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

Title	Web閲覧履歴およびE-mail履歴からの興味抽出による思考支援システム
Sub Title	A system for supporting conception by extracting user's interests underlying web pages and e-mails
Author	河津, 功典(Kawazu, Kosuke) 前野, 隆司(Maeno, Takashi)
Publisher	慶應義塾大学大学院システムデザイン・マネジメント研究科
Publication year	2009
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2009年度システムデザイン・マネジメント学 第10号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40002001-00002009-0013

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.

修士論文　　2009 年度

Web 閲覧履歴および E-mail 履歴からの興味抽出による思考支援システム

河津　功典
（学籍番号：80833154）

指導教員　教授　前野　隆司

2010 年 3 月

慶應義塾大学大学院システムデザイン・マネジメント研究科
システムデザイン・マネジメント専攻

論　　文　　要　　旨

学籍番号	80833154	氏　名	河津　功典

論 文 題 目：

Web 閲覧履歴および E-mail 履歴からの興味抽出による思考支援システム

（内容の要旨）

本研究は，ユーザの興味を把握し，その興味に関連する事柄を提示することで，ユーザの思考を支援するための，新たなシステムを提案するものである．

まず，人の思考を支援する方法について述べる．本研究では，人の興味拡大に焦点を当て，新たな知識獲得を促すことにより，思考の支援を行う．人は，未知の情報と自分の持っている知識をインタラクションさせ，思考を行う．この思考を支援するためには，3 つの方法がある．3 つの方法とは，未知の情報を効率的に抽出する方法，人に新たな知識を獲得させる方法および未知の情報と知識を結び付けやすくする方法である．本研究では，新たな知識の獲得に焦点を当て，人の思考を支援する．新たな知識の獲得には，興味を拡大することが有効である．したがって，本研究は，ユーザの興味を把握し，それに即した情報提示を行うことで，ユーザの興味を拡大し，思考を支援する．本研究では，Web 閲覧履歴および E-mail 履歴を用いてユーザの興味を抽出する．

つぎに，提案するシステムの概要について述べる．本システムの特徴は，新たに興味を抱くきっかけに基づき，ユーザの興味を拡大することである．ユーザの興味拡大には，その目的に即した適切な情報提示が有効である．興味拡大に適切な情報をユーザに提示するため，新たに興味を抱くきっかけを調査した．その結果，新たに興味を抱くきっかけは，(1)自分の興味に関連する事柄，(2)所属するグループ内の共通事項，(3)自分の感性に一致する事柄，(4)社会的に重要な事柄，(5)予想外の事柄，(6)予想通りの事柄の 6 種類があることがわかった．(1)を提示するためには，ユーザの興味をシステムに把握させる必要がある．そのため，ユーザの興味を把握するためのサブシステムを構築する．また，(4)，(5)および(6)は，単語の社会的重要度およびユーザの興味との関連度を用いて表すことができる．したがって，これら 2 つを用いて提示する情報を決定するためのサブシステムを構築する．さらに，興味拡大をさせやすくするように，情報提示方法も考慮する．

以上より，本システムは，以下の 3 つのサブシステムを用いて構築した．

- Web 閲覧履歴および E-mail 履歴からユーザの興味を抽出する，ユーザモデル構築システム
- ユーザの興味に関連した単語を抽出し，単語の社会的重要度およびユーザの興味との関連度を用いて提示する情報を決定する，外部情報整理システム
- マインドマップを模倣した情報提示を行うユーザインターフェース

構築したシステムの検証を行った結果，ユーザの興味を考慮することの有効性，提案する情報提示法の有効性およびマインドマップを模倣した情報提示法の有効性を示した．

最後に，構築したシステムの評価を行った．この評価では，本システムと既存のシステムである Yahoo!関連検索および 5 分間のネットサーフィンと比較した．その結果，本研究で提案するシステムの有効性を示した．

SUMMARY OF MASTER'S DISSERTATION

Student Identification Number	80833154	Name	Kousuke Kawazu

Title

A System for Supporting Conception by Extracting User's Interests Underlying Web Pages and E-mails

Abstract

This study suggests new system for supporting a user's conception by figuring out his interests and displaying words that relate to his interests.

Firstly, we describe about a way to support the user's conception which is helped by expanding his interests. Human's conception is impacted on interacting between his unknown information and his own knowledge. There are three ways to support this conception. Three ways are to extract information that he doesn't know effectively, to be acquired his new knowledge and to relay the information to his knowledge easily. In this study, it is chosen to be acquired his new knowledge for supporting the user's conception. In order to be acquired new knowledge it is effectual to expand the user's interests. Therefore, this study suggests new system for supporting a user's conception by figuring out his interests and displaying words that relate to his interests. The user's interests are extracted from web pages and e-mails.

Secondly, we describe about a system summary. The user's interests are expanded by using triggers. In order to expand his interests it is effectual to display appropriate information. Six triggers to expand his interests are cleared up. The six triggers are (1)something that relates to own interests, (2)something common to own groups, (3)something that fits own emotion, (4)something important, (5)something unexpected and (6)something expected. In order to display (1) a subsystem in the system needs to figure out his interests. (4), (5) and (6) are presented by using importance of the words and degree of association to the user's interests. Therefore, this system has a subsystem for deciding information to display by using these two marks. In order to expand his interests easily it is considered the way to display information.

The system is constructed by three subsystems based on above architecture.

- A subsystem to construct a user's model
- A subsystem to get information from the Internet
- A subsystem to indicate outputs by user interface

First subsystem constructs a user's model to figure out his interests. These interests are extracted from web pages and e-mails that he is interested in. Second subsystem gets information from the Internet. This extracts words that relate to the user's interests and defines outputs by using importance of the words and degree of association to the user's interests. Third subsystem is a user interface. He inputs necessary parameters into this. This presents information by arranging radially like mind map.

Verification of this system showed efficacy of considering the user's interests, way of extracting information, way of presenting information like mind map.

Finally, as the result of validation of this system, the proposed system is more useful about expanding the user's interests than two existing systems, which one is surfing websites for five minutes, another is Unit Search that searches related words to one word by using Yahoo! Japan.

目 次

第 1 章　序論 **1**
1.1 背景 1
1.2 目的 4
1.3 従来研究 5
1.3.1 興味喚起型システム 5
1.3.2 連想検索型システム 11
1.3.3 創造性支援型システム 13

第 2 章　要求分析 **15**
2.1 興味拡大における現状の問題点 15
2.2 提案するシステムが生み出す価値 19

第 3 章　思考支援システムの設計 **23**
3.1 新たに興味を抱くきっかけの分類 23
3.2 提案するシステムの概要 25
3.3 ユーザモデル構築システム 26
3.3.1 内容解析部 28
3.3.2 グラフ生成部 33
3.3.3 モデル修正部 33
3.4 外部情報整理システム 34
3.4.1 情報抽出部 34
3.4.2 出力決定部 38
3.5 ユーザインターフェース 39
3.5.1 入力部 40
3.5.2 提示部 40

第 4 章　思考支援システムの構築 **42**
4.1 ユーザモデル構築システム 42
4.1.1 内容解析部 42
4.1.2 グラフ生成部 45
4.1.3 モデル修正部 45
4.1.4 動作確認 47
4.2 外部情報整理システム 49
4.2.1 情報抽出部 49
4.2.2 出力決定部 51
4.2.3 動作確認 53
4.3 ユーザインターフェース 56
4.3.1 入力部 56
4.3.2 提示部 57
4.3.3 動作確認 60

第 5 章　思考支援システムの検証 **62**
5.1　ユーザモデル構築システムの精度の検証 62
5.1.1　実験目的 62
5.1.2　実験方法 62
5.1.3　結果および考察 63
5.2　情報抽出タイプおよび出力決定タイプによる影響の検証 65
5.2.1　実験目的 65
5.2.2　実験方法 66
5.2.3　結果および考察 66
5.3　構築したサブシステムの有効性の検証 69
5.3.1　実験目的 69
5.3.2　実験方法 70
5.3.3　結果および考察 71
5.4　ユーザの興味の考慮が興味拡大に与える影響の検証 73
5.4.1　実験目的 73
5.4.2　実験方法 74
5.4.3　結果および考察 75

第 6 章　思考支援システムの評価 **77**
6.1　実験目的 77
6.2　実験方法 78
6.3　結果および考察 78

第 7 章　結論および今後の展開 **82**
7.1　結論 82
7.2　今後の展開 83

謝辞 **86**

参考文献 **87**

付録 **91**

図 目 次

1.1 情報推薦システムの変遷 ([2] を改変) 1
1.2 協調フィルタリングの概要 [2] 2
1.3 WebWatcher[8] 2
1.4 Amazon のトップページ [11] 3
1.5 GoogleNews のトップページ [12] 3
1.6 従来の情報推薦システムの提示結果と本システム提示結果の違い 3
1.7 本稿の構成 4
1.8 吹き出し表示 [15] 5
1.9 人型表示 [15] 5
1.10 MineBlog のシステム概要 [16] 6
1.11 MineBlog を用いた推薦例 [16] 6
1.12 クリップボード共有システムの概要 [17] 6
1.13 データの表示方法 [17] 7
1.14 ブラウジング支援システムの概要 [18] 8
1.15 ブラウジング支援例 [18] 8
1.16 意外な Web ページの提示例 [19] 9
1.17 QueReSeek 実行時の Web ブラウザ画面 [21] 10
1.18 検索キーワード提示画面 [21] 10
1.19 GETA の実行例 [22] 11
1.20 WAM[24] 12
1.21 連想検索の流れ ([25] を改変) 12
1.22 文書-単語連想検索の模式図 [24] 12
1.23 記憶の想起に基づく創造性支援システム [26] 12
1.24 発想のヒントを提示するシステム [27] 13
1.25 発想システムの構成 [27] 13

2.1 連想検索エンジン reflexa[30] 15
2.2 連想検索エンジン ASSOCIE[31] 16
2.3 想 IMAGINE BOOK SEARCHE[32] 16
2.4 ASSOCIE の表示方法 [31] 17
2.5 興味拡大における現状の問題点の分類 17
2.6 情報量と人の関心の関係図 [35] 18
2.7 刺激強度と人の関心の関係図 [35] 18
2.8 QRIO SDR-4X [39] 20
2.9 投票率の推移 [40] 20
2.10 検索結果に表示される広告 21

3.1 新たな興味を抱くきっかけの分類 23
3.2 知的好奇心の概念図 [44] 24
3.3 興味拡大における CVCA 25

3.4 本研究で想定する CVCA 25
3.5 システム全体の機能分解 26
3.6 システム全体の流れ 27
3.7 ユーザモデル構築システムの機能分解 27
3.8 Pugh Selection の結果 29
3.9 検索キーワードの分類結果 29
3.10 メタデータの含有率 31
3.11 E-mail に必要なメタデータのアンケート結果 32
3.12 外部情報整理システムの機能分解 33
3.13 Satori 検索の例 34
3.14 新たに興味を抱く要素 35
3.15 情報抽出タイプ 1 の概念図 35
3.16 情報抽出タイプ 2 の概念図 36
3.17 情報抽出タイプ 3 の概念図 36
3.18 社会的重要度と興味との関連度で表す分類 38
3.19 入力部の機能分解 39
3.20 提示部の機能分解 39
3.21 Mind Map の例 [63] 40

4.1 SA の概念図 ([65] を改変) 43
4.2 単語および関係性の抽出例 43
4.3 実際に出力される XML の例 46
4.4 実際に出力される興味モデルの例 46
4.5 用意した Web ページ 47
4.6 用意した E-mail 47
4.7 構築されたユーザの興味モデル 47
4.8 モデルの拡大の動作確認 48
4.9 モデルの縮小の動作確認 48
4.10 頂点の移動の動作確認 49
4.11 Yahoo! API の出力結果例 49
4.12 検索ヒット数および単語の出現回数を載せた XML 51
4.13 同じ単語が出力されるマップ 52
4.14 情報抽出タイプ 1 の抽出結果 53
4.15 論文に関連する単語 (1) 53
4.16 情報抽出タイプ 2 の抽出結果 53
4.17 論文に関連する単語 (2) 53
4.18 慶應義塾大学に関連する単語 54
4.19 情報抽出タイプ 3 の抽出結果 54
4.20 出力決定タイプ 1 の結果 55
4.21 出力決定タイプ 2 の結果 55
4.22 出力決定タイプ 3 の結果 55
4.23 出力決定タイプ 4 の結果 55
4.24 入力部の画面 56
4.25 情報抽出タイプ 1 の出力例 58
4.26 情報抽出タイプ 2 の出力例 58
4.27 情報抽出タイプ 3 の出力例 58
4.28 実装した検索機能 59

4.29 動作確認に用いた提示結果 60
4.30 「この単語を検索する」の検索結果 60
4.31 「2 つの関係性を検索する」の検索結果 61
4.32 「3 つの関係性を検索する」の検索結果 61

5.1 評価項目 (a) における検定結果 67
5.2 評価項目 (b) における検定結果 67
5.3 評価項目 (c) における検定結果 67
5.4 評価項目 (d) における検定結果 67
5.5 評価項目 (e) における検定結果 67
5.6 評価項目 (f) における検定結果 67
5.7 評価項目 (g) における検定結果 68
5.8 情報提示方法を変更した本システム 70
5.9 余計なキーワードに関するアンケート結果 73

7.1 実験に用いた評価シート 91

表 目 次

3.1 定義する単語間の関係性のメタデータ 32

4.1 メタデータと背景色の対応 46
4.2 想定した単語とメタデータの対応 47
4.3 各出力決定タイプの想定する出力 54

5.1 Web ページおよび E-mail の解析精度 63
5.2 12 パターンを比較した実験結果 66
5.3 情報抽出タイプ別の平均得点および標準偏差 68
5.4 出力決定タイプ別の平均得点および標準偏差 69
5.5 サブシステムの有効性検証の実験結果 71
5.6 興味ある単語の含有率 74
5.7 モデルの違いを比較した実験結果 75

6.1 本システムと既存のシステムを比較した実験結果 78
6.2 ネットサーフィンの方法別に比較した実験結果 79

第1章　序論

本研究では，興味拡大を用いた思考支援を行う新たなシステムを提案する．まず，本章では，本研究での問題意識および本研究の位置づけについて述べる．1.1節では，従来の情報推薦システムの変遷について述べるとともに，本研究における問題意識を述べる．1.2節では，本研究の目的を述べる．1.3節では，従来の興味喚起型システムといった，現在までに提案されてきたシステムを簡単にまとめ，本研究の位置づけを述べる．

1.1　背景

人の思考は，心の中の情報処理の過程である．人の思考を支援するためには，未知の情報と自分の持っている知識のインタラクションを支援することが有効である [1]．人は，未知の情報と自分の持っている知識をインタラクションさせ，思考を行うからである．

まず，未知の情報について考えてみる．近年，インターネットの発展および普及に伴い，人の取得できる情報量が莫大に増えた．世界のインデックス可能な Web ページは，2005 年 1 月の段階で 115 億ページを超えたと推定されている [3]．また，Web ページの数は日に日に増加している．このように，インターネット上には，膨大な情報が存在し，ユーザにとって未知の情報が多く存在する．また，ユーザが望めば，様々な事柄に関する情報を取得することが可能である．情報を取得するツールとして，google[4] および Yahoo[5] といった検索エンジンやユーザの望む情報を適切に提供するための様々なシステムが考案されてきた．これは，扱える情報が多すぎて，本当に必要な情報を取得できないという問題や検索エンジンでは，効率的に情報を取得できないという問題があったからである．これらのシステムは，図 1.1に示すように変遷を遂げてきた．以下に，従来提案されてきたいくつかの情報推薦システムを挙げる．

土方 [2] によれば，ユーザの嗜好を抽出するための先駆的な研究は，1986 年に MIT の Malone らが開発した Information Lens[6] である．このシステムでは，電子メールやネットニュースのメッセージのヘッダに，配送や返信に関する情報だけでなく，内容に関連する場所，時間およびトピックなどの情報も構造化した．この研究では，半構造化文書の特性を知的処理に応用するという斬新な考え方を提案してい

	1980年代後半	1990年代前半	1990年代後半	2000年代前半
研究のトレンド	インターネットプロトコルを利用した情報フィルタリングシステムの提案	半構造化文書と情報検索からのアプローチによる情報フィルタリングの研究	協調フィルタリングおよび機械学習による内容に基づくフィルタリング方式の研究	実用的な協調フィルタリング方式に関する研究
代表的研究	・Information Lens(87)	・GroupLens(94)	・WebWatcher(97)	・推薦根拠の提示(00) ・アイテムベースの推薦方式(01)
時代背景	・研究機関へのインターネットの普及	・WWW創成期	・一般ユーザへのWWWの普及 ・Amazon.comの誕生	・一般企業のWWWの本格的ビジネス利用 ・ネット企業における情報サービスの導入

図 1.1: 情報推薦システムの変遷 ([2] を改変)

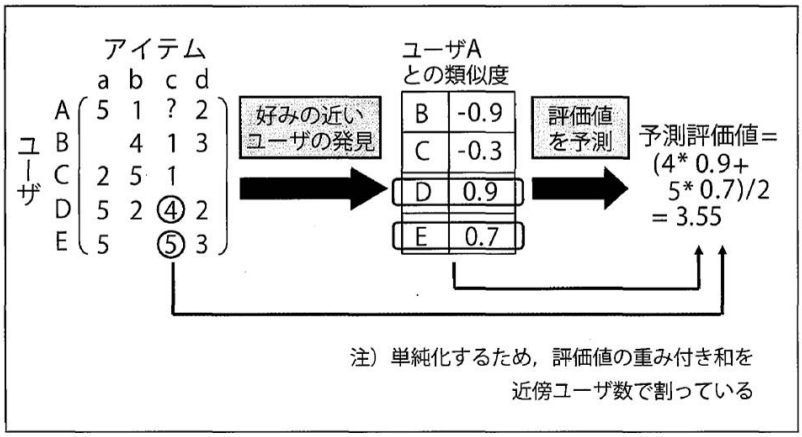


図 1.2: 協調フィルタリングの概要 [2]

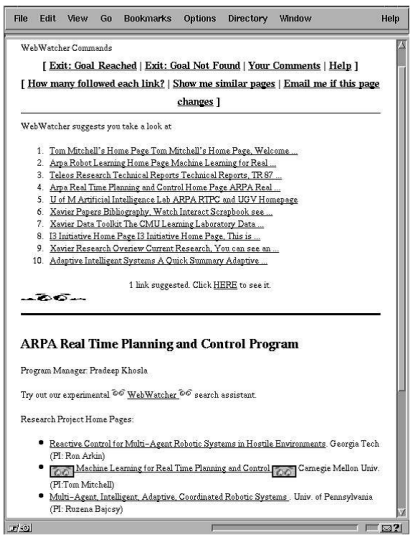


図 1.3: WebWatcher[8]

る．Malone らの研究を皮切りに，様々な情報推薦システムが提案されてきた．

Resnick らは，図 1.2に示す GroupLens を提案した [7]．GroupLens は，ある記事に対するユーザの評価をもとに，そのユーザと評価の傾向が似ている他のユーザを自動的に抽出し，ユーザに推薦を行う．この GroupLens は協調フィルタリングの基礎となっている．

Joachimes らは，図 1.3に示す WebWatcher を提案した [8]．WebWatcher は，ユーザの Web ページへのアクセス履歴からユーザプロファイルを作成し，そのユーザプロファイルに基づいて情報推薦を行うものである．この研究では，機械学習アルゴリズムを適用し，ユーザプロファイルを更新するという概念を提案している．

Sarwar らは，アイテムベースの協調フィルタリングを提案した [9]．アイテムベースの協調フィルタリングは，アイテム間の類似度を計算するアルゴリズムである．この研究では，ユーザ間の類似度ではなく，推薦対象の類似度を計算するという手法を提案した．この研究は，現在の推薦サービスの基となっている．

情報推薦システムが提示した結果に対して，ユーザの納得感を高めるため，情報推薦システムには，心的根拠やユーザの状況といった推薦に至った根拠を示すことが望まれ始めた．そのため，Herlocker らは，ユーザに情報を推薦するだけでなく，なぜその情報を推薦したのかという，推薦の根拠を示す試みを行った [10]．

他にも，ユーザは一人でいるのか恋人といるのかといった，現在のユーザの状況とユーザの過去の行

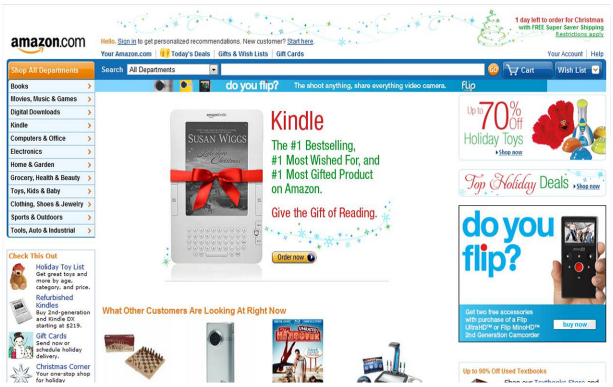


図 1.4: Amazon のトップページ [11]

図 1.5: GoogleNews のトップページ [12]

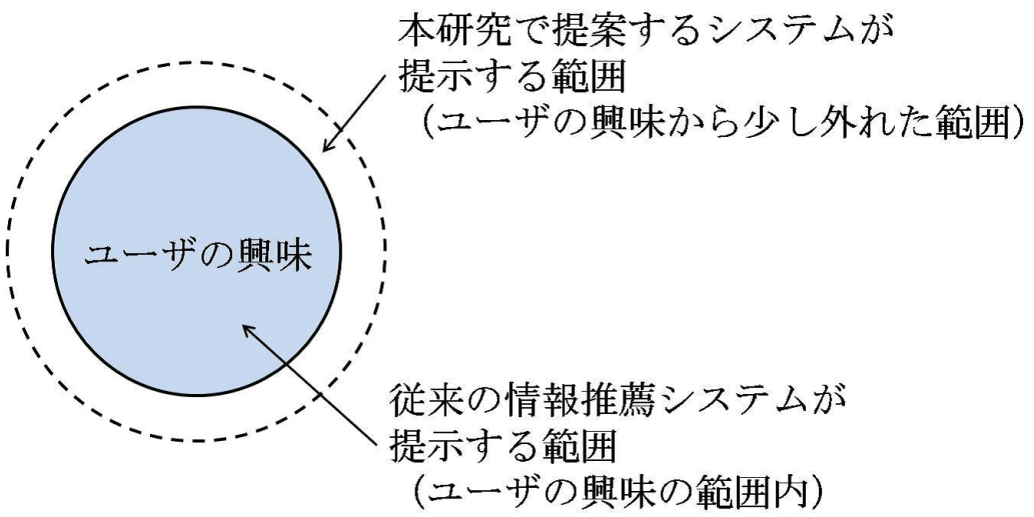


図 1.6: 従来の情報推薦システムの提示結果と本システム提示結果の違い

動履歴に基づく情報推薦システムを提案した奥らの研究 [13] などが提案されている．また，実際に推薦サービスを行っているものに，図 1.4に示す Amazon.com[11]，図 1.5に示す Google News[12] およびユーザの好みの番組を自動的に録画するソニーのスゴ録 [14] といったものがある．

以上のように，現在までに，様々な情報推薦システムが提案されてきた．すなわち，インターネット上には，ユーザにとって未知の情報が膨大に存在し，それらを効率的に抽出するシステムも整備しつつある．

つぎに，知識について考えてみる．人は，新たな知識を得たいという知的好奇心を持っている．新たな知識の獲得には，知的好奇心を刺激するといった支援が有効である．知的好奇心を刺激するための 1 つの方法として，人の興味の拡大が挙げられる．人の興味が拡大すれば，人は積極的に情報を収集し，新たな知識を獲得することができると考えられる．したがって，新たな知識の獲得には，興味の拡大が有効であると考えられる．

このように，思考支援の 1 つの方法として，新たな知識の獲得を目的とした興味拡大が挙げられる．思考支援のための興味拡大には，従来の正確さを重視した情報推薦システムではなく，興味の拡大を目的とした情報推薦システムを構築することが有効である．今までのシステムでは推薦するには不適切および不正確な情報とされていた情報が，興味の拡大には必要であると考えられるからである．図 1.6に示すように，従来の情報推薦システムは，ユーザが興味を抱いている情報のみをユーザへ提示していた．そのため，本研究では，ユーザの興味から少し外れた情報を提示する，興味拡大を目的とした情報推薦システムに焦点を当てる．

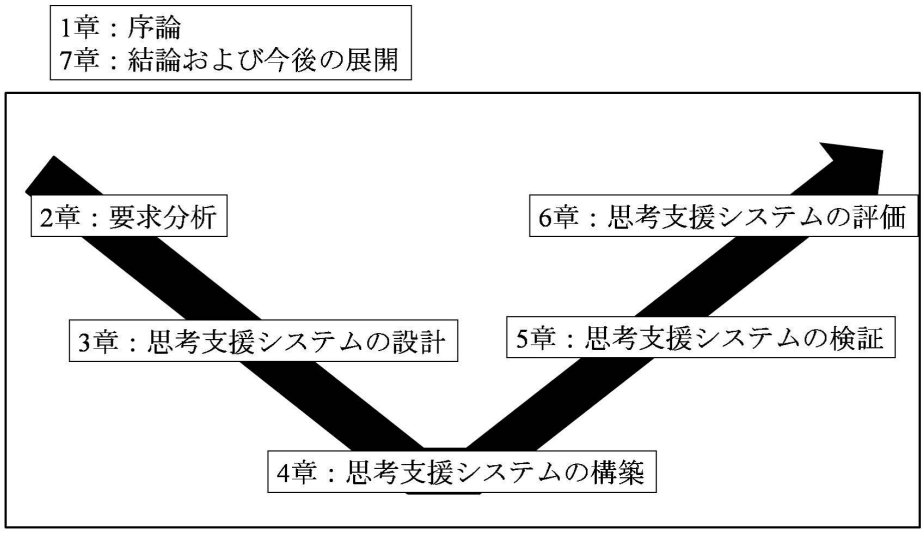


図 1.7: 本稿の構成

1.2 目的

本研究では，要求分析を行い，興味拡大における現状の問題点を明らかにした．その結果，

- 外部の情報を取り入れることができない
- ユーザが次の興味の対象を予測することができない

という 2 つの問題があることがわかった．現状分析については，2.1節で詳しく述べる．これら 2 つの問題を解決するため，本研究では，以下の 2 つの特徴を有するシステムを提案する．まず，ユーザが効率よく外部の情報を取り入れるため，ユーザはシステムにキーワードを入力するだけで，システムがユーザの興味拡大に必要な情報を抽出し，ユーザへ提示する．つぎに，ユーザの次の興味の対象を予測するために，Web 閲覧履歴および E-mail 履歴に基づき，現在のユーザの興味をモデル化する．そして，情報提示の内容として，ユーザの興味に関連し，ユーザが思いもよらぬ事柄を，システムが予測し，出力する．

現状の問題点を明確にした結果，興味拡大における問題の 1 つに,「ユーザが次の興味の対象を予測することができない」という項目があった．システムがユーザの次の興味の対象を予測するため，新たに興味を抱くきっかけを洗い出し，分類した．その結果，

- 自分の興味に関連する事柄
- 自分の感性と一致する事柄
- 所属するグループ内の共通の事柄
- 社会的に重要な事柄
- 予想外の事柄
- 予想通りの事柄

の 6 種類に分類することができた．そのため，本研究では，この分類に基づき，ユーザの興味に関連する単語を抽出し，興味の拡大を支援する．新たに興味を抱くきっかけの分類およびこの分類に基づいたシステムの設計については，3章で詳しく述べる．

以上に基づき，本研究では，

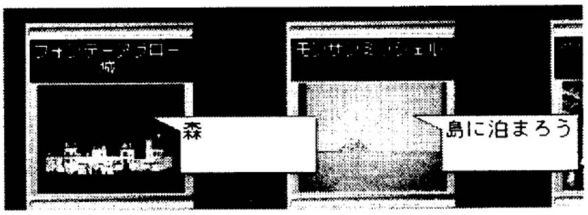


図 1.8: 吹き出し表示 [15]

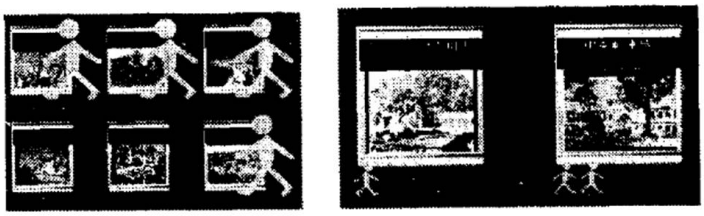


図 1.9: 人型表示 [15]

- Web 閲覧履歴および E-mail 履歴を用いて，ユーザの興味モデルを構築する
- 新たに興味を抱くきっかけの分類に基づき，インターネット上の情報から，ユーザの興味に関連する単語を抽出する
- 抽出された情報を，マインドマップのようにユーザへ提示する

という 3 つの機能を有したシステムを構築することで，ユーザの興味拡大を促し，ユーザの思考を支援することを目的とする．

本稿は，図 1.7に示すように，本章を含めて 7 章で構成される．2 章では，要求分析について述べる．この章では，提案するシステムの特徴を明らかにするための，興味拡大における現状の問題点の分析および本研究で提案するシステムが生み出す価値の分析について述べる．

3 章では，提案するシステムの設計について述べる．この章では，提案するシステムの全体像およびシステムを構成する 3 つのサブシステムについて述べる．

4 章では，提案するシステムの構築について述べる．この章では，提案するシステムの機能を満たすために行う処理およびシステムの実装について述べる．また，それぞれのサブシステムの動作確認について述べる．

5 章では，提案するシステムの検証について述べる．この章では，ユーザの興味を把握するための情報源の解析精度の検証および提案するシステムを構成する 3 つのサブシステムの有効性の検証について述べる．そして，その結果について考察し，それらが意味するところを述べる．

6 章では，提案するシステムの評価について述べる．この章では，提案するシステムと既存のシステムを比較して行った官能評価実験について述べる．そして，それらの結果について考察し，提案するシステムの有効性を述べる．

7 章では，本研究で述べた内容および本研究で得られた成果について総括し，今後の展開について述べる．

1.3 従来研究

本節では，本研究に関連する従来研究について述べる．1.3.1項では，興味拡大を目的とした情報推薦システムである，興味喚起型システムについて述べる．1.3.2項では，ユーザが入力した単語に関連する単語を出力する連想検索型システムについて述べる．1.3.3項では，新たな気づきを与えるという意味で興味拡大と関連すると思われる創造性支援型システムについて述べる．

1.3.1 興味喚起型システム

宮原らは，ユーザのにぎわい感を利用した興味喚起型のコンテンツガイドシステムを提案した [15]．このシステムでは，コンテンツ探索には，ユーザ自身の興味だけでなく，他のユーザの振る舞いおよび発言が影響することに着目している．ここで，ユーザが満足するコンテンツの発見につながる他のユーザの存在をにぎわいと定義し，宮原らのシステムでは，ユーザ同士の会話によるにぎわいと人だかりによ

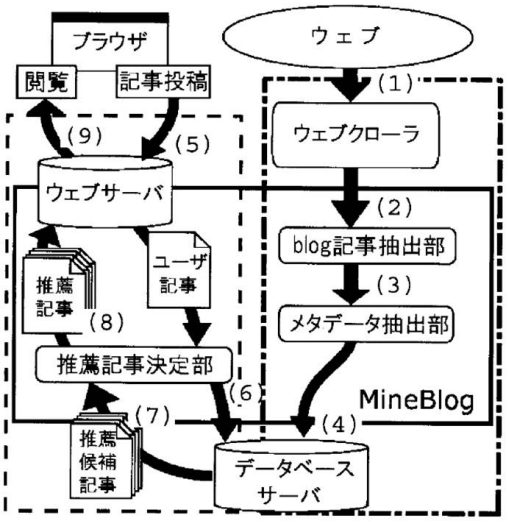


図 1.10: MineBlog のシステム概要 [16]

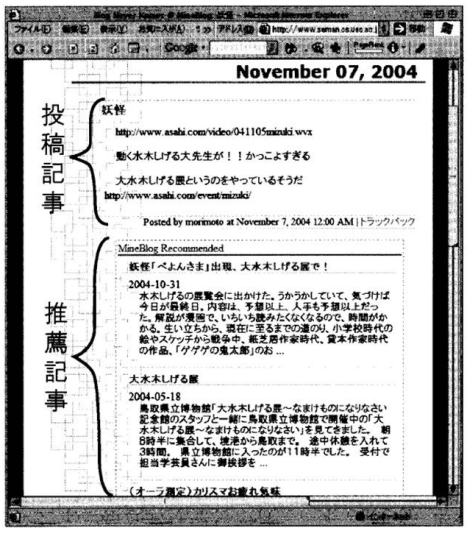


図 1.11: MineBlog を用いた推薦例 [16]

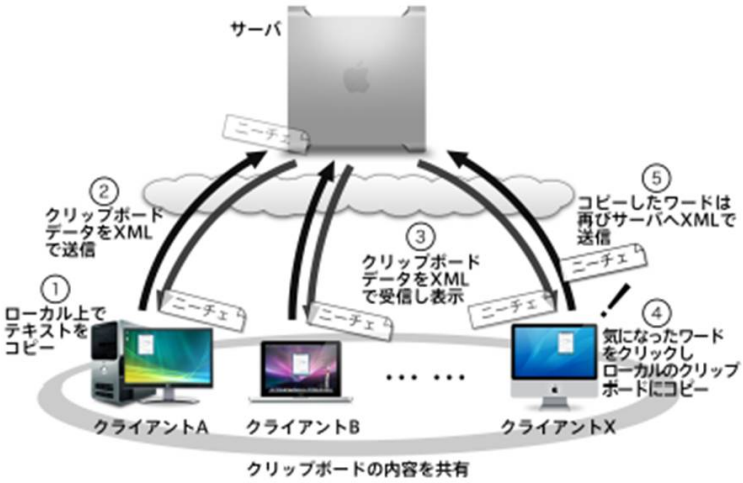


図 1.12: クリップボード共有システムの概要 [17]

るにぎわいの 2 つについて検討している．会話によるにぎわいの表現は，1 つの掲示板を用意し，その掲示板内にある記事のタイトルを，図 1.8に示すような吹き出しを用いて，会話的に表現する．人だかりによるにぎわいの表現は，コンテンツのアクセス情報からランキング情報を算出することで，コンテンツの注目度を表現し，図 1.9に示すように，人の形を模倣したグラフィックを表示することで，人だかりによるにぎわいを表現する．このように，各コンテンツの他のユーザの注目度を示すことにより，ユーザに興味喚起を促す．このシステムでは，ユーザ自身の興味だけでなく，他のユーザの振る舞いや発言が影響するということに注目しているため，ユーザのモデルは構築せず，他のユーザの注目度のみでユーザの興味喚起を促している．もちろん，他人の言動および注目度といったものもユーザの興味に影響を与える．しかし，ユーザの感性および今までの経験といったものもユーザの興味に影響するため，ユーザモデルを考慮しないことは，ユーザの興味喚起の効果を薄れさせると考えられる．また，このシステムでは，ユーザに多くのコンテンツを閲覧させることを目的とし，ユーザが次に興味を持つであろう興味の対象を，明確に提示していない．そのため，結局多くのコンテンツを閲覧しなければいけないため，効率的に興味の対象を発見することができるとは考えにくい．

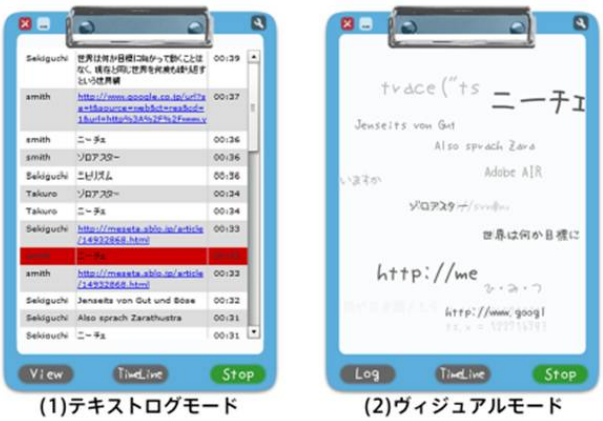


図 1.13: データの表示方法 [17]

森本らは，興味発見を支援する blog 記事推薦システム，MineBlog を提案した [16]．このシステムでは，ユーザが書いた文章と，関連性，相違性および話題性が高い文章をユーザに提示することで，ユーザの新たな興味の発見を促す．MineBlog は，図 1.10に示すシステム構成を持つ．このシステムは，以下の手順で処理を行う．

1. ウェブクローラを用いて，blog ページを収集する
2. blog 記事抽出部で，blog ページを記事単位に分割する
3. メタデータ抽出部で各記事に対し，メタデータとして記事中に存在する単語と記事の書かれた日付を抽出する
4. 記事とそれに付随するメタデータをデータベースに格納する
5. ユーザが投稿した記事をウェブサーバーから受け取る
6. ユーザが投稿した記事から特徴語を抽出し，データベースサーバへ送信する
7. データベースより推薦候補記事を抽出する
8. 推薦記事決定部で推薦記事を決定し，ウェブサーバへ送信する
9. ユーザの党記事の下に推薦記事を表示し，ユーザは推薦記事を閲覧する

以上の処理を行い，図 1.11のように blog 記事が推薦される．以上のように，ユーザは，MineBlog を用いて，新たに文章を投稿することにより，新たな興味の発見を効率的に行うことができる．しかし，MineBlog では，ユーザが新規に投稿した記事だけしか見ておらず，過去に投稿した記事の情報を利用していない．そのため，ユーザの興味の変遷などを追うことができない．また，ユーザが投稿した記事との相違性の重要性を述べているが，この相違性は，ユーザが書いた文章以外の単語がどのくらい含まれているかということで決まる．そのため，ユーザが書いた文章が非常に短く，単語数が少なければ，文

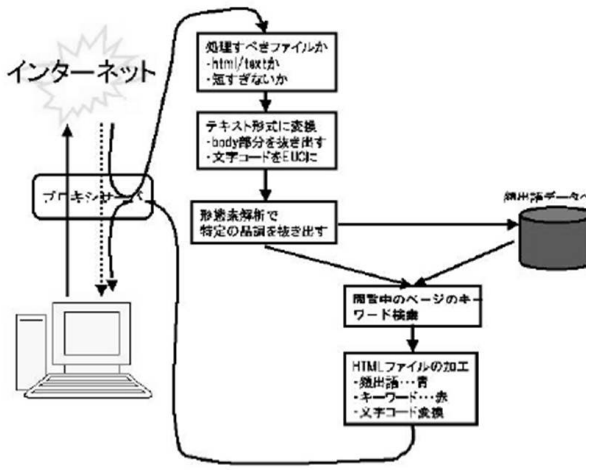


図 1.14: ブラウジング支援システムの概要 [18]

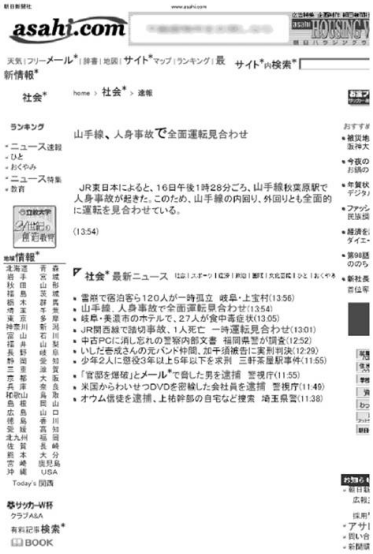


図 1.15: ブラウジング支援例 [18]

章が長い記事との相違性が高く出てしまう．したがって，本来はユーザにとって既知の情報にもかかわらず，文章を書くことを怠ったために，相違性が高いとして，表示されてしまう可能性がある．すなわち，ユーザが書いた文章の質により，blog 記事の推薦精度が大きく変動してしまう．相違性に関しては，単に記事に未出現の単語の含有率を用いるのではなく，例えば，サッカーというスポーツから，似ているフットサルというスポーツを推薦するといったように，ある概念では共通しているが，本質は異なるといった相違性を考慮すべきである．さらに，MineBlog では，blog 記事の推薦を受けるために，ユーザ自らが記事を投稿しなければならない．そのため，ユーザにとって，推薦を受けるための負荷が高いといえる．

関口らは，図 1.12に示すリアルタイムなクリップボード共有による関心拡張システムを提案した [17]．このシステムでは，コンピュータ上で頻繁に利用されるコピーコマンドによるクリップボードの収集を通して，それをユーザ間で共有することにより，リアルタイムにユーザ同士の興味および関心の共有を行う．システムの流れを以下に述べる．まず，ユーザがある単語をクリップボードへコピーを行う．つぎに．そのコピーした単語をサーバへ送信する．各ユーザには，図 1.13に示すように，その単語が表示され，自分のコンピュータのクリップボードへ単語をコピーし，そのまま検索エンジンのキーワードといったものに用いることができる．その結果，他者の興味および関心を知覚することによって，自分の中にある潜在的な興味および関心の拡張を促す．以上のように，他人がクリップボードへコピーした単語を利用することにより，ユーザの興味および関心を拡張させることができる．しかし，この研究では，クリップボードへコピーした単語が，ユーザの興味および関心を表すものであるという議論はされていない．コンピュータの使用中にコピーアンドペーストを使用する場合，多くは手間を省くために行われていると思われる．例えば，検索エンジンを用いてある単語を調べようとユーザが思って，その単語をクリップボードへコピーする場合は，その単語はユーザの興味を表している可能性が高い．しかし，文書作成の際に，文章をコピーし，その文章を再利用するという場合には，ユーザの興味を表しているとは考えにくい．そのため，他のユーザが，そのコピーされた文章を閲覧しても，興味を喚起するとは考えにくい．

福田らは，ユーザ個人の閲覧履歴からのキーワード抽出によるブラウジング支援システムを提案した [18]．このシステムでは，ユーザが閲覧している文章からユーザの興味を推定し，興味に関連する単語をハイライトすることによりユーザに対し，ブラウジング時の興味支援を行う．このシステムでは，図 1.14に示す処理を行う．まず，ユーザが閲覧するページの内容に応じて，処理すべきファイルであるかどうかを判断する．つぎに，Web ページの Body 部分から Web ページの文章を抽出し，文字コードを変換する．また，抽出された文章を形態素解析を用いて，特定の品詞を抜き出し，抽出キーワードを頻

図 1.16: 意外な Web ページの提示例 [19]

出語データーベースへ送る．さらに，今まで保存されていたユーザにとっての頻出語の上位 30 %の語との共起の偏りを計算し，15 語のキーワードを決定する．最後に，ユーザが閲覧しているページの中から決定されたキーワードを検索し，そのキーワードがあった場合は，文字の色などを変化させることにより，ユーザへキーワードを提示する．その結果，図 1.15のように情報が提示される．以上のように，頻出語との共起を計算し，ユーザが新たに興味を抱く可能性のあるキーワードをユーザへ提示する．このシステムでは，すでにユーザが閲覧している Web ページの中からキーワードを抽出し，提示する．そのため，提示されるキーワードは，ユーザが閲覧中の Web ページの中の単語のみとなる．すでにユーザが閲覧しているページは，ユーザにとって何かの目的があり，そのページを閲覧しているため，そのページ中の単語をハイライトしても，興味拡大支援の効果が薄いと考えられる．ユーザの興味拡大を支援するには，ユーザが予想できない語句を提示することも必要である．すなわち，システムを用いることによって，ユーザが閲覧する Web ページを決定するための検索キーワードとして思いつかないような語句を提示することが，ユーザの興味拡大支援には有効であると考えられる．

斎藤らは，図 1.16に示す，意外な Web ページと出会うきっかけを与えるエージェントシステムを提案した [19]．このシステムでは，電子メールという Web ページの閲覧履歴以外で得られた興味を Web ページの検索に用いる．その結果，Web ページ検索時にはユーザにとって思いつきにくい，意外な Web ページを提示する．このシステムは，興味抽出部および推薦情報提示部の 2 つのサブシステムを有する．興味抽出部では，ユーザのメールを用いてユーザの興味を把握し，Web ページを検索する．推薦情報提示部では，興味抽出部で検索された Web ページを特定のタイミングで提示する．2 つのサブシステムの処理を以下に述べる．

- 興味抽出部

  1. Topic 化
     まず，送信フォルダ内の全メールから，送信頻度の低い相手を除いた相手のアドレスリストを作成する．つぎに，作成されたリスト中のアドレスを含む全メールから，同じ話題をやり取りしていると考えられるメールをグループ化する．このグループを Topic と呼ぶ．最後に，返信のための全文引用および署名部分といった部分を削除する．
  2. Topic 選抜 1
     興味がある単語を多く含む Topic を選抜するために，ユーザの送信頻度の低い Topic を除外する．
  3. Topic 選抜 2
     理由として適切な Topic を選抜するために，送信単語数の少ない Topic を除外する．
  4. 単語の抽出
     まず，これらの Topic に対して形態素解析を行い，名詞を抽出する．つぎに，tf-idf 値に基づく式を使い，単語の重みを計算する．tf-idf 値とは，Salton らが提案した，ベクトル空間モデルで用いる値である [20]．文書 D のベクトルを，$\boldsymbol{D} = (d_1, d2, d3, ..., dn)$ とするとき，このベ

図 1.17: QueReSeek 実行時の Web ブラウザ画面 [21]

図 1.18: 検索キーワード提示画面 [21]

クトルの各成分である単語 i の重み d_i は,

$$d_i = \log(tf_i) * \log(N/df_i) \tag{1.1}$$

で表される．ここで，tf_i は，単語 i が文書 D 内に出現する回数，df_i は，単語 i が出現する対象文書数，N は総文書数を表す.

5. 推薦用 Topic 選抜
   最近 1 週間に送受信したメールの内容を現在の興味とみなし，手順 2 で選抜された Topic と類似度が最大となる Topic を選択する．
6. Web ページ検索
   類似度により選択された Topic に含まれている単語のうち，重みが高い上位 2 つを検索語として検索エンジンへ送る．その結果，Topic の関連情報として，20 個の URL を取得する．

- 推薦情報提示部

1. 推薦情報提示
   システムの起動中，メーラがアクティブになるタイミングで，情報を提示する．提示する情報は，興味抽出部で取得した Web ページのタイトルと，なぜその Web ページを提示したかを示す理由である．
2. Web アクセス
   ユーザが，提示された Web ページタイトルをクリックし，提示された Web ページへアクセスする．

以上のように，斎藤らが提案したシステムは，メールから，ユーザの興味を把握し，ユーザへ Web ページを提示する．このシステムでは，ユーザとのインタラクションなしに自動的にユーザの興味を把握し，その結果から自動的にユーザへ提示する情報を決定する．そのため，ユーザの興味モデルが変化しない限り，ある一定の推薦結果しかでてこない．興味拡大には，様々な情報が必要なため，ユーザとシステムがインタラクションを行い，推薦結果を大きく変化させることが有効であると考えられる．

以上に述べたことから，本研究で提案するシステムは，従来提案された興味喚起型システムと以下の 3 点で異なる．

- ユーザの興味モデルに利用するソースが Web 閲覧履歴および E-mail 履歴である

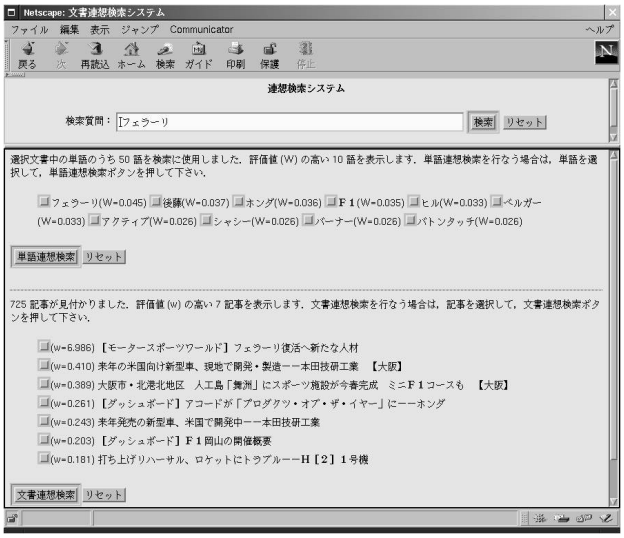


図 1.19: GETA の実行例 [22]

- ユーザは，システムにキーワードを入力し，インタラクションを行うことで，情報の提示を受ける
- ユーザが Web ページを閲覧しなくとも，情報の提示を受けることができる

1.3.2 連想検索型システム

連想検索とは，文章群同士の類似度を計算することで，文章の関連性を求める検索システムである．本研究とは，あるキーワードに関連する事柄を出力するという点が類似している．本項では，従来提案されてきた連想検索型システムについて述べる．

丹らは，図 1.17および図 1.18に示す，検索履歴共有によるコミュニティ指向の連想検索である QueReSeek を提案した [21][23]．QueReSeek の流れを以下に述べる．まず，ユーザが属するコミュニティ内で用いられた検索キーワードを共有する．つぎに，各検索キーワードにより得られた Web ページ群を，検索キーワードと Web ページの関係として，データーベースに保持する．この関係に基づき，ユーザが閲覧している Web ページから，関連する検索キーワードを求め，ユーザへ提示する．その結果，閲覧中の Web ページに関連する新たな着想を与える．このシステムでは，ユーザが閲覧している Web ページに関連したキーワードが出力されるため，Web ページを閲覧している必要がある．本研究で提案するシステムは，ユーザが閲覧する Web ページを見つけていない場合でも使用することができる．また，このシステムは，検索を不得意とするユーザによるキーワードが思いがけない結果となって，検索を得意とするユーザへ気づきを与える．すなわち，一部のユーザにしか気づきや新たな着想を与えることができない．本研究で提案するシステムは，すべてのユーザに対して，新たな興味喚起を促す．

高野らは，図 1.19に示す，汎用連想計算エンジン Generic Engine for Transposable Association （以下，GETA）を提案した [22][24][25]．GETA では，文書-単語行列を元に tf-idf 値など様々な統計的類似性計算を行うことで，文書間および単語間の関係を求めることができる．GETA の連想計算では，図 1.20に示す Word-Article Matrix(以下，WAM) と呼ばれる行列を用いて，計算を行う．WAM は，行には文書，列には索引語が格納されている．また，行列の要素は整数値であり，文書に含まれる索引語の頻度を示す．GETA は，図 1.21に示すように，文書から単語を検索するステップ（以下，文書-単語連想検索）とその検索された単語から文書を検索するステップ（以下，単語-文書連想検索）の 2 つのステップを行う．文書-単語連想検索では，図 1.22に示すように，まず，与えられた文書群中に少なくとも 1 回出現するようなすべての単語について，それぞれの文書中での頻度に基づき，スコアを計算する．つぎに，それらを集計し，スコアの大きい順に上位数個をスコアとともに出力する．これは，文書群から内容的に要約

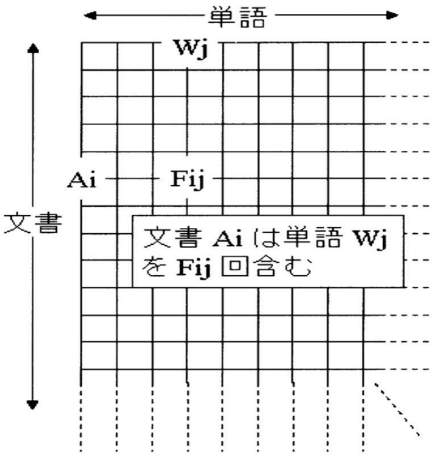


図 1.20: WAM[24]

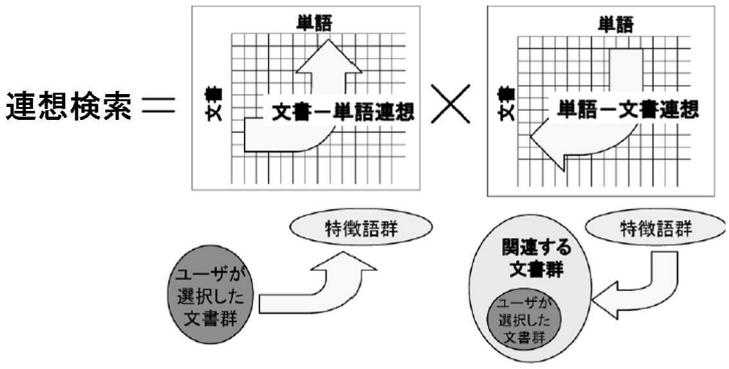


図 1.21: 連想検索の流れ ([25] を改変)

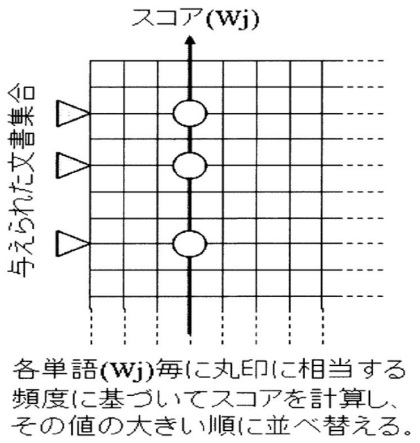


図 1.22: 文書-単語連想検索の模式図 [24]

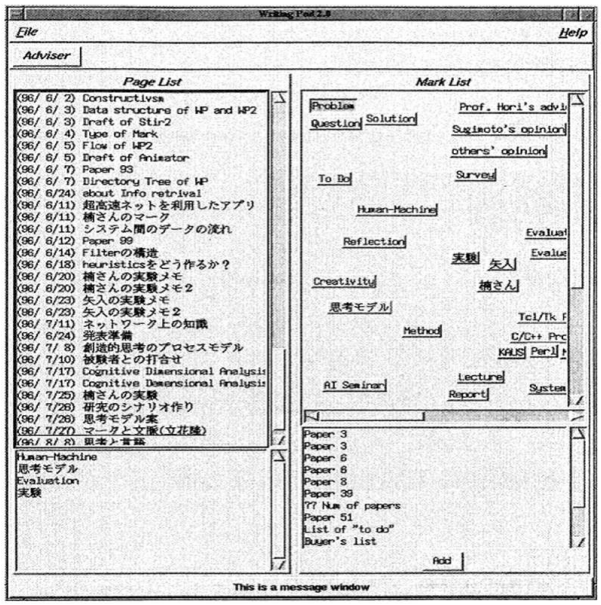


図 1.23: 記憶の想起に基づく創造性支援システム [26]

する特徴語群を抽出することを意味する．単語-文書連想検索では，文書-単語連想検索で抽出された特徴語群を用いて，文書方向へスコアを計算し，関連文書群を抽出する．以上のように，GETA は，ある文書から特徴語を抽出し，それに関連する文書を検索することができる．また，文書中の特徴語の頻度が少ないものを抽出することで，ユーザへ新たな気づきを与える可能性がある．しかし，このシステムは，どのユーザが検索を行っても，ある文書に関連した単語は変わらない．ユーザの興味拡大には，ユーザの興味を考慮して，情報を提示することが有効であると考えられる．

以上に述べたことから，本研究で提案するシステムは，従来提案された連想検索型システムと以下の 2 点で異なる．

- 一部のユーザでなく，システムを使用する全てのユーザに対して，興味拡大を支援する
- システムが，ユーザの興味を把握するために，ユーザの興味モデルを構築する

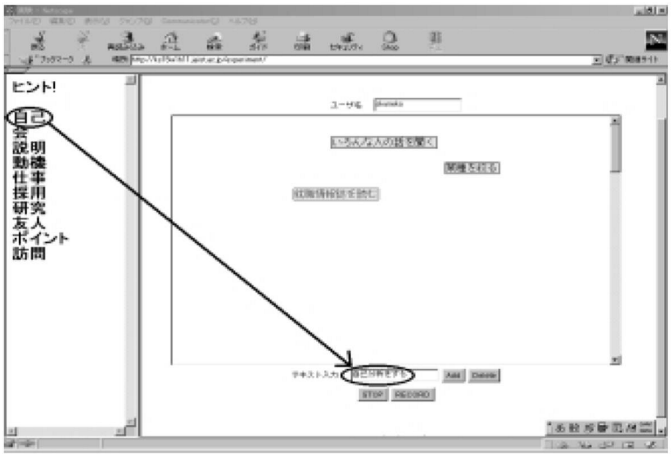


図 1.24: 発想のヒントを提示するシステム [27]

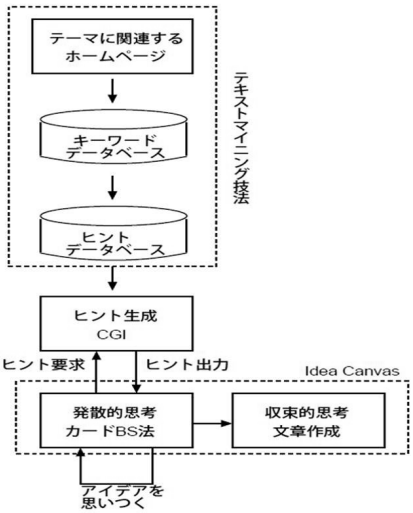


図 1.25: 発想システムの構成 [27]

1.3.3 創造性支援型システム

創造性支援とは，ユーザの思考の制約を変更し，ユーザに新たな気づき，新たな考え方および新たな物の捉え方といったものを与えることである．創造性支援は，興味拡大には焦点を当てていないが，ユーザへ新たな気づきを与えるという点で，本研究と類似している．本項では，従来提案されてきた創造性支援型システムについて述べる．

相原らは，図 1.23に示す，記憶の想起に基づく創造性支援システムを提案した [26][28]．このシステムは，ユーザが過去に蓄積したメモを見せ，過去の記憶を想起させることで，新たな概念の創発を促す．このシステムの流れを以下で述べる．まず，紙のノートと同じ利便性を持つ仮想的なノートを用意する．つぎに，ユーザは，その仮想的なノートに，研究メモや参考文献などをシステムに入力する．また，それぞれのメモを特徴づけるために，各メモに任意のインデックスを付ける．さらに，ユーザにメモ間の関連度を見せたり，メモの検索などを行わせたりする．その結果，ユーザが忘却したメモおよび想起されにくいメモを探し出すことができ，新たなの思考を促すことができる．以上のように，このシステムでは，ユーザが残した過去のメモを，ユーザに見せることによって，新たな思考を促す．すなわち，このシステムは，ユーザが過去に興味を抱いた単語をもう一度提示していることになり，新たな情報提示を行っているわけではない．新たな概念の創発やユーザの興味拡大には，過去に興味を抱いた単語を出力するのみでなく，興味に関連した単語を新たに提示することも必要であると考えられる．

金子らは，図 1.24に示す，テキストマイニング技法を活用した発想支援システムを提案した [27]．このシステムは，ユーザの発想が行き詰ったときに，ヒントを与え，新たな発想を促す．このシステムは，図 1.25に示すシステム構成を持つ．まず，あるテーマに関するヒントを取得するため，テーマに関連するホームページを人手で収集する．つぎに，収集したホームページのテキストに対して，形態素解析を行い，Apriori アルゴリズムを用いて相関ルールを導出し，ヒントデータベースを構築する．Apriori アルゴリズムとは，IBM アルマデン研究所の Rakesh Agrawal らが提案したアルゴリズムである [29]．このアルゴリズムは，相関ルールを高速に生成するためのアルゴリズムであり，相関ルールを導出するための代表的なアルゴリズムとして広く利用されている．ユーザの発想が行き詰った際には，ユーザはヒントを生成してほしいキーワードを入力する．その結果，ヒントデータベースから，入力したキーワードと相関のある単語を抽出し，ユーザへ提示し，ユーザの発想を促す．以上のように，このシステムは，ユーザが入力したキーワードと関連する単語を提示することにより，ユーザの発想を支援する．このシステムは，興味拡大支援ではなく，発想支援を行うため，ユーザのモデルを構築していない．また，関連するキーワードを取得するための源となるホームページは，人手で収集している．そのため，収集できるホームページは限定されてしまい，ユーザにとって本当に有効なヒントを提示することができない

可能性がある．

以上に述べたことから，本研究で提案するシステムは，従来提案された創造性支援型システムと以下の 2 点で異なる．

- ユーザにとって既知である過去の情報を提示するだけでなく，ユーザにとって未知の情報を提示する
- 限定されたホームページからでなく，インターネット上に存在する情報の中から，ユーザの興味に関連する情報を抽出する

第2章　要求分析

本章では，本研究で提案するシステムに至る要求分析ついて述べる．本研究では，要求分析として，興味拡大の現状分析および本研究で提案するシステムの価値分析を行った．2.1節では，現状分析について述べ，興味拡大における現状の問題について明らかにする．2.2節では，価値分析について述べ，本システムの有用性および本システムが今後生み出す価値について述べる．

2.1　興味拡大における現状の問題点

本研究では，興味拡大に関わる現状分析を行った．本節では，興味拡大における現状の問題点について述べる．

現在，ユーザが興味を拡大するためには，

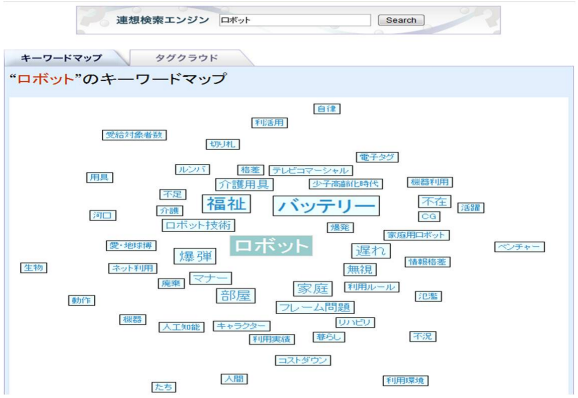
- ネットサーフィン
- テレビおよびラジオの視聴
- 本および雑誌の閲覧
- 人との会話
- 新聞の閲覧

といった手段が挙げられる．また，1.3節で述べたように，興味拡大を目的としたいくつかの情報推薦システムも提案されているが，現状では実用されていない．さらに，実用されている連想検索エンジンとして，図2.1(a)および図2.1(b)に示す，Preferred Infrastructure が開発した「reflexa」[30]，図2.2(a)および図2.2(b)に示す，野村総合研究所（以下，NRI）が開発した「ASSOCIE」[31] および図2.3(a)および図2.3(b)に示す，国立情報学研究所が開発した「想 IMAGINE BOOK SEARCH」[32] が挙げられる．

(a) reflexa のトップページ　　(b) reflexa での検索結果例

図 2.1: 連想検索エンジン reflexa[30]

(a) ASSOCIE のトップページ



(b) ASSOCIE での検索結果例

図 2.2: 連想検索エンジン ASSOCIE[31]

(a) IMAGINE BOOK SEARCH のトップページ

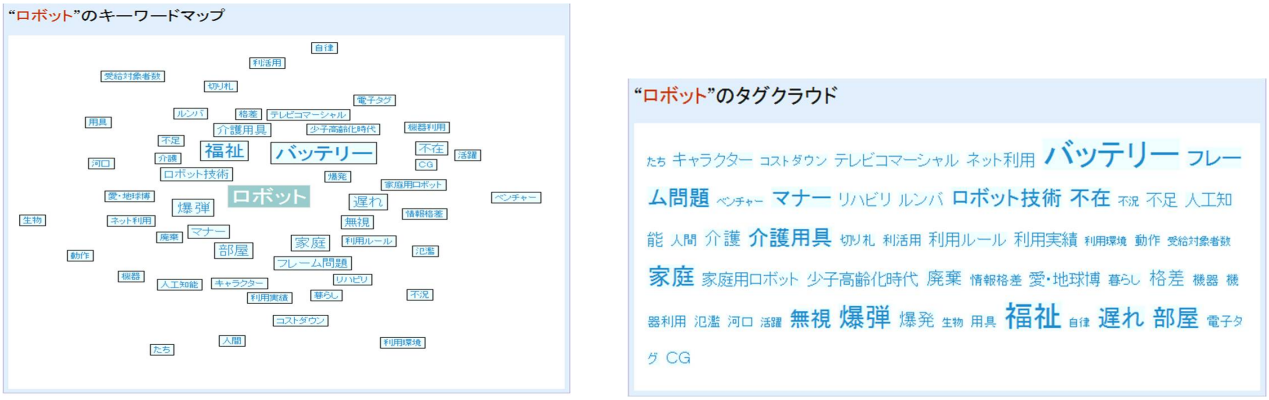
(b) IMAGINE BOOK SEARCH での検索結果例

図 2.3: 想 IMAGINE BOOK SEARCHE[32]

「reflexa」は，ユーザが入力したキーワードと関連の深い単語を自動的に抽出し，表示する連想検索エンジンである [30]．また，図2.1(b)に示すように，連想検索の結果以外に，入力したキーワードを用いて Google で検索した結果および入力したキーワードに関連する Wikipedia[33] の情報を表示する．「reflexa」では，Wikipedia およびアンサイクロペディア [34] の情報に基づき，単語間の関連を学習し，連想検索を行う．

「ASSOCIE」は，ユーザが入力したキーワードから連想される話題を自動的に抽出し，表示する連想検索エンジンである [31]．また，図2.4(a)および図2.4(b)に示すように，キーワードマップおよびタグクラウドの 2 種類の結果表示方法がある．キーワードマップでは，中心に入力された検索キーワードを表示し，その周囲に結果が表示される．タグクラウドでは，結果がリスト形式で表示される．「ASSOCIE」では，2007 年以降に刊行された NRI 出版物に基づき，連想辞書を作成し，単語間の関連を定義している．

「想 IMAGINE BOOK SEARCH」は，ユーザが入力したキーワードに関連する書籍を抽出し，表示する連想型書籍検索エンジンである [32]．「想 IMAGINE BOOK SEARCH」は，連想検索エンジン GETA を利用しており，図書館の蔵書データベース（以下，DB），神保町古書店の在庫 DB，テーマ別新書ガイド，百科事典，博物館収蔵品 DB，大学図書館 DB およびジュンク堂書店の在庫 DB といった様々なジャンルの情報源から，関連する情報を収集する．



(a) マップ表示　　(b) タグ表示

図 2.4: ASSOCIE の表示方法 [31]

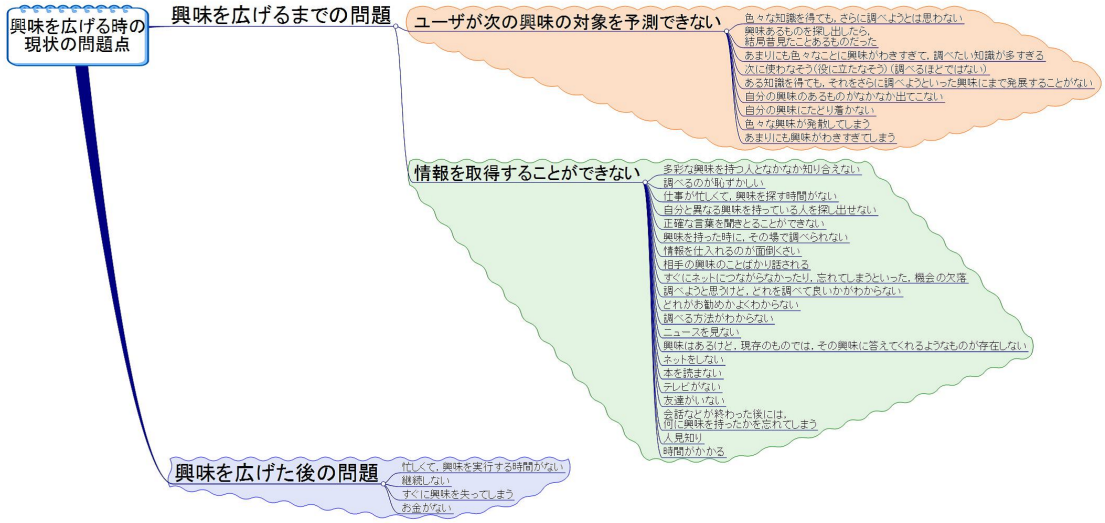


図 2.5: 興味拡大における現状の問題点の分類

これらの連想検索エンジンは，実用化されているものの，実際に使用し，興味拡大の効率化を図る人は少ない．したがって，本研究では，現状で興味拡大のために用いられている，

- ネットサーフィン
- テレビおよびラジオの視聴
- 本および雑誌の閲覧
- 人との会話
- 新聞の閲覧

という 5 つの手段を用いた興味拡大における問題点を明らかにする．

本研究では，インタビューおよびブレインストーミングを行い，問題点を洗い出した．インタビューは，1 対 1 の形式で，3 人にインタビューを行い，興味拡大における問題点を質問した．ブレインストーミングは，3〜5 人の少人数の集団で，30 分の時間制限を設けて行った．ブレインストーミングは，数回行った．その結果，34 個の問題点が挙がった．これらの問題点を分類した結果，図 2.5に示すように，

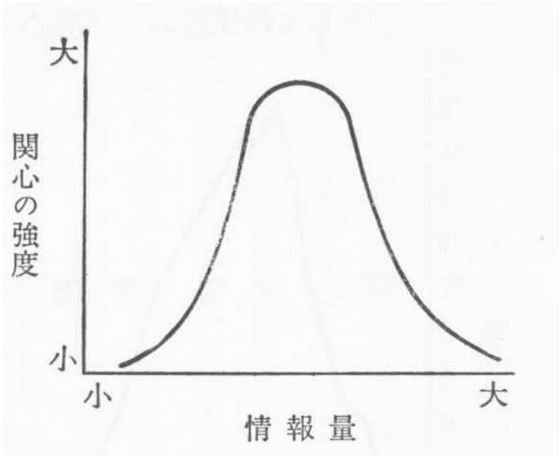


図 2.6: 情報量と人の関心の関係図 [35]

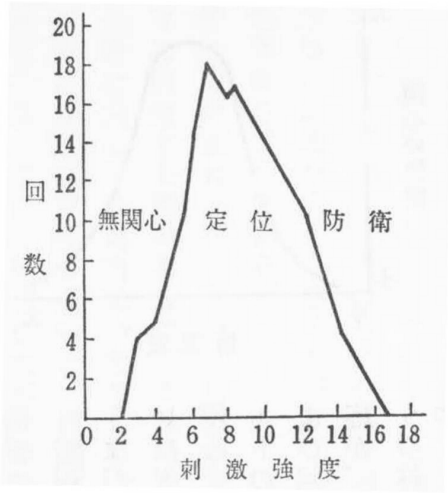


図 2.7: 刺激強度と人の関心の関係図 [35]

- 興味を広げるまでの問題
  - 情報を取得できない
  - ユーザが次の興味の対象を予測できない
- 興味を広げた後の問題

の 2 つに分類することができた.「興味を広げるまでの問題」には,「情報を取得できない」および「ユーザが次の興味の対象を予測できない」という 2 つの問題が含まれる.「情報を取得できない」という分類には,「本を読まない」や「ニュースを見ない」といった情報量に関わる問題が含まれる．また,「ユーザが次の興味の対象を予測できない」という分類には,「興味のあるものを探し出したら昔見たものだった」や「あまりにも興味を抱きすぎてしまい，結局興味をひかない」といった興味の対象に関わる問題が含まれる．さらに,「興味を広げた後の問題」という分類には,「お金がなくて興味あることを行うことができない」や「忙しくて興味あることを行うことができない」といった興味の実行に関わる問題および「すぐに興味を失ってしまう」といった興味の継続に関わる問題が含まれる．

本研究では，興味拡大に焦点を当てるため，興味の実行および興味の継続に関わる問題である,「興味を広げた後の問題」という分類は，対象外とする．したがって，本研究では,「興味を広げるまでの問題」を対象とし,「情報を取得できない」および「ユーザが次の対象を予測できない」という 2 つの問題を改善する．

ここで，東らによると，情報量と関心の間には，図 2.6に示す関係がある [35]．少なすぎる情報量および多すぎる情報量に対しては，人は関心を示さない．すなわち，人の関心を引くためには，適度な情報量が必要である．また，刺激強度と関心の間には，図 2.7に示す関係がある [35]．情報量と同様に，弱すぎる刺激強度および強すぎる刺激強度に対しては，人は関心を示さない．すなわち，人の関心を引くためには，適度な刺激強度が必要である．このように，人の関心には，情報量および刺激強度が関連する．したがって，人の関心には，対象との量的関係および質的関係が存在するといえる．

前述したように,「興味を広げるまでの問題」には,「情報を取得することができない」および「ユーザが次の興味の対象を予測することができない」という問題があることがわかった．これらの問題は，興味と対象との量的関係に関わる問題および質的関係に関わる問題であるといえる．関心は興味の一種であると考えられるため，興味にも，関心と同様に，対象との量的関係および質的関係が存在すると考えられる．したがって，本研究で行った，現状での興味拡大における問題点の分類は妥当であると考えられる．

以上のように，本研究で対象とする，現状における興味拡大における問題点は,

- 情報を取得できない

- ユーザが次の興味の対象を予測できない

という 2 つがあることがわかった．したがって，本研究では，これらの問題点を改善するため，

- 効率よく外部の情報を取り入れる
- ユーザの次の興味の対象を予測する

という 2 つの特徴を有したユーザの興味拡大支援システムを構築する．ユーザが効率よく外部の情報を取り入れるため，ユーザはシステムにキーワードを入力するだけで，システムがユーザの興味拡大に必要な情報を抽出し，ユーザへ提示する．また，ユーザの次の興味の対象を予測するために，現在のユーザの興味をモデル化する．そして，情報提示の内容として，ユーザの興味に関連し，ユーザが思いもよらぬ事柄を，システムが予測し，出力する．

2.2 提案するシステムが生み出す価値

本研究では，ユーザの興味拡大を図り，ユーザの思考を支援することに焦点を当てる．本節では，本研究で提案するシステムによりもたらされる価値について述べ，本システムの対象ユーザが多岐にわたることを示す．

人が友人などとコミュニケーションをする場合

人がコミュニケーションを行う場合，1 つの話題のみを用いて，コミュニケーションを行うことは少なく，多くの場合，複数の話題を用いて，コミュニケーションを行う．その際，自分だけが興味を抱いている話題を行うのではなく，所属するグループの全員が興味を抱く話題を提供することが，一般的であると考えられる．

所属するグループが興味を抱く可能性がある話題を推測し，選定することが困難である場合，本研究で提案するシステムを用い，グループメンバーの興味モデルから，新たに興味を抱く可能性のある話題を検索する．その結果，効率よく，適切な話題を選定することができる．

研究者が新たな研究アイデアを発見する場合

研究者の研究アイデアは，様々な場面から生まれてくる．例えば，今までの研究の中で，新たな疑問がわいたものは，新しい研究アイデアとなる．また，今まであると便利だと感じていたシステムを新たに開発したいと思うことも，新たな研究アイデアになる．このように，研究者の新たな研究アイデアは，様々なきっかけで浮かんでくる．しかし，全ての研究者が，常に新たな研究のアイデアを有しているとは限らない．さらに，研究は，自分が全く興味のわかない事柄に関して深く研究し，良い結果を出すことは難しい．新たな研究アイデアは，研究者の興味に即したものが適切である．

本研究で提案するシステムを用い，キーワードと自分の興味を照らし合わせることで，研究者の興味に即した，新たな研究アイデアを発見することができる．

教師が生徒の学習意欲を向上させる場合

近年，生徒の学力低下が問題となっている．長瀬らは，学習が成果を上げる最大の要因は学習者自身の主体的な学習意欲や態度であると述べている [36]．すなわち，生徒の学力を向上させるためには，生徒の学習意欲を向上させることが有効である．

ここで，学習意欲は，内発的動機づけと外発的動機づけの 2 種類に分けられる [37]．内発的動機づけとは,「好奇心に基づき，行動すること自体を目標とするもの」である [38]．また，外発的動機づけは,「賞または罰によって動機づけられるもので，行動は目標に到達するための手段」とするものである [38]．賞または罰を与える外発的動機づけによる学習を行うと，賞または罰がなくなっ

図 2.8: QRIO SDR-4X [39]

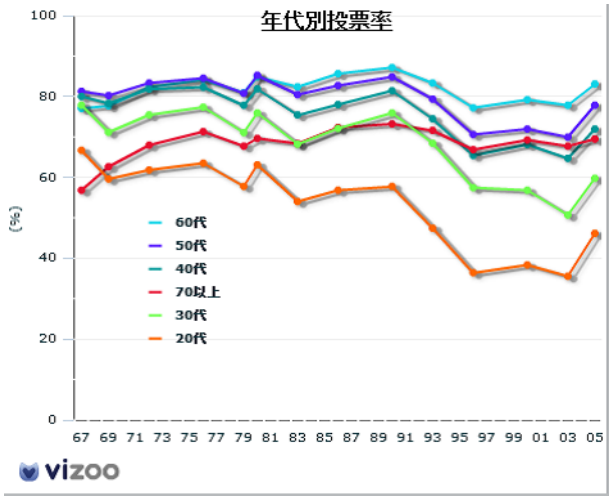


図 2.9: 投票率の推移 [40]

た時点で，生徒の学習意欲は低下する．したがって，内発的動機づけによる学習意欲の向上が有効であると考えられる．

教師が本研究で提案するシステムを用い，生徒の興味モデルと学習に関するキーワードを照らし合わせることで，生徒の興味に関連したキーワードを検索することができる．教師がそのキーワードを用いて授業を行うことにより，生徒は学習科目に対する興味を抱くと考えられる．その結果，生徒は学習科目の必要性を感じ，主体的に学習し，学力向上に有効であると考えられる．

効率的な営業を行う場合

製品およびシステムを販売する場合，どの企業でも，同じことを言い，手当たり次第に販売していくことは，効果的ではない．同じシステムでも，導入する企業によって，得られる効果は異なる．また，企業の方針によっても，求めるシステムは異なる．すなわち，製品およびシステムの販売では，企業の方針により提案するシステムを変え，それを導入した場合の効果を明確に説明することが有効である．

企業は組織であるため，企業の方針をモデル化することはできないが，組織に属している人，経営者およびシステム導入の最終的な決定権を持つ人をモデル化することにより，企業の方針を把握することができる．そのため，本研究で提案するシステムを用いて，どのシステムを提案すればよいかといったことを見つけることができる．その結果，効率的に営業を行うことができ，コスト削減や販売成績の向上へ役立つと考えられる．

効率的なマーケティングを行う場合

商品開発において，ユーザのニーズを調べるマーケティングは非常に重要である．マーケティングの方法にはさまざまなものがあるが，一般的には，インタビューおよびアンケートを用いて，既存の商品の問題点を洗い出し，ユーザの欲する商品を聞き出すことが多い．しかし，インタビューおよびアンケートには，様々な問題点がある．例えば，回答者は，現在存在していない商品を思い描くことは困難であり，全く新しい商品に関する情報は仕入れにくい．また，回答者がアンケートであるということを意識してしまい，本当のニーズを見極めることが困難である．さらに，インタビューおよびアンケートを行うためには，多くのコストがかかってしまう．

本研究で提案するシステムを用いることで，ユーザ群の興味を把握することができ，インタビューやアンケートをすることなく，ユーザのニーズを把握することが可能となる．その結果，コストをかけず，効果的な販売戦略を練ることが可能となる．

ロボットの親和性を向上させる場合

近年，計算機の高機能化に伴い，図 2.8に示す QRIO SDR-4X [39][41] に代表される，生命を模倣した様々なロボットが開発されている．これらのロボットは，今後，医療分野，教育現場およ

図 2.10: 検索結果に表示される広告

びサービスといった場において人の代替としての利用が期待されている．これらの場でロボットを利用するためには，ロボットには決められたタスクを行う能力だけでなく，人と円滑にコミュニケーションを行う能力が求められる．人と円滑にコミュニケーションを行うためには，人との円滑な会話といった，ロボットの親和性の向上が有効である．

円滑な会話には，相手に合わせて適切に話題を選定し，適度に話題を変えることが必要である．本研究で提案するシステムをロボットに実装することで，相手の興味を把握し，適切な話題を選定することができる．したがって，話の流れを切ることなく，適切な話題変換を行うことができる．その結果，ロボットの親和性を向上させることができると考えられる．

選挙の投票率を向上させる場合

近年，若者の投票率の低下が問題となっている．図 2.9に示すように，投票率は年々低下しており，特に 20 代〜30 代の投票率の低下が著しい．投票率の低下の原因の 1 つとして，政治への興味の低下が挙げられる．斎藤らは，現行の政治がわかりにくいと感じていることが政治への興味を低下させていると述べている [42]．そのため，政治をわかりやすく説明する書籍およびコンテンツが非常に多く存在する．しかし，若者は，そもそも政治に興味を抱いていないため，いくらコンテンツが存在しても，それらを閲覧する機会がない．

本研究で提案するシステムを用いることで，ユーザの興味と政治の内容を照らし合わせ，少しずつ政治へ興味を抱かせるキーワードを出力することができる．その結果，政治への興味を高め，前述したコンテンツを閲覧する機会を得ることができると考えられる．また，若者の政治への興味を高めることで，若者の投票率を向上させることができると考えられる．

Web ページの広告への応用

Google などの検索エンジンを用いて，検索を行うと，図 2.10に示す入力した検索キーワードに対応した広告が表示される．これらは，キーワード広告とよばれ，企業が契約し特定のキーワードが検索された際に，画面右側の欄に「スポンサー」という見出しとともに表示されるものである．閲覧者によってこの部分のリンクがクリックされるたびに，契約した企業から広告料が支払われる．例えば，Google のキーワード広告は特にアドワーズ（AdWords）とよばれ，同社の主な収入源となっている．現在のキーワード広告は，特定の検索キーワードを入力した際に表示される，「検索連動型広告」が主流となっている [43]．以上のように，キーワード広告は，大きな収入源となっており，閲覧者にクリックさせることが重要である．

本研究で提案するシステムを用い，ユーザの興味を把握することができれば，今まで特定の検索キーワードでしか表示されなかった広告が，検索キーワードが一致しなくても，ユーザが入力し

た検索キーワードとユーザの興味に関連した広告を表示することができる．その結果，キーワード広告を，ユーザに効率よくクリックさせ，収入の増加につながると考えられる．

以上のように，本研究で提案するシステムを応用することで，様々な分野で価値を生み出すことができる．本研究では，今後これらの価値を見出すために必要な，最も基礎的なシステムを構築する．

第3章　思考支援システムの設計

本章では，本研究で提案するシステムの設計について述べ，システムの全体像および機能を明確にする．3.1節では，本研究で提案するシステムの設計に密接に関わる，新たに興味を抱くきっかけの分類について述べる．また，本研究で提案するシステムは，ユーザモデル構築システム，外部情報整理システム，ユーザインターフェースの3つのサブシステムから構成される．3.2節では，システム全体の構造を述べ，提案するシステムの全体像を明らかにする．3.3節では，ユーザの興味を把握するための，ユーザモデル構築システムについて述べる．3.4節では，インターネット上から情報を抽出するための，外部情報整理システムについて述べる．3.5節では，ユーザインターフェースについて述べる．

3.1　新たに興味を抱くきっかけの分類

本研究では，2.1節で述べた,「ユーザが次の興味の対象を予測することができない」という問題を受け，システムがユーザの次の興味の対象を予測するために，人の新たに興味を抱くきっかけを調査した．本節では，新たに興味を抱くきっかけの調査結果および分類について述べる．

本研究では，インタビューおよびブレインストーミングを行い，新たに興味を抱くきっかけを洗い出した．インタビューは1対1の形式で，5人に行い，新たに興味を抱くきっかけを質問した．ブレインストーミングは，3〜5人の少人数の集団で，30分の時間制限を設けて行った．ブレインストーミングは，数回行った．その結果，33個のきっかけが挙がった．これらのきっかけを分類した結果，図3.1に示すように分類できた．新たな興味を抱くきっかけの分類について以下に述べる．

新たな興味を抱くきっかけは，大きく分けて，

- 自分の情報から派生
- 外部の情報から派生

という2つに分類される.「自分の情報から派生」する分類の中には，

(a) 自分の興味に関連する事柄

(b) 自分の感性と一致している事柄

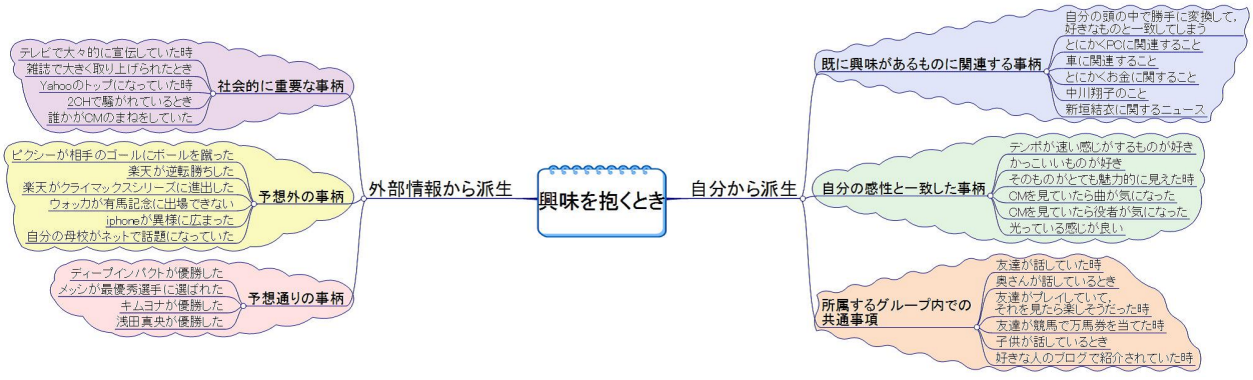


図 3.1: 新たな興味を抱くきっかけの分類

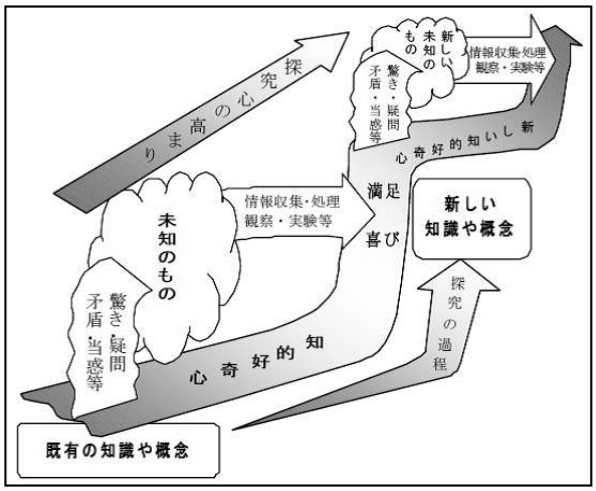


図 3.2: 知的好奇心の概念図 [44]

(c) 所属するグループ内の共通の事柄

という 3 つの事柄がある.「自分の興味に関連する事柄」には，中川翔子に関する情報全般や車に関することといった自分が既に持っている興味に関する事柄が含まれる.「自分の感性と一致している事柄」には，かっこいいものやテンポの早い感じがするものといった対象から感じられる印象と自分が有する感性が一致する事柄が含まれる.「所属するグループ内の共通の事柄」には，友達が話していた話題や友達がしていたゲームといった自分が所属するグループ内での話題や共通認識に関する事柄が含まれる.

また,「外部の情報から派生」する分類には,

(d) 社会的に重要な事柄

(e) 予想外の事柄

(f) 予想通りの事柄

という 3 つの事柄がある.「社会的に重要な事柄」には，テレビで大きく取り上げられていた事柄や Yahoo ニュースでトップになっていた事柄といった多くの人が関心を持ち，社会的に重要であると考えられる事柄が含まれる.「予想外の事柄」には,「楽天がクライマックスシリーズに進出した」や「iphone が非常に広まった」といった世間の考えていたことと逆の結果になった事柄および自分にとって予想外であった事柄が含まれる.「予想通りの事柄」には,「ディープインパクトが優勝した」や「キムヨナが優勝した」といった世間の考えていたことと同じ結果になった事柄および自分にとって予想通りであった事柄が含まれる.

ここで，従来の知見と照らし合わせることで，これらの分類の妥当性を述べる．森永らは，知的好奇心を図 3.2に示すように表した [44]．この図から，知的好奇心を抱かせるものとして，既知の知識や概念といった情報から派生するものと未知の情報から派生するものの 2 つがあることがわかる．知的好奇心は，興味の 1 種であると考えられる．したがって，本研究で分類した興味のきっかけの大分類である,「自分の情報から派生」するものと「外部の情報から派生」するものは，妥当な分類であると考えられる．

1.3.1項で述べたように，森本らによると，新たな興味を抱くためには，関連性，話題性および相違性を持つ記事が重要である [16]．この関連性，話題性および相違性は，それぞれ本研究の分類の「自分の興味に関連する事柄」,「社会的に重要な事柄」および「予想外の事柄」に相当すると考えられる．

Freitas らによると，興味研究において Surprisingness という概念が重要である [45]．この概念が提唱されたことにより，興味拡大における意外性の重要性が注目されている [19]．これらの知見は，本研究の分類の「予想外の事柄」に相当すると考えられる．

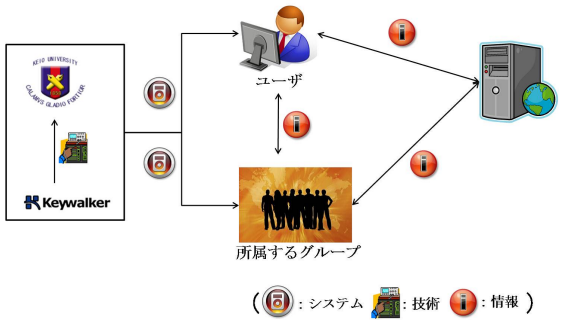


図 3.3: 興味拡大における CVCA

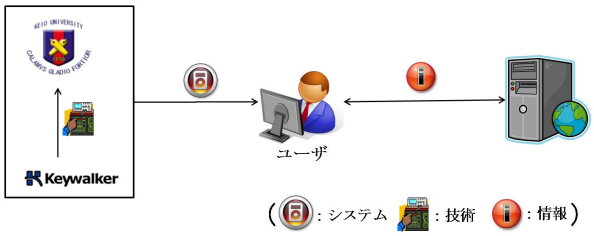


図 3.4: 本研究で想定する CVCA

宮原らの研究および関口らの研究では，所属するグループや他人の興味および関心を知ることにより，ユーザの興味が拡大すると述べられている [15][17]．この知見は，本研究の分類の「所属するグループ内の共通の事柄」に相当すると考えられる．

中小路らによると，デザイナは，椅子をデザインする際に，桃の花の写真を見ると，その画像がひらめきを与え，創造的発想を行うことができた [46]．これは，桃の花から受けた印象とデザイナが有する感性が一致し，デザインに創造的発想をもたらしたと考えられる．この研究は，発想の拡大に関する研究であるが，興味の拡大にも同じきっかけがあると考えられる．そのため，この知見は，本研究の分類の「自分の感性と一致している事柄」に相当すると考えられる．

波多野らは，人は自分の中に作り上げてきた枠組みを壊されることを嫌うと述べている [47]．すなわち，人は，自分が予想している結果を望んでいると言える．これは，本研究の分類の「予想通りの事柄」に相当すると考えられる．

以上に述べたように，本研究で分類した新たな興味を抱くきっかけは，従来の知見を統合した形であるため，妥当な分類であると考えられる．

本研究では，Customer Value Chain Analysis（以下，CVCA）を行い，ステークホルダーを明確にした．CVCA とは，製品や工程の開発において，決定権者や中間業者をはっきりさせ，誰が購入を決定するのか，複雑な市場の構造と各業者との関係を明確にするための手法である [48]．CVCA の結果，図 3.3に示すようなステークホルダーが存在することが明らかになった．本研究では，興味拡大における最も基礎的なシステムを構築するため，ユーザと所属するグループの情報のやり取りは行わないこととする．したがって，本研究では，図 3.4に示すステークホルダーを想定し，新たな興味を抱くきっかけの分類の 1 つである「所属するグループ内の共通の事柄」は，本研究では対象外とする．

また，新たな興味を抱くきっかけに,「自分の感性と一致している事柄」という分類が挙がった．感性は，美や善などの評価判断に関する印象の内包的な意味を知覚する能力である．これは，直感的なものであり，言語では表すことが困難である．本研究では，言語を用いて表現された Web ページおよび E-mail を用いてユーザの興味モデルを構築し，インターネット上の情報を用いて，ユーザの興味を拡大する．すなわち，本研究は，言語で表された情報を用いて，ユーザモデルの構築および情報の提示を行う．言語で表すことが困難な感性は，言語で表された情報を用いてモデル化するのではなく，他の情報源を用いてモデル化することが有効である．したがって,「自分の感性と一致している事柄」は，本研究では対象外とする．

以上に述べたように，本研究では，前述した 2 種類を除く，4 種類の分類に基づき，システムの設計および構築を行う．

3.2 提案するシステムの概要

本節では，3.1節で述べた新たな興味を抱くきっかけに基づき提案する，システムの全体像について述べる．

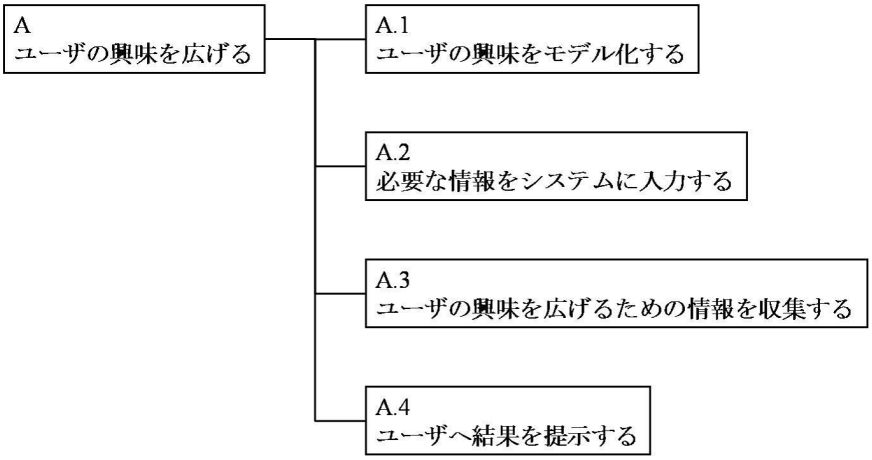


図 3.5: システム全体の機能分解

新たな興味を抱くきっかけを分類した結果，大きく分けて，

- 自分の情報から派生
- 外部の情報から派生

の 2 つに分類できることが分かった．この分類の結果を受け，本システムは，ユーザの情報を把握するためのユーザモデル構築システムおよび外部の情報を抽出するための外部情報整理システムの 2 つのサブシステムを有する．また，2.2節で述べたように，本システムを用いることによって，今後，様々な場面で価値を見出すことを想定している．そのため，ユーザとシステムがインタラクションするためのユーザインターフェースを実装することで，今後，様々な場面で応用できるようにする．以上より，本システムは，図 3.5に示す機能を有し，その機能を満たすために，

- ユーザモデル構築システム
- 外部情報整理システム
- ユーザインターフェース

の 3 つのサブシステムを有する．図 3.5中の A.1 は，ユーザモデル構築システムで行う．図 3.5中の A.3 は，外部情報整理システムで行う．図 3.5中の A.2 および A.4 は，ユーザインターフェースで行う．

ユーザモデル構築システムでは，共同研究を行っている株式会社キーウォーカー（以下，KW 社）が有する技術を用いて，Web ページおよび E-mail を解析し，その中にある単語を取得することにより，ユーザの興味モデルを構築する．ユーザモデル構築システムについては，3.3節および 4.1節で詳しく述べる．

外部情報整理システムでは，ある単語に関連する単語を取得するという KW 社が有する技術を用いて，ユーザが入力したキーワードおよびユーザの興味モデル内にある単語に関連する単語を取得する．外部情報整理システムについては，3.4節および 4.2節で詳しく述べる．

ユーザインターフェースでは，ユーザ ID の入力およびキーワードの入力を行う．また，情報提示方法を考慮し，Mind Map を模倣した形で，ユーザへ結果を提示する．ユーザインターフェースについては，3.5節および 4.3節で詳しく述べる．

また，本システムは，図 3.6に示す流れで処理を行う．

3.3 ユーザモデル構築システム

前述したように，本システムでは，ユーザの興味を把握するために，ユーザの興味モデルを構築する．本節では，ユーザの興味モデルを構築するためのユーザモデル構築システムについて述べる．

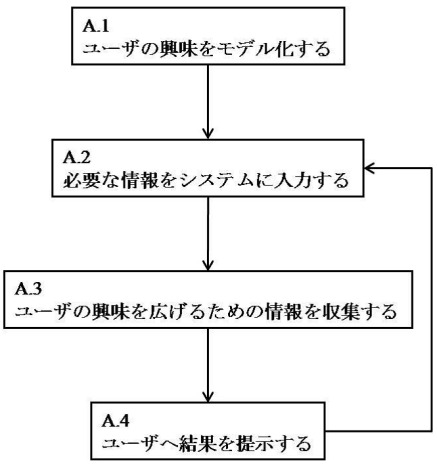


図 3.6: システム全体の流れ

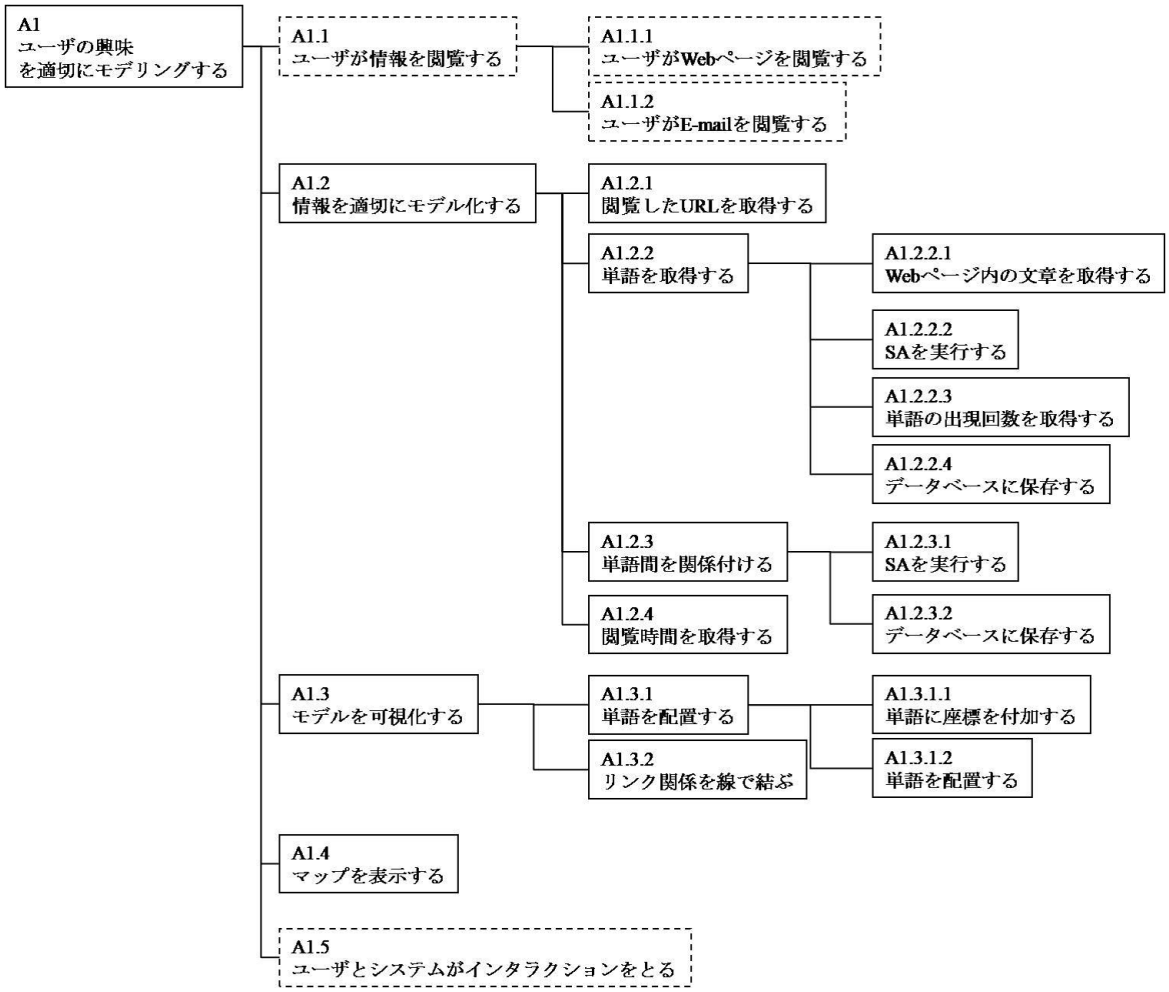


図 3.7: ユーザモデル構築システムの機能分解

ユーザモデル構築システムは，図 3.7に示す機能を有する．このシステムでは，まず，情報源からユーザの興味のある単語を抽出する．つぎに，抽出された単語をつなげることで，ユーザの興味をマップとして表現する．ここで，図中の SA とは，Situation Analyze の略であり，KW 社が有する技術の 1 つである．SA は，

- 抽出された本文に対して，形態素解析および構文解析を行い，文章を単語単位に切り分け，単語のメタデータを付加する
- 閲覧した Web ページ内の本文を解析し，意味ネットワークを用いて，単語間の関係性へメタデータを付加する

という 2 つの機能を有した解析である．ある Web ページに対して SA を行うと，単語，単語間の関係性，単語のメタデータ，単語間の関係性のメタデータおよび単語の出現回数を取得することができる．図 3.7中の破線で囲われている機能は，ユーザが行うものである．図 3.7中の A1.2 は，内容解析部で行う．内容解析部については，3.3.1項で述べる．図 3.7中の A1.3 は，グラフ生成部で行う．グラフ生成部については，3.3.2項で述べる．図 3.7中の A1.4 および A1.5 は，モデル修正部で行う．モデル修正部については，3.3.3項で述べる．

3.3.1 内容解析部

本項では，ユーザの興味モデルを構築するための情報源の解析を行う内容解析部について述べる．

本研究では，ユーザの興味モデルを扱いやすくするため，単語および単語間の関係性にメタデータを付加する．ここで，扱いやすい興味モデルとは，ユーザの興味の対象を整理することが容易なことおよびカテゴライズすることが容易なことである．これにより，ユーザの興味モデルを用いて，ユーザの要求を推測することが可能となる．内容解析部は，

- 閲覧した Web ページの URL の取得
- 閲覧した日時の取得
- Web ページおよび E-mail 中の単語および単語間の関係性の取得
- 取得した単語および単語間の関係性に対するメタデータの付加

の 4 つの機能を有する．ユーザの興味を把握するためには，用いる情報源と付加するメタデータの種類が重要である．そのため，3.3.1.1目では，ユーザの興味モデルを構築するための情報源の選定について述べる．また，3.3.1.2目では，単語および単語間の関係性に付加するメタデータについて述べる．

3.3.1.1 モデル構築のための情報源の選定

本目では，ユーザの興味モデルを構築するための情報源の選定について述べる．

ユーザの特徴をモデル化するための方法として，様々なユーザから情報を取得し，ユーザをモデル化する協調フィルタリングを用いる方法および対象とするユーザの情報だけを用いユーザをモデル化する情報フィルタリングを用いる方法が挙げられる [49]．協調フィルタリングは，ユーザの一般的な行動パターンの抽出や類似した興味や行動パターンを有するユーザコミュニティを抽出する．すなわち，協調フィルタリングはユーザ群の振る舞いに注目した手法である [50][51]．協調フィルタリングにより構築されるモデルはユーザ群のモデルであり，ユーザ個々の細かい特徴をモデル化することができない．したがって，本研究では，コンピュータが各ユーザの特徴を細部までモデル化することができる情報フィルタリングを用いる．

情報フィルタリングを用いる手法は，関心のあるキーワードや分野などをユーザに明示させる明示的手法と Web 閲覧時の挙動などから自動的にキーワードを抽出する暗黙的手法に分けられる [52]．明示的手法は，毎回ユーザに解析対象を明示させるため，ユーザにとって負担が大きい．また，解析対象が限定されるため，ユーザの特徴のモデルも限定的となってしまう [53]．したがって，本研究では，暗黙的手法を用いる．

	個人の差異	興味を反映しているかどうか	一般に使用するかどうか	解析のしやすさ
Web履歴	◎	◎	◎	◎
メール	◎	○	◎	◎
Wordなどのテキスト	◎	△	○	△
PowerPointのスライド	◎	△	○	△
写真	○	△	◎	×
音楽	◎	○	◎	×
動画	◎	○	◎	×
ゲーム	◎	○	○	×
プログラム	○	×	×	×

図 3.8: Pugh Selection の結果

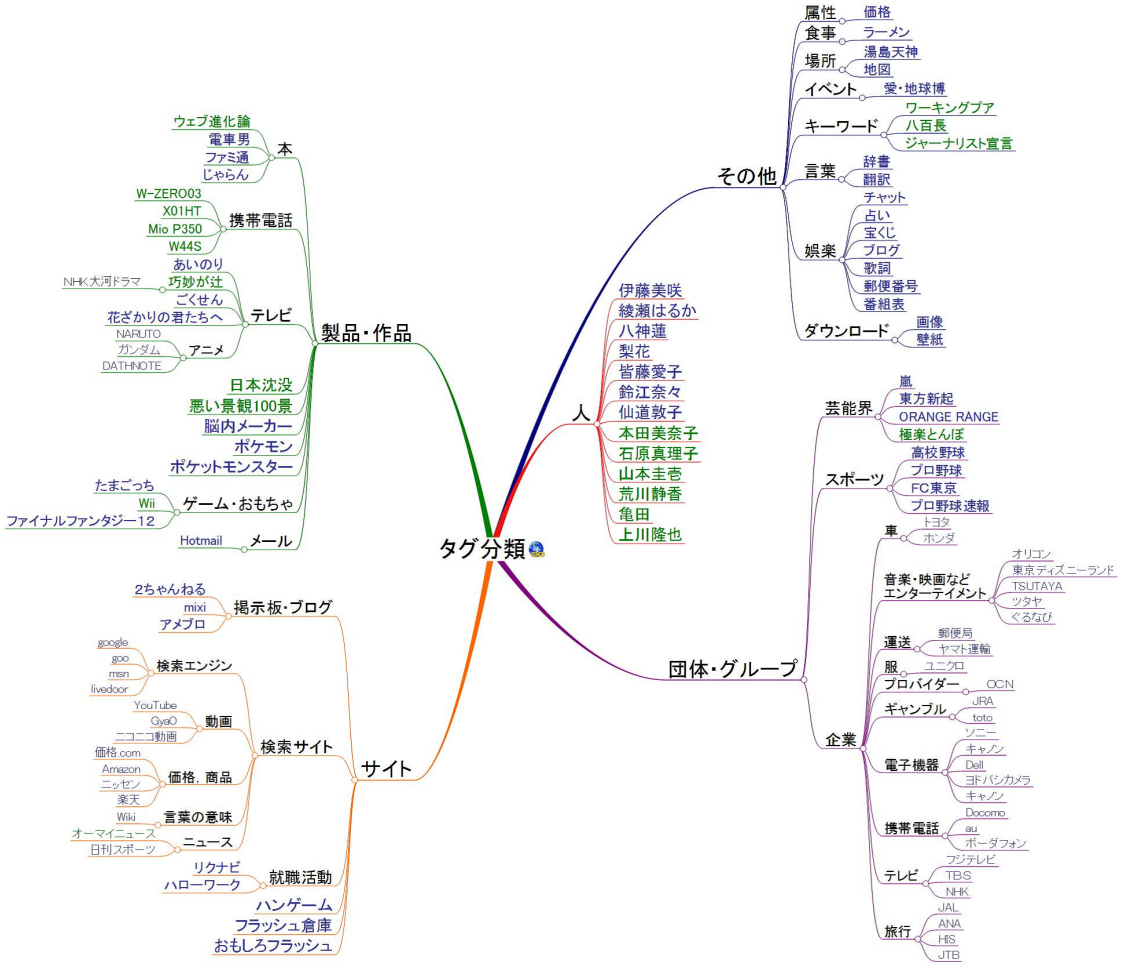


図 3.9: 検索キーワードの分類結果

ユーザの興味を把握するための情報源は，Web の閲覧履歴，E-mail 履歴およびテキストファイルといった多くの情報が挙げられる．本研究では，ユーザの興味を把握するための情報源を決定するために，Pugh Concept Selection を行った．Pugh Concept Selection とは，設定した Criteria を用いて，対象群を比較し，優劣をつける手法である [54]．この Pugh Concept Selection では，Criteria として，

- 個人の差異
- 興味の反映度
- 一般的な使用度
- 解析の容易さ

の 4 つを用いた．個人の差異は，収集する内容，閲覧する内容および記載する内容が，各ユーザによって異なることを評価するための項目である．興味の反映度は，収集する内容，閲覧する内容および記載する内容が，ユーザの興味を反映していることを評価するための項目である．一般的な使用度は，情報源は特定の人だけが使用するのではなく，一般的に使用する人が多いことを評価するための項目である．また，本研究は，KW 社と共同で研究を行っており，KW 社が有する技術を用いて研究を行う．そのため，解析の容易さという Criteria を用いて，KW 社の技術を用いた解析が容易であることを評価する．その結果，図 3.8に示すような結果となった．そのため，本研究では，Web ページおよび E-mail を用いて，ユーザの興味モデルを構築する．

3.3.1.2 付加するメタデータ

前述したように，ユーザの興味の対象のグルーピングを容易にするため，本研究では，単語および単語間の関係性にメタデータを付加する．本目では，単語に付加するメタデータおよび単語間の関係性に付加するメタデータのそれぞれについて述べる．

単語に付加するメタデータ

前述したように，種類を識別したり，グルーピングをするためには，Google といった単語だけでなく，Google は企業名であるといった単語の上位概念が必要である．この上位概念をメタデータとして単語に付加することにより，ユーザの興味モデルを扱いやすいモデルとする．

世の中には，多種多様な単語が存在する．そのため，それらの上位概念であるメタデータの種類も非常に多様である．これらのメタデータを全て把握し，システムに取得させることは困難である．そこで，様々な検索エンジンから発表されている検索ランキング [55][56][57][58][59] を参照し，図 3.9に示すように検索キーワードを分類した．その結果，ユーザが多く検索を行うキーワードは

- George Washington といった人の名前
- SONY，NASA といった企業やグループを示す団体名
- Google，Yahoo といったサイト名
- Xbox360，Freelander といった機種や車種を示す商品名

に分類できることがわかった．

これらの 4 種類のメタデータは，ユーザの興味をモデル化するために満足であることを検証するために，予備実験を行った．予備実験は，20 代〜30 代の男性 8 名に行い，以下の手順で行った．

1. 検索キーワードを自由に決定させ，被験者に閲覧する Web ページを決定させた
2. その Web ページの中で被験者の興味のある単語を抜き出させ，それらの関係性をマップに書かせた
3. 被験者が作成するマップは，名詞を頂点とし，頂点間を結ぶ関係性を明記させた
4. (1)〜(3) の手順を 30 分間自由に行わせた

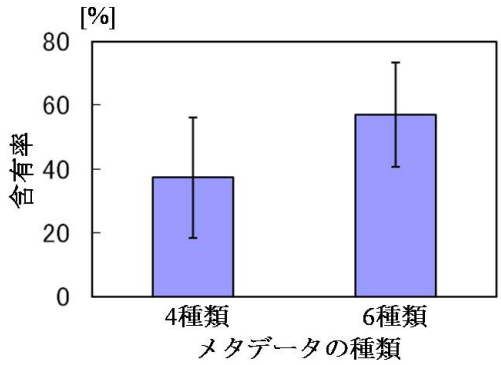


図 3.10: メタデータの含有率

この予備実験では，興味の対象には一切の制限を設けず，自由にマップを作成させた．

被験者が記入した全単語に人手でメタデータを付加し，4 種類のメタデータが付加された単語を抽出することで，メタデータの満足度を検証する．メタデータの種類の満足度は，被験者が作成したマップの中の，これら 4 種類のメタデータの含有率 $R_{inclusion}$ により求める．含有率 $R_{inclusion}$ は，

$$R_{inclusion} = \frac{N_{META}}{N_{ALL}} \tag{3.1}$$

で求める．ここで，N_{META} は被験者が記入した単語のうち，4 種類のメタデータと判断される単語の数，N_{ALL} は被験者が記入した単語の総数である．その結果，図 3.10に示すように，4 種類のメタデータでは，被験者が記入した単語のうち，37.4%しかモデル化できないことがわかった．そこで，被験者が次に多く記入した

- America，日本といった場所を示す地名
- カメラや車といった商品の概念を表す製品名

の 2 種類のメタデータを追加した．その結果，被験者が記入した単語のうちの 57.1%をモデル化することが可能となった．これら 2 つ以外は，被験者によって様々なメタデータが書かれていたため，ユーザの興味を表すために有効なメタデータを見つけることができなかった．したがって，人名，団体名，サイト名，商品名，製品名および地名の 6 種類をメタデータが，ユーザの興味のモデル化には必要であるとわかった．

本研究では，ユーザの興味モデルの構築のための情報源として Web ページだけでなく E-mail も用いる．そのため，E-mail に必要なメタデータを調べるために，アンケートを行った．被験者は，20 代〜30 代の男女 7 人である．アンケートでは，E-mail に必要であると思われるメタデータを被験者に選択させた．また，選択肢以外で必要だと思うメタデータがあれば，自由に追加させた．その結果，図 3.11に示すように，日時，イベント名および人名の順に平均得点が高く，3 つのメタデータを重要であると人は認識していることがわかった．

角谷らによると，E-mail には，通知内容を示す表現，時間表現および感情表現の 3 種類が重要である [60]．通知内容を示す表現は，本研究のイベント名のメタデータに相当し，時間表現は，本研究の日時のメタデータに相当する．また，感情表現は，本アンケートの結果には存在しないが，感情表現は，名詞ではなく，文章全体および動詞で表現されることが多い．そのため，名詞のみを抽出する本研究では，感情表現は扱わないこととする．以上の知見より，日時およびイベント名のメタデータは，E-mail に重要であることがわかる．

また，E-mail は人と人の情報交換の手段の 1 つである．したがって，人を表す人名のメタデータは，E-mail にとって重要であることは明白である．

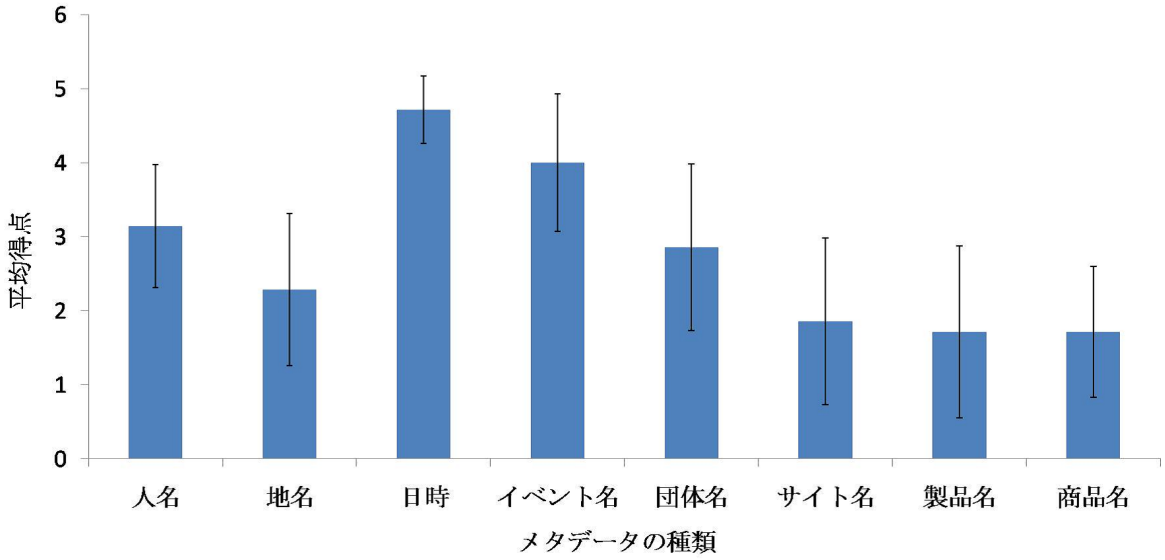


図 3.11: E-mail に必要なメタデータのアンケート結果

表 3.1: 定義する単語間の関係性のメタデータ

メタデータ－メタデータ	関係性
人名—人名	• 共演
人名—団体名	• メンバー
人名—商品名	• 関連作品
団体名—団体名	• 競合
	• 提携
	• 関連団体
	• 同一業界
	• グループ会社
団体名—商品名	• 開発，販売
サイト名—商品名	• 関連サイト
地名—イベント名	• 開催

ここで，日時のメタデータは，E-mail には非常に重要であるが，ユーザの興味を表すものではない．そのため，本研究では，日時のメタデータは扱わないこととする．さらに，インタビューを行った結果，E-mail には，意思決定のメタデータが必要であるという意見があった．意思決定は E-mail の内容を示す重要なメタデータであるが，意思決定は名詞ではなく，文章全体および動詞で表現されることが多い．したがって，名詞のみを抽出する本研究では，意思決定は扱わないこととする．

以上より，本研究では，人名，団体名，サイト名，商品名，製品名，地名およびイベント名の 7 種類のメタデータを用い，これらに関わる単語を取得する．

単語間の関係性に付加するメタデータ

従来，鷹城らの研究に代表されるように，単語間の関係性は，単語の出現頻度を用いて定義していた[61]．種類の識別およびグルーピングをするためには，単語と単語の間に何らかの関係性があるといった情報だけでなく，その関係性を明確にするための単語間の関係性のメタデータを付加することが有効である．本研究では，単語間の関係性を

- 単語と単語の出現頻度に基づいた関係性

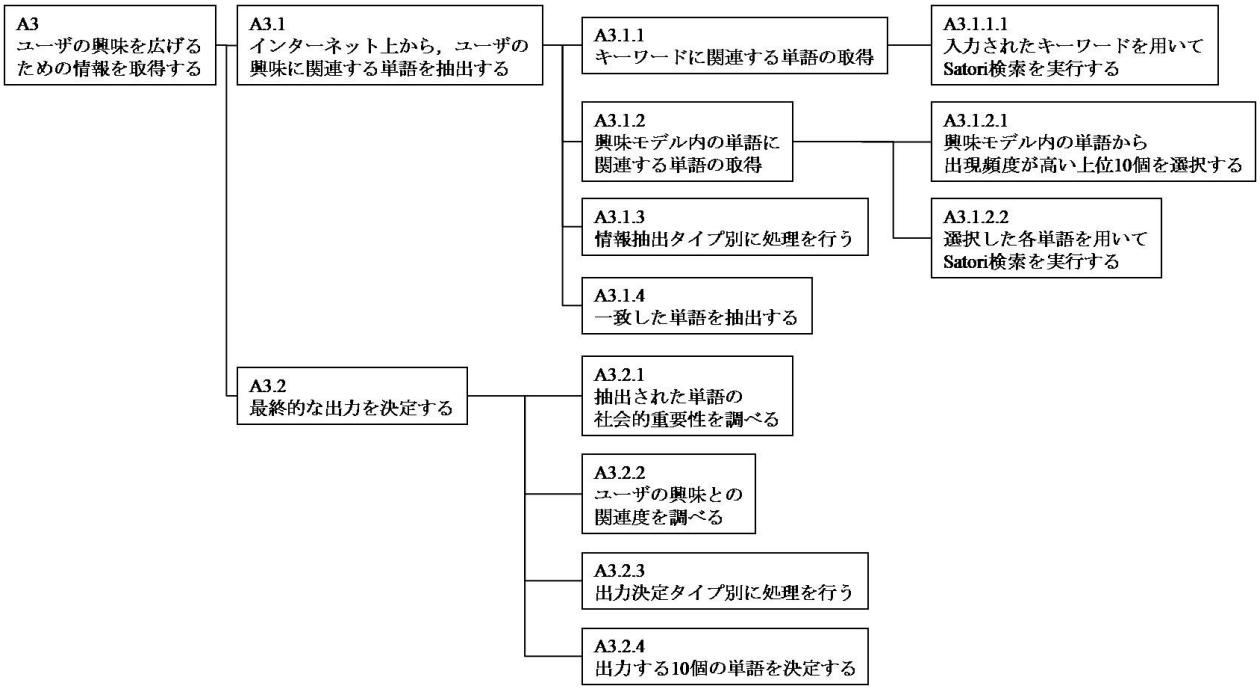


図 3.12: 外部情報整理システムの機能分解

- 単語と単語の意味に基づいた関係性

の 2 種類のアルゴリズムを用いて定義し，単語と単語の意味に基づいた関係性にメタデータを付加する．単語と単語の意味に基づいた関係性は，ユーザが閲覧した Web ページおよび E-mail 内の文章を解析し，文章中に明示してある意味に基づき決定する．

本研究では，取得する単語を，人名，団体名，サイト名，商品名，製品名，地名およびイベント名に関わる単語とした．したがって，これらを結ぶための関係性を定義する必要がある．これらの関係性は多種多様であるため，すべての関係性を定義することは困難である．本研究では，Web ページ内および E-mail 内に比較的明示してあり，取得することが容易である関係について定義する．本研究で定義した単語間の関係性のメタデータを表 3.1に示す．

3.3.2 グラフ生成部

本項では，内容解析部で取得された単語および単語間の関係性を 2 次元空間内に配置するためのグラフ生成部について述べる．グラフ生成部は，

- 取得された各単語に座標を付加する
- 与えられた座標情報に基づき，各単語を 2 次元空間内に配置する
- 取得された単語間の関係性に基づき，単語同士を線で結ぶ

という 3 つの機能を有する．

3.3.3 モデル修正部

本項では，ユーザが自分の興味モデルを閲覧し，システムとインタラクションするためのモデル修正部について述べる．モデル修正部は，

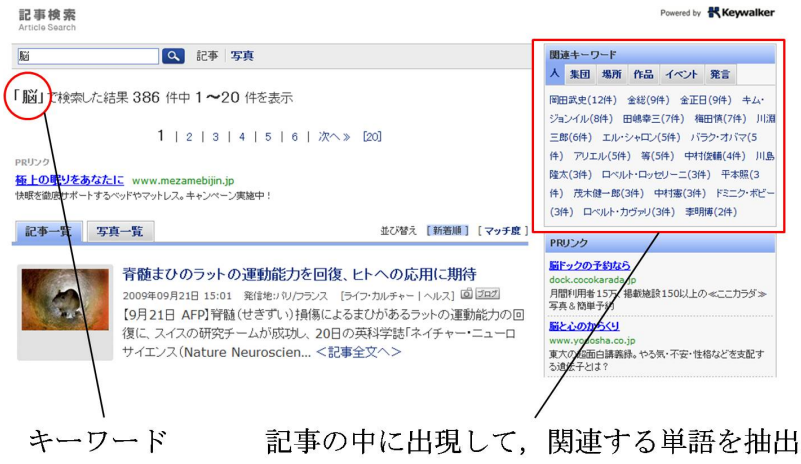


図 3.13: Satori 検索の例

- 各単語に付加されたメタデータの情報に基づき，各単語の背景を着色する
- ユーザの興味モデルを表示する
- ユーザとシステムがインタラクションする

という 3 つの機能を有する．

また，本研究では，情報源から，暗黙的手法を用いてユーザの興味を抽出する．そのため，ユーザの興味を全て正確に抽出することができるとは限らない．ユーザの興味モデルが不正確であった場合，モデル修正部で修正を行うことを想定している．本研究では，ユーザとシステムのインタラクションとして，モデルを修正するための最も基礎的な機能であるマップの拡大，マップの縮小および頂点の移動の 3 つを実装する．

3.4 外部情報整理システム

前述したように，本システムでは，インターネット上からユーザの興味を拡大するための情報を抽出する．本節では，情報抽出のための外部情報整理システムについて述べる．

外部情報整理システムは，図 3.12に示す機能を有する．図 3.12中の A3.1 は，情報抽出部で行う．情報抽出部については，3.4.1項で述べる．図 3.12中の A3.2 は，出力決定部で行う．出力決定部については，3.4.2項で述べる．

3.4.1 情報抽出部

本項では，インターネット上からユーザの興味に関連する単語群を抽出する，情報抽出部について述べる．情報抽出部は，

- 入力されたキーワードおよびユーザの興味モデル内の単語を用いて関連する単語を検索する
- 情報抽出タイプ別に処理を行う

という 2 つの機能を有する．

3.1節で述べたように，新たに興味を抱くきっかけを分類した結果，自分の興味に関連する事柄という分類が挙がった．そのため，情報抽出部で，ユーザが入力したキーワードおよびユーザの興味モデル内の単語を用いて，それぞれに関連する単語をインターネット上から取得する．本研究では，ある単語に

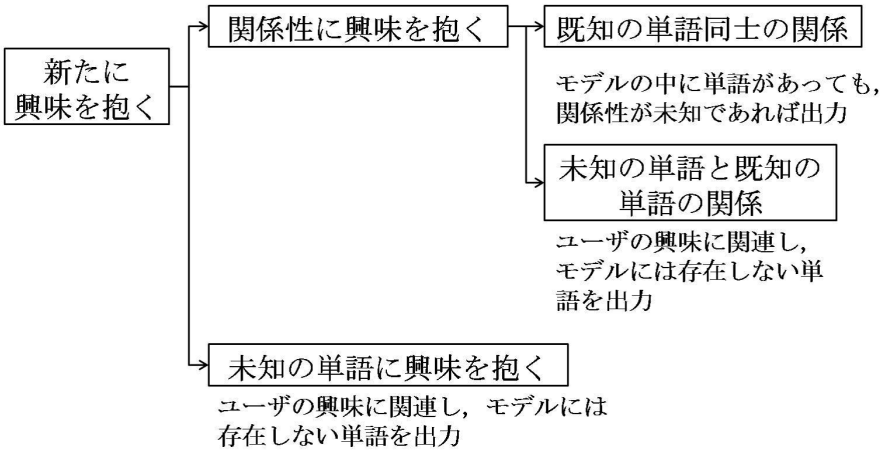


図 3.14: 新たに興味を抱く要素

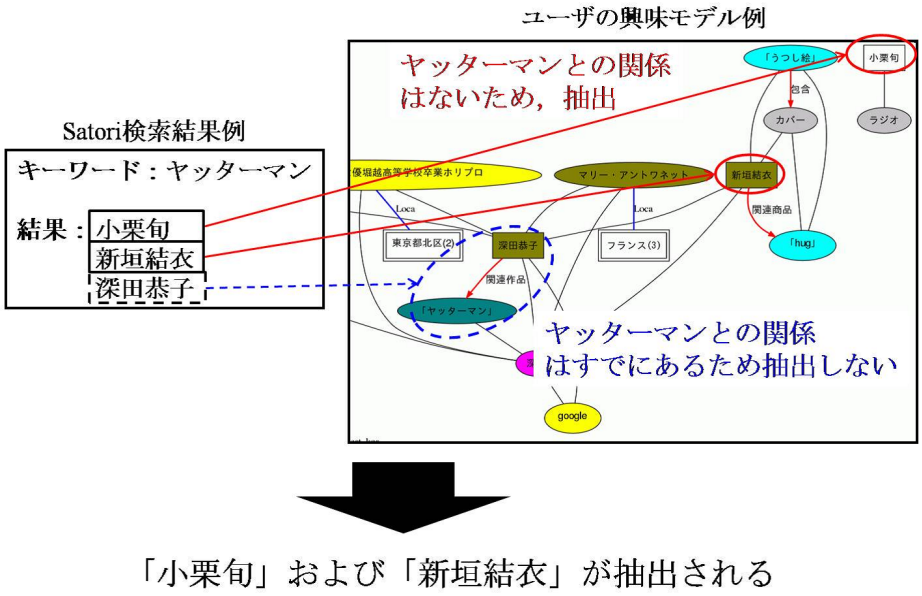


図 3.15: 情報抽出タイプ 1 の概念図

関連する単語をインターネット上から抽出する方法として，Satori 検索を用いる．Satori 検索とは，KW 社が有する技術の 1 つである．この技術は，図 3.13に示すように，まず，入力されたキーワードを用いて，通常の検索を行い，キーワードが出現する Web ページを抽出する．つぎに，抽出された記事に SA を行い，記事の中に出現し，キーワードに関連する単語を抽出する．

ここで，新たに興味を抱く対象について考える．新たに興味を抱く対象は，図 3.14に示すように，

- ユーザにとって未知の単語に興味を抱く場合
- 未知の関係性に興味を抱く場合
  - 既知の単語同士の関係性に興味を抱く場合
  - 既知の単語と未知の単語の関係性に興味を抱く場合

と分類することができる.「ユーザにとって未知の単語に興味を抱く場合」は，ユーザが単語自体に興味を抱く場合である．また，既にある興味に関連している未知の単語だけでなく，興味と関連していなくとも，未知の単語に興味を抱く場合もあると考えられる.「未知の関係性に興味を抱く場合」は，単語自

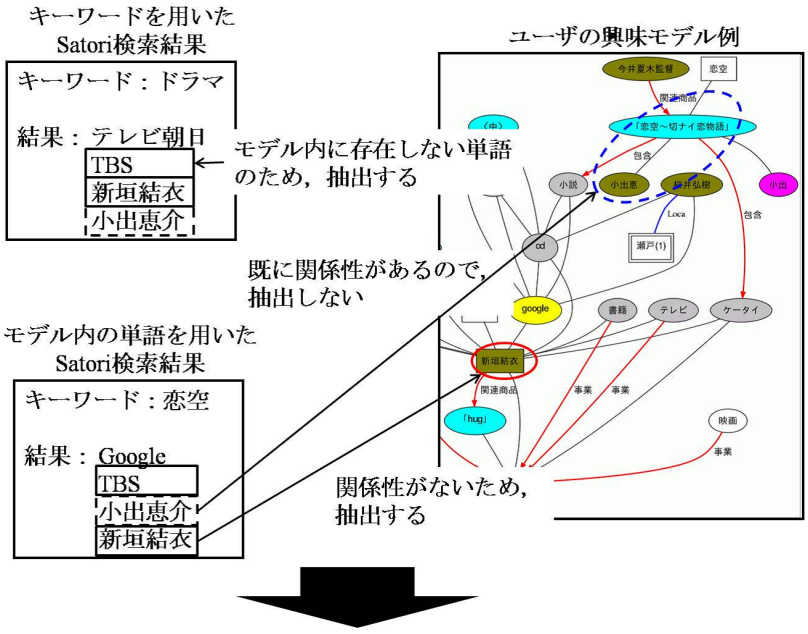


図 3.16: 情報抽出タイプ 2 の概念図

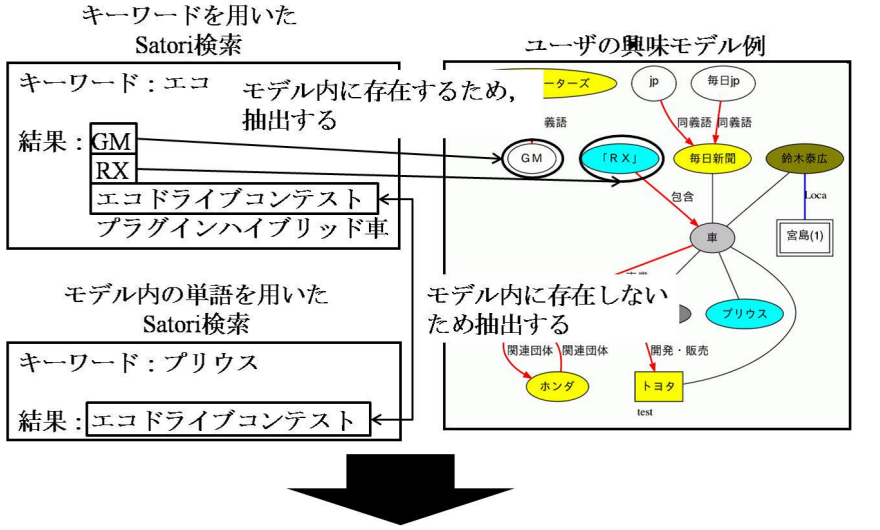


図 3.17: 情報抽出タイプ 3 の概念図

体に興味を抱くのではなく，単語と単語の繋がりに興味を抱く場合である．この分類の中には,「既知の単語同士の関係性に興味を抱く場合」および「未知の単語と既知の単語の関係性に興味を抱く場合」の 2 つが存在する.「既知の単語同士の関係性に興味を抱く場合」は，複数の単語があり，それらの単語自体は既知であったが，それらの単語間の関係性は未知であり，その関係性に興味を抱く場合である.「未知の単語と既知の単語の関係性に興味を抱く場合」は既知の単語と未知の単語が繋がっており，未知の単語自体には興味を抱かないが，既に興味ある単語と繋がっているため，その関係性に興味を抱く場合である.

そのため，本研究では，これらの要素に基づいて，ユーザの興味拡大に必要な情報をインターネット上から抽出するために，以下の 3 種類の情報抽出のタイプを実装する．本研究では，前述した既知の単語を，ユーザの興味モデル内に存在する単語と定義し，未知の単語を，ユーザの興味モデル内に存在しない単語と定義する．また，未知の関係性を，ユーザの興味モデル内には存在しない関係性と定義する．

(a) 情報抽出タイプ 1

このタイプは,「既知の単語同士の関係性に興味を抱く場合」に対応する情報抽出タイプである．このタイプの概念図を図 3.15に示す．図 3.15に示すように，このタイプでは，ユーザが入力したキーワードに関連するユーザの興味モデル内の単語を調べるしたがって,「ユーザの興味モデル内の単語」のみが出力されるため，ユーザの興味により近い結果が出力される．

(b) 情報抽出タイプ 2

このタイプは,「未知の単語に興味を抱く場合」および「既知の単語と未知の単語の関係性に興味を抱く場合」の 2 つに対応する情報抽出タイプである．このタイプの概念図を図 3.16に示す．図 3.16に示すように，このタイプでは，ユーザが入力したキーワードとユーザの興味モデル内の単語の両方に関連する単語を調べる．したがって,「ユーザが入力したキーワードとユーザの興味に共通して関連する単語」が出力されるため，ユーザの興味から少し離れた結果が出力される．

(c) 情報抽出タイプ 3

このタイプは，情報抽出タイプ 1 と情報抽出タイプ 2 を合わせたタイプである．このタイプの概念図を図 3.17に示す．図 3.17に示すように，このタイプでは，ユーザが入力したキーワードに関連するユーザの興味モデル内の単語，ユーザが入力したキーワードとユーザの興味モデル内の単語の両方に関連する単語を調べる．したがって,「ユーザの興味モデル内の単語」および「ユーザが入力したキーワードとユーザの興味に共通して関連する単語」の 2 種類が出力されるため，ユーザの興味に近い結果およびユーザの興味から少し離れた結果が出力される．

本研究では，通常の Web 検索のように，1 つのキーワードを用いて検索を行うのではなく，ユーザが入力した単語およびユーザの興味モデル内の単語をキーワードとして，検索を行う．すなわち，1 回の情報抽出のために，多くの検索回数を必要とし，検索時間が通常よりも長くかかってしまう．長い検索時間はユーザへストレスを与える要因となるため，製品化や実用化を目指すためには大きな問題となる．本研究では，性能に関しては考慮しないが，原田ら [62] が提案しているように，検索の高速化については，様々な知見が存在する．そのため，これらの知見を実装することにより，検索時間の短縮を行うことができ，十分実用化を目指すことができる．

興味モデル内の単語を全て用いて Satori 検索を行うと，検索時間が膨大となり，情報を抽出することができない．そのため，本システムでは，検索の高速化の代用として，興味モデルの中から，ユーザが一番興味を抱いていると考えられる，単語の出現頻度の高い上位 10 個の単語を用いて Satori 検索を行う．

前述したように，興味モデル内には 7 種類のメタデータが混在している．Web ページの中に含まれているメタデータを調査したところ，人名および団体名のメタデータが多く含まれていた．例えば，ユーザが Web ページ内のイベント名に興味を持ったとしても，同一ページ内にある人名および団体名も取得されてしまう．イベント名は，他の Web ページに書かれていることは少ないが，人名および団体名は他の Web ページにも多く書かれていることが多く，人名および団体名の単語の出現頻度と比べると，イベント名の出現頻度は低くなる．したがって，このイベント名の単語は，出現頻度の高い上位 10 個の単語には含まれず，Satori 検索が行われない．また，人名，団体名およびイベント名といった 7 種類のメタデータのうち，どのメタデータを用いることがユーザの興味拡大において有効かということは，明らかではない．以上より，本研究では，この問題を解決するために，ユーザに使用するメタデータを選択させることとした．

	興味との関係性 小	興味との関係性 大
単語の社会的 重要度 小	予想外な事柄	予想外な事柄
単語の社会的 重要度 大	社会的に重要な事柄 又は 予想外な事柄	社会的に重要な事柄 又は 予想通りな事柄

図 3.18: 社会的重要度と興味との関連度で表す分類

3.4.2 出力決定部

本項では，ユーザへ提示する最終的な出力を決定するための，出力決定部について述べる．出力決定部は，

- 抽出された単語の社会的重要性を調べる
- 抽出された単語とユーザの興味の関連度を調べる
- 出力決定タイプ別に処理を行う

という 3 つの機能を有する．

3.1節で述べたように，新たに興味を抱くきっかけの分類のうち，外部情報から派生するものには，

- 社会的に重要な事柄
- 予想外の事柄
- 予想通りの事柄

という 3 種類があった．ここで，この 3 種類は，

- 単語の社会的重要度
- 単語と興味との関連度

の 2 軸で表すことができると考えた．すなわち，図 3.18に示すように，単語の社会的重要度が高ければ，社会的に重要な事柄となり，社会的重要度が低く興味との関連度が低いもの，または社会的重要度が低く興味との関連度が高いものは，予想外の事柄となる．また，社会的重要度が高く興味との関連度が高いものは，予想通りの事柄となる．したがって，本研究では，ユーザの興味を拡大させる情報を決定するための軸として，単語の社会的重要度および単語と興味との関連度の 2 つを用いる．

以上より，出力決定部では，単語の社会的重要度の大・小および単語と興味との関連度の大・小を組み合わせた，以下の 4 種類の出力決定タイプを実装する．

(a) 出力決定タイプ 1
このタイプは，単語の社会的重要度が小さく，興味との関連度が小さいものを抽出するタイプである．予想外の事柄を出力することができる．

(b) 出力決定タイプ 2
このタイプは，単語の社会的重要度が大きく，興味との関連度が小さいものを抽出するタイプである．社会的に重要な事柄または予想外の事柄を出力することができる．

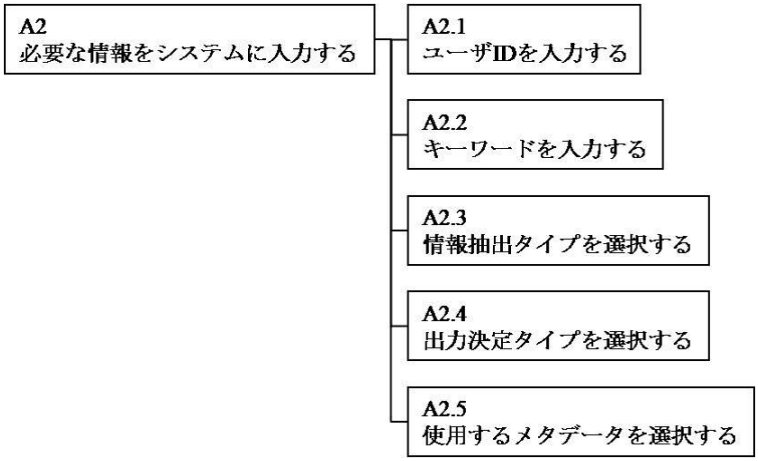


図 3.19: 入力部の機能分解

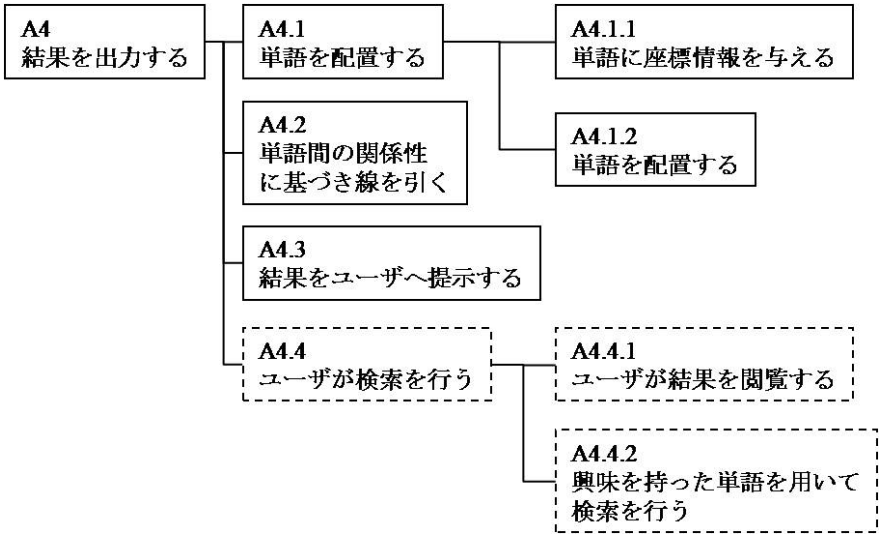


図 3.20: 提示部の機能分解

(c) 出力決定タイプ 3
このタイプは，単語の社会的重要度が低く，興味との関連度が大きいものを抽出するタイプである．予想外の事柄を出力することができる．

(d) 出力決定タイプ 4
このタイプは，単語の社会的重要度が大きく，興味との関連度が大きいものを抽出するタイプである．社会的に重要な事柄または予想通りの事柄を出力することができる．

また，本研究では，提示する結果の見やすさを考慮し，10 個の単語を出力することとした．そのため，例えば，出力決定タイプ 1 であれば，単語の社会的重要度が小さく，興味との関連度が小さいものから 10 個の単語を最終的な出力とする．

3.5 ユーザインターフェース

本システムのユーザインターフェースは，ユーザ ID およびキーワードといった情報提示に必要な情報を入力するための入力部および結果を提示するための提示部の 2 つのサブシステムで構成される．3.5.1項では，入力部について述べる．3.5.2項では，提示部について述べる．

図 3.21: Mind Map の例 [63]

3.5.1 入力部

入力部は，ユーザの興味拡大支援に必要な情報を入力するためのインターフェースである．入力部は，図 3.19に示す機能を有する．そのため，入力部は，ユーザ ID，情報抽出のタイプ，出力決定のタイプ，情報提示に使用するメタデータの種類およびキーワードの 5 種類のパラメータを入力するためのインターフェースを有する．

3.5.2 提示部

提示部は，システムが出力する情報を提示するためのインターフェースである．提示部は，図 3.20に示す機能を有する．図 3.20中の破線で囲われた機能は，ユーザが行う機能である．

本システムで提示される情報は，情報抽出タイプによって異なる．以下に，各情報抽出タイプで出力される情報を述べる．

(1) 情報抽出タイプ 1 で提示される情報
このタイプでは，提示される情報は,「ユーザが入力したキーワード」および「入力されたキーワードと関連し，ユーザの興味モデル内に存在する単語」の 2 種類である．

(2) 情報抽出タイプ 2 で提示される情報
このタイプでは，提示される情報は,「ユーザが入力したキーワード」,「興味に関連する単語を検索するためのユーザの興味モデル内の単語」および「入力されたキーワードに関連し，かつユーザの興味モデル内の単語に関連している単語」の 3 種類である．

(3) 情報抽出タイプ 3 で提示される情報
このタイプでは，提示される情報は,「ユーザが入力したキーワード」,「入力されたキーワードと関連し，ユーザの興味モデル内に存在する単語」,「興味に関連する単語を検索するためのユーザの興味モデル内の単語」および「入力されたキーワードに関連し，かつユーザの興味モデル内の単語に関連している単語」の 4 種類である．

本研究では，これらの出力される情報を Mind Map を模倣した放射状に配置し，ユーザへ提示する．Mind Map とは，図 3.21に示すように，放射状に言葉やイメージを描き，脳の中の考えを見える化し，思考空間を広げる技術である [64]．その結果，自分の考えの整理および理解を促し，その考えを発達させることが可能となる．Mind Map を模倣した情報提示を行うことにより，ユーザの興味拡大を促すと考えられる．

また，本研究では，図 3.20中の A4.4 にあるように，ユーザがシステムから提示された結果に興味を持った場合，検索を行い，更なる情報を取得してもらうという狙いがある．そのため，検索を行う際のユーザにかかる負担をなるべく減らし，多くの検索を行ってもらう必要がある．したがって，Web ブラウザに，ユーザが直接検索するためのキーワードを入力しなくとも，提示された結果を選択することで検索を行うことを可能にする．本研究では，新たに抱く対象は単語および単語間の関係性であるとした．本システムには，興味を抱いた単語を検索する機能，興味を抱いた関係性を検索する機能を実装する．

第4章　思考支援システムの構築

本章では，3章で述べた 3 つのサブシステムが有する機能を満たすために行う処理について述べる．4.1節では，ユーザモデル構築システムが有する機能を満たすための処理および構築について述べるとともに，動作確認を行った結果について述べる．4.2節では，外部情報整理システムが有する機能を満たすための処理および構築について述べるとともに，動作確認を行った結果について述べる．4.3節では，ユーザインターフェースが有する機能を満たすための処理および構築について述べるとともに，動作確認を行った結果について述べる．

4.1　ユーザモデル構築システム

本節では，図 3.7に示す，ユーザモデル構築システムが有する機能を満たすための処理について述べる．ユーザモデル構築システムは，内容解析部，グラフ生成部およびモデル修正部の 3 つのサブシステムで構成される．4.1.1項では，内容解析部の構築について述べる．4.1.2項では，グラフ生成部の構築について述べる．4.1.3項では，モデル修正部の構築について述べる．

4.1.1　内容解析部

3.3.1項で述べたように，本研究では，Web ページおよび E-mail を用いてユーザの興味モデルを構築する．本研究では，既存のブラウザである Firefox に自作のプラグインをインストールすることによって，Web ページおよび E-mail の文章を取得し，データベースで解析を行う．本項では，Web ページの解析および E-mail の解析についてそれぞれ述べる．

4.1.1.1　Web ページの解析

本目では，Web ページを解析し，単語および単語間の関係性を取得し，それぞれにメタデータを付加するための処理について述べる．Web ページの解析は，以下のように処理を行う．

1. Firefox から URL を受信する
2. 閲覧している Web ページの HTML を取得する
3. タグを整理する
   (a) 取得した HTML の中から，不要なタグやコメントを除去する
   (b) タグの表記を統一する
4. 本文を取得する
   (a) 特定の文字 (‘。‘，‘、‘等) を中身に含むタグを探す
   (b) 上記の文字の出現位置と出現数に基づき，タグにスコアを付ける
   (c) 最もスコアが高いタグとその子孫タグを本文と判定し，取得する
5. その他のタグを除去する

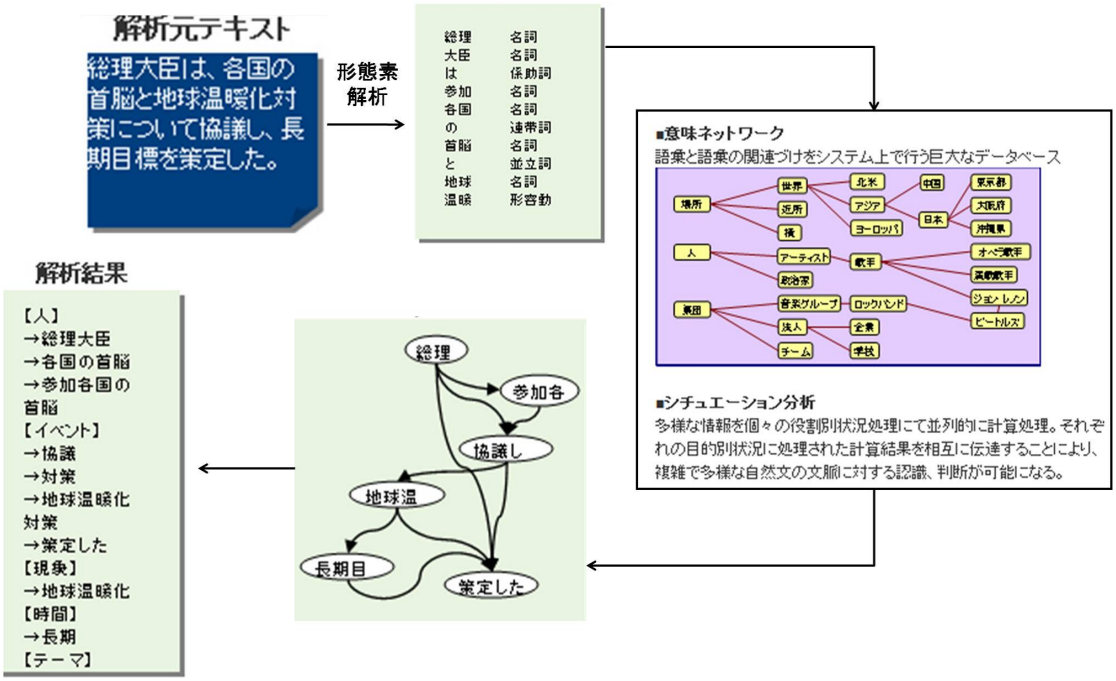


図 4.1: SA の概念図 ([65] を改変)

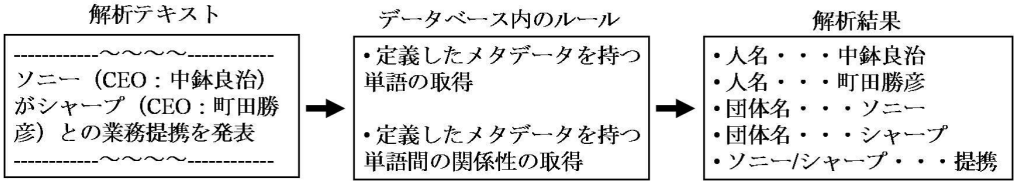


図 4.2: 単語および関係性の抽出例

6. 取得した文章に対して SA を行う
7. その結果，単語および単語間の関係性を抽出し，単語および単語間の関係性にメタデータを付加する

SA は，図 4.1に示すように，

1. 解析元の文章に形態素解析を行う
2. データベースに構築されているルールである意味ネットワークを用いて，シチュエーション分析を行う
3. 定義されたメタデータに関わる名詞を抽出する

という処理を行う．シチュエーション分析とは，形態素解析の結果抽出された単語と意味ネットワークを照らし合わせ，単語にメタデータを付加し，単語と単語をつなぎ合わせるものである．本研究では，形態素解析には，Mecab[66] を用いた．

3.3節で述べたように，本研究では，人名，団体名，サイト名，商品名，製品名，地名およびイベント名のメタデータが付加された単語を取得する．これらの単語を取得するための準備として，人名および団体名といったメタデータを定義するための独自のルールを作成し，そのルールをデータベースに保存した．この独自のルールが意味ネットワークと呼ばれ，KW 社が有している技術である．意味ネットワーク内に含まれる，単語にメタデータを付加するためのルールをいくつか例に示す．

- 社長や代表といった役職が付加されている名詞は人名とする
- 「{ 名詞 } の発売」と書かれている場合，名詞を商品名とする
- 「サイト { 名詞 }」と書かれている場合，名詞をサイト名とする
- 「{ 名詞 } 大会」と書かれている場合，名詞をイベント名とする
- 「{ 名詞 } に社名を変更した」と書かれている場合，名詞を団体名とする
- 「{ 名詞 } 大学」と書かれている場合，名詞を地名とする

そして，Web ページ内の本文を抽出し，図 4.2に示すように，その本文とルールを照合し，それに合致した場合，その単語にはメタデータが付加され，SA の結果として出力される．

3.3節で述べたように，本研究では，2 種類のアルゴリズムを併用し，単語間の関係性を定義する．まず，単語と単語の出現頻度に基づいた関係性について述べる．単語と単語の出現頻度に基づいた関係性は，単語と単語の関係性の重みにより定義される．単語 1 に対する単語 2 の関係性の重みは，

$$Weight_{W1\&W2} = \log(N_{W1\&W2}) \times \log(1 + \frac{N_{W1\&W2}N_{URL}}{N_{W1}N_{W2}}) \tag{4.1}$$

に基づいて計算する．ここで，$Weight_{W1\&W2}$ は単語 1 に対する単語 2 の関係性の重み，N_{URL} はユーザが今まで閲覧した Web ページの総数，$N_{W1\&W2}$ はユーザが今まで閲覧した Web ページの中に単語 1 および単語 2 が同時に含まれるページ数，N_{W1} はユーザが今まで閲覧した Web ページの中に単語 1 が含まれるページ数，N_{W2} は今まで閲覧した Web ページの中に単語 2 が含まれるページ数を示す．この計算を，SA を行った結果取得された単語に対して行う．さらに，単語 1 に関係性を有する全単語の重みの平均値および標準偏差を求め，平均値と標準偏差の和以上の重みを持つ単語を関係性があるとみなすこととした．

つぎに，単語と単語の意味に基づいた関係性について述べる．この関係性は，重みの計算を行うのではなく，データベースに保存されたルールに基づき，単語と単語の意味に基づいた関係性を付加する．単語と単語の意味に基づいた関係性も，単語の取得と同様，関係性を定義するための独自のルールを定義し，意味ネットワークとしてデータベースに保存した．意味ネットワーク内に含まれる，単語間の関係性にメタデータを付加するためのルールをいくつか例に示す．

- 「{ 団体名 } が発売した { 商品 }」と書かれている場合，団体名と商品の関係性は開発・販売とする
- 「{ 団体名 } の子会社 { 団体名 }」と書かれている場合，団体名と団体名の関係性はグループ会社とする
- 「{ 団体名 } を抜いて { 団体名 }」と書かれている場合，団体名と団体名の関係性は競合とする
- 「{ 団体名 }{ 役職 } に就任した { 人名 }」と書かれている場合，団体名と人名の関係性はメンバーとする
- 「{ 人名 } と { 人名 } が共演」と書かれている場合，人名と人名の関係性は共演とする

そのデータベースと Web ページ内の文章を照合し，関連作品といった関係性が取得できた場合，単語間のつながりに単語間の関係性のメタデータを付加する．

4.1.1.2 E-mail の解析

本目では，E-mail を解析し，単語および単語間の関係性を取得し，それぞれにメタデータを付加するための処理について述べる．E-mail の解析は，以下のように処理を行う．本研究では，E-mail の中でも，Gmail を用いて解析を行う．

1. Firefox から，閲覧中の Gmail の HTML ソースを取得する
2. 取得した HTML ソースに URL エンコードを行い，ユーザ ID を URL に載せる
3. その URL をメールサーバに HTTP リクエストする
4. メールサーバは，リクエストされた URL から，ユーザ ID および HTML ソースを取得する
5. HTML ソースの中から，特定のタグで囲まれたメールの本文を抽出する
6. Web ページの解析に使用するデータベースに HTTP リクエストを送る
7. データベースは HTTP リクエストを受け取り，Web ページの解析と同様に SA を行う
8. その結果，単語および単語間の関係性を抽出し，単語および単語間の関係性にメタデータを付加する

E-mail の解析は，Gmail からメールの本文を抽出した後，Web ページの解析に用いるデータベースを用いて解析を行う．したがって，単語および単語間の関係性の抽出や単語および単語間の関係性へのメタデータへの付加は，Web ページと同様のルールを用いる．

4.1.2 グラフ生成部

本項では，内容解析部で抽出された単語に座標を与え，2 次元空間内に配置するための処理について述べる．グラフ生成部は，以下のように処理を行う．

1. Graphviz[67] を用いて，単語の座標を決定する
2. その結果，各単語に x 座標および y 座標を割り当てる
3. これらの情報を XML 形式で出力する

本研究では，Graphviz の形式は，なるべく単語間を結ぶ線が重ならないように，dot 形式を用いて単語の座標を決定した．

実際に出力される XML の一例を図 4.3に示す．図 4.3に示すように，XML には，ユーザ ID，単語の ID 番号，単語，単語の出現回数，単語の座標情報，単語のメタデータ名，単語間の関係性および単語間の関係性のメタデータの情報が含まれている．ここで，＜ graph ＞タグの中の’user_id’ はユーザ ID，＜ node ＞タグの中の’id’ は単語の ID 番号，’name’ は単語，’count’ は単語の出現回数，＜ pos ＞タグの中の’x’ は単語の x 座標，’y’ は単語の y 座標，＜ meta ＞タグの中の’class_name’ は単語のメタデータ名をそれぞれ示す．また，＜ link ＞タグの中にある’name’ に開発といった言葉が書かれているものが，単語と単語の意味に基づいた関係性であり，空白になっているものが，単語と単語の出現頻度に基づいた関係性である．

4.1.3 モデル修正部

本項では，データベースで加工された情報を取得し，その情報を用いてユーザとインタラクションを行うための処理について述べる．本研究では，モデル修正部の開発環境として，Adobe Flash CS3 Professional を用いた．モデル修正部は，以下のように処理を行う．

1. データベースから出力された XML 形式の情報を Flash で受け取る
2. その XML の中の＜ nodes ＞タグの中から，単語に関する情報を取得する

```
-<graph user_id="thesis">
  -<nodes>
    -<node id="0" name="「プリウス」" count="2">
      <pos x="66" y="-444"/>
      -<attr>
        <meta class_name="商品" value="" class_num="2"/>
      </attr>
    </node>
    -<node id="1" name="車" count="2">
      <pos x="66" y="-356"/>
      -<attr>
        <meta class_name="製品" value="" class_num="0"/>
      </attr>
    </node>
    -<node id="2" name="トヨタ自動車" count="2">
      <pos x="310" y="-268"/>
      -<attr>
        <meta class_name="企業名" value="" class_num="1"/>
      </attr>
    </node>

                ⋮

    -<node id="9" name="名古屋" count="2">
      <pos x="231" y="-40"/>
      -<attr>
        <meta class_name="場所" value="" class_num="5"/>
      </attr>
    </node>
  </nodes>
-<links>
    <link id="10" from="0" to="1" name="包含" direct="forward"/>
    <link id="11" from="4" to="5" name="開発・販売" direct="forward"/>
    <link id="12" from="1" to="2" name="事業" direct="forward"/>
    <link id="13" from="7" to="8" name="開発・販売" direct="forward"/>
    <link id="14" from="0" to="2" name="" direct="none"/>
    <link id="15" from="0" to="3" name="" direct="none"/>
    <link id="16" from="2" to="3" name="" direct="none"/>
```


図 4.3: 実際に出力される XML の例

表 4.1: メタデータと背景色の対応

メタデータ	背景色
人名	水色
団体名	黄色
商品名	赤色
製品名	緑色
地名	紫色
サイト名	灰色
イベント名	ピンク色

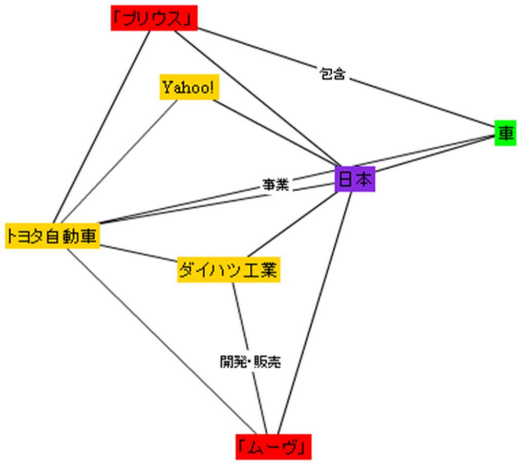


図 4.4: 実際に出力される興味モデルの例

3. 単語に関する情報に基づき，頂点の内容，頂点の座標，頂点の背景色および頂点の大きさを決定する

4. XML の中の＜ links ＞タグの中から，単語間の関係性に関する情報を取得する

5. 単語間の関係性に関する情報に基づき，単語間に線を引き，メタデータを付加する

本研究では，表 4.1に示すように単語のメタデータに基づき，頂点の背景色を決定した．また，前述したように，SA を行った結果，ユーザが今まで閲覧したページの中で単語が何回出現したかという単語の出現回数を求めることができる．この単語の出現回数に基づき，マップに表示する単語の大きさを決定した．本研究では，単語の出現回数が 1 の場合，文字のサイズを 13pt とし，単語の出現回数が 1 増えるごとに文字のサイズを 3pt ずつ大きくした．単語の出現回数が 5 を超えると，それ以上はすべて文字のサイズを 25pt とした．実際に作成されるユーザの興味モデルの一例を図 4.4に示す．

3.3.3項で述べたように，本研究では，ユーザとシステムのインタラクションとして，マップの拡大，マップの縮小および頂点の移動の 3 つの機能を実装した．マップの拡大および縮小は，ボタンを設置し，

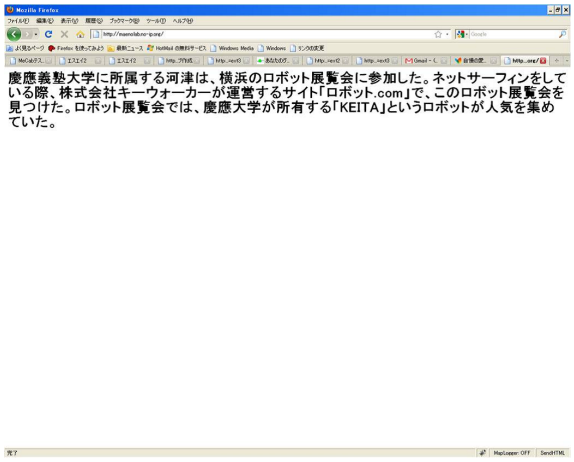


図 4.5: 用意した Web ページ

図 4.6: 用意した E-mail

表 4.2: 想定した単語とメタデータの対応

用意したテキスト	想定するメタデータ
慶應義塾大学	団体名
河津	人名
横浜	地名
ロボット展覧会	イベント名
キーウォーカー	団体名
ロボット.com	サイト名
KEITA	商品名
ロボット	製品名

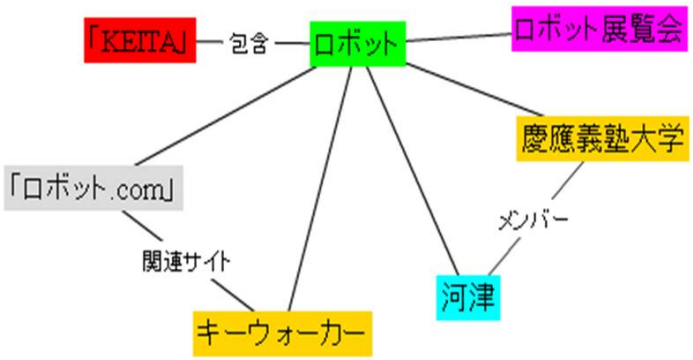


図 4.7: 構築されたユーザの興味モデル

ボタンを押すごとに拡大および縮小を行うものとした．本研究では，拡大率を 1.1 倍とし，縮小率を 0.9 倍とした．頂点の移動は，頂点をドラッグすると，マウスに追従するようにした．

4.1.4 動作確認

本項では，構築したユーザモデル構築システムの動作確認について述べる．確認する項目は，

(a) 用意された Web ページおよび E-mail を解析し，本研究で使用する 7 種類のメタデータを取得し，正しくユーザの興味モデルを構築することができること

(b) 提示されたマップの拡大を正しく行うことができること

(c) 提示されたマップの縮小を正しく行うことができること

(d) 頂点を移動させることができること

の 4 つである．

まず，(a) について動作確認を行った．確認方法は，

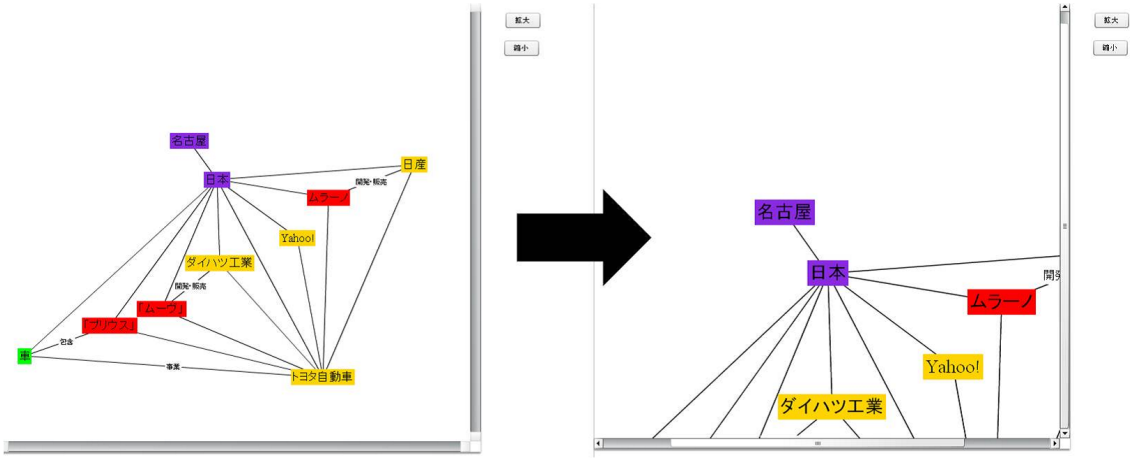


図 4.8: モデルの拡大の動作確認

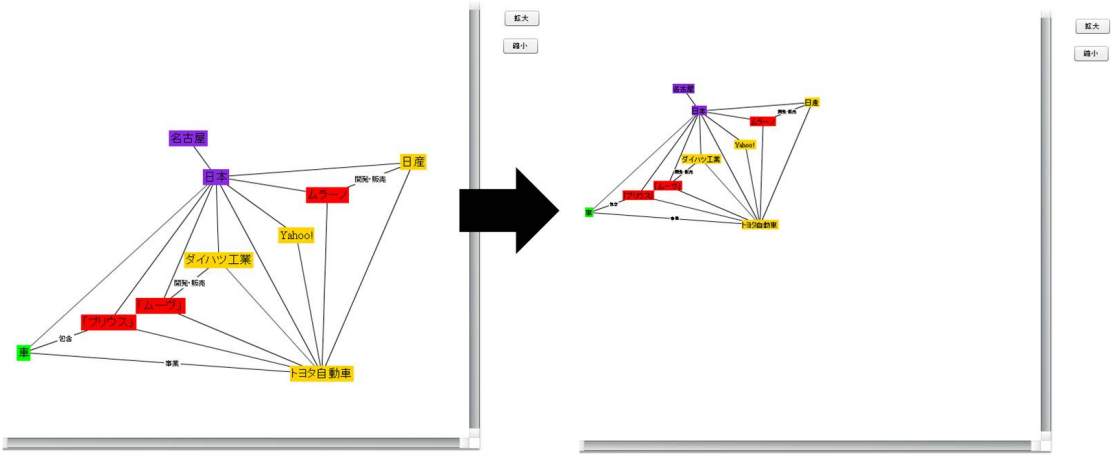


図 4.9: モデルの縮小の動作確認

1. 人名，団体名，商品名，製品名，サイト名，地名およびイベント名が含まれる Web ページおよび E-mail を用意した
2. その Web ページおよび E-mail をシステムに解析させた
3. ユーザの興味モデルが正しく構築されていることを確認した

とした．図 4.5に用意した Web ページ，図 4.6に用意した E-mail をそれぞれ示す．この動作確認で用意した Web ページおよび E-mail は，同じテキストを用い，同じように解析できることを確認した．この動作確認では，表 4.2のようにメタデータが付加されることを想定した．Web ページおよび E-mail の解析の結果，両方とも図 4.7に示すように，同じ興味モデルが構築された．したがって，正しくモデルが構築されていることを確認した．

つぎに，(b) および (c) について動作確認を行った．確認方法は，

1. 拡大ボタンおよび縮小ボタンを押した
2. モデルが正しく拡大および縮小されていることを確認した

とした．この結果を図 4.8および図 4.9に示す．したがって，モデルが正しく拡大および縮小されていることを確認した．

最後に，(d) について動作確認を行った．確認方法は，

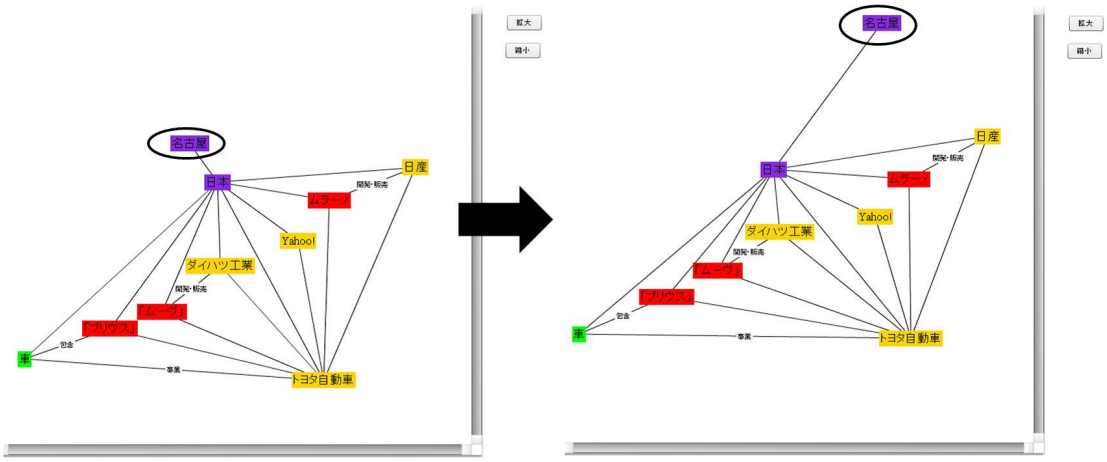


図 4.10: 頂点の移動の動作確認

図 4.11: Yahoo! API の出力結果例

1. モデル内の頂点をドラッグした
2. マウスに追従することを確認した

とした．この結果を図 4.10に示す．したがって，正しく頂点を移動させることができることを確認した．

4.2 外部情報整理システム

本節では，図 3.12に示す，外部情報整理システムが有する機能を満たすための処理について述べる．外部情報整理システムは，情報抽出部および出力決定部の 2 つのサブシステムで構成される．4.2.1項では，情報抽出部の構築について述べる．4.2.2項では，出力決定部の構築について述べる．

4.2.1 情報抽出部

本項では，ユーザの興味モデル内の単語を用いて，インターネット上から関連した単語を抽出するための処理について述べる．本研究では，3.4.1項で述べたように，3 種類の情報抽出タイプを実装した．そ

れぞれの情報抽出タイプは，以下のように処理を行う．

本研究では，Yahoo!Japan が提供する Web 検索 API[68](以下，Yahoo! API) を用いて，関連する単語を抽出する．Yahoo! API は，1 つの検索キーワードを用いた検索結果を XML として出力する．例えば，Yahoo! API で「トヨタ自動車」を検索キーワードとして検索を行うと，図 4.11に示す XML が出力される．Yahoo! API では，1 つのキーワードにつき，100 件から 1000 件までの検索結果を出力することができる．本研究では，検索時間および検索結果を考慮し，1 つのキーワードで 300 件の検索結果を出力させることとした．

- 情報抽出タイプ 1 で行う処理

  1. ユーザが入力部で入力したユーザ ID，キーワード，情報抽出タイプおよび検索に用いるメタデータを，システムは取得する
  2. 入力されたキーワードを検索キーワードとして，Yahoo! API を用いて検索を行う
  3. Yahoo! API が返す XML の中から，タイトルおよびサマリを抽出する
  4. 抽出したタイトルおよびサマリに SA を行い，人名，団体名，製品名，商品名，サイト名，地名およびイベント名に関する単語を抽出する
  5. 抽出された単語群とユーザの興味モデル内の単語を照合し，一致するものを抽出する

- 情報抽出タイプ 2 で行う処理

  1. ユーザが入力部で入力したユーザ ID，キーワード，情報抽出タイプおよび検索に用いるメタデータを，システムは取得する
  2. 選択されたメタデータが付加されているユーザの興味モデル内の単語の中から，出現回数が多い上位 10 個を抽出する
  3. 入力されたキーワードを検索キーワードとして Yahoo! API を用いて検索を行う（この検索結果をキーワード結果と呼ぶ）
  4. 抽出された 10 個のそれぞれの単語を検索キーワードとして Yahoo! API を用いて検索を行う（この検索結果をモデル結果と呼ぶ）
  5. Yahoo! API が返す XML の中から，タイトルおよびサマリを抽出する
  6. 抽出したタイトルおよびサマリに SA を行い，人名，団体名，製品名，商品名，サイト名，地名およびイベント名に関する単語を抽出する
  7. キーワード結果から抽出された単語群とモデル結果から抽出された単語群を照合し，一致するものを抽出する

- 情報抽出タイプ 3 で行う処理

  1. ユーザが入力部で入力したユーザ ID，キーワード，情報抽出タイプおよび検索に用いるメタデータを，システムは取得する
  2. 選択されたメタデータが付加されているユーザの興味モデル内の単語の中から，出現回数が多い上位 10 個を抽出する
  3. 入力されたキーワードを検索キーワードとして Yahoo! API を用いて検索を行う（この検索結果をキーワード結果と呼ぶ）
  4. 抽出された 10 個の単語それぞれを検索キーワードとして Yahoo! API を用いて検索を行う（この検索結果をモデル結果と呼ぶ）
  5. Yahoo! API が返す XML の中から，タイトルおよびサマリを抽出する

図 4.12: 検索ヒット数および単語の出現回数を載せた XML

6. 抽出したタイトルおよびサマリに SA を行い，人名，団体名，製品名，商品名，サイト名，地名およびイベント名に関する単語を抽出する
7. キーワード結果から抽出された単語群とユーザの興味モデル内の単語を照合し，一致するものを抽出する
8. キーワード結果から抽出された単語群とモデル結果から抽出された単語群を照合し，一致するものを抽出する

4.2.2 出力決定部

本項では，情報抽出部で抽出された単語群の中から，ユーザへ提示する単語を決定するための処理について述べる．3.4.2項で述べたように，本研究では，ユーザへ提示する単語を決定するために，単語の社会的重要度およびユーザの興味との関連度の 2 つの軸を用いて，4 種類の出力決定タイプを実装した．本研究では，単語の社会的重要性を Yahoo!の検索ヒット数とし，ユーザの興味との関連度を SA を行った結果得られる単語の出現回数と定義した．また，単語の出現回数が 1 のものは，ユーザの興味との関連度が小さいと定義し，単語の出現回数が 2 以上のものは，ユーザの興味との関連度が大きいと定義した．

各出力決定タイプは，まず，共通の処理を行い，つぎに個別の処理を行う．共通の処理は，

1. 情報抽出部で抽出された単語群をそれぞれ Yahoo! API を用いて検索し，検索ヒット数を取得する
2. SA を行った結果得られた，単語の出現回数を取得する
3. 情報抽出部で取得した情報，単語の検索ヒット数および単語の出現回数を XML 形式で出力する

という処理である．実際に出力される XML を図 4.12に示す．ここで，＜ keyword ＞タグの中はユーザが入力したキーワードに関する情報，＜ output ＞タグの中は情報抽出部で取得した単語群に関する情報および＜ links ＞タグの中は単語と単語のつながりに関する情報を示す．また，＜ graph ＞タグの中の’user_id’ はユーザ ID，＜ node ＞タグの’id’ 単語の ID 番号，’name’ は単語，’count’ は単語の出現回数，’total_hits’ は Yahoo!での検索ヒット数および＜ meta ＞タグの中の’class_name’ は単語のメタデータ名を示す．さらに，図 4.12中の’id’ が 1 であり’name’ が三菱重工業である単語には，’total_hits’ は表示されていない．これは，三菱重工業という単語は，検索結果ではなく，検索に用いたユーザの興味モデル内のノードであることを表す．

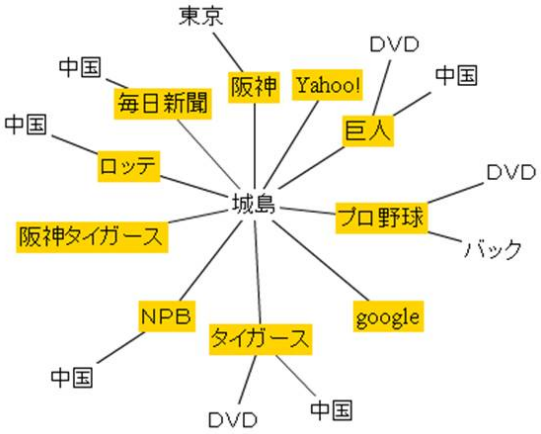


図 4.13: 同じ単語が出力されるマップ

前述した，全ての出力決定タイプに共通する処理を行い，XML が出力されると，本システムは Flash で XML を受け取り，出力決定タイプに基づき処理を行う．それぞれの出力決定タイプは，以下のように処理を行う．

- 出力決定タイプ 1 で行う処理
  1. 出力された XML の中から，'count' が 1 である単語を抽出する
  2. 抽出された単語の'total_hits' に基づき，単語をソートする
  3. 'total_hits' が小さいものから順に 10 個を抽出する
- 出力決定タイプ 2 で行う処理
  1. 出力された XML の中から，'count' が 1 である単語を抽出する
  2. 抽出された単語の'total_hits' に基づき，単語をソートする
  3. 'total_hits' が大きいものから順に 10 個を抽出する
- 出力決定タイプ 3 で行う処理
  1. 出力された XML の中から，'count' が 2 以上である単語を抽出する
  2. 抽出された単語の'total_hits' に基づき，単語をソートする
  3. 'total_hits' が小さいものから順に 10 個を抽出する
- 出力決定タイプ 4 で行う処理
  1. 出力された XML の中から，'count' が 2 以上である単語を抽出する
  2. 抽出された単語の'total_hits' に基づき，単語をソートする
  3. 'total_hits' が大きいものから順に 10 個を抽出する

ここで，上記の処理通りに出力すると，図 4.13に示すように，同じ単語ばかり出てしまうという現象があった．これは，製品名および地名といった一般的な単語は，どのような単語とも関連しやすく，出力されやすいということが原因であった．これらは同じ単語ではあるが，繋がる単語が異なるため，異なる意味を持っている．しかし，本研究では，製品名および地名を出力せず，人名および団体名といった固有名詞を出力することとした．予備実験を行ったところ，製品名および地名といった一般的な単語に比べ，人名および団体名といった固有名詞のほうが，ユーザの興味を喚起させやすいことがわかったからである．また，一般名詞より固有名詞のほうが，興味因子として望ましいという沢井らの知見 [69] もあるからである．

図 4.14: 情報抽出タイプ 1 の抽出結果

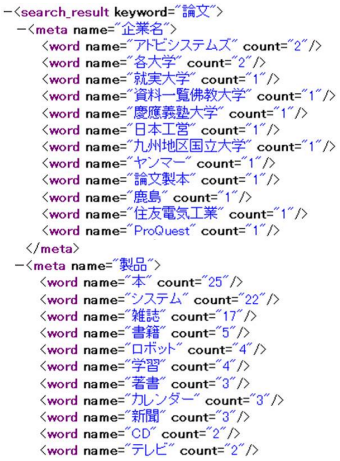


図 4.15: 論文に関連する単語 (1)

図 4.16: 情報抽出タイプ 2 の抽出結果

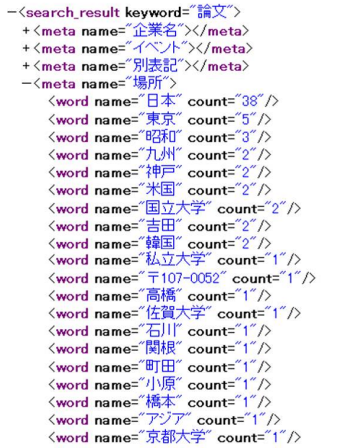


図 4.17: 論文に関連する単語 (2)

4.2.3 動作確認

本項では，構築した外部情報整理システムの動作確認について述べる．確認する項目は，

(a) 各情報抽出タイプで，インターネット上から関連する単語を正しく取得することができること

(b) 各出力決定タイプで，正しく出力を決定できていること

の 2 つである．

まず，(a) について動作確認を行った．確認方法は，

1. 検索に必要なパラメータを統一し，各情報抽出タイプで検索を行った
2. 検索結果の XML を取得した
3. 関連する単語を調べる API を用いて，関連する結果の XML を取得した
4. 2 つの XML を照合し，正しく関連する単語を抽出できていることを確認した

とした．この動作確認では，図 4.7に示すユーザの興味モデルを使用した．入力するキーワードは,「論文」とした．また，検索に使用するメタデータを団体名とした．すなわち，興味モデル内から検索に用いる単語は,「慶應義塾大学」および「キーウォーカー」の 2 つである．この動作確認の結果を以下に示す．

図 4.18: 慶應義塾大学に関連する単語

図 4.19: 情報抽出タイプ 3 の抽出結果

表 4.3: 各出力決定タイプの想定する出力

出力決定タイプ	想定する出力
出力決定タイプ1	長瀬智也，内田恭子，松井秀喜，タイガース，中村紀洋，近藤春菜，亀梨和也，ゴルフダイジェスト・オンライン，九州朝日放送，杉山愛
出力決定タイプ2	大和，天皇，近鉄，松井秀喜，タイガース，亀梨和也，ぴあ，JR，九州朝日放送，ゴルフダイジェスト・オンライン
出力決定タイプ3	日本代表，水川あさみ，フジテレビ，松岡昌宏，西武，レッドソックス，イチロー，メジャーリーグ，イチロー選手，阪神タイガース
出力決定タイプ4	TBS，プロ野球，MLB，西武，ソフトバンク，フジテレビ，結婚式，テレビ朝日，阪神，巨人

- 情報抽出タイプ 1

  動作確認の結果は，図 4.14に示す結果となった．このタイプは,「論文」に関連する単語のうち,ユーザの興味モデル内に存在する単語を抽出する.「論文」に関連する単語を抽出した XML を図 4.15に示す．図 4.15と図 4.7を比較すると,「論文」に関連する単語のうち，ユーザの興味モデル内に存在する,「慶應義塾大学」および「ロボット」を抽出していることがわかる．したがって，正しく情報抽出が行われていることを確認した．

- 情報抽出タイプ 2

  動作確認の結果は，図 4.16に示す結果となった．このタイプは,「論文」に関連し「慶應義塾大学」に関連する単語および「論文」に関連し「キーウォーカー」に関連する単語を抽出する．ここでは,「論文」に関連し「慶應義塾大学」に関連する単語が正しく抽出されていることを示す.「論文」に関連する単語を抽出した XML を図 4.17に示し,「慶應義塾大学」に関連する単語を抽出した XML を図 4.18に示す．図 4.17と図 4.18を比較すると,「論文」に関連し「慶應義塾大学」に関連する,「アジア」および「日本」を抽出していることがわかる．したがって，正しく情報抽出が行われていることを確認した．

- 情報抽出タイプ 3

  動作確認の結果は，図 4.19に示す結果となった．このタイプは,「論文」に関連しユーザの興味モデル内に存在する単語,「論文」に関連し「慶應義塾大学」に関連する単語および「論文」に関連し「キーウォーカー」に関連する単語を抽出する．すなわち，情報抽出タイプ 1 および情報抽

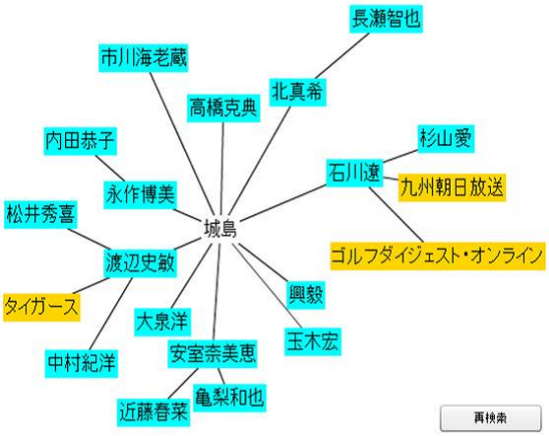


図 4.20: 出力決定タイプ 1 の結果

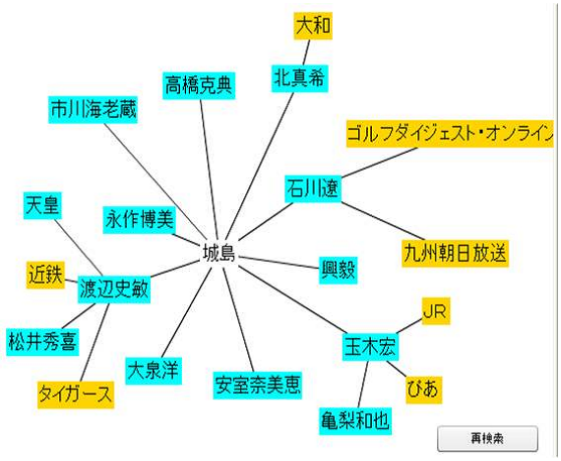


図 4.21: 出力決定タイプ 2 の結果

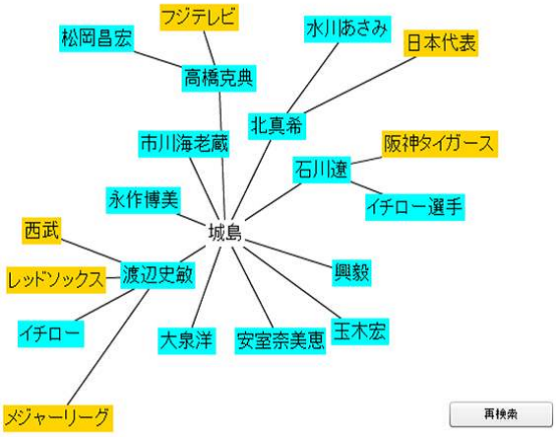


図 4.22: 出力決定タイプ 3 の結果

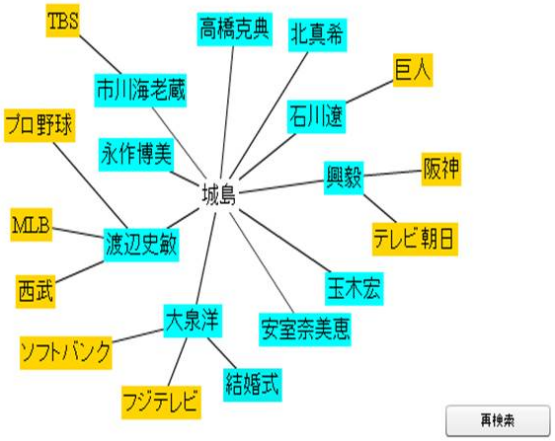


図 4.23: 出力決定タイプ 4 の結果

出タイプ 2 が混合された結果となる．図 4.19は図 4.14および図 4.16を混合した結果となっている．したがって，正しく情報抽出が行われていることを確認した．

つぎに，(b) について動作確認を行った．確認方法は，

1. 検索に必要なパラメータおよび情報抽出タイプを統一し，各出力決定タイプで検索を行った
2. 検索結果として出力される XML と表示される結果を比較し，正しく出力を決定できていることを確認した

とした．動作確認に用いた興味モデルは，多くの情報が提示されるように任意に作成した興味モデルである．入力するキーワードは,「城島」とし，検索に用いるメタデータは人名とした．また．情報抽出のタイプは，情報抽出タイプ 2 を用いた．この条件で出力される XML から，各出力決定タイプで想定される出力は表 4.3に示すとおりである．この動作確認の結果を以下に示す．

- 出力決定タイプ 1
  このタイプは，図 4.20に示すように出力された．図 4.20中の最も外側に表示されている単語は，想定通りの単語となっている．したがって，正しく出力決定ができていることを確認した．
- 出力決定タイプ 2
  このタイプは，図 4.21に示すように出力された．図 4.21中の最も外側に表示されている単語は，想定通りの単語となっている．したがって，正しく出力決定ができていることを確認した．

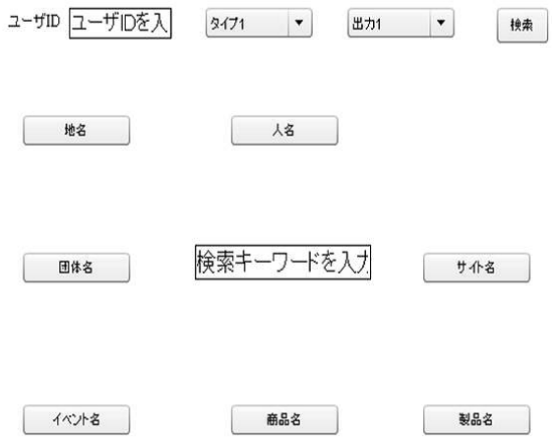


図 4.24: 入力部の画面

- 出力決定タイプ 3
  このタイプは，図 4.22に示すように出力された．図 4.22中の最も外側に表示されている単語は，想定通りの単語となっている．したがって，正しく出力決定ができていることを確認した．

- 出力決定タイプ 4
  このタイプは，図 4.23に示すように出力された．図 4.23中の最も外側に表示されている単語は，想定通りの単語となっている．したがって，正しく出力決定ができていることを確認した．

4.3 ユーザインターフェース

本節では，ユーザとシステムのインタラクションを司るユーザインターフェースが行う処理について述べる．本研究では，ユーザインターフェースの開発環境として，Adobe Flash CS3 Professional を用いた．4.3.1項では，情報提示に必要な情報を入力するための入力部で行う処理について述べる．4.3.2項では，結果を提示するための提示部で行う処理について述べる．

4.3.1 入力部

本項では，図 3.19に示す，入力部が有する機能を満たすための処理について述べる．前述したように，本システムでは，ユーザ ID，情報抽出のタイプ，出力決定のタイプ，情報提示に使用するメタデータの種類およびキーワードの 5 種類のパラメータを入力する必要がある．入力部では，以下のようにそれぞれのコンポーネントを構築した．

- ユーザ ID の入力
  ユーザ ID を入力するためのテキストボックスを配置した．テキストボックスには，ユーザ ID の入力を促すように,「ユーザ ID を入力」という言葉が書かれている．

- キーワードの入力
  キーワードを入力するためのテキストボックスを配置した．テキストボックスには，キーワードの入力を促すように,「検索キーワードを入力」という言葉が書かれている．

- 情報抽出のタイプの選択
  情報抽出のタイプを選択するためのプルダウンメニューを配置した．情報抽出タイプは 3 種類あるため，それぞれタイプ 1，タイプ 2 およびタイプ 3 と表示した．

- 出力決定のタイプの選択
  出力決定のタイプを選択するためのプルダウンメニューを配置した．出力決定のタイプは 4 種類あるため，それぞれ出力 1，出力 2，出力 3 および出力 4 と表示した．

- 情報提示に使用するメタデータの種類の選択
  メタデータを選択するためのボタンを，キーワードを入力するテキストボックスの周りに配置した．本研究では，人名，団体名，製品名，商品名，地名，サイト名およびイベント名の 7 種類を用いるため，7 つのボタンを配置した．

また，入力部には，検索ボタンを配置した．この検索ボタンをクリックすると，入力されたパラメータを用いて，情報抽出を開始し，画面が提示部へと切り替わる．
実際に構築した入力部の画面を図 4.24に示す．

4.3.2 提示部

本項では，図 3.20に示す，提示部が有する機能を満たすための処理について述べる．3.5.2項で述べたように，本システムでは，情報抽出タイプによって，提示される結果が異なる．提示部は，各情報抽出タイプにより，以下のように処理を行う．

- 情報抽出タイプ 1

  1. ユーザが入力したキーワードを配置する
  2. 出力決定部で決定された単語をキーワードの周りに配置するように，単語に座標情報を与える
  3. 単語間のつながりに基づき，単語同士を線で結ぶ
  4. マップを表示する

- 情報抽出タイプ 2

  1. ユーザが入力したキーワードを配置する
  2. 検索に用いたユーザの興味モデル内のノードをキーワードの周りに配置するように，単語に座標情報を与える
  3. 出力決定部で決定された単語をユーザの興味モデル内のノードの周りに配置するように，単語に座標情報を与える
  4. 単語間のつながりに基づき，単語同士を線で結ぶ
  5. マップを表示する

- 情報抽出タイプ 3

  1. ユーザが入力したキーワードを配置する
  2. 検索に用いたユーザの興味モデル内のノードをキーワードの周りに配置するように，単語に座標情報を与える
  3. 提示するユーザの興味モデル内のノードをキーワードの周りに配置するように，単語に座標情報を与える
  4. 出力決定部で決定された単語をユーザの興味モデル内のノードの周りに配置するように，単語に座標情報を与える
  5. 単語間のつながりに基づき，単語同士を線で結ぶ
  6. マップを表示する

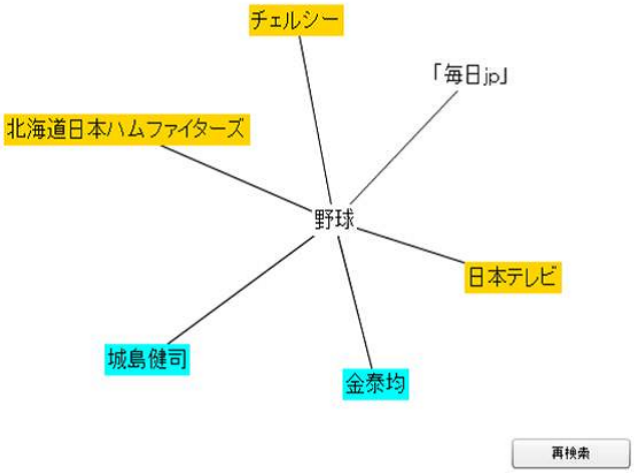


図 4.25: 情報抽出タイプ 1 の出力例

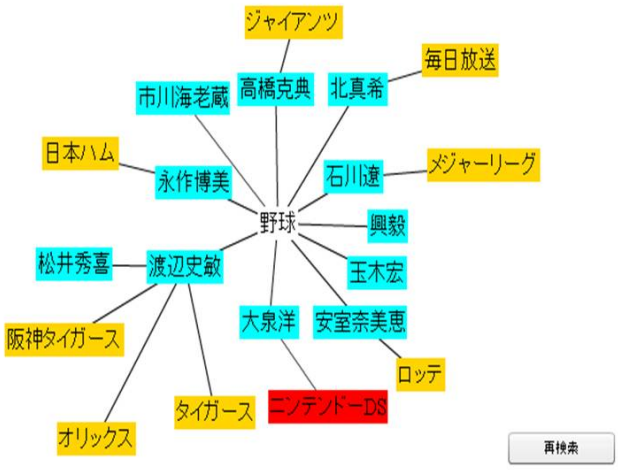


図 4.26: 情報抽出タイプ 2 の出力例

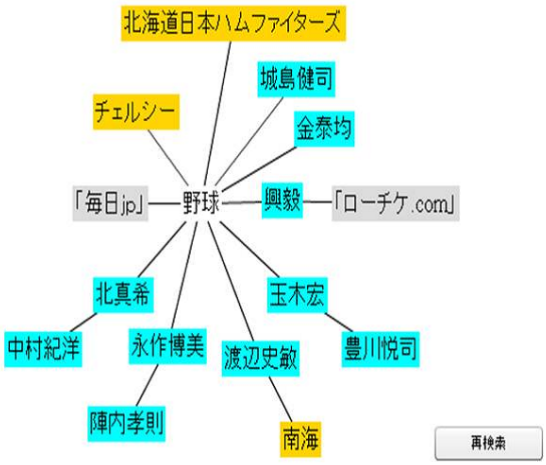


図 4.27: 情報抽出タイプ 3 の出力例

前述したように，情報抽出タイプ 1 では，ユーザが入力したキーワードの周りに出力決定部で決定された単語を配置する．単語の座標は，

$$X_{output} = X_{keyword} + R\cos\theta_{output} \tag{4.2}$$

$$Y_{output} = Y_{keyword} + R\sin\theta_{output} \tag{4.3}$$

で計算する．ここで，X_{output} は出力決定部で決定された単語の x 座標，Y_{output} は出力決定部で決定された単語の y 座標，$X_{keyword}$ はユーザが入力したキーワードの x 座標，$Y_{keyword}$ はユーザが入力したキーワードの y 座標，R は半径および θ_{output} は単語を配置する角度を表す．角度の単位は，rad である．本研究では，$X_{keyword}$ を 200，$Y_{keyword}$ を 200 および R を 100 と定義した．また，θ_{output} は，表示する単語数によって変化させた．情報抽出タイプ 2 では，ユーザが入力したキーワードの周りに，検索に用いたユーザの興味モデル内のノードを配置し，その外に出力決定部で決定された単語を配置する．それぞれ単語の座標は，

$$X_{model} = X_{keyword} + R_{model}\cos\theta_{model} \tag{4.4}$$

$$Y_{model} = Y_{keyword} + R_{model}\sin\theta_{model} \tag{4.5}$$

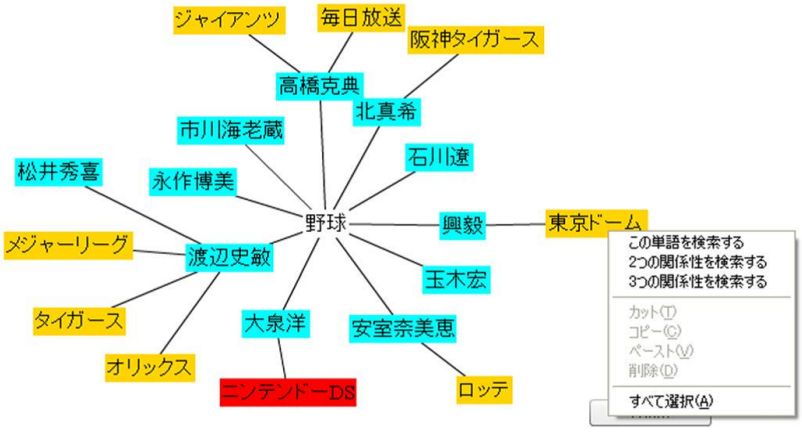


図 4.28: 実装した検索機能

$$X_{output} = X_{keyword} + R_{output} \cos\theta_{output} \tag{4.6}$$

$$Y_{output} = Y_{keyword} + R_{output} \sin\theta_{output} \tag{4.7}$$

で計算する．ここで，X_{model} は検索に用いたユーザの興味モデル内のノードの x 座標，Y_{model} は検索に用いたユーザの興味モデル内のノードの y 座標，R_{model} は興味モデル内のノードを配置する半径，R_{output} は出力決定部で決定された単語を配置する半径および θ_{model} は興味モデル内のノードを配置する角度を表す．本研究では，R_{model} を 100 および R_{output} を 200 と定義した．また，θ_{model} は，単語と単語の間が $\frac{1}{5}\pi$ となるようにした．情報抽出タイプ 3 では，ユーザが入力したキーワードの周りに，提示するユーザの興味モデル内のノードおよび検索に用いたユーザの興味モデル内のノードを配置し，その外に出力決定部で決定された単語を配置する．提示するユーザの興味モデル内のノードおよび検索に用いたユーザの興味モデル内のノードの座標は，4.4および 4.5式を用いて計算する．また，出力決定部で決定された単語の座標は，4.6および 4.7式を用いて計算する．

各情報抽出タイプで，実際に出力される情報提示結果を図 4.25，図 4.26および図 4.27にそれぞれ示す．

提示部には，再検索を行うためのボタンを配置した．このボタンをクリックすると，画面が入力部へ切り替わり，続けて再検索を行うことができるようになる．

3.5.2項で述べたように，本研究では，ユーザが興味を持った対象の検索補助を行う．そのため，図 4.28に示すように,「この単語を検索する」,「2 つの関係性を検索する」および「3 つの関係性を検索する」という 3 つの機能を実装した．以下で，3 つの機能の処理について述べる．

- この単語を検索する

  これは，提示された単語自体に，ユーザが興味を持った場合に使用する機能である．この機能を満たすために，

  1. ユーザが興味を持った単語をクリックする
  2. システムは，クリックされた単語のテキストを取得する
  3. 取得したテキストを検索キーワードとして，Google で検索を行う
  4. 検索結果を表示する

  という処理を行う．

- 2 つの関係性を検索する

  これは，ユーザが，提示された単語と入力したキーワードの関係性に興味を持った場合に使用する機能である．この機能を満たすために，

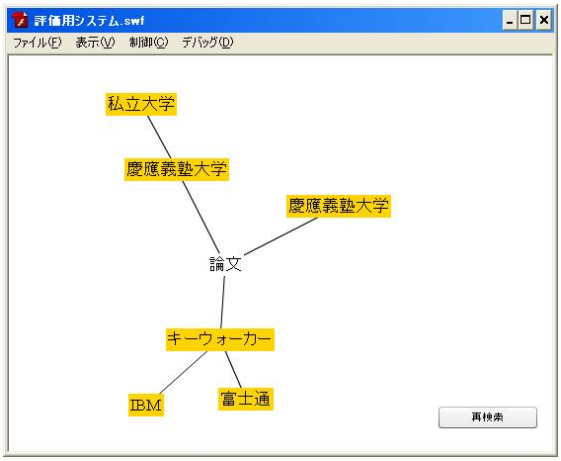


図 4.29: 動作確認に用いた提示結果

図 4.30: 「この単語を検索する」の検索結果

1. ユーザは，提示された単語の中で興味を持った単語をクリックする
2. システムは，クリックされた単語のテキストを取得する
3. 取得したテキストおよび入力部でユーザが入力したキーワードを検索キーワードとして，Google で AND 検索を行う
4. 検索結果を表示する

という処理を行う．

- 3 つの関係性を検索する
  これは，ユーザが，提示された単語と興味モデル内のノードと入力したキーワードの関係性に興味を持った場合に使用する機能である．この機能を満たすために，

1. ユーザは，提示された単語の中で興味を持った単語をクリックする
2. システムは，クリックされた単語のテキストを取得する
3. クリックされた単語につながっている興味モデル内のノードのテキストを取得する
4. 取得した 2 つのテキストおよび入力部でユーザが入力したキーワードの 3 つを検索キーワードとして，Google で AND 検索を行う

という処理を行う．

4.3.3 動作確認

本項では，構築したユーザインターフェースの動作確認について述べる．入力部および提示部は，4.2.3項で示したように，検索に必要なパラメータが正しく取得され，正しく結果が表示される．そのため，確認する項目は，ユーザとシステムのインタラクションである

(a) 「この単語を検索する」をクリックすると正しく検索を行うことができるかどうか

(b) 「2 つの関係性を検索する」をクリックすると正しく検索を行うことができるかどうか

(c) 「3 つの関係性を検索する」をクリックすると正しく検索を行うことができるかどうか

図 4.31: 「2 つの関係性を検索する」の検索結果

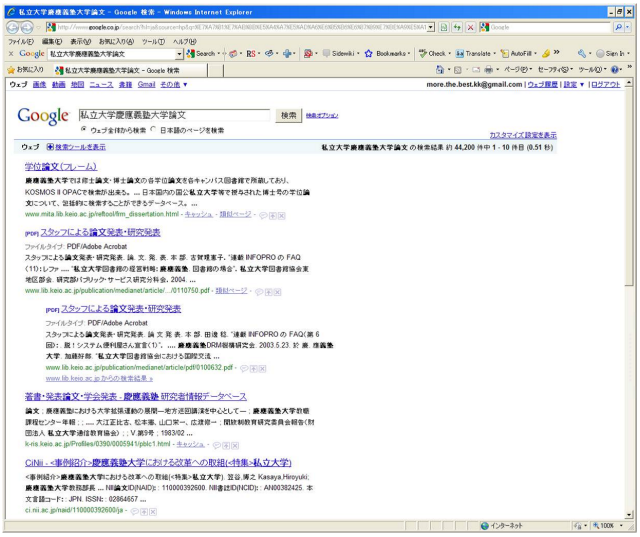


図 4.32: 「3 つの関係性を検索する」の検索結果

の 3 つとする．この動作確認では，図 4.29に示す提示結果を用いて確認を行った．

まず (a) について動作確認を行った．確認方法は，

1. 検索を行う対象を決めた
2. 「この単語を検索する」をクリックした
3. Google で正しく検索が行われていることを確認した

とした．検索の対象は，慶應義塾大学とした．図 4.30に示すように，慶應義塾大学が検索されている．したがって，正しく動作していることを確認した．

つぎに，(b) について動作確認を行った．確認方法は，

1. 検索を行う対象を決めた
2. 「2 つの関係性を検索する」をクリックした
3. Google で正しく検索が行われていることを確認した

とした．検索の対象は，慶應義塾大学とした．図 4.31に示すように，慶應義塾大学および入力したキーワードである論文が検索されている．したがって，正しく動作していることを確認した．

最後に，(c) について動作確認を行った．確認方法は，

1. 検索を行う対象を決めた
2. 「3 つの関係性を検索する」をクリックした
3. Google で正しく検索が行われていることを確認した

とした．検索の対象は，私立大学とした．図 4.32に示すように，私立大学，私立大学と繋がりを持つ慶應義塾大学および入力したキーワードである論文が検索されている．したがって，正しく動作していることを確認した．

第5章　思考支援システムの検証

本章では，構築した思考支援システムの検証について述べる．この検証では，本研究で提案したシステムを構成するサブシステムの有効性を示す実験を行う．まず，ユーザの興味モデルを構築するために重要な Web ページの解析の精度および E-mail の解析の精度の検証について述べる．つぎに，情報抽出タイプおよび出力決定タイプが興味拡大に与える影響の検証について述べる．さらに，構築した 3 つのサブシステムの有効性の検証について述べる．最後に，ユーザの興味が興味拡大に与える影響の検証について述べる．5.1節では，ユーザモデル構築システムの精度を検証するために行った実験について述べる．5.2節では，情報抽出タイプおよび出力決定タイプが興味拡大に与える影響を検証するために行った実験について述べる．5.3節では，構築したサブシステムの有効性を示すために行った実験について述べる．5.4節では，ユーザの興味が興味拡大に与える影響を検証するために行った実験について述べる．

5.1　ユーザモデル構築システムの精度の検証

本節では，ユーザモデル構築システムの解析によって抽出される単語の取得精度の検証について述べる．本研究では，Web ページおよび E-mail を用いて，ユーザの興味モデルを構築した．そのため，本節では，Web ページおよび E-mail の解析精度の検証について述べる．5.1.1項では，Web ページおよび E-mail の解析精度を検証する実験の目的について述べる．5.1.2項では，Web ページおよび E-mail の解析精度の検証に行った実験の方法および検証項目の検証方法について述べる．5.1.3項では，得られた結果および結果から得られた考察について述べる．

5.1.1　実験目的

本項では，Web ページおよび E-mail の解析精度を検証する実験の目的について述べる．この実験は，

(a) ユーザの興味の中のメタデータ含有率

(b) 本研究で用いたメタデータが付加される単語の取得精度

の 2 つを検証するために行った．ユーザの興味の中のメタデータ含有率を求めることで，本研究で用いた人名，団体名，製品名，商品名，地名，サイト名およびイベント名の 7 種類のメタデータで，ユーザの興味をどれだけモデル化することができるかを検証する．また，メタデータが付加される単語の取得精度を求めることで，本研究で用いた 7 種類のメタデータが付加され，取得されるべき単語の取得精度を検証する．

5.1.2　実験方法

本項では，Web ページおよび E-mail の解析精度の検証に行った実験の方法について述べる．
まず，Web ページの解析精度の検証に行った実験方法について述べる．実験は，以下の方法で行った．

1. 被験者に閲覧する Web ページを決定させた

2. その Web ページの中で被験者の興味のある名詞を抜き出させ，それらの関係性をマップに書かせた

表 5.1: Web ページおよび E-mail の解析精度

	メタデータの含有率 [%]	単語取得率 [%]
	平均値 (標準偏差)	平均値 (標準偏差)
Webページ	57.59 (16.17)	68.52 (8.34)
E-mail	46.58 (15.29)	60.18 (12.01)

3. 被験者が作成するマップは，名詞を頂点とし，マインドマップを模倣した形とした

4. 被験者には，15 分間自由に Web ページの閲覧をさせた

この実験では，興味の対象には一切の制限を設けず，自由にマップを作成させた．

つぎに，E-mail の解析精度の検証に行った実験方法について述べる．実験は，以下の方法で行った．

1. 被験者に興味がある内容を含む E-mail を決定させた

2. その E-mail の中で被験者の興味のある名詞を抜き出させ，それらの関係性をマップに書かせた

3. 被験者が作成するマップは，名詞を頂点とし，マインドマップを模倣した形とした

4. 被験者には，5 通の E-mail を選択させた

この実験では，興味の対象には一切の制限を設けず，自由にマップを作成させた．本研究では，日本語の文章しか解析を行えないため，英語で書かれた E-mail は選択しないように指示した．また，機密性の高い E-mail を選択しないように指示した．

2 つの検証項目は，それぞれ以下のように検証する．

(a) ユーザの興味の中のメタデータ含有率

まず，被験者が作成したマップの中の全単語に人手でメタデータを付加する．つぎに，本研究で用いる 7 種類のメタデータが付加される単語を抽出する．最後に，3.1式を用いて，メタデータの含有率を計算する．

(b) 本研究で用いたメタデータが付加される単語の取得精度

まず，被験者が作成したマップの中から，7 種類のメタデータが付加される単語を抽出する．つぎに，被験者のマップの中から抽出した単語と構築された被験者の興味モデル内の単語を照合し，一致した単語を抽出する．最後に，

$$R_{match} = \frac{N_{system}}{N_{META}} \tag{5.1}$$

を用いて，被験者のマップの中から抽出した単語と構築された被験者の興味モデル内の単語の一致率を計算する．ここで，R_{match} は単語の一致率，N_{system} はシステムが取得した単語の数および N_{META} は被験者が作成したマップの中で 7 種類のメタデータが付加された単語の数を示す．

被験者は 20 代から 30 代の男女 9 人である．

5.1.3 結果および考察

本項では，行った実験より得られた結果および結果から得られた考察について述べる．この実験で得られた結果を表 5.1に示す．

以下に，2 つの検証項目についての考察を述べる．

(a) ユーザの興味の中のメタデータ含有率

まず，Web ページの解析精度の検証実験から得られた実験結果および考察について述べる．

表 5.1に示すように，7 種類のメタデータを用いることで，ユーザの興味の 57.59%を表せることがわかる．3.3.1項で述べたように，予備実験では，7 種類のメタデータを用いることで，ユーザの興味の 57.1%を表せた．したがって，本実験では，ユーザの興味を予備実験と同程度モデル化することができたと言える．

ユーザの興味を 100%表すためには，膨大なメタデータを用意する必要がある．被験者が作成したマップの中には,「噴火」といった現象を表すメタデータや「スポーツ」といった分野を表すメタデータなど様々なメタデータが含まれていた．しかし，本研究で用いた 7 種類のメタデータ以外は，人によりばらつきが大きく，モデル化に有効なメタデータを見つけることができなかった．したがって，ユーザの興味をモデル化するためには，少なくとも 7 種類のメタデータは有効であることがわかった．ユーザの興味をよりモデル化するためには，検証を引き続き行い，ユーザの興味を表現するために有効なメタデータを見つける必要がある．

つぎに，E-mail の解析精度の検証実験から得られた実験結果および考察について述べる．

表 5.1に示すように，7 種類のメタデータを用いることで，ユーザの興味の 46.58%を表せることがわかる．

E-mail の場合も，Web ページの解析と同様の考察が考えられる．ただし，E-mail の場合は，3.3.1項で述べたアンケート結果のように，日時を表すメタデータを書く被験者が多かった．本研究では，日時はユーザの興味拡大には影響しにくいと考え，対象外として扱った．そのため，メタデータの含有率も低下したと考えられる．

また，被験者が作成したマップの中には，添付ファイル名といった E-mail のやり取り特有の名詞が含まれていた．これらは，ユーザの興味を抱かせ，E-mail の内容を把握するためには重要な要素であるが，インターネット上から関連する単語を抽出することができない．さらに，たとえ同じファイル名に関連する単語を抽出したところで，その単語は，ユーザの興味を拡大するきっかけにはなりにくい．これらの単語を除外し，ユーザの興味拡大のためのメタデータと考えると，メタデータの含有率は向上する．したがって，ユーザの興味拡大に必要な，E-mail を情報源とした興味モデルの構築には，7 種類のメタデータは有効であると考えられる．

(b) 本研究で用いたメタデータが付加される単語の取得精度

まず，Web ページの解析精度の検証実験から得られた実験結果および考察について述べる．

表 5.1に示すように，7 種類のメタデータが付加され，取得されるべき単語のうち，68.52%の単語をシステムは取得できることがわかる．

本研究では，書かれた文章の構造に基づくルールを構築することで，単語を取得し，メタデータを付加する．そのため，文章の構造が異なる場合は，正しく単語を取得することができない．KW 社で構築しているルールは，多くの Web ページの文書構造を参考として，汎用性の高いものとなっている．しかし，Web ページで書かれている文章の構造は様々であるため，100%正確に解析することは困難である．単語の取得精度を向上させるためには，より汎用性が高いルールを構築する必要があると考えられる．

つぎに，E-mail の解析精度の検証実験から得られた実験結果および考察について述べる．

表 5.1に示すように，7 種類のメタデータが付加され，取得されるべき単語のうち，60.18%の単語をシステムは取得できることがわかる．

E-mail の解析では，Web ページの解析に比べて単語の取得精度が低い．これは，E-mail の文書構造が Web ページの文書構造と異なるためであると考えられる．本研究では，Web ページの解析も E-mail の解析も，同様のルールを用いた．そのため，E-mail のような書く人により文書構造が

異なる文章を正確に解析することができず，Web ページの解析の単語取得に比べて，E-mail の解析の単語の取得は精度が低くなったと考えられる．E-mail の単語取得精度の向上のためには，決まった文書構造がない E-mail を解析することが可能なルールを構築する必要があると考えられる．

5.2 情報抽出タイプおよび出力決定タイプによる影響の検証

本節では，3 種類の情報抽出タイプと 4 種類の出力決定タイプを組み合わせた 12 パターンを比較し，情報抽出タイプおよび出力決定タイプによる興味拡大への影響を検証するために行った官能評価実験について述べる．5.2.1項では，官能評価実験の目的および評価項目について述べる．5.2.2項では，官能評価の実験方法について述べる．5.2.3項では，実験から得られた結果および結果から得られた考察について述べる．

5.2.1 実験目的

本項では，官能評価実験の目的および評価項目について述べる．官能評価とは，人の主観的な感覚に依存する計量や評価のことであり，JIS では,「官能検査 (評価) とは，人間の感覚を用いて品質特性を評価し，判定基準と照合して判定を下す検査のことである」と定義されている [70]．官能評価を行う際には，疲労感は被験者の心理状態に影響を与えるため，被験者に疲労感を与えないようにしなければならない．そのため，疲労感を与えないことに留意し，本研究では官能評価を行った．

前述したように，この評価では，3 種類の情報抽出タイプと 4 種類の出力決定タイプが，結果に与える影響を調査する．そのため，情報抽出タイプと出力決定タイプを組み合わせた 12 パターンを比較し，評価を行う．

本研究は，ユーザの興味拡大を促し，新たな知識を獲得させることで，ユーザの思考支援を目的としている．そのため，評価項目は，

(a) 興味が拡大されたように感じたか

(b) 気になる結果が提示されたように感じたか

(c) AND 検索には使用しない組み合わせが提示されたように感じたか

(d) 新たな知識が増えたように感じたか

(e) 自分にとって意外な結果が出てきたように感じたか

(f) 提示される結果に満足したか

(g) システムに面白みを感じたか

の 7 つとした．(a) は興味拡大に関する評価項目，(b) および (c) は興味を拡大するためのきっかけに関する評価項目，(d) は知識獲得に関する評価項目，(e) は興味拡大に重要とされている意外性に関する評価項目，(f) および (g) はシステムの継続利用に関わる満足度および面白さに関する評価項目である．これら 7 つの評価項目に対して，非常に感じた，感じた，やや感じた，どちらでもない，やや感じない，感じないおよび全く感じないの 7 段階で被験者に評価させた．また，12 パターンの順番が評価に影響しないように，被験者によって，12 パターンの選択の順番をランダムとした．実験に用いた評価シートを付録として掲載する．

表 5.2: 12 パターンを比較した実験結果

	1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)
評価項目(a)	-1.22 (0.30)	-1.53 (0.34)	-0.76 (0.58)	-0.96 (0.23)	0.92 (0.36)	0.68 (0.44)	0.70 (0.66)	0.69 (0.56)	0.97 (0.59)	-0.11 (0.55)	0.24 (0.56)	0.38 (0.79)
評価項目(b)	-1.05 (0.41)	-1.27 (0.51)	-0.96 (0.36)	-1.00 (0.46)	0.68 (0.47)	0.76 (0.35)	0.83 (0.42)	0.62 (0.70)	1.02 (0.42)	-0.24 (1.01)	0.35 (0.60)	0.28 (0.79)
評価項目(c)	-0.50 (0.68)	-1.10 (0.55)	-0.93 (0.59)	-0.79 (0.55)	0.59 (0.60)	0.64 (0.43)	0.75 (0.67)	0.45 (0.68)	0.83 (0.49)	-0.16 (0.90)	0.01 (0.73)	0.20 (0.79)
評価項目(d)	-0.44 (0.89)	-1.61 (1.01)	-0.49 (0.93)	-0.49 (0.93)	0.23 (0.78)	0.36 (0.65)	0.79 (0.86)	0.05 (0.70)	0.53 (0.80)	-0.10 (0.48)	-0.01 (0.59)	-0.02 (1.04)
評価項目(e)	-0.75 (0.76)	-1.05 (1.00)	-1.02 (0.71)	-1.05 (0.30)	0.68 (0.65)	0.49 (0.59)	0.74 (0.49)	0.55 (0.78)	0.78 (0.73)	0.28 (0.73)	0.11 (0.76)	0.23 (0.69)
評価項目(f)	-0.80 (0.41)	-1.10 (0.77)	-1.17 (0.34)	-1.00 (0.30)	0.72 (0.60)	0.55 (0.71)	1.13 (0.40)	0.72 (0.83)	0.49 (0.90)	0.37 (0.74)	-0.21 (0.46)	0.30 (0.65)
評価項目(g)	-0.77 (0.75)	-1.28 (0.49)	-1.11 (0.47)	-1.02 (0.41)	0.79 (0.55)	0.46 (0.70)	0.81 (0.54)	0.35 (0.68)	0.63 (0.93)	0.30 (0.68)	0.23 (0.72)	0.61 (0.61)

5.2.2 実験方法

本項では，官能評価で行った実験の方法について述べる．実験は，以下の方法で行った．

1. 被験者に，検索に用いるメタデータ，キーワード，情報抽出タイプおよび出力決定タイプを決定させた
2. 提示される結果を見て，気になった単語を検索させた
3. 1 つのパターンにつき，2 回以上キーワードを変えて，検索するように指示した
4. 12 パターンそれぞれで検索を行わせた
5. それぞれのパターンを比較し，評価させた

被験者には，実験を行う前に，15 分程度のネットサーフィンと興味ある 5 通の E-mail を用いて，興味モデルを構築させた．本実験では，このあらかじめ構築させた興味モデルを用いて，実験を行った．被験者は，20 代から 30 代の男女 8 人である．

5.2.3 結果および考察

本項では，官能評価で得られた実験結果および結果から得られた考察について述べる．

まず，官能評価で用いた非常に感じた，感じた，やや感じた，どちらでもない，やや感じない，感じないおよび全く感じないをそれぞれ 3，2，1，0，-1，-2 および-3 と数値化し，得られた結果を正規化した．正規化して得られたデータより，各評価項目に対しての平均得点および標準偏差を算出した．実験により得られた各評価対象の各評価項目に対する平均得点および標準偏差値を表 5.2に示す．表中の「1-1」という表記は，情報抽出タイプと出力決定タイプを示す．すなわち,「1-1」は，情報抽出タイプが 1 であり，出力決定タイプが 1 であることを示す．表 5.2に示すとおり，全てのパターンの平均得点に差が出ている．これらの差が統計的に有意な差であることを検証しなければならない．平均値に有意な差があることを確かめる統計的手法に，ウェルチ検定および t 検定といったものがある．しかし，複数の標本に対して，2 つの標本を検定する t 検定といった統計的手法を繰り返し用いると，全体の有意水準が大きくなってしまい，有意差が出る確率が上がってしまう．したがって，複数の標本に対して，検証するためには，全体の有意水準を上げないための多重比較法を用いる必要がある．多重比較法とは，3 つ以上の

	1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
1-1					**	**	**	**	**	*	**	**
1-2					**	**	**	**	**	**	**	**
1-3					**	**	**	**	**			*
1-4					**	**	**	**	**		**	**
2-1												
2-2												
2-3												
2-4												
3-1										*		
3-2												
3-3												
3-4												

図 5.1: 評価項目 (a) における検定結果

	1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
1-1					**	**	**	**	**		**	**
1-2					**	**	**	**	**		**	**
1-3					**	**	**	**	**		*	*
1-4					**	**	**	**	**		**	*
2-1												
2-2												
2-3												
2-4												
3-1										*		
3-2												
3-3												
3-4												

図 5.2: 評価項目 (b) における検定結果

	1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
1-1									*			
1-2					**	**	**	**	**			
1-3					**	**	**	*	**			*
1-4					*	*	**		**			
2-1												
2-2												
2-3												
2-4												
3-1												
3-2												
3-3												
3-4												

図 5.3: 評価項目 (c) における検定結果

	1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
1-1												
1-2												
1-3												
1-4												
2-1												
2-2												
2-3												
2-4												
3-1												
3-2												
3-3												
3-4												

図 5.4: 評価項目 (d) における検定結果

	1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
1-1					*		*		*			
1-2					**	*	**	*	**			
1-3					**	*	**	*	**			
1-4					**	*	**	*	**			
2-1												
2-2												
2-3												
2-4												
3-1												
3-2												
3-3												
3-4												

図 5.5: 評価項目 (e) における検定結果

	1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
1-1					**	*	**	**	*			
1-2					**	**	**	**	**	**		*
1-3					**	**	**	**	**	**		**
1-4					**	**	**	**	**	*		*
2-1												
2-2												
2-3												
2-4												
3-1												
3-2												
3-3												
3-4												

図 5.6: 評価項目 (f) における検定結果

複数の標本に対して，それぞれの差が有意であることを確かめる統計的手法である [71]．すなわち，多重比較法とは，全体としての有意水準を公称の値にコントロールできるように，1 回の検定における棄却限界値を調整する方法である．本研究では，多重比較法の中のボンフェローニ法を用いて検定を行う．

ボンフェローニ法とは，

$$Pr(\bigcup_{i=1}^{k} E_i) \leq \sum_{i=1}^{k} Pr(E_i) \tag{5.2}$$

に示すボンフェローニの不等式に基づく多重比較法のことである．式 5.2の左辺は，k 個の事象 E_i のうち少なくとも 1 つが成り立つ確率を表していて，右辺は各事象 E_i の成り立つ確率を加えたものである．この方法では，標本に含まれる帰無仮説はどのようなものでもよく，仮説間に複雑な数学的関数関係があってもよい．また，検定統計量の形は任意でよく，検定統計量間に複雑な数学的関係あるいは相関関係があってもよい．さらに，検定統計量に関する分布は個々の帰無仮説の下で対応する検定統計量の周

	1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
1-1					**		**		*			*
1-2					**	**	**	**	**	**	**	**
1-3					**	**	**	*	**	*	*	**
1-4					**	*	**	*	**	*		**
2-1												
2-2												
2-3												
2-4												
3-1										*		
3-2												
3-3												
3-4												

図 5.7: 評価項目 (g) における検定結果

表 5.3: 情報抽出タイプ別の平均得点および標準偏差

	情報抽出タイプ1	情報抽出タイプ2	情報抽出タイプ3
	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)
評価項目(a)	-1.12 (0.20)	0.75** (0.19)	0.37** (0.29)
評価項目(b)	-1.07 (0.28)	0.72** (0.18)	0.35** (0.38)
評価項目(c)	-0.83 (0.42)	0.61** (0.36)	0.22** (0.39)
評価項目(d)	-0.51 (0.93)	0.36 (0.55)	0.15 (0.46)
評価項目(e)	-0.96 (0.32)	0.61** (0.48)	0.35** (0.40)
評価項目(f)	-1.02 (0.27)	0.78** (0.35)	0.24** (0.44)
評価項目(g)	-1.04 (0.22)	0.6** (0.32)	0.44** (0.23)

辺分布さえわかっていればよく，複数の検定統計量の同時分布は必要ない．したがって，有意水準の調整を行うことにより，標本としての多重比較を個々の検定問題のように扱うことができる．

多重比較法を用いた検定の結果を，図 5.1から図 5.7に示す．図中の’**’ は 99%信頼区間において平均得点の差が統計的に有意であることを示し，’*’ は 95%信頼区間において平均得点の差が統計的に有意であることを示す．また，’*’ がないものは，平均得点の差が統計的に有意でないことを示す．

これらの結果をみると，情報抽出タイプ 1 の場合は，平均得点が低いことがわかる．そのため，表 5.3に示すように，情報抽出タイプ別に平均得点および標準偏差を算出した．これらの差が統計的に有意な差であることを多重比較法を用いて検定した．表中の’**’ は情報抽出タイプ 1 と比較した結果，99%信頼区間において平均得点の差が統計的に有意であることを示す．また，’*’ がないものは，情報抽出タイプ 1 と比較した結果，平均得点の差が統計的に有意でないことを示す．

情報抽出タイプ 1 および情報抽出タイプ 2 を比較した結果，(a),(b),(c),(e),(f) および (g) の評価項目で，99%信頼区間において平均得点の差が統計的に有意であることがわかった．また，情報抽出タイプ 1 および情報抽出タイプ 3 を比較した結果，(a),(b),(c),(e),(f) および (g) の評価項目で，99%信頼区間において平均得点の差が統計的に有意であることがわかった．

情報抽出タイプ 1 は，ユーザが入力したキーワードに関連する単語のうち，ユーザの興味モデル内に存在する単語を提示する．一方，情報抽出タイプ 2 および情報抽出タイプ 3 は，ユーザの興味モデル内に存在する単語を出力するわけではない．したがって，自分の興味モデルの中に存在する単語を提示するタイプは，興味モデル内に存在しない単語を提示するタイプに比べて，興味拡大をしにくく，システ

表 5.4: 出力決定タイプ別の平均得点および標準偏差

	出力決定タイプ1	出力決定タイプ2	出力決定タイプ3	出力決定タイプ4
	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)
評価項目(a)	0.22 (0.26)	-0.32** (0.23)	0.06 (0.20)	0.04 (0.18)
評価項目(b)	0.21 (0.20)	-0.25** (0.20)	0.07 (0.23)	-0.04 (0.22)
評価項目(c)	0.31 (0.16)	-0.20** (0.20)	-0.06 (0.31)	-0.05 (0.22)
評価項目(d)	0.11 (0.31)	-0.05 (0.15)	0.10 (0.32)	-0.15 (0.33)
評価項目(e)	0.24 (0.14)	-0.09 (0.37)	-0.06 (0.24)	-0.09 (0.25)
評価項目(f)	0.14 (0.41)	-0.06 (0.49)	-0.08 (0.31)	0.01 (0.16)
評価項目(g)	0.22 (0.42)	-0.17 (0.49)	-0.02 (0.27)	-0.02 (0.19)

ムに面白みを感じないことがわかった．

つぎに，表5.4に示すように，出力決定タイプ別に平均得点および標準偏差を算出した．これらの差が統計的に有意な差であることを多重比較法を用いて検定した．表中の’**’ は出力決定タイプ 1 と比較した結果，99%信頼区間において平均得点の差が統計的に有意であることを示す．また，’*’ がないものは，出力決定タイプ 1 と比較した結果，平均得点の差が統計的に有意でないことを示す．出力決定タイプ 1 と出力決定タイプ 2 を比較した結果，(a),(b) および (c) の評価項目で，99%信頼区間において平均得点の差が統計的に有意であることがわかった．出力決定タイプ 1 は，単語の社会的重要度が小さく，興味との関連度が小さい単語を提示する．一方，出力決定タイプ 2 は，単語の社会的重要度が大きく，興味との関連度が小さい単語を提示する．したがって，興味との関連度が小さい場合は，単語の社会的重要度が小さい場合に比べて，単語の社会的重要度が高いほうが，興味拡大をしにくいことがわかった．ただし，この実験を行っている際，出力決定タイプを変更しても，提示される結果の変化がわかりにくいという意見があった．そのため，興味との関連度が小さい場合は，単語の社会的重要度が小さい場合に比べて，単語の社会的重要度が高いほうが興味拡大をしにくいと断言するには，より詳細な検証を行う必要があると考えられる．

5.3 構築したサブシステムの有効性の検証

本研究で提案したシステムは，ユーザモデル構築システム，外部情報整理システムおよびユーザインターフェースの 3 つのサブシステムで構成される．本節では，3 つのサブシステムの有効性を示すための官能評価実験について述べる．5.3.1項では，官能評価実験の目的および評価項目について述べる．5.3.2項では，官能評価の実験方法について述べる．5.3.3項では，得られた実験結果および結果から得られた考察について述べる．

5.3.1 実験目的

本項では，官能評価実験の目的および評価項目について述べる．前述したように，本システムは，3 つのサブシステムで構成されている．これら 3 つのサブシステムが，興味拡大にそれぞれ有効であることを示すために，官能評価実験を行った．

本実験では，

- 本研究で提案するシステム

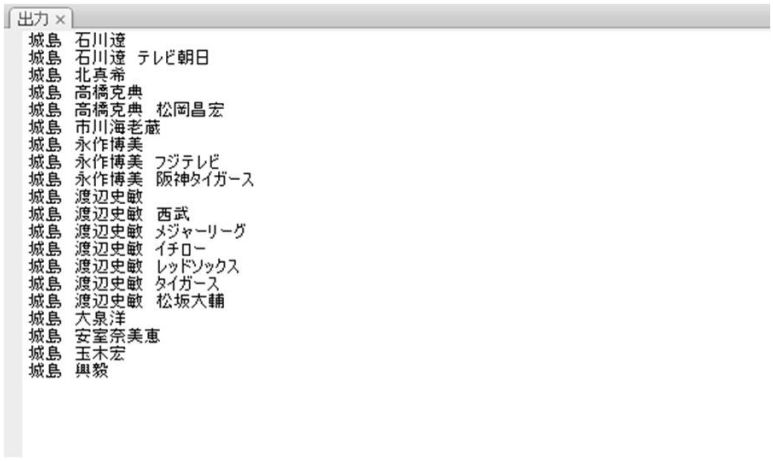


図 5.8: 情報提示方法を変更した本システム

- 本研究で提案するシステムの情報提示方法を図 5.8のように変更したシステム
- Yahoo!関連検索 [72]

の 3 つを比較することで，3 つのサブシステムの有効性を示す．Yahoo!関連検索は，ユーザの興味は考慮せず，入力したキーワードに関連する情報を，インターネット上から抽出し，出力結果をリストで提示するシステムである．本システムと情報提示方法を変更したシステムを比較することで，ユーザインターフェースが興味拡大に有効であることを示す．情報提示方法を変更したシステムと Yahoo!関連検索を比較することで，ユーザモデル構築システムおよび外部情報整理システムが興味拡大に有効であることを示す．

評価項目は，ユーザの興味拡大に与える影響を検証するため，5.2節で行った実験と同様，

(a) 興味が拡大されたように感じたか

(b) 気になる結果が提示されたように感じたか

(c) AND 検索には使用しない組み合わせが提示されたように感じたか

(d) 新たな知識が増えたように感じたか

(e) 自分にとって意外な結果が出てきたように感じたか

(f) 提示される結果に満足したか

(g) システムに面白みを感じたか

の 7 つとした．これら 7 つの評価項目に対して，非常に感じた，感じた，やや感じた，どちらでもない，やや感じない，感じないおよび全く感じないの 7 段階で被験者に評価させた．

5.3.2 実験方法

本項では，官能評価で行った実験の方法について述べる．実験は，以下の方法で行った．

1. 被験者に，本システムに必要な，検索に用いるメタデータ，キーワード，情報抽出タイプおよび出力決定タイプを決定させた
2. 本システムが提示する結果を見て，気になった単語を検索させた

表 5.5: サブシステムの有効性検証の実験結果

	本研究で提案したシステム	情報提示方法を変更したシステム	Yahoo!関連検索
	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)
評価項目(a)	0.98 (0.22)	0.09 (0.62)	-1.07 (0.67)
評価項目(b)	1.04 (0.24)	0.21 (0.43)	-1.26 (0.22)
評価項目(c)	0.89 (0.26)	0.41 (0.48)	-1.30 (0.23)
評価項目(d)	0.66 (0.41)	0.19 (0.63)	-0.85 (0.70)
評価項目(e)	0.95 (0.26)	0.07 (0.72)	-1.02 (0.69)
評価項目(f)	0.80 (0.61)	0.12 (0.53)	-0.93 (0.91)
評価項目(g)	1.18 (0.13)	-0.13 (0.53)	-1.05 (0.45)

3. 情報提示方法を変更したシステムにも，同様の操作を行わせた

4. Yahoo!関連検索に，本システムに入力したキーワードを入力させた

5. Yahoo!関連検索が提示する結果を見て，気になった単語を検索させた

6. キーワードは 2 つ以上使うように指示した

7. 3 つのシステムを比較し，評価させた

被験者には，実験を行う前に，15 分程度のネットサーフィンと興味ある 5 通の E-mail を用いて，興味モデルを構築させた．本実験では，このあらかじめ構築させた興味モデルを用いて，実験を行った．被験者は，20 代から 30 代の男女 8 人である．

5.3.3 結果および考察

本項では，官能評価実験で得られた結果および結果から得られた考察について述べる．

まず，官能評価で用いた非常に感じた，感じた，やや感じた，どちらでもない，やや感じない，感じないおよび全く感じないをそれぞれ 3，2，1，0，-1，-2 および-3 と数値化し，得られた結果を正規化した．正規化して得られたデータより，各評価項目に対しての平均得点および標準偏差を算出した．正規化した各評価対象の各評価項目に対する平均得点および標準偏差値を表 5.5に示す．表 5.5に示す通り，平均得点に差が出ている．これらの差が統計的に有意な差であることを検証しなければならない．統計的に有意な差であることを検証するために，多重比較法の中のボンフェローニ法を用いて検証した．以下に，情報提示方法を変更したシステムと Yahoo!関連検索を比較し，検定した結果および得られた考察について述べる．また，本システムと情報提示方法を変更したシステムを比較し，検定した結果および得られた考察について述べる．

(α) 情報提示方法を変更したシステムと Yahoo!関連検索の比較

多重比較法を用いた検定の結果，(a),(b),(c),(e) および (g) の評価項目で 99%信頼区間において，平均得点の差が統計的に有意であることがわかった．また，(d) および (f) の評価項目で 95%信頼区間において，平均得点の差が統計的に有意であることがわかった．このため，情報提示方法を変更したシステムは，Yahoo!関連検索に比べ，興味拡大およびシステムの面白みにおいて有効であることを示した．

Yahoo!関連検索は，インターネット上に存在する情報からキーワードに関連した単語を抽出する．しかし，ユーザの興味を考慮していないため，誰が検索をしても同じ結果が提示される．一方，情報提示方法を変更したシステムは，キーワードとユーザの興味に関連する単語を抽出する．すなわち，情報提示方法を変更したシステムと Yahoo!関連検索は，ユーザの興味の考慮した情報抽出という点で異なる．情報提示は両方ともリスト形式であるため，インターフェースの影響はない．したがって，興味拡大およびシステムの面白さには，本研究で提案するように，ユーザの興味を考慮した情報抽出を行うことが有効であることがわかった．

Yahoo!関連検索の情報抽出方法は，明らかではないが，Yahoo!関連検索の結果は，多くのユーザが AND 検索を行うような組み合わせが多く提示された．一方，情報提示方法を変更したシステムは，多くのユーザが AND 検索を行うかどうかではなく，Web ページに出現しているかどうかを検索する．そのため，普段の検索では，AND 検索に用いないような単語間のつながりが出現しやすく，(c) および (e) の評価項目の平均得点が高くなったと考えられる．

前述したように，情報提示方法を変更したシステムと Yahoo!関連検索は，ユーザの興味の考慮および情報抽出の方法の 2 点で異なる．これら 2 つの比較では，興味拡大およびシステムの面白みに有効であったのは，ユーザの興味を考慮した影響か情報抽出方法の違いであるかは定かではない．どちらも興味拡大において有効な方法であると考えられるが，興味拡大への寄与を求めるためには，他の実験を行う必要がある．

以上より，本研究で提案したシステムのサブシステムである，ユーザモデル構築システムおよび外部情報整理システムの少なくともどちらかは，興味拡大に有効であったことを示した．

(β) 本システムと情報提示方法を変更したシステムの比較

多重比較法を用いた検定の結果，(b) および (g) の評価項目で 99%信頼区間において，平均得点の差が統計的に有意であることがわかった．また，(a) および (e) の評価項目で 95%信頼区間において，平均得点の差が統計的に有意であることがわかった．このため，本システムは，情報提示方法を変更したシステムと比べ，興味拡大およびシステムの面白みにおいて有効であることを示した．

情報提示方法を変更したシステムは，出力結果をリストとしてユーザへ提示する．一方，本研究で提案したシステムは，出力結果をマインドマップを模倣した形でユーザへ提示する．すなわち，情報提示方法を変更したシステムと本研究で提案したシステムは，情報提示方法のみが異なる．したがって，興味拡大およびシステムの面白さには，マインドマップを模倣した情報提示方法が有効であることがわかった．

実験後には，リストとして情報を提示されると見やすいが，単語間に関係性があるという印象をあまり受けないため，興味をひかないというコメントがあった．一方，マップとして情報を提示されると，自分の興味のあるものを見つけやすく，単語間に関係性があるという印象を受けたというコメントがあった．そのため，(a),(b) および (e) の評価項目のように，単語間のつながりが重視される評価項目では，インターフェースによる影響があることがわかった．本研究では，マインドマップが持つ効果を得るために，マップとして情報を提示した．本実験の結果は，マインドマップを模倣した情報提示法は，発散思考が促進され，ユーザの興味拡大を促すことを示す．したがって，興味拡大には，提示する情報を考慮するだけでなく，情報提示の方法の考慮も重要であることがわかった．

(c),(d) および (f) の評価項目では，平均得点に有意な差が見られなかった．これは，評価項目が，情報提示方法ではなく，提示される情報が重視される評価項目であったためであると考えられる．すなわち，提示される情報を重視しなければならないものは，インターフェースによる影響が少ないことがわかった．

以上より，本研究で提案したシステムのサブシステムであるユーザインターフェースは，興味拡大に有効であることを示した．

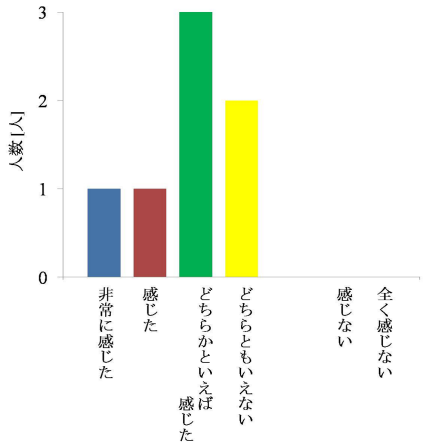


図 5.9: 余計なキーワードに関するアンケート結果

本実験では，被験者に，自分の興味モデルを見せ，自分の興味を把握させた後，実験を行わせた．その際，自分の興味モデルの中に，余計なキーワードが多いように感じるという意見が多かった．実際，余計なキーワードがあるように感じるかというアンケートを行った結果，図 5.9に示すように，8 人中 5 人の 62.5%の人が余計なキーワードがあるように感じていた．これは，Web ページの解析の際，本文だけでなく，Web ページ内にある広告を取得し，解析を行ってしまったことや本文中の単語でも，ユーザの興味のない単語を取得したことが原因であると考えられる．本実験では，(c) および (e) のような意外性を表す評価項目の平均得点が高かった．この結果は，前述した，モデル内の余計なキーワードの影響の可能性がある．そのため，余計なキーワードと意外性の関係を検証するためには，より詳細な実験が必要であると考えられる．

5.4 ユーザの興味の考慮が興味拡大に与える影響の検証

本節では，ユーザの興味拡大において，ユーザの興味を考慮することの影響を検証するための官能評価実験について述べる．5.4.1項では，官能評価実験の目的および評価項目について述べる．5.4.2項では，官能評価の実験方法について述べる．5.4.3項では，得られた実験結果および結果から得られた考察について述べる．

5.4.1 実験目的

本項では，官能評価実験の目的および評価項目について述べる．

本研究では，ユーザモデル構築システムで，ユーザの興味モデルを構築し，そのユーザの興味に関連する情報を提示する．ユーザの興味モデルは，暗黙的手法を用いて構築されるため，モデル構築のためにユーザに与える負荷は大きくないが，ユーザはモデル構築のために，何らかの作業を行う必要がある．もし，ユーザの興味の考慮が興味拡大に与える影響が少ないとすると，あらかじめ興味モデルを用意することができるため，ユーザに与える負荷の低減やコスト削減といったメリットがある．

また，5.3節で述べた実験では，ユーザの興味を考慮したことによる興味拡大に与える影響はわからなかった．そのため，本実験は，ユーザの興味の考慮が，ユーザの興味拡大に与える影響を検証するために行った．本実験では，

- ユーザの興味モデルを用いた本システム
- 社会的に重要な記事から構築したモデルを用いた本システム（以下，社会的重要度が大きいモデル）

表 5.6: 興味ある単語の含有率

被験者のモデル	社会的に重要度が大きいモデル	社会的に重要度が小さいモデル
平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)
42.36 (12.26)	33.13 (16.38)	0.00 (0.00)

- 社会的に重要でない記事から構築したモデルを用いた本システム（以下，社会的重要度が小さいモデル）

の 3 つを比較する．社会的重要度が大きいモデルは，ユーザの興味に関わらず，社会的に重要が大きいと思われる単語を用いて構築する．本研究では，BIGLOBE により発表されている 2009 年の重大ニュース [73] のトップ 50 から，ランダムに 10 個の記事を選び，その Web ページを解析し，モデルを構築した．社会的重要度が小さいモデルは，ユーザの興味に関わらず，社会的に重要度が小さいと思われる単語を用いて構築する．本研究では，一般的にはあまり知られておらず，マイナーであると言われている歌手に関する Web ページを選び，その Web ページを解析し，モデルを構築した．これら 3 つのモデルを比較することで，ユーザの興味の考慮が重要であるか，社会的に重要な単語が重要であるか，どんな単語でも本研究で提案する情報抽出方法であれば，興味拡大を行うことができるのかを検証する．

評価項目は，ユーザの興味拡大に与える影響を検証するため，5.2節で行った実験と同様，

(a) 興味が拡大されたように感じたか

(b) 気になる結果が提示されたように感じたか

(c) AND 検索には使用しない組み合わせが提示されたように感じたか

(d) 新たな知識が増えたように感じたか

(e) 自分にとって意外な結果が出てきたように感じたか

(f) 提示される結果に満足したか

(g) システムに面白みを感じたか

の 7 つとした．これら 7 つの評価項目に対して，非常に感じた，感じた，やや感じた，どちらでもない，やや感じない，感じないおよび全く感じないの 7 段階で被験者に評価させた．

5.4.2 実験方法

本項では，官能評価で行った実験の方法について述べる．実験は，以下の方法で行った．

1. 被験者に，本システムに必要な，検索に用いるメタデータ，キーワード，情報抽出タイプおよび出力決定タイプを決定させた
2. 本システムが提示する結果を見て，気になった単語を検索させた
3. 3 つの興味モデルでも，同様の操作を行わせた
4. キーワードは 2 つ以上使うように指示した
5. 3 つのモデルを比較し，評価させた

表 5.7: モデルの違いを比較した実験結果

	被験者のモデル	社会的重要度が大きいモデル	社会的重要度が小さいモデル
	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)
評価項目(a)	0.90 (0.57)	0.03* (0.65)	-0.93** (0.43)
評価項目(b)	0.87 (0.67)	0.18 (0.75)	-1.05** (0.31)
評価項目(c)	0.55 (0.94)	-0.32 (0.70)	-0.23 (1.09)
評価項目(d)	0.97 (0.34)	0.18* (0.68)	-1.15** (0.34)
評価項目(e)	0.70 (0.78)	-0.45* (0.55)	-0.25 (0.98)
評価項目(f)	1.02 (0.46)	-0.09** (0.81)	-0.94** (0.46)
評価項目(g)	1.12 (0.26)	0.03 ** (0.51)	-1.16** (0.27)

被験者には，実験を行う前に，15 分程度のネットサーフィンと興味ある 5 通の E-mail を用いて，被験者の興味モデルを構築させた．

また，検索に用いたそれぞれの興味モデルの中の，被験者の興味ある単語の含有率を調査するため，検索に用いた単語のうち，興味ある単語を抽出させた．

被験者は，20 代から 30 代の男女 8 人である．

5.4.3 結果および考察

本項では，官能評価実験で得られた結果および結果から得られた考察について述べる．

まず，検索に用いたそれぞれの興味モデル内の，被験者の興味ある単語の含有率を表 5.6に示す．表 5.6からわかるように，被験者に構築させたモデルが最も高い含有率を示している．

つぎに，官能評価で用いた非常に感じた，感じた，やや感じた，どちらでもない，やや感じない，感じないおよび全く感じないをそれぞれ 3，2，1，0，-1，-2 および-3 と数値化し，得られた結果を正規化した．正規化して得られたデータより，各評価項目に対しての平均得点および標準偏差を算出した．正規化した各評価対象の各評価項目に対する平均得点および標準偏差値を表 5.7に示す．表 5.7に示す通り，平均得点に差が出ている．これらの差が統計的に有意な差であることを検証しなければならない．統計的に有意な差であることを検証するために，多重比較法の中のボンフェローニ法を用いて検証した．表 5.7中の’**’ は，被験者の興味モデルを用いた時と比べ，99%信頼区間において平均得点に有意な差があることを示し，’*’ は，被験者の興味モデルを用いた時と比べ，95%信頼区間において平均得点に有意な差があることを示す．また，’*’ がないものは，平均得点に有意な差がみられないことを示す．以下で，多重比較法を用いて検定した結果と結果から得られた考察について述べる．

(α) ユーザの興味モデルと社会的重要度が大きいモデルの比較

多重比較法を用いた検定の結果，(f) および (g) の評価項目で 99%信頼区間において，平均得点の差が統計的に有意であることがわかった．また，(a),(d) および (e) の評価項目で 95%信頼区間において，平均得点の差が統計的に有意であることがわかった．このため，ユーザの興味モデルを用いた情報提示は，社会的重要度が大きいモデルを用いた情報提示に比べ，興味拡大およびシステムの面白さにおいて，有効であることを示した．

表 5.6に示したように，社会的重要度が大きいモデルに比べ，被験者の興味モデルの方が，興味ある単語の含有率が高い．そのため，モデルの中に興味ある単語が含まれている方が，興味拡大に有効であり，提示される結果の満足度およびシステムの面白さが向上すると考えられる．

また，(b) および (c) の評価項目では，被験者の興味モデルを用いた時と比べ，平均得点に有意な差がみられなかった．これは，情報抽出の方法は同一であり，普段 AND 検索を行わないような単語間のつながりが出力されたためであると考えられる．さらに，社会的に重要な単語につながっているため，被験者は提示結果が気になったと考えられる．

(β) ユーザの興味モデルと社会的重要度が小さいモデルの比較

多重比較法を用いた検定の結果，(a),(b),(d),(f) および (g) の評価項目で 99%信頼区間において，平均得点の差が統計的に有意であることがわかった．このため，ユーザの興味モデルを用いた情報提示は，社会的重要度が小さいモデルを用いた情報提示に比べ，興味拡大およびシステムの面白さにおいて，有効であることを示した．

表 5.6に示したように，社会的重要度が小さいモデルには，ユーザの興味ある単語は含まれていなかった．このため，ユーザの興味を考慮するほうが，興味拡大およびシステムの面白さに有効であることがわかった．

また，(c) および (e) の評価項目では，被験者の興味モデルを用いた時と比べ，平均得点に有意な差がみられなかった．これら 2 つの評価項目で有意な差がみられなかったのは，被験者が知らない言葉が，提示される結果の中に含まれていたためであると考えられる．すなわち，社会的重要度が小さいモデル内の単語を被験者が知らず，出力された結果は被験者が知っていたため，単語間のつながりに意外性を感じ，AND 検索に使わないと評価したと考えられる．

以上より，ユーザの興味を考慮するほうが，興味拡大において有効であることがわかった．また，ユーザの興味を考慮することは，提示される結果の満足度を高め，システムが面白いという印象を与えやすいことがわかった．しかし，本実験だけでは，ユーザの興味モデルをより正確にし，余計なキーワードを一切なくしたモデルの方が，興味拡大およびシステムの面白さに有効であるかどうかは定かではない．ユーザの興味ある単語の含有率が 100%になればよいのか，余計なキーワードとのバランスが大事であるのかは，より詳細な検証を行う必要がある．この検証を行うためには，余計なキーワードが入っていないユーザの興味モデルを構築する必要がある．現状のユーザモデル構築システムでは，任意の単語を入力し，モデルを構築することはできない．また，Web ページおよび E-mail の本文から，ユーザの興味ある単語のみを抽出する手法は，明らかにはなっておらず，実装することは現状では困難である．そのため，より詳細な検証を行うためには，余計なキーワードが入らずにモデルを構築する方法を検討する必要がある．

また，本実験で用いた社会的重要度が大きいモデルには，被験者の興味ある単語が 31%程度含まれていた．そのため，ユーザの興味を考慮することと社会的に重要な単語を用いることのどちらが，興味拡大において有効であるかということは，本実験ではわからなかった．これを検証するためには，社会的に重要であり，ユーザが興味を持っていない単語でモデルを構築する必要がある．3.1節で述べたように，社会的に重要な単語は，興味を抱くきっかけの 1 つである．社会的に重要な単語は，多くのユーザが興味を抱いているため，社会的に重要であり，ユーザが興味を持っていない単語を用いてモデルを構築することは困難である．社会的に重要な単語が興味拡大に与える影響を検証するためには，前述した問題を解決する方法を検討する必要がある．

第6章　思考支援システムの評価

本章では，構築した思考支援システムの評価について述べる．本システムの有効性を示すために，2つの既存のシステムと比較し，官能評価実験を行った．6.1節では，官能評価実験の目的および評価項目について述べる．6.2節では，官能評価の実験方法について述べる．6.3節では，実験から得られた結果および結果から得られた考察について述べる．

6.1　実験目的

本節では，官能評価の目的および評価項目について述べる．
本実験では，本システムの有効性を示すため，

- 本研究で提案したシステム
- Yahoo!関連検索
- 5 分間のネットサーフィン

の 3 つを比較した．ネットサーフィンでは，被験者に Google のトップページをスタートとして，自由に Web ページを閲覧させた．本システムと Yahoo!関連検索を比較することで，入力したキーワードに関連する情報を抽出するだけでなく，ユーザの興味モデルを考慮する方が，興味拡大に有効であることを示す．本システムとネットサーフィンを比較することで，現在ユーザの興味拡大の主な手段となっているシステムよりも，本システムの方が，興味拡大に有効であることを示す．
本研究は，ユーザの興味拡大を促し，新たな知識を獲得させることで，ユーザの思考支援を目的としている．そのため，評価項目は，

(a) 興味が拡大されたように感じたか

(b) 気になる結果が提示されたように感じたか

(c) AND 検索には使用しない組み合わせが提示されたように感じたか

(d) 新たな知識が増えたように感じたか

(e) 自分にとって意外な結果が出てきたように感じたか

(f) 提示される結果に満足したか

(g) システムに面白みを感じたか

の 7 つとした．(a) は興味拡大に関する評価項目，(b) および (c) は興味を拡大するためのきっかけに関する評価項目，(d) は知識獲得に関する評価項目，(e) は興味拡大に重要とされている意外性に関する評価項目，(f) および (g) はシステムの継続利用に関わる満足度および面白さに関する評価項目である．これら 7 つの評価項目に対して，非常に感じた，感じた，やや感じた，どちらでもない，やや感じない，感じないおよび全く感じないの 7 段階で被験者に評価させた．

表 6.1: 本システムと既存のシステムを比較した実験結果

	本研究で提案したシステム	Yahoo!関連検索	ネットサーフィン
	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)
評価項目(a)	0.98 (0.43)	-0.90 ** (0.56)	-0.07** (0.85)
評価項目(b)	0.70 (0.50)	-0.68** (0.82)	-0.01 (0.89)
評価項目(c)	1.26 (0.21)	-0.37 ** (0.52)	-0.89** (0.39)
評価項目(d)	0.30 (0.75)	-0.30 (0.99)	0.00 (0.78)
評価項目(e)	1.22 (0.22)	-0.91 ** (0.48)	-0.31** (0.56)
評価項目(f)	0.01 (0.92)	-0.81 (0.82)	0.80 (0.45)
評価項目(g)	1.19 (0.22)	-0.81 ** (0.56)	-0.38** (0.65)

6.2 実験方法

本節では，官能評価で行った実験の方法について述べる．実験は，以下の方法で行った．

1. 被験者に，本システムに必要な，検索に用いるメタデータ，キーワード，情報抽出タイプおよび出力決定タイプを決定させた
2. 本システムが提示する結果を見て，気になった単語を検索させた
3. Yahoo!関連検索に，本システムに入力したキーワードを入力させた
4. Yahoo!関連検索が提示する結果を見て，気になった単語を検索させた
5. キーワードは 2 つ以上使うように指示した
6. Google のトップページをスタートページとして，自由に 5 分間ネットサーフィンを行わせた
7. 3 つのシステムを比較し，評価させた

被験者には，実験を行う前に，15 分程度のネットサーフィンと興味ある 5 通の E-mail を用いて，興味モデルを構築させた．本実験では，このあらかじめ構築させた興味モデルを用いて，実験を行った．被験者は，20 代から 30 代の男女 9 人である．

6.3 結果および考察

本節では，官能評価で得られた実験結果および結果から得られた考察について述べる．

まず，官能評価で用いた非常に感じた，感じた，やや感じた，どちらでもない，やや感じない，感じないおよび全く感じないをそれぞれ 3，2，1，0，-1，-2 および-3 と数値化し，得られた結果を正規化した．正規化して得られたデータより，各評価項目に対しての平均得点および標準偏差を算出した．実験により得られた各評価対象の各評価項目に対する平均得点および標準偏差値を表 6.1に示す．表 6.1に示す通り，平均得点に差が出ている．これらの差が統計的に有意な差であることを検証しなければならない．統計的に有意な差であることを検証するために，多重比較法の中のボンフェローニ法を用いて検証した．表中の’**’ は，本システムと比較した結果，99%信頼区間において差が有意であることを示す．’*’ がないものは，差が有意でないことを示す．

以下に，本システムと既存のシステムを比較し，検定した結果および考察について述べる．

表 6.2: ネットサーフィンの方法別に比較した実験結果

	本研究で提案したシステム	ネットサーフィン (検索)	ネットサーフィン (特定)
	平均値 (標準偏差)	平均値 (標準偏差)	平均値 (標準偏差)
評価項目(a)	0.98 (0.43)	-0.71 ** (0.54)	0.73 (0.34)
評価項目(b)	0.70 (0.50)	-0.55 ** (0.78)	0.66 (0.44)
評価項目(c)	1.26 (0.21)	-1.04 ** (0.18)	-0.71** (0.50)
評価項目(d)	0.30 (0.75)	-0.30 (0.85)	0.38 (0.44)
評価項目(e)	1.22 (0.22)	-0.70 ** (0.36)	0.18** (0.36)
評価項目(f)	0.01 (0.92)	0.71 (0.34)	0.91 (0.53)
評価項目(g)	1.19 (0.22)	-0.78 ** (0.45)	0.11 ** (0.51)

(α) 本システムと Yahoo!関連検索の比較

多重比較法を用いた検定の結果，(a),(b),(c),(e) および (g) の評価項目で 99%信頼区間において平均得点の差が統計的に有意であることがわかった．このため，本システムは，Yahoo!関連検索に比べ，興味拡大およびシステムの面白みにおいて，有効であることを示した．

本システムは，Yahoo!関連検索と，ユーザの興味の考慮，情報抽出の方法およびインターフェースの 3 点で異なる．5.3節で述べたように，本システムを構成する 3 つのサブシステムは興味拡大において有効である．そのため，これら 3 つのサブシステムが興味拡大において有効に働き，Yahoo!関連検索に比べ，興味拡大およびシステムの面白みに有効であったと考えられる．

また，(d) の評価項目では，平均得点に有意な差が見られなかった．これは，自分にとって有益なものを学んだときに獲得するものが知識であると被験者が感じており，本システムおよび Yahoo!関連検索で提示される結果は，興味を拡大するが有益な情報で，何かを学んだと被験者が感じなかったため，有意な差が見られなかったと考えられる．どのようにして，ユーザにとって有益な情報を見極め，それを提示するかということは今後の課題である．

さらに，(f) の評価項目でも，平均得点に有意な差が見られなかった．これは，本システムで提示される結果のうちのいくつかは，関係性があるのではなく，同一の Web ページ内に書かれていた単語であったということがあった．そのため，このようなシステムの誤った結果の提示が，ユーザの満足度を低減させたと考えられる．今後，単語間の関係性の精度を向上させ．提示する結果を改良することで，ユーザの満足度を向上させることができると考えられる．

(d) および (f) の評価項目は，5.3節の検証実験では，平均得点に有意な差があった．本実験では，比較対象が 5 分間のネットサーフィンであり，ネットサーフィンに比べて，差が大きくないと被験者が感じたため，有意な差がみられなかったと考えられる．

(β) 本システムと 5 分間のネットサーフィンの比較

多重比較法を用いた検定の結果，(a),(c),(e) および (g) の評価項目で 99%信頼区間において平均得点の差が統計的に有意であることがわかった．このため，本システムは，5 分間のネットサーフィンに比べ，興味拡大およびシステムの面白さにおいて有効であることを示した．

ここで，被験者に 5 分間のネットサーフィンを行わせた際，ネットサーフィンには大きく分け，2 種類のタイプがあることがわかった．1 つは，被験者が思いついたキーワードを Google を用いて検索し，その検索結果の中から興味を抱く Web ページを探す方法である．この方法を以下，ネットサーフィン（検索）と呼ぶ．もう 1 つは，Yahoo!ニュースやサンケイスポーツといった気に入っ

ている Web ページを被験者が既に有しており，その Web ページの中でネットサーフィンを行う方法である．この方法を以下，ネットサーフィン（特定）と呼ぶ．このため，この 2 種類のネットサーフィン別に平均得点および標準偏差を算出した．この結果を表 6.2に示す．表 6.2に示す通り，平均得点に差が出ている．これらの差が統計的に有意な差であることを検証しなければならない．統計的に有意な差であることを検証するために，多重比較法の中のボンフェローニ法を用いて検証した．表中の'**' は，本システムと比較した結果，99%信頼区間において差が有意であることを示す．'*' がないものは，差が有意でないことを示す．

多重比較法を用いた検定の結果，本システムとネットサーフィン (検索) では，(a),(b),(c),(e) および (g) の評価項目で 99%信頼区間において平均得点の差が統計的に有意な差であることがわかった．また，本システムとネットサーフィン (特定) では，(c),(e) および (g) の評価項目で 99%信頼区間において平均得点の差が統計的に有意な差であることがわかった．

表 6.2からわかるように，ネットサーフィン（特定）の方法をとった被験者は，本システムとネットサーフィンの興味拡大のし易さは変わらないと評価している．ネットサーフィン (特定) の方法をとった被験者は，総合的な情報が書かれている Web ページを閲覧していた．ネットサーフィン (特定) は，ネットサーフィンをするだけで，様々な情報を取得することができるため，本システムとネットサーフィンは，興味拡大において違いがなかったと考えられる．

一方，ネットサーフィン（検索）の方法をとった被験者は，本システムの方が，ネットサーフィンに比べて，興味拡大をし易いと評価している．ネットサーフィン (検索) の方法をとった被験者は，自分の思いつくキーワードしか検索に用いることができない．ネットサーフィン (検索) は，新たな情報を得にくいため，ネットサーフィンに比べ，本システムの方が，興味が拡大されやすかったと考えられる．

本実験では，被験者 9 人のうち 5 人がネットサーフィン (検索) の方法をとった．ネットサーフィン (検索) の方法をとった被験者の方が多く，評価点が低かったため，2 つの方法を合わせて本システムと 5 分間のネットサーフィンを比較すると，表 6.1に示すように，平均得点に有意な差がみられたと考えられる．

また，(c),(e) および (g) の評価項目は，ネットサーフィン (検索) およびネットサーフィン (特定) の両方に比べて，本システムの方が平均得点が高い．ネットサーフィンは，被験者の思いつくキーワードや，被験者が知っている関係性を用いて検索を行う．一方，本システムは，被験者が知らない単語も提示する．そのため，意外性に関する評価項目に関しては，ネットサーフィンに比べ，本システムの方が平均得点が高くなったと考えられる．さらに，意外性のある結果の提示される方が，ユーザに新たな発見を与え，システムの面白みに繋がったと考えられる．

(f) の評価項目では，平均得点に有意な差はなかったが，本システムに比べ，5 分間のネットサーフィンの方が，高い平均得点を得た．ネットサーフィンは，ユーザが既に興味を抱いている事柄に関する結果が常に提示される．本システムは，ユーザの興味を拡大するための事柄を提示するため，既に興味を抱いている事柄と少し離れた結果を提示する．本システムは，全ての情報がユーザの興味拡大につながるとは限らないため，ユーザの興味ある情報を常に出すネットサーフィンに比べて，満足度の高い結果を出すことができなかったと考えられる．また，前述したように，本システムが提示する情報のうちのいくつかは，関係性の無い単語が誤って提示された．そのため，システムの誤った結果の提示が，ユーザの満足度を低減させたと考えられる．今後，単語間の関係性の精度を向上させ．提示する結果を改良することで，ユーザの満足度を向上させることができると考えられる．

本実験では，ネットサーフィンを行う時間を 5 分と限定した．しかし，ネットサーフィンを行う時間が長いほど，ネットサーフィンの平均得点は上昇する可能性がある．そのため，ネットサーフィンに比べ，本システムが本当に有効であることを示すためには，より詳細な検討が必要であると考えられる．

評価実験後には，本研究で提案したシステムは面白く，実際に常に使用していたいという意見があった．そのため，本システムを向上させ，製品として販売すれば，ニーズがあることがわかった．今後は，興味拡大だけでなく，2.2節で述べたような価値を見出すためのシステムへと改良する予定である．

第7章　結論および今後の展開

本章では，本研究の結論および今後の展開について述べる．7.1節では，本研究で述べた内容および得られた成果について述べる．7.2節では，今後の展開について述べる．

7.1　結論

本研究では，Web 閲覧履歴および E-mail 履歴からの興味抽出による思考支援システムの提案および構築を行った．本研究では，思考支援の中でも，新たな知識の獲得を促す興味拡大に焦点を当て，本システムを提案した．

本システムを提案するための要求分析として，まず，現状分析を行い，興味拡大における現状の問題点を明らかにした．インタビューおよびブレインストーミングを行い，問題点を洗い出し，分類することで，興味拡大には,「情報を取得できないこと」および「ユーザが次の興味の対象を予測できないこと」の 2 つの問題があることがわかった．そのため，本システムには,「効率よく外部の情報を取り入れること」および「ユーザの次の興味の対象を予測すること」の 2 つの特徴をもたせた．

つぎに，価値分析を行い，本システムが生み出す価値を明らかにした．本研究では，今後様々な分野で応用するために必要な，最も基礎的なシステムを構築することとした．

興味拡大における問題点の 1 つである「ユーザが次の興味の対象を予測できない」という問題を受け，システムがユーザの次の興味の対象を予測するために，新たに興味を抱くきっかけを調査した．その結果，新たに興味を抱くきっかけは,「自分の興味に関する事柄」,「所属するグループ内の共通事項」,「自分の感性に一致する事柄」,「社会的に重要な事柄」,「予想外の事柄」および「予想通りの事柄」の 6 種類があることがわかった．本研究では，この 6 種類のきっかけに基づいた設計を行った.「自分の興味に関する事柄」を提示するためには，ユーザの興味をシステムに把握させる必要がある．そのため，ユーザの興味を把握するためのサブシステムを構築する．また,「社会的に重要な事柄」,「予想外の事柄」および「予想通りの事柄」は，単語の社会的重要度およびユーザの興味との関連度を用いて表すことができる．したがって，これら 2 つを用いて提示する情報を決定するためのサブシステムを構築する．さらに，興味拡大をさせやすくするため，情報提示方法も考慮する．

以上より，本システムは，

- ユーザが閲覧する Web ページおよびユーザが興味をもった E-mail の内容から単語を抽出し，メタデータを付加することでユーザの興味の把握を行う，ユーザモデル構築システム
- インターネット上から，ユーザの興味に関連した単語を抽出し，単語の社会的重要度およびユーザの興味との関連度を用いて出力を決定する，外部情報整理システム
- 情報提示に必要なパラメータを入力させ，出力される単語をマインドマップのように放射状に配置し提示する，ユーザインターフェース

の 3 つのサブシステムを用いて構築した．

構築した思考支援システムの検証を行った．その結果，以下のことを検証した．

- Web ページおよび E-mail の解析精度

- 人名，団体名，サイト名，製品名，商品名，イベント名および地名の 7 種類のメタデータを用いることで，Web ページの解析では 57.59%，E-mail の解析では 46.58%，ユーザの興味をモデル化することができる
- Web ページの解析では，取得されるべき単語の 68.52%が取得でき，E-mail の解析では，取得されるべき単語の 60.18%が取得できる

- 情報抽出タイプおよび出力決定タイプが興味拡大に与える影響
  - ユーザの興味モデル内の単語のみを出力するタイプは，興味モデル内に存在しない単語を提示するタイプに比べて，興味拡大には有効ではない

- 3 つのサブシステムの有効性
  - 本システムと情報提示方法をリスト状に変更したシステムを比較した結果，マインドマップを模倣した情報提示方法の有効性を示した
  - 情報提示方法をリスト状に変更した本システムと Yahoo!関連検索を比較した結果，ユーザの興味を考慮した情報抽出方法の有効性を示した

- ユーザの興味の考慮が興味拡大に与える影響
  - ユーザの興味ある単語が含まれているモデルを用いて情報提示を行った方が，興味拡大において有効である

また，構築したシステムの評価を行った．この評価では，構築したシステムと既存のシステムである Yahoo!関連検索およびネットサーフィンを比較した．その結果，以下の知見を得た．

- Yahoo!関連検索と比べて，本システムの方が，興味が拡大するという印象およびシステムが面白いという印象をユーザに与える
- 5 分間検索を繰り返すネットサーフィンに比べて，本システムの方が，興味が拡大するという印象をユーザに与える
- 5 分間のネットサーフィンに比べ，本システムの方が，システムが面白いという印象をユーザに与える

以上より，本研究で提案した，Web 閲覧履歴および E-mail 履歴からの興味抽出による思考支援システムの有効性を示した．

7.2 今後の展開

本研究で提案したシステムは，様々な価値を生み出すための最も基礎的なシステムである．そのため，このシステムをより有用なシステムに発展させ，実用化するための課題および今後の展開を以下に述べる．

(a) ユーザモデル構築システムの改良

5.1節で述べたように，本研究では，7 種類のメタデータでは，ユーザの興味を正確に表現しているとは言えない．また，単語の取得精度も 65%程度であった．そのため，ユーザの興味をより正確に把握するための，追加するメタデータの種類の検討および単語の取得精度の向上をする必要がある．さらに，ユーザの興味をより正確に把握するために，興味モデルに用いる情報源を増やし，ユーザの興味を多面的に把握する必要がある．

また，Web ページおよび E-mail の本文を解析するためのルールをより一般化し，どのような文章にも対応できるようにする必要がある．本研究では，ルールを人手で構築した．あらゆる文章に対応するルールを構築するためには，システム自体が，文書構造を学習することが有効であると考えられる．そのため，ルールの一般化とともに，文書構造の学習も検討することで，より汎用性の高いルールを構築し，ユーザの興味モデルの精度を向上させることができると考えられる．

さらに，本研究では，ユーザとシステムのインタラクションとして，マップの拡大，マップの縮小および頂点の移動の 3 つの機能を実装した．これらは，今後ユーザが自分の興味モデルを修正するために，必要であると考えられる機能である．今後は，興味モデルの修正や頂点の数が多くなったときのためのカテゴライズといった機能を実装し，ユーザとシステムのインタラクションを多様にすることで，より正確なユーザの興味モデルを構築することができると考えられる．

本研究では，ユーザの興味が，興味拡大に与える影響を調べきることはできなかった．そのため，余計なキーワードがなく，より正確なユーザの興味モデルを構築し，詳細な検証を行い，ユーザの興味が，興味拡大に与える影響を検証する必要がある．検証結果によっては，コスト削減やユーザへ与える負荷の低減が望めるため，製品化するために，有利に働くと考えられる．

(b) ユーザの興味に関連する単語の抽出精度の向上

本研究では，ユーザの興味に関連する単語をインターネット上から抽出した．官能評価実験の最中に，ユーザが単語の関連を意外だと思い，それを検索しても，別々の記事が同じ Web ページ内にあっただけという状況がしばしば見られた．これは，同一の Web ページ内に単語だけが存在し，実際には関連していない単語が抽出されてしまったからである．そのため，今後，本当に関連している単語のみを抽出する方法を検討する必要がある．

(c) 興味を抱くきっかけに基づいた興味拡大方法の改良

本研究では，新たに興味を抱くきっかけの分類の中の「所属するグループ内での共通事項」および「自分の感性と一致する事柄」の 2 つを対象外とした．これら 2 つのきっかけも，ユーザの興味を拡大させると考えられるため，これら 2 つのきっかけに基づいたシステムを考案する必要がある．

まず,「所属するグループ内での共通事項」というきっかけに基づき，所属するグループまたは友人・知人の興味モデルを用いた情報推薦方法を検討するとともに，プライバシーの問題といった他人の興味モデルを用いることによるリスクを検討し，本システムを発展させる必要がある．

つぎに,「自分の感性と一致する事柄」というきっかけに基づき，ユーザの感性をモデル化する方法を検討し，インターネット上の情報からユーザの感性と一致する事柄を抽出する方法を検討する必要がある．

さらに，本研究では，新たに興味を抱くきっかけを 6 種類に分類した．しかし，興味拡大に与える，それぞれの分類の寄与率は算出していない．今後は，興味の因子分析を行い，興味拡大において，有効な因子を見つけることで，より効率的に興味拡大を行うことができると考えられる．

(d) システムの性能の向上

本研究で構築したシステムは，性能に関して考慮していない．現在，Google は，0.3 秒程度で検索を行うことができるが，本システムは，検索に 20 秒程度の時間を要する．これは，製品化を行ううえでは，大きな問題となる．3章で述べたように，検索速度を向上させる様々な研究が行われているため，これらの知見をもとに，検索の高速化を目指す必要がある．

また，本研究では，各ユーザのデータを全てデータベースに保存している．しかし，実際に製品化した場合，多くのユーザの全てのデータを保存することは困難である．したがって，データの保存方法および保存場所といった問題を解決する必要がある．

(e) 興味の継続性の検証

人は，提示される事柄の意外性が高いほど，長く関心を抱くと，波多野らは述べている [47]．本研究で提案したシステムでは，意外性の高い結果が多く提示された．そのため，2.1節で述べた興味拡大における現状の問題の 1 つである，すぐに興味を無くしてしまうという問題を解決できる可能性がある．

興味を長く継続させることができれば，ユーザにより多くの知識を獲得させることができ，思考をより支援することができる．そのため，本システムで提示される結果と興味の継続性の関係について，検証する必要がある．

(e) 様々な用途への対応

2.2節で述べたように，本システムは，様々な価値を見出すための最も基礎的なシステムである．そのため，ユーザの興味拡大といったエンターテイメント寄りの価値しか創出できなかった．今後は，様々な用途に対応するように，システムを再設計し，発展させる必要がある．

謝辞

本研究は，多くの方たちの支えなくしてやり遂げることはできませんでした．この場を借りて感謝の意を示します．

まず，研究室に配属されてからの約 3 年間に渡り，日頃から多大なる御指導，御鞭撻を賜りました慶應義塾大学大学院システムデザイン・マネジメント研究科前野隆司教授に心から感謝の意を表します．研究の方向性から研究の細部に渡るまで，未熟な私に対し辛抱強く御指導していただきました．本当にありがとうございました．

つぎに，副査としてご指導いただきました慶應義塾大学大学院システムデザイン・マネジメント研究科中野冠教授および浦郷正隆准教授に感謝の意を表します．中野教授および浦郷准教授は，貴重な時間を割き，副査として私の論文を指導していただきました．誠にありがとうございました．

また，私がこの研究に従事してからの約 2 年の間，共同研究を行った株式会社キーウォーカーの皆様に心から感謝の意を表します．代表取締役である真瀬正義様は，貴重な時間を割き，複数回にわたるミーティングを行っていただき，研究に関する鋭いご指摘をしてくださいました．太田健様は，お忙しい中，コンセプトの議論からプログラム構築まで，本研究の多くの部分に関わり，尽力してくださいました．本当にありがとうございました．

さらに，私の研究に対するご指摘のみならず，様々なご助言をくださった牧野泰才特別研究助教に感謝の意を表します．貴重な時間を割き，文章の添削および研究に関する相談にのってくださり，私に励ましの言葉をかけてくださいました．本当にありがとうございました．

そして，共同研究者として，研究に関する多くの議論を交わしてきた村尾将和君に感謝の意を表します．村尾君と議論することで，自分のアイデアや疑問をぶつけることができ，本研究をより良くすることができました．本当にありがとうございました．

互いの研究に関して議論することで切磋琢磨し，励ましあいつつ共に研究室生活を送った門並秀樹君，五味淳君，塩川雄太君，高峯聡一郎君，田代郁君，日高佑輔君，松本直仁君ら同輩に感謝の意を表します．

その他，研究室で一緒にやってきた皆様に深く感謝します．白坂成功君をはじめとして，今関一飛君，白土寛和君など様々な方から貴重なご意見をいただきました．

最後に，24 年間私のことを常に支えてくれた両親および家族に，心から感謝申し上げます．

2010 年 2 月 5 日
慶應義塾大学大学院システムデザインマネジメント研究科
システムデザイン・マネジメント専攻

参考文献

[1] 三宅芳雄. 個人知識の外化に基づく思考支援環境. 情報処理学会研究報告 94-HI-53, pp. 109–116, 1994.

[2] 土方嘉徳. 嗜好抽出と情報推薦技術. 情報処理, Vol. 48, No. 9, pp. 957–965, 2007.

[3] A.Gulli and A.Signorini. The indexable web is more than 11.5 billion pages. *Proceedings of the Fourtheenth International World Wide Web Conference*, pp. 902–903, 2005.

[4] Google. http://www.google.co.jp.

[5] Yahoo Japan. http://www.yahoo.co.jp.

[6] Malone, T.W, and et al. Semi-structured messages are surprisingly useful for computer-supported coodination. *Proc. of CSCW'86*, pp. 102–114, 1986.

[7] P.Resnick, N.Iacovou, and et al. Grouplens:an open architecture for collaborative filtering of netnews. *Proc. of CSCW'94*, pp. 175–186, 1994.

[8] Joachimes.T, Freitag.D, and et al. Webwatcher：a tour guide fro the world wide web. *Proc. of IJCAI'97*, Vol. 15, pp. 770–777, 1997.

[9] B.Sarwar, G.Karypis, and et al. Item-based collaborative filtering recommendation algorithms. *Proc. of WWW'01*, pp. 285–295, 2001.

[10] J.Herlocker, J.Konstan, and et al. Expaining collaborative filtering recommendations. *Proc. of CSCW'00*, pp. 241–250, 2000.

[11] Amazon. http://www.amazon.com.

[12] Google News. http://news.google.com.

[13] 奥健太, 中島伸介, 他. ユーザコンテキストを考慮した情報推薦方式に関する研究. 情報処理学会研究報告. 情報学基礎研究会報告, Vol. 2008, No. 56, p. 129, 2008.

[14] Sony. http://www.sony.jp.

[15] 宮原伸二, 安部伸治, 他. ユーザのにぎわい感を利用した興味喚起型コンテンツガイド方式について. 情報処理学会研究報告, Vol. 2005, No. 9, pp. 23–28, 2005.

[16] 森本和伸, 林貴宏, 他. Mineblog：興味発見を支援する blog 記事推薦システム. 情報処理学会論文誌, Vol. 47, No. 4, pp. 1171–1180, 2006.

[17] 関口厚次, 中西泰人. Slip board：リアルタイムなクリップボード共有による関心拡張システム. 情報処理学会シンポジウム論文集, Vol. 2009, No. 4, pp. 53–54, 2009.

[18] 松尾豊, 福田隼人, 他. ユーザ個人の閲覧履歴からのキーワード抽出によるブラウジング支援. 人工知能学会論文誌, Vol. 18, pp. 203–211, 2003.

[19] 斎藤真里, 山本則行. 電子メールからの興味抽出手法と意外な web ページと出会うきっかけを与えるエージェントシステム. 認知科学, Vol. 11, No. 3, pp. 252–261, 2004.

[20] Salton.G. Automatic Text Processing:The Transformation,Analysis,and Retrieval of Informationy Computer. Addison-Wesley, 1989.

[21] 丹英之, 大向一輝, 他. Quereseek：検索履歴の逆引きによるコミュニティベースの web ナビゲーション. 人工知能学会全国大会論文集, Vol. 21, pp. 2F4–1, 2007.

[22] 高野明彦, 西岡真吾, et al. 汎用連想計算エンジンの開発と大規模文書分析への応用. 情報処理振興事業協会「独創的情報技術育成事業」2001 年度成果報告論文, 2002.

[23] 丹英之, 大向一輝, 他. Quereseek：検索履歴共有によるコミュニティ指向の連想検索. 人工知能学会全国大会論文集, Vol. 22, pp. 1G1–01, 2008.

[24] 高野明彦, 西岡真吾, 他. 連想に基づく情報アクセス技術　-汎用連想計算エンジン geta を用いて-. 情報の科学と技術, Vol. 54, No. 12, pp. 634–639, 2004.

[25] 岩山真, 今一修. 汎用連想計算エンジン geta を用いた特許連想検索システム. *Japio 2007 YearBook*, pp. 152–155, 2007.

[26] 相原健郎. 概念空間の再構造化による創造性支援. デジタル図書館, No. 10, pp. 3–12, 1997.

[27] 金子修三. テキストマイニング技法を活用した発想支援システムの構築. 北陸先端科学技術大学院大学知識科学研究科知識社会システム学専攻修士論文, 2001.

[28] 相原健郎, 堀浩一. 記憶の想起に基づく創造性支援. 情報処理学会論文誌, Vol. 42, No. 6, pp. 1377–1386, 2001.

[29] R.Agrawal and R.Srikant. Fast algorithms for mining association rules. *Proc. the 20th VLDB Conference*, pp. 487–499, 1994.

[30] 連想検索エンジン reflexa. http://labs.preferred.jp/reflexa/.

[31] 連想検索エンジン ASSOCIE. http://www.nri.co.jp/renso/index.html.

[32] 想 IMAGINE BOOKSEARCH. http://imagine.bookmap.info/index.jsp.

[33] ウィキペディアフリー百科事典. http://ja.wikipedia.org/.

[34] アンサイクロペディア. http://ja.uncyclopedia.info/.

[35] 東洋, 大山正. 学習と思考. 大日本図書, 1969.

[36] 長瀬荘一. 関心・意欲・態度（情意的領域）の絶対評価. 明治図書出版, 2003.

[37] 松田浩平, 佐藤恵美. 文京学院大学生における学習への動機づけと試験成績の原因帰属. 文京学院大学研究紀要, Vol. 6, No. 1, pp. 149–166, 2004.

[38] 下山剛. 学習意欲の見方・導き方. 教育出版, 2000.

[39] T. Sawada, T. Takagi, and M. Fujita. Behavior selection and motion modulation in emotionally grounded architecture for qrio sdr-4x . *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 3, pp. 2514–2519, 2004.

[40] グラフで語ろう総選挙 2009. http://special.visualzoo.com/senkyo/.

[41] 星野由紀子, 高木剛, U. D. Profio, 他. パーソナルロボットにおける行動モジュールを用いた行動制御アーキテクチャ. 日本ロボット学会学術講演会予稿集, Vol. 21, p. 2B18, 2003.

[42] 齋藤竜太郎, 関谷直也. 若者はなぜ選挙に行かないのか. 東洋大学社会学部メディアコミュニケーション学科　卒業論文, 2008.

[43] 板村英典. インターネット空間における情報の信頼性：検索サイト「google」にみる「リンク」と「チャンネル荷重」をめぐって. 関西大学大学院人間科学：社会学・心理学研究, Vol. 66, pp. 1–27, 2007.

[44] 森永順治. 知的好奇心を喚起し問題解決能力を養う個別化実験教材の開発. 佐賀県教育センター長期研修報告書, 2002.

[45] A.A.Freitas. On rule interestingness measures. *Knowledge-Based Systems Journal*, Vol. 12, No. 5-6, pp. 309–315, 1999.

[46] 中小路久美代, 山本恭裕, 他. Collective creativity の支援へ向けて. 情報処理学会研究報告　ヒューマンインターフェース研究会報告, Vol. 99, No. 69, pp. 31–36, 1999.

[47] 波多野誼余夫, 稲垣佳世子. 知的好奇心. 中公新書, 1973.

[48] 石井浩介, 飯野謙次. 設計の科学　価値づくり設計. 養賢堂, 2008.

[49] 黄海湘, 藤井敦, 他. Web コミュニティに基づく情報検索の個人化手法. 言語処理学会年次大会発表論文集, Vol. 11, pp. 1006–1009, 2005.

[50] Satoshi NIWA, Takuo DOI, and et al. Web page recommender system based on folksonomy mining. *Transactions of information Processing Society of Japan*, Vol. 47, No. 5, pp. 1382–1392, 2006.

[51] M.Claypool, A.Gokhale, and et al. Combining content-based and collaborative filters in an online. *In Proc. ACM SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation, Berkeley, California*, 1999.

[52] 土方嘉徳, 青木義則, 他. マウス挙動に基づくテキスト部分抽出方式と抽出キーワードの有効性に関する検証. 情報処理学会論文誌, Vol. 43, No. 2, pp. 566–576, 2002.

[53] P. B Kantor. Capturing human inteligence in the net. *Comm. of the ACM*, Vol. 43, No. 8, pp. 112–115, 2000.

[54] Bijay K.Jayaswal and Peter C. Patton. *Design for Trustworthy Software: Tools, Techniques, and Methodology of Developing Robust Software*. Prentice Hall PTR, 2006.

[55] テクノラティジャパン検索キーワードランキング 2006. http://www.technorati.jp/ranking2006/.

[56] Google Press Center:Zeitgeist. http://www.google.com/press/zeitgeist2005.html.

[57] Yahoo Japan!2008 検索キーワードランキング. http://searchranking.yahoo.co.jp/ranking2008/general.html.

[58] Yahoo Japan!2007 検索キーワードランキング. http://searchranking.yahoo.co.jp/ranking2007/general.html.

[59] Yahoo Japan!2006 検索キーワードランキング. http://searchranking.yahoo.co.jp/ranking2006/general.html.

[60] 角谷由貴, 森田和宏, 他. 電子メールの重要文抽出と分類. 言語処理学会年次大会発表論文集, Vol. 11, pp. 321–324, 2005.

[61] Toru TAKASHIRO and Hideaki TAKEDA. Acquisition and organaization of personal knowledge through www browsing. *Institute of Electronics, Information, and Communication Engineers*, Vol. J85-D-1, No. 6, pp. 549–559, 2002.

[62] 原田昌紀, 佐藤進也, 他. 索引篩法 -大規模サーチエンジンのための高速なランキング検索法. In DEWS2003. 情報処理学会, 2003.

[63] マインドマップとその関連情報についてのサイト. http://mindmap.jp/.

[64] トニー・ブザン, バリー・ブザン. ザ・マインドマップ. ダイヤモンド社, 2005.

[65] 株式会社キーウォーカー. http://www.keywalker.co.jp/.

[66] 工藤拓. Mecab:yet another part-of-speech and morphorogical analyzer. 2006.

[67] AT&T 研究所. http://graphviz.org/. 2007.

[68] Yahoo! API. http://developer.yahoo.co.jp/.

[69] 沢井康孝, 峠泰成, 他. 因子組み合わせによる順位付け文書からの興味因子判別. 言語処理学会第 12 回年次大会, pp. 951–954, 2006.

[70] 天坂格郎, 長沢伸也. 官能評価の基礎と応用. 日本規格協会, 2000.

[71] 永田靖, 吉田道弘. 統計的多重比較法の基礎. サイエンティスト社, 1997.

[72] Yahoo!関連検索ワードサーチ「UnitSearch」ver.2.0. http://www.sem-analytics.com/lab/unitsearch.php.

[73] 2009 年重大ニュース (BIGLOBEMUSIC 調べ). http://music.biglobe.ne.jp/2009-2010/news.html.

付録

5章および6章で行った実験に用いた評価シートを図 7.1に示す.

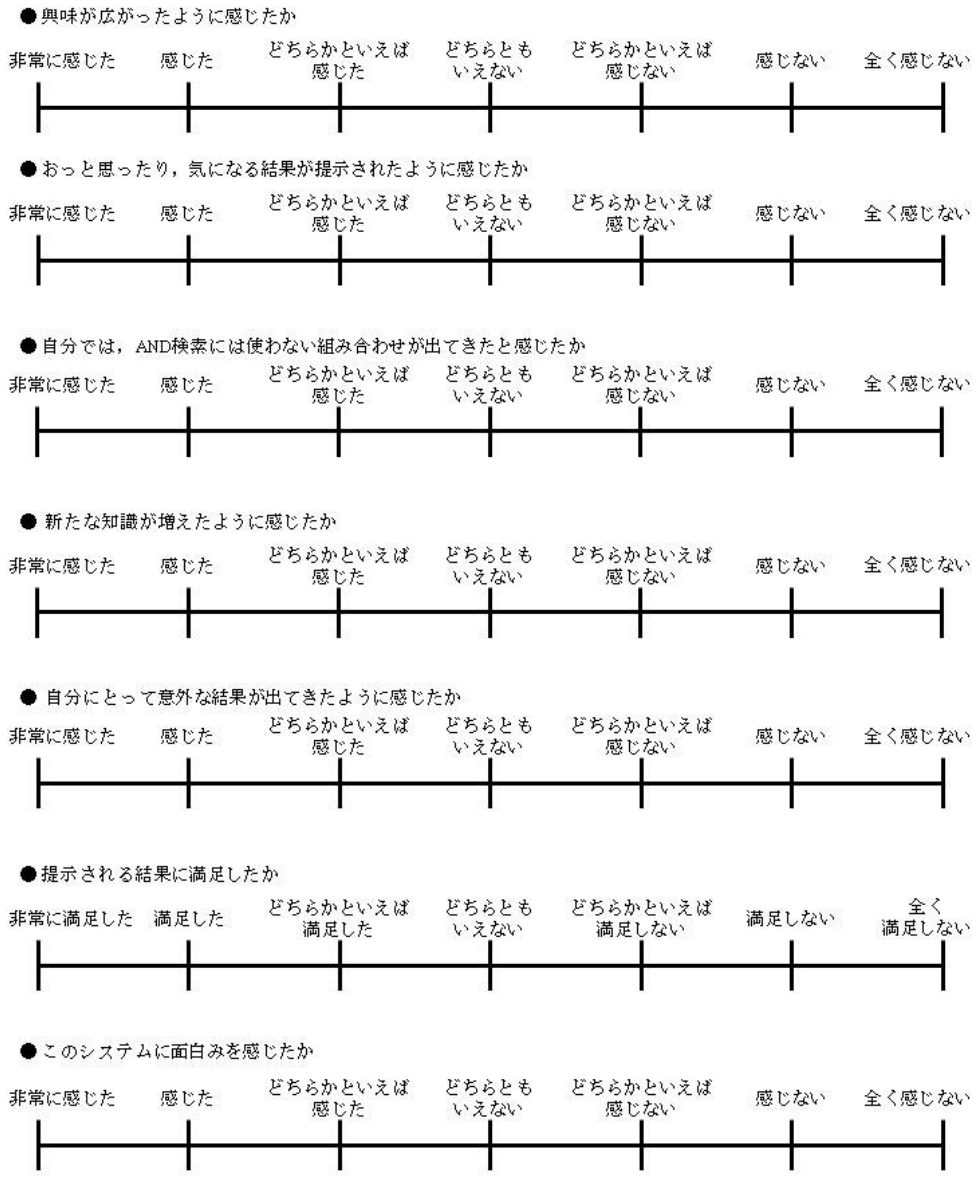


図 7.1: 実験に用いた評価シート