

TOPPAN

HitachiSystems

UNISYS



Knowledge Works

Metadata

XML Consortium

XML Consortium

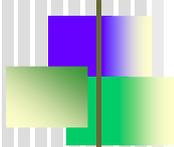
セマンティック・サービスの事例紹介

2009年5月13日

凸版印刷株式会社 松川昌洋
 株式会社日立システムアンドサービス 湯本正典
 ナレッジワークス株式会社 亀山悦治
 日本ユニシス株式会社 小林茂
 メタデータ株式会社 野村直之

© XML Consortium

1



目次

XML Consortium

1. 本講演の狙い、背景

- ➡ - なぜ今“セマンティック”が必要か?
- 現状の検索の問題点から考える
- Enterprise2.0 ,特にソーシャル関連市場予測

2. セマンティック・サービスの事例

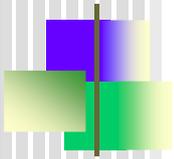
- Twine, Inform, Evri   
- hokia , Powerset, Zemanta   
- Faviki, Juice, iMa  
- TripIt! UpT  

3. セマンティック技術の実際

～5W1H Mextractr とOpenCalais  
 マッシュアップによる応用
 印刷・出版オンライン化で付加価値化を達成するヒント

© XML Consortium

2



検索の問題点



■ 検索者自身の知識の不足

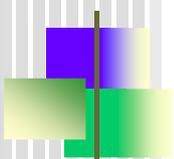
- 情報が断片的なもの
- カテゴリ/話す言葉が違うため、的外れな調べ方になるもの
- 専門的な判断が要求されるもの

■ 参照情報の不足/欠陥

- 別の単語が間違っ登録されたため、正しい判断が出来ないもの
- 知られていなさそうなもの
- どのように調べていいものかわからないもの

■ 参照情報間の連携がうまくリンクできていない

- 分布的なもの
- 情報の発生源
- 情報の時間軸を限定したもの



情報が断片的なもの

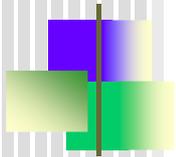


検索に必要な項目が不足しているために、対象が特定できないこと

例: **パスタ**と**カツ**が入った**郷土料理**って何?

答: 豚カツ・ピラフ・スパゲッティ・サラダが載せられている

長崎のローカルメニュー、**トルコライス**のこと



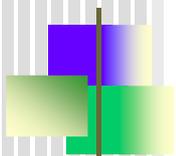
カテゴリ/話す言葉が違うため、 的外れな調べ方になるもの



以下のような問題のため、関係知識を知らない検索者はその情報を特定する用語に当たることが出来ない

- 専門家が知っているテクニカルターム/スラングを知らない
- (特に)スラングは地域によって定義が異なる

例: IT技術者が**デザイン**の知識を調べる場合、**金融**のことを調べる場合



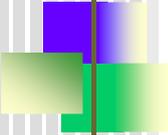
専門的な判断が要求されるもの



専門家が持ち合わせた**知識、経験**の検索

例:

- 自分が発明した内容と、他の関連特許との類似性
- 隣の家の窓がたまたま空いていたときに見てしまった場合の違法性
- 90円の製品に対して、消費税が4円と5円の店があるがなぜなのか



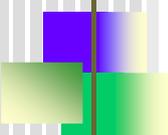
別の単語が間違っ て登録されたため、 正しい判断が出来ないもの

正しいだろう、という判断のもとで保存された情報がそもそも間違っている

例: 内臓、内蔵 (IMEの表示順)

森高千里、森下千里 (登録者の勘違い)

岸部四郎、岸部一徳 (登録者の見間違い)



知られていなさ そうなもの

一部の地域、集団、社会の中での利用頻度

例:

- 限定版CDが欲しいが、ほぼ完売している。人気のない(ここだったら売れ残ってるだろうと思われる)店がどこにあるか知りたい
- 見知らぬ土地でポスターを貼りたい。郊外で人通りの多い場所を知りたいQ

どのように調べていいものかどうか わからないもの

抽象的な表現しかできないもの

例:

フェルミ推定

- 砂浜に砂粒はどれくらいあるか
- 東京に美容師はどれだけ住んでいるか

等

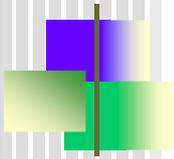
9

分布的なもの

特定の地域がどのような分布を持っているか

例:

- 今度～に出張するのだが、～にある店の雑煮はどんな味か
- 旅行先の服装を決めたいが、どんな格好で行くと良いだろう



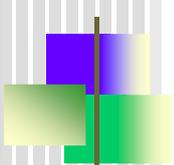
情報の発生源



不確かな情報に対しての根拠となる発生源の特定

例:

- PL法の話に出てくるレンジに猫を入れたという話は**どこの誰が行ったのか**
- ファイナルファンタジー(ゲーム)をだれが最初にファイファンと**呼んだか**



情報の時間軸を限定したもの



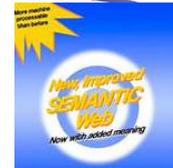
期間を限定した情報の探索

(サイトの作成、更新の日付だけで、当時のソースという情報がヒモ付けられていない)

例:

- 毒カレー事件の**発生当時**の状況と記事、背景を調べる
- ノストラダムスの大予言が信じられていたとき**ど**のようなことが起きていたのか

Semantic サービス群



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

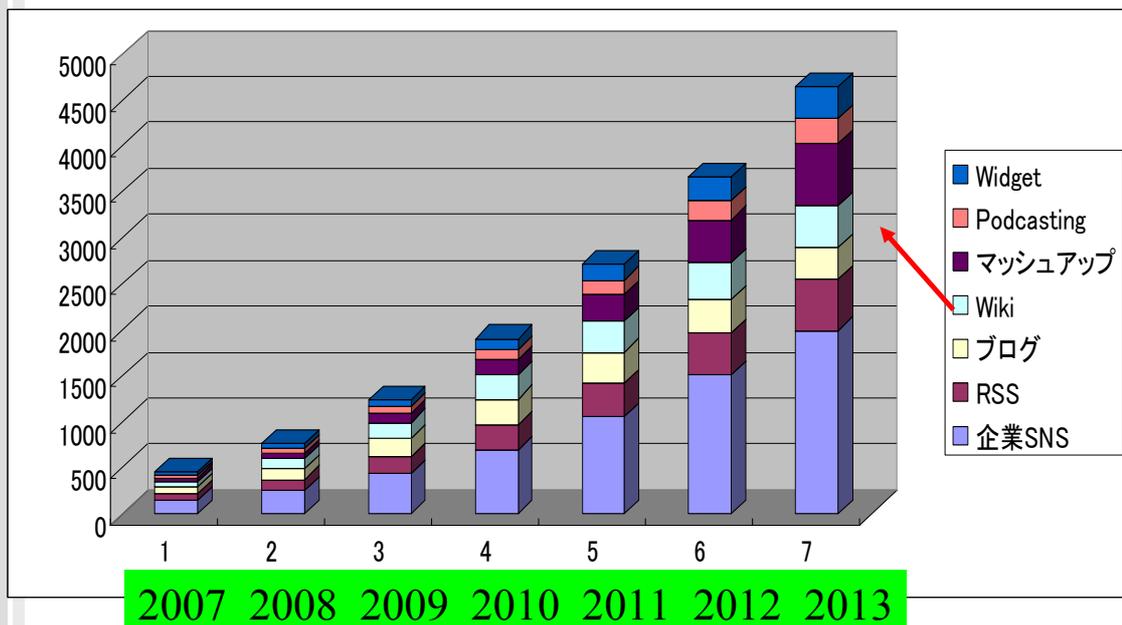
13

1. 背景

エンタープライズ2.0売り上げ予測

Forrester Research, Apr., 2008

単位: 百万米ドル



2007 2008 2009 2010 2011 2012 2013

企業SNS, ブログ, Wiki, RSS, S.ブックマークなどの「ソーシャル」が大半

14

1. 背景

“顧客2.0”対応企業戦略：グランズウェル ～ソーシャル・テクノグラフィで精密に2.0を導入



XML Consortium



ソーシャル技術が可能にした「集団としての人々の力（が高まりつつある状況）」のこと

企業が依って立つ大地がうねり隆起している..

独:8%, 日:22%, 韓:38% ←

日本では6%しかない ←
↓
ソーシャルブックマークが米国ほど流行らない！

ソーシャル・テクノグラフィ



Base: US adult online consumers
source: Forrester's NACTAS Q4 2006 Devices & Access Online Survey
12057 Source: Forrester Research, Inc.

15

© XML Consortium ↑ フォルスターGによるハーバード・ビジネススクールの講義録

1. セマンティック台頭の背景



XML Consortium

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？

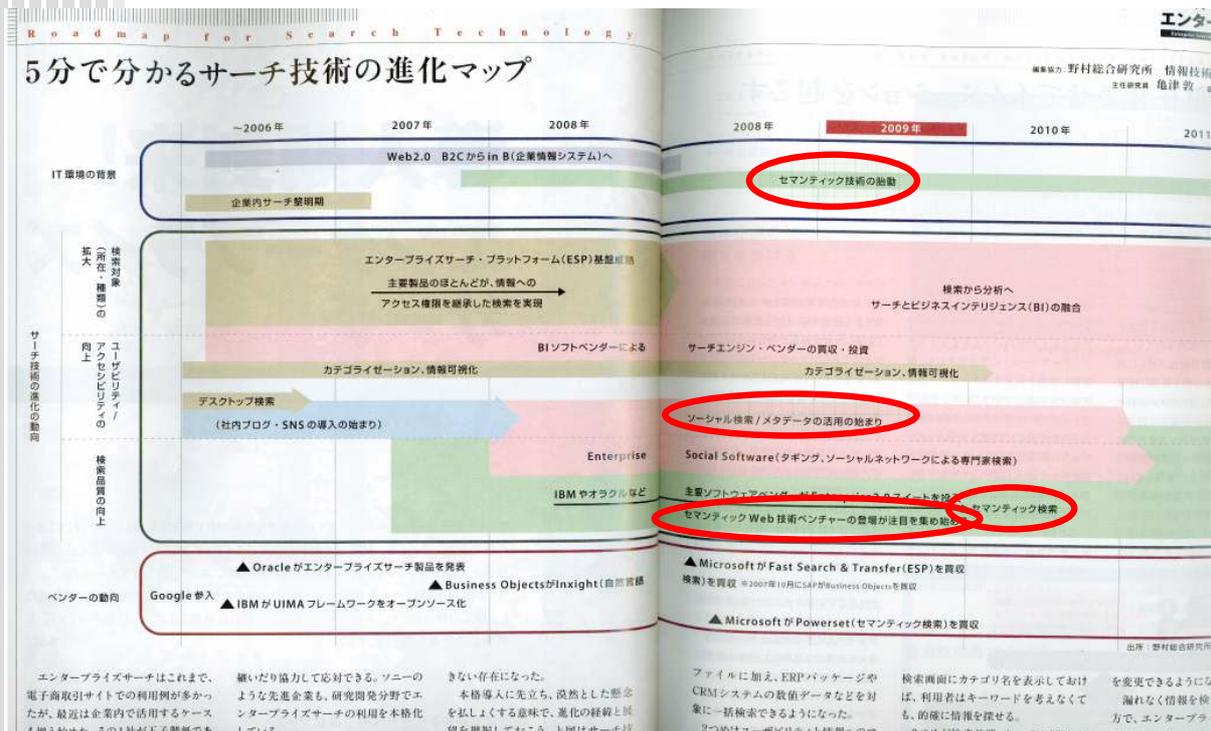
- ・ 懸念：圧倒的に少ない母集団 (4～5桁以上異なる)
- ・ ファシリテータ不足、活性化スキル不足
- ・ 社内外のリソースを横断で関連付けるニーズが大
- ・ キーワードレス(ゼロターム)検索など「ぞんざい」でリアルタイム性の高い要求

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。」

16

1. 背景

エンタープライズ・サーチが必要とする セマンティック技術、メタデータ活用



© XML Consortium

エンタープライズ・サーチあるいは、マスターデータを を整備しておくにもメタデータ活用が肝心



次世代情報システムを創造するリーダーのための専門誌
The Magazine for IT executives and professionals

IT Leaders

http://it.impression.co.jp/

CIO INSIGHT 掲載

特集
玉石混交から「玉」を見出す
エンタープライズ
サーチ

「特集」
並社の「人、モノ、金」を見いだす
今こそ実践!
マスターデータ
統合

2月号 No.05 2009

© XML Consortium

Check Point 4

文書検索を補完する機能

キーワードによる検索を補完する機能に過ぎないが、検索方法を工夫する製品もきちんと抑えておきたい。まずファストリサーチ&トランスファの「FAST ESP」は、インデックスを作成する際、抽出した語句を辞書に基づきカテゴリ分けする機能を備える(図8)。カテゴリには人名や地名、数値などのほか、企業固有の部品名や製品名、担当者などを設定できる。

検索結果を表示する際、カテゴリ分けして表示できるので、カテゴリから検索結果を絞り込める。契約書を探す場合、具体的な契約内容が分からなくても、カテゴリとして契約先企業名や契約した商品名が分かれば、そこから目的の契約書を探り込める。

ジャストシステムの「ConceptBase Enterprise Search」は、ビジネス文書によくあるフォーマットや表データの項目を指定して検索できる機能を持つ。表計算ソフトによる不具合報告書を例

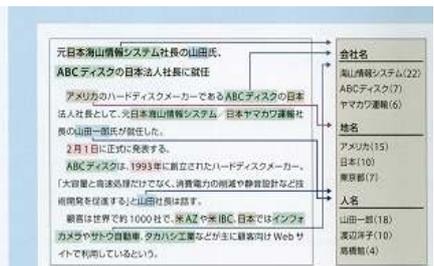


図8 「FAST ESP」の文書の自動カテゴリ機能の仕組み

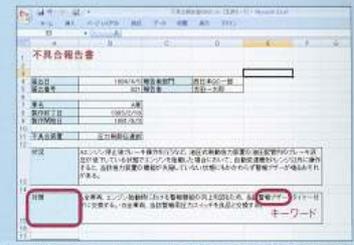


図9 「ConceptBase Enterprise Search」のエクセルの仕組みを利用した読み込み検索

ナレッジワークス株式会社による↑
FAST ESPのSaaS版、SCOPE
で利用可能 Knowledge Works 18

目次 (現在位置)



1. 本講演の狙い、背景

- Enterprise2.0 ,特にソーシャル関連市場予測
- なぜ今“セマンティック”が必要か?

2. セマンティック・サービスの事例

- Twine, Inform, Evri :



- hokia , Powerset, Zemanta



- Faviki, Juice, iMage



- TripIt! UpTake



3. セマンティック技術の実際

~5W1H Mextractr とOpenCalais

マッシュアップによる応用



印刷・出版オンライン化で付加価値化を達成するヒント

19

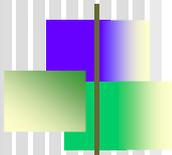
1. Twine

2007年10月に発表された、
セマンティックウェブサイトのさきがけ。

URL: <http://www.twine.com/>

米国 Radar Networks



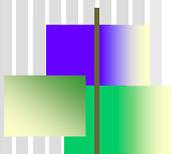


Twine 概要

一般Webユーザ向けの知識管理ツール。

個人専用のホームページが与えられ、情報を取り込み整理できる。

Semantic Webの技術(タグの自動付与)を用いている。



Twine トップページ

twine - Organize, Share, Discover Information Around Your Interests | Twine - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ScrapBook(S) ツール(T) ヘルプ(H)

http://www.twine.com/?%signout=t

Glue

twine - Organize, Share, Disco...

twine Take the Tour Explore Join Help

Welcome! Sign In

A smarter way to track interests

Collect — Add any bookmarks or any file type.
Share — Send what you find interesting to others.
Find — Search or get recommendations.

What are you into?

Go ▶
List a few of your interests and see what Twine has to offer.

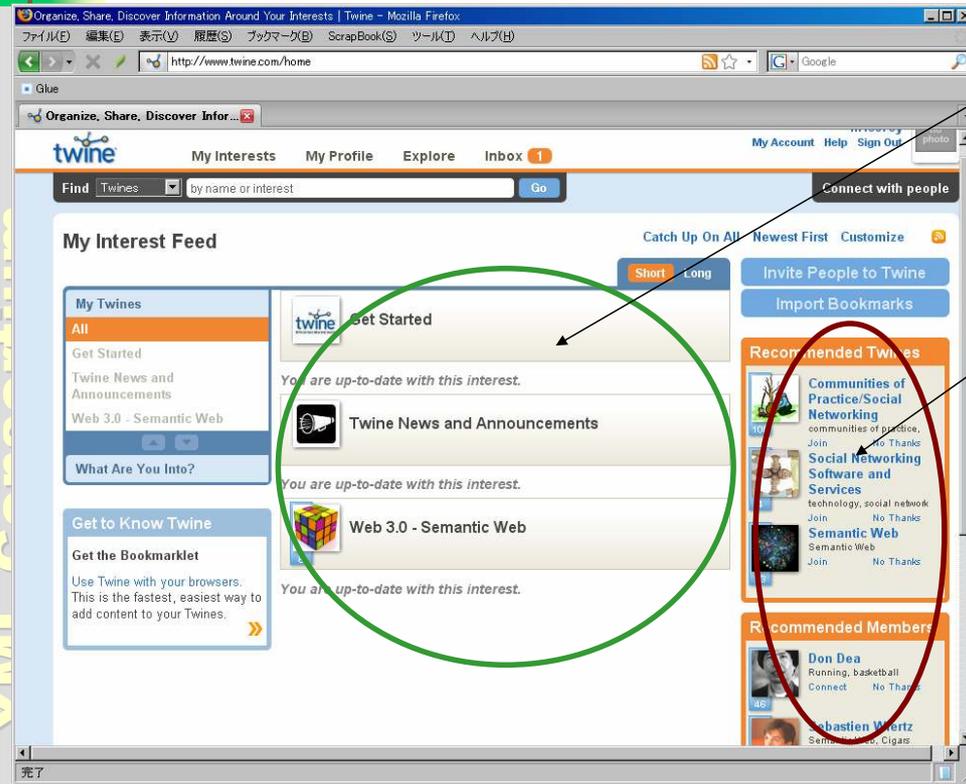
Join Twine!

Check out the [Most Active Twines](#) or the [Most Active People](#).

Featured Twines

- Web 3.0 - Semantic Web
- Cool/Funny TV Commercials
- Israel, Palestine and Peace
- Photography
- Visualization
- Current Science

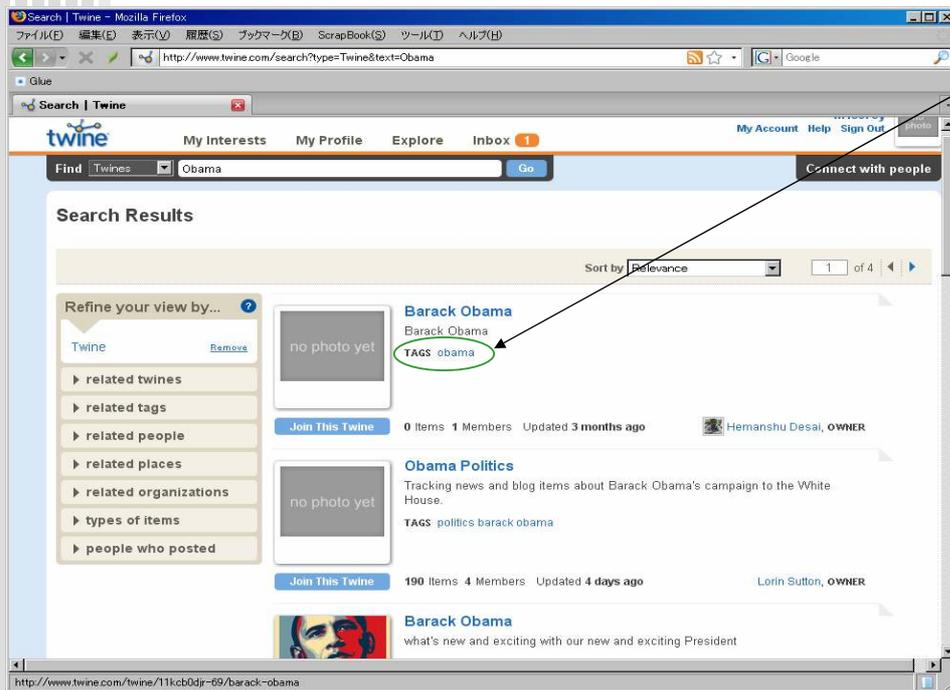
Twine ユーザの知識管理画面(ログイン後)



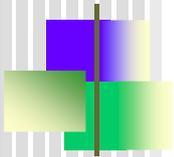
ユーザのお好みのコンテンツ

Twineからのお奨め

Twine 内での検索



タグ付き
TAGS
obama



Twine のその他の機能



・セマンティックさ

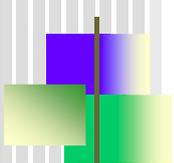
外部のコンテンツ(ブログやHPなど)をTwine内に登録できる。→タグが自動的に付与される。

・ソーシャルっぽさ

登録されたコンテンツは、Twine内のユーザで共有できる。

Twine内のユーザ同士が、「お友達」になれる。
→SNSのように。

25



Twine 所感



セマンティックさとソーシャルさの共存がうまく図られている。

ユーザにセマンティックの難しさやストレスを感じさせない。

→ 気軽に使えるセマンティックサイト

→ セマンティックのありがたみがはっきりとはわからない。

26

2. Inform

オンライン出版社のコンテンツを分析し、
出版社の独自のコンテンツアーカイブ、また関連したサイト、
ウェブ上の大量のコンテンツ
へのリンクを作成するサービス。

URL: <http://www.inform.com/>

米国 ニューヨーク Inform Technologies

27

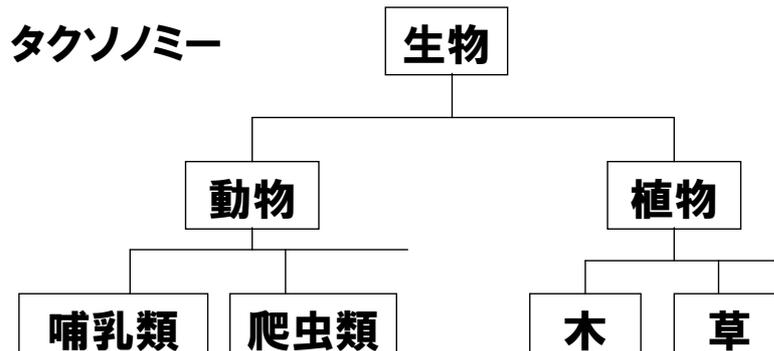
Inform 概要

毎日の大量のニュースコンテンツに含まれているキーワードを「セマンティックに」解析する。

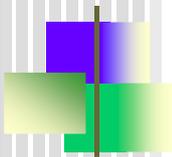
セマンティックに ← タクソノミーを利用している。

解析した結果、そのキーワードと関連の深いコンテンツをピックアップする。

ピックアップしたコンテンツとニュースにリンクをはる。



28



Inform トップページ



XML Consortium

Inform Technologies, Inc. - Mozilla Firefox

http://www.inform.com/

inform FOR **BLOGGERS** **PUBLISHERS** **DEVELOPERS**

Reinvigorate your website.

Inform API and Taxonomy

Here is a sample of what Inform's API can bring to your site: smart inline links, topic pages, related content, better SEO.

Get 10% to 20% growth to your site's traffic. We hold your hand to put it on your site, we customize it for your every need, and we have 100% uptime. [Learn more](#)

See it in action

Sign Up

- **Monetize** your web content.
- **Increase** your traffic organically.
- **Engage** users with relevant content
- **Optimize** your website for search engines

[Register Now! >>](#)

More than **100 leading media properties** use Inform's semantic technology to transform their websites.

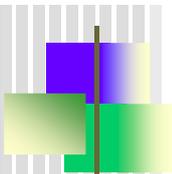
DAILY NEWS SI.COM Newsweek NBC nielsen IDG ZIFF DAVIS

Home - About - Terms © 2009 Inform Technologies, Inc.

http://www.inform.com/index.html?tab=2

© XML Consortium

29



Inform を実装したサイト



XML Consortium

Inform - Mozilla Firefox

http://www.inform.com/index.aspx

inform

Search News

Thursday, February 12, 2009 Login | Help | About Us

U.S. World Business Science & Technology Politics Sports Arts & Entertainment Health Culture & Lifestyle

My News

Top Stories

Flagged Articles

Sources

Thousands of your favorite news sources, all of their content, simply organized.

Subjects

Track subjects with unparalleled precision from news across the web.

Feeds

Add any RSS feed.

Continuing Coverage

US Presidential Election

Professional Football

2008 Hurricane Season

Information Technology Stocks

Mergers and Acquisitions

War in Iraq

Conflict in the Middle East

Oil Prices

Lead Stories

Taliban Storm Afghan Government Buildings

ABC News - My News

Three Afghan government buildings were attacked as eight armed Taliban militants stormed the Kabal offices, leaving at least 20 dead and 57 wounded.

related: stories subjects email flag

China's Lantern Festival, and an unfortunate ending

Boston Globe - My News

related: stories subjects email flag

Sudan dismisses Bashir arrest move

Al Jazeera - My News

related: stories subjects email flag

Historian: Lincoln ahead of his time

Chicago Sun-Times - My News

related: stories subjects email flag

American holds on for 52-50 win over Bucknell

Miami Herald - My News

related: stories subjects email flag

HSBC sees Asia recovering first from global slump

Reuters - My News

related: stories subjects email flag

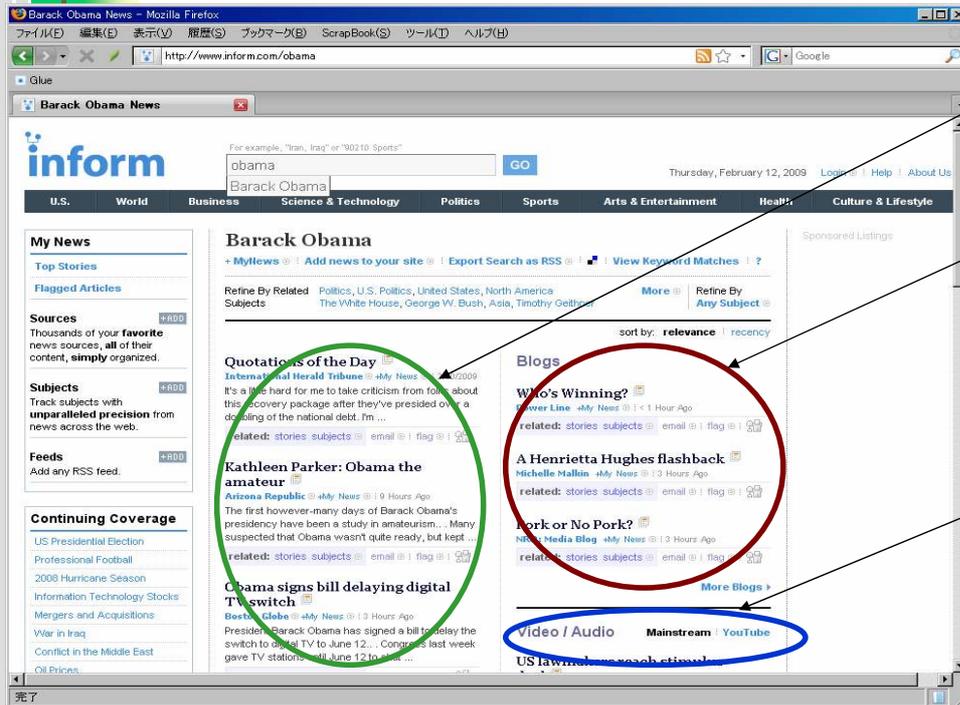
US World

完了

© XML Consortium

30

Informでの検索



ニュース

ブログ

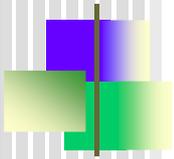
ビデオ

Inform のその他の事項

ブロガー向け、出版社向け、開発者向けの3種類のサービスが提供されている。

開発者向けのAPIが公開されている。

現時点で、100社を超えるクライアントを抱えている。→ CNN.com、Washington Postなど。



Inform 所感



関連するコンテンツとのリンクを「自動的に」はっ
てくれる。

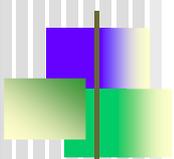
ブログ、動画、画像などいろいろなコンテンツに
対応している。

→ セマンティック技術ならではのサービス

→ 日本語のニュースサイトとリンクのはりかたが
異なる。

日本語では、まだサービスを開始していないのが
惜しまれる。

33



3. Evri



キャッチフレーズ:

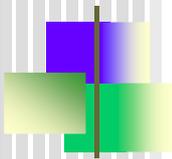
search less. understand more.

少ない検索で、より理解しよう。

URL: <http://www.evri.com/>

米国 シアトル Evri

34



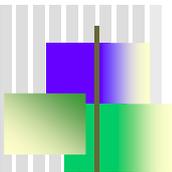
Evri 概要



XML Consortium

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。

cf. 日本語サイトの SPYSEE が似たサービスを提供してる。



Evri トップページ



XML Consortium

The screenshot shows the Evri website homepage. At the top, there's a navigation bar with menu items like 'ファイル(E)', '編集(E)', '表示(V)', '履歴(S)', 'ブックマーク(B)', 'ScrapBook(S)', 'ツール(T)', and 'ヘルプ(H)'. The address bar shows 'http://www.evri.com/'. The main content area features the Evri logo with the tagline 'search less. understand more.' and a sub-tagline 'Evri opens up a whole new way to explore connections -- between people, places, and things on the web and in the news.' Below this, there are sections for 'start with a popular person:' listing names like Barack Obama, Condoleezza Rice, and George W. Bush; 'find something you're interested in:' with a search input field; 'Are you a blogger or a publisher?' with a link to 'Learn more'; and 'or see what's going on in:' with sub-sections for 'business', 'entertainment', and 'sports', each listing various entities and names.

Evriでの検索

The screenshot shows a web browser window displaying the Evri website. The search term 'barackobama' is entered in the search bar. The results show a profile for Barack Obama, including a photo, a description, and a network diagram titled 'Top connections'. The network diagram shows Barack Obama connected to Fort Myers, Florida, George W. Bush, US Senate, John Dingell, and the GOP. The browser window shows the URL http://www.evri.com/person/barack-obama-0x16169.htm.

Evri 所感



Evriと似たサイトは、日本でも存在する。

SPYSEE

<http://spysee.jp/>

Evriのようなサービスが日本で実現されるのも、技術的には、近い未来のような気がする。

→ どういうビジネスモデルにするかが、課題。

SPYSEEはプライバシーの侵害との意見もあることに注意。

目次(現在位置)



1. 本講演の狙い、背景

- Enterprise2.0 ,特にソーシャル関連市場予測
- なぜ今“セマンティック”が必要か?

2. セマンティック・サービスの事例

- Twine, Inform, Evri   
- hokia , Powerset, Zemanta   
- Faviki, Juice  
- TripIt! UpTak  

3. セマンティック技術の実際

～5W1H Mextractr とOpenCalais  
マッシュアップによる応用

© XML Consortium

印刷・出版オンライン化で付加価値化を達成するヒント

39



・検索ワードの意味を解釈するセマンティック検索エンジン。

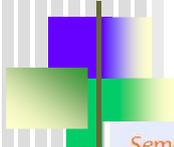
・検索結果はタブ表示され、「信頼できるサイト」タブには、hokia社が吟味した検索結果が表示される。

URL <http://www.hokia.com/>

米国 hokia, Inc.

© XML Consortium

40



hakia トップページ



XML Consortium

Semantic Search
on its way!



BETA
hakia®

 Ask

focus!

Focus better with clutter free,
organized, and quality search
results from the Web.



Contributed by
Librarians

hakia brings search results
from Credible Sites
recommended by Librarians.

How much should you stretch before exercise?



[Librarians' Corner](#)

[my hakia](#)

[Galleries](#)

[hakia API](#)

[h-CLUB](#)

[Compare to Google](#)

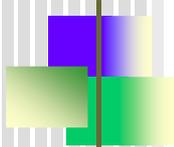
[health.hakia.com](#)

[pubmed.hakia.com](#)

[*make hakia your homepage](#) · [about](#) · [blog](#) · [terms of service](#) · [contact](#)

Copyright © 2007, 2008 hakia, Inc.

© XML Co



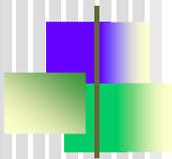
hakia 概要(1)



XML Consortium

- ・hakiaで検索を行う際、最初はWeb全体に対する検索結果が表示される。
- 「Credible Sites」のタブを開くと、司書やIT専門家の専門的な知識を生かした「信頼できるサイト」のコレクションから該当する検索結果が表示される。
- ⇒その結果「利用者にとって有益な情報」を返すことが出来る。

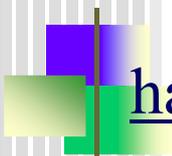
「われわれは、任意のページに含まれている概念やページ自体の信頼性といった要素の分析に、膨大な時間を費やしている。そうした要素を重視すると、返ってきた検索結果の数はそれほど意味を持たなくなる。大切なのは、検索結果が有している知識の断片の数だ」(バーカン氏)



hakia 概要(2)



- ・検索クエリ中の質問を検証するだけで、どのような概念が含まれているのかを理解できる。hakiaの開発者は、過去何年かにわたり手作業で概念マップを作り上げ、それらの関係性を把握したという。
- ・Webページ上の語句を基に検索結果を分類する代わりに、「QDEX(Query Detection and Extraction)」と呼ばれる技術を駆使して妥当性の高い結果を引き出す。
(QDEX: 日付、人名、出来事といった関連情報、すなわち知識の断片をWebページから収集し、より適切な検索結果を生成する手法)。
- ・「my Hakia」が用意されている、ユーザーはサインインし、自分の検索結果をパーソナライズすることができる。また、ブラウザのPlug-inとして使用できる「hakia ScoopBar」も用意されている。



hakiaとGoogleの検索結果を比較



When was John Lennon born?

(ジョンレノンは何年に生まれましたか?)

The image shows a side-by-side comparison of search results for the query "When was John Lennon born?".

- hakia (Left):** The search results are highly relevant and concise. The top result is "The Death of John Lennon -- A Cyclical Approach to F..." with a snippet mentioning his birth date. A yellow callout bubble points to the first result, stating "1番目に結果が表示されている" (The result is displayed in the first position).
- Google (Right):** The search results are more cluttered and less relevant. The top result is "John Lennon - Wikipedia, the free encyclopedia" with a snippet about his birth date. A yellow callout bubble points to the third result, stating "3番目に結果が表示されている" (The result is displayed in the third position).

カスタマイズ用メニュー

- add Headlines
- add News Search
- add Pubmed Search
- add YouTube Search
- add Alexa Traffic Chart
- add Weather
- add Quote of the Day
- add Cartoons (random)

「日本の今日のニュース」を検索条件として保存しておく、毎回表示される

こちらは「日本の昨日のニュース」

hakia ScoopBar と SearchBox

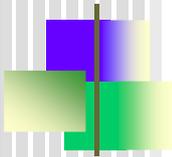
hakia ScoopBar

ブラウザにplug-inとしてインストールすれば、いつでもhakiaの機能が呼び出せる。

解説Video <http://company.hakia.com/scoopbar/scoopbarflash.html>

Search Box

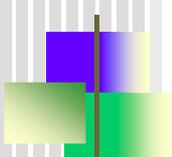
APIを直接埋め込まなくても簡単に自社サイトに組み込み出来る方法も提供されている。



hakiaが提供するAPI



- ・hakiaが用意するAPIを使うと、セマンティック検索アプリが提供できる。誰でもこのテクノロジーに乗り、自分だけのアプリが提供できる。
- ・API利用規約では、開発者とスタートアップは1日3万件まで検索が無料。超過するとライセンス契約が必要(料金は不明)。
- ・このAPIで実現できるアプリの例は下記の通り。
 - ・Web検索
 - ・ニュース検索
 - ・バーティカル検索(健康関連、エネルギー関連ほか)
 - ・Text Summarizer(テキスト概要作成)
 - ・Text Categorizer(テキストのカテゴリ分け)
 - ・Text Characterizer(テキスト特性化:“解説文やタグを特定・拡張するもの。検索エンジンマーケティング[SEM]には理想的”)
 - ・Text Meaning Representation (テキストの意味説明)



hakia 所感



- ・hakiaサイトでは、「信頼できるサイト」というタブが表示される。hakia社が厳選しているとのことだが、どう選別しているか分からない。
- ・「my hakia」というカスタマイズ機能も用意されている。ユーザーはサインインして、自分の検索結果をパーソナライズできることが便利だと感じた。
- ・APIを公開しており、Semantic Webを実現する基盤を提供している点はオープンで良い。これを使うと誰でもこのテクノロジーに乗り、自分だけのセマンティック検索アプリが開発できる。
- ・次々と新しいサービスを提供していること、英語以外の言語への対応にも取り組んでいることなど、今後の動向が期待出来る。

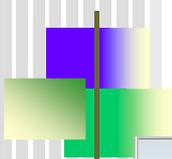
- ・単語による検索だけでなく、文章による質問形式で検索することが出来る。
- ・文章の内容を人工知能が理解し、知りたい事柄だけが検索できる。

URL

<http://www.powerset.com/>

米国 PowerSet, Inc.

49



PowerSet トップページ

PowerSet Powerlabs : Signup or Login

Ready. PowerSet. Go.

search wikipedia articles

Enter a topic, phrase or question Explore

search tips & examples

watch demo video

search & read
Wikipedia
articles

Discover Factz

For most people, places and things, PowerSet shows a summary of Factz from across Wikipedia.

Henry VIII	disolved :	monasteries	abbey,
	granted :	land, abbey, Thomas	
	married :	wife, Anne Boleyn,	
	founded :	professorships, Brist	

try these: [Henry VIII](#)
[David Hasselhoff](#)
[CIA](#)

less

Unlock Meaning

PowerSet finds articles related to the meaning of your query. And sometimes even direct answers.

Wikipedia

Items by David Lynch

try these: [what awards did No. Country for Old Men win](#)
[who discovered the neutrino](#)
[how many dimensions does](#)

less

Scan Summaries

PowerSet summarizes pages for easy browsing, with tools that follow you as you read and explore.

Show Outline Show Factz

The Daily Show

The Daily Show won Emmy Award
The Daily Show won Peabody

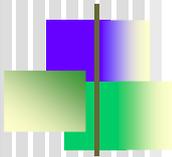
try these: [Harry Potter and the Deathly Hallows](#)
[Martha Stewart](#)
[The Simpsons Movie](#)

less

About PowerSet FAQ Blog News Careers Contact Feedback Terms of Use Privacy Policy Site Map

Microsoft

50



Powerset 概要



・Powersetは、キーワード検索ではなく、情報検索という考え方。

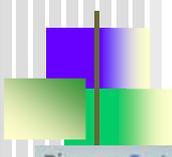
質問形式で入力すると、その答えを返す。

・概念や文章の構造を理解した上で適切な回答を探すという新しいアプローチ。自然言語検索であり、前置詞を正しく解読することで、文章の意味や文脈を読み取り検索結果を返す。

・現状は、Wikipedia、 α 版Freebaseのみが対象となっている。

・2008年7月に米国マイクロソフト社が買収

・別途APIも提供されている。



Powersetの主な機能



Discover Factz

For most people, places and things, Powerset shows a summary of Factz from across Wikipedia.

Henry VIII	dissolved :	monasteries, abbey,
	granted :	land, abbey, Thoma
	married :	wife, Anne Boleyn, .
	founded :	professorships, Brist

関連記事のリストだけでなく、クエリー対象についての情報を両サイトから集めて「サマリー」を表示。

「Factz」機能により、そのクエリー対象が「何をしたか」という事実を抽出し、一覧表示。

Unlock Meaning

Powerset finds articles related to the meaning of your query. And sometimes even direct answers.



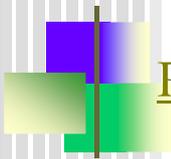
クエリーが質問文の場合、その答えとなる部分をWikipediaとFreebaseから抜き出し、ハイライト表示する。

Scan Summaries

Powerset summarizes pages for easy browsing, with tools that follow you as you read and explore.



ユーザーが検索結果のリストから、好きなWikipediaもしくはFreebaseの記事を選んでクリックすると、オリジナル記事のほか、右欄外にその記事の「概要」や「Factz」が表示され、ユーザーが望む回答をより見つけやすくしている。



PowersetとGoogleの検索結果を比較(1)



When was John Lennon born?

(ジョンレノンは何年に生まれましたか?)

Powerset
Wikipedia Articles

When was John Lennon born?

John Winston Ono Lennon: Date of birth

October 9 October 9, 1940 (68 years ago)

Wikipedia Articles: results 1 - 10 of 1736

- John Lennon** John Charles Julian Lennon was born in Sefton General Hospital on 8 April 1963. ... | Birth name | John Winston Lennon |
- Julian Lennon** John Charles Julian Lennon (born 8 April 1963), known universally as Julian Lennon, and by some fans as Jude, is an English singer, songwriter, musician, and first son of Beatle John Lennon and the only child of Lennon's first wife Cynthia Lennon.
- List of compositions for guitar** 1981 John Anthony Lennon (born 1950) Another's Fandango
- Cynthia Lennon** After Cynthia had been in labour for 24 hours, John Charles Julian Lennon was born on 8 April 1963, at 6:50 a.m. in the Sefton General Hospital in Liverpool.
- Death of John Lennon** It has become a gathering place for tributes on Lennon's birthday and on the anniversary of his death, as well as at other times of mourning, such as after the 11 September attacks and following Harrison's death on 29 November 2001. ... John Lennon
- Julia Baird** Baird had an older half-brother, John Lennon (9 October 1940 – 8 December 1980) who was born during her mother's marriage to Alfred Lennon, and an older half-sister, Ingrid Pederson (born Victoria Lennon, 19 June 1945, after Julia's affair with a Welsh soldier, but who was given up for adoption in 1945) and a younger sister, Jacqueline 'Jacqui' Dykins (born 26 October 1949).

先頭に結果が表示されている

Web Images Maps News Shopping Gmail more
etsuji.kameyama@gmail.com | Web History | My Account | Sign out

Google When was John Lennon born? Search Advanced Search Preferences

Web Results 1 - 10 of about 15,300,000 for When was John Lennon born?. (0.24 seconds)
Cloudlet: Tags Sites Net Off Tell a friend about Search Cloudlet

John Lennon - Wikipedia, the free encyclopedia
John Charles Julian Lennon was born in Sefton General Hospital on 8 April 1963. [107]. Lennon was on tour and would not see Julian for three days, ... en.wikipedia.org/wiki/John_Lennon - 335k - Cached - Similar pages -

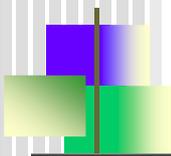
Julian Lennon - Wikipedia, the free encyclopedia
1 Feb 2009 ... John Charles Julian Lennon (born 8 April 1963), known universally as Julian Lennon, and by some fans as Jude, is an English singer, ... en.wikipedia.org/wiki/Julian_Lennon - 42k - Cached - Similar pages -

3番目に結果が表示されている

The story of the Liverpool Lenons: John Lennon
John Lennon was born to the sound of Hitler's bombs in Liverpool on the night of October 9th 1940. His mother Julia gave him the middle name Winston in ... www.lennon.net/lenons/john_lennon.shtml - 10k - Cached - Similar pages -

John Lennon Portfolio - Internet Beatles Album
20 Jan 2002 ... John Winston Lennon was born in Liverpool on October 9, 1940, during the height of WWII, his father, Fred Lennon, off at sea ... www.beatlesagain.com/bjohn.html - 14k - Cached - Similar pages -

WikiAnswers - Where was John Lennon born
History Politics and Society question: Where was John Lennon born? Answer John Winston Lennon was born in the Oxford Street Maternity Hospital in Liverpool ...



PowersetとGoogleの検索結果を比較(2)



Powerset
Wikipedia Articles

When was John Lennon born?

John Winston Ono Lennon: Date of birth

October 9 October 9, 1940 (68 years ago)

Wikipedia Articles: results 1 - 10 of 1736

- John Lennon** John Charles Julian Lennon was born in Sefton General Hospital on 8 April 1963. ... | Birth name | John Winston Lennon |
- Julian Lennon** John Charles Julian Lennon (born 8 April 1963), known universally as Julian Lennon, and by some fans as Jude, is an English singer, songwriter, musician, and first son of Beatle John Lennon and the only child of Lennon's first wife Cynthia Lennon.
- List of compositions for guitar** 1981 John Anthony Lennon (born 1950) Another's Fandango
- Cynthia Lennon** After Cynthia had been in labour for 24 hours, John Charles Julian Lennon was born on 8 April 1963, at 6:50 a.m. in the Sefton General Hospital in Liverpool.
- Death of John Lennon** It has become a gathering place for tributes on Lennon's birthday and on the anniversary of his death, as well as at other times of mourning, such as after the 11 September attacks and following Harrison's death on 29 November 2001. ... John Lennon
- Julia Baird** Baird had an older half-brother, John Lennon (9 October 1940 – 8 December 1980) who was born during her mother's marriage to Alfred Lennon, and an older half-sister, Ingrid Pederson (born Victoria Lennon, 19 June 1945, after Julia's affair with a Welsh soldier, but who was given up for adoption in 1945) and a younger sister, Jacqueline 'Jacqui' Dykins (born 26 October 1949).

検索結果の先頭のした↓をクリックすると、ページを移動することなく、内容を確認することが出来る

Julia Baird Baird had an older half-brother, John Lennon (9 October 1940 – 8 December 1980) who was born during her mother's marriage to Alfred Lennon, and an older half-sister, Ingrid Pederson (born Victoria Lennon, 19 June 1945, after Julia's affair with a Welsh soldier, but who was given up for adoption in 1945) and a younger sister, Jacqueline 'Jacqui' Dykins (born 26 October 1949).

Article

Julia Baird

Julia Baird (née Dykins) (born 5 March 1947) is the eldest daughter of John 'Bobby' Albert Dykins (1918 - December 1965) and Julia Lennon (12 March 1914 – 15 July 1958). Baird had an older half-brother, John Lennon (9 October 1940 – 8 December 1980) who was born during her mother's marriage to Alfred Lennon, and an older half-sister, Ingrid Pederson (born Victoria Lennon, 19 June 1945, after Julia's affair with a Welsh soldier, but who was given up for adoption in 1945) and a younger sister, Jacqueline 'Jacqui' Dykins (born 26 October 1949).

Lennon started visiting the Dykins' house in 1951. After the death of Julia Lennon in 1958, Harriet and Norman Birch were appointed guardians of Julia and Jacqui, ignoring Dykins' parentage, as he had never legally married their mother. Lennon invited the Dykins sisters to visit after the success of The Beatles, when he was living in Kenwood, Weybridge, with his wife, Cynthia Lennon.

Julia Dykins (Baird) married Allen Baird in 1968 and moved to Belfast. They had three children together but were divorced in 1981. Baird worked as a special needs teacher, and after Lennon's death she wrote John Lennon, My Brother (with Geoffrey Giuliano) and gave up working in 2004 to write Imagine

1 relevant passage

Google検索の何が問題になりそうか？



Google

世界中のページをクロールして、インデックス!! その結果大量のデータが溜まり...

利用者は必要な情報が見つけれられる？

Powerset

利用者が求めているのは情報量ではない!!

入力された検索条件と関連性が高い重要な情報が上位に表示されることが重要

Powersetは、それを解決する技術を提供

Powersetを使用した日本語の実験サイト



• PowersetがYahoo Pipesで利用できる。日本で、Powersetを利用した実験サイトが公開されている。

<http://www.machu.jp/jawerset/>

Jawerset

Search with Powerset in multi language - Powerset を多言語で検索

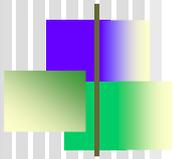
自然な言葉で検索できるPowersetを日本語で使えるようにしました。
検索語句を英語に変換してPowersetで検索しているだけなので、精度は高くありません。ネタとしてどうぞ。

JAPANESE

Search with Powerset!

Powered by Powerset, Yahoo! Pipes, and Google AJAX Language API.

- 翻訳エンジンにはGoogle AJAX Language API
- Powerset の検索には Yahoo! Pipesが使われている。
- Pipes では、Fetch Page を使用しPowerset のページを取得後JSONPに変換。Powerset のページを取得する Pipesと、呼び出し側JavaScript のみ実装。



Powerset 所感



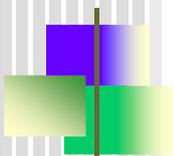
・検索対象は、Wikipedia(英語版)と、オンラインデータベース「Freebase」のみであるため、取扱い情報が限定されており、(少なくとも日本では)現時点での実用性は感じられない。

しかしながら、「自然言語検索の技術を使い、人間の言葉による質問に対して、それに適した回答を探し出してくれる」

「グーグルが検索キーワードに最も関連のある「ページ」を提示してくれるのに対して、Powersetは「情報そのもの」を提示してくれる」

という技術は、まさしくセマンティック的であり、次世代検索技術の方向性を示していることについて期待できる。

57



Zemanta



・入力情報に関連ある記事をウェブから探し出し、コンテンツを充実させることを手助けするプラットフォーム。

・入力された記事の内容を解析し、関連画像、リンク、記事をウェブ上から探し出す。

URL <http://www.Zemanta.com/>

英国 Zemanta Ltd.

58

Zemanta Popular! Super Bowl XLIII, Cooking, Pittsburgh Steelers, Business And Economy, Kids And | More popular >>

Zemanta+

Download | API | Demo | FAQ | Tour | Community | Blog

Get free blogging assistant in seconds!

Automatic links



Associated photos



Related articles



Automatic tags



Learn more



Download Zemanta
Firefox Extension
0.5.1 for Firefox 2.5.3

Other Browsers, Live Writer, or Server-side plugins

Zemanta API for your App or Business

You can now use Zemanta semantic API to boost your business with smarts

Read more about API and conditions



Sign up for our monthly newsletter: e-mail address

SHARE

© 2008 Zemanta Ltd. All rights reserved. [Terms of Service](#) | [Privacy](#)

[About](#) | [Contact](#) | [Press](#) | [Fan](#) | [Developer](#)

59

Zemanta 概要

・例えばブロガーが記事を入力すると、その自然言語を解析し、幾つかのサイトから関連情報(写真、文字情報)を自動抽出する。画面に表示された情報をブロガーが選択するだけで容易に自分のブログに取り入れられる。

・写真は、Flickr等から二次利用が許諾されているもののみ表示、文字情報は自分が書いているブログの特定の箇所と容易にリンクをはることができる。

・Plug-inをインストールすることによりブラウザで利用可能。APIも提供されている。

60

Zemanta Plug-inのインストール



- Firefox用とIE用の2種類
- インストールは容易

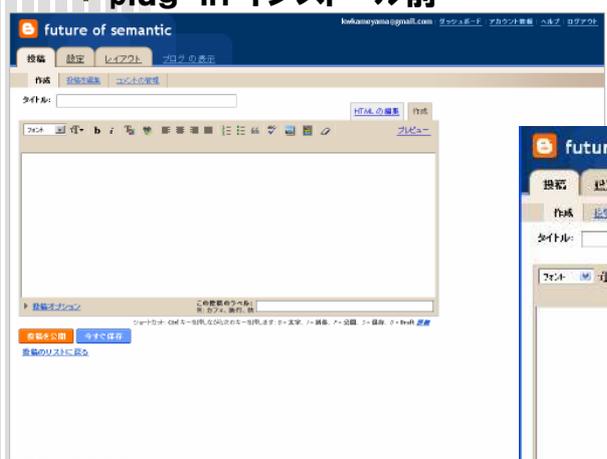


XML Consortium

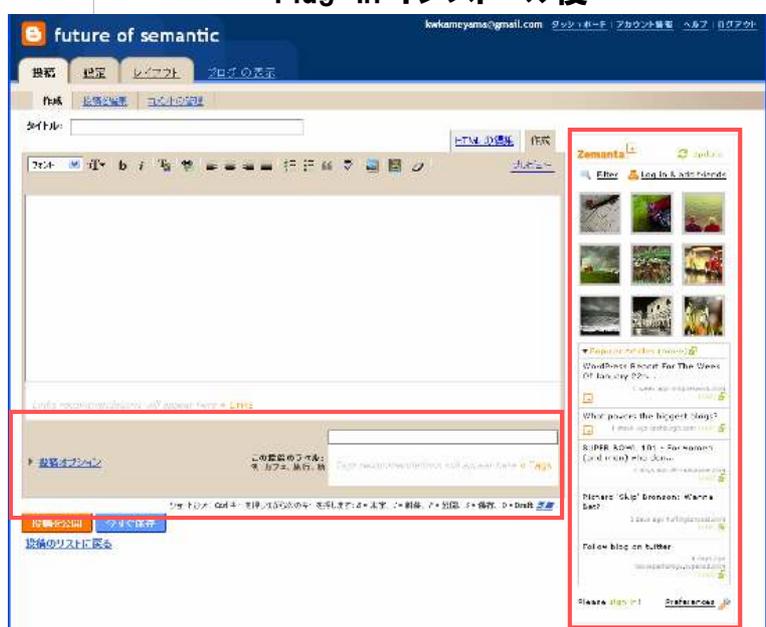
Zemanta Plug-in がインストールされると



plug-in インストール前

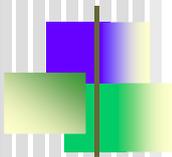


Plug-in インストール後



bloggerサイトでブログ記事を入力する画面に機能が追加される

XML Consortium



利用例:「blogger の場合」



完成!! 簡単に情報を集める上、見栄えの良いブログ記事を短時間で作ることができる。

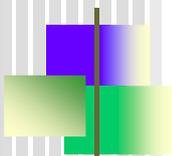
Filter by my custom keywords

Associated photos

Automatic links

Automatic tags

Related articles



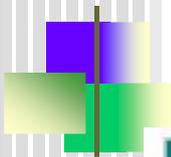
対応しているブログ・プラットフォーム(英語サイト)



- [Wordpress.com](#)
- [Self-hosted Wordpress.org 2.1+](#)
- [Blogger.com](#)
- [TypePad](#)
- [Ning](#)
- [MySpace](#)
- [LiveJournal](#)
- [MovableType](#)
- [Tumblr](#)
- [Drupal](#)

Connecting content from 10+ databases, 10.000+ sources and your preferred feeds

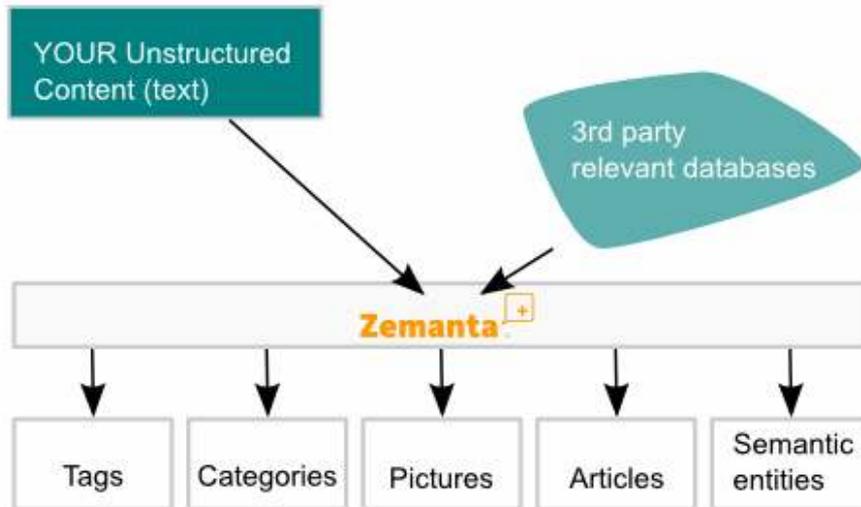




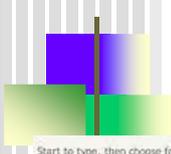
Zemanta-APIの構成



XML Consortium



Zemantaは同社のテクノロジーを広くブログ・プラットフォーム、メディア・サイト、その他スタートアップ、コンテンツ・データベース運営者に利用可能とした。
 Zemantaのセマンティック検索エンジンを利用すれば、構造化されていないコンテンツから関連ある情報を収集してサイトの内容を強化することが可能となる。



Zemanta-APIの利用事例



<http://www.faviki.com/>



<http://semantalyzr.com/>



<http://code.zemanta.com/bostjan/ubiquity/>

- ・Zemantaサイトでは、何が出来るかが明確に示されている。ブロッガー向けのツールであるため実用的である。Semantic Webの技術を利用していることを利用者自身が分かる。
- ・Zemantaサイトは、Wikipedia,Blogger,MovableTypeなど特定のブログツール、サイトを対象としているが、日本語に対応していないことは残念。
- ・Zemanta-APIを利用した幾つかのサイトが既に公開されており、Semantic Webを実現する基盤提供の影響度は大きいと感じる。

目次(現在位置)

1. 本講演の狙い、背景

- Enterprise2.0 ,特にソーシャル関連市場予測
- なぜ今“セマンティック”が必要か?

2. セマンティック・サービスの事例

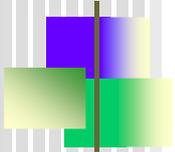
- Twine, Inform, Evri :   
- Zemanta, Powerset, hokia   
- Faviki, Juice, iMage   
- TripIt! UpTake  

3. セマンティック技術の実際

～5W1H Mextractr とOpenCalais
マッシュアップによる応用



印刷・出版オンライン化で付加価値化を達成するヒント



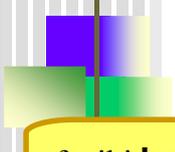
faviki (1): 概要



- ソーシャルブックマーク機能を提供する。
- ソーシャルタグの付け方に特徴がある。
 - ブックマークレットによりタグを付ける。
 - Wikipediaにある用語をタグとして利用する。

- URL <http://www.faviki.com/>
- 創設者 Vuk Milicic

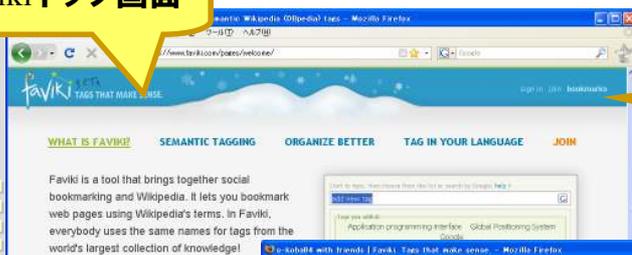
XI



faviki (2): 初期画面



faviki トップ画面



Join時に、ユーザ名とともに使用言語も選択する。
14言語選択可能

サインイン後の画面



日本語を選択すると、タグは日本語で表示される

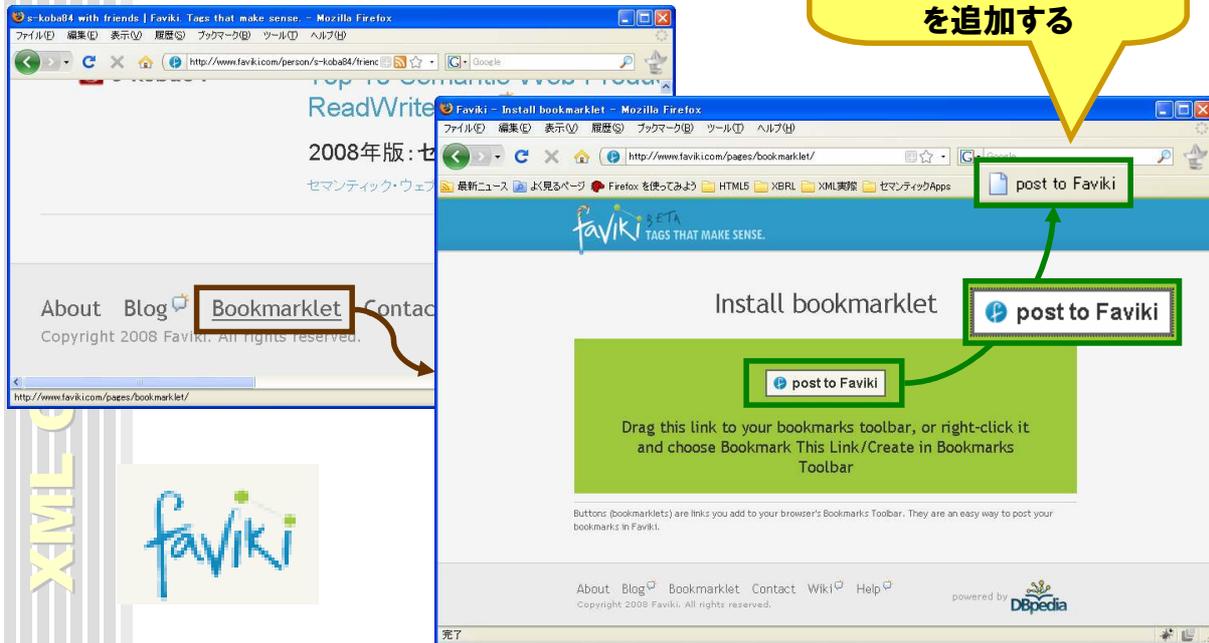
XI

faviki (3):ブックマークレットの登録



Webページにタグを付ける場合には、
事前にブックマークレットをインストールする

ドラッグして
ブックマークツールバーに
"post to Faviki"
を追加する



XML



faviki (4):タグの設定



Webページにタグを付けたい場合、
ブックマークレットからタグを設定する



Wikiにある項目が
提示される。
(次スライド参照)

②ブックマークレット
から登録する

①登録する
画面を表示
する

⑤④にない場合、ここ
からタグを指定する。

③指定したタグが
ここに表示される

④タグの候補が提示され
る。クリックすると、
③の領域に移動する。

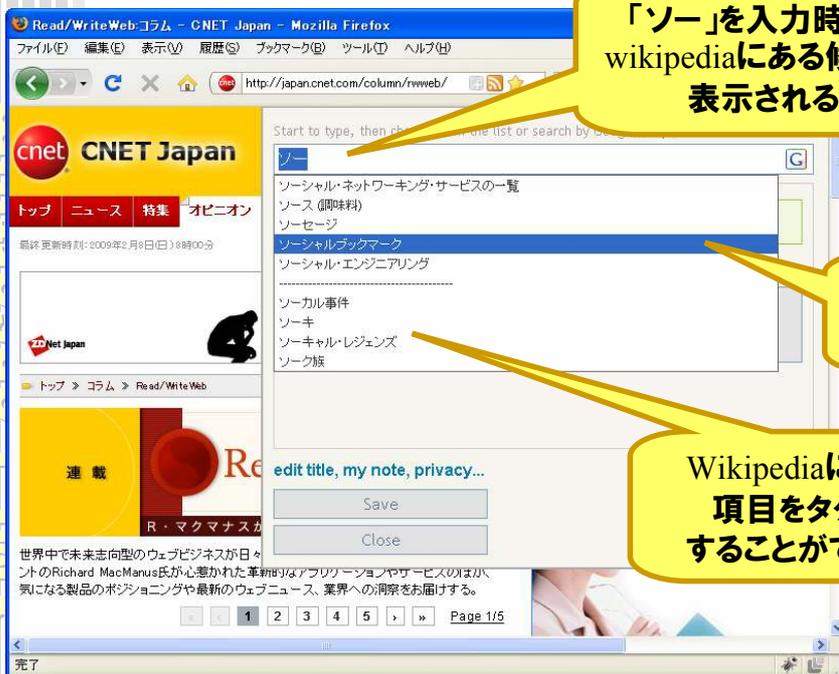
そのページに合った
タグが推奨される



XML Consortium

faviki (5): タグ候補の選択

付加できるタグは、wikipediaにあるものだけとなる。
タグのゆれが無くなる。



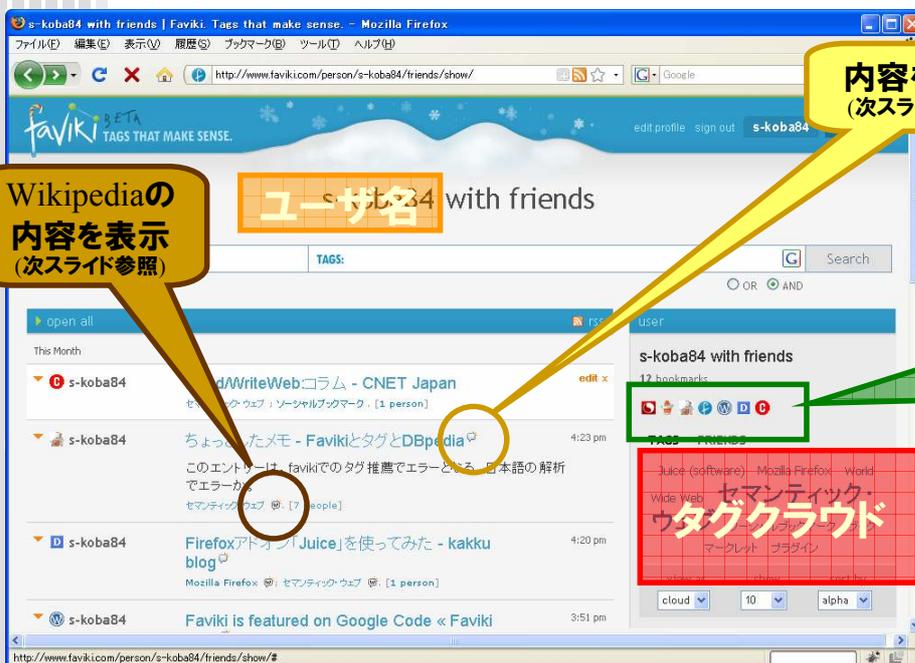
「ソー」を入力時に、
wikipediaにある候補が
表示される

候補を選択して
タグを決定する

Wikipediaにある
項目をタグと
することができる

faviki (6): 他情報へのリンク付け

自分で登録したタグを元に、Webページを見つける場合、
タイトル以外にも情報を捕捉してくれる。



Wikipediaの
内容を表示
(次スライド参照)

内容を表示
(次スライド参照)

ファビコンをク
リックしてその
項目だけを表
示可能

その項目だけ
を表示可能

faviki (7): 関連情報の表示



Wikipediaの内容を表示

内容を表示

faviki (8): タグ付けの特徴



- 自動タグ付けではない。Wikipediaにあるタグだけを指定する。
- タグ付けのゆれ（同義語の問題）を解決している。
- 日本語と英語の結びつけが可能である。
 - タグの一意性を英語版wikipediaで確保
 - 日本語版Wikipediaとの対応付けは、DBpediaによる。
 - Googleの言語翻訳APIも利用

faviki (9): 日本語ページの扱い



日本語のページを扱うために、まず英語に翻訳する。
 関連するタグを得るためにZemanta APIを利用する。
 推奨タグを日本語化するために、DBpediaを利用する。

XML Consortium



Google翻訳を利用

Zemanta APIを利用

DBpediaを利用
(次スライド参照)

翻訳の精度による
問題も含む

ページから
タグを抽出する

元になる各国のwiki
による相違も問題あり

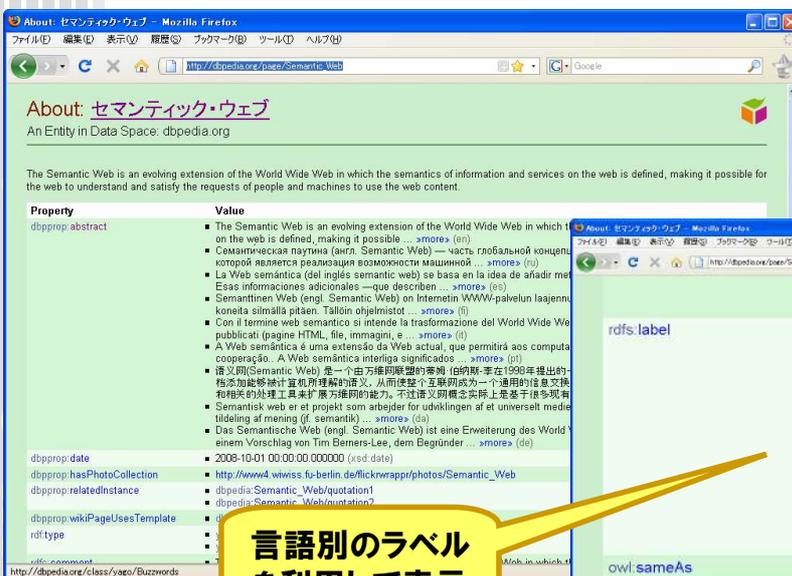
faviki (10): DBpediaの利用



DBpediaでは、項目の識別子に、URIを使う。
 同一識別子の各国語の表示ラベルを持つ

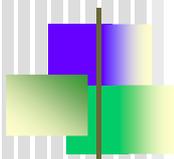
http://dbpedia.org/resource/Semantic_Web

識別子としてのURIは
言語によらない



言語別のラベル
を利用して表示

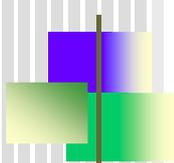




faviki (11): 所感



- 他のソーシャルブックマークと同じような使い方であるが、定まったタグが付けられるのが良い。
- 一方で、独自のタグを付けたい場合、wikiに登録がないできないのは使いづらい。
- 日本語サイトはうまく利用できないことがある。
 - SQLエラーとなってしまったケースあり。



juice (1): 概要



- Webページ上で選択した領域に関する関連情報をサイドバーに表示する
- 画像、動画をサイドバーに貯め込むことができる (ブックマーク)
- ドラッグによる簡単な操作
- Firefox上、アドオンのインストールが必要
- URL <http://juiceapp.com>
<http://linkool.biz>
- 北京 Linkool International Inc.

juice (2): 初期画面



XML Consortium

インストールが必要

ナビゲーションアイコン

チュートリアルの動画がある(英語)

サイドバー LiveMargin

juice (3): 簡単な操作



**選択してドラッグすると右側に
関連情報が表示される**

XML Consortium

Wikipedia

動画 YouTube

ビデオ購入 Amazon

ニュース Google News

**クリックすると左のペインにその
情報が表示される**



juice (4): 知的発見エンジン



XML Consortium

- インテリジェント・ディスカバリ・エンジン
- Linkoolラボ固有のエンジンを利用
 - キーワード, フレーズを対象
 - 自然言語解析により, 文脈に合わせた情報を提示
- 選択内容により, コンテンツやサービスを切り替える
 - Wikipedia, CrunchBase (企業, 人物), 百度, Amazon, YouTube, GoogleNews, ...
- 対応できない場合には, 既存の検索を利用する
 - Google, Yahoo, Live Search



juice (5): マジック



選択して内容に応じて
表示内容が変わる

Hepburn
を選択

Google
を選択

刘若英
を選択

Wikipedia

YouTube

Amazon

Google

CrunchBase

Wikivest

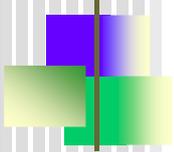
Yobo

Youku(优酷)

Amazon

Baido(百度)

インテリジェント
ディスカバリ
エンジン



juice (6): 所感



- 関連する情報をさり気なく見るには良さそう。
- 英語圏では使えるが、日本語でのサイトを表示している場合には、ほとんどが単純な検索となる。
- 日本人には使いにくそう。
 - 日本人の名前は中国語と見なされ、中国語で表示されるローカライズ機能が無く使いにくい。
 - 自動的に情報が選択されてしまうので、間違った場合など直すすべが無い。
- 日本人にとってはマジック機能が使えないのは残念である。

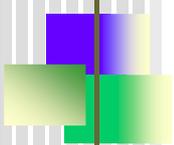


iMage (1): 概要

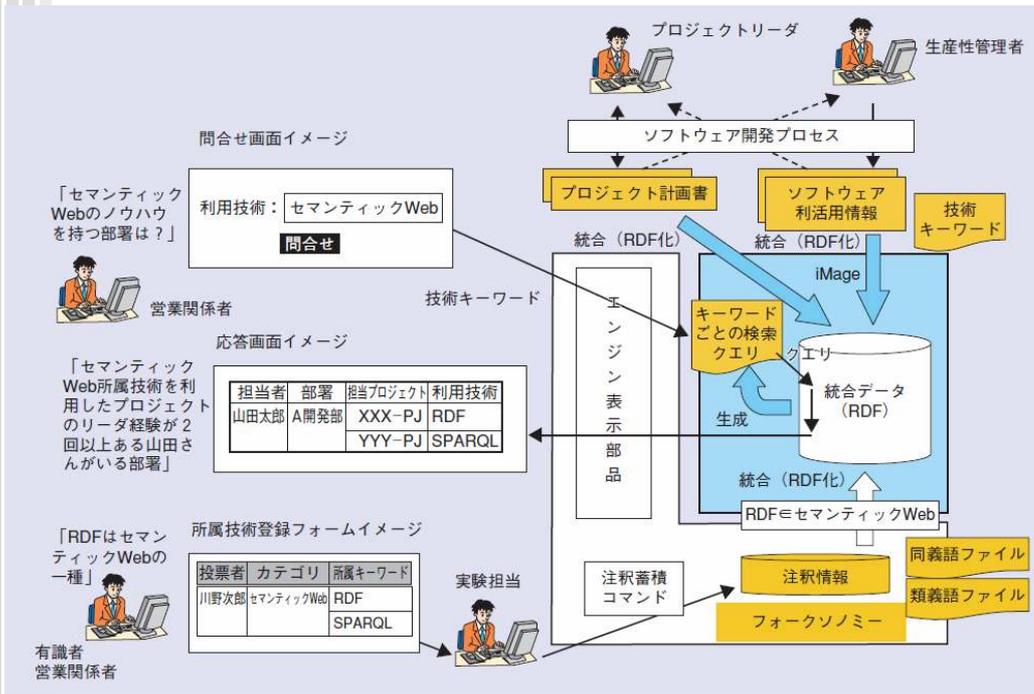


- information Mixable graph explorer: アイメージ
 - NTT情報流通プラットフォーム研究所, NTTソフトウェア:共同研究
 - メタデータに注目し, 新たな知識・情報の抽出・発見を行う次世代Web流通エンジン
 - 次世代Web情報流通エンジンを用いた人材マッチングシステム (NTT技術ジャーナル 2008.4)
<http://www.ntt.co.jp/journal/0804/files/jn200804062.pdf>
- 研究内容
 - メタ情報を大きなグラフとして表現し, そのグラフの「構造」に着目する。特徴的に現れるパターンが人間にとって意味のある有用な知識・情報を表す
 - 研究者や, 研究者が利用するSNSのデータを対象として実験
 - 「ある技術について詳しい人」
 - 「ある技術についてSNSにおいて積極的に交流している人」
 - 等の情報を得る
 - 人材マッチングへの適用について報告

iMage(4): 人材マッチングシステムの成



XML Consortium

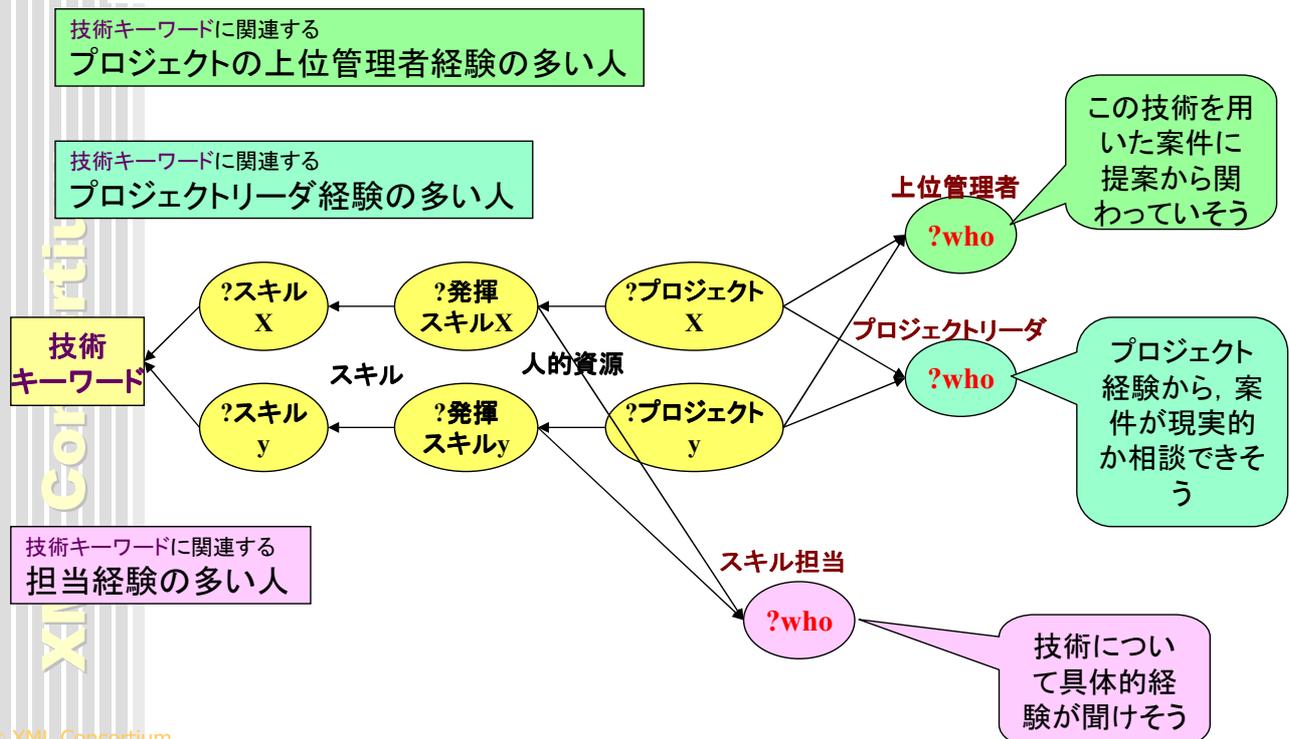


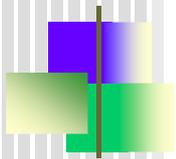
<http://www.ntt.co.jp/journal/0804/files/jn200804062.pdf> から

iMage(5): 人材マッチングシステムで利用したパターン



XML Consortium

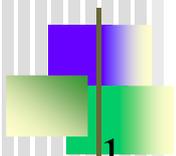




iMage(6): 所感



- メタデータをRDFグラフとして表現し, RDFパターンを抽出する技術をビジネスシーンにうまく結び付けている
- 実験, 評価では効果があると報告されており, 実用化まで進められることを期待したい
- エンジンを適用する具体的なアプリの開発が容易にできて, 現場での固有な利用シーンにも対応できると良い



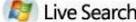
目次(現在位置)



1. 本講演の狙い、背景

- Enterprise2.0 ,特にソーシャル関連市場予測
- なぜ今“セマンティック”が必要か?

2. セマンティック・サービスの事例

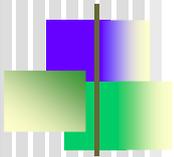
- Twine, Inform, Evri :   
- Zemanta, Powerset, hokia   
- Faviki, Juice, iMage   
- TripIt! UpTake  

3. セマンティック技術の実際

~5W1H Mextractr とOpenCalais
マッシュアップによる応用



印刷・出版オンライン化で付加価値化を達成するヒント



- TripIt! UpTake



XML Consortium

- 旅行業＝情報加工業：

5W1H (いつ、誰が誰と、どこで(どこからどこへ)、いくらで (how much)、何時間(how long))のアレンジが本業

- bulkで仕入れて小分けしパッケージング
- 身体が1つしかない人間が効率良く無理なく旅行を満喫できるよう一筆書きの制約条件を解く

- TripIt! は制約を自動解決

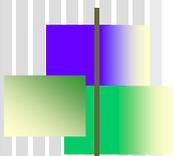
- UpTakeはWhat, Whereを入



Cnetブログ「セマンティック・ウェイブ」より:

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

93



TripIt!



XML Consortium

- オンライン予約の旅程表を1つにまとめるサービス

- 360種(2009.1現在)の予約確認メールを下記に転送

plans@tripit.com → 自動解析して旅程作成

- ソーシャルメディアで共有も可

- Eメールで仲間へ送信

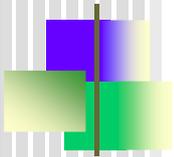
- 印刷、携帯で確認



3



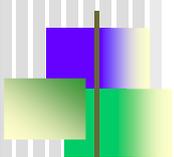
94



TripIT! のセマンティック技術

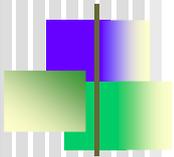
- Eメールのレイアウトや、意味的に特徴的なフレーズを解析
→5W1H: 何、いつ、どこ、誰の予約か抽出

Gregg Brockway社長:「TripIt は Eメールを構文解析し、文脈から情報を抽出してユーザーとその旅程にデータを結びつけ、関連コンテンツを追加」



TripIT! のビジネスモデル

- 有料サービス契約、広告、および共同ブランドで行なう旅程表作成サービス
- Trip Searchを通じた照会料 (サイトの検索サービスで予約が成立したら手数料を得る)
- 利用者から直接対価得る有料サービス契約
 - 2009年～個人/団体/企業向け有料サービス
 - 旅程に関する通知を行なうプッシュ型サービス
 - 企業が自社の出張業務方針と整合性を確保する支援サービス
- 広告モデルを広告主 (Courtyard Marriott～) にアピール
 - LinkedIn 上で自分の旅行情報を共有 (SNS上のソーシャルメディア アプリケーション)

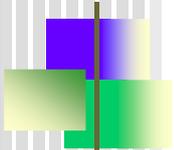


TripIt! APIを発表

2009.1.26

- “Add to TripIt”
 - Emailを介さず旅行社側でTripIT!旅程の構造データに追加できるAPI
- Expense Applications
 - 支出額の集計サービス
 - GDSなど複数種の複雑な裏側の計算のインタフェースをTripIt! APIでラッピング。One stop化
- Ground Transportation
 - 現地レンタカーやリムジン会社が旅程を参照。変更にも追隨して迎車時刻を自ら設定(pull)。

97



- TripIt! 所感

- 難しいところを回避して人海戦術でできるタスクに限定して、アプリ全体の完成をうまいこと前倒しにできたか。
 - 360種の一定の書式の予約確認メールのテンプレートを全部用意することで、そこから、5W1Hのパラメータを正確に(おそらく精度98%はいくでしょう)抽出。そこから計算し、割り出せる、旅程全体の最適化や制約条件の自動適用、管理あたりで頑張ったようだ。
 - 特定のEmailレイアウトに限定したことがビジネスのネックにならないか、確かめて欲しい。
 - その上で、汎用の5W1H抽出エンジンとUIでどこまで上流を代替できるか
- 中流・下流工程の要求精度(旅程の最適化や変更の精度)を、市場での実験でぜひ確かめて欲しい。
- 旅程作り・修正(とても大変!)の自動化率向上はどこまでいけるか。どの程度が近未来のユーザに支持されるか。
- ソーシャルな仕組みで旅程作りをするサービス(arukikataサイトの発展型など)との使い分け、融合は?

98

UpTake はWhat,Whereを入力したらお奨め訪問施設等を提案してくれるサービス



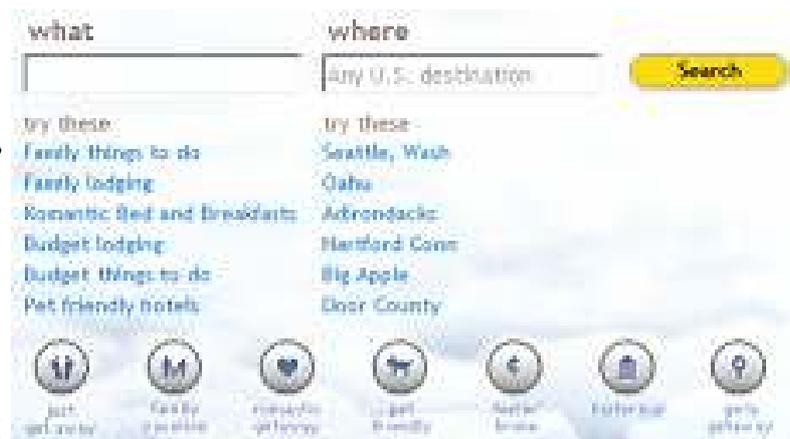
99

© XML

UpTake 概要



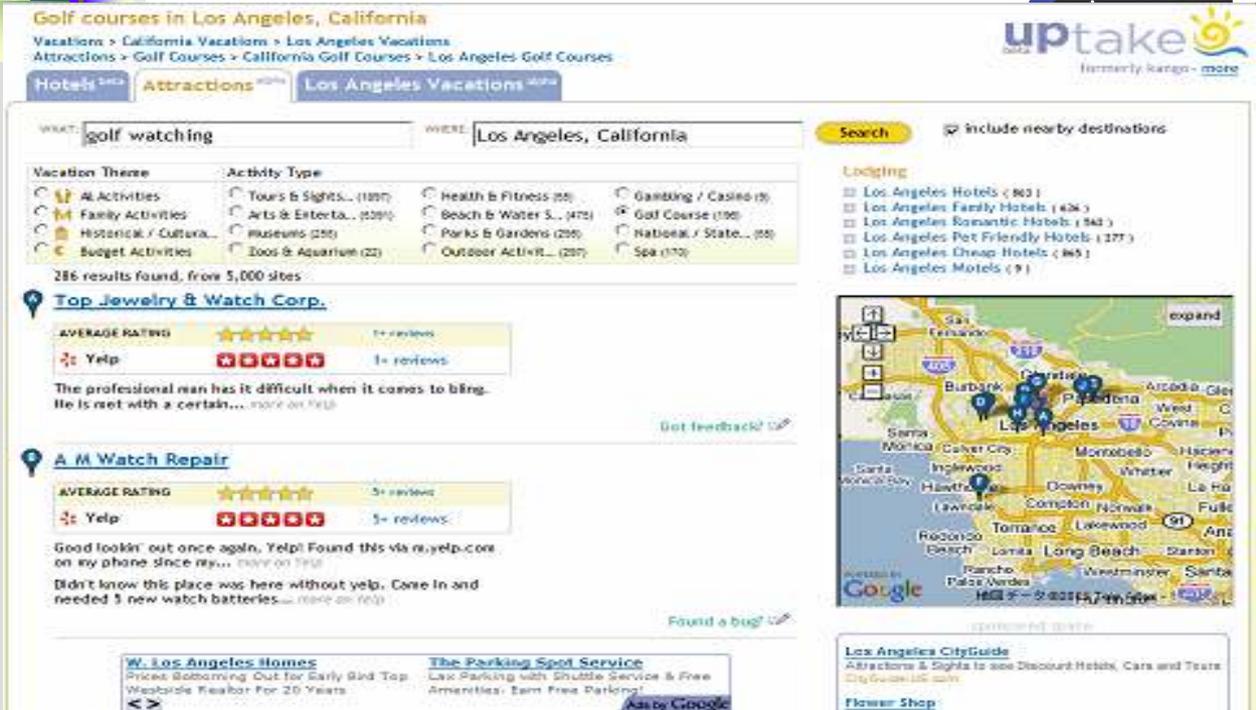
旅行プランのWhatとWhereを指定するフォーム



- “What”はWhyに近い：
 - 旅行目的、旅行コンセプト
 - 個々の旅程(訪問施設等)はシステムが検索、推薦
- “Where”は行きたい方面を大雑把に指定させるコンセプト
 - 旅程の中の個々のパーツ(飛行機とか)は、細かい位置の指定をFrom, Toで書いたり、移動時間とかの制約を入れる

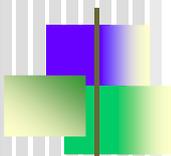
100

© XML Consortium



- ・golf watching と入れたら腕時計ショップの案内が上位に
 - ・ゴルフ場の案内は出現; ツアー観戦は出ず
- 単語間の意味の関連性は現状とらえていないか¹⁰¹

© XML Consortium



UpTake 所感



XML Consortium

- ・よくある素材検索の集約(マッシュアップ)サービスに、5W1Hの一部の構造を導入して利便性を高めようとした、という感じ。
- ・5W1Hに凝縮された事実情報を活用して、マッシュアップやリンク自動生成による付加価値増大は今後も発展の余地が大きい。

目次(現在位置)



1. 本講演の狙い、背景
 - Enterprise2.0 ,特にソーシャル関連市場予測
 - なぜ今“セマンティック”が必要か?

2. セマンティック・サービスの事例

- Twine, Inform, Evri :   
- Zemanta, Powerset, hokia   
- Faviki, Juice, iMage   
- TripIt! UpTake  

3. セマンティック技術の実際

~5W1H Mextractr とOpenCalais
マッシュアップによる応用



印刷・出版オンライン化で付加価値化を達成するヒント

103

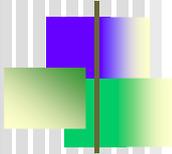
3. セマンティック技術の実際

~5W1H Mextractr とOpenCalais



日本語

英語



Mextractr とは？



- メタデータ株式会社が開発したMextractrは、テキストから5W1Hの記述(いつ、どこ、誰、何、どう、いくら etc)を抽出して活用できるサーバ・ソフトウェア
- 下記は5W1H情報を抽出した例

来る十二月二三日、定時株主総会を東京都文京区小石川の弊社にて開催いたします。是非ご出席いただけますようお願い申し上げます。

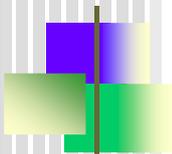
メタデータ株式会社
野村直之



	12月23日	いつ	<input type="checkbox"/>	
	定時株主総会	何	<input type="checkbox"/>	
	東京都文京区小石川	どこ	<input type="checkbox"/>	
	メタデータ株式会社	どの会社	<input type="checkbox"/>	
	野村直之	誰	<input type="checkbox"/>	

Metadata
Mextractr

105



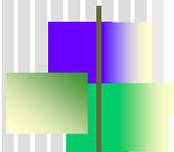
Metadata
Mextractr

Mextractr の動作



- MextractrはWebサーバとしてネット上で動作
- 様々なクライアントからのリクエストに応じて、入力テキストを加工して返す
- “フィード”という、Web上の通知、配信情報の形にして、日付や場所の数値や、予め登録した情報(例:上場会社の企業コード)を付加して配信可能
- 逆に、個人情報、機密情報と認識して、その部分を隠蔽可能。
- Mextractrで抽出した5W1H情報は、元文書のメタデータ(書誌情報)として活用したり、Web2.0の“マッシュアップ”という仕組みにより、地図や組織図、カレンダー上に自動的に配置、登録可能。
- データベースや社内ネット上で個人情報を発見したり、顧客情報を“名寄せ”したり、Eメール配信や掲示板登録に際して安全に再利用できるように加工できる。

106



Metadata
Mextractr

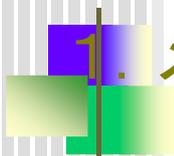
さまざまな用途



XML Consortium

1. メールからカレンダー自動登録、Webからクリッピング
2. マッシュアップ ”Mextractr for Mashup”
3. 個人情報保護
4. 5W1H Mextractr for IE8で、検索エンジン使わずにWebサービス間連携

107



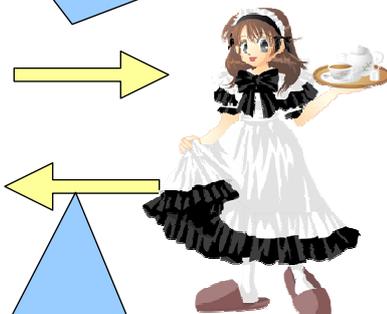
1. メールからカレンダー自動登録 “メイドめーる”



「秘書メール」と改名して、ケータイ・メールで予定の登録・確認・コラボができる次世代グループウェアに発展。



28日の予定。〇時からXX



Metadata
Mextractr

追加の予定を登録！



本日の予定は、
〇時からXXとなっております。
東京地方の天気予報は雨なので
傘をお忘れ無く。

さらに、

- ・メイドさんに返事が書ける！
- ・予定を登録してもらえる！

XML

108

2. 様々なマッシュアップの応用用途

XML Consortium



3. Mextractr 個人情報フィルタ

【営業日報】

メタデータ(株)の次期製品開発部署(東京都墨田区本所)については、窓口の野村主任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てくれば、発注の可能性高い。

平成二十年八月六日 訪問 鈴木太郎

```

<gd:who valueString="野村主任"/>
<gd:who valueString="松田課長"/>
<gd:who valueString="鈴木太郎"/>
<gd:where valueString="東京都墨田区本所"/>

```

営業日報から個人情報を「引き算」して社内掲示板で情報共有！

【営業日報社内公開版】

メタデータ(株)の次期製品開発部署(【...住所...】)については、窓口の【...個人名...】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【...個人名...】が出てくれば、発注の可能性高い。
平成二十年八月六日 訪問 【...個人名...】

```

<gd:when startTime="2008-08-06" valueString="平成二十年八月六日"/>

```



XML Consortium

Metadata Mextractr

4. 5W1H Mextractr for IE8で、検索エンジン使わずにWebサービス間連携



WHEN, WHERE, WHO, ORG, WHAT (event) と、未知語(新語など)の各々ごとに面白い結果が出るサービス数種の出力画面を直接開くことができます。

2009年1月26日に公開されたIE8 RC1 上に5W1H Mextractrを導入して、Webサービス間直結の新感覚を味わってみてください。検索ウィンドウを開いてコピーを繰り返すことがいかにか面倒であったか実感されることと思います。

XML

使い方など:

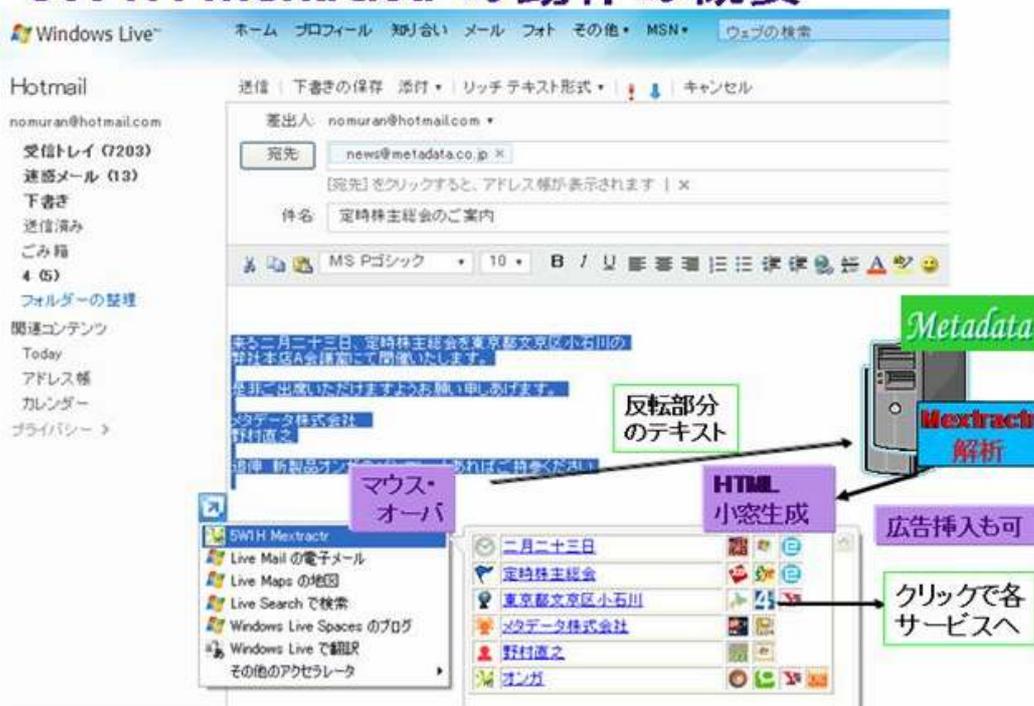
<http://www.mextractr.net/summary/5w1h-mextractr-for-ie8>

111

An Automatic Semantic tagger

5W1H Mextractr の動作の概要

XML Consortium



112

「5W1H Mextractr」の登録サービス一覧 (2009年2月3日版)



「5W1H Mextractr」の登録サービス一覧(2009年2月5日版)

- WHEN** 「今日は何の日」 Live Search EventCast (互換表示推奨)
- WHAT** ZDnet IT イベント検索 Event Eye EventCast
- WHERE** Live Map For Travel Yahoo ピンポイント天気
- WHO** SpySee Live 画像サーチ&関連人物
- ORG** ZDnet 企業紹介 104.com
- 未知語(新語) 瞬!コレ テクノラティ検索 Yahoo 知恵袋 OK Wave

「5W1H Mextractr」の特徴

- 手作業による単語の切り出し、コピー、検索窓の開閉が不要
- 事実情報で大事な5W1Hと未知語(新語など)を楽々自動抽出
- アクセラレータ対応のフォームにコピーすれば、メーラやワープロ、PDF等、Web以外のテキストからも5W1H抽出、および登録サービスへのダイレクト・アクセスが可能

113

Mextractr利用システム図: Mextractr for Mashup のエディタ

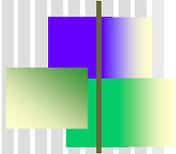
Mextractrエディタでコンテンツを加工し 本番サービスに配備・供給 (deploy)



Metadata Mextractr



114

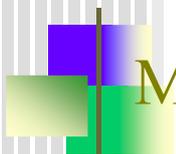


Mextractrの仕組み

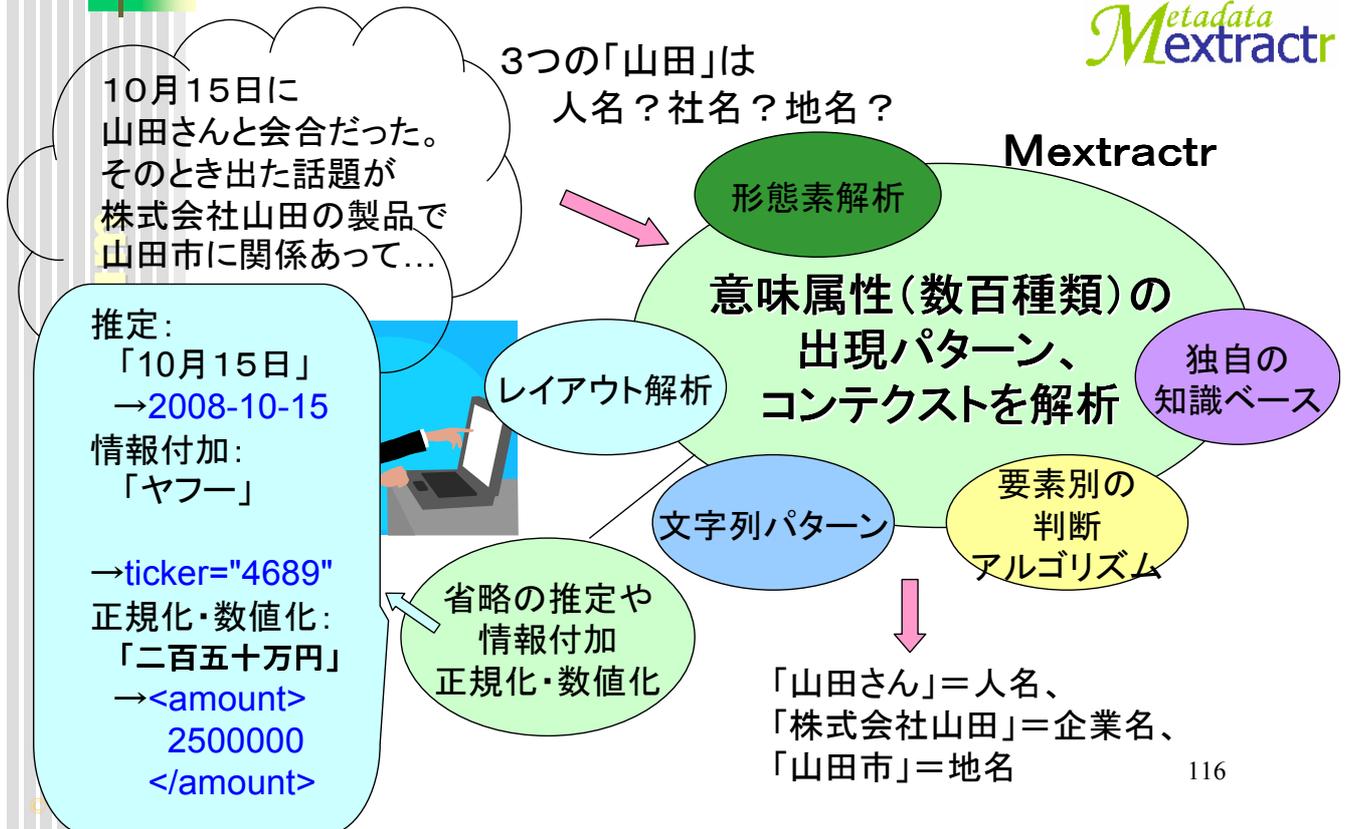


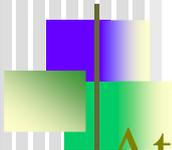
XML Consortium

- 入力:
 - ↓ Email流の改行パターン
 - 流し込みテキスト／毎行改行コード入りプレーンテキスト
- 出力:
 - AtomフィードにGData Event Type + 追加語彙
- 原理: 文字列パタン&単語列(意味素性列)パタンを併用した固有表現抽出と、曖昧性解消、正規化・数値化、情報付加
- 処理概要: 10数種のハイブリッドエンジン + 整合性確保&高速化ノウハウ



Mextractrのセマンティック技術





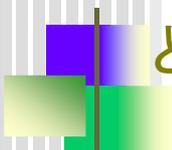
Mextractr出力: Atomベース、GData拡張+UBL2.0+α



XML Consortium

```
<entry xmlns:gd="http://schemas.google.com/g/2005">
  <category scheme="http://schemas.google.com/g/2005#kind"
  term="http://schemas.google.com/g/2005#event"/>
  <id>20080819_0001</id>
  <title>extracted event metadata</title>
  <author>
    <name>Mextractr WebAPI</name>
  </author>
  <updated>2008-08-19T07:46:37+09:00</updated>
  <content>来る五月二三日、定時株主総会を東京都墨田区本所の弊社本店A会議室にて開催いたします。是非ご出席いただけますようお願い申し上げます。メタデータ株式会社</content>
  <summary>来る5月23日、定時株主総会を【...住所...】の【...住所...】にて開催いたします。是非ご出席いただけますようお願い申し上げます。メタデータ株式会社</summary>
  <gd:who valueString="*メタデータ株式会社"/>
  <gd:when startTime="2008-05-23" valueString="五月二三日"/>
  <gd:where valueString="東京都墨田区本所"/>
  <gd:where valueString="弊社本店A会議室"/>
  <gd:extendedProperty name='what' value="定時株主総会"/>
</entry>
</feed>
```

117



どうしても必要な差別化、独自機能のために独自語彙(や拡張構造)を追加



- 例: Mextractr出力: Atomベース、GData拡張+UBL+α

```
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom"
  xmlns:gd="http://schemas.google.com/g/2005"
  xmlns:georss="http://www.georss.org/georss"
  xmlns:cac="urn:oasis:names:specification:ubl:schema:xsd:CommonAggregateComponents-2"
  xmlns:cbc="urn:oasis:names:specification:ubl:schema:xsd:CommonBasicComponents-2"
  xmlns:ext="urn:oasis:names:specification:ubl:schema:xsd:CommonExtensionComponents-2">
  <title>Mextractr WebAPI results</title>
  <updated>2008-08-19T07:46:37+09:00</updated>
  <author>
    <name>Mextractr</name>
  </author>
  <id>20080819_0001</id>
```

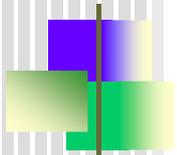
Atom

GData

georss

UBL2.0

118



日本語文章から年齢、年月日、金額などを機械が「理解」できるように数値化する世界初の機能群



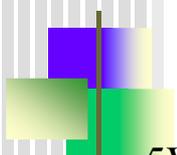
Who: 野村直之(46) → `<gd:extendedProperty name='age' value="46"/>`
 Who: リアルコム → `<gd:extendedProperty name='ticker' value="3856"/>`
 When: 一九九八年 → `<gd:when startTime="1998-01-01" valueString="一九九八年"/>`

Where: 東京都文京区小石川2-1-2 →
`<georss:point>139.751456 35.708823`
`</georss:point>`

How much: 二億三千万 →
`<gd:extendedProperty name='how-much' value="二億三千万円">`
`<cbc:Amount currencyID='JPY'>230000000</cbc:Amount>`
`</gd:extendedProperty>`

今後、How longなどに拡大

Next Step: Who, What の下位属性としての 5W1H
 How large/tall/heavy



メタデータ検索



5W1Hの組み合わせ、範囲指定による検索

－ 例: 目の前にある「年次報告書」の、前年のもの、***年のもの、別部門(WHO)が作った同年(この年次のみ)の報告書を検索する、など、メタデータを切り替えて、全文検索が得意な文書探索を行う。

- 5W1Hのメタデータをプルダウンやフォームにより、あるいは自動補完されたデフォルトにより指定
- 検索対象は、メールDB (Mail Box)や文書DB
- いわゆる、検索キーワードを全く指定しなくてもおぼろげな記憶から検索可能
 - － 「ほら、例の隣の事業本部の技術系の偉いさんが一昨年あたりに書いた、どこかヨーロッパの方の国から帰ってきて興奮して書いてた技術ビジョンの提言書みたいなやつだよ。すぐ出しておいてね。」 → 何のテーマかわからなくとも社内サーチならほぼ一意に確定。
- 全文検索ではできない→『3時間で専門家になる私の方法』by 佐々木俊尚
- エンドユーザが付加したメタデータも検索条件に加えた例 → Flickr



Open Calais



XML Consortium

1. 使用料 / 利用条件・制限

1.1 無料で使える範囲 (<http://www.opencalais.com/CommercialCalais> より)

40,000件/日 ; 1秒あたり4件まで

1.2 有料オプション(2種)

Option#1: Calais Professional (SaaS)

100,000件/日 ; 1秒あたり20件まで

(追加で、2,000,000件/日まで増やすことが可能)

Option#2: ClearForest solution (オン・プレミス)

直接サーバにサービス自体をインストール

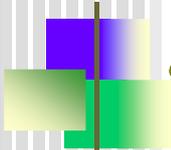
1.3 利用条件(<http://www.opencalais.com/terms> より)

1. Your Calais service.

calaisを使用したサイトやアプリケーションは、calaisのロゴを(サイズ変更せずに)表示し、そのロゴに、「<http://opencalais.com>」へのリンクを付けること。

5. Limitations on Use.

Calaisにより作られたメタデータを使用する場合、Globally Unique Identifier (GUID)もできるだけ一緒に付けること。
121



どんな5W1Hメタデータを抽出するか



XML Consortium

2.1 タグ

「Entities」と「Events and Facts」に大別(詳細は <http://www.opencalais.com/APIresponses>)

Entities:

Anniversary, City, Company, Continent, Country, Currency, EmailAddress, EntertainmentAwardEvent, Facility, FaxNumber, Holiday, IndustryTerm, MarketIndex, MedicalCondition, MedicalTreatment, Movie, MusicAlbum, MusicGroup, NaturalDisaster, NaturalFeature, OperatingSystem, Organization, Person, PhoneNumber, Product, ProgrammingLanguage, ProvinceOrState, PublishedMedium, RadioProgram, RadioStation, Region, SportsEvent, SportsGame, SportsLeague, Technology, TVShow, TVStation, URL

Events and Facts:

Acquisition, Alliance, AnalystEarningsEstimate, AnalystRecommendation, Bankruptcy, BonusShares, BusinessRelation, Buybacks, CompanyAffiliates, CompanyCustomer, CompanyEarningsAnnouncement, CompanyEarningsGuidance, CompanyInvestment, CompanyLegalIssues, CompanyLocation, CompanyMeeting, CompanyReorganization, CompanyTechnology, CompanyTicker, ConferenceCall, CreditRating, EmploymentRelation, FamilyRelation, FDAPhase, IPO, JointVenture, ManagementChange, Merger, MovieRelease, MusicAlbumRelease, PatentFiling, PatentIssuance, PersonAttributes, PersonCommunication, PersonEducation, PersonEmailAddress, PersonPolitical, PersonPoliticalPast, PersonProfessional, PersonProfessionalPast, PersonRelation, PersonTravel, Quotation, SecondaryIssuance, StockSplit

一部には属性として、「attribute」が付く:

Product >> Product, ProductType

PersonProfessional >> Person, Position, Company, Organization, Notes





サンプル出力



<http://www.opencalais.com/calaisAPI>

の下方:

「wikinews_08Nov5.zip」

http://www.opencalais.com/files/wikinews_08Nov5.zip

Wikipediaのニュースがソース

→カバレッジ(再現率:recall)がいまひとつか?

XML Consortium

• 対応言語

- 過去1年以上サービスされたのは英語 <http://opencalais.com/faq#faq6>
- 言語判定には長さが必要 c:language="InputTextTooShort"
- "It will be French in R4, to be released in Jan 09." <http://opencalais.com/node/2701>
近日中にフランス語対応

• ユーザー辞書 →今のところ無し

• 5W1H抽出結果の正規化・解釈 →ある程度行える模様

例: `<date>April 13, 2007</date>`

`<c:datestring>yesterday</c:datestring>`

から、

`<c:date>2007-04-12</c:date>`

123

The screenshot shows the Calais Web Service website. At the top, there is a navigation bar with links for Home, About, Gallery, Tools, Forums, Blog, News Room, Register, and a login section. The main content area is titled "API Overview" and contains the following sections:

- The Calais Web Service: Overview**: A paragraph explaining that the service allows automatic annotation of content with semantic metadata.
- Preprocessed Content**: A paragraph explaining that the service provides a preprocessed sample content set using Calais.
- WikiNews - Business and Economic**: A paragraph explaining that the content set contains Business & Economic news from WikiNews for the year 2007.

At the bottom of the main content area, there is a table with two columns: "Attachment" and "Size".

Attachment	Size
wikinews_08Nov5.zip	1.64 MB

On the right side of the page, there is a sidebar titled "Calais Web Service" with a search box and a list of links for "Calais API Overview", "Preprocessed Content", "API Calls", "API Responses", and "Entities".

© XML Consortium



その他特徴

接辞的語句の付加による抽出アシスト

～入力文中にヒントを与えることができる



- 1) 動詞に、“verbタグ” and/or “statusタグ” を付けて、「generic relation」やイベントの状態 (e.g. “IPO of ABC corp. will be planned on January 20.” →status: plannedと出力) を出力
- 2) タグが付かなかった単語の直前にカテゴリー名を付けてやることによってタグ出力を操作できるものがある

例1) The report, which I mentioned in my Microsoft analysis Thursday, is notable because it maps out the cloud landscape and puts Amazon in its place.

→ 何も抽出されない

["Amazon"の直前に"the company"を付加] The report, which I mentioned in my Microsoft analysis Thursday, is notable because it maps out the cloud landscape and puts the company Amazon in its place.

→ Company : Amazon が抽出される

例2) In explaining Twitter to people (something I seem to be a doing a lot lately), one of the questions I get the most is, "So who should I follow?" I've long wished for a tool, myself, that would let me bulk-add cool people to the network. Over the holidays I checked out two new tools designed to do that.

→ 何も抽出されない

["network"の直前に"the industrial term"、["Twitter"の直前に"the company"を付加]

In explaining Twitter to people (something I seem to be a doing a lot lately), one of the questions I get the most is, "So who should I follow?" I've long wished for a tool, myself, that would let me bulk-add cool people to the industrial term network. Over the holidays I checked out two new tools designed to do that.

→ Industry Term : industrial term network、Company : Twitter

125

ただし、[the industry term Twitter]では反応せず。



Mextractr 意味分類体系:

～モノと数量を中心に拡張固有表現分類2004+属性表現を5W1Hに階層分類



何らかの名称	zz	entity
時間表現	zz0	time_top
時点表現	zz01	point_in_time
時刻表現	zz011	time_oclock
日付表現	zz0110	date
祝日	zz0111	holiday
曜日表現	zz012	day_of_week
週表現(第3週、第42週)	zz013	week
月表現	zz014	month
年表現	zz015	year
時代表現	zz016	era
日本の年号表現	zz0161	japanese_era
時間_その他	zz019	time_g_oth
期間	zz02	period_g
時間(「3時間」勉強しました)	zz020	time_g
時刻期間(「3時台に」勉強しました)	zz021	time_period
日数期間	zz022	date_period
週期間	zz023	week_period
月期間	zz024	month_period
年期間	zz025	year_period
期間_その他	zz029	period_g_oth
時間表現_その他	zz09	time_top_oth

← 約370種の意味分類

誰が		zz1
組織において(誰とどこの中間)		zz2
どこで		zz3
何を(name)		zz4
どれだけ(how many)		zz5
出来事(event)名		zz6
「いつ」の属性		zy0
「誰」の属性		zy1
「組織」の属性		zy2
「どこ」の属性		zy3
「何(名前)」の属性		zy4
「どれだけ(数量)」の属性		zy5
出来事の属性		zy6

126

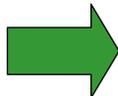
カレンダー情報の入力コスト削減

鈴木さんへ

本日の技術ミーティングは六本木のB社にて五月十日15時から執り行われます。開発部の後藤さんと同席してください。



日時
場所
人名
などを
抽出



Metadata Mextractr

タイトル	技術ミーティング		
日時	2009/05/10	15:00	か
	2009/05/10	16:00	<input type="checkbox"/>
	繰り返し: なし		
場所	六本木		
カレンダー	▼		
説明	鈴木さんへ 本日の技術ミーティングは六本木のB社にて。開発部の後藤さんと同席してください。		

・入力のコスト削減: 100人規模の企業なら年間2000万円分

(メール週平均40通、1通の処理3分間を1分に省略、時給平均3000円として)

・ダブルブッキングチェック(オプション機能)でスケジュールリングミスが減り業務効率UP

127

“Mextクリッパー”

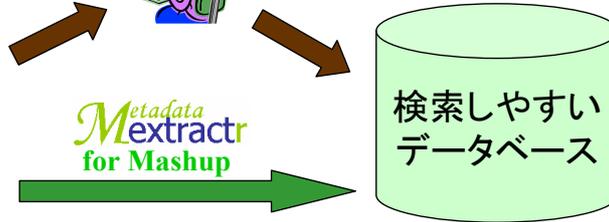
128

既存リソースにキーワード付与して構造化、
マッシュアップ

未整理のテキスト
データ



従来: 人手でキーワード付与



提案: Mextractrで5W1H等の抽出

- ・キーワード付与のコストダウン:
6000万字のテキスト処理のコストは
人手なら約4000時間、時給1000円として400万円。
Mextractrなら約12時間で完了。
- ・抽出の「ぶれ」の削減→100%の手作業は精度で機械に敵わない

業務フローの途中や、社内情報共有
で伏せ字化

山本株式会社 鈴木様との
会議で製品AB-01納入先変
更。
ABC社(東京都新宿区)から
XYG社(埼玉県さいたま市)
に



【企業名1】【個人名1】と
の会議で製品【製品名1】
納入先変更。
【企業名2】【住所1】から
【企業名3】【住所2】に

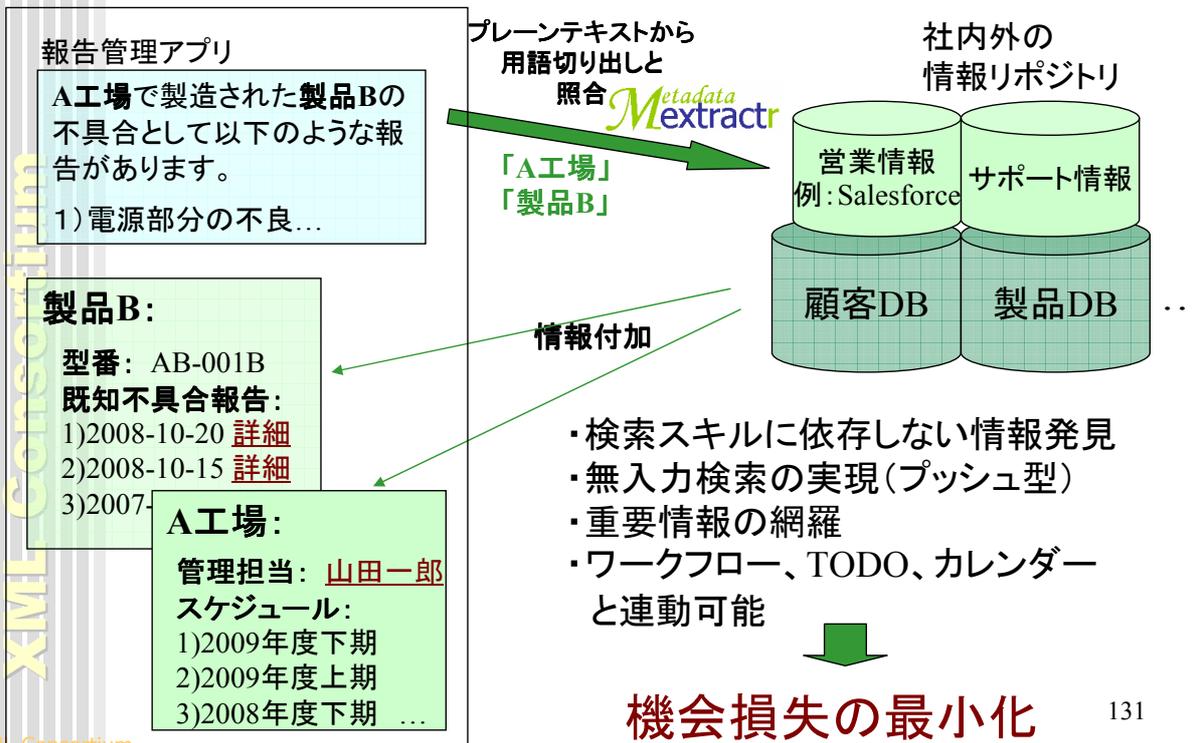
機密情報付の文書



伏せ字化したテキストに

- ・漏洩の危険性を軽減
- ・Mextractrを使うと: 6000万字/12時間 (新聞1年分のテキスト)
6000万字を処理した場合、人手なら400万円 (4000時間)
(1.5万字/1時間、時給1000円として。Mextractr初期費用は50万円)

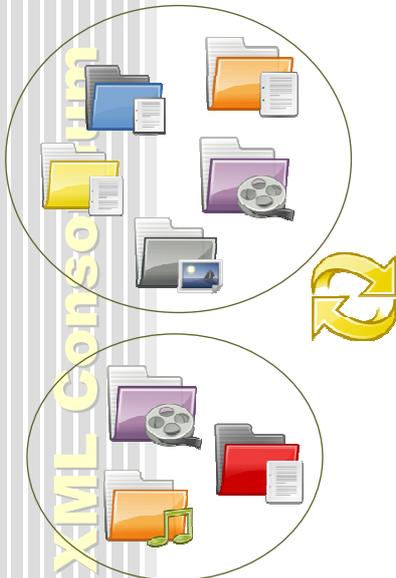
ナレッジマネジメントへの活用： 関連知識との照合・情報付加



まとめ: セマンティック技術の今後の可能性



様々な形式で、インターネット、イントラネット内で、意味付けされず散らばっている膨大なデジタルコンテンツ



情報を共有できる知識に変換
KnowWho、KnowHow?



直観的な条件指定



欲しい情報が欲しい形式で素早く得ることができる



日本はこれから!!