

探索意図の洗練を重視した情報探索支援システムの提案

Proposal of an Explore Support System
Focusing on Refinement of Search Intention

山崎辰也[†] 大沼亮[†] 神長裕明[†] 宮寺庸造[‡] 中村勝一[†]
Tatsuya Yamazaki Ryo Onuma Hiroaki Kaminaga Youzou Miyadera Shoichi Nakamura

1. はじめに

近年、インターネットの普及やWebページの爆発的な増加に伴い、Web上での情報探索が一般的となっている。中でも、研究活動における情報探索のような模索的な情報探索は、重要かつ難しい作業である。このような情報探索では、通常、複数の手がかり(Key)を用いた探索や、過去の探索意図(Context)の理解・見直し、得られた探索成果(Result)の管理・まとめを相互に繰り返し行う。

これに対し、探索支援に関する様々な研究が行われている。例えば、キーワードの検索を支援するものとして、検索キーワードの推薦[1, 11]、ブックマークの管理支援として、拡張ブックマークに関する研究[2]等が報告されている。これらの研究では、殆どの場合、探索の手がかり・探索意図理解・探索成果管理に対する支援がそれぞれ個別に検討される段階に止まっており、十分な実効性を得るには至っていない。

そこで本研究では、探索方法・探索意図理解・探索成果管理の連携の取れた支援システムを提案する。このシステムによって、Webにおける情報探索作業の効率化を目指す。

2. 問題点と支援方針

2.1. 模索的な情報探索

本研究では、研究活動における情報探索のような、一つの目的をもって繰り返し行われる情報探索活動を支援対象とする。この種の情報探索では、ある発想をもとに、探索を繰り返していく。それぞれの探索で得られる成果が存在し、それぞれ蓄積されていく。次の探索は、それまで行ってきた探索によって得られた成果、および経験を踏まえて行われていく。模索的な情報探索は、このような一連の流れを何度も何度も繰り返していくものである。

Web探索の過程で、ページを発見閲覧する手段として、検索エンジンやハイパーリンク等が考えられる。探索者はキーワードや、Webドキュメントなどの手がかりを用いて、探索を繰り返していく。(探索の手がかり:Key)

思考錯誤を繰り返しながら多様な探索が行われる。その際、様々なKeyを用いて探索を行っていくが、そのKeyの組合せや使用する順序等に意図が含まれていると考えられる。(探索の意図:Context)

探索は、様々なKeyを用いて、様々なContextを持って行われる。探索者は自分が求めているものや、重要だと思うページに対して、ブックマークを行ったり、自分なりの注釈の付け加え(アノテーション)を行う。それらは探索の成果として残されるものである。また、成果は次の探索に生かされるべきKeyでもある。(探索の成果:Result)

模索的な情報探索では、上述した手掛かり、意図、成果がより多く存在する。そのため、それらをうまく管理、活用することが求められ、その中で情報探索の状況、探索意図の理解・見直しを行うことや、自分にとって有益となるドキュメントを発見すること、不要となったドキュメントのみきわめを行うことが重要である。

これら考察から、模索的な情報探索における問題点は、以下の通り整理できる。

(問題点1) Web探索に関わる、Key・Context・Resultを上手く一緒に管理・活用することが困難である。

(問題点2) 情報探索の状況、探索意図の理解・見直しを行うことが難しい。

(問題点3) 探索者自身にとって有益なドキュメントを発見することが難しい。

(問題点4) 不要となったドキュメントのみきわめを行うことが難しい

2.2. 関連研究

情報探索の支援として、Webの構造に着目した探索ナビゲーション支援[3]などがある。これは、ユーザが閲覧中のWebページのリンク構造上の周辺情報を視覚的に提示し、ナビゲーションを行うものである。この研究では、現在見ているWebページのみを対象として支援を行うため、そのページに至るまでのユーザの意図が反映されたものではない。

また、検索キーの管理・蓄積に着目した支援[1][4][7]が報告されている。これらは、過去の検索履歴などから、ユーザの特徴や検索キーワード同士の関連性を抽出し、検索時にユーザに提供することで、検索キーワードの選定を支援しようとするものである。これらの研究では、探索意図のキーとしての表現に対する支援としては有効だが、そこでの発見や検索結果の管理活用などについては考慮していない。

さらに、探索意図の見直しやコンテキストを反映した情報探索支援を目指した研究[6][8]がある。これらの研究は、探索活動を継続的な視点で捉え、探索意図の変化を考慮している点で興味深い。しかし、検索結果の管理活用は十分には考慮されていない。

[†]福島大学[‡]東京学芸大学

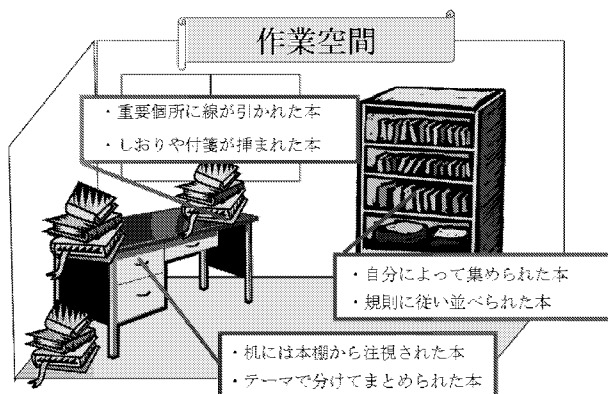


図1: 実世界での情報探索のしやすい作業空間

一方、探索結果の管理蓄積に対する支援として、ブックマークの拡張管理機構 [2] が提案されている。これは、ブックマークとその作成に至った状況などを共に保存することで、ブックマークの再利用時における状況の理解を容易にすることを指すものである。検索結果の管理と二次的な活用を考慮している点で興味深い取り組みだが、Webページの発見自体に対する支援は考慮していない。

このように、情報探索に関する支援は様々なものが存在しているが、情報探索における重要事項のいずれかのみを考慮したものである。そのため現状では、実際の模索的情報探索に対する支援効果は限定的である。本研究では、先述した問題点をトータルな視点から考慮し、実効性ある支援の実現を目指す。

2.3. 支援方針

本研究では、上記の4つの問題点に対し、以下の4つの手法を開発することにより解決を図る。

- Key・Context・Resultを一緒に扱うことができる「環境」の開発
- 探索意図の理解・見直しを可能とする空間の開発
- 有益となるドキュメントの推薦アルゴリズムの開発
- 不要なドキュメントの推薦アルゴリズムの開発

これら4つの手法を組み合わせた探索支援システムの開発を行い、Web探索におけるコンテキストの洗練を重視した支援を目指す。

3. コンテキストを重視した情報探索支援システム

探索の支援に関する研究は数多くある。しかし、それぞれ目的に応じて別個に行われているため、情報収集としての統括的な支援はなされていない。本研究では、Web上での情報探索を対象としている。図1は、Web上ではなく、実世界で本を利用した情報探索を行う場合を想定した作業空間のイメージである。本を

利用して自分の部屋で時間をかけて情報探索を行った場合、その場所には自分にしかわからない探索の形跡が残る。自分の理解しやすいように本を並べ、重要な箇所には線を引く、メモ書き等の作業を行う。そのような空間をWeb上での情報探索においても構成することで上述した問題点の解決を図る。

3.1. 設計

3.1.1. システム要件

本研究の目的を実現するためのシステム要件を以下の通りに整理する。

- 要件1：複数の手がかりを用いた探索ができること
- 要件2：探索の成果を作成できること
- 要件3：探索の流れを整理できること
- 要件4：探索における意図を組むことができること

3.1.2. 機能設計

以上の要件を満たすために次の機能を具備することを目指す。

- 複数の探索手段実行・管理機能
キーワードを用いた検索機能、ドキュメントを用いたリンク追跡機能、RSSによる最新情報の提示機能を有している。これら機能により様々な手がかりを用いた探索を可能とし、使用した履歴の管理も行う。この機能により、要件1を満たす。
- 探索成果の作成・管理機能
ブックマーク機能、アノテーション機能を有している。これらの機能により、探索の成果の作成を可能とする。この機能により、要件2を満たす。
- 探索経路把握機能
探索を繰り返すことで様々なKeyが生成される。そこには必ず流れが存在する。その流れの繋がりを提示する機能である。この機能によって要件3を満たす。
- 探索Key関連付け機能
Keyには探索者の意図した関連性があると考えられる。そのKeyの間の関連性を構成する機能である。この機能によって要件4を満たす。
- KeyContextResultの連携した探索空間提示機能
上述の機能によってシステム側は探索の流れ、探索者の意図を把握できる。本機能では、他の機能によって得られるものを考慮した探索空間の構築を行なう。

3.2. 実験システム概要

ユーザ自身が、検索において使用したキーワードや、残しておきたいWebページのブックマーク情報を蓄積し、それらをアイコンとして自由に配置・管理することを可能とする環境を構築し、そこでのユーザの振る舞いの分析を通じた有益ドキュメントの推薦につい

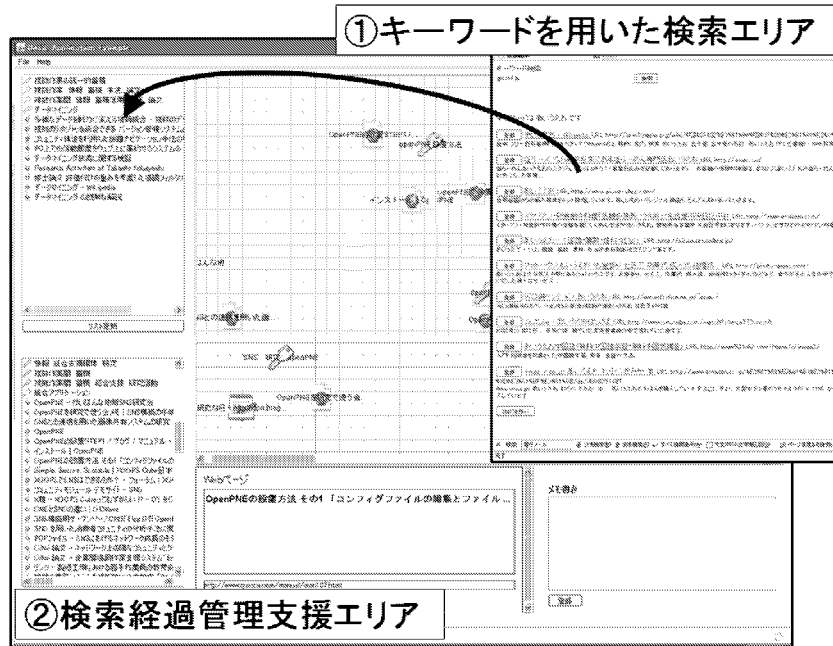


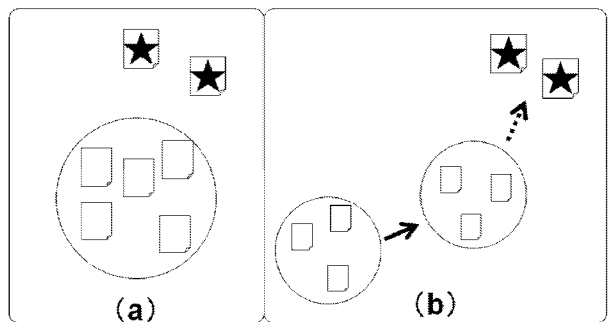
図 2: 実験システムのインターフェース

での検討を先行して実施するため、実験システムの開発を行った。

実験システムのインターフェースを図 2 に示す。本実験システムはキーワードを用いた検索エリア (図 2①)、および、検索経過管理支援エリア (図 2②) から構成される。検索結果の取得には Yahoo!Japan デベロッパーネットワーク [10] が提供するウェブ検索サービスを利用した。ユーザーはまず、キーワード検索エリアを利用して検索を行う。検索結果の提示方法は既存の検索エンジンサービスのものと相違ない形式で提供される。そこで見つけた Web ページはリンクの左端に配置されたボタンを押すことでデータベースに登録される。次に、検索経過管理支援エリア (図 3) に移る。ここでは、先のキーワード検索エリアによって登録されたデータを参照している。先ほど登録されたデータがこちらでリスト形式で維持されている (図 3①)。ユーザーはこのリストから保存された Web ページ情報、および検索キーワードを選び、横のアイコン配置マップ (図 3②) にドラッグすることで、情報がアイコン化され、配置することが可能となる。

3.3. 検索経過の解析による有益ドキュメントの推薦

探索ごとに探索者の意図が存在し、検索経過を配置することによって探索者の意図を表すクラスタが構成される。システムはその配置状況を分析することで、有益ドキュメントの推薦を行うことが可能であると考えられる。例えば、クラスタ内に存在するドキュメント群からそれぞれ特徴語を取り出し、既知のドキュメントを発見した際の検索キーワードとの比較を行うことで、探索者の気づいていないドキュメントの発見に繋がる。また、クラスタ同士の使用経過の連続性を見ることで、繋がりのある話題に対しての解析も行い、より選ばさ



(a) クラスタの内容と類似したドキュメントの推薦。
(b) クラスタの使用経過を考慮したドキュメントの推薦。

図 4: 検索経過の解析による有益ドキュメントの推薦

れた情報を得られることが期待できる。

4. 実験と考察

4.1. 実験概要

ユーザによる探索経過の配置状況に応じたクラスタリング、また、クラスタリングに基づいた有益ドキュメント推薦についての基本的有効性の確認、および、開発した実験システムの特徴、問題点の検証を目的としての評価実験を行った。実験協力者は、情報系の大学院生 1 名に「自身の研究に関する調査」としての情報探索と、その経過の管理を実験システムを用いて行ってもらった。具体的には、実験システムのキーワードを用いた検索エリアで情報探索を行っていただき、関

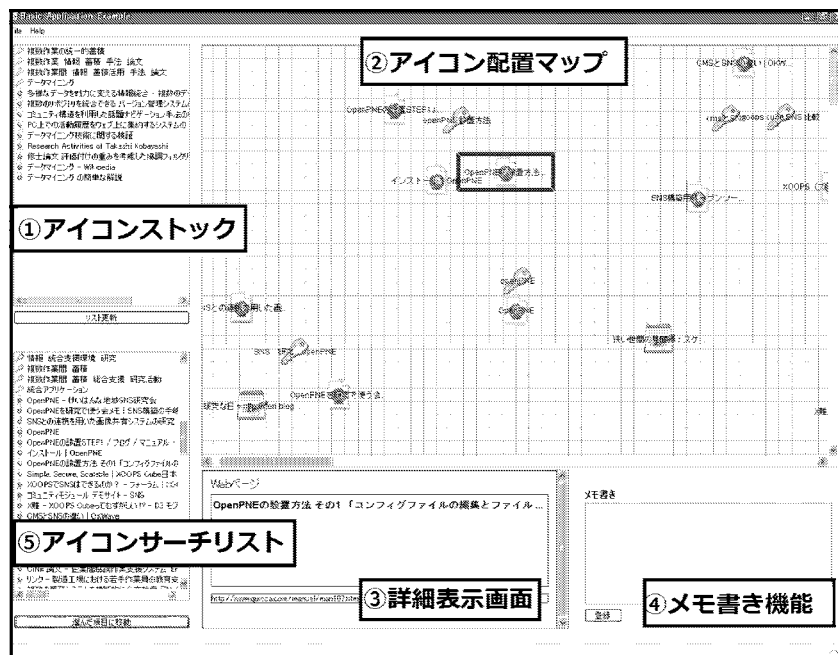


図 3: 探索経過管理支援エリア

覧した Web ページ情報や検索に使用したキーワードを検索経過管理エリアを利用して、配置・管理を行っていただいた。ここでは、目安として100件程度の Web ページの閲覧をお願いした。それによって、検索ごとに生じるキーワードのクラスターや、検索の流れの観察を行い、そのデータに対して上述したクラスターリングや、推薦の有効性を確認する。

【実験 1】

クラスターリングの有効性の確認の為、配置されたアイコン同士の距離に基づいたクラスターリングを行った。

【実験 2】

クラスターに基づいた有益ドキュメントの推薦の有効性の確認のため、ユーザの作成したクラスターに基づくドキュメント推薦を行った。具体的には、単一のクラスターに着目した推薦 (実験 2-A)、検索の順序をみたクラスターの遷移を考慮した推薦 (実験 2-B) をそれぞれ行った。

有益ドキュメントの推薦のイメージを図 4 に示す。

実際にシステムを利用した情報探索の後、協力者に、意図したクラスターを生成されたアイコン配置マップに記入していただいた。実験協力者が作成したクラスターを図 6 に示す。

ドキュメントの推薦はここで生成されたクラスターをもとに行う。今回の実験では、特徴語の抽出に "TermExtract" [9] を利用した。

また、実験 2 の分析の対象となるクラスターは、特に要素数の多い 3 つのクラスターを選択した。

選んだクラスターに対して特徴語の抽出・比較を行い、推薦するドキュメントを割り出し、それぞれが実験協力者にとって有益となるものかどうかを判定していただいた。

4.2. 実験結果と考察

表 1: 実験 1 における適合率, 再現率, F 値

適合率	再現率	F 値
0.91	0.94	0.93

表 2: 実験 2-A・2-B における適合率

実験 2-A	実験 2-B
0.53	0.43

実験 1 の結果を表 1 に、実験 2 の結果を表 2 に示す。

【実験 1】

適合率、再現率ともに良好な結果を得ることができた。クラスターリングの補佐として有効な結果を得ることができたが、これは協力者の配置が整ったものであったためであると考えられる。配置は人によって異なるものであり、距離だけでは推定できない場合も考えられる。例えば、一つの作業中に見つけたものをまとめて管理したい、などといった管理を行った場合、同じ話題が必ず近くに存在するとは限らない。そのため、今後も別の視点からクラスターリングを行う方法や、システムが行うクラスターリングの位置づけを検討していく必要があると考えている。

【実験 2-A】

適合率について概ね良好な結果を得ることができた。具体的には、今回対象とした 3 つのクラスターのうち、1 つについては高い適合率であるのに対し、他の 2 つのクラスターについては適合率が比較的低くなっている。実際の状況を精査してみると、適合率の低い結果となったクラスターから抽出された特徴語には、実験協力者が実際に使用したキーワードを部分的に含んでいるもの

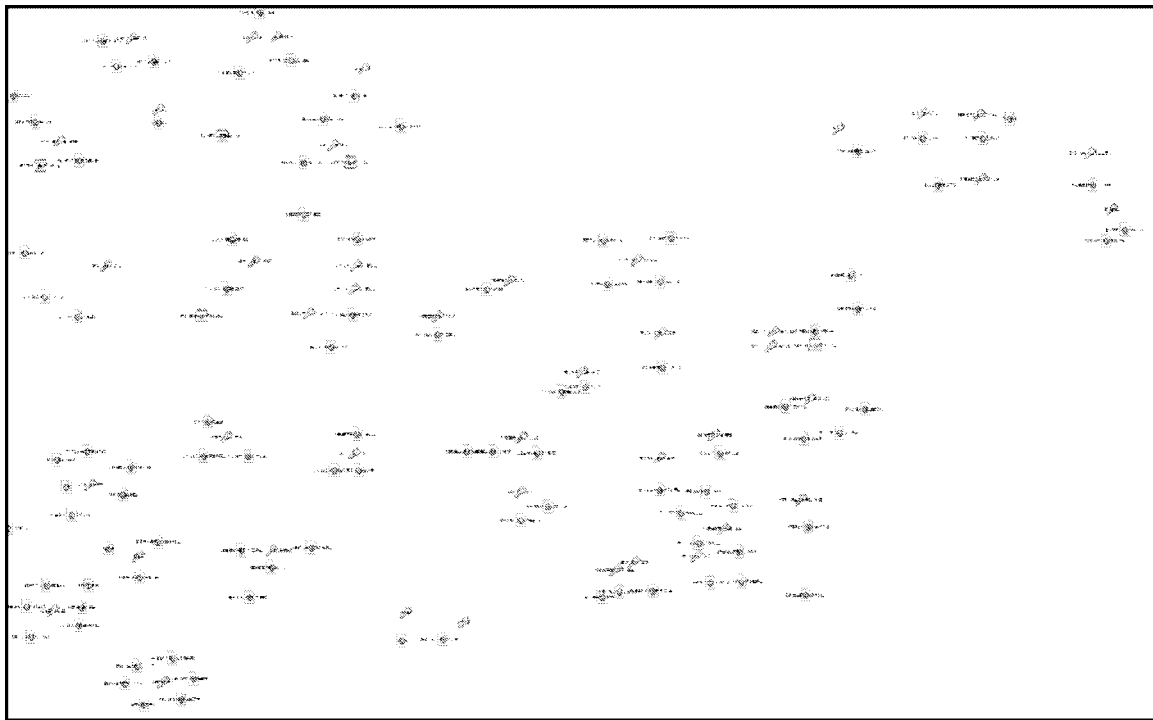


図 5: 実験協力者によって配置された探索経過

が多かった。そのため、推薦結果として与えられた情報に新規性がないものが多かったのではないかと考えている。クラスタ内に存在するページの中に、キーワードとして文が使われているものが存在しており、検索結果に大きな影響を与えている。そのため、特徴語からキーワードとして設定する条件を細かく設定することで、精度の向上が期待できるものと考えられる。

【実験 2-B】

適合率が、実験 A に比べるとやや低めの結果となった。しかし、より踏み込んだ推定であること、今回はシンプルな方法で推定を試みたことを合わせ考えると、十分今後に期待が持てる結果と考えられる。具体的な状況としては、今回対象にした 3 組のクラスタ対によって、適合度にかなり差が生じている。実際の状況を精査したところ、実験 A で考えられる原因と同じ現象が起きていることが見て取れる。また、別の原因として、今回の実験では、クラスタのつながりを、検索順序とクラスタ同士の距離のみから推定しているため、実験協力者が不意に異なる話題を検索し、前の検索と近い距離にアイコンの配置を行った場合、妥当な特徴語の抽出が行えなかったことが考えられる。今後、クラスタの連続性を定める手法について検討していきたい。例えば、蓄積された情報を見直す行為の参照情報や、アイコンの移動履歴などが手がかりとなり得るものと考えられる。

実験システムについて、「マップの全体像を把握しにくい」、「アイコンの情報がわかりづらく、なかなか次の探索に生かせない」「検索はしたけどページを残さない場合もキーワードは残して欲しい」などの意見を得る

ことができた。実験の様子を観察した結果からも、配置した後の支援にまだ不十分さが見受けられた。また、配置するマップについても、サイズに制限があることが、ユーザーの「これとこれは近くに置きたい」といった考えを制限してしまうことや、「ある程度埋まったからなるべく残さない」といった、残すことに対する制限を与えてしまうことが考えられた。これらは実験システムおよび目標とするコンテキストを重視した探索支援システムの改善を考える上で貴重な知見であり、今後の検討に活かしたい。

5. おわりに

本稿では探索における Key・Context・Result の連携を目指したシステムについて提案した。検索キーワードや、Web ページ情報の配置・管理の実験システムの開発し、構成されたクラスタの分析によるドキュメント推薦に関する評価実験を行った。その結果クラスタ単体の分析に基づくドキュメントの推薦の基本的な有効性について概ね良好な感触を得ることができた。また、クラスタの連続性の推定、実験システムのユーザビリティ等について、有益な知見を得ることができた。今後は、実際の情報探索におけるユーザのキーワード選定やアイコン配置に関する振る舞い等について分析を重ね、有益ドキュメントの推定手法、提示タイミングについて検討していきたい。

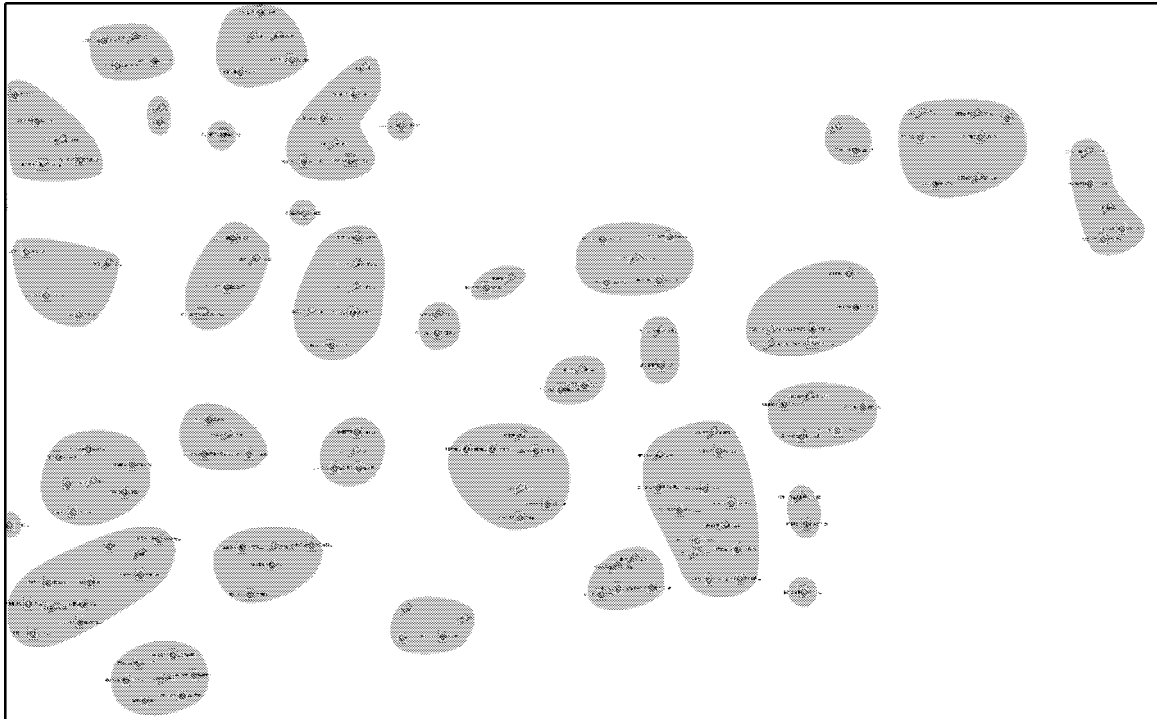


図 6: 実験協力者による探索結果のクラスタ概況

参考文献

- [1] 吉田大我, 小山聡, 中村聡史, 田中克己: “Web 探索結果におけるキーワード出現相関の可視化と対話的な質問変換”, 電子情報通信学会第 18 回データ工学 WS, 2007.
- [2] 日野洋一郎, 中島伸介, 小山聡, 田中克己: “Web Farm: 動物メタファを用いた Web ブックマーク再利用機構”, 電子情報通信学会第 14 回データ工学 WS, 2003.
- [3] 池田新平, 是津耕司, 小山聡, 田中克己: “Web コンテンツの周辺情報提示によるナビゲーション支援”, 日本データベース学会 letters Vol.2, No.1, 139-142, 2003.
- [4] 堀幸雄, 今井慈朗, 中山堯: “ユーザの Web 閲覧履歴を用いた検索支援システム”, 情報知識学会誌 Vol.17, No.2, 95-100, 2007.
- [5] 山田誠二, 村瀬文彦: “ページ情報エージェントの組織化による Web 検索エンジンの適応インターフェース”, 人工知能学会論文誌 Vol.16, No.1, 2001
- [6] 梶並知記, 高間康史: “Poker-Maker モデル: ユーザの検索意図を反映するキーワードマップと情報収集エージェントの連携による探索的情報検索”, 情報知識学会誌 Vol.20, No.3 2010
- [7] 奥村穂高, 田中二郎: “パーソナライズ表示画面を持つ Web 検索システムの構築”, 情報処理学会論文誌. プログラミング Vol.41, No.SIG-9, pp.113, 2000.
- [8] 奥健太, 中島伸介, 宮崎純, 植村 俊亮: “状況依存型ユーザ嗜好モデリングに基づく Context-Aware 情報推薦システム”, 情報処理学会論文誌. Vol.48, No.SIG-11, pp.162-176, 2007.
- [9] TermExtract: <http://gensen.dl.itc.u-tokyo.ac.jp/>
- [10] Yahoo!Japan デベロッパーネットワーク: <http://developer.yahoo.co.jp/>
- [11] Google サジェスト: <http://www.google.co.jp/>
- [12] Google アラート: <http://www.google.co.jp/alerts>