ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Θεωρία Υπολογισμού

Σημειώσεις – Λήμμα της άντλησης

Αλέξιος Καπόρης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Λήμμα της Άντλησης

Α. Κ. Καπόρης

12 Μαρτίου 2009

Περιεχόμενα

Κεφάλαιο 1

Το Λήμμα της Άντλησης για μη κανονικές γλώσσες

1.1 Μη κανονικές γλώσσες

Ένα πεπερασμένο αυτόματο, παρά την απλότητα του, είναι μια πολύ ισχυρή κατασκευή. Η δύναμη του-ς γίνεται φανερή επειδή τα αυτόματα αυτά μπορούν να εκφράσουν όλες τις κανονικές εκφράσεις. Ένα πεπερασμένο αυτόματο, παρά το πεπερασμένο πλήθος των καταστάσεων του, μπορεί να εκφράσει κανονικές γλώσσες με οσοδήποτε μεγάλο μήκος n. Για παράδειγμα την γλώσσα $A = \{(01)^*\}$ που αποτελείται από όλες τις δυνατές λέξεις που μπορούμε να γράψουμε με αλφάβητο $S = \{01\}$. Το αυτόματο που αναγνωρίζει τη A, θα εκμεταλλευτεί την κανονικότητα της εναλλαγής από 0 σε 1, και κάνοντας κάτι σαν ''φλιπ-φλοπ'' μεταξύ 2 καταστάσεων θα φθάνει σε αποδοχή.

Όμως, υπάρχουν γλώσσες που δεν αναγνωρίζονται από πεπερασμένα αυτόματα. Αυτές οι γλώσσες δεν είναι κανονικές. Για παράδειγμα μια μη κανονική γλώσσα είναι η $B = \{0^n 1^n : n \geq 0\}$. Διαισθητικά, ο λόγος που η B δεν είναι αναγνωρίζεται από πεπερασμένο αυτόματο, είναι ότι μια λέξη της B μπορεί να αποτελείται από οσοδήποτε μεγάλη αρχική ακολουθία από 0΄ς συνολικού μήκους n. Αυτό το πλήθος n από 0΄ς δεν είναι δυνατόν να απομνημονευτεί στις πεπερασμένες καταστάσεις ενός αυτομάτου ώστε να συγκριθεί αν είναι ίσο με τα 1΄ς που ακολουθούν.

Το παραπάνω διαισθητικό επιχείρημα δεν αποτελεί απόδειξη της μη κανονικότητας της B. Δηλαδή, η ανάγκη τη μετρήσεως του πλήθους των διαφορετικών εμφανίσεων στοιχείων που μας ενδιαφέρουν σε μια λέξη δεν συνεπάγεται πάντα μη κανονικότητα της γλώσσας. Για παράδειγμα ας θεωρήσουμε της γλώσσες στο αλφάβητο $S = \{0, 1\}$:

$$C = \{w : \text{ έχει ίσο πλήθος 0΄ς και 1΄ς}\}$$

και

$$D = \{w : \text{ έχει ίσο πλήθος εμφανίσεων των υπακολουθίων 01 και 10}\}$$

Διαισθητικά, επειδή και οι δυο γλώσσες φαίνεται να απαιτούν μέτρηση εμφανίσεων επιθυμητών υπακολουθίων, μας κάνει να νομίζουμε ότι δεν είναι κανονικές. Το εκπληκτικό είναι, και συνάμα πολύ διδακτικό στο να μην εμπιστευόμαστε τυφλά την διαίσθηση, ότι η γλώσσα D είναι κανονική.

1.2 Λήμμα άντλησης για μη κανονικές γλώσσες

Θα παρουσιάσουμε μια τεχνική αυστηρής αποδείξεως ότι μια δοσμένη γλώσσα δεν είναι κανονική. Η τεχνική αυτή ονομάζεται *λήμμα της άντλησης* και βασίζεται στο γεγονός ότι όλες οι κανονικές γλώσσες έχουν μια *ειδική* ιδιότητα. Αν μπορέσουμε να δείξουμε, με τη βοήθεια του λήμματος, ότι η προς εξέταση γλώσσα L δεν έχει την ειδική ιδιότητα τότε σίγουρα η L δεν είναι κανονική.

Παρατηρούμε ότι το λήμμα της άντλησης έχει *αρνητικό στόχο*: την απόδειξη ότι η προς εξέταση γλώσσα *δεν* είναι κανονική.

Η ειδική ιδιότητα κάθε κανονικής γλώσσας: Θα προσπαθήσουμε να περιγράψουμε διαισθητικά την ειδική ιδιότητα που εμφανίζεται σε όλες τις κανονικές γλώσσες. Καταρχήν παρατηρούμε ότι κάθε μία κανονική γλώσσα L αντιστοιχεί σε ένα πεπερασμένο αυτόματο M_L που την αναγνωρίζει. Επειδή το M_L είναι πεπερασμένο τότε θα έχει πεπερασμένο πλήθος καταστάσεων, έστω k αυτές $q_1, \ldots, q_k$. Η κρίσιμη παρατήρηση είναι ότι κάθε λέξη w της κανονικής γλώσσας L με μήκος $w > k$, δηλαδή μεγαλύτερο από το πλήθος των καταστάσεων του M_L, όταν δοθεί ως είσοδο στο M_L τότε αναγκαστικά το αυτόματο θα περιέλθει σε μια κατάσταση του > 1 φορά.

Διότι ξέρουμε ότι κάθε αυτόματο, σε κάθε γράμμα της εισόδου που διαβάζει μεταβαίνει σε αντίστοιχη κατάσταση. Το κρίσιμο συμπέρασμα είναι: το M_L ακόμα και *αν* διαβάζοντας κάθε γράμμα της λέξης w μεταβαίνει σε *διαφορετική* κατάσταση, θα οδηγηθεί σε επανάληψη επίσκεψης καταστάσεως, επειδή ισχύει $k < n$ (καταστάσεις του M_L είναι $<$ μήκος λέξεως $w \in L$)

Σχηματικά, έστω μια αρκετά μακρουλή λέξη $w \in L$:

$$w = \underbrace{\overbrace{\underbrace{x_1 \ldots x_t}_{x}\, \underbrace{x_{t+1} \ldots x_p}_{y}\, \underbrace{x_{p+1} \ldots x_n}_{z}}^{\text{μήκος: } n>k}}_{\text{μήκος: } p\leq k} = x_1 \ldots x_p z$$

ώστε x_t να είναι το γράμμα που οδήγησε το αυτόματο M_L για *πρώτη* φορά στην κατάσταση q_p και x_p να είναι το γράμμα που οδήγησε το αυτόματο M_L για πρώτη φορά σε *επανάληψη* επισκέψεως της καταστάσεως q_p. Άρα, αν μετά το γράμμα x_p που μας οδήγησε στην q_p παραθέσουμε όσες $i \geq$ φορές επιθυμούμε την υπολέξη

$$\underbrace{x_{t+1} \ldots x_p}_{y}$$

πάντα θα φθάνουμε στην κατάσταση q_p. Έτσι όταν το αυτόματο M_L διαβάσει στην συνέχεια την υπό-λέξη

$$\underbrace{x_{p+1} \ldots x_n}_{z}$$

θα οδηγηθεί σε αποδοχή της w (διότι η μακρουλή λέξη $w \in L$). Το σημαντικό είναι να αντιληφθούμε σε αυτό το σημείο ότι αν φουσκώσουμε την μακρουλή λέξη y για όσες i φορές θέλουμε τότε προκύπτει η λέξη:

$$w(i) = x_1 \ldots x_t \left(x_{t+1} \ldots x_p\right)^i x_{p+1} \ldots x_n = xy^i z, \text{ για κάθε } i \geq 0$$

η οποία ανήκει στη κανονική γλώσσα L. Διότι η ''αντλημένη'' υπό-λέξη $\left(x_{t+1} \ldots x_p\right)^i$ (''φουσκωμένη'' κατά i φορές, αν σας αρέσει) πάντα θα μας οδηγεί στην κατάσταση q_p από την οποία μας παραλαμβάνει η υπό-λέξη z και μας οδηγεί με βεβαιότητα σε κατάσταση αποδοχής.

Όσα περιγράψαμε διαισθητικά παραπάνω διατυπώνονται αυστηρά με το εξής λήμμα:

Λήμμα Άντλησης: Αν L είναι μια κανονική γλώσσα, υπάρχει ένας φυσικός αριθμός p (είναι το μήκος φουσκώματος που χαρακτηρίζεται από το πλήθος k καταστάσεων του πεπερασμένου αυτόματος M_L που αποδέχεται την L), ώστε κάθε μακρουλή λέξη $w \in L$, δηλαδή με μήκος $|w| \geq p$, να μπορεί να γραφεί $w = xyz$ ώστε να ισχύουν οι παρακάτω ιδιότητες:

1. Για κάθε φυσικό $i \geq 0$ η φουσκωμένη λέξη ανήκει στην γλώσσα: $xy^i z \in L$.
2. Υπάρχει η αρχική υπό-λέξη ώστε να φουσκωθεί: $|y| > 0$.
3. Η υπό-λέξη που φουσκώνουμε έχει μήκος το πολύ: $|xy| \leq p$.

Αν υπόθεση $\Rightarrow$ συμπέρασμα. Η διατύπωση του Λήμματος άντλησης είναι της μορφής $A \Rightarrow B$ και όχι της μορφής $A \Leftrightarrow B$. Η μοναδική υπόθεση του λήμματος είναι η κανονικότητα της γλώσσας και οι ιδιότητες 1, 2, 3 έπονται ως συμπέρασμα. Όμως, αν για μια προς εξέταση γλώσσα L και για κάθε μακρουλή λέξη $w \in L$ με μήκος $|w| > p$ ισχύουν οι ιδιότητες 1, 2, 3 τότε αυτό δεν συνεπάγεται ότι η γλώσσα L είναι *κανονική*. Για παράδειγμα ας θεωρήσουμε ότι είναι αληθινή η παρακάτω πρόταση:

"*Αν είμαστε στην Ελλάδα $\Rightarrow$ όλες οι μέρες είναι ηλιόλουστες.*"

Πείτε ότι μας κλείνουν τα μάτια και μας μεταφέρουν σε μια άγνωστη χώρα (η οποία μπορεί να είναι και η Ελλάδα). Ακόμα και αν μας φανούν ηλιόλουστες όλες οι μέρες της άγνωστης χώρας, από αυτό και μόνο, δεν είναι σωστό να συμπεράνουμε ότι η χώρα που οδηγηθήκαμε είναι η Ελλάδα (διότι η προς εξέταση χώρα μπορεί να μην έχει όμορφες παραλίες). Προσοχή όμως, αν βρούμε έστω και μία συννεφιασμένη μέρα τότε αυτό σίγουρα σημαίνει ότι η χώρα δεν είναι η Ελλάδα. Αυτή είναι ακριβώς και η χρησιμότητα του λήμματος της άντλησης: η εύρεση έστω και μίας λέξης που παραβιάζει τις ιδιότητες 1, 2, 3 μας οδηγεί με ασφάλεια στο συμπέρασμα ότι η προς εξέταση γλώσσα δεν είναι κανονική.

1.3 Χρήση του Λήμματος Άντλησης

Το Λήμμα το χρησιμοποιούμε για να δείξουμε ότι μια προς εξέταση γλώσσα B *δεν* είναι κανονική. Αυτό το επιτυγχάνουμε:

- υποθέτοντας ότι η B *είναι* κανονική, συνεπώς θα υπάρχει ο φυσικός αριθμός p που χαρακτηρίζει το μήκος της άντλησης
- προσπαθούμε να βρούμε μια *κακιά* μακρουλή λέξη $s = xyz \in B, |s| \geq p$ και να διακρίνουμε τις δυνατές μορφές $x_1y_1, \ldots, x_my_m$, μήκους $\leq p$, που μπορεί να έχει η προς φούσκωμα υπό-λέξη της *κακιάς* λέξης $s \in B$
- φουσκώνουμε i φορές, για κατάλληλο φυσικό i, κάθε μορφή: $x_1y_1^i, \ldots, x_my_m^i$ και δείχνουμε ότι οι φουσκωμένες λέξεις $x_1y_1^iz, \ldots, x_my_m^iz$ δεν ανήκουν στην γλώσσα B (άτοπο).

1.3.1 Δείξτε ότι γλώσσα $B = \{0^n1^n : n \geq 0\}$ είναι μη κανονική

Υποθέτουμε ότι η B είναι κανονική και έστω p το μήκος άντλησης που λόγω του Λήμματος Άντλησης υπάρχει. Πρέπει να βρούμε μια "κακιά" λέξη $s = xyz \in B$, που να είναι αρκετά μακρουλή $|s| > p$, ώστε φουσκώνοντας την i φορές να καταλήξουμε σε λέξη $s(i) \notin B$. Αυτό είναι *άτοπο*, διότι αν B κανονική γλώσσα τότε κάθε μακρουλή λέξη με μήκος $> p$ όσο και να φουσκώσει πρέπει να ανήκει στη B.

Σε αυτή την περίπτωση η κακιά λέξη s είναι εύκολο να ανακαλυφθεί. Λόγω της μορφής των λέξεων της γλώσσας B, όπου έχουμε ίσο-μοίρασμα των 0ς και 1ς, το παραμικρό φούσκωμα υπολέξης θα δημιουργήσει λέξη που καταστρέφει το ισομοίρασμα.

Έστω η κακιά λέξη:

$$s = 0^p1^p \in B, \text{ με } |s| > p$$

Το τμήμα y της s προς φούσκωμα μπορεί να είναι:

- $y = 0\ldots0$ ώστε $0 < |y| = k \leq p$. Δηλαδή το y αποτελείται μόνο από 0ς. Τότε η xy^iz θα περιέχει $(i-1) \times k$ περισσότερα 0΄ς από τα αντίστοιχα 1΄ς, δηλαδή δεν ανήκει στην B για κάθε $i \geq 2$.
- $y = 1\ldots1$ ώστε $0 < |y| = k \leq p$. Δηλαδή το y αποτελείται μόνο από 1΄ς. Τότε η xy^iz θα περιέχει $(i-1) \times k$ περισσότερα 1΄ς από τα αντίστοιχα 0΄ς, δηλαδή δεν ανήκει στην B για κάθε $i \geq 2$. (Ο έλεγχος αυτός είναι καταχρηστικός, διότι $|xy| \leq p$, και άρα υποχρεωτικά $y = 0\ldots0$.) $y = 0\ldots01\ldots1$ ώστε $0 < |y| = k \leq p$. Δηλαδή το y αποτελείται από 0΄ς ακολουθούμενα από 1΄ς.

Τότε η xy^iz θα περιέχει ίσο πλήθος 1′ς και 0′ς, όμως για κάθε $i \geq 2$ δημιουργεί εναλλαγές από 0′ς και 1′ς και συνεπώς δεν ανήκει στην B. (Ο έλεγχος αυτός είναι καταχρηστικός, διότι $|xy| \leq p$, και άρα υποχρεωτικά $y = 0 \ldots 0$.)

1.3.2 Δείξτε ότι γλώσσα $C = \{w : w$ έχει ίσο πλήθος από 0′ς και 1′ς$\}$ είναι μη κανονική

Υποθέτουμε ότι η C είναι κανονική και έστω p το μήκος άντλησης που λόγω του Λήμματος Άντλησης υπάρχει. Πρέπει να βρούμε μια "κακιά" λέξη $s = xyz \in C$, που να είναι αρκετά μακρουλή $|s| > p$, ώστε φουσκώνοντας την i φορές να καταλήξουμε σε λέξη $s(i) \notin C$. Αυτό είναι *άτοπο*, διότι αν C κανονική γλώσσα τότε κάθε μακρουλή λέξη με μήκος $> p$ όσο και να φουσκώσει πρέπει να ανήκει στη C.

Έστω η κακιά λέξη:

$$s = 0^p1^p \in C, \text{ με } |s| > p$$

Στόχος μας είναι να φθάσουμε σε άτοπο, δηλαδή να δείξουμε ότι η λέξη μπορεί να γραφεί $s = xyz$ ώστε η φουσκωμένη λέξη $xy^iz \notin C$ για κάποιο $i \geq 0$. Ένα έμπειρο, μα βιαστικό μάτι, θα παρατηρούσε ότι κάθε φούσκωμα y^i διατηρεί ίσα τα πλήθη των 0′1 και 1′ς στην s, όπως δηλαδή ήταν προ του φουσκώματος. Ούπς! δεν φαντάζει για άτοπο. Όμως, η ιδιότητα 3 του Λήμματος Άντλησης μας λέει ότι $|xy| \leq p$. Άρα η υπό-λέξη y της κακιάς λέξης $s = 0^p1^p$ πρέπει να έχει μόνο 0′ς και συνεπώς το φούσκωμα της σε $s(i)$ θα αυξήσει τα 0′ς και άρα $s(i) \notin C$.

Παρατήρηση. Σε αυτό το σημείο είναι σημαντικό να δείξουμε πόσο σημαντική είναι η σωστή επιλογή της κακιάς λέξης. Αν επιλέξουμε για παράδειγμα την λέξη

$$s' = (01)^p = xyz \in C, \text{ με } |s'| > p$$

Τότε έχουμε τις εξής επιλογές για την προς φούσκωμα υπολέξη y:

1. $y = (01)^j$ για $j > 0$, δηλαδή $s' = (01)^p = \underbrace{(01)^k}_{x}\underbrace{(01)^j}_{y}\underbrace{(01)^t}_{z}$, ώστε $k + j + t = p$ για $j > 0$ και $k, t \geq 0$. Εδώ δεν μπορούμε να πετύχουμε άτοπο, διότι για κάθε $i \geq 0$ η φουσκωμένη υπολέξη $y^i = \left((01)^j\right)^i$ περιέχει ίσο πλήθος από 0ς και 1ς, επειδή η $y = (01)^j$ περιέχει ίσο πλήθος από 0ς και 1ς. Άρα προκύπτει η λέξη $s(i)' = xy^iz$ θα ανήκει στην C για κάθε $i \geq 0$.

2. $y = 1(01)^j$ για $j \geq 0$, δηλαδή $s' = (01)^p = \underbrace{(01)^k0}_{x}\underbrace{1(01)^j}_{y}\underbrace{(01)^t}_{z}$, ώστε $k + j + t = p - 1$ για $j > 0$ και $k, t \geq 0$. Εδώ μπορούμε να πετύχουμε άτοπο. Διότι για κάθε $i \geq 2$ η φουσκωμένη λέξη $y^i = \left(1(01)^j\right)^i$ δεν περιέχει ίσο πλήθος από 0ς και 1ς, επειδή η $y = 1(01)^j$ δεν περιέχει ίσο πλήθος από 0ς και 1ς. Άρα η λέξη $s(i)' = xy^iz$ δεν θα ανήκει στην C για κάθε $i \geq 2$.

3. $y = 1(01)^j0$ για $j \geq 0$, δηλαδή $s' = (01)^p = \underbrace{(01)^k0}_{x}\underbrace{1(01)^j0}_{y}\underbrace{1(01)^t}_{z}$, ώστε $k+j+t = p-2$ για $j > 0$ και $k, t \geq 0$. Εδώ δεν μπορούμε να πετύχουμε άτοπο, διότι για κάθε $i \geq 0$ η φουσκωμένη υπολέξη $y^i = \left(1(01)^j0\right)^i$ περιέχει ίσο πλήθος από 0ς και 1ς, επειδή η $y = 1(01)^j0$ περιέχει ίσο πλήθος από 0ς και 1ς. Άρα λέξη $s(i)' = xy^iz$ θα ανήκει στην C για κάθε $i \geq 0$.

4. $y = (01)^j0$ για $j \geq 0$, δηλαδή $s' = (01)^p = \underbrace{(01)^k}_{x}\underbrace{(01)^j0}_{y}\underbrace{1(01)^t}_{z}$, ώστε $k + j + t = p - 1$ για $j > 0$ και $k, t \geq 0$. Εδώ μπορούμε να πετύχουμε άτοπο. Διότι για κάθε $i \geq 2$ η φουσκωμένη λέξη $y^i = \left((01)^j0\right)^i$ δεν περιέχει ίσο πλήθος από 0ς και 1ς, επειδή η $y = (01)^j0$ δεν περιέχει ίσο πλήθος από 0ς και 1ς. Άρα η λέξη $s(i)' = xy^iz$ δεν θα ανήκει στην C για κάθε $i \geq 2$.

Στις περιπτώσεις 2 και 4 πετυχαίνουμε να φτάσουμε σε άτοπο, γιατί το y έχει άνισο αριθμό 0ς και 1ς . Στις άλλες δύο περιπτώσεις δεν μπορούμε να πετύχουμε άτοπο. Δεν μπορούμε λοιπόν να εγγυηθούμε ότι κάποια από αυτές τις περιπτώσεις δεν θα ισχύσει και κατά συνέπεια δεν είμαστε σε θέση να ολοκληρώσουμε την απόδειξη χρησιμοποιώντας αυτή την επιλογή της κακιάς λέξης.

1.3.3 Δείξτε ότι γλώσσα $F = \{ww : \ w \in \{0,1\}^*\}$ είναι μη κανονική

Υποθέτουμε ότι η F είναι κανονική και έστω p το μήκος άντλησης που λόγω του Λήμματος Άντλησης υπάρχει. Πρέπει να βρούμε μια ''κακιά'' λέξη $s = xyz \in F$, που να είναι αρκετά μακρουλή $|s| > p$, ώστε φουσκώνοντας την i φορές, για κάποιο κατάλληλο i, να καταλήξουμε σε λέξη $s(i) \notin F$. Αυτό είναι *άτοπο*, διότι υποθέσαμε ότι η F είναι κανονική γλώσσα και τότε κάθε μακρουλή λέξη με μήκος $> p$ όσο και να φουσκώσει πρέπει να ανήκει στη F.

Έστω η κακιά λέξη[1]:

$$s = 0^p10^p1 \in F, \text{ με } |s| > p$$

Στόχος μας είναι να φθάσουμε σε άτοπο, δηλαδή να δείξουμε ότι η λέξη μπορεί να γραφεί $s = xyz$ ώστε η φουσκωμένη λέξη $xy^iz \notin C$ για κάποιο $i \geq 0$. Επειδή η συνθήκη 3 του Λήμματος 'Αντλησης λέει ότι $|xy| \leq p$ συμπεραίνουμε ότι η προς φούσκωση υπό-λέξη y αποτελείται μόνο από 0΄ς. Συνεπώς για $i = 2$ η φουσκωμένη λέξη $xyyz \notin F$.

1.3.4 Δείξτε ότι γλώσσα $D = \{1^{n^2} : \ n \geq 0\}$ είναι μη κανονική

Υποθέτουμε ότι η D είναι κανονική και έστω p το μήκος άντλησης που λόγω του Λήμματος Άντλησης υπάρχει. Πρέπει να βρούμε μια ''κακιά'' λέξη $s = xyz \in D$, που να είναι αρκετά μακρουλή $|s| > p$, ώστε φουσκώνοντας την i φορές, για κάποιο κατάλληλο i, να καταλήξουμε σε λέξη $s(i) \notin D$. Αυτό είναι *άτοπο*, διότι υποθέσαμε ότι η D είναι κανονική γλώσσα και τότε κάθε μακρουλή λέξη με μήκος $> p$ όσο και να φουσκώσει πρέπει να ανήκει στη D.

Το κλειδί για να κατασκευάσουμε την κακιά λέξη $s = xyz$ είναι να παρατηρήσουμε πως κλιμακώνονται τα μήκη των λέξεων της D. Τα μήκη κλιμακώνονται τετραγωνικά, δηλαδή:

$$0, 1 = 1^2, 4 = 2^2, 9 = 3^2, 16 = 4^2, 25 = 5^2, 36 = 6^2, 49 = 7^2, \ldots, n^2, (n+1)^2, \ldots,$$

Παρατηρούμε ότι όσο μεγαλώνει το τρέχον μήκος n μιας λέξης, οι λέξεις του αμέσως μεγαλύτερου μήκους $(n+1)^2$ έχουν μεγάλη διαφορά από τις αντίστοιχες του αμέσως μικρότερου μήκους. Με άλλα λόγια, το $(n+1)^2$ είναι πολύ μεγάλο σε σχέση με το n^2. Λαμβάνοντας υπόψιν αυτό το φαινόμενο, στόχος μας είναι καταφέρουμε να βρούμε κατάλληλο i για μια κακιά λέξη $s = xyz \in D$, ώστε η $s(i) = xy^iz$ να έχει τετραγωνικό μήκος $|s(i)| = n^2$ μα αν τη φουσκώσουμε 1 φορά ακόμη η $s(i+1) = xy^{i+1}z$ να αποκτήσει μήκος $|s(i+1)| < (n+1)^2$ και συνεπώς να μην ανήκει στην D. Δηλαδή για να αποτύχει το φούσκωμα κατά 1 φορά πρέπει να ισχύει η ανίσωση:

$$\begin{aligned} \underbrace{|xy^{i+1}z|}_{\text{μήκος: } n^2+|y|} - \underbrace{|xy^iz|}_{\text{μήκος: } n^2} &= |y| \\ &\text{οφείλει να είναι} \quad < \\ (n+1)^2 - n^2 = 2n+1 &= 2\sqrt{\underbrace{|xy^iz|}_{\text{μήκος: } n^2}} + 1 \end{aligned} \tag{1.1}$$

---

[1] Το ''κουμπί'' είναι η έξυπνη επιλογή της κακιάς λέξης. Για παράδειγμα η υποψήφια κακιά λέξη $s' = 0^p0^p \in F$ όσο και να φουσκώσει παραμένει στην F.

Παρατηρούμε ότι αν η κακιά λέξη είναι $s = xyz = 1^{p^2}$ με μήκος $|s| = p^2$, από συνθήκη 3 του Λήμματος Άντλησης έχουμε $|y| \leq |s| = p^2$. Συνεπώς, για να ισχύει η Ανίσωση (1.1), πρέπει να θέσουμε $i = p^4$ ώστε το $2\sqrt{|xy^iz|} + 1$ να υπερβεί το p^2 και κατ΄ επέκταση το $|y|$.

1.3.5 Δείξτε ότι γλώσσα $E = \{0^i1^j :\ i > j\}$ είναι μη κανονική

Σε αυτό το παράδειγμα, αντί να φουσκώσουμε τη κακιά λέξη $s = xyz = 0^{p+1}1^p$, θα τη ξεφουσκώσουμε. Επειδή $|xy| \leq p$ το τμήμα y θα αποτελείται μόνο από 0΄ς. Οπότε για κάθε $i > 0$ η φουσκωμένη $s(i) = xy^iz$ πάντα θα έχει περισσότερα 0΄ς από 1΄ς και συνεπώς θα ανήκει στην γλώσσα E. Ένα έμπειρο μάτι παρατηρεί ότι για $i = 0$ η ξεφουσκωμένη $s(0) = xz \notin E$ (γιατί;;;).

1.3.6 Δείξτε ότι γλώσσα $A = \{a(ab)^nc^n : n \geq 0\}$ δεν είναι κανονική

Έστω ότι η A είναι κανονική και p το μήκος άντλησης. Θεωρούμε τη λέξη

$$s = a(ab)^pc^p \in A$$

Επειδή $|s| = 3p + 1 \geq p$, σύμφωνα με το Λήμμα Άντλησης υπάρχουν x, y, z έτσι ώστε να ισχύουν οι ιδιότητες (i) $s = xyz$ (ii) $0 < |y| \leq |xy| \leq p$ και (iii) για κάθε $i \in \mathbb{N}$, $xy^iz \in A$. Επειδή $|xy| \leq p$, η συμβολοσειρά xy δεν περιέχει το χαρακτήρα c. Σχηματικά:

$$s = \underbrace{a(ab)^p}_{xy\subseteq}\underbrace{c^p}_{\subseteq z}$$

Διακρίνουμε περιπτώσεις για την μορφή του xy ώστε να αντιληφθούμε την μορφή του y.

1. Αν $xy = a$. Τότε $x = \varepsilon$ είναι η κενή συμβολοσειρά και $y = a$, δηλαδή:

$$s = \underbrace{\epsilon}_{x}\underbrace{a}_{y}\underbrace{(ab)^pc^p}_{z}$$

   Αλλά τότε η λέξη $s(0) = xy^0z = (ab)^pc^p \notin A$.

2. Αν $xy = a(ab)^q,\ q \leq p$. Τότε $s = \underbrace{a(ab)^q}_{xy}\underbrace{(ab)^{p-q}c^p}_{z}$ και διακρίνουμε τις υπο-περιπτώσεις

   - $x = a$. Τότε

$$s = \underbrace{a}_{x}\ \underbrace{(ab)^q}_{y:|y|=2q>0}\ \underbrace{(ab)^{p-q}c^p}_{z}$$

     Φουσκώνουμε για $i = 2$ και λαμβάνουμε

$$s(2) = xy^2z = a(ab)^{2q}(ab)^{p-q}c^p$$

     Παρατηρούμε ότι για να ανήκει η $s(2)$ στην A θα πρέπει να είναι της μορφής $a(ab)^pc^p$, δηλαδή θα πρέπει να ισχύει ότι $p = 2q + p - q \Rightarrow q = 0$, άτοπο διότι $|y| = 2q > 0$.

   - $x = a(ab)^t,\ 0 < t < q$. Τότε

$$s = \underbrace{a\overbrace{ab\ldots ab}^{(ab)^t}}_{x}\ \underbrace{(ab)^{q-t}}_{y:|y|=2(q-t)>0}\ \underbrace{(ab)^{p-q}c^p}_{z}$$

     Για να ανήκει στη γλώσσα A η $s(0) = xy^0z = a(ab)^t(ab)^{p-q}c^p$ θα πρέπει να είναι της μορφής $a(ab)^pc^p$, δηλαδή θα πρέπει να ισχύει $p = t + p - q \Rightarrow t = q$, οπότε $y = (ab)^{q-t} = \varepsilon$, άτοπο, διότι $|y| > 0$.

- $x = a(ab)^t a,\ 0 < t < q$. Τότε

$$s = \underbrace{a\,\overbrace{ab\ldots aba}^{(ab)^t a}}_{x}\ \underbrace{b(ab)^{q-t-1}}_{y:|y|=2(q-t)-1>0}\ \underbrace{(ab)^{p-q}c^p}_{z}$$

Όμως η λέξη $s(0) = xy^0z = a(ab)^t a(ab)^{p-q}c^p$ δεν μπορεί να ανήκει στη γλώσσα A.

3. Αν $xy = a(ab)^q a$ και $z = b(ab)^{p-q-1}c^p$, σχηματικά:

$$s = \underbrace{a(ab)^q a}_{xy}\underbrace{b(ab)^{p-q-1}c^p}_{z}$$

Διακρίνουμε πάλι τις παρακάτω υπο-περιπτώσεις.

- Αν $x = a$, τότε εργαζόμαστε όπως προηγούμενα.
- Αν $x = a(ab)^t,\ 0 < t < q$ οπότε $y = (ab)^{q-t}a,\ z = b(ab)^{p-q-1}c^p$, σχηματικά:

$$s = \underbrace{a(ab)^t}_{x}\underbrace{(ab)^{q-t}a}_{y}\underbrace{b(ab)^{p-q-1}c^p}_{z}$$

Παρατηρούμε ότι δεν μπορεί να ανήκει στην A η λέξη $s(0) = xy^0z = a(ab)^t b(ab)^{p-q-1}c^p$.

- Αν $x = a(ab)^t a,\ 0 < t < q$ οπότε $y = b(ab)^{q-t-1}a,\ z = b(ab)^{p-q-1}c^p$, σχηματικά:

$$s = \underbrace{a(ab)^t a}_{x}\underbrace{b(ab)^{q-t-1}a}_{y}\underbrace{b(ab)^{p-q-1}c^p}_{z}$$

Παρατηρούμε ότι για να ανήκει στην A η λέξη $s(0) = xy^0z = a(ab)^t ab(ab)^{p-q-1}c^p$ θα πρέπει να έχουμε $p = p - q - 1 + t + 1$ δηλαδή $q = t$, πράγμα που είναι σε αντίθεση με την υπόθεση της υπο-περίπτωσης που εξετάζουμε.

Συμπεραίνουμε ότι η A δεν είναι κανονική.

1.3.7 Έστω L η γλώσσα στο αλφάβητο $\{a, b\}$ που ορίζεται αναδρομικά ως εξής:

- $L_0 = \{ab\}$
- $L_{n+1} = L_n \cup \{bxx | x \in L_n\} \cup \{xabx | x \in L_n\}$
- $L = \bigcup_n L_n$

Δείξτε ότι η L δεν είναι κανονική.

Υποθέτουμε ότι η L είναι κανονική και έστω p το μήκος άντλησης. Θα προσπαθήσουμε να βρούμε ''κακιά'' λέξη s που παράγεται από τους κανόνες της γλώσσας L ώστε αν φουσκωθεί να μην ανήκει στην γλώσσα και να καταλήξουμε σε άτοπο. Έστω ότι η κακιά λέξη είναι της μορφής

$$s = (ab)^{s_p}$$

όπου $s_0 = 1, s_{n+1} = 2s_n + 1$ και $s \in L$ (δηλαδή η s παράγεται από το ab με διαδοχικές εφαρμογές του κανόνα $xabx$) και το μήκος της είναι

$$|s| = 2s_p > s_p \geq p$$

Η λέξη θα αποτελείται από τα μέρη:

$$s = xyz,\ 0 < |y| \leq |xy| \leq p$$

όπου κάθε μέρος είναι δυνατόν να έχει την μορφή:

$$s = \underbrace{\underbrace{(ab)^r a^t}_{x}\ \underbrace{b^t (ab)^{m-r-t} a^j}_{y}}_{|xy|=2m+j\leq p}\ \underbrace{b^j (ab)^{s_p - m - j}}_{z}$$

με $0 \leq r \leq m, 0 \leq t, j \leq 1$. Διαισθητικά, για να καταλάβουμε την άνω μορφή, αρκεί να δούμε ότι το προς φούσκωμα μέρος y μπορεί να είναι $T_1 ab \ldots ab T_2$ όπου $T_1 \in \{\epsilon, b\}$ και $T_2 \in \{\epsilon, a\}$. Θεωρούμε τη φουσκωμένη λέξη

$$s(2) = xy^2z = \underbrace{(ab)^r a^t}_{x}\ \underbrace{b^t (ab)^{m-r-t} a^j}_{y}\ \underbrace{b^t (ab)^{m-r-t} a^j}_{y}\ \underbrace{b^j (ab)^{s_p - m - j}}_{z}$$

Η $s(2)$ δεν είναι της μορφής bxx, ούτε μπορεί να είναι της μορφής $wabw$ με w της μορφής bvv, γιατί όλες αυτές οι λέξεις αρχίζουν με b. Επομένως η $s(2)$ μπορεί μόνο να παράγεται από το ab με διαδοχικές εφαρμογές του κανόνα $xabx$. Επομένως πρέπει

$$t = j \in \{0, 1\}$$

Κάνοντας με προσοχή της πράξεις στην άνω μορφή έχουμε:

$$s(2) = (ab)^{s_p + (m-r)} = (ab)^{s_n}$$

για κάποιο $n > p$, $n = p + i$, δηλαδή $s_n = s_{p+i} \geq s_{p+1}$ (όπου $r < m$, αλλιώς $y = \varepsilon$) άρα:

$$s_{p+1} \leq s_n = s_p + (m - r) \leq s_p + m < s_p + p \leq 2s_p < 2s_p + 1 = s_{p+1}$$

που είναι άτοπο.

1.3.8 Έστω L η γλώσσα στο αλφάβητο $\{a, b\}$ που ορίζεται αναδρομικά ως εξής:

1. $a \in L$
2. Αν $x \in L \Rightarrow xa^2bx \in L$
3. $x \in L$ μόνο από τα 1, 2, άνω.

Είναι η L κανονική;

Εύκολα βρίσκουμε ότι οι λέξεις της γλώσσας L έχουν την γενική μορφή:

$$L = \{(a^3b)^{2^n - 1} a | n \geq 0\}$$

Υποθέτουμε ότι η L είναι κανονική και έστω p το μήκος άντλησης. Θα προσπαθήσουμε να βρούμε ‘‘κακιά’’ λέξη s που παράγεται από τους κανόνες της γλώσσας L ώστε αν φουσκωθεί να μην ανήκει στην γλώσσα και να καταλήξουμε σε άτοπο. Έστω ότι η κακιά λέξη είναι της μορφής

$$s = (a^3b)^{2^p - 1} a, \text{ με } |s| \geq p$$

Θα δείξω ότι δεν μπορεί να γραφεί ως

$$s = xyz,\ 0 < |y| \leq |xy| \leq p$$

ώστε η λέξη

$$s(2) = xyyz \in L$$

Παρατηρώ ότι

$$|s| = 4 \times (2^p - 1) + 1 = 4 \cdot 2^p - 3$$

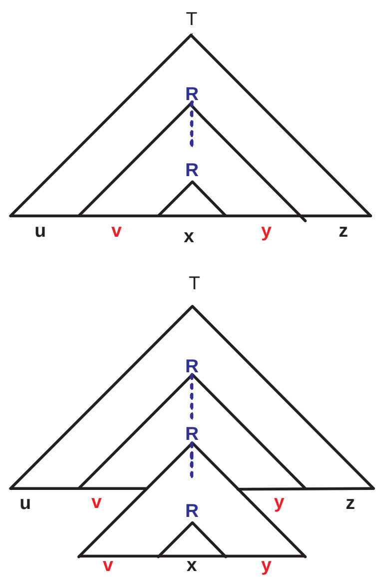
Οποιαδήποτε λέξη $x \in L$ με μήκος $|x| > |s|$ πρέπει να έχει μήκος τουλάχιστον όσο το αμέσως μεγαλύτερο μήκος από το $|s| = 4 \cdot 2^p - 3$, δηλαδή μήκος $|x| \geq 4 \cdot 2^{p+1} - 3$. Όμως η $s(2)$ αντιβαίνει αυτό τον κανόνα, διαισθητικά, αυξάνει λίγο σε σχέση με την s. Διότι παρόλο που $|s(2)| > |s|$ έχουμε $|s(2)| < 4 \cdot 2^{p+1} - 3$. Για να το δούμε αυτό, αρκεί να παρατηρήσουμε ότι $|s(2)| = |s| + |y|$ όπου $|y| \leq p$ και άρα $|s(2)| \leq 4 \cdot 2^p - 3 + p < 2^{p+1} - 3$. Άτοπο.

Κεφάλαιο 2

Το Λήμμα της Άντλησης για γλώσσες μη ανεξάρτητες συμφραζομένων

2.1 Διαισθητική παρουσίαση του Λήμματος με Δέντρο Παραγωγής

Ας παρατηρήσουμε το άνω και κάτω δέντρο του Σχήματος 2.1. Το άνω δέντρο με κορυφή τη μεταβλητή T



Σχήμα 2.1: Τα φύλλα του άνω δέντρου παραγωγής σχηματίζουν την λέξη $s = uvxyz$. Τα φύλλα του κάτω δέντρου παραγωγής σχηματίζουν την λέξη $s(2) = uv^2xy^2z$.

σχηματίζει στα φύλλα του τη λέξη $uvxyz$. Το υποδέντρο του με κορυφή τη πρώτη εμφάνιση της μεταβλητής

R σχηματίζει στα φύλλα του τη λέξη vxy (παρατηρήστε ότι το υποδέντρο αυτό δεν έχει στα φύλλα του τις υπολέξεις u και z). Πιο χαμηλά εμφανίζεται η δεύτερη εμφάνιση της μεταβλητής R και δημιουργεί υποδέντρο με τη λέξη x στα φύλλα του (παρατηρήστε ότι το υποδέντρο αυτό δεν έχει στα φύλλα του τις υπολέξεις v και y). Αν στο άνω δέντρο, ακριβώς μετά από τη δεύτερη εμφάνιση της R, τοποθετήσουμε όλο το υποδέντρο που υπάρχει ακριβώς μετά την πρώτη εμφάνιση της R, τότε θα λάβουμε το κάτω δέντρο. Με αυτό το τρόπο στο κάτω δέντρο στα φύλλα του εμφανίζονται οι υπολέξεις: $uvvxyyz = uv^2xy^2z$.

2.2 Παραδείγματα με το Λήμμα της Άντλησης

2.2.1 Δείξτε ότι γλώσσα $B = \{a^n b^n c^n : \ n \geq 0\}$ δεν είναι ανεξάρτητη συμφραζομένων.

Έστω ότι η B είναι ΓΑΣ και έστω ο φυσικός p που είναι το μήκος της άντλησης. Επιλέγω για υποψήφια κακιά λέξη την:

$$s = a^p b^p c^p = uvxyz \in B$$

και θα δείξω ότι για κατάλληλο i δεν μπορεί να φουσκωθεί, καταλήγοντας σε άτοπο.

- Αν τα v, y περιέχουν μόνο ένα είδος γραμμάτων δεν μπορούν να περιέχουν ταυτόχρονα τα a, b ή b, c. Συνεπώς η uv^2xy^2z δεν μπορεί να περιέχει το ίδιο πλήθος από a, b, c.
- Αν τα v, y περιέχουν δύο είδη γραμμάτων η uv^2xy^2z μπορεί να περιέχει το ίδιο πλήθος από a, b, c, όμως δεν μπορεί να τα εμφανίζει με την σωστή σειρά (δηλαδή θα δημιουργούνται μη επιτρεπόμενες εναλλαγές από γράμματα).

2.2.2 Δείξτε ότι γλώσσα $C = \{a^i b^j c^k : \ 0 \leq i \leq j \leq k\}$ δεν είναι ανεξάρτητη συμφραζομένων.

Έστω ότι η C είναι ΓΑΣ και έστω ο φυσικός p που είναι το μήκος της άντλησης. Επιλέγουμε για υποψήφια κακιά λέξη την:

$$s = a^p b^p c^p = uvxyz \in C$$

και θα δείξω ότι για κατάλληλο i δεν μπορεί να φουσκωθεί, καταλήγοντας σε άτοπο.

1. Έστω ότι τα v, y περιέχουν μόνο ένα είδος γραμμάτων, δηλαδή είναι της μορφής $a \ldots a$ ή $b \ldots b$ ή $c \ldots$. Συνεπώς ένα είδος γραμμάτων από τα a, b, c δεν μπορεί να εμφανίζεται ταυτόχρονα στα v και y.
    - Αν το είδος a δεν εμφανίζεται ταυτόχρονα στα v και y, τότε η λέξη $s(0) = uv^0xy^0z = xyz$ θα εμφανίζει ίδιο πλήθος a με την s όμως τα b και c οφείλουν να μειωθούν.
    - Αν το είδος b δεν εμφανίζεται ταυτόχρονα στα v και y, τότε το είδος a η b πρέπει να εμφανίζεται στα v, y, διότι $|vy| > 0$. Αν το είδος c εμφανίζεται στα v, y, τότε η λέξη $s(0) = uv^0xy^0z$ περιέχει περισσότερα b από c και άρα δεν ανήκει στην C. Αν το είδος a εμφανίζεται, τότε η λέξη $s(2) = uv^2xy^2z$ περιέχει περισσότερα a από b και άρα δεν ανήκει στην C.
    - Αν το είδος c δεν εμφανίζεται ταυτόχρονα στα v και y, τότε η λέξη $s(2) = uv^2xy^2z$ περιέχει περισσότερα a ή b από c, και άρα δεν ανήκει στην C.
2. Αν τα v, y περιέχουν περισσότερο του ενός είδους γραμμάτων τότε η uv^2xy^2z μπορεί να περιέχει το ίσο πλήθος από a, b, c, όμως δεν μπορεί να τα εμφανίζει με την σωστή σειρά (δηλαδή θα δημιουργούνται μη επιτρεπόμενες εναλλαγές από γράμματα).

2.2.3 Δείξτε ότι γλώσσα $D = \{ww : \ w \in \{0,1\}^*\}$ δεν είναι ανεξάρτητη συμφραζομένων.

Έστω ότι η D είναι ΓΑΣ και έστω ο φυσικός p που είναι το μήκος της άντλησης. Επιλέγουμε για υποψήφια κακιά λέξη την:

$$s' = 0^p10^p1 = uvxyz \in C$$

και θα δείξω ότι για κατάλληλο i δεν μπορεί να φουσκωθεί, καταλήγοντας σε άτοπο. Όμως η λέξη s μπορεί να φουσκωθεί, αν γραφεί ως εξής:

$$\overbrace{\underbrace{0\ldots0}_{u}\underbrace{0}_{v}\underbrace{1}_{x}}^{0^p1}\overbrace{\underbrace{0}_{y}\underbrace{0\ldots0}_{z}}^{0^p1}$$

Πρέπει να βρούμε οπωσδήποτε άλλη κακιά λέξη. Ας δοκιμάσουμε το:

$$s = 0^p1^p0^p1^p = uvxyz \in C$$

και ας προσέξουμε την συνθήκη $|vxy| \leq p$ του Λήμματος.

1. Έστω ότι το τμήμα vxy εμφανίζεται στο αριστερό μισό τμήμα της s. Τότε η φουσκωμένη λέξη $s(2) = uv^2xy^2z$ σπρώχνει ένα 1 στη πρώτη θέση του δεύτερου μισού της $s(2)$, άρα η $s(2)$ δεν μπορεί να είναι της μορφής ww.

2. Έστω ότι το τμήμα vxy εμφανίζεται στο δεξιό μισό τμήμα της s. Τότε η φουσκωμένη λέξη $s(2) = uv^2xy^2z$ σπρώχνει ένα 0 στη τελευταία θέση του πρώτου μισού της $s(2)$, άρα η $s(2)$ δεν μπορεί να είναι της μορφής ww.

3. Τέλος, έστω ότι το τμήμα vxy εμφανίζεται κεντρικά στην s. Τότε η λέξη $s(0) = uv^0xy^0z$ έχει μορφή $0^p1^i0^j1^p$, όπου τα i, j δεν μπορούν να είναι ταυτόχρονα ίσα με p. Άρα η $s(0)$ δεν μπορεί να είναι της μορφής ww.