

卒業論文 2002年度（平成14年度）
WWW情報検索ナビゲーション
システムの設計と実装

慶應義塾大学

栗本亜実

ami@sfc.wide.ad.jp

指導教員

村井 純

徳田 英幸

楠本 博之

中村 修

南 政樹

概要

インターネットには様々な情報システムが接続され、多種多様な情報共有が実現されている。家庭へのインターネットの普及、および WWW による情報取得がユーザに浸透するにつれ、従来、専門家しか持たなかった医療や科学、工学等に関する知識や、最新の社会動向をはじめ、特定のコミュニティが共有しているような地域情報や趣味性の高い情報等を、一般の人間が容易に得ることが出来るようになった。

WWW 上のデータを取得する一般的な方法は検索エンジンと呼ばれる情報検索サービスであり、それぞれの検索エンジンは複雑な機能を拡張していくことで、ユーザが雑多な情報の中から必要な情報を取り出せるための仕組みを模索しつづけている。

しかし、検索エンジンがシステム内部において様々な機能を拡張しつづけている一方で、キーワードを元にした手法は以前から変わっていない。ユーザは、自らの知識や経験といったバックグラウンドを基に、目的に沿ったキーワードを設定する事が強いられている。

そのため、WWW 上での情報検索の得手・不得手は、ユーザが目的を表現するに適切な語彙を有しているかいないかにおのずと左右される。

本研究では、従来の検索エンジンの機構から、キーワードのインプット部分を切り離し、同じ目的を持ったユーザ間のキーワードを蓄積・提示した。

ユーザの潜在キーワードを引き出す機能によって、情報検索の不得手なユーザの情報検索をサポートするシステムを構築した。

abstract

Thanks to the Internet, electronic devices of any kind can be inter-connected. This allows us to share various types of information. Now that the Internet has spread to the household and that many data have been put on the WWW, users are able to access information that they were not granted access to before. For instance, medical treatment, science and engineering data has previously only been accessed by specialists; similarly the latest social trends and local information was only known by specialized communities.

The common way to get the information on WWW is to use information retrieval services called search engines. Search engines provide a complicated function for a user to search a specific information out of miscellaneous information. However, while the search engines continuously improve their functions, the way of searching information based on keywords hasn't changed. Users are forced to specify the keywords based on their own knowledge, background or experience. As a result, the WWW information retrieval process is influenced by whether the user using the suitable vocabulary for searching or not.

The purpose of this research study is thus to design a system which supports an efficient reference model. This reference model helps to pull out a user's potential keyword and to accumulate the valid portion of the keywords typed by the users.

目次

第1章	はじめに	1
1.1	本研究の背景	1
1.2	本研究の目的	1
1.3	本論文における言葉の定義	3
1.4	本論文の構成	4
第2章	既存の検索エンジン	5
2.1	WWW 情報検索において発生する情報	5
2.2	WWW 情報検索における立場と機能	6
2.2.1	基本的な検索エンジンの機能	6
2.2.2	既存の検索エンジンのまとめ	9
2.2.3	検索エンジンに関連するその他の技術・問題	10
第3章	モデル化	11
3.1	既存のモデル	11
3.1.1	ユーザの視点での情報検索	11
3.1.2	現状の問題点	13
3.2	本研究のモデル	14
3.2.1	ユーザ側でのアプローチ	15
3.2.2	ナレッジシェアの実現	16
3.3	モデルの仮定とその検証	18
3.3.1	キーワード設定能力には個人差があることの検証	18
3.3.2	同じ目的を持ったグループは同じようなキーワードを使用することの検証	19
3.3.3	仮定の検証のまとめ	19
第4章	設計	20
4.1	設置場所の自由度	20
4.2	動作概要	21
4.3	サーバのシステム要件	22
4.4	サーバの機能	22
4.4.1	検索エンジンとのインターフェース機能	22
4.4.2	使用キーワード蓄積機能	22

4.4.3	キーワード提示機能	22
4.4.4	重み付け機能	22
第5章	実装	23
5.1	システム図	23
5.2	検索エンジン	23
5.3	実装環境	24
5.4	サブレット	24
5.5	データベーススキーマ	24
5.5.1	input_keys テーブル	25
5.5.2	users テーブル	25
5.5.3	logs テーブル	26
5.6	キーワードの提示部	26
5.7	ユーザインタフェース	26
第6章	システムの評価	29
6.1	評価項目	29
6.2	システムの機能する条件の検証(実験)	29
6.2.1	被験者に関する事前調査	30
6.2.2	実験環境	30
6.2.3	実験手法	30
6.3	結果	31
6.3.1	問題の分野と提示キーワードの使用度	31
6.3.2	データベースの成長	32
6.3.3	自由記述	34
6.4	考察	36
6.4.1	得意分野を持ったユーザの存在	36
6.4.2	情報検索能力の個人差	37
6.4.3	考察のまとめ	37
6.4.4	今後の課題	37
第7章	おわりに	39
7.1	まとめ	39
7.2	今後の展望	39
.1	プレアンケート	43
.2	ポストアンケート	44

目 次

3.1 ユーザの視点で見た検索の流れ	12
3.2 情報検索	13
3.3 本研究のモデル	14
3.4 既存のシステムと本システムの比較	15
3.5 本システムにおけるユーザの視点で見た情報検索の流れ	16
3.6 既存の情報検索と本システムの情報検索	17
3.7 他者とキーワードを共有することでできる視野	18
4.1 設置場所の自由度	21
4.2 動作概要	21
5.1 フロー	23
5.2 データベーススキーマ	25
5.3 検索画面のスクリーンショット	27
5.4 検索結果のスクリーンショット	28
6.1 実験概要	31
6.2 テーマと提示キーワードの使用度	32
6.3 データベースの成長と提示キーワードの使用度	33
6.4 キーワードの共有は有益だと思うか (DB の成長ごとの推移)	34

表 目 次

2.1	検索エンジンが処理する情報	6
2.2	ランキングアルゴリズム	7
3.1	グループと問題の割り振り	19
5.1	使用したソフトウェア	24
5.2	使用したハードウェア	24
5.3	データベース	25
5.4	input_keys テーブル	25
5.5	users テーブル	26
5.6	logs テーブル	26
6.1	提示キーワードの使用度とその理由	35

第1章 はじめに

1.1 本研究の背景

インターネットには様々な情報システムが接続され、多種多様な情報共有が実現されている。家庭へのインターネットの普及、および WWW による情報取得がユーザに浸透するにつれ、従来、専門家しか持たなかった医療や科学、工学等に関する知識や、最新の社会動向をはじめ、特定のコミュニティが共有しているような地域情報や趣味性の高い情報等を、一般の人間が容易に得ることが出来るようになった。

internet.com の調査 [15] では、“85 %以上のインターネットユーザーは、検索エンジンを介して商品やサービスを調べている”という結果が出ている。また、博報堂の調査 [16] では、“ブロードバンドになって変わった生活”として 70 %のユーザが「知りたい事を気軽にインターネットで検索」と答えた。このように、インターネットの普及とともに、検索エンジンは日常に密着した場面で利用されていることがわかる。

今後は、あらゆる情報の検索活動を WWW で行う機会が増加すると想定され、膨大な情報の中から効率的に有益な情報を獲得する手段の確立が望まれる。

WWW 上のデータを取得する一般的な方法は、検索エンジンと呼ばれる情報検索サービスである。これには、大きく分けてディレクトリ型とロボット型がある。ディレクトリ型とはカテゴリ別に WWW の情報が階層的に保持されており、ユーザはカテゴリを選択することで欲しい情報に到達する。ロボット型は、あらかじめ WWW 上の情報をフラットに保持しており、ユーザが入力するキーワードに関連する情報を提供し、ユーザは提供された情報をもとに欲しい情報を取得する。

ロボット型による情報検索では、ユーザが任意のキーワードを入力して検索を行うため、検索結果には個人差が顕著に反映される。この個人差は、ユーザの、欲しい情報に対する専門知識や経験などのバックグラウンドや、情報検索そのものに関する経験に因るところが大きい。ユーザがまったく知らない情報をインターネットを介して取得する場合には、専門性があるか否かによって得られる情報が異なってしまうのが現状である。

1.2 本研究の目的

それぞれの検索エンジンは複雑な機能を拡張していくことで、ユーザが雑多な情報の中から必要な情報を取り出せるための仕組みを模索しつつしており、ランキングのアルゴリズムや形態素解析技術は、日々成長している。しかし、キーワード検索において、ユーザにキーワードの入力を求めるインプット部分に関しては、どれも共通している。

また、同じ検索エンジンを使用しているにも関わらず、情報検索能力に個人差がある。これらの事から、キーワード設定の部分で差が生じていると考えられる。

そこで本研究の目的は、ユーザ間のキーワードの設定能力に個人差がある事に着目し、検索キーワードを共有することで、自らの目的をキーワードとして表現出来ない場合でも、他者の使用したキーワードを用いて検索の機会損失を減少させることである。

1.3 本論文における言葉の定義

情報 キーワード, URL, web ページ, web ページ内のテキスト, 画像, タグ等すべて.

ユーザ 検索を行う者, 情報検索者

検索エンジン ユーザの検索要求に対して WWW から web ページの URL とその説明等を結果として表示するシステム.

ユーザの目的 ユーザが知りたい事, 情報検索によって得たい対象

検索キーワード 検索エンジンでユーザが入力する言葉. 検索に使われる言葉.

検索要求 ユーザが検索エンジンでキーワードを入力し, 検索エンジン内の処理に渡す事.

検索結果 検索エンジンがキーワードに基づいて、導き出した URL リスト.

情報検索 ユーザが自らの目的をキーワードとして検索エンジンに入力し, 検索エンジンから検索結果を得る一連の動作.

潜在キーワード ユーザが目的を検索キーワードとして使用したいが, 適切に表現できていないキーワード.

提示キーワード 本システムで提示するキーワード.

情報検索能力 ユーザが検索エンジンを用い, 目的を果たす情報を探す能力.

検索効率 求める情報に, 的確に短時間でたどり着くこと.

バックグラウンド ある分野に関する知識や経験, 全体像.

再検索 1 度以上検索をしたユーザがキーワードを追加して AND 検索をすること.

1.4 本論文の構成

本論文は, 第2章で現状の問題として既存の検索エンジンと情報検索の流れを整理し, その問題点をまとめる. 第3章では, 新しいモデルを提案する. 現状のモデルと比較することで, その利点を述べる.

第4章では, 解決されていない問題点をあげた上で, 本研究の動作場面とシステム設計を示し, 第5章でシステムの実装と, 具体的な動作について述べる. 第6章では本システムの評価を行い考察する. 第7章では結論と今後の課題についてまとめる.

第2章 既存の検索エンジン

本章では、既存の WWW 上での情報検索の要素およびその機能を整理する。まず、情報検索において発生する情報を整理する。次に、情報検索における立場を、情報提供者・検索エンジン・情報検索者の三種に分けて考え、これらの役割・機能を示す。

2.1 WWW 情報検索において発生する情報

WWW における情報検索で、検索エンジンが処理する情報を、表 tab:data にまとめる。大別すると、WEB ページ内のテキストや画像等のコンテンツ情報、タグ等で表現される WEB ページの構造や相互関係等のメタ情報、検索時にユーザが入力するキーワード情報、検索結果情報に分類される。

各情報の具体例を表 2.1 に挙げる。検索エンジンは、これらの情報を活かして様々な技術の開発を行っている。

google ツールバー [20] のように、付加サービスで web ページ評価を取得している検索エンジンもあるが、基本的に全ユーザから常に取得できる情報は限られる。上記の情報のうち、検索エンジンが利用できるのは、web ページから抽出できる情報と、ユーザの入力したキーワードのみである。

表 2.1: 検索エンジンが処理する情報

	情報
コンテンツ情報	テキスト
	画像
	音
	動画
	ファイルオブジェクト (pdf 等)
メタ情報	index
	リンク構造
	HTML タグ
検索結果情報	検索結果 URL リスト
	Web ページ内容概要
	ページ評価
キーワード情報	キーワード
	論理演算子

2.2 WWW 情報検索における立場と機能

現在の WWW における情報検索は, WWW 上に情報を公開する情報提供者, WWW 上のデータを収集し, 提示するサービスを提供する検索エンジン, それを利用する情報検索者であるユーザの三者が存在する. これらの WWW 情報検索における立場とその役割・機能を示す.

2.2.1 基本的な検索エンジンの機能

まず, 検索エンジンの基本的な機能を以下に示す. 検索エンジンは, 大きく, WWW 上のファイルの収集, キーワードとファイルとのマッピング, ユーザのクエリに対する結果の表示を行っている.

- 検索エンジンの備えている機能
 - － ロボット (登録 URL 数, 更新頻度)
 - － インデックス (検索スピード etc)
 - － クエリープロセス (キーワードとページの関連性)
 - * 検索結果のランキング
 - － インターフェース (プレビューの工夫)

現在, 情報検索エンジンにおいて重要視されている技術は, 検索結果のランキングアルゴリズムである. これは, 検索エンジンごとのアルゴリズムで行われている. 各検索エンジンの特徴が出る部分であり, 検索エンジンとしての質を左右する部分である. 最も基本的なランキングアルゴリズムは, リンク構造を利用した google[1] のものであり, この応用がいくつも出現している.

特徴的な検索エンジンのアルゴリズム

今日, ランキングアルゴリズムが重視されているが, 特徴的なアルゴリズムを表 2.2 に示す ([18] 参照).

表 2.2: ランキングアルゴリズム

ランキングアルゴリズム	説明
キーワード出現頻度	キーワードと web ページ内テキストの関連付けを行う. キーワードマッチを支える基本的なアルゴリズム. しかし, 情報提供者の意図的なキーワードの多用は避ける傾向にあり, 多すぎると除外される場合もある.
キーワードの出現位置による重み付け	キーワードの web ページ内テキスト内での出現位置によってキーワードの重み付けを行うアルゴリズム.
リンクポピュラリティ	被リンクが多いサイトは価値のあるサイトであるという理論に基づいたアルゴリズム. Google (PageRank), Wisenut (WiseRank) が採用.
サイトテーマによる重み付け	サイトが扱うテーマやトピックを解析し, 適切な検索結果を返すアルゴリズム. Wisenut, Teoma が採用.
HTML タグによる重み付け	視覚的強調要素である <H1> ~ <H6>, などの文字の大きさや強調を示す HTML タグを考慮するアルゴリズム.
クリック人気	検索結果リストの中のクリックしたか否かをランキングに活かすアルゴリズム. HotBot, goo が採用.

google が採用したリンク構造からページの価値を決めるランキングアルゴリズムが画期的なアルゴリズムであるとして, 大きなインパクトを与えた. それに対して最近, google に対抗しようと新しい機能を採用している検索エンジンは, カテゴリを提示, 関連キー

ワードを提示, また検索を支援するインターフェースシステムといった機能を採用している. これらの検索エンジンを以下に示す.

カテゴリを提示する検索エンジン (Teoma, Vivisimo, WiseNut)

カテゴリを提示する検索エンジンは, 従来の手作業で生成しているディレクトリサービスと違い, ユーザの入力したキーワードに対して, 動的にカテゴリを自動生成し, 提示する機能を持つ. ディレクトリサービスの画一的なディレクトリと比較して, 検索する度に計算してディレクトリを生成するので, 最新かつ日々の情報の変化に即したディレクトリを提供している.

Teoma Teoma [2] の特徴的な技術は”HITS Algorithm”と呼ばれる WWW 内のコミュニティ (同様のテーマを持ったサイト群) 発見アルゴリズムである. キーワードにマッチするページを index から探した後, そのページのリンク構造から, テーマごとの集合をサイトコミュニティとして認識する. さらにサイト間の関係を解釈し, 複数の検索エンジンからリンクを張られているサイトを権威あるサイト (authority サイト) と見なし, authority サイトにリンクを張っているサイトを hub サイトとして重み付けを行い, authority サイトを結果として表示する. また, サイトコミュニティの中で共通して使われている言葉を抽出し, 検索結果と同時に表示される”refine”項目に, 絞込みキーワードとして表示される.

Vivisimo Vivisimo [5] は, ひとつの検索エンジンを指定する事もできるが, 複数の検索エンジン (Yahoo!, MSN, Lycos, AOL, AskJeeves) にクエリを送信し, 結果を統合して出力しているメタサーチエンジンである. 検索エンジンからの雑多な結果を, ドキュメント・クラスタリングの技法で自動的に分類し, ウィンドウの左側に各分類のキーワードを表示し, 右側に分類単位で結果が表示される. ドキュメントクラスタリング技術に定評があり, 企業にイントラネット用にも提供している.

Wisenut Wisenut [4] は 15 億を超える収拾 Web ページ数を確保する (2002 年 1 月) ロボットの技術力がある. Wisenut は, リンク構造に加え, ハイパーリンクになった部分に記述されているテキストであるアンカーテキストと, その周囲のテキストを解析し, リンク元ページの関連語として重み付けを行っている. また, ”検索ガイド”と呼ばれる検索結果を自動分類する機能を備えており, 検索キーワードによる絞込み以外の絞込み方法を提供している.

その他, TITAN, Northern Light など.

関連キーワードを提示する検索エンジン (Altavista, webcrawler, infoseek つぼシーク)

関連キーワードを提示する検索エンジンは, ユーザが送信したキーワードに関連するキーワードと関連性のあるキーワードを抽出し, 検索結果とともに提示する検索エンジンである.

Altavista Altavista [3] は 2002 年 7 月に, 絞込みキーワードの提示機能を追加した. 絞込みキーワードを見出す技術を” Prisma Technology”と呼んでいる.

検索クエリを受けて, 検索結果を返す. 検索結果の上位 50 件の title と abstract の中から, 12 語の関連語 (” Prisma Term”) を抽出する.” Prisma Term” は, 単語, 文, 名前, 概念のいずれかである.” Prisma Term” をクリックする事で絞込み検索をするが, その際, 既にクエリ送信したキーワードに加える事も, ” Prisma Term” のみで使用する事も可能である. 提示キーワードのみで使う選択肢を与える事で, より適切なキーワード設定へと導いている.

Mondou RCAAU ” Retrieval loCation by weighted AssociAtion rUle” (Mondou)[6] は 京都大学で開発された, .jp ドメインに特化したロボットサーチエンジンである. データマイニング技術による検索絞り込み機能やクライアントからのネットワークコストを含む評価機能などを持っている. また, 日本語形態素解析システム JUMAN や重み付き相関ルール導出アルゴリズムを用い, 導出されたキーワードの絞り込みを行っている. データマイニング技術に基づいた関連キーワード表示とリンク逆引き機能を実装している.

その他, excite, infoseek つぼシークなど .

検索を支援するインターフェースシステム

ユーザの情報検索を支援するために, キーワードを提示したり, キーワード間の関係を把握しやすいインターフェースの構築を行っている .

DualNavi DualNAVI[7] はノードを用いた情報検索システムである. このシステムは特徴語グラフというもので検索対象の内容を表示する. そのために, サーバの持つデータに対して事前に各語の全体の出現頻度を算出しておき, 検索結果の語の出現頻度と比較することで, 特徴語を抽出している. また, 対話性を重視した検索インターフェースを用い, ユーザーが検索結果の概要を把握しながら, 多様なフィードバックを掛けられるようにデザインされている.

2.2.2 既存の検索エンジンのまとめ

検索エンジンの進歩はめざましく, ロボット収拾技術の開発にはじまり, テキストマッチ技術, ランキングアルゴリズムの開発が日々行われている . ロボット技術に関しては,

収拾ページ数とスピードの向上, テキストマッチ・ランキングアルゴリズムに関しては, 情報提供者の検索エンジンで, ランキング上位に結果表示されるようなページづくりに左右されないアルゴリズムの開発, また情報検索者のニーズに合った情報を検索結果として提示するか, という観点で開発を行っている.

2.2.3 検索エンジンに関連するその他の技術・問題

今日の情報検索における無視できない問題として, 情報提供者(サイト運営者)が SEO (Search Engine Optimization) と呼ばれる検索結果が上位に表示させる手法を用い, 工夫を凝らしているという問題がある. 自らのサイトを上位に表示させる事で, 特定キーワードで検索をしたユーザを取り込む事を目的としている. 実際には, ランディングページを増やすものが多いが, これはユーザにとっては, 必ずしも有益な情報であるとは限らない. 検索エンジンはユーザにとって有益な情報を検索結果の上位に表示するようなアルゴリズムを開発しているが, これに対して, ユーザができる事は少ない.

一方, 検索エンジン側がサービスとして行っているも増加してきている. *overture* や *google* がその主たるもので, サイト運営者がキーワードと自分のサイトとの関連付けを買うというサービスである. 実際に検索結果から URL がクリックされた数で値段が決まるサービスもある. これも, ユーザにとって必ずしも有益な情報ではないので, 検索エンジンにこういったサービスによって上位表示されている URL には, 広告であるという記述を求める消費者運動もあり, 自主規制を行っている検索エンジンが増えている.

このような検索エンジンと情報提供者とのやり取りは, ユーザの手の届かないところで行われている. ユーザは, 検索エンジンのアルゴリズムに公平さを求める事はできない事からも, 現在の情報検索はユーザと検索エンジンが個々に動作していると言える.

第3章 モデル化

検索エンジンは従来から、ページの収集ロボットの開発にはじまり、今日のリンク構造によるランキングアルゴリズムの開発など、検索結果の質の向上に力を注いでいる。しかし、検索エンジンの根本的な機能は、キーワードマッチである。ユーザにキーワード入力を求める機構は変化しておらず、キーワードのインプット部分はここ10数年、変化がない。

一方、ユーザはそれぞれ個々に情報検索を行っているので、バックグラウンドのない分野での情報検索では、効率的な情報検索を行う事ができない。

以上の問題を踏まえ、本研究では新しいモデル化をおこなった。

3.1 既存のモデル

まず、ユーザの視点で情報検索を把握する。

3.1.1 ユーザの視点での情報検索

ユーザの視点での情報検索の流れを示す。

- ユーザの情報検索
 - － 目的
 - － 目的を検索キーワードで表現
 - － 検索結果から目的に沿った web ページを選出
 - － web ページから、目的に沿った情報を得る

図3.1はユーザの視点での検索の流れである。ユーザは目的を検索する際と、再検索する際にキーワードを設定しなければならない。一つ目のキーワードは自分で考察し設定するが、再検索の際は検索結果や目を通した web ページを元に次のキーワード設定をする（図3.1矢印2）。検索結果の上位しか見ないユーザが大半なので、検索結果の上位に目的を達成する web ページを表示する事が、ユーザにとって最も重要な事である。つまり、絞込み検索におけるキーワード設定は検索の流れの中で最も重要な要素であり、ユーザの能力に左右される。

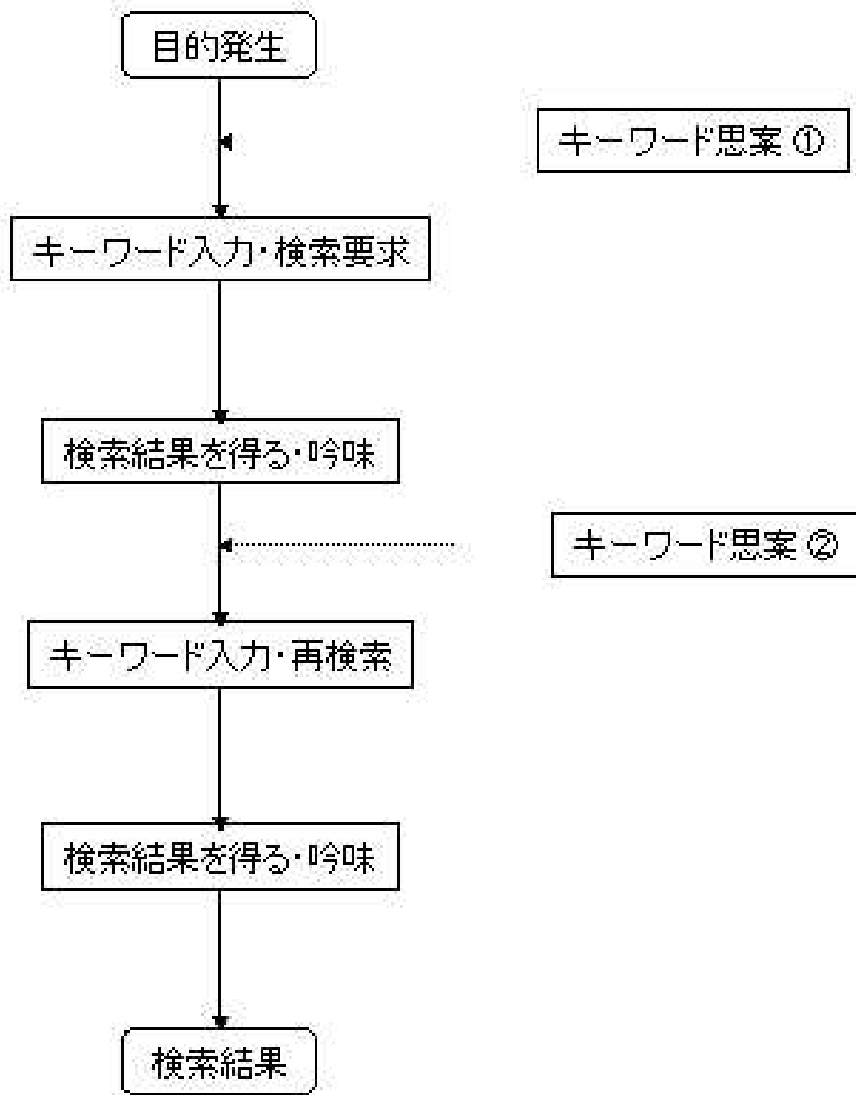


図 3.1: ユーザの視点で見た検索の流れ

3.1.2 現状の問題点

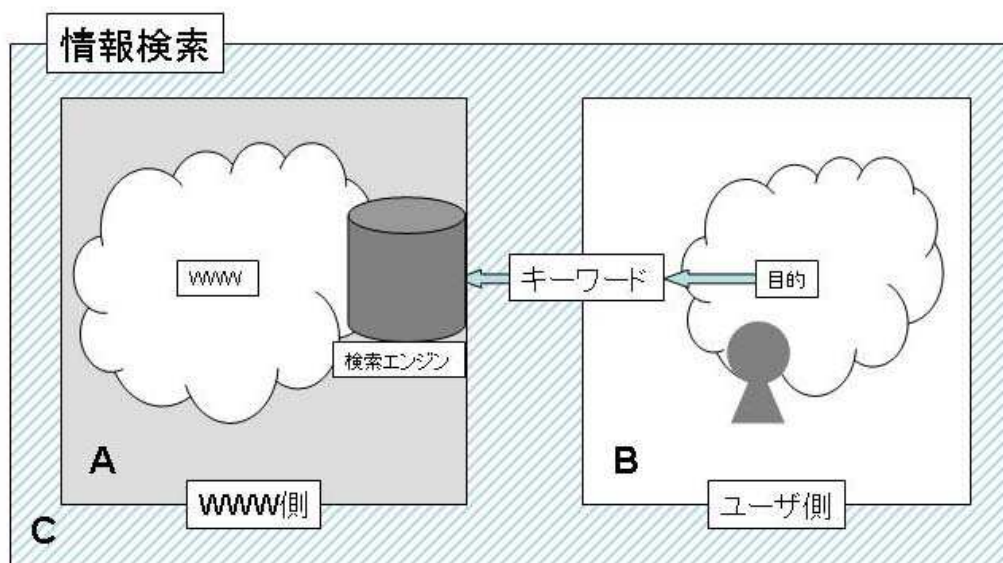


図 3.2: 情報検索

前述のように、既存の検索エンジンの改良はめざましいが、どの検索エンジンにも一貫して言える事は、検索エンジン側のみに焦点をあてている事である（図 3.2 中エリア A）。つまり、キーワードを受け取ったのちのプロセスの向上をはかっている。

情報検索は検索エンジン側の機能とユーザ側の行動をすべて合わせて成り立っている（図 3.2 中エリア C）。現状では、両者はそれぞれ別々に工夫をしており、両者をつなぐものは、キーワードのみである。唯一ユーザとつながりのある部分がキーワードであるにも関わらず、ユーザのキーワード設定を補助する機能はあまりない。検索エンジンが、この部分の開発をあまり行っていないことが問題である。

一方、ユーザの実際の検索活動を見てみると、同じ検索エンジンを使用しているにも関わらず、求めている情報にたどり着く時間や得た情報の質に個人差がある。つまり、検索エンジンの様々な機能以外の部分に問題があると考えられる。

ユーザの検索行動には、検索要求をキーワードで表現する他に、検索結果から必要な情報を選出する能力や、得た情報から、自らのほしい情報を整理・咀嚼する能力が求められる。しかし、ユーザの大半が検索結果は上位 1,2 ページめの結果のみを利用するという調査結果がある [8][11]。また、web ページ内から情報を整理・咀嚼する能力に関しては個人差はあるが、ここでは情報を掲載している web ページの URL が検索結果として表示されていれば、情報検索能力に大きな差はないと考える。

そこで重要となってくるのが、キーワード設定の能力である。検索エンジンを利用した経験の多い者は、AND 検索を多用するという調査 [12] も過去行われており、AND 検索による絞込みの有用性、またキーワード設定の適切さが検索効率を左右していると言え

る。ユーザの情報検索は、キーワード設定の部分が重要だとすると、検索効率の悪いユーザは検索要求をキーワードとして表現する能力が乏しいと考えられる。また、この能力は、ユーザの検索分野に関する知識や経験といったバックグラウンドに依存している。

検索エンジン側では、検索要求を受けた後の検索結果の向上のみを図り、ユーザ側では、自らのバックグラウンドを頼りに各々にキーワード設定のみを工夫しているというのが現状であり、このキーワード設定が情報検索のボトルネックとなっている。

3.2 本研究のモデル

現状のモデルを踏まえ、本研究のモデルとアプローチを提案する。本研究では、既存の検索エンジンが情報提供者側からのアプローチであるのに対して、ユーザ側でのアプローチを行うこと、また従来ユーザは個々に検索を行っていたのに対して、ユーザ間のナレッジシェアを実現する。

- ユーザ側でのアプローチ
- ナレッジシェアの実現

図 3.3は本モデルの全体像である。同じ目的を持ったユーザが一つのキーワードデータベースを共有し、キーワードを通して WWW にアクセスしている。

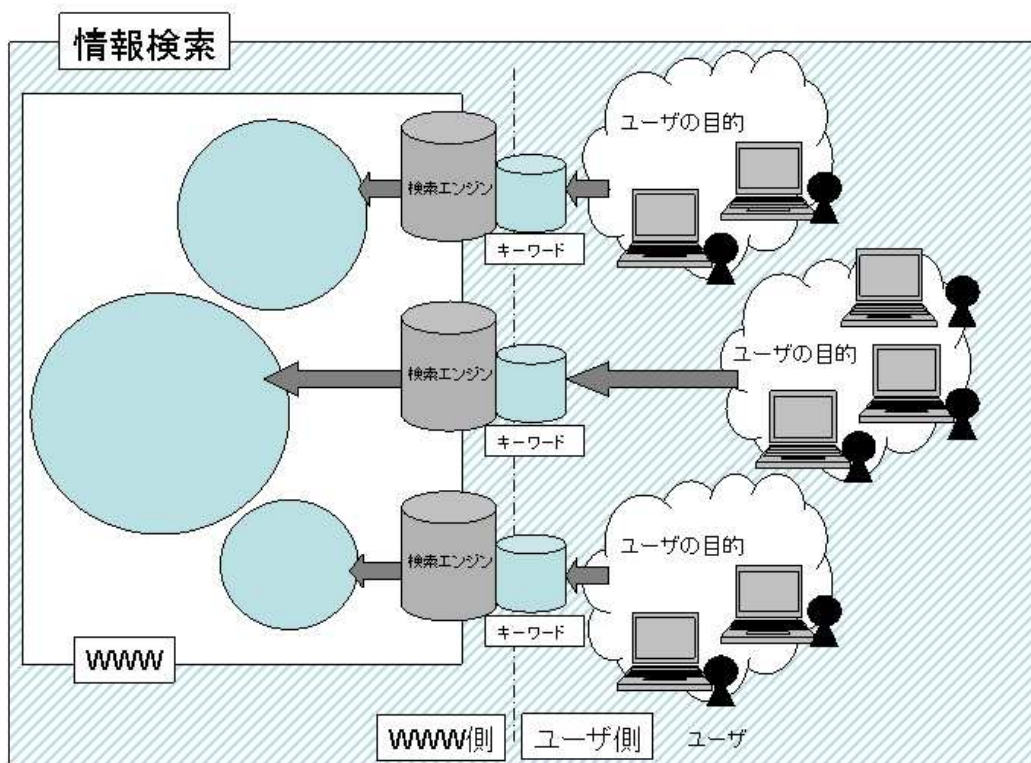


図 3.3: 本研究のモデル

3.2.1 ユーザ側でのアプローチ

既存の検索エンジンは、ランキングアルゴリズムの工夫や、動的カテゴライズ、関連キーワードの提示を行っているが、前章で示したように、多くは web ページ内でのキーワードの意味づけや、リンク構造による意味づけなど、情報提供者の作成した web ページを利用して、検索アルゴリズムの向上を図っている（図 3.4上）。つまり、情報検索者であるユーザの意図したキーワード設定に関わらず、キーワードと web ページとのマッピングを行っている（クエリプロセッシング）。これは、従来の検索エンジンのモデル（図 3.2エリア A）が、受け取るキーワードに関しては感知せず、キーワードを受け取った後の処理に焦点をあてていることに起因する。

本研究では、ユーザの視点で情報検索を捉え直し、情報検索能力の個人差はユーザのキーワード設定がボトルネックとなっていると考えた。

そこで、本システムでは、ユーザの入力したキーワードを蓄積する事によって、ユーザの視点に立った表現でキーワードを提示する事が可能である（図 3.4下）。よって、自らの目的に沿ったキーワード表現を利用しやすくすることを可能とする。

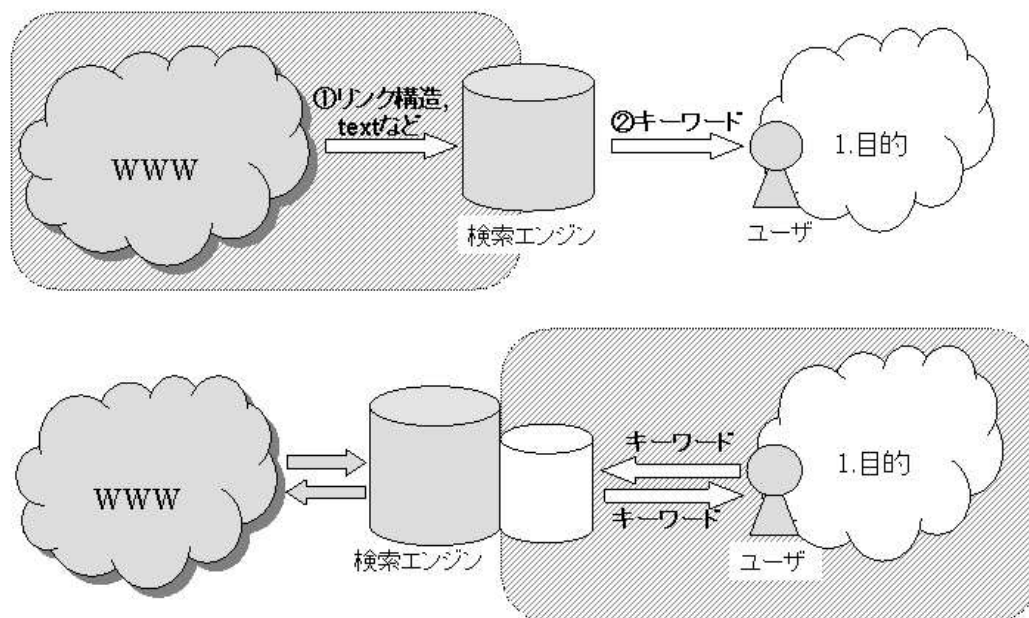


図 3.4: 既存のシステムと本システムの比較

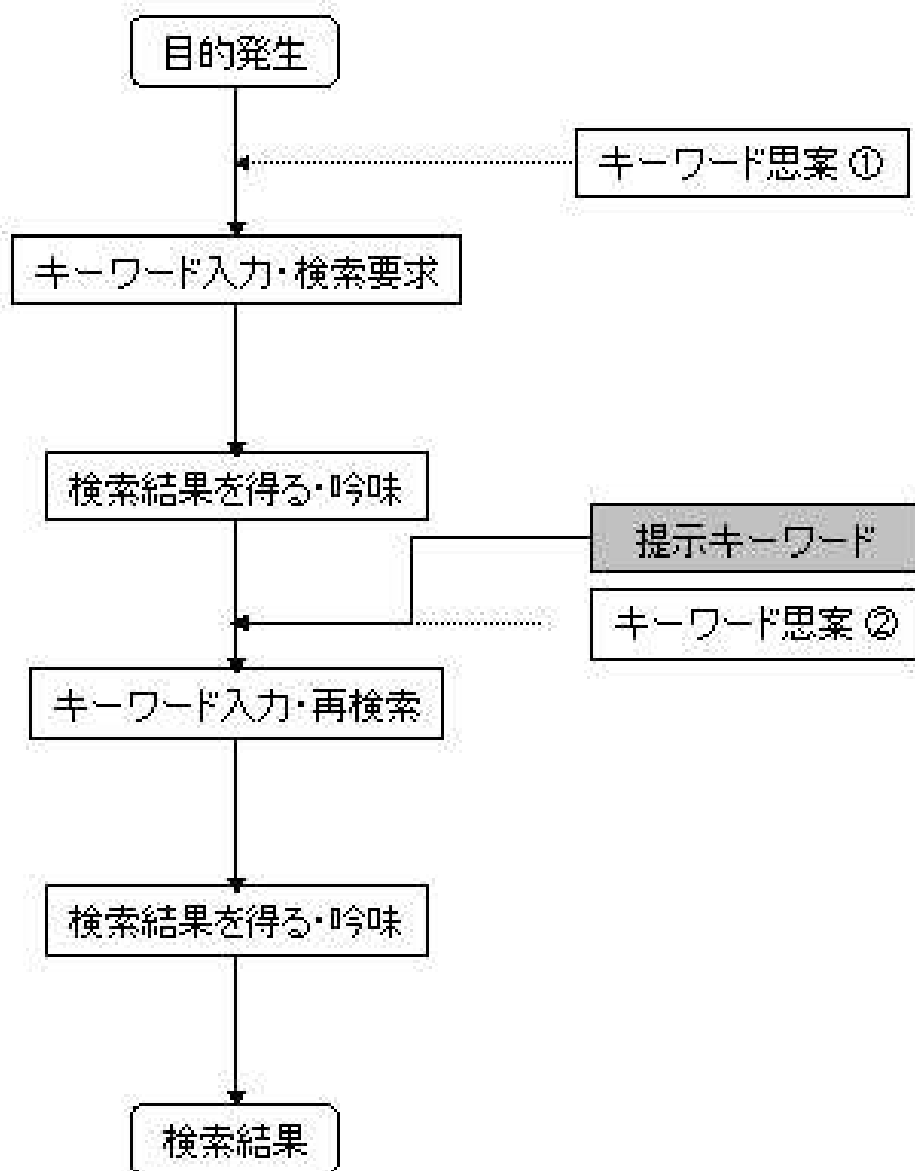


図 3.5: 本システムにおけるユーザの視点で見た情報検索の流れ

3.2.2 ナレッジシェアの実現

従来,たとえ同じ目標をもったユーザ同士でも,独自の手法で,個々に情報検索を行っていた(図 3.6上). 図 3.7では,A,B,Cは3ユーザの持つ視野を示す. この視野は,同じテーマであっても,各ユーザの知識や経験,全体像などといったバックグラウンドに依存し,個人差がある.

バックグラウンドのある分野であれば、その分野に関するある程度の知識や視野を持つ事ができ、バックグラウンドのあるユーザは、膨大な情報が存在する WWW からほしい情報を検索する際、自分なりの視野（図 3.7A,B,C）を持って、検索キーワードを設定し、検索する事もできる。

しかし、バックグラウンドのない分野に関しては、図 3.7A,B,C のような視野を持つことができない。また、適切な視野を持っている分野でも、他者の視野を知る事は、新たな発見につながる。

そこで本研究では、情報検索において、同じ目的を持ったグループで検索キーワードを共有する事によって、バックグラウンドのないユーザも、他ユーザの視野を共有する事を実現する。

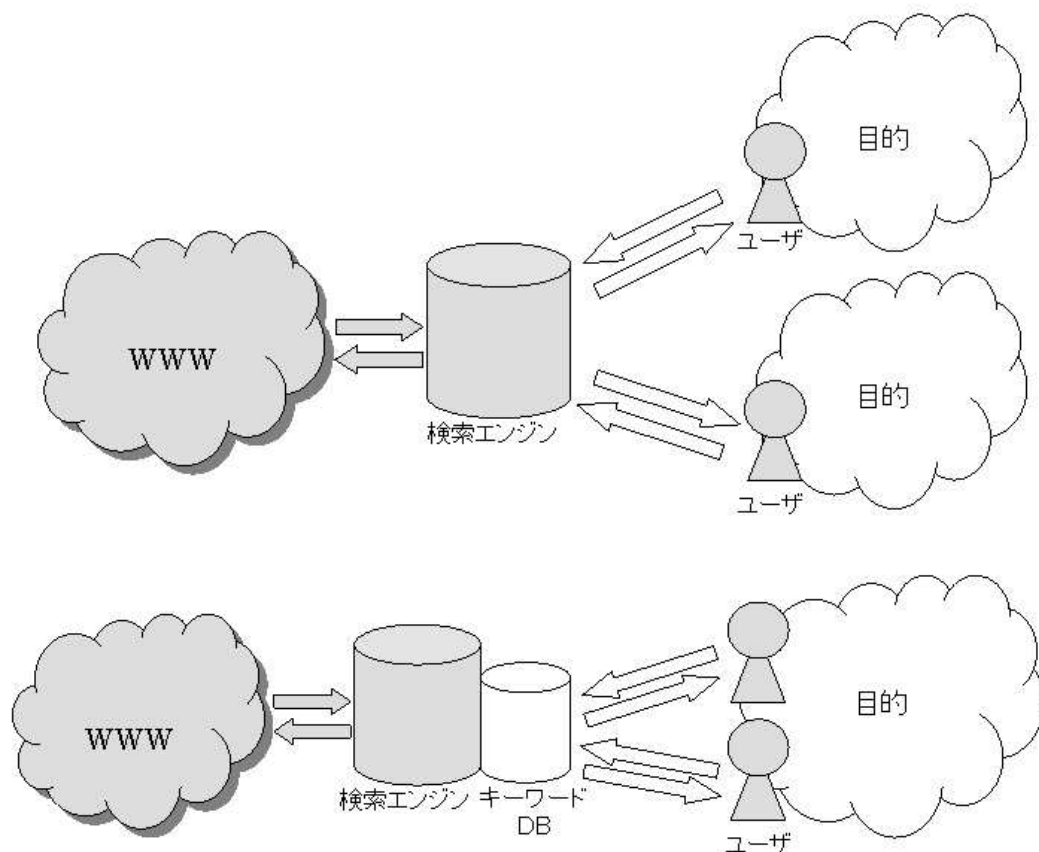


図 3.6: 既存の情報検索と本システムの情報検索

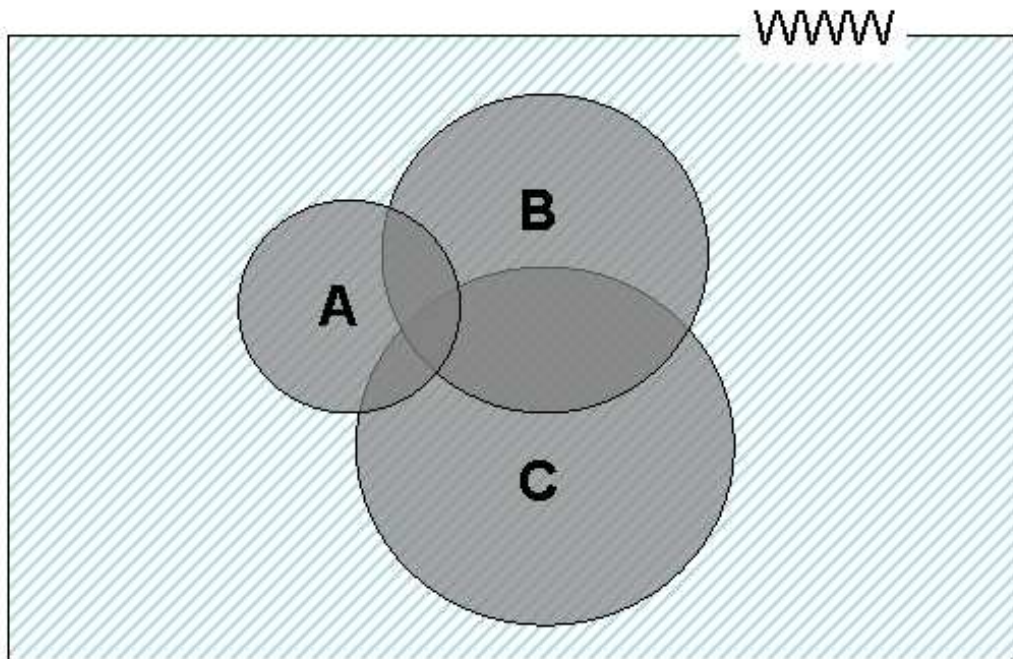


図 3.7: 他者とキーワードを共有することでできる視野

3.3 モデルの仮定とその検証

前節で述べたモデルは、いくつかの仮定に基づいている。そこで、このモデルの仮定となっている以下の点を 14 名の被験者に対して事前調査を行い、検証した。

- キーワード設定には個人差がある
- 同じ目的を持ったグループは同じようなキーワードを使う

3.3.1 キーワード設定能力には個人差があることの検証

同じ検索エンジンを使用しているにも関わらず、ユーザの情報検索能力には個人差がある。本研究では、これはキーワード設定に起因していると考えた。そこで、事前調査において、キーワード設定に個人差があるかどうかを検証をした。

問題を出し、答えを検索エンジンを用いて答えることを想定し、問題からキーワードを連想させた。その結果、キーワード設定能力は、表現能力と検索スキルの二種類に分けることができた。

表現能力は、バックグラウンドを活かして目的をキーワードに落とし込む能力、つまり問題からキーワードを連想する能力である。同じ問題に対して、設定できるキーワード数には、2~5 個の個人差があった。一方、検索スキルとは、検索エンジンの性質を念

頭に置いたものである．フレーズ検索や「とは」「いわゆる」といった語を前後につけるといった工夫が特徴的であった．

3.3.2 同じ目的を持ったグループは同じようなキーワードを使用することの検証

得意分野に対する問題と，不得意分野に関する問題で，キーワード設定をしてもらった．4人ずつ2つのグループに，それぞれの得意分野と得意でない分野の問題を出した．全部で4通り行った．

表 3.1: グループと問題の割り振り

グループ1	グループ2
得意分野の問題	得意でない分野の問題
得意でない分野の問題	得意分野の問題

検証の結果，同じ問題に対するキーワード設定は，7割以上のユーザが使用するキーワードが1～3個あり，キーワード設定に共通性があると言える結果であった．

3.3.3 仮定の検証のまとめ

事前調査では，本章で述べたモデルの前提となっている仮定を検証した．情報検索におけるキーワード設定の個人差，同じ目的を持ったグループにおけるキーワード設定の近似性を検証した．したがって，本研究のモデルである同じ目的を持ったユーザ間でキーワードを共有する事は，価値があると言える．

第4章 設計

前章で示したモデルの仮定の検証によって、情報検索におけるキーワード設定にはやはり個人差があることが言えた。また、同じ目的を持ったグループは同じようなキーワードを使用する事が言えた。個人差とは、良し悪しだけでなく、得意分野の守備範囲の差も含む。この事から、本研究の目指すナレッジシェアは、有効であると言える。そこで、キーワードを媒体に、個人が得意分野を相互補完するような機構を構築する事をめざす。

本システムでは、これを実現するために、同じ目標を持った検索者同士が、キーワードを共有する機構を構築する。検索機能に加えて、情報検索を行うユーザ間で使用したキーワードを蓄積・提示する機能を持たせる。ユーザがクエリの送信を繰り返すたびに、事前に検索者が使用したキーワードを、再検索に利用するためのキーワード案として提示する。

4.1 設置場所の自由度

同じ目的を持ったグループでキーワードを共有するためには、個々にデータベースを保持しなければならない。本システムでは、キーワード設定に着目した事で、情報検索の中の検索エンジンから入力部分を切り離れたシステムなので、システムの設置場所に自由度がある。

たとえば図 4.1 中、A,B,C グループといったように、目的に応じて自由に設置できる。

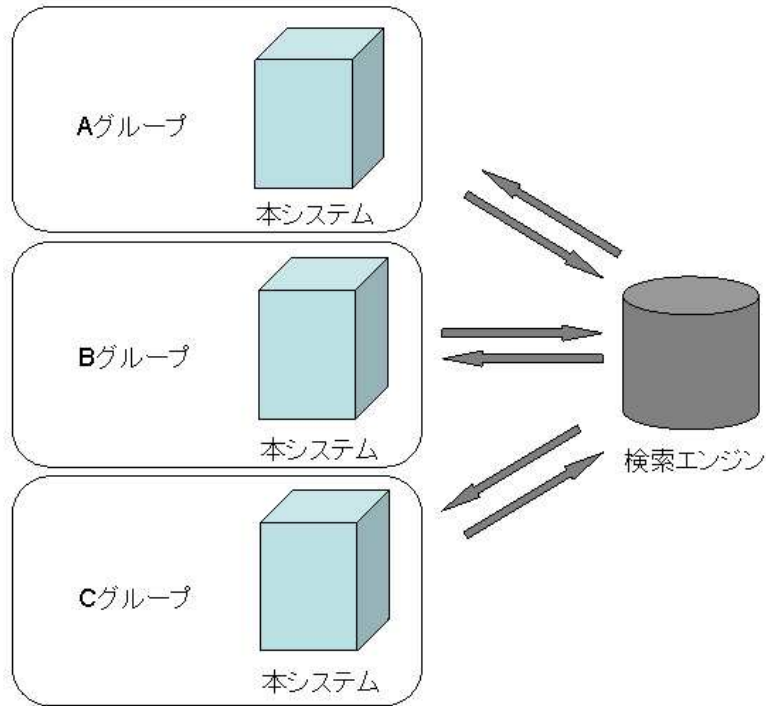


図 4.1: 設置場所の自由度

4.2 動作概要

システムの動作概要を示す。まず、ユーザが検索要求を出す(図 4.2 中 1)。すると、サーバは検索要求を検索エンジンにフォワードする(図 4.2 中 2)。その際、キーワードデータベースに使用したキーワードを登録する。検索エンジンは検索結果を返し(図 4.2 中 3)、サーバはユーザにこれをフォワードする(図 4.2 中 4)。この際、キーワードデータベースから絞込みキーワードを取り出し、検索結果とともに表示する(図 4.2 中 4)。

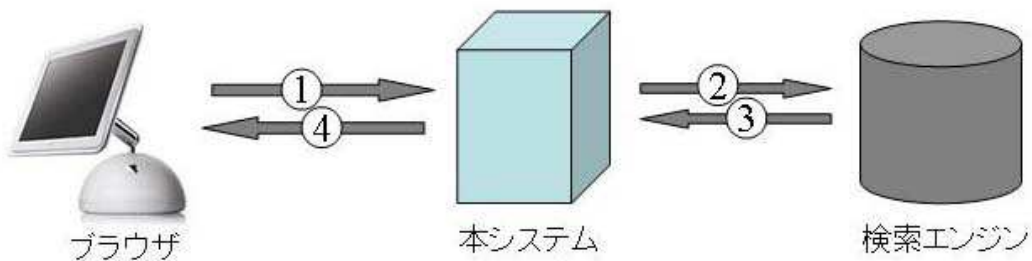


図 4.2: 動作概要

4.3 サーバのシステム要件

本システムに必要な要件は以下のとおりである。

- 検索エンジンへのインターフェース機能
- 使用キーワード蓄積機能
- 絞込みキーワード提示機能
- 重み付け機能

4.4 サーバの機能

本システムのサーバの機能は、以下の通りである。

4.4.1 検索エンジンとのインターフェース機能

前述の通り本システムは検索エンジンから独立したシステムである。つまり自身でweb 検索の機能を持たず、検索機能は既存の検索エンジンを利用して実現している。そのため、検索エンジンへのクエリー送付、結果セットの解析等のインタフェースを装備する。

4.4.2 使用キーワード蓄積機能

サーバは、ユーザの使用したキーワードをキーワードデータベースに登録する機能を持つ。ユーザがAND 検索した際、キーワードとキーワードの組を登録しておく。ユーザがAND 検索のクエリを送信する度にキーワードの組が蓄積される。

4.4.3 キーワード提示機能

検索エンジンから検索結果を受け、ユーザにフォワードする際に、次の一手である絞込みキーワードを表示する。リアルタイムでキーワードデータベースが更新されているので、ここで提示されるキーワードも動的に変わる。

4.4.4 重み付け機能

絞込みキーワードの提案をする際に、キーワードの表示順を決めるアルゴリズムが必要である。今回、ユーザは情報検索を繰り返すたびに、キーワード設定がうまくなるという仮説を元に、最新のキーワードを上位に表示した。また、使用頻度が高いものも、上位表示されるようにした。

第5章 実装

前章の設計のもと, 実装を行った.

5.1 システム図

本システムのフローを示す.

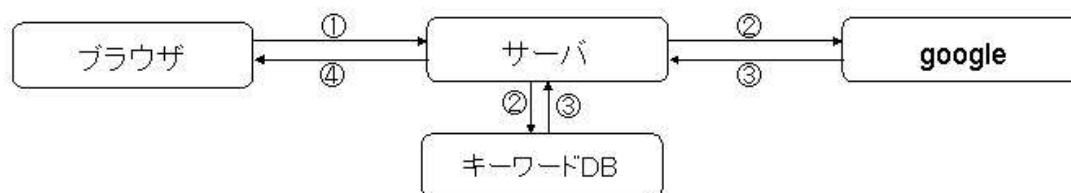


図 5.1: フロー

5.2 検索エンジン

今回、検索エンジンへと google を用いた。これはプレテストで明らかになったように検索エンジンとして google が圧倒的なシェアを持っていることと, google が外部ソフトウェアへのインタフェースを持っていることなどの理由による。google の外部インタフェースは, ユーザ登録しアカウントを取得することで1日あたり1000アクセスまでのクエリーを google に対して発行できる物である, WSDL が公開されており, SOAP を交換することでクエリーや結果セットをやりとりする。このインタフェースは SOAP さえ扱えればどんな言語からでもアクセス可能であるが XML の扱いやすさなどを考え java にて実装した。

google のインタフェースへアクセスするためには google Web API が google から提供されているが, これを用いると日本語を含んだキーによる検索が出来なかった。よって google Web API を用いずに, jaxp の機能を用いてサーブレット内で SOAP message の作成・解析を行った。

5.3 実装環境

使用したソフトウェアは以下の通りである。

表 5.1: 使用したソフトウェア

OS	Vine Linux(Kernel version 2.4.18)
Servlet エンジン	jakarta tomcat 4.1.12
データベース	mysql 3.23.49
コンパイラ	Java2SDK J2SE 1.4.1

同様にハードウェアは以下の通りである。

表 5.2: 使用したハードウェア

CPU	pentium2 400MHz
RAM	512Mbytes
HDD	15Gbytes

5.4 サブレット

サブレットは以下のクラスを作成した。

Login ユーザの認証を行う。セッション変数へユーザの id を登録。

Kserv ユーザの入力を元に google にクエリを発行し、結果の解析を行う。結果を HTML にしてユーザに提示。検索キーをデータベース (input_keys テーブル) へ登録。

DoodleList ユーザの入力を元にデータベースへを検索し、提示すべきキーワードのリストを HTML にしてユーザへ提供する。ユーザの入力と提示したキーワード、時刻などをデータベース (logs テーブル) へ登録。

5.5 データベーススキーマ

以下のようにデータベースを作成した。

表 5.3: データベース

データベース	doodle	
テーブル	input_keys	入力された検索キー
	logs	ログ (評価のため)
	users	ログ (評価のため)

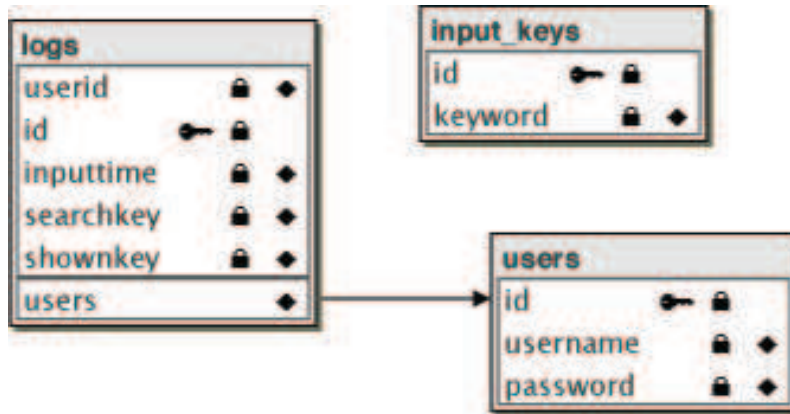


図 5.2: データベーススキーマ

5.5.1 input_keys テーブル

input_keys テーブルは実際にユーザが入力した検索キーを蓄積する。

表 5.4: input_keys テーブル

id	整数型	input_keys テーブルの primary key
keyword	文字列型	入力されたキーワード

5.5.2 users テーブル

users テーブルは本システムのユーザアカウントを蓄積する。ログの解析時にトレースを行うために使用するが、実際の運用では不要の場合もある。

表 5.5: users テーブル

id	整数型	users テーブルの primary key
username	文字列型	ユーザー名
password	文字列型	パスワード

5.5.3 logs テーブル

ユーザからの入力やユーザに対して提示したキーワードをログとして記録する。今回は次章の評価のために作成したが、実際の運用では不要の場合もある。

表 5.6: logs テーブル

id	整数型	logs テーブルの primary key
inputtime	時刻型	キーワードが入力された時刻
searchkey	文字列型	ユーザが検索他のために入力した文字列のリスト
shownkey	文字列型	カウンタとしてユーザに提示された語のリスト

5.6 キーワードの提示部

ユーザの入力のカウンタに提示するキーワードについて、データベース内のどの語をどの順番で提示するかについては、語と語の関連性に対する重み付けをするなど、いくつもそのポリシーが考えられる。今回はそのポリシー自体は研究の対象外とした。提示語は入力された検索キーとともに用いられた語を全て、順番は新しい(最後に検索された)順番とした。

5.7 ユーザインタフェース

ログイン画面でのユーザ認証の後、図 5.3 のような入力画面が表示される。ここで検索キーを入力すると図 5.4 のような画面に遷移する。この画面はフォームで以下のように 2 つに分けられている。

上半分の検索結果表示部には、さらに絞り込みを行うための絞り込みフォームと検索結果が表示されている。通常の検索エンジンのようにヒットページのタイトルや URL 等がリスト表示されている。リンクをクリックすることで別ウインドウが開いてそのページを読み込む。絞り込みフォームが post されるたびに新しい結果リストで更新される。

また、下半分のキーワード提示部にはキーワードをリスト表示する。検索結果表示部

が読み込まれる度に新しいキーワードリストで更新される。

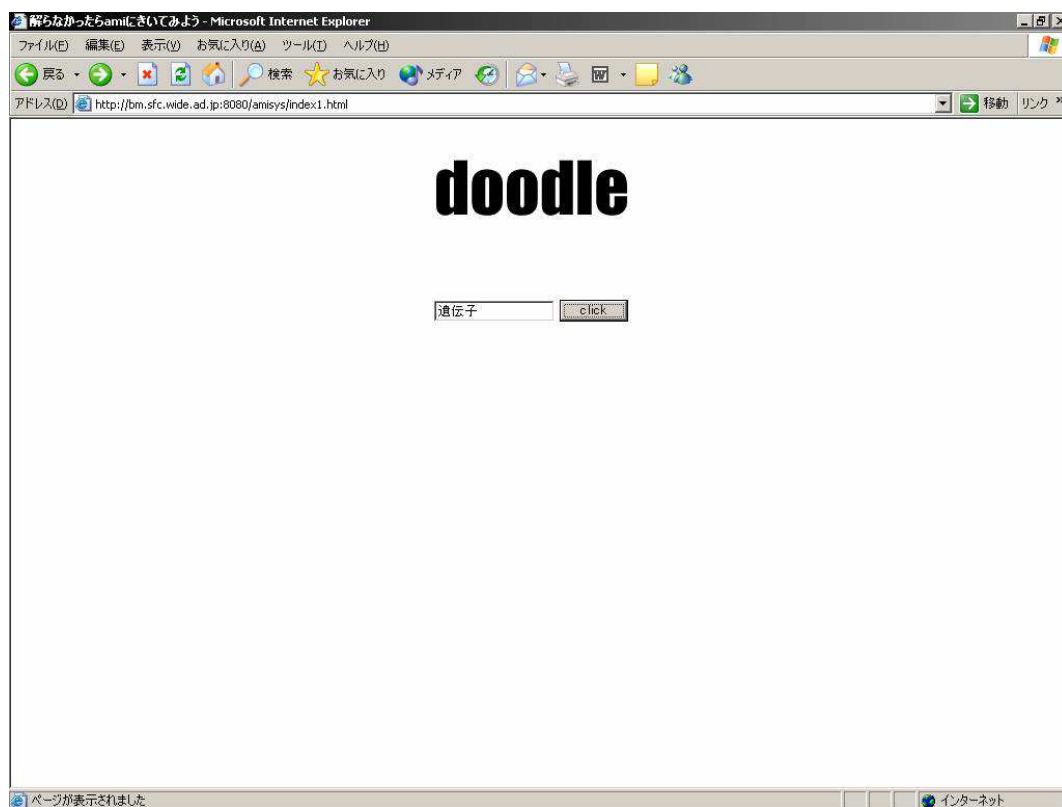


図 5.3: 検索画面のスクリーンショット

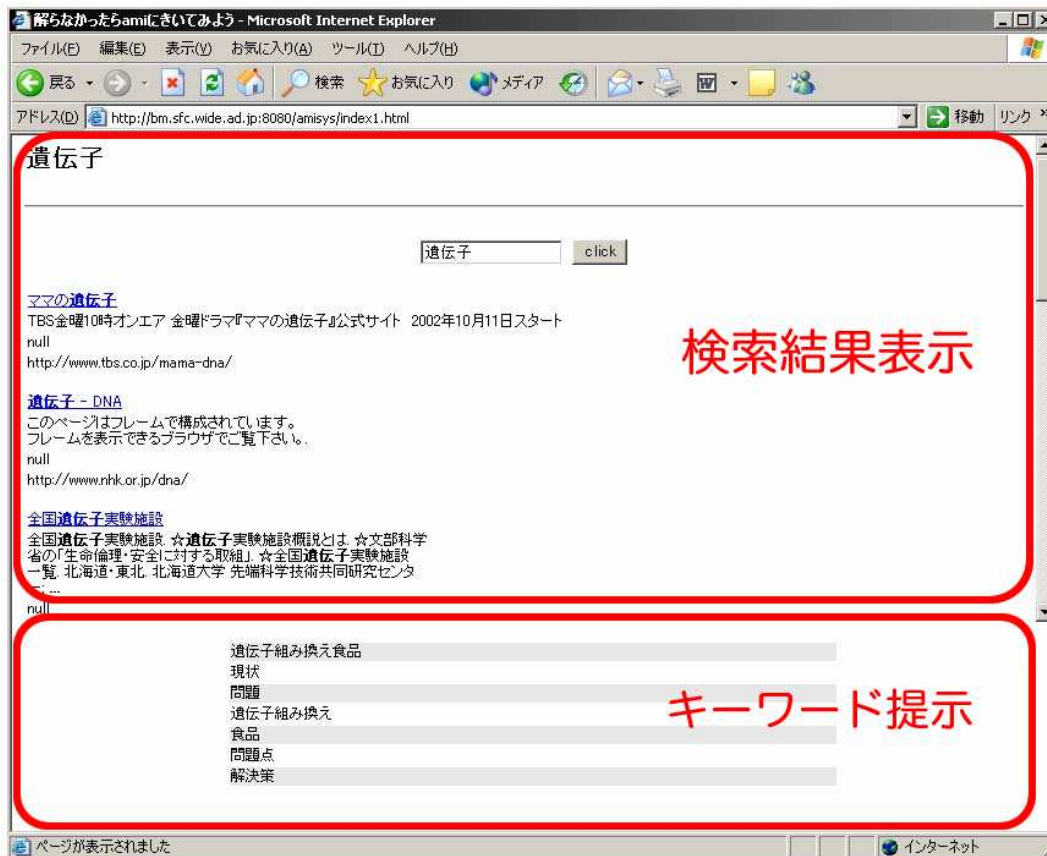


図 5.4: 検索結果のスクリーンショット

第6章 システムの評価

本章では、本システムの目標としていた、同じ目的を持ったグループ内で検索キーワードを共有する事は有益であるという仮定の検証を行う。また、本システムが価値を見出す条件をいくつか設定し、どの条件でどのように本システムが役立つかを評価する。

6.1 評価項目

以下に評価項目を示す。

システムの要件

まず、本システムの要件を満たしたかどうかを検証する。

1. ユーザ側でのアプローチ
2. キーワードを共有する機構
3. 設置場所の自由度
4. 目的の自由度

1. のユーザ側でのアプローチは、システムを検索エンジンの手前の部分に設置する事で実現した。同時に、2. に関しては、キーワードを共有する機構は提供する事ができた。後にシステムの work する条件に関して述べる。3. 設置場所の自由度は、検索エンジンからキーワードの入力部分を切り離す設計によって、達成した。これによって、同時に、4. の目的の自由度も確保された。

6.2 システムの機能する条件の検証（実験）

被験者 12 名に対して、事前調査を行い、得意分野ごとにグループ分けした。次にパラメータを設定し、システムが機能する条件を見出す実験を行った。

6.2.1 被験者に関する事前調査

メールを用い、以下の項目についてプレアンケートを行った。アンケート項目を以下に挙げる。

- 検索エンジンの使用頻度
- 普段使用している検索エンジン
- 検索エンジンの使用場面
- 情報検索に対する自信
- 得意分野

被験者は皆、使用頻度は一日数回から1時間に数回、と異なるが、少なくとも一日数回は検索エンジンを使用しており、普段使用している検索エンジンは、90%以上がgoogleであった。また、検索エンジンの使用場面として、わからないことがあった場合に検索エンジンを使用するかという質問に対して、80%以上がとてもあてはまると回答した。しかし、情報検索に対する自信は、様々であり、「とても自信がある」という回答はなかった。

6.2.2 実験環境

事前調査によって、被験者は検索エンジンの使用に関しては、ある程度慣れていることがわかっていた。被験者はgoogleの使用頻度が高いため、ユーザインターフェースの違いで、使い勝手に差が出ないように、デザインはgoogleと変わらないデザインにした。

6.2.3 実験手法

実験は条件を設定し、それぞれの条件における本システムの有効性を調査した。

趣味ごとに3つのグループをつくり、それぞれ3回ずつおこなった。その際に設定した条件の要素は、以下のとおりである。

グループ ユーザの属性を事前に調べ、得意分野が同じユーザ同士のグループを車・野球・料理のテーマで4人ずつ、3つのグループに分けた。

問題の分野 ユーザの得意分野の問題を3つ作成した ..

問題の種類 検索の種類によって、問題を3種類作成した。検索の種類とは、ある大きなトピックの全体像を把握する検索、ある小さなトピックの詳細知識を得る検索、答えが一意に決まっている問題の検索の3種類を想定して、問題を作成した。

データベースの成長 ユーザが使用するうちにDBのキーワードは蓄積される。蓄積の成長を、今回は3期に分けて実験した。

図6.1は実験の流れを示したものである。実験は、それぞれ得意分野ごとのグループに分け、得意分野の問題と得意でない分野の問題を本システムを使用して解く。この際、被験者は、グループごとにローテーションで同じ問題を解き、最終的には全ての問題を解くようにした。

キーワードデータベースは、問題の分野ごとに設置し、被験者の使用した検索キーワードを蓄積した。

	I	II	III
問題[車]	車グループ 得意分野	野球グループ	料理グループ
問題[野球]	料理グループ	車グループ	野球グループ 得意分野
問題[料理]	野球グループ	料理グループ 得意分野	車グループ

時間(DBの成長) →

図 6.1: 実験概要

6.3 結果

前節で示した条件ごとに、結果を分類した。

6.3.1 問題の分野と提示キーワードの使用度

得意な分野と得意ではない分野における、提示キーワードの使用度をグラフで表し(図6.2)、グラフから読み取れる事を以下に示す。図6.2のグラフは、問題の分野と提示キーワードの使用度との関係を、グループごとに示したものである。グラフの縦軸は、被験者に提示キーワードの使用度を5段階で示してもらったものである。横軸の上段はデータベース、下段はグループを示す。(グラフ中縦線は、データの最高値と最小値を示し、黒い点は、データの平均値を示す。)

- 得意分野での検索では、提示キーワードは使用しない

- 得意でない分野での検索では, 関連キーワードを使用する

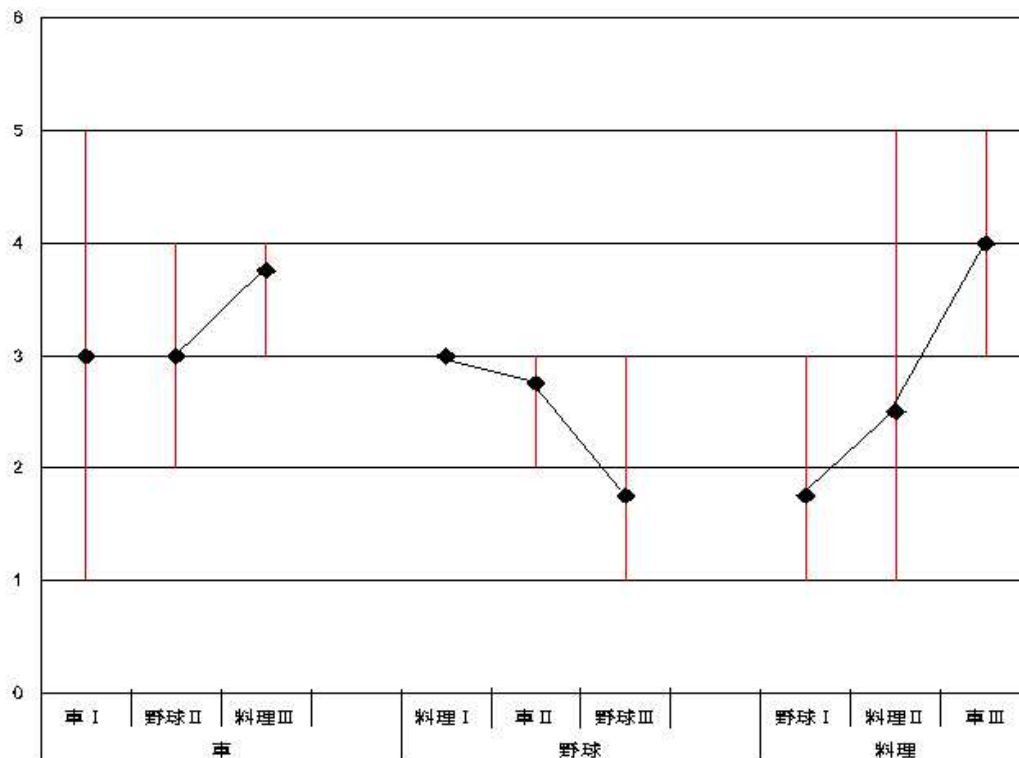


図 6.2: テーマと提示キーワードの使用度

6.3.2 データベースの成長

データベースの成長とユーザの提示キーワードの使用度の変遷や, キーワードの共有を有益だと思うかについて, データベースの成長と関わりがあるか, を調査した.

図 6.3はそれぞれのデータベースの時間ごとの使用度の変遷を示したものである. グラフの縦軸は, 被験者に提示キーワードの使用度を5段階で示してもらったものである. 横軸は上段が時間の変遷, 下段がデータベースの名前である. (グラフ中縦線は, 最高値と最小値を示し, 黒い点は, データの平均値を示す.)

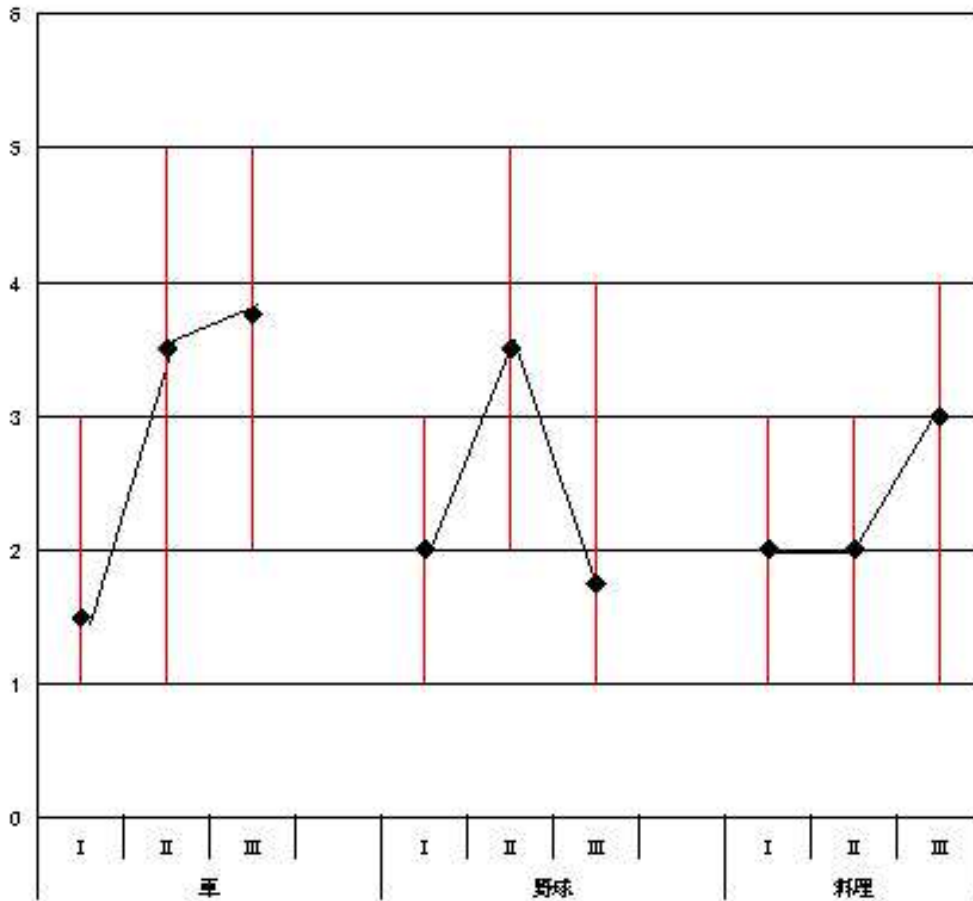


図 6.3: データベースの成長と提示キーワードの使用度

図 6.4 はそれぞれのデータベースの時間の変化と有益だと思うかどうかの変遷を示したものである。グラフの縦軸は、被験者に提示キーワードを有益だと思うかどうかを 5 段階で示してもらったものである。横軸は上段が時間の変遷、下段がデータベースの名前である。(グラフ中縦線は、最高値と最小値を示し、黒い点は、データの平均値を示す。)

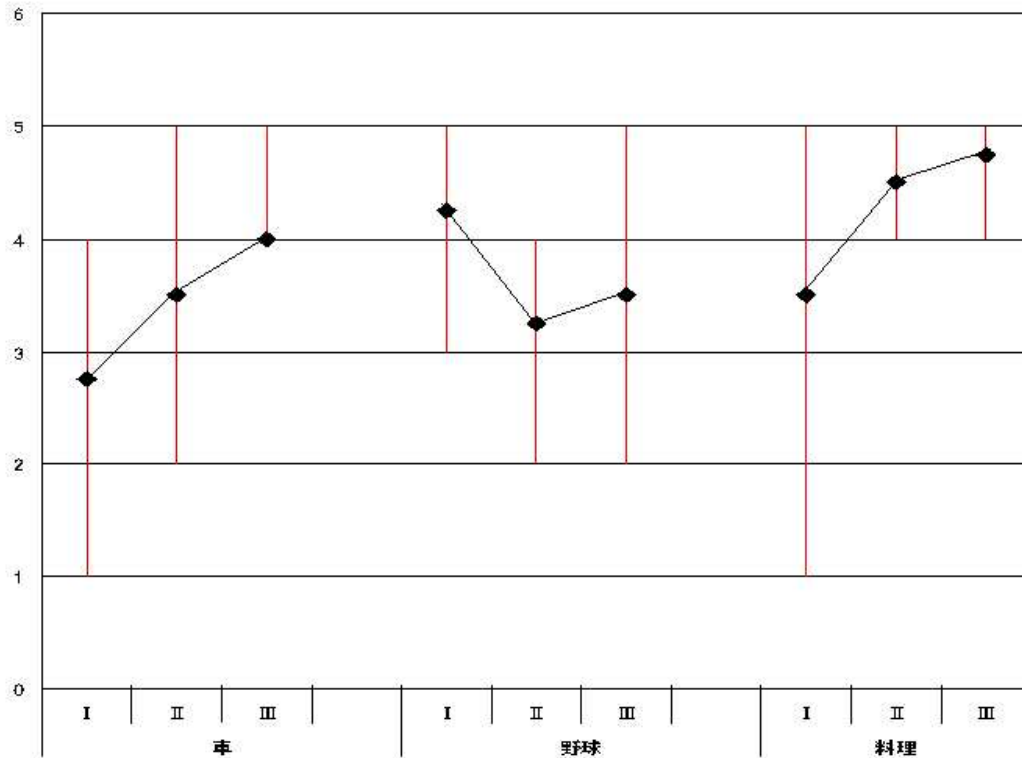


図 6.4: キーワードの共有は有益だと思うか (DB の成長ごとの推移)

6.3.3 自由記述

表 6.1 は、被験者に、提示キーワードを使用したか否かと、その理由をまとめたものである。1~5 の数字は使用頻度を 5 段階で回答したものである。提示キーワードの使用頻度の高い被験者は、「おもしろい」「便利」と感じ、使用頻度の低い被験者は、「(提示キーワードが) 表示されなかった」「必要なかった」というものが主であった。

表 6.1: 提示キーワードの使用度とその理由

提示キーワードを使ったか	理由
5	便利だから．おもしろいから．困った時は提示キーワードを使えばよい事を知っていたから（探し方が）全くわからなかったから．
4	使えそうなキーワードがあったから．検索目的に一致した．ヒントっぽかったから．あまり考えず，とりあえず使った．質問そのままのキーワードだった．自分の設定したキーワードで検索がうまくいかなかったら
3	表示されなかった．試しに使ってみたかった．
2	答えが見つかった時の検索で，提示キーワードが出たから（使う必要がなかった）．
1	表示されなかった．自分で検索できる話題だった．簡単に見つかったから．

以下は「キーワードを他者と共有することが有益だと思うか」という問いに対する回答である．

● 肯定的な評価 [4,5]

- － 自分では思いつかない単語が出てきたのうれしいかった
- － 他人が検索キーとして使っている単語が知れて参考になった
- － トリガになる
- － キーワードをある程度絞り込むと目的に一致した
- － キーワードが得られやすくなり，検索効率があがった
- － 自分が思いつかないキーワードを知ることができ，検索の効率は絶対的に上がると思う
- － クリック回数が減る
- － うまく検索できなかつたから他者のを参考にしたいと思った
- － たのしい
- － 答えられないのは悔しいので，他の人に聞きたくなった
- － 他者がどのような言葉で検索をかけているか，自分とは違った思考で書いたものなのでおもしろかった
- － キーワードがつきはてた時，他人のキーワードは有効だと思う

- 検索しようと思うときは、情報が多い方がよい(ヒントが多い方がよい)
 - 自分では思いつかないキーワードを思いつくことができるから
 - 自分が入力しないキーワードは新しい発見になる
 - 自分の思い浮かばないキーワードが出てくるため、ヒントが新たに、検索するたびに増えるから
 - 他者が問題をどのように捉えているかを知ることができるため、その問題をよりの確に把握できる
 - 他者のキーワードを使うと新しい視点から検索できる
 - 他者の視点を知ること、自分のインスピレーションが生まれた
- 否定的な評価 [3,2,1]
 - 普段の検索でも見つけ出せる内容だと思う
 - 検索は組み合わせで行うのでどうせなら組み合わせ提案をしてほしいと思った
 - 他人と探したいものは違うから、キーワードはきっかけにしかならないと思った
 - 簡単にみつかったから有効性がわからなかった
 - ヒントがでると、調べようという気になった。同時に緊張感もあった。

6.4 考察

本実験において、得意/不得意分野による提示キーワードの使用度・システムの満足度の差異は顕著であった。不得意分野における提示キーワードの有用性を示す事ができた。

また、データベースの成長に関しては、本実験の中でデータベースが習熟したと言える点まで、動かす事ができなかつたため、I期、II期、III期といったデータベースの成長時間に区切ったが、それらに差異はあまり見られなかつた。

しかし、提示キーワードを使用しなかつた理由として、キーワードが提示されない事を挙げる例が多かつた。これは、データベースの成長が多少あつた事と、1,2回のクエリ送信で目的を達成してしまう上級ユーザであつたという二点が原因であつた。

6.4.1 得意分野を持ったユーザの存在

得意分野の問題を調べる際、ユーザは提示キーワードを見る必要もなく、自らのバックグラウンドを活かして検索をしていた。しかし、得意分野でない問題を調べる際、キーワードが思い浮かばない、うまくいかない、自信がない等の理由で提示キーワードを利用していた。

6.4.2 情報検索能力の個人差

情報検索の個人差に関して、キーワード設定が問題であると仮定して、本システムを構築したが、実験を経て、キーワード設定を支える連想力以外の部分もあった。クイズ形式の問題において、正答を記載している web ページにたどり着いているにも関わらず、それを解答とすることができなかつた被験者がいた。また、同じ語幹・意味の言葉を使用している、用いる接頭・接尾語、フレーズ検索といった検索のための工夫を施すスキルを持つユーザもいた。

実験の様子から、効率的な情報検索のできないユーザを以下のように分類することができた。

- キーワードが思い浮かばない（連想力）
- web ページ内から欲しい情報を見つけることができない（読解力）
- キーワード設定において、AND 検索をする数が多すぎるためにフォーカスが外れるほど絞ってしまう

本研究では、適切なキーワードを設定することができないユーザをサポートする事が目的であったので、自分の不得意分野でのキーワード設定を、同じ目的をもった他者の視点でのキーワードを提示した点、自分の連想力で思い浮かばないキーワードによる新しい発見を与えたという点で価値があったと言える。

6.4.3 考察のまとめ

不得意分野の情報検索を行う際は、検索効率の悪いユーザとなり得ると考え、得意/不得意分野の問題を検索させた。また、データベースの成長とともに、有用なキーワードを提示することができると考えたが、データベースの成長による差異よりも、得意/不得意分野どちらかという条件による差異の方が顕著であった。

全体的に、どのグループも第一回の実験は、提示キーワードの使用度が低かった。これはデータベースの成長が未熟である事に加え、ユーザのシステムの習熟度も影響していた。

6.4.4 今後の課題

本実験では、データベースをゼロの状態から使用しはじめたので、データベースが習熟する点・閾値を超える点を計測する事はできなかった。データベースの成長と提示キーワードの質に関して、評価する事ができなかった。

また、キーワードを共有する範囲を同じ目的を持ったグループ内に絞ったが、データが蓄積されるにつれ、有用なキーワード、不用なキーワードという観点が必要になる。これらを考慮したキーワードのランキングアルゴリズムの開発が必要になる。

検索の上級者の用いたキーワードを上位に表示する事が望ましいが、テーマやシチュエーション、上級者の定義が困難である。人数の制約（価値あるキーワードが蓄積される人数から不用なキーワードが増えてしまう人数）、時間（キーワードデータベースの成長から不用なデータが蓄積されてしまうまで）のしきい値が存在するはずである。これらを検証することが、これからの課題である。

第7章 おわりに

7.1 まとめ

本論文では、情報検索における、検索エンジンでのキーワード設定を補助するシステムを構築した。

情報検索における検索能力の個人差は、個人のキーワード設定能力の差によると考え、そこから情報検索を新しいモデルで捉えなおし、ユーザ側でのアプローチを行った。

キーワード検索におけるキーワードこそが、共有すべきナレッジであると考え、同じ目的を持つグループで利用できる自由度のあるシステムを設計する事で、よりナレッジシェアを実現しやすいシステムを構築した。

7.2 今後の展望

検索の得手・不得手によって、使用したキーワードのポイント加算に重み付けをするなど、検索者の性質を考慮したシステムをつくると、より有益なキーワードの提示が可能になる。

また、検索結果に満足した事を知らせる機能を実装する事によって、一連の情報検索を括る事ができれば、最後のクエリが情報検索に最も効果的なキーワード組であると考え、重み付けをする必要がある。

謝辞

本研究を進めるにあたり、御指導を頂きました、慶應義塾大学環境情報学部教授の徳田英幸博士と村井純博士に感謝致します。また、絶えずご助言とご指導を頂きました、同学部助教授の中村修博士と楠本博之博士、専任講師の南政樹氏に深い感謝の意を表します。

テーマの定まらない時期からご指導とご助言を頂きました慶應義塾大学大学院理工学研究科開放環境科学専攻後期博士課程 江木啓訓氏、慶應義塾大学政策メディア研究科博士課程 石田剛朗氏をはじめとする先輩方に、心から感謝いたします。日頃の議論や相談に付き合ってくださいました研究グループ neco の皆様を筆頭に村井研究室の皆様へ深く感謝致します。

特に実装、論文執筆時にご助言を頂きました慶應義塾大学総合政策学部 久松慎一氏、須子善彦氏、英文の作成をお手伝い頂いた慶應義塾大学環境情報学部 金子紘子氏、SFC 研究所訪問研究員 Thierry Ernst 氏には大変お世話になりました。

研究への助言、励まし、気分転換の全面的援助を頂きました株式会社三菱総合研究所リサーチプロフェッショナル佐藤雅明氏に深く感謝致します。

さらに、忙しい中実験につきあってくれた慶應義塾大学環境情報学部 清水崇史氏をはじめ、高橋宏明氏ほか後輩達に感謝致します。様々な局面で励まし支えてくださった馬谷亜古氏、橋本和樹氏、成瀬大亮氏、千代佑氏をはじめ、楽しい研究室生活をともにした村井研究室の皆様へ感謝致します。

同じ卒論を書く立場でありながら、力を貸して下さった渡里雅史氏、入野仁志氏、論文執筆中励まし元気づけてくれた鈴木貴晶氏、吉村知夏氏、rg-99 の皆様へ感謝します。

そして最後まで研究・論文執筆を温かく見守り、生活を補助して下さった両親に心から感謝致します。

最後に、栗本央氏と山田靖子氏の結婚を心から祝福し、愛車エメロードと共に本論文を捧げます。

参考文献

検索エンジン

- [1] <http://www.google.co.jp>
- [2] <http://www.teoma.com>
- [3] <http://www.altavista.com>
- [4] <http://www.wisenut.com>
- [5] <http://www.vivisimo.com>
- [6] <http://www.kuamp.kyoto-u.ac.jp/labs/infocom/mondou/>
- [7] http://bace.hgc.jp/dn_intro/

論文

- [8] T.Kroeger, D.Long, and J.Mogul. Exploring the Bounds of Web Latency Reduction from Caching and Prefetching. In Proceedings of the USENIX Symposium on Internet Technologies and Systems, December 1997.
- [9] C.Maltzahn, K.Richardson, and D.Grunwald. Performance Issues of Enterprise Level Web Proxies. In Proceedings of the SIGMETRICS Conference on Measurement and Modeling of Computer Systems, June 1997.

書籍

- [10] 著: P. イングベルセン情報検索研究：認知的アプローチ 1995 トッパン

web

- [11] 著:japan.internet.com 編集部 ”検索結果は上から順に。上位表示サイトのクリック率高まる、87 %” 2002年3月20日 <http://japan.internet.com/research/20020320/1.html>
- [12] 著:japan.internet.com 編集部 ”膨大な検索結果に戸惑い、ネットでの情報検索、約6割が「自信ない」” 2002年2月1日 <http://japan.internet.com/research/20001214/1.html>
- [13] <http://searchenginewaatch.com>
- [14] <http://www.infotoday.com/newsbreaks/nb020708-1.htm>
- [15] 著:japan.internet.com 編集部 ”85%以上のインターネットユーザーは、検索エンジンを通じて商品やサービスを調べている” 2001年11月 <http://japan.internet.com/>
- [16] 著:博報堂 広報室 ”10～60代のブロードバンド生活者約1万人に調査” 2002年10月11日 <http://www.hakuhodo.co.jp/news/pdf/20021011.pdf>
- [17] 著:japan.internet.com 編集部 ”検索ポータルを使い分けは常識？44%が複数を利用” 2002年8月28日 <http://japan.internet.com/research/20020828/1.html>
- [18] 著: Search Engine Optimization.jp ”SEO 検索エンジン最適化 ” <http://www.searchengineoptimization.jp/index.html>
- [19] 著: google Web API **google Web API** <http://www.googleAPI.com>
- [20] 著: google Toolbar **google Toolbar** <http://toolbar.google.com/intl/ja/>

付録

.1 プレアンケート

1. 検索エンジンを利用する頻度
(1日1回以下・1日数回・1時間に数回以上)

答

2. 普段使っている検索エンジンは？
複数ある場合は、だいたいの割合を示してください。
[ex. google 80 % Yahoo! JAPAN 20 %]

答

3. 普段わからないことがあったら、検索エンジンを使う
(5. とても当てはまる 4. まあ当てはまる 3. どちらとも言えない
2. あまり当てはまらない 1. 全く当てはまらない)

答

4. 検索エンジンで情報を探し出す事に自信がある
(5. とても当てはまる 4. まあ当てはまる 3. どちらとも言えない
2. あまり当てはまらない 1. 全く当てはまらない)

答

5. 次の中から取ったことのある授業、もしくは得意分野を最も得意な順に選んでください。(3つ以内)

(認知科学 生物 国際政治 建築・環境デザイン 環境問題 メディア
論 人事・組織論)

答

.2 ポストアンケート

1. 提示キーワードを使用するか
2. 提示キーワードを見て新しい発見があるか
3. 自分の使用したキーワードに自信が持てるか
4. ある大きなトピックの全体像を把握するための検索に, 提示キーワードは役立つか
5. ある小さなトピックの詳細知識を得るための検索に, 提示キーワードは役立つか
6. 使用したキーワードが表示される事は, ユーザのキーワード設定の質の向上のモチベーションにつながるか
7. 普段の検索と比較して, 情報を探す事を諦めなくなるか
8. キーワードを他ユーザと共有する事は有益だと思うか