

トピックの特性を言語間で 比較・対照分析する 多言語ウェブテキストマイニングの研究

筑波山

筑波大学大学院
システム情報工学研究科 知能機能システム専攻
宇津呂武仁

2008年9月26日(金)、『言語処理技術の深化と理論・応用の新展開』科研・合同シンポジウム、@熱海金城館

科研費

- 基盤B, 平成20～22年度、
- 「トピックの特性を言語間で比較・対照分析する
多言語ウェブテキストマイニングの研究」

- 研究体制
 - 代表者: 宇津呂武仁(筑波大)
 - 分担者: 藤井敦(筑波大)
 - 連携研究者: 中川裕志(東大), 清田陽司(東大),
福原知宏(東大)

研究協力体制

- 科研特定情報爆発A01班関連「世界ニュース」研究グループ
 - メンバー：
 - 中川(東大)、神門(NII)、吉岡(北大)、
 - 森(横浜国大)、田浦(東大)、福原(東大)、
 - 宇津呂(筑波大)、他
- 乾(奈良先端大、科研特定情報爆発A01公募班)

背景

- 異文化交流の機会の増加
 - 旅行・留学・出張・・・
- 自分の知識と海外での知識・意見とのギャップ
- 言語間のギャップを見つけない！

言語間ギャップ: 竹島

- 日本
 - 2月22日を竹島の日に制定
 - 竹島は島根県の一部であると国際法的に認められている
 - 国民の話題に上る話ではない
- 海外
 - 韓国では10月を「独島の月」に制定
 - 「独島はわが領土」という歌が幼稚園などで歌われている



言語間ギャップを知る方法

- 共同通信の記事や他国テレビの映像を購入
- 特派員を派遣して現地取材
- シンクタンクによる調査
- ネットの情報を人海戦術



韓国のテレビCMを日本のテレビ局が報道
(2005年テレビ東京)

多言語・多メディア・多ジャンルの情報源からのマイニング

異なる国から発信された報道映像・ニュース・ブログを統合



多言語データの
収集基盤の構築

異なる特徴: 言語: 日本語・英語・中国語・韓国語……
興味(想定読者の興味:発行元の国の興味),
文化のギャップ, 事実の報道 v.s. ブログの意見

社会的貢献

- ・ 輸出入関連会社・海外情報アナリストは言うまでもなく、一般利用者までもが、言語・文化の壁を越えた情報収集・分析を手軽に実現
- ・ 異文化ギャップの理解を大幅に促進

話題:
イベント・人物
を選ぶ

多言語マルチメディア対照情報分析を
実現する情報アクセス

多言語マルチメディアデータの
メディア間の対応付け

地球温暖化に
関する世界中
の意見は?

様々な観点から
メディア・ブログを分析

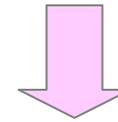
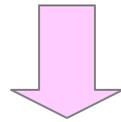
日本の主張
が海外で
あまり報道さ
れていないな

情報源の違いの分析

				
報道映像	国民の反応は概ね肯定的			世論の反対意見多数	
新聞(ウェブニュース)					
ブログ・電子掲示板					

	日本語	韓国語	中国語	英語
ニュース	多言語ウェブ文書収集: [世界ニュース研究 グループ・福原]			
ブログ・ 電子掲示板				

多言語のトピック体系Wikipediaを
利用したトピックの特性分析
[宇津呂・清田・福原]

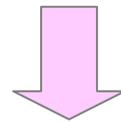


トピックの多様な特性の分析

スパムブログにおける
トピックのバースト特性の分析
[宇津呂・中川]

トピックの情報価値の質・持続性の
推定およびトピックの
視覚的二次元配置 [宇津呂]

トピックに関連する観点・
意見分布の抽出
[藤井・宇津呂]



トピックの言語間差異度の測定

トピックの情報価値の質・
持続性の差異度
[宇津呂・中川・福原・清田]

トピックの持つ観点の分布の
差異度 [宇津呂]

トピックの持つ意見分布の
差異度 [藤井]

Web2.0コンテンツの普及

- Web2.0的コンテンツの普及
 - ウェブログ(通称ブログ)
 - ウィキペディア
 - SNS(mixi,greeなど)
- 意見や情報などが多く書かれる
- 日々更新され、新しい情報がすぐに得られる

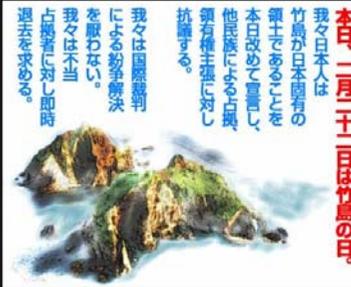
- 日本にいながらにして、
世界中の人が書いた意見や情報を知ることが出来る！

ブログで見る「竹島」

日本語ブログ



このページは「竹島プロジェクト2007」のために、提供していただいた画像、イラストを紹介しています。プロジェクトに参加したいが、適当な画像やイラストがなくて困っている方はぜひご利用ください。



「竹島の日」にネットで有志を集めてイベントを企画
「竹島」は日本の領土と主張
右翼的な著者が多い

英語ブログ



Kotaji 거타지

The Shimane assembly law's passing was applauded by right wing activists in paramilitary uniforms.

I didn't realise that Japan had its own neo-Nazi nutters. No picture:

韓国人ブロガーが多い。
「独島」を「竹島」として島根県の一部としていることに反発。反日的な感情が前面に出ている。

	日本語	韓国語	中国語	英語
ニュース				

多言語のトピック体系Wikipediaを

- ・ Wikipediaエン트리とブログサイトの対応付け [ICWSM2008, DEWS2008]
- ・ 日本語ブログ空間におけるトピック分布推定 [NL研2008年9月]

トピックの多様な特性の分析

スパムブログにおける

- ・ スプログデータセット作成・分析 [ICWSM2008, AIRWeb2008, DEWS2008]

トピックの情報価値の質・持続性の質およびトピックの次元配置 [宇津呂]

トピックに関連する観点・意見分布の抽出 [藤井・宇津呂]

トピ

- ・ 日英ブログ間の対照分析支援 [ICWSM2008, ISUC2008, DEWS2008, NLP2008]

トピックの情報価値の質・持続性の差異度 [宇津呂・中川・福原・清田]

意見分布の抽出 [藤井]

KANSHINシステム(福原)

図1のようにキーワードを入力することにより、そのキーワードを含む記事が更新された頻度を図2に示すような関心の時系列推移グラフに示すものである。このシステムはPerlとMySQLによって実装されており、Perlはブログ記事の取得と検索、MySQLはブログ記事のデータの格納に使用されている。システムはブログサービスの提供する新着記事リストを取得し、それぞれの記事に含まれるRSS・Atomを取得し、MySQLデータベースに格納する。

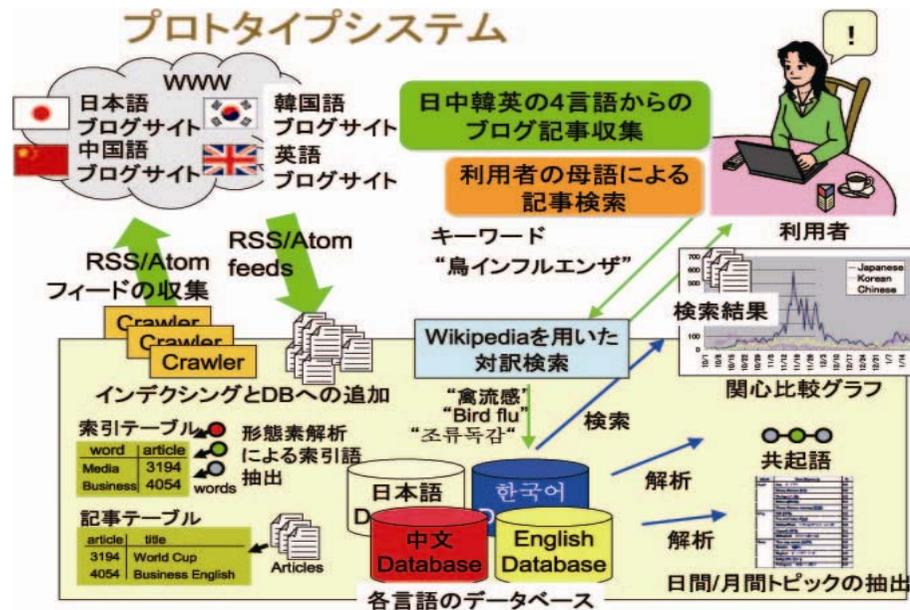
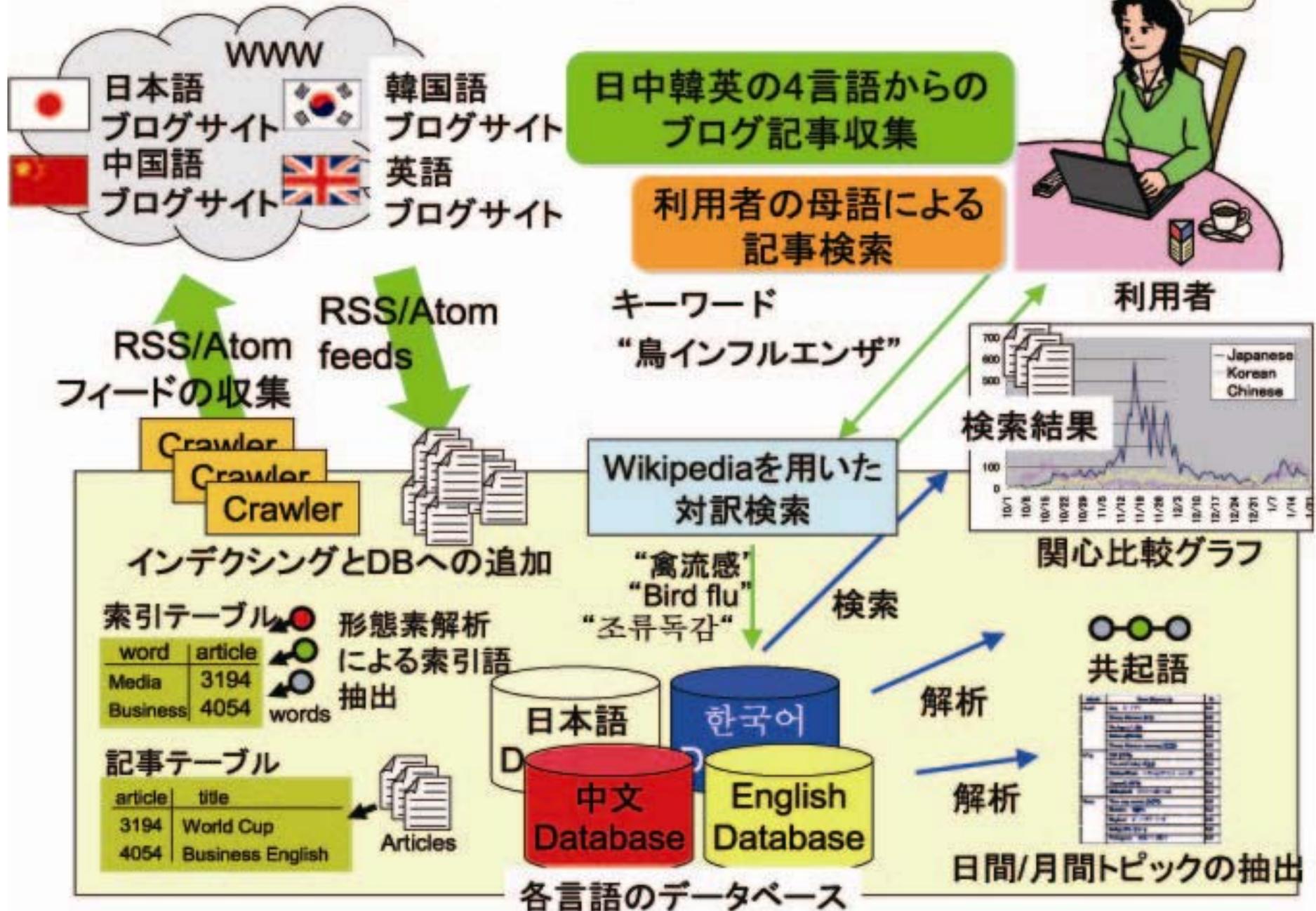


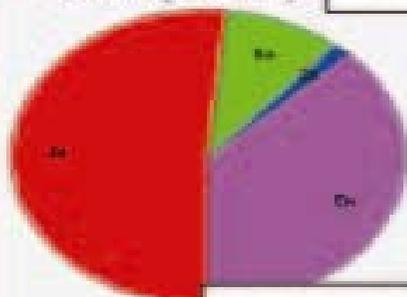
図1

図2

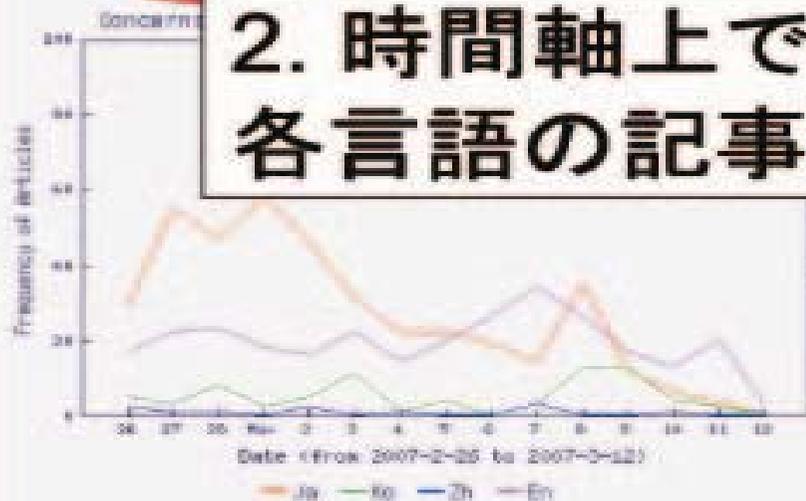
プロトタイプシステム



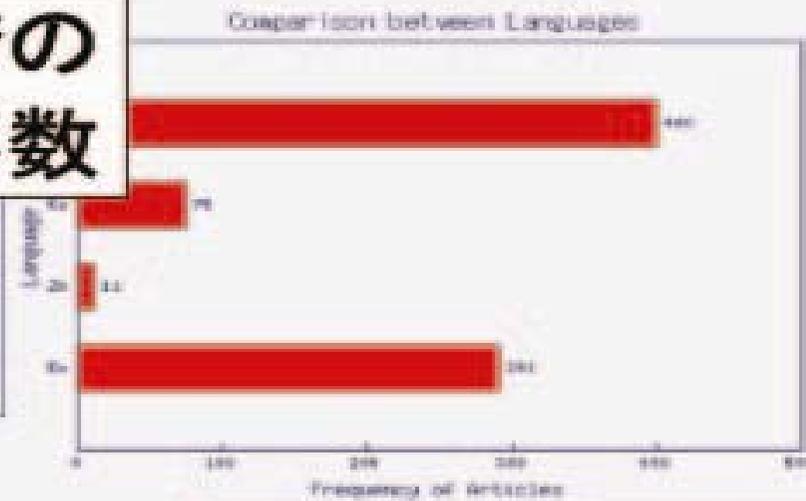
1. 言語の内訳



2. 時間軸上での各言語の記事数



3. 各言語の記事数

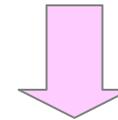
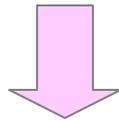


4. 検索された記事

date	fid	aid	lang	
2007-3-12 17:20:08	428320	4542	ko	가걸물을 채워해
2007-3-12 07:21:38	2559690	4115	ja	週刊誌月曜

	日本語	韓国語	中国語	英語
ニュース	多言語ウェブ文書収集: [世界ニュース研究 グループ・福原]			
ブログ・ 電子掲示板				

多言語のトピック体系Wikipediaを
利用したトピックの特性分析
[宇津呂・清田・福原]

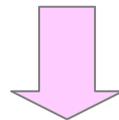


トピックの多様な特性の分析

スパムブログにおける
トピックのバースト特性の分析
[宇津呂・中川]

トピックの情報価値の質・持続性の
推定およびトピックの
視覚的二次元配置 [宇津呂]

トピックに関連する観点・
意見分布の抽出
[藤井・宇津呂]



トピ

・ 日英ブログ間の対照分析支援
[ICWSM2008, ISUC2008,
DEWS2008, NLP2008]

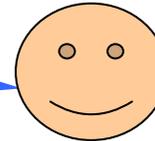
トピックの情報価値の質・
持続性の差異度
[宇津呂・中川・福原・清田]

意見分布の
抽出 [藤井]

提案システムの概念

しかし、どう探す??

「竹島」に関する日本語情報と英語情報を知りたい。



Wikipediaに記載されている知識や中立的観点から見た情報を
ブログ検索のために利用する!



韓国・北朝鮮は竹島を自国の最東端の領土(独島)であるとしているが、日本は国際法上適法な日本領土であるとしている

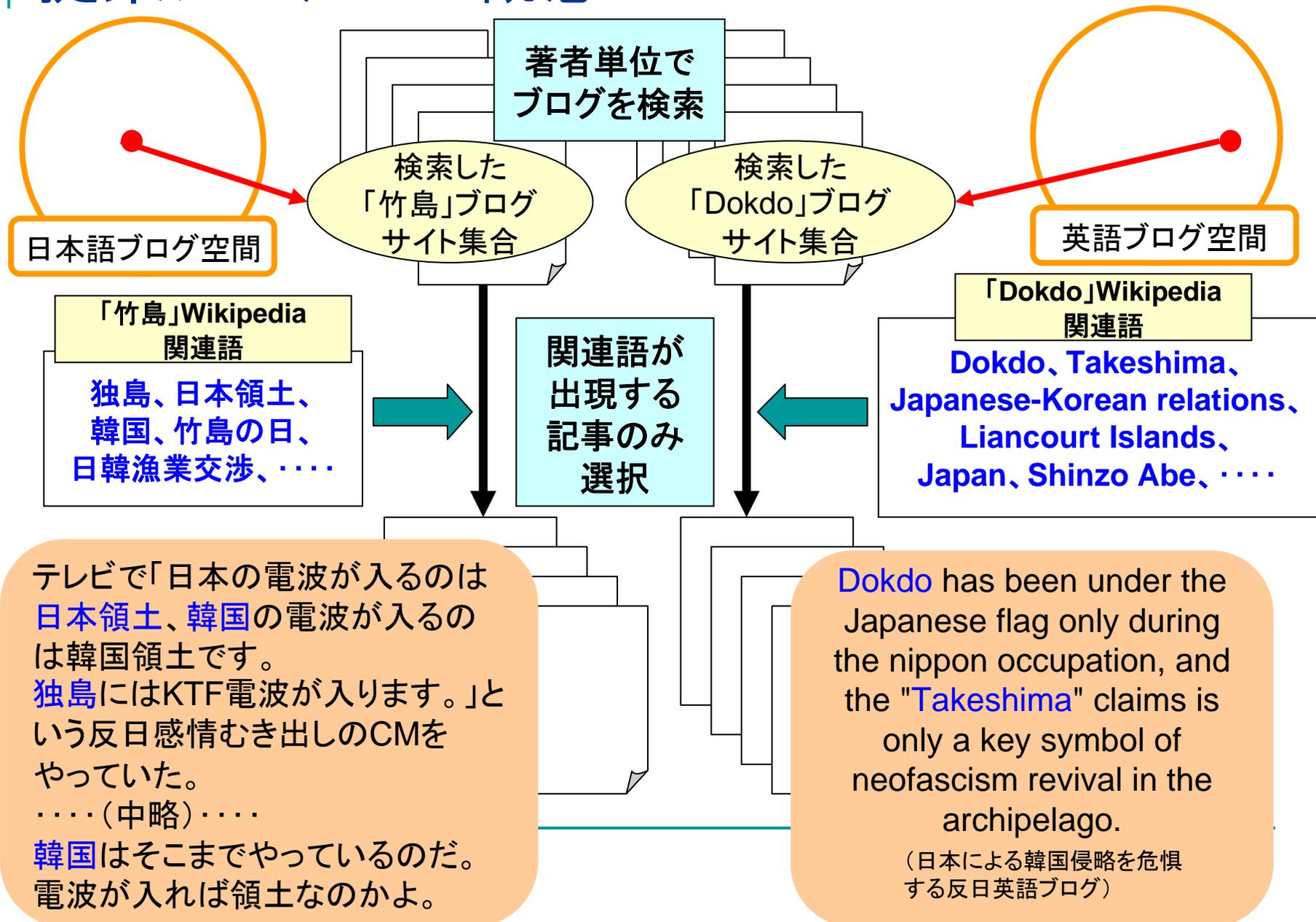
Sovereignty over the islands is disputed between Japan and South Korea. The islet is known as Dokdo or Tokto in Korean and Takeshima in Japanese.

「竹島」の関連語となる語句を
Wikipedia本文から抽出

独島、日本領土、
韓国、独島の月、
日韓漁業交渉、……

Dokdo, Takeshima,
Japanese-Korean relations,
Liancourt Islands,
Japan, Shinzo Abe, ……

提案システムの概念



提案システムの概念

テレビで「日本の電波が入るのは日本領土、韓国の電波が入るのは韓国領土です。独島にはKTF電波が入ります。」という反日感情むき出しのCMをやっていた。
……(中略)……
韓国はそこまでやっているのだ。電波が入れば領土なのかよ。

Dokdo has been under the Japanese flag only during the nippon occupation, and the "Takeshima" claims is only a key symbol of neofascism revival in the archipelago.

(日本による韓国侵略を危惧する反日英語ブログ)

ブログ記事本文から共起語を抽出

独島、日本領土、韓国、韓国領土、反日感情、……

Dokdo、Takeshima、nippon occupation、neofascism revival、……

Wikipediaの中立的観点の知識をもとに、文化間ギャップ発見支援となり得る偏った意見を表す共起語を取得

提案システムの概念

- Wikipediaの中立的観点の知識をもとに、ブログ本文に出現した偏った意見を表す共起語を取得
- 5つの社会問題系のトピックに対して実験
 - 捕鯨
 - バイオエタノール
 - 遺伝子組換え食品
 - 中絶
 - 臓器移植

提案システムの概要

特定トピックのブログサイト検索



Wikipediaからの関連語によるブログ記事検索



ブログ記事内の共起語抽出



共起語によるブログサイト・ブログ記事ランキング

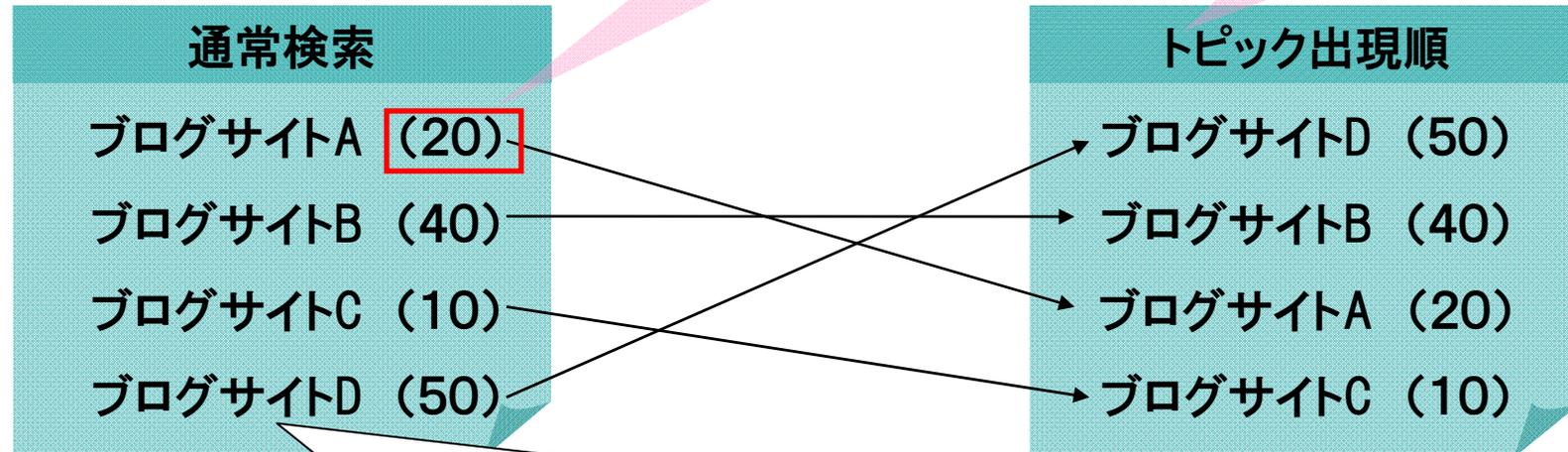
特定トピックのブログサイト検索手法

1. 様々な分野のWikipediaエントリから選定したトピックで検索し、ブログサイト集合を取得
 - 被リンク数の多い人気サイトから優先的に検索される（Yahoo!検索API使用）
2. 取得した各ブログサイト内でトピックが何件出現したかを調べる
3. ブログサイト内でトピック出現数が多い順に並び替える

人気度が高く、
検索トピックと適合する
ブログサイト

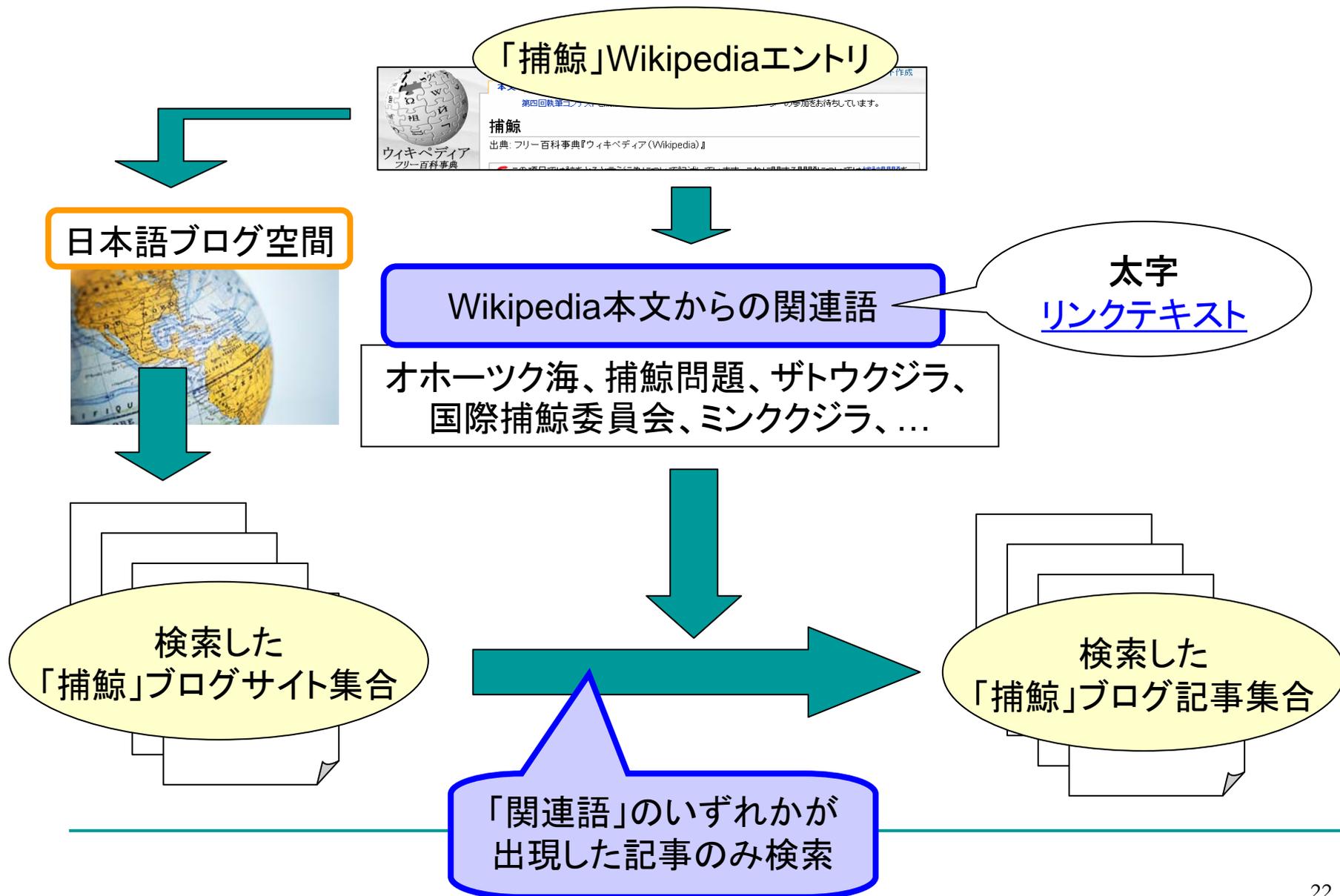
ブログサイト内での
トピック出現数

検索トピックの
出現数が多いブログサイト

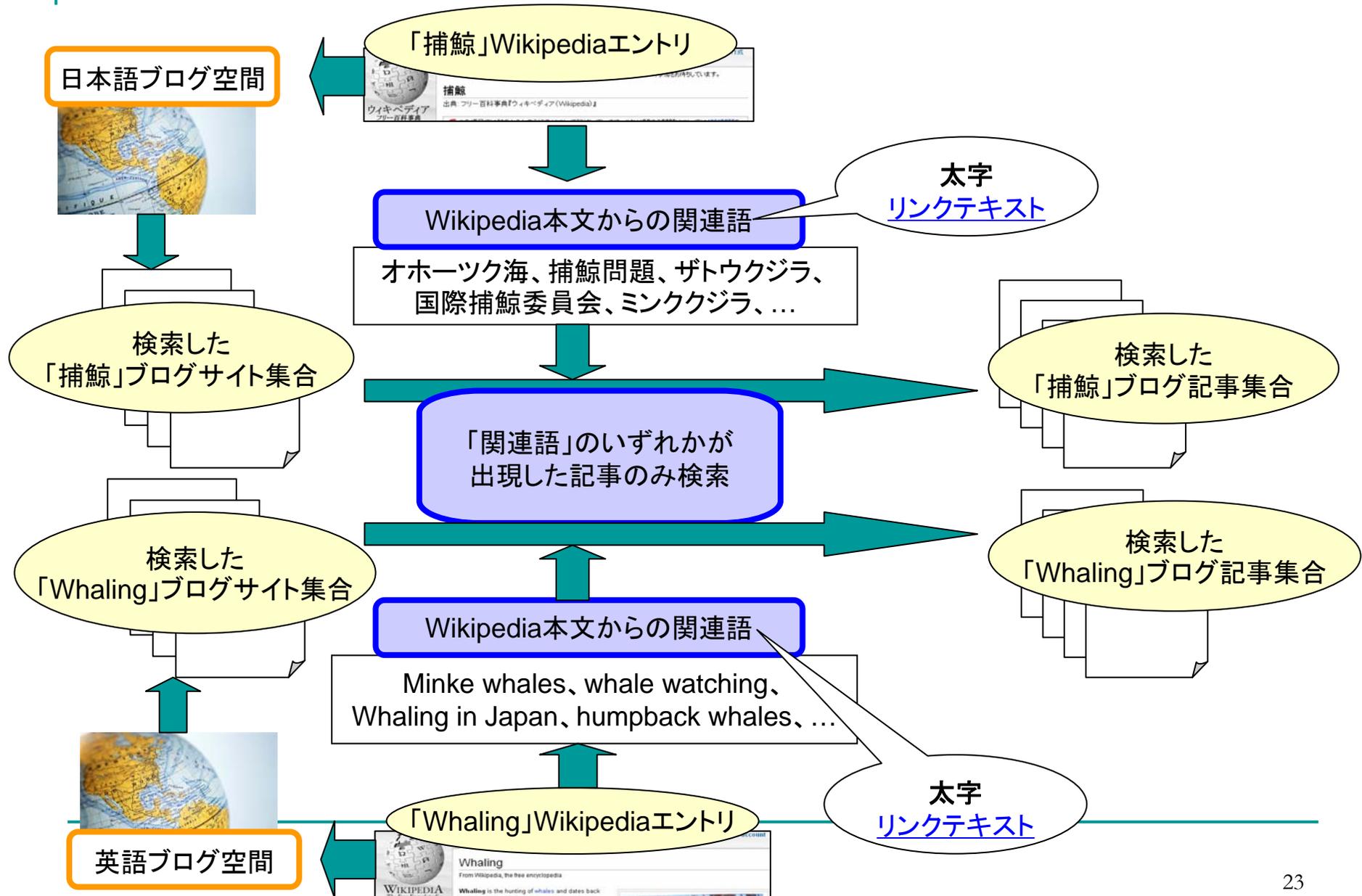


ブログサイト：人がブックマークするような
同一著者によって書かれたブログ記事集合のトップページ

Wikipediaからの関連語によるブログ記事検索:手法



Wikipediaからの関連語によるブログ記事検索:手法



ブログ記事内の共起語抽出

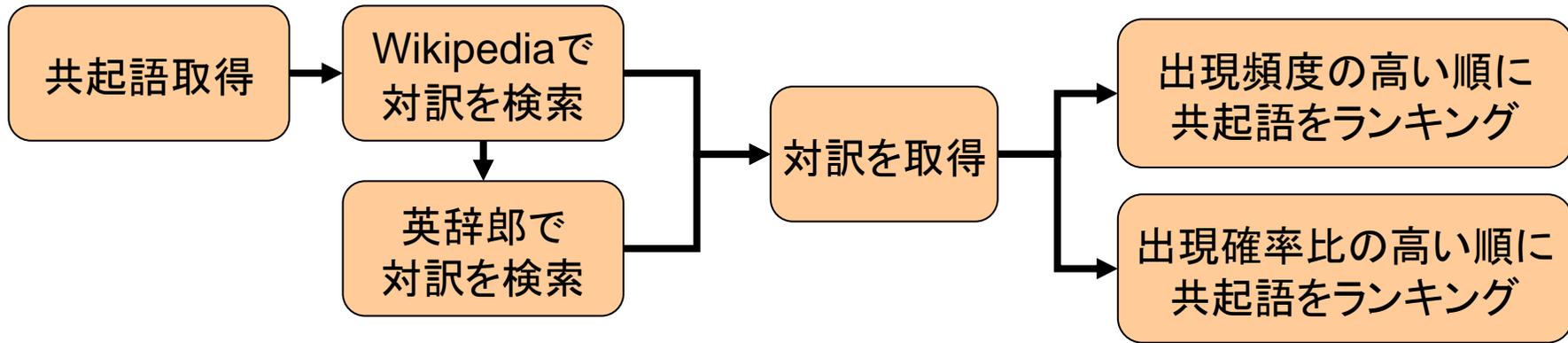
1. 同一トピックの日英ブログ記事集合から共起語を抽出。
 - 日本語ブログ記事集合からは名詞句を抽出。
 - 英語ブログ記事集合からは一単語・二単語連語・三単語連語を抽出。
2. 抽出した共起語をランキングするために、ランキング尺度として共起語に出現頻度と出現確率比を付与する。

出現頻度 = いくつ共起語が**単言語**のブログ記事集合内で出現したか？

出現確率比 = どれだけ共起語が**二言語間で比較して**
片方の言語のブログ記事集合内**のみ**で出現したか？

ブログ記事内の共起語抽出

■ 例: 捕鯨



共起語(日)	出現頻度(日)	出現確率(日)	共起語対訳	出現頻度(英)	出現確率(英)	出現確率比
反捕鯨国	136	0.000027	antiwhaling country	0	0	∞
調査捕鯨	620	0.000123	research whaling	59	2.259×10^{-5}	5.46
鯨肉	306	0.000061	whale flesh	1	3.829×10^{-7}	159.1
乱獲	81	0.000016	indiscriminate fishing	1	3.829×10^{-7}	42.1

ブログ記事内の共起語抽出

■ 例：捕鯨 (Whaling)

共起語(日)	出現頻度(日)	出現確率(日)	共起語対訳	出現頻度(英)	出現確率(英)	出現確率比
反捕鯨国	136				0	∞
調査捕鯨	620	0.000123	research whaling	59	2.259×10^{-5}	5.46
鯨肉	306	0.000061	whale flesh	1	3.829×10^{-7}	159.1
乱獲	81	0.000016	indiscriminate fishing	1	3.829×10^{-7}	42.1

反捕鯨国の団体が
調査捕鯨の邪魔をしている

共起語(英)	出現頻度(英)	出現確率(英)	共起語対訳	出現頻度(日)	出現確率(日)	出現確率比
Sea			シーシェパード	191	0.000038	7.63
Endangered species	302	6.12×10^{-6}	絶滅危惧種	47	9.35×10^{-6}	12.36
Animal welfare	188	0.000072	動物保護	7	1.39×10^{-6}	51.7
Whale sanctuary	101	0.000039	捕鯨禁止海域	2	3.98×10^{-7}	97.1

捕鯨によるクジラの絶滅を
危惧している

共起語によるブログサイト・ブログ記事ランキング

2種類の尺度でランキングされた共起語リストと
それを基に人手で選択した共起語リスト

人手選択

1. 調査捕鯨
2. 鯨肉
3. 反捕鯨国
- ・
- ・

出現確率比順

1. 批判
2. 責任
3. 鯨肉
- ・
- ・

出現頻度順

1. 捕鯨
2. 批判
3. ニュース
- ・
- ・

共起語抽出

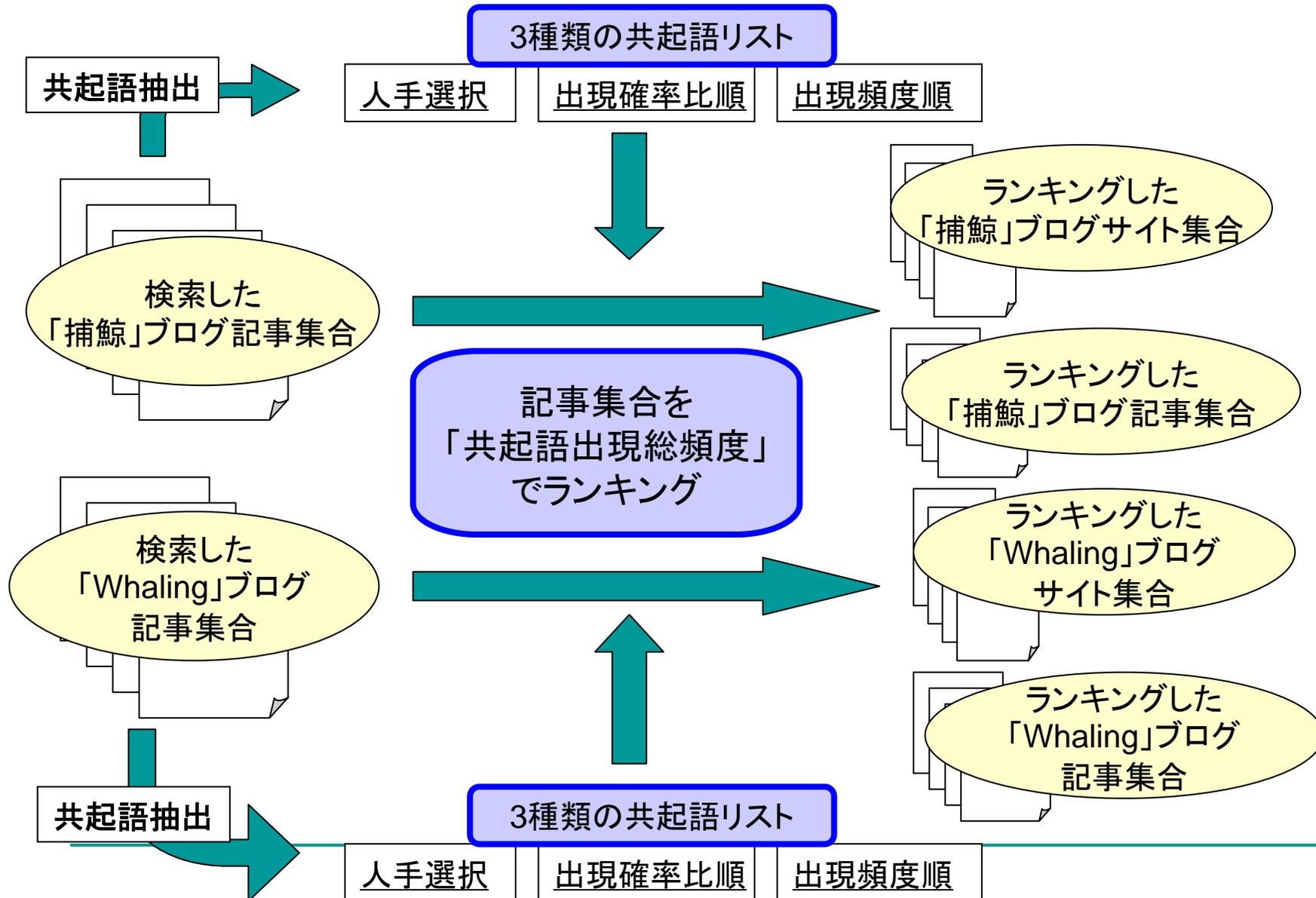
検索した
「捕鯨」ブログ記事集合

記事集合を
「共起語出現総頻度」
でランキング

ランキングした
「捕鯨」ブログサイト集合

ランキングした
「捕鯨」ブログ記事集合

共起語によるブログサイト・ブログ記事ランキング



「捕鯨」ブログ
サイトランキング

「Whaling」ブログ
サイトランキング

日本語ブログサイト1位

反捕鯨団体の反捕鯨論
を批判。
捕鯨は、日本の伝統的な
文化だ！！

1位

日本語ブログサイト2位

捕鯨反対派の主張は、科
学的根拠が無い！
捕鯨には賛成だが、伝統
的文化ではないだろう。

2位

日本語ブログサイト3位

反捕鯨団体の妨害に
屈してはならない！
捕鯨は続けるべきだ！

3位

・
・
・

英語ブログサイト1位

動物をこよなく愛する！
鯨も動物だから、
殺すべきではない！
反捕鯨団体を応援。

英語ブログサイト2位

動物にはみんな生きる
権利がある！
もちろん鯨も例外ではな
いので、捕鯨反対！

英語ブログサイト3位

捕鯨反対の意思を綴った
本の著者によるブログ。
オーストラリアの捕鯨廃
止に貢献

・
・
・

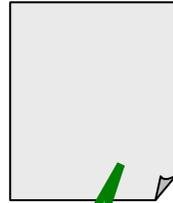
日本語ブログ記事1位

同じ団体や個人が欧米のメディアでは、あいかわらず**文化帝国主義**丸出しの主張や「日本の**調査捕鯨**は偽装した**商業捕鯨**だ」という国際法的には荒唐無稽な主張を繰り返しているからです。

「捕鯨」ブログ記事ランキング



1位



2位



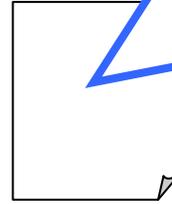
3位

日本語ブログ記事2位

オーストラリアのIan Campbell環境大臣への対応についても、**日本捕鯨協会**による非難声明も悪くは無いが、併せて**調査捕鯨**の正当性や、**捕鯨文化**の重要性をしっかりとアピールするべきだ。

- ・
- ・
- ・

「Whaling」ブログ記事ランキング



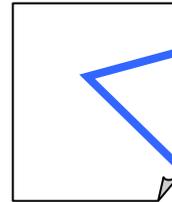
英語ブログ記事1位

the only organization that lately has done something to physically enforce the **whaling moratorium** and truly protect the whales is The **Sea Shepherd Society**



英語ブログ記事3位

Why does **Sea Shepherd** harass the **Japanese** and not the **Europeans**? Do they find it easier to bully Asians? Or do they know that the **Japanese** won't fight back?



- ・
- ・
- ・

共起語によるブログサイト・ブログ記事ランキング: 考察

■ ブログ記事ランキング

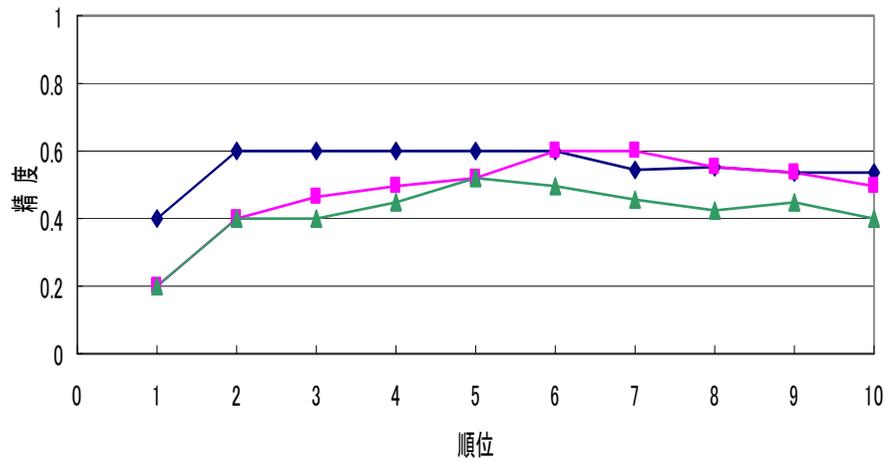
- 人手によって共起語を選択することで、ランキングの精度を大きく向上

■ ブログサイトランキング

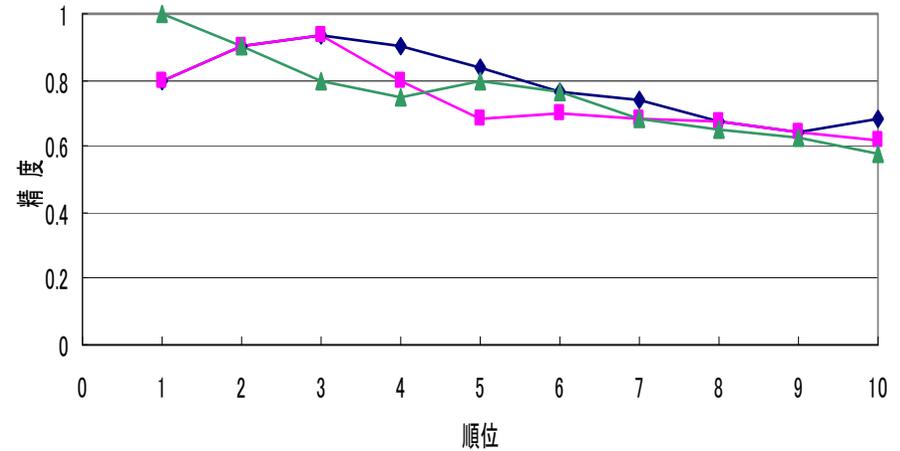
- 記事ランキングとは異なり、自動で選択した共起語を用いても人手で選択した共起語によるランキングと同等のランキング精度を取得できた

捕鯨、バイオエタノール、中絶、臓器移植、遺伝子組換え食品

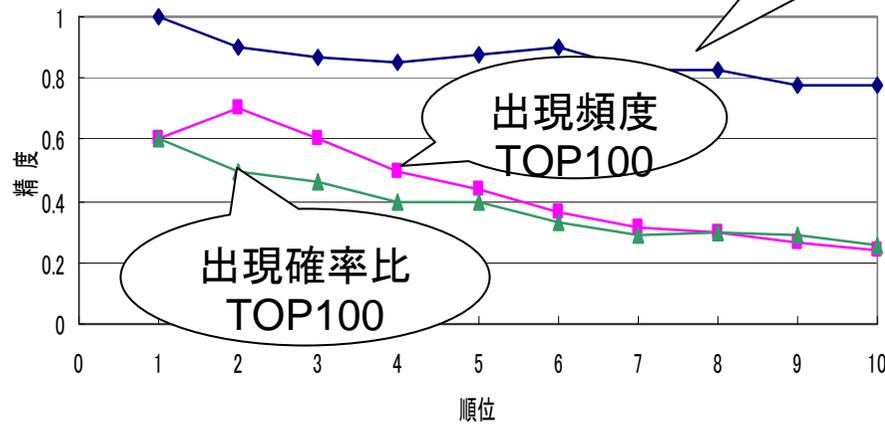
日本語サイトランキング(5トピック平均)



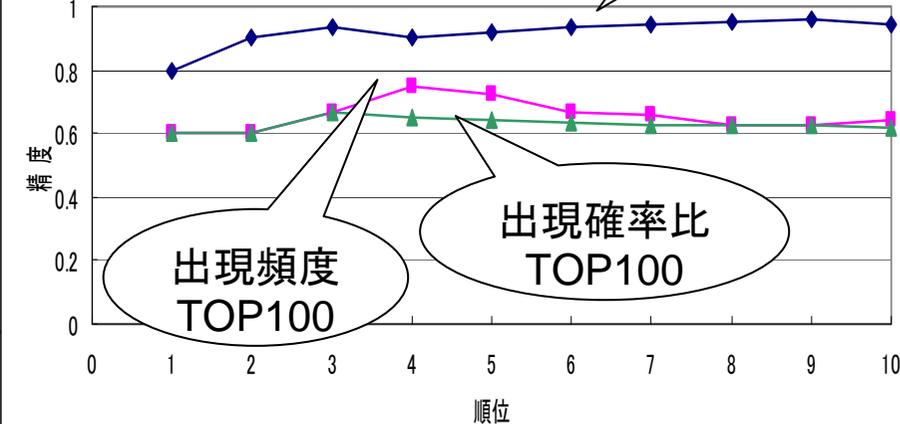
英語サイトランキング(5トピック平均)



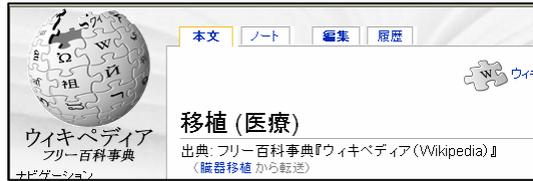
日本語記事ランキング(5トピック平均)



英語記事ランキング(5トピック平均)



Wikipedia関連語・ブログ共起語自動抽出例：臓器移植



日本語関連語

日本臓器移植ネットワーク、
脳死移植、再生医学、腎臓移植、
心臓移植、生体肝移植……

選択した「臓器移植」
ブログ記事集合

日本語共起語

病気腎移植、臓器移植法、
脳死移植、倫理委員会、脳死判定、
ドナーカード、生体肝移植、……



英語関連語

organ donation、heart transplant、
living donor、A Gift from Earth、
donor card、eye bank、……

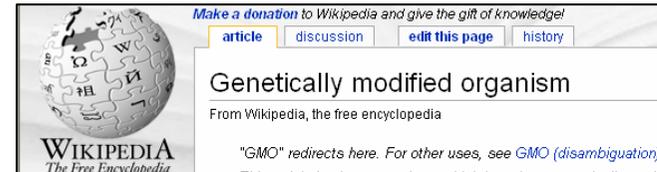
選択した「Organ transplant」
ブログ記事集合

英語共起語

Falun Gong、organ donation、
organ harvesting、living donor、
heart transplant、human rights、
donor waiting list、……

赤文字:
Wikipediaにない
関連語

Wikipedia関連語・ブログ共起語自動抽出例：遺伝子組換え食品



日本語関連語

遺伝子組換え作物、とうもろこし、
バイオテクノロジー、厚生労働省、
アメリカ、遺伝子組換え、……

英語関連語

GM crops、agriculture、
European Union、GMO foods、
genetic engineering、GMO、……

ブログ空間

検索した
「遺伝子組換え食品」
ブログ記事集合

検索した
「Genetically modified organism」
ブログ記事集合

日本語共起語

検査、安全性、厚生労働省、
遺伝子組換え、危険、賛成、
アメリカ、とうもろこし、脅威、……

英語共起語

damage biodiversity、GMO、
human health、European Union、
GM crops、danger、banned、
ethical、environmental consequence、
genetic engineering、…

赤文字:
Wikipediaに
ない関連語

言語間差異部分抽出例：遺伝子組換え食品

日本語記事1

以前は世間一般の評価を信じて、「**遺伝子組換え**」には反対であった。しかし、**厚生労働省**が許可したものは十分に**検査**しており、「逆にふつうの食品より**安全性**が高い」ということ、「**遺伝子組換え**商品の表示を、世間の人**危険**なものであると**誤解**している」ことなどを学び、私は**遺伝子組換え賛成**の考えを持っている。

世間一般には先行したイメージがあるが、それとは違ったため賛成

英語記事1

In the latest survey done across the **European Union**, majority of consumers, including myself, were against **genetically modified organism** (GMOs) due to fears that **GM crops** can **damage biodiversity** and be a **danger** to **human health**.

種の多様性と人間の健康に害を及ぼす恐れがあるので、EUではGMOに反対している。

日本語記事2

先日放送された番組では、**アメリカの遺伝子組換え**作物に関する情報はエタノール用**とうもろこし**などで触れられている部分のみであり、本当の**遺伝子組換え**作物の**危険性**については触れられていない。我々消費者は、実態のつかめない**遺伝子組換え**食品を食べなければならないといった**脅威**にさらされているのである

アメリカの遺伝子組換え作物の危険性が日本ではちゃんと報道されていない。

英語記事2

The more frivolous use **genetic engineering** of plants and animals needs to be **banned** and the rest **regulated** more strictly because of their significant **ethical**, health and **environmental consequences**.

食品の基となる植物や動物の遺伝子組換えは倫理的にも環境的にも良くないのでやめるべき！

Wikipedia関連語・ブログ共起語自動抽出例：バイオエタノール



日本語関連語

バイオエタノール、とうもろこし、
資源、温室効果ガス、
地球温暖化、二酸化炭素、.....

英語関連語

biofuel、greenhouse gas、
fertilizer、ethanol、liquid fuels、
hydrogen technology、potato、.....

ブログ空間

検索した
「バイオエタノール」
ブログ記事集合

検索した
「Ethanol fuel」
ブログ記事集合

日本語共起語

バイオエタノール、京都議定書、
利益、とうもろこし、高騰、アメリカ、
推進、排気ガス、地球温暖化、圧迫、
穀物価格、ブラジル、二酸化炭素、...

英語共起語

dependence on oil、biofuel、
grow corn、greenhouse gas、
fertilizer、genuine sustainability、
ultimately eliminate、ethanol、
destroy the ecosystem、...

言語間差異部分抽出例: バイオエタノール

日本語記事1

ガソリンにバイオエタノールを混ぜたからといって、**排気ガス**がクリーンに成る訳ではない。バイオエタノールの原料である**トウモロコシ**やサトウキビが、成長過程で**二酸化炭素(CO2)**を吸収するため、**地球温暖化**を招かないとされるだけの理由だ。むしろ食品・穀物**価格の高騰**を招き、生活への**圧迫・影響**が大きく、**マイナス面**が多いと思うのは自分だけだろうか。

バイオエタノールは食品価格高騰につながり、国民の生活を圧迫していると指摘

英語記事1

Biofuel from ethanol produces 70% less **greenhouse gas** emissions than gas or diesel, and cuts our **dependence on oil**. It should be one facet of a multi-faceted strategy to cut our **greenhouse gas** emissions, to move in the direction of **genuine sustainability**, and to reduce and **ultimately eliminate** our **dependence on oil**.

バイオエタノールを使用することで温室効果ガスも削減できる上に原油依存を解消するため賛成。

日本語記事2

地球温暖化なんてどうでもいいと考えている国が、**バイオエタノール**を**推進**しているのだ。その国とは、**アメリカとブラジル**です。ブラジルは**二酸化炭素**の削減を批准国に義務づける**京都議定書**の2012年からの第二次計画に参加することを「**約束**」しているだけ。アメリカは2001年に京都議定書から脱退済み。自分達の**利益**のみを追求しているのが明白だ。

環境を考えているようで利益のみを追求してバイオエタノールを推進する国を批判

英語記事2

It is my conclusion that we are going to **destroy the ecosystem** in the **Gulf of Mexico** and in rivers and lakes and seas and bays around the world if we continue to **grow corn** for **ethanol**. Just the **fertilizer** alone places far too great a burden on the rivers and streams.

バイオエタノール原料のトウモロコシを増産することで肥料の影響からメキシコ湾や世界中の海や河川の生態系を破壊することになる。

現状

- 各言語で共起語自動抽出をしているだけ
- 言語間で差があるトピックの場合に、
各言語に特有な記述が偶然上位に現れただけ

今後の展開

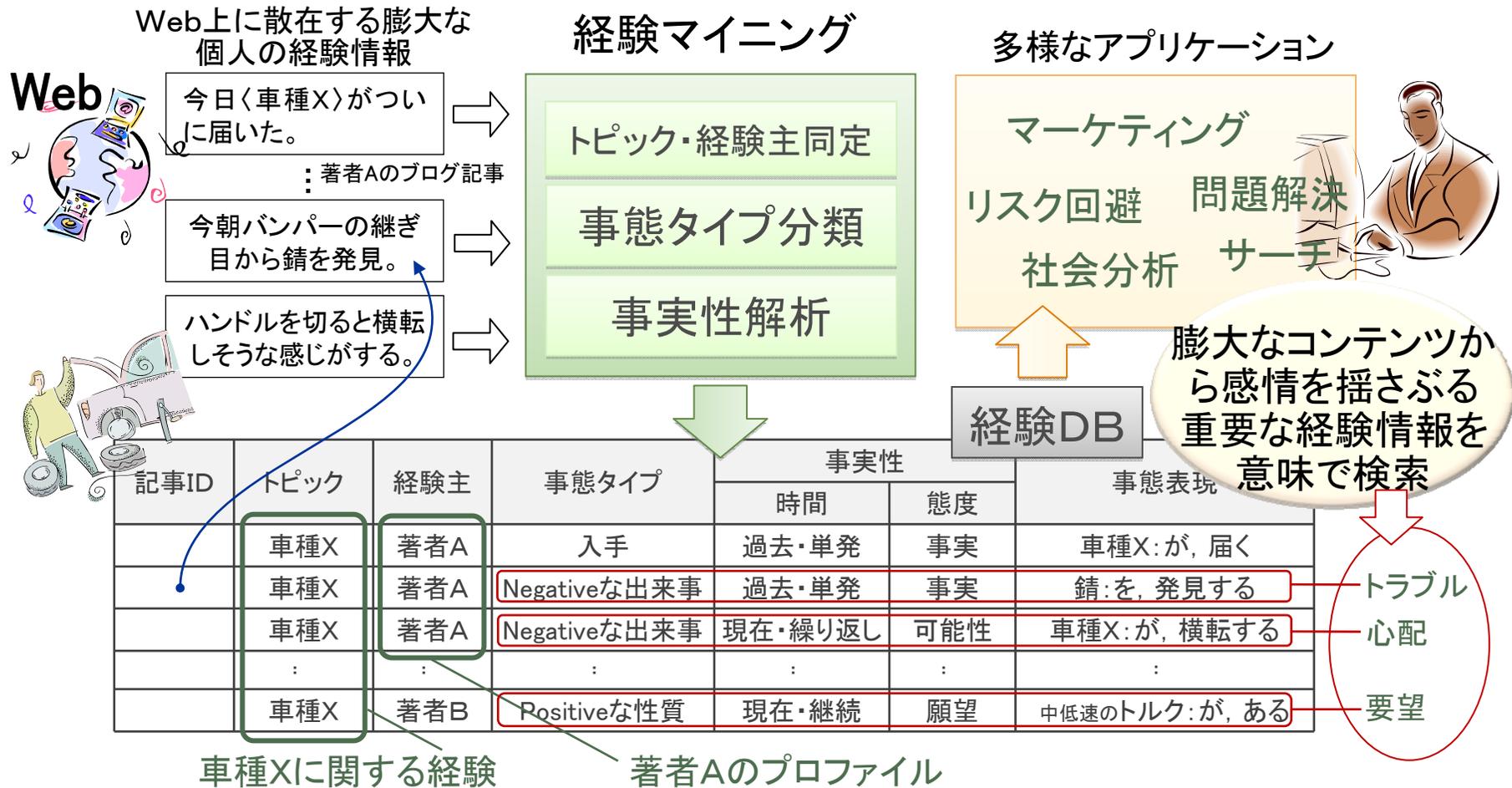
- 主観情報を多く含む箇所の抽出
- 日本語と英語で差のあるトピック、差の無いトピックを
自動検出する

主観情報抽出

⇒ 経験マイニング(乾)の要素技術を適用

経験マイニング (Inui&Abe+ 08)

Web上の膨大なUGCから人々の**経験情報**を収集し、
意味的に索引付けてDB化する



経験検索サイト「みんなの経験」

- ブログデータ: 1億5000万記事 (最近1年分)
- 対象トピック: 15万語 (Wikipedia)
- 収集した経験数: 6000万件
- 10月に一般公開予定



[使い方](#) | [概要\(PDF\)](#) | [詳細\(PDF\)](#) | [infoplusion](#)

i pod 検索

次のトピックと比較してみたいかでしょうか?

携帯型プレーヤー: ipod ウォークマン 電池バック 携帯音楽プレーヤー siren デジタルオーディオプレーヤー gigabeat creative rio wizpy roly アイリバー zune zerocore x-cute siren ビデオウォークマン

音響機器: サウンド アンプ ipod ヴァイオリン ヘッドフォン サイレン ステレオ ミキサー エフェクター my オルゴール ヘッドセット ガジェット

cdプレーヤー 補聴器 ホームシアター icレコーダー メガホン シンセサイザー eclipse

ipod: ipod ipod nano ipod shuffle ipodwizrd ipodlinux ipod管理ソフト

3個のトピックを比較した結果 [Prev](#) [Next](#)

🔍 **i p o d** [すべての経験を見る \(682人\)](#) [i p o dをWikipediaで調べる](#)

❤️	🔴 良さそう (経験前) 8人	良い 6 復活する 1 効く 1
👉	🔴 買う/利用するつもり 53人	買う 29 使う 8 購入する 4 行く 3 楽しむ 3 手に入れる 2 使用する 2
👉👉👉👉	🔴 買った/利用した 193人	入る 67 買う 56 購入する 19 使う 17 手に入れる 9 使用する 5 観る 3
🎵	🔴 役立った 10人	再生できる 2 足りる 1 楽しめる 1 持ち歩ける 1 手に入る 1 復活する 1 当選する 1
👉👉👉👉	🔴 良かった (感想) 165人	良い 38 安い 16 楽しい 15 嬉しい 12 使える 12 買える 8 素晴らしい 7
👉	🔴 お勧めします 4人	買う 3 使う 1
😭	🔵 悪そう (経験前) 11人	泣く 4 悪い 2 負ける 1 炎上する 1 滑る 1 汚れる 1 壊れる 1
⚠️⚠️⚠️⚠️⚠️	🔵 トラブった 145人	壊れる 36 悩む 12 切れる 12 故障する 6 飽きる 4 肺める 4 壊す 4
👎👎	🔵 悪かった (感想) 53人	悪い 30 痛い 5 寂しい 4 つまらない 3 汚い 2 見づらい 1 臭い 1
👎	🔵 薦めません 1人	買う 1

「みんなの経験」

■ 世界初のオープンドメイン経験データベース

みんなの経験

ホーム | 検索 | アンケート | 技術の詳細 | info@plaza.co.jp

白骨温泉 新平湯温泉

検索

2個のトピックを比較した結果

白骨温泉 [すべての経験を見る \(148人\)](#) [白骨温泉をWikipediaで調べる](#)

- 買う/利用するつもり** 22人
- 買った/利用した** 79人
- 良さそうな噂** 0人
- 良かった感想** 39人
- 良かった出来事** 37人
- 悪そうな噂** 0人
- 悪かった感想** 6人
- 悪かった出来事** 25人

新平湯温泉 [すべての経験を見る \(148人\)](#) [新平湯温泉をWikipediaで調べる](#)

- 買う/利用するつもり** 22人
- 買った/利用した** 79人
- 良さそうな噂** 0人
- 良かった感想** 39人
- 良かった出来事** 37人
- 悪そうな噂** 0人
- 悪かった感想** 6人
- 悪かった出来事** 25人

行く 21 断念 2 魅力 1 観光 1 見物 1 自慢 1 病気 1 気持ち 1
 日 1 旅行 1 好き 1 入る 1 人気 1 ため 1
 行く 55 入る 19 旅行 4 風呂 3 浸かる 3 効果 3 結構 2 無料 2
 最高 2 好き 2 偽装 2 事件 2 ハッピー 2 関節痛 1
 良い 14 楽しめる 6 有名 4 満腹 3 風情 3 楽しい 3 風情 3
 面白い 2 邪魔 1 有名になる 2 旅情 2 散歩 2 安い 2 一番 2
 気になる 4 騒動 3 落ち着く 3 胃腸病 3 病 3 効能 3 効果がある 3
 効果 3 風呂 2 風情 2 美しい 2 無料 2 楽しみ 2 匂 2
 苦しい 2 満腹 2 非常 1 静か 1 死 1 悪い 1 弱 1 友人 1 厄介 1
 なる 1 厄介 1 価値 1 乏しい 1 不老 1 やかましい 1
 問題 6 発生する 4 発生 4 源泉 4 偽装 4 事件 4 事件がある 4
 騒がれる 2 選れる 2 混む 2 ける 2 負 2 専門 2 畳 1

「入浴剤偽装」で信用を落とした白骨温泉

<http://blog.kansai.com/ukrockthebeates/> **良かった出来事 3経験**

「白骨温泉」に関するこのブロガーの経験

良かった出来事 解析結果: 経験=ポジ (経験) 時間=過去 (瞬間)

有名な**白骨温泉**へ向かう。今ではすっかり有名になって、もはや**秘湯**、**秘湯**を思わせる。

<http://blog.kansai.com/ukrockthebeates/129> 2007-11-09 22:14:47

良かった出来事 解析結果: 経験=ポジ (経験) 時間=現在 (状態)

今回は初めて**白骨温泉**の湯を**堪能**した。

<http://blog.kansai.com/ukrockthebeates/128> 2007-11-07 22:28:34

良かった出来事 解析結果: 経験=ポジ (経験) 時間=現在 (状態)

温泉**偽装**で**白骨温泉**はテレビでもかなり取り上げられたが、ある**意味**その**おかげ**で各地の温泉成分や掛け流しかどうかなど、正しい温泉表記を求められることになったのは**皮肉**だが**意義**があると思う。

<http://blog.kansai.com/ukrockthebeates/128> 2007-11-07 22:28:34

<http://www.cosmeticlife-hyper.com/> **良かった出来事 2経験**

「白骨温泉」に関するこのブロガーの経験

良かった出来事 解析結果: 経験=ポジ (経験) 態度=断定

白骨温泉を後にし、せっかくここまで来たのでもう少し長野県を**堪能**して帰ろうと松本を經由して目指すは長野市。

<http://www.cosmeticlife-hyper.com/2008/02/25/002554.php> 2008-02-25 02:55:04

良かった出来事 解析結果: 経験=ポジ (経験) 態度=断定

4年前の**白骨温泉**騒動の時も温泉の元を入れてなかったらしいし、**老舗**らしい**こだわり**を感じるきちんとしたお宿という印象でした。

<http://www.cosmeticlife-hyper.com/2008/02/24/192047.php> 2008-02-24 19:20:47

<http://onsenyoyzi.blog81.fc2.com/> **買った/利用した 1経験** **良かった出来事 1経験**

「白骨温泉」に関するこのブロガーの経験

良かった出来事 解析結果: 経験=ポジ (経験) 時間=現在 (状態) 態度=断定

まるかやな肌触りで浸かっているとボカボカと**温まってきます** (一では無いですがね。

<http://onsenyoyzi.blog81.fc2.com/blog-entry-100.html> 2007-06-24 23:00:00

<http://plaza.rakuten.co.jp/tanabata777/> **良かった出来事 1経験**

「白骨温泉」に関するこのブロガーの経験

良かった出来事 解析結果: 経験=ポジ (経験) 時間=過去 (瞬間) 態度=断定

白骨温泉独特のブルーの白濁の温泉に浸かりながら、白樺林を眺める**至福**のひと時が**味わ**えました。

信用回復後の利用客の生の声



主観情報抽出

⇒ 経験マイニング(乾)の要素技術を適用

- 検出したい情報・・・多様な主観情報
- 経験マイニングの要素技術
 - 事態タイプ分類(主観を含む)
 - 事実性解析(モダリティ含む)
- 多言語(英語)化、よりメタレベルの情報との対応付け

	日本語	韓国語	中国語	英語
ニュース				

多言語のトピック体系Wikipediaを

- ・ Wikipediaエン트리とブログサイトの対応付け [ICWSM2008, DEWS2008]
- ・ 日本語ブログ空間におけるトピック分布推定 [NL研2008年9月]

トピックの多様な特性の分析

スパムブログにおける
トピックのバースト特性の分析
[宇津呂・中川]

トピックの情報価値の質・持続性の
推定およびトピックの
視覚的二次元配置 [宇津呂]

トピックに関連する観点・
意見分布の抽出
[藤井・宇津呂]

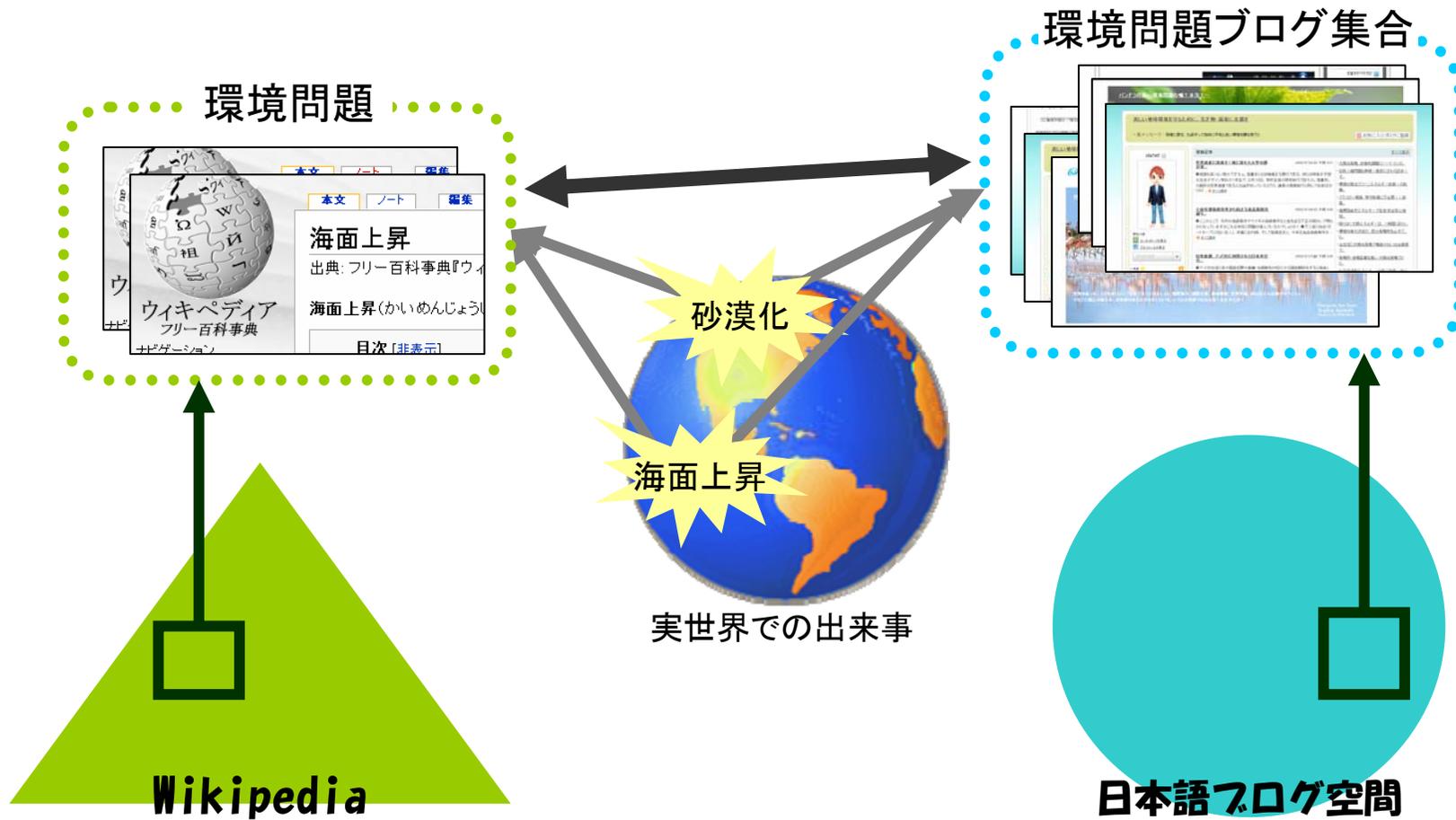
トピックの言語間差異度の測定

トピックの情報価値の質・
持続性の差異度
[宇津呂・中川・福原・清田]

トピックの持つ観点の分布の
差異度 [宇津呂]

トピックの持つ意見分布の
差異度 [藤井]

ブログ空間のトピック分布の推定



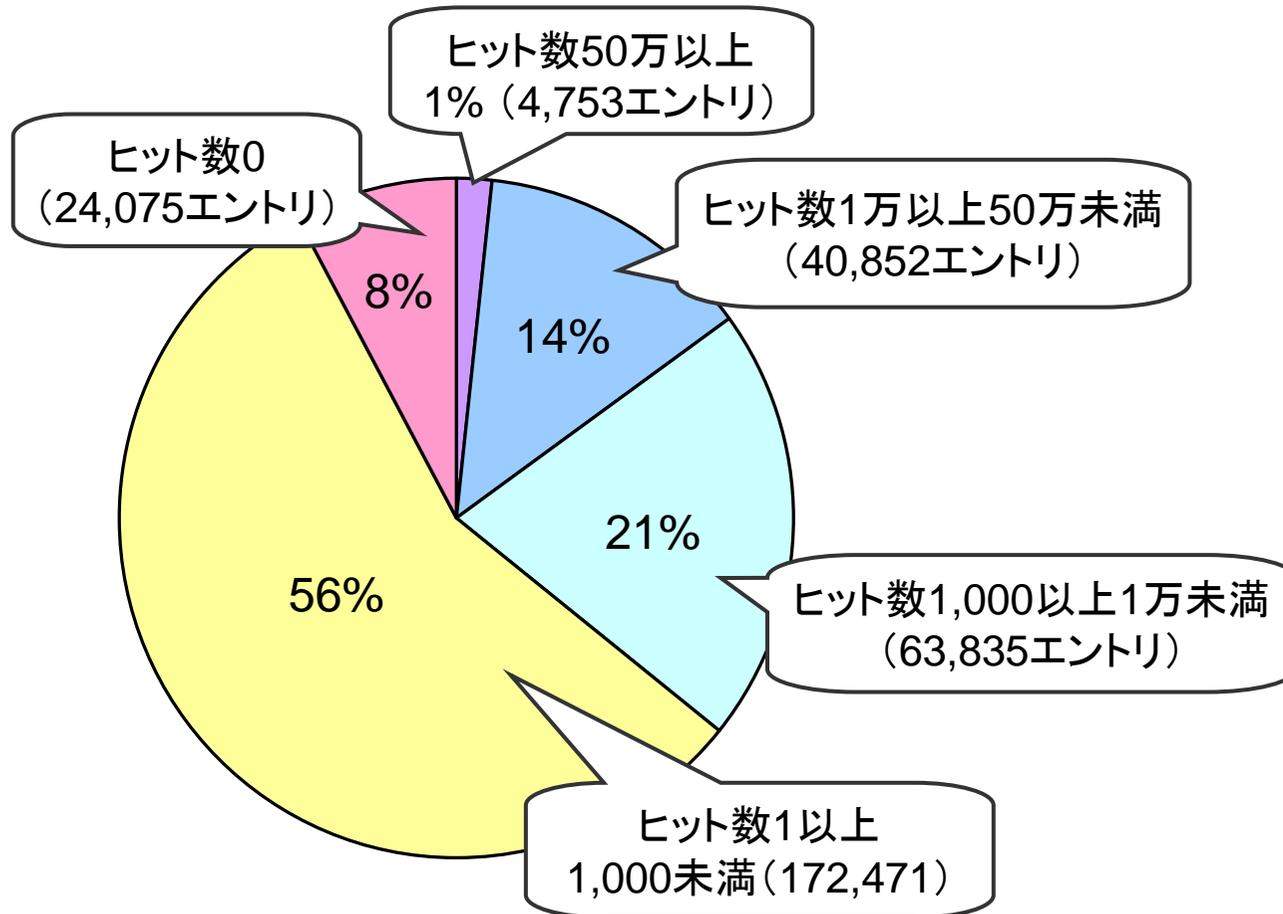
トピックのブログ検索ヒット数による ブログサイトの有無の推定

- ヒット数が大きいものには一般語が多い
- ヒット数が少ないものにはマイナーな語が多い

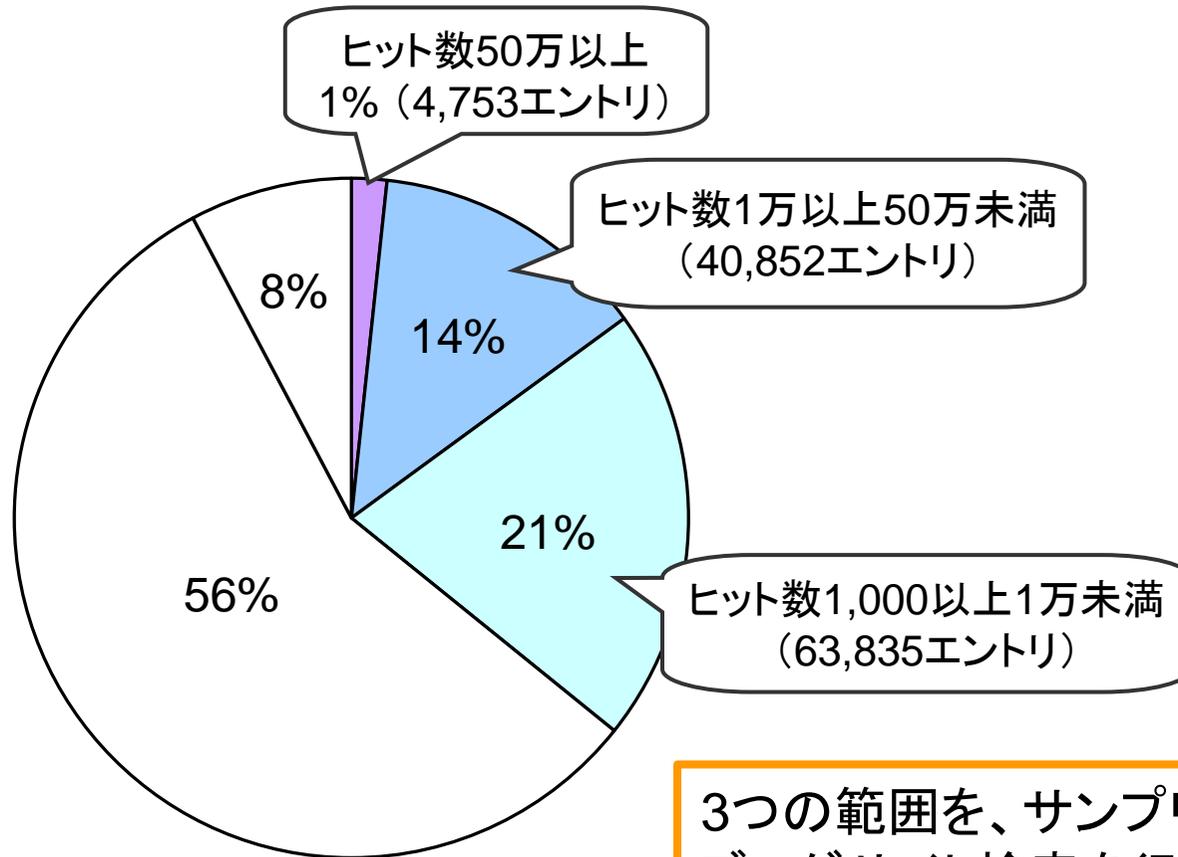


- ヒット数が**1万～50万**の範囲に、
ブログサイトがありそうなトピックが集中している

トピックのブログ検索ヒット数による ブログサイトの有無の推定

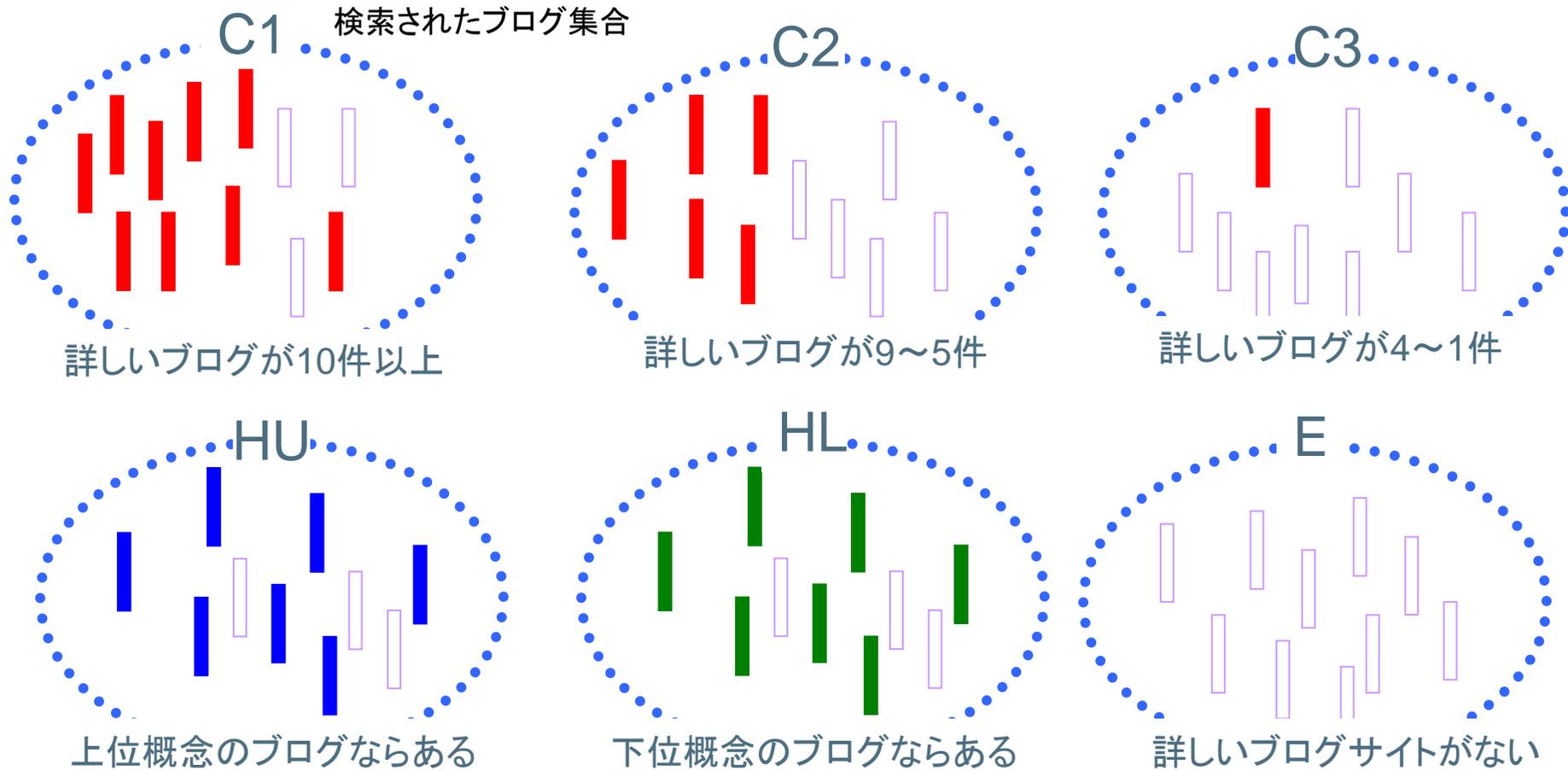


トピックのブログ検索ヒット数による ブログサイトの有無の推定



3つの範囲を、サンプリング。
ブログサイト検索を行い、6段階評価する。

評価基準



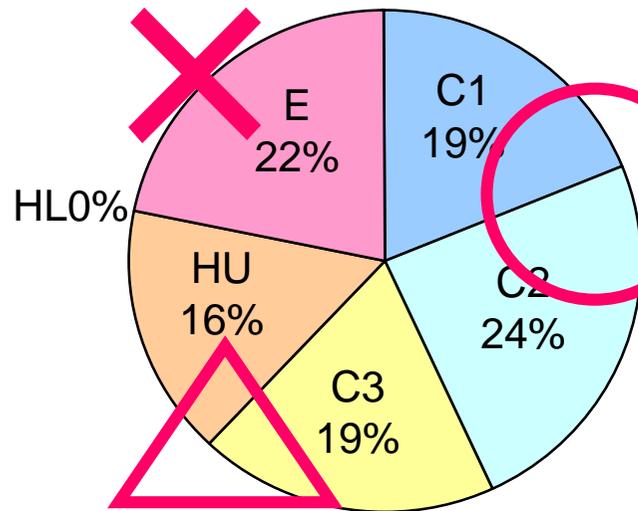
- トピックについて詳しく書かれたブログサイト
- 検索トピックの上位概念について書かれたブログサイト
- 検索トピックの下位概念について書かれたブログサイト
- 検索トピックの以外について書かれたブログサイト

評価結果

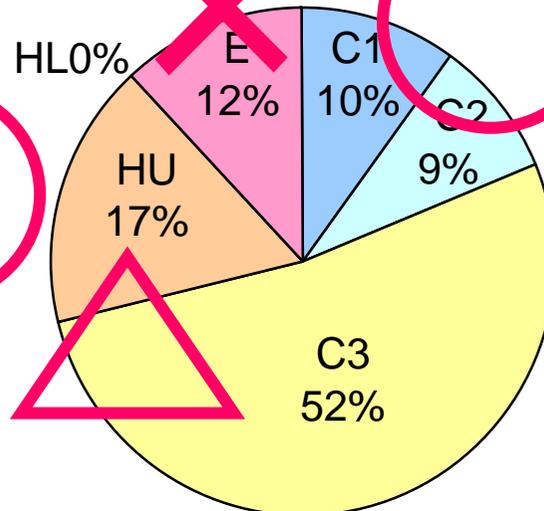
ブログサイトが5件以上あるトピックが多い

ブログサイトが
ごくわずかしかないトピックが多い

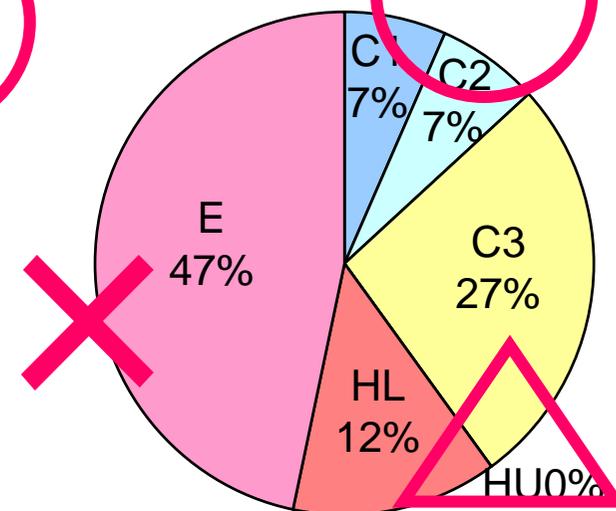
多くは一般語だが、「ピアノ」等のブログサイトが多いトピックも存在する



ヒット数1万~50万のエントリ



ヒット数1000~1万のエントリ



ヒット数50万以上のエントリ

4.6.1 ブログサイト例: ヒット数1万~50万

■ 生ゴミブログ

廃棄物分野での仕事をしています。ブログ歴ようやく2年に近くなりました。情報交換のため毎日の生活を通じて感じたことを発信しています。日々の日課として続いています。

自宅生ゴミ処理機 | 自宅生ゴミ処理機II | 生ゴミ処理機 | もみ殻段ボール | 剪定枝葉堆肥化 | コンポストの敵 | 職場の出来ごと | インドネシア | パキスタン | 家の出来事 | テレビ&ラジオ | 花や虫などのこと | 花の写真 | リサイクル | 未分類 |

ごみ問題取り組みの日々

▼ カテゴリ

- 自宅生ゴミ処理機 (243)
- 自宅生ゴミ処理機II (179)
- 生ゴミ処理機 (45)
- もみ殻段ボール (353)
- 剪定枝葉堆肥化 (15)

2008.05 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

▼ 最近の記事+コメント

6月26日もみ殻段ボール投入量235g
ヒベリカム(金糸梅)

6月27日もみ殻段ボール投入量145g
アジサイの花 +
トメロン♡ New!!
ト MataBola
トメロン♡

6月26日もみ殻段ボール投入量230g
6月24日もみ殻段ボール投入量185g

2008.06.28 Sat
6月26日もみ殻段ボール投入量235g



| HOME | 次の

年齢: 50代
職業: 地方公務員
業種: 試験検査機関の技術職員

「生ゴミ」について書かれたブログ
ゴミの処理方法や、家庭用生ゴミ処
理機について詳しく書かれている。

「生ゴミ」で検索したブログサイトは
エコ活動をしているブロガーによって
書かれたものが大半

4.6.2 ブログサイト例: ヒット数1000~1万

■ アフリカツメガエルブログ



4.6.3 ブログサイト例: ヒット数50万以上

■ ピアノブログ

ひとり語り ~松田真樹のピアノ語り~

日々のこと、好きなこと、音楽情報などを、ピアノを弾くようにお話できればと思っています。

カレンダー

<< 9月 >>

日	月	火	水	木	金	土
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

最近の記事一覧

- 秋肌・美肌
- 夢の家
- レディース神風
- 楽しい風邪

アルバム「PIANO FOLK SON...

店頭がない際はご予約い...

Amazon、TOWER RECORDSまた...

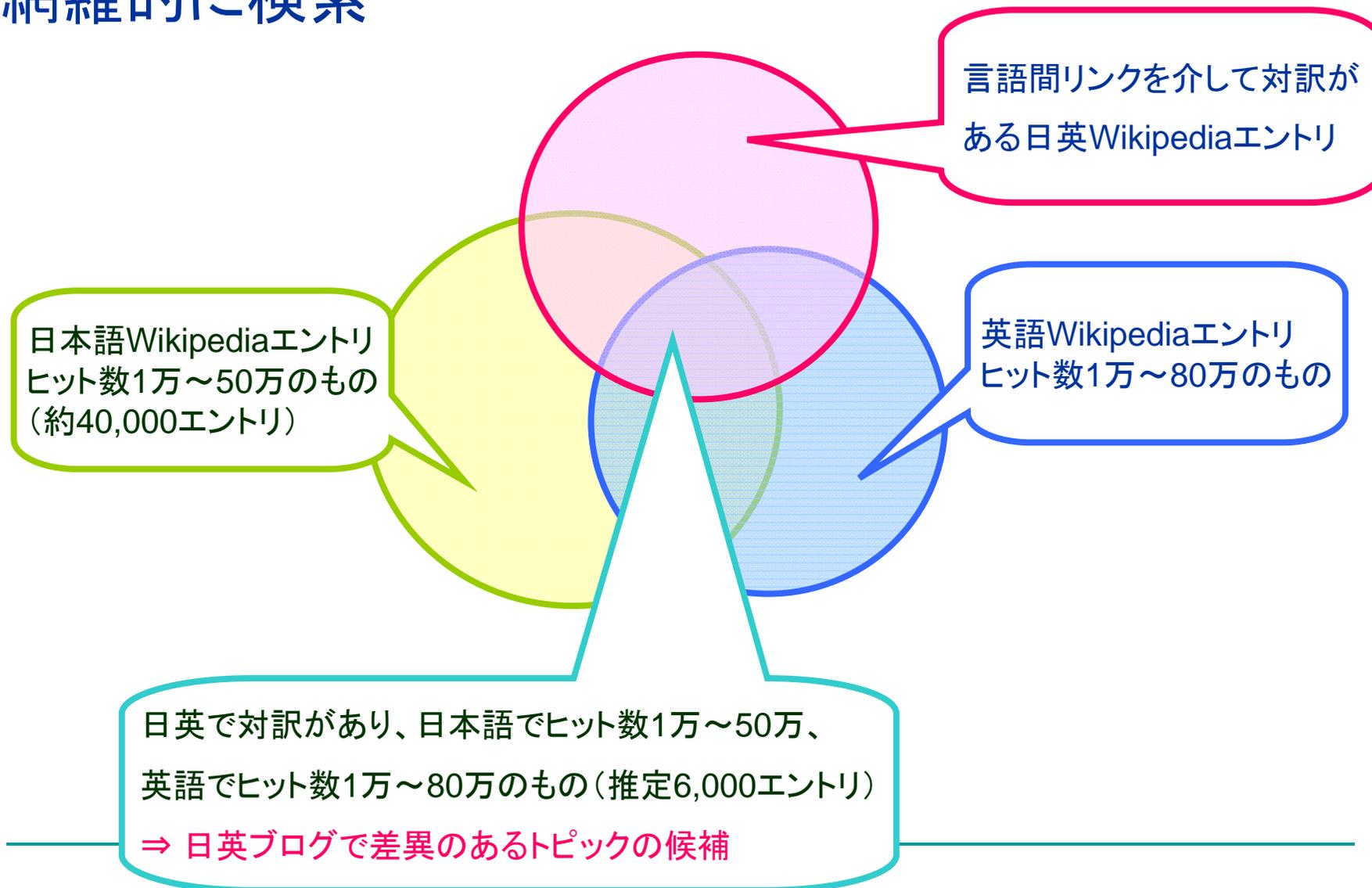
試聴・単曲購入はこちら...

1 | 2 | 3 | 4 | 5 | 最初 次ページ >>

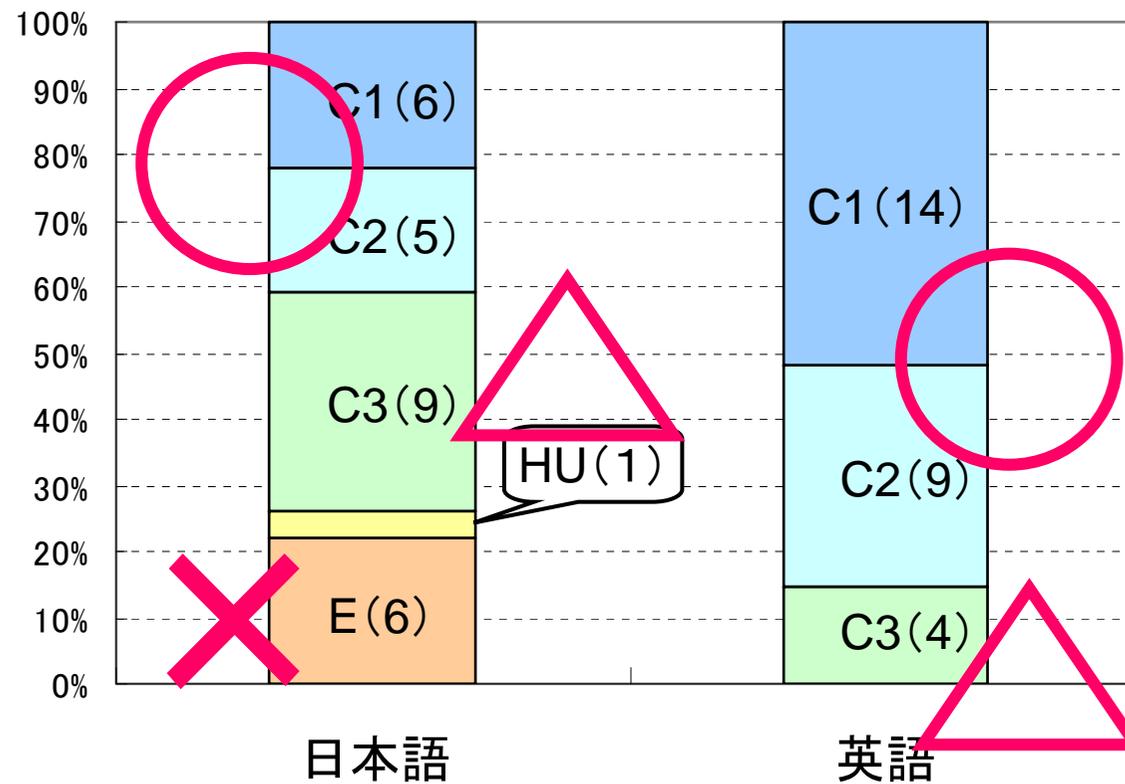
2008-09-23

ピアニストのブログ
ブロガーのコンサートやアルバムなどの
音楽情報を記載。
またピアノの弾き方などが
書かれている。

日本語と英語で差のあるトピック, 差の無いトピックを網羅的に検索



Wikipediaエントリに対応するブログサイト有無の 日英比較



Wikipediaエントリに対応するブログサイト有無の推定

- 日本語E/英語C1のブログサイト例
 - パンデミック
 - ある感染症や伝染病が世界的に流行することを表す用語



鳥インフルエンザや
Pandemic(伝染病)に
対する予防・対策・準備
の重要性を訴えている
ブログサイト

	日本語	韓国語	中国語	英語
ニュース				

多言語のトピック体系Wikipediaを

- ・ Wikipediaエン트리とブログサイトの対応付け [ICWSM2008, DEWS2008]
- ・ 日本語ブログ空間におけるトピック分布推定 [NL研2008年9月]

トピックの多様な特性の分析

スパムブログにおける
トピックのバースト特性の分析
[宇津呂・中川]

トピックの情報価値の質・持続性の
推定およびトピックの
視覚的二次元配置 [宇津呂]

トピックに関連する観点・
意見分布の抽出
[藤井・宇津呂]

トピックの情報価値の質・
持続性の差異度
[宇津呂・中川・福原・清田]

- ・ 日英ブログ間の対照分析支援
[ICWSM2008, ISUC2008,
DEWS2008, NLP2008]

意見分布の
抽出 [藤井]

世の中への出し方

- フリーのWebアプリケーションとして公開
- 国会議員に海外調査を頼まれる国会図書館員
 - 四川地震被災者への支援に関する意見
 - 日本支援の反省、中国人の反応
- 海外事情を知る必要のある企業
 - マーケティング
 - 国内と国外でのニーズの違いを調べる
- 海外事情に興味のある一般の人
 - 旅行計画
 - 現地でお勧めの場所などを調べる