慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	長期的に利用する情報へのアクセスに関する研究
Sub Title	Research on access to long term information
Author	上野, 大樹(Ueno, Taiki) 安村, 通晃(Yasumura, Michiaki)
Publisher	慶應義塾大学湘南藤沢学会
Publication year	2013-03
Jtitle	博士論文
JaLC DOI	
Abstract	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および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. 情報にはさまざま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を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, こういった情報が有用である可能性は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, 特別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, 長期的検索という新しい概念を提唱し, 長期的に利用される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」という2つの例に焦点を当てて,「セレクトブクマ」, 「MasteryEye」という2つのシステムを開発した. セレクトブクマは, 長期的に利用できるWebページに特化して検索するためのシステムである. ここではまず, 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, Webページがどのように利用され, ブックマークされているかを分析した. そして, その分析結果に基づき, 長期的なWebページを検索できるように設計・実装を行った. セレクトブクマでは, 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見つけることができるようにシステムの設計・実装を行った. MasteryEyeでは, 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. これら2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.
Notes	インタラクションデザインプロジェクト(iDP) 安村通晃研究会2012年度
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=0302-0000-0674

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

ISBN 978-4-87762-265-7
SFC-DT 2012-002

博士論文

長期的に利用する情報へのアクセスに関する研究

2012年度

上野 大樹　政策・メディア研究科 博士課程

インタラクションデザインプロジェクト(iDP)

慶應義塾大学湘南藤沢学会

Keio SFC Academic Society

推薦のことば

本論文は、Web アクセスにおいて、短期的な流行のみを追い求める、研究動向に対して、長期的なアクセス情報の提供を狙いとして、分析評価し、またそのため新たに、セレクトブクマとMasteryEyeという二つのシステムを構築した。すなわち、セレクトブクマは、はてなブクマののうちから長期的に有用な Web ページの検索を助けるものであり、また、MasteryEye の方は、Web 検索で長期的に有用なキーワードの提供を助けるものである。今後、Webの歴史が蓄積するに従い、ここで提示したような、長期的な情報のアクセスが必ず増えるものと考えられ、本研究の有用性が今後ますます高まるものと期待できる。

慶應義塾大学
環境情報学部教授
安村通晃

博士論文　平成24（2012）年度

長期的に利用する情報へのアクセスに関する研究

慶應義塾大学大学院政策・メディア研究科

上野 大樹

博士論文　平成24（2012）年度

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および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．

情報にはさまざま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を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，こういった情報が有用である可能性は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，特別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，長期的検索という新しい概念を提唱し，長期的に利用される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」という2つの例に焦点を当てて，「セレクトブクマ」，「MasteryEye」という2つのシステムを開発した．

セレクトブクマは，長期的に利用できるWebページに特化して検索するためのシステムである．ここではまず，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，Webページがどのように利用され，ブックマークされているかを分析した．そして，その分析結果に基づき，長期的なWebページを検索できるように設計・実装を行った．セレクトブクマでは，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見つけることができるようにシステムの設計・実装を行った．MasteryEyeでは，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．

これら２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，アクセスパターン

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

Summary

In this research, we proposed information retrieval method and information access method that find information used for a long-term. We call these methods "long-term retrieval". Moreover, we propose "degree of long-term" that is an index of which information was used for a long-term.

The people have used information in variety ways. For example, there are the information used for a short-term, the information used for a long-term and the information used periodically. We can predict information types and values by using these data. Thus, we consider information that used long-term in the past will be used for a long-term in the future too. Moreover, these information is a high possibility of the practical value. There are a lot of information retrieval methods that can search popular information and trend information, but there are few information retrieval methods that can search specialized long-term information. Therefore, now it is difficult to find long-term information if it is not popular. In this research, we propose a new concept that is "long-term retrieval", and we propose retrieval methods that can search long-term information.

For concrete examples, we developed two systems, "SelectBukuma" and "MasteryEye" that search long-term "Web pages" and "retrieval keywords".

"SelectBukuma" is a system that can search specialized long-term Web pages. First, we analyzed social bookmark time data, and we developed a retrieval system that can search long-term Web pages. "SelectBukuma" can search long-term Web pages by input keywords in the screen. For example, the system is aimed for search specialized Web pages that can get structured knowledge, Web service and some references.

"MasteryeEye" is a system that can search long-term retrieval keywords. To achieve the purpose, "MasteryEye" uses Google's time-series data of search times. In the case when a user wants to acquire new field knowledges, the user can get standard retrieval keywords about that field.

We developed these two systems, and we analyzed and evaluated about systems. In this paper, we indicate the result of analysis and the result of evaluation experiment, and discuss about availability and the signification of "long-term retrieval".

Key Word

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

Graduate School of Media and Governance, Keio University
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 6

第 2 章 背景と関連領域 **9**

2.1 はじめに 10

2.2 情報検索 10

2.3 情報フィルタリング 11

2.4 まとめ 11

第 3 章 長期的検索の提案 **13**

3.1 長期度 14

3.1.1 長期度とは 14

3.1.2 長期度 I 14

3.1.3 長期度 II 14

3.1.4 長期度 III 16

3.1.5 利用パターンの時系列モデル 16

3.1.6 長期度の分析 20

3.2 長期的な情報 27

3.2.1 長期的な情報の定義 27

3.2.2 長期的な情報と短期的な情報の特徴 28

3.2.3 本研究で対象とする情報 28

3.2.4 長期的な情報の具体例 30

3.2.5 長期的な期間について 31

3.3 長期的検索の提案 31

第 4 章 長期的な Web ページの検索 **33**

4.1 背景 34

4.2 目的 34

4.3 ソーシャルブックマーク 35

4.4 分析 35

4.4.1 ソーシャルブックマークデータの分析 36
4.4.2 時間情報に関する分析 37
4.5 セレクトブクマの提案 40
4.5.1 セレクトブクマの概要 40
4.5.2 セレクトブクマの設計思想 40
4.5.3 セレクトブクマ画面構成 41
4.5.4 セレクトブクマインタフェース 41
4.5.5 長期度と検索ランキングロジック 43
4.5.6 システム構成 44
4.6 セレクトブクマの運用と主な検索結果 44
4.6.1 セレクトブクマの運用 44
4.6.2 主な検索結果 50
4.6.3 セレクトブクマの利用経験 50
4.7 評価実験 50
4.7.1 実験目的 50
4.7.2 実験方法 51
4.7.3 実験結果 52
4.7.4 時間情報を利用する効果について 67
4.7.5 各実験結果からの考察 67
4.8 課題 69
4.9 まとめ 69

第 5 章 長期的な検索キーワードの提示 71
5.1 背景 72
5.1.1 背景 72
5.1.2 目的 72
5.2 分析 73
5.2.1 利用データ 73
5.2.2 時間情報に関する分析 73
5.3 MasteryEye の提案 77
5.3.1 システム概要 77
5.3.2 関連キーワードの取得 77
5.3.3 長期度の計算 78

5.3.4 システム表示 79
5.4 評価実験 79
5.4.1 実験目的 79
5.4.2 実験方法 80
5.4.3 実験結果 81
5.5 MasteryEye の利用経験 85
5.6 課題 87
5.7 まとめ 87

第 6 章 関連研究 **89**
6.1 情報検索 90
6.1.1 情報検索のアルゴリズム 90
6.1.2 関連ページの検索 90
6.1.3 ソーシャルな検索 91
6.1.4 検索インタフェース 92
6.1.5 さまざまな情報を利用した検索システム 92
6.2 情報フィルタリング 93
6.2.1 情報フィルタリングのアルゴリズム 93
6.2.2 ソーシャルな情報フィルタリング 93
6.2.3 情報フィルタリングのインタフェース 94
6.2.4 さまざまな情報フィルタリングシステム 94
6.3 ソーシャルブックマークデータの分析 94
6.4 検索キーワード 94
6.5 時間情報の利用 95

第 7 章 考察と展望 **97**
7.1 考察 98
7.1.1 人の行動から得られる時間情報の分析 98
7.1.2 長期度の提案 99
7.1.3 長期的な情報に着目した検索システム 101
7.1.4 長期的な情報の有用性 102
7.2 課題 103
7.2.1 長期度の統一 104
7.2.2 時間情報の収集 104

7.2.3 視覚化とユーザインタフェース 104
7.3 展望 105
7.3.1 さまざまな情報への応用 105
7.3.2 複数の長期的検索システムの連動と統合 105
7.3.3 長期的検索の広がり 106
7.3.4 長期的検索の未来 106

第 8 章 結論 **107**
8.1 研究の成果 108
8.1.1 長期的検索の提案 108
8.1.2 システムの開発と評価 108
8.1.3 長期的な情報と長期的検索の有効性の検討 109
8.2 総括 110

謝辞 **113**

研究業績 **115**

参考文献 **117**

付録 A **129**

付録 B **141**

図目次

1.1 本論文の構成 7

2.1 ターゲットとする情報 12

3.1 べき分布のデータ 17
3.2 ポアソン分布のデータ 17
3.3 べき分布より大きい値となる分布 18
3.4 べき分布より小さい値となる分布 18
3.5 2つのべき分布データ 19
3.6 2つのべき分布データの合計 19
3.7 長期度II 上位2ページ 21
3.8 長期度II 下位2ページ 21
3.9 長期度III 1位のページ 22
3.10 長期度III 50位のページ 22
3.11 長期度III 100位のページ 25
3.12 長期度Iと長期度IIの順位相関 26
3.13 長期度Iと長期度IIIの順位相関 26
3.14 長期度IIと長期度IIIの順位相関 27
3.15 対象とする情報 29
3.16 対象とする情報と提案システム 29

4.1 はてなブックマーク 36
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 37
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 38
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 38

4.5 Type1 の Web ページの種類 39
4.6 Type2 の Web ページの種類 40
4.7 セレクトブクマ画面 42
4.8 システム構成図 45
4.9 DB のテーブル構成図 45
4.10 セレクトブクマ人気の検索キーワード 46
4.11 3 人以上を適合文書とした場合の適合率 (java) 63
4.12 8 点以上を適合文書とした場合の適合率 (java) 63
4.13 3 人以上を適合文書とした場合の適合率 (健康) 64
4.14 8 点以上を適合文書とした場合の適合率 (健康) 64
4.15 3 人以上を適合文書とした場合の適合率 (映画) 65
4.16 8 点以上を適合文書とした場合の適合率 (映画) 65

5.1 キーワード「ハンカチ王子」の検索時系列データ 74
5.2 キーワード「ハンカチ王子」の検索時系列データ 74
5.3 キーワード「動画」の検索時系列データ 75
5.4 キーワード「動画」の検索時系列データ 75
5.5 キーワード「卒業論文」の検索時系列データ 76
5.6 キーワード「卒業論文」の検索時系列データ 76
5.7 「java」での検索結果画面 79

表目次

3.1 各長期度での上位10件 23
3.2 各長期度での下位10件 24
3.3 各長期度のスピアマンの順位相関係数 26
3.4 長期的な情報と短期的な情報の特徴 28

4.1 収集したデータ量 36
4.2 技術分野上位3キーワードでの各検索結果上位10件 46
4.3 趣味分野上位3キーワードでの各検索結果上位10件 47
4.4 社会・生活分野上位3キーワードでの各検索結果上位10件 48
4.5 その他分野上位3キーワードでの各検索結果上位10件 49
4.6 「java」で検索した場合の検索結果上位10件 52
4.7 「健康」で検索した場合の検索結果上位10件 53
4.8 「映画」で検索した場合の検索結果上位10件 54
4.9 「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） . 56
4.10 「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） . 57
4.11 「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） 58
4.12 「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） 59
4.13 「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） 60
4.14 「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） 61
4.15 全適合文書の件数 62
4.16 再現率(java) 63
4.17 再現率(健康) 64
4.18 再現率(映画) 65
4.19 「java」での検索結果の得点の合計値 66
4.20 「健康」での検索結果の得点の合計値 66
4.21 「映画」での検索結果の得点の合計値 67

5.1 関連キーワードデータベースのデータ量 78

5.2 被験者属性 81
5.3 選択されたキーワードの数 82
5.4 知っているかどうかの比較（手法1と手法2） 82
5.5 知っているかどうかの比較（手法1と手法3） 82
5.6 長期間利用かどうかの比較（手法1と手法2） 83
5.7 長期間利用かどうかの比較（手法1と手法3） 83
5.8 重要かどうかの比較（手法1と手法2） 84
5.9 重要かどうかの比較（手法1と手法3） 84
5.10 長期間利用と重要なものの関係 85
5.11 取得した関連キーワード一覧 86

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

情報やデジタルコンテンツの短命化が進んできている．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．中でも Web 上の情報は，急速に増加しているとともに，玉石混交化が進んできている [62] し，例えば Twitter[124] などのマイクロブログをはじめとした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 [57]．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．例えば，Twitter 情報の中には，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場合も多い．さらに，このような特性のある Twitter やさまざまなソーシャルメディアによって急速に広まる Web ページも増加してきているが，急速に広まるということは急速に廃れる可能性が高い．こういったことから，利用される期間の短い情報やデジタルコンテンツがだんだん増加してきていると言える．さらに，Web 以外に関しても，製品などのライフサイクルの短命化 [60] が起こっている．

このように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などでは，短期的な情報が提示される可能性が増加してきた．短期的な情報でも一時的に人気や流行となる場合があり，その場合多くのシステムやサービスによって提示される．こういった流行や人気の情報は，多くのメディアやソーシャルメディアで取り上げられることが多いため，人々の目に触れる機会は多い．このような背景に対して，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．

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 [101] 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てみる．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．筆者自身は，長期的に利用される情報に関心があるが，どちらの情報が欲しいかはユーザの好みだけでなく状況によっても変わってくる．そのため本研究では，商品だけでなく 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，こういった情報を見つけ

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，人々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．

***2** 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．

***3** 情報フィルタリング [29]

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．

***4** 情報レコメンデーション [40]

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．

***10** スピアマンの順位相関係数

順位データから求められる相関の指標である．ノンパラメトリックな指標であり，2 つの変数の分布について何も仮定せずに，変数の間の関係を評価するものである．

***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．

$$\text{適合率} = \frac{R}{N} \tag{1.1}$$

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．

$$\text{再現率} = \frac{R}{C} \tag{1.2}$$

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

***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** Google Insights for Search[117]

Google が提供している検索キーワードの検索ボリュームを分析することができるサービスである．指定した検索キーワードに対して，Google での 1 週間ごとの検索ボリュームを取得することができる．2004 年以降の検索ボリュームの時系列データを取得でき，また地域ごとのデータなどより詳細な分析も可能である．ここで検索ボリュームとは，1 週間ごとの検索回数の最大値を 100 としたときの 100 に対する割合の値である．2013 年 1 月 10 日現在，このサービスは Google トレンドと合併し，Google トレンドにて同様の機能を利用することができる．

***16** Google トレンド [118]

Google が提供している検索キーワードの検索ボリュームを分析することができるサービスである．Google Insights for Search とほぼ同等のサービスを提供している．

***17** 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．

***18** 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．

1.4. 論文の構成

本論文の構成は，以下の通りである（図 1.1）．

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「長期度」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，長期的に利用されている Web ページを検索するシステム「セレクトブクマ」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　関連研究

第 6 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 7 章　考察と展望

第 7 章では，本研究の考察と展望について述べる．

第 8 章　結論

第 8 章では，本研究の成果をまとめ，本論文を総括する．

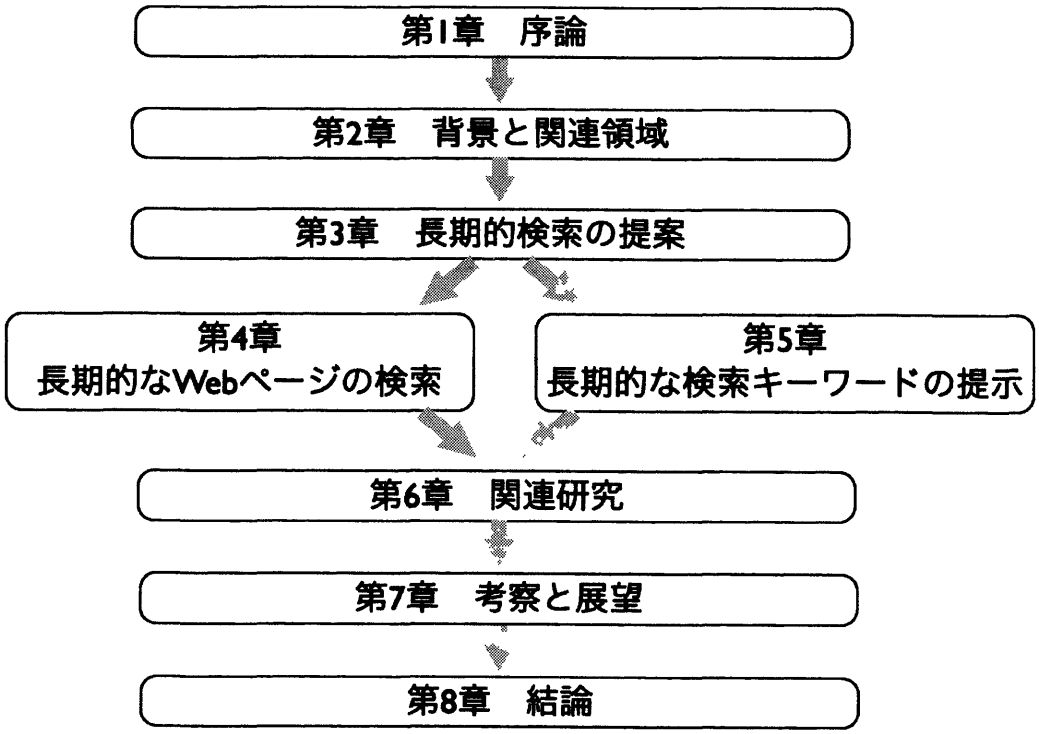

図 1.1: 本論文の構成

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

情報爆発により，人々は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，さまざまな情報検索手法，情報フィルタリング手法，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が主流であった [69]．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．例えば，現在の Google 検索 [116] や Bing[109] 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 [24][28] 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．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[35] 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，収集に関する研究 [51][33]，ソーシャルブックマークの情報を利用した情報検索に関する研究 [50][54]，レシピサービスの情報を利用した情報検索，情報推薦に関する研究 [102][71] などが幅広く行われている．

こういった理由からも筆者は，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 [59]．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 [41]．

情報フィルタリングや情報レコメンデーション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 [86][68][25][13]．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 [19][44]．また，はてなブックマーク [119] や食べログ [123] などのように人気の高い情報を見つけるシステムも多く存在する．

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ことが，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．

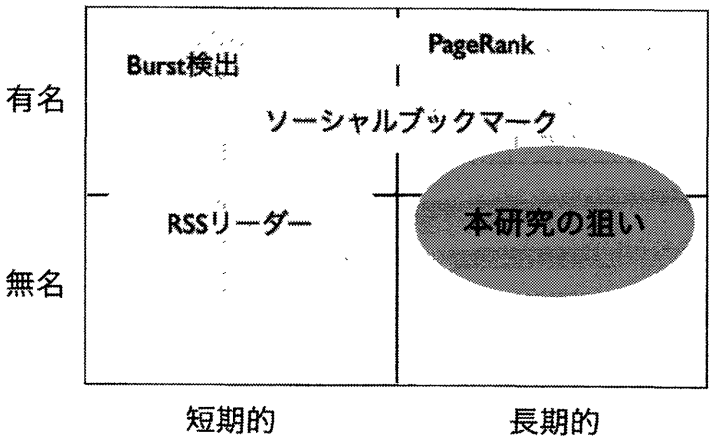

図 2.1: ターゲットとする情報

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，長期的な情報を検索するための手法である「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．長期度の指標は，長期度が高いほど長期的に利用されている情報とする．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．本研究では，長期度の具体的な計算手法として以下 3 つの手法を提案する．ここで，それぞれの長期度計算手法によって求められる長期度を長期度 I，長期度 II，長期度 III と呼ぶ．またこれらの長期度は，長期的な情報の順位を出すための尺度とする．

3.1.2. 長期度 I

長期度 I は，利用された日数が多いほど長期度が高くなる手法である．以下に長期度 I の数式を示す．

$$LongDegreeI = Days \tag{3.1}$$

$LongDegreeI$: 長期度 I
$Days$: 利用された日数

ここで日数とは，例えば，2011 年 4 月 20 日，2011 年 5 月 8 日，2012 年 7 月 13 日という 3 日間に利用された場合，3 とする．また，ここでの日付は，日本標準時（JST）での日付としている．

この手法のメリットは，非常にシンプルなため，ユーザにとってなぜこの情報が長期的なのか分かりやすいということと，データ量が少ないデータにも適用できることである．この手法のデメリットは，利用数自体が少ないものは，自然と除外されてしまうことと，1 日あたりの利用数が 1 でも 100 でも同じとみなしている点である．

3.1.3. 長期度 II

長期度 II は，単位時間ごとの利用数が，べき分布のデータよりどの程度大きいかで長期度を求める手法である．この手法では，最初に，実際の時系列データを利用数の多い順に並べる．次に，並び替えたデータからべき分布に基づくデータを生成する．以下に長期度 II の数式を示す．

$$LongDegreeII = \sum_{x=1}^{n} (Times_x - ax^k) \tag{3.2}$$

(3.3)

$LongDegreeII$: 長期度 II
n : データの合計数
x : 単位時間ごとの利用数の順位
$Times_x$: 単位時間ごとの利用数
ax^k : べき分布データ
a : べき分布の定数
k : べき分布のスケーリング指数（定数）

ここで ax^k は，両対数グラフにおいて y 軸の最大値と x 軸の最大値を結ぶ直線を描くようなべき分布データとし，a と k の値を設定する．このとき，y 軸の最大値の点は順位のため x=1 となり，x 軸の最大値の点は，y=1 とする．そして，実際の単位時間ごとの利用数データからこのようなべき分布モデルの値を引いた値の合計値を長期度とする．ここで，実際の時系列データと生成するべき分布に基づくデータは，単位時間ごとの利用数（縦軸）の最大値を固定して，最大値との割合の値として計算を行う．このように最大値を固定する理由は，最大値のボリュームが大きく異なると，長期度の値もそのボリュームに比例して変わってきてしまうからである．

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理論 [1] より，多くのデータは，単位時間ごとの利用数がべき分布になる可能性が高い．また，べき分布になるデータは「バースト」する [2]．ここで言う「バースト」とは，一定の範囲内では通常の正規分布をするが，突然急激な頻度の増加がみられ，また元の状態に戻るというような動きを意味する．Barabasi[2] によると，この「バースト」の仕組みを分析すると，人間の行動パターンの中でもべき分布になるものは多く，例えばメールの返信時間などもべき分布になることが示されている．

急速に流行したデータは，急激な利用数の増加が起こっているため，バーストしていることが分かる．これに対して，長期的に安定して利用されているデータは，利用数の変動が少ない．急激な利用数の増加が起こった場合は，単位時間ごとの利用数を多い順に並べたとき，急速に値が落ちる．これに対して，急激な利用数の増加が起こらなかった場合は，単位時間ごとの利用数を多い順に並べたとき，緩やかに値が落ちる．このため，急速に流行したデータと長期的に安定して利用されているデータでは，後者の方がべき分布との差分が大きくなる．

この手法のメリットは，単に長期間利用されているという指標だけでなく，長期間利用されて

いて，かつ安定して利用されている情報を取得できる点である．その代わり，長期間利用されていても，利用のされ方が安定していないような情報の長期度は小さくなる．またこの手法のデメリットとして，データ量が少ない場合適用しにくいことと，ユーザにとってなぜ長期的なのか分かりにくいことがあげられる．

3.1.4. 長期度 III

長期度 III は，べき分布の傾きの大きさを長期度とする手法である．この手法では，単位時間ごとの利用数データを利用数の多い順に並べる．そして，最小二乗法を利用して近似したものをべき分布とする．長期度 III の数式を以下に示す．

$$PowerLaw = ax^k \tag{3.4}$$

$$LongDegreeIII = k \tag{3.5}$$

$LongDegreeIII$：長期度 III
$PowerLaw$：べき分布データ
a：べき分布の定数
k：べき分布のスケーリング指数（定数）

単位時間あたりの利用数がべき分布に近いデータになる場合，最小二乗法によりべき分布データに近似して，傾き k を求める．そして，傾き k の値が大きいほど長期度が高いとする．ここで，べき分布では，傾きの値はマイナスになるため，値が大きいほど緩やかな傾きとなる．そして，傾きが緩やかなほどロングテールになる傾向があり，長期的に利用されていることになる．

この手法のメリットは，長期間利用されているという指標に重きをおきながらも，安定して利用されているかどうかという指標も含んでいることである．この手法のデメリットは，べき分布から大きく外れるようなデータに適用すると，信頼性が落ちることである．

3.1.5. 利用パターンの時系列モデル

ここでいくつかの利用パターンの時系列データモデルについて仮説を立て，長期度 II と長期度 III にどのような特性があるのか分析する．一般的にべき分布に基づくデータを両対数グラフにプロットすると，図 3.1 となる．また，ポアソン分布に基づくデータを両対数グラフにプロットすると，図 3.2 となる．

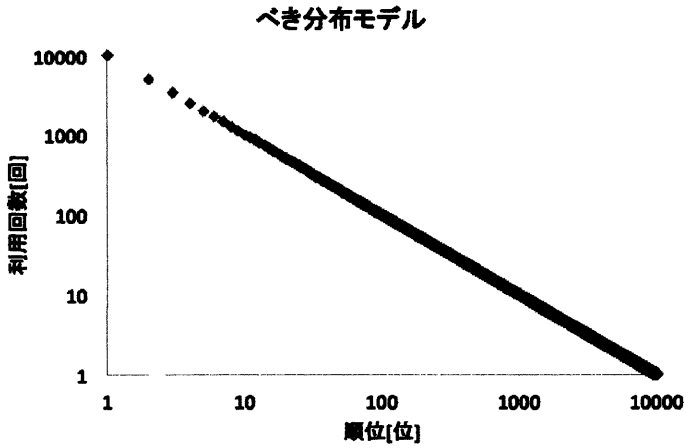

図 3.1: べき分布のデータ

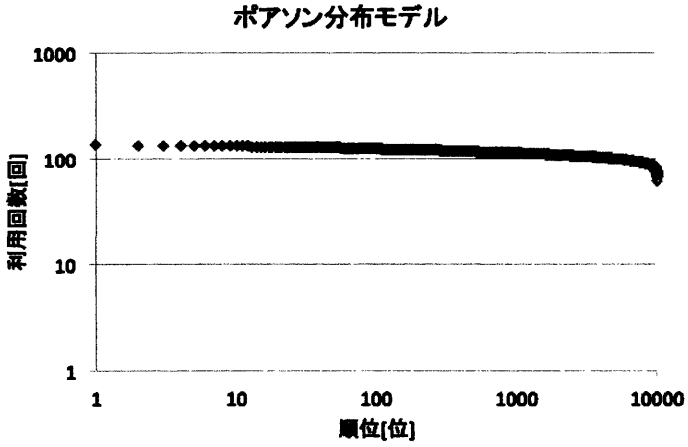

図 3.2: ポアソン分布のデータ

実際に Web ページへのアクセスパターンや，ブックマークパターン，検索キーワードの検索ボリュームのパターンの時系列データを値の大きい順に並べ両対数グラフにプロットしたとき，ある程度べき分布から外れるパターンも存在する．例えば，図 3.3 は，べき分布と比較して最初緩やかに落ちてその後落ち方が急になっている．こういった分布は，べき分布と比較して大きな値となる．また，図 3.4 は，ある Web ページのソーシャルブックマークによる 1 日ごとのブックマーク数を多い順に並べたもののグラフであるが，べき分布と比較して最初急激に落ちてその後落ち方が緩やかになっている．こういった分布は，べき分布と比較して小さな値となる．図 3.3 は，Google Insights for Search[117] から取得した「卒業論文」という検索キーワードの 1 週間ごとの検索ボリュームデータを利用している．図 3.4 は，はてなブックマーク [119] から取得した「Dreamweaver の代わりになるフリーソフト「Aptana」 - GIGAZINE」というタイトルの Web ページの 1 日ごとのブックマーク数データを利用している．

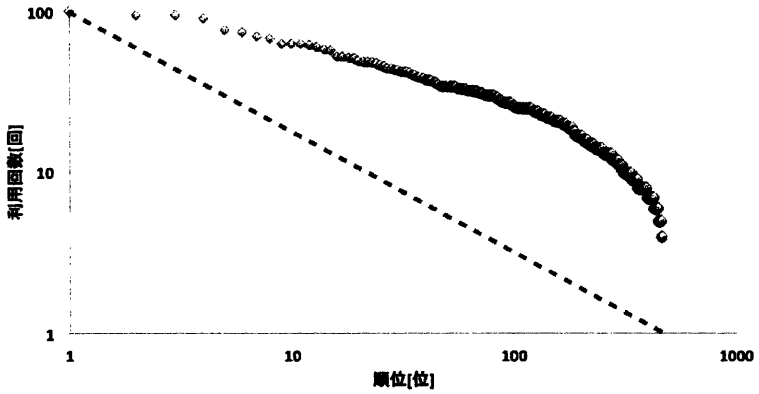

図 3.3: べき分布より大きい値となる分布

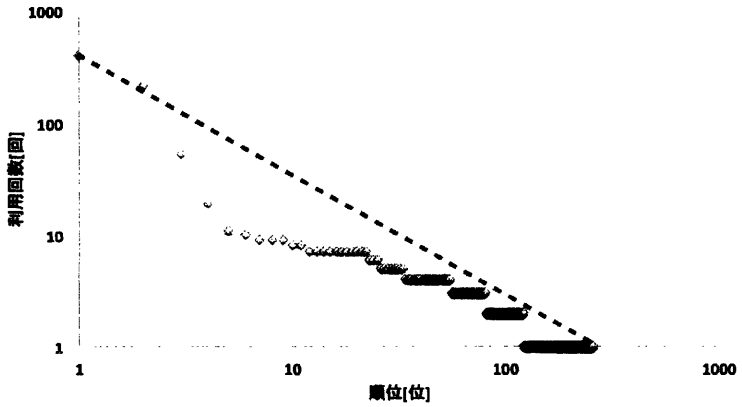

図 3.4: べき分布より小さい値となる分布

なぜ，このようにべき分布から外れるのか，モデルを考えて仮想データをプログラムによって生成し，実験してみる．

まず，べき分布より小さい値となる分布のモデルについて考える．これはなんらかの理由があって，一時的に急激に利用された可能性が高いと考えた．そのため，図3.5のような傾きが急なべき分布データモデルと傾きが緩やかなべき分布データモデルを生成し，これらを足してみる．そうすると，図3.6のようにべき分布の下に山がきて，べき分布より小さい値をとる分布となる．このことから小さな値となる分布のモデルは，利用されている期間自体は長くても，一時的に急激に利用され，流行り廃りのあったモデルであることが分かる．

また，このようなモデルでは，長期度IIの値が小さくなることが分かる．このことから長期度IIでは，利用された期間自体は長くても，流行り廃りが激しいような情報の長期度が低くなることが分かる．

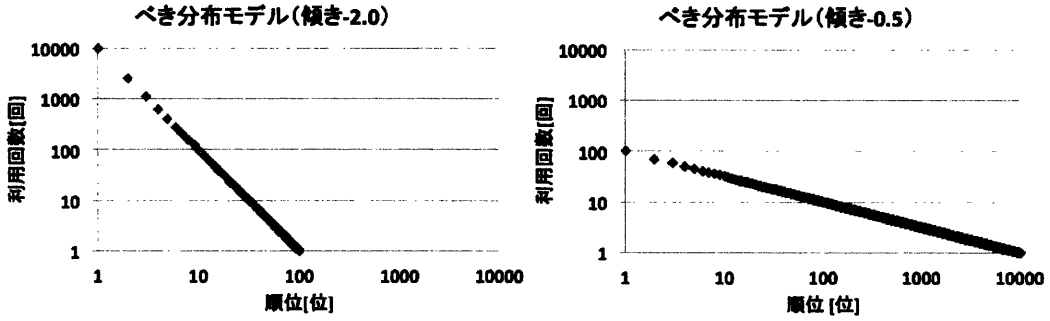

図 3.5: 2つのべき分布データ

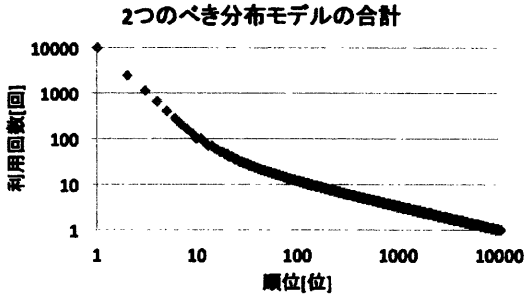

図 3.6: 2つのべき分布データの合計

次に，べき分布より大きい値となる分布について考える．これは，べき分布と比較して，平均値に近い値が多いことが分かる．そのため，べき分布とポアソン分布の間にあたるような分布ではないかと考えた．ポアソン分布では，時間の変化によって，利用される確率は変わらないが，

実際世の中のほとんどの情報は，時間の変化によって，利用される確率が変化すると考えられる．べき分布と比較して時間変化による確率の変化は少ないが，常に一定ではない場合にこのようなモデルになるのではないかと考えた．

また，このようなモデルでは，長期度 II の値が大きくなることが分かる．このことから長期度 II では，時間変化によってあまり利用される確率が変化しないような長期的に安定している情報の長期度が高くなることが分かる．

3.1.6. 長期度の分析

実データに各長期度を適用し，長期度の分析を行う．ここで分析に利用する実データは，日本で最大級のソーシャルブックマークサービスであるはてなブックマークのデータとする．ソーシャルブックマークサービスは，ブックマーク情報を共有するサービスであり，はてなブックマークでは，多くのユーザからブックマークされた Web ページが時間情報と共に公開されている．2008 年 9 月時点で，ブックマーク数 100 以上のページをランダムで 100 ページ取得して，それらのデータを分析対象とした．

取得した 100 ページに対して，長期度 I，長期度 II，長期度 III を利用して長期度を計算した．ここで，長期度 I，II，III の利用数は，ブックマーク数とした．また，長期度 II，III における単位時間を 1 日とした．そのため，単位時間あたりの利用数は，1 日あたりのブックマーク数となる．また，長期度 II における縦軸の最大値は 100 とした．

まず，1 日あたりのブックマーク数の時系列データが，べき分布になっているか調べるために，以下の図 3.7，図 3.8 に長期度 II で計算した，上位 3 件と下位 3 件のグラフを示す．これらのグラフは，1 日毎のブックマーク数を多い順に並べたものとする．また，グラフの縦軸が 1 日毎のブックマーク数，横軸が順位であり，グラフは両対数グラフである．長期度 II は，べき分布との差分の値を長期度としているため，この値の上位 3 件と下位 3 件は，取得した 100 ページの中でも，べき分布との差分が大きいページとなる．これらのグラフから分かるように，べき分布からあまり大きく離れたグラフになっていないことが分かる．このことから，ブックマーク数が 100 以上多い場合，1 日毎のブックマーク数がべき分布に近づくことが分かった．この結果から，ブックマーク数が多い場合，データがべき分布に近い値になるため，長期度 III のべき分布に近似したときの傾きによって長期度を求める手法も効果的であると考えられる．

そこで，長期度 III の値が 1 位，50 位，100 位のページのデータとそのデータを最小二乗法によって，べき分布に近似したデータを以下の図 3.9，図 3.10，図 3.11 に示す．これらのグラフから分かるように，元データと最小二乗法によってべき分布に近似したデータで，大きな違いが出ないことと，Web ページによって傾きの値に大きな違いが出ることとが分かった．

また，それぞれのべき分布の傾きは，小数点第 3 位以下を四捨五入して，1 位のページが −0.46，50 位のページが −1.46，100 位のページが −2.83 となった．

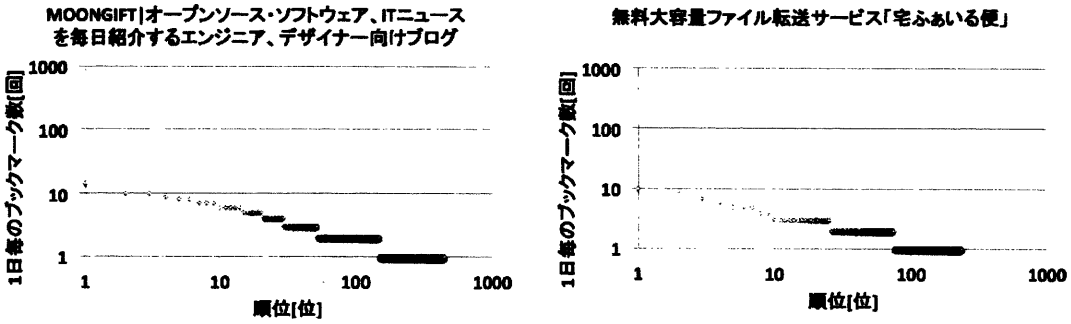

図 3.7: 長期度 II 上位 2 ページ

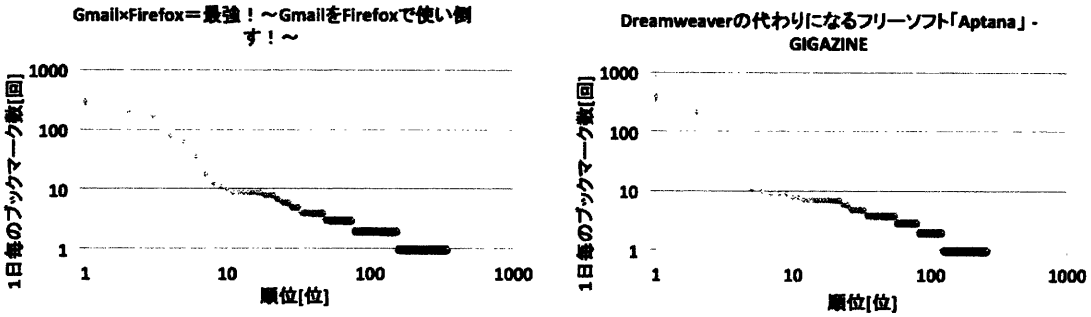

図 3.8: 長期度 II 下位 2 ページ

次に，表 3.1，表 3.2 にそれぞれの計算手法で計算した，長期度上位 10 件と下位 10 件のページのタイトルを示す．

表 3.1，表 3.2 の太字は，2013 年 1 月 9 日の時点でリンク切れになっていないページで，太字でないものがリンク切れになっているページとする．

これらの結果のタイトルから，長期度 I の上位 10 件には，Web サービスが 4 ページ，まとめサイトが 2 ページ，ツール紹介が 2 ページ含まれていたのに対して，下位 10 件は，ニュースや日記に関するものが 9 ページ含まれていた．また，リンク切れのページは，上位 10 件には 2 ページ含まれていたのに対して，下位 10 件には 6 ページ含まれていた．

長期度 II の上位 10 件には，Web サービスが 5 ページ，ブログトップが 2 ページ，まとめサイトが 2 ページ含まれていたのに対して，下位 10 件にはツールや Tips の紹介が 5 ページ，Web サービスが 3 ページ含まれていた．さらに，長期度 I の上位 10 件に含まれているページが，長期度 II の上位 10 件中に 6 ページ，下位 10 件中に 4 ページとすべて含まれていることが分かった．また，

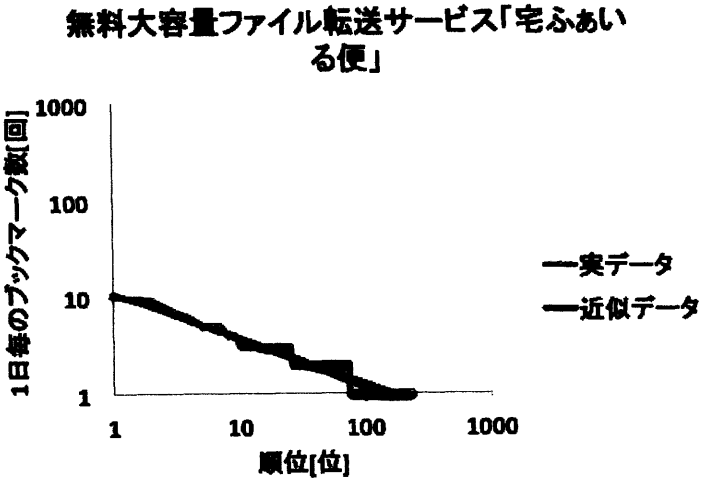

図 3.9: 長期度 III 1 位のページ

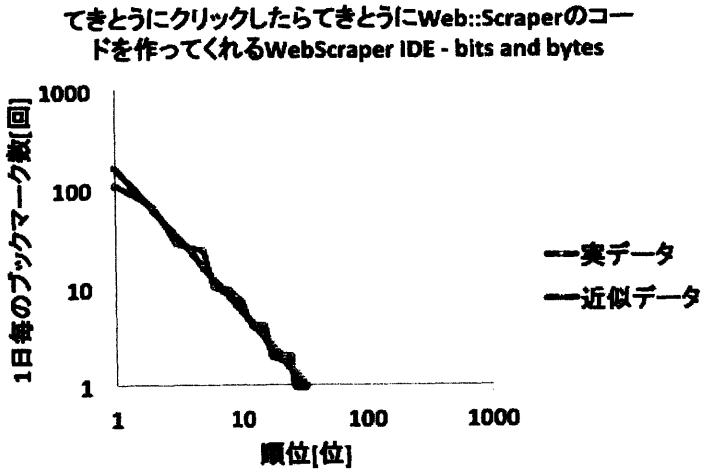

図 3.10: 長期度 III 50 位のページ

表 3.1: 各長期度での上位 10 件

順位	長期度 I	長期度 II	長期度 III
1	MOONGIFT—オープンソース・ソフトウェア，ITニュースを毎日紹介するエンジニア，デザイナー向けブログ	MOONGIFT—オープンソース・ソフトウェア，ITニュースを毎日紹介するエンジニア，デザイナー向けブログ	無料大容量ファイル転送サービス「宅ふぁいる便」
2	Gmail × Firefox =最強! ～ Gmail を Firefox で使い倒す！～	無料大容量ファイル転送サービス「宅ふぁいる便」	finalvent の日記
3	FrontPage - EclipseWiki	FrontPage - EclipseWiki	MOONGIFT—オープンソース・ソフトウェア，ITニュースを毎日紹介するエンジニア，デザイナー向けブログ
4	Dreamweaver の代わりになるフリーソフト「Aptana」- GIGAZINE	トラックフィード - リンクが張られたら瞬時に通知するアクセス解析ツール	FrontPage - EclipseWiki
5	無料大容量ファイル転送サービス「宅ふぁいる便」	お笑いコピペ選手権の記録	EyesPic - フリー画像素材
6	Google ソースコード検索	EyesPic - フリー画像素材	トラックフィード - リンクが張られたら瞬時に通知するアクセス解析ツール
7	トラックフィード - リンクが張られたら瞬時に通知するアクセス解析ツール	SEO/SEM 対策をするなら無料で使えるツールが豊富の FerretPLUS（フェレットプラス）	お笑いコピペ選手権の記録
8	EyesPic - フリー画像素材	コリス	コリス
9	白地図，世界地図，日本地図が無料【白地図専門店】	ウーマシャンプー【口コミで話題】	Apple - Support
10	finalvent の日記	finalvent の日記	SEO/SEM 対策をするなら無料で使えるツールが豊富の FerretPLUS（フェレットプラス）

表 3.2: 各長期度での下位 10 件

順位	長期度 I	長期度 II	長期度 III
91	宮崎勤死刑に思い出すこと.	高精度計算サイト	「私，ここで結婚できますか？」と聞く女子大生，どう思います?: NBonline(日経ビジネス オンライン)
92	mixi 規約改定問題　「ユーザーが著作者の時代」にまた繰り返す大騒動 (1/2) - ITmedia ニュース	私的録音録画小委員会:反対意見多数でも「ダウンロード違法化」のなぜ (1/2) - ITmedia ニュース	なぜプログラミングが楽しくなくなったのか・日本的ソフトウエア観 (1) -ビジネス-最新ニュース:IT-PLUS
93	総務省:通信・放送法制統合へ　ネット情報も規制 - 毎日 j p (毎日新聞)	ホームページ - Google Maps API - Google Developers	痛いニュース (∀ '): 「『私は手取り３４万円の薄給… 橋下知事は大阪のこと考えろ』…大阪府職員の甘えた考え」…デジタルマガジン
94	なぜプログラミングが楽しくなくなったのか・日本的ソフトウエア観 (1) -ビジネス-最新ニュース:IT-PLUS	Windows から Mac に"Switch"した人のための 10 の Tips	asahi.com:暗い所で本「目悪くなる」医学的根拠ないと米チーム
95	京．電車．現実と虚構の狭間で - 深町秋生のベテラン日記	バッハからワーグナーまで，無料のクラシック音楽を集めた「Classic Cat」 - GIGAZINE	[mixi] カズ←金欠 ('A') さん — 交通事故　なんで俺らが
96	マンガや雑誌に埋もれ，仙台市の男性死亡　地震による可能性（産経新聞） - Yahoo!ニュース	白地図，世界地図，日本地図が無料【白地図専門店】	京．電車．現実と虚構の狭間で - 深町秋生のベテラン日記
97	死刑執行の音をラジオで放送　文化放送が５月に特別番組（SANSPO.COM ）	Google ソースコード検索	おっくせんまんの歌詞ができるまで． - Myrmecoleon in Paradoxical Library. ニコニコ分室 - ニコニコ部
98	サッカーを毛嫌いする理由「読売テレビ」（ツカサネット新聞） - Yahoo!ニュース	黄金比をサイトのデザインに取り入れる簡単な３つの方法 — コリス	明石家さんまの眠らない哲学 - てれびのスキマ
99	明石家さんまの眠らない哲学 - てれびのスキマ	Dreamweaver の代わりになるフリーソフト「Aptana」 - GIGAZINE	死刑執行の音をラジオで放送　文化放送が５月に特別番組（SANSPO.COM ）
100	[mixi] カズ←金欠 ('A') さん — 交通事故　なんで俺らが	Gmail × Firefox ＝最強! ～ Gmail を Firefox で使い倒す！～	マンガや雑誌に埋もれ，仙台市の男性死亡　地震による可能性（産経新聞） - Yahoo!ニュース

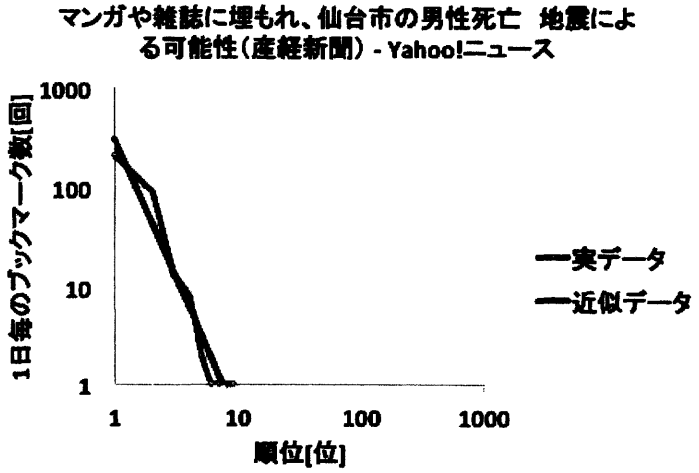

図 3.11: 長期度 III 100 位のページ

リンク切れのページは，上位 10 件には 1 ページ含まれていたのに対して，下位 10 件には 2 ページ含まれていた．

長期度 III の上位 10 件には，Web サービスが 4 ページ，まとめサイトが 2 ページ，ブログトップが 2 ページ含まれていたのに対して，下位 10 件には，ニュースや日記に関するものが 7 ページ含まれていた．また，リンク切れのページは，上位 10 件には 1 ページ含まれていたのに対して，下位 10 件には 5 ページ含まれていた．

このことから各長期度において，上位 10 件には長期的に利用できるようなタイプのページが多いことが分かる．これに対して，長期度 I と長期度 III の下位 10 件には，ニュース系や日記系のページが非常に多く，一時的にしか利用しないタイプのページが非常に多いことが分かった．また，長期度 II の下位 10 件は，ツールや Tips 紹介のページがもっとも多く，少し異なった特性になっていることが分かった．また，上位 10 件と下位 10 件を比較して，データを収集してから約 4 年 4ヶ月たった後，下位 10 件の方がリンク切れとなっているページが多いことが分かった．

また，長期度 I，長期度 II，長期度 III の関係性について分析を行った．ランダムに取得した 100 ページに対して，長期度 I〜III それぞれの値の大きさで，1 位〜100 位まで順位付けを行った．そして，それぞれの順位を利用して，スピアマンの順位相関係数を求めた．各長期的の順位の相関を取った場合の散布図を図 3.12〜図 3.14 に示す．ここで図の縦軸と横軸は各長期度の順位となり，図の点は各 Web ページとなる．また，スピアマンの順位相関係数の値を表 3.3 に示す．

表 3.3 を見ると分かるように，長期度 I と長期度 III は相関係数が 0.94 となり，強い相関があった．また長期度 I と長期度 II は，相関係数が 0.17 となり，ほとんど相関がなかった．長期度 II と長期度 III は，相関係数が 0.27 となり弱い相関があった．

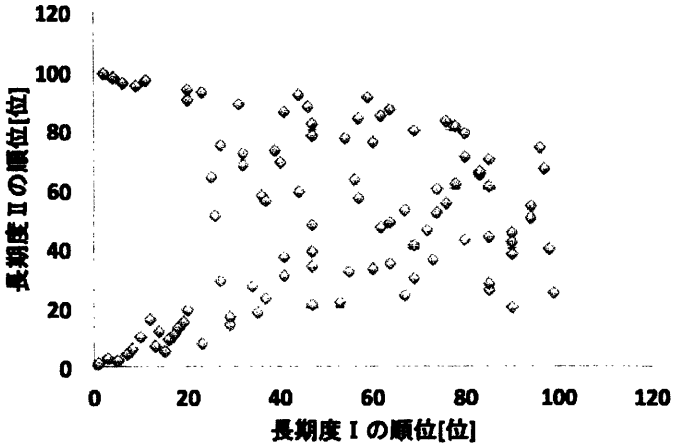

図 3.12: 長期度 I と長期度 II の順位相関

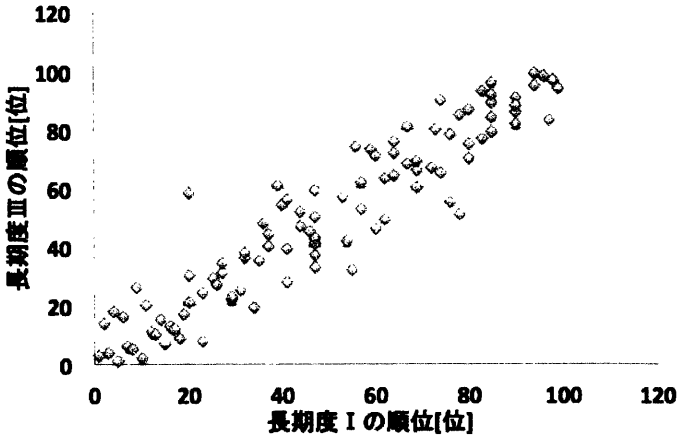

図 3.13: 長期度 I と長期度 III の順位相関

表 3.3: 各長期度のスピアマンの順位相関係数

比較する長期度	相関係数
長期度 I と長期度 II	0.17
長期度 I と長期度 III	0.94
長期度 II と長期度 III	0.27

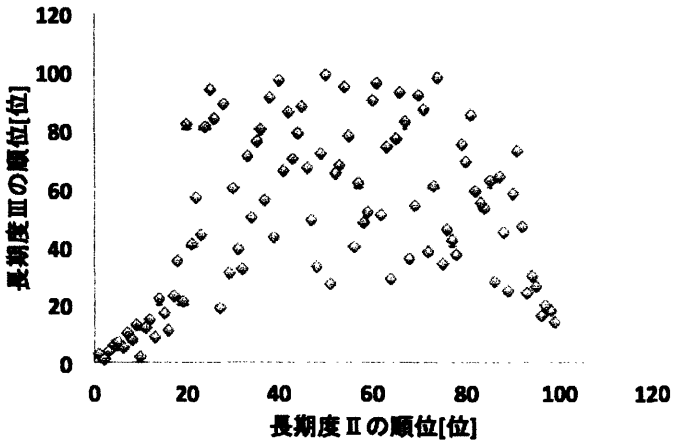

図 3.14: 長期度 II と長期度 III の順位相関

ソーシャルブックマークデータを利用する場合，ほとんどの Web ページについてブックマークされない日が多かったりブックマークされない期間が長かったりするため，安定して利用されてきたかどうかよりも，長期間利用されてきたかどうかを重視することとする．また，長期度 I と長期度 III は相関が強かったため，セレクトブクマでは，長期度 I を利用することとする．

また，検索回数の時系列データを利用する場合，多くのキーワードに関して，ほぼ毎日ある程度以上の回数の検索は行われている．そのため，検索されていた期間というよりもどれだけ安定して長期間検索されてきたかどうかを重視することとする．そこで，MasteryEye では，長期度 II を利用することとする．

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．

また，一言で「これまでに長期的に利用されてきた情報」と言っても，以下の 2 種類の情報に分類することもできる．

(1) 長期間安定して利用されてきた情報

(2) 長期間利用されてきたが，利用のされ方が不安定な情報

本研究全般では，これらの (1)，(2) 両方共長期的な情報と呼んでいるが，利用データの特性によって (1) だけの情報を取得する場合と (1)，(2) 両方の情報を取得する場合がある．常に利用されることが前提となっているデータに適用する場合は，利用のされ方が安定しているかどうかが長期的かどうかの重要な指標となるし，常に利用されないデータに適用する場合は，長期間利用されているかどうかが重要な指標となってくる．

3.2.2. 長期的な情報と短期的な情報の特徴

ここでは，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，以下の表 3.4 のような傾向がある．

表 3.4: 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

3.2.3. 本研究で対象とする情報

世の中のさまざまな情報に対して，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
- 検索キーワード

本研究では，以下 3 点の特徴を持った情報として，これら 2 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

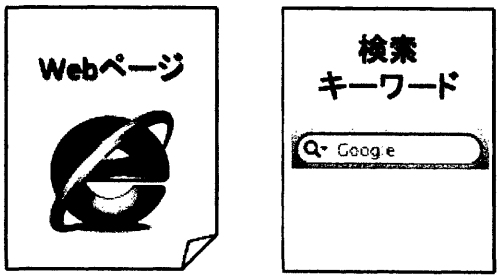

図 3.15: 対象とする情報

(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 3.16 のように，2 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，検索キーワードを選んだより詳細な理由に関しては，第 4 章，第 5 章で述べる．

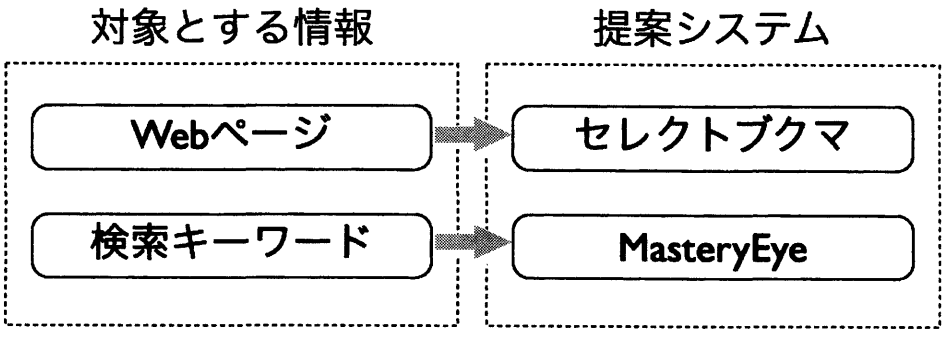

図 3.16: 対象とする情報と提案システム

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，趣味などさまざま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さまざまな共有サービスや Q & A サービスなどのサービス名である．例えば，「YouTube[126]」や「Flickr[115]」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．

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，どちらかというと相対的な表現として用いられる場合が多い．そのため，ど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．

そこで，例えば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ヶ月でも長期的とするし，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しない．

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，長期度に基づき順位付けを行い，システムにより結果を提示する．

第4章　長期的なWebページの検索

概要

本章では，長期的に利用されているWebページを検索するシステム「セレクトブクマ」について述べる．

4.1. 背景

Web 上のコンテンツは多種多様になってきており，Web 上に存在する体系だった知識や長期的に有用な Web ページを手軽に取得することが容易ではなくなってきている．これに対して，既存の Web 検索手法を用いてもある程度こういった情報を取得することは可能だが，既存の手法には，体系だった知識を得られる Web ページや長期的に有用な Web ページを中心に取得する手法はほとんど存在しない．また，自分では有用な Web ページや Web サービスの存在になかなか気付けない場合も多く，知人からの口コミによってその存在に気付かされる場合もある．口コミによってようやく Web の有用な情報に気付くことが多いという場合，ユーザがうまく Web を使いこなせていない可能性が高く，非常に有益な情報を逃している可能性も高い．自分の興味のある分野の有用な Web ページやコンテンツに関しては，Web 上で簡単に発見できるようになれば，こういった問題も解決できる．

これに対して，本研究では 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，長期的に利用できる Web ページや体系だった知識を得られる Web ページは，長期的に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いう仮説を立てた．この仮説を検証するため，ユーザのブックマーク情報を共有できるサービスであるソーシャルブックマークのデータを分析した．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，効率よく長期的に利用できる有用な Web ページや体系だった知識を得られる Web ページを検索するシステム「セレクトブクマ」を提案，実装し，評価実験を行った．

4.2. 目的

本研究では，Web 上から利用者が調べたい分野に対して，体系だった知識や有益な Web コンテンツ，Web サービスを手軽に発見できるようにすることを目的とする．本研究では，こういった Web ページや Web コンテンツは，一時的にのみ利用されるのではなく，長期的に利用されるものが多いと考えた．さらに，長期的に利用されているものは，多くの人々から長期間にわたって新たにブックマークされている可能性が高いという仮説を立てた．そこで，本研究では，このような仮説を検証するとともに，Web システムとして体系だった知識や有益な Web コンテンツ，Web サービスを手軽に発見できるシステムを提案し，実装する．

4.3. ソーシャルブックマーク

ソーシャルブックマークとは，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 サービスである．ソーシャルブックマークでは，folksonomy という情報の分類方法を利用しており，ユーザ各々がブックマークしたページに任意のタグをつけることができる．付与したタグを検索単語として，あとから自分がブックマークした Web ページを検索することも可能である．ソーシャルブックマークを用いることにより，被ブックマーク数が急激に増えたページから人気のブックマークを抽出し，興味深い情報や，最近旬な情報を発見することもできる．

日本国内では，主なソーシャルブックマークサービスとして，はてなブックマーク [119]，Yahoo! ブックマーク [125]，Buzzurl (バザール)[111] などがある．海外での主なソーシャルブックマークサービスとして，del.icio.us[113]，Digg[114] などがある．

ソーシャルブックマークのデータは，ユーザが気に入った Web ページをブックマークし，かつ，ブックマークした Web ページにタグやコメントを付与することが可能であり，ブックマークした時間情報も保持されている．ソーシャルブックマークによって，多くの Web ページには，以下のメタ情報が付与されている．

1. URL
2. タイトル
3. ブックマークしたユーザ ID
4. ブックマークした日時
5. ブックマークしたユーザが付与したタグ名

本研究では，ソーシャルブックマークデータを大量に利用し，上述のメタ情報を有効活用することによって，一般の Web 検索エンジンではできない，情報収集手法や情報フィルタリング手法を提案できないか考えた．そのために，まずソーシャルブックマークデータを大量に収集し，分析した．

4.4. 分析

ここでは，収集したソーシャルブックマークデータとそれに対して行った分析について述べる．

4.4.1. 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 (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，ユーザから約 1 億 5000 万ページ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．

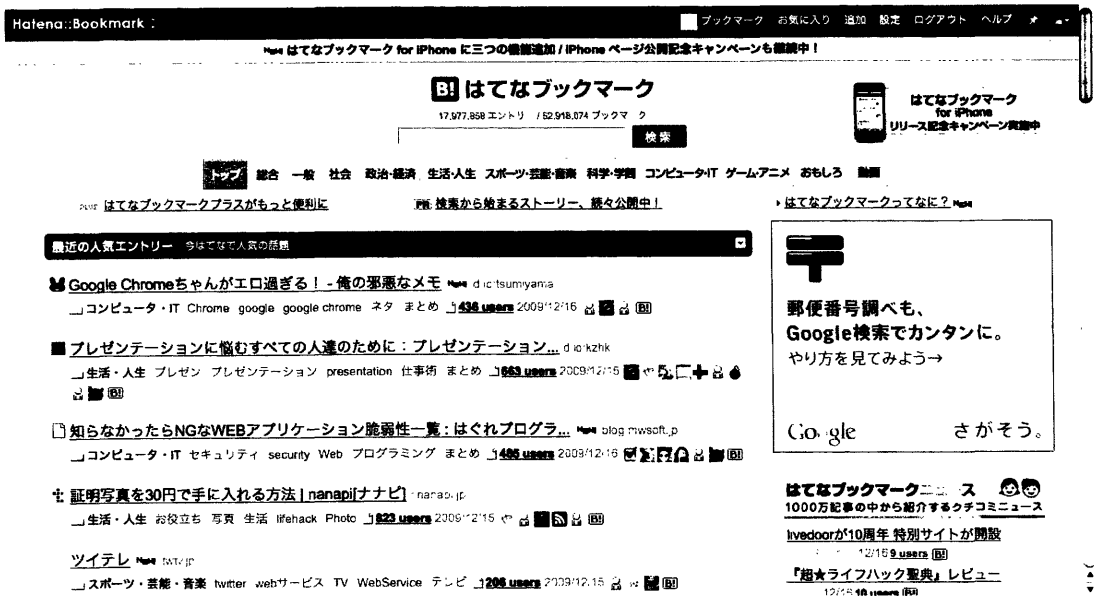

図 4.1: はてなブックマーク

はてなブックマークのデータ収集のために，まずブックマーク数 5 以上のページの URL 一覧を直接株式会社はてなのスタッフからもらった．そして，そのすべての URL に対して，はてなブックマーク API を利用してデータを収集した．データベースに収集したデータの量は，表 4.1 に示す．

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

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．

4.4.2. 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）

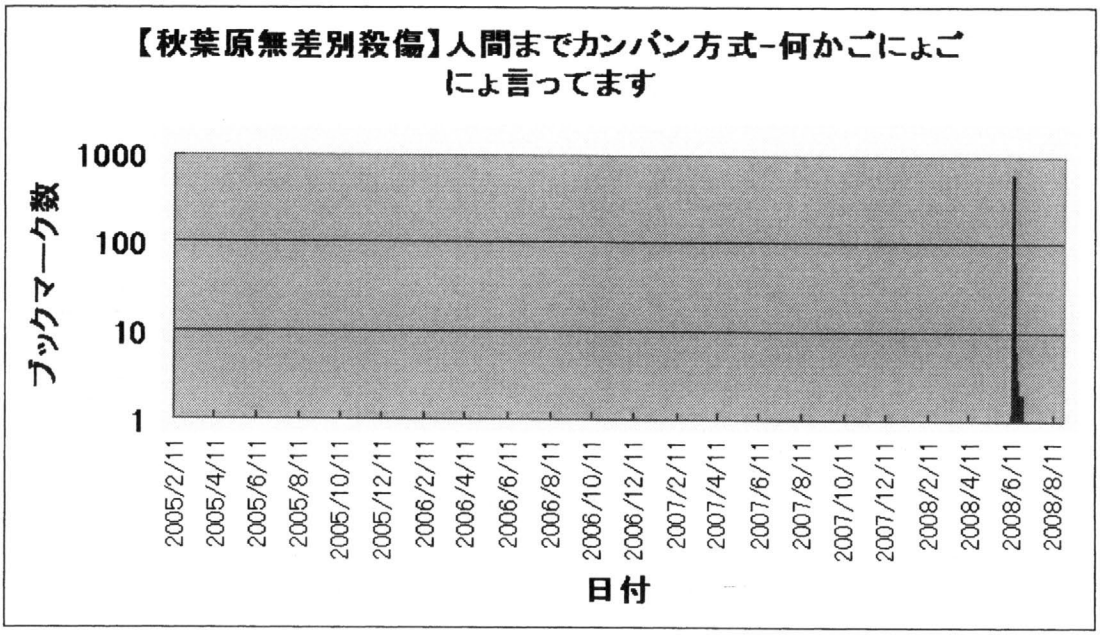

図 4.2: 一時的にブックマークされ，その後ほとんどブックマークされなくなるタイプのページ

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

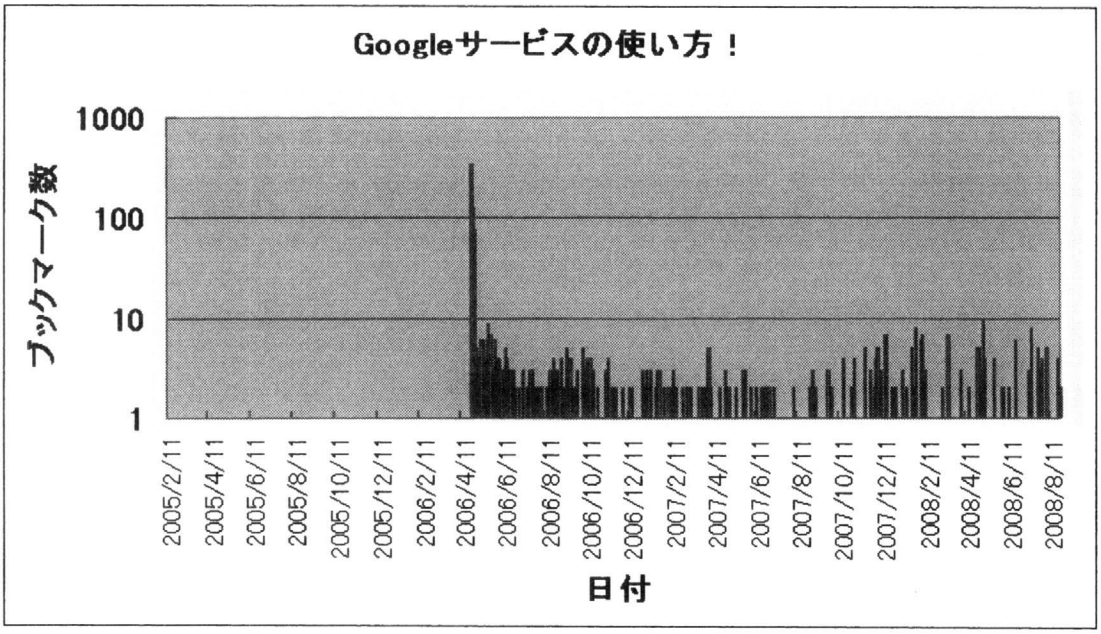

図 4.3: 一時的に大量にブックマークされ，その後も長い間ブックマークされ続けるタイプのページ

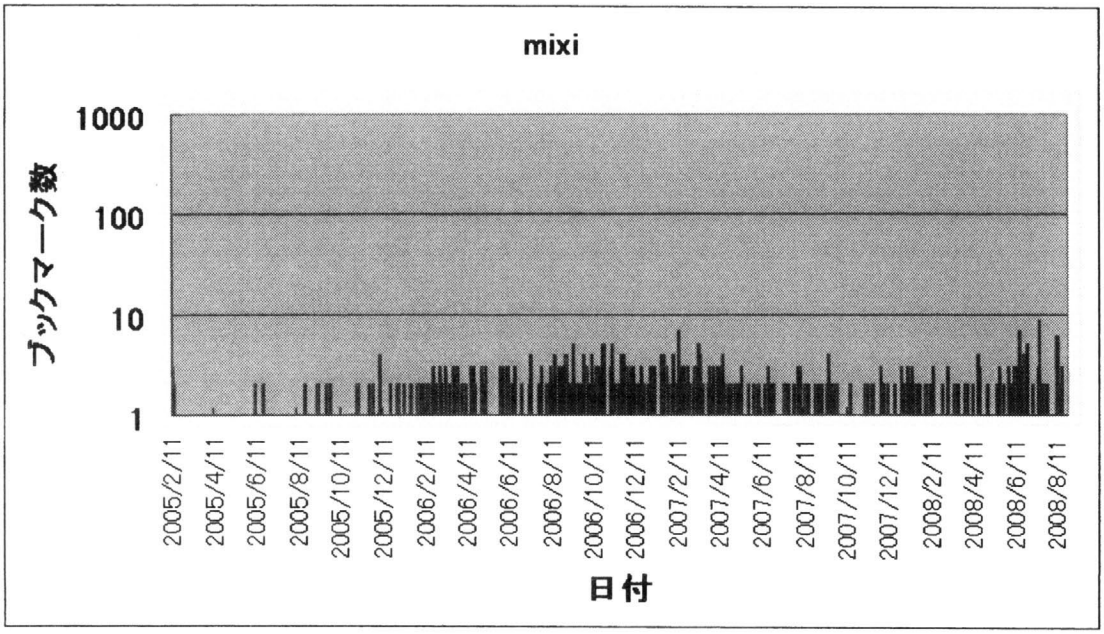

図 4.4: 大量にブックマークされる時期はないが，長い間ブックマークされ続けるタイプのページ

- 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 ページがどういった種類の 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．

- 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付の合計日数を表す．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．

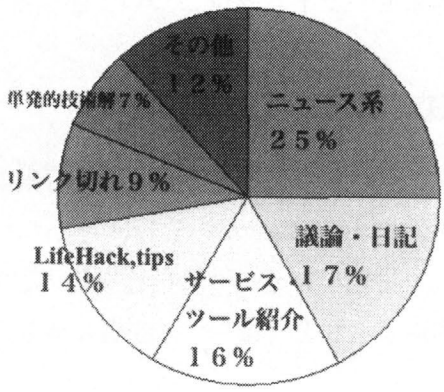

図 4.5: Type1 の Web ページの種類

図 4.5，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

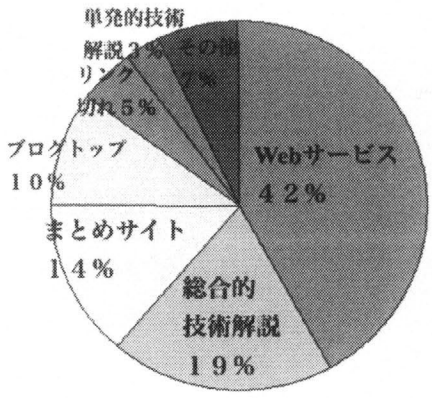

図 4.6: Type2 の Web ページの種類

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．

4.5. セレクトブクマの提案

ここでは，提案システムであるセレクトブクマについて述べる．

4.5.1. セレクトブクマの概要

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「セレクトブクマ」を提案・実装した．セレクトブクマでは，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．

4.5.2. セレクトブクマの設計思想

セレクトブクマでは，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 3 つのことを目的としている．

1. 長期的に利用できる Web ページを見つけること
2. 体系だった知識を得ること
3. 有用な Web サービスを見つけること

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，ユーザが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．

4.5.3. セレクトブクマ画面構成

セレクトブ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．

4.5.4. セレクトブクマインタフェース

セレクトブ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行なう．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．このインタフェースは，一般的な検索エンジンと同じである．

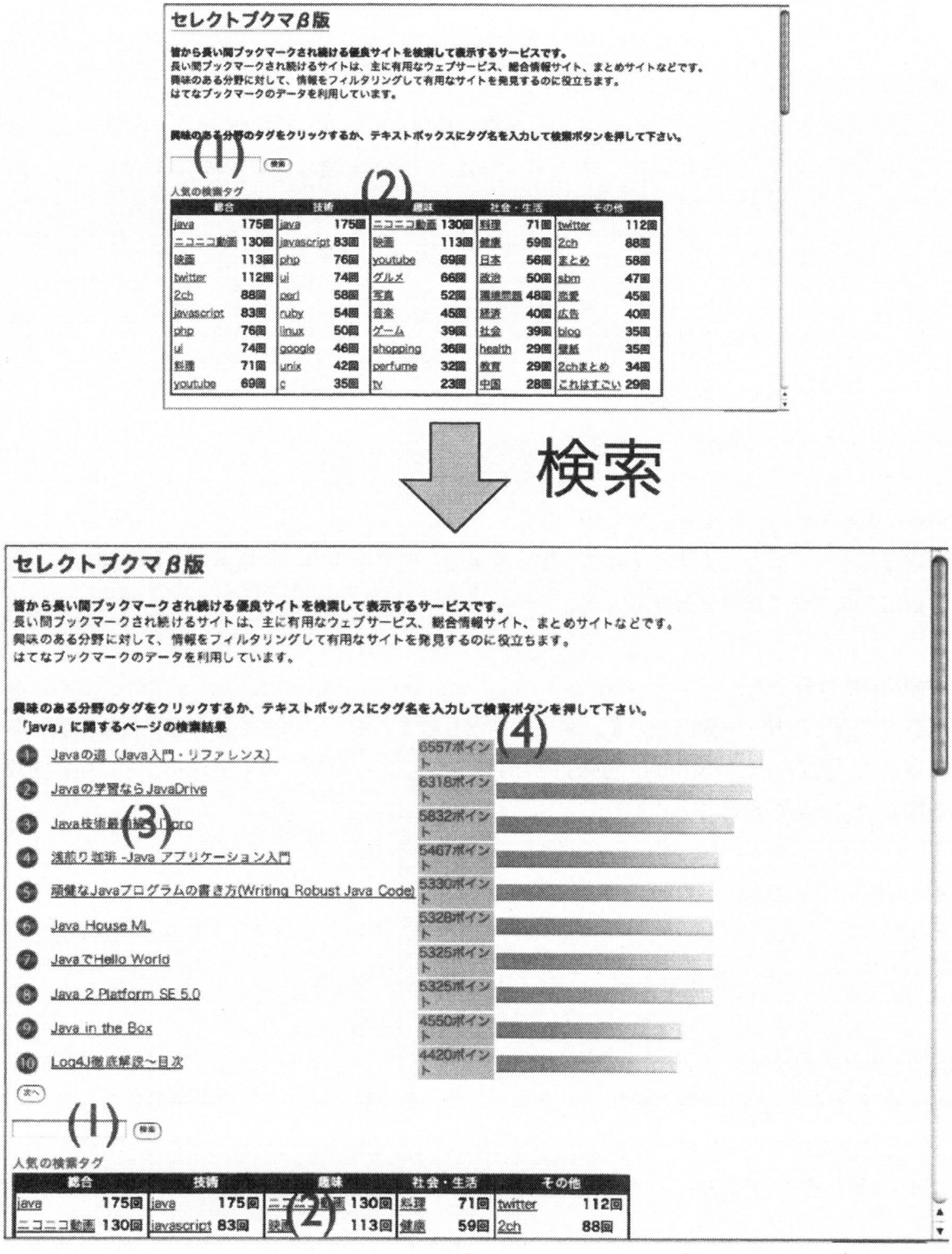

図 4.7: セレクトブクマ画面

上述に加えて，補助的な機能として，セレクトブ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．

4.5.5. 長期度と検索ランキングロジック

セレクトブクマでは，長期度 I を利用し，ブックマークされた日数を長期度とする．ブックマークされた日数が多ければ多いほど長期度が高くなる仕組みである．また，セレクトブクマの検索結果のランキングを出すにあたっては，長期度に加えて，ブックマーク数を利用した．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．

$$Ranking = Bookmarks \times Days^{\alpha} \tag{4.1}$$

Ranking：セレクトブクマのランキング
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付の合計日数
α：指数（セレクトブクマでは 1）

ここで，「指定したタグでブックマークされた日付の合計日数」とは，例えば 4 月 20 日，5 月 8 日，7 月 13 日にユーザの誰かからブックマークされたとしたら，3 とする．ここでの日付は，日本標準時（JST）での日付としている．また，セレクトブクマでは，指数 α の値は 1 としているが，α の値が大きければ大きいほど，時間情報の影響が強くなる．逆に α の値が小さいほど，そのページの人気（ブックマーク数）の影響が強くなる．

セレクトブクマでは，ブックマーク行為をページへの評価と考えているため，長期度だけでなく，ブックマーク数の指標も利用した．また，長期度 I を選んだ理由は，Web ページへのブックマークは常に行われているものではなく，ブックマークのされ方が多少不安定でも長期間ブックマークされていれば，本研究では長期的と考えることができるからである．そのため，長期度 II

のような長期的に安定している指標ではなく，長期度 I を選択し，長期度 I と長期度 III は，相関関係が強かったためよりシンプルな長期度 I を選択した．

4.5.6. システム構成

セレクトブクマは，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．

4.6. セレクトブクマの運用と主な検索結果

ここでは，セレクトブクマを運用した結果，よく検索されたキーワードとその検索結果を示す．

4.6.1. セレクトブクマの運用

セレクトブクマを Web 上に公開し，約 2 年間運用した．セレクトブクマでは，よく検索されているキーワードを以下の図 4.10 のように提示している．

図 4.10 から分かるように，特に「java」,「ニコニコ動画」,「twitter」などのキーワードがよく検索されていることが分かる．

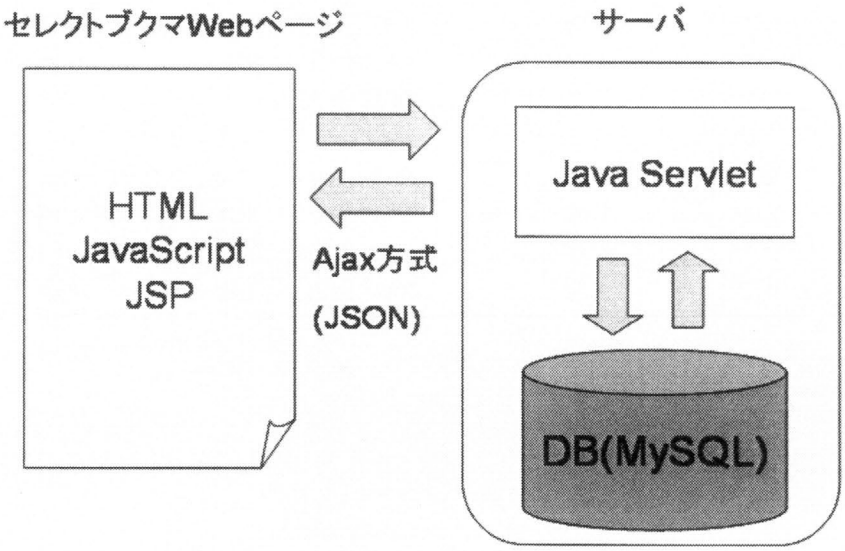

図 4.8: システム構成図

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

人気の検索タグ

総合		技術		趣味		社会・生活		その他	
java	201回	java	201回	ニコニコ動画	150回	料理	83回	twitter	134回
ニコニコ動画	150回	javascript	98回	映画	130回	健康	70回	2ch	100回
twitter	134回	php	89回	youtube	81回	日本	65回	まとめ	67回
映画	130回	ui	87回	グルメ	79回	政治	60回	sbm	57回
2ch	100回	perl	68回	写真	62回	環境問題	56回	恋愛	57回
javascript	98回	ruby	64回	音楽	53回	経済	48回	広告	48回
php	89回	linux	58回	ゲーム	47回	社会	47回	2chまとめ	43回
ui	87回	google	54回	shopping	44回	教育	39回	blog	43回
料理	83回	unix	50回	perfume	40回	health	37回	これはすごい	43回
youtube	81回	c	35回	tv	32回	中国	36回	壁紙	43回

図 4.10: セレクトブクマ人気の検索キーワード

表 4.2: 技術分野上位 3 キーワードでの各検索結果上位 10 件

順位	**java**	**javascript**	**php**
1	Java の道 (Java 入門:リファレンス)	一撃必殺 JavaScript 日本語リファレンス	レッツ PHP!
2	Java の学習なら，JavaDrive	オブジェクト指向プログラム言語としての JavaScript	symfony - open-source PHP5 web framework
3	Java 技術最前線：ITpro	IT 戦記 - JavaScript OOP におけるクラス定義方法	Maple Project
4	浅煎り珈琲-Java アプリケーション入門	檜山正幸のキマイラ飼育記 - プログラマのための JavaScript (0)	PEAR :: TOP ページ
5	頑健な Java プログラムの書き方	JavaScript によるオブジェクト指向プログラミング	PHP マニュアル
6	Java House ML	オブジェクトな JS の基礎講座	PEAR マニュアル
7	Java で Hello World	Core JavaScript 1.5 Guide - MDC	PHP 入門
8	Java 2 Platform SE 5.0	glayer.js - グレーアウト表示ライブラリ	Smarty
9	Java in the Box	高度な JavaScript 技集	初心者用 PHP 入門
10	Log4J 徹底解説〜目次	JavaScript で DOM を使う–オブジェクト指向入門の入門 - outsider reflex	symfony.jp PHP5 フレームワーク symfony の日本語情報

表 4.3: 趣味分野上位 3 キーワードでの各検索結果上位 10 件

順位	ニコニコ動画	映画	**youtube**
1	ニコニコチャート	★前田有一の超映画批評★	YouTube Flash Player
2	ニコニコ動画まとめ wiki	みんなのシネマレビュー	YouTube　ちゃんねる: YouTub ポータル
3	ニコニコ動画が言い続けた「アンチ集合知」とは何だったのか - Attribute=51	あの映画のココがわからない　まとめサイト	YouTube のムービーをダウンロードして MPEG に変換 - GIGAZINE
4	ニコニコ動画でプレイリストが使える Greasemonkey スクリプト - こたにき	allcinema ONLINE 映画データベース	[N] YouTube を便利に使うためのツール/サイト/情報のまとめ
5	時速ニコメートル-ニコニコ動画視聴率	映画生活 - 新作映画情報	ohaYouTube - おはようチューブ
6	nico in hatebu - はてブで人気のニコニコ動画が一目で分かる	超映画批評	削除された YouTube 動画を落とす「YouTube list to irv」 :にゅーあきばどっとこむ
7	ニコニコポン (α) - ニコニコ動画をちょっと便利にするツール集 [FLB ダウンロード]	CinemaScape-映画批評空間-	YouTube のストリーミング動画の保存法@ Angel Heart Club -Splash ☆ Star-
8	ニコ RSS	映画情報 - シネマカフェ cinemacafe.net	Tube ちゃんねる: 2 ch 風 YouTub ポータル
9	動画の保存も出来るニコニコ動画（γ）専用ブラウザ - nicoTool	eiga.com	OK Go - Here It Goes Again
10	エンコード設定 - ニコニコ動画まとめ wiki	いのちの食べかた	怖い YouTube

表 4.4: 社会・生活分野上位 3 キーワードでの各検索結果上位 10 件

順位	料理	健康	日本
1	男が作れる超簡単料理	health クリック	健康 生活習慣病 サプリメント	ベア速 日本が世界有数の先進国である理由
2	ボブとアンジー	基礎代謝を高めるための９９の技法	ヤバイ．日本ヤバイ．まじでヤバイよ，マジヤバイ．日本ヤバイ．:アルファルファモザイク
3	ぱくぱく名無しさんのレシピ大全	ゲンダイネット - 目の疲れをためない３大作戦	ぁやぃぃ(*ー)NEWS 2nd 日本文化は身内受けの凝り性文化
4	レシピ検索 No.1 ／料理レシピ載せるなら クックパッド	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	「韓国には白か黒しかない．灰色はないんだ」　活字中毒 R.
5	おいしいチャーハンの作り方@ぬけない２ちゃんねる	重い，痛いを吹き飛ばせ！肩コリ解消法 — Web 担当者 Forum	japan-guide.com
6	レシピブログ	体の歪みを治したい:アルファルファモザイク	さらば！「豊かな国日本」 - ビジネススタイル - nikkei BPnet
7	鳥はむの館	病院検索ならここカラダ	アンカテ (Uncategorizable Blog) - 利権が無い文化を育てれば日本は生き残れる
8	【味の素 KK】レシピ大百科	意外に効き目のある視力回復マッサージ	日本の暮らし二十四節気めぐり 言の葉草 — 伊勢志摩の旅 よいとこせ
9	ためしてガッテンに学ぶ定番料理17品の美味しい作り方 - Liner Note	蒸しタオルを使うと酷使した目の疲れが取れる - GIGAZINE	世界で一番美味いメシ - pal-9999 の日記
10	とろけるように柔らかい「豚の角煮」を手抜きで作る - GIGAZINE	MouRa	Net 現代	巷にはびこる「健康情報」50 のウソ・ホント	日本再発見マップ

表 4.5: その他分野上位 3 キーワードでの各検索結果上位 10 件

順位	**twitter**	**2ch**	**まとめ**
1	Twitter Friends Analyzer	必死チェッカーもどき	まとめ.com
2	twitterfeed.com : feed your blog to twitter - post RSS to twitter automatically	2 典 Plus　2 ちゃんねる用語サイト	2ch まとめ
3	[観] Twitter API 仕様書 (勝手に日本語訳シリーズ)	2chnavi - 2 ちゃんねるスレ紹介／レス紹介 blog まとめ	ストーリーを教えてもらうスレ暫定 Wiki
4	buzztter	2ch DAT 落ちスレ ミラー変換機 ver. 4	80 個以上のはてブ関連の Greasemonkey まとめ - 聴く耳を持たない (片方しか)
5	TwitterIrcGateway - misuzilla.org	2 ちゃんねる発の逮捕者	お金を出す前に見るべきクチコミ情報サイトまとめ
6	TweetStats :: Graphin' Your Stats	2ch まとめ	なにかと使える『名言ポータルサイト』まとめ — P O P * P O P
7	Twitter をはじめよう！ - GreenSpace	2 ちゃんねる検索	なつみかん@はてな - 便利な文房具&文房具の使い方まとめ
8	日本語 Twitter 検索 : Search	2 ちゃんねる情報の RSS 一覧	FrontPage - FoN まとめ wiki
9	Twitter まとめ Wiki	uraNews4VIP * 俺の先祖は恐ろしい人物かも知れない …	Twitter まとめ Wiki
10	P3:PeraPeraPrv - とかいろいろ	www.tv2ch.info	マルチディスプレイ総合スレッドまとめサイト

4.6.2. 主な検索結果

セレクトブクマを運用した結果，よく検索されたキーワードとその検索結果上位 10 件を以下の表 4.2〜表 4.5 に示す．

この検索結果を見ると，それぞれの分野で，まとまった情報が載っている Web ページが多いことが分かる．このため，セレクトブクマを利用したユーザが，本研究の目的の 1 つである体系だった知識を得られる Web ページを取得できていることが分かる．

4.6.3. セレクトブクマの利用経験

筆者は，約 1 年〜2 年程度セレクトブクマを利用して，調べたい分野の体系だった Web ページ，まとめページ，有用な Web サービスなどの検索を行ってきた．筆者が利用した経験では，データ量が約 76 万 URL とあまり多くないため，入力するキーワードによっては，なかなか良い結果を得られないが，よく利用されるようなキーワードを入力すると所望のタイプの Web ページを得られる場合が多かった．例えば，筆者はセレクトブクマで「映画」と入力してみんなのシネマレビュー [121] という映画レビューサイトを見つけ，現在でもこのサイトをよく利用している．また，セレクトブクマの利用者から「料理」で検索したときにボブとアンジー [110] というレシピ共有サービスが見つかって良かったという意見をもらった．その人によると，ボブとアンジーは，有名なレシピ共有サービスである COOCKPAD[112] よりも使いやすいとのことであった．

4.7. 評価実験

セレクトブクマの有用性を評価するため，セレクトブクマに対して評価実験を行なった．

4.7.1. 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．

- 長期的に利用できる Web ページを見つけること
- 体系だった知識を得ること
- 有用な Web サービスを見つけること

そのため，以下のようなことを目的とした実験を行なった．

- セレクトブ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

4.7.2. 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 ページの中で重複したものを除いた Web ページを，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 ページを 1 位〜10 位まで選んでもらう．

(A) セレクトブクマ（α=1）

(B) ブックマーク数順

(C) Google 検索

ここで言うブックマーク数順とは，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順位付けしたものとする．上述の 3 種類の手法を利用して得られた Web ページの中から，被験者が選んだ Web ページを適合文書として，適合率，再現率を求める．

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行った．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．

4.7.3. 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

(A) セレクトブクマ，(B) ブックマーク数順，(C)Google 検索での検索結果上位 10 件のタイトルを以下の表 4.6〜4.8 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．

表 4.6: 「java」で検索した場合の検索結果上位 10 件

順位	セレクトブクマ	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)

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 ページが上位にランキングされやすい傾向にあるのに対して，逆にセレクトブクマでは，個人で作成したような Web ページ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

表 4.7: 「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 9 9 の技法	体の歪みを治したい:アルファルファモザイク	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 9 9 の技法	ケンコーコム - 健康メガショップ

表 4.8: 「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? - ベ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	映画情報 - シネマカフェ cinemacafe.net	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」

クマでは，ある程度ランキングされる Web ページが近い傾向にあることがわかる．これは，セレクトブクマのロジックが，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

(A) セレクトブクマ，(B) ブックマーク数順，(C)Google 検索の検索結果から，被験者が今後も利用したいと思う Web ページを 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 4.9〜表 4.14 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 4.9，表 4.10,「健康」に関しては，表 4.11，表 4.12,「映画」に関しては，表 4.13，表 4.14，に示す．

表 4.9: 「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 4.10: 「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 Evolution 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 4.11: 「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９９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） —ｉｄｅａ*ｉｄｅａ	**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１４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 — 睡眠不足だと仕事がはかどらない理由，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 4.12: 「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） —ｉｄｅａ*ｉｄｅａ	**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	指をポキポ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 Silicon 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 4.13: 「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 4.14: 「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:観るまでバカにして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 ー 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・再現率を比較した．適合率・再現率の意味に関しては，第 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 種類とした．

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 4.15 に示す．

表 4.15: 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 4.11，図 4.13，図 4.15 に，得点から全適合文書を作成した場合の適合率を図 4.12，図 4.14，図 4.16 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 4.16〜表 4.18 に示す．適合率の値は，セレクトブクマ，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．

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

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．

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

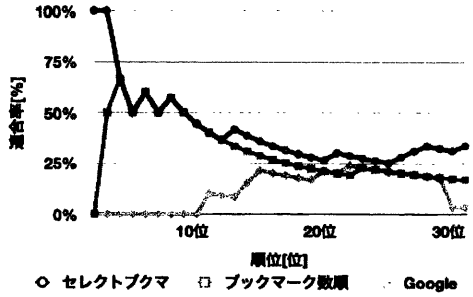

図 4.11: 3 人以上を適合文書とした場合の適合率 (java)

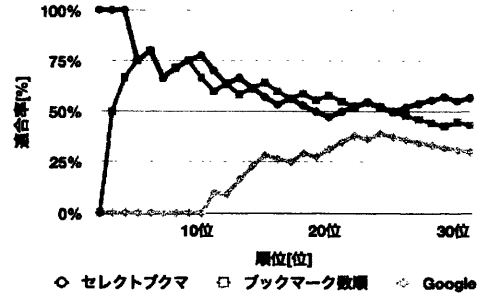

図 4.12: 8 点以上を適合文書とした場合の適合率 (java)

表 4.16: 再現率 (java)

	セレクトブクマ	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．

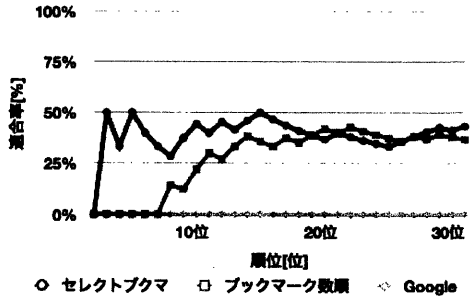

図 4.13: 3 人以上を適合文書とした場合の適合率 (健康)

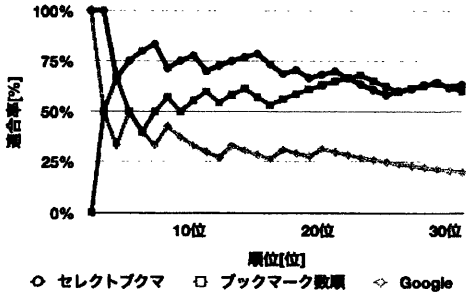

図 4.14: 8 点以上を適合文書とした場合の適合率 (健康)

表 4.17: 再現率 (健康)

	セレクトブクマ	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.

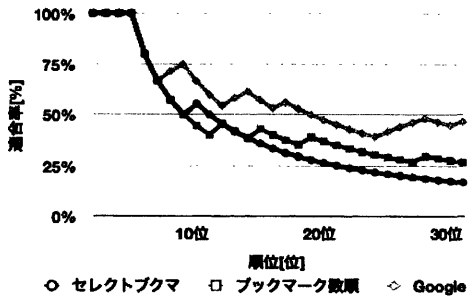

図 4.15: 3 人以上を適合文書とした場合の適合率 (映画)

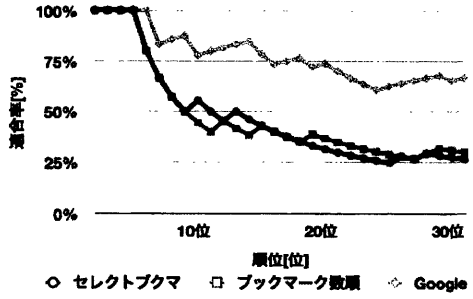

図 4.16: 8 点以上を適合文書とした場合の適合率 (映画)

表 4.18: 再現率 (映画)

	セレクトブクマ	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 4.15 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 4.19〜表 4.21 に示す．

表 4.19: 「java」での検索結果の得点の合計値

	セレクトブクマ	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 4.20: 「健康」での検索結果の得点の合計値

	セレクトブクマ	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

表 4.21: 「映画」での検索結果の得点の合計値

	セレクトブクマ	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

結果を見ると「java」と「健康」においては，セレクトブクマで上位 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．

4.7.4. 時間情報を利用する効果について

今回被験者実験を行なった検索単語は，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，現状セレクトブクマのランキングロジックでは，ブックマーク数順は，長期度を表すブックマーク日数にかかる指数 α の値が 0 のときと同義である．今回の実験は，α=1 として実験を行なったが，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．

4.7.5. 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・収集するのには，向かないと考えた．

これに対して，「java」で検索した場合，セレクトブ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 ページ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 ページや 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 ページ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 ページの中から，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．

4.8. 課題

現時点のセレクトブクマでは，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．

はてなブックマークからのデータの収集の手法にも課題がある．本研究では，2008 年 9 月時点でブックマーク数 5 以上のすべてのデータを収集したが，データが大量であったためクローリングに非常に時間がかかった．現状では，常にはてなブックマークから最新のブックマーク情報を収集し続けることは難しいが，はてなブックマークサービスに負荷をかけないように，常に少しずつデータを収集する仕組みを開発する必要もある．また，はてなブックマークのデータに依存しているため，API の変更やサービスの終了などによって，新しくデータが収集できなくなる可能性もある．

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排除できているは言えない．セレクトブクマの検索結果や被験者の選んだ Web ページを見ると分かるが，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，セレクトブクマは，Web 検索システムというより，どちらかというと，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クマの Web ページが一つあり，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．

4.9. 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「セレクトブクマ」を提案した.「セレクトブクマ」の評価実験として,「セレクトブクマ」,「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 ページ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 ページを排除し，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ことがわかった.

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

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は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．検索エンジンでの検索は，一般的に検索キーワードに大きく依存し，適切な検索キーワードを知らないと適切な検索結果を得ることは難しい．実際に多くの人は，良い検索キーワードが思い浮かばないといった場面に遭遇することがある．良い検索キーワードが思い浮かばない理由としては，大きく分けて次の 2 つの理由が考えられる．

1 つ目の理由は，調べたい内容に関するキーワードを知らないという理由である．慣れない分野や新しく学ぼうと思った分野に関しては，重要なキーワードや定番なキーワードを知らないことも多い．そういったキーワードを知らないと的はずれなキーワードで検索してしまったり，そもそもどうやって検索したらいいか分からなかったりする場合がある．

2 つ目の理由は，検索リテラシーが不足しているという理由である．検索リテラシーが不足していると，適切な検索キーワードが入力できないことが多い．検索キーワードとして文章を打ってしまうユーザや，ある言葉の意味を調べたい場合に「○○とは」や「○○　意味」のようなキーワードを打つことができないユーザもいる．

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 [37] 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 [92] などが行われてきた．また，Google サジェストなどの検索キーワードを推薦する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．

こういったアプローチも存在するが，新しい分野や慣れない分野を手軽に学びたいといったような状況については，その分野の定番キーワードを素早く無駄なく取得できると，学び始めの速度が向上する．こういった背景に対して，本研究では，長期的に検索され続けている検索キーワードは，定番のキーワードでその分野のことを知るための鍵となるようなキーワードが多いのではないかと考えた．

5.1.2. 目的

本研究では，何か新しい分野や慣れない分野について，調べたり学ぼうとしたりするときに，その分野を調べるにあたって，鍵となるような重要なキーワード一覧を手軽に取得できるように

することを目的とする.

例えば, 新しくユーザインタフェース分野について学んだり広く調査したりしようと思ったときに, どうやって検索したらいいのかというのは難しい問題である. 例えば,「実世界指向インタフェース」や「ユーザエクスペリエンス」などのキーワードは, ユーザインタフェース分野を学ぶ上で重要であるが,「ユーザインタフェース」というキーワードからこれらのキーワードに簡単にたどりつけるとは限らない. そのため, Web 上での検索だけでは, こういった概念の存在にすらなかなか気付けない可能性もある.

本研究では, こういった問題を解決するために, 手軽に素早く慣れない分野の長期的な定番キーワードを取得できることを目的とする.

5.2. 分析

長期的な検索キーワードを取得するために, 検索キーワードの時系列データの分析を行った.

5.2.1. 利用データ

検索キーワードの時系列データは, Google Insights for Search[117] というサービスのデータを利用した. 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 週間ごとの検索ボリュームを提示している. ここで検索ボリュームとは, そのキーワードでの 1 週間ごとの検索回数の最大値を 100 としたときの 100 に対する割合の値である.

5.2.2. 時間情報に関する分析

検索キーワードの検索ボリュームの時系列データについて, どのような分布になっているか分析を行った. その結果, 非常に大まかに分けると次の 3 種類のタイプの分布があることが分かった.

(1) 短期的に急激に検索されるキーワード

(2) 長期的に安定して検索されるキーワード

(3) 周期的に検索されるキーワード

(1) の特徴が強いキーワードとして, 急激に流行したキーワードがあげられる. 例として「ハンカチ王子」というキーワードの Google 検索の時系列データのグラフを図 5.1 に示す. グラフから, 一時期に急激に検索されて, その後はあまり検索されていないことが分かる. こういったキーワードは, 流行り廃りが激しく, 短期的なキーワードであることが分かる.

また，この時系列データを 1 週間ごとの検索ボリュームが多い順に並び替え，両対数グラフにプロットしたものを図 5.2 に示す．このグラフから，一時的に急激に流行してバーストしているような検索キーワードの時系列データは，べき分布に近い分布になることが分かる．

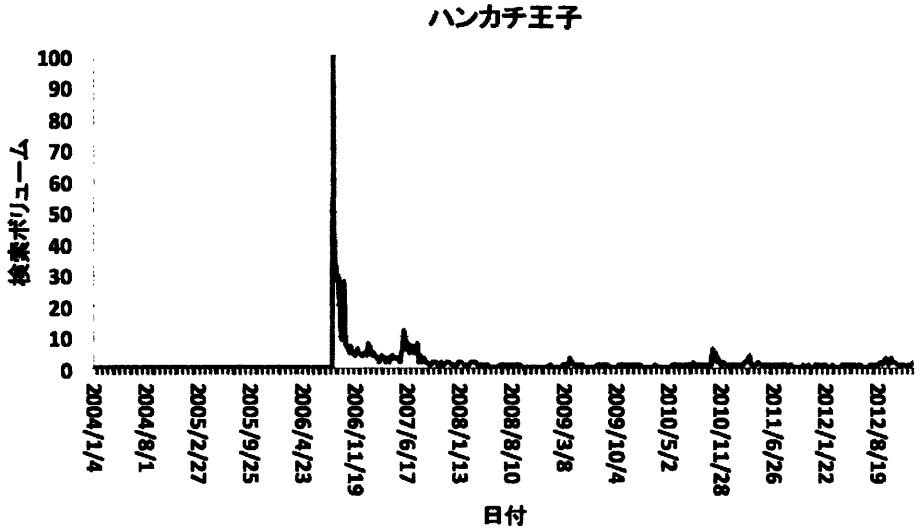

図 5.1: キーワード「ハンカチ王子」の検索時系列データ

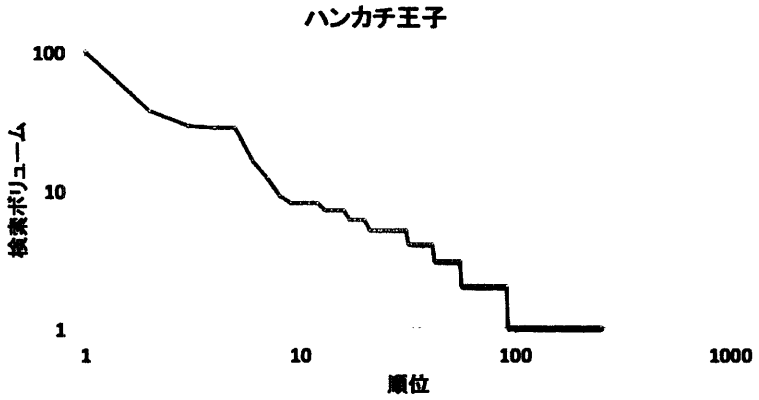

図 5.2: キーワード「ハンカチ王子」の検索時系列データ

(2) の特徴が強いキーワードとして，長期的に安定して必要とされているキーワードがあげられる．例として「動画」というキーワードの Google 検索の時系列データのグラフを図 5.3 に示す．グラフから，長期的に安定して検索されていることが分かる．こういったキーワードは，長期的に安定して検索されているので，長期的な検索キーワードであると言える．

また，この時系列データを検索ボリュームの多い順に並び替え，両対数グラフにプロットしたものを図 5.4 に示す．このグラフから長期的なデータはべき分布と比較して大きい値をとる分布となり，検索ボリュームの大きい順に並べたときの減少率が非常に小さいことが分かる．

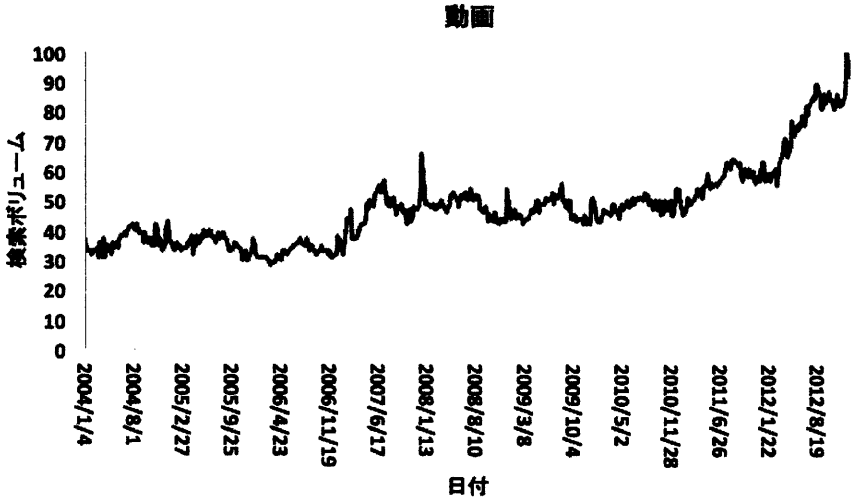

図 5.3: キーワード「動画」の検索時系列データ

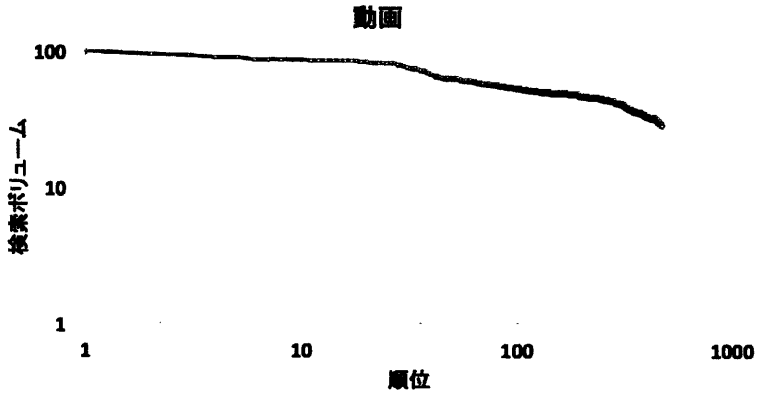

図 5.4: キーワード「動画」の検索時系列データ

(3) の特徴が強いキーワードとして，ある季節などの時期によって，必要となるキーワードがあげられる．例として「卒業論文」というキーワードの Google 検索の時系列データのグラフを図 5.5 に示す．グラフから，周期的に山ができており，1 年に 1 度の周期で多く検索される時期があることが分かる．

また，この時系列データを検索ボリュームの多い順に並び替え，両対数グラフにプロットしたものを図 5.6 に示す．このグラフから周期的なデータはべき分布と比較して，大きい値をとる分

布となるが，長期的に安定している「動画」の時系列データと比較して順位の値が大きくなるにつれて検索ボリュームの値の落ち方が大きくなることが分かる．

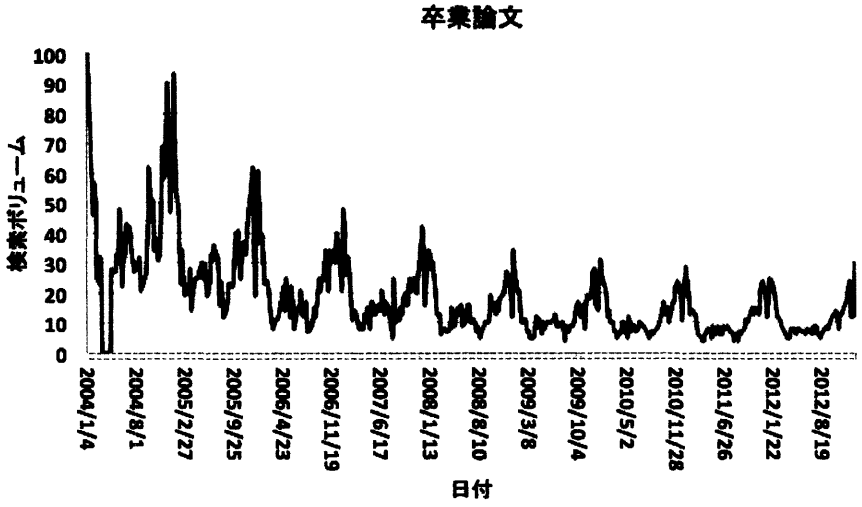

図 5.5: キーワード「卒業論文」の検索時系列データ

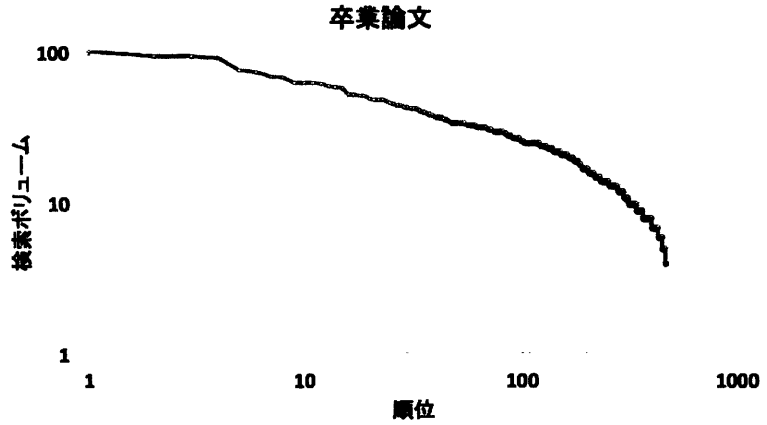

図 5.6: キーワード「卒業論文」の検索時系列データ

以上のことから，安定して検索され続けているようなキーワードほど，両対数グラフにプロットしたとき，べき分布と比較して大きな値をとることが分かる．逆に一時的に急激に検索されたキーワードは，べき分布に近い分布になることが分かる．このことからべき分布と比較して，大きな値となるキーワードほど長期的なキーワードではないかと考えた．

5.3. MasteryEye の提案

ここでは，長期的な検索キーワードを取得するシステム，MasteryEye の提案を行う．

5.3.1. システム概要

MasteryEye では，以下の手順で処理を行う．

(1) ユーザがキーワードを入力する

(2) 入力したキーワードに関連するキーワード一覧を取得する

(3) 関連キーワード一覧の長期度を計算する

(4) 長期度の高いキーワードを提示する

まず，MasteryEye の画面上でユーザが調べたい分野の中自分が知っている何か有名なキーワードを入力する．そうすることによって，そのキーワードに関連するキーワード一覧を取得し，その中で長期的に検索され続けている定番キーワードを提示する．

5.3.2. 関連キーワードの取得

関連キーワード取得は，既存の手法である，Lingua-JA-Expand[120] 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

この手法でさまざま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問題があった．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．

そこで本研究では，Wikipedia 日本語版のすべての単語を入力キーワードとして，Lingua-JA-Expand から関連キーワードを取得し関連キーワードデータベースを作成した．そして，検索者が入力したキーワードに対して，このデータベース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する手法を適用した．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

5.3.3. 長期度の計算

長期度の計算は，Google Insights for Search から取得した，Google における各検索キーワードの検索ボリュームの時系列データを利用して計算する．MasteryEye では，第 3 章で述べた長期度計算手法の長期度 II を適用した．長期度 II は，単位時間ごとの利用回数が，べき分布のデータよりどの程度大きいかで長期度を求める手法である．

最初に，Google 検索キーワードの時系列データを 1 週間ごとの検索ボリュームの多い順に並べる．次に，並べ替えたデータに基づき，べき分布データを生成する．そして，並べ替えた実データから生成したべき分布データを引いたものを合計し，その値の大きさを長期度とする．ここで，実際に長期度 II の数式にあてはめて示すと以下のようになる．

$$LongDegreeII = \sum_{x=1}^{n} (Times_x - ax^k) \tag{5.1}$$

(5.2)

$LongDegreeII$: 長期度 II
n : データの合計数
x : 1 週間ごとの検索ボリュームの順位
$Times_x$: 1 週間ごとの検索ボリューム
ax^k : べき分布データ
a : べき分布の定数
k : べき分布のスケーリング指数（定数）

ここで ax^k は，両対数グラフにおいて y 軸の最大値と x 軸の最大値を結ぶ直線を描くようなべき分布データとし，a と k の値を設定する．このとき，y 軸の最大値の点は順位のため x=1 となり，x 軸の最大値の点は，y=1 とする．そして，実際の単位時間ごとの利用数データからこのようなべき分布モデルの値を引いた値の合計値を長期度とする．

MasteryEye において，なぜ長期度 II を採用したのかの理由について述べる．MasteryEye では，長期的に安定している定番の検索キーワードを取得することを目的としている．そのため，長期的に利用されているということと安定して利用されているという 2 点が重要になってくる．長期

的に利用されているという観点だけならば，長期度 I でも良いが，長期度 I では単純に利用された期間で長期度を計算しているため，安定性は測れない．また，Google での検索キーワードの時系列データの場合，5.2 の例で示したように，長期的に安定して利用されているキーワードは，べき分布から外れる傾向にある．べき分布から大きく外れてしまうと，長期度 III は適用しにくい．これに対して，長期度 II では，長期的に利用されているという指標と安定して利用されているという指標，双方を含む．さらに，長期度 II はデータがある程度べき分布から外れていることも想定しているため，長期度 II を採用した．

5.3.4. 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 5.7 は，例として「java」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．

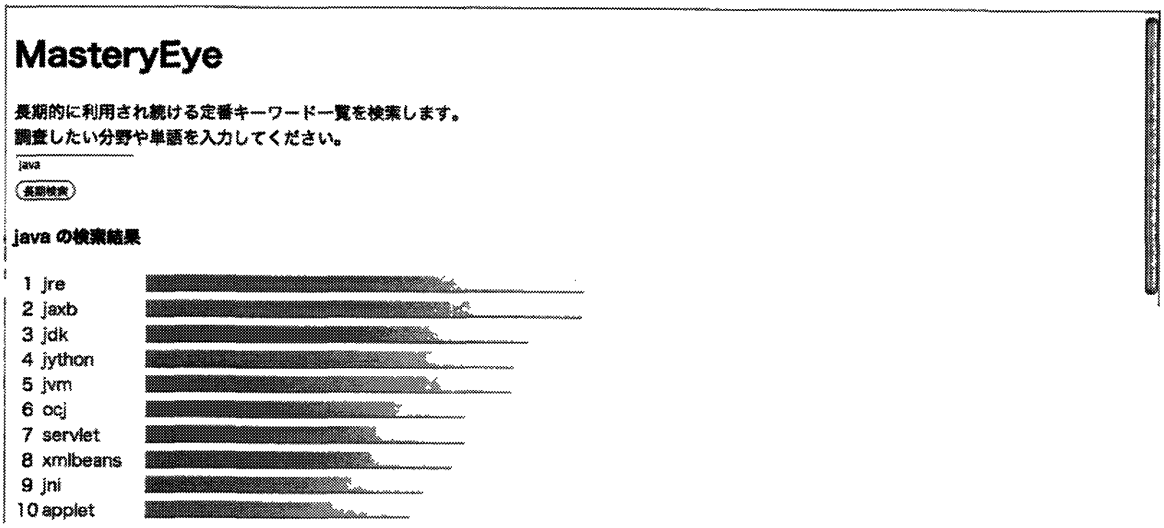

図 5.7: 「java」での検索結果画面

5.4. 評価実験

5.4.1. 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. 提案手法の有用性

5.4.2. 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（MasteryEye）
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[122]）

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	20 代：15 名，30 代：2 名，40 代：2 名，50 代：1 名

5.4.3. 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 5.11 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．reflexa 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によって評価を行った．まず，質問 1 の知っているかどうかについて提案手法で長期度が高かったもの（手法 1）と提案手法で長期度が低かったもの（手法 2），reflexa（手法 3）を比較した (表 5.4，表 5.5).

表 5.4: 知っているかどうかの比較（手法 1 と手法 2）

観測度数	選択あり	選択なし	計
手法 1	35	4	39
手法 2	24	16	40
計	59	20	79
期待度数	選択あり	選択なし	計
手法 1	29.1	9.9	39
手法 2	29.8	10.1	40
計	59	20	79
p 値	2.37×10^{-3}		

表 5.5: 知っているかどうかの比較（手法 1 と手法 3）

観測度数	選択あり	選択なし	計
手法 1	35	4	39
手法 3	24	15	39
計	59	19	78
期待度数	選択あり	選択なし	計
手法 1	29.5	9.5	39
手法 3	29.5	9.5	39
計	59	19	78
p 値	2.71×10^{-3}		

表 5.4 の結果より，選択されたキーワードは手法 1 の方が多かった．また，p 値は 2.37×10^{-3} 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，長期度が低いキーワードより長期度が高いキーワードの方が，被検者から知られている一般的なキーワードが多いこと

が分かった．表 5.5 の結果より，選択されたキーワードは手法 1 の方が多かった．また，p 値は 2.71×10^{-3} 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，reflexa より MasteryEye の方が被検者から知られている一般的なキーワードが多いことが分かった．

次に，質問 2 の長期的かどうかについて提案手法で長期度が高かったもの（手法 1）と提案手法で長期度が低かったもの（手法 2），reflexa（手法 3）を比較した (表 5.6，表 5.7)．

表 5.6: 長期間利用かどうかの比較（手法 1 と手法 2）

観測度数	選択あり	選択なし	計
手法 1	32	7	39
手法 2	14	26	40
計	46	33	79
期待度数	選択あり	選択なし	計
手法 1	22.7	16.3	39
手法 2	23.3	16.7	40
計	46	33	79
p 値	2.24×10^{-5}		

表 5.7: 長期間利用かどうかの比較（手法 1 と手法 3）

観測度数	選択あり	選択なし	計
手法 1	32	7	39
手法 3	21	18	39
計	53	35	78
期待度数	選択あり	選択なし	計
手法 1	26.5	12.5	39
手法 3	26.5	12.5	39
計	53	25	78
p 値	7.61×10^{-3}		

表 5.6 の結果より，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 2.24×10^{-5} 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，長期度が高いキーワードの方が，長期度が低いキーワードよりも長期間利用されていることが分かった．また，表 5.7 の結果より，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 7.61×10^{-3} 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，MasteryEye で取得したキーワードの方が，reflexa よりも長期間利用されていることが分かった．

次に，質問 3 の重要かどうかについて提案手法で長期度が高かったもの（手法 1）と提案手法で長期度が低かったもの（手法 2），reflexa（手法 3）を比較した (表 5.8，表 5.9)．

表 5.8: 重要かどうかの比較（手法 1 と手法 2）

観測度数	選択あり	選択なし	計
手法 1	29	10	39
手法 2	12	28	40
計	41	38	79
期待度数	選択あり	選択なし	計
手法 1	20.2	18.8	39
手法 2	20.8	19.2	40
計	41	38	79
p 値	7.97×10^{-5}		

表 5.9: 重要かどうかの比較（手法 1 と手法 3）

観測度数	選択あり	選択なし	計
手法 1	29	10	39
手法 3	18	21	39
計	47	31	78
期待度数	選択あり	選択なし	計
手法 1	23.5	15.5	39
手法 3	23.5	15.5	39
計	47	31	78
p 値	1.09×10^{-2}		

表 5.8 の結果より，選択されたキーワードは手法 1 の方が多かった．また，p 値は 7.97×10^{-5} 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，長期度が低いキーワードより長期度が高いキーワードの方が，被検者が重要だと思うキーワードが多いことが分かった．表 5.9 の結果より，選択されたキーワードは手法 1 の方が多かった．また，p 値は 1.09×10^{-2} となり，有意水準 5 %以下で有意差を求めることができた．このことから，reflexa より MasteryEye の方が，被検者が重要だと思うキーワードが多いことが分かった．

次に，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 5.10)．

表 5.10: 長期間利用と重要なものの関係

	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分かった．

5.5. MasteryEye の利用経験

筆者は，MasteryEye を利用して，学びたい分野の定番キーワードを取得してきた．筆者は，プレゼンテーションなどを作るときの素材を見つけるのが苦手で，素材を見つけるためにどうやって検索すればいいか困っていた．そこで，MasteryEye で「素材」と入力したら,「クリップアート」というキーワードを知ることができた.「クリップアート」は素材を探すためには定番のキーワードであると考えられるが，偶然筆者が知らなかったキーワードであったため，素材に関して知識の少なかった筆者には非常に有益なキーワードであった．また，筆者が新しく学ぼうと思ったプログラミング言語である「python」というキーワードを入力することにより,「numpy」という python の数学関数ライブラリ名を知ることができた．

表 5.11: 取得した関連キーワード一覧

元キーワード	長期度の高いもの	長期度の低いもの	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・バーツ	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
	動画プレーヤー	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

5.6. 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，さまざまな種類の情報が考えられる．本研究では，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．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るべきである．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．

また関連キーワードのデータと検索ボリュームの時系列データの収集に関しても課題がある．関連キーワードデータベースのデータ量は，約120万キーワードの関連キーワード約2500万キーワード収集したが，このデータ量だとマイナーなキーワードを入力した場合，関連キーワードがほとんど取得できない場合もある．検索ボリュームの時系列データを取得していたGoogle Insights for Searchというサービスが，2013年1月13日の時点では，Googleトレンドというサービスに移行している．このようにデータ取得元のサービスに変更があれば，システムが正常に動作しなくなること可能性もある．そのため，安定してサービスを稼働させるためには，独自で検索回数などの時系列データを取得できるようにする必要がある．

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．

5.7. 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，長期的に安定して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　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．

6.1. 情報検索

6.1.1. 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する必要がある．

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 [45] 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:Latent Semantic Indexing）[15][8] 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 [58] 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 [31] が幅広く応用されている．

機械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）[4][14] 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．

さまざま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 [5][59][97] などで詳しく解説されている．

6.1.2. 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 [12]．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 [81]．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 [64].

6.1.3. ソーシャルな検索

近年では，ソーシャルな情報や人力情報などを利用した検索手法が注目を浴びている．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 [6]．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 [87]．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 [52].

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 [50]．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 [54][55]．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 [104]．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 [49]．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る．

6.1.4. 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したりすることによって，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，入力のたび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 [46][7][36]，インクリメンタル検索の有効性も主張されている [3]．インクリメンタル検索に関する研究として，インクリメンタル検索を日本語に応用する研究 [99] や，音声インクリメンタル検索システムの提案 [90] などもされている．

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，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 [32]．この研究では，例えば，検索キーワードを共有できる手法などを提案している．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 [84]．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 [103]．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 [106]．

6.1.5. さまざまな情報を利用した検索システム

上述の他にもさまざまな情報を利用した検索手法が提案されている．ここでは，さまざま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 [43] 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 [28]．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 [116]

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 [82].

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 [77].

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 [88].

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 [63].

荘司らは，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 [98].

6.2. 情報フィルタリング

6.2.1. 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 [41].

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 [85]，[73] 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 [39]，[10]，[66] 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 [108].

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 [72] に関する研究もさかんになってきている [27].

6.2.2. 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行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

ページの推薦を行なう手法を提案している [34]．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 [95]．

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 [96]．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α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 [65]．

6.2.3. 情報フィルタリングのインタフェース

6.2.4. さまざま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，さまざま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．

6.3. 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 [18]．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 [22]．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 [75]．

6.4. 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 [42]．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 [30]．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 [56]．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 [11]．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 [93]．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 [91]．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 [70]．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 [79]．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 [80]．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 [105]．

6.5. 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 [26]．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 [44]．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 [23]．

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 [19]．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 [20]．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 [83]．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 [89]．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 [76]．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 [67]．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 [16]．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．Strange は，UNIX システムにおける長期的なアクセスパターンを分析することによって，ファイルの移動や削除を自動的に行うシステムの提案をしている [48]．また，Gibson らも UNIX システムにおいて長期的に利用されるファイルを分析することによって，2 度と利用されないファイルを発見する手法を提案している [17]．この研究では，分析により 90 %のファイルは 1 度作成された後には 2 度と利用されないファイルであり，1 %のファイルのみが毎日利用されるファイルであることを示している．

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 [53]．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 [74]．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 [107]．

また，情報のライフサイクルに関する研究も行われている．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 [100]．

第7章　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．

7.1. 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，セレクトブクマ，MasteryEye という 2 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

7.1.1. 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．

(1) ソーシャルブックマークの時間データの分析

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，長い期間新たにブックマークされ続けてきた Web ページは，Web サービスが 42 %，総合的技術解説サイトが 19 %，まとめサイトが 14 % と，長期的に利用されるようなタイプの Web ページの割合が非常に多いことが分かった．また，短い期間しか新たにブックマークされなかった Web ページは，ニュース系が 25 %，議論や日記が 17 %，サービスやツール紹介が 16 % と一時的に利用されるタイプの Web ページの割合が非常に多いことが分かった．Web ページのタイプを完全に分類することは難しいが，ブックマーク期間の長短で，非常にはっきりとしたタイプの違いが現れたため，ブックマーク期間を利用した Web ページの分類手法や検索手法が有効であると考えた．そして，ソーシャルブックマークによる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と考えた．

(2) 検索ボリュームの時間データの分析

Google 検索での検索ボリュームの時系列データを分析した．1 週間毎の検索ボリュームの時系列データを見てみると，ほとんどのキーワードについて毎週ある程度以上は検索されていたため，単純に検索された期間の長さで長期的かどうかを判断することが難しいことが分かった．

べき分布になるものが多い可能性があると考え，両対数グラフに検索ボリュームの多い順にプロットして分析を行った．その結果，べき分布となるキーワードも多かったが，べき分布と比較して平均値に近い値が多いキーワードも多かった．こういった平均に近い値が多くなるような分布は，安定して検索されている期間が長いため，長期的に安定して検索されているキーワードとなる．これと比較して，べき分布に近いような分布となるキーワードは，急激に検索ボリュームが多くなる期間があり，かつ，ピークと比較して検索ボリュームが非常に少ない期間が長いということが分かる．そのため，長期的に利用されている検索キーワードは，長期的に安定して検索されているキーワードであると考えた．

以上のことから，検索ボリューム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，長期的に利用されている情報を見つけることができると考えた．

7.1.2. 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．長期度の計算手法として，長期度 I，長期度 II，長期度 III という 3 つの手法を提案した．それぞれの手法をランダムに取得したソーシャルブックマークのブックマークデータに適用して分析を行った．

(1) データがべき分布かどうか

長期度 II の値が大きかった上位 2 ページと値が小さかった下位 2 ページの日付ごとのブックマーク数が多い順に並べたデータを両対数グラフにプロットした．結果のグラフを見てみると，長期度 II はべき分布との差分の大きさであるにも関わらず，上位 2 ページ，下位 2 ページ共にべき分布から大きく外れていないことが分かった．このことからある程度以上ブックマーク数が多いページは，1 日ごとのブックマーク数がべき分布に近づくものと考えた．そのため，長期度 III も上手く適用できているものと考え，長期度 III の値が 1 位のもの 50 位のもの 100 位のものを両対数グラフにプロットし，同時に最小二乗法によってべき分布に近似したデータもグラフにプロットした．その結果，それぞれ実データとべき分布に近似したデータに大きな相違はなく，また傾きの違いも顕著に出ていることが分かった．このことから，長期度 III はブックマーク数がある程度以上多いページには，適用できるものと考えた．

(2) 各長期度での順位付け

取得した 100 ページを長期度 I, II, III それぞれに対して長期度が高い順に並べた．そして，上位 10 件のページと下位 10 件のページにどんなページが含まれているか分析を行った．

長期度 I については，上位 10 件には長期的なタイプのページが多く，下位 10 件には短期的なタイプのページが多かった．この理由は，いつ見ても有用なページは，ブックマークされる日数が長くなり，ニュースや日記など古くなるとあまり有用ではなくなる可能性の高いページは，ブックマークされる日数が短くなるからであると考えた．そのため，単純にブックマーク数が多いページが有用であるということは言えないと考えた．こういったことからも，長期度 I によってある程度長期的なページを測定できていることが分かった．

長期度 II については，上位 10 件には，長期的なタイプのページが多く，下位 10 件には長期的なタイプのページとそうでないページが混じっていた．さらに，長期度 I の上位 10 件に含まれているページが，長期度 II の上位 10 件中に 6 ページ，下位 10 件中に 4 ページとすべて含まれていることが分かった．このように全部で 100 ページあるにも関わらず，ブックマーク日数が多いページが長期度 II の上位 10 件と下位 10 件にすべて含まれた．

長期度 III については，上位 10 件には長期的なタイプのページが多く，下位 10 件には短期的なタイプのページが多かった．この理由は，いつ見ても有用なページは，1 日あたりのブックマーク数のピークの値に対して，ブックマーク日数が多く，一時的に有用なページは，1 日あたりのブックマーク数のピークの値に対して，ブックマーク日数が少なくなるからであると考えた．こういったことから，長期度 III によってある程度長期的なページを測定できていることが分かった．

(3) 各長期度の関係性と特徴

3 つの長期度の順位を用いて，それぞれ相関係数を求めた．その結果，長期度 I と長期度 III には強い相関があり，長期度 III と長期度 I，長期度 II には明確な相関が見られなかった．

長期度 I と長期度 III に強い相関があった理由は，長期度 III の傾きに影響するもっとも大きい要因として，ブックマークされた日数の多さと 1 日辺りのブックマーク数のピークの大きさの 2 つがあげられる．もっとも影響の強い 2 つの要因のうち，1 つが長期度 I の指標一致しているため，相関が強かったと考えられる．このため，長期度 I と長期度 III は，利用された期間の長さによる影響が強く，長期間利用されているものが長期度が高くなる傾向にあることが分かった．

これに対して，長期度 II と他の長期度があまり相関がなかった理由について考察する．長期度 II の値の大きさは，ブックマーク日数が多いほどプラス方向への差分も大きくなりやすいが，逆にマイナス方向への差分も大きくなりやすい．このことから，長期度 I や長期度 III の値が大きかったものは，プラスの値になっても，マイナスの値になっても値が大きくなりやすい．そのた

め，順位相関係数の値が小さくなったのだと考えられる．この事実だけ見ると長期度 II は，長期的な情報を取得するのに向かないのではないかとも取れるが，長期度 II には，もう 1 つの特性が存在する．それは，利用数の安定性を測る特性である．一般的にブックマーク数の変化が少なく安定して利用されているものはべき分布と比較してプラスの値を取り，逆に急激なバーストをしているものはべき分布と比較してマイナスの値を取る．そのため，長期度 II では，長期的に利用されていてかつ安定して利用されているものが，もっとも高い長期度になる．この特性を利用して，長期度 I とは少し異なった，長期的に利用されていてかつ安定して利用されている情報を取得できる手法であることが分かった．同時に，長期的に利用されていても安定して利用されていないものは，長期度が低くなる手法であることが分かった．

7.1.3. 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，セレクトブクマと MasteryEye を開発した．

(1) セレクトブクマの考察

実験結果より，セレクトブクマにおいて,「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．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．この要因として，大手企業が運営している権威ある Web ページの評価が「映画」という分野において高かったこと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 ページが多数あり，充実した内容で長期的に利用できるような Web ページも多い．そのため，Google 検索でそういった大手企業の Web ページ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 ページ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 検索での検索結果の Web ページを選択したユーザが多く，Google 検索結果で高い適合率・再現率となり，セレクトブクマで低い適合率・再現率となったと考えられる．逆に「java」や「健康」といったような分野では，個人が作成した Web ページが多く，Web ページが玉石混交となっている．以上の結果から，Google 検索と比較してセレクトブクマでは，大手企業が運営している Web ページ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．

また，ブックマーク数順と比較して，時間情報を利用したセレクトブクマでは，特に検索結果上位の適合率・再現率が高かった．このことから，セレクトブクマにおいて，被検者が今後も利用

したいような長期的な Web ページを取得できていることが分かる．また，実験結果の Web ページを見ると，セレクトブクマの検索結果上位には，体系だった知識を得られる Web ページや有用な Web サービスが多いことが分かる．ブックマーク数順でもある程度こういった傾向は見られるが，時間情報を利用することによりこの傾向は強くなっている．以上のことにより，セレクトブクマは，特に長期的な Web ページ，体系だった知識を得られる Web ページや有用な Web サービスなどを手軽に取得できるものと考えられる．

(2) MasteryEye

MasteryEye では，長期度 II によってこれまで長期間安定して利用されてきた検索キーワードを取得している．評価実験の結果，MasteryEye で取得したキーワードが長期度が低かったものや reflexa と比較して，被検者が長期間利用してきたキーワードであったため，長期度 II の有効性を示すことができた．そして，こういった検索キーワードのように，ほとんどのキーワードが普段から検索されているような情報に対しては，長期度 II のような長期間利用されているという指標だけでなく，安定して利用されているという指標が有効であると考えられる．

また，MasteryEye で取得できる関連キーワードの中で，長期度が低いキーワードと長期度が高いキーワードを比較した結果，長期度が高いキーワードの方が，被検者が知っているキーワード，長期的に利用しているキーワード，重要であると思うキーワードが有意に多かった．このことから，長期度 II によって長期的で定番なキーワードを取得できる可能性が高いことが分かり，関連キーワードの中から長期的に検索され続けているキーワードを選択する手法を他の関連キーワード取得システムに適用した場合も有効であると考えた．

また，MasteryEye を既存の関連キーワード取得システム reflexa と比較実験をした結果，MasteryEye において被検者が知っているキーワード，長期的に利用しているキーワード，重要だと思うキーワードが多かった．このことから，既存の関連キーワード取得システムと比較して，より長期的で定番なキーワードが取得できているものと考えた．以上のことにより，新しい分野を手軽に学ぶにあたって提案手法の有効性を示唆することができる．

7.1.4. 長期的な情報の有用性

ここでは，これまでの分析や実験結果をふまえ，長期的な情報の有用性について考察する．

(1) 長期的な Web ページ

ソーシャルブックマークデータを分析することにより，長期度の高い Web ページの方が，長期的に利用できるタイプのページが多かった．また長期度の高いページの方が，データ収集後約 4

年4ヶ月経過した時点で，リンク切れとなっているページの割合が少なかった．

また評価実験結果より，セレクトブクマとブックマーク数順を比較した場合，時間情報を利用したセレクトブクマにおいて，被検者が今後も利用したいWebページがより多かった．また，特に上位10件の適合率が，セレクトブクマで高い値になっていたことからも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ある程度排除できていることがわかった．以上のことから，これまでに長期的に利用されてきたWebページは，今後も長期的に有用である可能性が高いことが分かった．

(2) 長期的な検索キーワード

MasteryEyeの評価実験の結果，被検者が長期的に利用している検索キーワードは，被検者にとって重要なキーワードであるものが多かった．被検者にとって重要なキーワードであるということとその分野を調べるための重要キーワードであることは完全に一致はしない．完全に一致はしないのだが，人が重要だと思うこととその分野を調べるために重要であることは関係性が高いと考えられる．また，長期度の高かったキーワード群と長期度の低かったキーワード群を実際に調べてみると，長期度の高かったキーワード群の方がその分野を調べるために有用で，今後も長期的に利用できる可能性が高いキーワードが多かった．以上のことから，長期的に利用されている検索キーワードは，多くの人にとって重要であり定番であるため，その分野を調べるために有用であることが分かった．

(3) 長期的な情報の有用性

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，本研究で定義した長期的な情報は，有用性が高いと考えられる．また，長期的な情報の有用性が高いということから，本研究で提唱した長期的検索は，重要であると示唆される．

7.2. 課題

本研究の課題について述べる．

7.2.1. 長期度の統一

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システムを提案したが，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，さまざま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，さまざま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，今回提案した長期度計算手法を改良して，より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，もっとさまざまなデータを利用し，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なってきたが，もっとさまざまなデータに対して，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，さまざまなデータに対して，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．

7.2.2. 時間情報の収集

さまざまなデータに対して分析をし，実験をするためには，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．そのため，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ことになる．

7.2.3. 視覚化とユーザインタフェース

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．

7.3. 展望

本研究の今後の展望について述べる．

7.3.1. さまざまな情報への応用

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だけでなく他のさまざま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，ファイル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，このようなさまざま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可能性があると考える．

7.3.2. 複数の長期的検索システムの連動と統合

複数の長期的検索システムを連動させて利用する仕組みを作ることも考えている．例えば，本研究で実装したセレクトブクマと MasteryEye を連動して利用することも可能である．MasteryEye で取得した長期的な検索キーワードを利用して，セレクトブクマ検索をすることによって，ユーザが入力したキーワードに対してより長期的な情報を得やすくなると考えている．この手法は 2 つの仕組が考えらる．1 つ目は，MasteryEye で取得した関連キーワード一覧すべてを利用してセレクトブクマ検索を行う手法である．2 つ目は，MasteryEye で取得した関連キーワードの中から検索意図にあったキーワードをユーザが選択し，セレクトブクマ検索を行うという手法である．

セレクトブクマでは，ユーザが良い検索キーワードを思いつかないといったような課題もあるし，長期的な Web ページを検索するために，短期的にしか利用されないようなキーワードを入力してしまうと，そのキーワードに関するページだけでは，長期的な Web ページがあまり取得できない可能性も高い．だが，例えユーザが入力したキーワードがあまり長期的で定番なキーワードでなくても，MasteryEye で取得した長期的で定番な関連キーワードを利用すれば，長期的な Web ページを取得できる可能性が高い．

また，もっと多くの情報に対して統合的に利用できる長期的検索のプラットフォームを作るということも考えられる．長期的検索システムが，対象とする情報ごとに分散しているよりも，1

つに統合されていて，1 つのサービスでキーワード，Web ページ，動画，レシピ，論文，本や商品などを検索できると非常に有用なシステムになる可能性があると考えられる．

7.3.3. 長期的検索の広がり

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，さまざま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．

7.3.4. 長期的検索の未来

10 年後，20 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．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ことが予測される．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，現状ではまだまださまざまなデータを保存し蓄積するようになってからの歴史は浅く，蓄積されているデータは，長くても 5 年〜10 年程度のものが多い．だが，10 年後，20 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，現状の長期的検索よりももっと精度や信頼性を向上させることができ，より有益な検索システムとすることが可能になるし，より多くの情報に対して長期的検索システムを開発することが可能になっているのではないかと考えている．また，情報の種類によるが，長期的な情報とみなす利用期間を 10 年〜20 年といったような非常に長いスパンにすることも可能となる．以上のようなことから，筆者は将来，長期的検索というビジョンが非常に重要になってくると考えている．

第8章　結論

概要

本研究の成果をまとめるとともに，本論文の総括を行なう．

8.1. 研究の成果

本研究では，長期的な情報に特化して検索する手法として，「長期的検索」のコンセプトを提案し，そのコンセプトに基づいたシステムを提案，実装，評価し，その有効性を検証した．本研究の成果は以下のようにまとめられる．

8.1.1. 長期的検索の提案

本研究では，長期的な情報を以下のように定義付けた．

(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

そして，(1) の情報は，(2) である可能性が高いという仮説を立てた．このような定義と仮説に基づき，長期的な情報を検索する手法，長期的検索の提案を行った．さらに，長期的検索を実現するために，情報がどれだけ長期的に利用されているかの指標である長期度の提案を行った．

また，本研究では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した．以下 3 つの長期度計算手法を提案した．

- 長期度 I：利用された日数による指標
- 長期度 II：べき分布よりどの程度大きいかよる指標
- 長期度 III：べき分布の傾きによる指標

8.1.2. システムの開発と評価

長期的検索の応用例として，長期的な Web ページを検索するシステムと長期的な検索キーワードを検索するシステムの提案，実装，評価を行った．

(1) 長期的な Web ページの検索

長期的な Web ページを検索するシステムとしてセレクトブクマを提案した．セレクトブクマでは，長期的な Web ページに特化して検索することにより，手軽に体系だった知識を得られる Web ページや今後何度も利用できるような Web ページを見つけたり集めたりすることができる．このような手法を実現するために，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ところ，分析結果から，長期的にブックマークされる Web ページは，長期的に利用されるようなタ

イプの Web ページであることが分かった．この分析結果に基づき，ブックマーク日数を利用した長期度計算手法を提案した．また，この長期度計算手法を用いて，セレクトブクマを実装し，評価実験を行った結果，以下のことが分かった．

- セレクトブクマと Google 検索を比較
  分野によって，セレクトブクマの方が優位な分野と Google 検索の方が優位な分野があった．この結果を考察することにより，セレクトブクマは Web2.0 のような玉石混合の Web ページ群から長期的な Web ページを取得するのに向いているのものであると考えた．

- 長期度を利用
  長期度を利用することによって，高順位に被検者が今後も利用した Web ページが占める割合が増加した．このことにより，長期度を利用することの有効性を示すことができた．

(2) 長期的な検索キーワードの検索

長期的な検索キーワードの検索システムとして MasteryEye を提案した．MasteryEye では，長期的な検索キーワードを見つけることにより，新しい分野について調べたり学んだりしようと思ったときに，その分野の定番キーワードを得ることができる．この手法を実現するために，Google 検索におけるキーワード検索ボリュームの時系列データを利用し，データがべき分布よりどれだけ大きいかを指標とした長期度計算手法を提案した．

また，MasteryEye について評価実験を行った結果，以下のことが分かった．

- 長期度 II の高いキーワードと低いキーワードを比較
  長期度 II の高いキーワードの方が被検者が長期間利用しているキーワードであることが分かった．このことから，長期度 II の計算手法の妥当性を示すことができた．

- 長期的に利用されているキーワードとそうでないキーワードを比較
  長期的に利用されているキーワードの方が被検者にとって重要であった．このことから，長期的な検索キーワードの有用性を示すことができた．

- 既存の関連キーワード検索システムである reflexa と MasteryEye を比較
  MasteryEye の方が被検者によって長期間利用されているキーワードと被検者が重要だと思うキーワードが多かった．このことから MasteryEye の有効性を示すことができた．

8.1.3. 長期的な情報と長期的検索の有効性の検討

本研究では，セレクトブクマと MasteryEye という 2 つのシステムを実装し，長期的な情報と長期的検索の有効性について検討した．評価実験により，長期的な Web ページと検索キーワード

は，双方とも有用であることを示すことができた．そのため，長期的な情報は，有用である可能性が高いといえる．さらに，評価実験によりセレクトブクマと MasteryEye の有効性を示すことができた．このようなことから，長期的検索の有効性や展望を示すことができた．

8.2. 総括

本研究では，長期的に利用できる情報に特化して検索するための概念として，「長期的検索」という概念を提案した．また，長期的検索を行うために「長期度」というその情報がどれだけ長期的に利用されているかの指標を提案した．

情報化社会の発展により，情報の短命化が進んできている．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アクセスシステムでは，長期的に利用できる情報を見つけることが難しくなってきた．これに対して，長期的に利用でき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．以上のような背景から，本研究では長期的検索という概念を提案した．

長期的検索の概念の有効性を示すために，本研究では，以下 3 点を目的とした．

- 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

以上のような目的を実現するために，長期的に利用されている情報に特化して検索するシステムを提案した．そして，具体的なシステムとして「Web ページ」，「検索キーワード」という 2 つの情報を対象にし，「セレクトブクマ」，「MasteryEye」という 2 つのシステムを提案し，実装した．

セレクトブクマは，長期的に利用できる Web ページに特化して検索するためのシステムである．ここではまず，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，Web ページがどのように利用され，ブックマークされているかを分析した．そして，その分析結果に基づき，長期的な Web ページを検索できるようにシステムの設計・実装を行った．セレクトブ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．また，システムの評価実験を行い，長期的 Web ページ検索の有効性を確認した．

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 検索における検索ボリューム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

キーワードを見つけることができるようにシステムの設計・実装を行なった．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．また，システムの評価実験を行い，長期的検索キーワードを得ることの有効性を確認した．

このようなシステムを提案し，本研究に関連する研究領域を紹介し，本研究の特徴や位置づけについて整理した．そして最後に，長期的検索の特徴や有効性について考察し，今後の課題と展望を示した．

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，インタラクション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．樋口先生には，評価実験に関しての多くのアドバイスと本論文や本研究に関する論文に関しての多くのコメントやアドバイスをして頂きました．

また，同じ安村研究室の先輩である児玉哲彦氏，渡邉恵太氏，永田周一氏，神原啓介氏に感謝いたします．児玉氏には，研究や技術の未来ビジョンなど多くの貴重なお話をして頂きました．渡邉氏には，アイデア発想方法や研究に対する考え方など多くの貴重な知見を頂きました．永田氏には，身近で多くの面白いシステムを開発して頂き，非常に大きな刺激を与えて頂きました．神原氏には，セレクトブクマのデータを集めるのにあたって，非常に有益なアドバイスを頂きました．

また，研究室の後輩である吉原建氏，元良龍太郎氏，藤沢和哉氏，秋山博紀氏，山本伶氏，白崎琢也氏，吉椿薫氏に感謝いたします．吉原氏は，もっとも長時間一緒に研究活動を行った一人で，同じ安村研の中では，もっとも研究や人生に対して大きな影響を与えて頂いた一人となりました．元良氏には，研究だけでなく研究室内のイベントなどに関しても非常に面白いアイデアを出して頂き，研究室に良い空気をもたらして頂きました．藤沢氏は，いつも真面目で見習うべき点が非常に多かったです．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，研究テーマ発想に関して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ンデザインプロジェクトに所属するみなさん，そして，私と共に研究生活を過ごしたみなさんには，日頃から研究に関するアドバイスや議論をして頂き，ときには共に食事や雑談などをして頂き，本当に感謝しております．

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．

2013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", *International Conference WWW/INTERNET (IADIS'11)*, 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'11)*, March 2011.

3. Taiki Ueno, Michiaki Yasumura, "Social Search with Shared Keywords", *International Conference WWW/INTERNET (IADIS'09)*, 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", インタラクション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. 上野大樹, 安村通晃, “セレクトブクマ：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.

参考文献

[1] Albert-laszlo Barabasi, Jennifer Frangos. *Linked: The New Science of Networks Science of Networks.* Basic Books, pp.288, 2002.

[2] Albert-laszlo Barabasi. *Bursts: The Hidden Patterns Behind Everything We Do, from Your E-mail to Bloody Crusades.* Plume Reprint, pp.320, 2011.

[3] J Raskin. *The humane interface: new directions for designing interactive systems.* Addison-Wesley Professional, pp.256, 2000.

[4] B. Scholkopf, C. J. C. Burges, A. J. Smola. *Advances in kernel methods : support vector learning.* The MIT Press, 1999.

[5] Ricardo Baeza-yates, Berthier Ribeiro-neto. *Modern Information Retrieval.* Addison Wesley, pp.544, 1999.

[6] 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.

[7] 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.

[8] 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.

[9] Andrew D. Booth. A Law of Occurrences for Words of Low Frequency. *Information and Control, Vol.10, No.4*, pp.386-393, 1967.

[10] John S. Breese, David Heckerman, Carl 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.

[11] 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.

[12] 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.

[13] 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.

[14] N. Cristianini, J. Shawe-taylor. An introduction to support vector machines and other kernel-based learning methods. *Cambridge University Press*, 2000.

[15] 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.

[16] Micah Dubinko, Ravi Kumar, Joseph Magnani, Jasmine Novak, Prabhakar Raghavan, Andrew Tomkins. Visualizing tags over time. *World Wide Web Conference Series (WWW'06)*, pp.193-202, 2006.

[17] Colonel Tim Gibson, Ethan L. Miller. Long-term File Activity and Inter-Reference Patterns. *Computer Measurement Group Conference (CMG'98)* , 1998.

[18] Scott A. Golder, Bernardo A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science Volume 32 Issue 2*, pp.198-208, 2006.

[19] Granovetter M. Threshold Models of Collective Behavior. *American Journal of Sociology, Vol.83(6)*, pp.1420-43, 1978.

[20] Granovetter M, R Soong. Threshold Models of Interpersonal Effects in Consumer Demand. *Journal of Economic Behavior and Organization*, 7, pp.83-99, 1986.

[21] Simon Hermann. Dynamics of Price Elasticty and Brand Life Cycle. *Journal of Marketing Research* ,16(4), pp.439-45, 1979.

[22] Paul Heymann, Georgia Koutrika, and Hector Garcia-Molina. Can social bookmarking improve web search? *Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM'08)*, 2008.

[23] 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.

[24] Glen Jeh, Jennifer Widom. Scaling personalized web search. *World Wide Web Conference Series (WWW'03)*, pp.271-279, 2003.

[25] Thorsten Joachims, Dayne Freitag. WebWatcher: A Tour Guide for the World Wide Web. *TInternational Joint Conference on Artificial Intelligence (IJCAI'96)*, 1996.

[26] Jon Kleinberg. Bursty and hierarchical structure in streams. *Proceedings of the 8th ACM SGKDD International Conference on Knowledge Discovery and Data Mining (SGKDD'02)*, pp.91-101, 2002.

[27] Jon Kleinberg and Mark Sandler. Using mixture models for collaborative filtering. *Journal of Computer and System Sciences Volume 74*, pp.49-69, 2008.

[28] 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.

[29] S. Loeb, D. Terry. Information filtering. *Communications of the ACM, Vol.35, No.12*, pp.26-81, 1992.

[30] 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.

[31] 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.

[32] 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.

[33] Meenakshi Nagarajan, Karthik Gomadam, Amit P. Sheth, Ajith Ranabahu, Raghava Mutharaju, Ashutosh Jadhav. Spatio-Temporal-Thematic Analysis of Citizen Sensor Data: Challenges and Experiences. *Web Information Systems Engineering (WISE'10)* , pp. 539-553, 2009.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] Jorg Rech. Discovering trends in software engineering with google trend. *ACM Sigsoft Software Engineering Notes, Vol.32, No.2*, pp.1-2, 2007.

[39] 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.

[40] Paul Resnick, Hal R. Varian. Recommender systems. *Communications of the ACM (CACM'97) , Vol.40, No.3*, pp.56-89, 1997.

[41] D. Riecken. Personalized views of personalization. *Communications of the ACM, Vol.43, No.8*, pp.26-158, 2000.

[42] S. E. Robertson. On term selection for query expansion. *Journal of documentation, Vol.46, No.4*, pp.359-364, 1990.

[43] J. J. Roccio. Relevance feedback in information retrieval. *The Smart Retrieval System: Experiments in Automatic Document Processing, ed. G.Salton*, pp.313-323, 1971.

[44] 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.

[45] 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.

[46] Ben Shneiderman. Dynamic Queries for Visual Information Seeking. *IEEE Software - SOFTWARE , Vol.11, No.6*, pp.70-77, 1994.

[47] 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.

[48] Stephen Strange. Analysis of Long-Term UNIX File Access Patterns for Application. *Technical Report, UCB/CSD 92/700*, University of California, Berkeley, 1992.

[49] 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.

[50] 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.

[51] Jianshu Weng, Ee-Peng Lim, Jing Jiang, Qi He. TwitterRank: finding topic-sensitive influential twitterers. *Proceedings of the third ACM international conference on Web search and data mining (WSDM '10)* , pp. 261-270, 2010.

[52] 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.

[53] Justin Wolfers, Eric Zitzewitz. Prediction Markets. *Financial Analysts Journal - FINANC ANAL J, Vol.18, No.2*, pp.107-126, 2004.

[54] 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.

[55] 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.

[56] 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.

[57] 枝洋樹, 林信行, 小林弘人, 津田大介, 武田徹, 高須賀宣, 岡野原大輔, 片瀬京子, 高橋秀和, 亀津敦, 日経 BP 社出版局. Twitter の衝撃 - 140 文字がビジネスからメディアまで変える. 日経 BP 社, pp.208, 2009.

[58] 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.

[59] 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.

[60] 経済産業省, 文部科学省 (編集), 厚生労働省. ものづくり白書〈2007 年版〉. ぎょうせい, 2007.

[61] 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, Vol.83, No.6, pp.460-466, 2000.

[62] 梅田望夫. ウェブ進化論 - 本当の大変化はこれから始まる. 筑摩書房, pp.256, 2006.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] 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. Vol.57, No.6, pp.302-307, 2012.

[68] 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. Vol.19, No.3, pp.365-372, 2004.

[69] 久野高志, 安形輝, 上田修一. 情報検索システムとしてみたサーチエンジン. 第49回日本図書館情報学会研究大会, 2001.

[70] 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.

[71] 岩根良輔 , 藤本典幸 , 萩原兼一. 協調フィルタリングを用いて個人の嗜好を反映するレシピ検索手法の提案. 情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告 2008 (126), pp.163-166, 2008.

[72] 神嶌敏弘. 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.

[73] 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, Vol.46, pp.14-25, 2005.

[74] 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.

[75] 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.

[76] 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.

[77] 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. Vol.50, No.4, pp.1251-1261, 2009.

[78] 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.

[79] 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, Vol.9, No.1, pp.12-17, 2010.

[80] 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.

[81] 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.

[82] 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, Vol.69, pp.53-58, 2003.

[83] 松田光司. 流行現象とベクトル型閾値モデル. 理論と方法. Vol.11, No.2, pp.113-128, 1996.

[84] 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.

[85] 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.

[86] 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, Vol.37, No.8, pp.751-758, 1996.

[87] 村田眞哉, 戸田浩之, 松浦由美子, 片岡良治. クリックログ解析による情報要求ベースの検索結果ランキング. 日本データベース学会論文誌, Vol.7, No.4, pp.37-42, 2009.

[88] 中村聡史, 田中克己. 印象に基づく動画検索. 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 2009, Vol.5, pp.77-84, 2009.

[89] 中山晶一朗. 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, Vol.18, No.1, pp.89-102, 2003.

[90] 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, Vol.11, pp.9-14, 2008.

[91] 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.J 92-D, No.3, pp.281-292, 2009.

[92] 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, 2005.

[93] 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.

[94] 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, 2004.

[95] 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, Vol.48, pp.14-27, 2007.

[96] 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.

[97] 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.

[98] 荘司慶行, 田中克己. 印象語をクエリとするリアクションに基づくウェブ情報検索. 情報処理学会論文誌, Vol.52, No.12, pp.3515-3526, 2011.

[99] 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. Vol.43, No.12, pp.3698-3705, 2002.

[100] 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, Vol.17, No.1, pp.11-31, 2004.

[101] 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, 1995.

[102] 上田真由美, 石原和幸, 平野靖, 梶田将司, 間瀬健二. 食材利用履歴に基づき個人の嗜好を反映するレシピ推薦手法. 日本データベース学会 letters, Vol.6, No.4, pp.29-32, 2008.

[103] 渡邊恵太. ユビキタス環境における眺めるインタフェースの提案と実現. 情報処理学会論文誌, Vol.49, No.6, pp.1984-1992, 2008.

[104] 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, Vol.2, No.3, pp.130-140, 2009.

[105] 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.

[106] 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.

[107] 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.

[108] Amazon.com. Amazon. http://www.amazon.com/, (参照 2012-12-13).

[109] Microsoft. Bing. http://www.bing.com/, (参照 2012-12-13).

[110] 株式会社エルネット. ボブとアンジー. http://www.bob-an.com/, (参照 2013-2-12).

[111] 株式会社ECナビ. Buzzurl [バザール]. http://buzzurl.jp/, (参照 2012-12-13).

[112] COOCKPAD. レシピ検索 No.1 ／料理レシピ載せるなら クックパッド. http://cookpad.com/, (参照 2013-2-12).

[113] Yahoo! Delicious. http://delicious.com/, (参照 2012-12-13).

[114] Digg. Digg. http://digg.com/, (参照 2012-12-13).

[115] Yahoo!. Flickr. http://www.flickr.com/, (参照 2012-12-13).

[116] Google. Google. http://www.google.com/, (参照 2012-12-13).

[117] Google. Google Insights for Search. http://www.google.com/insights/search/, (参照 2012-5-31).

[118] Google. Google トレンド. http://www.google.com/trends/, (参照 2013-1-10).

[119] 株式会社はてな. はてなブックマーク. http://b.hatena.ne.jp/ , (参照 2012-12-26).

[120] Takeshi Miki. Lingua-JA-Expand. http://search.cpan.org/ miki/Lingua-JA-Expand/, (参照 2012-12-13).

[121] JTNEWS. みんなのシネマレビュー. http://www.jtnews.jp/, (参照 2013-2-12).

[122] Preferred Infrastructure. 連想検索エンジン reflexa. http://labs.preferred.jp/reflexa/, (参照 2012-12-13).

[123] Kakaku.com. 食べログ ランキングと口コミで探せるグルメサイト. http://tabelog.com/, (参照 2012-12-26).

[124] Twitter. Twitter. https://twitter.com/, (参照 2012-12-26).

[125] ヤフー株式会社. Yahoo!ブックマーク. http://bookmarks.yahoo.co.jp/all, (参照 2012-12-26).

[126] Google. YouTube. http://www.youtube.com/, (参照 2012-12-13).

付録A
検索結果一覧

ここでは，セレクトブクマの評価実験における検索結果一覧を提示する．Google検索，ブックマーク数順に検索，セレクトブクマ(α=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.mspx
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/ codingStandard/writingrobustjavacode.html
4	Java 技術最前線：ITpro	http://itpro.nikkeibp.co.jp/article/ COLUMN/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/ COLUMN/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/ urlrrewrite/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/COLUMN/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/codingStandard/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/COLUMN/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 > Japanese > 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ｊｐ (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９９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/COLUMN/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 時間以下の睡眠続け死亡率 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...	http://rate.livedoor.biz/archives/ 50679219.html
20	指をポキポ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/llife01/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 — 睡眠不足だと仕事がはかどらない理由、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/~aiaihide/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だ（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！ 肩コリ解消法 — 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/~aiaihide/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　眼の疲れを取る：ITpro	http://itpro.nikkeibp.co.jp/article/COLUMN/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もらいました。 — 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/fareastblog/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	メルクマニュ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・シネマズー映画館、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/tsumiyama/ 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？ - ベ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、映画的ライフスタイル・エンターテイメント・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:観るまでバカにして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://wwws.warnerbros.co.jp/skycrawlers/

セレクトブクマで「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、映画的ライフスタイル・エンターテイメント・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://wwws.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

ここではセレクトブクマの評価実験の結果一覧を示す.「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 — 睡眠不足だと仕事がはかどらない理由、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もらいました。— 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 > Japanese > 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	シネマぴあ			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「ぴあ映画生活」			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	

長期的に利用する情報へのアクセスに関する研究

2013年3月25日　　初版発行

著者　上野大樹

監修　安村通晃

発行　慶應義塾大学 湘南藤沢学会
〒252-0816　神奈川県藤沢市遠藤5322
TEL:0466-49-3437

Printed in Japan　　印刷・製本　　ワキプリントピア

ISBN 978－4－87762－265－7
SFC-DT 2012-002

■本論文は博士論文において優秀と認められ、出版されたものです。