

平成 24 年度

慶應義塾大学

博士論文

時間情報を利用した長期的に利用する
情報へのアクセスに関する研究

政策・メディア研究科

上野 大樹

2012 年 1 月

時間情報を利用した長期的に利用する 情報へのアクセスに関する研究

論文要旨

本研究では、長期的に利用される情報に特化した情報検索手法と情報アクセス手法の提案を行う。ここでは、このような手法を「長期的検索」と呼ぶ。また、長期的検索を行うために、情報がどれだけ長期的に利用されているかの指標として、「長期度」という指標を提案する。

情報には様々な利用のされ方がある。例えば、ある一定の短い期間でしか利用されないもの、長期的に利用されるもの、周期的に利用されるものなどである。このような利用のされ方によって、情報の種類や今後の有用性などがある程度予測できる。ここで筆者は、長期的に利用されている情報は、これまで長期間にわたって利用されてきたという実績があるため、今後も長期的に利用される可能性が高いと考えた。そのため、こういった情報の有用である可能性が高いのではないかと考えた。情報の検索手法として、人気の情報や今流行りの情報を検索する手法は数多く提案され、システムとして実現されているが、長期的に利用される情報に特化して検索する手法やシステムはほとんど存在しない。そのため、現状では、長期的に利用されている有用な情報でも、特别人気が高くない場合、見つけることが難しい場合もある。そこで本研究では、情報検索の新しいビジョンとして、長期的検索という概念を提唱し、こういった情報を見つける手法を提案する。

長期的に利用される情報の具体的な例として、「Web ページ」、「検索キーワード」、「ファイル」という 3 つの例に焦点を当てて、「セレクトブックマ」、「MasteryEye」、「タームストア」という 3 つのシステムを開発した。

セレクトブックマは、長期的に利用され続ける Web ページに特化して検索するためのシステムである。ソーシャルブックマークのブックマーク時間情報を利用することによって、その Web ページがどのように利用、ブックマークされているかを分析し、長期的な情報を発見している。セレクトブックマの画面上でキーワードを入力して検索を行うことによって、入力したキーワードに関連した Web ページの中で、長期的に利用されている Web ページを検索することができる。例えば、体系的だった知識を得られる Web ページ、Web サービス、リファレンスなどに特化して検索することを目的としている。

MasteryEye は、長期的に検索され続ける検索キーワードを見つけるためのシステムである。Google 検索における検索回数の時系列データを利用することによって、このような検索キーワードを発見する。MasteryEye の画面上で入力したキーワードに関連するものの中で、長期的に検索され続けている検索キーワードを提示する。ある分野について新しく学びたい場合に、定番の検索キーワードを手軽に取得することを目的としている。

タームストアは、システムの利用者がこれまでに長期的にアクセスしてきたファイルと Web ページを提示するシステムである。タームストアでは、ユーザがこれまでアクセスしてきたファイルと Web アクセスの履歴に基づき、統合的にファイルや Web ページを提示する。また、ファイルや Web ページの管理を必要とせずに、これらの情報に手軽にアクセスできるようにすることを目的としている。

これら 3 つのシステムを試作し、分析および評価実験を行った。本論文では、分析および評価実験の結果を示し、長期的検索の意義や有効性についても議論する。

キーワード

情報検索、情報フィルタリング、情報レコメンデーション、長期的利用、ソーシャルブックマーク、検索キーワード、ファイルアクセス、アクセス履歴

慶應義塾大学 大学院政策・メディア研究科

上野 大樹

Abstract of Doctor's Thesis Academic Year 2012

Research on access to Long Term Information Focusing Time Scale

Summary

In this research, we proposed information retrieval method and information access method that find information used for a long-term. We call those method "Long-term Retrieval". Moreover, we proposed "Degree of Long-term" that is an index of which information was used for a long-term.

Key Word

Information Retrieval, Information Filtering, Information Recommendation, Long Term Usage, Social Bookmark, Retrieval Keyword, File Access, Access History

Keio University Graduate School of Media and Governance

Taiki Ueno

目 次

第 1 章	序論	1
1.1	研究の背景・動機	2
1.2	研究の目的	3
1.3	用語定義	3
1.4	論文の構成	5
第 2 章	背景と関連領域	7
2.1	はじめに	8
2.2	情報検索	8
2.3	情報フィルタリング	9
2.4	まとめ	9
第 3 章	長期的情報検索の提案	11
3.1	長期度	12
3.1.1	長期度とは	12
3.2	長期的な情報	12
3.2.1	長期的な情報の定義	12
3.2.2	長期的な情報と短期的な情報の特徴	12
3.2.3	本研究で対象とする情報	12
3.2.4	長期的な情報の具体例	14
3.2.5	長期的な期間について	15
3.3	長期的検索の提案	16
第 4 章	長期的な Web ページの検索	17
4.1	背景	18
4.2	分析	18
4.2.1	ソーシャルブックマークデータの分析	18
4.2.2	時間情報に関する分析	20
4.3	セレクトブックマの提案	23
4.3.1	セレクトブックマの概要	23
4.3.2	セレクトブックマの設計思想	23
4.3.3	セレクトブックマ画面構成	25
4.3.4	セレクトブックマインタフェース	25
4.3.5	検索ランキングロジック	26
4.3.6	システム構成	26
4.4	評価実験	26

4.4.1	実験概要	26
4.4.2	実験目的	28
4.4.3	実験方法	28
4.4.4	実験結果	29
4.4.5	時間情報を利用する効果について	44
4.4.6	各実験結果からの考察	44
4.5	考察と課題	46
4.5.1	実験結果からの考察	46
4.5.2	時間情報を利用することの有効性	46
4.5.3	本研究の有効性	46
4.5.4	課題	47
4.6	まとめ	47
第 5 章	長期的な検索キーワードの提示	49
5.1	背景	50
5.1.1	背景	50
5.1.2	目的	50
5.2	MasteryEye の提案	51
5.2.1	長期度の計算手法	51
5.2.2	関連キーワードの収集	52
5.2.3	検索キーワードの時系列データ収集	53
5.2.4	システム表示	53
5.3	評価実験	54
5.3.1	実験目的	54
5.3.2	実験方法	54
5.3.3	実験結果	55
5.4	課題と展望	57
5.4.1	課題	57
5.4.2	今後の展望	59
5.5	まとめ	59
第 6 章	長期的なファイルと Web 履歴へのアクセス	60
6.1	背景	61
6.2	長期的に利用されるファイルへのアクセス手法の提案	61
6.3	評価実験	61
6.4	考察	61
6.5	まとめ	61

第7章 関連研究	62
7.1 情報検索	63
7.1.1 情報検索のアルゴリズム	63
7.1.2 関連ページの検索	63
7.1.3 ソーシャルな検索	64
7.1.4 検索インタフェース	64
7.1.5 様々な情報を利用した検索システム	65
7.2 情報フィルタリング	66
7.2.1 情報フィルタリングのアルゴリズム	66
7.2.2 ソーシャルな情報フィルタリング	66
7.2.3 情報フィルタリングのインタフェース	66
7.2.4 様々な情報フィルタリングシステム	66
7.3 ソーシャルブックマークデータの分析	67
7.4 検索キーワード	67
7.5 ファイルアクセスに関する研究	68
7.6 時間情報の利用	68
第8章 長期的検索の考察と展望	70
8.1 考察	71
8.1.1 人の行動から得られる時間情報の分析	71
8.1.2 長期度の提案	71
8.1.3 長期的な情報に着目した検索システム	72
8.1.4 長期的な情報の有用性	72
8.2 議論	73
8.2.1 長期的な情報の特徴	73
8.2.2 長期的検索の有効性	73
8.2.3 長期的検索の未来	73
8.3 課題	73
8.4 展望	74
第9章 結論	75
9.1 研究の成果	76
9.2 総括	76
謝辞	100
研究業績	101

参考文献	102
付録 A	110
付録 B	110

目 次

1.1	本論文の構成	6
2.1	ターゲットとする情報	10
3.1	対象とする情報	13
3.2	対象とする情報と提案システム	14
4.1	はてなブックマーク	19
4.2	一時的にブックマークされ、その後ほとんどブックマークされなくなるタイプのページ	20
4.3	一時的に大量にブックマークされ、その後も長い間ブックマークされ続けるタイプのページ	21
4.4	大量にブックマークされる時期はないが、長い間ブックマークされ続けるタイプのページ	21
4.5	Type1 の Web ページの種類	22
4.6	Type2 の Web ページの種類	22
4.7	セレクトブックマ画面	24
4.8	システム構成図	27
4.9	DB のテーブル構成図	27
4.10	3人以上を適合文書とした場合の適合率 (java)	40
4.11	8点以上を適合文書とした場合の適合率 (java)	40
4.12	3人以上を適合文書とした場合の適合率 (健康)	41
4.13	8点以上を適合文書とした場合の適合率 (健康)	41
4.14	3人以上を適合文書とした場合の適合率 (映画)	42
4.15	8点以上を適合文書とした場合の適合率 (映画)	42
5.1	べき乗則のデータと長期的利用データ	52
5.2	「yahoo」での検索結果画面	53

表 目 次

3.1	長期的な情報と短期的な情報の特徴	13
3.2	長期的と短期的の期間	16
4.1	収集したデータ量	19
4.2	「java」で検索した場合の検索結果上位 10 件	30
4.3	「健康」で検索した場合の検索結果上位 10 件	31
4.4	「映画」で検索した場合の検索結果上位 10 件	32
4.5	「java」で検索した場合の検索結果の中から被験者が選んだ Web ページ（人数）	33
4.6	「java」で検索した場合の検索結果の中から被験者が選んだ Web ページ（得点）	34
4.7	「健康」で検索した場合の検索結果の中から被験者が選んだ Web ページ（人数）	35
4.8	「健康」で検索した場合の検索結果の中から被験者が選んだ Web ページ（得点）	36
4.9	「映画」で検索した場合の検索結果の中から被験者が選んだ Web ページ（人数）	37
4.10	「映画」で検索した場合の検索結果の中から被験者が選んだ Web ページ（得点）	38
4.11	全適合文書の件数	39
4.12	再現率（java）	40
4.13	再現率（健康）	41
4.14	再現率（映画）	42
4.15	「java」での検索結果の得点の合計値	43
4.16	「健康」での検索結果の得点の合計値	43
4.17	「映画」での検索結果の得点の合計値	44
5.1	関連キーワードデータベースのデータ量	53
5.2	被験者属性	55
5.3	選択されたキーワードの数	56
5.4	カイ二乗検定（知っているかどうか）	56
5.5	カイ二乗検定（長期間利用かどうか）	56
5.6	カイ二乗検定（重要かどうか）	57
5.7	長期間利用かどうか：長期度高と長期度低の比較（カイ二乗検定）	57
5.8	長期間利用と重要なものの関係（カイ二乗検定）	57
5.9	取得した関連キーワード一覧	58

第1章 序論

概要

本章では，本研究の背景・目的および論文の構成について示す．

1.1. 研究の背景・動機

近年、情報の短命化が進んできている [1]。その理由として、アクセス可能な情報量の急激な増大と情報伝播速度の速いサービスの一般化があげられる。近年、電子機器の記憶容量の増大により、Web 上や個人の PC 上の情報量は急激に増加しており、このように大量に記録された情報、特に Web 上の情報は、玉石混交であると言われている [2]。そのため、あまり利用されない情報も大量に存在する。また、マイクロブログ¹のような情報伝播速度が速いサービスも一般的に利用されるようになってきた。このようなサービスの一般化やネットワーク速度の向上などによって、情報が流行する速度、さらには情報が廃れる速度も速くなってきた [3]。例えば、代表的なマイクロブログサービスである Twitter[?] の情報などは、急速に流行して、短い期間に多くの人から見られるが、急速に廃れて 1~2 週間経過するとほとんど見られなくなるような情報も多い [4]。こういったことから、寿命の短い情報やデジタルコンテンツがだんだんと増加してきている。

このように寿命が短く短期的にしか利用しないような情報が増えてきたため、これまでの情報検索システムや情報提示システムでは、長期的に利用可能な情報を見つけることが難しくなってきた。これまでの情報検索システムや情報提示システムでは、特に長期的な情報に焦点をあてていないため、短期的な情報が増大した場合、短期的な情報が提示される確率があがる。こういった背景に対して、長期的に利用する情報に特化して取得するようなシステムはほとんど存在しない。そこで、本研究では、長期的に利用可能な情報を見つけようと思ったときに、そういった情報を見つけることが容易でないという課題がある。

ところで、本や商品に関しては、ベストセラーとロングセラー [84] という指標がある。ロングセラーと長期的に利用されるということは、「売れる」と「利用される」の違いはあるものの、類似した指標であると考えられる。そこでロングセラーの有用性を例として、長期的に利用される情報の有用性について考える。本や商品に関して言えば、ベストセラーを好む人もいれば、ロングセラーを好む人もいる。以上のような特徴がある一方、Web 上の情報はベストセラーといったような観点か、他人からの推薦といったような観点からの評価が主で、ロングセラーという観点での評価はほとんどされていない。Web の歴史はまだ浅く、商品の売れ方と Web では異なるところも多いが、Web 上の情報についてもベストセラーのような人気の高い情報を好む人もいれば、ロングセラーのように長期的に利用される情報を好む人もいる。また、どちらの情報が欲しいかは状況によって変わってくる。そのため、本研究では、Web 上の情報に関しても、長期的に利用されているかどうかという指標も有用であると考えた。

そこで、本研究では、長期的に利用される情報という視点に着目し、長期的に利用できる情報を見つけるための手法「長期的検索」を提唱する。このような情報を取得するためには、今までよく利用されていたリンク数、アクセス数、評価数、関連度などの指標だけではなく、時間情報を利用し、情報がどれだけの期間利用されてきたかという新しい評価指標が必要である。本研究では、この指標のことを「長期度」と呼ぶ。

1.2. 研究の目的

本研究では、主に以下の3点を目的とする。

- 検索者が長期的に利用できる情報を手軽に見つけられるようにすること
- 長期的に利用されている情報の有用性が高いことを示すこと
- 本研究で提案する長期的な情報を検索するシステムの有効性を示すこと

本研究の一番の目的は、自分が長期的に利用できる情報を手軽に見つけられるようにすることである。その理由は、長期的に利用できる情報は有用である可能性が高いにもかかわらず、見つけるための手法が確立されておらず、発見することが容易ではないからである。

次に、本研究では、長期的に利用されている情報が有用である可能性が高いことを示す必要がある。そのために、これまで長期的に利用されてきた情報は、今後も長期的に利用できるかどうかを調査する必要がある。今後も長期的に利用できるということは、有用性が高いということが言える。

さらに、実際に提案するシステムが有効であるかどうかを示す必要もある。このために、他の検索システムなどと比較して、提案システムの優れている点を明確にする必要がある。

1.3. 用語定義

本論文で使用するいくつかの用語に対して、用語の意味を説明する。以下に示すのは、本論文で利用する際の用語の意味である。

*1 マイクロブログ

主に140文字程度の短い文章を書いてWebに公開する短いブログ。代表的なマイクロブログサービスとしては、Twitterがあげられる。ブログとの違いは、手軽である点とユーザー同士のコミュニケーションを支援するソーシャルネットワーク的な側面が強いところである。

*3 長期度

その情報がどれだけ長期的に利用されているかを表す指標。

*4 情報フィルタリング [8]

大量の情報の中から、ユーザーにとって必要な情報を取り出し、不要な情報を除外する処理を自動的に行う技術のこと。本論文では、広義の意味として情報フィルタリングという用語を利用しており、情報の収集と排除双方の意味を含む。

*5 情報レコメンデーション [61]

ユーザの興味や嗜好に応じて、お勧めの情報を提供すること。

*5 パーソナライズド検索

パーソナライズド検索とは，検索ユーザの過去の検索履歴などから，ユーザの興味や関心を推定し，それぞれのユーザに合わせた検索結果を提示する検索手法のことである．

*6 folksonomy

Web 上のデータにおいて，ユーザ自らが情報の分類・収集を行うこと．「人々」(folks) と「分類」(taxonomy) とを掛け合わせた造語である．例えば，ソーシャルブックマークなどでは，ユーザがブックマークしたページに自由にタグを付与できるが，このユーザ自身が自由にタグをつける行為も folksonomy の一つである．タグをつけることにより，タグを付けた情報にアクセスするための検索や分類に役立てることを目的としている．

*7 体系だった知識を得られる Web ページ

ある分野に対して，基礎的な内容から応用的な内容までの幅広い知識を得られる Web ページ．また，ある分野に関する重要な知識がまとめて記述されているような Web ページ．

*8 RSS リーダー

指定した Web サイトの更新情報を一定時間ごとに自動的にダウンロードし，更新があると記事へのリンクを表示してユーザに知らせるツールのこと．

*9 ソーシャルブックマーク

インターネット上で自分のブックマークを不特定多数のユーザに公開し，有益な Web ページを共有する Web サービスのこと．

*11 適合率

検索結果として得られた文書中にどれだけ検索に適合した文書（適合文書）を含んでいるかという正確性の指標である．情報検索システムの評価を行う際に一般的に利用される指標であり，これとともに再現率^{*9}が利用される場合が多い．適合率は以下の数式で表される．

$$\text{適合率} = \frac{R}{N}$$

R : 検索された適合文書の数

N : 検索結果の文書の数

*12 再現率

検索対象としている文書の中で検索結果として適合している文書（適合文書）のうちで，どれだけの文書を検索できているかという網羅性の指標である．情報検索システムの評価を行う際に一般的に利用される指標であり，これとともに適合率^{*8}が利用される場合が多い．再現率は以下の数式で表される．

*13 バースト

爆発する，破裂する，急に起きる，勢い良くでるなどの意味を持つ．

***14 Google サジェスト**

Google 検索の補助機能で，ユーザが検索キーワードを入力しているときに，よく検索されるキーワードの候補を提示する機能である．

***15 スニペット**

スニペットとは，一般的には「切れ端」「断片」といったような意味の英語であるが，IT 用語としては，検索エンジンの検索結果の一部として表示される Web ページの要約文のことを表す．

***16 カイ二乗検定**

カイ二乗検定とは，帰無仮説が正しければ検定統計量がカイ二乗分布に従うような統計学的検定法の総称である．

$$\text{再現率} = \frac{R}{C}$$

R : 検索された適合文書の数

C : 全対象文書中の適合文書の数

1.4. 論文の構成

図 1.1 に本論文の構成図を示す．

第 2 章 背景と関連領域

第 2 章では，本研究の背景と関連領域について述べる．

第 3 章 長期的情報検索の提案

第 3 章では，本研究で述べている「長期的な情報」の定義と「長期間」の定義を行う．また，長期的な情報を検索するための手法を提案する．

第 4 章 長期的な Web ページの検索

第 4 章では，長期的な Web ページを検索するシステム「セレクトブックマ」について述べる．

第 5 章 長期的な検索キーワードの提示

第 5 章では，長期的な検索キーワードを提示するシステム「MasteryEye」について述べる．

第6章 長期的なファイルと Web 履歴へのアクセス

第6章では、自分が長期的に利用してきたファイルと Web ページへ手軽にアクセスするためのシステム「タームストア」について述べる。

第7章 関連研究

第7章では、本研究に関連する研究領域について整理し、本研究の特徴や位置づけについて述べる。

第8章 考察

第8章では、本研究の考察と展望について述べる。

第9章 結論

第9章では、本研究の成果をまとめ、本論文を総括する。

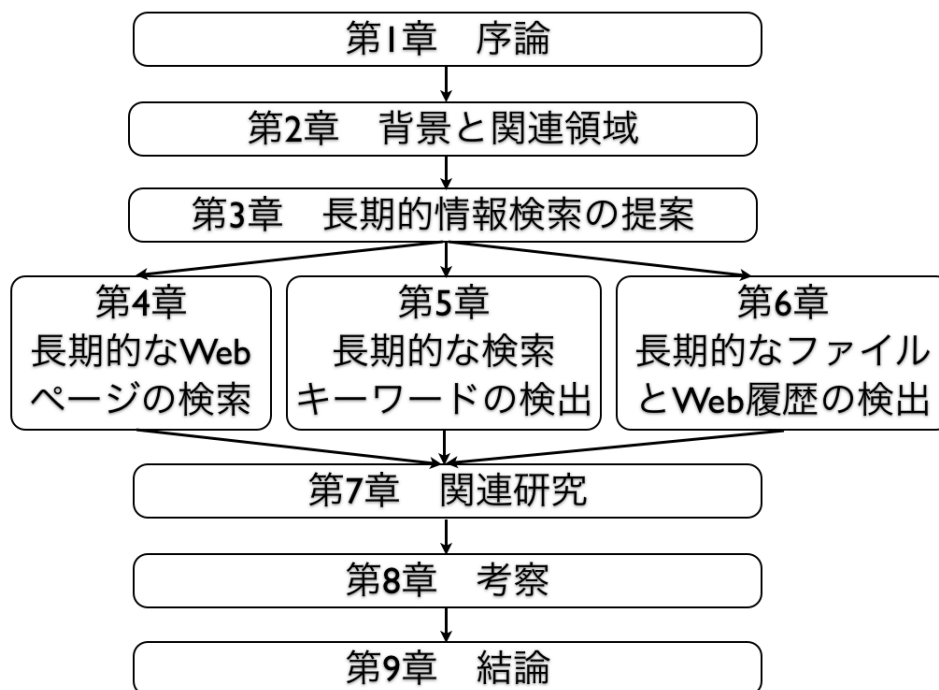


図 1.1: 本論文の構成

第2章 背景と関連領域

概要

本章では、本研究の背景と関連領域について述べる。

2.1. はじめに

近年、情報爆発により、膨大な量の情報にアクセスすることが可能になってきた。このような膨大な量の情報にアクセスするための手法として、様々な情報検索手法、情報フィルタリング^{*4}手法、情報提示手法が提案されてきた。ここでは、これらの情報検索手法や情報フィルタリング手法の発展と共に本研究の背景について述べる。

2.2. 情報検索

従来、Web 検索エンジンは、人手で検索結果を作成する、ディレクトリ型の検索エンジン [1] が主流であった。例えば、初期の頃の Yahoo の検索エンジンなどがこれにあたる。だが、Web ページの急激な増加に伴い、人手だけによる Web ページの選択・分類が難しくなってきた。そこで、近年では、ロボットが自動的に Web ページを巡回して、検索結果を生成するロボット型の検索エンジンが主流となっている [2]。例えば、現在の Google 検索 [22] や Bing[21] などの検索エンジンなどがこれにあたる。

だが、現在のロボット型の検索エンジンでは、検索結果が大量に取得されるため、検索キーワードの選択が重要であるが、その適切な選択は容易でないことも多い。また、検索者にとって不要な情報も大量に検索結果としてひっかかってしまい、検索結果の中から所望の Web ページを見つけるのも容易ではない場合がある。そのため、近年では、情報フィルタリングの重要性が問われている。

また、近年では、パーソナライズド検索^{*5}の研究も多くなされている。パーソナライズド検索とは、検索ユーザの過去の検索履歴などから、ユーザの興味や関心を推定し、それぞれのユーザに合わせた検索結果を提示する検索手法のことである。パーソナライズド検索の研究例として、例えば [53][54] があげられる。

一般的に、Web 検索の問題を解決するために、主に以下 2 種類のアプローチが取られている。

- アルゴリズムの改善による解決
- 人力利用による解決

近年では、アルゴリズムだけによる検索手法の提案ではなく、人力も利用した検索手法が注目されてきている [3]。その主な理由として、以下 2 つの理由があげられる。

1 つ目の理由は、現在成功している Web 情報検索システムや Web 情報フィルタリングシステムの多くが、人力を利用しているからである。例えば、Google 検索で利用されている PageRank[7] も被リンクを利用しているため、人力情報を存分に利用している。

2 つ目の理由は、近年、ソーシャルなシステムが広く普及してきたからである。このため、多くのユーザから発信された様々な情報を容易に取得できるようになってきた。このような情報を利用した研究例として、Twitter[?] の情報を利用した情報検索の研究 [4]、ソーシャルブックマー

ク^{*9}の情報を利用した情報検索の研究 [10]、レシピサービスの情報を利用した情報検索の研究 [11] などが幅広く行われている。

こういった理由から筆者は、機械的なアルゴリズムだけでなく人力も利用した検索手法の方が、有用な検索システムが作れる可能性が高いと考えている。そのため本研究では、機械的なアルゴリズムだけでなく、人力情報も利用したアプローチを行う。

2.3. 情報フィルタリング

情報フィルタリングとは、大量の情報の中から、ユーザにとって必要な情報を取り出し、不要な情報を排除する処理を自動的に行う技術のことである。情報検索と情報フィルタリングの違いは、検索質問が動的であるか静的であるかという点である。情報フィルタリングでは、ユーザの興味や関心などをプロファイルと呼ばれるファイルに記述しておき、これを用いて新規に生み出される情報を選別し、不要な情報を削除することが主な目的となる [72]。情報フィルタリングには、様々な手法が存在するが、情報推薦や収集に関しては、「内容に基づくフィルタリング」と「協調フィルタリング」という手法がよく利用される [13]。

情報フィルタリングや情報レコメンデーション^{*4}の技術を用いることによって、各ユーザにあった情報を見つけることを目的としている研究例は多い [94][109][63][?]。また、今流行の情報を見つけるための研究やシステムも多い [45][35]。また、人気の高い情報を見つけるシステムも多く存在する [12]。

これに対して、長期的に利用されている情報を見つけることを目的としている研究例はほとんど存在しない。このため、長期的に利用される情報に特化して見つけたり、収集したりすることは難しいのではないかと考えた。さらに、こういった情報の中で、あまり人気の高くない情報を探すが、特に難しいのではないかと考えた。こういった背景もあり、本研究では、長期的検索の提案をする。本研究でターゲットとする情報の位置づけを図 2.1 に示す。

2.4. まとめ

以上のような情報検索と情報フィルタリングの研究背景に基づき、本研究では長期的に利用される情報に焦点をあて、情報を見つけるための研究を行う。

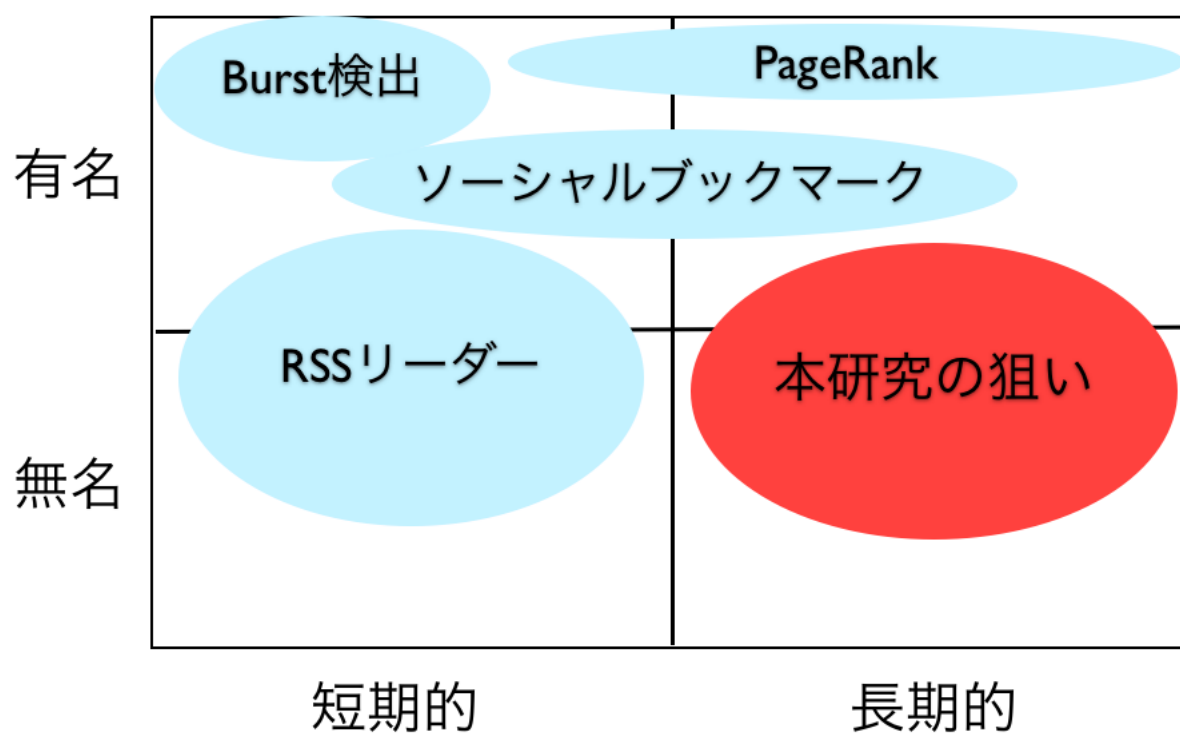


図 2.1: ターゲットとする情報

第3章 長期的情報検索の提案

概要

本章では、本研究で述べている「長期的な情報」の定義と「長期度」の定義を行う。また、長期的検索を提案する。

3.1. 長期度

本研究では、長期的な情報を取得するために、「長期度」という指標を提案する。ここではその長期度について述べる。

3.1.1. 長期度とは

本研究では、その情報がどれだけ長期的に利用されてきたかを表す指標として、「長期度」という指標を定義する。長期度が高い程、長期的に利用されている情報、長期度が低い程、短期的にしか利用されていない情報とする。本研究では、長期的に利用されている情報を取得するシステムを実装するため、この長期度が情報を評価するための主な軸となる。現状では、長期度の計算手法は、情報の種類によって異なる手法を用いている。その理由として、情報の種類によって、情報の利用のされ方や一般的な利用数や利用される期間などが大きく異なることがあげられる。長期度の具体的な計算手法については、第4章～第6章で述べる。

3.2. 長期的な情報

ここでは、本研究における長期的な情報の定義を示す。

3.2.1. 長期的な情報の定義

本研究で述べている長期的な情報とは、以下の2種類の意味を含む。

- (1) これまでに長期的に利用されてきた情報
- (2) 情報を取得した人が、今後長期的に利用できる情報

本研究では、(1)の情報は、(2)である可能性が高いという仮説を立てる。この仮説の元、(1)の情報を取得するシステムを開発し、システムの有効性と仮説の検証を行う。

3.2.2. 長期的な情報と短期的な情報の特徴

表3.1に、長期的な情報と短期的な情報の特徴をあげる。それぞれの情報は、以下のような傾向が強い。

3.2.3. 本研究で対象とする情報

世の中の様々な情報に対して、長期的に利用されている情報を取得することの有効性が考えられるが、本研究では、特に以下のような情報を取り上げて、研究を行う。

- Web ページ

表 3.1: 長期的な情報と短期的な情報の特徴

長期的な情報の特徴	短期的な情報の特徴
定番情報	流行情報
体系だった情報	断片的な情報
安定している	不安定である
じわじわと利用されるようになる	急速に利用されるようになる
流行り廃りが少ない	流行り廃りが激しい
熟練者から利用されやすい	初心者から利用されやすい

- 検索キーワード
- ファイル



図 3.1: 対象とする情報

本研究では、以下 3 点の特徴を持った情報として、これら 3 つの情報を取り上げた。

- (1) 膨大な量の情報に手軽にアクセスすることができる
- (2) 長期的に利用できる情報にアクセスすることが困難である
- (3) 利用履歴など、情報の利用のされ方に関するデータを入手できる

以上の特徴を持った情報を対象とした理由を述べる。そもそもアクセス可能な情報量が少ない場合、所望の情報にアクセスすることがあまり困難ではないため、(1) の特徴を選んだ。また、本研究では、長期的に利用できる情報へのアクセスが困難になってきているという問題意識があるため、(2) の特徴を選んだ。さらに、情報の利用履歴など、情報の利用のされ方に関するデータがないと、長期的に利用されているかどうか判定できないため、(3) の特徴を選んだ。

また、これらの情報の中から長期的に利用されるものを見つけるためのシステムとして、以下の図 3.2 のように、3 つのシステムを提案し、設計・実装した。これらのシステムについての詳細および Web ページ、検索キーワード、ファイルを選んだより詳細な理由に関しては、第 4 章～第 6 章で述べる。

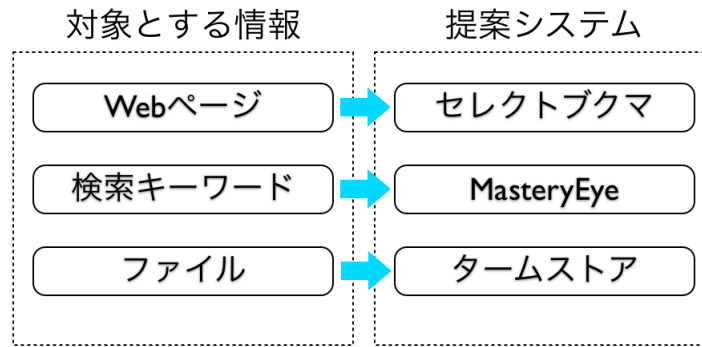


図 3.2: 対象とする情報と提案システム

3.2.4. 長期的な情報の具体例

本研究で述べている長期的な情報とは、具体的にどんな情報なのか、実際に具体例を述べる。

(1) Web ページ

長期的に利用する Web ページとして、以下のようなものなどが考えられる。

- まとめサイト
- リファレンス的なサイト
- Web サービス

まとめサイトとは、技術や趣味などについてまとめて情報が記載されているサイトのことである。例えば、OS やソフトウェアのショートカットキーをまとめたサイト、美味しいレストランをまとめたサイト、論文の書き方などをまとめたサイトなどがあげられる。

リファレンス的なサイトとは、文章やプログラムを書く際に、リファレンス的に利用するサイトである。例えば、プログラムを書くときに利用する各プログラミング言語のリファレンスやプレゼン資料を作るときやデザインをするときに利用する画像や写真の素材集などがあげられる。

Web サービスは、例えば、検索サービス、動画共有サービス、レシピ共有サービスなどがあげられる。

(2) 検索キーワード

長期的に利用する検索キーワードとして、以下のようなものなどが考えられる。

- 各分野における重要キーワード
- ロングセラーとなっている商品名

- 長期間利用されている Web サービス名

各分野における重要キーワードとは、その分野について詳しく調べるときに必要となるキーワードである。例えば、学問、技術、趣味など様々な分野に対して重要キーワードが存在する。

ロングセラーとなっている商品名は、例えば食品や日用品などの各分野でロングセラーになっている商品名である。

長期間利用されている Web サービス名は、例えば写真や動画など様々な共有サービスや Q & A サービスなどのサービス名である。例えば、「YouTube[26]」や「Flickr[27]」などのキーワードは、検索キーワードとしては、長期的に利用されているだろう。

(3) ファイル

長期的に利用するファイルとして、以下のようなものなどが考えられる。

- テンプレートファイル
- リファレンス的なファイル
- 画像などの素材

テンプレートファイルとは、仕事などで書類などを作成する際に、雛形となるファイルである。例えば、Word、Excel、PowerPoint などの雛形ファイルがあげられる。何かの申請をするときや何かフォーマットの決まった書類などを作成するときに利用されることが多い。

リファレンス的なファイルとは、何かファイルを作成するときに、何度も参照するファイルのことである。例えば、文章を書く参照する以前書いた文章ファイルやデータファイルなどがあげられる。

画像などの素材とは、例えばよくプレゼン資料や Web サービスに登録するアイコンなどの素材である。こういった素材の中には長期的に利用されるものも多いだろう。

3.2.5. 長期的な期間について

本研究では、長期的に利用されている情報という表現をしているが、長期的とはどの程度の期間なのかについて述べる。一般的に、長期的とは絶対的に期間が決められているものではなく、どちらかという相対的な表現として用いられる場合が多い。そのため、こういった情報について述べるかによって、期間が異なってくる。

そのため、1ヶ月以内しか利用されないものが多いタイプの情報ならば、3ヶ月でも長期的となるし、逆に半年ぐらい利用することが当たり前のタイプの情報ならば、半年利用していても長期的とはならない。そこで、本研究では、以下のような平均値と標準偏差を用いて、長期的な情報と短期的な情報を示す。

長期的な情報：長期度 $>$ 長期度の平均値 $+$ 長期度の標準偏差

短期的な情報：長期度 $<$ 長期度の平均値 $-$ 長期度の標準偏差

ここで長期度の平均値と標準偏差は、各種の情報ごとに計算する

例：Web ページ、検索キーワード、ファイルごと

以下の表 3.2 に、各情報についての具体的な期間を示す。

表 3.2: 長期的と短期的の期間

情報の種類	長期的	短期的
Web ページ	XXXX	XXXX
検索キーワード	XXXX	XXXX
ファイル	XXXX	XXXX

3.3. 長期的検索の提案

以上を踏まえて、本研究では、長期的な情報を検索するシステムを提案する。これは、検索者が今後も長期的に利用できる情報を見つけることができるようにすることを目指すものである。そのために、これまでに長期的に利用されてきた情報に特化して検索可能なシステムを実装する。

長期的に利用されてきた情報に特化して検索するために、まず、長期度の計算手法を提案する。そして、長期度の高い順に順位付けを行い、結果を提示する。

第4章 長期的な Web ページの検索

概要

本章では，長期的な Web ページを検索するシステム「セレクトブックマ」について述べる。

4.1. 背景

近年 Web 上のコンテンツは多種多様になってきており，一般の検索エンジンでは，Web 上に存在する体系だった知識や有益なコンテンツを手軽に取得することが容易ではなくなっている．これに対して，筆者は Web 上に存在する多種多様で玉石混合の Web コンテンツの中でも，いつ見ても有用な体系だった知識を得られる Web ページや有益なコンテンツは，長い間多くのユーザからアクセスされたり，ブックマークされたりすると考えた．

そこで，本研究では膨大で多種多様な情報の中から手軽に有益なコンテンツを取得するために，皆のブックマーク情報を共有できるソーシャルブックマークのデータの活用方法を考案した．本研究ではまず，ソーシャルブックマークデータを分析した．データの分析に基づき，一時期に限りブックマークされる Web ページは，一時的に必要とされる種類の Web ページが多いのに対し，長い間多くのユーザからブックマークされ続ける Web ページは，いつ見ても有用な情報を得られる種類の Web ページが多いことを示した．この特性に基づき，効率よく体系だった知識の得られる Web ページや有益な Web コンテンツを発見・収集する情報収集支援システム「セレクトブックマ」を提案，実装し，評価実験を行った．

4.2. 分析

4.2.1. ソーシャルブックマークデータの分析

国内最大規模のソーシャルブックマークサービスを提供しているはてなブックマーク [68] (図 4.1) のデータを収集した．はてなブックマークは，2012 年 12 月現在、約 1 億 5000 万件程ブックマークされている．この中から，2005 年 5 月～2008 年 9 月までにブックマークされたデータの中でブックマーク数 5 以上のページの以下のデータをすべてデータベースに収集した．

- URL
- タイトル
- ブックマークしたユーザ ID
- ブックマークした日時
- ブックマークしたユーザが付与したタグ名

はてなブックマークのデータ収集には，はてなブックマーク API を利用した．データベースに収集したデータの量は，表 4.1 に示す．

収集したデータから計算すると，ひとりが 1 つのブックマークをするときに平均して約 1.35 個のタグを付与していることがわかる．

表 4.1: 収集したデータ量

データ名	データ量
URL 数	762,239URL
ブックマーク数	12,751,661 ブックマーク
ユーザ数	87,898 人
タグ数	17,168,666 タグ
タグ数 (種類)	252,512 種類
レコード数	21,686,536 レコード



図 4.1: はてなブックマーク

4.2.2. 時間情報に関する分析

ユーザからいつ、どれくらいブックマークされるか、ブックマーク数と時間の関係について分析を行った。その結果、大まかに分けて次の3種類のタイプのWebページがあることが分かった。

1. 一時的にブックマークされ、その後ほとんどブックマークされなくなるタイプのページ（図 4.2）
2. 一時的に大量にブックマークされ、その後も長い間ブックマークされ続けるタイプのページ（図 4.3）
3. 大量にブックマークされる時期はないが、長い間ブックマークされ続けるタイプのページ（図 4.4）

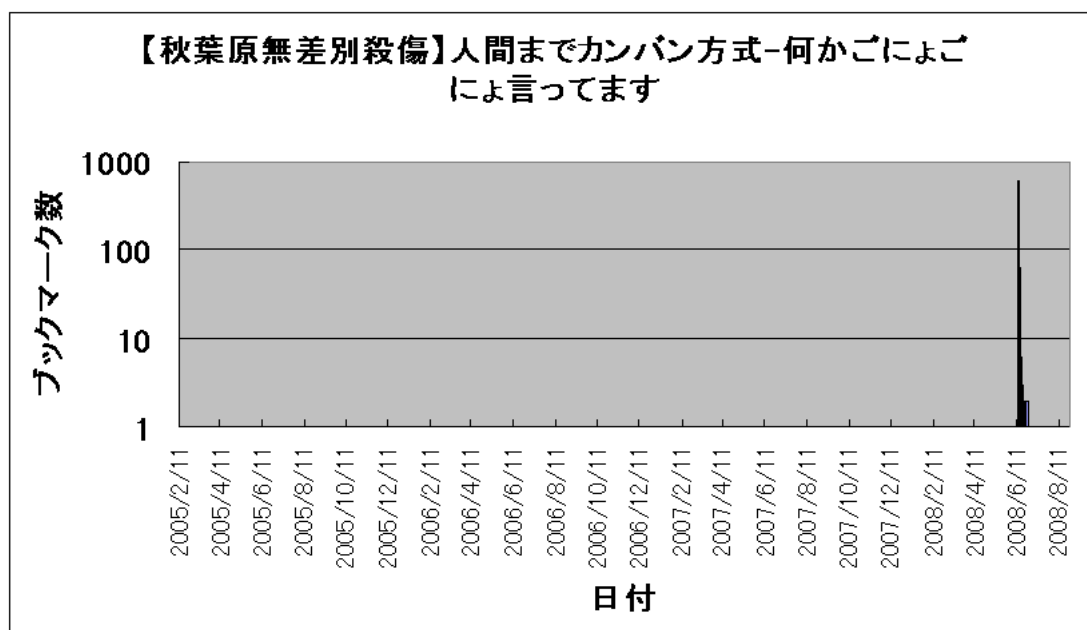


図 4.2: 一時的にブックマークされ、その後ほとんどブックマークされなくなるタイプのページ

以上の3種類のタイプのWebページをさらに、大まかに分類すると、以下の2種類のタイプのWebページに分類できる。

- Type1：一時期しかユーザからブックマークされないページ
- Type2：長い間ユーザからブックマークされ続けるページ

以上のType1とType2のWebページに対して、そのWebページがこういった種類のWebページであるかを分析した。その分析結果を以下の図 4.5 と図 4.6 に示す。分析対象のページは、以下の条件とした。

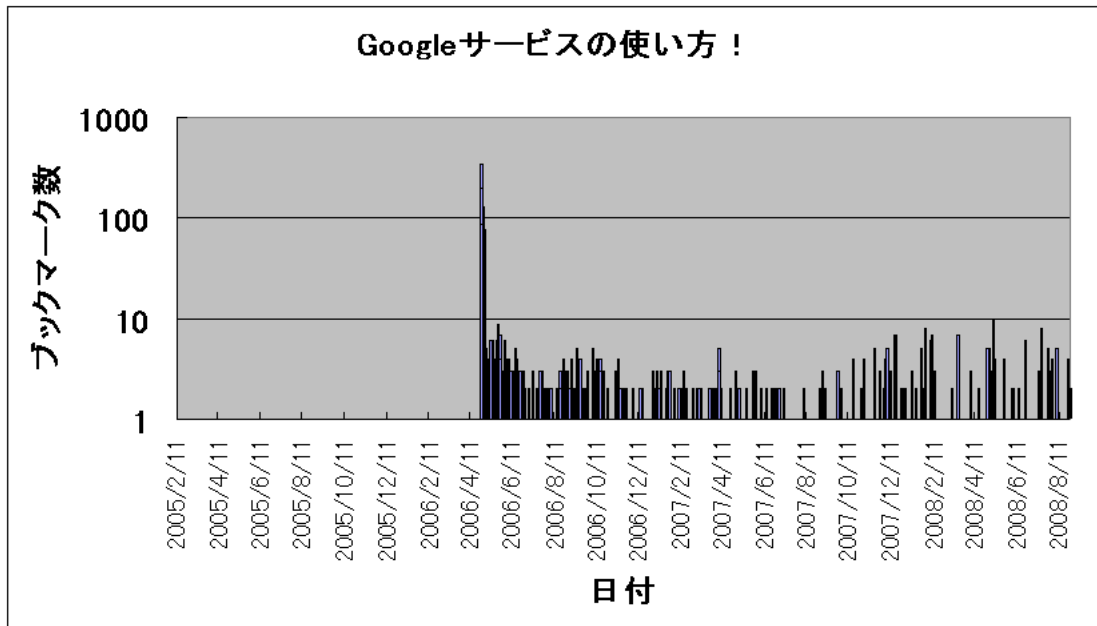


図 4.3: 一時的に大量にブックマークされ、その後も長い間ブックマークされ続けるタイプのページ

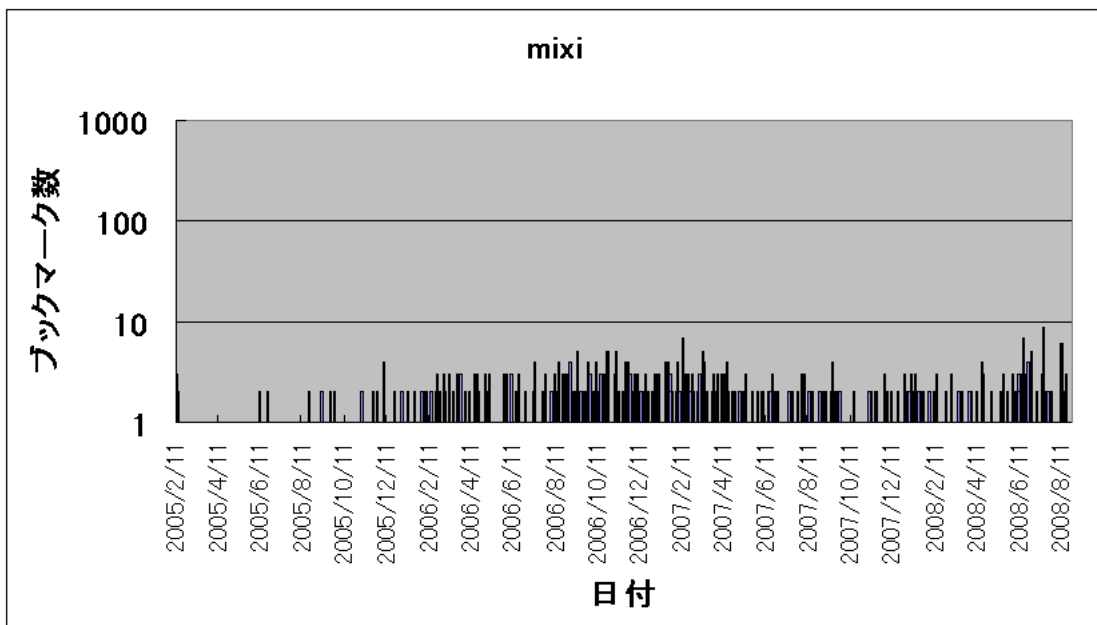


図 4.4: 大量にブックマークされる時期はないが、長い間ブックマークされ続けるタイプのページ

- Type1 は、全日数/全ブックマーク数=0.2 以下
- Type2 は、全日数/全ブックマーク数=0.8 以上
- Type1 と Type2 に対して、ブックマーク数 100 以上のページをランダムに 100 ページずつ取得

ここで全日数とは、ユーザからブックマークされた日数を表す。例えば、2007 年 1 月 3 日と 2007 年 2 月 10 日と 2008 年 10 月 10 日にそれぞれ異なったユーザからブックマークされた場合、3 日とする。ここで、全日数/全ブックマーク数=0.2 以下と 0.8 以上で分類した理由は、ブックマーク数 100 以上のページ数が、双方で近い値、かつ、双方とも 100 ページを大きく上回るページ数を確保できたからである。

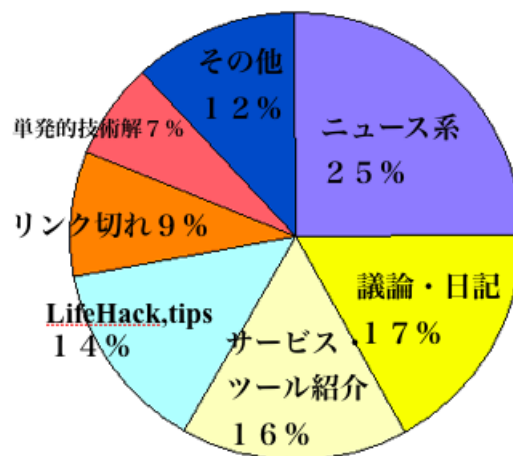


図 4.5: Type1 の Web ページの種類

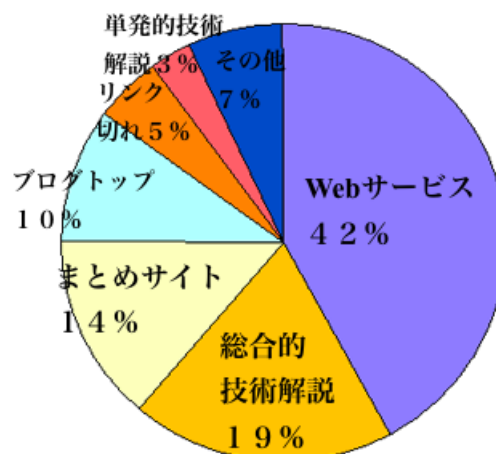


図 4.6: Type2 の Web ページの種類

図 4.5，図 4.6 から分かるように，Type1 の Web ページでは，「ニュース・話題」，「議論・日記」，「サービス・ツール紹介」が上位を占めており，一時的に利用される傾向の強い Web ページが大半を占めている．これに対して，Type2 の Web ページでは，「Web サービス」，「総合的技術解説サイト」，「まとめサイト」が上位を占めており，長期間にわたって利用される傾向の強い Web ページが大半を占めていることが分かった．このことから，Type2 のような長期間にわたってブックマークされ続けるような Web ページを優先的に取得することによって，Type1 のような一時的に利用される傾向の強い Web ページをフィルタリングして，いつ見ても有用な Web ページのみを検索できる可能性が高いことがわかった．

4.3. セレクトブックマの提案

4.3.1. セレクトブックマの概要

本研究では，ソーシャルブックマークのデータを利用した情報収集システム「セレクトブックマ」を提案・実装した．セレクトブックマでは，調べたい分野に対して，ソーシャルブックマークのブックマーク数とブックマークされた日数という二つの指標を利用して，Web ページをランキング化している．セレクトブックマを利用することによって，手軽に体系だった知識や有用な Web サービスを収集できる．

4.3.2. セレクトブックマの設計思想

セレクトブックマでは，特に情報収集の手軽さを重視している．また，調べたい分野に対して，以下 2 つのことを目的としている．

- 体系だった知識を得ること
- 有用な Web サービスを発見すること

情報を収集する際に，上記以外の Web ページが表示されないように，情報フィルタリングを行うことに注力している．

そのために，1 つ目の指標として，ソーシャルブックマークのブックマーク数という指標を利用している．これは，ユーザのブックマークするという行為が Web ページへの評価であるという考えに基づいている．

2 つ目の指標として，ブックマークされた日数を利用している．これは，4.2. の分析結果に基づき，長い間ブックマークされ続ける Web ページは，長期間必要とされる種類の Web ページが多いことを利用している．このブックマークされた日数という指標を利用することによって，一時的にしか利用しない Web ページをフィルタリングすることができる．

セレクトブックマβ版

皆から長い間ブックマークされ続ける優良サイトを検索して表示するサービスです。
長い間ブックマークされ続けるサイトは、主に有用なウェブサービス、総合情報サイト、まとめサイトなどです。
興味のある分野に対して、情報をフィルタリングして有用なサイトを発見するのに役立ちます。
はてなブックマークのデータを利用しています。

興味のある分野のタグをクリックするか、テキストボックスにタグ名を入力して検索ボタンを押して下さい。

(1) (2)

人気の検索タグ

総合	技術	趣味	社会・生活	その他
java 175回	java 175回	ニコニコ動画 130回	料理 71回	twitter 112回
ニコニコ動画 130回	javascript 83回	映画 113回	健康 59回	2ch 88回
映画 113回	php 76回	youtube 69回	日本 56回	まとめ 58回
twitter 112回	ui 74回	グルメ 66回	政治 50回	sbn 47回
2ch 88回	perl 58回	写真 52回	環境問題 48回	恋愛 45回
javascript 83回	ruby 54回	音楽 45回	経済 40回	広告 40回
php 76回	linux 50回	ゲーム 39回	社会 39回	blog 35回
ui 74回	google 46回	shopping 36回	health 29回	壁紙 35回
料理 71回	unix 42回	perfume 32回	教育 29回	2chまとめ 34回
youtube 69回	c 35回	tv 23回	中国 28回	これはすごい 29回



検索

セレクトブックマβ版

皆から長い間ブックマークされ続ける優良サイトを検索して表示するサービスです。
長い間ブックマークされ続けるサイトは、主に有用なウェブサービス、総合情報サイト、まとめサイトなどです。
興味のある分野に対して、情報をフィルタリングして有用なサイトを発見するのに役立ちます。
はてなブックマークのデータを利用しています。

興味のある分野のタグをクリックするか、テキストボックスにタグ名を入力して検索ボタンを押して下さい。

「java」に関するページの検索結果

(3) 1 [Javaの道 \(Java入門・リファレンス\)](#) 6557ポイント

2 [Javaの学習ならJavaDrive](#) 6318ポイント

3 [Java技術最新情報 JPro](#) 5832ポイント

4 [浅煎り珈琲 -Java アプリケーション入門](#) 5467ポイント

5 [頑健なJavaプログラムの書き方\(Writing Robust Java Code\)](#) 5330ポイント

6 [Java House ML](#) 5328ポイント

7 [JavaでHello World](#) 5325ポイント

8 [Java 2 Platform SE 5.0](#) 5325ポイント

9 [Java in the Box](#) 4550ポイント

10 [Log4J徹底解説～目次](#) 4420ポイント

(1) (2)

人気の検索タグ

総合	技術	趣味	社会・生活	その他
java 175回	java 175回	ニコニコ動画 130回	料理 71回	twitter 112回
ニコニコ動画 130回	javascript 83回	映画 113回	健康 59回	2ch 88回

図 4.7: セレクトブックマ画面

4.3.3. セレクトブックマ画面構成

セレクトブックマの検索前の画面と検索後の画面を図 4.7 に示す。

図 4.7 の (1) ~ (4) の説明を以下に示す。

(1) 検索単語 (タグ) 入力ボックス

検索単語 (タグ) を入力するテキストボックスで、タグを入力し、検索ボタンを押すことにより、指定したタグで検索を行う。

(2) 検索回数の多いタグ

人気のタグであり、検索回数の多い順に並べたものである。すべてを総合して検索回数の多い順に並べた「総合」と「技術」、「趣味」、「社会・生活」、「その他」のカテゴリごとに検索回数の多い順に並べたものがある。

(3) 検索結果の Web ページのタイトル

検索結果の Web ページのタイトルを表示したもので、タイトルのリンクをクリックすると、クリックした Web ページを表示する。

(4) ランキングの値

後に示すランキングの計算式を用いて計算した値とその値を棒グラフで可視化したものである。

4.3.4. セレクトブックマインタフェース

セレクトブックマの基本的なインタフェースは、検索窓になにか単語を入力し、検索ボタンを押すことによって検索をおこなう。検索結果は、1 ページに上位 10 件表示され「次へ」ボタンを押すと、11 位 ~ 20 位までが表示される。このあたりは、一般的な検索エンジンと同じである。

上述に加えて、補助的な機能として、セレクトブックマを利用したユーザの検索回数の多い単語をカテゴリごとに表示し、そのリンクをクリックすることによっても検索可能としている。セレクトブックマは、検索エンジンというより、情報収集支援システムという位置付けで、運用をしている。ユーザには、ある分野ごとに情報を収集するという目的で利用してもらう想定である。そのため、検索単語の選び方が少し一般的な検索エンジンと異なり、検索単語が複数例示されていると、新規ユーザにとっても取り付きやすいのではないかと考えた。

検索回数の多い単語は、興味のある分野を見つけやすいように「技術」、「趣味」、「社会・生活」、「その他」の 4 分野にカテゴリわけを行っている。さらに、すべてのカテゴリを含めた検索回数に対して順位を付け、それを「総合」として表示している。

4.3.5. 検索ランキングロジック

本研究では、検索結果のランキング手法について、さまざまな手法を考案し、試作した。ひとつひとつの手法についての詳細な評価は行っていないが、以下に示す手法が現状ではもっとも効果的であった。

セレクトブックマでは、検索結果のランキングを出すにあたって、検索単語（タグ）として指定したタグでのブックマーク数に、指定したタグでブックマークされた日数で重み付けをして、値の大きいものほど順位が高くなるようにランキングを行っている。ランキングの計算式を以下に示す。

$$Bookmarks \times Days^{\alpha} \quad (4.1)$$

Bookmarks : 指定したタグでのブックマーク数

Days : 指定したタグでブックマークされた日数

: 任意の係数

4.3.6. システム構成

セレクトブックマは、Web サービスとして実装した。画面の表示部分は、HTML、JSP、JavaScript を利用し、計算などやデータベースとの連携は、主に Java を利用している。Java と JavaScript の連携は、Ajax 方式を用いて、JSON 形式でデータの受け渡しをしている。ユーザが検索を行う場合、主に以下の手順でシステムが動作する。

1. ユーザが検索する
2. 検索単語がサーバへ送られる
3. 検索単語でデータベースを検索する
4. データベースの検索結果から、ランキングを計算する
5. ランキングに基づき、ユーザに検索結果を返す

セレクトブックマのシステム構成図を以下の図 4.8 に示す。

また、DB のテーブル構造は、以下の図 4.9 に示す。

<http://plazman.chi.mag.keio.ac.jp/sbm/summary.jsp>

4.4. 評価実験

4.4.1. 実験概要

セレクトブックマの有用性を評価するため、以下 3 種類の手法において比較実験を行った。

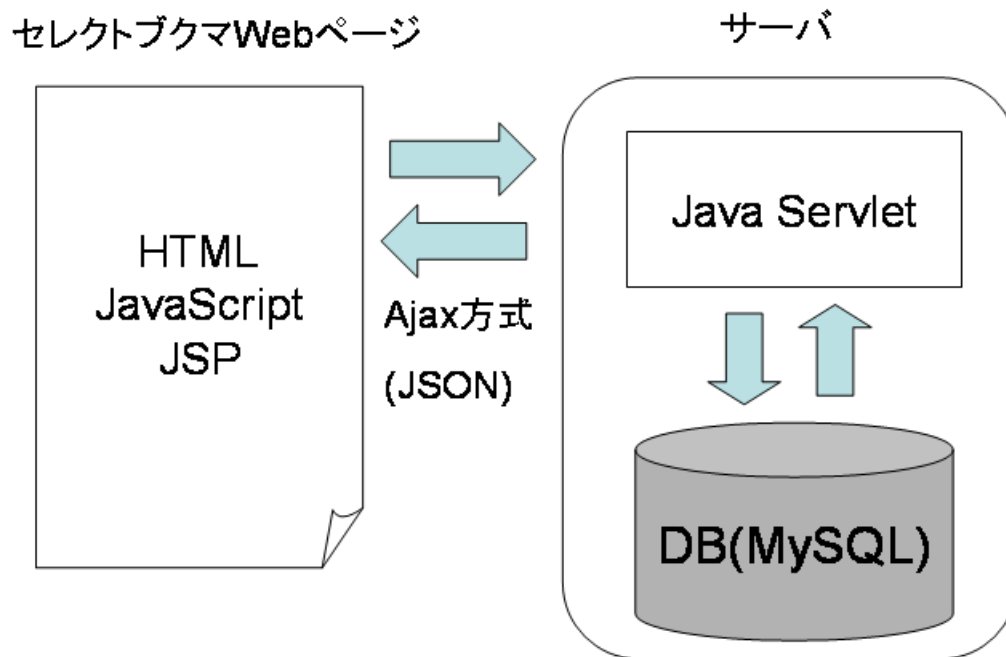


図 4.8: システム構成図

テーブル 1

カラム名	id	url
説明	ID	URL
例	1000000	http://xxx.xxx

テーブル 2

カラム名	url	title	count	user	time	tag
説明	URL	タイトル	ブックマーク数	ブックマークした ユーザのID	ブックマーク された時間	ブックマーク したタグ名
例	http://xxx.xxx	ページA	1232	wilfue	20070730	java

図 4.9: DB のテーブル構成図

- Google 検索
- 取得したはてなブックマークのデータの中で、タグで検索した場合のブックマーク数が多いものから順にランキングしたもの（以降、ブックマーク数順）
- セレクトブックマ（ $n=1$ ）

4.4.2. 実験目的

本実験の目的としては、本研究の目的である、Web 上から以下の情報を手軽に取得できるかどうかについて評価することである。

- 体系だった知識を得ること
- 有益な Web コンテンツを発見すること

そのため、以下のようなことを目的とした実験をおこなった。

- セレクトブックマと既存の検索エンジンにおいて、どちらがより今後も利用するような体系だった知識を得られる Web ページを手軽に取得できるか評価する
- セレクトブックマで利用している手法である、時間情報による重み付けによって、一時的に必要とされる情報をフィルタリングできていることを実証する

4.4.3. 実験方法

以下、3 種類の手法で検索を行い、検索結果上位各 30 件を取得した。取得した合計 90 件の Web サイトの中で重複したものを除いた Web サイトを、順番をランダムにしたリストとして示した。そのリストの中から被験者に今後も利用したいと思う Web サイトを 1 位～10 位まで選んでもらう。

- Google 検索
- ブックマーク数順
- セレクトブックマ（ $n=1$ ）

被験者が選んだ Web サイトを適合文書として、適合率^{*8}、再現率^{*9}を求める。
実験条件は以下とする。

- 被験者数：30 人（1 単語につき 10 人 × 3 単語）
- 検索単語（タグ）：「java」「健康」「映画」
- 「java」：男性 6 人，女性 4 人
- 「健康」：男性 5 人，女性 5 人

- 「映画」: 男性 3 人, 女性 7 人

検索単語 (タグ) は, 技術系の分野から「java」, 生活系の分野から「健康」, 娯楽系の分野から「映画」と 3 つの異なった分野から 1 つずつ選んだ。それぞれ, 技術系, 生活系, 娯楽系の中でも, 特別はてなブックマークデータを利用した場合に有利になるような単語ではなく, できるだけ一般的な単語を選んだ。被験者の負担を軽減するため, 1 人 1 単語を目安に実験をおこなった。また, Google 検索結果は, 2009 年 12 月 23 日に検索した検索結果を利用した。

4.4.4. 実験結果

以上のような実験方法に基づき、実験を行った結果について示す。

(1) 各検索手法による検索結果上位 10 件

Google 検索とブックマーク数順とセレクトブックマでの検索結果上位 10 件のタイトルを以下の表 4.2 ~ 4.4 に示す。すべての検索結果上位 30 件のタイトルと URL は, 付録 A に示す。

検索結果を見ると, Google 検索の場合は, 大手企業が作成した Web サイトが上位にランキングされやすい傾向にあるのに対して, 逆にセレクトブックマでは, 個人で作成したような Web サイトが上位にランキングされやすい傾向にあることがわかる。また, ブックマーク数順とセレクトブックマでは, ある程度ランキングされる Web ページに近い傾向にあることがわかる。これは, セレクトブックマのロジックが, ブックマーク数に対して日付で重みを付けていることに起因している。

(2) 被験者が選んだ Web ページ

Google 検索、ブックマーク数順、セレクトブックマの検索結果から, 被験者が今後も利用したいと思う Web サイトを 1 位 ~ 10 位までを被験者に選んでもらった結果を以下の表 4.5 ~ 表 4.10 に示す。「java」, 「健康」, 「映画」という各単語に対して, 以下 2 種類の結果を示す。

1. 人数順

被験者が何人が選んだか人数順に示したものである。3 人以上選んだ Web ページのみ記載する。

2. 得点順

被験者が選んだ順位によって得点を付与し, 得点順に並べたものの 2 種類を示す。得点は, 1 位 10 点, 2 位 9 点, 3 位 8 点 … 10 位 1 点のようにつけた。得点の上位 30 件のみを記載する。

「java」に関しては, 表 4.5, 表 4.6, 「健康」に関しては, 表 4.7, 表 4.8, 「映画」に関しては, 表 4.9, 表 4.10, に示す。

表 4.2: 「java」で検索した場合の検索結果上位 10 件

順位	セレクトブックマ	ブックマーク数順	Google 検索
1	Java の道 (Java 入門: リファレンス)	Java のクラスアンロード (Class Unloading)	java.com: あなたと Java
2	Java の学習なら、JavaDrive	Java の道 (Java 入門・リファレンス)	無料 Java ソフトウェアをダウンロード - Sun Microsystems
3	Java 技術最前線: ITpro	頑健な Java プログラムの書き方 (Writing Robust Java Code)	Java - Wikipedia
4	浅煎り珈琲-Java アプリケーション入門	Java 技術最前線: ITpro	Java テクノロジ - サン・マイクロシステムズ
5	頑健な Java プログラムの書き方	Java の学習なら JavaDrive	サン・マイクロシステムズ
6	Java House ML	浅煎り珈琲-Java アプリケーション入門	Java とは - 意味/解説/説明/定義 : IT 用語辞典
7	Java で Hello World	Java で Hello World	日本 Java ユーザグループ
8	Java 2 Platform SE 5.0	Java 2 Platform SE	Java とは - はてなキーワード
9	Java in the Box	【レポート】Java 初学者には最適!? 解説から実行までブラウザでコンプリート - Javala (MYCOM ジャーナル)	Sun Developer Connection - Java Developer Connection
10	Log4J 徹底解説 ~ 目次	Java House ML	Java の道 (Java 入門・リファレンス)

表 4.3: 「健康」で検索した場合の検索結果上位 10 件

順位	セレクトブックマ	ブックマーク数順	Google 検索
1	health クリック 健康 生活習慣病 サプリメント	ゲンダイネット - 目の疲れをためない3大作戦	健康 - Wikipedia
2	基礎代謝を高めるための99の技法	体の歪みを治したい:アルファアルファモザイク	家庭の医学・健康 - goo ヘルスケア
3	ゲンダイネット - 目の疲れをためない3大作戦	重い、痛いを吹き飛ばせ！肩コリ解消法 — Web 担当者 Forum	健康ネット
4	視力回復マッサージは本当に効き目があるのだ（映像付） — i d e a * i d e a	眼精疲労を治すには:アルファアルファモザイク	病院情報、家庭の医学、病気の検索、薬の情報、健康情報 - Yahoo ...
5	重い、痛いを吹き飛ばせ！肩コリ解消法 — Web 担当者 Forum	金も時間も掛からない花粉症予防の仕方:アルファアルファモザイク	厚生労働省：健康
6	体の歪みを治したい:アルファアルファモザイク	health クリック 健康 生活習慣病 サプリメント	健康 ON-LINE
7	病院検索ならここカラダ	蒸しタオルを使うと酷使した目の疲れが取れる - GIGAZINE	asahi.com(朝日新聞社):医療・健康・ニュース
8	意外に効き目のある視力回復マッサージ	疲れ目：パソコン作業で肩こり、視力低下など眼精疲労 対策は...（上） - 毎日 j p (毎日新聞)	いきいき健康 NIKKEI NET
9	蒸しタオルを使うと酷使した目の疲れが取れる - GIGAZINE	ビジネスマンの不死身力：「夜食は太る」の科学 (1/2) - ITmedia エンタープライズ	[健康管理] All About
10	MouRa Net 現代 巷にはびこる「健康情報」50のウソ・ホント	基礎代謝を高めるための99の技法	ケンコーコム - 健康メガシヨップ

表 4.4: 「映画」で検索した場合の検索結果上位 10 件

順位	セレクトブックマ	ブックマーク数順	Google 検索
1	前田有一の超映画批評	前田有一の超映画批評	Yahoo!映画 - 映画情報
2	みんなのシネマレビュー	みんなのシネマレビュー	映画情報 - goo 映画
3	あの映画のココがわからない まとめサイト	あの映画のココがわからない まとめサイト	映画・DVD と映画館の上映時間を完全網羅 — Movie Walker
4	allcinema ONLINE 映画データベース	allcinema ONLINE 映画データベース	映画のことなら eiga.com
5	映画生活 - 新作映画情報	超映画批評	映画「サマーウォーズ」公式サイト
6	超映画批評	段ボールで『2001 年宇宙の旅』をリメイクできるか? - バイエリア在住町山智浩アメリカ日記	映画 - Wikipedia
7	CinemaScape?映画批評空間?	映画生活 - 新作映画情報	TOHO シネマズ
8	cinemacafe.net シネマカフェ?映画ファンによる、映画ファンのための、映画的生活スタイル・エンターテイメント・サイト?	CinemaScape?映画批評空間?	ワーナー・マイカル・シネマズ - 映画館、映画情報、上映スケジュール ...
9	eiga.com	痛いニュース (‘):もっとも感動した映画トップ 10 発表	大ブーイング! ぐだぐだ最終回「JIN」映画化?TBS 意外な反応 ...
10	いのちの食べかた	404 Blog Not Found:アマとプロとが選んだディストピア映画歴代トップ 26 - 1	新作映画情報「ぴあ映画生活」

表 4.5: 「java」で検索した場合の検索結果の中から被験者が選んだ Web ページ (人数)

人数	タイトル	セレクトブックマ順位	ブックマーク数順位	Google 検索順位
7	Java の道 (Java 入門・リファレンス)	1	2	10
7	とほほの Java 入門			14
7	Java の学習なら JavaDrive	2	5	
5	初心者が Java を“超高速”で学ぶためのコツ:ITpro	12	22	
4	Java 入門	27		
4	TECHSCORE(テックスコア) ?C 言語/JAVA/デザインパターン/CORBA/XML/SQL/UML を基礎から丁寧に解説します?	26		
4	Java FAQ: よくある質問とその回答集			21
4	Java で Hello World	7	7	13
4	MYCOM ジャーナル - エンタープライズ - コラム - ライトニング Java	20		
3	Java を JavaScript に変換するグーグルのツールを使ってみよう ? @ IT	25		
3	Java -TECHSCORE-			19
3	頑健な Java プログラムの書き方 (Writing Robust Java Code)	5	3	
3	@ IT : Hibernate で理解する O/R マッピング (1)	30		

表 4.6: 「java」で検索した場合の検索結果の中から被験者が選んだ Web ページ (得点)

得点	タイトル	セレクトブ クマ順位	ブックマー ク数順位	Google 検索順位
55	とほほの Java 入門			14
45	Java の道 (Java 入門・リファレンス)	1	2	10
33	Java の学習なら JavaDrive	2	5	
27	Java FAQ: よくある質問とその回答集			21
23	Java 入門	27		
20	MYCOM ジャーナル - エンタープライズ - コラ ム - ライトニング Java	20		
19	Java で Hello World	7	7	13
19	Java 技術最前線 : ITpro	3	4	
18	Java -TECHSCORE-			19
18	頑健な Java プログラムの書き方 (Writing Robust Java Code)	5	3	
18	JavaA2Z	22		
15	TECHSCORE(テックスコア) ?C 言語/JAVA/デ ザインパターン/CORBA/XML/SQL/UML を基 礎から丁寧に解説します?	26		
15	ITmedia エンタープライズ : 矛盾を抱えつつ進化 する“ Java ”??黒船となった Ruby on Rails (1/2)		29	
14	初心者が Java を“ 超高速 ”で学ぶためのコツ:ITpro	12	22	
14	Java 2 Platform SE 5.0	8	8	
13	Java を JavaScript に変換するグーグルのツールを 使ってみよう ? @ IT	25		
13	Java におけるコード進化パターン (Code Evolu- tion Patterns in Java)		14	
12	C/C いっさいなし、Java だけで開発された OS - JNode (MYCOM PC WEB)		19	
11	@ IT : Hibernate で理解する O/R マッピング (1)	30		
10	Java アプレット - Wikipedia			20
9	Java Solution ? @ IT	28		12
9	IBM developerWorks Japan : Resources for Java developers			17
9	Java でゲーム作りますが何か?	16	17	
8	Ruby よりも Java が好きな理由		11	
8	Java in the Box	9	13	23
8	Ja-Jakarta Project	21		

表 4.7: 「健康」で検索した場合の検索結果の中から被験者が選んだ Web ページ (人数)

人数	タイトル	セレクトブックマ順位	ブックマーク数順位	Google 検索順位
6	意外に効き目のある視力回復マッサージ	8	28	
6	視力検査 - あなたの視力 今いくつ?: 視力回復のアイポータル	30		
5	蒸しタオルを使うと酷使した目の疲れが取れる - GIGAZINE	9	7	
5	なかなか眠れない人のための簡単に眠る 10 の方法 - GIGAZINE	20		
4	第 1 回 眼の疲れを取る: ITpro	11	12	
3	基礎代謝を高めるための 9 9 の技法	2	10	
3	視力回復とレーシックのアイポータル	14	16	
3	ビジネスマンの不死身力: 「夜食は太る」の科学 (1/2) - ITmedia エンタープライズ		9	
3	5 時間以下の睡眠続け死亡率 1.7 倍に 7 時間寝よう Ameba News	13	18	
3	視力回復マッサージは本当に効き目があるのだ (映像付) — i d e a * i d e a	4	21	
3	睡眠時間を記録するサイト — ねむログ	25		
3	なんでも評点: 空腹は幸福?? ストレスで腹が減ったときは何も食べずに我慢した方がストレスに打ち克てる...		19	
3	シゴタノ! - 睡眠時間を短くする 1 4 のコツ < 前編 >	26		
3	「うつ」にならない、繰り返さない? @ IT 自分戦略研究所	27		
3	スラッシュドット・ジャパン — 睡眠不足だと仕事はかどらない理由、科学的に明らかになる		26	
3	姿勢をよくするための運動 - Tech Mom from Silicon Valley	28	13	

表 4.8: 「健康」で検索した場合の検索結果の中から被験者が選んだ Web ページ (得点)

得点	タイトル	セレクトブ クマ順位	ブックマー ク数順位	Google 検索順位
36	視力検査 - あなたの視力 今いくつ?: 視力回復の アイポータル	30		
33	基礎代謝を高めるための 99 の技法	2	10	
30	意外に効き目のある視力回復マッサージ	8	28	
30	なかなか眠れない人のための簡単に眠る 10 の方法 - GIGAZINE	20		
24	蒸しタオルを使うと酷使した目の疲れが取れる - GIGAZINE	9	7	
23	睡眠時間を記録するサイト — ねむログ	25		
21	第 1 回 眼の疲れを取る: ITpro	11	12	
21	視力回復マッサージは本当に効き目があるのだ (映 像付) — i d e a * i d e a	4	21	
19	Medical Tribune あなたの健康百科			16
18	ビジネスマンの不死身力: 「夜食は太る」の科学 (1/2) - ITmedia エンタープライズ		9	
18	スラッシュドット・ジャパン — 睡眠不足だと仕事 がはかどらない理由、科学的に明らかになる		26	
17	5 時間以下の睡眠続け死亡率 1.7 倍に 7 時間寝 よう Ameba News	13	18	
17	「うつ」にならない、繰り返さない? @ IT 自分 戦略研究所	27		
14	視力回復とレーシックのアイポータル	14	16	
14	シゴタノ! - 睡眠時間を短くする 14 のコツ < 前 編 >	26		
14	精神状態の健康がピンチになった時の 3 つの対処 - koe だめ		27	
14	重い、痛いを吹き飛ばせ! 肩コリ解消法 — Web 担当者 Forum	5	3	
14	割れた腹筋を手に入れるトレーニングを教えても らいました。 — その他 (ライフ) — とりあえず...	17		
13	[健康管理] All About			19
12	医学都市伝説: 暗いところで本を読んでも目は悪 くならない	19	22	
11	asahi.com (朝日新聞社): 医療・健康・ニュース			7
10	なんでも評点: 空腹は幸福?? ストレスで腹が減っ たときは何も食べずに我慢した方がストレスに打 ち克てる...		19	
10	健康 - Wikipedia			1
10	病院情報、家庭の医学、病気の検索、薬の情報、健 康情報 - Yahoo ...			4
10	PC で眼が疲れない方法 - 萌え理論 Blog	12	17	
9	指をボキボキ鳴らすと太くなるって本当?		20	
9	姿勢をよくするための運動 - Tech Mom from Sil- icon Valley	28	13	
9	NHK 健康ホームページ: トップページ			12
8	health クリック 健康 生活習慣病 サプリメント	1	6	
8	体の歪みを治したい: アルファルファモザイク	6	2	

表 4.9: 「映画」で検索した場合の検索結果の中から被験者が選んだ Web ページ (人数)

人数	タイトル	セレクトブ クマ順位	ブックマー ク数順位	Google 検索順位
5	Yahoo!映画 - 映画情報			1
4	映画情報 - goo 映画			2
4	シネマぴあ			26
4	TOHO シネマズ			7
4	映画、映画館検索 - TSUTAYA online			16
4	@nifty 映画 - 映画情報			27
4	ワーナー・マイカル・シネマズ - 映画館、映画情 報、上映スケジュール ...			8
3	あの映画のココがわからない まとめサイト	3	3	
3	【2ch】ニュー速クオリティ:一生のうちに一度は 見ておくべき映画		11	
3	もっと知られていい映画:アルファルファモザイク		14	
3	前田有一の超映画批評	1	1	25
3	みんなのシネマレビュー	2	2	
3	シネマスクランブル 映画予告編・映画ランキング・ 映画上映 ...			13
3	映画なら GyaO! 【映画】 無料映画の映像視聴 GyaO![ギャオ] 映画			30
3	【2ch】ニュー速クオリティ:観るまでバカにして たのに観たら面白かった映画		27	
3	allcinema ONLINE 映画データベース	4	4	12
3	eiga.com	9	18	4
3	109 シネマズ公式ホームページ			24
3	映画・DVD と映画館の上映時間を完全網羅 — Movie Walker			3

表 4.10: 「映画」で検索した場合の検索結果の中から被験者が選んだ Web ページ (得点)

得点	タイトル	セレクトブックマ順位	ブックマーク数順位	Google 検索順位
29	Yahoo!映画 - 映画情報			1
28	ワーナー・マイカル・シネマズ - 映画館、映画情報、上映スケジュール ...			8
26	TOHO シネマズ			7
25	映画、映画館検索 - TSUTAYA online			16
25	【2ch】ニュー速クオリティ:観るまでバカにしたのに観たら面白かった映画		27	
23	あの映画のココがわからない まとめサイト	3	3	
21	allcinema ONLINE 映画データベース	4	4	12
20	109 シネマズ公式ホームページ			24
19	映画・DVD と映画館の上映時間を完全網羅 — Movie Walker			3
18	eiga.com	9	18	4
18	著作権の切れた過去の名作映画やドキュメンタリーをダウンロード - GIGAZINE	27		
17	みんなのシネマレビュー	2	2	
17	痛いニュース (':大人が選ぶ泣ける洋画ベスト30 1「タイタニック」 2「アルマゲドン」		28	
17	映画、ビデオ - Yahoo!カテゴリ			17
16	シネマぴあ			26
15	@nifty 映画 - 映画情報			27
15	【2ch】ニュー速クオリティ:一生のうちに一度は見ておくべき映画		11	
15	映画「サマーウォーズ」公式サイト			5
15	映画館・シネコンの【ムービックス-MOVIX】			11
14	映画なら GyaO! 【映画】 無料映画の映像視聴 GyaO![ギャオ] 映画			30
12	映画情報 - goo 映画			2
12	もっと知られていい映画:アルファルファモザイク		14	
12	シネマスクランブル 映画予告編・映画ランキング・映画上映 ...			13
12	無料映画館			28
11	新作映画情報「ぴあ映画生活」			10
10	前田有一の超映画批評	1	1	25
10	『アバター』は映画の未来に iPhone 登場なみの衝撃を与える			19
9	時をかける少女	12		
8	flowerwild.net - 蓮實重彦インタビュー リアルタイム批評のすすめ vol.1	25		

(3) 適合率・再現率による比較

各手法においての、適合率^{*8}・再現率^{*9}を比較した。適合率^{*8}・再現率^{*9}の意味に関しては、1章の用語定義に示した。

適合率・再現率を出すために必要な適合文書は、以下2種類作成した。

1. 3人以上の被験者が上位10位以内に選択したWebページ
2. 被験者が選んだWebページの中で、得点の合計値が8点以上のWebページ

全適合文書の件数を以下の表4.11に示す。

表 4.11: 全適合文書の件数

単語名	java	健康	映画
人数 [件]	13	16	19
得点 [件]	26	30	29

人数から全適合文書を作成した場合の適合率を図4.10、図4.12、図4.14に、得点から全適合文書を作成した場合の適合率を図4.11、図4.13、図4.15に示す。また、人数から全適合文書を作成した場合と得点から全適合文書を作成した場合の再現率を表4.12～表4.14に示す。適合率の値は、セレクトブックマ、ブックマーク数順、Google 検索での検索結果1位～30位までに対する以下の値を示している。

$$\text{適合率} = \frac{\text{適合文書の数}}{\text{検索結果の文書の数}} \quad (4.2)$$

ここでいう検索結果の文書の数とは、1位の場合1、2位の場合2、..., 30位の場合30となる。また、再現率の値は、以下の値を示している。

$$\text{再現率} = \frac{\text{検索結果中の適合文書の数}}{\text{全適合文書の数}} \quad (4.3)$$

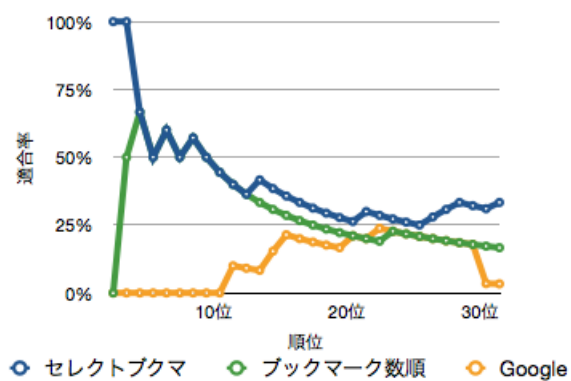


図 4.10: 3 人以上を適合文書とした場合の適合率 (java)

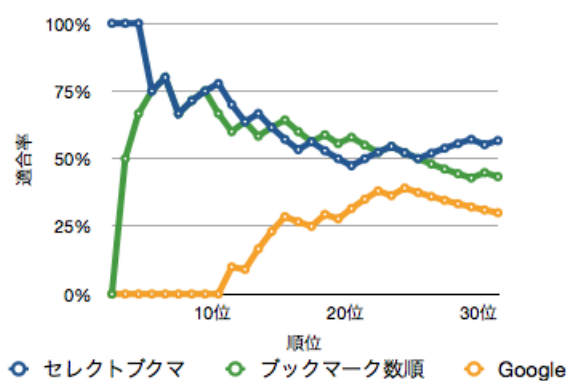


図 4.11: 8 点以上を適合文書とした場合の適合率 (java)

表 4.12: 再現率 (java)

	セレクトブックマ	ブックマーク数順	Google 検索
再現率 (人数)	77 %	38 %	38 %
再現率 (得点)	65 %	50 %	35 %

再現率の値は，小数点第一位以下を四捨五入している．

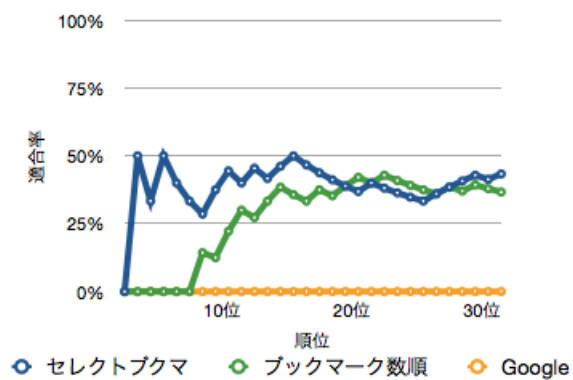


図 4.12: 3 人以上を適合文書とした場合の適合率 (健康)

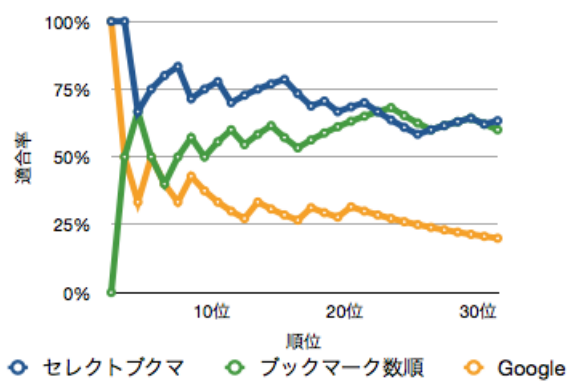


図 4.13: 8 点以上を適合文書とした場合の適合率 (健康)

表 4.13: 再現率 (健康)

	セレクトブックマ	ブックマーク数順	Google 検索
再現率 (人数)	81 %	69 %	0 %
再現率 (得点)	63 %	60 %	20 %

再現率の値は，小数点第一位以下を四捨五入している．

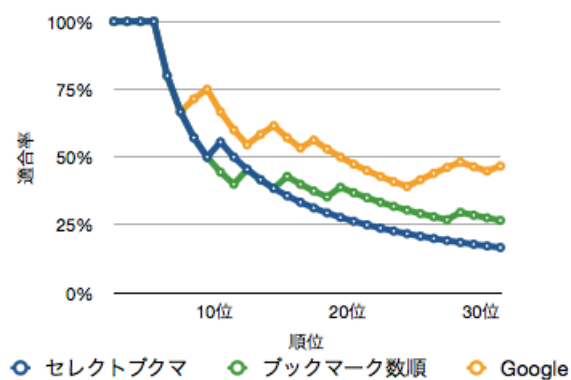


図 4.14: 3 人以上を適合文書とした場合の適合率 (映画)

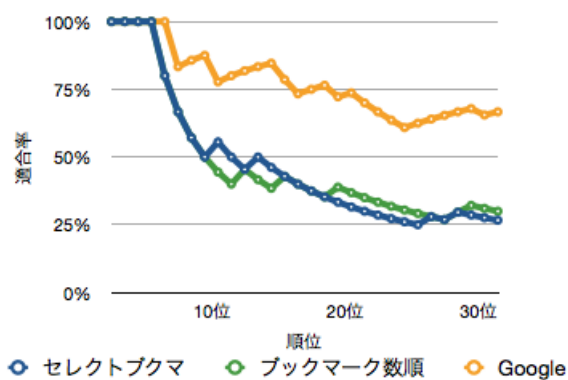


図 4.15: 8 点以上を適合文書とした場合の適合率 (映画)

表 4.14: 再現率 (映画)

	セレクトブックマ	ブックマーク数順	Google 検索
再現率 (人数)	26 %	42 %	74 %
再現率 (得点)	28 %	31 %	69 %

再現率の値は，小数点第一位以下を四捨五入している．

全体的に3人以上を適合文書とした場合の適合率が低いのは、表 4.11 からわかるように、全適合文書の件数が少ないからである。逆に、再現率は、8 点以上を適合文書としたときより、3 人以上を適合文書とした場合の方が全体的に高い値となっている。これは、一般的に再現率は、全適合文書の数が少ないと高い値になる傾向があるためである。

「java」に関しては、全体的にセレクトブックマにおいてもっとも高い適合率となっているが、特に高い順位においてその傾向は強い。また、再現率も人数で適合文書を作った場合、得点から適合文書を作った場合の両方において、セレクトブックマでもっとも高い値となった。

「健康」に関しても「java」と類似した結果となり、全体的にセレクトブックマにおいてもっとも高い適合率となっている。さらに、高い順位においてその傾向が強い。また、再現率も人数で適合文書を作った場合、得点から適合文書を作った場合の両方において、セレクトブックマでもっとも高い値となった。

「映画」に関しては「java」「健康」とはまったく異なる結果となり、Google 検索においてもっとも高い適合率となった。また、再現率も人数で適合文書を作った場合、得点から適合文書を作った場合の両方において、Google 検索でもっとも高い値となった。

(4) 得点の合計値による比較

被験者が選んだ Web ページには、順位によって得点がついている。その得点がセレクトブックマ、ブックマーク数順、Google 検索の検索結果の Web ページに合計何点入っているか、比較する。セレクトブックマ、ブックマーク数順、Google 検索において、検索結果上位 10 件、上位 20 件、上位 30 件の 3 つの得点の合計値を以下の表 4.15～表 4.17 に示す。

表 4.15: 「java」での検索結果の得点の合計値

	セレクトブックマ	ブックマーク数順	Google 検索
上位 10 件の合計値 [点]	162	152	23
上位 20 件の合計値 [点]	225	222	148
上位 30 件の合計値 [点]	322	279	195

表 4.16: 「健康」での検索結果の得点の合計値

	セレクトブックマ	ブックマーク数順	Google 検索
上位 10 件の合計値 [点]	147	119	48
上位 20 件の合計値 [点]	269	208	92
上位 30 件の合計値 [点]	378	311	95

結果を見ると「java」と「健康」においては、セレクトブックマで上位 10 件、上位 20 件、上位 30 すべてにおいて、もっとも高い合計得点となった。これに対して、「映画」においては、Google 検索で上位 10 件、上位 20 件、上位 30 すべてにおいて、もっとも高い合計得点となった。

表 4.17: 「映画」での検索結果の得点の合計値

	セレクトブックマ	ブックマーク数順	Google 検索
上位 10 件の合計値 [点]	89	77	163
上位 20 件の合計値 [点]	114	125	276
上位 30 件の合計値 [点]	153	169	371

4.4.5. 時間情報を利用する効果について

今回被験者実験をおこなった検索単語は、技術系、生活系、娯楽系と 3 つの異なった分野から 1 つずつ、特別セレクトブックマが有利になるような単語を選ばずに、できるだけ一般的な単語を選んだ。そのため、「java」、「健康」、「映画」という単語を選ぶこととなったが、これらの単語はセレクトブックマとブックマーク数順の結果において、上位 10 件の中だけ見ても、その順位に変化はあるが、半数程度同一の結果が含まれる。

これに対して、他の単語で検索した場合、もっとセレクトブックマとブックマーク数順で検索結果に相違がでるものも多い。例えば、「UI」で検索した場合は、上位 10 件の中に 3 件しか同一の結果が含まれない。そのため、「UI」の場合は、時間情報による影響がもっと強いといえる。

さらに、現状セレクトブックマのランキングロジックは、以下のようになっている。

$$Bookmarks \times Days^{\alpha} \quad (4.4)$$

Bookmarks : 指定したタグでのブックマーク数

Days : 指定したタグでブックマークされた日数

: 任意の係数

ブックマーク数順は、日付にかける係数 α の値が 0 のときと同義である。今回の実験は、 $\alpha = 1$ として実験をおこなったが、 α の値をもっと大きくすれば、時間の影響が強くなり、セレクトブックマとブックマーク数順の検索結果の違いを大きくすることができる。 α の値を変化させ、実験・評価をおこなうと時間情報を利用する効果についてもっと明確にできると考えられる。

4.4.6. 各実験結果からの考察

ここまでの各実験結果から、以下のようなことが考察できる。

(1) 適合率からの考察

「java」で検索した場合、興味深いのは、Google 検索の検索結果の順位が下がるにつれて、適合率が上昇傾向にあることである。このことから「java」という分野においては、Google 検索の検索結果の上位が、一時的にしか利用しない Web ページが多く、ユーザにとって今後も利用したいような Web ページが少ないということがわかる。この傾向は、他の一般的な検索エンジンを用

いて「java」で検索しても Google 検索と類似した結果が出るため、一般的な検索エンジン全体にいえるのではないかと考えられる。このため、少なくとも「java」という分野においては、一般的な検索エンジンが体系だった知識を得られる Web ページや有益な Web サービスを手軽に発見・収集するためには、向かないのではないかと考えられる。

これに対して、「java」で検索した場合、セレクトブックマでもっとも高い適合率が得られており、さらに高い順位程高い適合率が得られている。そのため、セレクトブックマが体系だった知識を得られる Web ページや有益な Web サービスを手軽に発見するために有用であることを確認できた。ブックマーク数順と比較しても、高い順位程適合率の差が大きい傾向にあるため、時間情報を利用することによるフィルタリング効果も確認できた。

「健康」や「映画」で検索した場合は、Google 検索の検索結果の順位が高い程、適合率が高い傾向にあるので、検索結果上位の方に今後も利用したいような Web ページが多いことがわかる。特に「映画」といったような、大手企業が Web サイトを作ることにより利益を見込めるような分野では、多くの大手企業が Web サイトや Web サービスを作成し、内容も充実している場合が多い。こういった Web ページが Google 検索において上位にランキングしたため、「映画」における適合率が高い値になったのではないかと考えられる。そのため、このような分野においては、Google 検索で有益な Web コンテンツを手軽に発見することができると考えられる。

セレクトブックマにおいても、同様に検索結果の順位が高い程、高い適合率が得られているため、検索結果の上位に今後も利用したいような Web ページが多いことがわかる。ただし、Google 検索とは、検索結果の傾向が異なり、検索結果には大手企業が作成したような Web サイトはほとんど出てこない。そのため、個人で作成したような Web サイトの中から、体系だった知識を得られる Web ページや有益な Web サービスを発見したい場合、セレクトブックマが有効なのではないかと考えられる。

(2) 再現率、得点の合計値からの考察

「java」、「健康」においては、人数から適合文書を作成した場合も得点から適合文書を作成した場合も、セレクトブックマの再現率がもっとも高い。さらに、得点の合計値も上位 10 件、上位 20 件、上位 30 件においてセレクトブックマがもっとも高い得点となっている。このことから、セレクトブックマにおいて、検索結果上位 30 件中には、Google 検索と比較してもブックマーク数順と比較しても、体系だった知識を得られる Web ページや有益な Web サービスの数が多いのではないかと考えられる。

「映画」においては、Google 検索において、すべての再現率、得点の合計値がもっとも高い値となっている。このことから、「映画」においては、Google 検索において、検索結果上位 30 件中に体系だった知識を得られる Web ページや有益な Web サービスの数がもっとも多いと考えられる。

4.5. 考察と課題

ここで、本研究における考察と課題について述べる。

4.5.1. 実験結果からの考察

実験結果より、セレクトブックマにおいて「java」「健康」という単語に関しては、Google 検索やブックマーク数順に並べたものと比較して、もっとも高い適合率、再現率を得ることができた。このことから、セレクトブックマにおいて、今後も見たいような体系だった知識を得られる Web ページや有益な Web コンテンツを手軽に収集できる可能性が高いと考えられる。

だが、逆に「映画」という単語においては、Google 検索と比較して適合率、再現率は大幅に低い値となった。この原因の 1 つとして、大手企業の Web サイトの満足度が依存していると考えられる。「映画」という分野においては、大手企業の Web サイトが多数あり、内容も充実している。そのため、Google 検索でそういった大手企業の Web サイトが検索結果の上位に表示される傾向にあり、被験者がそういった Web サイトを選択することが多かった。個人で作成している映画関連の Web ページでも、いくつか有益なレビューサイトやまとめサイトなどが存在するが、その数自体が少ない。そのため「映画」という単語においては、Google 検索において、高い適合率・再現率となり、セレクトブックマで低い適合率・再現率となったと考えられる。

4.5.2. 時間情報を利用することの有効性

セレクトブックマでは、ソーシャルブックマークが長い間ブックマークされてるかどうかという指標を利用することによって、一時的に面白い情報をフィルタリングし、体系だった知識を得られる Web ページやいつ見ても有益な Web コンテンツを中心に取得することを目的としている。セレクトブックマとブックマーク数順を比較した場合、セレクトブックマにおいて高い適合率・再現率を得ることができた。また、特に上位 10 件の適合率が、ブックマーク数順と比較してセレクトブックマで高い値になっていることから時間情報を利用することによって、一時的に必要とされる情報をフィルタリングできていることがわかった。

4.5.3. 本研究の有効性

検索結果や実験結果から、Google 検索と比較してセレクトブックマでは、大手企業が作成している Web サイトだけでユーザの満足が得られない分野において、有効性が高いと考えられる。個人で作成している Web ページは、数多くあり玉石混合である。そのため、既存の PageRank などの手法では、Web コンテンツ作成者がリンクを貼った場合に、PageRank があがるため、コンテンツ作成者しか Web ページの評価をすることができない上、コンテンツ作成者が良いと思った Web ページにどんどんリンクを貼って行くような Web サイトやコンテンツ作成者の数はそれほど多くない。これに対して、ソーシャルブックマークを利用した場合、コンテンツ消費者がブックマー

クをするという簡単な行為によって、コンテンツが評価される。そのため、個人が作成した玉石混合の Web ページ群の中から、特に有益な Web ページを発見するのに有用であると考えられる。

4.5.4. 課題

現時点のセレクトブックマでは、収集したデータ量不足の問題やはてなブックマークユーザのデータの偏りの問題がある。また、タグ数というものを指標にしているため、はてなブックマークユーザが、タグをつけにくい分野や単語での検索は、十分な結果が得られない場合が多い。同様にマイナーな分野では十分な結果が得られない場合が多く、万能な情報収集支援ツールとはなっていない。

メジャーな分野においても、現在利用しているセレクトブックマのランキングロジックでは、体系だった知識を得られる Web ページや有益な Web サービス以外を完全にフィルタリングできているは言えない。セレクトブックマの検索結果や被験者の選んだ Web ページやを見てもわかるように、上位 10 件の Web ページの中にも体系だった知識を得られる Web ページや有益な Web サービスではないものが存在する。さらに、上位 10 件の中にあるような Web ページよりも、あきらかに被験者にとって人気の高く有益である可能性の高い Web サービスなどが、上位 20 件以降になっている場合もある。

また、セレクトブックマは、Web 検索システムというより、どちらかという、Web からの情報収集支援システムという位置付けとなっている。そのため、現在のようにセレクトブックマの Web サイトが一つあり、そこで検索するというアーキテクチャが良いとは限らない。情報収集支援システムとして、もっとよいインタフェースやアーキテクチャを模索していく必要がある。

4.6. まとめ

本章では、ソーシャルブックマークの以下の特性を利用し、それに基づいた長期的に利用できる情報を検索するシステムの提案・実装・評価を行った。

- 短い期間しかブックマークされない Web ページ
一時的に必要とされる種類の Web ページが多い。
- 長い期間多くのユーザからブックマークされる Web ページ
いつ見ても有用な情報を得られる種類の Web ページが多い。

上述の特性に基づき、ソーシャルブックマークにおいて、ブックマークされる期間の長短という指標を利用した、情報収集支援システム「セレクトブックマ」を提案した。「セレクトブックマ」の評価実験として「セレクトブックマ」、「Google 検索」、「はてなブックマークのデータをブックマーク数順に並べたもの」の 3 つを比較した。その結果、「Google 検索」と比較した場合、特に大手企業の作成した Web サイトでユーザが満足できず、個人が作成したような Web ページをユーザが

利用している場合において「セレクトブックマ」が有用であることがわかった．さらに、「はてなブックマークのデータをブックマーク数順に並べたもの」と比較した場合にも，特に検索結果上位 10 件において，より適合率の高い値を得ることができた．ゆえに，時間情報を利用することによって，一時的に利用する Web ページをフィルタリングし，今後も利用したいと思うような体系だった知識を得られる Web ページや有益な Web コンテンツを取得することができるとわかった．

以上の結果から，本研究のソーシャルブックマークの時間情報を利用した情報フィルタリング手法により，個人が作成している大量の Web コンテンツ群の中から，体系だった知識を得られる Web ページや有益な Web サービスを中心に取得する一つの手法を提案できた．

第5章 長期的な検索キーワードの提示

概要

本章では，長期的な検索キーワードを提示するシステム”MasteryEye”について述べる。

5.1. 背景

ここでは、長期的な検索キーワードを提示するシステムの開発に至った背景や目的について述べる。

5.1.1. 背景

近年、Web 検索エンジンを利用して必要な情報を探すという行為が一般的に行われているが、Web の情報はどんどん膨大になってきているため、容易に必要な情報にたどり着けない場合もある。検索エンジンのアルゴリズムは、様々な手法が提案されているが ([72]、[86]) Google の PageRank[7]) が特に有名である。PageRank は主に被リンクを用いて、人気のページをランキングしている。また、この他にも流行の情報を発見する手法も多く提案されている。[29] は、時系列データからバースト^{*13}を検出する手法を提案しているし、こういった手法を応用して、実際に流行の情報を発見する研究も行われている ([87])。

このように人気の情報や流行の情報の発見する様々な手法が広く提案・利用されている一方で、長期的に利用され続けるような情報に特化して取得する手法はほとんど存在しない。これに対して筆者らは、長期的に利用されているということは、長期的に見て有用な情報である可能性が高いと考えた。また、近年では製品やコンテンツの寿命が短命化してきているというという背景があり、このことから長期的に利用できるモノや情報は見つけにくくなってきているのではないかと考えた。

ただ、一口に長期的に利用され続ける情報といっても、それは Web ページだけにとどまらずに、多くの情報が考えられる。そこで筆者らは、本論文でどれだけ長期的に利用されているかを表す指標である長期度という指標を提案し、これを検索キーワードに適用して、長期的に検索され続けている検索キーワードを取得するシステムを提案する。本論文で、検索キーワードを対象とした理由と長期的に利用する検索キーワードを取得する目的については、5.1.2 に示す。

5.1.2. 目的

Web 上からユーザが有用な情報を検索するためには、適切な検索キーワードを入力する必要がある。だが、誰もが適切な検索キーワードを選択できるとは限らないし、検索キーワードが思い浮かばないという問題も存在する。

こういった問題を解決するために、検索クエリを拡張する手法の研究 ([30]) や検索キーワードに関連するキーワード一覧を提示する研究 ([88]) が行われてきた。また、Google サジェスト^{*14}などのサービスも提供されてきた。これらは、検索キーワードの類似性や検索回数の多さを利用して、推薦するキーワードを選出している。

こういった背景に対して、筆者らは長期的に検索され続けている検索キーワードは定番の検索キーワードであり、有用な検索キーワードである可能性が高いと考えた。また、自分が不慣れな

分野を調べるときに、その分野で長期的に利用されている検索キーワードが分かれば、その分野を体系的に調べることが可能になるのではないかと考えた。

そこで本研究では、キーワードがどれだけ長期的に検索され続けているかに着目し、調べたい分野の中で定番の検索キーワードを取得する手法を提案する。

5.2. MasteryEye の提案

5.2.1. 長期度の計算手法

情報がどれだけ長期的に利用され続けているかの指標として、長期度という指標を定義し、長期度を計算する手法を提案する。長期度を計算するために、ここでは検索回数の時系列データを利用する。最初に、実際の時系列データを回数の多い順に並べる。次に、実際の時系列データの値の大きさから、べき乗則に基づくデータを生成する。そして、実際の時系列データとべき乗則に基づくデータの差分の大きさを計算し、これを長期度とする。実際に数式で示すと以下のようになる。

$$\text{長期度} = \sum_{k=1}^n (\alpha_k - \beta_k) \quad (5.1)$$

$$\alpha_k : \text{単位時間ごとの検索回数やアクセス回数} \quad (5.2)$$

$$\beta_k : \text{べき乗則に基づくデータ} \quad (5.3)$$

ここでは、両対数グラフにおいて縦軸の最大値と横軸の最大値を結ぶ直線を描くようなデータとし、べき乗則の指数係数の値を設定した。

ここで、このような手法で長期度を求める理論を説明する。まず、複雑な条件に基づいて形成されるデータは、べき乗則に基づくという仮定 ([31]) や単語の使用頻度は、べき乗則に基づく ([32]) などの理論を利用して、多くのキーワードは、単位時間ごとのアクセス数がべき乗則に基づく可能性が高いと仮定する。ここで、説明のため、横軸を単位時間ごとのアクセス数の順位、縦軸をアクセス数として、グラフを書く (図 5.1)。べき乗則に基づくデータは、図の実線のようになり、長期的に利用されるデータの分布は、破線のようになる。べき乗則に基づく分布は、ロングテールとなる。このため、単位時間ごとのアクセス数順に並べた場合、長期的に利用される分布は、べき乗則に基づく分布と比較して、大きな値となる期間が長いということが分かる。逆に、急速に流行って廃れていくものは、グラフの落ち方が急激になるため、べき乗則に基づく分布と比較して小さな値となる期間が長い。

以上のような法則を利用して、実際のデータとべき乗則に基づくデータの差分が大きければ大きいほど、長期度が高いとすることができる。また、分野によって、べき乗則との差分の値も変わってくる。そのため長期度は、べき乗則との差分の値が同じ分野においてどの程度大きいかで計算する。

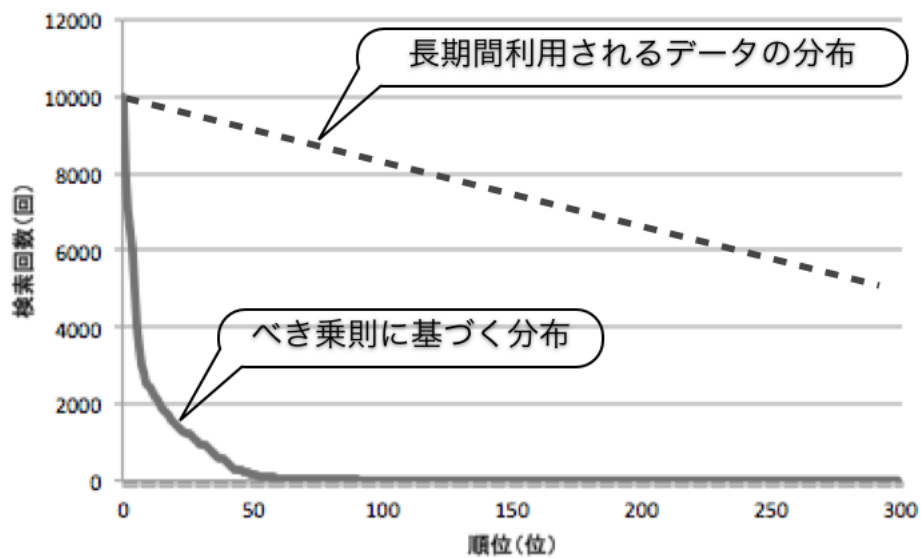


図 5.1: べき乗則のデータと長期的利用データ

なお、指数係数の値を変数とし、この値を小さな値にすることで長期的に利用されるデータの分布を近似し、指数係数の大きさを長期度の指標とすることも考えられる。だが、今回の場合は長期間検索され続ける検索キーワードの中には、平均値に近い検索回数が多くなり、べき乗則に従わないものも存在するため、べき乗則に基づく分布との差分とした。

5.2.2. 関連キーワードの収集

関連キーワード取得は、既存の手法である、Lingua-JA-Expand[23] を利用する。Lingua-JA-Expand は、以下の手順で関連キーワードを取得している。

- キーワードを受け取る
- Yahoo Search API を利用して Yahoo 検索結果のスニペット^{*15} を取得
- TF-IDF による計算を利用して、関連キーワードと関連度を取得

この手法で様々なキーワードで試してみたところ、関連キーワードの精度に不足を感じた。具体的に言うと、スニペットを利用しているため、どうしても取得したキーワードの特徴を表す単語が関連キーワードとして多く出てきてしまう。そのため、この関連キーワード一覧からキーワードを取得する逆引きによって、関連キーワードを取得した方が、より本システムに向いている関連キーワードが取得できるのではないかと考えた。

そのため、Wikipedia から取得したキーワードと Lingua-JA-Expand を利用して、関連キーワードデータベースを作成した。関連キーワードデータベースのデータ量は以下の表 5.1 のようになっている。

表 5.1: 関連キーワードデータベースのデータ量

キーワード	関連キーワード
約 120 万キーワード	約 2500 万キーワード

5.2.3. 検索キーワードの時系列データ収集

検索キーワードの時系列データは、Google Insights for Search[24] というサービスのデータをクロールする。Google Insights for Search では、2004 年以降の Google 検索におけるキーワードの 1 週間ごとの検索量を提示している。

この検索キーワードの時系列データに 2.1 で示した長期度の計算方式を適用して、長期度を計算する。そして、長期度が高い順にランキングする。

5.2.4. システム表示

本システムでは、ユーザがキーワードを入力するとその分野に関連するキーワード一覧が表示される。図 5.2 は、例として「yahoo」というキーワードを入力して関連キーワード一覧を取得した画面である。画面には、長期度が高い順から上位 10 件のキーワードを提示する。また、検索キーワードは、リンクとして、リンク先はそのキーワードでの検索結果とする。図の横棒は、長期度の大きさを表している。

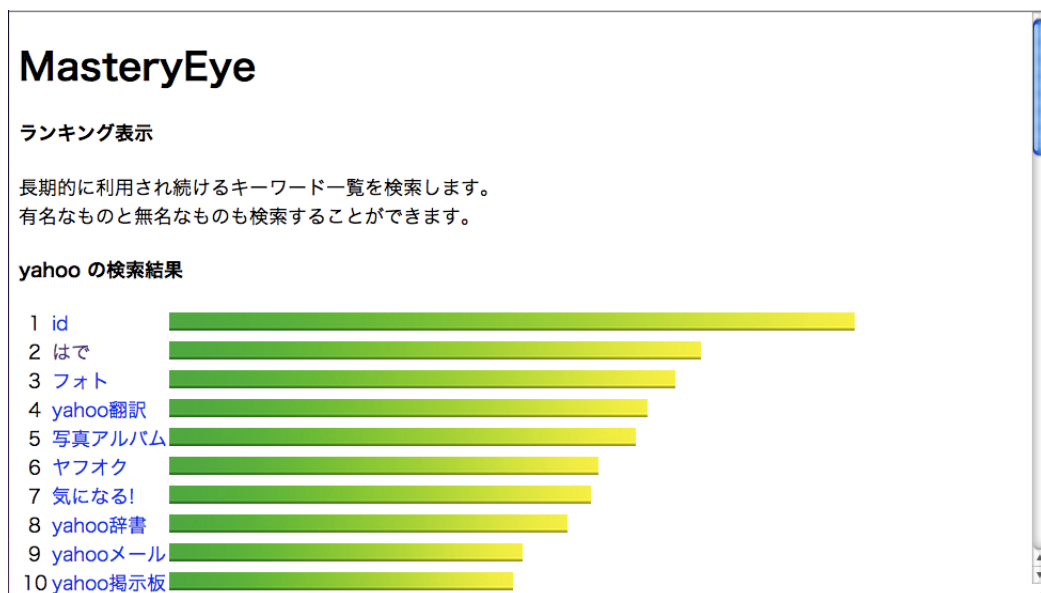


図 5.2: 「yahoo」での検索結果画面

5.3. 評価実験

5.3.1. 実験目的

評価実験は、以下 3 つを示すことを目的として行った。

1. 長期度の計算手法の正当性
2. 長期間利用するものは重要である可能性が高いこと
3. 開発したシステムの有用性

5.3.2. 実験方法

評価実験では、以下の 3 種類の手法で関連キーワードを取得して比較実験を行った。

手法 1：提案手法で長期度が高かったもの

手法 2：提案手法で長期度が低かったもの

手法 3：既存手法 (reflexa[25])

手法 1 と手法 2 については、提案手法で長期度が高かった関連キーワード、低かった関連キーワードをそれぞれ上位 10 件ずつ利用した。また、手法 3 の reflexa とは、連想検索エンジンで、入力したキーワードと関連の深いキーワードを提示するシステムである。既存手法に reflexa を選んだ理由は、一般公開されているシステムであるからと、非常に多くのキーワードに対して関連キーワードを取得できるからである。

以上の 3 種類の手法それぞれについて、Google で検索回数の多い上位 4 位のキーワードを利用して実験を行った。検索回数の多いキーワードは、Google Insights for Search の情報を参照した (2012 年 4 月 5 日)。

実験を行うにあたって、Google での検索回数が多いキーワードを選んだ理由は、本システムを利用するユーザ層として想定しているのが、検索キーワードがあまり思い浮かばないようなユーザを想定しているからである。そういった検索リテラシーがあまり高くないユーザは、少しマイナーなキーワードよりも、Google での検索回数が多いようなメジャーなキーワードの方が思い浮かぶ可能性も高いであろうと考えた。そして、こういったメジャーなキーワードから関連するキーワードを取得して、それらの元のキーワードに関連していて、かつ、長期間利用され続けているかを評価したいと考えた。例えば、以下の検索回数上位 2 位の「動画」というキーワードに関して、本システムで「動画」と入力すると動画に関連する検索キーワードの中で、長期間検索され続けているキーワードが提示できれば、動画に関連する定番の検索キーワードを知ることができるのではないかと考えた。

- 1 位 : Yahoo
- 2 位 : 動画
- 3 位 : YouTube
- 4 位 : 画像

以上の4つのキーワードの関連キーワードを10キーワードずつ3種類の手法で、合計120キーワード取得した。重複したキーワードを1つにまとめて、合計116キーワードをランダムに並べて被験者に提示した。被験者の属性は、表5.2に示す。被験者に合計116キーワードそれぞれに対して、以下の3つの質問項目に当てはまるものを選んでもらった。

質問1：この中であなたが知っているキーワードを選んで下さい

質問2：この中であなたが長期間利用してきたキーワードを選んで下さい

質問3：この中であなたが重要だと思うキーワードを選んで下さい

これらは複数選択可とし、選択数に制限は設けなかった。

表 5.2: 被験者属性

人数	20 名
性別	男性：12 名、女性：8 名
年代	0 代：15 名、30 代：2 名、40 代：2 名、50 代：1 名

5.3.3. 実験結果

評価実験のために取得した関連キーワード一覧は、表5.9に示す。取得した関連キーワード一覧を見てみると、提案手法で長期度が高いものは、一部関連性が低そうなキーワードもあるが、ほとんどのキーワードは、元のキーワードに関連していることが分かる。提案手法で長期度が低いものは、関連性が高いキーワードが多いが、あまり多くの人から利用されていないキーワードも多いことが分かる。また、少し詳しく見てみると、例えば「動画」に対しての「アメーバビジョン」や「YouTube」に対しての「字幕.in」など、以前少し流行ったが、現在はあまり利用されていないサービスなども多い。relexaについては、元のキーワードと関連していて、なおかつ一般的なキーワードが多いが、特に「Yahoo」に関連するキーワードで、あまり利用されていないキーワードが多いことが分かる。

これらに対して、評価実験を行った結果を以下に示していく。評価実験で各質問項目について、ユーザが選択したキーワード数の評価を行った。3種類の手法に対して、選択されたキーワード数の合計数を表5.3に示す。ここで、被験者のうち誰か1人でも選択したキーワードに対して、選択されたキーワードとした。

表 5.3: 選択されたキーワードの数

	手法 1	手法 2	手法 3
質問 1:知っている	35	24	24
質問 2:長期間利用	32	14	21
質問 3:重要である	29	12	18

これらの選択されたキーワードの数が確率分布に基づく期待値と有意差があるかどうかカイ二乗検定 *16 によって評価を行った。

最初に、3 種類の手法に対して、各質問で選択されたキーワードが期待値と有意差があるかどうか、カイ二乗検定を行なった。質問 1 の「知っているかどうか」、質問 2 の「長期間利用しているかどうか」、質問 3 の「重要かどうか」の 3 種類の質問項目に対して行った (表 5.4 ~ 表 5.6)。

その結果、「知っているかどうか」については、p 値は 0.233 (小数点第 4 以下四捨五入) となり、有意差は求められなかった。「長期間利用しているかどうか」については、p 値は 0.025 (小数点第 4 以下四捨五入) となり、有意水準 5 % 以下で有意差を求めることができた。「重要かどうか」については、p 値は 0.023 (小数点第 4 以下四捨五入) となりこちらも有意水準 5 % 以下で有意差を求めることができた。

表 5.4: カイ二乗検定 (知っているかどうか)

	手法 1	手法 2	手法 3	計
観測度数	35	24	24	83
期待度数	27.67	27.67	27.67	83
p 値	0.233			

表 5.5: カイ二乗検定 (長期間利用かどうか)

	手法 1	手法 2	手法 3	計
観測度数	32	14	21	67
期待度数	22.33	22.33	22.33	67
p 値	0.025			

次に、長期度の計算手法の正当性を示すため、提案手法で長期度が高かったもの (手法 1) と低かったもの (手法 2) で取得したキーワードに対して、「長期間利用しているかどうか」に選択されたキーワード数の比較を行った (表 5.7)。

その結果、選択されたキーワード数は手法 1 の方が多くなった。また、p 値は 0.0004 となり、有意水準 1 % 以下で有意差を求めることができた。このことから、システムで取得した長期度が高いものの方が、長期度が低いものよりも長期間利用されているキーワードであることが分かる。

また、長期間利用されているものが重要であるかどうかを調べるために、質問 2 の「長期間利用しているかどうか」と質問 3 の「重要であるかどうか」の質問項目に回答されたキーワードの

表 5.6: カイ二乗検定 (重要かどうか)

	手法 1	手法 2	手法 3	計
観測度数	29	12	18	59
期待度数	19.67	19.67	19.67	59
p 値	0.023			

表 5.7: 長期間利用かどうか：長期度高と長期度低の比較 (カイ二乗検定)

	選択あり：手法 1	選択あり：手法 2	選択なし：手法 1	選択なし：手法 2	計
観測度数	32	14	7	26	79
期待度数	23	23	16.5	16.5	79
p 値	0.0004				

関係性を調査した。質問 2 のみ回答されたもの、質問 3 のみ回答されたもの、質問 2 と質問 3 に重複して回答されたもの、どちらにも回答されなかったものの 4 種類に対してカイ二乗検定を行なって、有意差を求めた (表 5.8)。

表 5.8: 長期間利用と重要なものの関係 (カイ二乗検定)

	長期	重要	長期かつ重要	選択なし	計
観測度数	15	7	52	44	118
期待度数	33.5	25.5	33.5	25.5	118
p 値	3.04×10^{-10}				

その結果、「長期間利用している」かつ「重要である」キーワード数が期待値より多くなった。また、p 値は 3.04×10^{-10} となり有意水準 1 % 以下で有意差を求めることができた。この結果から、長期間利用しているものは重要である可能性が高いということが言える。

5.4. 課題と展望

ここで、本研究の課題と今後の展望について述べる。

5.4.1. 課題

検索キーワードは、主に必要であったり有用であったりする情報を検索するためのものだが、長期間利用される情報と言った場合、様々な種類の情報が考えられる。本論文では、検索キーワードのみを対象としているが、そのキーワードを利用して、本当に長期間利用できる Web ページを発見できる可能性が高くなるのか評価を行う必要がある。さらに、システムで取得したキーワードを元にして長期間利用できる、Web 上の情報を提示するように改良していくことも考えられる。

表 5.9: 取得した関連キーワード一覧

元キーワード	長期間の高いもの	長期間の低いもの	reflexa
Yahoo	id	リローンチ	OL 蔡桃桂
	はで	月刊 4b	ポアロのあと何分あるの?
	フォト	アリババグループ	すときゃ!
	yahoo 翻訳	東京めたりっく	漫畫
	写真アルバム	津乃村真子	関戸優希
	ヤフオク	キャロル・パーツ	Pheonix
	気になる!	リアルタイム検索	新浪
	yahoo 辞書	ヤフコメ	FQDN
	yahoo メール	みんなの検定	寶輔
	yahoo 掲示板	ポケモンガーデン	蔡桃
動画	動画サイト	アニタン	請
	動画編集	いじめ動画	原画
	サンプル動画	佳山三花	MPEG-4
	YouTube	日本動画協会	作画監督
	3gp	車載動画	YouTube
	デジタル動画	ニコニコ動画物語	東映動画
	ようつべ	動画大陸	H.264
	動画プレイヤー	ニコニコ組曲	コーデック
	日本動画	アメーバビジョン	MPEG
	ニコ動	グロ動画	DivX
YouTube	orbit	the 八犬伝	Stage6
	動画	楽珍トリオ	ニコニコ動画
	話	ヒピラくん	Google
	ようつべ	ユーチューブ xl	DivX
	3d 動画	陳士駿	PayPal
	ビデオソフト	ろうきゅうぶ	MOCO
	c-8	アキラブ	チャド・ハーリー
	fooooo	著作権管理	GUBA
	woopie	字幕.in	Veoh
	動画投稿サイト	私が恋愛できない理由	政見放送
画像	画像編集	死亡時画像診断	Ferrari
	画像掲示板	磁気共鳴画像	darkgreen
	おもしろ画像	衛星画像	ビットマップ
	画像診断	蓮画像	トランスミッション
	デジタル画像	綾波セナ	ビットマップ画像
	画像安定装置	ビットマップ画像	ピクセル
	画像処理	バカ画像	ストラット
	おすすめ画像	グロ画像	JPEG
	画像ビューア	レタッチソフト	クロスオーバー SVC
	画像認識	画像作成ソフト	トールワゴン

そして、実際にユーザが長期間利用され続けている検索キーワード一覧を取得する利用シーンに合わせて、ユーザインタフェースを整え、Web 上にサービスとして公開すべきである。

また現状のシステムでは、入力するキーワードに対しても、提示するキーワードに対しても、キーワードのゆらぎの問題が存在する。この問題にも対応していく必要がある。さらに、今回の実験では、被験者 20 人に対して、Google で検索数が多い 4 つのキーワードに関連するキーワードの評価実験を行った。だが、本来はもっと多く被験者に対して、もっと多くのさまざまなキーワードを選定して実験を行うことが望ましい。そこで、今後はユーザインタフェースを整え Web サービスとして公開した上で、実際に多くの人に利用してもらい評価を行なっていく。

5.4.2. 今後の展望

今回は 1 つの実装例として、検索キーワードの関連キーワードを取得するシステムを実装して実験した。だが、世の中にはさまざまな種類の情報があふれている。本研究の最終目的は、長期間利用される情報を取得する手法を確立し、さまざまな情報に対して応用可能にすることである。

そのため今後は、他にもさまざまな情報に対して、提案した長期度計算手法を適用して、長期間利用される情報を取得するシステムを開発していく。これは例えば、長期間多くのユーザから再生され続けている動画や長期間レビューされ続けている商品などを取得するシステムである。こういったさまざまなシステムを統合して、最終的に長期間利用する様々な情報を取得するシステムを開発していく。

5.5. まとめ

情報検索手法や情報フィルタリング手法として、多くの手法が提案されているが、長期的に利用されるという利用のされ方に着目して情報を探す手法はほとんど存在しない。また、Web 上の情報を検索する上で、適切な検索キーワードが思いつかないという問題もある。

そこで、ここでは、長期的に検索され続ける検索キーワードを取得するシステムを開発した。開発したシステムに対して評価実験を行った結果、提案手法により、一般的な検索キーワードと比較して、長期的に利用されているキーワードを取得することができた。

第6章 長期的なファイルと Web 履歴へのアクセス

概要

本章では、自分が長期的に利用してきたファイルと Web ページへ手軽にアクセスするためのシステム「タームストア」について述べる。

- 6.1. 背景
- 6.2. 長期的に利用されるファイルへのアクセス手法の提案
- 6.3. 評価実験
- 6.4. 考察
- 6.5. まとめ

第7章 関連研究

概要

本章では、本研究に関連する研究領域について整理し、本研究の特徴や位置づけについて述べる。

7.1. 情報検索

7.1.1. 情報検索のアルゴリズム

検索アルゴリズムを提案することによって、検索システムの性能や精度を向上させようとしている研究は多い。本研究では、新しい検索アルゴリズムを提案することを主眼としていないが、ある程度既存の検索アルゴリズムも参考に必要がある。

古くから情報検索の代表的な検索モデルとして利用されてきた手法として、ベクトル空間モデル [43] がある。ベクトル空間モデルとは、ベクトルの内積を利用して文書の類似度を計算する手法である。ベクトル空間モデルの手法を改善するために、潜在的意味インデキシング (LSI) [55][56] という手法が幅広く利用されている。潜在的意味インデキシングとは、高次元の空間にある文書ベクトルを低次元の空間へ射影することにより、検索精度の向上を図る技術である。

また、情報検索の代表的な技術として確率モデルが幅広く利用されている。確率モデルの中で、文字列の確率的な傾向を利用して検索精度を上げるために利用されるモデルとして、確率的言語モデル [95] がある。確率的言語モデルとは、言語の文字列の傾向を数値的に表すモデルである。確率的言語モデルの中でも、代表的なアルゴリズムとして、N-gram モデルがある。N-gram モデルとは、単語の生起確率を表したモデルである。例えば、N 個の単語があった場合、それらを 1 次元の単語列に配置する。そして、各単語の生起確率は、直前の (N-1) 単語に依存する。ベイズ確率に基づくベイズ推定も幅広く利用されている。この中でも文書を分類するための手法として、単純ベイズ分類器 [57] が幅広く応用されている。

機会学習の技術を利用して、検索精度を向上させるための技術として、サポートベクターマシン (SVM) [58][59] が、幅広く利用されている。サポートベクターマシンは、教師あり学習を用いる識別手法の 1 つであり、現在知られている多くの手法の中で最も認識性能が優れた学習モデルの 1 つである。

様々な情報検索手法や情報検索のアルゴリズムについては、文献 [28][72][77] などで詳しく解説されている。

7.1.2. 関連ページの検索

ある Web ページに関連したページを検索するための手法も提案されている。Chakrabarti らは、Web ページの全テキスト情報を対象として、Web ページ間の類似度を判定する手法を提案している [47]。栗原らは、ユーザのアクセス履歴を利用して、関連する Web ページ集合を検出するシステムを提案している [101]。この研究では、ユーザのアクセス履歴として、クリックしたリンクの URL とラベル文字を利用している。珍田らは、弱い紐帯を手がかりとして関連する Web ページを抽出する手法を提案している [100]。

7.1.3. ソーシャルな検索

近年では、ソーシャルな情報や人知情報などを利用した検索手法が注目をあびている。本研究でもソーシャルブックマークの情報を利用した検索システムを提案している。

Agichtein らは、大規模な実験により、ユーザの振る舞いの情報を利用することによって、Web 検索の結果を改善することが可能であることを示している [41]。村田らは、検索エンジンのクリックログと検索結果として表示されるページのタイトル、スニペットを利用して、検索結果ランキングを生成するシステムを提案している [?]。White らは、ユーザのアクセス履歴を利用して、ページの重要度を決定し、人気の高いページをランキングする手法を提案している [40]。

また、ソーシャルブックマークの情報を利用した検索システムも提案されている。Xu らは、ソーシャルブックマークデータを利用したパーソナライズド検索手法を提案している [17]。パーソナライズド検索とは、それぞれの検索者にとって最適な情報を提示しようとする検索手法である。さらに Xu らは、ソーシャルブックマークのタグ情報を利用することによって、パーソナライズド検索を自動的に評価する手法も提案している。Yanbe らは、ソーシャルブックマークのブックマーク数を新たな指標 SBRank を提案し、PageRank と SBRank を統合して、Web 検索ランキング精度の向上を計っている [19][20]。さらに、Yanbe らは、ソーシャルブックマークデータを分析することで、周期的に利用されるコンテンツが人気となる時期などを予測し、検索が行われた時期に応じた検索結果のランキングを可能にする時期連動型ランキング手法を提案している [?]。この研究は、ソーシャルブックマークの時間情報を利用している点で、第 4 章で提案したセレクトブックマの研究に近いが、周期的な情報に着眼点を置いているか、長期的な情報に着眼点を置いているかという点が異なっている。Takahashi らは、ソーシャルブックマークデータの時間データを利用して、鮮度の高い Web ページを取得する検索手法を提案している [15]。ここでは、ブックマーク日時の散らばりの大きさから、Web ページの賞味期限を判定して、賞味期限を過ぎていない鮮度の高い Web ページを取得している。この研究も、ソーシャルブックマークの時間データに着目した情報検索手法を提案しているため、セレクトブックマの研究と近い。ただし、この研究では、鮮度の高い Web ページを取得することを目的としているのに対して、セレクトブックマでは、体系だった知識をえられ、長期に渡って役立つ Web ページを取得することを目的としている点が異なっている。

7.1.4. 検索インタフェース

新しい検索インタフェースを提案したり、インタラクティブな検索システムを提案することによって、ユーザにとってより利用しやすい検索システムを提案しようとする研究事例も多い。近年、よく導入される検索インタフェースとして、インクリメンタル検索がある。インクリメンタル検索とは、入力のためごとに即座に検索候補を提示する検索手法である。インクリメンタル検索に関連する研究は、1990 年代から幅広くされており [49][50][51]、インクリメンタル検索の有効性も主張されている [52]。インクリメンタル検索が日本語に応用されたり [106]、音声インクリメ

ンタル検索システムの提案 [107] もされている。

検索システムのユーザ同士のコミュニケーションを利用して、検索効率の向上を目指す研究もある。Morris らは、検索キーワードを共有したりなど、数人の小規模なユーザ間での協調作業により、情報検索の効率向上を目指している [42]。松井らは、情報と人を瞬時に発見することができ、検索サービスとソーシャルサービスの双方の利点を活用できる検索システムを提案している [93]。このシステムを利用することにより、ユーザは、検索結果と同時にページの閲覧者も発見できる。さらに、閲覧者とチャットによるコミュニケーションを行うこともできる。

渡邉らは、眺めるインタフェースを用いた検索システムを実装している [?]。このシステムでは、あらかじめ興味のあるキーワードを入力しておくことにより、入力したキーワード同士で AND 検索を行ない、検索結果が提示される。そのため、ユーザは画面を眺めているだけで、興味のある情報を得ることができる。吉田らは、検索結果ページ中に現れる重要語を話題語として抽出し、マウスによる単語のドラッグ&ドロップによって、AND 検索、OR 検索、NOT 検索を直感的に行い、検索結果の再ランキングをできる手法を提案している [78]。

7.1.5. 様々な情報を利用した検索システム

上述の他にも様々な情報を利用した検索手法が提案されている。ここでは、様々な検索手法について述べる。

検索ユーザが検索システムに情報をフィードバックすることにより、検索結果を改善する手法として、適合性フィードバック [60] がある。適合性フィードバックとは、ユーザが、得られた検索結果のうち、どの文書が検索意図に適合し、どの文書が適合しないかを検索システムに教えることにより、システムの検索精度を改善する手法である。

Liu らは、検索者の過去の検索履歴を利用することによって、パーソナライズド検索を行う手法を提案している [54]。また、検索履歴を利用したパーソナライズド検索は、Google 検索 [22] でも利用されている。

増井らは、ユーザが見ている情報に近い情報を利用し、ユーザが関連リンクをクリックするだけで連想的な検索を行うことができる近傍検索システムを提案している [108]。

ユーザの行動履歴を利用した検索手法も提案されている。川田らは、ユーザの位置、時間、性別、年齢、行動の状態などの情報を匿名性を保持したまま利用することによって、似た行動をするユーザを探し、そこからコンテンツを提供する検索システムを提案している [105]。

中村らは、動画に付与されたアノテーション情報を利用して、印象に基づく動画検索システムを提案している [?]

旭らは、2つの検索キーワードが与えられたときに、それぞれの検索キーワードの間の情報を検索するシステムを提案している [104]。

莊司らは、発見した Web ページを読んだ際に、読み手がどのように感じるかをクエリとして入力可能な Web 情報検索システムを提案している [?]

7.2. 情報フィルタリング

7.2.1. 情報フィルタリングのアルゴリズム

情報フィルタリングの手法を分類すると、主にコンテンツに基づくフィルタリングと協調フィルタリングに分類できる [13] .

コンテンツに基づくフィルタリングとは、コンテンツの内容に基づき情報の取捨選択を行う手法のことである。コンテンツに基づくフィルタリングに関する研究の例として、文献 [79] , [82] があげられる。

協調フィルタリングとは、ユーザの過去の行動を記録し、そのユーザと類似した行動をとっているユーザの嗜好情報から、ユーザに情報を推薦する手法のことである。協調フィルタリングに関する論文の例としては、文献 [11] , [2] , [81] があげられる。また、協調フィルタリングの実用例としてもっとも有名なのが Amazon の推薦システムである [1] .

また、近年では、コンテンツに基づくフィルタリングと協調フィルタリングを組み合わせたハイブリッド法 [76] に関する研究もさかんになってきている [6] .

7.2.2. ソーシャルな情報フィルタリング

ソーシャルブックマークのデータを利用した情報推薦に関する研究があり、大別すると Web ページを推薦するものとソーシャルブックマークユーザを推薦するものがある。

Web ページを推薦する研究として、Niwa らは、タグのクラスタリングをおこなうことによりタグの表記ゆれの問題の解決をはかり、ユーザのブックマーク情報からユーザの趣向に沿った Web ページの推薦をおこなう手法を提案している [10] . また、Sasaki らは、タグを表象とする Web コンテンツ群の類似性に基づいた Web コンテンツ推薦システムを提案している [74] .

ソーシャルブックマークユーザを推薦する研究として、白土らは、ソーシャルブックマークユーザのブックマーク情報からユーザの関連度を解析した結果から興味の類似したユーザを推薦し、ネットワーク図として表示するシステムを構築している [71] . 大力らは、ソーシャルブックマークユーザの中のイノベータ、いわば ブックマーカーに注目した情報推薦手法を提案している [70] .

7.2.3. 情報フィルタリングのインタフェース

7.2.4. 様々な情報フィルタリングシステム

情報フィルタリングの研究も情報検索と同様に、様々な分野に適用されている。位置情報を利用した情報推薦の研究、画像のフィルタリングに関する研究、メールのフィルタリングに関する研究、ニュースサイトのフィルタリングに関する研究、ブログのフィルタリングに関する研究など多方面にわたる研究が実施されている。

7.3. ソーシャルブックマークデータの分析

ソーシャルブックマークや Folksonomy の分析に関する研究としては、以下のような研究事例がある。

Golder らは、ソーシャルブックマークのユーザやタグ、ブックマークの性質について分析し、各 Web ページに対する各タグの出現頻度は一定値に収束することを証明している [4]。Paul らは、del.icio.us のデータを収集して、ソーシャルブックマークが Web 検索において大きな改革を起こせるかどうか検討している。その結果、現状ではソーシャルブックマークのデータ量不足の問題やタグのゆらぎの問題から、現時点のデータでは、Web 検索に関して劇的な改革は起こせないが、今後ソーシャルブックマークのデータ量が急激に増えたりした場合は、Web 検索に革新を起こせる可能性があると結論づけている [5]。川中らは、あるタグと共起関係の強いタグを取得し、出現時期の早いほうを親タグとする手法を用いることによって、タグの時系列の関係性をグラフ化している [83]。

7.4. 検索キーワード

検索結果ではなく、検索キーワードを推薦したり、改善したりすることによって、Web からの検索を楽にしようという研究事例も多い。本研究でも長期的な検索キーワードを取得する手法を提案している。

クエリ拡張の手法として、Robertson らが提案した、検索結果から適当な単語を抽出して検索質問拡張を行う手法の一つである RSV(Robertson's Selection Value) がある [14]。正田らは、ユーザが与えたクエリでの検索結果上位 R 件を適合文書、それ以下を不適合文書として、上述の RSV を用いてクエリ拡張を行い、新たなクエリの重みにより初期の検索結果をソートする手法を提案している [9]。Zhang らは、検索エンジンのキーワード補完のように関連するクエリを推薦するために、クエリの分類と関連クエリを発見するための手法を提案している [38]。この研究では、キーワード間の TFIDF 類似度の代わりに、SF(Search Frequency) を用いた SFIDF を基づく Content Similarity の線形結合を類似度としたクラスタリング手法を提案している。Cao らは、ユーザーの入力したクエリとクリックされた URL をクラスタリングしたものと同一セッション内で入力されたクエリの関連性に基づいてクエリ推薦を行う手法を提案している [39]。大塚らは、大規模なアクセスログから抽出されたユーザの Web 検索の検索単語と、それによって閲覧したページを解析することで、関連語を抽出する手法を提案している [75]。大石らは、ユーザの意図する検索クエリを生成するための方法として、センテンス間の距離に注目した関連単語抽出アルゴリズムを提案する。このアルゴリズムは重要な語の近くに出現する単語は重要であるという考えに基づいている [80]。今井らは、入力した検索クエリが多義語の場合、選択された URL に基いて推薦クエリを生成する手法を提案している [89]。木田らは、時間とともに変化するクエリ間類似度を利用して、検索クエリをクラスタリングする手法を提案している [90]。甲谷らは、検索クエリとクリックされた URL 情報を利用して、Web サイトに到達するために頻繁に使用されるクエリを発

見することで、クエリ推薦を行う手法を提案している [91]。安川らは、検索クエリのログから検索語の関連語を取得して、関連語のみに限定した単語方向のクラスタを生成し、クラスタと Web ページ群との対応をユーザに提示する Web 検索の手法を提案している [92]。

7.5. ファイルアクセスに関する研究

デスクトップ上のファイルアクセスを容易にするための研究も多くなされている。本研究でも、第 6 章で長期的なファイルにアクセスするための手法を提案している。

Rekimoto らは、PC 上のファイルは、すべてデスクトップに置かれ、時間とともに消えていくシステムを提案している [36]。このシステムでファイルを探す場合は、過去のデスクトップに遡っていく。そして、そのときデスクトップに置かれていたファイル群と一緒にファイルを見つけることができる。

Soules らは、複数のファイル同士の関連性を利用して、ファイル検索の精度をあげる手法を提案している [37]。

7.6. 時間情報の利用

時間情報を利用することによって、有益な情報を見つけようとする研究例も多い。その中でも、流行を発見するための研究は、これまでたくさん行われてきた。

Kleinberg らは、トレンドを発見するために、時系列データの中からバーストを検出する手法を提案している [29]。Roy らは、大量のテキストデータから流行の単語を抽出する研究を行っている [35]。Ishikawa らは、時系列データを利用して、文書集合中のトレンドをとらえるための可視化システムを提案している [44]。

流行現象のモデル化を行なっている研究例もある。Granovetter は、しきい値モデルを提案し、他者がどれほどの割合で流行を採用しているのか、商品を購入しているのか、という他者の採用率によって、流行現象が起こったかどうかを判定している [45]。さらに、このしきい値モデルを利用し、実際の流行現象を分析している [46]。松田は、しきい値モデルを用いて、流行がはやったり廃れたりを繰り返す流行の循環、流行が不規則にはやる流行のカオス的挙動、一度流行がはやると急激に廃れ、その後全く顧みられなくなる流行の一過性の現象などを明らかにしている [97]。中山は、流行・普及現象を、流行を採用するのか否か、流行商品を購入するのか否か、という個々人の二者択一の離散選択の集合と捉え、その離散選択をロジットモデルにより定式化を行っている。このモデルは、同調や差別化という他者の影響及び購入価格を含む採用するためのコストを考慮し、流行の採用率を算出するものである [96]。河根らは、商品などの流行現象の開始と終了を求めるために、2 次元しきい値分布を利用したモデルを提案している [98]。また、服部らは、インターネット小売業などについて、商品の流行度を反映するランキングに関する数理モデルについて解説している。そして、商品の売り上げのような社外秘に属するデータ分布をランキングと

いう公開されたデータから分析する仕組みを提案している [99]。このように流行現象のモデル化を行ったり、流行の情報を発見したりする研究は、これまでたくさん実施されてきた。

また、情報検索や情報フィルタリングを行うために、情報が利用される時間に着目した研究例も存在する。

Dubinko らは、Flickr のタグが時間とともに発展する様を視覚化する手法を提案している [?]。視覚化手法として、川をメタファにしたものと滝をメタファにしたものを提案している。

未来の情報を予測したり、検索したりする手法も提案されている。Wolfers らは、群衆の叡智を利用して、近い未来を予測する手法を提案している [48]。河合らは、年度を明示的に含む文を検索対象とすることによって、未来情報と過去情報を検索する検索エンジンを提案している [102]。吉田らは、Web ニュースの情報を利用して、未来情報の年表を自動で構築する手法を提案している [103]。

竹井らは、情報のライフサイクルについて研究しており、文書のライフサイクルと情報のライフサイクルを対比させることにより、価値ある情報のライフサイクル管理について提案している [85]。

第8章 長期的検索の考察と展望

概要

本章では、本研究の考察と展望について述べる。

8.1. 考察

本研究では、長期的に利用される情報に特化して検索する手法である「長期的検索」を提案した。また、本研究では、長期的検索における応用例として、Web ページ、検索キーワード、ファイルの3つの情報に応用した。そして、具体的なアプローチとして、セレクトブックマ、MasteryEye、タームストアの3つのシステムを開発した。本研究のアプローチの特徴は、以下のようにまとめることができる。

- 人の行動から得られる時間情報の分析
- 長期度の提案
- 長期的な情報に着目した検索システムの提案
- 長期的な情報の有用性

8.1.1. 人の行動から得られる時間情報の分析

本研究では、人々の行動から得られる時間情報を分析することによって、この時間情報を検索システムに活かせないか考えた。そこで、ソーシャルブックマークでの人々からブックマークされた時間情報と Google 検索における各検索キーワードの検索された時間情報を取得し、利用した。

最初に、ソーシャルブックマークのブックマーク時間データを分析した。その結果、長い期間新しくブックマークされ続ける Web ページは、長期的に利用される種類の Web ページの割合が多いことが分かった。また、短い期間しかブックマークされていない Web ページは、一時的に利用される種類の Web ページの割合が多いことが分かった。このことから、ソーシャルブックマークという人々のブックマーク行為の時間情報から、長期的に利用される Web ページを取得できることが分かった。

次に、Google 検索での検索回数の時系列データを分析した。この分析では、長期的に多くの人から検索され続けていたキーワードと短期的にのみ多くの人から検索されたキーワードを比較した。その結果、長期的に検索され続けていたキーワードは、短期的にのみ検索されたキーワードと比較して、長期的に利用できるキーワードが多いことが分かった。このことから、人々の検索履歴を利用することによって、長期的に利用できるキーワードを取得できることが分かった。

このように、履歴情報やブックマーク情報など、人の行動から得られる情報の時間情報を利用することによって、長期的に利用できる可能性の高い情報を見つけることができることが分かった。

8.1.2. 長期度の提案

本研究では、その情報がどれだけ長期的に利用されているかの指標である長期度の提案をした。長期度の計算方式として、ブックマーク日数に基づいた計算手法とべき乗則とのズレの大きさに基づいた計算手法を提案した。

セレクトブックマでは、ブックマーク日数に基づき、長期度を計算し、評価実験を行った。その結果、長期度の低い Web ページよりも、長期度が高い Web ページの方が、有意に被検者が今後も利用したいページが多かった（要確認）。

MasteryEye では、べき乗則とのズレの大きさに基づいた計算手法により、長期度を計算し、評価実験を行った。その結果、長期度の低い検索キーワードよりも、長期度の高い検索キーワードの方が有意に被検者が長期間利用してきたキーワードが多かった。

このことから、提案したブックマーク日数に基づいた計算手法とべき乗則とのズレの大きさに基づいた計算手法双方の長期度計算手法において、長期的な情報が取得できていることを示すことができた。

8.1.3. 長期的な情報に着目した検索システム

これまでに述べた、分析結果や長期度の計算手法に基づき、長期的な情報に特化して検索するシステムを開発した。セレクトブックマと MasteryEye である。

セレクトブックマは、Google 検索とブックマーク数順検索の 2 種類の手法と比較した。その結果、Google 検索と比較して、「java」、「健康」などの分野で検索した場合、より今後も利用できる Web ページを取得することができた。また、ブックマーク数順検索と比較して、ある程度長期的な Web ページが多かった。ソーシャルブックマークのデータ量やデータ特性の影響により有効な分野は限られるが、Google 検索と比較して、有効性を示すことができた。また、長期的な Web ページに特化して検索するシステムとしての有効性も示すことができた。

MasteryEye は、ある分野について少し詳しく調べたいときに、有用な検索キーワードを取得することを目的としている。MasteryEye を既存の関連キーワード取得システム reflexa と比較実験をした。reflexa と比較して、被検者が長期的に利用していおり、かつ、被検者が重要であると思うキーワードを取得することができた。こういった結果から、MasteryEye の有効性を示すことができた。

8.1.4. 長期的な情報の有用性

長期度の高い Web ページ 今後も利用したい Web ページが多かった（要確認）

長期的に利用するようなタイプの Web ページ 今後も利用したい Web ページが多かった（要確認）

このため、本研究の長期度の計算方法の有効性と長期的に利用する Web ページの有用性を示すことができた。

長期度の高い検索キーワード 長期的に利用しているキーワード

長期度の高い検索キーワード 被検者にとって重要なキーワード

長期的に利用している検索キーワード 被検者にとって重要なキーワード

このため、MasteryEye での長期度の計算手法の有効性と長期的に利用されている検索キーワードの有用性を示すことができた。

このように、Web ページと検索キーワードという 2 つの異なった情報に関して、長期的な情報の有用性が高いことが分かった。また、長期度の高い情報は、長期的に利用できる可能性が高いことが分かった。以上のようなことから、長期的な情報の有用性が高いと考えられる。

8.2. 議論

8.2.1. 長期的な情報の特徴

8.2.2. 長期的検索の有効性

8.2.3. 長期的検索の未来

8.3. 課題

本研究では、長期的な Web ページと長期的な検索キーワードを取得する手法を提案したが、それぞれで異なった長期度の計算手法を用いている。だが、様々な情報に対する利用期間、利用頻度、利用回数などのデータを分析することによって、様々な情報の長期度計算手法を統一することができる可能性もある。例えば、係数などのパラメータを変更することにより、各情報の利用期間、利用頻度、利用回数などに合わせた係数を自動的に算出できる可能性もある。今後はこういった分析を行い、多くの情報に適用できる長期度計算手法を提案していく。

そのためには、もっと様々なデータを利用し、より詳細な分析や実験を行う必要もある。本研究では、ソーシャルブックマークのデータと Google 検索キーワードの時系列利用データの 2 つのデータで分析や実験を行ってきたが、もっと様々なデータに対して、分析や実験を行い、統合的に利用できる長期度の計算手法の提案をしていく。さらに、様々なデータに対して、長期度の計算手法の有効性や長期的な情報と長期的検索の有用性を示す必要がある。

様々なデータに対して分析したり、実験を行うためには、大規模で長期的な利用時間情報などを取得する必要があるが、このような情報を取得することが難しいという問題も存在する。例えば、Web サービスなどの利用時間情報などは、各 Web サービスを提供している企業が持っているが、これらのデータが公開されているとは限らない。公開されていない場合は、その企業と直接交渉してデータを貰うか、もしくはその企業が長期的検索のシステムを開発しない限り、実現できない。近年では公開されているデータも増えてきて、その結果、本研究でもソーシャルブックマークや Google の検索キーワード情報を取得することができた。だが、まだまだ利用などの時間情報が公開されているサービスは少ない。また、個人の利用履歴などの情報は、記録することも難しいという場合もある。例えば、Web ページやファイルへのアクセス履歴や検索キーワードの履歴などは、公開したくないと思っている人も多い。こういったプライバシーの問題に配慮しつつ、上手くデータを記録し、利用する必要がある。また、現状では入手できる情報を元に長期度の提案と長期的検索の開発を行なっていくしかない。

視覚化手法やアクセス手法を検討する必要もある。現状では、セレクトブックマも MasteryEye も一般的な検索システムと類似したインタフェースになっている。要するにキーワードを入力して検索することによって、長期的な情報を検索結果として提示する仕組みである。だが、長期的な情報を検索するのに特化して考えた場合、それ特有のインタフェースがあることで、よりユーザにとって利用のしやすいシステムにできると考えられる。

8.4. 展望

長期的検索の今後の展望として、Web ページや検索キーワード、ファイルだけでなく他の様々な情報に応用していくことができる。例えば、動画、レシピ、wikipedia、論文などの情報に応用することによって、長期的に利用できる動画、レシピ、wikipedia のページ、論文などを取得することができる。また、本や商品などに応用することによって、ロングセラーとなっている本や商品を検索によって見つけることができるようになる。さらに、このような様々な情報を統合して長期的検索ができるようになれば、非常に有用なサービスになるのではないかと考えられる。

また、本研究で提唱している「長期的な情報」や「長期的検索」といったような概念をもっと世の中に広め、長期的な情報の有用性を理解してもらうことができれば、様々なサービスに長期的検索の仕組みが導入されていくかもしれない。現状では、検索結果のソート方法として、人気順や新着順といったようなものが主となっているが、筆者は、長期順のようなソート方法があると、埋もれてしまった良コンテンツの発掘に役立つと考える。

さらに、5 年後 10 年後といったような未来における長期的検索の可能性について考えてみる。近年、情報やコンテンツの短命化と玉石混交化が進んでいるが、Web や PC 上の情報量は今後も増大していくため、この傾向はしばらく続くと予測される。そうすると、アクセス可能な情報はどんどん増大していくが、その中の情報は賞味期限の過ぎたような情報や玉石混交の情報が大量に存在することになる。そのときに、現状の検索手法だけだと、長期的に利用できる情報へのアクセスがさらに難しくなってしまう。また、現状ではまだまだ情報社会の歴史が浅く、蓄積されているデータは、長くて 5 年～10 年程度のものが多い。だが、5 年後、10 年後の未来では、蓄積されている利用時間情報の期間や量が増大していると予測できる。そのため、筆者は、こういった情報を利用した長期的検索というビジョンが、将来もっと重要になってくるのではないかと考える。

第9章 結論

概要

本研究の成果をまとめるとともに，本論文の総括をおこなう．

9.1. 研究の成果

9.2. 総括

謝辞

本研究を進めるにあたり，修士時代から5年間，主査として研究を基礎から指導して頂いた慶應義塾大学 安村通晃教授に深く感謝いたします。安村先生には，非常に伸び伸びと研究をさせて頂き，5年間幸せで充実した研究生活を送ることができました。

また，副査として，かつ，インタラクショナルデザインプロジェクトに参加して頂き，本研究に関して多くの的確なコメントとアドバイスをして頂きました慶應義塾大学 増井俊之教授に深く感謝いたします。増井先生の研究に対する姿勢や考え方などには非常に共感できるところがあり，今後自分の研究活動における考え方の基礎となっていくことでしょう。

また，副査として，本研究に関して，多くのコメントをいただき，特に評価実験に関してご指導いただきました慶應義塾大学 小川克彦教授および慶應義塾 萩野達也教授に感謝いたします。小川先生から頂いた本研究に対する意見やコメントにより，本研究は大きな影響を受けています。

さらに，所属する研究会において，多くのコメントとアドバイスや研究に関する相談に乗って頂きました慶應義塾大学 樋口文人先生に感謝いたします。

また，同じ安村研究室に所属する秋山氏，山本氏，白崎氏，吉椿氏に感謝いたします。秋山氏には，ブレインストーミングなどで面白いアイデアをたくさん出して頂き，研究に対する大きな刺激を受けました。山本氏には，Webに関する新しい技術などの話をたくさんして頂き，技術に対する大きな刺激を受けました。白崎氏とは，よく夜遅くまで一緒に研究室に残って研究を行い，よく車で湘南台駅まで送って頂きました。吉椿氏には，研究室にいつも明るい雰囲気をもたらして頂き，研究室としてのまとまりやつながりを与えて頂きました。

安村研究室およびインタラクショナルデザインプロジェクトに所属するみなさんには，日頃から本研究に関するアドバイスをして頂いたり，研究に関する議論をして頂き，大変感謝いたします。

最後に，長い研究生活をいつも暖かい目で見守ってくれた両親と弟に感謝いたします。本当にありがとうございました。

2012年1月

慶應義塾大学 大学院政策・メディア研究科博士課程3年

上野 大樹

研究業績

原著論文

1. 上野大樹, 樋口文人, 安村通晃, ”長期間利用される検索キーワード検出手法の提案”, 情報通信学会誌, Vol.30 No.2, pp57-65, September 2012.
2. 上野大樹, 樋口文人, 安村通晃, ”ソーシャルブックマークの時間スケールに着目した長期間利用する Web ページ収集支援システムの研究”, 情報社会学会誌, Vol.5 No.1, pp43-52, June 2010.

国際会議発表

1. Taiki Ueno, Michiaki Yasumura, ”Using The Time Course of Social Bookmarking to Enhance Search”, IADIS International Conference WWW/INTERNET 2011, November 2011.
2. Taiki Ueno, Michiaki Yasumura, ”MINDROAD - Route Memory Support System using a Smart Phone”, International Conference on Pervasive and Embedded Computing and Communication Systems(PECCS 2011), March 2011.
3. Taiki Ueno, Michiaki Yasumura, ”Social Search with Shared Keywords”, IADIS International Conference WWW/INTERNET 2009, November 2009.

国内学会発表

1. 上野大樹, 樋口文人, 安村通晃, ”時間情報を利用したロングセラー情報発見システム”, 情報知識学会誌, Vol.22 No.2, pp107-114, May 2012.
2. 上野大樹, 安村通晃, ”VoiTwi : スマートフォンのジェスチャー操作を用いた音声 Twitter システムの提案と実装”, インタラクシオン 2011 論文集, pp401-404, March 2011.
3. 上野大樹, 安村通晃, ”スマートフォンを用いた音声 Twitter システムの試作と研究”, 第 56 回福祉情報工学研究会, IEICE Technical Report WIT2010-57, January 2011.
4. 樋口文人, 渡邊恵太, 上野大樹, 菅原圭, 安村通晃 ”入力デバイスの特徴づけに関する考察”, ヒューマンインタフェースシンポジウム 2009, September 2009.

5. 上野大樹, 安村通晃, ”セレクトブックマ: ソーシャルブックマークの時間情報を用いた情報フィルタリング検索”, 第 50 回プログラミング・シンポジウム pp 9-16, January 2009.
6. 上野大樹, 泉谷昭二, 足立武彦 ”負性抵抗増大回路付き高周波水晶発振回路の狭帯域化”, 電気学会電子回路研究会資料, ECT-04-92, December 2004.
7. 上野大樹, 竹内圭太, 泉谷昭二, 足立武彦 ”高周波水晶発振回路のバイアス電流による周波数制御”, 電気学会電子回路研究会資料, ECT-04-91, December 2004.
8. 上野大樹, 泉谷昭二, 足立武彦 ”負性抵抗増大回路を付加した 100MHz ~ 数 GHz 帯コルピッツ発振回路の動作解析”, 電気学会電子回路研究会資料, ECT-04-84, November 2004.
9. 上野大樹, 野村記央, 青柳勇二, 泉谷昭二, 足立武彦 ”負性抵抗増大回路を付加した 600MHz-SAW 発振器の試作”, 電子情報通信学会 2004 ソサイエティ大会, A-11-3, September 2004.

参考文献

- [1] Amazon.com. Amazon. <http://www.amazon.com/>, (参照 2010-01-11) .
- [2] JS. Breese, D. Heckerman, C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the 14th conference on Uncertainty in artificial intelligence (UAI-98)*, pp. 43-52, 1998.
- [3] Digg. Digg. <http://digg.com/>, (参照 2010-01-11) .
- [4] Scott A. Golder, Bernardo A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science Volume 32 Issue 2*, pp.198-208, 2006.
- [5] Paul Heymann, Georgia Koutrika, and Hector Garcia-Molina. Can social bookmarking improve web search? *Proceedings of the international conference on Web search and web data mining (WSDM '08)*, pp. 195-206, 2008.
- [6] Jon Kleinberg and Mark Sandler. Using mixture models for collaborative filtering. *Journal of Computer and System Sciences Volume 74*, pp. 49-69, 2008.
- [7] Page L, Brin S, Motwaniand R, Winograd T. The pagerank citation ranking: Bringing order to the web. *Technical report, Stanford Digital Library Technologies Project*, 1998.
- [8] S. Loeb, D. Terry. Information filtering. *Communications of the ACM, Vol. 35, No. 12*, pp. 26-81, 1992.
- [9] T. Masada, T. Kanazawa, A. Takasu, J. Adachi. Improving web search by query expansion with a small number of terms. *Proceedings of the 5th NTCIR Workshop Meeting*, pp.486-493, 2005.
- [10] S. Niwa, T. Doi, S. Honiden. Web page recommender system based on folksonomy mining. *Proceedings of the 3rd International Conference on Information Technology : New Generations (ITNG '06)*, pp. 388-393, 2006.
- [11] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, John Riedl. Grouplens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW'94)*, pp. 175-186, 1994.

- [12] P. Resnick, H. Varian. Recommender systems. *Communications of the ACM*, Vol. 40, No. 3, pp. 56-89, 1997.
- [13] D. Riecken. Personalized views of personalization. *Communications of the ACM*, Vol. 43, No. 8, pp. 26-158, 2000.
- [14] S.E. Robertson. On term selection for query expansion. *Journal of documentation*, Vol. 46, No. 4, pp. 359-364, 1990.
- [15] Tsubasa Takahashi, Hiroyuki Kitagawa. S-bits: Social bookmarking induced topic search. *Proceedings of the 9th International Conference on Web-Age Information Management (WAIM2008)*, pp. 25-30, 2008.
- [16] Twitter. Twitter. <https://twitter.com/>, (参照 2010-01-11) .
- [17] Shengliang Xu, Shenghua Bao, Ben Fei, Zhong Su, Yong Yu. Exploring folksonomy for personalized search. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '08)*, pp.155-162, 2008.
- [18] Yahoo! Delicious. <http://delicious.com/>, (参照 2010-01-11) .
- [19] Yusuke Yanbe, Adam Jatowt, Satoshi Nakamura, Katsumi Tanaka. Can social bookmarking enhance search in the web? *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries(JCDL 2007)*, pp. 107-116, 2007.
- [20] Yusuke Yanbe, Adam Jatowt, Satoshi Nakamura, Katsumi Tanaka. Towards improving web search by utilizing social bookmarks. *Proceedings of the 7th International Conference on Web Engineering(ICWE 2007)*, 2007.
- [21] Microsoft. Bing. <http://www.bing.com/> (参照 2012-12-9) .
- [22] Google. Google. <http://www.google.com/> (参照 2012-12-9) .
- [23] Takeshi Miki. Lingua-JA-Expand. <http://search.cpan.org/~miki/Lingua-JA-Expand/> (参照 2012-12-13) .
- [24] Google. Google Insights for Search. <http://www.google.com/insights/search/> (参照 2012-5-31) .
- [25] Preferred Infrastructure. 連想検索エンジン reflexa. <http://labs.preferred.jp/reflexa/> (参照 2012-12-13) .
- [26] Google. YouTube. <http://www.youtube.com/> (参照 2012-12-13) .

- [27] Yahoo!. Flickr. <http://www.flickr.com/> (参照 2012-12-13) .
- [28] Ricardo Baeza-yates, Berthier Ribeiro-neto. Modern Information Retrieval. Addison Wesley, pp. 544, 1999.
- [29] Jon Kleinberg. Bursty and hierarchical structure in streams. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'02)*, pp. 91-101, 2002.
- [30] Yonggang Qiu, Hans-Peter Frei. Concept Based Query Expansion. *Proceedings of ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR'93)*, pp. 160-169, 1993.
- [31] Albert-laszlo Barabasi, Jennifer Frangos. Linked: The New Science Of Networks Science Of Networks. Basic Books, pp. 288, 2002.
- [32] Andrew D. Booth. A Law of Occurrences for Words of Low Frequency. *Information and Control Vol. 10, No. 4*, pp. 386-393, 1967.
- [33] Jorg Rech. Discovering trends in software engineering with google trend. *ACM Sigsoft Software Engineering Notes, vol. 32, No. 2*, pp. 1-2, 2007.
- [34] Simon Hermann. Dynamics of Price Elasticity and Brand Life Cycle. *Journal of Marketing Research* ,16(4), pp. 439-45, 1979.
- [35] Soma Roy, David Gevry, William M. Pottenger. Methodologies for trend detection in textual datamining. *Textmine '02 Workshop, Second SIAM International Conference on Data Mining*, 2002.
- [36] Jun Rekimoto. Time-machine computing: a time-centric approach for the information environment. *Proceedings of the 12th annual ACM symposium on User interface software and technology (UIST'99)*, pp. 45-54, 1999.
- [37] Craig A. N. Soules, Gregory R. Ganger. Connections: using context to enhance file search. *Proceedings of the twentieth ACM symposium on Operating systems principles*, pp. 119-132, 2005.
- [38] Z. Zhang, O. Nasraoui. Mining search engine query logs for query recommendation. *Proceedings of the 15th international conference on World Wide Web (WWW'06)*, pp. 1039-1040, 2006.

- [39] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen and H. Li. Context-aware query suggestion by mining click-through and session data. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and datamining (SIGKDD'08)*, pp. 875-883, 2008.
- [40] Ryen W. White, Mikhail Bilenko, Silviu Cucerzan. Studying the use of popular destinations to enhance web search interaction. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR'07)*, pp. 159-166, 2007.
- [41] Eugene Agichtein, Eric Brill, Susan Dumais. Improving web search ranking by incorporating user behavior information. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information (SIGIR'06)*, pp. 19-26, 2006.
- [42] Meredith Ringel Morris, Eric Horvitz. SearchTogether: an interface for collaborative web search. *Proceedings of the 20th annual ACM symposium on User interface software and technology (UIST'07)*, pp. 3-12, 2007.
- [43] G. Salton, A. Wong, C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM, Vol. 18, No. 11*, pp. 613-620, 1975.
- [44] Yoshiharu Ishikawa, Mikine Hasegawa. T-Scroll: Visualizing Trends in a Time-Series of Documents for Interactive User Exploration. *European Conference on Digital Libraries (ECDL'07)*, pp. 235-246, 2007.
- [45] Granovetter M. Threshold Models of Collective Behavior. *American Journal of Sociology, Vol.83(6)*, pp. 1420-43, 1978.
- [46] Granovetter M, R Soong. Threshold Models of Interpersonal Effects in Consumer Demand. *Journal of Economic Behavior and Organization*, 7, pp. 83-99, 1986.
- [47] Soumen Chakrabarti, Byron Dom, Prabhakar Raghavan, Sridhar Rajagopalan, David Gibson, Jon M. Kleinberg. Automatic Resource Compilation by Analyzing Hyperlink Structure and Associated Text. *Computer Networks and Isdn Systems - CN, Vol. 30, No. 1-7*, pp. 65-74, 1998.
- [48] Justin Wolfers, Eric Zitzewitz. Prediction Markets. *Financial Analysts Journal - FINANCIAL J, Vol. 18, No. 2*, pp. 107-126, 2004.
- [49] Ben Shneiderman. Dynamic Queries for Visual Information Seeking. *IEEE Software - SOFTWARE, Vol. 11, No. 6*, pp. 70-77, 1994.

- [50] Christopher Ahlberg, Ben Shneiderman. Visual information seeking: tight coupling of dynamic query filters with starfield displays. *Computer Human Interaction (CHI'94)*, pp. 222-317, 1994.
- [51] Joseph C. Pemberton, Richard E. Korf. Incremental Search Algorithms for Real-time Decision Making. *International Conference on Automated Planning and Scheduling/Artificial Intelligence Planning Systems - ICAPS(AIPS)*, pp. 140-145, 1994.
- [52] J Raskin. The humane interface: new directions for designing interactive systems. Addison-Wesley Professional, pp. 256, 2000.
- [53] Glen Jeh, Jennifer Widom. Scaling personalized web search. *World Wide Web Conference Series (WWW'03)*, pp. 271-279, 2003.
- [54] Fang Liu, Clement T. Yu, Weiyi Meng. Personalized Web Search For Improving Retrieval Effectiveness. *IEEE Transactions on Knowledge and Data Engineering (TKDE'04)* , Vol. 16, No. 1, pp. 28-40, 2004.
- [55] Scott C. Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, Richard A. Harshman. Indexing by Latent Semantic Analysis. *Journal of The American Society for Information Science and Technology - JASIS* , Vol. 41, No. 6, pp. 391-407, 1990.
- [56] Michael W. Berry, Susan T. Dumais, Gavin W. O'Brien. Using Linear Algebra for Intelligent Information Retrieval. *Siam Review - SIAM REV* , Vol. 37, No. 4, 1995.
- [57] Andrew K. McCallum, Kamal Nigam. A comparison of event models for Naive Bayes text classification. *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*, pp. 41-48, 1998.
- [58] B. Scholkopf, C. J. C. Burges, A. J. Smola. Advances in kernel methods : support vector learning. *The MIT Press*, 1999.
- [59] N. Cristianini, J. Shawe-taylor. An introduction to support vector machines and other kernel-based learning methods. *Cambridge University Press*, 2000.
- [60] J. J. Roccio. Relevance feedback in information retrieval. *The Smart Retrieval System: Experiments in Automatic Document Processing*, ed. G.Salton, pp. 313-323, 1971.
- [61] Paul Resnick, Hal R. Varian. Recommender systems. *TCommunications of The ACM - CACM* , Vol. 40, No. 3, pp. 56-58, 1997.
- [62] Thorsten Joachims, Dayne Freitag. WebWatcher: A Tour Guide for the World Wide Web. *TIInternational Joint Conference on Artificial Intelligence (IJCAI'96)* , 1996.

- [63] Liren Chen, Katia P. Sycara. WebMate: a personal agent for browsing and searching. *Autonomous Agents & Multiagent Systems International Conference on Autonomous Agents (AAMAS'98)*, pp. 132-139, 1998.
- [64] サイボウズ・ラボ株式会社. Pathtraq. <http://pathtraq.com/>, (参照 2010-01-11).
- [65] ニフティ株式会社. @nifty クリップ. <http://clip.nifty.com/>, (参照 2010-01-11).
- [66] ヤフー株式会社. Yahoo!ブックマーク. <http://bookmarks.yahoo.co.jp/all>, (参照 2010-01-11).
- [67] 株式会社 EC ナビ. Buzzurl [バザール]. <http://buzzurl.jp/>, (参照 2010-01-11).
- [68] 株式会社はてな. はてなブックマーク. <http://b.hatena.ne.jp/>, (参照 2010-01-11).
- [69] 株式会社ライブドア. livedoor クリップ - ソーシャルブックマーク. <http://clip.livedoor.com/>, (参照 2010-01-11).
- [70] 大力慶祐, 大向一輝, 武田英明. ソーシャルブックマークにおけるイノベータに注目した情報推薦手法の提案. 人工知能学会第 22 回全国大会 (JSAI2008), 2008.
- [71] 白土慧, 吉井伸一郎, 古川正志. ソーシャルブックマークサービスを利用した情報レコメンデーション. 情報処理学会研究報告. ICS, 知能と複雑系, pp.15-20, 2006.
- [72] 北研二, 津田和彦, 獅々堀正幹. 情報検索アルゴリズム. 共立出版, 2002.
- [73] 津田宏治. サポートベクターマシンとは何か. 電子情報通信学会誌 83(6), pp. 460-466, 2000.
- [74] 佐々木祥, 宮田高道, 稲積泰宏, 小林亜樹, 酒井善則. Social Bookmark におけるコンテンツクラスタ間の類似度を用いた web コンテンツ推薦システム. 情報処理学会論文誌 データベース 48, pp. 14-27, 2007.
- [75] 大塚真吾, 喜連川優. 大規模アクセスログを用いた検索語想起支援システムの提案とその評価. 情報処理学会研究報告. データベース・システム研究会報告, pp. 25-32, 2006.
- [76] 神鷹敏弘. 推薦システムのアルゴリズム (3). 人工知能学会誌, pp. 248-263, 2008.
- [77] 兼宗進. 検索エンジンの検索アルゴリズム. 情報の科学と技術. Vol. 54, No. 2, pp. 78-83, 2004.
- [78] 吉田大我, 小山聡, 中村聡史, 田中克己. Web 検索結果におけるキーワード出現相関の可視化と対話的な質問変換. 電子情報通信学会第 18 回データ工学ワークショップ (DEWS2007), pp. 1-8, 2007.

- [79] 松本知弥子, 馬強, 田中克己. Web ページのローカル度検出に基づく情報フィルタリング. 情報処理学会研究報告, Vol. 125, No. 36(DBWS2001), pp. 273-280, 2001.
- [80] 大石哲也, 倉元俊介, 峯恒憲, 長谷川隆三, 藤田博, 越村三幸. 関連単語抽出アルゴリズムを用いた web 検索クエリの生成. 電子情報通信学会論文誌 D Vol. J92-D, No. 3, pp. 281-292, 2009.
- [81] 福原知宏. 協調フィルタリングに関する研究動向. 奈良先端科学技術大学院大学学内研究報告資料, 1998.
- [82] 河合由起子, 官上大輔, 田中克己. 個人の選好に基づく複数ニュースサイトの記事収集・閲覧システム. 情報処理学会論文誌. データベース 46, pp. 14-25, 2005.
- [83] 川中翔, 佐藤周行. ソーシャルブックマークにおけるタグの派生関係の抽出. 修士論文, 2009.
- [84] 上原征彦, 福田亮. ロングセラーの論理とその戦略. マーケティング・ジャーナル, 第 59 号, pp. 6-15.
- [85] 竹井潔. 情報の価値とライフサイクル管理. 聖学院大学論叢 17(1), pp. 11-31.
- [86] 兼宗進. 検索エンジンの検索アルゴリズム. 情報の科学と技術, Vol. 54, No. 2, pp. 78-83.
- [87] 奥村学, 南野朋之, 藤木稔明, 鈴木泰裕. blog ページの自動収集と監視に基づくテキストマイニング. 人工知能学会, セマンティックウェブとオントロジー研究会, SIG-SWO-A401-01.
- [88] 大塚真吾, 豊田正史, 喜連川優. 大域ウェブアクセスログを用いた関連語の発見法に関する一考察. 情報処理学会論文誌: データベース, Vol. 46, No. SIG8(TOD26), pp. 82-92.
- [89] 今井良太, 戸田浩之, 関口裕一郎, 望月崇由, 鈴木智也, 今井桂子. Web 検索サービスにおける多義的なクエリ推薦手法. 日本データベース学会論文誌 Vol. 9, No. 1, pp. 7-11, 2010.
- [90] 木田巧, 豊田正史, 喜連川優. トレンドを考慮した検索クエリの分類手法の一検討. 日本データベース学会論文誌 9(1), pp. 12-17, 2010.
- [91] 甲谷優, 湯本高行, 小山聡, 田中克己. Web ページに対する典型的なクエリの発見. 情報処理学会研究報告. データベース・システム研究会報告 2007(65), pp. 49-54, 2007.
- [92] 安川美智子, 横尾英俊. クエリログから獲得した関連語のクラスタリングに基づく Web 検索. 電子情報通信学会論文誌. D, 情報・システム J90-D(2), pp. 269-280, 2007.
- [93] 松井優也, 河合由起子. 人と情報の検索および相互作用を目指したソーシャルサーチシステムの研究開発. 日本ソフトウェア科学会論文誌, コンピュータソフトウェア, Vol. 28, No.4, pp. 196-205, 2011.

- [94] 森田昌宏, 速水治夫. 情報フィルタリングシステム: 情報洪水への処方箋. 情報処理 37(8), pp. 751-758, 1996.
- [95] 北 研二, 辻井潤一. 言語と計算 (4) 確率的言語モデル. 東京大学出版会, pp. 230, 1999.
- [96] 中山 晶一郎. ロジット型離散選択に基づいた流行現象モデル. 理論と方法 18(1), pp. 89-102, 2003.
- [97] 松田光司. 流行現象とベクトル型閾値モデル. 理論と方法 11(2), pp. 113-128, 1996.
- [98] 河根拓文, 村重淳, 合原一幸. 2次元しきい値分布を利用した流行現象の数理モデルとその解析. 電子情報通信学会論文誌. A, 基礎・境界 J83-A(3), pp. 284-293, 2000.
- [99] 服部哲弥. 確率的ランキング: 流行度の順位付けとロングテール分析. オペレーションズ・リサーチ: 経営の科学 57(6), pp. 302-307, 2012.
- [100] 珍田計幸, 大沢英一. 弱い紐帯の概念を利用した WWW 上からの情報抽出手法の提案. 電子情報通信学会技術研究報告. AI, 人工知能と知識処理 108(456), pp. 99-104, 2009.
- [101] 栗原聡, 廣津登志夫, 高田敏弘, 明石修, 菅原俊治. ユーザのアクセス履歴を利用した類似 Web サイト発見手法. 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理 J87-D-I(6), pp. 741-742, 2004.
- [102] 河合英紀, Jatowt, A., 田中克己, 國枝和雄, 山田敬嗣. ChronoSeeker: Web からの過去・未来情報のオンデマンド検索エンジン. Web とデータベースに関するフォーラム (WebDB Forum) 2009, 1B-4, 2009.
- [103] 吉田光男, 乾孝司, 山本幹雄. Web ニュースを用いた未来情報年表の自動構築. 第 3 回楽天研究開発シンポジウム, 2010.
- [104] 旭直人, 山本岳洋, 中村聡史, 田中克己. 挟みこむ検索:2 オブジェクトからの補間オブジェクト発見. 情報処理学会研究報告. データベース・システム研究会報告 2009-DBS-149(4), pp. 1-8, 2009.
- [105] 川田正明, 小川克彦. アノニマイズされた行動履歴に基づく行動パターン検索方式の提案. 情報処理学会論文誌 50(4), pp. 1251-1261, 2009.
- [106] 高林哲, 小松弘幸, 増井俊之. Migemo: 日本語のインクリメンタル検索. 情報処理学会論文誌 43(12), pp. 3698-3705, 2002.
- [107] 西本卓也, 岩田英三郎, 櫻井実, 廣瀬治人. 探索的検索のための音声入力インタフェースの検討. 情報処理学会研究報告. HCI, ヒューマンコンピュータインタラクション研究会報告 2008(11), pp. 9-14, 2008.

- [108] 増井俊之. 近傍関係を活用した情報検索. 情報処理学会研究報告. HI, ヒューマンインタフェース研究会報告 2003(69), pp. 53-58, 2003.
- [109] 土方嘉徳. 情報推薦・情報フィルタリングのためのユーザプロファイリング技術. 人工知能学会論文誌 19(3), pp. 365-372, 2004.
- [110] 山家雄介, 中村聡史, アダムヤトフト, 田中克己. ソーシャルブックマーキングの周期性発見と時期連動型検索ランキングへの適用. 情報処理学会論文誌. データベース 2(3), pp. 130-140, 2009.

付録 A

検索結果一覧

Google 検索，ブックマーク数順に検索，セレクトブックマ (=1) で検索において，「java」，「健康」，「映画」の3種類の単語で検索した場合の検索結果上位 30 件のタイトルと URL を以下の表に示す。

Google で「java」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	java.com: あなたと Java	http://www.java.com/ja/
2	無料 Java ソフトウェアをダウンロード - Sun Microsystems	http://www.java.com/ja/download/
3	Java - Wikipedia	http://ja.wikipedia.org/wiki/Java
4	Java テクノロジ - サン・マイクロシステムズ	http://jp.sun.com/java/
5	サン・マイクロシステムズ	http://jp.sun.com/
6	Java とは - 意味/解説/説明/定義 : IT 用語辞典	http://e-words.jp/w/Java.html
7	日本 Java ユーザグループ	http://www.java-users.jp/
8	Java とは - はてなキーワード	http://d.hatena.ne.jp/keyword/Java
9	Sun Developer Connection - Java Developer Connection	http://sdc.sun.co.jp/java/
10	Java の道 (Java 入門・リファレンス)	http://www.javaroad.jp/
11	JAVA 動物実験の廃止を求める会	http://www.java-animal.org/
12	Java Solution - @ IT	http://www.atmarkit.co.jp/fjava/
13	Java で Hello World	http://www.hellohiro.com/
14	とほほの Java 入門	http://www.tohoho-web.com/java/index.htm
15	Java Group ジャヴァグループ	http://www.java.gr.jp/
16	Java 2 Platform, Standard Edition, v 1.4.2 (J2SE) のダウンロード	http://java.sun.com/j2se/1.4.2/ja/download.html
17	IBM developerWorks Japan : Resources for Java developers	http://www.ibm.com/developerworks/jp/java/
18	Java 2 プラットフォーム SE v1.4.0	http://sdc.sun.co.jp/java/docs/j2se/1.4/ja/docs/ja/api/
19	Java -TECHSCORE-	http://www.techscore.com/tech/Java/
20	Java アプレット - Wikipedia	http://ja.wikipedia.org/wiki/Java
21	Java FAQ: よくある質問とその回答集	http://javafaq.jp/
22	Java News (Java に関する最新ニュース)	http://javanews.jp/
23	Java in the Box	http://www.javainthebox.net/
24	Microsoft Java Virtual Machine (MSJVM) サポート	http://www.microsoft.com/japan/java/default.msp
25	Java SE ダウンロード - Previous Release - JDK 5	http://java.sun.com/j2se/1.5.0/ja/download.html
26	IAJapan Java(TM) 研究部会	http://www.iajapan.org/bukai/java/
27	Java	http://kmuto.jp/java/
28	Java Tips	http://www.asahi-net.or.jp/~dp8t-asm/java/tips/index.html
29	窓の杜 - Java Runtime Environment (JRE)	http://www.forest.impress.co.jp/lib/arc/runtime/runtimeapp/javaruntime.html
30	FreeBSD Java? Project	http://www.freebsd.org/ja/java/

ブックマーク数順で「java」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	Java のクラスアンロード (Class Unloading)	http://www.nminoru.jp/~nminoru/java/class.unloading.html
2	Java の道 (Java 入門・リファレンス)	http://www.javaroad.jp/index.htm
3	頑健な Java プログラムの書き方 (Writing Robust Java Code)	http://www.alles.or.jp/~torutk/oojava/coding-Standard/writingrobustjavacode.html
4	Java 技術最前線: ITpro	http://itpro.nikkeibp.co.jp/article/COL-UMN/20060915/248243/
5	Java の学習なら JavaDrive	http://www.javadrive.jp/index.html
6	浅煎り珈琲 -Java アプリケーション入門	http://www.nextindex.net/java/index.html
7	Java で Hello World	http://www.hellohiro.com/
8	Java 2 Platform SE 5.0	http://java.sun.com/j2se/1.5.0/ja/docs/ja/api/index.html
9	【レポート】Java 初学者には最適!? 解説から実行までブラウザでコンプリート - Javala (MYCOM ジャーナル)	http://journal.mycom.co.jp/articles/2007/01/09/javala/
10	Java House ML	http://java-house.jp/ml/
11	プログラマの思索: Ruby よりも Java が好きな理由	http://forza.cocolog-nifty.com/blog/2007/05/rubyjava_5c11.html
12	Java の性能を分析する 10 のツール — エンタープライズ — マイコミジャーナル	http://journal.mycom.co.jp/news/2008/07/28/003/index.html
13	Java in the Box	http://www.javainthebox.net/
14	Java におけるコード進化パターン (Code Evolution Patterns in Java)	http://ncf.sakura.ne.jp/asato/doc/evolution.html
15	【ハウツー】クラス構造がまる見えに! UDoc で Java をダイナミックに分析する (1) JavaDoc、バイナリファイル、ソースコードからクラス図を生成 (MYCOM ジャーナル)	http://journal.mycom.co.jp/articles/2007/03/20/udoc/
16	Log4J 徹底解説?目次	http://www.nurs.or.jp/~sug/soft/log4j/index.htm
17	Super Csv - A super CSV reader/writer...	http://supercsv.sourceforge.net/
18	【レビュー】絶対使うべき! FindBugs で Bug 退治 (1) FindBugs とは (MYCOM ジャーナル)	http://journal.mycom.co.jp/articles/2006/09/23/findbugs1/
19	C/C いっさいなし、Java だけで開発された OS - JNode (MYCOM PC WEB)	http://pcweb.mycom.co.jp/news/2006/03/03/341.html
20	【ハウツー】Java だって Ruby に負けちゃいない - Java でも Rails クイックスタート Grails (MYCOM ジャーナル)	http://journal.mycom.co.jp/articles/2006/07/20/grails/
21	web.paulownia.jp - JDBC ドライバと接続 URL	http://wiki.paulownia.jp/java/jdbc
22	初心者が Java を“超高速”で学ぶためのコツ: ITpro	http://itpro.nikkeibp.co.jp/article/COL-UMN/20060927/249181/
23	Java News	http://javanews.jp/
24	InfoQ: ディベート: 何故ほとんどの大規模 web サイトは Java で構築されないのか?	http://www.infoq.com/jp/news/2007/10/big-java
25	【ハウツー】Java Web アプリでもわかりやすい URL を! - Url Rewrite Filter の使い心地 (1) わかりやすい URL の重要性 — エンタープライズ — マイコミジャーナル	http://journal.mycom.co.jp/articles/2007/05/21/urlrewrite/index.html
26	J2SE 5.0 虎の穴 Java2 SE v5.0 Tiger の新機能	http://www.javainthebox.net/laboratory/J2SE1.5/contents.html
27	James 君!	http://www.nurs.or.jp/~sug/soft/james/index.htm
28	概要 (Java 2 プラットフォーム SE v1.4.0)	http://sdc.sun.co.jp/java/docs/j2se/1.4/ja/docs/ja/api/index.html
29	ITmedia エンタープライズ: 矛盾を抱えつつ進化する“Java”??黒船となった Ruby on Rails (1/2)	http://www.itmedia.co.jp/enterprise/articles/0708/16/news006.html
30	Java Mozilla Html Parser 登場、ブラウザと同じクオリティを実現 — エンタープライズ — マイコミジャーナル	http://journal.mycom.co.jp/news/2008/01/22/024/index.html

セレクトブックマで「java」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	Java の道 (Java 入門・リファレンス)	http://www.javaroad.jp/index.htm
2	Java の学習なら JavaDrive	http://www.javadrive.jp/index.html
3	Java 技術最前線 : ITpro	http://itpro.nikkeibp.co.jp/article/COL-UMN/20060915/248243/
4	浅煎り珈琲 -Java アプリケーション入門	http://www.nextindex.net/java/index.html
5	頑健な Java プログラムの書き方 (Writing Robust Java Code)	http://www.alles.or.jp/~torutk/oojava/coding-Standard/writingrobustjavacode.html
6	Java House ML	http://java-house.jp/ml/
7	Java で Hello World	http://www.hellohiro.com/
8	Java 2 Platform SE 5.0	http://java.sun.com/j2se/1.5.0/ja/docs/ja/api/index.html
9	Java in the Box	http://www.javainthebox.net/
10	Log4J 徹底解説 ~ 目次	http://www.nurs.or.jp/~sug/soft/log4j/index.htm
11	Java News	http://javanews.jp/
12	初心者が Java を“超高速”で学ぶためのコツ : ITpro	http://itpro.nikkeibp.co.jp/article/COL-UMN/20060927/249181/
13	J2SE 5.0 虎の穴 Java2 SE v5.0 Tiger の新機能	http://www.javainthebox.net/laboratory/J2SE1.5/contents.html
14	概要 (Java 2 プラットフォーム SE v1.4.0)	http://sdc.sun.co.jp/java/docs/j2se/1.4/ja/docs/ja/api/index.html
15	【レポート】Wicket 入門 - Java と HTML だけで作る Web アプリケーション (1) Wicket とは? — エンタープライズ — マイコミジャーナル	http://journal.mycom.co.jp/articles/2006/05/08/wicket/
16	Java でゲーム作りますが何か?	http://javagame.main.jp/
17	Archive: Java[tm] Technology Products Download	http://java.sun.com/products/archive/index.html
18	Eclipse ではじめるプログラミング	http://www.atmarkit.co.jp/fjava/rensai3/eclipsejava01/eclipse01.html
19	Java のクラスアンロード (Class Unloading)	http://www.nminoru.jp/~nminoru/java/class.unloading.html
20	MYCOM ジャーナル - エンタープライズ - コラム - ライトニング Java	http://journal.mycom.co.jp/column/java/
21	Ja-Jakarta Project	http://www.jajakarta.org/
22	JavaA2Z	http://www.kab-studio.biz/Programing/JavaA2Z/
23	JavaWorld Online - 最先端 Java テクノロジーの実践活用誌	http://www.javaworld.jp/
24	Click Framework	http://click.sourceforge.net/
25	Java を JavaScript に変換するグーグルのツールを使ってみよう? @ IT	http://www.atmarkit.co.jp/fwcr/rensai/googlewebtoolkit01/01.html
26	TECHSCORE(テックスコア)?C 言語/JAVA/デザインパターン/CORBA/XML/SQL/UML を基礎から丁寧に解説します?	http://www.techscore.com/index.html
27	Java 入門	http://www5c.biglobe.ne.jp/~ecb/java/java00.html
28	Java Solution ? @ IT	http://www.atmarkit.co.jp/fjava/
29	Restlet - Lightweight REST framework for Java	http://www.restlet.org/
30	@ IT : Hibernate で理解する O/R マッピング (1)	http://www.atmarkit.co.jp/fjava/rensai3/ormap01/ormap01.html

Google 検索で「健康」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	健康 - Wikipedia	http://ja.wikipedia.org/wiki/健康
2	家庭の医学・健康 - goo ヘルスケア	http://health.goo.ne.jp/
3	健康ネット	http://www.health-net.or.jp/
4	病院情報、家庭の医学、病気の検索、薬の情報、健康情報 - Yahoo ...	http://health.yahoo.co.jp/
5	厚生労働省：健康	http://www.mhlw.go.jp/bunya/kenkou/index.html
6	健康 ON-LINE	http://www.webkenko.com/
7	asahi.com (朝日新聞社): 医療・健康・ニュース	http://www.asahi.com/health/
8	いきいき健康 NIKKEI NET	http://health.nikkei.co.jp/
9	[健康管理] All About	http://allabout.co.jp/health/
10	ケンコーコム - 健康メガショップ	http://www.kenko.com/
11	健康とは - はてなキーワード	http://d.hatena.ne.jp/keyword/健康
12	NHK 健康ホームページ：トップページ	http://www.nhk.or.jp/kenko/
13	健康と医学 - Yahoo!カテゴリ	http://dir.yahoo.co.jp/Health/
14	asahi.com (朝日新聞社): 治療薬で健康被害、入院患者の 5 人に 1 人 3 ...	http://www.asahi.com/national/update/1220/TKY200912200254.html
15	健康と医療の総合情報サイト — 健康 Salad	http://www.k-salad.com/
16	Medical Tribune あなたの健康百科	http://www.medical-tribune.co.jp/kenkou/kenkou-index.htm
17	医学・健康情報サイト J-Medical	http://www.j-medical.net/
18	健康美容 EXPO-健康食品、美容の商材、原料、製造探しに	http://www.e-expo.net/
19	健康ウォーキング	http://gnl.cplaza.ne.jp/
20	主婦の友社 雑誌情報 「健康 2010 年 1 月号」	http://www.shufunotomo.co.jp/magazine/magazine_detail.php?id=3
21	健康日本 21	http://www.kenkounippon21.gr.jp/
22	Google ディレクトリ - World ぐ Japanese ぐ 健康	http://www.google.com/Top/World/ Japanese/健康
23	(財)日本健康・栄養食品協会	http://www.jhnfa.org/
24	ホーム - 全国健康保険協会	http://www.kyoukaikenpo.or.jp/
25	独立行政法人 国立健康・栄養研究所：トップページ	http://www.nih.go.jp/eiken/
26	健康倶楽部	http://www.nagara.com/
27	あいち健康プラザ あいち健康の森健康科学総合センター	http://www.ahv.pref.aichi.jp/
28	健康 - 雑誌のネット書店 Fujisan.co.jp	http://www.fujisan.co.jp/Product/673
29	健康情報館健康でドンドン！	http://www.isuoai.net/
30	健康食品ナビ<東京都福祉保健局>	http://www.fukushihoken.metro.tokyo.jp/anzen/supply/index.html

ブックマーク数順で「健康」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	ゲンダイネット - 目の疲れをためない 3 大作戦	http://gendai.net/?m=view
2	体の歪みを治したい:アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51205892.html
3	重い、痛いを吹き飛ばせ！肩コリ解消法 — Web 担当者 Forum	http://web-tan.forum.impressrd.jp/e/2007/07/10/1055
4	眼精疲労を治すには:アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51148963.html
5	金も時間も掛からない花粉症予防の仕方:アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51238144.html
6	health クリック 健康 生活習慣病 サプリメント	http://www2.health.ne.jp/
7	蒸しタオルを使うと酷使した目の疲れが取れる - GIGAZINE	http://gigazine.net/index.php?/news/comments/20070831_tired_eye/
8	疲れ目：パソコン作業で肩こり、視力低下など眼精疲労 対策は... (上) - 毎日 j p (毎日新聞)	http://mainichi.jp/life/electronics/news/20080601mog00m100003000c.html
9	ビジネスマンの不死身力：「夜食は太る」の科学 (1/2) - ITmedia エンタープライズ	http://www.itmedia.co.jp/enterprise/articles/0806/15/news002.html
10	基礎代謝を高めるための 9 の技法	http://taisya.denze.net/
11	MouRa Net 現代 巷にはびこる「健康情報」50 のウソ・ホント	http://moura.jp/scoop-e/mgendai/back/200511/
12	第 1 回 眼の疲れを取る：ITpro	http://itpro.nikkeibp.co.jp/article/COL-UMN/20071029/285771/
13	姿勢をよくするための運動 - Tech Mom from Silicon Valley	http://d.hatena.ne.jp/michikaifu/20071121/1195671034
14	口臭をなくしたい!:アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51235579.html
15	【2ch】ニュー速クオリティ:疲労が劇的に回復する食べ物	http://news4vip.livedoor.biz/archives/51176871.html
16	視力回復とレーシックのアイポータル	http://eyeportal.jp/
17	PC で眼が疲れない方法 - 萌え理論 Blog	http://d.hatena.ne.jp/sirouto2/20080420/p1
18	5 時間以下の睡眠続死亡率 1.7 倍に 7 時間寝よう Ameba News	http://news.ameba.jp/2007/09/7433.php
19	なんでも評判:空腹は幸福? ストレスで腹が減ったときは何も食べずに我慢した方がストレスに打ち克てる...	http://rate.livedoor.biz/archives/50679219.html
20	指をボキボキ鳴らすと太くなるって本当?	http://nikkeibp.jp/wcs/leaf/CID/onair/kenkou/plus/423239
21	視力回復マッサージは本当に効き目があるのだ (映像付) — idea * idea	http://www.ideaxidea.com/archives/2006/10/post_156.html
22	医学都市伝説: 暗いところで本を読んでも目は悪くならない	http://med-legend.com/mt/archives/2006/07/post_896.html
23	体に良い座り方、いす選びのポイント? @ IT 自分戦略研究所	http://jibun.atmarkit.co.jp/life01/special/ergo/ergo01.html
24	極東ブログ: 素人風邪対策	http://finalvent.cocolog-nifty.com/fareastblog/2007/10/post_a7bd.html
25	病院検索ならここカラダ	http://www.cocokarada.jp/index.html
26	スラッシュドット・ジャパン — 睡眠不足だと仕事はかどらない理由、科学的に明らかになる	http://slashdot.jp/science/article.pl?sid=08/05/22/2326234
27	精神状態の健康がピンチになった時の 3 つの対処 - koe だめ	http://d.hatena.ne.jp/bash0C7/20080302/1204425675
28	意外に効き目のある視力回復マッサージ	http://www3.tky.3web.ne.jp/~aihide/bilbbr_Ma.htm
29	【2ch】ニュー速クオリティ:肌の綺麗な人がしてる事	http://news4vip.livedoor.biz/archives/50965236.html
30	タバコと同じくらい有害—レーザープリンタに健康上のリスク:ニュース - CNET Japan	http://japan.cnet.com/news/biz/story/0,2000056020,20353976,00.htm

セレクトブックマで「健康」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	health クリック 健康 生活習慣病 サプリメント	http://www2.health.ne.jp/
2	基礎代謝を高めるための 99 の技法	http://taisya.denze.net/
3	ゲンダイネット - 目の疲れをためない 3 大作戦	http://gendai.net/?m=view
4	視力回復マッサージは本当に効き目があるのか (映像付) — idea * idea	http://www.ideaxidea.com/archives/2006/10/post_156.html
5	重い、痛いを吹き飛ばせ! 肩こり解消法 — Web 担当者 Forum	http://web-tan.forum.impressrd.jp/e/2007/07/10/1055
6	体の歪みを治したい:アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51205892.html
7	病院検索ならここカラダ	http://www.cocokarada.jp/index.html
8	意外に効き目のある視力回復マッサージ	http://www3.tky.3web.ne.jp/~aihide/bilbbr_Ma.htm
9	蒸しタオルを使うと酷使した目の疲れが取れる - GIGAZINE	http://gigazine.net/index.php?/news/comments/20070831_tired_eye/
10	MouRa Net 現代 巷にはびこる「健康情報」50 のウソ・ホント	http://moura.jp/scoop-e/mgendai/back/200511/
11	第 1 回 目の疲れを取る: ITpro	http://itpro.nikkeibp.co.jp/article/COL-UMN/20071029/285771/
12	PC で眼が疲れない方法 - 萌え理論 Blog	http://d.hatena.ne.jp/sirouto2/20080420/p1
13	5 時間以下の睡眠続け死亡率 1.7 倍に 7 時間寝よう Ameba News	http://news.ameba.jp/2007/09/7433.php
14	視力回復とレーシックのアイポータル	http://eyeportal.jp/
15	分裂勘違い君劇場 - シャンプーとリンスと石鹸は使わない方がいい	http://d.hatena.ne.jp/fromdusktildawn/20070202/1170403306
16	眼精疲労を治すには:アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51148963.html
17	割れた腹筋を手に入れるトレーニングを教えてくださいました。 — その他 (ライフ) — とりあえず...	http://blog.pasonatech.co.jp/masuda/399/3460.html
18	極東ブログ: 素人風邪対策	http://finalvent.cocolog-nifty.com/fareast-blog/2007/10/post_a7bd.html
19	医学都市伝説: 暗いところで本を読んでも目は悪くならない	http://med-legend.com/mt/archives/2006/07/post_896.html
20	なかなか眠れない人のための簡単に眠る 10 の方法 - GIGAZINE	http://gigazine.net/index.php?/news/comments/20071230_better_sleep/
21	【2ch】ニュー速クオリティ: 疲労が劇的に回復する食べ物	http://news4vip.livedoor.biz/archives/51176871.html
22	BMI 判定	http://www.ahv.pref.aichi.jp/taikei/chap1_nn.html
23	金も時間も掛からない花粉症予防の仕方:アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51238144.html
24	メルクマニユアル医学百科最新家庭版	http://mmh.banyu.co.jp/mmhe2j/index.html
25	睡眠時間を記録するサイト — ねむログ	http://www.nemulog.jp/
26	シゴタノ! - 睡眠時間を短くする 14 のコツ < 前編 >	http://cyblog.jp/modules/weblog/details.php?blog_id=90
27	「うつ」にならない、繰り返さない? @ IT 自分戦略研究所	http://jibun.atmarkit.co.jp/ljibun01/column/stress/stress34.html
28	姿勢をよくするための運動 - Tech Mom from Silicon Valley	http://d.hatena.ne.jp/michikaifu/20071121/1195671034
29	口臭をなくしたい!:アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51235579.html
30	視力検査 - あなたの視力 今いくつ?: 視力回復のアイポータル	http://eyeportal.jp/60siryokukensa/siryokukennsa.html

Google 検索で「映画」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	Yahoo!映画 - 映画情報	http://movies.yahoo.co.jp/
2	映画情報 - goo 映画	http://movie.goo.ne.jp/
3	映画・DVD と映画館の上映時間を完全網羅 — Movie Walker	http://movie.walkerplus.com/
4	映画のことなら eiga.com	http://eiga.com/
5	映画「サマーウォーズ」公式サイト	http://s-wars.jp/
6	映画 - Wikipedia	http://ja.wikipedia.org/wiki/映画
7	TOHO シネマズ	http://www.tohotheater.jp/
8	ワーナー・マイカル・シネマズ - 映画館、映画情報、上映スケジュール ...	http://www.warnermycal.com/
9	大ブーイング! ぐだぐだ最終回「JIN」映画化?TBS 意外な反応 ...	http://headlines.yahoo.co.jp/hl?a=20091222-00000007-ykf-ent
10	新作映画情報「ぴあ映画生活」	http://pia-eigaseikatsu.jp/
11	映画館・シネコンの【ムービックス-MOVIX】	http://www.movix.co.jp/
12	映画データベース - allcinema	http://www.allcinema.net/
13	シネマスクランブル 映画予告編・映画ランキング・映画上映 ...	http://cinesc.cplaza.ne.jp/
14	チネチッタ【川崎・映画館】公式サイト	http://cinecitta.co.jp/index_flash.shtml
15	新宿バルト 9	http://wald9.com/
16	映画、映画館検索 - TSUTAYA online	http://www.tsutaya.co.jp/cinema.html
17	映画、ビデオ - Yahoo!カテゴリ	http://dir.yahoo.co.jp/entertainment/movies_and_films/
18	映画とは - はてなキーワード	http://d.hatena.ne.jp/keyword/映画
19	『アバター』は映画の未来に iPhone 登場なみの衝撃を与える	http://jp.techcrunch.com/archives/20091219avatar-is-like-the-iphone-of-movies/
20	映画・映画館・試写会 — ウーマンエキサイト シネマ (映画)	http://woman.excite.co.jp/cinema/
21	日本映画のガラパゴス化、インフェルノ鎖国 - 俺の邪悪なメモ	http://d.hatena.ne.jp/sumiyama/20091222/p1
22	東宝 WEB SITE	http://www.toho.co.jp/
23	Yahoo!映画 - 映画館検索	http://movies.yahoo.co.jp/preflist.html
24	109 シネマズ公式ホームページ	http://109cinemas.net/
25	前田有一の超映画批評	http://movie.maeda-y.com/
26	シネマぴあ	http://www.pia.co.jp/cinema.html
27	@nifty 映画 - 映画情報	http://movie.nifty.com/
28	無料映画館	http://freemoviefree.blog31.fc2.com/
29	映画 — WOWOW オンライン	http://www.wowow.co.jp/movie/
30	映画なら GyaO!【映画】 無料映画の映像視聴 GyaO![ギャオ] 映画	http://www.freebsd.org/ja/java/

ブックマーク数順で「映画」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	前田有一の超映画批評	http://movie.maeda-y.com/
2	みんなのシネマレビュー	http://www.jtnews.jp/
3	あの映画のココがわからない まとめサイト	http://www.movie-faq.com/
4	allcinema ONLINE 映画データベース	http://www.allcinema.net/prog/index2.php
5	超映画批評	http://movie.maeda-y.com/index.html
6	段ボールで『2001 年宇宙の旅』をリメイクできるか? - ペイエリア在住町山智浩アメリカ日記	http://d.hatena.ne.jp/TomoMachi/20080105
7	映画生活 - 新作映画情報	http://www.eigaseikatu.com/
8	CinemaScape?映画批評空間?	http://cinema.intercritique.com/
9	痛いニュース (':) :もっとも感動した映画トップ 10 発表	http://blog.livedoor.jp/dqnplus/archives/1031610.html
10	404 Blog Not Found: アマとプロとが選んだディストピア映画歴代トップ 26 - 1	http://blog.livedoor.jp/dankogai/archives/50921614.html
11	【2ch】ニュー速クオリティ: 一生のうちに一度は見ておくべき映画	http://news4vip.livedoor.biz/archives/51161310.html
12	cinemacafe.net シネマカフェ?映画ファンによる、映画ファンのための、映画のライフスタイル・エンターテインメント・サイト?	http://www.cinemacafe.net/
13	「少林少女」はこんなにすごい! - かむかむごっくん	http://d.hatena.ne.jp/kossetsu/20080501/p1
14	もっと知られていい映画: アルファルファモザイク	http://alfalfa.livedoor.biz/archives/51140013.html
15	どうしてハリウッド映画は予算に何百億円もかけられるのか? - GIGAZINE	http://gigazine.net/index.php?/news/comments/20070808_movie_money/
16	超映画批評『それでもボクはやってない』98 点 (100 点満点中)	http://movie.maeda-y.com/movie/00854.htm
17	flowerwild.net? 蓮實重彦インタビュー リアルタイム批評のすすめ vol.1	http://flowerwild.net/2006/11/2006-11-08_133443.php
18	eiga.com	http://eiga.com/
19	2007 年の映画をふりかえる / 結果発表 - 空中キャンプ	http://d.hatena.ne.jp/zoot32/20071223#p1
20	超映画批評『明日、君がいない』97 点 (100 点満点中)	http://movie.maeda-y.com/movie/00895.htm
21	X51.ORG: 脳性マヒの殺人鬼? 『おそいひと』主演・住田雅清インタビュー	http://x51.org/x/07/12/2050.php
22	いのちの食べかた	http://www.espace-sarou.co.jp/inochi/
23	超映画批評『僕の彼女はサイボーグ』採点不能 (100 点満点中)	http://movie.maeda-y.com/movie/01116.htm
24	破壊屋 2007 年の映画の感想 恋空	http://hakaiya.web.infoseek.co.jp/html/2007/20071111_1.html
25	超映画批評『靖国 YASUKUNI』0 点 (100 点満点中)	http://movie.maeda-y.com/movie/01102.htm
26	超映画批評『シッコ』96 点 (100 点満点中)	http://movie.maeda-y.com/movie/00948.htm
27	【2ch】ニュー速クオリティ: 観るまでバカにしたのに観たら面白かった映画	http://news4vip.livedoor.biz/archives/50953673.html
28	痛いニュース (':) : 大人が選ぶ泣ける洋画ベスト 30 1 位「タイタニック」、2 位「アルマゲドン」	http://blog.livedoor.jp/dqnplus/archives/1079189.html
29	超映画批評『ミスト』90 点 (100 点満点中)	http://movie.maeda-y.com/movie/01104.htm
30	押井守監督最新作 映画「スカイ・クロラ The Sky Crawlers」公式サイト	http://www.warnerbros.co.jp/skycrawlers/

セレクトブックマで「映画」を検索した場合の検索結果上位 30 件とその URL

順位	タイトル	URL
1	前田有一の超映画批評	http://movie.maeda-y.com/
2	みんなのシネマレビュー	http://www.jtnews.jp/
3	あの映画のココがわからない まとめサイト	http://www.movie-faq.com/
4	allcinema ONLINE 映画データベース	http://www.allcinema.net/prog/index2.php
5	映画生活 - 新作映画情報	http://www.eigaseikatu.com/
6	超映画批評	http://movie.maeda-y.com/index.html
7	CinemaScape?映画批評空間?	http://cinema.intercritique.com/
8	cinemacafe.net シネマカフェ ?映画ファンによる、映画ファンのための、映画のライフスタイル・エンターテインメント・サイト?	http://www.cinemacafe.net/
9	eiga.com	http://www.eiga.com/index.shtml
10	いのちの食べかた	http://www.espace-sarou.co.jp/inochi/
11	The Internet Movie Database (IMDb)	http://www.imdb.com/
12	時をかける少女	http://www.kadokawa.co.jp/tokikake/
13	破壊屋_2007 年の映画の感想_恋空	http://hakaiya.web.infoseek.co.jp/html/2007/20071111_1.html
14	ダーウィンの悪夢	http://www.darwin-movie.jp/
15	DEATH NOTE -デスノート-	http://www.warnerbros.co.jp/deathnote/
16	X51.ORG : 脳性マヒの殺人鬼 『おそいひと』主演・住田雅清インタビュー	http://x51.org/x/07/12/2050.php
17	映画「日本以外全部沈没」公式サイト	http://www.all-chinbotsu.com/
18	段ボールで『2001 年宇宙の旅』をリメイクできるか? - ベイエリア在住町山智浩アメリカ日記	http://d.hatena.ne.jp/TomoMachi/20080105
19	Yahoo! JAPAN - 秒速 5 センチメートル	http://5cm.yahoo.co.jp/
20	不都合な真実	http://www.futsugou.jp/
21	映画『さくらん』公式サイト	http://www.sakuran-themovie.com/
22	痛いニュース (´):もっとも感動した映画トップ 10 発表	http://blog.livedoor.jp/dqnplus/archives/1031610.html
23	超映画批評『それでもボクはやってない』98 点 (100 点満点中)	http://movie.maeda-y.com/movie/00854.htm
24	～木曜洋画劇場の CM～	http://www.geocities.jp/thursday_movie/
25	flowerwild.net - 蓮實重彦インタビュー リアルタイム批評のすすめ vol.1	http://flowerwild.net/2006/11/2006-11-08_133443.php
26	映画「恋愛睡眠のすすめ」公式サイト	http://renaisuimin.com/
27	著作権の切れた過去の名作映画やドキュメンタリーをダウンロード - GIGAZINE	http://gigazine.net/index.php?/news/comments/20060801_freemovies/
28	404 Blog Not Found: アマとプロとが選んだディストピア映画歴代トップ 26 - 1	http://blog.livedoor.jp/dankogai/archives/50921614.html
29	どうしてハリウッド映画は予算に何百億円もかけられるのか? - GIGAZINE	http://gigazine.net/index.php?/news/comments/20070808_movie_money/
30	映画「太陽」公式サイト	http://taiyo-movie.com/

付録B

評価結果一覧

「java」、「健康」、「映画」の分野において、被験者が選んだ Web ページの一覧を以下の表に示す。

「java」で検索した場合の検索結果の中から被験者が選んだ Web ページ

人数	得点	タイトル	セレクトブックマ順位	ブックマーク数順位	Google 検索順位
7	55	とほほの Java 入門			14
7	45	Java の道 (Java 入門・リファレンス)	1	2	10
7	33	Java の学習なら JavaDrive	2	5	
5	14	初心者が Java を“超高速”で学ぶためのコツ：ITpro	12	22	
4	27	Java FAQ: よくある質問とその回答集			21
4	23	Java 入門	27		
4	20	MYCOM ジャーナル - エンタープライズ - コラム - ライトニング Java	20		
4	19	Java で Hello World	7	7	13
4	15	TECHSCORE(テック ス コ ア) ?C 言語/JAVA/デザインパターン/CORBA/XML/SQL/UML を基礎から丁寧に解説します?	26		
3	18	Java -TECHSCORE-			19
3	18	頑健な Java プログラムの書き方 (Writing Robust Java Code)	5	3	
3	13	Java を JavaScript に変換するグーグルのツールを使ってみよう？ @ IT	25		
3	11	@ IT : Hibernate で理解する O/R マッピング (1)	30		
2	10	Java アプレット - Wikipedia			20
2	18	JavaA2Z	22		
2	19	Java 技術最前線：ITpro	3	4	
2	15	ITmedia エンタープライズ:矛盾を抱えつつ進化する“Java ”??黒船となった Ruby on Rails (1/2)		29	
2	14	Java 2 Platform SE 5.0	8	8	

2	13	Java におけるコード進化パターン (Code Evolution Patterns in Java)		14	
2	12	C/C いっさいなし、Java だけで開発された OS - JNode (MYCOM PC WEB)		19	
2	9	Java Solution ? @ IT	28		12
2	8	Ruby よりも Java が好きな理由		11	
2	7	Java の性能を分析する 10 のツール — エンタープライズ — マイコミジャーナル		12	
2	7	Java とは - 意味/解説/説明/定義 : IT 用語辞典			6
2	5	Java			27
1	9	IBM developerWorks Japan : Resources for Java developers			17
1	9	Java でゲーム作りますが何か？	16	17	
1	8	Ja-Jakarta Project	21		
1	8	Java in the Box	9	13	23
1	7	Java テクノロジ - サン・マイクロシステムズ			4
1	6	Archive: Java[tm] Technology Products Download	17		
1	6	James 君！		27	
1	6	Log4J 徹底解説～目次	10	16	
1	6	Java Mozilla Html Parser 登場、ブラウザと同じクオリティを実現 — エンタープライズ — マイコミジャーナル		30	
1	5	Eclipse ではじめるプログラミング	18		
1	5	Java 2 Platform,Standard Edition, v 1.4.2 (J2SE) のダウンロード			16
1	4	【レポート】Java 初学者には最適!? 解説から実行までブラウザでコンプリート - Javala (MYCOM ジャーナル)	15	9	
1	4	Java - Wikipedia			3
1	4	Java News	11	23	22
1	3	【ハウツー】Java Web アプリでもわかりやすい URL を! - Url Rewrite Filter の使い心地 (1) わかりやすい URL の重要性 — エンタープライズ — マイコミジャーナル		25	
1	3	IAJapan Java(TM) 研究部会			26
1	3	【ハウツー】Java だって Ruby に負けちゃいない - Java でも Rails クイックスタート Grails (MYCOM ジャーナル)		20	

1	1	J2SE 5.0 虎の穴 Java2 SE v5.0 Tiger の 新機能	13	26	
---	---	---	----	----	--

「健康」で検索した場合の検索結果の中から被験者が選んだ Web ページ

人数	得点	タイトル	セレクトブックマ順位	ブックマーク数順位	Google 検索順位
6	36	視力検査 - あなたの視力 今いくつ?: 視力回復のアイポータル	30		
6	30	意外に効き目のある視力回復マッサージ	8	28	
5	24	蒸しタオルを使うと酷使した目の疲れが取れる - GIGAZINE	9	7	
5	30	なかなか眠れない人のための簡単に眠る 10 の方法 - GIGAZINE	20		
4	21	第 1 回 眼の疲れを取る: ITpro	11	12	
3	33	基礎代謝を高めるための 99 の技法	2	10	
3	23	睡眠時間を記録するサイト — ねむログ	25		
3	21	視力回復マッサージは本当に効き目があるのだ (映像付) — idea * idea	4	21	
3	18	ビジネスマンの不死身力: 「夜食は太る」の科学 (1/2) - ITmedia エンタープライズ		9	
3	18	スラッシュドット・ジャパン — 睡眠不足だと仕事はかどらない理由、科学的に明らかになる		26	
3	17	5 時間以下の睡眠続け死亡率 1.7 倍に 7 時間寝よう Ameba News	13	18	
3	17	「うつ」にならない、繰り返さない? @ IT 自分戦略研究所	27		
3	14	視力回復とレーシックのアイポータル	14	16	
3	14	シゴタノ! - 睡眠時間を短くする 14 のコツ < 前編 >	26		
3	10	なんでも評点: 空腹は幸福? ? ストレスで腹が減ったときは何も食べずに我慢した方がストレスに打ち克てる...		19	
3	9	姿勢をよくするための運動 - Tech Mom from Silicon Valley	28	13	
2	19	Medical Tribune あなたの健康百科			16
2	14	精神状態の健康がピンチになった時の 3 つの対処 - koe だめ		27	
2	14	重い、痛いを吹き飛ばせ! 肩コリ解消法 — Web 担当者 Forum	5	3	
2	14	割れた腹筋を手に入れるトレーニングを教えてくださいました。 — その他 (ライフ) — とりあえず...	17		
2	13	[健康管理] All About			19

2	12	医学都市伝説: 暗いところで本を読んでも目は悪くならない	19	22	
2	11	asahi.com (朝日新聞社): 医療・健康・ニュース			7
2	10	病院情報、家庭の医学、病気の検索、薬の情報、健康情報 - Yahoo ...			4
2	9	指をポキポキ鳴らすと太くなるって本当?		20	
2	5	タバコと同じくらい有害-レーザープリンタに健康上のリスク:ニュース - CNET Japan		30	
2	3	MouRa Net 現代 巷にはびこる「健康情報」50 のウソ・ホント	10	11	
1	10	健康 - Wikipedia			1
1	10	PC で眼が疲れない方法 - 萌え理論 Blog	12	17	
1	9	NHK 健康ホームページ: トップページ			12
1	8	health クリック 健康 生活習慣病 サプリメント	1	6	
1	8	体の歪みを治したい:アルファルファモザイク	6	2	
1	7	家庭の医学・健康 - goo ヘルスケア			2
1	6	ゲンダイネット - 目の疲れをためない3大作戦	3	1	
1	6	ケンコーコム - 健康メガショップ			10
1	6	金も時間も掛からない花粉症予防の仕方: アルファルファモザイク			24
1	4	健康 ON-LINE			6
1	4	口臭をなくしたい!:アルファルファモザイク	29	14	
1	3	眼精疲労を治すには:アルファルファモザイク	22	5	
1	3	健康ウォーキング			1
1	3	体に良い座り方、いす選びのポイント? @ IT 自分戦略研究所		23	
1	2	【2ch】ニュー速クオリティ:疲労が劇的に回復する食べ物	21	15	
1	2	Google ディレクトリ - World 健康			22
1	1	ホーム - 全国健康保険協会			24
1	1	分裂勘違い君劇場 - シャンプーとリンスと石鹸は使わない方がいい	15		

「映画」で検索した場合の検索結果の中から被験者が選んだ Web ページ

人数	得点	タイトル	セレクトブック マ順位	ブックマーク 数順位	Google 検 索順位
5	29	Yahoo!映画 - 映画情報			1
4	28	ワーナー・マイカル・シネマズ - 映画館、 映画情報、上映スケジュール ...			8
4	26	TOHO シネマズ			7
4	25	映画、映画館検索 - TSUTAYA online			16
4	16	シネマびあ			26
4	15	@nifty 映画 - 映画情報			27
4	12	映画情報 - goo 映画			2
3	25	【2ch】ニュー速クオリティ:観るまでバ カにしてたのに観たら面白かった映画		27	
3	23	あの映画のココがわからない まとめサ イト	3	3	
3	21	allcinema ONLINE 映画データベース	4	4	12
3	20	109 シネマズ公式ホームページ			24
3	19	映画・DVD と映画館の上映時間を完全網 羅 — Movie Walker			3
3	18	eiga.com	9	18	4
3	17	みんなのシネマレビュー	2	2	
3	15	【2ch】ニュー速クオリティ:一生のう ちに一度は見ておくべき映画		11	
3	14	映画なら GyaO! 【映画】 無料映画の映 像視聴 GyaO![ギャオ] 映画			30
3	12	もっと知られていい映画:アルファルファ モザイク		14	
3	12	シネマスクランブル 映画予告編・映画ラ ンキング・映画上映 ...			13
3	10	前田有一の超映画批評	1	1	25
2	18	著作権の切れた過去の名作映画やドキュ メンタリーをダウンロード - GIGAZINE	27		
2	17	映画、ビデオ - Yahoo!カテゴリ			17
2	17	痛いニュース ('):大人が選ぶ泣ける洋 画ベスト 30 1「タイタニック」、2「ア ルマゲドン」		28	
2	15	映 画 館 ・ シ ネ コ ンの【ム ー ビッ クス-MOVIX】			11
2	15	映画「サマーウォーズ」公式サイト			5
2	12	無料映画館			28
2	11	新作映画情報「びあ映画生活」			10
2	9	時をかける少女	12		
2	5	The Internet Movie Database (IMDb)	11		

2	3	映画・映画館・試写会 ― ウーマンエキサイト シネマ (映画)			20
1	10	『アバター』は映画の未来に iPhone 登場 なみの衝撃を与える			19
1	8	flowerwild.net - 蓮實重彦インタビュー リアルタイム批評のすすめ vol.1	25		
1	6	404 Blog Not Found:アマとプロとが選 んだディストピア映画歴代トップ 26 - 1	28	10	
1	6	映画 ― WOWOW オンライン			29
1	6	映画生活 - 新作映画情報			17
1	6	Yahoo! JAPAN - 秒速 5 センチメートル	19		
1	5	東宝 WEB SITE			22
1	5	大ブーイング! ぐだぐだ最終回「JIN」映 画化? TBS 意外な反応 ...			9
1	4	映画『さくらん』公式サイト	21		
1	4	チネチッタ【川崎・映画館】公式サイト			14
1	3	どうしてハリウッド映画は予算に何百億 円もかけられるのか? - GIGAZINE	29	15	
1	3	日本映画のガラパゴス化、インフェルノ 鎖国 - 俺の邪悪なメモ			21
1	3	X51.ORG : 脳性マヒの殺人鬼 『おそ いひと』主演・住田雅清インタビュー	16		
1	2	破壊屋_2007 年の映画の感想_恋空	13	24	