

# ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИДОБУВАННЯ ЗНАНЬ З ДОКУМЕНТАЦІЙНИХ ПОТОКІВ КОРПОРАТИВНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

Тарловський В.О.

*Херсонський національний технічний університет, м. Херсон*

Більшість корпоративних інформаційних систем, працюють з неструктурованою інформацією у формі документів та з даними. За останні роки в організаціях накопичилися величезні обсяги неструктурованого контенту, що включає документи, повідомлення електронної пошти, відеоролики, миттєві повідомлення, веб-сторінки і т. д. Накопичені масиви електронної інформації є найціннішим інтелектуальним активом організації, її корпоративної пам'яттю. При цьому, інформацію необхідно не тільки зберегти, але і зробити легко доступною в процесі прийняття управлінських рішень.

У роботі розглянута інформаційна технологія, що базується на інтелектуальному аналізі даних, яка дозволяє видобувати знання з потоків неструктурованих текстових документів корпоративних інформаційних систем. Враховуючи, що електронні документи, які надходять до організації (електронна пошта, HTML-документи, XML-документи і ін.) мають різноформатне представлення, запропонована технологія перекладу електронної інформації до інтелектуального активу компанії, яка є основою бази знань компанії, включає сім підсистем: передлінгвістична обробка, графемна обробка, морфологічна обробка, контекстний аналіз, статистична обробка, логіко-алгебрична обробка навчальної виборки, динамічна класифікація.

На етапі формування класів еквівалентності термінологічних понять, які представляють структуровані динамічно наповнювані інтелектуальні активи компанії, використовуються алгебро-логічні моделі, побудовані на основі застосування метода компараторної ідентифікації інформаційних об'єктів.

У розділі наведено формальну модель оцінювання ефективності підсистеми управління знаннями інтелектуальної корпоративної системи. Проведена оцінка ефективності запропонованих моделей та методів обробки текстової інформації. Для кількісної оцінки ефективності створеної моделі формування бази знань інтелектуальної корпоративної системи, використано показники ефективності, які затверджено міждержавним стандартом з інформації, бібліотечної й видавничої справи: коефіцієнт точності (*precision*), повноти (*recall*), акуратності (*accuracy*), а також міра Ван Ризбергена, яка визначає середнє гармонійне точності та повноти.