

特許文献検索キーワードアシスト機能にむけた 単語分散表現による類義語辞書作成手法の検討

株式会社 日立製作所

近藤 明宏

網羅的な特許検索における課題

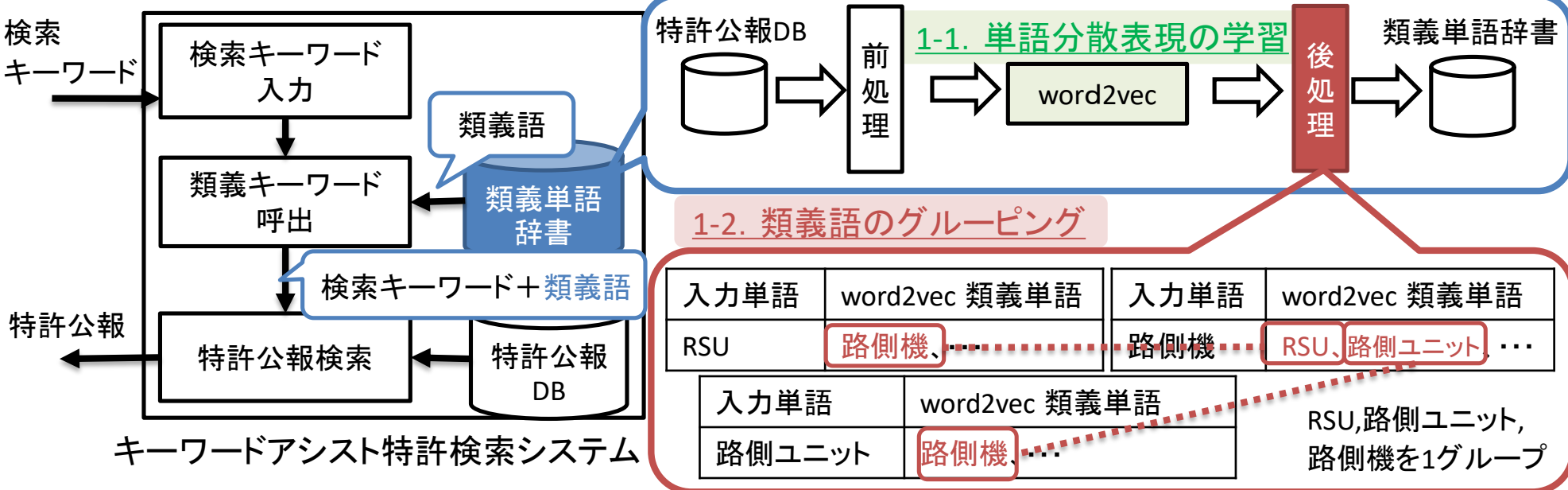
- ・特許文献検索時には、類義語を含めた適切な検索キーワードにて検索する必要がある。
例) 路側機, 路側装置, RSU, 路側ユニットなど
- ・適切なキーワードが含まれない場合には、検索漏れが生じる。この類義語の知識が文書化され共有されていることは少ない。この知識を人手でドキュメント化し続けるのも、多大な工数を必要とする。

分散表現による類義語辞書作成

- 検索キーワード入力をアシストするシステムを検討
1. 公開特許公報から、類義語辞書を自動で生成
 - 1-1. word2vecを用いて特許公報から単語の分散表現を学習
 - 1-2. word2vecの類義語抽出機能を用いて、分散表現空間上にて、相互に類義語が含まれる単語群をひとつの類義語としてグループ化

単語分散表現による類義語辞書作成手法

○ 手法概要



結果例※

結果の一例を右表に示す。

○ 「RSU」, 「路側機」, 「路側装置」, 「路側器」, 「路上装置」が相互に類義語として学習でき、類義語としてグルーピングできた

× 「路側センサ」, 「路側ユニット」は一方向の類似性しか確認されず、類義語セットに含められなかった

入力単語	word2vec Top15類義単語
○ RSU	路側機、路側ユニット、...
○ 路側機	路側装置、路側器、路上装置、RSU、...
○ 路側装置	路側機、路上装置、路側器、...
○ 路側器	路側機、路上装置、路側装置、...
○ 路上装置	路側機、路側器、路側装置、...

相互に類義語 一方に類義語 類義語以外

入力単語	word2vec Top15類義単語
× 路側センサ	路側装置、路上装置、路側機、...
× 路側ユニット	ランキングモジュール、修正モジュール、集中状態推定部、車載支援装置、ネットワークアクセステバイス、最低コスト経路モジュール、ローカルハブ、通報要因領域、電氣的接続部品、...

※キーワードに「車両」を指定した16,621件の特許公報から185,594単語が学習できた。その結果から特徴的なものを紹介