

平成20年度 情報工学コース卒業研究報告要旨

| | | |
|-----------|----------------------|-------|
| 宮尾・八槇 研究室 | 氏 名 | 鵜飼 真充 |
| 卒業研究題目 | メタデータによる多言語資源相互利用の実現 | |

現在、辞書、用例対訳、機械翻訳といった多言語資源が様々な組織毎に蓄積されている。言語グリッドは、このような分散した多言語資源を Web 上で共有・連携させるための基盤である。言語グリッドにおいて多言語資源は、Web サービスへとラッピングされる(多言語 Web サービスと呼ぶ)。各組織は必要な多言語 Web サービスに対してアクセスし、自分の持つ多言語資源へと統合することで、資源不足を解決できる。また、様々な資源を組み合わせ、利用現場に則した新しいサービスを開発することが可能である。

本研究では、言語グリッドを用いて対訳資源(複数の言語で表現された例文集)の相互利用を試みた。実際に、対訳資源を用いた翻訳サービスを提供するシステム(用例対訳アプリケーション)の間で相互利用実験を行ったところ、不要な対訳文まで取得されることがわかった。これは言語グリッドの現行の枠組みでは、キーワード検索やカテゴリマッチングといった単純な検索方法しか提供されておらず、対訳文の細かな内容を考慮した検索ができないためである。

この問題を解決するために、対訳文のメタデータを検索に利用するというアーキテクチャを提案した(図1)。このアーキテクチャでは、全ての対訳文のメタデータをディレクトリ機構である RDF ストアサービスで管理する。用例対訳アプリケーションはこのサービスを利用することで、必要な対訳文のみを取得することが可能となる。

一方、提案アーキテクチャが有効に働くにはメタデータ体系が重要である。そこで防災ドメインに限定して、ファセット分類に基づくメタデータの設計を提案し、複数の対訳資源に共通なカテゴリ語彙を設計した(表1)。全ての対訳文をこのカテゴリ語彙を用いて1つの体系のもとにファセット分類し、メタデータを付与する。これにより、主題や状況といった細かな内容を指定して、対訳文を検索することが可能となる。また、ファセットを組み合わせた検索によって、不要な対訳文を抑制できるようになる。実際に設計したカテゴリ語彙を用い、対訳文の検索実験を行ったところ、適合率・再現率ともに十分な結果が得られた。本研究により防災ドメインの対訳資源相互利用が実現可能となった。

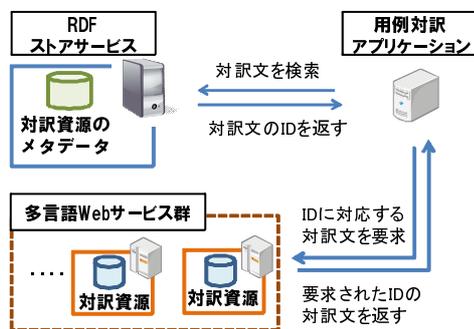


図1. メタデータを利用した検索アーキテクチャ

表1. 設計したカテゴリ語彙

| ファセット | 語 |
|-------|--|
| 災害の種類 | 地震, 余震, 火災, 津波, 台風, 洪水, 噴火, 土砂災害, 地形変化, etc |
| 対象時期 | 災害発生前, 災害発生後 |
| 文の性質 | 伝達, 依頼, 注意, 警戒, 身体的危険, 損失の危険, 行為要求, 行為禁止 |
| 文の主題 | 生活情報, 情報源, テレビ, ラジオ, 交通, 鉄道, 航空, 行政発表, 観測, 予知, 避難情報, 避難指示, 避難場所, etc |

発表実績:

- M. Gotou, H. Yamaki, D. Yanagisawa, M. Ukai, M. Tanaka, T. Ishida.
Resource Sharing by Multilingual Expression Services, APSCC-08, pp.470-475, 2008.
- 竹内, 八槇, 鵜飼. 言語グリッドのためのメタデータオーサリング環境の開発, 信学技報, to be published, 2009.
- 古田, 柳澤, 鵜飼, 八槇. 言語グリッドにおけるシステム統合のためのディレクトリ機構の実現, 信学技報, to be published, 2009.
- 鵜飼, 竹内, 八槇. メタデータを用いた多言語資源相互利用の実現, 情報処理学会第71回全国大会, to be published, 2009.