

КОНКУРЕНТНА РОЗВІДКА ТА УПРАВЛІННЯ ЗНАННЯМИ/ BUSINESS INTELLIGENCE & KNOWLEDGE MANAGEMENT

ПОШУК СТОРІНОК ДИСКУСІЙ В СОЦІАЛЬНИХ МЕРЕЖАХ ГЛОБАЛЬНИМИ ПОШУКОВИМИ СИСТЕМАМИ

Андрій Пелешишин¹, Руслан Гумінський²,
Оксана Тимовчак-Максимець¹

¹Національний університет «Львівська політехніка», Україна

²Академія сухопутних військ, Україна



ПЕЛЕШИШИН Андрій Миколайович, д.т.н., професор

Рік та місце народження: 1973 рік, м. Львів, Україна.

Освіта: Львівський Державний університет ім. І.Франка, 1995 рік.

Посада: завідувач кафедри соціальних комунікацій та інформаційної діяльності з 2011 року.

Наукові інтереси: системотворчі процеси WWW, методи побудови інформаційного суспільства, позиціонування сайтів у WWW, соціальні мережі у WWW та інформаційні технології соціальних комунікацій.

Публікації: більше 100 наукових публікацій, серед яких монографії, підручники, навчальні посібники, наукові статті, матеріали та тези доповідей на конференціях.

E-mail: apele@ridne.net



ГУМІНСЬКИЙ Руслан Вікторович

Рік та місце народження: 1975 рік, м. Житомир, Україна.

Освіта: Харківський військовий університет, 1997 рік, Національна Академія оборони України, 2007 рік.

Посада: старший науковий співробітник науково-дослідного відділу (моделювання бойових дій) Наукового центру Сухопутних військ Академії сухопутних військ з 2009 року.

Наукові інтереси: імітаційне моделювання бойових дій, інформаційні загрози в Інтернет середовищі.

Публікації: більше 20 наукових публікацій, серед яких наукові статті, матеріали та тези доповідей на конференціях.

E-mail: GRuslan@meta.ua



ТИМОВЧАК-МАКСИМЕЦЬ Оксана Юріївна, к.т.н.

Рік та місце народження: 1984 рік, м. Львів, Україна.

Освіта: Національний університет «Львівська політехніка», 2006 рік.

Посада: асистент кафедри соціальних комунікацій та інформаційної діяльності Національного університету «Львівська політехніка» з 2007 року.

Наукові інтереси: пошук інформації в масивах текстів природної мови, text mining, opinion mining.

Публікації: більше 25 наукових публікацій, серед яких монографія, наукові статті, матеріали та тези доповідей на конференціях.

E-mail: linoks@gmail.com

Анотація. На сьогоднішній день проходить бурхливий розвиток соціальних мереж, які пропонують для своїх користувачів багату кількість інструментів для задоволення їх потреб. Один із таких інструментів є створення сторінок дискусій, які об'єднують людей за тематикою спілкування. Виникає проблема в здійсненні пошуку необхідних сторінок дискусій в соціальних мережах, яка значно ускладнюється необхідністю проведення пошуку відповідно до тематики їх інформаційного наповнення. Було проведено

дослідження щодо використання формалізованих запитів глобальних пошукових систем та особливостей побудови сторінок дискусій. Розглянуті алгоритми пошуку за допомогою формалізованих запитів глобальної пошукової системи Google та аналізу Html коду сторінки дискусії надають можливість проведення пошуку дискусій в соціальних мережах відповідно до їх інформаційного наповнення.

Ключові слова: дискусії, соціальні мережі, глобальні пошукові системи, формалізовані запити, інформаційне наповнення.

Постановка проблеми

На сьогоднішній день проходить бурхливий розвиток соціальних мереж, які пропонують для своїх користувачів багату кількість інструментів для задоволення їх потреб. Один із таких інструментів є створення сторінок дискусій, які об'єднують людей за тематикою спілкування. За характером інформаційного наповнення цих сторінок вони широко використовуються, як користувачами соціальних мереж для обговорення відповідних тематик, так і для проведення соціальних досліджень, оцінки ставлення людей до тематик різною направленості (політичних, культурних та інші.) та вивчення потреб користувачів. Тим часом, пошук сторінок дискусій інструментами соціальних мереж здійснюється по назвам дискусій, або їх короткому змісту, що не завжди відповідає інформаційному наповненню цих сторінок.

Виникає проблема в здійсненні пошуку необхідної інформації на сторінках дискусій в соціальних мережах, яка значно ускладнюється необхідністю проведення пошуку відповідно до тематики інформаційного наповнення та актуальності сторінки дискусії з урахуванням особливостей функціонування цих сторінок дискусій [2], а саме:

- сторінки мають малий ранг в алгоритмах ранжування сторінок;
- велика кількість Веб-сторінок дискусій не ранжуються глобальними пошуковими системами;
- значна частина відвідувачів потрапляє на сайт за безпосередньою рекомендацією інших користувачів;
- взаємозв'язність веб-сторінок дискусій;
- збереження дискусій не актуальної тематичної направленості.

Аналіз останніх досліджень

У [3] проведено дослідження для виявлення релевантних спільнот та дискусій на спеціалізованих

Інтернет сервісах з використанням формалізованих запитів глобальних пошукових систем (ГПС) з урахуванням їх інформаційного наповнення, але не враховуються особливості функціонування сторінок дискусій в соціальних мережах та не дають можливості їх пошуку.

В інших дослідженнях [1] розглядаються використання інструментів ГПС для пошуку сторінок в соціальних мережах за загальними ознаками (сторінки користувачів, офіційні сторінки політичних партій, організацій) але без урахування тематики інформаційного наповнення.

Таким чином **цілями статті** є дослідження проведення пошуку сторінок дискусій в найбільш популярних соціальних мережах в Україні facebook та Вконтакте за допомогою глобальних пошукових систем з урахуванням їх інформаційного наповнення.

Основна частина

Структура формалізованого запиту

У дослідженні [3] визначено, що найефективнішим та найзручнішим є здійснення пошуку релевантних спільнот та дискусій у глобальній пошуковій системі Google. Для пошуку веб-сторінок спільнот за допомогою Google використовуємо метод формування формалізованих запитів до пошукових систем для виявлення релевантних веб-сторінок спільнот з використанням спеціальних операторів пошукової системи [4].

Запит до глобальної пошукової системи на виявлення релевантних дискусій складається з таких операцій:

- операції локалізації пошукового запиту;
- операції виявлення інформаційного наповнення;
- часові обмеження.

Загальна структура запиту зображена на рис. 1.



Рис. 1. Формалізований запит для виявлення релевантних дискусій веб-сторінок

Операція локалізації пошукового запиту дозволяє обмежити пошук веб-сайтом або доменом. Функціонально ця операція складається з **оператора локалізації** і **специфікатора веб-адреси**.

Оператор локалізації – службовий оператор ГПС для обмеження пошуку веб-сайтом або доменом.

Специфікатор веб-адреси – це вказівник на ресурс, на якому здійснюватиметься пошук.

Операція виявлення інформаційного наповнення дозволяє виявити серед веб-сторінок ті, які можуть стосуватися об'єкту пошуку. Ця операція складається з оператора **аналізу інформаційного наповнення** і **специфікатора релевантності**.

Оператор аналізу інформаційного наповнення – службові оператори ГПС, які аналізують інформаційне наповнення веб-сторінки за заданою тематикою. Інформаційним наповнення сторінки вважатимемо слово або словосполучення, яке знаходиться у тілі веб-сторінки.

Специфікатор релевантності – це ознака, яка характеризує об'єкт пошуку. Ознакою релевантності виступає слово або словосполучення у інформаційному наповненні веб-сторінки.

Часові обмеження – під час проведення пошуку необхідно використовувати обмеження результатів за визначений діапазон часу, з метою пошуку актуальних веб-сторінок спільнот. Для цього використовуються опції пошуку: за минулий час, останні 24 години, за минулий тиждень, місяць, рік або інтервал часу використовуючи інструменти пошуку.

У зв'язку з особливостями побудови та структури сторінок дискусій в соціальних мережах, їх пошук має свої характерні особливості.

Особливості пошуку спільнот та дискусій в facebook

Соціальна мережа facebook дозволяє користувачам створювати свої групи і сторінки спільноти, в яких можна об'єднувати людей за інтересами. Група (далі дискусії) – це спільнота, в якій люди можуть спілкуватися один з одним, ділитися інформацією та взаємодіяти, але тільки в рамках заданої теми або ідеї, тобто має ознаки тематичних дискусій. Одним із основних недоліків

дискусій є те, що вони не індексуються ГПС. Сторінки, в основному, організовуються різними організаціями або

окремими особистостями, як відкриті джерела для всіх, когочо якось цікавить тема сторінки, вони, як правило, зв'язані з тематично відповідними дискусіями. У відмінності від груп, сторінки індексуються ГПС.

Таким чином, пошук дискусій проводимо за допомогою пошуку сторінок в facebook, використовуючи формалізований запит (рис. 1), який має вигляд (рис. 2) з подальшим аналізом Html коду сторінки.

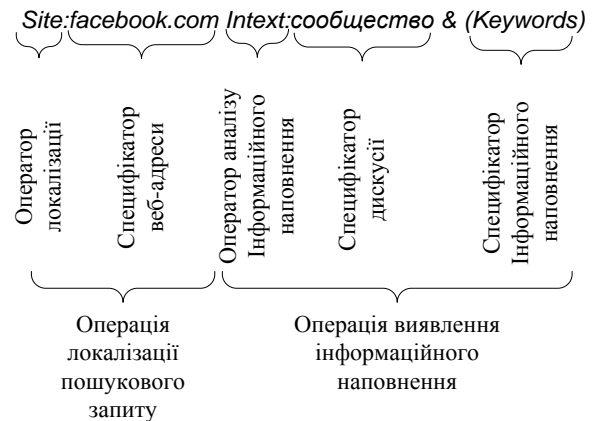


Рис. 2. Формалізований запит для виявлення сторінок в соціальній мережі facebook

Як і при пошуку в соціальній мережі Вконтакте до формалізованого запиту для виявлення сторінок в соціальній мережі facebook в операції виявлення інформаційного наповнення вводиться специфікатор сторінки «сообщество». Пошук проводиться відповідно до специфікатора інформаційного наповнення (Keywords) – множині ключових слів.

Сторінки в facebook строго структуровані і для визначення наявності зв'язаних дискусій проводимо аналіз html коду на присутність пункту меню «групи» (рис. 7).

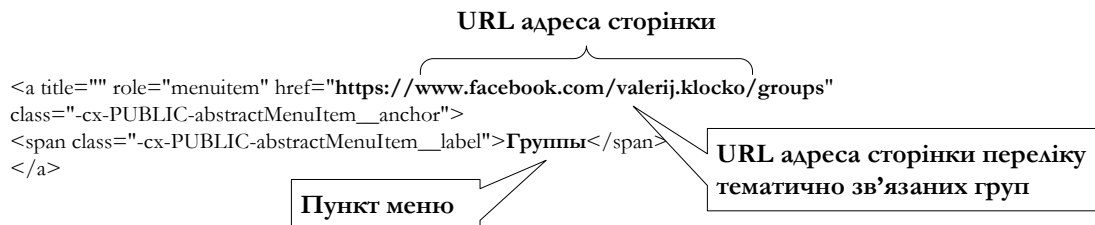


Рис. 3. Аналіз Html коду сторінки для виявлення URL адреси сторінки переліку тематично зв'язаних груп

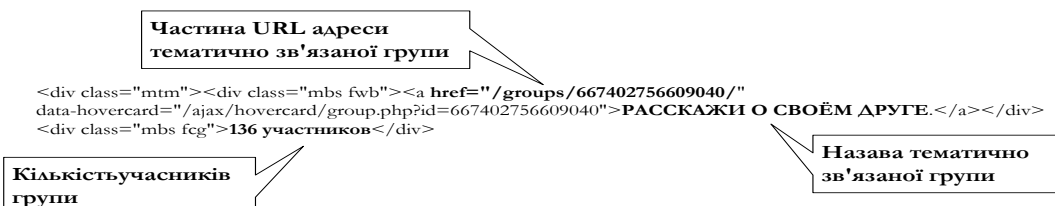


Рис. 4. Аналіз Html коду сторінки переліку тематично зв'язаних дискусій для виявлення частини URL адреси дискусій

Визначаємо частини URL адреси дискусій аналізуючи html код сторінки переліку тематично зв'язаних дискусій (рис. 4).

За результатами виявлених частин URL адреси формуємо URL адресу тематично зв'язаних груп. Приклад формування URL адреси тематично зв'язаних груп наведено в табл. 1.

Таблиця №1

Формування URL адреси тематично зв'язаних груп

Адреса сторінки	Частина адреси тематично зв'язаної групи	Адреса тематично зв'язаної групи
www.facebook.com/valerij.klocko	/groups/667402756609040/	www.facebook.com/groups/667402756609040/
www.facebook.com/valerij.klocko	/groups/Ivan.SOS/	www.facebook.com/groups/Ivan.SOS/
www.facebook.com/valerij.klocko	/groups/180978166552/	www.facebook.com/groups/180978166552/

Алгоритм пошуку зображений на рис. 5.

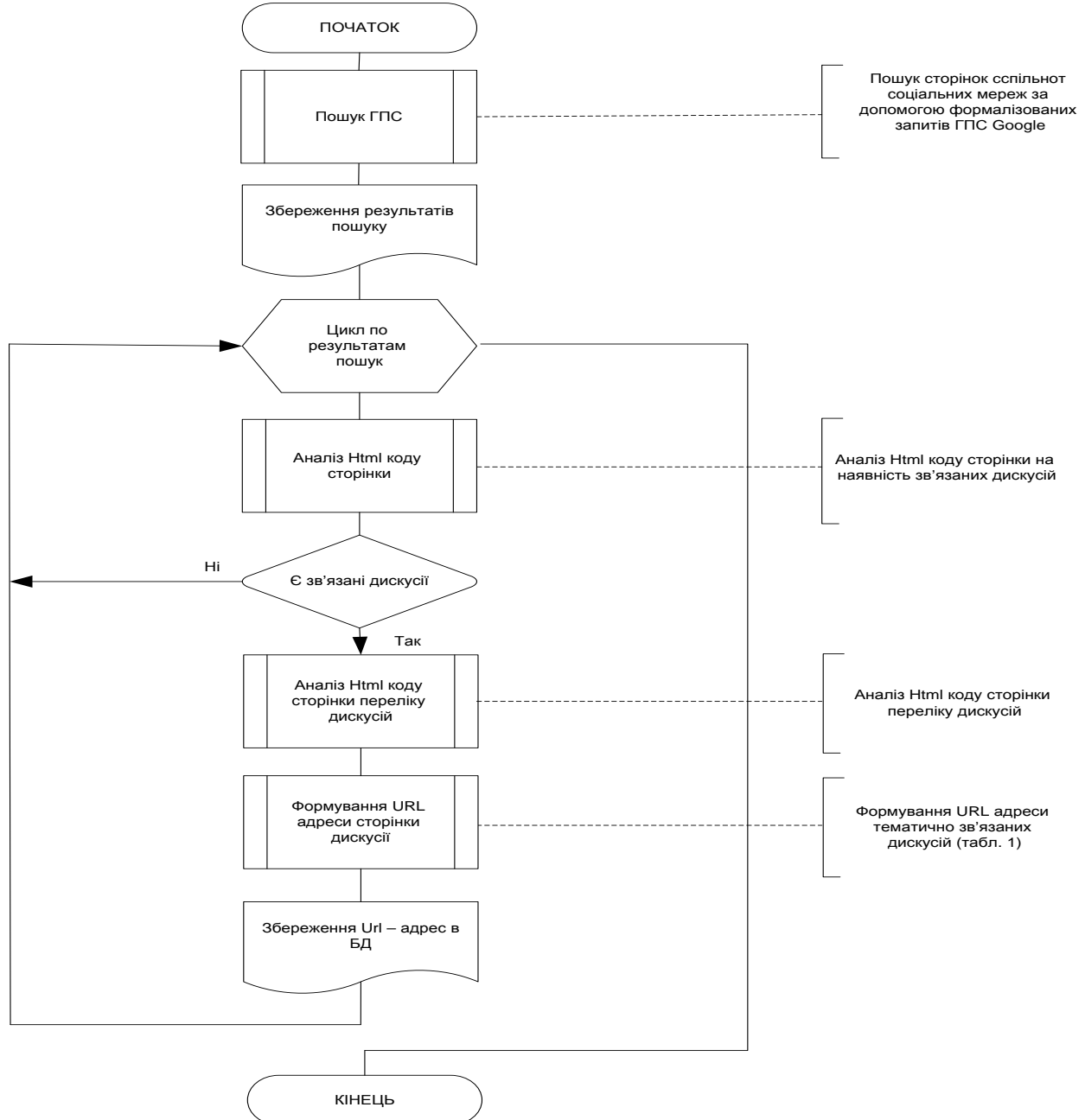


Рис. 5. Алгоритм пошуку дискусій в соціальній мережі facebook

Особливості пошуку спільнот та дискусій в Вконтакте

Соціальна мережа в Вконтакте має багато інструментів для задоволення потреб користувачів.

Одним із таких інструментів є створення груп (далі - дискусії), які мають ознаки тематичних дискусій. Таким чином для пошуку дискусій спільнот в

Вконтакте використовуємо формалізований запит (рис. 1), який має вигляд (рис. 6).

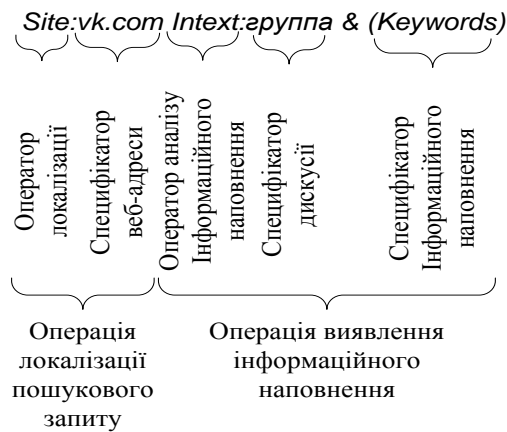


Рис. 6. Формалізований запит для виявлення дискусій в соціальній мережі Вконтакте

Додатково до формалізованого запиту для виявлення релевантних дискусій в соціальній мережі Вконтакте в операції виявлення інформаційного наповнення вводиться специфікатор дискусії «група», що дозволяє проводити пошук по сторінкам дискусій. Пошук проводиться відповідно до специфікатора інформаційного наповнення (Keywords) – множині ключових слів.

На сторінках дискусій в соціальній мережі Вконтакте можливе створення декількох тематик дискусій. Url адреси дискусій визначаємо за результатами аналізу html коду сторінки на наявність ідентифікатора внутрішніх дискусій (рис. 7) з визначенням їх адреси.

ідентифікатора внутрішніх дискусій

```

</div><div id="group_wide_topics"><div class="module clear topics_module" id="group_topics">
<a href="http://vk.com/board15667008" class="module_header">
<div class="header_top clear_fix">
  <span class="right_link fl_r" onmouseover="this.parentNode.parentNode.href='/board15667008'"
onmouseout="this.parentNode.parentNode.href='/board15667008'"></span>
  Обсуждения
</div>
<div class="p_header_bottom">
  <span class="fl_r"></span>
  14 тем
</div>
</a>
<div class="module_body clear_fix">
  <a class="clear_fix topic_row first" href="http://vk.com/topic-15667008_25846075">
<div class="info fl_l">
  <div><span class="topic_title">Что делать с кавказским беспределом?</span></div>
  <small>
    2132 сообщения.
  </small>
  Последнее от <span class="topic_inner_link" onmouseover="this.parentNode.parentNode.parentNode.href='/yesaanfa'"
onmouseout="this.parentNode.parentNode.parentNode.href='/topic-15667008_25846075'">Ерванда Саркисяна</span>, <span
class="topic_date">сегодня в 16:43</span>
  <span class="topic_to_last" onmouseover="this.parentNode.parentNode.parentNode.href='/topic-
15667008_25846075?offset=last&scroll=1'" onmouseout="this.parentNode.parentNode.parentNode.href='/topic-
15667008_25846075'">&#8594;</span>
  </small>
</div>
  ...
</a>
</div>
    
```

кількість дискусій

URL адреса дискусії

тема дискусії

Рис. 7. Аналіз Html коду сторінки дискусії для виявлення Url адрес внутрішніх дискусій

Результатом пошуку сторінок дискусій в соціальній мережі в Вконтакте є адреси знайдених сторінок дискусій, за допомогою формалізованих

запитів ГПС та адреси внутрішніх дискусій. Алгоритм пошуку зображений на рис. 8.

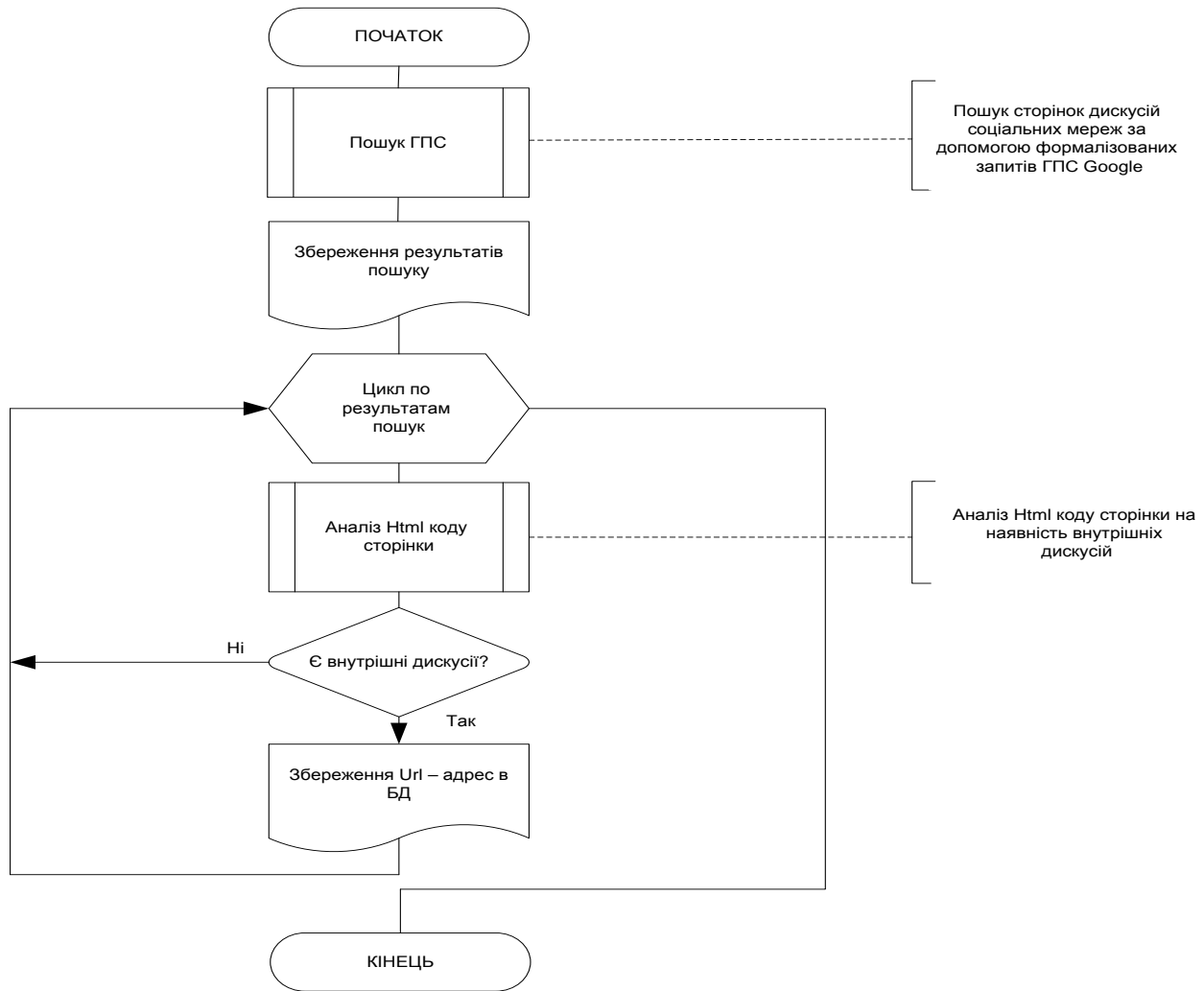


Рис. 8. Алгоритм пошуку дискусій в соціальній мережі Вконтакте

Висновки

Досліджено пошук сторінок дискусій в соціальних мережах facebook та Вконтакте з урахуванням особливостей їх функціонування за допомогою глобальної пошукової системи Google.

Було проведено ряд експериментів з використанням методу пошуку сторінок дискусій з використанням формалізованих запитів глобальної пошукової системи Google та аналізу Html коду сторінок дискусій в соціальних мережах з різної кількістю ключових слів (Keywords) від 3 до 6, оцінено релевантність пошуку яка становить для facebook 0,6 - 0,8 та Вконтакте 0,7 - 0,9. Оцінка проводилась за:

$$Relevation = \frac{N_{Keywords \subset D}}{N}$$

де $N_{Keywords \subset D}$ - кількість дискусій в інформаційному наповненню яких присутні ключові слова формалізованого запиту;

N - загальна кількість знайдених дискусій.

Таким чином запропоновані алгоритми, які використовують формалізовані запити глобальної пошукової системи Google та аналізу Html коду сторінок дискусій в соціальних мережах дають можливість їх пошуку у відповідності до їх інформаційного наповнення.

Література

[1] Губанов Д.А. Социальные сети информационного влияния: модели информационного влияния, управления и противоборства : монография / Д.А.Губанов, Д.А.Новиков, А.Г. Чхартишвили; под ред. Д.А. Новикова. - М.: Издательство физико-математической литературы, 2010. - 228 с.

[2] Пелецишин А.М. Процеси управління інтерактивними соціальними комунікаціями в умовах розвитку інформаційного суспільства : монографія / Ю.О. Серов, О.Л. Березко, О.П. Пелецишин, О.Ю. Тимовчак-Максимець, О.В. Марковець; за заг. ред. А.М.Пелецишина. - Львів : Видавництво Львівської політехніки, 2012. - 368 с.

[3] Тимовчак-Максимець О.Ю. Методи використання розширених можливостей глобальних пошукових систем в задачі пошуку споживацького досвіду в онлайн середовищах / О.Ю.Тимовчак-Максимець // Вісник Національного університету «Львівська політехніка»: Інформаційні системи та мережі. - № 689. - 2010. - С. 323-331.

[4] Google Guide Quick Reference: Google Advanced Operators [Електронний ресурс]. - Режим доступу: http://www.googleguide.com/using_advanced_operators.html

УДК 004.738.5 (045)

Пелецишин А.М., Гуминский Р.В., Тимовчак-Максимець О.Ю. Поиск страниц дискуссий в социальных сетях глобальными поисковыми системами

Аннотация. На сегодняшний день проходит бурное развитие социальных сетей, которые предлагают для своих пользователей богатую количество инструментов для удовлетворения их потребностей. Один из таких инструментов является создание страниц дискуссий, объединяет людей по тематике общения. Возникает проблема в осуществление поиска необходимых страниц дискуссий в социальных сетях, значительно осложняется необходимостью проведения поиска в соответствии с тематикой их информационного наполнения. Было проведено исследование по использованию формализованных запросов глобальных поисковых систем и особенностей построения страниц дискуссий. Рассматриваемый методы поиска с помощью формализованных запросов глобальной поисковой системы Google и анализа Html кода страницы дискуссии предоставляют возможность проведения поиска дискуссий в социальных сетях в соответствии с их информационного наполнения.

Ключевые слова: дискуссии, социальные сети, глобальные поисковые системы, формализованные запросы, информационное наполнение.

Peleshchyshyn A., Huminskyi R., Tymovchak-Maksymets O. Search for discussion pages in social networks by global search engines

Abstract. Nowadays a rapid development of social networks that offer a wide range of instruments to satisfy needs of their users is taking place. Creation of discussion pages, which unite people by the subject of discussion, is one of such instruments. A problem of needed discussion pages search in social networks emerges, and is much complicated by the necessity for search on the subject of their informational content. Research has been done on usage of formalized requests of global search engines and peculiarities of discussion pages design. Given search algorithms of formalized requests using global search engine Google and analysis of discussion page Html code enable discussion search in social networks according to their informational content.

Key words: discussions, social networks, global search engines, formalized requests, informational content.

Отримано 26 вересня 2013 року, затверджено редколегією 22 жовтня 2013 року