

контролю динаміки характеристик якості електроенергії. Запропоновано створювати автономні апаратно-програмні засоби контролю для підприємств і організацій, основним видом енергоресурсів яких є використання електроенергії. Наведено структуру апаратно-програмного засобу для контролю динаміки характеристик якості електроенергії і наведено ряд рекомендацій по реалізації такого засобу.

1. *Стогній Б.С.* Еволюція інтелектуальних електричних мереж та їхні перспективи в Україні / Б.С. Стогній, О.В. Кириленко, А.В. Праховник. С.П. Денисюк // Технічна електродинаміка. – 2012. – № 5. – С. 52 – 67.
2. *Праховник А.В., Жаркін А.Ф., Новський В.О., Гриб О.Г., Калінчик В.П., Карасінський О.Л., Довгалюк О.М., Лазуренко О.П., Ходаківський А.М., Васильченко В.І., Светелік О.Д.* Автоматизовані системи обліку та якості електричної енергії в оптовому ринку/Під ред. Гриба О.Г. – Харків: ПП «Ранок-НТ», 2012. – 516 с.
3. *Кузнецов В.А.* Измерения в электронике: Справочник/ В.А.Кузнецов, В.А. Долгов, В.М. Коневских и др. – М.: Энергоатомиздат, 1987. – 512 с.
4. *Волков О.І.* Метрологія: теорія і нормативне забезпечення: Навч. Посіб. / О.І. Волков, О.М. Величко, Г.І. Хімичева та ін. – К.: Вища шк., 2008. – 335 с.:іл.
5. *Орнатський Д.П.* Структура автономного засобу контролю характеристик якості електроспоживання. / Д.П. Орнатський, Л.М. Щербак, В.А. Готович, О.В. Кармазін // Авіа-2013: XI міжнародна науково-технічна конференція, 21-23 травня 2013 р. – К.: НАУ, 2013.
6. Каталог електронних компонентів ELFA. – Sweden: Elfa AB, 2010. – 2067 с.

Поступила 25.9.2013р.

УДК 004.652.4+004.827

Н.Б.Шаховська, Р.Ю.Нога

Національний університет «Львівська політехніка», Львів, Україна

МЕТОД КЛАСТЕРИЗАЦІЇ НАУКОВИХ ПУБЛІКАЦІЙ ДЛЯ ФОРМУВАННЯ НАУКОВОЇ ШКОЛИ ТА ПРОГНОЗУВАННЯ ДИНАМІКИ ЇЇ РОЗВИТКУ

У статті описано метод визначення приналежності наукової публікації до певної наукової школи та прогнозування швидкості розвитку цієї наукової школи.

Вступ. Науковий напрям — це сфера наукових досліджень наукового колективу, спрямованих на вирішення певних значних фундаментальних проблем.

Наукова школа – науковий колектив, діяльність якого спрямована на вирішення проблем наукового напрямку.

У цьому дослідженні наукова школа визначатиметься множиною наукових публікацій Sch , які характеризуються множиною ключових слів Key , множиною авторів $Author$ та множиною основоположників школи $Main$:

$$Sch = \langle Key, Author, Main \rangle, Main \in Author.$$

Поставлено задачі:

1. визначити за публікаціями, які наукові школи функціонують;
2. класифікувати нові надходження публікацій за науковими школами;
3. прогнозування динаміки появи нових публікацій наукової школи.

Основний матеріал

Вхідною інформацією для віднесення публікації до наукової школи є файл з вмістом публікації. З файлу необхідно визначити базові характеристики публікації:

1. Автор(и) публікації (A).
2. Наукова установа (B).
3. Тема публікації (C).
4. Ключові слова (D).
6. Текст статті

Алгоритм формування бази даних характеристик публікації передбачає наступні кроки [6]:

Крок 1. Наукова стаття, подана як структурована текстова інформація, розбивається на речення та слова;

Крок 2. Відкидаються слова, що містять менше трьох символів;

Крок 3. Здійснюється класифікація слів, шляхом видалення з загального списку слів, які містяться в базі даних «Стоп-слова» та неінформативних слів і словосполучень.

Крок 4. Формується загальний список слів у документі, при цьому зберігається інформація про їх форматування та місце в тексті;

Крок 5. Загальний список слів модифікується в процесі стеммінгу, тобто відкидаючи закінчення слів, ми також видаляємо однакові слова з бази даних, але збільшуємо значення, що відповідає за кількість вживань цього слова в тексті, а ваги, що були попередньо присвоєні цим словам, додаються. Таким чином утворюється база даних «Ключові слова тексту»;

Крок 6. Автори статті та їх наукові установи шукаються на початку файлу за ознакою форматування.

Після формування бази даних публікації даної публікації відбувається розбиття публікацій за науковими школами методом k -найближчих сусідів:

1. Задаємо кількість сусідів k .

Оскільки ознаки кластеризації (автор, наукова установа, тема, ключові слова) невпорядковані, то використовуватимемо метрику d ізольованих точок:

$$l(X.x, Y.x) = \begin{cases} 1, X.x = Y.x \\ 0, X.x \neq Y.x \end{cases}$$

$$d(X, X_i) = \sum_i^p l(X.A_i, Y.A_i) + \sum_j^r l(X.D_j, Y.D_j) + \sum_t^w l(X.B_t, Y.B_t) + l(X.C, Y.C),$$

де p - кількість авторів обох статей,

r - сумарна кількість ключових слів,

w - сумарна кількість наукових установ,

$X.A_i$ - значення автора з номером i для наукової статті X і т.д.

2. Для кожного об'єкта знаходимо його k найближчих сусідів. Об'єкт X_i називається найближчим сусідом об'єкта X , якщо $d(X_i, X) = \max_i d(X_i, X), i = \overline{1, N}$, де N – кількість публікацій

3. Об'єкт X зараховується до того класу, до якого належить більшість з його K сусідів.

Якщо об'єкт не зарахований до жодного з кластерів, то шукаються слабкі зв'язки об'єкта з кластером.

Слабким назвемо зв'язок між об'єктами X_i та X , якщо значення відстані між ними менше, ніж третина від максимальної:

$$d_i(X, X_i) \leq \frac{\max d(X, X_i)}{3}.$$

Далі розроблено метод визначення спільних ознак у назві публікації.

Нехай маємо деякі назви С1, С2, С3. Для прикладу:

С1= «Пошук та збереження інформації за допомогою пошукової системи»

С2= «Перегляд та збереження файлів у файловій системі»

С3= «Пошук інформації у всесвітній мережі інтернет»

Умовно розіб'ємо назви на дві частини: праву та ліву. Розбиття здійснюватиметься шляхом симетричного поділу по довжині. Вважається, що ліва частина є більш інформативно важливою, ніж права [1, 2]. При цьому не треба брати до уваги слова-коннектори, такі як «і», «та» і т.д. Проте не слід відкидати слова, написані великими літерами, так як це може бути абрєвіатура. Також здійснюється відсікання закінчень (стеммінг). Зв'язок, який утворюється між публікаціями, для яких співпадає більше половини слів у лівій частині назви, назвемо *сильним зв'язком назв*.

Для наведеного вище прикладк розіб'ємо назви статей Р1, Р2 та Р3 на ліву та праву частини та виберемо спільне. Тоді отримаємо:

С1л=С3л= «пошук, інформації»

С1л=С2л= «збереження»

Отже, оскільки С1л та С3л мають два спільних слова, то між публікаціями Р1 та Р3 утворився сильний зв'язок назв. Відповідно в назвах С1л та С2л утворено слабкий зв'язок назв.

Такі зв'язки між назвами можна використовувати для додаткового навантаження зв'язків між публікаціями, що, у свою чергу, може вплинути на прийняття рішення, в яку з існуючих шкіл відносити публікацію, чи залишати її для створення нової школи.

Прогнозування зміни динаміки публікацій здійснюється за допомогою часових рядів, а саме методом ковзаючого середнього.

Тоді задачею прогнозування є знаходження залежності між кількістю публікацій по кожній із знайдених наукових шкіл, частотою появи нових ключових слів та частотою отримання наукових ступенів представників цієї школи. Динаміка зміни кількості ключових слів може бути охарактеризована стосовно якогось базисного (зазвичай першого) спостереження і величиною зміни сусідніх рівнів. У якості статистичних характеристик часового ряду Y_i , $i = 1..n$ використовуються наступні величини: середнє арифметичне

$Y = \frac{1}{N} \sum_{j=1}^N Y_j$, середній абсолютний приріст $\bar{Y} = (Y_n - Y_1)/(N - 1)$, де N – число

рівнів ряду, Y_i – рівні ряду.

Відповідно до методу перевірки істинності різниці середніх початковий часовий ряд розбивається на дві однакові (або майже однакові) частини, після чого перевіряється гіпотеза про істотність різниці середніх для цих частин. Недолік методу полягає в неможливості правильно визначити наявність тренда у тому випадку, коли часовий ряд містить точку зміни тенденції в районі середини ряду.

У методі Форстера-Стюарта гіпотеза про відсутність тренда перевіряється за допомогою допоміжних функцій:

$$L = \sum_{t=2}^N I_t, \quad I_t = U_t - V_t, \quad u_t = \begin{cases} 1, & Y_t < Y_{t-1}, \dots, Y_1 \\ 0 & \end{cases}, \quad v_t = \begin{cases} 1, & Y_t > Y_{t-1}, \dots, Y_1 \\ 0 & \end{cases}.$$

Гіпотеза про відсутність тенденції не приймається, якщо розрахункове t -значення більше табличного на вибраному рівні значеннєвості 0.95.

Перевірка однорідності даних здійснимо на основі критерію Ірвіна, що заснований на порівнянні сусідніх значень ряду. Відповідно до нього

розраховується характеристика t : $t = \frac{Y_t - Y_{t-1}}{Y}$.

Отримані значення порівнюються з табличними значеннями. Оцінка властивостей зводиться до дослідження автокореляційної функції вихідного і різницевих рядів. Аналіз автокореляції виконується за допомогою графіка і критичних значень коефіцієнтів, встановлених експертно.

Параметри цього рівняння знаходять за методом найменших квадратів. Ковзну середню в обраному інтервалі визначають як зважене середнє усіх попередніх рівнів, причому ваги спостережень мають неоднакові значення.

Для апробації ми проаналізували 208 наукових публікацій за 2008 – 2011 рр. і отримали результат кластеризації публікацій:

Реляційні бази даних	Кількість
Національний університет «Львівська політехніка»	42
Харківський національний університет радіоелектроніки	
Тернопільський національний університет	27
Хмарні обчислення	
Національний технічний університет КПІ	45
Національний аерокосмічний університет	32
Національний університет «Львівська політехніка»	28

Прогнозування здійснювалось на 2012 р., отримані прогнознi кількості статей порівнювалися з реальною кількістю статей за цей рік. Відхилення прогнозу – 27%.

Висновки

Дослідження присвячене вивченню можливостей виділення наукових шкіл за аналізом публікацій та прогнозуванню їх розвитку. Результати доцільно використовувати для прогнозування актуальності наукових напрямів.

1. *Mani I., Bloedorn E.* Summarizing Similarities and Differences Among Related Documents // Information Retrieval. – 1999. – 1, № 1. – P. 35–67.
2. *Radev D. R., McKeown K. R.* Generating Natural Language Summaries from Multiple Online Sources // Computational Linguistics. – 1998. – 24, № 3. – P. 469–500.
3. *Ноза Р.* Метод формування наукових шкіл на основі аналізу елементів публікацій // Математичні машини і системи, 2013, No 4. – Київ. – С. 107 – 113

Поступила 30.9.2013р.

УДК 517.9+534.1

П.Я. Пукач, А.В. Чабан, м. Львів

МАЛІ ПОПЕРЕ ЧНІ КОЛИВАННЯ МЕМБРАНИ З УРАХУВАННЯМ НЕЛІНІЙНИХ ДИСИПАТИВНИХ СИЛ

Abstract. Interdisciplinary approach to the modeling of elastic vibrations in isotropic nonlinear medium as an example of the membrane is used. Mechanical equations of an object without any decomposition processes based solely on the energy approaches and theories of mechanical fields in continuum media are formed. The adequacy of received model of the real prototype is confirmed.

Keywords: Ostrogradskii-Hamilton principle, extended Lagrangian, Galerkin method.

© П.Я. Пукач, А.В. Чабан

117