



出来事を軸にしてデータを見る

松原 伸人

matubara@sra.co.jp

時間情報を持つ大量のデータを見て、データの特性を探索するようなツールの開発を行っています。

時間情報を持つデータは、テキストを書いた時刻、ニュース記事を読んだ時刻、写真を撮った時刻、絵を描いた時刻、ビデオを撮影した時刻などを扱っています。

Vol.100、102、103「時間情報を持つテキストの年表化」で紹介したプログラムは、テキスト、画像、動画で表される時系列データを読み込んで、時間情報に基づいたデータ操作を行えるようになっています。

現在、`chronicle.js` という名前で Web ブラウザ上で動作するプログラムを JavaScript で記述しています。

Vol.116 に書いたように、Chronicle オブジェクトを年表のように表示することもできれば、プログラムで検索したり順に見て行ったりもできるようになっています。

この時系列データを扱う Chronicle オブジェクトの対話方法もふくめた表現方法について様々な試作を行っています。

今回は試作の 1 つコードネーム `datarepeat` を紹介します。

◆ datarepeat

`datarepeat` は試作中の時系列データの可視化方法で、指定した出来事群を等間隔の目盛とする軸において、目盛間の時刻に応じた位置に全データを配置します。

`datarepeat` では、重要な出来事群を抜き出して軸にして、その前後でのデータの量とか種類の違いが見えるような可視化方法を目指しています。

上記のページを開くとデータファイル `gsletterneochronicle-publish_timeline.md` を読み込んだ Chronicle オブジェクトを指定して、ChronicleDatarepeat を作成します。

`gsletterneochronicle-publish_timeline.md` には、GSLetterNeo が発行された日、タイトル、著者、tfidf で抽出した名詞のうち数値が高い上位 10 単語が書かれています。

ChronicleDatarepeat は時系列データを HTML の Canvas で 2 次元描画します。[図 1]

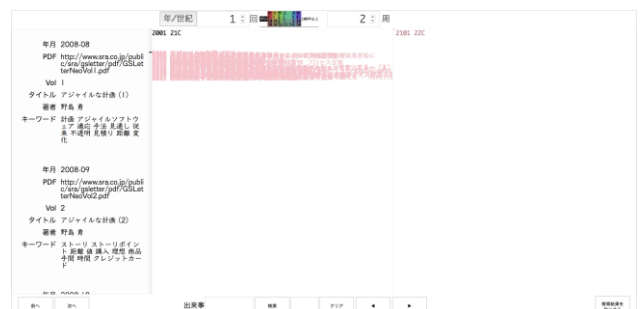


図 1 datarepeat を開いた直後の画面

画面左側三分の一くらいには出来事の内容を時間順に表示します。

画面右三分の二くらいにデータ全体を表示します。

縦が時間方向で、最初は西暦 2001 年から 2100 年の 1 世紀の期間を表しています。

薄赤色で描かれているのが GSGLetterNeo 記事の発行日とタイトルです。

「年/世紀」と書かれたボタンを押すと、「四季/年」にボタンがかわり、縦軸が春夏秋冬を目盛とする 1 年間を表すようになります。[図 2]

GSLetterNeo のデータは 2008 年 8 月から 2018 年 2 月までであるので横に年数分の 11 本の軸が並びます。

続けてボタンを押すと「月/年」[図 3]「検索結果」「クラスタ ID1」「クラスタ ID2」...「年/世紀」と順に切り替わります。

「周」を変えると、画面に表示する縦の軸の本数が変わります。図 4 は 3 周に変更した画面です。

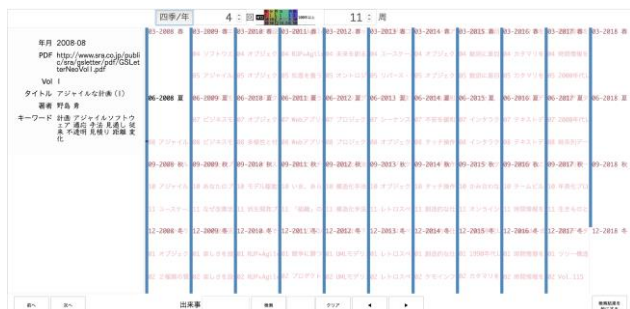


図 2 四季/年で表示した画面

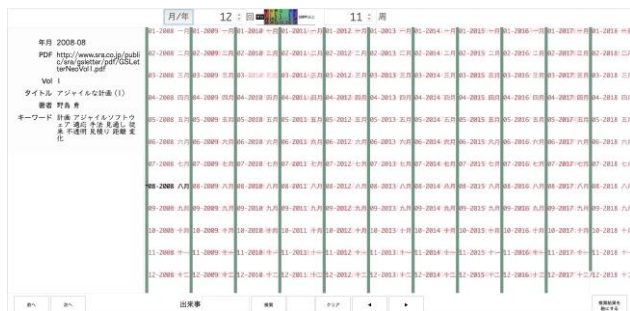


図 3 月/年で表示した画面

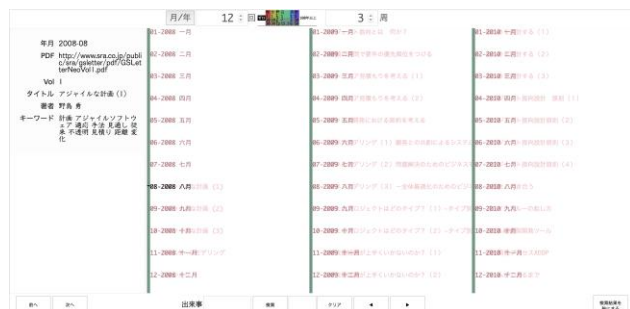


図 4 3 周に変更した画面

画面左下の「前へ」「次へ」ボタンを押すと、出来事を時間順に選び、画面左側に選んだ出来事を表示して、軸上で選ばれた出来事を黒色で表します。

画面中央下で出来事をテキスト検索できます。

「アジャイル」で検索すると、vol116 でプログラムで検索したのと同様に 6 個の出来事が見つかり、それらを黒色で表示します。[図 5]

◀▶□ボタンを押して検索にマッチした出来事を選んで順に見ていきます。

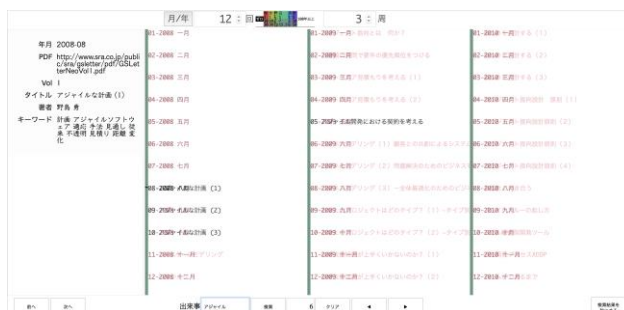


図 5 アジャイルで検索した画面

「検索結果を軸にする」ボタンを押すと、「アジャイル」で見つけた 6 個の出来事を等間隔に並べた目盛とする軸を配置し、全出来事を目盛が表す時間に応じた位置に配置します。[図 6]

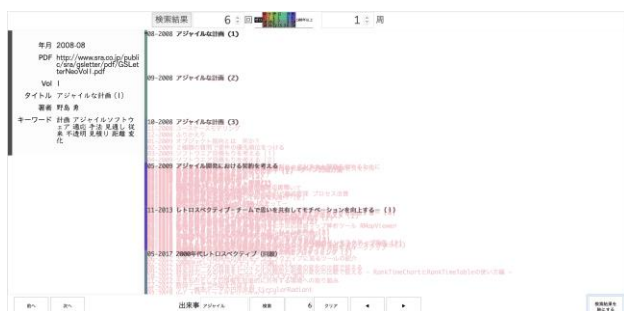


図 6 アジャイルの検索結果を軸にした画面

「アジャイル」で検索して見つかった出来事 6 個を軸にした場合だと、「アジャイルな計画が連作」で 3 本発行されたことがわかり、期間があいて 2009 年 5 月に「アジャイル開発における契約を考える」が発行され、それ以降

はしばらくアジャイルな記事は発行されずに、4年後の2013年11月に発行されたレトロスペクティブシリーズの中でアジャイル関連の内容に触れられていることがわかります。

軸の左側の色は、目盛間の時間の長さ、スケールに応じて変わって変わっています。

100年間以上は白色、1年から100年は紫系色、1ヶ月から12ヶ月は青系色、1日から31日はシアン系色、1時間から24時間は緑系色、1分から60分間は黄系色、1秒間から60秒間は赤系色、0秒は黒色といったようになっています。

◆GSLetterNeoのクラスタ

[GSLetterNeoのクラスタ](#)は、GSLetterNeoの全記事のテキストをMeCabで形態素解析して名詞を抽出し、名詞のTFiDFを特徴量とする記事のベクトルをk-meansで16のグループにクラスタリングした結果です。

日付を押すと、発行した日におけるクラスタリング結果を辿れるようになっています。

各グループには、クラスタの平均ベクトルの数値の大きい語10語を表示しています。

グループには、記事の中で使われている名詞の使用頻度が似ている記事をリストしています。

[datarepeat](#)では、[2018年2月20日のクラスタリング結果のデータファイル](#)を読み込んであり、グループを切り替えて軸にして見ていけるようにしてあります。

datarepeatの「クラスタID」は、2018年2月20日

に115号分の記事を16グループにクラスタリングした結果のグループを表しています。

「クラスタID」にすると、クラスタに含まれる記事群を縦に等間隔に並ぶ目盛とする軸を置いて、115号分の記事を配置します。[図7]



図7 クラスタID1を選んだ画面

クラスタID1の最初の2つの目盛になっている、09-2008 アジャイルな計画(2)と09-2012 構造化手法のススメの間に現れる記事群を、上述したようにプログラムで見ると、たとえば、目盛りになっている2個の出来事を取り出し、2つ出来事の出来事の間にある出来事を抜き出すコードで行えます。

```
var event1 = chronicle.findTimesWithDate(new Date('2008-09-01')).events()[0]
    event2 = chronicle.findTimesWithDate(new Date('2012-09-01')).events()[0],
    startIndex = chronicle.events().indexOf(event1),
    endIndex = chronicle.events().indexOf(event2),
    result = [];
for (index = startIndex; index <= endIndex; index += 1) {
    result.push(chronicle.events()[index]);
}
result
```

GSLetterNeo Vol.117
2018年4月20日発行
発行者●株式会社SRA 先端技術研究所
編集者●土屋正人

バックナンバを公開しています●<http://www.sra.co.jp/gsletter>
ご感想・お問い合わせはこちらへお願いします●gsneo@sra.co.jp

株式会社SRA

〒171-8513 東京都豊島区南池袋2-32-8

夢を。



夢を。Yawaraka Innovation
やわらかいのべーしょん