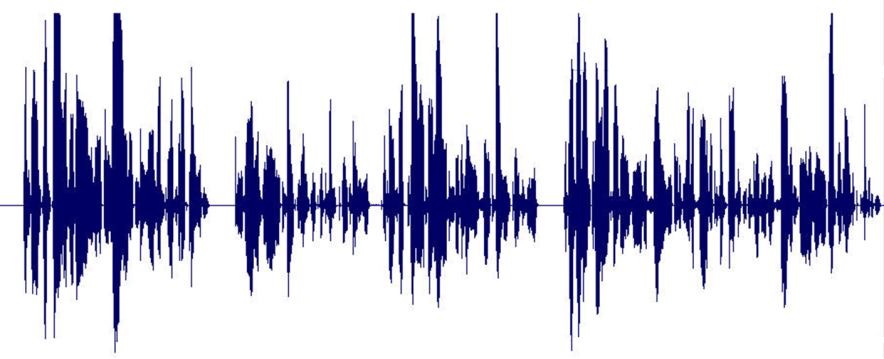
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Π.Μ.Σ. «Προηγμένα Συστήματα Πληροφορικής - Ανάπτυξη Λογισμικού και Τεχνητής Νοημοσύνης»



**Αναγνώριση Ομιλίας και Ήχου**

****

**Δαβράδου Αγάπη**

**Πάνου Γιώργος**

**Σπυρόπουλος Κωνσταντίνος**

Καθηγητής: Πικράκης Άγγελος

Αθήνα  
30/11/2019

##### Περίληψη

Η παρούσα εργασία υλοποιήθηκε στα πλαίσια του μαθήματος «Αναγνώριση Ομιλίας και Ήχου» και αποσκοπεί στην υλοποίηση ενός Automatic Speech Recognition (ASR) συστήματος. Το τελικό σύστημα δέχεται ως είσοδο μία ηχογράφηση, η οποία συνιστά πρόταση αποτελούμενη από 4-10 ψηφία, από το 0 έως το 9, και τα οποία έχουν ειπωθεί με αρκούντως μεγάλα διαστήματα παύσης, περίπου ένα ψηφίο ανά δευτερόλεπτο. Το σύστημα έχει υλοποιηθεί έτσι, ώστε να μην εξαρτάται από τα χαρακτηριστικά της φωνής του ομιλητή. Στην έξοδό του το σύστημα, αφού κατατμήσει την πρόταση και αναγνωρίσει κάθε λέξη, παράγει σε μορφή κειμένου τα αποτελέσματα που αναγνωρίζει.

##### Περιεχόμενα

[Περίληψη vi](#_Toc25878325)

[Περιεχόμενα vii](#_Toc25878326)

[Κεφάλαιο 1: Εισαγωγή 1](#_Toc25878327)

[1.1 Ψηφιακή επεξεργασία σήματος 1](#_Toc25878328)

[1.1.1 Φιλτράρισμα 2](#_Toc25878329)

[1.1.2 Δειγματοληψία (Sampling) 2](#_Toc25878330)

[1.2 Τεχνητά Νευρωνικά Δίκτυα (Neural Networks - NN) 2](#_Toc25878331)

[1.2.1 Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN) 2](#_Toc25878332)

[Κεφάλαιο 2: Σχεδιασμός και Μοντελοποίηση 3](#_Toc25878333)

[2.1 Απλό νευρωνικό δίκτυο 3](#_Toc25878334)

[2.1.1 Προεπεξεργασία δεδομένων 3](#_Toc25878335)

[2.1.2 Εξαγωγή χαρακτηριστικών (feature extraction) 4](#_Toc25878336)

[2.1.3 Δομή 4](#_Toc25878337)

[2.2 Συνελικτικό νευρωνικό δίκτυο 4](#_Toc25878338)

[2.2.1 Προεπεξεργασία δεδομένων 4](#_Toc25878339)

[2.2.2 Δομή 4](#_Toc25878340)

[Κεφάλαιο 3: Αποτελέσματα 6](#_Toc25878341)

[3.1 Απλό νευρωνικό δίκτυο 6](#_Toc25878342)

[3.2 Συνελικτικό νευρωνικό δίκτυο 6](#_Toc25878343)

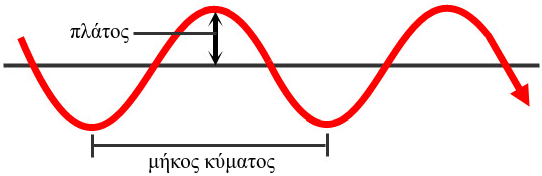
[Κεφάλαιο 4: Συμπεράσματα 8](#_Toc25878344)

[Βιβλιογραφία 9](#_Toc25878345)

# Εισαγωγή

## Ψηφιακή επεξεργασία σήματος

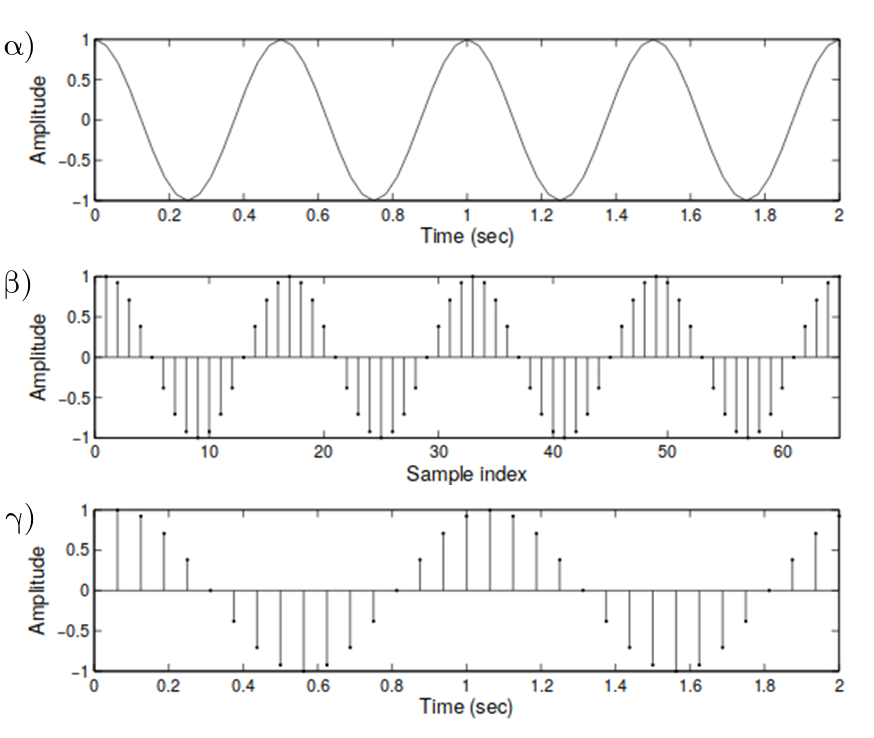
Το πρώτο βήμα της ψηφιακής επεξεργασίας σήματος είναι η μετατροπή ενός αναλογικού συστήματος σε ψηφιακό (δηλαδή η μετατροπή ενός συνεχούς σήματος σε διακριτό), κάνοντας χρήση ενός μετατροπέα αναλογικού σήματος σε ψηφιακό (analog-to-digital converter - ADC), ο οποίος μετασχηματίζει το αναλογικό σήμα σε μια ακολουθία από αριθμούς.



Σχήμα **1.1:** Παράμετροι ηχητικού σήματος.

Η μετατροπή πραγματοποιείται μέσω της δειγματοληψίας (sampling), η οποία συνήθως πραγματοποιείται σε δύο στάδια, στο στάδιο της διακριτοποίησης (discretization) και της κβάντισης σήματος (quantization). Η διακριτοποίηση σημαίνει ότι το σήμα χωρίζεται σε ίσα χρονικά διαστήματα και κάθε διάστημα αντιπροσωπεύεται από μία μόνο μέτρηση πλάτους (amplitude). Η κβάντιση σήματος σημαίνει ότι η κάθε μέτρηση πλάτους προσεγγίζεται με μία τιμή από ένα πεπερασμένο σύνολο.

Η συχνότητα δειγματοληψίας η ρυθμός δειγματοληψίας μετριέται σε Hz και εκφράζει το πλήθος δειγμάτων που λαμβάνονται σε διάρκεια ενός δευτερολέπτου. Για παράδειγμα, εάν η συχνότητα = 10.000Hz, σημαίνει ότι ο δειγματολήπτης λαμβάνει 10.000 δείγματα ανά δευτερόλεπτο σήματος. Η χρονική απόσταση των δειγμάτων υπολογίζεται από τον τύπο = , όπου Τ η περίοδος της δειγματοληψίας. Αντίστοιχα, για μία συχνότητα δειγματοληψίας = 10.000Hz, η περίοδος θα είναι Τ = = 0.0001 δευτερόλεπτα.



Σχήμα **1.2:** Αναπαράσταση μίας συνεχής κυματομορφής ημιτονοειδούς κύματος (α), του αντίστοιχου δειγματοληπτικού σήματος (β) και του ανασχηματισμού του δειγματοληπτικού σήματος στο μισό του αρχικού ρυθμού δειγματοληψίας.

Σύμφωνα με το θεώρημα δειγματοληψίας Nyquist–Shannon [1], [2], το οποίο ουσιαστικά αποτελεί μία γέφυρα μεταξύ των συνεχών και των διακριτών σημάτων, ένα αναλογικό σήμα μπορεί να αναπαραχθεί από το αντίστοιχο διακριτό όταν έχει χρησιμοποιηθεί συχνότητα δειγματοληψίας η οποία είναι τουλάχιστον διπλάσια από την μέγιστη συχνότητα του αρχικού σήματος.

Για παράδειγμα, αν οι συχνότητες ενός συνεχούς σήματος εκτείνονται μέχρι τα 10000Hz, τότε πρέπει να χρησιμοποιηθεί συχνότητα δειγματοληψίας τουλάχιστον ίση με 20KHz. Συνηθισμένες συχνότητες δειγματοληψίας για την περίπτωση των ηχητικών σημάτων είναι 44.1KHz για μουσική και 32, 16 ή και 8KHz για ομιλία.

Επιπλέον, αν και μεγάλες συχνότητες δειγματοληψίας βελτιώνουν την ποιότητα του ψηφιακού σήματος, μπορεί να οδηγήσουν σε μεγάλες υπολογιστικές πολυπλοκότητες στους αλγορίθμους ανάλυσης, καθώς αυξάνουν το πλήθος δειγμάτων ανά δευτερόλεπτο προς επεξεργασία.

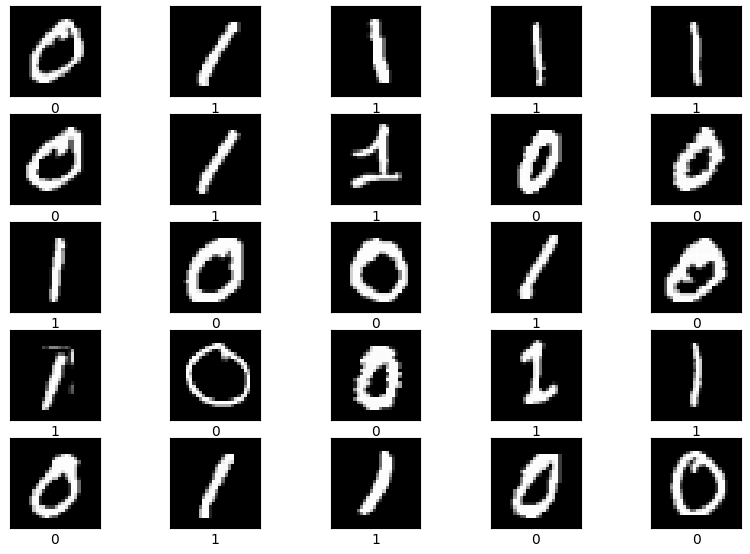
ΦΙΛΤΡΑΡΙΣΜΑ

## Τεχνητά Νευρωνικά Δίκτυα (Neural Networks - NN)

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.

ΕΠΙΤΗΡΟΥΜΕΝΗ ΜΑΘΗΣΗ;;;;;

### Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN)



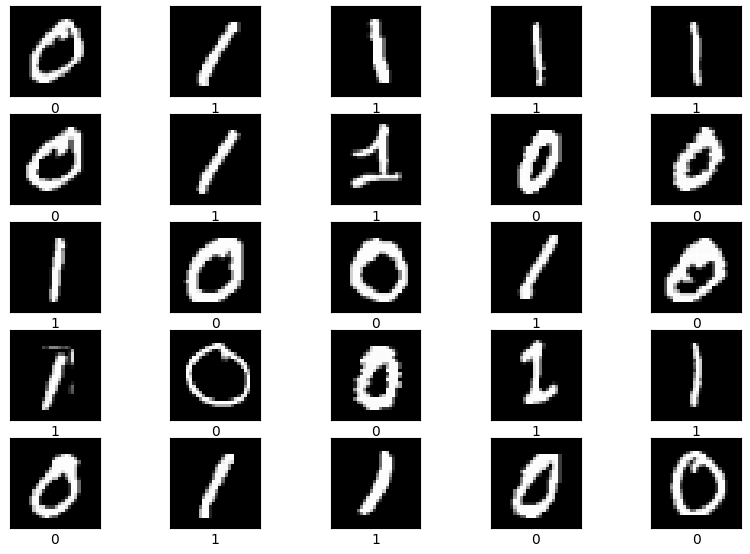
Σχήμα **1.1:** la bla bla bla bla blab la blab la

# Σχεδιασμός και Μοντελοποίηση

Αρχικά δημιουργήθηκε ένα απλό νευρωνικό δίκτυο, το οποίο δε λειτουργούσε καλά. Έτσι υλοποιήθηκαν δύο διαφορετικές μεθοδολογίες, οι οποίες αποσκοπούν στην επίτευξη του καλύτερου αποτελέσματος, αλλά και στο να διευρυνθεί η γνώση για την αποδοτικότητα των νευρωνικών δικτύων σε εφαρμογές αναγνωρισης ομιλίας και ήχου.

## Απλό νευρωνικό δίκτυο

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.



Σχήμα 2.1: la bla bla bla bla blab la blab la

### Προεπεξεργασία δεδομένων

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.

### Εξαγωγή χαρακτηριστικών (feature extraction)

### Δομή

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.

## Συνελικτικό νευρωνικό δίκτυο

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.

### Προεπεξεργασία δεδομένων

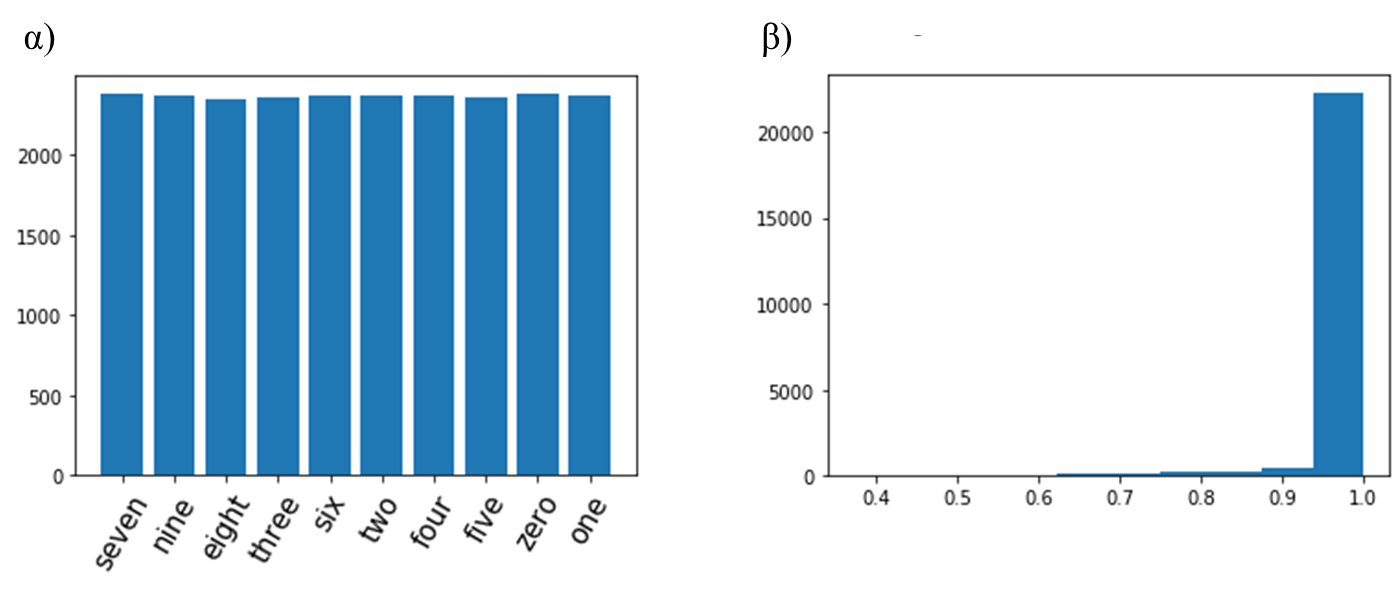
Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.

### Δομή

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.

# Αποτελέσματα

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε ένα σύνολο δεδομένων με εγγραφές εκφωνούμενων ψηφίων από 997 διαφορετικούς ομιλητές []. Συγκεκριμένα, το σύνολο περιλαμβάνει 10.000 εγγραφές για εκπαίδευση και 1.000 για αξιολόγηση, διαθέτοντας 10 κατηγορίες (classes) που αντιστοιχούν στα ψηφία από το 0 μέχρι και το 9.



Σχήμα 3.1: Γραφική αναπαράσταση του πλήθους κάθε κατηγορίας (α) και της διάρκειας (β) του συνόλου δεδομένων που χρησιμοποιήθηκε για εκπαίδευση και αξιολόγηση.

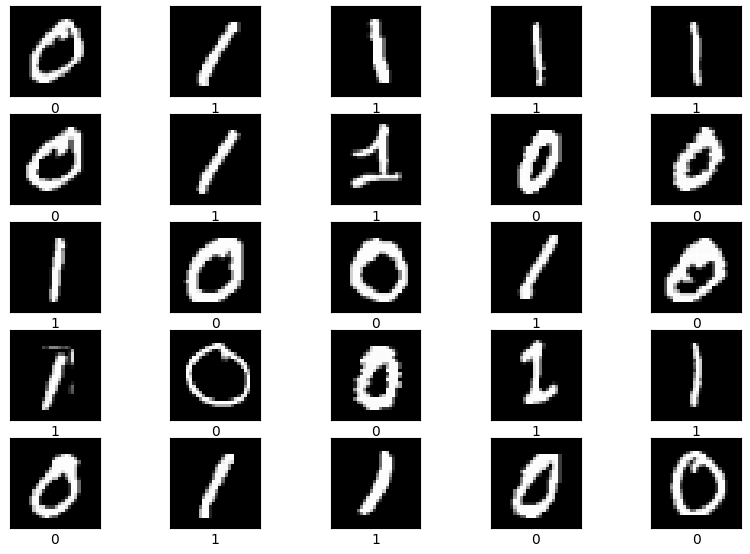
## Απλό νευρωνικό δίκτυο

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.

## Συνελικτικό νευρωνικό δίκτυο

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.

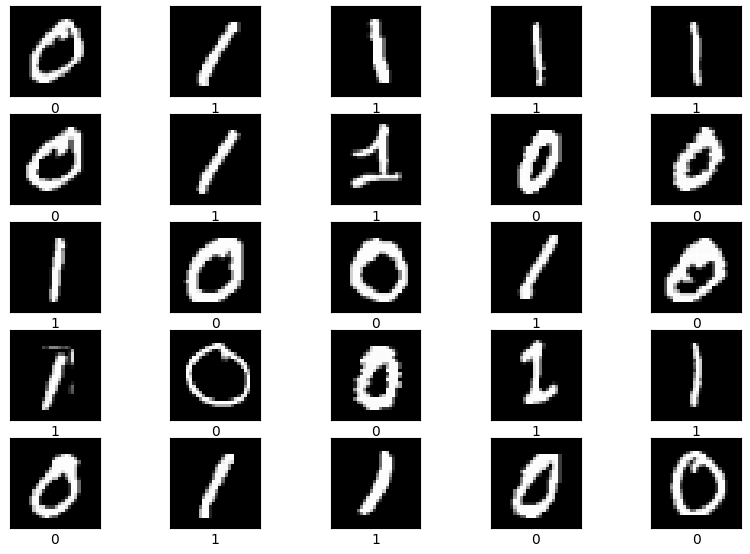
.



Σχήμα 3.1: bla bla bla bla blab la blab la

# Συμπεράσματα

Blab la bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla bla blab la blab la blab la bla bla bla.



Σχήμα 4.1: bla bla bla bla blab la blab la

##### Βιβλιογραφία

[1] Marks, R.J.(II): [*Introduction to Shannon Sampling and Interpolation Theory*](http://marksmannet.com/RobertMarks/REPRINTS/1999_IntroductionToShannonSamplingAndInterpolationTheory.pdf), Springer-Verlag, 1991.

[2] Marks, R.J.(II), Editor: [Advanced Topics in Shannon Sampling and Interpolation Theory](http://marksmannet.com/RobertMarks/REPRINTS/1993_AdvancedTopicsOnShannon.pdf), Springer-Verlag, 1993.

###### Παράρτημα A

**Κώδικες**