

Joshua Tanner

NLP エンジニア

職務経歴

NLP Lead

開発チーム

Mantra 株式会社

2021 年 12 月 - 現在

- 複数のプロジェクトにおいて、NLP の研究開発をリード（エンジニア 5 名）：
 - LLM を漫画翻訳に応用し、用語やストーリーを考慮した 3 万話以上の翻訳を達成
 - ノイズの多い OCR データを学習データから取り除き、単言語およびバイリンガルの LLM ファインチューニングを実施
 - LangChain を用いた小説翻訳の LLM パイプラインの構築により、150 冊以上の小説の翻訳を部分的に自動化
 - Mantra の翻訳技術を複数の数十億ドル規模の企業に統合
 - Fairseq を用いて多言語かつ用語認識可能な翻訳モデルを訓練し、複数の言語での BLEU スコアを 10 点以上向上
 - プロの翻訳者と協力し、評価方法を改善
- Django と AWS を使用してアプリケーションのバックエンドを開発・保守
- 看板商品モバイルアプリケーションを Flutter で一から書き直し

NLP コンサルタント

フリーランス

数社

2020 年 1 月 - 現在

- 翻訳モデルが提案する単語から繰り返し選択して翻訳を構築できるインタラクティブな翻訳システムの開発をリード
- Google の「CANINE」論文の超効率トランスフォーマーを PyTorch で実装し、事前学習されたモデルを公開[\[link\]](#)
- spaCy を用いた情報抽出のパイプラインを書き直し、F 値のマイクロ平均を 20 点以上改善

研究実習生

生産技術研究所

東京大学

2020 年 12 月 - 2021 年 12 月

- 説明可能な検索拡張文法誤り訂正の手法の研究

ソフトウェアエンジニア

新商品開発部 & グローバル成長部

eBay

2016 年 11 月 - 2019 年 9 月

- eBay のチャットボットと中国版 iPhone アプリの自発的なメッセージを送信する Scala の分散スケジューラとその自動テストの開発を率い、1 人で保守
- Python で spaCy を利用し、商品レビューから有用な名詞と形容詞の組み合わせを抽出するブートストラップ自然言語処理を実装
- Ruby on Rails を利用し、非技術系社員が顧客へのメッセージをカスタマイズし送信をスケジュールできるウェブアプリケーションをゼロから開発

ソフトウェアエンジニア

国際配送管理部

eBay

2015 年 9 月 - 2016 年 11 月

- 提携先の運送会社からデータをインポートするパイプラインを再構築し、eBay の検索結果ページに載る国際配送料の精度を最大で 40% 改善
- Java と Spring で大規模（毎時数百万のリクエスト）なウェブサービスにおける保守と新機能の開発をし、JUnit と Mockito を利用してその自動テストも開発
- インフラやデータ整理などの手作業を自動化する取り組みを率い、国際出荷管理部チームの 1 週間当たりの手作業時間を数十時間減少させる事に成功

インターン

国際データインフラ部

eBay

2013 年 6 月 - 2014 年 9 月

- Hadoop での権限変更や運用タスクを自動化する Ruby on Rails ウェブアプリケーションの担当者となり、様々な新機能を開発

研究

Project MOSLA: Recording Every Moment of Second Language Acquisition
M. Hagiwara, J Tanner
LREC-COLING 2024 (forthcoming) [\[link\]](#)

Utilizing Longer Context than Speech Bubbles in Automated Manga Translation
H Kaino, S Sugihara, T Kajiwar, T Ninomiya, J Tanner and S Ishiwatari
LREC-COLING 2024 (forthcoming)

MWE as WSD: Solving Multiword Expression Identification with Word Sense Disambiguation
J Tanner, J Hoffman

EMNLP Findings+Poster 2023 [\[link\]](#)

GrammarTagger: A Multilingual, Minimally-Supervised Grammar Profiler for Language Education
M Hagiwara, J Tanner, K Sakaguchi

arxiv preprint 2021 [\[link\]](#)

学歴

計算言語学の理学修士

ワシントン大学 / 2021 年

計算機科学の理学士

エバークリーン州立大学 / 2015 年

交換留学

兵庫県立大学 / 2015 年

個人開発

MWE Dictionary

2023 年

Pytorch Lightning で訓練されたモデルと gRPC バックエンドを使用し、言語学習者向けの複単語表現を検出する辞書アプリを開発

アフィリエイトリンク生成器

2016 年

関連性の高い n-グラムをアフィリエイトリンクに変換するハッカソンプロジェクトが社内イノベーション博覧会での発表に選ばれた (採択率 14%)

Parasite

2009 年

一週間当たりにプレイされた回数が 2 百万回に達した StarCraft 2 の MOD を開発し、その後維持するためにもう 1 人の開発者と 2 人のコミュニティ管理者でチームを結成

言語

英語

母語

日本語

日本語能力試験 N1