

データマイニング技術の応用によるウェブからの知識発見

情報科学研究科/社会情報科学部 大島 裕明



キーワード

情報検索、ウェブ、知識発見、データマイニング

研究概要

ウェブページやソーシャルメディアには、書き手が意識的または無意識的に含ませた様々な知識の断片が存在している。そのような知識の断片を効率よく集め、集約することによって、様々な応用が可能となる知識を取得することが可能となる。知識の断片を効率よく集めるに、ウェブ検索エンジンを活用する。たとえば、「兵庫県立大学」と並び称されるようなものは何かという疑問に対する回答が知りたいとき、ウェブ検索エンジンで「や兵庫県立大学」というキーワードで検索すると、「や」の前には正解がいくつか現れる。さらに、「兵庫県立大学や」というキーワードでも検索すると、今度は「や」の後ろに正解がいくつか現れる。さらに、少し工夫することで、8割程度の精度で並び称されるような関係にある語を取得することが可能である。これは、あくまで提案手法で得られる知識の一例でしかなく、用いるパターンを変化させることで、上位関係にある語、下位関係にある語、与えられた語の定義、与えられた地域名の名産品など、様々な種類の知識を取得することが可能である。ある種の知識発見のために必要な文書を効率よく取得し、オンデマンドに知識発見を行うことができるため、様々な応用に用いることができる。

アピールポイント

ウェブ検索エンジンを利用しているため、本技術を用いて得られる知識も日々更新され、また、適用範囲が広い。そのアプリケーションに合わせて、そのアプリケーションにしか必要がないような最新の辞書を気軽に構築することが可能である。

応用分野

情報を取得するということにかかわる情報アクセスシステム全般において適用可能。その中には、情報検索システム、情報推薦システム、情報分析システムなどを含む。