

発明の名称

専門用語抽出装置、専門用語抽出方法及びプログラム

～文書中から網羅的かつ精度よく専門用語を抽出する技術～

発明者

プタシンスキ ミハウ エドモンド（北見工業大学）

技術概要

文書のテキストデータからNグラム（任意の文字列を連続したN個の文字、単語、チャンク等で分割する手法）の文節を抽出後、文節内で助詞の前にある部分全体を専門用語の候補語として判別し、その候補語の出現頻度をカウントするとともに、候補語の出現頻度と候補語が判別されたNグラムの長さに基づいて候補語の「重み」を算出し、その重みが閾値以上である候補語を専門用語として抽出する技術です。これにより、文書中から網羅的に精度よく専門用語を抽出できます。

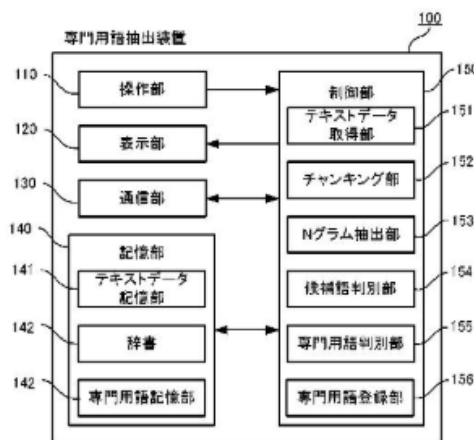
発明の目的

文書中のテキストデータから専門用語を抽出する精度を向上させた、専門用語抽出装置、抽出方法、プログラムを提供することを目的としています。

効果

専門分野に関する文書を入手できれば、専門分野に関する事前知識を必要とせずに任意の専門分野に関する専門用語を抽出できます。また、認識された候補語が専門用語であるかどうかを精度よく判別できます。

図面



詳細はこちら

[工業所有権情報・研修館のサイトに移動します](#)

出願・登録

出願日	設定登録日	存続期間満了日
2020/11/09	2024/09/19	2040/11/09