

汎用連想計算エンジン GETA とそれに基づく連想検索システム

西岡 真吾, 今一 修
株式会社日立製作所 中央研究所
{nis,imaichi}@harl.hitachi.co.jp

GETA は、計算コストの高い連想計算を高速実行する汎用連想計算エンジンである。GETA を利用することにより、従来は実現困難であった大規模文書を対象とした文書連想検索システムの実現が可能になる。また、連想計算に用いる類似度を容易に差し替えることができるため、様々な類似度を試すことができる。本デモでは、GETA の基本機能および GETA に基づいた高速文書連想検索システムを紹介する。

[キーワード] 情報検索, 文書連想, 大規模文書

An Associative Search System based on a Generic Engine for Transposable Association (GETA)

Shingo NISHIOKA, Osamu IMAICHI
Central Research Laboratory, Hitachi, Ltd.

GETA is a generic engine for association calculation, which is able to execute association calculation very fast. GETA enables us to construct an associative search system for large corpora that is hard to be realized. Moreover, similarity measures for association can be easily changed, so we can try various types of measures. This demo introduces basic functions of GETA and an associative search interface based on it.

[keywords] information retrieval, document association, large corpus

動的クラスタリング、話題の自動抽出、類似文書を検索する連想検索などは計算コストの高さからこれまで大規模文書データへの適用が困難であった。GETA は、データ圧縮と高速データアクセスにより、複雑な文書情報処理計算を現実的な速度で実行でき、従来のキーワード検索と比較して極めて高精度な検索システムを構築することを可能とした。

GETA の特長は、1. 大規模コーパスを対象とした文書連想検索など、計算コストが非常に高い計算を高速実行可能、2. 様々な尺度に基づく文書情報処理方式を実現できる高い汎用性、3. 実用規模のシステムの構築が可能な強力かつ軽量の連想計算エンジンを装備、4. 文書情報処理の研究者を対象とした洗練されたツール群の提供などである。

本研究は IPA 独創的 PJ の一環として行われた。

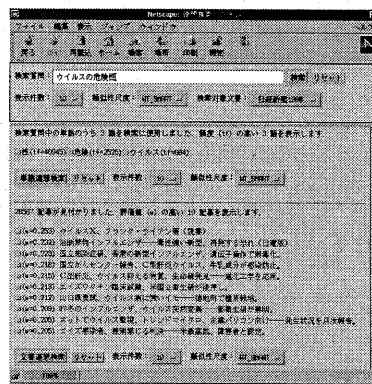


図 1: GETA を利用した連想検索システム