



ΑΚΑΔΗΜΙΑ ΠΛΑΤΩΝΟΣ – ΑΝΑΠΤΥΞΗ ΤΗΣ ΓΝΩΣΗΣ ΚΑΙ ΚΑΙΝΟΤΟΜΩΝ ΙΔΕΩΝ

Υπόεργο 6 «Τροφεία μεταπτυχιακών φοιτητών, θεμελίωση και λειτουργία Αποθετηρίου και
Επιστημονικοτεχνική Υποστήριξη του Έργου»

Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας
Πλάτωνος»

Π 6.2.3.1 Υποδομή Ψηφιακού Αποθετηρίου

Σεπτέμβριος, 2015

Ακαδημία Πλάτωνος – Ανάπτυξη της Γνώσης και καινοτόμων ιδεών
Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος»
Π 6.2.3.1 Υποδομή Ψηφιακού Αποθετηρίου

Σελίδα 2 από 12



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Εικόνα Εγγράφου

Αριθμός Παραδοτέου	6.2.3.1
Τίτλος Παραδοτέου	Υποδομή Ψηφιακού Αποθετηρίου
Αριθμός Πακέτου Εργασίας	Π.Ε 6.2.3
Τίτλος Πακέτου Εργασίας	Δημιουργία Υποδομών Ψηφιακού Αποθετηρίου
Υποέργο	6 - Τροφεία μεταπτυχιακών φοιτητών, θεμελίωση και λειτουργία Αποθετηρίου και Επιστημονικοτεχνική Υποστήριξη του Έργου
Δράση	2- Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος
Πράξη	1 - Ανάπτυξη της Γνώσης και καινοτόμων ιδεών
Μήνας Συμβατικής Παράδοσης	Σεπτέμβριος 2015 (Σταδιακή Παράδοση)
Ημερομηνία Ολοκλήρωσης	Σεπτέμβριος 2015
Κατάσταση	Τελική
Έκδοση	1
Σελίδες	12

Ακαδημία Πλάτωνος – Ανάπτυξη της Γνώσης και καινοτόμων ιδεών
Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος»
Π 6.2.3.1 Υποδομή Ψηφιακού Αποθετηρίου

Σελίδα 4 από 12



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Περιεχόμενα

Εικόνα Εγγράφου	3
Περιεχόμενα.....	5
1. Εισαγωγή	7
2. Τεχνολογία και Συνιστώσες Λογισμικού	8

Ακαδημία Πλάτωνος – Ανάπτυξη της Γνώσης και καινοτόμων ιδεών
Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος»
Π 6.2.3.1 Υποδομή Ψηφιακού Αποθετηρίου

Σελίδα 6 από 12



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

1. Εισαγωγή

Το αποθετήριο της Ακαδημίας Πλάτωνος είναι βασισμένο στο πλέον προηγμένο και ανοιχτό σύστημα αποθετηρίων Fedora Commons 3.5, πάνω από το οποίο χτίζεται μια σειρά υπηρεσιών για την άντληση και διάθεση περιεχομένου καθώς και ευκολιών για την διαχείριση και επιμέλεια του περιεχομένου αυτού.

Η εγκατάσταση και διαμόρφωση των διαφόρων συνιστωσών επιτυγχάνει την αδιάλειπτη λειτουργία του συστήματος, την διάθεση των παρεχομένων δεδομένων και μεταδεδομένων με αξιοπιστία στα, ενώ ταυτόχρονα διατηρούνται χαμηλές τιμές χρόνου απόκρισης για την παροχή εγγράφων αλλά και αναζήτησης μέσα στο κειμενικό σώμα του αποθετηρίου.

Το σύνολο του περιεχομένου του αποθετηρίου εμφανίζεται κάτω από το portal www.plato-academy.gr ενώ το διαχειριστικό μέρος βρίσκεται προστατευμένο με διαπιστευτήρια στη διεύθυνση repository.plato-academy.gr, όπου βρίσκεται και ο πυρήνας της αποθετηρίασης.

Ο πηγαίος κώδικας των συνιστωσών που αναπτύσσονται διατίθεται ως λογισμικό ανοιχτού κώδικα κάτω από την διεύθυνση:

<https://svn.madgik.di.uoa.gr/code/platoacademy/>

Ακαδημία Πλάτωνος – Ανάπτυξη της Γνώσης και καινοτόμων ιδεών
Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος»
Π 6.2.3.1 Υποδομή Ψηφιακού Αποθετηρίου

Σελίδα 8 από 12



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

2. Τεχνολογία και Συνιστώσες Λογισμικού

Τεχνολογία

Η τεχνολογία πάνω στην οποία βασίζεται το αποθετήριο της Ακαδημίας Πλάτωνος είναι κατά κύριο λόγο βασισμένη στην πλατφόρμα της Java με όλες τις απαραίτητες συνιστώσες για την παροχή webservices και HTML5 διαδικτυακών διεπαφών.

Το σύνολο των υπηρεσιών του συστήματος διατίθενται μέσω τυποποιημένων webservices.

Το σύνολο της διεπαφής του συστήματος είναι προσβάσιμο μέσω διαδικτυακής διεπαφής.

Πρότυπα

Τα πρότυπα που καλύπτει η υλοποίηση του λογισμικού υποδομής είναι τα εξής:

- OAI-PMH για την άντληση και διάθεση περιεχομένου
- SPARQL για την διερεύνηση του περιεχομένου
- XML για την αποθήκευση και διάθεση μεταδεδομένων
- JSON για την διάθεση δεδομένων και υπηρεσιών σε διαδικτυακούς πελάτες
- REST WS για την διάθεση υπηρεσιών σε τρίτους και σε διαδικτυακές διεπαφές

Συνιστώσες Λογισμικού Τρίτων

Παρακάτω αναφέρονται οι συνιστώσες λογισμικού του αποθετηρίου οι οποίες προέρχονται από την κοινότητα λογισμικού ανοιχτού κώδικα, κάτω από διάφορες άδειες χρήσεις οι οποίες είναι συμβατές με την ανοιχτού κώδικα αδειοδότηση του λογισμικού του έργου.

- [Fedora Commons 3.5](#) (Digital Repository Framework): Σύστημα αποθετηρίασης περιεχομένου (άνευ διεπαφής χρήστη και υπηρεσιών διαχείρισης του μοντέλου δεδομένων)
 - ο Link: <http://fedora-commons.org/>
- [Postgresql](#) : Σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων για την αποθήκευση μεταδεδομένων και άλλων στοιχείων του περιεχομένου, καθώς και υποστήριξη των περιφερειακών εφαρμογών.
 - ο Link: <http://www.postgresql.org/>
- [SOLR](#) Σύστημα για ευρετηρίαση και αναζήτηση δεδομένων, με πολλές δυνατότητες παραμετροποίησης και πολύ ταχύτατη απόκριση αναζήτησης.
 - ο Link: <http://lucene.apache.org/solr/>
- [Tomcat](#) (Java Application Server): εξυπηρετητής για την φιλοξενία των εφαρμογών.
 - ο Link: <http://tomcat.apache.org/>
- Tesseract OCR: Λογισμικό για την εξαγωγή κειμένου από εικόνες (OCR)
 - ο Link: <https://code.google.com/p/tesseract-ocr/>

Μια σειρά από βιβλιοθήκες χρησιμοποιούνται για την υλοποίηση των επιμέρους συνιστωσών του συστήματος:

- [Tess4J / jak1](#) Γέφυρα για την εξαγωγή πληροφοριών κειμένου από δεδομένα εικόνες μέσω του συστήματος Tesseract
 - ο Link: <http://tess4j.sourceforge.net/>
- [itextpdf](#) για την επεξεργασία αρχείων pdf
 - ο Link: <http://itextpdf.com/>
- [pdfbox](#) για την επεξεργασία και μετατροπή αρχείων pdf (πχ σε αρχεία μικρότερου μεγέθους)
 - ο Link: <https://pdfbox.apache.org/>

Τέλος, μια σειρά συνιστωσών απαιτούνται για την λειτουργία του Fedora Commons και την υλοποίηση των απαραίτητων για το έργο υπηρεσιών.

- [Mulgara](#): RDF triplestore (αποθηκεύονται οι σχέσεις μεταξύ αντικειμένων του fedora commons)
 - ο Link: <http://mulgara.org/>
- [Akubra](#) : χαμηλού επιπέδου file system για την αποθήκευση των ψηφιακών τεκμηρίων και του σώματος των μεταδεδομένων.
 - ο Link: <https://wiki.duraspace.org/display/AKUBRA/Akubra+Project>
- [Saxon](#) : επεξεργαστής XSLT και XQuery για τον μετασχηματισμό και την προβολή μεταδεδομένων του περιεχομένου
 - ο Link: <http://saxon.sourceforge.net/>
- [Apache FOP](#): βιβλιοθήκη και την μορφοποίηση XSL.
 - ο Link: <http://xmlgraphics.apache.org/fop/>

Σελίδα 10 από 12

- **ImageJ**: Υπηρεσία για τον χειρισμό και την παραγωγή εικόνων
 - Link: <http://rsbweb.nih.gov/ij/>
- **OAIProvider** για την εξυπηρέτηση (άλλων αποθετηρίων) εξαγωγής πληροφορίας από το αποθετήριο με το πρωτόκολλο διαμοιρασμού μεταδεδομένων OAI-PMH
 - Link: <https://wiki.duraspace.org/display/FCSVCS/OAI+Provider+Service+1.2.2>

Συνιστώσες Λογισμικού που Αναπτύχθηκαν

Το λογισμικό του αποθετηρίου περιλαμβάνει τις παρακάτω συνιστώσες:

- **OAI Harvester**: Συλλέγει περιεχόμενο για άντληση στο αποθετήριο, με δυνατότητες φιλτραρίσματος εξωτερικού περιεχομένου, περιοδικής εκτέλεσης, εντοπισμού του περιεχομένου κλπ
- **Custom Harvesters**: Μαζεύουν περιεχόμενο από αποθετήρια που δεν ακολουθούν το πρότυπο OAI-PMH (πχ web pages) και εντοπίζουν και τα μεταδεδομένα αλλά και το περιεχόμενο, αν διατίθεται.
- **Administration UI**: Πρόκειται για ένα πλήρες διαχειριστικό περιβάλλον για όλες τις λειτουργίες του ψηφιακού αποθετηρίου, το οποίο υλοποιείται μέσα στο Liferay Portal. Ενδεικτικά περιλαμβάνει λειτουργίες:
 - Εισαγωγής και επιμέλειας περιεχομένου
 - Διαμόρφωση μηχανισμού harvesting
 - Διαμόρφωση λειτουργίας αποθετηρίου
- **Web Front End**: Πρόκειται για διαδικτυακή διεπαφή του συστήματος η οποία αναφέρεται σε τελικούς χρήστες / επισκέπτες. Μέσω αυτών δίνονται δυνατότητες:
 - Βασικής αναζήτησης περιεχομένου
 - Προηγμένης αναζήτησης περιεχομένου
 - Προβολής μεταδεδομένων

- ο Προβολής περιεχομένου
- ο Προβολής των στατιστικών στοιχείων που υπολογίζουν άλλα υποσυστήματα.
- Backend API: Πρόκειται για διεπαφή διαδικτυακών υπηρεσιών οι οποίες διαθέτουν το περιεχόμενο προς τρίτα συστήματα αλλά και την διαδικτυακή διεπαφή του συστήματος.

Πέραν των παραπάνω υποσυστημάτων, ένας μεγάλος αριθμός μικρότερης εμβέλειας συνιστώσες έχουν υλοποιηθεί για την ολοκλήρωση των εξωτερικών συστημάτων με το αποθετήριο αλλά και για την επίτευξη επιμέρους λειτουργικών στόχων. Οι συνιστώσες αυτές μπορούν να εντοπιστούν στο αποθετήριο κώδικα του έργου:
<https://svn.madgik.di.uoa.gr/code/platoacademy/>