Savollarga Javoblar

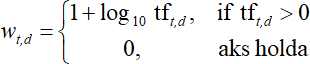
1. Nishonga 40 ta oʻq uzilgan, shundan 12 ta oʻq nishonga tekkani qayd qilingan. Nishonga tegishlar nisbiy chastotasini toping.

W= 0,3

2. “Biword index” turi qanday amalga oshiriladi?

Soʻzlar juft-juft boʻlib indekslanadi

1. Quydagi formulalardan qaysi biri terminni hujjatdagi logarifmik vaznini hisoblash uchun ishlatiladi?



1. Quyidagi keltirilgan formulalardan qaysi biri Bayes formulasini beradi?



5. Precision aniqlanish sohasini toping.

[0;1]

1. .Mantiqiy amallar yordamida ishlab chiqilgan qidiruv tizimiga Brutus or Caesar soʻrovi berilganda qidiruv natijasi qanday boʻladi?

Brutus yoki Caesar soʻzlaridan hech boʻlmaganda bittasi qatnashgan hujjatlar taqdim etiladi

7. Atributlar boʻyicha kataloglangan va ularga kirish qulay boʻlishini taʼminlash maqsadida indekslangan ma'lumotlar toʻplami nima deb ataladi?

Strukturalangan maʼlumot

8. Lugʻat tuzish usullari toʻgʻri koʻrsatilgan javobni belgilang

Heshlash, iyerarxik

9. Oltita tomoni bor oʻyin soqqasi tashlangan. Soqqaning tushgan yoqlaridagi ochkolar toq

son boʻlish ehtimolini toping?

1/2

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

0,1

11. Quyidagi soʻzlardan qaysi biri “Olam” soʻziga bir levenshteyn masofada farq qiladi?

Odam

12. Precision formulasi toʻgʻri koʻrsatilgan variantni toping. Bu yerda a – Q1 soʻrovga tegishli

va tizim toʻgʻri deb topgan hujjatlar soni, b – Q1 soʻrovga tegishli lekin tizim topa olmagan

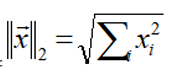
hujjatlar soni, c – Q1 soʻrovga tegishli boʻlmagan va tizim toʻgʻri deb topgan hujjatlar soni.

R=a/(a+c)

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

1/2

14. Vektor uzunligini aniqlash formulasini toping



15. Sahifada termin vazni formula boʻyicha hisoblanganda, agar hujjatda

soʻz 100 marta takrorlansa ushbu soʻzning 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 oʻyin soqqasi tashlangan. Soqqaning tushgan yoqlaridagi ochkolar juft

son boʻlish ehtimolini toping?

1/2

19. Vektorlarning skalyar koʻpaytmasi toʻgʻri koʻrsatilgan formulani toping?



20 . Joker koefisientini hisoblash toʻgʻri keltirilgan formulani aniqlang

J(A,B)=|A∩B|/|A∪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 oʻxshashligini baholashda ishlatiladi?

Cosinus oʻxshashlik

23 Hujat-soʻrov ju ligini baholash formulasi toʻgʻri koʻrsatilgan variantni tanlang.



24. Xtoy Xalq Respublikasida eng keng tarqalgan tizim qaysi?

Baidu

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

Barcha javoblar toʻgʻri

26 . “Ьгрфккшк (muharrir)” kabi hatoliklar qanday xatolik hisoblanadi?

Raskladka

27. Rokkio (Rocchio) algoritmi nima maqsadda ishlatiladi?

Soʻrov vektori 𝑞 ni tegishli hujjatlarga maksimal yaqinlashtirish va tegishli boʻlmagan

hujjatlardan maksimal uzoqlashtirish maqsadida

28 . . Quyidagi mulohazalardan qaysi biri toʻgʻri?

Levenshteyn masofasi soʻzlarni maʼnosi boʻyicha oʻxshashligini hisoblamaydi

29 . . Joker koefisienti aniqlanish sohasini belgilang

[0,1]

30 . Heaps' law qoniniyati nima?

Toʻplamda tokenlar soniga qarab termin sonini taxminiy aniqlash qonuniyati

31. . Recall formulasi toʻgʻri koʻrsatilgan variantni toping. Bu yerda a – Q1 soʻrovga tegishli

va tizim toʻgʻri deb topgan hujjatlar soni, b – Q1 soʻrovga tegishli lekin tizim topa olmagan

hujjatlar soni, c – Q1 soʻrovga tegishli boʻlmagan va tizim toʻgʻri deb topgan hujjatlar soni

R=a/(a+b)

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

kiradi?

Barcha javoblar toʻgʻri

33 . Quyidagi mulohazalardan qaysi biri toʻgʻri?

Terminning matnda uchrashlar sonimik logarifmik ifodalsh orqali uni maqbul vaznlashga erishish

Mumkin

34 . Qidiruv tizimlarida klasterlash orqali qanday maslalarni yechish mumkin

Toʻplam hujjatlarini guruhlarga ajratish uchun

35 . Quyidagi keltirilgan formulalardan qaysi biri Bayes formulasini beradi?



36. Test toʻplamda Q1 soʻrovga tegishli 30 ta hujjat bor edi. A1 qidiruv tizimi Q1 soʻrov

boʻyicha 50 ta hujjatni topdi. Topilgan hujjatlardan 20 tasi xaqiqatdan ham Q1 soʻrovga

tegishli, qolgan 10 tasi esa tegishli emas edi. A1 qidiruv tizimining Recal koʻrsatkichi

nechchiga teng boʻladi?

2/3

37 . Joker koefisientini hisoblash toʻgʻri keltirilgan formulani aniqlang

J(A,B)=|A∩B|/|A∪B|

38 . Maksimal ehtimollik qiymati nechchiga teng?

1

39 . Lugʻat tuzish usullari toʻgʻri koʻrsatilgan javobni belgilang

Heshlash, iyerarxik

40 . Oltita tomoni bor Ikkita oʻyin soqqasi tashlangan. Soqqalarning yoqlarida tushgan

ochkolar yigʻindisi 5 ga teng boʻlishi ehtimolini toping.

1/9

41 . Precision aniqlanish sohasini toping.

[0;1]

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

Alisa

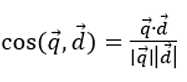
43. Dunyo boʻylab eng keng tarqalgan qidiruv tizimi qaysi?

Google

44 . F1 boholash qiymatini aniqlanish sohasini toping.

[0;1]

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



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?



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

6 ta

51. Oʻz formatida saqlanadigan va foydalanilgunga qadar qayta ishlanmaydigan

ma'lumotlar toʻplami nima deb ataladi?

Strukturalanmagan maʼlumot

52. Qidiruv tizimi e ektivlik koʻrsatkichlarni belgilang

Stemming qilinadigan soʻzlarni ajrata olishi darajasi;

53. Vektorlarning skalyar koʻpaytmasi toʻgʻri koʻrsatilgan formulani toping?



54. Mantiqiy amallar yordamida ishlab chiqilgan qidiruv tizimiga Brutus and Caesar soʻrovi

berilganda qidiruv natijasi qanday boʻladi?

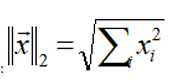
Brutus va Caesar soʻzlari birgalikda qatnashgan hujjatlar taqdim etiladi

55. Quyidagi xususiyatlarning qaysi biri Strukturalanmagan maʼlumotlar toʻplamiga

tegishli

Bulutli maʼlumotlar koʻllarida saqlanadi

56. 1. Vektor uzunligini aniqlash formulasini toping



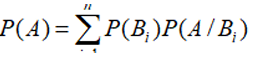
57. Toʻla ehtimollik yigʻindisi nechchiga teng?

 1

58. Maksimal ehtimollik qiymati nechchiga teng?

 1

59. Quyidagi keltirilgan formulalardan qaysi biri toʻla ehtimol formulasini beradi?



60. Test toʻplamda Q1 soʻrovga tegishli 30 ta hujjat bor edi. A1 qidiruv tizimi Q1 soʻrov

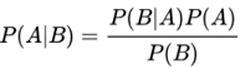
boʻyicha 50 ta hujjatni topdi. Topilgan hujjatlardan 20 tasi xaqiqatdan ham Q1 soʻrovga

tegishli, qolgani tegishli emas edi. A1 qidiruv tizimining Precision koʻrsatkichi nechchiga

teng boʻladi?

2/5

61. Quyidagi keltirilgan formulalardan qaysi biri Bayes formulasini beradi?



62. Axborotlarni izlashda soʻzlar tuzulmasi oʻrtasidagi farqni aniqlash uchun qanday usul

qoʻllaniladi?

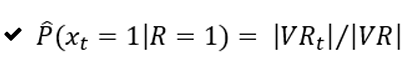
Levenshteyn masofa

63. Web qidiruv tizimlarida koʻproq qaysi indekslash algoritmidan faoydalaniladi?

taqsimlangan indekslash

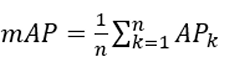
64. Ehtimoliy tegishlilik mulohazasiga koʻra 𝑡 terminning tegishli deb belgilangan hujjatda

uchrash ehtimoli qanday aniqlanadi?



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

formulasi qaysi variantda toʻgʻri koʻrsatilgan?



66. Qidiruv tizimi toʻplami deganda nima tushuniladi?

Qidiruv tizimidagi barcha hujjatlar toʻplami

67.  belgilash nimani anglatadi?

Toʻplamda termin qatnashgan hujjatlar sonini anglatadi

68. Nishonga 40 ta oʻq uzilgan, shundan 12 ta oʻq nishonga tekkani qayd qilingan. Nishonga

tegishlar nisbiy chastotasini toping.

W= 0,3

69. Web qidiruv tizimlarida koʻproq 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?

Soʻrov va hujjat cosinus oʻxshashligi hisoblash

72. Nuqtalar oʻrnini toʻldiring ... deb berilgan belgilar yoki xususiyatlarga ega boʻlgan

ob'ektlarning joylashuvini aniqlash bilan bogʻliq 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 qidirsh

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

hisoblanadi?



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‘кov-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  usuli qanday

hisoblanadi?



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‘кov-hujjat o‘xshashligini baholash uchun Cosinus o‘xshashlikdan

foydalaniladi

Vektorli fazo modelida so‘zov-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  ni tegishli hujjatlarga maksimal yaqinlashtirish va tegishli bo‘lmagan

hujjatlardan maksimal uzoqlashtirish maqsadida

83. Oʻrtacha Recall qiymatini aniqlash formulasi qaysi variantda toʻgʻri koʻrsatilgan?

Bunday tushuncha mavjud emas

84. Web qidiruv tizimlarida koʻproq qaysi indekslash algoritmidan faoydalaniladi?

taqsimlangan indekslash

85. A hodisaning sodir boʻlish ehtimoli qanday oraliqda aniqlanadi?

[0;1]

86. Quyidagi mulohazalardan qaysi biri toʻgʻri?

Terminning matnda uchrashlar sonimik logarifmik ifodalsh orqali uni maqbul vaznlashga erishish

mumkin

87. Berilgan soʻrovni hujjatga oʻxshashligini baholashning  usuli qanday hisoblanadi?



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

ChatGPT

 Alisa

Yandex

Google

89. Quyidagi mulohazalardan qaysilari toʻgʻri?

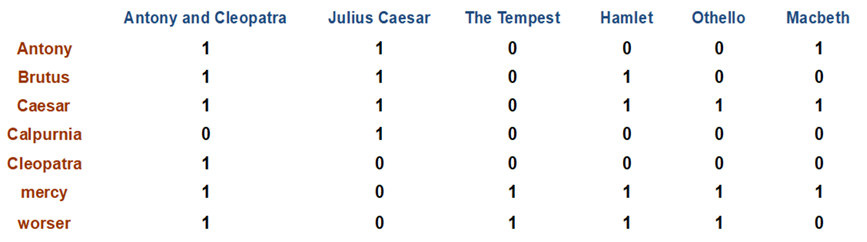
Cosinus qiymatining kamayishi soʻrov-hujjat oʻxshashligini ortishiga olib keladi

Cosinus oʻxshashlikning eng katta qiymati 0 boʻladi

Vektor fazoli modelda terminlar koordinata oʻqlari sifatida ifodalanadi

Cosinus oʻxshashlikning eng katta qiymati 1 boʻladi

90.



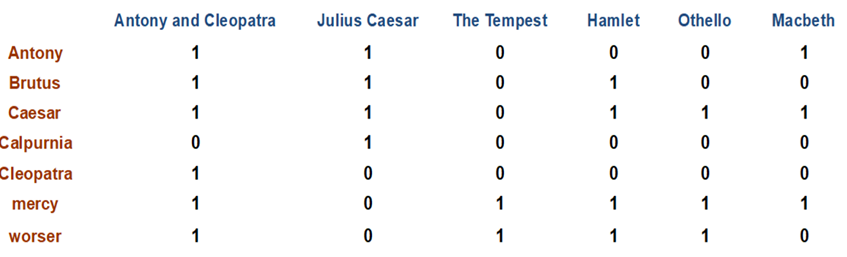
Ushbu termin-dokumen matritsasiga “not Caesar” soʻrovini berilsa natija qanday boʻladi

001000

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

 ½

92.



Ushbu termin-dokumen matritsasiga “Antony and Caesar” soʻrovini berilsa natija qanday

boʻladi

 110001

93. Nishonga 40 ta oʻq uzilgan, shundan 10 ta oʻq nishonga tekkani qayd qilingan. Nishonga

tegishlar nisbiy chastotasini toping.

W=0,25

94. Mantiqiy qidiruv tizimlarini kamchiliklari toʻgʻri koʻrsatilgan qatorni belgilang

Qidiruv soʻrovini aynan yozish kerak (yaʼni agar bitta belgi hato terilsa ham qidirilayotgan

maʼlumot topilmaydi) #Qidiruv natijalari soniga chegara qoʻyish imkoni yoʻq

95. Nuqtalar oʻrnini toʻldiring ... deb berilgan belgilar yoki xususiyatlarga ega boʻlgan

ob'ektlarning joylashuvini aniqlash bilan bogʻliq operatsiyalar majmuiga aytiladi.

Qidiruv

96. Termin vazni  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.