

УДК 004.78

В.М. ЛОКАЗЮК¹, В.Я. ЛЯШКЕВИЧ², О.Я. ОЛАР²¹ Хмельницький національний університет, Україна² Чернівецький національний університет імені Юрія Федьковича, Україна

РОЗРОБЛЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ОПРАЦЮВАННЯ ТЕКСТІВ З ДІАГНОСТИЧНОЮ ІНФОРМАЦІЄЮ

В статті представлено програмне забезпечення для опрацювання текстів з діагностичною інформацією. Програма реалізує взаємозв'язок ключових слів текстів із експертними знаннями, а також представляє їх у зручному вигляді для опрацювання штучною нейронною мережею, що розташована в структурі інформаційно-пошукової системи.

діагностична інформація, комп'ютерні системи, пошук інформації, експертні знання, тестування, діагностування

Вступ

Однією з найважливіших вимог щодо комп'ютерних систем (КС) є надійність їх функціонування. Забезпечення надійності КС на протязі життєвого циклу здійснюється різними заходами, зокрема, відновленням, інформаційною надлишковістю, резервуванням. Для визначення ж технічного стану КС використовуються методи та засоби тестування чи діагностування. Успішне проведення процесу тестування передбачає наявність об'єкту тестування, засобів тестування й знань інженера-діагноста щодо методів тестування [1].

Серед різноманіття інформації, котра у величезних об'ємах знаходиться в гіпермережі Інтернет або у локальних мережах різних організацій, можна виділити ту, що цікавить інженерів-діагностів для проведення тестування чи діагностування комп'ютерних систем. Характерність цього типу інформації не виходить за рамки взагалі визначення інформації та базується на трьох основних засадах технічної діагностики, які пов'язані з розпізнаванням технічного стану комп'ютерних пристроїв.

Діагностичною інформацією (ДІ) для тестування комп'ютерних систем вважатимемо таку інформацію, котра дає можливість розпізнавати технічний стан комп'ютерних систем як об'єктів діагностування [2].

Постановка задачі

Текстологічні методи пошуку інформації ґрунтуються на пошуку спеціальних текстів з підручників, монографій, статей, технічної документації та інших джерел знань з відповідної предметної галузі.

Взагалі, задачу пошуку інформації із зазначених джерел можна розглядати як задачу розуміння і виділення смислу тексту. Головними моментами процесу розуміння тексту є формування смислової структури виділення ключових слів чи "смислових віх" та прикінцеве зв'язування "смислових віх" в єдину семантичну структуру [3]. При аналізі тексту важливо виявити внутрішні зв'язки між окремими частинами тексту і поняттями. Основою для процесу пошуку діагностичної інформації є процедура розбиття тексту на частини, а потім стискання вмісту кожної смислової частини у "смилову віху". Це, по суті, відшукування набору ключових слів, що передають основний зміст тексту. У якості *ключового слова* може виступати будь-яка частина мови чи їх поєднання. Під *набором ключових слів* розумітимемо набір опорних точок, за якими розгортається текст при кодуванні і усвідомлюється при декодуванні.

Процес пошуку ДІ ґрунтується на описі об'єкта діагностування (ОД), що відповідає діагностичному забезпеченню (рис. 1), та взаємозв'язках складових

тексту, інтерпретованих інженером зі знань, із складовими опису ОД [4].

У роботі [5] представлено текстологічно-нейромережний метод пошуку діагностичної інформації, реалізація якого вимагає наявної бази знань, що використовується штучною нейронною мережею

(ШНМ). Відомо, що ШНМ не може працювати зі словами чи словосполученнями, тому виникає потреба вибрати та представити необхідні їй знання для оцінки корисності та достатності інформації, тобто представити слова та словосполучення у числовій формі, зрозумілій ШНМ.

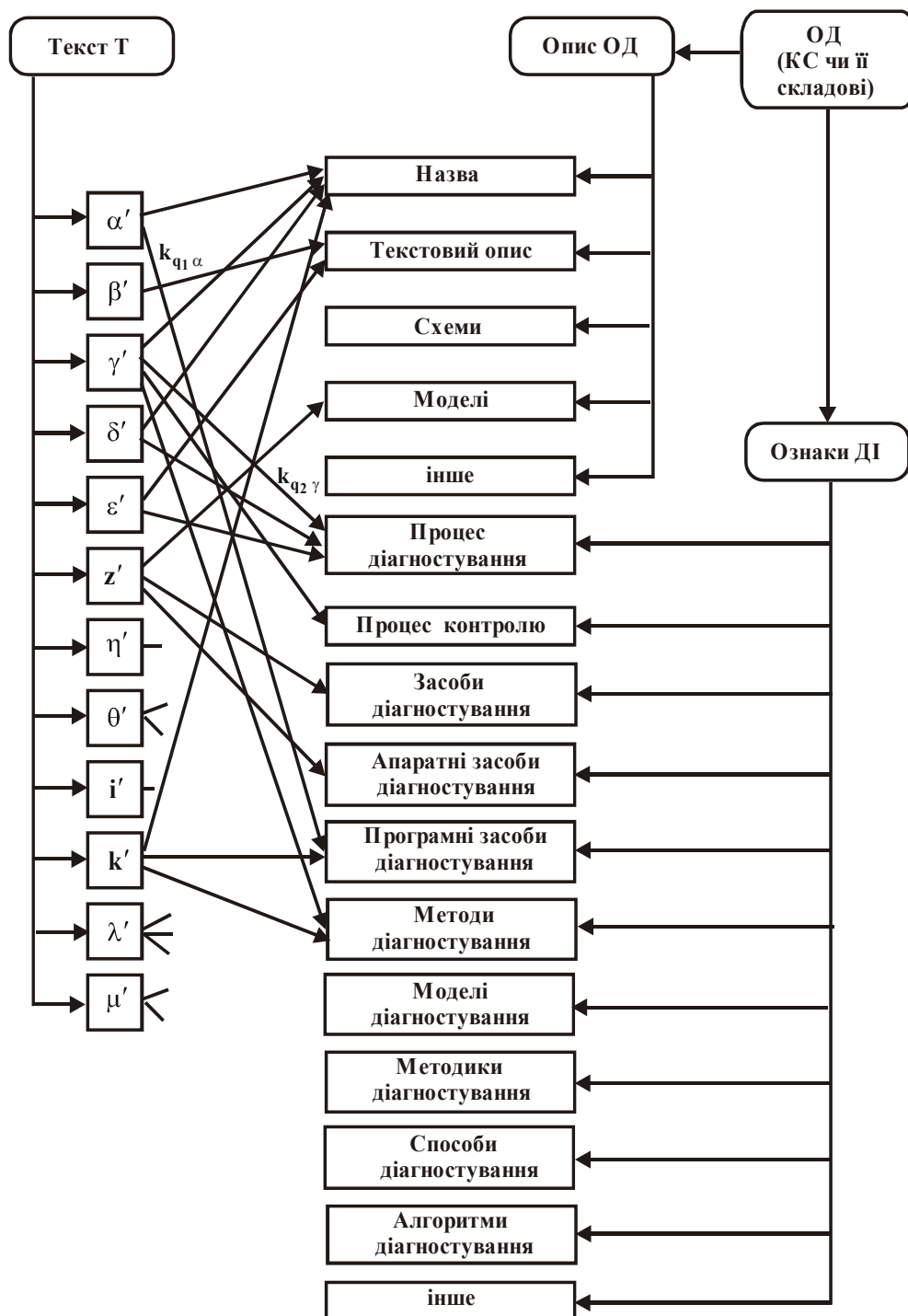


Рис. 1. Схема пошуку ДІ

Реалізація програмного забезпечення

Для реалізації програмного забезпечення опрацювання текстів з діагностичною інформацією розроблено алгоритм аналізу документів (рис. 2). Алгоритм описує порядок виконання дій експертом від моменту завантаження документа до моменту формування вхідних та вихідних навчальних векторів для ШНМ.

Відповідно до розробленого алгоритму, програмне забезпечення (ПЗ) складається з трьох модулів: Main, Edit, Zamina. Модуль Main містить 1084 командних рядки, Edit - 1007, Zamina - 58. Завантажувальний файл Analiz.exe займає 1,31 Мбайт на жорсткому диску. Експертні знання розміщені в чотирьох таблицях типу “Paradox 7”: Osn_base, Bazaznan, Statti, Koef.

Для використання розробленого ПЗ необхідно 75 Мбайт місця на жорсткому диску, 512 Мбайт оперативної пам'яті. Шлях до таблиць бази знань у вигляді Alias “Analiz”, прописується додатком BDE Administrator (Borland Delphi 6.0).

Головна форма програми Form1 описується модулем Main. Меню в головній формі програми організоване за допомогою дев'яти кнопок (об'єкт TButton), однієї мітки (об'єкт TLabel) й елемента ComboBox1 (рис. 3). Для зручності вони розташовані у різних місцях форми. Їхню функційність забезпечують відповідні процедури. Меню форми Form2 організоване за допомогою шести кнопок (об'єкт TButton) й однієї мітки (об'єкт TLabel) (рис. 4).

Процедура, викликана натисненням кнопки, може викликати додатково необхідну функцію чи процедуру.

Програма працює згідно функційної схеми, зображеної на рис. 5. Головна форма (рис. 6) надає можливість експерту встановити експертні оцінки для слів та словосполучень, вибраних у якості ключових слів, що відповідають визначенню ДІ. Оцінки виставляються від 01 до 16 відповідно до ознак ДІ (рис. 1) у правому верхньому куті форми.

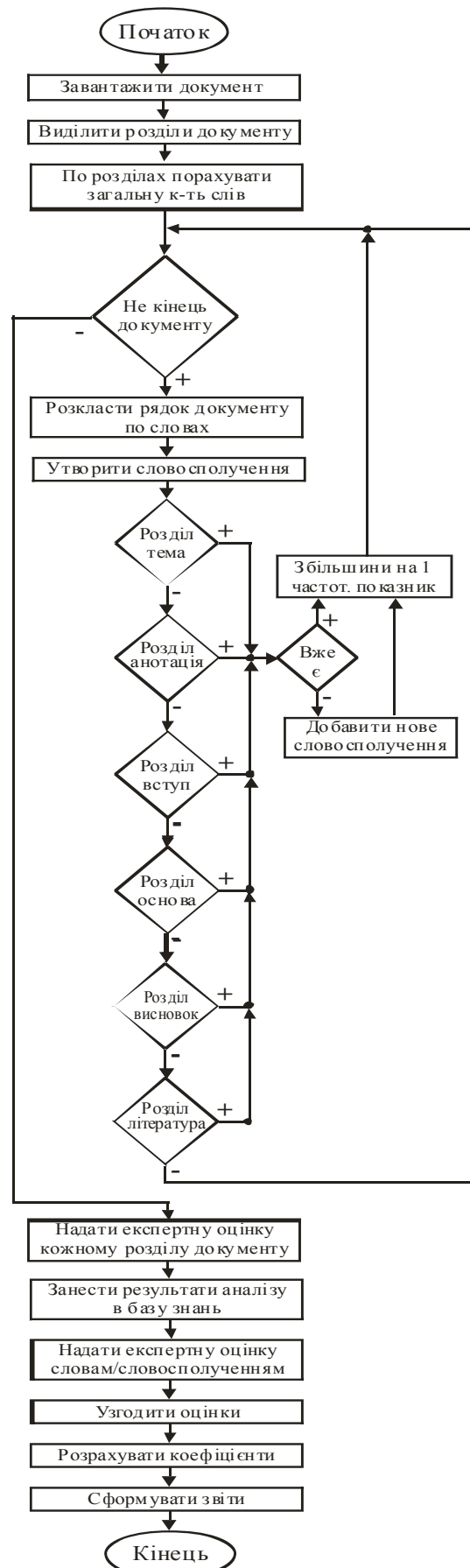


Рис. 2. Алгоритм аналізу документа

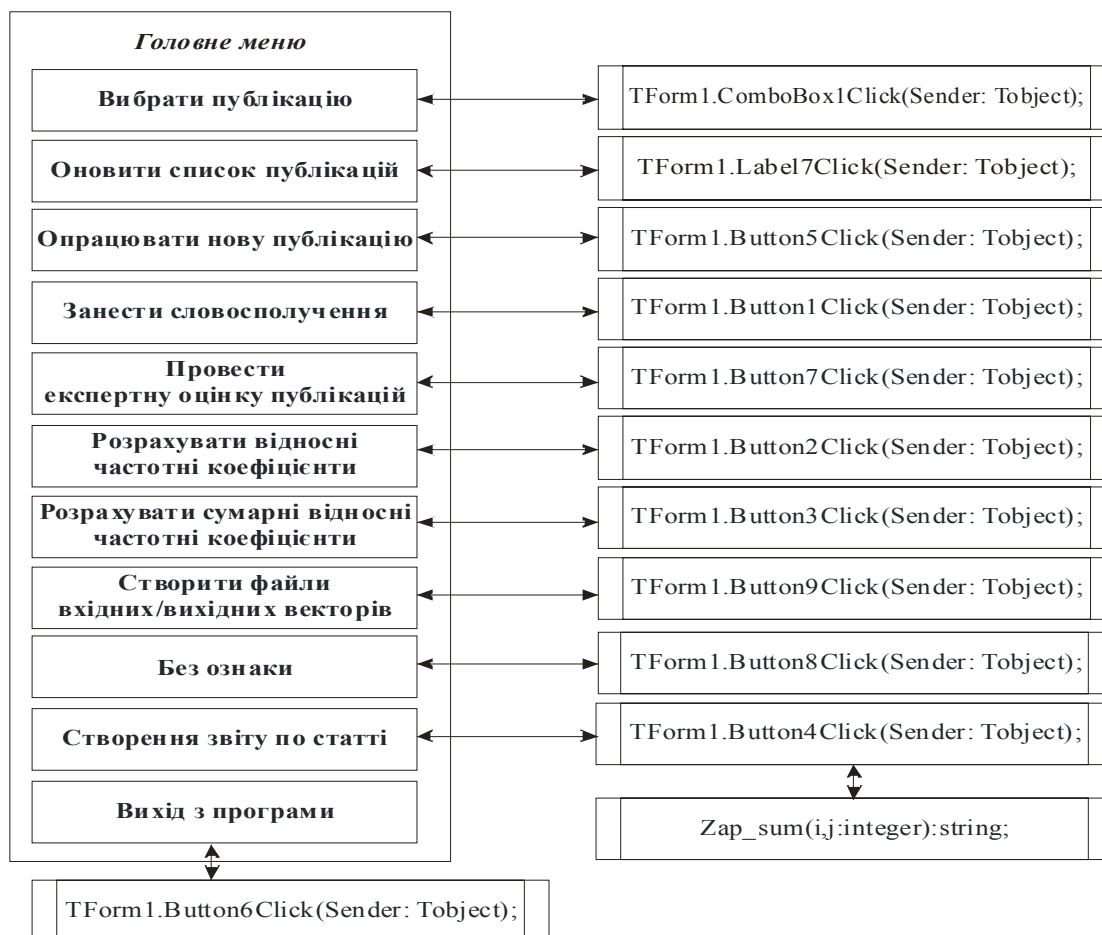


Рис. 3. Структура меню головної форми Form1 (Main)

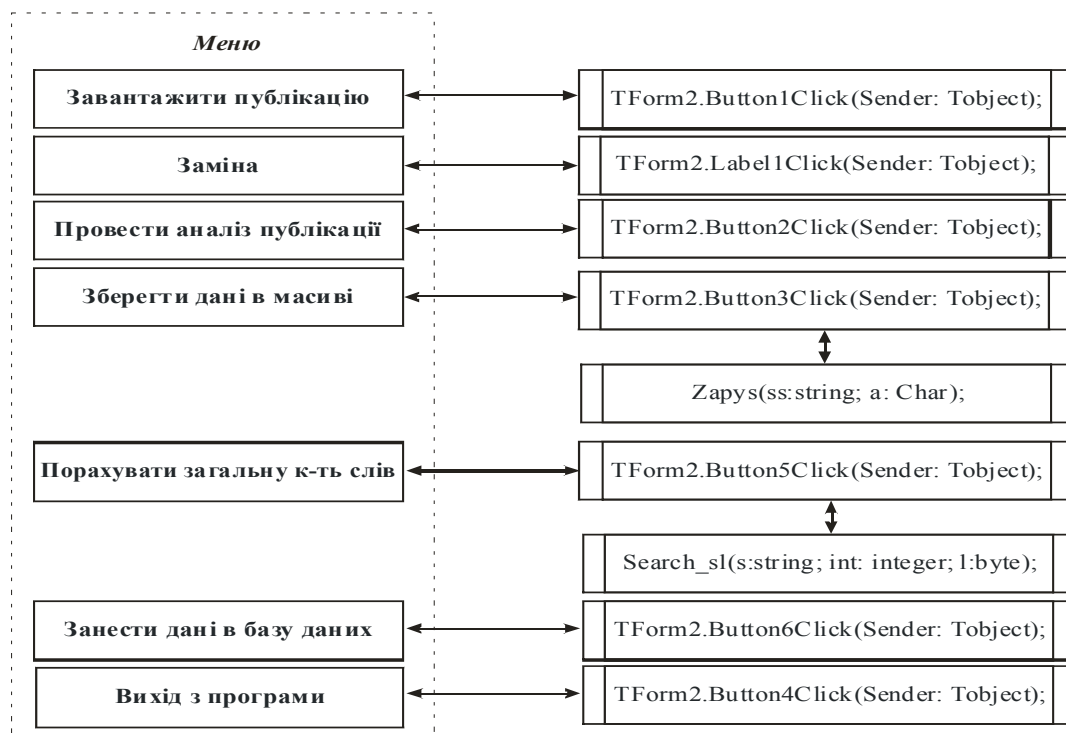


Рис. 4. Структура меню форми Form2 (Edit)

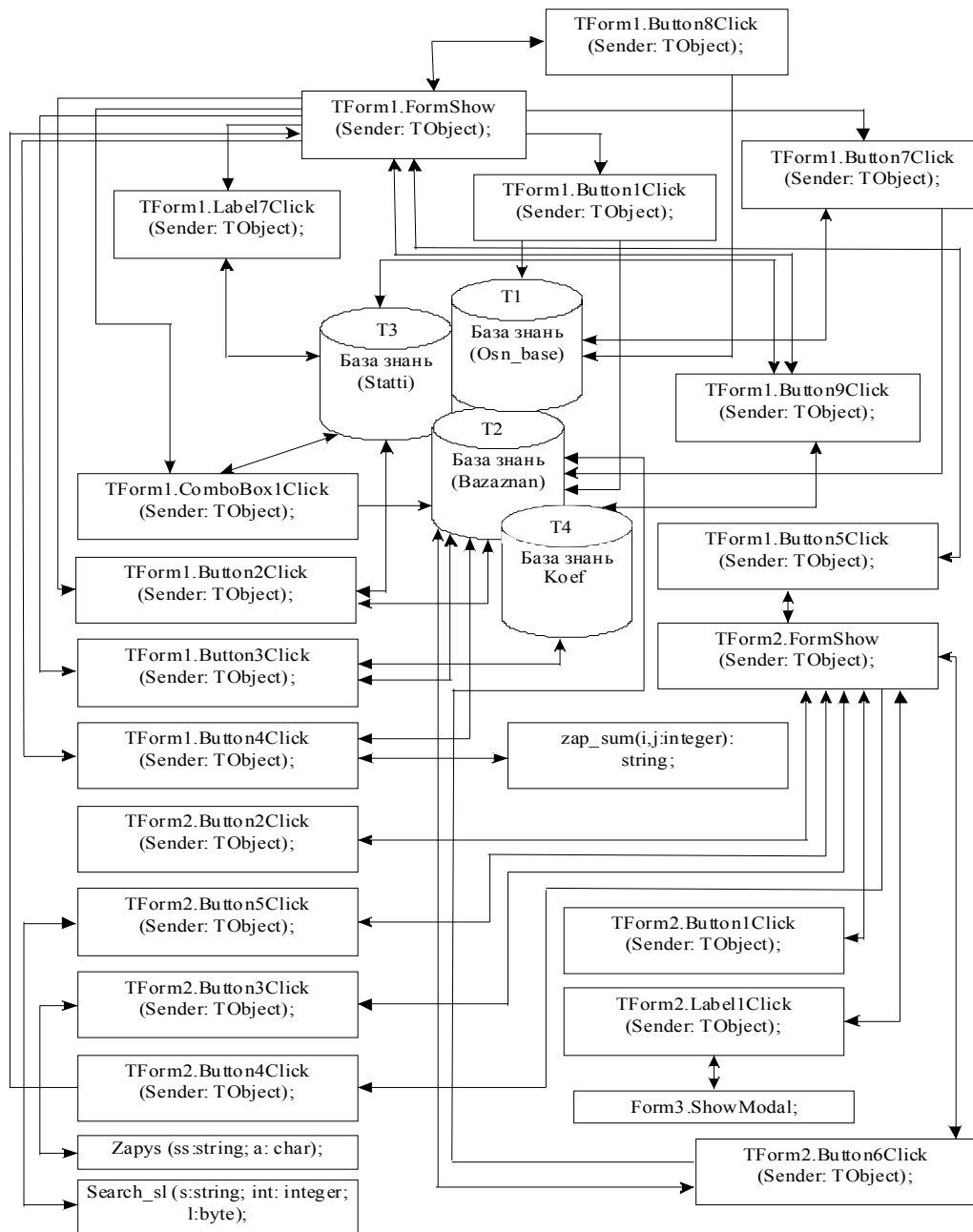


Рис. 5. Функційна схема програми

Програмно передбачено, що одному ключовому слову чи словосполученню експерт може призначити відповідність від 1-єї до 7-ми ознак одночасно.

Далі, згідно праці [5], розраховуються відносні частотні коефіцієнти, сумарні відносні частотні коефіцієнти кожного розділу документу окремо, а саме: тема, анотація, вступ, основа, висновок, література. У результаті, набір ключових слів представляється у вигляді числових значень (рис. 6), що можуть надходити на входи ШНМ, котра визначає корисність розділів джерела ДІ та загальну корисність тексту.

Розбиття тексту на слова та словосполучення для їх подальшого експертного аналізу реалізовано з використанням форми Form2 (рис. 7). Перед аналізом тексту експерт повинен розбити його на розділи за допомогою підготовлених міток. Після цього йому необхідно порахувати загальну кількість слів у документі.

Експерт, використовуючи програмне забезпечення, може створювати словосполучення до 4 слів. При необхідності деякі скорочення можна замінити їхніми повними назвами. Для цього використовується форма Form3.

Експертна система оцінки опорних точок текстів публікації

Нп Інформація про статтю

10 Локасько В.М., Поморова О.В., Тогова В.Ю. Система підтримки прийняття рішень для оперативно-чергових служб // (інет) Вальтер Троян. Высокая готовность в виде полного решения // LAN :: Приложения. 2003. - №07-08

12 Ляшкевич В.Я. Модель пошуку діагностичної інформації для тестування комп'ютерних пристроїв // Науковий Вісник

Стаття № 12

Ляшкевич В.Я. Модель пошуку діагностичної інформації для тестування комп'ютерних пристроїв // Науковий Вісник ЧНУ. Фізика. Електроніка. -2006. - Вип. 303. - С.94-100.

Виберіть наукову працю для перегляду

Оновити список публікацій

12 Ляшкевич В.Я. Модель пошуку діагност.

Вкажіть знайдене слово/словосполучення

Занести словосполучення

Опрацювати нову публікацію

Для всіх статей

Провести експертну оцінку публікацій

Код ознаки	Словосполучення	Кт	Тема	Від. оцінка	Анотація	Від. оцінка	Вступ	Від. оцінка	Основа	Від. оцінка	Висновок	Від. оцінка	Л-ра	Від. оцінка	Стат
020409101416	Шарі штучних нейронних мереж	4							1	0.00191					12
020409101416	Штучна нейронна мережа	3							13	0.01863					12
020409101416	Штучної нейронної мережі	3			1	0.03191									12
020409101416	Штучною нейронною мережею	3							1	0.00143					12
0516	Якісний	1							1	0.00048					12
020607	Якісний пертинентний пошук	3							1	0.00143					12

Розрахувати відносні частотні коефіцієнти

Розрахувати сумарні відносні частотні коефіцієнти

Створити файли вхідних/вихідних векторів

Вихід з програми

Тема	Озн. №1	Озн. №2	Озн. №3	Озн. №4	Озн. №5	Озн. №6	Озн. №7	Озн. №8	Озн. №9	Озн. №10	Озн. №11	Озн. №12	Озн. №13	Озн. №14	Озн. №15	Озн. №16
Тема	0	0.25	0.25	0.25	0.25	0	0.375	0	0.375	0.375	0	0.25	0.375	0.625	0.25	0.25
Анот.	0.11702	0.29787	0.20212	0.22338	0.06384	0.35106	0.31916	0.09574	0.23403	0.35106	0.04256	0.06382	0.21277	0.29787	0.0851	0.19149
Вступ	0.1616	0.2323	0.0303	0.0202	0.0909	0.202	0.1919	0.1111	0.101	0.0606	0.1111	0.0707	0.1919	0.1818	0.101	0.0606
Озн.	0.02825	0.23097	0.04306	0.14724	0.16693	0.19363	0.19407	0.09092	0.11852	0.16446	0.10661	0.08318	0.20175	0.21703	0.11955	0.2687
Висн.	0.04839	0.30646	0.04839	0.18549	0.16129	0.17742	0.26613	0.07258	0.19395	0.26613	0.03226	0.03226	0.17742	0.29033	0.12097	0.26613
Л-ра	0.01681	0.08404	0.04202	0.05883	0.06723	0.08404	0.08404	0.07563	0.05883	0.10925	0.0252	0.05042	0.0252	0.07563	0.04201	0.06723

Створення звіту по статті

Рис. 6. Результати опрацювання текстового файлу

C:\Kand\sts\TXT\12.txt

Заміна

Завантажити публікацію

Виб

Словосполучення

Словосполучення	8	90	87	1687	104	46
Модель пошуку	1					
Діагностичної інформації	1	2				
Тестування комп'ютерних пристроїв	1	2				
Архітектура інформаційно-пошукової с	1					
Інтелектуальні компоненти	1					
Системи	1					
Модуль штучної нейронної мережі	1					
База знань	1					
Нейромережні моделі пошуку	1					
Інтелектуалізувати процес пошуку	1					
Informative - searching system	1					
Computer devices	2					
Testing	2					
Intellectual components	1					
System	1					
Knowledge base	1					

Експертні оцінки розділів тексту

0.15

0.2

0.2

0.55

0.15

0.2

комп'ютерних систем, зокрема персональних комп'ютерів, та комп'ютерних мереж. На сьогодні без них важко уявити сучасне виробництво та управління. Однією з найважливіших властивостей комп'ютерних пристроїв є надійність їх функціонування. Забезпечення надійності комп'ютерних пристроїв на протязі життєвого циклу здійснюється різними заходами, зокрема, відновленням,

Скопіюйте і розставте помітки в тексті

Назва/тема /жжНжж/

Анотація /жжАжж/

Вступ /жжВжж/

Основна частина /жжОжж/

Висновок /жжВжж/

Література /жжЛжж/

Продовжити аналіз публікації

Зберегти дані в масиві

Порахувати загальну к-ть слів

Занести дані в базу даних

Вихід

Рис. 7. Процес аналізу публікації

Після опрацювання кожного розділу, експерт встановлює його оцінку корисності в межах від 0 до 1 для проведення процесу тестування КС. Оцінки корисності використовуються у якості вихідних значень при навчанні ШНМ. Програма автоматично підраховує кількість разів входження слова чи словосполучення відповідно у кожному з розділів тексту. Такий підхід надає можливість не відволікати експерта під час аналізу тек-

сту. Після опрацювання документа експерт повертається до головної форми програми, де може виставити експертним шляхом коди ознак слів чи словосполучень.

Результат роботи експерта можна представити у вигляді 4 типів звітів: звіт згенерований генератором Qreport, звіт у вигляді текстового файлу документу (рис. 8), файл вхідних векторів для ШНМ, файл експертних ознак для навчання ШНМ (рис. 9).

30
Хаханов В.И., Елисеев В.В. Применение IEEE стандартов для тестирования программно-технических комплексов // Радиоэлектронні і комп'ютерні системи. – 2006. – № 6 (18). – С.163–171.

кількість слів по розділах: Тема = 8; Анотація = 45; Вступ = 943; Основна частина = 1704; Висновок = 202; Література = 360.

Код ознак	Словосполучення	Тема	Анотація	Вступ	Основа	Висновок
Література						
04060711121314	ARM processor				1 0,00117	
04060711121314	ARM-процессора				2 0,00235	
111316	ASIC			1 0,00106		
04060708111214	Access Port				1 0,00117	
04060711121314	Ad hoc технологий			1 0,00318		1 0,01485
04060711121314	Ad-hoc моделей тестирования			1 0,00424		
04060711121314	Ad-hoc решения	1 0,06667				
04060711121314	Ad-hoc технологии	1 0,06667			1 0,00176	
04060711121314	Ad-hoc технологиями				1 0,00176	
04060711121314	Advanced Digital Networks				1 0,00176	1
0,00833						
04060711121314	Advanced Microcontroller Bus Architecture				2 0,00469	
04060711121314	Analyzer System				1 0,00117	
111316	BIST				4 0,00235	1
0,00278						

Рис. 8. Файл-звіт з результатами аналізу тексту

Р_вухид - Блокнот

0.15	0.15	0.2	0.7	0.2	0.15
0.1	0.1	0.2	0.5	0.2	0.1
0.1	0.15	0.2	0.55	0.2	0.1
0.15	0.15	0.2	0.65	0.15	0.05
0.1	0.1	0.15	0.45	0.1	0.2
0.1	0	0.2	0.6	0.2	0
0.05	0.15	0.25	0.55	0.1	0.025
0.15	0.15	0.2	0.6	0.2	0.1
0.15	0.1	0.15	0.55	0.15	0.1
0.1	0.1	0.15	0.4	0.1	0.15
0.05	0.05	0.2	0.6	0.2	0
0.1	0.1	0.15	0.5	0.1	0.1
0.2	0.2	0.2	0.7	0.1	0.1
0.15	0.2	0.2	0.6	0.2	0.1
0.1	0.2	0.2	0.6	0.15	0.25
0.1	0.1	0.2	0.6	0.2	0.15
0.1	0.15	0.2	0.65	0.05	0.25
0.1	0.15	0.2	0.65	0.1	0.15

Рис. 9. Текстовий файл з вихідними векторами ШНМ

Висновок

За допомогою розробленого програмного забезпечення було опрацьовано 80 публікацій. Серед них є публікації з фахових науково-технічних журналів і частина з мережі Інтернет. В результаті, опрацьовано 14,5 тис. різних слів і словосполучень. При опрацюванні останніх 10-ти публікацій в базу вносились порядку 50 нових слів і словосполучень із 250-350 визначених експертним шляхом при аналізі публікації.

Отже, враховуючи таку тенденцію, можна зробити висновок, що при застосуванні даної бази знань тематично схожі публікації й частково будь-які, що містять в собі діагностичну інформацію, відшукуються із високою відповідністю до пошукового запиту користувача.

Література

1. Локажук В.М., Савченко Ю.Г. Надійність, контроль, діагностика і модернізація ПК / За ред. В.М. Локажука. – К.: Академія, 2004. – 376 с.
2. Локажук В.М., Ляшкевич В.Я. Концептуальна модель пошуку діагностичної інформації для тестування комп'ютерних пристроїв // Інформаційні технології та комп'ютерна інженерія. – Вінниця: ВНТУ, 2005. – № 3. – С. 221-228.
3. Мельникова Е.А., Радионов А.Н. Построение классификационной иерархии документов на основе лексико-частотных характеристик с применением функций риска // Методы и средства обработки информации. – М.: МГУ, 2005. – 651 с.
4. Локажук В., Ляшкевич В. Текстологічні методи пошуку діагностичної інформації для тестування комп'ютерних пристроїв // Тези доповідей XIII МНТК Автоматика-2006, 25-28 вересня. – Вінниця: УНІВЕРСУМ, 2006. – С. 352.
5. Ляшкевич В.Я. Оцінка корисності та достатності діагностичної інформації для тестування комп'ютерних засобів у системі Simulink // Вісник Хмельницького національного університету. – Хмельницький: ХНУ, 2006. – № 6 (87). – С.77-85.

Надійшла до редакції 5.03.2007

Рецензент: д-р техн. наук, проф. О.Г. Шайко-Шайковський, Чернівецький національний університет ім. Ю. Федьковича, Чернівці.