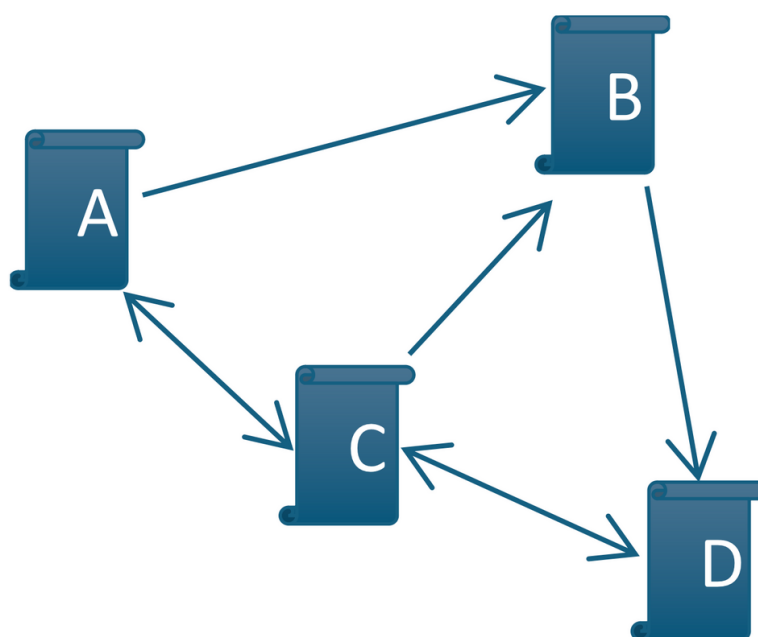


Personalized PageRank, RankBrain

Καθώς ο όγκος των ιστοσελίδων γινόταν ακόμα μεγαλύτερος αλλά και η ανάγκη για ακόμα γρηγορότερη απόκριση και πιο συγκεκριμένα αποτελέσματα αναζήτησης ο αλγόριθμος PageRank εξελίχθηκε ώστε να περιλαμβάνει τις προσωπικές προτιμήσεις του χρήστη.

Ο αλγόριθμος εξατομικευμένης ταξινόμησης ιστοσελίδων (**personalized PageRank**) βασίζεται στην παρατήρηση ότι: όταν ένας χρήστης πραγματοποιεί μια αναζήτηση, τα αποτελέσματα της αναζήτησης δεν βασίζονται μόνο στη συνάφεια κάθε ιστοσελίδας με τον όρο αναζήτησης, αλλά και στις ιστοσελίδες τις οποίες έχει επισκεφθεί κατά το παρελθόν ο χρήστης (ή κάποιος άλλος που χρησιμοποιεί το ίδιο πρόγραμμα περιήγησης) μέσω προηγούμενων αποτελεσμάτων αναζήτησης. Όλες οι αναζητήσεις οι οποίες έχουν πραγματοποιηθεί μέσω μιας εφαρμογής φυλλομετρητή, σε ένα συγκεκριμένο μηχάνημα, η μηχανή αναζήτησης τις έχει αποθηκεύσει. Οπότε όταν ο χρήστης αναζητά έναν όρο τα αποτελέσματα είναι ταξινομημένα με βάση πληροφορίες που γνωρίζει η μηχανή αναζήτησης για τον συγκεκριμένο φυλλομετρητή και το συγκεκριμένο μηχάνημα. Τα αποτελέσματα τα οποία προτείνονται στο χρήστη είναι εξατομικευμένα και στην πλειοψηφία των περιπτώσεων 'πετυχημένα'. Ο χρήστης βρίσκει στις πρώτες θέσεις των ιστοσελίδων των αποτελεσμάτων αυτό που ψάχνει. Από μαθηματικής άποψης ο personalized PageRank λειτουργεί όπως και ο PageRank με τη διαφορά ότι οι σχέσεις μεταξύ των σελίδων δεν έχουν την ίδια αξία. Έτσι στο παράδειγμα της εικόνας " Αλγόριθμος PageRank παράδειγμα» για να βρούμε το βαθμό της σελίδας B δεν θα πραγματοποιήσουμε τους υπολογισμούς λαμβάνοντας τις σελίδες A και C ισάξιας, αλλά θα δώσουμε παραπάνω 'αξία' στη σελίδα π.χ. A. Η επιπλέον αξία προκύπτει από τις πληροφορίες που διαθέτει η μηχανή αναζήτησης για το χρήστη. Οπότε η σελίδα A θα εμφανιστεί στις πρώτες θέσεις των αποτελεσμάτων.



Αλγόριθμος PageRank παράδειγμα

Οι πληροφορίες οι οποίες συλλέγονται από τις μηχανές αναζήτησης περιλαμβάνουν:

- **Δεδομένα τοποθεσίας:**

Χρησιμοποιώντας δεδομένα από τη λειτουργία GPS του κινητού τηλεφώνου είτε από την IP διεύθυνση του υπολογιστή γίνεται γνωστή η γεωγραφική τοποθεσία του χρήστη, οπότε τα αποτελέσματα της αναζήτησης περιλαμβάνουν πρώτα ιστοσελίδες οι οποίες είναι 'κοντά' σε αυτή την τοποθεσία.

- **Ιστορικό αναζήτησης και ιστοσελίδων:**

Περιλαμβάνει τους όρους της αναζήτησης, καθώς και τις ιστοσελίδες που έχουν επισκεφτεί. Εφόσον ο χρήστης έχει συνδεθεί/αυθεντικοποιηθεί (διαδικασία login) στην εφαρμογή μιας μηχανής αναζήτησης, τότε διατηρείται στο προφίλ το ιστορικό με όλες τις λέξεις κλειδιά/φράσεις που έχει αναζητήσει. Εάν ο χρήστης δεν αυθεντικοποιηθεί (login) τότε η μηχανή αναζήτησης αποθηκεύει στο φυλλομετρητή του ένα μοναδικό ανώνυμο κωδικό (αλφαριθμητικό) (cookie) και αποθηκεύει το ιστορικό αναζήτησης του με βάση αυτόν τον αριθμό. Τα αποτελέσματα μελλοντικών αναζητήσεων προσαρμόζονται σε αυτό το ιστορικό δίνοντας προτεραιότητα σε ιστοσελίδες τις οποίες έχει ξαναεπισκευτεί είτε είναι συναφείς προς αυτές.

Με την πάροδο των χρόνων και έπειτα από μελέτες έγινε φανερό ότι ο καλύτερος παράγοντας ώστε να αποφασιστεί ποια αποτελέσματα είναι καταλληλότερα/συναφέστερα προς τον χρήστη δεν είναι οι προσωπικές πληροφορίες του, αλλά οι ίδιοι οι όροι της αναζήτησης. Αναπτύχθηκαν αλγόριθμοι όπως ο **RankBrain**. Ο **RankBrain** είναι ένας αλγόριθμος μηχανών αναζήτησης ο οποίος βασίζεται στη μηχανική μάθηση (Machine learning - ML). Η μηχανική μάθηση (ML) είναι πεδίο μελέτης στο χώρο της τεχνητής νοημοσύνης (artificial intelligence - AI) και ασχολείται με την ανάπτυξη και τη μελέτη στατιστικών αλγορίθμων οι οποίοι μπορούν να μάθουν από τα δεδομένα και να τα συνδυάσουν ώστε να εκτελέσουν εργασίες χωρίς ρητές οδηγίες. Έτσι εάν ο χρήστης χρησιμοποιήσει λέξεις ή φράσεις για τις οποίες ο αλγόριθμος δεν έχει προγενέστερη γνώση τότε προτείνει άλλες με παρόμοια σημασία και παρέχει τα ανάλογα αποτελέσματα. Στη συνέχεια μελετάει τη συμπεριφορά του χρήστη. Εάν ο χρήστης είναι ικανοποιημένος από τα αποτελέσματα τότε αυξάνει τη 'σπουδαιότητα' τους, οπότε σε μελλοντικές αναζητήσεις θα εμφανιστούν ψηλότερα στη λίστα. Εάν ο χρήστης δεν είναι ικανοποιημένος τότε στο μέλλον δεν θα του προτείνει αυτές τις ιστοσελίδες ως αποτέλεσμα.

Παράγοντες όπως η διάρκεια της επίσκεψης, ο αριθμός των κλικ στην ιστοσελίδα και το ποσοστό των επισκεπτών που εγκατέλειψαν την ιστοσελίδα χωρίς να επιλέξουν κάτι, αποτελούν δείκτες της ικανοποίησης (ή μη) των χρηστών από τα αποτελέσματα αναζήτησης.