

# 関連度から文書連想

## 情報学研 検索エンジン公開

国立情報学研究所は、ある文書に関連した文書を探し出す新しい連想検索エンジン「GETAs soc」(<http://get>

[assoc.cs.nii.ac.jp](http://assoc.cs.nii.ac.jp))を公開した。複数のデータベース(DB)にある文書を、あたかも同一のDBから引っ張り出すようにして「関連度」から調べたい内容の文書を見つけて出す。従来の連想検索エンジンより運用が柔軟で、検索した文書の編集も可能という。

特化した連想エンジン。「GETAs soc」と連想検索アプリケーションを緩やかに結び、独立に運用する異種アプリを連携できるようにした。複数DB間の横断検索が可能で、分散して配置した「連想計算」を連動して行う。横断検索する場合も、本文などコンテンツの権利は保護する。「連想計算」は、単語

が文書に現れる頻度に基づき、類似文書や関連単語を統計的に計算する手法。一方「連想検索」は連想計算を応用した文書検索アプリ。横くして検索し、結果を統合する「キーワード」検索に比べ、効率的に関連文書を探し出せる利点がある。