

テキストデータ中の個人情報情報をマスクする —「TRUE TELLER個人情報フィルタ」の必要性—

顧客からの問い合わせ履歴データベースなどには、氏名、電話番号、メールアドレスなどの個人情報が多く格納されており、その保護は個人情報保護対策で見過ごされやすい部分となっている。本稿では、そのような個人情報を抽出しマスクする野村総合研究所（以後、NRI）の「TRUE TELLER（トゥルーテラー）個人情報フィルタ」の機能や特徴について紹介する。

情報共有による情報漏洩の危険

近年、企業では顧客情報を社内の共有データベースに蓄積し、関連部署で共有することが主流になってきている。たとえば、コンタクトセンターでは、電話やメールで寄せられた顧客からの問い合わせ履歴をデータベースに蓄積し、「いつ、どのような顧客が、どのような発言をしたか」といった情報を、社内の関連部署の担当者が参照できるようになっている。

このように顧客に関する大量の情報が蓄積される一方で、2005年4月から全面的に施行された「個人情報保護法」に対応するため、各企業は顧客情報漏洩防止のためのさまざまな対策に着手しはじめている。たとえば、顧客データベースにアクセスする際に、住所欄や電話番号欄は参照できないようにアクセス制限をかけることや、アクセスログを残しておき、後で漏洩の原因を特定できるような仕組みを作り、漏洩を抑止することなどがあげられる。

個人情報保護対策の落とし穴

しかし、このような対策と並んで大きな課

題となるのが、メールなどのテキストデータ中に記載されている氏名、電話番号、メールアドレスなどの個人情報の扱いである。この対策の不備は、個人情報保護対策の落とし穴になりかねない。

あるコンタクトセンターでは、数日を費やして、データベース内の数万件の問い合わせメールを全部チェックし、データの修正を行うという。問い合わせメールには氏名や連絡先を記した署名が付されていることが多く、またメールの本文中にも連絡先の電話番号が書かれている場合もある。そのため、そのまま社内で情報共有することは危険である。また、このまま外部委託業者とデータのやり取りを行うことは、非常に大きなリスクを伴う。そこで、テキストデータの本文中の個人情報に該当するデータを、前述のように手作業で修正することになる。

この修正は、担当者が1つひとつ目で見、個人情報となる文字列を探しながら「※※※」というような別の文字列に置き換えるか、または削除することで行う。しかしすべてのデータをチェックするのは大変な作業で、膨大な時間がかかってしまう。

上記のようなデータベース内のデータだけ



でなく、問い合わせに返信する際にも、本文中にメールアドレスなどの個人情報を含まないよう、署名を削除したり、引用部分からも個人情報を削除するようにしている企業も増えている。

テキストデータ中の個人情報をマスクする

以上のような問題を解決するため、NRIでは「TRUE TELLER個人情報フィルタ」を開発した。これは、入力されたテキストデータ中の個人情報の可能性が高い文字列を抽出し、これに対して自動的にマスク処理を施すものである（図1参照）。代表的なマスクの機能に次の3つがある。

①個別チェック機能（逐次処理機能）

読み込んだテキストデータ中の個人情報を、1つひとつ確認しながらマスク処理する。「メールアドレス」「電話番号」「人名」「地名」「組織名」などをリストアップし、任意の文字列に変換することができる。

②一括処理機能（バッチ処理機能）

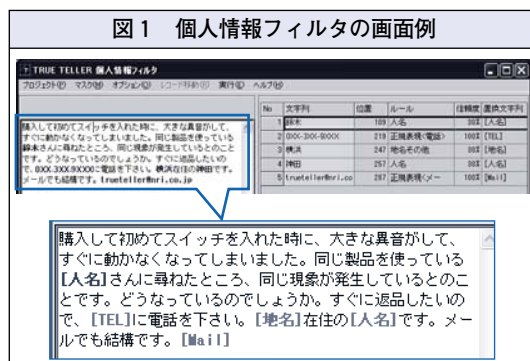
データ中の個人情報を一括してマスクするもので、バッチ設定を行えば処理を自動化することができる。大量データを定期的に更新する際に有効である。

③常駐モード

クリップボードにデータをコピーすると自動的に置換処理を行う。おもにメール対応などで効果を発揮する機能である。

処理にあたっては、NRIのテキストマイニ

図1 個人情報フィルタの画面例



ングシステム「TRUE TELLER」に搭載されている構文解析エンジンが用いられていることも大きな特徴で、単純な文字列一致では判定しにくい個人情報を、精度高く抽出することができる。たとえば「神田」のように、人名にも地名にも用いられる単語の場合、本ツールではその単語の前後の構文関係を把握し、人名だけマスクすることができる。「神田に伝えて下さい」は「人名」、「神田へ行って下さい」は「地名」という具合である。また、通常は人名と判断できないものも、依頼の言葉や敬称などを手がかりに、人名として判定することができる（一部特許申請中）。

いま企業内では、さまざまなドキュメントがいくつものサーバーに散在している。「TRUE TELLER個人情報フィルタ」は、社内ネットワークのどこに個人情報が存在するかを特定してマスクしたりするような機能を備えるなど、ナレッジマネジメントツールとしても進化していく予定である（<http://www.trueteller.net/>）。