

BÀI TẬP ÔN TẬP THI CUỐI KÌ

Môn: Lý thuyết ngôn ngữ hình thức và ôtômat

1 Ngôn ngữ hình thức và văn phạm

Phần bắt buộc ôn tập với tất cả các sinh viên.

1. Ôn tập các khái niệm liên quan tới ngôn ngữ và văn phạm hình thức, ôtômat.
2. Ôn lại kĩ năng xây dựng văn phạm sinh ngôn ngữ đã cho.
3. Ôn lại kĩ năng tìm ngôn ngữ biết văn phạm sinh ngôn ngữ đó.

2 Ngôn ngữ chính quy

2.1 Ôtômat hữu hạn trạng thái

1. Xây dựng otomat đơn định, hữu hạn trạng thái đoán nhận các ngôn ngữ sau trên bảng chữ cái $\{0, 1\}$:
 - (a) Tập các xâu trong đó số chữ cái 0 chia hết cho 3, số chữ cái 1 chia hết cho 5.
 - (b) Tập các xâu bắt đầu bằng 1, khi đổi xâu nhị phân đó sang số nguyên thì được một số chia hết cho 5.
 - (c) Tập các xâu trong đó số chữ cái 0 chia hết cho 2, số chữ cái 1 chia hết cho 3.
 - (d) Tập các xâu bắt đầu bằng 1, khi đổi xâu nhị phân đó sang số nguyên thì được một số chia hết cho 4.
2. Cho otomat đơn định, hữu hạn $A = (S, \Sigma, \{s_0\}, \delta, F)$ và một chữ cái $a \in \Sigma$. Giả sử $\forall s \in S$ đều có $\delta(s, a) = s$.
 - (a) Chứng minh bằng quy nạp theo n rằng $\forall n \geq 0$, ta luôn có giá trị hàm chuyển mở rộng $\delta(s, a^n) = s$.

- (b) Chứng minh rằng hoặc $\{a\}^* \subseteq L(A)$, hoặc $\{a\}^* \cap L(A) = \emptyset$.
3. Xây dựng otomat không đơn định, hữu hạn trạng thái đoán nhận các ngôn ngữ trên bảng chữ cái $\{0, 1, 2, \dots, 5\}$ sau (tận dụng tối đa lợi thế của việc xây dựng otomat không đơn định):
- (a) Tập các xâu trong đó chữ số cuối cùng đã xuất hiện ít nhất 1 lần trước đó
 - (b) Tập các xâu trong đó chữ số cuối cùng chưa hề xuất hiện trước đó

Thực hiện thuật toán đơn định hóa hai otomat trên.

2.2 Biểu thức chính quy

1. Cho biểu thức chính quy sau: $(bc \cup cb)(a^*)bccb$. Vẽ đồ thị chuyển, viết dưới dạng hình thức ôtomat và viết văn phạm chính quy tương đương với biểu thức chính quy này.
2. Cho cú pháp của biểu thức chính quy trên Unix như sau:
 - Mỗi kí tự biểu diễn chính nó, trừ các kí tự điều khiển (*metacharacter*) bao gồm các kí tự sau:
 $? + - * . \{ \} [] () \backslash | ^ \$$
 - Để biểu diễn các kí tự điều khiển, thêm dấu \backslash vào trước
 - $[03a-c] \equiv \{0, 3, a, b, c\}$, $[\wedge 15] \equiv$ tập các kí tự khác 1 và 5
 - $.$ biểu diễn kí tự bất kì, $^$ và $\$$ đánh dấu đầu và cuối dòng
 - $\backslash<$ và $\backslash>$ đánh dấu đầu và cuối từ, $\backslash b$ đánh dấu biên từ, $\backslash B$ đánh dấu xâu rỗng không ở biên từ
 - Phép lặp: $?$, $*$, $+$, $\{n\}$, $\{n, \}$, $\{n, m\}$
 - Phép lấy tích ghép, phép hợp $|$
 - $\backslash n$ biểu diễn xâu con nằm giữa cặp $()$ thứ n trước đó.

Hãy viết biểu thức chính quy và xây dựng otomat (nguồn) tương đương cho các xâu có dạng sau:

- (a) Đoạn văn bản HTML nằm giữa cặp mở đóng thẻ $<p>$ và $</p>$
- (b) Số thực dấu phẩy động
- (c) Địa chỉ Internet IPv4 (32 bits) biểu diễn bằng chuỗi số trong hệ thập phân.

- (d) Chú thích dòng trong chương trình Java: xâu nằm trên 1 dòng sau kí hiệu //
- (e) Chú thích khối trong chương trình: các xâu nằm giữa /* và */

2.3 Bổ đề bơm

1. Phát biểu và chứng minh điều kiện cần của ngôn ngữ chính quy (bổ đề bơm).
2. Chứng minh các ngôn ngữ sau không phải là ngôn ngữ chính quy:
 - (a) Tập các biểu thức số học trên tập số tự nhiên với 4 phép toán +, -, *, / và cặp dấu ngoặc đơn.
 - (b) $\{0^n 1^m | n \leq m\}$
 - (c) $\{0^n | n \text{ là số chính phương}\}$
 - (d) $\{0^n | n \text{ là lũy thừa của } 2\}$

2.4 Định lí Myhill-Nerode

1. Phát biểu và chứng minh điều kiện cần và đủ của ngôn ngữ chính quy (định lí Myhill - Nerode).
2. Xét tính chính quy của các ngôn ngữ sau dựa vào định lí Myhill-Nerode:
 - (a) $\{0^n 1^m | 0 \leq n \leq m\}$
 - (b) $\{0^n 1^m | m, n \geq 0\}$

3 Ngôn ngữ phi ngữ cảnh

3.1 Tính nhập nhằng

Chứng minh rằng văn phạm phi ngữ cảnh với các quy tắc sinh sau đây là nhập nhằng (quy ước các chữ cái in hoa là kí hiệu phụ):

$$S \rightarrow aS \mid aSbS \mid \epsilon$$

Tìm một văn phạm không nhập nhằng tương đương với văn phạm trên.

3.2 Điều kiện cần của ngôn ngữ phi ngữ cảnh

1. Phát biểu và chứng minh điều kiện cần của ngôn ngữ phi ngữ cảnh (bổ đề bơm).
2. Sử dụng định lý về điều kiện cần của ngôn ngữ phi ngữ cảnh, chứng minh các ngôn ngữ sau không phải là ngôn ngữ phi ngữ cảnh:

(a) $L_1 = \{w \in \{a, b, c\}^* : |w|_a = |w|_b = |w|_c\}$

(b) $L_2 = \{0^n 1^n 0^n 1^n : n \geq 0\}$

(c) $L_3 = \{0^{2^i}, i \geq 1\}$

3.3 Ôtômat đẩy xuống

Xây dựng ôtômat đẩy xuống đoán nhận các ngôn ngữ sau và cho ví dụ hoạt động của từng ôtômat:

1. $L_1 = \{a^n b^{2n}, n \geq 1\}$
2. $L_2 = \{x\tilde{x}, x \in \{a, b\}^*\}$

4 Bài tập lập trình

4.1 Thuật toán trên ôtômat

1. Bài tập không bắt buộc: Cài đặt thuật toán đơn định hoá otomat. Sinh viên làm bài tập này không cần ôn tập bài tập phần 2.1 và 2.2.
2. Bài tập bắt buộc: Xây dựng chương trình thực hiện tối thiểu hoá một ôtômat hữu hạn đơn định đã cho (thuật toán đã minh hoạ trong tài liệu bài giảng)

4.2 Thuật toán phân tích cú pháp

Sinh viên làm bài tập này không cần ôn tập bài tập phần 3.

1. Xây dựng chương trình thực hiện biến đổi một văn phạm phi ngữ cảnh đã cho về dạng chuẩn Chomsky.
2. Viết chương trình phân tích cú pháp sử dụng một văn phạm phi ngữ cảnh dạng chuẩn Chomsky theo thuật toán CKY.

4.3 Chương trình dịch

Sinh viên làm bài tập này không cần ôn tập bài tập phần 2 và 3.

Sử dụng công cụ Flex và Bison (<https://www.youtube.com/watch?v=54bo1qaHAfk>, https://www.youtube.com/watch?v=__wUHG2rfM, http://aquamentus.com/flex_bison.html) xây dựng chương trình dịch cho phép dịch một ngôn ngữ mới.

- Chương trình nguồn: Gồm 4 dòng lệnh
 - Dòng 1 - 3 gán giá trị cho 3 biến khác nhau: <tên biến> = <số tự nhiên>
 - Dòng 4: <tên biến> <phép toán> <tên biến>
<tên biến> là tên 1 trong 3 biến đã gán giá trị ở dòng 1-3. 2 <tên biến> ở đây có thể trùng nhau hoặc khác nhau.
<phép toán> là một trong 5 phép toán +, -, *, / (lấy phần nguyên), hoặc % (chia lấy phần dư)
- Kết quả chương trình: Kết quả biểu thức ở dòng 4.