

情報ナビゲーションシステム

沖電気工業株式会社
研究開発本部
ヒューマンインタフェースラボラトリ
知識処理チーム
森田 幸伯

欲しい情報を望む形で

人間の嗜好情報、状況(位置情報、端末情報、背景、...)に応じて、適切な情報対話を行う



このためには、情報の内容理解が課題



情報の構造化 + メタ情報で代替



セマンティックWeb技術

メタ情報: 情報の情報。
メタ情報の半自動付与に情報抽出技術を活用

オントロジー: 情報の構造や語彙を定める。
オントロジー構築技術(方法論)を開発する

知識表現言語: 知識を記述する言語。ここではRDF、OWL等、推論機能なども含む
これは、W3C等のツールを活用する

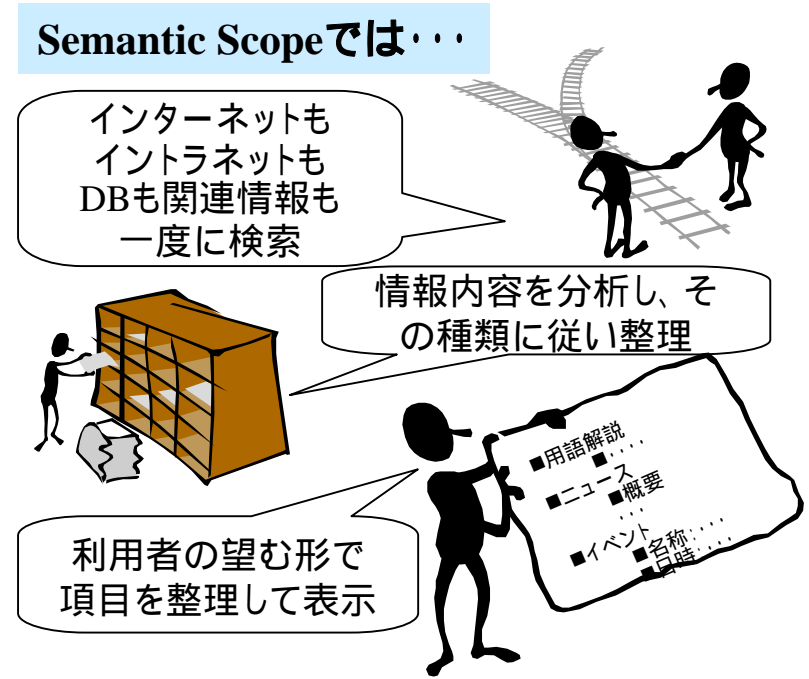
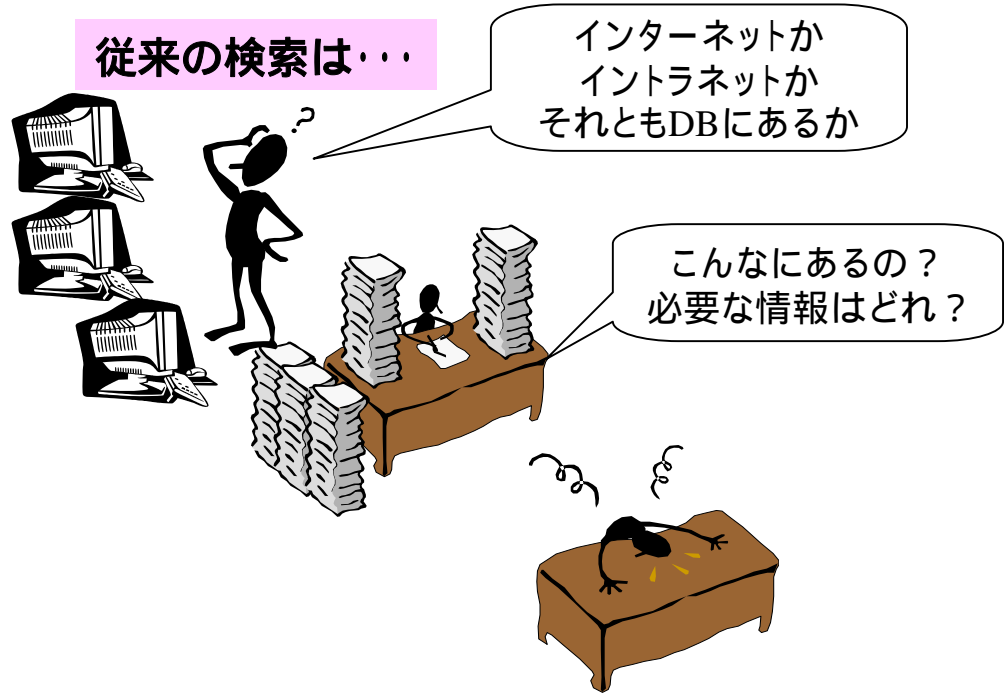


この2つに注力

SemanticScope (仮称) のコンセプト

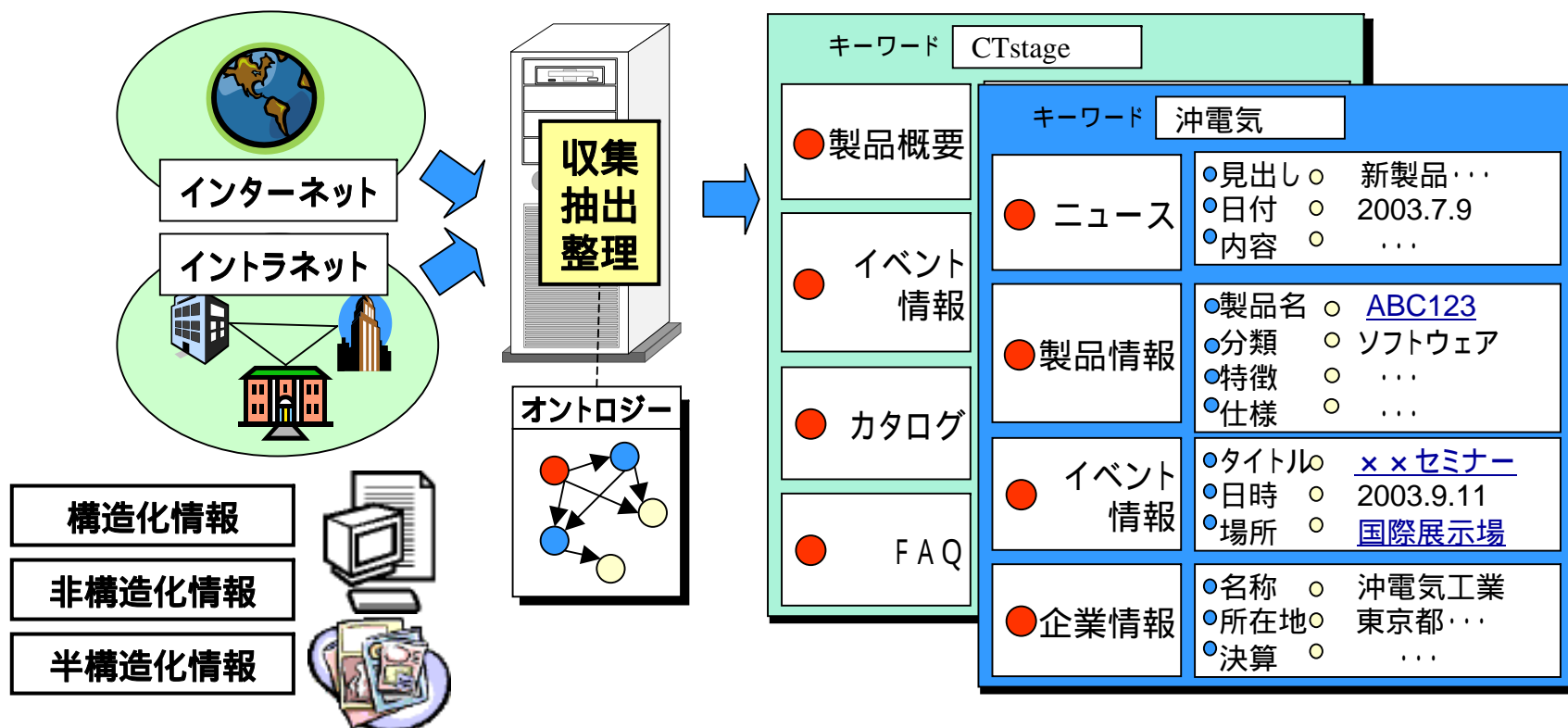
情報を理解し、欲しい情報を望む形でお届けする情報構造化エンジン

- < 特長 >
- オントロジー(概念とその関係に関する意味的(Semantic)な記述)を利用して、
 - (1) 収集: 広く情報にアクセスし、多様な情報を収集
 - (2) 分類・統合: 情報内容を分析し、情報間の関連、項目の属性を把握し体系化
 - (3) パーソナライズ: 体系化された情報のうち、利用者の望む部分だけを抽出して提供

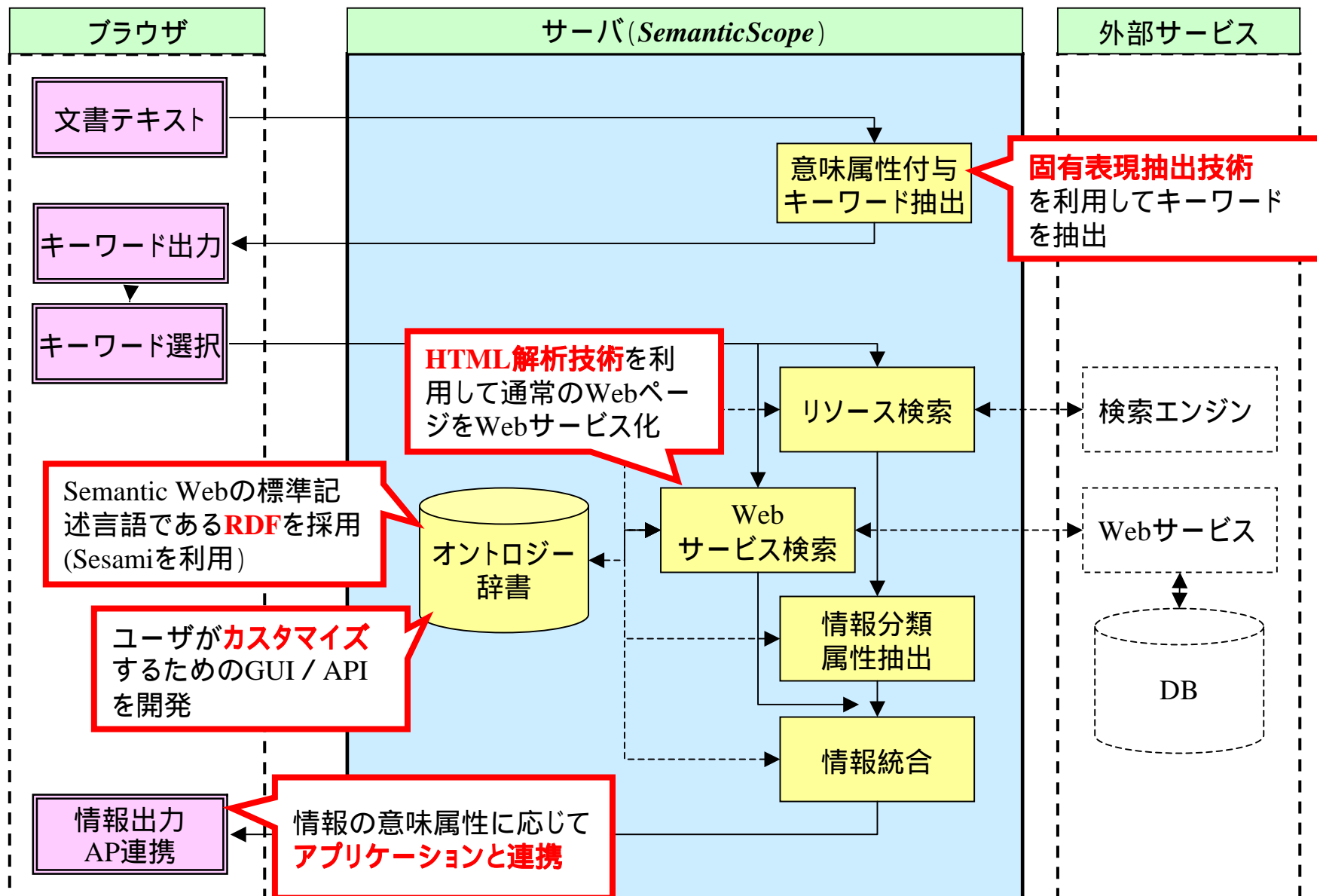


SemanticScopeのコンセプト(続き)

- **必要な情報だけを収集** **情報の洪水からの脱却**
例： ある技術に関する標準化動向だけをまとめて出力
- **抽出結果の体系化** **情報間の関連が明確になり、他の情報との連携も容易**
例： Webサービスで提供される企業情報と最新の決算情報との連携



プロトタイプ版V1 システム全体構成



プロトタイプ版V1 画面出力例

製品名、組織名、人名、技術名を抽出

指定したキーワードの情報を収集

組織情報

沖電気工業

《企業》, 《組織》

- 沖電気工業(株)

概要: 通信機器の
代表者: (社長)篠塚
設立年月: 1949.1
所在地: 105-8480
電話番号: 03-

- 製品
- Oracle Application
 - Oracle Collaborati
 - Oracle E-Business
 - OracleWorld Tokyo
 - Oracle Human Resou
 - CenterStage AS
- 組織
- 沖電気工業株式会社
 - 日本オラクル株式会社
 - 沖電気
 - オラクル
 - AP
 - Oracle
 - 報道機
 - 広報部
 - マネー
 - ネット
 - マネー

指定した属性の情報を収集

製品名、組織名、人名、技術名を抽出

指定したキーワードの情報を収集

指定した属性の情報を収集

製品

- Oracle Application
- Oracle Collaborati
- Oracle E-Business
- OracleWorld Tokyo
- Oracle Human Resou
- CenterStage AS

組織

- 沖電気工業株式会社
- 日本オラクル株式会社
- 沖電気
- オラクル
- AP
- Oracle
- 報道機
- 広報部
- マネー
- ネット
- マネー

沖電気と日本オラクル、業務アプリケーションとVoIPの連携に向けて協業
～人事アプリケーションとカレンダー機能、VoIP技術の融合により、IP電話を拡張～

2003年12月4日

沖電気工業株式会社
日本オラクル株式会社

沖電気工業株式会社（本社：東京都港区 代表取締役社長：篠塚 勝正、以下沖電気）と日本オラクル株式会社（本社：東京都千代田区 代表取締役社長：新宅 正明、以下日本オラクル）は、オラクルが提供する豊富な業務アプリケーションと沖電気を持つVoIP技術の連携に向けて協業することを発表します。

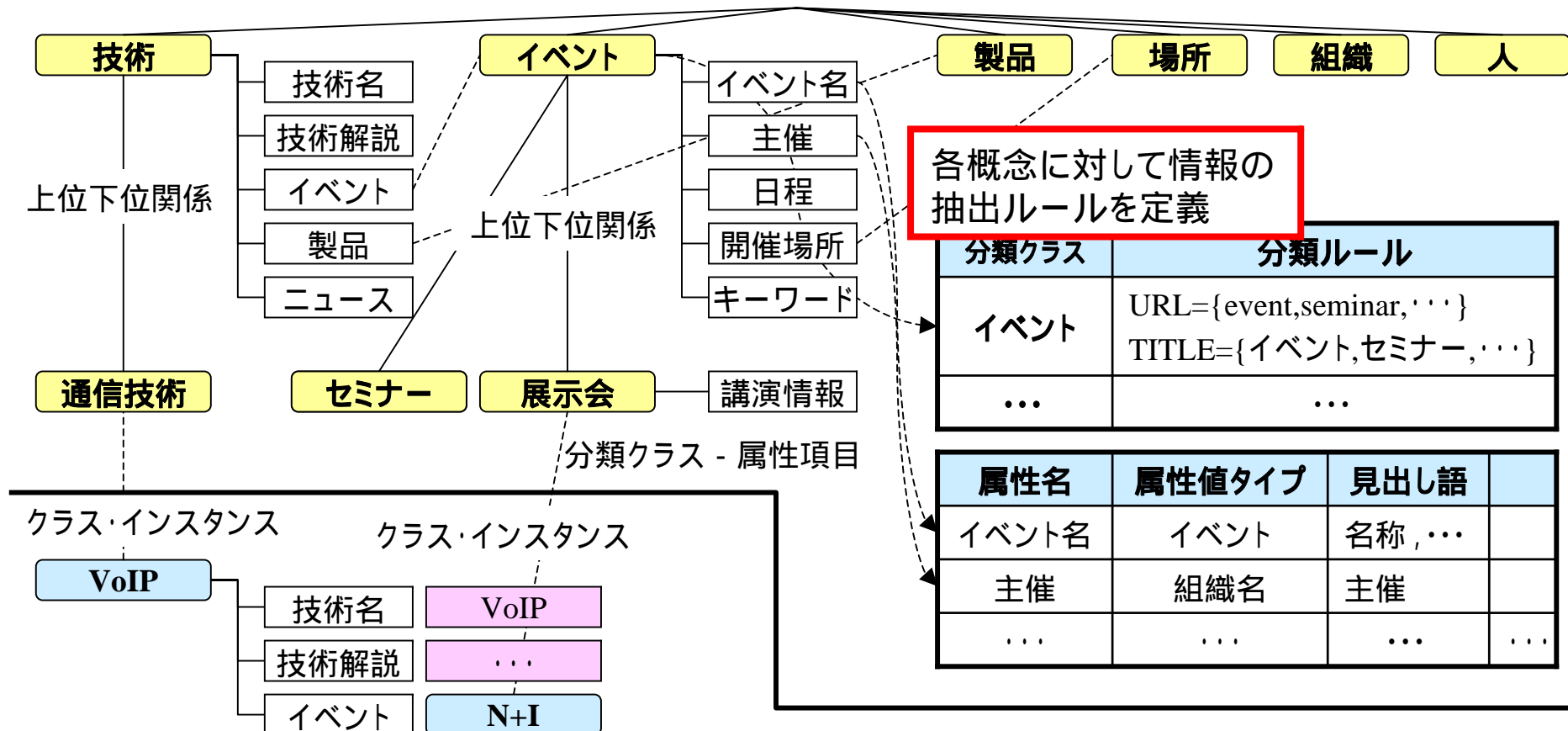
Voice over Internet Protocol) の導入が進んでいます。現状では、抑制など、VoIPをコスト削減の用途に利用しているのが主流です。一も統合が進められていますが、業務遂行の中では、依然として電話にける課題となっています。

アプリケーションとの融合を図りたい沖電気と、拡大しつつあるVoIPラクルとの戦略が一致したことによるものです。両社は、VoIPと業務IP技術の活用が、業務効率向上の用途に拡大することを目指して

オントロジーとは？

- 概念と概念の意味的な関係についての明示的な記述
- それにより、機械があたかも情報の意味を理解したように振舞うことができる

技術 / コミュニケーションに関する概念を網羅



固有表現抽出技術

文書から人名、組織名、日付、イベントなどを抽出し、属性を付与

- 正規表現や辞書を利用
 - 「～さん」「～部長」「田中」 人名
 - 「～大学」「株式会社～」「沖電気」 組織名
- 高いメンテナンス性
- IREX(*)にない固有表現抽出項目拡張

元文書:

```
<table><tr><td>
第1回 セミナー
</td><td></td></tr>
<tr><td> 2003年7月23日
</td><td>
14:00-18:30
</td></tr><tr><td>
大学
第3講義室
<br> 〒 123-4567
市 × × 町1-1
</td><td></td></tr></table>
```



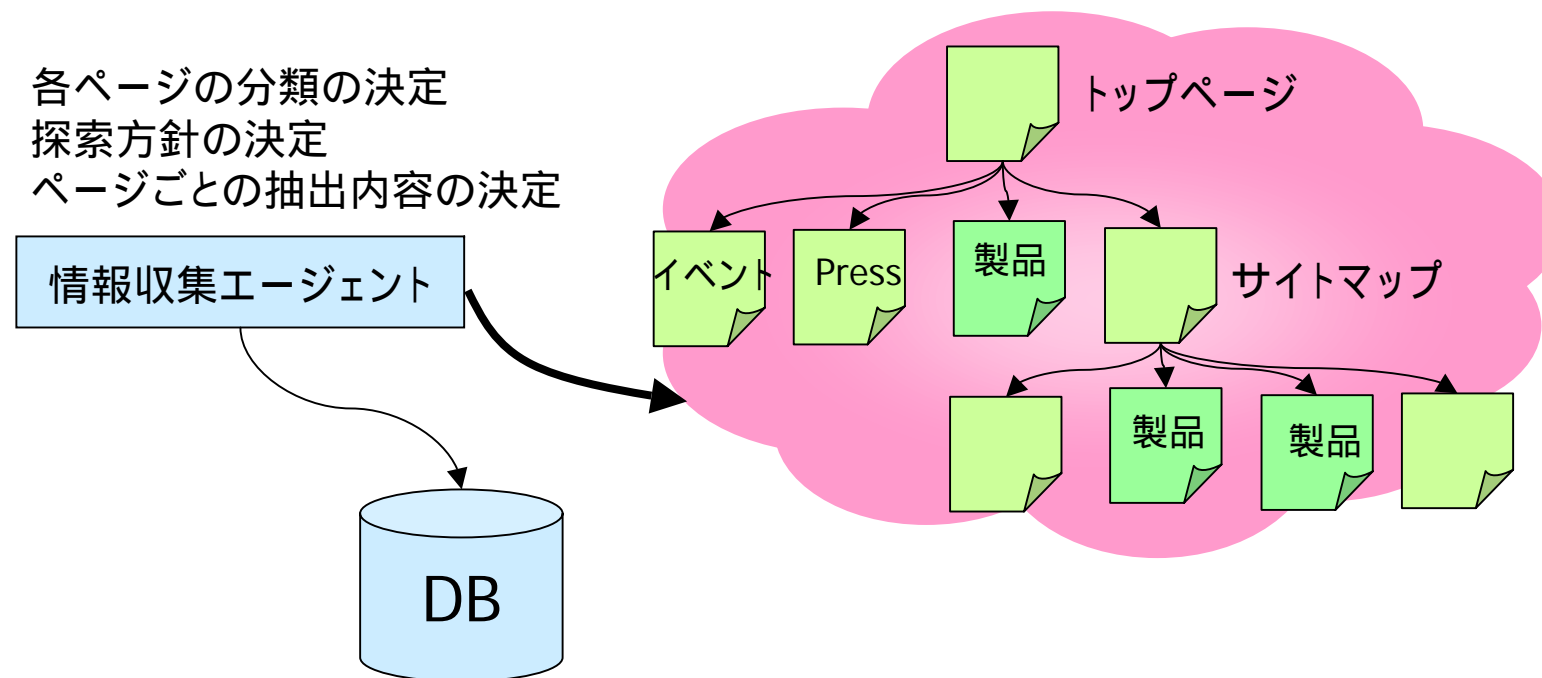
固有表現抽出結果:

```
<table><tr><td>
<イベント>第1回 セミナー </イベント>
</td><td></td></tr>
<tr><td><日付>2003年7月23日</日付>
</td><td>
<時刻>14:00-18:30</時刻>
</td></tr><tr><td>
<組織名> 大学</組織名>
<施設名>第3講義室</施設名>
<br> 〒 <郵便番号>123-4567</郵便番号>
<住所> 市 × × 町1-1 </住所>
</td><td></td></tr></table>
```

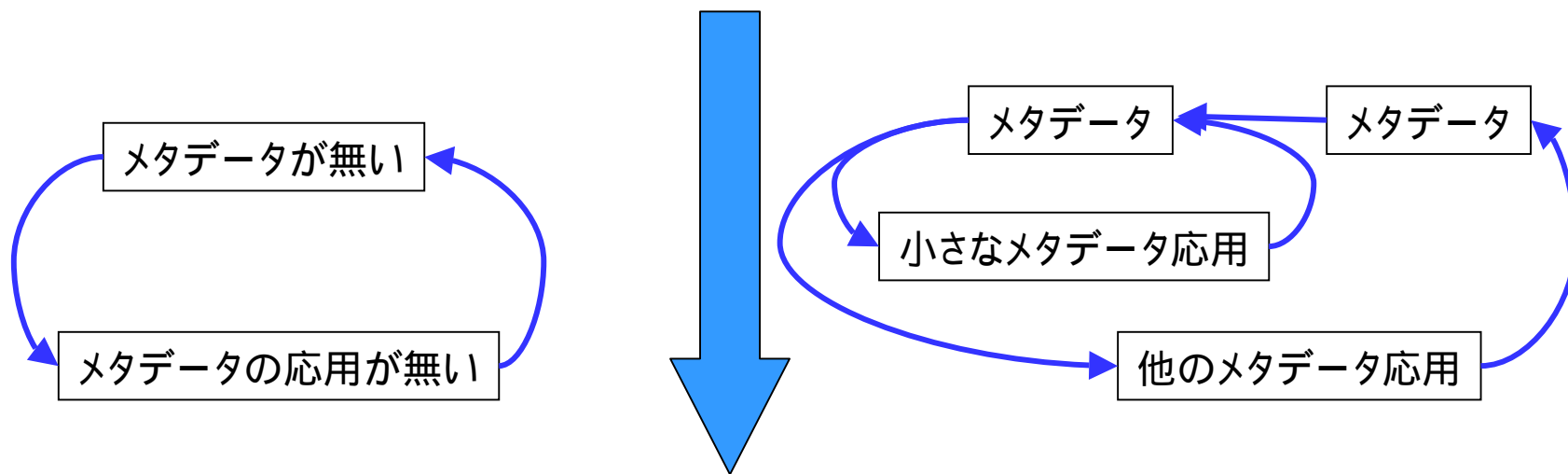
IREX: Information Retrieval and Extraction Exercise 情報検索の抽出コンテスト(1999)

個々の企業サイトから情報収集

- 企業のトップページから企業サイトの情報抽出
 - 製品カテゴリ、プレスリリース、イベント、...



メタデータとその適用事例は鶏と卵



理想は高く、適用は簡単なところから積み上げる

情報の構造記述(オントロジー)と情報抽出技術により
情報を整理するところから

ご清聴ありがとうございました