

セマンティックなWebとLODクラウド によせて

～ソーシャルを支えるSemantic技術、
賢いクラウドの要件



2010.3.5

メタデータ株式会社

野村直之

“00”年代の情報爆発



“00”年代の前半はブロードバンドとWeb化
後半はWeb2.0とソーシャル

2009年初め：Post Web2.0

ネット全体で成功した“ソーシャル”は
企業でも本当にうまくいくのか？

- 懸念：圧倒的に少ない人口 (Web全体より4~5桁少ない)
- ファシリテータ不足、活性化スキル不足 (属人的資質に依存)
- 社内外のリソースを横断で関連付けるニーズが大
- キーワードレス(ゼロターム)検索など「ぞんざい」でリアルタイム性の高い要求

例：「去年の暮れに隣の事業部の偉いサンが出してインパクトのあった10頁位のレポートを出しておいてね。」

フランクで、本音で、'軟式に(ユルク)' つかるTwitterがウケる下地??

Semantic サービス群
2007-2009年初

<http://japan.cnet.com/blog/nomura/>

2009年の”Semantic Wave”より

<http://japan.cnet.com/blog/nomura/>



2010年はソーシャルを支えるセマンティックか

学会研究会もソーシャル花盛り 本日は、情報処理学会の自然言語処理研究会 & 情報学基礎研究会の合同研に久々に出席しました。オフィスから徒歩圏内で助かりましたが、内容は1年前よりずっと充実している印象。ソ...

同義語検索はシンプルなセマンティック検索

Switch on → スイッチオン 20日午後、慶大の坪田知巳教授がオフィスに来られ、メタデータとセマンティック技術が拓く、「編集」の新次元について、大変有益な議論ができました。そ...

2009初頭の10の予想・期待は当たったか？

ソーシャルの大波に洗われた2009年が終わろうとしています。私も、3年近く前にとった Twitterのアカウントnomuranを再開。Mashup Award 5thの土壇場で、Twitt...

“セマンティック・カフェ” by Metadata Inc.

“セマンティック・カフェ” by Metadata Inc.

The screenshot shows the 'Semantic Cafe' website, powered by Metadata Inc. The interface includes a search bar at the top, a grid of service icons, and a sidebar with navigation links. A central section displays a calendar for the month of January 2010, and another section shows a list of items with titles and descriptions, including 'IPAX' and 'DBpedia'.

2010年初の構図 ReadWriteWeb Top 10 Semantic Services



Google site:yelp.com yogurt silicon valley Search Advanced Search Database

Web Show options...

Swirls Frozen Yogurt - San Jose, CA
 4.5 stars 5 reviews - Price range: \$
 As mentioned in another review, the owner is an ex Silicon Valley techie who...
 We had an interesting talk with him concerning his yogurt...
 Show map of 6505 Meridian Ave, San Jose, CA 95120
 www.yelp.com/biz/swirls-frozen-yogurt-san-jose - 59K - Cached - Similar pages



大規模な
オンロジ
サービス



SBI HG ReadWriteWeb日本語要約

<http://corporate.sbihg.co.jp/category/staff-blog/2009/>

- [Top 10 Semantic Web Products of 2009](#)

- (1)**Googleの検索オプションとリッチスニペット**

- ⇒「[Googleに新しい検索機能追加](#)」、「[Google、リッチスニペットの提供開始](#)」

- (2)**Feedly: 雑誌風フィードリーダー**

- ⇒Google Readerなどのフィードリーダーに追加する形で動作するプラグイン型のアプリケーション。…各種ソーシャルサービスとの連携なども実現。

Feedly Miniではウェブサーフィン中にツールバーが出現。現在見ているページからお勧めページを Recommend してくれる機能なども搭載 (類例: StumbleUpon)。

- (3)**Apture: インテキスト型セマンティックエンジン**

- ⇒「[ショートカット型セマンティックサービス展望](#)」でも紹介。記事中の重要単語がハイライトされ、クリックすることで関連する情報を Wikipedia や GoogleMap や Flickr、YouTube などから取得・表示。

- (4)**Zemanta: ショートカット型セマンティックエンジン**

- ⇒WordPress や Blogger、MovableType など様々なブログに組み込み可能なショートカット型セマンティックエンジン。ブログの新規作成時、記事に関連する「リンク」、「画像」、「コメント」を自動的に取得し、ワンクリックで記事中に埋め込むことが可能。英語記事限定。

- (5)**Open Calais 4.0: メタデータ自動抽出エンジン**

- ⇒コンテンツを解析しメタデータを付与することで、そのメタデータに関連した情報を Wikipedia などのウェブ上から取得することが可能な API を提供。

- (6)**BBC's Semantic Music Project: BBCによるセマンティックプロジェクト**
⇒[MusicBrainz](#) という音楽に関連する情報を構造化したデータベースを作るプロジェクトと協力して BBC の音楽サイトで推進。サイト上で様々な関連情報が閲覧可能になった。

- (7)**Glue: ショートカット型レコメンドエンジン**

- ⇒「[Adaptive BlueのGlueに新機能追加](#)」でも紹介。閲覧中のページの内容をセマンティック解析し、そのページやそのページに関連のある項目・アイテムに対する友人の意見などを見ることができるアドイン機能を提供。友人や過去のお気に入り解析し、レコメンドを行ってくれる機能も。

- (8)**Freebase: セマンティックデータベース**

- ⇒Wikipedia のような様々なトピックに関する情報を構造化してデータベースに格納し各社に提供。顧客として Microsoft、The Wall Street Journal、Zemanta など。09年11月に1000万トピックを超えた。

- (9)**DBpedia: セマンティックデータベース**

- ⇒英語版を中心に Wikipedia から構造化されたデータを抽出し、RDF の形で提供。約300万トピックを収録。

- (10)**Data.gov: 政府機関の情報データベース**

- ⇒連邦政府の様々な機関が扱う情報・データを入手できるサイト。営利・非営利を問わず任意利用が可能。英国でも同様の取り組みを検討中。

- SBI HG:「セマンティックウェブが重要なのではなく、ユーザーがウェブ上でより快適に、便利に活動するためのサポートをするための技術の一つがセマンティックウェブ」

2010年のSemanticはSocialの仕組みを裏で
支え、ドライブするエンジン、連携の仕組みか

twitter

セーフティネットとして:



<https://twitter.com/SankeiShakaibu/status/3651489549>

Semantic Servicesの実用性が進化:



LinkedInの実用ソーシャルアプリへ



Twitterの'ボット'に!

Everi: Web上に掲載されている、コンテンツ、ニュースなどに登場する、人、場所、モノなどを関連付けて表示する検索エンジン。



Barack Obama Academic, Author, Country Leader, Lawyer, and U.S. Politician



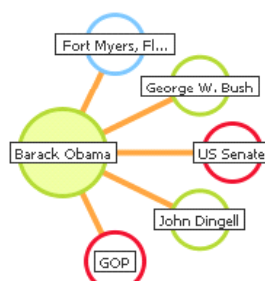
Barack Hussein Obama II is the 44th and current President of the United States. He is the first African American to hold the office. Obama was the junior United States Senator from Illinois from 2005 until he resigned following his election to the presidency. He was inaugurated as President on January 20, 2009. Obama is a graduate of Columbia University and Harvard Law School.

Description from Wikipedia

Show more

Find: barackobama

Barack Obama	Academic, Author, Country Lead...
Barack Obama, Sr.	Economist
Barack Obama presidential camp...	Election
George W. Bush	As part of his camp...
US Senate	34 minutes ago Ye...
John Dingell	
GOP	
US	
Abraham Lincoln	
Filter by category	
Politician	
US Politician	



Triplt! は5W1Hを元に旅程を最適化、提案してくれるサービス

http://japan.cnet.com/blog/nomura/2009/01/25/entry_27019732/

- 個人の旅程管理から、出張仲間の旅程共有へ

Basic Account: Upgrade Welcome, naoyuki nomura - Add Connections - Settings - Help - Sign Out

LinkedIn Home Profile Contacts Groups Jobs Inbox More... People Advanced

Applications FAQ Feedback Browse More Applications

My Travel

Application Preview

Your upcoming trips

December 2008
New York City, NY — Dec 13 - Dec 28
Flying to NYC for work, then down south to visit friends and family on the east coast.

Who's close to you

Scott Hintz Travel
Nov 18 - Nov 20, 08

Andy Denmark Travel

See where your entire professional network is traveling and when you will be in the same city as your colleagues. Meet up at the next industry event or re-connect with old friends. Add the My Travel application to display your current location, upcoming trips and travel stats within your network.

Raves about Triplt:

- "Triplt's straightforward approach makes it addictive..." - Walt Mossberg Solution, Wall Street Journal
- "Best for organizing travel details..." - CNN, Travel + Leisure
- "Easy and damn useful..." - DailyCandy
- "A terrific site that has changed my travel life..." - WNBC New York TV
- "It's kind of magical..." - Joel on Software

[Terms of Use](#) [Privacy Policy](#)

Application Info

Developed By **Triplt Inc.**
Category **Utility**

When you add an Application, you are allowing it to access profile information about you and your connections which the Application uses to provide the service. This and other information you provide to the Application may be displayed publicly depending on the Application's functionality.

By proceeding, you confirm your acceptance of the LinkedIn Terms of Use and you acknowledge that your use of the Application is subject to the Application developer's terms of use and to the Application developer's privacy policy.

Display on my profile
 Display on LinkedIn homepage

[Add application](#)

Triplt!

Welcome, Naoyuki
You're at home

Let's set up your application

Do you have an upcoming trip? (Optional)

Destination

Description

Trip dates

Do you have a Triplt account?

No, I want to sign up.
 Yes, I do.

Welcome, Naoyuki
You are at home
You have 0 upcoming trips

Who are the travelers in your LinkedIn Network?

Invite them to use My Travel so you can easily set up meetings and networking opportunities. You can invite up to 50 contacts at a time. **50 left.**

Jeff Char Japan	Pina Hirano Japan	Keiki Ichikawa Japan
Jun Iihara San Francisco Bay Area	Atsushi Nakatsugawa Japan	Makoto Nakazato Japan
Yuki Naotori Japan	Akiko Ohnuma Japan	Hiroyuki Okamoto Japan
Akio Shimada Japan	Kazuo Shimokawa Japan	Motohiko Tokuriki Japan
TATSUAKI YOSHIHIRO Japan		

[Skip](#) [Next](#)

ニーズに密着したSemantic応用の動向

【個人情報漏えい抑止】や【有害情報監視】もセマンティックの重要な領域：
<http://jp.techcrunch.com/archives/20100129tagged-erik-vogeler/>

【行動ターゲティング】ケータイが持ち主の行動先読み、情報を表示 東芝が開発
http://www.mmnews.jp/2010/01/post_919.html

→「次世代パーソナルサービス活用コンソーシアム」

140 ProofがTwitter広告の新境地:「色のついた」キーワードやフレーズを取り

出しているSuper Tweetは、有望なTwitter用広告モデルだ。

そこへ、140 Proof ...

<http://jp.techcrunch.com/archives/20100120140-proof-rolls-out-ad-network-for-twitter-clients/>

LOD クラウド:

本物の(smartな)クラウドか？

- ・ 「第24回産業構造審議会情報経済分科会つぶやきログ」 2010-02-16
- ・ 村上委員(グーグル): クラウドでキモなのはハードウェア。日本で広がっているのはクラウドのおもちゃ。前のままのデータセンターで名前を変えただけ。#sksjk
- ・ <http://twitter.com/HYamaguchi/status/9167168280> 村上委員: データセンターのおき方は変わってきている。環境問題もあって、データセンターの電力の2割はコンピューティングに使っていい。エフェクティブネスという考え方。グーグルは1.2を切り始めている。寒冷地で電力料金の安いところに設置する方向。
- ・ <http://twitter.com/HYamaguchi/status/9167245954> 村上委員: グーグルのデータセンターを日本に設置しようとする各種法令に違反することになり、できない。もはやデータセンターはコンピュータをどのくらい壊しながら使うか、というのが最先端。日本は考え方が3~5年遅れている。#sksjk
- ・ <http://twitter.com/HYamaguchi/status/9167309347>

NYTが過去100年分だかの記事のインデクシングに、24時間だけ100台のサーバを利用。そのためのSIもハード購入も必要無し:

Virtual Technolgy社より: <http://reflex.sourceforge.jp/scalablepdf.html>

米ニューヨークタイムス(New York Times)は2007年末、過去100年分の新聞記事をPDF化するためにAmazon EC2を使用しました。合計で数TBに及ぶ40万5000個のTIFF画像をPDF変換するのにAmazon EC2の仮想マシン100台を24時間使用したといいます。その際にAmazonに支払った料金は、EC2の使用料が240ドルで、データ転送量が1000ドル程度でした。(参考) [大規模インフラ個人運用: AWS+Hadoopの成功例](#)

老廃部品を自己廃棄したりして、自律的に健全に成長し、良いサービスを提供し続けられるのが本物のクラウド

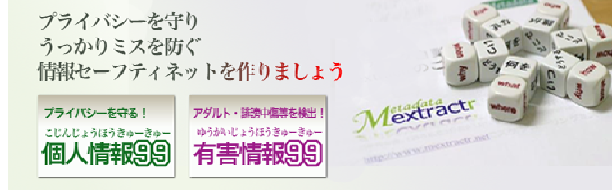
→情報も知識も新陳代謝と成長が必要

Azure と OGD I - Windows Live

- Azure をベースにした *Open Government Data Initiative* であり、ひとつの具体的なクラウド運用モデルを提示しているよう ...
- OGD I uses the Azure Services Platform External *Link* to make it easier to publish and use a wide variety of public *data* ...
→ ビジネスモデル、利用者課金モデルは？

結論に代えて

- ピンポイント検索、5W1Hでタイムリーに情報届ける (キャッチセールスの如く) 潜在ニーズは高い
- 行動ターゲティング用(Docomo, Toshiba)のメタデータは、システム由来に加えてコンテンツ(テキスト)解析由来が必要。
- 高精度なアドネットワーク、ビジネスマッチングへ。
- 個人情報漏えいや有害監視もセマンティックの重要な領域；



- LOD クラウドも本物の(smartな)クラウドのように自己成長し、老廃部品を自己廃棄したりして、ナレッジマネジメント領域で、常時役立って欲しい

Metadata

03-3813-5447
news@metdata.co.jp
メタデータ株式会社

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

<http://www.mextractr.net>

<http://www.emetadata.net>

<http://www.api-match.com>