**KHOA CÔNG NGHỆ THÔNG TIN**

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH**

Thực hiện:

20C11035 – Trương Thế Kiệt

20C11040 – Đặng Nhật Minh

BÁO CÁO ĐỒ ÁN CÀI ĐẶT ỨNG DỤNG

NĂM HỌC 2020-2021

**Xử lý ngôn ngữ tự nhiên nâng cao**

**ĐẠI HỌC QUỐC GIA TPHCM**

**BẢNG THÔNG TIN CHI TIẾT NHÓM**

|  |  |  |  |
| --- | --- | --- | --- |
| **Mã đề tài:** | **UD07** | | |
| **Số lượng:** | **2** | | |
| **MSSV** | **Họ tên** | **Email** | **Điện thoại** |
| 20C11035 | Trương Thế Kiệt | truongthekiet709@gmail.com |  |
| 20C11040 | Đặng Nhật Minh | minhdangnhat685@gmail.com |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **BẢNG PHÂN CÔNG & ĐÁNH GIÁ HOÀN THÀNH CÔNG VIỆC** | | | |
| **Người thực hiện** | **Công việc thực hiện** | **Mức độ hoàn thành** | **Đánh giá của nhóm** |
| 20C11035  Trương Thế Kiệt | Tìm hiểu các thuật toán có thể sử dụng được | 80% | 8/10 |
| Tìm hiểu và cài đặt thuật toán và web cho chương trình |
| Viết báo cáo |
| Quay demo |
| 20C11040  Đặng Nhật Minh | Tìm hiểu các thuật toán có thể sử dụng được | 80% | 8/10 |
| Tìm hiểu và cài đặt thuật toán và web cho chương trình |
| Viết báo cáo |
| Quay demo |

MỤC LỤC

[I. TỔNG QUAN 2](#_Toc76582244)

[1. Mô tả bài toán 2](#_Toc76582245)

[2. Các thư viện sử dụng trong project 2](#_Toc76582246)

[II. THUẬT TOÁN SỬ DỤNG 3](#_Toc76582247)

[1. Giới thiệu các thuật toán hiện có 3](#_Toc76582248)

[2. Giới thiệu về thuật toán Earley 3](#_Toc76582249)

[III. CHI TIẾT CÀI ĐẶT 4](#_Toc76582250)

[1. Giới thiệu code và chức năng từng class 4](#_Toc76582251)

[IV. CHẠY CHƯƠNG TRÌNH 5](#_Toc76582252)

[1. Các bước cần thực hiện 5](#_Toc76582253)

[2. Kết quả 5](#_Toc76582254)

[VI. KẾT LUẬN 7](#_Toc76582255)

[TÀI LIỆU THAM KHẢO 8](#_Toc76582256)

# TỔNG QUAN

## Mô tả bài toán

* Dựa vào kiến thức đã học xây dựng một ứng dụng giúp phân tích cú pháp của ngôn ngữ tiếng anh. Việc phân tích dựa vào các dữ liệu được nhập vào như bộ luật và bộ tự vựng của ngôn ngữ đó. Phân tích cú pháp là việc chia nhỏ một câu câu thành các thành phần của nó để câu có thể hiểu được. Việc phấn tích cú pháp sẽ mang lại nhiều lợi ích trong việc xây dựng một trình biên dịch, nó giúp xác định một giá trị đầu vào có đúng với cú pháp của ngôn ngữ đang xét hay không.

## Các thư viện sử dụng trong project

|  |  |
| --- | --- |
| **Tên thư viện** | **Mục đích sử dụng** |
| Spring Boot | Xây dựn ứng dụng web |
| Library | Giúp trực các từ sau khi đã phần tích thành câu cú pháp |
|  |  |

# THUẬT TOÁN SỬ DỤNG

## Giới thiệu các thuật toán hiện có

* Có nhiểu thuật toán giúp phân tích cú pháp của một câu không dựa vào ngữ cảnh. Kĩ thuật phân tích cấu trúc được chia làm loại chính:
  + Top Down Parsing
  + Bottom Up Parsing
* Top Down Parsing: cây cú pháp được xây dựng bắt đầu từ gốc(root) sau đó đi dần tới các node lá. Một số thuật toán sử dụng kĩ thuật này: LL Parser, **Earley Parser**, Pratt Parser, Recursive Descent Parser, Parser Combinator
* Bottom Up Parsing: Cây cú pháp được xây dựng dựa từ các node lá tiến dần về các node gốc. Một sô thuật toán sử dụng kĩ thuật này: CYK Parser, LR Parser, Generalized LR Parser

## Giới thiệu về thuật toán Earley

* Earley Parser là một thuật toán được sử

# CHI TIẾT CÀI ĐẶT

## Giới thiệu code và chức năng từng class

# CHẠY CHƯƠNG TRÌNH

## Các bước cần thực hiện

* Để tiến hành chạy được chương trình cần chỉnh sửa một số cấu hình trong file **application.yml**

server:

port: 7080 #port run BE API

host: http://localhost # host hiện tại đang chạy back end API

grammar:

is-read-default: false #Đọc giá trị grammar đã được đóng gói trong jar hay không

grammar-file-path: ./Grammar.txt #khi giá trị field is-read-default là true thí không cần set field này.

* Một ví dụ cho giá trị của file **Grammar.txt**:

S>NP VP

VP>V NP

NP>Conj NP

NP>N NP

NP>N

NP>Pron

NP>N Conj VP

* Thư mục hiện tại sẽ bao gồm các file:

application.yml

grammar.parser-0.0.1-SNAPSHOT.jar

Grammar.txt

* Thực hiện mở terminal và chạy dòng lệnh sau:

**java -jar grammar.parser-0.0.1-SNAPSHOT.jar**

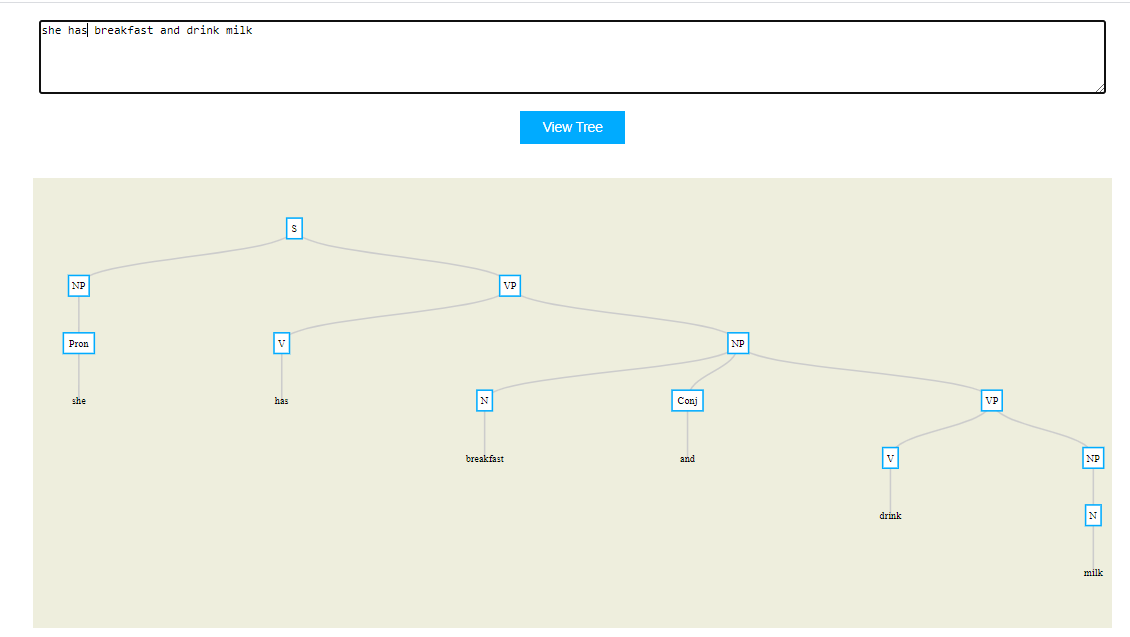
* Sau khi đã start thành công mở browser và truy cập vào đường dẫn: localhost:7080 để sử dụng app.

## Kết quả

* Kết quả sau khi start chương trình thành công với config như trên, và nhập giá trị text input là :

**she has breakfast and drink milk**

* Kết quả nhận được là:



# KẾT LUẬN

* Với mục tiêu tìm hiểu và cài đặt thuật toán để tiến hành phân tích cấu trúc của một câu dựa vào các luật đã nhập vào nhóm đã hoàn thành được.
* Đã cài đặt được một chương trình tương đối hoàn chỉnh có thể sử dụng được.

# TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | https://www.guru99.com/syntax-analysis-parsing-types.html |
| [2] | https://en.wikipedia.org/wiki/Earley\_parser |
| [3] | https://tomassetti.me/guide-parsing-algorithms-terminology/ |
| [4] | "NN-SVG," [Online]. Available: http://alexlenail.me/NN-SVG/LeNet.html. |
| [5] | S. I. M. K. M. &. T. S. Saha, "An Efficient Traffic Sign Recognition Approach Using a Novel Deep Neural Network Selection Architecture," 2019. |
| [6] | V. H. Tiệp, "Bài 36. Giới thiệu về Keras," machinelearningcoban.com, [Online]. Available: https://machinelearningcoban.com/2018/07/06/deeplearning/. |