

# AI を用いたセマンティック検索システムに関する研究

AI を用いた自然言語処理による文書データからの情報抽出技術の研究（令和 3～4 年度）

産業システム部 ○全 慶樹、近藤正一 開発推進部 堀 武司

## 1. はじめに

近年、テキストデータを対象とした新しい AI 技術が提案され、書類やメールなどのテキストデータを扱う業務の効率化が期待されている。例えば、問い合わせ対応を行うヘルプデスク業務の自動化には、AI を用いたセマンティック検索と呼ばれる検索技術が活用できると考えられている。この技術は、検索キーワード「スマホ」に対して「携帯電話」や「smartphone」といった類義語や同義語の検索や、キーワードではなく文章の意味による検索を可能とする。しかし、日本語を対象とした研究事例はまだ少なく、道内における活用は進んでいない状況にある。

そこで本研究では、企業や自治体の業務効率化支援に向けて日本語を対象とした AI によるセマンティック検索が可能な文書検索システムの開発を行った。

## 2. セマンティック検索 AI の開発

単語や文章の意味に基づく検索を可能とするアルゴリズムを用いて、日本語を対象としたセマンティック検索 AI を開発した。概要を図 1 に示す。このアルゴリズムでは、汎用言語モデルに追加の学習を行うことで、文章の意味的な類似性を比較できる埋め込み特化モデルを作成する。今回、日本語モデルを学習するためのデータセットとして公開されている JSNLI、JSICK、JGLUE (JNLI) の 3 つを選定し、学習を行った。

開発した AI と他機関から公開されている同様の日本語対応 AI との性能比較を図 2 に示す。評価用データには北海道庁で使用されている FAQ データセットを、検索性能には FAQ を検索した際の順位の高さに基づく指標(MRR)を用い、開発した AI が高い性能を示すことを確認した。



図 1 日本語を対象としたセマンティック検索 AI の開発

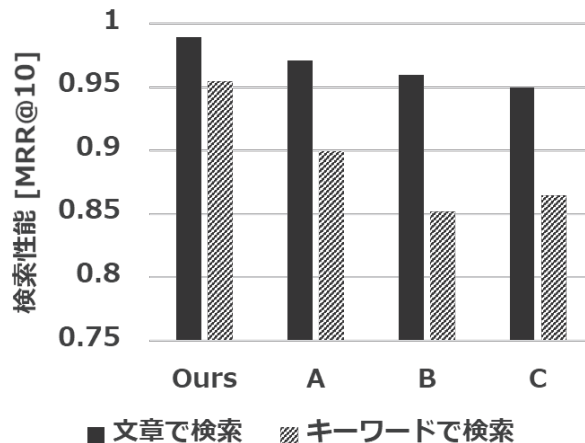


図 2 開発した AI と他機関の AI の性能比較 (A, B, C: Sentence Transformers 公式モデル)

## 3. AI を用いた文書検索システムの開発

ウェブブラウザによるセマンティック検索の利用を可能とするため、AI を用いた文書検索システムを開発した。システムの概要を図 3 に示す。サーバーへ前述の FAQ データセットを検索対象としたシステムを実装し、標準的な CPU で毎秒約 7 回の問い合わせに応答できることを確認した。

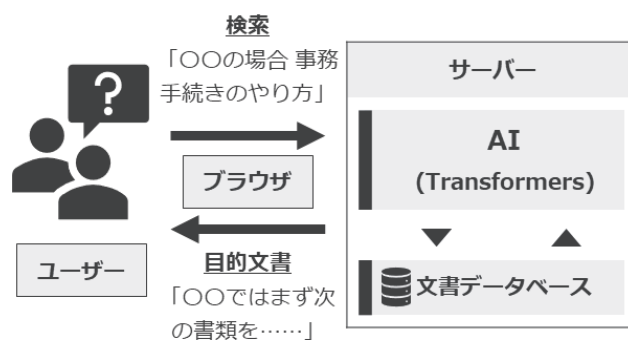


図 3 AI を用いた文書検索システム

## 4. おわりに

本研究では、日本語を対象とした高性能なセマンティック検索 AI を開発し、ユーザーの実利用を想定した事例開発として文書検索システムの開発に取り組んだ。今後、AI の検索性能や検索システムの操作性の改善に取り組むとともに道内企業等への技術移転を進める。

(連絡先: zen-keiki@hro.or.jp)