



УДК 65.012+004.9

WEB-MINING: ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ В МЕРЕЖІ INTERNET

Студ. В.П.Мінакова

Наук. керівник доц. Н.В. Геселева

Київський національний університет технологій та дизайну

Web-mining – це використання методів інтелектуального аналізу даних для автоматичного знаходження веб-документів та послуг, вилучення інформації із веб-ресурсів та виявлення загальних закономірностей в Інтернеті.

В WebMining можна виділити наступні етапи:

- Вхідний етап (англ. inputstage) – отримання необроблених даних з джерел (логи файлів, тексти електронних документів)
- Етап попередньої обробки даних (англ. preprocessingstage) – дані надаються у формі, необхідній для успішної побудови той чи іншої моделі
- Етап моделювання (англ. patterndiscoverystage)
- Етап аналізу моделі (англ. patternanalysisstage) – інтерпретація отриманих результатів

Це загальні кроки, які необхідно пройти для аналізу даних мережі Інтернет. Конкретні процедури кожного етапу залежать від поставленого завдання. У зв'язку з цим виділяють різні категорії WebMining.

Категорії WebMining:

- Аналіз використання веб-ресурсів (WebUsageMining)– процес вилучення корисної інформації з користувальницьких журналів доступу, журналів проксі-сервера, браузерних журналів, призначених для користувача сесійних даних..
- Вилучення веб-структур (WebStructureMining) - процес виявлення структурної інформації в Інтернеті.
- Вилучення веб-контенту (WebContentMining) - процес вилучення знань з контенту документів або їх опису, доступних в Інтернеті.

WebUsageMining має ряд переваг, що робить цю технологію привабливою для корпорацій. Ця технологія дозволила електронній торгівлі створити персоналізований маркетинг, який в кінцевому підсумку призвів до збільшення обсягів торгівлі. Компанії можуть краще зрозуміти потреби клієнта і швидше реагувати на потреби клієнтів; знайти, залучити й утримати клієнтів, заощадити на собівартості продукції за рахунок використання придбаного розуміння вимог замовника. Також компанії підвищують рентабельність за рахунок цільового ціноутворення на основі створених профілів.

Проте, найкритикованіше етичне питання, пов'язане з WebUsageMining, є питання про вторгнення в приватне життя. Захист вважається втраченим, коли отримана інформація про окремого користувача використовують або розповсюджують без їх відома і згоди. Отримані дані будуть проаналізовані і кластеризовані в формі профілів або будуть анонімними до кластеризації без створення особистих профілів. Таким чином, ці додатки де-індивідуалізують користувача, судячи про них тільки по їх clickstream-даних (англ. Clickstream - маршрут переміщення користувача по веб-сайту). Іншою важливою проблемою є те, що компанії зі збору даних можуть їх використовувати для абсолютно різних цілей, що істотно порушує інтереси користувачів. Зростаюча тенденція використання персональних даних в якості товару закликає власників веб-сайтів до торгівлі цими даними, розташованими на їх сайтах.