

科学データのデータ属性の探索的検索を利用した検索

重川 美咲子 (指導教員: 渡辺 知恵美)

1 はじめに

近年地球流体データは爆発的に増加しており、一般の科学者が低コストで尚且つ簡単に検索・公開を実現できるツールが必要とされている。そこで我々は、地球流体物理科学者のためのデータアーカイブサーバ構築支援ツール: Gfdnavi の開発を進めている。

本稿では、データ属性の探索的検索を利用する新しいインタラクションモデルを採用したデータの検索方法について、そして検索結果を取り扱う機能について述べる。

2 Gfdnavi

地球流体科学者のためのデータアーカイブサーバ構築支援ツール: Gfdnavi は、地球流体物理科学者が個人で持つ膨大な科学データをローカルで検索したり、共同研究者や同分野の科学者同士でデータを共有することを簡単に実現することを目的としたツールである。また、Web サーバ上で Web サイトとして利用することも可能である。

Gfdnavi 上では科学データを検索・可視化・分析することが可能であり、我々はデータ検索部の開発を行っている。

2.1 検索インタフェースの問題点

Gfdnavi では全ての科学データが持つ属性として時間属性・空間属性、そしてそれ以外の属性値をキーワード属性として、各科学データに3種類の属性を定義している。そして検索インタフェースでは、空間領域、時間領域、キーワードのどれかを指定すると画面の下部に検索結果のリストが表示される。

しかし、この先行研究 [1] のバージョンの検索インタフェースには、

- ・公開者により異なる属性名の検索が困難
 - ・空間領域以外の条件では絞り込み検索を重ねることができない
- という改良すべき二点がある。

3 探索的検索を利用する新しい検索インタラクションモデルの実装

ユーザにとってのより簡単なデータの検索のためには、前述した二点を中心とした改良が必要である。

3.1 探索的検索

求めるものが明確にどのようなものかわからない場合、試行錯誤しながら徐々に必要とするものに近づいていく検索方法を本稿では「探索的検索」と呼ぶ。

探索的検索の代表的な手法に Faceted Navigation [4] がある。絞り込み条件の選択肢としてあらかじめ設定されている属性値を、結果を見ながら変えるだけで容易に探索的検索を行うことが可能である。ユーザは自分で問合せ文を記述する必要がなく、システムにナビゲーションされる様にして簡単に検索が行える。

新しい検索インタフェース(これ以降 Gfdexplorer と呼ぶ)は、Faceted Navigation を取り入れた図 1 に示す新しい検索インタラクションモデルを採用し、ユーザに探索的検索を促す。

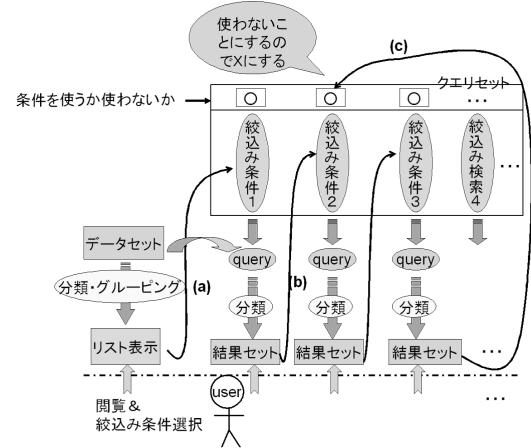


図 1: 新しい検索モデルのインタラクションイメージ

top ページを開いてから次に top ページを開くまでに行った絞り込みまでを一回の検索セットとし、その中で絞り込みに使われた条件を、一つのクエリセットとしてまとめて記録する。データを属性によって分類・グルーピングし、属性リストとして表示された中からユーザ絞り込み条件を選択すると、その条件はクエリセットに追加される(図 1 中 (a), (c)) クエリセット中の条件全てを満たすデータが結果として返され、検索が行われる度にデータは分類され、リストが更新される。リストからの選択を繰り返していくと複数の条件での絞り込みができる。

また、得られた結果を見てユーザが以前使ったある条件を取り消すということもある(図 1 中 (c))。その場合はクエリセット中の取り消す条件以外の条件を満たす検索結果を返す。

このモデルにより、ユーザはリスト中から属性値を選んでいくだけで簡単に必要とするデータを取得でき、容易な検索結果の比較も可能である。

以下、このモデルを実装した後の各属性部分と、新たに追加した機能について述べる。前提として、全ての属性について探索的検索が可能となっている。また、図 2 に Gfdexplorer を示す。

3.2 キーワード属性

絞り込み対象である各データのキーワード属性すべてをリストアップする。ユーザはリストアップされている値をクリックして選択すると絞り込み検索が行われる。リスト部分も更新され、今絞り込んだキーワード以下、更にどのようなキーワードで絞り込めるかをまたリスト表示する。

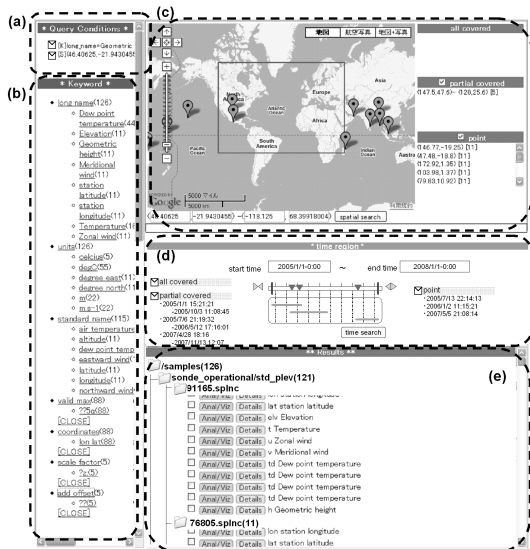


図 2: 新しい検索インターフェース: Gfdexplorer

3.3 空間属性と時間属性

空間属性による検索に関してはデータは領域を持つので表示する際に, "all covered", "partial covered", "point" の 3 種類に分けた. データの種類を選ぶとそれぞれのデータが地図上に現れる (図 2(c)) GoogleMap 上で範囲を指定し, 検索を行う.

時間属性の検索では, GoogleMap の下にあるバーで start time と end time を指定し, 時間の範囲の絞り込みを行う. また空間属性と同様に領域を持つデータなので, データを空間属性と同様の 3 種類に分けた. バーで絞り込んだ時間領域の中に属するデータを, データ別に表示する (図 2(d))

3.4 検索条件: Query Conditions

検索条件の編集を行うことができる. まず絞り込み検索を行うと, Gfdexplorer の左上にどのような条件で絞り込みを行ったかが表示される. (図 2(a))

各条件の左にあるボックスにチェックを入れるか外すかで, その条件を絞り込みの条件として使用するかどうかを変更できる. よって一度複数条件で絞り込んだ結果からある条件を除いた検索結果や, ある検索結果からいくつかの条件を他の条件に入れ替えた際の検索結果を容易に導くことができる.

3.5 検索結果の表示: Results

Gfdexplorer 下部に, 絞り込み検索を行った結果が表示される. データは意味のあるディレクトリ別に分けられていることが多いので, パスの階層構造を利用して表示する. 図 2(e) の場合は, QueryConditions 部でチェックの入っている二つの条件で絞り込みを行った結果が表示されている. この場合は, /samples/sonde.operational/std.plev の更に下にある 91165.splnc ディレクトリと 76805.splnc ディレクトリ内のデータが結果として表示されており, 図 2(e) のようにデータ同士の階層構造における位置がわかる.

3.6 クエリヒストリとキャッシュ機能

我々はデータベースの中に, query_histories というテーブルを用意した. その中身を図 3 に示す. 但し図では見易さの為に表として表示する.

id	querytype	user_id	time	description	queryset_id	conditions	count	cached
1	keyword	1	2007-12-30 18:16:46	[K]long_name=Dew point temperature	1	name = 'long_name' and value = 'Dewpoint temperature'	15	1
2	space	2	2007-12-31 11:17:32	[S]88.59375, -33.72433966174759, -258.75, 47.517200697839414	2	longitude_lb <= 88.59375 and longitude_rb <= -33.72433966174759 and latitude_lb <= 47.517200697839414 and latitude_rb >= -33.72433966174759 and not(node_id IS NULL)	63	0
3	keyword	1	2007-12-31 16:27:52	[K]long_name=station latitude	3	name = 'long_name' and value = 'station latitude'	15	1
4	keyword	1	2007-12-31 16:28:47	[K]standard_name=lattitude	3	name = 'standard_name' and value = 'lattitude'	15	1
5	keyword	1	2007-12-31 16:29:06	[K]units=degree north	3	name = 'units' and value = 'degree north'	15	1

図 3: query_histories の中身

また, クエリセットの id のみをキャッシュに保存することとした. 例えば図 2(a) で二つの条件にチェックを入れ, 更に絞り込みを重ねる場合を考える. この検索のクエリセット id は 3 とすると, キャッシュテーブルに保存されているクエリセット id が 3 のもの (図 3 中の id が 3 と 4 の絞り込み) との AND を取り, それを検索の結果とする. 新しく加えられた絞り込み条件は query_histories に追加される (図 3 中 id5)

しかし現在キャッシュ機能を使うにあたって, キャッシュデータがいたずらに増えてしまう可能性があるという問題点がある. この問題に対して二つの解決策を考案中である.

また, ディレクトリ構造を考慮したキャッシュの圧縮方法についても考案中である.

4 まとめと今後の課題

本稿では地球流体物理科学者のためのデータアーカイブサーバ構築支援ツール: Gfdnavi のシステム構成のうち, データ検索部におけるデータ属性の探索的検索を利用した, より簡単に低コストでユーザがデータの検索を行うことのできるインターフェースの実装について述べた.

今後, 更に考察・実装を進めていく予定である.

参考文献

- [1] 柳平有美, 渡辺知恵美, 堀之内武: "地球流体物理科学者のためのデータアーカイブサーバ構築支援ツール: Gfdnavi におけるデータベース設計と検索インターフェースの実装," DBSJ Letters vol.6, No.1, JUNE2007, pp.153-156
- [2] 堀之内武, 西澤誠也, 渡辺知恵美, 森川靖大, 神代剛, 石渡正樹, 林祥介, 塩谷雅人: "地球流体データベース・解析・可視化のための新しいサーバ兼デスクトップツール Gfdnavi の開発," 第 18 回データ工学ワークショップ (DEWS2007), D2-8(2007)
- [3] 地球流体電脳倶楽部: "http://www.gfd-dennou.org/,"
- [4] M.Hearst Clustering versus Faceted Categories for Information Exploration: Communications of ACM, Vol.49, No.4, pp.59-61, 2006.