Name Disambiguation of Japanese Researchers: A Case Study with Statistics Research Community

Masao Takaku[†], Yasumasa Baba[††], Akiko Aizawa[†††]

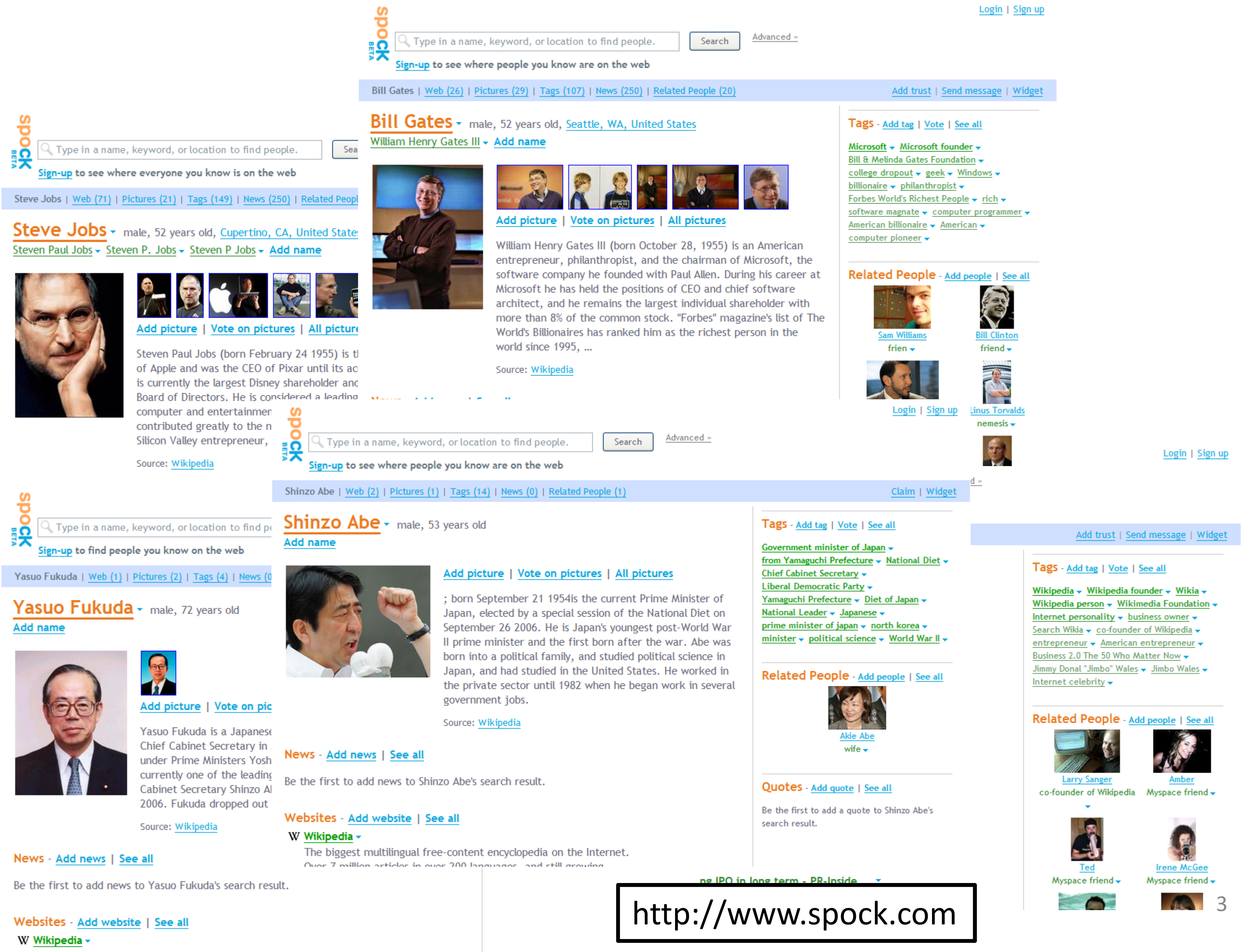
[†] National Institute for Materials Science

[††] Research Organization of Information and Systems / The Institute of Statistical Mathematics

[†††] National Institute of Informatics

Motivation

- There are many types of information on the Web
- Entity-based search
  - Categories for types of information
  - Events, place, **person**, etc.
- Searching for a person
  - Not sufficient to search Google with the name of the person
  - E.g. Expert search
  - Through the service of
    - SNS
    - Web search engine



spock BETA

Login | Sign up

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

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) | Add trust | Send message | Widget

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

Sam Williams frien

Bill Clinton friend

Linus Torvalds nemesis

---

spock BETA

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

Sign-up to see where everyone you know is on the web

Steve Jobs | Web (71) | Pictures (21) | Tags (149) | News (250) | Related Peopl

Steve Jobs male, 52 years old, Cupertino, CA, United State

Steven Paul Jobs Steven P. Jobs Steven P Jobs Add name

Add picture | Vote on pictures | All picture

Steven Paul Jobs (born February 24 1955) is t of Apple and was the CEO of Pixar until its ac is currently the largest Disney shareholder anc Board of Directors. He is considered a leading computer and entertainmen contributed greatly to the n Silicon Valley entrepreneur,

Source: Wikipedia

---

spock BETA

Login | Sign up

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

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) | Claim | Widget

Shinzo Abe male, 53 years old

Add name

Add picture | Vote on pictures | All pictures

; born September 21 1954is 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

W Wikipedia

The biggest multilingual free-content encyclopedia on the Internet.

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

Akie Abe wife

Quotes - Add quote | See all

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

---

spock BETA

Type in a name, keyword, or location to find p

Sign-up to find people you know on the web

Yasuo Fukuda | Web (1) | Pictures (2) | Tags (4) | News (0

Yasuo Fukuda male, 72 years old

Add name

Add picture | Vote on pic

Yasuo Fukuda is a Japanese Chief Cabinet Secretary in under Prime Ministers Yosh currently one of the leading Cabinet Secretary Shinzo A 2006. Fukuda dropped out

Source: Wikipedia

News - Add news | See all

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

Websites - Add website | See all

W Wikipedia

Login | Sign up

Add trust | Send message | Widget

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

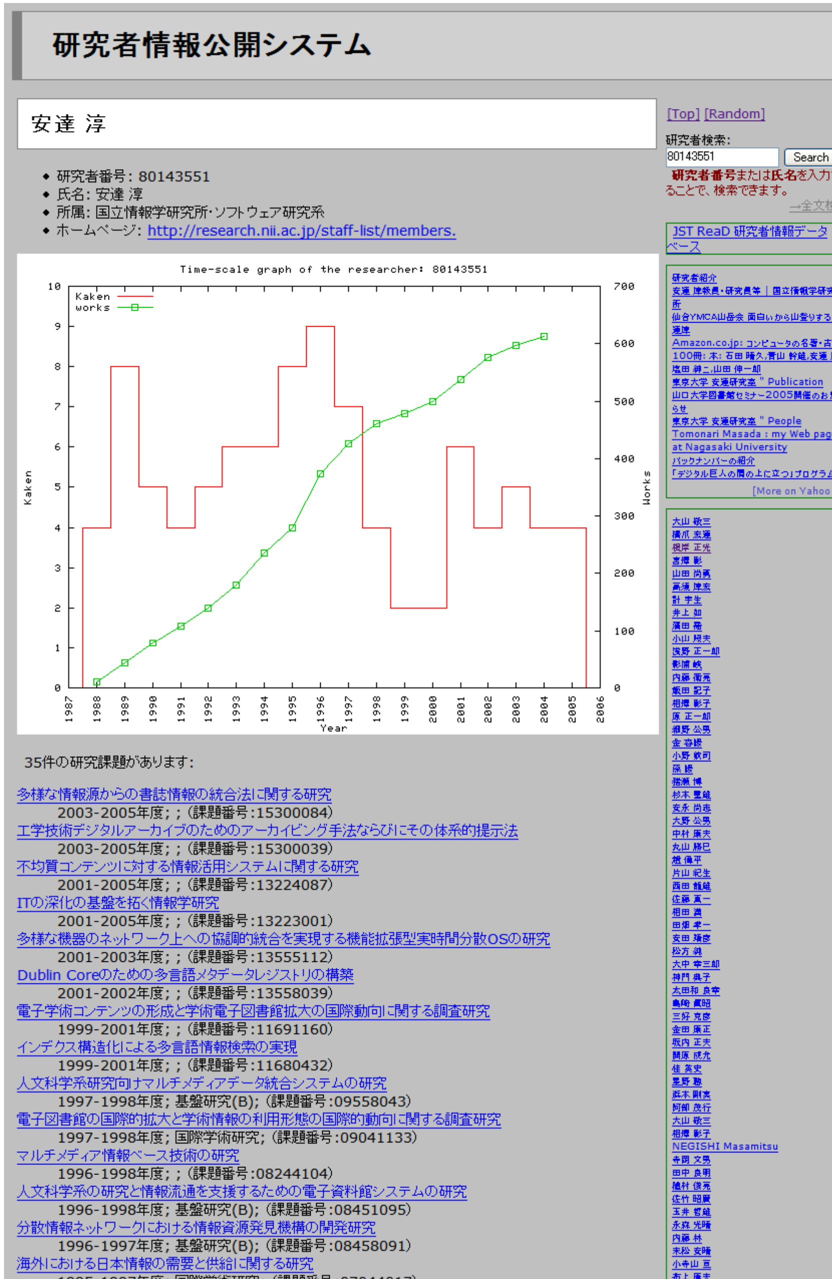
Objective

- Name identification
  - We integrated persons from different resources.
  - For identifying a different person with the same name, it is necessary to use several resources.
  - Estimating the costs
- Researcher’s information：
  - KAKEN Researcher ID（for researchers in Japan）

  （→ We developed “Researcher Information Server”.）
- We have made a dataset for statistics related researchers.
  - We intend to help evaluating and developing identification data and search system for researchers.
  - Data: Member lists of three academic associates for statistics related field in Japan

Researcher Information Server

- Data taken from KAKEN DB
  - # of projects (Funded by KAKEN): 247,745
    - Annual reports of the projects from 1989 to 2004
  - # of researchers (project leader and co-researchers): 133,067
    - Individual researcher information
- Basic information for researcher
  - Researcher name, affiliation name, and position (job title).
- Visualization
  - Timeline: # of KAKEN projects and publications
  - Co-researcher network:
    - Past KAKEN project's co-researcher
    - Using Google Maps



研究者情報公開システム

安達 淳

[Top] [Random]

研究者検索:

80143551 Search

研究者番号または**氏名**を入力することで、検索できます。

→全文検

- 研究者番号: 80143551
- 氏名: 安達 淳
- 所属: 国立情報学研究所･ソフトウェア研究系
- ホームページ: http://research.nii.ac.jp/staff-list/members.

Time-scale graph of the researcher: 80143551

JST ReaD 研究者情報データベース

研究者紹介
安達 淳 教員・研究員等 | 国立情報学研究所
仙台YMCA山岳会 面白いから山登りする 達淳
Amazon.co.jp: コンピュータの名著・古典100冊: 本: 石田 晴久,青山 幹雄,安達 淳,塩田 紳二,山田 伸一郎
東京大学 安達研究室 " Publication
山口大学図書館セミナー2005開催のお知らせ
東京大学 安達研究室 " People
Tomonari Masada : my Web page at Nagasaki University
バックナンバーの紹介
「デジタル巨人の肩の上に立つ」プログラム
[More on Yahoo

大山 敬三
橋爪 宏達
根岸 正光
吉澤 彰
山田 尚勇
高須 淳宏
計 宇生
井上 如
濱田 喬
小山 照夫
浅野 正一郎
影浦 峡
内藤 衛亮
飯田 記子
相澤 彰子
原 正一郎
細野 公男
金 容媛
小野 欽司
孫 媛
猪瀬 博
杉本 重雄
安永 尚志
大野 公男
中村 康夫
丸山 勝巳
趙 偉平
片山 紀生
西田 龍雄
佐藤 真一
相田 満
田畑 孝一
安田 靖彦
松方 純
大中 幸三郎
神門 典子
太田和 良幸
島崎 眞昭
三好 克彦
金田 康正
坂内 正夫
関原 成允
佐 英史
星野 聡
浜本 剛実
阿部 茂行
大山 敬三
相澤 彰子
NEGISHI Masamitsu
寺田 文男
田中 良明
植村 俊亮
佐竹 昭廣
玉井 哲雄
永森 光晴
内藤 林
末松 安晴
小寺山 亘
布上 康夫

35件の研究課題があります:

多様な情報源からの書誌情報の統合法に関する研究
　2003-2005年度; ; (課題番号:15300084)
工学技術デジタルアーカイブのためのアーカイビング手法ならびにその体系的提示法
　2003-2005年度; ; (課題番号:15300039)
不均質コンテンツに対する情報活用システムに関する研究
　2001-2005年度; ; (課題番号:13224087)
ITの深化の基盤を拓く情報学研究
　2001-2005年度; ; (課題番号:13223001)
多様な機器のネットワーク上への協調的統合を実現する機能拡張型実時間分散OSの研究
　2001-2003年度; ; (課題番号:13555112)
Dublin Coreのための多言語メタデータレジストリの構築
　2001-2002年度; ; (課題番号:13558039)
電子学術コンテンツの形成と学術電子図書館拡大の国際動向に関する調査研究
　1999-2001年度; ; (課題番号:11691160)
インデクス構造化による多言語情報検索の実現
　1999-2001年度; ; (課題番号:11680432)
人文科学系研究向けマルチメディアデータ統合システムの研究
　1997-1998年度; 基盤研究(B); (課題番号:09558043)
電子図書館の国際的拡大と学術情報の利用形態の国際的動向に関する調査研究
　1997-1998年度; 国際学術研究; (課題番号:09041133)
マルチメディア情報ベース技術の研究
　1996-1998年度; ; (課題番号:08244104)
人文科学系の研究と情報流通を支援するための電子資料館システムの研究
　1996-1998年度; 基盤研究(B); (課題番号:08451095)
分散情報ネットワークにおける情報資源発見機構の開発研究
　1996-1997年度; 基盤研究(B); (課題番号:08458091)
海外における日本情報の需要と供給に関する研究
　1995-1997年度; 国際学術研究; (課題番号:07044017)

Statistics related field researchers (in three academic societies)

- We used member lists of academic societies (total: 3,031)
  - The Japan Statistical Society (2003): 1,545
  - The Behaviormetric Society of Japan (2005): 1,070
  - The Japanese Society of Computational Statistics (2002): 416
- Registered information: name, affiliation name, job title, address, zip-code
  - Only name and affiliation name were used.

Identification (Step 1) Name matching

- Matching the researcher with the same name researcher in Research Information Server (KAKEN-DB)
  - A total of 1,400 candidates:
    - The Japan Statistical Society : 697 (45%)
    - The Behaviormetric Society of Japan : 519 (49%)
    - The Japanese Society of Computational Statistics: 184 (44%)
  - After removing duplicates: 1,307 candidates
    - Candidates of having exactly the same name and affiliation were considered as an identical researcher.
- We manually identify the following data to KAKEN researchers ID.

Identification (Step 2)
Manual judgments

- Adding a KAKEN ID for each member manually.

Candidates

Members of societies	KAKEN IDs
BSJ0004	20024581
BSJ0005	50305313
JSS0012	70303047 09246528
BSJ0007	90184332
JSS0014	90184332
JSS0015	90132696

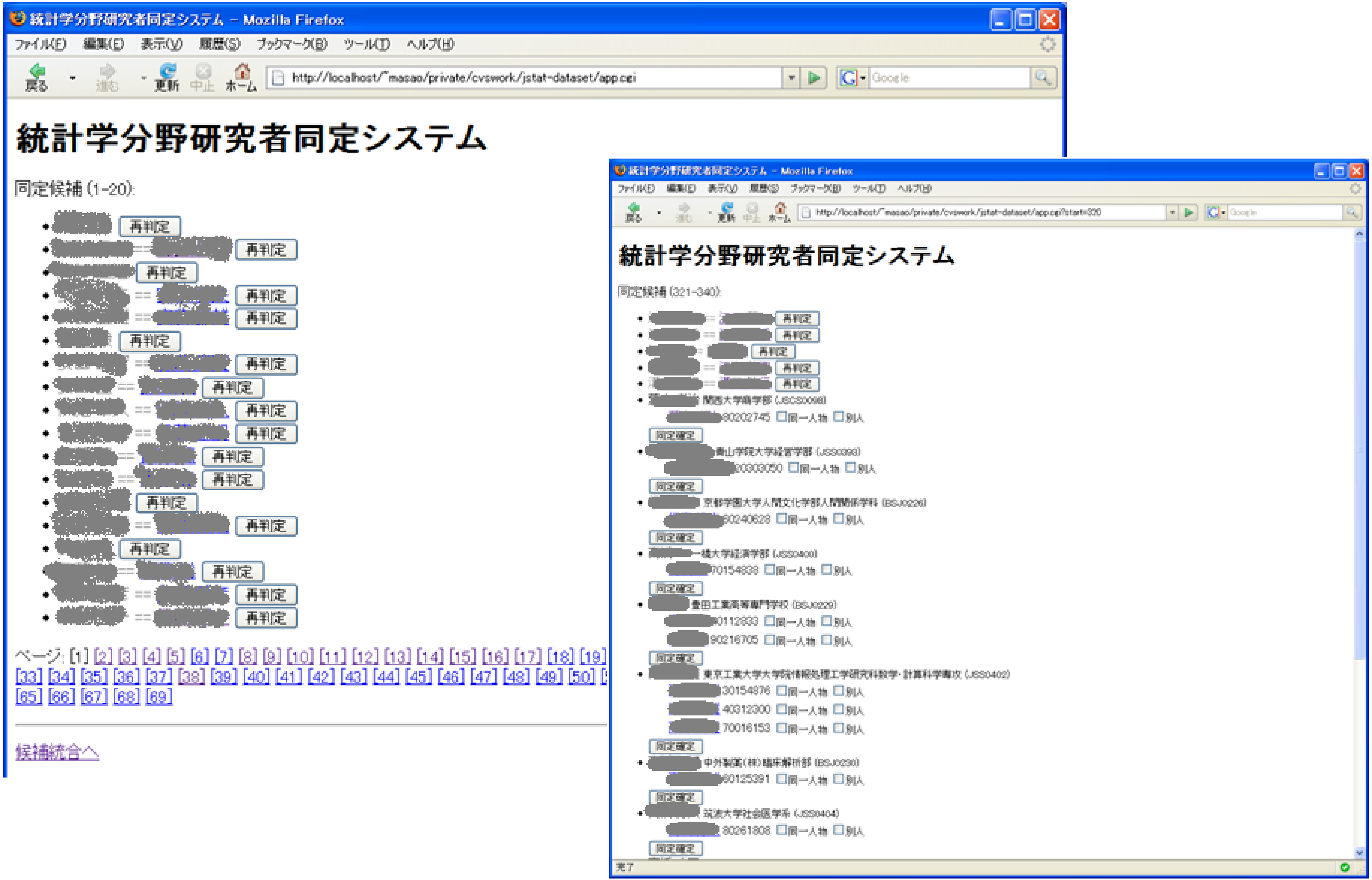
A member of multiple societies

Judged by manual

Results

Members of societies	KAKEN IDs
BSJ0004	Different ~~20024581~~
BSJ0005	Identical 50305313
JSS0012	~~70303047~~ Different ~~09246528~~
BSJ0007 JSS0014	Identical 90184332
JSS0015	Different 90132696

Judgment system for identification



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):

- [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 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

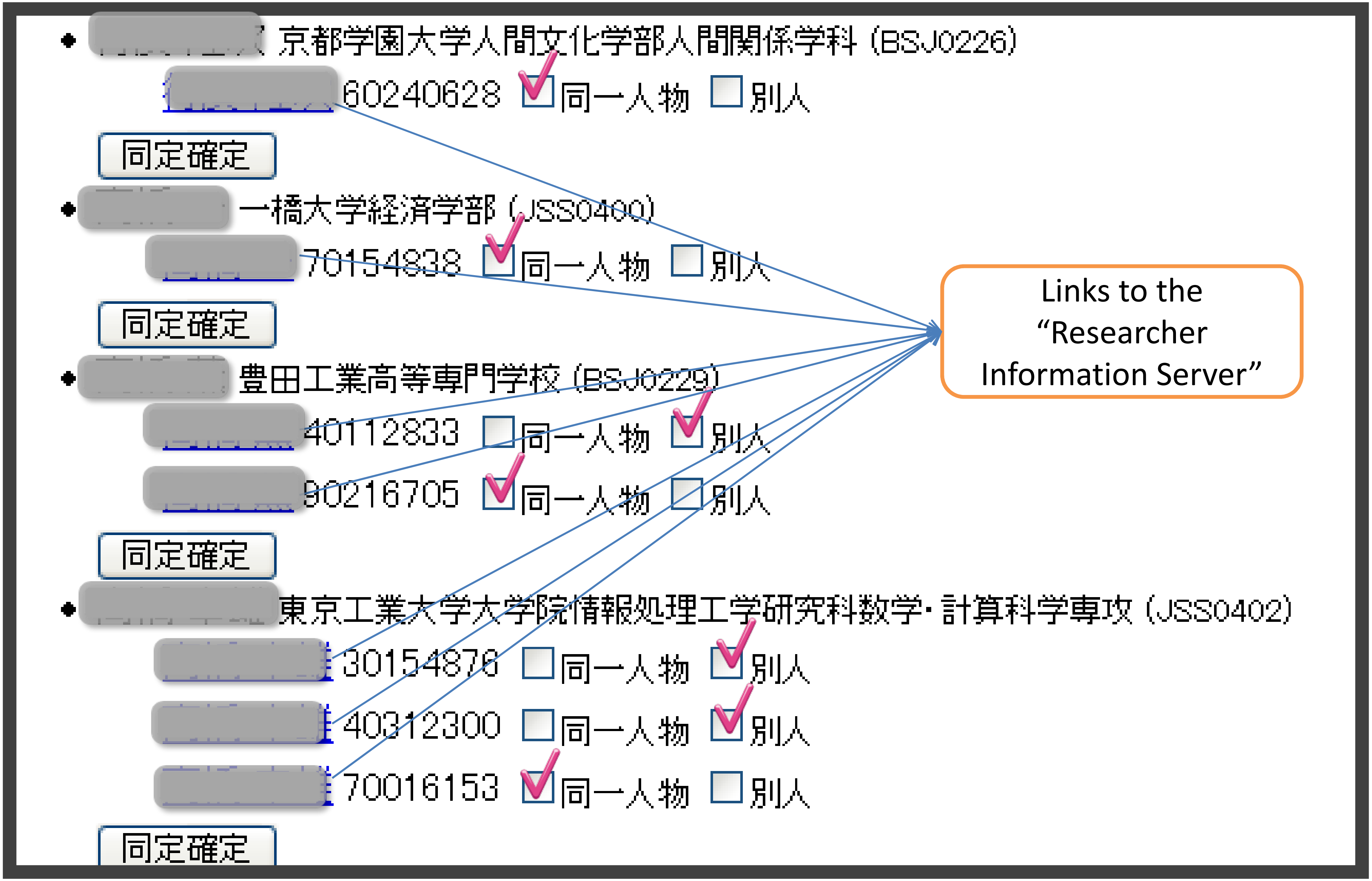
http://localhost/~masao/private/cvswork/jstat-dataset/app.cgi?start=320

統計学分野研究者同定システム

同定候補 (321-340):

- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] = [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] == [redacted] 再判定
- [redacted] 関西大学商学部 (JSCS0098)
  [redacted] 80202745 □同一人物 □別人
  同定確定
- [redacted] 青山学院大学経営学部 (JSS0393)
  [redacted] 20303050 □同一人物 □別人
  同定確定
- [redacted] 京都学園大学人間文化学部人間関係学科 (BSJ0226)
  [redacted] 80240628 □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] 80112833 □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
- [redacted] 中外製薬(株)臨床解析部 (BSJ0230)
  [redacted] 80125391 □同一人物 □別人
  同定確定
- [redacted] 筑波大学社会医学系 (JSS0404)
  [redacted] 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

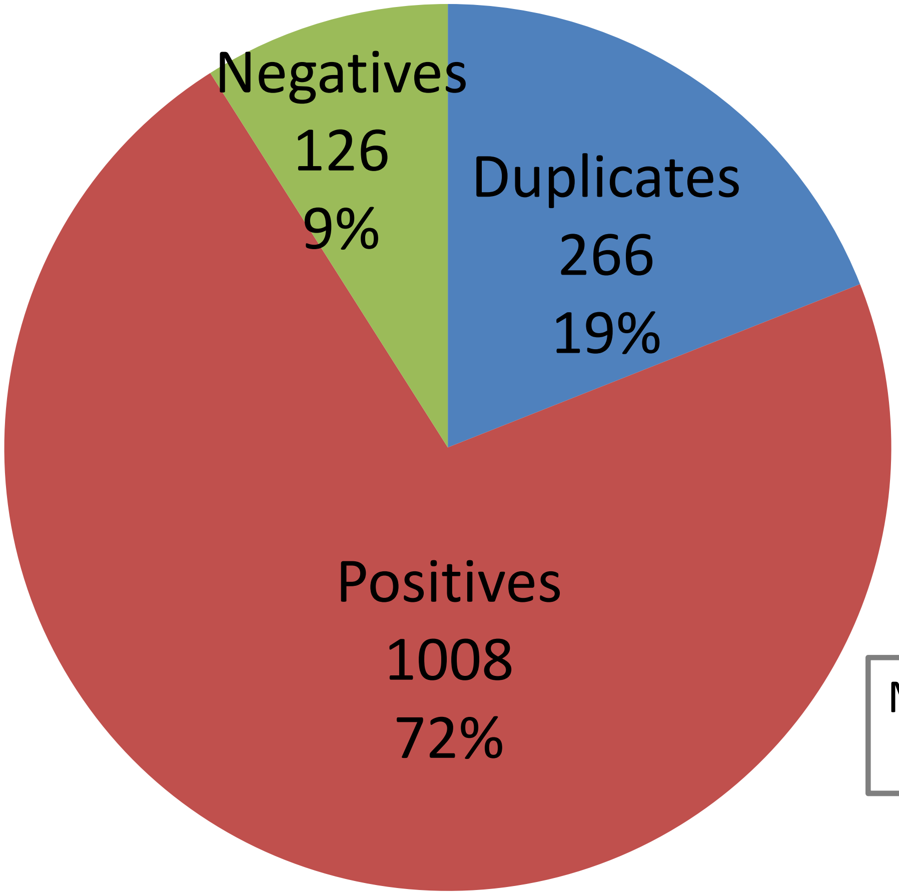
  同定確定
- [redacted] 東京工業大学大学院情報処理工学研究科数学・計算科学専攻 (JSS0402)
  - [redacted] 30154876 □同一人物 ☑別人
  - [redacted] 40312300 □同一人物 ☑別人
  - [redacted] 70016153 ☑同一人物 □別人

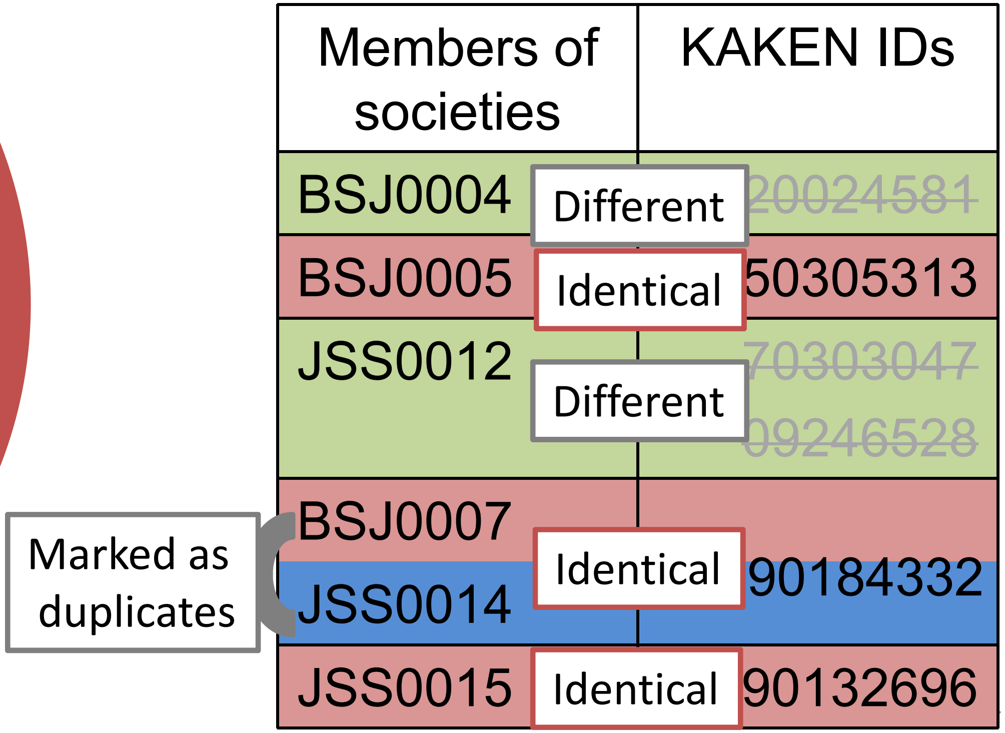
  同定確定

Links to the "Researcher Information Server"

Results

- 266 candidates were marked as “duplicates” (members of multiple societies)
- 1008 candidates were identified with a researcher within KAKEN-DB.





Members of societies		KAKEN IDs
BSJ0004	Different	~~20024581~~
BSJ0005	Identical	50303313
JSS0012	Different	~~70303047~~ ~~09246528~~
BSJ0007 JSS0014	Identical	90184332
JSS0015	Identical	90132696

Marked as duplicates

Researchers having the same name

- (a) Within KAKEN-DB (for positives)
  - A member’s name matched with KAKEN ID, and a different researcher having the same names within KAKEN-DB.

  → 105 candidates (10%)

Members of societies	KAKEN IDs
JSS0035	Identical 10308679
	Different ~~30241595~~

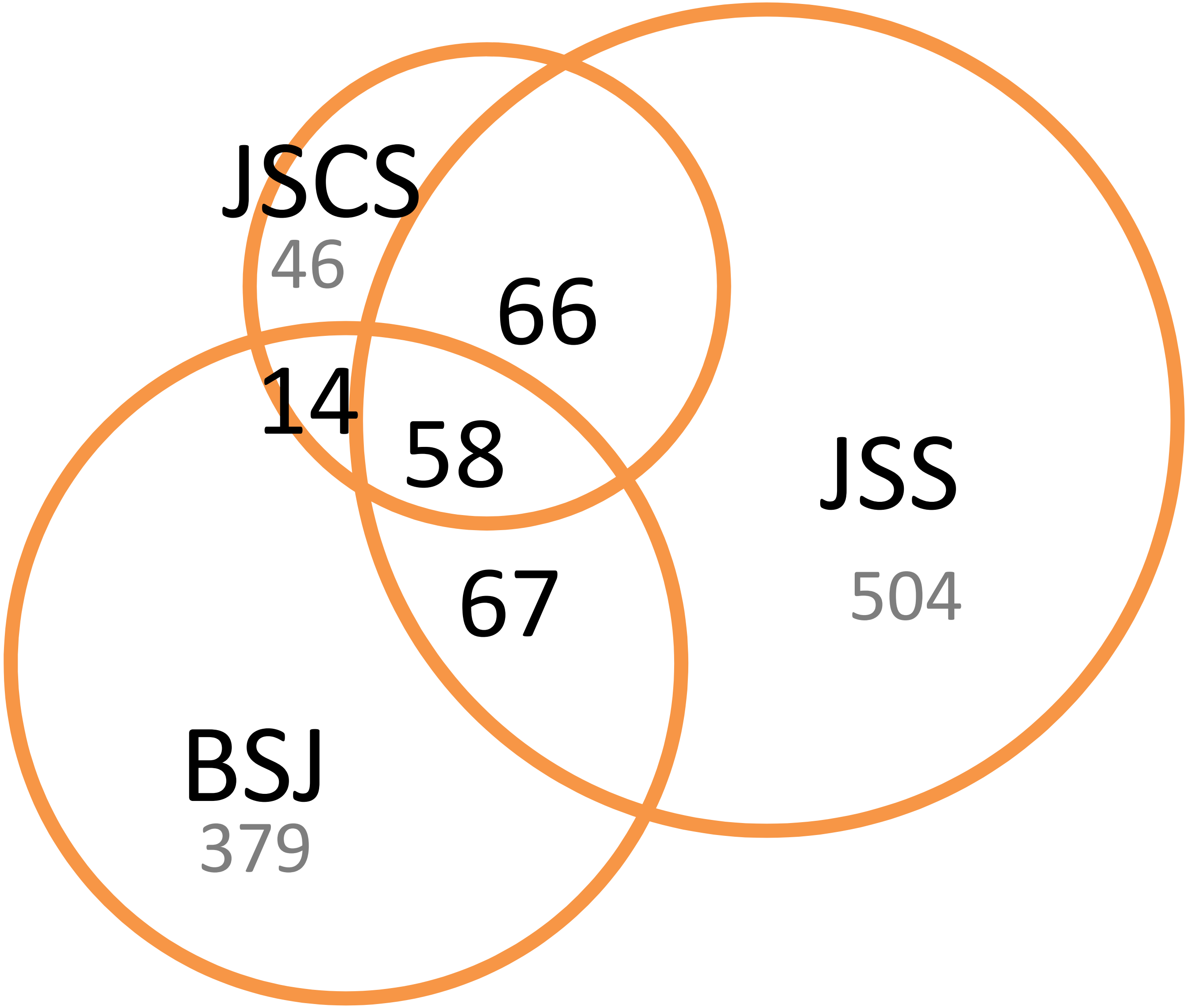
- (b) Concerning outside of KAKEN-DB (for negatives)
  - A member’s name matched with KAKEN ID, but that represented a different researchers.

  → 126 candidates

Members of societies	KAKEN IDs
BSJ0004	Different ~~20024581~~
JSS0012	~~70303047~~ Different ~~09246528~~

- (a) + (b) → 231 candidates (20%)

Duplicates (overlaps among the Japanese academic societies in Statistics-related field)



Discussion: Criteria for identification

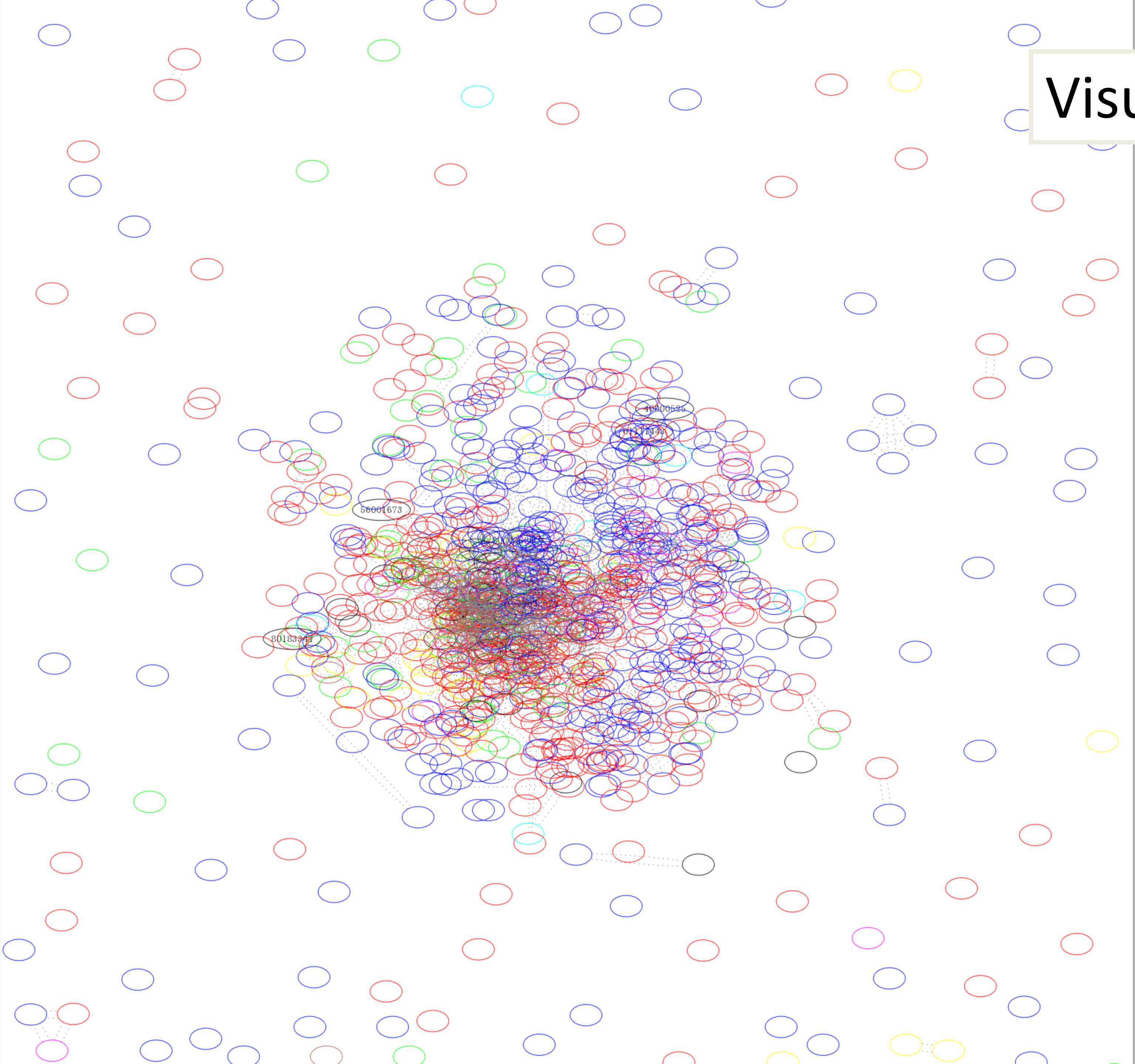
- How to determine two researchers are identical or not?
  - Researcher Information Server
    - Research area
    - Affiliations
    - Timeline based info.
  - Other resources:
    - JST ReaD (researcher directory)
    - On the Web
      - Organization website the researcher affiliates
      - Homepage provided by the researcher him/herself
  - Biography information provided as author introduction in his/her publications

- For statistics related field, left-side info are useful.
  - (Of course) His/her research fields are related to Statistics. It is a main hints for
- But, Sometime there is a difficult case to determine
  - By lack of information
  - It depends on the available resources and costs .
    - Need to clarify to utilize after the identification
- (Another possibility) Saving criteria and process for identification for the further analysis (afterwards).

Discussion: Visualizations

- Visualization of co-researchers network
  - Provides summary of researcher's community.
  - Connects researchers with having joint-research in the past.
  - Shows one more hop into the other researchers.
- Applications (It will be helpful for …)
  - Judgment of Identification
  - Summarization of researchers network (community)

Visualization



- JSS
- BSJ
- JSCS

Conclusions

- Identification of Japanese researchers in the field of Statistics
  - Identification between academic societies and KAKEN-DB
  - Identification of about 3,000 Statistics-related researchers with about 130,000 database records leads to about 1,000 identified persons.
- Future works
  - Semi-automatic process for identification (by using machine learning techniques)
  - Another dataset and resources
    - Other databases, transcriptions, and so on.
  - Comparison with other scientific fields
  - Represent a domain knowledge from researcher community
  - Entity-based search engine by identification results