

動的リンク機能を有するハイパーテキストシステム

5L-4

小川泰嗣 森田哲也 金矢光久
(株)リコー 中央研究所

1 はじめに

最近、ユーザインタフェースの改良・柔軟な思考をサポートする方法としてハイパーテキスト(ハイパーメディアという言葉も良く用いられるがここではハイパーテキストを用いる)の研究・開発が進んでいる[Tan]。ハイパーテキストはユーザに使い易いインタフェースを提供するが、リンク情報の生成が困難・大量の文書内で自分の居場所がわからなくなるといった問題も指摘されている。これはハイパーテキストでは従来のデータベースが持つ集合操作機能・検索機能が不十分なためである。これらの問題を解決するために、ハイパーテキストにデータ検索機能を付加し、動的にリンクを生成する方式が提案されている。例えば、神戸大で開発されたTextLinkでは、リンクの出発点・到着点をSQL風に宣言的に記述することでこの問題を解決している[Hir]。

われわれはオフィスなどに氾濫する大量の文書を効率的に管理する文書検索システムのプロトタイプを開発した[Mor, Oga]。このシステムは文書の内容に基づく検索をキーワードコネクションという知識ベースにファジ理論を応用することで実現した。現在このシステムの機能強化をはかっているが、新しいシステムではユーザインタフェースとしてハイパーテキストを採用する。ハイパーテキストの使用に当たって上述の問題を解決するため、われわれは文書管理システムに用いたキーワードコネクションを動的リンク生成に応用する。本発表ではこの方式について報告する。

2 キーワードコネクションによる動的リンク

本システムの特徴であるキーワードコネクションによる動的リンクにはつぎの3種類がある。

- キーワードリンク: キーワードから文書へのリンク
- ORリンク: 文書から文書へのリンク
- ANDリンク: 文書から文書へのリンク

いずれの方式もキーワードあるいは複数のキーワードと論理演算子で構成される検索条件をもとにファジ検索[Mor]を行ない、関連のある文書にホップしていく。なお、この方式は各文書にはキーワードが付与されている。

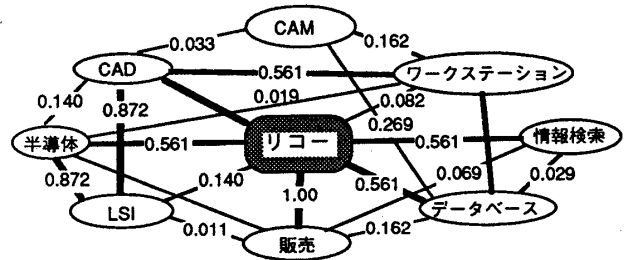


図1 キーワードコネクション

ることを前提としているが、われわれはキーワードの自動抽出法[Anw]を開発済である。

2.1 キーワードコネクションとファジ検索

まず、キーワードコネクションについて説明する。キーワードコネクションとはキーワード間の関連の情報を持つシソーラスの一種である[Mor]。通常のシソーラスとの相違点はつぎの2点である。

- 上位語・下位語といった階層構造を持たず、関連語の関係のみを表す。
- 関連の大きさを0~1の数値で表現する。(関連の大きい場合には関連度も大きな値となる。)

キーワードコネクションを図1に例示する。

我々が提案したファジ検索方式では、データベース内の全ての文書について文書確度という検索条件に対する適切さを表す数値を計算し、文書確度順に文書をユーザに提示する。文書確度の計算にキーワードコネクションを使用し、キーワードのうろ覚え等の場合にも検索洩れを無くすことが可能である。

2.2 キーワードリンク

キーワードリンクでは、あらかじめTextLinkのリンク定義言語LDL[Hir]の形式でリンクの出発点を定義する。TextLinkとの相違はリンクの到達点をLDLで定義せず、ファジ検索で求める点にある。

キーワードリンクではリンク強度(ファジ検索における文書確度に相当)は以下のように求められる。各文書のリンク強度 l_i は次式により計算される。

$$l_i = 1 - \prod_{K_k \in A_i} (1 - W_{jk})$$

A Hypertext System with a Dynamic Link Generation Mechanism

Yasushi Ogawa, Tetsuya Morita, Mitsuhsa Kanaya
Research and Development Center, RICOH Co., Ltd.

ここで、 K_j はLDLで指定されたキーワード、 A_i は*i*番目の文書に付与されているキーワードの集合、 W_{jk} は*j*番目と*k*番目のキーワード間のキーワードコネクションの値である。リンクをたどる際には、システムが全文書のリンク強度を計算・リンク強度順に文書（エントリ）をユーザに提示するので、ユーザはその中から目的の文書を探せばよい。

2.3 ORリンク

ORリンクでは、現在参照している文書に対し付与されているキーワードの集合から関連した文書への参照を生成する。現在参照している文書に対し付与されているキーワードの集合を $Q = \{K_{d_1}, \dots, K_{d_m}\}$ とすると、ORリンクはつぎの条件でファジィ検索を行なう。

$$Query = K_{d_1} \vee \dots \vee K_{d_m}$$

文書確度計算式 [Mor] にこの検索条件を当てはめると、ORリンクにおけるリンク強度計算式が得られる。

$$l_i = 1 - \prod_{K_j \in Q} \{ \prod_{K_k \in A_i} (1 - W_{jk}) \}$$

ORリンクの様子を例示したのが図2である。図2に示したように、各文書に付与されたキーワードがキーワードコネクションを介して関連付けられる。

2.4 ANDリンク

この場合、現在参照している文書に付与されているキーワードをANDで結合した検索条件が作成され、ファジィ検索が実施される。

$$Query = K_{d_1} \wedge \dots \wedge K_{d_m}$$

したがって、リンク強度計算式はつぎのようになる。

$$l_i = \prod_{K_j \in Q} \{ \prod_{K_k \in A_i} (1 - W_{jk}) \}$$

2.5 キーワードコネクションの学習

キーワードコネクションには学習機能 [Oga] が用意されており、ユーザによるキーワードの使われ方を学習しキーワード間の関連度が動的に変更される。ユーザによって学習効果が異なるため、キーワードコネクションをユーザごとに用意する。

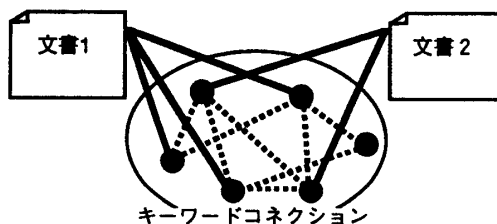


図2 キーワードコネクションによる動的リンク

3 システム構成

本システムには前節で説明したものを含め3種類のリンクを用意し、ユーザに便宜をはかる。

- 静的リンク
- LDLによる動的リンク
- キーワードコネクションによる動的リンク

静的リンクは通常のハイパーテキストで用いられているリンクであり、リンクの出発点・到達点がユーザ（ハイパーテキストの作者）によって設定される。LDLによる動的リンクはTextLinkの動的リンクである。また、ユーザ名によるリンクの管理を行なう。静的リンク・LDLによる動的リンクではリンクをパブリックとプライベートに分け、プライベートリンクはリンクを定義したユーザにしか見えない。キーワードコネクションによる動的リンクにはプライベートしか存在しない。

現在、上記の機能をWS上にインプリメント中である。文書はテキストベースであるが、グラフィックス・イメージ等の情報も扱えるようになっており、イメージスキャナが接続される。ソフトウェアはC言語で記述し、ユーザインタフェースはOSF/Motifを採用してX11上に構築する。

4 おわりに

従来のハイパーテキストシステムの問題点であるリンクの生成における問題点を解決するシステムを提案した。キーワードコネクションという知識ベースを用いて文書間のリンクを動的に生成し、関連のある文書間をユーザに合った形でホップすることが可能となる。現在、本発表で提案したキーワードコネクションに基づく動的リンク生成機能をWS上にインプリメント中である。

参考文献

- [Anw] P. Anwyl 他, "Automatic Keyword Assignment in English Documents", 情報処理学会第41回全国大会, pp. 3-87-3-88 (1989)
- [Hir] 平山伸一他, "リンク定義言語を有するハイパーテキストシステム: TextLink-III", 情報処理学会データベースシステム研究会 No. 78-7, pp. 63-70 (1990)
- [Mor] 森田哲也他, "ファジィ文書検索システム(1)", 情報処理学会第39回全国大会, pp. 1067-1068 (1989)
- [Oga] 小川泰嗣他, "ファジィ文書検索システム(2)", 情報処理学会第39回全国大会, pp. 1069-1070 (1989)
- [Tan] 田中克巳, "ハイパーメディアとは何か", Computer Today No. 38, pp. 4-12 (1990)