

Japanese Virtual Observatory (JVO)の研究開発

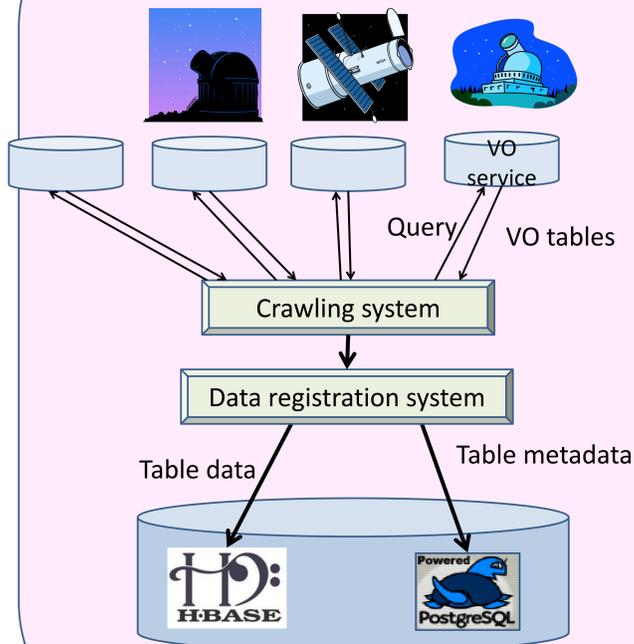
～ VO Crawler ～



ヴァーチャル天文台(VO)は、世界各地にある多くの天文データのシームレスな利用を可能にするものである。JVOプロジェクトでは、ポータルサイト(<http://jvo.nao.ac.jp/portal>)を通じて国内外のデータを検索・解析するシステムを開発し、運用を行っている。
今年度、JVOの新たな機能の一つとして、VOデータのクローリングシステム、VO Crawlerと、これにより取得したデータの統計情報収集機能、クローリングデータを活用した可視化・検索システムの開発を行った。

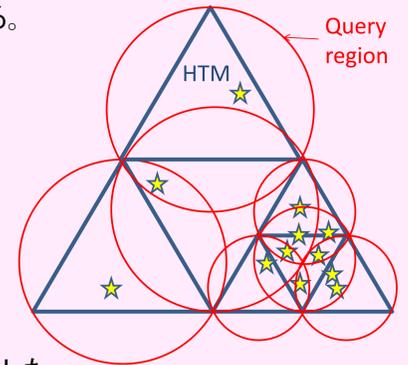
○小宮悠, 白崎裕治, 江口智士, 大石雅寿, 水本好彦 (国立天文台), 石原康秀, 大西隆史, 堤純平, 檜山貴博 (富士通), 中本啓之, 坂本道人(セック)

VO Crawler



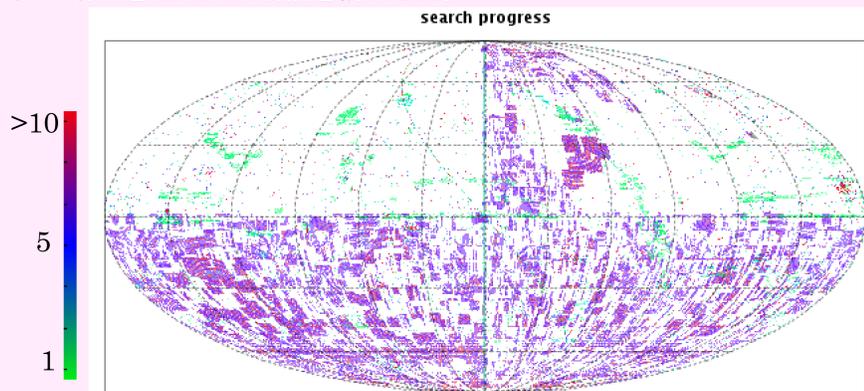
世界各国のVOサービスにある全天のデータを検索し、自動収集する、VO Crawlerの開発を行った。収集したデータを単一のシステム内に持つことにより、多波長・大量データの高速な検索が可能になる。
VO Crawlerは、全てVO対応サービスのカタログデータを収集し、画像・スペクトルデータについてはメタデータと画像・スペクトルへのアクセスURIを収集する。分散処理システム Hadoop を用いており、収集データはHadoop用分散データベースHBaseに格納される。メタデータはPostgreSQLに登録される。

VO Crawlerは、radial searchを繰り返してデータを収集する。各HTM indexごとに、その領域をカバーする検索半径でのradial searchを行い、全天を覆うまで繰り返す。データ数密度の高い領域では、自動的に細かいHTM levelに分けて検索が走る。こうした検索を、全てのVOサービスに対して行っていく。検索は最大10並列で走る。



2月時点で、4722サービスから～5.8億件のデータを収集した。(VizieR以外の全サービスに対しては、1ヶ月程度で収集は完了する。同一サイトには並列検索をかけないようにしているため、カタログ数の多いVizieRの検索には数カ月かかり、完了していない。)現在は、システムのアップデートに伴い、データの再収集中である。

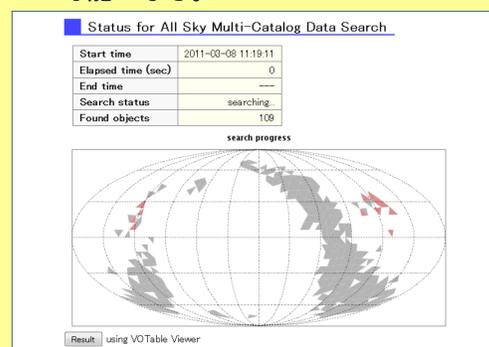
VO Crawlerの収集したデータについて、領域(HTM index)毎に、また波長および観測装置毎に、何件のデータがあるか、統計情報を収集する機能を構築した。



HTM index level 9 の各領域毎の、VO Crawlerが取得した画像データ数の天球上分布(現時点までに取得されたデータについて)。

全天検索

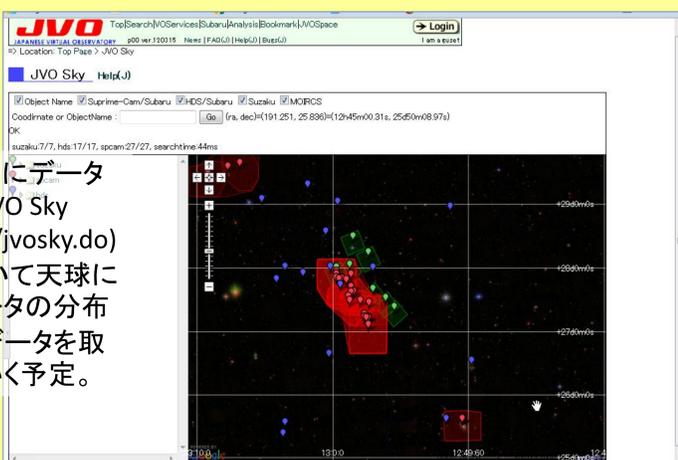
VO Crawlerの収集したデータを活用し、天体の色・等級を検索条件として、条件を満たす天体を全天の多波長データから検索する、全天検索機能を開発中である。様々なサービスから測光データを抜き出したカタログを作成しておき、これに対して検索を行う。これまでのVOでは専ら座標を指定しての検索であったが、天体の性質からの全天検索が可能になる。



SDSSと2MASS catalogをクロスマッチしたものからの全天検索(この例は、VO Crawlerのデータを用いたものではない)。グレー: 検索済み領域、赤: 検索にヒットした領域。

JVO Sky

昨年度、天球画像上にデータの分布を表示するJVO Sky (jvo.nao.ac.jp/portal/jvosky.do)を開発した。これを用いて天球にVO Crawler取得データの分布を表示し、そこからデータを取得できるようにしていく予定。



マイカタログ作成

VO Crawlerの収集したデータを基に、新たなカタログを作成する機能を開発中である。ユーザーがカラム名、天体種別を指定することで、クローリングデータの中から候補データを取得し、新たなカタログに統合していく。

RA	Dec	V

マイカタログの構造と、元カタログとの対応・変換関係を定義する。

VO Crawler Data Storage

Table A			
id	RA	Dec	V
1	3:24:18	43:17:56	11.2
2	5:23:47	11:24:34	10.7

Table B			
id	R.A.	Dec.	Vmag
1	52.413	24.526	10.4
2	129.423	42.315	9.32