

リクルートのAI研究機関が、高速・高精度な日本語の解析を実現する 日本語自然言語処理ライブラリ「GiNZA version 4.0」を公開

株式会社リクルートホールディングスの中間持ち株会社である株式会社リクルート（本社：東京都千代田区、代表取締役社長：北村吉弘、以下リクルート）は、このたび、当社のAI研究機関であるMegagon Labsより、高速・高精度な日本語解析を実現する日本語自然言語処理オープンソースライブラリ「GiNZA version 4.0」を無料公開しました。

1 日本語自然言語処理ライブラリ「GiNZA」について

自然言語処理技術とは、私たちが日常的に使っている言語（自然言語）をコンピューターに処理させる一連の技術を指し、検索エンジンや機械翻訳、対話システム、顧客の声分析など、生活・ビジネスにおけるさまざまなシーンで利用されています。

リクルートの研究開発機関・Megagon Labsが開発・提供する「GiNZA」は、機械学習を利用した日本語の自然言語処理に関心があり、日本語に特化した自然言語解析を行いたいと考えている研究者や企業のエンジニア・データサイエンティストに向けて開発されたオープンソースライブラリです。誰でも無料でご利用いただくことができ、ビジネスの現場でも広く利用されることを想定し、ワンステップでの導入、高速・高精度な解析処理、単語依存構造解析レベルの国際化対応などの特長を備えています。

2020年1月にversion 3.0を公開して以降、6ヶ月で23,000ダウンロードを超え、Universal Dependencies (UD、※1)の日本語解析系として、学術機関だけでなく、頑健かつ柔軟な応用が可能な実用的ライブラリを望む産業界の多くの方々にご利用いただいています。



2 「GiNZA version 4.0」アップデートの主な特長 ～高速・高精度な日本語解析を実現～

1. より高速で高精度な解析が可能

最先端の機械学習技術を取り入れ、産業利用を想定して高速・軽量化された多言語対応自然言語処理フレームワークであるspaCy version 2.3は、Megagon Labsが公開したコーパス（※2）を利用し公式日本語対応モデルをリリースしました。このspaCy version 2.3のフレームワークアーキテクチャーに合わせてライブラリを再設計したことにより、GiNZA 4.0はこれまで以上に高速かつ高精度な解析が可能となっています。

2. 日本語の文節構造にも対応

日本語を含む15言語に対応したspaCyの多言語解析フレームワークを用いることで、構文構造の解析から人名・組織名・地名・商品名・数値といった固有表現の抽出に至るまで、全ての言語をリソース差し替えのみで統一的に解析できます。さらに、「GiNZA version 4.0」では、UD体系のもとでMegagon Labs独自のチューニングを施した日本語解析モデルを搭載することで、日本語文法に根ざした解析機能をより一層強化し、日本語の文節を単位とする解析結果が容易に得られるようになっていきます。



従来のUD体系では、単語（トークン）単位での依存関係を定義していたのに対し、「GiNZA version 4.0」は、文節内の主語にあたる単語を区別して学習することにより、文節単位での依存構造を捉えることが可能

3. 複合語の区切り方も三段階から選択可能

SudachiPy（※3）の機能を利用しており、複合語の区切り方を三段階から選択可能です。

例：機能性食品（デフォルト）→機能性+食品（モードB）→機能+性+食品（モードA）

Megagon Labsは今後、「GiNZA」をさらにアップデートしていくことで、より速く、高精度な日本語の自然言語処理を可能にし、あらゆる産業において自然言語処理の活用が促進される世界をめざし、研究活動を進めてまいります。

※1 全世界の多様な言語を一貫した文法・品詞体系で解析可能にすることを旨とした国際的な学術プロジェクト

※2 国立国語研究所からリリースされたUD_Japanese-GSD v2.6に固有表現ラベルを追加し、Megagon Labsが2020年5月に公開したコーパス（正式名：UD_Japanese-GSD v2.6-NE）

※3 株式会社ワークスアプリケーションズの自然言語処理研究に特化したAI研究機関「ワークス徳島人工知能NLP研究所」が開発するオープンソースソフトウェア

About Recruit Group

リクルートグループについて

1960年の創業以来、リクルートグループは、就職・結婚・進学・住宅・自動車・旅行・飲食・美容などの領域において、一人ひとりのライフスタイルに応じたより最適な選択肢を提供してきました。現在、HRテクノロジー、メディア&ソリューション、人材派遣の3事業を軸に、45,000人以上の従業員とともに、60を超える国・地域で事業を展開しています。2019年度の売上収益は23,994億円、海外売上比率は約45%になります。リクルートグループは、新しい価値の創造を通じ、社会からの期待に応え、一人ひとりが輝く豊かな世界の実現に向けて、より多くの『まだ、ここがない、出会い。』を提供していきます。詳しくは <https://recruit-holdings.co.jp/> をご覧ください。