検索結果の概要を表すキーワード式生成による質問修正支援

松生　泰典[†]　　是津　耕司[††,†]　　小山　　聡[†]　　田中　克己[†]

† 京都大学大学院情報学研究科社会情報学専攻　〒 606-8501 京都府京都市左京区吉田本町
†† 独立行政法人　情報通信研究機構　〒 619-0289 京都府相楽郡精華町光台 3-5
E-mail: †{matsuike,oyama,tanaka}@dl.kuis.kyoto-u.ac.jp, ††zettsu@nict.go.jp

あらまし　検索エンジンを用いて Web 情報検索を行う際，ユーザが求めるページを得るための質問を生成することは困難である．この理由として，ユーザ自身の知識不足，また既存の検索エンジンの多くは，提示される結果がタイトルやテキストの単純な羅列であり，実際に１ページずつ開いてみなければ内容が理解しづらいことが挙げられる．本稿では，ユーザの質問から得られる検索結果を複数のキーワードを組み合わせたキーワード式で簡潔に表すことで，その概要を理解し，ユーザの知識発見と質問修正を支援するシステムを提案する．また，今後の課題の１つとして，話題構造や ”アスペクト ”の考え方を用いた検索結果の周辺のページ集合を扱うことについても述べる．

キーワード　Web 情報検索，視覚化，質問修正，周辺情報

Supporting the Query Modification by Making Keyword Formula of an Outline of Retrieval Result

Yasunori MATSUIKE[†], Koji ZETTSU[††,†], Satoshi OYAMA[†], and Katsumi TANAKA[†]

† Department of Social Informatics,Graduate School of Informatics,Kyoto University
Yosidahonmati,Sakyou-ku,Kyoto,606-8501 Japan
†† National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289 Japan
E-mail: †{matsuike,oyama,tanaka}@dl.kuis.kyoto-u.ac.jp, ††zettsu@nict.go.jp

Abstract　It is difficult to decide the query which can get Web pages that a user really wants in Web information retrieval by using the search engine. It is because that user do not have enough knowledge. And it is also because it is difficult to understand the context of page without peruse actually page by page so that the retrieval result is a simple enumeration of its title and the text by many of existing search engine.In this paper, it is proposed the system that support user to understand the outline of retrieval result, discover of knowledge and modify the query. In this system, retrieval result is concisely shown by the keyword formula that combines two or more words. Moreover, it is described to treat page set of surrounding of the retrieval result by using a topic structure and the idea of ”Aspect” as one of the problems in the future.

Key words　Web information retrieval,Visualization,Query Modification,circumference information

1. はじめに

昨今，爆発的な増加を続ける Web ページの中から得たい情報を求めて検索する際，ユーザがいかに目的を達成できるかを重要視する search experience という考え方が広まってきている．この考え方を実現するためには，既存の検索エンジンのようにユーザの質問に対して適合性の高い検索結果をリストするだけでなく，検索結果を含めた Web ページ集合から様々な情報を発見し，ユーザ自身がその検索目的を確認しながら漸進的に検索を進め，ユーザが本来求めている情報を得られるようにすることが重要である．既存の検索エンジンの問題点として，検索結果の膨大な量とその表示方法，そして質問の決定が困難である点が挙げられる．

つまり，情報検索には向いているが，情報探索には不向きであるといえる．上記の問題点を改善するために，ページ集合のクラスタリング [2] や表示する結果をグラフや画像を用いて視覚化するような検索エンジンもいくつか出てきている．しかし，検索結果から目的の情報を探すことは容易なことではない．

ここで本論文で取り上げるべき課題として，次の２点を挙げる．

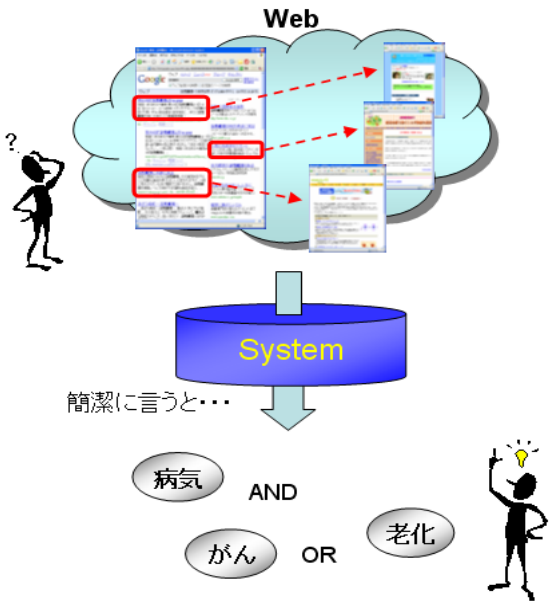

図 1　システムの概要

（1）　検索結果の概要を理解する

（2）　検索結果とその周辺のページ集合内の情報による質問修正を支援する

まず１つめは，検索結果から目的の情報を見つけ出すよりは，検索結果がどのような情報を含んでいるのかを一瞥するだけでおおまかにではあっても理解することができれば，ユーザ自身が入力した質問と目的との差異を知ることができ，質問を修正して再度検索することができるということである．また，２つめは，情報探索を目的としているような場合は，絞込み検索のように詳細な情報のみを見ていくよりは，選択したキーワードと関係のあるさまざまな情報を自動的に提示した上で, その中からよりユーザ自身の興味のある内容のものを選んだり, そこから質問をより詳細なキーワード集合に置き換えて再検索するほうが，新しい知識を発見することができたり，検索のために初めに選択したキーワードを，よりユーザの嗜好に沿った情報を取得するための質問に修正することができると考えられる．

そこで本研究では, ユーザがはじめに入力した質問の検索結果を取得し，その内容をいくつかのキーワードを論理和と論理積によって結合して生成するキーワード式をグラフ化して, 新たな検索のための質問を得るための情報としてユーザに提示し，新たな知識の発見と質問修正を支援するシステムを提案する．システムの概要を図 1 に示す．

本論文の２章では本研究の関連研究について述べ，３章では本提案手法の概要として，ユーザに提示するキーワード式とはどのようなものであるかについて述べる．４章では３章で紹介したキーワード式の生成プロセスとそのアルゴリズムについて，５章ではシステムの実装とその結果の考察について述べ，６章でまとめと今後の課題について述べる．

2.　関 連 研 究

2.1　Focus + Context View

Web 検索において，一度に提示する情報の量はブラウザやモニタのサイズなどに制限があるのに加え，ユーザが一度に把握できる情報の量にも限度がある．そのため，情報の量が膨大である場合は，その一部分だけを見せ，その上でユーザがさらに見たいと思った情報を選んで順に見ていくことで，情報を効率よく取得することができると考えられる．

情報をキーワード集合のグラフなどの簡潔に表示したものからその中の一つを選択すると，その選択したものを中心とした表示に変えるようにする手法がある．このように，選択したものを中心として, その周辺だけを見せる可視化のアプローチは，”Focus + Context View ”と呼ばれている．この例としては，FishEye View [3] [4] などが有名である．

2.2　KeyGraph

大澤ら [5] は，ある文章がどのような内容であるかを調べるために，その文章の中から重要であると考えられる単語を複数抽出してきて，それぞれの単語同士の関係を線でつないでグラフ表示するシステムを構築している．KeyGraph では一つの文章が対象であるので，Web ページであれば最大１つの Web ページという単位でグラフを生成するが，抽出してきた語を AND と OR で結合した複数通りのキーワード式という形でユーザに提示する点が，本研究が異なっている点である．

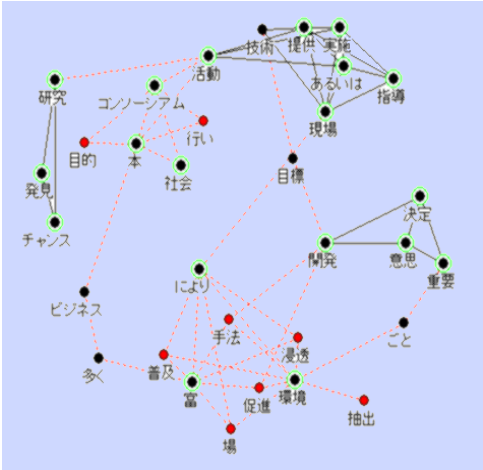

図 2　KeyGraph

2.3　検索結果とその参照ページ集合の近似的内包表現による Web 情報検索支援

筆者ら [6] はこれまでに，検索結果とその参照ページ集合それぞれの概要を表すために，検索結果とその参照ページ集合の中にそれぞれ含まれるキーワードを複数組み合わせ，それらのキーワードから得られる検索結果のページ集合により初めの質問から得られる検索結果を近似的に表すシステムを提案した．複数のキーワードの OR をとったものをキーワード式とし，その評価関数として再現率と適合率を足し合わせたものを用いて，図 3 のようなプロトタイプを作成し，キーワード式の提示を行った．

図 3　プロトタイプ

この手法では，近似的内包表現を用いて検索結果を近似的に表している．しかし，キーワード式に含める語が一般的な名詞などである場合などによく検索結果の集合とキーワード式で表されるページ集合との適合率が非常に小さくなってしまい，事実上検索結果の概要とならない場合が出てきてしまう．そのため，本論文での提案手法はこれとは別の手法によりキーワード式を生成することを考慮してのものとしている．

3.　検索結果の概要を表すキーワード式

本研究では，検索結果の概要を表すために，いくつかのキーワードを論理積（AND）と論理和（OR）を用いて接続したキーワード式を生成する．このキーワード式によって，ユーザが検索結果の概要を理解した上で新たな知識の発見とそこから質問の修正を支援することができると考える．

3.1　キーワード式とは

検索結果の概要を表そうとする際，最も単純な表し方としては，１つの語で表す方法がある．ここで，概要を表すということから，検索結果のページ集合の中で多くのページに含まれている語が評価としても高いと考えられる．しかし，１つの語ではページ集合全体を言い表すことは困難である．これは，検索エンジンで検索を行う場合に，１つの語を質問として検索することで目的のページを見つけ出すことが困難であることと同様である．そこで，検索結果の中に含まれる語を複数組み合わせ，それらの語の，出現ページ数による関係を AND と OR を用いた論理式で表したキーワード式として検索結果のページ集合の概要を表すことを考える．

3.2　キーワード式の例

生成されるキーワード式の例を次に示す．”イラク ”という質問から得られたキーワード式の一つとして，

「イラン OR（日本 AND（(政府 AND 自衛隊）OR 大統領))」

というものが挙げられる．これは，”イラク ”という質問の検索結果を表している次の，語を AND で結合した３つの式の OR をとったものである．

- 「イラン」
- 「日本 AND 政府 AND 自衛隊」
- 「日本 AND 大統領」

3.3　キーワード式の表示

生成したキーワード式は，生成した文字列のままでは，一瞥するだけで理解することは困難であるため，図 4 のように，決定木 [7] のような形でグラフ表示することにする．初めの質問を根のノードとして，その下にノードを順につなげていく．親ノードとその唯一の子ノードの関係を AND 結合，同じ親ノードに複数の子ノードが接続されている場合，その子ノード同士の関係を OR 結合としている．

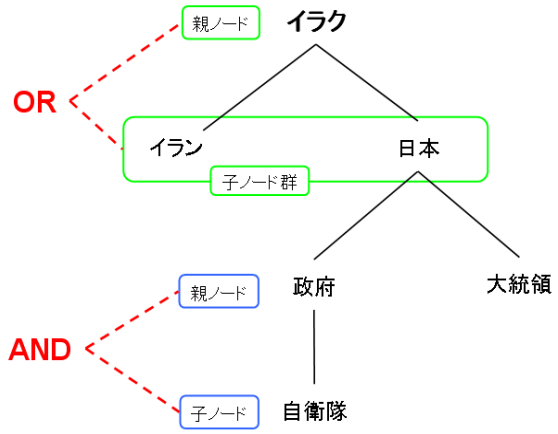

図 4　キーワード式の具体例

例えば上記の例の AND で結合した式の「日本 AND 政府 AND 自衛隊」という式は，”イラク ”という根ノードから ”自衛隊 ”という葉のノードまでの道のりを指している．

3.4　キーワード式からの質問修正

このグラフとして表示された中に含まれる語の中でユーザが興味のある語をランダムに抽出したり，根ノードからある葉ノードまでの語を複数選択してそれらを次の質問とすることができるようにすることで，ユーザの質問修正の支援ができると考えられる．例えば，”イラク”という質問から得られた上記のキーワード式の中から，”自衛隊”という語に興味を持ったとすると，”イラク”という根ノードから”自衛隊”というノードまでの語をすべて AND で結合した式を生成する．検索エンジンに入力する場合は次のように変換して入力する．

「イラク AND 日本 AND 政府 AND 自衛隊」
↓
「イラク　日本　政府　自衛隊」

という質問で検索することで,「イラクや日本という国の政府，自衛隊に関すること」についてのページを得ることができ，”イラク”という質問からさらに詳細な内容を知ることができると考えられる．

4. キーワード式の生成手順

この章では，３章で提案したキーワード式の生成のプロセスについて記述する．

提案手法のシステムの大まかな流れを次に示す．

（1） 検索結果の Web ページ集合を取得

（2） Web ページそれぞれのテキストを抽出

（3） 取得した Web ページのテキストの中に含まれる語の抽出

（4） 抽出した語群の中で，検索結果に含まれる割合の高い語それぞれについて

（a） キーワード式の基底のノードとする

（b） 基底の語に逐次的に語を追加していき，キーワード式を生成する

（5） 生成したキーワード式をユーザに木構造のグラフとしてそれぞれ提示する

次に，それぞれについて詳細に記述する．

4.1 検索結果の Web ページ集合の取得

まず，ユーザが入力した質問を検索エンジン（Google [1]）に投げ，検索結果のページ集合を取得し，それぞれのページのテキストを抽出する．検索結果は膨大な数が出てくることが多く，その全てを扱うことは現実的に困難であるため，近似的に検索結果全体を表すものとして，検索結果のページ集合のの上位 100 件を抽出し，この 100 ページを検索結果全体と近似的に表すこととする．

4.2 取得ページのテキストからの語の抽出

取得してきた Web ページのテキストをそれぞれ形態素解析する．形態素解析には，茶筌 [8] を用いる．解析結果の中から名詞と固有名詞を取り出し，キーワード式を生成するためのキーワードの候補とする．その中で，より多くのページに含まれている語を優先的にキーワード式に含めるキーワードとする．

4.2.1 語の抽出方法

検索結果の各ページを形態素解析し，その中から名詞と固有名詞を取り出してくる．ここで，連続して出現する名詞は，その品詞の役割によって接続して名詞句とするかどうかを判断する．連続して出現した場合に接続する名詞の関係の一例を下に示す．

- 一般名詞 ⇔ 一般名詞
- 一般名詞 ⇔ サ変接続名詞
- 一般名詞 ⇔ 接尾名詞
- 一般名詞 ⇔ 接続詞的名詞
- 固有名詞 → 一般名詞
- 数名詞 → 接続詞的名詞

これらの接続を用いることで，「日本－代表」や「北朝鮮－戦」，「親善－試合」などの名詞句を抽出することができる．このような名詞句も，ページ集合から抽出した語群に含めるものとする．

4.2.2 キーワード式に含める候補である語の評価基準

検索結果の概要を表す語として評価が高いと考えられるものは，少数のページだけに集中して数多く現れる語よりも，検索結果のページ集合の中でより多くのページに含まれているものであると考えられる．そこで，各語について検索結果のページ集合中の存在ページ数を求め，その値の高いものからキーワード式に含める語の候補とする．図 5 でその例を示す．語 A と語 B は検索結果のページ集合内での出現頻度（全ページ中の存在数）は同じである．しかし，語 A は多くのページに含まれているが，語 B は少ないページに多く含まれている．このような場合，ある部分のみに集中している語 B よりも，多くのページに含まれている語 A のほうが，検索結果のページ集合の概要を表すのに有用であると考えることができる．

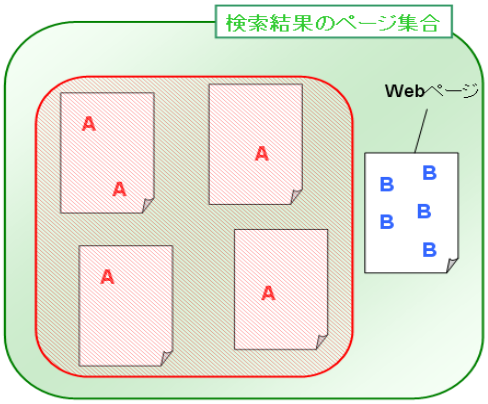

図 5 ページ集合を端的に表す語

4.3 キーワード式の生成

取得した Web ページのテキストをそれぞれ形態素解析して抽出した語群の中から，基底のノードとなる語を決める．基底のノードにそれぞれノードとして評価の高い語を逐次的に追加していく．

検索結果の概要を表すということから，親のノードに接続する場合に最も評価の高い語を選択するこのとき，親のノードの存在するページに関するページ集合として，正例と負例の二種類の集合を用意する．

4.3.1 ページ集合の正例と負例

あるノードの子として語を追加する場合を考える．検索結果のページ集合の多くのページに存在する語をその評価として，検索結果のページに数多く含まれる語を選ぶ．その際，あるノードの存在しているページ集合を正例，あるノードの存在していないページ集合を負例とし，キーワード式に追加する語の候補として，正例のページ集合に多く含まれていて，且つ負例のページ集合にはなるべく含まれていない語を選択する．例えば図 6 の例の場合，”サッカー ”というノードの子として ”日本 ”というノードが存在し，その ”日本 ”というノードに新しいノードを接続する場合を考える．ここで，”日本 ”というノードについて，

（1） 正例

”サッカー ”という語を含んでおり ”日本 ”という語を含

んでいるページ集合

（2） 負例

”サッカー ”という語を含んでおり ”日本 ”という語を含んでいないページ集合

とする．このとき，正例に多く含まれてはいるが，負例にはあまり含まれていない語（図 6 の例では ”中田 ”という語）は，”サッカー ”に関する話題であり，その中で ”日本 ”という語を含んでいるページ集合を特徴づけているということができると考えられる．

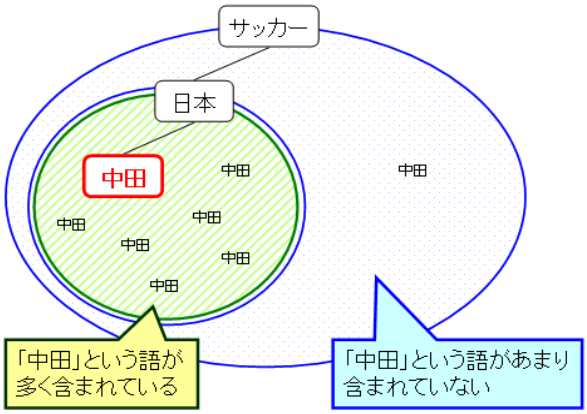

図 6　正例と負例

4.3.2　キーワード式への語の追加

キーワード式のあるノードを親としてその下に子ノードを接続するために，正例と負例を定める．抽出した語群の中からある親ノードに接続する際の評価として，正例の集合での出現ページ数が多く，また負例の集合での出現ページ数が最も少ない語をキーワード式の親ノードに接続する子ノードとして決定した場合を考える．追加する語がその親ノードの存在ページ中に高い割合で存在する場合，追加する語は親ノードの存在ページの概要をすでに表すことができていると考えることができる．親ノードの子としてはその追加する語のみでよいと考えられるため，

「親ノードの語 AND 追加する語」

のように接続する．この様子を図 7 に示す．

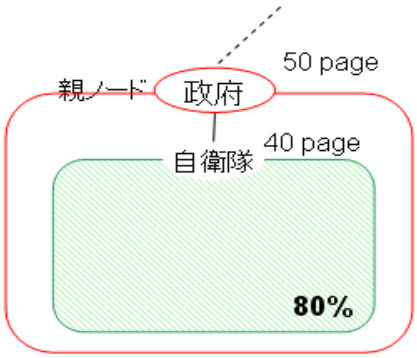

図 7　単一の子ノードの接続

それに対して，追加する語がその親ノードの存在ページ中に低い割合で存在する場合，追加する語は親ノードの存在ページの概要を一語で表すことができているとは考えることができないため，親ノードの子としてその追加する語以外にも子のノードが必要であると考えられる．そのため，他の子ノードを決定し，親ノードに追加する．このとき，新たな子を決定するときの正例と負例として，

（1） 正例

「親ノードを含んでいるページのうち，他の子ノードの語を含まないページ集合」

（2） 負例

「親ノードを含んでいるページのうち，他の子ノードの語を含んでいるページ集合」

とする．すなわち，親ノードの語を含んでいるページ集合内での，すでに追加された語と新たに追加する語の差異を明確にするということである．その例として，次のように接続する．

「親ノードの語 AND (追加する語 X OR 追加する語 Y …)」

この様子を図 8 に示す．

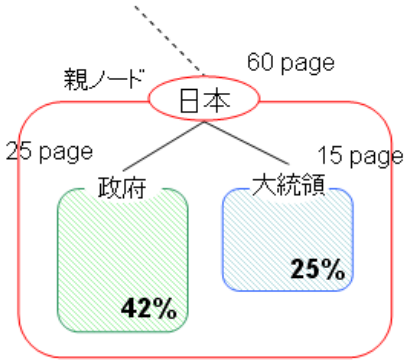

図 8　複数の子ノードの接続

このような場合，”日本 ”という語を含んでいるページ集合をさらに新たな語で表そうとする場合，”政府 ”という語だけでは表すことのできているページの割合が少なく，概要を表しているとは言いがたいため，さらにもうひとつ ”大統領 ”という語を付け加えている．”大統領 ”という語は，”政府 ”という語を含んでいないページ集合を正例，含んでいるページ集合を負例としている．そのため，”日本 ”という親ノードの語の表すページ集合は，”政府 ”という語もしくは ”大統領 ”という語で表すことができるということである．

図 7 の場合，この部分のキーワード式内での表し方は

「政府 AND 自衛隊」

となり，また図 8 の場合では，

「日本 AND (政府 OR 大統領)」

となる．

4.3.3　キーワード式生成のアルゴリズム

ユーザの質問から得られる検索結果のページ集合からキーワード式の要素の候補となる語群を抽出してきたあと，その語群の中の語を組み合わせてキーワード式を生成する．このとき，キーワード生成についてのアルゴリズムを次に示す．

（1） 検索結果のページ集合から抽出してきた語群の中から，キーワード式の最初のノードとなる語を選択し，Start_Node とする．

（2） Start_Node から次のノードの接続を開始する．

（3） Repeat(次に接続するノードの候補がなくなるまで)

（a） 親ノード p を展開する．親ノード p の子として新しいノードを接続する．このとき，検索結果のページ集合から抽出してきた語群の各語 k について，次の集合を定義する．

Pos ： 親ノード p の正例の集合
Neg ： 親ノード p の負例の集合
Pos(k) ： Pos の中で語 k が含まれる集合
Neg(k) ： Neg の中で語 k が含まれない集合

これらの集合を用いて，語 k について次の評価関数 F(k) の値をそれぞれ求める．

$$\mathrm{F(k)} = \frac{Pos(k)}{Pos} * \left(1 + \log \frac{Neg(k)}{Neg}\right) \qquad (1)$$

この評価関数の値が最も高い語 k を，親ノード p に接続する次のノードの語 n とする．

（b） AND と OR の判定のための閾値を $\mathrm{TH}_{\mathrm{and_or}}$ とする．
If (Page(n) / Page (p) > $\mathrm{TH}_{\mathrm{and_or}}$)
n を p の唯一の子ノードとして，AND 結合する．
Else
n を p に子ノードの一つとして AND 結合し，親ノード p を次の展開するノードとする．

（c） ノードの中から，次に展開するノードを p とする．
i. 接続した子が親ノードの唯一の子であった場合，未展開のノードの中から存在ページ数の最も多いものを次に展開する親ノード p とする．
ii. 接続した子が親ノードの子の中の一つであった場合，同じ親ノード p を次に展開するノードとする．

（4） 接続したノード集合を AND と OR をつないだ文字列に変換し，キーワード式とする．

このアルゴリズムを繰り返すことで，ユーザの質問から複数のキーワード式を生成し，ユーザに提示する．ユーザが選択したキーワード式について，木構造でグラフ表示する．

4.4 キーワード式の評価

生成したキーワード式は，取得してきた検索結果の概要を表すことを目的としてしているため，検索結果のページ集合をどれだけ的確に表すことができているかがその評価の基準となる．

複数の語を組み合わせてキーワード式を生成しているため，あるキーワード式 h の中に含まれている語 i を含んでいるページの数を P(k)，検索結果全体のページ数を pagenumber とすると，次のような評価関数からキーワード式の評価値 F_set(h) を求めることとする．この場合，値が 1 を超える場合も出てくるが，違う語で同じページを表していたとしても，その重なりの数が多いほうが，検索結果の概要をより強く表すことができていると考え，どのキーワード式がどの程度の割合で検索結果のページ集合を表すことができているかを比べるための指標とすることを考える．

$$\mathrm{F_set(h)} = \sum_{i=0}^{k} \frac{P(i)}{pagenumber} \qquad (2)$$

5. システムの実装

5.1 システムの概観

本論文の 4 章で提案した手法のシステムを作成した．ユーザが質問を入力すると，そのキーワードを検索エンジンである Google に入力し，その検索結果の上位 100 件分の Web ページを取得し，それぞれを形態素解析する．解析結果からキーワード式の要素の候補であるキーワードを取得し，そこからキーワード式を生成する．システムの実装画面を図 9 に示す．

図 9 システムの実装画面

5.2 システムの動作

まず，ユーザは得たい情報を表していると考えられる質問を入力する．入力した質問から検索結果のページを 100 ページ取得しそのテキストを形態素解析することによってページ集合に含まれる名詞と固有名詞，または名詞句を抽出し，語群とする．キーワード式の根ノードをユーザの最初の質問とし，そこに接続する一番目の語としては，語群の中で検索結果集合の中での出現ページ数の多い語をとする．出現ページ数の多い語の数は複数あると考えられる．語群の中に含まれる語を評価値に従って逐次的に組み合わせ，複数通りのキーワード式を生成する．

生成されたキーワード式の中で，まず最も評価値の大きなも

のをフォームにグラフ描画することで，ユーザに提示する．他のキーワード式もリスト表示しており，選択したものをフォームにグラフ描画する．

5.2.1　強 調 表 示

キーワード式はそのつくりによって非常に長いものであるため，ユーザはその目的によってキーワード式から情報を絞り込む．キーワード式をグラフ描画したものの中から，興味の深い語を選択すると，そのノードから根ノードまでの道のりが強調表示されるようにしている．ユーザの興味を誘引するために，ある程度多くの情報としてキーワード全体を見せ，その中からユーザの興味のより深い部分を限定できるようにすることで，ユーザの興味対象を明確なものとし，次の検索につなげることを目的としている．強調表示の様子を図 10 に示す．

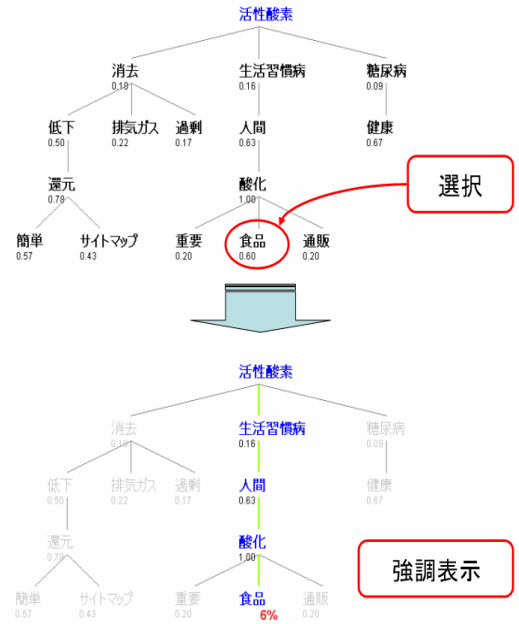

図 10　強 調 表 示

5.2.2　キーワード式の利用

ユーザが選択したノードから根のノードまでの道のりを強調表示した部分は，検索結果の中でも特にユーザの興味として表れている部分であると考えられる．選択して強調された部分の複数の語を利用し，その語を全て含んでいるページのみを閲覧するフォームを用意して，実際にページを閲覧することができるようにしている．それとともに，強調表示された部分の語群を次の質問として再度検索しなおすこともできるようにしている．この二通りの行動によって，ユーザは目的に漸進的に近づくように検索を続けることができると考えられる．

5.3　提示する情報の出力結果と考察

システムの結果を検証するために，実験を行った．実験は，最初の質問として ”活性酸素 ”を用いた．

5.3.1　結果の一例

[評価値が最も高いもの]

(病気 AND ((老化 AND ((原因 AND ((人間 AND 生活習慣病) OR (糖尿病 AND 動脈硬化))))))) OR (検索 AND (ガン OR 買い物 OR 質問)) OR (ビタミン AND ((ミネラル AND (カラダ OR 説明))))

: KeywordsetValue = 2.56

[評価値が最も低いもの]

(非常 AND ((電子 AND ((安定 AND ((病気 AND (ウイルス OR 白内障 OR サプリメント)))))))) OR (注目 AND (体内 AND (紫外線 AND (酸化力 AND 要因)))) OR (検索 AND (物質 OR サイトマップ OR 活性化))

: KeywordsetValue = 1.45

5.3.2　結果に対する評価

上の結果は，”活性酸素 ”というキーワードから得られたキーワード式の中で，その評価の最も高いものと最も低いものを抽出してきたものである．それぞれの式を主に比べてみた場合，どちらも語としても ”活性酸素 ”と関係が深い語で構成されているが，評価値の高さとして差が生まれている．

5.3.3　考　　察

実験による結果について，初めの質問と関連の深い語がその上下関係とともに出力されるため，興味のある語を用いて質問修正をすることができると考えられる．ただ，その上下関係がまだ明確なもので定義できていないため，それほどの意味を持たないという問題点が挙げられる．

また，複数回の検索をするにあたって問題となってくるのが実行時間であるが，インターネットの接続環境によってページの取得にかかる時間を除けば，それほど支障にならない時間での検索が可能である．そのため，システムとしては実行時間として複数回の検索に利用できるものであると考えられる．

6.　まとめと今後の課題

本研究では，ユーザの質問によって得られた検索結果の概要を理解することで，ユーザの知識発見と次の質問修正を支援するシステムを考案し，そのシステムを開発した．今後の課題として，検索結果の ”周辺情報 ”の導入，さらにはその周辺情報を取り扱う部分のシステムの実装を大きな課題の一つとして検討している．また，検索結果の概要を表すキーワード式の評価基準となるものを明確にすること，さらにシステムの評価としては，実際にユーザにこのシステムを使用してもらうことによって，より精度の高いキーワード式の生成方法を考案すること

6.1　検索結果の周辺情報

本研究では，検索結果として得られた Web ページ集合の概要を表し，その中にどのような知識や情報が含まれているのかを理解することでの，質問修正につなげることの支援を目的としている．ここで，検索結果以外にもユーザの求める情報を含むページや，直接的にはユーザの質問には関係ないとしても，ユーザが潜在的に興味を抱いているような内容を含んだページがあると考えられる．これらのページを取り扱うことで，ユーザの知識発見と質問修正のためにさらなる情報が提供できると考えている．そのため，次の 2 つの概念を用いて抽出してきたページ集合を検索結果の ”周辺情報 ”とし，それらについても検索結果と同様にキーワード式とて提示することも考慮して

いる．

6.1.1 ページ内での話題構造

検索結果の概要は，検索結果のページ集合の中に含まれるおおまかな内容をつかむためのものではあるが，ページはタイトルによってその内容を理解できると考えられる．また，さらに詳細なページの内容をつかむためには，テキストの部分に含まれるキーワードをページのタイトルと組み合わせることで，同じタイトルのページ間での差異を表すことができると考えられる．

6.1.2 ページ集合におけるリンク構造

ユーザの質問から得られた検索結果のページは，すべてユーザの質問を含んでいるが，ユーザが質問を的確に決定することは難しく，質問が目的の情報とずれてしまっていることも考えられる．さらに，ユーザは様々な情報の提供を求めて Web 情報探索を行うことも多いと考えられる．そこで，ユーザが入力した質問から得られる結果集合がどのような見られ方をしているのかを考えるために，リンク元のページ集合を取得し，それらを ”周辺情報 ”とすることを考案している．

また，検索結果のページ内のリンクアンカーとそのリンク先のページ集合とともにリンク元のページ集合を，検索結果集合との差異を明確化して扱うことで，ページ集合全体においての話題構造を考慮することを考えている．

謝　　辞

本研究の一部は，文部科学省科学技術振興費プロジェクト「異メディア・アーカイブの横断的検索・統合ソフトウェア開発」（代表：田中克己）および，平成 16 年度科研費特定領域研究 (2)「Web の意味構造発見に基づく新しい Web 検索サービス方式に関する研究」（課題番号：16016247，代表：田中克己）および，21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」によるものです．ここに記して謝意を表すものとします．

文　　献

[1] http://www.google.co.jp/

[2] Zeng H. et. al.: Learning to Cluster Web Search Results, Proceedings of the 27th International ACM SIGIR Conference on Research and Development of Information Retrieval (SIGIR‘ 04), 210-217,2004.

[3] Sarkar, Manojit and Marc H. Brown, “graphical fisheye views,” Comm. of the ACM, Vol.37, No.12, pp.73-83, 1994.

[4] Furnas , G. W., “Generalized Fisheye Views,” Proc. ACM SIGCHI ’86 Conference on Human Factors in Computing Systems, pp.16-32, 1986.

[5] 大澤幸生, Nels E. Benson, 谷内田正彦： KeyGraph:単語共起グラフの分割統合によるキーワード抽出，電子情報通信学会論文誌 J82-D1, No.2, pp. 391–400, 1999.

[6] 松生 泰典, 是津 耕司, 角谷 和俊, 田中克己： 検索結果とその参照文脈の近似的内包表現による Web 情報検索支援，第 15 回データ工学ワークショップ (DEWS2004), 2004 年 3 月

[7] 小久保 卓　小山 聡　山田 晃弘　北村 泰彦　石田　亨　”検索隠し味を用いた専門検索エンジンの構築 ”，情報処理学会 Vol.43 No.6，2002 年 6 月

[8] 奈良先端科学技術大学松本研究室 茶筌ホームページ: http://chasen.naist.jp/hiki/ChaSen/