

Ερώτηση: 1 (Βαθμός: -0.50 / 1.00)

(Πολλαπλής Επιλογής (Μοναδική Απάντηση))

Υπολογιστική Γλωσσολογία ονομάζεται η επιστημονική περιοχή μεταξύ:

Απάντηση	
<input checked="" type="checkbox"/>	Γλωσσολογίας και Ιατρικής (λάθος)
<input type="checkbox"/>	Ιατρικής και Πληροφορικής (λάθος)
<input type="checkbox"/>	Πληροφορικής και Γλωσσολογίας (σωστό)

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

Απάντηση	
<input checked="" type="checkbox"/>	American Computational Linguistics (λάθος)
<input type="checkbox"/>	Association for Computational Linguistics (σωστό)
<input type="checkbox"/>	Association for Computer Languages (λάθος)
<input type="checkbox"/>	Administration for Computer Languages (λάθος)

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

Απάντηση	
<input checked="" type="checkbox"/>	Γλωσσικός Υπολογισμός (λάθος)
<input type="checkbox"/>	Επεξεργασία Φυσικής Γλώσσας (σωστό)
<input type="checkbox"/>	Φυσική Επεξεργασία Γλώσσας (λάθος)
<input type="checkbox"/>	Γλωσσική Τεχνολογία (σωστό)
<input type="checkbox"/>	Τεχνική Γλωσσολογία (λάθος)

Η κωδικοποίηση Unicode...

Απάντηση	
<input checked="" type="checkbox"/>	Μπορεί να αναπαραστήσει πάνω από 1.000.000 διαφορετικούς χαρακτήρες. (σωστό)
<input type="checkbox"/>	Περιλαμβάνει 17 επίπεδα με πάνω από 60.000 code points το καθένα. (σωστό)
<input type="checkbox"/>	Μπορεί να χρησιμοποιηθεί για την συνύπαρξη χαρακτήρων από περισσότερες από δύο γλώσσες στο ίδιο κείμενο. (σωστό)
<input type="checkbox"/>	Αφορά μόνο αριθμητικά ψηφία. (λάθος)
<input type="checkbox"/>	Είναι παλαιότερη από την κωδικοποίηση ASCII. (λάθος)

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

Απάντηση

☒

✕

X

(λάθος)

☐

Θ

Θ

(λάθος)

☐

Ω

Ω

(λάθος)

☐

Ξ

Ξ

(λάθος)

☐

Κανένα από τα παραπάνω.

(σωστό)

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

Απάντηση

☒

✓

Σωστό (σωστό)

☐

Λάθος (λάθος)

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 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

☒

✕

1

(λάθος)

☐

2

(λάθος)

☐

3

(σωστό)

☐

4

(λάθος)

☐

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"/> <b>✓</b>	<b>1</b> (σωστό)
<input type="checkbox"/>	<b>2</b> (λάθος)
<input type="checkbox"/>	<b>3</b> (λάθος)
<input type="checkbox"/>	<b>4</b> (λάθος)
<input type="checkbox"/>	Μεγαλύτερο από 4. (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	
<input checked="" type="checkbox"/> <b>✓</b>	Μεταξύ 3 και 5 (σωστό)
<input type="checkbox"/>	Μεταξύ 1 και 4 (λάθος)
<input type="checkbox"/>	<b>6</b> (λάθος)
<input type="checkbox"/>	Κανένα από τα παραπάνω. (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	
<input checked="" type="checkbox"/> <b>✗</b>	<b>4</b> (λάθος)
<input type="checkbox"/>	Μεταξύ 3 και 5 (λάθος)
<input type="checkbox"/>	Μεταξύ 2 και 8 (σωστό)
<input type="checkbox"/>	Κανένα από τα παραπάνω. (λάθος)

Έστω η κανονική έκφραση: **/a[(bc)(cb)a]/**  
Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;

Απάντηση	
<input checked="" type="checkbox"/> <b>✓</b>	<b>abc</b> (σωστό)
<input type="checkbox"/>	<b>aba</b> (λάθος)
<input type="checkbox"/>	<b>acb</b> (σωστό)
<input type="checkbox"/>	<b>aca</b> (λάθος)
<input type="checkbox"/>	<b>cba</b> (λάθος)
<input type="checkbox"/>	<b>aa</b> (σωστό)

Έστω η κανονική έκφραση: /(ab)\*[ca]\*/

Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;

Απάντηση	
<input checked="" type="checkbox"/> ✖	baca (λάθος)
<input type="checkbox"/>	aba (σωστό)
<input type="checkbox"/>	cacaca (λάθος)
<input type="checkbox"/>	abca (σωστό)
<input type="checkbox"/>	ababcca (σωστό)
<input type="checkbox"/>	aabc (λάθος)

Έστω η κανονική έκφραση: /(abc)+[a(ab)c]\*/

Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;

Απάντηση	
<input checked="" type="checkbox"/> ✖	abcac (λάθος)
<input type="checkbox"/>	abcabc (σωστό)
<input type="checkbox"/>	aabcc (λάθος)
<input type="checkbox"/>	abcab (σωστό)
<input type="checkbox"/>	aababc (λάθος)
<input type="checkbox"/>	abcabcc (σωστό)

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

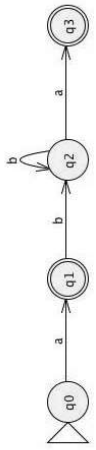
Ποιές από τις παρακάτω συμβολοσειρές περιλαμβάνονται στη γλώσσα που ορίζει;

Απάντηση	
<input checked="" type="checkbox"/> ✖	abcb (λάθος)
<input type="checkbox"/>	abcc (σωστό)
<input type="checkbox"/>	abccb (λάθος)
<input type="checkbox"/>	abcbcb (σωστό)
<input type="checkbox"/>	abcb (σωστό)
<input type="checkbox"/>	abb (σωστό)

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

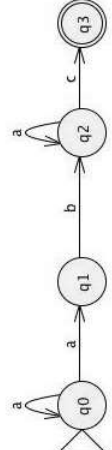
Απάντηση	
<input checked="" type="checkbox"/> ✔	/(ab)+b/ (σωστό)
<input type="checkbox"/>	/c*(ab)*b/ (σωστό)
<input type="checkbox"/>	/ab?ab+/ (σωστό)
<input type="checkbox"/>	/a(ba)+b/ (λάθος)

Το παρακάτω αυτόματο είναι:



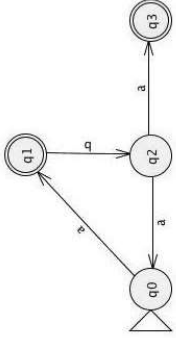
Απάντηση	
<input checked="" type="checkbox"/>	Κυκλικό (σωστό) Υπάρχει μια κυκλική μετάβαση από την κατάσταση q2 στον εαυτό της.
<input type="checkbox"/>	Ακυκλικό (λάθος)
<input type="checkbox"/>	Αιτιοκρατικό (σωστό)
<input type="checkbox"/>	Μη-αιτιοκρατικό (λάθος)

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



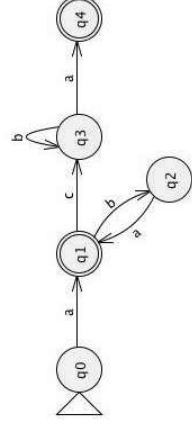
Απάντηση	
<input checked="" type="checkbox"/>	$a+b^*ac/$ (λάθος)
<input type="checkbox"/>	$a^*ba^*c/$ (λάθος)
<input type="checkbox"/>	$a^*ba+c/$ (λάθος)
<input type="checkbox"/>	$a+ba^*c/$ (σωστό)

Το παρακάτω αυτόματο είναι:



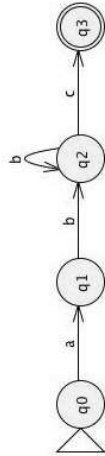
Απάντηση	
<input checked="" type="checkbox"/>	Κυκλικό (σωστό)
<input type="checkbox"/>	Ακυκλικό (λάθος)
<input type="checkbox"/>	Αιτιοκρατικό (λάθος)
<input type="checkbox"/>	Μη-αιτιοκρατικό (σωστό) Υπάρχουν δύο μεταβάσεις από την q2 με το ίδιο σύμβολο.

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



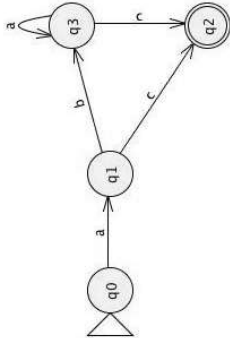
Απάντηση	
<input checked="" type="checkbox"/>	$a(ba)^*(cb^*a)?/$ (σωστό)
<input type="checkbox"/>	$a(ba)^*cb^*a/$ (λάθος)
<input type="checkbox"/>	$aba+(cb)^+a/$ (λάθος)
<input type="checkbox"/>	$a(b+a)^+(c^*b^*a)?/$ (λάθος)

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



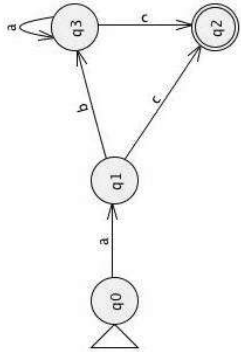
Απάντηση	
<input checked="" type="checkbox"/> <b>x</b>	$/abb+ c/$ (λάθος)
<input type="checkbox"/>	$/ab^*c/$ (λάθος)
<input type="checkbox"/>	$/ab+c/$ (σωστό)
<input type="checkbox"/>	$/ab^*c+ /$ (λάθος)

Ο πίνακας που αντιστοιχεί στο παρακάτω αυτόματο έχει διαστάσεις:



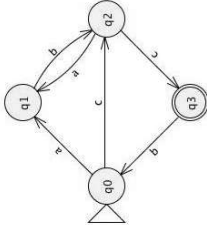
Απάντηση	
<input checked="" type="checkbox"/> <b>x</b>	3 γραμμές, 4 στήλες (λάθος)
<input type="checkbox"/>	4 γραμμές, 3 στήλες (σωστό) Οι γραμμές είναι όσες οι καταστάσεις. Οι στήλες είναι όσες τα σύμβολα του αλφάβητου.
<input type="checkbox"/>	4 γραμμές, 4 στήλες (λάθος)
<input type="checkbox"/>	3 γραμμές, 3 στήλες (λάθος)

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



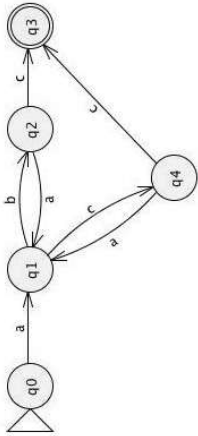
Απάντηση	
<input checked="" type="checkbox"/> <b>x</b>	$/a[c(b+a+c)]/$ (λάθος)
<input type="checkbox"/>	$/a[ac(bac)]/$ (λάθος)
<input type="checkbox"/>	$/a+[ca(bac)]/$ (λάθος)
<input type="checkbox"/>	$/a[c(ba^*c)]/$ (σωστό)

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



Απάντηση	
<input checked="" type="checkbox"/> <b>x</b>	ctbab (λάθος) Δεν καταλήγει σε τελική κατάσταση.
<input type="checkbox"/>	ababc (σωστό)
<input type="checkbox"/>	ctbcb (λάθος)
<input type="checkbox"/>	cabc (σωστό)
<input type="checkbox"/>	abc (σωστό)
<input type="checkbox"/>	abcbcab (σωστό)

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



Απάντηση

☒ ✖

/a[bc]\*c/

(λάθος)

☐

/abc+c/

(λάθος)

☐

/a[bc]+c/

(σωστό)

☐

/abc)\*c/

(λάθος)

Απάντηση

☒ ✔

Σωστό (σωστό)

☐

Λάθος (λάθος)

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

Απάντηση

☒ ✔

Σωστό (σωστό)

☐

Λάθος (λάθος)

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

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

Απάντηση

☒ ✔

ac

(σωστό)

☐

acd

(σωστό)

☐

ab

(σωστό)

☐

abb

(λάθος)

☐

abc

(σωστό)

☐

aca

(λάθος)

Επιλέξτε τον σωστό τύπο για τον κανόνα επανεγγραφής: Q -> a B f 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:  $a \vee b \rightarrow a \vee b \wedge c$

Απάντηση

☒ **x**

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 "συμμεταξύμενα".

☐

2

(λάθος)

☐

3

(λάθος)

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

Απάντηση

☒ **✓**

Σωστό (σωστό)

Ναι, με εξαίρεση κάποια συντακτικά φαινόμενα που απαιτούν γραμματική τύπου 1.

☐

Λάθος (λάθος)

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

Απάντηση

☒ **✓**

0

(σωστό)

☐

1

(λάθος)

☐

2

(λάθος)

☐

3

(λάθος)

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

Απάντηση

☒ **x**

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

(λάθος)

☐

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

(λάθος)

☐

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

(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)

Ποιοί από τους παρακάτω κανόνες είναι σε μορφή CNF:

Απάντηση

☒  $A \rightarrow B \vee C$   
(σωστό)

☐  $B \rightarrow C \vee A$   
(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)

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)

Τί τύπου γραμματικές μπορούν να μετατραπούν σε CNF;

Απάντηση

☒ Τύπου 0  
(λάθος)

☐ Τύπου 1  
(λάθος)

☐ Τύπου 2  
(σωστό)

☐ Τύπου 3  
(σωστό)  
Ναι, παρότι δε συνήθίζεται, 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.)



Το μέγεθος του πίνακα στον αλγόριθμο CKY εξαρτάται από:

Απάντηση	
<input checked="" type="checkbox"/> ✖	Το πλήθος των κανόνων της γραμματικής. (λάθος)
<input type="checkbox"/>	Το πλήθος των μη-τερματικών συμβόλων. (λάθος)
<input type="checkbox"/>	Το πλήθος των λέξεων της πρότασης που αναλύεται. (σωστό)
<input type="checkbox"/>	Το μέγεθος του λεξικού. (λάθος)
<input type="checkbox"/>	Τίποτα από τα παραπάνω. (λάθος)

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

Απάντηση	
<input checked="" type="checkbox"/> ✔	Τα τετράγωνα του δεν αντιστοιχούν σε σωστά διαστήματα της προς ανάλυση πρότασης. (σωστό)
<input type="checkbox"/>	Αν υπήρχαν, θα περιείχαν την ίδια πληροφορία με τον άλλο μισό πίνακα. (λάθος)

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

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

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

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

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)

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

Απάντηση

☒ ✔

Σωστό (σωστό)

☐

Λάθος (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)

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)

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

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	
<input checked="" type="checkbox"/> ✔	$A \rightarrow AS [0,14]$ $A \rightarrow SA [0,56]$ (σωστό)
<input type="checkbox"/>	$A \rightarrow AS [0,2]$ $A \rightarrow SA [0,8]$ (λάθος)
<input type="checkbox"/>	$S \rightarrow AS [0,14]$ $S \rightarrow SA [0,56]$ (λάθος)
<input type="checkbox"/>	$S \rightarrow AS [0,2]$ $S \rightarrow SA [0,8]$ (λάθος)

Έστω το σύνολο κανόνων PCFG:

S → a

[0,3]

S → b

[0,2]

S → b

[0,5]

Η σωστή συγχώνευση είναι:

Απάντηση

☒

✖

S → a

[0,3]

S → b

[0,2]

(λάθος)

☐

S → a

[0,3]

S → b

[0,5]

(λάθος)

☐

S → a

[0,6]

S → b

[0,4]

(λάθος)

☐

S → a

[0,3]

S → b

[0,7]

(σωστό)

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 (δηλαδή τρεις είναι οι αριθμοί που πολλαπλασιάζονται).

Έστω το σύνολο κανόνων PCFG:

S → a

[0,3]

S → b

[0,1]

S → S

[0,6]

Με την απαλοιφή της αυτοαναφοράς το σύνολο γίνεται:

Απάντηση

☒

✖

S → a

[0,3]

S → b

[0,1]

(λάθος)

☐

S → a

[0,75]

S → b

[0,25]

(σωστό)

☐

S → a

[0,18]

S → b

[0,06]

(λάθος)

☐

S → a

[0,6]

S → b

[0,4]

(λάθος)

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

a

[0,5]

A → c

c

[0,3]

A → d

d

[0,2]

B → b

b

[0,4]

B → d

d

[0,6]

0,5

A

0,4

B

0,3

A

a

b

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;

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

Απάντηση

☒ ☐ μόνο από τις προηγούμενες της.  
(σωστό)

☐ μόνο από τις επόμενες της.  
(λάθος)

☐ και από τις προηγούμενες και από τις επόμενες της.  
(λάθος)

☐ από το μοντέλο μονογραμμάτων.  
(λάθος)

☐ από το πλήθος εμφανίσεών της στο σώμα κειμένων.  
(λάθος)

Τα μοντέλα διγραμμάτων περιέχουν συντακτική πληροφορία.

Απάντηση

☒ ☐ Σωστό (σωστό)

☐ Λάθος (λάθος)

Τα μοντέλα μονογραμμάτων περιέχουν συντακτική πληροφορία.

Απάντηση

☒ ☐ Σωστό (λάθος)

☐ Λάθος (σωστό)

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

Απάντηση

☒ ☐ Σωστό (λάθος)

Το αντίθετο συμβαίνει. Όσο μεγαλύτερη τάξης είναι το μοντέλο, τόσο πιο αραιός είναι ο πίνακας. Αυτό συμβαίνει επειδή οι συνδυασμοί των λέξεων αυξάνονται πολύ, ενώ ελέγχεται από αυτούς είναι δοκίμια.

☐ Λάθος (σωστό)



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

Απάντηση	
<input checked="" type="checkbox"/> ✖	σπαταλούν μνήμη στον αποθηκευτικό χώρο του υπολογιστή. (λάθος)
<input type="checkbox"/>	μηδενίζουν τις πιθανότητες ακόμα και σωστών προτάσεων. (σωστό)
<input type="checkbox"/>	δεν μεταφέρουν καθόλου πληροφορία. (λάθος)
<input type="checkbox"/>	δε μπορούν να υπολογιστούν τα logprobs για αυτές τις πιθανότητες. (λάθος)

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 ✔

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 ✔

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

Απάντηση	
<input checked="" type="checkbox"/> ✖	-1 (λάθος)
<input type="checkbox"/>	0 (λάθος)
<input type="checkbox"/>	1 (σωστό)
<input type="checkbox"/>	2 (λάθος)
<input type="checkbox"/>	3 (λάθος)
<input type="checkbox"/>	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 αποφεύγουμε 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)

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)

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)

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)

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

Απάντηση

☒✔

Σωστό (σωστό)

☐

Λάθος (λάθος)

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

Απάντηση

☒✔

Σωστό (σωστό)

☐

Λάθος (λάθος)

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

Απάντηση

☒✕

0 bit  
(λάθος)

☐

0,5 bit  
(λάθος)

☐

1 bit  
(σωστό)

☐

2 bit  
(λάθος)

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

Απάντηση	
<input checked="" type="checkbox"/> ✕	0,17 bit (λάθος)
<input type="checkbox"/>	0,39 bit (λάθος)
<input type="checkbox"/>	0,5 bit (λάθος)
<input type="checkbox"/>	1 bit (λάθος)
<input type="checkbox"/>	2,58 bit (σωστό)

Απάντηση	
<input checked="" type="checkbox"/> ✓	σχετίζεται με την εντροπία. (σωστό)
<input type="checkbox"/>	υπολογίζεται απευθείας από πιθανότητες. (σωστό)
<input type="checkbox"/>	εκφράζει τον μέσο παράγοντα διακλάδωσης των κόμβων του γράφου που περιγράφει τη γλώσσα. (σωστό)
<input type="checkbox"/>	υπολογίζεται από το WordNet. (λάθος)

Η εντροπία...  
(επιλέξτε όλα όσα ταιριάζουν)

Απάντηση	
<input checked="" type="checkbox"/> ✓	δείχνει πόσο "τακτοποιημένο" ή "ανακατεμένο" είναι ένα σύστημα. (σωστό)
<input type="checkbox"/>	καθορίζει πόση πληροφορία χρειάζομαστε για να περιγράψουμε πλήρως ένα σύστημα. (σωστό)
<input type="checkbox"/>	μετρείται σε bits. (σωστό)
<input type="checkbox"/>	υπολογίζεται από πιθανότητες (και λογάριθμους πιθανοτήτων). (σωστό)
<input type="checkbox"/>	σχετίζεται με την περιπλοκή (perplexity). (σωστό)
<input type="checkbox"/>	εφαρμόζεται μόνο σε γραμματικές τύπου 2. (λάθος)

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

Απάντηση	
<input checked="" type="checkbox"/> ✓	Σωστό (σωστό) Προκύπτει άμεσα από τον τύπο υπολογισμού της περιπλοκής από την εντροπία (αντικαθιστώντας την εντροπία με 0).
<input type="checkbox"/>	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

Οι κύριες εννοιολογικές συσχετίσεις που κωδικοποιεί το WordNet αφορούν:

Απάντηση

☒

✓

συνωνυμία

(σωστό)

☐

ανωνυμία

λάθος

☐

επωνυμία

λάθος

☐

μερωνυμία

(σωστό)

☐

τροπωνυμία

(σωστό)

☐

παρωνυμία

λάθος

☐

αντινομία

λάθος

☐

υπωνυμία

(σωστό)

☐

παρωνομία

λάθος

☐

αντωνυμία

(σωστό)

Το 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

Ταιριάζτε τις έννοιες με τις ερμηνείες τους:

Απάντηση

συνωνυμία

λεξίζες με αντίθετες έννοιες / λεξίζες με παροξμοίες ερμηνίες ✕

αντωνυμία

ρηγμάτα που το ερνα ερναι εξειδικευση του αλλου / λεξίζες με αντιθετες ερνοιες ✕

υπωνυμία

θυσιαστικά? με σχεση μερος προς ολλον / ουσιαστικά? που το ερνα ειρναι εξειδικευση του αλλου ✕

μερωνυμία

λεξίζες με παροξμοίες έννοιες / ουσιαστικά? με σχεση μερος προς ολον ✕

τροπωνυμία

θυσιαστικά? που το ερνα ερναι εξειδικευση του αλλου / ρηγμάτα που το ερνα ειρναι εξειδικευση του αλλου ✕

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