

情報科学を図書館の現場で 実践する !?

2010年10月28日

高久雅生

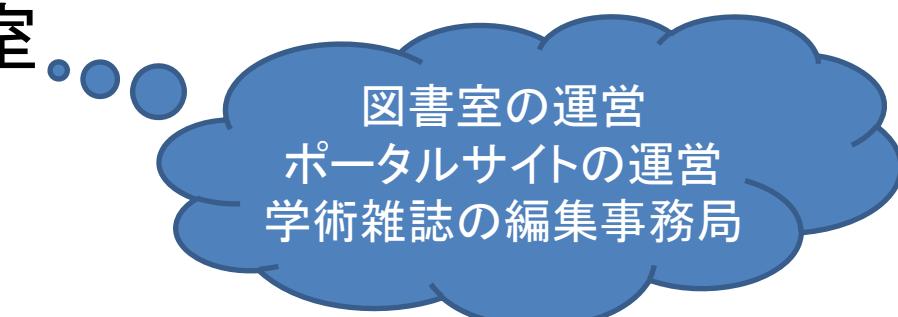
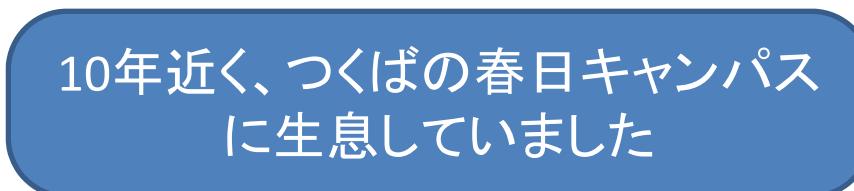
物質・材料研究機構 科学情報室

TAKAKU.Masao@nims.go.jp

本日のお品書き

- 自己紹介
 - これまでのキャリア
 - 学部生、大学院生 → ポストドク研究員 → 専門図書館運営
- いまやっていること
 - 日常 + 企画・開発
 - ウェブサービス
 - 機関リポジトリソフトウェア eSciDoc - 海外との共同開発
 - 研究者総覧サービス
 - 図書館利用者向けサービス、業務サービス
 - 研究的に
 - ウェブサービス: ふわっと検索
 - 被験者実験: CRES共同研究
- まとめ
- おまけ: 学生時代にやっておいてよかったこと

自己紹介

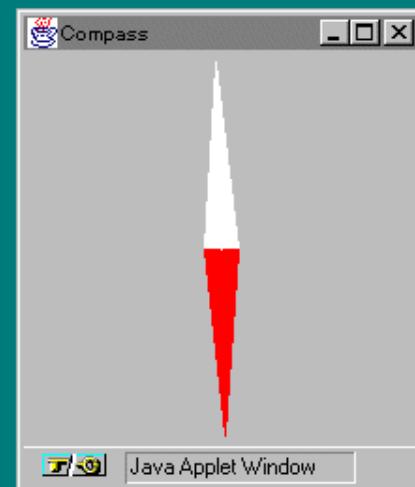
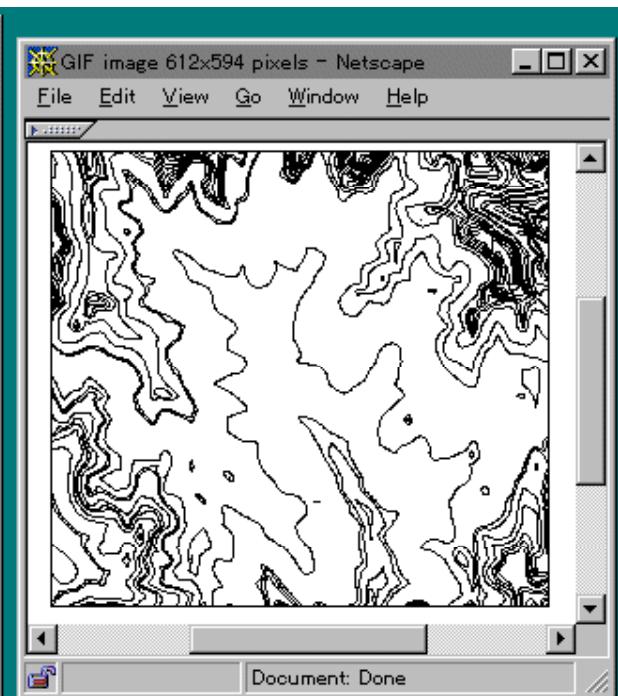
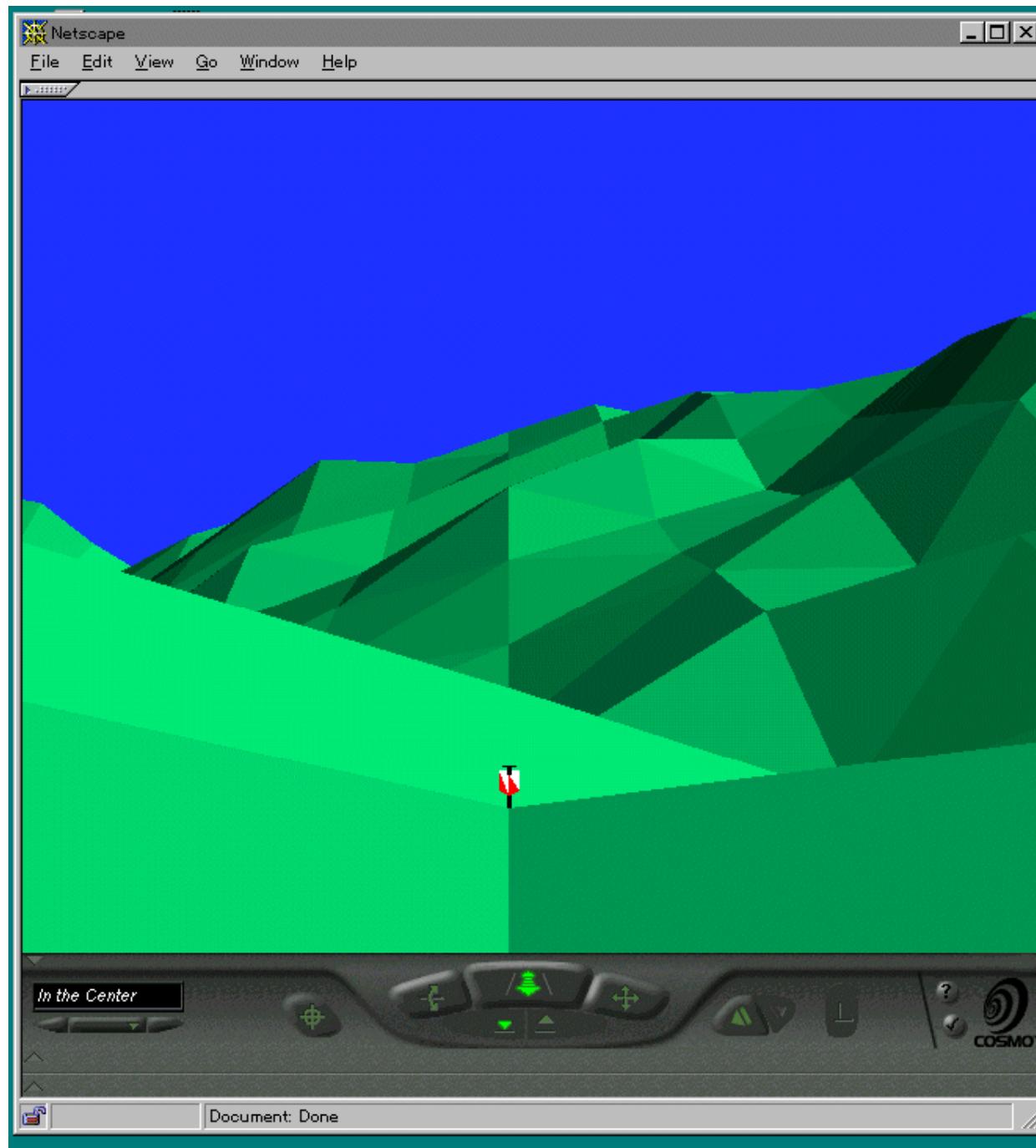
- 高久 雅生 (たかくまさお)
 - 所属: 独立行政法人 物質・材料研究機構 (NIMS)
 - 部署: 企画部 科学情報室
 - 職位: 主任エンジニア
 - 経歴:
 - 1994年4月 図書館情報大学 図書館情報学部 入学
 - 2004年9月 筑波大学図書館情報メディア研究科博士後期課程 修了
 - 博士(情報学)
- 
- 

これまでのキャリア(学生時代)

- 学部生
 - 入学当初: WWWの黎明期: Mosaic世代?(Mosaic, Netscape, ...)
 - 日本語のウェブページがすべて見れた時代!
 - 卒研: オリエンテーリングをウェブで実現
 - 標高付き地図データ+3次元リアリティ
- 院生
 - 学内ページを対象にした検索エンジン(ULIS-Search)
 - 検索エンジンNamazu開発プロジェクトに参加
 - オープンソースソフトウェア
 - サイト内検索として当時もっともよく使われていたツール
 - 大規模書誌データ検索システム構築(JAPAN/MARC; 約100万件)
 - 博士論文: 「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」



最初期のブラウザ Mosaic



スタート

Paint Shop Pro - exempl...

GIF image 612x594 pixel...

Netscape

Compass

A CAPS

5:00

これまでのキャリア(就職後)

- ポスドク研究員(国立情報学研究所)
 - NTCIR-5 WEB: 日本のウェブページ収集と検索システム構築
 - 1億ページ(1.5TB)、データ処理
 - KAKEN: プロトタイプシステムの構築
 - 15万人の日本の大学・研究機関研究者、マイニング
 - <http://kaken.nii.ac.jp>
- エンジニア(2008年9月～現在)
 - 所内向けウェブサービスの運営
 - 専門図書室の運営



いま、何をやっているのか？

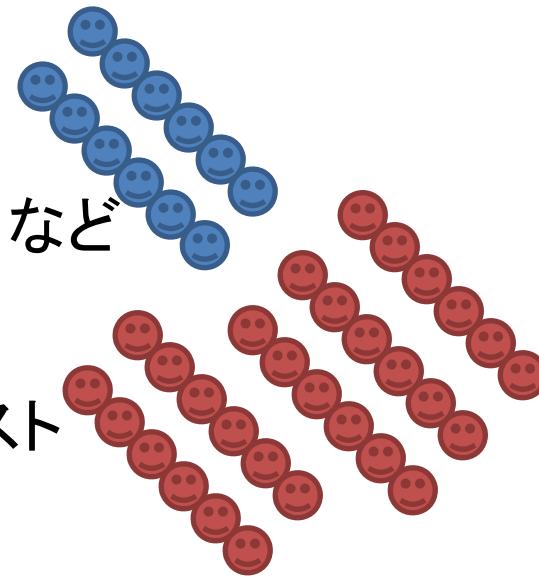
- 物質・材料研究機構
 - (材料研究: 鉄鋼材、ガラス、セラミックス、超電導、生体材料など)
 - 約1,200名の研究員(非常勤、外来、研修生なども含め)
- 主任エンジニア
- 所内向けウェブサービスの開発
 - 機関リポジトリ: NIMS eSciDoc
 - 研究者総覧サービス
- 図書室の運営
 - 電子リソースのサポート対応

機関リポジトリ

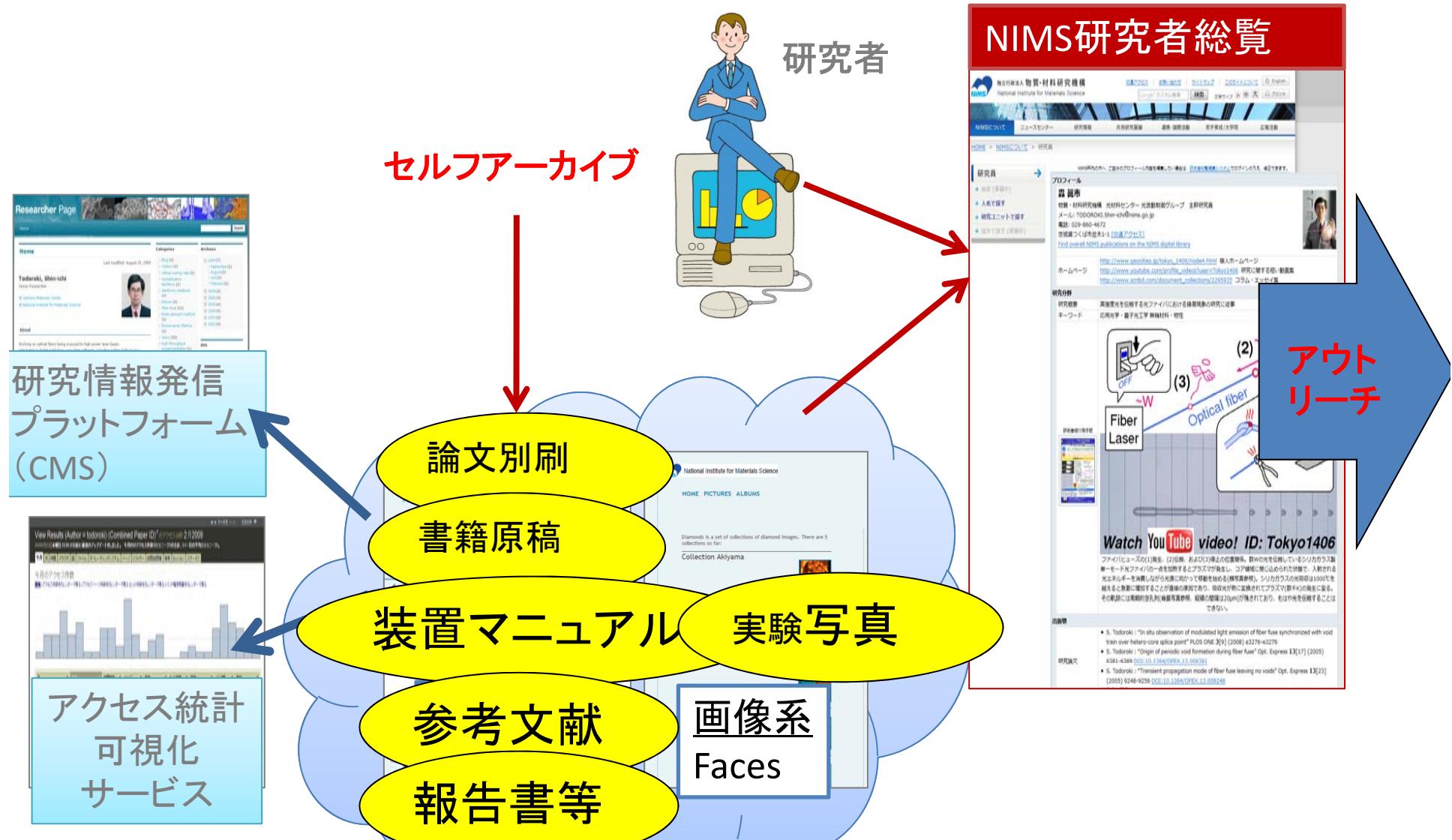
- 所内のひとが書いた論文、書籍などを保存し、公開・発信するプラットフォーム
 - セルフアーカイブ
- 開発しているシステム : NIMS eSciDoc
 - Max-Planck Digital Library (MPDL)との共同開発
- 私がやっていること
 - (システム開発) 例えば、日本語対応機能のやり取り
 - インタフェース言語の追加
 - 日本語データ処理の追加
 - 検索機能: 日本語分かち書き
 - 著者名表記
 - (運営) NIMS研究者アカウントの管理、論文の一括登録
 - MPDLとの相互訪問や情報交換

共同開発の実態

- eSciDocプロジェクト
 - FIZ: コアサービス
 - 認証、データストア、検索まわりなど
 - MPDL: アプリケーション
 - 仕様策定、実装、GUI設計、テスト
- NIMS: 開発&運用 
 - 日本で必要な機能の追加実装 = 日本語化(ローカライズ)
 - NIMSで必要な新機能提案 = 仕様設計



NIMS eSciDocの全体像

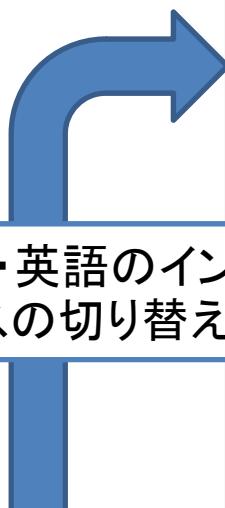


一体的な相互連携・相互リンクからセルフアピールまで

2010年10月28日

12

日本語・英語のインターフェースの切り替え



e SciDoc.PubMan
PROJECT

ホーム 一時保存 (0)

PubMan 6.1.12 - eSciDoc文献管理ソリューション

研究成果文献と関連資料を半永久的に保存提供できます。
研究者個人の“研究者ポートフォリオ”として論文情報を一箇所にまとめておくことができます。
マックスプランク研究所が提供する長期保存アーカイブのサービスを存分にお使いください。

[詳しく知るには | サポート](#)

文献検索

Login Contact us Policy&Disclaimer Help

English

新着アイテム一覧

The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects
by Meyer, Antje S. - 3 時間前
The goal of the study was to examine whether speakers naming pairs of objects would retrieve the names of the objects in ...

Electrophysiological evidence of semantic interference in visual search
3 時間前
were monitored while a target (e.g., bird) in a ...

... es: Effects of frontal lobe distraction
3 時間前
out visual search, distractors tested, rather than unrelated, to ...

d SMS shortcuts in sentence investigation
前
Participants read SMS shortcuts and ...

Start

Last Released Items

The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects
by Meyer, Antje S. - 3 時間前
The goal of the study was to examine whether speakers naming pairs of objects would retrieve the names of the objects in ...

Electrophysiological evidence of semantic interference in visual search
by Humphreys, Glyn W. - 3 時間前
Visual evoked responses were monitored while participants searched for a target (e.g., bird) in a four-object display that ...

Distracted by relatives? Effects of frontal lobe ...
9月13日
those who are working with ...

with PubMan Release 6.1
9月13日
those who are working with ...

日本語著者名表記の問題

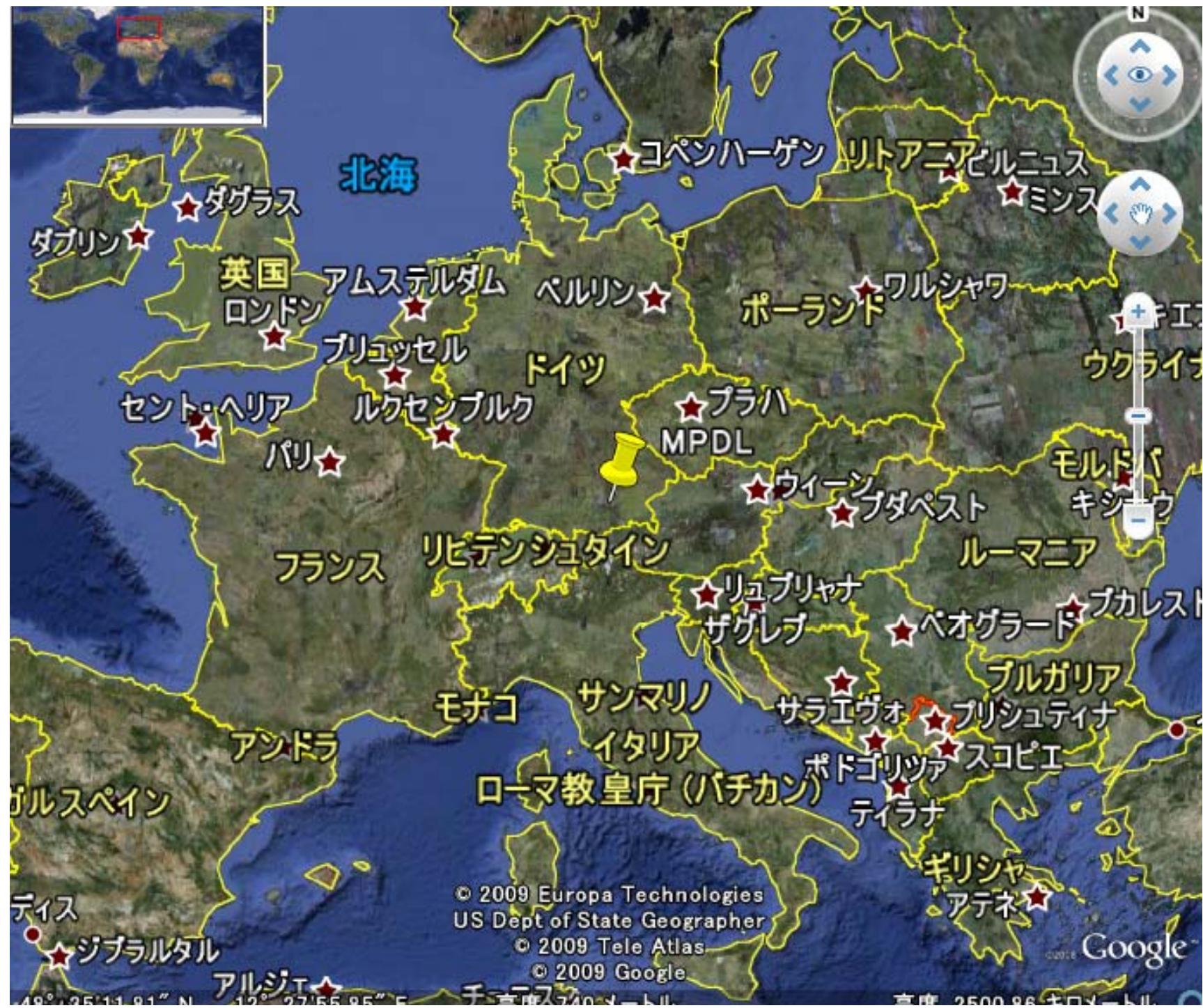
- 文献リストの中で:
 - 小西, 智., & 轟, 眞. (2010). BibTeX文献リストを用いた研究成果発信・管理支援システム ...
 - → イニシャルになってしまっている。欧米文化がそのまま適用されてしまった例。
 - 本当は: 「小西 智也 & 轟 真市 (2010)...」とすべき...。
- 私がやっていること:
 - 問題の発見、報告、修正(そのための議論)
 - ネイティブでないとわからない問題はいくつもある。

The screenshot shows a researcher profile page from eSciDoc.CoNE. The main title is '研究者ポートフォリオ' (Researcher Portfolio). Below it, the name '轟 真市' is displayed, followed by '物質・材料研究機構 光材料センター'. A small portrait photo of the researcher is shown. The page includes sections for '研究者プロフィール' (Researcher Profile) and '研究分野' (Research Fields), listing fields like Computer Sciences, Engineering, and Physics. At the bottom, there are links for '外部参照' (External References) and 'WorldCat Search for 轟 真市'.

皮大学 オムニバス講義

MPDLとの共同開発 打ち合わせ、情報交換

- 相互訪問
 - MPDL 訪日: 5回
 - 2008年7月, 2008年11月, 2009年3月
 - 2009年9月
 - 2010年7月
 - NIMS訪独: 5回
 - 2008年9月(2週間), 2009年1月(1週間)
 - 2009年6月(1週間), 2010年3月(1週間)
 - 2010年11月(2週間)
- ビデオ会議: 年に6回程度







MPDLエントランス



MPDLミーティングルーム



2010年10月28日

筑波大学 オムニバス講義

21

MPDL風景



2010年10月28日

筑波大学オムニバス講義

開発チームの部屋2

22

MPDL風景



もちろん、ミュンヘンと言えば...



Octoberfest 2008



2010年10月28日

筑波大学 オムニバス講義

26

研究者総覧サービス

- 研究所にどんな研究員がいるか、その情報を公開して、ひろく共同研究相手を探したりなどしてもらうための公開サービス
- NIMSでは、新しい研究者総覧サービスを機関リポジトリと連携して公開するための準備を進めてきた
 - 2010年10月に、研究者総覧サービスを公開。
 - 常勤研究職員 約700名を対象。
 - <http://samurai.nims.go.jp/>



NIMSについて

ニュースセンター

研究情報

共用研究基盤

連携・国際活動

若手育成/大学院

広報活動

HOME > NIMSについて > 研究者 > 研究ユニットで探す > 国際ナノアーキテクtonics研究拠点(MANA)

研究者 →

国際ナノアーキテクtonics研究拠点(MANA)

→ 検索 (準備中)

→ 人名で探す

→ 研究ユニットで探す

→ 論文で探す (準備中)

氏名	研究ユニット	キーワード
 青野 正和	国際ナノアーキテクtonics研究拠点(MANA)	薄膜・表面界面物性 重子デバイス・電子機器 無機材料・物性
 青柳 隆夫	領域コーディネーター	刺激応答性材料、DDS、再生医学
 荒川 秀雄	国際ナノアーキ研究拠点(MANA) ナノシステム分野ナノシステム構築グループ	
 荒船 竜一	国際ナノアーキテクtonics研究拠点(MANA)	
 有賀 克彦	国際ナノアーキテクtonics研究拠点(MANA)	ナノ材料・ナノバイオサイエンス
 アシュラフルイ スラム	次世代太陽電池センター	Organic/inorganic interface, Photochemistry and Photovoltaics
 魁崎 浩平	国際ナノアーキ研究拠点(MANA)	



研究者



- 検索 (準備中)
- 人名で探す
- 研究ユニットで探す
- 論文で探す (準備中)

プロフィール

小林一昭 (コバヤシ カズアキ)

物質・材料研究機構 計算科学センター 第一原理反応グループ 主幹研究員

メール: KOBAYASHI.Kazuaki@nims.go.jp

電話: 029-860-4511

茨城県つくば市並木1-1 [\[交通アクセス\]](#)



ホームページ	http://www.nims.go.jp/cm_sc/staff/kobayak/NCPS/ncps.html 擬ポтенシャルデータベースNCPS2K紹介ページ http://www.nims.go.jp/cm_sc/staff/kobayak/BN/bne.html 筆者の研究対象であるBN関連 (AIN,SiCを含む) ページ http://www.nims.go.jp/cm_sc/staff/kobayak/INFO/metal.html 筆者が雑誌「金属」に2009年10月～2010年9月に連載した記事の支援ページ
--------	---

研究分野

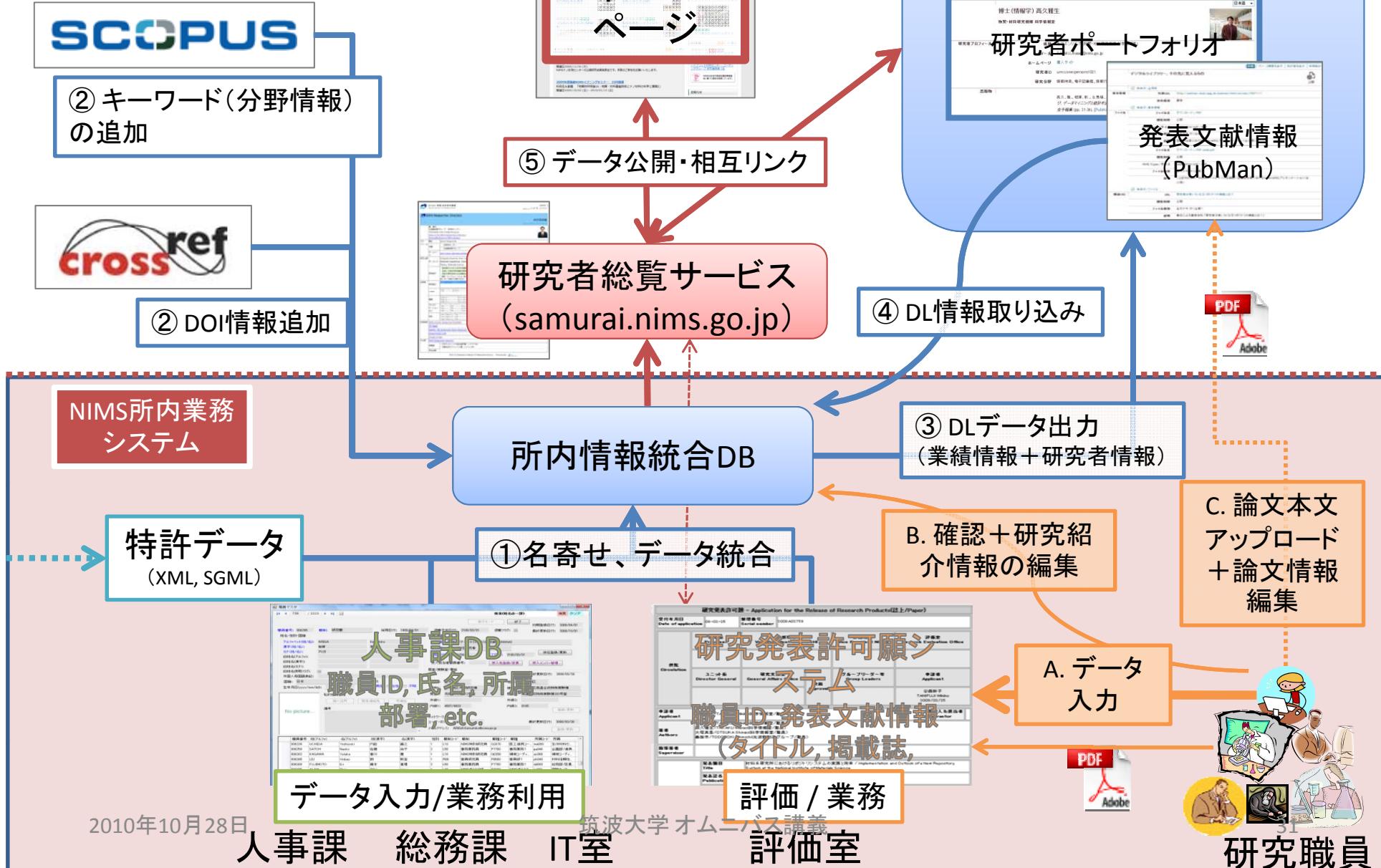
研究概要	I calculate the BN, SiC, AlN polytypes using the electronic structure calculation (DFT-LDA). I have investigated their electronic and structural properties. Calculated polytypes are 2H - 6H, 10H, 12H and 30H. I have already calculated 30H-BN.
キーワード	第一原理、バンド計算、未知物質探索（物質設計）、BN、AlN、SiC、ポリタイプ、層状構造、電子状態、電子構造、バンド構造、ワイドバンドギャップ、カーバリエロ法、擬ポテンシャル、平面波基底、密度汎関数法、局所密度近似、擬ポテンシャルデータベース、NCPS、NCPS2K



研究者総覧サービス

- 私がやっていること
 - (システムの設計、開発、運営)
 - データフローの構築
 - 人事情報、研究業績情報、機関リポジトリNIMS eSciDoc
 - 所内各部署とのやり取り、調整
 - データの整備
 - システムの開発
 - 書誌データ、人物データの修正など

研究者総覧サービスにおけるデータフロー



システム開発

- 人事情報データベースの情報を整理して表示
 - ユニット単位データ、人名インデックスを元に一覧表示
 - 日本語版・英語版
 - 研究ユニットごとにリクエストに応じたカスタマイズ
 - ポスドク研究員も一覧に加えてほしい
 - 研究者を限定して表示してほしい
 - 一部職種は表示しないように、外部からきているひとは表示しないように、など。
- 外部データベースとのやり取り
 - トムソン・ロイター社 ResearcherID.com サービスや国立情報学研究所KAKENサービスへの一括登録機能の開発

人事データ・論文書誌データの修正、確認など

- 論文DOI情報の取得
 - ウェブAPI : crossref.org へのクエリ、照合
- 著者名部分の情報の正確性
 - あとから修正できるようにする必要あり。
- 職位、部署名など、人事情報の正確さを確認
 - 各部署とやり取り、調整
- NIMS eSciDocとのデータ連携統合

図書館サービス

- 専門図書室の運営
 - 物質・材料分野の専門書、専門雑誌
 - 研究情報の大部分はEジャーナル、Eブックに移行
 - おもに3拠点(千現、並木、桜)
 - 総蔵書数16,000冊程度
- 図書担当スタッフ: 2.5名 (?)
 - 業務は大部分無人化: 貸出・返却はセルフなど
 - 利用者対応に加え、所蔵・購読していない論文を外部から複写・貸出してくる業務がメイン
- わたしがやっていること
 - 電子リソース担当
 - 論文データベース/Eジャーナル/リンクリゾルバー(SFX)周りの不具合・サポート対応
 - 図書室端末PC(千現2台、並木1台) の管理、サポート
 - (図書貸出返却システム、OPACシステムの開発)

NIMS図書室



(研究者として)

- ふわっと関連検索ツール
 - 文書検索のサービス
 - <http://fuwat.to>
- CRES共同研究
 - ウェブ探索行動を被験者を使って実証研究



「ふわっと関連検索」

- 文書類似度に基づく文書検索
 - 検索キーワードを考えなくともよい
 - 文書特徴量をそのまま利用できる
- 文書類似度の計算エンジンを積んでいない普通のキーワード検索だけの検索ツールで文書検索したい！
- 一回の検索要求に対して、キーワード検索を複数回/複数パターン実行

利用サイト公開中
<http://fuwat.to>



列車運行実績データの可視化

検索

[検索オプション](#)

ウェブ全体から検索 日本語のページを検索

[カスタマイズ設定を表示](#)

ウェブ [検索ツールを表示](#)

列車運行実績データの可視化 の検索結果 約 2,750 件中 1 - 10 件目 (0.18 秒)

[PDF] [列車運行実績データの可視化](#)

ファイルタイプ: PDF/Adobe Acrobat - [クリックビュー](#)

列車運行実績データの可視化を検討するに際しては、ユーザにとまどいを与えないように、これまでに使用されて 可視化手法が有効に用いられうることを確認することがで、きた。6 おわりに、列車運行実績データの可視化手法を提案し、それらが列 ...

www.tomii.cs.it-chiba.ac.jp/kashikaJRAIL.pdf -

[Twitter / hidetan: 列車運行実績データの可視化](#) <http://www.t...>

Check out m.twitter.com! Skip to navigation; Skip to sign in form. Loader. Twitter.com · Login · Join Twitter! 列車運行実績データの可視化 <http://www.tomii.cs.it-chiba.ac.jp/kashikaJRAIL.pdf> #nhk about 2 hours ago from web ...

twitter.com/hidetan/status/8542455794 - [キャッシュ](#) -

[Twitter / shiyu: 牛田さん共著論文「列車運行実績データの可視化 ...](#)

牛田さん共著論文「列車運行実績データの可視化」可視化の手法がクロマティックダイヤグラム <http://www.tomii.cs.it-chiba.ac.jp/kashikaJRAIL.pdf> #nhk.

twitter.com/shiyu/status/8542431922 - [キャッシュ](#) -

[TweetBuzz - kashika](#)

<http://www.google.co.jp/search?q=%E5%88%97%E8%BB%8A%E9%81%8B%E8%A1%8C%E5%AE%9F%E7%B8%BE%E3%83%87%E3%83%BC%E3%82%BF%E3%81%AE%E5%8F%AF%E8%A6%96%E5%8C%96>

tweetbuzz.jp/entry/9262701/www.tomii.cs.../kashikaJRAIL.pdf - [キャッシュ](#) -

スポンサーリンク

[データ可視化で競争に勝つ](#)

費用対効果・俊敏性・効率の向上!
ビジネスインテリジェンスで経営改革
spotfire.tibco.jp

[富士通可視化事例](#)

IT投資による改善効果をご紹介。
木村佳乃がコスト削減をナビゲート
jad.fujitsu.com

[DICOMデータなどを可視化](#)

数値だけでは理解、把握できない
3次元データを視覚化するソフトウェア
www.lightstone.co.jp/slicerdicer

[広告掲載はこちら »](#)

ふわっと CiNii 関連検索

文章から検索

ウェブページから検索

<http://www.tomii.cs.it-chiba.ac.jp/kashikaJRAIL.pdf>

論文検索

列車運行遅延ダイヤ Ci 295件中 1- 20件目 検索にかかった時間2.85秒

1. [第93回運輸政策コロキウム 都市鉄道の運行ダイヤ過密化に伴う列車遅延の波及に関する研究](#)
仮屋崎,圭司; 岩倉,成志; 森地,茂
運輸政策研究, 2009/Win.
2. [スイス連邦鉄道における接続を重視した新しい運行管理手法：戦略的施策から実際の運営の場に至るまでの余裕時分の活用手法\(<特集>鉄道のスケジューリング問題\)](#)
Laube,Felix; Luthi,Marco; 富井,規雄
オペレーションズ・リサーチ：経営の科学, 2008-08-01
3. [スプレッドシートを用いた協調推論型知識調整方式：列車運行予測の高精度化へ向けて](#)
江口,俊宏; 鶴田,節夫
全国大会講演論文集, 1991-02-25
4. [列車運行予測のための協調推論型知識調整方式](#)
鶴田,節夫; 江口,俊宏; 松本,邦顕
全国大会講演論文集, 1990-09-04
5. [論文部会報告 第46回鉄道サイバネ・シンポジウムの論文を査読して\(列車運行システム、新しい交通システム、情報伝送・処理システム\)](#)
平栗,滋人
サイバネティクス, 2010/1
6. [近鉄名古屋地区列車運行遅延対策](#)
高道,和博
鉄道と電気技術, 2010/1

<http://fuwat.to/cinii?url=http%3A%2F%2Fwww.tomii.cs.it-chiba.ac.jp%2FkashikaJRAIL.pdf>

「ふわっとCiNii関連検索」

最終更新日：2010年02月23日
[\(更新履歴\)](#)

ふわっと CiNii 関連検索

[文章から検索](#)

[ウェブページから検索](#)

[論文検索](#)

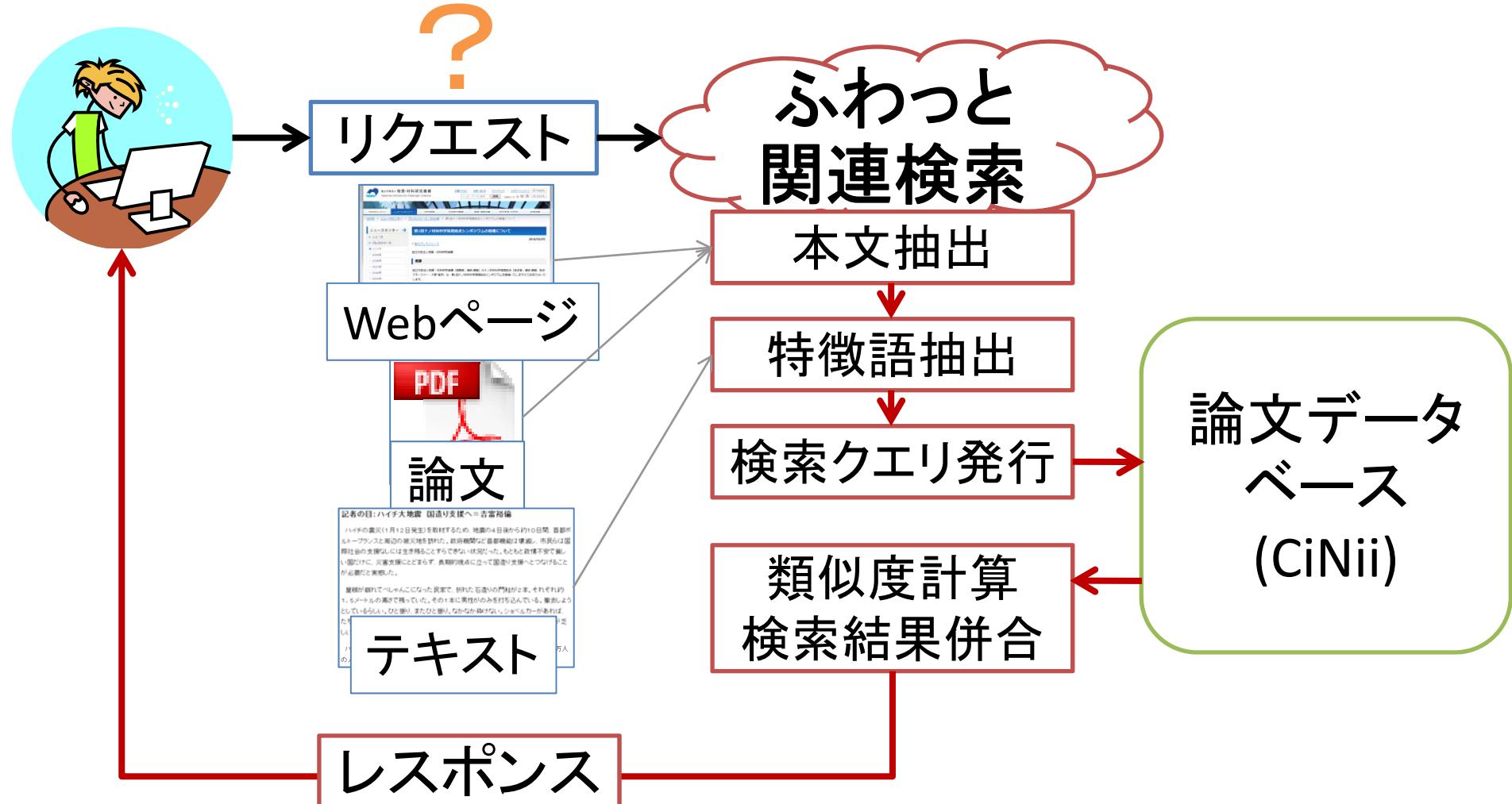
入力したテキストまたはウェブページに関連した論文をCiNiiで検索します。長いテキストやURLで指定したページからでも関連キーワードを自動的に抜き出して論文検索できるのが特徴です。

例: [朝日新聞社説](#) (元記事(asahi.com)), [日本経済新聞社説](#) (元記事(nikkei.co.jp)), [毎日新聞「記者の目」](#) (元記事(mainichi.jp))

<http://fuwat.to/cinii>

高久 雅生 (Masao Takaku)
<http://masao.jpn.org/>, tmasao@acm.org

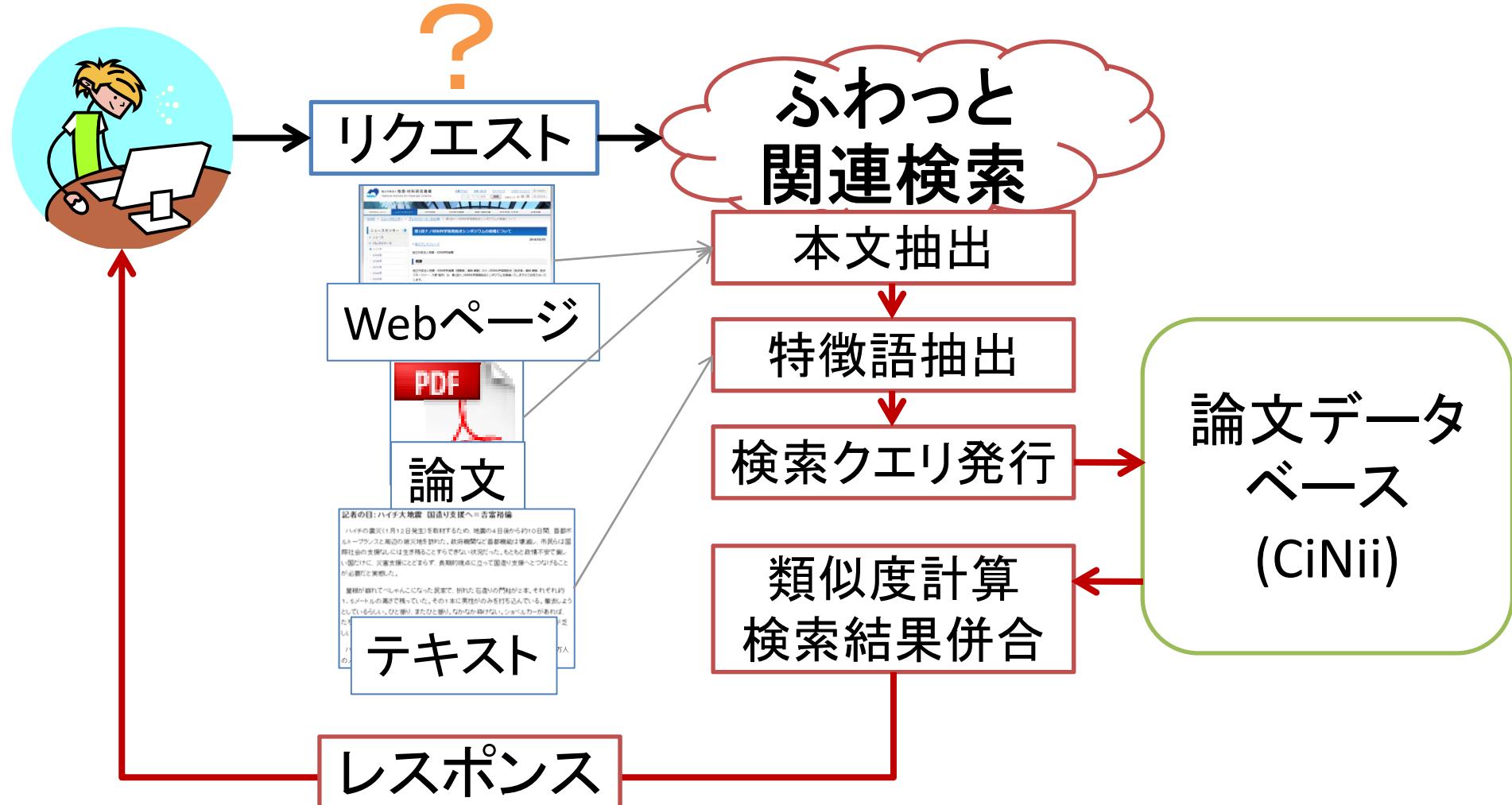
「ふわっと関連検索」システムの概要



実装: 本文抽出

- 類似検索の対象文書: Webページ(HTML)の場合
 - ヘッダ・フッタやナビゲーション部分など、特徴語抽出にノイズとなる部分を除去し、対象の本文テキストを取得
- PDFフルテキストの場合
 - 本文中のテキスト抽出
- 既存のツールを利用
 - Webページ(HTML): contentextract.rb
 - PDF: pdftotext (xpdf)

「ふわっと関連検索」システムの概要

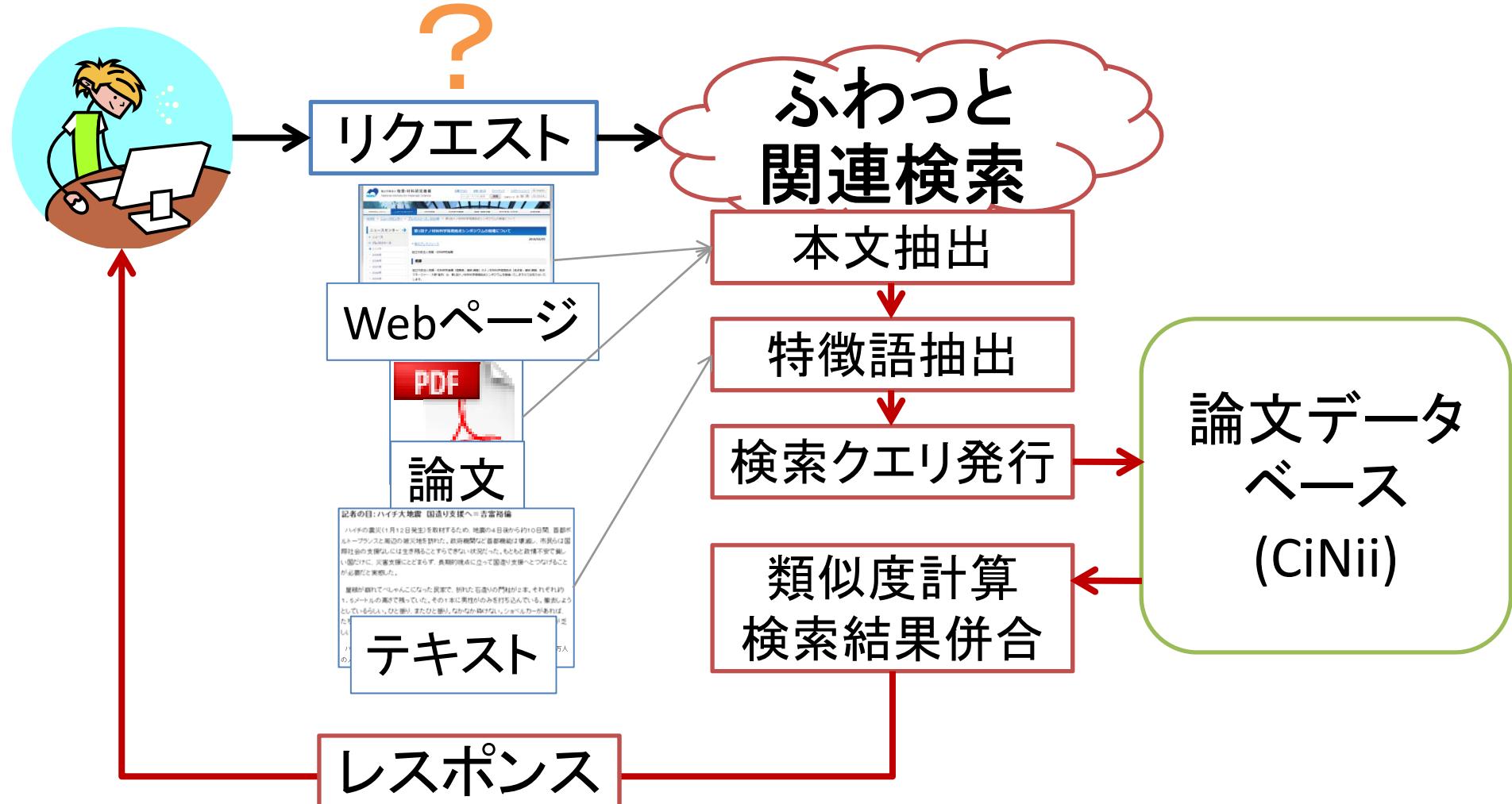


実装: 特徴語抽出

- テキスト中の出現語 w に対して各語の重みを計算
 - 2段階での簡易TF-IDFを採用
 - テキスト内の出現回数: $TF(w)$
 - テキスト中の単語の生起コスト: $Cost(w)$ ※
 - データベース文書群中のヒット文書数: $DF(w)$
1. テキスト中の特徴語上位n件 (*Logcost*)
 - $weight(w) = TF(w) \cdot LogCost(w)$
 2. データベース問合せ結果を使って特徴語スコアを調整 (*IDF*)
 - $weight'(w) = weight(w) / \log(DF)$

※生起コスト $Cost(w)$: 形態素解析ツールMeCabのコスト値

「ふわっと関連検索」システムの概要



実装: クエリ発行と検索結果併合

- 特徴語上位 n 件をクエリ発行
 - n 回のクエリ発行
 - ゼロヒットになるものを除外
 - (w_1, \dots, w_n) の各語をAND結合し、クエリ発行 **(AND)**
 - ゼロヒットになってしまう場合は、重みの低い特徴語から除外して、 (w_1, \dots, w_{n-1}) の各語として、クエリ発行。
 - 以下、指定した表示件数が得られるまで続行。
 - ※「CiNii関連検索」においては $n = 10$ を採用
 - ※CiNiiが提供するOpenSearch APIを通じて検索
- 得られた検索結果を重複除去し、最終的な検索結果とする。

文書類似度に基づく関連論文検索 「ふわっと関連検索」

利用サイト公開中

<http://fuwat.to>

(対応検索サービス募集中)

- 検索キーワードを考えなくともよい
- 文書特徴量をそのまま利用できる
- キーワード検索しかサポートしていない検索ツールを対象とした類似文書検索ができるようになる
- 一回の検索要求に対して、キーワード検索を複数回/複数パターン実行
- 新聞記事を対象とした評価実験
 - 適合度順ランキング10位時点での精度：平均0.25

まとめ

- 研究所内外の学術情報の取り出しから発信まで
 - 機関リポジトリ、研究者総覧、専門図書室運営
 - 研究者の活動、ニーズによりそったシステムの要件とその研究開発
 - 電子的なリソース管理やシステム開発を中心に
 - 部署間の調整やデータ交換の仕組みを設計、開発
 - (おまけ的に)研究活動も忘れずに。国内の情報科学研究の最前線と付いていく。

(おまけ) 学生時代にやっておいてよかったこと

- 失敗もしてみる
 - データディスク消失
 - 学内Googleブロック事件
- 外部のプロジェクトと関わる
 - オープンソースソフトウェア
 - 力試し
- 大は小を兼ねる: 大きなデータを扱う癖を知る
 - 3日くらいかかるプログラムを実行する
 - 最初によく考えてから処理に取り組む
- 図書館運営に親近感
 - なんとなくあれ全方位的にどんなサービスがあるか分かる

Thank you!

共同開発の実態（日本語化）

- eSciDocプロジェクト
 - FIZ: コアサービス
 - 認証、データストア、検索まわりなど
 - MPDL: アプリケーション
 - 仕様策定、実装、GUI設計、テスト
- NIMS: 開発&運用 
 - 日本で必要な機能の追加実装 = 日本語化（ローカライズ）
 - NIMSで必要な新機能提案 = 仕様設計

