“Software Repository” におけるデータの取り扱い

— Matrix Market, 行列工房 を例に —

高久 雅生

1999 年 3 月 1 日

1 はじめに

科学技術計算の分野で使われるソフトウェア (Software) を集め、共有するサービスが行なわれている。このようなサービスはソフトウェアレポジトリ (Software Repository) と呼ばれている。ソフトウェアは

- プログラム (Programs)
- データ (Data)
- ドキュメント (Documents)

からなるものと考える。

ここでは特に Matrix Market や行列工房などの科学計算用データのレポジトリを例にとり、データの対象、利用方法、利用形態などについての概説を行なう。

2 Matrix Market

Matrix Market[1] は線形代数の数値計算アルゴリズムの比較・検証に利用されるテスト用行列データのレポジトリを提供している。このレポジトリでは行列と行列生成ツールの両方が提供されている。

2.1 概要 [2]

Matrix Market では数値計算ソフトウェアや統計ソフトウェアの研究者、開発者、ユーザがテストデータを容易に利用できるように、行列アルゴリズム用のテストデータの収集、保存を行なっている。

科学計算の領域では新しいアルゴリズムやソフトウェアが開発される際の線形方程式の分解と解法、固有値分解は重要な問題である。新しいアルゴリズムに対する信頼できる、再利用可能な定量的な評価を行なうためには、その手法を比較するための典型問題のための共通コレクションの保持が有効である。疎行列では Harwell-Boeing Matrix Collection がこの目的のために提供されている。この行列コレクションは様々な適用分野の 9 ~ 35558 のオーダーのおよそ 300 個の行列を含んでいる。このようなコレクションでの問題点の一つは、行列のサイズと多様性のために管理・利用が難しいという点である。

一方、 WWW(World Wide Web) をはじめとする通信ネットワークの急速な発展により、このようなテスト用行列へのアクセスと使いやすさを向上させる新たな可能性が開けつつある。

Matrix Market は Harwell-Boeing コレクションに基づいた様々なサービスの提供を行なう WWW 上の行列レポジトリである。 1999 年 2 月現在、全部で 488 の行列が提供されている。

Matrix Market では行列はセット (set) にまとめられており、一つのセット中の行列は適用分野や配布元に関連のあるものである。さらにセットは一つのグループとして管理されるコレクション (collection) にまとめられている。

個々の行列は疎行列・密行列として系統的に、または生成コードを介して入手され、蓄積される。行列成分には、実数、複素数、対称、非対称、 Hermitian のように幅広い様々な種類のものがある。行列データは一ファイルに一行列という形で蓄積し、非ゼロのパターンのみで表現してファイルサイズを減らしたり、反復解法の解ベクトル、初期ベクトルなどの補助データを付加したりする。

各行列に対して、 HTML ページを提供し、行列の特性の概説やその構造の図形表現の表示を行なっている。同様に、セットについても WWW ページを作成し、配布元や適用分野などの背景や参照文献や各行列の非ゼロパターンのアイコンなどを提示している。さらにこれらの WWW ページの情報全てを含むデータベースを運営し、行列 WWW ページ・セット WWW ページを自動的に生成したり、キーワードによる全文の検索といったデータの様々な操作を行なえる。

2.2 データ形式

Matrix Market で提供されている行列のデータファイルの形式には次の 3 種類がある。

1. Matrix Market Exchange Formats[3]（資料 1 参照）

   Matrix Market で利用するために開発されたデータ形式で、行列データのやりとりするためのテキスト形式のファイルフォーマットである。表現が簡単にでき、解析しやすいように設計されている。

   Matrix Market Exchange Formats には次の 2 種類の表現法がある。

   (a) Coordinate Format：疎行列の表現に用いる。

   (b) Array Format：密行列の表現に用いる。

2. Harwell-Boeing Exchange Format（資料 2 参照）

   疎行列の表現用の形式で固定長フィールドを持つテキスト形式のファイルフォーマットである。特に、 Fortran での扱いに適した形式になっている。

3. Coordinate Text File Format（廃止予定）

さらに、これらのデータファイルは gzip 形式で圧縮されて提供される。

2.3 Matrix Homepage

Matrix Market では、各行列データと各行列セットは HTML 形式での "Homepage"（ホームページ）を持つ。

2.3.1 行列データのホームページ

行列ホームページの一例を資料 3 に示す。
ここでは以下の情報が提供されている。

1. 行列の ID
2. 適用分野
3. 行列セット名

4. コレクション名

5. データファイルへのリンク

6. 視覚的表現（Structure Plot, City Plot, Spectral Plot, VRML City Plot（資料 4 参照））

7. 行列の特性・統計情報（サイズ、種類、非ゼロ部の構成、列データ、行データなど）

8. セット情報

2.3.2 行列セットのホームページ

行列セットのホームページの一例を資料 5 に示す。

1. 行列セットの ID

2. 適用分野

3. コレクション名

4. 配布元

5. 追加日付

6. セット中の各行列の ID、タイトル、サイズ、“thumbnail sketch”(GIF 形式のアイコン)

7. セット中の全ての行列を含むファイルへのリンク

8. 参照文献

の情報が提供される。行列の ID やサムネイルをクリックすると、その行列のホームページが表示できる。

2.4 利用者へのインタフェース

Matrix Market のトップページ（`http://math.nist.gov/MatrixMarket/`）の構成を図 1 に示す。行列データへのインタフェースとしては、

- Browse(閲覧) ...資料 6 ~ 10 参照。
- Search(検索) ...資料 11 ~ 14 参照。
- Interactive Generation(対話的データ生成) ...資料 15 ~ 18 参照。

などによる方法が提供されている。

3 行列工房

3.1 概要 [5][6]

行列演算アルゴリズムでは特定のサイズの行列に対して機能することがあらゆるサイズの行列に正しく機能する保証とはならない。そこで、行列データのサイズ変更に自由度が必要となる。しかし、Matrix Market で提供されている行列データはサイズ固定であるため、アルゴリズムのテスト・比較に十分とはいえない。そこで、行列のサイズ変更の自由度を高めた、テスト用行列データの作成と配布を行い、信頼性の高いアルゴリズムのテストをサポートするシステムとして、「ネットワーク計算用行列工房：Matrix Workshop」を開発した。（図 2 参照）

提供しているデータは、“A Collection of Matrices for Testing Computational Algorithms”（テスト行列集）と村田らによる純拡散問題（非線形定常問題と非線形時間依存問題）の 2 つのコレクションから取られたもので、全行列データ生成系の数は約 30 となっている。

3.2 データ形式

Matrix Market Exchange Formats を利用している。

3.3 インタフェース

CGI（資料 19・20 参照）や Java（資料 21・22 参照）により、対話的に行列を生成できるシステムが提供されている。

参照文献

[1] Matrix Market. (Last Change: January 22, 1998)
<http://math.nist.gov/MatrixMarket/>

[2] R. Boisvert, R. Pozo, K. Remington, R. Barrett, and J. Dongarra. The Matrix Market: A web resource for test matrix collections. In Ronald F. Boisvert, editor, Quality of Numerical Software, Assessment and Enhancement, p. 125-137, London, 1997. Chapman & Hall.
<http://math.nist.gov/MatrixMarket/reports/MMpaper.ps.gz>

[3] R. Boisvert, R. Pozo, and K. Remington. The Matrix Market exchange formats: Initial design. Technical Report NISTIR 5935, National Institute of Standards and Technology, Gaithersburg, MD, USA, December 1996.
<http://math.nist.gov/MatrixMarket/reports/MMformat.ps.gz>

[4] 行列工房. (Last change: Wed Nov 4 18:35:18 JST 1998)
<http://ninf.etl.go.jp/~ymatsuba/MatrixWorkshop/>

[5] 松原有里, 高木浩光, 関口智嗣, 長嶋雲兵. WWW を用いた行列生成システム: Matrix Workshop の開発. 情報処理学会第 55 回全国大会, 1997.
<http://www.sap.is.ocha.ac.jp/~matubara/MYRES/MW/swopp98ym.ps.gz>

[6] 松原有里, 中田秀基, 関口智嗣, 建部修見, 長嶋雲兵. ネットワーク計算用行列工房: Matrix Workshop による性能評価システム. 情報処理学会研究報告 98-HPC-72.
<http://www.sap.is.ocha.ac.jp/~matubara/MYRES/MW/swopp98ym.ps.gz>

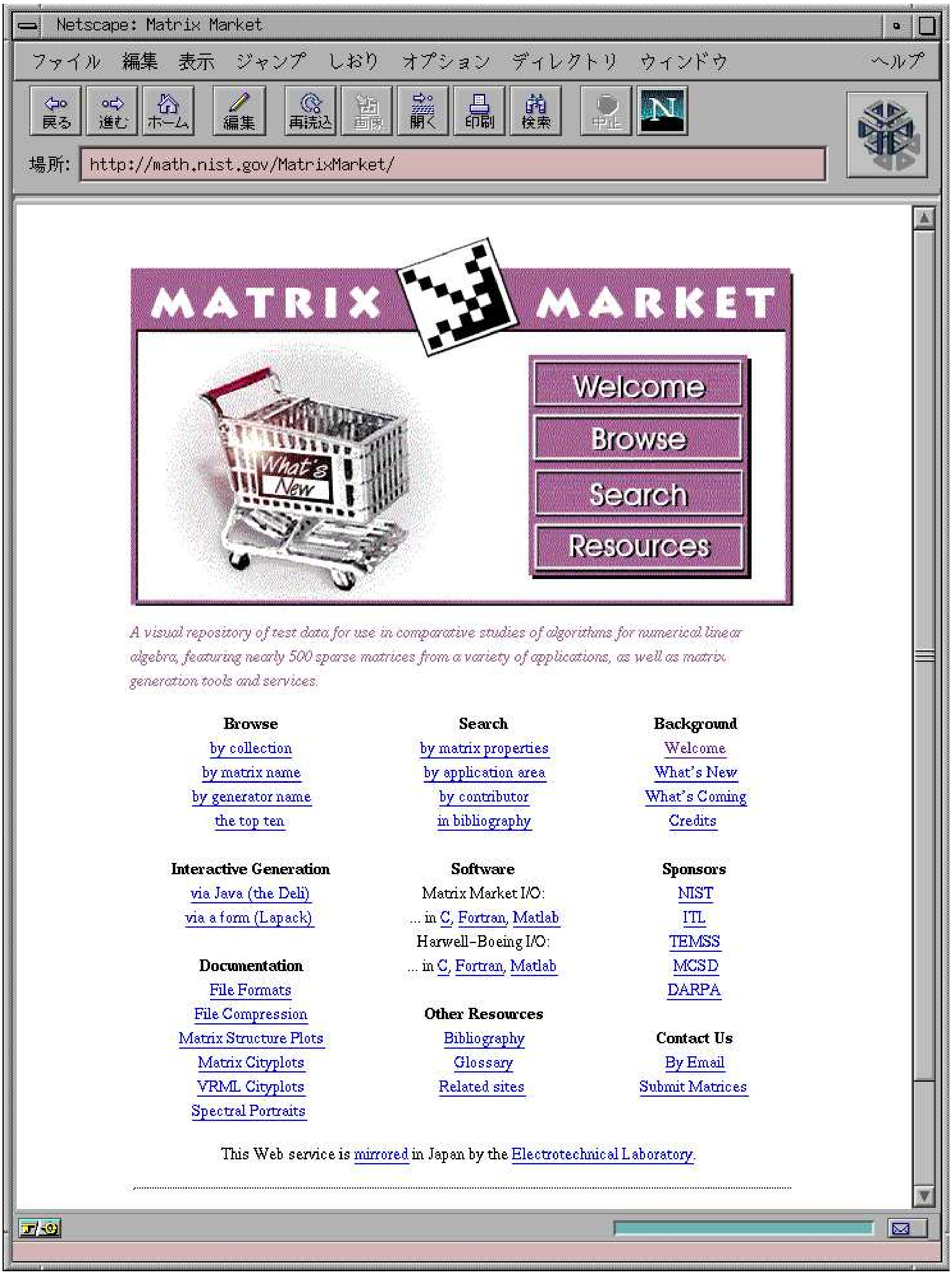

図 1: Matrix Market[1]

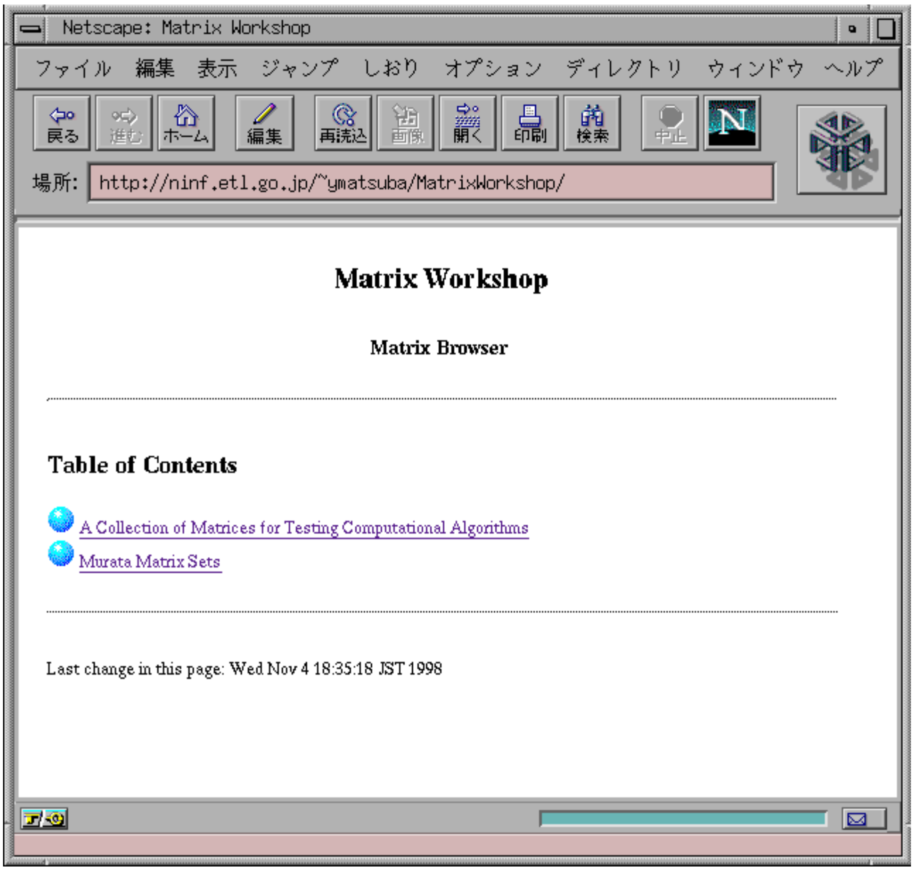
Netscape: Matrix Workshop
ファイル 編集 表示 ジャンプ しおり オプション ディレクトリ ウィンドウ ヘルプ
戻る
進む
ホーム
編集
再読込
画像
開く
印刷
検索
中止
場所:
http://ninf.etl.go.jp/~ymatsuba/MatrixWorkshop/
Matrix Workshop
Matrix Browser
Table of Contents
A Collection of Matrices for Testing Computational Algorithms
Murata Matrix Sets
Last change in this page: Wed Nov 4 18:35:18 JST 1998

図 2: 行列工房 [4]