

人間の記憶の拡張を目指した知識情報管理基盤の開発

1. 背景

情報を発信、受信するメディアの多様化、IT化によって、情報の流通速度や流通量は増加し続けており、以前に比べて、はるかに容易な方法でより多くの情報が得られるようになった。その結果、人びとが毎日の生活を通じて得る情報は膨大な量となり、もはや得られる情報の全てを人間が直接目にしてその一つ一つを記憶しておくことは不可能といつてよい。

一方で、企業などの組織では、個々のメンバーが電子メールや Web アクセスなどを通じて様々な情報収集・提供活動を行っている。そしてこれらの活動を通じて授受した情報の集大成によってその組織の知識や記憶、経験といったものが形成されている。しかしながら、組織に属している個人が組織活動の一環として収集した個人の知識や組織外とのインタラクションの履歴が組織としては管理されておらず、個人の異動と共に失われてしまうことが多い。こうした情報を組織の記憶として共有し管理する必要が求められている。

このような背景から、我々が授受した情報を我々に代わって分析・記録し、また個人が授受した情報を組織の知識や経験として共有するための仕組みが求められている。

2. 目的

我々の日常生活や組織活動においてコンピュータの役割は非常に大きくなっている。その中でも電子メールや Web アクセスなどの手段によるネットワークを介した情報の授受は、コンピュータの利用目的の大きな割合を占めている。また、われわれが手に入れられる情報の増加をもたらした最大の要因もコンピュータネットワークを介した情報の授受にあると考えられる。新聞やテレビ、ラジオなど既存のメディアによって配信される情報の量には大きな変化がないのに対して、インターネットを介して配信され授受される情報量が累進的に増加していることからそのことがうかがえる。

これらのコンピュータネットワークを介した情報の授受は、個人や組織が外部に向けたコミュニケーションや情報収集手段として広く用いられており、その履歴は一種の記憶の一部であると考えられる。そこで本プロジェクトでは、利用者がコンピュータネットワークを介して授受されるメールおよび Web アクセスに着目した。これらを用いて交換された情報を収集し分析・蓄積することで、利用者やその組織が潜在的に目にした情報を一種の記憶とみためて検索できるようなソフトウェアの基盤を開発することが本プロジェクトの目的である。

知識情報の分析や蓄積・検索を効果的に行うためには、その用途や目的に応じたカスタマイズが欠かせない。そこで本プロジェクトではいまままで様々な形で開発されてきた知識情報処理技術を集約し、これらを容易にカスタマイズして組み合わせられるようなフレームワークを構築し、これにより知識情報処理アプリケーションが開発できるようなミドルウェアを開発した。またミドルウェアの有用性を検証するために、ミドルウェアを活用したアプリケーションとしてメールの授受や Web アクセスの履歴を時系列などに応じて検索、閲覧するアプリケーションも開発した。

3. 開発の内容

本ソフトウェアは Web 閲覧とメールの送受信を通して利用者が授受した情報や、利用者が行った操作を収集・分析し、その結果をリポジトリに保存する。また、リポジトリに対する検索機能を提供する。

本ソフトウェアはシステムの中核となるコアエンジンと、様々なアプリケーション機能を実現するモジュール群からなる。コアエンジンは、モジュール間のデータの流れと、その蓄積および検索全般を管理する。各モジュール内のコンポーネントがどのように連携するかの定義をコアエンジンに与えると、その定義に従ってモジュール間のデータフローの橋渡しをし、さらにその履歴をコアエンジンが内包する XML DB に記録する。

モジュール群はコアエンジンに対してプラグイン形式で提供する仕組みとなっており、必要に応じてモジュールを追加したり変更したりすることが可能である。

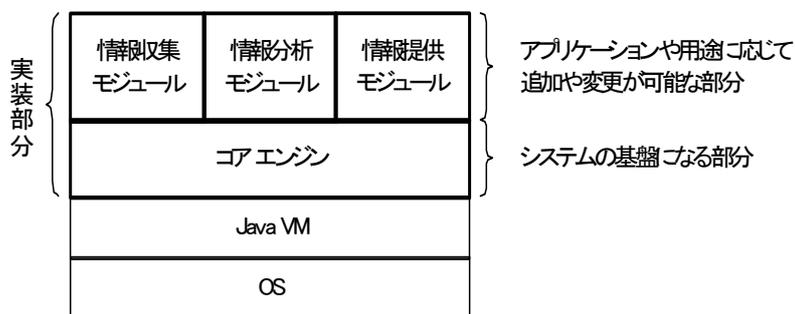


図 1: 本システムの基本構成

アプリケーション固有の動作は、個々のモジュール群の機能とその機能の組み合わせ方によって実現される。本ソフトウェアのモジュールは大別すると「情報収集モジュール」、「情報解析モジュール」、「情報提供モジュール」の3種類に分類される(図 2)。

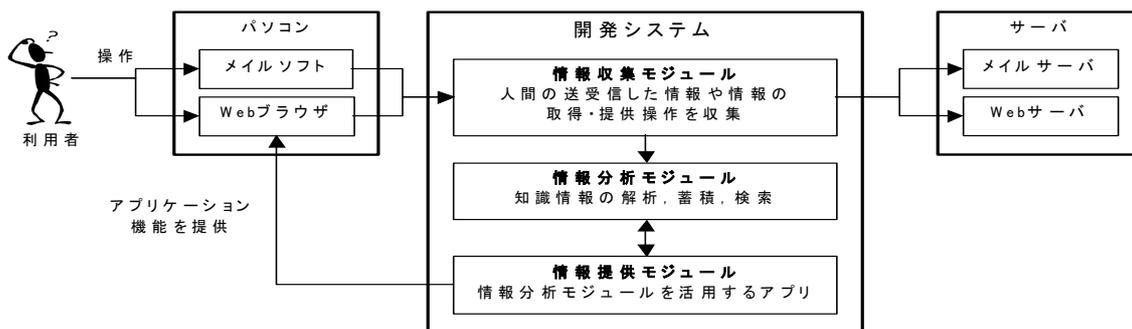


図 2: システムを構成するモジュール

情報収集モジュールは利用者によるメールの送受信や Web アクセスの履歴やその内容を収集するモジュールである。このモジュールは SMTP, POP, IMAP および HTTP のプロキシ機能を持った4種類のモジュールによって構成されている。プロキシを介して授受された情報を全て入力として受け取り、処理する。

情報分析モジュールは、個々の情報を整形したり解析したりする機能を持っている。メールや HTTP のヘッダとボディとの分割や、ヘッダの解析などプロトコルに固有の分析モジュ

ールのほか、HTML のパースなど特定のフォーマットに特化した分析モジュール、また文や段落の境界の判定や日本語の形態素解析など自然言語を対象としたモジュールなどが含まれる。

情報提供モジュールは利用者の検索問い合わせに対応して XML DB を検索し、結果を返すモジュールで、アプリケーションとしての性格が強いモジュール群である。本プロジェクトではメールや Web の時系列検索やキーワード検索などを実装した。

移植性を考慮してコアエンジンと大部分のモジュールは Java で実装した。

4. 従来技術(または機能)との相違

組織内やコンピュータ上の情報検索システムとしては、全文検索システムが広く普及している。全文検索は簡単なキーワードで関連する文書が検索・収集できるという点で優れている。一方で、個々の情報の関係に基づく検索や、単語よりも文脈や内容に踏み込んだ検索は困難である。既にドキュメント量によっては全文検索だけでは必要十分な情報にたどり着くことが困難になっており、全文検索と組み合わせられるような様々な技術が求められている。本システムでは全文検索機能は内包した上で、その上に文脈情報の分析や関連文書とのリレーションの記録などが可能であり、またこうした分析のカスタマイズが出来る点で従来の全文検索だけでは困難であった情報へのアプローチも可能とする。

この他に、組織内での情報共有ツールとしては、ナレッジベースなどのシステムが良く知られている。これらのシステムの多くは利用者があらかじめ知識を登録するものが多い。一方で、本システムは人手で知識を登録するようなシステムでは、知識情報の正確性の点では非常に優れているが、情報の網羅性の点では劣っている。また、ユーザが情報登録の手間を嫌ってあまり有益な知識情報が収集できないといった問題も散見される。本システムでは、情報は利用者の情報授受の履歴から自動的に分析・抽出するため、このような手間は発生しない。

プロキシを介してメールや Web アクセスを記録するシステムとしては、組織のネットワークのゲートウェイに配置されたフィルタなどがあるが、これらは主に情報統制を行う目的で使われており、本システムが目的とする記憶や経験の共有という目的とは本質的に異なる。

5. 期待される効果

今後、知識情報処理技術を活用したアプリケーションへのニーズが高まってくることが想定されるが、現時点では一方で知識情報処理アプリケーションを開発する際に有用なミドルウェアとしては全文検索エンジンくらいしかない。現状では、知識情報処理システムを開発する際には、既存の様々なツールや独自の実装を組み合わせるほぼフルスクラッチに近い形で実装が行われている。

今後、より高度な知識情報処理アプリケーションの必要性が高まるに従って、全文検索エンジンよりもより有用なミドルウェアが必要不可欠となる。本ソフトウェアはこうしたニーズに対する一つのソリューションとして利用されることを想定している。本ソフトウェアにより、基盤となる部分がパッケージ化されて利用できるようになるので、システム開発等の際に人

的資源をドメインごとの処理に集中できるようになり、開発期間やコストの低減に貢献する。

これにより知識情報処理アプリケーションの開発や普及が促進され、これらのアプリケーションソフトウェアを通じて企業や組織内などでの知識共有や情報管理がより効果的に行われることが期待できる。

6. 普及(または活用)の見通し

本ソフトウェアは、アプリケーション等を開発するための基盤となるものなので、本システムをエンドユーザが直接利用するケースよりも、アプリケーション開発者が利用するケースが多い。具体的な利用形態として下記の形態を想定している。

- ✓ スタンドアロンプログラムの基盤
- ✓ Web ベースのプログラムの基盤
- ✓ ビジネスアプリケーション向けのミドルウェア
- ✓ ソーシャルネットワーク等の活動支援ツール

スタンドアロンプログラムの基盤としての利用方法は、パソコンなどにインストールするようなアプリケーションソフトを開発する際に、本ソフトウェアをライブラリのひとつとして活用するという方法である。たとえば、個人向けの情報収集エージェントソフトや秘書ソフトなど、またキャラクター物の対話型ゲームの基盤としての利用が考えられる。

Web ベースプログラムの基盤とは、Web サーバやアプリケーションにインストールし、組織内などの複数の人間が共同して利用するソフトウェアである。グループウェアやポータルなどに本ソフトウェアを導入し、メールや Web アクセスの検索などの拡張機能と用途に特化した Web ベースのユーザインタフェースと併せて提供する。

3つめのビジネスアプリケーション向けのミドルウェアとしての用途は、個別の顧客に特化したシステム開発での利用である。システムインテグレータが、特定顧客の情報システムを開発する際に、アプリケーションサーバやリレーショナルデータベースと同じように、本ソフトウェアをミドルウェアのひとつとして利用し、その上に顧客の要件に従ったアプリケーションを開発するというものである。

4つめの用途は、他の用途とは少し異なりエンドユーザが本プロジェクトの成果を直接利用する形態である。本年度開発したアプリケーション部分のうち、主にメールの分析と検索機能の部分をアプリケーションサービスとして不特定多数のユーザに提供する。e-groups などのように、ネットワーク上のバーチャルなコミュニティの活動を支援する一種のグループウェアとして活用してもらうというものである。

7. 開発者名(所属)

鈴木 泰山 (株式会社ピコラボ)

八木 豊 (株式会社ピコラボ)

(参考)

株式会社ピコラボ (<http://www.picolab.jp/>)