

# 自己紹介：當山仁健

- 大学図書館で3年前まで利用者教育とシステムの担当
  - 利用者講習会でCiNiiの紹介、利用方法の指導
  - 利用者のプロフィールを考慮する連想検索OPACの構築
  - 機関リポジトリの構築
- 現在は琉球大学の情報工学系博士課程の社会人学生
- 翻訳させていただいた本
  - 「集合知プログラミング」オライリージャパン
  - 「Pythonクックブック」オライリージャパン

# CiNiiへの期待

メタデータに「集合知プログラミング」  
してみた雑感を通して

学校法人沖縄国際大学総務部会計課  
琉球大学理工学研究科総合知能工学専攻

當山仁健

# 今日お話する内容

[前半]

CiNiiのAPIを使ってメタデータを取得し、  
「集合知プログラミング」した雑感について

[後半]

CiNiiに期待する事

# CiNiiのWebAPI体験

集合知の宝庫、メタデータの

「つながり」を手軽に(それっぽく)

「見える化」してみた

※参考:「集合知プログラミング」Toby Segaran(著) オライリージャパン

# メタデータの「つながり」に着目する理由

- 検索漏れの少ないシステムを目指して
  - 新語、訳語の確定していないキーワード同士をレコメンドできるシステム
- レコメンデーション、適合度基準として利用
  - 類似の論文をまとめてレコメンドできるシステム
- 活発な研究分野、重要な論文の可視化
  - 論文間の関係进行分析する事で可視化
- 「偶然の出会い」の創出
  - ブール検索では難しいが、より創出したい

# 「つながり」がありそうな メタデータ項目の例

- 執筆者

→ 人間関係の可視化？

- キーワード

→ サジェスト機能等

- 論文

→ レコメンデーション等

# —執筆者間の「つながり」の例—

- 同じキーワードを使っている人たち
- 共著の回数が多い人たち

# —「つながり」の「見える化」の手順—

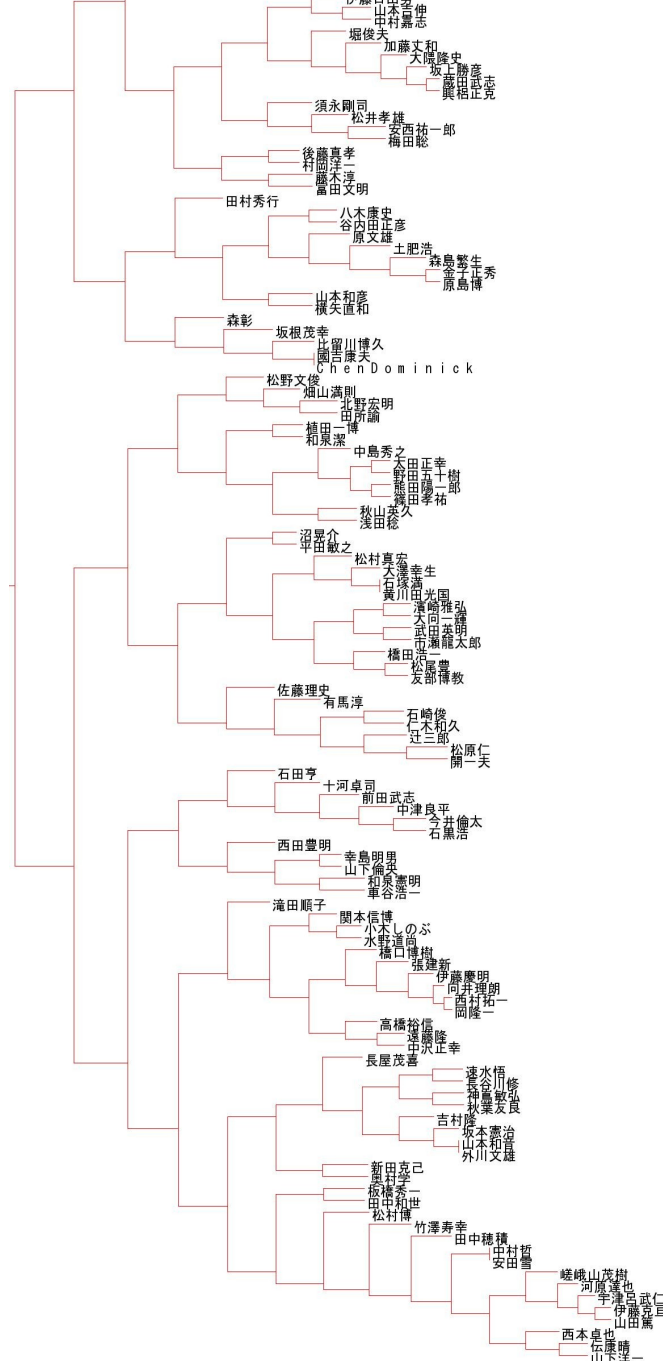
- APIを利用して共著者を芋蔓式に抽出
- APIを利用してそれぞれの方が書いた論文(最大20件)のキーワードを抽出
- 著者×キーワードの行列を作成
- 行列を利用して「つながり」を「見える化」



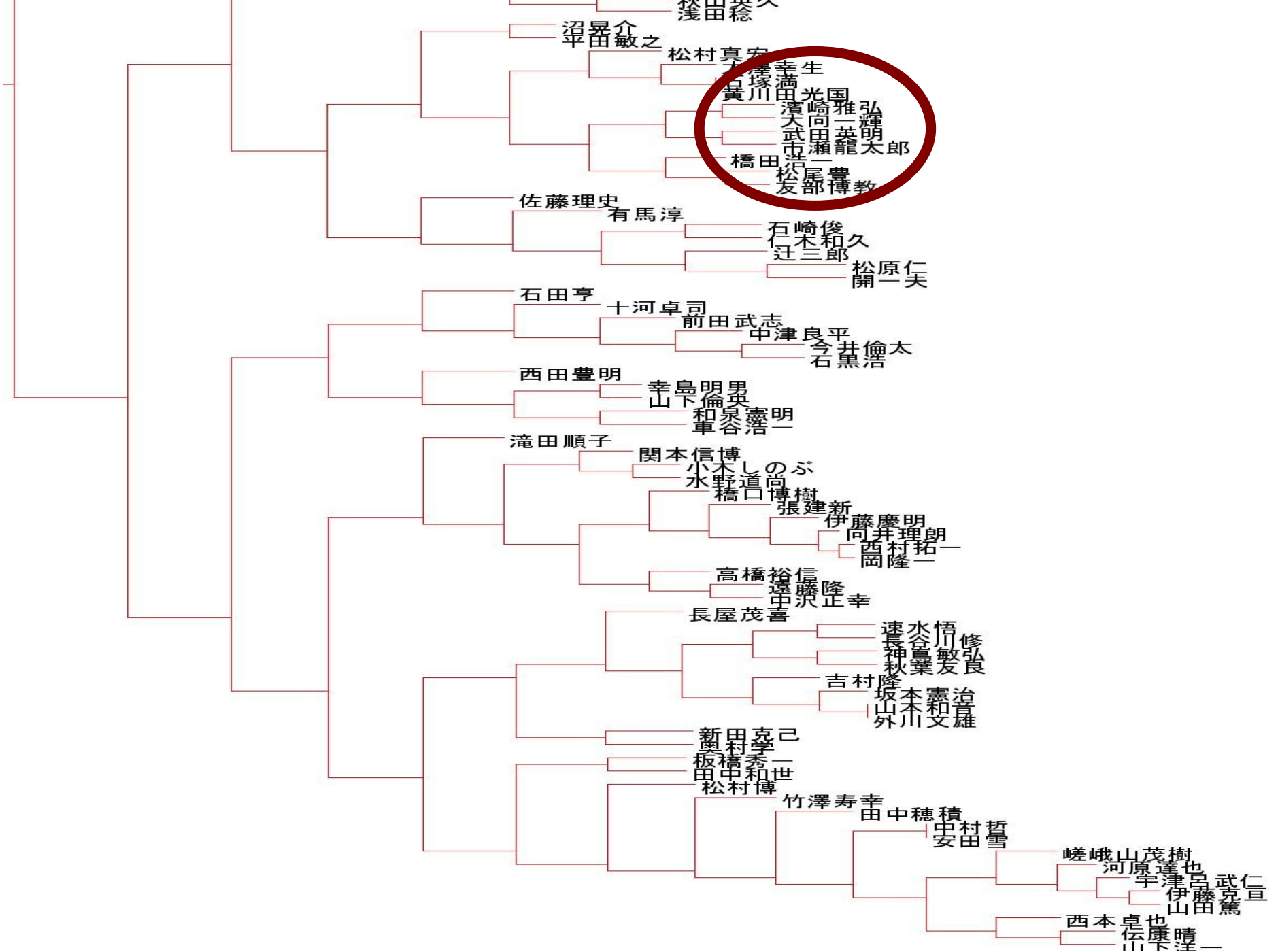
# 著者×キーワードの行列

author	Object_Stacking	FWA	マルチハイブリッド発電	時系列関数	エージェントシステム	インタラクティビティ	サトウキビ
RashidMohd	0	0	0	0	0	0	0
長田智和	0	0	0	0	0	0	0
吉浦陽介	0	0	0	0	0	0	0
玉城史朗	0	0	1	0	0	0	0
木本守	0	0	0	0	0	0	0
谷口祐治	0	0	0	0	0	0	0
神里大	0	0	0	0	0	0	0
知念真也	0	0	0	0	0	0	0
比嘉哲也	0	0	0	0	0	0	0
城間政司	0	0	0	0	1	0	0
新垣秀雄	0	0	0	0	1	0	0
中村文也	0	0	0	0	0	0	0
名嘉村盛和	1	0	1	1	0	0	1

各行(著者)間の類似度を計算、樹形図を描いてみる



「大向一輝」を基に130名の著者を抽出。それぞれが執筆した論文のキーワードを基に樹形図にしてみた(データ取得にすごく時間がかかった)



「大向一輝」を基に130名の著者を抽出。それぞれが執筆した論文のキーワードを基に樹形図にしてみた(データ取得にすごく時間がかかった)

# —キーワード間のつながり—

- キーワードもクラスタリングしてみた



26名の著者を抽出。それぞれが執筆した論文のキーワードを基にした樹形図

# さらに分かりやすく2次元上にプロット

分散協調 連携管理・運用

サーバー構築  
仮想化技術 Benchmark

教育支援

Server Benchmark Construction  
Virtualization

モバイルP2Pエージェント

mobile\_file\_delivery\_system  
mobile\_agent  
ファイル配信システム

輻輳制御 Congestion Control

システム設計

QoS

ネットワーク分散システム

総合学習

RED

DiffServ

管理システム

データベース

小学校教育

教材開発

**教育関係  
の単語群**

エージェントシステム  
Photovoice  
Wireless-LAN  
Multipoint  
重回帰分析  
Agent-system  
自律的ITシステム  
Autonomous System  
太陽光発電システム

学校と家庭のコミュニケーション

# メタデータの「つながり」に着目する理由

- 検索漏れの少ないシステムを目指して
  - 新語、訳語の確定していないキーワード同士をレコメンドできるシステム
- レコメンデーション、適合度基準として利用
  - 類似の論文をまとめてレコメンドできるシステム
- 活発な研究分野、重要な論文の可視化
  - 論文間の関係を分析する事で可視化
- 「偶然の出会い」の創出
  - ブール検索では難しいが、より創出したい

# —論文間のつながり—

- 引用/被引用を基にクラスタリング

## Web APIで未提供

(ログイン後にしか見えない情報の一つ)

技術的にはHTMLから取得可能だが、  
無断でやってはまずい

「システムやプログラム等で動的な処理をされる場合は、事前にまでご連絡ください。  
折り返し、手続き方法について連絡させていただきます。」CiNii著作権とリンクより



# 「集合知プログラミング」してみたの雑感

- 適当な分析でも意味がありそうな「つながり」を発見する事ができ、面白かった
- 検索の際にこの情報を使う(適合度として、レコメンドとして)と効果がありそう
- メタデータを大量にWeb APIから取るのは時間がかかる
- APIで取得できない情報がある
- そもそもメタデータ大量取得のためにAPIを使うのは間違い？
- メタデータをもっと(丸ごと)見てみたい

CiNiiへの期待

# CiNiiへの期待

## 公開部分の拡大

### –引用/被引用論文情報の公開

- 論文間の関係「つながり」をみたい

### –API経由の認証機能の実装

- 登録ユーザはログイン時と同様の機能が利用可能にしてほしい(PPV等)

# CiNiiへの期待

## 公開部分の拡大

—本文へのリンクの公開

- せっかくアプリケーションを作るなら  
利用者へ一次情報を提供したい

# CiNiiへの期待

## メタデータを丸ごと提供

- それ自体が研究対象になる
- 独自の検索システムを作る
  - 連想検索したり、レコメンドしたり・・・、  
独自のインタフェース作ったり

# CiNiiへの期待

## メタデータを丸ごと提供

- それ自体が研究対象になる
- 独自の検索システムを作れる
  - 連想検索したり、レコメンドしたり・・・、  
独自のインタフェース作ったり
  - 検索システムに正解はない
  - 多様化できる仕組が欲しい

# CiNiiへの期待

## —持っている情報のさらなるオープン化

- 新たなサービスの可能性は果てしない。革新的なサービスの出現を期待してオープン化を進めて欲しい

## —無料範囲の拡大

- PPVの低価格化(せめてILL以下に…)
- 「普通の人」への恩恵をさらに広げてほしい!

## —わくわくするような新規技術の積極導入

- 新技術がサービスとして見れることを楽しみにしています

# 参考資料

「集合知プログラミング」  
オライリージャパン



# APIコンテスト応募者へ

マッシュアップもいい  
ですが、集合知として  
メタデータを使っても  
面白いと思います