

山中国広* 白川正知† 古川泰男†

*豊橋技術科学大学 エコロジー工学系

†豊橋技術科学大学 未来技術流動研究センター

1. はじめに

地域におけるインターネット利用推進のために、豊橋地域発の WWW ページを生活情報の観点からカテゴリ別にデータベース化し、豊橋コミュニティ・ナビゲータとして公開している^[1]。インターネットの利用を考える上で、WWW ページがどの程度新鮮な情報を提供しているかが重要である。そこで豊橋コミュニティ・ナビゲータに登録されている WWW ページの更新頻度を調査する。更新頻度を調べるために、WWW サーバにアクセスして、WWW ページのヘッダ情報を取得する。数 100 ページの更新調査を効率よく行うために、WWW サーバを巡回してヘッダ情報を自動的に取得する巡回ロボットを構築する。このロボットを毎日定時に巡回させ、ヘッダ情報から更新日時を取得し、更新頻度を調査する。

2. 巡回ロボットの構築

作成する巡回ロボットのフローチャートを図 1 に示す。巡回ロボットのプログラミングでは、サーバとのソフトウェア上の互換性を考え、Perl 言語を用いる。調査時刻になると、自動的に巡回プログラムが起動し、最初のカテゴリの URL 群を入力する。Perl の LWP (Library for

WWW access in Perl) によって、サーバと HTTP 回線接続する。次に入力された URL に対してヘッダ情報をリクエストする。取得したヘッダ情報の中から更新日時を抜き出し (図 1:A)、前回調査した結果と比較する。しかし、更新日時が何らかの理由で取得できなかった場合は、ヘッダ情報に含まれる“ファイルの大きさ”を調べる。ファイルの大きさに変動があれば、更新されたと判断することにする。

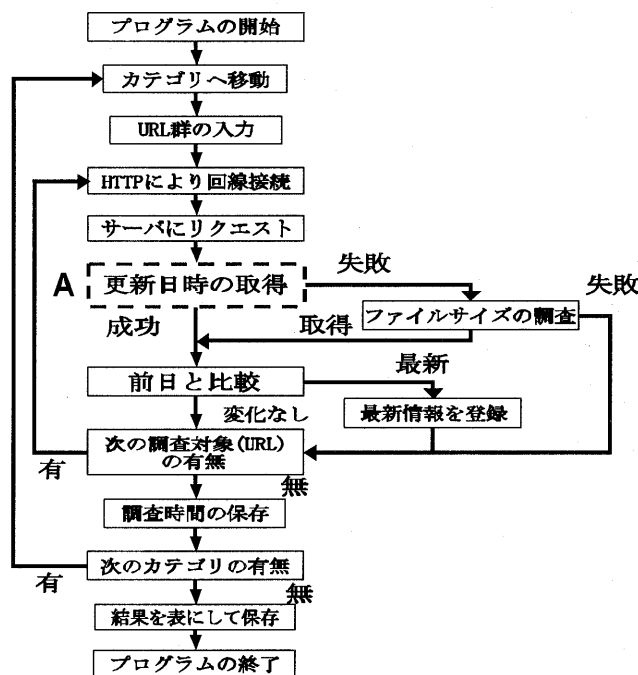


図 1 巡回ロボットのフローチャート

本ロボットは、Pentium III (700MHz) AT 互換パソコンの Windows2000 上で動作し、本学のネットワークからインターネット接続した。

3. 適切な巡回時刻の調査

WWW ページの更新頻度を知るためには、ロボットを毎日巡回させる必要がある。そのた

Research on renewal of WWW pages of a regional community by a traveling robot

Kunihiro Yamanaka*, Masatomo Shirakawa†, and Yasuo Furukawa†

*Department of Ecological Engineering, Toyohashi University of Technology, Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, Japan

†Research Center for Future Technology, Toyohashi University of Technology

めに、ロボットが最も短い時間で巡回できる時刻を選定する調査を行った。

約1ヶ月間、毎日2時間ごとにロボットを巡回させ、所要時間を測定した。図2に各調査時刻における平均の所要時間と標準偏差を示す。午前4時から6時頃が最も巡回時間が短くなるのが分かり、巡回開始を5時とした。

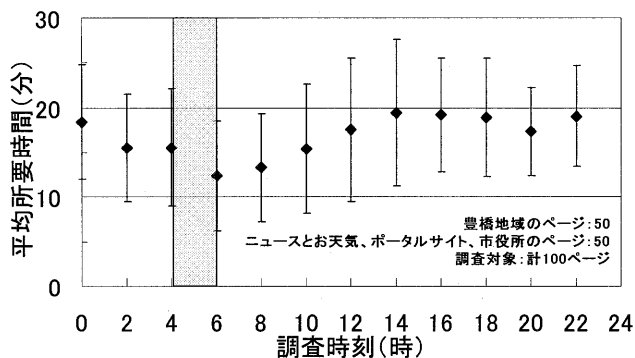


図2 時刻調査結果 (1ヶ月平均)

4. 更新調査の結果と考察

図3は10、11月の平均の更新回数(月あたり)と該当するページ数の割合を、図4は豊橋地域のページから住居、教養、個人のページに関するカテゴリに注目し、同様に示したものである。地域のWWWページでは0~2回の更新回数の占める割合が75%を超えていて、総じて更新頻度が低い。しかし、季節ごとの行事や連絡事項の多い教養のカテゴリでは、若干ではあるが、更新頻度が高いように思える。個人のページのカテゴリでは、更新回数が多いものも散見される。

県庁所在市役所のページでは、更新回数が低いところもあるが、10回以上の更新は全体の21%で他の2つに比較して良く更新されている。ニュースやお天気のページは、毎日更新されると考えられる。しかし、調査結果では、更新回数が0または1回のページが49%あった。これらのページは多層ファイル構造になっており、下層ファイルの情報が更新されても、トッ

ページのヘッダ情報に反映されていないためと考えられる。この場合、調査URLを下層ファイルのページに変更する等の処理を施す必要がある。

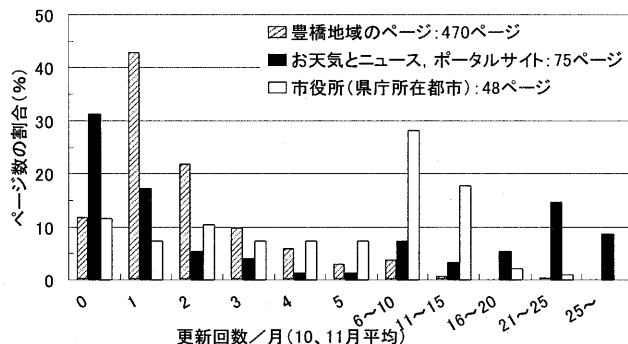


図3 更新回数に対するページ数の相対頻度

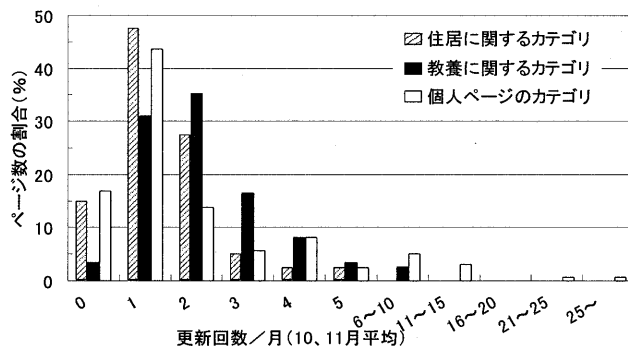


図4 更新回数の相対頻度 (カテゴリ別)

5. おわりに

巡回ロボットを作成し、地域WWWページの更新頻度を調査した。地域のインターネット利用を推進する上で有用である。課題として、下層ファイルの更新がトップに反映されない多層ファイル構造のページに対して、巡回ロボットを対応させる必要がある。一方、更新情報は重要であるから、トップのヘッダ情報等に反映させることもページ管理の上では必要である。

本研究は電気通信普及財団の助成を受けたことを記して感謝する。

参考文献

[1]古川：地域のWWWページのディレクトリ集(コミュニティ・ナビゲータ)の構築, 情報処理学会第60回全国大会, 4-243 (2000).