

NLP若手の会第3回シンポジウム
ポスターセッション発表

Wikipedia解析ツール **Wik-IE**

森 竜也、増田 英孝 (東京電機大学)
清田 陽司、中川 裕志 (東京大学)

辞書作成の情報源としての Wikipedia

- ◆ ユーザによる階層的なカテゴリ付け
- ◆ リダイレクト設定
- ◆ ページ間リンク
- ◆ 他言語版Wikipedia記事への言語間リンク

辞書作成の情報源として
有用で使いやすい情報豊富

現状

- ◆ 使ってみたいけどやり方が分からない？
- ◆ 自分でプログラムを作る必要がある？
- ◆ 環境を整えるのが大変そう？
- ◆ データが更新されるたびにデータベースを更新するのは面倒？

誰でも簡単に使えるツールがあれば便利！

Wikipedia解析ツール Wik-IE

- ◆ Wikipediaのデータファイルを直接解析
- ◆ 単純なテキスト形式で出力
- ◆ Javaプログラム
- ◆ ファイルと機能を指定するだけで実行
- ◆ Hadoopプラットフォーム上での分散処理
- ◆ マスタ1台、スレーブ3台で処理した場合
- ◆ 日本語版で約20分
- ◆ 英語版で約2時間
- ◆ スタンドアロン版もあり

ぜひお立ち寄りください

- ◆ 「Wikipediaの情報は信用できるの？」
- ◆ 「実際に何に使えるの？」
- ◆ 実行結果やそれを使って何ができるかなど
実例を交えて紹介します
- ◆ 16:30-17:30 ポスターセッション2
- ◆ 発表7. Wikipediaデータ抽出ツールWikIE