

# WikipediaとWebの解析による大規模関係辞書構築

大阪大学 大学院情報科学研究科 マルチメディア工学専攻 西尾研究室 白川 真澄

**目的** 整理された情報を持つWikipediaと、膨大な量のデータを持つWebを解析することにより、大規模な関係辞書を構築する。

**内容** Wikipediaから抽出した関連度情報をもとにWeb検索クエリを作成し、検索結果の文書を解析することで動作関係を抽出する。

**結果** 提案手法により、既存の統計的なアプローチでは抽出できない（あるいは精度よく抽出できない）ような関係についても精度よく抽出できることを確認した。実際に大規模な解析実験を行い、30万弱の動作関係を持つ辞書を構築した。構築した動作関係辞書のプロトタイプをWeb上 (<http://sigwp.org/>) で公開している。

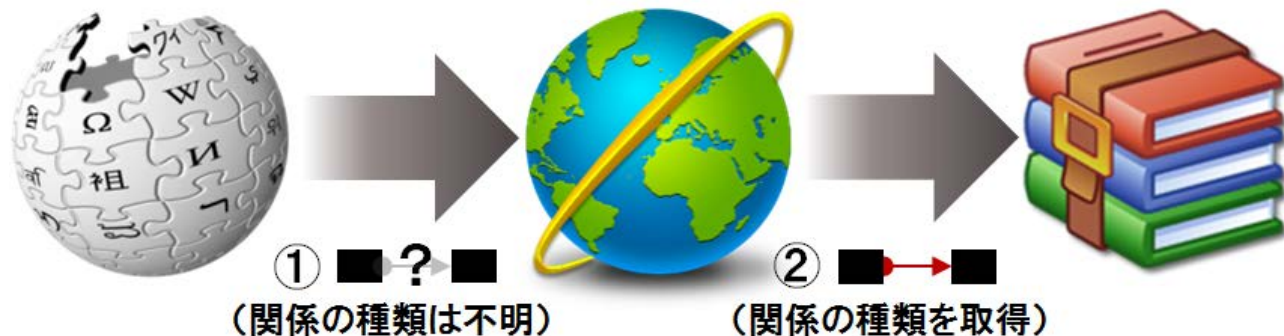


図: 関係辞書構築のアプローチ