

フロー情報を対象にした情報検索システム (1)

4F-6

- 概要 -*

上田隆也 柴田昇吾 伊藤史朗 廣田 誠 池田裕治 藤田 稔
キヤノン(株) 情報メディア研究所

1 はじめに

インターネットをはじめとするコンピュータネットワークの整備により種々の情報がオンラインで利用可能になり、情報収集が容易になってきた。しかしその反面、情報の量が多すぎるため有用な情報を見逃す危険性も増大してきている（“情報洪水”の問題）。特に電子メール・NetNews・新聞等の“フロー情報”では情報の“新しさ”が重要であり、時間が経過してしまうことによりユーザにとっての情報の価値が失われてしまうことも少なくない。

こうした問題への対処として、我々は、フロー情報からの情報収集・整理を支援するシステムの研究開発を進めている。本システムは、フロー情報の中からユーザにとって有用な情報を選別し、大意・キーワード等を付与し、ユーザの設定した分類体系に振り分けて、適切なタイミングで提示する。柔軟な処理を行なうために、マルチエージェントモデルに基づくアーキテクチャを採用した。

本稿ではこのシステムの概要について述べる。

2 フロー情報に対する情報検索

情報は、利用する目的・形態から見ると、フロー情報とストック情報とに分けられる。

フロー情報 電子メール・NetNews・新聞のように刻々と更新される情報。日常的に利用する。情報の新しさが要求される。

ストック情報 過去の特許公報や技術文献のように蓄積されている情報。必要が生じた時に利用する。情報の新しさが要求されるとは限らない。

従来の情報検索は主にストック情報を対象にしてきた。ストック情報の中からユーザの検索意図に合致した情報を得る技術に関してはさまざまな研究が行なわれている [1]。

*An Information Retrieval System for Flow-type Information (1) - Overview -

UEDA Takaya, SHIBATA Shogo, ITOH Fumiaki, HIROTA Makoto, IKEDA Yuji and FUJITA Minoru (Media Technology Laboratory, Canon Inc.)

これに対して、フロー情報に対する検索は、ユーザの検索意図に合った情報を探すという点では、ストック情報に対する検索と共通の部分がある。しかし、情報の“新しさ”やユーザの興味の変化、ユーザが置かれている状況といった時間軸に関係した要素を考慮しなければならない点で異なっている。また、単に情報を検索するだけでなく、短時間での情報の把握・整理を支援することが必要である。したがって、フロー情報を対象にした情報検索の枠組みを新たに作る必要がある。

フロー情報に対する情報検索を特徴づける要因として、我々は次の三点が重要だと考えている。

1. 情報のライフサイクル

ユーザにとってその情報が有効である期間を判断し、その期間内にユーザに情報を提示する。

2. ユーザへの適応

各ユーザの興味に合わせて処理を行なう。興味が変わった場合はその変化に追随する。また、ユーザの状況にも適応し、ユーザにとって適切なタイミングで情報を提示する。

3. 複数情報源への対応

一般に複数の情報源から情報が得られるので、特定の情報源に依存せず、それぞれの情報を同様に扱えるようにする。また、情報間の重複・矛盾にも対処する。

これらの要因を考慮して、フロー情報に対する情報検索に求められる機能を考えると、次の三種類の機能を挙げることができる。

1. 数の削減

大量のフロー情報の中からユーザの興味に合った情報だけを選び出す。また、古くなった情報は提示せず、重複した情報はマージする。

ex. 情報フィルタリング、文書分類

2. 量の圧縮

選び出された情報の中から重要な部分を抜き出し、適当な形態に変換して提示する。

ex. 文章圧縮 (大意抽出、キーワード抽出)

3. 適時性

情報のライフサイクルとユーザの状況に応じて、適切なタイミングで情報をユーザに提示する。

3 システム構成

前節で述べた機能を実現するシステムを NeXT 上に試作している。まず数の削減と量の圧縮を実現する機能をインプリメントした。システム構成を図1に示す。

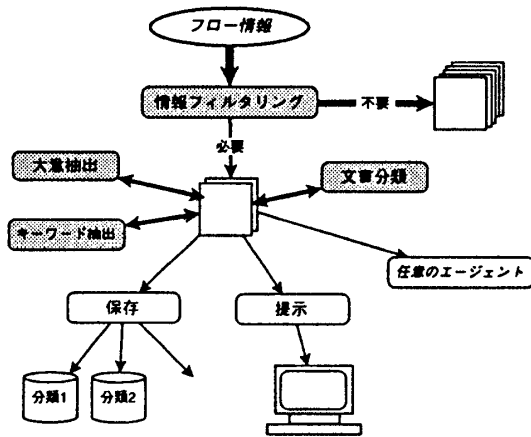


図1: システム構成

本システムは、フロー情報の中からユーザの興味に合致した情報だけを選び出す(情報フィルタリング)。なお今回はフロー情報として、オンラインデータベースからダウンロードした新聞記事(2紙)を用いた。それぞれの記事に、ユーザが設定した分類カテゴリを付与し(文書分類)、さらに大意やキーワードを付与する(文章圧縮)。大意やキーワードはそれだけで情報を伝えるというよりも、ユーザが記事の重要性を判断するために用いる。記事は分類カテゴリに応じて、すぐにユーザに提示したり、すぐに提示せずいったん保存したりする。また、他の処理に回す場合もある。

「ユーザへの適応」「複数情報源への対応」という観点からは、各ユーザ・情報源に合わせて動的に処理を行なう必要がある。そこで、システム構築の基盤として、マルチエージェントモデルに基づいたアーキテクチャーを採用し、各機能をエージェントとして実現した。

4 システムの特徴

以下に本システムのアーキテクチャーと機能の特徴について簡単に述べる。

4.1 アーキテクチャー [2]

マルチエージェントモデルに基づくアーキテクチャーを採用した。本システムにおけるエージェントは次の特徴を持つ。

- 実時間に応じて自律的に行動する。
- 外部の状況を把握する能力を持つ。
- 予め規定することができない状況を処理するための推論機能を持つ。
- モジュール同士が緩やかなプロトコルで通信する。

4.2 機能

4.2.1 文章圧縮 [3]

- 文章中の語句について、出現位置や頻度を考慮し、重要なものをキーワードとして抽出する。
- 付属語の出現頻度の偏り等から文章のタイプを判定する。
- 文章のタイプに応じて、適切な手法で大意を抽出する。例えば、社説や解説の場合、文章構造や特徴的な表現を利用して重要文を抽出する。

4.2.2 情報フィルタリング/文書分類 [4]

- ユーザが「必要/不要」の判断を下した文書、分類カテゴリを付与した文書を事例として、ユーザの必要としている情報やユーザの分類体系を学習する。
- 文書内容を適切に表現するために、分類カテゴリを特徴づける単語を事例から自動抽出する。また、単語の言語的な役割も考慮する。
- 学習結果に基づいて、入力文書から必要な文書だけを選別し、分類カテゴリを付与する。

5 おわりに

本稿では、フロー情報を対象にした情報検索システムの概要について述べた。今後は、

- 情報の緊急度・重要度の判定手法を検討し、“情報配信の適時性”を実現する
- ユーザの興味の変化に対応する

ことにより、情報洪水の緩和にとって必要な機能を充実させる。

参考文献

- [1] 小川：テキストデータベース技術の最近の動向，アドバンスト・データベースシステム・シンポジウム '93, 1993.
- [2] 伊藤他：フロー情報を対象にした情報検索システム(2) -アーキテクチャー-, 情報処理学会第50回全国大会 4F-7, 1995.
- [3] 柴田他：フロー情報を対象にした情報検索システム(3) -文章圧縮-, 情報処理学会第50回全国大会 4F-8, 1995.
- [4] 廣田他：フロー情報を対象にした情報検索システム(4) -文書分類-, 情報処理学会第50回全国大会 4F-9, 1995.