

4.

Vaziyat (TF-IDF orqali):

Oliy attestatsiya komissiyasi (OAK) har kuni turli fan sohalariga oid katta hajmdagi hujjatlar bilan shug'ullanadi. Ushbu hujjatlarni samarali boshqarish va tasniflash uchun OAK mashinali o'rgatish usullaridan foydalangan holda hujjatlarni tasniflash tizimini ishlab chiqishni maqsad qilgan. Tizim kiruvchi hujjatlarni "Matematika", "Adabiyot", "Geografiya" kabi toifalarga aniq tasniflashi kerak.

Amaliy masala: Quyida har bir matn bitta xujjatda kelgan deb faraz qilinadi:

1. Uchburchakning ikki tominini yig'indisi har doim uchunchi tomonidan katta degan aksioma har doim bajariladi.
2. Zamonaviy o'zbek adabiyotining yorqin nomoyondalaridan biri bu Erkin Vohidov hisoblanadi.
3. O'zbekiston xududida sug'oriladigan yerlarning aksariyatiga suv Amudaryo va Sirdaryo kabi yirik daryolar orqali boradi.

So'raladi:

Hujjatdagi so'zning butun korpusga nisbatan ahamiyatini ifodalovchi har bir hujjatdagi har bir so'z uchun TF-IDF ballini hisoblang.