

論文内容の要旨

博士論文題目 Case Studies on Cross-Language Information Retrieval
and Bilingual Terminology Acquisition from Comparable Corpora

氏名 Fatima Sadat

(論文内容の要旨)

インターネットの急速な普及にともない、オンライン文書を始めとするオンライン情報資源の手軽な交換、相互利用が可能になりつつある。国際協調の可能性が拡大する一方で、各国文字、言語による電子的なテキストは増加の一途をたどり、英語を母国語としない利用者も増え続けているので、非ラテンアルファベット圏の利用者を中心に、言語間の障壁をなくするための言語横断型情報検索システム(CLIR: Cross Language Information Retrieval system)の研究開発が盛んである。CLIRとは、利用者がある言語で質問をすると、それとは別のさまざまな言語による解答文書を検索することができるシステムである。

CLIRでは、まだ十分に解決されていない研究課題が多い。問合せの翻訳におけるあいまいさの解消、対訳辞書における語彙の欠如などは、その例である。本研究では、次の課題に明快な見通しを与えることを目指した。

- (1) 問合せ翻訳におけるあいまいさを解消する技術 — 辞書だけを手がかりとする問合せ翻訳に統計的な手法を導入して、あいまいさを解消する技法を提案した。仏英2か国語間の TREC テストデータによる評価では、もとの用語の対をまずランク付けし、目的の言語への翻訳におけるあいまいさを解消するという2段がまえの手法が CLIR にきわめて有効であることを示した。
- (2) 問合せを拡張するために複数の言語資源を結合する技術の評価。適合性フィードバック、ドメインを手がかりとするフィードバック、シソーラスの利用により問合せを拡張する手法を、それぞれ問合せを翻訳する前か後に適用する手法を比較検討した。TREC テストデータによる実験では、シソーラスの利用と、ドメインによるフィードバックの組み合わせが効果的であることがわかったが、最適な拡張を行うには、それぞれの手法を適

切にランク付けしておく必要がある。

- (3) 比較可能コーパスから用語の対訳を獲得する技術。対訳辞書に見出し語があるかどうかわからない固有名詞の対訳語を、前述の単語対の共起頻度などを手がかりとして正しく獲得する手法を提案して、検討した。NTCIR テストデータによって、この方法がよい結果を生むことを確認した。

本研究では、以上の問題を中心に、言語横断型情報検索システムの基本課題に明快な見通しを与えることができた。

氏名	Fatiha Sadat
----	--------------

(論文審査結果の要旨)

平成15年6月24日に開催した公聴会の結果を参考に、全審査委員により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。

Fatiha Sadat は、本博士論文において、言語横断型情報検索システムの多言語処理を取り上げ、二つの技術課題を核に問題と解決法を論じている。一つは、問合せの翻訳におけるあいまいさの解消であり、もう一つは、対訳辞書における語彙の欠如である。

問合せ翻訳におけるあいまいさを解消する技術は、言語横断型情報検索の基本課題である。検索対象の全文を翻訳することは物理的に不可能なので、問合せを翻訳する。しかし、問合せは、短く、翻訳するとあいまいになりがちである。本研究では、まず辞書だけを手がかりとする問合せ翻訳に、統計的な手法（共起頻度など）を導入して、あいまいさを解消する方式を提案した。とくに、複数の単語からなる問合せにおいて、すべての単語対の組合せを共起頻度によりランク付けし、目的の言語への翻訳におけるあいまいさを解消するという2段がまえの手法を提案し、実験して、仏英2か国語間の TREC テストデータによる評価で、CLIR にきわめて有効であることを示した。

問合せを拡張して、よりよい検索結果を産む研究も盛んである。そうした研究で利用されてきたさまざまな手法、すなわち適合性フィードバック、ドメインを手がかりとするフィードバック、シソーラスの利用により問合せを拡張する手法を、それぞれ問合せを翻訳する前か後に適用して、比較検討した。TREC テストデータによる実験では、シソーラスの利用と、ドメインフィードバックの組合せが効果的であることがわかったが、最適な拡張を行うには、それぞれの手法を適切にランク付けしておく必要がある。

最後に、比較可能コーパスから用語の対訳を獲得する技術についても、同様の共起頻度を応用する手法を検討し、対訳語の自動獲得手法を提案している。

以上のように本研究は、独創性が高く、しかも実用的であり、言語横断型情報検索の分野に大きく貢献するものと評価する。