

УДК 004.67

Ланде Д.В.¹, д.т.н.

Страшной Л.²

Балагура І.В.¹, к.т.н.

¹ *Інститут проблем реєстрації інформації НАН України*

² *Університет каліфорнії в Лос-Анджелесі*

НАПРАВЛЕНА КОРЕЛЯЦІЙНА МЕРЕЖА ПОНЬТЬ НА ОСНОВІ ДИНАМІКИ ПУБЛІКАЦІЙ

Запропоновано методику формування, кластеризації та візуалізації спрямованих кореляційних мереж. Зв'язки між вузлами таких мереж відповідають модифікованим певним чином значенням крос-кореляцій між векторами - наборами пара-метрів, що відповідають цим вузлам. Для побудови мережових структур для кожного вузла (тематики) формуються вектори - масиви чисел, відповідні деяким часовим рядам. В якості при-кладу розглядаються часові ряди, що формуються сервісом Google Books Ngram Viewer. Кожному року ставиться у відповідність число - кількість появ поняття в публікаціях, які охоплюються системою Google Books. Розмірність цього вектора відповідає кількості років, довжині інтервалу часу, протягом якого аналізувався масив публікацій.

Мета роботи – представити методику формування, кластеризації, ранжирування вузлів і візуалізації так званих спрямованих кореляційних мереж, графових структур, зв'язку між вузлами (поняттями, сутностями) яких відповідають значенням кореляцій між наборами параметрів, які відповідають цим поняттям.

Наведений підхід, на відміну від існуючих має такі переваги, як інтуїтивні, близькі до реальності правила визначення ваги вузлів і зв'язків; надійна математична основа кореляційного аналізу; облік невикористовуваних раніше параметрів, часових рядів динаміки публікацій, які відповідають сутностям і дозволяють групувати сутності по тенденціям їх розвитку в часі; об'єктивність і відносна простота реалізації. Наведена методика може базуватися на даних, отриманих, наприклад, від систем контент-моніторингу, використовуватися в аналітичних системах різного призначення з метою узагальнення безлічі сутностей без явно виражених зв'язків між ними.