

QIDIRUV ALGORITMLARI VA ULARNI OPTIMALLASHTIRISH

Nabijonov Ravshanbek Muxammadjon o‘g‘li

Azamov Shohruhmirzo Alisher o‘g‘li

Nazirov Qodirjon Nodirovich

Muhammad al-Xorazmiy nomidagi TATU Farg‘ona filiali

E-mail: rnabijonov19@gmail.com

Annotatsiya

Ushbu maqolada Qidiruv algoritmlari va optimallashtirish masalalari bilan bog‘liq bo‘lgan barcha usullari ko‘rib chiqildi. Shu bilan bir qatorda qidiruvning afzalliklari va kamchiliklari tahlil qilinib, qaysi qidiruv tizimda kerakli ma‘lumot olish haqida ma‘lumotlar keltirilgan.

Kalitso‘zlar: www, qidiruv algoritmlari, interpretasiya, Ketma-ket qidiruv algoritmi, SequentialSearch.

Qidiruv tizimi (inglizcha: search engine) — kompyuterda, kompyuter tarmog‘ida yoki butunjahon web tarmog‘ida World Wide Web saqlanayotgan ma‘lumotlarni qidirishga mo‘ljallangan dasturdir. Internet qidiruv tizimi Information Retrieval systemdan kelib chiqqan. Bu sistema ma‘lumotlar bazasi uchun kalit-so‘zlar indeksini yaratib, natijada ushbu kalit-so‘z qatnashgan qidiruv so‘rovnomasiga javob sifatida natijalar ro‘yxatini ko‘rsatish imkoni paydo bo‘ladi. Qidiruv tizimining afzalliklaridan biri bu kerakli ma‘lumot yoki axborotni qisqa vaqt ichida kompyuter xotirasidan topib uni foydalanuvchiga yetkazishdir. Qidiruv tizimining asosiy vazifalari hamda tarkibiy qismlari quyidagilardir:

- Indeks yaratish va yangilab turish (hujjatlardagi ma‘lumotlar strukturasi)
 - Qidiruv so‘rovnomalarini bajarish
 - Qidiruv natijasini imkon qadar mazmunli, ya‘ni tushunarli shaklda ko‘rsatish
- Odatda ma‘lumotlar qidirish avtomatik ravishda amalga oshiriladi, masalan WWWda Webcrawler, alohida kompyuterda esa foydalanuvchi belgilagan Indexlar ro‘yxatidan barcha ma‘lumotlarni o‘qish.

Qidiruv tizimi turlari Qidiruv tizimlarini bir qancha belgilariga ko‘ra turlarga ajratish mumkin. Quyidagi belgilar misol sifatida tanlangandir. Yangi qidiruv tizimini yaratishda ushbu belgilarning hammasini olish shart emas.

Axborotlar turi Turli xil qidiruv tizimlari turli xil axborot turlarini qidira oladi. Avvaliga umumiy qilib ma’lumotlar turini ko‘rsatish mumkin, masalan, matn, tasvir, videotasvir, audiofayl va b. Natijalar ham ushbu turlarga bog‘liq holda tartibga solinadi. Matn qidirilganda natija sifatida ushbu matndan kalit-so‘z qatnashgan bir qism ko‘rsatiladi.

Axborotlar manba’i Klassifikatsiyaning yana bir turi bu qidiruv tizimi topgan axborotning manba’i hisoblanadi. Ko‘pincha qidiruv tizimining nomi axborotlani qaysi manba’dan qidirishini ko‘rsatadi.

- Web qidiruv tizimi — WWWdan ma’lumot qidiradi
- Usenet qidiruv tizimi — butun dunyoga taqsimlangan munozaralar mediumi (vositasi) hisoblanadigan Usenetdan ma’lumot qidiradi
- Intranet qidiruv tizimi — alohida tashkilotning ichki tarmog‘idagi kompyuterlardan ma’lumot qidiradi
- Desktop qidiruv tizimi — alohida kompyuterning ma’lumotlar bazasidan ma’lumot qidiradigan dastur.

Amalga oshirish usuli

- Indeksga asoslangan qidiruv tizimi — hozirda eng muhim tizim. Bu tizim tegishli ma’lumotlarni topib, indeks yaratadi. Indeks shunday bir strukturaki, bunda saqlangan ma’lumotlar keyingi qidiruv vaqtida ham qo‘llaniladi. Lekin bu tizimda indeksni yangilab turish va saqlash muammoli bo‘lsa ham, qidiruv jarayoni tez ekanligi bu tizimning ustunligidir. Bu strukturani ko‘pincha invers indeks deb ham atashadi.
- Meta qidiruv tizimi — qidiruv so‘rovnomasini bir vaqtning o‘zida bir necha indeksga asoslangan qidiruv tizimlariga jo‘natadi va natijani turli kombinatsiyada ko‘rsatadi. Ustunlik tomoni — ko‘p miqdorda ma’lumotlar topishi va implementatsiya jarayoni osonroqligidir. Ushbu qidiruv tizimi kam uchraydigan ma’lumotlarni qidirishda yaxshi hisoblanadi.
- Gibrid formadagi qidiruv tizimi — odatda nisbatan kichik indeksga ega, lekin boshqa qidiruv tizimlariga ham so‘rovnoma jo‘natib, natijani kombinatsiyalangan



tarzda ko'rsatishi mumkin. Kam uchraydigan ma'lumotlarni qidirishda yaxshi emas.

- Guruhlantirilgan qidiruv tizimi — nisbatan yangi, bunda qidiruv so'rovnomasi o'z qidiruv tizimiga ega bir necha kompyuterlarga jo'natiladi va natija umumlashtiriladi. Markazlashtirilmagani uchun uzilib qolish xavfi yo'q, ishonchli, lekin markaziy tsenzura (cheklashlar) qo'yishning iloji yo'q.

Interpretasiya Foydalanuvchining qidiruv so'rovnmasi qidirish jarayoni boshlanishidan oldin kerakli qidiruv algoritmiga tushunarli qilib interpretasiya (tarjima) qilinadi. Bu esa o'z navbatida so'rov sintaksisini imkon qadar soddaroq qilish va murakkab so'rovlarni bajarish imkonini yaratishga xizmat qiladi. Ko'plab qidiruv tizimlari matematik mantiqdagi bul operatsiyalari yordamida turli kalit-so'zlarning mantiqiy birikmasini hosil qila oladi. Natijada esa ma'lum kalit-so'zga ega web-sahifani topish imkoni paydo bo'ladi. Eng yangi qidiruv tizimi tabiiy tilni tushunadigan qidiruv tizimini bo'lib, uni semantik qidiruv tizimi deb atashadi. **Eng ko'p qo'llaniladigan qidiruv tizimlarining belgilari** Eng ko'p qo'llaniladigan belgilar kombinatsiyasi WWW dan axborot manbai sifatida, matnli ma'lumotlar uchun esa HTML-format va qidiruvni amalga oshirishda esa indeks yaratishdan foydalanadi. So'rovnimani interpretasiya qilish ko'pincha soddagina sintaksisga asoslanga holda amalga oshiriladi. Eng yirik qidiruv tizimi hosoblanmish Google, Yahoo Search va Microsoft Live Search mana shu uslubda ishlaydi. **Natijani ko'rsatish** Qidiruv natijasi ko'rsatadigan sahifa ko'pgina qidiruv tizimlari tomonidan ikkiga: tabiiy ro'yxat va homiylar linklariga ajratiladi. Homiylar linklari to'lov asosida qidiruv indeksiga kiritilgan bo'lsa, tabiiy ro'yxatda qidirilayotgan kalit-so'z qatnashgan bo'lsagina ko'rsatiladi. Foydalanuvshiga qidiruv tizimidan foydalanishni osonlashtirish maqsadida natijalar muvofiqlik boy'icha (Ranking) saralanadi, lekin har bir qidiruv tizimi saralash uchun o'z me'yorlariga egadir. Bu me'yorlarga quyidagilar kiradi:

- ma'lumotning ahamiyati (Google uchun PageRank-ko'rsatkichi)
- har bir topilgan malumotda qidirilayotgan kalit-so'zning qanchalik ko'p uchrashi va joylashgan o'rni
- so'ralgan ma'lumotni baholash va soni
- boshqa sahifalardan ushbu kalit-so'z qatnashgan sahifaga qanchalik ko'p linklar berilganligi



- link beriladigan sozlardir.

Ro'yxatdan zarur axborotni qidirish - nazariy dasturlashning fundamental masalalaridan biri hisoblanadi. Qidiruv algoritmini tahlil qilishda, qidirilayotgan axborot kompyuterda ma'lumotlar massivi ko'rinishida joylashgan deb faraz qilamiz. Yozuvlar yoki ro'yxat elementlari massivda ketma-ket joylashadi va ular o'rtasida bo'shliq yo'q. Ro'yxatdagi yozuvlar 1 dan N gacha tartiblangan bo'ladi. Aslida yozuvlar maydonlardan tashkil topgan bo'ladi, bizni kalit deb ataluvchi maydonlardan bittasining qiymati qiziqtiradi. Ro'yxatlar kalit maydonlari qiymati bo'yicha tartiblangan yoki tartiblanmagan bo'lishi mumkin. Aniq qiymatni qidirish masalasi biror elementni tanlash masalasi bilan bog'liq. Aytaylik, bizga kattaligi bo'yicha beshinchi element, oxiridan yettinchi yoki o'rta qiymatli element kerak.

Qidiruv - bu kompyuter xotirasida ma'lumotlarni qayta ishlash jarayonida qo'llaniladigan asosiy amallardan biri hisoblanadi. Bu amalning mazmuni shundan iboratki, berilgan argument bo'yicha massiv elementlari orasidan shu argumentga mos ma'lumotni (elementni) aniqlashdan iborat.

Ixtiyoriy ma'lumot (yoki tuzilma) elementi qandaydir belgisi bilan boshqasidan farq qiladi. Ushbu belgi **kalit** deyiladi. Kalit takrorlanmas bo'lishi mumkin, ya'ni tuzimadagi mavjud bitta element uchun aynan shu kalit qo'llaniladi. Bunday kalit **birlamchi** deb ataladi. Elementlarning ma'lum guruhi uchun takrorlanishi mumkin bo'lgan kalit **ikkilamchi** kalit deyiladi, ushbu kalit bo'yicha ham qidiruv tashkil qilish mumkin. Ma'lumot elementlarining kalitlari alohida joyda to'plangan (boshqa jadvalda) bo'lishi yoki o'zi alohida yozuvdagi biror bir maydonda saqlanishi mumkin. Bunday ajratib olingan yoki alohida faylda saqlanadigan **tashqi kalit** deyiladi. Agar kalit yozuvda joylashgan bo'lsa, u holda **ichki kalit** deyiladi.

Berilgan argument bo'yicha qidiruv berilgan argument **kaliti bilan aniqlangan algoritm** deyiladi. Qidiruv algoritmi ishlashi natijasi ushbu ma'lumotda joylashishi mumkin yoki u jadvalda mavjud bo'lmasligi mumkin. Ma'lumotning mavjud bo'lmaslik holati ikkita amalda yuz berishi mumkin:

- 1) berilgan qiymatning yo'qligini aniqlashda;
- 2) berilgan qiymatni jadvalga qo'yishda.

Masalan, bizga ro'yxat elementi kaliti sifatida **k** berilgan bo'lsin. Har bir **k(i)** kalitda **r(i)**-ma'lumot mavjud. Qidiruv argumenti - **Key**. Unga yozuv ma'lumoti **rec** mos keladi. Jadvalda ma'lumotlar tuzilmasiga bog'liq holda qidiruvning bir nechta shakllari farqlanadi.

Bu masalani yechish uchun ikki xil yondoshuv mavjud: ketma-ket qidiruv va



indeksli ketma-ket qidiruv. Ushbu qidiruv algoritmlarini o'rganib chiqamiz.

Ketma-ket qidiruv algoritmi

Qidiruv algoritmlarida biror aniq elementni mavjud ro'yxat elementlarini birma-bir qarab chiqish orqali qidirib topish masalasi hal qilinadi. Ketma-ket qidiruv algoritmidan ro'yxatning saralanganligi ahamiyatga ega bo'lmas-da, lekin saralangan ro'yxatda eng yaxshi natijaga erishiladi. Odatda qidiruv kerakli elementning ro'yxatda bor yoki yo'qligini shunchaki tekshirish uchun emas, balki shu kalit-qiyamatga ega bo'lgan ma'lumotni ajratib olish uchun ham qo'llaniladi. Masalan, kalit-qiyamat qidirilayotgan elementning tartib raqami yoki boshqa unikal (yagona) identifikator bo'lishi mumkin. Kerakli kalit topilgandan so'ng dastur shu kalitga mos ma'lumotlarni o'zgartirishi yoki shunchaki barcha yozuvlarni ajratib chiqarishi mumkin. Har qanday holatda ham qidiruv algoritmi kalitning joylashgan o'rnini aniqlash masalasini yechish uchun qo'llaniladi. SHuning uchun ham qidiruv algoritmlari kerakli kalitdan tarkib topgan yozuv indeksini natija sifatida ajratib beradi. Agar kalit-qiyamat topilmasa, u holda qidiruv algoritmi massivning yuqori chegarasidan katta bo'lgan indeks qiymatini qaytaradi. Maqsadga erishish uchun ro'yxat elementlari 1 dan N gacha bo'lgan sonlar yordamida raqamlangan bo'lsin deb faraz qilamiz. Bu holatda qidirilayotgan kalit ro'yxatda mavjud bo'lmasa, algoritm 0 qiymatni beradi (1-rasm). Sodda uchun ajratib olinadigan kalit-qiyamatlar ro'yxatda takrorlanmaydi deb qabul qilinadi.

	i	k	r
индекс	1	8	...
	2	136	...
	3	4	...

КЛИТ	n-1	17	...
	n	234	...

Ma'lumot (axborot) maydoni

1-rasm

Ketma-ket qidiruv algoritmi ro'yxatning birinchi elementidan boshlab oxirgi elementgacha qidirilayotgan elementni topilmaguncha qarab chiqiladi. Bundan kelib

chiqadiki, kalit qiymati ro'yxatda qancha uzoqda turgan bo'lsa, qidiruv shuncha uzoq davom etadi (vaqtga nisbatan). Bu holatni ketma-ket qidiruv algoritmi tahlilida e'tiborga olish zarur bo'ladi.

Ketma-ket qidiruv algoritmining to'liq ko'rinishi quyidagicha:

SequentialSearch (list, target,N)

list // qidirish amalini bajarish uchun ro'yxat

target //kalit qiymati

N //ro'yxatdagi elementlar soni

for i=1 to N do

if (target=list[i]) return

i

end if end

for return 0

Algoritmining tahlili. Ketma-ket qidiruv algoritmidagi ikki eng yomon holat mavjud. Birinchi holatda qidiruvdagi element ro'yxat oxirida joylashgan bo'ladi. Ikkinchi holatda esa, qidiruvdagi element ro'yxatda mavjud bo'lmaydi. Har ikki holatda ham necha marta taqqoslash amali bajarilishini qarab chiqamiz. Ro'yxat elementlarining kalit qiymati unikal (takrorlanmas) deb qaraydigan bo'lsak, qidiruvdagi elementga moslik oxirgi yozuvda uchrasa, u holda oxirgi taqqoslash keraksiz bo'lishi mumkin. Lekin algoritm oxirgi elementga qadar taqqoslash amalini bajaradi. Natijada N ta taqqoslash amalga oshiriladi (bu yerda N- ro'yxatdagi elementlar soni).

Qidirilayotgan qiymatning ro'yxatda yo'qligini tekshirish uchun uni ro'yxatning barcha elementlari bilan taqqoslab chiqishga to'g'ri keladi. Agar biror element bilan taqqoslash qoldirilsa, u holda qidirilayotgan element rostdan ham ro'yxatda yo'qligini aniqlab bo'lmaydi. Bu esa ro'yxatning barcha elementlarini qarab chiqishni talab etadi va bu holatda ham N marta taqqoslash amalga oshiriladi.

Umuman olganda, qidirilayotgan element ro'yxatning oxirgi elementi yoki unda mavjud emasligini aniqlash uchun N marta taqqoslash amali bajariladi. N har qanday qidiruv algoritmining eng yuqori chegarasi hisoblanadi, agar taqqoslashlar N dan ko'p bajarilsa, hech bo'lmaganda ro'yxatning bitta elementi bilan ikki marta taqqoslash bajarilganligini bildiradi. Bundan kelib chiqadiki, ortiqcha ish bajarilgan va algoritmni yaxshilash talab etiladi.

O'rtacha holat tahlili. Qidiruv algoritmi uchun ikkita o'rtacha holat mavjud. Birinchi holatda qidiruv muvaffaqiyatli, ikkinchi holatda qidiruv samarasiz, ya'ni qidirilayotgan element ro'yxatda mavjud bo'lmasligi mumkin.



Agar qidirilayotgan element ro'yxatda bo'lsa, u holda N marta mumkin bo'lgan holatlardan biriga teng bo'ladi: birinchi, ikkinchi, uchinchi va h.k. Bu holatlarning barchasini teng ehtimolli deb faraz qilamiz, ya'ni qidirilayotgan element mavjud holatlarning birida uchrashi ehtimoli $1/N$ ga teng bo'ladi.

Ushbu xulosalardan kelib chiqqan holda quyidagi savollar tug'iladi:

1. Agar qidirilayotgan element ro'yxatda birinchi turgan bo'lsa, necha marta taqqoslash bajariladi?
2. Ikkinchi bo'lsa-chi?
3. Uchinchisi bo'lsa-chi?
4. N -element bo'lsa-chi?

Agar yuqorida aytilgan fikrlardan kelib chiqadigan bo'lsak, ushbu savollarga javob tayyor, chunki har bir holat uchun mos ravishda 1, 2, 3 va N marta taqqoslash bajariladi. Bu esa N ta holatdan har biri qidirilayotgan elementning ro'yxatda joylashish tartibiga mos kelishini bildiradi. Natijada, o'rtacha holatlar uchun quyidagi tenglikga erishamiz:

$$A(N) = 1 + \frac{1}{N} \sum_{i=1}^N i = \frac{N(N+1)}{2N} = \frac{N+1}{2} \quad (i)$$

Agar qidirilayotgan element ro'yxatda mavjud bo'lmasa taqqoslashlar soni $N+1$ gacha oshib ketadi. Element ro'yxatda mavjud bo'lmasa, u holda qidiruv uchun N ta taqqoslash zarurligini yuqorida ko'rib chiqdik. Agar barcha $N+1$ taqqoslashlar teng ehtimolli deb faraz qilsak, u holda quyidagi formula kelib chiqadi:

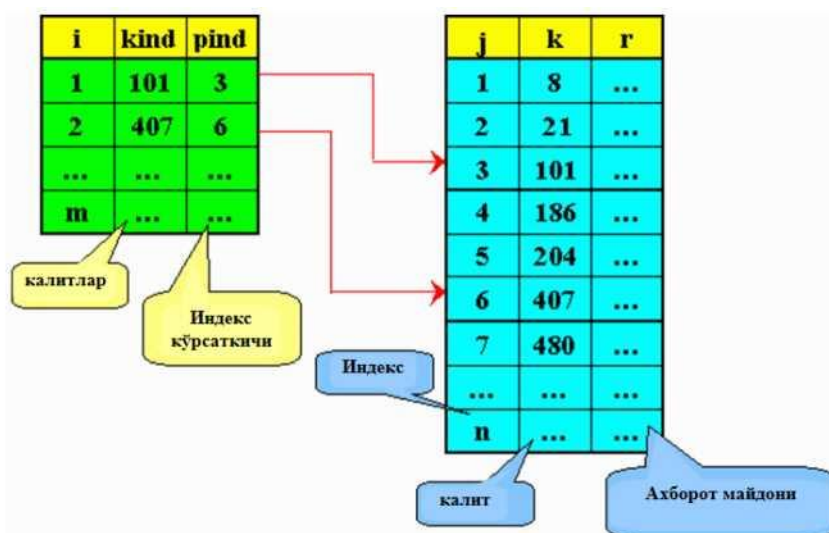
$$\sum_{i=1}^N i = \frac{N(N+1)}{2} \quad (2)$$

Agarda ro'yhat elementlari soni N juda katta son bo'lsa u holda $1/(N+1)$ ning qiymati 0 ga yaqin son bo'ladi.

Indeksli ketma-ket qidiruv algoritmi

Bu qidirish usulida ikkita jadval tashkil etiladi: o'z kalitlariga ega bo'lgan ma'lumotlar jadvali (o'sish tartibida joylashtirilgan) va ma'lumotlar kalitlaridan

tashkil topgan indekslar jadvali. Bu kalitlar asosiy jadvaldan oldindan aniqlangan intervalda olinadi (2-rasm).



2-rasm. Indeksli ketma-ket qidirish usuliga misol.

Bu yerda birinchi berilgan argument bo'yicha indekslar jadvalidan ketma- ketlikda qidirish amalga oshiriladi. Kalitlarni ko'rib chiqishda berilgan kalitdan kichigi topilsa, u holda ushbu kichik kalitni asosiy jadvaldagi qidirishning eng quyi chegarasi - **low** ga joylashtiramiz, xuddi shunday berilgan kalitdan katta deb topilgan kalitni (**kind** > **key**) yuqori **hi** ga joylashtiramiz. Misol uchun, **key** = 101.

Algoritmning tahlili. Agar barcha holatlar uchun teng ehtimollik deb hisoblansa, u holda qidirish samaradorligini quyidagicha aniqlash mumkin:

Belgilashlar kiritamiz: **m** - indeks o'lchami; **m = n / p**;

p - qadamlar o'lchami

$$Q = \frac{m+1}{2} \cdot \frac{p+1}{2} + 1$$

Yuqoridagi formulada **Q** ni **p** bo'yicha differetsiallaymiz va nolga tenglashtiramiz:

$$\frac{d^d}{dp} (Qp + y + 1) = 0$$

Bu yerda

$$p^2=n; p =4n$$

Q ning ifodasida **popt** ning qiymatini o'rniga qo'yib, quyidagicha taqqoslashlar soniga ega bo'lamiz:

$$Q = 4n + 1$$

Q(Vn) indeksli ketma-ket qidirishning samaradorlik darajasi hisoblanadi.

Asosiy adabiyotlar

1. Alisher o'g'li, Azamov Shohruhmirzo. "MA'LUMOTLARNI MANTIQUIY IZLASH MODEL, USULI VA ALGORITMLARI." Journal of new century innovations 15.2 (2022): 41-43.
2. Muxammadjon o'g'li, Nabijonov Ravshanbek, Mamirxo'jayev Muxammadamin Mavlonjon o'g', and To'ychiboyev Abbasjon Erali o'g'. "ZAMONAVIY TIBBIYOTDA QO'LLANADIGAN STATISTIK MA'LUMOTLARNI KOMPYUTER YORDAMIDA QAYTA ISHLASH ALGORITMLARI VA DASTURLARI." (2022): 147-158.
3. Khalilov, D. "СУНЬИЙ ИНТЕЛЛЕКТ ВА РАДИАЛ НЕЙРОН ТАРМОҚЛАРИНИНГ МАТЕМАТИК АСОСЛАРИ." Science and innovation 1.A6 (2022): 664-671.
4. e-Government Applications, by Nag Yeon Lee and Kwangsok Oh, printed in Scand-Media Corp., Ltd., Republic of Korea, in 2011, pages - 109.
5. Gorana Celebic, Dario Ilija Rendulic. Basic Concepts of Information and Communication Technology. Zagreb, 2011
6. S.S.Kosimov. Axborot texnologiyalari. T.: Aloqachi, 2006.