

## АДАПТИВНИЙ МЕТОД МАРШРУТИЗАЦІЇ ПОШУКУ ДОКУМЕНТІВ В НЕСТРУКТУРОВАНИХ PEER-TO-PEER МЕРЕЖАХ

Запропоновано адаптивний алгоритм маршрутизації в неструктурованих децентралізованих Peer-to-Peer мережах на основі вивчення зовнішнього локального стану мережі. Був використаний метод подоби косинуса, для модифікації структури індекса маршрутизації та алгоритму вибору експертних сусідів, що збільшує подібність знайдених документів до теми запиту і зменшує кількість згенерованих повідомлень у мережі.

An adaptive routing algorithm in unstructured decentralized Peer-to-Peer networks, based on external network learning is offered in the article. Cosine similarity method is used for modifying the routing index structure and expert neighbors selection algorithm, which increases the similarity of found documents to the request topic and reduces the number of generated messages in the network.

### Вступ

Неструктуровані системи P2P пропонують гнучке і автономне середовище, вони потребують не великої кількості управління для розміщення ресурсів і пірів. Стратегія маршрутизації в неструктурованій системі P2P повинна розглядати динамічні аспекти систем P2P; піри динамічні та постійно приєднуються і від'єднуються від системи, мережеве завантаження дуже мінливе, і протягом усього часу додаються нові ресурси, видаляються старі. Тому, стратегія маршрутизації повинна адаптуватися до таких змін, щоб підтримувати її продуктивність.

### Огляд існуючих рішень

Найбільш популярними неструктурованими системами є Gnutella [1] і Kazaa [2]. Оригінальний алгоритм Gnutella використовує ширококомове поширення запитів для пошуку ресурсів. В [3] пропонується модифікована схема ширококомової передачі, що випадково вибирає сусідів для передачі запиту. Два аналогічних підходи, які використовують BFS пошук зі збільшенням глибини, описані в [4, 5].

У підході [6] вводиться поняття «суперпір», в якому вузли розділені на суперпіри і листові вузли. Кожен суперпір пов'язаний з іншими суперпірами і з набором листових вузлів. У першому підході GUESS [7], пошук виконується за допомогою ітеративного взаємодії суперпірів, які в свою чергу перенаправляють запит своїм листовим вузлам, доки не виконається деяка умова. У другому підході, Gnutella2 [8], коли суперпір отримує запит від листового вуз-

ла, він ініціює еквівалентний Gnutella пошук з TTL = 2. Сусідні суперпіри регулярно обмінюються інформацією локальних таблиць для фільтрації трафіку. Даний підхід має слабку структуру і централізацію, що є слабким місцем в мережах з сильно динамічною поведінкою вузлів.

В [9] описується адаптивний до внутрішніх станів алгоритм маршрутизації для ефективного пошуку з використанням індексів маршрутизації. Недолік структури індекса маршрутизації та алгоритму його використання в роботі [9] – недостатня точність вибору потенційних сусідів. Індекс маршрутизації зберігає тільки подоби теми поточного вузла до однієї з тем сусіднього вузла, що не є достатньою підставою вважати, що подібність теми запиту пошуку перевищить необхідний поріг, щоб вважати теми подібними. Так само запит буде перенаправлений тільки пірам, посилення на які зберігаються в обраній темі індекса маршрутизації, але на практиці, більшої результативності можна добитися, вибираючи зв'язки з декількох рядків індекса. Даний недолік може бути усунутий з використанням методу подоби косинуса і його вихідних даних для розрахунку подоби тем, зміни структури індекса маршрутизації та алгоритму вибору сусідів.

### Алгоритм маршрутизації на основі модифікованого індекса маршрутизації

За основу даного алгоритму було взято підхід, запропонований в роботах [9, 10]. Розглядається однорангова архітектура, пошук якої заснований на використанні ключових слів для

знаходження документів. Піри зберігають набір документів. Запити формуються у формі списку умов і ключових слів. Кожен пір підтримує індекс маршрутизації для направлення запитів. Структура індекса та алгоритм його обробки в [9,10] були модифіковані в даній роботі, для покращення ефективності та результативності

пошукових запитів в мережі, а також зменшення кількості повідомлень в мережі.

Розглянемо процес створення записів індекса маршрутизації при підключенні піра до мережі та вибору найкращих сусідів для передачі запиту пошуку. Мережа зображена на рис. 1.

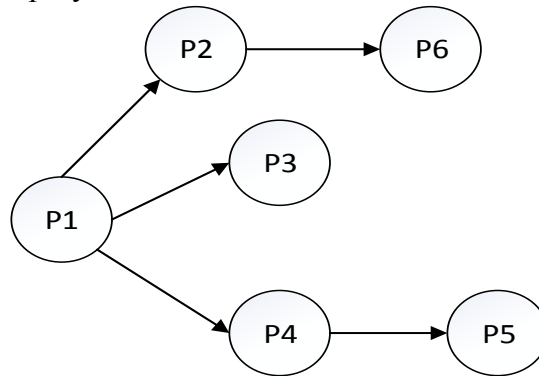


Рис. 1. P2P мережа

Припустимо пір P1 є експертом у двох темах,  $t_{11}$  і  $t_{12}$ :

$t_{11} = \{A, B, C, D, E, F, G, H, U, P\}$

$t_{12} = \{E, F, G, H, U, X, Z, W, Y\}$

Пір P2 має ресурси в темою  $t_{21}$ , P5 з темою  $t_{51}$ :

$t_{21} = \{A, B, C, D, E, F, X, W, O, J\}$

$t_{51} = \{E, F, G, H, U, P, X, Z, R, S\}$

Для простоти прикладу припустимо, що піри P3, P4, P6 не мають ресурсів і відповідно не є експертами в жодній з тем. У процесі підключення пір ширококомовно розсилає вектори тем  $t_{11}$  і  $t_{12}$ , в яких він є експертом з деяким TTL і порогом подібності тем  $\delta$ , досягнення якого говорить про те, що теми подібні. У даному

прикладі візьмемо TTL запити пошуку сусідів = 2 і  $\delta = 0.55$ .

Спочатку пір P1 ширококомовно передає теми  $t_{11}$  і  $t_{12}$  пірам P2, P3 та P6. Так як піри P3 і P4 не мають ресурсів, обчислення подібності тем не відбувається, і якщо TTL запити після декременту не дорівнює 0, то запит буде переданий далі його сусідам. Пір P2 є експертом в темі  $t_{21}$ . У процесі пошуку подібних до тем, будуть порівняні між собою  $t_{11}$  з  $t_{21}$  і  $t_{12}$  з  $t_{21}$ . При порівнянні двох відбувається за використанням методу подоби косинуса. Процес зображено в табл. 1 на прикладі  $t_{11}$  і  $t_{21}$ .

Табл. 1. Обчислення подібності між векторами тем  $t_{11}$  і  $t_{21}$

R =	A	B	C	D	E	F	G	H	U	P	X	Z	R	S
$t_{11}$	A	B	C	D	E	F	G	H	U	P	-	-	-	-
$t_{21}$	-	-	-	-	E	F	G	H	U	P	X	Z	R	S
$t_{11b}$	1	1	1	1	1	1	1	1	1	1	0	0	0	0
$t_{21b}$	0	0	0	0	1	1	1	1	1	1	1	1	1	1

$$\text{Cos}(t_{11}, t_{21}) = \frac{1 + 1 + 1 + 1 + 1 + 1}{\sqrt{10} * \sqrt{10}} = \frac{6}{10} = 0.6$$

Аналогічно порівнюються теми  $t_{12}$  з  $t_{21}$ . Було отримано два значення подібності,  $\text{Sim}(t_{11}, t_{21}) = 0.6$  та  $\text{Sim}(t_{12}, t_{21}) = 0.42$ . Так як поріг подібності  $\delta = 0.55$ , теми  $t_{11}$  та  $t_{21}$  вважаються подібними і піри P1 та P2 запишуть дані один про одного в індекс маршрутизації відповідно до тем  $t_{11}$  та  $t_{21}$ .

З використанням модифікованої структури індекса маршрутизації, для створення запису береться та частина відображення результуючо-

го вектора тем, яка відповідає ключовим словам локальної теми вузла. Додаткові ключові слова в індексі маршрутизації є різницею множини вектора об'єднання і вектора теми запиту. Так як дані теми є подібними, то кількість відсутніх ключових слів не велика, що дозволяє економити пам'ять для підтримки більшої кількості записів в індексі маршрутизації. Приклад запису показано в табл. 2.

Табл. 2. Запис в індексі маршрутизації вузла P1.

Ідентифікатор теми	Вектор	Ключові слова	Адреса
t <sub>11</sub>	63(0 0 0 0 1 1 1 1 1 1)	X, Z, R, S	P2

Так як значення TTL не було досягнуто, пошук експертних сусідів продовжується. На другому кроці запит пошуку сусідів отримує пір P6, що не містить ресурсів, та пір P5 що зберігає документи під темою t<sub>51</sub>. В даному вузлі будуть проведені операції аналогічні з тими, що відбувалися в пірі P2.

В результаті обчислень подібності тем, будуть отримані такі значення:  $\text{Sim}(t_{11}, t_{51}) = 0.52$  та  $\text{Sim}(t_{12}, t_{51}) = 0.73$ . Так як поріг

подібності  $\delta = 0.55$ , теми t<sub>12</sub> та t<sub>51</sub> вважаються подібними і піри P1 та P5 запишуть дані один про одного в індекс маршрутизації відповідно до тем t<sub>12</sub> та t<sub>51</sub>.

Після другого кроку максимальне значення TTL досягнуто и пошук припинено. В результаті пір P1 буде мати 2 записи в індексі маршрутизації. Таблиця маршрутизації піра P1 в результаті буде виглядати так, як показано в табл. 3.

Табл. 3. Індекс маршрутизації піра P1

Ідентифікатор теми	Вектор	Ключові слова	Адреса
t <sub>12</sub>	508	P, R, S	P5
t <sub>11</sub>	63	X, Z, R, S	P2

Припустимо пір P1 ініціює пошук і створює запит, що містить вектор теми  $t_q = \{E, F, G, H, U, P\}$ . При виборі сусідів для передачі запиту P1 має обчислити подібність запиту теми  $t_q$  до усіх тем в індексі маршрутизації. Дане обчислення також використовує метод подоби косинуса, так як і в порівнянні тем при пошуку експертних сусідів.

Подібність між темами  $\text{Sim}(t_{11}, t_q) = 0.778$ ,  $\text{Sim}(t_{12}, t_q) = 0.68$ . Обидві теми перевищують поріг  $\delta = 0.55$  і вузли, що записані в рядках даних тем, додаються до списку потенціальних вузлів для направлення запиту пошуку. Далі, для усіх вузлів даного списку відновлюються їхні теми. Процес відновлення полягає у відборі

ключових слів з вектора локальної теми за допомогою використання числового вектора в якості маски. Якщо ключовому слову відповідає одиниця, це означає що тема сусіднього піра містила дане ключове слово у своєму векторі ключових слів, якщо відповідає 0, значить даного ключового слова не було у векторі сусіда. Після знаходження спільних ключових слів, у відновлений вектор додаються ключові слова з вектора доповнення. В результаті отримуємо повністю відновлений вектор ключових слів сусіднього піра, що дозволить більш точно вибрати сусіда для перенаправлення запиту. Процес для запису t<sub>12</sub> в індексі маршрутизації показаний в табл. 4.

Табл. 4. Відновлення теми сусіда з даних в індексі маршрутизації

t <sub>12</sub>	E	F	G	H	U	X	Z	W	Y	
Числовий вектор = 508	1	1	1	1	1	1	1	0	0	
Вектор доповнення	P, R, S									
Відновлення вектора теми сусіда	E	F	G	H	U	X	Z	-	-	
Відновлений вектор t <sub>p5</sub>	A	B	C	D	E	F	X	P, R, S		

В результаті отримаємо такі теми для пірів P2 і P5:

$$t_{p5} = \{E, F, G, H, U, P, X, Z, R, S\}.$$

Аналогічно для  $t_{p2} = \{A, B, C, D, E, F, X, W, O, J\}$ .

Після відновлення тем пірів, наступним кроком вираховується подібність між темою запиту та пірів,  $\text{Sim}(t_{p2}, t_q) = 0.259$  і  $\text{Sim}(t_{p5}, t_q) = 0.683$ .

Так як тема вузла P2 не перевищує поріг подібності, він не буде використаний в якості експертного вузла, а P5 перевищує поріг і потрапляє в список вузлів кандидатів.

Необхідно звернути увагу, що подоба теми запиту до теми t<sub>11</sub> була вищою, ніж до теми t<sub>12</sub> і

у випадку стандартного алгоритму експертних груп були б вибрані сусіди з запису в темі  $t_{11}$ , тобто запит був би направлений до вузла P2, а вузол P5 взагалі не був би вибраний. Але з прикладу можна побачити, що пір P2 насправді містить тему, подібність якої до теми запиту дуже мала, а саме, має значення 0.259. В той самий час пір P5 має тему, подібність якої до теми запиту має високе значення і перевищує поріг, а саме, має значення 0.683. Тобто запит був би направлений до вузла, який не має теми з необхідним порогом подібності, а вузол з темою, що перевищує поріг, взагалі не був би розглянутий в якості потенційного експертного вузла. Використовуючи модифікацію структури індекса маршрутизації та алгоритму вибору експертних сусідів, було не тільки враховано вузол з кращою подобою теми до запиту, але й виключено потенційно задовільний експертний вузол, який насправді не має теми з необхідним порогом подібності. В результаті, розроблений в даній роботі алгоритм, не тільки дозволяє врахувати всі можливі експертні вузли, тим самим покращивши результативність та

ефективність системи, а й зменшити кількість згенерованих повідомлень в мережі, за рахунок вилучення потенційно задовільних експертних вузлів.

## Висновки

В даній роботі було запропоновано та описано алгоритм адаптивної маршрутизації в неструктурованих P2P мережах. Були виправлені недоліки алгоритму обслуговування та структури індекса маршрутизації, за допомогою метода подоби косинуса та зберігання додаткових даних, що з'являються в процесі виконання даного алгоритму. Приклад роботи алгоритму наглядно показав переваги даного підходу.

В результаті, розроблений в даній роботі алгоритм, не тільки дозволяє врахувати всі можливі експертні вузли, тим самим покращивши результативність та ефективність системи, а й зменшити кількість згенерованих повідомлень в мережі, за рахунок вилучення потенційно задовільних експертних вузлів, які насправді не можуть задовільнити потреби запиту.

## Список посилань

1. Gnutella, "The Gnutella Protocol Specification" [Електронний ресурс] // – Режим доступу: <http://dss.clip2.com/GnutellaProtocol04.pdf> – Дата доступу: 27.04.2015.
2. Kazaa website [Електронний ресурс] // – Режим доступу: <http://www.kazaa.com>. – Дата доступу: 27.04.2015.
3. V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-Yazti. A local search mechanism for peer-to-peer networks. // Proceedings of the eleventh international conference on Information and knowledge management, 2002. – P. 300-307.
4. B. Yang and H. Garcia-Molina. Improving search in peer-to-peer networks. // In 22nd International Conference on Distributed Computing Systems (ICDCS), July 2002.
5. C. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and replication in unstructured peer-to-peer networks. // In Integrated Computer Solution, 2002.
6. A. Singla and C. Rohrs. Ultrapeers: Another step towards gnutella scalability. // IEEE Transactions on Network and Service Management, 2007.
7. S. Daswani and A. Fisk. Gnutella UDP extension for scalable searches (GUESS) v0.1. // Proc. of the 3rd International Workshop on Peer-to-Peer Systems (IPTPS'04), 2004.
8. M. Stokes. Gnutella2 specifications part one [Електронний ресурс] // – Режим доступу: <http://www.gnutella2.com> – Дата доступу: 27.04.2015.
9. Hidayanto, A.N., Bressan, S.: Towards a Society of Peers: Expert and Interest Groups in Peer-to-Peer System. // On The Move (OTM) Workshop, Portugal, (2007).
10. Achmad Nizar Hidayanto and Stephane Bressan. Adaptive Routing Algorithms in Unstructured Peer-to-Peer (P2P) Systems. // International Journal on Computer Science and Engineering (IJCSE), 2007.