

データの説明

データ番号 (コモンズセンターで記入します)	
タイトル	日本の 47 都道府県に関する Wikipedia 記事コーパス ① 書き込み記事のテキスト情報 ② 記事に含まれる名詞、形容詞の頻度別データ
作成者	筑波大学シスム情報系社会工学域 助教 佐野 幸恵 佐野幸恵
問い合わせ先	佐野 幸恵 029-853-5552 sano@sk.tsukuba.ac.jp
概要(データの内容や作成方法)	2016 年 3 月までに Wikipedia 日本語版に書き込まれた、日本の 47 都道府県に関する記事コーパス。詳細は README.txt を参照。 ① Wikipedia 本文のテキスト ② ①のテキストを形態素解析し、名詞、形容詞(原型)のみを抽出して、頻度別に整理したもの
更新履歴(版・ファイル名・年月日)	Ver.1.0 (2017/04/10)
データの形式	Txt (UTF-8)
データのサイズ	① 47 ファイル、合計 3.0MB ② 47 ファイル、合計 2.0MB
利用上の注意(メタデータの利用など)	詳細は Wikipedia の著作権方針に従う。 (https://ja.wikipedia.org/wiki/Wikipedia:%E8%91%97%E4%BD%9C%E6%A8%A9)
関連報告書・論文等	
備考	オリジナルデータの URL https://dumps.wikimedia.org/jawiki/latest/ このファイルは、クリエイティブ・コモンズ 表示-継承 3.0 非移植のライセンスで配布されている。 