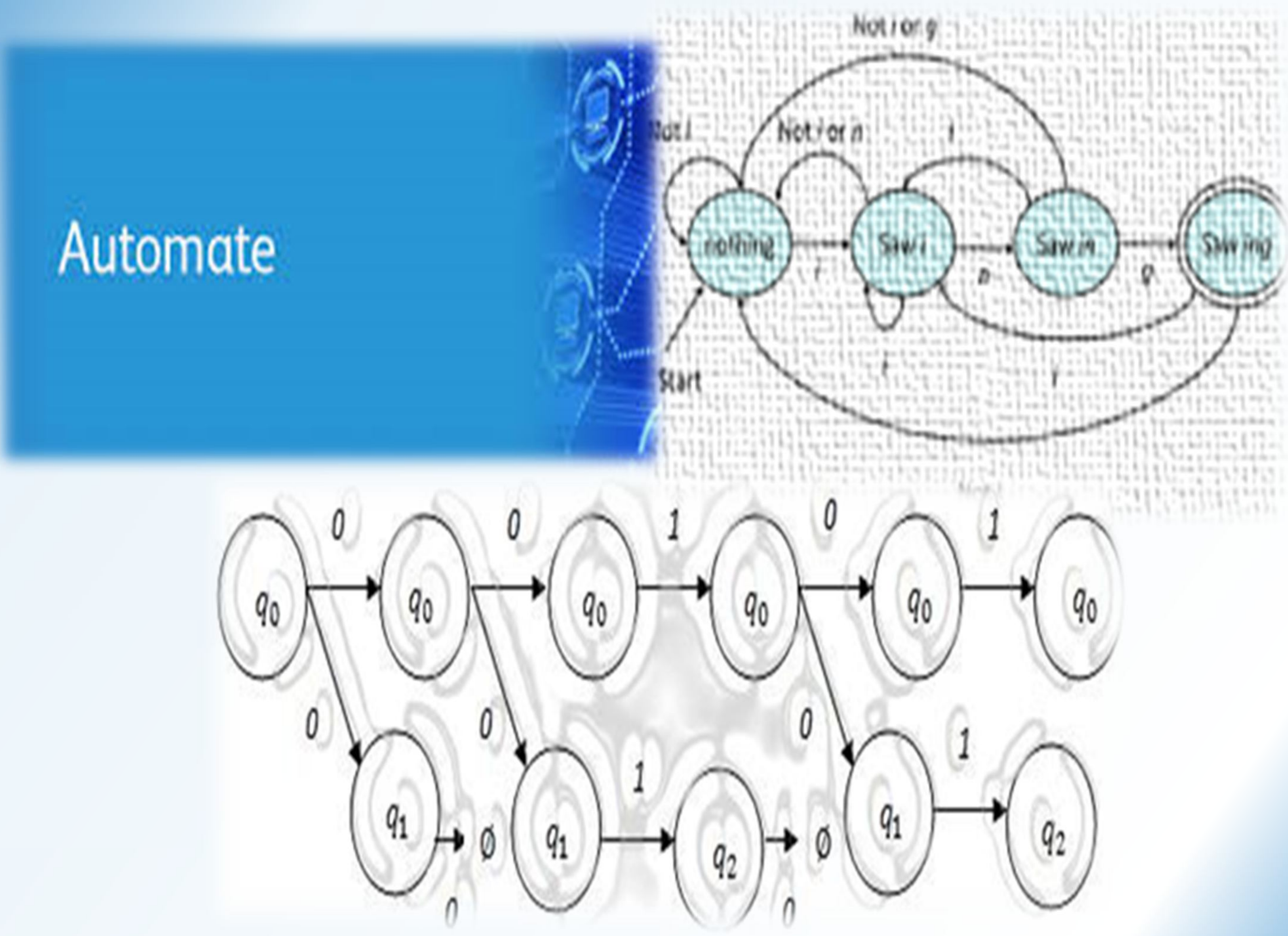
جامعة دمشق – كلية العلوم

قسم الرياضيات التطبيقية

نظرية الاتومات واللغات الصورية



2015/2014

تعتبر نظرية الاتومات واللغات الصورية من علوم الحاسوب النظرية وتدرس النمذجة الرياضية للحاسوب وعلاقة الاتومات بالغات الصورية وكيفية توليداللغة والتعرف عليها .

للمادة تطبيقات عديدة وهامة في تصحيح المترجمات او معالجة النصوص .

الاتومات : هي النمذجة الرياضية للحاسب .

مفاهيم اساسية :

❖ الرمز (Symbol) :

هو كائن غير قابل للتجزئة مثل الحروف العربية اواللاتينية او الارقام ......

❖ الابجدية (Alphabet) :

هي مجموعة منتهية ومحدودة وغير خالية من الرموزتتميز بانه لايمكن توليد اي رمز فيها بواسطة بقية الرموز ونرمز لها ب (سغما) Σ .

مثال

- الابجدية في الدارات المنطقية 0 و1 اي $\Sigma=\{0,1\}$
- الابجدية في اللغة الانكليزية $\Sigma=\{A,B,\dots\dots,Z, a, b,\dots\dots,z\}$
- الارقام العربية والهندية هي ايضا ابجدية $\Sigma=\{0,1,\dots\dots,9\}$
- ان $\Sigma=\{0,1,01\}$ ليست ابجدية لان: الرمز 01 مولد من رموزالابجدية Σ التي هي 0 و1

❖ السلسلة (String) :

هي تسلسل منتهي من رموزالابجدية متوضعة بجانب بعضها البعض بدون فراغات او فواصل .

عادة مانرمز للسلاسل بالرمز $\dots\dots, w, v, x, y$

مثال

$x=alphabet$ و $v=cat$ و $w=abbb$

مثال

لتكن لدينا الابجدية $\Sigma=\{a, b\}$ فإن

- السلسلة $w=aabbb$ سلسلة مولدة من الابجدية Σ
- السلسلة $v=aabbc$ سلسلة ليست مولدة من الابجدية Σ

❖ طول السلسلة (Length of String) :

وهي عدد الرموز المشكلة للسلسلة ونرمز لها ب $|\quad|$.

في المثال السابق $w=aabbb$ ومنه $|w| = 5$

❖ السلسلة الفارغة (Empty String) :

هي سلسلة لا تحتوي اي رمز وطولها يساوي الصفر ونرمز لها ب (ابسلون) ε .

العمليات على السلاسل :

- التعاقب (الوصل) : تعاقب سلسلتين x وy هي سلسلة مشكلة من وضع رموز السلسلة الاولى متبوعة مباشرة برموز السلسلة الثانية ويرمز لهذه العملية ب (.) وتحذف في معظم الاحيان .

مثال

لدينا الابجدية $\Sigma=\{0,1\}$ ولدينا $x=000$ و $y=0100$ عندئذ $x.y=0000100$
$y.x=01000000$

- ان عملية التعاقب ليست تبديلية لكن تجميعية والعنصر الحيادي للعملية هو السلسلة الفارغة اي $w\varepsilon=\varepsilon w=w$

❖ بادئة سلسلة (Prefix) :

تكون السلسلة u بادئة السلسلة v اذا وجدت السلسلة w بحيث يتحقق $v=uw$.

مثال

$V=ababbbaa$
فتكون مجموعة البادئات هي $\{\varepsilon,a,ab,aba,.....\}$

❖ لاحقة سلسلة (Suffix) :

تكون السلسلة u لاحقة السلسلة v اذا وجدت سلسلة w حيث يتحقق $v=wu$.

في المثال السابق

مجموعة اللاحقات هي $\{\varepsilon,a,aa,baa,....\}$

❖ السلسلة الجزئية (Substring) :

تكون السلسلة u سلسلة جزئية من v اذا كان كل رمز من u موجود في v وجميع الرموز بنفس الترتيب او التسلسل .

مثال

لتكن لدينا السلسلة $v=ababbbaa$
ان كل من $u=abbb$ و $x=bbbaa$ سلاسل جزئية من v
و$y=aabbb$ ليست سلسلة جزئية من v .

ملاحظة : يرمز لعدد ظهور الرمز a في السلسلة x بالرمز $\|x\|_a$.

مثال

لدينا السلسلة $v=0100$ عندئذ $|v|_0 = 3$, $|v|_1 = 1$

❖ قوة الابجدية (Power of Alphabet) :

هي مجموعة كل السلاسل اللتي يمكن تشكيلها او تركيبها باستخدام عملية التعاقب بين رموز الابجدية Σ

- ان Σ^* هي : مجموعة كل السلاسل من رموز الابجدية Σ وهي مجموعة غير منتهية وتشمل دوما ε.

مثلا : لتكن الابجدية $\Sigma=\{0,1\}$ عندئذ
$\Sigma^* =\{\varepsilon,0,1,01,10,00,11,001,100,010,........\}$

ملاحظة :   نرمز لمجموعة الكلمات المعرفة على ابجدية ولها الطول n بالرمز Σ^n اي ان كل سلسلة مولدة من هذه الابجدية طولها n .

بالعودة للمثال السابق نجد ان :

$\Sigma^0=\{\varepsilon\}$, $\Sigma^1=\{0,1\}$,
$\Sigma^2=\Sigma^1\,\Sigma^1=\{o,1\}\,\{o,1\}=\{00,11,01,10\}$,
$\Sigma^3=\Sigma^1\,\Sigma^1\,\Sigma^1=\Sigma^2\,\Sigma^1=\{00,10,01,11\}\,\{o,1\}=\{000,111,100,.........\}$

نتيجة : $\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup .... =\{\varepsilon,0,1,00,11,...........\}$

ملاحظة : احيانا نحتاج الى تعريف جميع السلاسل على الابجدية عدا السلسلة الفارغة عندئذ نرمز لذلك ب Σ^+ اي : $\Sigma^+= \Sigma^*-\{\varepsilon\} =\Sigma^* - \Sigma^0 \Longrightarrow \Sigma^* =\Sigma^+ \cup \Sigma^0$

❖ اللغة (language) :

هي مجموعة السلاسل المختارة من المجموعة Σ^* المعرفة على الابجدية Σ ونرمزلها L.
اذاكانت $L \subseteq \Sigma^*$ وكانت Σ ابجدية فان L هي لغة مبنية على Σ (مأخوذة من الابجدية Σ)

ملاحظة :
ان اي لغة على ابجدية لا تحوي بالضرورة كل السلاسل المولدة من الابجدية

مثال
لتكن الابجدية $\Sigma = \{a , b\}$ بحيث $L \subseteq \Sigma^*$
و L هي اللغة التي تحوي التعاقب aa اي
$L = \{aa, aab, aabb, aaba, aaa, \ldots\ldots\ldots\}$

مثال
لتكن الابجدية $\Sigma = \{a , b ,c \}$ بحيث

- L_1 هي اللغة المكونة من الكلمات التي تحوي c مرتين على الاقل (موشرط التعاقب هنا ل c) .

$L_1 = \{cc, abcc, cbc, aabbcbac , \ldots\ldots\ldots\}$

- L_2 هي اللغة المكونة من الكلمات التي تحوي b مرتين على الاكثر.

$L_2 = \{\varepsilon, a, b, c, ab, abc, abbc , abcb , \ldots\ldots\ldots\}$

انتهت المحاضرة ☺
Tasneem Shalabi