

5L-2

# Fish View: 概念体系を用いた Fish Eye マッチング による視点を考慮した文書整理支援ツール\*

高間 康史      石塚 満

東京大学工学部電子情報工学科

## 1 はじめに

インターネットに代表される情報環境の急速な整備・拡大により、研究や仕事などに必要となる情報を収集する過程はますます容易になりつつある。その反面、入手可能な情報量が人間の情報処理能力を越え、かえって効率が低下するという、いわゆる「情報過多 (information overflow)」が問題となってきた。この様な事態に対し、人間の情報処理能力に関する知的増幅器 (intelligent amplifier) として計算機が果たす役割は、今後ますます重要となっていくことが想定される。

我々は、インターネット等を通じて大量に収集された文書を熟読し、有効に活用するためには、各文書間の関係を整理しながら漸進的に読み進めていくことが有効であり、この過程を計算機によって効率的に支援するためには、その時点におけるユーザの興味・視点で捉えた場合の各文書間の関係を見出すことが必要であると考えた。

この様な観点から、我々は Fish Eye マッチングと呼ぶ、概念体系を利用した新しい動的ベクトル生成・マッチング機構を提案している [2, 1]。Fish Eye ベクトルは、単語間に存在する意味関係を利用して、ユーザの興味・視点を反映して動的に生成される特徴ベクトルであり、ドキュメント間の類似性、差異を多様な視点から判断するのに有効であると考えた。

本稿では Fish Eye マッチングの概要について簡単に紹介した後、開発中の文書整理支援システムおよびその文書整理過程について説明する。

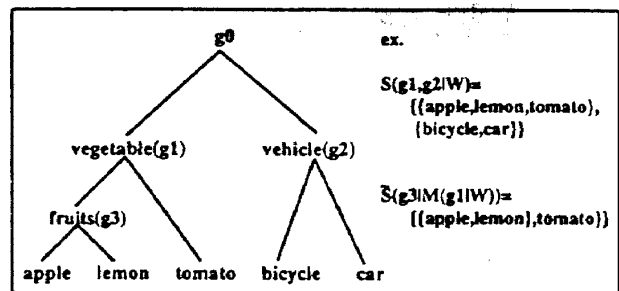


図 1: 概念構造と特徴生成の例

## 2 Fish Eye マッチング

Fish Eye マッチングでは、ドキュメント中に含まれる単語をそのまま特徴として用いた基本特徴ベクトル (Basic Feature Vector) から、ユーザの視点において同様とみなせる単語は一つの特徴に縮退 (Shrink) したり、その視点に関係のある単語のみを選択 (Magnify) したりといった操作を加えることにより、その時点に応じた特徴ベクトルを動的に生成する (図 (1))。この、ユーザの視点を反映して動的に生成される特徴ベクトルのことを Fish Eye ベクトルと呼ぶ。

本研究では、EDR 電子化辞書中の概念体系中に存在する概念を、ユーザの視点・興味を表すものとして扱い、この概念単位で特徴の縮退、選択を行う。ある概念に属する単語の集合を意味グループと呼ぶ。意味グループに関しては、ユーザの手によって興味に従い分類された、小数の正、負テキスト集合から、特徴として有効な単語を多く含んでいる意味グループを求めるアルゴリズムが提案されており、検索精度に関して有効な結果が得られている [1]。

\*FISH VIEW, a Document Organization Support Tool Using Concept Structure-Based Fish-Eye Matching.  
Yasufumi Takama, Mitsuru Ishizuka  
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan  
takama@miv.t.u-tokyo.ac.jp

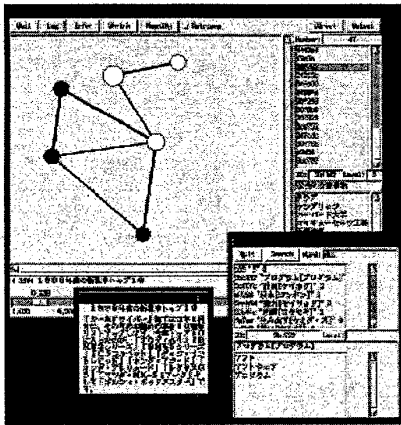


図 2: 文書整理支援システムの概観

### 3 Fish Eye マッチングを用いた文書整理プロセス

我々は、Fish Eye マッチングを基盤技術として用いた文書整理支援システムのプロトタイプを開発中である(図2)。このツールを用いて文書整理を行うプロセスについて、簡単な例をあげて説明する。

まず最初にインターネットに関する3つの記事(3中のノード0, 1, 2)を正例、それ以外の2つの記事を負例として意味グループの抽出および Fish Eye マッチングを行った結果、現在の視点において正例と関係があると思われる新たなドキュメントが検索され、それらの関係がリンクで表現された(図3)。

この図において新たに登場したドキュメントの大半はユーザの視点に関する記事であると判断されたが、ドキュメント19, 112, 187は車に関する記事であった。これらの記事が検索された理由の一つとして、Fish Eye ベクトル生成に使用された意味グループに不適切なものがあり、単語 driver と guide が「指導者」という見出しを持つ意味グループにまとめられてしまっていた。この結果、ドキュメント2は、「インターネット上でウェブサイトガイドする」ことについての記事であるため、車関係の記事と関係があると判断されたのであろう。しかし、他の意味グループとしては、「証書」や「情報処理関係の属性」といった見出しを持った、視点に関する意味グループが抽出されている。

ここで、興味は車関係の記事に移り、ドキュメント19, 112, 187, 3を正例にし、ドキュメント0, 1, 2, 4, 31, 62, 145を負例として新たに Fish Eye マッチングを行ってみた結果を図4に示す。

ドキュメント146は、ドキュメント109, 149とそれぞれ太いリンクで結ばれているが、これらは全

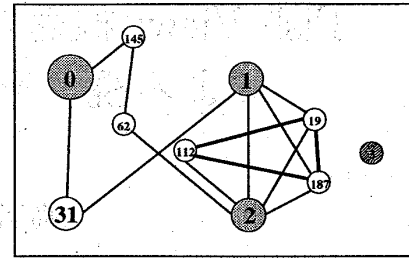


図 3: 1回目の検索終了後

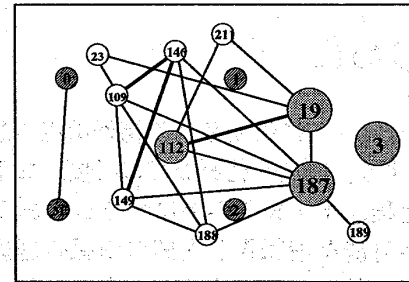


図 4: 2回目の検索終了後

て、「新しい乗物(146)」や「全地形車(149)」,「新車(トラック)情報(109)」など、特徴ある車両についての記事である。これら以外にも、ドキュメント187, 188, 189はそれぞれ、新型車や輸入車などのランキングに関する記事であり、その関係がリンクにより表現されているといえよう。意味グループとしては、「大通り」や「vehicle」といった車に関するものや、「音声」など、新しい視点に対応するものが抽出されている。

### 4 最後に

本稿では、概念体系を用いた新しいドキュメントマッチング手法である Fish Eye マッチングの概要および、それを用いた情報整理支援プロセスについて紹介した。今後は、単一の興味による整理だけでなく、より複雑な文書間の構造を扱える様にする予定である。

### 参考文献

- [1] 高間 康史, 石塚 満: 概念体系を用いた Fish Eye マッチングによる視点を考慮した文書整理支援機能の実現, 情処研資 SIG-ICS, (1998).
- [2] 高間 康史, 石塚 満: 概念体系を用いた Fish Eye ベクトルの情報整理支援ツールへの応用, 人知研資 SIG-FAI-9702, pp. 97-102(1997).