

相互連携可能な情報提供システムの研究

— 用語体系および検索システムへの適用 —

高久雅生

筑波大学 図書館情報メディア研究科

博士後期課程

目次

- 背景と目的
- 検索システム間の連携
 - ①JAPAN/MARC検索システム
 - ②Dublin Core検索システム
- 用語体系間の連携
 - ③複数用語体系ブラウジングシステム
- 異種情報提供システム間の連携
 - ④Webサービスによる用語体系提供システム
- まとめ

背景

- Webの普及、情報量の増大
- 多様な利用者
- ユーザの多様な情報要求
 - キーワードに関連する情報が欲しい
 - 分野・カテゴリに属する情報が欲しい
 - 特定のWebページ、サイトにある情報が欲しい

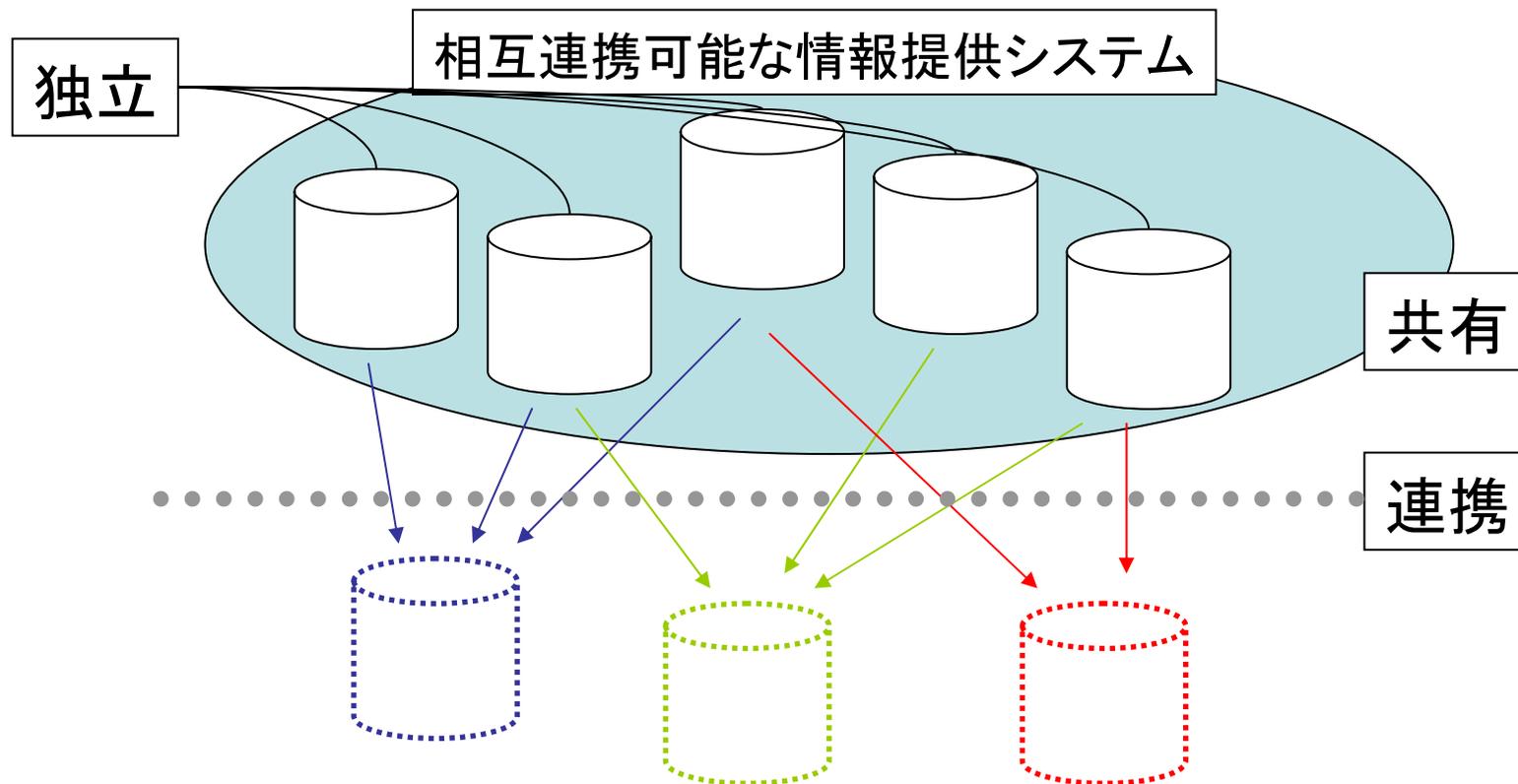
背景(2)

- 提供者側のニーズ
 - 多様な要求に応える情報提供システム
- 既存のシステムでは不十分
 - 情報提供システムの長所・短所
 - サーチエンジン: キーワード選択が困難
 - Webディレクトリ: 階層構造の把握が困難
 - 単一のシステムだけで多様な要求全てに対応するのは現実的ではない
- 多様な情報提供システム
 - 多様な要求に応じて、新たなシステム構築が必要
 - 構築コストが問題

目的

- 新たな情報提供システム構築の支援
 - － 構築にかかるコスト軽減が必要
- 既存の情報提供システム同士の連携
 - － 長所を生かしたまま、短所を補いあう
- 2種類の情報提供システムを対象
 - － 用語体系
 - － 検索システム
 - － 両者とも情報提供システムにおける基盤

情報提供システムの相互連携 イメージ図

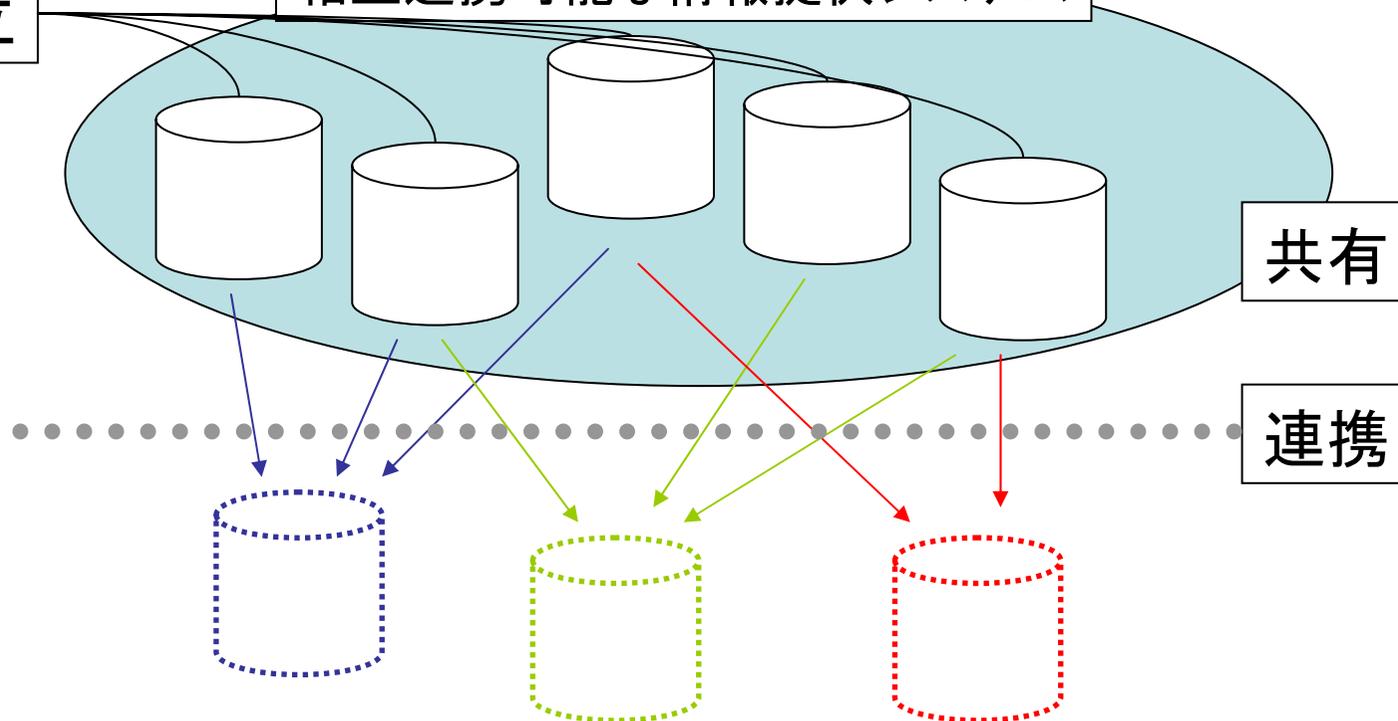


共有

- 異なる情報提供システムを共通のやり方で使える
- 類似の情報提供システム間を切り替えて利用できる

独立

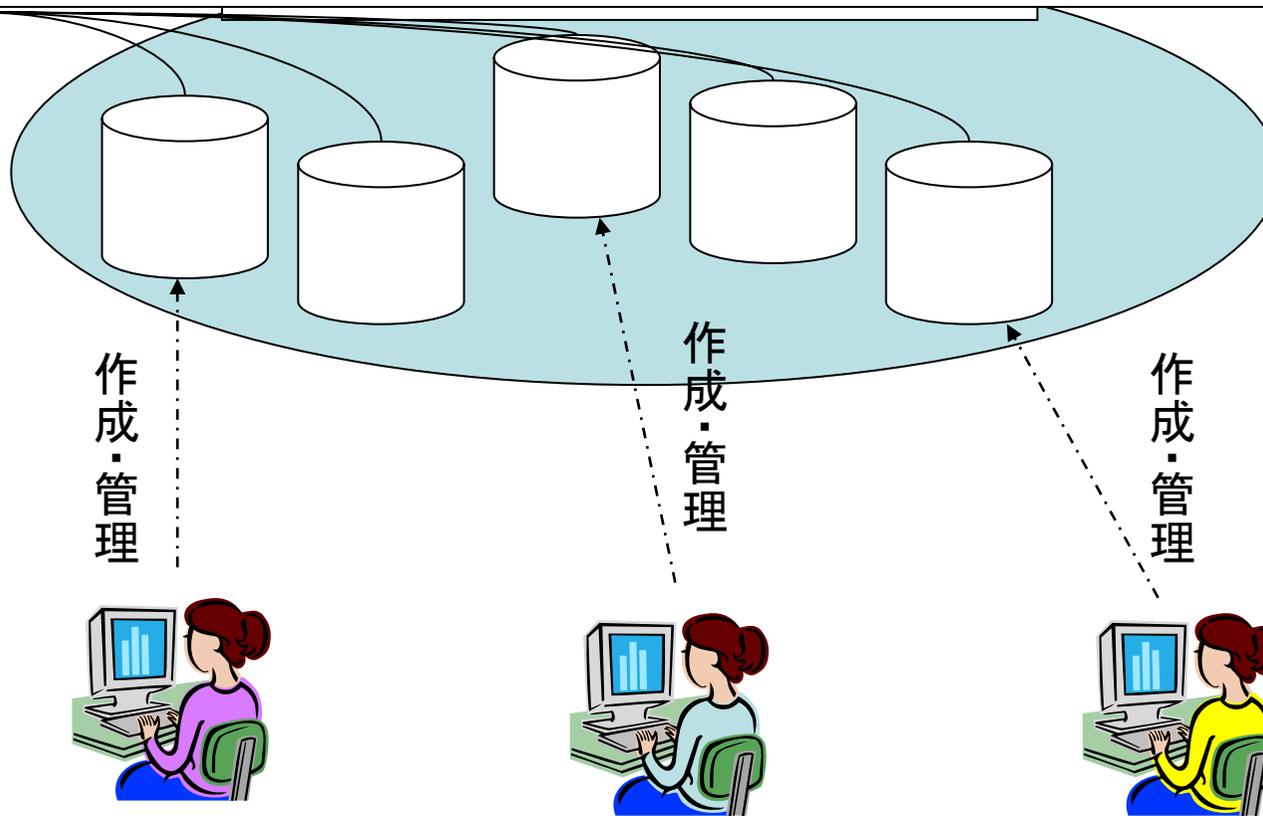
相互連携可能な情報提供システム



独立

- 情報を提供する組織や個人ごとに作成、提供が自由に行える
- 情報提供者毎の得意分野などを活かした情報提供が可能になる

独立

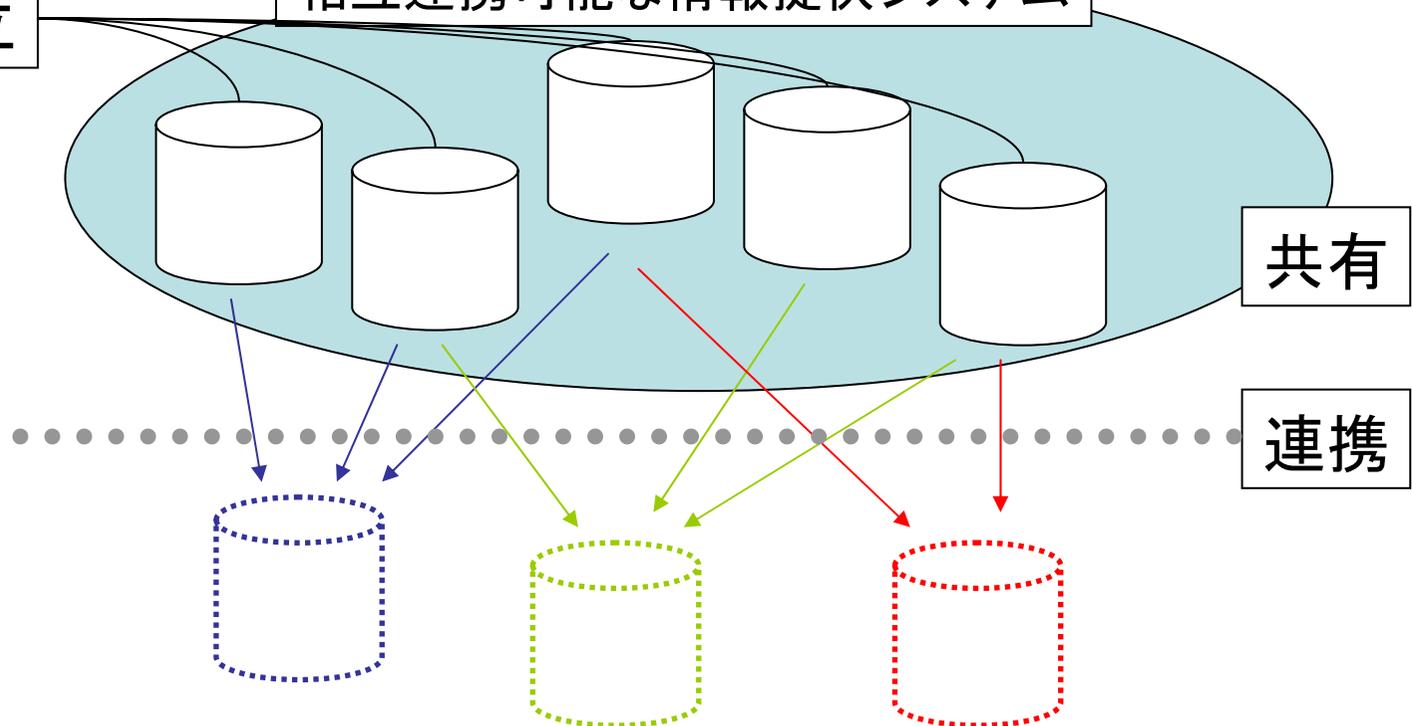


連携

- 他で作成された情報提供システムと相互に利用できる、自由な組み合わせで連携できる
- 自身の提供するシステムには無い機能を補うことが可能

独立

相互連携可能な情報提供システム



目次

- 背景と目的
- 検索システム間の連携
 - ①JAPAN/MARC検索システム
 - ②Dublin Core検索システム
- 用語体系間の連携
 - ③複数用語体系ブラウジングシステム
- 異種情報提供システム間の連携
 - ④Webサービスによる用語体系提供システム
- まとめ

情報提供システムにおける 検索システムの特徴

- 情報提供における重要な要素
- 検索システムがより身近な存在に
 - 例: サイト内検索
- 検索システム間での違い
 - インタフェース
 - 検索式の指定方法
 - アクセスポイント

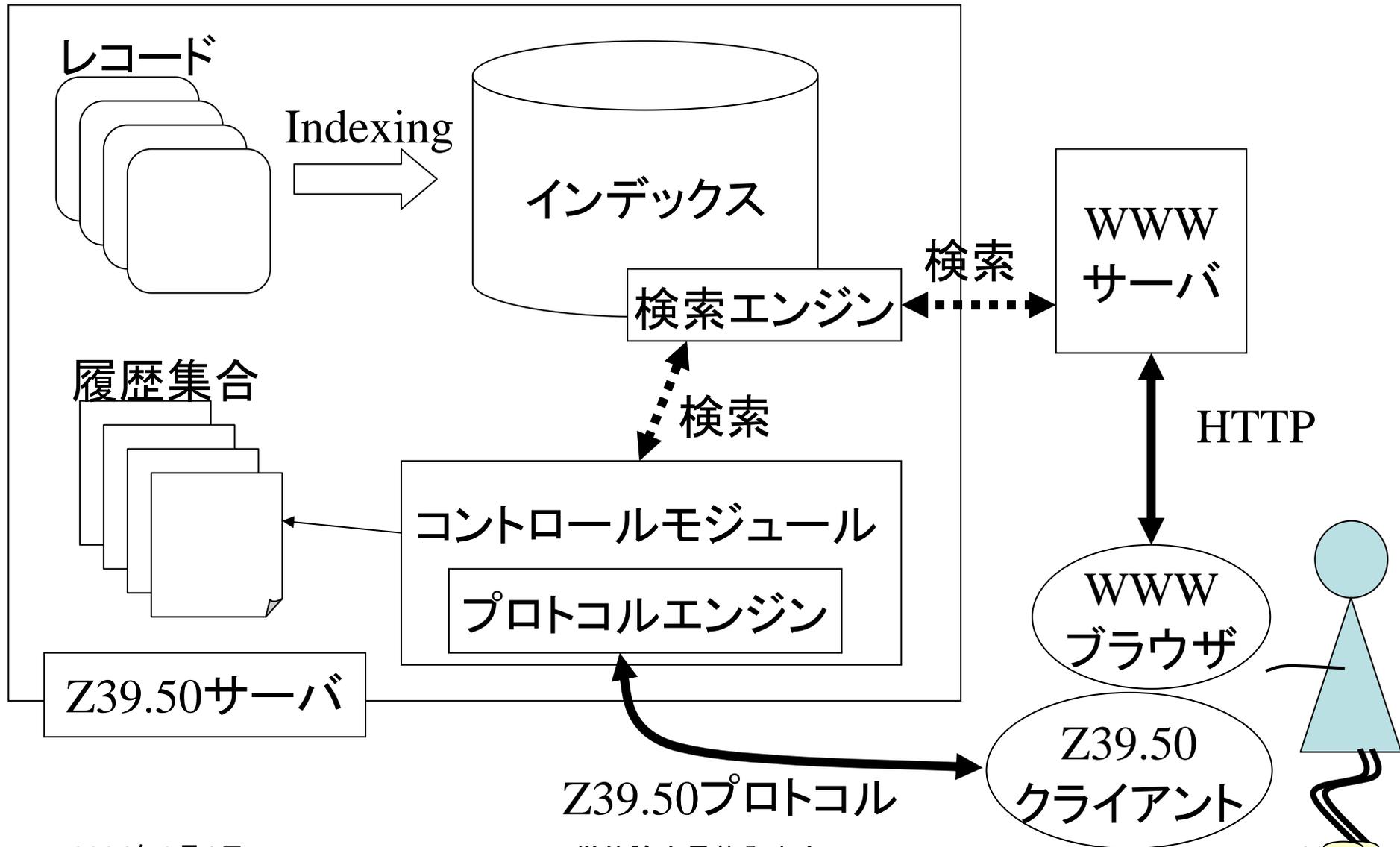
Z39.50検索システム

- 異なる検索システム間での利用の問題
- Z39.50 (ISO 23950)
 - 情報検索のための通信プロトコル
 - 検索式の際のやり取りを標準化し、多種多様な検索システム利用を共通に行える
 - 同一インタフェースでの利用や横断検索等の相互連携が可能
 - 欧米を中心に普及：図書館OPAC, 商用データベース
 - 1999年当時、日本語書誌データを対象としたサーバ・クライアント双方を備えたZ39.50システム構築の報告例なし

本研究で構築した Z39.50システムの概要

- 2つのZ39.50システムの構築
 - JAPAN/MARC検索システム
 - Dublin Core検索システム
- 特徴
 - 日本語書誌データを対象
 - 大規模データに対応
 - Z39.50とWWWの両者で利用可能
 - 相互運用性の向上

システムの構成



2004年9月1日

学位論文最終発表会

16

目次

- 背景と目的
- 検索システム間の連携
 - ①JAPAN/MARC検索システム
 - ②Dublin Core検索システム
- 用語体系間の連携
 - ③複数用語体系ブラウジングシステム
- 異種情報提供システム間の連携
 - ④Webサービスによる用語体系提供システム
- まとめ

用語体系とは

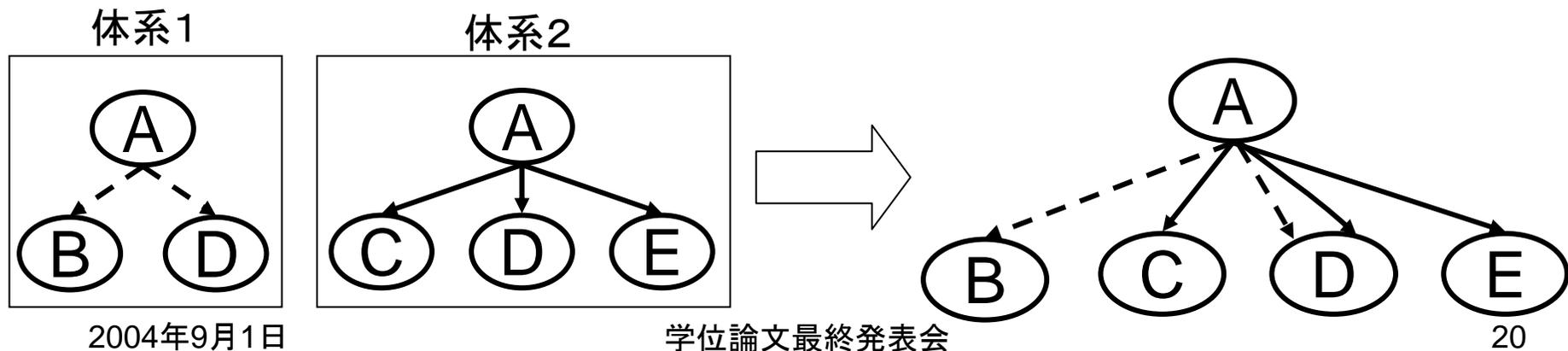
- 見出し語とそれらの間の関係を記述
 - 階層関係、同義関係、関連関係など
 - 例: シソーラス(情報検索)、辞書、分類体系、Webディレクトリ、オントロジー(AI)
- 本研究では:
 - 提供機能をもつ情報提供システムの一つ
 - 見出し語と階層関係を共通の要素
 - 用語体系間の相互連携 ⇒ ③ブラウジングシステム
 - 用語体系と他システム間の相互連携 ⇒ ④Webサービスシステム

複数用語体系ブラウジングシステムの概要

- ブラウジングシステム
 - 用語間の関係をグラフ構造として表示
 - Web上のクリックابلマップ
- 複数の用語体系に対応
 - 5つの用語体系に適用
- 複数用語体系の重ね合わせ
 - 様々な組み合わせによる重ね合わせ
 - 複数体系の違った観点からの把握
 - 語彙の収集、再利用

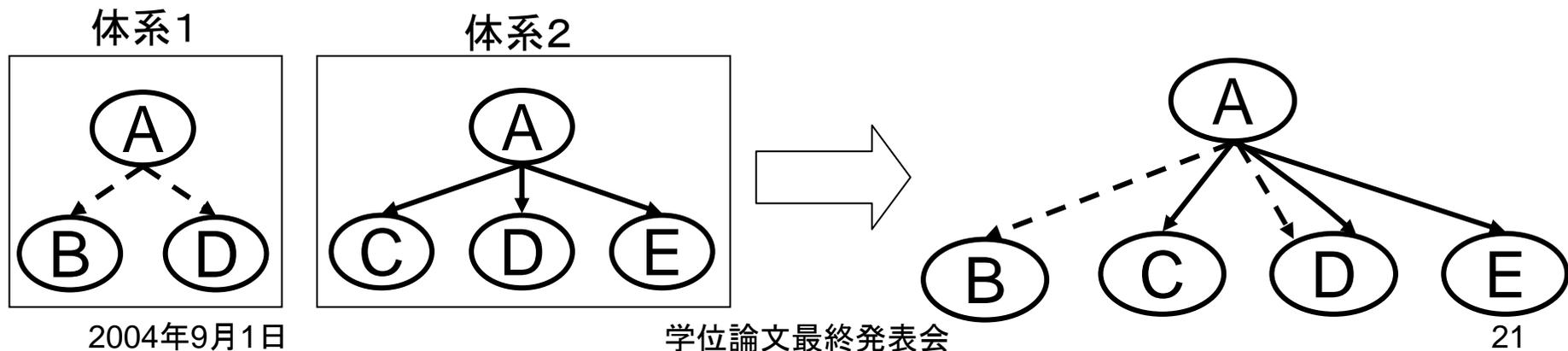
用語体系の重ね合わせ

- 複数の用語体系を重ね合わせる
- 体系間に共通する「見出し語」「上位・下位関係」を対象とする
- 用語体系を有向グラフとしてみなし、同一の見出し語からなるノードを統合する



用語体系の重ね合わせ(2)

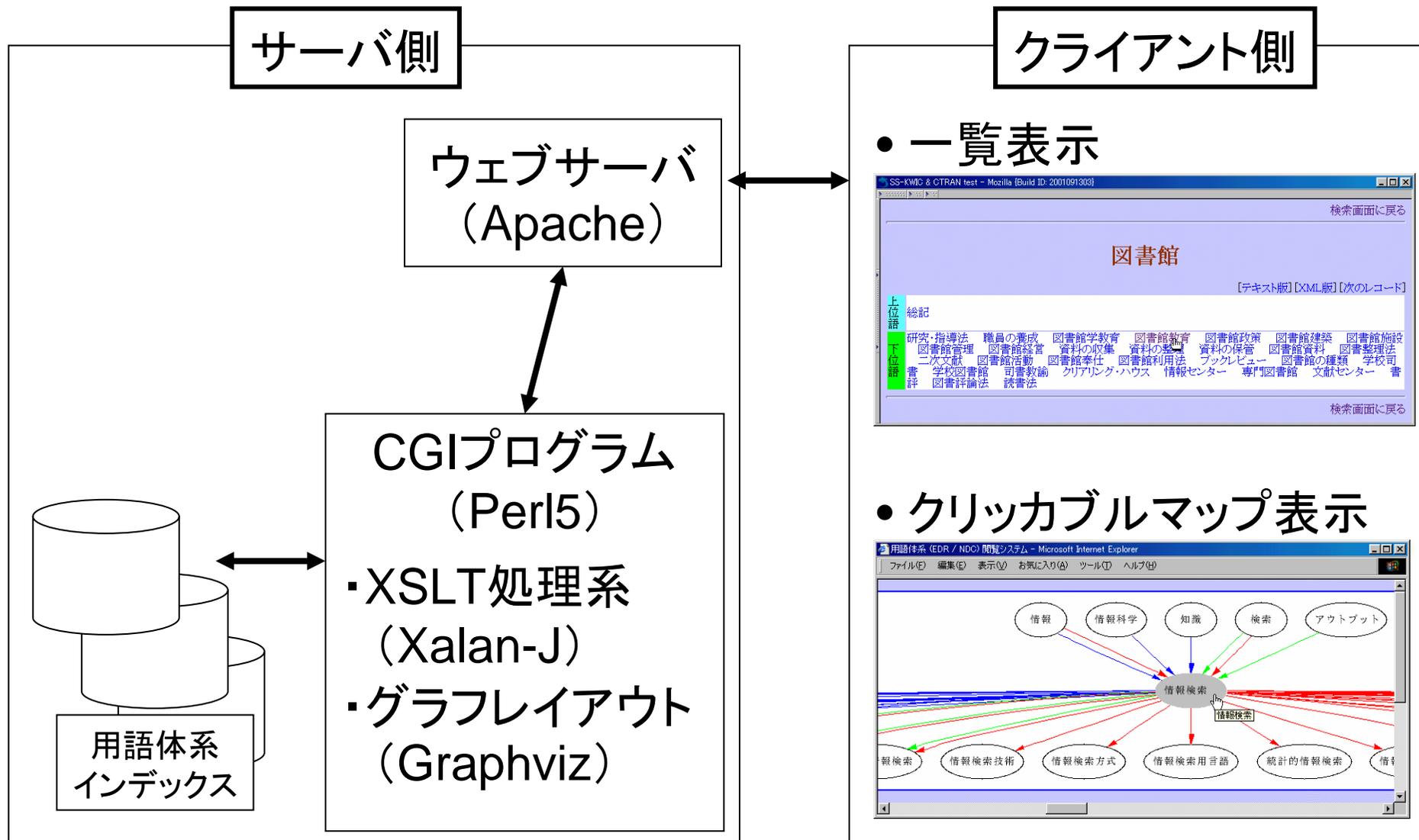
- 重ね合わせ結果
 - 人間が判断し、取捨選択を行う
 - 用語の収集
 - 異なる観点の発見
 - 別の用語体系への再利用



対象データ

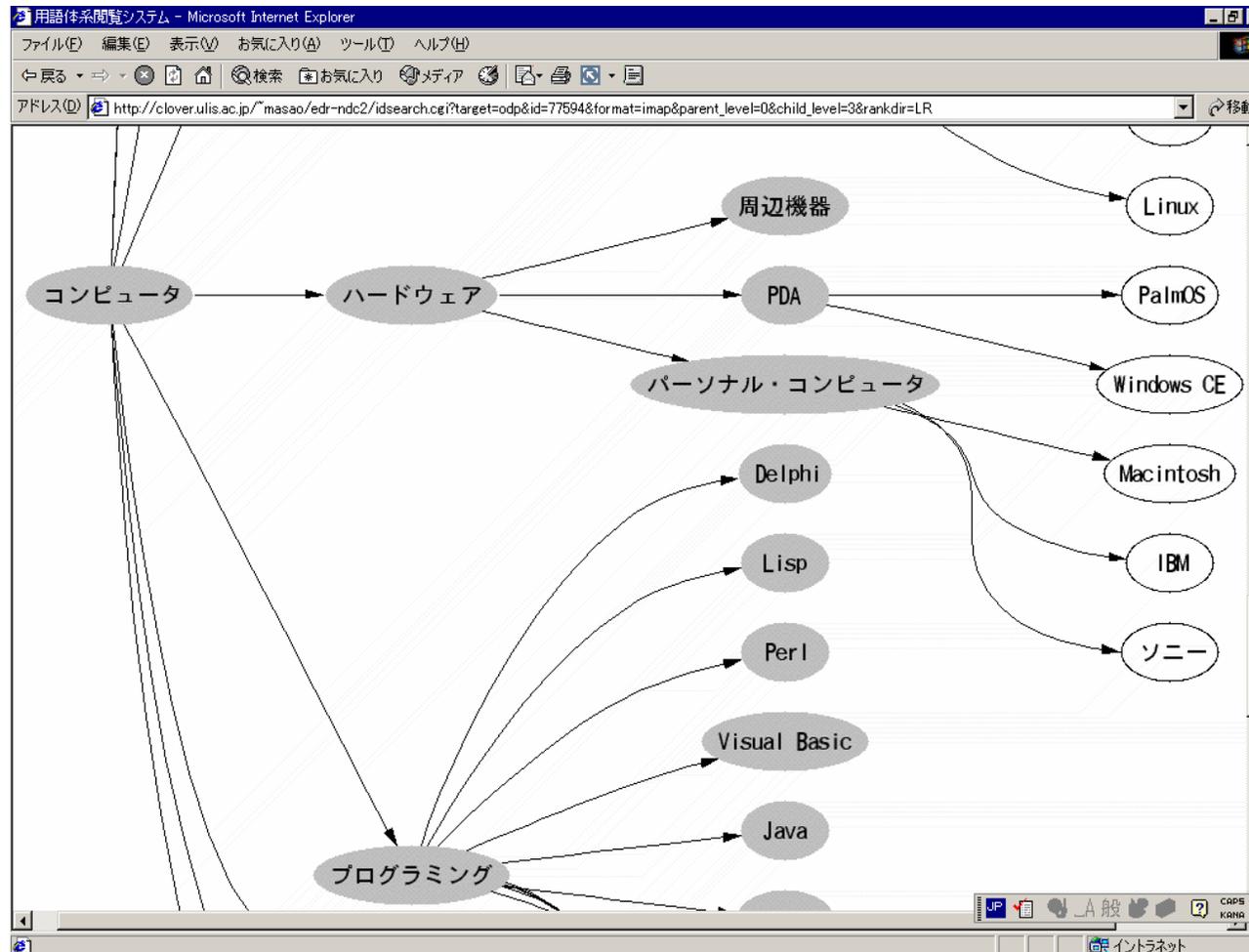
用語体系	用語数	異なり語数	階層関係数	特徴
EDR	118,782	118,134	184,958	情報処理分野の専門用語を対象とした大規模な体系
NDC	30,659	25,410	102,639	全学問分野を対象とした大規模な体系
EIC	391	361	380	生物・化学・気象などの複合領域から構成される環境分野を対象とした体系
ODP	446,926	150,387	446,928	WWW上に存在する情報全てを対象とした大規模な体系
ODP-kt	2,059	1,617	2,058	WWW上に存在する情報のうち、子ども向けの情報に限定した部分を対象とした体系
(総計)	--	292,479	--	

システムの構成



表示例(クリックابلマップ)

- 複数階層表示 -

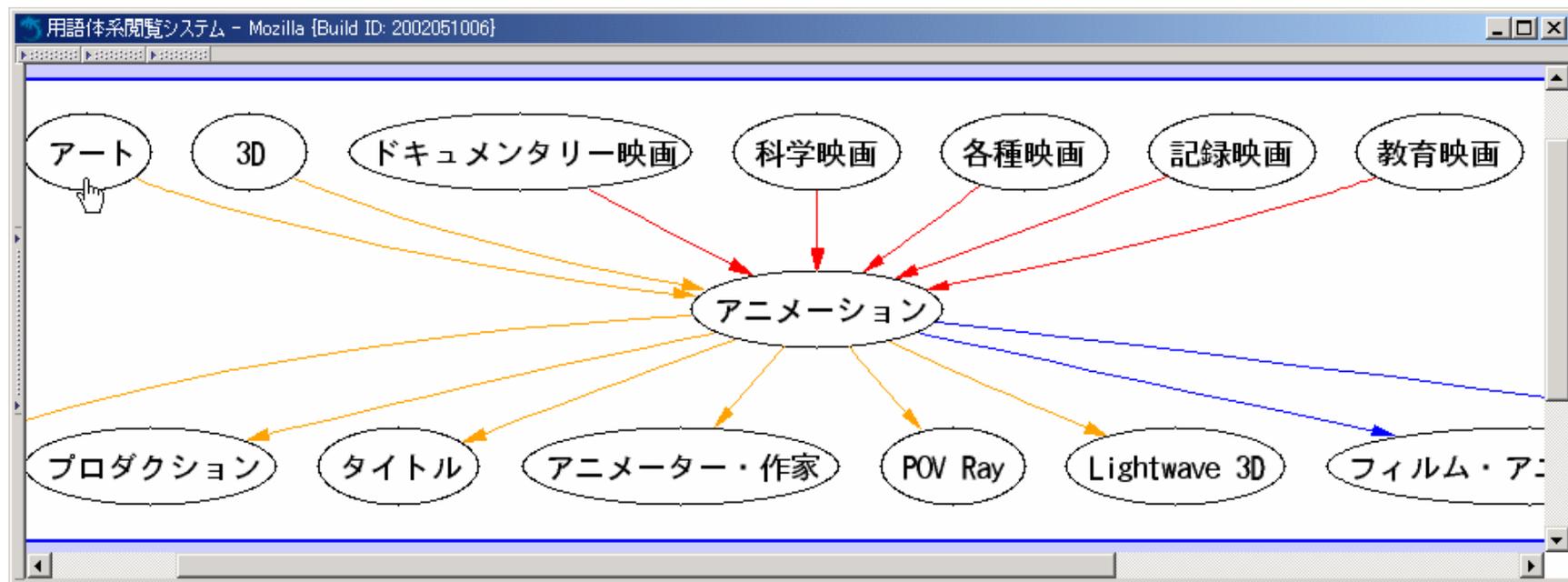


2004年9月1日

「コンピュータ」(ODP)

25

表示例(クリックブルマップ) - 重ね合わせ -



「アニメーション」

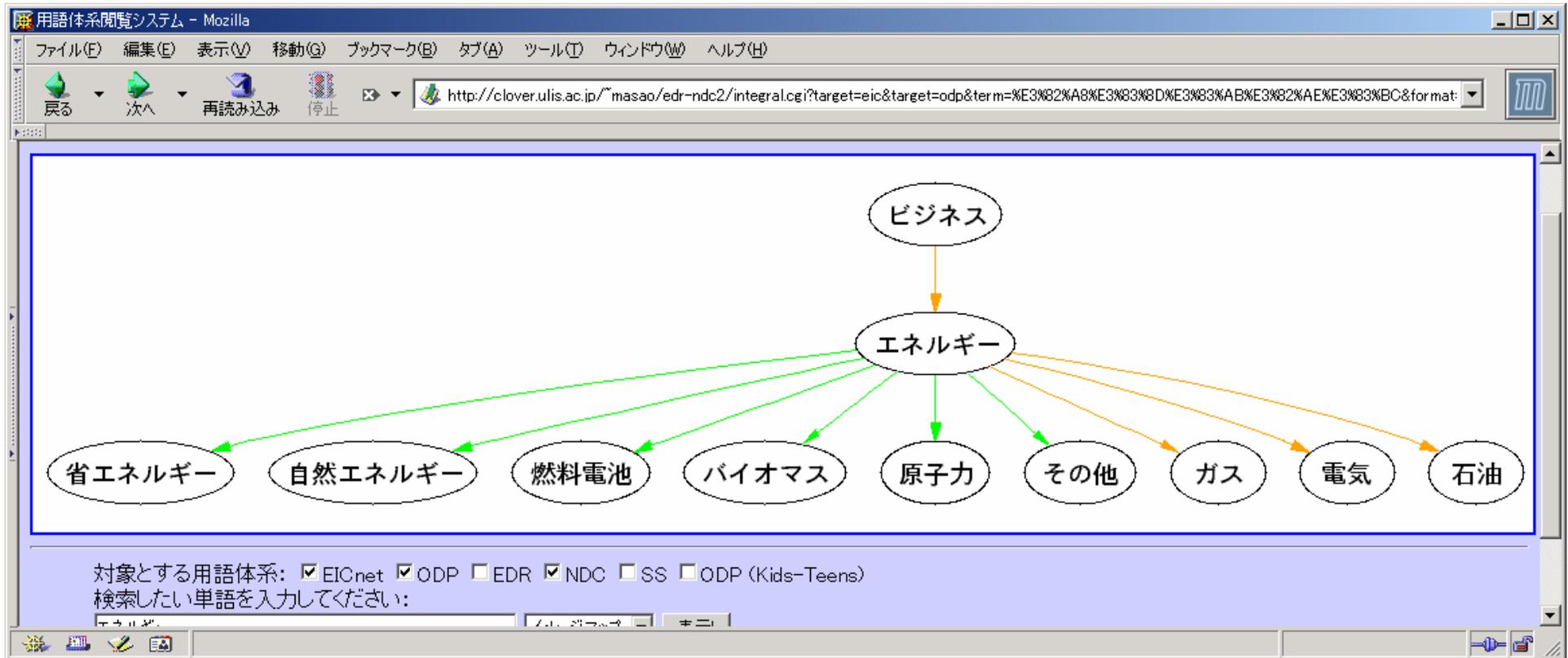
— NDC
— ODP
— EDR

様々な組み合わせによる重ね合わせ

- 重ね合わせる体系を自由に選択できる
- 重ね合わせの動機
 - 類似性: 複数体系同士の特徴(作成用途など)が似ている
 - 習熟度: ある用語体系に精通している
 - 特性: ある体系の特徴
- 例
 - EIC + ODP ⇒ 環境分野、全般的分野の体系
 - NDC + ODP ⇒ 全般的な分野の分類体系

活用例1

- エネルギー産業における「ガス」「電気」「石油」



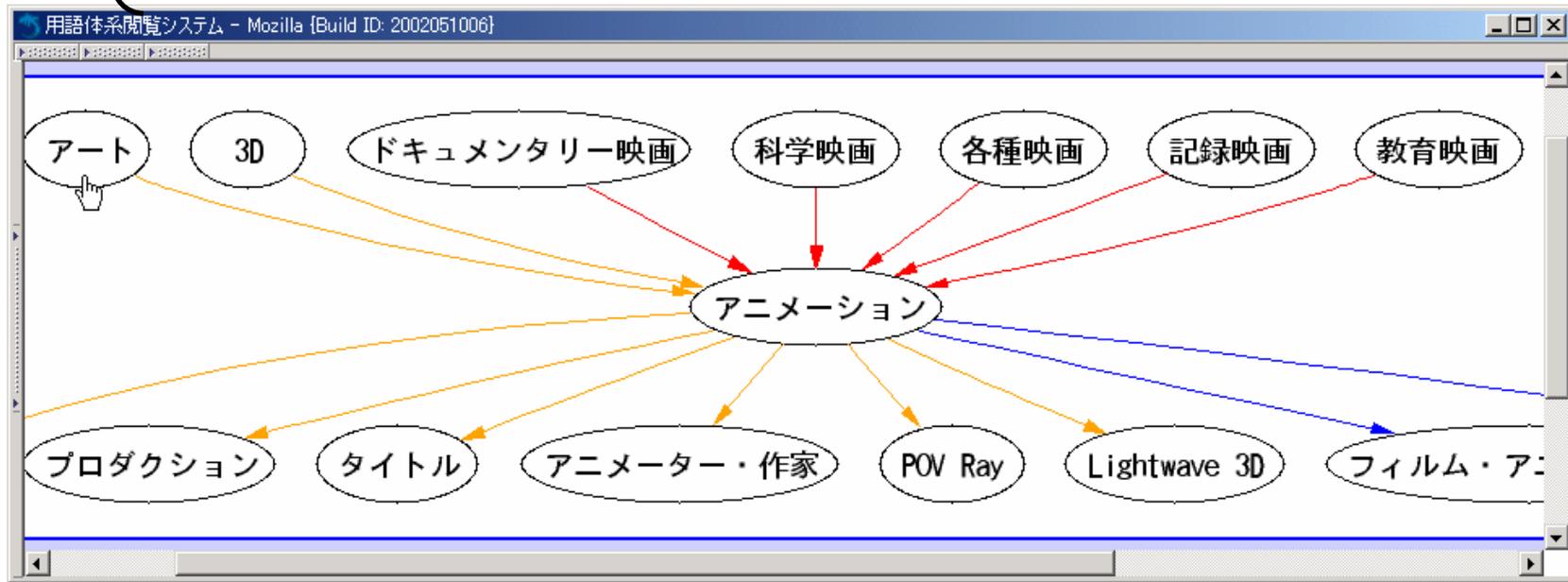
2004年9月1日

「エネルギー」

— ODP 28
— EICネット

活用例2

- 「映画」の中のアニメーション
- アニメーション「ソフトウェア」
- アニメーション「作品」 (タイトル、作家...)



2004年9月1日

「アニメーション」

— NDC
— ODP
— EDR

活用例2

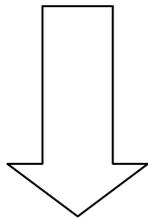
- 「アニメーション」分野のポータルサイト作者
 - 多くの体系間の観点を見ることで、「用語・分野」の観点を眺められる
 - ⇒ 語彙の収集・分析

目次

- 背景と目的
- 検索システム間の連携
 - ①JAPAN/MARC検索システム
 - ②Dublin Core検索システム
- 用語体系間の連携
 - ③複数用語体系ブラウジングシステム
- 異種情報提供システム間の連携
 - ④Webサービスによる用語体系提供システム
- まとめ

Webサービスによる 用語体系提供システム

- 用語体系の独立性
 - 様々な組織・個人が独自に構築・管理
- 他システムとの連携
- 再利用性



- Webサービス
 - ネットワーク上で独立して提供
 - 様々な利用(再利用、連携)

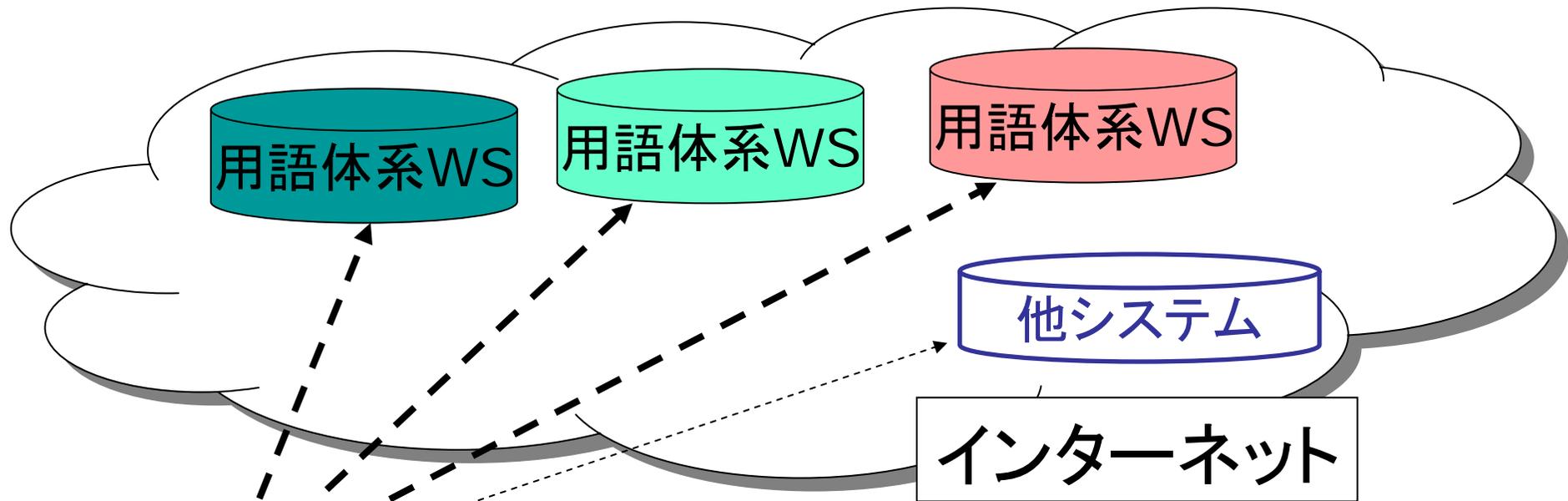
Web サービスとは

- Web+XMLによるRPC (Remote Procedure Call)
- 規格: SOAP+WSDL (+UDDI)
- サーバ{への/からの} {要求/返答}
→ すべてXML形式
- 既存の情報提供サービス
 - Google (サーチエンジン)
 - Amazon.com (書店)

本研究のアプローチ

- 独立して運営可能な提供システム
 - 一つの用語体系は一つのWebサービスに対応する
 - 体系の管理者が独立して提供可能
 - アクセス方式は統一
- 用語体系Webサービス
 - 見出し語の検索
 - 見出し語のもつ情報(構造)を提供
- 他システムとの連携
 - 用途に合わせたフロントエンドシステム
 - 例: 検索サービスとの連携など
 - 動作環境・プラットフォームを問わず、様々な利用が可能

Webサービスによる用語体系提供 イメージ図



フロントエンド
(利用システム)

2004年9月1日



用語体系提供Webサービスシステム

- 用語体系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
 - グラフ表示サービスとの連携例

用語体系提供Webサービス (TermService)

- 見出し語検索

情報検索

- 引数: 検索語

- 返値: 検索に一致した用語の一覧が返る

情報検索 (47888)
光学的情報検索 (30645)
選択的情報検索(55142)
文献情報検索言語(95639)
.....

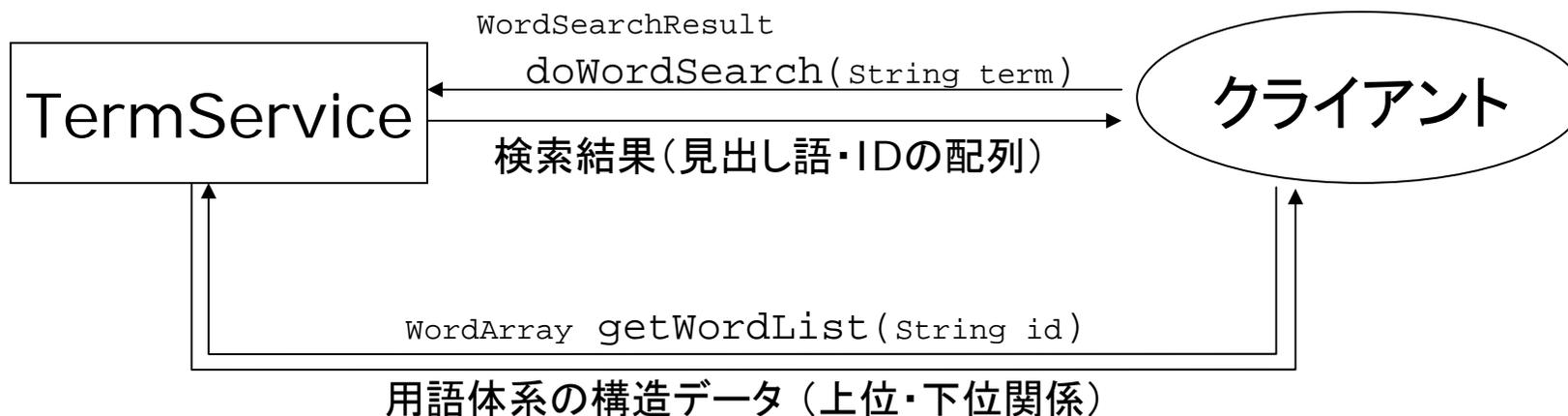
- 用語の情報(構造)を取得

- 引数: 用語ID

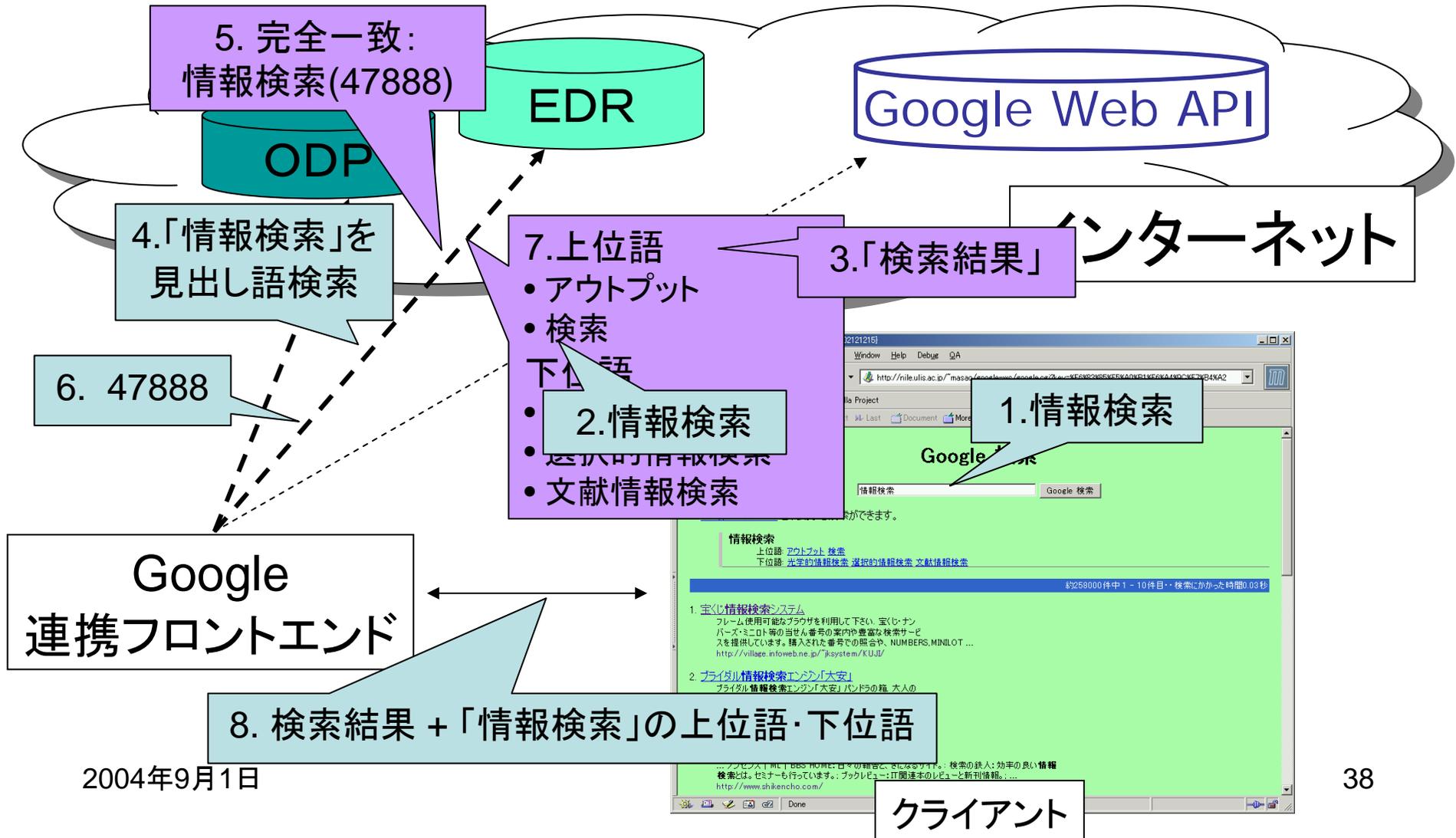
47888

- 返値: その用語のもつ関係を返す

上位語
• アウトプット
• 検索
下位語
• 光学的情報検索
• 選択的情報検索
• 文献情報検索



フロントエンド Google Web API 連携システム



2004年9月1日

まとめ

- 目的
 - 新たな情報提供システム構築の支援
 - 既存の情報提供システム同士の連携
- 検索システム間の相互連携
 - Z39.50プロトコルにもとづく検索システム
- 用語体系間の相互連携
 - 重ね合わせ機能を持つブラウジングシステム
 - 5つの分野・規模の異なる用語体系
- 異種情報提供システム間の相互連携
 - Webサービスにもとづく用語体系の提供
 - フロントエンドシステムの構築

まとめ(連携可能性)

		用語体系WS		検索システム			Graphviz WS
		EDR (WS)	ODP (WS)	Z39.50-JP	Z39.50-DC	Google WS※	
用語体系WS	EDR (WS)	-	◎	○	○	◎	◎
	ODP (WS)	◎	-	○	○	◎	◎
検索システム	Z39.50-JP	○	○	-	◎	○	○
	Z39.50-DC	○	○	◎	-	○	○
	Google WS※	◎	◎	○	○	-	○
	Graphviz WS	◎	◎	○	○	○	-

◎:システム構築を通じて、実証済

○:原理上可能(システム構築は行っていない)

※: Google社による提供