

国立情報学研究所は、検索タスクに特化した新しい連想検索エンジン「GETAssoc」を公開した(<http://getassoc.cs.nii.ac.jp/>)。検索に頻繁に用いられる機能を中心に構成しており、様々な用途に応じた連想検索アプリケーションを、従来に比べて柔軟かつ容易に構築することができるとしている。

国立情報学研が新たな検索エンジン

連想計算は、文書に現れる単語の頻度に基づいて、類似文書や関連単語を統計的手法によって計算する技術。連想検索は、連想計算を応用した文書検索アプリケーションで、ある文書から関連文書を探し出すといった高度な検索ができる。さらに、文書が複数のデータベースに分かれて登録されている場合でも、あたかもそれらが同一データベースに登録されているかのようになり、文書同士の関連度で検索が行える。そのため、キーワード串刺し検索などのように検索結果を統合する手法に比べて、より効率的に関連文書を探すことができる。

GETAssocは、分散して配置された連想計算ライブラリ間の連携を可能にする通信規約(プロトコル)を備えている。

文書の関連度で探す「連想」アプリケーション簡単構築

データベースを連携させ、これらのデータベース間のやり取りを行う。gss3はGETAssoc用に新しく設計されたプロトコルで、連想計算の特長の一つである複数のサーバに分散して配置されたデータベースを横断的に検索できるという機能を効率的に利用できる。gss3プロトコルにより複数データベースの連携を行う際には、GETAssocが通信を仲介することで、連想計算ライブラリ間での直接の通信は行わない。この設計によって、コンテンツの権利保護が可能になった。例えば、あるデータベースを他のデータベースと横断的に連想検索できるようにするために公開しなければならぬ情報は、データベース中の各文

書の特徴づける単語リストだけであり、本文そのものは非公開のままでの運用ができる。また、連想計算ライブラリ同士が直接通信する必要がないため、アクセス制御などの連想計算ライブラリ運用の負荷が低減される。gss3はこうした特長を備えているため、gss3に対応した連想検索アプリケーションであれば、既にgss3を採用している他のアプリケーション

などとも連携させて運用することが容易だ。gss3のもう一つの特長は、連想検索アプリケーションから連想計算ライブラリを効率的に呼び出すことができるよう設計されていること。そのため、煩雑な呼び出し手順を記述することなく、容易に連想検索アプリケーションを構築することができる。例えば、これまでは連想検索アプリケーションを実装する際、一回の連想検索のために複数回の連想計算ライブラリを呼び出す手続きを記述する必要があったが、gss3プロトコルを利用すれば一回の呼び出しで必要とする計算を一気に行える。