

# 統計分野における 研究者の氏名同定とその応用

高久雅生<sup>†</sup>, 相澤彰子<sup>††</sup>, 大山敬三<sup>††</sup>, 馬場康維<sup>†††</sup>

<sup>†</sup> 情報・システム研究機構

<sup>††</sup> 国立情報学研究所 / 総合研究大学院大学

<sup>†††</sup> 統計数理研究所

masao@nii.ac.jp

# 背景

- エンティティ検索
- 人物情報検索
  - 専門家探し (Expert search) などの応用
  - サービス
    - SNS
    - Webサーチエンジン

**spock** BETA

Type in a name, keyword, or location to find people.

[Sign-up](#) to see where everyone you know is on the web

Steve Jobs | [Web \(71\)](#) | [Pictures \(21\)](#) | [Tags \(149\)](#) | [News \(250\)](#) | [Related People \(20\)](#)

**Steve Jobs** - male, 52 years old, [Cupertino, CA, United States](#)  
[Steven Paul Jobs](#) - [Steven P. Jobs](#) - [Steven P. Jobs](#) - [Add name](#)



[Add picture](#) | [Vote on pictures](#) | [All pictures](#)

Steven Paul Jobs (born February 24 1955) is the co-founder, chairman, and CEO of Apple and was the CEO of Pixar until its acquisition by Disney. He is currently the largest Disney shareholder and a member of the Board of Directors. He is considered a leading figure in the computer and entertainment industries and has contributed greatly to the development of the modern Silicon Valley entrepreneur.

Source: [Wikipedia](#)

**spock** BETA

Type in a name, keyword, or location to find people.

[Sign-up](#) to find people you know on the web

Yasuo Fukuda | [Web \(1\)](#) | [Pictures \(2\)](#) | [Tags \(4\)](#) | [News \(6\)](#)

**Yasuo Fukuda** - male, 72 years old  
[Add name](#)



[Add picture](#) | [Vote on picture](#)

Yasuo Fukuda is a Japanese politician who served as the 9th Prime Minister of Japan from September 16, 2006, to September 26, 2006. He was the first Prime Minister to be born after World War II. Fukuda dropped out of politics in 2006.

Source: [Wikipedia](#)

[News](#) - [Add news](#) | [See all](#)

Be the first to add news to Yasuo Fukuda's search result.

[Websites](#) - [Add website](#) | [See all](#)

[Wikipedia](#)

**spock** BETA

Type in a name, keyword, or location to find people.

[Sign-up](#) to see where people you know are on the web

Bill Gates | [Web \(26\)](#) | [Pictures \(29\)](#) | [Tags \(107\)](#) | [News \(250\)](#) | [Related People \(20\)](#)

**Bill Gates** - male, 52 years old, [Seattle, WA, United States](#)  
[William Henry Gates III](#) - [Add name](#)



[Add picture](#) | [Vote on pictures](#) | [All pictures](#)

William Henry Gates III (born October 28, 1955) is an American entrepreneur, philanthropist, and the chairman of Microsoft, the software company he founded with Paul Allen. During his career at Microsoft he has held the positions of CEO and chief software architect, and he remains the largest individual shareholder with more than 8% of the common stock. "Forbes" magazine's list of The World's Billionaires has ranked him as the richest person in the world since 1995, ...

Source: [Wikipedia](#)

[Tags](#) - [Add tag](#) | [Vote](#) | [See all](#)

[Microsoft](#) - [Microsoft founder](#) - [Bill & Melinda Gates Foundation](#) - [college dropout](#) - [geek](#) - [Windows](#) - [billionaire](#) - [philanthropist](#) - [Forbes World's Richest People](#) - [rich](#) - [software magnate](#) - [computer programmer](#) - [American billionaire](#) - [American](#) - [computer pioneer](#)

[Related People](#) - [Add people](#) | [See all](#)



[Login](#) | [Sign up](#) | [Linus Torvalds nemesis](#)

[Login](#) | [Sign up](#)

**spock** BETA

Type in a name, keyword, or location to find people.

[Sign-up](#) to see where people you know are on the web

Shinzo Abe | [Web \(2\)](#) | [Pictures \(1\)](#) | [Tags \(14\)](#) | [News \(0\)](#) | [Related People \(1\)](#)

**Shinzo Abe** - male, 53 years old  
[Add name](#)



[Add picture](#) | [Vote on pictures](#) | [All pictures](#)

Shinzo Abe (born September 21 1954) is the current Prime Minister of Japan, elected by a special session of the National Diet on September 26 2006. He is Japan's youngest post-World War II prime minister and the first born after the war. Abe was born into a political family, and studied political science in Japan, and had studied in the United States. He worked in the private sector until 1982 when he began work in several government jobs.

Source: [Wikipedia](#)

[News](#) - [Add news](#) | [See all](#)

Be the first to add news to Shinzo Abe's search result.

[Websites](#) - [Add website](#) | [See all](#)

[Wikipedia](#)

The biggest multilingual free-content encyclopedia on the Internet. Over 7 million articles in over 200 languages, and still growing.

[Tags](#) - [Add tag](#) | [Vote](#) | [See all](#)

[Government minister of Japan](#) - [from Yamaguchi Prefecture](#) - [National Diet](#) - [Chief Cabinet Secretary](#) - [Liberal Democratic Party](#) - [Yamaguchi Prefecture](#) - [Diet of Japan](#) - [National Leader](#) - [Japanese](#) - [prime minister of Japan](#) - [north Korea](#) - [minister](#) - [political science](#) - [World War II](#)

[Related People](#) - [Add people](#) | [See all](#)



[Quotes](#) - [Add quote](#) | [See all](#)

Be the first to add a quote to Shinzo Abe's search result.

[Tags](#) - [Add tag](#) | [Vote](#) | [See all](#)

[Wikipedia](#) - [Wikipedia founder](#) - [Wikia](#) - [Wikipedia person](#) - [Wikimedia Foundation](#) - [Internet personality](#) - [business owner](#) - [Search Wikia](#) - [co-founder of Wikipedia](#) - [entrepreneur](#) - [American entrepreneur](#) - [Business 2.0 The 50 Who Matter Now](#) - [Jimmy Donal 'Jimbo' Wales](#) - [Jimbo Wales](#) - [Internet celebrity](#)

[Related People](#) - [Add people](#) | [See all](#)



[Larry Sanger](#) - [co-founder of Wikipedia](#) | [Amber](#) - [Myspace friend](#)



[Ted](#) - [Myspace friend](#) | [Irene McGee](#) - [Myspace friend](#)

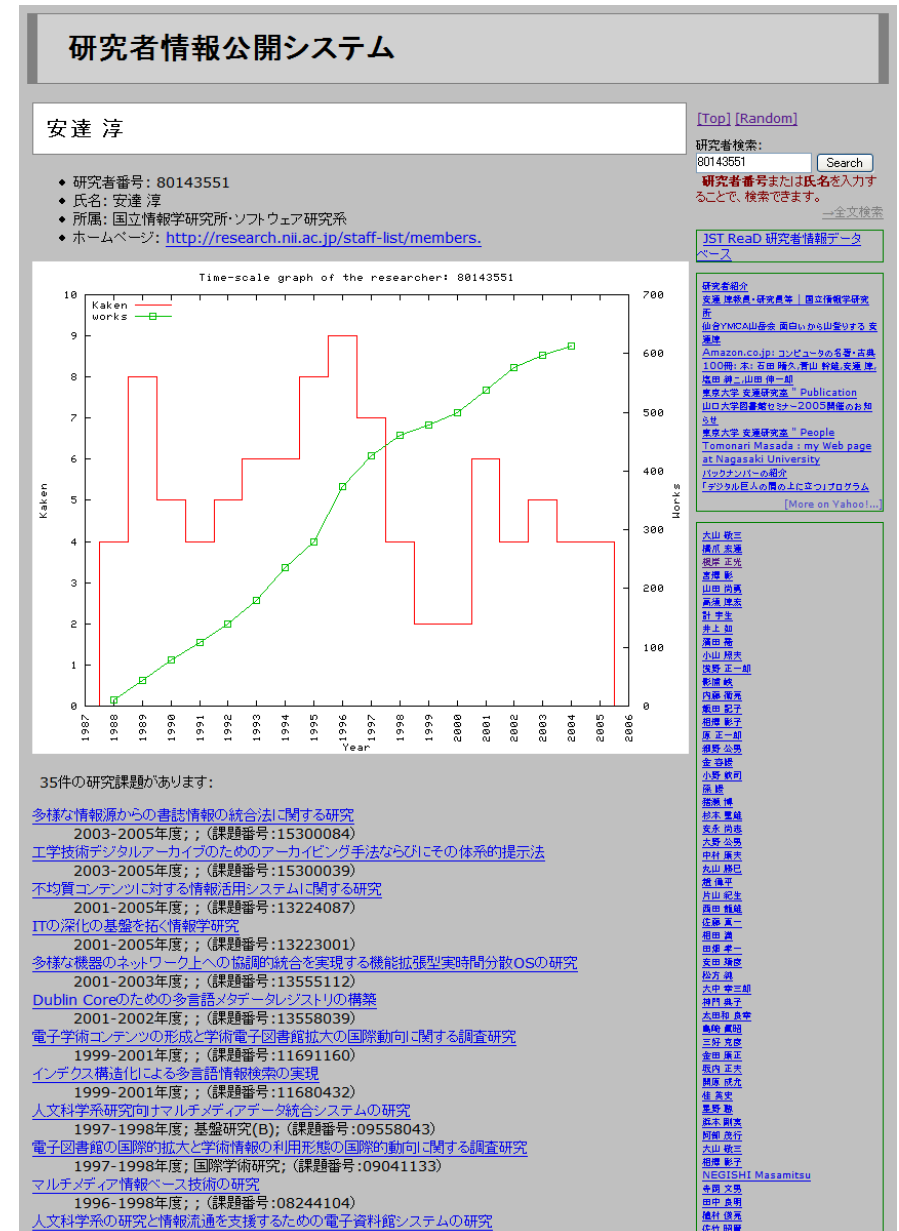
<http://www.spock.com>

# 目的

- 名寄せ
  - － 別々のリソースから得られた人物情報を照合して統合
  - － 同姓同名は様々なアプリケーションで同定が必要
  - － コスト見積もり
- 研究者に対して：
  - － 科研費研究者番号（日本国内の研究者）  
（→ これまでに「研究者情報サーバ」を構築）
- 統計分野に焦点を当てたデータセットの構築
  - － 研究者の同定・検索において、その評価や実装を助けるため「統計分野」を対象とした（統計3学会の学会員名簿を元にした）データセットを作成

# 研究者情報サーバ

- 元データ(科研費DB)
  - (科研費採択課題) 247,745件
    - 主に1989年度から2004年度報告分まで
  - (研究代表者・分担者) 133,067名分
    - 個別研究者情報
  - (報告書記載文献数) 1,896,070 件(重複含)
- 研究者の基本情報
  - 氏名、所属、職位
- 採択課題情報
- 関連する発表文献情報
  - 報告書記載分
- Web情報源とのリンク
  - ReaDデータベース
  - Yahoo!
- 経年的な活動の様子を可視化
  - 採択課題・関連発表論文数
- 共同研究者ネットワークを可視化
  - 過去の研究課題における共同研究者
  - GoogleMaps連携



# 統計分野(3学会)データセット

- 学会名簿を対象(延べ3,031名)
  - － 日本統計学会(2003年): 1,545名
  - － 日本行動計量学会(2005年): 1,070名
  - － 日本計算機統計学会(2002年): 416名
- 登録情報: 氏名、所属、住所、郵便番号等
  - － 今回は氏名と所属のみを用いる

# 名寄せ作業

- 氏名表記による研究者情報サーバ登録者と照合：
  - 延べ1,399件
    - 日本統計学会： 696件 (45%)
    - 日本行動計量学会： 519件 (49%)
    - 日本計算機統計学会：184件 (44%)
    - なお、簡易に重複を除去すると延べ1,307件
      - (所属・氏名表記が一致するもののみ)
- これらの同定対象候補と研究者番号とを人手で同定する

# 研究者名寄せ・同定

- 氏名表記に対して一意となる識別IDを付与

同定候補集合

統計3学会 名簿	科研費 研究者番号
BSJ0004	20024581
BSJ0005	50305313
JSS0012	70303047 09246528
BSJ0007	90184332
JSS0014	90184332
JSS0015	90132696

同定結果

統計3学会 名簿		科研費 研究者番号
BSJ0004	別人	20024581
BSJ0005	同一	50305313
JSS0012	別人	70303047 09246528
BSJ0007	同一	90184332
JSS0014	同一	90184332
JSS0015	同一	90132696

重複

判定作業



# 統計分野データセット・判定作業

統計学分野研究者同定システム - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

戻る 進む 更新 中止 ホーム http://localhost/~masao/private/cvswork/jstat-dataset/app.cgi

## 統計学分野研究者同定システム

同定候補 (1-20):

- 再判定
- 野口 隆士 == 野口 隆士 再判定
- 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定
- 山田 隆士 == 山田 隆士 再判定

ページ: [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [65] [66] [67] [68] [69]

[候補統合へ](#)

統計学分野研究者同定システム - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

戻る 進む 更新 中止 ホーム http://localhost/~masao/private/cvswork/jstat-dataset/app.cgi?start=320

## 統計学分野研究者同定システム

同定候補 (321-340):

- 再判定
- 再判定
- 再判定
- 再判定
- 再判定
- 関西大学商学部 (JSS0098)  
80202745 ☐ 同一人物 ☐ 別人  
同定確定
- 青山学院大学経営学部 (JSS0393)  
20303050 ☐ 同一人物 ☐ 別人  
同定確定
- 京都学園大学人間文化学部人間関係学科 (BSJ0226)  
60240628 ☐ 同一人物 ☐ 別人  
同定確定
- 一橋大学経済学部 (JSS0400)  
70154838 ☐ 同一人物 ☐ 別人  
同定確定
- 豊田工業高等専門学校 (BSJ0229)  
40112833 ☐ 同一人物 ☐ 別人  
90216705 ☐ 同一人物 ☐ 別人  
同定確定
- 東京工業大学大学院情報理工学系研究科数学・計算科学専攻 (JSS0402)  
30154876 ☐ 同一人物 ☐ 別人  
40312300 ☐ 同一人物 ☐ 別人  
70016153 ☐ 同一人物 ☐ 別人  
同定確定
- 中外製薬(株)臨床解析部 (BSJ0230)  
60125391 ☐ 同一人物 ☐ 別人  
同定確定
- 筑波大学社会医学系 (JSS0404)  
80261808 ☐ 同一人物 ☐ 別人  
同定確定

完了

- ◆ [redacted] 京都学園大学人間文化学部人間関係学科 (BSJ0226)

[redacted] 60240628 ☒ 同一人物 ☐ 別人

同定確定

- ◆ [redacted] 一橋大学経済学部 (JSS0400)

[redacted] 70154838 ☒ 同一人物 ☐ 別人

同定確定

- ◆ [redacted] 豊田工業高等専門学校 (BSJ0229)

[redacted] 40112833 ☐ 同一人物 ☒ 別人

[redacted] 90216705 ☒ 同一人物 ☐ 別人

同定確定

- ◆ [redacted] 東京工業大学大学院情報処理工学研究科数学・計算科学専攻 (JSS0402)

[redacted] 30154876 ☐ 同一人物 ☒ 別人

[redacted] 40312300 ☐ 同一人物 ☒ 別人

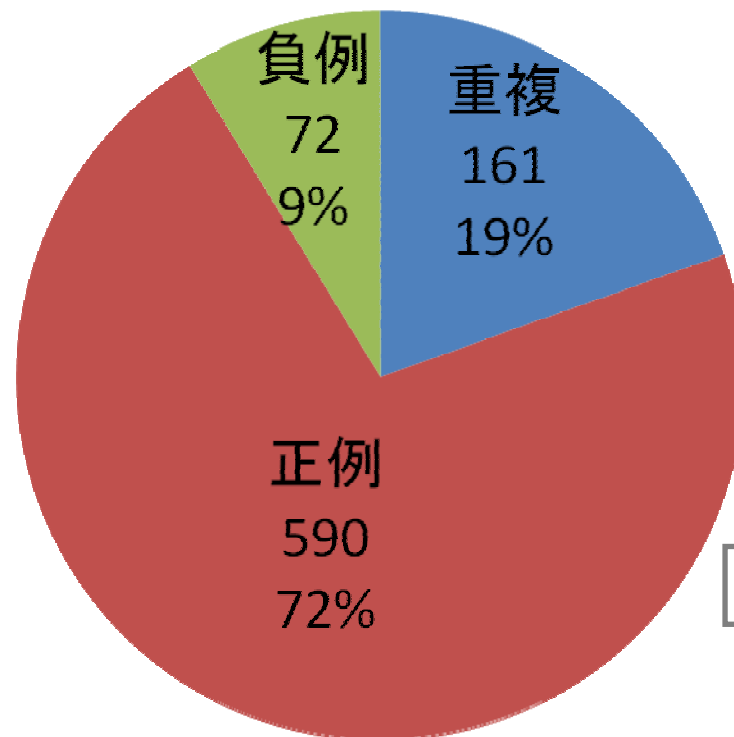
[redacted] 70016153 ☒ 同一人物 ☐ 別人

同定確定

「研究者情報サーバ」  
へのリンク

# 同定結果 (823候補・判定終了時点)

- 氏名表記が同一の161候補(19%)は重複分(複数学会に所属)として除去
- このうち、590候補は、少なくとも一つ以上の研究者サーバ上の研究者番号と同定できた(89%)。



統計3学会名簿		科研費 研究者番号
BSJ0004	別人	20024581
BSJ0005	同一	50305313
JSS0012	別人	70303047 09246528
BSJ0007	同一	90184332
JSS0014		
JSS0015	同一	90132696

重複

# 同姓同名候補

- 研究者情報サーバ内での同姓同名
  - 氏名表記が複数の研究者番号と一致した候補で、その判定が同一人物・別人と分かれたケース

→ 82候補

統計3学会名簿	科研費研究者番号
JSS0035	10308679
	80241595

- 研究者情報サーバ外を考慮した同姓同名
  - 氏名一致対象者が別人と判定されたケース

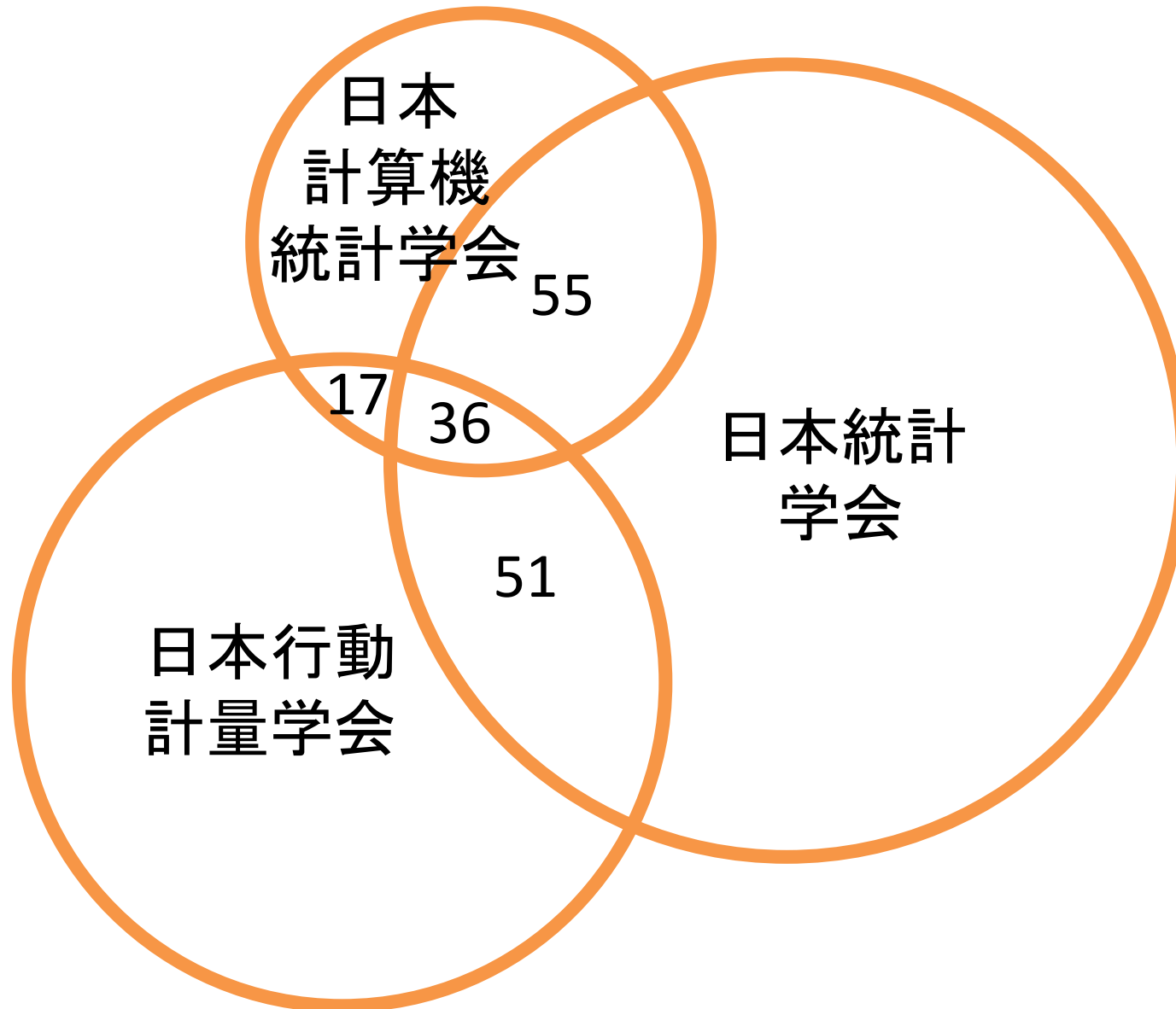
→ 72候補

統計3学会名簿	科研費研究者番号
BSJ0004	20024581
JSS0012	70303047
	09246528

- 上記2つの同姓同名を合わせると:

→ 154候補

# 学会間の重複登録者



# 考察(1) 同定基準

- 同定の際の基準は？
  - 研究者情報サーバの情報
    - 研究分野
    - 所属組織
    - 経年変化
  - JST ReaD研究者ディレクトリ上の情報
  - Web上の情報
    - 所属先Webサイト
    - 研究者自身が提供するWebサイト
  - 著書などの著者紹介
- 今回は左記のような情報を手掛かりに同定
  - 場合によっては、同定が困難なケースもあり、どのような情報源を用いるか、作業に使えるコスト、見積もりなど...
    - 想定する応用領域ごとに検討する必要あり
  - また、同定根拠を記録しておくことも検討

# 考察(2) 可視化

- 同定結果を可視化
  - 研究者の共同研究ネットワークによるもの
    - 科研費において共同研究したことがある者同士を接続
    - 関連研究者をリストアップ
  - 1ホップ先までの研究者を対象にネットワークを描画
- 応用
  - 同定対象候補の絞り込み
  - 研究分野の概観



## 判定結果の可視化

- : 同定済み研究者
- : 未同定判定候補者
- : 上記研究者の隣接研究者 (1-hop)

8410ノード, 43874エッジ



# まとめ

- 統計分野3学会と科研費DB研究者間を人手によって名寄せ同定
- 今後の課題
  - 分野全体像の把握
  - 同定の半自動化
  - 他分野データセットへの適用
  - エンティティ検索への同定結果の適用