

# セレンディピティを促す論文検索ツール 「ふわっと関連検索」

高久雅生

物質・材料研究機構

[TAKAKU.Masao@nims.go.jp](mailto:TAKAKU.Masao@nims.go.jp)

江草由佳

国立教育政策研究所

[yuka@nier.go.jp](mailto:yuka@nier.go.jp)

# 背景と目的

- サイエンスコミュニケーション
  - 市民社会の議論の接続点としての論文
  - 知的好奇心を満たすための専門文献としての論文
- ディープウェブ
  - 検索されなければ見えないコンテンツ
- 論文との出会いを増やす
  - 明示的なキーワード入力を必要としない検索手法
    - 通常のWebサーフィンで興味を持ったテキスト内容をそのまま検索に使えるようにする！

## これまでの放送

第135回 2010年2月2日放送

＜これまでの放送TOP＞ プロフェッショナルの道具

# サラリーマンは、スジを通せ

鉄道ダイヤ作成・牛田貢平



## スジは机で引くな、‘現場’で引け

牛田たちスジ屋の大きな仕事はダイヤ改正のために、列車運行表を作ること。中には1日のすべての列車の動きが書き込まれ、600以上の‘スジ’がひかれている。牛田が表を作る際に最も大切にしているのは、これまでの実績ややり方にとらわれず、駅や客(現場)の変化に常に向き合い、そこから考えること。

牛田は、ここ2年の東西線のダイヤ改正の際、朝ラッシュの時間帯のすべての電車にのり、すべての駅の状況を把握した。現場の実態を克明に記録し、それにあわせて各列車の駅での停車時間を5秒単位で振り分けし直したのだ。

その結果、朝ラッシュの時間帯の慢性的な遅延は半分以上になり、遅れに対する客からの苦情も10分の1に減った。牛田のダイヤはこうした徹底した現場主義から生まれた。

プロフェッショナル「仕事の流儀」2010年2月2日

<http://www.nhk.or.jp/professional/backnumber/100202/>

牛田は足繁く現場に通い、現場の状況に特に、数字に表れない変化に目をこらす。

# 列車運行実績データの可視化

学[電]○稲川真範（千葉工大） 正[電]富井規雄（千葉工大） 牛田貢平（東京地下鉄（株））

## Visualization of Train Operation Record Data

Masanori INAGAWA† Norio TOMII† Kohei USHIDA‡

†Chiba Institute of Technology 2-17-1 Tsudanuma Narashino, Chiba 275-0016 Japan

‡Tokyo Metro Co., Ltd. 3-19-6 Higashi Ueno Taito-ku Tokyo 110-0015 Japan

Several visualization methods of train operation record data obtained from train control systems are introduced. Visualization of the data is helpful to identify overall situation of occurrence of delays and the propagation to other trains and to confirm effectiveness of revision of timetables to make them more robust. We present several visualization methods together with our experience to apply them to actual train operation record data of Tokyo Metro, the largest subway company in Tokyo.

*Keywords:* train operation record, train operation, visualization, robustness

### 1. はじめに

近年、特に大都市圏においては、ラッシュ時に列車の遅延が頻繁に発生し、それを受けた[1]。文献[1]においては、遅延の発生もさることながら、最初に発生した遅延（初期遅延）が他の多数の列車に伝播

して、運行計画の改善作業を、「概観」「特定」「提案」「予測」「評価」の5つのフェーズに分け、特にそのうちの「概観」が列車運行実績データの表示形

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

### 2. 列車運行実績データの活用

[\[PDF\] 列車運行実績データの可視化](#)ファイルタイプ: PDF/Adobe Acrobat - [クイックビュー](#)

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](http://www.tomii.cs.it-chiba.ac.jp/kashikaJRAIL.pdf) - [コメント](#) [共有](#) [閉じる](#)[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](https://twitter.com/hidetan/status/8542455794) - [キャッシュ](#) - [コメント](#) [共有](#) [閉じる](#)[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図 <http://www.tomii.cs.it-chiba.ac.jp/kashikaJRAIL.pdf> #nhk.

[twitter.com/shiyu/status/8542431922](https://twitter.com/shiyu/status/8542431922) - [キャッシュ](#) - [コメント](#) [共有](#) [閉じる](#)[TweetBuzz - kashika](#)

番組で出ていた牛田さん  
列車運行実績データの可  
視化」可視化の手法が

[tweetbuzz.jp/entry/9262701/www.tomii.cs.it-chiba.ac.jp/kashikaJRAIL.pdf](https://tweetbuzz.jp/entry/9262701/www.tomii.cs.it-chiba.ac.jp/kashikaJRAIL.pdf) - [キャッシュ](#) - [コメント](#) [共有](#) [閉じる](#)

## 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](http://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](http://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/slicerdicom](http://www.lightstone.co.jp/slicerdicom)

[広告掲載はこちら »](#)

<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>

# 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](#)  
飯屋崎,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\(列車運行システム、新しい交通システム、情報伝送・処理システム\)](#)  
平粟,滋人  
サイバネティクス, 2010/1
6. [近鉄 名古屋地区列車](#)  
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>



# 参考: GoogleAdSense


## 「すべて物理の法則で動く」=野口さん、理科教材用に宇宙実験

国際宇宙ステーション (ISS) に長期滞在中の野口聡一さん(44)が6日夜、日本実験棟「きぼう」と筑波宇宙センターを衛星回線で結んで「おもしろ宇宙実験」を行った。無重力状態で赤いリンマーや紅茶パックを回して重心を示したり、テープの輪同士の衝突や水が入った白い大きなタンクをひもで引っ張る実験で、運動量や作用・反作用を説明した。

黒いTシャツ姿の野口さんは次に、ひもを手すり結びつけて1回引くと、自分の体がいつまでも動くほか、引くたびに加速していくのは、自身が船外活動(宇宙遊泳)で体験した動きに似ていると紹介。「すべて物理の法則にのっとって動いていることが分かります」と話した。

映像は宇宙航空研究開発機構のホームページで8日夜以降に見られるようになるほか、同機構の宇宙教育センターが小・中・高校向けの理科教材に編集する。中学の理科教諭を長く務めたセンター長の中村日出夫さん(61)は「子供たちの興味や関心を喚起できるのではないかな。教材は授業の導入やまとめに使える」と話した。(2010/03/06-22:00)

- 松井秀、オープン準決勝出場へ
- 一人暮らしの男性の部屋にあ...



インテル新プロセ  
ssor dv6i 春モデル

HP Pavilionシリーズ  
**春モデル発売開始**

妥協のない最強モデル  
HPEシリーズ

おすすめ 特集一覧

 笑顔の帰国 女子フィギュア 浅田真央	 刺繍と 海の土都 長崎・巻岐	 混乱続く 被災地 チリ大地震
 あの名場面を もう一度 ハンター・バー五輪	 YS-11、DC-8 航空博物館 航空博物館	 2010-11秋冬 パコレ 開幕!

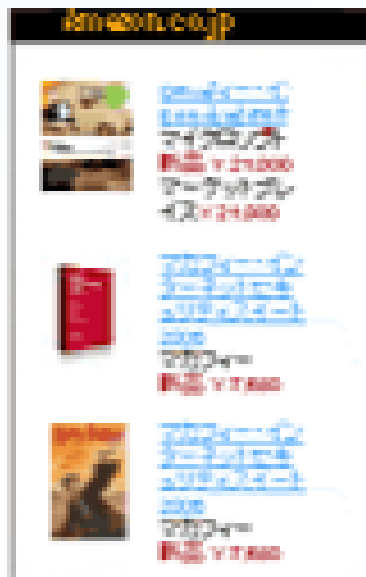
### Ads by Google

突然、口から英語が飛び出す ever-e.com  
第1位/日本一売れてる英会話教材EE 今、最も売れている英語教材とはEE?  
小学生の学習教材なら china.china.com  
日能研の“知の翼”。学び進める チカラを広げる2色添削!

### 関連ニュース

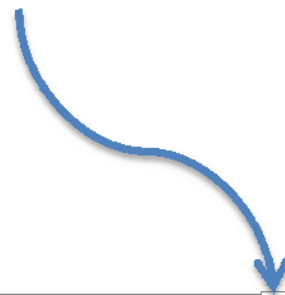
# 参考: Amazonおまかせリンク

## Amazonおまかせリンク(R)



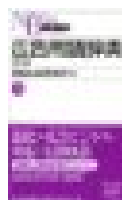
Amazonおまかせリンク(R)は、あなたのWebサイトの内容に沿った商品を自動的に表示する便利なリンク方法です。

- コンテンツマッチ広告
- Amazonの商品アイテムとWebページの内容コンテンツ



Amazon.co.jp

プライバシーについて

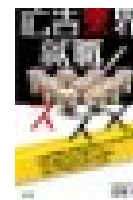


広告用語辞典

日経広告研究所

新品 ¥945

ベストプライス ¥50



広告業界就職ノススメ...

吉開 章

ベストプライス ¥82



# 参考 : Science and You

<http://you.costep.jp/>



つながりのある科学の話題にジャンプ！

## Science and You とは？

知識データベース



ウェブサイトに表示されたテキストと、科学に関するコンテンツを集めた「知識データベース」を照らし合わせ、サイトに関連する科学コンテンツを紹介する新しいサービスです。科学に興味のある方には、より幅広い知識を、あまり興味のない方には、新鮮な発見をお届けします。どんな体験ができるか、ぜひお試しください。

[\[→詳しく\]](#)

▼ 試してみたいサイトのURLを入力してください(どんなサイトでも結構です)。

[試してみる](#)

Science and You の  
ブログパーツを使おう！



あなたの  
ブログから  
科学の  
話題へ！

[新規作成](#)



ブログパーツを  
利用できない方はこちら

[ユーザの方はこちら](#)



# 「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。

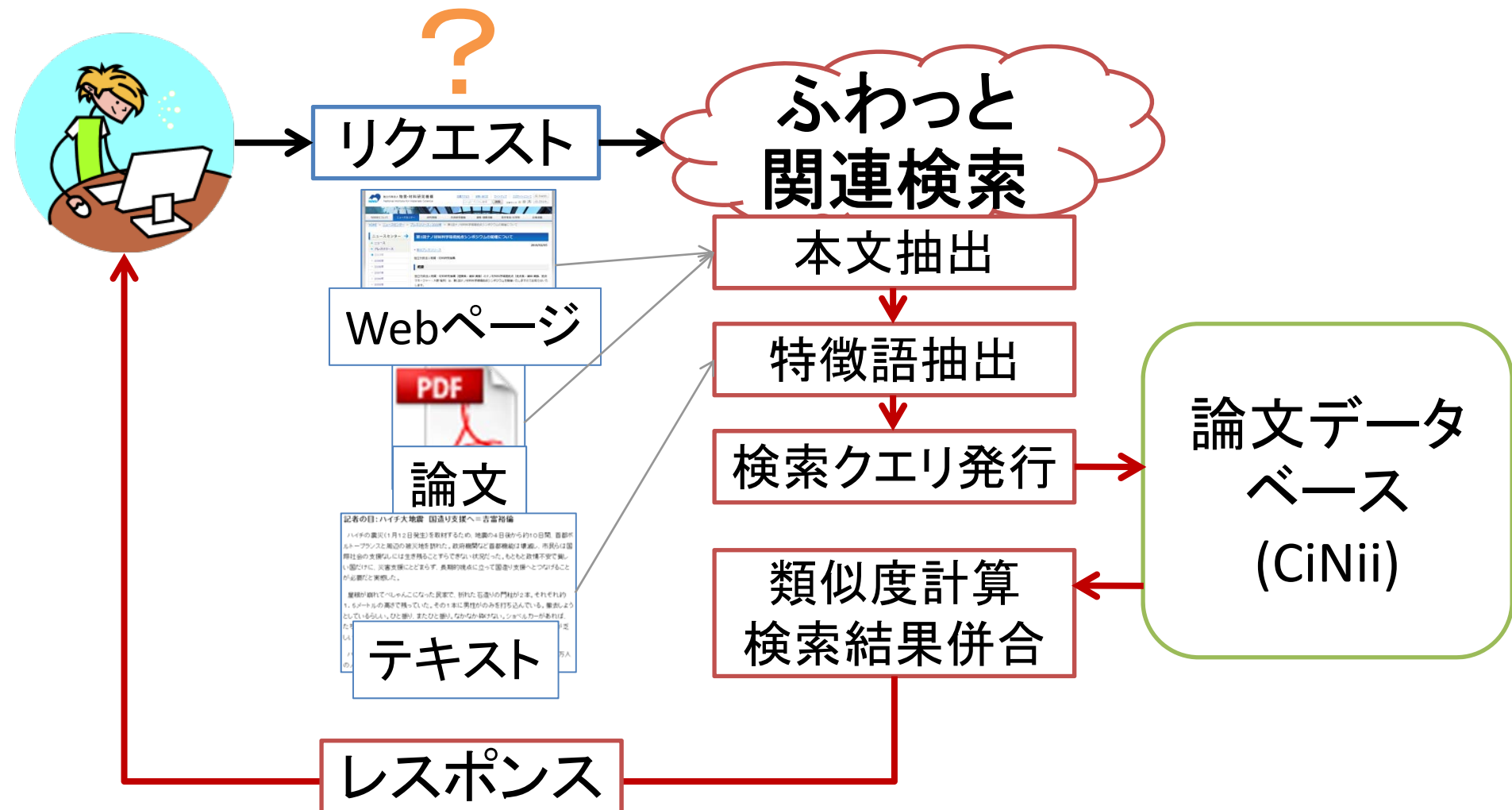
例: [朝日新聞社説](#) (元記事([asahi.com](#))), [日本経済新聞社説](#) (元記事([nikkei.co.jp](#))), [毎日新聞「記者の目」](#) (元記事([mainichi.jp](#)))

<http://fuwat.to/cinii>

高久雅生 (Masao Takaku)

<http://masao.jpn.org/>, [tmasao@acm.org](mailto: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)

# 実装: 特徴語抽出

- テキスト中での出現語 $w$ に対して各語の重みを計算
- 2段階での簡易TF-IDFを採用
  - テキスト内での出現回数:  $TF(w)$
  - テキスト中での単語生起確率:  $C(w)$
  - データベース文書群中でのヒット文書数:  $DF(w)$
- 1. テキスト中での特徴語上位 $n$ 件
  - $TF \times C(w)$
- 2. データベース問合せ結果を使って再ランク
  - $TF \div \log(DF)$

※生起確率 $C(w)$ : 形態素解析ツールMeCabのコスト値

# 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
    - 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。



# 「ふわっとCiNii関連検索」利用状況

- 2010年1月10日に公開

twitter

ホーム プロフィール 友だちを検索 設定 ヘルプ ログアウト

年末年始に作った、CiNii APIベースの論文検索ツール「ふわっとCiNii関連検索」を公開します。何かお気づきの点などありましたら、お気軽にご連絡いただければうれしいです。 <http://bit.ly/4JfHho> #cinii #fuwatto ☆

7:29 AM Jan 10th via Echofon  
2人がリツイート



tmasao  
Masao Takaku

twitter

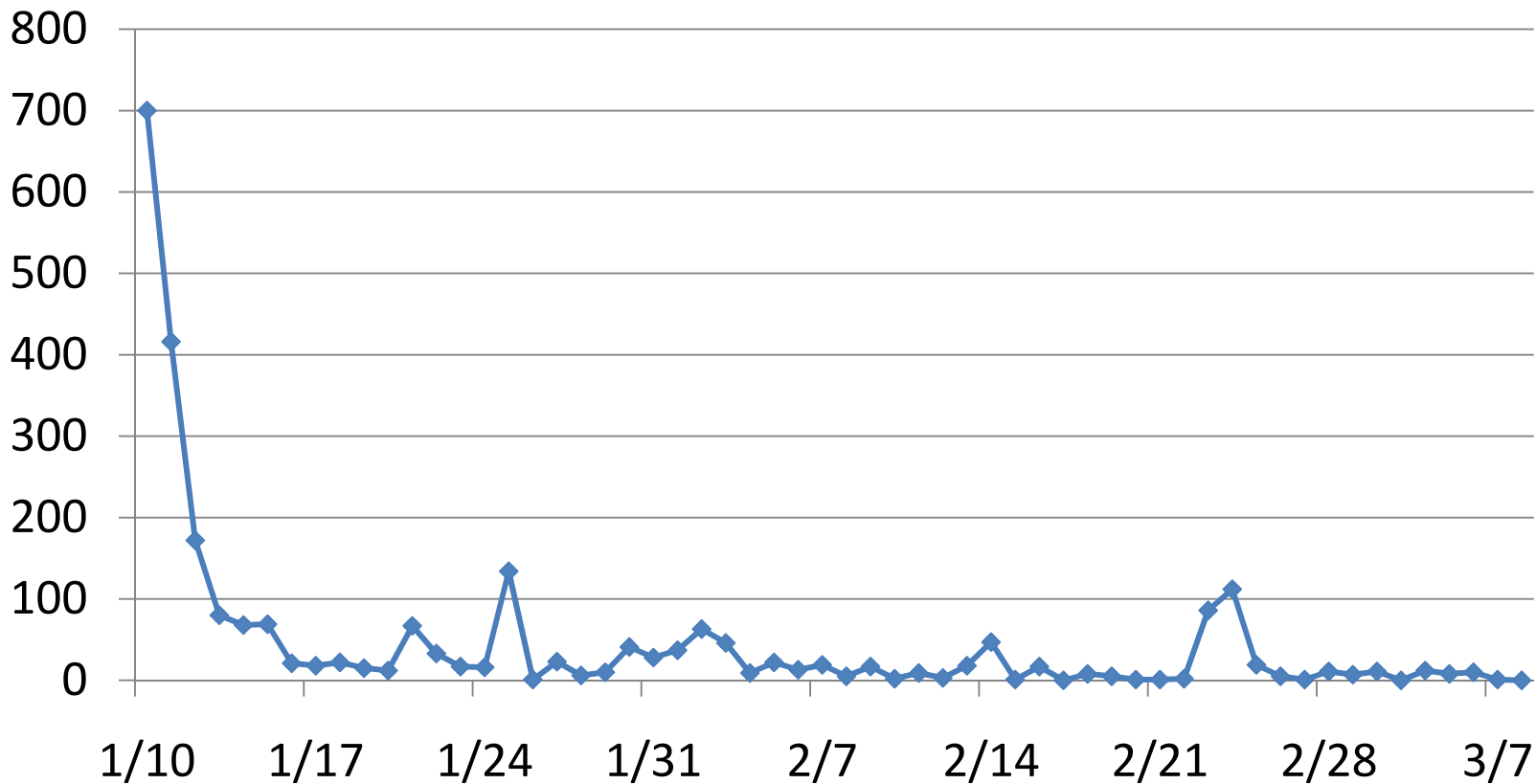
ホーム プロフィール 友だちを検索 設定 ヘルプ ログアウト

いわゆる連想検索風のツールなので、既知文献検索などよりは、テキストなどを元に気づきを得るための検索という使い方に適しているとおもいます。 <http://bit.ly/4JfHho> #cinii #fuwatto

7:29 AM Jan 10th via Echofon

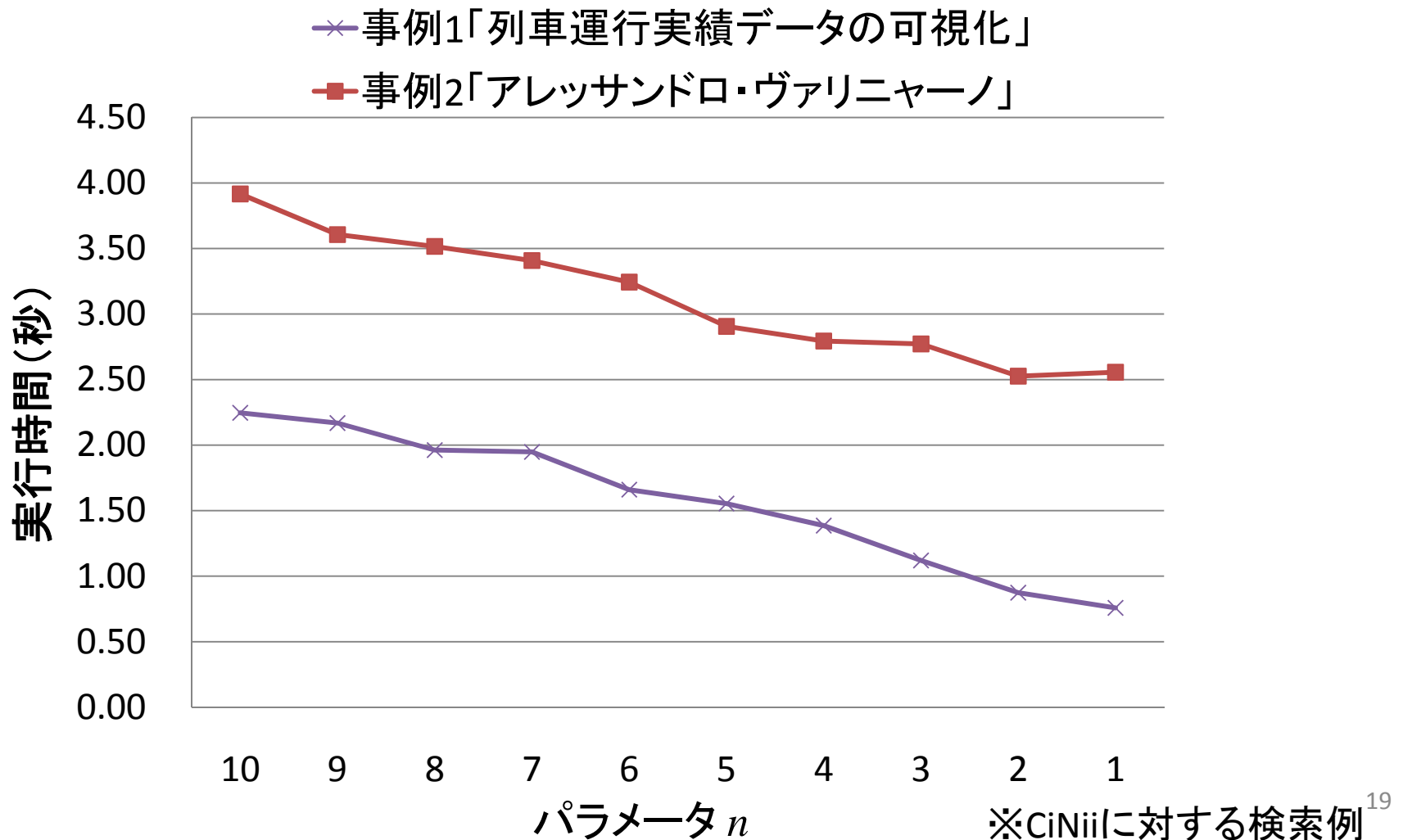
# 「ふわっとCiNii関連検索」利用状況

- 2010年1月10日に公開
  - Twitter / Blogで宣伝
- 1/10～3/7: 約2,500件のページビュー (Google Analytics)



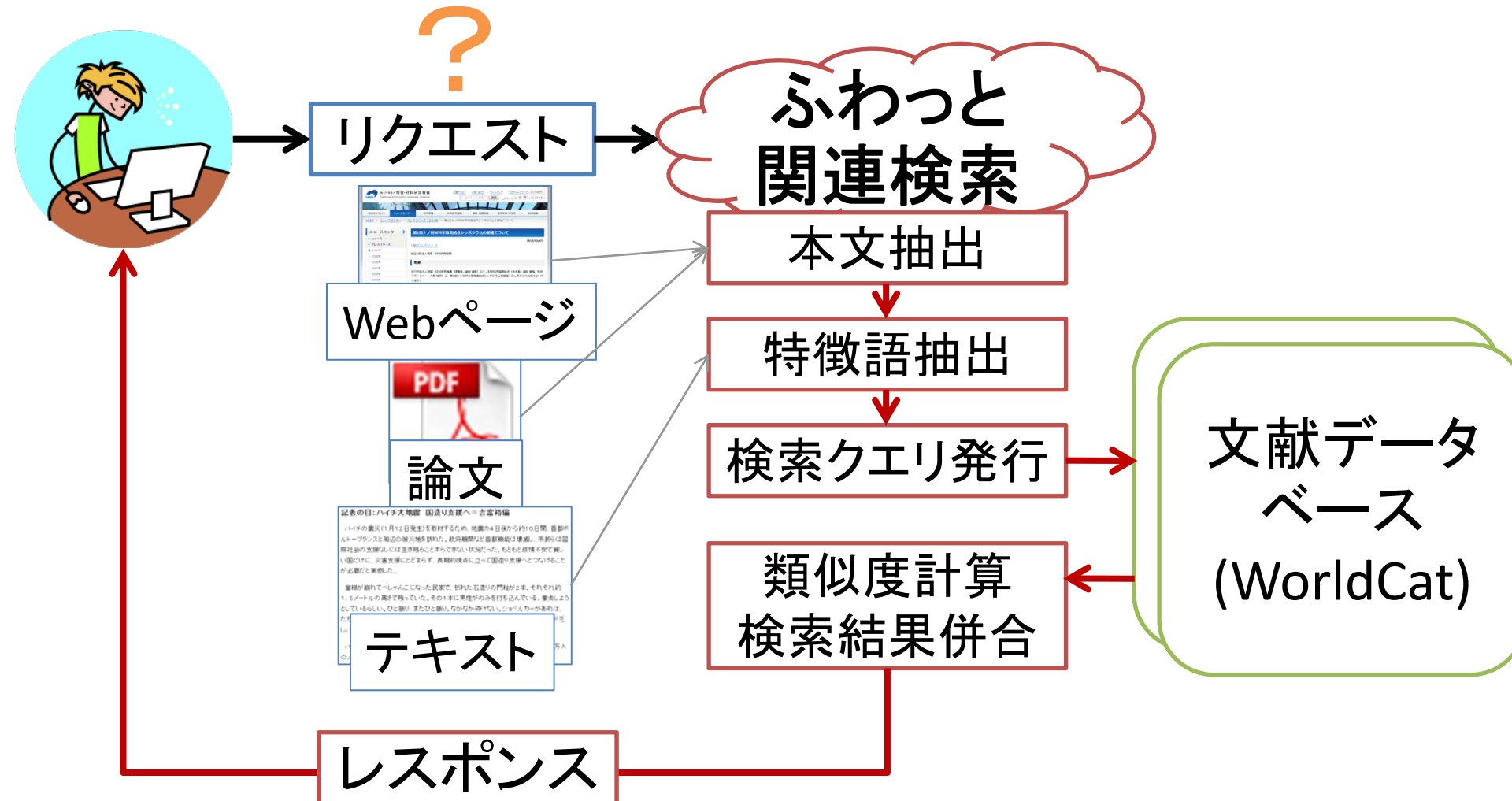
# 応答速度

- 特徴語数パラメータ  $n$  にほぼ比例
- 対応データベース側の応答速度に依存



# 「ふわっと関連検索」システム

## 対象データベース: WorldCat



# 「ふわっとWorldCat関連検索」

最終更新日：2010年01月22日  
(更新履歴)

## ふわっと WorldCat 関連検索

文章から検索

ウェブページから検索

<http://news.bbc.co.uk/2/hi/americas/8528100.stm>

論文検索

obama healthcare proposals Ci 3380件中 1- 20件目・ 検索にかかった時間24.83秒

1. [California health care almanac children's health coverage facts and figures](#)  
Cousineau, Michael R.
2. [Healthcare policy in an Obama administration : delivering on the promise of universal coverage.](#)  
PricewaterhouseCoopers LLP. Health Research Institute.
3. [Howard Dean's prescription for real healthcare reform : how we can achieve affordable medical care for every American and make our jobs safer](#)  
Dean, Howard, 1948-
4. [Now or Never : Getting Down to the Business of Saving Our American Dream.](#)  
Cafferty, Jack.
5. [The Georgia gang.](#)  
Williams, Dick. Dickerson, Jeff, 1955- Scott, Alexis. McKinney, Erica. Kent, Phil, 1951-
6. [Obamanomiksũ : Obama chǒngbu ũi segye kyǒngje chǒnmang](#)  
Talbot, John R., 1955-
7. [A day late and a dollar short : high hopes and deferred dreams in Obama's "postracial" America](#)

<http://fuwat.to/worldcat>

# まとめ

- 文献DBに対する連想検索的なマッシュアップ活用
  - － 簡易検索APIを活用
  - － 専門的文献をよりエンドユーザとつなげる
  - － DB提供者にとってはブラウジング機能強化の一方策
- 理論的には、簡易な類似文書検索手法  
(document-by-document retrieval)
  - － 対象DBによらない



# 今後の予定

- 特徴語抽出および類似度計算アルゴリズムの改良
  - 特徴語生起確率
  - 疑似適合フィードバック
- 適合性評価
- ふわっと関連検索API提供
  - JSON形式による提供を実装済み
- 対象データベース: NDL PORTA