

## 動的シソーラスを用いた連想検索

巖寺俊哲 木本晴夫

NTT情報通信処理研究所

文書検索システムでは、利用者の検索意図を明確にすることが重要である。本稿では、利用者の検索意図に最も適合した文書を状況に応じて高精度で検索する方法として、動的シソーラスを導入し、これを利用して検索意図を明確化した最適なキーワード集合である連想キーワードを生成する方法を述べる。さらに、利用者が検索を意図した文書に付与されているキーワードに対する連想キーワードの適切性をキーワード再現率/適合率により評価する。また、連想キーワードを検索に利用し、文書検索精度を文書再現率/適合率により評価する。これによって連想検索法により文書検索精度の向上を図ることが可能であることを示す。

### Associated information retrieval using Dynamic-thesaurus

Toshiaki Iwadera Kimoto Haruo

NTT Communications and Information Processing Laboratories

1-2356, Take, Yokosuka-Shi, Kanagawa 238-03, Japan

For a document retrieval system, it is important to disambiguate what user's queries mean. This paper presents a method to retrieve documents which user needs efficiently according to the situation in which user is placed. Dynamic-thesaurus is introduced to disambiguate user's queries, and then generating Associated-keyword that is a set of keywords relevant to user's queries is discussed. In addition, its relevance is evaluated by keyword recall/precision factor, and results of using Associated keyword are evaluated by document recall/precision factor. Some experiments proved "Associated information retrieval" approach to work well.

## 1. はじめに

近年、計算機技術、通信技術のめざましい進歩に伴い、OA化やデータベースシステムの構築が進み、また、コンピュータネットワークが一般に普及している。このため、電子化ファイルや大容量データベースが、個人で簡単に利用できるようになってきている。

しかし、これらのシステムでは、文書蓄積時、文書検索時にそれぞれ問題点が存在する。

文書蓄積時の問題点としては、①大量文書の分類作業が煩雑である、②文書へのキーワード付与作業が煩雑である、等がある。

これらの問題点の解決をめざして、①に対しては、文書自動分類の研究[1]が、②に対しては、キーワード自動抽出の研究[2]が行われている。文書検索時には、① unnecessaryな文書まで検索されてしまう、②必要な文書が検索されない、等の検索誤差が生じるという問題点がある。

この問題は、現在の検索技術が、相変わらず、検索意図を表現するものとして利用者が思い付いた言葉を検索用のキーワードとして入力し、このキーワードを使ったAND/OR検索の域を脱しておらず、利用者の検索意図を明確に表現し、システムに伝えることが、困難なことによって生じる。

従って、文書検索システムには、利用者のあいまいな検索要求を明確にする機能が必要である。この機能を実現するためのアプローチとして、自然言語理解の手法を利用したもの[3]やハイパーテキストの考え方を利用したもの[4]の研究が進められている。我々は、利用者の検索意図に最もよく適合した文書を、状況に応じて、容易に、かつ、高精度で検索することをめざして連想検索技術の研究を進めている。

今回は、利用者の指定した情報を使用してシソーラス上のノードに重みを付与したり、リンクを生成することによって動的に変化させ、このシソーラスを利用して利用者の入力したキーワードに対して、検索要求を明確にするためのキーワードを補完する試みを報告する。また、今回用いた方法の新聞記事を対象にした実験による評価結果を示す。

さらに、今回用いた方法によって検索精度向上を図ることが可能であることを示す。

## 2. 連想検索の概要

### 2.1 キーワード検索の問題点

従来の文書検索システムでは、利用者は、システムへの検索要求を単一、または、複数のキーワードから構成されるキーワード検索式として与える。以下、システムへ検索要求として与えられるキーワード検索式を入力キーワードと呼ぶ。

利用者は、この入力キーワードによって、検索意図を表現する。しかし、システムにとっては、必ずしも利用者の検索意図が明確ではない。

このため、利用者は、入力キーワードによって検索された文書を見て、入力キーワードを変更して、さらに検索を繰り返す必要があった。

この入力キーワードがシステムにとって不明確である要因には、次のようなものがある。

- (1) 目的とする文書に付与されていないキーワードの使用
- (2) 検索範囲を限定するのに不十分なキーワードの使用
- (3) 一般的な意味のキーワードの使用
- (4) 検索範囲の限定過剰なキーワードの使用
- (5) キーワードの不足、特殊な意味のキーワー

ドの使用)

このような要因を含んだ利用者の入力キーワードから利用者の検索意図を明確に定めるためには、次のような問題を解決する必要がある。

(1) に対しては、入力キーワードを利用者の検索意図に沿って別のキーワードに置き換えることが必要である。また、(2)、(3)に対しては、入力キーワードに直接表現されていない情報を検索意図に沿って補完する必要がある。

利用者の入力キーワードに情報を補完するアプローチとして、次のような研究が行われている。

(1) シソーラスやユーザ辞書、類義語辞書を使用したもの[5, 6, 7, 8, 9]

(2) 知識工学的手法を使用したもの[10]

(1)のアプローチは、利用者が入力するであろうキーワードに関連のある言葉をシソーラスや辞書としてあらかじめ登録しておき、検索時にこれらのシソーラスや辞書を使って入力キーワードに情報を補完するものである。

(2)のアプローチは、人間があるキーワードから別のキーワードを連想する知識等をあらかじめ用意しておき、この知識を使って入力キーワードに情報を補完して検索を行うものである。

これらのアプローチでは、あらかじめ想定した検索時の状況や視点においては、入力キーワードへの最適な情報の補完が可能である。しかし、状況が変化したり、あらかじめ想定したとは異なった視点から検索をしようとした場合、最適な情報を補完できない。これは、蓄積されている情報を有効に利用しようとしたとき、無視できない問題である。

状況や視点の変化に応じて最適な検索結果を得るために、我々は、次のようなアプローチを採用した。  
・利用者の入力キーワードに対して、検索時に最適な情報を補完するキーワードを自動的に生成し、このキーワードを用いて検索をする。

・このとき、検索時の状況に応じて動的に変化するシソーラス上のリンクとノードをキーワードの生成に利用する。

上記のアプローチを採用することによって後述するように、入力キーワードのみによる文書検索結果では、再現率73%、適合率47%であったものが、本アプローチを用いることで再現率91%、適合率91%と著しく向上した実験結果も得られている。また、この結果は、上記(1)のアプローチの1つである類義語辞書による検索要求の類義語展開手法[9]による検索結果(再現率91%、適合率12%)を上回っている。

このアプローチを採用するにあたって以下に述べる概念を導入した。

#### (1) 環境情報の導入

環境情報は、利用者の検索時の状況や利用者の視点を示す情報である。この環境情報を文書としてシステムへ提供することによって、システムに検索時の状況や利用者の視点を知らせることを可能にする。

#### (2) 動的シソーラスの導入

動的シソーラスは、環境情報を利用することによって状況に応じて変化するシソーラスであり、利用者毎の一種の言葉の体系である。

この動的シソーラスを使用することによって、入力キーワードに状況や視点に応じた最適な情報の補完を可能にする。さらに、この動的シソーラスと検索結果文書を照合することによって結果文書を順位付けして出力することを可能にする。

#### (3) 連想キーワードの導入

連想キーワードは、利用者による入力キーワードに対して動的シソーラスを用いてキーワードを生成して必要な情報を補完した最適キーワード集合であり、利

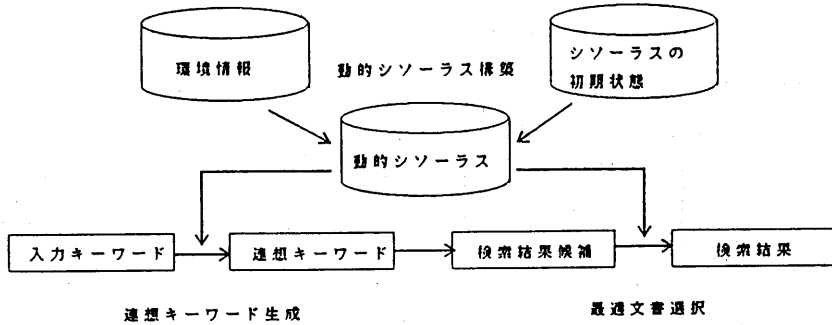


図1 連想検索方式の概要

利用者の検索要求意図を表現している。

連想キーワードを検索に使用することで、従来からのキーワード検索と同様の手法で、かつ、高精度の検索を行うことが可能になる。

## 2.2 連想検索方式

図1に、我々が連想検索と呼んでいる方式の概要を示す。四角いボックス内に処理過程で現れるデータを記してあり、矢印は、処理の方向を示す。

この連想検索方式の処理の流れの概要を以下に述べる。

まず、①検索時の状況を表す情報、具体的には、利用者によって提供される利用者の検索意図に合致した文書のサンプルから環境情報を抽出する。②これを利用してあらかじめ与えられたシソーラスを動的に変化させる。③このシソーラス(動的シソーラス)を用いて入力キーワードに対して情報を捕完したキーワード集合(連想キーワード)を生成する。④これを利用して文書DBを検索し、⑤検索された文書を再度この動的シソーラスを使用して順位付けする。

以下で、今回報告する環境情報、動的シソーラスと連想キーワードの生成アルゴリズムについて述べる。

### (1) 環境情報

環境情報は、利用者の検索意図に合致した文書(適合文書)のサンプルである文書(サンプル文書)の集合から抽出する。

図2に、環境情報の構成を示す。環境情報は、キーワード情報とキーワード間の関係を表す関係情報から構成されている。キーワード情報は、キーワードそのものとそのキーワードのサンプル文書中での重要度とで構成されている。関係情報は、キーワード間の関係の種類とその関係の強さを表す関連度で構成されている。

次に、環境情報の抽出の手順について述べる。

#### ① キーワードの抽出と重要度の算出

各サンプル文書毎にキーワードを抽出し、各キーワードのその文書中での重要度を算出する。この重要度は、文書中でのキーワードの出現頻度、出現位置、あるいは、構文情報等を用いて算出される。この処理には、我々が開発したキーワード自動抽出システム Indexer [11] を利用している。

#### ② キーワード間関係の種類決定と関連度の算出

①の結果を利用してキーワード間の関係の種類決定と関連度の算出を行う。現在、関係の種類としては、共起関係のみを使用している。この共起関係は、同一

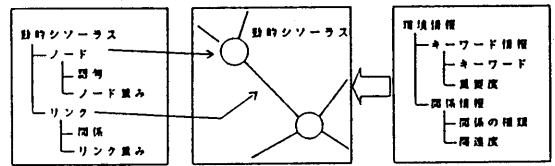


図2 環境情報と動的シソーラスの構成

のサンプル文書中に現れる異なるキーワード間にある関係と定義している。また、関連度は、今回使用していない。

### (2) 動的シソーラス

図2に動的シソーラスの構成を示す。

動的シソーラスは、ノードとノード間の関係を表すリンクを持つネットワークである。

ノードは、語句とノード重みから構成されている。語句は、環境情報として抽出されたキーワードに対応している。また、ノード重みは、環境情報全体中でのキーワードの重要度に対応する。これは、環境情報をもとに算出され、ノードに付与されている数値である。リンクは、関係とリンク重みから構成されている。リンクの示す関係は、次の4種である。

- ① 下位関係
- ② 上位関係
- ③ 同義関係
- ④ 共起関係

リンク重みは、2つのノード間の関係の強さをしめす(今回は、使用していない)。

動的シソーラスは、初期状態として上記の①～③の関係のみを持ち、また、利用者に依存しない典型的な言葉の関係を示す体系として与えられる。この状態では、少数のノードからなる複数のネットワークが互いに孤立して存在しており、文書中で実際に使用されているキーワード間の関係を十分に表現していない。従って、環境情報として得られる共起関係を動的シソーラス上に導入することで、動的シソーラスの状態を実際の利用者のキーワードの使用状況により適応したものに変わることが可能になる。

次に動的シソーラスの構築手順について述べる。動的シソーラスは、動的シソーラスの初期状態に対してリンクを生成し、ノード重みを付与することによって構築される。

#### ① リンクの生成

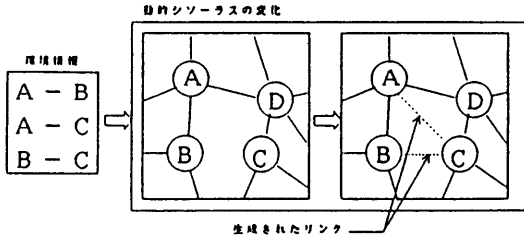


図3 環境情報中のキーワード間関係と生成されたリンクの関係

リンクの生成には、環境情報として得られた各サンプル文書毎のキーワード間の関係を利用する。図3に、環境情報中のキーワード間関係と生成されるリンクの関係を示す。図3に示すようにサンプル文書集合内のキーワード間の関係のうち動的シソーラス上に存在していない関係のみについてリンクを生成する。このリンクは、前述した同一文書内のキーワードの共起関係を示すために生成されるリンクなので共起性リンクと呼ぶ。

② ノード重みの付与

ノード重みは、環境情報として抽出された各サンプル文書中のキーワードの重要度をそのキーワードの環境情報全体中での重要度に正規化することによって算出される数値である。この数値をノード重みとして環境情報中の各キーワードに対応する動的シソーラス上の各ノードに付与する。

ノード重みの算出方法を以下に示す。

サンプル文書集合全体Dは、n個のサンプル文書Tから構成されているとする(式1)。また、Tは、Indexerの出力したm個のキーワードK<sub>i</sub>から構成されているとする(式2)。

$$D = \{T_i\} \quad (i=1 \dots n) \quad (1)$$

$$T = \{K_i\} \quad (i=1 \dots m) \quad (2)$$

まず、T内でのキーワードKの重要度KW(T)を式3から計算する。

$$KW(T) = \frac{2TW}{n(n+1)} \quad (m+1-j) \quad (3)$$

ここで、TWは、D内でのTの点数、jは、KのT内での順位でIndexerの出力する順位値をもとに計算した数値である。また、TWは、次のように定義する。

$$TW = \frac{DW}{n} \quad (4)$$

ここで、DWは、Dの点数である。あるノードN(=K)の各T内での重要度の計算の後、最終的なノード重みNWを計算する。

$$NW = \begin{cases} \sum_{T \in D} KW(T) : N \in D \\ \text{デフォルト値} : N \notin D \end{cases} \quad (5)$$

この計算方法は、各サンプル文書中でのキーワード重要度とキーワード数をノード重みに反映している。

(3) 連想キーワード

連想キーワードは、入力キーワードに対応する動的シソーラス上のノードを出発点として後述する最適ノード選択アルゴリズム、最適リンク選択アルゴリズムによって選択されたリンクとノードをトラバースすることによって生成される。ここで、出発点となるノードを生成開始ノード、トラバースしたリンクとノードからなるサブネットワークを生成経路、あるいは、単に経路と呼ぶ。

最適ノード選択アルゴリズムとは、トラバースする経路上に存在する各ノードを連想キーワードとして生

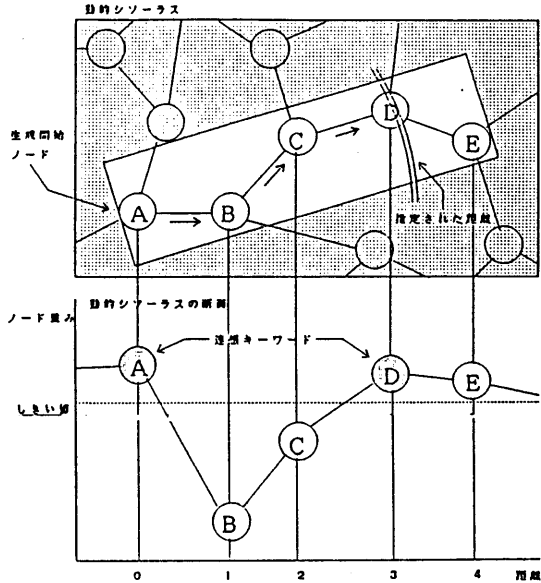


図4 連想キーワード生成アルゴリズム

成するか否かを選択するアルゴリズムである。また、最適リンク選択アルゴリズムは、あるノードからリンクで結合している他のノードへの経路(方向、関係)を選択するアルゴリズムである。

次に、連想キーワード生成アルゴリズムについて述べる。今回の実験では、リンクとノード重みの使用の有効性を確認するために、以下の手順で連想キーワードを生成している。

①生成に利用するリンクの種類とトラバースする範囲を限定する距離をあらかじめ指定する。また、生成を許すノード重みの範囲もしきい値としてあらかじめ指定する。ここで、距離とは、動的シソーラス上でのキーワード間の距離のことで、あるノードとリンクで直接結合しているノードを距離1と定義する。

②入力キーワードに対応するノード(生成開始ノード)から指定したリンクを、指定した距離にあるノードまでたどる。

③経路上に存在するノードを連想キーワードの候補とする。

④これらの連想キーワードの候補のうちノード重みが指定したしきい値以上のノードを連想キーワードとして生成する。

以上の手順の一例として図4の動的シソーラスの白抜き部分に適用したときの連想キーワードの生成過程を以下に示す。

①まず、距離を3、リンクはすべて使用すると指定する。また、適当なしきい値を指定する。生成開始ノードは、Aとする。

②ノードA、B、C、Dが連想キーワードの候補となる。

③ノード重みとしきい値の関係からA、Dが連想キーワードとして生成される。Eは、しきい値以上のノード重みを持っているが距離が3より大きいので生成されない。

3. 連想検索の実験と評価

連想検索方式の評価を下記の2つの実験によって行った。

(1) 連想キーワード生成精度評価実験

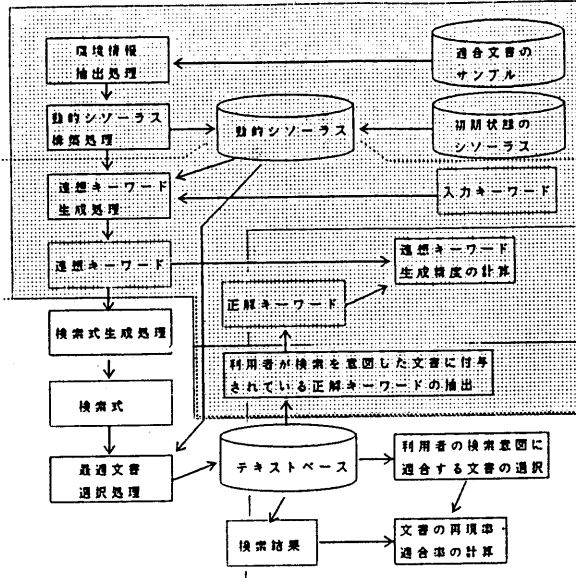


図5 実験システムの構成

(2) 検索精度評価実験

3. 1 実験システムの構成

図5に実験システムの全体構成を示す。図中の網掛けの部分が、今回の実験に使用した部分である。それ以外の部分は、連想検索方式全体の性能の評価を目的として現在開発中である。

3. 2 実験対象

実験に使用する動的シソーラスの初期状態として、ある新聞用シソーラスを選択した。これは、十分な語彙数(8000語)を持ち、また、実用に供されており、比較的使用する安定した語彙から構成されている、等の理由による。また、検索対象としては、新聞記事を選択した。これは、今回使用した新聞用シソーラスとの語彙的整合性のよさ、等の理由による。

3. 3 連想キーワード生成精度評価実験

これは、入力キーワードに対してどの程度正確な情報が補充されているかを評価する実験である。

3. 3. 1 実験方法

利用者が入力キーワードを検索システムに与えることによって手に入れようとしている文書(適合文書)にあらかじめ付与されているキーワード(正解キーワード)と連想キーワードとを比較することによって評価する。このとき、キーワード一致率を実験の指標として用いる。

キーワード一致率は、キーワード集合Aの異なるキーワード集合Bに対する一致の度合を測る指標である。これは、キーワード適合率とキーワード再現率で構成されている。これらを下記の式で定義する。

$$Kp = \frac{Nab}{Na}$$

$$Kr = \frac{Nab}{Nb}$$

ここで、

Kp: キーワード適合率

Kr: キーワード再現率

Na: あるキーワード集合A中のキーワードの数

Nab: あるキーワード集合Aに含まれていて、かつ、異なるキーワード集合Bにも含まれているキーワードの数

Nb: キーワード集合Bに含まれているキーワードの数

連想キーワード方式に使用する条件と生成精度の関係を調査するために以下の3つの条件を変化させて実験を行った。

①シソーラスの構造

②環境情報

③しきい値

実験は、一人の被験者に対して行った。また、次のような手順で実験を行った。

(1) 入力キーワードの収集

①被験者に複数文書(1回につき10記事)を読んでもらう。以後この文書を既知文書と呼ぶ。

②1日後に、各既知文書を検索するための入力キーワードを被験者に作成してもらう。この入力キーワードは、各既知文書毎に10個のキーワードで構成されている。

(2) キーワード一致率の比較

各既知文書毎に(1)で収集した入力キーワードから連想キーワードを生成する。次に正解キーワードに対する連想キーワードのキーワード一致率と正解キーワードに対する入力キーワードのキーワード一致率を比較する。

3. 3. 2 実験結果

(1) シソーラスの構造の種類と生成精度の関係  
連想キーワードの生成に使用したシソーラスを以下に示す。

A: 静的シソーラス(通常のシソーラス)  
(ノード重み、共起性リンク、ともになし)

B: 動的シソーラス  
(ノード重みあり、共起性リンクなし)

C: 動的シソーラス  
(ノード重みなし、共起性リンクあり)

D: 動的シソーラス  
(ノード重みあり、共起性リンクあり)

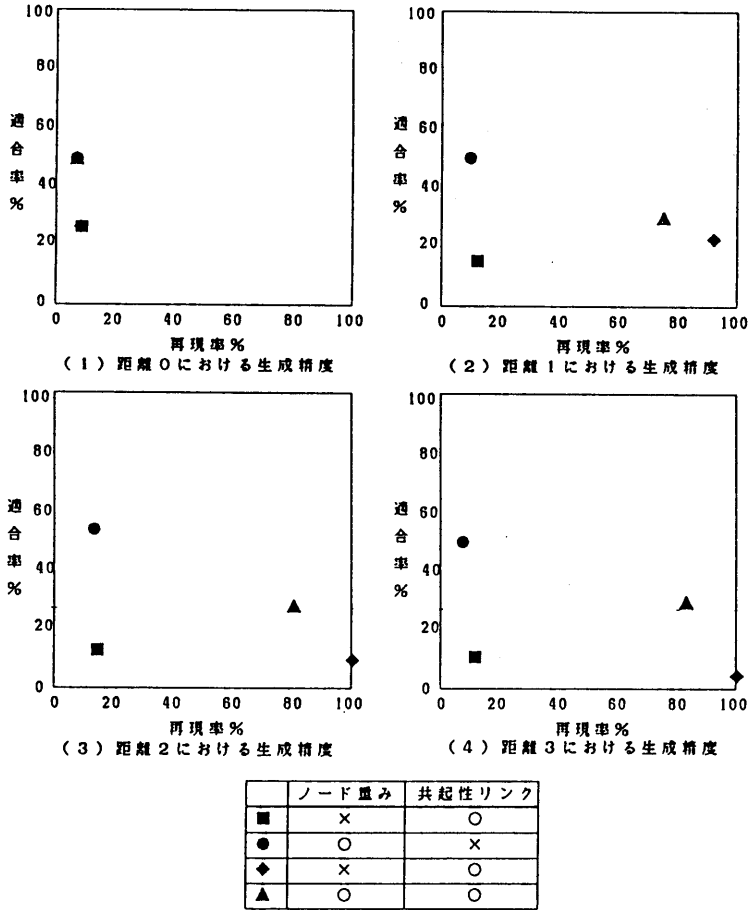


図6 シソーラスの構造の種類と生成精度の関係

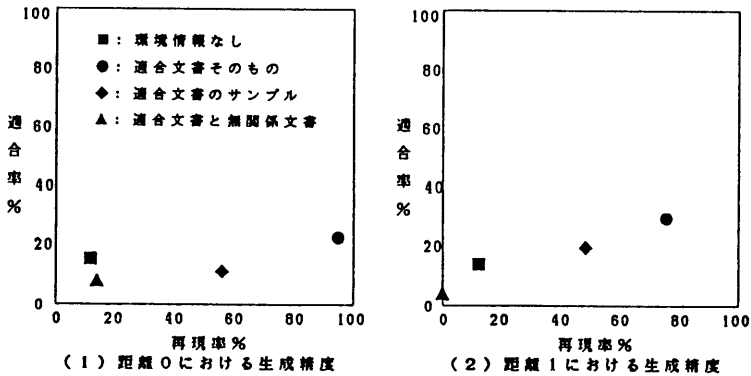


図7 環境情報と生成精度の関係

ここでは、環境情報として適合文書（「中東問題に関する5つの新聞記事」）そのものを与えている。また、ノード重みを利用する場合には、しきい値0.005以上のものを連想キーワードとして生成するように指定している。

図6に、シソーラスの構造の種類とキーワード再現

率、キーワード適合率の関係を示す。実験結果から、入力キーワードのみ（距離0のとき）の場合、約10%だった再現率が、共起性リンクを利用している動的シソーラス（C、D）を使用して生成した連想キーワードが距離1以上で、再現率70%以上の結果を得ている。

また、入力キーワードのみ（距離0のとき）の場合、24%だった適合率がノード重みを利用しているシソーラス（B、D）を利用することによって適合率約50%～29%に向上している。ノード重みを利用しているシソーラス（B、D）を比較した場合、共起性リンクを利用していないシソーラスBの方が、Dよりも適合率が高い。また、距離が大きくなっても適合率が低下していない。

以上のことから次のことが言える。  
 ・共起性リンクは、再現率向上に寄与する。  
 ・ノード重みは、適合率向上に寄与する。  
 ・共起性リンクを持つシソーラスを使用して連想キーワードを生成する場合、生成する距離が大きくなるに従って適合率が低下する。

(2) 環境情報と生成精度の関係

図7に、環境情報と生成精度の関係を示す。ここで、図7(1)は、ノード重みを利用していない動的シソーラスを使用した場合の結果であり、図7(2)は、ノード重みを利用した場合の結果である。ここで使用した環境情報は、以下の通りである。

- A: 環境情報なし
- B: 適合文書そのもの
- C: 適合文書のサンプル
- D: 適合文書と無関係文書

ここでは、適合文書とそのサンプルには、「中東問題」に関する記事を、また、適合文書と無関係な文書として「いじめ」に関する記事を使用した。

図7からノード重みの利用の有無にかかわらず、環境情報としてB、C、Dをそれぞれ与えた場合を比較したとき、B、C、Dの順でよい結果を得ていることがわかる。

以上のことから次のことが言える。  
 ・適合文書に、より関連の強い文書を環境情報として与えることによって精度の向上を図ることが可能。  
 ・適合文書に無関係な文書を環境情報として与えた場合、環境情報を利用しない場合よりもさらに、適合率が低下する。

(3) しきい値と生成精度の関係

入力キーワード、動的シソーラスを固定して連想キーワードとして生成を許すノード重みのしきい値を変化させて生成精度の変化を測定した。

しきい値と生成精度の関係を図8に示す。図8からわかるようにしきい値が、大きくなるにしたがって再現率は減少し、適合率は向上している。

3.4 検索精度評価実験

検索精度評価実験は、連想キーワードを使用して検索したとき、どの程度の正確さで適合文書を検索結果として得られるかを示す検索精度を測定する実験である。

この実験によって連想検索方式の有効性の評価ができる。

3.4.1 実験方法

この実験の評価指標には、通常の検索システムの評価に用いる再現率、適合率を使用する。

連想キーワードの生成に利用するシソーラスの構造を替えて文書検索精度の測定を行った。環境情報、しきい値は、一定である。

連想キーワードの生成には、下記のシソーラスを使用した。

- 1: 入力キーワード
- 2: 静的シソーラス（通常のシソーラス）
- 3: 動的シソーラス（共起性リンクのみ使用）
- 4: 動的シソーラス（ノード重み、共起性リンクともに使用）

3.4.2 実験結果

シソーラスの構造と文書検索精度の関係を図9に示す。図中の各点の番号は、上記のシソーラスに対応している。

実験結果から以下のことが言える。

- ・入力キーワードのみの結果に比較して静的シソーラスを利用した場合と共起性リンクのみを利用した動的シソーラスを利用した場合は、再現率は、そのままか、あるいは、向上するが、適合率は、減少している。
- ・ノード重み、共起性リンクを利用した動的シソーラスを利用した場合、再現率、適合率ともに、向上している。
- ・連想キーワード生成精度とほぼ同様の挙動を示している。

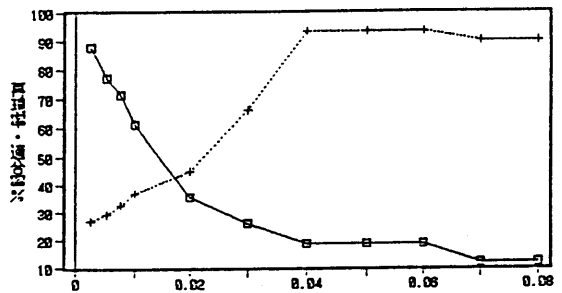


図8 しきい値と生成精度の関係

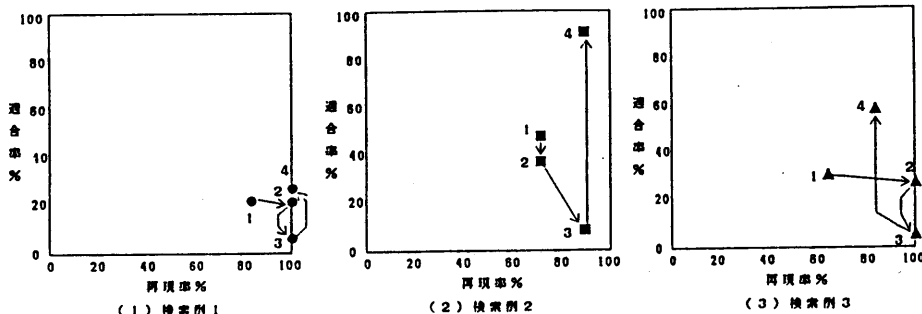


図9 シソーラスの構造と文書検索精度の関係

以上のことから、現在、検討中の本方式によって文書検索精度の向上が期待できる見通しが得られた。

#### 4. 考察

実験結果については前述した。連想キーワード生成精度評価実験、検索精度評価実験のどちらの結果からも次のことが言える。

- ① 共起性リンクは、再現率向上に寄与する。
- ② ノード重みは、適合率向上に寄与する。
- ①、②から共起性リンクは、必要なキーワードの生成に寄与し、また、ノード重みは、不必要なキーワードの生成の防止に寄与することがわかる。このことから、入力キーワードに情報を補完するために動的シソーラスを使用する我々のアプローチは、状況に応じた高精度の検索に対して有効であることがわかる。

以下で今回実験を試みた方式の問題点と今後の課題について述べる。

##### 4. 1 問題点

実験結果から明らかになった問題点を以下に述べる。

- ① 共起性リンクを連想キーワードの生成に使用した場合、動的シソーラス上の生成時にトラバースする範囲を限定する距離が大きくなるにしたがって生成される不必要なキーワードの数が増加する。これは、適合率の低下を招く。
- ② しきい値の決め方によって生成精度が左右される。
- ③ 適合文書と無関係の文書を環境情報として与えた場合、環境情報を利用しない場合より適合率が低下してしまう。

##### 4. 2 今後の課題

上記問題点の①、②は、動的シソーラスの構築法とこれを利用した連想キーワード生成アルゴリズムに関する問題点である。

今回の実験では、動的シソーラスを構築する際に、環境情報中の共起関係を利用してリンクを生成した。このため、共起はしているが、あまり関連のないキーワード間にも一様にリンクが生成されてしまった。また、今回の生成アルゴリズムでは、あらかじめ指定したしきい値と距離で、連想キーワードの生成、非生成を決定するため、ノード重みや距離によっては、必要なキーワードが生成されなかったり、不必要なキーワードが生成されたりしてしまう。

これを解決するためには、以下のことを検討する必要がある。

- ① 現在、共起関係として一括して扱っている関係を詳細化(たとえば、原因-結果関係、時間関係、位置関係、無関係)の方法
- ② ①で詳細化した関係とノード重みを有効に利用した連想キーワード生成アルゴリズム

上記問題点の③は、我々が検討を進めている連想検索方式の利用方法に関する問題点である。

問題点③は、適合文書とは無関係な文書である「いじめ」に関する記事を環境情報として与えたことによって、連想キーワードとして「いじめ」に関するキーワードも生成してしまったことに起因している。

この問題点③を別の観点からみれば、これは、非常によく環境情報を反映したキーワードの生成の仕方である、と考えられる。たとえば、「いじめ」に関して関心を持っている利用者が、日頃、「いじめ」に関する記事を環境情報として与えていたとする。これによって構築された動的シソーラスを利用して「中東問題」について検索しようとした場合、結果として「中東問

題」と「いじめ」の複合領域の結果を得ることができると可能性を示している。つまり、日頃の「いじめ」に関する関心を反映した結果が得られる可能性を示している。

これは、我々が検討を進めている連想検索方式の目的の1つである「状況に応じた検索」の実現の可能性を示している。

#### 5. おわりに

利用者の検索要求に最もよく適合した文書を状況に応じて容易に、かつ、高精度で検索することを可能にする連想検索方式について検討した。

利用者の入力キーワードに対して適切な情報の補完を行うために、新しい概念として環境情報と動的シソーラスを導入した。さらに、これらを利用して、最適なキーワード集合である連想キーワードを生成する方式を提案した。

適合文書に付与されているキーワード群に対する本方式による連想キーワードの適切性を実験によって評価した。この結果、従来の入力キーワードを直接用いる方法に比べ、適合文書に付与されているキーワードに対する検索に利用するキーワードの再現率が約60ポイント、適合率が約10ポイント向上することを示した。

さらに、連想キーワードを検索に利用することによって検索精度の向上を図ることが可能であることを実験により示した。

#### 【参考文献】

- [1] Hamill, K.A. and Zamora, A.: "The Use of Titles for Automatic Document Classification," *Journal of the American Society for Information Science*, Nov. 1980.
- [2] 木本: 「キーワード自動抽出における分野特性の利用」信学会春季全国大会、1989.
- [3] 杉山、秋山、伊吹、川崎、内田: 「自然言語理解に基づく情報検索システムIRIS」、情報処理学会自然言語処理研究会資料58-8、1986.
- [4] 原、田村、笠原: 「ハイパーメディア電子ブックにおけるトラバース主導型情報検索方式」、電子情報通信学会データ工学研究会資料88-17、1988.
- [5] 仲谷、築山: 「事例ベース型概念設計支援システムSUPPORT(2)」、情報処理学会第38全国大会、1989.
- [6] 河合、中馬、榎木: 「知識処理機構を用いた知的文書管理機能」、情報処理学会第37回全国大会、1988.
- [7] 三村、中埜、今井、榎木: 「パーソナル・マルチメディアデータベースを用いた文書ファイルシステム」、情報処理学会データベース・システム研究会資料62-5、1987.
- [8] 中村、柴田: 「表層上の手掛かりを用いた日本語テキスト検索・分類システムの試作」、情報処理学会第36回全国大会、1988.
- [9] 福永、斎藤: 「語の類義性と結合関係を考慮したテキスト検索」、電子情報学会春季全国大会、1989.
- [10] 辻、小堀、橋本: 「ソフトウェア常識集IRシステムSOCKS(2)」、情報処理学会第35回全国大会、1987.
- [11] 木本、永田、河合: 「キーワードの自動抽出と重要度評価」、研究実用化報告、第38巻、第1号、1989.