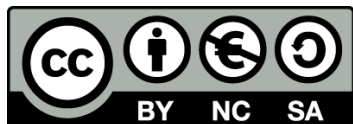




ΑΛΓΟΡΙΘΜΟΙ

Ενότητα 7: Χωρικοί – Χρονικοί Συμβιβασμοί

Ιωάννης Μανωλόπουλος, Καθηγητής
Αναστάσιος Γούναρης, Επίκουρος Καθηγητής
Τμήμα Πληροφορικής ΑΠΘ



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





Χωρικοί – Χρονικοί Συμβιβασμοί

Ταξινόμηση με απαρίθμηση, Ταύτιση
Προτύπου, Κατακερματισμός



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
πρόγραμμα για την ανάπτυξη

Χωρικοί-χρονικοί συμβιβασμοί

Σε πολλά προβλήματα, ο επιπλέον χώρος ευνοεί τις χρονικές επιδόσεις (extra χώρος = χώρος για αναπνοή)

- Ενίσχυση εισόδου
 - Ταξινόμηση μη βασισμένη σε συγκρίσεις
 - Βοηθητικοί πίνακες ολίσθησης στην ταύτιση προτύπου
- Προ-δόμηση
 - κατακερματισμός
 - κατάλογοι (πχ, B-δένδρα)
- Πίνακες με πληροφορία, όπου γίνεται η διεργασία
 - Δυναμικός προγραμματισμός



Ταξινόμηση με απαρίθμηση (1)

Algorithm ComparisonCountSort(A[0..n-1])

```
for i ← 0 to n-1 do Count[i] ← 0
```

```
for i ← 0 to n-2 do
```

```
    for j ← i+1 to n-1 do
```

```
        if A[i] < A[j] then Count[j] ← Count[j] + 1
```

```
        else Count[i] ← Count[i] + 1
```

```
for i ← 0 to n-1 do S[Count[i]] ← A[i]
```

- Παράδειγμα: 62 31 84 96 19 47
- Αποδοτικότητα



Ταξινόμηση με απαρίθμηση (2)

Algorithm DistributeCountSort(A[0..n-1])

for $j \leftarrow 0$ to $u-1$ do $D[j] \leftarrow 0$

for $i \leftarrow 0$ to $n-1$ do $D[A[i]-1] \leftarrow D[A[i]-1] + 1$

for $j \leftarrow 1$ to $u-1$ do $D[j] \leftarrow D[j-1] + D[j]$

for $i \leftarrow n-1$ down to 0 do

$j \leftarrow A[i]-1$; $S[D[j]-1] \leftarrow A[i]$; $D[j] \leftarrow D[j]-1$

- Παράδειγμα: 13 11 12 13 12 12
- Αποδοτικότητα



Ταύτιση προτύπου

- πρότυπο: string m χαρακτήρων προς αναζήτηση
- κείμενο: string $n \gg m$ χαρακτήρων, όπου θα γίνει αναζήτηση
- Αλγόριθμος ωμής βίας:
 1. Ευθυγραμμίζουμε το πρότυπο με την αρχή του κειμένου
 2. Κινούμενοι από αριστερά προς τα δεξιά, συγκρίνουμε κάθε χαρακτήρα του προτύπου με τον αντίστοιχο του κειμένου μέχρι
 - να ταυτίζονται όλοι οι χαρακτήρες (επιτυχής αναζήτηση), ή
 - να βρεθεί μία μη-ταύτιση
 3. Όσο δεν βρίσκεται το πρότυπο και το κείμενο δεν έχει τελειώσει, επανα-ευθυγραμμίζουμε το πρότυπο κατά μία θέση και πηγαίνουμε στο βήμα 2.



Ταύτιση προτύπου – ιστορικό

- 1970: ο Cook δείχνει (finite-state machines) ότι το πρόβλημα μπορεί να επιλυθεί σε χρόνο ανάλογο του $n+m$.
- 1976: ο Knuth και ο Pratt βρίσκουν έναν αλγόριθμο βασισμένο στην ιδέα του Cook. Ο Morris ανεξάρτητα βρίσκει τον ίδιο αλγόριθμο στην προσπάθεια να αποφύγει τις “επιστροφές” στο ίδιο κείμενο.
- Περίπου την ίδια περίοδο οι Boyer και Moore βρίσκουν έναν αλγόριθμο που εξετάζει μόνο ένα τμήμα των χαρακτήρων του κειμένου στις περισσότερες περιπτώσεις (συγκρίνοντας χαρακτήρες του προτύπου και του κειμένου από δεξιά προς τα αριστερά, αντί από αριστερά προς τα δεξιά).
- 1980: Ένας άλλος αλγόριθμος των Rabin και Karp πρακτικά τρέχει πάντα σε χρόνο ανάλογο του $n+m$, ενώ έχει το πλεονέκτημα ότι γενικεύεται εύκολα στις 2 διαστάσεις και είναι τόσο απλός όσο αυτός της ωμής βίας.



Αλγόριθμος Horspool

- Είναι μία απλή εκδοχή του αλγορίθμου Boyer-Moore (κρατώντας τα βασικά):
 - Συγκρίνει από δεξιά προς τα αριστερά τους χαρακτήρες του προτύπου προς το κείμενο
 - Δοθέντος ενός προτύπου, δημιουργεί ένα πίνακα ολίσθησης, που δίνει το μέτρο ολίσθησης σε περίπτωση μη-ταύτισης (*ενίσχυση εισόδου*)



Πόσο μακριά να γίνει η ολίσθηση;

Βλέπουμε τον πρώτο (δεξιότερο) χαρακτήρα κειμένου που συγκρίνεται:

- Ο χαρακτήρας δεν υπάρχει στο πρότυπο

.....C..... (το c δεν υπάρχει στο πρότυπο)

BAOBAB

- Ο χαρακτήρας υπάρχει στο πρότυπο αλλά όχι στη δεξιότερη θέση

.....O..... (το O υπάρχει 1 φορά στο πρότυπο)

BAOBAB

.....A..... (το A υπάρχει 2 φορές στο πρότυπο)

BAOBAB

- Ο δεξιότερος χαρακτήρας προκάλεσε ταύτιση

.....B.....

BAOBAB

- Πίνακας ολίσθησης: Αποθηκεύει το πλήθος των χαρακτήρων που θα ολισθήσουν με βάση τον πρώτο συγκρινόμενο χαρακτήρα



Πίνακας ολίσθησης (1)

- Δημιουργείται σαρώνοντας το πρότυπο πριν την αναζήτηση
- Όλες οι θέσεις αρχικοποιούνται με το μήκος του προτύπου
- Ενημερώνουμε τη θέση του χαρακτήρα c που υπάρχει στο πρότυπο με την απόσταση της δεξιότερης εμφάνισης του c από το τέλος του προτύπου

Algorithm ShiftTable(P[0..m-1])

```
for  $i \leftarrow 0$  to  $size-1$  do  $Table[i] \leftarrow m$   
for  $j \leftarrow 0$  to  $m-2$  do  $Table[P[j]] \leftarrow m-1-j$   
return Table
```



Πίνακας ολίσθησης (2)

Παράδειγμα για το πρότυπο ΒΑΟΒΑΒ:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6

Τότε:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
<u>1</u>	<u>2</u>	6	6	6	6	6	6	6	6	6	6	6	6	<u>3</u>	6	6	6	6	6	6	6	6	6	6	6



Αλγόριθμος Horspool

- Horspool Matching(P[0..m-1],T[0..n-1])

ShiftTable (P[0..m-1])

$i \leftarrow m-1$

while $i \leq n-1$ do

$k \leftarrow 0$

 while $k \leq m-1$ and $P[m-1-k] = T[i-k]$ do

$k \leftarrow k+1$

 if $k = m$ return $i-m+1$

 else $i \leftarrow i + \text{Table}[T[i]]$

return -1



Αλγόριθμος Boyer-Moore

- Βασίζεται στις ίδιες 2 ιδέες:
 - συγκρίνουμε τους χαρακτήρες του προτύπου με το κείμενο από δεξιά προς τα αριστερά
 - δεδομένου ενός προτύπου, δημιουργεί ένα πίνακα ολίσθησης για την περίπτωση μη ταύτισης (*ενίσχυση εισόδου*)
- Χρησιμοποιεί έναν επιπλέον πίνακα ολίσθησης, ο οποίος εφαρμόζεται ανάλογα με το πλήθος των χαρακτήρων που συμπίπτουν



Ολίσθηση «κακού» συμβόλου

- Βασίζεται στην ιδέα του Horspool για τη χρήση ενός επιπλέον πίνακα, ωστόσο αυτός ο πίνακας δημιουργείται διαφορετικά:
 - Αν ο χαρακτήρας c του κειμένου που αντιστοιχεί με τον τελευταίο χαρακτήρα του προτύπου δεν υπάρχει στο πρότυπο, τότε ολισθαίνουμε κατά m χαρακτήρες (το c είναι «κακό» σύμβολο)
 - Αν ο μη ταυτιζόμενος χαρακτήρας (το «κακό» σύμβολο) δεν εμφανίζεται στο πρότυπο, τότε ολισθαίνουμε ώστε να τον προσπεράσουμε
 - Αν ο μη ταυτιζόμενος χαρακτήρας (το «κακό» σύμβολο) εμφανίζεται στο πρότυπο, τότε ολισθαίνουμε ώστε να ευθυγραμμίσουμε το κακό σύμβολο με τον ίδιο χαρακτήρα (που βρίσκεται αριστερά της θέσης όπου υπήρξε μη-ταύτιση).



Ολίσθηση «κακού» συμβόλου – παράδειγμα

Το κακό σύμβολο δεν βρίσκεται στο πρότυπο

...**S**ER.....

BAR**B**ER

BARBER

ολισθαίνουμε 4 θέσεις

Το κακό σύμβολο βρίσκεται στο πρότυπο

...**A**ER.....

BA**R**BER

BARBER

ολισθαίνουμε 2 θέσεις

Η ολίσθηση αυτή δίνεται από: $d = \max[t_1(c) - k, 1]$, όπου το t_1 είναι ο πίνακας του Horspool, k είναι η απόσταση του κακού συμβόλου από το τέλος του προτύπου



Ολίσθηση «καλού» επιθέματος – παράδειγμα (1)

- Τι συμβαίνει αν το ταυτισμένο επίθεμα εμφανίζεται και πάλι στο πρότυπο (πχ. **ABRACADABRA**)
- Είναι σημαντικό να βρούμε ένα άλλο επίθεμα με διαφορετικό προηγούμενο χαρακτήρα. Η ολίσθηση ισούται με την απόσταση μεταξύ των δύο εμφανίσεων του επιθέματος
- Επίσης, είναι σημαντικό να βρούμε το μακρύτερο πρόθεμα μεγέθους $l < k$ που να ταυτίζεται με το επίθεμα μήκους l . Η ολίσθηση ισούται με την απόσταση μεταξύ επιθέματος-προθέματος



Ολίσθηση «καλού» επιθέματος – παράδειγμα (2)

K	pattern	d2
1	ABCBA <u>B</u>	2
2	ABCBA <u>B</u>	4
3	ABC <u>B</u> AB	4
4	ABC <u>B</u> AB	4
5	<u>ABC</u> BAB	<u>4</u>

K	pattern	d2
1	BAOBA <u>B</u>	2
2	BAOBA <u>B</u>	5
3	BAO <u>B</u> AB	5
4	BAO <u>B</u> AB	5
5	BAO <u>B</u> AB	5



Τελικός κανόνας του Boyer-Moore

- Υπολογίζουμε την ολίσθηση ως
$$d = \begin{cases} d1 & \text{αν } k=0 \\ \max(d1, d2) & \text{αν } k>0 \end{cases}$$
όπου $d1 = \max(t1(c) - k)$

- Παράδειγμα
BESS_KNEW_ABOUT_BAOBABS
BAOBAB



Κατακερματισμός

- Είναι μία πολύ αποτελεσματική μέθοδος για την υλοποίηση του λεξικού, δηλαδή το σύνολο των πράξεων:
 - εισαγωγή
 - αναζήτηση
 - διαγραφή
- Εφαρμογές:
 - βάσεις δεδομένων
 - πίνακες συμβόλων



Συναρτήσεις Κατακερματισμού

- Πίνακας κατακερματισμού: ένας πίνακας με θέσεις που αντιστοιχούν σε buckets
- Συνάρτηση κατακερματισμού: δίνει το bucket για κάθε record
- Παράδειγμα: έστω ότι κλειδί είναι το SSN. Η συνάρτηση κατακερματισμού είναι:

$$h(k) = k \bmod m$$

(k είναι το κλειδί και m το πλήθος των buckets)

- αν $m=1000$, που αποθηκεύεται το record SSN= 315-17-4251 ;
- Οι συναρτήσεις κατακερματισμού πρέπει:
 - να υπολογίζονται εύκολα
 - να κατανέμουν τα κλειδιά ομοιόμορφα σε όλο τον πίνακα



Συγκρούσεις

- Αν $h(k_1)=h(k_2)$, τότε υπάρχει σύγκρουση
- Μία καλή συνάρτηση προκαλεί λιγότερες συγκρούσεις
- Οι συγκρούσεις δεν μπορούν να εξαλειφθούν
- Οι συγκρούσεις λύνονται με δύο τρόπους:
 - *Ανοικτός κατακερματισμός* – το bucket δείχνει σε μία λίστα με όλα τα κλειδιά που έρχονται σε αυτή τη θέση
 - *Κλειστός κατακερματισμός* – ένα κλειδί ανά bucket, σε περίπτωση σύγκρουσης, βρες ένα άλλο bucket για το ένα κλειδί
 - *Γραμμική αναζήτηση*: πήγαινε στο επόμενο bucket
 - *Διπλός κατακερματισμός*: χρησιμοποίησε μία άλλη συνάρτηση κατακερματισμού για να βρεις το άλμα



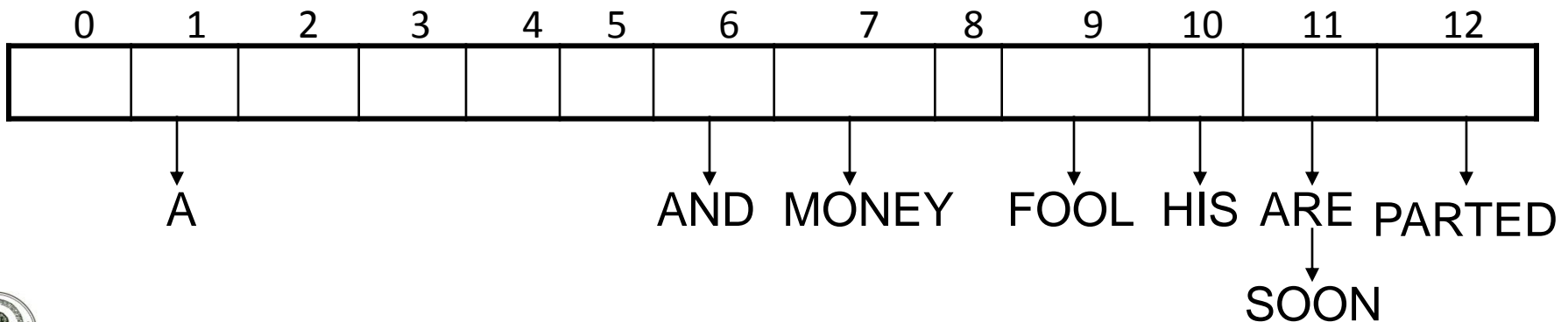
Ανοικτός κατακερματισμός (1)

Τα κλειδιά απόθηκεύονται σε λίστες έξω από τον πίνακα κατακερματισμού, που περιέχει μόνο τις κορυφές των λιστών.

Παράδειγμα: A, FOOL, AND, HIS, MONEY, ARE, SOON, PARTED

$h(K) = \text{άθροισμα των θέσεων των γραμμάτων του } K \text{ MOD } 13$

Key	A	FOOL	AND	HIS	MONEY	ARE	SOON	PARTED
$h(K)$	1	9	6	10	7	11	11	12



Ανοικτός κατακερματισμός (2)

- Αν η συνάρτηση κατακερματισμού κατανέμει τα κλειδιά ομοιόμορφα, το μέσο μήκος της λίστας είναι $a=n/m$ (a =παράγοντας φόρτωσης)
- Το μέσο πλήθος των συγκρίσεων είναι $S=1+a/2$, $U=a$
- Η χειρότερη περίπτωση είναι γραμμική!
- Ο ανοικτός κατακερματισμός λειτουργεί ακόμη και αν $n>m$



Κλειστός κατακερματισμός (1)

Τα κλειδιά απόθηκεύονται μέσα στον πίνακα κατακερματισμού.

Key	A	FOOL	AND	HIS	MONEY	ARE	SOON	PARTED
$h(K)$	1	9	6	10	7	11	11	12

	0	1	2	3	4	5	6	7	8	9	10	11	12
	A												
	A								FOOL				
	A					AND			FOOL				
	A					AND			FOOL	HIS			
	A					AND	MONEY		FOOL	HIS			
	A					AND	MONEY		FOOL	HIS	ARE		
	A					AND	MONEY		FOOL	HIS	ARE	SOON	
PARTED	A					AND	MONEY		FOOL	HIS	ARE	SOON	



Κλειστός κατακερματισμός (2)

- Δεν λειτουργεί αν $n > m$ και αποφεύγει τους δείκτες
- Οι διαγραφές δεν είναι άμεσες
- Το πλήθος των συγκρίσεων για μια εισαγωγή/ αναζήτηση/ διαγραφή ενός κλειδιού εξαρτάται από το α (=πυκνότητα του πίνακα κατακερματισμού)
 - επιτυχής αναζήτηση: $(\frac{1}{2}) (1 + \frac{1}{(1-\alpha)})$
 - ανεπιτυχής αναζήτηση : $(\frac{1}{2}) (1 + \frac{1}{(1-\alpha)^2})$
- καθώς γεμίζει ο πίνακας ($\alpha \rightarrow 1$), το πλήθος των συγκρίσεων αυξάνει δραματικά

α	$-\left(\frac{1}{1-\alpha} \right)$	$-\left(\frac{1}{(1-\alpha)^2} \right)$
50%	1.5	2.5
75%	2.5	8.5
90%	5.5	50.5



Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, **Ιωάννης
Μανωλόπουλος, Αναστάσιος Γούναρης**. «Αλγόριθμοι. ». Έκδοση: 1.0.
Θεσσαλονίκη 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:
<http://eclass.auth.gr/courses/OCRS417/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά - Μη Εμπορική Χρήση - Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>





Τέλος ενότητας

Επεξεργασία: Ανδρέας Κοσματόπουλος
Θεσσαλονίκη, Αύγουστος 2015



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση **1.00**.



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

