

ΠΕΡΙΛΗΨΗ

Αρχειοθέτηση ιστοπεριεχομένου και διατήρηση ψηφιακής μνήμης: η εμπειρία του ΟΠΑ”.

Μιχάλης Βαζιργιάννης

Η αναγκαιότητα της αρχειοθέτησης του περιεχομένου των ιστότοπων έχει αναγνωριστεί από το ξεκίνημα του Παγκόσμιου Ιστού, λόγω της ευκολίας με την οποία μπορεί να υπάρχει απώλεια της διαθέσιμης πληροφορίας. Στα πλαίσια του έργου ψηφιακής σύγκλισης αναπτύχθηκε σύστημα το οποίο επεξεργάζεται σε μηνιαία βάση όλο το περιεχόμενο των ιστότοπων του Οικονομικού Πανεπιστημίου Αθηνών. Το σύστημα αρχειοθέτησης δίνει την δυνατότητα στη Βιβλιοθήκη του Οικονομικού Πανεπιστημίου Αθηνών για την υλοποίηση ενός από τους βασικούς στόχους, που είναι η αρχειοθέτηση και διατήρηση του ψηφιακού περιεχομένου των ιστοσελίδων του πανεπιστημίου.

Τα πιο σημαντικά οφέλη για τη Βιβλιοθήκη που λειτουργεί ως κεντρικός κόμβος πληροφόρησης του πανεπιστημίου είναι τα ακόλουθα:

- Διατήρηση της ιστορίας του ινστιτούτου.
- Η χρήση ιστορικών δεδομένων για μελλοντική έρευνα
- Διατήρηση και προστασία της πολιτιστικής κληρονομιάς και της επιστημονικής παραγωγής

Το περιεχόμενό αποθηκεύεται και στη συνέχεια ευρετηριοποιείται τόσο με βάση το χρόνο όσο και με βάση το περιεχόμενο (γλωσσική ευρετηρίαση). Έτσι ο χρήστης μπορεί να αναζητήσει εύκολα το περιεχόμενο που τον ενδιαφέρει και να παρακολουθήσει πως αυτό εξελίχθηκε στον χρονικό διάστημα που τον ενδιαφέρει. Για παράδειγμα κάποιος θα μπορούσε να αναζητήσει το πρόγραμμα σπουδών και να μελετήσει την εξέλιξή του και κατά πόσο το πρόγραμμα μεταβάλλεται ανάλογα με τις ανάγκες της εποχής ή οικονομικά στοιχεία ώστε να δει πόσο επηρέασε η κρίση και πώς τη λειτουργία του Πανεπιστημίου. Η σημασία της αρχειοθέτησης είναι εμφανής καθώς μεγάλο μέρος του αρχειοθετημένου περιεχομένου δεν είναι διαθέσιμο στη σημερινή έκδοση των ιστοσελίδων.

Το σύστημα συλλογής ιστοπεριεχομένου επισκέπτεται κάθε μήνα περίπου 500.000 συνδέσμους από 38 κύρια domain. Το συνολικό μέγεθος της πληροφορίας που έχει αρχειοθετηθεί σε συμπιεσμένη μορφή αυτή τη στιγμή είναι περίπου 850GB. Αυτή τη στιγμή η συλλογή περιέχει υλικό από τον Φεβρουάριο του 2010 έως και σήμερα. Η αρχιτεκτονική του συστήματος φαίνεται σε παρακάτω σχήμα και βασίζεται σε 3 βασικά υποσυστήματα για την αποτελεσματική ευρετηρίαση και ανακατασκευή του περιεχομένου. Ακόμα χρησιμοποιείται αλγόριθμος κατάταξης για την πιο αποτελεσματική αναζήτηση του περιεχομένου.



Η αρχειοθέτηση ιστοπεριεχομένου μπορεί να πραγματοποιηθεί τόσο σε παγκόσμιο επίπεδο, όσο και σε εθνικό και τοπικό επίπεδο. Η εμπειρία μας στην αρχειοθέτηση των ιστοσελίδων του ΟΠΑ δείχνει ότι είναι δυνατόν να αρχειοθετηθεί το ιστοπεριεχόμενο μεγάλων ιδρυμάτων με ένα βιώσιμο τρόπο χρησιμοποιώντας την υπάρχουσα τεχνολογία και η αποτελεσματική ανάκτηση των διαθέσιμων πληροφοριών.