

**УДК 004.912**

**Б.І. Мороз докт. техн. наук, проф., Д.Є. Костенко, О.М.Спалило**  
Університет митної справи та фінансів, Україна.

## **ВИКОРИСТАННЯ ОНТОЛОГІЧНОГО ПІДХОДУ ДО ПОБУДОВИ ЕФЕКТИВНИХ МОДЕЛЕЙ ПОШУКУ**

**B.I.Moroz Dr., Prof.; D.E. Kostenko, O.M.Spalylo**  
**USING THE ONTOLOGICAL APPROACH FOR EFFECTIVE SEARCH PATTERNS  
CONSTRUCTION**

На сьогоднішній день людина все більше стикається з проблемою пошуку інформації. Зростаючий розмір кількості документів робить більш важливим завдання збільшення якості інформаційного пошуку.

Інформаційним пошуком називається деяка послідовність операцій, що виконуються з метою пошуку документів, що містять певну інформацію або з метою видачі фактичних даних, що представляють собою відповіді на питання. Термін був введений американським математиком Келвіном Муерсом в 1947 році.

Інформаційний пошук здійснюється засобами інформаційно-пошукової системи. Основними критеріями якості пошуку результатів інформаційного пошуку є: повнота, точність і оперативність. Пошук виконується в чотири етапи:

- 1) Визначення інформаційної потреби і власника інформаційного масиву;
- 2) Формулювання запиту;
- 3) Вилучення інформації з інформаційного масиву;
- 4) Ознайомлення з отриманою інформацією і оцінка результатів пошуку.

Основним засобом передачі інформації є документ. Документ – це матеріальний об'єкт, що містить інформацію в зафіксованому вигляді і спеціально призначений для її передачі в часі і просторі. За рівнем узагальнення інформації документи ділять на первинні і вторинні. Первинний документ містить інформацію, яка є викладенням результатів вивчення, дослідження, розробок. Вона оригінальна по своєму характеру. Такий документ являє собою результат безпосереднього відображення наукової, виробничої, управлінської, творчої інформації. Його завдання полягає в забезпеченні суспільства первинною, початковою інформацією. Вторинний документ є результатом аналітико-синтетичної переробки одного або декількох первинних документів. Вторинний документ (інформаційний) містить відомості про первинні документи, отримані шляхом їх інформаційного аналізу.

Існують такі моделі пошуку інформації: адресний пошук, семантичний пошук, документальний пошук, фактографічний пошук.

Семантичний пошук інформації – це процес пошуку документів по їх сенсовому змісту. Метою семантичного пошуку є розширення стандартного словникового значення слова або фрази для того, щоб зрозуміти наміри користувача в рамках конкретного контексту. Інформація в процесі “зіставлення” повинна оброблятися з використанням знань (про користувача, ресурси і т.д.). При розробці онтологічної моделі інформаційної потреби користувача враховуються не тільки формальні відомості про запит, але і більш складноструктуровані знання про них. Наявність додаткових знань про те, що саме шукає користувач, дозволяє структурувати знайдену інформацію і надавати її користувачу в більш зручному вигляді. Концепція вирішення проблеми інформаційного пошуку на основі онтологічного походу передбачає використання декількох видів онтологій і онтологічних структур. Онтологія - це формальний опис результатів концептуального моделювання предметної області, представлена у формі, яка сприймається людиною і комп'ютерною системою.

Об'єктом дослідження є розробка покращеного алгоритму пошуку інформації у сукупності документів. Необхідність вдосконалення процесу пошуку обумовлена наступними причинами:

1) Неструктурований характер інформації більшості електронних документів. Неструктуровані дані становлять більшу частину інформації, з якою мають справу користувачі. Це – не менш 80-90% інформації, а 10-20% – це структуровані дані.

2) Необхідно зробити пошук динамічним і зручним для користувача. Для будь-якого типу запиту, що виникає в практичній діяльності, повинні бути знайдені адекватні знання в інформаційному просторі.

3) Експонентне зростання кількості документів. При збільшенні простору пошуку пропорційно зростає і кількість документів у відгуку пошукової системи.

4) Відсутність стандартизованих механізмів семантичного індексування.

Для вирішення подібного роду проблем з документом необхідно пов'язати метадані, що дозволяють інтерпретувати й обробляти інформацію, яка зберігається в цьому документі, тобто включити в документ інформацію, яка описує структуру і семантику його змісту. Пропонується використання семантичного пошуку як одного з можливих варіантів вирішення проблем пошуку інформації. Онтологічна модель може бути використана для повнотекстового пошуку і для окремої класифікації. Для побудови онтології потрібно формальне декларативне подання чітко організованих конструкцій, які містять у собі словник термінів тематичної області, опис визначень цих термінів, існуючі взаємозв'язки між ними, і взагалі – теоретично можливі й неможливі взаємозв'язки.

Обов'язковим у системі пошуку повинен бути процес “самонавчання” системи. Вважається, що цей процес дозволить ліквідувати ситуації з термінами або назвами, які записуються некоректно у базі даних.

Важливість дослідження авторів цих тез, пов'язаного з онтологіями, обумовлена також тим, що знання, яке не описане і не тиражоване, в кінцевому рахунку стає застарілим і непотрібним. А знання, яке поширюється, є генератором нових знань.

Звідси виникає ще один з майбутніх напрямків дослідження – вирішення проблеми старіння інформації.

Авторами здійснюється процес практичної реалізації програмних засобів семантичного пошуку, який би надавав результати не тільки за заданими словами з запиту, але й за еквівалентними за сенсом. При цьому приділяється увага засобам пошуку по тексту на основі шаблонів – регулярним виразам. Шаблон описує закономірність, якій повинні підкорятися шукані послідовності символів у тексті.

На даний момент розроблено деякі компоненти системи, функціональне призначення яких – надавати можливість пошуку в неструктурованих текстових документах за різними критеріями, а також формувати регулярні вирази та N-грами.

### **Література**

1. Губин М.В. Исследование качества информационного поиска с использованием пар слов // Труды RCDL-2003 – 2003 – с. 186-191.

2. Маннинг К.Д., Рагхаван П., Шютце Х. Введение в информационный поиск: пер. с англ. – М: ООО “И.Д. Вильямс” – 2011. – 528 с.

3. Шатовская Т., Каменева И. Интегрированный подход текстовой кластеризации для неструктурированных документов/ Т.Шатовская, И.Каменева // “INTERNET – EDUCATION – SCIENCE”: материалы 6-й Международной конференции (Винница, Украина, 7-11 октября, 2008 г.). Винница – 2008. – с.504-506.