

# POLYPHONET

## Human Network Search Engine

100430021

オウ セイ

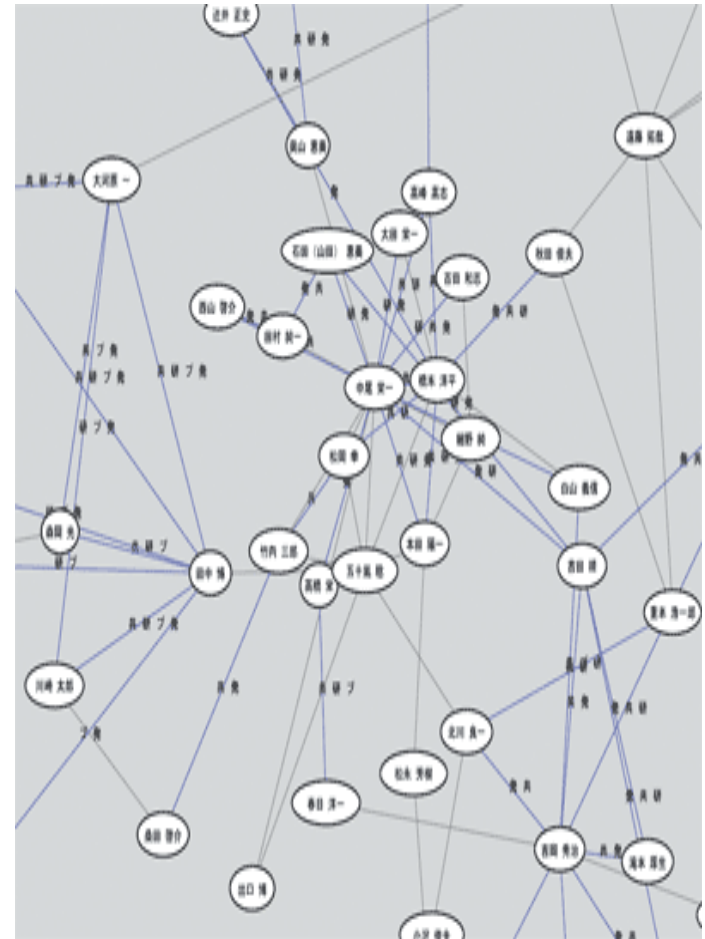
2013/05/03

# 文献

- **タイトル** : POLYPHONET: An advanced Social Network Extraction System from the web
- **作者** : Yutaka Matsuo Junichiro Mori Masahiro Hamasaki
- **発行日** : 2006/05

# 発表内容

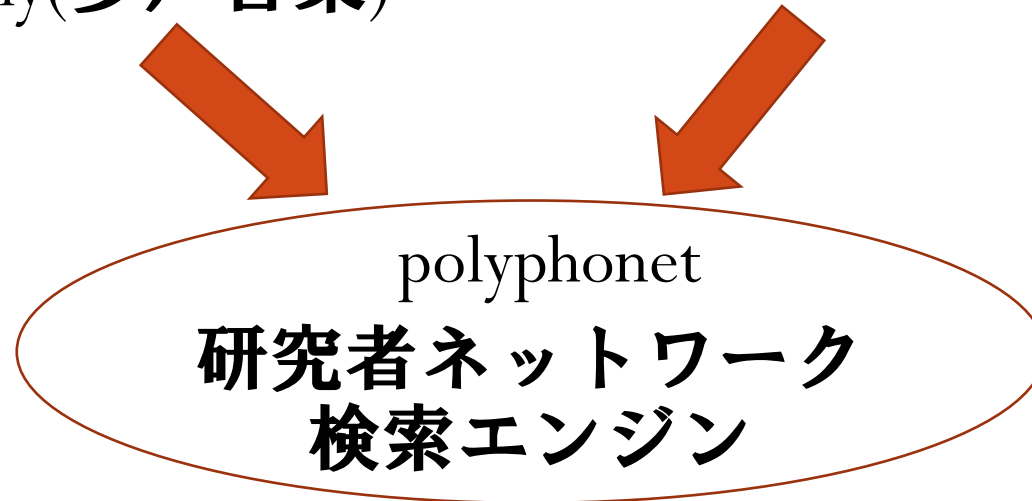
- Polyphonetとは  
定義、背景
- 抽出方法
  - 研究者の関係の強さの計算
  - 研究者の関係の種類抽出
- Polyphonetでできること
- 今後の課題



# Polyphonet(ポリフォネット)

Polyphony(**多声音楽**)

network



研究者及び産業者のダイナミックな変化に対応でき、詳細な  
専門分野を特定した上で共同研究者を探すことが可能な検索  
システム

# 背景

- **研究に関する情報のオンライン化**  
研究者のホームページ、学会プログラム、業績リスト
- **融合領域、産学官連携の重要性**  
ロボット、ナノテク、防災
- **最新研究情報を更新し続けるのは困難**

# 抽出方法

- **Webマイニング**

**情報検索、自然言語処理、機械学習などの技術を組み合わせて、Web上の情報を自動的に加工・処理し、有用な知識を抽出する**

- **処理流れ**

- **ノード**：研究者のリストを与える（名前＋所属）
- **エッジ抽出**：すべての2人間の関係の強さを測定
- **エッジラベル抽出**：関係の種類を測定

# 詳細

## 人工知能学会（JSAI）の研究者の例

### ① ノードを与える

- 研究者のリスト
- 例ではJSAIの著者・共著者をネットワークのノードとする
- 個人情報として氏名と所属を利用

## ② エッジ抽出

検索エンジンを使って共起関係の強さを測る

- アルゴリズム：「X AND Y」と入力するだけ
- Jaccard係数、相互情報量などさまざまな尺度があるが  
閾値つき Overlap係数（もしくはSimpson係数）を用いる

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

「X」、 「Y」 の単独でのヒット件数が閾値  $k$   
以下の場合、エッジが張られない



例)

“松尾豊 石塚満” 1740件 **強い** Google

“松尾豊 溝口理一郎” 184件 **弱い**

“石塚満” 9090件

“溝口理一郎” 9850件



## 同姓同名の問題

➤ 氏名 AND 所属でクエリ

➤ 複数の所属名、過去の所属名がすべてORでつ  
なく

例)、松尾豊 AND (産業技術総合研究所 OR  
産総研 OR 東京大学 OR 東京大 OR 東  
大)

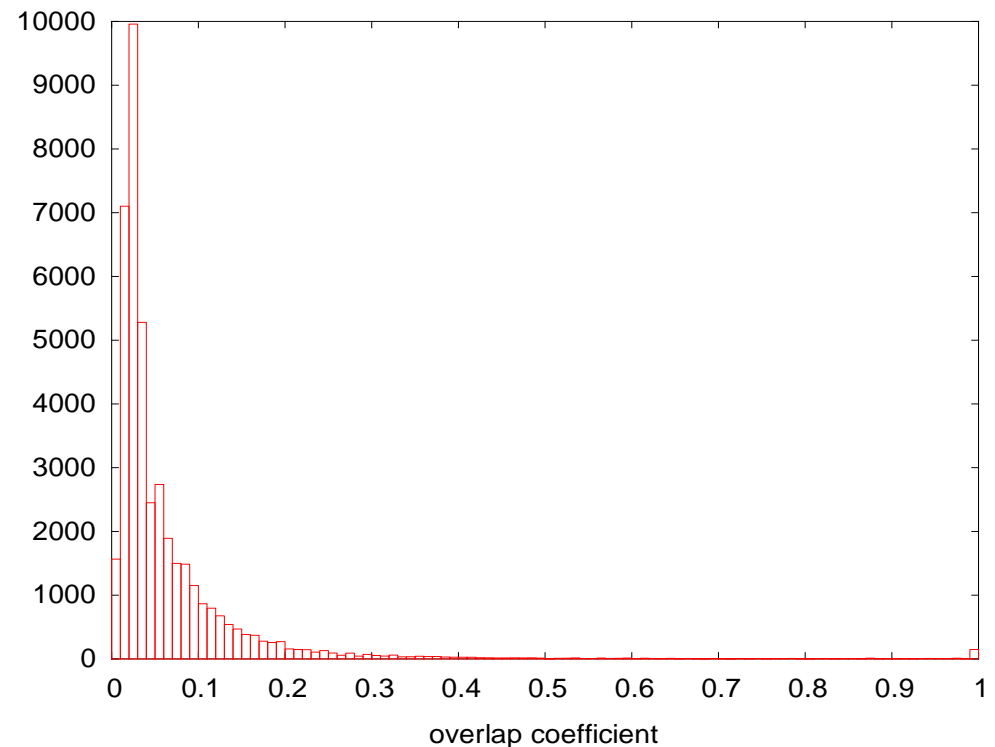
### ③ エッジラベルの抽出

- **関係が強いと判断された2人間の関係の種類を測定する**
- 「X AND Y」クエリ⇒上位5ページ取得⇒属性抽出⇒機械学習で判別
- Polyphonetにおいて定めた関係
  - 共著
  - 同研究室
  - 同プロジェクト
  - 同発表

# 検索エンジンに対する負荷問題

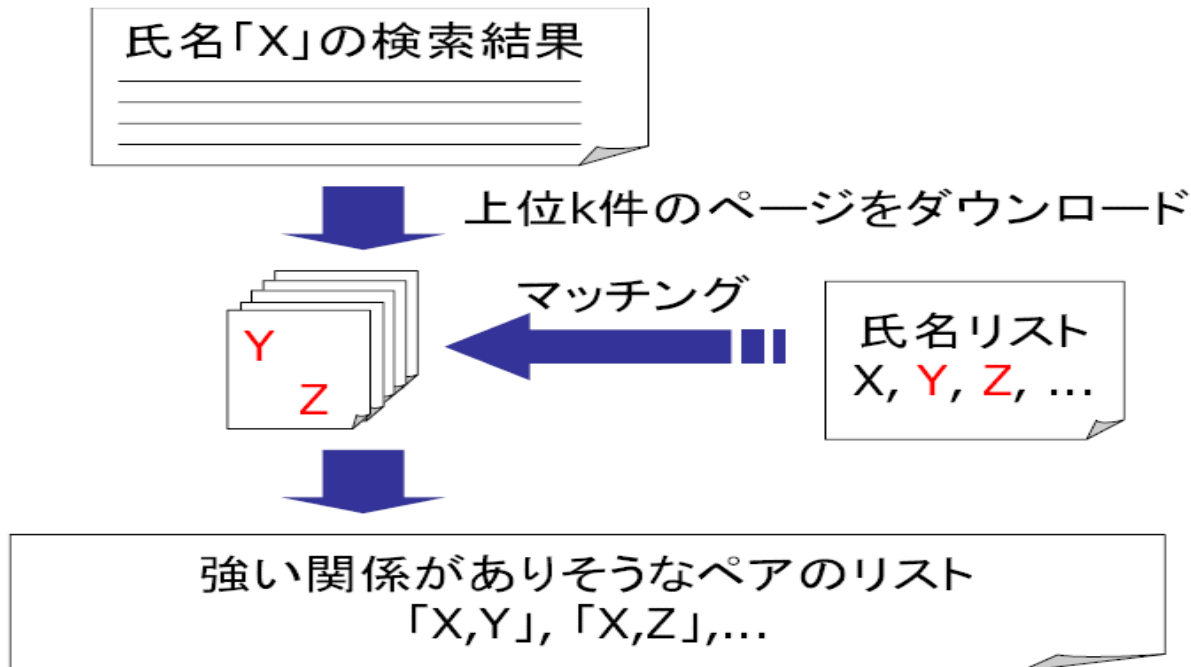
- 氏名リストに $n$ 個の氏名があるとすると、共起ページ数の検索に必要なクエリ数は ${}_n C_2$ 個（ほぼ $O(n^2)$ ）
- Overlap係数分布
  - $0 \rightarrow$  約67%
  - $0 \sim 0.2 \rightarrow$  約98%

ほとんどが弱  
い関係



# 提案

**強い関係がなさそうな氏名のペアを除き、検索エンジンに与えるクエリ数を減らす**



**85%のクエリを削減できる**

# Polyphonetでできること

- **研究者ネットワーク表示**
- **研究者の検索**  
氏名、所属、研究キーワード、研究分野
- **共著・同研究室の研究者の閲覧**
- **分野全体研究者やグループ構成把握**
- **ある研究者から別の研究者へのパスの検索**

# 今後課題

- **Webからの高次情報のマイニング**
  - Blog、掲示板、SNS
  - **検索エンジンの進歩**
- **融合領域での活用**  
**産業との連携のツールとして**
- **日常生活の場面での情報支援**
  - **社会性：有用な情報を提供**
  - **ロボットやユビキタスといった実世界での情報支援**
- **海外の研究者ネットワークの抽出**

# まとめ

- Polyphoneは研究者ネットワーク検索エンジン
- Webマイニング技術利用
- 研究者ネットワーク作成、研究者検索など可能
- 今後、融合領域での活用と様々分野の適用

# 参考文献

- **Webからの人物に関するキーワード抽出**

<http://ymatsuo.com/papers/jsai05jmori.pdf#search='web%E3%81%8B%E3%82%89%E6%8A%BD%E5%87%BA%E3%81%99%E3%82%8B%E3%82%BD%E3%83%BC%E3%82%B7%E3%83%A3%E3%83%AB%E3%83%8D%E3%83%83%E3%83%88%E3%83%AF%E3%83%BC%E3%82%AF+%E6%9D%BE%E5%B0%BE%E3%82%86%E3%81%9F%E3%81%8B'>

- POLYPHONET [http://ymatsuo.com/papers/pamphlet\\_polyphonet\\_robot.pdf](http://ymatsuo.com/papers/pamphlet_polyphonet_robot.pdf)

- **Semantic Webのための人の社会ネットワーク抽出と利用** [http://www.miv.t.u-tokyo.ac.jp/papers/Tomobe-](http://www.miv.t.u-tokyo.ac.jp/papers/Tomobe-ipsj05.pdf#search='web%E3%81%8B%E3%82%89%E6%8A%BD%E5%87%BA%E3%81%99%E3%82%8B%E3%82%BD%E3%83%BC%E3%82%B7%E3%83%A3%E3%83%AB%E3%83%8D%E3%83%83%E3%83%88%E3%83%AF%E3%83%BC%E3%82%AF')

[ipsj05.pdf#search='web%E3%81%8B%E3%82%89%E6%8A%BD%E5%87%BA%E3%81%99%E3%82%8B%E3%82%BD%E3%83%BC%E3%82%B7%E3%83%A3%E3%83%AB%E3%83%8D%E3%83%83%E3%83%88%E3%83%AF%E3%83%BC%E3%82%AF'](http://www.miv.t.u-tokyo.ac.jp/papers/Tomobe-ipsj05.pdf#search='web%E3%81%8B%E3%82%89%E6%8A%BD%E5%87%BA%E3%81%99%E3%82%8B%E3%82%BD%E3%83%BC%E3%82%B7%E3%83%A3%E3%83%AB%E3%83%8D%E3%83%83%E3%83%88%E3%83%AF%E3%83%BC%E3%82%AF')

- **楽天技術研究所におけるウェブマイニング**

<http://el.jibun.atmarkit.co.jp/rakuten/2009/11/post-7414.html>

- **Web上の情報から人間関係ネットワークの抽出**

<http://www.miv.t.utokyo.ac.jp/papers/matsuoJSAI05.pdf#search='Web+%E4%B8%8A%E3%81%AE%E6%83%85%E5%A0%B1'>

- **ロボット分野版「研究者ネットワーク検索エンジン」の実用化試験スタート**

[http://www.aist.go.jp/aist\\_j/press\\_release/pr2005/pr20050418/pr20050418.html](http://www.aist.go.jp/aist_j/press_release/pr2005/pr20050418/pr20050418.html)



ご清聴ありがとうございました！

# 共起の指標

- Matching Coefficient (**共有度係数**)  $|X \cap Y|$
- Mutual Information (**相互情報量**)  $\log N |X \cap Y| / |X| |Y|$
- Dice Coefficient (**さいころ係数**)  $2|X \cap Y| / (|X| + |Y|)$
- Jaccard Coefficient (**ジャカルル係数**)  $|X \cap Y| / |X \cup Y|$
- Simpson coefficient (**シムソン係数**)  $|X \cap Y| / \min(|X|, |Y|)$
- Cosine (**余弦**)  $|X \cap Y| / (\sqrt{|X| |Y|})$

# 閾値付きOverlap係数

- 共起頻度を表す係数
- Overlap 係数が人の協働関係の強さを表すのに最も適しているという知見を示している
- 単独ヒット件数が極端に少ない人との関係の強さを正確に測れるため閾値が必要

例)  $X=1, Y=100, X \cap Y=1$       Overlap係数=1