TÖL403G GREINING REIKNIRITA

9. Kvik bestun 3

Hjálmtýr Hafsteinsson
Vor 2022

Í þessum fyrirlestri

- Breytingafjarlægð (*edit distance*)
  - Endurkvæm framsetning
  - Rakningarformúla
  - Kvik bestun

3.7 − 3.8

- Hlutmengissumma (*subset sum*)
  - Rakningarformúla
  - Kvik bestun

Breytingarfjarlægð (*edit distance*)

- Viljum vita hversu líkir tveir strengir *A* og *B* eru
- Notum breytingarfjarlægð (*Levenshtein dist.*, *Ulam dist.*)
  - Hversu margar breytingar þarf að gera á *A* til að fá *B*
  - Minnsti fjöldi breytingar gefur fjarlægð á milli *A* og *B*

3 gerðir breytinga:

Innsetning (á staf í *B*)

Eyðing (á staf í *A*)

Breyting (á staf í *A* yfir í staf í *B*)

Dæmi

Fjarlægðin á milli orðanna FOOD og MONEY er 4:

FOOD → MOOD → MON D → MONED → MONEY

Hér eru þrjár stafabreytingar (F → M, O → N, D → Y)
og ein innsetning (á E)

Auðveldari leið til að sjá breytingarnar:

F	O	O		D
M	O	N	E	Y
Breyting		Breyting	Innsetning	Breyting

Autt bil í efra orði: Innsetning
Autt bil í neðra orði: Eyðing

Ef stafir í sama dálki eru ólíkir: Breyting á staf

Stærra dæmi

óeigingjarn

A = ALGORITHM, *B* = ALTRUISTIC

Ekki augljóst hversu mikil fjarlægðin er á milli *A* og *B*

Hér er ein leið til að breyta *A* í *B:*

```
A L G O R   I   T H M
A L   T R U I S T I C
```

Eyðing
Breyting
Innsetning
Innsetning
Breyting
Breyting

1 eyðing
2 innsetningar
3 stafabreytingar

Samtals 6 breytingar

svo fjarlægðin er a.m.k. 6

Er hægt að gera betur?

Endurkvæm framsetning

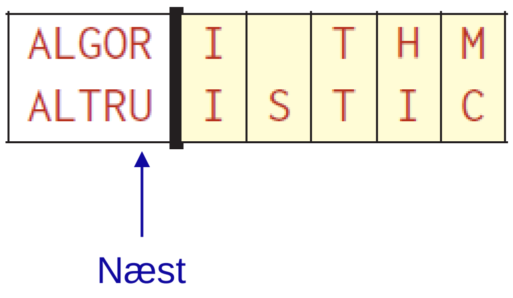
- Höfum strengina $A[1..m]$ og $B[1..n]$
- Byrjum aftast í strengjunum og vinnum okkur fram eftir þeim

- Skilgreinum:

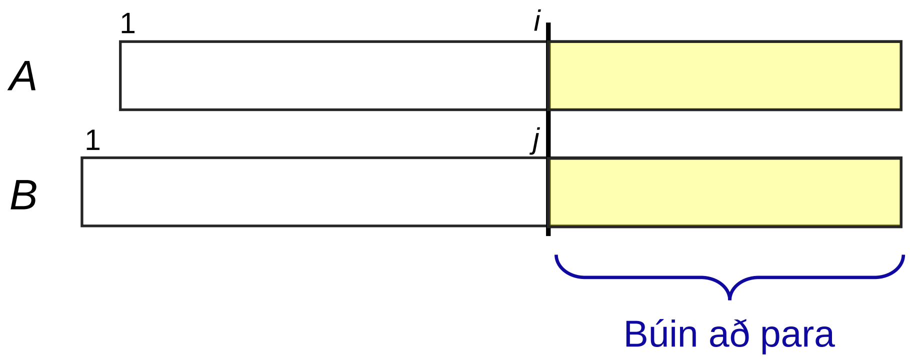
  Edit(i, j): Breytingafjarlægð á milli $A[1..i]$ og $B[1..j]$

  Við viljum finna *Edit*(m, n)

Til dæmis:



Almennt:



Rekjum okkur fram strengina tvo

Rakningarvensl (*recurrence*)

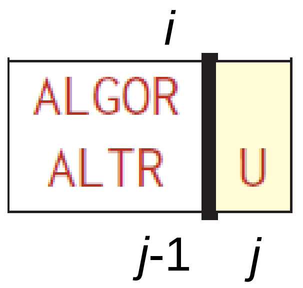
- Þrír möguleikar á því hvað getur gerst í síðasta dálkinum

- **<u>Innsetning</u>**
  - Síðasta stakið í efri línunni er tómt (stafur fer í *B*, sem var ekki í *A*)
  - Þá er $Edit(i, j) \;=\; Edit(i, j-1) + 1$

Fækkum stöfum í *B*

Kostn. við innsetningu

Dæmi:



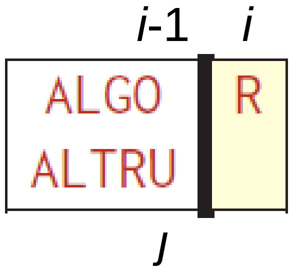
Rakningarvensl (*recurrence*)

- **<u>Eyðing</u>**
  - Þá er eyða í síðasta sætinu í neðri línunni (stafur í *A*, sem fer ekki í *B*)
  - Þá er $Edit(i, j) \;=\; Edit(i-1, j) + 1$

Fækkum stöfum í *A*

Kostn. við eyðingu

Dæmi:

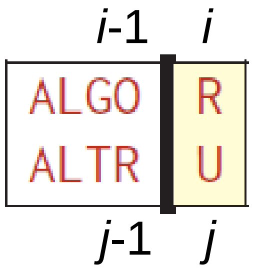


- **<u>Stafabreyting</u>**
  - Þá er stafur í aftasta dálki í báðum línum (stafur í *A* breytist í annan staf í *B*)
  - Þá er $Edit(i, j) \;=\; Edit(i-1, j-1) + 1$

Fækkar í báðum

Kostn. við breytingu

Dæmi:



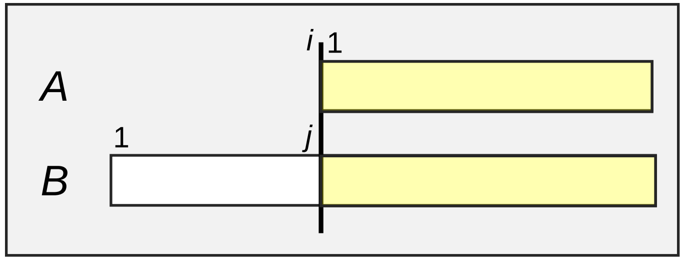
Rakningarvenslin

- Jaðartilvik:

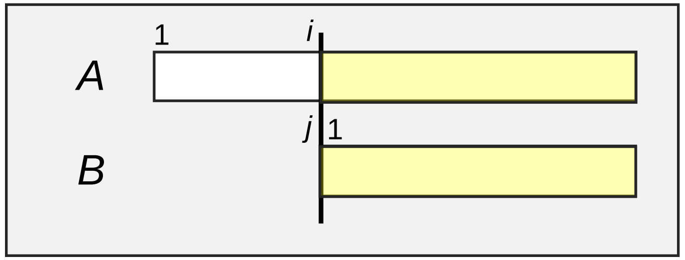
  Ef annar strengurinn er orðinn tómur (i=0 eða j=0), þá þarf að

  <u>setja inn *j* stafi</u> (ef i=0), svo að $Edit(0, j) = j$

  <u>eyða út *i* stöfum</u> (ef j=0), svo að $Edit(i, 0) = i$



- Rakningarvensl:



$$Edit(i, j) = \begin{cases} i & \text{ef } j = 0 \\ j & \text{ef } i = 0 \\ \min \left\{ \begin{array}{c} Edit(i, j-1) + 1 \\ Edit(i-1, j) + 1 \\ Edit(i-1, j-1) + [A[i] \neq B[j]] \end{array} \right\} & \text{annars} \end{cases}$$

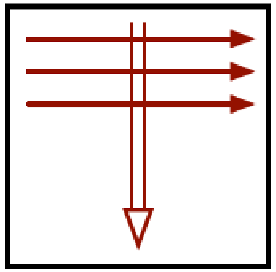
innsetn.

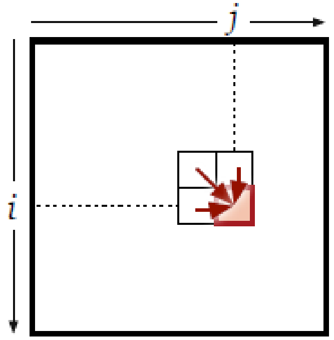
eyðing

1 ef satt,
0 annars

Kvik bestun

- Breytum yfir í kvika bestun með venjulegu aðferð:
  - **Undirverkefni:**
    - Hvert undirverkefni er auðkennt með vísunum $0 \leq i \leq m$ og $0 \leq j \leq n$
  - **Gagnagrind:**
    - Til að muna öll möguleg gildi á *Edit*(i, j) notum við tvívítt fylki $E[0..m, 0..n]$
  - **Tengsl:**
    - Til að reikna $E[i, j]$ þurfum við aðeins gildin $E[i-1, j]$, $E[i, j-1]$ og $E[i-1, j-1]$
  - **Útreikningsröð:**
    - Getum reiknað fylkið línu fyrir línu, frá efstu línu niður





Reikniritið *EditDistance*

```
EDITDISTANCE(A[1..m], B[1..n]):
    for j ← 0 to n
        Edit[0, j] ← j

    for i ← 1 to m
        Edit[i, 0] ← i
        for j ← 1 to n
            ins ← Edit[i, j − 1] + 1
            del ← Edit[i − 1, j] + 1
            if A[i] = B[j]
                rep ← Edit[i − 1, j − 1]
            else
                rep ← Edit[i − 1, j − 1] + 1
            Edit[i, j] ← min {ins, del, rep}

    return Edit[m, n]
```

Fylla í efstu línu

Fylla í fyrsta dálk

Innsetning

Eyðing

Ef sami stafur þá enginn kostnaður

Stafabreyting

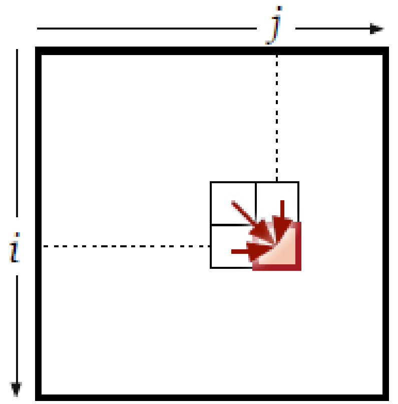
Velja það sem gefur minnsta kostnað

Höfum að $E[0, j] = j$ og $E[i, 0] = i$

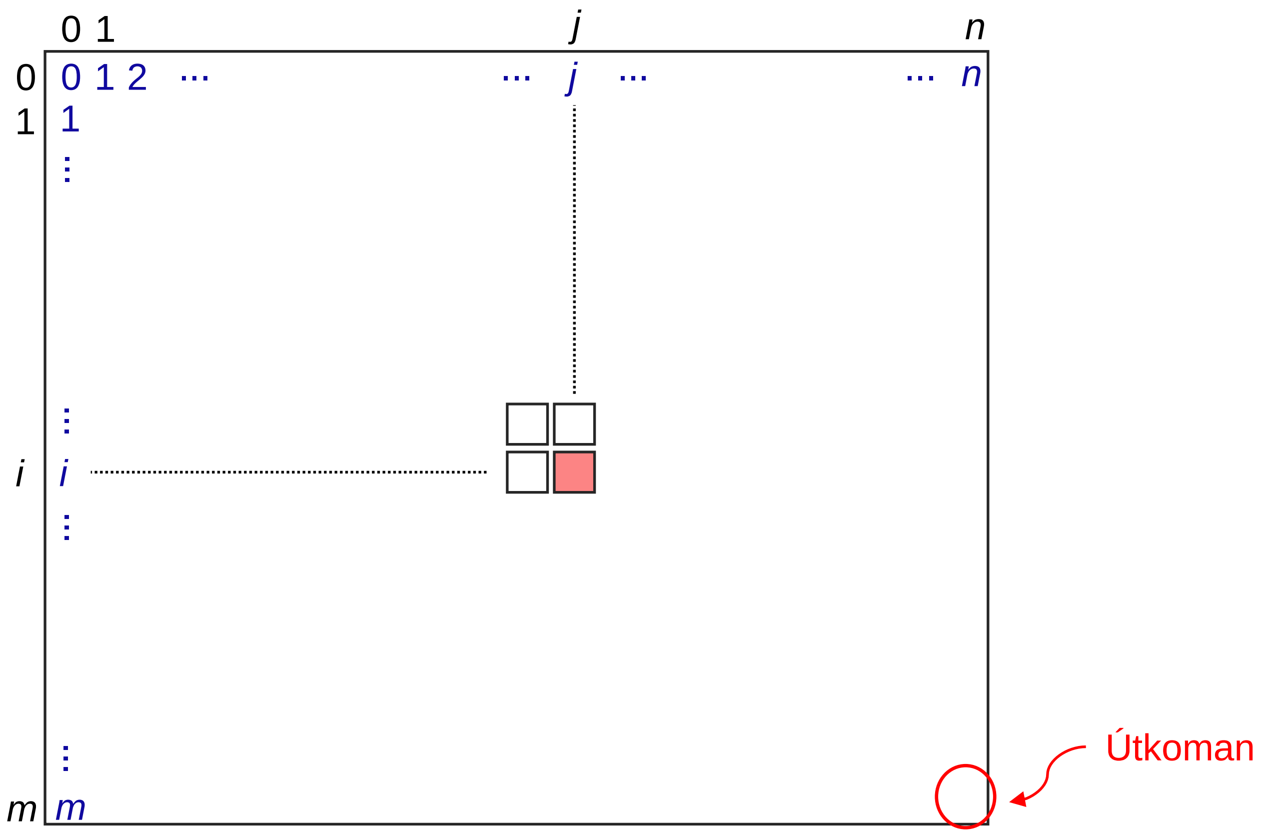
setja inn j stafi í B

eyða út i stöfum úr A

$$E[i, j] = \mathbf{min}\{\ E[i-1, j] + 1,\ E[i, j-1] + 1,\ E[i-1, j-1] + 0/1\}$$



Dæmigerður útreikningur



Byrjum á að fylla í línu 0 og dálk 0

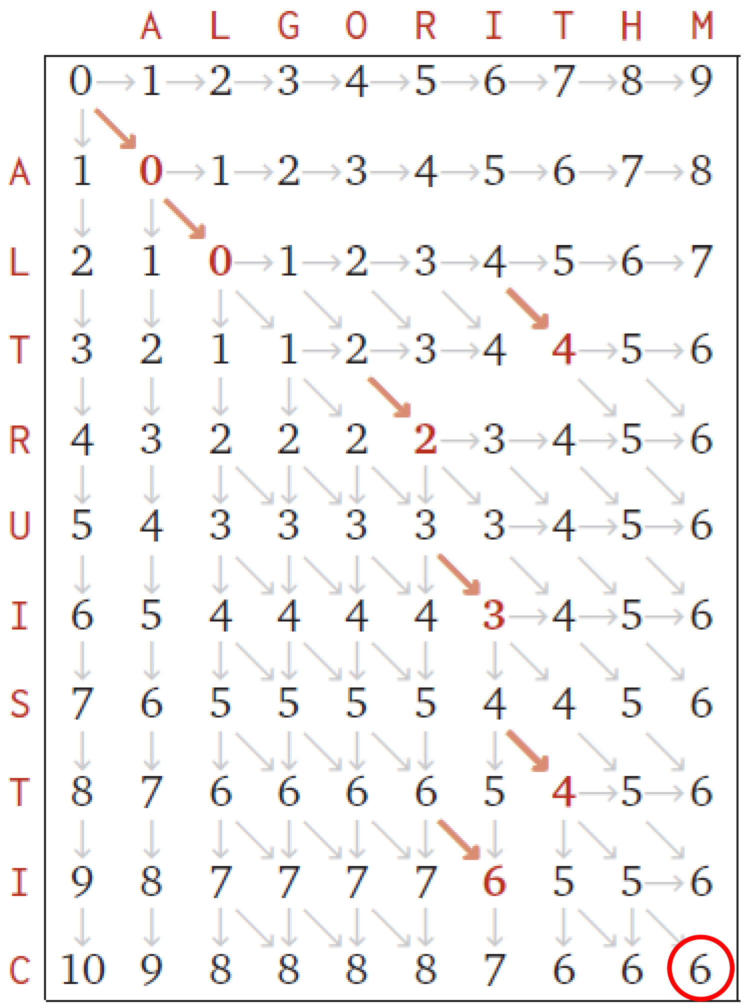
Reikna svo $E[i, j]$ út frá:

$E[i-1, j]$, $E[i, j-1]$ og $E[i-1, j-1]$

Niðurstaðan kemur í $E[m, n]$

Getum gert útreikninginn niður línurnar eða út dálkana

Stórt sýnidæmi



Ör sýnir hólfið sem gaf gildið í næsta hólfi
(stundum eru tvær örvar, því gildin voru þau sömu)

↓ : innsetning

→ : eyðing

↘ : stafabreyting

Rauð hornalínuör þýðir sami stafur

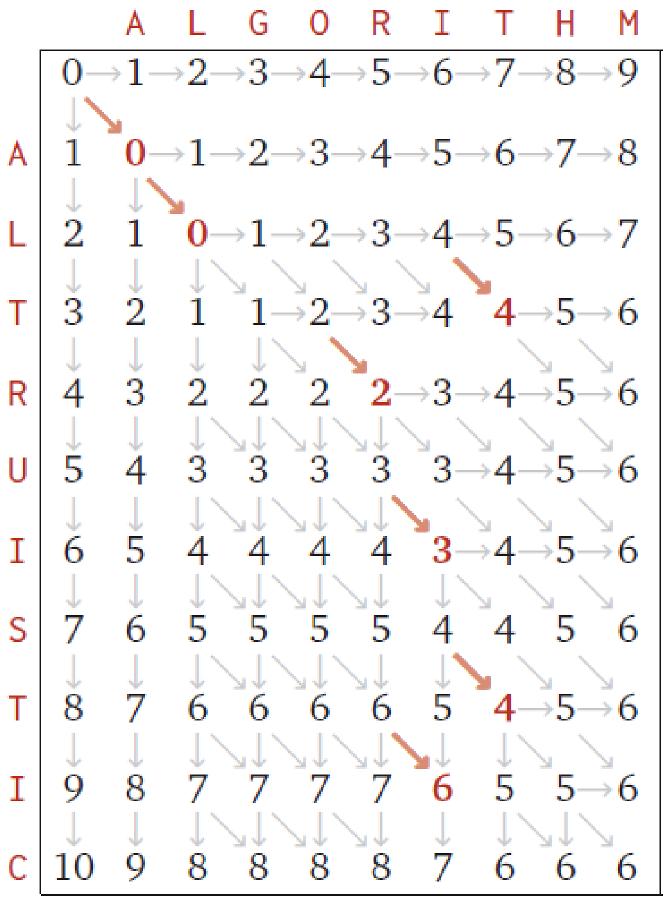
Sérhver vegur eftir örvum frá efra vinstra horni til neðra hægra horns samsvarar bestu leið til að breyta *A* yfir í *B*

Sjáum að fjarlægðin er 6 (hólfið neðst til hægri)

Þurfum að rekja okkur "aftur á bak" frá hólfinu neðst til hægri

Æfingar

- Finnið leiðina í gegnum *Edit*-töfluna fyrir
  - ```
    A L G O R I   T H M
    A L T R U I S T I C
    ```
  - ```
    A L G O R   I   T H M
    A L   T R U I S T I C
    ```



Hlutmengissumma (*subset sum*)

- Sáum í kafla 2.3 (og fyrirlestri 5) verkefnið:

> Gefið mengi $X[1..n]$ af jákvæðum heiltölum og heiltalan T, er til hlutmengi í X með summuna T?

- Skilgreinum nú fallið:

skilar **satt/ósatt**

$$SS(i, t) = \text{til hlutmengi í } X[i..n] \text{ með summu } t$$

- Endurkvæm framsetning:

þá nægir tóma mengið

þetta er ómögulegt

$$SS(i,t) = \begin{cases} \textbf{satt} & \text{ef } t = 0 \\ \textbf{ósatt} & \text{ef } t < 0 \text{ eða } i > n \\ SS(i+1, t) \vee SS(i+1, t-X[i]) & \text{annars} \end{cases}$$

X[i] **ekki** í summunni

X[i] **er** í summunni

Umbreyting í kvika bestun

- **Undirverkefni:**
  - Skilgreint með vísinum i, $1 \leq i \leq n+1$, og heiltölunni $t \leq T$.
- **Gagnagrind:**
  - Notum tvívíða fylkið $S[1..n, 0..T]$, þar sem $S[i, t]$ geymir gildið á $SS(i, t)$
- **Útreikningsröð:**
  - Sérhvert stak $S[i, t]$ er aðeins háð tveimur öðrum stökum, sem eru bæði af gerðinni $S[i+1, \cdot\,]$. Það er því hægt að reikna fylkið S línu fyrir línu uppávið, því lína i byggir aðeins á línu $i+1$.
- **Tími og minni:**
  - Gagnagrindin nota $O(nT)$ minnispláss. Ef $S[i+1, t]$ og $S[i+1, t-X[i]]$ eru bæði þekkt þá er hægt að reikna $S[i, t]$ á $O(1)$ tíma. Heildartími reikniritsins er þá $O(nT)$.

Reiknirit fyrir Hlutmengissummu

$\underline{\textsc{FastSubsetSum}(X[1..n], T):}$
$\quad S[n+1, 0] \leftarrow \textsc{True}$
$\quad$ for $t \leftarrow 1$ to T
$\quad\quad S[n+1, t] \leftarrow \textsc{False}$

$\quad$ for $i \leftarrow n$ downto 1
$\quad\quad S[i, 0] = \textsc{True}$
$\quad\quad$ for $t \leftarrow 1$ to $X[i] - 1$
$\quad\quad\quad S[i, t] \leftarrow S[i+1, t]$ ⟨⟨Avoid the case $t < 0$⟩⟩
$\quad\quad$ for $t \leftarrow X[i]$ to T
$\quad\quad\quad S[i, t] \leftarrow S[i+1, t] \vee S[i+1, t - X[i]]$

$\quad$ return $S[1, T]$

<u>Upphafsstillum fyrst línu *n*+1:</u>

Ef T = 0 þá gengur tóma mengið

Ef T > 0 þá er þetta ómögulegt

<u>Rekjum okkur svo upp línurnar:</u>

Ef T = 0 þá gengur tóma mengið

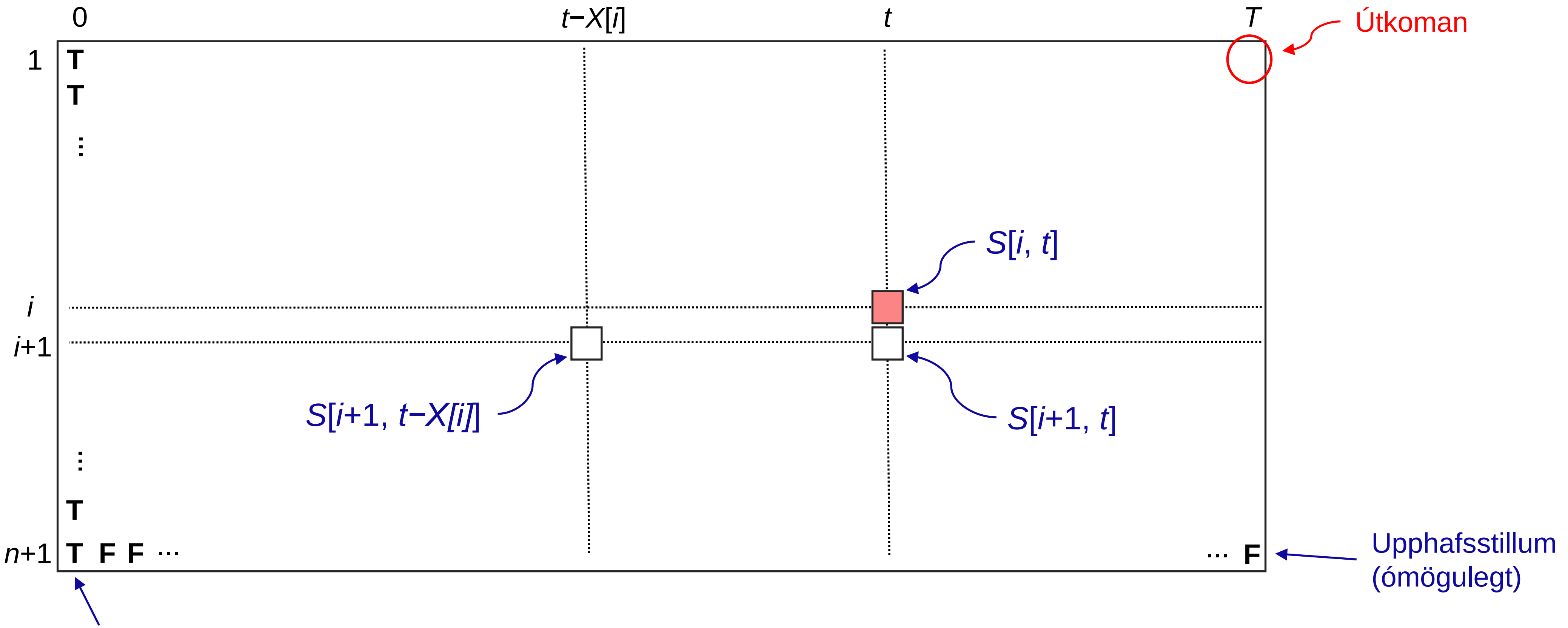
Ef t < $X[i]$ þá getur X[i] ekki verið hluti af summunni

Annars flettum við upp á báðum möguleikum
og EÐ-um þá saman

Svar við:

"Er til hlutmengi í X[1..n] með summu T?"

Gagnagrindin



Upphafsstillum
(tóma mengið dugar)

Reiknum svo:

$$S[i, t] = S[i+1, t] \vee S[i+1, t-X[i]]$$

X[i] **ekki** í summunni

X[i] **er** í summunni

Nánar um tímaflækju

- Endurrakningareiknirit hafði tímann $O(2^n)$
- Þetta reiknirit hefur tímann $O(nT)$

Athugið að T er gildið á tölunni

T er táknað með inntaki af stærð $\log(T)$

Þetta reiknirit er því veldistíma í stærð inntaksins!

- Svona reiknrit eru oft sögð vera gervimargliðutíma (*pseudo polynomial*)
  - Ef T er lítil tala þá er $O(nT)$ betra en $O(2^n)$
  - En T getur líka verið stærra en 2^n og þá er þetta reiknirit verra!

Verkefnið Hlutmengissumma er *NP-complete*, svo ekki er skrýtið að tímaflækjan sé slæm!

Fyrirlestraæfingar

1. Hver er minnsta breytingafjarlægð (*edit distance*) á milli orðanna 'tölva' og 'kölska'?

2. Búið til *Edit*-fylkið fyrir strengina 'abra' og 'bara'.

3. Gefið safnið $S = \{3, 1, 1, 2, 2, 1\}$, er hægt að skipta því upp í tvennt þannig að summa stakana í báðum hópunum sé sú sama?