

# 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<sup>1)</sup> 竹内 孔一<sup>1)</sup>  
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。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ものである。有名なコンコーダンスの例として、Sketch Engine<sup>1)</sup>がある。Sketch Engine<sup>1)</sup>はクエリ言語として CQL (Corpus Query Language)<sup>2)</sup>が使用されており、コーパス内で正規表現や演算子を組み合わせることでパターンマッチを行うことができる。しかし Sketch Engine は、コーパスベースの言語分析や統計的な情報抽出が主な役割であるため、直接的に依存関係解析の直接的な機能を提供していない。

ユーザ自らがこれらを考慮してテキスト中の特定フレーズや表現を抽出するようなシステムを構築することは容易ではない。そこで本研究では意味役割付与システムで解析した結果をユーザ自身が求める表現をあらかじめ用意された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 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

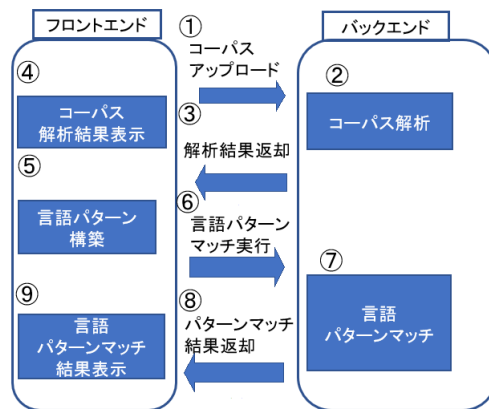


図1 システムの構成図

## 2.1 言語パターンマッチシステムの概要

表1 Prolog の述語一覧

Predicates	第2引数	第3引数
chunk(SENTENCE_ID, 0, _)	根ノード ID	文節 ID
morph(SENTENCE_ID, _, _)	文節 ID	形態素 ID
main(SENTENCE_ID, _, _)	文節 ID	主形態素
part(SENTENCE_ID, _, _)	文節 ID	副形態素
role(SENTENCE_ID, _, _)	文節 ID	意味役割
semantic(SENTENCE_ID, _, _)	文節 ID	概念
surf(SENTENCE_ID, _, _)	ノード ID	ノードの表層
surfBF(SENTENCE_ID, _, _)	形態素 ID	形態素の基本形
sloc(SENTENCE_ID, _, _)	文節/形態素 ID	出現位置
pos(SENTENCE_ID, _, _)	形態素 ID	品詞
dep(SENTENCE_ID, _, _)	文節 ID	文節 ID

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。

## 2 提案する言語パターンマッチシステム

### 1) 岡山大学

1) <https://www.sketchengine.eu/>

2) <https://www.sketchengine.eu/documentation/corpus-querying/>

## 3 評価実験

### 3.1 実験内容

### 3.2 結果

## 4 考察と結論

### 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。