تتمة في اللغات خارج السياق

د. محمد الأحمد

24/5/2023

نظري

ستتحدث في هذه المحاضرة عن:

تبسيط القواعد خارج السياق CFG

صيغة تشومسكي المعيارية CNF

خوارزمية العضوية CYK

خواص اللغات خارج السياق

تذكرة:

ذكرنا سابقا أن القواعد الخاصة باللغات خارج السياق تمثل من خلال الرباعية التالية:

$$(V, \Sigma, R, S)$$

تكون قواعد الإنتاج الخاصة باللغات خارج السياق على الشكل التالي:

$$A \longrightarrow \alpha \; : \; \alpha \in (V \cup \Sigma)^*$$

حيث يكون الطرف اليساري عبارة عن متحول والطرف اليميني عبارة عن أي تشكيلة من الرموز الغير منتهية (المتحولات) والرموز المنتهية Terminal ومن الممكن أن تكون α عبارة عن ε ، والآن سنتحدث عن تبسيط هذه القواعد

تبسيط القواعد خارج السياق CFG

تبسيط القواعد هو إعادة كتابة القواعد بطريقة أبسط بحيث نحصل على قواعد مكافئة تعرف نفس اللغة.

تشمل عملية تبسيط القواعد ثلاث عمليات رئيسية:

1. **إزالة الرموز التافهة** removal of nullable variables
2. **إزالة الرموز الغير نافعة** removal of useless variables
3. **إزالة القواعد الأحادية** removal of unit-production

إزالة الرموز التافهة:

- نقول عن رمز ما وليكن N أنه رمز تافه $nullable$ إذا استطعنا أن نصل منه لـ ε (بشكل مباشر أو غير مباشر) بحيث: $N \xrightarrow{*} \varepsilon$
- تسمى القاعدة ($N \xrightarrow{*} \varepsilon$) بالقاعدة الخالية $\varepsilon - production$ وهي القاعدة التي ينتج عنها ε
- الهدف من إزالة الرموز التافهة هو إزالة القواعد الخالية التي تنتج الرمز ε

فينتج لدينا لغة بدون الرمز ε

$$L` = L \backslash \{\varepsilon\}$$

لماذا من الضروري إزالة الرموز التافهة:

إن إزالة الرموز التافهة يؤدي إلى تقليل الغموض مثلا وليكن لدينا هذه القاعدة $S \rightarrow aSb|\varepsilon$
في هذه القاعدة من الممكن بأي لحظة أن نستبدل S ب ε وهذا يؤدي إلى عدة شجرات اشتقاق لدخل واحد وعلى مستوى الكومبايلر ستكون عملية توقع وجود الرمز ε في أي لحظة عملية مكلفة.

طريقة إزالة الرموز التافهة:

1. نبحث عن الرموز التافهة التي ينتج عنها بشكل مباشر ε أو بشكل غير مباشر
2. نقوم بإزالة الرمز ε من القواعد عن طريق التعويض بـ أثر ε في هذه القواعد. وللتعويض بـ أثر ε نضع في كل قاعدة تحوي متحول يولد ε كل التشكيلات الممكنة لهذا المتحول بوجود ε أو بعدم وجود ε

توضيح: ليكن لدينا القواعد التالية:

$$X \rightarrow \alpha N \beta$$
$$N \rightarrow \varepsilon$$

نلاحظ أن الرمز N هو رمز تافه لأنه يعطي الرمز ε بشكل مباشر ، نقوم بإزالة الرمز ε عن طريق التعويض بأثره في القواعد التي تحوي الرمز التافه N.

فتصبح القواعد بعد إزالة الرمز التافه N بحيث يكون مرة $N = \varepsilon$ ومرة $N \neq \varepsilon$

$$X \rightarrow \alpha N \beta | \alpha \beta$$

حالة $N \neq \varepsilon$ ← ... → حالة $N = \varepsilon$

وهكذا يصبح N رمز غير تافه

مثال1: ليكن لدينا مجموعة القواعد التالية :

$$S \to AB$$
$$A \to aA|\varepsilon$$
$$B \to bB|\varepsilon$$

الحل:

لدينا A و B رموز تافهة لأنهما يعطيان الرمز ε بشكل مباشر ولدينا الرمز S هو رمز تافه لأنه يحتوي على كلا المتحولين A و B وكلاهما يعطيان الرمز ε ومنه فإن S رمز تافه بشكل غير مباشر

والآن نبدأ بتعويض بأثر ε في القواعد السابقة بحيث نفرض بأن الرمز التافه مرة يساوي ε ومرة لا يساوي ε

- نبدأ من القاعدة الأولى S التي تحوي الرمزان التافهان A و B :
  - عندما يكون $A = \varepsilon$ نحصل على القاعدة: $S \to B$
  - عندما يكون $B = \varepsilon$ نحصل على القاعدة: $S \to A$
  - عندما يكون $A \neq \varepsilon$ و $B \neq \varepsilon$ نحصل على القاعدة: $S \to AB$

فتكون القاعدة S بالشكل التالي:

$$S \to AB|A|B$$

- والآن نكرر نفس العملية على القاعدتين A و B:
  - لدينا القاعدة الثانية A تحوي الرمز التافه A:
  - عندما يكون $A = \varepsilon$ نحصل على القاعدة: $A \to a$
  - عندما يكون $A \neq \varepsilon$ نحصل على القاعدة: $A \to aA$

فتكون القاعدة A بالشكل التالي:

$$A \to aA|a$$

- القاعدة الثالثة B تحوي الرمز التافه B:
  - عندما يكون $B = \varepsilon$ نحصل على القاعدة: $B \to b$
  - عندما يكون $B \neq \varepsilon$ نحصل على القاعدة: $B \to bB$

فتكون القاعدة B بالشكل التالي:

$$B \rightarrow bB|b$$

بعد أن قمنا بإزالة جميع الرموز التافهة تصبح القواعد بالشكل التالي:

$$S \rightarrow AB|A|B$$
$$A \rightarrow aA|a$$
$$B \rightarrow bB|b$$

وهذه القواعد ينتج عنها نفس الكلمات التي تنتجها القواعد السابقة ولكن بدون وجود ε أي:

$$L` = L \setminus \{\varepsilon\}$$

مثال2: ليكن لدينا مجموعة القواعد التالية:

$$S \rightarrow AB$$
$$A \rightarrow aAA|\varepsilon$$
$$B \rightarrow bBB|\varepsilon$$

الحل:

A و B رموز تافهة بشكل مباشر

S رمز تافه بشكل غير مباشر

بعد تطبيق الخطوات لسابقة نفسها تنتج القواعد التالية:

$$S \rightarrow AB|A|B$$
$$A \rightarrow aAA|aA|a$$
$$B \rightarrow bBB|bB|b$$

- في القاعدة A يمكن أن تكون A الأولى والثانية معاً تساويان ε فنحصل على القاعدة $A \rightarrow a$
- ومن الممكن أن تكون واحدة منهم فقط تساوي ε فنحصل على القاعدة $A \rightarrow aA$
- ومن الممكن ألا يساوي أياً منهما ε فنحصل على القاعدة $A \rightarrow aAA$

لماذا في القاعدة S من غير الممكن أن يكون $A = \varepsilon$ و $B = \varepsilon$ في نفس الوقت؟
لأنه سينتج لدينا فقط الرمز ε والهدف لدينا هو الوصول لقواعد لا تحوي الرمز ε

إزالة الرموز الغير نافعة:

- **الرمز النافع** Useful Variables: **نقول عن X أنه رمز نافع إذا استطعنا الوصول له من رمز البداية S وإذا استطعنا الوصول من هذا الرمز بعد القيام بعملية اشتقاق أو أكثر إلى كلمة تنتمي إلى Σ^***

$$S \overset{*}{\rightarrow} \alpha X \beta \overset{*}{\rightarrow} w \in \Sigma^*$$

- **أي نقول عن متحول أنه نافع عندما يكون مولد للكلمات (أي نستطيع من خلاله الوصول إلى رمز نهائي** Terminal **) ويمكن الوصول له من رمز البداية.**
- **ومنه فإن الرموز تكون غير نافعة** Useless Variables **إذا كانت:**
  - **غير مولدة** Not generative
  - **غير قابلة للوصول** Not reachable
  - **غير مولدة وغير قابلة للوصول في نفس الوقت.**

مثال1: ليكن لدينا مجموعة القواعد التالية :

$$S \rightarrow BC$$
$$B \rightarrow a$$
$$C \rightarrow T$$

الحل:

نلاحظ أن الرمز B هو رمز نافع لأنه مولد و استطعنا من خلاله الوصول إلى رمز نهائي أما الرمز C فهو رمز غير نافع لأنه غير مولد لم نستطع من خلاله الوصول إلى رمز نهائي
وبالتالي يمكن حذف القاعدة C وكل قاعدة متعلقة بـ C يتم حذفها أيضا لأننا فقط نهتم بالقواعد التي تولد كلمات فتصبح لدينا القواعد بالشكل التالي:

$$B \rightarrow a$$

بما أنه في القاعدة S لدينا القاعدة BC متعلقة بـ C لذلك قمنا بحذفها

مثال2: ليكن لدينا مجموعة القواعد التالية حيث A هو رمز البداية:

$$A \rightarrow a$$
$$E \rightarrow b$$

الحل:

نلاحظ أن الرمز E هو رمز غير نافع لأنه غير قابل للوصول إذ لا نستطيع الوصول إليه انطلاقا من رمز البداية A وبالتالي على الرغم من أن الرمز E مولد ولكنه غير قابل للوصول نقوم بحذف قاعدته فتصبح لدينا القواعد بالشكل التالي:

$$A \rightarrow a$$

ملاحظات:

- يؤدي إزالة القواعد غير المولدة والغير قابلة للوصول الى جعل القواعد أكثر كفاءة.
- في حال اجتمع لدينا اثناء اختصار القواعد رمز غير مولد ورمز اخر غير قابل للوصول عندها نحذف اولاً الرمز الغير مولد ثم نحذف الرمز الغير القابل للوصول.

مثال3: ليكن لدي مجموعة القواعد التالية حيث S هو رمز البداية:

$$S \rightarrow AB|a$$

$$A \rightarrow b$$

الحل:

نلاحظ ان الرمز B هو رمز غير مولد وبالتالي نقوم بحذف قاعدته فأصبحت لدينا القواعد بالشكل التالي:

$$S \rightarrow a$$

$$A \rightarrow b$$

نلاحظ أن الرمز A هو رمز غير قابل للوصول نقوم بحذف قاعدته أيضاً فتصبح لدينا القواعد بالشكل التالي:

$$S \rightarrow a$$

إزالة القواعد الأحادية

القاعدة الأحادية (unit production): هي كل قاعدة يكون فيها متحول وحيد يعطي متحول وحيد أخر

$$X \rightarrow Y$$

مثل:

$$\left.\begin{array}{l} X \rightarrow Y \\ Y \rightarrow Z \end{array}\right\} X \rightarrow Z$$

مثال 1 : ليكن لدينا القواعد التالية :

$$S \to A$$

$$A \to B$$

$$B \to c$$

الحل : إن القاعدة $S \to A$ هي قاعدة أحادية نقوم باختصارها بحيث يصبح لدينا :

$$S \to B$$

$$B \to c$$

إن القاعدة $S \to B$ هي قاعدة أحادية أيضا نقوم باختصارها بحيث يصبح لدينا:

$$S \to c$$

حيث c هي Terminal

فكرة: إذا كان لدينا حلقة من القواعد الأحادية بحيث

$$A_1 \to A_2 \to \cdots \to A_k \to A_1$$

نقوم بحذف الحلقة ونستبدل كل شيء بـA_1

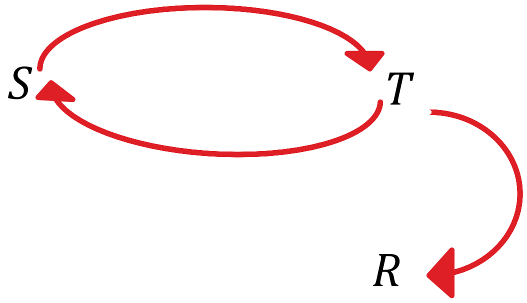
مثال 2: ليكن لدينا القواعد التالية:

$$S \to 0S1|1S0S1|T$$

$$T \to S|R$$

$$R \to 0SR$$

الحل: نلاحظ أنه لدينا حلقة بين الرمز S والرمز T .



نحاول حذف هذه الحلقة واستبدال T بـ S على الشكل التالي:

$$S \to 0S1|1S0S1|\cancel{T}$$

$$T \to \cancel{S}|R$$

$$R \to 0SR$$

فتصبح لدينا القواعد على الشكل التالي:

$$S \to 0S1|1S0S1$$

$$S \to R$$

$$R \to 0SR$$

لدينا القاعدة الأحادية $S \to R$ نقوم باختصارها:

$$S \to 0S1|1S0S1|0SR$$

ملاحظة: إذا كان لدينا في القواعد رموز تافهة وقواعد أحادية نقوم بالتخلص أولاً من الرموز التافهة ثم من القواعد الأحادية.

صيغة تشومسكي المعيارية Chomsky Normal Form (CNF)

يقول تشومسكي أن جميع القواعد خارج السياق يكون الطرف اليميني عبارة عن تشكيلة لا نهائية من الرموز الغير منتهية (المتحولات) والرموز المنتهية وعملية توقع هذه التشكيلة بالنسبة للكومبايلر ستكون مكلفة جداً .

لذلك اقترح تشومسكي أن يتم كتابة هذه القواعد بشكل معياري بحيث يكون الطرف اليميني للقاعدة عبارة عن تتالي متحولين أو رمز منتهي.

فيكون الشكل المعياري للقواعد:

$$A \rightarrow BC|b$$

BC : تتالي متحولين

b : رمز منتهي

مثال 1: ليكن لدينا القاعدة التالية:

$$S \rightarrow ABC$$

الحل: نلاحظ ان هذه القاعدة هي قاعدة خارج السياق ولكنها ليست بصيغة تشومسكي المعيارية، لتحويلها لصيغة تشومسكي يجب ان يكون الطرف اليميني عبارة عن متحولين او رمز منتهي:

$$S \rightarrow TC$$

$$T \rightarrow AB$$

مثال 2: حول القواعد التالية لصيغة تشومسكي المعيارية:

1) $S \rightarrow ab$

$S \rightarrow AB$

$A \rightarrow a$

$B \rightarrow b$

2) $S \rightarrow aA$

$S \rightarrow BA$

$B \rightarrow a$

$$3)\ S \to aBC$$

$$S \to TZ$$

$$T \to a$$

$$Z \to BC$$

ملاحظات :

- جميع القواعد خارج السياق يمكن تحويلها لصيغة تشومسكي المعيارية .
- معظم الخوارزميات والمترجمات تعتمد على صيغة تشومسكي لأنها تقدم سهولة اثناء عملية الـ parsing لأن الكومبايلر لن يضطر لفحص جميع التشكيلات الهائلة التي كانت سابقاً .

خوارزمية فحص العضوية Testing Membership Algorithm (CYK)

- تفيد خوارزمية CYK باتخاذ القرار فيما إذا كانت هناك كلمة تنتمي الى اللغة اعتماداً على قواعدها التي تحقق صيغة تشومسكي المعيارية.
- ذكرنا سابقاً انه لبرهان ان الكلمة تنتمي للغة L أولاً من خلال الاشتقاق بحيث نبدأ من رمز البداية ونشتق وفقاً لقواعد الاشتقاق حتى نصل للكلمة وبالتالي نكون قد رسمنا شجرة الاشتقاق Parse Tree.
- في طريقة الـ Parse Tree تكون الخوارزمية $(Top \to Down)$ لأننا نبدأ رمز البداية من جذر الشجرة ثم ننتقل الى أسفل الشجرة (الأوراق).
- أما خوارزمية CYK فهي $(Bottom \to Up)$ وهي أيضاً خوارزمية ديناميكية لأننا نقوم بتقسيم المشكلة لأجزاء لنحصل على حلول جزئية تساعدنا بإيجاد الحل الكلي للمسألة.

توضيح لفكرة خوارزمية CYK الديناميكية :

- في البداية نولد السلاسل الجزئية ذات الطول 1 من الكلمة w ومن ثم نوجد طريقة لتوليد جميع السلاسل الجزئية التي طولها 2 اعتماداً على السلاسل ذات الطول 1 ومن ثم نوجد طريقة لتوليد السلاسل التي طولها 3 اعتماداً على السلاسل التي طولها 1 والسلاسل التي طولها 2 ونتابع حتى نصل الى توليد سلسلة طولها n وننظم ذلك في جدول كما سيتم الشرح لاحقاً .

كيف تعمل خوارزمية CYK ؟

سيتم الشرح مع مثال فوراً:

مثال: لتكن لدينا مجموعة القواعد التالية:

$$S \to AB|BC$$
$$A \to BA|a$$
$$B \to CC|b$$
$$C \to AB|a$$

استخدم خوارزمية CYK للتحقق من انتماء الكلمة التالية $baaba$ إلى اللغة خارج السياق المعرفة بالقواعد السابقة.

الحل:

1. الخطوة الأولى: نتأكد من أن جميع القواعد المعطاة مكتوبة بصيغة تشومسكي المعيارية (الطرف اليميني عبارة عن متحولين فقط أو رمز منتهي) وأن تكون جميع القواعد أيضا مبسطة أي نزيل لرموز التافهة والغير نافعة والقواعد الأحادية

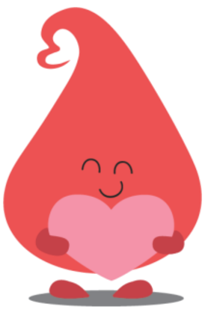
تطبيق الخطوة الأولى:

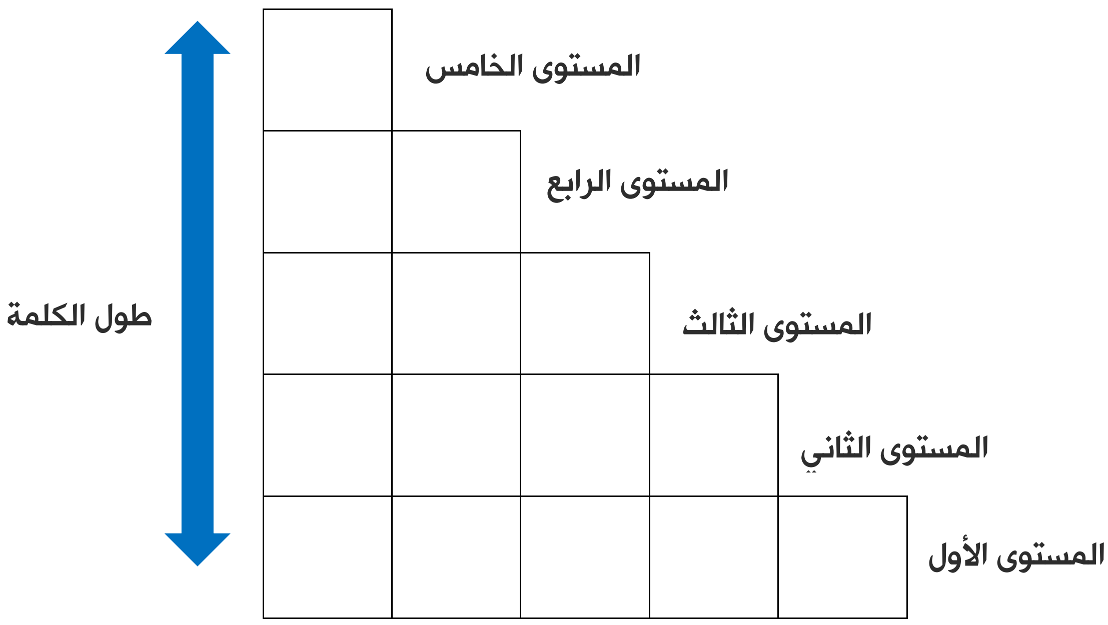
في هذا المثال لدينا جميع القواعد مكتوبة بصيغة تشومسكي ومبسطة لذلك ننتقل للخطوة الثانية

2. الخطوة الثانية: نحسب طول الكلمة n ونقوم ببناء جدول مثلثي سفلي بحيث يكون طول الجدول عبارة عن n (عدد محارف الكلمة)

تطبيق الخطوة الثانية:

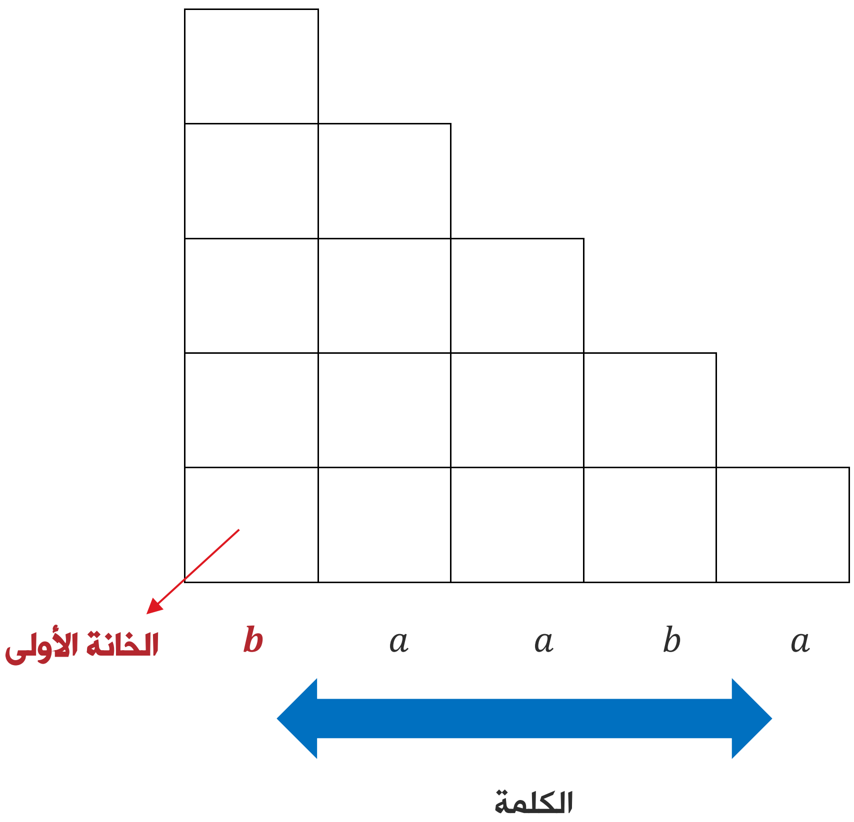
لدينا $w = baaba$ فإن طول الكلمة $|w| = 5$ أي $n = 5$ ومنه فإننا نقوم برسم جدول بحيث يكون طوله 5

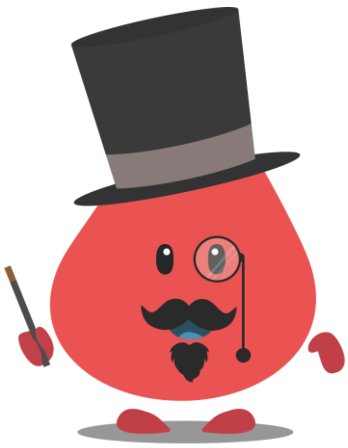




نلاحظ أن الجدول مدرَّج أي السطر الأول يحوي 5 خانات والثاني 4 خانات وهكذا.. وصولا إلى قمة المثلث حيث يحوي خانة واحدة

3. الخطوة الثالثة: نضع رموز الكلمة تحت القاعدة السفلى للمثلث فيصبح لدينا الجدول بالشكل التالي:





والآن نبدأ ببقية الجدول كالتالي:

4. الخطوة الرابعة: في هذه الخطوة سنقوم بتعبئة الصف الأول

a. نبدأ من اليسار ونذهب للخانة الأولى ننظر مباشرة إلى الرمز الموجود أسفل الخانة الأولى وهو b نذهب إلى القواعد المعطاة ونبحث عن القاعدة التي تعطي الرمز b بشكل مباشر وهي القاعدة B فنضع الرمز B في الخانة الأولى

b. والآن نذهب للخانة الثانية ننظر إلى الرمز الموجود أسفل هذه الخانة وهو a

نبحث عن القواعد التي تعطي الرمز a بشكل مباشر وهي A و C

فنضع A, C في الخانة الثانية

القواعد الخاصة بنص السؤال:

$S \rightarrow AB|BC$

$A \rightarrow BA|a$

$B \rightarrow CC|b$

$C \rightarrow AB|a$

ويتم تعبئة باقي الخانات في السطر الأول بنفس الطريقة ويصبح لدينا الجدول بعد الخطوة الرابعة بالشكل التالي:

B	A, C	A, C	B	A, C
b	a	a	b	a

في هذه الخطوة قمنا بإيجاد القواعد التي تولد سلاسل جزئية من الطول 1

5. الخطوة الخامسة: في هذه الخطوة سنقوم بتعبئة الصف الثاني من الجدول حيث في الصف الثاني سنقوم بإيجاد القواعد التي تولد السلاسل الجزئية من الطول 2 وبالتالي نقوم بتقسيم الكلمة المعطاة إلى سلاسل جزئية من الطول 2

$$\underbrace{b\ a}_{1}\ \underbrace{a\ b}_{3}\ a \qquad b\ \overbrace{a\ a}^{2}\ \overbrace{b\ a}^{4}$$

نلاحظ أن السلاسل الجزئية من الطول 2 هي عبارة عن سلسلتين جزئتين من الطول 1 وبالتالي نأخذ من الصف الأول كل خانتين معا ونأخذ تشكيلاتهم مع مراعاة الترتيب:

والآن نبدأ بتعبئة الصف الثاني:

نبدأ السلسلة الجزئية رقم 1 وهي ba :

في **السطر الأول** تحوي الخانة الخاصة بالرمز b على B وفي الخانة الخاصة بالرمز a لدينا A, C نأخذ التشكيلات الناتجة عن ba فينتج لدينا بالترتيب BA و BC

والآن نبحث عن القواعد التي تعطي التشكيلات السابقة. نعود للقواعد فنجد أن القاعدة A تعطينا التشكيل BA وبالتالي نضع A في الخانة الأولى من اليسار في الصف الثاني كما نجد أن القاعدة S تعطينا التشكيل BC فنضع **S بعد A** في الخانة نفسها كما هو موضح في الجدول المجاور

A,S				
B	A,C	A,C	B	A,C
b	a	a	b	a

1 2 3 4

وهكذا نكون قد انتهينا من تعبئة الخانة الأولى في الصف الثاني وننتقل للخانة الثانية

الآن نأخذ السلسلة الجزئية رقم 2 وهي aa

في السطر الأول تحوي الخانة الخاصة بالرمز a الأول A, C

وتحوي الخانة الخاصة بالرمز a الثاني A, C أيضا

نأخذ التشكيلات الناتجة عن aa فينتج لدينا بالترتيب AA و AC و CA و CC والآن نبحث عن القواعد التي تعطينا التشكيلات السابقة

نعود للقواعد فنجد أنه لا يوجد أي قاعدة تعطي AA كما لا يوجد أي قاعدة تعطي AC ولا يوجد أي قاعدة تعطي CA ولكن يوجد قاعدة تعطي CC وهي القاعدة **B** وبالتالي نضع **B** في الخانة الثانية من الصف الثاني كما هو موضح في الجدول

A,S	B			
B	A,C	A,C	B	A,C
b	a	a	b	a

1 2 3 4

A,S	B	S,C	A,S	
B	A,C	A,C	B	A,C

الآن نأخذ السلسلة الجزئية رقم 3 وهي ab:

في السطر الأول تحوي الخانة الخاصة بالرمز a (الخانة الثالثة) تحوي A,C الخانة الخاصة بالرمز b تحوي B نأخذ التشكيلات الناتجة عن ab فينتج لنا بالترتيب AB و CB **نعود للقواعد** فنجد أن القاعدة S تعطينا AB فنضع S في الخانة الثالثة من الصف الثاني ونجد أيضا ان القاعدة C **تعطينا AB أيضا فنضع C بعد S** و نأخذ السلسلة الجزئية رقم 4 و الأخيرة وهي ba وهي نفس السلسلة الجزئية رقم 1 فنضع A,S في الخانة الرابعة والأخيرة من الصف الثاني.

$\underbrace{b \quad a}_{1}$ $\underbrace{a \quad a}_{2}$ $\underbrace{a \quad b}_{3}$ $\underbrace{b \quad a}_{4}$

وهكذا نكون قد انتهينا من تعبئة الصف الثاني من الجدول

6. الخطوة السادسة: في هذه الخطوة سنقوم بتعبئة باقي الصفوف في الجدول بدءا من الصف الثالث حتى الصف الخامس بنفس الطريقة

تعبئة الصف الثالث:

A,S	B	S,C	A,S	
B	A,C	A,C	B	A,C
b	a	a	b	a

نبدأ من الخانة الأولى من اليسار ونقوم برسم متجهين متجه عامودي يتجه من السطر الأول حتى الخانة الأولى من السطر الثالث ومتجه مائل يتجه من الخانة الأولى للسطر الثالث وصولا إلى السطر الأول كما هو موضح في الرسم ثم نأخذ الخانة الأولى التي يمر بها المتجه العامودي وهي B مع الخانة الأولى التي يمر بها المتجه المائل وهي B فيتشكل لدينا من الخانتين السابقتين BB

ثم ننتقل للخانة الثانية من المتجه العامودي وهي A,S مع الخانة الثانية من المتجه المائل وهي A,C ونقوم بأخذ التشكيلات الناتجة من الخانتين السابقتين فينتج لدينا بالترتيب AA و AC و SA و SC

وتكون التشكيلات بالكامل التي حصلنا عليها من المتجهين هي BB و AA و AC و SA و SC والآن نبحث عن القواعد التي تعطينا التشكيلات السابقة ونلاحظ انه لا يوجد أي قاعدة تعطي هذه التشكيلات فنضع ϕ في الخانة الأولى من السطر الثالث كما في الجدول

$\emptyset$				
A,S	B	S,C	A,S	
B	A,C	A,C	B	A,C
b	***a***	***a***	***b***	***a***

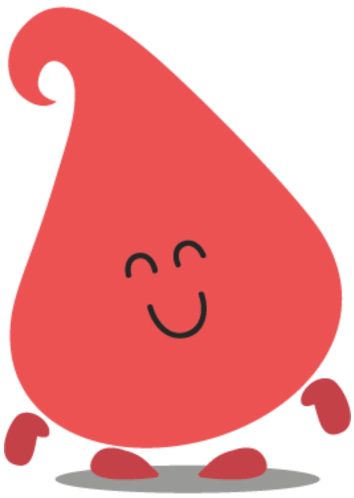
في الصف الثالث نحاول إيجاد القواعد التي تولد السلاسل الجزئية من الطول 3 أي عبارة عن وصل سلاسل جزئية من الطول 2 و 1 أو 1 و 2 لذلك نلاحظ أن الخانة الأولى التي يمر بها المتجه العامودي هي بالسطر الأول والخانة الأولى التي يمر بها السطر المائل هي بالسطر الثاني أي اخذنا تشكيلات سلاسل جزئية من الطول 1 و 2 ونلاحظ ان الخانة الثانية التي يمر بها المتجه العامودي هي بالسطر الثاني والخانة الثانية التي يمر بها المتجه المائل هي بالسطر الأول أي أخذنا تشكيلات سلاسل جزئية من الطول 2 و 1.

ومن اجل تعبئة الخانة الثانية نقوم باتباع الطريقة ذاتها فتنتج لدينا التشكيلات التالية:

AS و AC و CS و CC و BB

نلاحظ فقط ان التشكيل CC يوجد في القاعدة B وباقي التشكيلات غير موجودة وبالتالي نضع B في الخانة الثانية من السطر الثالث.

$\emptyset$	B			
A,S	B	S,C	A,S	
B	A,C	A,C	B	A,C
b	***a***	***a***	***b***	***a***



وسيكون الجدول بعد تعبئة الخانة الأخيرة في السطر الثالث بعد اتباع الخطوات نفسها بالشكل التالي:

$\emptyset$	B	B		
A,S	B	S,C	A,S	
B	A,C	A,C	B	A,C
b	a	a	b	a

والآن نبدأ بتعبئة الصف الرابع من الجدول:

ان طريقة تعبئة السطر الرابع هي نفس طريقة تعبئة السطر الثالث ولكن في السطر الرابع نحاول إيجاد القواعد التي تولد السلاسل الجزئية من الطول 4 أي عبارة عن وصل سلاسل جزئية من الطول (1 و 3) أو (3 و 1) أو (2 و 2) .

لنقوم بتعبئة الخانة الأولى من السطر الرابع:

$\emptyset$				
$\emptyset$	B	B		
A,S	B	S,C	A,S	
B	A,C	A,C	B	A,C

نقوم برسم متجه عامودي يتجه من السطر الأول باتجاه الخانة الأولى من السطر الرابع ومتجه مائل من الخانة الأولى في السطر الرابع باتجاه السطر الأول، نقوم بأخذ الخانة الأولى التي يمر بها المتجه العامودي وهي B "سطر أول " مع الخانة الأولى التي يمر بها المتجه المائل وهي B "سطر ثالث" فتكون لدينا التشكيلة BB ونقوم بأخذ الخانة الثانية التي يمر بها المتجه العامودي وهي A,S "سطر ثاني" مع الخانة الثانية التي يمر بها المتجه المائل وهي S,C "سطر ثاني" فتظهر لدينا التشكيلات التالية على التتالي: AS و AC و SS و SC كما نقوم بأخذ الخانة الثالثة التي يمر بها المتجه العامودية وهي $\emptyset$ مع الخانة الثالثة التي يمر بها المتجه المائل وهي B فيصبح لدينا $\emptyset B$ ،
فيكون مجموع التشكيلات الناتج عن هذين المتجهين هو:

BB و AS و AC و SS و SC و$\emptyset B$

نلاحظ عدم وجود أي قاعدة تعطي هذه التشكيلات فنقوم بتعبئة الخانة الأولى من السطر الرابع بـ$\emptyset$

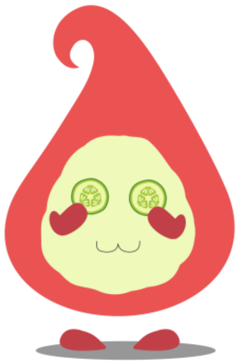
والآن لتعبئة الخانة الثانية من السطر الرابع:

بعد رسم المتجهات واخذ الخانات مع بعضها سيظهر لدينا:

AB و CB و BA و BS و BA و BC

نلاحظ ان التشكيلان AB و BC ينتجان من القاعدة S، و AB ينتج من القاعدة C والتشكيل BA ينتج من القاعدة A وبالتالي نملأ الخانة بـS,C,A

$\emptyset$	S,C,A			
$\emptyset$	B	B		
A,S	B	S,C	A,S	
B	A,C	A,C	B	A,C



ولتعبئة السطر الخامس نعيد نفس العملية المطبقة على السطرين الثالث والرابع، سينتج لدينا التشكيلات التالية:

BS و BA و BC و AB و SB و $\emptyset A$ و $\emptyset S$ و $\emptyset C$

ونلاحظ ان BA ينتج من القاعدة A و AB ينتج من القاعدة S والقاعدة C فيصبح لدينا الجدول كما في الشكل المجاور

S,A,C				
$\emptyset$	S,A,C			
$\emptyset$	B	B		
A,S	B	S,C	A,S	
B	A,C	A,C	B	A,C

ولمعرفة أن الكلمة $baaba$ تنتمي الى اللغة خارج السياق المعرفة بالقواعد الموجودة في نص السؤال اعتماداً على خوارزمية CYK يجب ان تكون الخانة الموجودة في أعلى المثلث تحوي الرمز S (رمز البداية)، وبما انه في مثالنا هذا الخانة الموجودة في اعلى المثلث S,A,C تحوي رمز البداية الخاص بالقواعد السابقة وهو S فإن الكلمة $baaba$ تنتمي الى اللغة.

مثال: تحقق من انتماء الكلمة $aaab$ الى اللغة المعرفة بقواعد المثال السابق.

الحل:

نبدأ بتطبيق الخطوات السابقة نفسها بدءاً من رسم الجدول حتى تعبئة اخر خانة موجودة في المثلث فينتج لدينا الجدول التالي: وبما ان الخانة الموجودة في اعلى المثلث تحوي رمز البداية الخاص بالقواعد السابقة وهو S فإن الكلمة $aaab$ تنتمي الى اللغة.

C,S			
S,C,A	B		
B	B	C,S	
A,C	A,C	A,C	B
$\boldsymbol{a}$	$\boldsymbol{a}$	$\boldsymbol{a}$	$\boldsymbol{b}$

ان تعقيد خوارزمية CYK هو n^3 حيث n هو طول الكلمة.

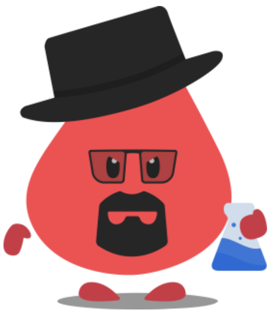
خواص اللغات خارج السياق

تذكرة: في اللغات المنتظمة جميع الخواص محققة (الاجتماع – التقاطع – المتمم – المعكوس – الوصل – تكرار – الفرق).
اما في اللغات خارج السياق فجميع الخواص السابقة محققة عدا (التقاطع – المتمم – الفرق)

إثبات الاجتماع:

إن اجتماع لغتين خارج السياق يعطي لغة خارج السياق، لأنه وبفرض لدي اللغة L_1 هي لغة خارج السياق ومعرفة بمجموعة من القواعد G_1 ورمز البداية لهذه القواعد هو S_1.

ولدينا اللغة L_2 هي لغة خارج السياق ومعرفة بالقواعد G_2 ورمز البداية لهذه القواعد هو S_2 فإن اجتماع L_1 و L_2 يعطي لغة خارج السياق ولتكن L ويصبح لدينا رمز بداية جديد وهو S يعطي إما رمز البداية S_1 أو رمز البداية S_2 .



$$\left.\begin{matrix} L_1 \\ L_2 \end{matrix}\right\} \quad S \to S_1|S_2$$

- ✓ أما بالنسبة للتكرار يصبح رمز البداية الخاص باللغة L الناتجة عن تكرار اللغة L_1 :

$$S = S.S_1$$

- ✓ أما في حالة المعكوس نحافظ على المتحولات بالطرف الأول ونعكس المتحولات في الطرف الثاني، مثلاً:

$$A \to aB$$

تصبح : $A \to Ba$

إثبات التقاطع:

لنبرهن أن تقاطع لغة خارج السياق لا يعطي لغة خارج السياق، سنورد مثال يعاكس فرضية ان تقاطع لغتان خارج السياق هي لغة خارج السياق:

ليكن لدينا اللغة L_1 المعرفة بالشكل التالي:

$$L_1 = \{a^i b^i c^j, i \neq j\}$$

وليكن لدينا اللغة L_2 المعرفة بالشكل التالي:

$$L_2 = \{a^i b^j c^j, i \neq j\}$$

كلتا اللغتين هما لغات خارج السياق لانهما تحتاجان الى مكدس لتخزين قيمة i في اللغة الأولى ولتخزين قيمة j في اللغة الثانية، ولنفرض اللغة L هي اللغة الناتجة عن تقاطع اللغتين وبالتالي كل كلمة من كلماتها يجب ان تنتمي الى كلا اللغتين معا وبالتالي يجب ان يكون شكل الكلمات التي تنتمي الى اللغة L من الشكل:

$$L_2 = \{a^i b^j c^k, i = j, j = k\}$$

وهذه اللغة ليست لغة خارج السياق لان كل الأسس متساوية وذكرنا سابقاً ان هذه اللغة ليست لغة خارج السياق.

- بما انه يوجد مثال واحد على الأقل ينقض خاصية تقاطع اللغات خارج السياق هي لغة خارج السياق فيمكننا الاستنتاج ان هذه الخاصية غير صحيحة .

وبما ان الفرق والتكامل يمكن تعريفهم من خلال التقاطع:

حيث بفرض لدينا اللغتين خارج السياق CFL_1 و CFL_2

فيكون فرق CFL_1 عن CFL_2 كالتالي:

$$CFL_1 \backslash CFL_2 = CFL_1 - (CFL_1 \cap CFL_2)$$

ويكون تكامل اللغة CFL_1 كالتالي:

$$CFL_1' = CFL_2 - (CFL_1 \cap CFL_2)$$

بالتالي يمكن برهان ان اللغة الناتجة الفرق بين لغتين خارج سياق او تكامل لغة خارج السياق هي ليست لغة خارج السياق.