الاتومات واللغات الصورية

المحاضرة الثانية — الاثنين 2017/10/16

اللغات الصورية: هي اللغة التي يمكن تعريفها بوساطة عدد من القواعد مثل: لغات البرمجة.

مفاهيم أساسية:

الرمز symbol: هو كائن غير قابل للتجزئة مثل الحروف الإنجليزية {a, b, c, ..., z} الأرقام العربية {0, 1, 2, ..., 9} الحروف العربية {أ, ب, ت, ..., ي}

ab ليس رمز

الأبجدية Alphabet: هي مجموعة منتهية وغير خالية من الرموز وتتميز بأنه لا يمكن توليد أي رمز منها بوساطة بقية الرموز ونرمز لها بـ Σ

* $\phi = \{\}$ ليست أبجدية.
* $\Sigma = \{a, b\}$
* $\Sigma = \{a, b, c, ..., z\}$ أبجدية اللغة الانكليزية
* $\Sigma = \{0, 1\}$ أبجدية نظام العد الثنائي
* $\Sigma = \{$أ, ب, ت, ..., ي$\}$ أبجدية اللغة العربية
* $\Sigma = \{0, 1, 2, ..., 9\}$ أبجدية نظام العد العشري
* $\Sigma = \{0, 1, 01\}$ ليست أبجدية لأن 01 ليس رمز وهو قابل للتجزئة ومولد من رموز المجموعة Σ.
* مجموعة الأعداد الطبيعية ليست أبجدية لأنها غير منتهية ($\mathbb{N}$ ليست أبجدية)

السلسلة String: هي تسلسل «تعاقب» «تتالي» عدد منتهي من الرموز المأخوذة من أبجدية ما دون وجود فراغات بينها.

* لتكن $\Sigma = \{a, b\}$ أبجدية عندئذٍ aba هي سلسلة مشكلة من الأبجدية Σ و babbbaa هي سلسلة أخرى مشكلة من الأبجدية Σ.

نرمز للسلاسل عادة برموز صغيرة x, y, z, w, r, v, u

$x = aa$

$y = abaabbb$

$w = bbbb$

$r = b$ (قد تكون السلسلة مشكلة من رمز واحد)

طول السلسلة: Length of String هو عدد الرموز المشكلة للسلسلة ونرمز لطول السلسلة w ب $|w|$

* مثلا: إذا كان لدينا السلسلة $v=0111$ المشكلة من أبجد النظام الثنائي $\Sigma=\{0,1\}$ عندئذ $|v|=4$

* $\Sigma=\{2,7,8\}$ ولتكن $x=77827$ سلسلة مشكلة من الأبجدية Σ عندئذ $|x|=5$

السلسلة الفارغة: Empty String هي السلسلة التي لا تحوي أية رموز وطولها يساوي الصفر ونرمز لها ب ϵ (ابسيلون) $|\epsilon|=0$

ملاحظة: ϵ هي سلسلة تنتمي إلى أي أبجدية.

بعض العمليات على السلاسل:

التعاقب بين السلاسل: concatenation

تعاقب سلسلتين x و y هو سلسلة مشكلة من توضع رموز السلسلة الأولى x متبوعة مباشرة برموز السلسلة الثانية y ونرمز لعملية التعاقب ب . وتحذف في معظم الأحيان

* مثال: لتكن لدينا الأبجدية $\Sigma=\{a,b\}$ ولتكن

$$\left.\begin{array}{l} x=aba \\ y=bbb \end{array}\right\} \Rightarrow \begin{array}{l} x.y=xy=ababbb \\ y.x=yx=bbbaba \end{array}$$

تكتب هيك للاختصار

* عملية تعاقب السلاسل عملية غير تبديلية

* السلسلة الفارغة ϵ هي عنصر حيادي بالنسبة لتعاقب السلاسل

$\epsilon x=x\epsilon=x$

$abaa\epsilon=abaa$

$\epsilon aabbb=aabbb$

* عملية تعاقب السلاسل عملية تجميعية.

بادئة سلسلة: prefix تكون السلسلة u هي بادئة السلسلة v إذا وجد سلسلة w بحيث يتحقق $v=u.w$

* مثال: $v=ababbbaa$

عندئذٍ مجموعة بادئات هذه السلسلة:

$\{\epsilon, a, ab, aba, abab, ababb, ababbb, ababbba, ababbbaa\}$

لاحقة سلسلة suffix: تكون السلسلة u هي لاحقة للسلسلة v إذا وجد سلسلة w بحيث يتحقق $v = w.u$

* مثال: $v = ababbbaa$

مجموعة لاحقات هذه السلسلة

$\{\epsilon, a, aa, baa, bbaa, bbbaa, abbbaa, babbbaa, ababbbaa\}$

السلسلة الجزئية: substring تكون السلسلة u سلسلة جزئية من سلسلة v إذا كانت كل رموز السلسلة موجودة في السلسلة v مع المحافظة على الترتيب

* في المثال السابق عندما $v = ababbbaa$

$$\left.\begin{array}{l} babbb \\ bab \\ ab \\ baa \end{array}\right\}$$ سلاسل جزئية

* aabb ليست سلسلة جزئية من v

* إن كل من البادئات واللاحقات هي سلاسل جزئية.

قوة سلسلة: power of String قوة سلسلة w من الدرجة n هو عبارة عن تعاقبها n مرة

$$w^n = \underbrace{w.w.w.....w}_{n \text{ مرة}}$$

* $x = abb = ab^2$

$x^1 = abb$, $x^2 = abbabb = ab^2ab^2$

$x^3 = x^2x = xx^2 = xx = abbabbabb = ab^2ab^2ab^2 = (ab^2)^3$

* $y = a^3b^4a$

$y^3 = a^3b^4a\,a^3b^4a\,a^3b^4a = a^3b^4a^4b^4a^4b^4a$

قوة أبجدية: power of Alphabet: هي مجموعة السلاسل المولدة من الأبجدية Σ ذات الطول n ونرمز لها بـ Σ^n

* مثال: $\Sigma = \{0,1\}$

$\Sigma^0 = \{\epsilon\}$

$\Sigma^1 = \{0,1\}$, $\Sigma^2 = \Sigma.\Sigma = \{0,1\}.\{0,1\} = \{00, 01, 10, 11\}$

$\Sigma^3 = \Sigma.\Sigma.\Sigma = \Sigma^2.\Sigma = \Sigma.\Sigma^2 = \{0,1\}.\{00,01,10,11\}$

$= \{000, 001, 010, 011, 100, 101, 110, 111\}$

Σ^*: نعرف Σ^* على أنها مجموعة كل السلاسل التي يمكن توليدها من رموز الأبجدية Σ وهي مجموعة غير منتهية وتحوي ϵ دوماً

$$\Sigma^* = \{\epsilon, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \ldots\ldots\}$$

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \Sigma^4 \cup \ldots$$

* ملاحظة: أحياناً يلزمنا تعريف كل السلاسل المولدة من الأبجدية Σ عدا السلسلة ϵ عندئذٍ نرمز لذلك بـ Σ^+ هي اجتماع كل السلاسل المولدة على Σ عدا Σ^0

$$\Sigma^+ = \Sigma^* - \Sigma^0 = \Sigma^* - \{\epsilon\}$$

$$\Sigma^+ \cup \{\epsilon\} = \Sigma^*$$

$$\{\epsilon\} = \Sigma^* - \Sigma^+$$

اللغة Language هي مجموعة السلاسل المختارة من المجموعة Σ^* والمولدة من الأبجدية Σ ونرمز لها بـ L أي أنه إذا كانت Σ أبجدية وكانت L لغة مولدة من Σ فإن $L \subseteq \Sigma^*$ أي أن اللغة المعرفة على أبجدية ما لا تحوي بالضرورة كل السلاسل المولدة من هذه الأبجدية.

* مثال:

لتكن $\Sigma = \{a, b\}$ أبجدية عندئذٍ Σ^* هي مجموعة كل السلاسل المولدة من Σ ولنفرض اللغة $L \subseteq \Sigma^*$ والتي هي لغة مولدة من الأبجدية Σ وكل كلماتها تحوي التعاقب aa عندئذٍ:

$$L = \{aa, baa, aab, aaaab, \ldots, aaaabaaaab, \ldots\}$$

وهي مجموعة غير منتهية في هذه الحالة.

انتهت المحاضرة