

京都学デジタル図書館プロジェクト — 京都学コンテンツに対する情報アクセスの研究 —

前田 亮
理工学部 情報学科

1 はじめに

近年、デジタル図書館やデジタルアーカイブが注目され、さまざまな文化的資料のデジタル化や保存に関する研究が盛んに行われている。しかしながら、それらのコンテンツに対して容易で効率的なアクセス手段を提供するという観点からの研究はまだ多くはない。コンテンツの量が膨大になればなるほど高度なアクセス手段が要求されることは、現在の Web の状況を見ても明らかである。本研究プロジェクトでは、主に文字情報を対象として、高度な情報アクセスを実現する手法について研究を行っている。

2 京都学のための情報アクセス技術

2.1 古文書の内容検索

本研究プロジェクトでは、典籍・古文書などに対して、単なる文字列マッチングではなく、文書全体あるいは単語単位、さらには文字単位で意味を解析することにより、たとえば現代語による検索などの機能を提供することを目指している。このためには、XML、メタデータ、セマンティック Web などの技術が不可欠である。また、古文書には現在の文字コードに含まれない文字が多く含まれるが、これら「外字」を含む文書に対して効率的な検索を行う手法についても研究を行う。現在のところ、古文書および外字を含む文書に対する内容検索の技術について基礎的な検討を行っている段階である。

2.2 言語の壁を越える検索

京都学に関するコンテンツの研究では、国内だけではなく東アジアなどの近隣諸国との関係も重要である。そのため、研究対象となる資料が書かれている言語が必ずしも日本語だけとは限らない。また、本 COE プログラムにおける研究成果を公開する際には、日本だけではなく世界のより多くの人々が容易にアクセスできる手段を提供するべきである。この実現のために、コンテン

の翻訳版を用意することなく検索を可能とする言語横断情報検索技術について研究を行っている [1]。現在のところ、複数の言語版を持つ Web ディレクトリの階層構造をコーパスとして使用し、言語横断情報検索における訳語曖昧性を解消する手法について研究を進めている [2, 3]。

3 おわりに

本研究は、古文書の現代語による検索と、言語間を跨って検索する言語横断検索とを、概念検索という共通の枠組みを用いて融合することで、京都学コンテンツに対して時代や言語の壁を越えた検索を可能にすることを目指している。

本 COE プログラムにおいてデジタル化が進められている、京都に関わる過去／現在の膨大かつ多様な文化・アートの集積にアクセスする手段を提供するには、単なる文字列マッチングではなく、より高度な検索技術が必要不可欠である。本研究プロジェクトの成果によって、世界中のより多くの人々が、京都に関わる過去から現在に至る膨大な知識の集積に容易にアクセスすることが可能になる。

参考文献

- [1] Maeda, A.: Multilingual Information Processing for Digital Libraries, *Proc. PNC2002*, pp. 80-81 (2002).
- [2] 木村文則, 前田亮, 吉川正俊, 植村俊亮: Web ディレクトリの階層構造を利用した言語横断情報検索, 日本データベース学会 Letters, Vol. 2, No. 1 (2003). (掲載予定) .
- [3] Kimura, F., Maeda, A., Yoshikawa, M. and Uemura, S.: Cross-Language Information Retrieval Using Web Directories, *Proc. IEEE PACRIM'03* (2003). (to appear).