

平成20年度

筑波大学第三学群情報学類

卒業研究論文

題目

Web上のキーワードを辿る
視覚的な情報探索インターフェースの開発

主専攻 情報科学主専攻

著者 結城 崇

指導教員 三末和男, 志築文太郎, 高橋伸, 田中二郎

要 旨

検索結果を比較することで、様々な情報が見えてくる。例えば、検索結果の一部である検索ヒット数を比べることで、言葉の意味や信頼性を読み取れることがある。しかしながら、現在の Web 検索サービスには、複数のクエリに対する検索結果を比較するインターフェースが提供されていない。本研究では、検索ヒット数の比較を可能にする、円を用いた Web 検索ヒット数の視覚化手法を提案する。

また、Web を用いた情報探索の際に、目的のコンテンツを指す適切なキーワードを想起できないという場面がよく見られる。そこで本研究では、キーワードを探索するインターフェースを提案する。探索の対象を抽象度の高いキーワードとすることで、発散的な探索を行うことができる。発散的な探索により、より多くのキーワードと遭遇でき、適切なキーワードの想起を可能にする。

本研究では、先に提案した視覚化手法とインターフェースを組み合わせ、Web での情報探索を支援するツール Searchle を開発した。また、Searchle に対して評価実験を行い、その結果から、本ツールが Web での情報探索に有効であることを示した。

目次

第 1 章 序論	1
1.1 情報探索	1
1.2 現在の Web の状況	1
1.3 Web 検索結果の比較	2
1.4 本研究の目的	2
1.5 本研究の貢献	2
1.6 本論文の構成	2
第 2 章 Web 情報探索に関する既存のアプローチ	3
2.1 既存のアプローチの分類	3
2.2 Web ページを探索空間とするアプローチ	3
2.2.1 Web 検索サービス	3
Web 検索エンジン	3
検索クエリの拡張	4
2.3 コンテンツを探索空間とするアプローチ	4
2.3.1 グラフ構造での視覚化	4
2.3.2 グラフ構造以外の視覚化	4
2.4 キーワードを探索空間とするアプローチ	4
2.4.1 ディレクトリ型検索	4
2.4.2 Web 上のキーワードの視覚的な提示	5
2.4.3 インタラクションによる探索	5
2.5 その他	5
2.5.1 Web ページの閲覧を支援するアプローチ	5
2.5.2 自己完結型の Web サービス	6
第 3 章 本研究のアプローチ	7
3.1 キーワードの発散的な探索	7
3.2 Web 上のキーワードのボリューム感の提示	7
3.3 円を用いた検索ヒット数の視覚化	7
第 4 章 Searchle:Web を用いた情報探索を支援するツール	9
4.1 Searchle の概観	9

4.2	Searchle の構成	10
4.2.1	円へのインタラクション	10
	ビュー内での再帰的な Web キーワード探索	10
	図の移動	11
	ズームング	11
4.2.2	詳細情報の提示	12
	表での Web 検索ヒット数の提示	12
	Web ページタイトルの提示・閲覧	13
4.2.3	円のレイアウト法の変更	13
	表示する円の個数の切り替え	14
	円の間隔	14
	Log スケール	15
	円の自動サイズ調整のモード切り替え	16
4.3	Web 情報の視覚化手法	18
4.3.1	面積から円の中心間の距離を求める計算法	20
4.3.2	複数の円のレイアウト	21
4.4	Searchle の利用シーン	22
4.4.1	辞書的な Web の利用	22
4.4.2	曖昧な探索	22
4.4.3	流行の把握	22
第 5 章	実装	23
5.1	本システムの処理の流れ	23
5.1.1	YahooApi からの検索結果の取得部	24
5.1.2	レイアウト	25
第 6 章	評価実験	26
6.1	評価実験の目的	26
6.1.1	評価実験の狙い	26
6.2	評価実験の概要	26
6.2.1	被験者	26
6.2.2	評価実験の手順	27
6.3	評価実験の結果	27
6.4	考察	28
6.4.1	ツールの有効性	28
6.4.2	各モードの評価	28
6.4.3	今後の課題	29
	モードの切り替え	29
	複数クエリの対応	29

円へのインタラクションの拡張	29
検索結果を取得するサービスの切り替え・拡張	29
第7章 結論	30
謝辞	31
参考文献	32
脚注	34

図目次

3.1	ベン図の例	8
4.1	開発したツールの概観	9
4.2	キーワードの選択例	11
4.3	ズームングの例	12
4.4	詳細情報を示す表	13
4.5	Web ページタイトルの提示例	13
4.6	円の個数切り替えの例	14
4.7	円の集合の間の距離の変更	15
4.8	Log スケールでの描画例	15
4.9	ローカルモードでの描画例	16
4.10	ローカルモードの説明	17
4.11	グローバルモードでの描画例	17
4.12	グローバルモードの説明	18
4.13	映画をクエリとした時の描画例	19
4.14	円配置の計算法の説明 (1)	20
4.15	円配置の計算法の説明 (2)	21
4.16	円の配置順序	21
5.1	全体のフロー図	23
5.2	関連語と検索結果を取得する際のフロー図	24
5.3	レイアウトのフロー図	25

表 目 次

4.1	映画を入力クエリとした時の Yahoo 関連語上位 10 件と検索ヒット数	19
6.1	タスク 1 の設問に対する各被験者の回答	27
6.2	タスク 2 の設問に対する各被験者の回答	28

第1章 序論

1.1 情報探索

本研究では、情報を探し自身の知識として吸収し、その知識を基に新たな情報を探すという一連の流れを、情報探索と呼ぶことにする。言いかえると、再帰的に情報を収集することである。それに対して、情報検索は、データベースや辞書から必要な情報を取得することを指す。したがって、情報検索は、情報探索をする際の一つの手段と考えることができる。情報探索をする場面は、日常生活でもよく見られる。例えば、「新商品が発売された」という情報を得た時に、その商品を販売している「会社の名前」を Web 検索で調べ、その「商品の名前」を知るといった一連の流れが、情報探索である。

また今日では、情報探索の手段として、Web が用いられるようになってきた。これは、Web の持つ情報量の多さとオンデマンド性のサービスのためである。

1.2 現在の Web の状況

現在、世界に存在する Web ページ数が膨大なものとなっている。また、年々 Web ページの数は増加している。2008 年 7 月時点では、世界に 10 兆もの URL が存在していると GoogleBlog¹ が公表している。

また、Web ページには、Blog や日記といった個人的なコンテンツ、コンピュータ言語のサンプルや Tips のような教科書のようなコンテンツ、国や企業のホームページなどの公的なコンテンツなどの様々な種類のコンテンツが混在している。多種多様な情報を持ち、膨大な数が存在する Web ページの中から信頼性の高い Web ページ や得たいの Web ページを見つけ出すのは、困難である。

Web 上で信頼性の高い Web ページを見つけたい時には、検索エンジンが利用される。検索エンジンでは、特定のクエリに対して検索結果の上位しか閲覧されていない。そのため、検索結果で上位になるということが、重要視されるようになった。そして、検索結果で上位を獲得するために、検索エンジン最適化 (SEO) 技術が競われるようになった。SEO とは、Web ページの内容を向上させるのではなく、html ソースの書き方などを変えることで検索上位を取得するというものである。そのため、必ずしも Web ページの信頼性と検索結果の順位がかみ合わなくなってしまった。

このような状況で、Web 上で目的の情報を探索することは、困難になってきている。

1.3 Web 検索結果の比較

基本的な Web 検索結果は、クエリに対する検索ヒット数、Web ページ (順位、タイトル、URL) で構成される。Web 検索結果の比較により、読み取ることのできる情報は数多く存在する。例えば、類義語をクエリとした時に、検索上位の Web ページを見比べることで、Web ページから類義語の意味の違いを読み取ることができることがある。これは、検索結果を比較することで、クエリに対する Web ページの概観を把握することができるためである。

1.4 本研究の目的

本研究では、Web を用いた情報探索を支援することを目的とする。Web 情報探索の支援のためには、クエリに対する Web ページの概観の把握が必要であると考えた。そこで、Web を用いた情報探索支援のために、検索結果の比較を容易にする視覚的なインターフェースの開発を目指した。

1.5 本研究の貢献

本研究で提案したインターフェースによって、これまでには困難であった複数のクエリに対する検索結果の比較を容易にした。また、複数のクエリに対する関連語とその検索ヒット数を視覚によって比較できるようになった。複数クエリの検索結果の比較によって、これまでよりもキーワードの概観をつかむことが容易になった。これらにより、目的に対して適切な Web 検索クエリを想起させると共に、Web での情報探索の機会を増加させ、ユーザが目的に適した自分の意図通りのクエリを構築できるようになった。

1.6 本論文の構成

第 2 章では、1.4 節で述べた目的に対する既存の Web サービスを紹介する。第 3 章で、Web での情報探索を支援するアプローチについて述べる。第 4 章では、本研究で開発した Searchle の機能を紹介する。第 5 章では、Searchle の実装について述べる。第 6 章では Searchle に対する評価実験について、第 7 章に関連研究について、第 8 章に結論を述べる。

第2章 Web 情報探索に関する既存のアプローチ

2.1 既存のアプローチの分類

また、各アプローチを探索空間により分類する。分類は以下の通りである。

- Web ページを探索空間とするアプローチ
- コンテンツを探索空間とするアプローチ
- キーワードを探索空間とするアプローチ
- その他

2.2 Web ページを探索空間とするアプローチ

2.2.1 Web 検索サービス

莫大な数の Web ページの中から、目的の情報の探索を支援する。Web 検索サービスは、Web 探索を行う際、日常的に使用される。

Grokker²では、検索結果に対してリランキングを行うことができる。リランキングとは、入力したクエリに対する検索結果を、さらに絞り込むことである。このサービスでは、日付による Web ページの絞り込み、テキスト検索を提供している。

Clusty.com³は、vivisimo 社の提供するクラスタリングメタサーチエンジンである。クラスタリングとは、検索結果に対し、出現頻度の高いキーワードによる分類を行い、キーワードごとに検索結果の Web ページをまとめることである。Hyperbolic Tree[9] は、検索結果をクラスタリングし、木でそのリンク構造を視覚化した。

これらのサービスは、一つのクエリに対する検索結果を詳細に把握するのに有効である。また、収束型の探索であり、検索結果の中から目的の Web ページに辿りつくことが目的である。

Web 検索エンジン

リンク構造や内容から Web ページにランキングを付け、そのランキング情報から探索を行う。Web 検索サービスを提供している Google, Ask.com は、RageRank や HITS などの代表的な検索アルゴリズムを採用している。信頼性の高い Web ページを探索する時に有効である。

検索クエリの拡張

これまで、Web 検索で用いるクエリはテキストであった。しかしながら、これから紹介するサービスは、画像や色などをクエリとして検索を行うことができる。like.com⁴では、絵を表す言葉をクエリに用いることができる。例えば、「尖った靴」で検索すると、尖った靴を履いている有名人の写真などが結果として返される。xcavator⁵では、色をクエリにして、色調がクエリと似ている画像を検索できる。retriever⁶では、描いた絵をクエリに用いることができる。

2.3 コンテンツを探索空間とするアプローチ

2.3.1 グラフ構造での視覚化

Web 上のコンテンツに関する情報を視覚的に提示することで、コンテンツの相関や関連などの把握を支援する。書籍を視覚的に見せる amazonode⁷は、書籍同士の関連度をエッジ、書籍をノードとするグラフ構造で表現する。Touch Graph⁸は、Google, Amazon, FaceBook から情報を取得し、関係のある Web ページや商品や人が線で結ばれるグラフ構造で表示する。Web ページとリンク構造をノードとエッジを用いて視覚的に表現する研究 [2] は、Web ページの持つリンク構造が複雑な時にその Web ページを一つのノードにまとめ、エッジの交差を少なくする。LOGEMON[5] は、Web ページとそのリンク構造を視覚的に表示する。Web ページを用いた授業での利用を考えて開発された。Web ページの可視化の Web 教材に対する適応例である。PIRCS[16] は、3 軸を持つデカルト座標系のグラフで 3 D 表現する。3 つの指標まで自由に選択することができる。また、ノード間を線で結ぶ。

2.3.2 グラフ構造以外の視覚化

liveplasma⁹は、音楽や映画の情報を円を用いて視覚的に表示する。また、円は重なりを持たず関連のあるものは、線によって結ばれる。searchCrystal¹⁰は、5 つの座標系を持つ円内にコンテンツを配置し、その配置場所によって、属性を表現する。画像、動画、Web ページなどに対してサービスを提供している。oSkope visual search¹¹は、写真や動画を並べて配置し、現実世界で行うように、写真などを自分の好みの通りに写真を配置し、画像の探索を行う。[10] は、Web ページに関する 2 つの指標のデカルト座標系グラフにランキング情報を対応させ、コンテンツを表す点を配置することで、視覚的にランキング情報を提示する。

2.4 キーワードを探索空間とするアプローチ

2.4.1 ディレクトリ型検索

J ディレクトリー¹² は、日本国内のポータルサイトに導入されているディレクトリ探索検索エンジンである。Web ページをファイル、Web を一つのディレクトリと見たて、探索を行

うディレクトリ型検索がある。木構造で構成されるディレクトリのように、キーワードを選択し、再帰的に下の階層へとキーワードを辿っていき、最終的に一つの Web ページに到達する。目的が、明確でない時に有効である。これは、再帰的に下の階層へキーワードを辿る、収束型の探索である。

2.4.2 Web 上のキーワードの視覚的な提示

Mooter¹³は、検索結果をクラスタリングした結果を円、円を結ぶ線を用いて表現する。この手法では、クラスタリングした結果を収束的に探索していくため、クエリの想起には有効ではない。Quintura¹⁴は、キーワードに大きさ、位置の情報を持たせて、視覚的に提示する。また、視覚化したキーワードを用いて、動的なクエリ生成を提供する。Quintura のキーワードの大きさには 3 種類しかなく、一度に読み取れる情報量が少ない。KartOO¹⁵は、検索結果をキーワード、Web ページ、その二つの相関を視覚的に表示する。Blog keyword Vizualizer¹⁶は、Blog で頻繁に使われるキーワードをグラフ構造を用いて可視化している。エッジ間の繋がりの強さを色で表現しているが、色では値を明確には読み取れない。

2.4.3 インタラクションによる探索

Carta[8] は、クエリとそのクエリに対する関連語をカルタに模した形で提示し、カルタの重ね合わせ方で AND,OR 検索の検索式を作成するインターフェースである。Web 検索用のクエリ作成の支援を目的として、関連語を視覚的に表示した研究[1]がある。まず、入力したクエリに対する話題語とその話題語に対する共起語を Web 上から抽出する。そして、抽出したクエリを相関度ごとに円型の相関グラフ上に配置し、関係のあるものは、線で結ばれる。また、話題語や共起語と入力したクエリの距離で AND,OR,NOT の演算子を変更する。

2.5 その他

2.5.1 Web ページの閲覧を支援するアプローチ

Web ブラウザは、Web を用いて情報探索をする際に、Web ページの閲覧に使用される。InternetExplorer を代表に、Firefox、Sleipnir、Chrome など様々なブラウザの開発が行われている。SpaceTime¹⁷は、3D 空間でのタブブラウザを開発した。3D 空間上に配置された Web ページのサムネイルから見たいページを選択し、閲覧することができる。このように、Web ブラウザの開発が盛んに行われている。

Searchme¹⁸は、Web ページのサムネイルをアルバムのアートワークに見立てた iTunes 風のブラウジングサービスを提供している。InfoLeed[6] は、3DWeb ブラウジングを提供する。街を模した 3D 空間に配置された柱に Web ページが貼り付けてあり、街を歩くようにして Web ブラウジングを行う。Web ページをウィンドウショッピングしているような感じを受ける。

2.5.2 自己完結型の Web サービス

近年では、一つの Web ページだけを見続けるという利用が多くなってきている。Youtube やニコニコ動画などの動画共有サイトや SNS によるものである。これらは、リンクを辿ることなく自身のサイト内だけで情報探索を完結させる。ソーシャルネットワーキングサービス (SNS) は、人との繋がりから情報を得る。Flickr などの画像投稿サイトや mixi などがある。mixi では、日記やコミュニティや友達を辿ることで情報探索を行う。

第3章 本研究のアプローチ

ここでは、Web を用いた情報探索を如何にして支援するのかについて説明する。

3.1 キーワードの発散的な探索

情報探索において Web 上のキーワード間を移動することの意義をここで説明する。Web での情報探索と言えば、Web ページやコンテンツを探すことに直結していた。そのため、Web においてキーワードは、Web から目的のコンテンツを見つけ出すための道具として考えられてきた。これは、目的のコンテンツを指すキーワードが明確な時には、有効な手段である。しかしながら、Web ページや画像などのコンテンツを指す適切な検索クエリが分かっているなければ、探索を行うことができない。そのため、本研究では、目的に対する適切なクエリを想起させるために、抽象度の高いキーワードを発散的に探索することで、目的の情報の概観をつかみ、情報を明確にさせる。キーワードを発散的に探索することで、一度の探索において検索回数が増え、より多くの情報を取得することができ、目的の情報の理解を深めることができると考えている。

3.2 Web 上のキーワードのボリューム感の提示

検索結果の比較のためには、キーワードに関する情報が必要である。本研究では、キーワードの持つボリューム感を提示することを考えた。キーワードの持つボリューム感とは、キーワード自体、検索ヒット数などのキーワードの概観のことである。キーワードのボリューム感を提示し比較することで、調べている言葉の Web 上での使用のされ方、意味といった概観の把握が容易になると考えた。

3.3 円を用いた検索ヒット数の視覚化

本研究では、キーワードのボリューム感の一つである Web 検索ヒット数を視覚的に表現する。キーワードの Web 検索ヒット数を円の面積に割り当て、視覚化する。円を用いるのは、一般的に集合の関係を表すことに使用される Venn 図を想起させると考えたからである。一般的な Venn 図を想起させることで、Web ページの集合の関係を柔軟に理解できると考えた。表で、 n 個のキーワードの間の AND, OR, NOT の関係を全て示そうとすると、 $2^{(2^n-1)} - 1$ 個の要素を表現しなくてはならない。それに対し、円を用いた表現では、一枚の絵ですべてを表現

することができる。図 3.1 は、2つの円の Venn 図である。この2円とその重なりから構成される関係の全て $A, B, A-B, B-A, A \cap B, A \cup B, (A \cup B) - (A \cap B)$ の7つの任意の領域に注目できる。

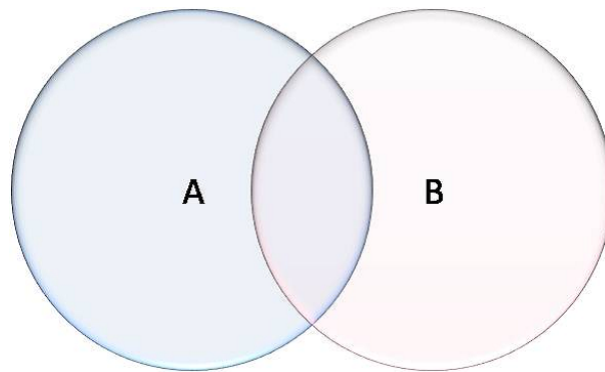


図 3.1: ベン図の例

第4章 Searchle:Webを用いた情報探索を支援するツール

本章では、本研究で開発した Web を用いた情報探索を支援するツール Searchle(Search+Circle)を紹介する。Searchle は、3 章で述べたアプローチに基づいて開発を行った。Searchle は、入力されたクエリの関連語とその検索ヒット数を視覚化して図を提示し、その図からのインタラクションによって情報探索を行うツールである。

4.1 Searchle の概観

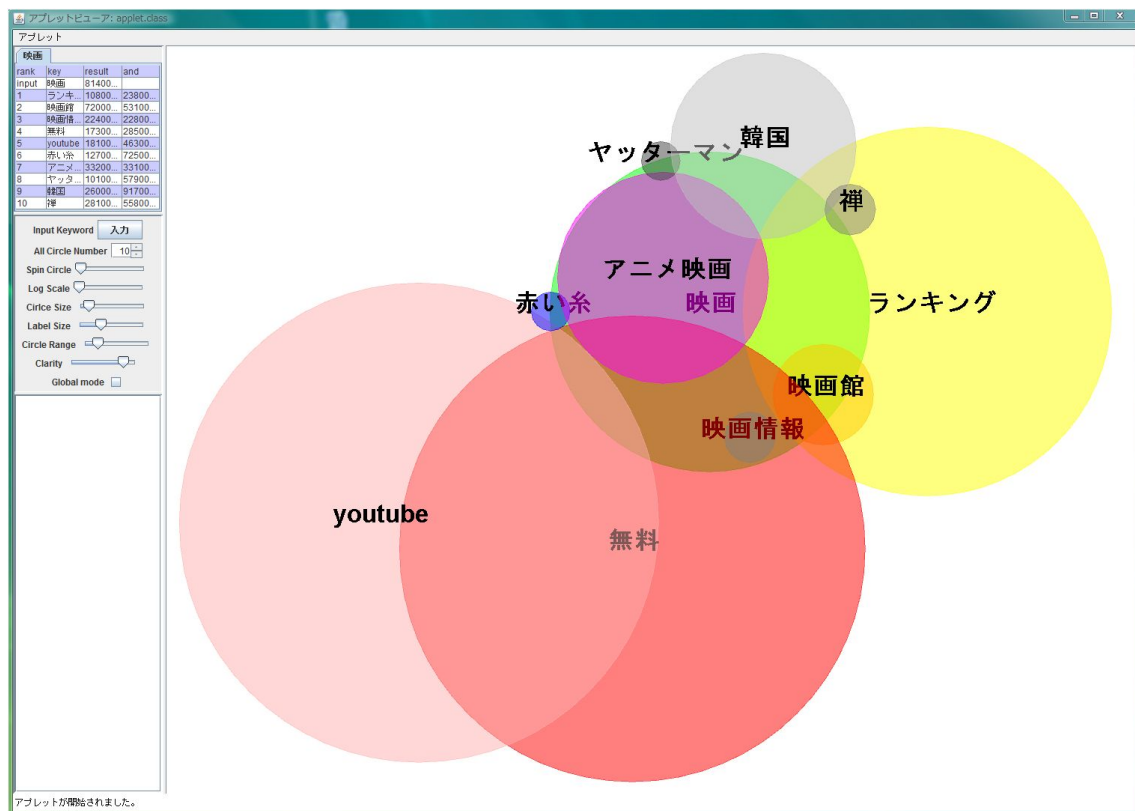


図 4.1: 開発したツールの概観

本ツールは、画面右側の「ビューパネル」、画面左側の「設定パネル」で構成される。「ビューパネル」では、円の描画、円の操作を行う。「設定パネル」では、描画される円のサイズ、モードなどの設定を行う。図 4.1 は、円が関連語を表し、円の大きさは検索ヒット数を表す。また、円の重なりは、AND 検索の検索ヒット数を表している。

4.2 Searcle の構成

4.2.1 円へのインタラクション

ここでは、円へのインタラクション法を紹介する。

ビュー内での再帰的な Web キーワード探索

本ツールでは、ビュー内での再帰的な Web キーワード探索機能として、以下の 3 つの機能を備えている。

- クエリ選択
- 図の移動
- ズーミング

これらの機能により、一つのビューパネル内での再帰的な Web キーワード探索を実現する。ビュー内での再帰的な検索は、本ツールを利用した Web 情報探索中に、入力したクエリの比較、履歴の閲覧、任意の情報への注目などを可能にする機能である。

クエリ選択 クエリ選択には、キーボードからの入力、円からの選択の二つの入力形式がある。入力と同じものならば、どちらで入力しても同じ結果が返ってくる。円からの選択についての詳細は、実装で説明する。

テキストでの入力 メニューパネルの上部にある入力ボタンを押すと、キーワード入力画面が出る。こちらでも円からの選択時と同様に、前に作成した円の右側に新しい円が描かれる。

表示されている円からの選択 描画されている円の任意の場所を右クリックすると、クエリ選択を行う。クエリを選択すると、選択した円のビュー内右側に選択した円を中心とする新しい円の集合が描画される。ツールを使用しているユーザが円を見て、興味をもった言葉について続けて探索を進めることができる。

円を選択した時のクリック位置とクエリの対応付けについて説明する。重なりのない円をクリックしたときには、そのままそのクエリが検索クエリとして選択される。円の重なりをクリックした時には、重なった全ての円に対応するキーワードを AND で結び検索クエリとし

ている。これは、関連検索ワードを用いるときに、キーワードを AND で結んだ検索では、取得できる関連検索ワードが極端に減ってしまい、関連検索ワードが見つからないということがあるためである。そこで、新しく円を作成するときのクエリと、Web 検索に対するクエリで、対応付けを変えている。新しい円の作成には、クリック位置と円の中心の距離が一番近いものをクエリとしている。

例えば、図 4.2 では、左側にある「映画」に対応する円の集合の中から、「赤い糸」というキーワードを選択し、「映画」に対応する円の右側に、新たに「赤い糸」に対応する円の集合が描画されたところである。

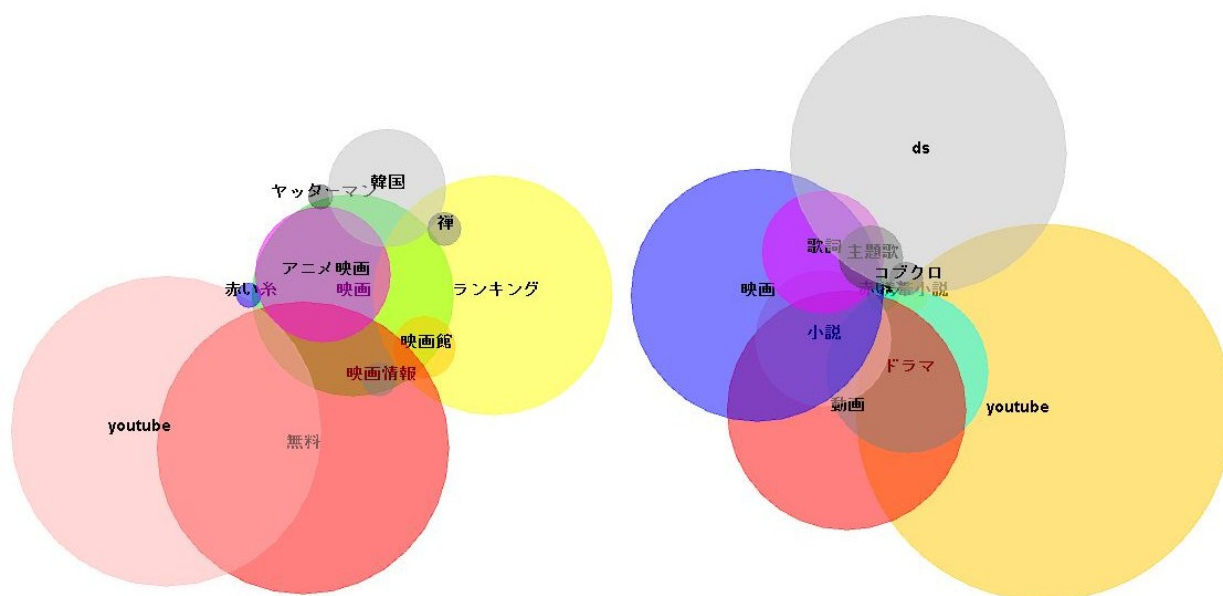


図 4.2: キーワードの選択例

図の移動

ビューパネル上で、マウスドラッグするとマウスの動きに合わせて、全ての円が移動する。複数の円を提示したときに、移動を行うことで、複数のクエリの比較を行うことができる。また、この機能により、X 軸に対して仮想的に無限の領域を確保することができる。これまでに自分の調べた履歴も閲覧することができる。また、これを用いて、見比べたい円に焦点を合わせることができる。

ズームイン

ビューパネル上で、マウスホイールを回転させると、ビュー内のズームインを行う。ズームアウトすると円全体の関係を見ることができる。複数の円の集合を見ることができ、クエ

りの比較がしやすい。ズームインすると、円の詳細な関係を見ることができる。ズームアウトした状態では、見えなかった詳細な情報を見ることができる。図 4.3 は、図 4.2 を作成した場面で、入力した「赤い糸」の検索ヒット数の関係を比較するために、ズームインしたところである。

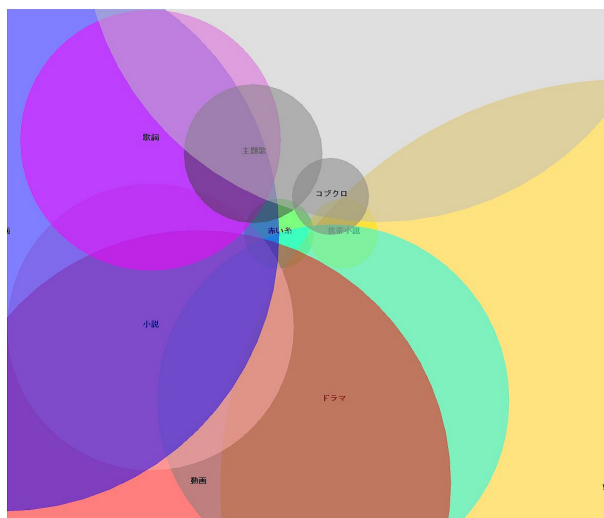


図 4.3: ズーミングの例

4.2.2 詳細情報の提示

詳細情報の提示機能として以下の 2 つの実装を行った。

- 表での Web 検索ヒット数の提示
- Web ページタイトルの提示

これらの機能により、Web 検索結果の比較、クエリに対する Web ページの概観の把握を支援する。

表での Web 検索ヒット数の提示

「設定パネル」の上部の表にて、詳細情報を提示する。詳細情報を見ることで、Web 検索ヒット数の明確な値を見ることができる。図 4.4 では、関連語の上位 10 件の検索ヒット数と入力したクエリと関連語の AND 検索ヒット数が表示されている。

映画			
rank	key	result	and
input	映画	814000000	
1	ランキング	1080000000	238000000
2	映画館	72000000	53100000
3	映画情報	22400000	22800000
4	無料	1730000000	285000000
5	youtube	1810000000	46300000
6	赤い糸	12700000	7250000
7	アニメ映画	332000000	331000000
8	ヤッターマン	10100000	5790000
9	韓国	260000000	91700000
10	禅	28100000	5580000

図 4.4: 詳細情報を示す表

Web ページタイトルの提示・閲覧

円の上でダブルクリックを行うと、選択したクエリに対する検索結果上位のタイトルを提示する。検索結果を見ることにより、選択したキーワードの概観をつかむことができる。また提示されたタイトルをダブルクリックすると、Web ブラウザが起動され、選択した Web ページを閲覧できる。Web ページを見ることで、さらに探索を進めることができる。図 4.5 は、ツールのメニュー領域の下部にある白いテキスト領域に、Web ページタイトルを表示しているところである。

映画 赤い糸
映画 赤い糸
1: 映画×ドラマ「赤い糸」公式サイト
2: 赤い糸 - フジテレビ
3: Yahoo映画 - 赤い糸
4: 赤い糸 - フジテレビ
5: 赤い糸 - goo 映画
6: 赤い糸 : 映画情報 - 映画のことなら eiga.com
7: 映画/赤い糸 映画作品情報 - cinemacafe.net
8: 赤い糸と THE KISS コラボレーション商品が登場!
9: 映画 赤い糸 : 象のロケット映画DVD総合ナビゲーター
10: 映画「赤い糸」特集 - TSUTAYA online

図 4.5: Web ページタイトルの提示例

4.2.3 円のレイアウト法の変更

以下のレイアウトに関する 4 つの要素の切り替えや調整方法を実装した。

- 表示する円の個数の切り替え
- 円の間隔の距離

- Log スケール
- 円の自動サイズ調整のモード切り替え

表示する円の個数の切り替え

一度に表示する円の数を変更する。円の個数を切り替えることで、概観をつかんだり、詳細を見ることができる。図 4.6 は、円の数を変更しているところである。

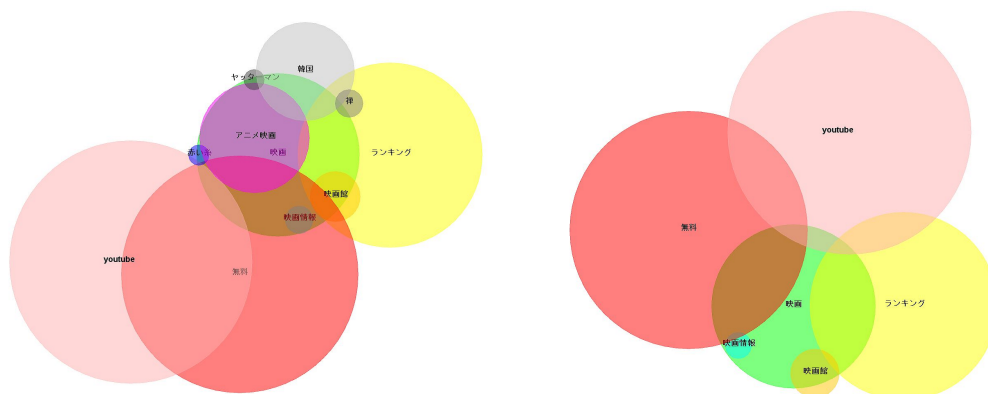


図 4.6: 円の個数切り替えの例

円の間距離

メニュー」内のスライダーを動かすことで、円の集合間の距離の設定をする。これを用いることで、注目している円の詳細を見ることができる。自分の好みの距離に合わせて円の集合を閲覧できる。

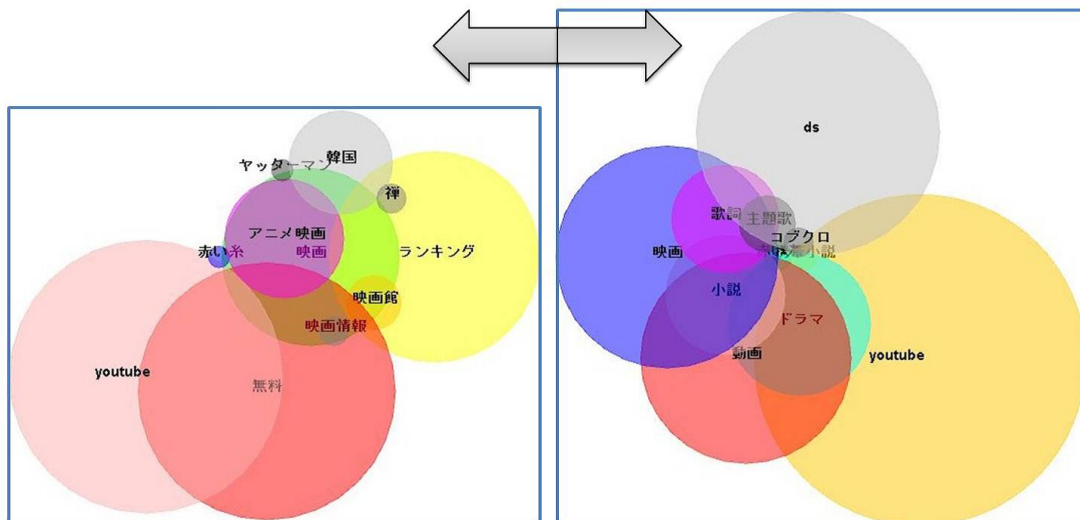


図 4.7: 円の集合の間の距離の変更

Log スケール

「メニュー」内のスライダーを動かすことで、検索ヒット数にスライダーの値を底とする Log の値を変化させる。検索ヒット数は、キーワードによって差があり、その差が大きい時には、円の面積がほぼ 0 になり、情報が読み取れない時がある。それを解消するために Log スケールを採用した。また、Log の値を変化させることで、これまでは見えなかった円の関係を見ることができる。図 4.8 は、「映画」に対応する円の集合を Log スケールで描画した図である。

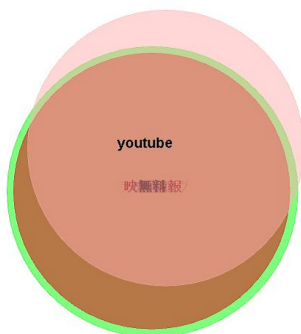


図 4.8: Log スケールでの描画例

円の自動サイズ調整のモード切り替え

ここでは、新しく描画される円の集合の大きさの自動調整の2つのモードについて説明する。モードの詳細については、後述する。

ローカルモード 一つのクエリで作成される円の集合の横幅を一定にし保持しながら描画するモードである。複数の検索結果を閲覧し、言葉の概観を知りたいという時に使用する。より目的に対する理解度が低い時に使用するモードである。

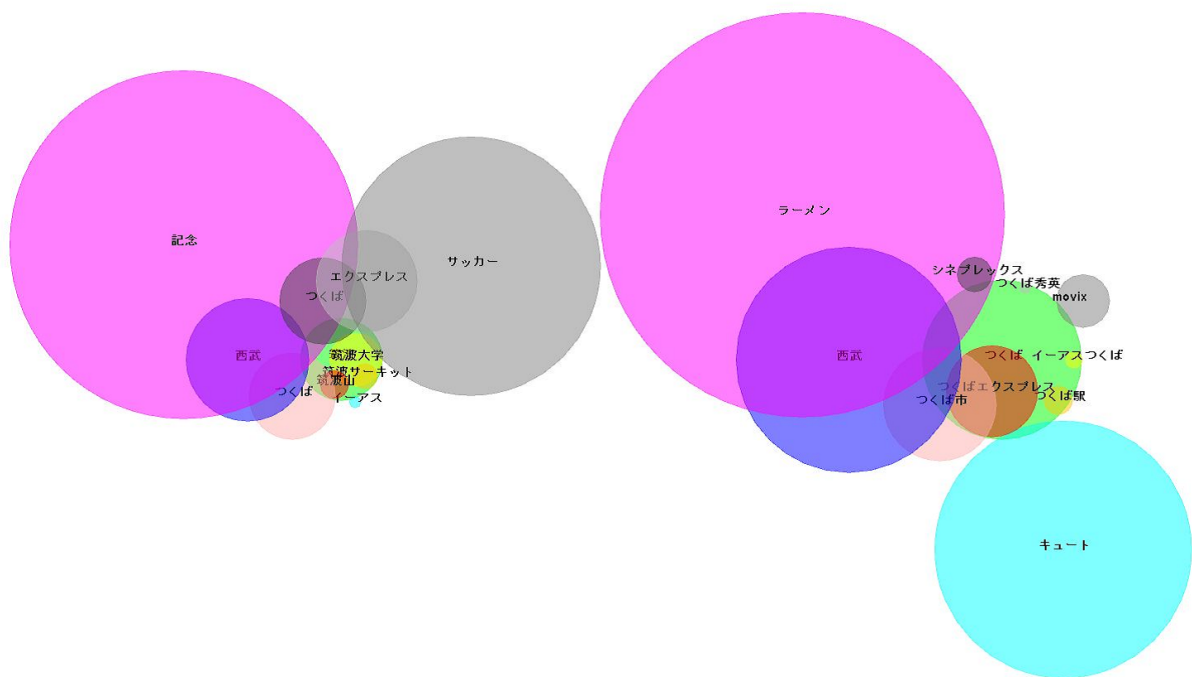


図 4.9: ローカルモードでの描画例

ローカルモードでは、水平方向に一定の長さに収まるように描画している。そのため、一つのクエリに対して作成される円に対して、検索ヒットと円サイズの比率に、ローカルな比率が用いられる。図 4.10 中の長さ l に保つように円サイズを計算する。一度により多くのキーワードを見ることができる。ローカルモードでは大きさが一定に保たれるため、新しく円が作成される時にビューの急な変更がなく、ユーザにストレスを感じさせないと考えた。

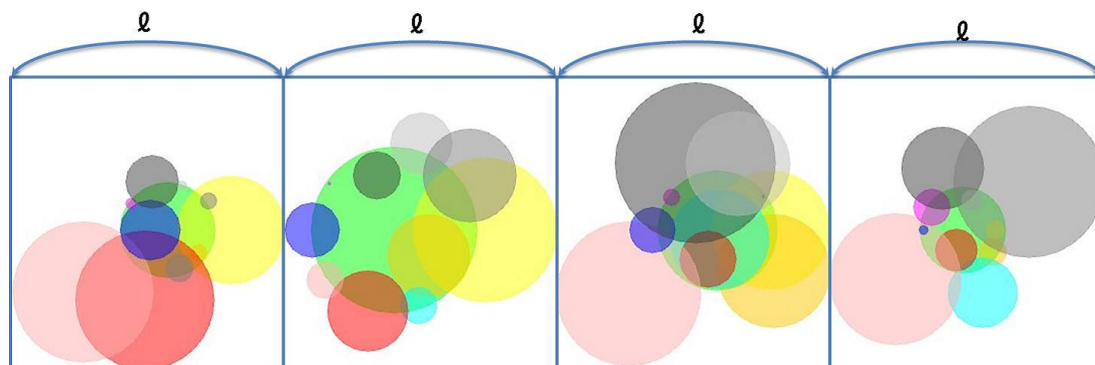


図 4.10: ローカルモードの説明

グローバルモード 描き出されるすべての円のサイズの比率を保ちながら、描画される。集合同士を比較するときには有効である。目的についてある程度知識を持っている時に、詳細に見比べるときに使用する。図 4.11 は、筑波とつくばの言葉の違いを調べようとしている所である。左側が筑波に対応する円の集合で、右側はつくばに対応する円の集合である。

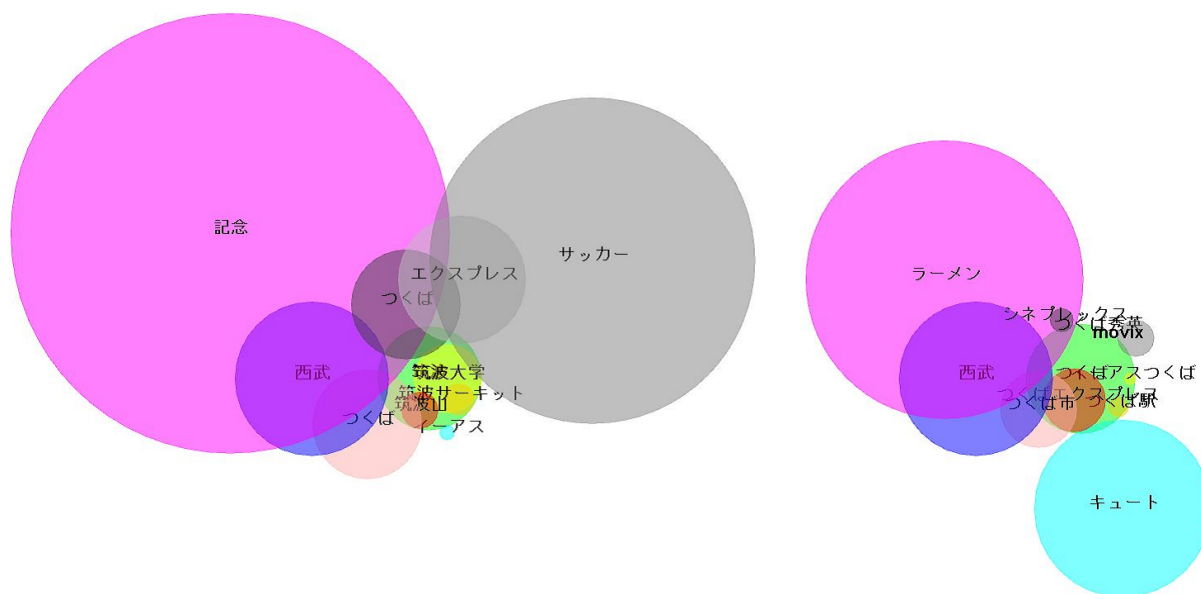


図 4.11: グローバルモードでの描画例

検索ヒット数と円サイズのグローバルな比率を用いて計算する。グローバルモードであれば、常に円サイズと検索ヒット数の比率が保たれるため、違うクエリから作成された円と円を比較することができる。また、一つのクエリに対する円全体を見た時に、一つのクエリの検索ヒット数とその関連語の検索ヒット数の大体の総和を把握できる。他に円が出ていない時は、ローカルモードを用いてサイズ計算が行われる。

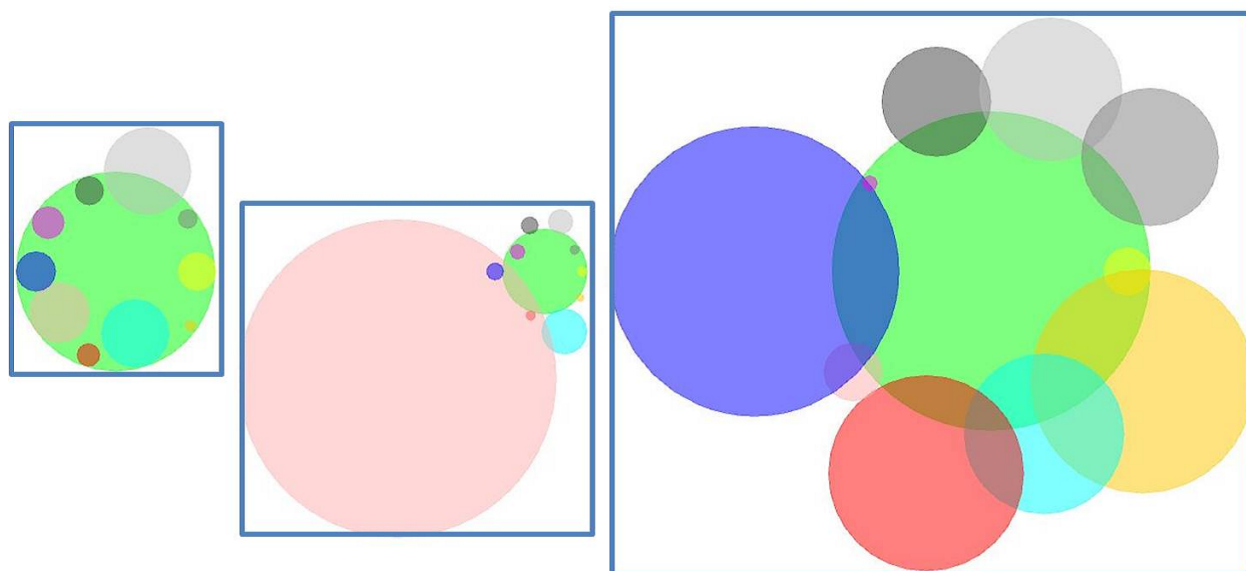


図 4.12: グローバルモードの説明

4.3 Web 情報の視覚化手法

入力したクエリに対する検索ヒット数を表す円と、その円の周辺に Yahoo 関連語検索の上位 10 件までの円を表示する。(デフォルトは 10 件) また、このとき周辺に表示される円は、中心の円(入力したクエリ)との関係を保つ。下の図(4.13)は、表 3.1 の情報を本手法で視覚的に表現した例である。中心にある緑の円が入力したクエリ(図中では”映画”)を示している。そして、Yahoo 関連語上位 10 件を、1 位のものを中心の緑の円の水平に右に配置し、そこを起点として時計回りで等間隔に、降順にキーワードを表す円を配置している。

表 4.1: 映画を入力クエリとした時の Yahoo 関連語上位 10 件と検索ヒット数

順位	キーワード	検索ヒット数 (page)	クエリとの AND 検索ヒット数 (page)
クエリ	映画	814000000	—
1	ランキング	1080000000	238000000
2	映画館	72000000	53100000
3	映画情報	22400000	22800000
4	無料	1730000000	285000000
5	youtube	1810000000	46300000
6	赤い糸	12700000	7250000
7	アニメ映画	332000000	331000000
8	ヤッターマン	10100000	5790000
9	韓国	260000000	91700000
10	禅	28100000	5580000

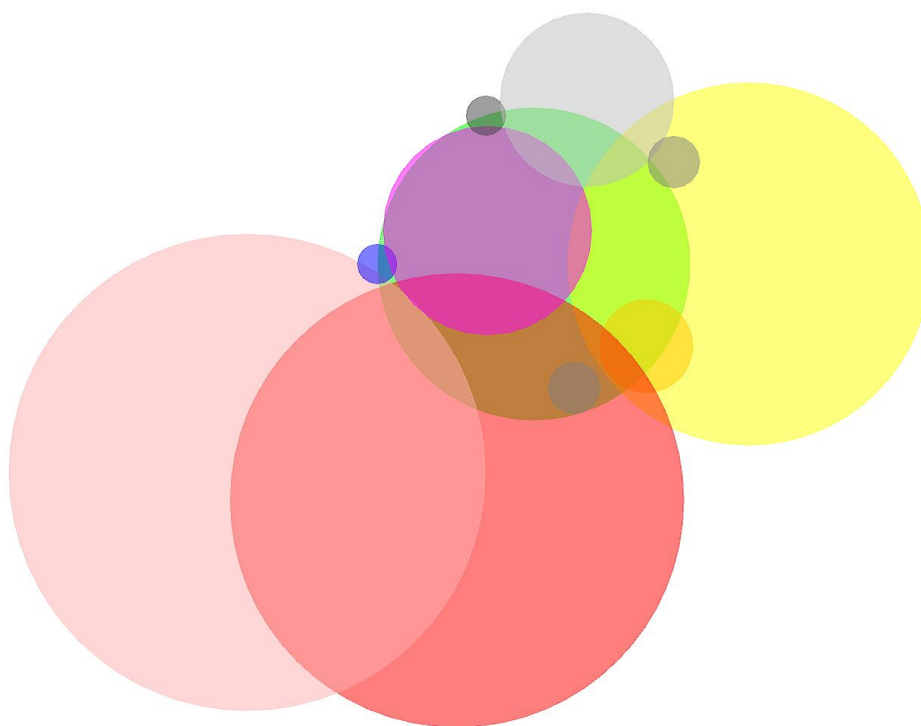


図 4.13: 映画をクエリとした時の描画例

4.3.1 面積から円の中心間の距離を求める計算法

ここでは、Web 検索ヒット数を円の面積とした時に、2つのクエリの検索ヒット数とその2つの And 検索のヒット数から円の配置を求める計算法について紹介する。

初めに、ここで使う文字を定義しておく。

S_1 := 1 の検索ヒット数, S_2 := 2 の検索ヒット数, S_3 := $1 \cap 2$ の検索ヒット数

d := 1,2 の円の中心を結ぶ線, h := 1,2 の円の交点から d への垂線

$a = \sqrt{S_1/\pi} = 1$ の半径, $b = \sqrt{S_2/\pi} = 2$ の半径

$a \geq b$

検索ヒット数から円の面積を求めるために、小さい円の中心が2円の交点を結ぶ線よりも外側にある場合の計算法を説明する。

$$if(S_1 \times \arcsin(b/a)/\pi + S_2/2 < b\sqrt{a^2 - b^2})$$

$$S_1 \times \arcsin(h/a)/\pi + S_2 \times \arcsin(h/b)/\pi - S_3 = h \times (\sqrt{a^2 - h^2} + \sqrt{b^2 - h^2}) \quad (4.1)$$

$$d = (\sqrt{a^2 - h^2} + \sqrt{b^2 - h^2}) \quad (4.2)$$

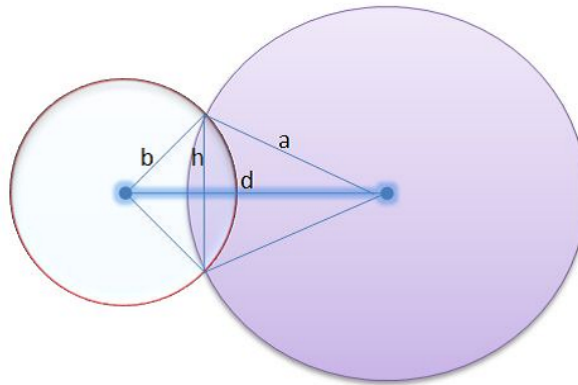


図 4.14: 円配置の計算法の説明 (1)

次に、小さい円の中心が2円の交点を結ぶ線よりも外側にある場合の計算法を説明する。

$$S_1 \times \arcsin(h/a)/\pi - S_2 \times (\arcsin(h/b) - 1)/\pi + S_3 = -h \times (\sqrt{a^2 - h^2} + \sqrt{b^2 - h^2}) \quad (4.3)$$

$$d = (\sqrt{a^2 - h^2} - \sqrt{b^2 - h^2}) \quad (4.4)$$

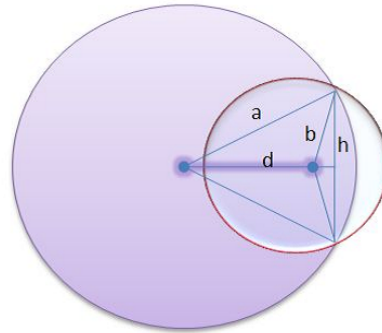


図 4.15: 円配置の計算法の説明 (2)

4.3.2 複数の円のレイアウト

ここでは、円のレイアウトについて説明する。先に求めた、入力したクエリと関連検索ワードを表す円の中心間の距離を保ちながら、入力したクエリに対する円の中心を軸に等間隔に配置している。また、水平方向右側を起点(1位)として、Yahoo 関連検索ワードでのランキング順に時計回りに配置していく。

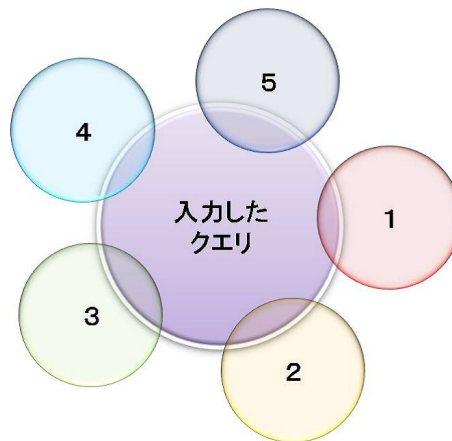


図 4.16: 円の配置順序

4.4 Searchle の利用シーン

ここでは、Searchle の利用シーンについて述べる。

4.4.1 辞書的な Web の利用

Web を用いて類義語の検索ヒット数を比べ、ヒット数がより多い言葉を正しい言葉として認識するという場面が見られる。本研究で開発したインターフェースでは、検索ヒット数の把握と比較が容易にできる。本ツールを用いて、入力した言葉が Web 上でどのように使われているのかを把握し、描画される図を見て、言葉の概観をつかむことができる。さらに、Searchle では複数の検索ヒット数を同時に把握することができるため、これまでには、より多くの言葉に対して比較を行うことができる。

4.4.2 曖昧な探索

曖昧な探索とは、目的や目的に対する適切なクエリが曖昧である時の探索ということである。本システムは、探したい Web ページやコンテンツがあるのだが、それを指す適切なクエリが思い当たらないという場面での利用シーンを考えている。キーワードに対して探索を行えるため、あまり詳しくない言葉に対して、本ツールを用い、その言葉の概観を把握し、適切なクエリを想起できる。

4.4.3 流行の把握

流行の把握とは、流行を知りたいという場面のことである。本ツールで採用している Yahoo 関連検索ワードは、Web 上で頻繁に使用されるキーワードが高順位であるという特性を持つ。そのため、入力したキーワードに関する流行しているキーワードを調べることができる。

第5章 実装

本章では、本システムの実装方法について説明する。Searchle は、Java6 を開発環境としており、Web 検索結果と関連語の取得に、YahooApi¹⁹を用いている。Web ページでの公開を考え、java アプレットとして実装した。

5.1 本システムの処理の流れ

本システムの全体の処理の流れについて説明する。まず、システムにクエリを入力することから始まる。入力されたクエリに対し、YahooApi から関連検索ワードと検索結果を取得する。取得した情報を円を用いて視覚化する。円が描画されてからは、視覚化した円を選択して、クエリとすることができる。図 5.1 が、全体のフロー図である。

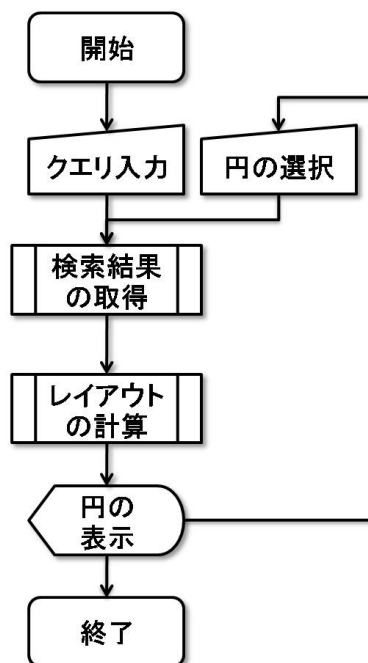


図 5.1: 全体のフロー図

5.1.1 YahooApi からの検索結果の取得部

本システムでは、YahooApi の関連検索ワードから情報を取得している。Yahoo 公式ページで、”[関連検索ワード]とは、Yahoo!検索の利用者が入力したキーワードや、その組み合わせを機械的に収集・処理した結果をもとに、検索キーワードの組み合わせなどを自動的に表示し、再検索を補助する機能です。”としている。Yahoo 関連検索ワードを用いるのは、クエリに対して関連度の高いキーワードを取得するためである。既存の Web 検索エンジンを用いることにより信頼性の高い情報を得ることができる。

本システムにクエリ入力が入力されると、入力したクエリに対する関連検索ワード上位 10 件と Web 検索結果を取得する。そして、上位 10 件の関連検索ワードに対する Web 検索ヒット数と AND 検索ヒット数と上位 10 件の Web ページタイトルを取得する。一度の試行で、計 22 回 YahooApi から情報を取得している。図 5.2 は、YahooApi にクエリを投げ、XML を取得、解析し検索結果と関連検索ワードを取得するフロー図である。

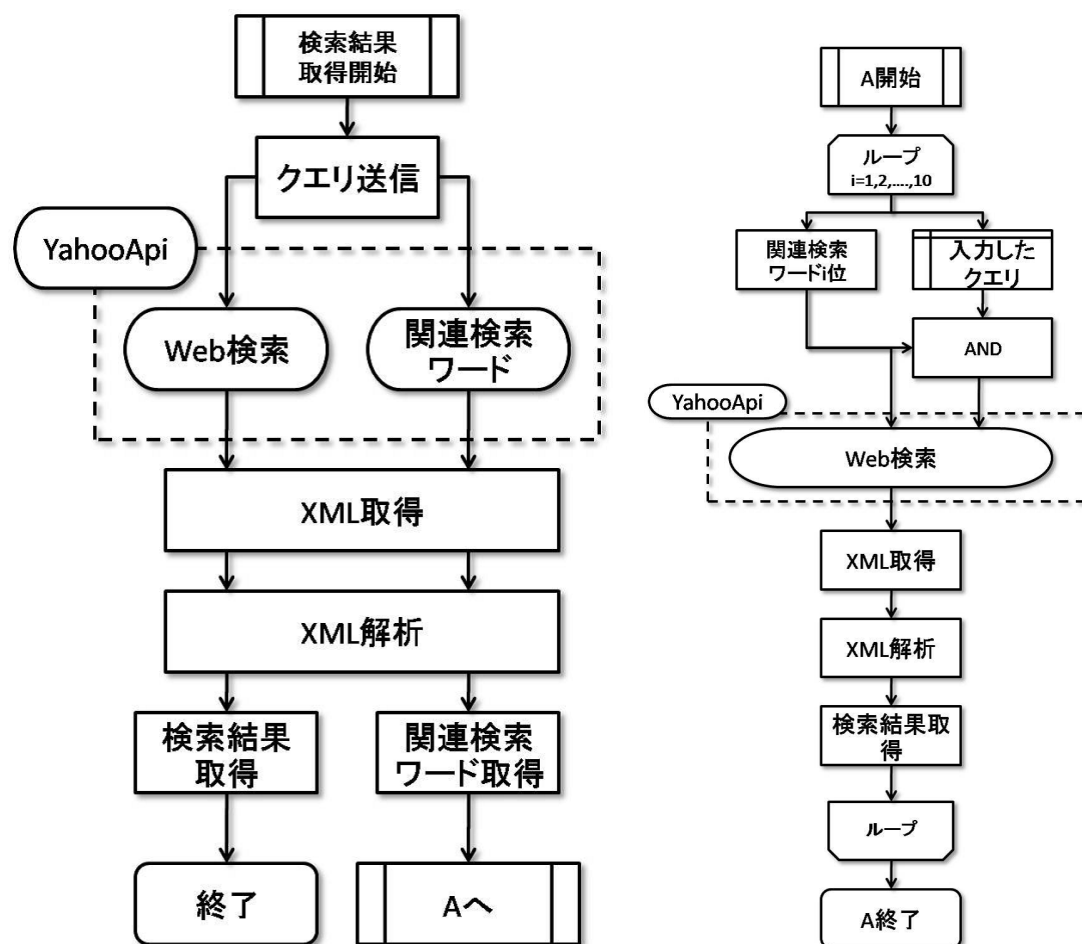


図 5.2: 関連語と検索結果を取得する際のフロー図

5.1.2 レイアウト

レイアウトは、先に YahooApi から取得した関連語とその検索ヒット数と入力したクエリと AND 検索ヒット数を元に計算を行う。また、グローバルモードの時だけ、ビューサイズの計算を行う。図 5.3 は、レイアウトの処理の流れである。

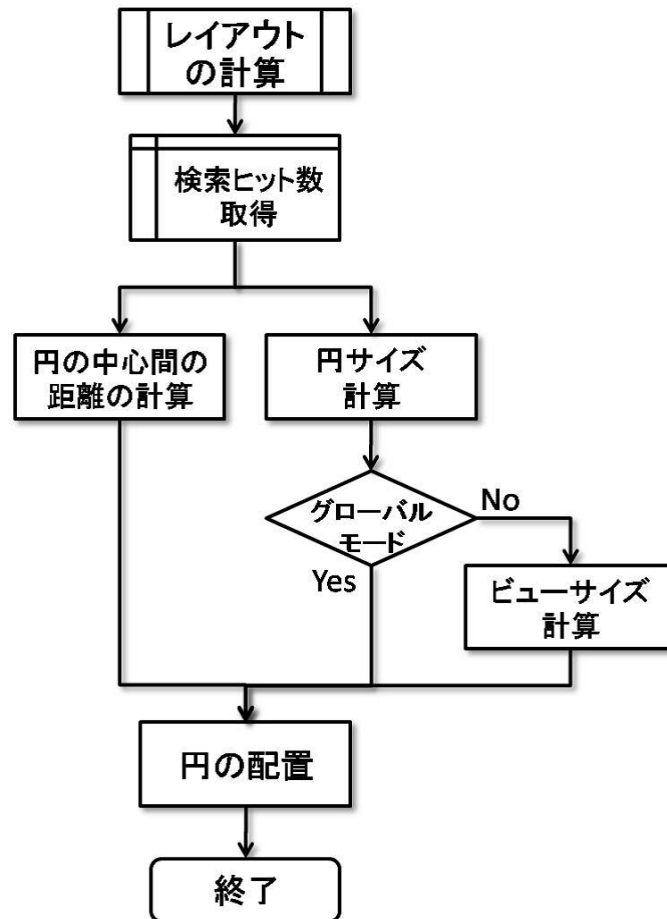


図 5.3: レイアウトのフロー図

第6章 評価実験

6.1 評価実験の目的

評価実験の目的は以下の通りである。

- 本ツールの Web での情報探索に対する有効性の評価
- 円サイズの自動調整の各モードの利用場面の調査と利用場面に対する各モードの有用性の評価

6.1.1 評価実験の狙い

初めに、「Web 理解度」をここで定義する。Web 理解度は、ある事柄に関する Web ページにはどのようなものがあるか、それらの Web ページに対する適切なクエリが分かるか、といったことを総合的に表す指標である。Web 理解度が高いとは、ある事柄について多くの Web ページを知っており、それらの Web ページに対する適切なクエリを作成できることを表す。評価実験では、Searchle の使用前と使用後の Web 理解度を計測し、本ツールが Web 理解度の上昇にどの程度寄与するかを調査した。

また調べる事柄は、各モードの利用場面の調査と利用場面に対する有用性を評価するために、検索結果に対する比較の必要性の違う利用場面を設定した。タスク 1 では比較の必要性が低い「流行を調べる」という利用場面を設定し、タスク 2 では比較の必要性が高い「言葉の違いを調べる」という利用場面を設定した。

6.2 評価実験の概要

被験者に、本ツールを用いて与えられたタスクをこなしてもらう。また、タスクを行いながら、解答用紙に記入を行ってもらう。そして、最後にタスクの達成度やツールの有用性について、アンケート用紙に記入してもらった。

6.2.1 被験者

コンピュータサイエンスを専攻する大学生と大学院生の 5 名を対象に評価実験を行った。彼らは、一般の人よりも Web を用いて情報探索をする機会が多い。そのため、本研究で扱う Web を用いた情報探索についてより多くの知識を有している。

6.2.2 評価実験の手順

まず、Searchle の目的と使い方を 10 分ほど説明し、説明の後に被験者の満足のゆくまで本ツールを使用してもらった。実際の Web 上の情報探索での利用シーンを模した以下のタスクを本ツールを用いて行ってもらった。

1. 流行している映画を探す
2. 「筑波」と「つくば」の利用され方の違いを調べる

という 2 つのタスクを行った。

タスク 1 の試験前に、映画に対する Web 理解度を尋ねた。タスク中は、流行している映画を見つけられたら解答用紙に書いてもらった。試験後に、「次に映画について探索を行う機会に、どのような検索クエリを作成するか」、「映画に対する Web 理解度」について質問した。

タスク 2 の試験前に、「筑波」と「つくば」にクエリとして違いがあると思うか尋ねた。タスク中は、違いを見つけたら解答用紙に書きとめてもらった。タスク終了後に、再び「筑波」と「つくば」にクエリとして違いがあると思うか尋ねた。

2 つのタスク終了後に、タスクの達成度とタスクに対するモードの有用性について尋ねた。

6.3 評価実験の結果

表は、各被験者 (A~E) の回答である。表の中にある数字は、以下の意味である。

(全般) 1:低い 2:少し低い 3:どちらでもない 4:少し高い 5:高い
(各モードの有用性)0:使用しなかった 1:低い 2:少し低い 3:どちらでもない 4:少し高い 5:高い
(筑波とつくばの違い) 1:ない 2:あまりない 3:どちらでもない 4:少しある 5:ある

表 6.1: タスク 1 の設問に対する各被験者の回答

評価項目	A	B	C	D	E	平均	0 を除いた平均
ツール使用前の「映画」の理解度	2	1	2	1	2	1.6	
ツール使用後の「映画」の理解度	3	3	4	3	3	3.2	
タスクの達成度	5	3	4	5	4	4.2	
Global モードの有用性	0	0	0	5	0	1	5
Local モードの有用性	4	0	3	0	0	1.4	3.5

表 6.2: タスク 2 の設問に対する各被験者の回答

評価項目	A	B	C	D	E	平均	0を除いた平均
ツール使用前の「つくば」と「筑波」の違い	4	2	4	2	4	3.2	
ツール使用後の「つくば」と「筑波」の違い	5	2	3	4	4	3.6	
タスクの達成度	5	5	4	3	5	4.4	
Global モードの有用性	4	5	3	5	0	3.4	4.25
Local モードの有用性	4	1	0	0	0	1	2.5

6.4 考察

6.4.1 ツールの有効性

タスク 1 で流行している映画を発見できない被験者はおらず、被験者全員が 2 つ以上の流行している映画を発見することができた。さらに、被験者があげた映画は未公開のものを除いて、いずれも 1/27 付の映画興行成績ランキング²⁰にて、Top10 にランクインしているものだった。このことから、本ツールを用いて流行を探ることができるということが分かる。

また、「本ツールの使用前と使用語の言葉の理解度」に対して t 検定を行ったところ、有意水準 1% で有意な差があることが示された。本ツールを使用することにより、検索した言葉に対する理解を深めることができることが分かった。理解を深めることで、検索したい事柄が明確になり、適切なクエリを想起しやすくなる。適切なクエリの構築により、目的の Web ページの発見に繋がる。これらのことから、Searchle が Web を用いた情報探索の支援に有効であることが分かる。

また、実験中に被験者は、Web ページを見ることをあまりしなかった。ほとんどの場面で、本ツールの特徴でもある円を用いた視覚的なインターフェースから情報を読み取っていた。

6.4.2 各モードの評価

「タスク 2 での各モードの有効性」、「各タスクに対するグローバルモードの有効性」について t 検定を行った。「タスク 2 での各モードの有効性」は、有意水準 5% で有意な差がある。「グローバルモードの各タスクに対する有効性」は、有意水準 5% で有意な差がある。これにより、比較の必要性が高い利用場面において、グローバルモードが有効であると示された。

6.4.3 今後の課題

モードの切り替え

今回モードの切り替えなかった被験者が見受けられた。これは、モード変更の煩わしさのためであると考えられる。今後は、評価実験で評価の良かったグローバルモードへの円サイズの自動調整法の統一やモード変更の自動化などの対策を行う。

複数クエリの対応

直接入力で複数のクエリを入力した時に、現在は Yahoo 関連語の仕様に検索語がヒットしないということが見受けられた。そこで、複数クエリの入力の場合は、入力されたクエリに対する円だけの表示、もしくは、それぞれ単一のクエリとして見なし円の集合を複数個作成することが求められる。

円へのインタラクションの拡張

現在、円を消したり、個別に動かすというインタラクションがなかった。円を見比べたりするときには、個別に動かしたり、円を消すことが必要である。また、円からのクエリ選択に関して、「円をマウスのクリックでつかんで、重ね合わせたりすることで、クエリを作成できたらおもしろい」との声があった。今後は、円へのインタラクションを拡張し、操作性を向上させる。

検索結果を取得するサービスの切り替え・拡張

現在は、検索結果を YahooApi からのみ取得しているが、今後は、Google などの検索結果も取得できるようにする。また、複数の検索エンジンから取得した結果に対して、本システムに合った検索結果を抽出するアルゴリズムを作成する。

第7章 結論

本研究では、Web 検索ヒット数を円の面積を用いて表現する視覚化手法、分散的な探索を実現したキーワードを辿るインターフェースを提案した。また、提案した手法とインターフェースを組み合わせた Searcle を開発し、それに対して評価実験を行った。

評価実験の結果から、Searcle がある事柄に関する Web 上の概観の把握を促し、目的に対する適切なクエリを想起させることを示した。また、検索ヒット数と円の面積の対応付けに関して、一定の比率を用いて計算するグローバルモードが、検索結果の比較に適していることを示した。検索結果の比較と Web の概観の把握は、Web を用いた情報探索の支援に有効であると示した。

謝辞

本研究を行うに当たり、指導教員である三末和男先生, 志築文太郎先生, 高橋伸先生, 田中二郎先生には、丁寧な指導と助言を頂き、心から感謝を申し上げます。また, インタラクティブプログラミング研究室の皆様には, ゼミでの発表や議論を通して貴重な意見を頂きました。特に NAIS チームの皆様には、研究内容、論文執筆などの研究の「いろは」をご教授いただきました。深く感謝しております。最後に、大学生活でお世話になった全ての方々に深く感謝しております。本当に有難うございました。

参考文献

- [1] 吉田大我, 小山聡, 中村聡史, 田中克己, Web 検索結果におけるキーワード出現相関の可視化と対話的な質問変換, DEWS2007 ,pp1-8,2007.
- [2] Alberto O. Mendelzon , Visualizing the World Wide Web, pp.13-19,1996.
- [3] 塩澤秀和, 西山晴彦, 松下温, 「納豆ビュー」の対話的な情報視覚化における位置づけ, 情報処理学会論文誌 Vol.38, pp. 2331-2342,1997.
- [4] 塩澤秀和, 西山晴彦, 相馬隆宏, 松下温, 情報の関連性と多人数アクセスに着目した WWW 空間の視覚化, 情報処理学会研究報告.[グループウェア] IPSJ SIG Notes Vol.96, pp. 61-66,1996.
- [5] 金西計英, 妻鳥貴彦, 矢野米雄, LOGEMON : Web 教材を使用した授業での教師支援システム : 学習者の閲覧履歴の視覚化による教師支援 (次世代教育 (学習) 支援システム論文特集), 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理 Vol.J83-D-I, pp. 658-670,2000.
- [6] 武藤哲幸, 新井克也, 情報探索におけるブラウジング作業を支援する WWW ブラウザの提案, マルチメディア通信と分散処理, pp.43-48,2006.
- [7] 種市淳子, 逸村裕, Web の探索行動と情報評価過程の分析, 2006 Library and Information Science. No.55, pp.1-23,2006.
- [8] 奥村穂高, ノードを用いた Web 検索インタフェースの研究, 修士論文, pp.1-33,2000.
- [9] 小林拓海, Web 検索結果の概観提示による情報収集支援システム, 卒業論文, pp1-29,2004.
- [10] 西原陽子, 赤井れい子, 砂山渡, 橘啓八郎, 積極的思考支援のためのキーワード選好インタフェース, 知能と情報 : 日本知能情報ファジィ学会誌 : journal of Japan Society for Fuzzy Theory and Intelligent Informatics Journal of Japan Society for Fuzzy Theory and Intelligent Informatics Vol.18, pp. 766-776,2006.
- [11] 三輪真木子, 情報検索のスキルー未知の問題をどう解くか, 中央公論新社, 2003/09/25.
- [12] 高久雅生, 寺井仁, 江草由佳, 齋藤ひとみ, 三輪真木子 神門典子, Web 情報探索における視線データの予備的分析, 情報知識学会誌 2008Vol.18,No2.pp181-188,2008.
- [13] 砂山渡, 大澤幸生, 谷内田正彦, ユーザの興味の構造を用いて関連検索キーを提示する検索支援インターフェイス, 人工知能学会誌 Vol.15, pp. 1117-1124,2000.

- [14] 戸川聡, 金西計英, 矢野米雄, Web 閲覧の個人特性に基づいた利用者動向可視化による管理者支援システムの構築, 情報処理学会研究報告. DSM [分散システム/インターネット運用技術] IPSJ SIG Notes Vol.2002, pp. 49-54, 2002.
- [15] 西原陽子, 砂山渡, 谷内田正彦, 有効な組合せの発見による創造活動支援 (人工知能, 認知科学), 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理 The transactions of the Institute of Electronics, Information and Communication Engineers. D-I Vol.J87-D-I, pp. 939-949, 2004.
- [16] 宮原秀友, 小林亜樹, 山岡克式 酒井善則, 複数評価値を用いた検索のための Web リンク空間の可視化, DEWS2005, pp1-8, 2005.
- [17] 池田新平, 中島伸介, 角谷和俊, 田中克己, Web 標識: 巡航履歴を反映した Web 情報空間の周辺案内, 電子情報通信学会技術研究報告. DE, データ工学 IEICE technical report. Data engineering Vol.102, pp. 1-6, 2002.

脚注

¹Google 社の提供する公式ブログ <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>

²<http://www.groxis.com/service/grok/>

³<http://clusty.com/>

⁴<http://www.like.com/>

⁵<http://www.xcavator.net/>

⁶<http://labs.systemone.at/retrievr/>

⁷<http://amaznode.fladdict.net/>

⁸<http://www.touchgraph.com/>

⁹<http://www.liveplasma.com/>

¹⁰<http://www.searchcrystal.com/>

¹¹<http://www.oskope.com/>

¹²<http://www.jlisting.jp/jdirectory/>

¹³<http://www.mooter.co.jp/>

¹⁴<http://www.quintura.com/>

¹⁵<http://www.kartoo.com/>

¹⁶<http://bkv.so-net.ne.jp/>

¹⁷<http://www.spacetime.com/>

¹⁸<http://www.searchme.com/>

¹⁹<http://search.yahooapis.jp/>

²⁰<http://movie.goo.ne.jp/ranking/boxoffice/>