リンク情報に基づく検索エンジンの比較

樺島結城† 廣川佐千男‡

† 九州大学大学院システム情報科学府

‡ 九州大学情報基盤センター

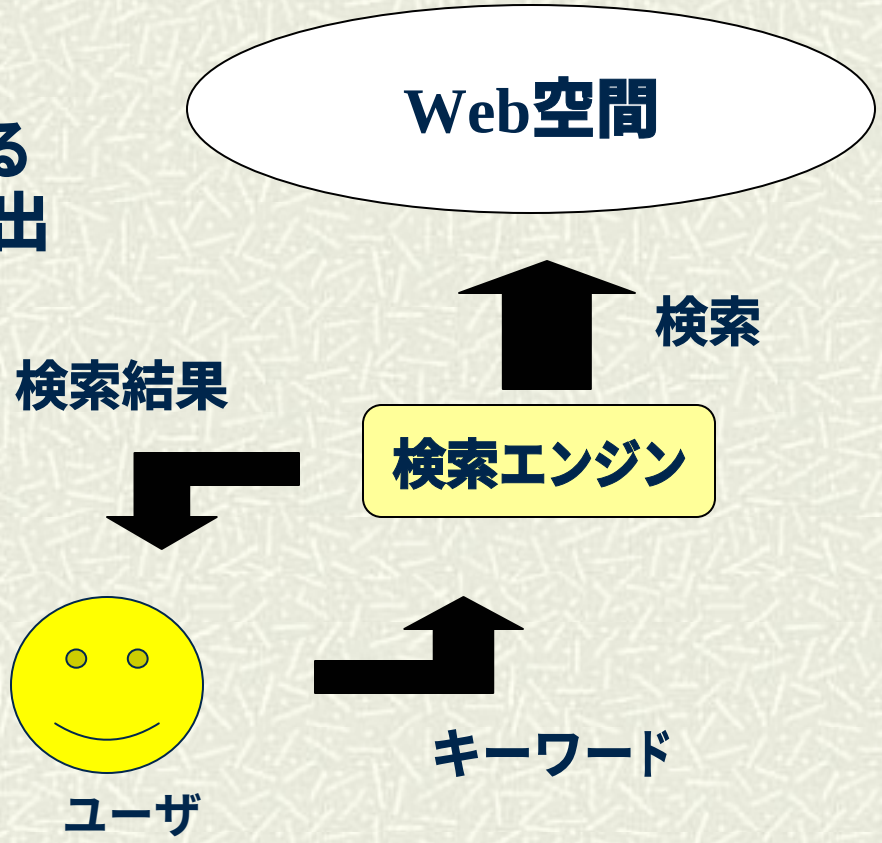
発表内容

- 背景
- 研究の目的
- 入次数と出次数での比較
- オーソリティー度とハブ度での比較
- まとめと今後の課題

背景

■Web空間には情報が氾濫
→有用な情報も多く存在するが、必要な情報を見つけ出すのが困難

ユーザは検索エンジンを用いて情報を探す



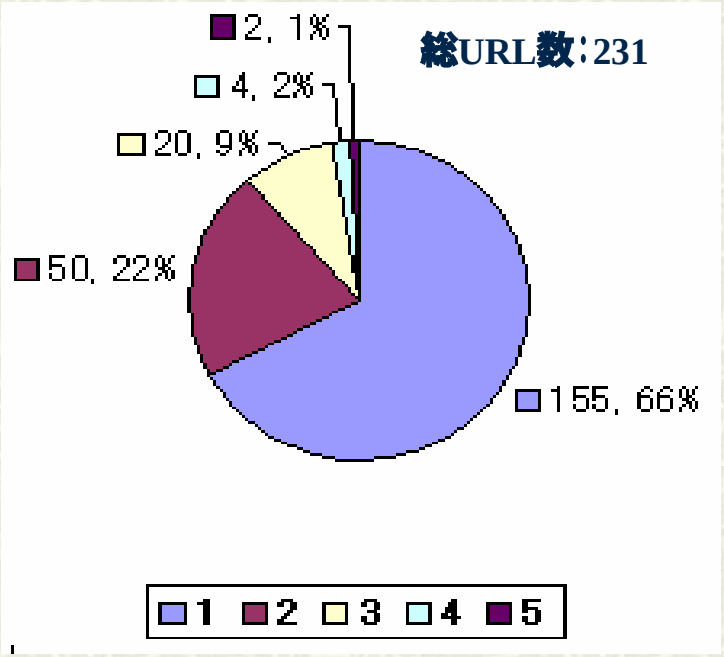
検索結果の重複

■検索エンジンも複数あり、
違いが大きい

■利用する際に感じる“違い”
を評価できないか？

検索結果上位10件の重複数

(4つの検索エンジンに対し7のキーワード)



使用した検索エンジン

Alta Vista, goo, google

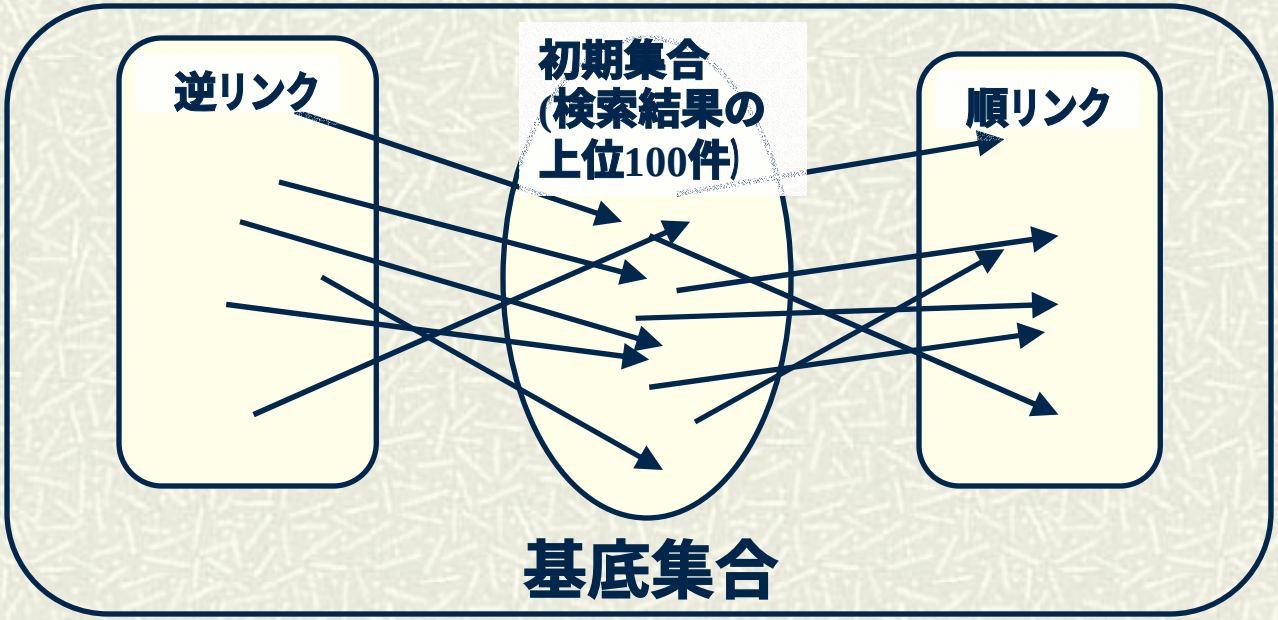
Yahoo!(カテゴリ検索,ページ検索)

本研究の目的

- 目的:リンク情報を利用して検索エンジンの比較を行う

- 実験内容
  - 四つの検索エンジンに対して比較
    1. 入次数、出次数での比較実験
    2. オーソリティー度、ハブ度での比較実験

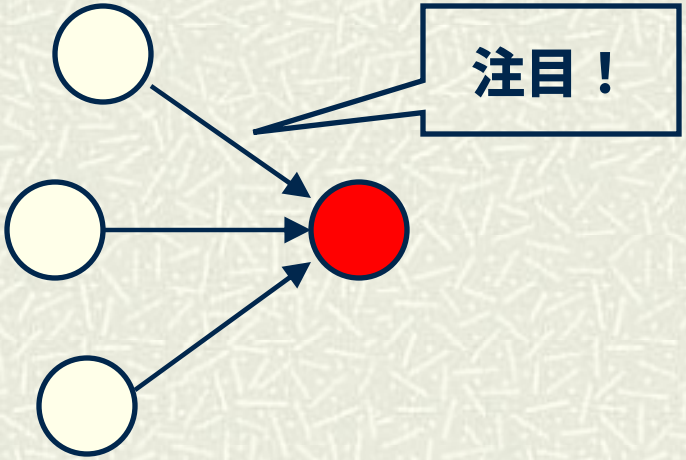
基底集合



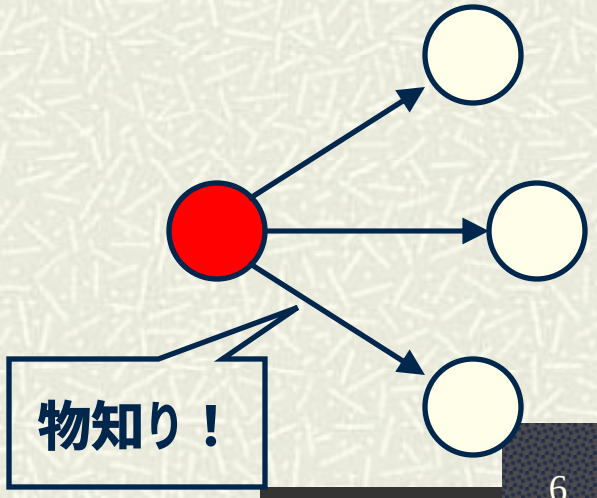
基底集合についての隣接行列
を分析の対象とする(検索空間)

入次数と出次数

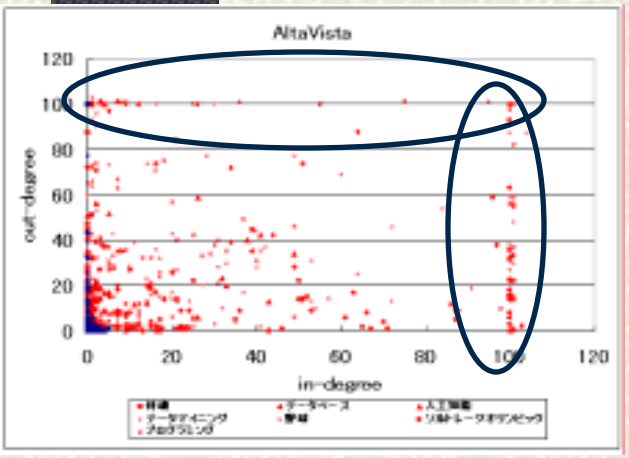
- **入次数(indegree)**
  - ー各ページに対するリンクの数
    多くのページからリンクされているページはそれだけ注目されているといえる

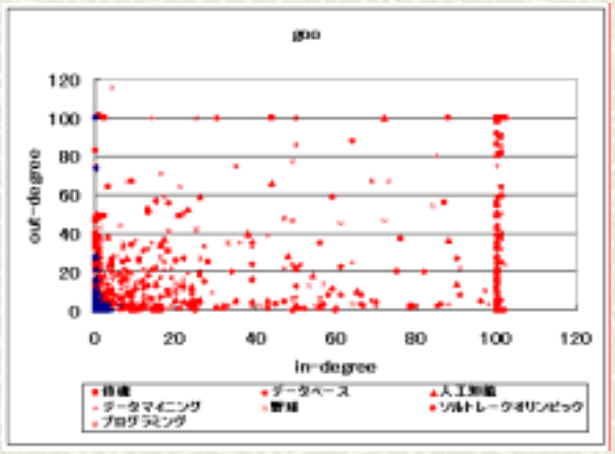


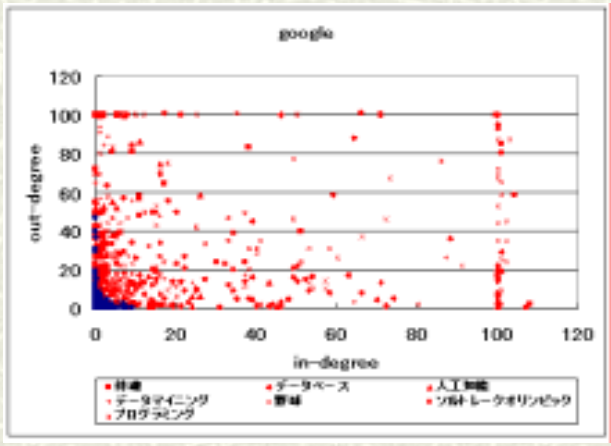
- **出次数(outdegree)**
  - ー各ページからのリンク数
    多くのページをリンクしているページは多くの情報を持っているといえる

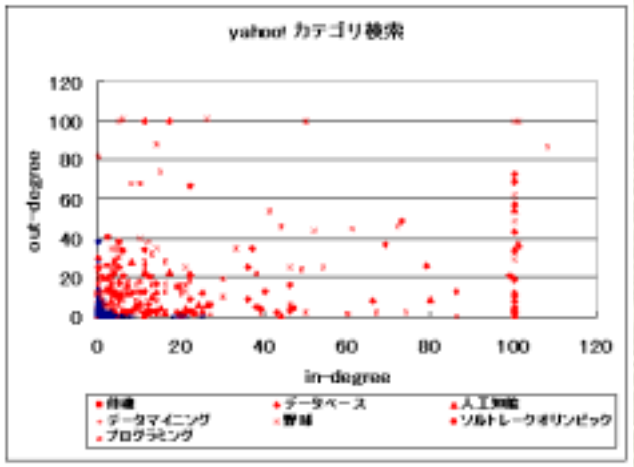


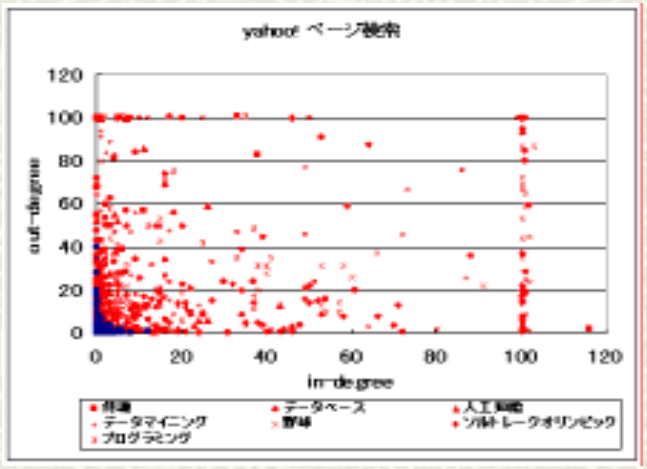
実験1:入次数、出次数での比較







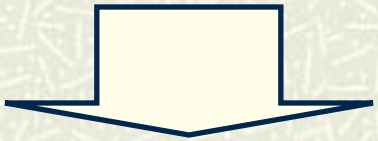




問題点

問題点

キーワードとは関係のない内容のリンクの数まで評価されてしまう



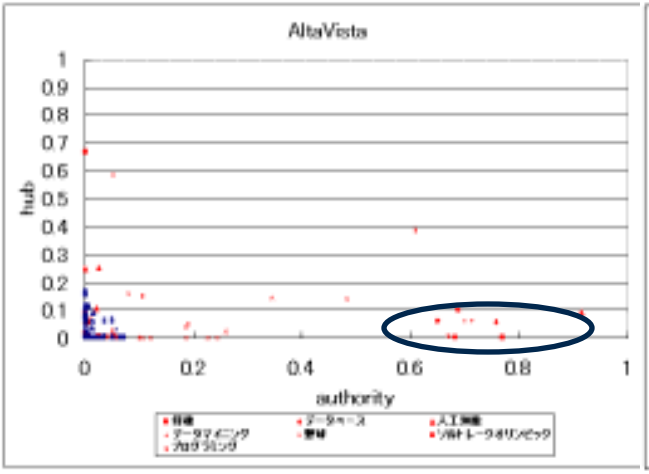
各リンクを客観的に評価する必要がある

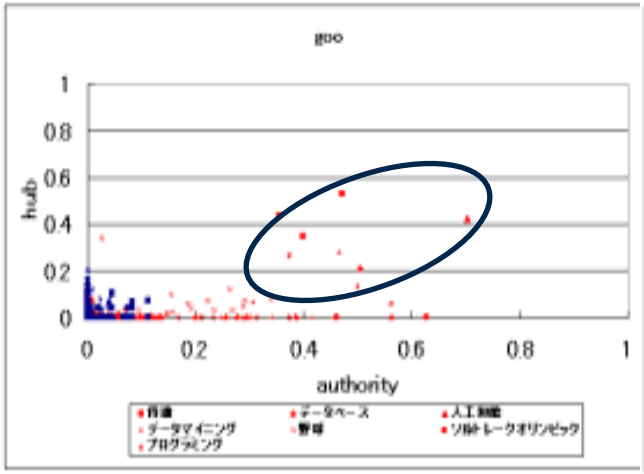
オーソリティー度とハブ度

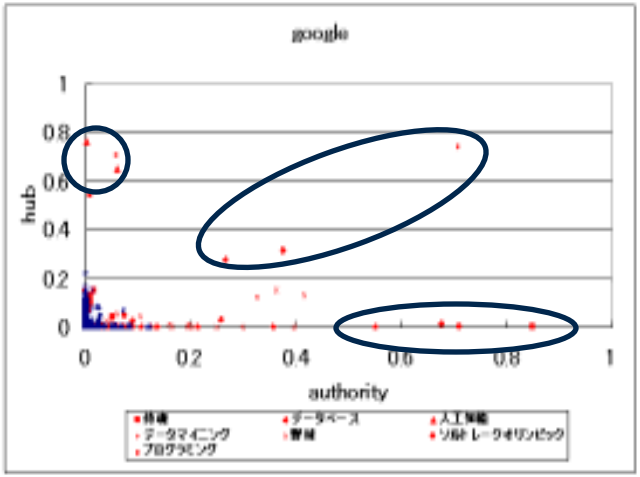
- **オーソリティー度**(authority)
  - **ー各ページをリンクしているページのハブ度の和。**
- **ハブ度**(hub)
  - **ー各ページがリンクしているページのオーソリティー度の和。**

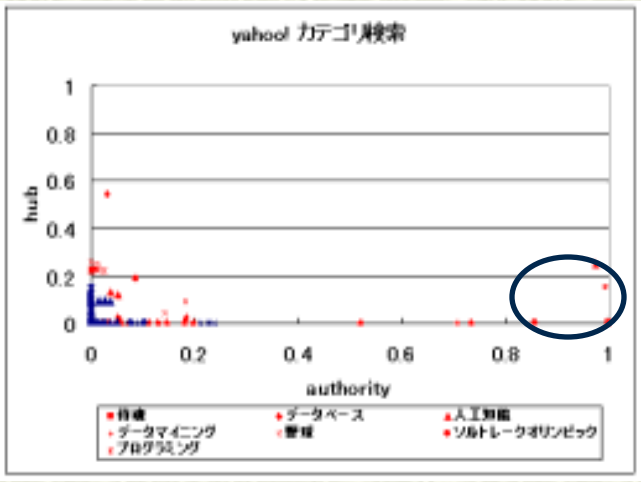
これらを反復計算により求めることで、各リンクに客観的な評価を与える

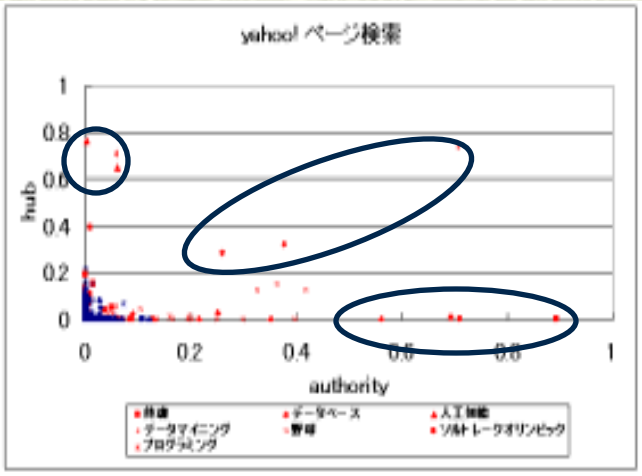
実験2：オーソリティー度、ハブ度での比較











まとめと今後の課題

まとめ

- 入次数、出次数を用いて検索エンジンを比較
- ランキングの手法であるオーソリティー度、ハブ度を検索エンジンの比較に導入

今後の課題

- より多くの検索空間での比較
- 検索空間でのクラスタの抽出
- リンク情報以外の、別な側面からの検索エンジンの比較

HITSアルゴリズム

HITSアルゴリズム

オーソリティー度およびハブ度を反復計算によって求める。
具体的には、検索空間におけるリンク情報の隣接行列をA、オーソリティー度、ハブ度のベクトルをそれぞれx,yとすると、下記1～3の手順をx,yが収束するまで反復することで求められる。
初期値として、x,yは値がすべて1のベクトルとする。

1：$x_k = A^t y_{k-1}$
2：$y_k = A x_k$
3：x_k，y_k　を正規化する