Εργασία 5η

Επεξεργασία Φυσικής Γλώσσας

Εαρινού Εξαμήνου 2017



Δημήτρης Λώλης

Βασιλική Χήρα

Για την συγκεκριμένη εργασία χρησιμοποιήσαμε την πρόταση "I saw the man with the telescope" απο τη διαφάνεια 41 ( <http://lxmls.it.pt/2015/cky.html> ). H γραμματική για την συγκεκριμένη πρόταση βρίσκεται στο φάκελο που είναι και ο κώδικας. Για την επεξεργασία της γραμματικής (με όνομα grammar.txt) αρκεί ένας απλός text editor.

Για να επιτευχθεί ο σκοπός της άσκησης ο κώδικας μας χρησιμοποιήσαμε τις εξής **συναρτήσεις** με τη σειρά εκτέλεσης τους:

**Grammar**: Η συγκεκριμένη συνάρτηση διαβάζει το αρχείο γραμματικής και μετατρέπει τους κανόνες σε 2 λίστες που περιέχουν τα τερματικά και τα μη τερματικά στοιχεία αντίστοιχα.

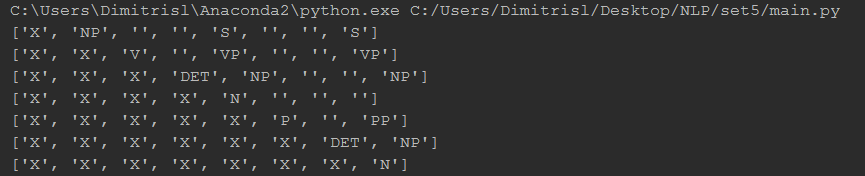
**Check\_grammar:** Η συγκεκριμένη συνάρτηση καλείται 2 φορές μέσα στον κώδικα α)για την αναγνώριση των στοιχείων της πάνω διαγώνιού β)για το matching των συνδυασμών των τερματικών στοιχείων μέσα στον πίνακα CKY.

**Left\_child** και **right\_child:** Πρόκειται για 2 αναδρομικές συναρτήσεις που καλούνται μέχρι να βρουν τα τερματικά στοιχεία του συντακτικού δέντρου. Παράγοντας κατ' αυτόν τον τρόπο συντακτικό δέντρο βάσει μιας δοσμένης πρότασης.

Οι **δομές** που χρησιμοποιήθηκαν είναι οι εξής:

**Matrix** : Δομή που διαμορφώνεται κατά την εκτέλεση του προγράμματος και περιέχει τις τιμές που διαμορφώνονται από τον αλγόριθμο του CKY . Το στοιχείο που ξεκινά την πρόταση είναι το S και βρίσκεται πάντα στην πρώτη γραμμή και στην τελευταία στήλη.

**Trails**: Πρόκειται για μια παράλληλη δομή με την matrix που περιέχει τους προγόνους του στοιχείου που έχει δημιουργηθεί από τον CKY, έτσι ώστε όταν τον τροφοδοτήσουμε στις αναδρομικές συναρτήσεις να δημιουργήσουν το συντακτικό δέντρο.

Ο αλγόριθμος CKY παράγει τον παρακάτω πίνακα: 

Στα αριστερά βρίσκεται το αποτέλεσμα του αλγορίθμου το οποίο μεταφράζεται στα δεξιά ως συντακτικό δέντρο .

|  |  |
| --- | --- |
|  |  |

Το δέντρο παράχθηκε από την πρόταση [S [NP I] [VP[V saw] [NP [NP [DET the] [N man]] [PP [P with] [NP [DET the] [N telescope]]].