

石濱雄志 伊藤吾一 阿部倫之 服部進実

金沢工業大学 情報工学科

1 はじめに

現在、膨大な量の Web 文書の中から有益なものを探し出す手段として、WWW 検索システムが多数提供されており、検索キーワードを入力すると、そのキーワードを含む文書を容易に発見することができる。

しかし、キーワードに基づく WWW 検索システムでは、探している情報とは関係のない文書が検索結果に現れる場合が多く存在するため、有益な情報の発見を困難なものにしている。

本稿では、検索行為の類似する他者の操作履歴を用いて、検索範囲の絞り込みとユーザが関心を持ちそうなページを推奨するシステムを提案する。このシステムは分散協調型の WWW 検索システムとして実装しており、本稿ではそのシステム概略について述べる。

2 機能概要

本システムでは、通常のキーワード検索に加え、ユーザの操作履歴を用いて以下に示した 4 つの機能を協調検索システムとして実現している。

- (1) ユーザが関心を持ちそうなページの推奨
- (2) 本人が過去に閲覧したページを対象とした検索
- (3) 類似ユーザが過去に閲覧したページを対象とした検索
- (4) 本人が登録したサイトを対象とした検索

3 システム構成

本システムの構成を図 1 に示す。

CGI : 個々のユーザを識別できるようにユニークな ID を生成し、Cookie を使用して割り当てる。また、ユーザからの検索要求をコーディネータに送り、その結果を返す。ユーザが検索結果のどのページを閲覧したかをモニタし、閲覧したページの URL をコーディネータに送る。

コーディネータ : 検索要求を検索サーバに伝え、検索結果を CGI に返す。また、検索式と閲覧 URL をユーザ推奨機構に送る。必要であればユーザ推奨機構からユーザが過去に閲覧したページの URL を所得し、検索式と共に検索サーバへ送る。要求

に応じて他の検索システムのコーディネータと協調し、コーディネータ間で検索結果やユーザ・プロファイルをやりとりする。

検索サーバ : 検索要求に一致したページ情報を返す。ページ情報には URL、タイトル、概要などが含まれる。また、URL を複数指定することで、指定した URL 群を対象とした検索機能を備えている。URL の指定方法には完全一致と前方一致がある。

ユーザ推奨機構 : ユーザが入力した検索式からユーザ・プロファイルを生成する。また、ユーザが閲覧したページの URL を蓄積する。ユーザ・プロファイルから傾向の類似するユーザを発見し、類似ユーザ群の閲覧 URL を用いてページの推奨を行う。

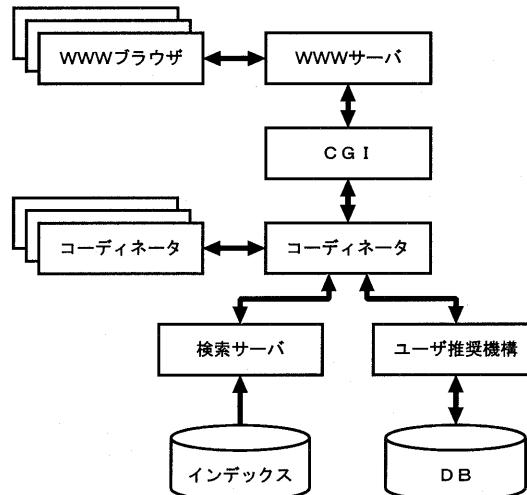


図 1: システム構成

4 システムの動作概要

$tf \cdot idf$ 法を用いて複数サーバからの検索結果を正しくスコアリングするには 2 パスのスコア付けが必要である [1]。本システムではページのスコアリングに $tf \cdot idf$ 法を使用しており、複数の検索サーバから検索結果を得る場合は、2 パス方式よりも通信量を削減した 3 パス方式を使用し、単一の検索サーバから検索結果を得る場合は 1 パス方式を使用している。

図 2 に検索時のプロトコルを、図 3 にページ推奨時のプロトコルをそれぞれ示す。

図 2、図 3 はローカル・システム内のプロトコルであるが、コーディネータ間で協調して他の検索サーバから検索結果を得る場合にも、これと同じことをコーディネー

タ間で行う。ただし、他のシステムから推薦ページを得る場合には、ローカル・システムのユーザ推薦機構からユーザのプロファイルを取得し、これを他のコーディネータに送る。

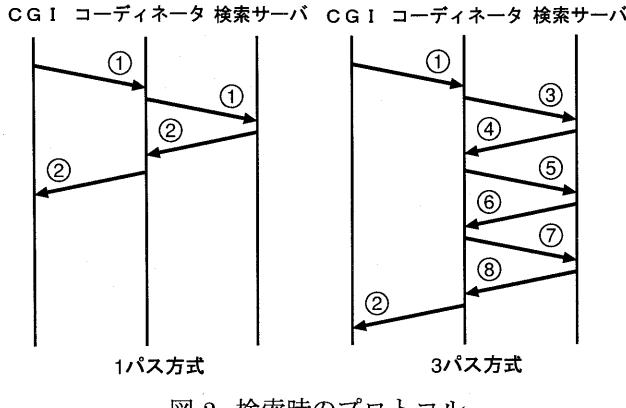


図 2: 検索時のプロトコル

- ① 検索式、取得開始位置、取得最大数
- ② 検索結果 (URL、タイトル、概要など)
- ③ 検索式内で使用している単語のリスト
- ④ 全文書数、単語ごとのヒット件数
- ⑤ 検索式、取得開始位置、取得最大数、全文書数の合計、単語ごとの合計ヒット件数
- ⑥ ヒット・ページのリスト (URL、時刻、スコア)
- ⑦ URL リスト
- ⑧ ヒット・ページの詳細 (URL、タイトル、概要など)

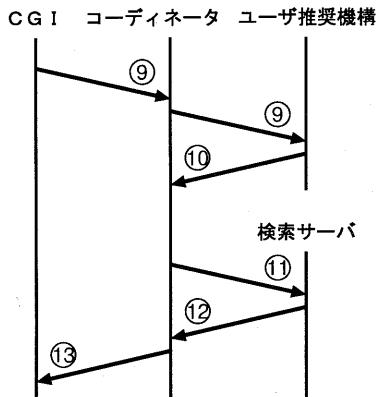


図 3: ページ推薦時のプロトコル

- ⑨ ユーザ ID、要求ページ数
- ⑩ 推奨ページのリスト (URL、スコア)
- ⑪ URL リスト
- ⑫ 推奨ページの詳細 (URL、タイトル、概要など)
- ⑬ 推奨ページ

5 ユーザ推薦機構

ユーザ推薦機構の構成を図 4 に示す。

履歴管理部：ユーザが閲覧したページの URL を、ユーザ ID と対にして閲覧履歴に格納する。要求があればユーザが過去に閲覧したページの URL を返す。

特徴抽出部：ユーザが入力した検索式を用いてユーザ・プロファイルを更新する。ユーザ・プロファイルは、語に対する重みを要素とするベクトル \mathbf{u} で、検索式から生成した 2 値ベクトル \mathbf{q} を用いて式(1)により更新する。 α は割引率、 β は学習率である。

$$\mathbf{u}(t+1) = \alpha \cdot \mathbf{u}(t) + \beta \cdot \mathbf{q} \quad (1)$$

ユーザ照合部：検索傾向の類似するユーザを発見する。2 ベクトル $\mathbf{u}_i, \mathbf{u}_j$ の類似度は式(2)により求める。

$$Sim(\mathbf{u}_i, \mathbf{u}_j) = \frac{\mathbf{u}_i \cdot \mathbf{u}_j}{|\mathbf{u}_i| \cdot |\mathbf{u}_j|} \quad (2)$$

ページ・スコアリング部：複数の類似ユーザが閲覧した URL 集合に対してスコアリングを行う。各類似ユーザの閲覧履歴に出現した URL のスコアを式(3)により求め、類似ユーザ間で URL の重複があったものはその合計を求める。C は定数、S は減少比、k は閲覧順序が過去に向うほど大きくなる。 $Score = C \cdot Sim(\mathbf{u}_i, \mathbf{u}_j) \cdot \frac{1}{S^k}$ (3)

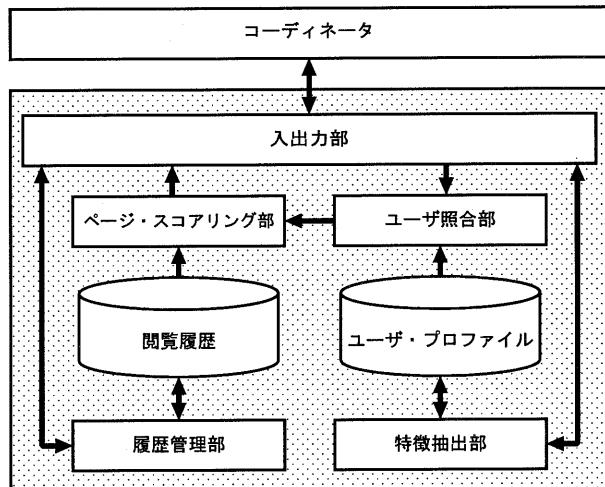


図 4: ユーザ推薦機構

6 おわりに

本稿では、検索行為の類似する他者の操作履歴を用いて、検索範囲の絞り込みとページ推薦を行うシステムの概略について述べた。本システムはこれらの機能を分散環境上で実装している。今後は、システムの有効性について評価を進めていく予定である。

参考文献

- [1] 國頭五郎, 相澤清晴, 森川博之, 青山友紀, “モバイルエージェントを用いた連携検索システムの実装に関する検討”, DICOMO 2000, pp.157-162, 2000