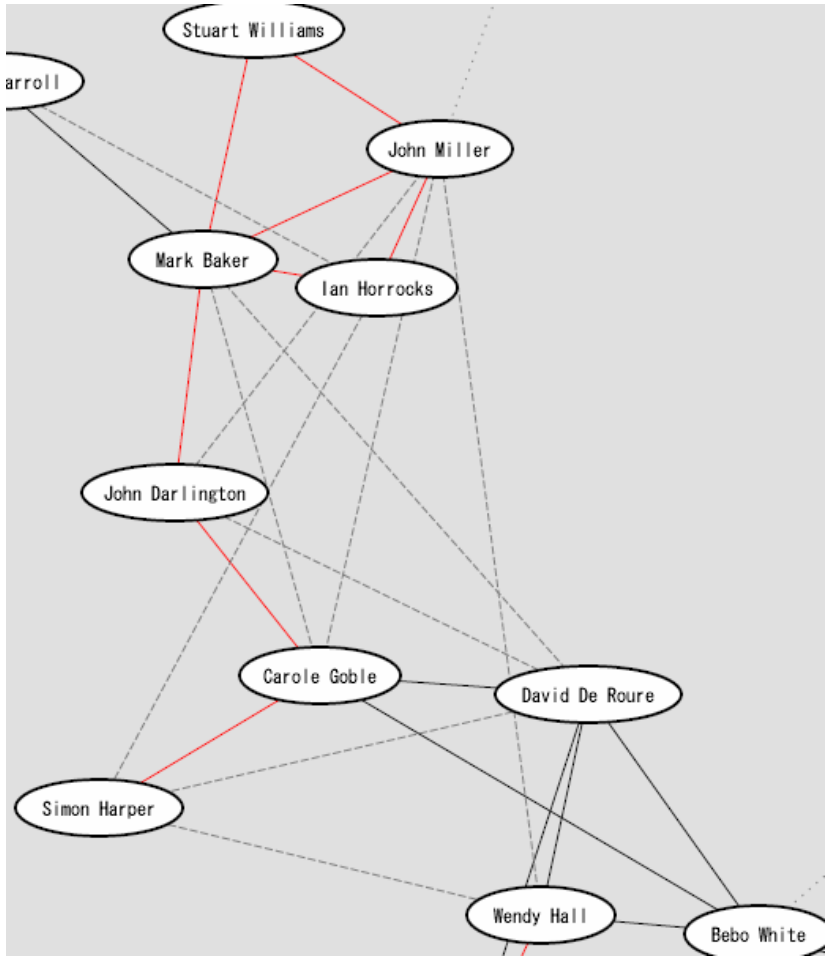
POLYPHONET: An Advanced Social Network Extraction System from the Web

Yutaka Matsuo
National Institute of
Advanced Industrial Science and Technology (AIST), Japan

J. Mori, K. Ishida, M. Hamasaki, H. Takeda, T. Nishimura, K. Hasida, and M. Ishizuka
Univ. of Tokyo and National Institute of Informatics

Social network

- Represent relations (*ties*) among persons (*actors*)
- Important for
  - Semantic Web, knowledge management, information retrieval, ubiquitous computing…
  - Our lives are enormously influenced by relations to others.
- Network data
  - Questionnaires in sociology
    - "Please name your four closest friends"
  - SNS
    - Friendster, orkut, myspace, LiveJournal, Yahoo!360…
  - FOAF documents
  - Web link analysis, citation analysis, blog analysis…
- **Our approach: Web mining approach using a search engine**

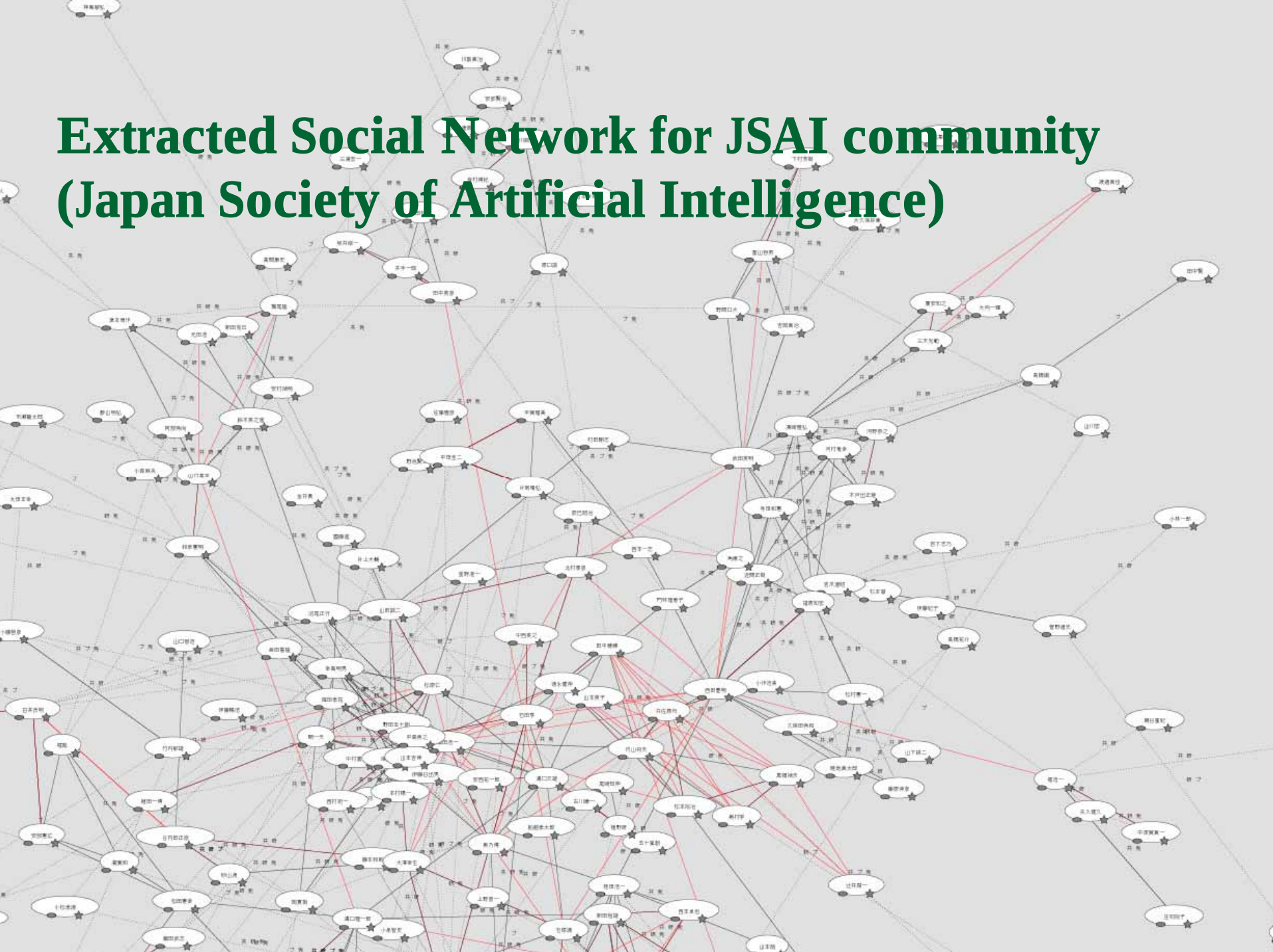


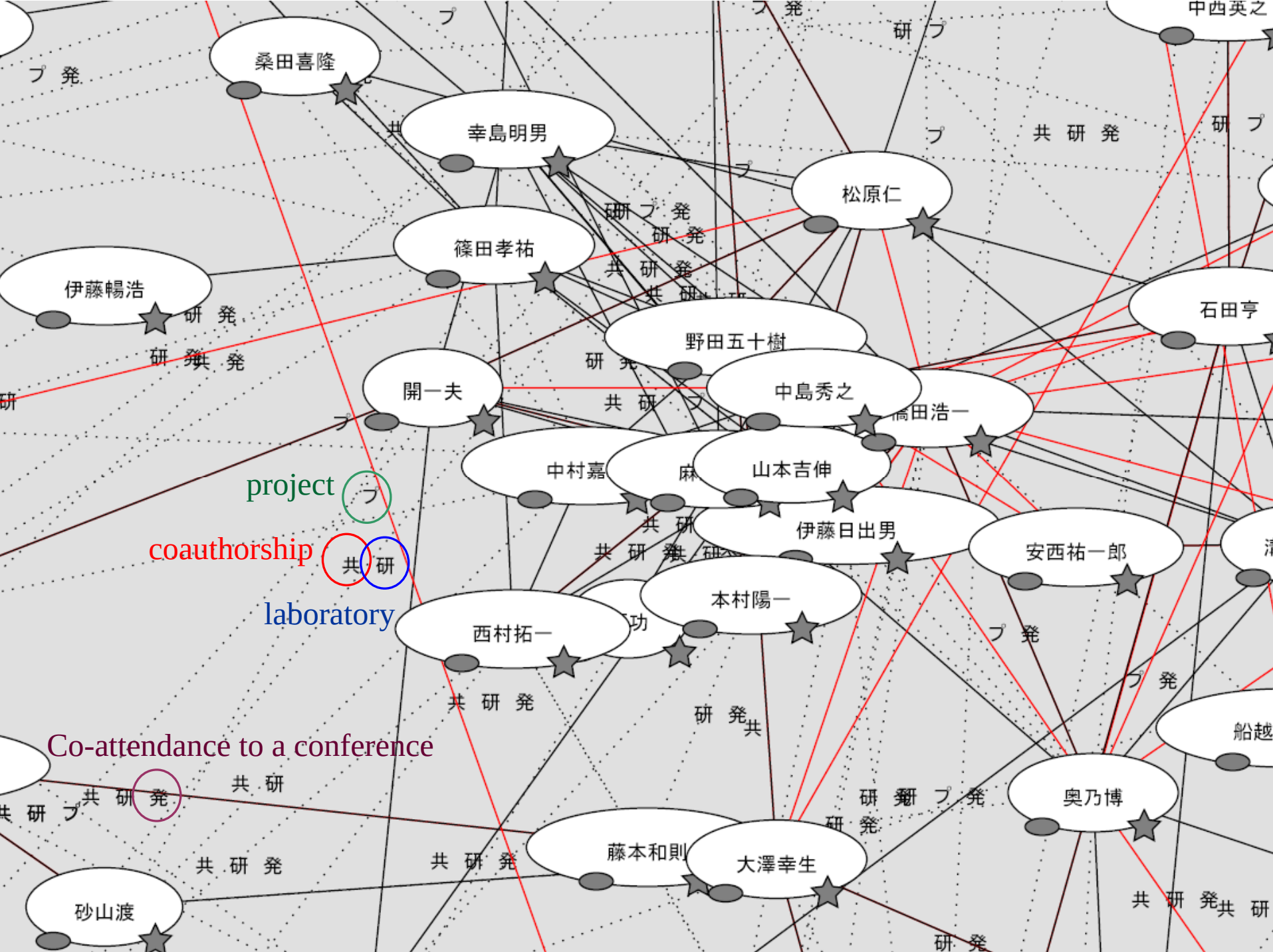
A social network of WWW2006 organizers

Content

- Social network mining from the Web
  - Algorithms and evaluations
  - Advanced algorithms
- POLYPHONET
  - Academic conference support system
- Future directions

Extracted Social Network for JSAI community (Japan Society of Artificial Intelligence)




桑田喜隆
幸島明男
松原仁
篠田孝祐
伊藤暢浩
石田亨
野田五十樹
開一夫
中島秀之
高田浩一
中村嘉
山本吉伸
伊藤日出男
安西祐一郎
本村陽一
西村拓一
船越
奥乃博
藤本和則
大澤幸生
砂山渡
project
coauthorship
laboratory
Co-attendance to a conference

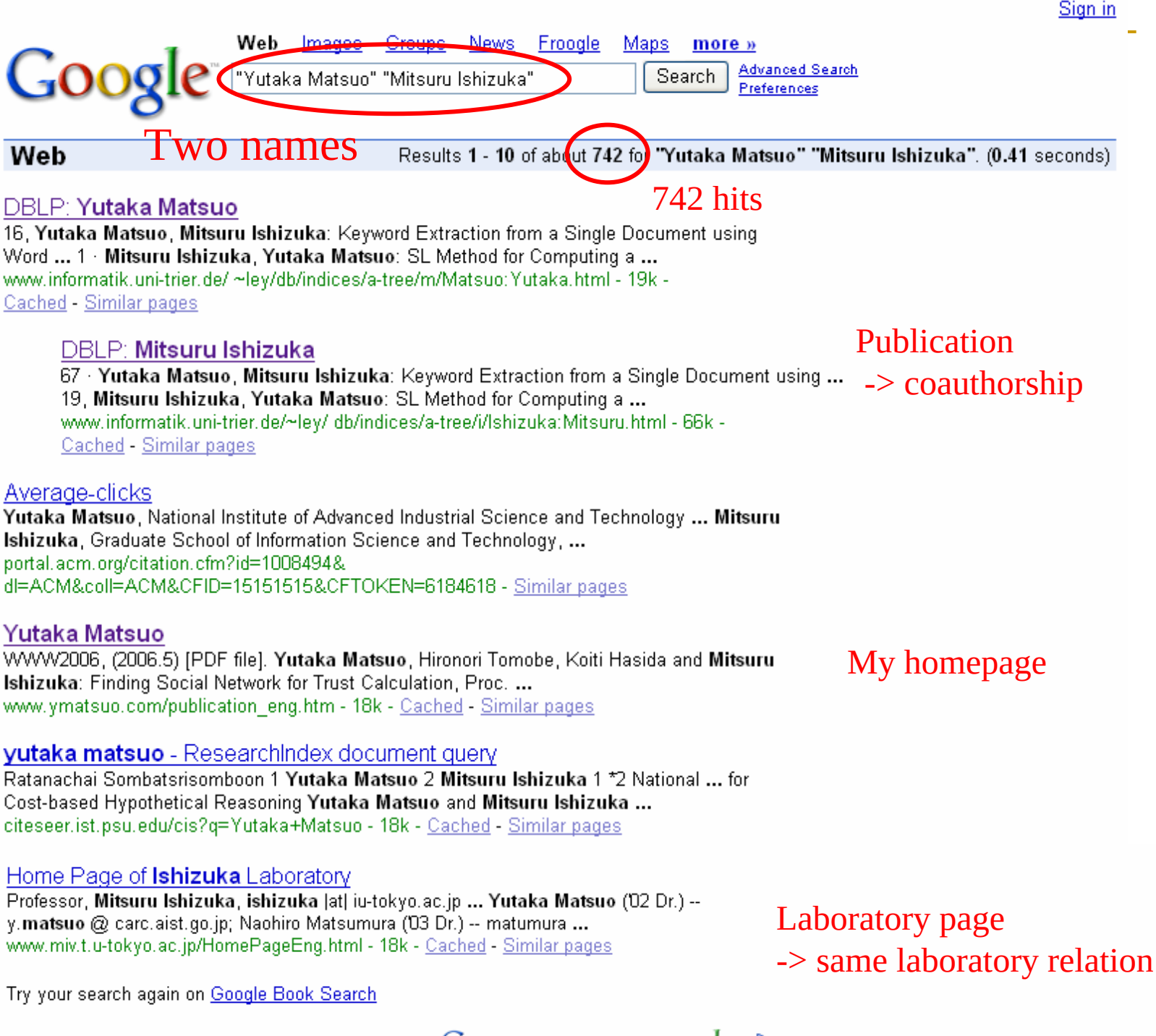
How to make the network?

- Node
  - Conference participants / members of a society
    - We need a list of names and affiliations.

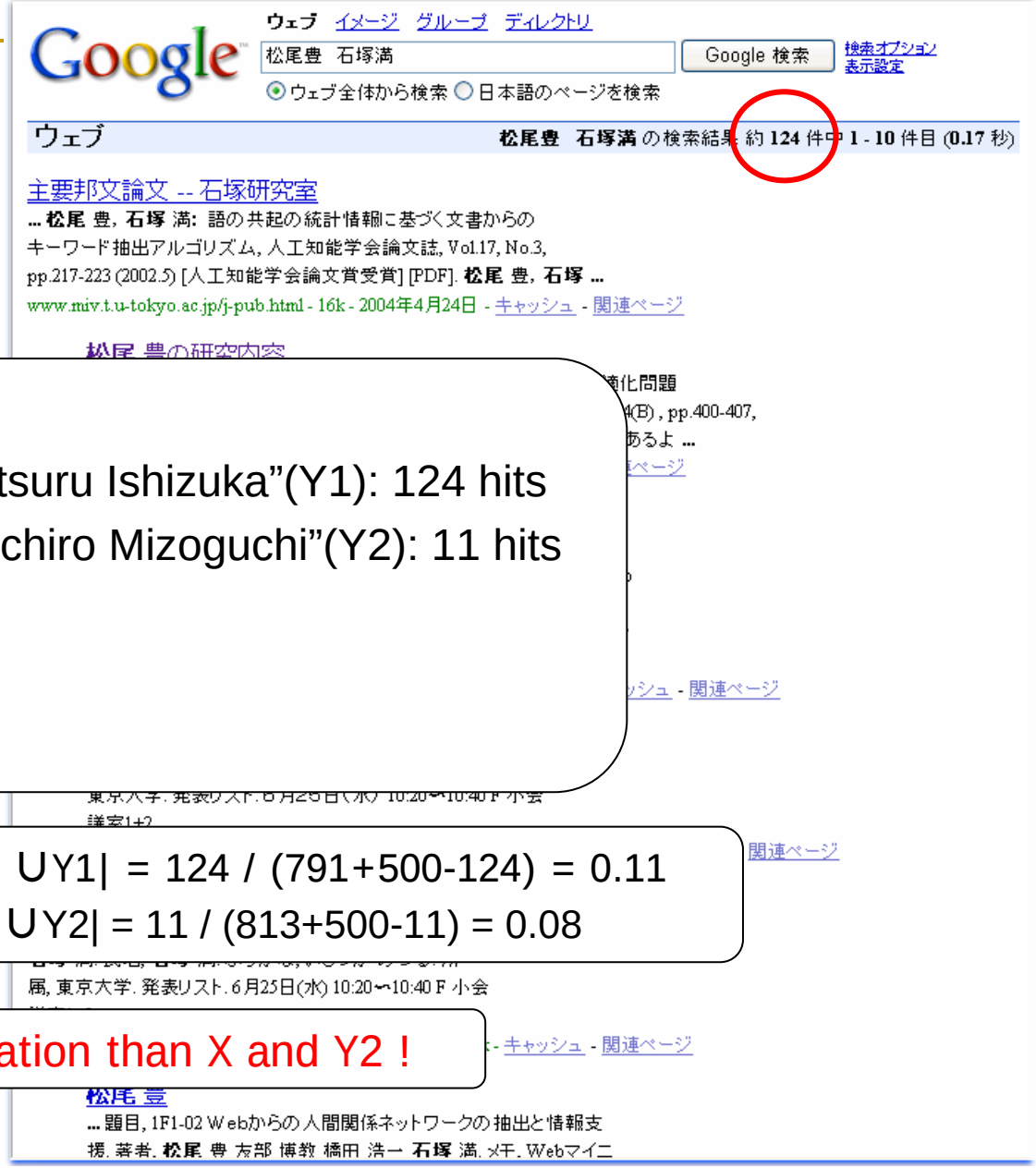
Given!

Automatically obtained!

- Edge
  - Obtained by Web information.
  - Using a search engine, *the co-occurrence of two persons' name* on the Web is measured.
  - Relation types(Edge labels)
    - Coauthor, Lab, Proj, Conf
  - Also, keywords of each researcher, clusters of reserachers


Sign in
Google
Web Images Groups News Froogle Maps more »
"Yutaka Matsuo" "Mitsuru Ishizuka"
Search
Advanced Search
Preferences
Two names
Web
Results 1 - 10 of about 742 for "Yutaka Matsuo" "Mitsuru Ishizuka". (0.41 seconds)
742 hits
DBLP: Yutaka Matsuo
16, Yutaka Matsuo, Mitsuru Ishizuka: Keyword Extraction from a Single Document using Word ... 1 · Mitsuru Ishizuka, Yutaka Matsuo: SL Method for Computing a ...
www.informatik.uni-trier.de/ ~ley/db/indices/a-tree/m/Matsuo:Yutaka.html - 19k - Cached - Similar pages
DBLP: Mitsuru Ishizuka
67 · Yutaka Matsuo, Mitsuru Ishizuka: Keyword Extraction from a Single Document using ... 19, Mitsuru Ishizuka, Yutaka Matsuo: SL Method for Computing a ...
www.informatik.uni-trier.de/~ley/ db/indices/a-tree/i/Ishizuka:Mitsuru.html - 66k - Cached - Similar pages
Publication
-> coauthorship
Average-clicks
Yutaka Matsuo, National Institute of Advanced Industrial Science and Technology ... Mitsuru Ishizuka, Graduate School of Information Science and Technology, ...
portal.acm.org/citation.cfm?id=1008494&dl=ACM&coll=ACM&CFID=15151515&CFTOKEN=6184618 - Similar pages
Yutaka Matsuo
WWW2006, (2006.5) [PDF file]. Yutaka Matsuo, Hironori Tomobe, Koiti Hasida and Mitsuru Ishizuka: Finding Social Network for Trust Calculation, Proc. ...
www.ymatsuo.com/publication_eng.htm - 18k - Cached - Similar pages
My homepage
yutaka matsuo - ResearchIndex document query
Ratanachai Sombatsrisomboon 1 Yutaka Matsuo 2 Mitsuru Ishizuka 1 *2 National ... for Cost-based Hypothetical Reasoning Yutaka Matsuo and Mitsuru Ishizuka ...
citeseer.ist.psu.edu/cis?q=Yutaka+Matsuo - 18k - Cached - Similar pages
Home Page of Ishizuka Laboratory
Professor, Mitsuru Ishizuka, ishizuka |at| iu-tokyo.ac.jp ... Yutaka Matsuo ('02 Dr.) -- y.matsuo @ carc.aist.go.jp; Naohiro Matsumura ('03 Dr.) -- matumura ...
www.miv.t.u-tokyo.ac.jp/HomePageEng.html - 18k - Cached - Similar pages
Laboratory page
-> same laboratory relation
Yutaka M
Try your search again on Google Book Search

Co-occurrence



In JP domain,

- “Yutaka Matsuo”(X) AND “Mitsuru Ishizuka”(Y1): 124 hits
- “Yutaka Matsuo”(X) AND “Riichiro Mizoguchi”(Y2): 11 hits
  - Y1: 791 hits
  - Y2: 813 hits
  - X: 500 hits

- Jaccard coefficient $|X \cap Y1| / |X \cup Y1| = 124 / (791+500-124) = 0.11$
- Jaccard coefficient $|X \cap Y2| / |X \cup Y2| = 11 / (813+500-11) = 0.08$

X and Y1 is a stronger relation than X and Y2 !

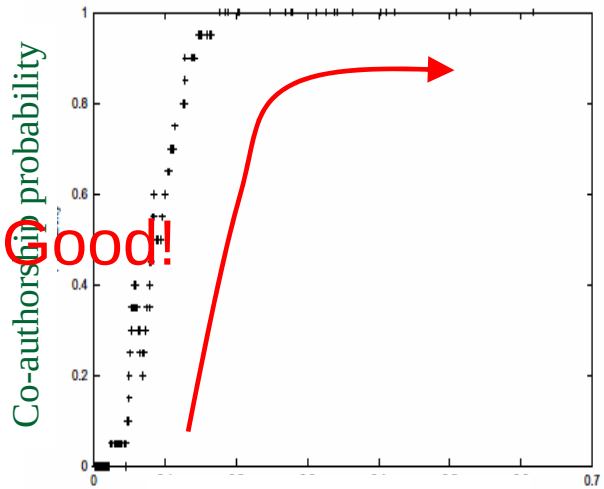
Measures of co-occurrence

- Matching coefficient $|X \cap Y|$
- Mutual Information $\log N|X \cap Y| / |X||Y|$
- Dice coefficient $2|X \cap Y| / (|X|+|Y|)$
- Jaccard coefficient $|X \cap Y| / |X \cup Y|$
- Cosine $|X \cap Y| / (\sqrt{|X||Y|})$
- Simpson coefficient (overlap coefficient) $|X \cap Y| / \min(|X|,|Y|)$

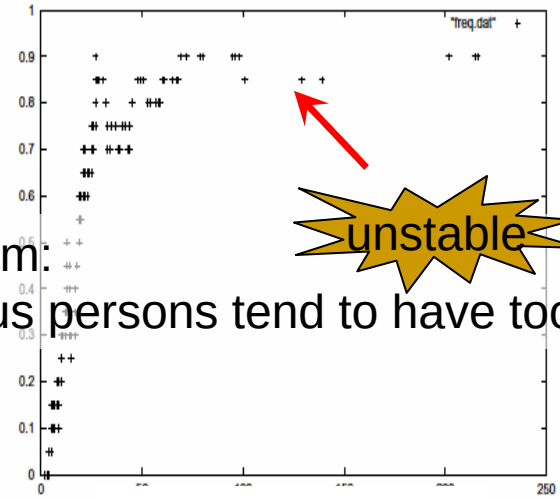
- with a cutoff threshold on |X| and |Y|.

$$f(X,Y) = \begin{cases} |X \cap Y| / \min(|X|,|Y|) & \text{if } |X| > k \text{ and } |Y| > k \\ 0 & \text{otherwise} \end{cases}$$

Which co-occurrence measure is good?
Co-oc value vs Coauthor probability

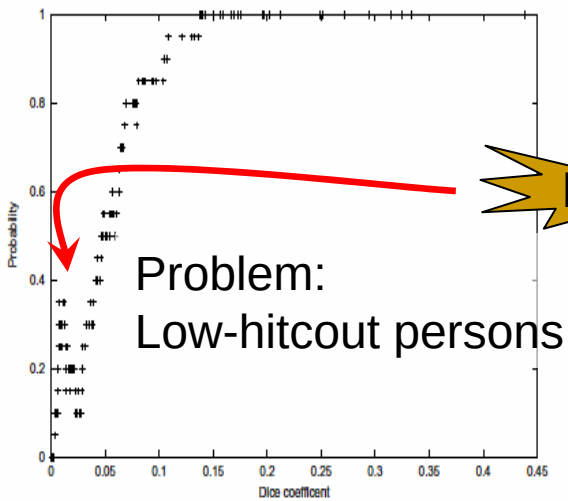


Good!

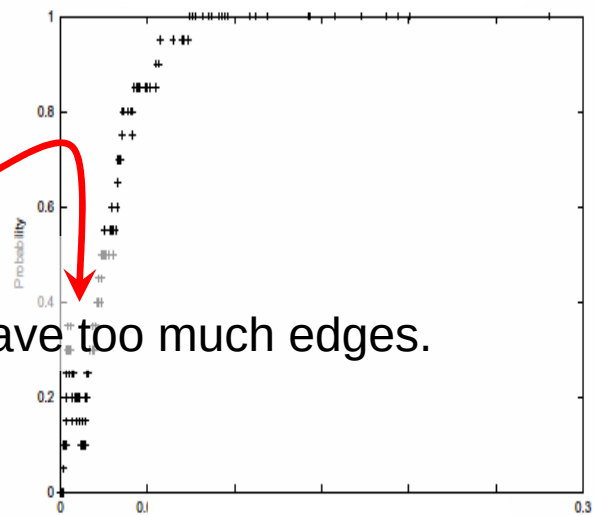


unstable

Problem:
Famous persons tend to have too much edges



bump



Problem:
Low-hitcout persons tend to have too much edges.

Yutaka

Fig. 5. Probabi coefficient

Related works on social network mining

- Referral Web (H. Kautz et al, 1997)
  - A name is given as input
  - Retrieve the name, and extract other names.
  - Measure co-occurrence (by Jaccard coefficient), and invent edges.
  - Ego-centric network within 2-3 radius
    - E.g, find a path from Henry Kautz to Marvin Minsky
- Flink (P. Mika, 2004)
  - Email messages, publications, FOAF documents, and Web mining
  - Web mining part
    - Similar to Referral Web
    - Jaccard coefficient
    - “Semantic Web OR Ontology” is added to a query for disambiguation.
- A. McCallum et al. (2004-)
  - Identify people in e-mail messages, and find homepages
    - Links are placed between the owner of Web page and persons discovered on the page.
  - They also use co-occurrence on the entire Web
    - with name-disambiguation probability model
- Other studies using co-occurrence information
  - [Harada04] [Faloutsos04] [Kees04]…

Small pseudocodes

- Among related studies, underlying procedures are similar and very simple
- To enhance versatility, we describe in small pseudocodes.
- Two basic functions:
  - *GoogleHit*: returns hit count
  - *GoogleTop*: returns (*k*) retrieved documents

Algorithm 3.1: GOOGLECOOC(X, Y)

comment: Given person names X and Y, return the co-occurrence.

$n_X \leftarrow GoogleHit(\text{“}X\text{”})$
$n_Y \leftarrow GoogleHit(\text{“}Y\text{”})$
$n_{X \wedge Y} \leftarrow GoogleHit(\text{“}X\ Y\text{”})$
$r_{X,Y} \leftarrow CoocFunction(n_X, n_Y, n_{x \wedge y})$
return $(r_{X,Y})$

Algorithm 3.1: G

comment: Given p

$n_X \leftarrow GoogleH$
$n_Y \leftarrow GoogleH$
$n_{X \wedge Y} \leftarrow Googl$
$r_{X,Y} \leftarrow CoocF$
return $(r_{X,Y})$

Algorithm 3.2: GOOGLECOOCTOP(X, Y, k)

comment: Given person nar

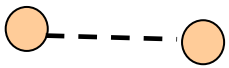
$D_X \leftarrow GoogleTop(\text{“}X$
$D_Y \leftarrow GoogleTop(\text{“}Y\text{”}$
$n_X \leftarrow NumEntity(D_)$
$n_Y \leftarrow NumEntity(D_Y$
$n_{X \wedge Y} \leftarrow NumCooc(D$
$r_{X,Y} \leftarrow CoocFunction$
return $(r_{X,Y})$

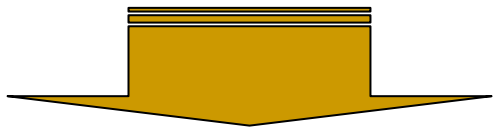
Algorithm 3.5: GOOGLECOOCCONTEXT(X, Y, W_X, W_Y)

comment: Given X, Y and word(s) W_X, W_Y, return co-occurrence.

$n_X \leftarrow GoogleHit(\text{“}X\ W_X\text{”})$
$n_Y \leftarrow GoogleHit(\text{“}Y\ W_Y\text{”})$
$n_{X \wedge Y} \leftarrow GoogleHit(\text{“}X\ Y\ W_X\ W_Y\text{”})$
$r_{X,Y} \leftarrow CoocFunction(n_X, n_Y, n_{X \wedge Y})$
return $(r_{X,Y})$

Name disambiguation

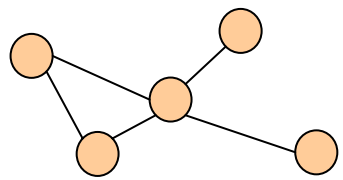




Algorithm 3.3: GETSOCIALNET(L)

comment: Given person list L, return a social network G.

for each $X \in L$
  do set a node in G
for each $X \in L$ *and* $Y \in L$
  do $r_{X,Y} \leftarrow GoogleCooc(X, Y)$
for each $X \in L$ *and* $Y \in L$ where $r_{X,Y} > thre$
  do set an edge in G
return (G)



Algorithm 3.4: EXPANDPERSON(X, k)

comment: Extract person names from the retrieved pages.

$D \leftarrow GoogleTop(\text{“}X'', k)$
$E \leftarrow ExtractEntities(D)$
return (E)

Node expansion

Content

- Social network mining from the Web
  - Algorithms and evaluations
  - Advanced algorithms
    - A1: Types of relation
    - A2: Scalability
    - A3: Person-to-word co-occurrence
- POLYPHONET:
  - Academic conference support system
- Future directions

A1: Types of relations

- It is more useful if we recognize types of relations.
- In JSAI case, we define four types (classes) of relations:
  - Coauthor: Co-author of a technical paper
  - Lab: Members of the same laboratory or research institute. It is not always a coauthor-relation, but it is also a strong relationship.
  - Proj: Members of the same project or committee
  - Cof: Participants of the same conference or workshop
- Multi-labels for an edge.

Algorithm 4.1: CLASSIFYRELATION(X, Y, k)

comment: Given person names X and Y, return the class of relation.

$D_{X \wedge Y} \leftarrow GoogleTop(\text{“}X\ Y\text{”}, k)$
for each $d \in D_{X \wedge Y}$
  do $c_d \leftarrow Classifier(d, X, Y)$
$class \leftarrow determine\ on\ c_d \in D_{X \wedge Y}$
return $(class)$

Text classification problem!!

Evaluation

- In JSAI case,

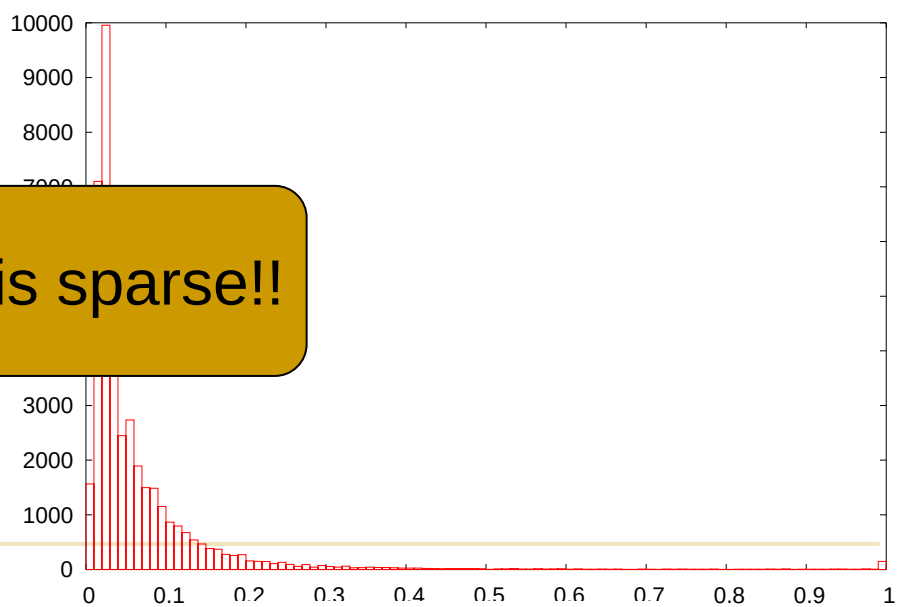
class	error rate	precision	recall
Coauthor	4.1%	91.8% (90/98)	97.8% (90/92)
Lab	25.7%	70.9% (73/103)	86.9% (73/84)
Proj	5.8%	74.4% (67/90)	91.8% (67/73)
Conf	11.2%	89.7% (87/97)	67.4% (87/129)

  - For JSAI2003 participants, 275 training samples and 200 evaluation samples are used. We use C4.5 as a learner.

A2: Scalability

- Too many queries to a search engine
  - Assume we have *n* names. Then, ${}_nC_2$, or $O(n^2)$ queries are necessary.
  - For 500 people, 124,750 queries...
    - cf) Google API (1000 queries/day), Yahoo! API (5000/day)
- Distribution of Simpson coefficient
  - 0: about 67%
  - 0-0.2: about 98%

A social network is sparse!!



Idea for scalability

- Filter out pairs of persons that seem to have no relation.
- Apply *GoogleCooc* only for promising pairs investigating by *GoogleTop*

Algorithm 4.2: GETSOCIALNETSCALABLE(L, k)

comment: Given person list L, return a social network G.

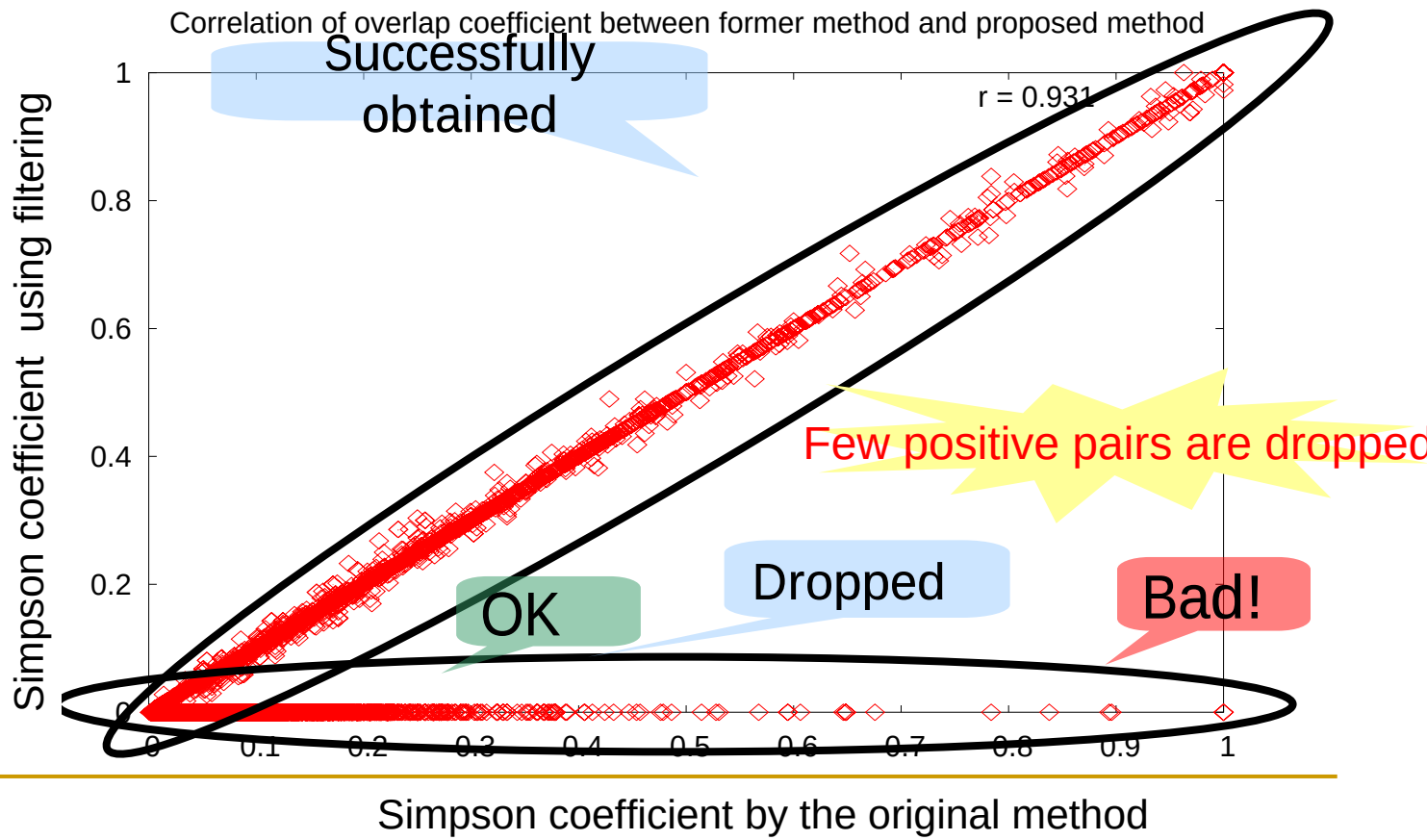
for each $X \in L$
  do set a node in G
for each $X \in L$
  do $D \leftarrow GoogleTop(\text{“}X\text{”}, k)$
  $E \leftarrow ExtractEntities(D)$
  for each $Y \in L \cap E$
    do $r_{X,Y} \leftarrow GoogleCooc(X, Y)$
for each $X \in L$ *and* $Y \in L$ where $r_{X,Y} > threshold$
  do set an edge in G
return (G)

Combination of *GoogleCooc* and *GoogleTop*

Evaluation for JSAI case

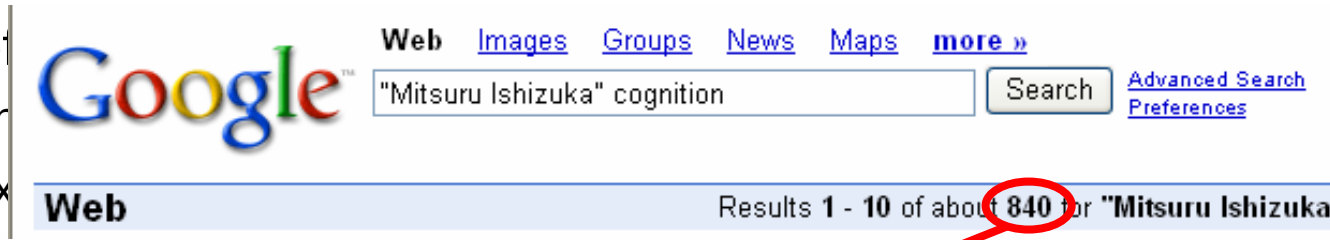
85% of queries are saved.

- For 504 researchers, originally 126,253 queries: $O(n^2)$
- By applying filtering, 19,182 queries: $O(n)$



A3: Word-to-person Co-occurrence

- Co-occurrence of
  - Not a person ar
  - Affiliation matrix



	agent	mining	communication	audio	cognition	...
Mitsuru Ishizuka	454	143	414	382	840	...
Koiti Hasida	412	156	1020	458	1150	...
Yutaka Matsuo	129	112	138	89	58	...
Nobuaki Minematsu	227	22	265	648	138	...
Yohei Asada	6	6	6	2	0	...
...	...	...	...	...	...	...

Affiliation matrix

Adjacent matrix

- Keyword extraction for a person
- Clustering of persons

Clustering of persons

- Context similarity of persons
  - using an affiliation matrix

Algorithm 4.4: CONTEXTSIM(X, Y, W_L)

comment: Given names X, Y and word list W_L, return the similarity.

for each $W \in W_L$

do $\begin{cases} a_W \leftarrow GoogleCooc(X, W) \\ b_W \leftarrow GoogleCooc(Y, W) \end{cases}$

$s_{X,Y} \leftarrow$ similarity of two vectors $a = \{a_W\}$ and $b = \{b_W\}$

return $(s_{X,Y})$

- In JSAI case, We cluster 540 researchers into 30 groups
  - using cosine similarity, and
  - complete-link clustering.

Result of extracted researcher clusters

Cluster1：神嶌敏弘, 角田祐一, 新村昭好, 岡崎直観, 倉田岳人, 村田剛志, 相原健郎, 豊田正史, 久保山哲二, 藤村滋, 仲尾由雄, 藤澤瑞樹, 橋本泰一, 池原悟, 徳永健伸, 荒木健治, 奥村学, 鈴木雅実, 熊本忠彦, 太田公子, 徳久雅人, 村上仁一, 賀沢秀人, 岩垣守彦, 川野洋, 藪内佳孝, 峯松信明, 嵯峨山茂樹, 西本卓也, 中沢正幸, 新田恒雄, 桐山伸也, 駒谷和範, 伊藤敏彦, 桂田浩一, 天野成昭, 中川聖一, 古塩貴行, 池ヶ谷有希, 鈴木夕紀子, 野口靖浩, 松本泰明, 小玉智志

NLP

Weighted words: 自然言語(NLP), 音声(audio), 対話(dialogue), 言語(language), 抽出(extraction), 処理(processing), クラスタリング(clustering), 生成(generation), 検索システム(retrieval system), 情報抽出

Cluster 2: 中西英之, 角康之, 間瀬健二, 西本一志, 藤城卓己, 田森裕邦, 安念克洋, 天野竜太, 櫻井晴章, 岡夏樹, 今井倫太, 田島敬士, 畠山誠, 大村英史, 劉耿盈, 正要将文, 楠本昌弘, 大谷尚史, 高畠政実, 高橋昌史, 小出義和, 岩澤昭一郎, 伊藤禎宣, 竹林洋一, 堀聡, 広瀬公太, 小泉智史, 矢入健久, 木村春彦, 青島大悟, 雨宮陽介

Ubiquitous computing and robotics

Weighted words: インタラクション(interaction)、センサ(sensor)、ヒューマン(human)、ユビキタス(ubiquitous)、エージェント(agent)、コミュニケーション(communication)、ロボット(robot)、支援(support)、実装(implementation)

Cluster 10: 小出誠二, 官上大輔, 武田英明, 和泉憲明, 岩爪道昭, 小路悠介, 垂見晋也, 來村徳信, 古崎晃司, 溝口理一郎, 渡邉英一, 池田満, 酒井隆道, 西原陽子, 森田武史, 見置孝昌, 繁田佳宏, 田中庸平, 武内雅宇

Semantic Web

Weighted words: オントロジー(Ontology), 支援環境(support environment), 知能(intelligence), 支援システム(support system), 学習支援(learning), 構築(development), エージェント, 設計, 支援, 人工

Keyword extraction

- Using the person-to-word matrix, we can extract keywords for each person.

- Example of keywords for D. Brickley

Dan Brickley	Dan Connolly
Libby Miller	Jan Grant
FOAF	RDF Interest Group
Semantic Web	xmlns.com=foaf
Dave Beckett	RDF
RDFWeb	Eric Miller
ILRT	FOAF Explorer

Algorithm 4.3: EXTRACTKEYWORDS(X, k_1, k_2)

$D \leftarrow GoogleTop(X, k_1)$
$words \leftarrow ExtractWords(D)$
for each $W \in words$
  do $score_W \leftarrow GoogleCooc(X, W)$
$K \leftarrow \{W | score_W \ is \ top \ k_2\}$
return (K)

- Comparison to standard document-to-word matrix (on retrieved documents): *tf* and *tfidf*

Method	tf	tf-idf	*ExtractKeywords*
precision	0.13	0.18	0.60
recall	0.20	0.24	0.48

For 567 JSAI researchers. For tf and tfidf, we collect 10 documents for each person. Termex is used to recognize to recognize terms.

Content

- Social network mining from the Web
  - Algorithms and evaluations
  - Advanced algorithms
    - Types of relation
    - Scalability
    - Person-to-word co-occurrence
- POLYPHONET:
  - Academic conference support system
- Future directions

Polyphonet

- Web-based system for an academic community
  - to facilitate communication and understanding of each other
  - based on a social network extracted from the Web
- Operation
  - JSAI (the Japan Society of Artificial Intelligence) Annual Conferences (JSAI03-05), and International Conference on Ubiquitous Computing (UbiComp05)

	JSAI03	JSAI04	JSAI05	UbiComp05
#participants	558	639	about 600	about 500
#users	276	257	217	308

JSAI2003

UbiComp05

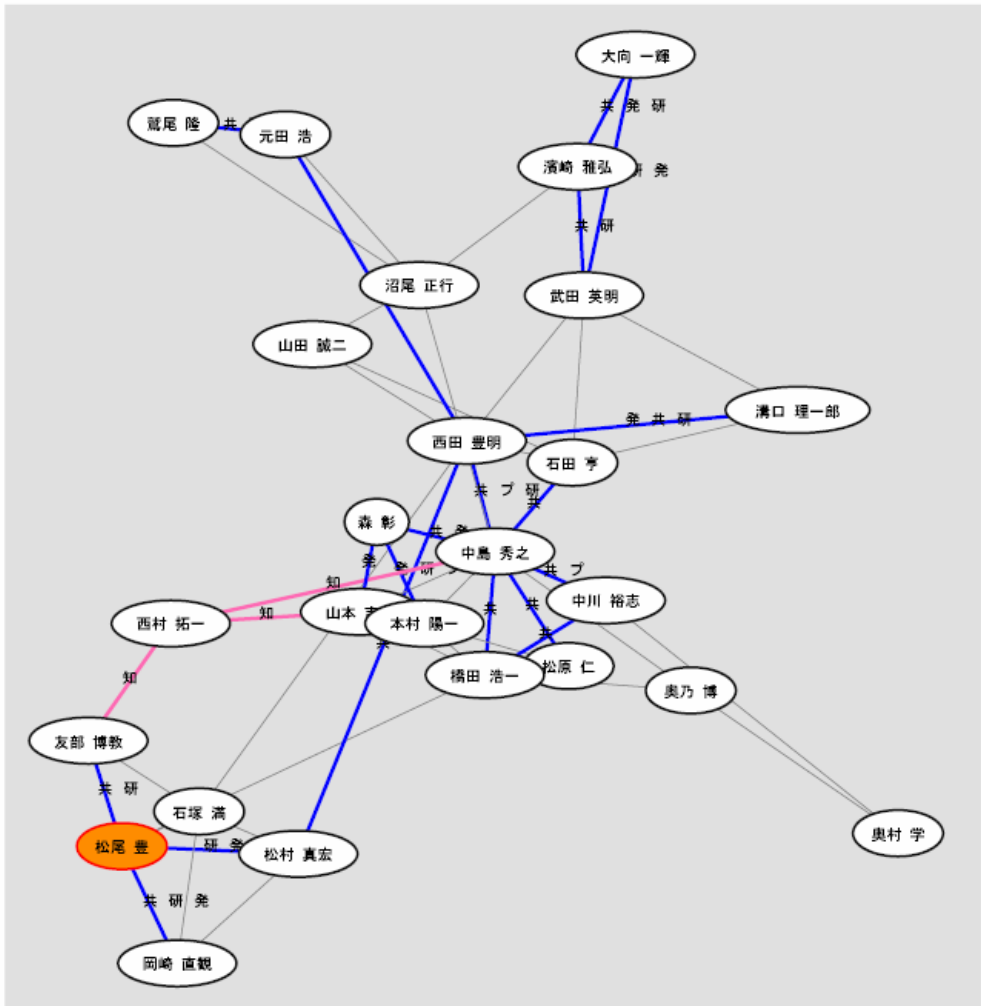
Polyphonet JSAI2005: Person view (Mypage)

POLYPHONET NETWORK VIEWER - 松尾 豊 さんの周辺の人間関係ネットワーク -

ネットワーク規模(ノード数) / 10 / 25 / 50

人間関係ネットワークのヘルプ

[松尾 豊 さん のページに戻る]



研究者

松尾 豊
産業技術総合研究所

操作方法

拡大: [Ctrl]+クリック
縮小: [Ctrl]+[Shift]+クリック
全体をずらす: [Alt]+ドラッグ
ノードを動かす: ノードをドラッグ

図の説明

名前の付いた楕円が研究者を表し、関係のある他の研究者と線でつながっています。関係が強ければ強いほど線は短くなります。青い線には関係ラベルが付随しており、種類は以下のとおりです。

共 ... 共著関係
研 ... 同研究室関係
プ ... プロジェクト関係
発 ... 同発表関係
知 ... 知り合い関係

インターネット

Presentation view

Title and abstract

Add to my schedule

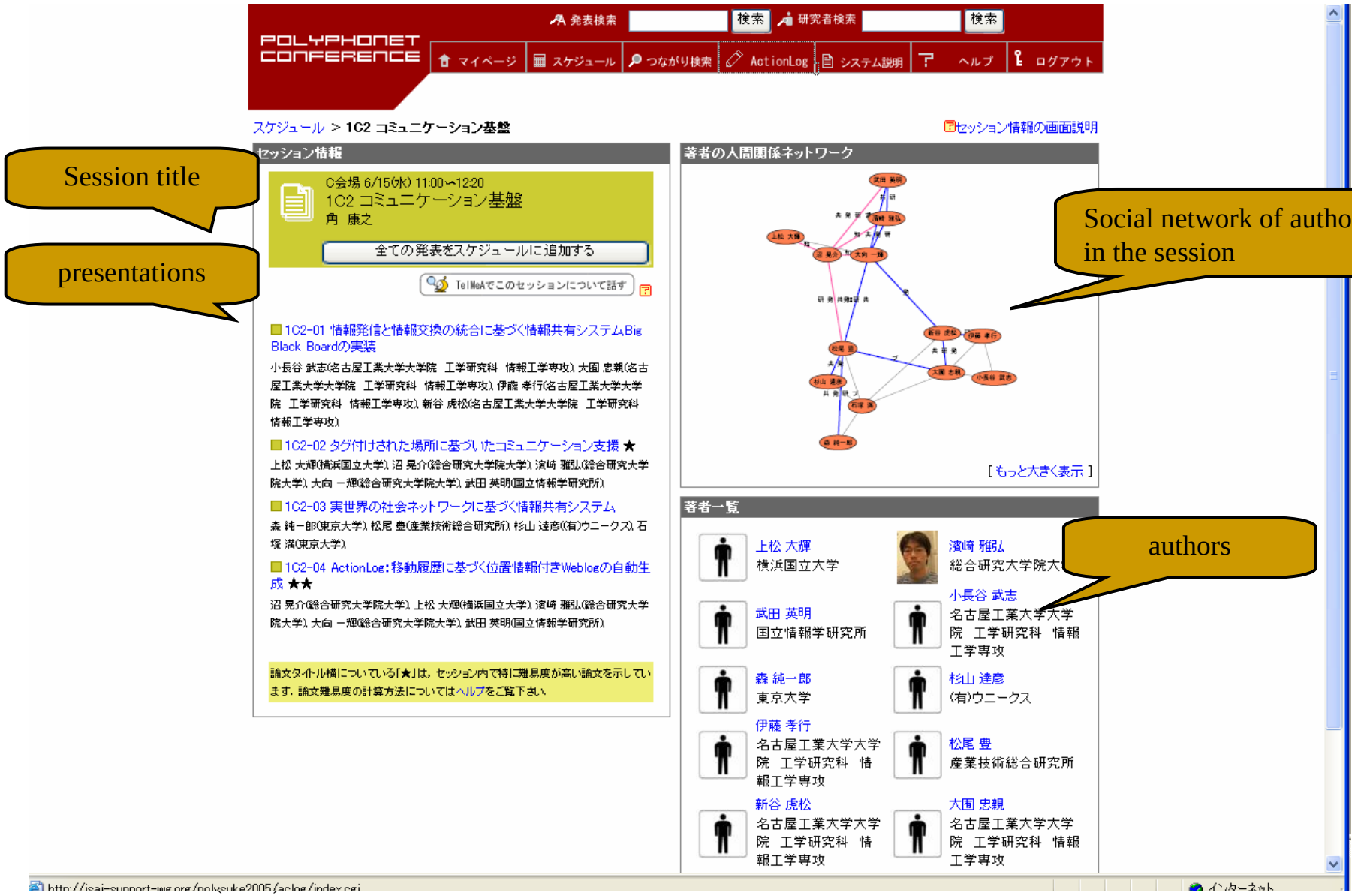
authors

Social network among authors and persons who checked the presentation

Persons who checked this presentation

other presentations by the author

Session view



Researcher Information | Bookmark | ActionLog | Ironomy

Help for This Page

Yutaka Matsuo

AIST

- Related people
- Edit your profile

Status of link with Information Clip

(unavailable) Link Code: [ ] Link make it available

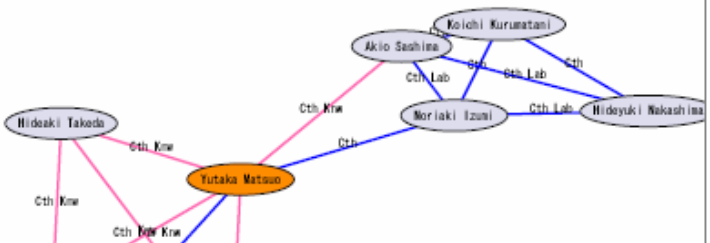
Acquaintances

Know each other (2) / Who you know (5) / Who knows you (9)
[Show All]

Yutaka Matsuo's Presentation

Invited Demo : ID03
Ubiquitous Community Support System for UbiComp2005

Neighbor Human Network



Related People

Coauthor Laboratory Project Presentation Know

Takuichi Nishimura
AIST

Masahiro Hamasaki
AIST

Helmut Prendinger
NII

Akio Sashima
AIST

Hideaki Takeda
National Institute of Informatics

Yasuyuki Sumi
Kyoto University

Tim Kindberg
HP Labs

Noriaki Izumi
AIST

Koichi Kurumatani
AIST

Hideyuki Nakashima
Future University - Hakodate

[Show All]

Yuta

ubicomp '05

Ubiquitous Community Support System

Presentation search | search | Researcher search | search

My page | Schedule | Connection search | Community | Help | Logout

Researcher Information | Bookmark | ActionLog | Ironomy

Help for This Pag

Personal Information

Yutaka Matsuo

Cool People for You

Junichiro Mori bookmark this paper...

R06: Self-Mapping in 802.11 Location Systems

Next >>>

Hot Paper in Your Acquaintance

K01: Opening Keynote

Next >>>

Bookmark

R02: u-Texture: Self-Organizable Universal Panels for Creating Smart Surroundings

R03: Fast and Robust Interface Generation for Ubiquitous Applications

R05: Preventing Camera Recording by Designing a Capture-Resistant Environment

R13: Control, Deception, and Communication: Evaluating the Deployment of a Location-Enhanced Messaging Service

R17: From Interaction to Participation: Configuring Space Through Embodied Interaction

Demonstration (9/12 16:40-18:30)

ID03: Ubiquitous Community Support System for UbiComp2005

Poster (9/12 16:40-18:30)

Yutaka

ubicomp'05
Ubiquitous Community Support System

Presentation search | search | Researcher search | search

My page | Schedule | Connection search | Community | Help | Logout

Help for This Page

Origin Node

Yutaka Matsuo
AIST
[Cancel]

Target Node

Hideyuki Tokuda
Keio University
[Cancel]

Connection

Coauthor Laboratory Project Presentation Know

1 / 2 / Next

Yutaka Matsuo
AIST

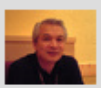
0.81 km ↑ ↓ Know [Detail]

Takuichi Nishimura
AIST
[Exclude this person]

1.06 km ↓ Know [Detail]

Hideyuki Nakashima
Future University - Hakodate
[Exclude this person]

0.65 km ↓ Know [Detail]



Hideyuki Tokuda
Keio University

1 / 2 / Next

Exclusion List

nobody

In Ubiquitous Environment…

- An IC card is distributed as a namecard in JSAI2005 and UbiComp05
  - A participant can also use his/her own IC card.
- If two users put their cards…

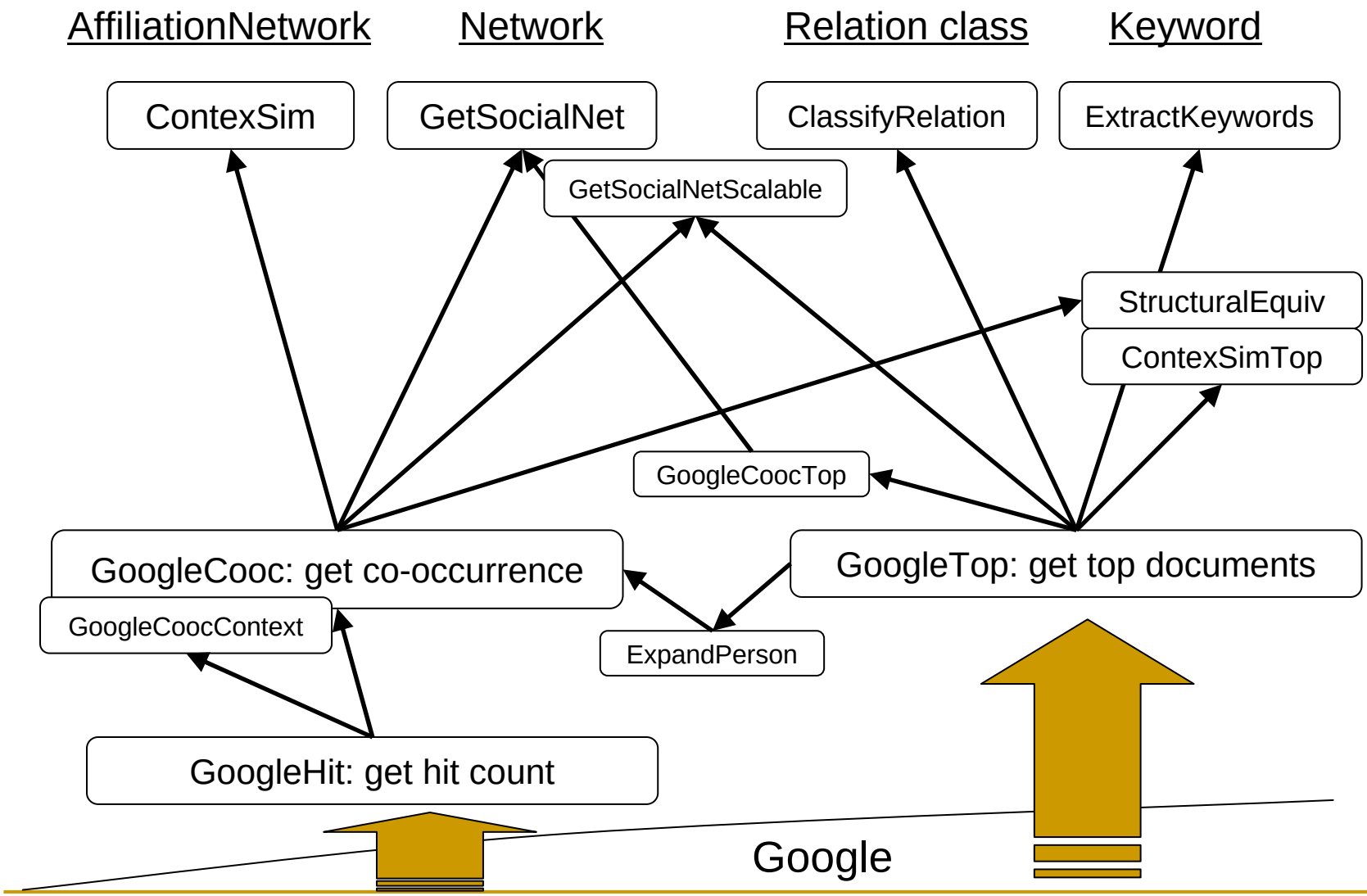
- Then, the two will be connected by a *face-to-face link*, resulting in a more connected social network during the conference.

Future directions

- Mine structures of entities using a search engine
  - Cross-field among AI, Web and social network analysis

- Scalability
  - Simple modules to be built up
  - A goog combination of *GoogleCooc* and *GoogleTop* is necessary.
  - As the Web grows…,
    - *GoogleTop* (which retrieves document and enables text analysis) gets less effective: it returns relatively smaller and smaller sizes of Web pages.
      - Sampling [Anagnostopoulos05,Bar-Yossef06] is important.
    - *GoogleCooc* yields a more precise number because the low-frequency problem is improved.
      - Search engine for NLP [Cafarella05]

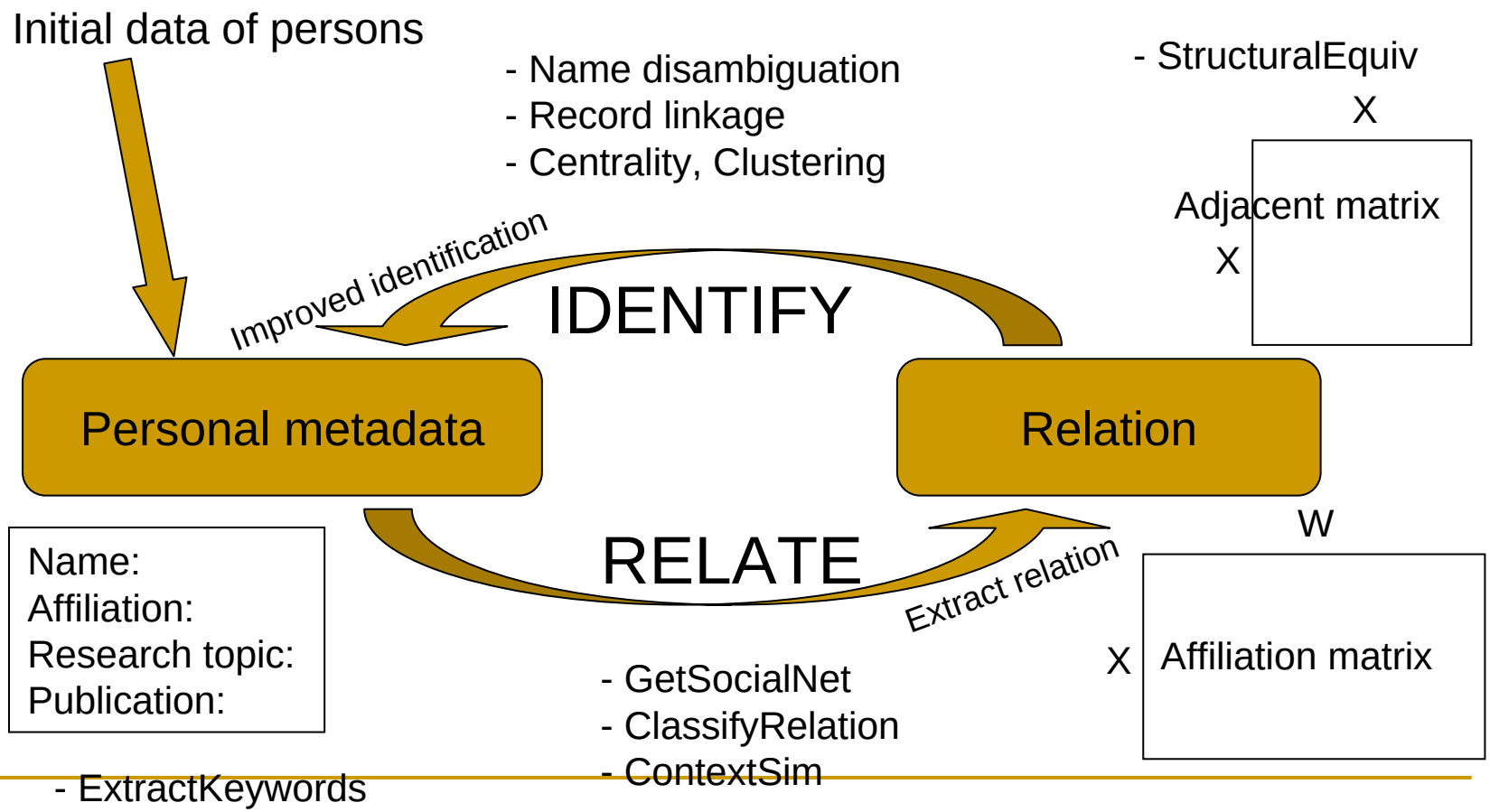
Module dependencies



A *GoogleHit* and *GoogleTop* provides a variety of modules though they are very simple

Relate-Identify process

We identify entities, extract relations of entities, and based on the whole network structure and statistics, we improve the way to identify entities. The process is repeated iteratively, and is gradually improved.



For integration of multiple social networks, see
Y.matsuo et. al, "Spinning social networks for Semantic Web" to appear in AAAI-06

Conclusion

- Social network extraction from the Web
  - Small modules using a search engine are built up for a variety of functions.
  - Algorithms for types of relations, scalability, a person-to-word matrix
- Polyphonet
  - JSAI conference series(2003-06), and UbiComp05
  - Many positive comments from participants
- Extraction of structural information among entities
  - Scalability
  - Identify-Relate process

Thank you!

- Yutaka Matsuo
  - y.matsuo@aist.go.jp
  - National Institute of Advanced Industrial Science and Technology (AIST)

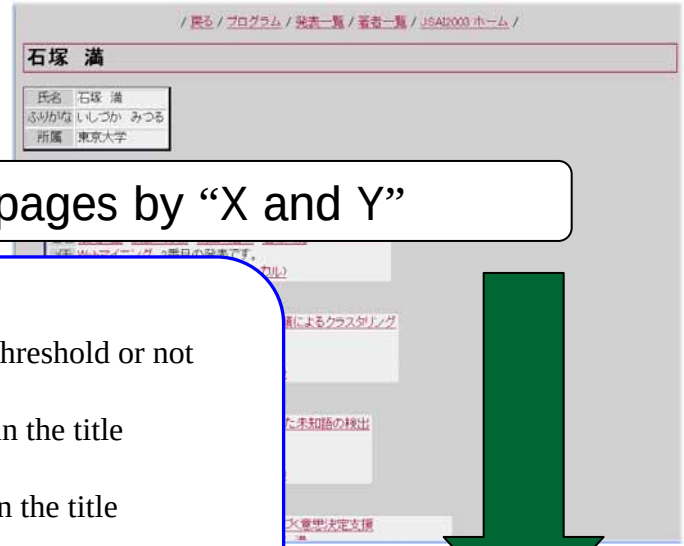
Thanks to
Mr. J. Mori, Dr. M. Hamasaki,
Mr. K. Ishida, Drs. T.Nishimura,
K. Hasida, and Prof. M. Ishizuka,
and other colleagues.

AIST Office at Akihabara

Name disambiguation

- Community-based approach
  - “ubiquitous OR pervasive” for UbiComp community
  - “Semantic Web OR Ontology”
- Individual-based approach
  - We use affiliations for JSAI community
    - Yutaka Matsuo (“University of Tokyo” OR “National institute of Advanced Industrial Science and Technology” OR “AIST”)
- Many related works
  - [Bekkerman05][Li05][Guha][Lloyd05][Marlin05]
  - [Mann03][Wacholder97]
- Procedures
  - Retrieved by a query “X”
  - Cluster the results -> different clusters (hopefully) correspond to different persons
    - Similarity is defined using link structures, text similarities (or snippet similarities)
  - Obtain the characteristic keywords for the person X.

Proceedure



Retrieved pages by “X and Y”

Attributes

・Co-occurrence number of two persons ・Whether Simpson coefficient is over the threshold or not
・Number of X occurrence ・Number of Y occurrence
・Either {publication, paper, presentation, activity, theme, award, author} is included in the title
・Either {member, laboratory, research institute, team} is included in the title
・Either {workshop, conference, seminar, meeting, sponsor, symposium} is included in the title
・・・・
・Either {publication, paper, presentation, activity, theme, award, author} is included in the first five line
・・・

Attributes of a page

(more than one, yes, yes, more than one, more than one, no, no, no, no, no, no, yes, no, no, no, yes, no

Classification rule

NumCo = more_than_one → coauthor
NumCo = more_than_one & GroFFive(F)=no → laboratory
(Rel=yes & GroTitle(E)=no & GroFFIve(C)=no → laboratory
・・・・

Class of relation: coauthor, laboratory, project, conferen