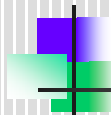




XML活用事例収集



XMLコンソーシアムWeek
関西部会
日本アイ・ピー・エム 伊藤 健太郎
ブレインワークス 松尾 萌



参加メンバー

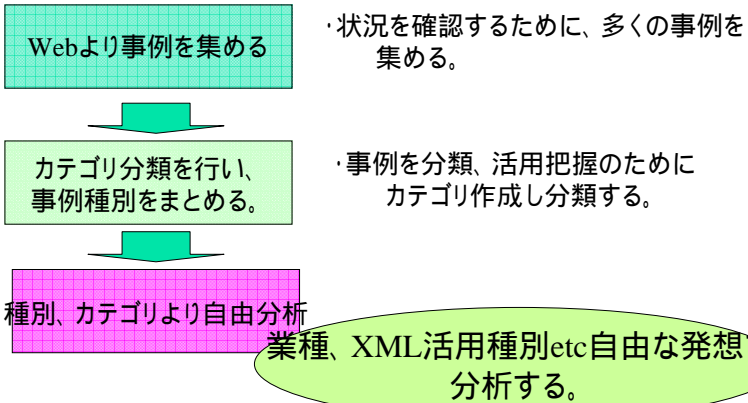


- | | |
|--------------|-------------|
| ■ ブレインワークス | 松尾 萌 |
| ■ ブレインワークス | 芦田 尚人(リーダー) |
| ■ 日本アイ・ピー・エム | 伊藤 健太郎 |
| ■ 日本システムウェア | 名津井 誠 |
| ■ 日本システムウェア | 中澤 奈美 |

どこで使ってますねん?! 概要



XML Consortium



2006年度実績

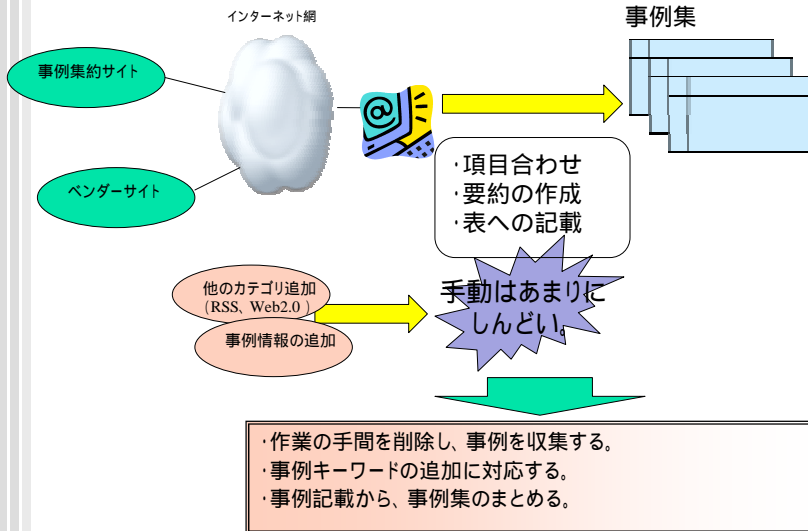
調査件数: 163件

独自分析: 6種類(業種、年代etc)

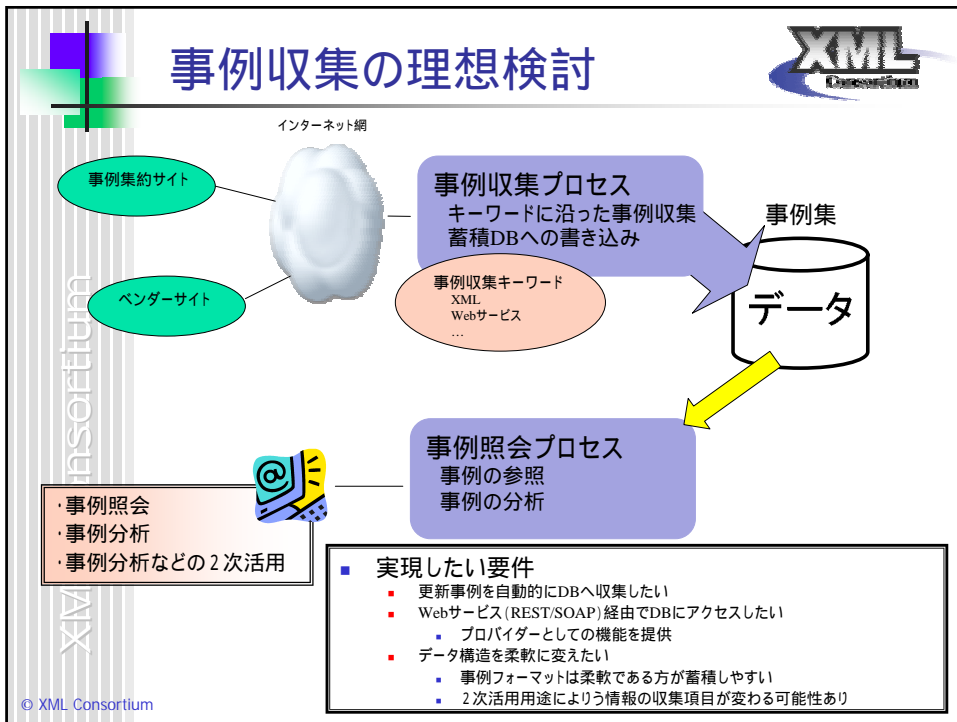
昨年度の調査方法



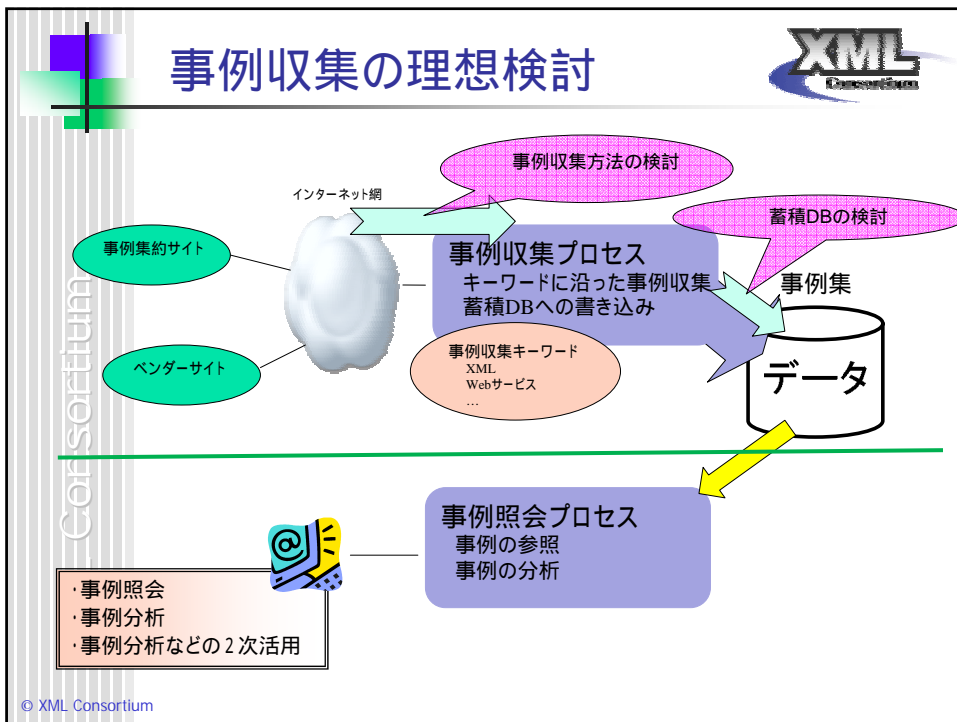
XML Consortium



事例収集の理想検討



事例収集の理想検討





検討項目



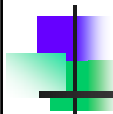
1. 新規カテゴリ事例の調査
RSS、BLOG等新しいカテゴリで調査
2. 管理DBの調査
XMLDB及びRDBで蓄積DBの検討
3. DBへのデータ蓄積
DBへのデータのLoadの仕組み検討
4. データの照会
DBからデータを照会し、自由分析できる仕組みを検討

目的

- 楽しんで集める
 - ロボットに関する情報収集(活用方法)
 - 数を増やす:事例情報
- 蓄積方法
 - 過去のデータを入れてみる
DBを入れてみる、つかってみる。
-----以上2006年度-----
- 次のステップ
 - 分析・活用方法 XMLの進化・変化など...
 - 事例データ、分析の2次活用



事例収集方法について



XMLコンソーシアムWeek

関西部会

株式会社ブレインワークス 松尾 萌

事例収集方法



ロボット型検索

webロボット、クローラーと呼ばれるプログラムがwebページを自動巡回し、情報を収集する仕組み

RSS

最新のニュースやブログの更新情報などwebサイトの更新情報を簡単にまとめ、配信するXMLベースのフォーマット

アンテナサービス

サイトの更新状況を通知するネットサービス
ユーザーが良く見るサイトを登録しておく、それらのサイトを定期的に自動巡回する

事例収集方法



ロボット型検索

【特徴】

- 収集する情報量が多く、情報の新鮮度が高い
- 関連するキーワードの出現頻度が高いwebページを優先する
- 情報が細かく選別されていないため、必要としない情報が含まれている可能性がある

【活用例】

- **ロボット型検索エンジン**
 - Google
 - goo (NTT)
 - MSN (マイクロソフト)
 - Yahoo! Japan
- **求人情報収集「ジョブエンジン」**

事例収集方法



RSS

【特徴】

- 多数のwebサイトの更新情報を効率的に把握できる
- RSSリーダーと呼ばれるソフトにサイトを登録しておく、更新されるたびに最新記事の見出しや概要などが自動的に表示される

【RSS 活用例】

- | | |
|-------------|------------------|
| ■ Web型 | はてなRSS |
| ■ ブラウザ・メール型 | gooRSSリーダー |
| | Firefox |
| ■ ティッカー型 | Headline-Deskbar |

事例収集方法

アンテナサービス



【特徴】

- ユーザーが登録したサイトの更新状況を通知する
- 前回の更新内容との差分が表示される
- 登録ページの状況に応じてお薦めページを教えてくれる
- 他のユーザーとアンテナデータを共有することができる

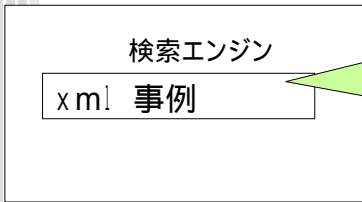
【活用例】

- はてなアンテナ

ロボット型検索 調査結果



【調査方法】



ロボット型検索エンジンにて
Webサービス、xml、事例、
事例導入などの
関連キーワードを用いて
事例を収集

第2、第3検索キーワードの絞込みにより、適切な事例の収集が可能

	google	goo	msn	yahoo
xml 事例 導入	812,000	56,000	93,175	506,000
webサービス 事例 導入	396,000	27,700	477,152	290,000

© XML Consortium

ロボット型検索 調査結果



【情報量】 収集できる情報量が多い

【情報の新鮮度】 情報の新鮮度が高い

【収集できるジャンル】 さまざまなジャンルのサイトから収集が可能

【利便性】 検索したいキーワードを入力するだけで事例を収集できる

【適切性】 xml事例関連書籍を紹介するサイトなど必要のない情報も収集される

【プログラム組み込みの容易性】 組み込みは容易であるが、検索キーワードの検討が必要

© XML Consortium



RSS 調査結果



【調査方法】 goo RSSリーダーを使用
「xml」関連のキーワードを登録し、関連記事を収集

【情報量】 情報量が少ない

【情報の新鮮度】 最新の情報が収集できる

【収集できるジャンル】

RSSに対応したサイトからの情報収集が可能

【利便性】 ユーザー登録が必要となり、手間がかかる

【適切性】 ブログの記事がほとんどであり、
適切な情報が収集できるとは限らない

【プログラム組み込みの容易性】

デスクトップで使用できるソフトもあり、比較的容易



アンテナサービス 調査結果



【調査方法】 はてなアンテナを使用し、情報を収集

- アンテナページからxml関連の用語を検索
- ユーザー登録をし、IT関連のURLを登録

【情報量】 情報量が少ない

【情報の新鮮度】 最新の記事を収集することができる

【収集できるジャンル】

ユーザーが登録したwebサイトのみから収集が可能

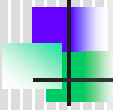
【適切性】

自分でサイトを登録することができるため、適切な
ジャンルのサイトから情報収集することができる
(ただし、xml関連の記事のみ収集できるとは限らない)

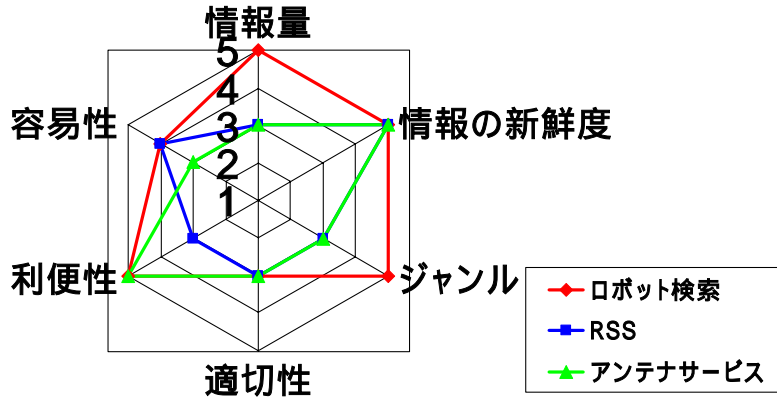
【利便性】 xml関連の用語を検索するだけであれば

ユーザー登録がいらない

【プログラム組み込みの容易性】 登録するサイトの検討が必要



考察



ロボット型検索が事例収集に適しているのではないかと？



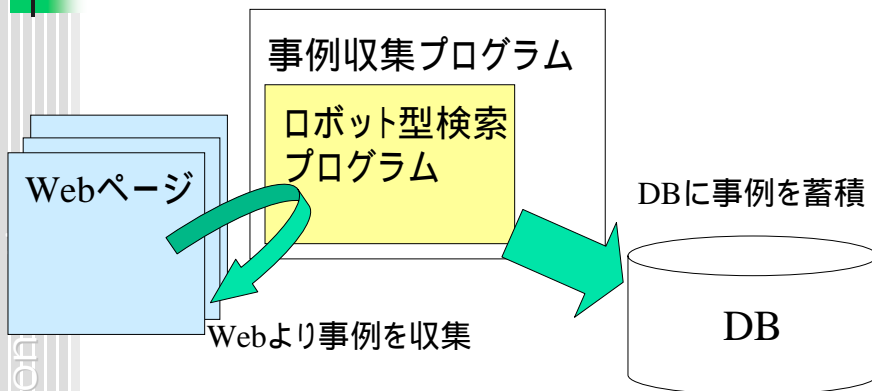
適切な事例収集方法は？



- ロボット型検索を用いた場合のメリット
 - 多くの情報を容易に収集できる
 - 新たな事例キーワードの追加が可能
 - 将来的には、RSS・アンテナサービスの機能を組み込むことが可能



ロボット型検索を用いた事例収集の検討



次年度の課題

- ・ロボット検索プログラムを使用した事例収集プログラムの構築

事例集の蓄積について

XMLコンソーシアムWeek

関西部会

日本アイビーエム 伊藤 健太郎

なぜXML DBが必要？



前提：事例は、非定形データ

XMLDBの調査

- XMLの普及がますます進んできた
 - データ交換、ドキュメント管理で拡大
 - 電子商取引
 - FPML：金融派生商品の取引および情報共有のための規格
 - ebXML：企業間取引のための規格
 - cXML：受発注やカタログ配信などの企業間取引のための規格
 - 電子カルテ、電子決済など
 - 身近で使われているXML
 - ブログの標準技術(RSS,トラックバック)はXML
 - オフィス文書の標準規格ODF(Open Document Format)はXML

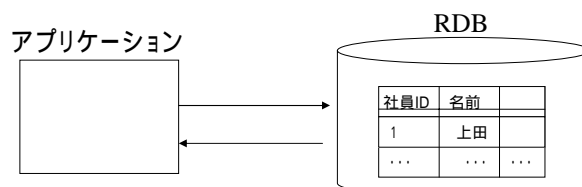
XMLの保管先としてのXML DBが注目されている

なぜXML DBが必要？



- データの変化に対応できるDBが必要
 - ビジネス変化への対応は、ITシステムの柔軟性に依存
 - ITシステムの柔軟性には、データ変更の柔軟性も1つ
 - RDBにおいて、データの変更が求められた場合、
 - テーブル構造の変更、データの再登録が必要
 - アプリケーションの変更も必要

データ変更柔軟に対応できるDBとしてXML DB
事例など、非定型データは、XML DBが適している

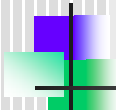
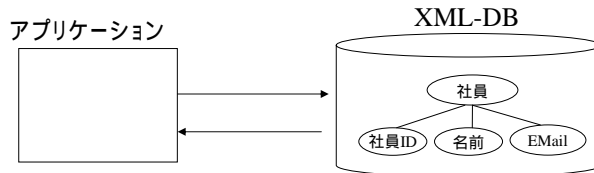




XML DBとは



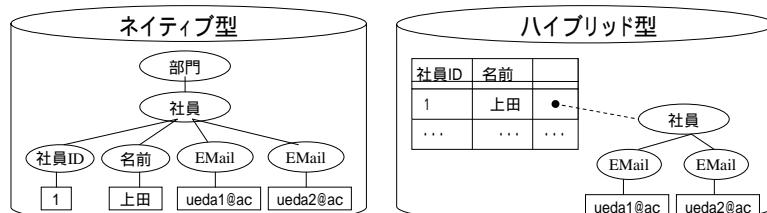
- XML構造でデータを扱うことが可能
 - 階層構造でデータを格納、操作できる
(RDBでバイナリー形式での格納とは異なる)
 - XQuery/Xpathを使用してデータを操作
- データの変化に柔軟に対応可能
 - データ項目を容易に追加できる
 - RDBでの列追加がXML-DBでは容易
 - アプリケーションの変更は最小限
 - 半定型文書への適用



XML DBとは



- XML DBのテクノロジー
 - ネイティブ型
 - XML専用のデータベース
 - XML文書をそのままデータベースに格納
 - XMLのツリー構造に合わせたデータモデル
 - eXist (OSS GNU LGPL), XPriori (Xpriori), etc.
 - ハイブリッド型
 - RDBによるXMLサポート
 - XMLデータを表データにマッピングして格納
 - テキスト(CLOB)としてXML全体を格納
 - DB2 9 (IBM), Oracle 10g R2 (Oracle), etc.





XML DBの操作



- XML DBの操作
 - XQueryを使用してXML DBを検索
 - W3Cで Xquery、XPathを定義
 - XMLデータの更新はXML DB機能を使用
 - XQueryの表現式 FLWOR
 - FOR: 繰り返しを扱う
 - LET: 変数を割り当てる
 - WHERE: 条件を設定する
 - ORDER: 並び替えを行う
 - RETURN: 結果セットを作成する

(例) 事例集から要素技術に"SOAP"を含むデータを年度順で検索

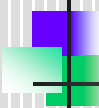
```
for $p in doc('/db/jirei/jirei.xml')
let $year := xs:decimal($p/year)
where contains($p,"SOAP")
order by $year
return $p
```



事例集のXML DBへの蓄積



- 事例集のXML DBへの入力手順
 - 昨年度に収集した事例集の確認
 - 事例集(Excelデータ)のXMLへの変換
 - 事例蓄積システムの構成
 - XML DBへのデータ登録と検索



事例集のXML DBへの蓄積



XML Consortium

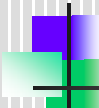
- 昨年度に収集した事例集の内容
 - VBマクロ付きのExcelデータ
 - データ項目
 - 以下の10項目から構成

項目	URL	概要	ユーザー	年度	業種	分類	通信	文書	要素技術	コメント
内容	事例のURL	事例の概要	事例のユーザー名	1998-2006	24業種に分類	14種類に分類	か空白	か空白		

昨年度の事例集の集計データ(抜粋)

年度	業種	分類	通信	文書	要素技術	コメント
2001	放送	放送局	0	0	BML	
2004	サービス	コンテンツ管理・配信	0	0		
2005	放送	放送局	0	0		Web TV, Inc. [GMA]
2004	放送	放送局	0	0		
2001	サービス	コンテンツ管理・配信	0	0	SOAP	2次元情報[1]
2002	サービス	コンテンツ管理・配信	0	0		2次元情報[1]
2001	放送	放送局	0	0		2次元情報の電子化/送信
2004	放送	放送局	0	0	SOAP	
2001	放送	放送局	0	0		
2004	サービス	コンテンツ管理・配信	0	0	SOAP	放送局

© XML Consortium



事例のXML DBへの蓄積



- 事例集のXMLへの変換
 - 2001年の放送局の事例をXMLに変換した例

項目	URL	概要	ユーザー	年度	業種	分類	通信	文書	要素技術	コメント
内容	http://xml/tv_net/	データ放送	放送局	2001	報道	コンテンツ管理・配信			BML	【データ放送】

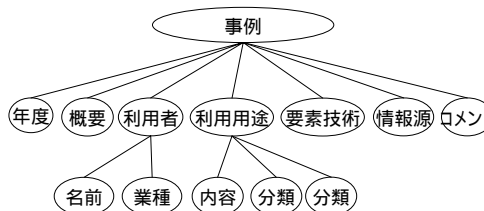
XMLのテキスト表現

```

<?xml version="1.0" encoding="UTF-8"?>
<事例>
  <年度>2001</年度>
  <概要>データ放送</概要>
  <利用者>
    <名前>放送局</名前>
    <業種>報道</業種>
  </利用者>
  <利用用途>
    <内容>コンテンツ管理・配信</内容>
    <分類>通信</分類>
    <分類>文書</分類>
  </利用用途>
  <要素技術>BML</要素技術>
  <情報源 タイプ=URL>http://xml/tv_net/</情報源>
  <コメント>【データ放送】</コメント>
</事例>

```

階層構造(要素のみ)



©

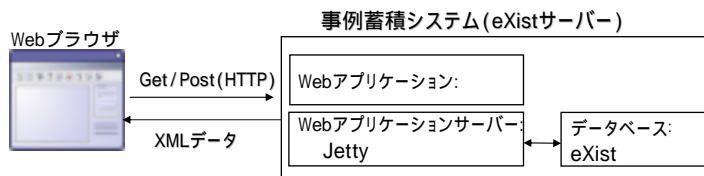
事例集のXML DBへの蓄積



■ 事例入力システムの構成

■ システム構成

- XML DB: OSSであるeXist 1.1.1を採用
 - XQuery, Xpathに対応
 - SOAP, XML-RPC, HTTPインターフェースなどを提供
 - データ管理ツール(GUI)を提供
- Webアプリケーションサーバー: Jettyを使用
- Webアプリケーション: eXist提供のモジュールを使用
- クライアント: Webブラウザ
- eXistとの連携: HTTPインターフェース(Get/Post)を使用



事例集のXML DBへの蓄積



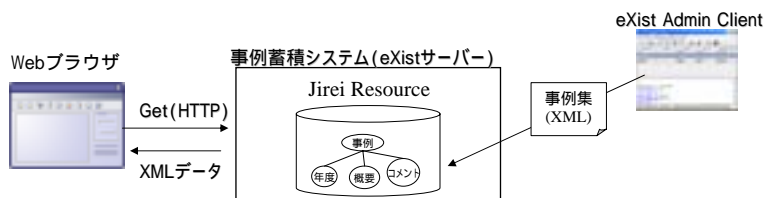
■ XML DBへのデータ登録と検索

■ eXistへのXMLデータの登録

- eXist Admin Clientを使用しXMLデータを登録

■ XMLデータの検索

- eXistのHTTPインターフェースを使用
 - Webブラウザでの実行例
`http://10.11.1.1:8080/exist/rest/db/jirei?_query=for%20$p%20in%20doc("/db/jirei/jirei.xml")%20return%20$p`
 - 実行結果
<事例><年度>2001</年度>...</事例>



事例蓄積 今後検討すべき点



新たな事例の追加、登録した事例の更新の方法

- XML DBへの事例データの追加
- XML DBの事例データの更新

事例活用のための、事例分析の方法

- 普及状況の把握
- 事例から新たなルールの発見

2006年度のまとめ

