

# 図書館AIのファーストステップ



株式会社カール  
代表取締役エンジニア  
吉本龍司

CC0 パブリック・ドメイン



# 日本最大の図書館蔵書検索サイト

- 複数の図書館をまとめて検索
- 全国6700館以上の図書館に対応
- 民間企業(株式会社カーリル)が運営
- オープンデータ・ビックデータ活用

# 使える図書館はたくさんある

- 住んでいるところ
- 働いているところ
- 大学図書館
- 国立国会図書館
- 専門図書館
- そもそも公共図書館は入場制限なし



# すてきな三にんぐみ

トミー=アンゲラー (著)、いまえよしも (翻訳)

石川県 ▼ の図書館をまとめて探す

33館の図書館で見つかりました！

## 蔵書あり

 かほく市	高松 貸出可	予約する	
 内灘町	内灘図 貸出可	予約する	
 石川県立図書館	本館 貸出可	予約する	
 能美市	根上図書館 貸出可	辰口図書館 貸出可	予約する
 羽咋市	羽咋 貸出可	予約する	
 志賀町	志賀図書館 貸出可	富来図書館 貸出可	予約する
 津幡町	津幡町立図書館 貸出中	予約する	
 金沢大学	研究室 蔵書あり	予約する	
 加賀市	中央 貸出可	山中 貸出可	予約する
 金沢美術工芸大学	図書館 貸出可	21世紀美術館 館内のみ	予約する

# カーリルの働き方

- 本社は岐阜県中津川市
- 基本はリモートワーク  
(中津川・京都・東京)
- 月に1回の開発合宿
- エンジニア・デザイナー  
図書館司書のチーム



ゼロステップ

人工知能と人間知能

# ゼロステップ

## カーリルの中の人間知能 の自動化

- 検索結果のバランス調整
- キャッシュ時間の制御
- スクレイピングエラーの統計的検出
- 閉館図書館の統計的検出

経験から得た閾値で自動化

ファーストステップ

AIっぽい



# 選書AIの開発

(ハッカソンでやってみた！)



**3月  
11日** 【応募締め切りしました】 町の本屋さんをも  
っと元気にするハッカソン

公開 · 主催者: バリューブックス valuebooks 古本で社会を  
変える会社

✓ 参加予定 ▾

...

# 選書のニーズ

- 図書館

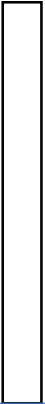
購入・除籍・配架・展示・利用促進

- 書店

- 出版社

# 選書API

本棚の幅



— ここに入れる本を返す  
—

はじめの本

おわりの本

# いい感じに本を並べる

—川・土・森—

海

山

どうやってテーマをたどるか

# Wikipediaの出典情報

## 脚注

1. <sup>^</sup> 詳しくは寄付のページを参照
2. <sup>^</sup> <sup>*ab*</sup> 「【インタビュー】有志で取り組む「史上最大の百科事典」——Wikipedia日本語版の舞台裏 Wikipedia日本語版 管理者・ビューロクラット 今泉 誠氏」 日経パソコン、2007年3月13日。今泉のアカウントは利用者:Suisui (会話 / 投稿記録 / 記録)。
3. <sup>^</sup> No.278 2003.12.20 CA1510 - オープンコンテンツの百科事典ウィキペディア / 福田亮 (カレントアウェアネス・ポータル)
4. <sup>^</sup> <sup>*ab*</sup> Terdiman, Daniel (2005年1月13日). “拡大するフリー百科事典『ウィキペディア』の課題”. *WIRED.jp*. WIRED.jp. 2009年8月19日閲覧。
5. <sup>^</sup> <sup>*abc*</sup> アンドリュー・リー 『ウィキペディア・レボリューション——世界最大の百科事典はいかにして生まれたか』 千葉敏生 翻訳、早川書房、2009年、284-286頁。ISBN 978-4-15-320005-0。
35. <sup>^</sup> 茂木健一郎 『脳を活かす勉強法 - 奇跡の「強化学習」』 (PHP研究所 2007年12月) P92-p94、ISBN 978-4569696799
36. <sup>^</sup> 火山爆発 2007年1月12日 (クオリア日記)
37. <sup>^</sup> 五千嵐茂さんからのコメント 2007年1月12日 (クオリア日記)
38. <sup>^</sup> クライン孝子の日記 5月分より「フリー百科事典『ウィキペディア』への疑問と抗議」。
39. <sup>^</sup> 本人ブログよりウィキペディアのでたらめ
40. <sup>^</sup> お互いに出口の見えない論争になるため、終いには馬鹿馬鹿しくなって放置され、記事も保護されたままとなる。
41. <sup>^</sup> <sup>*ab*</sup> どこまで載せる？ウィキペディア 進行中の項目巡り議論 (朝日新聞2006年10月20日)
42. <sup>^</sup> ネット上で壮絶バトル勃発!! キムタク&Hydeの身長は？ サイズー2006年3月号 (ウェブアーカイブ)

# Pythonのネットワーク解析ライブラリ

## NetworkX

[NetworkX Home](#) | [Documentation](#) | [Download](#) | [Developer \(Github\)](#)

### High-productivity software for complex networks

NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.



[Documentation](#)

*all documentation*

[Examples](#)

*using the library*

[Reference](#)

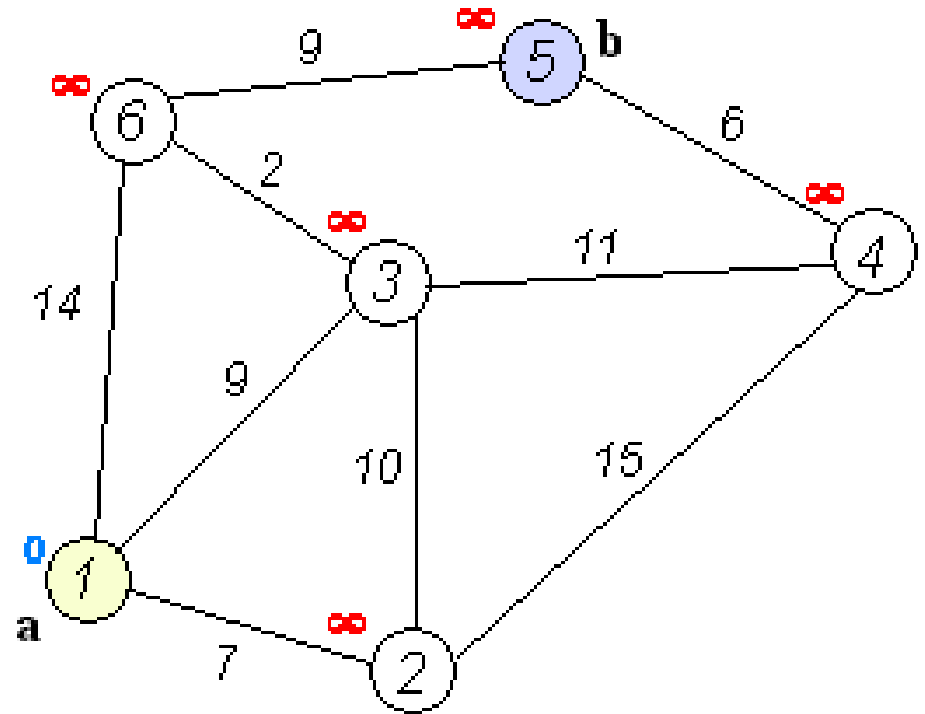
*all functions and methods*

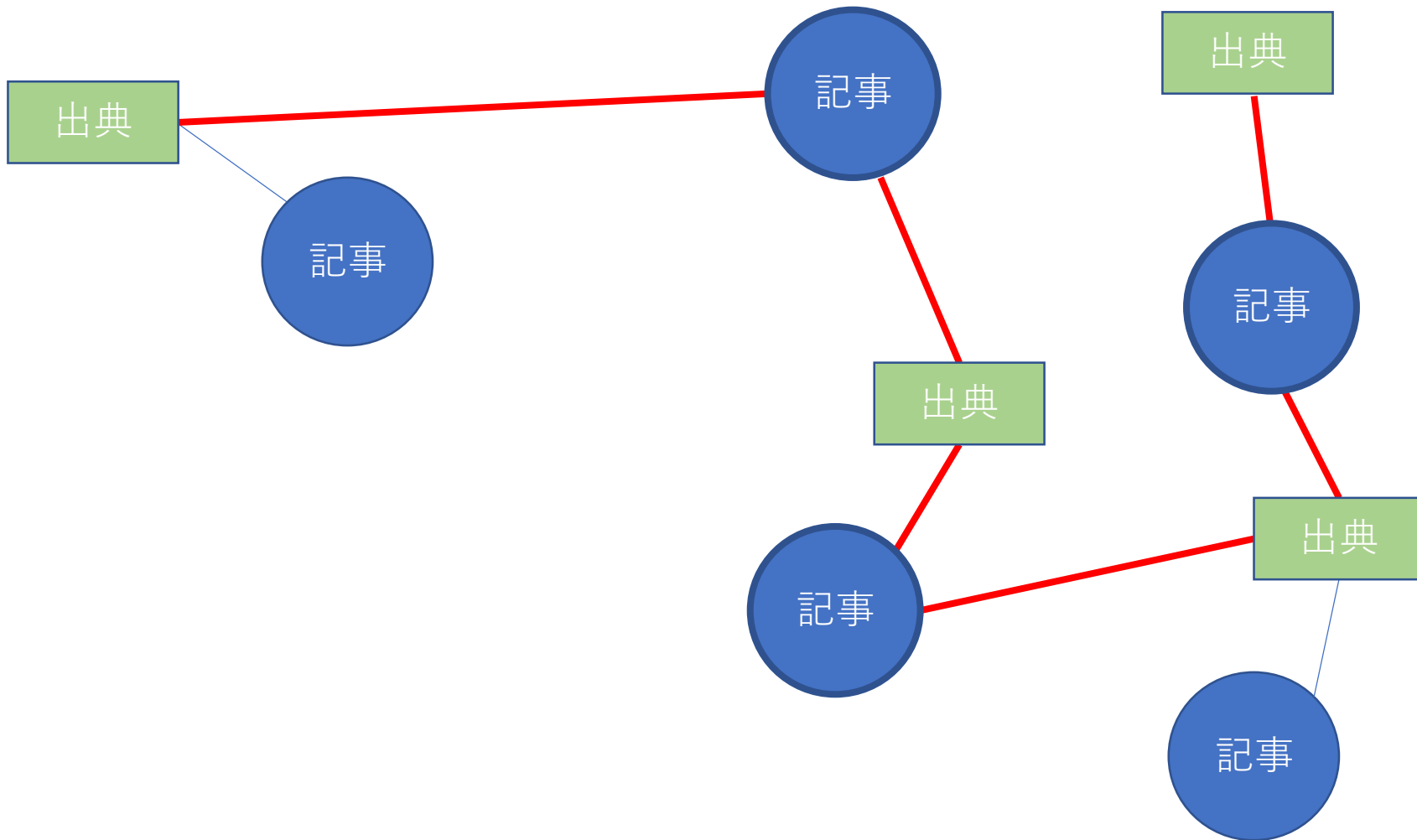


最短経路を探  
す (Dijkstra's  
algorithm)

ダイクストラ  
法

---





Wikipediaの出典にある本しか出ない

いい感じにごまかす

山と

# 出典資料は高速道路

書棚をご利用いただく  
Adobe Systemsの



- 脇道は書誌情報から構成する
- 連想検索

山と川



**こんなに楽しい岐阜の山旅100コース 美濃下**  
 吉川幸一 編著  
 風媒社  
 2003.1

No Photo

**石川 山と川十人集**  
 石川忠三ほか著  
 山と川  
 1978



**山と川と海と**  
 森山京文; 太田大八 絵  
 小峰書店  
 2005.12

No Photo

**美濃の山 第1巻 (揖斐川水系の山 1)**  
 大垣山岳協会 編  
 アカニシヤ  
 1996.12



**君の自然**  
 鈴木欣也, 藤平量郎, 吉原洋 著  
 うぶ書房  
 2003.10

No Photo

**美濃の山 第2巻**  
 大垣山岳協会 編  
 ナカニシヤ  
 1997.10

# Elasticsearch (全文検索エンジン) + Okapi BM25

- 普及している類似検索のアルゴリズム

## The ranking function [\[ edit \]](#)

---

BM25 is a [bag-of-words](#) retrieval function that ranks a set of documents based on the [relationship](#) between the query terms within a document (e.g., their relative proximity). It is not a [simple](#) function of the different components and parameters. One of the most prominent instantiations of the [function](#)

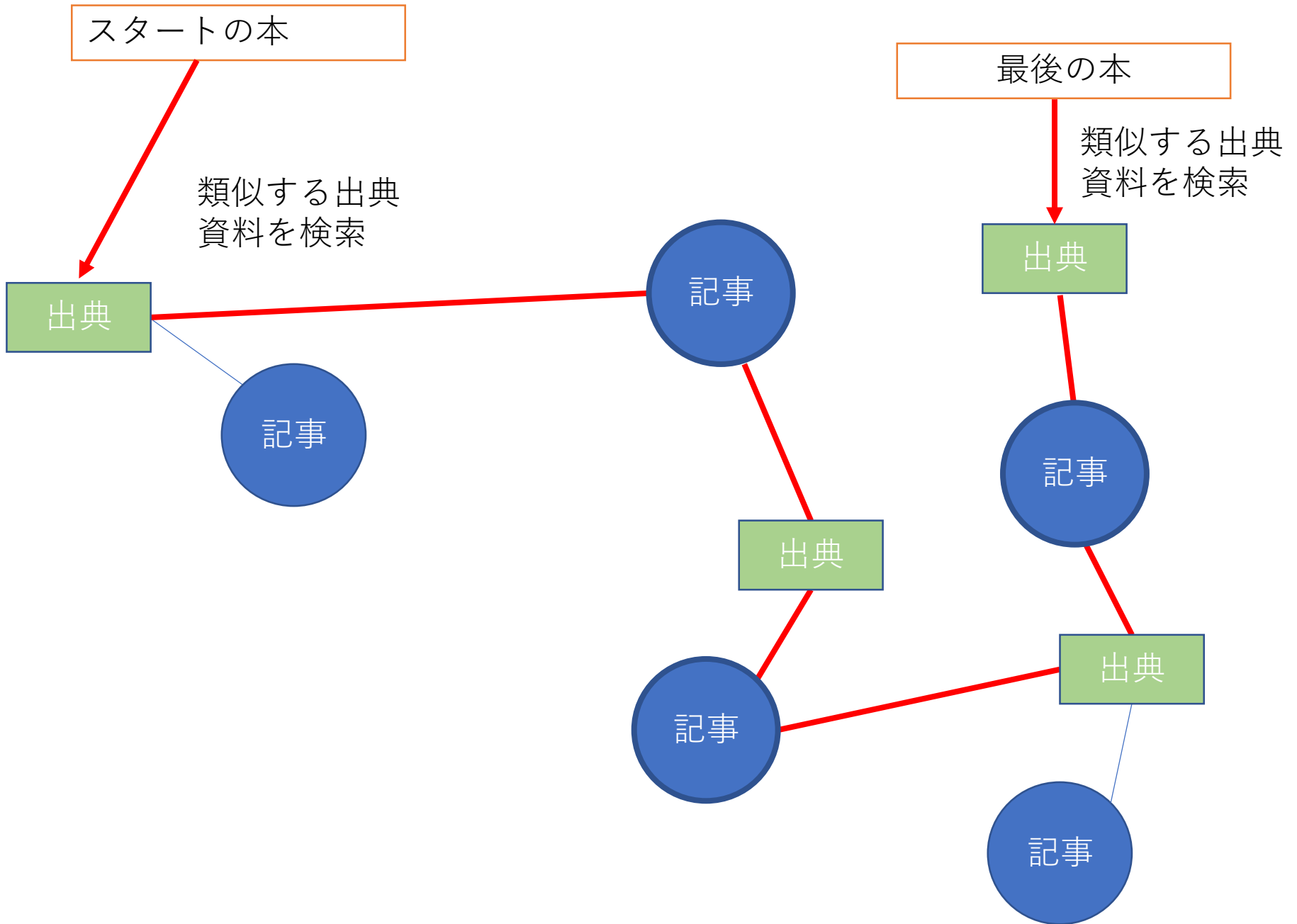
Given a query  $Q$ , containing keywords  $q_1, \dots, q_n$ , the BM25 score of a document  $D$  is

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)},$$

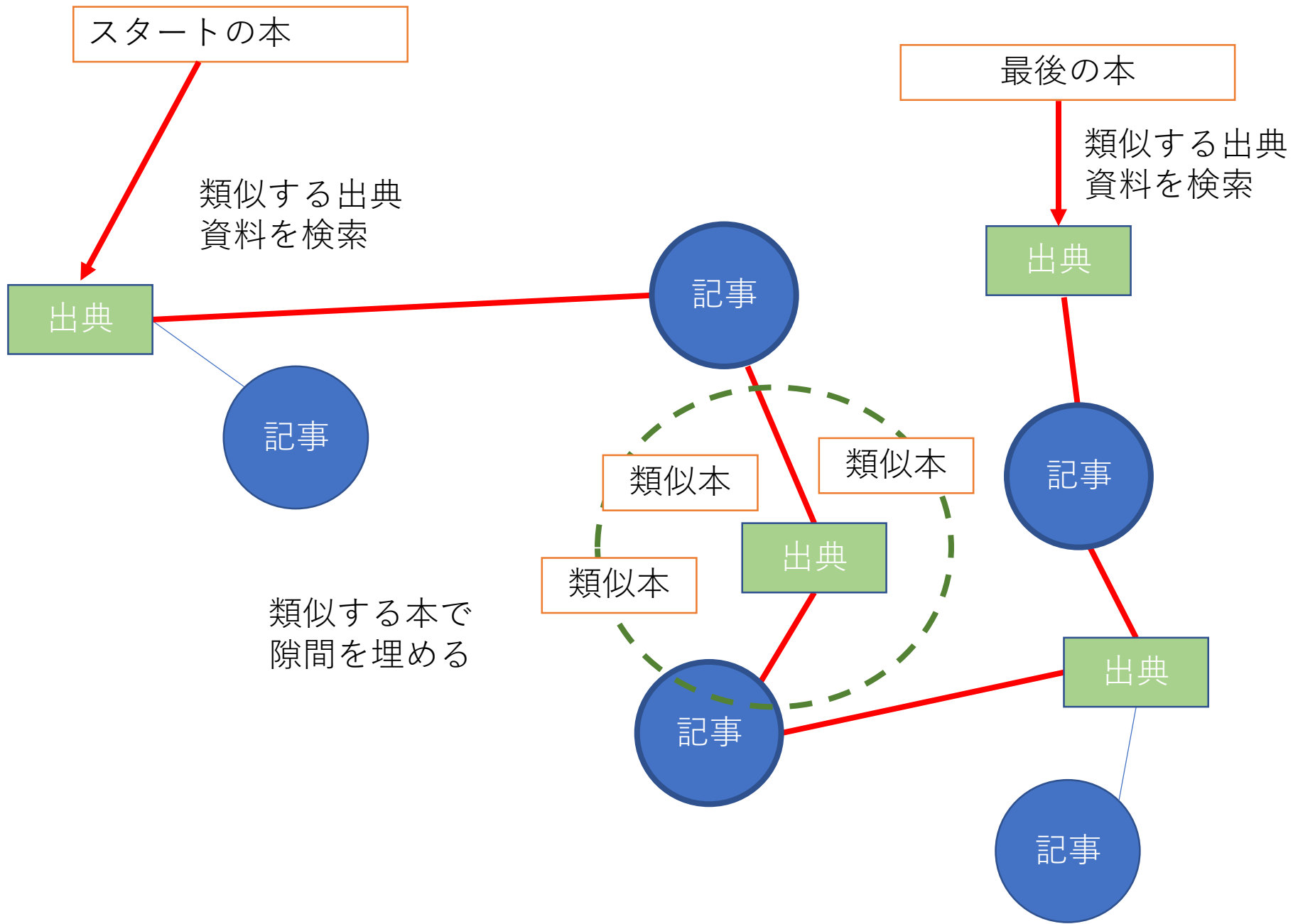
where  $f(q_i, D)$  is  $q_i$ 's [term frequency](#) in the document  $D$ ,  $|D|$  is the length of the doc collection from which documents are drawn.  $k_1$  and  $b$  are free parameters, usually chosen as  $k_1 = 1.2$  and  $b = 0.75$ .<sup>[1]</sup>  $\text{IDF}(q_i)$  is the [IDF \(inverse document frequency\)](#) weight of the query term

$$\text{IDF}(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5},$$

NDLサーチの全書誌を投入







NDL+Wikipedia

あとは見てみよう

```
{
  "publisher": "立風書房",
  "isbn": "4651700799",
  "title": "家閥：地方豪族のネットワーク",
  "author": "佐藤朝泰／著",
  "cover": "http://images-jp.amazon.com/images/P/4651700799.09.LZZZZZZZ",
  "relation": null,
  "relation_title": null
}
```

```
{
  "publisher": "立風書房",
  "isbn": "4651700756",
  "title": "漢方が救う人体危機：西洋医学一辺倒からの脱出",
  "author": "渡辺武 著",
  "cover": "http://images-jp.amazon.com/images/P/4651700756.09.LZZZZZZZ",
  "relation": "Wikipedia",
  "relation_title": "人体"
}
```

```
{
  "publisher": "紀伊国屋書店",
  "isbn": "4314010037",
  "title": "利己的な遺伝子",
  "author": "Dawkins,Richard／著 垂水雄二／著 岸由二／翻訳 日高敏隆／翻訳",
  "cover": "https://cover.openbd.jp/9784314010030.jpg",
  "relation": "Wikipedia",
  "relation_title": "血縁選択説"
}
```

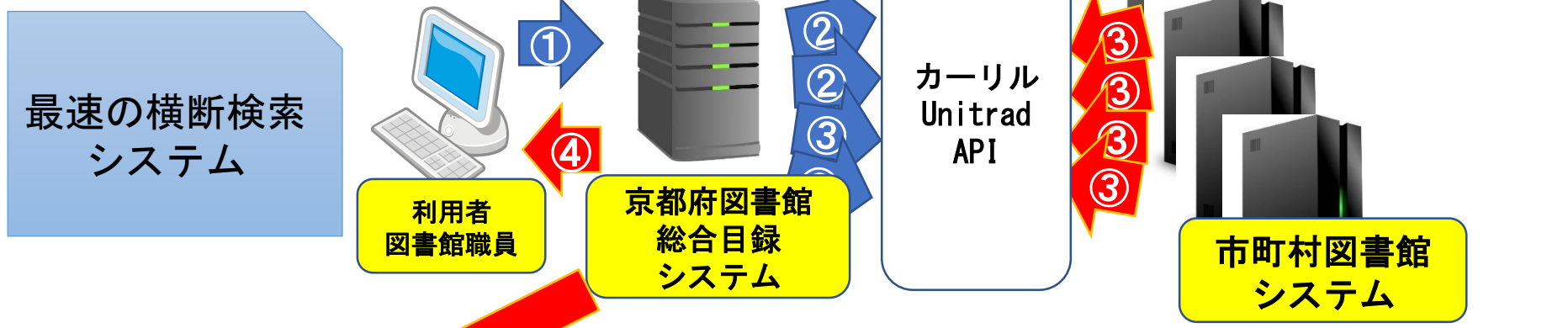
```
{
  "publisher": "東海大学出版会",
  "isbn": "4486017374",
  "title": "社会：社会生物学・行動生態学入門",
  "author": "菅昭 編著",
  "cover": "http://images-jp.amazon.com/images/P/4486017374.09.LZZZZZZZ",
  "relation": "Wikipedia",
  "relation_title": "竹内久美子"
}
```

```
{
  "publisher": "講談社 KADOKAWA",
  "isbn": "4040800202",
  "title": "人間を超えるか",
  "cover": "https://cover.openbd.jp/9784040800202.jpg",
  "relation": null,
  "relation_title": null
}
```

安倍晋三  
⇒  
ディープラーニング

これからの話

日本全国の全書誌を同定する



### 共同研究の仕組み

(国会図書館 (予定) ・同志社大 ・筑波大 ・千葉大 ・カーリル ・府立図書館)

蓄積されたデータ 【約713万冊】

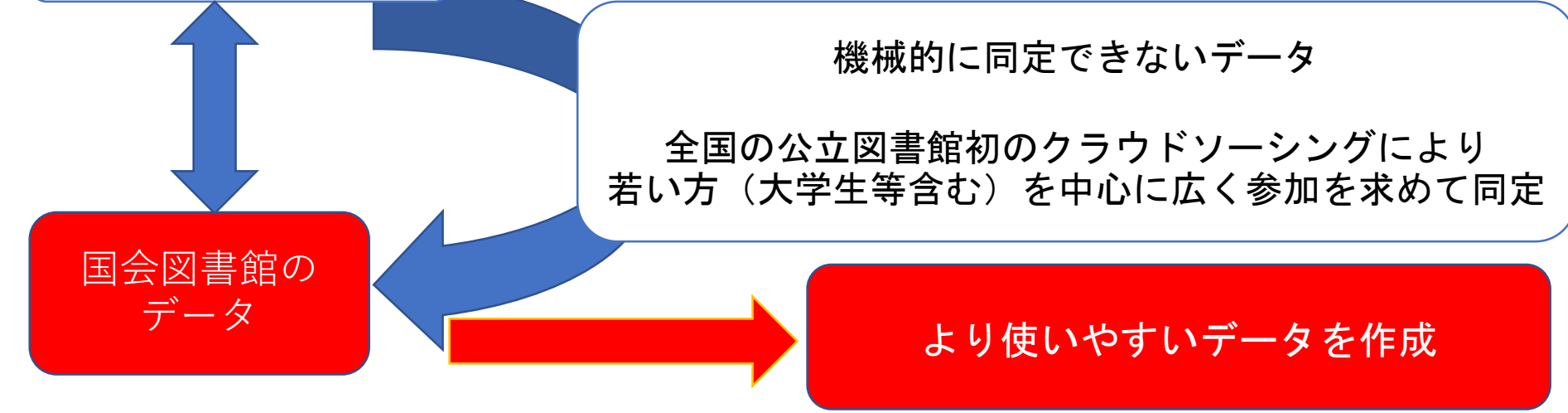
機械的に同定

機械的に同定できないデータ

全国の公立図書館初のクラウドソーシングにより若い方 (大学生等含む) を中心に広く参加を求めて同定

国会図書館のデータ

より使いやすいデータを作成



# 府民サービス上のメリット

# 横断検索の精度の向上

## • 現状

京都府の公共図書館 平成4年版	京都府立図書館/編集	京都:京都府立図書館1993
京都府の公共図書館 平成4年版	京都府立図書館/編集	京都府立図書館 1993
京都府公共図書館等所蔵雑誌・新聞総合目録	京都府図書館等連絡協議会/編集	京都:京都府図書館等1989
点字・録音図書目録 1990年1月現在	(京都府立図書館/編)	(京都):京都府立図1990
京都図書館紀行	勝山 俊光/編集	玄光社 2013.4
日本のテラコッタ建築		LIXIL出版 2012.4

同一資料がばらばらに

## • 改善後

京都府公共図書館等所蔵雑誌・新聞総合目録 1989	京都府図書館等連絡協議会/編者	京都:京都府図書館等198910
京都府の公共図書館 平成4年版	京都府立図書館/編者	京都府立図書館 1993
京都府の公共図書館 平成4年版	京都府立図書館/編者	京都府立図書館 1993
京都府の公共図書館 1987年版	京都府立図書館/編集	京都:京都府立図書館1987
京都府の公共図書館	京都府立図書館/編集	京都:京都府立図書館1983

同一資料は一つに  
さらに、検索語に適合  
して上位に表示

シリーズものは続けて表示

# 人間知能

- 地域資料は1年くらいいずれる
- タイトル全然違うのもある
- タイトルの取り方も違う



スコアリングの工夫

同定のための機械学習

# 図書館AIの未来

案外騙せる

ブラックボックスよくない

(exp.配架図デジタル化)