

ブラウザ上でユーザが編集可能な言語パターンマッチシステムの構築 Building a user-editable language pattern matching system in the browser

桂 辰弥¹⁾ 竹内 孔一¹⁾
Tatsuya Katsura Koichi Takeuchi

概要

テキスト中の特定の表現を見つけることは、言語および教育分野において必要となることがある。例えば語学学習において、言語を検索するのに用いるコンコーダンスがある。本研究ではユーザ自身が求める表現を検索ブロックで組み合わせてシステムに投入し、事例を検索できるシステムの開発を行っている。先行研究において JavaScript と Python を利用した基本システムを構築したが、システムの本格利用にはいくつかの課題が残されている。そこで本報告では検索エンジンの中心部分である Prolog データベースの実装の改良、および、大規模なテキストが扱えるためにデータベースをシステムに導入したので、この改良について報告する。

1 はじめに

テキスト中の特定のフレーズや表現を見つけることは、言語および教育分野において必要となることがある。テキストデータから特定のキーワードやフレーズの出現位置や文脈を抽出するためのプログラムとしてコンコーダンスがある。コンコーダンスは語学学習において特定のフレーズや表現の使用例を実際の文脈で把握することで、語彙や文法の理解、単語の使用法や文脈の把握に役立ち、学習者の語彙や表現力の向上に役立つ。例えばではが提案している。

パターンマッチングはテキストの表層で検索を行う正規表現とは異なり、情報を抽出したい文を対象に予め関係する文や文の一部に対応する文構造のパターンを用意し、そのパターンに合致する結果を取得する。本研究ではユーザ自身が求める表現を検索ブロックで組み合わせてシステムに投入し、事例を検索できるシステムの開発を行っている。

先行研究において WEB アプリケーションとして

1) 岡山大学

JavaScript と Python を利用した基本システムを構築したが、システムの本格利用にはいくつかの課題が残されている。そこで本報告では検索エンジンの中心部分である Prolog データベースの実装の改良、および、大規模なテキストが扱えるためにデータベースをシステムに導入したので、この改良について報告する。本研究の評価実験を行い、本システムでは 10000 文程度の比較的大きなコーパスを扱えることを確認できた。

2 提案する言語パターンマッチシステム

2.1 言語パターンマッチシステムの概要

3 評価実験

謝辞

謝辞の文章を `\acknowledgment` で指定します。使わなければ謝辞は出力されません。

/

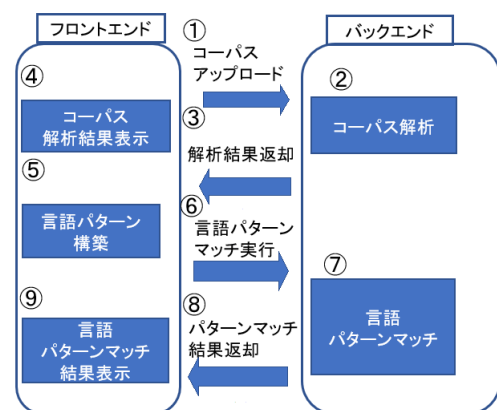


図1 システムの構成図