

相互運用可能な用語体系の構築を 目指して

— 用語辞書, 分類表, Webディレクトリを
対象として —

高久雅生

国立情報学研究所

masao@nii.ac.jp

はじめに

- Webにおける情報の氾濫
- 概念体系・用語体系による情報提供
- 体系構築における労力の軽減
 - 再利用

- 既存の用語体系: どのようなものがあるか?
- 用語体系の再利用に向けた手法

用語体系とは

- 見出し語とその見出し語間の関係を記述したもの
- 見出し語間の関係
 - 同義関係、階層関係(上位・下位関係)、関連関係など
- 例:
 - 用語集 学術用語集, EICネット
 - 辞書 EDR電子化辞書
 - シソーラス JICSTシソーラス, MeSH
 - 分類表 NDC
 - Webディレクトリ Yahoo!, ODP
 - etc.

Webにおける用語体系

- Webページ群の組織化
- 自らが提供するリソースに関するもの
 - 用語集
 - サイトマップ
- 外部提供のリソースに関するもの
 - リンク集
 - Webディレクトリ

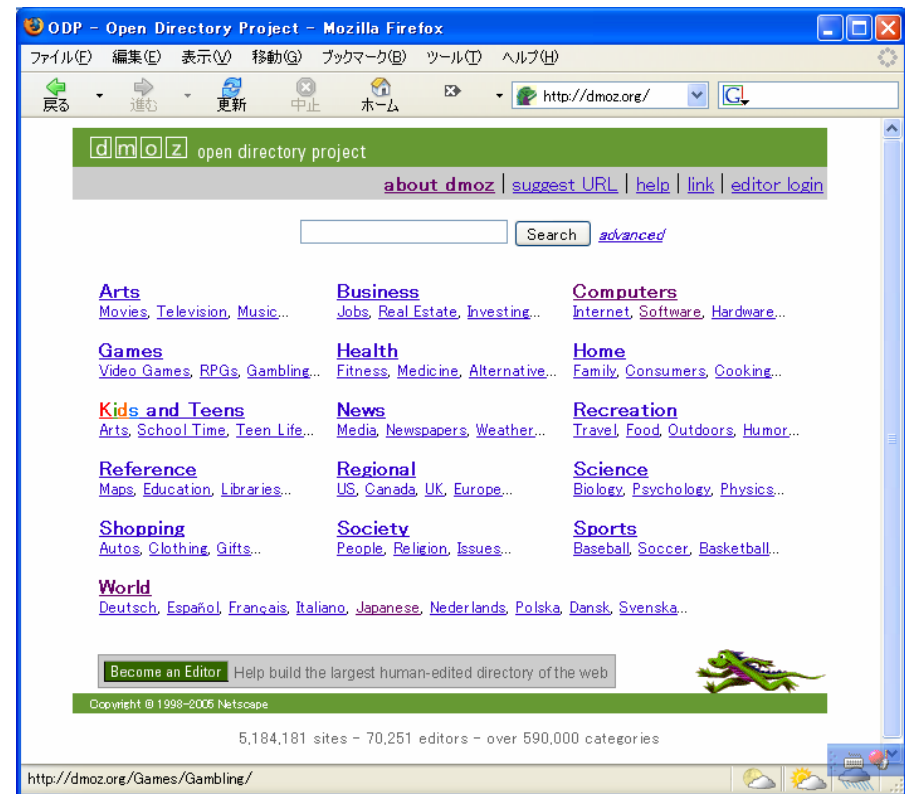
サイト内用語集：EICネット

- 国立環境研究所が運営するポータルサイト
 - 環境分野に関する最新情報
 - 研究者・一般向けに提供
 - <http://www.eic.or.jp/>
 - 環境分野の用語集
 - 2段階のカテゴリ(階層)
 - 自然環境, 地球環境, 大気環境, 水・土壌環境, 健康・化学物質, ごみ・リサイクル, エネルギー, エコビジネス, 環境教育・学習, 環境行政, 環境一般
- (最上位カテゴリ11項目)



Webディレクトリ: ODP

- Open Directory Project
 - <http://dmoz.org/>
 - Netscape社により開始されたプロジェクト
 - ボランティア編集者による盛んな更新
- 階層(ディレクトリ)構造
 - 最上位16レベル
 - 総計約66万階層
- リンク
- 多言語(76言語)
- RDF相当の記述によるディレクトリデータの供給



用語体系構築の特徴

- 分野や用途に応じて作成される
- 同じ分野や用途では
 - 既存の体系を流用
 - または、一部改変により作成されることも多い
- 分野や用途が異なると
 - 新規に作成

個々のニーズに応じて、
それぞれの組織で作成されることも多い

用語体系の構築プロセス

- 方針(どんな体系が必要か?)の決定
- 用語の収集、選定
- 構造の設定
- 表現(どういう形式で表現するか?)
- 実装(計算機で扱う場合の手法)
- 実際の利用
- 評価

用語体系の構築プロセス

類似の体系を参考にする

- 方針(どんな体系が必要か?)の決定

- 用語の収集、選定

資料や文献から

- 構造の設定

分野や利用に関する知識

- 表現(どういう形式で表現するか?)

- 実装(計算機で扱う場合の手法)

- 実際の利用

- 評価

再利用

連携

体系の再利用と連携

- 用途が違ってても再利用できるとうれしい
- 再利用
 - ネットワーク構造を表示しながらブラウジング
 - 複数体系の重なる部分を見ながら検討
- 連携
 - 既存の体系をそのまま(もしくは一部を)利用
 - 他の機能と組み合わせて利用

体系の再利用と連携

- 用途が違ってても再利用できるとうれしい

複数用語体系のブラウジング

- 再利用

- ネットワーク構造を表示しながらブラウジング
- 複数体系の重なる部分を見ながら検討

- 連携

- 既存の体系をそのまま(もしくは一部を)利用
- 他の機能と組み合わせて利用

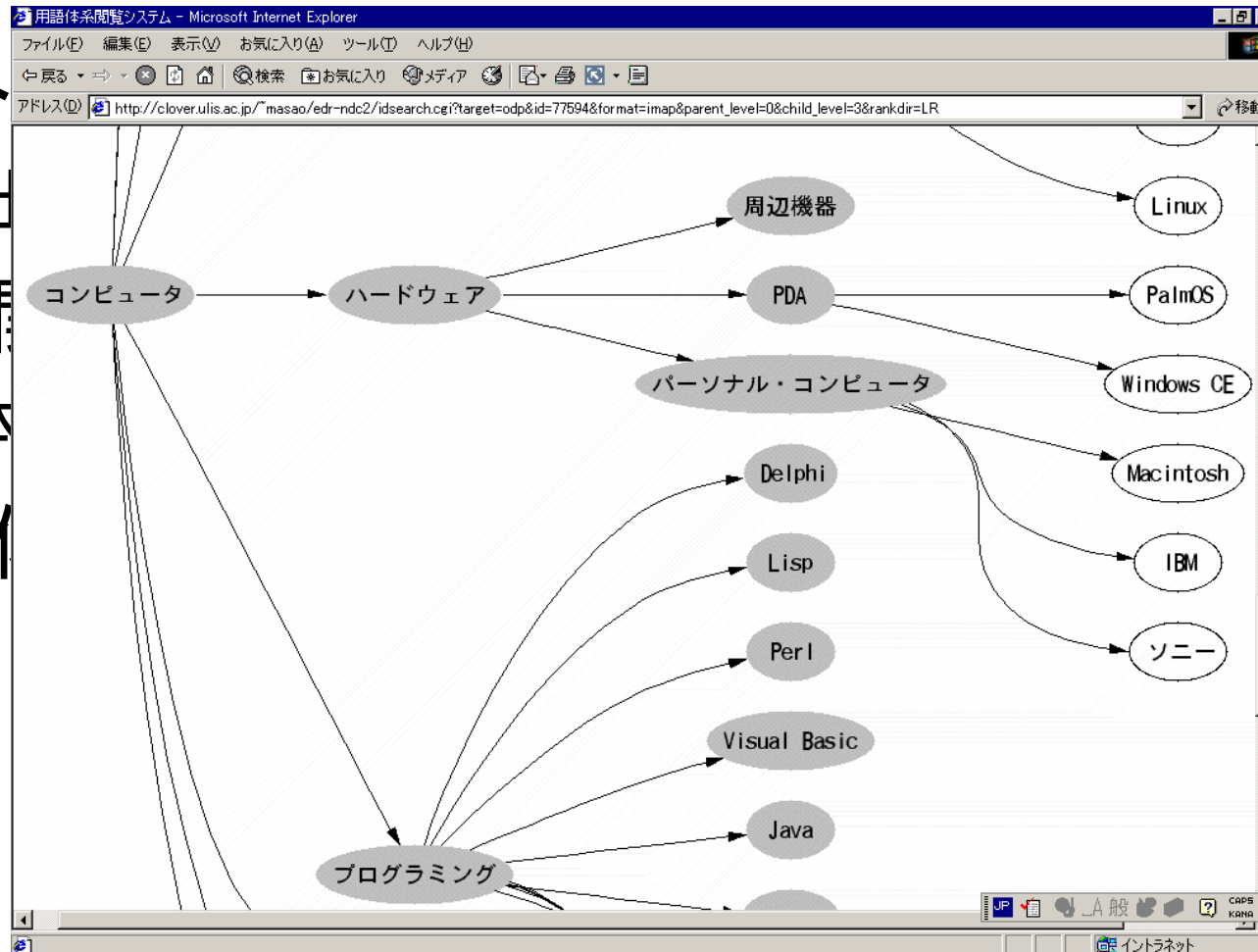
用語体系を提供するWebサービス

複数用語体系ブラウジング

- ネットワーク構造を表示
 - 見出し語をノード、階層関係をリンク
- 階層関係を持つ複数の用語体系に対応
 - 5体系で実験(ODP(+kt)、NDC、EDR、EIC)
- 用語体系間に共通する見出し語を重ね合わせ

複数用語体系ブラウジング (2)

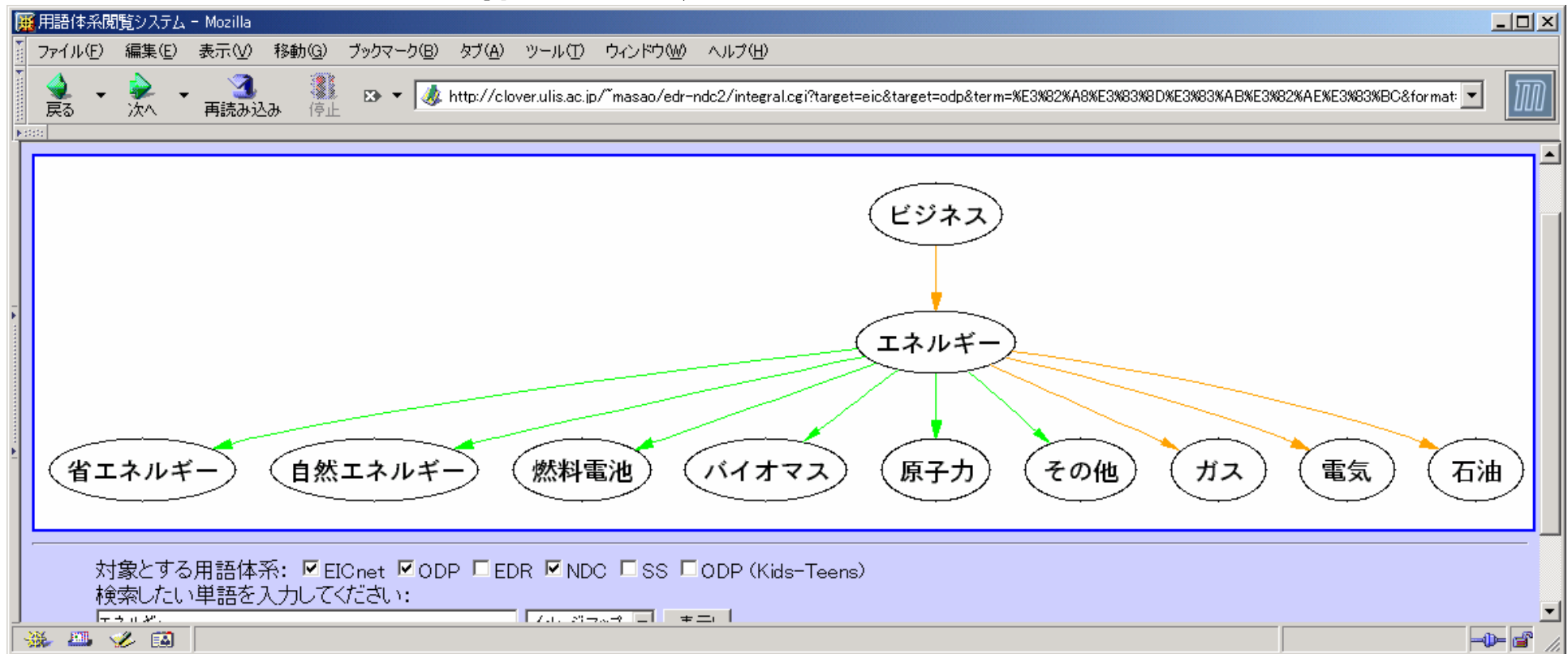
- ネット
- 見出し
- 階層
- 5体
- 用語体



わせ

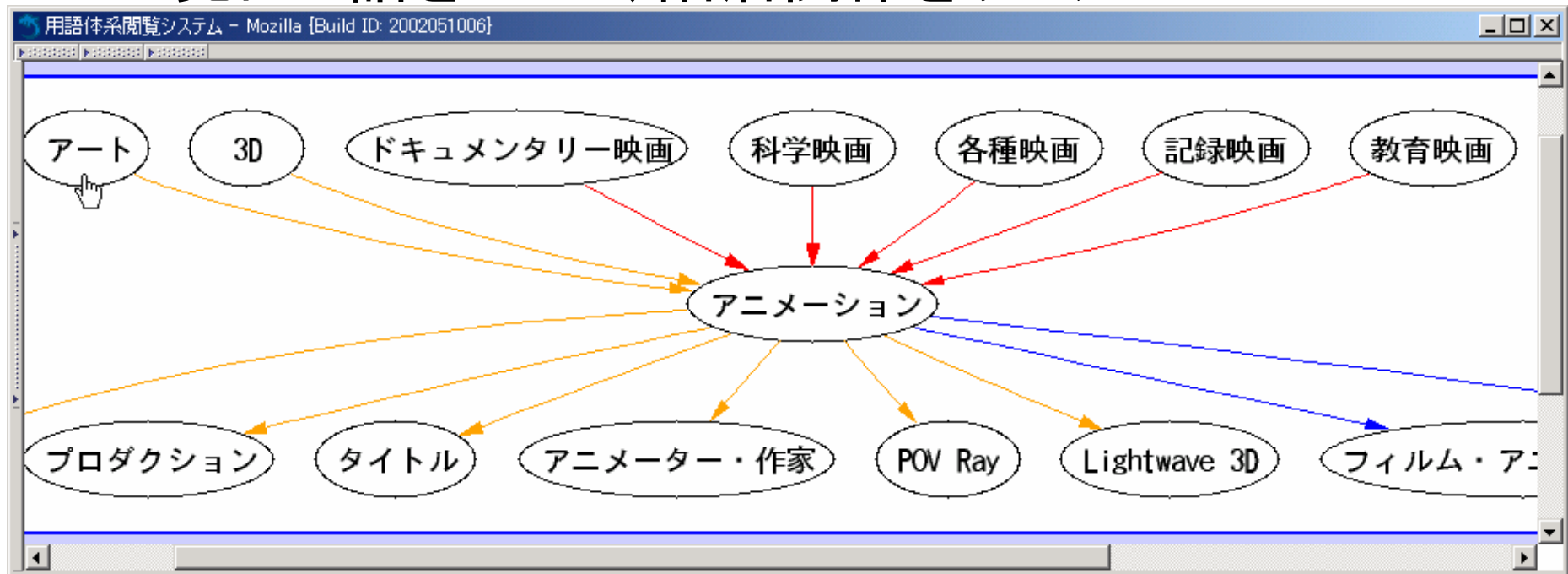
複数用語体系ブラウジング (3)

- ネットワーク構造を表示



複数用語体系ブラウジング (4)

- ネットワーク構造を表示
 - 見出し語をノード、階層関係をリンク



連携利用

- ネットワーク上で独立して提供
- 様々な利用(再利用、連携)
- 例
 - データ配布(ODP)
 - 体系データをRDF形式で提供
 - API提供(Google、Amazon)
 - 問い合わせ可能なWebサービスAPIを提供

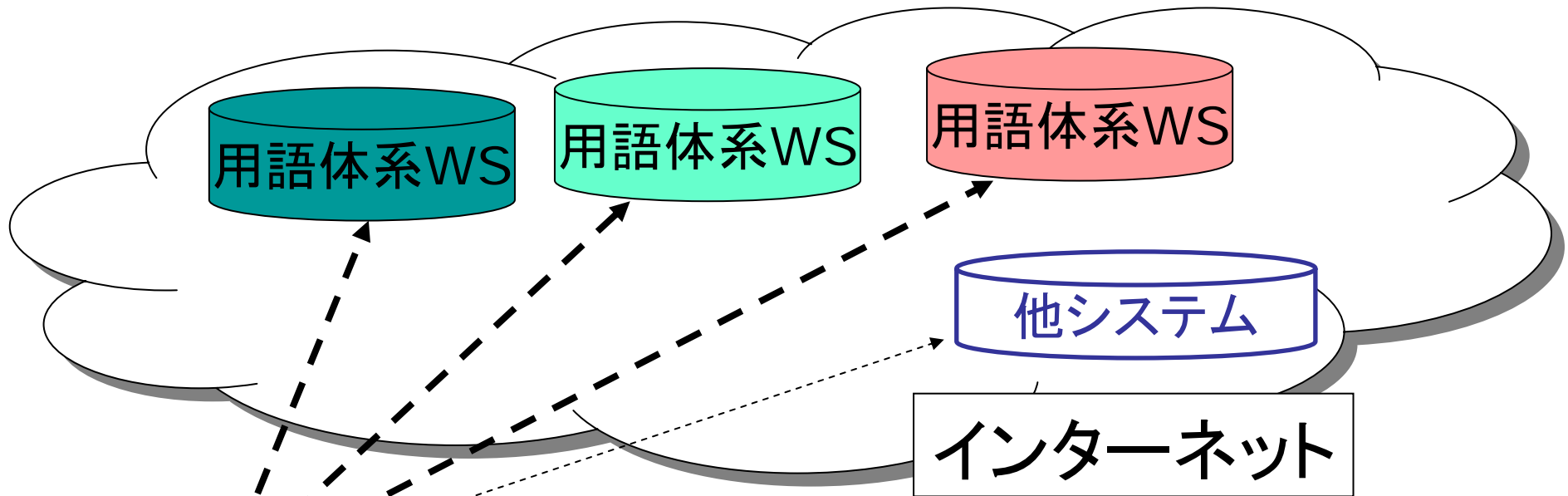
ODPデータの配布

```
<Topic r:id="Top/World/Japanese/コンピュータ/ソフトウェア/インターネット/クライアント">
  <catid>1124401</catid>
  <d:Title>クライアント</d:Title>
  <d:Description>このカテゴリでは、インターネットに関するクライアントソフトに関する情報を
    扱うサイトを掲載します。 </d:Description>
  <lastUpdate>2005-07-21 09:07:04</lastUpdate>
  <narrow r:resource="Top/World/Japanese/コンピュータ/ソフトウェア/インターネット/クライアント/ブラウザ"/>
  <narrow r:resource="Top/World/Japanese/コンピュータ/ソフトウェア/インターネット/クライアント/FTP"/>
  <narrow r:resource="Top/World/Japanese/コンピュータ/ソフトウェア/インターネット/クライアント/NTP"/>
  <narrow r:resource="Top/World/Japanese/コンピュータ/ソフトウェア/インターネット/クライアント/メーカー"/>
  <narrow r:resource="Top/World/Japanese/コンピュータ/ソフトウェア/インターネット/クライアント/ファイル共有"/>
  <narrow r:resource="Top/World/Japanese/コンピュータ/ソフトウェア/インターネット/クライアント/コミュニケーション"/>
</Topic>
```

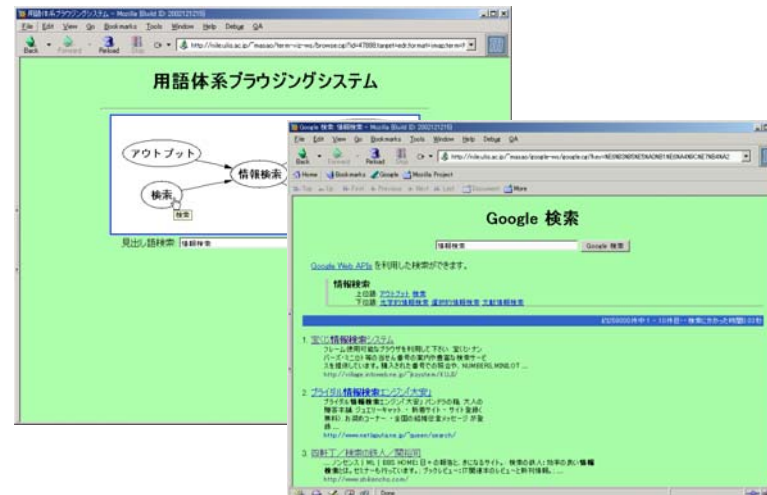
用語体系提供API(アクセス手段)

- 用語体系提供Webサービス
 - ネットワーク上に分散して存在する用語体系の欲しい部分を動的に取得可能
 - ODP、EDRの2体系を対象にしたものを構築
- 2種類のフロントエンドシステム
 - 検索サービス(Google)との連携例
 - ブラウジングツールとの連携例

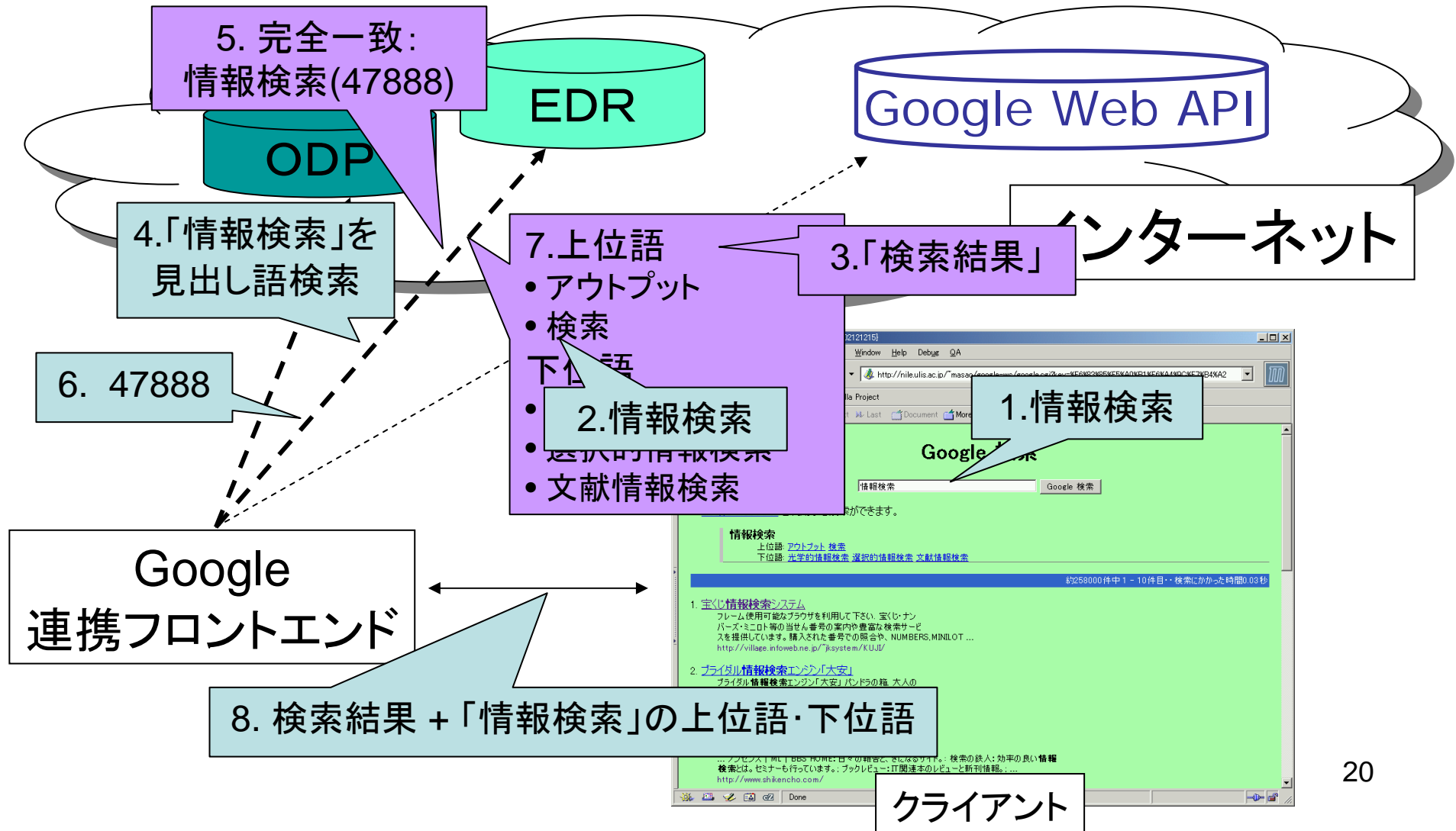
用語体系提供とその連携システム 概念図



フロントエンド
(連携利用システム)



フロントエンド Google Web API 連携システム



まとめ

- 用語体系
 - 情報提供の際に重要な要素
- 構築の労力・コストを低減
- 再利用と連携を可能にする
 - データ配布、アクセス手段(API)の提供
 - 新規に作成する他の体系でも流用可能に
 - 検索システムとの連携などサイト内要素との連携