

Профілювання текстів у технології інформаційного моніторингу

Сергій Голуб¹, Марія Голуб²

1. Кафедра інтелектуальних систем прийняття рішень, Черкаський національний університет імені Богдана Хмельницького, УКРАЇНА, м. Черкаси, бул.

Шевченка, 81

E-mail: fpkpk@ukr.net

2. Кафедра журналістики, реклами і PR-технологій, Черкаський національний університет імені Богдана Хмельницького, УКРАЇНА, м. Черкаси, бул.

Шевченка, 81

E-mail: Mas-golub@yandex.ru

Described and the mathematically set forth task of profiling of texts in technology of the multilevel informative monitoring

Ключові слова – інформаційний моіторинг, профілювання текстів, постановка задачі.

Вступ

Технологія інформаційного моніторингу реалізує забезпечення інформацією процесів прийняття рішень шляхом обробки та перетворення результатів умовно неперервних спостережень за характеристиками профілю об'єкта досліджень [1]. Технологія багаторівневого моніторингу передбачає реалізацію глобальної задачі опису властивостей об'єктів спостереження шляхом розв'язання ієрархії локальних задач моніторингу окремих ознак їх профілю [2].

Технології інформаційного моніторингу, на відміну від інших видів моніторингу, забезпечують виявлення та обробку ознак профілю об'єкта, поданого не тільки у вигляді масивів чисельних характеристик, а і у вигляді текстових повідомлень, відео- та аудіо- файлів. Тому технологія багаторівневого інформаційного моніторингу забезпечує консолідацію результатів обробки різнорідних даних, що описують ознаки профілю об'єкта дослідження [3]. Таким чином передбачаються розробка та реалізація методів перетворення різнорідних даних до вигляду масиву чисельних характеристик. Перетворення текстових повідомлень до вигляду масиву чисельних характеристик забезпечується застосуванням частотних ознак [4]. Розв'язання послідовності задач розпізнавання ознак профілю автора текстового повідомлення реалізує профес профілювання автора тексту [5].

Постановка задачі профілювання текстів

З метою формування ієрархії локальних задач виявлення окремих ознак профілю об'єкта дослідження необхідно забезпечити пошук

текстових повідомлень, що несуть необхідну інформацію. Оскільки об'єкт досліджень відомий, ґрунтуючись на результатах розв'язання задачі профілювання автора окремого тексту [5] сформулюється задача пошуку інформативних текстових повідомлень у вигляді задачі профілювання текстів.

Профілювання текстів – це процес пошуку текстових повідомлень, що мають ознаки заданого профілю об'єкта досліджень.

Нехай задана множина ознак профілів тестових повідомлень

$$X = \{X_1, X_2, \dots, X_l\}, X_i = \{x_{i1}, x_{i2}, \dots, x_{ik}\}, i = 1, l \quad (1)$$

та множина текстів що характеризуються індивідуальними профілями

$$A = \{A_1, A_2, \dots, A_n\}, A_i = \{a_{i1}, a_{i2}, \dots, a_{im}\}, i = 1, n \quad (2)$$

де x_{ik} – сукупність характеристик, що дозволяють описати перелік ознак X_i профілю A_i ; a_{im} – текстове повідомлення, що характеризується переліком $x_{ik} \in X_i$, має набір ознак $X_i \in X$, що відповідають профілю A_i . При цьому існує відповідність між елементами множин A і підмножинами множини X :

$$y: X \rightarrow A. \quad (3)$$

Необхідно створити множину багатопараметричних моделей – правил

$$y = \{y_1, y_2, \dots, y_z\}, y_i = f(X), i = 1, z, \quad (4)$$

які б забезпечували групування текстових повідомлень $a_{im} \in A_i$.

Висновок

Розв'язання поставленої задачі дозволяє реалізувати пошук текстової інформації заданої тематики для підтримки прийняття рішень

Література

1. Рыжов А.П. Информационный мониторинг сложных процессов: технологические и математические основы / А.П. Рыжов // Интеллектуальные системы, Том 11, вып. 1-4, 2008, с. 101-136.
2. Голуб С.В. Багаторівневе моделювання в технологіях моніторингу оточуючого середовища / С.В. Голуб. Черкаси: Вид. від. ЧНУ імені Богдана Хмельницького, 2007. – 220 с.
3. С.В. Голуб Консолідація моделей в процесі багаторівневого опрацювання даних / С.В. Голуб // Інформація, комунікація, суспільство 2014: матеріали 3-ї Міжнар. наук. конференції ICS-2014. – Львів: Видавництво Львівської політехніки, 2014. – С. 162-163.
4. Голуб М. Вибір ознак в процесі інтелектуальної обробки текстових повідомлень / М. Голуб // Інформація, комунікація, суспільство 2014: матеріали 3-ї Міжнар. наук. конференції ICS-2014. – Львів: Видавництво Львівської політехніки, 2014. – С. 148-149.
5. Литвинова Т.А. Профилірование автора письменного текста / Т.А. Литвинова // Язык и культура. - 2013. - № 3 (23). - С. 64-72.