

Πτυχιακή Εργασία της Μαυρόγιαννη Μαριαλένα



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΙΟΝΙΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«Επισκόπηση e recruitment εργαλείων και πρακτική εφαρμογή τους»

ΕΠΙΒΛΕΠΟΥΣΑ ΚΑΘΗΓΗΤΡΙΑ : κα Κάτια Λήδα Κερμανίδου

ΦΟΙΤΗΤΡΙΑ : Μαριαλένα Π. Μαυρόγιαννη

ΑΜ: Π2016011

ΚΕΡΚΥΡΑ, 2021

ΠΡΟΛΟΓΟΣ

Η παρούσα Πτυχιακή Εργασία με τίτλο «Επισκόπηση e recruitment εργαλείων και πρακτική εφαρμογή τους» εκπονήθηκε στα πλαίσια της ολοκλήρωσης των προϋποθέσεων, για τη λήψη του πτυχίου μου από το Τμήμα Πληροφορικής του Ιονίου Πανεπιστημίου με έδρα την Κέρκυρα. Σκοπός μου ήταν να μελετήσω σε βάθος τα διαθέσιμα e recruitment εργαλεία και να προτείνω τυχόν βελτιώσεις αλλά και να παρουσιάσω μια κατανοητή και πλήρης ανάλυση του θέματος.

ΠΝΕΥΜΑΤΙΚΑ ΔΙΚΑΙΩΜΑΤΑ

Copyright © Μαυρόγιαννη Μαριαλένα, 2021

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ

Δηλώνω υπεύθυνα πως είμαι η συγγραφέας αυτής της πτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην πτυχιακή εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Καταλαβαίνω πως η λογοκλοπή θα θεωρηθεί λόγος αποτυχίας, σε αυτή την Πτυχιακή Εργασία και στο συνολικό βαθμό της. Δηλώνω πως αυτή η γραπτή εργασία, προετοιμάστηκε από εμένα προσωπικά και αποκλειστικά και πως θα αναλάβω πλήρως τις συνέπειες εάν η εργασία αυτή αποδειχθεί ότι δεν μου ανήκει.

Όνομα:

Μαυρόγιαννη Μαριαλένα

Ημερομηνία:

27/09/2021

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ θερμά την Επιβλέπουσα καθηγήτρια κα Κάτια Λήδα Κερμανίδου για την ανιδιοτελή προσφοράς της, για τον χρόνο που αφιέρωσε και τις πολύτιμες πληροφορίες, ιδέες και κριτικές που μου έδωσε καθ' όλη τη διάρκεια της συγγραφής αυτής της εργασίας.

Ευχαριστώ επίσης όλους τους καθηγητές και καθηγήτριες του Τμήματος Πληροφορικής για όσα με δίδαξαν όλα αυτά τα χρόνια.

Θα ήθελα να απευθύνω τις ευχαριστίες μου στα μέλη της Εξεταστικής Επιτροπής, οι οποίοι με εμπιστεύτηκαν και ευγενικά δέχτηκαν να αξιολογήσουν την παρούσα πτυχιακή εργασία.

Τέλος το μεγαλύτερο ευχαριστώ το δίνω στην οικογένεια μου για όσα μου προσφέρει καθημερινά καθώς και για το γεγονός πως είναι πάντα δίπλα μου και με στηρίζει σε κάθε νέο βήμα, σε κάθε νέο ξεκίνημα.

Περίληψη

Το E-Recruitment, που ονομάζεται επίσης Online Recruitment, είναι η διαδικασία πρόσληψης των πιθανών υποψηφίων για τις κενές θέσεις εργασίας, με τη χρήση ηλεκτρονικών πόρων, ιδίως του Διαδικτύου.

Οι αιτούντες εργασίας στέλνουν τις αιτήσεις ή το βιογραφικό τους σημείωμα μέσω email. Εναλλακτικά τοποθετούν τα βιογραφικά τους στον παγκόσμιο ιστό, τα οποία μπορούν να αντληθούν από υποψήφιους εργοδότες ανάλογα με τις απαιτήσεις τους.

Το E-Recruitment περιλαμβάνει ολόκληρη τη διαδικασία εύρεσης των υποψηφίων υποψηφίων, αξιολόγησης, συνέντευξης και πρόσληψής τους, σύμφωνα με την απαίτηση εργασίας κάθε φορά. Μέσω αυτού, η πρόσληψη γίνεται πιο αποτελεσματικά και αποδοτικά και παράλληλα εξοικονομείται χρόνος και χρήμα.

Στην παρούσα πτυχιακή γίνεται μια εμπεριστατωμένη επισκόπηση των μεθόδων που εστιάζουν στην διαδικασία του E-Recruitment. Αναλύονται επίσης και 3 διαδεδομένες τεχνικές κατηγοριοποίησης οι οποίες μπορούν να εφαρμοστούν σε προβλήματα E-Recruitment. Τέλος, περιγράφεται ένα ενδεικτικό πρόβλημα E-Recruitment μαζί με την επίλυση του και αποτυπώνονται θεωρητικά τα επιδιωκόμενα αποτελέσματα στην εφαρμογή των τριών αλγορίθμων κατηγοριοποίησης που αναγράφονται ανωτέρω.

Λέξεις κλειδιά

Ηλεκτρονική Πρόσληψη, Μηχανική Μάθηση, Μέθοδοι Πρόβλεψης

Abstract

E-Recruitment, also called Online Recruitment, is the process of recruiting potential job vacancies, using electronic resources, especially the Internet.

Job seekers send their applications or CV via email. Alternatively, they post their CVs on the World Wide Web, which can be obtained from prospective employers according to their requirements.

The E-Recruitment includes the entire process of finding candidates, evaluating, interviewing and recruiting them, according to the job requirement each time. Through this, the recruitment becomes more efficient and effective and at the same time saves time and money.

This dissertation provides a comprehensive overview of the methods that focus on the E-Recruitment process. Three common classification techniques that can be applied to E-Recruitment problems are also analyzed. Finally, an indicative E-Recruitment problem is described together with its solution and the desired results are theoretically reflected in the application of the three classification algorithms described above.

Keywords

E-Recruitment, Machine Learning, Prediction Methods

Περιεχόμενα:

Περίληψη	4
Abstract	5
Εισαγωγή	8
1. Τεχνικές Ηλεκτρονικής Εύρεσης Εργασίας	9
1.1. Η Έννοια της Ηλεκτρονικής Πρόσληψης	9
1.1.1. Παράγοντες που Επηρεάζουν την Πρόσληψη	10
1.1.2. Μειονεκτήματα Ηλεκτρονικής Πρόσληψης	11
1.1.3. Πλεονεκτήματα Ηλεκτρονικής Πρόσληψης	11
1.2. Person-Job Fit	12
1.3. Text Matching Τεχνικές	13
1.4. Αποφυγή Αδικίας του Αλγορίθμου Κατά την Εξέταση των Non-Textual Features	14
1.5. Recommender Systems	19
1.6. Deep Learning Models for Text Classification and Matching	21
1.7. Transfer Learning Methods	21
1.8. Αλγόριθμοι Εξαγωγής Πληροφοριών	22
1.9. Comparison Methods	22
2. Ηλεκτρονικά Συστήματα Ευρέσεως Εργασίας	24
2.1. Recruitment Platforms	24
2.2. Web Based Systems	24
2.3. Πλατφόρμες Εύρεσης Εργασίας	26
2.3.1. Freelancers	26
2.3.2. oDesk	28
2.3.3. Elance	28
2.3.4. People Per Hour	28
2.3.5. Fiverr	29
2.3.6. LinkedIn	29
2.3.7. Η Πλατφόρμα Xing	32
2.3.8. Indeed	35
2.3.9. Facebook	35
2.3.10. Twitter	36
2.4. Πλεονεκτήματα Μέσων Κοινωνικής Δικτύωσης ως Εργαλεία Πρόσληψης	36
2.5. Κίνδυνοι και Κόστος από τη Χρήση των Μέσων Κοινωνικής Δικτύωσης για Σκοπούς Πρόσληψης	37

3. Βιβλιογραφική Ανασκόπηση Μελετών Εύρεσης Εργασίας με Χρήση Μηχανικής Μάθησης	39
3.1. Μηχανική Μάθηση Αναγνώρισης Βιογραφικού για Εύρεση Εργασίας	39
3.2. Μηχανική Μάθηση για τον Προσδιορισμό Σχετικών Δεξιοτήτων από Περιγραφή Εργασίας	40
3.2.1. Αρχιτεκτονική WORD2VEC	41
3.2.2. Συλλογή και επεξεργασία δεδομένων	42
3.2.3. Αρχιτεκτονική Skill2vec	43
3.3. Μηχανή Αναζήτησης Εργασίας Βάσει Εταιρικών Χαρακτηριστικών Δεξιοτήτων	46
4. Εφαρμογή Αλγορίθμων Μηχανικής Μάθησης σε Διαδικτυακό Σύστημα Πρόσληψης	48
4.1. Περιγραφή του Προβλήματος	48
4.2. Δέντρο Αποφάσεων	50
4.3. Μηχανές Διανυσμάτων Υποστήριξης	51
4.4. Αλγόριθμος K-Nearest Neighbors	52
5. Συγκριτική Αξιολόγηση τεχνικών Μηχανικής Μάθησης στην ηλεκτρονική εύρεση εργασίας	54
6. Συμπεράσματα	57
Βιβλιογραφία	61

Εισαγωγή

Στην παρούσα εργασία αναλύεται το πρόβλημα της ηλεκτρονικής πρόσληψης και παρουσιάζονται σχετικές τεχνικές και αλγόριθμοι καθώς επίσης και ένα ενδεικτικό πρόβλημα ηλεκτρονικής πρόσληψης με υπολογιστική προσέγγιση. Στο πρώτο κεφάλαιο περιγράφονται οι τεχνικές ηλεκτρονικής εύρεσης εργασίας, πιο συγκεκριμένα η έννοια της ηλεκτρονικής πρόσληψης και διάφορες υπολογιστικές τεχνικές που σχετίζονται με αυτό το πρόβλημα. Στο δεύτερο κεφάλαιο αναφέρονται τα ηλεκτρονικά συστήματα ευρέσεως εργασίας και οι διάφορες πλατφόρμες εύρεσης εργασίας. Στο επόμενο κεφάλαιο γίνεται βιβλιογραφική ανασκόπηση μελετών εύρεσης εργασίας με χρήση μηχανικής μάθησης και στο τέταρτο κεφάλαιο η εφαρμογή αλγορίθμων μηχανικής μάθησης σε διαδικτυακό σύστημα πρόσληψης. Το πέμπτο κεφάλαιο περιγράφει ένα ενδεικτικό πρόβλημα e-πρόσληψης και τα επιδιωκόμενα αποτελέσματα από διάφορες τεχνικές επιβλεπόμενης μηχανικής μάθησης. Στο τελευταίο κεφάλαιο παραθέτουμε τα τελικά συμπεράσματα της εργασίας καθώς επίσης και τις μελλοντικές κατευθύνσεις στα πλαίσια της εργασίας.

1. Τεχνικές Ηλεκτρονικής Εύρεσης Εργασίας

1.1.Η Έννοια της Ηλεκτρονικής Πρόσληψης

Η ηλεκτρονική πρόσληψη αποτελεί ζήτημα ενδιαφέροντος τα τελευταία δέκα χρόνια. Το Διαδίκτυο θεωρείται το πιο πρόσφατο εργαλείο πρόσληψης. Είναι μια πραγματική επανάσταση που εξαπλώνεται στον κόσμο της αναζήτησης και της πρόσληψης εργασίας. Ο όρος διαδικτυακή πρόσληψη, ηλεκτρονική πρόσληψη ή κυβερνοεξέταση, συνεπάγεται την επίσημη προμήθεια πληροφοριών εργασίας στο διαδίκτυο. Οι πρώτες αναφορές σε ηλεκτρονικές προσλήψεις εμφανίζονται σε άρθρα των μέσων της δεκαετίας του 1980. Η ηλεκτρονική πρόσληψη μπορεί να χωριστεί σε δύο τύπους χρήσεων: α. εταιρική ιστοσελίδα για προσλήψεις και πίνακες εμπορικών θέσεων εργασίας (όπως το monster.com) για την ανάρτηση αγγελιών εργασίας, β. ιστότοπος της εταιρείας με σύνδεσμο για την απόσπαση θέσης εργασίας/επιλογές καριέρας όπου οι υποψήφιοι μπορούν να συνδεθούν για νέες θέσεις απασχόλησης. Εάν η εταιρεία διαφημίσει τις κενές θέσεις της σε άλλους ιστότοπους που

ειδικεύονται σε προσλήψεις όπως - naukri.com, timesjob.com, monster.com κ.λπ., οι εταιρείες θα υιοθετούσαν εμπορικούς πίνακες θέσεων εργασίας για προσλήψεις.

Η πρόσληψη είναι η διαδικασία αναζήτησης των υποψηφίων για απασχόληση και η ενθάρρυνσή τους να υποβάλουν αίτηση για θέσεις εργασίας στον οργανισμό. Η πρόσληψη είναι η δραστηριότητα που συνδέει τους εργοδότες και τους αναζητούντες εργασία. Ένας ορισμός της πρόσληψης είναι: Μια διαδικασία εύρεσης και προσέλκυσης ικανών αιτούντων για απασχόληση. Η διαδικασία ξεκινά όταν αναζητούνται νεοσύλλεκτοι και τελειώνει όταν υποβάλλονται οι αιτήσεις τους. Η διαδικτυακή πρόσληψη προσφέρει στους εργοδότες έναν από τους πιο ισχυρούς και οικονομικά αποδοτικούς τρόπους πρόσληψης προσωπικού για έναν οργανισμό. Η διαδικτυακή πρόσληψη, ηλεκτρονική πρόσληψη ή η πρόσληψη μέσω διαδικτύου είναι η χρήση διαδικτυακής τεχνολογίας για την προσέλκυση και την προμήθεια υποψηφίων και την ενίσχυση της διαδικασίας πρόσληψης.

Αυτό που συνήθως συμβαίνει είναι η χρήση της ιστοσελίδας μιας εταιρείας, ενός ιστότοπου εργασίας τρίτων μερών ή ενός πίνακα εργασίας, μιας βάσης δεδομένων βιογραφικών, μάρκετινγκ μηχανών αναζήτησης ή πλατφορμών κοινωνικών μέσων για την κάλυψη κενών θέσεων. Πολλοί μεγάλοι οργανισμοί χρησιμοποιούν το Διαδίκτυο ως πηγή πρόσληψης. Η ηλεκτρονική πρόσληψη είναι η χρήση τεχνολογίας για να βοηθήσει τη διαδικασία πρόσληψης. Διαφημίζονται κενές θέσεις εργασίας μέσω παγκόσμιου ιστού. Οι αιτούντες εργασία στέλνουν τις αιτήσεις τους ή το βιογραφικό τους σημείωμα μέσω ηλεκτρονικού ταχυδρομείου στο διαδίκτυο. Εναλλακτικά, οι αιτούντες εργασία τοποθετούν τα βιογραφικά τους στον παγκόσμιο ιστό, τα οποία μπορούν να συντάξουν οι υποψήφιοι εργαζόμενοι ανάλογα με τις απαιτήσεις τους.

1.1.1. Παράγοντες που Επηρεάζουν την Πρόσληψη

Η λειτουργία πρόσληψης των επιχειρήσεων επηρεάζεται και διέπεται από ένα μείγμα διαφόρων εσωτερικών και εξωτερικών δυνάμεων. Οι εσωτερικές δυνάμεις ή παράγοντες είναι οι παράγοντες που μπορούν να ελεγχθούν από την επιχείρηση. Και οι εξωτερικοί παράγοντες είναι εκείνοι οι παράγοντες που δεν μπορούν να ελεγχθούν από αυτή.

Η σημασία της μελέτης των Kapse et al., (2012) αφορά την ανάπτυξη του έργου που δίνει την ιδέα για τη δημιουργία διαδικτυακής πύλης πρόσληψης θέσεων εργασίας που θα αντικαταστήσει τον παραδοσιακό τρόπο πρόσληψης υποψηφίου δημοσιεύοντας διάφορες αγγελίες σε εφημερίδες. Η κύρια σημασία για τη δημιουργία διαδικτυακής πύλης εργασίας

που προσφέρει το διαδικτυακό σύστημα για διάφορα πανεπιστήμια είναι : 1) Να εξαλειφθεί η χειροκίνητη διαδικασία πρόσληψης υποψηφίου 2) Η πιο οικονομική μέθοδος 3) Μέθοδος εξοικονόμησης χρόνου, δηλαδή γρήγορη και βολική διαδικασία 4) Διαθεσιμότητα θέσεων εργασίας 24/7.

Η ηλεκτρονική πρόσληψη στοχεύει στην ανάπτυξη ενός διαδικτυακού και κεντρικού συστήματος διαδικασιών πρόσληψης για τον Όμιλο Ανθρώπινου Δυναμικού σε μια εταιρεία.

Το σύστημα επιτρέπει τόσο σε άτομα που αναζητούν εργασία όσο και σε εταιρείες να αλληλεπιδρούν και να ολοκληρώνουν επιτυχώς τη διαδικασία πρόσληψης. Το σύστημα παρέχει καλά καθορισμένες διεπαφές με όλες τις απαιτούμενες λειτουργίες.

Για να παρέχεται ευελιξία στους χρήστες, έχουν αναπτυχθεί οι διεπαφές που είναι προσβάσιμες μέσω προγράμματος περιήγησης. Τα GUI "S στο ανώτερο επίπεδο έχουν κατηγοριοποιηθεί ως:

- Διοικητική διεπαφή χρήστη η λειτουργική ή γενική διεπαφή χρήστη. Η «διαχειριστική διεπαφή χρήστη» επικεντρώνεται στις συνεπείς πληροφορίες που είναι πρακτικά μέρος των οργανωτικών δραστηριοτήτων και οι οποίες χρειάζονται κατάλληλο έλεγχο ταυτότητας για τη συλλογή δεδομένων. Αυτές οι διεπαφές βοηθούν τους διαχειριστές με όλες τις καταστάσεις συναλλαγής, όπως εισαγωγή δεδομένων, διαγραφή δεδομένων και ενημέρωση δεδομένων μαζί με τις εκτεταμένες δυνατότητες αναζήτησης δεδομένων.
- Η «λειτουργική ή γενική διεπαφή χρήστη» βοηθά τους τελικούς χρήστες του συστήματος στις συναλλαγές μέσω των υπαρχόντων δεδομένων και των απαιτούμενων υπηρεσιών. Η λειτουργική διεπαφή χρήστη βοηθά επίσης τους απλούς χρήστες να διαχειρίζονται τις δικές τους πληροφορίες με προσαρμοσμένο τρόπο, σύμφωνα με τις ευελιξίες που περιλαμβάνονται.

1.1.2. Μειονεκτήματα Ηλεκτρονικής Πρόσληψης

Εκτός από τα διάφορα οφέλη, η ηλεκτρονική πρόσληψη έχει το δικό της μερίδιο ελλείψεων και μειονεκτημάτων. Κάποια από αυτά είναι:

- Ο έλεγχος της χαρτογράφησης δεξιοτήτων και της γνησιότητας εκατομμυρίων βιογραφικών αποτελεί πρόβλημα και χρονοβόρα άσκηση για τους οργανισμούς.

- Υπάρχει μικρή διείσδυση στο Διαδίκτυο, καμία πρόσβαση και έλλειψη ευαισθητοποίησης για το Διαδίκτυο σε πολλές τοποθεσίες του κόσμου και σε όλη την Ινδία.
- Περιορισμός αιτούντων από ομάδες που δεν μπορούν να χρησιμοποιήσουν πλήρως την τεχνολογία όπως κάποια άτομα με ειδικές ανάγκες
- Οι οργανισμοί δεν μπορούν να εξαρτώνται αποκλειστικά και μόνο από τις διαδικτυακές μεθόδους πρόσληψης.
- Η ευκολία αίτησης για εργασία ίσως προσελκύσει μεγαλύτερο αριθμό ακατάλληλων αιτούντων
- Η διαδικασία μπορεί να θεωρηθεί ως απρόσωπη από κάποιους
- Στην Ινδία, οι εργοδότες και οι εργαζόμενοι εξακολουθούν να προτιμούν μια προσωπική αλληλεπίδραση αντί να στέλνουν e-mail.(Vinky Sharma “Impact of E-recruitment on human resource supply chain management”)

1.1.3. Πλεονεκτήματα Ηλεκτρονικής Πρόσληψης

Στα πλεονεκτήματα της ηλεκτρονικής πρόσληψης συγκαταλέγονται τα εξής:

- Ευρύτερη πρόσβαση και γεωγραφική εξάπλωση
- Μεγαλύτερο κοινό
- Μεγαλύτερη πιθανότητα να βρεθεί ο κατάλληλος υποψήφιος γρηγορότερα/με μεγαλύτερη αποτελεσματικότητα
- 24/7 - καμία αναμονή για τις ημερομηνίες συμφωνίας
- Εξοικονόμηση χρόνου/εξοικονόμηση κόστους
- Σχετικά οικονομικό
- Καλύτερη αντιστοίχιση εργαζομένων - κενές θέσεις
- Αύξηση αποδοτικότητας
- Εξοικονόμηση κόστους/εξοικονόμηση κόστους προσωπικού
- Απόκτηση πρόσβασης σε αναζητούντες εργασία
- Μείωση των μη προσόντων
- Περισσότερες ευκαιρίες για μικρότερες εταιρείες

1.2. Person-Job Fit

Το Person-Job Fit στοχεύει στην αυτόματη αντιστοίχιση θέσεων εργασίας με τους κατάλληλους υποψηφίους υπαλλήλους. Η αντιστοίχιση των θέσεων εργασίας με τους υποψηφίους που είναι οι χρήστες ουσιαστικά καλείται job fit, δηλαδή συνταίριασμα (Shen et al., 2018). Οι υπάρχουσες μελέτες σχετικά με το Person-Job Fit επικεντρώνονται συνήθως στη μέτρηση του βαθμού αντιστοίχισης μεταξύ προσόντων-ταλέντων και απαιτήσεων εργασίας, βασιζόμενες κυρίως στο χειροκίνητη εξέταση από εμπειρογνώμονες, οι οποίοι όμως θα μπορούσαν εύκολα να παραπλανηθούν από την υποκειμενική, ελλιπή και αναποτελεσματική φύση της ανθρώπινης κρίσης. Αυτό ουσιαστικά σημαίνει πως τα κριτήρια επιλογής μπορεί να διαφέρουν ανάλογα με τη φύση και τη κρίση του ατόμου που είναι υπεύθυνο για τις προσλήψεις κάθε φορά. Οι πρώτες ερευνητικές προσπάθειες του Person-Job Fit μπορούν να χρονολογηθούν από τους (Malinowski et al., 2006). Ωστόσο η ραγδαία ανάπτυξη διαδικτυακών πλατφορμών προσλήψεων όπως το LinkedIn, και το Lagou, έδωσε έναυσμα για περαιτέρω έρευνα. Πιο συγκεκριμένα το 2017, υπήρχαν 467 εκατομμύρια χρήστες και 3 εκατομμύρια ενεργές λίστες θέσεων εργασίας στο LinkedIn από περίπου 200 χώρες και περιοχές σε όλο τον κόσμο. Αργότερα σύμφωνα με τον (Iqbal, 2019), υπήρχαν 590 εκατομμύρια χρήστες και 11 εκατομμύρια λίστες θέσεων εργασίας στο LinkedIn από περίπου 200 χώρες και περιοχές σε όλο τον κόσμο. Το παραδοσιακό Person-Job Fit έχει μελετηθεί από διαφορετικές οπτικές γωνίες, όπως το μοντέλο πέντε παραγόντων (διαφάνεια, συνείδηση, εξωστρέφεια, συμφωνία και νευρωτισμός), οι χαρακτηριστικές θέσεις εργασίας (job characteristic beliefs) και η χειροτεχνία εργασίας (job crafting) τα οποία έχουν συνδεθεί με την απόδοση των ατόμων. Δεδομένου ότι πολλαπλές ικανότητες μπορεί να ταιριάζουν στην ίδια απαίτηση και διαφορετικοί υποψήφιοι μπορεί να έχουν διαφορετικές ικανότητες, όλες οι ικανότητες πρέπει να ζυγίζονται για μια ολοκληρωμένη βαθμολογία με σκοπό να γίνει σύγκριση του βαθμού αντιστοίχισης μεταξύ διαφορετικών υποψηφίων (Chuan Q. et al., 2020).

Ως σημαντικό καθήκον στην εξόρυξη δεδομένων πρόσληψης (Li et al., 2017a; Xu et al., 2018; Oentaryo et al., 2018), η προσαρμογή προσωπικού-εργασίας έχει μελετηθεί εκτενώς στη βιβλιογραφία. Οι πρώτες μέθοδοι περιλαμβάνουν την αντιμετώπιση της αντιστοίχισης ατόμων-θέσεων εργασίας ως πρόβλημα σύστασης θέσης εργασίας / υποψηφίου (Lu et al., 2013; Diaby et al., 2013) και επέκταση της ιστορικής συμπεριφοράς των χρηστών με αρχεία

αιτήσεων εργασίας για καλύτερο ταίριασμα (Hong et al., 2013). Πρόσφατα, χρησιμοποιήθηκαν μέθοδοι βαθιάς μάθησης για το σχεδιασμό μιας δομής ιεραρχικής αναπαράστασης βάσει ιστορικών αρχείων προσλήψεων (Qin et al., 2018; Zhu et al., 2018; Ramanath et al., 2018). Μεταξύ των πρώτων έργων, το Structural Correspondence Learning (SCL) (Blitzer et al., 2007, 2006) είναι ένας κλασικός αλγόριθμος που μαθαίνει μεταβιβάσιμες αναπαραστάσεις χαρακτηριστικών σε διάφορους τομείς. Με την αναβίωση της βαθιάς μάθησης, αρκετές μελέτες επικεντρώνονται στην επέκταση των μοντέλων νευρωνικών δικτύων για προσαρμογή τομέα (Yu and Jiang, 2016; Ziser and Reichart, 2017).

1.3.Text Matching Τεχνικές

Με την πάροδο του χρόνου, τα υπολογιστικά συστήματα έχουν γίνει όλο και πιο έξυπνα και μπορούν να λάβουν ανεξάρτητες αποφάσεις αναλύοντας πληθώρα ανεπεξέργαστων δεδομένων. Το πρόβλημα της μέτρησης της καταλληλότητας των υποψηφίων για εργασία έχει τεθεί σε ένα πεδίο έρευνας από την εμφάνιση των πρώτων συμβούλων εργασίας στο διαδίκτυο τη δεκαετία του 1990. Αναφερόμενοι στην τεχνική του text matching, ουσιαστικά αναφερόμαστε σε έναν αλγόριθμο ο οποίος χρησιμοποιείται για πολλαπλούς σκοπούς ένας εκ των οποίων είναι να αντιστοιχίζει τα άτομα που αναζητούν εργασία με τις θέσεις εργασίας στις οποίες ανταποκρίνονται καλύτερα με βάση τις γνώσεις και τις ικανότητές τους. Ουσιαστικά η αντιστοίχιση κειμένου είναι η διαδικασία αναγνώρισης και εντοπισμού συγκεκριμένων αντιστοιχιών κειμένου σε ακατέργαστα δεδομένα. Η αντιστοίχιση κειμένου είναι ένα ζωτικό συστατικό σε πρακτικές εφαρμογές και μια ουσιαστική διαδικασία σε διάφορους τομείς. Το ίδιο ισχύει και για την διαδικασία της πρόσληψης η οποία αποτελεί μια ιδιαίτερη και ζωτικής σημασίας διαδικασία για κάθε εταιρεία. Η βελτίωση της ποιότητας αντιστοίχισης για τη σύνδεση υποψηφίων και εταιρειών είναι σημαντική όχι μόνο για την απόδοση του προσωπικού αλλά και για την κοινωνία ολόκληρη. Δεδομένου ότι εκατοντάδες υποψήφιοι υποβάλλουν αίτηση για μία μόνο θέση πολλές φορές, το να υπάρξει ένα φιλτράρισμα στους υποψηφίους προκειμένου να προκύψει ο ιδανικός υποψήφιος, είναι δύσκολη υπόθεση. Ταυτόχρονα, εάν η διαδικασία διαρκεί πολύ, οι υποψήφιοι ενδέχεται να τείνουν να αποχωρούν ή να βρίσκουν αλλού εργασία στο ενδιάμεσο διάστημα. Αυτός είναι και ένας από τους λόγους για τον οποίο οι αλγόριθμοι αυτοί βρίσκονται στην επικαιρότητα και απασχολούν πολλούς ερευνητές. Είναι εμφανές πως οι αλγόριθμοι που βασίζονται στις text matching τεχνικές έχουν βοηθήσει τους αιτούντες εργασία να βρουν τις καλύτερες θέσεις

εργασίας για να υποβάλουν αίτηση και αντίστοιχα τους υπεύθυνους προσλήψεων να βρουν τον κατάλληλο εργαζόμενο για κάθε θέση. Οι συγκεκριμένοι αλγόριθμοι μπορούν να παρέχουν προτάσεις εργασίας σε αιτούντες και ταυτόχρονα, μπορούν να στείλουν αντίστοιχους αιτούντες στις εταιρείες με τις οποίες ταιριάζουν. Αυτό μπορεί να βοηθήσει και τα δύο μέρη, ενώ παράλληλα εξυπηρετεί την εταιρεία που παρέχει τις υπηρεσίες με δύο ξεχωριστές ροές εσόδων. Επιπλέον με βάση τα εξαγόμενα αποτελέσματα ο υποψήφιος μπορεί να λάβει μια ειδοποίηση για μια θέση εργασίας στην οποία θα μπορούσε να ήταν κατάλληλος και έπειτα να αποφασίσει αν θα υποβάλλει αίτηση. Οι matching algorithms (αλγόριθμοι αντιστοίχισης εργασίας) λοιπόν, είναι αλγόριθμοι που χρησιμοποιούνται για την επίλυση προβλημάτων αντιστοίχισης γραφημάτων στη θεωρία γραφημάτων. Ένας τέτοιος αλγόριθμος είναι ο Boyer-Moore-Horspool. Τα προβλήματα αντιστοίχισης γραφημάτων είναι πολύ κοινά στις καθημερινές δραστηριότητες. Οι matching algorithms αντιστοιχίζουν ένα άτομο που έκανε αίτηση για εργασία δοκιμαστικά και στη πορεία μπορεί να το αφαιρέσουν για να κάνουν χώρο για έναν αιτών με υψηλότερη βαθμολογία. Οι διαδικασίες αντιστοίχισης που αναφέρουμε εκτελούνται από πολλούς υπαλλήλους που κατέχουν το ρόλο του στρατολογητή χρησιμοποιώντας τα συστήματα αναζήτησης. Πιο συγκεκριμένα τα συστήματα αναζήτησης δημιουργούν έναν μακρύ κατάλογο υποψηφίων κατόπιν αιτήματος του στρατολογητή. Με βάση τη λίστα, ο υπεύθυνος προσλήψεων προσφέρει θέσεις εργασίας στους υποψηφίους μέσω email ή τηλεφώνου. Πριν από αυτή την επικοινωνία, ο εκάστοτε στρατολόγος επαληθεύει δεδομένα ελεύθερου κειμένου από το υποψήφιο σύνολο δεδομένων που καταγράφηκε από τον οργανισμό προσωπικού μετά από πρόσωπο με πρόσωπο συνέντευξη ή τηλεφωνική συνέντευξη (Kino et al., 2017). Το ελεύθερο κείμενο περιλαμβάνει ειδικές απαιτήσεις, χαρακτηριστικά, συνθήκες και άλλα ειδικά θέματα. Ο τύπος δεδομένων ελεύθερου κειμένου είναι ένας από τους πιο συχνά χρησιμοποιούμενους τύπους δεδομένων, καθώς μπορεί να συλλέξει οποιονδήποτε τύπο χαρακτήρα κειμένου. Γενικά με την αυξανόμενη ποσότητα δεδομένων κειμένου στο διαδίκτυο, υπήρξε η απαίτηση για την δημιουργία καλύτερων εργαλείων ανάλυσης κειμένου προσαρμοσμένα σε εξειδικευμένες περιπτώσεις όπως αυτή της ηλεκτρονικής πρόσληψης και αναζήτησης εργατικού δυναμικού. Η λύση που έδωσε η τεχνική του text matching είναι η χρήση μέτρων ομοιότητας κειμένου για τον εντοπισμό αντίστοιχων τμημάτων κειμένου (Landthaler et al., 2018). Θα πρέπει ωστόσο να τονίσουμε ότι υπάρχουν και κάποιοι περιορισμοί στους αλγόριθμους που βασίζονται στις text matching τεχνικές. Ενώ οι αλγόριθμοι αντιστοίχισης εργασίας έχουν

σημειώσει τεράστια βήματα στον τομέα της αυτόματης αντιστοίχισης εργασίας, εξακολουθούν να υπάρχουν πολλά ζητήματα ή περιορισμοί που πρέπει να ξεπεραστούν. Ένας αλγόριθμος υποτίθεται ότι στερείται οποιασδήποτε προκατάληψης, αλλά δεδομένου ότι τα δεδομένα εκπαίδευσης εισάγονται από τον άνθρωπο, η μεροληψία συχνά αποτελεί θέμα σε αυτούς τους αλγόριθμους στο χρόνο εκτέλεσης. Για παράδειγμα, εάν χρησιμοποιήσουμε έναν αλγόριθμο αντιστοίχισης θέσεων εργασίας με ιστορικά δεδομένα, τότε υπάρχει η τάση του αλγορίθμου να επιλέγει περισσότερους άντρες παρά γυναίκες, δεδομένου ότι ο κλάδος της τεχνολογίας αποτελούνταν κυρίως από άνδρες την προηγούμενη δεκαετία. Τέτοιες ανεπάρκειες μπορούν να προκύψουν όταν τα δεδομένα στα οποία γίνεται αναζήτηση δεν αναλύονται σωστά εκ των προτέρων. Ένα άλλο πρόβλημα με την κατάρτιση αλγορίθμων αντιστοίχισης θέσεων εργασίας είναι ότι οι υπεύθυνοι πρόσληψης εισάγουν δεδομένα στους αλγόριθμους βασιζόμενοι στο υπάρχον δυναμικό της εταιρείας. Αυτό προκαλεί σημαντική μεροληψία στο σύστημα, καθώς ο αλγόριθμος αναζητά μόνο άτομα με πολύ παρόμοια χαρακτηριστικά και έτσι διαφορετικά προφίλ ή προφίλ που ξεχωρίζουν συνήθως καταλήγουν να απορρίπτονται. Αυτό με τη σειρά του, θα έχει αντίκτυπο στην ποικιλομορφία των αιτούντων που καταλήγουν να προσληφθούν, όσον αφορά τις δεξιότητες, τα χαρακτηριστικά της προσωπικότητας και την εργασιακή εμπειρία.

1.4.Αποφυγή Αδικίας του Αλγορίθμου Κατά την Εξέταση των Non-Textual Features

Παρότι η ακρίβεια του αλγορίθμου είναι απαραίτητη, ένα ακόμα σημαντικό ζήτημα είναι η ενδυνάμωση των σωστών τιμών του συστήματος πρόσληψης και η διασφάλιση της ορθότητας του αλγορίθμου. Αυτό σημαίνει πως για να γίνει δίκαιος οποιοσδήποτε αλγόριθμος που βασίζεται στη μηχανική μάθηση, πρέπει πρώτα να αποφύγει την προκατάληψη των δεδομένων κατάρτισης, όπως η σημαντική διαφορά στην αναλογία απασχόλησης γυναικών και ανδρών. Δυστυχώς στην πραγματική ζωή οι προκαταλήψεις φαίνεται δύσκολο να αποφευχθούν εντελώς. Το πρόβλημα επιδεινώνεται περαιτέρω από ποικίλα, μη τυποποιημένα χαρακτηριστικά του βιογραφικού ή συνοδευτικών επιστολών κάθε αιτούντος. Με γνώμονα τα παραπάνω καταλήγουμε πως, δεν πρέπει να προσθέσουμε τη δυνατότητα φύλου για να εκπαιδεύσουμε κάποιο μοντέλο έτσι ώστε να αποφευχθούν οι φυλετικές διακρίσεις. Καταλαβαίνουμε πως όλα τα μοντέλα αποδίδουν καλύτερα χωρίς τη χρήση πληροφοριών φύλου. Επομένως, μπορούμε να συμπεράνουμε ότι όταν το σύνολο δεδομένων προσλήψεων ιστορικού περιέχει την προκατάληψη της κατανομής δεδομένων,

όπως η διάκριση λόγω φύλου, δεν πρέπει να χρησιμοποιούμε τα αντίστοιχα χαρακτηριστικά για να εκπαιδεύσουμε το μοντέλο. Σε γενικές γραμμές, η χρήση μεθόδων αυτόματης αντιστοίχισης έχει σημαντικές δυνατότητες μείωσης της κοινωνικής ανισότητας, επιτρέποντας στους κατάλληλους υποψηφίους να κληθούν για συνέντευξη βασιζόμενη αποκλειστικά στην ικανότητά τους και όχι σε επιφανειακούς παράγοντες όπως η ηλικία, το φύλο ή η φυλή. Ακόμα κι αν ο ερευνητής έχει προκατάληψη, η πρόσκληση για συνέντευξη είναι πιθανό να βελτιώσει σημαντικά την πιθανότητα πρόσληψης ενός υποψηφίου.

Σε συνέχεια με τα παραπάνω, πολλές δυναμικές τεχνικές έχουν εισαχθεί σε αυτό το πλαίσιο προκειμένου να δημιουργηθεί ευκολία στη δημιουργία προτύπων από λέξεις.

Η διαδικασία περιλαμβάνει:

- Matching of text files(αντιστοίχιση αρχείων κειμένου)
- Text mining(εξόρυξη κειμένου) περιλαμβάνει (keyword matching, sequence matching και clustering). Το Text mining ορίζεται ως η διαδικασία ανάλυσης και διερεύνησης μεγάλων συνόλων μη δομημένων δεδομένων για τον εντοπισμό και την απομόνωση θεμάτων, λέξεων-κλειδιών, εννοιών, προτύπων και άλλων στοιχείων που είναι εμφανή στα σύνολα δεδομένων. Η διαδικασία γίνεται πιο πρακτική λόγω της εμφάνισης μεγάλων πλατφορμών δεδομένων και διαφόρων αλγορίθμων βαθιάς μάθησης με την ικανότητα ανάλυσης μεγάλων συνόλων μη δομημένων δεδομένων. Μέσω της εξόρυξης κειμένων, είναι δυνατή η εστίαση σε δεδομένα κειμένου αντί για άλλες πιο δομημένες μορφές δεδομένων που εξετάζονται στην εξόρυξη δεδομένων. Ωστόσο, και οι δύο διαδικασίες περιλαμβάνουν την οργάνωση και τη δομή των δεδομένων με συγκεκριμένο τρόπο πριν από την υποβολή τους σε ποιοτική και ποσοτική ανάλυση. Η εξόρυξη κειμένων αποτελείται από ένα ευρύ φάσμα αλγορίθμων και θεμάτων για την ανάλυση κειμένων και την κάλυψη διαφόρων κοινοτήτων. Μερικά από τα κοινά εργαλεία εξόρυξης κειμένων περιλαμβάνουν text clustering, text visualization, και association rule extraction.
- Text clustering(ομαδοποίηση κειμένων): Βασίζεται στην υπόθεση συμπλέγματος που υποστηρίζει ότι η ομαδοποίηση είναι δυνατή όταν τα σχετικά έγγραφα έχουν περισσότερες ομοιότητες μεταξύ τους σε σύγκριση με τα μη σχετικά έγγραφα. Παρόλο που η ομαδοποίηση κειμένων παραμένει ζωτικής σημασίας στις διαδικασίες

περιήγησης και πλοήγησης, διάφορες στρατηγικές ομαδοποίησης κειμένων έχουν μειονεκτήματα όπως η πολυπλοκότητα του χώρου και ο υψηλός υπολογιστικός χρόνος, η μειωμένη ευρωστία(αντοχή), οι κίνδυνοι απορρήτου και η αδυναμία κατανόησης των στοιχείων των λέξεων με βάση τα συμφραζόμενα. Ως θετικό αποτελεί μια αξιόπιστη τεχνική που εφαρμόζεται σε μεγάλο βαθμό στην ανάλυση και την εξέταση μεγάλων συνόλων δεδομένων παρόμοια με την εξόρυξη δεδομένων. Ως αρνητικό στοιχείο, αποτελεί λιγότερο αποτελεσματική προσέγγιση.

- Association rule extraction(εξαγωγή κανόνων συσχέτισης): Οι προσεγγίσεις που σχετίζονται με την εξαγωγή κανόνων συσχέτισης επικεντρώνονται στον εντοπισμό των σχέσεων σε ένα μεγάλο σύνολο μεταβλητών σε σύνολα δεδομένων(Association-Rule Mining (ARM) strategies). Οι στρατηγικές ARM προσδιορίζουν τις ρυθμίσεις μεταβλητής αξίας που συμβαίνουν επανειλημμένα. Οι διαδικασίες συσχέτισης για την εξόρυξη κειμένου επικεντρώνονται στην διερεύνηση των σχέσεων μεταξύ διαφόρων πραγματικών εννοιών ή θεμάτων που χρησιμοποιούνται για την περιγραφή ενός συνόλου δεδομένων. Ο πρωταρχικός στόχος είναι να ανακαλυφθούν οι ζωτικοί κανόνες συσχετισμού συγκριτικοί με ένα σύνολο δεδομένων με τέτοιο τρόπο ώστε η ύπαρξη συγκεκριμένων θεμάτων στα άρθρα να μοιάζει με την ύπαρξη άλλων θεμάτων. Μερικοί ερευνητές υποστηρίζουν ότι η εξαγωγή κανόνων συσχέτισης παραμένει μια αποτελεσματική προσέγγιση για την εξεύρεση ακριβών και ζωτικών γνώσεων από μια ποικιλία εγγράφων. Ωστόσο, ένα σημαντικό πρόβλημα που αντιμετωπίζει η αντιστοίχιση κειμένου(text matching) είναι ότι η αναζήτηση και εύρεση ακριβών και κρίσιμων πληροφοριών μπορεί να είναι χρονοβόρα ή να οδηγήσει σε παρεξηγήσεις. Η δημιουργία μιας μεθόδου για την ανακάλυψη γνώσεων μέσω της εξαγωγής κανόνων συσχέτισης αποδεικνύεται αποτελεσματική τόσο για την εξοικονόμηση χρόνου όσο και για την αποφυγή παρεξηγήσεων.
- Word cloud(παγκόσμιο cloud): Μια άλλη τεχνική εξόρυξης κειμένου σε κοινή χρήση είναι το word cloud. Όπως σημειώθηκε, ένα σύννεφο λέξεων ή σύννεφο ετικετών(tag cloud) είναι η οπτική αναπαράσταση μιας λέξης για μια συγκεκριμένη γραπτή διάταξη περιεχομένου σύμφωνα με τη συχνότητά της. Είναι μια προκαταρκτική φάση κατά τη διεξοδική ανάλυση συγκεκριμένου υλικού κειμένου και των απαραίτητων

πληροφοριών. Είναι απλή και η πιο κοινή μέθοδος για την γραφική παρουσίαση δεδομένων κειμένου. Για μαθήματα γραφής και ανάγνωσης, τα σύννεφα λέξεων είναι ζωτικής σημασίας για τη μείωση του χρόνου ανάγνωσης και τη βοήθεια με το λεξιλόγιο και την κατανόηση γραφής. Επιπλέον, τα σύννεφα λέξεων χρησιμοποιούνται επίσης για την αντιμετώπιση ρητορικών στοιχείων στις περιγραφές των καθηκόντων των εκπαιδευτικών, στα προγράμματα σπουδών και στις εκθέσεις των μαθητών. Ωστόσο, αποτυγχάνει να μελετήσει τη γλωσσική τεχνογνωσία σχετικά με τις λέξεις κειμένου και τις αντίστοιχες συσχετίσεις με ένα συγκεκριμένο θέμα όταν προσφέρει μια αριθμητική περίληψη στις απομονωμένες λέξεις. Ως αποτέλεσμα αυτού, στην πλειονότητα των συστημάτων, τα σύννεφα λέξεων χρησιμοποιούνται συνήθως με αριθμητικό τρόπο για να συνοψίσουν τα κείμενα και να προσφέρουν ελάχιστη ή καθόλου πιθανότητα να συσχετιστούν τα δεδομένα.

- Natural language processing-NLP(επεξεργασίας φυσικής γλώσσας): Το NLP είναι ένα υπο-πεδίο της τεχνητής νοημοσύνης και της επιστήμης των υπολογιστών που ασχολείται με αλληλεπιδράσεις και συσχετίσεις μεταξύ ανθρώπινων και υπολογιστικών γλωσσών. Είναι επίσης εφαρμόσιμο σε αλγόριθμους μηχανικής μάθησης για την αποκρυπτογράφηση κειμένων και ομιλίας. Έχει αποκτήσει ιδιαίτερη προσοχή στον τομέα της επιστήμης των υπολογιστών και ειδικότερα της αντιστοίχισης κειμένων. Σε αντίθεση με τους συνηθισμένους ρόλους μηχανικής μάθησης που ασχολούνται με δομημένα δεδομένα, ασχολείται με δεδομένα μη δομημένου κειμένου. Οι αλγόριθμοι NLP βασίζονται όπως προαναφέραμε συνήθως σε αλγόριθμους μηχανικής μάθησης. Αντί για χειροκίνητη κωδικοποίηση μεγάλων συνόλων κανόνων, το NLP μπορεί να βασιστεί στη μηχανική εκμάθηση για να μάθει αυτόματα αυτούς τους κανόνες αναλύοντας ένα σύνολο παραδειγμάτων. Επίσης, ένας από τους τύπους NLP που ονομάζεται ανάλυση συναισθημάτων, είναι μια νέα υβριδική προσέγγιση για την ανάλυση των κοινωνικών προβλημάτων. Οι αλγόριθμοι αναζήτησης συμβολοσειρών είναι τα πιο κοινά εργαλεία NLP για χρήση στην αντιστοίχιση κειμένου. Μερικοί από τους πιο κοινούς αλγόριθμους είναι Brute Force, Boyer-Moore Horspool, Rabin-Karp και Knuth-Morris-Pratt.
- Text similarity measures(μέτρα ομοιότητας κειμένου) (knowledge-based, corpus-based, string-based, and hybrid similarities): Η προσέγγιση string-based

αποτελεί την πιο εμφανή μορφή εξόρυξης κειμένου που εφαρμόζεται σε διαφορετικές περιπτώσεις. Υπάρχουν τέσσερις μεγάλες κατηγορίες μέτρων ομοιότητας κειμένου. Το μέτρο ομοιότητας βάσει συμβολοσειράς είναι μια από τις παλαιότερες μεθόδους. Οι αλγόριθμοι string-based για την αντιστοίχιση της λεξικής ομοιότητας εμπίπτουν σε δύο κατηγορίες. Η πρώτη κατηγορία αποτελείται από αλγόριθμους που βασίζονται σε χαρακτήρες όπως LCS, Damerau Levenshtein, Jaro και N gram. Η δεύτερη κατηγορία αποτελείται από term-based αλγόριθμους, όπως Cosine similarity, Block Distance, Euclidean Distance, Jaccard similarity, and Matching Coefficient. Η σημασιολογική ανάλυση χωρίζεται σε Corpus-based και knowledge-based αλγόριθμους. Οι τεχνικές αντιστοίχισης ομοιότητας Corpus-based έχουν αποκτήσει ευρεία εφαρμογή. Η πιο δημοφιλής τεχνική μοντελοποίησης θεμάτων είναι η λανθάνουσα σημασιολογική ανάλυση (LSA-Latent Semantic Analysis), η οποία υποθέτει ότι λέξεις με παρόμοιο νόημα εμφανίζονται σε παρόμοια κομμάτια κειμένου. Είναι μια τεχνική στην επεξεργασία φυσικής γλώσσας η οποία χρησιμοποιεί clustering. Παρόλο που το LSA είναι εύκολο να εφαρμοστεί και προσφέρει καλή απόδοση, έχει εγγενείς περιορισμούς για μη γραμμικές καταστάσεις λόγω της γραμμικής προσέγγισής του και το πιθανοτικό μοντέλο ενδέχεται να μην ταιριάζει με τα παρατηρούμενα δεδομένα. Οι αλγόριθμοι που χρησιμοποιούνται για την αντιστοίχιση της εργασιακής εμπειρίας και δεξιοτήτων του χρήστη με τα έγγραφα περιγραφής εργασίας είναι Latent Semantic Analysis (LSA) και Latent Dirichlet Allocation (LDA) που εφαρμόζονται χρησιμοποιώντας το Gensim. Το LDA είναι ένα διαδικτυακό μοντέλο μάθησης, το οποίο διασφαλίζει την online ενημέρωση (ενημέρωση σε παρτίδες) του πίνακα / διανυσμάτων κάθε φορά που μια νέα παρτίδα εγγράφων απορροφάται από το μοντέλο. Μια εναλλακτική μέθοδος είναι το Latent Semantic Indexing (LSI), το οποίο ξεπερνά τη συνωνυμία. Ο όρος LSI χρησιμοποιείται για εργασίες που αφορούν την ευρετηρίαση ή την ανάκτηση πληροφοριών, ενώ ο όρος LSA χρησιμοποιείται για εργασίες που σχετίζονται με την ανάλυση κειμένων όπως η αυτόματη βαθμολόγηση έκθεσης και η σύνοψη κειμένων. Το LSA λειτουργεί καλά στις προτάσεις καθώς διατηρεί τη σειρά των λέξεων στην πρόταση. Λόγω της ποικιλίας των λέξεων που χρησιμοποιούνται για την περιγραφή των βιογραφικών σημείων και των θέσεων εργασίας, δηλαδή συνώνυμα, τα βιογραφικά και οι μέθοδοι αντιστοίχισης θέσεων εργασίας συχνά αποτυγχάνουν να ανακτήσουν τις πληροφορίες που

χρησιμοποιούνται για τη μέτρηση της πραγματικής ομοιότητας. Μπορούν επίσης να χρησιμοποιηθούν και άλλα μέτρα ομοιότητας, όπως η ομοιότητα Jaccard. Ως εκ τούτου, το LSA είναι καλύτερο να ερμηνεύσει σημασιολογικά, σε αντίθεση με το LDA που προσπαθεί να δημιουργήσει θέματα και στη συνέχεια να υπολογίσει την ομοιότητα. Οι σημασιολογικές μέθοδοι προέρχονται από την ιδέα ότι μια περιγραφή μπορεί να παρέχεται με πολλούς τρόπους. Μια άλλη τεχνική corpus-based είναι η Explicit Semantic Analysis (ESA). Παρόλο που το ESA έχει αποδειχθεί χρήσιμο στο χειρισμό διαφορετικών πλαισίων, βρέθηκε ότι ο αλγόριθμος εμφανίζει απροσδόκητη συμπεριφορά. Οι προσεγγίσεις knowledge-based βασίζονται σε πληροφορίες από σημασιολογικά δίκτυα για τον προσδιορισμό του βαθμού ομοιότητας. Το πιο δημοφιλές knowledge based εργαλείο είναι το WordNet, το οποίο αποτελείται από μια λεξική βάση δεδομένων αγγλικών λέξεων που συνδέονται με εννοιολογικές σχέσεις. Πρόσφατα, έχουν γίνει προσπάθειες συνδυασμού πολλαπλών προσεγγίσεων. Το εργαλείο SimAll συνδυάζει χαρακτηριστικά string, knowledge, and corpus similarity. Αποτελεί το καλύτερο εργαλείο για τη σημασιολογική ανάλυση, επειδή συνδυάζει τη λεξική και τη σημασιολογική ανάλυση σε διαφορετικά επίπεδα βαθμού λεπτομέρειας(ευαισθησίας). Το TextRank είναι ένα graph-based μοντέλο για την επεξεργασία κειμένου, το οποίο μπορεί να χρησιμοποιηθεί για την εύρεση των πιο σχετικών προτάσεων στο κείμενο και επίσης για την εύρεση λέξεων-κλειδιών. Στον LexRank αλγόριθμο όλα τα ουσιαστικά και τα επίθετα διαχωρίζονται για να σχηματίσουν ένα σύμπλεγμα εγγράφων. Το LexRank είναι μια μη εποπτευόμενη προσέγγιση για τη σύνοψη κειμένων και είναι μια μέθοδος που βασίζεται σε γραφήματα. Η κύρια ιδέα είναι ότι οι προτάσεις "προτείνουν" άλλες παρόμοιες προτάσεις στον αναγνώστη. Έτσι, εάν μια πρόταση είναι πολύ παρόμοια με πολλές άλλες, πιθανότατα θα είναι μια πρόταση μεγάλης σημασίας. Μια κοινή μέτρηση ανάκτησης πληροφοριών, είναι το tf-idf(Το TF-IDF είναι ένα στατιστικό μέτρο που αξιολογεί πόσο σχετική είναι μια λέξη με ένα έγγραφο σε μια συλλογή εγγράφων. Αυτό γίνεται πολλαπλασιάζοντας δύο μετρήσεις: πόσες φορές μια λέξη εμφανίζεται σε ένα έγγραφο και τη αντίστροφη συχνότητα του κειμένου της λέξης σε ένα σύνολο εγγράφων)

1.5.Recommender Systems

Τα συστήματα σύστασης είναι μια υποκατηγορία συστημάτων φιλτραρίσματος πληροφοριών. Οι αλγόριθμοι που χρησιμοποιούν τα συστήματα σύστασης εξοικονομούν χρόνο και χρήμα στις εταιρείες, μειώνοντας αρκετά το κόστος εύρεσης του ιδανικού υποψηφίου. Το μόνο που χρειάζεται να κάνουν οι εταιρείες είναι να τροφοδοτούν τα δεδομένα του λογισμικού για διαφορετικές παραμέτρους μιας θέσης εργασίας, όπως είναι τα χρόνια εργασιακής εμπειρίας, η γνώση των γλωσσών προγραμματισμού, η εκπαίδευση, οι ξένες γλώσσες και άλλα. Ο εκάστοτε αλγόριθμος θα περάσει από προφίλ στο LinkedIn ή σε άλλες δημόσιες βάσεις δεδομένων και θα βρει βιογραφικά που ταιριάζουν καλύτερα στις πληροφορίες που παρέχονται από τους υπεύθυνους πρόσληψης. Μόλις γίνει αυτό, τα επιλεγμένα βιογραφικά αποστέλλονται αυτόματα στους υπεύθυνους προσλήψεων (HR managers), ώστε να μπορούν να κάνουν έναν χειροκίνητο έλεγχο και στη συνέχεια να περάσουν στο στάδιο των συνεντεύξεων. Οι ερευνητές που σχεδιάζουν συστήματα σύστασης συχνά τα ταξινομούν σε δύο κύριες προσεγγίσεις, δηλαδή content-based filtering και collaborative filtering. Οι content-based μέθοδοι προβλέπουν προτιμήσεις σχετικά με ένα σύνολο στοιχείων με βάση την ομοιότητά τους με τα στοιχεία που ένας χρήστης έχει προτιμήσει στο παρελθόν (δηλαδή με το προφίλ προτιμήσεων ενός χρήστη). Εξελεγχμένες μέθοδοι χρησιμοποιούν τεχνικές μηχανικής εκμάθησης όπως Bayesian Classifiers, cluster analysis, decision trees, and artificial neural networks. Οι δημοφιλείς προσεγγίσεις opinion-based χρησιμοποιούν διάφορες τεχνικές όπως text mining, information retrieval, sentiment analysis και deep learning. Στο collaborative filtering τα στοιχεία συνιστώνται σε έναν χρήστη με βάση τις προτιμήσεις παρόμοιων χρηστών. Συνεπώς τα μοντέλα αυτά κάνουν προβλέψεις σχετικά με τα ενδιαφέροντα των χρηστών βάσει των ιστορικών αλληλεπιδράσεων στοιχείων-χρηστών είτε με άμεσα / έμμεσα σχόλια. Επομένως, στην πρώτη περίπτωση, τα προφίλ προτιμήσεων των χρηστών πρέπει να συγκρίνονται με τις περιγραφές των στοιχείων και στην τελευταία περίπτωση, πρέπει να προσδιοριστεί η ομοιότητα μεταξύ των προφίλ των χρηστών. Οι πιο συνηθισμένες επιλογές είναι η ομοιότητα Jaccard και συνημίτονο. Μεταξύ άλλων δημοφιλών μέτρων, υπάρχει ο συσχετισμός Pearson, kNN και η τυπική απόσταση Ευκλείδειας (αντιμετωπίζεται ως μέτρο ανομοιότητας. Νέα μοντέλα μπορούν να κατασκευαστούν σε περιβάλλον R, Python. (Janusz et al., 2018). Υπάρχουν επίσης οι υβριδικές μέθοδοι σύστασης οι οποίες είναι οι προσεγγίσεις που

ενσωματώνουν δύο ή περισσότερους τύπους στρατηγικών προτάσεων. Οι Yao et al. παρουσίασε ένα hybrid recommender system που ενσωμάτωσε τη σχέση content-based και interaction-based, για τα μη δομημένα δεδομένα, όπως ομοιότητα μεταξύ εργασίας και προφίλ χρήστη, η λανθάνουσα σημασιολογική ανάλυση εφαρμόζεται στο σύστημα.

Πιο συμπυκνωμένα:

- Content-based Recommendation (CBR) Η αρχή μιας σύστασης βάσει περιεχομένου είναι να προτείνει στοιχεία που έχουν παρόμοιες πληροφορίες περιεχομένου με τους αντίστοιχους χρήστες, όπως το Prospect.
- Collaborative Filtering Recommendation (CFR) Η πρόταση συνεργατικού φιλτραρίσματος βρίσκει παρόμοιους χρήστες που έχουν τις ίδιες προτιμήσεις με τον χρήστη-στόχο και προτείνει στοιχεία βασισμένα στις προτιμήσεις παρόμοιων χρηστών, όπως το CASPER.
- Knowledge-based Recommendation (KBR) Στη σύσταση που βασίζεται στη γνώση, κανόνες και μοτίβα που λαμβάνονται από τη λειτουργική γνώση του τρόπου με τον οποίο ένα συγκεκριμένο αντικείμενο πληροί τις απαιτήσεις ενός συγκεκριμένου χρήστη, χρησιμοποιούνται για τη σύσταση στοιχείων, όπως το Proactive.
- Hybrid recommender systems Τα υβριδικά συστήματα σύστασης συνδυάζουν δύο ή περισσότερες τεχνικές προτάσεων για να αποκτήσουν καλύτερη απόδοση και να ξεπεράσουν τα μειονεκτήματα μεμονωμένων τεχνικών. Συνήθως, το συλλογικό φιλτράρισμα συνδυάζεται με κάποια άλλη τεχνική. (σε μια προσπάθεια να αποφευχθεί το πρόβλημα ramp-up).

1.6.Deep Learning Models for Text Classification and Matching

Για εργασίες αντιστοίχισης κειμένου, τα μοντέλα βαθιάς μάθησης μπορούν να χωριστούν σε δύο κύριες κατηγορίες: representation-focused model και interaction-focused model. Συνήθως, τα μοντέλα που εστιάζουν στην αναπαράσταση δημιουργούν πρώτα την αναπαράσταση προτάσεων με νευρωνικά δίκτυα και στη συνέχεια μετρούν τον βαθμό αντιστοίχισης μέσω διαφορετικών τύπων λειτουργιών βαθμολογίας. Αντίθετα, τα μοντέλα που εστιάζουν στην αλληλεπίδραση επιβάλλουν την εξαγωγή των τοπικών διαδραστικών

χαρακτηριστικών με βάση τις βασικές αναπαραστάσεις λέξεων ή φράσεων, ακολουθούμενα από τα βαθιά νευρωνικά δίκτυα για να μάθουν τα περίπλοκα μοτίβα αλληλεπίδρασης για αντιστοίχιση.

Το Recurrent Neural Network (RNN) ωφελεί την εργασία ταξινόμησης κειμένου, καθώς μπορεί να συλλάβει τις πληροφορίες σειριοποίησης και να μάθει τις σχέσεις μακράς διάρκειας ή την παγκόσμια σημασιολογική αναπαράσταση. Οι Tai et al. [2015] εισήγαγε ένα ειδικό δίκτυο RNN - συγκεκριμένα, δέντρο-βραχυπρόθεσμης μνήμης (LSTM) - για ταξινόμηση συναισθημάτων. Παραλλαγή του LSTM αποτελεί το BiLSTM το οποίο είναι σε θέση σε οποιοδήποτε σημείο της σειράς να διατηρήσει πληροφορίες τόσο από το παρελθόν όσο και από το μέλλον. Το BiLSTM μπορεί να κατανοήσει καλύτερα το περιβάλλον από το LSTM και να δείξει τα πλεονεκτήματά του σε πολλές διαφορετικές εργασίες εξόρυξης κειμένου, όπως ανάλυση συναισθημάτων [Chen et al. 2017a], ομοιότητα πρότασης [Neculoiu et al. 2016] και αναγνώριση οντοτήτων ονόματος [Lample et al. 2016].

1.7. Transfer Learning Methods

Ενώ οι οργανισμοί προσλήψουν διαθέτουν δεδομένα σχετικά με τα βιογραφικά των αιτούντων και τις περιγραφές διαφημίσεων εργασίας, τα βιογραφικά είναι συχνά εμπιστευτικά, περιορίζοντας έτσι τη χρήση μεθόδων άμεσης βαθιάς μάθησης. Ωστόσο, με την έλευση των μεθόδων εκμάθησης μεταφοράς (transfer learning methods), είναι δυνατόν να εκπαιδεύσουμε γλωσσικά μοντέλα ανεξάρτητα από την εργασία που υπάρχει, και συνεπώς ανεξάρτητα από τα διαθέσιμα δεδομένα. Καθώς το NLI μπορεί να ισχύει για αυθαίρετα ζεύγη προτάσεων, εξετάζεται επίσης η χρήση τμηματοποίησης κειμένου για την ενίσχυση της απόδοσης των μεθόδων. Τα αποτελέσματα δείχνουν ότι η προσέγγιση NLI είναι σημαντικά καλύτερη από έναν τυχαίο ταξινομητή καταλληλότητας (suitability classifier). (Τα Support Vector Machine (SVM) είναι εποπτευόμενα μοντέλα μάθησης με συναφείς αλγόριθμους μάθησης που αναλύουν δεδομένα για ταξινόμηση και ανάλυση παλινδρόμησης. για την ταξινόμηση διαφορετικών τμημάτων μιας αγγελίας εργασίας).

1.8.Αλγόριθμοι Εξαγωγής Πληροφοριών

Τα Linear-chain Conditional Random Fields (CRFs) σχετίζονται στενά με τα μοντέλα Hidden Markov (HMMs) και συγκαταλέγονται στους αλγόριθμους εξαγωγής πληροφοριών με την καλύτερη απόδοση. Σε αντίθεση με τα HMM, τα CRF έχουν αμφίδρομες εξαρτήσεις.

1.9. Comparison Methods

Για την αξιολόγηση ενός τομέα, θεωρούμε τρεις βασικές μεθόδους ως βασικές γραμμές.

- DSSM (Huang et al., 2013): χρησιμοποιεί ένα βαθύ νευρωνικό δίκτυο (DNN) για να χαρτογραφήσει αραιά χαρακτηριστικά υψηλών διαστάσεων σε πυκνά χαρακτηριστικά χαμηλής διάστασης και υπολογίζει τη σημασιολογική ομοιότητα του ζεύγους κειμένων
- BPJFNN (Qin et al., 2018): Εφαρμόζει τα BiLSTM για τη λήψη της σημασιολογικής αναπαράστασης κάθε λέξης σε αγγελίες εργασίας και βιογραφικά και θεωρεί το περιεχόμενο του κειμένου ως μία μόνο ακολουθία.
- PJFNN (Zhu et al., 2018): Προτείνει ένα bipartite Convolutional Neural Network που μπορεί να μάθει αποτελεσματικά την κοινή αναπαράσταση της ικανότητας PersonJob από ιστορικές εφαρμογές(applications) εργασίας.
- APJFNN (Qin et al., 2018): Μαθαίνει μια σημασιολογική αναπαράσταση σε επίπεδο λέξεων τόσο για τις απαιτήσεις εργασίας όσο και για τα βιογραφικά που βασίζονται σε RNN και τέσσερις ιεραρχικές στρατηγικές προσοχής σχετικά με την ικανότητα.

Μεταξύ των βασικών γραμμών, η μέθοδος APJFNN αποδίδει καλύτερα από τις άλλες μεθόδους.

Το σύστημα E-Gen πραγματοποιεί ανάλυση και κατηγοριοποίηση των μη δομημένων προσφορών εργασίας (δηλαδή, με τη μορφή εγγράφων χωρίς δομή), καθώς και ανάλυση και κατάταξη των υποψηφίων. Το πλαίσιο CommOn εφαρμόζει τεχνολογίες Σημασιολογικού Ιστού στον τομέα της Διαχείρισης Ανθρώπινου Δυναμικού. Σε αυτό το πλαίσιο, τα χαρακτηριστικά προσωπικότητας του υποψηφίου, που καθορίζονται μέσω ενός διαδικτυακού ερωτηματολογίου που συμπληρώνεται από τον υποψήφιο, εξετάζονται για πρόσληψη. Προκειμένου να ταιριάζουν με τους υποψηφίους με θέσεις εργασίας, τα συστήματα ηλεκτρονικής πρόσληψης συνήθως συνδυάζουν τεχνικές από κλασικά συστήματα IR και συστημάτων σύστασης, όπως relevance feedback, semantic matching and Analytic

Hierarchy Process. Αυτές οι μέθοδοι, αν και χρήσιμες, υποφέρουν από τις αποκλίσεις που σχετίζονται με ασυνεπείς μορφές βιογραφικού, δομή του βιογραφικού και πληροφορίες που προκύπτουν με βάση τα συμφραζόμενα. Επιπλέον, δεν είναι σε θέση να αξιολογήσουν κάποια δευτερεύοντα χαρακτηριστικά που σχετίζονται με βιογραφικά σημειώματα, όπως το στυλ και τη συνοχή, τα οποία είναι πολύ σημαντικά στην αξιολόγηση βιογραφικού. Οι μέθοδοι που μαθαίνουν πώς να συνδυάζουν προκαθορισμένα χαρακτηριστικά για την κατάταξη μέσω εποπτευόμενων αλγορίθμων μάθησης ονομάζονται μέθοδοι «μάθηση-κατάταξη(learning-to-rank)». Οι δυνατότητες του υποψηφίου μπορούν να αξιολογηθούν είτε σε αριθμητική κλίμακα (π.χ. έτη επαγγελματικής εμπειρίας) είτε με μια μεταβλητή Boolean, η οποία αντιπροσωπεύει εάν ο υποψήφιος αναφέρει μια συγκεκριμένη ικανότητα ή όχι στο προφίλ του στο LinkedIn. Το σετ εκπαίδευσης τροφοδοτείται σε έναν αλγόριθμο μάθησης που κατασκευάζει το μοντέλο κατάταξης, έτσι ώστε η έξοδος του να προβλέπει την κρίση του προσλήπτη όταν δίνεται το διάνυσμα χαρακτηριστικών του υποψηφίου ως είσοδος. Τα χαρακτηριστικά της προσωπικότητας των αιτούντων είναι κρίσιμα για την επιλογή τους σε πολλές θέσεις εργασίας, αλλά συνήθως παραβλέπονται στα υπάρχοντα συστήματα ηλεκτρονικής πρόσληψης. Συνήθως, η προσωπικότητα των υποψηφίων αξιολογείται κατά τη διάρκεια της φάσης της συνέντευξης, η οποία προορίζεται για τους υποψηφίους που πέρασαν τη φάση πριν από τον έλεγχο. Ωστόσο, η συλλογή ορισμένων προκαταρκτικών δεδομένων για την προσωπικότητα του υποψηφίου στη φάση πριν από τον έλεγχο θεωρείται πολύτιμη και αυτές οι πληροφορίες λαμβάνονται συχνά μέσω αναζητήσεων στο διαδίκτυο. Προηγούμενα έργα έχουν δείξει ότι με την εφαρμογή γλωσσικής ανάλυσης σε αναρτήσεις ιστολογίου, μπορούν να εξαχθούν τα χαρακτηριστικά της προσωπικότητας του συγγραφέα, καθώς και η διάθεση και τα συναισθήματά του. Η ανάλυση κειμένου σε αυτά τα έργα φαίνεται να πραγματοποιείται με το σύστημα LIWC (Linguistic Inquiry and Word Count), το οποίο εξάγει γλωσσικά χαρακτηριστικά που λειτουργούν ως δείκτες της προσωπικότητας του συγγραφέα. Το LIWC χρησιμοποιεί ένα λεξικό στελεχών λέξεων που ταξινομούνται σε ορισμένες ψυχολογικές σημασιολογικές και συντακτικές κατηγορίες λέξεων. Αναλύει γραπτά δείγματα κειμένου μετρώντας τις σχετικές συχνότητες των λέξεων που εμπίπτουν σε κάθε κατηγορία λέξεων. Οι Pennebaker και King έχουν βρει σημαντικές συσχετίσεις μεταξύ αυτών των μετρήσεων συχνότητας και των χαρακτηριστικών της προσωπικότητας του συγγραφέα, όπως μετράται από τις

διαστάσεις της προσωπικότητας Big-Five. Οι αναρτήσεις ιστολογίου εισάγονται στο εργαλείο TreeTagger για λεξική ανάλυση και λεματοποίηση.

2. Ηλεκτρονικά Συστήματα Ευρέσεως Εργασίας

2.1.Recruitment Platforms

Οι πέντε πιο εντυπωσιακές πλατφόρμες πρόσληψης σύμφωνα με τις προτιμήσεις των χρηστών είναι οι (e-jobs, hipo, bestjobs, LinkedIn, Freelancer). Οι διαδικασίες πρόσληψης και επιλογής ανθρώπινου δυναμικού είναι ζωτικής σημασίας για έναν οργανισμό που θέλει όχι μόνο να επιβιώσει στην αγορά αλλά και να αυξήσει τη θέση του στην αγορά στην οποία δραστηριοποιείται. Η γεωγραφική θέση δεν έχει όριο πια. Δεν περιορίζει κανέναν να υποβάλει αίτηση για εργασία σε άλλη χώρα ή ακόμα και ήπειρο. Το E-Recruitment επιτρέπει την άμεση αλληλεπίδραση σε πραγματικό χρόνο και πρόσληψη / αναζήτηση εργασίας. Επίσης, οι δημοσιεύσεις είναι διαθέσιμες για περίπου 30 ημέρες και μπορούν να συνεχίσουν να λαμβάνουν αιτούντες μετά τη λήξη τους. Το Bayt.com μπορεί να προσφέρει έως και 33 φίλτρα για να διακρίνει τους υποψηφίους. Η SAT telecom ανέφερε 44% εξοικονόμηση κόστους και μείωση του μέσου χρόνου που απαιτείται για να καλυφθεί μια κενή θέση από 70 σε 37 ημέρες μετά την ανάπτυξη ενός recruitment συστήματος. Η νεολαία προσελκύεται περισσότερο από τις διαδικτυακές προσλήψεις και τη χρησιμοποιεί συχνότερα από τους ηλικιωμένους. Αυτό οφείλεται στο γεγονός ότι οι νέοι αναζητούν νέες θέσεις εργασίας και επειδή βρίσκονται σε συνεχείς αλλαγές για να βρουν την καλύτερη δουλειά για αυτούς. Οι πιο γνωστές πλατφόρμες φαίνεται να είναι το LinkedIn και bestjobs.ro. Και οι δύο έχουν χαρακτηριστικά που είναι σημαντικά για τους αιτούντες εργασία. Οι περισσότεροι από τους χρήστες εξέθεσαν την προτίμησή τους να διαθέτουν μια δυνατότητα που τους ενημερώνει πραγματικά για νέες ευκαιρίες. Ένα εξαιρετικό χαρακτηριστικό που μπορεί να εφαρμόσει ένας ιστότοπος ηλεκτρονικής πρόσληψης είναι η λειτουργικότητα ειδοποίησης που εκτιμάται ιδιαίτερα από τους περισσότερους από τους ερωτηθέντες. Ο ιστότοπος πρέπει να διαθέτει ένα σύστημα περιοδικών ειδοποιήσεων που να ενημερώνει τους χρήστες για νέες ευκαιρίες ή μηνύματα που έχουν λάβει από τους ενδιαφερόμενους εργοδότες. Το LinkedIn είναι το πιο ισχυρό σε αυτό το κεφάλαιο με τις περισσότερες ψήφους, με βάση το γεγονός ότι διαθέτουν ένα πραγματικά αποτελεσματικό σύστημα ειδοποιήσεων για νέες εργασίες, νέα μηνύματα και νέα.

2.2.Web Based Systems

Ένα web based σύστημα είναι μια εφαρμογή στην οποία υπάρχει πρόσβαση μέσω HTTP. Στο πεδίο εφαρμογής τους τα συστήματα web based βρίσκουν εφαρμογή σε πληθώρα ιστοσελίδων που αφορούν την διαδικτυακή αναζήτηση εργασίας, είτε από την πλευρά των εταιριών είτε από αυτή των ατόμων που αναζητούν εργασία. Ο πρώτος ιστότοπος αυτής της μορφής στον οποίο θα αναφερθούμε είναι ίσως και ο πιο γνωστός το LinkedIn. Ιδρύθηκε στην Αμερική το 2002. Η δημιουργία λογαριασμού για τα άτομα είναι δωρεάν, εκτός αν θέλουν την premium εφαρμογή. Αντίθετα οι εταιρίες χρειάζεται να πληρώσουν ακόμα και για τη δημοσίευση μιας αγγελίας. Δεν έχει σημασία αν κάποιος είναι απλά φοιτητής που ακόμα δεν αναζητά εργασία ή ιδιοκτήτης κάποιου μικρού καταστήματος, το LinkedIn είναι για οποιονδήποτε ενδιαφέρεται να πάρει την επαγγελματική του ζωή πιο σοβαρά αναζητώντας νέες ευκαιρίες για την ανάπτυξη της καριέρας του. Ουσιαστικά το LinkedIn είναι μια πλατφόρμα κοινωνικών μέσων που απευθύνεται σε επαγγελματίες. Προωθεί την δικτύωση και δημιουργία ενός επαγγελματικού χαρτοφυλακίου. Όπως και άλλα δίκτυα κοινωνικών μέσων, το LinkedIn χρησιμοποιεί έναν αλγόριθμο που ενημερώνεται συνεχώς έτσι ώστε να παρέχει στους χρήστες αυτό που αποφασίζει σαν καλύτερη δυνατή εμπειρία, δηλαδή στην συγκεκριμένη περίπτωση θέσεις εργασίας που αντιστοιχούν σε άτομα. Ο χρήστης μπορεί να παρατηρήσει ότι η "ροή" του, η οποία λειτουργεί όπως η ροή στο Facebook, παρέχει ένα συνεχώς ανανεωμένο περιεχόμενο βάσει ενδιαφερόντων (Cunnington, 2019). Μάλιστα είναι σημαντικό να αναφέρουμε εδώ ότι το Facebook έχει αρχίσει τα τελευταία χρόνια να χρησιμοποιεί έναν αλγόριθμο αντίστοιχο με αυτόν του LinkedIn, αφού πλέον μπορούν να δημοσιεύονται θέσεις εργασίας στον ιστότοπό του τις οποίες προβάλλει ως διαφήμιση σε άτομα που οι σπουδές και οι δεξιότητες που έχουν δηλώσει στο προφίλ τους ταιριάζουν με την θέση (Qin et al., 2019). Προτού όμως φτάσει αυτή η ροή στο προφίλ, περνά μέσα από φίλτρα. Μερικά από αυτά τα φίλτρα αξιολόγησης είναι δομημένα με τέτοιο τρόπο ώστε να αναζητούν ανεπιθύμητο περιεχόμενο, χαμηλής ποιότητας περιεχόμενο ή να παρέχουν τη δυνατότητα να επιτρέπουν το περιεχόμενο. Αυτό που εμφανίζεται είναι πιθανό να έχει αξιολογηθεί ως καλό από ένα bot υπολογιστή, δηλαδή έναν κεντρικό υπολογιστή. Η ταχεία ανάπτυξη του LinkedIn τα τελευταία χρόνια αντανακλάται στο γεγονός ότι πλέον υπάρχουν 590 εκατομμύρια μοναδικοί χρήστες από 200 χώρες και 11 εκατομμύρια θέσεις εργασίας από διάφορες εταιρίες ανά τον κόσμο. Λόγω της σπουδαιότητας της αντιστοίχισης ατόμου με

εργασία, πολλοί έχουν καταβάλει προσπάθειες για τη βελτίωση του αλγόριθμου αντιστοίχισης. Με βάση ένα σύνολο δεδομένων με ετικέτα (δηλ. αρχεία αντιστοίχισης προσωπικού-εργασίας), ο αλγόριθμος στοχεύει στην πρόβλεψη της ετικέτας αντιστοίχισης με βάση το κείμενο των θέσεων εργασίας και των βιογραφικών των υποψηφίων (Bian et al., 2019). Ουσιαστικά μέσω του αλγόριθμου αντιστοίχισης έχει δημιουργηθεί ένα μοτίβο για την ομοιότητα βιογραφικών με θέσεις εργασίας.

2.3.Πλατφόρμες Εύρεσης Εργασίας

2.3.1. Freelancers

Η προοπτική της καριέρας δίχως όρια (freelance) αναπτύχθηκε πραγματικά στη δεκαετία του 1990, ειδικά μέσω των προσπαθειών των (DeFillippi & Arthur, 1994, 1996) και (Arthur & Rousseau, 1996) που καθόρισαν την καριέρα χωρίς όρια ως «ακολουθίες ευκαιριών εργασίας που ξεπερνούν τα όρια του single» ρύθμιση της απασχόλησης ». Στα μέσα της δεκαετίας του 1990, (Arthur, Claman, & DeFillippi, 1995) έδειξαν ισχυρούς δεσμούς μεταξύ (Quinn, 1992) ευφυών επιχειρήσεων και της έννοιας της επαγγελματικής απασχόλησης χωρίς όρια. Υποστηρίζουν ότι η δημιουργία της ευφυούς επιχείρησης απαιτεί έξυπνες ενασχολήσεις, οι οποίες βασίζονται σε αρχές από την έννοια της καριέρας δίχως όρια. Οι (Marler, Woodard Barringer, & Milkovich, 2002) επιπλέον υποστηρίζουν, ότι εμφανίζεται ένας νέος τύπος εργαζομένου: ο χωρίς όρια προσωρινός εργαζόμενος. Αυτό που διακρίνει αυτά τα όρια είναι η προτίμησή του εργαζομένου για προσωρινή εργασία και το υψηλό επίπεδο δεξιοτήτων και εμπειρίας του.

Η καριέρα του χαρτοφυλακίου είναι ίσως η πιο ακραία, απροσδόκητη μορφή της σύγχρονης καριέρας χωρίς όρια. Μια καριέρα χαρτοφυλακίου συνδέεται με μια ισχυρή προτίμηση κινητικότητας. Η ατομική πορεία καριέρας δεν επηρεάζεται από οποιεσδήποτε οργανωτικές δραστηριότητες διαχείρισης επαγγελματικής ενασχόλησης, όπως: προγράμματα καθοδήγησης και κατάρτισης ή άλλες μορφές οργανωτικής υποστήριξης. Σε αυτόν τον «ελεύθερο» κόσμο, μόνο οι ατομικές επιλογές και οι ευκαιρίες αγοράς καθορίζουν την καριέρα κάποιου και την καριέρα του. Ο Charles Handy ονόμασε αυτούς τους υπαλλήλους εργαζόμενους χαρτοφυλακίου, καθώς δημιουργούν ένα χαρτοφυλάκιο εργασίας για τον εαυτό τους.

Οι ελεύθεροι υπάλληλοι, γνωστοί και ως ανεξάρτητοι επαγγελματίες, είναι άτομα που εργάζονται μόνοι τους, χωρίς μακροπρόθεσμη συμβατική δέσμευση σε κανέναν εργοδότη.

Ένας ελεύθερος επαγγελματίας συνήθως εκτελεί υπηρεσίες ή ολοκληρώνει εργασίες βάσει βραχυπρόθεσμων συμβάσεων με διάφορους εργοδότες ή πελάτες, οι οποίοι έχουν το δικαίωμα να ελέγχουν μόνο το τελικό αποτέλεσμα της εργασίας του ατόμου, και όχι τα συγκεκριμένα μέσα που χρησιμοποιούνται για την ολοκλήρωση της εργασίας. Παραδείγματα θέσεων που κατέχουν ανεξάρτητοι επαγγελματίες κυμαίνονται από γιατρούς και προγραμματιστές υπολογιστών έως υπηρέτριες και εργαζόμενους σε αγροκτήματα. Η ελεύθερη απασχόληση μπορεί να προσφέρει πολλά πλεονεκτήματα σε άτομα, συμπεριλαμβανομένων των ευέλικτων εργασιακών ρυθμίσεων, της ανεξαρτησίας, της ποικιλίας και ορισμένων φορολογικών εκπτώσεων. Μπορεί επίσης να έχει κάποιες παγίδες, όπως η ανάληψη κινδύνου σε επιχειρηματικές συναλλαγές, η πληρωμή φόρων αυτοαπασχόλησης και η ανάληψη προσωπικής ευθύνης για την ασφάλιση υγείας, την αναπηρία και την κάλυψη συνταξιοδότησης.

Προς το παρόν, οι ακαδημαϊκοί ερευνητές χρησιμοποιούν συνήθως τον όρο ελεύθερος επαγγελματίας για να αναφερθούν σε εργαζόμενους για λογαριασμό προσωπικού στη ψυχαγωγία και τα μέσα ενημέρωσης, συμπεριλαμβανομένης της δημοσιογραφίας (Baines, 1999), της τηλεόρασης και του ραδιοφώνου (Dex, Willis, Paterson, & Sheppard, 2000; Saundry, Stuart, & Antcliff, 2007), στις ταινίες (Davenport, 2006), στις εκδόσεις (Stanworth & Stanworth, 1995), στις δημόσιες σχέσεις (Tench, Fawkes, & Palihawadana, 2002), στους προσωρινοί διευθυντές (Goss & Bridson, 1998) και καλλιτέχνες (Menger, 1999). Οι εργαζόμενοι για άλλους λογαριασμούς σε άλλα επαγγέλματα τείνουν να περιγράφονται διαφορετικά, ως εργολάβοι ή σύμβουλοι, και εντοπίζονται στον τομέα της πληροφορικής (S. Barley & Kunda, 2004; SR Barley & Kunda, 2006), και στον τομέα της κοινωνικής εργασίας (Kirkpatrick & Hoque, 2006).

Ενώ η περισσότερη έρευνα έχει επικεντρωθεί σε ψυχαγωγικές ενασχολήσεις και επαγγέλματα στα μέσα μαζικής ενημέρωσης, υποστηρίζουμε ότι οι εργαζόμενοι σε διευθυντικά, επαγγελματικά, επιστημονικά και τεχνικά επαγγέλματα θα πρέπει επίσης να αντιμετωπίζονται ως μέρος του ελεύθερου εργατικού δυναμικού.

Ο Fabio Rosati, Διευθύνων Σύμβουλος της Elance-oDesk λέει ότι οι αμερικανικές επιχειρήσεις θα ξοδέψουν μισό δισεκατομμύριο δολάρια για την πρόσληψη απομακρυσμένων ανεξάρτητων εργαζομένων το 2014. «Η ανάπτυξη είναι σαφής δείκτης της μετάβασης της διαδικτυακής εργασίας», Business Insider Malaysia, 7 Ιουνίου 2014. Το Elance-oDesk χρησιμεύει ως διαμεσολαβητής μεταξύ εταιρειών που χρειάζονται δεξιότητες

και ελεύθερων επαγγελματιών που επιθυμούν να τους τις προσφέρουν, και κατέχει 2,7 εκατομμύρια ελεύθερες θέσεις εργασίας που δημοσιεύονται ετησίως. Σύμφωνα με την εταιρεία, τα συνολικά κέρδη των ελεύθερων επαγγελματιών αυξήθηκαν κατά 50% ετησίως τα τελευταία πέντε χρόνια. Από την άποψη του Rosati, τόσες πολλές επιχειρήσεις προσλαμβάνουν αυτούς τους εργαζόμενους, επειδή χρειάζονται εργασία κατόπιν αιτήματος, συνήθως για ένα σύντομο χρονικό διάστημα. Σε αυτές τις περιπτώσεις, μια αγορά, όπως το Elance-oDesk γίνεται σαν το Amazon Prime για δουλειές. Είναι μια νίκη τόσο για τον εργοδότη, ο οποίος μπορεί να προσλάβει ανάλογα με τις ανάγκες του, όσο και για τον εργαζόμενο, ο οποίος έχει ευελιξία στο πρόγραμμά του και την τοποθεσία του. Επιπλέον, οι ελεύθεροι επαγγελματίες με τις κατάλληλες δεξιότητες μπορούν να κερδίσουν ένα υπέροχο μεροκάματο από την άνεση των σπιτιών τους, σύμφωνα με τις πιο ακριβοπληρωμένες θέσεις εργασίας που είναι αυτές των ελεύθερων επαγγελματιών.

Η ελεύθερη απασχόληση είναι ένας πολύ καλός τρόπος εργασίας. Μπορεί κανείς να εργαστεί από οπουδήποτε, να επιλέξει πόσο θέλει να εργαστεί και τι ώρα θέλει να εργαστεί. Δεν χρειάζεται να ανησυχεί αν πρέπει να ξυπνήσει νωρίς ή να κάνει μεγάλες διακοπές. Είναι το αφεντικό του εαυτού του. Καθώς η ανεξάρτητη κουλτούρα έγινε διάσημη όλα αυτά τα χρόνια, έχουν δημιουργηθεί εκατοντάδες ιστότοποι που παρέχουν ελεύθερες υπηρεσίες. Ψάχνοντας στον ιστό, μπορεί κανείς να βρει μερικούς ανεξάρτητους ιστότοπους που είναι πραγματικά πολύ καλοί και έχουν ήδη καθιερωθεί ως μεγάλες πηγές freelancing και outsourcing.

2.3.2. oDesk

Ξεκίνησε το 2005, αυξήθηκε τόσο γρήγορα και τώρα είναι ένας από τους πιο προτιμώμενους ανεξάρτητους ιστότοπους για ελεύθερους επαγγελματίες εκκίνησης. Το oDesk έχει σχεδιαστεί για ελεύθερους επαγγελματίες που προτιμούν ωριαίες αμοιβές σε αντίθεση με τη πάγια χρηματική αμοιβή. Είναι λίγο πιο ακριβά στις προμήθειές τους, αλλά καλύπτουν τα φορολογικά έξοδα του εργοδότη εάν ο εργοδότης θέλει να προσλάβει κάποιον μέσω της υπηρεσίας μισθοδοσίας του. Πρέπει να συνδεθεί κανείς στον ιστότοπο ενώ εργάζεται για να μπορέσει να καταγράψει στιγμιότυπα οθόνης.

2.3.3. Elance

Είναι ο παλαιότερος και ένας από τους πιο αξιόπιστους ανεξάρτητους ιστότοπους. Ξεκίνησε το 1999 και έκτοτε έχει πραγματοποιήσει μια συνολική ελεύθερη συναλλαγή περίπου μισού δισεκατομμυρίου δολαρίων. Με μεγάλο αριθμό ελεύθερων επαγγελματιών σε διάφορους τομείς όπως σε αυτήν της ανάπτυξης ιστοσελίδων, του σχεδιασμού γραφικών, της σύνταξης περιεχομένου κ.λπ. και περίπου 50.000 θέσεων εργασίας που δημοσιεύονται κάθε μήνα, υπάρχει πάντα ένα τεράστιο περιθώριο για να βρει κανείς νέες δουλειές σε αυτόν τον ιστότοπο. Υπάρχουν πέντε μεγάλες κατηγορίες θέσεων εργασίας στο Elance: Προγραμματιστές, προγραμματιστές κινητών, σχεδιαστές, συγγραφείς και έμποροι. Αλλά μπορείτε επίσης να μπειτε σε μια ευρεία λίστα. Ως ένας από τους κορυφαίους ιστότοπους freelancing, διατηρούν έναν εξαιρετικό ιστότοπο και υπηρεσία. Η προμήθεια είναι υψηλότερη από τις περισσότερες. Προσφέρουν πληρωμή μεσεγγύησης και επίλυση διαφορών, για τις εργασίες που δεν πάνε σωστά.

2.3.4. People Per Hour

Αυτός ο ιστότοπος είναι σχετικά νέος σε σχέση με άλλους παλαιότερους που αναφέρθηκαν παραπάνω. Αν και κυκλοφόρησε το έτος 2008, έχει αποκτήσει σημαντική δημοτικότητα τώρα. Είναι ένα καλό μέρος για προγραμματιστές, γραφίστες, συγγραφείς περιεχομένου και εικονικούς βοηθούς για να προσφέρουν τα αξιολόγητα έργα τους με πληρωμή. Εάν κανείς έχει ένα ταλέντο (όπως το να μπορεί να σχεδιάσει μια επαγγελματική κάρτα σε μία ώρα), το PeoplePerHour είναι ο ιστότοπος freelancing που του ταιριάζει. Λειτουργεί όπως οι άλλοι ιστότοποι freelancing, αλλά επιτρέπει επίσης να δημοσιεύει κανείς "ωριαία" για να προσελκύει επιχειρήσεις. Σε γενικές γραμμές, υπάρχουν λιγότερες θέσεις εργασίας, αλλά είναι οι υπηρεσίες είναι υψηλότερης ποιότητας.

2.3.5. Fiverr

Είναι η μεγαλύτερη αγορά στον κόσμο για μικρές υπηρεσίες. Το πιο ενδιαφέρον μέρος είναι ότι μπορεί κανείς να προσφέρει τις υπηρεσίες του για όσο χαμηλά όσο \$ 5. Μπορεί να αποδειχθεί ότι είναι ένα σύστημα κερδοφορίας για κάποιον, ακόμη και αν δεν γνωρίζει κωδικοποίηση ή σχεδιασμό. Απλά ρίχνοντας μια ματιά στις υπηρεσίες και δουλειές που προσφέρουν οι άλλοι μπορεί κανείς να δει αν μπορεί να συμμετέχει και ο ίδιος. Εάν μπορεί

να κάνει κάτι πολύ συνηθισμένο για το οποίο οι άνθρωποι μπορεί να είναι πρόθυμοι να πληρώσουν 5 \$, τότε μπορεί να κερδίσει σημαντικά χρήματα σε αυτόν τον ιστότοπο.

2.3.6. LinkedIn

Ο αρχικός ιστότοπος κοινωνικής δικτύωσης για επαγγελματίες μπορεί να θεωρηθεί στην απλούστερη μορφή του, ως «κατάστημα» για έναν υπεύθυνο πρόσληψης προσωπικού, όπου μπορεί να δει, μεταξύ άλλων, το βιογραφικό, τα προσόντα, την εκπαίδευση, τις δεξιότητες και τις αναφορές των δυνητικών υπαλλήλων¹.

Ο διανομέας ηλεκτρονικών ειδών “RS Components” προσπάθησε να χρησιμοποιήσει το δίκτυο LinkedIn για να διατηρήσει μια ευρωπαϊκή εργασία το 2010 και ήταν ικανοποιημένος με τη μείωση του κόστους, της ταχύτητας απασχόλησης και των προνομίων που προσφέρει. Με το LinkedIn, η εταιρεία ανέπτυξε μια στρατηγική για την απόκτηση ενός ευρύτερου ανθρώπινου δυναμικού, συνδυάζοντας εργαλεία LinkedIn με τους δικούς της ιστότοπους RS Components. Ο Jon Hull, ο παγκόσμιος επικεφαλής πόρων στην RS Components, απέδειξε ότι, χρησιμοποιώντας τα μέσα κοινωνικής δικτύωσης, το μέσο κόστος ανά απασχόληση είναι 714 £ για ηγετικές θέσεις, σε σύγκριση με 7.500 £ με τις παραδοσιακές μεθόδους, αναγνωρίζοντας ότι, αν και ο αρχικός λόγος για την προσέλκυση στα μέσα κοινωνικής δικτύωσης ήταν η εξοικονόμηση χρημάτων, τα οφέλη που αποκτήθηκαν αποδείχθηκαν πιο κοινότυπα, καθώς η υιοθέτηση των μέσων κοινωνικής δικτύωσης επέτρεψε όχι μόνο τη διαμόρφωση μιας εντελώς νέας στρατηγική πρόσληψης, αλλά επιπλέον, οι ομάδες, τα φόρουμ και τα προφίλ βοήθησαν στην ανάπτυξη της επικοινωνίας, το μάρκετινγκ και την οργάνωση με πιο θετικό τρόπο από ό, τι αρχικά αναμενόταν².

Το LinkedIn είναι ένα ισχυρό επαγγελματικό εργαλείο δικτύωσης. Έχει σχεδιαστεί για να φέρνει κοντά τους ανθρώπους, να διευρύνει τους κύκλους δικτύωσης και να δίνει τη δυνατότητα στους χρήστες να ανταλλάσσουν πληροφορίες. Ωστόσο, αυτό το ισχυρό εργαλείο μπορεί να είναι ένα μυστήριο για μερικούς ανθρώπους. Δεν έχουν πλήρη κατανόηση της χρήσης του και πώς μπορεί να αξιοποιηθεί αποτελεσματικά σε αναζήτηση εργασίας για έρευνα, δικτύωση, προβολή και αυτο-μάρκετινγκ. Επιπλέον, οι άνθρωποι συχνά σκέφτονται το LinkedIn ως ένα μέρος όπου αναζητούν εργασία. Ενώ είναι αλήθεια ότι το LinkedIn έχει μια ισχυρή έλξη για όσους αναζητούν εργασία, είναι έξυπνο για όποιον

¹ ***<http://www.payscale.com/career-news/2013/12/the-state-of-social-recruiting-and-why-you-should-care>;

² ***<http://www.hrmagazine.co.uk/hr/features/1020862/social-media-recruitment-coherent-hr-strategy>;

προσπαθεί να προχωρήσει τη σταδιοδρομία του μακροπρόθεσμα, να προσπαθήσει να πουλήσει ένα προϊόν ή μια ιδέα ή να προσπαθήσει να συνδεθεί με συναδέλφους ειδικούς για να μείνει ενεργός στο LinkedIn. Το LinkedIn είναι ένας ιστότοπος δικτύωσης που ανοίγει έναν κόσμο ευκαιριών για οποιονδήποτε έχει βραχυπρόθεσμους και μακροπρόθεσμους στόχους σταδιοδρομίας.

Ένας χρήστης μπορεί να αξιοποιήσει το LinkedIn για να έρθει σε επικοινωνία με άτομα που γνωρίζει για να ζητήσει συμβουλές, να ζητήσει πληροφορίες, να ρωτήσει για μια ανοιχτή θέση εργασίας που είδε να διαφημίζεται ή να διεξάγει έρευνα σε μια εταιρεία ή στους υπαλλήλους της. Μερικές βασικές ιδέες που πρέπει να κατανοήσει κανείς για το LinkedIn είναι:

- **Εικόνα:** Η ανάλυση έδειξε ότι οι εικόνες με ανοιχτό το στόμα και μια ελαφρύ κλίση του κεφαλιού είναι οι καλύτερες φωτογραφίες.
- **Επικεφαλίδες:** Αξιοποιώντας τη δήλωση επικεφαλίδας για να επικοινωνήσει κανείς την επαγγελματική του εμπειρία, δείχνει ότι εστιάζει στην καριέρα του και ότι δεν είναι απλώς υπάλληλος μιας εταιρείας.
- **Σύντομες παράγραφοι:** Με τον περιορισμό των παραγράφων σε 2-3 προτάσεις, οι πληροφορίες θα αναδειχθούν καλύτερα και η αναγνωσιμότητα του προφίλ θα ενισχυθεί
- **Γλώσσα σε πρώτο πρόσωπο:** Επειδή το LinkedIn είναι ιστότοπος δικτύωσης, πρέπει να είναι συμπληρωματικό σε μια ζωντανή συνομιλία. Όλοι οι χρήστες του LinkedIn θα πρέπει να γράψουν το προφίλ χρησιμοποιώντας τη γλώσσα σε πρώτο πρόσωπο και όχι το στυλ βιογραφικού ή του τρίτου προσώπου για να επικοινωνήσουν με φιλικό, ευγενικό - αλλά επαγγελματικό - τρόπο.
- **Συνδέσεις:** Δεν υπάρχουν κανόνες για το ποιος κάνει μια καλή σύνδεση. Ένας χρήστης μπορεί να συνδεθεί με καθηγητές, συναδέλφους, συνομηλίκους, διευθυντές, επόπτες, συνεργάτες, πελάτες, πωλητές, φίλους και, όπως λέει το τραγούδι του Sesame Street, «τους ανθρώπους που συναντά κάθε μέρα».
- **Ποσότητα Συνδέσεων:** Όσο περισσότερες συνδέσεις έχει κάποιος, τόσο πιο εύκολο είναι για το άτομο αυτό να βρεθεί από έναν στρατηλάτη ή έναν διευθυντή προσλήψεων που πραγματοποιεί μια αναζήτηση με βάση λέξεις -κλειδιά.

Ένα από τα εργαλεία που έχει δημιουργήσει το LinkedIn και είναι ιδιαίτερα χρήσιμο για όσους αναζητούν εργασία και για όσους αναζητούν δικτύωση είναι η λειτουργία Advanced Search. Με αυτήν τη λειτουργία, ένας χρήστης του LinkedIn θα μπορεί να αναζητήσει ένα άτομο με πολύ κοινό όνομα και να το συνδέσει με μια συγκεκριμένη εταιρεία. Μπορεί ακόμη και να έχει τη δυνατότητα να βρεί ένα άτομο με μια συγκεκριμένη θέση μέσα σε μια εταιρεία σε μια συγκεκριμένη τοποθεσία. Η σύνθετη λειτουργία αναζήτησης είναι εξαιρετικά χρήσιμη όταν κάποιος πραγματοποιεί επιθετική αναζήτηση εργασίας και μπορεί να κάνει καλή προσέγγιση.

Στη γραμμή αναζήτησης, ένας χρήστης μπορεί να αναζητήσει το όνομα ενός ατόμου, τον τίτλο εργασίας ή το όνομα της εταιρείας. Η λειτουργία Advanced Search ανοίγει την επιλογή αναζήτησης συνδυασμών πληροφοριών. Αυτός ο τύπος αναζήτησης αναφέρεται επίσης ως "Boolean search". Για να χρησιμοποιήσει κανείς τη λειτουργία Advanced Search, κάνει κλικ στη λέξη "Advanced" δίπλα στην τυπική γραμμή αναζήτησης. Μόλις γίνει η επιλογή, ανοίγει το παράθυρο Αναζήτηση για προχωρημένους. Τώρα, ένας χρήστης μπορεί να αναζητήσει μια σειρά από λέξεις -κλειδιά, ονόματα ατόμων, τίτλους θέσεων εργασίας, ονόματα εταιρειών και τοποθεσίες. Τα δεδομένα μπορούν να φιλτραριστούν κατά βαθμό σχέσης και πολλές άλλες επιλογές που θα βοηθήσουν στη στόχευση συγκεκριμένων αποτελεσμάτων.

Αυτά τα αποτελέσματα μπορούν στη συνέχεια να χρησιμοποιηθούν για να βοηθήσουν έναν επαγγελματία ή ένα άτομο που αναζητά εργασία να εντοπίσει άτομα με τα οποία επιθυμεί να επικοινωνήσει. Στις παρακάτω ενότητες υπάρχουν πληροφορίες και προτάσεις σχετικά με το τι μπορεί να ειπωθεί και τι μπορεί να ζητηθεί από αυτές τις επαφές. Ένας τομέας στον οποίο οι άνθρωποι συχνά δεν εκπροσωπούνται καλά είναι το αίτημα για συγκεκριμένη βοήθεια και κοινοποίηση συγκεκριμένων πληροφοριών. Ενώ πιθανότατα υπάρχουν πολλοί ψυχολογικοί λόγοι για τους οποίους οι άνθρωποι ντρέπονται για αυτό, το συμπέρασμα είναι ότι όσο πιο συγκεκριμένες είναι οι πληροφορίες που προσφέρονται ή ζητούνται, τόσο πιο εύκολο είναι για τους άλλους να εκπληρώσουν αυτό το αίτημα και να κατανοήσουν τους στόχους.

Αυτό ισχύει ιδιαίτερα για άτομα που αναζητούν εργασία. Οι αναζητούντες εργασία χρειάζονται συχνά πληροφορίες, πόρους, εισαγωγές ή άλλη βοήθεια, αλλά δεν τις ζητάνε άμεσα, φοβούμενοι ότι θα συναντήσουν πολύ επιθετική ανταπόκριση ή θα αποκλείσουν τον εαυτό τους από ενδεχόμενες ευκαιρίες. Το αντίστροφο ισχύει στην πραγματικότητα. Επικοινωνώντας συγκεκριμένα, ο αναζητητής εργασίας εμφανίζεται ως εστιασμένος και

σίγουρος και διασφαλίζει ότι οι προθέσεις του γίνονται κατανοητές και δίνει στο άλλο άτομο ακριβή και ενεργή αιτήματα που μπορεί να είναι σε θέση να εκπληρώσει γρήγορα.

Το LinkedIn έχει σχεδιαστεί για επιχειρηματική δικτύωση. Δεν απαιτεί ώρες προσοχής και δεν προορίζεται να είναι μια κοινωνική πρίζα ή ένας ιστότοπος με περιεχόμενο που αλλάζει από λεπτό σε λεπτό. Μεταξύ των πολλών χρήσεών του, εννοείται ότι είναι ένας χώρος στον οποίο ένας χρήστης μπορεί να πάει για να συνδεθεί με συνομηλίκους και άλλους με σκοπό να επιδιώξει φιλοδοξίες καριέρας ή απλώς να μεγιστοποιήσει την επαγγελματική του προσέγγιση. Ανεξάρτητα από το επίπεδο χρήσης που χρησιμοποιεί κάποιος με το LinkedIn, είναι σημαντικό να έχει μια παρουσία έτσι ώστε όταν έρθει η ώρα για πιο επιθετική δικτύωση (δηλαδή όταν κάποιος μπαίνει σε αναζήτηση εργασίας ή έχει σημαντικές συνδέσεις να κάνει), το άτομο να είναι προετοιμασμένο. Χρησιμοποιώντας το LinkedIn σε τακτική βάση για να συνδεθεί ο κάθε χρήστης με συναδέλφους, να επικοινωνήσει με γρήγορες σημειώσεις για να έρθει σε επαφή ή να συνδεθεί με ειδικούς του κλάδου, θα μπορεί μακροπρόθεσμα ή βραχυπρόθεσμα να βασίζεται σε αυτές τις σχέσεις σε περιόδους ανάγκης.

Τα προτεινόμενα θέματα για την έρευνα στο LinkedIn περιλαμβάνουν τη δημιουργία προφίλ, τη δικτύωση μέσω ομάδων, την αξιοποίηση των συνδέσεων άλλων, την ανάρτηση άρθρων, την παρακολούθηση εταιρειών, την επικοινωνία μέσω της λειτουργίας ανταλλαγής μηνυμάτων και τη χρήση του πίνακα εργασίας, μεταξύ άλλων θεμάτων.

2.3.7. Η Πλατφόρμα Xing

Παρόμοια ιστοσελίδα αποτελεί και το Xing(πρώην OpenBC). Είναι ένας ιστότοπος επιχειρηματικής δικτύωσης για επαγγελματίες. Ιδρύθηκε στο Αμβούργο στα τέλη του 2003 από τον επιχειρηματία Lars Hinrichs. Αυτός ο ιστότοπος υποστηρίζει 16 γλώσσες, έχει περισσότερα από 10 εκατομμύρια μέλη, αυξάνοντας τα συνεχώς και αποτελεί μια ιστοσελίδα που χρησιμοποιείται πιο συχνά από εταιρείες στην Γερμανία. Επιτρέπει σε κάθε άτομο να συνδεθεί με κάποιο άλλο άτομο που σχετίζεται με τον επαγγελματικό του τομέα. Το Xing σάς επιτρέπει να είστε συνδεδεμένοι με άτομα στο δίκτυο μέσω των δικών σας επαφών εμφανίζοντας αυτόματα τη συντομότερη διαδρομή προς το άτομο που επιλέγετε να δείτε. Για τον λόγο αυτό χρειάζεται η συμπλήρωση κάποιων στοιχείων του βιογραφικού όπως στοιχεία επικοινωνίας, εκπαίδευση και ενδιαφέροντα. Έπειτα αρχίζει η διαδικασία της πρόσθεσης ατόμων στη λίστα επαφών. Η συμμετοχή στο Xing είναι δωρεάν, αν και όπως και το LinkedIn το Xing προσφέρει Premium συνδρομή, με περισσότερες δυνατότητες, όπως

αναλυτικότερες αναζητήσεις ή τη δυνατότητα αποστολής μηνυμάτων σε άτομα εκτός της λίστας επαφής. Το Xing, βασιζόμενο σε αλγόριθμο αντίστοιχο με αυτόν του LinkedIn βοηθά τους ανθρώπους να λαμβάνουν ενημερώσεις και μηνύματα σε καθημερινή βάση από τα ακόλουθα μέλη σχετικά με τις νέες θέσεις εργασίας ή πρότζεκτ στα οποία θα μπορούσαν να συμμετέχουν.

Το Xing είναι ένας ευρωπαϊκός ιστότοπος κοινωνικής δικτύωσης για επαγγελματίες. Όπως και με το LinkedIn, ο στόχος του XING είναι να βοηθήσει τα μέλη του να βρουν δουλειά ή να έρθουν σε επαφή με άτομα με τα οποία έχουν συνεργαστεί. Στο Xing μπορεί κανείς να βρεί επαγγελματίες κάθε κατηγορίας.

Το Xing δεν είναι ιστότοπος γνωριμιών, αλλά είναι ένας ιστότοπος δικτύωσης επιχειρήσεων για επαγγελματίες. Το Xing έχει επιχειρηματική αίσθηση. Επιτρέπει σε οποιοδήποτε αυθαίρετο άτομο να συνδεθεί με ένα άλλο αυθαίρετο άτομο που σχετίζεται με το επάγγελμα και τον τομέα του. Το Xing επιτρέπει να συνδεθεί ο κάθε χρήστης με άτομα στο δίκτυο μέσω των δικών του επαφών, εμφανίζοντας αυτόματα τη συντομότερη διαδρομή προς το άτομο που επιλέγει να δει.

Μόλις συμπληρώσει τα στοιχεία του σε ένα βιογραφικό σημείωμα με στοιχεία επικοινωνίας, το επίπεδο εκπαίδευσής του και τα ενδιαφέροντά του, μπορεί στη συνέχεια να αρχίσει να προσθέτει άτομα στη λίστα επαφών του. Η συμμετοχή στο Xing είναι δωρεάν, αν και το Xing προσφέρει μια premium συνδρομή, με περισσότερες δυνατότητες, όπως προηγμένες "αναζητήσεις ισχύος", δυνατότητα αποστολής μηνυμάτων σε άτομα και έλεγχο για το ποιος επισκέφθηκε πρόσφατα τη σελίδα σας.

Το Xing είναι επίσης μια πλατφόρμα όπου οι άνθρωποι μπορούν να παραμείνουν συνδεδεμένοι είτε ζουν μαζί είτε εργάζονται μαζί σε μια εταιρεία ή εργάζονται σε ένα έργο. Συχνά δεν συμβαίνει πλέον οι άνθρωποι να εργάζονται στην ίδια εταιρεία και να ζουν στο ίδιο μέρος για όλη τους τη ζωή. Το Xing παρέχει κινητικότητα και ευελιξία στους ανθρώπους να παραμένουν συνδεδεμένοι για τη μελλοντική τους αλληλεπίδραση και επιτρέπει στα μέλη να έρχονται σε επαφή με τους σχετικούς συνεργάτες με την επιχείρησή τους στο νέο τους περιβάλλον.

Το Xing βοηθά τους ανθρώπους να λαμβάνουν ενημερώσεις και μηνύματα σε καθημερινή βάση από τα ακόλουθα μέλη σχετικά με τις νέες θέσεις εργασίας ή λεπτομέρειες για ενδιαφέροντα έργα. Με την επεξεργασία των αποτελεσμάτων της έρευνας, πολλά μέλη

δήλωσαν ότι έχουν δημιουργήσει πραγματικά έσοδα από νέες επιχειρήσεις ως άμεσο αποτέλεσμα της χρήσης του XING. Πολλά μέλη ανέφεραν ότι έχουν κάνει περισσότερες συμφωνίες χρησιμοποιώντας το Xing. Τα μέλη του Xing μπορούν να αποκτήσουν πλεονεκτήματα κάνοντας νέες επαφές και επίσης να επωφεληθούν από συμβουλές σε χιλιάδες ομάδες που έχουν δημιουργηθεί στον ιστότοπο. Με τη σύνδεση στην ενότητα "Αγορά" βλέπει κανείς ακόμη και αναρτήσεις εργασίας που ταιριάζουν με το προφίλ του.

Το Xing είναι ένα ισχυρό εργαλείο. Σήμερα, πολλοί χώροι φιλοξενίας παρέχουν διαφορετικά εργαλεία για την εξαγωγή δεδομένων διαφορετικών ανθρώπων για τη χρήση του κοινωνικού μάρκετινγκ και για το ηλεκτρονικό ταχυδρομείο

Στα χαρακτηριστικά του Xing συγκαταλέγονται:

- **Εργασία:** Βρίσκει κανείς τη σωστή δουλειά για τον τρόπο ζωής του και στέλνονται νέες προσφορές σε αυτόν αυτόματα.
- **Xing News:** Κορυφαία καθημερινά άρθρα από τον κλάδο του κάθε χρήστη. Ο κάθε χρήστης μαθαίνει περισσότερα για τον κλάδο του και προσθέτει τη δική του οπτική.
- **Εκδηλώσεις:** Ανακάλυψη σπουδαίων σεμιναρίων, συνεδρίων, εμπορικών εκθέσεων και πολλών άλλων στη μεγαλύτερη πλατφόρμα επιχειρηματικών εκδηλώσεων.
- **Εταιρείες:** Κάθε χρήστης έχει τη δυνατότητα να βρει τους καλύτερους και πιο δημοφιλείς εργοδότες με κορυφαία αξιολόγηση στην πόλη ή την επιχείρησή του.

Το Social Marketing είναι ο καλύτερος τρόπος για να προωθήσει κανείς και να κοινωνικοποιήσει την επιχείρησή του στο σημερινό μάρκετινγκ. Το δίκτυο Xing είναι το καλύτερο που μπορεί να προσελκύσει στοχευμένο κοινό για την επιχείρηση καθενός χρήστη. Είναι πολύ δύσκολο να αναζητήσει κανείς και να εξαγάγει χρήσιμες πληροφορίες για το Xing όπως (αριθμός τηλεφώνου, σύνδεσμος προφίλ, όνομα, email, φαξ, τίτλος εργασίας, ακαδημαϊκός τίτλος, αριθμός συνδέσεων, επάγγελμα, προφίλ εταιρείας, διεύθυνση, βιομηχανία, , ενδιαφέροντα , θέσεις προσόντων, αναγνωριστικό προφίλ, κλπ). με αναζήτηση από το google και την επίσκεψη σε πολλούς άλλους ιστότοπους για τη λήψη αυτών των πληροφοριών. Είναι πολύ οδυνηρό να συλλέγονται πληροφορίες από έναν - έναν άνθρωπο με μια επίσκεψη στα προφίλ τους και το κόστος του να είναι πολύ υψηλό, επομένως μπορεί να ειπωθεί ότι πρόκειται για σπατάλη χρόνου, χρημάτων και πόρων. Σήμερα, η γενιά του 21ου αιώνα κινείται προς την αυτοματοποίηση. Σήμερα, χρειάζεται κάποιο λογισμικό για την εξαγωγή πληροφοριών επικοινωνίας από το κοινωνικό δίκτυο για το κοινωνικό μάρκετινγκ.

Έχουν αναζητηθεί λογισμικά για να εξαχθούν οι πληροφορίες των χρηστών του Xing, αλλά δεν παρέχουν τα σχετικά αποτελέσματα έναντι της αναζήτησης.

2.3.8. Indeed

Επιπλέον είναι σημαντικό να αναφερθούμε και στην ιστοσελίδα indeed.com όπου βασικός της στόχος είναι να δημοσιεύει θέσεις εργασίας. Τα άτομα μπορούν και εδώ να δημιουργήσουν δωρεάν ένα προφίλ και να ενημερώνονται για τις θέσεις εργασίας που τους ταιριάζουν καλύτερα, με την διαφορά ότι δεν μπορούν να συνδεθούν με άλλους επαγγελματίες. Το indeed έχει δικό του (ιδιόκτητο) αλγόριθμο που καταχωρεί και ταξινομεί τις αγγελίες κατά συνάφεια ή ημερομηνία. Έχει την κατηγορία των χορηγουμένων θέσεων εργασίας (sponsored jobs) όπου έχουν κορυφαία θέση στα αποτελέσματα αναζήτησης και προσελκύουν μεγαλύτερο κοινό, ωστόσο κοστίζουν περισσότερο αν μια επιχείρηση επιλέξει αυτή την κατηγορία για την αγγελία της. Είναι ορατές μόνο στους υποψηφίους που εισάγουν όρους αναζήτησης που σχετίζονται με τον τίτλο της εργασίας, την περιγραφή, την τοποθεσία και άλλες πληροφορίες εντός της εργασίας.

2.3.9. Facebook

Αν και το LinkedIn αναφέρεται συχνά ως το πιο προφανές δίκτυο για την πρόσληψη, οι περισσότεροι απόφοιτοι και νέοι προτιμούν το Facebook και το Potentialpark πράγμα που οδηγεί τους υπεύθυνους πρόσληψης προσωπικού να οδηγούνται σε αυτό το περιβάλλον, ενώ το Jobvite infographic κατατάσσει το facebook στη δεύτερη θέση της λίστας (65%). Πάνω από το ένα τρίτο των 100 κορυφαίων Ευρωπαίων εργοδοτών έχουν σελίδες καριέρας στο Facebook, μερικά παραδείγματα σελίδων «κοινωνικής» καριέρας στο Facebook, συμπεριλαμβανομένων των Unilever και Oracle, όπου οι ζωντανές συζητήσεις στον «τοίχο» βοηθούν τους αιτούντες από όλο τον κόσμο να ανακαλύψουν τις ευκαιρίες που ανοίγονται μπροστά τους, πώς να τις αξιοποιήσουν και τι να περιμένουν. Η Unilever χρησιμοποιεί τη σελίδα της στο Facebook για να δώσει στους αποφοίτους την ευκαιρία να κερδίσουν μια τσάντα με δώρα και να μπουν στην κλήρωση για βραβεία, αν κάνουν ένα "like" στη σελίδα και συμπληρώσουν μια μίνι έρευνα για τις προσδοκίες που έχουν όσον αφορά την καριέρα τους.

Για την Ένωση Δασικών Προϊόντων του Καναδά (FPAC), το Facebook ήταν ένα ουσιαστικό εργαλείο για την προσέλκυση φοιτητών να συμμετάσχουν στον διαγωνισμό κοινωνικών

μέσων με τίτλο Green Dream Internship, προκειμένου να κερδίσουν την ευκαιρία να «βρουν μια θέση σε ένα μία από τις εννέα κορυφαίες εταιρείες δασικών προϊόντων, για τέσσερις μήνες, το καλοκαίρι του 2014»³. Ο Ντέιβιντ Λίντσεϊ, Πρόεδρος και Διευθύνων Σύμβουλος της FPAC, εξήγησε ότι, "Δεν υπάρχει καλύτερος τρόπος για να προσεγγίσουμε τους νέους, παρά μόνο εάν τους προσεγγίσουμε μέσω των μέσων κοινωνικής δικτύωσης. Λάβαμε σημαντικό αριθμό αιτήσεων υψηλού επιπέδου από φοιτητές σε όλη τη χώρα."⁴. Το Facebook επιτρέπει στους εργοδότες να συνδέονται μαζικά με ομάδες ανθρώπων.

2.3.10. Twitter

Όταν το Twitter πρωτοεμφανίστηκε, ο κόσμος των μέσων κοινωνικής δικτύωσης το απέφυγε επειδή φαινόταν άχρηστο, ωστόσο, αποδείχθηκε γρήγορα ότι το Twitter ήταν ένας πολύ καλός τρόπος για άμεση ενημέρωση σε πραγματικό χρόνο. Στον επαγγελματικό τομέα, ο ιστότοπος micro-blogging έχει αποκτήσει δημοτικότητα, επειδή επέτρεψε στους χρήστες να παρακολουθούν ό,τι και όποιον θέλουν-διασημότητες, άτομα με επιρροή, κλπ. Επί του παρόντος, το Twitter λειτουργεί ως μια εξαιρετική υπηρεσία για επαγγελματίες που θέλουν να συνδεθούν και να συνεργαστούν με κορυφαίους στον τομέα τους, να δημιουργήσουν δίκτυα με άλλα άτομα αντίστοιχου επαγγέλματος και να παραμείνουν ενημερωμένοι όσον αφορά τα νέα που σχετίζονται με τον κλάδο τους.

2.4.Πλεονεκτήματα Μέσων Κοινωνικής Δικτύωσης ως Εργαλεία Πρόσληψης

Η χρήση των κοινωνικών μέσων μαζικής ενημέρωσης ως μέρος της στρατηγικής πρόσληψης των εταιρειών προσφέρει πραγματικά πλεονεκτήματα ενώ εστιάζει κυρίως στη μείωση του κόστους, στον εντοπισμό μεγαλύτερου συνόλου των εν δυνάμει αιτούντων και στην ικανότητα να κατευθύνουν πιο αποτελεσματικά τις προσλήψεις⁵.

Το πιο σημαντικό πλεονέκτημα της χρήσης των μέσων κοινωνικής δικτύωσης ως μέρος της στρατηγικής προσλήψεων, είναι η μείωση του κόστους ενώ ταυτόχρονα οι εταιρείες που χρησιμοποιούν ιστότοπους όπως το LinkedIn για αναζήτηση εξειδικευμένου προσωπικού μπορούν να εξοικονομήσουν μεγάλα χρηματικά ποσά από τις αμοιβές των εταιρειών εύρεσης προσωπικού.

³ ***<http://www.digitaljournal.com/pr/1620697>;

⁴ ***<http://thegreenestworkforce.ca/index.php/en/contest2-home/>;

⁵ CIPD, *Employers are from Mars, young people are from Venus: Addressing the young people/Job mismatch*, Research Report, April 2012, p. 25

Επίσης, η χρήση του Facebook εμφανίζεται ως ένας σχετικά γρήγορος και εύκολος τρόπος για την προσέλκυση μεγάλου αριθμού εν δυνάμει αιτούντων, ειδικά μεταξύ των νέων.

Από την άλλη, οι νέοι είναι πολύ εξοικειωμένοι με τον ιστότοπο κοινής χρήσης βίντεο YouTube και ορισμένοι εργοδότες χρησιμοποιούν αυτήν τη μέθοδο για να δώσουν στους υποψήφιους αιτούντες μια πραγματική ιδέα για την κουλτούρα ενός χώρου εργασίας κάνοντας χρήση βίντεο εργαζομένων και διευθυντών. Για τους νέους, αυτός ο τύπος προσέγγισης είναι πιθανό να είναι ιδιαίτερα αποτελεσματικός, καθώς είναι πιο πιθανό να επηρεαστούν από παράγοντες όπως η οργανωτική κουλτούρα και η εύρεση ενός οργανισμού που θα τους επιτρέψει να συνδυάσουν την εργασία και τον τρόπο ζωής που επιθυμούν.

2.5.Κίνδυνοι και Κόστος από τη Χρήση των Μέσων Κοινωνικής Δικτύωσης για Σκοπούς Πρόσληψης

Εκτός από τα πλεονεκτήματα, η χρήση των μέσων κοινωνικής δικτύωσης μπορεί επίσης να προκαλέσει μια σειρά προβλημάτων. Στην εργασία τους οι (Davison et al., 2011) σημείωσαν ότι "ένα άτομο που δημιουργεί μια ιστοσελίδα μπορεί να προσπαθήσει να εντυπωσιάσει τους φίλους, την οικογένεια, τους συναδέλφους και/ή τους εργοδότες του και ο βαθμός στρέβλωσης ή «μεταμόρφωσης» μπορεί να διαφέρει ανάλογα με τον «ενδιαφερόμενο θεατή» (Davison et al., 2012). Οι άνθρωποι για παράδειγμα έχουν την τάση να «παραποιούν» ορισμένες πληροφορίες υπέρ τους, εάν πιστεύουν ότι οι γονείς ή οι εργοδότες τους θα δουν την ιστοσελίδα.

Ένα άλλο πρόβλημα που μπορεί να προκύψει σχετίζεται με το κόστος συντήρησης τοποθεσιών/ροών. Δεδομένου ότι οι χρήστες ενθαρρύνονται να δημοσιεύουν ερωτήσεις μέσω αυτών των ιστότοπων, είναι σαφές ότι οι εταιρείες πρέπει να διασφαλίσουν ότι απαντούν σχετικά γρήγορα και ότι ο ιστότοπος ενημερώνεται τακτικά, διαφορετικά υπάρχει κίνδυνος οι εν δυνάμει αιτούντες να χάσουν το ενδιαφέρον τους για τον εν λόγω ιστότοπο. Κατά συνέπεια, οι εργοδότες πρέπει να διαθέσουν προσωπικό που θα έχει ως αντικείμενο εργασίας τη διαχείριση των ιστοτόπων κοινωνικής δικτύωσης και την παρακολούθηση της επισκεψιμότητάς τους.

Το Διαδίκτυο παρέχει μια αποτελεσματική μέθοδο για τη λήψη πρόσθετων πληροφοριών σχετικά με τους υποψηφίους και η προβολή των δραστηριοτήτων των ατόμων στα μέσα κοινωνικής δικτύωσης μπορεί να παρέχει στον εργοδότη πολύτιμες πληροφορίες για το προφίλ συμπεριφοράς τους (Ghoshray, 2013). Ωστόσο αυτή ακριβώς η πιθανή παρέμβαση

μεταξύ της εργασίας και της ιδιωτικής ζωής είναι που μπορεί να καταστεί η αιτία πιθανών προβλημάτων. Αποτελέσματα έρευνας που πραγματοποιήθηκε από τον ιστότοπο CareerBuilder.com στις ΗΠΑ έδειξαν ότι το 37% έχει εξετάσει τα ηλεκτρονικά προφίλ των υποψηφίων και μεταξύ αυτών που δεν το έχουν κάνει, το 15% δήλωσε ότι ο οργανισμός δεν επέτρεπε κάτι τέτοιο ενώ το 11% απάντησε ότι παρόλο που μέχρι στιγμής δεν έχει κάνει κάτι τέτοιο, σχεδιάζει να το ξεκινήσει. Από τους εργοδότες που επισκέπτονται διαδικτυακά προφίλ, το 34% δήλωσε ότι το υλικό που βρήκε στο προφίλ (φωτογραφίες (49%) ή ακατάλληλες πληροφορίες ή πληροφορίες που έδειχναν ότι τα προσόντα ήταν πλαστά (22%)) ήταν η αιτία που το συγκεκριμένο άτομο δεν προσελήφθη.⁶

Παρόλο που προς το παρόν δεν υπάρχει συγκεκριμένη νομοθεσία για αυτό το ζήτημα, ενδέχεται να υπάρχουν δυσκολίες όσον αφορά την γνώση τόσο πολλών πληροφοριών για κάποιο άτομο.

Ένα ακόμη μειονέκτημα την μεθόδου πρόσληψης μέσω μέσων κοινωνικής δικτύωσης, αποτελεί το γεγονός ότι τα μέσα αυτά προσελκύουν ορισμένες ομάδες ανθρώπων, όπως εκείνους κάτω των 40 ετών και εκείνους που είναι «εξοικειωμένοι» με αυτήν την τεχνολογία κάτι που έχει σαν αποτέλεσμα τον αποκλεισμό πολλών υποψηφίων που δεν εμπίπτουν σε αυτές τις κατηγορίες. Ως εκ τούτου, συνιστάται τα εργαλεία κοινωνικής δικτύωσης να μην χρησιμοποιούνται μεμονωμένα, αλλά ως μέρος μιας στρατηγικής πρόσληψης.

⁶ ***<http://www.careerbuilder.com/share/aboutus/pressreleasesdetail.aspx?id=pr691&sd=4%2F18%2F2012&ed=4%2F18%2F2099>;

3. Βιβλιογραφική Ανασκόπηση Μελετών Εύρεσης Εργασίας με Χρήση Μηχανικής Μάθησης

3.1. Μηχανική Μάθηση Αναγνώρισης Βιογραφικού για Εύρεση Εργασίας

Το σύστημα αντιστοίχισης εργασίας είναι ένα είδος συστήματος σύστασης. Εισήχθη για πρώτη φορά από τους (Resnick & Varian, 1997) που επεσήμαναν ότι σε ένα τυπικό σύστημα σύστασης οι άνθρωποι παρέχουν συστάσεις ως εισόδους, τις οποίες το σύστημα συγκεντρώνει και κατευθύνει στους κατάλληλους παραλήπτες. Τα συστήματα σύστασης άρχισαν να γίνονται ιδιαίτερα αποδεκτά σε διάφορους κλάδους και ακαδημαϊκούς τομείς και κερδίζουν δυναμική όλα αυτά τα χρόνια. Γενικά, τα συστήματα σύστασης εφαρμόζονται σε διάφορους τομείς (όπως βιβλία, ψηφιακά προϊόντα, ταινίες, μουσική, τηλεοπτικά προγράμματα και ιστότοπους) και βοηθούν τους χρήστες να βρουν περιεχόμενο, προϊόντα, ή υπηρεσίες συγκεντρώνοντας και αναλύοντας προτάσεις και συμπεριφορές από άλλους χρήστες (Lu et al., 2015). Η αναλυτική ερευνητική μελέτη των (Lu et al., 2015) παρείχε στους ερευνητές τις τελευταίες γνώσεις σχετικά με το σύστημα σύστασης, συμπεριλαμβανομένων εφαρμογών πραγματικού κόσμου, μεθόδων σύστασης, τομέων εφαρμογών πραγματικού κόσμου και πλατφορμών εφαρμογών.

Για την αντιστοίχιση θέσεων εργασίας, έχουν διεξαχθεί πολλές έρευνες προς συζήτηση διαφορετικών συστημάτων σύστασης που σχετίζονται με το πρόβλημα της πρόσληψης (Al-Otaibi, & Ykhlef, 2012). Μεταξύ αυτών, οι (Malinowski et al., 2006) συζήτησαν συστήματα διμερούς αντιστοίχισης συστάσεων για να αντιστοιχήσουν τις θέσεις εργασίας με τους ενδιαφερόμενους χρησιμοποιώντας τον αλγόριθμο Expectation Maximization (EM), ενώ οι (Golec & Kahya, 2007) σκιαγράφησαν ένα ασαφές μοντέλο για αξιολόγηση και επιλογή υπαλλήλων βάσει ικανοτήτων με ασαφείς κανόνες. Στην εργασία τους οι (Paparrizos et al., 2011) χρησιμοποίησαν τον πίνακα απόφασης/Naive Bayes (DTNB) ως υβριδικό ταξινομητή. Αν και αυτά τα συστήματα χρησιμοποίησαν πολλά χαρακτηριστικά όχι με αυτόματο τρόπο καθώς και διάφορες τεχνικές ανάκτησης πληροφοριών, σε σύγκριση με την εργασία των (Lin et al., 2016) όπου χρησιμοποιήθηκαν μέθοδοι βαθιάς μάθησης για να επιταχυνθεί η διαδικασία εύρεσης των πιο κατάλληλων θέσεων εργασίας, ωστόσο απέτυχαν να συμβαδίσουν με τις ραγδαίες αλλαγές στην ικανότητα των υπολογιστών και τη νοημοσύνη των μηχανών. Όπως και η εργασία των (Zhang et al., 2015), έτσι και οι (Lin et al., 2016) προσπάθησαν επίσης να βελτιστοποιήσουν τη αντιστοίχιση θέσης

εργασίας-γνώσης εργαζομένων, λαμβάνοντας υπόψη διάφορα χαρακτηριστικά των γνώσεων των εργαζομένων. Σε σύγκριση με τη μελέτη των (Guo et al., 2016), η λύση την οποία παρουσίασαν οι (Lin et al., 2016) βασίστηκε εντελώς σε δεδομένα, χωρίς την ύπαρξη εξωτερικών σημασιολογικών τρόπων (NLTK και DBpedia). Εκτός αυτού, τα δεδομένα βιογραφικού συλλέχθηκαν ευρέως από διάφορους τομείς που καθιστούν τη συγκεκριμένη λύση των (Lin et al., 2016) πιο καθολική και στιβαρή.

3.2.Μηχανική Μάθηση για τον Προσδιορισμό Σχετικών Δεξιοτήτων από Περιγραφή Εργασίας

Οι υπεύθυνοι πρόσληψης στον τομέα της τεχνολογίας της πληροφορίας αντιμετώπισαν το πρόβλημα της εύρεσης κατάλληλων υποψηφίων από τις δεξιότητές τους. Στο βιογραφικό, ο υποψήφιος μπορεί να περιγράψει μια δεξιότητα με διαφορετικούς τρόπους ή οι δεξιότητες θα μπορούσαν να αντικατασταθούν από άλλα. Οι υπεύθυνοι πρόσληψης ενδέχεται να μην έχουν τη γνώση του τομέα ώστε να μπορούν να εκτιμήσουν εάν οι δεξιότητες κάποιου είναι κατάλληλες ή όχι, επομένως μπορούν να βρουν μόνο εκείνους με αντίστοιχες δεξιότητες.

Για να αντιμετωπιστεί το πρόβλημα, πρέπει να γίνει μία προσπάθεια περί ευρέσεως της σχετικότητας των δεξιοτήτων. Υπάρχουν μερικές προσεγγίσεις, όπως δημιουργία ενός λεξικού με μη αυτόματο τρόπο, προσέγγιση οντολογίας, μέθοδοι επεξεργασίας φυσικής γλώσσας, κ.λπ. Σε αυτήν τη μελέτη των (Le Van-Duyet et al., 2019), εφαρμόστηκε μια μέθοδος ενσωμάτωσης λέξεων Word2Vec, χρησιμοποιώντας δεξιότητες από περιγραφές σε απευθείας σύνδεση θέση εργασίας. Οι δεξιότητες αντιμετωπίζονται ως όροι, οι θέσεις εργασίας ως έγγραφα και σκοπός είναι να βρεθεί η συνάφεια αυτών των δεξιοτήτων.

Για να εντοπίσει τη σχετικότητα των δεξιοτήτων, ο (Hughes, 2015) από το Dice πρότεινε μια προσέγγιση χρησιμοποιώντας Λανθάνουσα Σημασιολογική Ανάλυση (Latent Semantic Analysis) με την υπόθεση ότι οι δεξιότητες σχετίζονται με δεξιότητες που εμφανίζονται στο ίδιο πλαίσιο, όπου στην συγκεκριμένη περίπτωση τα πλαίσια αντιστοιχούν σε θέσεις εργασίας. Αυτή η προσέγγιση θα δημιουργήσει μια μήτρα όρου-εγγράφου και θα χρησιμοποιήσει την Αποσύνθεση Μοναδικής Αξίας (Singular Value Decomposition) για να μειώσει τη διάσταση. Τα μειονεκτήματα αυτής της προσέγγισης είναι ότι όταν εισέρχονται νέα δεδομένα, δεν καθίσταται δυνατή η ενημέρωση του εγγράφου του παλαιού πίνακα, το

οποίο οδηγεί σε δυσκολίες όσον αφορά τη διατήρηση του μοντέλου, καθώς η αλλαγή τάσης σε αυτόν τον τομέα είναι υψηλή.

Οι Επιστήμονες δεδομένων της Google αντιμετωπίζουν επίσης τα ίδια προβλήματα στο Cloud Jobs API (Posse, 2016). Η λύση τους είναι να χτίσουν μια οντολογία δεξιοτήτων που να ορίζει περίπου 50.000 δεξιότητες σε 30 τομείς εργασίας. Αυτή η προσέγγιση μπορεί να αντιπροσωπεύει περίπλοκες σχέσεις μεταξύ δεξιοτήτων και θέσεων εργασίας, αλλά η οικοδόμηση μιας τέτοιας οντολογίας κοστίζει πολύ χρόνο και προσπάθεια.

Για να ξεπεραστεί το πρόβλημα του σχετικού όρου, (Huang et al., 2003) παρουσιάζεται μια νέα, αποτελεσματική προσέγγιση βασισμένη σε λογαριθμική βάση σχετική με την εξαγωγή όρων και την πρόταση όρων. Ο στόχος του (Bhat et al., 2004) είναι να αναπτύξει ένα αυτοματοποιημένο σύστημα που ανακαλύπτει πρόσθετα ονόματα για μια οντότητα δεδομένου μόνο ενός από τα ονόματά της, χρησιμοποιώντας την Λανθάνουσα Σημασιολογική Ανάλυση (Latent semantic analysis-LSA) (Berry & Fierro, 1996). Στη μελέτη των (Le Van-Duyet et al., 2019), η πόλη της Ινδίας που αναφέρεται ως Bombay σε ορισμένα έγγραφα μπορεί να αναφέρεται ως Mumbai σε άλλα επειδή το όνομά της άλλαξε επίσημα από το πρώτο στο δεύτερο το 1995. Στην εργασία τους οι (Lau & Sure, 2002) παρουσίασαν την εισαγωγή ενός συστήματος διαχείρισης δεξιοτήτων που βασίζεται στην οντολογία στο Swiss Life και τα διδάγματα που αντλήθηκαν από τη χρήση της μεθοδολογίας, παρουσιάζουν μια μεθοδολογία για την ανάπτυξη οντολογιών που βασίζονται στην εφαρμογή, η οποία βασίζεται σε μια μελέτη περίπτωσης.

3.2.1. Αρχιτεκτονική WORD2VEC

Το Word2Vec είναι μια ομάδα μοντέλων που πρότειναν οι (Mikolov et al., 2013). Αποτελείται από 2 μοντέλα: συνεχές bag-of-words και συνεχές Skip-gram, και τα δύο είναι ρηχά νευρωνικά δίκτυα που προσπαθούν να μάθουν κατανεμημένες αναπαραστάσεις λέξεων με στόχο να μεγιστοποιηθεί η ακρίβεια με παράλληλη ελαχιστοποίηση της υπολογιστικής πολυπλοκότητας. Στη συνεχή αρχιτεκτονική bag-of-words, το μοντέλο προβλέπει την τρέχουσα λέξη από ένα παράθυρο σχετικών με την τρέχουσα λέξεων. Από την άλλη πλευρά, το μοντέλο Skip-gram προσπαθεί να προβλέψει τις σχετικές λέξεις με βάση την τρέχουσα. Στην εργασία των (Le Van-Duyet et al., 2019) αναλύεται το μοντέλο Skipgram, καθώς είναι

γνωστό ότι είναι καλύτερο όσον αφορά τις σπάνιες λέξεις και δίνει ελαφρώς καλύτερο αποτέλεσμα.

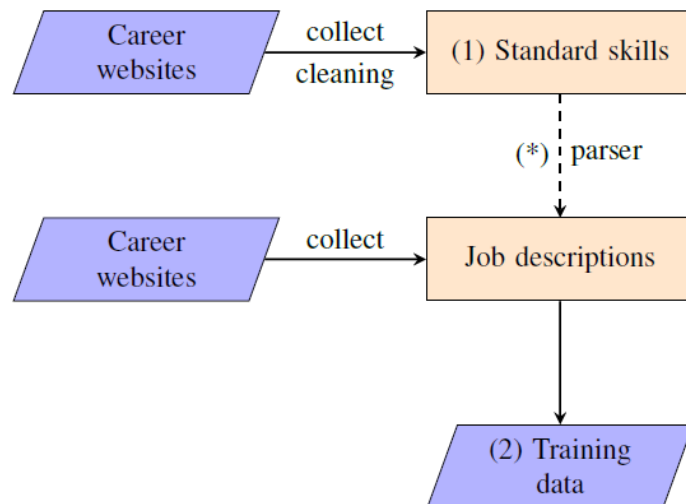
Το μοντέλο αποτελείται από τρία επίπεδα: στρώμα εισόδου, ένα κρυφό στρώμα και στρώμα εξόδου. Το επίπεδο εισόδου παίρνει μια λέξη κωδικοποιημένη χρησιμοποιώντας κωδικοποίηση 1-of-V (επίσης γνωστή ως κωδικοποίηση one-hot), όπου το V είναι μέγεθος του λεξικού. Στη συνέχεια, η λέξη τροφοδοτείται μέσω του γραμμικού κρυφού επιπέδου στο επίπεδο εξόδου, το οποίο είναι ένας ταξινομητής softmax. Το επίπεδο εξόδου θα προσπαθήσει να προβλέψει λέξεις εντός μεγέθους παραθύρου πριν και μετά την τρέχουσα λέξη. Χρησιμοποιώντας στοχαστική κλίση καθόδου και οπίσθια διάδοση, πραγματοποιείται εκπαίδευση του μοντέλου έως ότου επιτευχθεί η σύγκλιση.

Αυτό το μοντέλο λαμβάνει ως παραμέτρους τις διαστάσεις του φορέα και το μέγεθος του παραθύρου. Οι (Le Van-Duyet et al., 2019) διαπίστωσαν ότι η αύξηση του μεγέθους παραθύρου βελτιώνει την ποιότητα του διανύσματος λέξεων, αλλά αυξάνει την υπολογιστική πολυπλοκότητα.

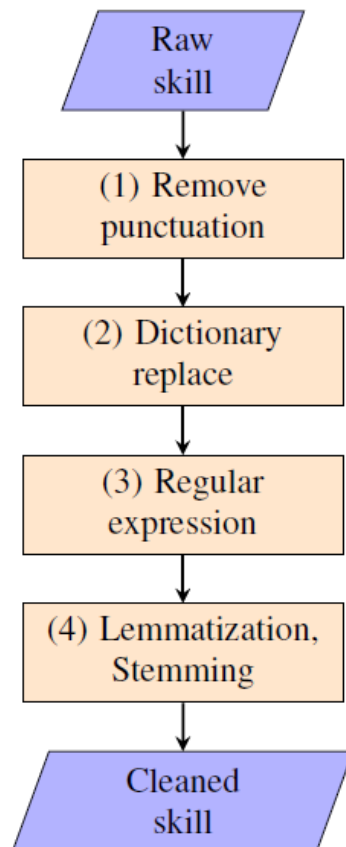
3.2.2. Συλλογή και επεξεργασία δεδομένων

Η επιλογή ενός καθολικού συνόλου δεδομένων για το μοντέλο είναι εξαιρετικά σημαντική, τα δεδομένα πρέπει να είναι αρκετά μεγάλα και πρέπει να έχουν ισορροπημένες κατανομές σε λέξεις (δηλαδή δεξιότητες). Υπάρχουν δύο σύνολα δεδομένων για τα οποία πρέπει να υπάρχει ενδιαφέρον, ένα (1) είναι το τυπικό λεξικό δεξιοτήτων για τον αναλυτή και ένα άλλο (2) είναι το λεξικό δεξιοτήτων για το μοντέλο εκπαίδευσης, σύμφωνα με το σχήμα 1.

Αρχικά συλλέχθηκε και ετοιμάστηκε ένα μεγάλο λεξικό δεξιοτήτων. Με αυτό το λεξικό, καθίσταται δυνατή η εξαγωγή ενός συνόλου δεξιοτήτων από περιγραφές εργασίας. Οι δεξιότητες πρέπει να καθαριστούν σε μοναδικές δεξιότητες, επειδή υπάρχουν πολλοί τρόποι για να παρουσιαστεί μια δεξιότητα στην περιγραφή εργασίας (π.χ. OOP ή αντικειμενοστραφής προγραμματισμός). Το σχήμα 2 απεικονίζει εν συντομία την ιδέα της διαδικασίας καθαρισμού.



Σχήμα 1. Αγωγός σύνδεσης και επεξεργασίας δεδομένων



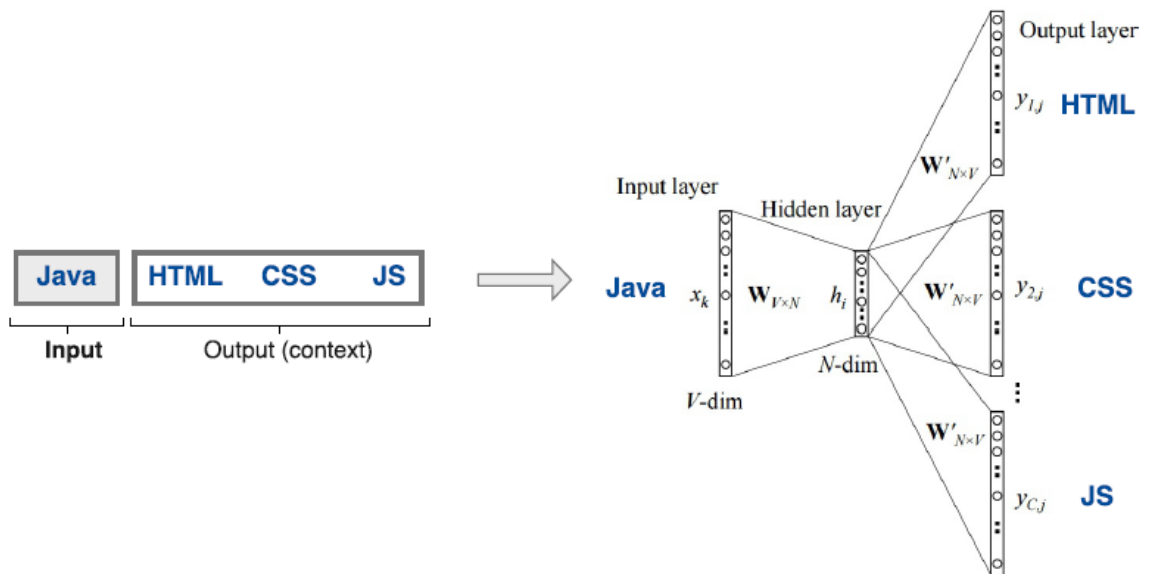
Σχήμα 2. Δεξιότητες καθαρισμού αγωγών

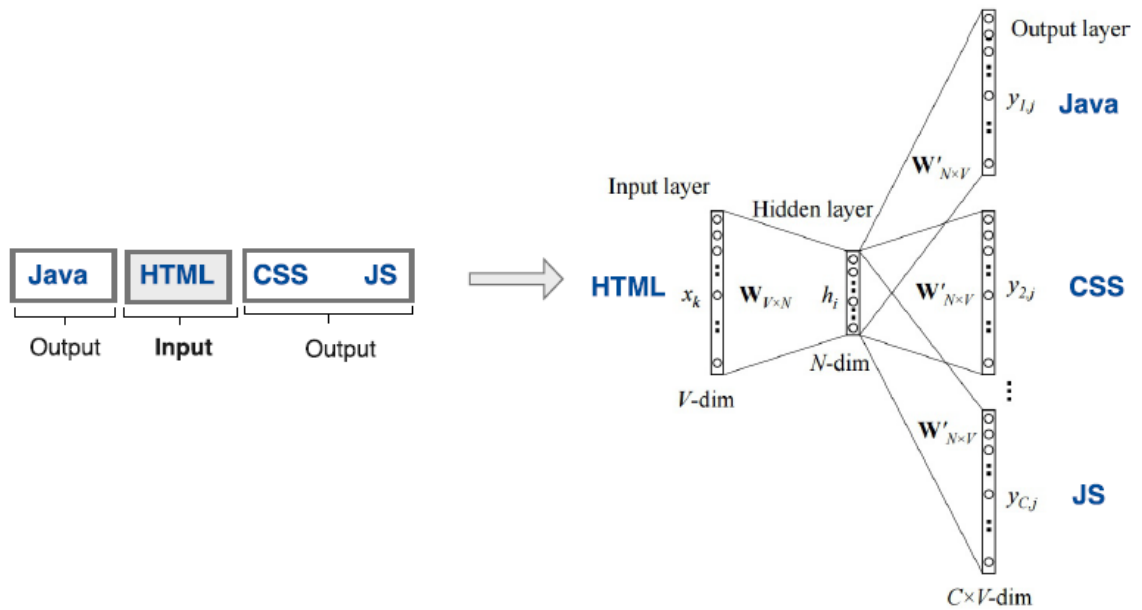
Εν συνεχεία, το λεξικό δεξιοτήτων ήταν έτοιμο για ανάλυση. Συλλέχθηκε ένας τεράστιος αριθμός περιγραφών θέσεων εργασίας από το Dice.com - έναν από τους πιο δημοφιλείς ιστοτόπους σταδιοδρομίας σχετικά με τις θέσεις εργασίας Tech στις ΗΠΑ. Από αυτές τις περιγραφές θέσεων εργασίας, εξάγονται δεξιότητες για κάθε μία χρησιμοποιώντας το λεξικό

δεξιότητων (1). Το σύνολο δεδομένων προκύπτει από μια λίστα συλλογών δεξιότητων με βάση τις περιγραφές εργασίας. Μετά την ανίχνευση, ελήφθησαν συνολικά 5 GB με περισσότερες από 1.400.000 περιγραφές εργασίας. Από αυτά τα δεδομένα, εξήχθησαν οι δεξιότητες και εκτελέστηκαν με τη μορφή λίστας δεξιότητων στο ίδιο πλαίσιο. (το πλαίσιο περιελάμβανε δεξιότητες για την ίδια περιγραφή εργασίας).

3.2.3. Αρχιτεκτονική Skill2vec

Στη μελέτη των (Le Van-Duyet, et al., 2019), για την εκπαίδευση του συνόλου δεδομένων, χρησιμοποιήθηκε νευρωνικό δίκτυο εμπνευσμένο από το μοντέλο Word2Vec. Οι δεξιότητες αντιμετωπίστηκαν ως λέξεις στο μοντέλο Word2Vec. Επιπλέον επιλέχθηκε το μέγιστο μέγεθος παραθύρου με τα έγγραφα να περιέχουν μόνο τις δεξιότητες, υπονοώντας ότι οι δεξιότητες στην ίδια περιγραφή εργασίας σχετίζονται μεταξύ τους. Για τις διαστάσεις του διανύσματος, μετά από κάποιο σημείο, η προσθήκη περισσότερων διαστάσεων παρέχει μειούμενες βελτιώσεις, επομένως η παράμετρος αυτή επιλέχθηκε εμπειρικά. Το μοντέλο ονομάστηκε Skill2Vec το οποίο περιγράφεται εν συντομία στο σχήμα 3.





Σχήμα 3. Αρχιτεκτονική Skill2vec

Με σκοπό την αξιολόγηση της μεθόδου, μια ομάδα εμπειρογνομόνων αξιολογεί το αποτέλεσμα ακολουθώντας αυτά τα βήματα:

- Τυχαία επιλογή 200 δεξιοτήτων από το λεξικό.
- Το σύστημά θα επιστρέψει τις κορυφαίες 5 "πλησιέστερες" δεξιότητες για καθένα.
- Η ομάδα εμπειρογνομόνων θα ελέγξει εάν αυτές οι κορυφαίες 5 "πλησιέστερες" δεξιότητες είναι σχετικές ή όχι.

Το πείραμα έδειξε ότι το 78% των δεξιοτήτων που ελήφθησαν από το συγκεκριμένο μοντέλο ήταν πραγματικά σχετικές με τα εισαχθέντα προσόντα.

Τα πειραματικά αποτελέσματα παρουσιάζονται στον ακόλουθο πίνακα

Πίνακας 1. Ερωτηματολόγιο των 5 πιο σχετικών δεξιοτήτων

Δεξιότητες ερωτήματος	Οι πιο σχετικές δεξιότητες
HTML5	css3
	bootstrap
	front end
	angular
	Responsive
OOP	OOD
	Objective
	Java
	Multi Thread
	Software Debug
Hadoop	Pig
	Hive
	HBase
	Big Data
	Spark
Scala	Zookeeper
	Spark
	Data System
	Sqoop
	solrcloud
Hive	Pig
	HDFS

	Hadoop
	Spark
	Impala

3.3.Μηχανή Αναζήτησης Εργασίας Βάσει Εταιρικών Χαρακτηριστικών Δεξιοτήτων

Στην εργασία των Faliagka et al., (2012) προτείνεται μια μέθοδος ώστε αυτή η αναζήτηση να αναχθεί σε πολλαπλά επίπεδα όπου ένας χρήστης μπορεί να αναζητήσει θέσεις εργασίας με βάση διαφορετικές κατηγορίες όπως τεχνολογίες, δεξιότητες, έσοδα, μέγεθος εταιρείας, βιομηχανία και πολλά άλλα και τα αποτελέσματα αναζήτησης κατατάσσονται λαμβάνοντας υπόψη πολλά χαρακτηριστικά, όπως η κατάταξη Alexa (μετρά την κίνηση στο διαδίκτυο για ιστότοπους, δηλ. δημοτικότητα της εταιρείας), κοινωνικούς οπαδούς, έσοδα, λέξεις-κλειδιά που δίνονται από τον χρήστη (δεξιότητες, βιομηχανία, τεχνολογίες) και πολλά άλλα.

Σε υψηλό επίπεδο, το σύστημά ασχολείται με κείμενο που σχετίζεται με δημοσιεύσεις θέσεων εργασίας στο διαδίκτυο, για να συναγάγει και να κατατάξει το σύνολο δεξιοτήτων για κάθε εργασία χρησιμοποιώντας τις περιγραφές εργασίας και τον τίτλο εργασίας. Το σύνολο δεξιοτήτων που εξάγεται για κάθε ανάρτηση εργασίας συνδυάζεται με τη βάση γνώσεων του Everstring ώστε να το βελτιώσει σύμφωνα με τα χαρακτηριστικά της εταιρείας. Τα συνδυασμένα δεδομένα χρησιμοποιούνται για την αναζήτηση εργασιών με πολύ συγκεκριμένα ερωτήματα αναζήτησης που δίνονται από τον χρήστη/άτομο που αναζητά εργασία. Τα αποτελέσματα αναζήτησης κατατάσσονται τελικά χρησιμοποιώντας βαθμολογία συνάφειας. Η βάση δεδομένων που δημιουργείται μπορεί να χρησιμοποιηθεί για να βρει τις θέσεις εργασίας που ταιριάζουν ακριβώς στις απαιτήσεις των χρηστών και επίσης να δώσει στους χρήστες τις επαφές των σχετικών ατόμων που εργάζονται επί του παρόντος στις εταιρείες των αντίστοιχων θέσεων εργασίας.

Η διαδικασία περιγράφεται παρακάτω.

- Εξαγωγή δεδομένων θέσεων εργασίας και δεξιοτήτων από σελίδες εργασίας εταιρειών και επαγγελματικά κοινωνικά δίκτυα (Malherbe & Aufaure, 2016).
- Εξαγωγή χρησίων διακριτικών δεδομένων και αποθήκευσή τους σε κατανεμημένη βάση δεδομένων - ETL (Extract Transform and Load).

- Ομαλοποίηση, καθαρισμός των δεδομένων και λήψη του ονόματος του τομέα από το όνομα της εταιρείας.
- Εξαγωγή και βαθμολόγηση του συνόλου δεξιοτήτων που απαιτείται για μια εργασία.
- Συγχώνευση με τα δεδομένα Everstring χρησιμοποιώντας API για εξαγωγή χρήσιμων πληροφοριών.
- Εκτέλεση ερωτημάτων αναζήτησης συγκεκριμένου χρήστη στη βάση δεδομένων ώστε να ληφθούν σχετικές θέσεις εργασίας με ταξινομημένη σειρά.
- Εκτέλεση αναλυτικών στοιχείων.

4. Εφαρμογή Αλγορίθμων Μηχανικής Μάθησης σε Διαδικτυακό Σύστημα Πρόσληψης

4.1. Περιγραφή του Προβλήματος

Σύμφωνα με την έρευνα των (De Meo et al., 2007), τα ηλεκτρονικά συστήματα πρόσληψης έχουν σημειώσει μια ραγδαία επέκταση τα τελευταία χρόνια, επιτρέποντας στις υπηρεσίες ανθρώπινου δυναμικού (Human Resources - HR) να στοχεύουν σε ένα πολύ ευρύ κοινό με μικρό κόστος. Ωστόσο, αυτή η κατάσταση μπορεί να είναι αφόρητη για τους οργανισμούς ανθρώπινου δυναμικού, οι οποίοι πρέπει να διαθέσουν δυναμικό για τη χειροκίνητη αξιολόγηση των βιογραφικών των υποψηφίων και την αξιολόγηση της καταλληλότητας των αιτούντων για τις συγκεκριμένες θέσεις.

Η αυτοματοποίηση όμως της διαδικασίας ανάλυσης των προφίλ των αιτούντων για τον προσδιορισμό αυτών που ταιριάζουν στις προδιαγραφές της αντίστοιχης θέσης, θα μπορούσε να οδηγήσει σε αυξημένη αποδοτικότητα. Για παράδειγμα, η εταιρεία SAT telecom ανέφερε εξοικονόμηση κόστους της τάξης του 44% και πτώση του μέσου χρόνου που απαιτείται για την κάλυψη μιας κενής θέσης από 70 σε 37 ημέρες μετά την ανάπτυξη συστήματος ηλεκτρονικής πρόσληψης. Έχουν προταθεί αρκετά συστήματα ηλεκτρονικής πρόσληψης με στόχο την επιτάχυνση της διαδικασίας πρόσληψης, οδηγώντας σε μία καλύτερη συνολική εμπειρία χρήστη. Όπως προκύπτει από την έρευνα των (Kessler et al., 2007), το σύστημα E-Gen πραγματοποιεί ανάλυση και κατηγοριοποίηση αδόμητων προσφορών εργασίας

(δηλαδή, με τη μορφή μη δομημένων εγγράφων κειμένου) καθώς και ανάλυση και κατάταξη σχετικότητας των υποψηφίων.

Μία άλλη περίπτωση συστήματος είναι αυτή που αναφέρεται σε μελέτη των (Radevski & Trichet, 2006) και αφορά το κοινό πλαίσιο (common framework), το οποίο εφαρμόζει τεχνολογίες Σημασιολογικού Ιστού στον τομέα της Διαχείρισης Ανθρώπινου Δυναμικού. Σε αυτό το πλαίσιο, εξετάζονται για πρόσληψη τα χαρακτηριστικά της προσωπικότητας του υποψηφίου, τα οποία καθορίζονται μέσω ενός διαδικτυακού ερωτηματολογίου που συμπληρώνεται από τον ίδιο τον υποψήφιο. Για να γίνει η αντιστοίχιση των αιτούντων με τις θέσεις εργασίας, αυτά τα συστήματα συνδυάζουν συνήθως τεχνικές από κλασικά συστήματα ανάκτησης πληροφορίας (information retrieval – IR) και συστήματα προτάσεων, όπως η ανατροφοδότηση σχετικότητας, η σημασιολογική αντιστοίχιση και η διαδικασία αναλυτικής ιεραρχίας. Μια άλλη προσέγγιση που προτείνεται από τους (Amdouni & Ben Abdessalem Karaa, 2010), χρησιμοποιεί την τεχνολογία επεξεργασίας φυσικής γλώσσας (Natural language processing – NLP) για να αντιπροσωπεύει αυτόματα τα βιογραφικά σημειώματα σε μια τυπική γλώσσα μοντελοποίησης.

Αυτές οι μέθοδοι, αν και είναι χρήσιμες, υποφέρουν από αποκλίσεις που σχετίζονται με ασυνεπείς μορφές βιογραφικού, με τη δομή αυτών αλλά και με σχετικές τους πληροφορίες. Επιπλέον, δεν μπορούν να αξιολογήσουν κάποια δευτερεύοντα χαρακτηριστικά που σχετίζονται με βιογραφικά σημειώματα, όπως είναι το ύφος και η συνοχή, τα οποία είναι πολύ σημαντικά στην αξιολόγηση του εκάστοτε βιογραφικού. Για την επίλυση του προβλήματος κατάταξης των υποψηφίων, δόθηκε βάση στην έρευνα που έγινε από τους (Faliagka et al., 2012), οι οποίοι προτείνουν την εφαρμογή εποπτευόμενων αλγορίθμων μάθησης σε αυτοματοποιημένα συστήματα ηλεκτρονικής πρόσληψης. Έχουν ήδη εφαρμόσει και δοκιμάσει ένα ολοκληρωμένο σύστημα ηλεκτρονικής πρόσληψης προσανατολισμένο στην εταιρεία που αυτοματοποιεί τη διαδικασία προεπιλογής και κατάταξης των υποψηφίων. Στο προτεινόμενο σύστημα, η αξιολόγηση των αιτούντων βασίζεται σε ένα προκαθορισμένο σύνολο αντικειμενικών κριτηρίων, τα οποία εξάγονται απευθείας από το αντίστοιχο προφίλ LinkedIn του αιτούντος. Επιπλέον, τα χαρακτηριστικά της προσωπικότητας του υποψηφίου, τα οποία εξάγονται αυτόματα από την κοινωνική του παρουσία, λαμβάνονται υπόψη στην αξιολόγηση. Στόχος λοιπόν των συγγραφέων (Faliagka et al., 2012), είναι να περιορίσουν τη συνέντευξη και τη διερεύνηση υποψηφίων μόνο στους κορυφαίους υποψηφίους που προσδιορίζονται από το σύστημα, έτσι ώστε να αυξήσουν την αποτελεσματικότητα της

διαδικασίας πρόσληψης. Το σύστημα έχει σχεδιαστεί με στόχο να ενσωματωθεί στην υποδομή Διαχείρισης Ανθρώπινου Δυναμικού των εταιρειών, βοηθώντας και όχι αντικαθιστώντας, τους υπεύθυνους πρόσληψης στη διαδικασία λήψης αποφάσεων.

Στη συνέχεια, αξιοποιούνται τεχνικές μηχανικής μάθησης για την επίλυση του προβλήματος κατάταξης υποψηφίων σε συστήματα ηλεκτρονικής πρόσληψης. Στο πρόβλημα κατάταξης υποψηφίων, μια συνάρτηση μοριοδότησης $h(x)$ εξάγει τη βαθμολογία συνάφειας του υποψηφίου, η οποία αντικατοπτρίζει πόσο καλά ένα προφίλ υποψηφίου ταιριάζει στις απαιτήσεις μιας δεδομένης θέσης εργασίας. Δεδομένου ότι η βαθμολογία συνάφειας είναι μια συνεχής μεταβλητή, το πρόβλημα κατάταξης υποψηφίου μπορεί να μειωθεί σε πρόβλημα παλινδρόμησης, στο οποίο πρέπει να διδαχθεί η λειτουργία βαθμολόγησης υποψηφίου χρησιμοποιώντας τεχνικές μάθησης με επίβλεψη. Έπειτα, το σύστημα εξάγει την τελική λίστα εφαρμόζοντας την ήδη διδαγμένη λειτουργία για να ταξινομήσει τους υποψηφίους. Η συνάρτηση βαθμολογίας $h(x)$ αντλεί το βαθμό συνάφειας του υποψηφίου y_i από τις τιμές του διανύσματος χαρακτηριστικών x_i . Στην εργασία αυτή (Faliagka et al., 2012), το διάνυσμα χαρακτηριστικών x_i αποτελείται από ένα σύνολο χαρακτηριστικών $\{a_1, ..., a_m\}$ που αντιστοιχούν στα κριτήρια επιλογής του υποψηφίου. Αυτές μπορεί να είναι είτε συνεχείς μεταβλητές (που αντιπροσωπεύουν το χαρακτηριστικό ενός υποψηφίου που έχει αξιολογηθεί σε αριθμητική κλίμακα) είτε Boolean μεταβλητές (δηλώνοντας εάν έχει την επιθυμητή ικανότητα ή όχι). Η πραγματική συνάρτηση βαθμολόγησης είναι συνήθως άγνωστη και γίνεται εκμάθηση μιας προσέγγισης μέσω του συνόλου εκπαίδευσης D . Στο προτεινόμενο σύστημα το σύνολο εκπαίδευσης αποτελείται από ένα σύνολο N παραδειγμάτων επιλογής προηγούμενων υποψηφίων, που δίνονται ως είσοδος στο σύστημα:

$$D = \left\{ (x_i, y_i) \mid x_i \in R^m, y_i \in R \right\}_{i=1}^N \quad (\text{Συνάρτηση 1})$$

Στους αλγορίθμους που ακολουθούν, παρουσιάζεται ένα σύνολο αντιπροσωπευτικών αλγορίθμων «εκμάθησης κατάταξης» που χαρτογραφούν το εκπαιδευτικό σύνολο των προηγούμενων αποφάσεων πρόσληψης σε ένα μοντέλο παλινδρόμησης που χρησιμεύει ως προγνωστικό στοιχείο μελλοντικών αποφάσεων πρόσληψης. Οι συγγραφείς υλοποίησαν 4 αλγορίθμους παλινδρόμησης Linear Regression (LR), M5' model tree (M5'), REPTree decision tree (REP), and Support Vector Regression (SVR) και κατέληξαν ότι καλύτερα αποτελέσματα είχε ο SVR. Εμείς για να επεκτείνουμε την εργασία, θεωρούμε το πρόβλημα

μας ότι έχει κατηγορικές μεταβλητές και προτείνουμε μεθόδους ταξινόμησης. Παρουσιάζονται 3 παραδοσιακοί και διαδοσμένοι αλγόριθμοι πρόβλεψης, οι οποίοι αναλύονται στα επόμενα υποκεφάλαια.

4.2. Δέντρο Αποφάσεων

Το δέντρο αποφάσεων είναι ένα εργαλείο υποστήριξης αποφάσεων, που χρησιμοποιεί ένα μοντέλο αποφάσεων το οποίο έχει τη μορφή δέντρου και τις πιθανές συνέπειες των αποφάσεων αυτών, συμπεριλαμβανομένων αποτελεσμάτων τυχαίων γεγονότων, κόστους πόρων και χρησιμότητας (Goodfellow et al., 2016). Είναι ένας τρόπος εμφάνισης ενός αλγορίθμου που περιέχει μόνο δηλώσεις ελέγχου υπό όρους. Τα δέντρα αποφάσεων χρησιμοποιούνται συνήθως στην επιχειρησιακή έρευνα, συγκεκριμένα στην ανάλυση αποφάσεων, για να βοηθήσουν στον εντοπισμό μιας στρατηγικής που είναι πιθανότερο να επιτευχθεί ένας στόχος, αλλά αποτελούν επίσης ένα δημοφιλές εργαλείο στη μηχανική μάθηση.

Ένα δέντρο αποφάσεων είναι μια δομή που μοιάζει με ένα διάγραμμα ροής, στην οποία κάθε εσωτερικός κόμβος αντιπροσωπεύει ένα "τεστ" σε ένα χαρακτηριστικό (π.χ. αν στο στρίψιμο ενός νομίσματος έρθει κορώνα ή γράμματα), κάθε κλάδος αντιπροσωπεύει το αποτέλεσμα της δοκιμής και κάθε κόμβος φύλλου αντιπροσωπεύει μία ετικέτα κλάσης (απόφαση που ελήφθη μετά τον υπολογισμό όλων των χαρακτηριστικών). Οι διαδρομές από τη ρίζα στο φύλλο αντιπροσωπεύουν κανόνες ταξινόμησης.

Στην ανάλυση αποφάσεων, ένα δέντρο αποφάσεων και το στενά συνδεδεμένο διάγραμμα επιρροής χρησιμοποιούνται ως οπτικό και αναλυτικό εργαλείο υποστήριξης αποφάσεων, όπου υπολογίζονται οι αναμενόμενες τιμές (ή αναμενόμενη χρησιμότητα) των ανταγωνιστικών εναλλακτικών λύσεων.

Ένα δέντρο αποφάσεων αποτελείται από τρεις τύπους κόμβων:

1. Κόμβοι απόφασης - συνήθως αντιπροσωπεύονται από τετράγωνα
2. Κόμβοι πιθανότητας - συνήθως αντιπροσωπεύονται από κύκλους
3. Τερματικοί κόμβοι - συνήθως αντιπροσωπεύονται από τρίγωνα

Τα δέντρα αποφάσεων χρησιμοποιούνται συνήθως στην έρευνα επιχειρήσεων και τη διαχείριση επιχειρήσεων. Εάν, στην πράξη, οι αποφάσεις πρέπει να λαμβάνονται διαδικτυακά χωρίς ανάκληση υπό ελλιπείς γνώσεις, ένα δέντρο αποφάσεων θα πρέπει να

παραλληλίζεται με ένα μοντέλο πιθανότητας ως μοντέλο καλύτερης επιλογής ή αλγόριθμος μοντέλου επιλογής στο διαδίκτυο. Μια άλλη χρήση των δέντρων αποφάσεων είναι ως ένα περιγραφικό μέσο για τον υπολογισμό των πιθανοτήτων υπό όρους.

4.3. Μηχανές Διανυσμάτων Υποστήριξης

Ο στόχος του αλγόριθμου Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines -SVM), είναι να βρει ένα υπερεπίπεδο σε χώρο N -διαστάσεων (N , ο αριθμός των χαρακτηριστικών) που να ταξινομεί ευδιάκριτα τα σημεία δεδομένων (Goodfellow et al., 2016). Τα SVM είναι ένα σύνολο σχετικών μεθόδων για εποπτευόμενη μάθηση, που εφαρμόζονται τόσο σε προβλήματα ταξινόμησης όσο και σε παλινδρόμηση. Η δύναμη των SVM προέρχεται από την αναπαράσταση του πυρήνα, η οποία επιτρέπει μια μη γραμμική χαρτογράφηση του χώρου εισόδου σε έναν χώρο χαρακτηριστικών υψηλότερης διάστασης. Ο στόχος της Υποστήριξης Διανυσματικής Παλινδρόμησης είναι να βρεθεί μια συνάρτηση f που ελαχιστοποιεί το αναμενόμενο σφάλμα, δηλαδή το ολοκλήρωμα μιας συγκεκριμένης συνάρτησης απώλειας - σύμφωνα με την άγνωστη κατανομή πιθανότητας των δεδομένων. Αυτό ελαχιστοποιεί τον εμπειρικό κίνδυνο ότι η εκτιμώμενη συνάρτηση διαφέρει από την αρχική (ακόμη άγνωστη). Υποθέτοντας ότι το N είναι τα σημεία δεδομένων και το K ο πυρήνας, τα διανύσματα υποστήριξης και οι τιμές υποστήριξης της λύσης ορίζουν την ακόλουθη συνάρτηση παλινδρόμησης:

$$f(x) = \sum_{i=1}^N a_i K(x, x_i) + b \mid b, a_i \in R \quad (\text{Συνάρτηση 3})$$

Για παράδειγμα, για να διαχωριστούν δύο κατηγορίες σημείων δεδομένων, υπάρχουν πολλά πιθανά υπερεπίπεδα που θα μπορούσαν να επιλεγούν. Στόχος είναι να βρεθεί ένα επίπεδο που να έχει το μέγιστο περιθώριο, δηλαδή τη μέγιστη απόσταση μεταξύ των σημείων δεδομένων και των δύο κλάσεων. Η μεγιστοποίηση της απόστασης περιθωρίου παρέχει κάποια ενίσχυση, έτσι ώστε τα μελλοντικά σημεία δεδομένων να μπορούν να ταξινομηθούν με μεγαλύτερη εμπιστοσύνη. Τα υπερεπίπεδα είναι όρια αποφάσεων που βοηθούν στην ταξινόμηση των σημείων δεδομένων. Τα σημεία δεδομένων που πέφτουν εκατέρωθεν του υπερεπιπέδου μπορούν να αποδοθούν σε διαφορετικές κλάσεις. Επίσης, η διάσταση του υπερεπιπέδου εξαρτάται από τον αριθμό των χαρακτηριστικών. Εάν ο αριθμός των χαρακτηριστικών εισόδου είναι 2, τότε το υπερεπίπεδο είναι απλώς μια γραμμή. Εάν ο

αριθμός των χαρακτηριστικών εισόδου είναι 3, τότε το υπερεπίπεδο γίνεται δισδιάστατο επίπεδο. Είναι δύσκολο να φανταστεί κανείς τι μπορεί να συμβεί όταν ο αριθμός των χαρακτηριστικών υπερβεί το 3.

Τα διανύσματα υποστήριξης είναι σημεία δεδομένων που βρίσκονται πιο κοντά στο υπερεπίπεδο και επηρεάζουν τη θέση και τον προσανατολισμό του. Χρησιμοποιώντας αυτά τα διανύσματα υποστήριξης, μεγιστοποιείται το περιθώριο του ταξινομητή. Σε περίπτωση διαγραφής των διανυσμάτων υποστήριξης, θα αλλάξει και η θέση του υπερεπιπέδου.

4.4. Αλγόριθμος K-Nearest Neighbors

Ο αλγόριθμος k-πλησιέστερων γειτόνων (k-nearest neighbors - k-NN) είναι μια μη παραμετρική μέθοδος ταξινόμησης που χρησιμοποιείται για ταξινόμηση και παλινδρόμηση. Και στις δύο περιπτώσεις, η εισαγωγή αποτελείται από τα πλησιέστερα παραδείγματα εκπαίδευσης στο σύνολο δεδομένων (Goodfellow et al., 2016). Η έξοδος εξαρτάται από τον τρόπο που χρησιμοποιείται ο αλγόριθμος. Δηλαδή είτε για ταξινόμηση είτε για παλινδρόμηση:

- Στην ταξινόμηση k-NN, το αποτέλεσμα είναι μια ιδιότητα μέλους. Ένα αντικείμενο ταξινομείται με πολλαπλή ψήφο των γειτόνων του, με το αντικείμενο να εκχωρείται στην πιο κοινή κλάση μεταξύ των k πλησιέστερων γειτόνων του (το k είναι ένας θετικός ακέραιος αριθμός, συνήθως μικρός). Εάν $k = 1$, τότε το αντικείμενο απλώς εκχωρείται στην κλάση του ενιαίου πλησιέστερου γείτονα.
- Στην παλινδρόμηση k-NN, η έξοδος είναι η τιμή ιδιότητας για το αντικείμενο. Αυτή η τιμή είναι ο μέσος όρος των τιμών των k πλησιέστερων γειτόνων.

Ο k-NN είναι ένας τύπος ταξινόμησης όπου η συνάρτηση προσεγγίζεται μόνο τοπικά και όλος ο υπολογισμός αναβάλλεται μέχρι την αξιολόγηση της συνάρτησης. Δεδομένου ότι αυτός ο αλγόριθμος βασίζεται στην απόσταση για ταξινόμηση, εάν τα χαρακτηριστικά αντιπροσωπεύουν διαφορετικές φυσικές μονάδες ή έρχονται σε πολύ διαφορετικές κλίμακες, τότε η ομαλοποίηση των δεδομένων εκπαίδευσης μπορεί να βελτιώσει δραματικά την ακρίβειά του.

Τόσο για την ταξινόμηση όσο και για την παλινδρόμηση, μια χρήσιμη τεχνική μπορεί να είναι ο καθορισμός βαρών στις συνεισφορές των γειτόνων, έτσι ώστε οι πλησιέστεροι γείτονες να συνεισφέρουν περισσότερο στο μέσο όρο από τους πιο μακρινούς. Για

παράδειγμα, ένα κοινό σχήμα στάθμισης συνίσταται στο να δοθεί βάρος σε κάθε γείτονα $1/d$, όπου d είναι η απόσταση από τον γείτονα.

Οι γείτονες λαμβάνονται από ένα σύνολο αντικειμένων για τα οποία είναι γνωστή η κλάση (για ταξινόμηση k-NN) ή η τιμή ιδιότητας αντικειμένου (για παλινδρόμηση k-NN). Αυτό μπορεί να θεωρηθεί ως το σύνολο εκπαίδευσης για τον αλγόριθμο, αν και δεν απαιτείται ρητό βήμα εκπαίδευσης.

5. Συγκριτική Αξιολόγηση τεχνικών Μηχανικής Μάθησης στην ηλεκτρονική εύρεση εργασίας

Το προτεινόμενο σύστημα δοκιμάστηκε σε πραγματικό σενάριο πρόσληψης για να αξιολογηθεί η αποτελεσματικότητά του στην κατάταξη των αιτούντων εργασίας. Η αξιολόγηση της απόδοσης του συστήματος βασίζεται στο πόσο αποτελεσματική είναι στην εκχώρηση συνεπών βαθμολογιών συνάφειας στους υποψηφίους, σε σύγκριση με αυτές που έχουν ανατεθεί από τους υπεύθυνους προσλήψεων.

Στο σενάριο προσλήψεων που χρησιμοποιήθηκε στις δοκιμές των συγγραφέων (Faliagka et al., 2012), συντάχθηκε ένα σύνολο 100 αιτούντων με λογαριασμό LinkedIn και προσωπικό ιστολόγιο, καθώς αυτές είναι οι βασικές απαιτήσεις του προτεινόμενου συστήματος. Οι αιτούντες επιλέχθηκαν τυχαία μέσω της διεπαφής προγραμματισμού εφαρμογών (application programming interface – API) του ιστολογίου της Google, με μοναδική απαίτηση να έχουν τεχνικό υπόβαθρο, όπως υποδεικνύεται από τα μεταδεδομένα ιστολογίου (λίστα ενδιαφερόντων), καθώς και από ένα προφίλ LinkedIn. Το σώμα των αιτούντων εργασίας δημιουργήθηκε επιλέγοντας τα πρώτα 100 ιστολόγια που πληρούσαν τις προϋποθέσεις του σεναρίου μέσω της API. Συλλέχθηκαν επίσης τρεις αντιπροσωπευτικές τεχνικές θέσεις που ανακοινώθηκαν από μια ανώνυμη εταιρεία πληροφορικής με διαφορετικές απαιτήσεις, δηλαδή, θέση μηχανικού πωλήσεων, θέση κατώτερου προγραμματιστή και θέση ανώτερου προγραμματιστή. Η θέση του μηχανικού πωλήσεων ευνοεί έναν υψηλό βαθμό εξωστρέφειας, ενώ το πιο σημαντικό χαρακτηριστικό για ανώτερους προγραμματιστές είναι η εμπειρία. Οι νεότεροι προγραμματιστές κρίνονται κυρίως από την αφοσίωση τους (επειδή μια εταιρεία δεν θα επενδύσει στην εκπαίδευση ενός ατόμου που έχει τη τάση να αλλάζει συχνά θέσεις) καθώς και στην εκπαίδευση. Επιπλέον, κάθε θέση έχει το δικό της επιθυμητό σύνολο δεξιοτήτων, οι οποίες ταιριάζουν με το σύνολο δεξιοτήτων που αναφέρει ο κάθε χρήστης στο προφίλ του στο LinkedIn. Συγκεκριμένα, η κατώτερη θέση απαιτεί δεξιότητες προγραμματισμού σε γλώσσες ανάπτυξης C++ ή Java, ενώ η ανώτερη θέση απαιτεί εμπειρία 5 ετών σε τεχνολογίες J2EE. Η χρήση διαφορετικών απαιτήσεων ανά θέση αναμένεται να δοκιμάσει την ικανότητα του προαναφερθέντος συστήματος. Δηλαδή να ταιριάζει τα προφίλ των υποψηφίων με την κατάλληλη θέση εργασίας.

Στα πειράματα των συγγραφέων (Faliagka et al., 2012), εκλαμβάνεται η υπόθεση ότι κάθε υποψήφιος έχει υποβάλει αίτηση και για τις τρεις διαθέσιμες θέσεις εργασίας. Για κάθε θέση

εργασίας, οι υποψήφιοι κατατάχθηκαν ανάλογα με την καταλληλότητά τους τόσο από το σύστημα (αυτοματοποιημένη κατάταξη) όσο και από έναν έμπειρο υπεύθυνο πρόσληψης. Οι υπεύθυνοι προσλήψεων είχαν πρόσβαση στις ίδιες πληροφορίες με το σύστημα, δηλαδή στο ιστολόγιο του εκάστοτε υποψηφίου αλλά και στο προφίλ του στο LinkedIn. Πρέπει ωστόσο να σημειωθεί ότι παρά το γεγονός ότι τα κριτήρια επιλογής είναι ήδη γνωστά στο σύστημα, η ερμηνεία των δεδομένων και η ακριβής διαδικασία λήψης αποφάσεων, είναι άγνωστες ως προς τους υπεύθυνους προσλήψεων και κατ' επέκταση θα ήταν φρόνιμο να τα διδαχθούν προς διευκόλυνση της διαδικασίας. Σε ένα από τα πειράματά τους και σύμφωνα με τη μελέτη των (Hall et al., 2009), χρησιμοποιείται το λογισμικό Weka για να αξιολογηθούν τα μοντέλα κατάταξης της μάθησης. Συγκεκριμένα, δοκιμάζεται η συσχέτιση των αποτελεσμάτων που προκύπτουν από το σύστημα (δηλαδή, προβλέψεις μοντέλων) με τις πραγματικές βαθμολογίες που έχουν εκχωρηθεί από τους υπεύθυνους προσλήψεων, χρησιμοποιώντας τη μέτρηση του συντελεστή συσχέτισης Pearson.

Στην παρούσα εργασία γίνεται μια συγκριτική αξιολόγηση των τριών παραδοσιακών αλγορίθμων κατηγοριοποίησης σχετικά με την εφαρμογή τους σε προβλήματα ηλεκτρονικής πρόσληψης. Ο SVM λειτουργεί σχετικά καλά όταν υπάρχει σαφές περιθώριο διαχωρισμού μεταξύ των κλάσεων και είναι πιο αποτελεσματική σε χώρους υψηλών διαστάσεων. Επίσης, είναι αποτελεσματικός σε περιπτώσεις όπου ο αριθμός των διαστάσεων είναι μεγαλύτερος από τον αριθμό των δειγμάτων. Ωστόσο τα δεδομένα σε προβλήματα ηλεκτρονικής πρόσληψης συνήθως δεν είναι πολυδιάστατα μιας και για κάθε υποψήφιο εργαζόμενο, τα χαρακτηριστικά που έχουμε δεν είναι μεγάλης κλίμακας. Σε τέτοιες περιπτώσεις, δε θα επιλέγαμε τον αλγόριθμο SVM. Επίσης, ο αλγόριθμος SVM δεν είναι κατάλληλος για μεγάλα σύνολα δεδομένων. Ο SVM δεν αποδίδει πολύ καλά όταν το σύνολο δεδομένων έχει αρκετό θόρυβο, κάτι που όμως δεν ισχύει στα δεδομένα που σχετίζονται με προβλήματα ηλεκτρονικής πρόσληψης, συνεπώς είναι ένα χαρακτηριστικό που δε μας επηρεάζει. Ένα άλλο χαρακτηριστικό είναι ότι σε περιπτώσεις όπου ο αριθμός των χαρακτηριστικών για κάθε σημείο δεδομένων υπερβαίνει τον αριθμό των δειγμάτων δεδομένων εκπαίδευσης, ο SVM θα υπολειτουργεί.

Σχετικά με τα δέντρα αποφάσεων, εκτός από απλά για κατανόηση και ερμηνεία, έχουν αξία ακόμη και με ελάχιστα δεδομένα. Μπορούν να παραχθούν σημαντικές γνώσεις και σε δεδομένα με λίγα δείγματα, κάτι που δεν ισχύει πχ. στα τεχνητά νευρωνικά δίκτυα βαθιάς μάθησης. Από την άλλη, είναι ασταθή, πράγμα που σημαίνει ότι μια μικρή αλλαγή στα

δεδομένα μπορεί να οδηγήσει σε μια μεγάλη αλλαγή στη δομή του βέλτιστου δέντρου αποφάσεων. Αυτό μας αποτρέπει στην χρησιμοποίησή του, όπως και το ότι είναι σχετικά ανακριβή. Πολλοί άλλοι προγνωστικοί παράγοντες έχουν καλύτερη απόδοση με παρόμοια δεδομένα. Αυτό μπορεί να διορθωθεί με την αντικατάσταση ενός μεμονωμένου δέντρου αποφάσεων με ένα τυχαίο δάσος αποτελούμενο από δέντρα αποφάσεων, αλλά με το μειονέκτημα ότι ένα τυχαίο δάσος δεν είναι τόσο εύκολο να ερμηνευτεί όσο ένα δέντρο αποφάσεων. Για δεδομένα που περιλαμβάνουν κατηγορικές μεταβλητές με διαφορετικό αριθμό επιπέδων, το κέρδος πληροφοριών στα δέντρα αποφάσεων είναι προκατειλημμένο υπέρ αυτών των χαρακτηριστικών με περισσότερα επίπεδα.

Ο KNN κατά βάση δεν έχει διαδικασία εκπαίδευσης μιας και αποθηκεύει το σύνολο δεδομένων εκπαίδευσης και μαθαίνει από αυτό μόνο τη στιγμή που κάνει προβλέψεις σε πραγματικό χρόνο. Αυτό καθιστά τον αλγόριθμο KNN πολύ πιο γρήγορο από άλλους αλγόριθμους που απαιτούν εκπαίδευση π.χ. SVM, Γραμμική παλινδρόμηση κλπ. Δεδομένου ότι ο αλγόριθμος KNN δεν απαιτεί εκπαίδευση πριν κάνει προβλέψεις, μπορούν να προστεθούν νέα δεδομένα χωρίς προβλήματα, τα οποία δεν θα επηρεάσουν την ακρίβεια του αλγορίθμου. Ο KNN είναι πολύ εύκολο να εφαρμοστεί. Απαιτούνται μόνο δύο παράμετροι για την εφαρμογή του, δηλαδή η τιμή του K και η συνάρτηση απόστασης (π.χ. Ευκλείδεια ή Μανχάταν κ.λπ.). Από την άλλη μεριά, ο KNN δεν λειτουργεί καλά με μεγάλα σύνολα δεδομένων γιατί σε μεγάλα σύνολα δεδομένων, το κόστος υπολογισμού της απόστασης μεταξύ του νέου σημείου και κάθε υπάρχοντος σημείου είναι τεράστιο, γεγονός που υποβαθμίζει την απόδοση του αλγορίθμου.

Επίσης, δεν λειτουργεί καλά με υψηλές διαστάσεις: Ο αλγόριθμος KNN δεν λειτουργεί καλά με δεδομένα μεγάλων διαστάσεων, επειδή με μεγάλο αριθμό διαστάσεων, καθίσταται δύσκολο για τον αλγόριθμο να υπολογίσει την απόσταση σε κάθε διάσταση. Τέλος, είναι ευαίσθητος σε θορυβώδη δεδομένα, σε τιμές που λείπουν και σε υπερβολικές τιμές: Ο KNN είναι ευαίσθητος στο θόρυβο στο σύνολο δεδομένων. Οι τιμές που λείπουν πρέπει να καταγραφούν μη αυτόματα και οι υπερβολικές τιμές να αφαιρεθούν.

6. Συμπεράσματα-Μελλοντικές κατευθύνσεις

Ένα σύστημα σύστασης είναι μια υποκατηγορία συστημάτων φιλτραρίσματος πληροφοριών. Οι αλγόριθμοι που χρησιμοποιούν τα συστήματα σύστασης εξοικονομούν χρόνο και χρήμα στις εταιρείες, μειώνοντας αρκετά το κόστος εύρεσης του ιδανικού υποψηφίου. Το μόνο που χρειάζεται να κάνουν οι εταιρείες είναι να τροφοδοτούν τα δεδομένα του λογισμικού για διαφορετικές παραμέτρους μιας θέσης εργασίας, όπως είναι τα χρόνια εργασιακής εμπειρίας, η γνώση των γλωσσών προγραμματισμού, η εκπαίδευση, οι ξένες γλώσσες και άλλα. Ο εκάστοτε αλγόριθμος θα περάσει από προφίλ στο LinkedIn ή σε άλλες δημόσιες βάσεις δεδομένων και θα βρει βιογραφικά που ταιριάζουν καλύτερα στις πληροφορίες που παρέχονται από τους υπεύθυνους πρόσληψης. Μόλις γίνει αυτό, τα επιλεγμένα βιογραφικά αποστέλλονται αυτόματα στους υπεύθυνους προσλήψεων (HR managers), ώστε να μπορούν να κάνουν έναν χειροκίνητο έλεγχο και στη συνέχεια να περάσουν στο στάδιο των συνεντεύξεων. Οι ερευνητές που σχεδιάζουν συστήματα σύστασης συχνά τα ταξινομούν σε δύο κύριες προσεγγίσεις, δηλαδή content-based filtering και collaborative filtering. Οι content-based μέθοδοι προβλέπουν προτιμήσεις σχετικά με ένα σύνολο στοιχείων με βάση την ομοιότητά τους με τα στοιχεία που ένας χρήστης έχει προτιμήσει στο παρελθόν (δηλαδή με το προφίλ προτιμήσεων ενός χρήστη). Εξελιγμένες μέθοδοι χρησιμοποιούν τεχνικές μηχανικής εκμάθησης όπως Bayesian Classifiers, cluster analysis, decision trees, and artificial neural networks. Οι δημοφιλείς προσεγγίσεις opinion-based χρησιμοποιούν διάφορες τεχνικές όπως text mining, information retrieval, sentiment analysis και deep learning. Στο collaborative filtering τα στοιχεία συνιστώνται σε έναν χρήστη με βάση τις προτιμήσεις παρόμοιων χρηστών. Συνεπώς τα μοντέλα αυτά κάνουν προβλέψεις σχετικά με τα ενδιαφέροντα των χρηστών βάσει των ιστορικών αλληλεπιδράσεων στοιχείων-χρηστών είτε με άμεσα / έμμεσα σχόλια. Επομένως, στην πρώτη περίπτωση, τα προφίλ προτιμήσεων των χρηστών πρέπει να συγκρίνονται με τις περιγραφές των στοιχείων και στην τελευταία περίπτωση, πρέπει να προσδιοριστεί η ομοιότητα μεταξύ των προφίλ των χρηστών. Οι πιο συνηθισμένες επιλογές είναι η ομοιότητα Jaccard και συνημίτονο. Μεταξύ άλλων δημοφιλών μέτρων, υπάρχει ο συσχετισμός Pearson, kNN και η τυπική απόσταση Ευκλείδειας (αντιμετωπίζεται ως μέτρο ανομοιότητας. Νέα μοντέλα μπορούν να κατασκευαστούν σε περιβάλλον R, Python. (Janusz et al., 2018). Υπάρχουν επίσης οι υβριδικές μέθοδοι σύστασης οι οποίες είναι οι προσεγγίσεις που ενσωματώνουν δύο ή περισσότερους τύπους

στρατηγικών προτάσεων. Οι Yao et al. παρουσίασαν ένα hybrid recommender system που ενσωμάτωσε τη σχέση content-based και interaction-based, για τα μη δομημένα δεδομένα, όπως ομοιότητα μεταξύ εργασίας και προφίλ χρήστη, η λανθάνουσα σημασιολογική ανάλυση εφαρμόζεται στο σύστημα.

Υπάρχει έκκληση για υποστήριξη των δραστηριοτήτων σχεδιασμού πολιτικής και αξιολόγησης μέσω της λήψης αποφάσεων βάσει τεκμηριωμένων στοιχείων, όπως τονίζεται στην ανακοίνωση της Ευρωπαϊκής Επιτροπής «Νέες δεξιότητες για νέες θέσεις εργασίας» και σε μία από τις εμβληματικές πρωτοβουλίες της στρατηγικής «Ευρώπη 2020», «Ατζέντα για νέες δεξιότητες και θέσεις εργασίας». Επιπλέον, το 2010 η Ευρωπαϊκή Επιτροπή δημοσίευσε την ανακοίνωση «Μια νέα ώθηση για την ευρωπαϊκή συνεργασία στην επαγγελματική εκπαίδευση και κατάρτιση (Vocational Education and Training-VET) για την υποστήριξη της Ευρωπαϊκής Στρατηγικής 2020», με στόχο την προώθηση των εκπαιδευτικών συστημάτων εν γένει, και ιδίως την VET. Το 2016 η Ευρωπαϊκή Επιτροπή έχει επισημάνει τη σημασία των δραστηριοτήτων της VET, καθώς "εκτιμώνται για την προώθηση ειδικών θέσεων εργασίας και δεξιοτήτων, διευκολύνοντας τη μετάβαση στην απασχόληση και διατηρώντας και ενημερώνοντας τις δεξιότητες του εργατικού δυναμικού σύμφωνα με τις τομεακές, περιφερειακές και τοπικές ανάγκες".

Εστιάζοντας στον τομέα της αγοράς εργασίας, η εξαγωγή σημαντικών πληροφοριών από μη δομημένα κείμενα έχει αφιερωθεί κυρίως στην υποστήριξη της διαδικασίας ηλεκτρονικής πρόσληψης (Lee, 2011) π.χ. αντιστοίχιση των προφίλ των υποψηφίων με τις περιγραφές εργασίας (Hong et al., 2013). Αυτό το πρόβλημα είναι επίσης σχετικό για επιχειρηματικούς σκοπούς και έχει παρακινήσει την ανάπτυξη πολλών εμπορικών προϊόντων παρέχοντας σε άτομα που αναζητούν εργασία και στις εταιρείες εργαλεία ικανότητας αντιστοίχισης δεξιοτήτων. Όσον αφορά τις εταιρείες, είναι έντονη η ανάγκη τους να αυτοματοποιήσουν τις δραστηριότητες του τμήματος Ανθρώπινου Δυναμικού (HR). Κατά συνέπεια, τα τελευταία χρόνια έχει αναπτυχθεί ένας αυξανόμενος αριθμός προϊόντων που ταιριάζουν με τις δεξιότητες, όπως για παράδειγμα ισχύει με τις εταιρείες BurningGlass, Workday, Pluralsight, EmployInsight και TextKernel. Μέχρι σήμερα, η μόνη εμπορική λύση που χρησιμοποιεί τόσο τα συστήματα ISCO όσο και τα συστήματα ESCO(Η ταξινόμηση ESCO είναι πρωτοβουλία που εντάσσεται στην «Ευρώπη 2020».) ως θησαυρούς είναι η Janzz: μια διαδικτυακή πλατφόρμα για την αντιστοίχιση της ζήτησης και της προσφοράς εργασίας τόσο στον

δημόσιο όσο και στον ιδιωτικό τομέα. Παρέχει επίσης πρόσβαση API στη βάση γνώσεων, αλλά δεν αποσκοπεί στην ταξινόμηση των θέσεων εργασίας. Αξίζει επίσης να σημειωθεί το Google Job Search API, μια υπηρεσία pay-as-you-go που ανακοινώθηκε το 2016 για την ταξινόμηση των κενών θέσεων εργασίας μέσω της υπηρεσίας Google Machine Learning στο O*NET (Επαγγελματικό Δίκτυο Πληροφοριών το οποίο είναι μια δωρεάν ηλεκτρονική βάση δεδομένων που περιέχει εκατοντάδες επαγγελματικούς ορισμούς και προορίζεται για χρήση από άτομα που αναζητούν εργασία, ανάπτυξη εργατικού δυναμικού και επαγγελματίες ανθρώπινου δυναμικού, φοιτητές, ερευνητές και πολλά άλλα), η οποία αποτελεί τυπική ταξινόμηση των ΗΠΑ. Αν και αυτή η εμπορική υπηρεσία εξακολουθεί να είναι κλειστή *alpha*, είναι αρκετά ελπιδοφόρα και ρίχνει επίσης φως στις ανάγκες συλλογισμού με κενές θέσεις εργασίας στο Διαδίκτυο χρησιμοποιώντας μια κοινή ταξινόμηση ως βασική γραμμή.

Από τις αρχές της δεκαετίας του '90, η ταξινόμηση κειμένου (text classification-TC) υπήρξε ενεργό ερευνητικό θέμα. Έχει οριστεί ως «η δραστηριότητα επισήμανσης κειμένων φυσικής γλώσσας με θεματικές κατηγορίες από ένα προκαθορισμένο σύνολο» (Sebastiani, 2002). Οι πιο δημοφιλείς τεχνικές βασίζονται στο παράδειγμα της μηχανικής μάθησης, σύμφωνα με το οποίο δημιουργείται ένας αυτόματος ταξινομητής κειμένου χρησιμοποιώντας μια επαγωγική διαδικασία ικανή να μάθει, από ένα σύνολο προ-διαβαθμισμένων εγγράφων, τα χαρακτηριστικά των κατηγοριών ενδιαφέροντος. Η περίπτωση κατά την οποία μια κατηγορία πρέπει να εκχωρηθεί σε κάθε έγγραφο ονομάζεται ταξινόμηση μιας ετικέτας, ενώ η ταξινόμηση πολλαπλών ετικετών είναι η περίπτωση κατά την οποία πολλές κατηγορίες μπορούν να εκχωρηθούν στο ίδιο έγγραφο.

Στην πρόσφατη βιβλιογραφία, η ταξινόμηση κειμένου έχει αποδειχθεί ότι δίνει καλά αποτελέσματα στην κατηγοριοποίηση πολλών πραγματικών δεδομένων που βασίζονται στον Ιστό, όπως, για παράδειγμα, ειδήσεις και κοινωνικά μέσα (Zubiaga et al., 2015) και ανάλυση συναισθημάτων (Melville et al., 2009). Οι ταξινομητές κειμένων δεν έχουν εφαρμοστεί ακόμη στην ταξινόμηση των θέσεων εργασίας στο Διαδίκτυο οι οποίες έχουν δημοσιευτεί σε αρκετές τοποθεσίες Web για την ανάλυση της αγοράς εργασίας στο Διαδίκτυο μιας γεωγραφικής περιοχής και το σύστημα είναι το πρώτο παράδειγμα προς αυτήν την κατεύθυνση.

Από την άλλη πλευρά, η εξαγωγή δεξιοτήτων από κενές θέσεις εργασίας στο Διαδίκτυο μπορεί να πλαισιωθεί στο πεδίο Εξαγωγή πληροφοριών (Chang et al., 2006) και Αναγνώριση

οντοτήτων (Sang & De Meulder, 2003). Το τελευταίο έχει εφαρμοστεί για την επίλυση πολυάριθμων ειδικών προβλημάτων τομέα στους τομείς της Εξόρυξης Πληροφοριών και της Ομαλοποίησης (Zhao et al., 2015). Αξίζει να αναφερθούν επίσης σχετικά έργα με την εξαγωγή δεξιοτήτων από βιογραφικά (Zhao et al., 2015).

Η εργασία που περιγράφεται από τους (Boseli et al., 2018) μπορεί να πλαισιωθεί στο πεδίο εξόρυξης ιστού, σύμφωνα με «Η εξόρυξη Ιστού στοχεύει στην ανακάλυψη χρήσιμων πληροφοριών ή γνώσεων από υπερσυνδέσμους Ιστού, περιεχόμενα σελίδας και αρχεία καταγραφής χρήσης». Ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στο (Han et al., 2011) για περισσότερες λεπτομέρειες σχετικά με την εξόρυξη δεδομένων και την εξόρυξη δεδομένων ιστού.

Όλες αυτές οι προσεγγίσεις είναι αρκετά σχετικές και αποτελεσματικές, και αποδεικνύουν επίσης τη σημασία του Διαδικτύου για πληροφορίες σχετικές με την αγορά εργασίας. Ωστόσο, διαφέρουν από την συγκεκριμένη προσέγγισή των (Boseli et al., 2018) σε δύο πτυχές. Πρώτον, στόχος είναι η ταξινόμηση των κενών θέσεων εργασίας σύμφωνα με ένα σύστημα ταξινόμησης στόχων για τη δημιουργία μιας (ανεξάρτητης από τη γλώσσα) γνωσιακής βάσης για σκοπούς ανάλυσης, και όχι το ταίριασμα των βιογραφικών με τις κενές θέσεις εργασίας. Επιπλέον, τα βιογραφικά γράφονται συνήθως με ακρίβεια από τους υποψηφίους, ενώ οι διαφημίσεις στο Διαδίκτυο γράφονται με λιγότερο ακριβή τρόπο και αυτό το ζήτημα ποιότητας μπορεί να έχει απρόβλεπτες επιπτώσεις στις παρεχόμενες πληροφορίες (Mezzanzanica et al., 2015). Δεύτερον, η προσέγγισή των (Boseli et al., 2018) στόχευε στην παραγωγή αναλύσεων που βασίζονται σε τυποποιημένες ταξινομίες (δηλαδή, ISCO και ESCO) για την υποστήριξη δραστηριοτήτων λήψης αποφάσεων βάσει γεγονότων πολλών ενδιαφερομένων.

Σχετικά με την αξιολόγηση της πρόβλεψης, παρότι η ακρίβεια του αλγορίθμου είναι απαραίτητη, ένα ακόμα σημαντικό ζήτημα είναι η ενδυνάμωση των σωστών τιμών του συστήματος πρόσληψης και η διασφάλιση της ορθότητας του αλγορίθμου. Αυτό σημαίνει πως για να γίνει δίκαιος οποιοσδήποτε αλγόριθμος που βασίζεται στη μηχανική μάθηση, πρέπει πρώτα να αποφύγει την προκατάληψη των δεδομένων κατάρτισης, όπως η σημαντική διαφορά στην αναλογία απασχόλησης γυναικών και ανδρών. Δυστυχώς στην πραγματική ζωή οι προκαταλήψεις φαίνεται δύσκολο να αποφευχθούν εντελώς.

Από τους αλγορίθμους που μελετήσαμε, ως καλύτερη επιλογή θεωρούμε τον αλγόριθμο KNN, ο οποίος είναι απλός, γρήγορος και με μεγάλη γενίκευση. Στην περίπτωση που τα δεδομένα είναι υψηλών διαστάσεων τότε προτείνεται η εφαρμογή ενός αλγορίθμου μείωσης διαστατικότητας. Στην περίπτωση που τα δείγματα είναι μεγάλου όγκου και αργούν οι υπολογισμοί, προτείνεται και η προσεγγιστική αναζήτηση, με ένα δεδομένο σφάλμα. Σε κάθε περίπτωση, ο KNN θεωρούμε ότι ταιριάζει στο πρόβλημα της ηλεκτρονικής πρόσληψης μιας και τα δεδομένα είναι χωρίς θόρυβο και χωρίς υψηλή διαστατικότητα.

Μελλοντικά, τέτοιου είδους σετ δεδομένων θα αυξάνονται μιας και θα ενσωματώνουν πολλές εταιρείες τέτοιου είδους συστήματα πρόσληψης. Αυτό σημαίνει ότι δημιουργείται η ανάγκη νέων αλγορίθμων κατηγοριοποίησης για την αποτελεσματική ανάλυση τους. Επίσης, σε περιπτώσεις όπου οι χρήστες είναι πολυάριθμοι (δηλαδή το σετ δεδομένων έχει πολλά δείγματα), προτείνονται οι αλγόριθμοι που στηρίζονται σε δέντρα, όπως ο decision tree, ειδικότερα αλγόριθμοι που σχετίζονται με ensemble τεχνικές με δέντρα, όπως ο Random Forest ο οποίος είναι μια bagging τεχνική η οποία υλοποιεί πολλά δέντρα αποφάσεων, με το κάθε δέντρο να ψάχνει σε ένα μέρος του σετ δεδομένων τόσο στα δείγματα όσο και στα χαρακτηριστικά του. Τέτοιου είδους αλγόριθμοι λειτουργούν αποδοτικά σε μεγάλου όγκου δεδομένα. Αυτό θα μπορούσε να είναι και μια μελλοντική κατεύθυνση της παρούσας πτυχιακής, δηλαδή η εφαρμογή ensemble αλγορίθμων σε δεδομένα ηλεκτρονικής πρόσληψης μεγάλου όγκου.

Βιβλιογραφία

- AlJadda, K., Korayem, M., Ortiz, C., Russell, C., Bernal, D., Payson, L., ... & Grainger, T. (2014). Augmenting recommendation systems using a model of semantically-related terms extracted from user behavior. *arXiv preprint arXiv:1409.2530*.
- Al-Otaibi, S. T., & Ykhlef, M. (2012). A survey of job recommender systems. *International Journal of Physical Sciences*, 7(29), 5127-5142.
- Alqahtani A., Alhakami H., Alsubait T., Baz A. (2021). A Survey of Text Matching Techniques. *Engineering, Technology & Applied Science Research*, vol 11. no. 1. pages. 6656-6661. doi: <https://doi.org/10.48084/etasr.3968>
- Amdouni, S., & abdessalem Karaa, W. B. (2010, May). Web-based recruiting. In *ACS/IEEE International Conference on Computer Systems and Applications-AICCSA 2010* (pp. 1-7). IEEE.
- Arthur, Michael B, & Rousseau, Denise M. (1996). A Career Lexicon for the 21st Century. *The Academy of Management Executive*, 10(4), 28–39.
- Arthur, Michael B, Claman, Priscilla H, & DeFillippi, Robert J. (1995). Intelligent Enterprise, Intelligent Careers. *The Academy of Management Executive*, 9(4), 7–20.
- Avinash S. Kapse, Assit. Prof & Head Computer Science and engineering,, Anuradha Engg. College Chikhli, Chikhli, India, 9730895352, (e-mail: askapse@gmail.com). 20/03/2012
- Barley, SR, & Kunda, G. (2004). Gurus, Hired Guns and Warm Bodies: Itinerant Experts in a Knowledge Economy.
- Barley, Stephen R, & Kunda, Gideon. (2006). Contracting: A New Form of Professional Practice. *The Academy of Management Perspectives*, 20(1), 45–66.
- Berry, M. W., & Fierro, R. D. (1996). Low-rank Orthogonal Decompositions for Information Retrieval Applications. *Numerical linear algebra with applications*, 3(4), 301-327.
- Bhat, V., Oates, T., Shanbhag, V., & Nicholas, C. (2004). Finding aliases on the web using latent semantic analysis. *Data & Knowledge Engineering*, 49(2), 129-143.
- Bian S., Xin Zhao W., Song Y., Zhang T., Wen J. (November 2019), pages 4810–4820 Domain Adaptation for Person-Job Fit with Transferable Deep Global Match Network. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. doi:10.18653/v1/D19-1487
- Chang, C. H., Kaye, M., Girgis, M. R., & Shaalan, K. F. (2006). A survey of web information extraction systems. *IEEE transactions on knowledge and data engineering*, 18(10), 1411-1428.

- Chiplunkar N.N., Fukao T., (2020). *Advances in Artificial Intelligence and Data Engineering*. Springer
- Chuan Qin, Hengshu Zhu, Tong Xu, Chen Zhu, Chao Ma, Enhong Chen, and Hui Xiong. 2020. An Enhanced Neural Network Approach to Person-Job Fit in Talent Recruitment. *ACM Trans. Inf. Syst.* 38, 2, Article 15 (March 2020), 33 pages. DOI:<https://doi.org/10.1145/3376927>
- Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines: Information retrieval in practice* (Vol. 520, pp. 131-141). Reading: Addison-Wesley.
- Davenport, John. (2006). UK Film Companies: Project-Based Organizations Lacking Entrepreneurship and Innovativeness? *Creativity and Innovation Management*, 15(3), 250–257.
- Davison, H. K., Maraist, C., & Bing, M. N. (2011). Friend or foe? The promise and pitfalls of using social networking sites for HR decisions. *Journal of Business and Psychology*, 26(2), 153-159.
- De Meo, P., Quattrone, G., Terracina, G., & Ursino, D. (2007). An XML-based multiagent system for supporting online recruitment services. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(4), 464-480.
- DeFillippi, Robert J, & Arthur, Michael B. (1994). The Boundaryless Career: A Competency-based Perspective. *Journal of Organizational Behavior*, 15(4), 307–324.
- DeFillippi, Robert J, & Arthur, Michael B. (1996). Boundaryless Contexts and Careers: A Competency-based Perspective. *The Boundaryless Career*, 116–131.
- Dex, Shirley, Willis, Janet, Paterson, Richard, & Sheppard, Elaine. (2000). Freelance Workers and Contract Uncertainty: The Effects of Contractual Changes in the Television Industry. *Work, Employment & Society*, 14(2), 283–305.
- Faliagka E., Ramantas K., Tsakalidis A., Tzimas G. (2012, May). Application of machine learning algorithms to an online recruitment system. In *Proc. International Conference on Internet and Web Applications and Services*.
- Faliagka, E., Ramantas, K., Tsakalidis, A., & Tzimas, G. (2012, May). Application of machine learning algorithms to an online recruitment system. In *Proc. International Conference on Internet and Web Applications and Services* (pp. 215-220).
- Feng, J., Xie, F., Hu, X., Li, P., Cao, J., & Wu, X. (2011, August). Keyword extraction based on sequential pattern mining. In *proceedings of the third international conference on internet multimedia computing and service* (pp. 34-38).
- Ghoshray, S. (2013). The Emerging Reality of Social Media: Erosion of Individual Privacy Through Cyber-vetting and Law's Inability to Catch Up, 12 J. Marshall Rev. Intell. Prop. L. 551 (2013). *UIC Review of Intellectual Property Law*, 12(3), 4.

- Golec, A., & Kahya, E. (2007). A fuzzy model for competency-based employee evaluation and selection. *Computers & Industrial Engineering*, 52(1), 143-161.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Machine learning basics. *Deep learning*, 1(7), 98-164.
- Goss, David, & Bridson, Joanna. (1998). Understanding Interim Management. *Human Resource Management Journal*, 8(4), 37-50.
- Guo, S., Alamudun, F., & Hammond, T. (2016). RésumMatcher: A personalized résumé-job matching system. *Expert Systems with Applications*, 60, 169-182.
- Guo, S., Alamudun, F., & Hammond, T. (2016). RésumMatcher: A personalized résumé-job matching system. *Expert Systems with Applications*, 60, 169-182
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hong, W., Zheng, S., & Wang, H. (2013, April). Dynamic user profile-based job recommender system. In *2013 8th International Conference on Computer Science & Education* (pp. 1499-1503). IEEE.
- Huang, C. K., Chien, L. F., & Oyang, Y. J. (2003). Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology*, 54(7), 638-649.
- Hughes, S. How we data-mine related tech skills (2015). URL: <http://insights.dice.com/2015/03/16/how-we-data-mine-relatedtech-skills/>(visited on 09/12/2017).
- Jain L., Harsha Vardhan M.A., Kathiresan G., Narayan A. (2021) Optimizing People Sourcing Through Semantic Matching of Job Description Documents and Candidate Profile Using Improved Topic Modelling Techniques. In: Chiplunkar N., Fukao T. (eds) *Advances in Artificial Intelligence and Data Engineering. Advances in Intelligent Systems and Computing*, vol 1133. Springer, Singapore. https://doi.org/10.1007/978-981-15-3514-7_66
- Janusz A., Stawicki S., Drewniak M., Ciebia K., Slezak D., Stencel K., (2018). How to Match Jobs and Candidates - A Recruitment Support System Based on Feature Engineering and Advanced Analytics. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. pages 503-514. Springer, Cham. University of Warsaw
- Joachims, T. (1996). *A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization*. Carnegie-mellon univ pittsburgh pa dept of computer science.
- Kapse, A. S., Patil, V. S., & Patil, N. V. (2012). E-recruitment. *International Journal of Engineering and Advanced Technology*, 1(4), 82-86.

- Kara, S., Alan, Ö., Sabuncu, O., Akpınar, S., Cicekli, N. K., & Alpaslan, F. N. (2012). An ontology-based retrieval system using semantic indexing. *Information Systems*, 37(4), 294-305.
- Kessler, R., Torres-Moreno, J. M., & El-Bèze, M. (2007, November). E-Gen: automatic job offer processing system for human resources. In *Mexican international conference on artificial intelligence* (pp. 985-995). Springer, Berlin, Heidelberg.
- Kino Y., Kuroki H., Machida T., Furuya N., Takano K., (2017). Text analysis for job matching quality improvement. *Procedia Computer Science* Volume 112, Pages 1523-1530. doi: <https://doi.org/10.1016/j.procs.2017.08.054>
- Kirkpatrick, Ian, & Hoque, Kim. (2006). A Retreat from Permanent Employment? Accounting for the Rise of Professional Agency Work in UK Public Services. *Work, Employment & Society*, 20(4), 649–666.
- Landthaler J., Glaser I., Matthes F., (2018). Towards Explainable Semantic Text Matching. Conference: JURIX: International Conference on Legal Knowledge and Information Systems. pages 200-204
- Lau, T., & Sure, Y. (2002). Introducing ontology-based skills management at a large insurance company. *Modellierung 2002, Modellierung in der Praxis–Modellierung für die Praxis*.
- Lee, I. (2011). Modeling the benefit of e-recruiting process integration. *Decision Support Systems*, 51(1), 230-239.
- Lian, Y., Peng, J. Y., Xie, H. M., & Feng, X. Y. (2011). Concepts similarity algorithm based on Google and KL distance. *Computer Engineering*, 37(19), 291-293.
- Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74, 12-32.
- Malherbe, E., & Aufaure, M. A. (2016, August). Bridge the terminology gap between recruiters and candidates: A multilingual skills base built from social media and linked data. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 583-590). IEEE.
- Malinowski, J., Keim, T., Wendt, O., & Weitzel, T. (2006, January). Matching people and jobs: A bilateral recommendation approach. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)* (Vol. 6, pp. 137c-137c). IEEE.
- Marler, Janet H, Woodard Barringer, Melissa, & Milkovich, George T. (2002). Boundaryless and Traditional Contingent Employees: Worlds Apart. *Journal of Organizational Behavior*, 23(4), 425–453.
- Melville, P., Gryc, W., & Lawrence, R. D. (2009, June). Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM*

- SIGKDD international conference on Knowledge discovery and data mining* (pp. 1275-1284).
- Menger, Pierre-Michel. (1999). Artistic Labor Markets and Careers. *Annual Review of Sociology*, 25(1), 541–574.
- Mezzanzanica, M., Boselli, R., Cesarini, M., & Mercorio, F. (2015). A model-based evaluation of data quality activities in KDD. *Information Processing & Management*, 51(2), 144-166.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Nikhil V. Patil, Student, Department of Computer Science & Engg. Anuradha Engg. College Chikhli, Chikhli, India, 8793718084, (e-mail: nikhilpatil17@gmail.com). 20/03/2012
- Paparrizos, I., Cambazoglu, B. B., & Gionis, A. (2011, October). Machine learned job recommendation. In *Proceedings of the fifth ACM Conference on Recommender Systems* (pp. 325-328).
- Posse, C. (2016). Cloud Jobs API: machine learning goes to work on job search and dISCOvery.
- Quinn, James Brian. (1992). The Intelligent Enterprise A New Paradigm. *The Executive*, 6(4), 48-63.
- Radevski, V., & Trichet, F. (2006, October). Ontology-based systems dedicated to human resources management: an application in e-recruitment. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 1068-1077). Springer, Berlin, Heidelberg.
- Research paper by Vinky Sharma on “Impact of E-recruitment on human resource supply chain management”www.jiit.ac.in/uploads/Synopsis%20-%20Vinky%20Sharma.pdf
- Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58.
- Rosoiu O., Popescu C. (2016) E-recruiting Platforms: Features that Influence the Efficiency of Online Recruitment Systems. *Informatica Economica*, vol. 20, no.2. doi: 10.12948/issn14531305/20.2.2016.05
- Sang, E. F., & De Meulder, F. (2003). Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050*.
- Saundry, Richard, Stuart, Mark, & Antcliff, Valerie. (2007). Broadcasting Discontent—Freelancers, Trade Unions and the Internet. *New Technology, Work and Employment*, 22(2), 178–191.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.

- Spirin, N., & Karahalios, K. (2013, May). Unsupervised approach to generate informative structured snippets for job search engines. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 203-204).
- Stanworth, Celia, & Stanworth, John. (1995). The Self-employed without Employees—Autonomous or Atypical? *Industrial Relations Journal*, 26(3), 221–229.
- Tench, Ralph, Fawkes, Johanna, & Palihawadana, Dayan. (2002). Freelancing: Issues and Trends for Public Relations Practice. *Journal of Communication Management*, 6(4), 311–322.
- Thun A., (2020). Matching Job Applicants to Free Text Job Ads Using Semantic Networks and Natural Language Inference. KTH
- Varshney, K., Wang, J., Mojsilović, A., Fang, D., & Bauer, J. (2013, June). Predicting and recommending skills in the social enterprise. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 7, No. 1).
- Vishal S. Patil, Student, Department of Computer Science & Engg. Anuradha Engg. College Chikhli, Chikhli, India, 8275331855., (e-mail: vpatil1180@gmail.com). 20/03/2012
- Yih, W. T., & Meek, C. (2007, July). Improving similarity measures for short segments of text. In *AAAI* (Vol. 7, No. 7, pp. 1489-1494).
- Zhang, K., Xu, H., Tang, J., & Li, J. (2006, June). Keyword extraction using support vector machine. In *international conference on web-age information management* (pp. 85-96). Springer, Berlin, Heidelberg.
- Zhang, L., Fei, W., & Wang, L. (2015). Pj matching model of knowledge workers. *Procedia computer science*, 60, 1128-1137.
- Zhao, M., Javed, F., Jacob, F., & McNair, M. (2015, March). SKILL: A system for skill identification and normalization. In *Twenty-Seventh IAAI Conference*.
- Zubiaga, A., Spina, D., Martínez, R., & Fresno, V. (2015). Real-time classification of twitter trends. *Journal of the Association for Information Science and Technology*, 66(3), 462-473.