

ソーシャルメディアの時代に産業上の重要さを増す メタデータ自動抽出技術

野村 直之*

Web2.0 と総称された IT への人々の関わりと IT 自身の変化により、前世紀にはありえなかったスケールの情報爆発が起こった。新興 Web サービスに押される形で、先に変化した顧客側に引っ張られる形で企業が「2.0」的変革（エンタープライズ 2.0 ともグランズウェルとも呼ばれる）を起こし始めた。ライフログに象徴される大量の行動履歴や不定形のクチコミ、多量の写真や動画が一般ユーザからネットに投稿されるようになり、もはや産業界で、メタデータによる交通整理が待たなしの状況になっている。本稿では、そこで必要となる 5W1H メタデータ自動抽出技術、その応用としての自動匿名化やカレンダー自動連携サービス、ツイッターを初めとするソーシャルメディアのサービス向上の可能性、そして、パーソナル広告など、関連の将来市場について論ずる。

キーワード：Web2.0, 情報爆発, 新興 Web サービス, エンタープライズ 2.0, グランズウェル, ライフログ, 行動履歴, クチコミ, 5W1H, メタデータ自動抽出, 匿名化, カレンダー連携, ツイッター, ソーシャルメディア, パーソナル広告

1. 産業応用～ビジネスや企業情報システムにおけるメタデータの重要性

Web2.0 と総称される、IT への人々の関わりと IT 自身の変化により、前世紀にはありえなかったスケールの情報爆発が起こった。新興 Web サービスに押される形で、また、先に変化した顧客側に引っ張られる形で、企業が「2.0」的変革（エンタープライズ 2.0 ともグランズウェルとも呼ばれる）を起こし始めた。ライフログに象徴される大量の行動履歴や不定形のクチコミ、多量の写真や動画が一般ユーザからネットに投稿されるようになり、もはや産業界では、メタデータによる交通整理が待たなしの状況になっている。

本稿では、そこで必要となる 5W1H メタデータ自動抽出技術、その応用としての自動匿名化やカレンダー自動連携サービス、ツイッターを初めとするソーシャルメディアのサービス向上の可能性、そして、パーソナル広告など、関連の将来市場について論ずる。

1.1 ‘マッシュアップ’を支える軸足メタデータ

GoogleMaps に象徴される、大規模で整備されたデータが無償公開されるようになったのが“Web2.0”の一大特徴であった。この公開の技術的仕組みが、WebAPI というも



図1 マッシュアップを支える軸足メタデータ

のである。互いに遠隔拠点にあり、それぞれ独立に企画・設計・運営される Web サーバが、REST, SOAP, JSON, その他 Feed や一般 XML 形式の WebAPI を介して連携し、協調する。このように構築される Web アプリケーションは、「混ぜ合わせた料理（例：マッシュポテト）」に例えて、「マッシュアップ」と呼ばれる。

優れて有用なマッシュアップを創る鍵は、これらの部品をつなぎ合わせ、連携させる「メタデータ」である。

異種コンテンツが、メタデータにより次々と結びつき、連携していく姿を図1「マッシュアップを支える軸足メタデータ」に示した。右下の乗用車ダッシュボードの写真から反時計回りに連携処理が進む。

ダッシュボード上に置いた GPS デバイスは、一定間隔

*のむら なおゆき メタデータ（株）

〒112-0002 東京都文京区小石川 2-1-2 山京ビル 7F

Tel. 03-3813-5447

（原稿受領 2010.10.6）

で刻々と現在時刻と緯度経度の組み合わせをと JPEG 写真を記録し続ける。ドライブコースの各地点で普通の (GPS 機能のない) デジカメ画像で撮影した写真を PC に読み込み、先の GPS デバイスを USB 接続。すると、専用ソフトウェアにより、各 JPEG 写真の EXIF メタデータに、緯度経度が流し込まれる。これは、日付時刻という軸足メタデータを介して、別々に取得された緯度経度データと、写真画像データが結びつけられたことになる。

次に、緯度経度情報入りとなった JPEG 写真を、同様に、別途作られていた、緯度経度入りの地図データ (例えば GoogleMaps) の上に、サムネイル化して地図の上にプロットする。今度は、緯度経度という軸足メタデータを介して、別々に作られた写真画像データと地図データが結びつけられたことになる。

このように、あたかも球技バスケットボールのビボットターン (軸足転回) のように、軸足メタデータを切り替えて、様々な個性的な WebAPI を結びつけ、魅力的なアプリケーションを構築するのが、優れたマッシュアップ制作のコツとなる。もう 1 つのコツは優れた素材そのもの、すなわち WebAPI を適切に選ぶことである。

そのための検索サービスとして唯一国立国会図書館データベース検索サービス Dnavi¹⁾に登録されているのが、メタデータ株式会社が無償提供している、“API 比較・マッチング” サービス²⁾である。国内外の 2,000 超の公開 WebAPI が登録され、日英語でキーワード検索し、WebAPI のプロフィール (WebAPI 自体の「メタデータ」のセット) を比較、吟味し、選別しやすくしている。WebAPI のメタデータとしては、ライセンス条件の種類や実行制限などもあり、組み合わせ、相性 (マッチング) をよく吟味する必要がある。中でも最重要なのは、入出力データ形式であり、直列につなぐ WebAPI 間で、直前の出力と直後の入力インタフェース、データ形式が、Text, REST, SOAP, JSON, Atom etc. のいずれかでそろっている必要がある。これを表形式で一目瞭然に比較、マッチング評価しやすいよう、“API 比較・マッチング” サービスは設計、構築されている。

1.2 新世代企業情報システム「エンタープライズ 2.0」とメタデータ

一方、企業情報システムは少し遅れて 2007 年初め頃から、Web2.0 の影響で低コスト化、短納期化と高機能化、使い勝手向上を迫られることになった。その内容、背景、理論 (モデル)、将来ビジョンを描いた書籍がある。『エンタープライズ 2.0 ～次世代ウェブがもたらす企業変革』である³⁾。著者の吉田によれば、私人として Web2.0 の便利さに馴れ、目覚めてしまった人々が所属企業に出勤して、企業情報システムの使いにくさに反乱を起こすようになったのが、この変革の原動力になったとされる。

反乱、逆転は、自宅で Web 検索することで耳年増になった消費者と、営業マン・販売員との間でも起きている。自社の新商品・サービスの学習に熱心でない彼らは、情熱的

に情報を探索・比較・習得する消費者に、知識量で凌駕されかけている。その結果、昔の MIL 規格 (Military Standard: アメリカ国防総省が制定した、アメリカ軍が必要とする様々な物資の調達規格) に象徴されるように官公庁向けの特種・専用製品が最先端で、そのテクノロジーを何年かかけてコストダウンして消費者向けに適用、という流れが逆転しつつある。すなわち、まず最新、最先端の高度なテクノロジーが多数の一般大衆向けに投入され、そのエッセンスの一部を活かした専用機やサービスが企業向けに、遅れて開発されるケースが増えている。これを IT コンシューマライゼーションと呼ぶ。

IT コンシューマライゼーションの小規模なものは社内でも起きている。情報システム部門が数年かけてトップダウンに企画、設計したシステムが拒否反応にあう。代わりに、ほんの数日で、マッシュアップにより機能追加した Web アプリが使われ、すぐに評価・改良される、という、常時変化 (永遠の β 版) の世界が社内に出現したのである。もともと個人用だった小さなアプリ、サービス、部品が、「勝手に」社内ですぐに使われ、普及した結果、プラグインや社内版などの形で、既存の情報システムに組み込まれ、だいぶ後になってから正式承認される。このようなことが、アジャイル (俊敏) な企業では当たり前になりつつある。

情報システムが使いやすくなるとともに、当然情報爆発も企業内に侵入する。大量のデータを、ますます多忙になる社員が迅速にこなさねばなくなっている。多量の情報は、判断速度、処理速度を低下させがちであるため、これは大変深刻な問題である。

この問題を解決する鍵は、社内外の情報に適切なメタデータ (分類のための属性情報) を付与し、意味的に適切に分類された手がかりをもとに情報探索、比較・評価、マッチングできるようにすることである。この点を、『エンタープライズ 2.0 ～次世代ウェブがもたらす企業変革』では、5.4 節「メタデータ：情報洪水のカオスを救う救世主」にまとめている。

社内メタデータの供給源、付与手法として、著者の吉田は次を挙げる：

- ・ LotusNotes をはじめとする古き良きレガシー基盤のメタデータの変換、流用
- ・ MS Share Point Server のように文書操作を機に強制付与
- ・ Web2.0 流にユーザが手作業で付与 (忙しい社内では「?」)
- ・ 利用状況から自動付与 (例：アクセス人数、「この文書を読んだ人はこの文書も読んでいます」等のソーシャルフィードバック)

しかし、なかなか、低コストで迅速に付与したメタデータが統一できなかったり不備だったり、という問題が残りがち、とされる。

そこで、メタデータ株式会社他、自然言語処理技術で間

題解決をはかる企業は、自然言語による文章、プレーンテキストを解析して自動抽出する、という道を追求している。

2. 最近のメタデータ自動抽出技術

自然言語テキストからの商用メタデータ自動抽出技術としては、ロイター社がベンチャーを買収して立ち上げた、OpenCalais⁴⁾が有名である。英語の文章から個人名、法人名、地名等の固有名詞類を抽出するという基本機能に加えて、会社の M&A や株式公開関係の事実情報を抽出するという特徴的な機能をもつ。

WebAPI 化された、テキストからのメタデータ抽出エンジンを調査すべく、“API 比較・マッチング”サービスの検索フォームに、「メタデータ」と入力すると、メタデータ関連の API 24 件がヒットする（2010 年 9 月 30 日現在）。

この中から、最初から何らかのメタデータを集積してデータベース化し、検索サービスを提供しているものや、画像ファイルの Exif など、既存のメタデータを出力させるものを除くと、個人情報フィルタ API と、Mextractr API の 2 件となる。この他、「メタデータ」という言葉は使っていないが、地理情報システム関連の API として、LocoSticker 位置表現特定 API が、テキスト中の一表現を抽出し、Yahoo!テキスト解析は特徴的なキーワードの抽出、なずき EX は感情表現を抽出する機能をもつ。

かように、日本語テキストから特徴的な情報、メタデータを抽出する公開 WebAPI は豊富に提供されているといえる。

2.1 テキストからの 5W1H の抽出・数値化・正規化

MextractrAPI は、日本語文章中から 5W1H の意味属性をもったフレーズを抽出し、数値化、正規化して出力することができる。この MextractrAPI をもとに、名前や住所を分解してその一部を伏せ字化、仮名化できる制御を付加し、電話番号や銀行口座番号、クレジットカード番号、住基ネット ID 等に拡張したものが、前記個人情報フィルタ API の後継 API、5W1H メタデータ（個人情報）抽出 API である⁵⁾。

2005 年 4 月施行の個人情報保護法により、日本のビジネス慣習は、世界的にも類例のないほどセンシティブに個人情報を管理する方向にシフトした。各種内部要因による個人情報漏えいが漏えい事故の原因の上位を占めるため、基本的には従業員を疑う「性悪説に基づく」企業ガバナンスが求められるようになった。そこで企業情報システムでは、個人の自主管理ではなく、

- ・サーバー型の集中管理システムで活動ログをチェック
 - ・抜き打ち的、強制的に個人情報の管理状況を監査
- という解決法が求められている。

5W1H メタデータ（個人情報）抽出 API は、5W1H（いつ、どこで、誰が、何を、どう、いくらで、etc.）のメタデータ自動抽出 API である。日本語テキストを受け取ると、その中の日時・地名・人名といったメタデータを抽出。名前や住所だけでなく、電話番号やカード番号、銀行口座

番号などの個人情報も抽出できる。数値に変換して Atom-gData 形式（Google カレンダー互換）で出力を生成するのを基本としている。

例えば、本年 2010 年の 8 月 24 日に「再来週金曜朝八時」という表現を解析すると“2010-09-10T08:00:00+9:00”と出力する。解析の結果、このように年月が省略されていれば、システムクロックから現在の日付時刻を取得して補完する。さらに、翌々週の年月日を計算。入力が日本語であることから GMT（グリニッジ標準時）+9 時間と同定し、時分秒を確定させ、内部表現の数値を形成する（データ型は日付時刻型）。これを改めて国際標準規格に準拠してシリアルライズ（可搬性・汎用性の高い文字列として出力）したのが、“2010-09-10T08:00:00+9:00”である。

文献 5) のデモ画面の最下部の [Mextractr 意味体系分類名一覧を開く] を押下すると、5W1H を数百のノードをもつ木構造に体系化した、意味体系分類コードが表示される。これらの全てを解析、出力可能としているわけではないが、多彩な固有表現の解析と、数量表現を単位付（意味付）で抽出可能なことが示唆されている。

いわゆる個人情報としては、姓名、住所、電話番号、e-Mail、銀行口座番号、クレジットカード番号、生年月日等が共通的に認識されている。そこで、テキスト中からこれらの記述を特定し、その全体または一部を伏せ字や、仮名にする、といった対応が、テキストの再利用の局面で求められることが多い。

図 2 は、5W1H メタデータ（個人情報）抽出 API を用いて構築した CSV フィルタ・アプリケーションの画面例である。

図 2 5W1H メタデータ（個人情報）抽出 API を用いて構築した CSV フィルタ・アプリケーションの画面例

読み込む (upload する) CSV ファイルを指定し、下方のオプション指定画面で、どんな個人情報を対象にするか、その全体を伏せるのか、一部だけ伏せるのか、そして、仮名化する際の文字列を指定できるようになっている。

2.2 メタデータ駆動型の新世代アプリ “メイドめーる”

5W1H メタデータを抽出した結果を、そのまま「隠す」だけでは、先述の「軸足メタデータを活用した賢いマッシュアップ」アプリケーションとはいえない。そこで、この 5W1H メタデータ (個人情報) 抽出 API を用いて作成された数十のアプリケーションのうち、マッシュアップのコンテストで受賞した Salestrack⁶⁾; Mashup Awards 4 にて 2 賞を受賞)、メイドめーる⁹⁾; Award on Rails 2008 にて受賞)、Tmeeting⁷⁾; Mashup Awards 5 にて受賞)、OpenSocial Dashboard⁸⁾; Mashup Awards 5 にて受賞)の中から、利用者の多いカレンダー SaaS と連携させた、メイドめーるを紹介する。

メイドめーるは、Google カレンダーの予定を指定した時刻に PC メールや携帯電話にメールしてくれるサービス。メールに返信することで、予定を登録することもできる。返信メールの文章を解析して日付時刻に言及したフレーズを抽出するのに、5W1H メタデータ (個人情報) 抽出 API を活用している。

図 3 「メイドめーるの利用イメージ」に図示したように、登録ユーザは携帯メールを含む任意のメーラに一定時刻にその日の Google Calendar の予定をサマライズしたものを受信する。それを見て追加すべき予定を思いついたら、およそ日本語を解する読み手に理解しやすい自由な書き方で、例えば、「*月*日に誰誰とビール」とか、「再来週火曜夜八時に打ち合わせ」などを書いてメールを返信。すると、10 数秒程度で「ご主人様、10 月 12 日予定ですね。ちゃんと登録しておきましたから！」云々の確認メールが送られるとともに、当人の Google Calendar アカウントの予定当該の日付の予定が自動的に登録完了となる。



図 3 メイドめーるの利用イメージ

3. リアルタイム SNS でのメタデータ活用と付加機能による拡張

国内最大級の SNS, mixi では、ほぼ全ての書き込みの際に、コメント執筆者のハンドルネーム、日付時刻、誰 (何) に向けて書かれたか等のメタデータが自動記録される。これは旧来の Wiki 等の情報共有・共創システムにはなかった特徴であり、省力化と利便性向上に貢献する。このため、企業内でも社内 SNS のブームが起きたり、通常のブログ、CMS (Contents Management System) に対し、SNS で実証されたメタデータ活用機能を付加することが行われてきた。

この流れをさらに進化、徹底させた、ともいえるのが、マイクロブログ、リアルタイム SNS とも呼ばれる、ツイッターをはじめとするサービスである。

3.1 ツイッター検索は ‘5W1H メタデータ検索’

ツイッター隆盛の理由は、1 つには、API を幅広く開放して、第三者にビジネスチャンスを与え、文字通り、アプリケーションが 5 万と作られたことにあるといわれる。この際、プログラミングのしやすさだけを論じるのは片手落ちである。シンプルで扱いやすいインタフェースを支えるもう 1 つの鍵は、140 字以内の短いコメントの全てが独立した ID (URL) をもち、図 4 に示すように、発言の日時、

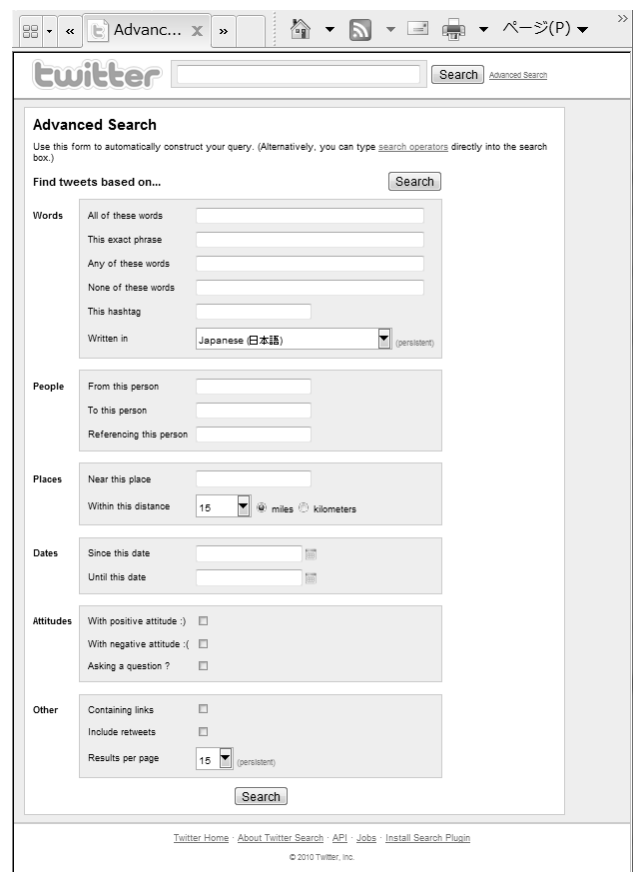


図 4 ツイッターの拡張検索画面

場所、主体、相手、等のメタデータを完備したオブジェクトとなっていることにある。

図4は、ツイッターのコメント（ツイート）群を検索するAPIをWebフォームの形で提供された拡張検索画である。

<http://search.twitter.com/advanced>

図中で、平易な英単語で表現されている項目名を、5W1Hに対応付けると次のようになる。

Words: WHAT

People: WHO

Places: WHERE

Dates: WHEN

Attitudes: HOW

さすがに、Why（「なぜこんなツイートを書いたかの発言意図」か）はない。また、ビジネス、消費で重要な、How much（いくら）、How long（納期）もない。ともあれ、発言（ツイート）の基本属性として、上記4W1H+α（リンクやリツイート（何かの引用であるか）の有無）が検索時に指定可能となっているのは興味深い。

5W1Hをもつのは、ツイートが、発言行為というイベントとしてとらえられているからである。僅か140字の短い曖昧な文字列は、文脈から切り離されて独立させられると意味をなしにくい。単なる文字列というモノではなく、フロー情報らしく、イベントの属性を与えられたことで、他のフロー情報やストック情報と結び付けて新たな価値を生みやすくなった。このことこそ、多彩なアプリケーションを生み出した本質的な理由ではないか、と筆者は考える。

3.2 最新の‘ツイート’から関心・意図を抽出可能なTwitSafe99

このようなマイクロ・ブログのシステムが、個々のコメント（ツイッターの場合には‘ツイート’）等に自動付与するメタデータだけで利便性は十分か、と問われれば、十分でない場合がある、というのが適切な回答であろう。例えば、ツイッターで、発言者の現在地は多くの場合自動で知ることができるのに対し、プロフィール欄にべた書きテキストで書かれた出身地の情報他、発言者に関する基本データ（メタデータ）は、自然言語解析、意味抽出を行わなければアプリケーションが自動認識して活用することはできない。

最新のツイート中に含まれる5W1Hをも抽出し、他のツイートのメタデータ等と対比、照合して、情報連携をはかることで、さらに利便性を向上させることができるのではなかろうか。

TwitSafe99は、企業ツイッターアカウント向け対話監視代行サービスである¹⁰⁾。ツイッターアカウント@kigyoへの書き込みや、指定したハッシュタグ、検索文字列等、を365日24時間リアルタイムで自動監視する。内部には有害情

報99¹¹⁾の意味解析エンジンと知識ベースを備え、オントロジーと構文解析、文脈解析により、ある程度有人監視に比肩できる性能を有する（有人vs自動で一長一短がある）。

現状、TwitSafe99による監視の判定結果には、有害度の大小2種類と無害扱いの3段階ある。有害度大の場合（明らかに違法ないしは公序良俗に反するもの）、書き込み者をブロックすることで様々なタイムラインへの表示を抑止する（書き込み者自身のタイムラインには表示される）。と共に担当者、関係者にその旨をリアルタイムでメール通知する。有害度小の場合、メール通知のみを行う。

有害度大と判定される書き込みは全体の0.1%未満。有害度小と判定される書き込みの数は、アカウント間で大きく異なるものの平均2.8%程度である。これらはネガティブな表現を伴うクレームだったり公開質問の類であることが多い。そのため、決してブロック対象にしてはならない（ブロックするという行為が炎上の引き金をひいてしまう）。迅速に慎重な対応を要する書き込みに気づかない事態を激減させ、最速で問題を解決し、公衆の信頼を回復、強化できるように支援するのが、TwitSafe99の主眼である。

TwitSafe99は、有害情報の一種として、詳細過ぎる個人情報（個人のフルネームや丁目以下を含む住所など）を、5W1Hメタデータ（個人情報）抽出APIにより検出可能である。このAPIを活用することにより、前述のように、最新のツイート中に含まれる5W1Hをも抽出し、他のツイートのメタデータ等と対比、照合して、情報連携をはかることできる。また、オントロジーとして、何か有用なフレーズ表現（例：「富士急／ハイランド／エヴァオフ会…行きたい／連れてって／出撃だ！」）を登録することにより、構文解析エンジンが、これらの意味内容に限定されたテキストマイニングを高精度に遂行できる。

このようにして、TwitSafe99は、最新の‘ツイート’から興味・関心、行動の意図・理由、といったものをある程度網羅的に、高精度で抽出することができる。

4. 個人情報（匿名化）活用と、メタデータ自動抽出・連携の将来市場

ユーザの最新の興味・関心、行動（発言）の意図・理由が得られたなら、それらをユーザのプロフィールや、過去の行動履歴と照合することにより、従来にない高い精度で直近の行動を予測し、支援情報が提供できるようになる。商業ベースでこのような情報を提供することは、パーソナル広告のリアルタイム提供に他ならない。

行動履歴×最新の発言の解析結果 = パーソナル広告

このように個人的な書き込み（ツイートなど）を活用するには、プライバシーへの十分な配慮が必要である。具体的には、その個人を特定できるような情報を伏せ字化し、十分な匿名化をはかってから、取扱い業者に渡す必要がある。

この際に、5W1Hメタデータ（個人情報）抽出API等により、一律に詳細部分を匿名化するアプローチが1つには

存在する。もし、一律な部分匿名化で不十分な場合には、 k -匿名化のエンジン¹¹⁾を使うアプローチがある。ユーザ情報データベースを活用し、対象となるユーザ群の所在や様々な属性情報をチェックし、ある部分を匿名化することによって、 k 人以下には個人が特定できないことを保証するアプローチである。

以上のような技術開発、運用管理ノウハウの蓄積、ならびに制度基盤（法制度、商慣習等）を整備することにより、パーソナル情報活用を応用した巨大な市場が開ける、という見通しが報告されている¹²⁾。SNS やマイクロブログにおけるメタデータ、属性情報の集合体を「ソーシャル・グラフ (Social Graph)」と呼ぶならば、SEO (Search Engine Optimization) に代わって、SGO (Social Graph Optimization) が巨大な市場を形成する可能性が十分ある。これは、急速に進展するソーシャル技術時代において、今後は、巧みなメタデータ活用こそがビジネスの成否を握る、という見通しに他ならない。

参 考 文 献

- 1) 国立国会図書館 データベース・ナビゲーション・サービス
http://dnavi.ndl.go.jp/bnnv/servlet/bnnv_user_top.jsp
 [accessed 2010-09-30].
- 2) API 比較・マッチングサービス
<http://www.api-match.com> [accessed 2010-09-30].
- 3) 吉田健一著. エンタープライズ 2.0 ～次世代ウェブがもたらす企業変革. 2007 年 9 月 1 日, インプレス R&D
- 4) OpenCalais
<http://www.opencalais.com/> [accessed 2010-09-30].
- 5) 5W1H メタデータ (個人情報) 抽出 API
<http://ma6api.mashupaward.jp/entry/70673/>
<http://ap.mextractr.net/demo> [accessed 2010-09-30].
- 6) Salestrackr
<http://code.google.com/p/sfdc-ubiquity/>
 [accessed 2010-09-30].
- 7) Tmeeting
<http://ma5.mashupaward.jp/oubo/201/>
 [accessed 2010-09-30].
- 8) OpenSocial Dashboard
<http://d.hatena.ne.jp/shinichitomita/20091130/1259576873>
 [accessed 2010-09-30].
- 9) メイドめーる
<http://www.maidmail.jp/top/about>
http://gigazine.net/index.php?/news/comments/20081111_maidmail/ (解説記事) [accessed 2010-09-30].
 ※片仮名と平仮名を逆にした「めいどメール」は、「冥途からのメール」という、異なるサービスのことである。
- 10) TwitSafe99
<http://www.metadata.co.jp/twitsafe99.html>
 [accessed 2010-09-30].
- 11) 「パーソナル情報保護・解析基盤」のご紹介
http://www.meti.go.jp/policy/it_policy/daikoukai/igvp/cp_jp/cat305/post-3.html [accessed 2010-09-30].
- 12) “4. 国際標準化及び市場動向に関する調査研究”. 平成 21 年度情報大航海プロジェクト報告書 (第 4 分冊)
http://www.meti.go.jp/policy/it_policy/daikoukai/igvp/index/h22_report/main/04.pdf [accessed 2010-09-30].

Special feature: Metadata. A survey of metadata extraction & exploitation in Japan with special emphasis on social media marketing. Naoyuki NOMURA, PhD (Metadata Inc., 702 Sankyo Bldg., 2-1-2 Koishikawa, Tokyo 112-0002 JAPAN)

Abstract: A real information explosion occurred when Web 2.0 reentered corporate CRM systems and workflows, so organizing unstructured or semi-structured UGC data became almost essential. A key technology to solving this problem is automatic Metadata extraction, which enables personal information hiding, automatic posting of calendar data, and creating huge market of personalized advertisement among others.

Keywords: Web2.0 / fact metadata / extraction / unstructured / semi-structured / UGC / personal information / calendar / personalized advertisement