

**ΑΝΑΛΥΤΙΚΟ ΥΠΟΜΝΗΜΑ
ΕΠΙΣΤΗΜΟΝΙΚΩΝ ΔΗΜΟΣΙΕΥΣΕΩΝ**

ΛΕΩΝΙΔΑΣ ΑΚΡΙΤΙΔΗΣ

Περιεχόμενα

1. Αλγόριθμοι μηχανικής εκμάθησης σε δεδομένα μεγάλης κλίμακας.....	1
2. Δομές δεδομένων και ευρετήρια σε δεδομένα μεγάλης κλίμακας - Τεχνικές ανάκτησης πληροφορίας - Μετα-αναζήτηση.....	4
3. Εξόρυξη γνώσης από κοινωνικά και ακαδημαϊκά δίκτυα.....	7
4. Συστήματα.....	10
5. Ανασκόπηση τεχνολογιών ιστού.....	12
6. Ετεροαναφορές.....	13

1. Αλγόριθμοι μηχανικής εκμάθησης σε δεδομένα μεγάλης κλίμακας

Οι αλγόριθμοι μηχανικής εκμάθησης αποτελούν ένα από τα κύρια ενδιαφέροντα της ερευνητικής δραστηριότητας του Λ. Ακριτίδη. Κατά την ενασχόληση του με αυτά τα ζητήματα, δόθηκε ιδιαίτερη έμφαση στην επίλυση προβλημάτων ταξινόμησης (classification) και της συσπείρωσης (clustering) διαφορετικών τύπων οντοτήτων (entities) μεταξύ των οποίων προϊόντα, ερευνητικά άρθρα, ερευνητές, και άλλα. Οι αλγόριθμοι που έχουν προταθεί βασίζονται τόσο σε εποπτευόμενες (supervised), όσο και σε μη εποπτευόμενες (unsupervised) τεχνικές μηχανικής εκμάθησης. Έχουν σχεδιαστεί ώστε να λαμβάνουν υπόψη τους τον μεγάλο όγκο των κατά περίπτωση εμπλεκόμενων δεδομένων. Συγκεκριμένα, έχουν δημοσιευθεί μέθοδοι ανάπτυξης αυτών των αλγορίθμων στις τρέχουσες μοντέρνες πλατφόρμες παραλληλοποίησης όπως το Apache Spark και το Hadoop MapReduce. Οι μέθοδοι αυτοί έχουν σχεδιαστεί με γνώμονα τόσο τον μεγάλο όγκο των δεδομένων, όσο και τον υψηλά διαστατικό (high-dimensional) χώρο χαρακτηριστικών (feature space) και συνδυάζονται με τις κατάλληλες στρατηγικές μοντελοποίησης των υπό εξέταση δεδομένων. Συνοδεύονται δε από στρατηγικές αντιμετώπισης του προβλήματος της μείωσης των διαστάσεων αυτών των χώρων (dimensionality reduction) στις προαναφερθέντα συστήματα παραλληλοποίησης, καθώς και της αποδοτικής αναπαράστασης ως διανύσματα χαρακτηριστικών (feature vectors).

- [c5] L. Akritidis, P. Bozanis, "Computing Scientometrics in Large-Scale Academic Search Engines with MapReduce", *In Proceedings of the 13th International Conference on Web Information System Engineering (WISE)*, Lecture Notes in Computer Science (LLNCS), Springer, vol. 7651, pp. 609-623, 2012.

Στην εργασία αυτή επιχειρείται να αντιμετωπιστεί το ενδιαφέρον πρόβλημα υπολογισμού επιστημονομετρικών (scientometrics) με τη χρησιμοποίηση του υποδείγματος MapReduce, ένα καταναεμημένο, με ανοχή σε σφάλματα, πλαίσιο που χρησιμοποιείται για την επίλυση προβλημάτων σε μεγάλες κλίμακες, χωρίς να λαμβάνονται υπόψη πολύπλοκες λεπτομέρειες διαδικτυακού προγραμματισμού. Αποδεικνύεται ότι, διαρθρώνοντας το πρόβλημα με έναν τρόπο που είναι συμβατό με το MapReduce, μπορούμε να επιτύχουμε μία αποτελεσματική και επεκτάσιμη λύση. Προτείνονται τέσσερις αλγόριθμοι για τον υπολογισμό τριών μετρικών (h-index, contemporary h-index και trend h-index), οι οποίες εκμεταλλεύονται τα χαρακτηριστικά του πλαισίου και συγκρίνεται η αποδοτικότητά τους με τη διεξαγωγή πειραμάτων σε ένα μεγάλο σύνολο δεδομένων, αποτελούμενο από περίπου 1,8 εκατομμύρια επιστημονικά έγγραφα, τα οποία ελήφθησαν από το CiteSeerX.

- [c6] L. Akritidis, P. Bozanis, "A Supervised Machine Learning Classification Algorithm for Research Articles", *In Proceedings of the 28th ACM Symposium on Applied Computing (SAC)*, pp. 115-120, 2013.

Το θέμα της αυτόματης ταξινόμησης των ερευνητικών άρθρων σε μία ή

περισσότερες επιστημονικές περιοχές είναι πρωταρχικής σημασίας για τις επιστημονικές βάσεις δεδομένων και τις ψηφιακές βιβλιοθήκες. Μία αποτελεσματική τεχνική ταξινόμησης μπορεί να καταστήσει την αναζήτηση περισσότερο αποτελεσματική και μπορεί να υποβοηθήσει τους χρήστες στον εντοπισμό παρόμοιων σχετικών εγγράφων. Αν και οι περισσότεροι εκδότες απαιτούν από τους συγγραφείς να ταξινομήσουν οι ίδιοι τα έργα τους, εξακολουθούν να υπάρχουν περιπτώσεις όπου παλαιότερα έγγραφα παραμένουν αταξινόμητα, ή η ταξινομία αλλάζει με την πάροδο του χρόνου. Στην παρούσα εργασία επιχειρείται να αντιμετωπιστεί αυτό το ενδιαφέρον πρόβλημα με την εισαγωγή ενός αλγορίθμου μηχανικής μάθησης, ο οποίος συνδυάζει πολλές παραμέτρους και τα μετα-δεδομένα ενός ερευνητικού άρθρου. Ειδικότερα, το εν λόγω μοντέλο εκμεταλλεύεται το σύνολο εκπαίδευσης για να συσχετίσει λέξεις-κλειδιά, συγγραφείς, συν-συγγραφή, και τα περιοδικά των δημοσιεύσεων σε μια σειρά από ετικέτες της ταξινομίας. Στην συνέχεια, εφαρμόζει αυτή την πληροφορία για να ταξινομήσει το υπόλοιπο των εγγράφων. Τα πειράματα, που έχουν γίνει επί ενός μεγάλου συνόλου δεδομένων που αποτελείται από περίπου 1,5 εκατομμύρια άρθρα, αποδεικνύουν ότι, σε αυτήν την εφαρμογή, το προτεινόμενο μοντέλο ξεπερνά τις μεθόδους AdaBoost.MH και SVM κατά περίπου 6%.

- [c8] L. Akritidis, P. Bozanis, "Effective Unsupervised Matching of Product Titles with k-Combinations and Permutations", *In Proceedings of the 14th IEEE International Conference on Innovations in Intelligent Systems and Applications*, July 3-5, 2018.

Το ζήτημα το οποίο πραγματεύεται η παρούσα εργασία έγκειται στο φαινόμενο της εμφάνισης ενός προϊόντος υπό διαφορετικές παραλλαγές στην περιγραφή του. Σκοπός είναι το ταίριασμα αυτών των διαφορετικών περιγραφών και η αναγνώριση του προϊόντος που αυτές εκπροσωπούν. Πρόκειται για ένα ιδιαίτερα προκλητικό πρόβλημα στο οποίο τα παραδοσιακά μετρικά ομοιότητας (similarity metrics) δεν αποδίδουν ικανοποιητικά. Αυτό οφείλεται στο γεγονός ότι δύο τίτλοι που διαφέρουν ελάχιστα μπορεί να αναφέρονται σε διαφορετικό προϊόν, και αντίστροφα, τίτλοι με σημαντικές διαφοροποιήσεις δύνανται να εκπροσωπούν το ίδιο προϊόν. Πέραν αυτού, η πολυπλοκότητα που συνεπάγεται η σύγκριση ενός τίτλου με όλους τους υπόλοιπους, καθιστά επιτακτική την ανάγκη επινόησης ενός πιο αποδοτικού αλγορίθμου. Εδώ παρουσιάζεται μια μη εποπτευόμενη μέθοδος μηχανικής εκμάθησης (unsupervised learning) η οποία βασίζεται στη μορφολογική ανάλυση των τίτλων και στην κατασκευή των συνδυασμών των λέξεων ενός τίτλου καθώς και των αντιμεταθέσεών τους. Για κάθε συνδυασμό καταγράφεται μια σειρά από στατιστικά στοιχεία σε ειδικές δομές λεξικού και ορίζεται μία συνάρτηση βαθμολόγησης που βασίζεται σε αυτά τα στατιστικά. Ο συνδυασμός που σημείωσε την υψηλότερη βαθμολογία ορίζεται ως ετικέτα του προϊόντος. Η ιδέα έγκειται στο ότι πολλαπλά προϊόντα μπορούν να έχουν

ως ετικέτα τον ίδιο συνδυασμό, οπότε θεωρείται ότι ταιριάζουν. Η πειραματική αποτίμηση επί ενός συνόλου πραγματικών δεδομένων κατέδειξε ότι σε σύγκριση με τρία βασικά μετρικά ομοιότητας, η προταθείσα μέθοδος επιτυγχάνει πάνω από 36% καλύτερο ταίριασμα, ενώ είναι τουλάχιστον 13 φορές γρηγορότερη.

- [c9] L. Akritidis, P. Bozanis, A. Fevgas, "Supervised Papers Classification on Large-Scale High-Dimensional Data with Apache Spark", *In Proceedings of the 4th IEEE International Conference on Big Data Intelligence and Computing (DataCom)*, accepted, to appear, 2018.

Δεδομένου του μεγάλου όγκου των ακαδημαϊκών δεδομένων που χειρίζονται οι σύγχρονες ψηφιακές βιβλιοθήκες και οι μηχανές αναζήτησης, το πρόβλημα της ταξινόμησης των ερευνητικών άρθρων σε περιοχές της επιστήμης καθίσταται εκ νέου προκλητικό. Στην εν λόγω εργασία παρουσιάζεται αρχικά μία βελτιωμένη εκδοχή του αλγορίθμου ταξινόμησης εποπτευόμενης μηχανικής εκμάθησης (supervised machine learning classification algorithm) ερευνητικών άρθρων που εισήγαγε πρώτη η δημοσίευση [c2]. Εν συνεχεία, ο αλγόριθμος επανασχεδιάζεται ώστε να είναι εφικτή η παραλληλοποίηση του με χρήση του Apache Spark, ενός καινοτόμου και σχετικά νέου συστήματος που επιτρέπει τη διαμοίραση μεγάλων ποσοτήτων δεδομένων στην κύρια μνήμη των κόμβων μιας ομάδας δικτυωμένων υπολογιστών. Αναλύεται λεπτομερώς η μεθοδολογία παραλληλοποίησης τόσο της φάσης εκπαίδευσης του μοντέλου (training phase), όσο και της φάσης δοκιμής (test phase) με χρήση των τεχνολογιών και των δομών δεδομένων που διαθέτει το Spark. Επιπλέον, παρουσιάζονται τεχνικές μοντελοποίησης των δεδομένων που έχουν γενική χρησιμότητα και εκτείνονται πέραν των αναγκών και των συνθηκών του συγκεκριμένου προβλήματος. Οι τεχνικές αυτές σχετίζονται με την αναπαράσταση των δεδομένων (data modelling & representation), καθώς και με το πρόβλημα της μείωσης της μεγάλης διαστατικότητας (dimensionality reduction) του εμπλεκόμενου χώρου χαρακτηριστικών (feature space). Τα πειράματα που πραγματοποιήθηκαν με χρήση της συλλογής Open Academic Graph (170 εκατομμυρίων ακαδημαϊκών άρθρων) σε ένα δίκτυο με 128 υπολογιστικούς πυρήνες, έδειξαν ότι ο προτεινόμενος αλγόριθμος είναι αρκετά πιο ακριβής και αποδοτικός σε σχέση με τις εποπτευόμενες τεχνικές μηχανικής εκμάθησης ταξινόμησης πολλαπλών ετικετών που έχουν υλοποιηθεί εντός της βιβλιοθήκης μηχανικής εκμάθησης του Spark, την mlLib.

- [c10] L. Akritidis, A. Fevgas, P. Bozanis, "Effective Product Categorization with Importance Scores and Morphological Analysis of the Titles", *In Proceedings of the 30th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2018)*, under review, 2018.

Ο πρόσφατος εμπλουτισμός των υπηρεσιών που παρέχονται από τις τρέχουσες

πλατφόρμες ηλεκτρονικού εμπορίου (παρουσιάσεις παρόμοιων προϊόντων, προσωποποιημένες προτάσεις, αλγόριθμοι κατανόησης ερωτημάτων, κλπ), κατέστησε το ζήτημα της αποτελεσματικής κατηγοριοποίησης των προϊόντων ζωτικής σημασίας. Σε αυτή την εργασία παρουσιάζουμε έναν εποπτευόμενο αλγόριθμο μηχανικής εκμάθησης για την αυτόματη κατηγοριοποίηση προϊόντων, που βασίζεται στη μορφολογική ανάλυση των τίτλων τους. Συγκεκριμένα, η προταθείσα τεχνική υπολογίζει και εξάγει συνδυασμούς λέξεων και n-grams στα οποία αργότερα προσδίδει βαθμούς σημαντικότητας. Η σημαντικότητα ποσοτικοποιείται λαμβάνοντας υπόψη κριτήρια που αντανακλούν τη δύναμη συσχέτισης ενός συνδυασμού λέξεων ή n-grams με μία κατηγορία προϊόντων και πέραν της κατηγοριοποίησης, δύναται επίσης να χρησιμοποιηθεί για τη μείωση της διαστατικότητας του υπάρχοντος χώρου χαρακτηριστικών (dimensionality reduction). Η πειραματική αποτίμηση της αποτελεσματικότητας της μεθόδου σε ένα αληθινό σύνολο δεδομένων αποτελούμενο από περίπου 320 χιλιάδες τίτλους προϊόντων, έδειξε ότι η εισαχθείσα τεχνική επιτυγχάνει ακρίβεια κατηγοριοποίησης που προσεγγίζει το 95%. Επιπλέον, η χρήση της προταθείσας dimensionality reduction μεθόδου, οδηγεί σε μείωση του μεγέθους του παραχθέντος εκπαιδευμένου μοντέλου κατηγοριοποίησης κατά ποσοστό μεγαλύτερο του 50%, που συνοδεύεται από μία μείωση της ακρίβειας του αλγορίθμου κατά μόλις 2%.

- [c11] L. Akritidis, A. Fevgas, Y. Tsompanopoulou, P. Bozanis, "Efficient MapReduce Implementations of Machine Learning Algorithms on SSDs", In *Proceedings of the 30th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2018)*, under review, 2018.

Ο μεγάλος όγκος των δεδομένων του Παγκόσμιου Ιστού αλλά και η ταχεία μεταβολή τους, κατέστησαν το ζήτημα της εξεύρεσης αποδοτικών μεθόδων ανάλυσης τους πρωταρχικής σημασίας. Υπό αυτές τις συνθήκες δημιουργήθηκε μία νέα μεγάλη γραμμή έρευνας για την παράλληλη εκτέλεση αλγορίθμων μηχανικής εκμάθησης και ανάλυσης στα μοντέρνα συστήματα κατανομής και παραλληλισμού εργασιών. Το Hadoop MapReduce αποτελεί ένα από τα πιο διαδεδομένα συστήματα παραλληλισμού και μεγάλη ποσότητα έρευνας έχει αφιερωθεί στη μελέτη και τη βελτίωση της αρχιτεκτονικής του. Σε αυτή την εργασία μελετάται ο τρόπος που οι παραδοσιακοί αλγόριθμοι μηχανικής εκμάθησης εκτελούνται σε ένα υπολογιστικό κέντρο βασισμένο στο MapReduce και οι κόμβοι του είναι εξοπλισμένοι με μοντέρνους δίσκους στερεάς κατάστασης (SSDs). Δεδομένου ότι το κλασικό MapReduce εγγράφει και διαβάσει μεγάλες ποσότητες προσωρινών δεδομένων στην εξωτερική μνήμη, επιδεικνύουμε πειραματικά ότι η χρήση τέτοιων δίσκων σε αναδρομικούς αλγορίθμους μηχανικής εκμάθησης αποφέρει σημαντικά οφέλη στις επιδόσεις εκτέλεσης.

2. Δομές δεδομένων και ευρετήρια σε δεδομένα μεγάλης κλίμακας - Τεχνικές ανάκτησης πληροφορίας - Μετα-αναζήτηση

Σε αυτή την ενότητα παρουσιάζονται οι εργασίες που σχετίζονται με το ερευνητικό πεδίο της ανάκτησης πληροφορίας. Οι δημοσιεύσεις στην περιοχή δύνανται να ταξινομηθούν σε δύο κατηγορίες. Στην πρώτη εντάσσονται τα άρθρα που σχετίζονται με τις δομές δεδομένων που υποστηρίζουν τις μηχανές αναζήτησης του Ιστού. Το αντεστραμμένο ευρετήριο (*inverted index*) αποτελεί την κύρια δομή που κατασκευάζει και χρησιμοποιεί μία μηχανή αναζήτησης προκειμένου να αποκρίνεται με ταχύτητα στα ερωτήματα των χρηστών. Δεδομένου ότι τα συστήματα αυτά αποθηκεύουν δεκάδες δισεκατομμύρια έγγραφα και δέχονται ερωτήματα με ρυθμούς δεκάδων χιλιάδων ανά δευτερόλεπτο, το αντεστραμμένο ευρετήριο οφείλει να είναι κατασκευασμένο με το βέλτιστο δυνατό τρόπο ώστε να χειρίζεται αυτόν τον τεράστιο φόρτο εργασίας αποδοτικά. Ο τρόπος αναπαράστασης, οργάνωσης και συμπίεσης του αντεστραμμένου ευρετηρίου καθώς και των συνοδευτικών δομών του (π.χ. πίνακας υπερπήδησης - *skip table*), είναι το αντικείμενο των άρθρων αυτής της κατηγορίας. Στη δεύτερη κατηγορία εντάσσονται άρθρα που πραγματεύονται το ζήτημα της συσσωμάτωσης κατατάξεων (*rank aggregation*). Οι αλγόριθμοι αυτοί χρησιμοποιούνται σε ένα διαφορετικού τύπου σύστημα ανάκτησης πληροφορίας, τις μηχανές μετα-αναζήτησης. Τα συστήματα αυτά αναλαμβάνουν να υποβάλλουν το ερώτημα ενός χρήστη σε όλες τις κύριες μηχανές αναζήτησης, και στη συνέχεια να συλλέξουν, να ενοποιήσουν, και να ανακατατάξουν τα αποτελέσματα που επιστρέφονται από αυτές.

- [c2] L. Akritidis, D. Katsaros, P. Bozanis, "Effective Ranking Fusion Methods for Personalized Metasearch Engines", *In Proceedings of the 12th Panhellenic Conference on Informatics (PCI), IEEE*, pp. 39-43, 2008.

Η μετα-αναζήτηση (*meta-searching*) αναφέρεται στην ανάκτηση πληροφορίας στον Παγκόσμιο Ιστό μέσω της κατάλληλης συγχώνευσης των αποτελεσμάτων που μεμονωμένα προσφέρουν οι μηχανές αναζήτησης, όπως Google, Live Search κλπ. Κεντρικό ζήτημα στην μετα-αναζήτηση αποτελούν οι μέθοδοι συσσωμάτωσης κατάταξης (*rank aggregation methods*), οι οποίες κατασκευάζουν την (μοναδική) λίστα εξόδου συγχωνεύοντας τις μεμονωμένες λίστες κατατάξεως των μηχανών αναζητήσεως. Στην εργασία αυτή έμφαση δόθηκε στις ακόλουθες περιπτώσεις: (α) την εκμετάλλευση τοπικών (*regional*) πληροφοριών για τον χρήστη, (β) την βεβαρημένη συμμετοχή των μηχανών αναζήτησης στην συσσωμάτωση των αποτελεσμάτων βάσει της εμπειρίας του χρήστη, (γ) την κατηγοριοποίηση των σελίδων βάσει του υποτομέα (*subdomain*) που ανήκει και, τέλος, (δ) έναν γενικό τύπο συσσωμάτωσης που συνδυάζει τα (α) ως (γ). Όλες οι προτεινόμενες μέθοδοι υλοποιήθηκαν και αξιολογήθηκαν μέσω της πειραματικής μηχανής μετα-αναζήτησης QuadSearch, η οποία φιλοξενείται στη διεύθυνση <http://quadsearch.csd.auth.gr>.

- [j2] L. Akritidis, D. Katsaros, P. Bozanis, "Effective Rank Aggregation for Metasearching", *Journal of Systems and Software, Elsevier*, vol. 84, no 1, pp 130-143, 2011.

Η παρούσα εργασία αφορά στις μηχανές μετα-αναζήτησης και συγκεκριμένα στην ανάπτυξη μεθόδου συσσωμάτωσης κατάταξης (*rank aggregation methods*), η οποία παράγει την λίστα εξόδου αποτελεσμάτων, συγχωνεύοντας τις επιμέρους λίστες κατατάξεως των υποκειμένων μηχανών αναζήτησεως. Ειδικότερα, εισάγεται η μέθοδος QuadRank, η οποία μεταχειρίζεται κατά την διαδικασία συσσωμάτωσης, μία σειρά παραμέτρων, όπως ο τίτλος, τα αποσπάσματα των αποτελεσμάτων (*snippets*), το URL κά Επιπλέον, η μέθοδος υλοποιήθηκε και ελέγχθηκε στα πλαίσια μίας πραγματικής μηχανής μετα-αναζήτησης, της QuadSearch, όπως και με την χρήση πραγματικών δεδομένων του TREC-2009. Τα πειραματικά αποτελέσματα καταδεικνύουν ότι στην συντριπτική πλειοψηφία η προτεινόμενη μέθοδος υπερτερεί τόσο των τεσσάρων υποκειμένων μηχανών αναζήτησης, όσο και της μηχανής μετα-αναζήτησης Dogpile, αλλά και των πετυχημένων μεθόδων συσσωμάτωσης Borda Count και Outranking Approach.

- [j4] L. Akritidis, D. Katsaros, P. Bozanis, "Improved Retrieval Effectiveness by Efficient Combination of Term Proximity and Zone Scoring: A Simulation-based Evaluation", *Simulation Modelling: Practice And Theory, Elsevier*, vol. 22, no. 3, pp. 74-91, 2012.

Σε αυτή την εργασία μελετάται το πρόβλημα της οργάνωσης ενός ανεστραμμένου ευρετηρίου (*index*) όπου αποθηκεύονται πρόσθετες πληροφορίες. Ειδικότερα, εξετάζει πώς οι φυσικές θέσεις ενός εγγράφου, που ονομάζεται ζώνες, μπορούν να ενσωματωθούν αποτελεσματικά με μια τέτοια δομή ευρετηρίου. Εισάγεται ο TZP, ένας κωδικοποιητής που συμπιέζει τις ζώνες αυτές, σε συνδυασμό με τις θέσεις των λέξεων σε ένα έγγραφο, χρησιμοποιώντας ένα σταθερό αριθμό δυαδικών ψηφίων για κάθε τμήμα μίας λέξης στην ανεστραμμένη λίστα. Αποδεικνύεται ότι η μέθοδος επιτρέπει την άμεση πρόσβαση στις συμπιεσμένες ζώνες και θέσεις χωρίς ακριβά ψαξίματα (*look-ups*), αποφεύγει την αποκωδικοποίηση κάθε περιττή πληροφορία, ενώ το συνολικό μέγεθος του ευρετηρίου είναι ανάλογο ή ακόμα και καλύτερο όταν συγκρίνεται με τα τρέχοντα καλύτερα (*state-of-the art*) συστήματα. Επιπλέον, εξετάζει πώς μπορούν να συνδυαστούν οι θέσεις κάθε λέξης με τις ζώνες για να βελτιωθεί η αποτελεσματικότητα της ανάκτησης. Εισάγεται το BM25TOPF, ένα σχήμα το οποίο ενσωματώνει την εγγύτητα των όρων και την στάθμιση ζώνης σε ένα ενιαίο τύπο κατάταξης. Σε αντίθεση με τις άλλες προσεγγίσεις, το BM25TOPF λαμβάνει επίσης υπόψη την διάταξη των όρων ερωτήματος, «επιβραβεύοντας» τα έγγραφα που τους περιέχουν με τη σωστή σειρά. Τα πειράματα με το Web Adhoc Task του συνεδρίου ανάκτησης κειμένου TREC 2009, καθώς και μια σειρά από ερωτήματα που υποβλήθηκαν από τους συγγραφείς του άρθρου, δείχνουν ότι το BM25TOPF ξεπερνά τις τρέχουσες

καλύτερες (state-of-the-art) προσεγγίσεις κατά ένα περιθώριο που κυμαίνεται μεταξύ 6% και 11%.

- [c4] L. Akritidis, P. Bozanis, "Positional Data Organization and Compression in Web Inverted Indexes", *In Proceedings of the 23rd International Conference on Database and Expert Systems Applications (DEXA)*, Lecture Notes in Computer Science (LLNCS), Springer, vol. 7446, pp. 422-429, 2012.

Προκειμένου να ανταποκριθούν στον τεράστιο καθημερινό φόρτο εργασίας, οι μηχανές αναζήτησης του Παγκόσμιου Ιστού χρησιμοποιούν υπερσυμπιεσμένες δομές ευρετηρίων γνωστές και ως ανεστραμμένα ευρετήρια. Έχει καταδειχθεί ότι η οργάνωση των ανεστραμμένων λιστών ενός ευρετηρίου σε επιμέρους τμήματα αναρτήσεων οδηγεί σε σημαντικές βελτιώσεις της αποτελεσματικότητας. Επιπλέον, η πρόσφατη βιβλιογραφία, έχει δείξει ότι οι τρέχουσες καλύτερες (state-of-the-art) στρατηγικές συμπίεσης, όπως οι PForDelta και VSEncoding, αποδίδουν καλά όταν χρησιμοποιούνται για να κωδικοποιήσουν τις λίστες με τα docIDs. Σε αυτή την εργασία εξετάζεται η απόδοσή τους όταν χρησιμοποιούνται για την συμπίεση των τιμών θέσης. Επιδεικνύονται τα μειονεκτήματά τους και εισάγεται το PFBC, ένα απλό αλλά αποτελεσματικό σχήμα κωδικοποίησης, το οποίο κωδικοποιεί τα δεδομένα θέσης από ένα μπλοκ ανεστραμμένης λίστας, χρησιμοποιώντας ένα σταθερό αριθμό bit. Το PFBC επιτρέπει την άμεση πρόσβαση στα αναγκαία δεδομένα, αποφεύγοντας δαπανηρά ψαξίματα (look-ups) και αποκωδικοποίηση περιττής πληροφορίας, επιτυγχάνοντας γρηγορότερη αποσυμπίεση θέσεων κατά περίπου 5 φορές, όπως επιδεικνύουν πειράματα που έγιναν σε μία μεγάλου όγκου συλλογή 50 εκατομμυρίων εγγράφων.

3. Εξόρυξη γνώσης από κοινωνικά και ακαδημαϊκά δίκτυα

Η διάδοση της χρήσης των κοινωνικών δικτύων είχε ως αποτέλεσμα την ανάπτυξη νέων δυνατοτήτων που με τη σειρά τους όρισαν νέες γραμμές έρευνας. Ένα από τα σημαντικότερα προβλήματα στην περιοχή αποτελεί ο εντοπισμός των μελών των κοινωνικών δικτύων που έχουν μεγάλη επιρροή σε άλλα μέλη. Τέτοιου είδους χρήστες παρουσιάζουν ιδιαίτερο ενδιαφέρον, καθώς καθορίζουν ως ένα βαθμό την εμπορική επιτυχία ενός προϊόντος, μιας επιχείρησης ή ενός γεγονότος, ή ακόμα επηρεάζουν ψηφοφόρους, εκλέκτορες, κλπ. Εξίσου σημαντικό είναι και το πρόβλημα της αποτίμησης και της ποσοτικοποίησης της δουλειάς ενός επιστήμονα. Τα τελευταία χρόνια έχει καταστεί αρκετά ενδιαφέρον, καθώς διευκολύνει στη λήψη αποφάσεων για την εξέλιξη των καριέρων ή για την εκτίμηση της ερευνητικής δραστηριότητας των ακαδημαϊκών. Συνεπώς, σημαντική ποσότητα έρευνας έχει αφιερωθεί για την αποτελεσματική επίλυση του. Οι δημοσιεύσεις που παρατίθενται σε αυτή την ενότητα ασχολούνται με αυτού του είδους τα προβλήματα, καθώς επίσης και με την εκμετάλλευση της γνώσης που πηγάζει από τις προταθείσες λύσεις.

[c3] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying Influential Bloggers: Time Does Matter", *In Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 76-83, 2009.

Τα μπλογκ (blogs) έχουν χαρακτηριστεί ως μορφή συμμετοχικής δημοσιογραφίας, με άμεση επίδραση στο κοινό του παγκόσμιου διαδικτύου, καθώς έχει ειπωθεί πως «στους 10 αμερικάνους, ο ένας λέει στους υπόλοιπους εννέα πώς να ψηφίσουν, πού να γευματίσουν και τι να ψωνίσουν». Επομένως, καθίσταται αναγκαία η μελέτη μεθόδων και τεχνικών προσδιορισμού των μπλόγκερ με την μεγαλύτερη επιρροή (influential bloggers). Στην εργασία αυτή προτάθηκαν δύο εύκολα υπολογίσιμα μετρικά συστήματα κατάταξης (ranking) των μπλόγκερ βάσει της επιρροής τους. Η καινοτομία της προτάσεως έγκειται στην εισαγωγή του χρόνου ως μία νέα παράμετρο στην διαδικασία αξιολόγησης της «δραστηριότητας» των μπλόγκερ. Συγκεκριμένα, η πρώτη μετρική MEIBI χρησιμοποιεί για την κατάταξη το πλήθος των δεσμών που καταδεικνύουν ένα μπλογκ (inlinks) και των σχολίων του, όπως και την ημερομηνία δημοσίευσης του μπλογκ. Η δεύτερη μετρική MEIBIX κατατάσσει τα μπλοκ βάσει του πλήθους και της χρονολογίας των εισερχόμενων δεσμών (inlinks) και των σχολίων του. Οι δύο μέθοδοι αξιολογήθηκαν πειραματικά, με πραγματικά δεδομένα της «μπλογκόσφαιρας», σε σχέση με την προηγούμενη πρόταση [N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. *In Proceedings of ACM WSDM Conf.*, p.207–218, 2008]. Τα ληφθέντα αποτελέσματα καταδεικνύουν την αποτελεσματικότητα της πρότασης στην ανακάλυψη των τρεχόντων μπλόγκερ με επιρροή, ενώ επιπρόσθετα είναι εύκολα εφαρμόσιμη αφού δεν προϋποθέτει παραμέτρους συντονισμού (tuning parameters).

- [j1] D. Katsaros, L. Akritidis, P. Bozanis, "The f-index: Quantifying the Impact of Coterminal Citations in Scientists' Ranking". *Journal of the American Society for Information Science and Technology*, vol. 60, no. 5, pp. 1051–1056, 2009.

Παγκοσμίως, οι επιδόσεις ενός ακαδημαϊκού αξιολογούνται και λαμβάνονται υπόψη κατά την διάρκεια προαγωγών, την ανάθεση κονδυλίων κλπ. Εργαλείο στην αξιολόγηση αποτελούν οι διάφορες επιστημονο-μετρικές (scientometrics), όπως το πλήθος των εργασιών, το πλήθος των αναφορών, ο δείκτης h (h-index). Στην παρούσα εργασία εισάγεται και αξιολογείται μία καινούργια μετρική, ο δείκτης f (f-index), ο οποίος προσπαθεί να αντιμετωπίσει το φαινόμενο των συν-οριακών αναφορών (co-terminal citations), δηλαδή των αναφορών από εργασίες στις οποίες παρουσιάζονται ομάδες ιδίων συσσυγγραφέων. Προκειμένου να αναδειχθεί η χρησιμότητα του νέου δείκτη, διενεργήθηκε εκτενής πειραματική μελέτη με επιστήμονες της πληροφορικής, διαθέτοντες μεγάλη τιμή δείκτη h.

- [j3] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying the Productive and Influential Bloggers in a Community", *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 41, no 5, pp. 759-764, 2011.

Η παρούσα δημοσίευση διερευνά το θέμα του προσδιορισμού των bloggers που είναι και παραγωγικοί και έχουν μεγάλη επιρροή στις διάφορες διαδικτυακές κοινότητες, εισάγοντας τον δείκτη παραγωγικότητας του blogger και τον δείκτη επιρροής του blogger. Η πρώτη μετρική, (BP-index), χρησιμοποιείται για την αξιολόγηση της παραγωγικότητας των blogger σε σχέση με τον πρόσφατο χρόνο (recency). Η δεύτερη μετρική, (BI-index) αντισταθμίζει την επιρροή ενός blogger εντός και εκτός της κοινότητας, λαμβάνοντας υπ'όψιν τον αριθμό και την «ηλικία» των εισερχόμενων συνδέσεων και των σχολίων. Ο συνδυασμός αυτών των δύο τιμών χρησιμοποιείται στη συνέχεια για χαρακτηρίσει τους bloggers. Οι προτεινόμενες μετρήσεις αξιολογήθηκαν συγκρινόμενες με τις τρέχουσες καλύτερες (state-of-the-art) μεθόδους με τη χρήση των δεδομένων που συλλέχθηκαν από έναν πραγματικό δικτυακό τόπο blog. Τα ληφθέντα αποτελέσματα επιβεβαιώνουν ότι οι νέες προτεινόμενες μέθοδοι είναι σε θέση να προσδιορίσουν τα σημαντικά μοτίβα (patterns) στην συμπεριφορά των bloggers.

- [j5] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying Attractive Research Fields for New Scientists", *Scientometrics (Springer)*, vol. 91, no. 3, pp. 869-894, 2012.

Η εργασία αυτή αντιμετωπίζει και εισάγει ένα νέο πρόβλημα: Πριν από την έναρξη της επιστημονικής του σταδιοδρομίας, κάθε νέος επιστήμονας είναι υποχρεωμένος να αντιμετωπίσει το κρίσιμο ζήτημα του προσδιορισμού της περιοχής όπου θα διεξάγει την έρευνά του. Ανεξάρτητα από τις δυνατότητες του, μια λανθασμένη επιλογή μπορεί να οδηγήσει σε σπατάλη ενέργειας,

χρόνου και πόρων. Σε αυτό το άρθρο γίνεται προσπάθεια να εντοπιστούν τα ερευνητικά πεδία που είναι ελκυστικά για τους νέους ερευνητές. Η προτεινόμενη προσέγγιση μας συγκρίνεται με ένα στατιστικό μοντέλο, το οποίο αποκαλύπτει δημοφιλείς τομείς έρευνας. Η σύγκριση αυτή μας οδηγεί στο συμπέρασμα ότι δεν είναι όλες οι μοντέρνες (“trendy”) περιοχές έρευνας κατάλληλες για τους νέους επιστήμονες. Ως δευτερεύον συμπέρασμα, αναδεικνύεται η ύπαρξη επιστημονικών τομέων που αν και δεν είναι τόσο αναδυόμενοι ή δημοφιλείς, είναι πολλά υποσχόμενοι για τους επιστήμονες που τώρα ξεκινούν τη σταδιοδρομία τους.

- [j6] L. Akritidis, P. Bozanis, "Improving Opinionated Blog Retrieval Effectiveness with Quality Measures and Temporal Features", *World Wide Web Journal (Springer)*, vol. 17, no. 4, pp 777–798, 2014.

Ένα από τα πιο ενδιαφέροντα προβλήματα στην ανάκτηση πληροφορίας που αφορά τα blog είναι το ζήτημα της ανάκτησης γνώμων (opinionated retrieval), που είναι η ανάκτηση των καταχωρήσεων blog που περιέχουν γνώμες/απόψεις σχετικά με ένα θέμα. Παρόλο που υπάρχουν προσεγγίσεις ανάκτησης καταχωρήσεων blog που είναι σχετικές με ένα συγκεκριμένο ερώτημα και περιέχουν απόψεις, δημιουργώντας μια καταταγμένη λίστα εγγράφων, καμμία δεν λαμβάνει υπόψη το ζήτημα της σημασίας των ανακτημένων απόψεων. Σε αυτήν την εργασία, εισάγεται ένα μοντέλο κατάταξης που συνδυάζει τις υπάρχουσες στρατηγικές ανάκτησης, με ανεξάρτητες του ερωτήματος πληροφορίες για τη βελτίωση της κατάταξης των εγγράφων με γνώμες. Πιο συγκεκριμένα, το νέο μοντέλο λαμβάνει υπ’όψιν την επιρροή του blogger που δημοσίευσε μια γνώμη, η φήμη του δικτυακού τόπου του blog όπου δημοσιεύτηκε η συγκεκριμένη ανάρτηση blog, και η επίδραση της ίδιας της ανάρτησης. Επιπλέον, επεκτείνονται οι τρέχουσες στρατηγικές βαθμολόγησης που βασίζονται στην εγγύτητα, εξετάζοντας τις φυσικές θέσεις των όρων του ερωτήματος και της γνώμης μέσα σε ένα έγγραφο. Η προτεινόμενη προσέγγιση αξιολογείται με εκτενή πειράματα βάσει του συνόλου δεδομένων TREC Blogs08, που αποδεικνύουν ότι η εφαρμογή των μεθόδων βελτιώνει την ακρίβεια ανάκτησης κατά ένα σημαντικό περιθώριο, που φθάνει μέχρι και το 7.2%.

4. Συστήματα

Στην κατηγορία αυτή εμπίπτουν εργασίες που αφορούν πληροφοριακό σύστημα υποστήριξης μεταφοράς με αποτελεσματική ανάθεση εργασιών και δρομολόγηση μεταφορών, καθώς και πειραματική μηχανή μετα-αναζήτησης για την ανάπτυξη και τον έλεγχο αλγορίθμων συσσωμάτωσης κατατάξεων.

- [c1] L. Akritidis, G. Voutsakelis, D. Katsaros, P. Bozanis, "QuadSearch: A novel metasearch engine", *In Proceedings of the 11th Panhellenic Conference on Informatics (PCI)*, pp. 453-466, 2007.

Στην εργασία αυτή περιγράφεται η QuadSearch, μία πειραματική μηχανή μετα-αναζήτησης, η οποία παρέχει αποτελέσματα διενεργώντας ταυτόχρονη προσπέλαση σε τέσσερις γνωστές μηχανές αναζήτησης, βασισμένες σε crawlers, τις Google, Yahoo!, Live Search και Ask Jeeves/Teoma. Προκειμένου να καταστεί κάτι τέτοιο δυνατό, προτάθηκαν δύο νέοι αλγόριθμοι συσσωμάτωσης κατά κατάταξη (rank-based aggregation), η μέθοδος *ke*, οποία μεταχειρίζεται τον τύπο $S/(n^m(k/10+1)^n)$, όπου n το πλήθος των μηχανών που περιέχουν το στοιχείο, m το πλήθος των μηχανών που ρωτήθηκαν, k το πλήθος των στοιχείων που η QuadSearch χρησιμοποιεί και S το συνολικό σκορ που έλαβε το ερωτούμενο αντικείμενο, και η μέθοδος *antispam ke*, η οποία φιλτράρει την *ke*, ανάλογα με τον πληθάρημο των μηχανών που περιέχουν κάθε στοιχείο ερώτησης. Ο χρήστης, μάλιστα, είναι σε θέση να επιλέξει την μέθοδο και να διαμορφώσει τον τρόπο με τον οποίο επιστρέφονται τα αποτελέσματα, δίδοντας, εάν επιθυμεί, έμφαση σε επιπρόσθετα στατιστικά στοιχεία. Η προτεινόμενη μηχανή είναι προσπελάσιμη στην διεύθυνση <http://quadsearch.csd.auth.gr>.

- [c7] L. Akritidis, P. Bozanis, "Efficient Urban Transportation(s) with IoT Devices and Robust Workers Allocation", *In Proceedings of 2018 South Eastern European Conference on Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, to appear, 2018.

Κατά τη διάρκεια των τελευταίων ετών, οι φορητές συσκευές και οι ασύρματοι αισθητήρες διαδραμάτισαν έναν αυξανόμενης σημασίας ρόλο στην εξέλιξη των μεταφορών. Οι συσκευές αυτές συλλέγουν μία ευρεία ποικιλία δεδομένων σχετικά με τους μεταφορείς (όπως την τοποθεσία, την ταχύτητα, την κατεύθυνση, την κατανάλωση καυσίμου, την κατάσταση του οχήματος, κλπ) και τα μεταδίδουν μέσω του Παγκόσμιου Ιστού σε εξειδικευμένους κόμβους επεξεργασίας. Στην τρέχουσα βιβλιογραφία έχουν προταθεί διάφορες μέθοδοι για την εκμετάλλευση αυτών των δεδομένων. Το αντικείμενο αυτών των εργασιών επικεντρώνεται στη βελτίωση της αποδοτικότητας και του οικονομικού αντίκτυπου των συστημάτων μεταφοράς. Σε αυτό το άρθρο παρουσιάζεται το *Anything*, ένα σύστημα σχεδιασμένο με σκοπό τη βελτίωση

της αποδοτικότητας των διαδικασιών παράδοσης πακέτων μέσα στα γεωγραφικά όρια μίας πόλης. Αποδεικνύεται ότι με προσεκτική σχεδίαση ενός κεντρικού συστήματος που επικοινωνεί με μεταφορείς που είναι εξοπλισμένοι με φορητές συσκευές, καθώς και με επινόηση ενός αποτελεσματικού αλγορίθμου ανάθεσης εργασιών, είναι εφικτή η ταχεία λήψη και διεκπεραίωση εργασιών μεταφοράς. Συγκεκριμένα, παρουσιάζονται αληθινά δεδομένα από τη βάση δεδομένων του συστήματος Anything, τα οποία καταδεικνύουν ότι ο μέσος χρόνος που μεσολαβεί από τη λήψη μιας εργασίας μεταφοράς πακέτου/ων μέχρι την ολοκλήρωση της, δεν ξεπερνά την μία ώρα.

5. Ανασκόπηση τεχνολογιών ιστού

Η ενότητα αυτή περιλαμβάνει μία δημοσίευση σε κεφάλαιο βιβλίου το οποίο παρουσιάζει όλες τις μοντέρνες τεχνικές που διέπουν σήμερα τον Παγκόσμιο Ιστό.

- [b1] L. Akritidis, D. Katsaros, P. Bozanis. "Modern Web Technologies", Chapter 4 in *New Directions in Web Data Management 1*, (L. Jain, A. Vakali, Eds.), pp. 83–107, Springer-Verlag, ISBN 978-3-642-17550-3, 2011.

Καθώς ο Παγκόσμιος Ιστός αναδείχθηκε τα τελευταία χρόνια ως ένα από τα σημαντικότερα εργαλεία αναζήτησης πληροφορίας, επικοινωνίας και ανταλλαγής ιδεών και εμπειριών ή ακόμη και εμπορικών συναλλαγών, όπως αγορές και κρατήσεις on-line, το παρόν κεφάλαιο παρουσιάζει τις τεχνολογίες που κατέστησαν εφικτές αυτές τις μοντέρνες εφαρμογές. Μεταξύ άλλων συζητούνται το μοντέλο πελάτη-εξυπηρετητή, το πρωτόκολλο HTTP, η ασύγχρονη επικοινωνία, οι κατακευματισμένες εφαρμογές, η νεφοϋπολογιστική και οι κινητές εφαρμογές ιστού. Παράλληλα, παρουσιάζεται το Web 2.0 και οι εφαρμογές του, όπως τα κοινωνικά δίκτυα, οι κοινότητες ιστού, οι σουίτες γραφείου, οι υπηρεσίες διαμοίρασης αρχείου και πολυμέσων, αλλά και το τελευταίο προκύψαν ζήτημα του ιστού πραγματικού χρόνου.

6. Ετεροαναφορές

6.1. Συνοπτικός Πίνακας Αναφορών

Στον παρακάτω πίνακα παρατίθενται συνοπτικά τα δημοσιευμένα άρθρα ομαδοποιημένα κατά έτος, συνοδευόμενα από το πλήθος των ετερο-αναφορών που έχουν λάβει μέχρι την ημερομηνία σύνταξης του παρόντος.

Έτος	Δημοσίευση	Αναφορές
2007	[c1] L. Akritidis, G. Voutsakelis, D. Katsaros, P. Bozanis, "QuadSearch: A novel metasearch engine", In <i>Proceedings of the 11th Panhellenic Conference on Informatics (PCI 2007)</i> , pp. 453-466, 2007.	3
2008	[c2] L. Akritidis, D. Katsaros, P. Bozanis, "Effective Ranking Fusion Methods for Personalized Metasearch Engines", In <i>Proceedings of the 12th Panhellenic Conference on Informatics (PCI 2008)</i> , IEEE Press, pp. 39-43, 2008.	11
2009	[c3] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying Influential Bloggers: Time Does Matter", In <i>Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2009)</i> , pp. 76-83, 2009.	34
	[j1] D. Katsaros, L. Akritidis, P. Bozanis, "The f-index: Quantifying the Impact of Coterminal Citations in Scientists Ranking", <i>Journal of the American Society for Information Science and Technology</i> , vol. 60, no. 5, pp. 1051-1056, 2009.	25
2011	[j2] L. Akritidis, D. Katsaros, P. Bozanis, "Effective Rank Aggregation for Metasearching", <i>Journal of Systems and Software (Elsevier)</i> , vol. 84, no 1, pp 130-143, 2011.	24
	[j3] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying the Productive and Influential Bloggers in a Community", <i>IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews</i> , vol. 41, no 5, pp. 759-764, 2011.	38
	[b1] L. Akritidis, D. Katsaros, P. Bozanis. "Modern Web Technologies", Chapter 4 in <i>New Directions in Web Data Management 1</i> , (L. Jain, A. Vakali, Eds.), pp. 83–107, Springer-Verlag, ISBN 978-3-642-17550-3, 2011.	1
2012	[j4] L. Akritidis, D. Katsaros, P. Bozanis, "Improved Retrieval Effectiveness by Efficient Combination of Term Proximity and Zone Scoring: A Simulation-based Evaluation", <i>Simulation Modelling: Practice And Theory (Elsevier)</i> , vol. 22, no. 3, pp. 74-91, 2012.	4
	[j5] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying Attractive Research Fields for New Scientists", <i>Scientometrics (Springer)</i> , vol. 91, no. 3, pp. 869-894, 2012.	3
	[c4] L. Akritidis, P. Bozanis, "Positional Data Organization and Compression in Web Inverted Indexes", In <i>Proceedings of the 23rd International Conference on Database and Expert Systems Applications (DEXA 2012)</i> , Lecture Notes in Computer Science (LLNCS), vol. 7446, pp. 422-429, 2012.	1
	[c5] L. Akritidis, P. Bozanis, "Computing Scientometrics in Large-Scale Academic Search Engines with MapReduce", In <i>Proceedings of the 13th International Conference on Web Information System Engineering (WISE 2012)</i> , Lecture Notes in Computer Science (LLNCS), vol. 7651, pp. 609-623, 2012.	3
2013	[c6] L. Akritidis, P. Bozanis, "A Supervised Machine Learning Classification Algorithm for Research Articles", In <i>Proceedings of the 28th ACM Symposium on Applied Computing (SAC 2013)</i> , pp. 115-120, 2013.	4
2014	[j6] L. Akritidis, P. Bozanis, "Improving Opinionated Blog Retrieval Effectiveness with Quality Measures and Temporal Features", <i>World Wide Web Journal (Springer)</i> , vol. 17, no. 4, pp. 777-798, 2014.	8
2018	[c7] L. Akritidis, P. Bozanis, "Efficient Urban Transportation(s) with IoT Devices and Robust Workers Allocation", In <i>Proceedings of the 2018 SouthEast European Conference on Design Automation, Computer Engineering, Computer Networks and Social Media (SEEDA-CECNSM 2018)</i> , to appear, 2018.	0
	[c8] L. Akritidis, P. Bozanis, "Effective Unsupervised Matching of Product Titles with k- Combinations and Permutations", In <i>Proceedings of the 14th IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA 2018)</i> , 3-5 July 2018.	0
	[c9] L. Akritidis, P. Bozanis, A. Fevgas, "Supervised Papers Classification on Large-Scale High-Dimensional Data with Apache Spark" In <i>Proceedings of the 4th IEEE International Conference on Big Data Intelligence and Computing (DataCom 2018)</i> , to appear, 2018.	0
ΣΥΝΟΛΟ		159

6.2. Λεπτομερής Πίνακας Αναφορών

Σε αυτή την ενότητα παρουσιάζονται αναλυτικά οι ετερο-αναφορές που έχει λάβει κάθε δημοσιευμένο άρθρο. Τα άρθρα είναι διατεταγμένα κατά αύξουσα χρονολογία δημοσίευσης.

- [c1] L. Akritidis, G. Voutsakelis, D. Katsaros, P. Bozanis, "QuadSearch: A novel metasearch engine", In *Proceedings of the 11th Panhellenic Conference on Informatics (PCI 2007)*, pp. 453-466, 2007.

[R1] N.M.S. Amador "MetaCluster.PT: Um Meta-Motor de Pesquisa para a Web Portuguesa". Grau de Mestre thesis, Universidade Técnica de Lisboa, Portugal, 2009.

[R2] D. Puspitaningrum, J. Apriansyah Pagua, A. Erlansari, Fauzi, R. Efendi, D. Andreswari, I.S.W.B. Prasetya, "The Analysis of Rank Fusion Techniques to Improve Query Relevance". *TELKOMNIKA*, 13(4):1495-1504, Dec 2015.

[R3] D. Puspitaningrum, F. Boko Susilo, J. Apriansyah Pagua, A. Erlansari, D. Andreswari, R. Efendi, I.S.W.B. Prasetya, "An MDL-Based Frequent Itemset Hierarchical Clustering Technique to Improve Query Search Results of an Individual Search Engine". In *Proceedings of the 11th Asia Information Retrieval Societies Conference, AIRS 2015*, (Zuccon G., et al. Eds.), Lecture Notes in Computer Science, vol.9460, pp. 279-291, 2015.

- [c2] L. Akritidis, D. Katsaros, P. Bozanis, "Effective Ranking Fusion Methods for Personalized Metasearch Engines", In *Proceedings of the 12th Panhellenic Conference on Informatics (PCI 2008)*, IEEE Press, pp. 39-43, 2008.

[R4] Jadidoleslamy, H. Introduction to Metasearching Engines and Result Merging Strategies: A Survey. *International Journal of Advances in Engineering & Technology*, 1(5):30-40, Nov 2011.

[R5] Jadidoleslamy, H. Search's result merging strategies in meta-search engines: a survey. *Journal of Advanced Research in Computer Engineering*, 5(2):123-134, 2001.

[R6] Paredes-Valverdea, M.A., Alor-Hernández, G., Rodríguez-González, A. and Hernández-Chan, G. Developing Social Networks Mashups: An Overview of REST-Based APIs. *Procedia Technology*, 3:205-213, 2012.

[R7] Lalnunsanga, M. *An Introduction to a Meta-meta-search Engine*. MSc Thesis, Andrew University, 2012.

[R8] Pfalzer, M. *Potentially Occuring Mismatches due to Travel-Related Information Retrieval from Search Engines: A comparative analysis of information demand of travellers and the corresponding information supply by stakeholders in the tourism industry*. MSc Thesis, IMC University of Applied Sciences Krems, May 2012.

[R9] Jadidoleslamy, H. Search Result Merging and Ranking Strategies in Meta-Search Engines: A Survey. *International Journal of Computer Science Issues*, 9(4):239-251, July 2012.

[R10] Manzini, M.V. *Realizzazione di un Motore di Ricerca Semantico Basato sul Contesto*. MSc Thesis, Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università degli studi di Modena e Reggio Emilia, 2013-2014.

[R11] Puspitaningrum, D., Apriansyah Pagua, J., Erlansari, A., Fauzi, Efendi, R., Andreswari, D., and Prasetya, I.S.W.B. The Analysis of Rank Fusion Techniques to Improve Query Relevance. *TELKOMNIKA*, 13(4):1495-1504, Dec 2015.

[R12] Mala, V., and Lobiyal, D.K. Learning to Rank Methods for Information Retrieval and Natural Language Processing. *International Journal of Advance Foundation and Research in Computer*, 2(12):34-41, December 2015.

- [R13] Kumar, N., Mittal, S., and Kataria, S. Ranking Techniques Challenges and Solutions in Meta Search Engine. *International Journal of Control Theory and Applications*, 10(30):105-113, 2017.
- [R14] Xu, D., and Shaikh, N.I. A Heuristic Approach for Ranking Items Based on Inputs from Multiple Experts. *International Journal of Information Systems and Social Change*, 9(3), 1-22, 2018.
- [c3] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying Influential Bloggers: Time Does Matter, In *Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2009)*, pp. 76-83, 2009.
- [R15] Sun, C., Liu, B.-Q., Sun, C.-J., Zhang, D.-Y., and Wang, X.-L. SimRank: A link analysis based blogger recommendation algorithm using text similarity. In *Proceedings of the 2010 International Conference on Machine Learning and Cybernetics (ICMLC 2010)*, 11-14 July, 2010, Qingdao, China, vol.6, pp. 3368-3373, July 2010.
- [R16] Kim, E.-H., and Chung, Y.M. Enhancing the Performance of Blog Retrieval by User Tagging and Social Network Analysis. *Journal of the Korean Society for information Management*, 27(1):61-77, March 2010.
- [R17] Aziz, M., and Rafi, M.N. Identifying influential bloggers using blogs semantics. In *Proceedings of the 8th International Conference on Frontiers of Information Technology (FIT'10)*, December 21-23, 2010, Islamabad, Pakistan, 2010.
- [R18] Muller, M., Hupfer, S., Levy, S., Gruen, Sempere, A., and Priedhorsky, R. Circles of Crowdsourcing: The Social Organization of Participatory Sensing. In *Proceedings of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2011)*, Aug 30–Sept 2, 2011, Stockholm, Sweden, ACM, 2011.
- [R19] Shiratuddin, N., Hassan, S., Hashim, N.L., Sakdan, M.F., and Sajat, S.S. Blog Influence Index: A Measure of Influential Weblog. *International Journal of Virtual Communities and Social Networking*, 3(3): 35-45, IGI Global, July-September 2011.
- [R20] Eirinaki, M., Monga, S.P.S., Sundaram, S. Identification of Influential Social Networkers. *International Journal of Web Based Communities*, 8(2):136-158, April 2012.
- [R21] Vernet, E., Bertrandias, L., Galan, J.-P., and Vignolles, A. Identification d'un leader d'opinion : état des controverses. In *Actes du 28ème Congrès Afm (Association Française du Marketing)*, Brest, vol.28, May 2012.
- [R22] Vernet, E., Bertrandias, L., Galan, J.-P., and Vignolles, A. Construit et concept de leader d'opinion dans les réseaux sociaux : proposition d'un agenda de recherche. In *Actes du 28ème Congrès Afm (Association Française du Marketing)*, Brest, vol.28, May 2012.
- [R23] Gao, D. Opinion influence and diffusion in social network. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'12)*, August 12-16, 2012, Portland, Oregon, pp. 997-997, 2012.
- [R24] Gliwa, B., Koźlak, J., Zygmunt, A., and Cetnarowicz, K. Models of Social Groups in Blogosphere Based on Information about Comment Addressees and Sentiments. In *Proceedings of Social Informatics (SocInfo 2012) 4th International Conference*, (K. Aberer et al., Eds.), 5–7 December 2012, Lausanne, Switzerland, LNCS, vol. 7710, 2012, pp 475-488, Springer 2012.
- [R25] Varlamis, I., Eirinaki, M., and Louta, M. Application of Social Network Metrics to a Trust-Aware Collaborative Model for Generating Personalized User Recommendations. Chapter 3, T. Özyer et al. (eds.), *The Influence of Technology on Social Network Analysis and Mining*, Lecture Notes in Social Networks 6, Springer-Verlag, Wien 2013.
- [R26] Shalaby, M., and Rafea, A. Identifying the topic-specific influential users and opinion leaders in Twitter. In *Proceedings of the 12th IASTED International Conference on Artificial Intelligence and Applications*, (AIA 2013), Innsbruck, Austria, 11 - 13 February 2013, pp. 106-113, IASTED Multiconferences, 2013.

- [R27] Shalaby, M. Identifying the topic-specific influential users in Twitter. Msc Thesis, American University in Cairo, 2014.
- [R28] Varlamis, I., Eirinaki, and M., Louta, M. Application of Social Network Metrics to Trust-Aware Collaborative Model for Generationg Personalized User Recommendations. Chapter 3, In The Influence of Technology on Social Network Analysis and Mining, LNSC 6, pp. 49-73, 2013.
- [R29] Shiratuddin, N., Hassan, S., Hashim, N.L., Sakdan, M.F. and Sajat, M.S. Blog Influence Index: A Measure of Influential Weblog. Chapter 11, In Studies in Virtual Communities, Blogs, and Modern Social Networking: Measurements, Analysis, and Investigations, pp. 156-166, IGI-Global 2013.
- [R30] Moh, T.-S., and Shola, S.P. New factors for identifying influential bloggers. In Proceedings of IEEE International Conference on Big Data, Big Data 2013, Santa Clara, CA, United States, 6-9 October 2013, pp. 18-27, 2013.
- [R31] Jimoh, R. G., Awotunde Joseph Bamidele, Enikuomihin, O. A. Identifying Influential Bloggers on the Web. Computing, Information Systems, Development Informatics and Allied Research Journal, IISTE, 5(1):107-118, 2014.
- [R32] Губанов, Д.А., and Чхартишвили, А.Г. Акциональная модель влияния пользователей социальной сети. Control Sciences, 4:20-25, 2014.
- [R33] Abnar, A. Structural Role Mining in Social Networks. MSc Thesis, Department of computing Science, University of Alberta, 2014.
- [R34] Gubanov, D.A., and Chkhartishvili, A. An actional model of user influence levels in a social network. Automation and Remote Control, 76(7):1282-1290, July 2015.
- [R35] Khan, H., Daud, A., and Malik, T.A. MIIB: A Metric to Identify Top Influential Bloggers in a Community. PLoS ONE, 10(9):e0138359, 2015.
- [R36] Kayes, I., and Chakareski, J. Retention in Online Blogging: A Case Study of the Blogster Community. IEEE Transactions on Computational Social System, 2(1):1-14, March 2015.
- [R37] Khan, H.U., and Daud, A. Finding the top influential bloggers based on productivity and popularity features. New Review in Hypermedia and Multimedia, 23(3), 2016.
- [R38] Sadaba, T., and SanMiguel, P. Fashion Blog's Engagement in the Customer Decision Making Process. Chapter 9 in Handbook of Research on Global Fashion Management and Merchandising, pp. 211-259, IGI-Global, 2016.
- [R39] Губанов, Д.А., and Чхартишвили, А.Г. Влиятельность пользователей и метапользователей социальной сети. Control Sciences, 6:12-17, 2016.
- [R40] Hussien, W.A., Tashtoush, Y., Al-Ayyoub, M., and dAL-Kabi, M.N. Are Emoticons Good Enough to Train Emotion Classifiers of Arabic Tweets? In Proceedings of the 7th International Conference on Computer Science and Information Technology (CSIT 2016), Amman, Jordan, 13-14 July, 2016, IEEE, 2016.
- [R41] Ishfaq, U., Khan, H.U., and Iqbal, K. Modeling to find the top bloggers using Sentiment Features. In Proceedings of the 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube 2016), Quetta, Pakistan, 11-12 April, 2016, pp. 227-233, 2016.
- [R42] Zhou, J., Zhang, Y., Wang, B., and Yan, Y. Predicting user influence in microblogs. In Proceedings of the First IEEE International Conference on Computer Communication and the Internet (ICCCI 2016), Wuhan, China, 13-15 October, 2016, pp. 292-295, 2016.
- [R43] Zhou, J. Wu, G., Tu, M., Wang, B., Zhang, Y., and Yan, Y. Predicting user influence under the environment of big data. In Proceedings of the 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA 2017), Chengdu, China, 28-30 April, 2017, pp. 133-138, 2017.
- [R44] Jungnickel, K. Interdisziplinäre Meinungsführerforschung: Eine systematische

Literaturanalyse. Springer VS, Wiesbade, 2017.

[R45] Khan, H.U., Daud, A., Ishfaq, U., and Alowibdi, J.S. Modeling to identify influential bloggers in the blogosphere: a survey. *Computer in Human Behavior*, 68:64-82, Elsevier, 2017.

[R46] SanMiguel, P. and Sádaba, T. Nice to be a fashion blogger, hard to be influential: An analysis based on personal characteristics, knowledge criteria, and social factors. *Journal of Global Fashion Marketing*, DOI: 10.1080/20932685.2017.1399082, 2017.

[R47] Shalaby, M. And Rafea, A. Identifying the Topic-Specific Influential Users in Twitter. *International Journal of Computer Applications*, 179(18):34-39, 2018.

[R48] Gubanov, D.A., and Chkhartishvili, A.G. Influence Levels of Users and Meta-Users of a Social Network. *Automation and Remote Control*, 79(3):545-553, March 2018.

[j1] D. Katsaros, L. Akritidis, P. Bozanis, "The f-index: Quantifying the Impact of Coterminal Citations in Scientists Ranking", *Journal of the American Society for Information Science and Technology (Wiley)*, vol. 60, no. 5, pp. 1051-1056, 2009.

[R49] Yao, L.W. A Study of author self-citation and journal self-citation in environmental engineering. PhD Thesis, Taiwan University, 2009.

[R50] Silagadze, Z.K. Citation entropy and research impact estimation. *Acta Physica Polonica B*, 41(11):2325-2333, 2010.

[R51] De Visscher, A. An index to measure a scientist's specific impact. *Journal of the American Society for Information Science and Technology*, 61(2):319-328, 2010.

[R52] Ajiferuke, I., Lu K., and Wolfram, D. A comparison of citer and citation-based measure outcomes for multiple disciplines. *Journal of the American Society for Information Science and Technology*, 61(10):2086-2096, 2010.

[R53] Lu, K., and Wolfram, D. Delineating Citation Concepts for Author Studies. In 2010 Annual Meeting of the American Society for Information Science and Technology (ASIS&T 2010), October 22-27, Pittsburgh, PA, 2010.

[R54] Bras-Amorós, M., Domingo-Ferrer, J., and Torra, V. A Bibliometric Index Based on Collaboration Distances. In 7th International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2010), (V. Torra, Y. Narukawa, and M. Daumas, Eds.), Perignan, France, October 27-28, 2010, LNAI 6408, pp. 5–6, 2010.

[R55] Bras-Amorós, M., Domingo-Ferrer, J., and Torra, V. A Bibliometric Index based on the Collaboration Distance between Cited and Citing Authors. *Journal of Informetrics*, 5(2):248-264, 2011.

[R56] Ajiferuke, I., Lub, K., and Wolfram, D. Who are the research disciples of an author? Examining publication recitation and oeuvre citation exhaustivity. *Journal of Informetrics*, 5(2):292-302, 2011.

[R57] Takeda, H. Examining Scholarly Influence: A Study in Hirsch Metrics and Social Network Analysis. PhD Thesis, Georgia State University, January 2011.

[R58] Huang, M.-H., and Lin, W.-Y. C. Probing the effect of author self-citations on h index: A case study of environmental engineering. *Journal of Information Science*, 37(5):453-461, 2011.

[R59] Peri, I. Quasi-convex Risk Measures and Acceptability Indices. Theory and Applications. PhD Thesis, University of Milan-Bicocca, January 2012.

[R60] Mazov, N.A., and Gureyev, V.N. Studying the information needs of scientists using bibliometric analysis for acquisition optimization. *БИБЛИОСФЕРА*, 4:57–66, 2012.

[R61] Iskander, D. The EDU-Index: A Way to Objectively Quantify an Individual's University Teaching Output. *IEEE Signal Processing Magazine*, p. 152, September 2013.

[R62] Rosenberg MS. A Biologist's Guide to Impact Factors. *PeerJ PrePrints*, 2:e477v1,

2014.

[R63] Daud, A., and Muhammad, F. Consistent Annual Citations based Researcher Index. Collnet Journal of Scientometrics and Information Management, 8(2):209-216, 2014.

[R64] Bowen, J.B. Provably Correct Systems: Community, connections, and citations, In Proceedings of Festschrift Symposium in Honour of Ernst-Rüdiger Olderog, At Oldenburg, Germany, Sept. 2015.

[R65] Amjad, T., Ding, Y., Daud, A., and Malic, V. Topic-based heterogeneous rank. Scientometrics, 104(1):313-334, May 2015.

[R66] Amjad, T., Daud, A., Che, D., and Akram, A. MuICE: Mutual Influence and Citation Exclusivity Author Rank. Information Processing & Management, 52(3):374-386, 2016.

[R67] Todeschini, R., and Baccini, A. Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research, Wiley-VCH Verlag GmbH & Co. KGaA, Chapter 6, 2016.

[R68] Kun, X.C. Diversity of Forward Citation of Highly Cited Researchers in the Field of Chemistry and Materials Science. Msc Thesis, Taiwan University, 2017.

[R69] Bowen J.P. Provably Correct Systems: Community, Connections, and Citations. Book Chapter in Provably Correct Systems. NASA Monographs in Systems and Software Engineering. Springer, Cham, pp. 313-328, 2017.

[R70] Amjad T., and Daud, A. Indexing of Authors According to their Domain of Expertise. Malaysian Journal of Library and Information Science, 22(1):69-82, April 2017.

[R71] Farooq, M., Khan, H.U., Iqbal, S., Munir, E.U., and Shahzad, A. DS-Index: Ranking Authors Distinctively in an Academic Network. IEEE Access, 5:19588–19596, September 2017.

[R72] Rehman, G., Lee, J., Abbasi, R., Kabir, A., and bin Ubaid, F. Quantifying the Impact of Hot-paper on new Researchers. In Proceedings of the 2017 2nd International Conference on Communication and Information Systems (ICCIS 2017), Wuhan, China, November 07 - 09, 2017, pp. 329-334, 2017.

[R73] Amjad, T., Daud, A., and Aljohan, N.R. Ranking authors in academic social networks: a survey. Library Hi Tech, 2018.

[j2] L. Akritidis, D. Katsaros, P. Bozanis, "Effective Rank Aggregation for Metasearching", *Journal of Systems and Software (Elsevier)*, vol. 84, no 1, pp 130-143, 2011.

[R74] Li, X., and Chen, Li. Recommendations based on network analysis. In the 2011 International Conference on Advanced Computer Science and Information System (ICACSIS 2011), Jakarta, Indonesia, 17-18 December 2011, pp. 9-16, IEEE 2011.

[R75] Amina, G.R., Emrouznejada, A., and Sadeghia, H. Metasearch information fusion using linear programming. RAIRO - Operations Research, 46(4):289-303, October 2012.

[R76] Priya, R.V., and Vadivel, A. Capturing Semantics of Web Page using Weighted TAG- Tree for Information Retrieval. International Journal of Asian Business and Information Management, 3(4):7-24, IGI-Global, 2012.

[R77] Kanawati, R. Mining the dynamics of scientific publication networks for collaboration recommendation. In Proceedings of The Second International Workshop on Mining Communities and People Recommenders (COMMPER 2012), (J. Hollmen et al. Eds.), ECML/PKDD 2012, September 28, 2012, Bristol, UK, pp. 10-23, 2012.

[R78] Marine-Roig, E. A Webometric Analysis of Travel Blogs and Review Hosting: The Case of Catalonia. Journal of Travel & Tourism Marketing, 31(3): 381-396, 2014.

[R79] Moulahi, B., Tamine, L. and Yahia, S. BiAggregator: Multidimensional relevance aggregation based on a fuzzy operator. Journal of the Association for Information Science

and Technology, 2014.

[R80] Dayile, S. Mobile Medical Information for the Deaf. Bsc Thesis, Department of Computer Science, University of the Western Cape, July 2014.

[R81] Baccianella, S., Esuli, A., and Sebastiani, F. Feature Selection for Ordinal Text Classification. *Neural Computation*, 26(3):557-591, 2014.

[R82] Moulahi, B., Tamine, L. and Yahia, S. B. Leveraging Temporal Query-Term Dependency for Time-Aware Information Access. In *Proceedings of the 2015 IEEE / WIC / ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT2015)*, Singapour, Singapore, 6-9 December, 2015.

[R83] Moulahi, B. Definition and evaluation of aggregation models for multidimensional relevance estimation in information retrieval. PhD Thesis, Universite Toulouse III Paul Sabatier, 2015.

[R84] Keyhanipour, A.H., Moshiri, B., and Rahgozar, M. CF-Rank: Learning to rank by classifier fusion on click-through data. *Expert Systems with Applications*, 42(22):8597–8608, Elsevier, December 2105.

[R85] Ansari, M.Z., Sufyan Beg, M.M., and Kumar, M. Enhancement of Fuzzy Rank Aggregation Technique. In *Proceedings of the Second International Conference on Computer and Communication Technologies (IC3t 2015)*, *Advances in Intelligent Systems and Computing* (Eds. M.Z. Ansari, M. M. Sufyan Beg, M. Kumar), vol. 381, pp. 127-135, Springer, 2016.

[R86] Kanagasabapathy, J., and Swaraj Paul, C. Secure and Trusted Information Brokering In Cloud Computing. *International Journal of Scientific Research in Science and Technology*, 2(2):210-217, 2016.

[R87] Mohammed, M., Amine, C.M., and Fethallah, H. Leveraging fuzzy dominance relationship and machine learning for hybrid web service discovery. *International Journal of Web Engineering and Technology*, 11(2):107-132, 2016.

[R88] Keyhanipour, A.H., Moshiri, B., Rahgozar, M., Oroumchian, F., and Ansari, A.A. Integration of data fusion and reinforcement learning techniques for the rank-aggregation problem. *International Journal of Machine Learning and Cybernetics*, 7(6): 1131–1145, December 2016.

[R89] Vijaya, P., Raju, G., and Ray, S.K. Artificial neural network-based merging score for Meta search engine. *Journal of Central South University*, 23(10):2604-2615, October 2016.

[R90] Rao, P.S., and Vasumanthi, D. Web Personalization using the Efficient Fuzzy Cluster Based Multi Objective Social Spider Algorithm. *International Journal of Advanced Scientific Technologies in Engineering and Management Sciences*, 2(12):104-107, 2016.

[R91] Sibony, E. Multiresolution analysis of ranking data. PhD Thesis, Télécom ParisTech, 2016.

[R92] Rao, S., and Vasumathi, D. Utilization of Co-occurrence Pattern Mining with Optimal Fuzzy Classifier for Web Page Personalization. *Journal of Intelligent Systems*, January 2107.

[R93] Aledo, J.A., Gámez, J.A., and Rosete-Suárez. A. Utopia in the solution of the Bucket Order Problem. *Decision Support Systems*, 97:69-80, March 2017.

[R94] Kaur, P, Singh, M., and Josan, G.S. Comparative analysis of Rank Aggregation techniques for metasearch using genetic algorithm. *Education and Information Technologies*, 22(3): 965–983, May 2017.

[R95] Kaur, P, Singh, M., and Josan, G.S., and Dhillon, S.S. Rank aggregation using ant colony approach for metasearch. *Soft Computing*, 1-16, July 2017.

[R96] Maisano, D., and Mastrogiacomo, L. Checking the Consistency of Solutions in Decision-Making Problems with Multiple Weighted Agents. *International Journal of Decision Support System Technology*, 10(1):39-58, 2018.

- [R97] Srinivasa Rao P., Vasumathi D., and Suresh K. The Adaptive Strategies Improving Web Personalization Using the Tree Seed Algorithm (TSA). Chapter 3 in: *Cognitive Science and Artificial Intelligence*, pp. 23-29, SpringerBriefs in Applied Sciences and Technology. Springer, Singapore, 2018.
- [j3] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying the Productive and Influential Bloggers in a Community", *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 41, no 5, pp. 759-764, 2011.
- [R98] Uzunoglu, E., Kip S.M., and Yaman, B. How brand messages are disseminated through blogger endorsement: Descriptive case studies from food sector. In *Proceedings of the 10th International Symposium Communication in the Millennium*, Instabul, 24-26 May, 2012, pp. 272-284, 2012.
- [R99] Magnani, M., Montesi, D., and Rossi, L. Conversation retrieval for microblogging sites. *Information Retrieval*, 15(3-4):354-372, June 2012.
- [R100] Uzunoglu, E., and Öksüz, B. New opportunities in social media for ad-restricted alcohol products: The case of 'Yeni Raki'. *Journal of Marketing Communications*, Taylor & Francis, July 2012.
- [R101] Luo, W., and Tay, W.P. Identifying multiple infection sources in a network. In *Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers (ASILOMAR 2012)*, Pacific Grove, CA, US, 4 - 7 November 2012, pp. 1483-1489, 2012.
- [R102] Luo, W. and Tay, W.P. Identifying infection sources in large tree networks. In *Proceedings of the 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON 2012)*, pp. 281-289, IEEE, 2012.
- [R103] Shalaby, M., and Rafea, A. Identifying the topic-specific influential users and opinion leaders in Twitter. In *Proceedings of the 12th IASTED International Conference on Artificial Intelligence and Applications, (AIA 2013)*, Innsbruck, Austria, 11 - 13 February 2013, pp. 106-113, IASTED Multiconferences, 2013.
- [R104] Luo, W., Tay, W.P. and Leng, M. Identifying Infection Sources and Regions in Large Networks. *IEEE Transactions on Signal Processing*, 61(11): 2850-2865, 2013.
- [R105] Moh, T.-S., and Shola, S.P. New factors for identifying influential bloggers. In *Proceedings of IEEE International Conference on Big Data, Big Data 2013*, Santa Clara, CA, United States, 6-9 October 2013, pp. 18-27, 2013.
- [R106] Xu, Z, Luo, X., Wei, X., and Mei, L. Investigating the Impact Factors for Event Detection Using Web Search Engines. In *Proceedings of 16th IEEE International Conference on Computational Science and Engineering (CSE 2013)*, Sydney, NSW, 3-5 December 2013, pp. 580-587, IEEE 2013.
- [R107] Shalaby, M., and Rafea, A. Identifying the topic-specific influential users and opinion leaders in Twitter. In *Proceedings of the 12th IASTED International Conference on Artificial Intelligence and Applications, (AIA 2013)*, Innsbruck, Austria, 11 - 13 February 2013, pp. 106-113, IASTED Multiconferences, 2013.
- [R108] Chen, K.-Y., Liao, H.-Y., Chi, H.-C., and Liu, D.-R. An analysis of Social Activity Influences in On-line Virtual Worlds. In *Proceedings of the 2013 International DSI and Asia Pacific DSI*, Bali, Indonesia, July 9-13, 2013.
- [R109] Zhou, E., Zhong, N., and Li, Y. Extracting news blog hot topics based on the W2T Methodology. *World Wide Web*, 17(3): 377-404, May 2014.
- [R110] Bui, D.L., Nguyen, T.-T., and Ha, Q.-T. Measuring the Influence of Bloggers in Their Community Based on the H-index Family. In *Proceedings of the 2nd International Conference on Computer Science, Applied Mathematics and Applications (ICCSAMA 2014)*, pp 313-324, AISC 282, Springer, 2014.
- [R111] Agarwal, N., Mahata, D., and Liu, H. Time-and Event-Driven Modeling of Blogger Influence. Book chapter, *Encyclopedia of Social Network Analysis and Mining*, pp. 2154-

2165, Springer New York, Jan 2014.

[R112] Raymond, B.E. Food Safety Communication in Social Media. Msc Thesis, North Carolina State University, 2014.

[R113] Shalaby, M. Identifying the topic-specific influential users in Twitter. Msc Thesis, American University in Cairo, 2014.

[R114] Khan, H., Daud, A., and Malik, T.A. MIIB: A Metric to Identify Top Influential Bloggers in a Community. PLoS ONE, 10(9):e0138359, 2015.

[R115] Vasanthakumar, G.U., Prajakta, B., Shenoy, P.D., and Patnaik, L.M. PIB: Profiling Influential Blogger in Online Social Networks, A Knowledge Driven Data Mining Approach. In Proceedings of the 9th International Multi-Conference on Information Processing, Bangalore, India, 21-23 August, 2015, Procedia Computer Science, 54, pp. 362–370, 2015.

[R116] Luo, W. Identifying infection sources in a network. PhD Thesis, Nanyang Technological University, 2015.

[R117] Zhou, E., Zhong, N., Li, Y., and Huang, J. Hot Topic Detection in News Blog Based on W2T Methodology, Book Chapter, pp. 237-238, In Wisdom Web of Things, 2016.

[R118] Sadaba, T., and SanMiguel, P. Fashion Blog's Engagement in the Customer Decision Making Process. Chapter 9 in Handbook of Research on Global Fashion Management and Merchandising, pp. 211-259, IGI-Global, 2016.

[R119] Vasanthakumar, G.U., Deepa Shenoy, P., and Venugopaln, K.R. PTIB: Profiling Top Influential Blogger in Online Social Networks. International Journal of Information Processing, 10(1): 77-91, 2016.

[R120] Vasanthakumar, G.U., Priyanka, R., Vanitha Raj, K.C., Bhavani, S., Asha Rani, B.R., Deepa Shenoy, P., Venugopaln, K.R. PTMIB: Profiling top most influential blogger using content based data mining approach. In Proceedings of the 2016 International Conference on Data Science and Engineering (ICDSE), Cochin, Kerala, India, 23-25 August 2016.

[R121] Lü, L., Chen, D., Ren, X.-L., Zhang Q.-M., Zhang, Y.-C., and Zhou, T. Vital nodes identification in complex networks. Physics Reports, 650:1-63, 2016.

[R122] Ishfaq, U., Khan, H.U., and Iqbal, K. Modeling to find the top bloggers using Sentiment Features. In Proceedings of the 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube 2016), Quetta, Pakistan, 11-12 April, 2016, pp. 227-233, 2016.

[R123] Khan, H.U., and Daud, A. Finding the top influential bloggers based on productivity and popularity features. New Review in Hypermedia and Multimedia, 23(3), 2016.

[R124] Agarwal, N., and Liu, H. Time- and Event-Driven Modeling of Blogger Influence. Encyclopedia of Social Network Analysis and Mining, pp. 1-13, January 2017.

[R125] Khan, H.U., Daud, A., Ishfaq, U., and Alowibdi, J.S. Modeling to identify influential bloggers in the blogosphere: a survey. Computers in Human Behavior, 68:64-82, Elsevier, 2017.

[R126] Vasanthakumar, G.U., Sunithamma, K., Deepa Shenoy, P., and Venugopal K.R. An Overview on User Profiling in Online Social Networks. International Journal of Applied Information Systems, 11(8):25-42, January 2017.

[R127] Vasanthakumar, G.U., Priyanka, R., Vanitha Raj, K.C., and Venugopal, K.R. PTMIBSS: Profiling Top Most Influential Blogger using Synonym Substitution Approach. ICTACT Journal on Soft Computing, 79(2):1408-1420, January 2017.

[R128] Xu Y., Li F., Liu J., Zhang R., Yao Y., Zhang D. Detecting False Information of Social Network in Big Data. In Proceedings of the 12th International Conference on Collaborate Computing: Networking, Applications and Worksharing (CollaborateCom 2016), (S.Wang, A.Zhou Eds.), Beijing, China, November 10–11, 2016, Lecture Notes of the

- Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol.201, pp. 642-651, Springer, 2017.
- [R129] Zhou, C., Lu, W.-X., Zhang, J., Li, L., Hu, Y., and Guo, L. Early detection of dynamic harmful cascades in large-scale networks. *Journal of Computational Science*, DOI 10.1016/j.jocs.2017.10.014, 2017.
- [R130] Garcias, A.P., Tur, G., Bennassar, F.N., and Lizana, A. Factores de éxito de las comunidades virtuales universitarias basadas en redes sociales. *Análisis de XarFED. Revista Complutense de Educación*, 28(2):497-515, 2017.
- [R131] Jungnickel, K. *Interdisziplinäre Meinungsführerforschung: Eine systematische Literaturanalyse*. Springer VS, Wiesbaden, 2017.
- [R132] Shalaby, M. and Rafea, A. Identifying the Topic-Specific Influential Users in Twitter. *International Journal of Computer Applications*, 179(18):34-39, 2018.
- [R133] Li, Y., Wang, Z., Zhong, X., and Zou, F. Identification of influential function modules within complex products and systems based on weighted and directed complex networks. *Journal of Intelligent Manufacturing*, 1-16, Springer, 2018.
- [R134] Qiao, T., Shan, W., Ganjun, Yu G., and Chen Liu, C. A Novel Entropy-Based Centrality Approach for Identifying Vital Nodes in Weighted Networks. *Entropy* 20(4):261, April 2018.
- [R135] Zhai, L., Yan, X., and Zhang, G. Bi-directional h-index: A new measure of node centrality in weighted and directed networks. *Journal of Informetrics*, 12(1): 299-314, 2018.
- [b1] L. Akritidis, D. Katsaros, P. Bozanis. "Modern Web Technologies", Chapter 4 in *New Directions in Web Data Management 1*, (L. Jain, A. Vakali, Eds.), pp. 83–107, Springer-Verlag, ISBN 978-3-642-17550-3, 2011.
- [R136] Anderson, P. *Web 2.0 and Beyond: Principles and Technologies*. CRC Press, 2012.
- [j4] L. Akritidis, D. Katsaros, P. Bozanis, "Improved Retrieval Effectiveness by Efficient Combination of Term Proximity and Zone Scoring: A Simulation-based Evaluation", *Simulation Modelling: Practice And Theory (Elsevier)*, vol. 22, no. 3, pp. 74-91, 2012.
- [R137] Li, M.-J., Chen, G.-H., Chen, B., and Xu, L.-M. A Dynamic Combinational Evaluation Method Based on Integrated Methods Set. *Chinese Journal of Management Science*, V(2): 132-137, 2013.
- [R138] Li, M.-J. Chen, G.-H., Xu, L., M., and Ou, Z.-H. A dynamic combinatorial method based on dirft. *Chinese Journal of Management Science*, 23(1):141-145, 2015.
- [R139] Li, M.-J., Xu, L.-M., and Chen, G.-H. A dynamic combined evaluation method based on consistency. *Chinese Journal of Management Science*, 24(10):149-155, 2016.
- [R140] Yang, J., Tong, J., Stones, R.J., Zhang, Z., Ye, B., Wang, G., and Liu, X. Selective Term Proximity Scoring Via BP-ANN. *Neu-IR '16, SIGIR Workshop on Neural Information Retrieval*, July 21, 2016, Pisa, Italy, 2016.
- [j5] L. Akritidis, D. Katsaros, P. Bozanis, "Identifying Attractive Research Fields for New Scientists", *Scientometrics (Springer)*, vol. 91, no. 3, pp. 869-894, 2012.
- [R141] Braga H.A. Uma análise cienciométrica das subáreas da ciência da computação. MSc Thesis, Instituto de Informática, Universidade Federal de Goiás, 2013.
- [R142] Mryglod, O., Holovatch, Y., Kenna, R., and Berche, B. Quantifying the evolution of a scientific topic: reaction of the academic community to the Chornobyl disaster. *Scientomertrics*, 106(3):1151–1166, November 2015.
- [R143] Lingfang, W. A Comparative Study of Bibliographic Analysis and Research Front

- between Anesthesiology and Anesthesia-related Institutions. PhD Thesis, Taiwan University, 2016.
- [c4] L. Akritidis, P. Bozanis, "Positional Data Organization and Compression in Web Inverted Indexes", In *Proceedings of the 23rd International Conference on Database and Expert Systems Applications (DEXA 2012)*, Lecture Notes in Computer Science (LLNCS), vol. 7446, pp. 422-429, 2012.
- [R144] Kouris, I., and Makris, C. Indexing and Compressing Text. Chapter 173 in *Encyclopedia of Information Science and Technology*, Third Edition, IGI-Global, pp. 1800-1808, 2015.
- [c5] L. Akritidis, P. Bozanis, "Computing Scientometrics in Large-Scale Academic Search Engines with MapReduce", In *Proceedings of the 13th International Conference on Web Information System Engineering (WISE 2012)*, Lecture Notes in Computer Science (LLNCS), vol. 7651, pp. 609-623, 2012.
- [R145] Senger, H., Gil-Costa, V., Arantes, L., Marcondes, C.A.C., Marin, M., Sato, L.M., and da Silva, F.A.B. BSP Cost and Scalability Analysis for MapReduce Operations. *Concurrency and Computation: Practice and Experience*. 2015.
- [R146] Hellig, L., and Voss, S. A Scientometric Analysis of Public Transport Research. *Journal of Public Transportation*, 18 (2):111-14, 2015.
- [R147] Senger, H., Gil-Costa, V., Desanti, D., and Marcondes, C.A.C. Black-Box Optimization of Hadoop Parameters Using Derivative-Free Optimization. In *Proceedings of the 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP 2016)*, Heraklion, Greece, 17-19 February, 2016, pp. 43-50, IEEE 2016.
- [c6] L. Akritidis, P. Bozanis, "A Supervised Machine Learning Classification Algorithm for Research Articles", In *Proceedings of the 28th ACM Symposium on Applied Computing (SAC 2013)*, pp. 115-120, 2013.
- [R148] Munteanu, D. CONTRIBUȚII LA REALIZAREA SISTEMELOR DE INFORMARE INTELIGENTE. PhD Thesis, Universitatea „Dunărea de Jos” din Galați, 2015.
- [R149] Fukuda, S., Nanba, H., and Takezawa, T. Automatic Classification of Research Papers Focusing on Elemental Technologies and Their Effects. *Journal of Japan Library and Information Science Association*, 62(9):145-162, September 2016.
- [R150] Melethadathil, N., Nair, B.G., Diwakar, S., and Pazhanivelu, S. Assessing Short-Term Social Media Marketing Outreach of a Healthcare Organization using Machine Learning. In *Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI 2017)*, Karnataka, India, 13-16 September 2017, pp. 387-392, IEEE 2017.
- [R151] Wu, C.W. Can machine learning identify interesting mathematics? An exploration using empirically observed laws. arXiv:1805.07431 [cs.LG], 2018.
- [j6] L. Akritidis, P. Bozanis, "Improving Opinionated Blog Retrieval Effectiveness with Quality Measures and Temporal Features", *World Wide Web Journal (Springer)*, vol. 17, no. 4, pp. 777-798, 2014.
- [R152] Luo, Z., Osborne, M., and Wang, T. An effective approach to tweets opinion retrieval. *WWW Journal*, to appear, published on line, DOI: 10.1007/s11280-013-0268-7.
- [R153] Yun, U., Lee, G., and Pyun, G. Correlated blog-page retrieval with structural

- characteristics. In Proceedings of the 6th FTRA International Conference on Computer Science and its Applications (CSA 2014), Guam, US, 17-19 December 2014, LNEE, vol. 330, pp. 191-196, 2015.
- [R154] Khan, H., Daud, A., and Malik, T.A. MIIB: A Metric to Identify Top Influential Bloggers in a Community. PLoS ONE, 10(9):e0138359, 2015.
- [R155] Diplán, C.R. Consumers' Communication Channels Preferences: High-Stake Versus Low-Stake Brands. Msc Thesis, The Rochester Institute of Technology, 2015.
- [R156] Yadav, A., Sharma, D.K., and Pradhan, R. Implicit queries based Temporal Information Retrieval using temporal taggers. In Proceedings of the 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, India, 2-4 September, 2015.
- [R157] Luo, Z., Osborne, M., and Wang, T. An effective approach to tweets opinion retrieval. WWW Journal, 18(3):545-566, May 2016.
- [R158] Khan, H.U., and Daud, A. Finding the top influential bloggers based on productivity and popularity features. New Review in Hypermedia and Multimedia, 23(3), 2016.
- [R159] Li, L., Lin, X., Zhou, M.C., and Fu, L.L. Sociability-based Influence Diffusion Probability Model to evaluate influence of BBS post. Neurocomputing, 293:18–28, 2018.
- [c7] L. Akritidis, P. Bozanis, "Efficient Urban Transportation(s) with IoT Devices and Robust Workers Allocation", In *Proceedings of the 2018 SouthEast European Conference on Design Automation, Computer Engineering, Computer Networks and Social Media (SEEDA-CECNSM 2018)*, to appear, 2018.
- [c8] L. Akritidis, P. Bozanis, "Effective Unsupervised Matching of Product Titles with k-Combinations and Permutations", In *Proceedings of the 14th IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA 2018)*, 3-5 July, 2018.
- [c9] L. Akritidis, P. Bozanis, A. Fevgas, "Supervised Papers Classification on Large-Scale High-Dimensional Data with Apache Spark" In *Proceedings of the 4th IEEE International Conference on Big Data Intelligence and Computing (DataCom 2018)*, to appear, 2018.