



Ε. Γαλιώτου

Ερευνητικό έργο
“ΠΟΛΥΤΙΜΟ”

7

ΕΡΕΥΝΑ
ΣΥΝΕΡΓΑΣΙΕΣ

Ερευνητικό έργο “ΠΟΛΥΤΙΜΟ”



της Ε. Γαλιώτου*



**Σύστημα Επεξεργασίας,
Διαχείρισης και
Παροχής Πρόσβασης
στο Περιεχόμενο
Πολύτιμων Βιβλίων και
Χειρογράφων της Ιεράς
Μονής Ευαγγελισμού
της Θεοτόκου Σκιάθου.**

Ο στόχος του ερευνητικού έργου “ΠΟΛΥΤΙΜΟ” ήταν η χρησιμοποίηση καινοτόμων τεχνικών επεξεργασίας εικόνας, αναγνώρισης προτύπων και επεξεργασίας φυσικής γλώσσας για την ανάπτυξη ενός πρωτοποριακού πληροφοριακού συστήματος για την διαχείριση και παροχή πρόσβασης στο περιεχόμενο της συλλογής βιβλίων και χειρογράφων της Ιεράς Μονής Ευαγγελισμού της Θεοτόκου Σκιάθου. Ιστορική αυτή Μονή ιδρύθηκε το 1794 και στη βιβλιοθήκη της βρίσκονται -μεταξύ άλλων - έντυπα βιβλία του 16^{ου} έως και του 19^{ου} αιώνας, χειρόγραφα των πέντε τελευταίων αιώνων, πατριαρχικά μολυβδόβουλα, σιγίλια και επιστολές.

Το έργο χρηματοδοτήθηκε από την Γενική Γραμματεία Έρευνας και Τεχνολογίας στα πλαίσια του επιχειρησιακού προγράμματος “Κοινωνία της Πληροφορίας” (πρόγραμμα: “Επεξεργασία Εικόνων, Ήχου και Γλώσσας”) και υλοποιήθηκε από κοινοπραξία που απαρτιζόταν από την εταιρεία BSI A.E. (ανάδοχος του έργου), το Εργαστήριο Υπολογιστικής Ευφυΐας του Ε.Κ.Ε.Φ.Ε. “Δημόκριτος” και το Τμήμα Πληροφορικής του ΤΕΙ Αθήνας.

Συνολικός προϋπολογισμός του Έργου

559.595,00 €

Προϋπολογισμός για το τμ. Πληροφορικής

65.000,00 €

Επιστημονική υπεύθυνη του Έργου

Ε. Γαλιώτου,

επικ. καθηγήτρια του τμήματος Πληροφορικής

Μέλη της ερευνητικής ομάδας

από το τμήμα Πληροφορικής

Χ. Σκουρλάς, καθηγητής

Ν. Βασιλάς, καθηγητής

Γ. Πάντζιου, καθηγήτρια

Συμμετείχαν επίσης οι κ. κ. Ι. Μανωλέσσου

και Α. Ράλλη από το Τμήμα Φιλολογίας του

Πανεπιστημίου Πατρών καθώς και οι υποψήφιοι

διδάκτορες Πληροφορικής Α. Λαμπρόπουλος,

Δ. Σωτηρόπουλος και Γ. Ζήκος.

Διάρκεια του Έργου

3 Ιουλίου 2006 έως 31 Μαρτίου 2008

Σκοπός του Έργου

Η ανάπτυξη ενός καινοτόμου συστήματος

*Η κ. Ε.Γαλιώτου είναι αναπληρώτρια καθηγήτρια στο Τμήμα Πληροφορικής

ανοιχτής αρχιτεκτονικής για την επεξεργασία, διαχείριση και παροχή πρόσβασης στο περιεχόμενο πολύτιμων βιβλίων και χειρογράφων το οποίο βασίζεται στο ψηφιοποιημένο ιστορικό υλικό της Ιεράς Μονής Ευαγγελισμού της Θεοτόκου Σκιάθου, αλλά μπορεί να επεκταθεί και σε αντίστοιχες ιστορικές συλλογές βιβλίων και χειρογράφων άλλων φορέων στην Ελλάδα και στο εξωτερικό. Η συγκεκριμένη Μονή είχε ήδη προχωρήσει στην ψηφιοποίηση μεγάλου μέρους της συλλογής της (80 ψηφιοποιημένα βιβλία και χειρόγραφα τα οποία αντιστοιχούν σε 52.000 σελίδες -μερικά εκ των οποίων είναι δίγλωσσα: Ελληνικά και Λατινικά-) με τη βοήθεια του Τμήματος Πληροφορικής.

Στα πλαίσια του έργου, χρησιμοποιήθηκαν καινοτόμες τεχνικές επεξεργασίας εικόνας, αναγνώρισης προτύπων και επεξεργασίας γλώσσας για την ανάπτυξη ενός πρωτοποριακού πληροφοριακού συστήματος για την διαχείριση και παροχή πρόσβασης στο περιεχόμενο της συλλογής βιβλίων και χειρογράφων της Μονής καθώς και την ανάπτυξη περιβάλλοντος ηλεκτρονικής πλοήγησης στο υλικό.

Το σύστημα δίνει - μεταξύ άλλων - τη δυνατότητα στους χρήστες:

- Να βελτιώσουν την ποιότητα του ψηφιοποιημένου βιβλίου ή χειρογράφου και να αναδείξουν τις περιοχές του κειμένου. Με τον τρόπο αυτό παρακάμπτονται οι δυσκολίες πρόσβασης στο περιεχόμενο των ιστορικών εγγράφων που οφείλονται στην παλαιότητά τους.
- Να αναζητήσουν λέξεις απευθείας στο περιεχόμενο των ψηφιοποιημένων εγγράφων παρακάμπτοντας συμβατικές μεθόδους οπτικής αναγνώρισης χαρακτήρων που αποδεικνύονται αναποτελεσματικές στην περίπτωση των ιστορικών εγγράφων της Μονής.
- Να χρησιμοποιήσουν προηγμένες τεχνικές επεξεργασίας φυσικής γλώσσας για την πραγματοποίηση "έξυπνων" αναζητήσεων.

➤ Να εξάγουν το πλήρες κείμενο από ψηφιοποιημένα δίγλωσσα εκκλησιαστικά κείμενα (Ελληνικά - Λατινικά) για περαιτέρω μελέτη.

Στο σχήμα 1 φαίνεται η δομή του συστήματος «ΠΟΛΥΤΙΜΟ».

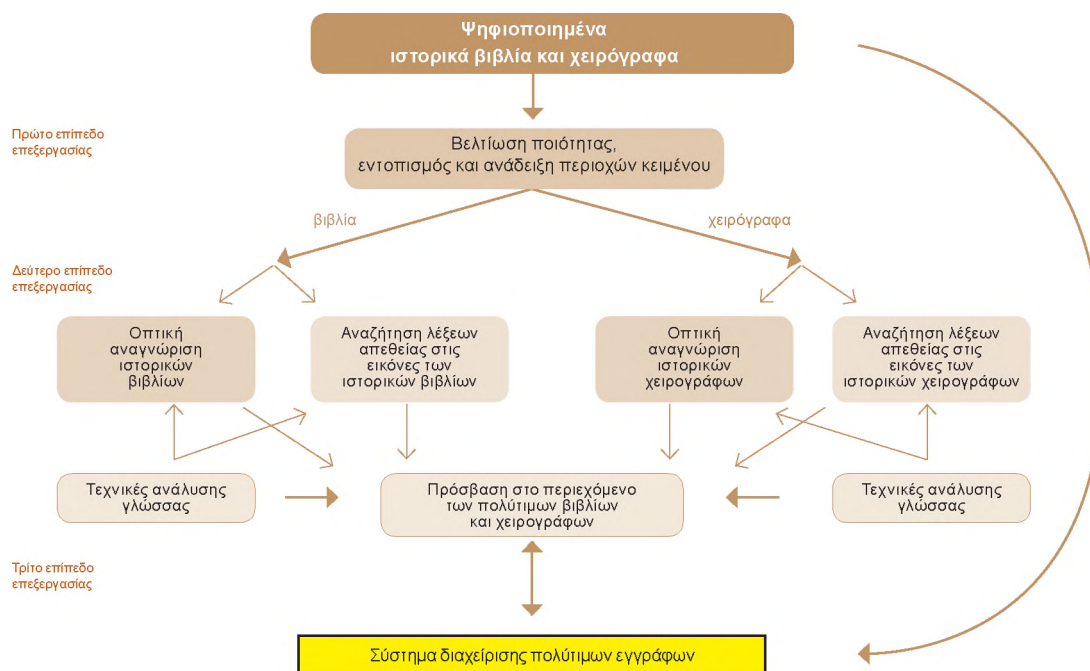
Η μεθοδολογία υλοποίησης του έργου αφορά τέσσερις διακριτές φάσεις και συγκεκριμένα:

- Ανάλυση απαιτήσεων χρηστών και προδιαγραφών συστήματος
- Ανάπτυξη εργαλείων επεξεργασίας και παροχής πρόσβασης στο περιεχόμενο των ιστορικών βιβλίων και χειρογράφων
- Ανάπτυξη περιβάλλοντος διαχείρισης ιστορικών εγγράφων, ολοκλήρωση και έλεγχος του συστήματος
- Διάχυση και εκμετάλλευση των αποτελεσμάτων

Οι φάσεις αυτές κατανεμήθηκαν στις ακόλουθες οκτώ ενότητες εργασίας:

- EE1:** Απαιτήσεις και προδιαγραφές
- EE2:** Βελτίωση της ποιότητας και ανάδειξη περιοχών κειμένου
- EE3:** Οπτική αναγνώριση ιστορικών τυπωμένων κειμένων
- EE4:** Οπτική αναγνώριση ιστορικών χειρογράφων
- EE5:** Αναζήτηση λέξεων σε ιστορικά βιβλία και χειρόγραφα
- EE6:** Βελτίωση με χρήση μεθόδων επεξεργασίας γλώσσας
- EE7:** Ανάπτυξη περιβάλλοντος διαχείρισης ιστορικών εγγράφων - ολοκλήρωση και έλεγχος του συστήματος
- EE8:** Διάχυση και εκμετάλλευση των αποτελεσμάτων

Στα πλαίσια του έργου, η ομάδα του ΤΕΙ Αθήνας - εκτός από τη συμμετοχή της και σε άλλες ενότητες εργασίας - ήταν ο υπεύθυνος φορέας



Σχήμα 1: Δομή συστήματος "ΠΟΛΥΤΙΜΟ"

υλοποίησης της ενότητας εργασίας με τίτλο "Βελτίωση με Χρήση Μεθόδων Επεξεργασίας Γλώσσας". Η ενότητα εργασίας αυτή, είχε στόχο τη βελτίωση της διαδικασίας της αναζήτησης και κινήθηκε γύρω από τέσσερις άξονες:

1. Δημιουργία καταλόγων λέξεων για τη βελτίωση της διαδικασίας του εντοπισμού των λέξεων.
2. Δημιουργία γεννήτριας λέξεων με στόχο τον εντοπισμό όλων των κλιτών μορφών στο κείμενο.
3. Δημιουργία και χρήση λεξικού συνωνύμων για αποτελεσματικότερη πρόσβαση στο σημασιολογικό περιεχόμενο των ιστορικών εγγράφων.
4. Αναζήτηση σε δίγλωσσα ιστορικά κείμενα (Ελληνικά - Λατινικά).

Τα εργαλεία και οι μεθοδολογίες που έχουν αναπτυχθεί στα πλαίσια της ενότητας εργασίας αυτής από την ομάδα του ΤΕΙ Αθήνας, αναμένεται να αξιοποιηθούν και εκτός των αυστηρών

πλαισίων του έργου. Πιο συγκεκριμένα:

➤ Η γεννήτρια λέξεων που χαρακτηρίζουν τα βιβλία που απαρτίζουν το σώμα κειμένων είναι το πρώτο - από όσο είμαστε σε θέση να γνωρίζουμε - σύστημα αυτόματης μορφολογικής επεξεργασίας της γλώσσας της περιόδου αυτής. Παρουσιάζει λοιπόν ενδιαφέρον τόσο από την άποψη της μεθοδολογίας όσο και από την άποψη της μελέτης της Ελληνικής γλώσσας της εποχής. Για τον λόγο αυτό, μπορεί να επεκταθεί και να συμβάλει στην αναζήτηση και άλλων ιστορικών κειμένων αλλά και να αποτελέσει εργαλείο που θα διευκολύνει το έργο των μελετητών της διαχρονικής εξέλιξης της Ελληνικής γλώσσας.

➤ Η επέκταση του λεξικού συνωνύμων που στα πλαίσια του έργου καλύπτει λέξεις που χαρακτηρίζουν το σώμα κειμένων, μπορεί να συμβάλει στην αναζήτηση σε ιστορικά κείμενα με διαφορετική θεματική αλλά και στη δημιουργία ηλεκτρονικών γλωσσικών πόρων χρήσιμων στη μελέτη της διαχρονικής εξέλιξης της Ελληνικής γλώσσας.

➤ Τέλος, η επέκταση του συστήματος ευθυγράμμισης δίγλωσσων κειμένων και αναζήτησης σ' αυτά αναμένεται να συμβάλλει στη μελέτη εκκλησιαστικών κειμένων και στη δημιουργία αντίστοιχων ηλεκτρονικών γλωσσικών πόρων. Το τριήμερο 6 έως 8 Μαρτίου 2008 εκπρόσωποι των τριών φορέων της κοινοπραξίας (BSI A.E., Ε.Κ.Ε.Φ.Ε «Δημόκριτος», ΤΕΙ Αθήνας) ταξίδεψαν στη Σκιάθο για να εγκαταστήσουν το "ΠΟΛΥΤΙΜΟ" στην Ι.Μ. Ευαγγελίστριας. Το σύστημα εγκαταστάθηκε στη βιβλιοθήκη της Μονής και απέσπασε πολύ θετικά σχόλια κατά την επίδειξή του στους παρισταμένους και εν δυνάμει χρήστες του. Έκτοτε είναι στη διάθεση κάθε ενδιαφερομένου επισκέπτη της Μονής. Ο τελικός έλεγχος του έργου «ΠΟΛΥΤΙΜΟ» έγινε στις 9 Σεπτεμβρίου 2008 από εκπροσώπους της ΓΓΕΤ οι οποίοι διαπίστωσαν την πλήρη επίτευξη των στόχων του έργου μέσω της συνεργασίας των φορέων της σύμπραξης, η οποία υπήρξε εξαιρετικά αρμονική και γόνιμη. ☑

Αναφορές

- [1] A. Lampropoulos, E. Galiotou, I. Manolessou, A. Ralli, "A Finite-State Approach to the Computational Morphology of Early Modern Greek", Proceedings of the 7th IEEE International Conference on Applied Computer Science (ACS'07), pp. 242-245, Venice, Italy, 2007.
- [2] A. Lampropoulos, D. Sotiropoulos, G. Zikos, I. Manolessou, A. Ralli, C. Skourlas, E. Galiotou, "Improving the access to the content of Greek historical documents", Proceedings of the 2nd International Scientific Conference eRA, Piraeus, Greece, 2007.
- [3] D. Sotiropoulos, E. Galiotou, C. Skourlas, "Application of a Word-Alignment Algorithm to Bilingual Greek-Latin Documents", Proceedings of the 7th IEEE International Conference on Applied Computer Science (ACS'07), 238-241, Venice, Italy, 2007.
- [4] D. Turcato, F. Popowich, J. Toole, D. Fass, D. Nicholson, D. Tisher, "Adapting a synonym database to specific domains", J. Klavans and J. Gonzalo J. (eds.) Proceedings of the ACL Workshop on Recent Advances in Natural Language Processing and Information Retrieval, pp. 1-11, 2000.
- [5] H. Schmid, "A Programming Language for Finite State Transducers", Proc. FSMNLP 2005, Helsinki, Finland, 2005.
- [6] P. Fung, K. Church, "K - vec A new approach for Aligning Parallel texts", Proc. 15th International Conference on Computational linguistics, Kyoto, Japan, 1994, pp. 1096 - 1102.



Το Ιατρείο του ΤΕΙ Αθήνας λειτουργεί καθημερινά

Προσωπικό:

Πρωί: 8.00 π.μ. έως 2.30 μ.μ.

Παναγιώτης- Κωνσταντίνος Χάικος, Καρδιολόγος

Παρασκευή Κόρρα - Μανδελιά,
Νοσηλεύτρια - Επισκέπτρια Υγείας

Απόγευμα: 2.30 μ.μ. έως 5.30 μ.μ.

Δευτέρα και Τρίτη

Νικόλαος Ζιώρης, Χειρουργός

Τετάρτη

Πάυλος Πετρίδης, Ορθοπαιδικός

Πέμπτη

Νικηφόρος Κλήμης, Γυναικολόγος

Δευτέρα έως Παρασκευή:

Κωνσταντίνος Μπούρας, Νοσηλευτής