

論文内容の要旨

博士論文題目 多言語情報アクセスのための言語資源の構築と活用
に関する研究

氏名 木村文則

(論文内容の要旨)

本論文は、ある言語で書かれた文書群を別の言語による問合せで検索することを可能とする多言語情報アクセスあるいは言語横断型情報検索において、ドメインに依存しない言語資源を構築する技術及びその活用について論じる。

世界的なインターネットの発展とともに、激増する利用者の国籍も、Web 文書の記述に用いられる言語も多様化している。利用者の検索要求によっては、利用者の母国語以外の言語で記述された情報の方が豊富である場合も考えられ、これらを検索したいというニーズは少なくない。しかし多くの利用者は母国語以外の言語に精通していない。従来のWeb 検索エンジンは、問合せと同一言語の文書群が検索対象であるため、外国語文書に対する検索は効率的ではなく、利用者自身が辞書などを用いて問合せを翻訳する必要があった。この作業は利用者に負担を強いるだけでなく、訳語の選択を誤る可能性があった。

このような要求から、言語の壁を越えた多言語情報アクセス（あるいは言語横断型情報検索）に関する研究が盛んになり、問合せの翻訳や訳語の曖昧性解消などにコーパスを利用する手法などが提案されて、検索精度の向上に一定の成果が得られてきた。しかしコーパスを利用した手法では、学習に用いるコーパスのドメインに対する依存が大きいため、それ以外のドメインに対しては検索精度が低くなる可能性がある。Web 文書の言語横断検索では文書内容の分野は広範囲に渡っているため、このドメイン依存の問題を改善しなければならない。

本研究は、ドメイン依存の問題を解決するための言語資源の構築および利用を中心テーマとしている。Web ディレクトリを利用してある言語のオントロジーを別の言語に翻訳することにより、二つの言語を対象としたオントロジーを構築する手法を提案した。また、Web ディレクトリを言語資源として用いることを提案し、これを利用した言語横断情報検索システムの構築を行った。さらに、これらのシステムの実証実験を行い、Web ディレクトリを言語資源として利用することが多言語アクセスシステムに有効であることを確認した。

氏名	木村文則
----	------

(論文審査結果の要旨)

1月29日に開催した公聴会の結果を参考に、2月15日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

木村文則は本博士論文において、多言語情報アクセスの検索精度を向上させる方法を提案し、実験してその結果を評価した。多言語情報アクセスあるいは言語横断型情報検索は、ある言語で書かれた文書群を別の言語による問合せで検索することを可能とする技術である。このために、様々の研究が行われており、現在は、それぞれの言語のコーパスを利用する手法が主流である。しかし、コーパスを利用した手法では、学習に用いるコーパスのドメインに対する依存が大きいため、それ以外のドメインに対しては検索精度が低くなる可能性がある。とくに、Web文書の言語横断検索では文書内容の分野が広範囲に渡っているため、ドメイン依存の問題を改善しなければならない。

本論文は、こうしたドメイン依存問題を解決するための言語資源の構築および利用を中心テーマとしている。具体的には、Webディレクトリを利用してある言語のオントロジーを別の言語に翻訳することにより、二つの言語を対象としたオントロジーを構築する手法を提案した。さらに、Webディレクトリを言語資源として用いることを提案し、これを利用した言語横断情報検索システムの構築を行った。具体的には、利用者の問合せを別の言語の問合せに翻訳するときに、Webディレクトリを用いて、翻訳におけるあいまいさを解消する。こうした考え方に基づくシステムの実証実験を行い、Webディレクトリを言語資源として利用することが多言語アクセスシステムに有効であることを確認した。とくに、Webディレクトリをどの深さまで利用するべきかについて、実験結果をもとに具体的に論じた。

多言語情報アクセスにおけるドメイン依存の問題を解決するために、具体的な方式を提示して実験評価を行った本研究は、独創性が高く、しかも実用的であり、多言語情報アクセスの分野において高い貢献があると評価する。

よって、本論文は、博士(工学)の学位論文として価値あるものと認める。