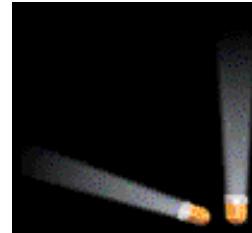


Толь бичиг



- Хосуудын цуглуулга.
 - (key, element)
 - Хос бүр өөр түлхүүртэй.
- Үйлдлүүд.
 - `get(theKey)`
 - `put(theKey, theElement)`
 - `remove(theKey)`

Хэрэглээ

- CS204 –г сонгосон оюутнууд.
 - (key, element) = (оюутны нэр, бие даалт болон шалгалтын дүнгийн шугаман жагсаалт)
 - Бүх түлхүүр ялгаатай.
- Ө.Дөлгөөн гэсэн түлхүүртэй элементийн авах
- Д.Туяа гэсэн түлхүүртэй элементийг өөрчлөх
 - put().
 - remove().

Давхцалттай толь бичиг

- Түлхүүр давхцаж болно.
- Үгсийн толь бичиг.
 - Хос нь (үг, утга).
 - Нэг үг хэд хэдэн утгатай байж болно.
 - (гар, хүний эрхтэн)
 - (гар, гадагшлах хөдөлгөөн)
 - (гар, компьютерийн оруулах төхөөрмж)
 - ГЭХ МЭТ.

Шугаман жагсаалтаар дүрслэх

- $L = (e_0, e_1, e_2, e_3, \dots, e_{n-1})$
- e_i бүхэн (key, element).
- 5-хостой толь бичиг $D = (a, b, c, d, e)$.
 - $a = (aKey, aElement), b = (bKey, bElement),$
Г.М.
- Массив эсхүл Холбоост дүрслэл.

Массив дүрслэл

| | | | | | | | | | | | | | | |
|---|---|---|---|---|--|--|--|--|--|--|--|--|--|--|
| a | b | c | d | e | | | | | | | | | | |
|---|---|---|---|---|--|--|--|--|--|--|--|--|--|--|

- `get(theKey)`
- `put(theKey, theElement)`
- `remove(theKey)`

Эрэмбэлэгдсэн массив

| | | | | | | | | | | | | | | |
|---|---|---|---|---|--|--|--|--|--|--|--|--|--|--|
| A | B | C | D | E | | | | | | | | | | |
|---|---|---|---|---|--|--|--|--|--|--|--|--|--|--|

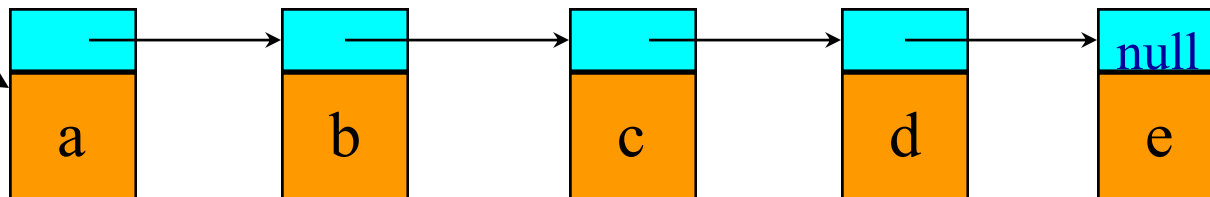
- элементүүд түлхүүрийн өсөх дарааллаар байрлана.
- `get(theKey)`
 - $O(\log \text{ size})$
- `put(theKey, theElement)`
 - $O(\log \text{ size})$ давхцлыг шалгахад, $O(\text{size})$ нэмэхэд.
- `remove(theKey)`
 - $O(\text{size})$.



Эрэмбэлэгдээгүй гинж

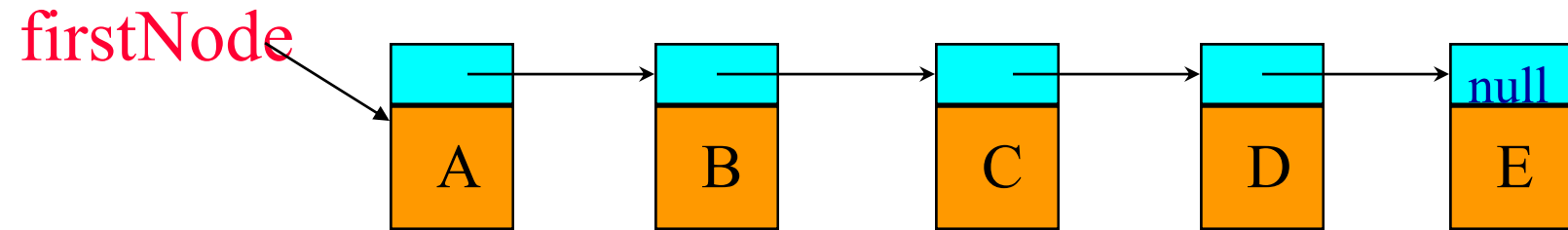


firstNode



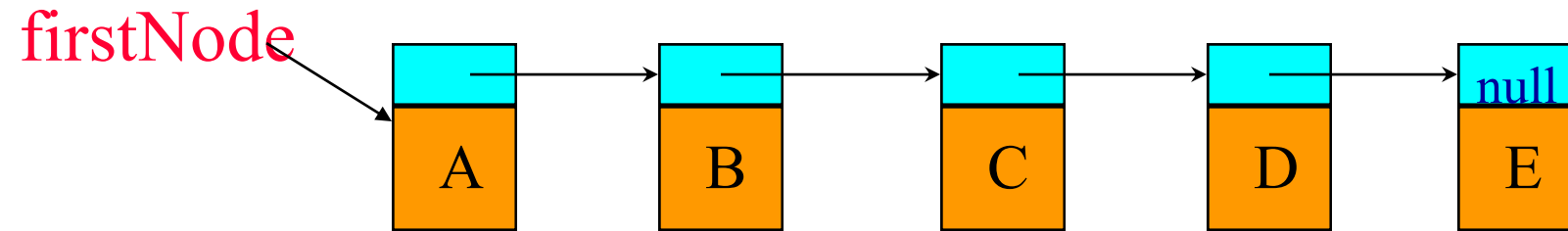
- `get(theKey)`
 - $O(\text{size})$
- `put(theKey, theElement)`
 - $O(\text{size})$ давхцлыг шалгахад, $O(1)$ зүүн талд нь НЭМЭХЭД.
- `remove(theKey)`
 - $O(\text{size})$.

Эрэмбэлэгдсэн гинж



- Элементүүд түлхүүрийн өсөх дараалалтай.
- `get(theKey)`
 - $O(\text{size})$
- `put(theKey, theElement)`
 - $O(\text{size})$ давхцлыг шалгахад, $O(1)$ зөв газар нь НЭМЭХЭД.

Эрэмбэлэгдсэн гинж



- Элементүүд түлхүүрийн өсөх дараалалтай.
- `remove(theKey)`
 - $O(\text{size})$.

Алгасах жагсаалт

- Муу тохиолдолд **get, put, remove** - **$O(\text{size})$** .
- Хүссэн хугацаа - **$O(\log \text{ size})$** .
- **Skip lists** – алгасах жагсаалтыг бид үзэхгүй

Хэш хүснэгт

- Муу тохиолдолд **get, put, remove** - **$O(\text{size})$** .
- Хүссэн хугацаа **$O(1)$** .

Хүслийн Хэш

- 1D массив(хүснэгт) ашиглая $table[0:b-1]$.
 - Массивын байршил бүр багц.
 - Ер нь багц толь бичгийн зөвхөн нэг хосыг хадгалах ёстой.
- Ашиглах хэш функц f түлхүүр u k хүснэгтийн индекст хувиргах ёстой $[0, b-1]$.
 - $f(k)$ бол түлхүүр k -ийн багцны үүр
- Толъ бичгийн $(key, element)$ хос бүр $table[f[key]]$ гэсэн багцны үүрэнд хадгалагдана

Хүслийн хэшийн жишээ

- Хосууд: (22,a), (33,c), (3,d), (73,e), (85,f).
- Хэш хүснэгт `table[0:7]`, `b = 8`.
- Хэш функц `key/11`.
- Хосууд хүснэгтэд дараах байдлаар хадгалагдана:

| | | | | | | | |
|-------|-----|--------|--------|-----|-----|--------|--------|
| (3,d) | | (22,a) | (33,c) | | | (73,e) | (85,f) |
| [0] | [1] | [2] | [3] | [4] | [5] | [6] | [7] |

- `get`, `put`, `remove` - ажиллах хугацаа $O(1)$.

Яавал буруу тийш явах вэ?

| | | | | | | | |
|-------|-----|--------|--------|-----|-----|--------|--------|
| (3,d) | | (22,a) | (33,c) | | | (73,e) | (85,f) |
| [0] | [1] | [2] | [3] | [4] | [5] | [6] | [7] |

- (26,g) хаашаа явах вэ?
- Нэг багцны үүртэй түлхүүрүүдийг **СИНОНИМ** гэнэ
 - 22 ба 26 бол ашиглаж байгаа хэш функцийн хувьд синонимууд.
- (26,g) –д харгалзах багцны үүр эзлэгдсэн байна.

Яавал буруу тийш явах вэ?

| | | | | | | | |
|-------|--|--------|--------|--|--|--------|--------|
| (3,d) | | (22,a) | (33,c) | | | (73,e) | (85,f) |
|-------|--|--------|--------|--|--|--------|--------|

- Өөр түлхүүртэй шинэ хосод харгалзах багцны үүр эзлэгдсэнээс **collision-зөрчил** үүснэ.
- Шинэ хосод харгалзах багцны үүрэнд зай байхгүйгээс **overflow-халилт** үүснэ.
- Багцны үүр зөвхөн нэг хосыг хадгалдаг бол зөрчил болон халилт зэрэг үүснэ.
- Халилтыг шийдэх арга хэрэгтэй.

Хэш хүснэгтийн асуудлууд

- Хэш функцийг сонгох.
- Халилтыг зохицуулах арга.
- Хэш хүснэгтийн хэмжээ (Багцны тоо).

Хэш функцүүд

- Хоёр хэсэг:
 - Түлхүүр бүхэл биш бол бүхэл болгох.
 - Шийдэх арга нь `hashCode()`.
 - Бүхлийг багцны үүрийн тусгал болгох.
 - $f(k)$ функц $[0, b-1]$ мужид бүхэл утгатай, үүнд b болхүснэгт дэх багцны тоо.

Хэлхээсийг бүхэлд

- Java –ийн тэмдэгт бүр 2 байт урттай.
- `int` бол 4 байт.
- 2 тэмдэгтийн `s` хэлхээсийг давтагдашгүй 4 байт `int` –д хувиргахдаа:

```
int answer = s.charAt(0);
```

```
answer = (answer << 16) + s.charAt(1);
```

- 2 тэмдэгтээс урт хэлхээст давтагдашгүй `int` дүрслэл байхгүй.

Хэлхээсийг сөрөг биш бүхэлд

```
public static int integer(String s)
{
    int length = s.length();
    // s –ийн тэмдэгтийн тоо
    int answer = 0;
    if (length % 2 == 1)
    { // урт нь сондгой бол
        answer = s.charAt(length - 1);
        length--;
    }
}
```

Хэлхээсийг сөрөг бус бүхэлд

// урт нь тэгш бол

```
for (int i = 0; i < length; i += 2)
```

```
{// нэг удаа 2 тэмдэгтийг
```

```
    answer += s.charAt(i);
```

```
    answer += ((int) s.charAt(i + 1)) << 16;
```

```
}
```

```
return (answer < 0) ? -answer : answer;
```

```
}
```

Багцны үүрийн тусгал

| | | | | | | | |
|-------|-----|--------|--------|-----|-----|--------|--------|
| (3,d) | | (22,a) | (33,c) | | | (73,e) | (85,f) |
| [0] | [1] | [2] | [3] | [4] | [5] | [6] | [7] |

- Хамгийн нийтлэг арга бол хуваалт(**divisor**).

homeBucket =

Math.abs(theKey.hashCode()) % divisor;

- divisor** багцны тоо **b** -тэй тэнцүү
- $0 \leq \text{homeBucket} < \text{divisor} = b$**

Жигд Хэш функц

| | | | | | | | |
|-------|-----|--------|--------|-----|-----|--------|--------|
| (3,d) | | (22,a) | (33,c) | | | (73,e) | (85,f) |
| [0] | [1] | [2] | [3] | [4] | [5] | [6] | [7] |

- **keySpace** бол байж болох түлхүүрийн олонлог болог.
- Хэш функц **keySpace** олонлогийн түлхүүрийг багцад тусгахдаа дунджаар ижил тооны түлхүүр нэг багцад тусч байвал функцийг **жигд хэш функц** гэдэг

Жигд Хэш функц

| | | | | | | | |
|-------|-----|--------|--------|-----|-----|--------|--------|
| (3,d) | | (22,a) | (33,c) | | | (73,e) | (85,f) |
| [0] | [1] | [2] | [3] | [4] | [5] | [6] | [7] |

- Өөрөөр хэлбэл, санамсаргүй сонгосон түлхүүр багц i –д тусах магадлал $1/b$, $0 \leq i < b$.
- Жигд хэш функц түлхүүр санамсаргүй сонгогдох тохиолдолд халилтыг минимум болгодог.

Хуваалтын хэш

- $\text{keySpace} =$ бүх int -үүд
- Ямар ч b -ийн хувьд дунджаар $2^{32}/b$ тооны int -үүд багц i –д тусдаг (хуваагддаг).
- Иймд, $\text{keySpace} =$ бүх int -үүд бол хуваалтын арга нэгэн жигдийн хэш функц болно
- Амьдралд, түлхүүрүүд хамааралтай байдаг.
- Тэгэхээр, хуваагч b –ийн сонголт багцын үүрийн тусгалд нөлөөлдөг.

Хуваагчийг сонгох

- Түлхүүрүүд ер нь хамааралтай байдгаас тэгш, сондгой багцны үүрт тусгал давамгайлах тал байдаг.
- Хуваагч тэгш тоо бол, сондгой бүхэл сондгой багцад, тэгш бүхэл тэгш багцад тусдаг.
 - $20\%14 = 6, 30\%14 = 2, 8\%14 = 8$
 - $15\%14 = 1, 3\%14 = 3, 23\%14 = 9$

Хуваагчийг сонгох

- Хуваагч сондгой бол, сондгой (тэгш) бүхэл дурын үүрд хуваагдаж болдог.
 - $20\%15 = 5$, $30\%15 = 0$, $8\%15 = 8$
 - $15\%15 = 0$, $3\%15 = 3$, $23\%15 = 8$
- Багцны үүрт нэгэн жигд тараах илүү боломжтой.
- Иймд тэгш хуваагчийг битгий ашигла.

Хуваагчийг сонгох

- Амьдралд жигд биш тархалт хуваагчийг 3, 5, 7, ... мэтийн анхны тооны үржвэр байдлаар сонгосноос болдог
- Хэрвээ p нь b -ийн хуваагч бол p өсөх тутам энэ нөлөөлөл багасдаг.
- Зөв сонолтоор b анхны тоо байх нь чухал.
- Эсхүл, b сонгохдоо 20 -оос доош тоонд хуваагддаггүй байх явдал

Java.util.HashMap



- Сондгой тоог хуваагч болгон ашигладаг.
- Ингэснээр олон хосыг толь бичигт оруулах зорилгоор хэш хүснэгтийн хэмжээг өөрчлөх боломж олгодог.
 - Жишээ нь, массивыг 2 дахин ихэсгэхэд, b (сондгой) урттай 1D массив $table$ –ийн уртыг $2b+1$ болгоно(мөн сондгой).



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 | 31 | | |

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 |

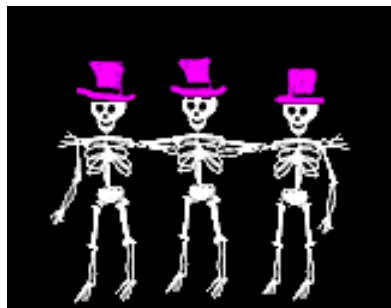
- Чөлөөлөгдсөн багцыг ашиглаж болох хосыг хайх.

| | | | | | | | | | | | | | | | | |
|----|---|----|--|--|---|---|----|---|----|--|----|----|----|--|----|----|
| 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 |

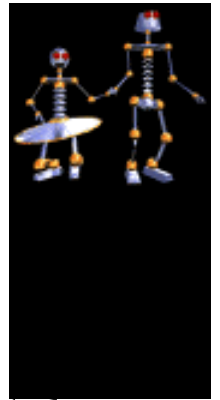
Шугаман тандалтын үзүүлэлт



| | | | | | | | | | | | | | | | | |
|----|---|----|--|--|--|---|----|---|--|--|----|----|----|----|----|----|
| 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.

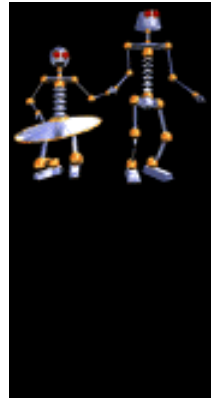
Хүссэн үзүүлэлт



| | | | | | | | | | | | | | | | | |
|----|---|----|--|--|--|---|----|---|--|--|----|----|----|----|----|----|
| 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 (хүссэн) тоо (n - том тоо бол)
- U_n = амжилтгүй хайлтаар шалгах багцын (хүссэн) тоо (n - том тоо бол)
- тэгвэл put , remove үйлдлүүдийн хугацаа U_n -аар тодорхойлогдоно

Хүссэн 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
зөвлөж байна.

Хэш хүснэгтийн зохиомж

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

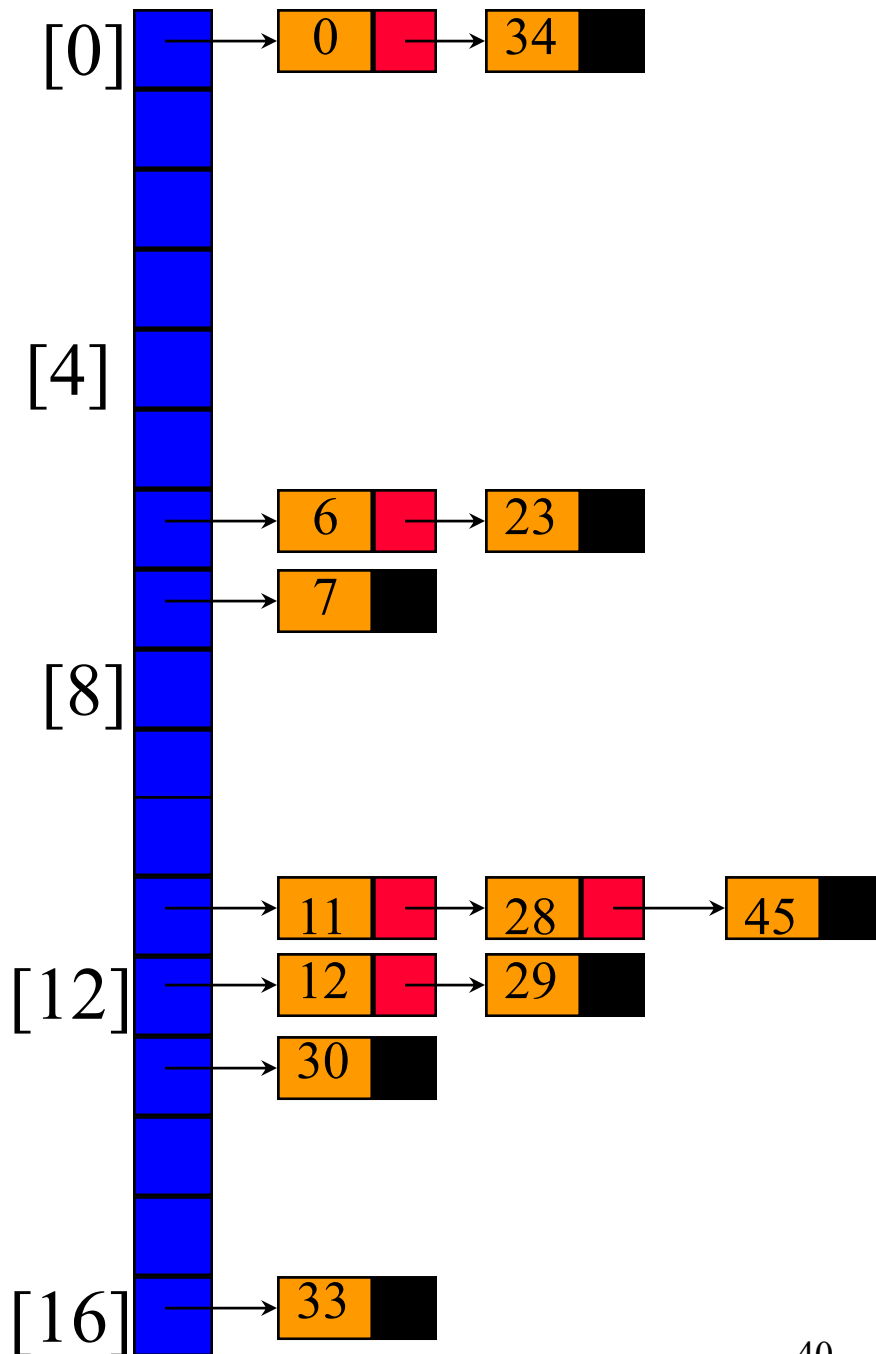
Эрэмбэлэгдсэн

ГИНЖ

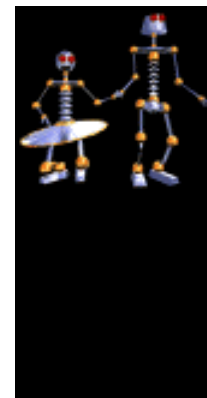
- 6, 12, 34, 29,
28, 11, 23, 7, 0,
33, 30, 45

түлхүүртэй
хосуудыг
хийх

- Багцны үүр =
 $\text{key \% } 17$.



Хүссэн үзүүлэлт



- $\alpha \geq 0$ байна.
- Гинжний хүссэн 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 = **abababbabababbaabba**
- 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 = **abababbabaabbabbaabba**
- $p = a$
- $pCode = 0$
- $c = b$
- **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**babaabbabbaabba**
- 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 = abababbabababba**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**ababbababbaabba
- 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**babaabbabbaabba
- 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 = **abababbabababbaabba**
- 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 = abababbabababbaabba
- Шахсан текст = 012233588
- 8 бол abba
- Задалсан текст = abababbabababbaabba.
- 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.