

6R-1

電子メール情報利用支援のための 統合利用モジュールの設計開発*

宮下一博 後藤文太郎§

北見工業大学 情報システム工学科¶

1. はじめに

インターネットの普及が進むにつれ、電子メールの利用形態の幅がますます広がってきている。PIM やグループウェアとの連携がその一例にあたり、そのような機能を持ったメールアプリケーションが現れてきている¹⁾。

一方、SPIRAL²⁾では、情報の収集と編集および発信を繰り返し行っていくというワークフローを WWW の利用形態に導入し、WWW における情報の再利用性を高めている。

また、オプトインメールやメールマガジンといった利用形態も近年盛んに使われており、これらは個人や団体にとって重要な情報収集、プッシュ型発信ツールになっている。

このようなメールの利用形態についても、上記のワークフローを導入することにより、電子メール情報の再利用性が高まると考えられる。そこで、本研究では、電子メール情報利用支援のためのモジュールを設計開発し、システムを構築した(図 1)。

本モジュールは、メールマガジン等の電子メール

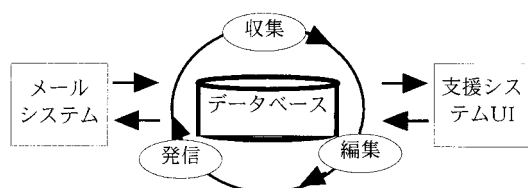


図 1: SPIRAL ワークフローにおけるメールの利用

閲読および新規作成時における過去の関連電子メール内容の情報統合を行う。これにより、旧来のメールには不可能な情報収集、編集手段が実現する。また、メールマガジン作成をターゲットとしたスケジュール設定による連載記事管理を行い、支援を試みた。これで、ユーザはプッシュ型の利点を活かした情報発信ができるようになった。

2. メールモデル化

Title 属性と Content 属性を持つオブジェクトを Article オブジェクトと定義する。Article オブジェクト O の Content 属性の属性値がテキストである時、Article オブジェクト O は Primitive であるという。Article オブジェクト O の Content 属性の属性値が Article オブジェクト O_1, \dots, O_n のリスト $[O_1, \dots, O_n]$ である時、Article オブジェクト O は Composite であるという。

次に、メールから Article オブジェクト O への変換を行うフィルタリングについて説明する。

3. フィルタリング

メールから Article オブジェクト O への変換を行うフィルタリングは次のように行われる。

- (i) メールがメールニュースでない場合
件名を Title 属性の属性値、本文を Content 属性の属性値とする、Primitive な Article オブジェクト O に変換される。
- (ii) メールがメールニュースの場合
件名を Title 属性の属性値、本文中の記事 X_1, \dots, X_n を Article オブジェクト O_1, \dots, O_n に変換し、そのリスト $[O_1, \dots, O_n]$ を Content 属性の属性値とする、Composite な Article オブジェクト O に変換される。

ただし、ここでメールニュースの本文中の各記事は、各記事のタイトルを Title 属性の属性値、記事内容を Content 属性の属性値とする Article オブジェクトに変換される。その際に、記事内容に URL が含まれていれば、その URL をアクセスして得られるデータを、記事中の URL と置き換える。

現在、フィルタリングの実装に関しては、メールニュース「INTERNET Watch Flash³⁾」への対応が完了している。メールニュースか否かは差出人メールアドレスで判断を行っている。記事の抽出は、飾り文字や空行などのテキスト構造に着目して処理を行っている。

4. 類似度計算によるリンク生成

Article オブジェクトからのキーワード抽出を行う。各 Article オブジェクトのキーワード間の類似度計算を行い、リンク生成を行った。

キーワードには Title 属性の属性値に含まれる固有名詞を選んだ。属性値からの名詞抽出には茶筌³⁾を用いた。

* Development of an Integrated E-mail Library Module for Information-Use Support

§ Kazuhiro MIYASHITA and Fumitaro GOTO

¶ Department of Computer Sciences, Kitami Institute of Technology

類似度計算にはベクトル空間法を適用し、重みはtf-idfを用いた^[1]。

5. モジュールアーキテクチャ

モジュールは以下のように4層から成り立っている。Articleデータベース（以下：DB）との連携や、情報統合およびArticleの管理を行う（図2）。

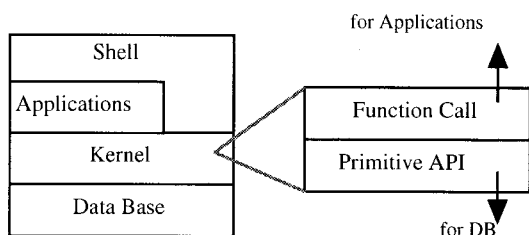


図2:モジュールアーキテクチャ

Kernel層はさらに2つに分かれる。Primitive API層はDBデータとArticle間の相互変換を担う。また、多彩なDBの実装を可能にする。Function Call層はオブジェクトに対する操作を担当し、特にComposite Articleの階層管理を実現する。

また、Applications層はArticleに対するメモ付けやメールのダウンロード、フィルタリングなどの高次の処理を行うKernel拡張機能の集合体である。拡張機能は比較的小さなモジュールとして存在し、動的に追加および削除が可能である。

そして、Shell層はユーザや他システムとのインタラクションを担う。TCP/IPに対応した物にすることで分散環境でのモジュールの使用が可能になる。

6. 電子メール情報利用支援システム

実現した機能の例として、現在閲読中のArticleに対する自動リンク付け処理を説明する。図2中のShellをCGIとした。図3にシステム概念図を示す。

自動リンク処理の流れは以下の通りである。

- (i) ユーザからのリクエストでCGIが起動される(図3(1))。
- (ii) CGIはユーザプロファイルを読み、今どのArticleを閲読しているのかを調べる(図3(2))。
- (iii) 上記手続きにより得た情報を元に、閲読Articleのkeywordを得て類似度計算を行い、リンクを生成する(図3(3,4))。
- (iv) 以上の出力結果をHTMLに整形しブラウザに返す(図3(5,6))。

実行画面を図4に示す。画面下部に閲読中Article、上部に類似Articleへのリンクが示されている。

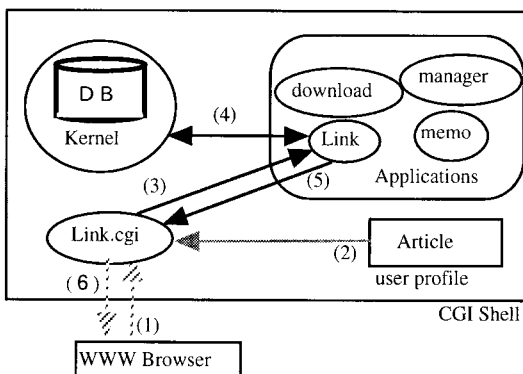


図3:システム概念図

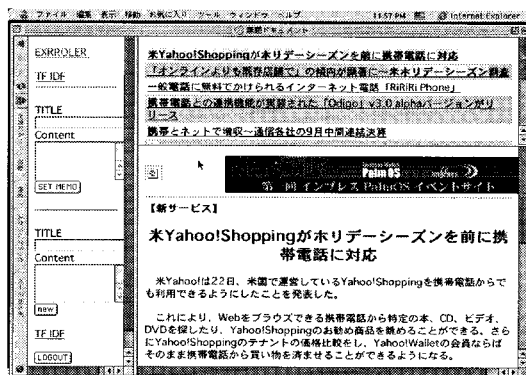


図4:閲読中Articleに対するリンク提示

7. おわりに

情報の収集と編集および発信をくり返し行うというワークフローにおけるメールの利用をサポートするモジュールを設計開発した。結果、メールマガジン等閲読時の類似情報検索と、新規記事作成およびメールマガジン発信支援が実現できた。

今後はモジュール運用時の利便性に関する評価と、他システムとの連携を行う。

参考文献

- [1] 長谷川隆明,浅野久子,堀井統之: "電子メールのインテリジェントサービス",人工知能学会誌 Vol.14 No.6 pp.951-958,1999.
- [2] 松井英昭,後藤文太郎: "WWW のシームレスな利用環境の構築",人工知能学会研究会資料 SIG_KBS_9803, pp.37-41, 1999
- [3] "形態素解析システム「茶筌」", <http://chasen.aist-nara.ac.jp/index.html> ja
- [4] 長尾真: "自然言語処理", 岩波書店, 1996.
- [5] "INTERNET Watch Flash", <http://www.watch.impress.co.jp/internet/>