

電子メールを利用した住所録管理システム

3L-1

浅野 久子 加藤 恒昭 高木 伸一郎
NTT情報通信研究所

1. はじめに

近年、電子メールを用いた情報伝達の機会が増大し、各種指示、連絡や挨拶等にも電子メールが利用されている。この電子メールには、送信者の姓名や電話番号等の住所録情報をもつsignatureとよばれる署名が付与される場合が多い。また、電子メールの配送に必要な情報をもつheaderは、通常、送信者のメールアドレスを含み、さらに姓名等の情報を含む場合もある。受信する電子メールからこれらの送信元の住所録情報を自動的に抽出して住所録の維持更新に利用できれば、非定期に起こる住所録の変更の必要性やそのタイミングをユーザが留意している必要がなくなり非常に有益である。しかしsignatureは必ず存在する訳ではなく、その表現形式は多彩であり、含まれる情報も様々である。このため、電子メールからの住所録情報の抽出は単純なパターンマッチでは困難である。

そこで我々は、レイアウト情報に基づくsignatureの検出、および、signatureの局所構造を単位としたパターンマッチによる住所録情報の抽出を特徴とした、signatureとheaderからの送信元住所録情報の抽出を提案している[1]。本稿では、その一つの応用として、住所録の維持管理の負荷を軽減することを目的とした住所録管理システムを提案する。本システムは、ユーザが指定した条件を満たす電子メールが到着すると自動的に[1]の手法で住所録情報を抽出し、既存住所録との照合を行い、差分が存在する場合には特定のタイミングでユーザに住所録の更新を促し、ユーザの簡単なGUI操作で住所録の更新を行える。

2. システム構成

住所録管理システムはPerlとTcl/Tkにより実装され、SunOS4.1.4上で動作する。図1にそのシステム構成を示す。本システムは、メール分類部、個人情報サーバ、送信元情報抽出部、住所録情報更新部から構成される。以下、各部の処理概要を説明する。

(1) メール分類部

メール分類部は、headerの各fieldやbodyに対する特定文字列の有無条件に従って、到着した電子メールを任意のプログラムに引き渡すことができる。これにより「To fieldに自分のメールアドレスが存在する場合のみ住所録情報の抽出を行う（メーリングリスト宛やCcされている電子メールは処理しない）」

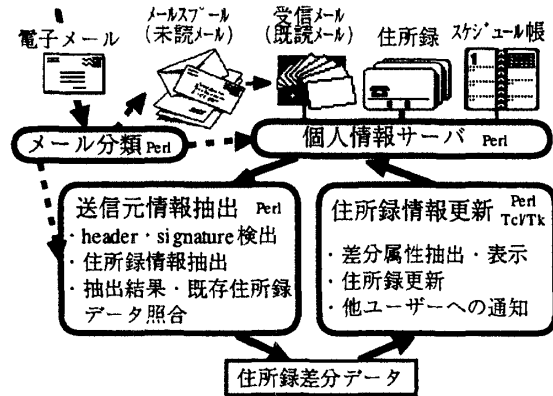


図1 システム構成

といった、ユーザの必要に応じた条件設定が可能となる。また、個人情報サーバ用の電子メールを個人情報サーバに引き渡す役割ももつ。

(2) 個人情報サーバ

個人情報サーバ[2]は、受信メール、住所録、スケジュール帳等のすべての個人情報を共通的な形式で管理するサーバであり、各利用者毎に起動される。他ユーザの個人情報サーバへ、電子メールを用いてアクセスすることも可能である。住所録へのアクセスは、この個人情報サーバを介して行う。

(3) 送信元情報抽出部

送信元情報抽出部では、[1]の手法を用いて、signatureの検出を行い、signatureを検出した場合には、そのsignatureとheaderのFrom fieldから定型度の高い属性順に送信元住所録情報を抽出する。抽出する住所録情報は、姓名、郵便番号、住所、メールアドレス、電話番号、FAX番号、会社名、所属名（会社名より下位の組織名）、役職の9属性である。

住所録情報の抽出では、オプション機能で既存住所録データ¹を利用することもできる。この時、既存住所録データが存在した場合には、各属性毎に、抽出ルールを適用する前に既存住所録データの該当する属性データがsignature内に存在するかをチェックし、存在する場合にはその属性の値は変更なしと判断する。

ここで、住所録情報の抽出精度は、送信人がすべて異なる200通の電子メールに対する評価において、既存住所録データを利用しない抽出の場合、全抽出対象属性に対する適合率92.7%、再現率81.3%

¹ 電子メールのFrom fieldから抽出したメールアドレスと同一のメールアドレスをもつ既存住所録のデータ

(signature未検出・誤検出による誤りも含む)であり、実用レベルであることを確認している。

住所録情報の抽出後、抽出した送信者のデータが既に住所録に登録されているか²、そして、登録されている場合には、既存住所録データと抽出データに差分があるかを調べる。既存住所録データでは空であるが、抽出データでは値をもつ属性、または、既存住所録データ、抽出データとも値をもつが、その内容が異なる属性が存在した場合には、住所録情報の変更が生じた判断し、抽出データを住所録差分データファイルに書き出す。また、既存の住所録に抽出した送信者のデータが存在しない場合には、新規データと判断し、抽出データを住所録差分データに書き出す。

この差分チェックおよび既存住所録を利用した抽出では、電話番号、FAX番号、住所、会社名、所属名については抽出データの正規化を行い、既存住所録データとの照合を行う。例えば、signature内で「(03)3456-7890」と表記されている電話番号は、「03-3456-7890」と正規化して照合する。

(4) 住所録更新部

住所録更新部は、ユーザが指定したタイミング（ログイン時や特定時刻）での自動起動、あるいはユーザ自身により起動される。

起動後、住所録差分データと現在の住所録との照合を行う。これは、住所録差分データ作成時（メール到着時）には差分が存在しても、その後、手作業等で住所録の更新が行なわれる可能性があるためである。照合後、新規データ件数、変更データ件数を表示する。

ユーザは、確認画面を表示させ、新規、変更データの内容を確認し、場合によっては修正を加えて住所録を更新することができる。確認画面例を図2に示す。

図2の画面上部の「抽出Signature部」には、抽出されたsignatureが表示されている。画面中部には、右側に旧データ（既存住所録データ）、左側に新データ（抽出データ）が表示され、差分が分かりやすく表示されている。ここではグレースケールで表示されているが、実際にはカラー表示であり、既存住所録データとの差分が存在する所属、電話番号1、および、signatureから抽出されなかったため既存住所録データがそのまま抽出データとなった住所1は、画面背景色とは異なる背景色で表示されている。既存住所録データと抽出データで値が異なる電話番号1の新旧データ間の矢印をクリックすると、新データに旧データの値（03-3456-1234）が埋め込まれ、電話番号1の値は変更されない。また、画面下部の備考には、住所録情報として抽出されなかった

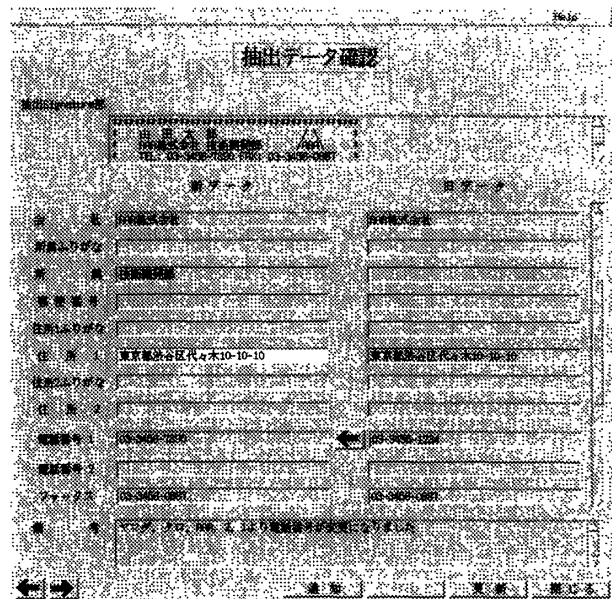


図2 確認画面例

signature内の文字列が列挙されている。これにより、抽出できなかった住所録情報が存在した場合にも、ユーザが該当する文字列を備考からCut&Pasteして、新データの該当属性欄に簡単に情報を埋め込むことができる。そして、画面右下の「更新」ボタンを押すことにより、現在の住所録の内容が、「新データ」として表示されている内容に更新される。また、「通知」ボタンを押すと、他のユーザの住所録に、更新した住所録データを通知（上書き）することができる。ただし、受け側のユーザが、通知を行うユーザによる住所録の書き込みを許可している場合のみ実行される。

3. おわりに

到着する電子メールから送信元の住所録情報を自動的に抽出し、既存住所録との差分が存在した場合にはユーザに住所録の変更を促す住所録管理システムを提案した。本システムにより、ユーザは非定期に起こる住所録の変更の必要性やそのタイミングを留意している必要がなくなり、住所録の更新も簡単なGUI操作のみで行えるので、従来の住所録管理の負荷が大幅に軽減されることが期待できる。

今後は、GUIの使いやすさ等のシステム評価を行う予定である。

参考文献

- [1]浅野, 大山: 電子メールからのパーソナル情報抽出方法の検討, 第52回情処全大4J-4 (1996)
- [2]加藤, 浅野, 中野: 情報交換に電子メールを用いる個人情報サーバの実現, 第55回情処全大1X-7 (1997)

² 既存住所録データを利用した住所録情報の抽出を行った場合には省略される（抽出段階でチェック済みのため）