

# 卒業論文概要書

2010年2月提出

|          |                                  |      |                  |          |         |
|----------|----------------------------------|------|------------------|----------|---------|
| 学科名      | コンピューター・ネットワーク工学科                | 氏名   | 久保田 大動           | 指導<br>教員 | 大石進一 教授 |
|          |                                  | 学籍番号 | CD<br>1G06R065-1 |          |         |
| 研究<br>題目 | PageRank に基づいた新たな技術価値の付加及びその簡易実装 |      |                  |          |         |

## 1. はじめに

日本に限らず世界中において使用されているインターネット検索エンジンとして非常に高いシェアを誇る Google 社の考案した、そして現在でも検索結果のページ表示順序の決定する手段として主に用いられている PageRank 技術の概要について簡潔に説明する。それに並行して現状の検索エンジンを使用者側に立った特徴及び解決すべき問題点や課題についても同時に説明する。

卒業論文の趣旨については、既存技術である PageRank の概念を今回の論文を通じて包括的に理解した上で、そこにいかなる新しい技術価値を上乗せできるかについて重点的に研究する。

## 2. PageRank の理論について

概要に引き続き、PageRank についてどのような数学的モデルがベースとなっている概念なのかを説明していく。

基本となる定常分布ベクトル  $\vec{\pi}$  及び推移確率行列  $G$  を元にしたべき乗法による PageRank 導出の説明、Web ページのリンク構造から隣接行列への変換手順、及びその行列の正規化を経て得られる推移確率行列  $G$  の生成方法など PageRank の基本的な考え方を説明する。

### 2.1 PageRank の特性について

定常分布ベクトル  $\vec{\pi}$  が上べき乗法によって収束するか否か(エルゴード性)についての説明や、マルコフ連鎖の簡潔な説明の後、PageRank を導出するにあたってどのようなコンピューティング環境が求められるかを簡潔に説明する。

## 3. 既存の PageRank にどのような新しい価値が付加できるかを考察

Google 検索エンジンを向上するにあたってどのような技術を応用でき、また PageRank とは全く別のいかなる技術を付加すれば既存の検索エンジンとは違う一線を画したシステムとなるかを考察する。

## 4. PageRank に対する時間概念の付加

Google における WWW 上のリンク構造分析にはクローラー(Crawler)と呼ばれる巡回ロボットが定期的に巡回して分析していると言われており、おもに HTTP 経由で情報を取得している。

既存技術である PageRank に対して時間といった検索条件を付加しようと試みた場合、そういったことは可能なのだろうか。可能な場合、Crawler のように HTTP 経由で必要な各種情報を取得できるかといった技術的な問題についても同時に触れる。

## 5. ローカル環境下で用意されたリンク構造を分析し PageRank を導出するシミュレーション

インターネットに接続されている環境下ではなく、ローカル上の仮想環境下においてリンクファイルの保有するリンク情報から隣接行列  $S$  及び正規化した推移確率行列  $G$  の導出、それからその行列  $G$  から固有値・固有ベクトルを導出するための行列計算ライブラリである Lisys の導入、その他計算するにあたり有効な HouseHolder 法や QR 分解法、LU 分解等といった各種アルゴリズムについても取り上げる。

## 6. 総括

上記で述べた PageRank の概念を元にして派生した研究及びローカル環境での全体的な計算シミュレーションを通じて最終的に得られた結果や課題、卒論趣旨である新たな技術を付加することができたか、また今後の展望についてを総括的に説明する

## 7. 付録: ソースコード

上記 HouseHolder 法・QR 分解のアルゴリズム等、また行列固有値の導出、さらに行列を直接表示できるよう改良したグラフィックインターフェースなどのソースコード群を付録として添付