情報処理学会論文誌　データベース　Vol. 3　No. 2　123–137 (June 2010)

SearchLife：単語の特徴量を考慮した多視点クラスタリング検索エンジン

村松亮介[†1,*1]　福田直樹[†2]
横山昌平[†2]　石川博[†2]

本論文では Web ページ検索結果の閲覧性向上のための，検索結果のクラスタリングとラベリング手法を実現したシステムについて述べる．クラスタ閲覧の手がかりとなるクラスタラベルが単一の手法で生成されている場合，ラベル一覧にユーザの求めている語が存在しない場合，クラスタラベルが閲覧のための有用な手がかりとならず，目的の Web ページがクラスタ内に埋もれたまま見つからないことがある．本論文で示すシステムでは，閲覧性の向上のためにクラスタに対するラベリング手法を複数用意することで，ユーザが迅速に情報収集を行えるようにする．クラスタリングによる検索結果表示に対する提案手法の有効性を評価し，本システムを用いた場合に，いくつかの種類の情報収集タスクを効率的に行えることを示す．

SearchLife: A Multiple Viewpoints Clustering Search Engine Based on Document Features

Ryosuke Muramatsu,[†1,*1] Naoki Fukuta,[†2]
Shohei Yokoyama[†2] and Hiroshi Ishikawa[†2]

Web search engines play important roles as a tool to obtain useful information from the web. Since keyword-based search engines produce a long list of results that contain links to relevant web pages, approaches for improving visibility of results to grasp an overview of the search results has investigated. In this paper, we propose a system that clusterizes search results with multiple cluster labels for improved view of results. To make labels be a useful guide, we provide a multi way labeling method that considers statistical features of words in documents. We show that the usability of our labeling method for clustering search system has advantages in some complex searching tasks.

1. まえがき

検索エンジンは，Web 上の膨大な情報を検索する手段の 1 つとして，重要な役割を果たしている．代表的な検索エンジンである Google[1] や Yahoo![2] が提供している検索エンジンの検索結果は，ランキングに基づくリスト表示に基づいているため，検索対象やそれに対する検索クエリの与え方によっては，ユーザが検索結果の概観をとらえることや必要な情報をすぐに探し出すことが難しい場合がある．

検索エンジンにおける検索結果提示の改善方法として，クラスタリングに基づくアプローチがあり，その効果についての検証も試みられている[3)–15)]．しかしながら，クラスタリングに基づく検索結果の提示手法には，クラスタリングを高精度に行うということ以外にも，クラスタへの適切なラベル付けや，目的に応じて異なったラベル付け手法を用い，それらの表示を選択可能とすることで，改善が行える余地がある．

同一の検索クエリによって検索を行う場合であっても，それぞれのユーザの検索意図，目的，および検索対象に関する理解度は多様である．本論文では，これらのユーザの検索意図や検索対象に対する理解度に対して適切となるような検索結果の提示のされ方を，1 つの「視点」として定義する．クラスタリング結果から複数の Web ページを収集したり，多義性のある検索クエリの持つ複数の意味を分析したりするような，クラスタリング結果を精査する必要があるタスクの場合，クラスタリングを別の視点で何度もやり直すと，クラスタの構造が変化することにより，どこまでが調べた内容なのかが分からなくなってしまい，効率的な精査ができない場合がある．そこで，本論文ではクラスタの構造は保存したまま，クラスタのラベルを切り替えることを考える．クラスタのラベルを複数の視点からそれぞれ有益であろうと考えられるものに切り替えながらクラスタリング結果をブラウズできるようにすることで，検索結果の精査が円滑に進められるようにする．

本論文では，閲覧性の向上のためにクラスタに対するラベリング手法を複数用意する．具体的には，Web 全体集合と，あるクエリに対する検索結果集合という異なる 2 種の集合に対する，それぞれの単語の特徴量の違いを考慮したラベリング手法を用意することで，検索

†1 静岡大学大学院情報学研究科
Graduate School of Informatics, Shizuoka University
†2 静岡大学情報学部情報科学科
Department of Computer Science, Faculty of Informatics, Shizuoka University
∗1 現在，株式会社エヌ・ティ・ティ・データ
Presently with NTT DATA Corporation

結果中の単語の専門性，一般的認知度の違いを反映させるような，複数のラベルを同一のクラスタリング結果に対して提示できるようにする．本論文では 4 種類のラベリング手法を用いる．手法 1 と手法 2 で生成されるラベルはクラスタ内文書のいくつかに共通して出現する単語である．前者は後者と比較して Web 上で使用頻度が高く，後者は前者と比較して Web 上で使用頻度が低い単語であるという特徴を持つ．手法 3 と手法 4 で生成されるラベルは統計的手法によって各文書から抽出される単語である．手法 3 で生成されるラベルは手法 4 のそれと比較して Web 全体集合では使用頻度は低いが，検索クエリに対する検索結果集合中では使用頻度が高くなる単語である．手法 4 で生成されるラベルは手法 3 のそれと比較して Web 全体集合では使用頻度は高いが，検索結果集合中では使用頻度が低くなる単語である．本論文では上記 4 種のラベリング手法を導入することで，ユーザが検索結果クラスタを複数の視点で閲覧可能とする多視点ラベルを構築する．

提案システムの評価に関して，Broder [16] が提案するユーザの検索意図に基づく検索クエリの分類モデルを本研究で提案する検索システム SearchLife に適用し，ユーザビリティ評価実験により本システムの有効性を確認する．

本論文の構成は，次のようになっている．2 章で，本研究と関連研究との差分を述べ，3 章で，本論文で提案する検索結果のクラスタリング手法および各種ラベルの生成方法を述べる．4 章で，本論文で提案する多視点ラベルに関するユーザビリティ評価を述べ，5 章で，その他の評価に関する議論を述べ，6 章で，本論文で得られた成果をまとめる．

2. 関 連 研 究

2.1 検索結果のクラスタリング

Web 検索結果のクラスタリングに関する研究は大きく 2 つに分類できる．1 つは，Web ページの内容に着目してクラスタリングを行うコンテンツマイニングであり，もう 1 つは，Web ページのリンク情報に基づいてクラスタリングを行うストラクチャマイニングである．コンテンツマイニングを行う研究として，たとえば Zamir ら[6] や成田ら[8] の研究がある．Zamir らの手法は Suffix Tree と呼ばれるデータ構造から共通して現れる単語を容易に発見し，その単語が出現する検索結果をまとめてクラスタを形成する手法であり，STC と呼ばれる．成田らは検索結果の階層型排他的クラスタリングを行うシステムとして METAL を開発した．成田らの研究では，クラスタリングの性能に関しては検討はされているが，生成されたクラスタに対するラベルの有用度に関して踏み込んだ検討はなされていない．

ストラクチャマイニングを行う研究として Imafuji ら[17] や大野ら[7] の研究がある．Imafuji らは Web の構造分析を目的として Web 全体を対象としてマイニングを行っており，大野らは検索結果内からの効率的な情報収集を目的として検索結果を対象としたマイニングを行っている．大野らの研究では高い再現率および適合率を示しているが，類似度判定手法の特性からクラスタに分類されないページが多くなる．この問題に対してはコンテンツマイニングによるクラスタリング手法との統合による改善の可能性が指摘されている．

また，現在 Web 上に公開されているクラスタリング検索エンジンとして Clusty [5] や Carrot2 [14] がある．Clusty はメタ検索エンジンの一種で，検索結果を階層的にクラスタリングして，画面左にクラスタをツリー型メニューとして表示し，画面右に選択したクラスタに属する Web ページがリスト表示される．Clusty は “Velocity” と呼ばれる独自クラスタリングエンジンを利用しており，文書を意味のあるグループに自動組織化する．また，2008 年 1 月より，新機能として remix 機能が追加された．remix とは提示されたクラスタリング結果がユーザの意図と一致しなかった場合，つまり，ラベル一覧に求めている情報が存在しなかった場合に，別の観点で再クラスタリングを行う機能である．remix はユーザにとって適切なクラスタが生成されるまで繰り返される．クラスタリング結果から複数の Web ページを収集したり，多義性のある検索クエリの持つ複数の意味を分析したりするような，クラスタリング結果を精査する必要があるタスクの場合，remix のようにクラスタリングを何度もやり直すと，そのたびにクラスタの構造が大きく変化することにより，すでに精査した内容がほとんどを占めるようなクラスタについても，その内容を改めて精査する必要が出てくるため，効率的な閲覧ができない可能性がある．本論文では，クラスタ構成を一定のままとするため，1 つの要素を複数のクラスタに所属可能とする非排他的クラスタリングを用いることで，複数の観点からのクラスタリング結果を 1 つのクラスタリング結果に統合して示すことを可能とし，再クラスタリングを不要とする．また，ユーザにとって有益なラベルが提示されないという問題への対処として，生成クラスタに対して性質の異なるラベリング手法を複数用意することで，ユーザが検索結果から目的とする Web ページ発見の支援を可能とする．

Carrot2 はオープンソースのクラスタリング検索エンジンであり，検索結果を意味を持ったカテゴリに自動組織化する．また，Carrot2 はクラスタリングアルゴリズムとして Lingo [13] および STC [6] を搭載している．Lingo は，まず入力文書からクラスタラベルを生成し，その後，各ラベルに文書を割り当てる手法である．

2.2 Web ページに対する複数スニペットの提示

検索エンジンから返される検索結果の Web ページそれぞれの要約文章（スニペット）を，複数の視点から構築しユーザに提示する研究として，Ahn ら[18] や高見ら[19] の研究がある．

Ahn らは，Web ページに対するパーソナライズ化されたスニペットをユーザに提示する手法を提案している．Ahn らの手法では，ユーザの現在直面している課題からタスクモデルを構築し，そのモデルに基づいて，検索結果に表示される各 Web ページのスニペットを構築する．タスクの干渉度を考慮して，3 種のスニペットを構築し，ユーザが状況に応じてスニペットを選択できる．高見らは，スニペットを，その生成方法により 2 種類の軸で分類している．さらに，Web ページに対して 4 種類のスニペットを生成できるようにすることで，ユーザの検索目的に適したスニペットを提示できるようにした．これら 2 つの研究は，1 つの Web ページに対する多角的なスニペットの提示を目指したものである．本論文では，検索結果クラスタという，ある類似性を持った Web ページ群に対するラベルを，複数の視点から構築できるようにすることを目的とする．

3. 多視点ラベルの構築

本章では多視点ラベルの構築手法に関して述べる．以下に本論文で提案する 4 種類のラベルの特徴を記す．

ラベルタイプ 1

クラスタ内文書のいくつかに共通して出現し，1 つの名詞からなる．ラベルタイプ 1 はラベルタイプ 2 と比較して Web 上で使用頻度が高い単語を用いる．ラベルタイプ 1 は検索対象においてよく使われる単語の把握やそれらの単語を手がかりとして必要とする文書を迅速に発見できることが期待される．

ラベルタイプ 2

クラスタ内文書のいくつかに共通して出現し，複数の名詞により構成される．ラベルタイプ 2 はラベルタイプ 1 と比較して Web 上で使用頻度が低い単語を用いる．ラベルタイプ 2 はラベルタイプ 1 と比較して，Web 上で使われない単語であるため，検索結果が抽象度の低い詳細な内容を示す単語を多く含む場合にはラベルタイプ 2 を見ることでより迅速に必要となる文書を発見できることが期待される．

ラベルタイプ 3

ラベルタイプ 1 とラベルタイプ 2 で用いられる単語はクラスタ内の文書どうしで共起する単語つまり，クラスタを概観できる可能性がある単語を用いる．一方，ラベルタイプ 3 では共起関係を考慮しない単語を用いている．ラベルタイプ 3 で用いられる単語はラベルタイプ 4 と比較して Web 全体集合では使用頻度は低いが，検索クエリに対する検索結果集合中では使用頻度が高くなる単語である．たとえば，ある分野に詳しいユーザなら常識的に知っているが，その分野に関して事前に触れたことのないユーザには理解の難しい語句を含む内容が検索結果に含まれることがある．そのような場合に，検索クエリ集合においてある種「常識的」な単語であると期待されるラベルタイプ 3 を閲覧することで，検索クエリの結果の閲覧に関してあらかじめ知っておくべき事項をユーザが把握するのに役立つ可能性がある．

ラベルタイプ 4

ラベルタイプ 4 では，ラベルタイプ 3 と同様に共起関係は考慮しない単語を用いている．ラベルタイプ 4 は，ラベルタイプ 3 と比較して Web 全体集合では使用頻度は高いが，検索結果集合中では使用頻度が低くなる単語である．同じような文書が多く検索結果に含まれるような状況で特徴的な文書を含むクラスタをすばやく探し出す場合や他のクラスタとの差異を際立たせることに利用できる可能性がある．

上記 4 タイプのラベル生成手法を用いることにより，ユーザは，各自の検索要求や検索対象に関連する事項の理解の程度に応じて，閲覧するラベルタイプを選択することで，より早く目的の Web ページを発見でき，閲覧性を向上させることができるのではないかと考えられる．

3.1 ラベルタイプ 1 およびラベルタイプ 2 の生成

本節ではラベルタイプ 1 およびラベルタイプ 2 の生成手法について，その概要を示す．なお，ラベルタイプ 1 およびラベルタイプ 2 の生成手法は本システムにおけるクラスタリング過程と密接に関わる部分があるため，あわせて本節では文献 9) で述べた本システムでのクラスタリング手法についても述べる．本システムの概要を**図 1** に示す．

3.1.1 検索結果の取得および形態素解析

本システムでは最初に，Yahoo! デベロッパーネットワーク[20] が提供するウェブ検索 Web サービスを利用して検索結果上位 100 件のタイトル，テキストサマリ，URL を取得する．次に，同デベロッパーネットワークが提供する日本語形態素解析 Web サービスを利用して，上記で取得した検索結果 100 件のタイトル，テキストサマリ，URL の形態素解析を行い，名詞のみを抽出する．ここで，本サービスを用いてたとえば人名 ‘村松亮介’ を形態素解析した場合，‘村松’，‘亮介’ のように 2 つの名詞として抽出されてしまう．そこで 2 回以上連続して名詞が出現した場合には，先頭から 2 つずつを 1 つの単位として，その 2 つの名詞を連結して 1 つの名詞としたものを順にテーブル 1 に保存する[⋆1]．また，サービスからの

⋆1 たとえば，5 回連続して名詞が出現した場合には，1 番目と 2 番目，3 番目と 4 番目をそれぞれ 1 つの名詞としてテーブル 1 に保存し，その後 5 番目の名詞を単体でテーブル 1 に保存する．これは，単語の共起のとりやすさと単語の抽出精度とのバランスをとるためである．

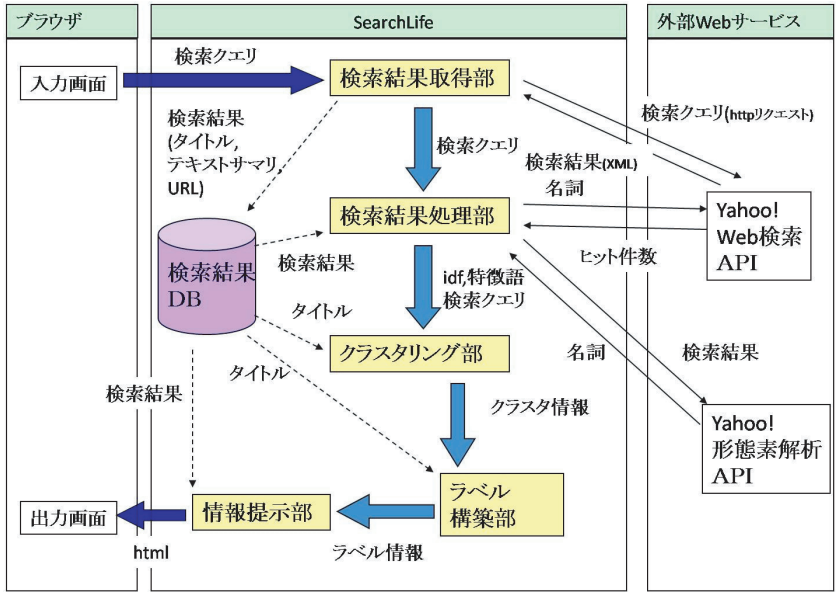

図 1　システム概要
Fig. 1　Overview of the system.

そのままの返却結果をテーブル 2 として保存する．

3.1.2　*idf* 算出およびラベルタイプ 2 の抽出

各文書においてその特徴を表すと思われる単語を抽出するためタイトルに出現する名詞の *idf* 値を算出する．単語 t が出現する文書数を $dt(t)$ とし，N を比較文書数とすると，式 (1) のように表すことができる．

$$idf(t) = \log \frac{N}{dt(t)} \tag{1}$$

ここで比較文書数 N は 548 億[⋆1]に設定する．また，$dt(t)$ はウェブ検索 Web サービスを利用したときの検索クエリ t に対する検索結果ヒット件数とする．上記式 (1) によって形態素解析結果であるテーブル 1 に保存される全名詞の *idf* を求め，各タイトルにおいて以下の 2 条件を満足する名詞を特徴語とし，これをラベルタイプ 2 とする．タイトル内に条件を満足する名詞が存在しない場合はテキストサマリ，URL の順で同様の処理を行い，条件を満足する名詞を探索する．

条件 1：各タイトルにおいて *idf* が最大値である．

条件 2：検索クエリの部分文字列ではない．

以上の条件を設定した理由は 3.1.3 項手順 (1) で述べる．

⋆1 この数字は Yahoo![2] の検索エンジンを用いて検索クエリとして「www」や「http」など多くの Web ページがヒットするような検索クエリで検索したときのヒット件数よりも多い値となるように仮に設定した値であり，Web 全体のページ数を正確に反映したものではない．

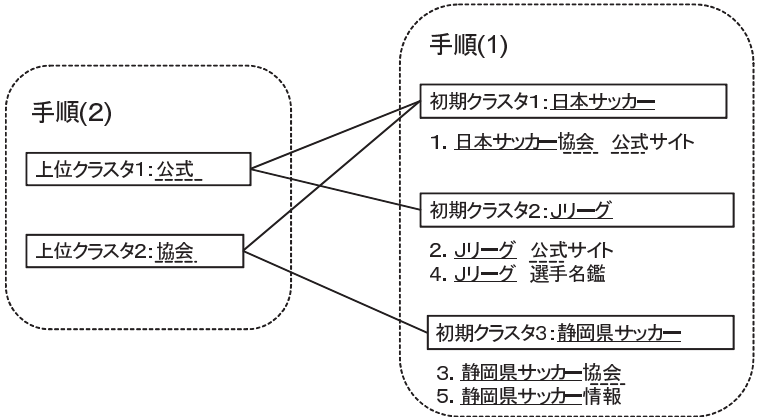

図 2　クラスタサイズの平均化とラベリング
Fig. 2　Equalization of cluster size and labeling.

3.1.3　クラスタリングおよびラベルタイプ 1 の抽出

本提案手法における検索結果のクラスタリングとラベリング手法の概略を**図 2** に示す．

手順 (1)　3.1.2 項の処理により求めた特徴語集合における各特徴語の出現回数を計測する．特徴語集合とは 1 つの検索タスクで扱われる各文書から 1 つずつ抽出された特徴語の集合である．生成されたクラスタにおける特徴語の表示順序は，*tf idf* によって調整される．*tf idf* は特徴語の出現回数 *tf* と *idf* を用いて式 (2) で定義される．

$$tf\,idf = tf \times idf \tag{2}$$

特徴語をタイトルに含む文書を集めて，クラスタを形成する．ここでは非排他的クラスタリングを行い，各文書が 2 個以上のクラスタに含まれることを許す．以下では，ここで作成されるクラスタを初期クラスタと呼ぶこととする．また，クラスタリングの指標とした特徴語

を各初期クラスタのラベルに設定し，これをラベルタイプ 2 とする．このとき 3.1.2 項の条件 2 を付加しない場合，たとえば検索クエリが “静岡大学” のとき特徴語として “静岡大学” や “静岡” が選択される可能性がある．たとえばこの例の場合，検索クエリ “静岡大学” に対する取得検索結果 100 件中 59 件がタイトル内に “静岡大学” を含み，76 件が “静岡” を含んでいた．同様に他の 5 個の検索クエリで行った結果，取得検索結果 100 件中平均 72 件がタイトル内に検索クエリを含んでいた．このような検索対象に対して上記のクラスタリングを行うと，タイトル内に検索クエリが存在する文書を 1 つのクラスタに集合させることになり 1 クラスタに膨大な文書が含まれてしまい，閲覧性が低下する．また，初期クラスタラベルとして検索クエリが設定されることになり，クラスタとしての有効性が低下する．そこで，我々の手法では，条件 2 を付加することで，クラスタ内文書数の平均化を図り，意味のあるラベルが設定されるようにする．作成される初期クラスタの例を**図 3** に示す．

手順 (2)　先の手順 (1) では得られなかったタイトル間における名詞のつながりを発見するため，形態素解析結果であるテーブル 2 に保存される名詞で以下の条件を満足する名詞を発見する．

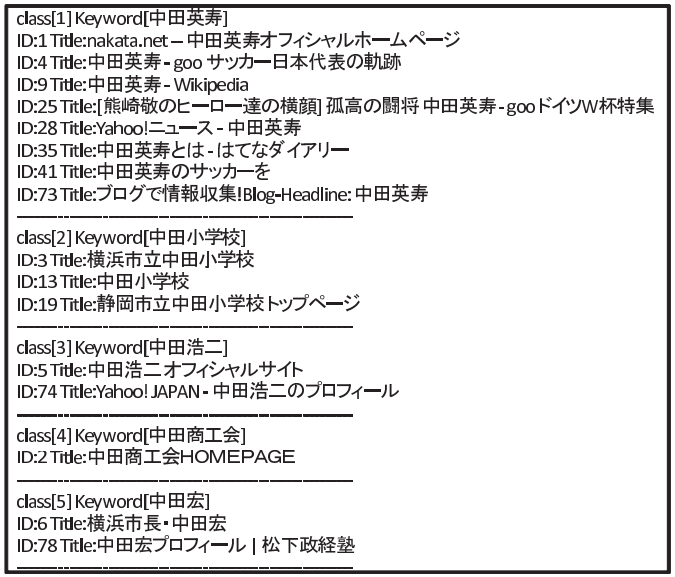

図 3　検索クエリ「中田」に対する初期クラスタの例
Fig. 3　An example of first clusters about search query 中田.

条件 1：特徴語ではなく，2 タイトル以上に出現する名詞
条件 2：その名詞が検索クエリの部分文字列ではない
条件 3：その名詞の idf が 1.5 以上

条件 2 については，手順 (1) での理由と同じである．条件 3 については，ラベルとして意味をなさないと思われる語，たとえば com，jp，co など多くの Web ページで使用される名詞を排除するため，経験的に設定した．以上の 3 つの条件を満たす名詞が使われているタイトルを含む初期クラスタを併合し，新たにクラスタを作成する．以下では，このクラスタを上位クラスタと呼ぶこととする．3 つの条件を満たす名詞を上位クラスタのラベルに設定し，これをラベルタイプ 1 とする．上位クラスタ内の初期クラスタラベルの表示順序は，式 (2) で求めた $tf\,idf$ によるランキングに従う．併合が行われなかったクラスタに関して，初期クラスタ内文書が 1 個のクラスタに関しては “その他” のクラスタに分類する．たとえば図 3 のような初期クラスタに対して上記の処理を行う場合，class[2] に所属する ID3 のタイトル内の “横浜” という名詞は class[5] の ID6 のタイトル内にも出現する．この場合，これら 2 個の初期クラスタを併合して “横浜” をラベルとする上位クラスタを作成する．

3.2　ラベルタイプ 3 とラベルタイプ 4 の生成

本節では，検索結果文書内からの，ラベルタイプ 3 およびラベルタイプ 4 の生成手法に関して述べる．本手法では単語の Web 全体集合における特徴量と検索クエリに対する検索結果集合における特徴量という 2 種の特徴量の差異を考慮することで，ラベルタイプ 3 およびラベルタイプ 4 の抽出を試みる．ラベルタイプ 3 のようなある集合において頻繁に使われるような単語の抽出は本提案手法以外にもカイ 2 乗統計値や相互情報量[21] を使うことも考えられる．本論文では，計算時間の短縮を目的として，各ラベルタイプの抽出には以下で述べる提案手法を適用した．

3.2.1　条件付き特徴量 $idf(t,q)$

3.1 節までに生成されたラベルタイプ 1 とラベルタイプ 2 は，Web 全体集合において各単語が一般的もしくは特徴的な単語であるかを考慮して生成された．しかし，その同じ単語が検索クエリに対する検索結果集合内においても同じように一般的もしくは特徴的であるとは限らず，その特性が逆転することがある．本研究では，2 個の異なる集合における特徴量の差異を考慮することで，ラベルタイプ 3 とラベルタイプ 4 の抽出を試みる．ここで，検索クエリ q に対する検索結果集合における単語 t の特徴量を，条件付き特徴量 $idf(t,q)$ と呼び，式 (3) により求める．

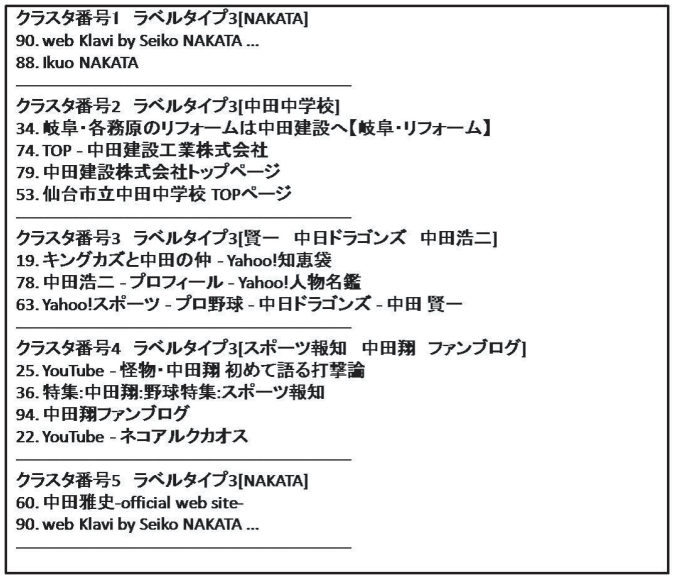

図 4　検索クエリ「中田」に対する上位クラスタに対するラベルタイプ 3 の例
Fig. 4　An example of label type 3 about top clusters about search query 中田.

$$idf(t,q) = \log \frac{dt(q)}{dt(q+t)} \tag{3}$$

$dt(q)$ は単語 q に対する検索結果ヒット件数，$dt(q+t)$ は単語 q と t の and 検索による検索結果ヒット件数を表す．

3.2.2　ラベルタイプ 3 とラベルタイプ 4 の抽出

以下にラベルタイプ 3 とラベルタイプ 4 の抽出手順を示す．

手順 (1)　条件付き特徴量の算出および単語の分類

まず，各文書のタイトル内の全名詞の条件付き特徴量を式 (3) により求める．次に，Web 全体集合および検索クエリ q に対する検索結果集合における，それぞれの特徴量の平均値によって，各単語が一般的もしくは特徴的であるかの判定を行う．手順 (1) で求めた取得した検索結果のタイトル内のすべての名詞の $idf(t,q)$ と 3.1.2 項で求めた取得した検索結果のタイトル内のすべての名詞の $idf(t)$ の，それぞれの平均値 $aveidf(t,q)$，$aveidf(t)$ を計算する．そして，平均値より高い単語を特徴的，低い単語を一般的であると仮定し，仮説に

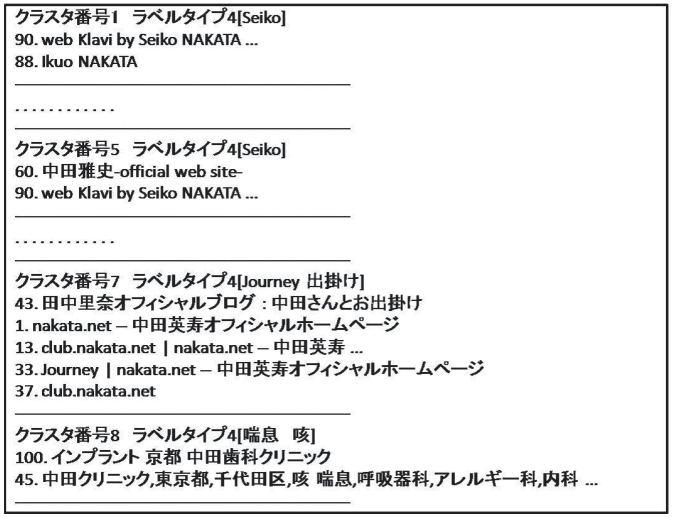

図 5　検索クエリ「中田」に対する上位クラスタに対するラベルタイプ 4 の例
Fig. 5　An example of label type 4 about top clusters about search query 中田.

沿って以下のように単語を分類する．

ラベルタイプ 3

$$idf(t) > aveidf(t) \cap idf(t,q) < aveidf(t,q)$$

ラベルタイプ 4 候補語

$$idf(t) < aveidf(t) \cap idf(t,q) > aveidf(t,q)$$

手順 (2)　ラベルタイプ 4 候補語の絞り込み

手順 1 の時点では内容的に関係のない単語もラベルタイプ 4 候補語として設定されている．そこでラベルタイプ 4 候補語が本文の html の body タグ内で使われているときに限りラベル 4 とする．**図 4** および**図 5** にラベルタイプ 3 およびラベルタイプ 4 の抽出例を示す．

4.　多視点ラベルに関するユーザビリティ評価

本研究では生成されたクラスタに対して複数の視点でラベリングを行っている．本章では，実験によりその効果を確認する．検索クエリの分類モデルに関して，Broder [16] や Rose ら [22] はユーザの検索意図に基づく検索クエリの分類モデルを提案した．Broder は検索エンジン利用者の検索クエリを navigational，informational，transactional の 3 種に分類し

表 1　本実験の検索クエリの種別
Table 1　Types of search query.

種別	説明
直接的クエリ	検索対象を直接的に示す．たとえば，政党「民主党」は検索クエリ「民主党」と表現される．
間接的クエリ	検索対象を間接的に示す．たとえば，自動車「プリウス」は検索クエリ「トヨタ　エコカー」と表現される．
多義的クエリ	検索クエリが多くの意味を持つ．たとえば，単語「さくら」は植物，人名，楽曲名などの意味を持つ

表 2　タスクタイプ 1（直接的クエリ）
Table 2　Task type1 (Direct query).

タスク番号	実験タスク	検索クエリ
1.1	2010 年サッカーワールドカップのグループリーグの組み合わせが書かれたページを 3 ページ集めよ	「2010 サッカーワールドカップ　グループリーグ組み合わせ」
1.2	鹿島アントラーズの公式ホームページを発見せよ	「鹿島アントラーズ」
1.3	探索アルゴリズムの計算量の比較に関して記述されているページを 3 ページ集めよ	「探索アルゴリズム　計算量　比較」
1.4	行政刷新会議における仕分けの対象となる事業一覧が記述されているページを 2 ページ集めよ	「行政刷新会議　仕分け対象一覧」

表 3　タスクタイプ 2（間接的クエリ）
Table 3　Task type2 (Indirect query).

タスク番号	実験タスク	検索クエリ
2.1	漫画「ホイッスル」に関するページを 3 ページ集めよ	「サッカー　漫画」
2.2	サッカーワールドカップの初代王者を答えよ	「サッカーワールドカップ　初代王者」
2.3	ドラマ「プライド」に関するページを 2 ページ集めよ	「フジテレビ　ドラマ　木村」
2.4	2009 年女子プロゴルフの賞金王を答えよ	「女子ゴルフ　2009 賞金王」

表 4　タスクタイプ 3（多義的クエリ）
Table 4　Task type3 (Ambiguous query).

タスク番号	実験タスク	検索クエリ
3.1	名字が「阿部」であるスポーツ選手を 3 人答えよ	「阿部」
3.2	地名に「栄」が含まれる都道府県を 5 個答えよ	「栄」
3.3	名字が「加藤」であるスポーツ選手を 3 人答えよ	「加藤」
3.4	名前が「大輔」であるスポーツ選手を 5 人答えよ	「大輔」
3.5	病気についてのページを 5 ページ集めよ	「ウイルス」
3.6	チーム名にジャイアンツが含まれるチームを 5 チーム答えよ	「ジャイアンツ」
3.7	人名にマイケルが含まれる人物を 4 人答えよ	「マイケル」
3.8	歌手に関するページを 4 ページ集めよ	「スティーブン」

た．また，Rose らは Broder が提案した informational をさらに階層的に再定義し，また，transactional の代わりに resource を定義した．本実験では Broder が提案した検索クエリの分類モデルを参考に検索タスクを作成する．本実験では分類モデルのうち，informational に関する検索タスクを作成し，実験を行う．さらに，本実験では検索クエリを直接的クエリ，間接的クエリ，多義的クエリの 3 種に分類する．本実験における検索クエリの分類を**表 1** に示す．本実験における計 3 個のタスクタイプごとのタスクを**表 2**，**表 3** および**表 4** に示す．本実験では，直接的クエリおよび間接的クエリは各 4 タスク，多義的クエリは他の 2 種のクエリと比較して，検索結果をクラスタリングすることの効果をより発揮するであろうと予測し，8 タスク設定した．

4.1 実験方法

本実験では提案システムである多視点ラベル型システム（以下「提案システム」）のほか，比較システムとして提案手法のラベルタイプ 1 型システム（以下「比較システム 1」），Carrot2 [14]（以下「比較システム 2」）およびリスト表示型システム（以下「比較システム 3」）の計 4 種のシステムを用意する．

比較システム 2 はオープンソースのクラスタリング検索エンジンであり，そのアルゴリズムについて文献 13)，15) で述べられているため，比較システムとして採用した．提案システムは**図 6** の右上，左下および右下の 3 つの画面を切り替える形式となっている．各画面の左フレームにはラベルタイプ 1 がつねに表示されており，中央フレームにラベルタイプ 2，ラベルタイプ 3 およびラベルタイプ 4 のいずれかが表示され，ラベル一覧の上部に他のラベルタイプへのリンクが貼られており，リンクをクリックすることで，表示ラベルタイプが切り替わる仕組みになっている．比較システム 1 は他のラベルタイプへのリンクは貼られておらず，つねにラベルタイプ 1 が表示されている．比較システム 1 は，クラスタに対して複数種のラベルタイプが提示されることの有効性を確かめるためのものである．比較システム 1 では，文献 9) で示された単一ラベルタイプによる表示と同一のものとなるように，ラベルタイプ 1 のみが表示され，他のラベルタイプのラベルは表示されないようになっている．比較システム 1 は図 6 の左上，比較システム 2 は**図 7**，比較システム 3 は**図 8** のように示される．提案システムは画面の左フレームおよび中央フレームに，比較システム 1 は左フレームにラベル一覧が表示され，ラベルをクリックすると画面の右フレームにそのクラスタ内文書が表示される仕組みになっている．比較システム 2 はクラスタリングアルゴリズムとしてデフォルトのアルゴリズムとして設定されている Lingo を適用し，ソースを

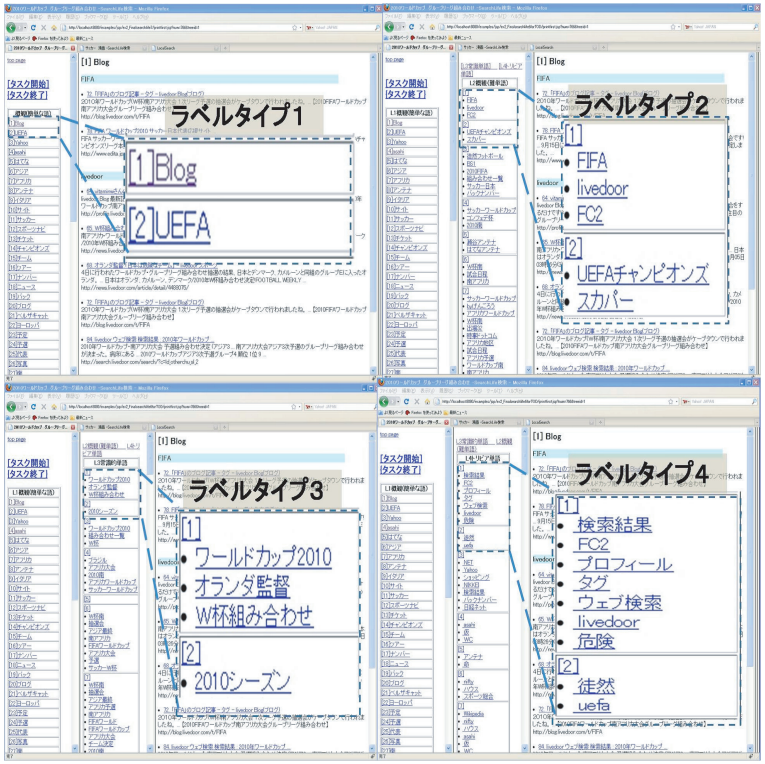

図 6　異なるラベル付け戦略によるクラスタリング結果提示
Fig. 6　Different labeling methods applied for clustered search results.

Yahoo!，言語を日本に設定した．なお，比較システム 3 は Yahoo! デベロッパーネットワーク[20] が提供するウェブ検索 Web サービスを利用したときの検索結果上位 100 件を順番に 10 件ずつ提示する．ユーザは画面下部のリンクをクリックすることで 10 件ずつブラウジングを行う．また，4 つの検索結果は共通して各文書のタイトル，テキストサマリ，URL が表示されている．被験者は情報学部 4 年生から情報学研究科修士 2 年生までの計 16 名である．いずれの被験者も情報系の学生であり，Web ブラウザと検索エンジンを用いた情報収集には，ある程度習熟している．本実験では，被験者を 4 つのグループに分け，1 つの実験タスクに対してはそれぞれが異なるシステムを用い，各被験者はすべてのシステムを同一

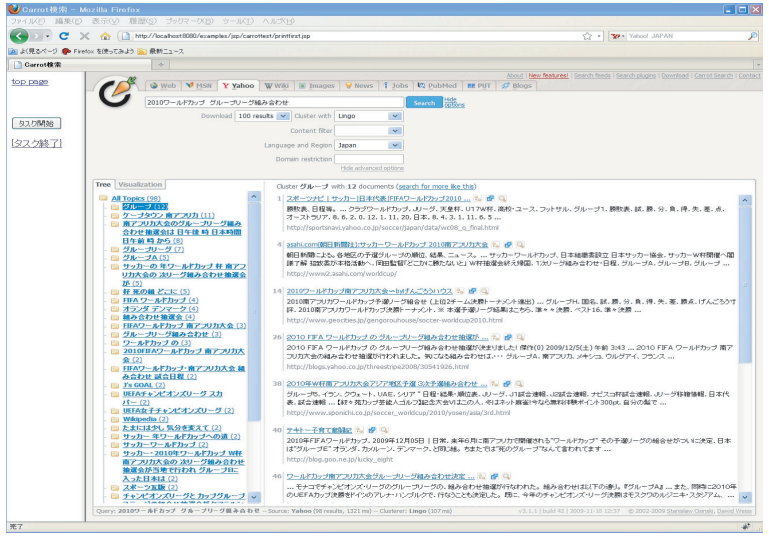

図 7　Carrot2 によるクラスタリング結果提示
Fig. 7　Example search results by Carrot2.

回数用いることになるようにした．グループ 1 は提案システム → 比較システム 1 → 比較システム 2 → 比較システム 3 の順番にシステムを使用する．同様に，グループ 2 は比較システム 1 → 比較システム 2 → 比較システム 3 → 提案システム，グループ 3 は比較システム 2 → 比較システム 3 → 提案システム → 比較システム 1，グループ 4 は比較システム 3 → 提案システム → 比較システム 1 → 比較システム 2 の順番にシステムを使用する．被験者はタスクタイプ 1 およびタスクタイプ 2 では各システムを 1 回ずつ，タスクタイプ 3 では，タスク数が 8 個なので各システムを 2 回ずつ使用する．

被験者に対しては，表 2 から表 4 に示される合計 16 タスクを順番に 1 タスクずつ提示する．各タスクの遂行では，本実験の設計者が設定した検索クエリを用い，それに対する各システムの結果を提示する．他の検索クエリによる再検索は行わず，提示した検索結果内でタスクに対する解答および Web ページを探す．また，各タスクにおいて制限時間は 5 分とし，制限時間を超過した場合の被験者のタスク達成時間は 300 秒とする．実験開始前にクラスタリング型検索エンジンや提案システムについての効果や使い方などの簡単な説明を行う．被験者は「タスク開始」ボタンをクリックして，1 つのタスクを開始し，そのタス

図 8　検索結果のリスト型表示の例
Fig. 8　Example search result in list structure.

クが終了したら画面左の「タスク終了」ボタンをクリックする．被験者は提案システムおよび比較システム 1 を使用する場合，解答 Web ページの文書番号および発見した解答 Web ページが含まれる上位クラスタの番号を記述し，比較システム 2 および比較システム 3 を使用する場合は解答 Web ページの文書番号のみを記述する．1 つのタスクタイプに対する実験は続けて行い，タスクタイプが 1 種類終了するごとに，各タスクにおける使用システムについてのアンケートを行う．各タスクにおける使用システムについてのアンケートではタスク達成のために役に立っていたと思われるシステムの順番を回答させる．本実験における評価指標はタスク達成時間およびアンケートより算出される支援率である．タスク達成時間は 1 つのタスクにおいて，被験者が「タスク開始」ボタンをクリックしてから「タスク終了」ボタンをクリックするまでの時間である．支援率はシステムがタスク達成のためにどれだけ役に立っていたかを示す値として本論文で導入した指標である．1 つのタスクに対する，あるシステムの支援率の算出手順は次のとおりである．まず，各タスクにおける使用システムについてのアンケートに対し，1 番役に立っていたと思われるシステムとされたものに 3 ポイント，2 番目に役に立っていたと思われるシステムとされたものに 2 ポイント，3 番目に役に立っていたと思われるシステムとされたものに 1 ポイント，4 番目に役に立っていたと思われるシステムとされたものには 0 ポイントを各システムに与えることとする．次に以下の式 (4) を用いて 1 つのタスクにおけるシステム A の支援率 support(A) を求める．TotalPoint(A) は 1 タスクにおいてシステム A が獲得したポイントの合計であり，Examinee(A) は 1 タスクにおいてシステム A を使用した被験者数である．

$$support(A) = \frac{TotalPoint(A)}{3 \times Examinee(A)} \tag{4}$$

4.2　実験結果と考察

表 5 に各システム 4 名の被験者の平均タスク達成時間，タスク達成人数および支援率を示す．平均タスク達成時間について，制限時間である 5 分以内にタスクを達成できない被験者の達成時間を 300 秒[⋆1]として平均タスク達成時間を算出する．タスクタイプごとの支援率を**表 6** に示す．**表 7** にタスクタイプ 1 の平均支援率に関するシステム間の t 検定結果を示す．**表 8** にタスクタイプ 2 の平均支援率に関するシステム間の t 検定結果を示す．**表 9** にタスクタイプ 3 の平均支援率に関するシステム間の t 検定結果を示す．

タスクタイプ 1 の直接的クエリにおいて，表 6 より，比較システム 1 および比較システム 3 の支援率が高くなっている．表 5 より，比較システム 3 は，4 タスクすべてにおいて平均タスク達成時間が一番短く，すべての被験者がタスクを達成している．これは，検索クエリが検索対象を詳しく表現しており，曖昧性が少ないことで，リスト表示において，解答ページが上位にランキングされた結果であると考えられる．また，提案システムおよび比較システム 1 のクラスタリング手法は巨大なクラスタが生成されることを防止するため，クラスタリングの際の指標となるラベルは検索クエリの部分文字列ではない名詞とした．そのため，システムが適切なラベルを提示することができず，被験者は適切なクラスタを選択することができなかったと考えられる．

タスクタイプ 2 の間接的クエリにおいて，提案システムの支援率が一番高くなっており，平均タスク達成時間は比較システム 1 と比較して，タイプ 2 の全タスクにおいて短くなっている．表 8 より，提案システムと比較システム 1 での平均支援率に関する有意差が認められた．これはラベルタイプ 1 以外のラベルタイプがタスクを達成するうえで有益であっ

⋆1 限られた時間内で評価実験を効率的に遂行するために，本実験では制限時間を 300 秒とした．なお，仮にこの値を 2 倍の 600 秒として再計算した場合でも，例外的にタスク 3.6 の比較システム 1 の達成時間が比較システム 3 よりも長くなることを除いては，タスク達成時間の順位が入れ替わることはなかった．この値を 300 秒としても，ここでの議論を進めるうえで深刻な問題はないと考える．

表 5　実験結果
Table 5　Experimental result on 3 task types.

タスク番号	提案システム（多視点ラベル）			比較システム 1（ラベルタイプ 1）			比較システム 2（Carrot2）			比較システム 3（リスト）		
	平均タスク達成時間（秒）	支援率	達成人数	平均タスク達成時間（秒）	支援率	達成人数	平均タスク達成時間（秒）	支援率	達成人数	平均タスク達成時間（秒）	支援率	達成人数
1.1	270	0.50	3	221	0.75	3	216	0.08	3	160	0.92	4
1.2	262	0.42	2	195	0.58	4	208	0.08	3	11	0.75	4
1.3	273	0.50	2	252	0.50	3	283	0.25	3	206	0.42	4
1.4	195	0.75	3	196	0.67	2	173	0.16	4	115	0.41	4
2.1	118	0.75	4	140	0.25	4	71	0.67	4	110	0.58	4
2.2	183	0.75	3	234	0.17	2	205	0.67	2	212	0.25	2
2.3	35	0.67	4	119	0.50	4	94	0.75	4	99	0.58	4
2.4	169	0.50	3	250	0.33	1	126	0.42	4	87	0.33	4
3.1	226	0.58	2	154	0.83	3	201	0.50	3	222	0.17	3
3.2	243	0.83	3	287	0.67	2	300	0.00	0	288	0.00	1
3.3	164	0.67	3	151	0.67	4	192	0.25	2	138	0.58	3
3.4	192	0.58	3	135	0.75	4	236	0.33	2	160	0.50	4
3.5	135	0.67	4	74	0.67	4	108	0.33	4	154	0.42	4
3.6	192	0.92	4	202	0.58	3	187	0.25	4	216	0.16	4
3.7	157	0.67	4	153	0.67	4	133	0.50	3	137	0.25	4
3.8	176	0.83	4	104	0.83	4	263	0.00	4	236	0.25	4

表 6　タスクタイプごとの各システムの支援率
Table 6　support(A) of compared systems in each task type.

タスクタイプ	提案システムの平均支援率	比較システム 1 の平均支援率	比較システム 2 の平均支援率	比較システム 3 の平均支援率
タイプ 1（直接的クエリ）	0.54	0.63	0.15	0.62
タイプ 2（間接的クエリ）	0.67	0.31	0.63	0.44
タイプ 3（多義的クエリ）	0.72	0.71	0.27	0.29

表 7　タスクタイプ 1 の平均支援率に関する t 検定結果（$p < 0.05$）
Table 7　Result of t-test about average support(A) of tasktype1 ($p < 0.05$).

標本 1	標本 2	有意差
提案システム	比較システム 1	なし
提案システム	比較システム 2	あり
提案システム	比較システム 3	なし
比較システム 1	比較システム 2	あり
比較システム 1	比較システム 3	なし
比較システム 2	比較システム 3	あり

表 8　タスクタイプ 2 の平均支援率に関する t 検定結果（$p < 0.05$）
Table 8　Result of t-test about average support(A) of tasktype2 ($p < 0.05$).

標本 1	標本 2	有意差
提案システム	比較システム 1	あり
提案システム	比較システム 2	なし
提案システム	比較システム 3	なし
比較システム 1	比較システム 2	あり
比較システム 1	比較システム 3	なし
比較システム 2	比較システム 3	なし

表 9　タスクタイプ 3 の平均支援率に関する t 検定結果（$p < 0.05$）
Table 9　Result of t-test about average support(A) of tasktype3 ($p < 0.05$).

標本 1	標本 2	有意差
提案システム	比較システム 1	なし
提案システム	比較システム 2	あり
提案システム	比較システム 3	あり
比較システム 1	比較システム 2	あり
比較システム 1	比較システム 3	あり
比較システム 2	比較システム 3	なし

たからであると考えられ，被験者のクリックログからも，ラベルタイプ 2 のラベルを基に正解ページに到達しているケースが多く確認された．

タスクタイプ 3 の多義的クエリにおいて，提案システムおよび比較システム 1 の支援率がリスト表示型である比較システム 3 と比較して高くなっている．また，表 9 より，提案システムと比較システム 3，比較システム 1 と比較システム 3 での平均支援率の有意差が認められた．リスト型表示においては，直接的クエリや間接的クエリと比較して，検索クエリの曖昧性が増加しており，検索結果内の解答ページ以外のページが増加する．その結果，解答ページが上位にランキングされにくい状態となり，線形的な探索を行う場合が多くなっている．これに対し，クラスタリング型表示である提案システムおよび比較システム 1 は類似文書がグループ化されているため，被験者が適切なラベル選択を行えた場合には短時間で複数の解答ページを収集できる．また，提案システムは比較システム 1 と比較して，タスク達成時間は長くなっているタスクが多くある．これは被験者に対する実験後のインタビューにおける「提示されるラベル情報が多すぎて提案システムが使いにくい場合があった」，「ラベルタイプ 2 のラベルを集中的に調べすぎてしまった」という指摘に関連している．ラベルタイプ 2 のラベルはラベルタイプ 1 と比較して，多くの名詞から構成され，名詞の Web 全体における頻出度も低い単語である．つまり，ラベルタイプ 2 は単語の意味の理解容易性が低いため，検索対象について知識が少ない被験者であると，ラベルの意味を理解することが困難であり，解答に関連のない不要なラベルをクリックしてしまう回数が多くなってしまったと考えられる．これに対し，ラベルタイプ 1 のみを表示する比較システム 1 のラベルは 1 つの名詞から構成され，意味の理解容易性が高いため，検索対象について知識が少ない被験者でも，適切なラベル選択が行えたと考えられる．たとえば，検索タスクとして「名字に「中田」が含まれるサッカー選手以外のスポーツ選手を答えよ」というタスクがあったとする．3.1.3 項図 3 に多義語「中田」に対する初期クラスタの例が示されている．class[1] のラベル「中田英寿」はラベルタイプ 2 のラベルであり，「中田英寿」はサッカー選手である．中田英寿という人物を知っている人であれば，そのクラスタ内文書の精査を回避することができるが，知らない場合には，クラスタ内文書の閲覧を行う必要が生じる．このような場合に，このクラスタに対して，「中田英寿」より一般的な言葉であると考えられる「サッカー」という単語がラベルタイプ 1 として設定されれば，「中田英寿」を知らないユーザでもクラスタ文書の精査を回避することができると考えられる．提案システムは他の 3 種と比較して，提示される情報量も多く，ユーザが各ラベルタイプの特性を理解し，自身がどのラベルタイプを集中的に閲覧するかを判断する必要があるため，ある程度の「慣れ」が必要である．よって，ユーザが提案システムに熟練し，自身に適切なラベルタイプを選択できるような場合や，逆にシステムがユーザの検索意図，目的，および検索対象に関する理解度を判断し，適切なラベルタイプを推薦できるような場合には，提案システムはより迅速な情報収集を行えると考えられる．

以上のことから，本実験で使用した 4 種のシステムと 3 種のタスクタイプの関係を評価すると，直接的クエリに関しては比較システム 3，間接的クエリに関しては提案システムが有効であると考えられる．多義的クエリに関しては，比較システム 1 が高い支援率を示し，平均タスク達成時間も短いタスクが多いことから，有効であると考えられる．また，多義的クエリに関して，提案システムも比較システム 1 とほぼ同様な高い支援率を示しており，平均タスク達成時間も比較システム 1 には及ばないが良好な結果を示している．提案システムの 4 種のラベルタイプの中には比較システム 1 で提示するラベルタイプ 1 が含まれており，ユーザがシステムに熟練した場合，もしくは，システムがユーザに適切なラベルタイプを推薦できるような場合には，提案システムは比較システム 1 より有効に機能することも考えられる．本提案システムでは，クラスタリングが有効となるようなタスクタイプ 2 および 3 のいずれでも安定して高い有効性を示している．従来のクラスタリング検索エンジンで課題であった，検索タスクの種類によって有効性が大きく変化してしまうという問題は，本手法により改善できたと考えられる．

5. 議　　論

5.1 ラベルタイプ 3 およびラベルタイプ 4 の効果

ラベルタイプ 3 およびラベルタイプ 4 は，クラスタ内の単語の共起関係には基づかず，クラスタ内に 1 つでもそのラベルが抽出された場合には，そのタイプのクラスタラベルとして提示するようになっている．逆に，クラスタ内に 1 つもラベルタイプ 3 あるいはラベルタイプ 4 に該当するラベルが抽出されない場合もある．そこで，4 章で用いた計 16 の評価用タスク中で，どの程度ラベルタイプ 3 あるいはラベルタイプ 4 のラベルが抽出できたかを調べた結果を示す．本クラスタリング手法は非排他的クラスタリングを用いており，同一の検索結果ページが複数のクラスタに属することがある．ラベルタイプ 3 およびラベルタイプ 4 で抽出できたラベルの数をクラスタ単位で数えた場合，同一の検索結果ページから抽出されたラベルを複数のクラスタで出現した回数分だけ数えてしまうことを避けるため，それらのラベルを抽出できた検索結果ページの数で示すことにする．ラベルタイプ 3 あるいはラベルタイプ 4 として提示されるラベルが抽出できた検索結果ページを，ここではそ

れぞれのラベルタイプに対するラベル抽出ページと呼ぶ．また，ラベル抽出ページに含まれた検索結果ページ中で，当該タスクの遂行に寄与する解答を含むページを，解答ラベル抽出ページ，その中で，ラベルタイプ 1 およびラベルタイプ 2 のいずれとも異なるラベルを抽出できた検索結果ページを，ユニーク解答ラベル抽出ページと呼ぶことにする．これらの数をタスクごとにまとめたものを，**表 10** に示す．なお，当該タスクの遂行に寄与する解答を含むページであるかどうかの判定は，直接的クエリおよび間接的クエリの計 8 タスクについては著者の 1 名が，多義的クエリの計 8 タスクについては被験者の 1 名が行った．

いずれのタスクについても，10 を超える数のラベル抽出ページが，ラベルタイプ 3 およびラベルタイプ 4 のそれぞれで見られた．ユニーク解答ラベル抽出ページは，数は多くないが 75 パーセントほどのタスクで抽出できており，タスク 2.1 や 3.1 のように，片方のラベルタイプでは抽出できなかった場合でも，他方のラベルタイプではユニーク解答ラベル抽出ページが得られた場合があった．

ユニーク解答ラベル抽出ページで提示されたラベルタイプ 3 あるいはラベルタイプ 4 のラベルが，どの程度実際のタスク遂行に寄与したのかの判定は，本評価実験では必ずしも容易ではない．本評価システムでは，ラベルタイプ 1 と並記する形でラベルタイプ 2 から 4 のうち 1 つを提示するようになっており，同時に見えているどちらのラベルタイプの効果で目的の解答ページにたどり着いたかが，必ずしも明確にならない場合がある．さらに，実際のタスク遂行時には，複数のラベルタイプのラベルを総合して見ることで，クラスタの特徴を被験者がつかんでいたことも考えられる．本来であれば，提案手法が比較的有効に働いたタスクに対してラベルタイプ 3 およびラベルタイプ 4 が貢献した事例を示すべきであるが，これらの理由から，評価システムに対する被験者らの操作ログから確実に判断可能であった事例として，タスク 1.2 および 1.3 の場合について示す．

表 10　ラベルタイプ 3 およびラベルタイプ 4 におけるユニーク解答ラベル抽出ページ
Table 10　Pages that include unique answer labels on label type 3 and label type 4.

タスク番号	ラベルタイプ 3			ラベルタイプ 4		
	ラベル抽出ページ数	解答ラベル抽出ページ数	ユニーク解答ラベル抽出ページ数	ラベル抽出ページ数	解答ラベル抽出ページ数	ユニーク解答ラベル抽出ページ数
1.1	49	9	3	55	6	3
1.2	22	0	0	25	0	0
1.3	19	5	2	32	4	3
1.4	63	6	1	42	2	0
2.1	32	1	0	23	4	4
2.2	42	0	0	26	3	3
2.3	67	15	2	47	3	2
2.4	60	1	0	63	3	1
3.1	36	5	2	15	0	0
3.2	28	7	4	14	2	2
3.3	33	1	1	13	1	0
3.4	49	21	6	14	2	2
3.5	23	4	1	22	3	2
3.6	40	31	7	17	10	10
3.7	47	36	5	15	11	11
3.8	55	8	1	25	1	1

ラベルタイプ 3 に関して，被験者のタスク遂行における，操作ログを基にその有効性を考察する．本実験において，操作ログから確実にラベルタイプ 3 のラベルをクリックして解答ページを発見したのは 2 回であった．直接的クエリの 2 タスクにおいて，それぞれ 1 名の被験者がラベルタイプ 3 のラベルを利用して，解答ページを発見している．タスク 1.2 において，提案システムを使用し，タスクを達成した被験者は 2 名であり，その 1 名はラベルタイプ 3 の「J リーグ公式 クラブガイド」というクラスタラベルを基に解答ページを発見している．このとき，同クラスタのラベルタイプ 1 は「サイト」，ラベルタイプ 2 は「Kashima サポーターズサイト　クラブガイド Antlers 日本代表 オフィシャルサイト」，ラベルタイプ 4 は「Cafe Fan」であった．タスク 1.3 において，提案システムを使用し，タスクを達成した被験者は 2 名であり，その 1 名はラベルタイプ 3 の「情報処理 性能評価 文字列探索」というクラスタラベルを基に解答ページを発見している．このとき，同クラスタのラベルタイプ 1 は「単位」，ラベルタイプ 2 は「教官コード 電子制御　基礎　2」，ラベルタイプ 4 は「選 科目」であった．

ラベルタイプ 4 の効果が端的に表れた事例として，タスク 3.7 における事例がある．タスク 3.7 における検索クエリは「マイケル」である．ラベルタイプ 3 として，27 個のラベルが生成された．その中には「マイケルジャクソン」，「マイケル中村」，「マイケル福島」，「マイケル長澤」などタスクに対する解答であるラベルも含まれていた．そのほかにも「ジョージマイケル」の「ジョージ」や「マイケルジョンソン」のジョンソンなど解答に結び付くラベルも含まれていた．ラベルタイプ 4 としては，15 個のラベルが生成された．ラベルタイプ 3 の「マイケルジャクソン」のようにラベルが解答そのものとなるようなラベルは生成されていなかった．一方，たとえばラベルタイプ 4 である「Achievement」というラベルを持つページの本文中に「マイケルボルダック」という人名が記述されていた．これはラベルタイプ 1，ラベルタイプ 2，およびラベルタイプ 3 では示されていない人名であり，「Achievement」という単語が

個人の業績を連想させるものであるととらえることもできることから，この「Achievement」はタスク遂行に貢献できるラベルとなっていた可能性がある．これ以外にラベルタイプ 3 およびラベルタイプ 4 の個々のラベリング手法によって得られた定性的評価については文献 23) で述べている．文献 23) の場合では固有名詞の一部分（たとえば，「道」など）が誤って抽出される例があった．これらの形態素解析の問題を含めたラベル生成手法の洗練は今後の課題である．

ラベルタイプ 3 およびラベルタイプ 4 は，クラスタリング手法によらないラベルの抽出方法であるため，他のクラスタリング手法との併用を行うことが可能である．これらのラベル抽出手法が他のクラスタリングアルゴリズムでどの程度の効果を発揮するかの検証は，今後の課題である．

5.2　クラスタリングの精度の影響

検索結果に対するクラスタリング精度は，閲覧の行いやすさにとって重要な要素の 1 つである．文献 9) では，排他的クラスタリングである成田らの手法[8] と，非排他的クラスタリングである本手法（ラベル 1 を用いた場合）とを，文献 8) の実験条件に従い，5 つの実験用検索クエリに対して，クラスタ再現率，クラスタ適合率，およびその調和平均をとった F 値を求め，比較した．成田らの手法と比較して，クラスタ再現率では成田らの手法が平均 0.287 であったのに対し，本手法では平均 0.409 と，再現率が上がる形になった．一方で，クラスタ適合率では成田らの手法が平均 0.833 であったのに対し，本手法では平均 0.678 となった．F 値では，本手法が 51.0 となったのに対し，成田らの手法では 42.7 となり，本手法の値が高くなっている．本定量的評価の詳細については，誌面の都合で割愛し，文献 9) に譲る．非排他的クラスタリングでクラスタ再現率に優れた性能が得られる傾向があることは，文献 12) などでも指摘されている．

本提案手法を含むいくつかのクラスタリング手法では，その手法そのものにクラスタに対するラベリング手法が（全部あるいは一部が）内包されているため，ラベリング手法とクラスタリング手法の組合せを検索対象に応じて柔軟に選択することを，必ずしも容易には行えない．目的に応じてクラスタ適合率とクラスタ再現率のバランスを調整しながら，適切なラベリング手法の組合せを選択できるようなクラスタリング手法の実現は，今後の課題である．

6. おわりに

本論文では，ユーザが特定タスクのための Web ページからの情報収集の効率化を目的として検索結果の概観把握が容易となるような検索結果をクラスタリングして提示し，生成クラスタに対して複数の手法でラベリングを行うシステムとして SearchLife を提案した．クラスタリングによる検索結果表示に対するユーザビリティ評価を行い，提案システムを用いた場合においてページ閲覧を効率的に行える場合があることを確認できた．今後の課題としては，ユーザの好みに応じてクラスタリング手法やラベリング手法を変化させるパーソナライズ手法の実現があげられる．本論文で示した実装では，複数のラベリング手法によるラベル提示を実現しているが，画面内でのラベルの効果的な提示方法については踏み込んだ検討をしていない．異なる手法によって得られたラベルどうしを融合して 1 つのラベルのように表示したり，検索語や検索タスクの種類などから適切と考えられるラベリング手法を自動的に選択して提示する手法の実現は，今後の課題である．

参　考　文　献

1) Google. http://www.google.com/
2) Yahoo!. http://www.yahoo.com/
3) Ferragina, P. and Gulli, A.: A Personalized Search Engine Based on Web-Snippet Hierarchical Clustering, *WWW'05: Special interest tracks and posters of the 14th international conference on World Wide Web*, ACM, pp.801–810 (2005).
4) Xu, S., Jin, T. and Lau, F.C.: A New visual Search Interface for Web Browsing, *Proc. 2nd ACM International Conference on Web Search and Data Mining*, pp.152–161, ACM (2009).
5) Clusty. http://www.clusty.com/
6) Zamir, O. and Etzioni, O.: Grouper: A Dynamic Clustering Interface to Web Search Results, *WWW'99: Proc. 8th international World Wide Web Conference*, pp.1361–1374, Elsevier North-Holland, Inc. (1999).
7) 大野成義，渡辺　匡，片山　薫，石川　博，太田　学：Max Flow アルゴリズムを用いた Web ページのクラスタリング方法の提案，日本データベース学会論文誌（DBSJ Letters），Vol.4, No.2, pp.13–16 (2005).
8) 成田宏和，太田　学，片山　薫，石川　博：階層的クラスタリングを利用したメタ検索エンジンの提案，技術報告，電子情報通信学会技術研究報告 DE2002-61 (2002).
9) 村松亮介，福田直樹，石川　博：分類階層を利用した検索エンジンの検索結果の構造化とその提示方法の改良，電子情報通信学会第 19 回データ工学ワークショップ，B6-3 (2008).
10) Hearst, M.A. and Pedersen, J.O.: Reexamining the Cluster Hypothesis: Scater/Gather on Retrieval Results, *SIGIR'96*, pp.76–84, ACM (1996).
11) Pantel, P. and Lin, D.: Document clustering with committees, *SIGIR'02*, pp.199–

206, ACM (2002).
12) 成田宏和，太田　学，片山　薫，石川　博：Web 文書検索のための非排他的クラスタリング手法の提案，電子情報通信学会第 14 回データ工学ワークショップ，2-p-01 (2003).
13) Osinski, S., Stefanowski, J. and Weiss, D.: Lingo: Search Results Clustering Algorithm Based on Singular Value Decomposition, *Proc. International IIS: IIPWM'04 Conference*, pp.359–368 (2004).
14) Carrot2. http://search.carrot2.org/stable/search
15) Stefanowski, J. and Weiss., D.: Carrot^2 and language properties in web search results clustering, *Proc. AWIC-2003, 1st International Atlantic Web Intelligence Conference*, Vol.2663 of Lecture Notes in Computer Science, pp.240–249, Springer (2003).
16) Broder, A.: A taxonomy of web search, *SIGIR Forum 36*, Vol.36, pp.3–10, ACM (2002).
17) Imafuji, N. and Kitsuregawa, M.: Finding Web Communities by Maximum Flow Algorithm Using Well-Assigned Edge Capacities, *IEICE Trans. Inf. Syst.*, Vol.E87-D(2), pp.407–415 (2004).
18) Ahn, J.-W., Brusilovsky, P., He, D., Grady, J. and Li, Q.: Personalized Web Exploration with Task Models, *Proc. 17th international conference on World Wide Web*, ACM, pp.1–10 (2008).
19) 高見真也，田中克己：検索目的に基づくスニペットの動的再生成によるウェブ検索結果の個人適応化，日本データベース学会論文誌（DBSJ Letters），Vol.6, No.2, pp.33–36 (2007).
20) Yahoo! デベロッパーネットワーク．http://developer.yahoo.co.jp/
21) Manning, C.D., Raghavan, P. and Schuetze, H.: *Introduction to Information Retrieval*, Cambridge University Press (2008).
22) Rose, D.E. and Levinson, D.: Understanding User Goals in Web Search, *WWW '03: Proc. 13th international conference on World Wide Web*, pp.13–19, ACM (2004).
23) 村松亮介，横山昌平，福田直樹，石川　博：単語の特徴量を考慮した検索結果クラスタに関する多視点融合型スニペットの構築，第 146 回データベースシステム研究発表会（iDB フォーラム 2008），pp.301–306 (2008).

(平成 21 年 12 月 20 日受付)
(平成 22 年 3 月 31 日採録)

（担当編集委員　小山 聡）

村松　亮介

2008 年静岡大学情報学部情報科学科卒業．同年同大学大学院情報学研究科修士課程入学．2010 年同大学院情報学研究科修士課程修了．同年 NTT データ入社．現在に至る．修士（情報学）．Web マイニング，Web 情報検索に興味を持つ．日本データベース学会会員．

福田　直樹（正会員）

1997 年名古屋工業大学工学部知能情報システム学科卒業．1999 年同大学大学院工学研究科電気情報工学専攻博士前期課程修了．2002 年同大学院博士後期課程修了．同年静岡大学情報学部情報科学科助手．平成 19 年より同助教．現在に至る．博士（工学）．IEEE-CS，ACM，人工知能学会，ソフトウェア科学会，情報システム学会各会員．

横山　昌平（正会員）

静岡大学情報学部助教．産業技術総合研究所を経て 2008 年より現職．2006 年東京都立大学大学院工学研究科修了，博士（工学）．データベース技術の研究開発に従事．電子情報通信学会，日本データベース学会各正会員．情報処理学会論文誌（データベース）幹事補佐．

石川　　博（フェロー）

静岡大学情報学部情報科学科教授．東京大学理学部情報科学科卒業．東京都立大学を経て 2006 年より現職．東京大学博士（理学）．著書に『次世代データベースとデータマイニング—DB&DM の基礎と Web・XML・P2P への適用』（CQ 出版社），『JavaScript によるアルゴリズムデザインオブジェクト指向から DB・Web・マイニングまで』（培風館），『データベース』（森北出版）等．国際論文誌 ACM TODS，IEEE TKDE，国際学会 VLDB，IEEE ICDE 等学術論文多数，1994 情報処理学会坂井記念特別賞，1997 科学技術庁長官賞（研究功績者）受賞．情報処理学会データベースシステム研究会主査，情報処理学会（データベース）共同編集委員長，International Journal Very Large Data Bases Editorial Board，日本データベース学会理事歴任．情報処理学会フェロー．ACM，IEEE 各会員．