

メタデータ自動抽出ソフトウェア

Mextractr for Mashup

のご紹介

2008.11.28

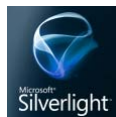
メタデータ株式会社

クライアント技術部長代理 佐藤 正志
サーバ技術部長代理 松田 圭子

Mextractr for Mashup



IdbA™



WebAPIのデータを

- ・マッシュアップ
- ・見やすく出力

→ユーザーに快適な環境をご提供

Mextractr for Mashup



IdbA™



XMLでもない
JSONでもない
→マッシュアップしにくい
でも何か**重要な情報**が入っている



日本語テキスト

Mextractr for Mashup



IdbA™



5W1H(いつ、どこで、誰が、何を等)
の抽出&XML化で
Mashup可能に

→WebAPI化

Metadata
Mextractr



日本語テキスト

Mextractrで何が出来る？

プレーンテキスト

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
野村直之



Metadata
Mextractr



5W1Hの抽出

When	2008-12-23
Where	東京都文京区小石川
Where	弊社本店A会議室
Who	*メタデータ株式会社
Who	野村直之
What	定時株主総会

※抽出対象を隠すの也可能

デモ

Metadata
Mextractr

宮崎県宮崎市で、株式会社宮崎の宮崎さんと意見交換会。

テキスト:
[解析]
[表示]

中間出力ON 中間出力OFF

テキスト解析の実行

[Mextractr意味体系分類名一覧を閲]

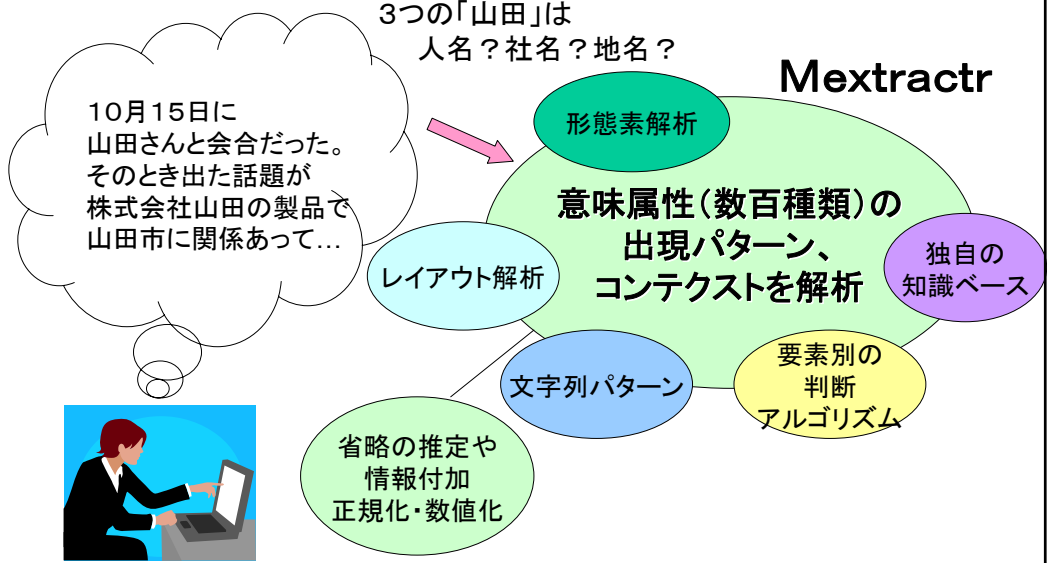
Mextractr出力

```
<?xml version="1.0" encoding="utf-8"?>
<entry xmlns:gd="http://schemas.google.com/2005">
  <gd:where valueString="宮崎県宮崎市" />
  <gd:who valueString="株式会社宮崎" />
  <gd:who valueString="*株式会社宮崎" />
  <gd:who valueString="宮崎さん" />
  <gd:extendedProperty name="what" value="意見交換会" />
</entry>
```

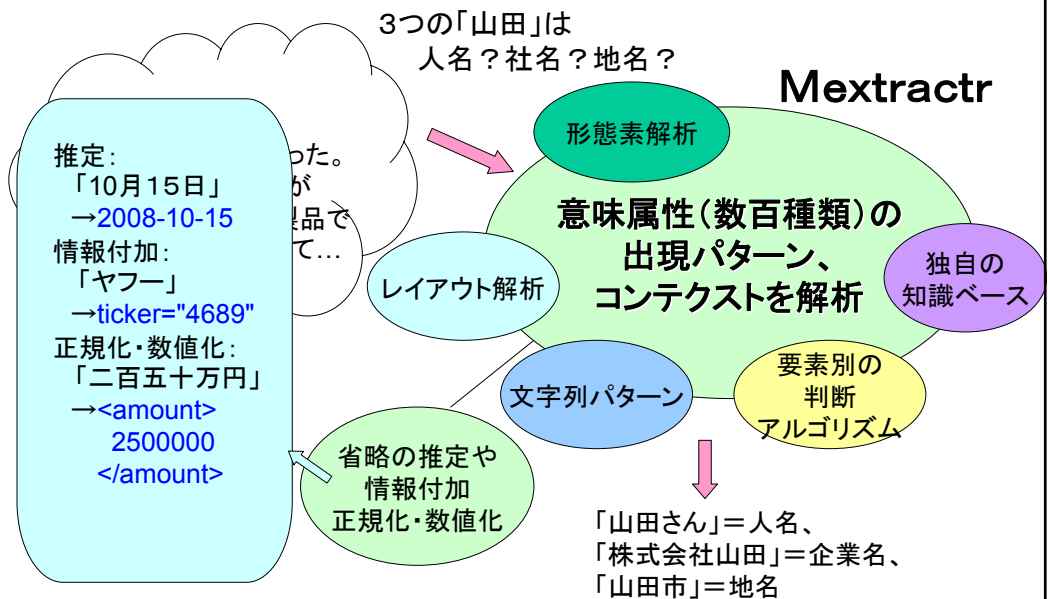
Googleランダーに登録する呼び出しの検証: (上記テキストサンプルの事例) [Googleランダー](#)

Mextractr および Mextractr 個人情報フィルタ(プライバシーフィルタ)に関するご質問、お問い合わせは[問い合わせセンター](#)をご利用下さい。
Mextractr WebAPI、Mextractr 個人情報フィルタ(プライバシーフィルタ)は [メタデータ株式会社](#) が提供しています。

Mextracrのセマンティック技術



Mextracrのセマンティック技術



WebAPI化によるMashupのしやすさ

サン×リクルート Mash up Award 4thに
Mextractr機能をAPI提供。
6作品に利用されました。

MASHUP AWARD
2008.06.03 - 09.16

Mashionary

Golazo MA4

Salestracr

ノリで宿検索!



Antenna for MA4

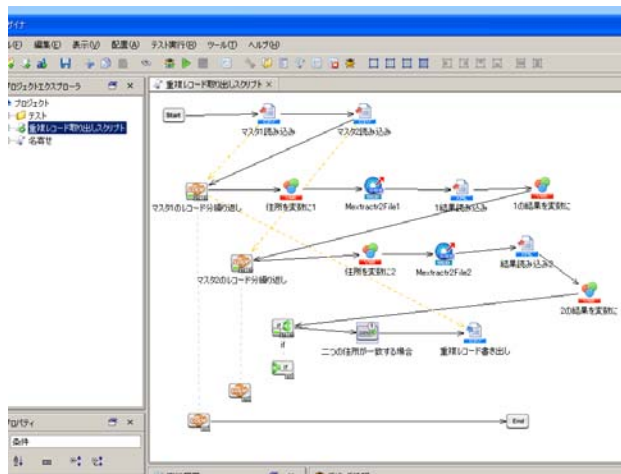
メイドめーる

Award on Railsにてメタデータ賞受賞



DataSpider上での試作

アプレッツ社によるEAI(Enterpriseアプリケーション統合)ツールである
DataSpider上での匿名化、名寄せの試作



他のツールに手軽に組み込むことで効果発揮&開発工数削減

Mextractの二つのタイミング

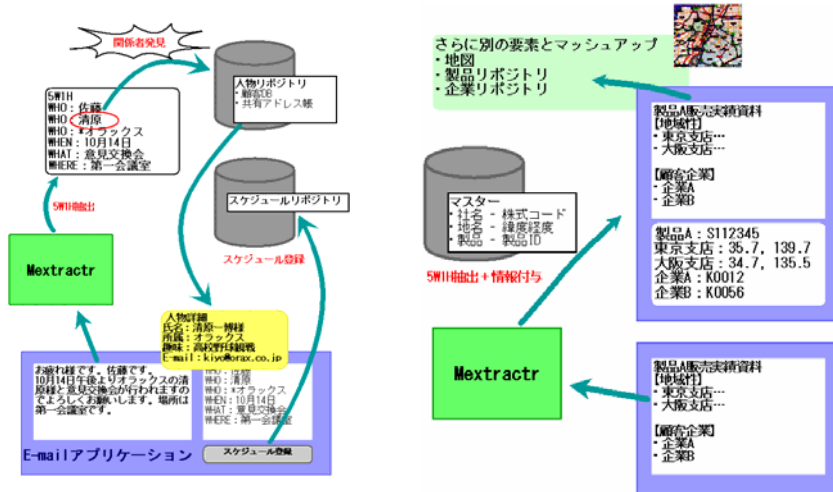
Mextractのタイミングは大きく分けて二つです。

- ・オンデマンド(フロー)
- ・バッチ(ストック)

※Mextractr = Metadata extractor(メタデータ抽出器)
Mextract = メタデータ抽出する

オンデマンド(フロー)

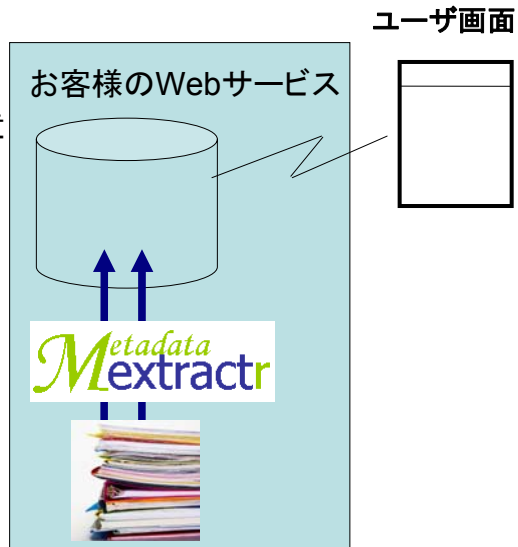
入力文章を即座にMextract
多くはクライアント寄りアプリからコール



バッチ(ストック)

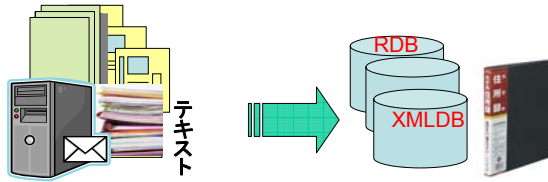
既に存在する大量の文章
を事前にMextract

多くはサーバに組み込み



命題 (顧客の課題)

大量のドキュメントをシステムのリソース (DBレコード、マッシュアップ部品) として扱いたい



手段

大人数を投入して恒常的に人手で加工したいか？
cf 人手の作業は100%正しいか？

No! だから



とはいえ

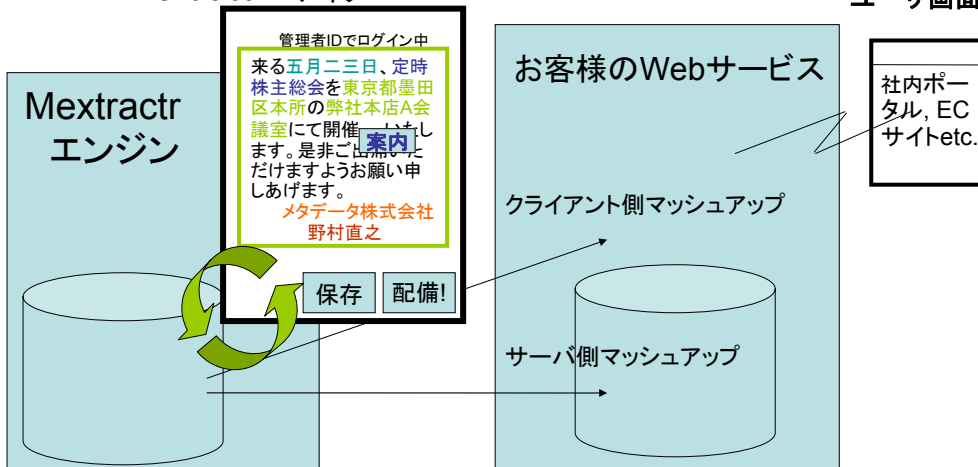
- 任意の (恣意的な) 正解を作りたい
- 文章中に無かったメタデータを付けたい
- 解析漏れ、勇み足への対処



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)

Mextractrエディタ

ユーザ画面



エディタのデモです

The screenshot shows a web browser window titled "http://ap.mextractor.net/editor/edit?id=test2.xml - Windows Internet Explorer". The page content is titled "5W1Hエディタ". It features a text area with introductory text about Mextractor for Mashup, followed by a table of 5W1H data. The table has columns for category (WHAT, WHEN, WHERE, WHERE, WHO), text, count, and a delete button. On the left, there are settings for output and extraction, and a list of location categories with checkboxes.

WHAT	WHAT	WHAT	WHAT	WHAT	WHAT	WHAT	WHAT	WHAT	WHAT
SW1H	文字列	位置							
WHAT	Mextractor for Mashup	125	削除						
WHAT	Mextractorエンジン	352	削除						
WHAT	Mextractorエディタ	404	削除						
WHEN	12月12日	111	削除						
WHERE	東京都文京区	20	削除						
ORG	メタデータ株式会社	7	削除						
WHO	豊科村直之	30	削除						

話は変わりますが

検索ワードを利用しない検索
キーワードレスサーチ

隣の部署の偉い人が去年アメリカで講演した10ページ位の文書はどこ？

- ・バッチ(ストック)で5W1Hのデータを作成
- ・オンデマンド(フロー)のクライアントから一致を調べにゆく

まとめ

- 日本語テキストをMashup可能にしたい場合は是非Mextractrをご検討ください。
- バッチ(ストック)のケースでの出力情報最適化にMextractrエディタをお役立てください。

まとめ

日本語のWebAPI化
ならMextractr!

- 日本語テキストをMashup可能にしたい場合は是非Mextractrをご検討ください。
- バッチ(ストック)のケースでの出力情報最適化にMextractrエディタをお役立てください。