MetaCommander: スクリプトに基づく WWW 情報収集システムの試作
MetaCommander: a script-based WWW information gathering system

野崎 哲也† 北村 泰彦† 辰巳 昭治†
Tetsuya NOZAKI Yasuhiko KITAMURA Shoji TATSUMI
† 大阪市立大学工学部
Faculty of Engineering,Osaka City University

1. まえがき

情報発信の手段としての WWW は情報発信の容易さと表現力の高さによって急速に普及した. 情報の発信者にとっては home page を作るだけで全世界に向けて自分の情報を発信することが簡単にでき, 文章だけでなく画像, 音声, 動画なども混在して扱うことができる. そして, ハイパーリンクを張ることで他のページの情報と関連づけることができる. 一方, 情報の受信者の観点に立つと, 全世界に分散し個別に構築されているサーバーから必要な情報を捜し出すのは困難である. また, 表現力の高さは逆にデータ構造の統一性をなくし, 機械的に情報を抽出することを難しくしている.

これらの問題を解決するために, キーワード検索によって必要なページを探し出すサーチエンジンがある. しかし, サーチエンジンでは必要な情報とは関係のないページが出力されることが多い. 更に, 検索はページ単位であり必要な情報だけを抽出することが困難である.

サーチエンジンの検索結果を加工して出力をする Internet softbot[1] が研究されている. しかし, 既存の Internet softbot は特定用途に限定され汎用性がないので, 個々のユーザーの要求を満たすことは不可能である. そこで, 利用者が目的に応じてスクリプトを記述することで必要な情報の収集を可能とする MetaCommander を提案する.

2. MetaCommander

MetaCommander はスクリプトを解釈し, 命令を実行するインタプリタである. ユーザーはスクリプトを記述することで MetaCommander を制御し, 収集した情報を WWW ブラウザで見ることができる.

スクリプトは複数の関数列からなり, 関数はその順番どおりに実行される. 検索, 抽出, 整形, 表示と保存, 制御の機能を実現するために表 1 に示すような関数を用意した. 検索関連の関数では,url で指定された HTML を開く関数だけでなく, インタラクティブな検索が可能となる CGI を呼び出す関数も用意した. また,HTML はタグによって文章が構造化されている. そこで, ページ中から必要な情報だけを抽出する方法として, タグで囲まれた部分を抽出するようにした. searchString 関数では文字列検索をし, それを含む部分を抽出する.

そして, スクリプトを用いることによって,

- サーチエンジンを通じて何度も同じような作業を繰り返さなければならない処理の自動化
- WWW 上の複数のデータベースの出力結果の統合や, 異種のデータベースを組み合わせた情報検索 [2]

このようなことが実現されている.

分類	関数	機能
検索	`getURL`	url を開く
	`postURL`	CGI に post する
抽出	`searchString`	文字列をタグ単位で検索する
	`getAnchor`	HTML 中の anchor を検索する
整形	`tag`	HTML のタグを出力する
表示, 保存	`print`	文字列や HTML を出力する
	`file`	ファイルをオープンする
	`close`	ファイルをクローズする
	`read`	ファイルから 1 行読み込む
	`write`	ファイルへ 1 行書き込む
制御	`set`	変数を設定する
	`if`	条件分岐をする
	`for`	指定回数のループをする
	`while`	条件ループをする
	`foreach`	繰返し処理をする

表 1:MetaCommander の関数

3. 今後の課題

ユーザーが必要な情報だけを自動的に抽出することはデータ構造の明確な関係データベースなどから行なうのは比較的簡単である. しかし, データ構造の統一性がない Web ページからは難しい. たとえば, ある大学教授の E-mail アドレスを知りたいとする. その大学の home page からリンクを辿って教授のページを見つけ出さなければならない. このとき, どのような構造で Web ページが構成されているかは大学によって異なるため, 単純なキーワード検索では困難である. そこで, このような情報抽出を行なうためには推論や探索機構が必要である.

参考文献

[1] O.Etzioni, “Moving Up the Information Food Chain: Deploying Softbots on the World Wide Web”, AAAI96 (1996).

[2] Y.Kitamura et al,“MetaViewer and MetaCommander: Applying WWW Tools to Genome Informatics”, Genome Informatics 1996, Universal Academy Press, pp.137–146 (1996).