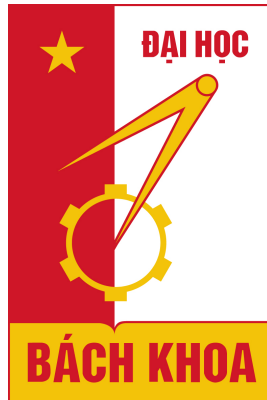


TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI



Báo cáo bài tập lớn môn học:
Lý thuyết ngôn ngữ và phương pháp dịch

**TÌM HIỂU BỘ SINH
TRÌNH PHÂN TÍCH TỪ VỰNG FLEX**

Sinh viên thực hiện:

Tạ Quang Tùng
MSSV: 20154280

Giáo viên hướng dẫn:

TS. Phạm Đăng Hải

Hà Nội, Ngày 23 tháng 10 năm 2018

Mục lục

1	Giới thiệu về bộ sinh trình phân tích từ vựng Flex	1
2	Cấu trúc của file chỉ dẫn trong Flex [1]	1
2.1	Vùng định nghĩa - Definitions Section	1
2.2	Vùng các luật - Rules Section	2
2.3	Vùng mã nguồn bổ sung - User Code Section	2
Tài liệu tham khảo		3

1 Giới thiệu về bộ sinh trình phân tích từ vựng Flex

Flex là bộ sinh chương trình phân tích từ vựng, là một phần mềm miễn phí và mã nguồn mở thay thế cho Lex. [2] Nó thường xuyên được sử dụng đi kèm với bộ sinh chương trình phân tích cú pháp Yacc (hoặc mã nguồn mở thay thế là GNU Bison) để tạo nên 2 khối chức năng cơ bản của một chương trình dịch.

Flex được viết bằng ngôn ngữ C bởi Vern Paxson vào khoảng năm 1987. Flex nhận đầu vào là một file chỉ dẫn, thông thường có đuôi là `.l`. Từ đó Flex có thể sinh ra mã C có thể biên dịch và thực thi mà không cần thêm bất kỳ thư viện ngoài nào.

Mã C sinh ra bởi Flex sử dụng Automata hữu hạn đơn định (Deterministic Finite Automation - DFA) để thực hiện việc tách xâu đầu vào thành các từ tố tương ứng. Thuật toán thường có độ phức tạp thời gian tính là $O(n)$ với n là độ dài xâu đầu vào.

Flex chỉ có thể sinh ra mã code C hoặc C++.

2 Cấu trúc của file chỉ dẫn trong Flex[1]

Mỗi file chỉ dẫn đầu vào cho Flex bao gồm 3 vùng, được phân tách nhau bởi các dòng chỉ chứa xâu `'%%'`:

```
|| definitions
|| %%
|| rules
|| %%
|| user code
```

Trong đó:

- **definitions:** Vùng chứa các định nghĩa.
- **rules:** Vùng chứa luật của các từ tố.
- **user code:** Vùng chứa mã C/C++ được thêm vào.

2.1 Vùng định nghĩa - Definitions Section

Vùng định nghĩa chứa các định nghĩa tên và các định nghĩa điều kiện bắt đầu. Định nghĩa tên có dạng:

```
|| name definition
```

'**name**' là các tên có cú pháp giống như các tên trong ngôn ngữ C. '**definition**' là phần được lấy từ kí tự không phải kí tự trắng đầu tiên sau '**name**' và đến hết dòng. '**name**' không được thụt lề đầu dòng. Các định nghĩa có thể được tham chiếu đến bằng cách sử dụng '**{name}**' và nó sẽ được chuyển thành '**(definition)**'.

Ví dụ:

```
|| DIGIT [0-9]
```

Thì '**{DIGIT}**' sẽ được chuyển thành '**[[0-9]]**'.

Bất kì một dòng nào trong vùng này được viết thụt lề đầu dòng hoặc được bao xung quanh bởi `%{` và `%}` thì sẽ được sao chép nguyên văn vào file mã nguồn sinh ra.

Ví dụ:

```
||      int number, count;  
||  %{  
||  #include <math.h>  
||  %}
```

Thì đoạn code sau sẽ được sao chép nguyên văn vào file đầu ra:

```
||  int number, count;  
||  #include <math.h>
```

2.2 Vùng các luật - Rules Section

2.3 Vùng mã nguồn bổ sung - User Code Section

Tài liệu tham khảo

- [1] Lexical Analysis With Flex, for Flex 2.6.2: Top, Oct 2016.
- [2] Flex (lexical analyser generator) - Wikipedia, Oct 2018.