

Πανεπιστήμιο Μακεδονίας

University of Macedonia

ΔΠΜΣ Πληροφοριακά Συστήματα

Master Information Systems

Δίκτυα Υπολογιστών

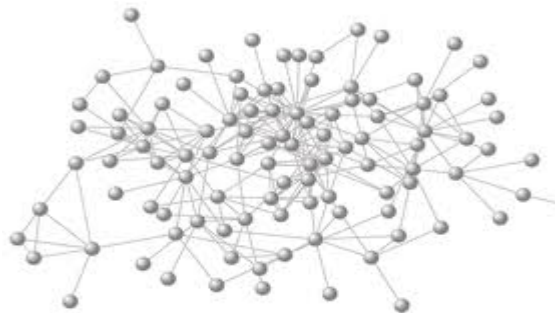
Computer Networks

Καθηγητής: Α.Α. Οικονομίδης

Professor: A.A. Economides



SEMANTIC OVERLAY NETWORKS
FOR P2P SYSTEMS



Παρασίδου Δόμνα

Θεσσαλονίκη, Ιανουάριος 2013

Περιεχόμενα

1. Περίληψη - Abstract	σελ. 3
2. Εισαγωγή	σελ. 5
3. P2P Systems	σελ. 7
4. SONs	σελ. 8
5. Κίνητρα για τη δημιουργία ενός Semantic Overlay Network	σελ. 13
6. Δημιουργία και χρήση των SONs	σελ. 14
7. Η descent προσέγγιση	σελ. 16
8. Ταξινόμηση εγγράφων και ερωτημάτων	σελ. 17
9. Κριτήρια καλής κατηγοριοποίησης και πηγές λαθών	σελ. 18
10. Στρατηγικές κατάταξης των κόμβων	σελ. 19
11. Layered SONs	σελ. 20
12. Πειράματα και συμπεράσματα	σελ. 25
13. Βιβλιογραφία	σελ. 29

Περίληψη

Σε ένα Peer-to-Peer δίκτυο οι κόμβοι συνήθως συνδέονται σε ένα μικρό σύνολο τυχαίων γειτονικών κόμβων αυτών, και τα ερωτήματα μεταδίδονται κατά μήκος αυτών των συνδέσεων. Με αυτό τον τρόπο όμως δημιουργείται καταγισμός ερωτημάτων στο σύστημα, που αφενός δημιουργεί κόστος, αφετέρου κάνει το σύστημα αργό στην ανταπόκριση σχετικά με μια ερώτηση συγκεκριμένου περιεχομένου. Έτσι προτείνεται οι συνδέσεις των κόμβων να επηρεάζονται από το περιεχόμενό τους, άρα σημασιολογικά σχετικοί κόμβοι σχηματίζουν ένα σημασιολογικά επικαλυπτόμενο δίκτυο (Semantic Overlay Network - SON). Με αυτό τον τρόπο, τα ερωτήματα δρομολογούνται στο κατάλληλο SON, αυξάνοντας τις πιθανότητες τα αρχεία που ταιριάζουν να βρεθούν γρήγορα και μειώνοντας το φόρτο αναζήτησης στους κόμβους που δεν έχουν σχετικό περιεχόμενο. Από σχετικές εκτιμήσεις που έχουν γίνει φαίνεται ότι τα SON μπορούν να βελτιώσουν σημαντικά τις επιδόσεις των ερωτημάτων επιτρέποντας ταυτόχρονα στους χρήστες να αποφασίζουν σχετικά με το περιεχόμενο αλλά και με ποιους θα συνδεθούν.

Abstract

In a peer-to-peer system nodes are usually connected to a small set of random nodes, and queries communicate along these connections. This way creates flooding of queries that on the one hand is quite expensive, on the other hand doesn't help the system to correspond quickly to a question of a certain content. Thus, is proposed a system by which the connection of the nodes is affected by their content, so semantically relative nodes create a semantic overlay network (SON). This way queries are routed to the appropriate SON, fact

that increases the chances that matching files will be found quickly and reduces the load on nodes that have irrelevant content. It seems that SONS can improve semantically the query performance while allowing the users to decide about the content and to whom to be connected.

Εισαγωγή

Η χρήση των συστημάτων Peer-to-Peer (P2P) έχει αυξηθεί πολύ, παρέχοντας τη δυνατότητα κάποιος να μπορεί να αποθηκεύσει και να μοιραστεί με χαμηλό κόστος πληροφορίες και δεδομένα, διατηρώντας παράλληλα την αυτονομία του. Παρ' όλα αυτά, υπάρχουν μειονεκτήματα στην επεξεργασία ερωτημάτων στα σημερινά P2P συστήματα, όπως η αναποτελεσματικότητα και η μη σωστή κλιμάκωση τους. Η αναποτελεσματικότητα οφείλεται στο ότι τα περισσότερα P2P συστήματα δημιουργούν ένα τυχαίο δίκτυο επικάλυσης όπου τα ερωτήματα αποστέλλονται από κόμβο σε κόμβο χωρίς κάποιο κριτήριο ή κάποια διάκριση. Εναλλακτικά υπάρχουν προτάσεις για "άκαμπτα" P2P συστήματα που βάζουν το περιεχόμενο σε κόμβους, διαχωρίζοντας το, βασιζόμενα σε συναρτήσεις κατακερματισμού, κάνοντας πιο εύκολο τον εντοπισμό του περιεχομένου αργότερα. Αυτά τα συστήματα έχουν καλή απόδοση για τα ερωτήματα των οποίων η αναζήτηση είναι γνωστή, αλλά δεν είναι τόσο αποτελεσματικά για προσεγγίσεις, ερωτήματα κειμένων κ.ά. (Arturo Crespo, Hector Garcia-Molina, 2003)

Τα SONs (Semantic Overlay Networks) αποτελούν μια προσέγγιση που οργανώνει τους ομότιμους σε θεματικές ομάδες. Ουσιαστικά, είναι μια ευέλικτη οργάνωση του δικτύου που βελτιώνει την απόδοση των ερωτημάτων, ενώ διατηρεί υψηλό βαθμό αυτονομίας του κόμβου. Με τα SONs οι κόμβοι με παρόμοιο περιεχόμενο "ενώνονται". Αξιοσημείωτο είναι ότι δεν μπορούμε να επιβάλλουμε το πώς θα γίνουν οι συνδέσεις σε ένα SON. Ακόμη, ένας κόμβος μπορεί να ανήκει σε περισσότερα από ένα SON. Σε αυτά τα συστήματα τα ερωτήματα υποβάλλονται σε επεξεργασία αναγνωρίζοντας ποιά ή ποιά SON ταιριάζουν καλύτερα στην απάντησή τους, και αναλόγως προωθούνται μόνο προς αυτά, ελαχιστοποιώντας έτσι το χρόνο που απαιτείται για να δοθεί η απάντηση. Οι υπόλοιποι

κόμβοι δεν ασχολούνται με το συγκεκριμένο ερώτημα, ελευθερώνοντας έτσι πόρους που μπορούν να χρησιμοποιηθούν στη βελτίωση της απόδοσης άλλων ερωτημάτων. (Arturo Crespo, Hector Garcia-Molina, 2003), (Xuemin Shen, Heather Yu, John Buford, Muralin Akon, 2010)

Υπάρχουν πολλά κριτήρια για τη δημιουργία ενός SON, πρέπει να μπορούμε να ταξινομούμε ερωτήματα και κόμβους, να αποφασίζουμε το επίπεδο διάκρισης για την ταξινόμηση, να αποφασίζουμε πότε ένας κόμβος θα ενταχθεί σε ένα SON καθώς και ποιό SON να χρησιμοποιήσουμε για να απαντήσουμε σε ένα ερώτημα.

Σκοπός της εργασίας είναι να παρουσιάσει τα SONs και τη χρησιμότητα τους, παραθέτοντας παράλληλα τη σπουδαιότητα της καλής κατηγοριοποίησης των ερωτημάτων, καθώς αυτά είναι που καθορίζουν τις ιδιότητες των SONs.

P2P Systems

Ένα δίκτυο υπολογιστών P2P είναι ένα δίκτυο που επιτρέπει σε δύο ή περισσότερους υπολογιστές να μοιράζονται τους πόρους τους ισοδύναμα. Το δίκτυο αυτό χρησιμοποιεί την επεξεργαστική ισχύ, τον αποθηκευτικό χώρο και το εύρος ζώνης των κόμβων, οι οποίοι όλοι έχουν ίσα δικαιώματα. Πληροφορίες που βρίσκονται στον ένα κόμβο, ανάλογα με τα δικαιώματα που καθορίζονται, μπορούν να διαβαστούν από όλους τους άλλους και αντίστροφα. (Wikipedia. (2012). *P2P*. Retrieved from <http://www.easypedia.gr/el/articles/p/e/e/Peer-to-peer.html>).

Τα P2P συστήματα βασίζονται στην ακόλουθη λογική : αντί να βασίζονται στους server, οι υπηρεσίες λειτουργούν από αποκεντρωμένες αρχιτεκτονικές επικάλυψης όπου οι τελικοί χρήστες συνδέονται δυναμικά μεταξύ τους. Διαχωρίζονται σε δύο μεγάλες κατηγορίες, τα δομημένα και τα αδόμητα. Στα αδόμητα P2P συστήματα κάθε κόμβος (ομότιμος), περιέχει έναν περιορισμένο-συγκεκριμένο αριθμό συνδέσεων με άλλους γειτονικούς κόμβους του δικτύου. Η αναζήτηση σε ένα τέτοιο αδόμητο P2P περιβάλλον οδηγεί συνήθως είτε σε καταιγισμό ερωτημάτων στο δίκτυο, είτε στην προώθηση των ερωτημάτων με βάση κατασκευασμένους δείκτες δρομολόγησης που δίνουν "κατεύθυνση" στην έρευνα. Στα δομημένα P2P συστήματα, μια συνάρτηση κατακερματισμού χρησιμοποιείται για να "ενώσει" το ερώτημα με τις πιθανές απαντήσεις. (Arturo Crespo, Hector Garcia-Molina, 2003)

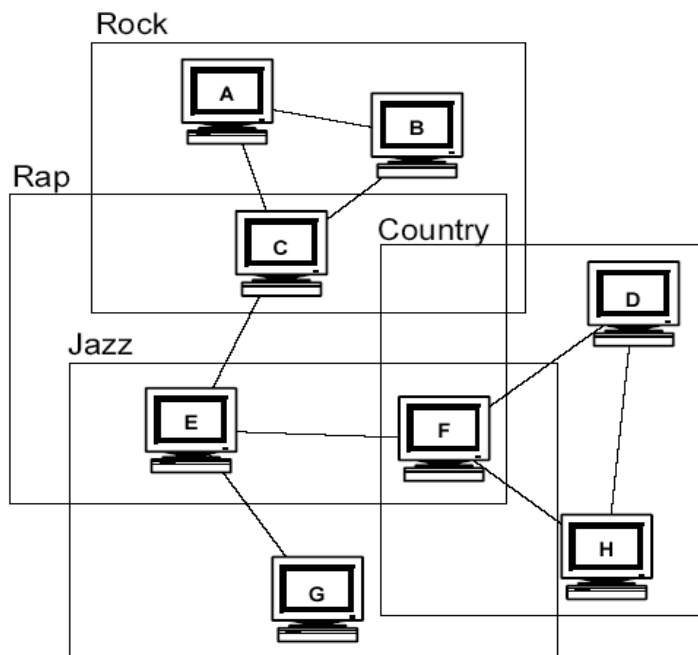
SONs

Τα τελευταία χρόνια τα συστήματα P2P έχουν γίνει αναπόσπαστο μέρος του διαδικτύου. Μετά από επιτυχίες που σχετίζονται με τη μουσική, όπου οι χρήστες μπορούν να μοιράζονται τραγούδια, άλμπουμ, φακέλους κτλ, (site όπως : Napster, Gnutella) αναπτύχθηκαν πολύ γρήγορα και έχουν εδραιωθεί σε διάφορες περιπτώσεις, από τη διανομή μεγάλης κλίμακας περιεχομένου (BitTorrent) έως την τηλεφωνία μέσω ίντερνετ (Skype). Πολλές έρευνες έχουν επικεντρωθεί στον εμπλουτισμό των P2P συστημάτων με πιο εκφραστικά μοντέλα δεδομένων και γλώσσες αιτημάτων-ερωτημάτων. Έτσι εμφανίστηκαν τα Semantic Overlay Networks, που στόχο έχουν τη διαχείριση δομημένων δεδομένων μεγάλης κλίμακας, αποκεντρωμένων σε πολύ δυναμικά περιβάλλοντα. Είναι ένας τρόπος, δηλαδή, για την οργάνωση του περιεχομένου σε ένα P2P δίκτυο, που ενώνει σε θεματικές ομάδες τους ομότιμους με παρόμοιο περιεχόμενο, ώστε τα ερωτήματα να προωθούνται επιλεκτικά μόνο σε κόμβους που σχετίζονται με το συγκεκριμένο θέμα. Στην περίπτωση των αδόμητων P2P συστημάτων, τα SONs βοηθούν στην αποτελεσματικότερη δρομολόγηση των ερωτημάτων σε συγκεκριμένες ομάδες ομότιμων, αντί να τα προωθούν "τυφλά". Παρότι αυτό το πρόβλημα είναι πιο ήπιο στα δομημένα P2P συστήματα, τα SONs χρησιμοποιούνται και σε αυτά αυξάνοντας την απόδοση της έρευνας. Χαρακτηριστικό τους είναι ότι ένας ομότιμος μπορεί να ανήκει παράλληλα σε περισσότερα από ένα SONs, καθώς μπορεί να έχει δεδομένα που ανήκουν σε διάφορες θεματικές κατηγορίες. (Arturo Crespo, Hector Garcia-Molina, 2003), (Josiane Xavier Parreira, Sebastian Michel, Gerhard Weikum), (Σταματελάτος,2011)

Κεντρική Ιδέα

Η κύριος σκοπός των SONS είναι η δημιουργία μιας ευέλικτης οργάνωσης του δικτύου που βελτιώνει την επίδοση των αιτημάτων-ερωτημάτων, και βασίζεται στις σημασιολογικές σχέσεις μεταξύ ομότιμων. Στηρίζονται στη λογική του ότι θα είναι αποδοτικότερο η δρομολόγηση των ερωτημάτων να γίνεται μόνο προς τους κόμβους που είναι πιο πιθανό να έχουν απαντήσεις. Βασική αρχή των SONS είναι το ότι οι ομότιμοι είναι συγκεντρωμένοι σύμφωνα με το περιεχόμενο τους σε ομάδες, οι οποίες επικαλύπτονται γιατί οι ομότιμοι μπορεί να περιέχουν διαφορετικό περιεχόμενο και να ανήκουν σε διαφορετικές ομάδες. Έτσι, ομότιμοι με παρόμοιο περιεχόμενο ενώνονται και "σχηματίζουν" ομάδες με βάση αυτό, βοηθώντας στην αποτελεσματικότερη ανάκτηση πληροφοριών καθώς τα ερωτήματα απευθύνονται μόνο στις σχετικές με αυτά ομάδες. (Arturo Crespo, Hector Garcia-Molina, 2003)

Για παράδειγμα στην εικόνα που ακολουθεί κάποιοι ομότιμοι μοιράζονται αρχεία μουσικής, που είναι διαφορετικό στυλ και ανάλογα με το στυλ χωρίζονται σε ομάδες. Έτσι όταν κάποιο αίτημα-ερώτημα τεθεί θα κατευθυνθεί μόνο στις ομάδες που είναι σχετικές και οι υπόλοιποι δεν θα λάβουν κανένα μήνυμα.



Σχήμα 1

Για ένα τέτοιο δίκτυο πρέπει να δημιουργηθεί ένα σύστημα εναλλακτικών δικτύων, με βάση το υπάρχον φυσικό δίκτυο. Το "νέο" δίκτυο θα έχει μόνο ομότιμους που θα συσχετίζονται μεταξύ τους και θα κάνουν μια ομάδα-σύμπλεγμα ομότιμων. Αυτό το δίκτυο (SON) έχει το ρόλο του διαμεσολαβητή μεταξύ των ερωτημάτων και των συγκεκριμένων ομότιμων και είναι αρμόδιο για να κατανοήσει το ερώτημα και να το προωθήσει στο αντίστοιχο τμήμα της ιεραρχίας προς απάντηση. Αν δεν βρεθεί απάντηση ή δεν βρεθούν όλα τα δεδομένα, τότε το ερώτημα προωθείται σε ανώτερα επίπεδα. . (Arturo Crespo, Hector Garcia-Molina, 2003) (Sungsu Kim, John Strassner, James Won-Ki Hong, 2011)

Τα SONs σχετίζονται με την ιδέα των online κοινοτήτων, όπως τα Yahoo Groups και τις MSN Communities. Σε μια online κοινότητα χρήστες με κοινά ενδιαφέροντα εντάσσονται σε γκρουπ και μοιράζονται πληροφορίες, αρχεία κτλ. Ωστόσο οι περισσότερες online κοινότητες έχουν ένα κεντρικό στοιχείο που συντονίζει και ελέγχει τις δραστηριότητες των μελών

Αρχική Δόμηση του SON

Ένα SON μπορεί να αναπαρασταθεί ως ένα σύνολο συνδέσμων ή τριάδων (n_i, n_j, L) όπου το n_i και το n_j είναι συνδεδεμένοι ομότιμοι και L είναι η έννοια τους, ο λόγος για τον οποίο δηλαδή οι ομότιμοι συνδέονται (το κοινό τους χαρακτηριστικό). Στο παράδειγμα της παραπάνω εικόνας ένας ομότιμος μπορεί να συνδεθεί με "γείτονες" του σε μια ομάδα Rock ή σε μια δυνητικά διαφορετική ομάδα ομότιμων Jazz ή Rap.

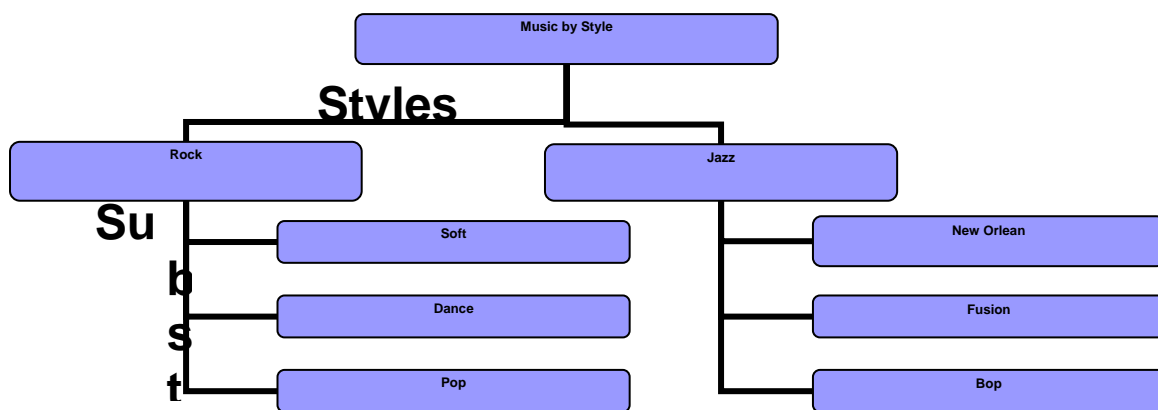
Ένα SON δίκτυο επικάλυψης πρέπει να υποστηρίζει τις ακόλουθες λειτουργίες :

- join-συμμετοχή, (n_i) όπου ένας ή περισσότεροι σύνδεσμοι της μορφής (n_i, n_j, L) δημιουργούνται,
- search-αναζήτηση, (q) όπου επιστρέφει ένα σύνολο ομότιμων που ταιριάζουν με την αναζήτηση (q) , και
- leave όπου βάζουμε όλους τους συνδέσμους στο SON που περιλαμβάνει το n_i . Η υλοποίηση αυτών των τριών λειτουργιών διαφέρει ανάλογα με το σύστημα. (Arturo Crespo, Hector Garcia-Molina, 2003)

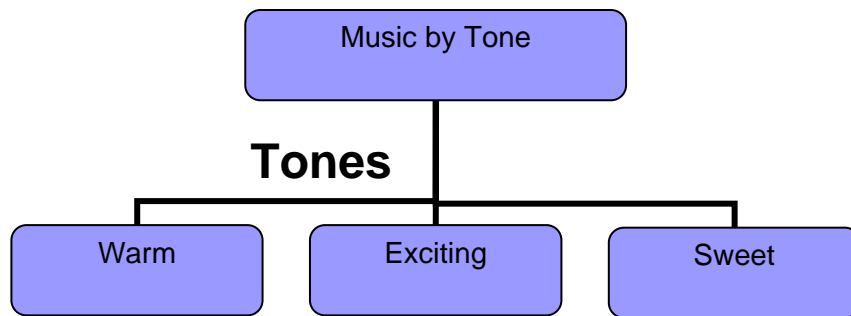
Ιεραρχική Δομή του SON

Ένας άλλος ορισμός των SON ενσωματώνει την έννοια της ιεραρχίας καθώς τα SONs συνδέονται με την έννοια της ιεραρχικής κατηγοριοποίησης. Στο σχήμα που ακολουθεί

υπάρχουν δύο πιθανές ιεραρχίες ταξινόμησης για αρχεία μουσικής. Αρχικά χωρίζονται ανάλογα με το στυλ και τις υποκατηγορίες του, και έπειτα ανάλογα με το ρυθμό που έχουν. Έτσι, σύμφωνα με την πρώτη ιεραρχία (Σχήμα 2) ορίζουμε 9 SONS με βάση το στυλ ή 4 με βάση το ρυθμό, δεύτερη ιεραρχία (Σχήμα 3). Συμπεραίνουμε ότι τα αρχεία ενός κόμβου πρέπει να συσχετίζονται με έννοιες, έτσι ώστε ο κόμβος να μπορεί να εισαχθεί στα αντίστοιχα SONS. (Arturo Crespo, Hector Garcia-Molina, 2003), (Carmela Comito, Simon Patarin, Domenico Talia), (Tao Gu, Daqing Zhang, Hung Keng Pung)



Σχήμα 2



Σχήμα 3

Κίνητρα για τη Δημιουργία ενός Semantic Overlay Network

Η αποστολή των αιτημάτων-ερωτημάτων σε όλους τους κόμβους δεν ευνοεί την κλιμάκωση. Τα SONs δημιουργήθηκαν με τη λογική του ότι είναι καλύτερο να στέλνουμε τις ερωτήσεις μόνο σε κόμβους που είναι πιο πιθανό να βρούμε απαντήσεις. Έτσι και απαντώνται γρηγορότερα, αλλά και αποφεύγεται η σπατάλη πόρων επιτρέποντας σε άλλα ερωτήματα να υποβάλλονται σε επεξεργασία συντομότερα.

Με αυτόν τον τρόπο τα ερωτήματα μπορούν να δρομολογηθούν μόνο σε σημασιολογικά επιλεγμένα υποσύνολα ομότιμων, που είναι ικανά να απαντήσουν στο ερώτημα. Εάν ένας ομότιμος δεν μπορεί να δώσει πλήρη απάντηση, προωθεί το ερώτημα μόνο σε γείτονες του, που μπορεί να έχουν την απάντηση κ.ο.κ. Έτσι καταφέρνουμε να μειωθεί ο καταγισμός ερωτημάτων. (Pavel Serdyukov Saarland University, Department of Computer Science)

Επιπρόσθετα, τα δεδομένα που μοιράζονται στα P2P συστήματα έχουν συχνά έντονη οντολογική δομή, εξ αιτίας της καταγωγής τους και των σχέσεων της με τις πραγματικές έννοιες στον κόσμο (πχ μουσική), και είναι πιθανό να ταξινομήσουν τα δεδομένα σε τμήματα, διαχωρίζοντας με κάποιο τρόπο το περιεχόμενό τους και δημιουργώντας σημασιολογικά όμοιες ομάδες. (Kamal Jain, László Louász, Philip A. Chou), (Karl Aberer, Philippe Cudr'e-Mauroux, Manfred Hauswirth, Tim Van Pelt)

Δημιουργία και Χρήση των SONs

Σε ένα SON τα ερωτήματα υποβάλλονται σε επεξεργασία ανάλογα με το ποιό/ά SON είναι κατάλληλα για να τα απαντήσουν. Έπειτα το ερώτημα στέλνεται σε έναν κόμβο αυτού του SON και διαβιβάζεται μόνο στα άλλα μέλη του συγκεκριμένου SON.

Παρά το ότι η ταξινόμηση εγγράφων και ερωτημάτων έχουν κοινά χαρακτηριστικά, διαφέρουν αρκετά στην ακρίβεια και τον χρόνο που απαιτεί ο αλγόριθμος που θα κάνει την ταξινόμηση τους.

Οι κύριες απαιτήσεις για τη δημιουργία ενός SON σε ένα δυναμικό περιβάλλον P2P είναι η προσέγγιση χωρίς επίβλεψη για την ομαδοποίηση, η επεκτασιμότητα, η αυτό-οργάνωση, η αυτονομία και η αποκέντρωση. Όλα αυτά βοηθούν στην αύξηση της αξίας, αλλά και του οφέλους ενός SON. (Michalis Vazirgiannis, , Kjetil Nørn'ag, Christos Doulkeridis)

Για να δημιουργηθεί ένα SON αρχικά θα πρέπει να καθοριστούν και να αξιολογηθούν τα κριτήρια ταξινόμησης που θα χρησιμοποιηθούν στο σύστημα. Αξιολογούμε τις πιθανές ιεραρχίες της κατάταξης χρησιμοποιώντας την πραγματική κατανομή των δεδομένων στους κόμβους και καταλήγουμε στη σωστή ιεραρχία. Η ιεραρχία

αυτή θα χρησιμοποιηθεί από τους κόμβους στο σύστημα και θα καθορίσει τα SONs . Κατά την είσοδο ενός από τους κόμβους στο σύστημα, ο κόμβος κατακλύζει το δίκτυο με ερωτήματα σχετικά με την ιεραρχία, "τρέχει" έναν ταξινομητή εγγράφων που βασίζεται στην ιεραρχία που έχει δημιουργηθεί λαμβάνοντας υπόψη όλα τα έγγραφα του και τέλος ο ταξινομητής βάζει τον κόμβο σε ένα συγκεκριμένο SON. Κάθε κόμβος μπαίνει σε ένα ξεχωριστό SON βρίσκοντας άλλους που ανήκουν στο συγκεκριμένο SON.

Για την απάντηση των ερωτημάτων, αφού ο κόμβος έχει εισάγει το ερώτημα (το οποίο έχει δοθεί από κάποιο χρήστη), το ταξινομεί και το προωθεί σε κάποια από τα SONs όπου μπορεί να υπάρχει η απάντηση και προσπαθεί να τη βρει σε αυτά χρησιμοποιώντας μηχανισμούς μετάδοσης. (Philippe Cudré-Mauroux, Karl Aberer, 2007)

Τα επιθυμητά χαρακτηριστικά της δημιουργίας ενός SON εξετάζονται ώστε να μπορούμε να αξιολογούμε την ποιότητα τους. Δύο είναι τα βασικά χαρακτηριστικά που βοηθούν στην αξιολόγηση : η intra- SON similarity και η inter- SON similarity. Η ποιότητα τους αξιολογείται ανεξάρτητα από τα άλλα SONs, μετρώντας το βαθμό ομοιότητας κάθε ζεύγους ομότιμων που ανήκουν στο SON. Στην intra- SON similarity, υψηλός βαθμός δείχνει ότι οι ομότιμοι περιέχουν σχετικό περιεχόμενο, άρα και η ποιότητα του SON είναι υψηλή. Αντίθετα, χαμηλός βαθμός δείχνει ότι οι ομότιμοι έχουν διαφορετικό περιεχόμενο, άρα η ποιότητα του SON δεν είναι αρκετά υψηλή. Στην inter- SON similarity, η ομοιότητα κάποιων SONs δείχνει πόσο όμοια είναι μεταξύ τους. Υψηλός βαθμός δείχνει ότι αυτά έχουν παρόμοιο περιεχόμενο και δεν είναι εύκολο να διακριθούν, ενώ χαμηλός βαθμός δείχνει ότι τα SONs είναι καλά διαχωρισμένα όσο αναφορά το περιεχόμενο τους, έτσι ώστε κάθε ερώτημα να μπορεί να απαντηθεί μόνο από κάποια (λίγα) συγκεκριμένα SONs. (Tao Gu , Hung Keng Pung , Daqing Zhang), (Xuemin Shen, Heather Yu, John Buford, Muralin

Akon, 2010), (Josiane Xavier Parreira, Sebastian Michel, Gerhard Weikum), (Heng Tao Shen, Yan Feng Shu, Bei Yu)

Η DESENT προσέγγιση

Η προσέγγιση DESENT βασίζεται στη δημιουργία τοπικών ομάδων ομότιμων, που λέγονται ζώνες, που σχηματίζουν συμπλέγματα-ομάδες, που βασίζονται στα δεδομένα που είναι αποθηκευμένα στους ομότιμους, και έπειτα συγχωνεύουν τις ζώνες και τις ομάδες κατ' επανάληψη, μέχρι να γίνουν παγκόσμιες ζώνες και ομάδες.

Κάποιοι αρχικοί ομότιμοι επιλέγονται με έναν τυχαίο τρόπο, δημιουργούν τοπικές ζώνες με τους γειτονικούς ομότιμους, και κάθε ένας συλλέγει την περιγραφή της ομάδας όλων των ομότιμων που είναι στη ζώνη αυτή και εκτελεί έναν αλγόριθμο ομαδοποίησης έτσι ώστε να δημιουργηθούν νέες ομάδες που καλύπτουν όλο το σύνολο της ζώνης. Όταν οι ομάδες δύο ή περισσότερων ομότιμων συγχωνευτούν σε μία νέα ομάδα, σημαίνει ότι αυτοί οι ομότιμοι γίνονται μέλη ενός SON, του οποίου τα περιεχόμενα είναι η περιγραφή μιας νέας ομάδας. Έπειτα, οι αρχικοί δημιουργούν ένα νέο αδόμητο P2P δίκτυο που παίζει το ρόλο των ομότιμων στο αρχικό στάδιο. Η διαδικασία ακολουθείται μέχρι να επιτευχθούν παγκόσμιες ομάδες. Προκειμένου να μπορούν να δημιουργηθούν ζώνες περίπου ίδιου μεγέθους στο δίκτυο, ένα υποσύνολο ομότιμων ορίζει το ρόλο της αρχικής ζώνης που μπορεί να πραγματοποιήσει τη δημιουργία καινούριας ζώνης, και κατά συνέπεια να εισάγει και να ελέγξει την πορεία της ομάδας μέσα σε κάθε ζώνη. Η διαδικασία επιλογής αρχικών είναι τελείως μοιρασμένη και ουσιαστικά βασίζεται σε κάθε ομότιμο που αναθέτει ή όχι στον

εαυτό του το ρόλο, χρησιμοποιώντας μια λειτουργία που βασίζεται στο συνδυασμό προσδιορισμού και χρόνου. (Christos Doulkeridis, Akrivi Vlachou, Kjetil Nørvag, Michalis Vazirgiannis)

Χαρακτηριστικό του αλγορίθμου αρχικής επιλογής είναι ότι εξασφαλίζεται διαφορετική αρχική τιμή κάθε φορά που τρέχει ο αλγόριθμος και το ότι όλοι οι ομότιμοι του δικτύου επικοινωνούν, όσο παραμένουν συνδεδεμένοι σε αυτό. (Arturo Crespo, Hector Garcia-Molina, 2003)

Ταξινόμηση εγγράφων και ερωτημάτων

Παρ' ότι η ταξινόμηση ερωτημάτων και εγγράφων είναι παρόμοια, οι απαιτήσεις για τους αντίστοιχους ταξινομητές μπορεί να διαφέρουν πολύ. Ειδικότερα, είναι λογικό το γεγονός ότι οι κόμβοι θα μπαίνουν σε ένα P2P δίκτυο με αργούς ρυθμούς, (κάποιοι το λεπτό), σε αντίθεση με τα ερωτήματα που η αναλογία τους στο λεπτό θα είναι αρκετά υψηλότερη (εκατοντάδες το δευτερόλεπτο). Αντίστοιχα, η ταξινόμηση των κόμβων μπορεί να είναι πιο "εκρηκτική", καθώς ένας κόμβος μπορεί να περιέχει εκατοντάδες αρχεία προς ταξινόμηση, ενώ των ερωτημάτων αισθητά λιγότερο. Έτσι ο ταξινομητής εγγράφων θα πρέπει να χρησιμοποιήσει έναν αλγόριθμο πολύ ακριβή, παρ' ότι χρονοβόρο, που θα μπορεί να επεξεργάζεται μεγάλο αριθμό εγγράφων, σε αντίθεση με τον ταξινομητή ερωτημάτων που θα πρέπει να κάνει χρήση ενός γρήγορου και συνάμα λιγότερο ακριβή αλγορίθμου.

Η ταξινόμηση (εγγράφων και ερωτημάτων) μπορεί να γίνει αυτόματα, με το χέρι ή με υβριδικές διαδικασίες. (Arturo Crespo, Hector Garcia-Molina, 2003) (Alexander Loser, Steffen Staab, Christoph Tempich, 2007)

Κριτήρια καλής κατηγοριοποίησης και πηγές λαθών

Η ταξινόμηση είναι πολύ σημαντική και καθορίζει τις ιδιότητες των SONs. Για να είναι σωστή και αποδοτική η ταξινόμηση θα πρέπει :

α)τα SONs να συνδέονται με λίγους κόμβους, καθώς όσο πιο λίγους κόμβους πρέπει να ψάξουμε, τόσο πιο αποδοτικά και γρήγορα δίνεται απάντηση στο ερώτημα που έχει δοθεί.

β)αντίστροφα και οι κόμβοι να συνδέονται με λίγα SONs, καθώς όσο περισσότερα είναι τα SONs, τόσο μεγαλύτερο είναι το κόστος για έναν κόμβο ώστε να τα παρακολουθεί όλα.

γ)και τέλος, να χρησιμοποιούνται εύκολοι στην εφαρμογή αλγόριθμοι ταξινόμησης που κάνουν λίγα λάθη.

Στην πραγματικότητα υπάρχουν πολλές πηγές λαθών κατά τη χρήση ταξινομητή εγγράφων. Πρακτικά οι διαδικασίες ταξινόμησης μπορεί να μην είναι ακριβείς καθώς ενδέχεται να μην μπορεί να προσδιοριστεί με ακρίβεια που ανήκει (σε ποιά κατηγορία) ένα ερώτημα ή αρχείο κτλ. Αν δεν γίνει σωστά η ταξινόμηση μπορεί να μειωθούν, ή ακόμη και να εξαλειφθούν τα πλεονεκτήματα της χρήσης των SONs. Για να αξιολογήσουμε την ταξινόμηση πρέπει αρχικά να βεβαιωθούμε ότι αυτή και μπορεί να εφαρμοστεί αλλά και ότι είναι αποδοτική. Μετά χρησιμοποιώντας τα δεδομένα των κόμβων στο σύστημα

προβλέπουμε τον αριθμό των SONs, καθώς και τον αριθμό των SONs που ένας κόμβος μπορεί να ανήκει. (Prusty, 2009)

Με τη χρήση του ταξινομητή εγγράφων μπορεί να προκύψουν διάφορα λάθη. Πηγές λαθών θεωρούνται οι ακόλουθες : ο τύπος των αρχείων να μην ακολουθεί τα αναμενόμενα πρότυπα, τα ορθογραφικά λάθη κτλ.

Η ταξινόμηση θεωρείται λανθασμένη για ένα δεδομένο αρχείο ή έγγραφο, όταν ο ταξινομητής επιστρέφει μία ή περισσότερες υποκατηγορίες στις οποίες δεν ανήκει αυτό. Πάραυτα, είναι προτιμότερο για τους ομότιμους να ταξινομούνται τουλάχιστον στα κατάλληλα SONs, εκτός από τα ακατάλληλα, γιατί πολλά μη προσιτά δεδομένα μπορούν να κάνουν αυτή την προσέγγιση να μην είναι χρήσιμη. (Pavel Serdyukov Saarland University, Department of Computer Science)

Στρατηγικές κατάταξης των κόμβων

Κάποια από τα χαρακτηριστικά της καλής ταξινόμησης εξαρτώνται από τις στρατηγικές που ενώνουν σε ομάδες τους ομότιμους.

Η συντηρητική στρατηγική τοποθετεί ένα κόμβο σε ένα συγκεκριμένο SON αν έχει έστω και ένα αρχείο που κατηγοριοποιείται στο συγκεκριμένο χαρακτηριστικό-διακριτικό του SON. Αντίθετα, η επιθετική στρατηγική βάζει έναν κόμβο σε ένα συγκεκριμένο SON αν αυτός έχει αρκετά έγγραφα που ανήκουν στην κατηγορία του χαρακτηριστικού του SON. Έτσι ενώ στη συντηρητική στρατηγική δημιουργούνται πάρα πολλές συνδέσεις, στην επιθετική δεν είναι δυνατή η εύρεση όλων των εγγράφων. Παρ' όλα αυτά, η επιθετική στρατηγική και μειώνει τον αριθμό των κόμβων σε ένα SON και μειώνει τον αριθμό των

SON στα οποία ένας κόμβος μπορεί να ανήκει. Έτσι, αφ' ενός αυξάνονται τα πλεονεκτήματα των SON καθώς λιγότεροι κόμβοι θα πρέπει να απασχοληθούν με ένα ερώτημα, αφ' ετέρου μειώνεται το κόστος των SON καθώς σε όσο περισσότερα SON ανήκει ένας κόμβος τόσο μεγαλύτερη είναι η επιβάρυνση του για το χειρισμό πολλών διαφορετικών συνδέσεων. (Pavel Serdyukov Saarland University, Department of Computer Science)

Μετά την κατηγοριοποίηση των κόμβων στα αντίστοιχα SONs μπορούμε να προσαρμόσουμε τα τελευταία με βάση τις πραγματικές κατανομές δεδομένων στους κόμβους, όπως πχ αν ένα SON περιέχει πολύ λίγους κόμβους μπορούμε να το ενσωματώσουμε με κάποιο άλλο.

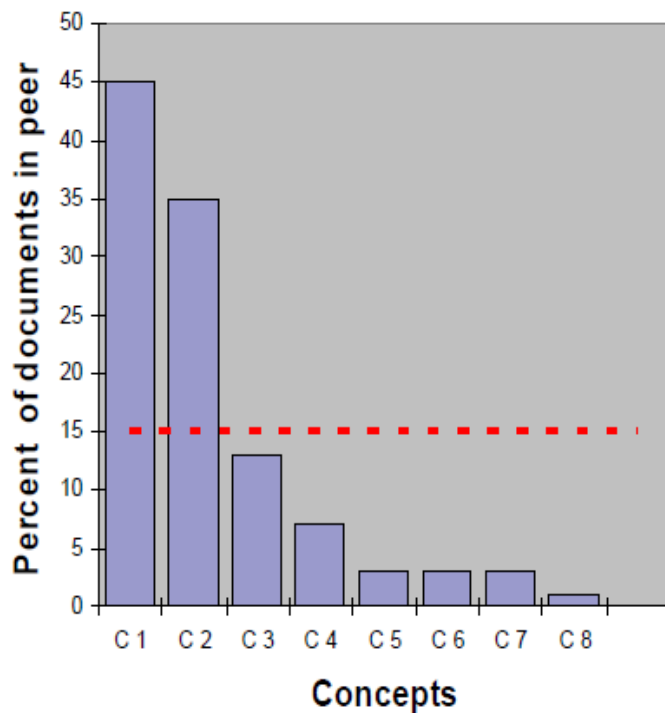
Κάποια λύση στα παραπάνω που προτείνεται από πολλούς επιστήμονες είναι η χρήση μίας πιο σύνθετης στρατηγικής, αυτής των πολύ-επίπεδων SONs - Layered SONs. (Kamal Jain, Laszl Lovouasz, Philip A. Chou), (Spyros Voulgaris, Maarten van Steen, Konrad Iwanicki, 2002)

Layered SONs

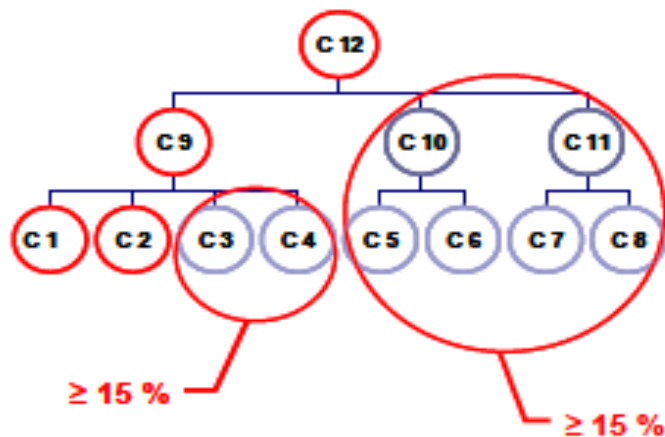
Μια όχι συντηρητική κατάταξη των κόμβων είναι τα Layered SONs (πολύ-επίπεδα SONs). Η προσέγγιση των Layered SONs εκμεταλλεύεται την πολύ συχνή διανομή δεδομένων zipfian σε συστήματα αποθήκευσης εγγράφων. Οι κόμβοι καθορίζουν σε ποιο SON θα ενταχθούν, βασιζόμενοι στον αριθμό των εγγράφων σε κάθε κατηγορία.

Μια παράμετρος της προσέγγισης των Layered SONs είναι το ελάχιστο ποσοστό που ένας κόμβος πρέπει να έχει σε μια κατηγορία για να ανήκει στο σχετικό SON. Η συντηρητική ανάθεση είναι ισοδύναμη με ένα Layered SON, όπου το όριο για να ενταχθεί στο SON έχει οριστεί στο 0%. Έτσι ο κόμβος θα ενωθεί με τα SONs που έχουν σχέση με τις βασικές έννοιες για τις οποίες έχει ένα ή περισσότερα έγγραφα.

Ένα πλεονέκτημα της χρήσης των Layered SONs είναι ότι μπορεί να μειώσει τον αριθμό των κόμβων σε ένα SON, κατά συνέπεια το μέγεθος του ίδιου του SON, σε αντίθεση με τη συντηρητική στρατηγική, η οποία μπορεί να οδηγήσει σε σπουδαίες βελτιώσεις όσο αναφορά την απόδοση του ερωτήματος. (Pavel Serdyukov Saarland University, Department of Computer Science), (Kamal Jain, Laszl Lovouasz, Philip A. Chou)



Σχήμα 4



Σχήμα 5

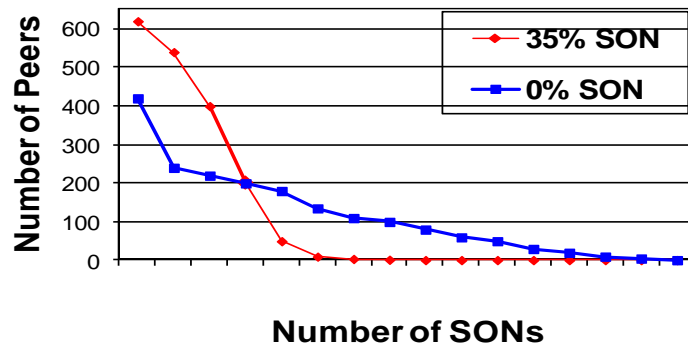
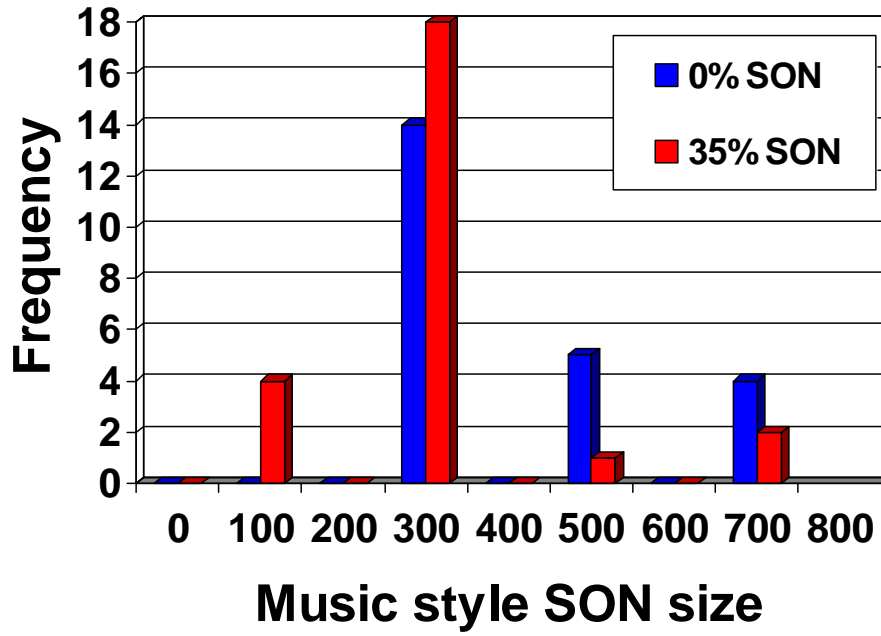
Στην παραπάνω εικόνα (Σχήμα 4) μπορούμε να δούμε την κατανομή και το εύρος (θεματικό) των περιεχομένων ενός ομότιμου. Εδώ, συγκεκριμένα, ο ομότιμος έχει αρχεία-έγγραφα οκτώ διαφορετικών κατηγοριών περιεχομένου. Μπορούμε υποθετικά να ορίσουμε ότι ένας ομότιμος μπορεί να μπει στο συγκεκριμένο SON, που έχει το C περιεχόμενο, εάν τουλάχιστον το 15% του περιεχομένου των αρχείων του ανήκουν στο θέμα C. Αυτή η υπόθεση όμως δεν μας επιτρέπει τα βρούμε τα έγγραφα που ενώ είναι στον κόμβο, δεν ανήκουν στις κατηγορίες C1 και C2. Για να λυθεί αυτό το πρόβλημα, μπορούμε να "δημιουργήσουμε ομάδες" που να πληρούν τα κριτήρια της αρχικής μας υπόθεσης, ενώνοντας το C3 και το C4 με άθροισμα μεγαλύτερο του 15% και εν συνεχεία τα C5, C6, C7 και C8. Έτσι, οι νέες "ομάδες" C9, C10 και C11 ανήκουν στο SON C12.

Η συντηρητική εκχώρηση ισοδυναμεί με τα πολύ-επίπεδα SONs (Layered SONs), όπου το όριο για να ενταχθεί ένας κόμβος σε ένα SON έχει οριστεί σε 0%, και έτσι ο κόμβος

θα μπει σε ένα SON που συσχετίζεται με το περιεχόμενό του, για το οποίο έχει ένα ή περισσότερα αρχεία - έγγραφα.

Οι αναζητήσεις στα πολύ - επίπεδα SONs ξεκινούν με την ταξινόμηση των ερωτημάτων, συνεχίζουν με την αποστολή του ερωτήματος στο/α SON(s) που συσχετίζονται με την ταξινόμηση του ερωτήματος και καταλήγουν με την αποστολή του ερωτήματος σταδιακά, σε ανώτερα ιεραρχικά επίπεδα, καθώς αυτά όπως προαναφέρθηκε μπορεί να έχουν σχετικά αποτελέσματα.

Αντιπαραβάλλοντας τα αποτελέσματα της σύγκρισης σχετικά με το μέγεθος των SON και τον αριθμό των κόμβων μέσα σε ένα SON, από την συντηρητική προσέγγιση και την πολύ-επίπεδη SON (Layered SONs) προσέγγιση, με τη χρήση των παρακάτω διαγραμμάτων παρατηρούμε ότι περίπου το 35% των συνολικών κόμβων πρέπει να ανήκουν σε ένα μόνο στυλ. Επιπλέον, το 97% των κόμβων πρέπει να ανήκουν σε τέσσερις ή λιγότερες κατηγορίες στυλ. Βλέπουμε ακόμη ότι η χρήση των Layered SONs, όπως έχει ήδη αναφερθεί, βοηθά στη μείωση του αριθμού των κόμβων ανά SON. Έτσι, έχουμε μεγάλο αριθμό μικρών SONs, γεγονός που με τη σειρά του οδηγεί σε σημαντικές βελτιώσεις στην ταξινόμηση και προώθηση των ερωτημάτων. (Pavel Serdyukov Saarland University, Department of Computer Science)



Σχήμα 6

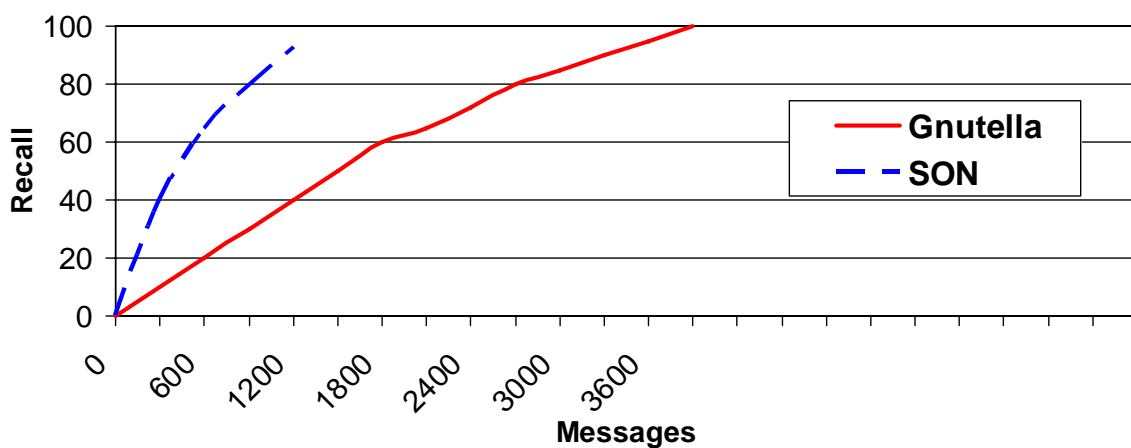
Αναζήτηση με Layered SONs

Οι αναζητήσεις στα Layered SONs γίνονται αρχικά ταξινομώντας τα διάφορα ερωτήματα. Έπειτα το ερώτημα αποστέλλεται στο SON, που συσχετίζεται με την βασική έννοια της ταξινόμησης του, και τέλος σταδιακά αποστέλλεται υψηλότερα στην ιεραρχία μέχρι να βρεθούν αρκετά αποτελέσματα. Αν επιστραφούν περισσότερες από μία έννοιες από τον ταξινομητή, γίνεται διαδοχική έρευνα για όλες τις επιστρεφόμενες έννοιες προτού προωθηθεί σε υψηλότερο επίπεδο. Ένα παράδειγμα για να γίνει αυτό πιο κατανοητό θα μπορούσε να είναι η αναζήτηση ενός αρχείου "Soft Rock". Ξεκινάμε την αναζήτηση από το Soft Rock SON, στέλνουμε το ερώτημα στο Rock SON εάν δεν επιστραφούν εξ' αρχής αρκετά αποτελέσματα και τέλος στο Music SON αν και η δεύτερη προώθηση δεν αποδώσει τα προσδόκιμα. (Arturo Crespo, Hector Garcia-Molina, 2003)

Πειράματα

Παρακάτω θα εξεταστούν δύο πιθανές διαμορφώσεις SON και θα αξιολογηθούν οι επιδόσεις τους σε σχέση με ένα σύστημα όπως το Gnutella. Υποθέτουμε ότι όλοι οι κόμβοι του δικτύου, τόσο αυτοί στο εσωτερικό των SON, όσο και αυτοί στο δίκτυο Gnutella είναι συνδεδεμένοι μέσω ενός άκυκλου γραφήματος και ότι κατά μέσο όρο κάθε κόμβος συνδέεται με τέσσερις άλλους κόμβους. Παρά το ότι η υπόθεση του άκυκλου γραφήματος δεν είναι ρεαλιστική, θεωρούμε άκυκλα τα δίκτυα καθώς η επίδραση των κύκλων είναι ανεξάρτητη της δημιουργίας των SON. Οι κύκλοι επηρεάζουν ένα P2P σύστημα δημιουργώντας επαναλαμβανόμενα μηνύματα που περιέχουν ερωτήματα τα οποία οι κόμβοι που τα δέχονται τα έχουν ήδη δει. Έτσι η ανάλυση ενός άκυκλου P2P δικτύου μας δίνει χαμηλότερη

εκτίμηση του αριθμού των μηνυμάτων που δημιουργούνται. Για το συγκεκριμένο πείραμα χρησιμοποιήθηκαν 50 διαφορετικά, τυχαία ερωτήματα που αποκτήθηκαν από ίχνη πραγματικών ερωτημάτων που στάλθηκαν σε έναν OpenNap server που λειτουργεί στο Stanford. Τα συγκεκριμένα ερωτήματα ταξινομήθηκαν με το χέρι. Αυτά που ταξινομήθηκαν στο επίπεδο substyle αποστάλθηκαν διαδοχικά στο αντίστοιχο SON(s) και έπειτα στο επίπεδο στυλ(είδος) του SON. Ερωτήματα που ταξινομήθηκαν στο επίπεδο στυλ, στάλθηκαν πρώτα διαδοχικά σε όλες τις υποκατηγορίες του στυλ και έπειτα στο επίπεδο στυλ. Τα ερωτήματα που ταξινομούνται στη "ρίζα" της ιεραρχίας αποστέλλονται σε όλους τους κόμβους. Μετράμε το επίπεδο της ανάκλησης κατά μέσο όρο για όλα τα ερωτήματα(50), σε σχέση με τον αριθμό των μηνυμάτων που στάλθηκαν στο σύστημα. Τα διαγράμματα αποκτήθηκαν "τρέχοντας" 50 προσομοιώσεις σε τυχαίες τοπολογίες δικτύου. Στο σχήμα που ακολουθεί φαίνονται τα αποτελέσματα του πειράματος. (Arturo Crespo, Hector Garcia-Molina, 2003)



Σχήμα 7

Το σχήμα δείχνει τον αριθμό των μηνυμάτων που στάλθηκαν σε σχέση με το επίπεδο της ανάκλησης. Τα πολυ-επίπεδα SONS είναι ικανά να επιτύχουν το ίδιο επίπεδο "ταιριασμάτων" με σημαντικά λιγότερα μηνύματα από ότι ένα σύστημα όπως το Gnutella. Παρ' όλα αυτά γενικά δεν πετυχαίνουν επίπεδα ανάκλησης 100% , μέση μέγιστη ανάκληση είναι το 93% , γεγονός που οφείλεται σε λάθη της ταξινόμησης των κόμβων.

Από την εικόνα φαίνεται η μέση απόδοση για όλους τους τύπους των ερωτημάτων. Ωστόσο, αν ένας χρήστης είναι σε θέση να ταξινομήσει με ακρίβεια το ερώτημα του, θα έχει συγκριτικά καλύτερη απόδοση. Η εικόνα δείχνει τον αριθμό των μηνυμάτων που στάλθηκαν σε σχέση με το επίπεδο της ανάκλησης για ερωτήματα που ταξινομήθηκαν στο επίπεδο των υποκατηγοριών του στυλ, που αντιστοιχεί στο χαμηλότερο επίπεδο της ιεραρχίας. Σε αυτή την περίπτωση υπάρχει σημαντική βελτίωση έναντι στο Gnutella. Για παράδειγμα, για να επιτευχθεί επίπεδο ανάκλησης της τάξης του 50% τα Layered SONS απαιτούν 461 μηνύματα σε αντίθεση με το Gnutella που απαιτεί 1731(!), δηλαδή 375% περισσότερα. Επιπρόσθετα, ακόμη και σε υψηλά επίπεδα ανάκλησης τα Layered SONS είναι ικανά να πετύχουν επίπεδο ανάκλησης της τάξης του 92% με περίπου το 1/5 των μηνυμάτων που απαιτεί το Gnutella. Το σχήμα της καμπύλης για την απόδοση του μηνύματος στο Gnutella είναι λίγο διαφορετικό για όλα τα ερωτήματα και για τα ερωτήματα που ταξινομήθηκαν στο επίπεδο υποκατηγορίας του στυλ. Στην προκειμένη περίπτωση έχουν ταξινομηθεί με ακρίβεια ερωτήματα για πολύ δημοφιλή τραγούδια. Εξαιτίας της δημοτικότητας τους, υπάρχουν πολλά αντίγραφα αυτών σε όλο το δίκτυο. Επομένως η προσέγγιση αναζήτησης τύπου Gnutella θα έχει μεγάλη πιθανότητα να βρει κάτι που να ταιριάζει σε πολλούς από τους κόμβους που επισκέφτηκε, κάνοντας την "πλημμύρα" στο δίκτυο μικρότερης σημασίας πρόβλημα από ότι θα ήταν τα "σπάνια" τραγούδια. Πάραυτα, και σε αυτή την περίπτωση τα Layered SONS έχουν καλύτερες επιδόσεις από ότι το Gnutella. (Arturo Crespo, Hector Garcia-Molina, 2003)

Συμπεράσματα

Αναλύθηκε πως μπορεί να βελτιωθεί η αποτελεσματικότητα ενός peer to peer συστήματος, ομαδοποιώντας κόμβους με παρόμοιο περιεχόμενο σε Semantic Overlay Networks (SON). Ακόμη πως τα SON μπορούν αποτελεσματικά να προωθήσουν τα ερωτήματα, διατηρώντας παράλληλα υψηλό βαθμό αυτονομίας του κόμβου. Παρουσιάστηκαν ακόμη τα πολύ - επίπεδα SON (Layered SONs), μια προσέγγιση που βελτιώνει την απόδοση των ερωτημάτων ακόμη περισσότερο με κόστος μικρή μείωση του μέγιστου εφικτού επιπέδου ανάκλησης. Από διάφορα πειράματα μπορούμε να πούμε ότι τα SON προσφέρουν σημαντικές βελτιώσεις σε σύγκριση με τα τυχαία επικαλυπτόμενα δίκτυα, κρατώντας παράλληλα το κόστος σε χαμηλά επίπεδα, και μπορούν να συμβάλλουν στη βελτίωση της απόδοσης των αναζητήσεων (ερωτημάτων) σημερινών, αλλά και μελλοντικών P2P συστημάτων όπου τα δεδομένα είναι φυσικά συγκεντρωμένα..

Βιβλιογραφία

Alexander Loser, Steffen Staab, Christoph Tempich (2007, JANUARY). *Semantic Social Overlay Networks*.

Arturo Crespo, Hector Garcia-Molina (2003). *Semantic Overlay Networks for P2P Systems*. Stanford University.

Carmela Comito, Simon Patarin, Domenico Talia. (n.d.). *A Semantic Overlay Network for P2P Schema-Based Data Integration*. Italy.

Christos Doulkeridis, Akrivi Vlachou, Kjetil Nørvag, Michalis Vazirgiannis (n.d.). *Distributed Semantic Overlay Networks*.

Heng Tao Shen, Yan Feng Shu, Bei Yu (n.d.). *Efficient Semantic-based Content Search in P2P*.

Josiane Xavier Parreira, Sebastian Michel, Gerhard Weikum (n.d.). *p2pDating: Real Life Inspired Semantic Overlay Networks for Web Search*. Germany.

Kamal Jain, Laszl Lovouasz, Philip A. Chou (n.d.). *Building Scalable and Robust Peer-to-Peer Overlay Networks for Broadcasting using Network Coding*.

Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth, Tim Van Pelt (n.d.). *GridVine: Building Internet-Scale*. Switzerland.

Michalis Vazirgiannis, , Kjetil Nørvag, Christos Doulkeridis (n.d.). *Peer-to-Peer Clustering for Semantic Overlay Network*.

Pavel Serdyukov Saarland University, Department of Computer Science (n.d.). http://www.mpi-inf.mpg.de/departments/d5/teaching/ws03_04/p2p-data/12-09-writeup2.pdf.

Philippe Cudré-Mauroux, Karl Aberer (2007, SEPTEMBER). *Message Passing in Semantic Peer-to-Peer*. *IEEE SIGNAL PROCESSING MAGAZINE* , σσ. 131-135.

Prusty, S. (2009, November). *Semantic Overlay Networks*. (Διπλωματική Εργασία).Guwahati.

Spyros Voulgaris, Maarten van Steen, Konrad Iwanicki (2002). *Proactive gossip-based management of semantic overlay networks*.

Sungsu Kim, John Strassner, James Won-Ki Hong (2011). *Semantic Overlay Network for Peer-to-Peer Hybrid Information Search and Retrieval*. , Korea.

Tao Gu , Hung Keng Pung , Daqing Zhang (n.d.). *A Hierarchical Semantic Overlay for P2P Search*. Singapore.

Tao Gu, Daqing Zhang, Hung Keng Pung (n.d.). *A Two-tier Semantic Overlay Network for P2P Search*. Singapore.

Wikipedia. (2012). *P2P*. Retrieved from <http://www.easypedia.gr/el/articles/p/e/e/Peer-to-peer.html>).

Xuemin Shen, Heather Yu, John Buford, Muralin Akon (2010). *Handbook of peer-to-peer networking*. USA: Springer.

Σταματελάτος,Γ. (2011, Απρίλιος). *Κατανεμημένη αναζήτηση απο δίκτυα agents με προστασία της ιδιοτηκότητας. (Διπλωματική εργασία)* . Ξανθη.

Μπαλιούσκας,Π. (n.d.). *Peer-to-Peer Συστήματα* (Πτυχιακή Εργασία). Κρήτη.

Οι εικόνες είναι από :

Arturo Crespo, Hector Garcia-Molina (2003). *Semantic Overlay Networks for P2P Systems*. Stanford University.

Εξώφυλλο :

<http://www.ijscn.com/Author-Instructions.php>

<http://connectedthebook.com/pages/links.html>