情 報 学 基 礎 51－4
自 然 言 語 処 理 127－4
（1998. 9. 17）

適応型 WWW 自動検索手法

塩見隆一 徳田 克己 青山 昇一 柿ヶ原 康二
松下電器産業（株） マルチメディア開発センター

効率的に Web ページを検索収集する適応型 WWW 自動検索手法を提案する。本手法は、検索情報として検索起点、検索キーワード、検索履歴を保持し、検索情報に基づいた検索アルゴリズムと検索情報を学習する学習アルゴリズムで構成される。検索アルゴリズムは、リンク集を優先的に検索するリンク集優先検索アルゴリズムを採用し、検索起点から検索キーワードを利用して効率よく自動的にネットサーフィンを行い、検索対象 Web ページへのハイパーリンクを収集する。学習アルゴリズムは、検索情報を学習し、幅広く新しい Web ページの検索を可能にする。本手法を実行する検索エージェントを開発中の WWW 閲覧支援システムに組み込み評価を行なった。

A Method of Adaptive and Automatic Infomation Retrieval on the WWW

Takakazu SHIOMI Katsumi TOKUDA Shoichi AOYAMA
Kouji KAKIGAHARA

Multimedia Development Center, Matsushita Electric Industrial Co., Ltd.

We suggest a new method of adaptive and automatic information retrieval on the WWW. The method maintains the retrieval information that consists of initial URLs, retrieval keywords and retrieval history. The method consists of a new retrieval algorithm which uses the retrieval information and a learning algorithm which updates the retrieval information. Our new retrieval algorithm retrieves link pages earlier than other web pages, netsurfs using retrieval keywords automatically and gets hyperlinks which correspond to the desired web pages. The learning algorithm updates the retrieval information and the retrieval algorithm can always retrieve new and many hyperlinks. We implemented and evaluated an agent software which executes our new method on the WWW browsing aided system.

1 はじめに

我々は個人がインターネット上の Web ページを閲覧するための WWW 閲覧支援システムの開発を行なっている。WWW 閲覧支援システムは、通信コストの削減と情報機器に不慣れなユーザーが簡単に使用できることを目標とし、

- 読みたい Web ページはローカルディスクに保存し、通信回線を切断してから内容を読む
- キーボード操作を少くする

を、基本方針としてシステム開発を行なっている。

図 1は、WWW 閲覧支援システムの構成図である。WWW 閲覧支援システムは、短時間で Web ページを保存するためのオートパイロット機能とブラウザ保存機能を用意している。オートパイロット機能は予め登録された保存起点の URL と保存範囲に従って Web ページを収集する。ブラウザ保存機能は、ブラウザが表示している Web ページを保存する。両機能とも、Web ページを構成する HTML テキスト及び画像データをファイリング機能を通してデータベースに保存する。

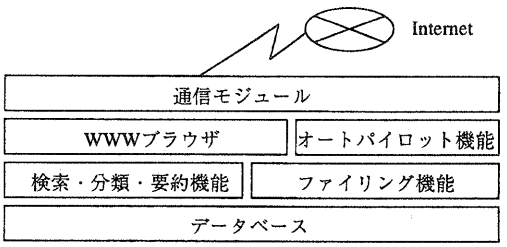

図 1: WWW 閲覧支援システム構成図

さらに、WWW 閲覧支援システムは、保存した Web ページの閲覧を支援するために、データベースの検索・分類・要約機能を備える。情報機器に不慣れなユーザーが簡単にデータを検索できるように、Web ページの自動分類結果を利用したメニュー検索機能 [2, 3] を用意している。また、迅速な内容把握のための要約機能 [4] も提供している。

しかし、ネットサーフィンを行ないブラウザ保存機能を使って興味のある Web ページを収集する場合、通信コストが削減されないことがわかった。理由は、Web ページの内容をある程度読まないと、ネットサーフィンできないからである。ネットサーフィンをせず、検索エンジンサイトの検索結果をオートパイロットする方法もある。しかし、検索結果には不要なものも多く含まれ、通信コストは増大し、ローカルディスクを浪費する。また、検索結果はほとんど変化しないので、毎日 Web ページを閲覧するには不向きである。

そこで、我々は上記問題を解決するための適応型 WWW 自動検索手法を提案する。また、適応型 WWW 自動検索手法を実行するエージェントソフトを WWW 閲覧支援システム上に実現し評価実験を行なったので、その結果を報告する。

関連研究

Fish search[5] は、Web ページの検索合致度によって Web ページ中のハイパーリンクを評価し、関連 Web ページを検索する。ハイパーリンク先の Web ページは似通っていることを前提に考案された手法で、実際のシステムで実証されている。

WebClawler[6, 7] はサーバー上のインデックスから Web ページを検索した後、Fish search 同様ハイパーリンクを辿ることで、関連 Web ページを検索する。検索の際、ハイパーリンクのアンカー文字列を検索キーワードとシソーラスを用いて評価する。

qbook[8] は、ハイパーリンクを評価に、キーワードによる Web ページの評価と、収集した Web ページ中での URL の出現頻度による評価を利用する。

我々の手法では、リンク集[1]へのハイパーリンクを優先的に検索するリンク集優先検索アルゴリズムを採用し Web ページ収集をより効率化している。また、検索のための情報を学習しユーザーへの適応を可能とし、検索の度に、未知の興味のある Web ページを閲覧できる点が既存手法と異なる。

2 適応型 WWW 自動検索手法

適応型 WWW 自動検索手法の基本アイデアを説明する。本手法は、図 2のように構成される。

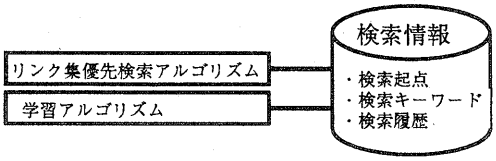

図 2: 適応型 WWW 自動検索手法の 3 要素

2.1 検索情報

検索に必要な検索起点と検索キーワード、検索結果表示に用いる検索履歴を検索情報として保持する。

2.2 リンク集優先検索アルゴリズム

検索起点から検索キーワードを利用して自動的にネットサーフィンを行い、検索対象 Web ページへのハイパーリンクを収集する。ただし、リンク集への検索を優先し、検索効率の向上を目指す。図 3は、本アルゴリズムの基本アイデアの動作を表している。

1. 検索起点の HTML テキストを取得する。
2. 取得した HTML テキストから参照されるリンク集の HTML テキストを優先して取得する。
3. 未取得のリンク集がなければ、リンク集から参照される HTML テキストの 1 つを取得する。
4. リンク集が見つかれば優先的に取得する。

このように HTML テキストの取得を繰り返し、リンク集に記述されているハイパーリンクを収集する。

ハイパーリンク収集後、ハイパーリンク先の Web ページからユーザーが興味のある Web ページを一括取得し、ローカルディスク上のデータベースに保存する。もう 1 つの方法として、対話的な検索機能も用意する。収集したハイパーリンクをユーザーが興味のある順に提示する。ユーザーはハイパーリンクを選択することで興味のある Web ページを閲覧または保存する。以降、本稿ではユーザーと対話的に検索を行なう場合を想定し、議論を進める。

[1] Web ページへのハイパーリンクを集めた Web ページ

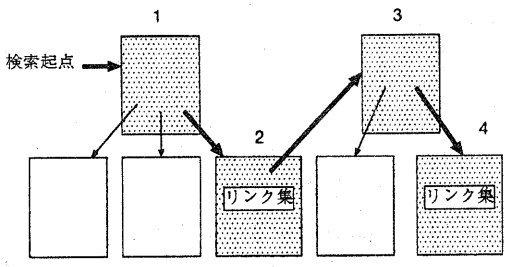

図 3: 検索手法の概念図

2.3 学習アルゴリズム

検索起点・検索履歴・検索キーワードからなる検索情報を学習する。検索起点に検索で得られたWebページを追加することで、次回の検索の際、新しいWebページを検索できる。検索履歴としてユーザーに提示したハイパーリンクを学習することで、新しく検索したハイパーリンクを優先的に表示できる。ユーザーが選択したハイパーリンクのアンカー文字列及び説明文字列 (3.2,3.5 参照) から検索キーワードを自動抽出し、以降の検索に利用することで、よりユーザーの興味あるWebページを検索することができる。

3 Webページの調査

リンク集優先検索アルゴリズムを実現するため、実際のWebページを調査し、検索戦略を抽出した。

3.1 調査対象

表1に、調査を行なった日本国内のサイト数[2]と各サイトから人手で抽出したリンク集数をまとめた。企業サイトにリンク集が少いことがわかる。

サイト種別	個人サイト	企業サイト	合計
サイト数	83	44	127
リンク集数	39	8	47

表 1: 調査対象

3.2 ハイパーリンクのHTML記述

ハイパーリンクに関する用語を定義する。

一般的に、ハイパーリンクは図4(1)のように、開始タグ、アンカー文字列、終了タグで構成される。開始タグ中には、ハイパーリンク先のURLが記述される。WWWブラウザは、アンカー文字列を下線

[2]本来「サイト」とは、WWWサーバーを指す。しかし、ここでは個人が作成したWebページの集合をサイトと呼ぶことにする。

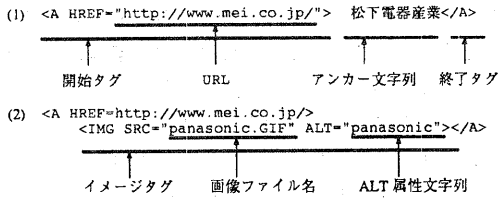

図 4: ハイパーリンクのHTML記述例

表示し、ユーザーがこれを選択すると開始タグ中に記述されたハイパーリンク先を検索して表示する。

また、図4(2)のように、アンカー文字列の代わりにイメージタグが記述される場合もある。イメージタグ中には表示する画像ファイル名、ALT属性文字列が定義されている。ALT属性文字列は、画像を表示できなかった時、代わりに表示される文字列で、省略可能である。

3.3 リンク集へのハイパーリンクの特徴

リンク集優先検索アルゴリズムを実現するには、リンク集へのハイパーリンクを自動認識しなければならない。そこで、リンク集へのハイパーリンク中の (1) アンカー文字列 (2)URL(3)ALT 属性文字列の特徴を調べた。図5は、調査結果をまとめたものである。

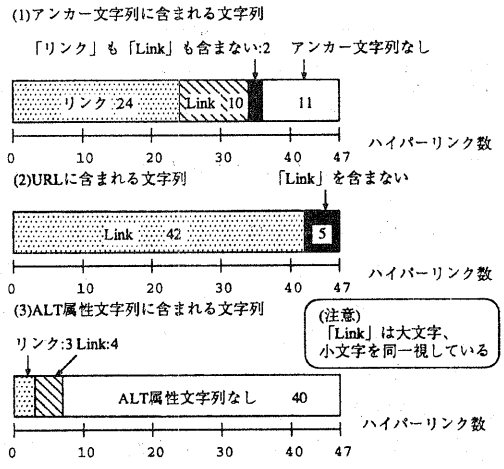

図 5: ハイパーリンクに含まれる文字列

3.4 リンク集が参照するWebページ

サイトに記述されている情報の主題と、そのサイトのリンク集に記述されているハイパーリンクの主題が一致しているのは、47サイトのうち7サイトだけであった。よって、リンク集のハイパーリンク

を無作為に選択すると検索効率が落ちることが明らかとなった。また、サイトによらず検索エンジンサイトや素材集サイトへのハイパーリンクが記述されている場合が多いことも判明した。

3.5 リンク集のハイパーリンクの説明文字列

ユーザーが興味のあるWebページへのハイパーリンクを抽出する一般的な方法として、ハイパーリンク中のアンカー文字列と、検索キーワード及びその類語を照合する方法がある。しかし、アンカー文字列は一般に短く、また、最近ではリンクバナーと呼ばれる画像データが用いられることが多くなっている。

一方、リンク集のハイパーリンクには、ハイパーリンク先のWebページの説明文字列（以降、「説明文字列」と記述する）が記述されていることが多い。アンカー文字列とともに、この説明文字列と検索キーワードを照合し、ハイパーリンクを選択することが有効と考えられる。

説明文字列を正しく抽出するため、リンク集内の説明文字列の記述方法について調べた。図6はハイパーリンクと説明文字列のHTMLテキスト内での記述順序について調べた結果である。説明文字列の90%以上がハイパーリンクの後に記述されている。

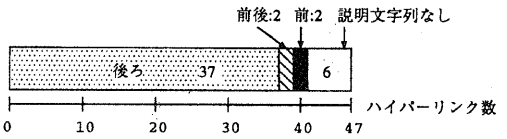

図6: ハイパーリンクに対する説明文字列の位置

また、説明文字列の終端を判別するため、ハイパーリンクと説明文字列を記述しているHTMLタグ構造も調べた。表構造、リスト構造が多く利用されていたが、説明文字列の終端を的確に判別できる特徴は得られなかった。

3.6 検索戦略

調査結果から、以下の検索戦略を検索アルゴリズムに採り入れることにした。

3.6.1 リンク集へのハイパーリンクの抽出

アンカー文字列、URL、ALT属性文字列に「リンク」または（英大文字小文字同一視で）「link」が含まれる時、リンク集へのハイパーリンクとする。

3.6.2 除外リストの採用

検索エンジンサイトなどを除外リストに登録し、収集するハイパーリンクから除外する。

3.6.3 説明文字列の抽出

ハイパーリンクの後続に記述される文字列を説明文字列として抽出する。抽出した説明文字列は検索キーワードと比較を行ない、その結果をハイパーリンク選択に利用する。また、説明文字列中からキーワードを抽出し学習する。

4 リンク集優先検索アルゴリズム

Webページを調査した結果を基に、以下のアルゴリズムを決定した。図7の具体例とともに説明する。

[ステップ0] 除外リストの作成と検索情報のセット

取得禁止のWebページ、WWWサイトを除外リストに登録する。検索キーワードと検索起点を検索情報に登録する。収集したハイパーリンクのURL、アンカー文字列、説明文字列、ALT属性文字列、取得フラグ、リンク集フラグ、リンクスコア、ページスコアを格納するハイパーリンクテーブルを用意する。

[ステップ1] ハイパーリンクテーブルの初期化

ハイパーリンクテーブルに検索起点のURLを登録する。取得フラグはTRUEにセットする（図7(1)）。

[ステップ2]HTMLテキスト取得と評価

ハイパーリンクテーブルから、取得フラグがTRUEのURLを取り出し、ネットワークを通して対応するHTMLテキストを取得する。取得が終了したら取得フラグをDONEにセットし、ページスコアを計算してハイパーリンクテーブルに登録する（図7(2)）。ページスコアは次式で計算する。

$$\text{ページスコア} = \frac{HTML\text{テキスト中の検索キーワード数}}{HTML\text{タグを取り除いたファイルサイズ}}$$

[ステップ3] ハイパーリンクの抽出と評価

取得したHTMLテキストからハイパーリンク(URL、アンカー文字列、ALT属性文字列、説明文字列)を抽出し、ハイパーリンクテーブルに登録する。ただし、除外リストに該当するハイパーリンクは登録しない。取得フラグはFALSEにセットする。リンク集へのハイパーリンクと判定した場合は、リンク集フラグをTRUEにセットする。リンクスコアを計算してハイパーリンクテーブルに登録する（図7(3)）。リンクスコアはアンカー文字列と説明文字列に出現する検索キーワード数である。

[ステップ4] ハイパーリンク選択

取得フラグがFALSEのハイパーリンクから、以下の選択基準の中で上位の選択基準に該当するハイパーリンクを1つ選択し、取得フラグをTRUEにする（図7(4)）。

選択基準1：リンク集へ
リンク集フラグがTRUEで、抽出元Webページのページスコアが最も高いハイパーリンク。

選択基準2：リンク集から外部サイトへ
抽出元Webページのリンク集フラグがTRUEで、リンクスコアが最も高い外部サイトへのハイパーリンク。

選択基準3：リンクスコア
リンクスコアが最も高いハイパーリンク。ただしリンクスコア0のハイパーリンクは選択しない。

選択基準4：抽出元Webページのページスコア
抽出元Webページのページスコアが最も高いハイパーリンク。

上記、基準はネットワークに大きな負荷を与えない設定となっている。また、特定のサイトに負担をかけないため、同一サイトへのハイパーリンクは6回以上選択しない。

[ステップ5] ハイパーリンクの収集

以下の条件を満たすまで、ステップ2～4を繰り返し、ハイパーリンクを収集する（図7(5)）。尚、括弧内は実際に用いている値で、ネットワークに大きな負荷を与えないようにしている。

- 有効ハイパーリンクを一定数(10)以上収集した。
- HTMLテキストの取得回数が一定数(10)に達した。
- 選択基準を満たすハイパーリンクがない。
- 選択基準4でハイパーリンクを選択した回数が一定数(5)を越えた。

[ステップ6] 提示

リンクスコアが0より大きいハイパーリンクを以下の基準で順に表示する。

表示基準1：抽出元Webページがリンク集。
表示基準2：過去に提示された回数が少い。
表示基準3：リンクスコアが高い。

5 WWW閲覧支援システム

学習アルゴリズムを説明する前に、適応型WWW自動検索手法を実行する検索エージェントが組み込まれたWWW閲覧支援システムについて説明する。

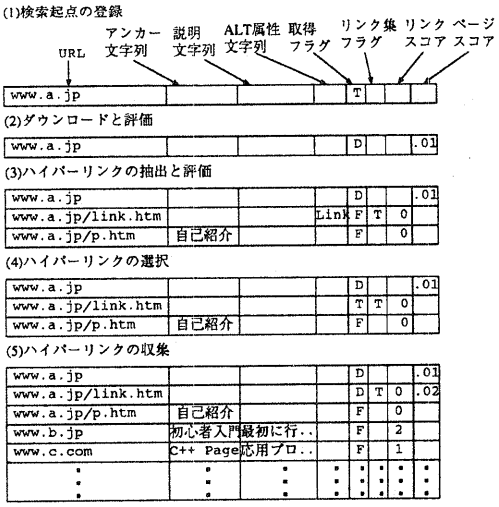

図7: 検索アルゴリズムとハイパーリンクテーブル

図8は、擬人化されたエージェントが一覧表示されている状態を表す。ユーザーは5つのエージェントに対して、それぞれ検索起点、検索キーワードを設定し、自動検索を指示することができる。

図8: WWW閲覧支援システム：エージェント一覧

各エージェントはユーザーからの指示で自動検索を行ない図9のように検索結果を表示する。検索後、画面右上の「更に検索」ボタンをクリックすると、その時点のハイパーリンクテーブルの状態から追加検索が実行される。追加検索は何度でも実行可能である。画面右上の「エージェント一覧」ボタンをクリックすると図8のエージェント一覧画面に戻り、ハイパーリンクテーブルの内容は破棄される。

ここで、最初の検索からエージェント一覧画面に戻ってハイパーリンクテーブルが破棄されるまでを**検索ターム**と呼ぶことにする。

図9の画面左には表示基準に基づいて、上位30個の検索されたハイパーリンクのアンカー文字列が一覧表示される。アンカー文字列をクリックすると、エージェントからのメッセージとして画面中央上にハイパーリンクの説明文字列が表示される。ダブル

クリックすると対応するハイパーリンク先の Web ページが画面中央の WWW ブラウザに表示される。

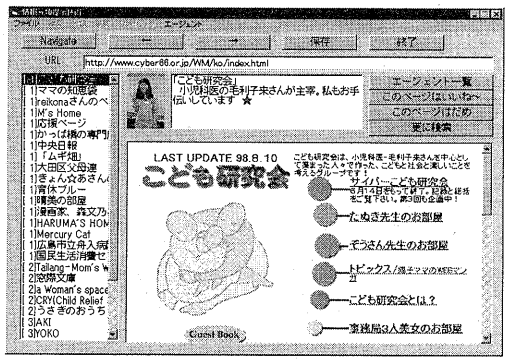

図 9: WWW 閲覧支援システム：検索結果表示

6 学習アルゴリズム

6.1 検索履歴の学習

ハイパーリンクの検索履歴は、以下の 3 つの情報で構成される。

1. URL
   ユーザーに提示されたハイパーリンクの URL
2. 表示回数
   検索結果としてユーザーに提示した延べ回数
3. 非表示連続回数
   検索結果として提示されなかった連続回数

学習は、検索タームが終了時に以下の通り行なう。

学習条件		検索履歴情報		
検索結果提示	検索履歴	URL	表示回数	非表示連続回数
○	既存		+1	0
○	なし	新規登録	1	0
×	既存			+1

なお、非表示連続回数が一定回数を越えたものは削除する。バッファが一杯になった時は、非表示回数が最多のものを削除する。検索履歴の表示回数は、検索結果の提示順の決定に利用される。

6.2 検索起点の学習

ユーザーから与えられた検索起点とは別に、以下の検索された Web ページを検索起点として学習する。

- 有効ハイパーリンクが抽出できた Web ページ
- ユーザーが選択したハイパーリンクが抽出されたリンク集

学習は、検索タームが終了した際に行なう。検索起点には、重要度が与えられる。重要度は登録時に、その Web ページから抽出された有効ハイパーリンク数が初期値として与えられる。重要度は毎学習毎に減点され、検索起点のバッファが一杯になった時、重要度の低い検索起点は削除される。

6.3 検索キーワードの学習

ユーザーが選択したハイパーリンクのアンカー文字列と説明文字列からキーワードの学習を行なう。学習したキーワードをトピックキーワードと呼び、ユーザーが与えたキーワードとは区別する。学習はユーザーがハイパーリンクを選択し Web ページを表示後、次の検索が行なわれる前、もしくはエージェント一覧画面に戻る前に行なう。トピックキーワードは重要度を持ち学習の度に一定値を減点する。

キーワードは、字種切りで 2 文字以上の漢字、カタカナ、アルファベットを抽出したあと、不要語を取り除いて抽出する。以下の 2 条件のいづれかを満たすキーワードをトピックキーワード候補として抽出する。

- ユーザーが選択した 2 つ以上のハイパーリンクから抽出された。
- ユーザーが選択したハイパーリンクから抽出され、ユーザーが選択しなかったハイパーリンクからは抽出されない。

トピックキーワード候補は既存のトピックキーワードと比較を行なう。一致した時は、重要度にトピックキーワード候補が出現したハイパーリンク数が加算される。不一致の時はトピックキーワード候補がトピックキーワードとして登録され、重要度には規定値にトピックキーワード候補が出現したハイパーリンク数を加えた値が与えられる。トピックキーワードを格納するバッファが一杯になった時は、重要度の低いトピックキーワードが削除される。

なお、ページスコアやリンクスコアを計算する際、ユーザーが入力した検索キーワードとトピックキーワードは同等に扱う。

7 評価実験

7.1 検索アルゴリズムの評価

検索実験を行ない、リンク集優先検索アルゴリズムの効果を調べた。

7.1.1 従来手法

比較する従来検索手法は、リンク集優先検索アルゴリズムの [ステップ 4] ハイパーリンク選択において、以下の選択基準を使用するものとした。

選択基準 1：外部サイトへ
リンクスコアが最も高い外部サイトへのハイパーリンク

選択基準２：リンクスコア

リンクスコアが最も高いハイパーリンク。ただしリンクスコア０のハイパーリンクは選択しない。

選択基準３：抽出元 Web ページのページスコア

抽出元 Web ページのページスコアが最も高いハイパーリンク。

7.1.2 検索条件

検索起点と検索キーワードは、ユーザーが集めたブックマークを用いた。表2は用いた検索キーワードと検索起点をまとめたものである。学習機能はオフとし、検索終了条件を HTML テキストの取得回数５０回のみとした。なお、従来アルゴリズムでは１サイトの取得回数上限を取り除いた。

No	キーワード	検索起点数	
		個人サイト	企業サイト
1	野球	2	1
2	カープ	2	0
3	競馬	1	0
4	テレビ, ラジオ	0	2
5	子供, こども, 幼稚園	1	0

表 2: 実験に用いた検索キーワードと検索起点

7.1.3 検索結果

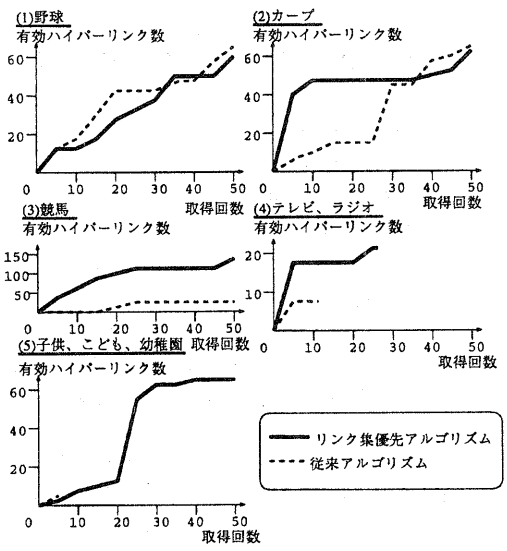

図 10: 取得回数と検索した有効ハイパーリンク数

図 10は、HTML テキストの取得回数に対して、検索で得られた有効ハイパーリンク数の変化をまとめたグラフである。折れ線が途中で途切れているものは、取得する HTML テキストがなくなり、検索が打ち切られたことを表している。なお、実験番号2,3 の従来アルゴリズムでは、１サイトの取得回数上限５を設定すると、取得回数１０回以内で検索が終了してしまう。

7.2 検索起点と検索履歴の学習の効果

新規の Web ページが見つかるまで追加検索を行なう検索タームを繰り返し、検索起点の自動学習について調べた。記憶できる検索起点数の上限はユーザー設定と合わせ５とした。検索起点と検索キーワードは表2の１を用いた。表 3は、各検索タームの検索結果と学習結果をまとめたものである。

検索ターム回数	追加検索回数	新規 Web ページ数	検索起点学習数
1	0	12	1
2	1	11	2
3	1	6	2
4	1	4	0
5	2	11	2
6	0	6	0
7	3	1	1
8	1	8	1
9	0	2	0
10	1	1	1

表 3: キーワード「野球」における検索起点の学習結果

7.3 検索キーワードの学習の効果

検索キーワード学習の効果を調べるため、収集される有効ハイパーリンク数について調べた。１回の自動検索後に、提示されたハイパーリンクを２つ選択し、その後に収集されたハイパーリンク数を学習の有無で比較した。検索起点と検索キーワードは表2の３を用いた。

図 11は、HTML テキストの取得回数に対して、検索で得られた有効ハイパーリンク数の変化をまとめたグラフである。

8 考察

我々が提案するリンク集優先検索アルゴリズムは従来手法に比べ同等以上の性能を有することがわかった。従来アルゴリズムでは、同一サイトからの HTML テキスト取得数の上限を設定すると、外部サイトへのハイパーリンクを発見できずに検索を終了してしまうことがあるが、リンク集優先検索アルゴリズムは、リンク集を優先検索することでこの問

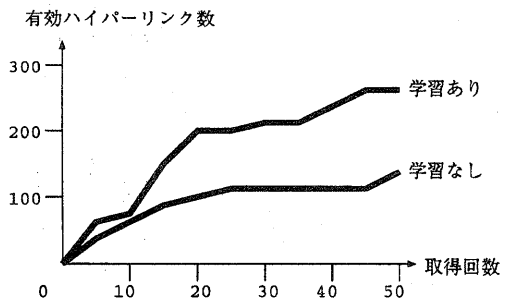

図 11: 検索キーワード学習の効果

題を回避できている。図 12は、表 2の 3の検索実験において、2つのアルゴリズムが取得した HTML テキストを示している。図中□、○が 1つの Web ページを表し、同一サイトの Web ページは四角で囲んでいる。中の数字が何回目に取得されたかを表す。→は Web ページの参照関係、○はリンク集、破線の○はリンク集と判断して取得したがリンク集ではなかった Web ページである。リンク集優先検索アルゴリズムは、従来アルゴリズムに比べ、多様なサイトへネットサーフィンできており、基本アイデア通りの動作を実現していることがわかる。

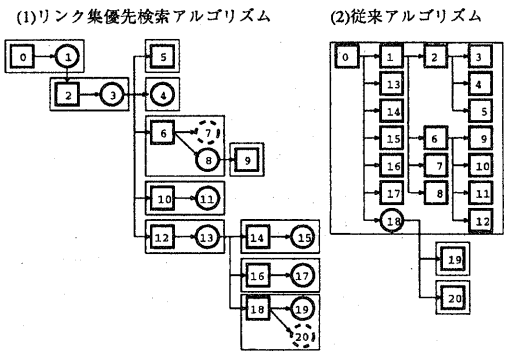

図 12: キーワード「競馬」にける検索過程の比較

検索起点を学習することで検索毎に未知の Web ページを検索できることも確認できた。検索キーワードの学習により、検索対象外と判断していた Web ページへのハイパーリンクも、有効ハイパーリンクとして多数抽出できるようになった。正確な数値は算出していないが、再現率は向上している。ただし、有効ハイパーリンクが検索目的に合致する適合率は低下しているはずである。

9 まとめ

提案する適応型 WWW 自動検索手法は、

- ユーザーが興味ある Web ページへのハイパーリンクを効率良く集められる。
- 学習により、検索するたびにより多くの新しい検索結果が得られる。

ことを確認できた。

今後は、非対話における Web ページの一括取得保存機能を実現する予定である。また、学習アルゴリズムの評価はまだ十分とはいえないため、被験者による評価を行ない学習アルゴリズムの改良を進める予定である。

参考文献

[1] http://town.hi-ho.ne.jp/market/result/internet/index.htm「インターネットユーザー調査結果」

[2] 徳田克己，塩見隆一，青山昇一，柿ヶ原康二:" 分類パターンを用いた文書データの自動分類法," 情報処理学会研究会報告，Vol.NL 123-9, pp. 65-72, 1998.

[3] 塩見隆一，徳田克己，青山昇一，柿ヶ原康二:" シソーラスを用いた文書データの自動分類法," 情報処理学会研究会報告，Vol.NL 117-14, pp. 99-104, 1997.

[4] 塩見隆一，徳田克己，青山昇一，柿ヶ原康二:" 視点を考慮した文書要約手法の提案," 第 5 6 回情報処理学会全国大会講演論文集，3-104～105, 1998.

[5] P.M.E. De Bra, R.D.J.Post: "Information retrieval the World-Wide Web," Proceedings of the First International Woeld Wide Web Conference, 1994.

[6] Pinkerton, Brain :"Finding What People Want:Experiences with the WebClawler," THe 2nd International Woeld Wide Web Conference '94:Mosaic and the Web, 1994.

[7] 大野浩之監訳:" インターネットエージェント," インプレス, 1998.

[8] 来住 伸子:" 分野を特定した自動収集による WWW 情報検索," 情報処理学会研究会報告, Vol.NL 124-12, pp. 87-94, 1998.