

時系列データの特長をビジュアルの類似と相違の変化の比較で捉える

松原 伸人

matubara@sra.co.jp

◆ はじめに

開発中のインタラクティブな Web アプリケーションプログラム、RankTimeChart と RankTimeTable を紹介します。

この2つはそれぞれ、大量の項目からなる時系列データ—各項目が日付と数値による—を読み込んで、ラインチャート、ランキング表で表示します。

◆ RankTimeChart

RankTimeChart は、複数の日毎の数値データを折れ線グラフにして表示します。[図 1]

日付を指定すると、指定した日の数値を比較してデータの順位をもとめ、上位 10 個に順位に応じた色をつけ、11 位以下は黒色にしてラインチャートを表示します。

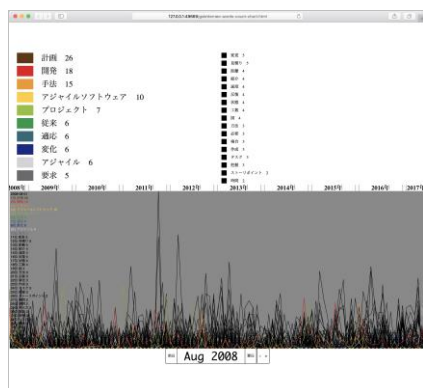


図 1 2008 年 8 月の単語数トップ 10 とその後の推移

◆ RankTimeTable

RankTimeTable は、日付を指定すると、指定した日の数値を比較してデータの順位をもとめ、上位 10 個に順位に応じた色をつけ、11 位以下は黒色にしてランキングを表示します。[図 2]

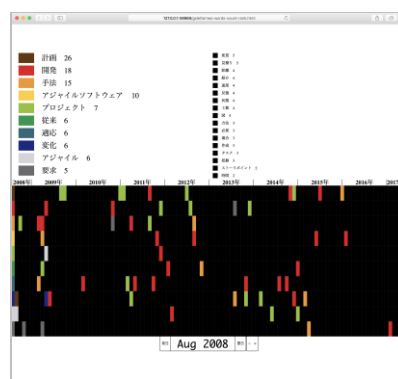


図 2 2008 年 8 月の単語数ランキングと推移

◆ トライアル&ビデオ

例として、GSLetterNeo の Vol.1 から Vol.107—2008 年 8 月から 2017 年 6 月—の期間に毎月発行されている全ドキュメントを形態素解析して単語に分解し、使用している名詞の数を月毎に集計し、各月で使用数の高い単語上位 10 個からなるデータを作りました。

次の URL で、それぞれ試すことができます。

リンク 1 [単語の出現数 top10 の推移](#)

リンク 2 [単語の出現数の順位の推移](#)

[単語の出現数 top10 の推移]はラインチャートで表示する、[単語の出現数の順位の推移]はランキング表で表示する、HTML と JavaScript で記述したプログラムです。

また、2008 年 8 月から 2017 年 6 月までを 約 1 月 /60 ミリ秒 で月を切り替えて表示するビデオを置いてあります。

ビデオ1 月別の GSLetterNeo 記事内で用いられている単語数と推移

ビデオ2 月別の GSLetterNeo 記事内で用いられている単語数のランキングの推移

ビデオを見る場合は、ダウンロードしてメディアプレイヤーなどで開き、再生せずに、タイムスライダを動かして見ると、実際に使っている雰囲気に近いと思います。

画面下半分のエリアでマウス動かして日付を指定すると、その日の上位 10 項目に色がつき、その日までという変化をしてきたか、その後どう変化したかを見比べられます。

画面最下部でも日付を指定できます。

RankTimeChart と RankTimeTable は常にデータの全期間を表示します。

◆ RankTimeChart で見る

RankTimeChart では、データ全体の変化を黒色で表し、上位 10 項目の変化をカラーで表すため、10 項目の単語の使用数について、どの時期に多く、どの時期に少ないかが見えやすいです。

例えば全体を見たときに一番とびぬけている箇所にマウスを持っていくと、2011 年 10 月に「管理」が 69 回使われ、他の時期ではほとんど使われていない、ということが分かります。

同じ 2011 年 10 月に 49 回使われている「構成」も、他の時期ではそれほど多く使われていないことが見て取れます。[図 3]

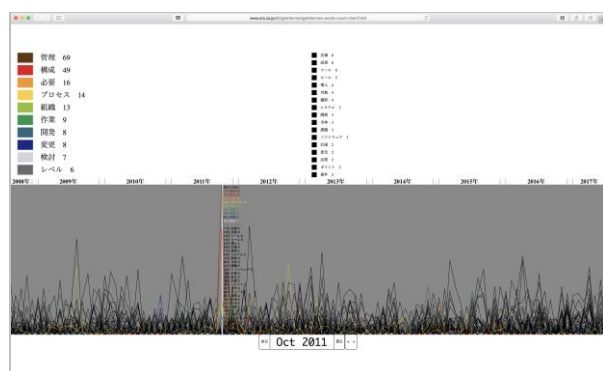


図 3 RankTimeChart 単語数 2011 年 10 月

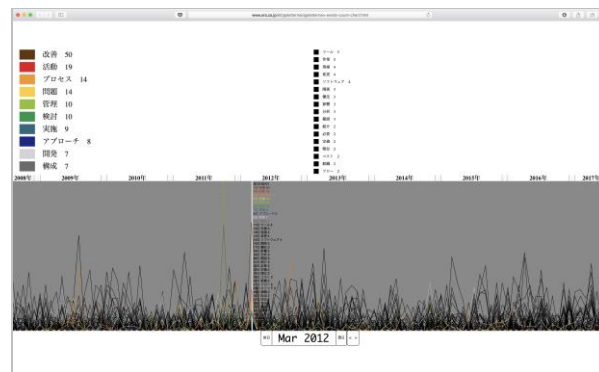


図 4 RankTimeChart 単語数 2012 年 3 月

次に飛び出しているところは、2012 年 3 月で 1 位が「改善」で 50 回使われていますが、他の時期ではほとんど使われていません。[図 4]

同月 4 位の「プロセス」は、他の時期にも割と使われていることが分かります。

また 2011 年 10 月で 1 位の「管理」がここでも 5 位になっていることから、「プロセス」と「管理」が共通の話題であることが想像できます。

2011 年 10 月は「Vol.39 いま、あらためて考える構成管理」、2012 年 3 月は「Vol.44 ボトムアップアプローチのソフトウェア構成管理 プロセス改善」というタイトルでした。[図 5]

2011年10月	2012年3月	2011年10月	2012年3月
管理 69	改善 50	改善 50	改善 50
構成 49	評価 19	改善 50	改善 50
必要 16	プロセス 14	改善 50	改善 50
組織 13	問題 14	改善 50	改善 50
作業 9	管理 10	改善 50	改善 50
開発 8	検討 10	改善 50	改善 50
変更 8	開発 9	改善 50	改善 50
検討 7	アプローチ 8	改善 50	改善 50
レベル 6	開発 7	改善 50	改善 50
	構成 7	改善 50	改善 50

図 5 GSLetterNeo 年表 2011 年 10 月から 2012 年 4 月の抜粋

◆ RankTimeTable で見る

RankTimeTable で同じ日付を見比べてみます。

[図 6]は 2011 年 10 月にフォーカスした状態です。

RankTimeTable では、10 位以下の項目は黒色で表し上位 10 項目だけ色をつけて目立たせます。

着目している 10 項目が現れる時期が見えやすくなります。

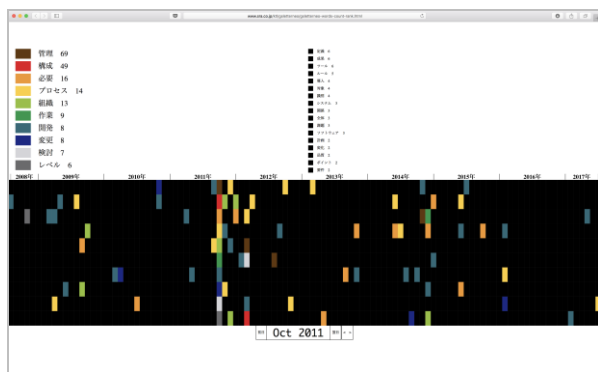


図 6 RankTimeTable 単語数 2011 年 10 月

例えば「管理」は 3 つの時期で多く使われている、「構成」は 2 箇所、「プロセス」は 16 箇所、「組織」7 箇所、「作業」2 箇所、「開発」27 箇所、「変更」5 箇所、「検討」2 箇所、「レベル」2 箇所、というように目で見ても数え上げられる程度に読み取れます。

[図 7]は 2012 年 3 月にフォーカスした状態です。

RankTimeChart で見たときは「改善」は他の時期に読み取れませんでした、10 箇所です使われていることが分かります。

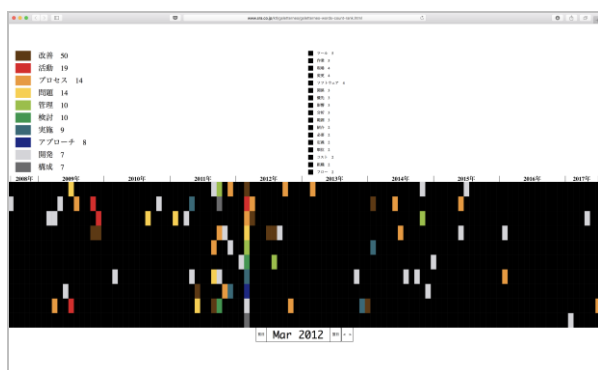


図 7 RankTimeTable 単語数 2012 年 3 月

「プロセス」と「開発」は、割と全体的に出てくるのが分かります。

翌月 2012 年 4 月でも「プロセス」と「改善」が他の単語と比べると多く使われているので、この 2 つが関係する話題なのだと想像できます。

2012 年 4 月は「Vol.45 未来を創る」というタイトルで一見するとプロセス改善と関係なさそうですが、内容を見てみると、プロセス改善について書かれていることが分かります。

◆ 連続的な現れ方

日付を変えてパラパラ見ていくと、頻繁に現れる単語と 1、2 回しか使われていない単語があるのが見えます。

例えば、最初の 2008 年 8 月から見ていくと、「開発」「プロジェクト」は 2014 年頃までは割と使われていて、「時間」が 2008 年 9 月と 2016 年頃に使われ、「クラス」は 2010 年から 2015 年初期まで、「モデル」は 2009 年 2 月から間を空けて何度か使用数が増える時期がある、といったことが分かります。

2008 年 9 月にフォーカスすると、翌月 2008 年 10 月で使われる 5 単語が両方の月で上位に入っていることから、「距離」「ストーリーポイント」「時間」「方法」「紹介」が共通な話題で連作になっていることが想像できます。[図 8]

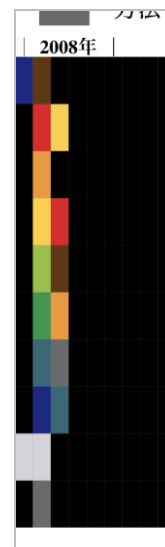


図 8 連続している部分の拡大画像

タイトルは「アジャイルな計画(2)(3)」でした。

同様に、ある月で多く使われる単語群が翌月も使われる箇所を抜き出したのが表 1 です。

ざっと見ただけでも、見落としがあつたりして精度的にはいい加減ですが、使用数の多い上位 10 単語が連続して現れる月の多くが、連作になっていることが見て取れました。

表 1 同一単語群が数ヶ月連続で現れる箇所

画像	時期	単語	タイトル
	2008 年 9 月 10 月	距離, ストーリーポイント, 時間, 方法, 紹介	アジャイルな計画 (2) (3)
	2009 年 3 月 4 月	工数, 開発, ポイント	ソフトウェア見積 もりを考える (1) (2)
	2009 年 9 月 10 月	プロジェクト, チーム, 組織, 自分	あなたのプロジェ クトはどのタイ プ? (1) (2)
	2009 年 11 月 12 月	ループ, 活動, フィードバック, 改善	なぜ改善活動が上 手くいかないの か? (1) (2)
	2010 年 1 月 2 月 3 月	注意, フロー, 目標	楽しさを設計する (1) (2) (3)
	2010 年 4 月 5 月	原則, クラス, 設計, 責任, オブジェクト, メソッド, モジュール	オブジェクト指向 設計原則 (1) (2)
	2012 年 7 月 8 月	プロジェクト, 診断, 改善, チャート, ふりかえり	プロジェクトの健 康診断してみま せんか? (1) (2)
	2013 年 7 月~10 月	設計, クラス, パターン	シーケンス図の階 層化 オブジェクト指向 モデリング (1) (2) (3)
	2015 年 3 月 4 月	クラス, 関連, 動詞, モデリング, 定義, 機能, ユースケース	動詞に着目したモ デリング手法 (1) (2)
	2017 年 2 月~4 月	年表, 矩形, 表示, 処理	時間情報を持つテ キストの年表化ス クリプト (2) 時間情報を有する データのビジュアル の Web プログラ ミング (1) (2)

◆ 単発的な現れ方

もう一度 2009 年 2 月の単語を見てみると、「モデル」以外はほとんど他の時期では使われていないことが分かります。

表 2 単語群がある月で上位、他の時期では少ない箇所

画像	時期	単語	タイトル
	2009 年 2 月	充足, 質問, 品質, フィーチャ, 満足, 優先, 順位, 魅力, モデル, 狩野	2 種類の質問で要件 の優先順位をつけ る
	2010 年 8 月	物語, ミュージカ ル, キャスト, 文 化, 想像, 大陸, 世界, 人, シーン, 自分	多様性と付き合う
	2011 年 11 月	ツール, 組織, 機能, 開発, 視点, 導入, 脆弱, 重要, プロセス, 結果,	組織のツールの視 点
	2015 年 2 月	反応, 経路, 把握, 物質, RMapView, 図, ProjDepot, 問い合わせ, 化学, 状況	統合開発管理プラ ットフォーム ProjDepot ケモインフォーマテ ィクスへの貢献~ 化学反応経路マッ プ解析ツール RMapView
	2015 年 8 月 9 月	円盤, 回転, 角度, 指, 模様, スポー ク, 中心, パイ, 実 装, wheely	インタラクティブ 体験アプリケー ションを作る (2) (3)
	2015 年 11 月	IMS, 学習, 教育, 規格, LD, Learning, 利用, モデル, 概 念, 教授	オンライン教育シ ステムの技術規格 IMS-LD
	2016 年 12 月	ログ, lambda, 出力, 式, 実行, Java, Apache, method, Commons, 場合	lambda 式の遅延実 行を利用したログ 出力

タイトルは「2 種類の質問で要件の優先順位をつける」でした。

同様に、ある月で上位を占め、他の時期では使われていないような単語数の変化の見られる時期とタイトルが表 2 です。

この 7 つは 1 回読み切りの内容でした。

2015 年 2 月は 2 冊発刊されていて、そのうちのケモインフォーマティクスに使われている単語が上位を占めています。

2015 年 8 月 9 月の「インタラクティブ体験アプリケーションを作る」は 7 月からの 3 回連続の内容ですが、1 回目はプロジェクトの紹介で、2 回目と 3 回目は実際に行ったことの紹介的な内容でした。よくみると 3 回に共通して「指」「円盤」が用いられています。

◆ 終わりに

GSLetterNeo の 2008 年 8 月から 2017 年 6 月のドキュメントで使われている単語数の変化から見えることを紹介しました。

このデータは月 1 回発行される 1 回が A4 サイズで 2 ページほどの文量で、データ量はそれほど多くはありません。

期間が長くなったり、発行間隔が短くなったり、ドキュメント数が多くなったりすると、見えてくるものが変わってくると思います。

データの種類がテキストではなく、画像であったり、取引数とか数値だったりするだけで、見えることが違ってくると思います。

次回は、RankTimeChart や RankTimeTable で自分のデータを見るプログラムの書き方を紹介します。

GSLetterNeo Vol. 109

2017 年 8 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

夢を。



株式会社SRA

〒171-8513 東京都豊島区南池袋 2-32-8

夢を。Yawaraka Innovation
やわらかいのべしょん