr/doc/UCD.htm>
- [6] VOTable <http://vizier.u-strasbg.fr/doc/VOTable/>
- [7] FITS Support Office <http://fits.gsfc.nasa.gov/>
- [8] Hierarchical Triangular Mesh <http://www.sdss.jhu.edu/htm/>
- [9] Globus Alliance <http://www.globus.org/>
- [10] Resource Metadata for the Virtual Observatory <http://www.ivoa.net/Documents/PR/ResMetadata/ResMetadata.html>

田中 昌宏 Masahiro TANAKA

国立天文台天文学データ解析計算センター研究員。1997 年名古屋大学大学院理学研究科博士課程修了, 博士(理学)。日本天文学会, 日本データベース学会正会員。

白崎 裕治 Yuji SHIRASAKI

国立天文台天文学データ解析計算センター上級研究員。1997 年東京工業大学大学院理工学研究科博士課程修了, 博士(理学)。日本物理学会, 日本天文学会正会員。

本田 敏志 Satoshi HONDA

国立天文台光赤外研究部研究員。2002 年総合研究大学院大学数物科学研究科博士課程修了, 博士(理学)。日本天文学会正会員。

大石 雅寿 Masatoshi OHISHI

国立天文台天文学データ解析計算センター助教授。1985 年東京大学大学院理学系研究科博士課程修了, 理学博士。日本天文学会, 日本データベース学会正会員。

水本 好彦 Yoshihiko MIZUMOTO

国立天文台光赤外研究部教授。1979 年東京工業大学大学院理工学研究科博士課程修了, 理学博士。日本物理学会, 日本天文学会, 情報処理学会正会員。

安田 直樹 Naoki YASUDA

東京大学宇宙線研究所助教授。1995 年東京大学大学院理学系研究科博士課程修了, 理学博士。日本天文学会正会員。

増永 良文 Yoshifumi MASUNAGA

お茶の水女子大学理学部情報科学科教授。1970 年東北大学大学院工学研究科博士課程修了, 工学博士。情報処理学会, 電子情報通信学会フェロー。日本データベース学会会長。