

Ερώτηση: 1 (Βαθμός: -0.50 / 1.00)

(Πολλαπλής Επιλογής (Μοναδική Απάντηση))

Υπολογιστική Γλωσσολογία ονομάζεται η επιστημονική περιοχή μεταξύ:

Απάντηση



Γλωσσολογίας και Ιατρικής
(λάθος)



Ιατρικής και Πληροφορικής
(λάθος)



Πληροφορικής και Γλωσσολογίας
(σωστό)

(Σχεδόν) συνώνυμοι όροι με την Υπολογιστική Γλωσσολογία είναι οι εξής:

Απάντηση



Γλωσσικός Υπολογισμός
(λάθος)



Επεξεργασία Φυσικής Γλώσσας
(σωστό)



Φυσική Επεξεργασία Γλώσσας
(λάθος)



Γλωσσική Τεχνολογία
(σωστό)



Τεχνική Γλωσσολογία
(λάθος)


ACL είναι τα αρχικά του οργανισμού, το πλήρες όνομα του οποίου είναι:

Απάντηση

- ☒ ☒ American Computational Linguistics
(λάθος)
- ☐ Association for Computational Linguistics
(σωστό)
- ☐ Association for Computer Languages
(λάθος)
- ☐ Administration for Computer Languages
(λάθος)

Η κωδικοποίηση Unicode...

Απάντηση

- ☒  Μπορεί να αναπαραστήσει πάνω από 1.000.000 διαφορετικούς χαρακτήρες.
(σωστό)
- ☐ Περιλαμβάνει 17 επίπεδα με πάνω από 60.000 code points το καθένα.
(σωστό)
- ☐ Μπορεί να χρησιμοποιηθεί για την συνύπαρξη χαρακτήρων από περισσότερες από δύο γλώσσες στο ίδιο κείμενο.
(σωστό)
- ☐ Αφορά μόνο αριθμητικά ψηφία.
(λάθος)
- ☐ Είναι παλαιότερη από την κωδικοποίηση ASCII.
(λάθος)

Ο δεκαεξαδικός αριθμός D8 στον κώδικα ISO 8859-7 αντιστοιχεί στον χαρακτήρα:

Απάντηση



Χ

(λάθος)



Θ

(λάθος)



Ω

(λάθος)



Ξ

(λάθος)




Κανένα από τα παραπάνω.

(σωστό)

Η κωδικοποίηση των Αγγλικών χαρακτήρων είναι ακριβώς ίδια και στον κώδικα ASCII και στους κώδικες ISO 8859 και στους κώδικες ANSI (Windows).

Απάντηση

- ☒  Σωστό (σωστό)
- ☐ Λάθος (λάθος)

Στην αρχή της κωδικοσελίδας CP 1253, υπάρχει ο κώδικας ASCII.

Απάντηση



Σωστό (σωστό)



Λάθος (λάθος)

Στην αρχή του κώδικα ISO 8859, υπάρχει ο κώδικας ASCII.

Απάντηση



Σωστό (σωστό)



Λάθος (λάθος)

Ο χαρακτήρας F στον κώδικα ASCII έχει δεκαεξαδική αναπαράσταση:

Απάντηση



46

(σωστό)



64

(λάθος)



F

(λάθος)



7F

(λάθος)



Κανένα από τα παραπάνω.

(λάθος)

Ο δεκαεξαδικός αριθμός 65 στον κώδικα ASCII αντιστοιχεί στον χαρακτήρα:

Απάντηση

☐

F

(λάθος)

☒ ✖

\$

(λάθος)

☐

x

(λάθος)

☐

e

(σωστό)

☐

Κανένα από τα παραπάνω.

(λάθος)

Η απόσταση Levenshtein μεταξύ των λέξεων ΚΑΛΥΚΑ και ΑΚΥΚΛΑ είναι:

Απάντηση

<input checked="" type="checkbox"/> 	1 (λάθος)
<input type="checkbox"/>	2 (λάθος)
<input type="checkbox"/>	3 (σωστό)
<input type="checkbox"/>	4 (λάθος)
<input type="checkbox"/>	5 (λάθος)

Το πλήθος των βέλτιστων μονοπατιών στον πίνακα Levenshtein για τον υπολογισμό της απόστασης μεταξύ των λέξεων ΗΛΙΚΙΑ και ΛΙΑΚΑΔΑ είναι:

Απάντηση


- ☒ ☒
- 1
- (λάθος)
- ☐
- 2
- (σωστό)
- ☐
- 3
- (λάθος)
- ☐
- 4
- (λάθος)
- ☐
- Μεγαλύτερο από 4.

Το πλήθος των βέλτιστων μονοπατιών στον πίνακα Levenshtein για τον υπολογισμό της απόστασης μεταξύ των λέξεων ΥΛΙΚΑ και ΗΛΙΚΙΑ είναι:

Απάντηση	
<input checked="" type="checkbox"/> ✓	1 (σωστό)
<input type="checkbox"/>	2 (λάθος)
<input type="checkbox"/>	3 (λάθος)
<input type="checkbox"/>	4 (λάθος)
<input type="checkbox"/>	Μεγαλύτερο από 4. (λάθος)

Έστω δυο λέξεις A και B που απέχουν 3 μονάδες (πλήθος διορθώσεων) σύμφωνα με την απόσταση Levenshtein, και μια τρίτη λέξη Γ που απέχει 5 από την B. Πόσο μπορεί να απέχει η Γ από την A;

Απάντηση

- ☒ 
- 4
- (λάθος)
- ☐
- Μεταξύ 3 και 5
- (λάθος)
- ☐
- Μεταξύ 2 και 8
- (σωστό)
- ☐
- Κανένα από τα παραπάνω.
- (λάθος)

Έστω δυο λέξεις A και B που απέχουν 4 μονάδες (πλήθος διορθώσεων) σύμφωνα με την απόσταση Levenshtein, και μια τρίτη λέξη Γ που απέχει 1 από την B. Πόσο μπορεί να απέχει η Γ από την A;

Απάντηση

☒  Μεταξύ 3 και 5
(σωστό)

☐ Μεταξύ 1 και 4
(λάθος)

☐ 6
(λάθος)

☐ Κανένα από τα παραπάνω.
(λάθος)

Έστω η κανονική έκφραση: $/a[(bc)(cb)a]/$

Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;

Απάντηση

☒  abc
(σωστό)

☐ aba
(λάθος)

☐ acb
(σωστό)

☐ aca
(λάθος)

☐ cba
(λάθος)

☐ aa
(σωστό)

Έστω η κανονική έκφραση: $/(ab)^+[ca]^*/$

Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;

Απάντηση



bbca

(λάθος)



aba

(σωστό)



cacaca

(λάθος)



abca

(σωστό)



ababcca

(σωστό)



aabc

(λάθος)

Έστω η κανονική έκφραση: $/a(bc)?b[bc]c?/$

Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;

Απάντηση



acbc

(λάθος)



abcc

(σωστό)



abccb

(λάθος)



abcbb

(σωστό)



abcbc

(σωστό)



abb

(σωστό)

Έστω η κανονική έκφραση: $/(abc)^+[a(ab)c]?/$

Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;

Απάντηση



abccac

(λάθος)



abcabc

(σωστό)



aabcc

(λάθος)



abcab

(σωστό)



aababc

(λάθος)



abcabcc

(σωστό)

Επιλέξτε όσες από τις παρακάτω κανονικές εκφράσεις αναγνωρίζουν τη συμβολοσειρά: **ababb**

Απάντηση

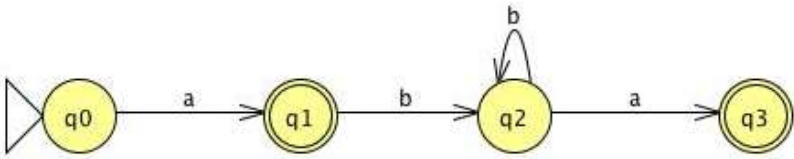
☒  $/(ab)^+b/$
(σωστό)

☐ $/c^*(ab)^*b/$
(σωστό)

☐ $/ab?ab^+/$
(σωστό)

☐ $/a(ba)^+b/$
(λάθος)

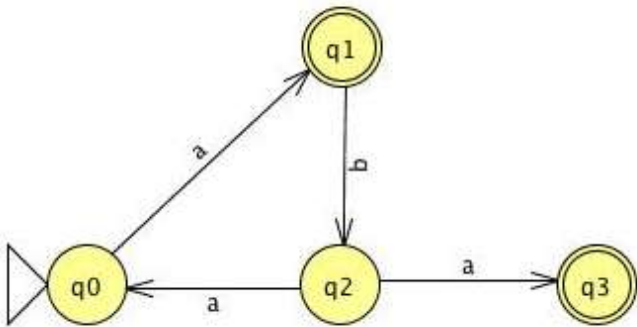
Το παρακάτω αυτόματο είναι:



Απάντηση

- ☒ ☒ Κυκλικό
(σωστό)
Υπάρχει μια κυκλική μετάβαση από την κατάσταση q2 στον εαυτό της,
- ☐ Ακυκλικό
(λάθος)
- ☐ Αιτιοκρατικό
(σωστό)
- ☐ Μη-αιτιοκρατικό
(λάθος)

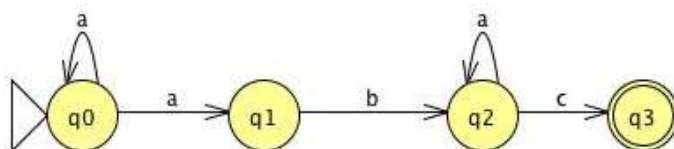
Το παρακάτω αυτόματο είναι:



Απάντηση

- | | |
|---------------------------------------|--|
| <input checked="" type="checkbox"/> ✓ | Κυκλικό
(σωστό) |
| <input type="checkbox"/> | Ακυκλικό
(λάθος) |
| <input type="checkbox"/> | Αιτιοκρατικό
(λάθος) |
| <input type="checkbox"/> | Μη-αιτιοκρατικό
(σωστό)
Υπάρχουν δύο μεταβάσεις από την q2 με το ίδιο σύμβολο. |

Με ποιά από τις παρακάτω κανονικές εκφράσεις είναι ισοδύναμο το ακόλουθο αυτόματο πεπερασμένων καταστάσεων;



Απάντηση

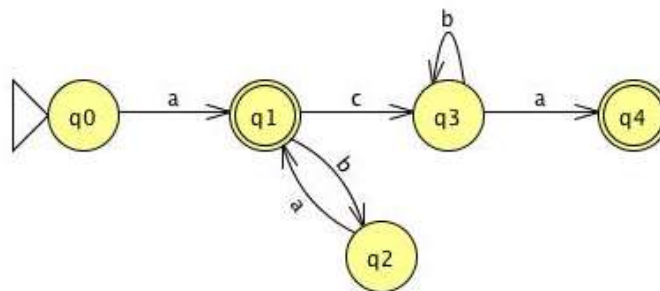
☒ ☒ /a+b*ac/
(λάθος)

☐ /a*ba*c/
(λάθος)

☐ /a*ba+c/
(λάθος)

☐ /a+ba*c/
(σωστό)

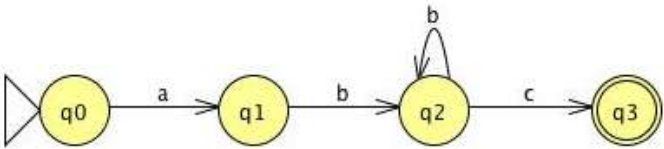
Με ποιά από τις παρακάτω κανονικές εκφράσεις είναι ισοδύναμο το ακόλουθο αυτόματο πεπερασμένων καταστάσεων;



Απάντηση

- | | |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <code>/a(ba)*(cb*a)?/</code>
(σωστό) |
| <input type="checkbox"/> | <code>/a(ba)*cb*a/</code>
(λάθος) |
| <input type="checkbox"/> | <code>/aba+[cb]+a/</code>
(λάθος) |
| <input type="checkbox"/> | <code>/a(b+a)+(c*b*a)?/</code>
(λάθος) |

Με ποιά από τις παρακάτω κανονικές εκφράσεις είναι ισοδύναμο το ακόλουθο αυτόματο πεπερασμένων καταστάσεων;



Απάντηση

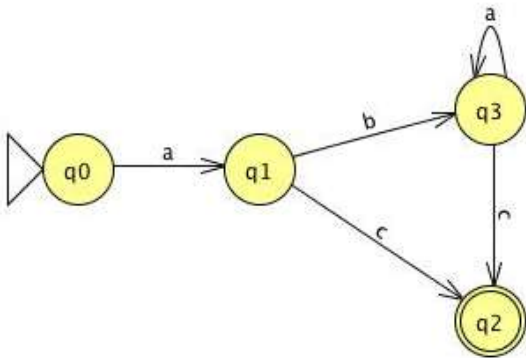
☒ ☒ /abb+c/
(λάθος)

☐ /ab*c/
(λάθος)

☐ /ab+c/
(σωστό)

☐ /ab*c+/
(λάθος)

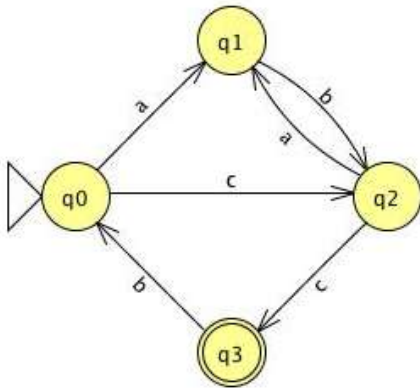
Ο πίνακας που αντιστοιχεί στο παρακάτω αυτόματο έχει διαστάσεις:



Απάντηση

- ☒ ☒ 3 γραμμές, 4 στήλες
(λάθος)
- ☐ 4 γραμμές, 3 στήλες
(σωστό)
Οι γραμμές είναι όσες οι καταστάσεις. Οι στήλες είναι όσες τα σύμβολα του αλφαβήτου.
- ☐ 4 γραμμές, 4 στήλες
(λάθος)
- ☐ 3 γραμμές, 3 στήλες
(λάθος)

Ποιές από τις παρακάτω συμβολοσειρές αναγνωρίζει το ακόλουθο αυτόματο πεπερασμένων καταστάσεων;



Απάντηση

☒ ☒ ccbab

(λάθος)

Δεν καταλήγει σε τελική κατάσταση.

☐ ababc

(σωστό)

☐ ccbcb

(λάθος)

☐ cabc

(σωστό)

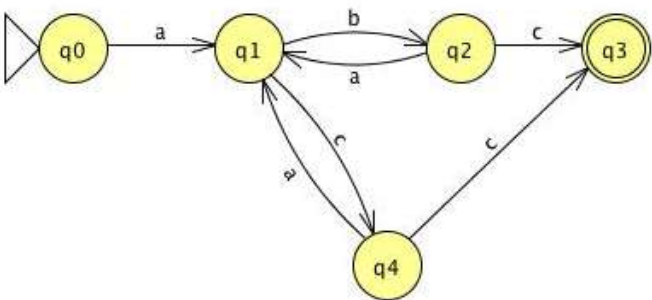
☐ abc

(σωστό)

☐ abcbcab

(σωστό)

Με ποιά από τις παρακάτω κανονικές εκφράσεις είναι ισοδύναμο το ακόλουθο αυτόματο πεπερασμένων καταστάσεων;



Απάντηση

- | | |
|--|---------------------------|
| <input checked="" type="checkbox"/> ✗ | $/[a[bc]]^*c/$
(λάθος) |
| <input type="checkbox"/> | $/(abc)+c/$
(λάθος) |
| <input type="checkbox"/> | $/(a[bc])+c/$
(σωστό) |
| <input type="checkbox"/> | $/(abc)]^*c/$
(λάθος) |

Ο ακόλουθος πίνακας ορίζει ένα αυτόματο πεπερασμένων καταστάσεων. Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;


FSA	a	b	c
>1	2	-	-
2	-	3,5	4
3	-	-	5
4:	-	5	-
5:	-	-	-

Απάντηση

- ☒  ac
(σωστό)
- ☐ acb
(σωστό)
- ☐ ab
(σωστό)
- ☐ abb
(λάθος)
- ☐ abc
(σωστό)
- ☐ aca
(λάθος)

Σύμφωνα με την Ιεραρχία Chomsky, οι γραμματικές Τύπου 3 εμπεριέχονται στις γραμματικές Τύπου 1.

Απάντηση

☒  Σωστό (σωστό)

☐ Λάθος (λάθος)

Επιλέξτε τον σωστό τύπο για τον κανόνα επανεγγραφής: $Q \rightarrow a B f E a$

Απάντηση

- ☒ ☒

0

(λάθος)

Όλοι οι κανόνες είναι ΚΑΙ τύπου 0. Γιαυτό μας ενδιαφέρει να σημειωθεί ο μεγαλύτερος αριθμός τύπου που αντιστοιχεί στον κανόνα.
- ☐

1

(λάθος)
- ☐

2


(σωστό)
- ☐

3

(λάθος)

Επιλέξτε τον σωστό τύπο για τον κανόνα επανεγγραφής: $a \ Q \ a \ \rightarrow \ a \ B \ f \ E \ a$

Απάντηση

- ☒ 

0
(λάθος)
Όλοι οι κανόνες είναι ΚΑΙ τύπου 0. Γιαυτό μας ενδιαφέρει να σημειωθεί ο μεγαλύτερος αριθμός τύπου που αντιστοιχεί στον κανόνα.
- ☐


1
(σωστό)
Τα a στα δεξιά και αριστερά του Q αποτελούν τα "συμφραζόμενα".
- ☐

2
(λάθος)
- ☐

3
(λάθος)

Επιλέξτε τον σωστό τύπο για τον κανόνα επανεγγραφής: $a \rightarrow b \rightarrow a \rightarrow b \rightarrow a$

Απάντηση

- ☒ 
- 0
- (σωστό)
- ☐
- 1
- (λάθος)
- ☐
- 2
- (λάθος)
- ☐
- 3
- (λάθος)

Συντακτικά, οι φυσικές γλώσσες μπορούν να εκφραστούν ικανοποιητικά με γραμματικές τύπου 2.

Απάντηση

- ☒ 
- Σωστό (σωστό)
- Ναι, με εξαίρεση κάποια σπάνια συντακτικά φαινόμενα που απαιτούν γραμματική τύπου 1.
- ☐
- Λάθος (λάθος)

Επιλέξτε τη σωστή μορφή CNF που αντιστοιχεί στον κανόνα: $W \rightarrow a \wedge b \wedge c$

Απάντηση



$W \rightarrow A \wedge B \wedge C$
 $A \rightarrow a$
 $B \rightarrow b$
 $C \rightarrow c$

(λάθος)



$W \rightarrow D \wedge E$
 $D \rightarrow A \wedge B$
 $E \rightarrow C$
 $A \rightarrow a$
 $B \rightarrow b$

(λάθος)



$W \rightarrow D \wedge C$
 $D \rightarrow A \wedge B$
 $A \rightarrow a$
 $B \rightarrow b$
 $C \rightarrow c$

(σωστό)



Είναι ήδη σε μορφή CNF.

(λάθος)

Επιλέξτε τη σωστή μορφή CNF που αντιστοιχεί στον κανόνα: $C \rightarrow A \vee B$

Απάντηση



$C \rightarrow a$

$C \rightarrow b$

$A \rightarrow a$

$B \rightarrow b$

(λάθος)



$C \rightarrow a$

$C \rightarrow b$

(λάθος)



$C \rightarrow a \vee b$

(λάθος)



Είναι ήδη σε μορφή CNF.

(σωστό)

Επιλέξτε τη σωστή μορφή CNF που αντιστοιχεί στον κανόνα: $C \rightarrow B$

Απάντηση

- ☐ $C \rightarrow b$
(λάθος)
- ☒ ☒ $C \rightarrow b$
 $B \rightarrow b$
(λάθος)
- ☐ Πρέπει να γνωρίζουμε και τους υπόλοιπους κανόνες της γραμματικής για να τον μετατρέψουμε.
(σωστό)
- ☐ Είναι ήδη σε μορφή CNF.
(λάθος)

Ποιοί από τους παρακάτω κανόνες είναι σε μορφή CNF;

Απάντηση



$A \rightarrow B \vee C$

(σωστό)



$B \rightarrow C \wedge a$

(λάθος)



$C \rightarrow D$

(λάθος)



$D \rightarrow b$

(σωστό)



$E \rightarrow D \vee A \vee C$

(λάθος)



$F \rightarrow a \vee b$

(λάθος)



$G \rightarrow F \vee C$

(σωστό)



$H \rightarrow a \vee b \vee c$

(λάθος)

Τί τύπου γραμματικές μπορούν να μετατραπούν σε CNF;

Απάντηση

<input checked="" type="checkbox"/> 	Τύπου 0 (λάθος)
<input type="checkbox"/>	Τύπου 1 (λάθος)
<input type="checkbox"/>	Τύπου 2 (σωστό)
<input type="checkbox"/>	Τύπου 3 (σωστό) Ναι, παρότι δε συνηθίζεται, και οι γραμματικές τύπου 3 μπορούν να μετατραπούν σε CNF. (Επειδή όλες οι γραμματικές τύπου 2 μετατρέπονται σε CNF, και οι τύπου 3 εμπεριέχονται στις τύπου 2.)

Έστω οι παρακάτω κανόνες μετεγγραφής. Ταιριάζτε καθέναν με τον τύπο της γραμματικής στον οποίο μπορεί να εμφανίζεται.
Σημείωση: Προφανώς ένας κανόνας π.χ. τύπου 3 μπορεί να εμφανίζεται και σε γραμματικές τύπου 2, 1, ή 0. Για κάθε κανόνα μας ενδιαφέρει να επιλέξετε τον πιο περιορισμένο τύπο (το μεγαλύτερο αποδεκτό νούμερο).

Απάντηση	
$F \rightarrow x F$	Τύπου 0 / Τύπου 3 ✖
$E F \rightarrow T x G$	Τύπου 1 / Τύπου 0 ✖
$E F \rightarrow E x T$	Τύπου 2 / Τύπου 1 ✖
$G \rightarrow W y E$	Τύπου 3 / Τύπου 2 ✖

Ποιές από τις παρακάτω συμβολοσειρές αναγνωρίζει η ακόλουθη γραμματική;

$$\begin{array}{lcl} S & \rightarrow & B \ C \\ A & \rightarrow & a \ c \\ A & \rightarrow & S \ c \\ B & \rightarrow & b \ A \\ B & \rightarrow & A \\ B & \rightarrow & C \ a \ b \\ C & \rightarrow & c \end{array}$$

Απάντηση

☒ ☒ abca
(λάθος)

☐ bacca
(λάθος)

☐ cabc
(σωστό)

☐ bacc
(σωστό)


Στον αλγόριθμο CKY, για να γεμίσουμε το τετράγωνο (1,4) θα πρέπει να συνδυάσουμε τα περιεχόμενα των τετραγώνων:

Απάντηση

- ☒  (1,3) με (2,4)
(λάθος)
- ☐ (1,2) με (3,4)
(λάθος)
- ☐ (1,2) με (2,3) και (2,3) με (3,4)
(λάθος)
- ☐ (1,2) με (3,4) και (1,3) με (2,4)
(λάθος)
- ☐ (1,2) με (2,4) και (1,3) με (3,4)
(σωστό)
- ☐ (1,2) με (2,3) και (1,3) με (2,4) και (2,3) με (3,4)
(λάθος)

Για να χρησιμοποιήσουμε τον αλγόριθμο CKY, η γραμματική τύπου 2 πρέπει να είναι σε μορφή CNF.

Απάντηση

- ☒  Σωστό (σωστό)
- ☐ Λάθος (λάθος)

Το μέγεθος του πίνακα στον αλγόριθμο CKY εξαρτάται από:

Απάντηση



Το πλήθος των κανόνων της γραμματικής.
(λάθος)



Το πλήθος των μη-τερματικών συμβόλων.
(λάθος)



Το πλήθος των λέξεων της πρότασης που αναλύεται.
(σωστό)



Το μέγεθος του λεξικού.
(λάθος)



Τίποτα από τα παραπάνω.
(λάθος)

Στα τετράγωνα του τριγωνικού πίνακα στον αλγόριθμο CKY βάζουμε τερματικά σύμβολα.


Απάντηση

☒ ☒ Σωστό (λάθος)

☐ Λάθος (σωστό)

Στον αλγόριθμο CKY, το τμήμα του πίνακα που βρίσκεται κάτω από την κύρια διαγώνιο παραλείπεται επειδή:

Απάντηση

- ☒  Τα τετράγωνά του δεν αντιστοιχούν σε σωστά διαστήματα της προς ανάλυση πρότασης.
(σωστό)
- ☐ Αν υπήρχαν, θα περιείχαν την ίδια πληροφορία με τον άλλο μισό πίνακα.
(λάθος)

Στα τετράγωνα του τριγωνικού πίνακα στον αλγόριθμο CKY δεν ξαναβάζουμε σύμβολα που ήδη υπάρχουν.


Απάντηση

☒ ☒ Σωστό (λάθος)

☐ Λάθος (σωστό)

Στον τριγωνικό πίνακα του αλγορίθμου CKY, το πάνω δεξιά τετράγωνο γεμίζει τελευταίο.

Απάντηση

☒  Σωστό (σωστό)

☐ Λάθος (λάθος)

Στον αλγόριθμο CKY, αν στο πάνω δεξιά τετράγωνο εμφανιστεί 2 φορές το σύμβολο S, αυτό σημαίνει ότι:

Απάντηση

- ☒ ☒ έχουμε κάνει κάποιο λάθος στη διαδικασία.
(λάθος)
- ☐ η γραμματική δεν είναι σε μορφή CNF.
(λάθος)
- ☐ η πρόταση μπορεί να αναλυθεί συντακτικά με δύο τρόπους.
(σωστό)
- ☐ η πρόταση δεν αναγνωρίζεται από τη γραμματική.
(λάθος)

Αν $P(X) = 0,5$ τότε $\log P(X) = \dots$

Απάντηση



-0,5

(λάθος)



-0,3

(σωστό)



0,3

(λάθος)



0,5

(λάθος)



1

(λάθος)

Αν $\log P(X) = -4,5$ τότε $P(X) = \dots$

Απάντηση



0,0316227

(λάθος)



0,00316227

(λάθος)



0,000316227

(λάθος)



0,0000316227

(σωστό)



0,00000316227

(λάθος)


Στις PCFG ισχύει ότι...
(επιλέξτε όλα όσα ταιριάζουν)

Απάντηση

- | | |
|---------------------------------------|---|
| <input checked="" type="checkbox"/> ✖ | οι κανόνες βρίσκονται πάντα σε μορφή CNF.
(λάθος) |
| <input type="checkbox"/> | οι πιθανότητες όλων των κανόνων είναι πάντα 1.
(λάθος) |
| <input type="checkbox"/> | οι πιθανότητες των κανόνων με το ίδιο head αθροίζουν στο 1.
(σωστό) |
| <input type="checkbox"/> | οι πιθανότητες όλων των κανόνων αθροίζουν στο 1.
(λάθος) |
| <input type="checkbox"/> | οι πιθανότητες όλων των κανόνων αθροίζουν στο πλήθος των μη-τερματικών συμβόλων.
(σωστό) |
| <input type="checkbox"/> | το πλήθος των κανόνων ισούται με το πλήθος των μη-τερματικών συμβόλων.
(λάθος) |

Αν κατά τη μετατροπή PCFG σε CNF εμφανιστούν ίδιοι κανόνες, τους συγχωνεύουμε αθροίζοντας τις πιθανότητές τους.

Απάντηση

☒  Σωστό (σωστό)

☐ Λάθος (λάθος)

Κανόνες με αυτοαναφορές στις PCFG απαλείφονται και οι πιθανότητες τους μοιράζονται αναλογικά στους υπόλοιπους κανόνες με το ίδιο head.

Απάντηση	
<input checked="" type="checkbox"/> ✓	Σωστό (σωστό)
<input type="checkbox"/>	Λάθος (λάθος)

Έστω η PCFG:

$S \rightarrow AS$ [0,2]

$S \rightarrow SA$ [0,8]

$A \rightarrow a$ [0,3]

$A \rightarrow S$ [0,7]

Για να μετατρέψουμε τον τελευταίο κανόνα σε μορφή CNF τον αντικαθιστούμε με τους κανόνες:

Απάντηση



$A \rightarrow AS$ [0,14]

$A \rightarrow SA$ [0,56]

(σωστό)



$A \rightarrow AS$ [0,2]

$A \rightarrow SA$ [0,8]

(λάθος)



$S \rightarrow AS$ [0,14]

$S \rightarrow SA$ [0,56]

(λάθος)



$S \rightarrow AS$ [0,2]

$S \rightarrow SA$ [0,8]

(λάθος)

Έστω το σύνολο κανόνων PCFG:

$S \rightarrow a$ [0,3]

$S \rightarrow b$ [0,2]

$S \rightarrow b$ [0,5]

Η σωστή συγχώνευση είναι:

Απάντηση



$S \rightarrow a$ [0,3]

$S \rightarrow b$ [0,2]

(λάθος)



$S \rightarrow a$ [0,3]

$S \rightarrow b$ [0,5]

(λάθος)



$S \rightarrow a$ [0,6]

$S \rightarrow b$ [0,4]

(λάθος)



$S \rightarrow a$ [0,3]

$S \rightarrow b$ [0,7]

(σωστό)

Έστω το σύνολο κανόνων PCFG:

$S \rightarrow a$ [0,3]

$S \rightarrow b$ [0,1]

$S \rightarrow S$ [0,6]

Με την απαλοιφή της αυτοαναφοράς το σύνολο γίνεται:

Απάντηση



$S \rightarrow a$ [0,3]

$S \rightarrow b$ [0,1]

(λάθος)



$S \rightarrow a$ [0,75]

$S \rightarrow b$ [0,25]

(σωστό)



$S \rightarrow a$ [0,18]

$S \rightarrow b$ [0,06]

(λάθος)



$S \rightarrow a$ [0,6]

$S \rightarrow b$ [0,4]

(λάθος)

Οι πιθανότητες στα τετράγωνα του πίνακα του πιθανοτικού CKY ισούνται με το γινόμενο πιθανοτήτων των δύο τερματικών συμβόλων στο σώμα του κάθε κανόνα.

Απάντηση

- ☒ ✖ Σωστό (λάθος)
- ☐ Λάθος (σωστό)
Ισούνται με το γινόμενο πιθανοτήτων των δύο τερματικών συμβόλων στο σώμα του κανόνα **επί την πιθανότητα του κανόνα** (δηλαδή τρεις είναι οι αριθμοί που πολλαπλασιάζονται).

Έστω η PCFG και ο πίνακας CKY που ακολουθούν:

S	→	A C	[0 , 4]
S	→	C A	[0 , 6]
C	→	A B	[0 , 2]
C	→	B A	[0 , 8]
A	→	a	[0 , 5]
A	→	c	[0 , 3]
A	→	d	[0 , 2]
B	→	b	[0 , 4]
B	→	d	[0 , 6]

0,5 A	C	S,S
a	0,4 B	C
	b	0,3 A
		c

Στο πάνω-δεξιά τετράγωνο του πίνακα εμφανίζονται δύο σύμβολα S, που σημαίνει ότι υπάρχουν δύο τρόποι να αναλυθεί συντακτικά η πρόταση "abc". Αν η πιθανότητα του πρώτου S είναι 0,0192, ποια είναι η πιθανότητα του δεύτερου;

Απάντηση	
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>	0,9808 (λάθος)
<input type="checkbox"/>	0,0384 (λάθος)
<input type="checkbox"/>	0,0072 (σωστό)
<input type="checkbox"/>	0,0192 (λάθος)

Η υπόθεση Markov λέει ότι κάθε λέξη στην πρόταση εξαρτάται...

Απάντηση

- ☒  μόνο από τις προηγούμενές της.
(σωστό)
- ☐ μόνο από τις επόμενές της.
(λάθος)
- ☐ και από τις προηγούμενες και από τις επόμενές της.
(λάθος)
- ☐ από το μοντέλο μονογραμμάτων.
(λάθος)
- ☐ από το πλήθος εμφανίσεών της στο σώμα κειμένων.
(λάθος)

Τα μοντέλα μονογραμμάτων περιέχουν συντακτική πληροφορία.

Απάντηση



Σωστό (λάθος)



Λάθος (σωστό)

Τα μοντέλα διγραμμάτων περιέχουν συντακτική πληροφορία.

Απάντηση



Σωστό (σωστό)



Λάθος (λάθος)

Στο ίδιο σώμα κειμένων, ο πίνακας εμφανίσεων διγραμμάτων είναι πιο αραιός από τον πίνακα εμφανίσεων τριγραμμάτων.

Απάντηση

- ☒ 
- Σωστό (λάθος)
- Το αντίθετο συμβαίνει. Όσο μεγαλύτερης τάξης είναι το μοντέλο, τόσο πιο αραιός είναι ο πίνακας. Αυτό συμβαίνει επειδή οι συνδυασμοί των λέξεων αυξάνονται πολύ, ενώ ελάχιστοι από αυτούς είναι δόκιμοι.
- ☐
- Λάθος (σωστό)

Τα μηδενικά στους πίνακες πιθανοτήτων των μοντέλων γλώσσας είναι ανεπιθύμητα επειδή:

Απάντηση

- ☒ **✗** σπαταλούν μνήμη στον αποθηκευτικό χώρο του υπολογιστή.
(λάθος)
- ☐ μηδενίζουν τις πιθανότητες ακόμα και σωστών προτάσεων.
(σωστό)
- ☐ δεν μεταφέρουν καθόλου πληροφορία.
(λάθος)
- ☐ δε μπορούν να υπολογιστούν τα logprobs για αυτές τις πιθανότητες.
(λάθος)

Θεωρήστε το σώμα:
<s> ποτέ μη λες ποτέ </s>
<s> μη λες ποτέ ποτέ </s>
<s> ποτέ ποτέ μη λες </s>

Υπολογίστε τις παρακάτω πιθανότητες, ταιριάζοντάς τες με τους σωστούς αριθμούς.

Απάντηση

P(ποτέ ποτέ)	0,000 / 0,333 ✖
P(λες μη)	0,000 / 1 ✖
P(μη λες)	0,000 / 0,000 ✔
P(ποτέ λες)	0,000 / 0,667 ✖
P(λες ποτέ)	0,000 / 0,000 ✔

Έστω το σώμα:
<s> ποτέ μη λες ποτέ </s>
<s> μη λες ποτέ ποτέ </s>
<s> ποτέ ποτέ μη λες </s>

Υπολογίστε τις παρακάτω πιθανότητες **μετά την εφαρμογή εξομάλυνσης Laplace**, ταιριάζοντάς τες με τους σωστούς αριθμούς.

Απάντηση

P(λες μη)	0,125 / 0,5 ✖
P(λες λες)	0,273 / 0,125 ✖
P(ποτέ ποτέ)	0,273 / 0,273 ✔
P(μη λες)	0,125 / 0,125 ✔

Στον πίνακα εμφανίσεων τριγραμμάτων ενός μοντέλου γλώσσας βλέπουμε ένα μηδενικό. Αν εφαρμόσουμε εξομάλυνση Laplace, αυτό το 0 θα γίνει:

Απάντηση

- ☒ ☒ -1
(λάθος)
- ☐ 0
(λάθος)
- ☐ 1
(σωστό)
- ☐ 2
(λάθος)
- ☐ 3
(λάθος)
- ☐ Πρέπει να γνωρίζουμε και τα υπόλοιπα στοιχεία του πίνακα για να απαντήσουμε.
(λάθος)

Στον πίνακα πιθανοτήτων τριγραμμάτων ενός μοντέλου γλώσσας βλέπουμε ένα μηδενικό. Αν εφαρμόσουμε εξομάλυνση Laplace, αυτό το 0 θα γίνει:

Απάντηση

- ☒ ☐ ☐ ☐ ☐ ☐
- 1

(λάθος)

0

(λάθος)

1

(λάθος)

2

(λάθος)

3

(λάθος)

Δεν έχουμε αρκετές πληροφορίες για να υπολογίσουμε ακριβώς.

(σωστό)

Για να αποφύγουμε τις μηδενικές δεσμευμένες πιθανότητες σε ένα μοντέλο γλώσσας, χρησιμοποιούμε:
(επιλέξτε όλα όσα ταιριάζουν)

Απάντηση

☒  παρεμβολή (interpolation).
(σωστό)

☐ μοντέλο ανώτερης τάξης.
(λάθος)

☐ αλγόριθμο CKY.
(λάθος)

☐ οπισθοχώρηση (backoff).
(σωστό)


☐ εξομάλυνση Laplace.
(σωστό)

☐ απόσταση Levenshtein.
(λάθος)

☐ μετατροπή σε CNF.
(λάθος)

Στα μοντέλα γλώσσας μεγάλης τάξης (τετραγράμματα, πενταγράμματα, κλπ) γίνεται εμφανές το είδος των κειμένων του σώματος που χρησιμοποιήθηκε για την κατασκευή τους.

Απάντηση

- ☒  Σωστό (σωστό)
- ☐ Λάθος (λάθος)

Στην τεχνική της οπισθοχώρησης (backoff), όταν η δεσμευμένη πιθανότητα ενός n-gram είναι 0, τότε...

Απάντηση

- ☒ ☒ προσθέτουμε 1 στον πίνακα εμφάνισης για αυτήν την πιθανότητα.
(λάθος)
- ☐ υπολογίζουμε την πιθανότητα του αμέσως μικρότερης τάξης n-gram.
(σωστό)
- ☐ αγνοούμε τελείως αυτήν την πιθανότητα στον υπολογισμό.
(λάθος)
- ☐ υπολογίζουμε νέα πιθανότητα με βάση τους συντελεστές λ.
(λάθος)

Οι συντελεστές λ στην τεχνική της παρεμβολής (interpolation)...
(επιλέξτε όλα όσα ταιριάζουν)

Απάντηση



έχουν γινόμενο ίσο με 1.

(λάθος)



έχουν άθροισμα ίσο με 1.

(σωστό)



είναι όλοι μικρότεροι (ή ίσοι) από 1.

(σωστό)



ενδέχεται να είναι αρνητικοί.

(λάθος)

Στο ARPA format οι πιθανότητες των n-grams εμφανίζονται με μορφή Logprob.

Απάντηση




Σωστό (σωστό)



Λάθος (λάθος)

Με τη μέθοδο του "Θορυβώδους Καναλιού" μπορούμε να κάνουμε ορθογραφική διόρθωση με βάση τα συμφραζόμενα.


Απάντηση

☒  Σωστό (σωστό)

☐ Λάθος (λάθος)

Η πληροφορία που αφορά το στρίψιμο ενός κέρματος (κορώνα-γράμματα, ισοπίθανα ενδεχόμενα) είναι:

Απάντηση

- | | |
|---|--------------------|
| <input checked="" type="checkbox"/>  | 0 bit
(λάθος) |
| <input type="checkbox"/> | 0,5 bit
(λάθος) |
| <input type="checkbox"/> | 1 bit
(σωστό) |
| <input type="checkbox"/> | 2 bit
(λάθος) |

Η πληροφορία που αφορά τη ρίψη ενός ζαριού (με έξι ισοπίθανα ενδεχόμενα) είναι:

Απάντηση



0,17 bit

(λάθος)



0,39 bit

(λάθος)



0,5 bit

(λάθος)



1 bit

(λάθος)



2,58 bit

(σωστό)

Η εντροπία...

(επιλέξτε όλα όσα ταιριάζουν)


Απάντηση

- ☒  δείχνει πόσο "τακτοποιημένο" ή "ανακατεμένο" είναι ένα σύστημα.
(σωστό)
- ☐ καθορίζει πόση πληροφορία χρειαζόμαστε για να περιγράψουμε πλήρως ένα σύστημα.
(σωστό)
- ☐ μετριέται σε bits.
(σωστό)
- ☐ υπολογίζεται από πιθανότητες (και λογάριθμους πιθανοτήτων).
(σωστό)
- ☐ σχετίζεται με την περιπλοκή (perplexity).
(σωστό)
- ☐ εφαρμόζεται μόνο σε γραμματικές τύπου 2.
(λάθος)

Η περιπλοκή (perplexity)...

(επιλέξτε όλα όσα ταιριάζουν)

Απάντηση

- ☒ 
- σχετίζεται με την εντροπία.
- (σωστό)
- ☐
- υπολογίζεται απευθείας από πιθανότητες.
- (σωστό)
- ☐
- εκφράζει τον μέσο παράγοντα διακλάδωσης των κόμβων του γράφου που περιγράφει τη γλώσσα.
- (σωστό)
- ☐
- υπολογίζεται από το WordNet.
- (λάθος)

Αν μια γλώσσα έχει εντροπία 0, τότε έχει περιπλοκή 1.

Απάντηση

- ☒ 
- Σωστό (σωστό)
- Προκύπτει άμεσα από τον τύπο υπολογισμού της περιπλοκής από την εντροπία (αντικαθιστώντας την εντροπία με 0).
- ☐
- Λάθος (λάθος)

Ένα αποτελεσματικό εργαλείο για την κατηγοριοποίηση κειμένων είναι:

Απάντηση



ο αλγόριθμος Levenshtein

(λάθος)



ο αλγόριθμος CKY

(λάθος)



οι κανονικές εκφράσεις

(λάθος)



ένα πρόγραμμα συμπίεσης αρχείων

(σωστό)



ο κώδικας ASCII

(λάθος)



ένας δεκαεξαδικός συντάκτης

(λάθος)

To WordNet...

(επιλέξτε όλα όσα ταιριάζουν)

Απάντηση



είναι ένα ηλεκτρονικό λεξικό.

(σωστό)



χρησιμοποιείται για συντακτική ανάλυση.

(λάθος)



έχει μορφή δικτύου με κόμβους και μεταβάσεις.

(σωστό)



χρησιμοποιείται για την εύρεση συνώνυμων λέξεων.

(σωστό)



περιέχει πληροφορίες για τη μορφολογία των λέξεων.

(λάθος)



είναι ένα είδος treebank.

(λάθος)

Οι κυ?ριες εννοιολογικε?ς συσχετι?σεις που κωδικοποιει? το WordNet αφορου?ν:

Απάντηση



συνωνυμία

(σωστό)



ανωνυμία

(λάθος)



επωνυμία

(λάθος)



μερωνυμία

(σωστό)



τροπωνυμία

(σωστό)



παρωνυμία

(λάθος)



αντινομία

(λάθος)



υπωνυμία

(σωστό)



παρανομία

(λάθος)



αντωνυμία

(σωστό)

Ταιριάξτε τις έννοιες με τις ερμηνείες τους:

Απάντηση	
συνωνυμία	λε?ξεις με αντι?θετες ε?ννοιες / λε?ξεις με παρο?μοιες ε?ννοιες ✖
αντωνυμία	ρη?ματα που το ε?να ει?ναι εξειδι?κευση του α?λλου / λε?ξεις με αντι?θετες ε?ννοιες ✖
υπωνυμία	ουσιαστικά? με σχε?ση με?ρος προς ο?λον / ουσιαστικά? που το ε?να ει?ναι εξειδι?κευση του α?λλου ✖
μερωνυμία	λε?ξεις με παρο?μοιες ε?ννοιες / ουσιαστικά? με σχε?ση με?ρος προς ο?λον ✖
τροπωνυμία	ουσιαστικά? που το ε?να ει?ναι εξειδι?κευση του α?λλου / ρη?ματα που το ε?να ει?ναι εξειδι?κευση του α?λλου ✖

Ταιριάζτε τις έννοιες		
Απάντηση		
Synset		Γραμματική τύπου 2 / WordNet ✖
Θορυβώδες κανάλι		FSA / Ορθογραφική διόρθωση ✖
Τρίγραμμα		FSA / Μοντέλο γλώσσας ✖
CNF	CNF	FSA / Γραμματική τύπου 2 ✖
Παράλληλα FSTs		Levenshtein / PC-KIMMO ✖
Βέλτιστο μονοπάτι		PC-KIMMO / Levenshtein ✖
Εντροπία		PC-KIMMO / Perplexity (Περιπλοκή) ✖
Τελική κατάσταση		WordNet / FSA ✖