*та інформа. – 2003. – Вип. 4. – С. 191–211. 6. Jean-Marc Adamo. Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel algorithms. Springer-Verlag. 2000. 7. Szczuka M. Rules as attributes in classifier construction // Proceedings of RSFDGrC'99, Yamaguchi, Japan, LNAI 1711, Springer-Verlag, Berlin. – P. 492–499. 8. Slowinski K., Stefanowski J. On limitations of using rough set approach to analyse non-trivial medical information systems // Proceedings of the 4th Int. Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, Tokyo 1996. – P. 176–184. 9. Richard O. Duda. Pattern Classification. John Wiley&Sons, Inc. 2000. 10. Christopher M. Bishop. Neural Networks for Pattern Recognition.Oxford University Press. 2003. 11. Komorowski J., Pawlak Z., Polkowski L., Skowron A. Rough Sets: A Tutorial // A New Trends in Decision Making, Springer-Verlag, 1999. – P. 3–98.*

**УДК 004.652.4+004.852**

**Ю.М. Оградіна, П.Ф. Павлов,**
**Р.В. Разуваєв, В.В. Шепелєв, П.І. Кузьменко**
Харківський національний університет радіоелектроніки,
кафедра "Соціальна інформатика"

# РОЗРОБКА СИСТЕМИ АНАЛІЗУ НАПІВСТРУКТУРОВАНИХ ІНФОРМАЦІЙНИХ ПОВІДОМЛЕНЬ

*© Оградіна Ю. М., Павлов П.Ф., Разуваєв Р.В., Шепелєв В.В., Кузьменко П.І., 2006*

***Paper describes the program system for analysis of natural language messages about emergencies. Using data mining methods (association rules) for analyzing data are proposed.***

***Описано програмну систему аналізу повідомлень природною мовою про надзвичайні ситуації. Для проведення аналізу пропонується використовувати методи інтелектуального аналізу даних, зокрема побудову асоціативних правил***

На початку XXI століття, з різким, «вибуховим» збільшенням розмірів баз даних та інформаційних потоків зросла актуальність ефективного використання даних. Без розвитку засобів і методів опрацювання інформації велика її частка може стати "мертвим вантажем", що "засмічує" канали передачі і носії.

До початку 90-х років для цих цілей застосовувалася, в основному, прикладна статистика. Однак її методи слабко пристосовані до реальних задач, оскільки оперують умовними "усередненими" величинами. Крім того, традиційні статистичні методи в силу своєї числової природи не можуть дати інструмента ані для аналізу текстової інформації, ані для пошуку складних, неочевидних залежностей між різними даними [1]. Виникнення інтелектуального аналізу даних (ІАД, Data Mining) по праву вважається продуктом природної еволюції інформаційних технологій в останні роки [2].

У роботі досліджується задача опрацювання та аналізу повідомлень про надзвичайні ситуації (НС), що були зібрані Міністерством Російської Федерації в справах Цивільної оборони, надзвичайних ситуацій та ліквідації наслідків стихійних лих (МНС РФ). Ці повідомлення складають великий обсяг слабко структурованої інформації, що викладена

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

природною мовою. У статті розглядаються методи та засоби технологій Data Mining, Text Mining і їх застосування для аналізу повідомлень. Також описується програмна система, розроблена з використанням цих методів.

## 1. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ ТА ТЕКСТІВ

За визначенням Г. П'ятецькі-Шапіро, "Data Mining – це процес виявлення в "сирих" даних раніше невідомих, нетривіальних, практично корисних, доступних для інтерпретації знань" [1]. Алгоритми, що використовуються в Data Mining, вимагають великої кількості обчислень. Раніше це було чинником, що стримував широке практичне застосування Data Mining, однак сьогоднішнє зростання продуктивності сучасних процесорів зняв гостроту цієї проблеми [3].

Задачі, що розв'язуються методами Data Mining [3]:

**1. Класифікація** – це віднесення об'єктів (подій) до одного із заздалегідь відомих класів.

**2. Кластерізація** – це групування об'єктів (подій) на основі даних (властивостей), що описують сутність об'єктів. Об'єкти всередині кластера повинні бути "схожими" один на одного і відрізнятися від об'єктів, що ввійшли до інших кластерів.

**3. Регресія**, **прогнозування –** це встановлення функціональних залежностей між залежними і незалежними змінними.

**4. Асоціація** – виявлення закономірностей між подіями.

**5. Послідовність** – установлення закономірностей між зв'язаними в часі подіями.

**6. Аналіз відхилень** – виявлення найбільш нехарактерних випадків.

Технологія ІАД призначена для аналізу структурованих даних. Однак, більшу частину (близько 80 %) наявних даних становлять текстові неструктуровані дані, наприклад, Web-сторінки, електронна пошта, патенти, документи та багато іншого. Тому збільшується кількість досліджень з використання технологій Data Mining для аналізу текстових даних. Цей напрямок отримав назву Text Mining.

Щодо значення цього терміна (Text Mining, видобуток даних з тексту) дослідники не прийшли до єдиної думки. Так, в New Zealand Digital Library стверджується, що "Text Mining – це виявлення закономірностей (patterns) у тексті природною мовою, іншими словами, видобуток інформації з нього (тексту) для конкретних цілей" [4]. Поширено також "пошукоцентричне" тлумачення цього терміна – "щось, що допомагає знаходити інформацію в Мережі" [5]. Професор Marti A. Hearst пояснює його як "масиво-орієнтовану обчислювальну лінгвістику" [5].

Далі будемо називати Text Mining (чи інтелектуальним аналізом тексту, ІАТ) процес, що дозволяє виявити в тексті природною мовою раніше невідомі, нетривіальні, доступні для інтерпретації закономірності. ІАТ можна визначити як відшукання смислових патернів, угруповань інформації, що несуть зміст, в неструктурованих, розрізнених текстових сукупностях. Ця технологія дозволяє аналітику використовувати обчислювальну техніку для аналізу тексту дуже великого обсягу, раніше недоступного; та дає можливість без читання величезної маси файлів досліджувати поняття, що являють зміст сукупності документів. Найбільш ефективним використання ІАТ буде для науково-технічної інформації, де тексти мають явне, недвозначне тлумачення, слова вживаються в прямому, а не переносному значенні.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

На даний час існують не тільки дослідницькі системи, такі як Patent Miner [6] та Athena [7], розроблені в IBM Almaden Research Center, система Phrasal Frontier [8, 9], розроблена в університеті Хельсінкі, але й комерційні системи, наприклад TextMiner, розроблений *SAS Institute,* та LexiQuestMine компанії SPSS (www.spss.com).

Система Patent Miner призначена для аналізу текстових БД. Її тестування проводилось на БД патентів U.S.Patent, що видавалися в різних категоріях. Для аналізу даних використовувалися запити на основі форми (shape-based query). Система Athena призначена для побудови та експлуатації ієрархічної організації текстових документів електронної пошти в діалоговому режимі. Для забезпечення виконання основних функцій системи використовуються два блоки – класифікації та кластерізації. Для класифікації текстів використовується простий баєсівський класифікатор (Naive Bayes classifier), для кластеризації алгоритм C-Evolve [7].

Система TextMiner призначена для порівняння граматичних та словесних рядів у письмовій мові людини, з якою користувач спілкується електронною поштою, з тим, що було написано раніше, та виявляти підозрілі розбіжності.

Система LexiQuestMine є доповненням до системи Clementine, що дозволяє використовувати текстову інформацію як джерело даних для проведення ІАД. Система LexiQuestMine будує перелік ключових понять, що містяться в тексті на основі контексту документів, потім автоматично класифікує терміни за групами, наприклад, продукти, організації чи люди.

Слід зазначити, що за прогнозами аналітичної компанії IDC, попит на програми Text Mining суттєво зросте в найближчі 4–5 років. До 2005 року очікується збільшення прибутків від такого програмного забезпечення з $540 млн. (в 2002 році) до 1,5 мільярдів доларів.

## 2. ПОСТАНОВКА ЗАДАЧІ

Необхідно побудувати програмну систему для аналізу коротких напівструктурованих інформаційних повідомлень і виявлення в них існуючих залежностей.

***Вхідні дані*** – набір коротких напівструктурованих текстових повідомлень природною мовою.

Джерелом інформації є офіційний сайт МНС РФ. Типове повідомлення має такий вигляд:

*ДАТА ЧС: ВРЕМЯ ЧС:*
*МЕСТО ЧС:*
*ТИП ЧС:*
*Короткий текстовий опис*

***Вихідні дані*** – існуючі закономірності та залежності між інформаційними одиницями текстів.

Для вирішення поставленої задачі необхідно:

- проаналізувати існуючі методи та алгоритми ІАД та обрати алгоритм, що відповідає вимогам задачі;
- проаналізувати існуючі методи та алгоритми морфологічного та синтаксичного аналізу тексту та обрати алгоритми, що відповідають вимогам ІАД;
- розробити схему взаємодії підсистем ІАД та обробки природної мови;
- програмно реалізувати розроблені алгоритми.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

Основні очікувані результати роботи системи:

1. Виявлення слів та словосполучень, які зустрічаються найчастіше. Це може бути корисно для автоматичного формування словника предметної області, що може стати основою для подальшого розвитку цієї та подібних систем.

2. Виявлення найчастіших взаємозв'язків між словами та словосполученнями, що може бути основою для виявлення шаблонів типових повідомлень.

3. Виявлення прихованих залежностей між різними повідомленнями, зв'язаними чи то місцем, де вони відбулися, чи часом відбування, чи типом ситуації.

## 3. ЗАГАЛЬНИЙ ОПИС МЕТОДУ

### 3.1. Аналіз природномовної інформації

Важливим кроком проведення ІАТ є опрацювання природномовної інформації, яка може займати значну частину часу проведення всього процесу аналізу. Від ефективності реалізації процедур аналізу природної мови багато в чому залежить ефективність проведення ІАТ взагалі [8, 9].

У цій роботі реалізовано два етапи аналізу текстів природною мовою – морфологічний та синтаксичний аналіз.

Морфологічний аналіз необхідний для отримання морфологічної інформації про кожне слово у тексті. Як морфологічна інформаця виступають синтаксичні класи слів (прислівник, прикметник, дієслово і т. п.), форми слів у цих класах (наприклад, у прикметника – відмінок, рід, число).

Слова у синтаксичних класах можна також розбити на підкласи – флективні класи. Флективні класи виділяються за допомогою аналізу закінчень (флексій) слів у синтаксичних класах [12]; це означає, що у всіх слів в одному флективному класі одна і та ж система закінчень, яка відрізняється від системи закінчень слів будь-якого іншого флективного класу. Отже, для кожного слова достатньо лише з'ясувати, до якого синтаксичного та флективного класу воно належить, і, виходячи з цього, можна отримати будь-яку форму синтаксичного класу цього слова. Винятки становлять лише ті слова, у яких зміна форми призводить до зміни основи і суфікса. Знання про флективні класи допомагає у створенні словника, який надалі буде використовуватися у морфологічному аналізі.

Синтаксичний аналіз передбачається використовувати тільки для виявлення іменних словосполучень, а не для повного синтаксичного розбору речення. Під іменними словосполученнями маються на увазі сполучення слів, у яких "головним словом (основним носієм змісту) є, як правило, перший ліворуч іменник, а інші слова служать для уточнення значення головного слова" [12]. У цих словосполученнях виражені поняття, що позначають "різного роду об'єкти, їхні ознаки, значення ознак і т. п." [12]. Припускаємо, що інтелектуальний аналіз таких понять нам дозволить виявити більше інформації, ніж аналіз набору слів.

В іменних словосполученнях нас цікавлять не самі другорядні слова, а вид зв'язку їх із головним словом. Наприклад, при синтаксичному аналізі було виявлено два словосполучення: "прорив нафтопроводу" і «прорив газопроводу». Тут головне слово – «прорив», а другорядні слова – "нафтопровід" і "газопровід". Для ІАД ці словосполучення будуть передані у вигляді: головне слово ("прорив") – вид зв'язку слів ("уточнення головного слова іменником у родовому відмінку") – і другорядне слово ("нафтопровід" чи "газопровід"). Другорядні слова будуть лише задавати приблизну область значень цього виду зв'язку.

Отже, застосування морфологічного і синтаксичного аналізу текстів повідомлень є ефективними методами попереднього опрацювання текстових описів НС, що не потребує

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

задавати всі можливі види словосполучень та структуру повідомлень, які можуть згодом змінюватися; так само немає гарантії, що "вручну" вдасться відшукати всі подібні словосполучення та правильно їх класифікувати.

### 3.2. Побудова асоціативних правил

Асоціативні правила (Association Rules) дозволяють знаходити закономірності між зв'язаними подіями. Перший алгоритм пошуку асоціативних правил, що називався AIS, було розроблено у 1993 році співробітниками дослідницького центру IBM Almaden. З того часу зріс інтерес до асоціативних правил; на середину 90-х років минулого століття припадає пік дослідницьких робіт у цій області, і з тих пір щороку з'являлося по декілька алгоритмів [11].

При використанні методу асоціативних правил метою аналізу є встановлення залежностей виду: якщо в структурній одиниці даних зустрівся деякий набір елементів X, то на підставі цього можна зробити висновок про те, що інший набір елементів Y також має з'явитися в цій одиниці. Ці правила мають такий вигляд (1):

$$X \to Y \tag{1}$$

Алгоритми пошуку асоціативних правил призначені для пошуку всіх правил виду (1), причому для кожного з них визначаються такі величини, як **підтримка** (скільки разів у всій множині зустрічаються одиниці даних, що містять X та Y одночасно) і **вірогідність** (який відсоток від всіх одиниць, що містять X, містить також і Y). Їх формули (2), (3), відповідно:

$$supp(X \to Y) = P(X, Y), \tag{2}$$

$$conf(X \to Y) = \frac{supp(X \to Y)}{supp(X)}. \tag{3}$$

де $P(X, Y)$ – ймовірність того, що $X$ та $Y$ зустрінуться в одній транзакції. Ці величини мають бути більшими за деякі наперед визначені пороги, називані відповідно мінімальною підтримкою (minsupport) і мінімальною вірогідністю (minconfidence) [11].

З нашого погляду цей метод найбільше відповідає поставленій задачі. У роботі був застосований найбільш ефективний з алгоритмів побудови асоціативних правил – Apriori. Була розроблена модифікація, що враховує специфіку даних.

Етапи розробленого нами алгоритму та оригінального алгоритму Apriori для порівняння зведені у таблиці. Зауважимо, що під набором елементів, які часто зустрічаються, ми будемо розуміти такий набір, підтримка якого більша за мінімальну.

**Алгоритм видобування асоціативних правил Apriori та його модифікація**

| Етап | Оригінальний алгоритм | Модифікований алгоритм |
|---|---|---|
| 1 | 2 | 3 |
| Формалізація даних | Кожна *транзакція* (набір *елементів*) перетворюється з вигляду $T=\{a_1,a_2,\ldots a_n\}$ на двійкову послідовність, яка має довжину, що дорівнює кількості елементів усіх транзакцій, причому на місці елемента, який присутній у транзакції, стоїть 1, а на місці елемента, який відсутній – 0 | Кожна *фраза* (набір *лексем*) з текстового вигляду перетворюється на двійкову послідовність, причому допоміжні слова, лексеми та символи відкидаються. Двійкова послідовність зберігається у пам'яті комп'ютера як послідовність бітів |

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

**Алгоритм видобування асоціативних правил Apriori та його модифікація** (*продовження*)

| 1 | 2 | 3 |
|---|---|---|
| Пошук 1-елементних наборів, що часто зустрічаються | Порівняння кожного елемента з кожною транзакцією для підрахування кількості входжень | Підсумовування усіх бітових наборів “у стовпчик”. Результат кожного “стовпчика” є частотою входження відповідного елементу до набору транзакцій |
| Пошук k+1-елементних наборів, що часто зустрічаються | *Примітка: за властивістю **антимонотонності** часто зустрічатися k+1-елементний набір може лише тоді, коли часто зустрічаються усі k-елементні набори, що до нього входять.* | |
| | 1. Побудова k+1-елементних наборів<br>2. Вилучення таких, у яких є k-елементні набори, що нечасто зустрічаються<br>3. Розрахунок підтримки для кожного набору (за допомогою хеш-дерева)<br>4. Вилучення таких, у яких значення підтримки нижче за мінімальну підтримку | Такі ж кроки, але з використанням побітових операцій (***і***, ***або***, ***не***), у тому числі й розрахунок підтримки |
| Побудова асоціативних правил | 1. Побудова правил<br>2. Розрахунок їх вірогідності<br>3. Вилучення таких, у яких значення вірогідності нижче за мінімальну вірогідність | Такі ж кроки, але крім того, для кожного набору правил, побудованого з одного набору елементів, вилучаються усі, окрім одного, що має максимальну підтримку |

## 4. ОПИС ПРОГРАМИ

Програма-аналізатор повідомлень про НС призначена для пошуку прихованих закономірностей у вигляді асоціативних правил. Стосовно предметної області це мають бути залежності вигляду: *набір лексем* $\rightarrow$ *набір лексем* (*підтримка*, *вірогідність*). Значення терміну “набір лексем” у даному контексті буде визначено нижче.

Процес аналізу повідомлень складається з таких етапів:

1. Підготовка словникової бази для аналізу тексту.

2. Попереднє опрацювання масиву текстів повідомлень з метою усунення зайвої (службової) інформації (препроцесінг).

3. Морфологічний аналіз підготованого тексту.

4. Синтаксичний аналіз тексту.

5. Побудова асоціативних правил.

Програма призначена для вивчення ефективності та застосовності обраних методів для вирішення поставленої задачі. Система має варіанти функціонування, коли етапи морфологічного чи синтаксичного аналізу можуть бути пропущені. Це зроблено для того, щоб виявити, як і якою мірою кожен із цих етапів впливає на остаточний результат. Структура системи наведена на рис. 1.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua


*Рис. 1. Структура системи аналізу текстових повідомлень*

Система зроблена у вигляді так званого майстра (wizard), тобто користувач проводить процес аналізу покроково з можливістю перегляду проміжних результатів та вибору наступних кроків.

**Крок 1. Завантаження файлів.** Користувач обирає файли, у яких є повідомлення про надзвичайні ситуації. Можна обрати один або більше файлів. На екрані відображається вміст обраних файлів у неопрацьованому вигляді.

**Крок 2. Препроцесінг.** Файл чи файли з інформаційними повідомленнями опрацьовуються з метою усунення зайвої інформації. Зайвою є службова інформація (наприклад, зворотна адреса служби розсилки файлів і т. ін.).

Кожен файл містить три основні розділи:

1) зведення про надзвичайні ситуації за останню добу;

2) надзвичайні ситуації, які знаходяться на контролі;

3) інші питання.

До аналізу був обраний лише розділ 1, інші розділи усуваються. Результати попереднього опрацювання наведені на рис. 2.

**Крок 3. Морфологічний аналіз.** Кожне слово опрацьовується за допомогою аналізу складу слова та порівняння його з наявними словниками. Визначається частина мови, до якої належить слово, його базова форма (для іменника – іменний відмінок, для дієслова – однина теперішнього часу і т. д.) та відмінювання. Цей крок може бути пропущеним. Тоді синтаксичний аналіз також пропускається.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

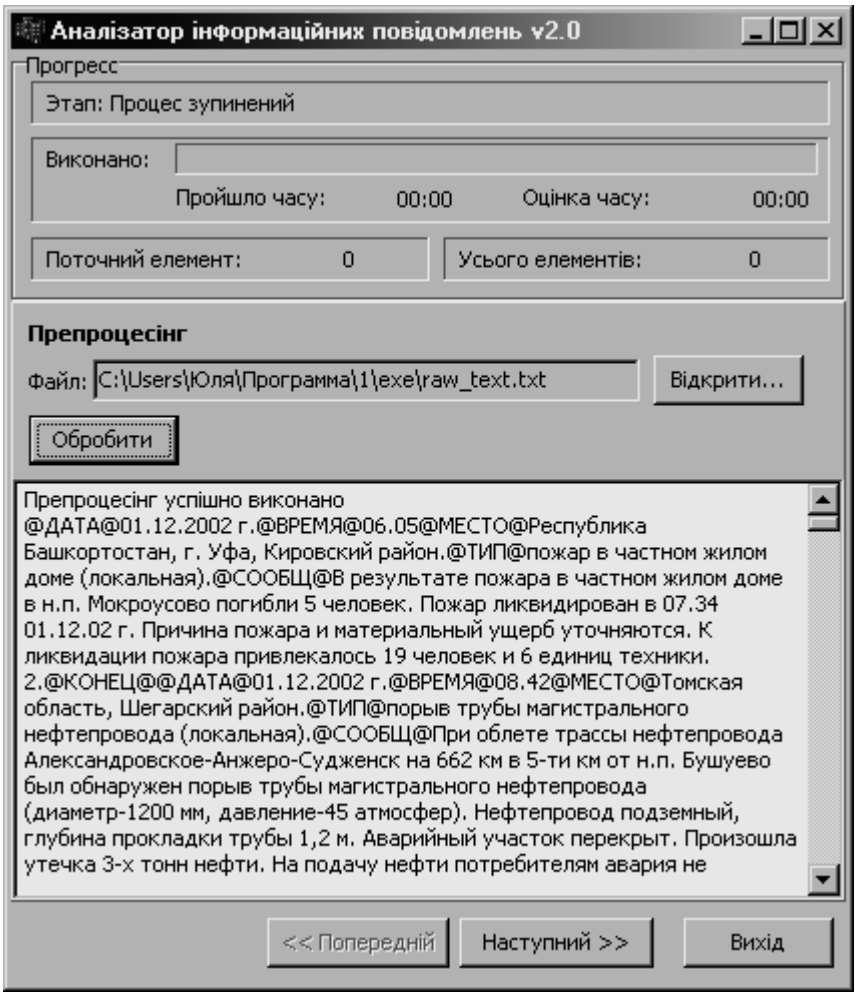


*Рис. 2. Результати попереднього опрацювання*

**Крок 4. Синтаксичний аналіз.** Визначаються так звані іменні словосполучення – такі словосполучення, у яких головним словом є іменник. Для побудови словосполучень використовуються результати морфологічного аналізу. Цей крок може бути пропущеним.

**Крок 5. Побудова асоціативних правил.** Для цього, по-перше, знаходяться найчастіші набори лексем. Набором лексем залежно від того, чи був пропущений якийсь з кроків 3–4 може бути:

- набір слів з попередньо опрацьованого тексту;
- набір слів, що виділені морфологічним аналізом;
- набір іменних словосполучень.

На знайдених наборах будуються правила вигляду: *набір лексем1* → *набір лексем2* (*підтримка*, *вірогідність*). Це може бути сприйнято як: “Якщо у повідомленні зустрівся *набір лексем1*, то в цьому ж повідомленні зустрінеться *набір лексем2* з імовірністю *вірогідність*. Це сполучення зустрічається у *підтримка* % усіх повідомлень”. Для обмеження кількості знайдених правил встановлюються пороги *мінімальної вірогідності* та *мінімальної підтримки*. Правила, параметри яких нижче встановлених, не враховуються. *Мінімальна вірогідність* відсіює правила, які не є закономірностями, а лише випадковим

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

угрупуванням лексем. *Мінімальна підтримка* відсіює правила, котрі зустрічаються настільки нечасто, що неможливо визначити, чи вони є закономірностями.

Результати роботи програми наведені на рис. 3.

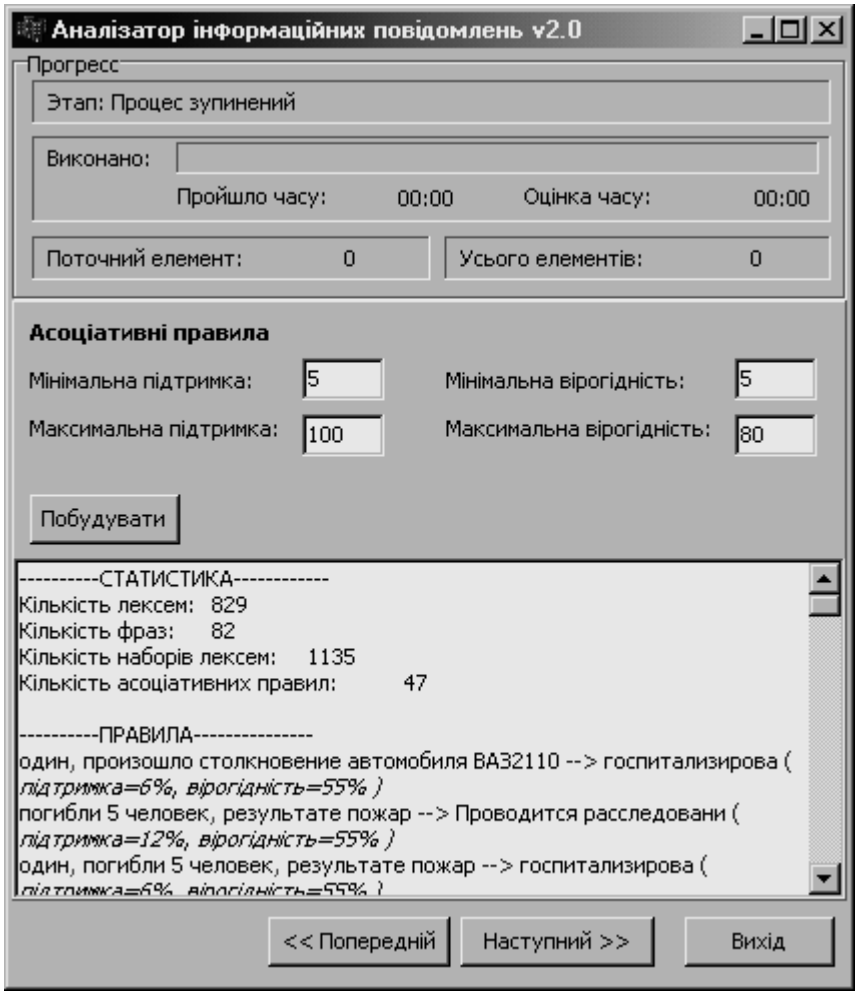


*Рис. 3. Результати інтелектуального аналізу*

## ВИСНОВКИ

Отже, побудована система дозволяє виявляти існуючі в даних закономірності, які є певним узагальненням досвіду, що відображений в описах НС. Аналіз отриманих закономірностей може покращити розуміння проблемної області і сприяти формуванню ефективніших розв'язань задач, що виникають в цій області.

Також показано корисність мовно-орієнтованих засобів аналізу тексту (морфологічного та синтаксичного аналізаторів). Застосування морфологічного аналізатора дозволяє різні форми одного і того ж слова розпізнавати як одне слово, використання синтаксичного аналізатора полегшує побудову асоціативних правил та покращує їх корисність, а також дозволяє зменшити час побудови правил.

Робота над системою триває, проводяться тестування, аналіз можливостей підвищення ефективності системи та оптимізації алгоритму для зменшення часу побудови правил.

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua

*1. Дюк В., Самойленко А. Data Mining: Учебный курс. – СПб: Питер, 2001. – 368 с. 2. Han J., Kamber M. Data Mining: Concepts and Techniques. – Simon Fraser University, 2000. 3. А. Шахиди. Data Mining – добыча данных // http://www.basegroup.ru. 4. New Zealand Digital Library // http://www.cs.waikato.ac.nz/~nzdl. 5. Marti A. Hearst. Untangling Text Data Mining. // http:www.sims.berkeley.edu/~hearst. 6. Lent B. Discovering trends in text databases // www.almaden.ibm.com. 7. Agrawal R., Bayardo R., Sricant R. Athena: Mining-based interactive management of text databases // www.almaden.ibm.com. 8. Ahonen H., Heinonen O., Klemettinen M., Verkamo I. Mining in the Phrasal Frontier // 1st European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD'97), Trondheim, Norway, 1997. 9. Ahonen H., Heinonen O., Klemettinen M., Verkamo I. Applying Data Mining Techniques in Text Analysis.// Report C-1997-23, University of Helsinki, Department of Computer Science, 1997. 10. Киселев М. Средства добычи знаний в бизнесе и финансах // Открытые системы. – 1997. – № 4. – С. 41–44. 11. Буров К. Обнаружение знаний в хранилищах данных // Открытые системы. – 1999. – № 5–6. – С. 67–77. 12. Белоногов Г.Г., Кузнецов Б.А. Языковые средства автоматизированных информационных систем. – М.: Наука, 1983. – 290 с. 13. Mark Dixon. An Overview of Document Mining Tecnology. 14. Шахиди А. Введение в анализ ассоциативных правил // http://www.basegroup.ru.*

**УДК 681.3**

**А. М. Пелещишин**
Національний університет "Львівська політехніка",
кафедра "Інформаційні системи та мережі"

# ПОШУК, КОНСОЛІДАЦІЯ ТА АНАЛІЗ ДАНИХ У ГЛОБАЛЬНІЙ СИСТЕМІ WORLD WIDE WEB

*© Пелещишин А.М., 2003*

***This paper considers main problems of information searching, consolidation and analysis in World Wide Web. Survey of present researchs has done. Some approaches to resolve these problems are proposed.***

***Розглядаються проблеми пошуку, консолідації та аналізу інформації, що існує у глобальній службі WWW. Зроблено огляд стану існуючих досліджень і запропоновано деякі підходи до вирішення вказаних проблем.***

## 1. ПОСТАНОВКА ЗАДАЧІ

Інтернет та глобальна система WWW сьогодні є безпрецедентним і унікальним інформаційним ресурсом, який об'єднує в собі величезні об'єми інформації. Проте сьогодні слабо розроблені реально працюючі механізми ефективного використання цього масиву інформації. Фактично існуючі системи обмежуються елементарним пошуком ресурсів, що, можливо, є релевантними запитам. Складніша синтетична обробка даних, що містяться у WWW, лише започатковується.

Кількість користувачів Інтернет зараз у світі сягає півмільярда і продовжує швидко зростати. Це зростання відбувається не тільки за рахунок традиційно високотехнологічних

Lviv Polytechnic National University Institutional Repository http://ena.lp.edu.ua