**Μελέτη Περίπτωσης: Ανάκτηση και αξιοποίηση των ανοικτών δεδομένων - Η περίπτωση των ανοιχτών συνδεδεμένων δεδομένων του Πρώτου Παγκοσμίου Πολέμου**

Η παρούσα εργασία εκπονήθηκε στα πλαίσια του μαθήματος «Τεχνολογίες Ιστού και Ανάλυση Δεδομένων Ιστού» της 4ης κατεύθυνσης του μεταπτυχιακού Εφαρμοσμένης Πληροφορικής με υπεύθυνο καθηγητή τον κ. Ταμπούρη Ευθύμιο από τους φοιτητές Καρακατσάνη Χρήστο και Καράμπελα Αναστασία. Σκοπός της εργασίας είναι να τονίσει τη χρησιμότητα των ανοιχτών συνδεδεμένων δεδομένων (LinkedOpenData), μέσα από την παρουσίαση και αξιοποίηση της πύλης ανοιχτών δεδομένων WorldWar 1 as Linked Open Data της σχεσιακής βάσης Datahub. Τα δεδομένα ανακτήθηκαν από την βάση Datahub με την γλώσσα SPARQ. Στο πρώτο σκέλος της εργασίας γίνεται η παρουσίαση της βάσης, οι οποία περιλαμβάνει δεδομένα που αφορούν Γεγονότα, Περιοχές, Πρόσωπα, Ημερομηνίες, Λέξεις κλειδιά και Θέματα του Πρώτου Παγκοσμίου Πολέμου. Βασικές πηγές άντλησης δεδομένων είναι το Imperial War Museum (IWM) First World War Centenary Partnership, the British Army’s Battle Nomenclatures Committee καθώς και πηγή με πληροφορίες που αφορούν τις θηριωδίες των γερμανικών στρατευμάτων στο Βέλγιο το 1914. Στο δεύτερο σκέλος γίνεται μια σύντομη περιγραφή της Πύλης και των δεδομένων της. Η Πύλη περιλαμβάνει 61,307 τριπλέτες, 38,188 RDF Κόμβους και συνολικά 72 κλάσεις. Ακολουθεί η παρουσίαση του μοντέλου στο οποίο βασίστηκαν τα SPARQL ερωτήματα που χρησιμοποιήθηκαν για την άντληση της χρήσιμης πληροφορίας με όλα τα αντικείμενα/υποκείμενα, τις κλάσεις και τις ιδιότητες που τα συνδέουν. Στη συνέχεια παραθέτουμε τα 8 ερωτήματα SQARQL μαζί με την αντίστοιχη περιγραφή των αποτελεσμάτων που επιστέφουν. Βασιζόμενοι στα αποτελέσματα των παραπάνω SPΑRQL ερωτημάτων, προχωρήσαμε σε οπτικοποιήσεις των δεδομένων με τη χρήση των εργαλείων Sgvizler και Exploratory οι οποίες μας βοήθησαν να καταλήξουμε σ χρήσιμα συμπεράσματα σχετικά με τα δεδομένα μας. Χαρακτηριστικό παράδειγμα αποτελεί η κατασκευή ενός χάρτη στον οποίον απεικονίζονται οι θηριωδίες του πεζικού της Γερμανικής Στρατιάς, το όνομα του κάθε συντάγματος και οι περιοχές που έλαβαν μέρος οι θηριωδίες από τον οποίο μπορούμε να δούμε τις πόλεις που υπέστησαν τις περισσότερες καταστροφές. Τέλος, ακολουθούν στατιστικές αναλύσεις τις οποίες πραγματοποιήσαμε στα αποτελέσματα των SPARQL ερωτημάτων με το εργαλείο Exploratory. Οι αναλύσεις αυτές αφορούν ομαδοποιήσεις στα δεδομένα μας με τη χρήση του αλγορίθμου k-means clustering για k=3. Παραδείγματα αποτελούν η ομαδοποίηση περιοχών με βάση τον αριθμό των στρατιωτικών μονάδων της 4ης γερμανικής στρατιάς που έλαβαν μέρος στις μάχες που πραγματοποιήθηκαν εκεί, καθώς και η ομαδοποίηση των θηριωδιών ανάλογα με το
μέγεθος τους. Ελπίζουμε η παρούσα εργασία να επιτελέσει τον σκοπό της και να παρουσιάσει με παραστατικό τρόπο την χρησιμότητα των ανοικτών διασυνδεδεμένων δεδομένων ώστε μελλοντικοί επαγγελματίες και ερευνητές να κάνουν πλήρη χρήση των δυνατοτήτων τους.