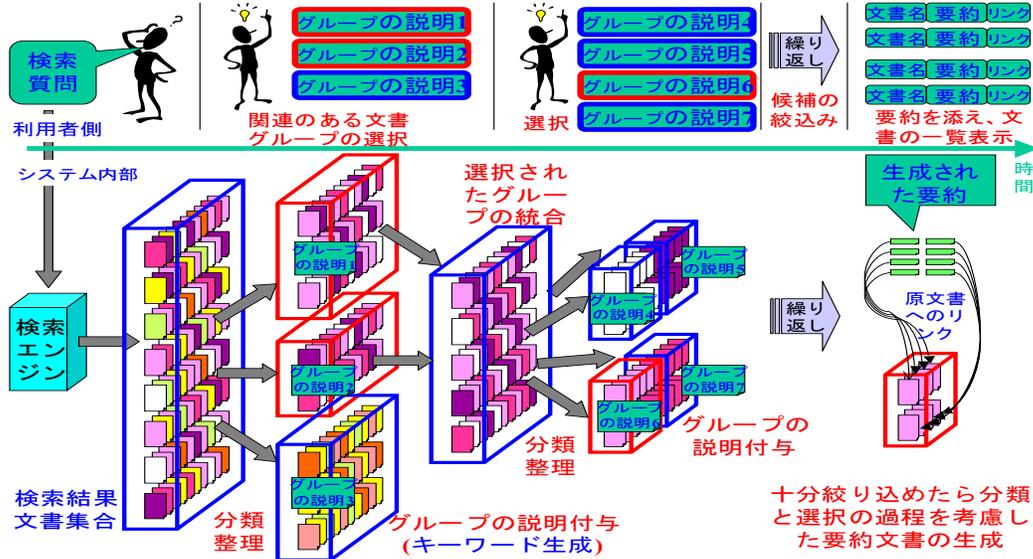


# (A02) 情報検索結果の知的提示のための 自動要約ならびにインタフェースに関する研究

研究代表者 森 辰則 (横浜国立大学大学院環境情報研究院)<sup>1</sup>  
 研究分担者 田村 直良 (横浜国立大学大学院環境情報研究院)



## 1 研究の目的

近年、WWW 検索エンジン等の情報検索システムが広く利用されているが、関連性の低い文書を完全には排除できない、検索結果文書の構造化がなされないなどの問題がある。その解決策の根幹をなすものが複数文書間の関係を考慮した重要語抽出であるという観点から、本研究では、複数文書の類似構造を文書中の語の重要度に写像するという新しい手法を確立することを目的とする。これに基づき、検索文書の要約、及び、情報ナビゲーションのための情報提示に関する技術を開発し、利用者が真に必要とする情報を情報検索結果から効率良く選択するための情報提示技法を検討する。

## 2 手法

上図は想定する利用者インタフェースの概要である。文書の分類整理部分には、動的なクラスタリングに基づく利用者インタフェースである Cutting らの Scatter/Gather を想定し、利用者との対話により検索対象を絞り込むことを考える。この時、如何にして有用な情報を利用者に提示できるかが重要である。我々は、分類整理された文書グループに対する適切な説明記述(キーワード)の生成と各文書に対する適切な要約の生成が必要不可欠であると考えている。キーワードの適切さは、そのグループにおいて特徴的であることのみならず、他グループとの違いを際立たせることも重要な観点となる。要約生成では、要約文書と検索質問との間の関連性判定が原文書における関連性判定に一致する程度が適切さの尺度となる。

本研究では、キーワードならびに要約の生成において、検索文書群の間の類似構造というマクロな情報を、文書内の各語の重みというミクロなものに射影する新しい技術を共通基盤として活用する。具体的には、まず、表層的類似性により検索文書を階層的クラスタに

分類し、文書間の類似性構造を抽出する。次に、クラスタ中の各語について確率分布を調べ、そのクラスタの下位分岐構造と整合性の高い語に高い重みを与える。特に、その重みに情報利得比(IGR)を用いる点が新しい。この IGR 値は、各語・各クラスタ毎に算出されるので、その値の統合手法により、提示するクラスタ構造の詳細度に応じた重みづけができる。これを TF 値、IDF 値等、従来提案の語の重要度と組み合わせ、総合的な語の重みを得る。

## 3 関連研究

検索結果文書の要約には、従来手法として、Tombras らの検索質問に基づくバイアス法がある。しかし、各種フィードバックや検索要求の拡張など検索エンジンにおける工夫の結果が反映されない。一方、本研究では、検索要求文を使わずに、検索文書集合のみを用いて語の重みをつけるので、この問題は生じない。

また、本手法は文書の分類・整理も同時に行なわれ、なおかつ、整理された構造間の特徴量として重要語を求めることができる点でも注目値する。各クラスタ毎に語の重みづけがなされるので、クラスタ毎に適切な重要語を選択し、説明記述を生成できる。一方、Eguchi らの研究や Fukuhara らの手法など従来手法においては各クラスタの内部の語の分布だけを見ている。

## 4 期待される成果

本研究の成果により、検索結果文書を、適切なキーワードの下に分類・整理をするとともに、必要に応じて原文書の要約を提示することにより、利用者に効率よく提示する手法が確立される。これは、情報検索システム等の大規模・多利用者システムのみならず、電子メールの知的整理システム等、個人利用システムにおいても、インタフェースの質的向上がなされ、利用者の負担が軽減する。

<sup>1</sup>〒 240-8501 横浜市保土ヶ谷区常盤台 79-7  
 mori@forest.eis.ynu.ac.jp