

# Wikipedia 知識の構造化プロジェクト 「森羅」

関根聡<sup>†1</sup>

**概要:** 「説明できる言語処理技術」を目指して、Wikipedia にある知識を構造化するプロジェクト「森羅」を推進している。「森羅」では「Resource by Collaborative Contribution (RbCC)」の方針のもと、多くの参加者が協力して知識を作成する方法を実践している。2018 年には 8 グループから 16 のシステムが参加し、有益な結果が得られた。また、その結果のアンサンブル学習からは RbCC の方針が有効である事が分かった。2018 年には 3 つのタスクを実施する予定である。

**キーワード:** 知識、構造化、Wikipedia、オントロジー、自然言語処理

## “SHINRA” A Project for Structuring Knowledge on Wikipedia

Satoshi SEKINE<sup>†1</sup>

**Abstract:** “SHINRA” project is a project aiming at structuring knowledge in Wikipedia so that we can embody the explainable AI technology. It is conducted under the principle of “Resource by Collaborative Contribution (RbCC)” and we have many participants and collaborators in the field. At SHINRA-2018, we have 16 systems from 8 group participated on the task of structuring Wikipedia on 5 categories. The ensemble learning experiments using the output demonstrated the idea of RbCC works very effectively. At SHINRA-2019, we are planning to conduct 3 tasks.

**Keywords:** Knowledge, Structured Knowledge, Wikipedia, Ontology, Natural Language Processing

**謝辞** MS-Word 「森羅」プロジェクトに参加、協力していただいている全ての皆様に謹んで感謝の意を表す。

### 参考文献

- [1] “「拡張固有表表現+Wikipedia」データ(2015 年 11 月版 Wikipedia 分類作業完成版)” 関根聡 (ランゲージ・クラフト/理研 AIP), 安藤まや (ランゲージ・クラフト), 小林暁雄 (理研 AIP), 松田耕史, 鈴木正敏 (東北大), Duc Nguyen (オルツ), 乾健太郎 (理研/東北大) 言語処理学会第 24 回年次大会
- [2] “Wikipedia 構造化データ「森羅」構築に向けて” 関根聡, 小林暁雄 (理研 AIP), 安藤まや (ランゲージ・クラフト), 馬場雪乃 (理研 AIP/京大), 乾健太郎 (理研 AIP/東北大) 言語処理学会第 24 回年次大会