

XÂY DỰNG SCANNER – PARSER CHO SIMPLECODE

Qui định

- Hạn nộp: xem trên trang web môn học
- Dạng bài tập: nhóm.
- Ngôn ngữ: Java
- Cấu trúc bài nộp:

<MSSV>

```

__ libs          // Thư viện hỗ trợ
__ Bao_cao       //chứa tập tin báo cáo (.pdf / .doc / .docx)
__ Source_code   // code của chương trình

```

Nén theo định dạng .rar hoặc .zip và nộp vào link tương ứng.

Yêu cầu

Sử dụng thư viện hỗ trợ, ANTLR, để xây dựng bộ phân tích từ vựng (scanner) và cú pháp (parser) cho ngôn ngữ SimpleCode (xem đặc tả trong tập tin đính kèm).

Chương trình chạy với tham số dòng lệnh như sau:

```
java -jar dist/<tên_chuong_trinh>.jar -target <option> -debug <tập tin >
```

Trong đó,

- **<option>**: nhận một trong 2 giá trị 0 hoặc 1. Nếu là 0, chương trình thực hiện phân tích từ vựng. Nếu là 1, thực hiện phân tích cú pháp.
- **<tập tin>**: đường dẫn tập tin chứa đoạn code được viết bằng ngôn ngữ SimpleCode.
- Chương trình xuất kết quả phân tích ra tập tin. Tập tin này có tên trùng với tên của tập tin đầu vào, nhưng phần mở rộng là .out

Định dạng của tập tin đầu vào

Tập tin văn bản bình thường có thể chứa đoạn mã viết bằng ngôn ngữ SimpleCode hoặc không. Sinh viên xem thêm trong tập dữ liệu cung cấp.

Định dạng của tập tin đầu ra

Phân tích từ vựng

Liệt kê các token phân tích được từ tập tin output, mỗi token nằm trên một dòng. Mỗi dòng liệt kê thông tin sau đây:

- Số thứ tự dòng mà token xuất hiện (bắt đầu từ 1)
- Loại token (type): chỉ liệt kê cho các loại token sau (theo đúng định dạng đã ghi): CHARLITERAL, INTLITERAL, BOOLEANLITERAL, STRINGLITERAL và IDENTIFIER. Các trường hợp còn lại không ghi
- Nội dung cụ thể của token(text): trong trường hợp token có loại là STRINGLITERAL và CHARLITERAL thì lần lượt phải đặt trong dấu nháy kép và dấu nháy đơn

Trong trường hợp bị lỗi khi phân tích token nào thì thực hiện in lỗi tương ứng ở dòng đó, theo định dạng sau:

<tên tập tin đầu vào> line <chỉ số dòng xuất hiện> : <chỉ số cột xuất hiện> : <nội dung lỗi>

Ví dụ tập tin đầu ra cho đoạn code print (“Hello, World!”);

```
1 IDENTIFIER print
1 (
1 STRINGLITERAL “Hello, World!”
1 )
1 ;
```

Phân tích cú pháp

- Trong trường hợp, đoạn chương trình trong tập tin đầu vào là hợp lệ thì tập tin đầu ra là rỗng (không ghi gì)

- Trong trường hợp, đoạn chương trình trong tập tin đầu vào không hợp lệ thì tập tin đầu ra liệt kê các lỗi đã xảy ra. Mỗi lỗi nằm trên một dòng. Sinh viên có thể xuất thông điệp lỗi theo định dạng do cá nhân tự định nghĩa.