

# ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ

## Φροντιστήριο 2



ΜΕΡΟΣ Α  
ΤΑΞΙΝΟΜΗΣΗ ΣΤΗ ΔΕΥΤΕΡΕΥΟΥΣΑ  
ΜΝΗΜΗ

# ΕΙΣΑΓΩΓΗ

- ▶ Όταν η χωρητικότητα της κύριας μνήμης δεν είναι ικανή να χωρέσει όλα τα δεδομένα που πρόκειται να ταξινομηθούν αναγκαζόμαστε να χρησιμοποιήσουμε τη δευτερεύουσα μνήμη (δίσκο, μαγνητικές ταινίες).

# Αλγόριθμοι Ταξινόμησης Δευτερεύουσας Μνήμης

- ▶ Εξωτερική Ταξινόμηση
- ▶ Εξωτερική Ταξινόμηση με Επιλογή Αντικατάστασης

Υποθέτουμε ότι έχουμε στη διάθεσή μας μαγνητικές ταινίες που χρησιμοποιούνται εναλλάξ ανά  $p$  για είσοδο και έξοδο δεδομένων.

Η κύρια μνήμη έχει μέγεθος  $M$  και ταξινομούμε  $N \gg M$  στοιχεία.

# Εξωτερική Ταξινόμηση

Έστω  $M = 3$ ,  $p = 3$

Ακολουθία εισόδου προς ταξινόμηση (αγνοούμε τα κενά)

“A SORTING AND MERGING EXAMPLE”

## Φάση Α

Εφόσον  $M = 3$ , η ακολουθία διαβάζεται κατά μπλοκ μήκους 3 ως εξής:

ASORTINGANDMERGINGEXAMPLE

Κάθε  $i$  – οστό ταξινομημένο μπλοκ εγγράφεται στην ταινία  $(i \bmod p)$

Π.χ. Μπλοκ 4 = DMN, το γράφουμε ταξινομημένα στην  $(4 \bmod 3 = 1)$  ταινία, άρα στην 1η ταινία δίπλα από το 1ο μπλοκ.

<b>ΤΑΙΝΙΑ 1</b>	AOS	DMN	AEX
<b>ΤΑΙΝΙΑ 2</b>	IRT	EGR	LMP
<b>ΤΑΙΝΙΑ 3</b>	AGN	GIN	E

# Εξωτερική Ταξινόμηση

Έστω  $M = 3$ ,  $p = 3$

Ακολουθία εισόδου προς ταξινόμηση (αγνοούμε τα κενά)

“A SORTING AND MERGING EXAMPLE”

## Φάση Β

Συγχώνευση (επαναληπτικά) των μπλοκ της Φάσης Α και εγγραφή στην ταινία  $(i \bmod p) + p$ .

Συγχωνεύουμε τα  $p = 3$  πρώτα μπλοκ από τις ταινίες εισόδου στην κύρια μνήμη και τα γράφουμε στην ταινία  $p+1$ . Στη συνέχεια, συγχωνεύουμε τα επόμενα  $p = 3$  μπλοκ από τις ταινίες εισόδου στην κύρια μνήμη και τα γράφουμε στην ταινία  $p+2$  κ.ο.κ.

<b>TAINIA 4</b>	AAGINORST
<b>TAINIA 5</b>	DEGGIMNR
<b>TAINIA 6</b>	AEELMPX

# Εξωτερική Ταξινόμηση

Έστω  $M = 3$ ,  $p = 3$

Ακολουθία εισόδου προς ταξινόμηση (αγνοούμε τα κενά)

“A SORTING AND MERGING EXAMPLE”

## Φάση Β

Οι  $p$  ταινίες (1-3) είναι οι ταινίες εισόδου και οι υπόλοιπες  $p$  (4-6) οι ταινίες εξόδου.

**Συγχώνευση:** Οι ταινίες εισόδου (1-3) “ξανατυλίγονται” σε μία ταινία εξόδου, την Ταινία 1. Οι ταινίες εξόδου (4-6) γίνονται ταινίες εισόδου.

Σαρώνουμε τα πρώτα στοιχεία της κάθε ταινίας και διώχνουμε το μικρότερο στην ταινία εξόδου. Στη θέση του στοιχείου που έφυγε ολισθαίνει το αμέσως επόμενο στοιχείο της ταινίας που αυτό ανήκει.

# Εξωτερική Ταξινόμηση

<b>A</b>	A	G	I	N	O	R	S	T
D	E	G	G	I	M	N	N	R
A	E	E	L	M	P	X		

Ταινία 1: **A**

<b>A</b>	G	I	N	O	R	S	T	
D	E	G	G	I	M	N	N	R
A	E	E	L	M	P	X		

Ταινία 1: **AA**

G	I	N	O	R	S	T		
D	E	G	G	I	M	N	N	R
<b>A</b>	E	E	L	M	P	X		

Ταινία 1: **AAA**

Έξοδος A,  
Ολίσθηση A μία  
θέση αριστερά

Έξοδος A,  
Ολίσθηση G μία  
θέση αριστερά

Έξοδος A,  
Ολίσθηση E μία  
θέση αριστερά



# Εξωτερική Ταξινόμηση

G	I	N	O	R	S	T		
D	E	G	G	I	M	N	N	R
E	E	L	M	P	X			

Ταινία 1: **AAAD**

G	I	N	O	R	S	T		
E	G	G	I	M	N	N	R	
E	E	L	M	P	X			

Ταινία 1: **AAADE**

G	I	N	O	R	S	T		
G	G	I	M	N	N	R		
E	E	L	M	P	X			

Ταινία 1: **AAADEE**

Έξοδος D,  
Ολίσθηση E μία  
θέση αριστερά

Έξοδος E,  
Ολίσθηση G μία  
θέση αριστερά

Έξοδος E,  
Ολίσθηση E μία  
θέση αριστερά

# Εξωτερική Ταξινόμηση

G	I	N	O	R	S	T		
G	G	I	M	N	N	R		
E	L	M	P	X				

Ταινία 1: **AAADEEEE**

G	I	N	O	R	S	T		
G	G	I	M	N	N	R		
L	M	P	X					

Ταινία 1: **AAADEEEEG**

I	N	O	R	S	T			
G	G	I	M	N	N	R		
L	M	P	X					

Ταινία 1: **AAADEEEEGG**

Έξοδος E,  
Ολίσθηση L μία  
θέση αριστερά

Έξοδος G,  
Ολίσθηση I μία  
θέση αριστερά

Έξοδος G,  
Ολίσθηση G μία  
θέση αριστερά

# Εξωτερική Ταξινόμηση

I	N	O	R	S	T			
<b>G</b>	I	M	N	N	R			
L	M	P	X					

Ταινία 1: **AAADEEEGGG**

<b>I</b>	N	O	R	S	T			
I	M	N	N	R				
L	M	P	X					

Ταινία 1: **AAADEEEGGGI**

N	O	R	S	T				
<b>I</b>	M	N	N	R				
L	M	P	X					

Ταινία 1: **AAADEEEGGGII**

Έξοδος G,  
Ολίσθηση I μία  
θέση αριστερά

Έξοδος I,  
Ολίσθηση N μία  
θέση αριστερά

Έξοδος I,  
Ολίσθηση M μία  
θέση αριστερά

# Εξωτερική Ταξινόμηση

- ▶ Συνεχίζουμε την προηγούμενη διαδικασία μέχρι να εμφανιστεί ένα και μόνο μπλοκ στην έξοδο, δηλαδή στην Ταινία 1 να εμφανιστούν τα γράμματα της ακολουθίας εισόδου ταξινομημένα.

Ταινία 1: **AAADEEEGGGIILMMNNNOPRRSTX**

# Εξωτερική Ταξινόμηση με Επιλογή Αντικατάστασης

- ▶ Παραλλαγή της Φάσης A του προηγούμενου αλγόριθμου.
- ▶ Η Φάση B είναι ίδια με του προηγούμενου αλγόριθμου
- ▶ Το μέγεθος των μπλοκ που παράγονται κατά την εφαρμογή του τρέχοντος αλγορίθμου στη Φάση A είναι κατά μέσο όρο 2-πλάσιο σε σχέση με τον προηγούμενο μειώνοντας τα βήματα συγχώνευσης.
- ▶ Ταξινόμηση Ακολουθίας Εισόδου: EXTERNAL SORTING

# Εξωτερική Ταξινόμηση με Επιλογή Αντικατάστασης- Φάση Α

HEAP	PIZA HEAP	ΕΠΟΜΕΝΟ	ΕΞΟΔΟΣ	ΣΧΟΛΙΑ
EXT	E	E	E	
EXT	E	R	EE	
RTX	R	N	EER	(N<R μπαίνει με *)
N*XT	T	A	EERT	(A<T μπαίνει με *)
N*XA*	X	L	EERTX	(1ο μπλοκ-Ταινία 1)
N*L*A*	A*	S	A*	
N*L*S*	L*	O	A*L*	
N*O*S	N*	R	A*L*N*	
R*O*S	O*	T	A*L*N*O*	
R*T*S*	R*	I	A*L*N*O*R*	(I<R μπαίνει χωρίς *)
IT*S*	S*	N	A*L*N*O*R*S*	(N<T μπαίνει χωρίς *)
IT*N	T*	G	A*L*N*O*R*S*T*	(2ο μπλοκ -Ταινία 2)
IGN		-	IGN	(3ο μπλοκ-Ταινία 3)

# ΑΝΑΖΗΤΗΣΗ ΠΡΟΒΛΗΜΑ ΛΕΞΙΚΟΥ

# ΠΕΡΙΕΧΟΜΕΝΑ

16

1. Εισαγωγή
2. Δυαδική Αναζήτηση
3. Αναζήτηση Παρεμβολής
4. Δυαδική Αναζήτηση Παρεμβολής

16



# Εισαγωγή

17

- Έστω σύνολο  $S \subseteq U$  (σύμπαν) ,  $x \in S, \text{inf}(x)$
- Το λεξικό είναι δομή δεδομένων που υποστηρίζει τις πράξεις:

*Access(x): if  $x \in S$  true return( $\text{inf}(x)$ ) else false* Στατικές Δομές

*Delete:  $S = S - \{x\}$*

*Insert:  $S = S \cup \{x\}$*

Δυναμικές Δομές

17

# Εισαγωγή

18

- **Insert(x) και Delete(x) είναι καταστρεπτικές πράξεις**
  - Εφήμερες Δομές: κάθε πράξη ενημέρωσης αλλάζει την δομή
- **Διαχρονικές Δομές (Persistent): Δομές που υποστηρίζουν Insert(x) και Delete(x) με Μη-Καταστρεπτικό τρόπο και διατηρούν όλες τις εκδόσεις της δομής**
  - Μόνο Access -> Μερικώς διαχρονικές
  - Insert, Delete σε προηγούμενες εκδόσεις -> Πλήρως διαχρονικές
- **Κατηγοριοποίηση ως προς το χώρο**
  - **Συνοπτικές Δομές:** ουρά με χρήση πίνακα
  - **Εκτενείς Δομές:** ουρά με χρήση δεικτών

18

# ΤΡΟΠΟΙ ΑΝΑΖΗΤΗΣΗΣ ΣΤΟ ΛΕΞΙΚΟ

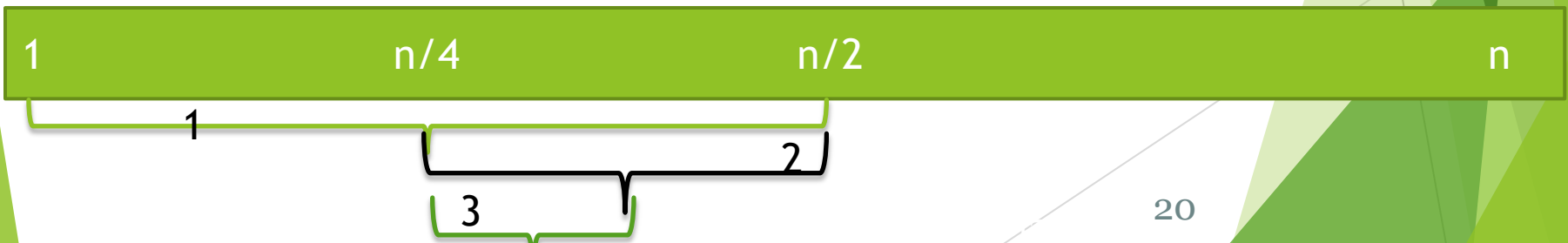
19

- Γραμμική:  $next \leftarrow left + 1$  (μη ταξινομημένα τα στοιχεία)
- Δυϊκή:  $next \leftarrow \left\lfloor \frac{right+left}{2} \right\rfloor$  (ταξινομημένα τα στοιχεία)
- Παρεμβολής:  $next \leftarrow \left\lfloor \frac{x-s[left]}{s[right]-s[left]} (right - left) \right\rfloor + left$

19

# ΔΥΙΔΙΚΗ ΑΝΑΖΗΤΗΣΗ

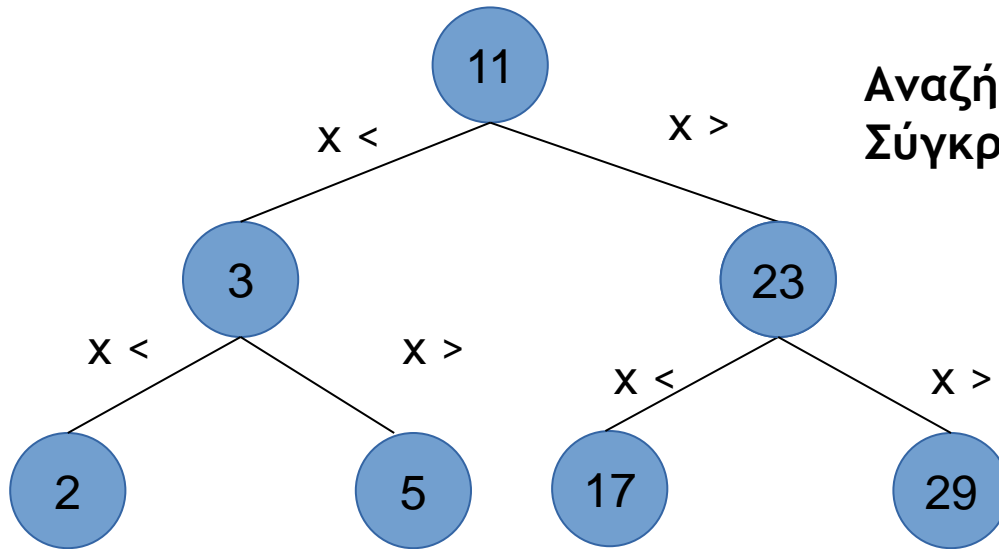
- ▶ Σε κάθε βήμα του while loop ελέγχεται το μέσο του διαστήματος.
- ▶ Προοδευτικά, καθώς το υπό εξέταση διάστημα υποδιπλασιάζεται, μετά από έναν αριθμό επαναλήψεων, αυτό θα αποκτήσει μήκος 1.
- ▶ Ο αλγόριθμος θα τερματίσει γιατί είτε το μοναδικό στοιχείο θα είναι το  $x$  (αυτό που ψάχνουμε) είτε το διάστημα θα γίνει μηδενικό και άρα το  $x$  δεν ανήκει στα στοιχεία του.



# ΔΥΙΚΗ ΑΝΑΖΗΤΗΣΗ

21

2	3	5	11	17	23	29
---	---	---	----	----	----	----



**Αναζήτηση του στοιχείου  $x = 7$   
Σύγκριση με 11,3,5**

**Αναζήτηση του στοιχείου  $x = 17$   
Σύγκριση με 11,23,17**

21

# ΑΝΑΖΗΤΗΣΗ ΠΑΡΕΜΒΟΛΗΣ

- ▶ Αναζήτηση ανάλογη της αίσθησης που έχει ο άνθρωπος όταν αναζητά σε ένα λεξικό.
- ▶  $next \leftarrow \left\lfloor \frac{x-s[left]}{s[right]-s[left]} (right - left) \right\rfloor + left$
- ▶ Πόσο μεγαλύτερο είναι το ζητούμενο στοιχείο από το αριστερό άκρο
- ▶ Πόσο μεγαλύτερο είναι το δεξί άκρο από το αριστερό
- ▶ Παίζει ρόλο στην απόδοση το πόσο διαφέρουν τα στοιχεία από τα γειτονικά.

# ΑΝΑΖΗΤΗΣΗ ΠΑΡΕΜΒΟΛΗΣ

## Αναζήτηση του 103

23

left = 1	2	3	4	5	6	right = 7
100	101	102	103	104	105	106

1 ↑  $Next = ((103-100)/(106-100))*(7-1) + 1 = 4$

100	101	103	107	108	111	1112
-----	-----	-----	-----	-----	-----	------

1 ↑  $Next = ((103-100)/(112-100))*(7-1) + 1 = 2$

2 ↑  $Next = ((103-101)/(112-101))*(7-2) + 2 = 3$

# ΑΝΑΖΗΤΗΣΗ ΠΑΡΕΜΒΟΛΗΣ

24

## Αναζήτηση του 103

1	2	3	4	5	6	7
100	101	102	103	214	1821	2004

1 ↑ Next =  $((103-100)/(2004-100))*(7-1) + 1 = 1$

2 ↑ Next =  $((103-101)/(2004-101))*(7-2) + 2 = 2$

3 ↑ Next =  $((103-102)/(2004-102))*(7-3) + 3 = 3$

4 ↑ Next =  $((103-103)/(2004-103))*(7-4) + 4 = 4$

24