

Wikipediaと検索エンジンサジェストを用いた Web検索語候補推薦システムの試作

宮下 竜太郎, 増田 英孝, 山田 剛一 (東京電機大学)
清田 陽司, 中川 裕志 (東京大学)

NLP若手の会 第5回シンポジウム
2010/09/15 国立情報学研究所

検索エンジンサジェストの問題

- 検索には適切なキーワード選択が必要
- 主要な検索エンジンには**サジェスト機能**がある

「電機大」の検索語候補
電機大学
電機大学 2ch
電機大学 偏差値
電機大学 理工学部
電機大手
電機大手 決算
電機大手8社
電機大手8社とは
電機大手9社
電機大手9社とは

Google サジェストより取得
電機大=**東京電機大学**を期待

半数以上が東京電機大学
とは無関係なサジェスト

- 前方一致出力のために**語の異表記やゆれに対応していない!**

Wikipediaの情報を使ってサジェストを整理する

- 「電機大」と入力した場合においても「東京電機大学」のサジェストが出力され、電機大に含まれる**電機大手は別物扱い**になる

-----[東京電機大学]-----

- 東京電機大学
- 東京電機大学 2ch
- 東京電機大学 オープンキャンパス
- 東京電機大学 メディアセンター
- ~
- 東京電機大学 高校
- 東京電機大学中学校
- 東京電機大学高校
- 東京電機大学高等学校

-----[電機大]-----

- 電機大 スレ
- 電機大 傷害
- 電機大 北千住
- 電機大 千住
- 電機大 浪人
- 電機大 移転
- ~

- 電機大手(固有名詞)
- 電機大手 決算(固有名詞)
- 電機大手8社(固有名詞)
- 電機大手8社とは(固有名詞)
- 電機大手9社(固有名詞)
- 電機大手9社とは(固有名詞)
- ~

是非ともお立ち寄りください

- **提案手法**
- **使用しているWikipediaの情報**
- **今後の方針に関する議論**
- 11:15-12:15 ポスターセッション2
- **発表10: Wikipediaと検索エンジンサジェストを用いた
Web検索語候補推薦システムの試作**