

MAILPIA[®] メールで受け取る情報収集支援サービス ～情報抽出技術を適用したネットワークサービスの開発～

川北 泰広

インターネットの普及と共に、容易に情報を発信することができるようになった。最新のトピックスとしては、BLOGの普及がある。BLOGサービス提供サイトにアクセスできれば、入力欄に書き込むだけで情報の発信が可能になる。情報発信の容易さと、互いに情報を連携しあう仕組み（トラックバック）により、利用者が急増している。その一方で、このような情報の増加、変化に、情報を受ける側がついていけなくなっている。Webブラウザ、電子メール、RSSリーダなどさまざまなツールを駆使してはいるものの、いわゆる情報の洪水は受け止められず、知らない、もしくは受けていたとしても気付かずに捨てていることが多い。

このような状況ではあるが、配信される情報には、プライベート、ビジネス問わず、有用なものが豊富に含まれている。したがって、大量に存在する有益、無益な情報の中から、いかに受け手が欲しい情報だけを抽出して活用するかというところに、情報を受ける側のニーズがある。

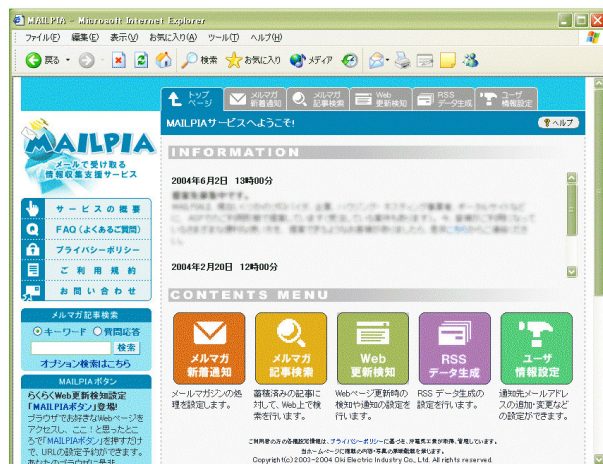


図1 MAILPIAのトップページ

我々は、膨大で雑多な情報の中から、利用者にとって有用な情報のみを、情報抽出という技術により選別して取得する研究に取り組んで来た。本稿では、我々が開発した

メールで受け取る情報収集支援サービス「MAILPIA」^{*1)}の技術概要を説明し、その機能および利用シーンを紹介する。また、サービス提供者への導入形態の具体例と、株式会社朝日ネット殿への導入事例を紹介し、今後の展開を説明する。

MAILPIAの技術

MAILPIAは、情報の変化を電子メールで知らせるサービスである（図1）¹⁾。現在、変化を検知できる情報源は、Webページ、RSSデータ、メールマガジンであり、利用者はこれらの情報源とキーワードを指定することで、情報の変化や新着情報のクリッピングを受けることができる。このサービスを提供するMAILPIAシステムは、メールのハンドリング技術、Webアクセスのスケジューリング技術、Webアプリケーション構築技術、そして情報抽出技術の4つの技術を融合して構築されている。

メールハンドリング技術は、メールの送受信を高効率に行うために必要である。MAILPIAに転送されるメールに対して、メールの受信、情報抽出処理、必要な情報のメール送信と滞ることなく処理を実行している。一般の転送処理だけでなく、受信するメール毎に内容を解析し、既に受信したメールとの比較を行うことから、高度なトランザクション管理を必要とする。

Webアクセスのスケジューリング技術は、様々なWebページの更新内容を、指定されている利用者のメールアドレス宛てに効率的に送信するために必要である。複数のメールアドレスと、複数のWebページと、キーワードの設定有無の組み合わせは無数に存在する。さらに同一ドメインへのWebアクセスのタイミング、一定時間以内に送信を完了する条件も考慮する必要がある。そのため、更新検知するWebページの順序や各処理単位へのWebページの割り当てや、送信するメールアドレスのスケジューリングの工夫などで効率化を図っている。

Webアプリケーション構築技術としては、データベース管理システムOracle^{*2)}、WebアプリケーションサーバWebLogic^{*3)}を組み合わせ、実装効率、汎

^{*1)} MAILPIA (Mail-based Personalized Information Accessing Service) は沖電気工業株式会社の登録商標です。 ^{*2)} Oracleは日本オラクル株式会社の登録商標です。

^{*3)} WebLogicは日本BEAシステムズ株式会社の登録商標です。

用性、拡張性のあるシステムを、Javaにより構築した。Javaを採用した理由は、拡張性の面から、MAILPIAをMVCモデルに従ったWebアプリケーションとして開発するのに適していたためである。また、サービス利用者の増加に応じたパフォーマンスを維持するためのスケールビリティ確保には、サーバ間プロセスが並列に動作するように工夫している。

情報抽出技術は、日本語入力文に対して、意味のある単位で文字列を抽出し、その属性を付与する技術である。全文一致検索に比べて、計算コストは増加するが、欲しい情報を絞り込んで的確に取得するという用途には有効に働く。具体的には、個々の文字列に対して、人名・地名・組織名、数値情報などの属性を付与する。抽出された連続する複数の文字列を一つに結合した上で新しい属性を付与することや、一つの文字列に複数の属性を付与することも可能である。たとえば、「大阪支店長」を「大阪支店:組織」、「支店長:役職」と属性が付与され、どちらのキーワードでも抽出することができるようになる。

MAILPIAの機能

MAILPIAには、5つの機能がある。顧客には、この中から必要なものだけを組み合わせ提供する。

(1) Webページ/RSSデータの更新検知機能

利用者がURLで指定するWebページまたはRSSデータの更新を定期的に検知し、その更新内容をメールで利用者に通知する。RSSデータの場合は、更新したitem要素の概要部分をメールで通知する。それぞれ、キーワードを指定することにより、キーワードに一致する更新があった場合にのみ通知するような設定が可能である（図2）。

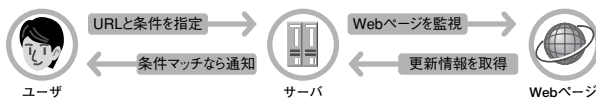


図2 Webページ/RSSデータの更新検知の動作

(2) メールマガジン新着通知機能

利用者が購読しているメールマガジンを、MAILPIAの所定のメールアドレス宛てに送信するだけで、利用者があらかじめ指定しているキーワードに一致する記事が含



図3 メールマガジンの新着通知の動作

まれていた場合に、その記事だけを利用者にメールで通知する（図3）。

(3) メールマガジン記事検索機能

メールマガジン新着通知機能と同じ方法で、転送されたメールマガジンを記事単位に分割し、その利用者だけが利用可能なデータベースに蓄積する。複数のメールマガジンの記事に対して、横断的にキーワード検索、または自然文による質問応答検索が可能である。たとえば、「沖電気の創業者は誰ですか?」と質問すると、「沖牙太郎です」と記事に含まれている情報から回答を導き出す（図4）。



図4 メールマガジン記事検索の動作

(4) 訳して通知機能

Webページ/RSSデータ更新検知機能において、検知先URLの内容が英語の場合に、その更新内容を日本語に翻訳する。そして翻訳結果を英語の更新内容に付加してメールで通知する。このとき翻訳エンジンには、「訳してねっと」をWebサービスとして利用している（図5）。

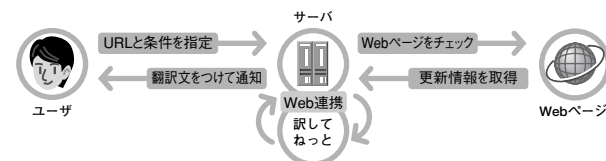


図5 訳して通知の動作

(5) RSSデータ生成機能

Web更新検知機能において、複数の更新情報を用いて、RSSフォーマットのデータを生成し、メールに添付するという形態で利用者に送信する。RSSフォーマットのデータは、XML形式のデータであり、メールに添付して利用者の元へ送信することで、利用者は自分のRSSリーダーでその内容を確認できる。また、利用者自身が開設しているWebサイトの更新内容をRSSデータに変換して、

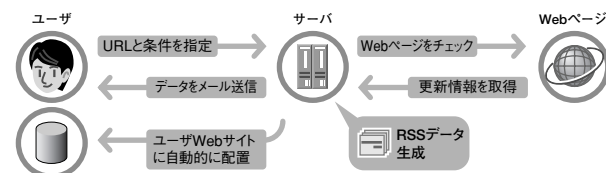


図6 RSSデータ生成の動作

同Webサイト上にRSSデータを掲載することにも利用できる (図6)。

MAILPIAの利用シーン

利用者の立場でのMAILPIAの3つの代表的な利用シーンを例示する。

(1) 自分でマイメールマガジンを作る。

お気に入りのアーティストのWebページや、ECサイトのレア商品を提供しているWebページ。また、自分の趣味に関連する情報が紹介されるBLOGページ。これら気になるWebページのURLを設定しておくだけで、MAILPIAはWebページの更新部分を検知し、更新内容からその利用者のためだけのメールを生成して通知する。受信するメールは、結果的に「マイメールマガジン」と呼ぶのが適切な形式となり、情報収集の手間が格段に減るとともに、自分が必要な情報を逃さず取得することができる (図7)。

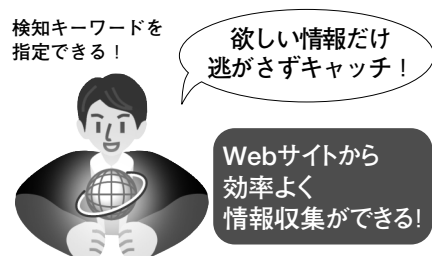


図7 マイメールマガジンの取得イメージ

(2) 最新のIT記事は携帯で

毎日届くITニュースに関するメールマガジンは、鮮度の高い情報が豊富であり、その中には、求めている情報が埋もれながらもいくつか含まれている。しかし、分量が多いため、興味のある記事を探すだけでも膨大な労力が必要とする。そこで、MAILPIAにあらかじめ読みたい記事のキーワードを設定する。そして通知先メールアドレスとして携帯電話のメールアドレスを設定する。このように設定することにより、受信したメールマガジンをMAILPIAに転送するだけで、キーワードに一致した記事だけが、携帯電話に送られる。たとえば、通勤途中の僅かなホームでの待ち時間に、最新のニュースを読むことができる。その結果、出社した時には、メールを開く前に、最新情報で同僚と会話ができるというような活用が可能である (図8)。



図8 メールマガ新着通知の活用イメージ

(3) 市場動向を読む

メールボックスに入ったままのITニュースに関するメールマガジンを、後から容易に活用することは困難である。メールソフトで検索する場合は、まずメールの検索を行い、続いて記事の検索を行うという2回のステップを踏む必要がある。そこで、利用者が受信したメールマガジンをMAILPIAに送信しておく。すると、MAILPIAは自動的にその利用者だけがアクセスできるデータベースを構築し検索可能にする。たとえば「IP電話」でキーワード検索すると、どんな企業がIP電話に関するリリースを行ってきたのかを、一目瞭然で掴むことができる (図9)。



図9 記事検索の活用イメージ

MAILPIAの提供形態

MAILPIAは、インターネットサービスプロバイダ、コンテンツプロバイダ、ポータルサイトなどと連携することで、その各プロバイダの個人利用者に対してサービスを提供する形態をとる。

インターネットサービスプロバイダや、ポータルサイト向けには、それぞれが持つ基本サービスと密に連動した形態をとる。たとえば、BLOGを提供しているポータルサイトでは、BLOGに書き込みが発生した場合に、MAILPIAを利用して通知するというサービス形態になる。ポータルサイトは、更新通知機能がない状態から新たにシステムを構築するよりも、安価にかつ早くサービスを開始することができる (図10)。

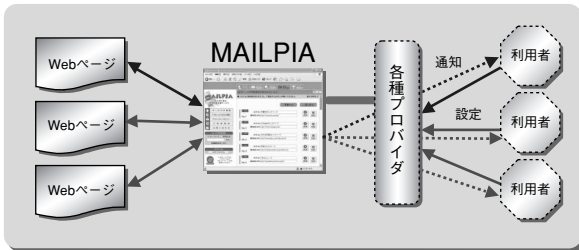


図10 MAILPIAの提供形態

コンテンツプロバイダ向けには、Web上で提供している有料コンテンツ情報のプロモーションに利用する形態をとる。利用者が通知先メールアドレスとチェックしたいサイト内ページを指定するだけで、更新の都度その内容が通知される。コンサートチケットや、レア商品などECサイトにて期限付きで不定期に発売するもののプロモーションに適している。

MAILPIAの現状と今後

MAILPIAは、2004年7月より、株式会社朝日ネット殿の会員向けにWeb/RSSデータ更新検知機能とRSSデータ生成機能の提供を開始した²⁾。半年間のトライアルにより有効性が判断され、必要な機能を組み合わせてサービス提供を開始した。会員は、朝日ネット殿にて会員認証を受けるだけで、MAILPIAのサービスを利用できる。新たなID、パスワードを覚える必要がないため、利便性は高いと考えている。

今後、MAILPIAは情報の変化を通知する機能を基軸に、その他の最新技術と融合させていく予定である。たとえば、多言語機械翻訳技術との融合がある。現在は英日翻訳だけのサポートだけであるが、今後、中日翻訳と連携

し中国語の情報の更新を容易に受信できるようにする。また、音声合成技術との連動も検討している。更新検知、新着通知のメッセージを音声で読み上げるという形態である。これにより、メッセージを注視することができない制限下（車の運転中など）など、何か別のことをしながらでも情報を取得することができるようになる。

我々は、MAILPIAによる通知インフラを整え、その中に必要な情報、関連する情報、周辺情報などを載せようと考えている。最初は利用者が知っているだけの情報源と求める情報のキーワードとを指定する。そして、様々な利用者が様々な情報を取得できるようになると、互いに関連する情報源や情報そのものを関連付けて提供することで、利用者が気付かなかった新しい情報に気付くようになる。情報の洪水の中から、簡単かつ安全に、利用者が求める情報を提供できる環境を築き上げていく予定である。◆◆

参考文献

- 1) MAILPIA紹介サイト <http://www.mailpia.jp/>
- 2) 株式会社朝日ネットMAILPIA提供サイト <http://www.asahi-net.or.jp/service/mail/mailpia/>

筆者紹介

川北泰広：Yasuhiro Kawakita. 研究開発本部 ユビキタスシステムラボラトリ

TIPS

【基本用語解説】

BLOG

利用者がWebブラウザから文章を入力するだけで、自分のWebページを作ることができるサービス。自分の思いや意見を日記風に書き連ねることが出来ることから、利用者が急増しており、多くのプロバイダがBLOG環境を提供し始めている。

トラッキング

あるBLOGに記載された文章に対する意見や関連情報を自分のBLOGに書きこむ際に、元文書のページに自分の文書へのリンクを張る仕組み。

RSSリーダー

RSSは、Rich Site SummaryまたはRDF Site Summaryを略したもの。XML文書の種類で、そのWebページの要約や更新内容を、アプリケーションが解釈できるフォーマットで記載したもの。RSSリーダーはこのフォーマットで記述されたデータを定期的に読み込んで表示するツール。

を、アプリケーションが解釈できるフォーマットで記載したもの。RSSリーダーはこのフォーマットで記述されたデータを定期的に読み込んで表示するツール。

MVCモデル

1960年ごろに提唱されたSmalltalkで、GUIを設計するための考え方であり、Webシステムの構成をM（Model:業務ロジック部分）、V（View:表示処理部分）、C（Controller:MとVを仲介する制御部分）の3つに論理的に分けることを基本とする。

訳してねっと

沖電気工業が提供する、コミュニティ型機械翻訳サービス。<http://www.yakushite.net/>にてトライアル利用可能。