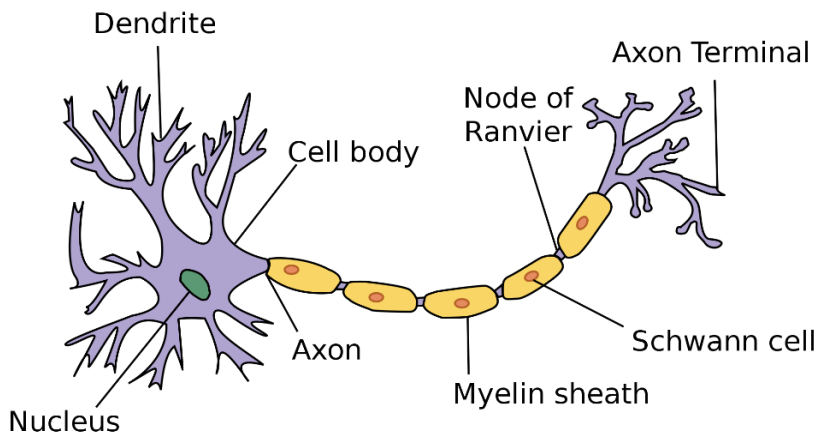


ΘΑΛΗΣ: Ένα Βαθύ Νευρωνικό Δίκτυο (Deep Neural Network - DNN) που αναπτύχθηκε εξ ολοκλήρου στον ΟΒΙ

Ο «**ΘΑΛΗΣ¹**» είναι ένα **Βαθύ Νευρωνικό Δίκτυο (Deep Neural Network - DNN)** το οποίο αναπτύχθηκε εξ ολοκλήρου στον ΟΒΙ και συγκεκριμένα στην Διεύθυνση Ελέγχου Τίτλων (ΔΕΤ), από επιστημονικό προσωπικό με γνώσεις ή σπουδές στο χώρο της **Τεχνητής Νοημοσύνης (Artificial Intelligence)**.

Σε πολύ γενικές γραμμές ένα **Νευρωνικό Δίκτυο** προσπαθεί να μιμηθεί κατά το δυνατόν τη λειτουργία του ανθρώπινου εγκεφάλου, ο οποίος συντίθεται από δισεκατομμύρια **Νευρώνες** διασυνδεδεμένους μεταξύ τους σε ένα πυκνό δίκτυο κατά μήκος δεσμών (ενώσεων) που αποκαλούνται **Συνάψεις**.

Όπως απεικονίζεται στο παρακάτω σχήμα της Wikipedia, κάθε νευρώνας δέχεται πληροφορίες (ηλεκτρικά σήματα) μέσω των Δενδριτών (Dendrite) από άλλους νευρώνες με τους οποίους είναι συνδεδεμένος, διεξάγει μια ορισμένη επεξεργασία, και εφόσον το αποτέλεσμα ξεπερνά ένα ορισμένο κατώφλι πυροδοτεί μέσω του Νευροάξονα ή Άξονα (Axon) ένα ηλεκτρικό παλμό σε μια άλλη ομάδα νευρώνων με τους οποίους είναι συνδεδεμένος.



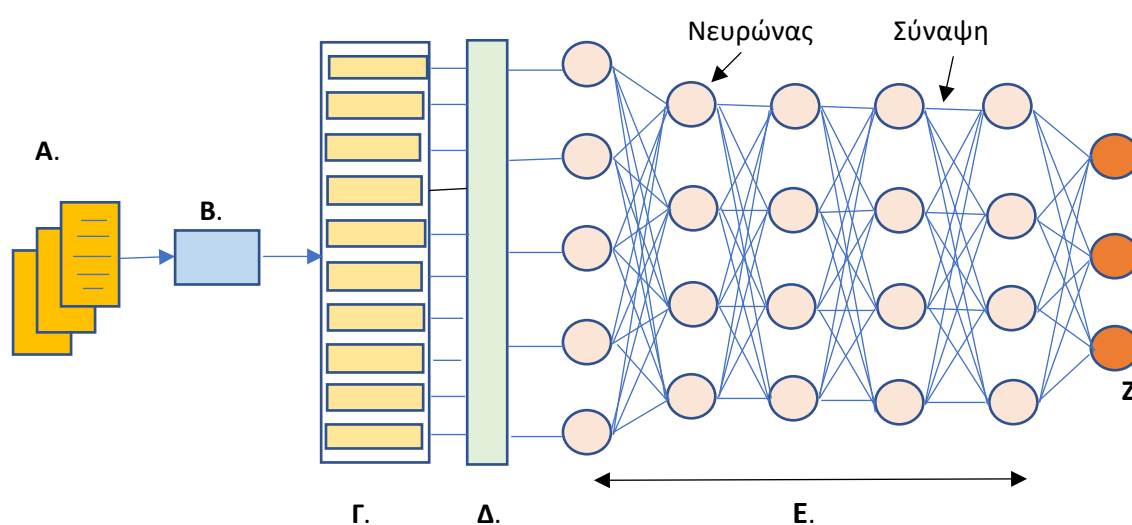
Στην παραπάνω λογική βασίζεται η αρχιτεκτονική και η λειτουργία του ΘΑΛΗ.

Μετά από μακροχρόνιες δοκιμές προσαρμοσμένες στα τεχνικά/επιστημονικά κείμενα των Διπλωμάτων Ευρεσιτεχνίας, καταλήξαμε σε μία αρχιτεκτονική η οποία (χωρίς να αναφερθούν τεχνικές λεπτομέρειες) περιλαμβάνει **483** συνολικά **Νευρώνες (Neurons ή Units)** κατανομημένους σε **5 Κρυφά Στρώματα ή Επίπεδα (Hidden Layers)**, με συνολικά **31** Εκατομμύρια **Εκπαιδευσιμες Παραμέτρους**.

Μια εντελώς απλοποιημένη σχηματική απεικόνιση της αρχιτεκτονικής του ΘΑΛΗ δίνεται στο παρακάτω σχήμα όπου κάθε κύκλος αντιπροσωπεύει ένα **Νευρώνα** ενώ οι γραμμές μεταξύ τους αντιπροσωπεύουν τις **Συνάψεις**. Οι νευρώνες βρίσκονται κατανομημένοι σε πέντε κατακόρυφα επίπεδα που αντιπροσωπεύουν τα Κρυφά

Στρώματα (Ε). Πρόκειται για ένα **Πλήρως Διασυνδεδεμένο Νευρωνικό Δίκτυο (Fully Connected Neural Network)** όπου ο κάθε νευρώνας ενός στρώματος συνδέεται και παίρνει πληροφορίες από όλους τους νευρώνες του προηγούμενου στρώματος τις οποίες επεξεργάζεται, και μέσω μιας **Συνάρτησης Ενεργοποίησης (Activation Function)** προωθεί το αποτέλεσμα της επεξεργασίας σε όλους τους νευρώνες του επόμενου στρώματος (επιπέδου)

Στο δεξί άκρο του σχήματος, το οποίο αντιστοιχεί στο Στρώμα Εξόδου (Ζ), κάθε νευρώνας (κύκλος) αντιπροσωπεύει ένα από τα Τμήματα Εξεταστών του Οργανισμού στο οποίο θα δρομολογηθεί/ανατεθεί το εισερχόμενο από αριστερά έγγραφο (Α).



Α. Τεχνικά κείμενα Διπλωμάτων Ευρεσιτεχνίας

Β. Επεξεργασία της Φυσικής Γλώσσας των κειμένων (Natural Language Processing – NLP)

Γ. Στρώμα Διανουματοποίησης Λέξεων (Embeddings Layer)

Δ. Flatten Layer

Ε. Πέντε Κρυφά Στρώματα με 483 συνολικά νευρώνες (Hidden Layers)

Ζ. Τμήματα Εξεταστών (Output Layer)

Ο ΘΑΛΗΣ, ολοκληρώθηκε και τέθηκε για πρώτη φορά σε δοκιμαστική εφαρμογή στον ΟΒΙ τον Απρίλιο του 2022.

Στόχος της ανάπτυξης του ΘΑΛΗ ήταν η αυτοματοποίηση της διαδικασίας **Αρχικής Μελέτης** και **Ταξινόμησης** των νεοεισερχόμενων αιτήσεων για Δίπλωμα Ευρεσιτεχνίας (ΔΕ) ανάλογα με το τεχνικό/επιστημονικό πεδίο στο οποίο αναφέρονται, και η εν συνεχεία **δρομολόγηση/ανάθεση** στα αρμόδια εξεταστικά τμήματα του ΟΒΙ.

Μέχρι τον Απρίλη του 2022 η συγκεκριμένη διαδικασία γινόταν από 2μελή ή 3μελή ομάδα εργασίας του επιστημονικού προσωπικού της ΔΕΤ, η οποία σε μηνιαία βάση συγκέντρωνε και μελετούσε μία προς μία όλες τις νέες αιτήσεις ΔΕ και τις ανέθετε σε εξεταστή ειδικότητας, ανάλογα με το τεχνικό/επιστημονικό πεδίο στο οποίο αναφερόταν. Η όλη διαδικασία απασχολούσε την ομάδα εργασίας για αρκετές ημέρες μέχρι την ολοκλήρωσή της.

Σήμερα, ο ΘΑΛΗΣ έχει αντικαταστήσει πλήρως την ως άνω ομάδα εργασίας, ενώ η όλη διαδικασία της μελέτης και ανάθεσης των νέων αιτήσεων ΔΕ διαρκεί 2-3 λεπτά.

Για την εκπαίδευσή του ΘΑΛΗ χρησιμοποιήθηκε η διαδικασία της **Επιτηρούμενης Μηχανικής Μάθησης (Supervised Machine Learning)** στα πλαίσια της οποίας δημιουργήσαμε και χρησιμοποιούμε *Βάση Δεδομένων* περιλαμβάνουσα τα κείμενα των Περιλήψεων ή/και Αξιώσεων 12.000 περίπου αιτήσεων ΔΕ και ΠΥΧ από το 2000 μέχρι σήμερα (όλα τα ΔΕ που Δημοσιεύτηκαν την τελευταία 20ετία).

Επιπλέον, για τις ανάγκες Επεξεργασίας της Γλώσσας των τεχνικών κειμένων (NLP) δημιουργήσαμε **3 Ελληνικά Λεξικά ΛΗΜΜΑΤΩΝ** με περίπου 10.000 τεχνικούς όρους στους τρεις βασικούς τεχνολογικούς/επιστημονικούς τομείς της *Μηχανολογίας/Ενέργειας, Ηλεκτρονικής/Πληροφορικής και Χημείας/Βιοτεχνολογίας*, καθώς και ένα *Ελληνικό Λεξικό με Λέξεις Χαμηλής Αξίας (Stopwords)*, ειδικό για τα κείμενα των ΔΕ. Τα Λεξικά αυτά αποτελούν σημαντικό κομμάτι της διαδικασίας Μηχανικής Μάθησης και αυξάνουν κατά πολύ την ακρίβεια των αποτελεσμάτων.

Με τα ανωτέρω δεδομένα η ακρίβεια του ΘΑΛΗ στη *Μελέτη/Ταξινόμηση/Ανάθεση* των αιτήσεων ΔΕ αγγίζει το **96-97%** στην τρέχουσα έκδοση, κάτι το οποίο είναι παραπάνω από ικανοποιητικό.

Είναι σημαντικό να αναφερθεί ότι κατά τη διάρκεια της σύλληψης, σχεδίασης και ανάπτυξης του ΘΑΛΗ από την ομάδα της ΔΕΤ έγινε φανερό μία άλλη δυνατότητά, την οποία δεν υποπτευόμασταν στην αρχική σύλληψη του συστήματος. Πρόκειται για τη δυνατότητα του ΘΑΛΗ να πραγματοποιεί **Νοηματική Έρευνα (Semantic Search)** καθώς είναι ικανός να αποδίδει νοηματικό περιεχόμενο σε κάθε τεχνικό ή επιστημονικό κείμενο.

Η κατανόηση αυτής της δεύτερης δυνατότητας, μάς οδήγησε στην ανάπτυξη μιας **δεύτερης έκδοσης του ΘΑΛΗ** όπου ο χρήστης μπορεί να εισάγει το τεχνικό/επιστημονικό κείμενο για το οποίο αναζητά προϋπάρχουσα σχετική τεχνολογία σε διπλώματα ευρεσιτεχνίας που έχουν κατατεθεί στην Ελλάδα την τελευταία 20ετία και αυτόματα να έχει τα αποτελέσματα.

Αυτό επιτυγχάνεται χωρίς να είναι πλέον αναγκαία η γνώση και χρήση *Λέξεων Κλειδιών (keywords), Συνωνύμων και Boolean Operators*.

Στη δεύτερη αυτή έκδοση του ΘΑΛΗ, η οποία χρησιμοποιείται αποκλειστικά και μόνο για την έρευνα, χρησιμοποιούμε διαφορετική αρχιτεκτονική και **Μη Επιτηρούμενη Μηχανική Μάθηση (Unsupervised Machine Learning)**.

Η συγκεκριμένη έκδοση, είναι επί του παρόντος σε δοκιμαστική φάση με πολύ καλά αποτελέσματα. Σε φάση κανονικής εφαρμογής θα μπορεί να χρησιμοποιηθεί τόσο από τους εξεταστές του ΟΒΙ κατά τη διαδικασία σύνταξης των Εκθέσεων Έρευνας, αλλά και να παρασχεθεί σαν **Νέα Υπηρεσία** προς το κοινό μέσα από το site του Οργανισμού, όπου ο εξωτερικός χρήστης θα μπορεί απλά να εισάγει το τεχνικό/επιστημονικό κείμενο για το οποίο αναζητά σχετική τεχνολογία μέσα από τα ΔΕ που έχουν κατατεθεί, δημοσιευτεί και προστατεύονται στην Ελλάδα, και να έχει αυτόματα τα αποτελέσματα.

Η ομάδα που ανέπτυξε το Βαθύ Νευρωνικό Δίκτυο αποτελείται από τους:

- Στράτο Κουτίβα, Μηχανολόγο Μηχανικό, Διευθυντή Ελέγχου Τίτλων (ΔΕΤ)
- Δημήτρη Σταφυλά, Μηχανικό Η/Υ, Προϊστάμενο Τμήματος Εφευρέσεων Πληροφορικής & Επικοινωνιών, Εξεταστή, ΔΕΤ
- Ιωάννη Ποντίκη, Μηχανολόγο Μηχανικό, Συνεργάτη ΔΕΤ
- Αλέξια Φιλίππου, Φυσικό, Συνεργάτη ΔΕΤ)

1. *Ο ΘΑΛΗΣ ο Μιλήσιος, αρχαίος Έλληνας φιλόσοφος, μαθηματικός, φυσικός, αστρονόμος και μηχανικός, θεωρείται από πολλούς ο πρώτος Έλληνας φιλόσοφος με μεγάλη επιρροή στους μεταγενέστερους φιλοσόφους. Είναι ο λόγος που επιλέχθηκε το όνομά του για το σύστημα που αναπτύξαμε, καθώς πρόκειται για την πρώτη προσπάθεια χρήσης της Μηχανικής Μάθησης και Τεχνητής Νοημοσύνης στις διαδικασίες του Οργανισμού Βιομηχανικής Ιδιοκτησίας.*