

Ο Μύθος των Υπερδεσμών

Ένα

Το Πρόβλημα της Τεχνητής Νοημοσύνης (ΤΝ), όπως ονομάζεται – το πρόβλημα να κάνουμε τις μηχανές να συμπεριφέρονται αρκετά κοντά με το πώς οι άνθρωποι συμπεριφέρονται ευφυώς – ... δεν έχει ακόμα λυθεί. Επιπλέον, δεν υπάρχει τίποτα στον ορίζοντα, που να λέει ότι βλέπουμε κάποιο φως προς την κατεύθυνση αυτή. Οι όροι, όπως ‘τεχνητή νοημοσύνη’, ‘ευφυείς πράκτορες’, ‘υπηρέτες’ – όλες αυτές οι μυθοποιημένες λέξεις, που διαβάζουμε στον τύπο – είναι όλες επανατοποθετήσεις της περιπλοκής και του προβλήματος, μέσα στο οποίο βρισκόμαστε.

Θα μας άρεσε να έχουμε μια μηχανή, η οποία να μπορεί να πηγαίνει να ψάχνει στον Ιστό και στα αγαπημένα μας μαγαζιά, γνωρίζοντας τις προτιμήσεις μας και γνωρίζοντας τι εννοούμε, όταν λέμε κάτι. Αλλά, απλώς, δεν έχουμε κάτι σε αυτό το επίπεδο.

Michael Dertouzos, Εργαστήριο Επιστήμης Υπολογιστών, MIT¹

Ο Παγκόσμιος Ιστός είναι τεράστιος και αναπτύσσεται γρήγορα. Σε μια πρόσφατη μέτρηση, είχε πάνω από ένα δισεκατομμύριο σελίδες και συνεχίζει να μεγαλώνει με ρυθμό τουλάχιστον ενός εκατομμυρίου σελίδων τη μέρα.² (Είναι χαρακτηριστικό του Ιστού ότι αυτές οι στατιστικές, καθώς τις διαβάζετε, είναι ήδη κατά πολύ ξεπερασμένες.) Υπάρχει ένα καταπληκτικό ποσό χρήσιμων πληροφοριών στον Ιστό, αλλά όμως αυτές γίνονται όλο και πιο δύσκολο να βρεθούν. Το πρόβλημα προέρχεται από τον τρόπο, με τον οποίο είναι οργανωμένες οι πληροφορίες (ή, καλύτερα, ανοργανώτες). Με τον τρόπο λειτουργίας του Ιστού, κάθε στοιχείο αυτού του συμφερτού των πληροφοριών είναι συνδεδεμένο με πολλά άλλα στοιχεία μέσω των υπερδεσμών. Τέτοιοι υπερδεσμοί μπορούν να διασυνδέουν κάθε στοιχείο πληροφορίας με οποιοδήποτε άλλο στοιχείο, για οποιοδήποτε λόγο τυχαίνει να έχει ο οποιοσδήποτε δημιουργός των υπερδεσμών. Καμία αρχή ή κανένα προσυμφωνημένο σύστημα ταξινόμησης δεν περιορίζει τους υπερδεσμούς, όταν κάποιος τους δημιουργεί.³

Οι υπερδεσμοί δεν εισήχθησαν, επειδή ήταν περισσότερο χρήσιμοι για την ανάκτηση των πληροφοριών, σε σχέση με το παλαιότερο ιεραρχικό σύστημα τοποθέτησης των πληροφοριών. Μάλλον, αποτελούν το φυσικό τρόπο εκμετάλλευσης της ταχύτητας και της επεξεργαστικής ισχύος των υπολογιστών, για να επιτευχθεί η αλληλοσυσχέτιση μιας τεράστιας ποσότητας πληροφοριών, χωρίς να χρειάζεται να τις έχει ήδη κατανοήσει κανείς ή χωρίς να χρειάζεται να επιβάλλει κανείς κάποια αυταρχική ή και γενικότερα αποδεκτή δομή στις πληροφορίες αυτές. Αλλά, όταν τα πάντα μπορούν να συνδεθούν με όλα τα άλλα, δίχως να λαμβάνεται υπόψη ο σκοπός ή το νόημα, το μέγεθος του Ιστού, η τυχαιότητα και η αυθαιρεσία των υπερδεσμών καθιστούν εξαιρετικά δύσκολο το έργο της αναζήτησης των συγκεκριμένων πληροφοριών.

Το πρόβλημα της ανάκτησης των σχετικών πληροφοριών από το πεδίο των υπερσυνδεδεμένων στοιχείων είναι τόσο νέο όσο και το Internet. Ο παραδοσιακός τρόπος ταξινόμησης των πληροφοριών βασίζεται στη δουλειά κάποιου

συγκεκριμένου ανθρώπου – π.χ., κάποιου ζωολόγου ή βιβλιοθηκάρου ή φιλόσοφου ή κ.λπ. – ο οποίος εκπονεί ένα συγκεκριμένο σχέδιο ταξινόμησης, σύμφωνα με τα νοήματα των όρων, που εμπλέκονται, και τα ενδιαφέροντα των ατόμων, τα οποία στη συνέχεια πρόκειται να χρησιμοποιήσουν την ταξινόμηση αυτή.⁴ Έτσι, οι νέες πληροφορίες εισάγονται σε ένα τέτοιο σχήμα ταξινόμησης, με βάση το πώς αντιλαμβάνεται ο δημιουργός της ταξινόμησης το νόημα των κατηγοριών του σχήματος της ταξινόμησης και το τι σημαίνει για αυτόν κάθε νέα εισαγόμενη πληροφορία. Όταν στη συνέχεια θέλει κανείς να χρησιμοποιήσει την πληροφορία, πρέπει να βασιστεί σε αυτούς, που αρχικά έχουν γράψει και εφαρμόσει το σχήμα της ταξινόμησης της και το έχουν οργανώσει με βάση το νόημα για αυτούς της πληροφορίας. Δηλαδή, στο παραδοσιακό μοντέλο, προϋπάρχει ένα σχήμα ταξινόμησης, μέσα στο οποίο μπορούν οι μετέπειτα χρήστες να αναζητήσουν εύκολα την πληροφορία, που είναι σχετική με τις επιδιώξεις τους.

Ο David Blair, Καθηγητής Υπολογιστικών και Πληροφοριακών Συστημάτων στο Πανεπιστήμιο του Michigan,⁵ επισημαίνει ότι τα περισσότερα ‘παραδοσιακά’ σχήματα ταξινόμησης ήταν ρητά ή εν δυνάμει συνδεδεμένα με μια συγκεκριμένη ‘πρακτική’ δραστηριότητα κάποιου είδους. Οι επιστήμες ζωής είναι το προφανές παράδειγμα, αλλά υπάρχουν και άλλες λιγότερο τυπικές πρακτικές δραστηριότητες, οι οποίες αποτελούν τη βάση τέτοιων σχημάτων ταξινόμησης, όπως είναι η πολύ παλιά πρακτική της αγροτικής τέχνης, όπου ο αγρότης πρέπει να είναι σε θέση να αναγνωρίζει διάφορα είδη φυτών, ζώων, ζιζανίων, ασθενειών, καιρικών συνθηκών, εποχών, κ.λπ. Ενώ κάποιος από τους δεσμούς στον Ιστό μπορεί να είναι μεταξύ Δικτυακών τόπων, που αφορούν συγκεκριμένες πρακτικές, οι περισσότεροι δεν έχουν να κάνουν τίποτε με καμία πρακτική δραστηριότητα. Χωρίς τις απαιτήσεις μιας συγκεκριμένης πρακτικής δραστηριότητας, η οποία μπορεί να καθορίζει και να περιορίζει τι θα πρέπει να συνδέεται με τι, ο πολλαπλασιασμός των υπερδεσμών γίνεται ανεξέλεγκτος.⁶

Από τον Αριστοτέλη έχουμε συνηθίσει να οργανώνουμε την πληροφορία μέσα σε μια ιεραρχία όλο και πιο ευρέων κλάσεων ή κατηγοριών, με την καθεμιά από αυτές να περιέχει τις πιο ειδικότερες κλάσεις ή κατηγορίες από κάτω της. Έτσι, κατεβαίνουμε από τα όντα, στα ζωντανά όντα, στα ζώα, στα θηλαστικά, στα σκυλιά, στα κόλλει, στη Λάσι. Όταν η πληροφορία είναι οργανωμένη σε μια τέτοια ιεραρχική βάση δεδομένων, ο χρήστης μπορεί να ακολουθήσει τους δεσμούς με βάση τη σημασία τους. Επιπλέον όμως, ο χρήστης είναι υποχρεωμένος να έχει αφοσιωθεί ή να έχει δεσμευτεί, για κάποιο δικό του σκοπό, σε μια συγκεκριμένη κλάση ή κατηγορία της πληροφορίας, προτού γνωρίσει τα ειδικότερα δεδομένα, που emπίπτουν σε αυτήν. Για παράδειγμα, προϋποτίθεται ότι έχει κανείς ένα συγκεκριμένο ενδιαφέρον για τα ζώα, προτού μπορέσει να ανακαλύψει κάτι που θέλει να μάθει για τις χελώνες, για παράδειγμα. Και έχοντας αυτό το συγκεκριμένο ενδιαφέρον, δηλαδή, έχοντας μια συγκεκριμένη ‘δέσμευση’ μέσα στη γραμμή των ζώων της βάσης δεδομένων, δεν μπορεί στη συνέχεια να εξετάζει όλα αδιακρίτως τα δεδομένα του προβλήματος επ’ άπειρο, χωρίς να μεταβάλλει τη δέσμευση εκείνη, που αρχικά έχει ακολουθήσει.

Όταν η πληροφορία είναι οργανωμένη μέσω υπερδεσμών, ωστόσο, όπως είναι στον Ιστό, στη θέση της συγκεκριμένης σχέσης, που έχει μια κλάση ή κατηγορία με τα μέλη της, η οργανωτική αρχή, που τώρα μπαίνει, είναι μόνο η απαίτηση μιας γενικής διασυνδετικότητας μεταξύ των στοιχείων της βάσης δεδομένων. Δεν υπάρχουν ιεραρχίες: τα πάντα είναι συνδεδεμένα με όλα τα άλλα, σε ένα μόνο – και αυτό

εντελώς γενικό – επίπεδο. Έτσι, οι υπερδεσμοί επιτρέπουν στο χρήστη να κινείται άμεσα από μια καταχώρηση της βάσης δεδομένων σε οποιαδήποτε άλλη, αρκεί οι δυο καταχωρήσεις να συσχετίζονται με τουλάχιστον κάποιον τρόπο, έστω και αν ο τρόπος αυτός είναι με πολύ μικρή σημασία από ουσιαστικής άποψης. Σχεδόν ολόκληρος ο Ιστός βρίσκεται μόνο μερικούς δεσμούς (ή μερικά κλικ) μακριά από οποιαδήποτε δικτυακή σελίδα. Σε μια υπερσυνδεδεμένη βάση δεδομένων, ο χρήστης ενθαρρύνεται να διασχίσει ένα τεράστιο δίκτυο πληροφορίας, όπου όλα είναι εξίσου προσβάσιμα και τίποτα δεν είναι προνομιούχο. Έτσι, για παράδειγμα, ανάμεσα στους δικτυακούς τόπους, που περιέχουν πληροφορίες για τις χελώνες, τους οποίους βρίσκει μια τυπική μηχανή αναζήτησης, μπορεί κανείς να επισκεφθεί έναν τόπο, που λέγεται, ας πούμε, ‘χελώνες και λαγοί,’ και από εκεί να μεταφερθεί στη στιγμή σε μια σελίδα για το παράδοξο του Ζήνωνα.

Μπορεί κανείς να κατανοήσει την αντίθεση του παλιού με το νέο τρόπο οργάνωσης και ανάκτησης πληροφοριών και το θέλημα του καθενός από αυτούς με μια σύγκριση της παλιάς κουλτούρας της βιβλιοθήκης με το νέο είδος ‘βιβλιοθηκών,’ που δημιουργήθηκαν με τους υπερδεσμούς. Οι αντιθέσεις δείχνουν το μετασχηματισμό μιας σημασιολογικής δόμησης των πληροφοριών, η οποία στηρίζεται στο νόημά τους, προς μια τυπική συντακτική δόμηση, στην οποία το νόημα δεν παίζει κανένα ρόλο. Ο Πίνακας 1 δείχνει μια συστηματικοποίηση κάποιων από τις αντιθέσεις αυτές.

ΠΑΛΙΑ ΚΟΥΛΤΟΥΡΑ ΒΙΒΛΙΟΘΗΚΗΣ	ΚΟΥΛΤΟΥΡΑ ΤΩΝ ΥΠΕΡΣΥΝΔΕΣΕΩΝ
Ταξινόμηση	Ποικιλότητα
A. Σταθερή	A. Εύκαμπτη
B. Ιεραρχικά οργανωμένα	B. Οριζόντια
Γ. Καθορίζεται από συγκεκριμένα ενδιαφέροντα/δεσμεύσεις	Γ. Επιτρέπει όλες τις συνδέσεις
Σκόπιμη επιλογή	Πρόσβαση σε όλα
A. Ποιότητα	A. Περιεκτικότητα
B. Αυθεντικότητα	B. Διαθεσιμότητα
Γ. Απομάκρυνση παλαιού υλικού	Γ. Τα πάντα σώζονται
Μόνιμες συλλογές	Δυναμικές συλλογές
A. Σταθερότητα κειμένου	A. Διακειμενική εξέλιξη
B. Φυλλομέτρηση για κάποιο σκοπό	B. Σερφάρισμα-παιχνίδι

Πίνακας 1: Σύγκριση παλαιών και νέων συστημάτων ανάκτησης πληροφοριών

Προφανώς, ο χρήστης μιας υπερσυνδεδεμένης βιβλιοθήκης δεν θα είναι πλέον ένα νεωτερικό υποκείμενο με μια σταθερή ταυτότητα, που επιδιώκει να βρίσκεται μέσα σε ένα περισσότερο ολοκληρωμένο και αξιόπιστο μοντέλο του κόσμου,⁷ αλλά μάλλον ένα μετανεωτερικό ευμετάβλητο ον, έτοιμο να ανοιχθεί σε συνεχώς νέους ορίζοντες. Ένα τέτοιο ‘νέο’ (ή, καλύτερα, μετανεωτερικό) ον δεν ενδιαφέρεται να συλλέξει ότι είναι σημαντικό, ότι έχει σημασία, αλλά να συνδεθεί σε όσο το δυνατόν ευρύτερο ιστό πληροφοριών.

Οι σέρφερ του Ιστού αγκαλιάζουν τις διαρκώς πολλαπλασιαζόμενες πληροφορίες σαν να ήταν κάτι, που συμβάλλει στη δημιουργία μιας νέας μορφής ζωής και μιας νέας κατάστασης πραγμάτων, όπου η έκπληξη και ο θαυμασμός θεωρούνται πιο σημαντικά από το νόημα και τη χρηστικότητα. Μια τέτοια προσέγγιση ελκύει ιδιαίτερα εκείνους, που αρέσκονται στην ιδέα της απόρριψης της ιεραρχίας και της αυθεντίας και δεν νοιάζονται για το πρακτικό πρόβλημα της εύρεσης των έγκυρων

πληροφοριών. Έτσι, οι μετανεωτερικοί θεωρητικοί και καλλιτέχνες δέχονται τους υπερδεσμούς σαν ένα τρόπο απελευθέρωσης από τους ανώνυμους εμπειρογνώμονες, που αυθαίρετα οργανώνουν τις βάσεις των δεδομένων και αποφασίζουν μονομερώς για το τι είναι σχετικό με τι. Η ποσότητα των συνδέσεων εκτιμάται οπωσδήποτε περισσότερο από την ποιότητά τους. Μια τέτοια δημοκρατική ιδέα συχνά θεωρείται ότι προέρχεται από τις παν-Αμερικάνικες πολιτικές αξίες. Όπως παρατηρεί ο Fareed Zakaria, διευθυντής εκδότης του *Foreign Affairs*: 'Το Internet είναι βαθιά ασεβές ως προς την παράδοση, την κατεστημένη τάξη και την ιεραρχία και αυτή είναι μια ιδιαίτερα Αμερικάνικη στάση ζωής.'⁸

Εκείνοι όμως, που θέλουν να χρησιμοποιήσουν τα διαθέσιμα δεδομένα, πρέπει να βρουν τις πληροφορίες, που είναι σημαντικές και σχετικές για αυτούς, στο πλαίσιο των τρεχόντων ενδιαφερόντων τους. Αλλά, λαμβάνοντας υπόψη ότι σε μια υπερσυνδεδεμένη βάση δεδομένων τα πάντα μπορούν να είναι συνδεδεμένα με όλα τα άλλα, αυτή είναι μια πολύ δύσκολη αποστολή. Αφού οι υπερδεσμοί μπορούν να δημιουργηθούν εξ αιτίας του οποιουδήποτε λόγου και αφού υπάρχει μόνον ένας βασικός τύπος υπερδεσμού, ο ερευνητής δεν μπορεί να χρησιμοποιήσει το νόημα των συνδέσεων αυτών, για να φτάσει στην πληροφορία που αναζητεί. Το πρόβλημα είναι ότι, αναφορικά με το νόημα, όλοι οι υπερδεσμοί είναι όμοιοι. Όπως το θέτει ένας ερευνητής, η δουλειά της ανάκτησης των πληροφοριών είναι χειρότερη από το ψάχνει κανείς μια συγκεκριμένη βελόνα μέσα στα άχυρα' είναι σαν να ψάχνει κανείς μια συγκεκριμένη βελόνα μέσα σε έναν τεράστιο σωρό από βελόνες. Δεδομένης της απουσίας οποιουδήποτε σημαντικού περιεχομένου (από σημασιολογικής πλευράς), που προσδιορίζει τις συνδέσεις, κάθε μέσο αναζήτησης στον Ιστό είναι καταδικασμένο να είναι απλά και μόνο μια τυπική συντακτική τεχνική για το χειρισμό συμβόλων χωρίς σημασία, όταν το πρόβλημα είναι ακριβώς το αντίθετο – ο εντοπισμός του σχετικού σημαντικού περιεχομένου.

Η δυσκολία χρησιμοποίησης των ανούσιων μηχανικών λειτουργιών για την ανάκτηση σημαντικών πληροφοριών είναι γνωστή από παλιά, πριν από τον ερχομό του Internet. Μια τέτοια δυσκολία εμφανίζεται, όταν αναζητά κανείς την ανάκτηση πληροφοριών από μια βάση δεδομένων, πληροφοριών οι οποίες είναι σχετικές με ένα συγκεκριμένο σκοπό, αλλά τυχαίνει η βάση δεδομένων να μην είναι οργανωμένη, έτσι ώστε να εξυπηρετείται εκείνος ο συγκεκριμένος σκοπός. Μια τυπική περίπτωση είναι όταν ένας ερευνητής ψάχνει για δημοσιευμένες εργασίες σε ένα θέμα, που του ενδιαφέρει, αλλά οι απλές λέξεις των τίτλων των εργασιών δεν μπορούν να δώσουν τη δυνατότητα στη μηχανή αναζήτησης να επιστρέψει μόνο εκείνα τα κείμενα ή τους δικτυακούς τόπους, που ανταποκρίνονται στις συγκεκριμένες ανάγκες του ερευνητή.

Για να κατανοήσουμε το πρόβλημα βοηθά να ξεχωρίσουμε την Ανάκτηση Δεδομένων (ΑΔ) από την Ανάκτηση Πληροφοριών (ΑΠ). Ο David Blair εξηγεί τη διαφορά:

Τα Συστήματα Διαχείρισης Βάσεων Δεδομένων έχουν επαναστατικοποιήσει τη διαχείριση και την ανάκτηση των δεδομένων – μπορούμε να τηλεφωνήσουμε στην υπηρεσία καταλόγου και να πάρουμε τον αριθμό τηλεφώνου σχεδόν οποιουδήποτε, οπουδήποτε μέσα στις ΗΠΑ ή στον Καναδά' μπορούμε να πάμε σε ένα ATM σε μια πόλη, πολύ μακριά από την πόλη μας, και να κάνουμε ανάληψη χρημάτων από τον τραπεζικό λογαριασμό μας' μπορούμε να πάμε σε ένα γραφείο εισιτηρίων στο Michigan και να κλείσουμε μια θέση για ένα θεατρικό έργο στο San Francisco κ.λπ. Όλα αυτά είναι δυνατά, εν μέρει, λόγω των μεγάλης-κλίμακας αξιόπιστων

συστημάτων διαχείρισης βάσεων δεδομένων, που έχουν αναπτυχθεί τα τελευταία 35 χρόνια.

Η ΑΔ λειτουργεί με οντότητες, όπως ‘ονόματα’, ‘διευθύνσεις’, ‘αριθμοί τηλεφώνου’, ‘υπόλοιπα λογαριασμών’, ‘αριθμοί κοινωνικής ασφάλισης’ κ.ο.κ. – όλα, δηλαδή, πράγματα που έχουν ξεκάθαρες, αναμφίβολες αναφορές. Αλλά, μολονότι κάποια από τα γενικά χαρακτηριστικά των πληροφοριών αντιστοιχούν σε ξεκάθαρες έννοιες και αναφορές – όπως ο συγγραφέας ή ο τίτλος ενός κειμένου – πολλές αναζητήσεις ΑΠ δεν βασίζονται σε συγγραφείς ή τίτλους, αλλά ενδιαφέρονται για το ‘θεματικό περιεχόμενο’ των κειμένων (π.χ., ‘βρείτε όλες τις αναφορές, που αναλύουν τις προοπτικές των επενδύσεων στην Κεντρική Ευρώπη σε βιομηχανίες υπηρεσιών’). Οι περιγραφές του *θεματικού περιεχομένου* δεν είναι σχεδόν ποτέ καθορισμένες με απόλυτη ακρίβεια και σε μεγάλα συστήματα ΑΠ, ειδικά στο WWW, οι θεματικές περιγραφές είναι συνήθως απειριστικά ανακριβείς ή αόριστες, για όλες εκτός από την πολύ γενική αναζήτηση.

Έτσι, η αναζήτηση ενός γνωστού URL στο WWW είναι απλή και εύκολη· έχει την ακρίβεια και την αμεσότητα της ΑΔ. Αλλά η έρευνα για μια Ιστοσελίδα με ένα συγκεκριμένο θεματικό περιεχόμενο, χρησιμοποιώντας τις μηχανές αναζήτησης του Ιστού, μπορεί να είναι πολύ δύσκολη, κάποιες φορές αδύνατη.⁹

Η διαφορά μεταξύ Ανάκτησης Δεδομένων και Ανάκτησης Πληροφοριών μπορεί να συνοψιστεί, όπως φαίνεται, στον Πίνακα 2.

ΑΝΑΚΤΗΣΗ ΔΕΔΟΜΕΝΩΝ	ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΩΝ
1. Άμεση (‘Θέλω να μάθω το X’)	1. Έμμεση (‘Θέλω να μάθω για το X’)
2. Απαραίτητη η σχέση μεταξύ του αιτήματος και μιας ικανοποιητικής απάντησης	2. Πιθανοτική η σχέση μεταξύ του αιτήματος και μιας ικανοποιητικής πληροφορίας
3. Κριτήριο επιτυχίας = ορθότητα	3. Κριτήριο επιτυχίας = χρηστικότητα
4. Η κλιμάκωση δεν είναι μεγάλο πρόβλημα	4. Η κλιμάκωση είναι ένα μεγάλο πρόβλημα

Πίνακας 2: Διαφορές μεταξύ ανάκτησης δεδομένων και ανάκτησης πληροφοριών

Πριν την έλευση του Ιστού και των μηχανών αναζήτησης στον Ιστό, είχε δοκιμαστεί η λύση στο πρόβλημα της ανάκτησης πληροφοριών με τη διαμεσολάβηση ανθρώπων – δηλαδή, των αρχειοθετών, που κατανοούν τις πληροφορίες και βοηθούν στην περιγραφή των περιεχομένων τους, έτσι ώστε να μπορούν αυτές να ανακτώνται από όσους τις αναζητούσαν. Αλλά προφανώς τώρα δεν μπορούν να υπάρξουν αρχειοθέτες-άνθρωποι, που θα δημιουργήσουν ένα ευρετήριο όλου του Ιστού, γιατί είναι πολύ μεγάλος και αναπτύσσεται πάρα πολύ γρήγορα.

Για να κατανοήσουμε το μέγεθος του προβλήματος της ανάκτησης των πληροφοριών, βοηθά να μελετήσουμε μια αναλογία, που έδωσε ο Blair:

Έστω ότι θέλουμε να βρούμε ένα βιβλίο, το οποίο είναι ένα από τις πολλές εκατοντάδες μεταξύ αυτών που μας είναι προσβάσιμα. Το πρόβλημα αυτό μοιάζει με την εύρεση ενός συγκεκριμένου ατόμου σε ένα συνωστισμένο δωμάτιο μετρίου μεγέθους. Όχι ένα ιδιαίτερα δύσκολο πρόβλημα, ακόμα και εάν η περιγραφή του βιβλίου ή του προσώπου, που ψάχνουμε, είναι αρκετά γενική. Αλλά υποθέστε ότι θέλουμε να βρούμε ένα βιβλίο σε μια σχετικά μικρή βιβλιοθήκη των 50.000 βιβλίων. Αν και έχουμε όλοι βρεθεί σε βιβλιοθήκες

αυτού του μεγέθους, μπορεί να είναι λίγο δύσκολο να φανταστούμε το μέγεθος του προβλήματος. Θεωρείστε ένα παρόμοιο πρόβλημα: Καθώς πολλά από τα στάδια του επαγγελματικού baseball στις ΗΠΑ χωράνε περίπου 50.000 θεατές, θα μπορούσαμε να οπτικοποιήσουμε καλύτερα το πρόβλημα αυτής της αναζήτησης, εάν φανταζόμασταν ότι ο στόχος μας θα ήταν να βρούμε ένα συγκεκριμένο άτομο, το οποίο παρακολουθεί έναν κατάμεστο αγώνα, ας πούμε, στο Fenway Park. Αλλά τώρα το πρόβλημα είναι πιο δύσκολο. Έστω επίσης ότι οι οδηγίες, για να βρούμε το πρόσωπο που ψάχνουμε, είναι αρκετά γενικές: ότι είναι μεσήλικας, έχει μαύρα μαλλιά, σκούρα μάτια, είναι περίπου 5'10'' ύψος και αδύνατος. Τώρα υποθέστε ότι ψάχνουμε ένα βιβλίο σε μια αρκετά μεγάλη βιβλιοθήκη μερικών εκατοντάδων χιλιάδων βιβλίων. Εδώ, η αναλογία θα ήταν να αναζητούσαμε κάποιο άτομο σε μια συναυλία των Rolling Stones στο Central Park της Νέας Υόρκης. Αλλά ακόμα και τότε, θα απέχαμε πολύ από το να αντιληφθούμε το μέγεθος του χώρου αναζήτησης στο Internet. Η αναζήτηση μέσα στα εκατομμύρια σελίδων με διαφορετικό θεματικό περιεχόμενο, που είναι επί του παρόντος διαθέσιμες στο Διαδίκτυο, χρησιμοποιώντας μόνο αυτά τα εργαλεία αναζήτησης, που είναι επίσης επί του παρόντος διαθέσιμα, είναι ανάλογη με την αναζήτηση μέσα στην πόλη της Νέας Υόρκης ενός συγκεκριμένου ατόμου με μόνο στοιχείο τη γενική περιγραφή ότι έχει σκούρα μαλλιά, σκούρα μάτια, είναι μεσήλικας και αδύνατος.¹⁰

Μπροστά σε ένα τόσο φοβερό πρόβλημα, οι ερευνητές που εργάζονται για την ανάκτηση των πληροφοριών στράφηκαν στην Τεχνητή Νοημοσύνη (TN). Από τη δεκαετία του 1960, οι ερευνητές της TN εργάζονται για να λύσουν το πρόβλημα του πώς θα κάνουν τους υπολογιστές, που είναι συντακτικές μηχανές ευαίσθητες μόνο στη μορφή ή τη σχηματική δομή των δεδομένων εισόδου, να συμπεριφέρονται σαν άνθρωποι, που είναι ευαίσθητοι στη σημασιολογία ή το νόημα. Έτσι, φυσικά, οι ερευνητές στράφηκαν στην TN για να καταλάβουν πώς πρέπει να προγραμματίζουν τους υπολογιστές ώστε να βρίσκουν ακριβώς εκείνες τις πληροφορίες, των οποίων η συνάφεια θα αναγνωριζόταν από ένα ανθρώπινο ον, που θα έκανε την αναζήτηση. Στην αρχή, οι ερευνητές της TN ήταν αισιόδοξοι ότι θα μπορούσαν να αναπαραστήσουν όλα τα στοιχεία του κόσμου, για τα οποία ενδιαφέρονται οι άνθρωποι, αναπαριστώντας μερικά εκατομμύρια πράγματα και γεγονότα και προσθέτοντας κατάλληλους κανόνες, ώστε να μπορούν να βρίσκουν ποια από αυτά είναι σχετικά σε κάθε δεδομένη κατάσταση. Αλλά στα τέλη της δεκαετίας του 1970 και στις αρχές της δεκαετίας του 1980, οι ερευνητές της TN απρόθυμα έφτασαν να αναγνωρίσουν ότι, για να παραχθεί η τεχνητή νοημοσύνη, θα έπρεπε να αποσαφηνίσουν και να οργανώσουν τη γνώση της κοινής λογικής, που μοιράζονται οι άνθρωποι, και αυτό ήταν ένα αζεπέραστο έργο.¹¹ Ο πιο γνωστός συνήγορος της προσέγγισης αυτής ήταν ο Douglas Lenat.¹²

Ο Lenat αντιλαμβανόταν ότι η γνώση της κοινής λογικής δεν είναι το είδος γνώσης, που βρίσκουμε στις εγκυκλοπαίδειες, αλλά μάλλον το είδος της γνώσης, που θεωρείται δεδομένη από αυτούς, που γράφουν τα άρθρα στις εγκυκλοπαίδειες. Μια τόσο στοιχειώδης (‘σιωπηρή’) γνώση-υπόβαθρο είναι εντελώς αυτονόητη, που δύσκολα κανείς την αντιλαμβάνεται. Ο Lenat τόνιζε ότι, για να καταλάβουμε ένα άρθρο για τον George Washington, για παράδειγμα, μπορεί να χρειάζεται να καταλαβαίνουμε τέτοια αυτονόητα πράγματα, όπως το ότι, όταν αυτός βρισκόταν στο Καπιτώλιο, το ίδιο ήταν και το αριστερό του πόδι, και το ότι, όταν αυτός πέθανε, έμεινε νεκρός. Έτσι, το 1985, ο Lenat ισχυρίστηκε ότι, μέσα στα επόμενα δέκα χρόνια, θα συλλάμβανε αυτήν την κοινή λογική δημιουργώντας ‘ένα μόνο νοήμονα

πράκτορα, του οποίου η βάση γνώσης θα περιλάμβανε ... εκατομμύρια καταχωρήσεις'.¹³

Ο Lenat έχει ως τώρα ξοδέψει 15 χρόνια και τουλάχιστον 15 εκατομμύρια δολάρια, για να αναπτύξει το CYC, μια βάση δεδομένων γνώσης κοινής λογικής, προσπαθώντας να καταστήσει ικανούς τους υπολογιστές να κατανοήσουν τέτοια ζητήματα κοινής λογικής, όπως τα πληροφοριακά αιτήματα. Αυτό θεωρείται ότι είναι ένα πρώτο βήμα προς την επίλυση του προβλήματος της ανάκτησης των πληροφοριών. Σαν μια συγκεκριμένη επίδειξη της χρήσης του CYC, ο Lenat ανέπτυξε ένα σύστημα ανάκτησης φωτογραφιών, σαν ένα παράδειγμα του πώς η γνώση της κοινής λογικής παίζει έναν ουσιαστικό ρόλο στην ανάκτηση των πληροφοριών. Το σύστημα αυτό υποτίθεται πως ανακτά on-line εικόνες, όταν δίνονται κάποια χαρακτηριστικά (ή επεξηγήσεις) για κάθε μια εικόνα μιας βάσης δεδομένων. Αντί ενός δισεκατομμυρίου εικόνων, όπως θα μπορούσε κανείς να βρει στον Ιστό, ο Lenat ξεκίνησε πιο συντηρητικά με είκοσι φωτογραφίες. Ένας καθηγητής του Stanford περιγράφει την εμπειρία του με το σύστημα αυτό ως εξής:

Η επίδειξη του CYC έγινε με είκοσι εικόνες. Το αίτημα, 'κάποιος που αναπαύεται', απέδωσε μια εικόνα 3 αντρών με μαγιό, που κρατούν σανίδες του σερφ. Το CYC βρήκε αυτήν την εικόνα συσχετίζοντας την 'ανάπαυση' με τα χαρακτηριστικά των εικόνων, που είχαν εισαχθεί από πριν στη βάση δεδομένων (των 20 εικόνων). Αλλά ακόμα και για 20 εικόνες, το σύστημα δεν λειτουργεί πολύ καλά.¹⁴

Στο βαθμό που το σύστημα λειτουργεί, αυτό γίνεται μόνο επειδή οι προγραμματιστές του CYC έχουν αποφασίσει σαν *γνώση* κάποιες από τις *αντιλήψεις*, που οι άνθρωποι έχουν για την ανάπαυση, την κίνηση, την προσπάθεια κ.λπ., αντιλήψεις προερχόμενες από τη σωματικότητα των ανθρώπων και τα βιώματα, που αυτή συνεπάγεται. Αλλά το μεγαλύτερο μέρος των αντιλήψεων των συνεπειών της σωματικότητας είναι τόσο ακαθόριστες και βιωματικές, ώστε να υπάρχει κάθε λόγος να αμφιβάλλουμε ότι θα μπορούσαν οι αντιλήψεις αυτές να συστηματοποιηθούν με ακρίβεια και να εισαχθούν σε μια βάση δεδομένων ενός 'ασώματου' υπολογιστή.

Αυτό, φυσικά, δεν είναι πρόβλημα για εμάς, στην καθημερινή ζωή μας. Μπορούμε να ανακαλύψουμε τις απαντήσεις για τα ερωτήματα, που αφορούν το σώμα, χρησιμοποιώντας το ίδιο το σώμα μας ή φανταζόμενοι πώς θα ήταν να κάναμε το ένα ή το άλλο πράγμα. Έτσι, για παράδειγμα, καταλαβαίνουμε ότι με τις κάμψεις δεν μπορούμε να αναπαυθούμε, απλώς φανταζόμενοι ότι εκτελούμε μια τέτοια δραστηριότητα. Αλλά, για το CYC, μια φωτογραφία κάποιου, που κάνει κάμψεις, θα έπρεπε να είχε εκ των προτέρων περιγραφεί από τον προγραμματιστή σαν η εικόνα κάποιου, που γυμνάζεται. Μόνο τότε θα μπορούσε το CYC να 'συμπεράνει' ότι το άτομο στη φωτογραφία αυτή δεν αναπαύοταν.

Γενικώς, επειδή έχουμε σώματα, μπορούμε να παράγουμε κατά βούληση έναν απεριόριστα μεγάλο αριθμό γεγονότων για το σώμα μας, τόσων πολλών, που δεν μπορούμε και ούτε ποτέ θα μπορούσαμε να τα αποθηκεύσουμε όλα σαν ρητή-αποσαφηνισμένη γνώση. Αλλά το CYC δεν έχει σώμα, οπότε, όπως έχουμε δει, πρέπει να του δοθούν όλα τα γεγονότα για το σώμα, που χρειάζεται να ξέρει, για να ανακτήσει πληροφορίες από τη βάση των δεδομένων του. Επιπλέον, το CYC δεν θα καταλάβαινε πώς θα έπρεπε να χρησιμοποιήσει τα γεγονότα, που πράγματι γνώριζε, για να απαντήσει κάποιο νέο ερώτημα για το σώμα. Για παράδειγμα, στην περίπτωση

που ρωτούσε κανείς το CYC, αν οι άνθρωποι μπορούν να μασούν τσίγλα και να σφυρίζουν ταυτοχρόνως, δεν θα έπαιρνε καμία απάντηση, ακόμα και όταν το CYC γνώριζε πολλά γεγονότα για το μάσημα και το σφύριγμα, μέχρις ότου ένα ενσώματο ανθρώπινο να φανταζόταν ότι προσπαθούσε να το κάνει και στη συνέχεια να πρόσθετε την απάντηση στη βάση δεδομένων του CYC. Αλλά είναι απερίριστος ο αριθμός τέτοιων γεγονότων για το σώμα, που θα χρειαζόταν κανείς πρώτα να αποσαφηνίσει και μετά να αποθηκεύσει, προβλέποντας ότι θα μπορούσαν να ήταν σχετικά με κάποιο ερώτημα. Ευτυχώς, έχοντας σώμα, απαλλασσόμαστε εξ ολοκλήρου από την ανάγκη να αποθηκεύουμε οποιαδήποτε τέτοια γεγονότα.

Αλλά ακόμα και εάν όλα όσα αντιλαμβανόμαστε, απλώς επειδή έχουμε σώμα, μπορούσαν να οριστούν ρητά και να εισαχθούν στη βάση δεδομένων του CYC, θα υπήρχε ακόμα το γενικότερο πρόβλημα της παρακολούθησης των αλλαγών στον κόσμο, που θα απαιτούσαν αλλαγές στη βάση δεδομένων. Ακόμα και ένας δικτυακός τόπος με ένα απλό και ξεκάθαρο τίτλο, όπως 'Bill Clinton,' θα μπορούσε να αφορά μια πολλαπλότητα θεμάτων, το καθένα από τα οποία θα μπορούσε να ενδιαφέρει μόνο κάποια ομάδα χρηστών. Επιπλέον, τα ενδιαφέροντα των χρηστών θα μεταβάλλονταν, καθώς θα άλλαζε η επικαιρότητα. Μια μέρα, η εξωτερική πολιτική θα μπορούσε να ήταν το μείζον θέμα των ενδιαφερόντων και, την επόμενη, τα φώτα της επικαιρότητας να εστιάζονταν στην Έκθεση του Starr. Αφού το τι είναι σχετικό για τον Clinton αλλάζει από μέρα σε μέρα, θα ήθελε κανείς να είχε κάποιο μηχανισμό, που θα παρακολουθούσε τις καθημερινές αλλαγές της επικαιρότητας, έτσι ώστε ο υπολογιστής να μπορούσε να ενημερώνεται για το πώς οι δικτυακοί τόποι αναφέρονται στον Clinton, καθώς η σημασία του περιεχομένου τους θα μεταβαλλόταν.

Αλλά ένας απερίριστα μεγάλος αριθμός αλλαγών στον κόσμο συμβαίνουν συνεχώς η ημερομηνία αλλάζει, το ίδιο και οι σχηματισμοί των νεφών, όπως επίσης και το βάρος του Clinton, η ηλικία του, ο τόπος που βρίσκεται, οι απόψεις του, κ.λπ. Μόνο όμως μερικές από αυτές τις αλλαγές είναι σχετικές με το τι οι άνθρωποι, κάθε δεδομένη μέρα, προσδοκούν να βρουν σε ένα δικτυακό τόπο για τον Clinton. Ένας μηχανισμός για την ενημέρωση του τρόπου, με τον οποίο παρουσιάζονται οι πληροφορίες στον Ιστό, θα έπρεπε, επομένως, να μπορεί να αγνοεί σχεδόν όλες τις αλλαγές, που γίνονται στον κόσμο και στην ζωή του Clinton, και να λαμβάνει υπόψη μόνο τις σχετικές.

Επειδή τα ανθρώπινα όντα ανταποκρίνονται μόνο στις σχετικές με αυτά αλλαγές, με βάση τα σώματά τους και τα ενδιαφέροντά τους, δεν θα έπρεπε να δημιουργήσει καμία έκπληξη το γεγονός ότι κανείς δεν έχει ποτέ καταφέρει να προγραμματίσει τον υπολογιστή, έτσι ώστε να ανταποκρίνεται σε ό,τι είναι σχετικό. Πράγματι, το πρόβλημα της αναγνώρισης των αλλαγών, που είναι σχετικές σε ένα δεδομένο πλαίσιο, έχει αναγνωριστεί ως ένα σοβαρό πρόβλημα, από τότε που πρωτοεμφανίστηκε στην Τεχνητή Νοημοσύνη, κατά την δεκαετία του 1960. Ονομάζεται 'πρόβλημα του πλαισίου' και παραμένει άλυτο ως σήμερα.

Ο Lenat αντιλαμβάνεται ότι το πρόβλημα της συνάφειας απειλεί όλο το πρόγραμμά του και προτείνει, όπως θα έπρεπε, την αντικατάσταση της έννοιας της συνάφειας, που βασίζεται στο νόημα, από κάποια τυπικά αξιώματα. Προτείνει δύο είδη αξιωμάτων συνάφειας: τα *συγκεκριμένα* και τα *γενικά*. Η ιδέα πίσω από τα συγκεκριμένα αξιώματα της συνάφειας είναι ότι οι διαφορετικοί τομείς της βάσης

δεδομένων των γνώσεων ‘μπορούν να βαθμολογηθούν σύμφωνα με τη συνάφειά τους με το έργο επίλυσης του προβλήματος, κάθε φορά’.¹⁵ Έτσι, για παράδειγμα, εάν το έργο, που δίνεται στο CYC, είναι να βρει πληροφορίες σχετικές με τη σχεδίαση ενός ολοκληρωμένου κυκλώματος, το πρόγραμμα θα καθοδηγηθεί στην αναζήτησή του από ένα αξίωμα, το οποίο υπαγορεύει ότι ο τομέας των υπολογιστών είναι πιο σχετικός από τον τομέα της βοτανικής (αν και ο τομέας της βοτανικής δεν μπορεί να αποκλεισθεί εντελώς, αφού μπορεί να είναι η πηγή κάποιων χρήσιμων αναλογιών).¹⁶

Αλλά δεν είναι απλώς η αναλογία, αυτό που κάνει κάποια γεγονότα να θεωρούνται σχετικά με άλλα, τα οποία φαινομενικά βρίσκονται πολύ μακριά. Θεωρείστε την περίπτωση κάποιου, που θέλει να στοιχηματίσει σε μια ιπποδρομία υπέρ ενός συγκεκριμένου αναβάτη, γνωρίζοντας όμως ότι ο αναβάτης αυτός υποφέρει από μια αλλεργία. Έτσι, θα αλλάξει το στοίχημά του, αν τυχόν παρατηρήσει ότι στο στίβο υπάρχει κάποιο φυτό, που ερεθίζει την αλλεργία του αναβάτη. Το να κάνουμε τον υπολογιστή να καταλάβει ότι αυτή η φαινομενικά άσχετη αλλαγή στον κόσμο είναι πολύ σχετική με το στοίχημα, που θα βάλουμε, σημαίνει ότι πάλι αντιμετωπίζουμε το πρόβλημα του πλαισίου. Στην πραγματικότητα, όλα όσα γνωρίζουμε μπορούν να συνδεθούν με όλα τα άλλα, με άπειρους τρόπους, που έχουν νόημα. Σε τέτοιες περιπτώσεις, μόνο η κατανόηση των νοημάτων, που εμπλέκονται, μας επιτρέπει να επιλέξουμε τι είναι σχετικό με το προς επίλυση πρόβλημα. Άρα, αντί να βοηθούν στην επίλυση του προβλήματος της συνάφειας, τα συγκεκριμένα αξιώματα συνάφειας απλώς εγείρουν το γενικό πρόβλημα της συνάφειας με δραματικότερη μορφή.

Για να λυθεί το γενικό πρόβλημα της συνάφειας, ο Lenat προτείνει τα γενικά αξιώματα συνάφειας. Αυτά είναι τυποποιήσεις τέτοιων προτάσεων, όπως ‘είναι απαραίτητο να θεωρούμε μόνο συμβάντα, που είναι χρονικά κοντά στο χρόνο του υπό συζήτηση συμβάντος ή κατάστασης’.¹⁷ Αυτό πράγματι θα περιλάμβανε το φυτό της αλλεργίας στο προηγούμενο παράδειγμα, αλλά, φυσικά, θα πρόσθετε έναν απεριόριστα μεγάλο αριθμό άλλων γεγονότων για το στίβο της ιπποδρομίας, έτσι ώστε θα ήταν αδύνατο το πρόβλημα της συνάφειας να λυνόταν. Επιπλέον, εξηγώντας και υπερασπίζοντας αυτό το αξίωμα, οι Guha και Levy λένε ‘είναι σπάνιο να συμβεί ένα γεγονός και ... [μετά από] μια σημαντική χρονική περίοδο ... ξαφνικά να εκδηλώνονται τα αποτελέσματα του’.¹⁸ Αλλά οι υποσχέσεις και όλα τα είδη των προβλημάτων υγείας, για να πάρουμε μόνο δύο παραδείγματα, έχουν ακριβώς αυτό το χαρακτηριστικό, ότι τα σχετικά αποτελέσματα μπορεί να εμφανισθούν μακριά στο μέλλον, και όλα τα είδη των ιστορικών και ψυχολογικών γεγονότων, που είναι σχετικά με το παρόν ενός ατόμου, μπορούν να αναζητηθούν στο περισσότερο ή λιγότερο μακρινό παρελθόν του.¹⁹

Όταν ο Lenat ξεκίνησε το πρόγραμμά του, πριν δεκαπέντε χρόνια, ισχυριζόταν ότι σε δέκα χρόνια το CYC θα μπορούσε να διαβάξει άρθρα εφημερίδων και να καταλογοποιεί τα νέα γεγονότα, που θα έβρισκε εκεί, μέσα στη βάση των δεδομένων του, χωρίς καμία ανθρώπινη βοήθεια. Αυτό είναι το όνειρο όλων εκείνων, οι οποίοι προσδοκούν ότι οι τεχνητά νοήμονες πράκτορες θα μπορούν να βρίσκουν και να παραδίδουν σε όλους τις πληροφορίες, που τους ενδιαφέρουν. Αλλά, όπως ο Μιχάλης Δερτούζος επισημαίνει στην επιγραφή της επικεφαλίδας του κεφαλαίου αυτού, ένα τέτοιο επίτευγμα δεν έχει ακόμα συμβεί. Το ηθικό δίδαγμα είναι, όπως τονίζει ο Don Swanson, ότι ‘οι μηχανές δεν μπορούν να αναγνωρίσουν το νόημα και, έτσι, δεν μπορούν να αντιγράψουν όλα εκείνα, τα οποία η ανθρώπινη κρίση, κατά κανόνα, μπορεί να φέρει στη διαδικασία της αρχειοθέτησης και ταξινόμησης των κειμένων’.²⁰

Η αποτυχία των προγραμμάτων της TN, όπως του Lenat, θα έπρεπε να προκαλέσει την προσοχή μας για το πόσο σημαντικά είναι τα σώματά μας στη διαδικασία της κατανόησης του κόσμου. Πράγματι, η μορφή της ζωής μας είναι οργανωμένη από και για όντα με σώματα σαν και εμάς: πλάσματα με σώματα, που έχουν χέρια και πόδια, εσωτερικό και εξωτερικό· που πρέπει να ισορροπούν μέσα σε ένα βαρυτικό πεδίο· που κινούνται ευκολότερα προς τα μπρος παρά προς τα πίσω· που κουράζονται· που, για να πλησιάσουν τα αντικείμενα, πρέπει να διασχίσουν τον ενδιάμεσο χώρο και να ξεπεράσουν εμπόδια, καθώς προχωρούν· κ.λπ. Οι δυνατότητες της σωματικότητάς μας είναι τόσο διάχυτες στον κόσμο μας, ώστε δεν αντιλαμβανόμαστε τον τρόπο, με τον οποίο το σώμα μας αφήνει να τον κατανοήσουμε.²¹ Θα το αντιλαμβανόμασταν μόνον, όταν θα χρειαζόταν να βιώσουμε τον αποπροσανατολισμό μας, αν, για παράδειγμα, μεταφερόμασταν σε ένα εξωγήινο κόσμο, του οποίου τα πλάσματα θα είχαν θεμελιωδώς διαφορετικά – ασ πούμε, σφαιρικά ή αεριώδη – σώματα, ή όταν θα παρατηρούσαμε την απελπιστική σύγχυση τέτοιων εξωγήινων πλασμάτων, αν τυχόν βρισκότουσαν στον κόσμο μας.

Θα παίρναμε προφανώς μεγάλη βοήθεια, αν μπορούσαμε να χρησιμοποιήσουμε τις σωματικές αισθήσεις μας για το τι είναι σχετικό, για όντα με σώματα και ενδιαφέροντα σαν τα δικά μας, ως το υπόβαθρο των αναζητήσεων των σχετικών πληροφοριών σε βάσεις δεδομένων και σε παγκόσμιους δικτυακούς τόπους. Αλλά, όπως έχει δείξει η αποτυχία του Lenat να πετύχει το στόχο του να αποσαφηνίσει τη γνώση της κοινής λογικής, δεν υπάρχει λόγος να ελπίζουμε ότι μπορούμε να τυποποιήσουμε τη γνώση, που έχουμε από τη σωματικότητά μας. Έτσι, η ελπίδα ότι η Τεχνητή Νοημοσύνη θα μπορούσε να λύσει το πρόβλημα της συνάφειας έχει τώρα κατά πολύ εγκαταλειφθεί. Υπάρχει πλέον ένας τεράστιος και συνεχώς αυξανόμενος όγκος πληροφοριών στον κόσμο και φαίνεται ότι η μόνη μας πρόσβαση σε αυτές θα πρέπει να είναι μέσω των υπολογιστών, που δεν έχουν σώματα, δεν μοιράζονται τον κόσμο μας και, άρα, δεν καταλαβαίνουν το νόημα των κειμένων και των δικτυακών τόπων.

Εάν αφήσουμε κατά μέρος την κοινή λογική και την ενσώματη αντίληψή μας του κόσμου, όπως η χρήση των υπολογιστών μας επιβάλλει, πρέπει να κάνουμε τα πράγματα με τον τρόπο του υπολογιστή, να επιχειρούμε, δηλαδή, να εντοπίζουμε τις σχετικές πληροφορίες αντικαθιστώντας τη σημασία με συσχετίσεις μεταξύ τυπικών συμβόλων. Τότε, βλέπουμε να υπάρχει στο Διαδίκτυο μια ολόκληρη βιομηχανία ανάκτησης πληροφοριών, της οποίας ο αποκλειστικός στόχος είναι η ανάπτυξη 'εξερευνητών' του Ιστού (*Web crawlers*) και μηχανών αναζήτησης, που προσπαθούν να προσεγγίσουν την αίσθηση συνάφειας των ανθρώπων μέσω της χρήσης συμβόλων, χωρίς νόημα, στον υπολογιστή.

Οι ερευνητές στην ανάκτηση των πληροφοριών ξεχωρίζουν δυο είδη ανάκτησης: την *επιστροφή* και την *ακρίβεια*. Σε μια ιδανική περίπτωση, ο ερευνητής θα μπορούσε ανακτήσει το 100% των πληροφοριών, σαν πληροφορίες σχετικές με την αναζήτησή του, και το 100% αυτών (δηλαδή, των ανακτηθέντων πληροφοριών) θα μπορούσαν να ήταν πράγματι σχετικές με τις συγκεκριμένες απαιτήσεις της αναζήτησής του. Με άλλα λόγια, στην ιδανική περίπτωση, θα μπορούσε να ανακτήσει κανείς όλες – και μόνο αυτές – τις σχετικές πληροφορίες. Στην πραγματικότητα, όμως, κάτι τέτοιο δεν είναι πολύ πιθανό να συμβεί, οπότε η 'επιστροφή' αναφέρεται στο ποσοστό των ανακτηθέντων πληροφοριών, ενώ η 'ακρίβεια' στο ποσοστό των ανακτηθέντων

πληροφοριών, οι οποίες είναι σχετικές με την αναζήτηση. Η επιστροφή και η ακρίβεια δεν είναι ανεξάρτητα μεγέθη, παρότι οι ερευνητές μπορεί κάποτε να αναγκάζονται να θεωρούν το ένα στη θέση του άλλου. Καθώς ο ερευνητής προσπαθεί να μεγιστοποιήσει την επιστροφή, η ακρίβεια τείνει να μειωθεί, και καθώς προσπαθεί να μεγιστοποιήσει την ακρίβεια, η επιστροφή τείνει να ελαττωθεί. Σαν αποτέλεσμα, μια έρευνα που ισχυρίζεται ότι έχει πετύχει 100% επιστροφή και 100% ακρίβεια δεν αποτελεί, εκτός από σπάνιες περιπτώσεις, παρά μια ανέφικτη ιδανική περίπτωση.

Η επιστροφή και η ακρίβεια γίνονται ακόμη πιο δύσκολες να μεγιστοποιηθούν, όταν ο όγκος της βάσης δεδομένων των πληροφοριών μεγαλώνει. Δοθέντος του τεράστιου μεγέθους του Internet, εκτιμάται ότι οι μηχανές αναζήτησης μπορούν να επιστρέψουν το πολύ το 2% των σχετικών δικτυακών τόπων. Ο Blair εξηγεί γιατί αυτό το σημαντικό γεγονός σπάνια σχολιάζεται:

Παρά το ότι ο συνολικός όγκος των πληροφοριών είναι τεράστιος και οι δυσκολίες αναζήτησης για συγκεκριμένο περιεχόμενο είναι ανυπέρβλητες, το μεγαλύτερο μέρος της δημοσιότητας, που πήρε ο Παγκόσμιος Ιστός, ήταν πράγματι θετικό. Ο πρωτοπόρος της ΑΠ (ανάκτησης πληροφοριών) Don Swanson παρατηρούσε αυτό το φαινόμενο δεκαετίες πιο πριν και το ονόμαζε 'απάτη της αφθονίας'. Η απάτη της αφθονίας είναι το λάθος, που κάνει ένας ερευνητής, όταν χρησιμοποιεί ένα μεγάλο σύστημα ΑΠ και πετυχαίνει να βρει κάποιες χρήσιμες πληροφορίες. Ο Swanson τόνιζε ότι σε ένα αρκετά μεγάλο σύστημα ... σχεδόν κάθε αίτημα θα επιστρέψει κάποιες χρήσιμες πληροφορίες. Το λάθος είναι να θεωρήσει κανείς ότι επειδή του επέστρεψε μερικές χρήσιμες πληροφορίες, το σύστημα ΑΠ λειτουργεί καλά. Αυτό που δεν γνωρίζει είναι το πόσες πολλές καλύτερες πληροφορίες το σύστημα παρέλειψε να επιστρέψει.²²

Πράγματι, η εμπιστοσύνη στη δυνατότητα ανάκτησης ακριβώς και μόνο εκείνων των πληροφοριών, που αναζητεί κανείς, έχει νόημα, μόνο εφόσον υπάρχει μια ταξινομία, όπως εκείνη του Αριστοτέλη ή του δεκαδικού συστήματος του Dewey, η οποία αντιλαμβάνεται με ένα συγκεκριμένο τρόπο τη δομή και τις υποδιαρέσεις του συνόλου της βάσης των δεδομένων της. Αλλά στον κόσμο των υπερδεσμών, δεν μπορεί να υπάρχει καμία μεταφυσική λύση, που να κάνει κάτι τέτοιο.

Οι πρώτες μηχανές αναζήτησης απλώς δημιουργούσαν έναν κατάλογο λέξεων-κλειδιών συνδεδεμένων με μια σειρά κειμένων-πληροφοριών, που τις περιείχαν, και έθεταν μια βαθμολογία αναλόγως του αν ή όχι μια λέξη-κλειδί βρισκόταν στον τίτλο, στο σώμα, στην περίληψη, κ.λπ. του κειμένου. Οι ερευνητές γενικώς συμφωνούν, ωστόσο, ότι τόσο στοιχειώδεις τεχνικές έχουν μόνο μια πιθανότητα 10% ανάκτησης μιας χρήσιμης πληροφορίας σε ένα δεδομένο αίτημα.

Οι αποκαλούμενες μηχανές δημοτικότητας, οι οποίες συνδέουν τις δικτυακές σελίδες με συγκεκριμένα αιτήματα, μετρώντας τον αριθμό των κλικ και το χρόνο παραμονής των επισκεπτών στις σελίδες, αύξησαν τον αριθμό αυτό περίπου στο 20%. Η ουσία της χρήσης του αριθμού των κλικ και του χρόνου παραμονής στους δικτυακούς τόπους, σαν παράγοντες που μπορούν να βοηθήσουν τους ερευνητές να βρουν αυτό που αναζητούν, βρίσκεται στην εύλογη υπόθεση ότι κάποιος που θα έθετε ένα αίτημα παρόμοιο με ένα αίτημα, που είχε πιο πριν τεθεί από άλλους χρήστες, πιθανώς θα ικανοποιούνταν από μια απάντηση, που είχε ικανοποιήσει τους προηγούμενους ερευνητές. Και η ικανοποίηση αυτή θα μπορούσε ενδεχομένως να μετρηθεί με τον

αριθμό των κλικ σε μια συγκεκριμένη δικτυακή σελίδα και το χρόνο παραμονής σε αυτήν.

Αλλά και αυτό δεν λειτουργεί τόσο καλά όσο αναμενόταν. Το πρόβλημα βρίσκεται στην έννοια της ομοιότητας. Τα πάντα μπορούν να είναι όμοια με όλα τα άλλα με βάση έναν απεριόριστα μεγάλο αριθμό πλευρών του θέματος – για παράδειγμα, αυτό το βιβλίο και εσείς είστε όμοιοι ως προς το ότι βρίσκεστε στη γη, είσαστε φτιαγμένοι από ύλη, αντανακλάτε το φως, μαζεύετε σκόνη, κ.λπ. Όμως, προσέχει κανείς μόνο εκείνες τις ομοιότητες, που έχουν σημασία για αυτόν, δοθείσης της σωματικότητάς του και των ενδιαφερόντων του. Αφού ό,τι κάποιο άτομο θεωρεί όμοιο με κάτι άλλο εξαρτάται από τα ενδιαφέροντα του ατόμου για αυτό, οι υπολογιστές δεν μπορούν να εκφράσουν χρήσιμες κρίσεις για την ομοιότητα. Έτσι, δεν θα έπρεπε να μας εκπλήσσει, όταν διαπιστώναμε ότι η μέθοδος της μέτρησης των κλικ λειτουργεί μόνο εάν τα αιτήματα, που συγκρίνονται, είναι *παρόμοια*, και, επομένως, η μέθοδος αυτή αποτυγχάνει να καλύψει την περίπτωση των ίδιων αιτημάτων, τα οποία όμως εκφράζονται με βάση διαφορετικές πλευρές του θέματος. Επιπλέον, όπως ο Gordon Rios αναφέρει, ‘η ανάλυση των μεγάλης κλίμακας αρχείων των αιτημάτων (πάνω από 100 εκατομμύρια) δείχνει γενικώς ότι τα μισά από τα αιτήματα είναι μοναδικά, οπότε, για αυτά, η μηχανή αναζήτησης δεν έχει προηγούμενα δεδομένα για τον αριθμό των κλικ. Αυτό απαιτεί να γίνει κάποια συνάθροιση των “όμοιων” αιτημάτων και οδηγεί πάλι πίσω στο πρόβλημα της ομοιότητας.’

Σύμφωνα με την τελευταία τεχνική για την εύρεση των σχετικών δικτυακών τόπων σε μια αναζήτηση, οι αντιδράσεις των χρηστών των μηχανών αναζήτησης (τα κλικ) αντικαθίστανται από την ανάλυση των υπερδεσμών μεταξύ των δικτυακών σελίδων. Χρησιμοποιώντας τα σχόλια, τα οποία οι συγγραφείς των σελίδων παραθέτουν, όταν συνδέουν τις δικτυακές σελίδες τους με άλλες, οι πιο προηγμένες μηχανές αναζήτησης έχουν βελτιώσει την ακρίβεια των αναζητήσεων για κάποια αιτήματα. Όμως, το πρόβλημα της ομοιότητας εμφανίζεται πάλι, αφού ο χώρος των αιτημάτων είναι πολύ μεγαλύτερος από το συγκεκριμένο κείμενο, που χρησιμοποιείται για το σχολιασμό των υπερδεσμών. Και, φυσικά, τα σχόλια των υπερδεσμών μπορούν να αξιοποιηθούν αποτελεσματικά από και για τους σκοπούς των spammers (αυτών που στέλνουν διαφημιστικά ή άλλα μηνύματα με e-mail σε παραλήπτες, που δεν έχουν ερωτηθεί αν επιθυμούν να παίρνουν τέτοια μηνύματα) όπως και με τη δημοτικότητα των σελίδων (τα κλικ), οι spammers έχουν εκμεταλλευθεί αμέσως τα σχόλια των υπερδεσμών, για να διαφημίσουν τις δικές τους ψεύτικες δικτυακές σελίδες.

Ένα τυπικό πρόβλημα είναι ότι με τη δημοτικότητα των σελίδων και με τις πλασματικές περιγραφές των υπερδεσμών τείνουν να επισκιασθούν οι λιγότερο δημοφιλείς σελίδες, οι οποίες όμως μπορεί να είναι σχετικές ως προς κάποια συγκεκριμένη πλευρά του αιτήματος. Για παράδειγμα, αν ψάχνει κανείς τις εργασίες του διάσημου ερευνητή Michael Jordan, κάποιες μηχανές αναζήτησης, οι οποίες βασίζονται στη δημοτικότητα των κλικ, τον αγνοούν τελείως και αντί για αυτόν προβάλλουν τον παίκτη της καλαθοσφαίρισης Michael Jordan. Άλλες πάλι μηχανές αναζήτησης, προορισμένες ειδικά για χρήστες με επιστημονικά ενδιαφέροντα, ενώ κάνουν καλά τη δουλειά τους για τους χρήστες αυτούς, δεν τα πάνε και τόσο καλά, όταν χρησιμοποιούνται από χρήστες με άλλα, π.χ., αθλητικά, ενδιαφέροντα. Είναι ολοφάνερο ότι η επιλογή των τεχνικών και των πληροφοριών, οι οποίες έχουν χρησιμοποιηθεί και εγκατασταθεί στη βάση δεδομένων της μηχανής αναζήτησης,

καθορίζουν την οπτική γωνία, που ακολουθείται, όταν παρουσιάζονται τα αποτελέσματα της αναζήτησης στο χρήστη.

Ο Gordon Rios, μέλος του επιστημονικού προσωπικού της Inktomi, ενός από τους μεγαλύτερους παροχείς υπηρεσιών αναζήτησης, συνοψίζει την κατάσταση ως εξής:

Έχουμε κάνει μελέτες μεγάλης κλίμακας για όλες τις κύριες μηχανές αναζήτησης και έχουμε βρει ότι η χρήση του κειμένου, της δημοτικότητας των κλικ των επισκεπτών και της δομής και των σχολίων των υπερδεσμών δίνουν μια ακρίβεια περίπου 20-30% για λογικά αιτήματα. Κατορθώσαμε να βελτιώσουμε λίγο παραπάνω αυτό το ποσοστό της ακρίβειας, χρησιμοποιώντας πολύπλοκα στατιστικά μοντέλα, με τα οποία επεξεργαζόμαστε όλες αυτές τις πηγές των πληροφοριών. Οι περισσότεροί μας σε αυτήν τη δραστηριότητα αντιλαμβανόμαστε, όμως, ότι *πέφτουμε πάνω σε ένα τοίχο* ως προς τις δυνατότητες, που τα συστήματά μας μπορούν να έχουν.²³

Αυτό το ποσοστό της ακρίβειας, περίπου στο 30%, που μπορεί κανείς να ελπίζει να του δώσουν οι μηχανές αναζήτησης, δεν θα πρέπει να δημιουργεί καμία έκπληξη. Έχουμε δει ότι δεν μπορεί να υπάρξει καμία κατανόηση της συνάφειας, χωρίς να υπάρχει η αντίληψη της κοινής λογικής, και δεν μπορεί να υπάρξει καμία αντίληψη της κοινής λογικής, χωρίς να υπάρχει η αίσθηση του πώς ο κόσμος αλληλοσυνδέεται με τη σωματικότητά μας. Όλες οι τεχνικές αναζήτησης στον Ιστό είναι καταδικασμένες σε αποτυχία εκ των προτέρων, επειδή προσπαθούν να προσεγγίσουν τις ανθρώπινες αντιλήψεις του νοήματος και της συνάφειας, αντιλήψεις που βασίζονται στο σώμα, χωρίς να έχουν ένα συγκεκριμένο σώμα να κάνει αυτή τη δουλειά και, επομένως, χωρίς να μπορούν να χρησιμοποιήσουν την κοινή λογική. Επομένως, οι μηχανές αναζήτησης, από πρακτικής πλευράς, δεν μπορούν να διεκπεραιώσουν τους στόχους της μηχανογραφημένης ανάκτησης ακριβώς εκείνων των πληροφοριών, που είναι σχετικές με τα ενδιαφέροντα των χρηστών· δεν μπορούν καθόλου να οδηγήσουν στην αναζήτηση της συνάφειας.

Ο Don Swanson συνοψίζει λακωνικά την ουσία των πραγμάτων ως εξής:

Η συνεπής, αποτελεσματική και πλήρως αυτοματοποιημένη αρχειοθέτηση και ανάκτηση των πληροφοριών δεν είναι δυνατή. Οι κρίσεις μας για τη συνάφεια ... εξαρτώνται από τη γνώση, που έχουμε για το ποιοι ή τι είμαστε, για το είδος του κόσμου, στον οποίο ζούμε, και για το γιατί θέλουμε αυτό που αναζητούμε. Είναι δύσκολο να φανταστούμε ότι ένας τέτοιος μηχανισμός, αν θα υπήρχε, ... θα μπορούσε να αποκτήσει μια τέτοια αυτογνωσία ή να κάνει τη δουλειά χωρίς αυτήν την αυτογνωσία.²⁴

Συνεπώς, στον κυβερνοχώρο, χωρίς την ενσώματη ικανότητά μας να πιάνουμε το νόημα, η συνάφεια γλιστράει μέσα από τα άυλα δάκτυλά μας. Αλλά, πώς τότε οι άνθρωποι μπορούν ποτέ να βρουν τι είναι σχετικό με τα ενδιαφέροντά τους; Η βασική ιδέα αυτού του κεφαλαίου ήταν ότι, για τους ανθρώπους, ο κόσμος δεν είναι μια σκέτη συλλογή από δισεκατομμύρια γεγονότα, μια συλλογή χωρίς νόημα. Αλλά αντίθετα είναι ένα πεδίο σημασίας, το οποίο οργανώνεται από και για ανθρώπους, που έχουν σώματα, επιθυμίες, ενδιαφέροντα και σκοπούς. Όχι ότι η ιδέα αυτή λύνει το μυστήριο του πώς ο εγκέφαλος των ανθρώπων καταφέρνει να συντονισθεί με ότι, σε οποιαδήποτε στιγμή, είναι σχετικό με αυτούς. Αλλά, έτσι, τουλάχιστον μπορούμε να διαπιστώσουμε ότι, αφού ο κόσμος είναι οργανωμένος από και για ενσώματους

ενεργούς πράκτορες και όχι από και για ασώματους υπολογιστές, οι άνθρωποι έχουν ένα τεράστιο προβάδισμα, απέναντι στις μηχανές, για να τον κατανοήσουν και να βρουν τις πληροφορίες που αναζητούν. Ένα πράγμα είναι σίγουρο, καθώς ο Ιστός και το Internet αυξάνονται: οι χρήστες του Δικτύου, οι οποίοι μπαίνουν στον κυβερνοχώρο αφήνοντας τα σώματά τους πίσω και γίνονται εξαρτημένοι από τους συντακτικούς 'εξερευνητές' του Ιστού (Web crawlers) και τις μηχανές αναζήτησης, θα πρέπει μάλλον να παραιτηθούν από το άχαρο έργο του ψαξίματος μέσα σε τεράστιους σωρούς σκουπιδιών, για να βρουν την πληροφορία που αναζητούν.