

## Savollarga Javoblar

1. Nishonga 40 ta o'q uzilgan, shundan 12 ta o'q nishonga tekkani qayd qilingan. Nishonga tegishlar nisbiy chastotasini toping.

$W = 0,3$

2. "Biword index" turi qanday amalga oshiriladi?

So'zlar juft-juft bo'lib indekslanadi

3. Quyidagi formulalardan qaysi biri terminni hujjatdagi logarifmik vaznini hisoblash uchun ishlatiladi?

$$w_{t,d} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{aks holda} \end{cases}$$

4. Quyidagi keltirilgan formulalardan qaysi biri Bayes formulasini beradi?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

5. Precision aniqlanish sohasini toping.

$[0;1]$

6. Mantiqiy amallar yordamida ishlab chiqilgan qidiruv tizimiga Brutus or

Caesar so'rovi berilganda qidiruv natijasi qanday bo'ladi?

Brutus yoki Caesar so'zlaridan hech bo'lmaganda bittasi qatnashgan hujjatlar taqdim etiladi

7. Atributlar bo'yicha kataloglangan va ularga kirish qulay bo'lishini ta'minlash maqsadida indekslangan ma'lumotlar to'plami nima deb ataladi?

Strukturalangan ma'lumot

8. Lug'at tuzish usullari to'g'ri ko'rsatilgan javobni belgilang

Heshlash, iyerarxik

9. Oltita tomoni bor o'yin soqqasi tashlangan. Soqqaning tushgan yoqlaridagi ochkolar toq

son bo'lish ehtimolini toping?

1/2

10. Qopda 45 ta qora va 5 ta oq shar bor. Tavakkaliga bitta shar olinadi. Olingan shar oq bo'lish ehtimolini toping.

0,1

11. Quyidagi so'zlardan qaysi biri "Olam" so'ziga bir levenshteyn masofada farq qiladi?

Odam

12. Precision formulasi to'g'ri ko'rsatilgan variantni toping. Bu yerda a – Q1 so'rovga tegishli

va tizim to'g'ri deb topgan hujjatlar soni, b – Q1 so'rovga tegishli lekin tizim topa olmagan

hujjatlar soni, c – Q1 so'rovga tegishli bo'lmagan va tizim to'g'ri deb topgan hujjatlar soni.

$R=a/(a+c)$

13. Tanga bir marta tashlangan. "Raqamli" tomon tushish ehtimolini toping.

1/2

14. Vektor uzunligini aniqlash formulasini toping

$$\|\vec{x}\|_2 = \sqrt{\sum_i x_i^2}$$

15. Sahifada termin vazni formula bo'yicha hisoblanganda, agar hujjatda so'z 100 marta takrorlansa ushbu so'zning shu hujjatdagi logarifmik vaznini hisoblang.

3

16. Tanga bir marta tashlangan. "Gerbli" tomon tushish ehtimolini toping.

1/2

17. belgilash nimani anglatadi?

Hujjatda termin necha marta kelganini anglatadi

18. Oltita tomoni bor o'yin soqqasi tashlangan. Soqqaning tushgan yoqlaridagi ochkolar juft

son bo'lish ehtimolini toping?

1/2

19. Vektorlarning skalyar ko'paytmasi to'g'ri ko'rsatilgan formulani toping?

$$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos \varphi$$

20. Joker koefisientini hisoblash to'g'ri keltirilgan formulani aniqlang

$$J(A,B) = |A \cap B| / |A \cup B|$$

21. Hozirda zamonaviy qidiruv tizimlari axborot qidirish bilan birgalikda yana qanday

xizmatlarni taklif qilmoqda?

Matnni tahrirlash xizmati

Tarjima xizmat

Xarita xizmati

Tavsiya qiluvchi tizimlar

22. Quyidagi mezonlardan qaysi biri temin-hujjat o'xshashligini baholashda ishlatiladi?

Cosinus o'xshashlik

23 Hujat-so'rov ju ligini baholash formulasi to'g'ri ko'rsatilgan variantni tanlang.

$$\sum_{t \in q \cap d} (1 + \log \text{tf}_{t,d})$$

24. Xtoy Xalq Respublikasida eng keng tarqalgan tizim qaysi?

Baidu

25. Quyidagilardan qaysi biri termin-document binar matritsasi kamchiliklari qatoriga kiradi?

Barcha javoblar to'g'ri

26. "бҗрфккшк (muharrir)" kabi hatoliklar qanday xatolik hisoblanadi?

Raskladka

27. Rokkio (Rocchio) algoritmi nima maqsadda ishlatiladi?

So'rov vektori  $q$  ni tegishli hujjatlarga maksimal yaqinlashtirish va tegishli bo'lmagan

hujjatlardan maksimal uzoqlashtirish maqsadida

28. . Quyidagi mulohazalardan qaysi biri to'g'ri?

Levenshteyn masofasi so'zlarni ma'nosi bo'yicha o'xshashligini hisoblamaydi

29. . Joker koefisienti aniqlanish sohasini belgilang

[0,1]

30. Heaps' law qoniniyati nima?

To'plamda tokenlar soniga qarab termin sonini taxminiy aniqlash qonuniyati

31. . Recall formulasi to'g'ri ko'rsatilgan variantni toping. Bu yerda  $a$  – Q1 so'rovga tegishli

va tizim to'g'ri deb topgan hujjatlar soni,  $b$  – Q1 so'rovga tegishli lekin tizim topa olmagan

hujjatlar soni,  $c$  – Q1 so'rovga tegishli bo'lmagan va tizim to'g'ri deb topgan hujjatlar soni

$$R=a/(a+b)$$

32. Quyidagilardan qaysi biri termin-document binar matritsasi kamchiliklari qatoriga

kiradi?

Barcha javoblar to'g'ri

33 . Quyidagi mulohazalardan qaysi biri to'g'ri?

Terminning matnda uchrashlar sonimik logarifmik ifodalsh orqali uni maqbul vaznlashga erishish

Mumkin

34 . Qidiruv tizimlarida klasterlash orqali qanday maslalani yechish mumkin

To'plam hujjatlarini guruhlariga ajratish uchun

35 . Quyidagi keltirilgan formulalardan qaysi biri Bayes formulasini beradi?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

36. Test to'plamda Q1 so'rovga tegishli 30 ta hujjat bor edi. A1 qidiruv tizimi Q1 so'rov

bo'yicha 50 ta hujjatni topdi. Topilgan hujjatlardan 20 tasi xaqiqatdan ham Q1 so'rovga

tegishli, qolgan 10 tasi esa tegishli emas edi. A1 qidiruv tizimining Recal ko'rsatkichi

nechchiga teng bo'ladi?

2/3

37 . Joker koefisientini hisoblash to'g'ri keltirilgan formulani aniqlang

$$J(A,B)=|A \cap B|/|A \cup B|$$

38 . Maksimal ehtimollik qiymati nechchiga teng?

1

39 . Lug'at tuzish usullari to'g'ri ko'rsatilgan javobni belgilang

Heshlash, iyerarxik

40 . Oltita tomoni bor Ikkita o'yin soqqasi tashlangan. Soqqalarning yoqlarida tushgan

ochkolar yig'indisi 5 ga teng bo'lishi ehtimolini toping.

1/9

41 . Precision aniqlanish sohasini toping.

[0;1]

42. Foydalanuvchi bilan dialog oʻrnatuvchi tizimlarga misol qaysi variantlarda keltirilgan?

Alisa

43. Dunyo boʻylab eng keng tarqalgan qidiruv tizimi qaysi?

Google

44. F1 boholash qiymatini aniqlanish sohasini toping.

[0;1]

45. Soʻrov-hujjat cosinus oʻxshashligi formulasi toʻgʻri koʻrsatilgan variantni toping.

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| |\vec{d}|}$$

46. A hodisaning sodir boʻlish ehtimoli qanday oraliqda aniqlanadi?

[0;1]

47. Quyidagi soʻzlardan qaysi biri "Olam" soʻziga bir levenshteyn masofada farq qiladi?

Odam

48. Oltita tomoni bor oʻyin soqqasi tashlangan. Soqqaning tushgan yoqlaridagi ochkolar ju

son boʻlish ehtimolini toping?

1/2

49. Oʻrtacha silliqlangan Precision (MAP- Mean average precision) qiymatini aniqlash

formulasi qaysi variantda toʻgʻri koʻrsatilgan?

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k$$

50. "Yashash yaxshi, Yaxshi yashash yanada yaxshi" matnida nechta token mavjud?

6 ta

51. Oʻz formatida saqlanadigan va foydalanilgunga qadar qayta ishlanmaydigan

ma'lumotlar to'plami nima deb ataladi?

Strukturalanmagan ma'lumot

52. Qidiruv tizimi ektivlik ko'rsatkichlarni belgilang

Stemming qilinadigan so'zlarni ajrata olishi darajasi;

53. Vektorlarning skalyar ko'paytmasi to'g'ri ko'rsatilgan formulani toping?

$$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos \varphi$$

54. Mantiqiy amallar yordamida ishlab chiqilgan qidiruv tizimiga Brutus and Caesar so'rovi

berilganda qidiruv natijasi qanday bo'ladi?

Brutus va Caesar so'zlari birgalikda qatnashgan hujjatlar taqdim etiladi

55. Quyidagi xususiyatlarning qaysi biri Strukturalanmagan ma'lumotlar to'plamiga

tegishli

Bulutli ma'lumotlar ko'llarida saqlanadi

56. 1. Vektor uzunligini aniqlash formulasini toping

$$\|\vec{x}\|_2 = \sqrt{\sum_i x_i^2}$$

57. To'la ehtimollik yig'indisi nechchiga teng?

1

58. Maksimal ehtimollik qiymati nechchiga teng?

1

59. Quyidagi keltirilgan formulalardan qaysi biri to'la ehtimol formulasini beradi?

$$P(A) = \sum_i P(B_i)P(A / B_i)$$

60. Test to'plamda Q1 so'rovga tegishli 30 ta hujjat bor edi. A1 qidiruv tizimi Q1 so'rov

bo'yicha 50 ta hujjatni topdi. Topilgan hujjatlardan 20 tasi xaqiqatdan ham Q1 so'rovga

tegishli, qolgani tegishli emas edi. A1 qidiruv tizimining Precision ko'rsatkichi nechchiga

teng bo'ladi?

2/5

61. Quyidagi keltirilgan formulalardan qaysi biri Bayes formulasini beradi?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

62. Axborotlarni izlashda so'zlar tuzulmasi o'rtasidagi farqni aniqlash uchun qanday usul

qo'llaniladi?

Levenshteyn masofa

63. Web qidiruv tizimlarida ko'proq qaysi indekslash algoritmidan faoydalaniladi?

taqsimlangan indekslash

64. Ehtimoliy tegishlilik mulohazasiga ko'ra  $t$  terminning tegishli deb belgilangan hujjatda

uchrash ehtimoli qanday aniqlanadi?

$$\checkmark \hat{P}(x_t = 1 | R = 1) = |VR_t| / |VR|$$

65. O'rtacha silliqlangan Precision (MAP- Mean average precision) qiymatini aniqlash

formulasi qaysi variantda to'g'ri ko'rsatilgan?

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k$$

66. Qidiruv tizimi to'plami deganda nima tushuniladi?

Qidiruv tizimidagi barcha hujjatlar to'plami

67.  $df_t$  belgilash nimani anglatadi?



To'plamda termin qatnashgan hujjatlar sonini anglatadi

68. Nishonga 40 ta o'q uzilgan, shundan 12 ta o'q nishonga tekkani qayd qilingan. Nishonga

tegishlar nisbiy chastotasini toping.

$W = 0,3$

69. Web qidiruv tizimlarida ko'proq qaysi indekslash algoritmidan faoydalaniladi?

taqsimlangan indekslash

70. Hozirda zamonaviy qidiruv tizimlari axborot qidirish bilan birgalikda yana qanday

xizmatlarni taklif qilmoqda?

Matnni tahrirlash xizmati

Tarjima xizmat

Xarita xizmati

Tavsiya qiluvchi tizimlar

71. Quyida nimani hisoblash algoritmi keltirilgan?

So'rov va hujjat cosinus o'xshashligi hisoblash

72. Nuqtalar o'rnini to'ldiring ... deb berilgan belgilar yoki xususiyatlarga ega bo'lgan

ob'ektlarning joylashuvini aniqlash bilan bog'liq operatsiyalar majmuiga aytiladi.

Qidiruv

73. Nuqtalar o'rnini to'ldiring .... deb Katta to'plamlardan (odatda kompyuter xotiralarida

saqlanadigan to'plamlardan) strukturalanmagan xarakterga ega kerakli axborotni

mazmunan qanoatlantiradigan (odatda matn) ma'lumotlarni qidirishga aytiladi.

Axborot qidirish

74. Berilgan so'rovni hujjatga o'xshashligini baholashning usuli qanday

hisoblanadi?

$$\text{Score}(q, d) = \sum_{t \in q \cap d} \text{tf.idf}_{t,d}$$

75. Recall me'zoni nima maqsadda ishlatiladi?

Qidiruv tizimi samaradorligini baholash uchun

76. Quyidagi mulohazalardan qaysi biri to'g'ri?

Terminning matnda uchrashlar sonimik logarifmik ifodalsh orqali uni maqbul vaznlashga

erishish mumkin

77. Quyidagi mulohazalardan qaysilari to'g'ri?

Cosinus o'xshashlikni baholash uchun so'zlar vektorli fazo modeli orqali ifodalanilishi kerak

Vektorli fazo modelida so'kov-hujjat o'xshashligini baholash uchun Cosinus o'xshashlikdan

Foydalaniladi

Vektorlar orasidagi burchak kichraygan sari cosinus o'xshashlik ortib boradi

78. Berilgan so'rovni hujjatga o'xshashligini baholashning  $\text{tf.idf}_{t,d}$  usuli qanday hisoblanadi?

$$\text{Score}(q, d) = \sum_{t \in q \cap d} \text{tf.idf}_{t,d}$$

79. Quyidagi mulohazalardan qaysi biri to'g'ri?

Terminning matnda uchrashlar sonimik logarifmik ifodalsh orqali uni maqbul vaznlashga

erishish mumkin

80. Quyidagi mulohazalardan qaysilari to'g'ri?

Cosinus o'xshashlikni baholash uchun so'zlar vektorli fazo modeli orqali ifodalanilishi kerak

Vektorli fazo modelida so'kov-hujjat o'xshashligini baholash uchun Cosinus o'xshashlikdan

foydalaniladi

Vektorli fazo modelida so'rov-hujjat o'xshashligini baholash uchun yevklid masofadan

foydalaniladi

Vektorlar orasidagi burchak kichraygan sari cosinus o'xshashlik ortib boradi

81. Quyidagi mulohazalardan qaysilari to'g'ri?

Cosinus o'xshashlikning eng katta qiymati 1 bo'ladi

Cosinus qiymatining kamayishi so'rov-hujjat o'xshashligini ortishiga olib keladi

Cosinus o'xshashlikning eng katta qiymati 0 bo'ladi

Vektor fazoli modelda terminlar koordinata o'qlari sifatida ifodalanadi

82. Rokkio (Rocchio) algoritmi nima maqsadda ishlatiladi?

So'rov vektori  $\vec{q}$  ni tegishli hujjatlarga maksimal yaqinlashtirish va tegishli bo'lmagan

hujjatlardan maksimal uzoqlashtirish maqsadida

83. O'rtacha Recall qiymatini aniqlash formulasi qaysi variantda to'g'ri ko'rsatilgan?

Bunday tushuncha mavjud emas

84. Web qidiruv tizimlarida ko'proq qaysi indekslash algoritmidan faoydalaniladi?

taqsimlangan indekslash

85. A hodisaning sodir bo'lish ehtimoli qanday oraliqda aniqlanadi?

[0;1]

86. Quyidagi mulohazalardan qaysi biri to'g'ri?

Terminning matnda uchrashlar sonimik logarifmik ifodalsh orqali uni maqbul vaznlashga erishish

mumkin

87. Berilgan so'rovni hujjatga o'xshashligini baholashning  $tf.idf_{t,d}$  usuli qanday hisoblanadi?

$$Score(q, d) = \sum_{t \in q \cap d} tf.idf_{t,d}$$

88. Foydalanuvchi bilan dialog oʻrnatuvchi tizimlarga misol qaysi variantlarda keltirilgan?

ChatGPT

Alisa

Yandex

Google

89. Quyidagi mulohazalardan qaysilari toʻgʻri?

Cosinus qiymatining kamayishi soʻrov-hujjat oʻxshashligini ortishiga olib keladi

Cosinus oʻxshashlikning eng katta qiymati 0 boʻladi

Vektor fazoli modelda terminlar koordinata oʻqlari sifatida ifodalanadi

Cosinus oʻxshashlikning eng katta qiymati 1 boʻladi

90.

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	1	1	0	0	0	1
Brutus	1	1	0	1	0	0
Caesar	1	1	0	1	1	1
Calpurnia	0	1	0	0	0	0
Cleopatra	1	0	0	0	0	0
mercy	1	0	1	1	1	1
worser	1	0	1	1	1	0

Ushbu termin-dokumen matritsasiga “not Caesar” soʻrovini berilsa natija qanday boʻladi

001000

91. Tanga bir marta tashlangan. “Raqamli” tomon tushish ehtimolini toping.

$\frac{1}{2}$

92.

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	1	1	0	0	0	1
Brutus	1	1	0	1	0	0
Caesar	1	1	0	1	1	1
Calpurnia	0	1	0	0	0	0
Cleopatra	1	0	0	0	0	0
mercy	1	0	1	1	1	1
worser	1	0	1	1	1	0

Ushbu termin-dokumen matritsasiga “Antony and Caesar” so’rovini berilsa natija qanday

bo’ladi

110001

93. Nishonga 40 ta o’q uzilgan, shundan 10 ta o’q nishonga tekkani qayd qilingan. Nishonga

tegishlar nisbiy chastotasini toping.

$W=0,25$

94. Mantiqiy qidiruv tizimlarini kamchiliklari to’g’ri ko’rsatilgan qatorni belgilang

Qidiruv so’rovini aynan yozish kerak (ya’ni agar bitta belgi hatto terilsa ham qidirilayotgan

ma’lumot topilmaydi) #Qidiruv natijalari soniga chegara qo’yish imkoni yo’q

95. Nuqtalar o’rnini to’ldiring ... deb berilgan belgilar yoki xususiyatlarga ega bo’lgan

ob'ektlarning joylashuvini aniqlash bilan bog’liq operatsiyalar majmuiga aytiladi.

Qidiruv

96. Termin vazni  $w_{t,d} = 1 + \log_{10} tf_{t,d}$  formula bo’yicha hisoblanganda, agar hujjatda so’z 100 marta takrorlansa ushbu so’zning shu hujjatdagi logarifmik vaznini hisoblang.

3

97. Matnni tokenlash deb nimaga aytiladi ?

Belgilar ketma-ketligidan so’zlarni ajratib olish

98. Naïve Bayes algoritmi qidiruv tizimlarida nima maqsadda foydalaniladi ?

Matnni tasniflash uchun

99.