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

リポジトリの有効性分析

馬場, 謙介
九州大学附属図書館研究開発室 : 准教授

https://hdl.handle.net/2324/18785

出版情報：DRF/ShaRe地域ワークショップ(九州地区), 2010-12-22
バージョン：
権利関係：

リポジトリの有効性分析

馬場 謙介
九州大学附属図書館 研究開発室
baba@lib.kyushu-u.ac.jp

本研究の目的

リポジトリへの文献登録数を増やす

まず，九州大学の機関リポジトリ（QIR）の登録数を増やす

大きな目標：学術情報へのオープンアクセス

- 電子ジャーナル購読料問題の解決
- そもそも，学術情報の公開は大学の義務では？

これが容易になることもリポジトリのメリットと考えるべき

リポジトリは利用されているか？

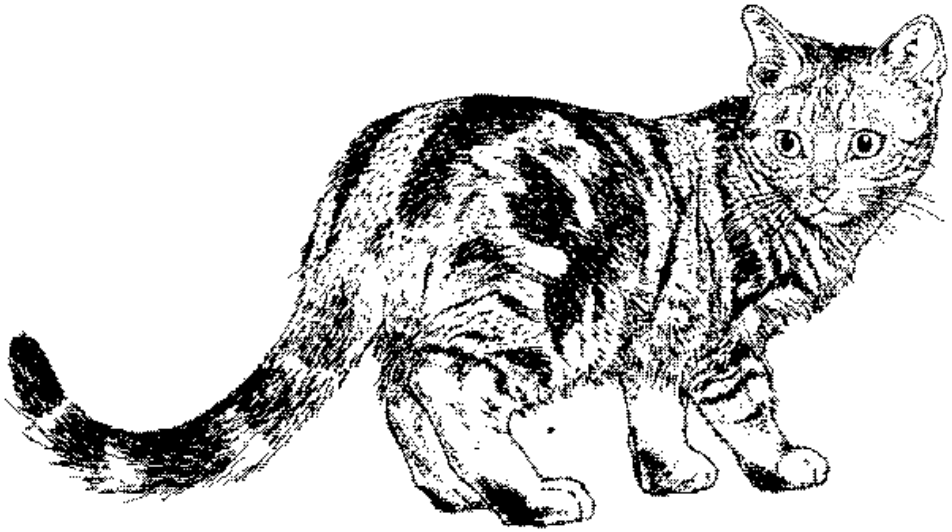
九州大学の場合

- QIRに登録されているフルテキスト：約16,000件
- 九州大学の研究者データベース（DHJS）に登録されている論文メタデータ：約70,000件（うち重複が1割程度 ）
- つまり，執筆されたがリポジトリに登録されていない論文が，少なくとも40,000件以上！

研究者の意識が変わるだけでも，登録数の大きな増加が期待できる

解決の切り口

こんなに好きなのに，どうして分かってくれない？



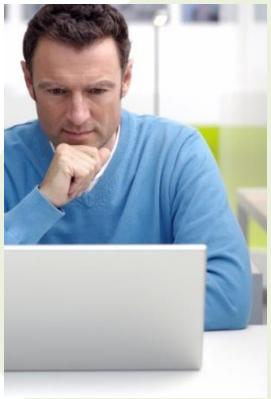
もしかして，こっちが分かっていないのでは？

これまでの工夫は管理者の視点

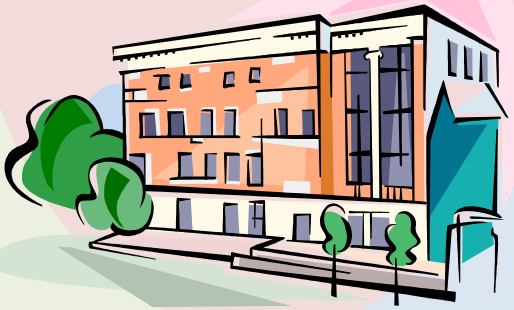
こんなに便利なのに，
どうして登録してくれ
ないの？

管理者

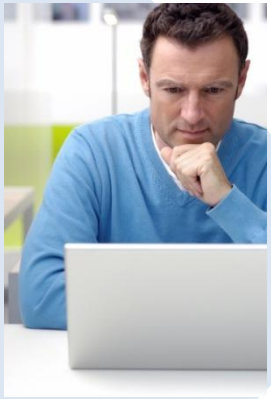
著者や閲覧者の視点から，
もっと質の違う工夫ができ
るはず！



著者



リポジトリ



閲覧者

ひとつの問題点

著者（研究者）にとっての登録のメリットが見えない

実感として

- 確かに面倒だ（色々と工夫はあるが．．．）
  - DHJSとの連携（ServCom2010, http://hdl.handle.net/2324/18775）
  - QIRからのきっかけ（H22年度CSI事業のアイデア）
- メリットがあれば多少面倒でもやる
  - DHJSの高い登録率
    デメリットをさけるため
  - 閲覧者からの反応があるなら
    質問のメールや論文での引用等

解決方法

QIRのアクセスログを解析して，有用な情報を著者にフィードバック

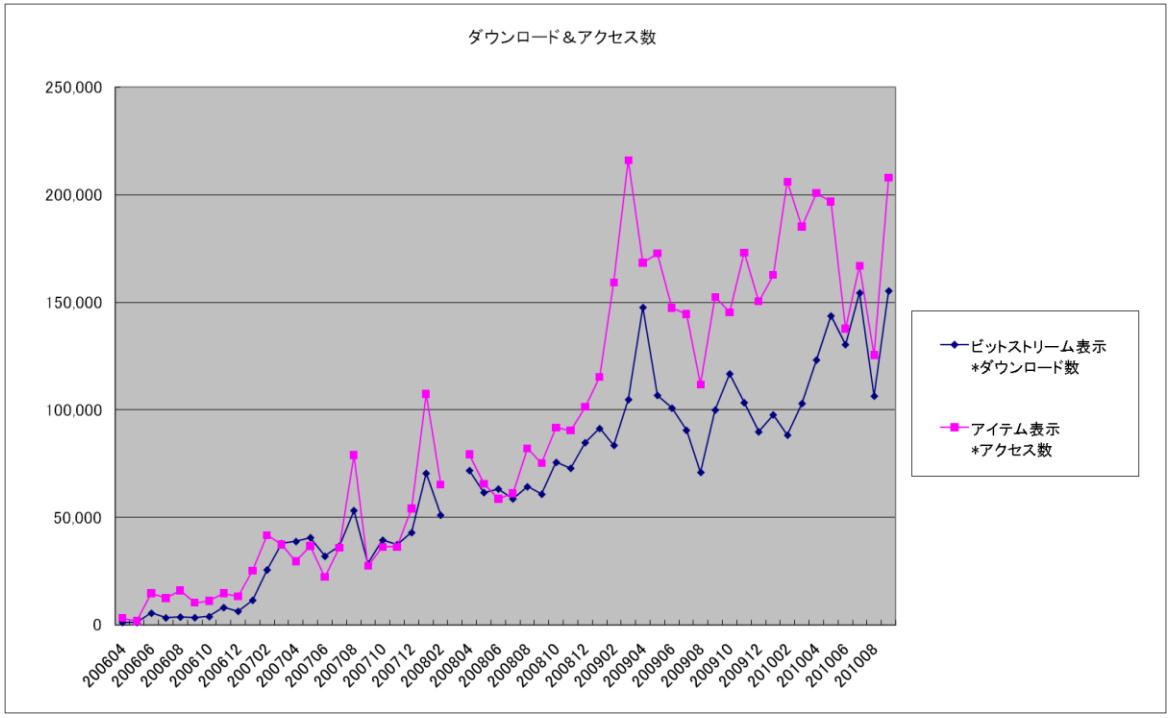
- 基本的な解析（これだけでも有効）
  - どれだけ見ている？
  - 誰が見ている？
  - 何を見ている？
- 詳しい解析（もっと有効な情報は？）
  - まじめに見ている？

    →前後の閲覧内容との関係

どれだけ見ている？

月のアクセス数が20万件を超えることも

- 「誰も見ない」なんてことはない
- botによる機械的なアクセスも二次利用されるはず
- もはや，1人や2人の担当者が片手間でやる規模でない
- 組織として定常的な経費を用意するべき



誰が見ている？

- botによるアクセス
  - ログの字面で判別できるだけでも3割強
  - 一般的に6割以上（標準化が千葉大等のCSI事業で）
- 地域別の分類
  - 英語の記事は海外からもよくアクセスされている（佐藤さんら＠筑波大）
  - 閲覧者の趣向が少し分かる：アクセス数・時間と地域の関係から，研究者や留学希望者の分類ができるのでは？（森先生＠評価情報室）
    - アジア：いくつかのサイトを短い時間，研究者を探している？
    - 欧米：少ないサイトを長い時間，論文を探している？

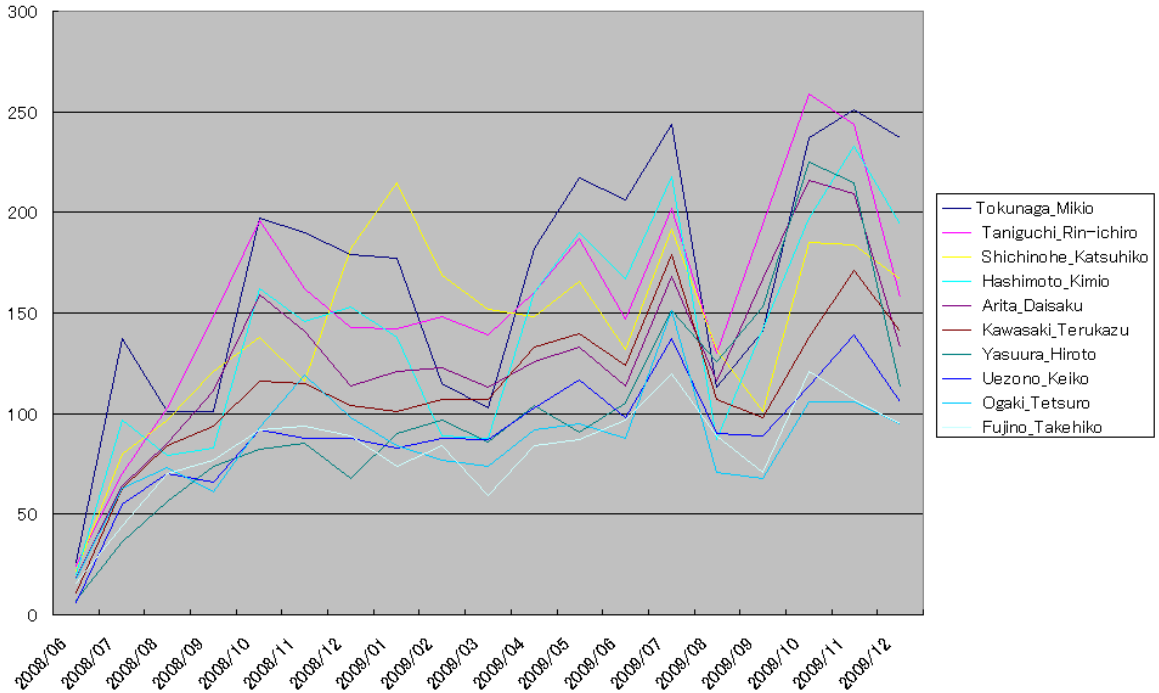
何を見ている？

アクセス数ランキング

- 記事について

  Dspace等の基本的な機能で実現可能

- 著者について
  - 公開を嫌がる研究者もいる（ログイン後の表示等で解決）
  - 理屈でなく奮い立つ研究者もいる



まじめに見ている？

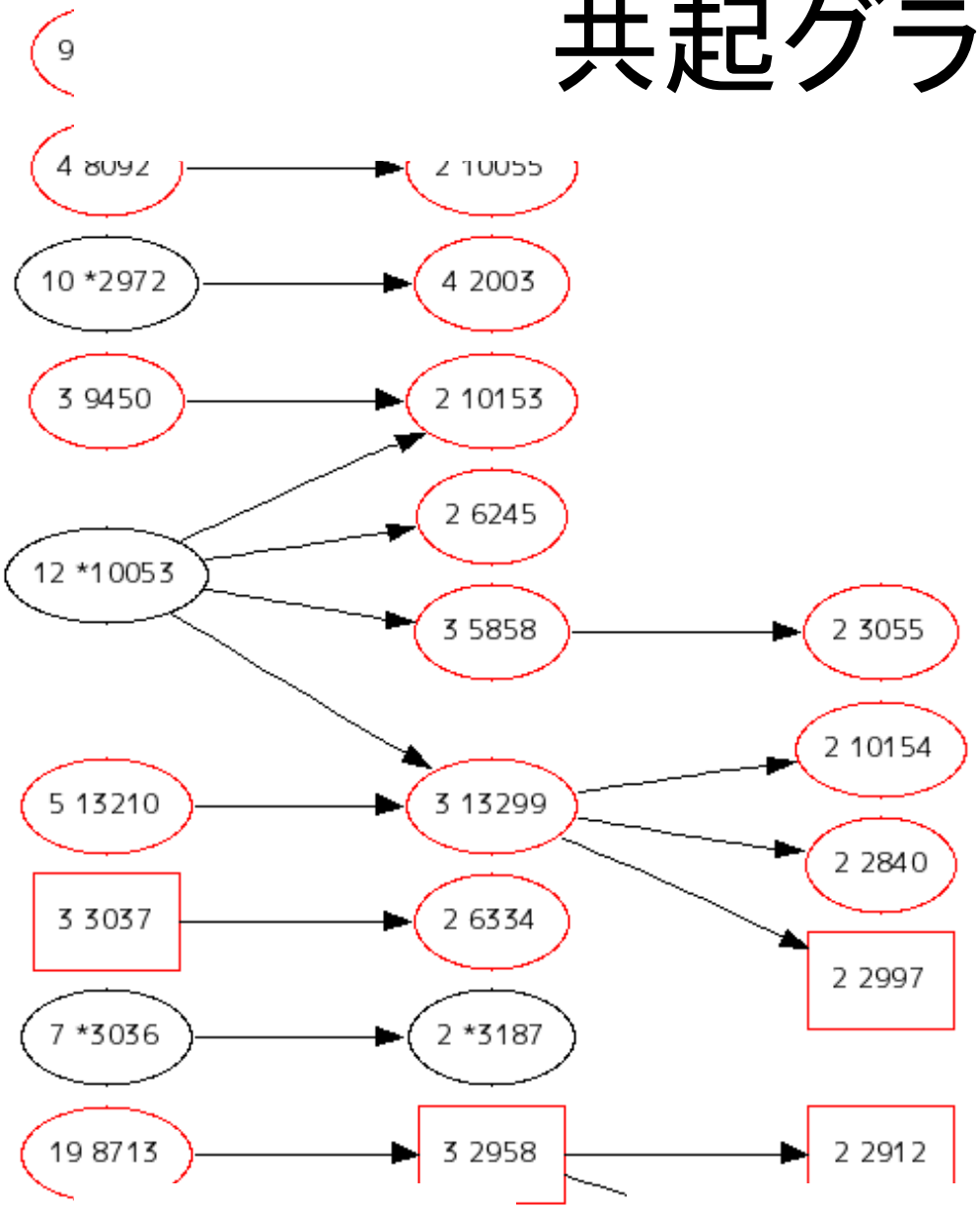
- アクセス数と被引用数との関係
  - いくつかの電子ジャーナルや主題（ある分野についての）リポジトリについて相関が報告されている（Watson[09]等）
  - 国内の機関リポジトリについてはまだ否定的（佐藤さんら＠筑波大）
- 共起解析new!
  - 同じ人がどの記事とどの記事を読んでいるか
  - botを除いた残りでも共起のあるアクセスが1割程度
  - 共起関係でできるグラフを分類（JCAICT2011で発表予定）
  - 具体的に有益な情報はこれからだが，何かありそうだ

共起グラフの作り方

- botによるアクセスを削除
- 同一アドレスからの同日のアクセスを同一人物のアクセスとみなす
- アクセス数が1回や，異常に多い人物は削除
- 記事の初期集合を決定し，そのアクセスとの共起を調べる

ほとんど共起が無いと予想していたが．．．

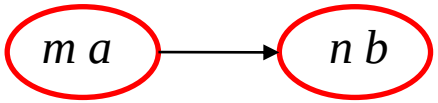
共起グラフの例



記事

リストセット（目次）

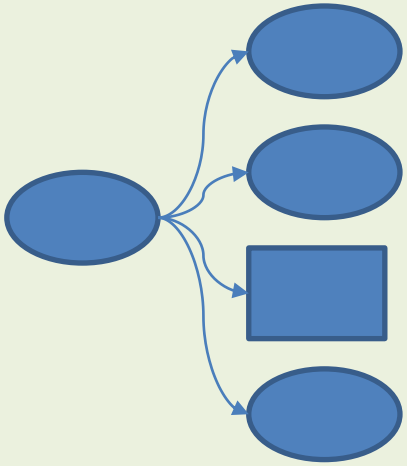
設定した著者による記事



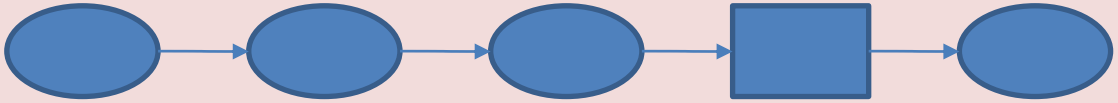
記事 a へアクセスした m 人のうち，
n 人が記事 b にアクセスしている．
* が付く記事は初期集合に含まれる

この記事を読んだ人は，この記事も読んでますよ

グラフの分類



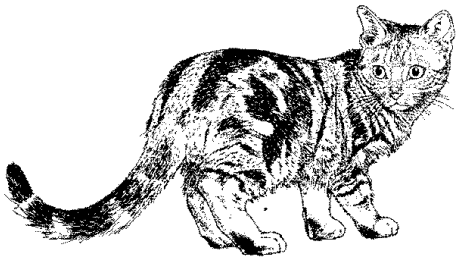
いくつかの記事がばらばらに
読まれている
→検索語による外部からの
アクセスが多いのでは？



ある程度決まった記事が読まれている
→リストセットを利用したアクセスが多いのでは？

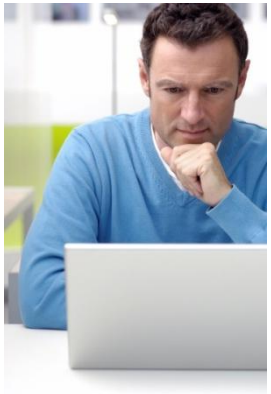
まとめ

- 目的：リポジトリへの登録数を増やしたい
- 問題点：著者にとっての登録のメリットが明確でない
- 解決方法：アクセスログの解析結果を著者にフィードバック
- 結論（事前の解析からの予測）：
  - 基本的な解析の結果でも有効
  - 詳しい解析でより有益な情報が得られそう
  - 現在，フィードバックの仕組みを開発中
  - 登録数の変化で効果を検証する予定



もうひとつの視点

「閲覧者」の気持ちも考える

- マーケティング

  アクセスログ解析結果を，記事の見せ方にも反映させる
- 質の保証
  - 独自の査読のしくみ
  - 信頼できるバーションとの差分の明確化
- リポジトリならではの付加価値
  - 著者による論文間の関係の入力
  - なるべく読まなくてすむ見せ方