



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**  
**ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ**  
**ΜΑΘΗΜΑ: Διαχείριση Δεδομένων στον Παγκόσμιο Ιστό**  
**ΔΙΔΑΣΚΩΝ: Χρήστος Δουλκερίδης ([cdoulik@unipi.gr](mailto:cdoulik@unipi.gr))**

Δευτέρα 5 Δεκεμβρίου 2016

**2<sup>η</sup> ΕΡΓΑΣΙΑ:** Προγραμματισμός με XML

Το DBLP είναι ένας ιστότοπος (<http://www.informatik.uni-trier.de/~ley/db/index.html>) που περιέχει βιβλιογραφικές αναφορές σχετικά με την επιστήμη των υπολογιστών και φιλοξενείται από το University of Trier της Γερμανίας. Το Νοέμβριο του 2012, το DBLP κατέγραφε πληροφορία για περίπου 2,1 εκατ. άρθρα σχετικά με την επιστήμη των υπολογιστών. Για παράδειγμα, οι δημοσιεύσεις του διδάσκοντα βρίσκονται στον ακόλουθο σύνδεσμο του DBLP: <http://www.informatik.uni-trier.de/~ley/pers/hd/d/Doulkeridis:Christos.html>

Τα δεδομένα του DBLP είναι διαθέσιμα για download σε μορφή XML (<http://dblp.uni-trier.de/xml/dblp.xml>) σαν ένα τεράστιο αρχείο μεγέθους μεγαλύτερο από 1 GB.

Στην παρούσα εργασία καλείστε να γράψετε ένα πρόγραμμα σε Java που να εξάγει από αυτό το τεράστιο XML αρχείο τις δημοσιεύσεις των μελών ΔΕΠ του τμήματος Ψηφιακών Συστημάτων (<http://www.ds.unipi.gr/en/academical/>). Το πρόγραμμά σας θα παίρνει ως είσοδο ένα σύνολο από ονόματα ερευνητών (στα αγγλικά). Στη συνέχεια θα επεξεργάζεται το XML αρχείο με έναν από τους τρόπους που μάθατε στο μάθημα για parsing και επεξεργασία XML αρχείων, και θα παράγει μια λίστα με τις δημοσιεύσεις του κάθε ερευνητή του οποίου το όνομα δόθηκε ως είσοδος. Συγκεκριμένα, η λίστα θα περιέχει για κάθε δημοσίευση/άρθρο τα ονόματα των συγγραφέων, τον τίτλο, το πού δημοσιεύτηκε, και το έτος δημοσίευσης. Στο ακόλουθο δείγμα (το XML αρχείο περιέχει χιλιάδες τέτοιους κόμβους) πρέπει να εξαχθούν τα author, title, booktitle, year.

```
<dblp>
<inproceedings mdate="2011-06-23" key="conf/aime/BianchiD95">
<author>Nadia Bianchi</author>
<author>Claudia Diamantini</author>
<title>Integration of Neural Networks and Rule Based Systems in the
Interpretation of Liver Biopsy Images.</title>
<pages>367-378</pages>
<year>1995</year>
<crossref>conf/aime/1995</crossref>
<booktitle>AIME</booktitle>
<url>db/conf/aime/aime1995.html#BianchiD95</url>
<ee>http://dx.doi.org/10.1007/3-540-60025-6\_152</ee>
</inproceedings>
</dblp>
```

**BONUS:** Θα δοθεί εάν η λίστα που θα παραχθεί θα είναι αναγνώσιμη με εύκολο τρόπο από έναν άνθρωπο, άρα θα πρέπει να έχει μορφοποιηθεί κατάλληλα, για παράδειγμα ως html σελίδα.

Η εργασία μπορεί να εκπονηθεί ομαδικά σε ομάδες μέχρι 2 ατόμων.

**ΠΑΡΑΔΟΣΗ ΕΡΓΑΣΙΑΣ: Τρίτη 31 Ιανουαρίου 2017**

**ΠΑΡΑΔΟΤΕΑ:**

Με την ολοκλήρωση της εργασίας, θα πρέπει να παραδώσετε ένα αρχείο **ΑΜ1\_Επώνυμο1\_Όνομα1\_ΑΜ2\_Επώνυμο2\_Όνομα2.zip** (όπου ΑΜ ο αριθμός μητρώου) που θα περιλαμβάνει:

1. Τα αρχεία πηγαίου κώδικα (κλάσεις .java) που θα περιέχουν επαρκή σχόλια.
2. Τις αντίστοιχες μεταγλωττισμένες κλάσεις (.class)
3. Ένα αρχείο με οδηγίες εκτέλεσης του προγράμματος (pdf)
4. Ένα αρχείο με τα στοιχεία των φοιτητών (αριθμό μητρώου, ονοματεπώνυμο, e-mail) που αποτελούν την ομάδα (pdf)

Η παράδοση της εργασίας θα γίνει μέσω της πλατφόρμας (ΕΥΔΟΞΟΣ) του τμήματος. Ανεβάστε το αρχείο (ΑΜ\_Επώνυμο\_Όνομα.zip) στην περιοχή «Εργασίες».

**Η ΠΑΡΑΔΟΣΗ ΘΑ ΓΙΝΕΙ ΑΠΟΚΛΕΙΣΤΙΚΑ ΜΕΣΩ ΤΗΣ ΠΛΑΤΦΟΡΜΑΣ ΕΥΔΟΞΟΣ**

**Καλή Επιτυχία**