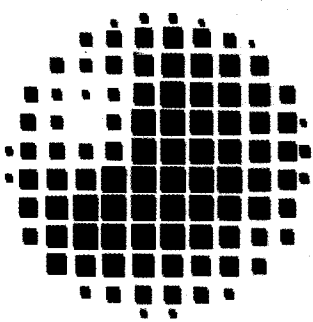




науковий журнал

НАУКОВО – ТЕХНІЧНА ІНФОРМАЦІЯ

SCIENTIFIC AND TECHNICAL INFORMATION

3(61)'2014



УДК 004.89

МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО ПОШУКУ Й АНАЛІЗУ ІНФОРМАЦІЇ НА ЕТАПАХ ЖИТТЄВОГО ЦИКЛУ ІНФОРМАЦІЙНИХ СИСТЕМ



Ю.Б. Беляєв, *докт. техн. наук*,
С.В. Запара

Постановка проблеми. Найбільш поширеною формою подання і збереження інформації є текстові документи на природній мові. Такий спосіб подання полегшує їхнє сприймання, тиражування й модифікування, а це спричинює швидке зростання обсягів інформаційних ресурсів (ІР). Упровадження

в життя комп'ютерних систем і мереж призвело до накопичення і постійного зростання великих обсягів ІР в електронному вигляді, що ускладнює доступ до необхідних даних і знань, коли в них виникає потреба.

У зв'язку з цим набувають актуальності системи пошуку інформації, виявлення даних

і знань.

Аналіз проблеми й існуючих рішень. Життєвим циклом інформаційної системи (ІС) вважається сукупність стадій і етапів, які вона проходить у своєму розвитку від моменту прийняття рішення про її створення до моменту припинення функціонування. Існує велика кількість стандартів і методик, які регламентують як життєвий цикл ІС у цілому, так і окремі його етапи [1]. Ми будемо вважати, що життєвий цикл ІС складається з таких етапів:

1. Виявлення й аналіз вимог – системний аналіз. Здійснюються дослідження й аналіз існуючих систем, виявляються вимоги до ІС, формуються техніко-економічне обґрунтування і технічне завдання на розробку ІС.

2. Проектування (технічне і логічне). Відповідно до вимог формуються переліки функцій системи (функціональна архітектура) і підсистем (системна архітектура), оформляється технічний проект ІС.

3. Реалізація (робоче і фізичне проектування, кодування). Розробка та налаштування програм, формування і наповнення баз даних, складання робочих інструкцій для персоналу, оформлення робочого проекту.

4. Упровадження (дослідна експлуатація). Комплексне налаштування підсистем ІС, навчання персоналу, поетапне впровадження ІС в експлуатацію по підрозділах організації, оформлення нормативних документів про впровадження в експлуатацію ІС.

5. Експлуатація ІС (супровід, модернізація). Збирання рекламаций і статистики функціонування ІС, виправлення помилок і недоробок, виявлення вимог до модернізації ІС і її виконання (повторення етапів 2–5).

На всіх етапах життєвого циклу ІС як розробникам, так і персоналу потрібно вирішувати задачі інформаційного пошуку: пошук джерел знань і виявлення знань предметної області (ПО)[2] (етапи 1–2); виявлення інформації для формування і наповнення баз даних ІС (етап 3); пошук і аналіз змін нормативно-

правового й інформаційного забезпечення ІС (етапи 4–5) тощо. Існує велика кількість методів інформаційного пошуку, аналізу інформації, виявлення даних і знань [3; 4]. Розроблені інструментальні засоби, у тому числі інформаційно-пошукові системи Інтернету, що можуть бути використані для вирішення окремих задач життєвого циклу ІС [5; 6]. Але на сьогодні відсутні інструментальні засоби і методики, які б дали змогу вирішувати ці задачі комплексно, на всіх етапах життєвого циклу ІС, тобто склалася ситуація, коли в Інтернеті і корпоративних мережах міститься величезна кількість інформації у вигляді текстів на природній мові. Отже, є можливість отримувати спеціалізовані видання і книги в електронному вигляді, але за існуючих інструментальних засобів обробки таких документів доводиться витрачати багато часу і зусиль для того, щоб здобути знання, необхідні для вирішення практичних завдань. Парадоксальність ситуації полягає в тому, що чим більші обсяги інформації в електронному вигляді отримує людина, тим більш недоступними стають знання. За відсутності адекватних і ефективних засобів обробки текстів на природній мові користувач змушений прочитувати великі масиви текстових документів, що досить складно. Отже, виникає необхідність скорочувати масив документів за рахунок виключення малоінформативних, повторюваних, неактуальних тощо. При цьому зростає ризик відбраковування корисних інформативних або навіть ключових документів.

Для вирішення цієї проблеми потрібно мати технології й інструменти виявлення даних і знань з масивів текстів на природній мові. Специфічним для України є те, що джерела даних і знань можуть бути подані як українською, так і російською мовами. У зв'язку з цим необхідно відмітити декілька фактів, які були нами виявлені, зокрема:

- інструментальні засоби обробки текстових документів, складених українською мовою, практично відсутні, за винятком

інформаційно-пошукових систем Інтернету;

- публікацій в Інтернеті стосовно комп'ютерної обробки текстів українською мовою вкрай мало;

- україномовні ресурси в електронному вигляді, у вільному доступі (словники, корпуси текстів тощо) практично відсутні.

Такий стан речей можна пояснити як відсутністю вказаних засобів і ресурсів, так і слабким поданням їх в Інтернеті.

Рішення, що пропонується. На основі аналізу потреб в інформаційному забезпеченні етапів життєвого циклу ІС та існуючих підходів до пошуку і аналізу інформації розроблено структуру системи пошуку інформації і виявлення знань з інформаційних ресурсів (див. рисунок).

Система складається з двох основних частин: пошукової машини (І); системи аналізу текстів і виявлення знань (ІІ).

Пошукова машина (ПМ) складається з підсистем: пошуку інформації в Інтернеті (ППІ МІ); пошуку інформації в корпоративній мережі (ППІ КМ); спостереження за діями

експерта (ПС ДЕ); спостереження за змінами документів предметної області (ПС ЗД); база даних (БД).

Підсистема аналізу текстів і виявлення знань, складається з таких підсистем:

сегментатор [7]; морфологічний аналізатор; синтаксико-семантичний аналізатор; база знань (БЗ).

Основною відмінністю системи, що пропонується, від аналогічних [5] є наявність підсистеми спостереження за діями експерта. Це надасть можливість отримувати первинний набір текстів предметної області, необхідний для налаштування і навчання системи, без чого її робота буде неможлива.

Слабким місцем при розробці і впровадженні ІС є взаємодія між аналітиками й експертами, але наявність підсистеми спостереження надасть можливість поліпшити їхню взаємодію.

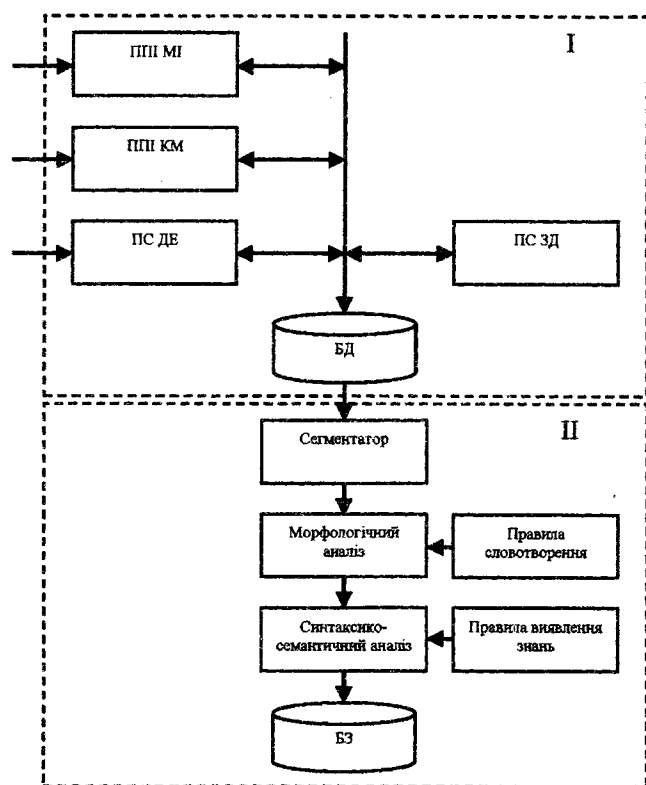
Згідно з класифікацією [6] система, що розробляється, належить до автоматизованих систем управління знаннями (АСУЗ).

У перспективі намічається отримати базу знань ПО [8], яка може бути використана як на етапі розробки ІС, так і на етапі експлуатації як складова частина ІС; скоротити час проведення етапу аналізу ПО, у тому числі і за рахунок скорочення часу спілкування між основними учасниками процесу; підвищити якість розробки і функціонування ІС за рахунок більш точного виявлення вимог до системи; накопичувати знання і дані ПО під час роботи над різними проектами.

Використання вказаних інструментальних засобів передбачається як доповнення до існуючих засобів аналізу ПО.

Висновки

Зважаючи на вищенаведене, дослідження методів інтелектуального пошуку й аналізу інформації, особливо методів аналізу документів на природній мові, є перспективним напрямом.



Структура системи пошуку та виявлення знань

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. *Maiaşek Leszek A.* Анализ требований и проектирование систем. Разработка информационных систем с использованием UML.; пер. с англ. – М.: «Вильямс», 2002. – 32 с.
2. *Гаерилова Т.А., Хорошевский В.Ф.* Базы знаний интеллектуальных систем. – СПб: «Питер», 2000. – 384 с.
3. *Christopher D. Manning, Prabhakar Raghavan, Hinrich Rütze,* Introduction to Information Retrieval. – Cambridge University Press, 2008 – 496 p.
4. *Ермаков А.Е.* Автоматизация онтологического инжиниринга в системах извлечения знаний из текста // Компьютерная лингвистика и интеллектуальные технологии: материалы Международной конференции «Диалог'2008», т. 7 (14). – М.: РГГУ, 2008. – С. 154–158.
5. *Бабин Д.В., Вороной С.М.* Повышение эффективности извлечения знаний на основе интеллектуального анализа и структурирования информации // Искусственный интеллект. – 2005. – № 3. – С. 259–264.
6. *Ермаков А.Е.* Извлечение знаний из текста и их обработка: состояние и перспективы / А.Е. Ермакова // Информационные технологии. – 2009. – № 7. – С. 50–55.
7. *Скатов Д.С., Ерехинская Т.Н., Окатьев В.В.* Модели и методы анализа иерархически структурированных текстов // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции «Диалог'2009». – М.: Наука, 2009.
8. Інтелектуальні системи підтримки прийняття рішень: навч. посіб. [Б.М. Герасимов, В.М. Локазюк, О.Г. Оксіюк, О.В. Поморова]. – К.: Видавництво Європейського університету, 2007. – 335 с.