



# TOSHKENT AMALIY FANLAR UNIVERSITETI

## Ma'lumotlar tuzilmasi va algoritmlar fani



“Kompyuter injiniring” kafedrası

Katta o'qituvchi Kendjayeva Dildora Xudayberganovna

**Satrlarda qismaniy satrlarni qidirish algoritmlar. Qismaniy satrlarni izlashda primitiv algoritmlarning kamchiligi, Qismaniy satrlarni qidirish algoritmlarining turlari.**





# *Asosiy adabiyotlar:*

1. Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). Introduction to algorithms. MIT press.
2. O. R. Yusupov, I. Q. Ximmatov, E. Sh. Eshonqulov. Algoritmlar va berilganlar strukturalari. Oliy o'quv yurtlari uchun o'quv qo'llanma. – Samarqand: SamDU nashri. 2021-yil, 204 bet.
3. Xayitmatov O'.T., Inogomjonov E.E., Sharipov B.A., Ruzmetova N., Ma'lumotlar tuzilmasi va algoritmlari fanidan o'quv qo'llanma
4. Rahimboboeva D. "Ma'lumotlar tuzilmasi va algoritmlari" fanidan o'quv qo'llanma – T.: TDIU, 2011.-135 bet.



# ***MA'RUZA REJASI***



Satlarda qismaniy satrlarni qidirish algoritmlar.



Qismaniy satrlarni izlashda primitiv algoritmlarning kamchiligi.



Primitiv algoritmning muvaffaqiyatsizligi.

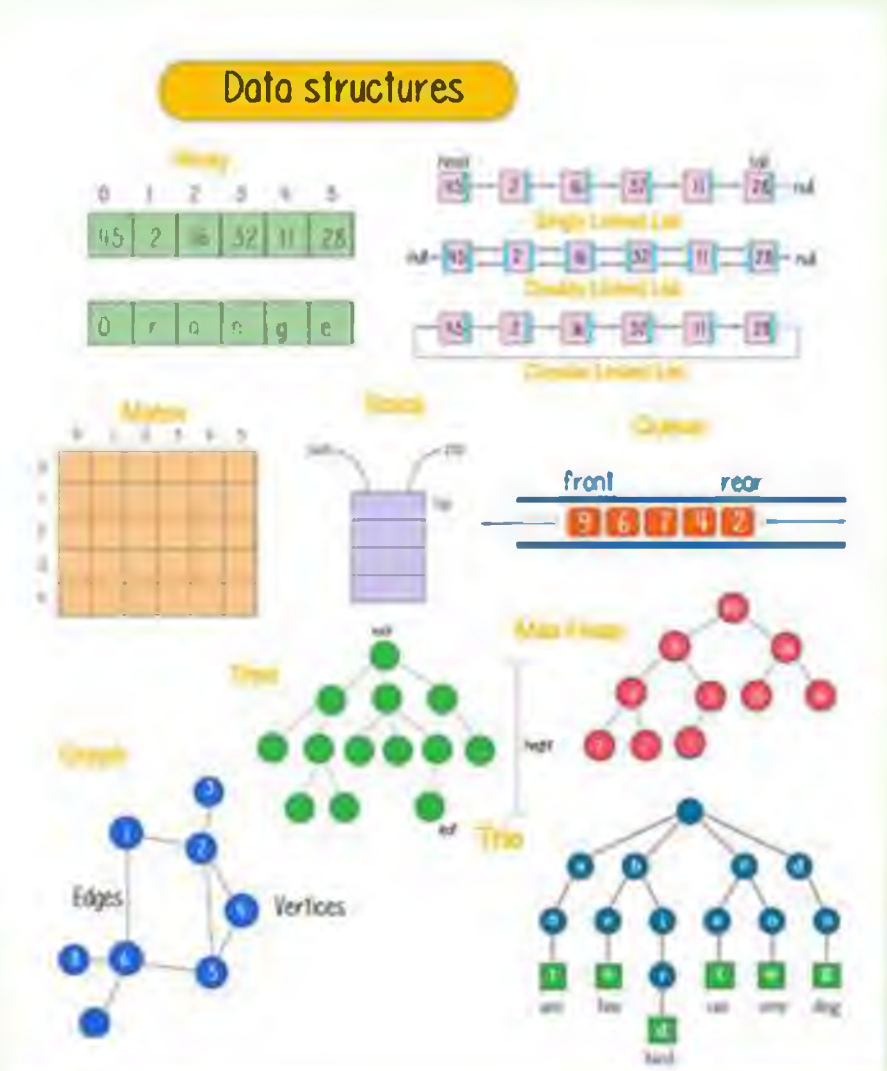


Qismaniy satrlarni qidirish algoritmlarining turlari.



# Satrlardan qismaniy satrni qidirish algoritmi

- **Satrlardan qismaniy satrni qidirish algoritmi** - bu matnda (*text*) qismaniy satr (*pattern*) topishga imkon beradigan satrlar ustidagi algoritmlar sinfi. U matn muharrirlari, MBBT, qidiruv tizimlari, dasturlash tillari va boshqalarda ***o'rnatilgan funksiya*** sifatida ishlatiladi.





# Primitiv algoritmning muvaffaqiyatsizligi

- Qidiruv vazifalarida qidiruv satrni **“igna”** (inglizchadan - **“needle”**) va qidiruv o'tkaziladigan satrni **“g'aram”** (ingliz tilidan - **“haystack”**) deb belgilash odat tusiga kirgan. Shuningdek, biz qidirish olib boriladigan alifboni  $\Sigma$  bilan belgilaymiz.
- Agar satrlar birdan boshlab raqamlangan deb hisoblasak, eng oddiy **“qo'pol kuch” (Brute force)** algoritmi (sodda algoritm) quyidagicha bo'ladi:

```
for i=0...|haystack|-|needle|  
  for j=0...|needle|  
    if haystack[i+j + 1] <> needle[j]  
      then goto 1  
output("Topildi: ", i+1)
```



# Satrlardan qismaniy satrni qidirish algoritmi

Bugungi kunda qismaniy satrlarni qidirish algoritmlarining xilma- xilligi mavjud. Dasturchi bunday omillarga qarab, mosini tanlashi kerak.

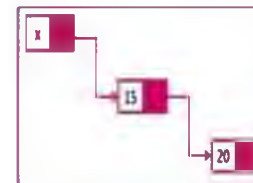
1. **Optimallashtirish kerakmi yoki primitiv algoritm yetarlimi?** Jimlik bo'yicha, bu dasturlash tillarining standart kutubxonalari amalga oshiradi.

2. **Foydalanuvchining "dushmanligi".** Boshqacha aytganda: foydalanuvchi ataylab algoritm sekin bajariladigan ma'lumotlarni aniqlaydimi? Eng oddiy holatda  $O(\text{haystack} \cdot |\text{needle}|)$  ball qo'yadigan juda oddiy algoritmlar mavjud.

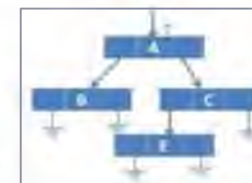
3. **Tilning grammatikasi qidiruvni "o'rtacha" tezlashtiradigan** ba'zi evristikalarga dushman bo'lishi mumkin.



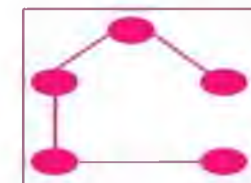
Sorting



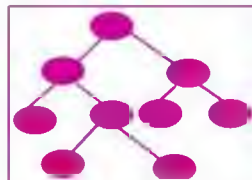
Link list



list



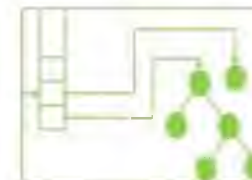
spanning tree



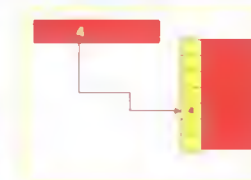
Tree



Graph

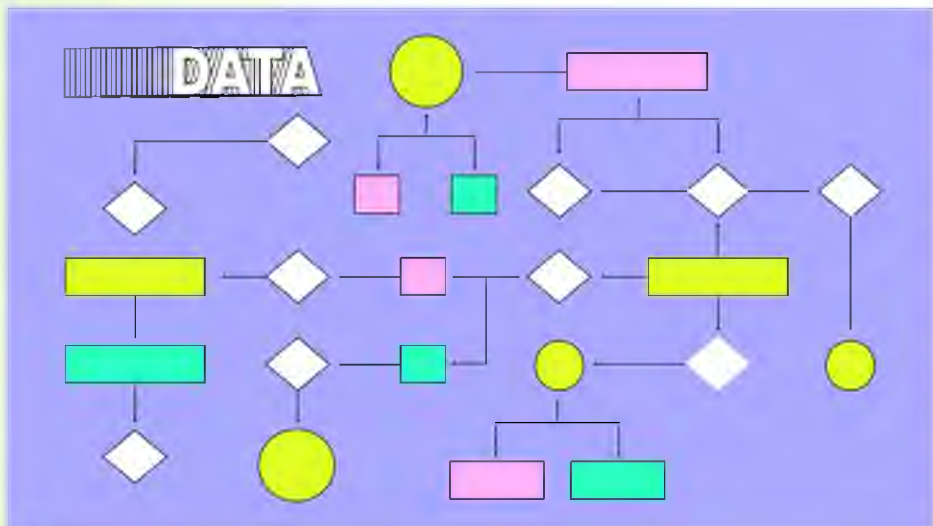


Stack



Hashing

# Satrlardan qismaniy satrni qidirish algoritmi



4. **Protessor arxitekturası.** Ba'zi protsessorlarda avtomatik kattalashtirish yoki SIMD amallari mavjud bo'lib, ular sizga ikkita operativ xotirani tez taqqoslashga imkon beradi.

5. **Alifbo o'lchami.** Ko'p algoritmlar (ayniqsa, oxirigacha taqqoslashga asoslangan), mos kelmaydigan belgi bilan bog'liq evristikaga ega.

6. **"haystack"ni indekslash qobiliyati.** Agar mavjud bo'lsa, qidiruv juda tezlashadi.

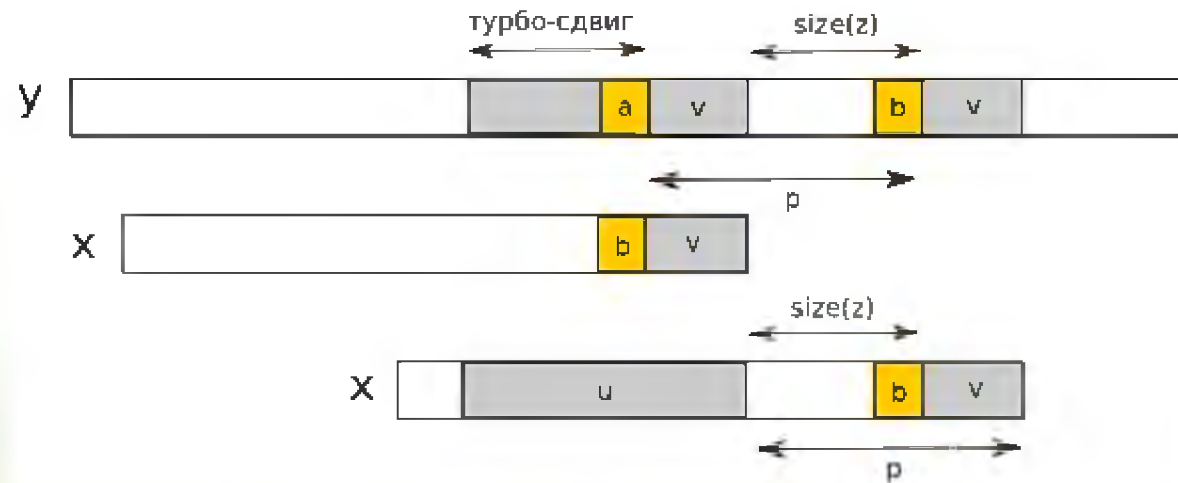
7. **Bir vaqtning o'zida bir nechta satrlarni qidirish kerakmi?** Ba'zi algoritmlarning yon xususiyatlari (Axo-Korasik, ikkilik algoritmi) bunga imkon beradi.





# Satrlardan qismaniy satrni qidirish algoritmi

- Qoida tariqasida, matn tahrirlovchisida **Boyer-Mur-Xorspul kabi eng oddiy evristik algoritmni** olish kifoya-hatto juda sekin kompyuter ham bir soniya ichida qidirishni amalga oshira oladi. Agar matn hajmi gigabaytda o'lchanadigan bo'lsa yoki qidiruv ko'plab so'rovlarni bajaradigan serverda ishlayotgan bo'lsa, siz eng muvaffaqiyatli algoritmni tanlashingiz kerak bo'ladi.



# Rabin-Karp algoritmi.

**Rabin-Karp algoritmi**-bu matnni xeshlash yordamida berilgan satrdan ichki satrni qidiradigan qidirish algoritmi. U 1987-yilda Maykl Rabin va Richard Karp tomonidan ishlab chiqilgan.

Algoritm kamdan-kam hollarda bitta qismaniy satrni topish uchun ishlatiladi, lekin muhim nazariy ahamiyatga ega va bir xil uzunlikdagi bir nechta qismaniy satr uchun moslikni topishda juda samarali.

Rabin Karp Algorithm

Step 1

String	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B
Pattern	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B

Step 2

String	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B
Pattern	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B

Step 3

String	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B
Pattern	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B

Step 4

String	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B
Pattern	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B

Step 5

String	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B
Pattern	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B

Step 6

String	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B
Pattern	0	1	2	3	4	5	6	7	8	9
	A	B	A	B	C	A	B	D	A	B



# Misol

Quyidagi misol orqali Rabin-Karp algoritmini ko'rib chiqamiz.

Berilgan matn S= "aevesapng"

Izlanadigan satr P= "esap"

0	1	2	3	4	5	6	7	8
a	e	v	e	s	a	P	n	g

0	1	2	3
e	s	a	P

Quyida simvollar uchun xesh-kod keltirilgan:

a	→	1
b	→	2
c	→	3
d	→	4
e	→	5
f	→	6
...	→	...
z	→	26





### 1-qadam

0	1	2	3
e	s	a	p
↓	↓	↓	↓
5	19	1	16

Xesh-kod  
qiymati:  
 $5+19+1+1$   
 $6=41$

### 2-qadam

0	1	2	3	4	5	6	7	8
a	e	v	e	s	a	p	n	
↓	↓	↓	↓					
1	5	22	5					

Xesh kod  
qiymati:  
 $1+5+22+5$   
 $= 33$

### 3-qadam

0	1	2	3	4	5	6	7	8
a	e	v	e	s	a	p	n	
	↓	↓	↓	↓				
	5	22	5	19				

Xesh kod  
qiymati:  
 $5+22+5+1$   
 $9 = 51$

### 4-qadam

0	1	2	3	4	5	6	7	8
a	e	v	e	s	a	p	n	
		↓	↓	↓	↓			
		22	5	19	1			

Xesh kod  
qiymati:  
 $22+5+19+$   
 $1 = 47$

### 5-qadam

0	1	2	3	4	5	6	7	8
a	e	v	e	s	a	p	n	
			↓	↓	↓	↓		
			5	19	1	16		

Xesh kod  
qiymati:  
 $5+19+1+1$   
 $6 = 41$

### 6-qadam

0	1	2	3	4	5	6	7	8
a	e	v	e	s	a	p	n	
				↓	↓	↓	↓	
				19	1	16	14	

Xesh kod  
qiymati:  
 $19+1+16+$   
 $14 = 50$

### 7-qadam

0	1	2	3	4	5	6	7	8
a	e	v	e	s	a	p	n	
					↓	↓	↓	
					1	16	14	7

Xesh kod  
qiymati:  
 $1+16+14$   
 $+7= 38$

**Eslatma:** Bu yerda xesh funksiyasini yaratish yoki aniqlashning turli usullari mavjud. Yaxshi tushunish uchun oddiy xesh funksiyasidan foydalanildi.



# Boyer-Mur algoritmi.

- 1977-yilda **Robert Boyer** va **Jey Mur** tomonidan ishlab chiqilgan, matnda oldindan ishlov berish imkoniyati bo'lmagan taqdirda, satrda qisman satrni topish algoritmlari orasida eng tezkori hisoblanadi.

Boyer-Moore search

THIS IS A TEST

TEST

Letters	T	E	S	*
Values	1	2	1	4



# Boyer-Mur algoritmi

1. q w t e **e** q e w q r w q w r q r  
q w r q **r**

2. q w t e e q **e r** q r w q w r q r  
q **w r** q r

3. q w t e e q e w q r w q w r **q** r  
q w r q **r**

4. q w t e e q e w q r w **q w r** q r  
**q w r** q r

*Algoritm g'oyasi* quyidagicha:

- ☐ Chapdan o'ngga skanerlash, o'ngdan chapga taqqoslash.
- ☐ To'xtash belgisini toppish
  - agar taqqoslanadigan birinchi harf mos kelmasa, shablon eng yaqiniga o'tkaziladi
  - to'xtash belgisi bo'lmasa, shablon uning orqasiga siljiydi
- ☐ Mos keladigan qo'shimchani toppish
  - agar 1 yoki undan ortiq belgi mos kelsa, shablon bu qo'shimchani birinchi mos kelishiga qadar o'ngga siljiydi





# To'xtatish belgisi jadvali.

- Qismaniy satrdagi elementning oxirgi o'rnini belgilaydi (oxirgi harfdan tashqari). Agar qismaniy satrda element bo'lmasa, jadvalga 0 kiritiladi (bittadan raqamlash uchun) Misol. Qismaniy satr: qwrqr

1. q t e e q r **w** q w r e e  
q w r q **r**  
2. q t e e q r **w** q w r e e  
q **w** r q r

Simvol	q	w	r	e	t
Oxirgi pozitsiya	4	2	3	0	0



# Suffiks jadvali.

- Mumkin bo'lgan barcha qo'shimchalar uchun jadvalda qismaniy satrni o'zgartirish kerak bo'lgan eng kichik miqdor ko'rsatilgan, u yana qo'shimchaga mos keladi. Agar bunday siljishning iloji bo'lmasa, satrning uzunligi ko'rsatilgan.
- Misol. Qismaniy satr: qwrqr

Suffiks	Bo'sh	r	qr	rqr	wrqr	qwrqr
qadam	1	2	5	5	5	5

1. q t e e q r **w** q w r e e q w r q **r**

2. q t e e q r **w** q w r e e q **w** r q r

3. q t e e q r w q **w** r e e q w r **q** r

4. q t e e q r w q w **r** e e q w **r** q r



# *Algoritmning murakkabligi.*

- $O(|\text{haystack}| + |\text{needle}| + |\Sigma|)$  davriy bo'lmagan qisman satrlar bo'yicha
- $O(|\text{haystack}| \cdot |\text{needle}| + |\Sigma|)$  davriy
- haystack - berilgan satr, needle - qisman satr,  $\Sigma$  - solishtirish uchun alifbo
- 1991-yilda Koul shuni ko'rsatdiki, davriy bo'lmagan sxemalar bo'yicha, algoritmlar satr bo'ylab to'liq o'tishda  $3 \cdot |\text{haystack}|$  tadan ko'p bo'lmagan taqqoslashni amalga oshiradi.





---

***Do you have  
any questions?***

---

