

## 学術情報サービスの メタデータ・デザイン

国立情報学研究所 (NII)  
大向 一輝

## 自己紹介：大向 一輝

- 国立情報学研究所 准教授
  - 次世代のウェブとコミュニケーション
    - セマンティックウェブ・ソーシャルウェブ
    - 「ウェブがわかる本」岩波ジュニア新書
    - 「セマンティックWebプログラミング」オライリー (監訳)
  - 学術情報流通
    - CiNii：論文情報ナビゲータ
- 株式会社グルコース 取締役
  - インタラクションデザイン・UIデザイン



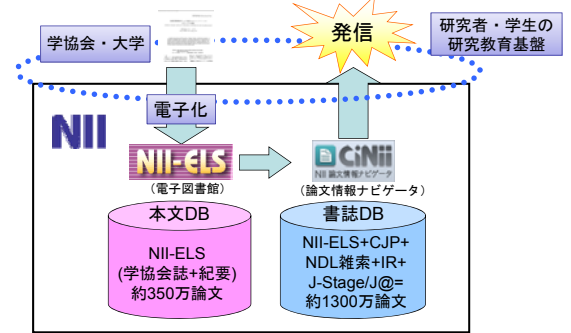
## 国立情報学研究所 (NII)

- 研究と事業の2本柱
- 学術情報インフラの整備機関
  - ネットワークサービス：SINET
  - コンテンツサービス
    - CiNii・Webcat Plus・KAKEN・JAIRO...



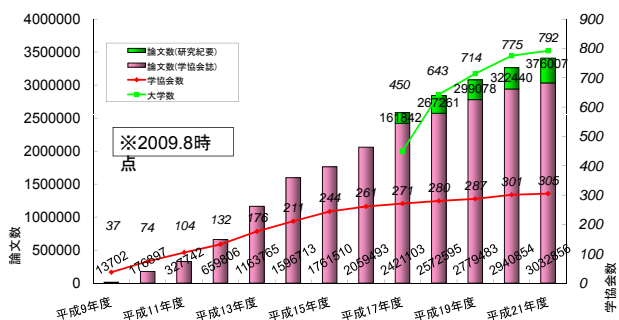
## CiNiiとは

- 国内最大規模の論文検索・提供サービス



## CiNii論文データベース

- 約350万論文のPDF提供

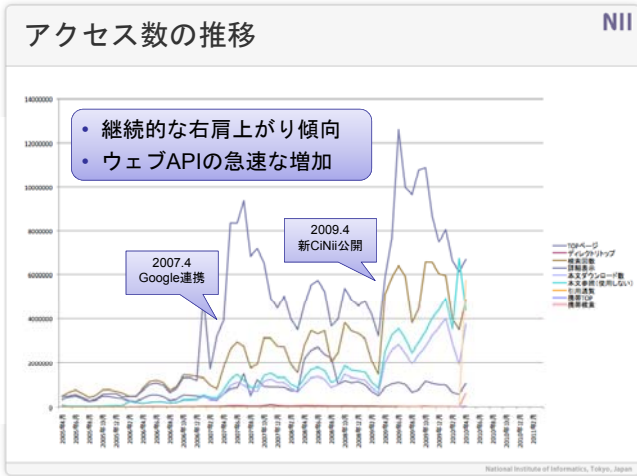


## CiNii書誌データベース

- 自動・手動処理を組み合わせず書誌同定・統合

データベース名	データ数	年間増加数	更新頻度	本文	料金
NII-ELS学協会刊行物	約350万件	約18万件	週次	○	一部有料
NII-ELS研究紀要	約87万件	約3.5万件	週次	△	無料
引用文献索引データベース	書誌:約154万件 引用:約1661万件	書誌:約14万件 引用:約161万件	10回/年	×	無料*1
雑誌記事索引データベース	約827万件	約40万件	週次	×	無料
機関リポジトリ	約30万件	不定	週次	○	無料
J-Stage/Journal@rchive	約8万件	不定	数回/年	○	無料
<b>CiNii合計*2</b>	<b>約1239万件</b>	<b>約70万件</b>	<b>週次</b>		

\*1 参考文献/被引用文献の閲覧は制限あり。  
\*2 重複データが統合されるため、単純合計とは一致しない。

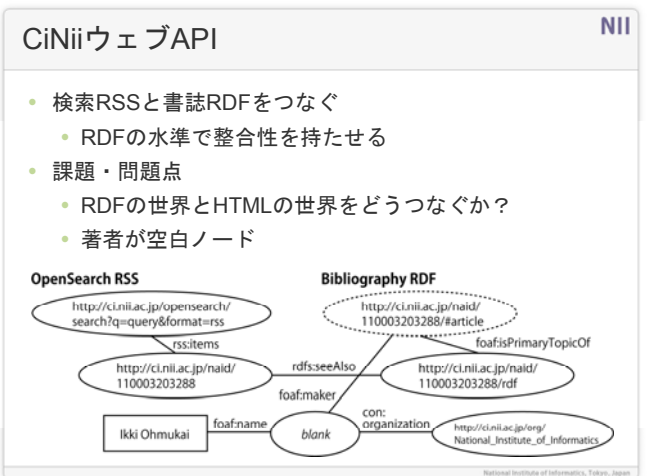


- ### セマンティックウェブの意義
- ウェブ情報の構造化・組織化
    - コンピュータにとっての意味の曖昧性を除去し、自動処理の可能性を拡大する
      - 横断検索・マッシュアップ
      - 推論・人工知能
  - 実現方法
    - ひとつのものごと・概念に対応するIDを与える
      - URI・パーマリンク
    - 構造化フォーマットで表現する
      - XML / RDF
    - 表現内容はURI同士の関係性
- National Institute of Informatics, Tokyo, Japan

- ### 学術情報とセマンティックウェブ
- 学術情報分野の特徴
    - 情報の構造化を生業とする職業集団・組織がある
      - 研究者 (大学・研究機関)・学会・図書館...
    - 情報の構造化フォーマットが共有されている
      - タイトル・著者名・抄録・本文・参考文献...
    - フォーマットを変換するだけでセマンティックウェブ対応に
- Weblogの現在と展望: セマンティックWebおよびソーシャルネットワークの基盤として  
The State-of-Art and Prospects of Weblog: An Infrastructure of Semantic Web and Social Networking
- 著者: 武田 英樹 (TAKEDA Hideaki) / 大島 一博 (OHMUKAI Ikki)  
所属: 国立情報学研究所 (National Institute of Informatics) / 総合研究大学院大学 (National Institute of Informatics)
- ▼ 参考文献: 21件 ▼ 被引用文献: 15件
- National Institute of Informatics, Tokyo, Japan

- ### CiNiiのセマンティックウェブ対応
- 設計方針をどうするか?
    - データ構造・語彙...
    - 厳密性と利便性のトレードオフ
    - 既存? 独自?
  - ウェブAPIとして考える
    - できるだけ開発者が使いやすいように
      - シンプルなデータ構造
      - ライブラリの普及状況を念頭に
      - 世界標準に準拠
    - どういう用途に使ってもらいたいのか
      - サービスの主要な機能に対応
- National Institute of Informatics, Tokyo, Japan

- ### CiNiiウェブAPI
- 検索: OpenSearch
    - RSS 1.0 (=RDF) と Atom 1.0
      - RSS/Atom基本語彙+Dublin Core (Legacy)
        - 一般のブラウザ・フィードリーダーで扱えるように
      - Prism Standard
        - 書誌情報は国際標準に準拠 (Natureなど)
  - 書誌: RDF
    - Dublin Core+Prism Standard (書誌情報)
      - XML時代の語彙はリテラルを値に持つように
    - FOAF (著者情報)
      - RDFならではの表現はRDF時代の語彙で
  - 独自語彙は使用しない
    - はじめての試みだったため
- National Institute of Informatics, Tokyo, Japan



## 学術情報とセマンティックウェブ

NII

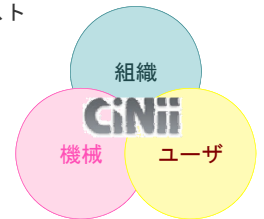
- 構造化されてこなかった情報への対応
  - 既存データに手をつけるコスト
    - 分量・ワークフロー
  - 完全性の保証
- 代表的な例：著者ID
  - 高まる重要性
    - 個人の業績管理
    - 国際競争 (ResearcherID・ORCID)
  - 著者名典拠がない
    - 論文の著者名は膨大かつロングテール
    - 同姓同名・旧姓・タイプミス...

National Institute of Informatics, Tokyo, Japan

## CiNii著者検索

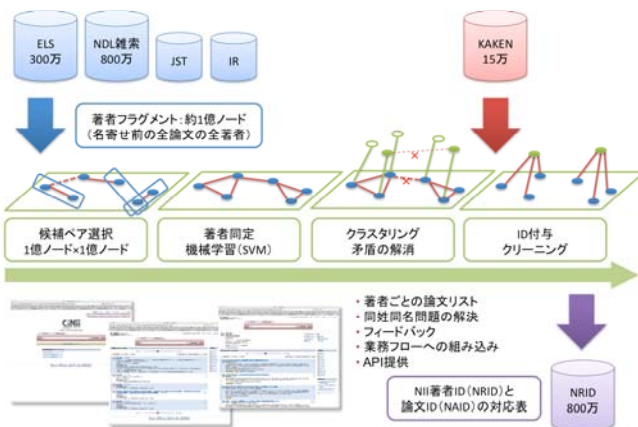
NII

- NII著者ID (NRID) の導入
  - 科研費番号+機械処理による著者へのID付与
  - 著者ごとにURIを発行 (著者パーマリンク)
- NRIDベースの論文検索機能
  - 著者名→IDリスト→論文リスト
  - APIの提供
- 新たなデータ生成・管理モデル
  - 研究成果の活用
  - ユーザーフィードバック



National Institute of Informatics, Tokyo, Japan

## CiNii著者検索の概要



## CiNii著者検索

NII

- ALS (Author Linking System)
  - i-Linkage (NII相澤教授)の大規模・実運用システム
  - CPU32コア・メモリ320GB・計算時間5日 (全件処理)
- フィードバック (同一人物の報告)
  - 機械処理だけで100%の精度を得ることは不可能
  - あらかじめフィードバックを織り込んだシステム・アルゴリズム設計
    - 例：過統合より未統合を指摘する方が簡単
- 実績：6217件 (4月1日～7月15日)
  - Researchmap経由で研究者本人からのフィードバックも可能に

National Institute of Informatics, Tokyo, Japan

## CiNiiとセマンティックウェブ

NII

- 他の情報源とのリンク
  - NIIのサービス：KAKEN・研究者リゾルパー
  - 外部のサービス：Wikipedia / DBpedia・Freebase...
- 今度の展望
  - 各種情報のID・パーマリンク化
    - 組織名・キーワード
  - リンクトデータの一部に
    - Web NDLISH
    - 「ことはぶ」

National Institute of Informatics, Tokyo, Japan

## 図書館とセマンティックウェブ

NII

- 図書館の役割
  - アーカイブ+構造化・組織化
  - 組織・機械・ユーザーの三位一体によるセマンティックウェブ環境の実現
- インセンティブのデザイン
  - 誰が何のために情報を構造化するのか？
    - 著者検索は著者IDのためのきっかけ
  - システムの外部にあるインセンティブ
    - ユーザー参加型サービスとの連携

National Institute of Informatics, Tokyo, Japan

NII

Thank you!

**CI*N*ii**  
<http://ci.nii.ac.jp/>

National Institute of Informatics, Tokyo, Japan