

地球流体データ解析・可視化ツールGfdnaviにおける 知見情報のデータベース化

伴林 晃紀[†] 堀之内 武[†] 津田 敏隆[†] 渡辺知恵美^{††} 西澤 誠也^{†††}

[†] 京都大学生存圏研究所 〒 611-0011 京都府宇治市五ヶ庄

^{††} お茶の水女子大学理学部情報科学科 〒 112-8610 東京都文京区大塚 2-1-1

^{†††} 京都大学大学院理学研究科 〒 606-8502 京都市左京区北白川追分町

E-mail: †{tomoba,horinout,tsuda}@rish.kyoto-u.ac.jp, ††chiemi@is.ocha.ac.jp, †††seiya@kagi.kyoto-u.ac.jp

あらまし 近年，地球流体の数値データは急速に増加している．膨大な数値データは，解析・可視化して初めて人間にとって意味を持つ．そこで，著者らを含むグループにより地球流体データ解析・可視化ツール Gfdnavi の開発が進められている．本研究では，Gfdnavi 内で，数値データから得られた知見を文章化して図とともに保存できるように開発を行った．知見文書は数値データ及び作図に使われたパラメータとともにデータベースにアーカイブされ横断的に利用できる．数値データと知見情報とが相互に関連を持つことにより，数値データから得られた知見を検索したり，解析を再現したり，検証したりできる．また，本システムは，知見情報がメタデータとして自動的に蓄積されていくものと言える．

キーワード 科学 DB，知識処理，データ可視化

Development of archiving knowledge in the database for Gfdnavi: analysis and visualization tool for geophysical fluid data.

Akinori TOMOBAYASHI[†], Takeshi HORINOUCHE[†], Toshitaka TSUDA[†],
Chiemi WATANABE^{††}, and Seiya NISHIZAWA^{†††}

[†] Research Institute for Sustainable Humanosphere, Kyoto University

^{††} Department of Information Sciences, Faculty of Science, Ochanomizu University

^{†††} Research Institute for Mathematical Sciences, Kyoto University

E-mail: †{tomoba,horinout,tsuda}@rish.kyoto-u.ac.jp, ††chiemi@is.ocha.ac.jp, †††seiya@kagi.kyoto-u.ac.jp

Abstract Recently, numerical data on geophysical fluids have been increasing explosively. Large-scale numerical data can have meaning only after visualizing and/or analyzing them. In this study, we extended Gfdnavi to support documentation and archival of knowledge and images obtained through data analysis. A Gfdnavi knowledge document consists of figures along with the parameters used to create them, and it is saved in the database holding the metadata of the numerical data. It enables the users to search knowledge obtained from particular data, to reproduce the analysis made, and to further verify it. Also, such a system can be regarded as an automatic collector of metadata produced by users, in many cases experts.

Key words Scientific DB, Knowledge Processing, Data Visualization

1. はじめに

1.1 背景

近年，人工衛星などによる地球観測は種類・観測数・解像度

のいずれも増大しつつあり，大気や海洋などの「地球流体」の数値データは急速に増加している．また，気候や水循環などの予測や診断における数値シミュレーションの重要性も増しており，大規模な計算により地球流体に関する大量のデータが生み

出されている．現在は，多様なデータを組み合わせた総合的な研究が必要になっている．

インターネットにおいて膨大な地球流体の観測データが提供されており，これらは多次元の大きな数値データである．そのような多次元のデータをそのままの状態人間が理解することは難しく，可視化（例えば一部のパラメータを取り出して2次元の図に変換する）を行うことによって初めて人間にとって意味を持った情報になる．そこで，

- サーバ側でデータの解析・可視化を行ってくれる
- データベースとしての検索機能を持つ
- 複数サーバ間に渡って横断的な利用ができる

という条件を満たすデータ提供サーバが求められる．従来，個別の Web でのデータ提供サーバ構築ツールはすでに存在していた．しかしそれらは独自のものが多く，良いデータ提供サーバを作っても応用性や移植性に乏しかった．

そこで，サーバ構築ツール Live Access Server (LAS) [3] が登場した．LAS は現在最も使われている地球流体データ提供サーバ構築ツールである．また，Web ベースのツールであり Web ブラウザを用いてアクセスするため専用クライアントが不要である．他に，可視化機能を持っていて，カスタマイズ性に優れるという特徴がある．一方で，公開サーバ向けのツールなので個人での運用には向いていない．また検索機能も弱い．

そこで，著者らを含むグループでは地球流体データ提供サーバ構築ツール Gfdnavi (Geophysical Fluid Data Navigator) [1] が開発されている．Gfdnavi には次のような特徴がある．

- サーバによる解析・可視化機能
Gfdnavi ではサーバ上の数値データを解析・可視化して図を作成することができる．
- データベースとしての検索機能
作成した図はデータベースに保存することができ，容易に検索できる．さらに，Google Map を用いたインタフェースによる検索も可能である．
- 横断的利用の機能 (実装予定)
個々のサーバでの検索だけではなく，横断的に他の公開サーバ内のデータも検索対象にすることができる．現在，ピア・ツー・ピア (P2P) を利用したネットワーク横断検索・共有システムの開発が進められている [4] ．

本研究では Gfdnavi 内で作成された図に対し，そこから得られた知見についての文章を作成し，図とともにドキュメントとして保存できるように開発を行った．なお，Gfdnavi 内で作成された図以外の図に対しても知見についての文章はつけることができるが一部の機能については利用することができない．

LAS や従来の Gfdnavi 他，地球物理関連分野でよく知られたツールの中にはこのような知見情報を作成する機能はなく，これは新しい機能だといえる．

1.2 目 的

この節では知見情報のデータベース化の開発の目的について述べる．

本論文では，文章化された知見と図を合わせたものを知見情報と呼んでいる．また，そのようなドキュメントを知見文書と呼ぶことにする．知見文書は，例えばホームページで気象イベントを紹介するための PR 文であったり，研究者が残す簡単なメモであったり，また論文といえる規模の文書であったりする．

本研究では，知見情報を Gfdnavi 内で作成しデータベースに数値データや図，解析プロセスとともに知見情報を格納できるよう開発を行った．これにより数値データと知見情報とを関連付けて相互に検索を行えるようにすることが本研究の目的である．

相互の検索が行えることで可能になることについて挙げていく．

(1) 数値データを元に図や知見情報を検索することができるようになる．例えば，研究者が降雨データを解析して集中豪雨の分布を見出して図を描き，その説明となる文章を書いて知見文書として保存したとする．その降雨データを研究したい他の研究者は当然その図や文書の存在や置き場所を知らないが，検索によってそれらをデータを元に見つけ出すことができる．また，解析を行った本人にとっても，結果を保存した場所をいちいち憶えておく必要が無くなる．

(2) また逆に図や知見情報からその元となった数値データや解析のプロセスを検索することもできる．例えば，台風について書かれた知見文書に対して，関連した降雨データと解析のプロセスを見つけることができるので，それを利用して解析を再現したり，また同じ解析手法を別の地点の降雨データに適用するとどうなるか検証したりすることができる．

(3) さらに，検索の手がかりとして数値データ・知見情報のどちらも手元に無い状態から言葉を用いて検索することもできる．例えば「集中豪雨」というキーワードに対して，そのキーワードをメタデータとして付加された降雨データや，集中豪雨について書かれた知見文書を検索によって見つけ出すことができる．

より詳しい内容については，後に 3.4 節で具体例を交えて述べる．

知見情報を作成する機能や，数値データと相互に検索を行える機能は以下のように様々な場面で役に立つと考えられる．

- 作った知見文書の中から発表したいものをホームページでまとめて公開するなど，情報の公開に役立てることができる．
- 複数の検索方法があるため，膨大な量の研究データの中から欲しいデータを見つけ出しやすい．
- データに関連した図や知見文書を検索結果として得ることで，グループでの研究において共有している観測データから他の共同研究者が得た知見を参考にすることができる．
- キーワードを元に検索するだけでそれに関連した様々な知見文書を読むことができる他，知見情報からそれを得た解析プロセスを知ることによってその解析手法について学ぶことができ，その手法を利用してさらなる解析を行うこともで

きる。

本論文では第2章で Gfdnavi の構成について紹介した後、第3章で知見情報のデータベース化についてより具体的に述べる。第4章ではこのシステムを別の角度から見てみる。最後に第5章でまとめとする。

2. Gfdnavi の構成

本章では知見情報のデータベース化開発以前の Gfdnavi の構成について述べる。詳細は、参考文献 [1] を参照のこと。

2.1 システム概要

Gfdnavi のシステム構成について解説する。図1にそのシステム構成を示す。

Gfdnavi では、ディスク上の科学データからメタデータを自動抽出し、関係データベースとしてメタデータのデータベースを作成する。ユーザ管理用等、メタデータ以外のデータも関係データベースに格納する。ユーザインタフェース (UI) は Web ベースであり、Web ブラウザで利用するため専用クライアントツールは必要としない。この UI は Ruby 埋め込みの html と JavaScript で実現される。システムはいわゆる Model-View-Controller (MVC) 構成であり、Ruby on Rails [5] という開発フレームワークを用いて開発されており、関係データベースに対応する “Model” となるクラス群を持つ。

Gfdnavi には Web サーバプログラムが同梱されており、図1の全体を利用者の PC1 台で運用することができる。また同梱 Web サーバプログラムだけでなく、Apache 等の汎用 Web サーバも用いることができる。複数サーバにまたがる利用についても、P2P を用いた実装の開発が行われている [4]。

Gfdnavi のデータベースは関係データベースマネジメントシステム (RDBMS) によって実装されている。先に述べた通り Gfdnavi では扱う科学データからメタデータを自動抽出することで、科学者が容易にデータベースを作成できるようになっている。

Gfdnavi には管理のためにユーザ及びグループの概念が導入

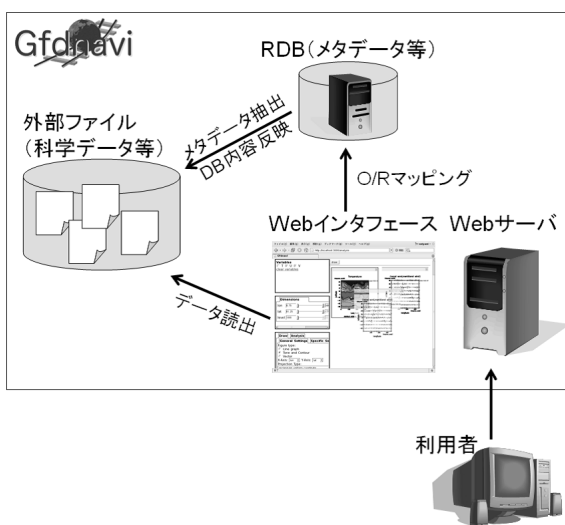


図1 システム構成概要

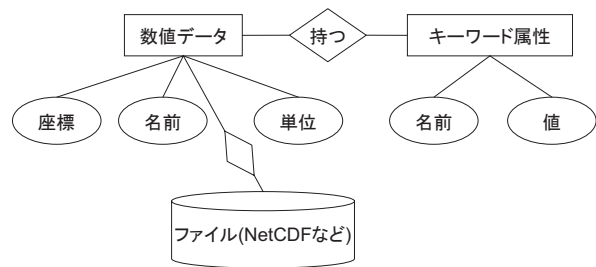


図2 Gfdnavi における数値データの構成

されており、これらもデータベースで管理されている。また、アクセス制御が可能である。

Gfdnavi における数値データは GPhys [6] オブジェクトに代表される実体である。GPhys は地球流体の多次元数値データを扱うための Ruby のクラスライブラリである。本ライブラリでは、数値データを「離散化された連続空間の物理量」として抽象化し、座標に関するメタデータと、単位、データ欠損、物理量の名前等に関するメタデータを持つ。他に、任意の名前と値の組で表されるメタデータを持てる。数値データの構成を図2に示す。GPhys のデータモデルは、地球流体分野でよく使われる NetCDF [7] というファイル形式 / 入出力ライブラリにおけるそれを抽象化したものである。GPhys を用いることで、多様な数値データを統一的に扱うことができる。

図3に、知見情報のデータベース化開発以前の Gfdnavi の Web ベース UI の構成概略を示す。矢印はコンポーネント間の遷移・相互作用を示す。検索・選択 UI には、ディレクトリ階層のツリー表示による選択 UI と、メタデータ検索 UI がある。

2.2 Gfdnavi で利用できる機能

図4に前節で述べたツリー表示 UI のスクリーンショットを示す。同 UI は、Microsoft Explorer 風の階層表示を行い、解析・可視化 UI やメタデータ詳細表示 UI と結びついている。

前節で述べた GPhys ライブラリ群を使えば、出版に相応しい高いクオリティの可視化ができ、また自在な数値処理による解析ができる。そこで、Gfdnavi では Web ベースで GPhys の全機能を提供できるようにすることで、図5のような解析と可視化を実現している。可視化によって作成された画像はデータベースに保存され、ディレクトリ構造を持つ。

解析・可視化 UI では以上のほかに、次の機能が実現されている。

解析:

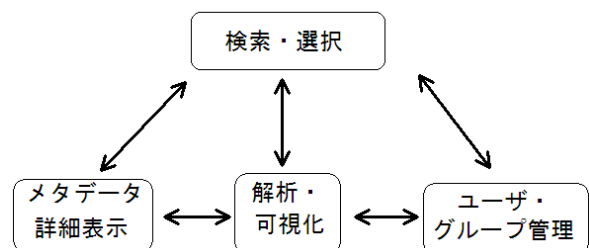


図3 UI の構成

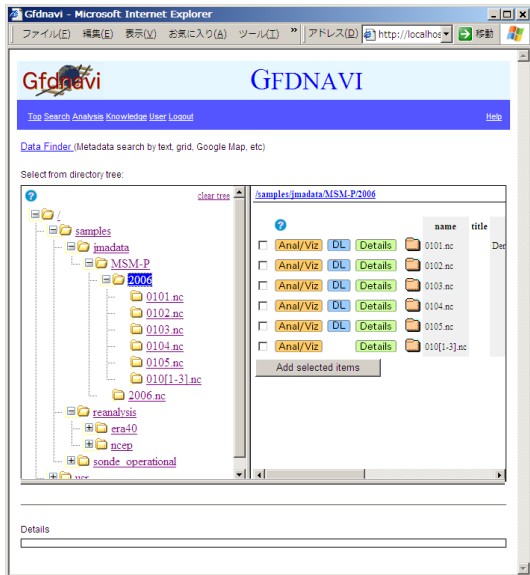


図 4 ツリー表示

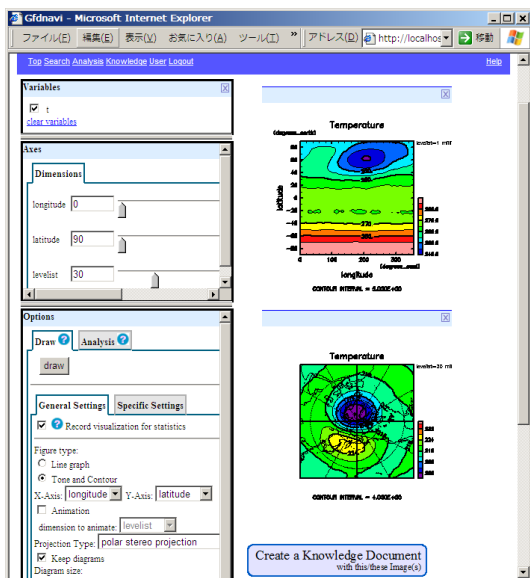


図 5 解析・可視化

- 数学・統計計算
- ユーザ定義関数の登録（要ログイン）
- 生成されたデータのサーバでの保存（要ログイン）またはダウンロード（同不要）

可視化：

- 各種可視化パラメータ設定
- 新規描画画像の上書きまたは追加（並べていく）の切り替え
- 画像の合成
- 描画した画像のサーバでの保存（要ログイン）またはダウンロード（同不要）

- 画像を再現する Ruby スクリプトと、描画に必要なサブセットのみを切り出したデータの組み合わせのダウンロード
- 画像を Gfdnavi 上で再描画するための各種パラメータを持った URL を取得する機能（この URL をブラウザのアドレスバーに入力すれば、同じ画像を描くことができる）

Gfdnavi は、解析・可視化で生成されたデータや画像を保存する際に、作成に使ったデータへの参照を自動的に記録する。これにより相互にリンクが貼られ、お互いを辿ることができるようになる。

3. 知見情報のデータベース化

第 2. 章で述べた Gfdnavi の構成に加えて知見情報を保存することができるように、開発を行った。本章では、これにより実現される機能について述べる。

3.1 知見情報の提案構成

この節では提案する知見情報の構成について述べる。

知見情報の構成は、図 6 の通りである。1 つの知見情報は表題を 1 つ持っていて、要約と本文を 1 つまで持つことができる。図は複数持つことができる。また、知見情報とその元となった数値データとの関連が存在し、これらは多対多で対応する。

現時点では以下に示す機能が未実装である。

- 知見情報に要約を持たせる
- 知見情報と数値データが互いに関連を持つ（現在は図を通してのみ関連を持つ）

また、図に関して現在の実装における詳細を図 7 に示す。図は 1 つの画像と 1 つのキャプションを持ち、図は知見情報の中に複数含まれる。画像と図との従属関係は弱く、図を消しただけでは画像は消えないが、図と知見情報を同時に作成・保存した場合にはそれらが強い従属関係を持つようになり、図とともに画像も消えるようになる。

画像は画像の名前、画像のディレクトリ内のパス、ディスク上に保存される画像ファイル、画像を作った際のスクリプト、数値データへのリンクを持つ。

3.2 データベース

Gfdnavi のデータベースは多数のテーブルから成っているが、その中から知見情報に関わっている知見情報、図、画像のテーブルについて説明する。テーブルは、前節の図 6、図 7 で示した構成に必要な属性に加え、実装する上で必要な属性を含んでいる。

- 知見情報のテーブル

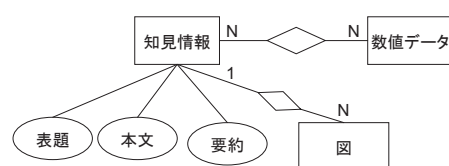


図 6 知見情報の構成

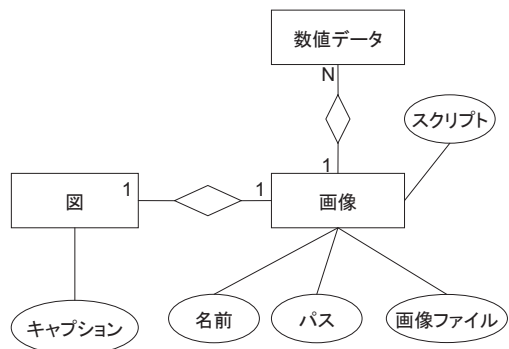


図 7 図に関する実装の詳細

知見文書を表すテーブルで、レコードが1つの知見文書に相当する。知見文書の属性として表題と本文を持つ。その他、知見文書が置かれるパスと所属するグループ、作成者のユーザ ID も持っている。

- 図のテーブル

図の情報を持つテーブルである。1つのレコードが1つの図に対応する。知見文書の ID と画像の ID を持ち、知見文書と画像を関連付ける役割を担っている。また、個々の図に含まれるキャプションを持つ。知見情報テーブルに付属している。

- 画像のテーブル

画像の ID、画像の名前、画像の置かれているディレクトリを指し示す ID、画像のパス、画像を作る元となった数値データの ID、画像の描画パラメータ(実装では実行可能な形式でスクリプトとして保存されている)を持っている。

3.3 UI の機能

知見情報に関する UI 関連の機能について解説する。

- 知見文書のリスト表示

存在する知見文書のリストを表示する。リスト表示の際には知見文書の表題、作成者、所属グループ、パスの他、本文の冒頭部分が表示される。

- 知見文書の閲覧

知見文書を閲覧する機能で、表題、本文、パス、図を見ることができる。表示は図を横に並べて置いたり、縦に並べて置いたりといったレイアウトが複数用意されており(図8に一例として、図を横に並べたレイアウトを示した)、その中から用途に合わせて選ぶことができる。

- 知見文書の作成

知見文書の作成の際にはフォームに必要な情報(タイトル、本文、知見文書を設置するパス、デフォルトのレイアウト、所属グループ、知見文書に含まれる図のパスとキャプション)を入力して投稿する(図9参照)。知見文書の中には図を複数持たせることができ、その順番も自由にえられる。また、知見文書に使用する図は既存のものを用いることができるほか、新たな図を作ってそれを用いることもで

きる。新たな図を用いる場合は知見文書の作成と同時に図の保存も行う。

- 知見文書の編集

新規作成と同様のフォームを使用、上書き保存と別名保存ができる

- 知見情報から数値データや解析プロセスを取得する
知見文書内に含まれる図に対し、その元になった数値データと解析プロセスをデータベースから引き出すことができる。それらを用いて知見文書内の図を再現することができる。同じ図を再現する URL を取得することもできる。

- データ・図に関連した知見情報の一覧の取得

3.4 データベース化のメリット

1.2 で述べたように Gfdnavi のデータベースには数値データ、解析・可視化により作成された図、図の解析プロセス(作図に使ったメソッドとパラメータ)、知見情報が保存されていて、データベースに一体的に保存されたこれらのデータは横断的に扱うことができる。このことにより生まれるメリットについて具体的な例を挙げながら解説する。

(1) 数値データと知見情報が一体的に保存されていることから、ある数値データについてそのデータを可視化した図や、図を用いて作られた知見文書にどのようなものがあるかを検索できるようになる。これにより、扱っているデータが膨大な量でどこに図や知見文書を保存してあるかが分からないときにも欲しい情報を見つけることが容易になる。

また Gfdnavi を共同研究に利用すると、データベースに保存された図や知見文書を参照するという形で、データから見出した知見情報を研究者間で共有することができるようになる。専門分野の研究者の間でもデータに対して着目する点は各々異なっているため、他の研究者の解析結果を見ることは有用である。異分野の研究者にとっても、自分が普段扱わないデータに



図 8 レイアウトの例

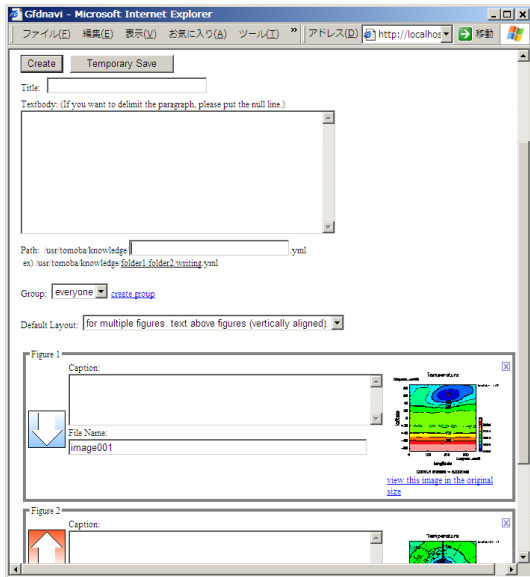


図 9 入力フォーム

対して専門の研究者が見出した知見を文章化したものを見て数値データの持つ意味（何を表すデータなのか、観測結果の妥当性、など）を知ることができる。

実際に数値データから知見情報を得るまでの流れを図 10 の実線部に示した。例えば、各時刻の雨量が格子点化されたデータがあるとする。このデータが何を表しているのかが分からないとき、それを元に作られた図と知見文書の一覧を取得し、その中から例えば台風について書かれた知見文書が見つければ、このデータは台風を表しているということが分かる。

(2) また、逆に知見情報に対しその知見を見出す元となった数値データやそこから作られた図および解析のプロセスを知ることができる。これにより、図が作られる元となった数値データと解析のプロセスを用いて調べたい図や知見情報が作られる元になった解析を再現することができる。解析の再現ができれば、見づらい部分を拡大してより細かく調べるなど、必要に応じて可視化のパラメータを変えて見るようになる。また再現するだけでなく、同様の解析手法を別のデータに適用することもできる。

知見情報から数値データ、解析プロセスを得て解析を再現したり検証したりする流れについて図 10 の点線部に示した。例

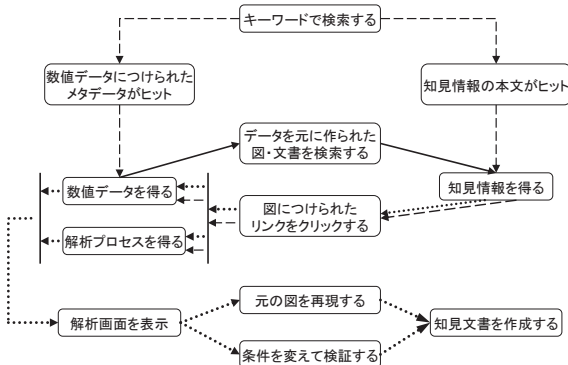


図 10 知見情報に関する流れ：実線 - (1) 破線 - (2) 点線 - (3)

えば、台風に関して書かれた知見文書があるときその中に含まれる図につけられたリンクをクリックすることで、画像を作る元になった数値データと描画パラメータ（例えば台風の発生した時期、場所など）が得られる。それらを用いて図を再現したり、同様の解析プロセスを別の場所のデータに対して適用したり、といったことができる。そして、このようにして得られた知見を知見文書として保存することもできる。

(3) これまで挙げた例は特定の数値データおよび知見情報から検索を始める話だったが、それらの手がかりが無いところから言葉（キーワード）で検索することもできる。データに付与されたメタデータは全て検索対象となるため、調べたいキーワードに関連したデータを検索によって見つけ出すことができる。同様に知見情報に含まれる単語も検索対象となるため、知見情報を經由して、関連した数値データを見つけ出すこともできる。

このことは図 10 の中では破線部にあたる。例えば、「台風」というキーワードで検索してメタデータがついていないために数値データが見つけれなかった場合でも、台風について書かれた知見情報が検索結果として得られればその知見情報の元となった数値データを見つけ出すことができる。

また一見関わりの無さそうなデータが検索結果として得られることがあり、それによって新たな知識を得ることがある。例えばキーワード「マイクロ波放射計」を含む知見情報が台風のデータにヒットした場合、その結果からマイクロ波放射計が台風の研究に使えるということが分かる。

4. 議 論

第 3 章で述べたように、知見情報は検索の対象となるため、言葉での検索を行った際に、検索語を含む知見情報を見つけて出すことができる。それに伴い、知見情報に関連した数値データも検索できる。つまり、もしも知見情報が無かったとしたら検索結果に現れなかったであろう数値データが検索で得られることになる。このことから、数値データに対して知見情報を作成することはデータに新たなメタデータを付与することだと捉えることができる。言い換えると Gfdnavi における知見情報のデータベースは、研究活動でドキュメントを書くとき自動的にメタデータが増えて検索性が増すシステムになっているといえる。他の公開サーバ内のデータの横断的利用と併せて Gfdnavi が普及すればするほどデータベースとして有用なものになっていく。

5. ま と め

本論文では、地球流体データ解析・可視化ツール Gfdnavi において知見情報を作成しデータベースに保存する機能について、その構成と実装を示した。

データベースには知見情報とともにその元となった数値データが格納されており、これらは横断的に扱うことができる。これにより、知見情報と数値データとの間で相互の検索が可能となり、様々なことができるようになる。さらに、言葉による検索も可能となる。

数値データに対して知見情報をつけることは、数値データに

対するメタデータを付与することになる。つまり、Gfdnavi を研究活動に使っていくと自動的にメタデータが増大してデータベースの検索性が増すことになる。

今後は将来的に実装したい構成や機能について開発を行っていく。

謝辞

本研究は、文部科学省科研費特定領域「情報爆発時代に向けた新しい IT 基盤技術の研究」の課題 (課題番号 19024039) により行われた。

文 献

- [1] 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”, Proceedings of Data Engineering Workshop(DEWS-2007), D2-8, pp8.
- [2] 柳平有美, 堀之内武, 渡辺知恵美, 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フェースの実装. 日本データベース学会 Letters Vol. 6, No. 1, pp. 153-156 (2007)
- [3] Hankin, S., J. Callahan, J. Sirott, “The Live Access Server and DODS: Web visualization and data fusion for distributed holdings”.
<http://ferret.wrc.noaa.gov/Ferret/LAS/LASovertime.html>, 2001
- [4] 齋藤真衣, 堀之内武, 渡辺知恵美, “対話的な絞り込み操作を考慮した P2P による地球流体データアーカイブサーバの横断検索”, Proceedings of Data Engineering Workshop(DEWS-2008)
<http://www.rubyonrails.org/>
- [5] Horinouchi T., R. Mizuta, S. Nishizawa, D. Tsukahara, and S. Takehiro, “GPhys – a multi-purpose class to handle gridded physical quantities”, <http://ruby.gfd-dennou.org/products/gphys/>, 2003.
- [6] Rew, R., and G. Davis, “NetCDF - An interface for scientific-data access,” IEEE Computer Graphics and Applications, 10(4), pp76-82, 1990.