# ソーシャルブックマーキングサービスのデータ分析 によるウェブマーケティング

プロジェクトマネジメントコース ソフトウェア開発管理グループ 矢吹研究室 1342029 遠藤一輝

# 謝辞

本研究の作成において,矢吹研究室矢吹太朗准教授には数多くのご指導,ご鞭撻を賜り深謝致します.また,並びに数多くの助言やご指摘をくださいました矢吹研究室の同期,後輩の皆様に感謝致します.

# 目次

謝辞		1
第 1 章	序論	4
第2章	背景	5
第3章	ソーシャルブックマーキングサービスについて	6
3.1	本研究におけるソーシャルブックマーキングサービスの使用の理由について	6
3.2	ソーシャルブックマーキングサービスとは	7
第4章	はてなブックマークについて	9
4.1	はてなブックマークとは	9
4.2	エントリとは	9
4.3	はてなブックマークの特徴	10
4.4	株式会社はてな	13
4.5	はてなブックマークの選出理由	13
4.6	ホットエントリ	13
4.7	はてなブックマークの利用方法と用語解説	14
第5章	研究環境の構築	21
5.1	Virtual Box のインストール	21
5.2	Virtual Box のインストール	24
5.3	計算機環境の構築・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	27
第6章	目的	28
第7章	手法	29
7.1	データの収集方法	29
7.2	API によるデータの収集範囲	30
7.3	API を利用したデータ取得プログラム	31
7.4	JQuery による取得したデータの整理	32
7.5	取得したデータの抽出	37
7.6	取得したデータの処理	42
77	時間の推移とブックマーク数の増加の推移の可視化	45

7.8	時間経過によるブックマーク数の増加	46
7.9	時刻ごとのブックマーク数の増加	47
7.10	分析	50
第8章	結果	51
8.1	時間の推移に伴うブックマーク数の増加の傾向のパターン	51
8.2	各時刻ごとの増加数	54
8.3	時間経過による増加数の変化・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	55
第 9 章	考察	57
9.1	ブックマークの伸びやすい時間帯についての考察	57
9.2	時間の推移に伴うブックマーク数の増加の傾向の考察	59
9.3	時間経過による増加数の変化の考察	62
9.4	3 種の分析結果からの考察	63
第 10 章	結論	64
参考文献		65

## 第1章

## 序論

本研究は,ウェブマーケティングに関するデータの分析を行い,ウェブマーケティングの効率的な手法についての考察と提案を行う.

インターネットの発展に伴う情報の入手手段の増加と媒体の多様化により,私たちの周りには多くの情報があふれるようになった.それに伴いマーケティングの手法も新たな体系としてインターネットを利用したウェブマーケティングがうまれた.様々な媒体から手軽に見られるインターネットは新しいマーケティングの場として適したものであり,多くのマーケティングがウェブ上で行われている.

インターネットでは手軽に情報の入手及び発信が行えるが,インターネット上にある情報は膨大なため必要とする情報の取捨選択が求められる.そのため自身の発信する情報を多くの人に閲覧してもらうためには,発信する情報を効果的に伝えることが重要となる.

そこでウェブマーケティングを効果的に行うための指標を作成することを目的とし, ウェブマーケティングの分析を行う.

## 第2章

# 背景

インターネットで情報を得ようとするとき,インターネット上には個人では扱いきれないほどの多くの情報があふれており,必要な情報のみを選択しなければ調査に多くの時間がかかってしまう.

そのため私たちは情報を得るとき,情報の様々な要素からその情報が自身に必要なものかを判断し,取捨選択をすることで本当にその情報が自身が必要としているものか見極めている.

情報の様々な要素とはタイトルや掲載日時,タグなどの情報発見に必要な情報から詳しい内容やコメント,画像などのそれが自身にとって必要か判断するための情報など多岐にわたる種類のものである.

本研究では数多くある要素のうち,時間の推移と閲覧した人数の関係に対して研究を行い,その結果をもとに効果的なウェブマーケティングの手法を考察することで,ウェブマーケティングの一つの指標とすることを目標とする.

## 第3章

# ソーシャルブックマーキングサービ スについて

本章ではソーシャルブックマーキングサービスについて解説する.

## 3.1 本研究におけるソーシャルブックマーキングサービスの 使用の理由について

本研究においてはウェブマーケティングの一例としてソーシャルブックマーキングサービスを用いる.

その理由を以下に挙げる.

- 1. 様々な利用者が様々な分野の記事をブックマークに登録するため,幅広い分野のデータが得られる.
- 2. 不特定多数の人物のブックマークを参考にするため,思惟が関わりにくく偏りが起こりにくい.
- 3. ウェブマーケティングは幅広い範囲の分野に対応するため、比較のための共通の数値が必要となる. ソーシャルブックマーキングサービスでは、共通の数値であるブックマーク数というデータが得られるため、比較が容易となる.

## 3.2 ソーシャルブックマーキングサービスとは

ソーシャルブックマーキングサービスとは,ユーザーのブックマークをローカルではなくインターネット上に保存するサービスである.

インターネット上にブックマークを保存するため,端末が異なってもブックマークを共有できる.

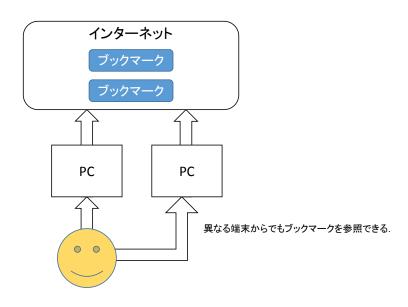


図 3.1 ソーシャルブックマーキングサービスのイメージ 1

また,他のユーザとブックマークを共有したり,タグをつけたり,コメントを残したり もできる.

自身のブックマークから推測し関心の持つ可能性の高いブックマークを発見しやすくなったり、注目のされているニュースなどを発見しやすくなる.また様々なユーザが使用しているため様々なジャンルのブックマークを発見しやすくなる.

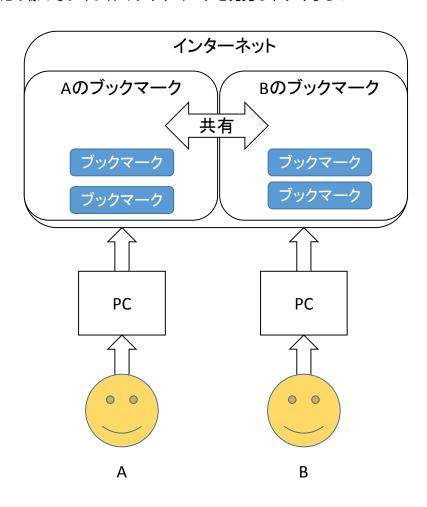


図 3.2 ソーシャルブックマーキングサービスのイメージ 2

## 第4章

# はてなブックマークについて

本章では本研究で使用するソーシャルブックマーキングサービスであるはてなブックマークについて解説する.

### 4.1 はてなブックマークとは

はてなブックマークとは,株式会社はてなの提供するソーシャルブックマーキングサービスである.

はてなブックマークは 2005 年 2 月にベータ版が開始され , 同年 8 月に正式版が開始された [1] .

## 4.2 エントリとは

はてなブックマークに登録されているブックマークのことをエントリという. エントリは同じサイトをブックマークしている人がつけたタグやコメントを共有できる.

## 4.3 はてなブックマークの特徴

株式会社はてなは,はてなブックマークの特徴として以下の3つの点を挙げている [2][3].

#### 4.3.1 保存

端末を選ばないオンラインブックマーキングサービスと検索,タグ機能による情報の取り出しの容易さ.

- 1. オンラインにさえつながっていればいつでもブックマークを参照できる.
- 2. 様々な媒体でも共通のブックマークを参照できる.
- 3. 1000 件以上のブックマークからでもタグなどの機能によって検索できる.

#### 保存

家で見たページを学校や会社でも見たい

そんな風に思った経験はありませんか?

はてなブックマークは、オンラインにあなたのブックマークを保存するツールです。インターネットに繋がっていれば、家でも学校や会社でもブックマークを見られます。タグやコメントを付けて整理しましょう。保存したブックマークは簡単に検索できます。

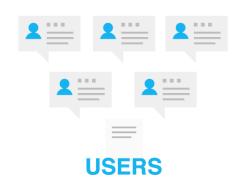


図 4.1 はてなブックマークの特徴 1-保存

#### 4.3.2 共有

ブックマークの公開による他のユーザーとのブックマークの共有ができる.

- 1. 公開されているほかの人のブックマークを参照できる.
- 2. 自身が広めたいと感じたブックマークは公開して共有できる.
- 3. コメントやタグをつけることで感想を共有できる.



#### 共有

このページを読んだ人の感想が見たい

ブックマークを公開して、多くの人と感想をシェアできます。このベージはぜひ他の人にも見て欲しい、そう思ったらどんどんブックマークしましょう。 感想を書いておけば、だれかの目に触れるはずです。 はてなブックマークは、インターネットのあらゆるベージを介して人と繋がるソーシャルネットワークです。

図 4.2 はてなブックマークの特徴 2-共有

#### 4.3.3 発見

人気エントリやキーワード検索などの機能によって情報の入手に幅ができる.

- 1. 人気エントリーや新着エントリーによって話題の情報が入手できる.
- 2. タグやキーワードで解く手のジャンルのみを検索できる.
- 3. 自分と似たユーザーをお気に入り登録すれば同じ趣味嗜好の情報を入手できる.

#### 発見

#### 今、話題になっているページが知りたい

はてなブックマークは、旬なニュースが見つかるメディアです。

人気ブックマークをチェックすれば、みんなが見ているベージがわかります。タグを使って、特定のトビックを追いかけられます。趣味が近いユーザーのブックマークから、自分好みの情報を簡単に見つけられます。おもしろいベージをどんどん発見しましょう!



図 4.3 はてなブックマークの特徴 3-発見

### 4.4 株式会社はてな

はてなブックマークを運営している株式会社はてなは 2001 年 7 月に有限会社はてなと して設立され, 2004 年 2 月に株式会社はてなに改組した [1].

人力検索はてなを始めとする様々なウェブサービスを提供している [4].

### 4.5 はてなブックマークの選出理由

ソーシャルブックマーキングサービスの中からはてなブックマークを選んだ理由を以下に挙げる.

- 1. 国内サービスの中では最大規模のソーシャルブックマーキングサービスであるため.
- 2. API を公開してるため,データの収集が容易である.
- 3. コメントなどのが機能があるため収集できるデータが多い.

### 4.6 ホットエントリ

ホットエントリとははてなブックマークで最近人気の勢いがあるブックマークである. 本研究ではこのホットエントリを取り扱い分析を進める.

### 4.7 はてなブックマークの利用方法と用語解説

はてなブックマークの画面について解説する.

#### 4.7.1 はてなブックマークのトップ画面

この画面では多くのユーザーに共有されているブックマークが表示される.



図 4.4 ブックマーク一覧での表示

#### 検索ウィンドウ (図 4.1 1)

ここにキーワードを打つことでその言葉に関連するブックマークを検索することができる.

図 4.2 はその結果である.

#### ジャンル一覧 (図 4.1 2)

はてなブックマークでのジャンル一覧であり,ジャンルの名前をクリックすることで ブックマークをジャンルごとに表示させる.

#### ユーザー数 (図 4.1 3)

そのブックマークを共有しているユーザー数である.クリックすることで詳細画面へ 行く.

#### タイトル (図 4.1 4)

そのブックマークのタイトルである.クリックすることでそのブックマークのサイトへ 行く.

#### 内容 (図 4.1 5)

そのブックマークの内容を表示する.表示しきれない場合は省略される.

#### 画像 (図 4.1 6)

そのブックマークにある画像が表示される.画像がない場合は表示されずそのスペース に内容が表示される.

#### ジャンル (図 4.1 7)

そのブックマークが分類されているジャンルが表示される.クリックすることでその ジャンルが表示される.

#### 掲載日時 (図 4.1 8)

そのブックマークが掲載された時間である.

#### 提供者タグ (図 4.1 9)

そのブックマークのエントリを提供しているサイトのタグである.クリックすることで そのサイトの関連エントリが表示される.エントリとははてなブックマークで登録されて いるブックマークのことである.

#### 関連エントリ (図 4.1 10)

そのエントリに関連付けられたブックマークがある場合表示される.関連がない場合は表示されない.

#### 4.7.2 はてなブックマークでの記事検索画面

この画面では検索したワードにヒットしたブックマークが表示される.

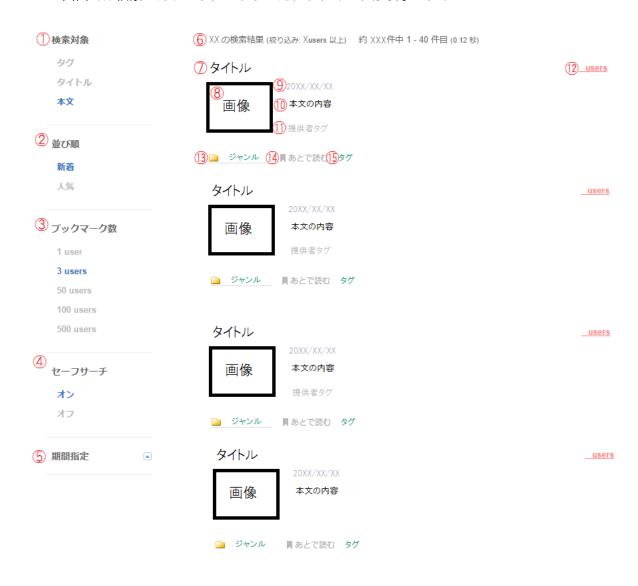


図 4.5 ブックマークの検索結果

#### 検索対象設定 (図 4.2 1)

検索ワードの対象を設定する.タグ,タイトル,本文の3つから選択できる.

#### 並び順 (図 4.2 2)

検索結果のソートを行う.時間と登録数でソートが行える.

#### ブックマーク数絞り込み (図 4.2 3)

ブックマーク数による絞り込みを行う.クリックした数以上のブックマークを表示する. 1,3,50,100,500の数値で絞り込みが行える.

#### セーフサーチ (図 4.2 4)

検索結果にフィルターをかける.オンにした場合検索結果から不適切な表現がある記事 を除外する.

#### 期間指定 (図 4.2 5)

表示するブックマークの掲載された期間の指定をする.プルダウンを押下すると期間指定のための入力ウィンドウが2つ表示される.1つ目の入力ウィンドウから2つ目の入力ウィンドウの間の期間が表示される.

#### 検索結果のヒット数 (図 4.2 6)

検索した結果,該当したブックマークの数が表示される.

#### タイトル (図 4.2 7)

検索にヒットしたブックマークのタイトルである.検索したワードがタイトルに含まれている場合,その部分がハイライトされる.クリックすることでブックマークのサイトが表示される.

#### 画像 (図 4.2 8)

そのブックマークにある画像が表示される.画像がない場合は表示されずそのスペース に内容が表示される.

#### 掲載日時 (図 4.2 9)

そのブックマークが掲載された時間である.期間指定をした場合この数値を参照する.

#### 内容 (図 4.2 10)

そのブックマークの内容を表示する.表示しきれない場合は省略される.

#### 提供者タグ (図 4.2 11)

そのブックマークの記事を提供しているサイトのタグである.クリックすることでその サイトの関連記事が表示される.

#### ユーザー数 (図 4.2 12)

そのブックマークを共有しているユーザー数である.クリックすることで詳細画面へ 行く.

#### ジャンル (図 4.2 13)

そのブックマークが分類されているジャンルが表示される.クリックすることでその ジャンルが表示される.

#### あとで読む (図 4.2 14)

そのブックマークに後で読むタグをつけ、未読の場合通知する.

#### タグ (図 4.2 15)

そのブックマークにつけられたタグを表示する.タグ検索ではここを参照する.タグを クリックすると同じタグをもつブックマークを表示する.

### 4.7.3 ブックマーク記事の詳細

この画面ではブックマークの詳細を見ることができる.

ジャンル	
①XXXX ②タイトル user	
3 ジャンル 20XX/XX/XX XXXX	
5 97	
⑥ 本文の内容	
(7)     (8)       人気コメント     新著コメント     すべてのブックマーク	
⑩コメント	_

図 4.6 ブックマークの詳細表示

#### ユーザー数 (図 4.3 1)

そのブックマークを共有しているユーザー数である.

#### タイトル (図 4.3 2)

そのブックマークのタイトルである.クリックすることでそのブックマークのサイトへ 行く.

#### ジャンル (図 4.3 3)

そのブックマークが分類されているジャンルが表示される.クリックすることでその ジャンルが表示される.

#### 掲載日時 (図 4.3 4)

そのブックマークが掲載された時間である.

#### タグ (図 4.3 5)

そのブックマークにつけられたタグを表示する.タグをクリックすると同じタグをもつ ブックマークを表示する.

#### 内容 (図 4.3 6)

そのブックマークの内容を表示する.動画などの場合はここで再生もできる.

#### 人気コメント (図 4.3 7)

コメント欄に表示するコメントを多くスターを集めているものだけにする.

#### 新着コメント (図 4.3 8)

コメント欄に表示するコメントを新着順にする.

#### すべてのブックマーク (図 4.3 9)

コメントの有無にかかわらずこのエントリをブックマークしたユーザを表示する.

#### コメント欄 (図 4.3 10)

図 4.2 7.8.9 の押したものによってそれに合うコメントを表示する.

## 第5章

# 研究環境の構築

本研究を進める上での環境を解説する.

## 5.1 Virtual Box のインストール

本研究を進めるためのプログラムを起動する仮想マシンの構築を行う.

### 5.1.1 Virtual Box とは

仮想マシンの構築,管理を行うソフトである.仮想マシン上ではホストとは異なる環境を構築することも可能であり,様々な作業を仮想マシン上で行うことが可能である[5].

#### 5.1.2 Chocolatey のインストール

Virtual Box をインストールするためにまず Chocolatey をインストールする. コマンドプロンプトを開き,以下の画像のコマンドを実行する.

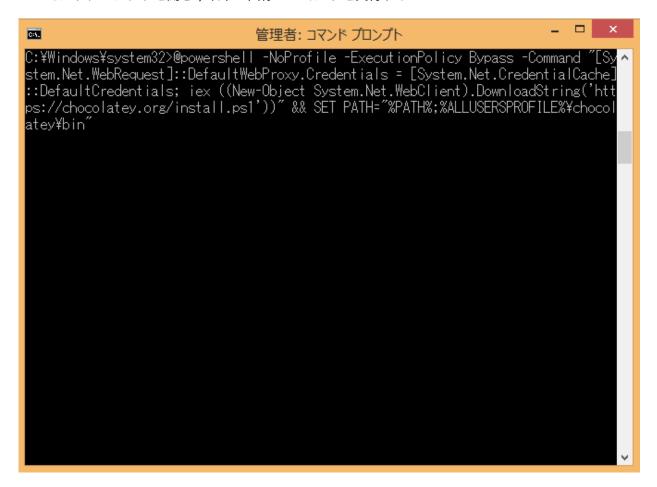


図 5.1 Chacolatey のインストール画面

以下の画像のようになったら成功である.

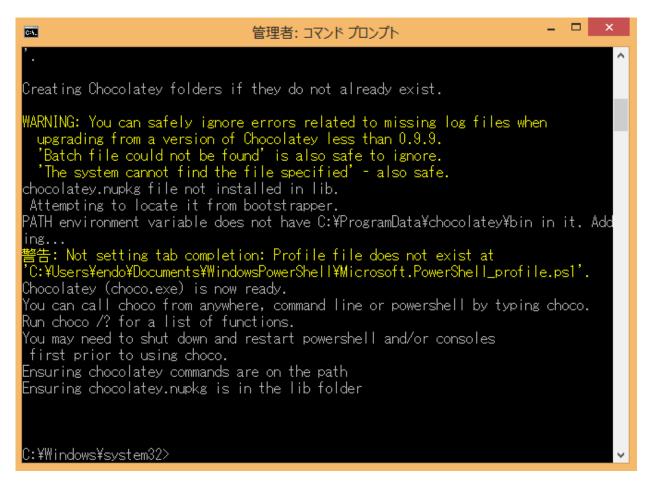


図 5.2 Chacolatey のインストール画面 2

## 5.2 Virtual Box のインストール

コマンドプロンプトを開き,以下の画像のコマンドを入力する.



図 5.3 Virtual Box のインストール画面

#### ここで Y を入力し Enter を押す.

```
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\footnote{Windows} (Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\footnote{Windows} (Version 6.3.9600)
(c) 2013 Microsoft Corporation. All rights reserved.

C:\footnote{Windows} (Version 6.3.9600)
(c) 2013 Microsoft Corporation. All rights reserved.

C:\footnote{Windows} (Version 6.3.9600)

All rights reserved.

All rights
```

図 5.4 Virtual Box のインストール画面 2

以下の画像のようになったら成功である.



図 5.5 Virtual Box のインストール画面 3

### 5.3 計算機環境の構築

#### 5.3.1 計算機環境の導入

本研究には矢吹研究室にて構築した計算機環境をクローンして使用する.

```
c:
cd \
mkdir vagrant
cd vagrant
git clone https://github.com/yabukilab/machine.git
```

図 5.6 導入コマンド

#### 5.3.2 コマンド

基本的なコマンドを以下に記す.

vagrant up

仮想マシンの起動を行う.

vagrant ssh

仮想マシンの再起動を行う.

vagrant halt

仮想マシンの停止を行う.

vagrant destroy

仮想マシンの削除を行う.

## 第6章

# 目的

ソーシャルブックマーキングサービスにおけるサイトのブックマーク数の推移を調査し 分析することで,ブックマーク数の増加傾向や伸びやすい時間帯の分析を行う.

その結果をもとに考察を行い,増加傾向のパターンの可視化及び分類をする.それによりウェブマーケティングを行う上での指標の一つを作成することを目的とする.

## 第7章

# 手法

本研究ではソーシャルブックマーキングサービスの 1 つであるはてなブックマークを使用しデータを収集する.

### 7.1 データの収集方法

はてなブックマークの API[6] を使用し, はてなブックマークのトップ画面からエントリ情報を収集する.

はてなブックマークの API 用の URL の後ろに対象の URL を記載することで対象のエントリの各データを json 形式で取得できる.



図 7.1 はてなブックマークトップ画面

### 7.2 API によるデータの収集範囲

はてなブックマークのトップページの人気のエントリ 100 件を対象に,はてなブックマークの API[7] を使用しデータを取得する. API を使用することにより対象エントリに含まれるデータが取得できる. 以下に取得できるデータの種類を列挙する.

#### 7.2.1 取得できる対象エントリのデーター覧

- 1. title
- 2. count
- 3. url
- 4. entry-url
- 5. screenshot
- 6. eid
- 7. bookmarks ユーザーが対象エントリに行ったブックマークごとの詳細である.
- 8. related 対象エントリに関連の深いエントリのデータである.

#### 7.2.2 取得データの詳細

対象エントリ取得した 8 種のデータのうち bookmarks と related に関しては複数の要素を含む.以下にその種類を列挙する.

#### bookmarks

- 1. user
- 2. tags
- 3. timestamp
- 4. comment

#### related

- 1. title
- 2. url
- 3. entry-url
- 4. eid

## 7.3 API を利用したデータ取得プログラム

はてなブックマークの API を利用し,エントリデータを取得するプログラムを作成した. 以下に作成したプログラムの画像とその一連の流れの説明である.

```
#for i in `cat urllist`;do
#for v in `seq 1 10`;do
#mkdir ./`date '+%F'`
#curl "http://b.hatena.ne.jp/entry/json/`$i`" | jq '.' > ./`date '+%F'`/$v.json
#grep "timestamp" ./`date '+%F'`/$v.json | gawk -f gsub.awk > ./`date '+%F'`/$v.csv

#done
#done

mkdir ./`date '+%F'`
for v in `seq 1 10`;do
curl "http://b.hatena.ne.jp/entry/json/`cat urllist$v.txt`" | jq '.' > ./`date '+%F'`/$v.json
grep "timestamp" ./`date '+%F'`\$v.json | gawk -f gsub.awk > ./`date '+%F'`\$v.csv

done
```

図 7.2 データ取得プログラム

事前に urllist という txt ファイルを作っておき,そこに対象のエントリのもとのページの URL を格納しておく.

curl によって対象の URL の前にはてなブックマークの API を接続し,取得したデータを JQuery で整理し, json 形式で出力する.

grep によって timestamp を抽出し, awk によって時間データのみに置換する.その後csv 形式に日付データを出力する.

## 7.4 JQuery による取得したデータの整理

APIによって取得したデータの JQueryによる処理の説明について記す.

#### 7.4.1 データの整理

はてなブックマークの API で取得したデータは json 形式で保存される. json 形式は整理を行わなくては閲覧が難しいためデータの整理を行い, 閲覧の可能な状態にする必要がある.

データの整理には JQuery を使用する.

### 7.4.2 JQuery について

Javascript のライブラリの一つであり,今回は json 形式のファイルの整理に使用する.

#### 7.4.3 整理の実行

取得した json 形式のデータに JQuery を実行することで閲覧可能な状態に整理することができる.

### 7.4.4 json 形式のデータの構造

以下にはてなブックマークの API を使用して取得した json 形式のデータの構造を記す.

#### エントリの基本情報

json の上部に記載されているエントリの基本情報である.

```
{
    "entry_url": "http://b.hatena.ne.jp/entry/www3.nhk.or.jp/news/html/20170105/k10010829301000.html",
    "screenshot": "http://screenshot.hatena.ne.jp/images/200x150/8/b/2/2/8/c5191a9289d61a0604cdadd6e5c3c58608d.jpg",
    "title": "住民票などコンビニでの交付導入促進 財政支援を拡充へ | NHKニュース",
    "eid": 314796750,
    "url": "http://www3.nhk.or.jp/news/html/20170105/k10010829301000.html",
```

図 7.3 json 形式のファイルの構造例 1

#### entry-url

はてなブックマーク内での対象エントリの概要ページの URL である. http://b.hatena.ne.jp/entry/... の形式で表示される.

#### screenshot

対象エントリのスクリーンショットの URL である. 画像は jpg 形式となる. http://screenshot.hatena.ne.jp/images/... の形式で表示される.

#### title

対象エントリのタイトルである.

#### eid

内部でふられている対象エントリの ID である. 9 桁の数字が羅列される.

#### url

対象エントリのもとのページの URL である.

#### 7.4.5 エントリのブックマーク情報

bookmarks は基本情報の下に記録されているブックマークの情報の一覧である. bookmarks には4種類のデータが含まれる.

```
{
    "tags": [],
    "user": "guoookoe",
    "timestamp": "2017/01/05 04:53:42",
    "comment": "住民票などコンビニでの交付導入促進 財政支援を拡充へ"
    }
],
"count": 189,
```

図 7.4 json 形式のファイルの構造例 2

#### tags

対象エントリにユーザーがつけたユーザーごとのタグである. 複数ある場合はそのすべてが列挙される.

#### user

対象エントリをブックマークしたユーザーの名前である.

#### timestamp

対象エントリがブックマークされた時間の記録である. 西暦/月/日 時:分:秒の形式で表記される.

#### comment

対象エントリのブックマークにつけられたコメントの内容である.

#### count

対象エントリをブックマークした合計人数である.

#### 7.4.6 エントリに関連の深いエントリ情報

related はブックマーク情報の下に記録されている関連エントリの情報である.related には種類のデータが含まれる.

```
"related": [
{
    "entry_url": "http://b.hatena.ne.jp/entry/www.jiji.com/jc/zc?k=201406/2014061400191&g=pol",
    "title": "時事ドットコム:住基力ード発行、3年ぶり増加=「コンビニで証明書」追い風に-総務省",
    "eid": 203313781,
    "url": "http://www.jiji.com/jc/zc?k=201406/2014061400191&g=pol",
    "count": 1
},
```

図 7.5 json 形式のファイルの構造例 3

#### entry-url

はてなブックマーク内での関連エントリの概要ページの URL である. http://b.hatena.ne.jp/entry/... の形式で表示される.

#### title

関連エントリのタイトルである.

#### eid

内部でふられている関連エントリの ID である.9 桁の数字が羅列される.

#### url

関連エントリのもとのページの URL である.

#### count

関連対象エントリをブックマークした合計人数である.エントリをブックマークした合計人数である.

### 7.4.7 その他のデータの記録

研究における参考やデータの整理のために使用するデータを以下に記す.

#### 記録日時

処理のため各エントリにつける番号である.

#### 通し番号

対象エントリのデータを取得した時間である.

#### ジャンル

対象エントリがはてなブックマークで分類されているジャンルの種類である.

# 7.5 取得したデータの抽出

はてなブックマークの API によって取得したデータから本研究に使用する要素のみを抽出する. 抽出には grep を使用する. また余分データの置換には awk を使用する. 以下に処理の手順について解説する.

### 7.5.1 grep とは

linux コマンドの一つで文字列の検索,抽出を行えるコマンドである.

#### 7.5.2 awk とは

unix で開発されたプログラム言語であり、テキストファイルの処理を行える。 区切りのあるデータの処理を特に得意とする。

### JQuery によるデータの整理

json ファイルを整理し,閲覧可能な状態にしている.

以下に画像は JQuery によって閲覧が可能となったデータである.

図 7.6 JQuery による整理の結果

# grep によるデータの抽出 grep で抽出したブックマークの時間データは以下の画像のようになる.

```
1
          "timestamp": "2016/11/24 19:05:11",
         "timestamp": "2016/11/24 18:00:17",
         "timestamp": "2016/11/24 10:36:10",
         "timestamp": "2016/11/23 11:33:11",
         "timestamp": "2016/11/23 01:20:58",
         "timestamp": "2016/11/21 17:31:14",
         "timestamp": "2016/11/21 17:19:37",
         "timestamp": "2016/11/20 21:02:29",
         "timestamp": "2016/11/20 05:51:34",
         "timestamp": "2016/11/19 20:55:24",
         "timestamp": "2016/11/19 19:58:11",
         "timestamp": "2016/11/19 14:22:32",
         "timestamp": "2016/11/19 13:30:23",
         "timestamp": "2016/11/18 12:29:42",
         "timestamp": "2016/11/18 12:08:41",
         "timestamp": "2016/11/17 17:59:11",
         "timestamp": "2016/11/16 22:37:54",
         "timestamp": "2016/11/16 18:11:46",
         "timestamp": "2016/11/16 12:49:59",
         "timestamp": "2016/11/16 09:14:05",
         "timestamp": "2016/11/16 01:41:30",
         "timestamp": "2016/11/16 00:50:52",
         "timestamp": "2016/11/15 23:58:55",
         "timestamp": "2016/11/15 23:07:53",
         "timestamp": "2016/11/15 22:50:05",
         "timestamp": "2016/11/15 18:35:31",
         "timestamp": "2016/11/15 14:14:34",
         "timestamp": "2016/11/15 11:13:56",
         "timestamp": "2016/11/15 08:55:31",
         "timestamp": "2016/11/15 05:45:51",
         "timestamp": "2016/11/14 23:37:40",
         "timestamp": "2016/11/14 21:03:42",
```

図 7.7 grep による時間データの抽出

#### awk による余分データの置換

余分データを置換することで時間データのみの抽出を行う.

以下の画像は今回処理に用いた awk である.

```
2 {gsub(/\"timestamp\": \"/,"",$0);
3 gsub(/\"\,/,"",$0);
4 gsub(/ /,"",$0)
5 gsub(/ /,",",$0)
6 print $0}
```

図 7.8 実行した awk

また,以下の画像は時間データのみを抽出した結果である.

2017/1/7	14:00:48
2017/1/6	11:48:46
2017/1/4	22:49:55
2017/1/3	0:07:01
2017/1/1	6:44:51
2016/12/29	14:03:06
2016/12/28	8:43:35
2016/12/27	6:13:29
2016/12/27	5:45:34
2016/12/25	17:43:28
2016/12/24	16:41:41
2016/12/24	12:20:38
2016/12/24	5:19:53
2016/12/24	3:10:52
2016/12/22	23:47:01
2016/12/21	3:02:24
2016/12/21	1:50:39
2016/12/20	15:03:48
2016/12/20	4:09:17
2016/12/19	8:58:04
2016/12/19	8:08:03
2016/12/19	7:43:53
2016/12/18	22:26:05
2016/12/18	7:19:39
2016/12/17	20:47:20
2016/12/17	12:53:34
2016/12/17	10:41:15
2016/12/17	10:23:09
2016/12/16	23:08:16
2016/12/16	7:23:53
2016/12/15	22:17:58
2016/12/15	20:19:04
2016/12/15	19:59:15
2016/12/15	19:28:59
2016/12/15	19:18:39
2016/12/15	18:22:53

図 7.9 awk による余分データの置換結果

# 7.6 取得したデータの処理

はてなブックマークの API を使用し取得したデータを分析するために行う処理について 記載する.以下の一連の作業をマクロとして記録し,処理を行う.

### 7.6.1 データの変換

はてなブックマークを使用して取得した json 形式のデータを csv 形式に変換し出力する.csv 形式で出力されたデータは以下のようになる.

2017/1/4	19:09:44
2017/1/4	18:50:27
2017/1/4	17:39:50
2017/1/4	15:13:26
2017/1/4	12:35:48
2017/1/4	12:24:42
2017/1/4	12:21:39
2017/1/4	12:05:43
2017/1/4	11:20:56
2017/1/4	11:15:44
2017/1/4	9:55:02
2017/1/4	9:31:29
2017/1/4	9:30:23
2017/1/4	8:54:53
2017/1/4	8:32:08
2017/1/4	8:28:56
2017/1/4	8:14:27
2017/1/4	7:59:39
2017/1/4	7:54:18
2017/1/4	7:29:30
2017/1/4	7:24:17
2017/1/4	6:58:33
2017/1/4	6:40:21
2017/1/4	6:33:18
2017/1/4	3:26:07
2017/1/4	2:50:06
2017/1/4	2:13:54
2017/1/4	2:05:26
2017/1/4	1:41:36
2017/1/4	1:03:49
2017/1/4	0:54:06
2017/1/4	0:48:37
2017/1/4	0:17:08
2017/1/4	0:01:39
2017/1/3	23:53:37
2017/1/3	23:50:24

図 7.10 csv 形式での出力結果

### 7.6.2 データの統合

csv 形での出力では初期状態では日付と時刻が同一セルに存在していないため,それらを統合する必要がある.

excel では日付及び時刻を数値に変換した場合,自然数部分が日付にあたり,小数点以下が時刻にあたる.そのため日付と時刻に対し加算を行えば統合することができる.

以下の画像は時間データの統合結果である.

2017/1/4	19:09:44	2017/1/4 19:09
2017/1/4	18:50:27	2017/1/4 18:50
2017/1/4	17:39:50	2017/1/4 17:39
2017/1/4	15:13:26	2017/1/4 15:13
2017/1/4	12:35:48	2017/1/4 12:35
2017/1/4	12:24:42	2017/1/4 12:24
2017/1/4	12:21:39	2017/1/4 12:21
2017/1/4	12:05:43	2017/1/4 12:05
2017/1/4	11:20:56	2017/1/4 11:20
2017/1/4	11:15:44	2017/1/4 11:15
2017/1/4	9:55:02	2017/1/4 9:55
2017/1/4	9:31:29	2017/1/4 9:31
2017/1/4	9:30:23	2017/1/4 9:30
2017/1/4	8:54:53	2017/1/4 8:54
2017/1/4	8:32:08	2017/1/4 8:32
2017/1/4	8:28:56	2017/1/4 8:28
2017/1/4	8:14:27	2017/1/4 8:14
2017/1/4	7:59:39	2017/1/4 7:59
2017/1/4	7:54:18	2017/1/4 7:54
2017/1/4	7:29:30	2017/1/4 7:29
2017/1/4	7:24:17	2017/1/4 7:24
2017/1/4	6:58:33	2017/1/4 6:58
2017/1/4	6:40:21	2017/1/4 6:40
2017/1/4	6:33:18	2017/1/4 6:33
2017/1/4	3:26:07	2017/1/4 3:26
2017/1/4	2:50:06	2017/1/4 2:50
2017/1/4	2:13:54	2017/1/4 2:13
2017/1/4	2:05:26	2017/1/4 2:05
2017/1/4	1:41:36	2017/1/4 1:41
2017/1/4	1:03:49	2017/1/4 1:03
2017/1/4	0:54:06	2017/1/4 0:54
2017/1/4	0:48:37	2017/1/4 0:48
2017/1/4	0:17:08	2017/1/4 0:17
2017/1/4	0:01:39	2017/1/4 0:01
2017/1/3	23:53:37	2017/1/3 23:53
2017/1/3	23:50:24	2017/1/3 23:50

図 7.11 日付と時刻の統合結果

### 7.6.3 番号の振り分け

json 形式では日付の新しい順でデータが並べられているので,ブックマーク数の増加の 推移を可視化するにあたり昇順でデータのソートを行う.

ソート後,時間データに番号をふる.

以下の画像は日付データのソート結果と番号の振り分けの結果である.

2017/1/3 9:39	1
2017/1/3 9:49	2
2017/1/3 10:06	3
2017/1/3 10:15	4
2017/1/3 10:18	5
2017/1/3 10:22	6
2017/1/3 10:24	7
2017/1/3 10:29	8
2017/1/3 10:34	9
2017/1/3 10:36	10
2017/1/3 10:37	11
2017/1/3 10:39	12
2017/1/3 10:40	13
2017/1/3 10:47	14
2017/1/3 10:54	15
2017/1/3 11:08	16
2017/1/3 11:08	17
2017/1/3 11:08	18
2017/1/3 11:21	19
2017/1/3 11:28	20
2017/1/3 11:33	21
2017/1/3 11:40	22
2017/1/3 11:59	23
2017/1/3 12:05	24
2017/1/3 12:06	25
2017/1/3 12:15	26
2017/1/3 12:21	27
2017/1/3 12:21	28
2017/1/3 12:21	29
2017/1/3 12:32	30
2017/1/3 12:37	31
2017/1/3 12:40	32
2017/1/3 13:09	33
2017/1/3 13:27	34
2017/1/3 13:33	35
2017/1/3 13:34	36

図 7.12 データのソート結果及び番号の振り分け

### 7.7 時間の推移とブックマーク数の増加の推移の可視化

縦軸にブックマーク数と横軸に日付データを設定し,時間の推移におけるブックマーク数の増加の傾向に関して散布図を作成し可視化する.

以下の画像は時間の推移におけるブックマーク数の増加の傾向を可視化した結果の例である.

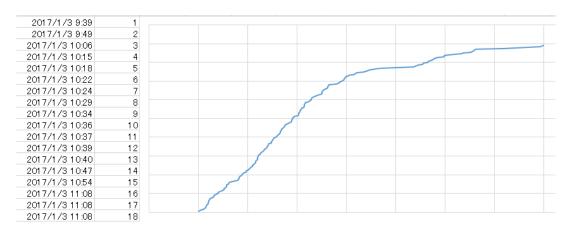


図 7.13 時間の推移とブックマーク数の増加の推移の可視化の結果

# 7.8 時間経過によるブックマーク数の増加

初めてエントリがブックマークがされてから 6 時間ごとのブックマーク数増加数を算出する.

	揭示日時	book	6	12	18	24
1	2017/1/3 9:39	178	52	127	154	167
2	2017/1/5 1:51	545	74	443	507	531
3	2017/1/3 12:22	210	25	94	124	167
4	2017/1/2 19:55	542	9	19	115	270
5	2017/1/2 18:20	703	22	93	308	491
6	2017/1/5 4:53	162	49	123	151	157
- 7	2017/1/2 17:57	143	4	5	11	49
8	2017/1/3 18:30	185	42	81	139	175
9	2017/1/5 4:53	162	49	126	150	157
10	2017/1/5 17:00	337	72	143	243	300
11	2016/12/13 5:18	886	224	551	726	767
12	2016/12/13 2:19	565	66	374	474	536
13	2016/12/13 1:00	525	8	175	338	438
14	2016/12/13 0:49	325	18	174	271	300
15	2016/12/12 8:14	103	40	73	91	93
16	2016/12/12 11:11	742	12	336	468	617
17	2016/12/24 20:49	301	88	145	242	275
18	2016/12/11 23:39	275	18	77	187	228
19	2013/6/16 12:07	82	37	62	67	75
20	2016/12/13 7:13	174	105	147	163	165

図 7.14 時間経過によるエントリのブックマークの増加数

# 7.9 時刻ごとのブックマーク数の増加

初めてエントリがブックマークがされてから 24 時間以内を対象に 1 時間ごとの時刻のブックマーク数の増加数を算出する.

### 7.9.1 時間データの変換

各時刻ごとの増加数を求めるため時間データを処理できる数値に変換する.

数値に変換した際,自然数部分が日付にあたり,小数点以下が時刻を表している.ここでは各時刻ごとの増加数を求めるため,小数点以下の数値のみを抽出する.

時刻データの抽出は INT 関数を使用する.以下は数値データへの変換と時刻データの抽出の結果である.左部が元データ,中央部が数値変換の結果,右部が抽出した時刻データである.

2017/1/3 9:39	42738.402	0.402
2017/1/3 9:49	42738.409	0.409
2017/1/3 10:06	42738.421	0.421
2017/1/3 10:15	42738.427	0.427
2017/1/3 10:18	42738.429	0.429
2017/1/3 10:22	42738.432	0.432
2017/1/3 10:24	42738.433	0.433
2017/1/3 10:29	42738.437	0.437
2017/1/3 10:34	42738.440	0.440
2017/1/3 10:36	42738.442	0.442
2017/1/3 10:37	42738.442	0.442
2017/1/3 10:39	42738.444	0.444
2017/1/3 10:40	42738.444	0.444
2017/1/3 10:47	42738.449	0.449
2017/1/3 10:54	42738.454	0.454
2017/1/3 11:08	42738.464	0.464
2017/1/3 11:08	42738.464	0.464
2017/1/3 11:08	42738.464	0.464
2017/1/3 11:21	42738.473	0.473
2017/1/3 11:28		0.478
2017/1/3 11:33	42738.481	0.481
2017/1/3 11:40		0.486
2017/1/3 11:59		0.499
2017/1/3 12:05		0.503
2017/1/3 12:06		0.504
2017/1/3 12:15		0.510
2017/1/3 12:21		0.515
2017/1/3 12:21		0.515
2017/1/3 12:21		0.515
2017/1/3 12:32		0.522
2017/1/3 12:37		0.526
2017/1/3 12:40		0.528
2017/1/3 13:09		0.548
2017/1/3 13:27		0.560
2017/1/3 13:33		0.565
2017/1/3 13:34	42738.565	0.565

図 7.15 数値への変換と時刻データの抽出

#### 7.9.2 時刻ごとの増加数のカウント

次に時刻ごとのブックマーク数の増加数を算出する.

1日を 1 として扱い , 24 時間ごとの数値にすると 1 時間当たり 0.04166... となる . よって各時刻の増加数を算出するには COUNTIFS 関数を使用し , 各時刻ごとに条件付けし , 該当する数をカウントする .

各時刻ごとの条件は対象の時刻をxとし、0.041667にxをかけ、0.041667x以上 0.041667(x+1)未満に該当する数とすることで対象の時刻の増加数が算出できる.

=COUNTIFS(\$I\$1:\$I\$1500,">="&L2,\$I\$1:\$I\$1500,"<"&M2)

図 7.16 countifs 関数

以下はあるエントリの各時刻ごとのブックマークの増加数である. 左部は時刻,中央部は条件数値,右部は増加数である.

0.041667	2
0.083333	3
0.125	1
0.166667	0
0.208333	0
0.25	3
0.291667	4
0.333333	4
0.375	4
0.416667	13
0.458333	8
0.5	9
0.541667	9
0.583333	9
0.625	16
0.666667	13
0.708333	12
0.75	11
0.791667	14
0.833333	7
0.875	8
0.916667	5
0.958333	8
1	4
	0.083333 0.125 0.166667 0.208333 0.25 0.291667 0.333333 0.375 0.416667 0.458333 0.5 0.541667 0.583333 0.625 0.666667 0.708333 0.75 0.708333 0.75 0.791667 0.833333 0.875 0.916667 0.958333

図 7.17 各時刻ごとの条件数値と増加数

# 7.10 分析

時間の推移に伴うブックマーク数の増加について可視化を行い,その増加の傾向からパターン分けを行う.

# 第8章

# 結果

100 件のデータに対し, 3 種類の分析を行った.

# 8.1 時間の推移に伴うブックマーク数の増加の傾向のパターン

時間の推移に伴うブックマーク数の増加について可視化を行ったところ大きく分けて 3 つのパターンに分類することができた.

以下に分類した推移のパターンと解説を記す.

#### 8.1.1 平均型

安定してブックマーク数が伸び続ける.急な上昇や大きな衰退もなく,ほぼ直線に近い 推移を見せる.

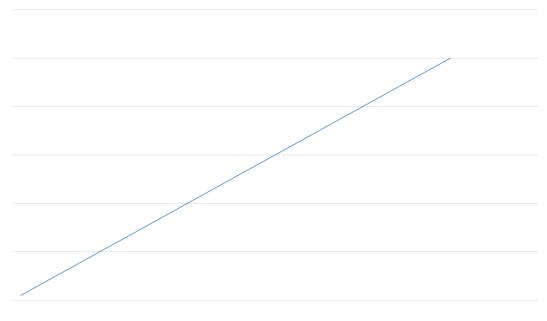


図 8.1 平均型

### 8.1.2 急上昇型

初期はあまり数値が伸びず停滞を続けるが,ある一点から数値が急速に伸び,上昇する.

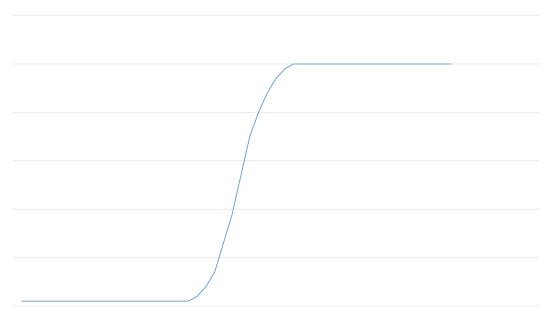


図 8.2 急上昇型

### 8.1.3 不規則型

数値が上昇と停滞を繰り返し,波のような曲線を描く.

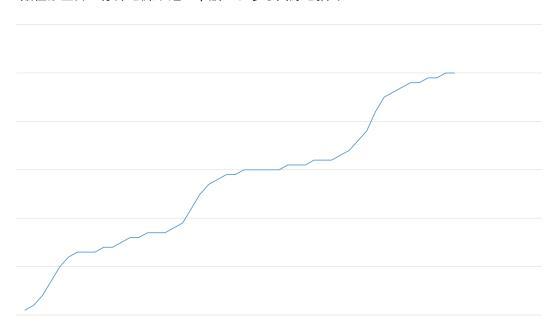


図 8.3 不規則型

# 8.2 各時刻ごとの増加数

初めてエントリがブックマークがされてから 24 時間以内を対象に 1 時間ごとの時刻の ブックマークの増加数を算出し,平均を求めた結果以下の表のような結果が得られた.

表 8.1 時刻ごとの増加数の平均

時刻	増加数平均
1	5.76
2	4.53
3	2.41
4	1.89
5	2.97
6	4.66
7	9.74
8	15.5
9	17
10	15.7
11	14.9
12	21.1
13	15.7
14	12.1
15	12.2
16	12
17	12.1
18	13.4
19	13.4
20	12.7
21	11.7
22	11.5
23	10.1
24	8.15

# 8.3 時間経過による増加数の変化

初めてエントリがブックマークがされてから 6 時間ごとのブックマーク数増加数を算出した. 初めてのブックマークから 24 時間後のブックマーク数を 100% とし, 各時間帯の増加率を求める. その結果, 以下の表のような結果となった.

表 8.2 時刻の経過と増加率の平均

時刻の経過 増加率平均 6 時間経過 22.39% 12 時間経過 32.741% 18 時間経過 27.854% 24 時間経過 17.02%

以下のグラフは時刻の経過と増加率の平均を可視化したものである.

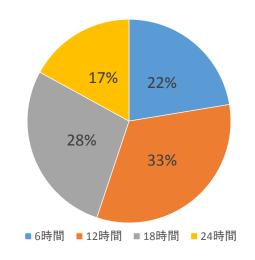


図 8.4 時刻の経過と増加率の平均のグラフ

以下は時間の経過とブックマーク数と増加の例である.

	掲載日時	bookmark	6	12	18	24
1	2017/1/3 9:39	178	52	127	154	167
2	2017/1/5 1:51	545	74	443	507	531
3	2017/1/3 12:22	210	25	94	124	167
4	2017/1/2 19:55	542	9	19	115	270
5	2017/1/2 18:20	703	22	93	308	491
6	2017/1/5 4:53	162	49	123	151	157
7	2017/1/2 17:57	143	4	5	11	49
8	2017/1/3 18:30	185	42	81	139	175
9	2017/1/5 4:53	162	49	126	150	157
10	2017/1/5 17:00	337	72	143	243	300
11	2016/12/13 5:18	886	224	551	726	767
12	2016/12/13 2:19	565	66	374	474	536
13	2016/12/13 1:00	525	8	175	338	438
14	2016/12/13 0:49	325	18	174	271	300
15	2016/12/12 8:14	103	40	73	91	93
16	2016/12/12 11:11	742	12	336	468	617
17	2016/12/24 20:49	301	88	145	242	275
18	2016/12/11 23:39	275	18	77	187	228
19	2013/6/16 12:07	82	37	62	67	75
20	2016/12/13 7:13	174	105	147	163	165
21	2016/12/13 7:21	348	135	243	308	310
_22	2016/12/12 22:35	289	10	50	176	243
_23	2016/12/16 5:50	625	169	405	506	530
24	2016/12/12 14:54	269	5	26	98	220
25	2016/12/12 20:22	214	28	78	172	191
26	2016/12/13 1:19	153	18	85	131	143
_27	2016/12/29 6:56	105	27	58	85	94
_28	2016/12/3 0:20	405	8	42	197	294
_ 29	2016/12/12 17:13	441	13	63	216	331
30	2016/12/25 20:17	279	60	146	202	228
31	2016/12/12 23:30	289	11	80	176	233
_32	2016/12/13 8:28	116	76	100	107	107
_33	2016/12/11 20:29	195	12	29	140	161
34	2016/12/12 15:35	384	98	225	273	339
_ 35	2016/12/13 6:25	212	41	132	185	189
36	2016/12/13 7:13	83	47	68	76	78

図 8.5 時間の経過とブックマーク数

# 第9章

# 考察

本章では分析結果をもとに考察を行っていく.

### 9.1 ブックマークの伸びやすい時間帯についての考察

分析結果からブックマーク数が伸びやすい時間帯は8時から13時まで伸びやすい時間帯であることが分かった.特に12時が最もブックマークの増加数が大きい時刻であった.朝起床してから新着のエントリを確認する人や昼の業務の合間の休憩時間にエントリを見る人が多いためだと推測される.

逆にブックマークの伸びにくい時間帯は2時から6時までであり,その中でも4時が最もブックマークの伸びにくい時刻であった.活動する人口が少ない深夜であることと,相対的にエントリをブックマークする人が少ないためエントリが共有されにくいことも要因と考えられる.

以下の画像は各時刻のブックマークの増加数の平均を可視化したものである.

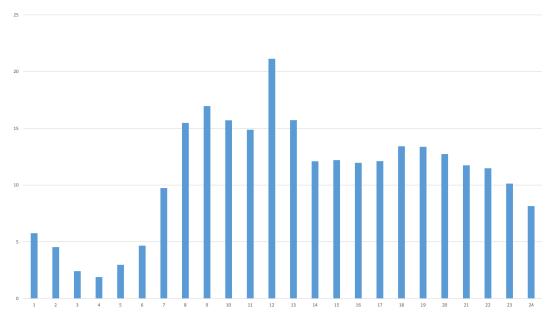


図 9.1 各時刻のブックマークの増加数の平均

## 9.2 時間の推移に伴うブックマーク数の増加の傾向の考察

3種類のグラフのパターンについて考察する.

### 9.2.1 平均型

この傾向は短期に急激にブックマーク数を伸ばしているものが多い.また日中にブックマークを伸ばし始めたものが多く,伸びにくい深夜に入る前の日中のうちにブックマーク数が伸びきってしまうためだと思われる.そのため停滞する区間がなく,変化の少ない曲線を描く.

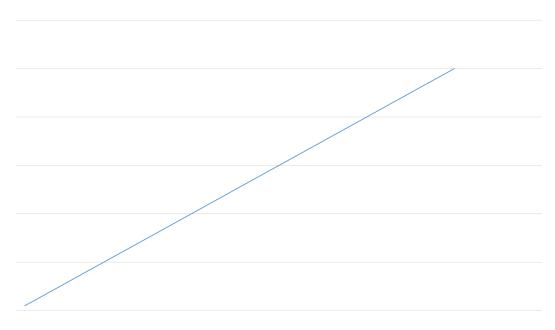


図 9.2 平均型

#### 9.2.2 急上昇型

この傾向は短期的にブックマーク数を伸ばしているもう一つの傾向であるが,平均型と 異なり掲載日時が深夜のものが多い.深夜のためブックマーク数が伸びにくく初期の増加 が停滞するが,人が活動を始める朝以降になるとブックマーク数を急速に伸ばし,結果急 激に伸びる曲線を描いている.

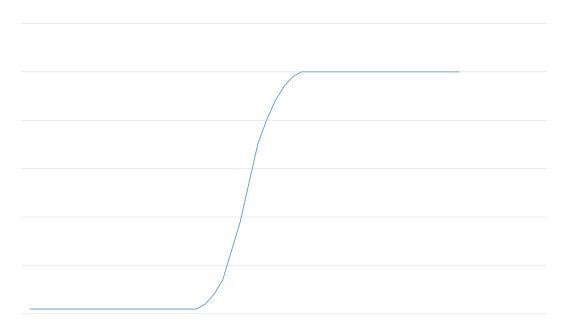


図 9.3 急上昇型

### 9.2.3 不規則型

この傾向は長期的にブックマーク数を伸ばしているものに多い、長期にわたってブックマークを伸ばしているためその期間中にブックマーク数の伸びにくい深夜を数度挟んでいる、その結果停滞と上昇を繰り返し、波のような曲線を描いている。

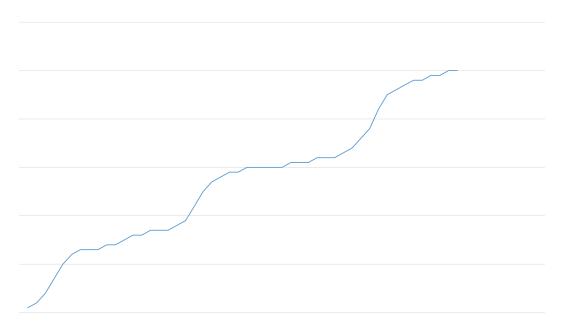


図 9.4 不規則型

### 9.3 時間経過による増加数の変化の考察

分析の結果からそれぞれの時間帯に大きな差はないものの,傾向として6時間以上18時間未満に多く増加する傾向があることが分かった.特に6時間以上12時間未満での増加率が最も大きいことが分かった.

逆に 18 時間を超えると増加率は最も下がり,また差が小さいとはいえ初めの 6 時間も増加率が小さいことが分かった.

初めてのブックマークからあまり時間がたたない間はそこまで数が伸びにくく,ブックマークが本格的につき始めるためには数時間を要することが多いため,6時間以上12時間未満の時間帯の増加率が大きいのだと予測される.

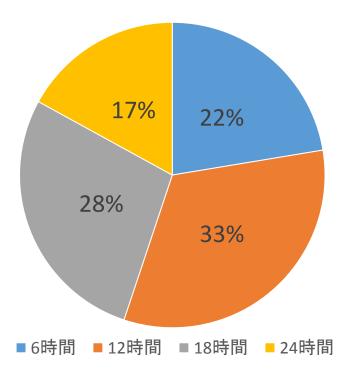


図 9.5 時刻の経過と増加率の平均のグラフ

### 9.4 3種の分析結果からの考察

三種の分析結果から時間帯がブックマークの増加数に最も顕著に影響が出ていることが 分かった.また時間の推移に伴うブックマーク数の増加の傾向についての可視化の結果と 初回のブックマークからの時間経過による増加率から考察を行った.

その結果,朝の5時から7時程度までの間に初回のブックマークがつくようにすれば, 昼には時間経過による増加率がピークを迎え,増加率の伸びにくい深夜に時間経過による 増加率の少ない時間帯を重ねることができる.

# 第 10 章

# 結論

本研究ではウェブマーケティングの例としてソーシャルブックマーケティングサービスを対象に,時間の推移に伴うブックマーク数の増加の傾向の分析を行い,ウェブマーケティングを行う上での指標の作成した.

分析の結果から時間帯によって大きくブックマーク数の増加の傾向が異なることが分かった.多く閲覧が見込める時間帯を意識しウェブマーケティングを行うことで,効果的なウェブマーケティングを行える確率が高まると思われる.

今回は時間とブックマーク数に着目し研究を行ったが,取得したが分析を行えなかった他の要素に関しても調査と分析を行い,それらの結果を組み合わせ多面的に考察を行うことでより精度の高いウェブマーケティングの予測と効果的なウェブマーケティングの手法の提案を行えることが期待できる.

# 参考文献

- [1] 株式会社はてな. 沿革 株式会社はてな. http://hatenacorp.jp/information/history.
- [2] 株式会社はてな. はてなブックマーク 3 つの特徴. http://b.hatena.ne.jp/guide.
- [3] 株式会社はてな. はてなブックマーク 3 つの特徴. http://b.hatena.ne.jp/help/entry/feature.
- [4] 株式会社はてな. 会社情報 株式会社はてな. http://hatenacorp.jp/information/.
- [5] Oracle. Oracle vm virtualbox. https://www.virtualbox.org/.
- [6] 株式会社はてな. はてなブックマークの api. http://b.hatena.ne.jp/help/entry/api.
- [7] 株式会社はてな. はてなブックマークエントリー情報取得 api. http://developer.hatena.ne.jp/ja/documents/bookmark/apis/getinfo.