

AWSの検索サービス「Amazon OpenSearch Service」で 「Sudachi」が利用可能に 国内最大規模の辞書で高精度な検索を実現

株式会社ワークスアプリケーションズ（本社：東京都千代田区、代表取締役最高経営責任者：秦修、以下 WAP）は、WAPが無償公開するオープンソースソフトウェアの日本語形態素解析器*1「Sudachi」について、2023年10月11日より、アマゾンウェブサービス（以下 AWS）が提供する「Amazon OpenSearch Service」において、日本語形態素解析のカスタムプラグインとして利用可能となったことをお知らせします。これまでAWSユーザーからは「Sudachi」を利用したいとの声が度々上がっていました。今後は検索エンジン内のカスタマイズ機能として、利用者が「Sudachi」を選択できるようになり、検索力の向上が期待されています。

powered by 



国内最大規模の語彙をもつ日本語形態素解析器「Sudachi」

Amazon OpenSearch Serviceは、AWSクラウドにおけるOpenSearchクラスターのデプロイ、オペレーション、スケーリングを容易にするマネージドサービスです*2。クラスターを作成するときに、どの検索エンジンを使用するかを選択できるオプションがあります。AWS利用者がSudachiをAmazon OpenSearch Serviceのカスタマイズ機能として選択すると、Sudachiがもつ高い言語解析精度や柔軟性により、AWSクラウド上で高精度な検索を実現させることができます。

Sudachiは、WAPのワークス徳島人工知能NLP研究所が開発しているオープンソースソフトウェア（OSS）型の日本語形態素解析器です。形態素解析は、テキストを意味を持つ最小単位に分割し、品詞などの情報を付与します。なかでもSudachiは、（1）290万語を超える国内最大規模の語彙を収録、（2）複数の単語分割単位を選択・併用することが可能、（3）文字種の違い、送り仮名の違いなど日本語単語がもつ表記の揺れを吸収可能、（4）プラグインにより様々な機能を追加可能といった特長をもっています。

Sudachiは解析に必要な辞書に最新の単語を取り込み、定期的な更新・メンテナンスを行うことで、商用利用に耐えられる高品質な性能を確保しています。SudachiのPython版である「SudachiPy」は、2019年6月にOSSとして公開し、2023年9月に1100万ダウンロードを突破しました。

<https://worksapplications.github.io/Sudachi/>

曖昧さや表記の揺れがあっても同じ言葉と判断

漢字の違い (異体字、代用表記、慣用表記)	芸術—藝術、驚歎—驚嘆、徳用—得用
文字種の違い	向日葵—ひまわり—ヒマワリ
送り仮名の違い	受け付け—受付け—受付
縮約形（くだけた言い方）	～ちゃあ～～ては

Sudachiに関してお客様専用のヘルプデスク・障害対応・情報発信をご希望の方へ

WAPIはSudachiに関する保守サービスを提供しています。お問合せは以下よりお願いいたします。

https://landing.worksap.co.jp/SaaS_LP_Sudachi_LP.html

※OpenSearchプラグイン（Sudachi）についてのお問合せはAWSへお願いいたします。

Sudachi GitHubスポンサー募集中

「GitHubスポンサー」は、2019年に世界30カ国でリリースされたGitHub*3が提供するスポンサーシッププログラムで、オープンソース・プロジェクトの開発者やチームを財政的に支援することを目的としています。WAPではSudachiの研究開発力を強化し、OSSとしての持続的開発を実現するため、Sudachi GitHubスポンサーを募集しています。

Sudachi GitHubスポンサーは、個人・組織を問わず受付しています。スポンサー料の金額や支援頻度は、1回1ドルから自由に設定できます。また、スポンサー料へのリワードとして「Githubページへのスポンサーバッチ表示、ウェブサイトへのロゴ掲載、開発ロードマップへのアクセス、スポンサー用チャットスペース参加、優先バグ対応、ワークショップ開催」などが用意されています。

スポンサー募集に関する詳細は、以下サイトよりご覧ください。

<https://github.com/sponsors/WorksApplications>

プロフェッショナルサービスとして「Sudachi有効活用における設定、辞書作成支援」を提供

WAPが提供する「Sudachi有効活用における設定、辞書作成支援」が、OpenSearch上のSudachi有効活用支援として、AWS Marketplace*4上にて購入可能となる予定です。本プロフェッショナルサービスでは、経験豊富なワークス徳島人工知能NLP研究所のメンバーが、検索精度向上のためのさまざまな支援を提供します。

*1:形態素解析

自然言語処理（NLP）の一部で、普段生活の中で一般的に使っている言葉である「自然言語」を形態素（言葉が意味を持つまとまりの単語の最小単位）にまで分割する技術

*2:Amazon OpenSearch Serviceの詳細についてはこちらをご覧ください。

https://docs.aws.amazon.com/ja_jp/opensearch-service/latest/developerguide/what-is.html

*3:GitHubは4000万人もの開発者に利用されており、オープンソースプロジェクトからビジネスユースまでソフトウェア開発の中心となる開発プラットフォーム

*4:AWSの詳細についてはこちらをご覧ください。

<https://aws.amazon.com/jp/mp/marketplace-service/overview/>

【ワークス徳島人工知能NLP研究所について】

WAPが2017年2月に開設した研究機関で、AI分野の中でもNLP（自然言語処理）に特化して曖昧な表現や意味の重複、表記の揺れがある言葉をコンピュータが正しく処理するための研究を行っています。研究成果の多くはチャットボットやAI-OCRなどの自社製品に活用しているほか、ほかの企業や研究機関で活用できるようにOSSとして無償公開しており、この分野の研究や企業のAI活用に役立てられています。産学官連携の取り組みのほか、無償公開したOSSが幅広く活用されている点などが評価され、地域情報化の推進に寄与したとして、2022年度の「徳島県地域情報化表彰(e-とくしま表彰)」を受賞しました*。

<https://nlp.worksap.co.jp/>

*1月18日付リリース：

ワークス徳島人工知能NLP研究所、「徳島県地域情報化表彰」受賞
産学官共同研究や、760万DLされた自然言語処理のOSSなどを評価

<https://www.worksap.co.jp/news/2023/0118/>

【ワークスアプリケーションズについて】

ワークスアプリケーションズは、1996年の創業以来、日本発の業務アプリケーションのパッケージソフトウェア会社として、主に国内の大手企業向けに製品・サービスを提供してまいりました。「働く」の概念を変え、仕事をより創造的なものへ、企業の生産性を高め、企業価値を拡大する、この企業理念のもと、ERPを軸としたソリューションプロバイダーとして、大手企業に加えて中堅・中小・スタートアップ企業のDX推進のパートナーとなれるよう、さらなる発展を目指していきます。

株式会社ワークスアプリケーションズ サイト <https://www.worksap.co.jp/>

*会社名、製品名およびサービス名は各社の商標または登録商標です。

*本リリースに掲載された内容は発表日現在のものであり、予告なく変更または撤回される場合があります。また、本リリースに掲載された予測や将来の見通しなどに関する情報は不確実なものであり、実際に生じる結果と異なる場合がありますので、予めご了承ください。

■本件に関するお問い合わせ先

TEL：03-3512-1400 FAX：03-3512-1401 Email：pr@worksap.co.jp

株式会社ワークスアプリケーションズ 広報担当