

УДК 343.1:65.012.8+004

Іван Вікторович Панасюк,

*аспірант Харківського національного
університету внутрішніх справ*

Робота з великими текстовими масивами у правоохоронних органах

Під час проведення розслідування нерідко доводиться справу з великими текстовими масивами, які представлені в різних форматах. Вказані дані окрім захисту [1, 2] також потребують належної обробки. Для їх приведення до зручного формату потрібно користуватися спеціальними програмними інструментами.

Якщо великі дані зберігаються у текстовому вигляді, то переглянути їх за допомогою неспеціалізованих програмних засобів є достатньо складним завданням. Алгоритм роботи стандартних засобів перегляду передбачає першочергове завантаження всього об'єму файлу до оперативної пам'яті. Якщо такий файл має розмір декілька гігабайт, то його відкриття триватиме довго. З метою перегляду змісту таких документів слід користуватися спеціалізованими програмами. Однією з таких програм є редактор EmEditor. За його допомогою досить зручно переглядати великі текстові документи, здійснювати в них пошук, розділяти їх на частини, вносити інші зміни. У разі потреби перетворення текстових файлів у формат бази даних, може знадобитися їх попередня обробка для приведення до певної форми. В цьому випадку спеціалізовані редактори можуть бути використані для швидкого перегляду файлу та вилучення з нього фрагменту даних для відпрацювання процесу перетворення.

Для імпорту текстових даних до якоїсь СУБД вони нерідко мають бути перетворені у певну форму, вимоги до якої визначаються алгоритмом роботи СУБД. З метою швидкого внесення відповідних змін можуть бути застосовані спеціалізовані інструменти, як от TextPipe.

Порядок роботи з вказаною програмою є достатньо простим. У лівому полі обирається відповідний фільтр, який налаштовується, а потім розміщується в тому порядку, в якому його слід застосувати до відповідного файлу. Для ефективного створення фільтрів потрібно знати головні шаблони для перетворень. З відповідними прикладами

можна ознайомитися, наприклад, за адресою datamystic.com/textpipe/manual/general_usage_easypatterns_reference.htm. По суті створення фільтру нагадує процес написання простої програми.

Існують випадки, коли імпортувати великі дані до СУБД є недоцільним, через суттєве зростання об'єму вихідних файлів банку. В такому випадку пошук можна здійснювати стандартними засобами або з використанням спеціалізованих утиліт.

У системі Windows, наприклад, для цього можна скористатися утилітою `findstr` з вказівкою потрібних параметрів. Наприклад,

`findstr /s «що шукаємо» де_шукаємо`

Крім того, з цією метою можна використовувати утиліти Grep, Folder Find Text, DocFetcher.

За результатами обробки великих текстових масивів, як правило, складається протокол, або експертний висновок, у якому в стислій формі наведено узагальнену інформацію, зрозумілу для суду та пересічних громадян. Але... Не є секретом наявність великих черг на проведення комп'ютерно-технічних експертиз. Інколи така черга може тривати більше ніж пів року. І це тільки до початку проведення експертизи. Якщо ж мова йде про огляд документів, то у багатьох підрозділах оперативні працівники або слідчі в ручному режимі вивчають документи, що, по-перше, сприяє виникненню помилок через потребу концентрації уваги, по-друге, великий час на проведення аналізу невиправдано збільшує тривалість розслідування.

Список бібліографічних посилань

1. Носов В. В. Манжай О. В. Організація та забезпечення інформаційної безпеки : навч. посіб. Харків : Вид-во Харків. нац. ун-ту внутр. справ, 2007. 216 с.
2. Манжай О. В. Правові засади захисту інформації : навч.-посіб. Харків : Ніка Нова, 2014. 104 с.

Одержано 25.10.2019