九州大学学術情報リポジトリ
Kyushu University Institutional Repository

2020年の学術情報流通と大学図書館を展望する ： 情報システムの観点から

池田, 大輔
九州大学大学院システム情報科学研究院 ： 准教授

https://hdl.handle.net/2324/25449

出版情報：2012-11-18
バージョン：
権利関係：

九州大学
KYUSHU UNIVERSITY

2020年の学術情報流通と大学図書館を展望する
～情報システムの観点から～

国立大学法人 九州大学
大学院システム情報科学研究院
情報学部門 知能科学講座
池田 大輔

2012.11.18 日本図書館情報学会

目次

自己紹介、研究室紹介

情報システムの特徴

- 分割とモジュール化
- ログ

これまでの取り組み

- SarabiWeko: 分割とモジュール化
- 機関リポジトリのアクセスログ解析
  - どんな検索語で論文に到達しているか?

将来的な学術情報流通と大学図書館

自己紹介

平成　6年3月　九州大学理学部物理学科卒業
平成　6年4月　九州大学大学院総合理工学研究科
情報システム学専攻修士課程入学
平成　8年4月　九州大学大学院システム情報科学研究科
情報理学専攻博士課程進学
平成　9年4月　九州大学大型計算機センター
平成16年５月　九州大学附属図書館
平成18年8月　第6回MIT-EFJ
ビジネスプランコンテンスト最優秀賞
平成18年8月　九州大学大学院システム情報科学研究院
平成18年10月～平成20年3月 ベンチャー社外取締役

研究室紹介：ビジョン

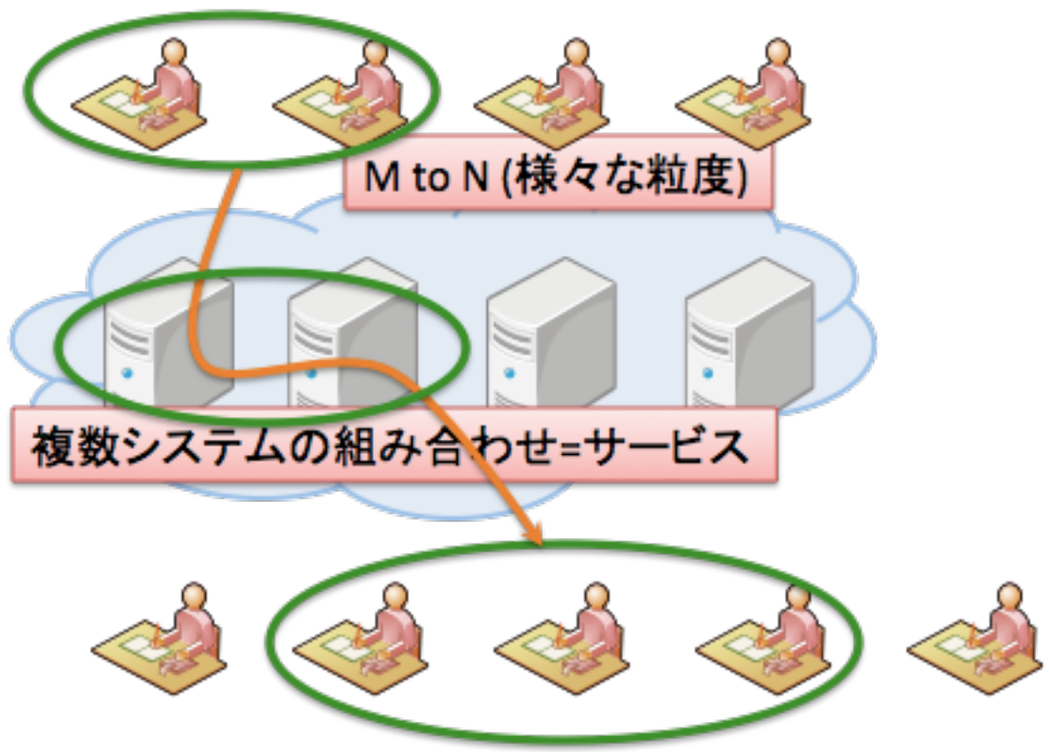
データの増加、でもシステム・機能が中心

- 例：MyLibrary
  - 附属図書館の個人向けWebサービス
  - 現在または過去に借りた本の確認、貸出の延長、本の予約配送依頼等が可能
  - 過去に借りた本のデータは重要
    - 一種のブックマークとして利用できる
  - しかし、一覧として見れるだけで、データとして活用できない
    - 過去に借りた本の検索をしたい、本にコメントをつけたい、本に星をつけたい、など利用者によってデータの使い方は異なる。
- 例：電子マネーの履歴
  - 様々な電子マネーが存在：Edy, WAON, nanaco等
  - それぞれのサービスごとに履歴が提供される
    - 利用者は、自分の履歴をまとめてみたり、家族と共有したりしたい

研究室紹介：ミッション

データの高度な利用やシェアを簡単に!!

- システムを改修することなく、利用者が複数のサービスのデータを統合して利用できると嬉しい
  - 複数のICカードサービス (nanako, WAON, nimoca等) から自動的に家計簿を作り、費目ごとのグラフにして欲しい。
  - 附属図書館と福岡市図書館の貸出履歴をまとめて、自分にピッタリの本を推薦して欲しい。



- あらかじめ提供された機能を使うのではなく、利用者が自由に使い方を決める。

民主主義的なデータの利活用を可能にする基盤システムの研究

研究内容：概略

データの利用：データを高度に利用するための研究

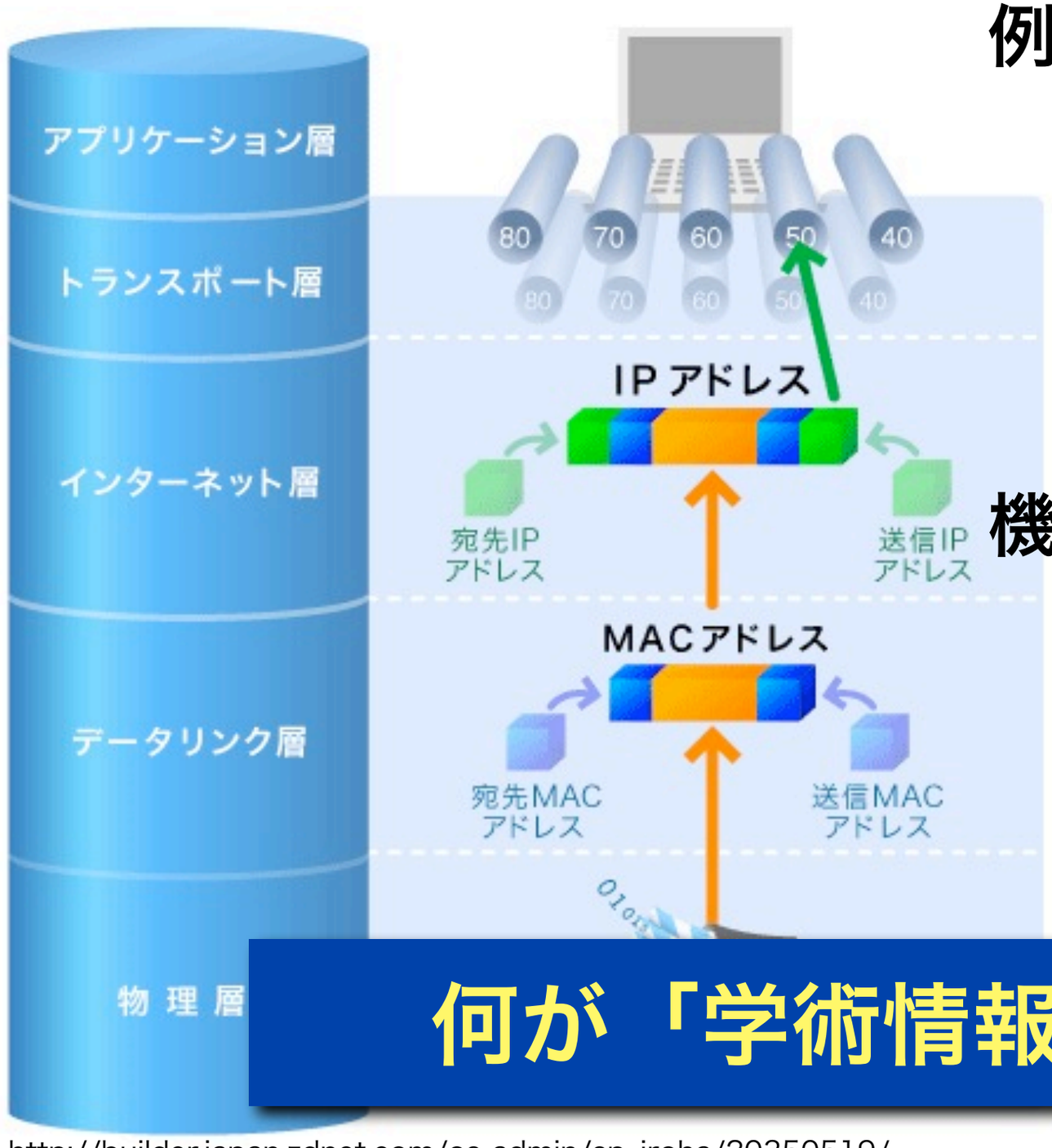
- 様々なデータを対象としたマイニング
  - テキストデータ、ゲノム配列、関係データ、履歴データ等
  - 時系列データ、マルチメディアデータ(画像、音楽等)
- マイニングのためのアルゴリズムやデータ構造の開発
- データ利用を妨げるもの(スパムやボットネット等)の検出

データ基盤：データを保存、操作、シェアするシステム

- データベースの基礎的な研究(データモデル、索引構造、アルゴリズム等)
- 認証や認可、データ連携、サービス連携(マッシュアップ)
- 時系列データやマルチメディアデータのデータベース構築
  - 検索アルゴリズムや索引構造

情報システムの特徴

特徴1：機能の分割とモジュール化



例：

- インターネットのプロトコル階層(左図)
- OS機能

機関リポジトリ

- Webサーバ、DB等をモジュールとして内部的に利用している

何が「学術情報流通」か?

http://builder.japan.zdnet.com/os-admin/sp_iroha/20350519/

ログの蓄積

(物理媒体の)貸出履歴

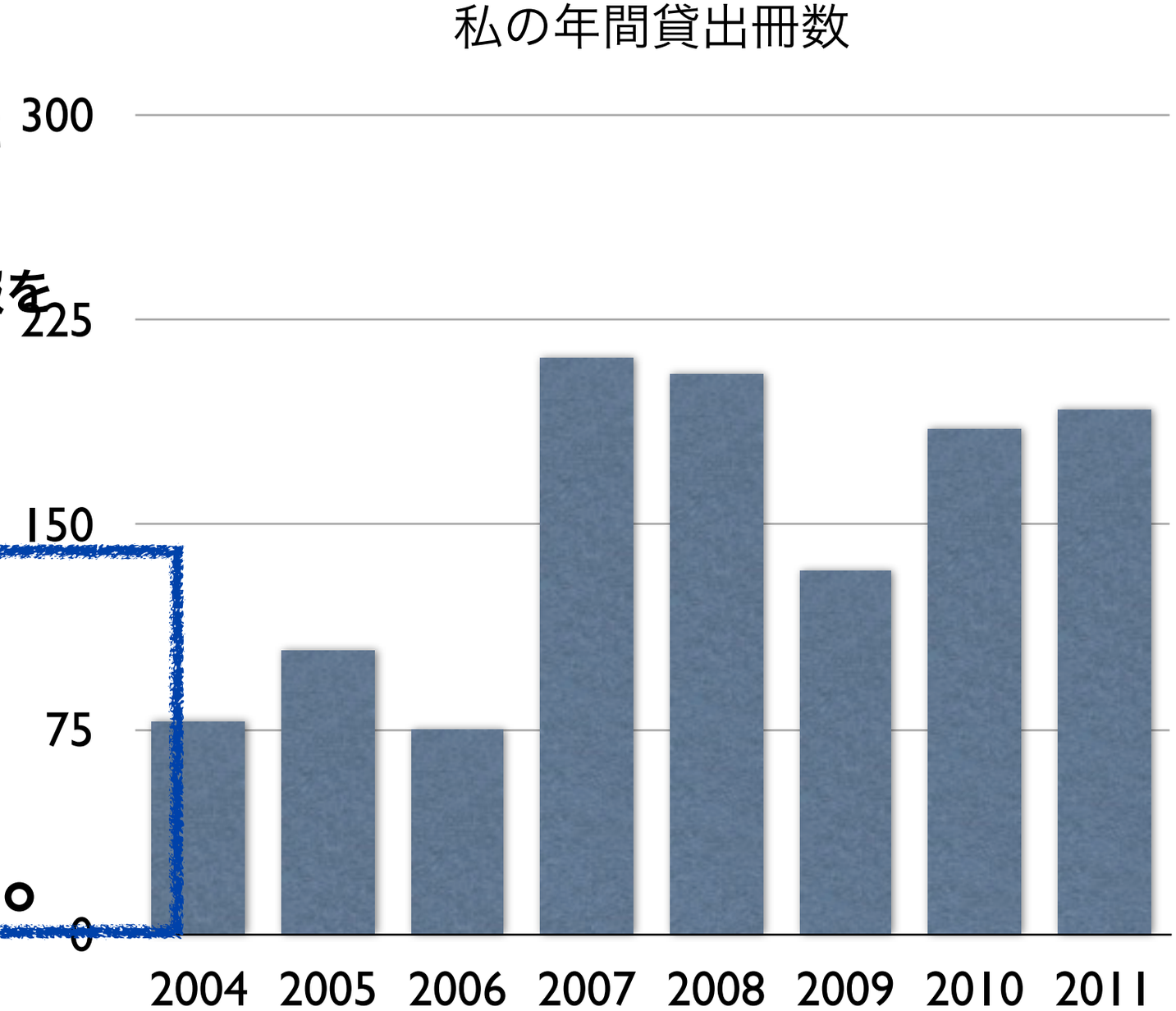
- 歴史は古いが十分に活用されてきたとは言い難い。

(電子媒体の)閲覧履歴

- 電子ジャーナルのログは出版社側のもの。

大量のログには価値がある!!

- 気になった本をブックマークするために借りる。
- 自分で本を管理せず、
  図書館に管理させるために
  備品扱いとする。
- どのくらい使ったかの情報を
  残すために借りる。



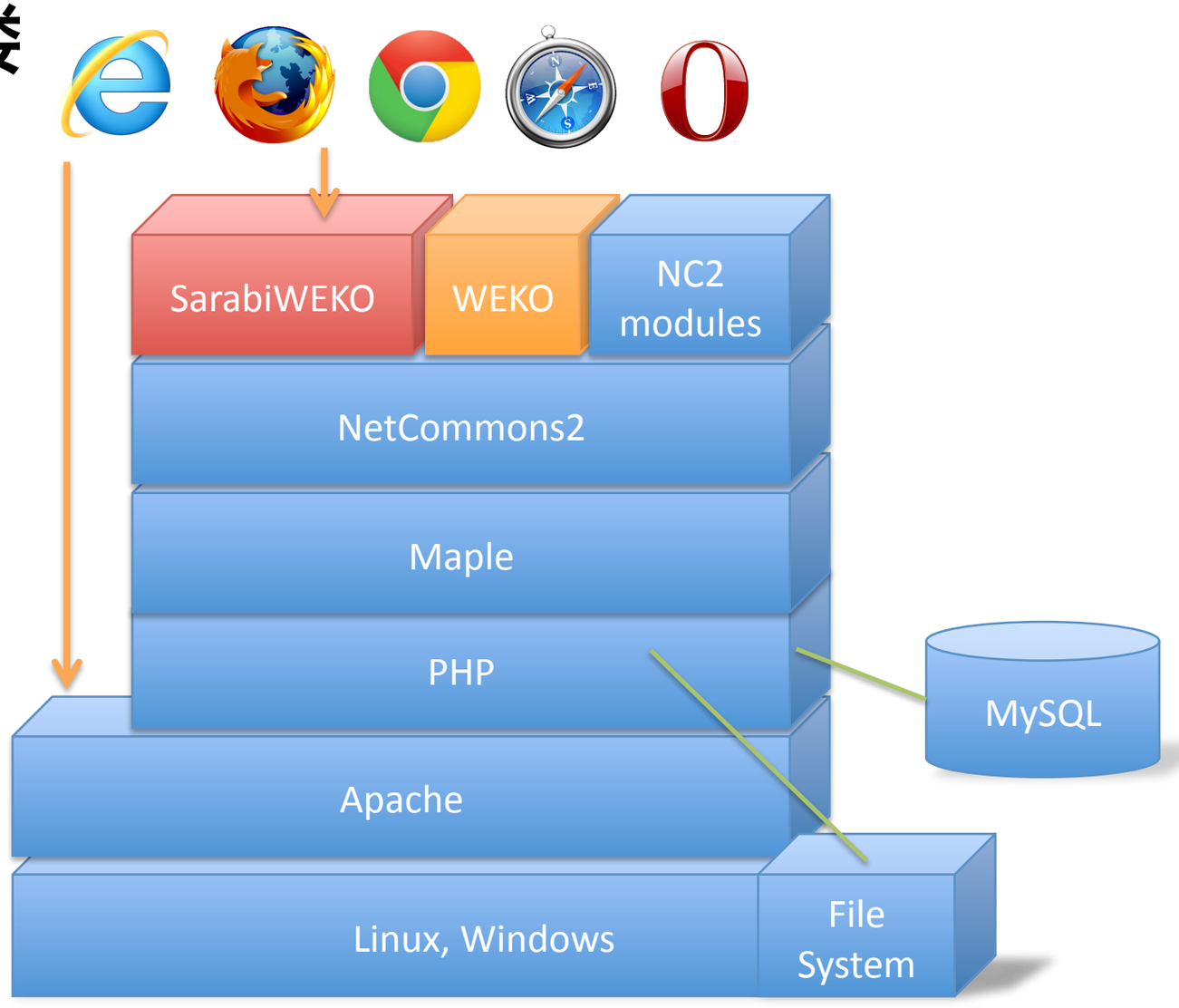
図書館は
本ではなく、
データを管理すべき。

これまでの取り組み

1. SarabiWekoの開発

Weko(リポジトリ)上の蜃気楼

- 実体であるリポジトリのコンテンツを様々な形で見せる。
- Wekoと協調して動作。
  - NC2上で動作。



卒業研究 | 編集

加齢に伴う色覚変化のシミュレーション
宮口[redacted]
鈴鹿工業高等専門学校 , 2010
発表スライド , 予稿集 , 2011-02-02

pdf

SVMによる文献分類に関する研究
松本[redacted]
鈴鹿工業高等専門学校 , 2006
発表スライド , 予稿集 , 2007-02-04

pdf

PDF 論文へのデジタルペンによるコメント編集・共有に関する研究
平田[redacted]
鈴鹿工業高等専門学校 , 2008
発表スライド , 予稿集 , 2009-02-03

pdf

ゲノムマイクロアレイデータベースに関する研究
田中[redacted]
鈴鹿工業高等専門学校 , 2007
発表スライド , 予稿集 , 2007-02-05

pdf

卒業研究発表会スライド

卒業研究発表会予稿集

発表年月日

WEKOが持つコンテンツ

WEKOが持つメタデータ

SarabiWEKOが持つメタデータ

卒業研究 編集

インデックス: 青山研究室

Permalink : http://id.nii.ac.jp/1035/00000034/

SVMによる文献分類に関する研究

File	Name	Size	License
	Matsumoto2006		
PDF	Matsumoto2006.pdf	182.28KB	Creative Commons 示

OAI-PMH

アイテムタイプ: 学位論文 / Thesis or Dissertation

言語: 日本語

著者: 松本，昇吾

寄与者: 青山，俊弘

抄録: Recently, researchers can get bibliography information from the Internet easily than before. However, the huge information on the Internet makes it difficult for researchers to get the infor- mation in which they are really interested. Researchers need efficient methods for managing and getting the information, i.e. interest important articles.
Support Vector Machine(SVM) is one of the techniques for data classification. Previous stud- ies have shown that SVM provides high classificational performance compared to other learning algorithms.
I applied SVM to article (abstract) categorization and evaluated the performance of classifica- tion. An abstract is converted into multi-dimensional vector which consists of term frequencies of the abstract. As a result, my system can achieve high performance even with small training data with the following conditions:
in low dimensions of input space to extract input space to maximize a difference of term frequency among categories with the Hard Margin method without biased number of training data in inner/outer category
Using the results, I developed a software which automatically categorizes and manages articles.
The application is merging into Concierge — electronic research notebook & digital resource management tool.

学位授与大学: 鈴鹿工業高等専門学校

学位授与年度: 2006

学位授与年月日: 2006-03-31

発表スライド: 発表スライド

予稿集: 予稿集

発表日: 2007-02-04

戻る メタ

この論文へのコメント

※登録されているコメントはありません

コメントする

名前*

Aoyama

e-mail

コメント*

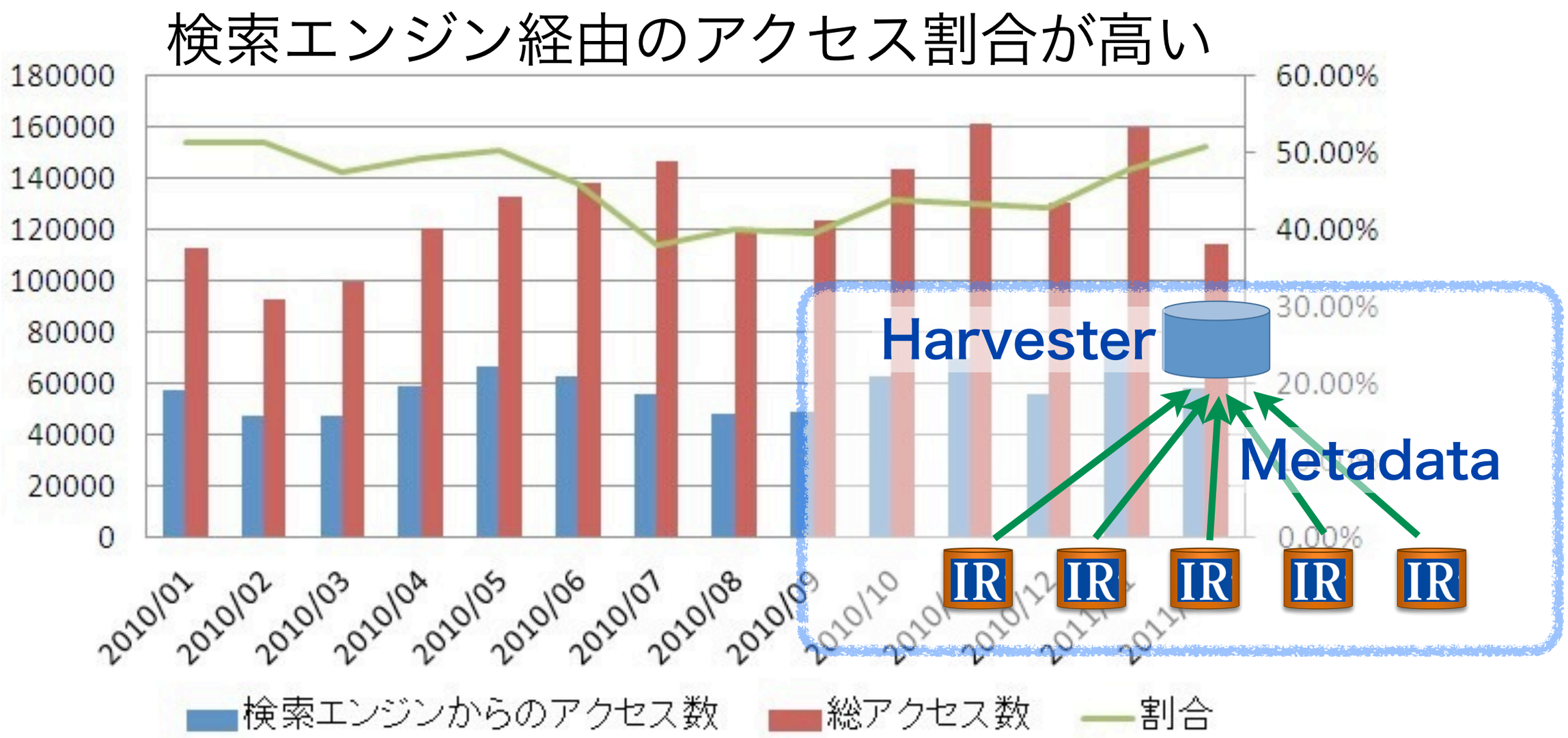
送信

2. アクセスログ解析

(標準的には)以下のような情報が記録される

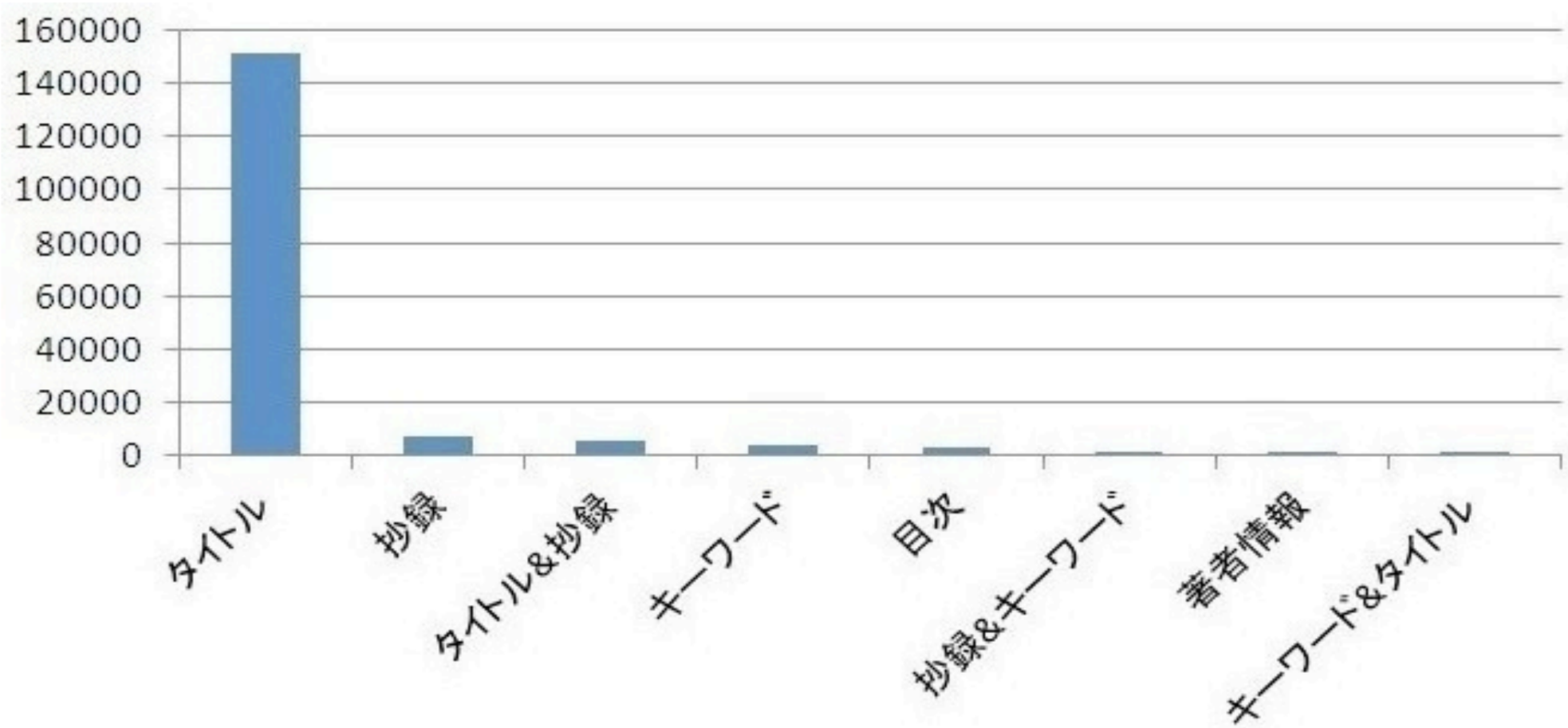
- クライアントのIPアドレス
- アクセス日時
- アクセス要求したファイル
- ブラウザの種類やバージョン
- リファラ(referer)：リンクをクリックして、移動した場合、移動する元のサイトのURLが含まれている。
  - 検索エンジン経由でくると、検索語が分かる!!
- 133.5.24.117 - - [20/Jun/2012:13:40:44 +0900] "GET /sns/ HTTP/1.1" 200 29113 "http://www.google.com/url?sa=t&rct=j&q=magnet%20sns&source=web&cd=1&v...A" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_8) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2"
  - 見やすさのために、適当に省略しています。

検索エンジン経由のアクセス



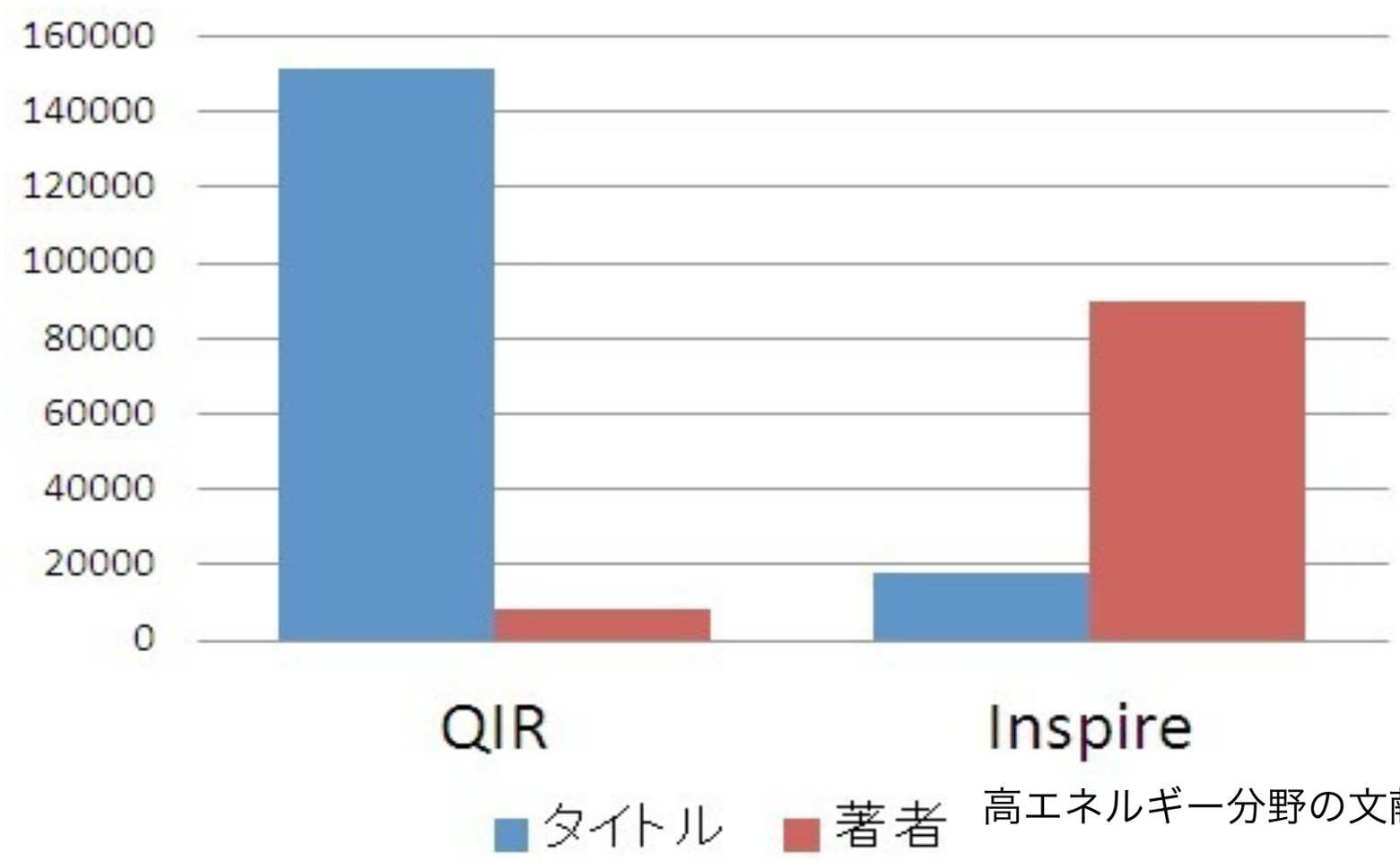
ハーベスタによる検索モデルは機能せず

検索クエリのメタデータへの分類



タイトル検索が圧倒的に多い

研究者向けのデータベースとの比較

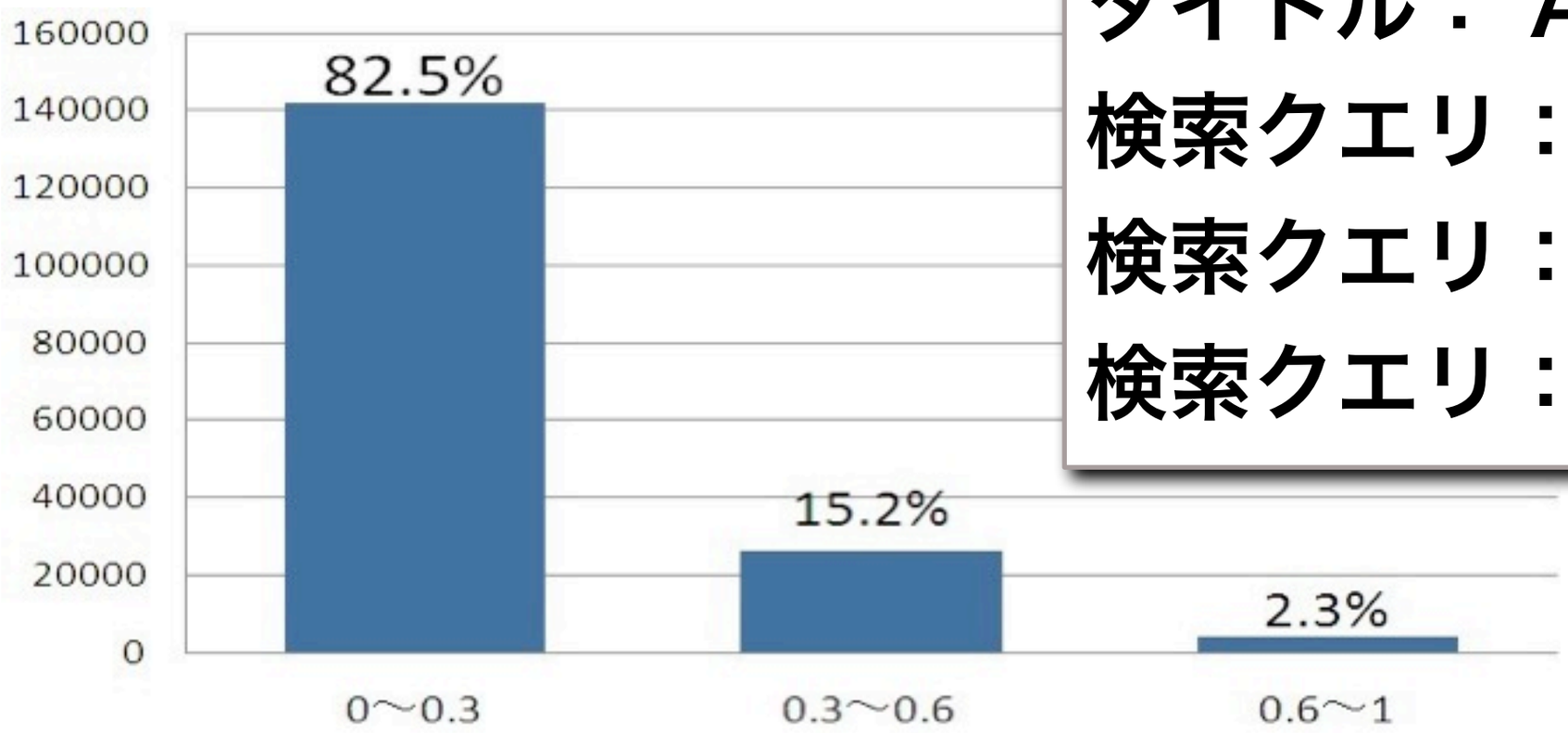


高エネルギー分野の文献DB

検索手法が逆 → 利用者層も逆？

タイトルとの一致率

$$k = \frac{\text{タイトルに一致する検索語の長さの和}}{\text{完全なタイトルの長さ}}$$



タイトル： ABCDEF

検索クエリ：A CD	0.5
検索クエリ：CEF	×
検索クエリ： AB G	0.33

利用者の専門性が低い可能性が高い

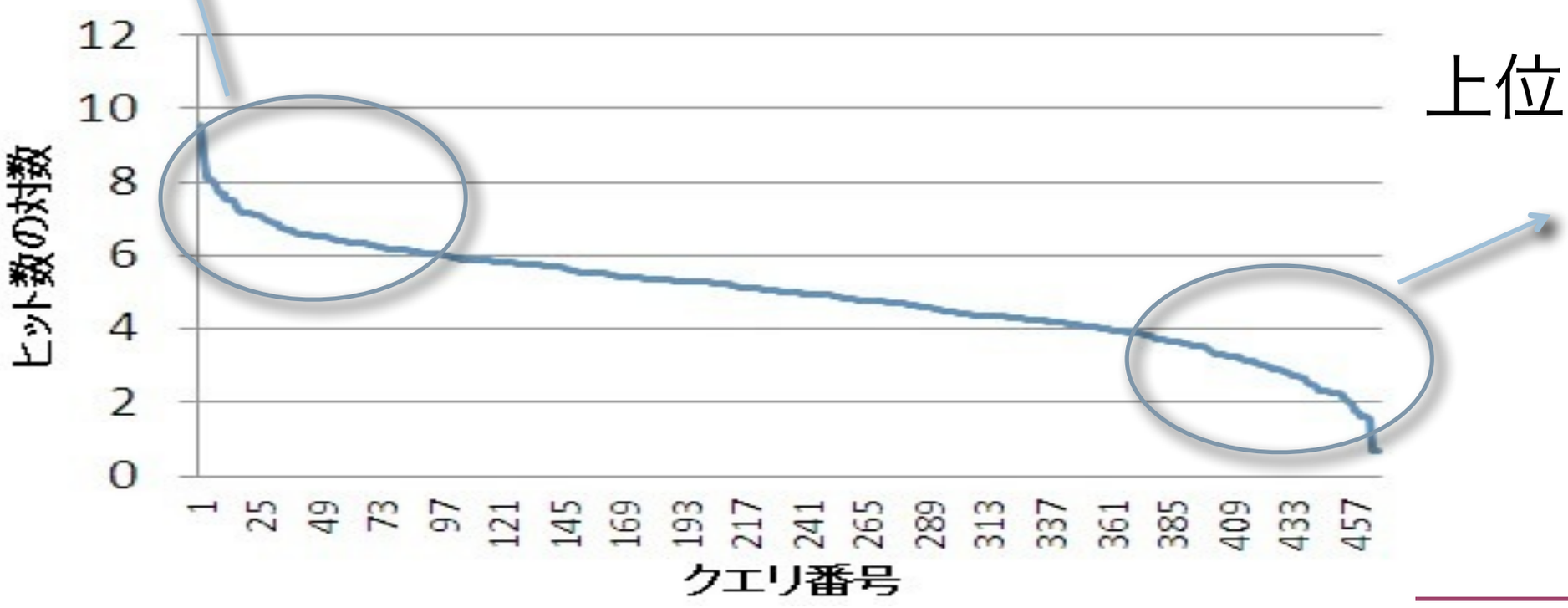
検索クエリの検索エンジンでのヒット数

クエリ数：Yahoo!検索経由のクエリ４５０件

例

クエリ	ヒット数
ロールシャッハ FC:CF+C	43
日本の原子力発電 政策	7690000

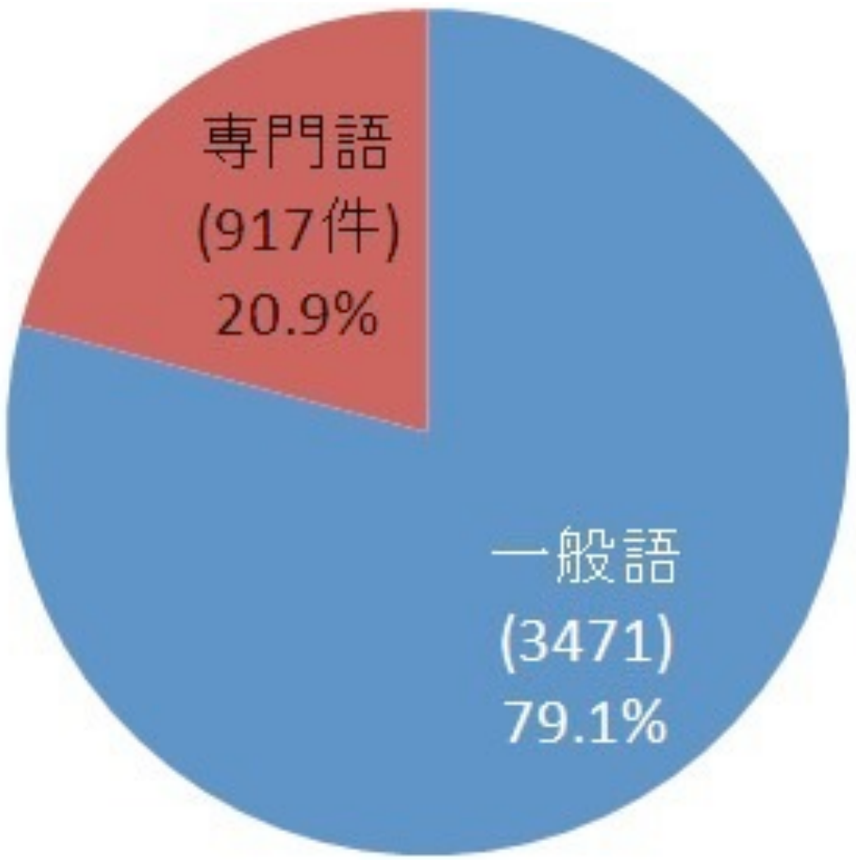
上位100件



上位100件

専門語と一般語の利用頻度

QIRの期間：2008年7月～2011年2月



クエリ	ヒット数
ロールシャッハ FC:CF＋C	43
日本の原子力発電 政策	7690000

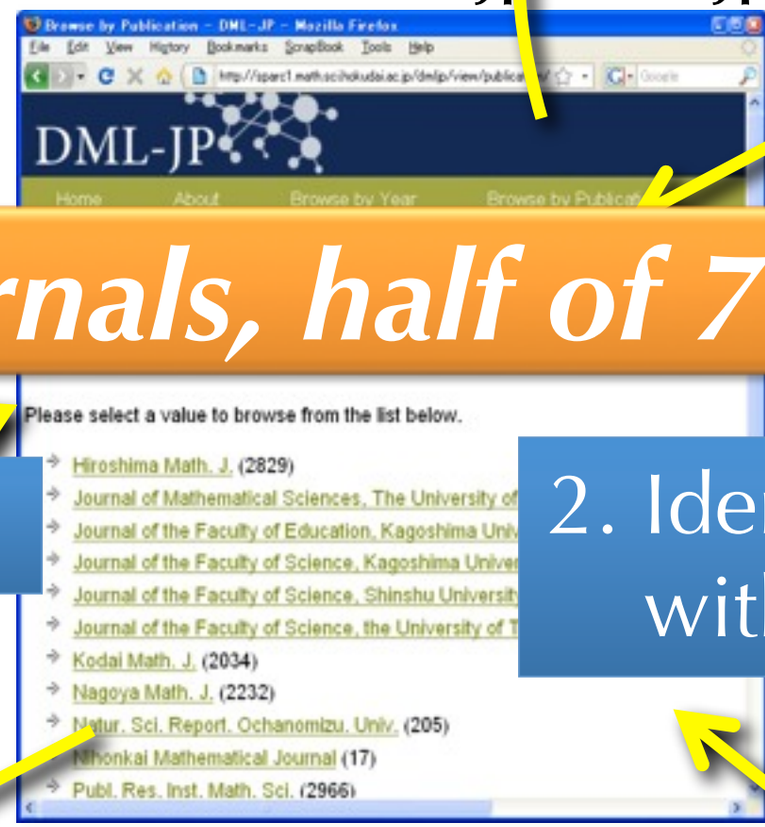
研究者以外の新規利用者の存在を強く示唆

将来的な学術情報流通と大学図書館

役割分担の例：数学分野のポータルサイト

http://sparc1.math.sci.hokudai.ac.jp/dmljp/

DML-JP



3. Merging metadata

68 of 200 journals, half of 70,000 articles

1. Metadata Harvesting

2. Identifying Papers on IRs with MR number

4. SWORD and OAI-ORE

40 Institutional Repositories

Math. Reviews/MathSciNet
2,000,000+ reviews

将来的なIRの役割

(DB的な)コンテンツ+付加データの流通

- 現在は情報検索的な再利用
  - オーバーレイ：コンテンツのみがread-onlyで再利用
- IRはコンテンツのアクセスのみを保証するシステムへ
  - 付加データや再利用は、IRを基盤とした別システムで
  - OAI-PMHやSWORDがより重要に
  - (利用者が管理できる)識別子がより重要に
- 個人ベースでのまとめサイトが可能に
  - 属性つきデータとしてまとめることが可能

様々な読者層へのコンテンツ提供

- (昔)研究者による、研究者のための学術情報流通
- (今)研究者と出版社と図書館による、研究者のための学術情報流通

図書館が扱うべきデータ

ログ

- 利用者のニーズに加え、自機関のコンテンツがどう利用されているのかを知る
  - 例： 原子力の利用に伴なう環境の放射能汚染について
    - 小樽商科大学 Barrel (http://hdl.handle.net/10252/3071)

付加データ(CGM)

- 付加データ自体を格納するプラットフォームを提供し、 自機関のコンテンツがどう利用されているのかを知る

メタデータ

- (情報)検索のためのメタデータではなく、再利用のためのDB的な属性としての重要性が高くなる

まとめ

分割から見た将来的な学術情報流通

- まだ分割しきれていない
  - まだIRとは何かがハッキリしていない、とも言える
- メタデータの役割の変更
  - 分割したモジュール間をつなぐために重要
    - 識別子
  - 探すものからデータの再利用へ
  - 情報検索からデータベースへ

大学図書館の役割

- 基本メタデータの管理
- 付加データを管理し、分析