

平成 18 年度 情報工学コース卒業研究報告要旨

石川 研究室	氏 名	長 谷 川 幹 根
卒業研究題目	T-Scroll:時系列文書のクラスタリングに基づく トレンド可視化システムの開発	

インターネット上の情報提供・配信サービスの進展により、今日では、ネットワークを介したニュース配信などが盛んに行われている。それに伴い、大量の情報を要約しフィルタリングするための、オンラインテキスト情報処理の重要性がさらに増してきている。大量のテキスト情報を内容に基づきグループ化する手法として文書クラスタリング手法が存在するが、時々刻々と配信される時系列的な文書データに適したクラスタリング手法についての情報の要約と提示に関する研究に関しては、新たな技術の開発が求められている。

本研究では、文書クラスタリングシステム上に構築を行った、トピックの推移を可視化するインタフェース T-Scroll(Topic/Trend-Scroll) について、システムの構成と機能の詳細を述べる。T-Scroll システムには主として以下の特徴がある。

1. 継続的なクラスタリングにより得られた各時点のクラスタリング結果を、時間軸上にトピックを表すラベルと表示することで、各時点における主要なトピックを把握可能とする。
2. クラスタを選択することで、より詳細な情報や元記事を対話的に参照可能とする。
3. ある時点で得られたクラスタ集合に対し、一つ前の時点で得られたクラスタ集合から、関連度の強さに応じてリンクを張ることで、隣接する時刻におけるクラスタ間の関連の把握を容易にする。

図 1 に、T-Scroll システムのインタフェースを示す。T-Scroll を利用するユーザは、表示対象の期間をインタフェース上で指定し、表示ボタンをクリックすることで、図 1 のような画面が表示される。クラスタ上のラベルは、クラスタ中の文書に含まれる語で、スコアが最大のものを選択して表示する。また、クラスタ上に書かれた楕円の面積はクラスタに含まれる文書の数の量に対応しており、トピックの規模を示している。さらに、クラスタの質の良さも容易に把握できるようにするため、T-Scroll ではクラスタの質の高さを色分けして表示する。一部のクラスタ間には左から右にリンクが張られている。これはクラスタ間の関連性の深さを示している。

T-scroll には、クラスタの内容を容易にブラウザできる機能もある。クラスタ上にマウスカーソルが乗ると、そのクラスタに関連の深いキーワードリストが表示されたり、クラスタの上をクリックすることでクラスタに含まれる文書を表示する機能がある。以上の機能の実行の様子を図 2 に示す。

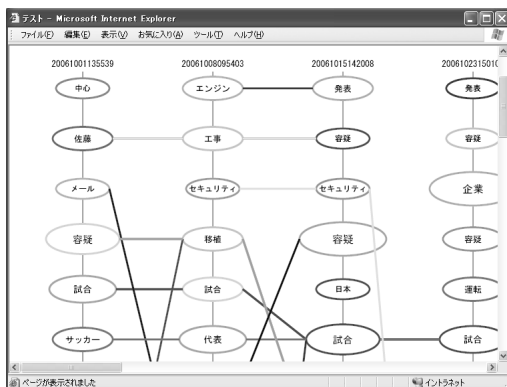


図 1 T-Scroll のインタフェース画面

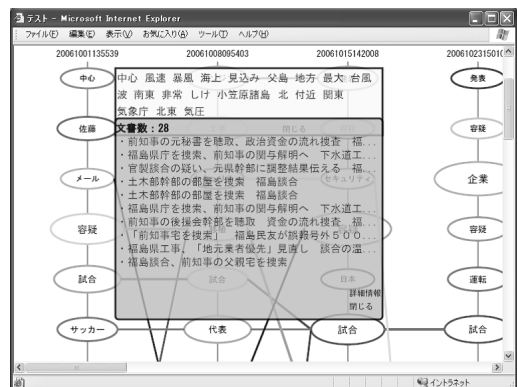


図 2 クラスタの詳細情報表示