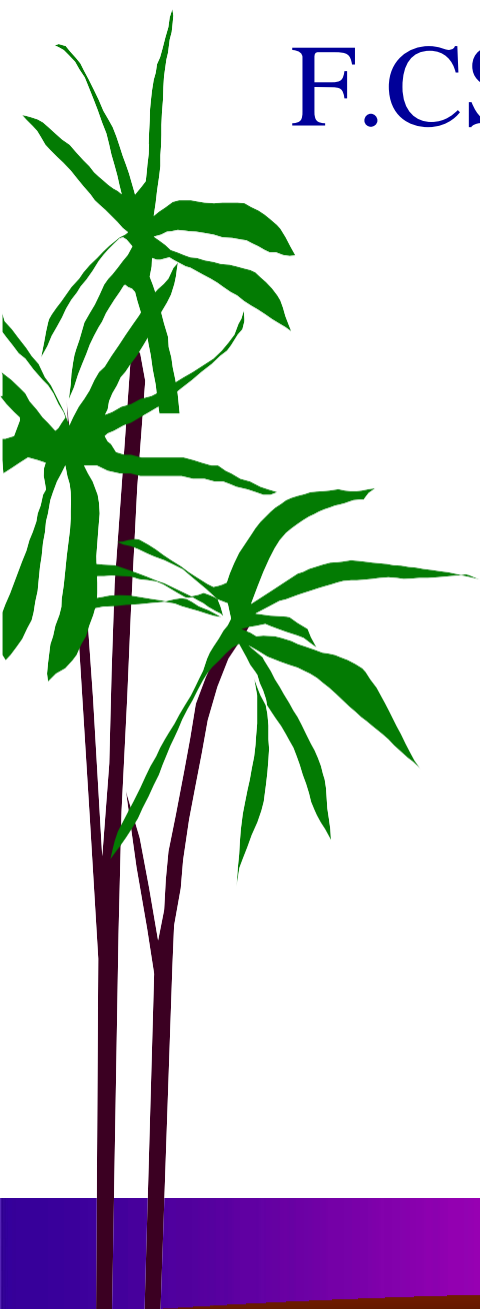
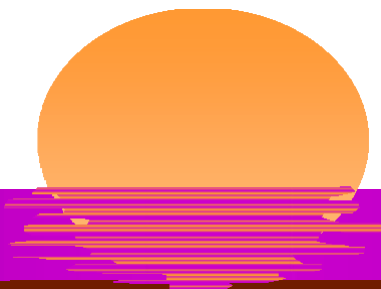


Ү.СS203 - Өгөгдлийн бүтэц ба алгоритм 2022-2023

Д.Золбоо
МТ-н салбарын багш





Халилтыг зохицуулах



- Шинэ хос (**key, element**) -ийн хувьд багцны үүр дүүрэн бол халилт үүсдэг.
- Халилтыг зохицуулахдаа:
 - Хэш хүснэгтээс голдуу дүүрэн бус байдаг багцыг хай.
 - Шугаман тандалт.
 - Квадрат тандалт.
 - Санамсаргүй тандалт.
 - Багц бүрт ижил үүртэй бүх хосуудын жагсаалтыг хадгалах замаар халилтаас зайлсхийж болно.
 - Массив шугаман жагсаалт.
 - Гинж.

Шугаман тандалт – Get ба Put

- $\text{divisor} = b$ (багцын тоо) = 17.
- Багцын үүр = $\text{key} \% 17$.

0	4				8				12				16				
34	0	45				6	23	7				28	12	29	11	30	33

- 6, 12, 34, 29, 28, 11, 23, 7, 0, 33, 30, 45 ГЭСЭН
түлхүүртэй хосуудыг хийлээ

Шугаман тандалт – Remove

0	4				8				12				16				
34	0	45				6	23	7				28	12	29	11	30	33

- **remove(0)**

0	4				8				12				16			
34		45				6	23	7			28	12	29	11	30	33

- Чөлөөлөгдсөн багцыг ашиглаж болох хосыг хайх.

0					4					8					12					16
34	45					6	23	7				28	12	29	11	30	33	4		

Шугаман тандалт – remove(34)

0	4				8				12				16			
34	0	45				6	23	7			28	12	29	11	30	33

0	4				8				12				16			
	0	45				6	23	7			28	12	29	11	30	33

- Чөлөөлөгдсөн багцыг ашиглаж болох хосыг хайх.

0	4				8				12				16			
0		45				6	23	7			28	12	29	11	30	33

0	4				8				12				16			
0	45					6	23	7			28	12	29	11	30	33

Шугаман тандалт – remove(29)

0	4				8				12				16			
34	0	45				6	23	7			28	12	29	11	30	33

0	4				8				12				16			
34	0	45				6	23	7			28	12		11	30	33

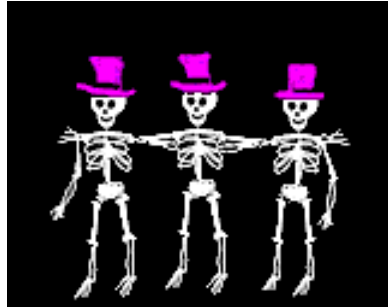
- Чөлөөлөгдсөн багцыг ашиглаж болох хосыг хайх.

0	4				8				12				16			
34	0	45				6	23	7			28	12	11		30	33

0	4				8				12				16			
34	0	45				6	23	7			28	12	11	30		33

0	4				8				12				16			
34	0					6	23	7			28	12	11	30	45	33 ₆

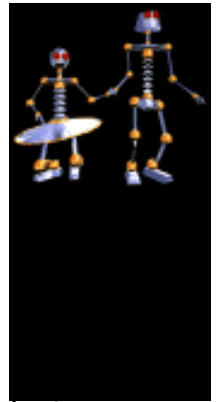
Шугаман тандалтын үзүүлэлт



0	4				8				12				16				
34	0	45				6	23	7				28	12	29	11	30	33

- get/put/remove үйлдлийн муу тохиолдлын хугацаа $\Theta(n)$, үүнд n – хүснэгт дэх хосын тоо.
- Бүх хос нэг үүрт ороход үүснэ.

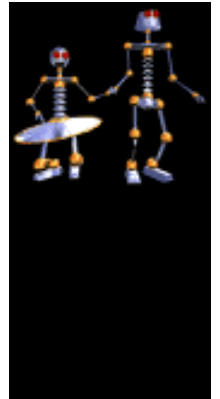
Дундаж үзүүлэлт



0	4				8				12				16			
34	0	45				6	23	7			28	12	29	11	30	33

- α = ачааллын нягтрал = (хосын тоо)/b.
 - $\alpha = 12/17$.
- S_n = амжилттай хайлтаар шалгах багцын (дундаж) тоо (n - том тоо бол)
- U_n = амжилтгүй хайлтаар шалгах багцын (дундаж) тоо (n - том тоо бол)
- тэгвэл put , remove үйлдлүүдийн хугацаа U_n -аар тодорхойлогдоно

Дундаж үзүүлэлт



- $S_n \sim \frac{1}{2}(1 + 1/(1 - \alpha))$
- $U_n \sim \frac{1}{2}(1 + 1/(1 - \alpha)^2)$
- $0 \leq \alpha \leq 1$.

α	S_n	U_n
0.50	1.5	2.5
0.75	2.5	8.5
0.90	5.5	50.5

$\alpha \leq 0.75$ байхыг
зөвлөдөг.

Хэш хүснэгтийн зохиомж

- Өгөгдсөн шаардлагаас хамаарч, ачааллын нягтралын зөвшөөрөгдөх дээд хэмжээг тогтоох.
- Амжилттай хайлт хийхэд 10 –с илүүгүй харьцуулалт хэрэгтэй бол (шаардлага).
 - $S_n \sim \frac{1}{2}(1 + 1/(1 - \alpha))$
 - $\alpha \leq 18/19$
- Амжилтгүй хайлт хийхэд 10 –с илүүгүй харьцуулалт хэрэгтэй бол (шаардлага).
 - $U_n \sim \frac{1}{2}(1 + 1/(1 - \alpha)^2)$
 - $\alpha \leq 4/5$
- Иймд $\alpha \leq \min\{18/19, 4/5\} = 4/5$.

Хэш хүснэгтийн зохиомж

- Динамик хүснэгтийн хэмжээ.
 - Ачааллын нягтрал хүссэн хэмжээнээс хэтэрвэл ($4/5$ манай жишээнд), хэш хүснэгтийн хэмжээг одоогийнхоос нь ойролцоогоор 2 дахин нэмэгдүүлнэ.
- Тогтсон хүснэгтийн хэмжээ.
 - Хосын максимум тоог мэднэ.
 - 1000 -с илүүгүй хостой.
 - Ачааллын нягтрал $\leq 4/5 \Rightarrow b \geq 5/4 * 1000 = 1250$.
 - b (ө.х. **divisor**) 20 –с доошхи анхны тоонд хуваагддаггүй сондгой тоо, эсхүл анхны тоо байхаар сонго.

Синонимын шугаман жагсаалт

- Багц бүрт ижил үүртэй хосуудын шугаман жагсаалтыг халдгалах.
- Шугаман жагсаалт эрэмбэлэгдсэн байж болно.
- Шугаман жагсаалт нь массив эсхүл гинж байж болно.

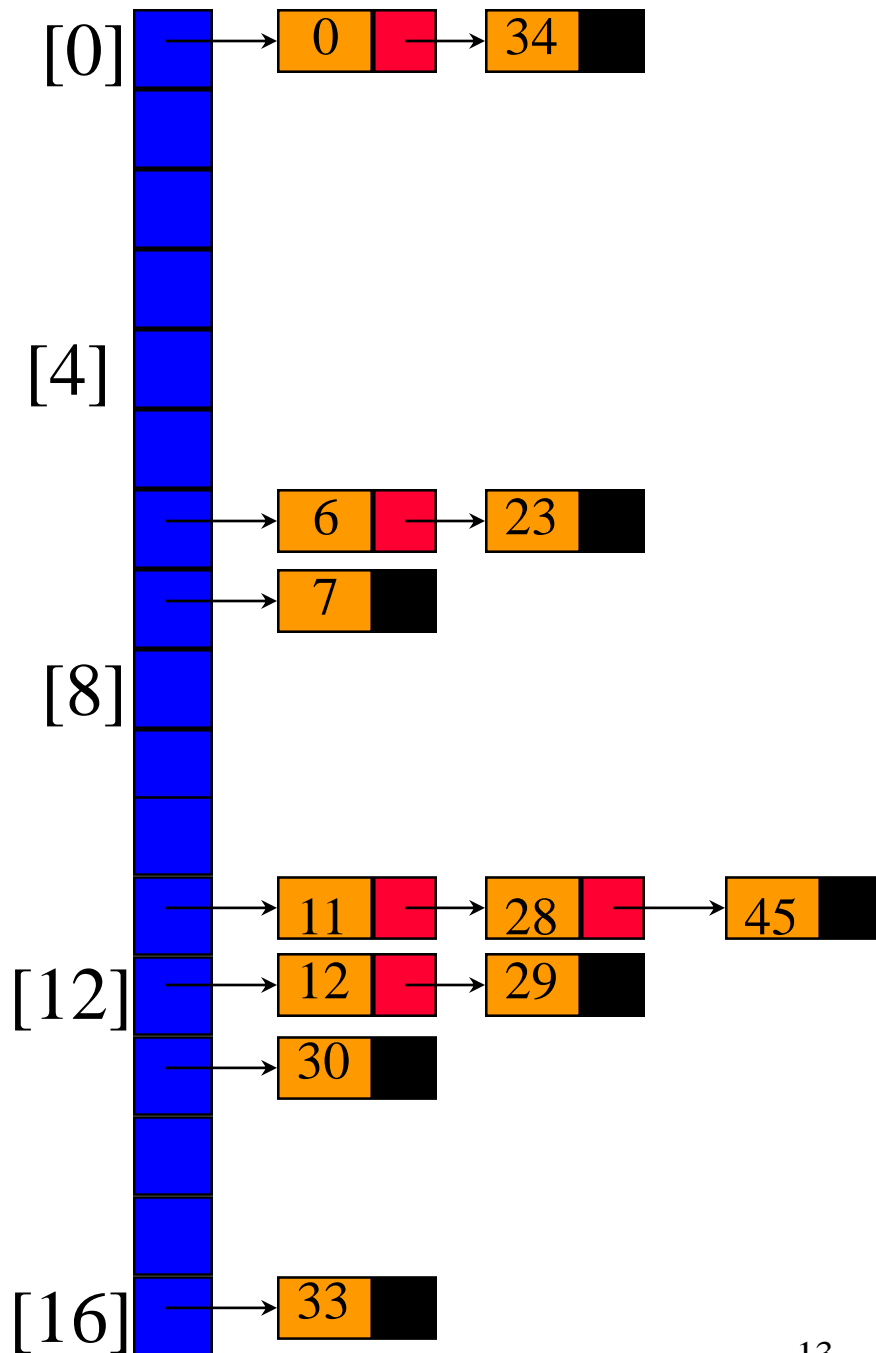
Эрэмбэлэгдсэн

ГИНЖ

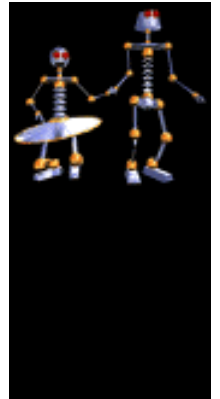
- 6, 12, 34, 29,
28, 11, 23, 7, 0,
33, 30, 45

түлхүүртэй
хосуудыг
хийх

- Багцны үүр =
 $\text{key \% } 17$.



Дундаж үзүүлэлт



- $\alpha \geq 0$ байна.
- Гинжний дундаж урт α .
- $S_n \sim 1 + \alpha / 2$.
- $U_n \leq \alpha$, $\alpha < 1$ бол
- $U_n \sim 1 + \alpha / 2$, $\alpha \geq 1$ бол

java.util.Hashtable



- Эрэмбэлээгүй гинж.
- Анхдагч хуваагч $b = \text{divisor} = 101$
- Анхдагч $\alpha \leq 0.75$
- Ачааллын нягтрал зөвшөөрөгдсөн максимум нягтралаас давбал хэш хүснэгтийн шинэ урт $\text{newB} = 2b + 1$.

Өгөгдлийг шахах



- Өгөгдлийн хэмжээг багасгах.
 - Санах ойг багасгаж, ингснээр санах зардлыг багасгана.
 - Шахалтын зэрэг = анхны хэмжээ/шахагдсан хэмжээ
 - Өгөгдөл дамжуулах, хүлээн авах хугацааг багасгана.



Хаягдалгүй, хаягдалтай шахалт



- $\text{compressedData} = \text{compress}(\text{originalData})$
- $\text{decompressedData} = \text{decompress}(\text{compressedData})$
- $\text{originalData} = \text{decompressedData}$, бол хаягдалгүй шахалт
- $\text{originalData} \neq \text{decompressedData}$, бол хаягдалтай шахалт



Хаягдалгүй, хаягдалтай шахалт



- Хаягдалтай шахагч, хаягдалгүй шахагчаас илүү өндөр шахалтын зэрэгтэй
 - магадгүй **100 : 2**.
- Хаягдалгүй шахалтыг жишээ нь текст файлын хэрэглээнд ашигладаг.
- Хаягдалтай шахалтыг дүрсний хэрэглээнд өргөн хэрэглэдэг.
 - Видео дамжуулахад бага зэргийн хаягдлыг хүний нүд ялгадаггүй.



Текст шахалт



- Хаягдалгүй шахах нь чухал.
- Түгээмэл тархсан **zip** болон Unix-н **compress** шахагч LZW(1984) (Lempel-Ziv-Welch) аргыг ашигладаг.



LZW шахалт



- Эх текстэд орсон тэмдэгтийн цувааг динамик байдлаар тодорхойлогддог кодоор сольдог.
- Кодын хүснэгтийг шахагдсан текстэд кодчилдоггүй. Учир нь буцааж задлахад хэрэг болдог.



LZW шахалт



- Текстийн тэмдэгтэд хязгаарлалт хийе $\{a, b\}$.
 - Амьдралд цагаан толгой ASCII олонлогийн 256 тэмдэгттэй.
- Цагаан толгойн тэмдэгтүүдэд олгох кодыг 0 -ээс эхлэн дугаарлая
- Анхны кодын хүснэгт:

code	0	1
key	a	b



LZW шахалт



code	0	1
key	a	b

- Эх текст = **abababbabaabbabbaabba**
- Эх текстийг зүүнээс баруун тийш гүйлгэх замаар шахья.
- Кодын хүснэгтэд код нь орсон хамгийн урт угтвар **p** –г хайна.
- **p** –г түүний код **pCode** –оор дүрсэлж, дараачийн боломжтой кодыг **pc** -д онооно. Үүнд **c** бол шахах текстийн дараачийн тэмдэгт.



LZW шахалт



code	0	1	2
key	a	b	ab

- Эх текст = abababbabababbaabba
- $p = a$
- $pCode = 0$
- $c = b$
- a -г 0 -ээр дүрслээд ab -г кодын хүснэгтэд нэмнэ.
- Шахсан текст = 0



LZW шахалт



code	0	1	2	3
key	a	b	ab	ba

- Эх текст = **a**bababbabaabbabbaabba
- Шахсан текст = 0
- $p = b$
- $pCode = 1$
- $c = a$
- **b** -г 1 а-ээр дүслээд **ba** —г кодын хүснэгтэд нэмнэ.
- Шахсан текст = 01



LZW шахалт



code	0	1	2	3	4
key	a	b	ab	ba	aba

- Эх текст = **ab**ababbabaabbabbaabba
- Шахсан текст = 01
- $p = ab$
- $pCode = 2$
- $c = a$
- **ab** -г 2 -оор дүслээд **aba** кодын хүснэгтэд нэмнэ.
- Шахсан текст = 012



LZW шахалт



code	0	1	2	3	4	5
key	a	b	ab	ba	aba	abb

- Эх текст = abababbabaabbabbaabba
- Шахсан текст = 012
- $p = ab$
- $pCode = 2$
- $c = b$
- ab -г 2 -оор дүрслээд abb —г кодын хүснэгтэд нэмнэ.
- Шахсан текст = 0122



LZW шахалт



code	0	1	2	3	4	5	6
key	a	b	ab	ba	aba	abb	bab

- Эх текст = ababab**bab**aabbabbaabba
- Шахсан текст = 0122
- $p = ba$
- $pCode = 3$
- $c = b$
- **ba** -г 3 –аар дүрслээд **bab** –г кодын хүснэгтэд нэмнэ.
- Шахсан текст = 01223



LZW шахалт



code	0	1	2	3	4	5	6	7
key	a	b	ab	ba	aba	abb	bab	baa

- Эх текст = abababbab**aabbabbaabba**
- Шахсан текст = 01223
- **p** = ba
- **pCode** = 3
- **c** = a
- **ba** -г 3 -аар дүрслээд **baa** –г кодын хүснэгтэд нэмнэ.
- Шахсан текст = 012233



LZW шахалт



code	0	1	2	3	4	5	6	7	8
key	a	b	ab	ba	aba	abb	bab	baa	abba

- Эх текст = abababbaba**abbabbaabba**
- Шахсан текст = 012233
- **p = abb**
- **pCode = 5**
- **c = a**
- **abb** -г **5** -аар дүрслээд **abba** —г кодын хүснэгтэд нэмнэ.
- Шахсан текст = 0122335



LZW шахалт



code	0	1	2	3	4	5	6	7	8	9
key	a	b	ab	ba	aba	abb	bab	baa	abba	abbaa

- Эх текст = abababbabaabb**abba**abba
- Шахсан текст = 0122335
- **p** = abba
- **pCode** = 8
- **c** = a
- **abba** -г **8** -аар дүрслээд **abbaa** –г кодын хүснэгтэд нэмнэ.
- Шахсан текст = 01223358



LZW шахалт



code	0	1	2	3	4	5	6	7	8	9
key	a	b	ab	ba	aba	abb	bab	baa	abba	abbaa

- Эх текст = abababbababba**abba**
- Шахсан текст = 01223358
- **p = abba**
- **pCode = 8**
- **c = null**
- **abba** -г **8** –аар дүрсэлнэ
- Шахсан текст = 012233588



Кодын хүснэгтийн дүрслэл



code	0	1	2	3	4	5	6	7	8	9
key	a	b	ab	ba	aba	abb	bab	baa	abba	abbaa

- Толь бичиг.
 - Хосууд $(key, element) = (key, code)$.
 - Үйлдлүүд : $get(key)$, $put(key, code)$
- Кодын хязгаар 2^{12} .
- Хэш хүснэгт ашиглах.
 - Хувьсах урттай түлхүүрийг ижил ижил урттай болгох.
 - Түлхүүр бүр pc хэлбэртэй. Үүнд: хэлхээс p бол хүснэгтэд өмнө нь байгаа түлхүүр.
 - pc -г $(pCode)c$ -оор сольно



Кодын хүснэгтийн дүрслэл



code	0	1	2	3	4	5	6	7	8	9
key	a	b	ab	ba	aba	abb	bab	baa	abba	abbaa

code	0	1	2	3	4	5	6	7	8	9
key	a	b	0b	1a	2a	2b	3b	3a	5a	8a

LZW задлалт

code	0	1
key	a	b

- Эх текст = abababbabababbaabba
- Шахсан текст = 012233588
- Кодыг зүүнээс баруун тийш текстэд хөрвүүлнэ
- 0 бол a.
- Задалсан текст = a
- pCode = 0 , p = a.
- p = a -ийн араас дараачийн тэмтэгтийг кодын хүснэгтэд оруулна.

LZW задлалт

code	0	1	2
key	a	b	ab

- Эх текст = **a**bababbababbaabba
- Шахсан текст = **0**12233588
- 1 бол **b**.
- Задалсан текст = **ab**
- **pCode** = 1 , **p** = **b**.
- **lastP** = **a** –ийн араас залгасан **p** –н эхний тэмдэгтийг кодын хүснэгтэд оруулна.

LZW задлалт

code	0	1	2	3
key	a	b	ab	ba

- Эх текст = **ab**ababbabababbaabba
- Шахсан текст = **01**2233588
- 2 бол **ab**.
- Задалман текст = **abab**
- **pCode** = 2 , **p** = **ab**.
- **lastP** = **b** –ийн араас залгасан **p** –н эхний тэмдэгтийг кодын хүснэгтэд оруулна.

LZW задлалт

code	0	1	2	3	4
key	a	b	ab	ba	aba

- Эх текст = **ab**abbabababbaabba
- Шахсан текст = **012233588**
- 2 бол **ab**
- Задалсан текст = **ababab**.
- **pCode** = 2 , **p** = **ab**.
- **lastP** = **ab** –ийн араас залгасан **p** –н эхний тэмдэгтийг кодын хүснэгтэд оруулна.

LZW задлалт

code	0	1	2	3	4	5
key	a	b	ab	ba	aba	abb

- Эх текст = **abababb**aabaabbabbaabba
- Шахсан текст = **012233588**
- 3 бол **ba**
- Задалсан текст = **abababba**.
- **pCode = 3** , **p = ba**.
- **lastP = ab** –ийн араас залгасан **p** –н эхний тэмдэгтийг кодын хүснэгтэд оруулна.

LZW задлалт

code	0	1	2	3	4	5	6
key	a	b	ab	ba	aba	abb	bab

- Эх текст = **abababb**ababbaabba
- Шахсан текст = **012233588**
- 3 бол **ba**
- Задалсан текст = **abababb**aba.
- **pCode = 3** , **p = ba**.
- **lastP = ba** –ийн араас залгасан **p** –н эхний тэмдэгтийг кодын хүснэгтэд оруулна.

LZW задлалт

code	0	1	2	3	4	5	6	7
key	a	b	ab	ba	aba	abb	bab	baa

- Эх текст = **abababbabaabbabbaabba**
- Шахсан текст = **012233588**
- **5** бол **abb**
- Задалсан текст = **abababbabaabb.**
- **pCode = 5** , **p = abb.**
- **lastP = ba** –ийн араас залгасан **p** –н эхний тэмдэгтийг кодын хүснэгтэд оруулна.

LZW задлалт

code	0	1	2	3	4	5	6	7	8
key	a	b	ab	ba	aba	abb	bab	baa	abba

- Эх текст = **abababbababbaabba**
- Шахсан текст = **012233588**
- 8 бол **???**
- Код хүснэгтэд байхгүй бол, түүний түлхүүр нь **lastP** –ий араас залгасан **lastP**-ий эхний тэмдэгт байх болно
- **lastP** = **abb**
- Тэгэхээр 8 бол **abba**.



LZW задлалт

code	0	1	2	3	4	5	6	7	8	9
key	a	b	ab	ba	aba	abb	bab	baa	abba	abbaa

- Эх текст = abababbababbaabba
- Шахсан текст = 012233588
- 8 бол abba
- Задалсан текст = abababbababbaabba.
- pCode = 8 , p = abba.
- lastP = abba –ийн араас орсон p –ийн эхний ТЭМДЭГТ КОДЫН ХҮСНЭГТЭД орно.

Кодын хүснэгтийн дүрслэл

code	0	1	2	3	4	5	6	7	8	9
key	a	b	ab	ba	aba	abb	bab	baa	abba	abbaa

- Толь бичиг.
 - Хосууд $(key, element) = (code, \text{what the code represents}) = (code, codeKey)$.
 - Үйлдлүүд : $get(key)$, $put(key, code)$
- Түлхүүрүүд бүхэл тоо 0, 1, 2, ...
- 1D **codeTable** массивыг ашиглая
 - $codeTable[code] = codeKey$.
 - Кодын түлхүүр бүр **pc** хэлбэртэй. Үүнд хэлхээс **p** бол хүснэгтэд өмнө нь орсон кодын түлхүүр.
 - **pc** -г **(pCode)c** -аар сольно

Хугацааны үзүүлэлт



- Шахалт.
 - $O(n)$ хүлээх хугацаа, үүнд n шахагдсан текстийн урт.
- Задлалт.
 - $O(n)$ хугацаа, үүнд n задалсан текстийн урт.