



2010年3月3日
メタデータ株式会社

メタデータ、内容フィルタリングソフト『個人情報99』、 『有害情報99』を発売

メタデータ株式会社（所在地：東京都文京区 代表取締役社長：野村直之）は3日、文章内容を解析して個人情報や有害情報を高精度に判別するフィルタリング・ソフト、『個人情報99』、『有害情報99』を提供開始いたします。

『個人情報99』は、メールや、ソーシャルメディアを介した個人情報漏洩の予防や、印刷時の個人情報の伏せ字化や出力者の追跡、開示用データの作成の効率化等に役立ちます。『有害情報99』は、青少年ネット規制法逃れで一般サイトに流入する有害投稿や、ネット上のいじめ・喧嘩、また、企業サイトの誹謗中傷対策に役立ちます。

両製品は、サーバ直販、代理店販売、メーカー・SIer への OEM 供給、SaaS 提供に加え、人的サービスによるチェックやデータ加工を請け負う事業者様との共同事業によるサービス提供にも活用してまいります。

エンドユーザにとっては、例えば会社で「日付・曜日の間違いや個人情報をうっかり残したままメールの送信ボタンを押して顔面蒼白になったけれど送信保留されて助かった」、「SNS やブログに記事書いて不快なコメントをもらうことが激減」といったメリットを享受していただけるようになります。

2005年4月に個人情報保護法が施行されて以来、個人情報保護の意識は高まっていますが、漏洩事故は一向に無くなりません。悪意の無い、うっかりミスであっても、メールや掲示板では大規模な漏洩事故につながる可能性があります。紙による漏洩は追跡しにくい、という問題点があります。

テキスト: ②

【営業日報】
メタデータ株式会社 (http://www.metadata.co.jp) の次期製品開発部署 (東京都墨田区本所 3-15-5; 03-5608-8352) については、窓口の野村主任 (nomura@metadata.co.jp) がソフトウェア部品調達、推薦の主担当であり、発注可否の鍵を握っている。次回訪問で上司の松田課長が出れば、予定通り五百万円で発注の可能性高い。
平成二十年八月六日 訪問
鈴木太郎

↓ 解所の実行 ? 伏せ字設定を変更する

Mextractr個人情報フィルタ (プライバシー・フィルタ) 出力 伏せ字内容の対応表を表示

【…法人・団体名1…】 (【…URL1…】) の次期製品開発部署 (【…住所1 (都道府県)…】 【…住所1 (市区町村)…】 【…住所1 (市区町村下の地名)…】 - 【…住所1 (丁以下の番地)…】 ; 【…電話番号1 (市外局番)…】 - 【…電話番号1 (市内局番)…】 - 【…電話番号1 (加入者番号)…】) については、窓口の【…個人名1…】 (【…email1 (前半)…】 @ 【…email1 (後半)…】) がソフトウェア部品調達、推薦の主担当であり、発注可否の鍵を握っている。次回訪問

詳細制御:

個人名:	<input checked="" type="checkbox"/> 姓	山田
	<input checked="" type="checkbox"/> 名	太郎
法人・団体名:	<input checked="" type="checkbox"/> すべて伏せ字	株式会社あいう
住所:	<input checked="" type="checkbox"/> 都道府県	東京都
	<input checked="" type="checkbox"/> 市区町村	あいう市
	<input checked="" type="checkbox"/> 市区町村下の地名	あいう
	<input checked="" type="checkbox"/> 丁以下の番地	1-1-1
	<input checked="" type="checkbox"/> 郵便番号	123-4567
地名・施設名:	<input checked="" type="checkbox"/> すべて伏せ字	あいう金庫
tel/fax:	<input checked="" type="checkbox"/> 前半を伏せ字	01
	<input checked="" type="checkbox"/> 真ん中を伏せ字	2345
	<input checked="" type="checkbox"/> 末尾4桁を伏せ字	6789
e-mail/URL:	<input checked="" type="checkbox"/> emailの前半を伏せ字	tmp
	<input checked="" type="checkbox"/> emailの後半を伏せ字	tmp.com
	<input checked="" type="checkbox"/> URLを伏せ字	http://www.tmp.com
金額表現:	<input checked="" type="checkbox"/> すべて伏せ字	100000円
クレジットカード:	<input checked="" type="checkbox"/> すべて伏せ字	
ID:	<input checked="" type="checkbox"/> すべて伏せ字	3555-1234-5678-9012
住民票コード:	<input checked="" type="checkbox"/> すべて伏せ字	12345678903

*伏せ字全体の機能制御:
 既に匿名化された情報は伏せ字にしない
 この文字列を匿名化済みとして伏せ字にしない (カンマ区切りで複数指定可)

一方、CRM 部門での顧客の声をはじめ、少しでも個人情報を含むからといって一切社内情報共有を禁止してしまえば、貴重な顧客の声が製品やサービスに反映しなくなり、競争力の低下や、甚だしくは欠陥製品の出荷につながりかねません。

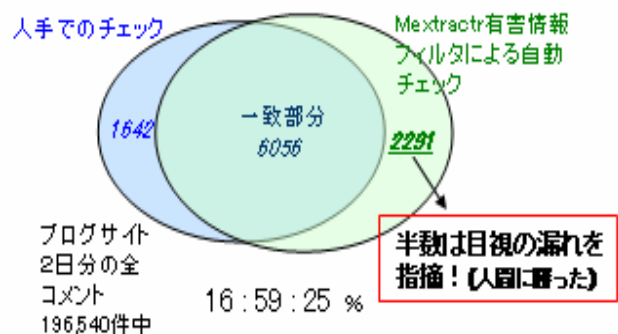
『個人情報 99』は、これらの問題を効率良く解決するため、次の特徴を備えました：

- ・ 個人名、法人名、住所、email、電話番号、金額、カード番号、住基 ID をマスキング。
- ・ 姓・名、住所、電話番号、email などの一部を指定してマスキング。
- ・ マスキング対象の通し番号を管理。マスキング前後の対応表を出力可。
- ・ マスキングの代わりに指定文字列に置き換える仮名化が可能。
- ・ 日本人の個人名の抽出率 99% (社内評価) など、文法、辞書、文脈処理により高精度。
- ・ オプションで「日付・曜日の対応誤り」チェック & 訂正等が可能。
- ・ WebAPI 経由 (安全な SSL 通信) で利用の他、CSV to CSV 変換のコマンド呼び出しも可能
- ・ CPU コア数に比例して処理能力が向上。標準的サーバ 1 台で数時間で新聞記事 1 年分のフィルタリングが可能。

業務での運用形態としては、メールフィルタや、グループウェア / SNS サーバ、文書管理サーバ、議事録管理システムへの組込み、Twitter 等ソーシャルメディアへの投稿管理システムからの呼び出しで、試作実績があります。「伏せ字プリンティング」ソリューションとして、印刷時の個人情報の伏せ字化や出力者の追跡の提案、また、マーケティングできめ細かいセグメント化を行う際に、顧客の住所の下位部分を適宜伏せてまとめ上げる提案が有望視されています。この他、SE さんが顧客データを匿名化・仮名化して、リアルさのある程度保ったまま自社へ持ち帰れるよう、お客保有の個人情報入り実データ (CSV) を、現場に持ち込んだノート PC で仮名化することもできます。

『有害情報 99』は、青少年ネット規制法逃れで分散する有害投稿や、ネット上のいじめ・喧嘩、また、誹謗中傷を含む企業サイトの炎上対策に役立つ製品です。昨年 7 月、SNS 運営者が「枕を高くして眠れる」よう、有害情報 99 のプロトタイプを搭載して安心運営できるマーケティング SNS 『Mext ウェル』を発売。以来、暴力 / アダルト表現、誹謗中傷、個人情報暴露、行動指示などの僭越な内容、なりすましや、人名・住所を含むリアルな犯行予告 (当局への届出義務あり) の各ジャンルについて、数百万記事の NG / OK 振り分けで実績を積んでまいりました。

監視基準をすり合わせ、辞書類をチューニングした結果、人手による目視チェックに比肩できる精度に肉薄。一部では人手によるチェック漏れをカバーできるようになりました。



中高生から中高年まで

特に 2009 年 4 月施行の青少年ネット規制法の対象サイトはもちろん、規制から逃れた

非対象のソーシャルメディア上で、中高生、小学生を守り抜く必要があります。また、意外に嫌がらせや喧嘩の多い、中高年向け SNS でも、トラブルを未然に、スムーズに解決する対応が求められています。

『有害情報 99』は、対象サイトごとに異なるフィルタリング・ポリシーや文章の特徴に応じて、半自動収集した有害表現オントロジー(= 階層分類されたシソーラス(類語辞書))を複数備えています。また、今後現れる新規サイト、新規ユーザー層に対しても、少数の正解データから、サイト特有の文体を学習して適切に対応することができます。

参加者の文体を”学んで”精度up!

• 新聞向け文体モデルで解析

- 「いちいち 苦情削除するのって どうよ。」 ✕
- 「投稿無意味か。」(1文節につながっちゃう)
- 「速攻削除反則だよな。」

• 有名人ブログ向け文体モデルで解析

- 「いちいち 苦情 削除するのって どうよ。」
 - 「投稿 無意味か。」
 - 「速攻 削除 反則だよな。」
- 「投稿が無意味」「削除は反則」の各文の意味を汲んでNG判定できる!



こうして、正しい分かち書き(形態素解析)、正しい文節の切り出し、正しい係り受け解析(構文解析)を実行し、有害表現オントロジーと併用することで、高精度な NG/OK の切り分けができます。例えば、下記の通り:

- NG** お前をいつか刺してやる。 **NG** 結局その日も体をしっかり重ねていた。
OK あれだけ釘を刺されても忘れたの。 **OK** 苦労をいくら重ねても学べなかった。

個々の文字列、単語だけでは、NG にしようがないものでも、離れて出現しながらフレーズとして NG な意味を構成するもののみ、NG にしています。この他、特定の人名に対してのキツイ表現のみ NG としたりすることも可能です。

リアルタイム監視も o k

『有害情報 99』のプロトタイプは、マーケティング SNS 製品『Mext ウェル』に組み込んで稼働させてきた以外にも監視業務に適用された実績があります。特に、365日 24時間、秒単位で判定しなければならない、リアルタイム・ソーシャルメディアの自動監視試行で実績があります。多めにフィルタリングしつつ問題無く扱われるような運用上の工夫や、監視を意図的にかいくぐるための破格な書き方(例:平仮名小文字『あいうえお』をランダムに混ぜるなど)も一網打尽にする新技術を投入し、実用化にこぎ着けました。

『個人情報 99』『有害情報 99』の両製品は、サーバ直販、代理店販売、メーカー・SIerへの OEM 供給、SaaS 提供に加え、人的サービスによるチェックやデータ加工を請け負う事業者様との共同事業によるサービス提供にも活用してまいります。

なお、『個人情報 99』『有害情報 99』の両製品は、3月5日午後、慶應大学三田校舎で開催のセマンティック Web コンファレンス 2010 (主催:慶應義塾大学 SFC 研究所)にて展示、ハンズオン・デモをご体験いただくことができます。

<http://s-web.sfc.keio.ac.jp/conference2010/>

動作環境

サーバ要件 CPU :Pentium Dual Core 2GHz 相当以上
メモリー : 2GB 以上 HDD: 空容量 40GB 以上
OS Red Hat Enterprise Linux, CentOS, Fedora
Windows Web Server2008 (64bit 版) など

初期代理店様一覧

eBrid

イーブリッドリレーションズ株式会社

<http://ebrid.co.jp/>

NXTG

株式会社ネクステージ

<http://www.nxtg.co.jp/>



株式会社エスアイアソシエイツ

<http://www.siaj.co.jp/>

価格・納期等

代理店様、SIer 様のサービス内容に応じ、システム買取初期費用 100 数 10 万円～。
『個人情報 99』は、簡単な電話インタビューの後、推奨設定を添えて最短で翌日出荷。
『有害情報 99』は、NG ポリシーを伺い実例をお預かりして 1 週間～ 3 週間程度チューニングの上、SaaS 提供またはサーバ出荷（リモート保守の設定を推奨）。

メタデータ株式会社について

メタデータ株式会社は、2005 年 12 月にメタデータ活用技術、セマンティック技術の応用ソフトウェア会社として設立。文書管理、情報系を中心とした企業への Web2.0 の導入を手がけると共に、WebAPI 化支援、セマンティック技術を駆使した次世代のデータ連携によりホワイトカラー業務の自動化率向上、創造性拡大の支援に努めています。

→ <http://www.metadata.co.jp>

本件に関するお問い合わせ先

メタデータ株式会社

pressrel@metadata.co.jp

TEL:03-3813-5447

FAX:03-3813-5447



<http://www.metadata.co.jp>