

Технології інформаційного менеджменту та інтелектуального опрацювання даних

Відображення властивостей текстового документа в багатовимірних моделях інформаційних систем комп'ютерного моніторингу

Сергій Голуб

Кафедра інтелектуальних систем прийняття рішень, Черкаський національний університет імені Богдана Хмельницького, УКРАЇНА, м. Черкаси, бульвар Шевченка, 81; e-mail: fpkpk@cdu.edu.ua

The technology of converting printed text into an array of numerical characteristics that can be used to automate the process of consolidation of heterogeneous information. Its efficiency tested by synthesis models of authorship printed texts and tests.

Ключові слова – друкований текст, багатовимірна модель, моніторинг.

I. Вступ

Подальша автоматизація процесів обробки даних в технологіях консолідації інформації пов'язана з усуненням проблем, пов'язаних із поєднанням описів досліджуваного об'єкта, що міститься в різних форматах – в масивах чисельних даних, текстових документах, звукових файлах та інших. Було запропоновано поєднувати різномірні моделі, що містять описи цього об'єкта, в ієрархічній структурі глобальної функції інформаційних систем багаторівневого комп'ютерного моніторингу [1]. В цій роботі подано метод розв'язку задачі перетворення текстової інформації до вигляду масиву чисельних характеристик із наступним синтезом багатовимірної моделі об'єкта моніторингу.

II. Результати досліджень

До масиву вхідних даних (МВД) інформаційної системи багаторівневого моніторингу висувається ряд вимог, задоволення яких дозволяє синтезувати якісну багатовимірну модель. МВД має вигляд матриці чисельних характеристик об'єкта, інформативність якої повинна бути достатньою для синтезу якісних моделей.

Для характеристики якості моделі крім точності, адекватності та стійкості в цих системах використовується показник стабільності динаміки вагових

коефіцієнтів (СДВК). Для відображення текстового документа в структурі МВД запропоновано використовувати частотні характеристики елементів тексту. Відповідно до поставленої задачі проводиться декомпозиція тексту, глибина якої визначається індивідуальними його особливостями і є об'єктом додаткових досліджень. Результати декомпозиції подаються у вигляді двовимірної таблиці, в якості стовпців якої слугують частотні характеристики окремого елемента, розраховані за результатом аналізу сегмента тексту. Строчки зводяться спостереженнями та містять частотну характеристику множини елементів, розрахованих для окремого сегмента.

Послідовність спостережень утворюють МВД, на основі якого розв'язуються комплексні задачі кластеризації спостережень агрегатами мікстрати, ідентифікації функціональних залежностей агрегатами макстрати та розпізнавання образів багатовимірними моделями метастрати [2]. На рис. 1 подані результати застосування інформаційної системи багаторівневого комп'ютерного моніторингу з метою розпізнавання авторства друкованого тексту.

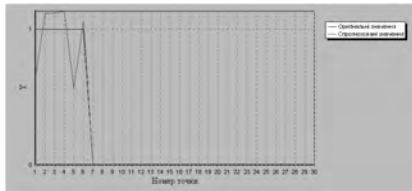


Рис. 1 Приклад рисунку

Результати випробувань доводять, що в структурі багатовимірної моделі відображені індивідуальні властивості автора, які дозволяють відрізнити його тексти від текстів інших авторів.

Висновок

Запропонована нова інформаційна технологія відображення властивостей автора друкованого тексту в структурі багатовимірної моделі інформаційної системи багаторівневого комп'ютерного моніторингу. Запропонована технологія може бути використана для автоматизації процесу консолідації різномірної інформації.

Література

1. Голуб С.В. Багаторівневе моделювання в технологіях моніторингу оточуючого середовища / С.В. Голуб. – Черкаси: Вид. від. ЧНУ імені Богдана Хмельницького, 2007. – 220 с.
2. Голуб С.В. Відображення консолідованої інформації економічних показників регіону у структурі багаторівневих моделей / С.В. Голуб, Н.О. Химиця // Вісник Східноукраїнського національного університету імені Володимира Даля. – 2012. – № 8 (179). – Частина 1. – С. 122-128.