شرح پروژه درس اصول طراحی کامپایلر

دانشگاه فردوسی مشهد – آبان‌ماه ۱۴۰۴

عنوان پروژه: پیاده‌سازی تحلیل‌گر لغوی (Lexical Analyzer) و جدول نماد (Symbol Table) برای زبان --Java

وزن پروژه: ۳ نمره (از ۷ نمره پروژه‌های درس)

مهلت تحویل: ۱۴ آبان‌ماه

هدف این پروژه، پیاده‌سازی یک تحلیل‌گر لغوی است که بتواند توکن‌های مختلف را از کدهای نوشته‌شده به زبان برنامه‌نویسی --Java شناسایی کرده و جدول نماد را برای شناسه‌ها ایجاد کند.

ورودی برنامه: یک فایل حاوی کد منبع به زبان --Java (مطابق گرامر ارائه شده)

خروجی مورد انتظار:

- **بخش اول:** استخراج توکن‌ها

برنامه باید بتواند انواع توکن‌های زیر را شناسایی و دسته‌بندی کند:

۱. شناسه‌ها (Identifiers):

- نام کلاس‌ها، متغیرها، متدها، اینترفیس‌ها

- مثال: "calculateSum"، "myVariable" و "Main"

۲. عملگرها (Operators):

- عملگرهای محاسباتی: `+` (جمع)، `-` (تفریق)، `*` (ضرب)، `/` (تقسیم)، `٪` (باقیمانده)، `**` (توان)

- عملگرهای رابطه‌ای: `==` (برابری)، `!=` (عدم برابری)، `<` (کمتر از)، `<=` (کمتر یا مساوی)، `>` (بیشتر از)، `>=` (بیشتر یا مساوی)

- عملگرهای منطقی: `&&` (AND)، `||` (OR)، `!` (NOT)

- عملگرهای انتساب و دسترسی: `=` (انتساب)، `.length` (طول آرایه)

۳. کلمات کلیدی (Keywords):

- کلمات کلیدی اصلی: class - interface - extends - implements - public - private - protected

internal - static - void - abstract - if - else - while - for - break
continue - return - new - this - import - true - false – null

- نواع داده: boolean - int - char - String

۴. جداکننده‌ها (Delimiters): `;` (نقطه ویرگول)، `,` (کاما)، `{` `}` (آکولاد)، `[` `]` (براکت)، `(` `)` (پرانتز)، `.` (نقطه)

۵. اعداد (Numbers):

- اعداد صحیح (IntegerLiteral)

- مثال: `۰`, `۱۲۳`, `۴۵_۶۷۸`

۶. مقادیر ثابت:

- StringLiteral: رشته‌ها درون دابل کوتیشن

- CharLiteral: کاراکترها درون سینگل کوتیشن

- BooleanLiteral: true, false

- NullLiteral: null

- **بخش دوم:** ساخت جدول نماد (Symbol Table)

در این بخش، می‌بایست یک ساختار داده‌ای برای نگهداری اطلاعات شناسه‌ها (Identifiers) پیاده‌سازی شود. توجه شود که ساخت جدول نماد باید با در نظرگرفتن حوزه‌های (Scope) مختلف در زبان JavaMinusMinus2 صورت گیرد.

انواع Scopeها در زبان JavaMinusMinus2:

- حوزه global (برای importها، کلاس‌ها و اینترفیس‌ها)

- حوزه کلاس (برای فیلدها، متدها و سازنده‌ها)

- حوزه متد (برای پارامترها و متغیرهای محلی)

- حوزه بلوک (برای متغیرهای داخل بلوک)

هر یک از حوزه‌های فوق دارای یک جدول نماد مجزا می‌باشد که اطلاعات مربوط به شناسه‌های تعریف‌شده در آن حوزه را ذخیره می‌کند. با توجه به این‌که ساختار حوزه‌های مختلف زبان بصورت تودرتو و درختی است، ساختار جداول نماد نیز به همین صورت خواهد بود.

اطلاعاتی که در رابطه با هر نوع شناسه، باید در جدول نماد ذخیره شود:

۱. کلاس‌ها (Classes): نوع شناسه = کلاس، نام کلاس، نام والد (در صورت وجود)، نام اینترفیسی که توسط کلاس پیاده‌سازی شده‌است (در صورت وجود)، آیا کلاس abstract است؟

۲. اینترفیس‌ها (Interfaces): نوع شناسه = اینترفیس، نام اینترفیس.

۳. متدها (Methods): نوع شناسه = متد، نام متد، نوع بازگشتی، لیست پارامترها (نام و نوع)، سطح دسترسی، آیا override است؟، آیا abstract است؟

۴. سازنده‌ها (Constructors): نوع شناسه = سازنده، نام سازنده، لیست پارامترها (نام و نوع).

۵. متغیرها و فیلدها (Variables/Fields): نوع شناسه = متغیر/فیلد، نام متغیر یا فیلد، نوع داده، حوزه تعریف (کلاس، متد، بلوک)، سطح دسترسی، مقدار اولیه (در صورت وجود).

۶. پارامترها (Parameters): نوع شناسه = پارامتر، نام پارامتر، نوع پارامتر.

ساختار پایه جدول نماد:

پیشنهاد می‌شود جدول نماد به صورت یک HashTable پیاده‌سازی شود که در آن:

- کلید (Key): نام شناسه

- مقدار (Value): مجموعه ویژگی‌های مربوط به شناسه

توابع اصلی جدول نماد:

- insert (identName, attributes): برای درج شناسه جدید با ویژگی‌هایش

- lookup (identName): برای جستجوی شناسه و بازیابی ویژگی‌هایش

- **نکات فنی:**

۱. جهت استخراج اطلاعات مرتبط با شناسه‌ها از ابزار ANTLR استفاده شود.

۲. از گرامر ارائه شده برای تعریف دقیق الگوی توکن‌ها استفاده شود.

۳. توکن‌های چندحرفی مانند `&&` و `||` به‌درستی شناسایی شوند.

۴. کامنت‌ها و فضاهای خالی نادیده گرفته شوند.

۵. در صورت مشاهده توکن‌های نامعتبر پیغام خطای مناسبی تولید شود.

۶. خروجی بصورت خوانا و فرمت‌شده ارائه شود.

- **تحویل پروژه:**

- کد منبع کامل

- فایل README با راهنمای اجرا

- نمونه‌های تست همراه با خروجی‌های مورد انتظار

- گزارش کوتاه از عملکرد بخش‌های مختلف کد

موفق باشید - خسروی