データベースシステム　122－18
(2000. 7. 26)

Web検索における意味的適合フィードバック機構

平田陽一[†]　松倉健志[†]
田島敬史[††]　田中克己[†]

従来のWeb検索における適合フィードバックでは，ユーザがサンプルページの内容を見て評価し，その評価をもとに再検索または検索結果の再構成を行なう．その際，ページの評価は「良い」または「悪い」の2種類であった．そのため，従来の適合フィードバックの手法は，「良い」と評価されたページに類似したページを獲得するのには有効だが，ユーザの要求が「この話題についてのより詳しいページが欲しい」などのように複雑になると，十分にユーザの意図を汲み取ることが難しく，うまく機能しないことがあった．そこで，本研究では，単に「良い」または「悪い」の二元論的な評価に基づいて，サンプルページに類似するページを集めるのではなくて，サンプルページと実際に欲しいページの違いを評価として与えることによって，そのような差異を持つページを探してくれるような，意味的な適合フィードバック機構を提案する．ページ間の差異としては，各ページ中の単語数や，お互いのページ内の単語の共起度を用いて，ページ間の情報量や話題の広がりの相対的な差を測定する．

Semantic Relevance Feedback for Web Document Retrieval

Youichi Hirata,[†] Takeshi Matsukura,[†] Keishi Tajima[††]
and Katsumi Tanaka[†]

In conventional relevance feedback for Web query systems, the user evaluates some sample pages, and then based on that evaluation, the original query is revised, or the query results are reorganized. In that evaluation, sample pages are classified as good or bad. This approach is effective to find pages similar to the pages evaluated as good. In some cases, however, the users want to specify their requirements more specifically, such as, "I want pages discussing this topic in more detail." In the usual relevance feedback, the users cannot express such requirements in the evaluation. In this paper, we propose a framework where the users can express such requirements, which we call semantic relevance feedback. In our framework, the users specify the difference between a sample page and pages they want. We estimate the difference between pages based on the amount of information and the extent of the topic in each page. We measure the former by the number of word occurrences, and measure the latter by the ratio of cooccurrence of words in pages.

1. はじめに

インターネットはここ数年で急速に我々の身近なものになり，それに伴いWWW（World Wide Web）に存在するページの量も増加し，いまやその規模は10億ページにも及んでいる．このようなページの増加により，莫大な規模となったWWWの中から，ユーザが的確に自分の欲しい情報を発見，獲得することはほとんど不可能である．そこで，WWWから必要な情報を検索するためのツールとして多くの検索エンジンが開発された．我々はそれらを使用することによって豊富な情報資源から必要な情報を検索することができる．現在では，情報を検索するためのキーワードとして数語入力するだけで，検索エンジンは検索結果としてユーザに大量のページのURL（Uniform Resource Locator）リストを返し，ユーザはそこから必要な情報を獲得することができる．しかし，検索エンジンの検索結果ですら，一般に膨大な量になり，その中から必要な情報を獲得することは大変な労力を要する．ユーザは大量のページのURLリストからひとつひとつページを選別していかなければならない．

このような検索の際にかかる労力を軽減させるための技術のひとつに適合フィードバックというものがある．適合フィードバックとは検索されたページ群の中

† 神戸大学大学院自然科学研究科
Graduate School of Science and Technology, Kobe University
†† 神戸大学工学部情報知能工学科
Department of Computer and Systems engineering, Kobe University

からユーザがページを選択し，選択されたページの特徴ベクトルを用いて，もとの質問ベクトルを修正するというものである．Web 検索における適合フィードバックにおいては，あるページをサンプルページとしてユーザが実際に見て評価し，その評価をもとに再検索，または検索結果の再構成を行う．ここでは適合フィードバックを用いることにより，ユーザはより少ない労力で自分の要求するページを獲得することができる．

しかし、従来の Web 検索における適合フィードバックではページの評価は「良い」あるいは「悪い」の 2 種類であった．そのため，従来の適合フィードバックの手法は,「良い」と評価されたページに類似したページを獲得するのには有効だが，ユーザの要求が「この話題についてのより詳しいページが欲しい」などのように複雑になると，十分にユーザの要求をシステムが把握できず，適合フィードバックの結果とユーザの求めていたものが一致しないことがある．

そこで，本研究ではユーザの要求を把握するために,「良い」または「悪い」の二元論的な評価に基づいて,「良い」と評価されたサンプルページに類似したページ集めるだけの従来の適合フィードバックではなく，サンプルページとユーザの要求しているページの違いを評価としてシステムに渡すことによって，そのような差異を持つページのスコアを高め，検索結果にそのスコアリングを反映させる意味的な適合フィードバック機構を提案する．

具体的な手法としては，検索結果内の各ページの単語情報をそのページを表す特徴量とみなし，サンプルページと判定するページのお互いの単語数を比較したり，お互いのページ内の単語の共起度を測定することによって，そのページの相対的な情報量や話題の広がりを測定する．そしてそれらに基づく特徴ベクトル空間に各ページをサンプルページとの関係がわかるように配置し，その特徴ベクトル空間を用いて，検索結果の再ランキングを行う手法をとる．(図 1)．また、この特徴ベクトル空間における検索結果の各ページの配置はユーザの応答により動的に再構成される．

これによってユーザの要求を把握し、ユーザの要求するページを的確に発見することが可能になる．

2. 適合フィードバック

適合フィードバックとは検索されたページ群の中からユーザがページを選択し，選択されたページの特徴ベクトルを用いて，もとの質問ベクトルを修正するというものである．一般的な適合フィードバックにおいては修正された質問ベクトルは，

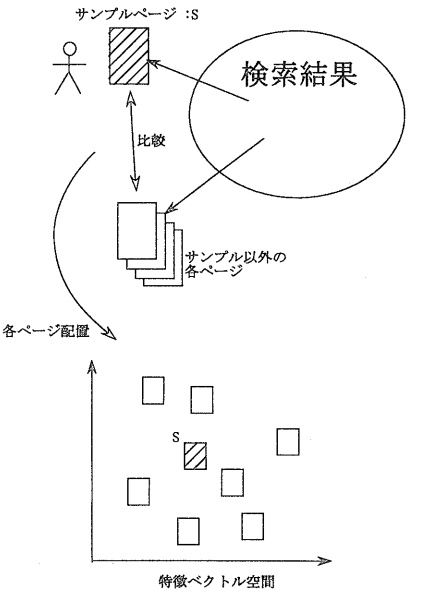


図 1　Web ページの特徴ベクトル空間への配置
Fig. 1　Arrangement of Web pages in feature vector space

- 修正された質問ベクトル = (もとの質問ベクトル) + (選択されたページのベクトルー選択されなかったページのベクトル)

という形で定義される．この考えは例えば，検索エンジン google[1] の Similar pages という機能に適用されている．厳密には google Similar pages では質問ベクトルの修正は

- 修正された質問ベクトル = (選択されたページのベクトル)

という形で定義される．ここではベクトルと書いたが，実際は数語のキーワードであると思われる．

我々が行ってきた研究では，まず何らかの検索キーワードで検索し，得られたページ群を，選択されたページの特徴ベクトルを用いて再スコアリングする．ここでは質問ベクトルの修正を行うのではなく，検索結果のスコアリングに使用される各ページの特徴ベクトルを作り直すのである．これは厳密には従来の適合フィードバックと異なるかもしれないが，ユーザの応答に応じて検索結果を再構成するという Web 検索における適合フィードバックの基本的な考えは同じなので，本研究では適合フィードバックと呼ぶことにする．

3. 意味的な適合フィードバックによる検索結果の再構成

従来の適合フィードバックは，あるページをユーザが実際に見て判定するのだが，その際のページの判定の評価は「良い」あるいは「悪い」の2種類であった．そしてユーザが判定したページの特徴ベクトルを利用して検索エンジンの検索結果を並び替え，あるいは分類する．適合フィードバックを利用することによってユーザの要求を検索結果に反映させることが可能になる．

しかし，従来の適合フィードバックでは十分にユーザの意図を汲み取ることは難しいと我々は考えた．そこで我々は以下のようなことを考え，以下のようなシステムを開発した．

- ユーザの要求を的確にシステムに伝えるために，かつユーザの応答をシステムに渡すための労力を最小限にするために，ユーザに情報の過不足を重視した3段階評価でページを評価させる．
- 3段階評価に関して各段階の距離をユーザの評価に呼応して有機的に変動させる．

以上の技術を用いることにより，よりユーザの要求に応じた検索結果の再ランキングが可能になった．なお，3段階評価というのはページの以下のような3段階評価を想定している．

(1) 完全にユーザの要求に当てはまったページ
(2) 部分的にはユーザの要求に当てはまるが，完全に当てはまっているわけではないページ
(3) ユーザの要求に全く当てはまらないページ

これによって，システムはユーザの要求を柔軟に理解し，検索結果をユーザの要求に応じたものに並べ替えることができる．

しかし，このシステムでは次のような問題点があった．

- システムは，部分的にはユーザの要求に当てはまるが，完全に当てはまっているわけではないページをユーザに教えてもらうが，ここでシステムはユーザの要求に当てはまっている部分，あるいはユーザの要求に当てはまっていない部分を把握することは可能だが，ユーザの要求に当てはまっていない部分が，どう当てはまっていないのかはシステムは理解できない．例えば，あるページが大体要求を満たしているがもう少しここの部分についての詳しい情報が欲しい場合と，あるページが大体要求を満たしているがもう少しここの部分については簡単な説明だけでいいという場合との区別はこのシステムにはできない．

そこで，ここでは，単に「良い」または「悪い」の二元論的な評価に基づいて，サンプルページに類似したページを集めるのではなくて，サンプルページと実際に欲しいページの違いを評価として与えることによって，そのような差異を持つページを探してくれるような，意味的な適合フィードバック機構について考える．

具体的な手法としては，検索結果内の各ページの単語情報をそのページを表す特徴量とみなし，サンプルページと判定するページのお互いの単語数を比較したり，お互いのページ内の単語の共起度を測定することによって，サンプルページに対するそのページの相対的な情報量や話題の範囲を測定する．そしてそれらに基づく特徴ベクトル空間にサンプルページを中心として各ページを配置し，その特徴ベクトル空間を用いて，検索結果の再ランキングを行う手法をとる．この特徴ベクトル空間はユーザの応答により動的に変化する．以上のようにして、ユーザが自分の要求するページを獲得する際の労力を軽減する。

本システムの構成を図2に示す．ここでは意味的な適合フィードバックによる検索結果の再構成について述べる．

3.1 フィードバックの種類

ユーザが必要とする情報を獲得する際の労力を軽減するために意味的な適合フィードバックを行うのだが，ページの情報量には大きく2種類のパターンがある．

(1) ある話題に関しての情報量が多い/少ない
(2) ページの中に書かれてある話題の数が多い/少ない

適合フィードバックを行う際には次のようになる．

(1) サンプルページと比較してもっとその話題についての詳細なページが欲しい．
(2) サンプルページと比較してもっとその話題についての簡潔なページが欲しい．
(3) サンプルページの話題に関連した話題が付加されてあるページ，すなわち話題の範囲が広がっているページが欲しい．
(4) サンプルページの複数の話題のうちの一部の話題が削除されてあるページ，すなわち話題の範囲が狭まっているページが欲しい．

例えば，図3の（1）はあるサッカーチームのページに対して，そのチームの情報がより詳細に書かれてあるページ，(2）はあるサッカーチームのページに対して，そのチームの情報がもっと簡潔にまとめられているページ，(3）はあるサッカーチームのページに対して，他のチームの情報も書かれてあるページ，(4）は

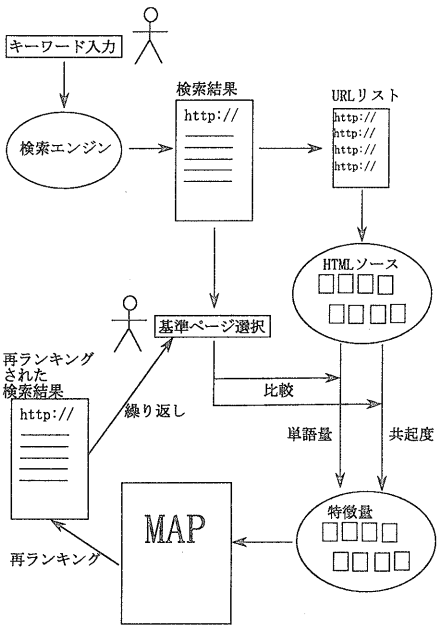


図 2　システムの構成
Fig. 2　System Architecture

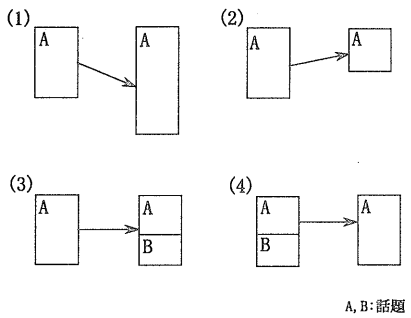


図 3　ページ間の関係
Fig. 3　Relation between pages

ある複数のサッカーチームについて書かれてあるページに対して，その中のいくつかのチームだけについて書かれてあるページなどである．

3.2　特徴ベクトル空間生成

適合フィードバックする際のユーザの要求としては

(1)　サンプルページと比較してもっとその話題についての詳細なページが欲しい．

(2)　サンプルページと比較してもっとその話題についての簡潔なページが欲しい．

(3)　サンプルページの話題に関連した話題が付加されてあるページが欲しい．

(4)　サンプルページの複数の話題のある話題が削除されてあるページが欲しい．

の 4 種類になると上で述べたが，この 4 種類の要求のうち上 2 つと下 2 つはそれぞれお互いに矛盾するので，それらを同時にみたすような要求がユーザからだされることはない．しかし，上 2 つのうちのどちらかと，下 2 つのうちどちらかの 2 つを同時に満たすページは存在し，そのようなページが欲しいとユーザが思っている時の適合フィードバックまで考えると適合フィードバックの種類は上の 4 つに加え，

1.　サンプルページと比較してもっとその話題についての詳細なページでかつ，サンプルページの話題に関連した話題が付加されてあるページが欲しい．

2.　サンプルページと比較してもっとその話題についての詳細なページでかつ，サンプルページの複数の話題のある話題が削除されてあるページが欲しい．

3.　サンプルページと比較してもっとその話題についての簡潔なページでかつ，サンプルページの話題に関連した話題が付加されてあるページが欲しい．

4.　サンプルページと比較してもっとその話題についての簡潔なページでかつ，サンプルページの複数の話題のある話題が削除されてあるページが欲しい．

の 4 つが加わり，合計 8 種類の適合フィードバックを行うことになる．ユーザにどのようなフィードバックを行いたいかシステムが聞く時にいちいち 8 種類の条件を出すのは、システムとして非常に使い勝手が悪い．

そこで，本研究では，縦軸にページの総単語数，横軸に話題のばらつきをとった特徴ベクトル空間を生成し，サンプルページを中心に検索結果の各ページを配置していく．それにより 8 種類の適合フィードバックを行うことが可能になる．8 種類の適合フィードバックがマップのどの部分に該当するか図 4 で示す．かっこ数字は基本的なもの 4 つであり，数字はその 4 つの派生である．

3.3　ページの総単語数

あるページの情報量を判定するために，そのページの総単語数を利用する．もとのページと比較して，総単語数の多いページはより情報量の多いページと考える．逆にサンプルページと比較して，総単語数の少ないページはより情報量の少ないページと考える．ここ

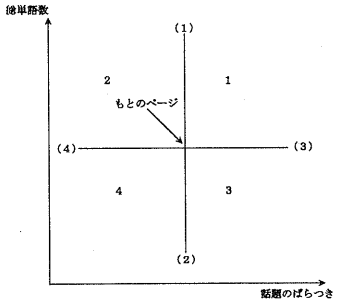


図 4　特徴ベクトル空間
Fig. 4　Feature vector space

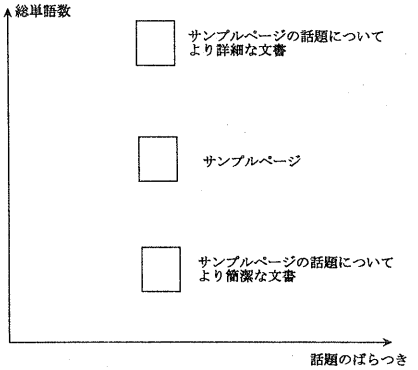


図 5　総単語数による分類
Fig. 5　A classification by word count

での総単語数というのは，同じ単語を重複して数えたときのページの総単語数のことである。例えばある単語がそのページに N 回出たとすると N 回総単語数が増える．単語の総種類（ボキャブラリー）ではなく単語の総出現回数で判定するのは次のような理由からである．

- 総単語出現回数からそのページの文章量がわかる．複数のページの情報量を比較するのにページの文章量は重要な手がかりになる．ボキャブラリーではそのページの文章量ではなくそのページの多様性が分かるが，単語の出現回数を無視することになるので，そのページの文章量がわからない．

総単語数を調べることによって，あるページに対して，そのページに書かれてある話題についてのより詳細な情報を保有しているページ，あるいは，あるページに対して，そのページに書かれてある話題についてのより簡略な情報を保有しているページを発見することができる．図 5 は情報量の詳細/簡潔なページが特徴ベクトル空間のどのあたりに位置するかを示している．

3.4　単語共起度による話題のばらつきの測定

サンプルページの話題に対して，そのページがサンプルページの話題に関連した他の話題も持っているかどうか，また，サンプルページの話題に対して，そのページがもとのページの話題の一部のみを持っているページかどうかを調べるために，もとのページの単語と判定するページの単語の共起度を利用する．

3.4.1　共　起　度

まず，共起度について説明する．共起度とは普通，次のような式で表される．ある集合内の単語 a が存在するページ数を $C(a)$，ある母集合内の単語 b が存在するページ数を $C(b)$，ある母集合内の単語 a と単語 b が存在するページ数を $C(a,b)$ とすると，単語 a と単語 b との共起度 $h(a,b)$ は，

$$h(a,b) = \frac{C(a,b)}{C(a)} \frac{C(a,b)}{C(b)} \tag{1}$$

となる．この共起度を使って検索結果の各ページのサンプルページに対する話題の範囲を求めて数値化し，その数値からユーザの要求しているページを発見するのである．

ここで，共起度について詳細に考察する．共起度を次のように定義した場合について考察する．

$$h1(a,b) = \frac{C(a,b)}{C(a)} \tag{2}$$

$$h2(a,b) = \frac{C(a,b)}{C(b)} \tag{3}$$

$$h3(a,b) = \frac{C(a,b)}{C(a)} \frac{C(a,b)}{C(b)} \tag{4}$$

$h1(a,b)$ の場合，単語 b に対して従属的な単語 a について，より高い値が出る．例えば「J リーグ」に対しての J リーグのチームのひとつである「ジェフ」の場合などである．この共起度 $h1(a,b)$ を採用した場合，サンプルページに比べてより従属的な話題が付加，あるいは削除されているページがより発見されやすくなる．

$h2(a,b)$ の場合，単語 b に対して支配的な単語 a について，より高い値が出る．例えば「ジェフ」に対しての「J リーグ」の場合などである．この共起度 $h2(a,b)$ を採用した場合，サンプルページに比べてより支配的な話題が付加，あるいは削除されているページがより発見されやすくなる．

$h3(a,b)$ の場合，単語 b に対して同等な単語 a について，より高い値が出る．例えば「ジェフ」に対しての同じ J リーグのチームのひとつである「マリノス」の場合などである．この共起度 $h3(a,b)$ を採用した場

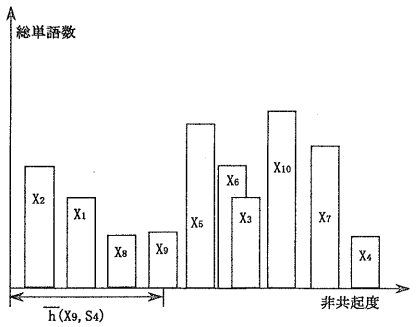


図 6 話題のばらつき
Fig. 6 Dispersion of topics

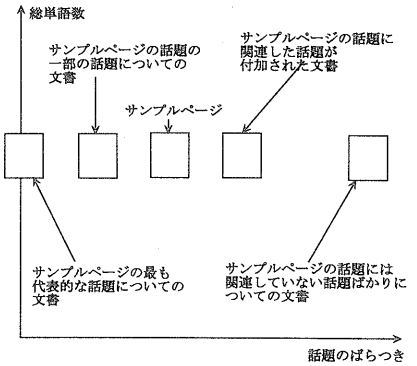


図 7 話題のばらつきによる分類
Fig. 7 A classification by a dispersion of topic

合，サンプルページに比べて同等な話題が付加，あるいは削除されているページがより発見されやすくなる．

3.4.2 話題のばらつきの測定

サンプルページに対する判定するページの話題のばらつきを測定する手法を述べる．まず，判定するページの代表的な単語を獲得する．また，サンプルページの代表的な単語を獲得する．判定するページを代表する単語 $X_i(i=1,2,\cdots)$ とサンプルページを代表する単語 $S_j(j=1,2,\cdots)$ の非共起度 $\overline{h}(X_i,S_j)=1-h(X_i,S_j)$ を調べる．それをグラフにすると図 6 のようになる．ただし，図 6 はあるひとつの単語 S_4 についてのグラフである．

図 6 について説明する．横軸は非共起度，縦軸は単語数である．$X_i(i=1,2,\cdots)$，$S_j(j=1,2,\cdots)$ はそれぞれ判定するページを代表する単語、サンプルページを代表する単語である．$\overline{h}(X_i,S_j)=1-h(X_i,S_j)$ は判定するページ内の単語 X_i とサンプルページを代表する単語 S_j の非共起度である．棒グラフの高さは単語 X_i が判定するページに出現した回数である．このグラフを利用して，サンプルページと判定するページの話題のばらつきを測定する．

判定するページのサンプルページに対する話題のばらつきは以下のように定義する．サンプルページを P_S，判定するページを P_X，サンプルページを代表する単語を S_j，単語 X_i 等が判定するページ P_X に出現した回数を N_i，判定するページ，サンプルページのボキャブラリー数をそれぞれ T_X，T_S，判定するページのサンプルページに対する話題のばらつき $B(P_X,P_S)$ は，

$$B(P_X,P_S)=\sum_{i=1}^{T_X}\sum_{j=1}^{T_S}\frac{N_i\overline{h}(X_i,S_j)}{\sum_{k=1}^{T_X}N_kT_S} \qquad (5)$$

のようになる．このようにして，話題のばらつきを測定し，マッピングすると，図 7 のようになる．

サンプルページと比べて，特徴ベクトル空間の右側にあるページは，サンプルページの話題に関連した話題が付加されたページであり，逆に特徴ベクトル空間の左側にあるページは，もとのページの話題の一部の話題についてのページである．また縦軸上にあるページはもとのページに書かれていた複数の話題のうち，もっとも代表的ないくつかの話題についてのみ書かれてあるページである．

ここで問題なのは，サンプルページの右側に位置した複数のページのうち，全てのページが，サンプルページの話題に関連した話題が付加されたページであるとは限らないことである．話題のばらつきが大きいページは，サンプルページの話題とは関係ない話題について書かれてあるページである可能性が高い．

4. 予備実験とその評価

ここでは，上で挙げたシステムの予備実験を行い，その結果についての考察を行う．今回の実験では本論で挙げた式を一部簡略化して使用している．よってその簡略化した式についてまず述べる．

4.1 総単語数の計算

まず，ページの単語数を測定するのだが，これについては以下の点が簡略化されている．

(1) ページ内の名詞以外の単語は全て無視する．
(2) ページ内の日本語以外の言語は全て無視する．ただし，これは検索結果を集める際の検索キーワードが日本語のみのときに限る．

まず，名詞以外の単語を無視することについては，一般に名詞以外の品詞を特徴としているようなページは

めったにないということと，たとえ存在したとしてもほとんどのページにとって，名詞以外の品詞を特徴量として考えることは無意味であることが理由にあげられる．次に，日本語以外の言語を無視することについては，日本語を検索キーワードとして代入した時の検索結果では，日本語以外の単語はほとんど特徴量としては無意味な単語がほとんどであり，また一部の特徴量として有益な単語についても，その単語の特徴量を付加することによって大きく結果が変わるわけではないので今回は無視した．

4.2 話題のばらつきの計算

次に，話題のばらつきの計算だが，これについては以下の点が簡略化されている．

(1) 共起度については式 (2) を採用する．

(2) ある単語 a に対してのある単語 b の共起度は本来ならば，ある母集団内の単語 a が存在するページ数を $C(a)$，単語 a と単語 b が存在するページ数を $C(a,b)$ とすると，単語 a と単語 b との共起度は $h1(a,b)$ は $h1(a,b) = \frac{C(a,b)}{C(a)}$ のようになるが，今回は母集団をある検索キーワードによって出された検索結果 400 件としている．

(3) サンプルページに対する判定するページの話題のばらつきを測定する際，お互いのページの単語間の非共起度を計算するのだが，その際，判定するページの単語は頻出上位 10 単語に限定し，サンプルページについては最頻出単語しか利用していない．

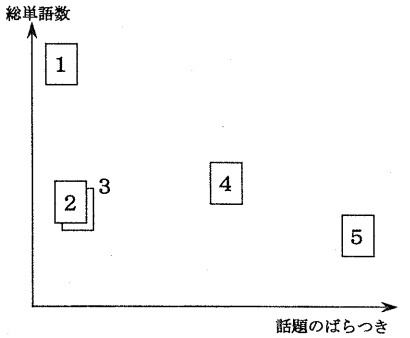
よってここでは，判定するページのサンプルページに対する話題のばらつきは以下のように定義する．サンプルページを P_S，判定するページを P_X，もとのページの最頻出単語を $W(P_S)$，単語 $X_i(i = 1, 2, \cdots, 10)$ が判定するページ P_X に出現した回数を N_i とすると，判定するページのサンプルページに対する話題のばらつき $B(P_X, P_S)$ は，

$$B(P_X, P_S) = \sum_{i=1}^{10} N_i \frac{\overline{h1}(X_i, W(P_S))}{\sum_{k=1}^{10} N_k} \qquad (6)$$

のようになる．

以上の簡略化した式を使用して，予備実験を行った．なお，Web ページの単語情報は形態素解析ツール『茶筌 (Chasen)』[2)3)4)] を利用して得られたものであり，ページの全単語のうち名詞のみを利用している．また，検索結果をとるための検索エンジンは goo[5)] を利用している．

今回の実験では検索結果の中からページ間の関係がわかりやすいものを選び出し，数値を計算したところ



もとのページ（ページ2）

図 8 実験結果

Fig. 8 An experiment result

表 1 のようになり，(ただし，話題のばらつきはページ 2 に対する話題のばらつきである) その関係どおりの特徴ベクトル空間を生成するかどうか確認したところ図 8 のようになった．ページ 1 はあるサッカー J リーグの試合の詳細な情報が書かれてあるページ，ページ 2，3 は同じ試合の簡潔な情報が書かれてあるページ，ページ 4 は J リーグについてのページ，ページ 5 は全くサッカーに関係ないページである．図 8 より，ページ 2 を中心とした関係が現れていることがわかる．

Target	総単語数	話題のばらつき
ページ 1	285	0.0758
ページ 2	69	0.1021
ページ 3	65	0.1105
ページ 4	77	0.695
ページ 5	43	0.9857

表 1 実験結果

Table 1 An experiment result

また，検索結果の上位 200 件について，ある特定のサンプルを決める．そして人間が検索結果の各ページ実際に見て，サンプルページと比べての差異からそのページがどのようなページか評価し，その評価とシステムが出した評価と一致するか調べた．その結果を表 2 に示す．

よって，表 2 の結果より，検索結果の各ページのサンプルページに対する差異を測定することによって，ユーザの細微な要求に応えることができるような意味的適合フィードバックが上手く働くことがわかる．

	精度
上位 100 件	0.73
上位 200 件	0.84

表 2　実験結果 2
Table 2　An experiment result 2

5. おわりに

本研究ではユーザの要求を満たすことができるように,「良い」または「悪い」の二元論的な評価に基づいて,「良い」と評価されたサンプルページに類似したページを集めるだけの従来の適合フィードバックではなく，ユーザがサンプルページと実際に欲しいページの違いを評価として与えることによって，そのような差異を持つページを探してくれるような，意味的な適合フィードバック機構を提案した.

またその実現方法としては検索結果内の各ページの単語情報をそのページを表す特徴量とみなし，あるページと判定するページのお互いの単語数を比較したり，お互いのページ内の単語の共起度を測定することによって，そのページの相対的な情報量や話題の広がりを測定する．そして，それらに基づく特徴ベクトル空間にサンプルページを中心として各ページを配置し，その特徴ベクトル空間を用いて，検索結果の再ランキングを行う方法をとった．また, この特徴ベクトル空間における各ページの配置はユーザの応答により動的に変化させた.

これによってユーザの意図を汲み取り，ユーザの要求するページを的確に発見することが可能になった.

また今後の課題として,

- 本研究では，ページの単語情報を利用したが，単語情報の中でも特に単語の頻度や単語間の共起度にのみ注目した．単語情報には他にも単語の位置情報というものがある．位置情報を利用すればさらに精度のいい適合フィードバックが可能になるはずであり，位置情報を考慮した意味的適合フィードバック機構の構築
- 話題のばらつきを調べる際に，サンプルページの話題に関連した話題を付加しているページと，全く関係ない話題についてのみ書かれてあるページとの区別化
- サンプルページを複数用意した場合、サンプルページ各々についてユーザの評価が異なることがある。例えば「ページ A より詳細なページでページ B より話題の広がりが狭いページが欲しい」などである。そのようなときの意味的適合フィードバックについての考察
- 意味的適合フィードバックの収束

などが挙げられる.

謝辞　本研究の一部は，日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」(プロジェクト番号 JSPS-RFTF97P00501) によっております．ここに記して謝意を表すものとします.

参考文献

1) google, http://www.google.com/
2) 茶筌ホームページ『茶室』, http://cl.aist-nara.ac.jp/lab/nlt/chasen/
3) 松本裕治 北内啓 山下達雄 平野善隆 松田寛 浅原正幸, 日本語形態素解析システム『茶筌』 version 2.0 使用説明書 第 2 版, NAIST Technical Report, NAIST-IS-TR99012, December 1999.
4) 山下達雄, パトリシア木を用いた形態素解析のための辞書検索, ChaSen Technical Report, CTR-1, 1996.
5) goo, http://www.goo.ne.jp/