# 論文内容の要旨

博士論文題目　Studies on Multilingual Information Processing on the Internet
（インターネット上における多言語情報処理に関する研究）

氏　名　　前田 亮

（論文内容の要旨）

世界的なインターネットの発展による各地域での利用者の拡大に伴い，Web 文書に用いられる言語も従来の英語中心から様々な言語に広がりつつある．しかしながら，これらの多言語文書群を統一的に扱う検索システムの実現には，解決されていない様々な課題がある．

利用者の観点から見ると，一般的なインターネットの利用に不可欠なテキスト処理機能は，表示，入力，検索の三つである．しかしながら，たとえば日本語，中国語，韓国語などの言語については，テキストの表示や入力に必要な文字フォントや入力メソッドが必ずしもクライアント側にインストールされているとは限らない．

一方，システム側の観点から最も面倒な問題は，Web 文書に用いられる文字符号系が言語により多種多様でありながら，多くの Web 文書にはそれ自体に符号系のメタ情報が含まれていない点である．これは，Web ブラウザ上での文字化けや，Web 検索エンジンにおける誤った索引付けの原因となる．

また，特に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，利用者が不慣れなあるいは読み書きができない言語で記述された文書を検索したい場合があると考えられる．このような情報を検索したいというニーズは少なくないと思われる．このため，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言語横断情報検索（Cross-Language Information Retrieval: CLIR）に関する研究が近年盛んになってきている．しかしながら，多様な言語およびドメインの文書が混在する Web 文書に対して相応の検索性能を得ることは困難である．

本論文では，これらの問題点に対するいくつかの解決策を示すことを目標とする．具体的には，次のような手法を提案する．

1. クライアント側にインストールされているフォントや入力メソッドに依存しない，多言語テキストの表示および入力機能
2. 統計的手法とヒューリスティクスの併用による，Web 文書の言語および符号系の自動識別アルゴリズム
3. Web 検索エンジンから得られた単語共起情報に基づく，多様なドメインの Web 文書に適した言語横断情報検索手法

これらの三つの手法を統合することで，利用者の母国語以外の言語で書かれた文書へのアクセスを支援するシステムを実現した．このシステムにより，インターネットに特有の多言語情報処理に関わるいくつかの問題点に対して，一定の解決策を提示することができた．

| 氏　名 | 前田　亮 |
|---|---|

（論文審査結果の要旨）

本論文は，インターネット上における多言語情報処理に関わる問題点について，利用者側とシステム側の観点からの解決を図った研究について述べたものである．本論文における研究成果は次の3点に要約される．

（1）多言語Webテキストの表示および入力機能：
本研究では，クライアント側へのフォントや入力メソッドのインストールを一切必要としない多言語テキストの表示および入力機能を実現している．本手法は，HTML文書にその文書中に出現する文字のみのフォントを付加した形のMHTML文書と呼ぶ形式のものを作り，それを表示するためのJavaアプレットと共にクライアントのWebブラウザに送ることにより，多言語テキストの表示および入力機能を実現している．本手法は，インターネットの一般的な利用者が容易に利用でき，かつ効率的な多言語テキストの表示および入力機能を実現する実用的な手法として高く評価できる．

（2）Web文書の言語および符号系の自動識別アルゴリズム：
本研究では，統計的手法とヒューリスティクスの併用による，文書の言語および符号系の効率的な自動識別アルゴリズムを提案している．本技術は，現状においてインターネット上の文書を扱う上で重要な基本技術である．提案手法は，アジア言語とヨーロッパ言語からなる12言語10符号系を対象とした実験において，98%以上の正解率が得られている．本研究は，従来の手法と比較して大幅な効率の向上を達成しながら，ほぼ同等の正解率を得ることに成功しており，評価できる．

（3）多様なドメインのWeb文書に適した言語横断情報検索手法：
本研究では，容易に利用できるWeb検索エンジンから得られた単語共起情報を用いて，言語横断情報検索における訳語の曖昧性解消を行う手法を提案している．実験の結果，一般的なWebの利用者が用いるような非常に短い問合せに対して本手法が有効であることを確認している．また，検索対象と一致するコーパスを用いることで，人手で翻訳した場合とほぼ同等の検索性能が得られることが示されている．本研究は，Web文書の言語横断情報検索における訳語曖昧性解消に対する実用的な解決策を示している点で評価できる．

以上のように，本論文における三つの研究成果は，インターネット上における多言語文書の処理に関わる問題点に対する実用的な解決策を提示しており，学術上，応用上寄与するところが多い．また，本研究の成果は，学術論文4件および国際会議3件において公表されているほか，研究成果の一部について平成11年度情報処理学会論文賞を受賞しており，本研究の学術面での貢献を認めることができる．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．