



ΑΚΑΔΗΜΙΑ ΠΛΑΤΩΝΟΣ – ΑΝΑΠΤΥΞΗ ΤΗΣ ΓΝΩΣΗΣ ΚΑΙ ΚΑΙΝΟΤΟΜΩΝ ΙΔΕΩΝ

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

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

Π 6.2.7.1 Λειτουργική Εικονική Υποδομή Ψηφιακού Αποθετηρίου

Οκτώβριος, 2015



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

«Ακαδημία Πλάτωνος – Ανάπτυξη της Γνώσης και καινοτόμων ιδεών»
Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος»
Π 6.2.7.1 Λειτουργική Εικονική Υποδομή Ψηφιακού Αποθετηρίου

Σελίδα 2 από 15



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Εικόνα Εγγράφου

Αριθμός Παραδοτέου	Π 6.2.7.1
Τίτλος Παραδοτέου	Λειτουργική Εικονική Υποδομή Ψηφιακού Αποθετηρίου
Αριθμός Πακέτου Εργασίας	Π.Ε 6.2.7
Τίτλος Πακέτου Εργασίας	Εγκατάσταση – Διαμόρφωση Υποδομής
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	Νοέμβριος 2015
Ημερομηνία Ολοκλήρωσης	Οκτώβριος 2015 (Παρούσα Αναφορά)
Κατάσταση	Τελική
Έκδοση	1
Σελίδες	15



«Ακαδημία Πλάτωνος – Ανάπτυξη της Γνώσης και καινοτόμων ιδεών»
Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος»
Π 6.2.7.1 Λειτουργική Εικονική Υποδομή Ψηφιακού Αποθετηρίου

Σελίδα 4 από 15



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Περιεχόμενα

Εικόνα Εγγράφου	3
Περιεχόμενα.....	5
1. Εισαγωγή	7
2. Περιγραφή Υποδομής	9
Τεχνολογία Virtualization	9
Τύποι εικονικών μηχανών	9
3. Κατάσταση Λειτουργίας.....	15

«Ακαδημία Πλάτωνος – Ανάπτυξη της Γνώσης και καινοτόμων ιδεών»
Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος»
Π 6.2.7.1 Λειτουργική Εικονική Υποδομή Ψηφιακού Αποθετηρίου

Σελίδα 6 από 15



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

1. Εισαγωγή

Το παρόν έγγραφο περιγράφει την υποδομή λειτουργίας του ψηφιακού αποθετηρίου, η οποία υλοποιείται με τεχνικές εικονικών μηχανών (server virtualization) και εξασφαλίζει την βέλτιστη δυνατή εκμετάλλευση του υλικού (hardware) που διατίθεται για την λειτουργία του ψηφιακού αποθετηρίου.

Με την τεχνική αυτή επιπλέον επιτυγχάνεται η δυνατότητα ευελιξίας στην μεταφορά του συστήματος σε άλλες συμβατές υποδομές χωρίς την μεταφορά υλικού ή την εμπλοκή σύνθετων διαδικασιών μεταφοράς τεχνολογίας.

Στο κείμενο παρουσιάζονται οι τύποι εικονικών μηχανών που δημιουργούνται και η τρέχουσα δομή της υποδομής.

«Ακαδημία Πλάτωνος – Ανάπτυξη της Γνώσης και καινοτόμων ιδεών»
Δράση 2 «Θεμελίωση, Ενίσχυση και Λειτουργία Ψηφιακού Αποθετηρίου Ακαδημίας Πλάτωνος»
Π 6.2.7.1 Λειτουργική Εικονική Υποδομή Ψηφιακού Αποθετηρίου

Σελίδα 8 από 15



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

2. Περιγραφή Υποδομής

Τεχνολογία Virtualization

Η τεχνολογία virtualization χρησιμοποιείται είναι αυτή του Microsoft Hyper-V Server 2012R2, ο οποίος διατίθεται δωρεάν από την Microsoft. Η συγκεκριμένη τεχνολογία, προσφέρει με πάρα πολύ οικονομικό και συνάμα τρόπο την δυνατότητα εκμετάλλευσης και διαχείρισης της υποδομής μέσα από υφιστάμενα συστήματα του χώρου εγκατάστασης των φυσικών εξυπηρετητών. Λόγω της συμβατότητας με υφιστάμενα συστήματα δημιουργούνται οι προοπτικές εκμετάλλευσης υφιστάμενων πόρων σε περίπτωση ανάγκης (φόρτου εργασιών ή αστοχίας υλικού).

Τύποι εικονικών μηχανών

Το αποθετήριο διαθέτει πέντε τύπους εικονικών μηχανών (VMs), οι οποίο περιγράφονται παρακάτω.

Τύπος 1 - repository

Ο πρώτος τύπος εικονικών μηχανών εξυπηρετεί όλες τις λειτουργίες που χρειάζεται για να λειτουργήσει το αποθετήριο, όσον αφορά το διαχειριστικό κομμάτι, το κομμάτι όπου εξυπηρετεί άλλα αποθετήρια για harvesting καθώς επίσης παρέχει όλες τις ζητούμενες πληροφορίες στις εικονικές μηχανές τύπου 2. Ενδεικτικά το repository.plato-academy.gr είναι μια τέτοια εικονική μηχανή.

Λειτουργικό σύστημα: [DebianGNU/Linux 6.0.9 \(Squeeze\)](#)

Η τελευταία έκδοση του πιο σύγχρονου και ευρέως χρησιμοποιούμενου δωρεάν λειτουργικού συστήματος για μηχανήματα εξυπηρετητή. Είναι εγκατεστημένες όλες οι τελευταίες ενημερώσεις ασφαλείας και λειτουργικότητας που παρέχονται.

Βιβλιοθήκες και Frameworks:

- [Java 1.6.0.26](#), σύγχρονη και ευρέως διαδεδομένη γλώσσα προγραμματισμού για εφαρμογές διαδικτύου

- [GWT 2.5.1](#), ένα framework από τη Google για δημιουργία σύνθετων εφαρμογών που τρέχουν χρησιμοποιώντας Javascript στο πρόγραμμα περιήγησης ενώ διαθέτει, επίσης, μηχανισμό επικοινωνίας με τον server.
- [Fedora Commons 3.7](#), το ισχυρότερο και ευρέως διαδεδομένο σύστημα αποθετηρίων, παρέχει πληθώρα λειτουργιών, ενώ ανταποκρίνεται με ικανοποιητική ταχύτητα σε δύσκολες περιπτώσεις αποθετηρίων με εκατομμύρια εγγραφών.
- [Apache SOLR 4.2.0](#), σύστημα αναζήτησης με πολλές λειτουργίες από την Apache. Λειτουργεί ταχύτατα με εκατομμύρια εγγραφές. Συνδυάζεται με το Fedora ώστε να παρέχεται παράλληλα ταχύτητα και αξιοπιστία στο αποθετήριο.
- [Tematres](#), σύστημα διαχείρισης λεξικών, οντολογιών και ταξινομιών, χρησιμοποιείται για την ορθή προβολή και αντικατάσταση όρων όπου αυτό κρίνεται απαραίτητο. Εφαρμογή του έγινε στα ονόματα συγγραφέων όπου υπήρχαν παραλλαγές (ανάλογα με τον εκδότη) στο πως γράφεται ένα όνομα, καθώς και στις ταξινομίες των εγγράφων όπου ισοδύναμοι όροι αντικαθίστανται από τον επικρατέστερο.
- [Tess4j](#), Java βιβλιοθήκη για την επικοινωνία με το TesseractOCR. Χρησιμοποιείται για την οπτική αναγνώριση κειμένου μέσα σε εικόνες. Γίνεται εξόρυξη του κειμένου μέσα από τις εικόνες που περιέχονται στα PDF.
- [Itextpdf](#) και [pdfbox](#), Java βιβλιοθήκες για την επεξεργασία (εξαγωγή περιεχομένου) και / ή μετατροπή αρχείων pdf σε μικρότερου μεγέθους (πχ συμπίεση εικόνων)

Εξυπηρετητής και βάση δεδομένων:

- [Apache Tomcat 6](#), σύγχρονος και διαδεδομένος εξυπηρετητής, λειτουργεί άψογα με τη Java 1.6 και το Debian λειτουργικό. Προσφέρει ταχύτητα και σταθερότητα στις υπηρεσίες του συστήματος.

[PostgreSQL 9.4](#), η τελευταία έκδοση της διαδεδομένης βάσης δεδομένων PostgreSQL, χρησιμοποιείται για αποθήκευση και ανάκτηση πληροφοριών

Σελίδα 10 από 15

γεωδεδομένων, καθώς επίσης και εσωτερικών πληροφοριών διαχείρισης του Fedora Commons.

Τύπος 2: front-end

Ο δεύτερος τύπος VM χρησιμοποιείται για την προβολή των πληροφοριών. Εκεί βρίσκεται η διεπαφή του αποθετηρίου (<http://www.plato-academy.gr/repository-app>).

Για τις ανάγκες του αποθετηρίου το παραπάνω vm έχει το ακόλουθο setup/configuration.

Λειτουργικό σύστημα: [DebianGNU/Linux 6.0.9 \(Squeeze\)](#)

Η τελευταία έκδοση του πιο σύγχρονου και ευρέως χρησιμοποιούμενου δωρεάν λειτουργικού συστήματος για μηχανήματα εξυπηρετητή. Είναι εγκατεστημένες όλες οι τελευταίες ενημερώσεις ασφαλείας και λειτουργικότητας που παρέχονται.

Βιβλιοθήκες και Frameworks:

- [Java 1.7.0.51](#), σύγχρονη και ευρέως διαδεδομένη γλώσσα προγραμματισμού για εφαρμογές διαδικτύου
- [GWT 2.5.1](#), ένα framework από τη Google για δημιουργία σύνθετων εφαρμογών που τρέχουν χρησιμοποιώντας Javascript στο πρόγραμμα περιήγησης ενώ διαθέτει επίσης μηχανισμό επικοινωνίας με το server.
- [Liferay 6.1.1](#), λογισμικό δημιουργίας ιστοσελίδων και εφαρμογών, βασισμένο σε Java, υποστηρίζει πληθώρα εφαρμογών, διαθέτει δωρεάν εφαρμογές προς χρήση, είναι φιλικό προς τον προγραμματιστή για τη δημιουργία προσωποποιημένων και εξειδικευμένων εφαρμογών και έχει μεγάλη κοινότητα υποστήριξης. Χρησιμοποιείται για την προβολή της βασική ιστοσελίδας της Ακαδημίας Πλάτωνος (www.plato-academy.gr), φιλοξενεί επίσης την ειδικά διαμορφωμένη εφαρμογή του αποθετηρίου (www.plato-academy.gr/repository-app)

Εξυπηρετητής εφαρμογών

- [ApacheTomcat 6](#), σύγχρονος και διαδεδομένος εξυπηρετητής, λειτουργεί άψογα με τη Java 1.7 και το Debian λειτουργικό. Προσφέρει ταχύτητα και σταθερότητα στις υπηρεσίες του συστήματος.

Τύπος 3: community

Ο τρίτος τύπος εικονικών μηχανών χρησιμοποιείται, επίσης για την προβολή των τεκμηρίων αλλά παράλληλα λειτουργεί ως πλατφόρμα συνεργασίας, όπου οι χρήστες μπορούν να αποθηκεύσουν τα τεκμήρια που τους ενδιαφέρουν, να σχολιάσουν πάνω σε αυτά, καθώς και να κοινοποιήσουν συνδέσμους προς άλλους χρήστες της κοινότητας. Το [community.plato-academy.gr](#) αποτελεί τέτοιο χώρο.

Για τις ανάγκες του αποθετηρίου το παραπάνω vm έχει το ακόλουθο setup/configuration.

Λειτουργικό σύστημα: [DebianGNU/Linux 6.0.9 \(Squeeze\)](#)

Η τελευταία έκδοση του πιο σύγχρονου και ευρέως χρησιμοποιούμενου δωρεάν λειτουργικού συστήματος για μηχανήματα εξυπηρετητή. Είναι εγκατεστημένες όλες οι τελευταίες ενημερώσεις ασφαλείας και λειτουργικότητας που παρέχονται.

Βιβλιοθήκες και Frameworks:

- [Java 1.7.0.51](#), σύγχρονη και ευρέως διαδεδομένη γλώσσα προγραμματισμού για εφαρμογές διαδικτύου
- [GWT 2.5.1](#), ένα framework από τη Google για δημιουργία σύνθετων εφαρμογών που τρέχουν χρησιμοποιώντας Javascript στο πρόγραμμα περιήγησης ενώ διαθέτει επίσης μηχανισμό επικοινωνίας με το server.
- [Liferay6.2CE](#), λογισμικό δημιουργίας ιστοσελίδων και εφαρμογών, βασισμένο σε Java, υποστηρίζει πληθώρα εφαρμογών, διαθέτει δωρεάν εφαρμογές προς χρήση, είναι φιλικό προς τον προγραμματιστή για τη δημιουργία προσωποποιημένων και εξειδικευμένων εφαρμογών και έχει μεγάλη κοινότητα υποστήριξης. Εξυπηρετητής και βάση δεδομένων:

- [ApacheTomcat7](#), σύγχρονος και διαδεδομένος εξυπηρετητής, λειτουργεί άψογα με τη Java 1.7 και το Debian λειτουργικό. Προσφέρει ταχύτητα και σταθερότητα στις υπηρεσίες του συστήματος.
- [PostgreSQL 9.4](#), η τελευταία έκδοση του συστήματος διαχείρισης σχεσιακών βάσεων δεδομένων Postgresql, χρησιμοποιείται για αποθήκευση πληροφοριών από το λογισμικό Liferay
- [annotatorJS](#), βιβλιοθήκη Javascript για τη δημιουργία annotations πάνω σε htmlstructures. Χρησιμοποιείται για την δημιουργία και επεξεργασία επισημάνσεων πάνω σε σελίδες Html
- [pdf2htmlEX](#) - ένα εργαλείο για τη μετατροπή PDF αρχείων σε css styled html αρχεία. Χρησιμοποιείται για τη μετατροπή Pdf αρχείων σε html και έτσι μπορούν να γίνουν πάνω τους επισημάνσεις.

Λοιπές βιβλιοθήκες

- [restdeasy-jaxrs](#), βιβλιοθήκη για τη δημιουργία restful interface για την επικοινωνία του annotatorjs με το backend (δημιουργία και χρήση tokens, μετασχηματισμό των annotations).
- [java-jwt](#), για τη παραγωγή JSON Web Tokens, χρησιμοποιείται από το annotatorjs.

Τύπος 4: storage

Πρόκειται για τύπο εικονικής μηχανής ο οποίος δίνει υπηρεσίες αποθήκευσης σε όλες τις άλλες μηχανές, με δυνατότητα συστοιχίας. Οι συγκεκριμένες εικονικές μηχανές δεν έχουν δημόσια διεύθυνση, για λόγους ασφαλείας.

Λειτουργικό σύστημα: [DebianGNU/Linux 6.0.9 \(Squeeze\)](#)

Επιπλέον φέρει:

- [PostgreSQL 9.4](#), η τελευταία έκδοση του συστήματος διαχείρισης σχεσιακών βάσεων δεδομένωνPostgresql, χρησιμοποιείται για αποθήκευση πληροφοριών από διάφορα επιμέρους υποσυστήματα.
- [GlusterFS](#), ένα εξαιρετικά κλιμακούμενο καταναμημένο σύστημα αρχείων.

Σελίδα 13 από 15

Τύπος 5: testing&development

Πρόκειται για τύπο εικονικής μηχανής ο οποίος διατίθεται για ελέγχους του συστήματος και για την ανάπτυξη επιμέρους συνιστωσών. Πέραν του λειτουργικού συστήματος, όλες οι άλλες συνιστώσες διαμορφώνονται κατά περίπτωση.

Οι συγκεκριμένες εικονικές μηχανές αποκοτούν δημόσια διεύθυνση κατά περίπτωση.

Λειτουργικό σύστημα: [DebianGNU/Linux 6.0.9 \(Squeeze\)](#)

3. Κατάσταση Λειτουργίας

Στην τρέχουσα κατάσταση λειτουργίας διατίθενται στους χρήστες οι εξής εικονικές μηχανές:

- Τύπου 1, 4 vCores, 8GB RAM, 80GB Storage: 1 instance
- Τύπου2, 4 vCores, 8GB RAM, 80GB Storage: 2 instances
- Τύπου3, 4 vCores, 8GB RAM, 80GB Storage: 2 instances
- Τύπου 4, 4 vCores, 8GBRAM, 120GBStorage: 2 instances (εγκατεστημένες σε διαφορετικόφυσικό εξυπηρετητή)
- Τύπου 5, 4 vCores, 8GBRAM, 30GBStorage: 3έως 10 instances αναλόγως απαιτήσεωνdevelopment / testing