

TOHOKU
UNIVERSITY

英語論文執筆のための 統合ライティング支援環境

伊藤拓海^{1,2}栗林 樹生^{1,2}萩原 正人³鈴木 潤^{1,4}乾 健太郎^{1,4}¹東北大学²Langsmith株式会社³Octanove Labs.⁴理化学研究所

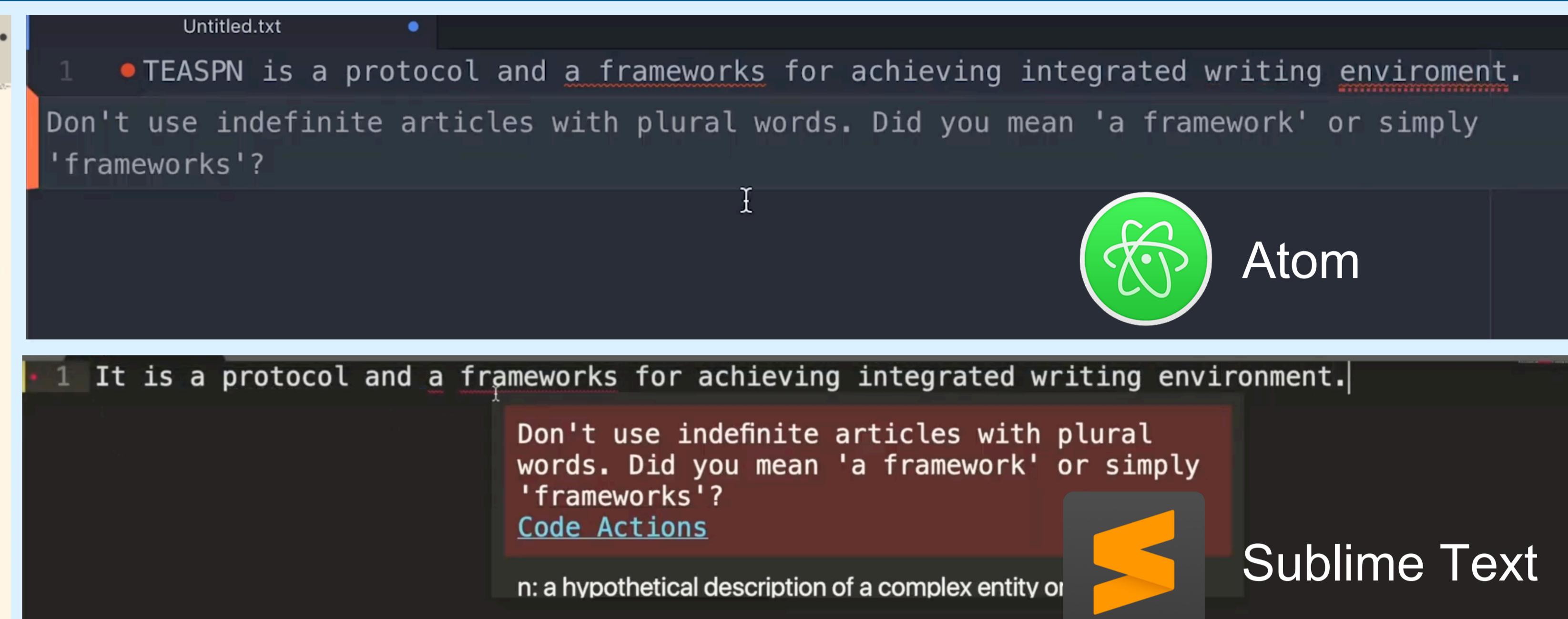
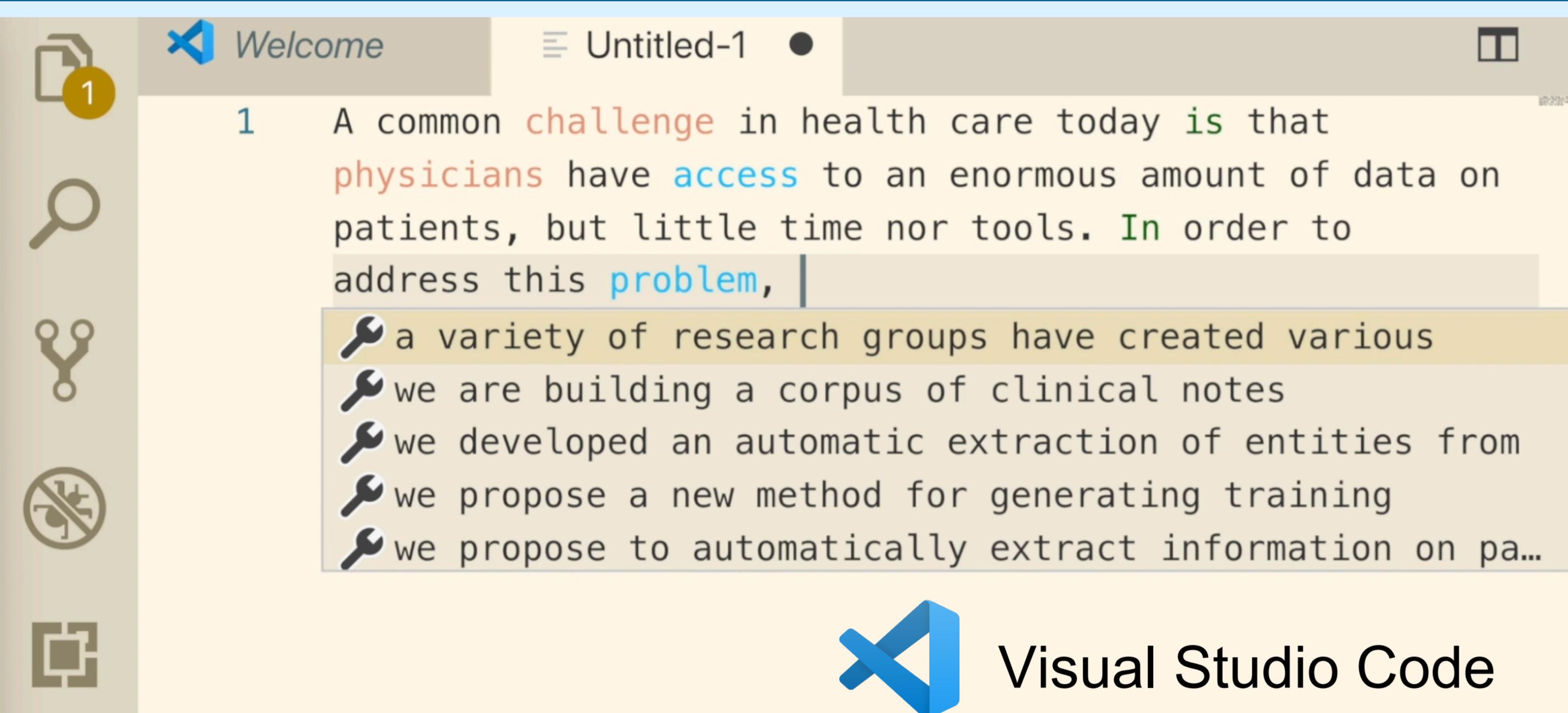
概要: 論文執筆用の自然言語エディタを作る

- 論文執筆に役立つ様々なモジュールを実装
- ライティング支援のためのプロトコルTEASPNを用いて、英語論文執筆支援環境を実現

背景

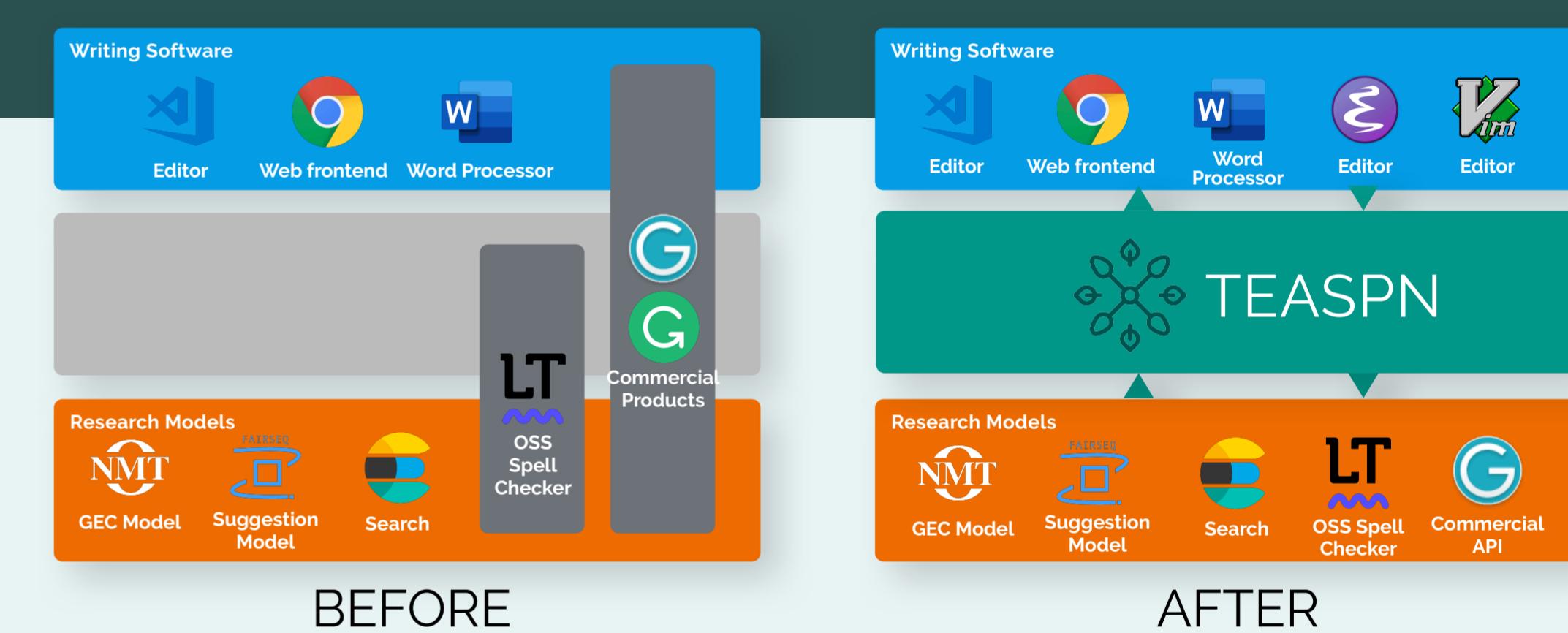
- 研究成果を世界に向けて発信するためには英語による論文執筆が必須
- 執筆支援に応用できる要素技術・リソースは多く存在するにも関わらず、執筆者(ユーザ)が使いやすい形で提供されていない

執筆支援環境の提供: 様々な要素技術を使い慣れたエディタで



TEASPN: 要素技術と執筆ソフトウェア間の効率的な橋渡し

- Language server protocolを自然言語用に拡張
- Document, SDK, デモ動画を公開済み
(EMNLP2019 system demo track accepted)



機能: ライティング支援に応用可能な要素技術

- アカデミックライティングに適した
言い換え提案/訂正エンジン

多様な言い換えを提案

入力

The global modeling using the reinforcement learning in all documents is our work in the future.

出力

*Global modeling using reinforcement learning for all documents is our future work.
We plan to extend our work to global modeling using reinforcement learning for all documents.
In future work, we plan to apply global modeling using reinforcement learning to all documents.*

*A more global model that uses reinforcement learning for all documents is our future work.
The global modeling using reinforcement learning for all documents is our future works.*

適切な単語、フレーズによる穴埋め

入力

Our parser () accuracy 82.5%, 14.7% improvement (*) state-of-the-art.*

出力

*Our parser achieves an accuracy of 82.5%, a 14.7% improvement over the state-of-the-art.
Our parser achieves a 82.5% accuracy, 14.7% improvement over the state-of-the-art.*

約20Mの擬似パラレルデータと
クラウドソーシングで集めた日本人が書いた
草文・リファレンスペア20Kを利用 [伊藤ら, 2019]

- 言語モデルによる補完

*We propose |
a method for identifying ...
a new translation system that ...*

GPT-2をNLPドメインでチューニング

- 文脈を考慮した引用推薦

We extend the LSTM-minus model _cite_

推薦
(Wang and Chang, 2016) Graph-based Dependency Parsing with Bidirectional LSTM

- アカデミックライティング
に適した単語・フレーズの提示

We suggest the ...

*propose
demonstrate
point out*

論文で表れにくいフレーズから
表れやすいフレーズへの言い換え

- 文法誤り検出・訂正

They proposes the ...

Fix: propose

- 辞書

カーソルを当てて辞書検索

- ハイライト

構文解析器を用いた、
シンタックスハイライト等

- 共参照解析

他人が書いた文章を読むときなど
(リーダーとして)役立つ可能性

- 用例・KWIC検索

コマンドメニューから (F1 or ⌘↑P)

- PoEC [高松ら, 2012] の検索

TEASPN: PoEC search



[adj] idea

"main idea"を含む用例: 274 件
• The main idea behind this heuristic is to find...

- 一般的な英語の例文検索*

TEASPN: Tatoeba search

保守的

* <https://tatoeba.org/jpn/>
のデータを使用

- 論文検索

TEASPN: Citation search

NLI alignment

今後の展望と課題

- システムの公開

- 要素技術の精度向上・高速化

- 自然言語執筆に特化した
エディタの開発