平成 12 年度 卒業論文概要

WWW サーバからの一括ダウンロードツールの作成

静岡大学 工学部 システム工学科 4 年
若林研究室 5071-3067 野末 佳伸

1 序論

ここ数年でインターネットは飛躍的に普及してきている．またインターネット上のホームページ (以下 HP) の数は膨大であり，全世界では数億の HP が存在する (1997 年 12 月で約 3 億 2000 万ページ)．今やインターネットは非常に大きな情報源になっている．その HP 上では文字や画像，ソフトウェアなど様々な情報が公開されている．それらの中には有用であるものも多く含まれている．これらのダウンロードの作業を簡略化する為にダウンロード支援ツールが存在する．

本研究では，既存のツールについて検討し，不便な点を改善する為に機能を追加・変更することで，既存のツールよりも簡単に，効率良く必要なファイルをダウンロードできるようにするツールを作成する．

2 ダウンロード支援ツール

既存のダウンロードの支援ツールの中に，必要なファイルのみを選択してダウンロードするツールが存在する．これらのツールでは，URL を入力，拡張子を選択し，HTML を解析することによりファイル名を抽出，そこからファイルを選択してダウンロードを行う．

このツールを使用することにより複数のファイルを一度にダウンロードすることができる．また必要なファイルのみをダウンロードするので効率が良い．だがダウンロードするファイルを選択する際どれがどのファイルか分かりにくい時があった．

3 ツールの概要

本ツールでは，上で述べたツールと同様，HTML ファイルの URL を入力，抽出したい拡張子を選択し，入力した URL の HTML ファイルから指定した拡張子のファイル名を抽出，その中からダウンロードするファイルを選択し，ダウンロードを行う．またその上でいくつかの点を改良している．

本ツールの流れを図 1 に示す．

4 改良点

このツールでは既存のツールに対して主に以下のような追加・変更をした．

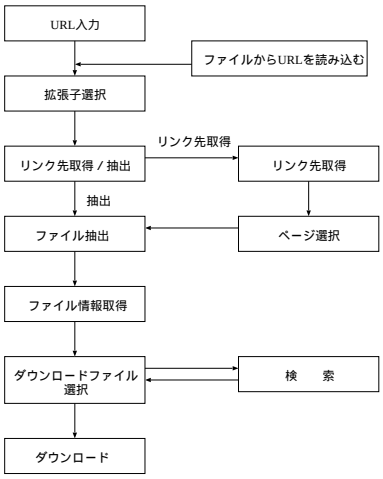

図 1 ツールの流れ

- ヘッダからのファイル情報の取得

  ヘッダを取得し，まずそのファイルが存在するかどうか判定する．その後，ヘッダ情報よりファイルのサイズやファイルの最終更新日を調べ，それらをダウンロードするファイルを選択する際の判断材料とする．

- 検索機能

  取得したファイル情報より抽出したファイルの検索を行う．検索対象はファイル名 (文字列)，ファイルのサイズ，ファイルの最終更新日である．

- ファイルからの読み込みやリンク取得による複数ページからの抽出

  URL を入力する代わりに，あらかじめいくつかの URL を入力しておいたファイルから URL を読み込めるようにする．また，HTML ファイルを解析し (リンクタグを識別する)，リンク先を取得，その中から抽出を行うページを選択し，抽出を行う．

ヘッダからのファイル情報の取得と検索機能により，ファイルを選択する為の手段を増やし，より確実に必要なファイルを選択できるようにした．また，複数ページから抽出することにより，各ページ毎に抽出する作業を一度に行えるようにした．

5 結論

既存のツールよりも簡単に，効率良く必要なファイルをダウンロードできるようにするツールを作成することができた．