

BÀI TẬP ÔN TẬP THI CUỐI KÌ

Môn: Lý thuyết ngôn ngữ hình thức và ôtômat

1 Ngôn ngữ hình thức và văn phạm

1. Định nghĩa văn phạm, quan hệ dẫn trực tiếp và gián tiếp, ngôn ngữ sinh bởi văn phạm.
2. Phân loại văn phạm, ngôn ngữ.
3. Ôn các bài tập viết văn phạm sinh ngôn ngữ, tìm ngôn ngữ sinh bởi văn phạm cho trước.

2 Ngôn ngữ chính quy

2.1 Ôtômat hữu hạn trạng thái

1. Định nghĩa otomat hữu hạn đơn định, ngôn ngữ đoán nhận bởi otomat hữu hạn đơn định.
2. 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.
3. 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 $\{a\}^* \cap L(A) = \emptyset$.
4. 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. Định nghĩa biểu thức chính quy.
2. Cho biểu thức chính quy sau: $(bc \cup cb)(a^*)bccb$. Vẽ đồ thị chuyển, viết dưới dạng hình thức ôtômat và viết văn phạm chính quy tương đương với biểu thức chính quy này.
3. 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
 - $[03a-c] \equiv \{0, 3, a, b, c\}$, $[^15] \equiv$ 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) Tập tên biến/tên hàm trong chương trình
- (b) Các số thập phân
- (c) Tập số thực dấu phẩy động
- (d) Tập các địa chỉ Internet IPv4 (32 bits) biểu diễn bằng chuỗi số trong hệ thập phân.

2.3 Bỏ đề bơm

1. Phát biểu và chứng minh điều kiện cần của ngôn ngữ chính quy (bỏ 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) $\{0^n 1^m | m, n \geq 0\}$

2.5 Thuật toán tối thiểu hóa otomat

Thuật toán được mô tả như sau.

Cho otomat hữu hạn, đơn định $A = (S, \Sigma, \delta, s_0, F)$, trong đó $F \neq \emptyset$.

Kí hiệu P là phân hoạch của S . Hai trạng thái s_i và s_j thuộc cùng một tập trong phân hoạch P được gọi là tương đương theo P , kí hiệu $s_i \equiv_P s_j$.

Mục đích của thuật toán là tìm P trong đó mỗi tập hợp trong phân hoạch chứa các trạng thái tương đương của otomat A (quan hệ tương đương bất biến phải đã giới thiệu trong bài học).

$P \leftarrow \{F, S - F\}$

repeat

```

done ← true
for Q ∈ P
  for si, sj ∈ Q, si ≠ sj
    for a ∈ Σ
      if δ(si, a) ≠P δ(sj, a)
        done ← false
    làm mịn Q theo a
until done;

```

Từ phân hoạch P thu được như trên, ta xây dựng otomat tối tiểu M tương đương với otomat A , trong đó mỗi trạng thái của M tương ứng với 1 tập trạng thái trong P , đỉnh vào, đỉnh kết và các cung được khôi phục từ A .

Viết chương trình cho phép tối tiểu hóa một otomat hữu hạn đơn định bất kì.

3 Ngôn ngữ phi ngữ cảnh

3.1 Tính nhập nhằng

Chúng 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 lí 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 làm thêm

4.1 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 epsilon tự do bất kì 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.

4.2 Chương trình dịch

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

Sử dụng công cụ Flex và Bison (http://aquamentus.com/flex_bison.html, http://www.capsl.udel.edu/courses/cpeg421/2012/slides/Tutorial-Flex_Bison.pdf) xây dựng chương trình dịch cho phép ngôn ngữ sau:

- Chương trình nguồn: Một xâu biểu diễn một danh sách số nguyên trong cặp ngoặc đơn. Ví dụ: $()$, $(1,-2,3)$.
- Kết quả chương trình: Tổng các số nguyên trong danh sách.