

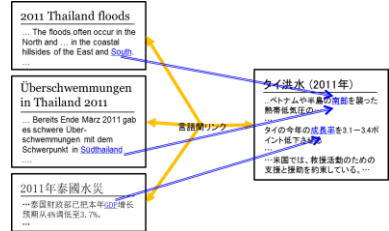
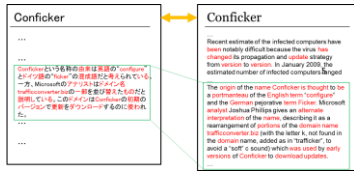
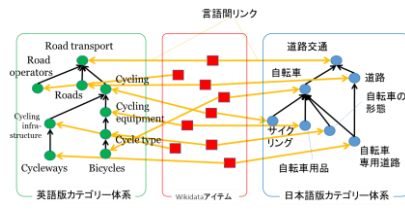
# Wikipediaの多言語利用における 言語障壁の解消に関する研究

新谷 誠, 綱川 隆司, 許山 秀樹, 梶 博行

## ● 研究の背景と目的

- Wikipediaでは異なる言語版の内容に整合性がない
- 他言語版へのアクセス、他言語版記事の理解に労力を要する
- クロス言語アプローチによるWikipediaの質・ユーザビリティの向上

## ● 開発内容

サブテーマ	多言語wikification (リンク自動付与)	記事対内部の対応関係の抽出	多言語タクソノミーの構築
概要	<ul style="list-style-type: none"> <li>他言語版記事のリンクの言語間変換により、記事のリンクを増強</li> <li>適切なアンカー選択方法、変換したリンクの順位付け方法を提案</li> </ul> 	<ul style="list-style-type: none"> <li>二言語の記事間で同じ内容のパスセージを抽出し、記事の増補・参照を支援</li> <li>単語対応可能性行列による方法を提案</li> </ul> 	<ul style="list-style-type: none"> <li>言語間で整合性のないカテゴリ体系から、汎言語カテゴリ体系を自動的に構築</li> <li>カテゴリ体系のわかりやすい視覚化</li> </ul> 
特長	通常wikificationと異なり、単純な方法で高精度にリンクを得られる	記事中の内容のまとめ(パスセージ)とそれらの対応関係を同時に判定	観点の異なる複数の上位カテゴリを保持したまま汎言語カテゴリ体系を構築
効果	アンカーの正解率は95~96%を保ち、リンク数は日本語記事1記事あたり12.8件増加	正解を用意した10記事の対応関係に対し、一致度は平均0.240(最大0.659、最小0)	汎言語カテゴリ体系の一部を構築 各言語の体系と比較する視覚化方法を提案

## ● 今後の展開

- 多言語Wikificationの一般テキストへの応用
- カテゴリ体系の言語間比較によるWikipediaの質向上