



## ПРОГРАМНО-ТЕХНІЧНІ КОМПЛЕКСИ

А.В. МОВЧАН, В.Ю. ТАРАНУХА

УДК 004.94:343.985

### ПОБУДОВА СИСТЕМИ АВТОМАТИЗАЦІЇ ДЛЯ ВПРОВАДЖЕННЯ МОДЕЛІ ПОЛІЦЕЙСЬКОЇ ДІЯЛЬНОСТІ, КЕРОВАНОЇ АНАЛІТИКОЮ, В ОРГАНАХ НАЦІОНАЛЬНОЇ ПОЛІЦІЇ УКРАЇНИ

**Анотація.** Проаналізовано актуальні проблеми впровадження моделі поліцейської діяльності, керованої аналітикою, в органах Національної поліції України. На основі результатів цього аналізу запропоновано архітектуру аналітичної системи та розглянуто математичний апарат побудови складових майбутньої системи. Успішне впровадження та застосування нових методів кримінального аналізу дасть змогу в майбутньому поширити його на всю систему Національної поліції України, а також активно застосовувати аналітичні методи та технології, які забезпечать ефективне розслідування кримінальних правопорушень.

**Ключові слова:** інформаційно-аналітична робота; кримінальний аналіз; модель поліцейської діяльності, керованої аналітикою; засоби автоматизації; комп'ютерна лінгвістика; умовні випадкові поля; глибоке навчання.

Обов'язковою запорукою своєчасної та ефективної протидії проявам злочинності є проведення копійної інформаційно-аналітичної роботи. Застосування новітніх інформаційних і телекомунікаційних технологій дає змогу інтегрувати й опрацювати величезний обсяг даних, що містяться у відкритих джерелах інформації та спеціалізованих автоматизованих інформаційних системах, і до того ж отримати нові знання кримінологічного та оперативно-розшукового характеру.

Метою статті є розроблення основних підходів, алгоритмів, моделей і методичних основ для опису та аналізу проблем поліцейської діяльності, керованої аналітикою (intelligence-led policing) [1], а також створення в майбутньому засобів автоматизації, що забезпечать підвищення результативності поліцейської роботи. Для цього проведено аналіз як сучасного закордонного досвіду, так і стану справ у системі Національної поліції України. На основі результатів цього аналізу запропоновано архітектуру аналітичної системи та розглянуто математичний апарат побудови складових майбутньої системи.

Наразі відбувається процес реформування поліцейської діяльності. Одним з його напрямків є впровадження в органах Національної поліції України моделі поліцейської діяльності, керованої аналітикою, що здійснюється за підтримки Консультативної місії Європейського Союзу в Україні. Поліцейська діяльність, керована аналітикою, є моделлю поліцейської діяльності, згідно з якою оперативно-аналітична інформація (intelligence) слугує підставою для проведення операцій/розслідувань, а не розслідування є єдиною підставою для збору оператив-

© А.В. Мовчан, В.Ю. Тарануха, 2018

но-аналітичної інформації [2]. У країнах Європейського Союзу, США та інших розвинених країнах світу використання можливостей кримінального аналізу є обов'язковим для всіх правоохоронних органів. Його зміст, правила та процедури чітко визначено й урегульовано в нормативно-правових актах.

У процесі кримінального аналізу здійснюється цілеспрямований пошук, виявлення, фіксація, вилучення, упорядкування, аналіз та оцінка кримінальної інформації, її представлення (візуалізація), передача та реалізація. Зокрема, в аналітичній роботі застосовують оперативний аналіз (даних телефонних дзвінків, злочинних угруповань, справ, порівняльний аналіз), тактичний аналіз (кримінальний, геопросторовий, часовий, аналіз кримінальних тенденцій, місць концентрації злочинності, кримінальних моделей, профілів підозрюваних/жертв, МО-аналіз), стратегічний аналіз (моделей/форм злочинності та профілювання, тенденцій, аналіз з використанням географічного профілювання, SWOT- та PEST-аналіз).

У системі Національної поліції існує багато джерел розрізненої інформації, що зберігається в різних базах даних. Цю інформацію автономно аналізують співробітники різних департаментів. Крім того, кожний оперативний працівник накопичує і зберігає власну оперативну інформацію, яка після його звільнення або переміщення по службі стає практично недоступною для інших оперативних працівників. При цьому порушується європейський принцип, згідно з яким інформація належить не конкретному поліцейському, а державі, як один з продуктів діяльності поліції. У зв'язку з цим постає завдання з об'єднання всієї оперативної інформації та її подальшого аналізу, що має сприяти розкриттю насамперед тяжких та особливо тяжких злочинів [3].

До того ж застосування моделі поліцейської діяльності, керованої аналітикою, окрім роздрібності даних для оперативного аналізу, тягне за собою низку проблем на інших рівнях аналізу. Зокрема, на етапі тактичного аналізу наявні інструменти забезпечують хороші можливості для візуалізації, але мають відчутний недолік — слабкість математичного апарату. Для проведення стратегічного аналізу часто бракує даних через те, що їх ніхто не узагальнив, або сам аналіз проводиться лише заради підвищення звітної результативності правоохоронного органу, при тому, що, наприклад, PEST- і SWOT-аналіз можна застосовувати для передбачення очікуваних змін у діяльності злочинних угруповань.

На жаль, найбільш поширеним аналітичним інструментом, що використовується у повсякденній роботі органів Національної поліції, є програмний продукт Microsoft Office, хоча в деяких департаментах застосовується аналітичне програмне забезпечення (зокрема, i2 Analyst's Notebook, E-Gis maps, ArcGIS тощо) [3].

З усіх наявних методик аналізу оперативної інформації, взятих на озброєння правоохоронними органами більшості розвинених країн світу, найчастіше використовуються програмні продукти i2 та ANACAPA. Зокрема, програмний продукт i2 являє собою комп'ютерне програмне забезпечення на базі рішення SQL-server, призначене для узагальнення, аналізу, побудови ймовірнісних зв'язків, а також візуалізації обміну інформацією в режимі реального часу. Це програмне забезпечення є набором різних програм, сумісних між собою, які виконують відповідні специфічні функції на всіх етапах розкриття і розслідування злочинів.

Аналітична лінія продуктів i2 представлена елементами, здатними до розв'язання переважної більшості аналітичних задач. Їх можна застосовувати як окремо, так і в різних комбінаціях. Це, до прикладу, Analyst's Notebook, Analyst's Notebook — Esri® Edition, Analyst's Workstation, iBase, iBase IntelliShare, iXv Visualizer, iBridge, iXa, TextChart, ChartReader, PatternTracer. У сфері кримінального аналізу продукт i2 зазвичай застосовується у поєднанні з програмними продуктами iBase, iBridge, iGlass, Analyst's Workstation. Серед користувачів такої аналітичної системи можна

виділити правоохоронні органи США та Канади (ФБР, ЦРУ, DEA, NSA, RCMP), правоохоронні органи країн Євросоюзу, Інтерпол та Європол [4]. Важливим компонентом аналітики є збирання відомостей з кількох джерел. При цьому підвищується ймовірність отримання ключової доказової інформації і забезпечується можливість підтвердження і перевірки достовірності відомостей [5].

Основними недоліками, що стають на заваді прямому використанню цих продуктів в аналітичній роботі для потреб українських правоохоронців, є значна слабкість підсистем, орієнтованих на оброблення текстової інформації мовами, що не належать до романо-германської групи, та слабкість математичного апарату. Продукт і2 TextChart підтримує ефективне виділення та об'єднання сутностей лише для англійської мови, низка функцій є доступною лише для англійської, французької, німецької, іспанської та італійської мов, а засоби оброблення кирилических текстів залишаються вельми примітивними [6]. Продукт і2 PatternTracer орієнтований у першу чергу на оброблення телефонних дзвінків з метою віднаходження кластерів, що відповідають специфічній активності, та виділення ключових учасників. Аналіз інших часових послідовностей, у тому числі різнорідних, цей продукт не підтримує.

Програмний продукт компанії «Anasara Sciences Inc.» є втіленням передової методики розслідування злочинів і аналізу оперативної інформації [7]. Подібно до продуктів лінійки і2 продукти ANASARA є зручними інструментами для розв'язання багатьох важливих задач, але залишаються нерозглянутими одні й ті самі проблеми, а саме: підтримка лінгвістичних засобів аналізу документів для української та російської мов (у тому числі, поміж іншим, автоматичне здобуття з цих документів даних для аналізу та ключових фактів) та підтримка математичних засобів для аналізу взаємозв'язків того чи іншого виду між сутностями.

Отже, потребують нагального розв'язання такі проблеми: забезпечення аналітиків доступом до спеціального аналітичного програмного забезпечення шляхом розроблення цілісної системи; доопрацювання або розроблення компонент, орієнтованих на використання з ухилом в українську та російську мови; створення захищеної мережі для організації обміну інформацією в електронній формі; створення ІТ-системи збирання та оброблення оперативно-аналітичної інформації, у тому числі автоматизованих засобів, що використовують машинне навчання для здобуття даних з текстів.

Поки що за межами розгляду залишається низка проблем, пов'язаних з природою джерел даних, а саме те, що зазвичай аналіз розподіляється на аналіз даних з відкритих джерел (OSINT), агентурних джерел (HUMINT), даних радіотехнічної розвідки (SIGINT), багатьох джерел (multi-source analysis) тощо, що так само зумовлює потребу в розробленні відповідних спеціалізованих інструментів. Для їхнього розв'язання запропоновано архітектуру системи віднаходження та встановлення взаємозв'язків (СВВВ) (рис. 1). У цій архітектурі умисно не вказано спосіб збереження документів справи та доступу до них, а також не вказано доступ до загальнодержавних баз даних, наприклад, інформаційно-пошукової системи (ІПС) «Розшук» [8], оскільки це залежатиме від особливостей реалізації автоматизованого робочого місця (АРМ) слідчого. Основну увагу у відображенні архітектури зосереджено на взаємодії з інтелектуальними системами автоматизації.

Вибір такої архітектури зумовлений потребою в ощадливому витрачання обчислювальних ресурсів, оскільки віднаходження та встановлення взаємозв'язків є ресурсовитратним, та потребою в економії часу роботи слідчого.

Відомо, що важливу роль в аналітичній роботі відіграють аналітичні схеми. Їх складають у разі розслідування складних багатоепізодних справ, пов'язаних з діяльністю організованих злочинних груп, коли доводиться встановлювати спо-

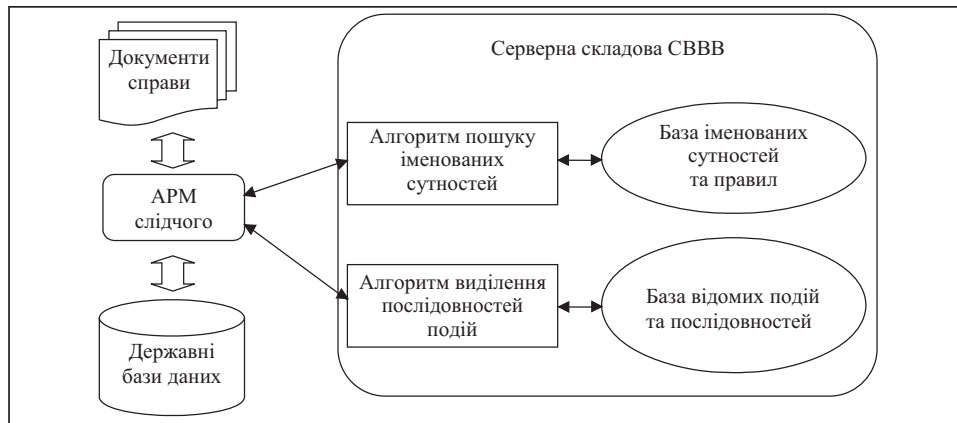


Рис. 1. Архітектура системи віднаходження та встановлення взаємозв'язків

соби вчинення ними злочинів, визначати складні зв'язки, які взаємно перетинаються між окремими учасниками злочину. У запропонованій архітектурі збереження та опрацювання аналітичних схем відбувається в АРМ слідчого, при цьому важливо, що база відомих подій та послідовностей є спільною та відкритою в межах поліції в цілому. Це дає змогу сподіватися на те, що розслідування тривалої та багатоєпізодної злочинної діяльності нарешті здійснюватиметься спільно, а не окремими відділами (звісно, після прийняття рішення про об'єднання справ). Але, щоб ухвалити таке рішення, слідчі повинні мати можливість побачити картину в цілому.

Суттєвим аспектом аналітичної роботи є аналіз географічних даних. У нинішніх умовах ці дані є джерелом цінної інформації на кожному рівні роботи правоохоронних органів. Зокрема, з метою профілактики і оперативного реагування на злочини аналітики Управління кримінального аналізу Національної поліції запроваджують географічну прив'язку до кожного житлового будинку з можливістю нанесення інформації на карту. Завдяки цьому надалі можна буде здійснювати якісний аналіз скоєних правопорушень, визначати зони вчинення злочинів, скеровувати у такі місця додаткові наряди патрульної поліції тощо. Для інтеграції географічних даних пропонується так само використовувати АРМ слідчого і передавати до серверної складової СВВВ лише текстові маркери як результат аналітичної роботи з географічними даними. У разі потреби можна доповнити їх примітками, наприклад, «вул. Хрещатик 1, поблизу від вул. Прорізна 16/2».

Як видно з рис. 1, взаємодія документів справи з автоматизованим інструментом віднаходження іменованих сутностей (П.І.Б, адреси, дати тощо) відбувається незалежно від взаємодії з інструментом виділення послідовностей подій. Це зумовлено тим, що на першому етапі слідчий з використанням АРМ має встановити всі важливі для конкретної справи сутності та згадування їх у текстах, бо крім важливих може бути дуже багато різних згадувань неважливих сутностей, аж до назви державних установ, і для якісної роботи дуже бажано одразу відкинути непотрібні дані. У разі потреби до них можна буде повернутися, якщо розслідування почне гальмуватись або зайде в глухий кут. На другому етапі система має віднайти послідовності подій та встановити можливі зв'язки та фактичні імена/назви для формальних сутностей «зловмисник», «інструмент злочину» тощо, використовуючи як дані, отримані з матеріалів слідства, так і дані, отримані з загальнодержавної бази відомих подій та послідовностей подій.

Слід зауважити, що для ефективного використання результатів процедури визначення іменованих сутностей необхідно, щоб вона виконувалась одночасно з про-

цедурою зіставлення різних згадувань відповідним сутностям. Але для розв'язання обох цих задач немає ані можливостей, ані засобів для отримання значного обсягу повністю та якісно розмічених текстів, потрібних для навчання відповідних систем. Таким чином, це змушує розробників або спиратися на правила, або застосовувати моделі тексту, що здатні використовувати частково розмічені дані [9].

За основну модель тексту для цих задач запропоновано взяти формалізм умовних випадкових полів. Умовне випадкове поле (conditional random field, CRF) — це дискримінативна ненаправлена ймовірнісна графічна модель [10], що використовується для представлення спільних розподілів набору декількох випадкових змінних. Суттєвою перевагою моделі є те, що вона не вимагає незалежності спостережуваних змінних.

Умовне випадкове поле визначається так.

1. Задано неорієнтований граф (або фактор-граф)  $G = (V, E)$ , де кожна вершина  $v \in V$  є випадковою змінною  $X$  і кожне ребро  $\{u, v\} \in E$  являє собою залежність між випадковими величинами  $u$  і  $v$ .

2. Задано набір індикаторних функцій або факторів  $\{\varphi_k\}$ , по одній для кожної кліки (кліка — це повний підграф неорієнтованого графу) у графі  $G$ . Функція  $\varphi_k$  ставить у відповідність кожному можливому стану елементів кліки певне невід'ємне дійсне число.

Вершини, які не є суміжними, повинні відповідати умовно незалежним випадковим величинам. Група суміжних вершин утворює кліку, і набір станів вершин є аргументом відповідної індикаторної функції.

Спільний розподіл набору випадкових величин обчислюють за формулою

$$P(x) = \frac{1}{Z} \prod_k (\varphi_k(x_k)),$$

де  $\varphi_k(x_k)$  — індикаторна функція, яка описувала стан випадкових величин у  $k$ -й кліці;  $Z$  — коефіцієнт нормалізації, який обчислюють за формулою

$$Z = \sum_{x \in X} \prod_k (\varphi_k(x_k)).$$

Множина вхідних лексем (слів текстового корпусу)  $X = \{x_t\}$  і множина відповідних їм типів  $Y = \{y_t\}$  разом утворюють множину випадкових змінних  $V = X \cup Y$ . Для того, щоб здобути інформацію з тексту, потрібно визначити умовну ймовірність  $P(Y | X)$ . Індикаторна функція має вигляд

$$\varphi_k(x_k) = \exp \left( \sum_k \lambda_k f_k(y_t, y_{t-1}, x_t) \right),$$

де  $\{\lambda_k\}$  — дійсний вектор параметрів,  $\{f_k(y_t, y_{t-1}, x_t)\}$  — множина функцій ознак.

Тоді можна задати лінійне умовне випадкове поле, при цьому умова лінійності впливає з властивостей предметної області (аналіз тексту) та бажання отримати задачу, що має прийнятну обчислювальну складність. Розподіл імовірностей задають так:

$$P(x | y) = \frac{1}{Z(x)} \prod_k \exp \left( \sum_k \lambda_k f_k(y_t, y_{t-1}, x_t) \right) \quad (1)$$

з відповідним коефіцієнтом нормалізації

$$Z(x) = \sum_y \prod_k \exp \left( \sum_k \lambda_k f_k(y_t, y_{t-1}, x_t) \right). \quad (2)$$



Подібно до іншого дискримінативного методу МЕММ [11] вибір фактор-ознак залежить від даних, але підхід з використанням умовних випадкових полів передбачає більшу свободу вибору і може враховувати практично будь-які залежності, які існують у вхідних даних. Також важливо, що для навчання моделі можна застосовувати добре відомий алгоритм Вітербі.

Крім цього, в умовних випадкових полях відсутній так званий зсув міток (label bias problem) [12] — ситуація, коли перевагу отримують стани з меншою кількістю переходів, оскільки здійснюється побудова єдиного розподілу ймовірностей, і нормалізація (коефіцієнт  $Z(x)$  з формул (1), (2)) проводиться в цілому, а не в межах окремого стану. Це, безумовно, є перевагою методу — алгоритм не вимагає припущення щодо незалежності спостережуваних змінних. Крім того, використання довільних факторів дає змогу описати різні ознаки об'єктів, що визначаються, і за рахунок цього послабити вимоги до повноти і обсягу навчальної вибірки. Залежно від розв'язуваної задачі, на практиці обсяг від декількох сотень тисяч до мільйонів термів є достатнім. При цьому точність визначатиметься не лише обсягом вибірки, а й обраними факторами, отже комп'ютерні лінгвісти матимуть простір для досліджень та оптимізації моделі. Недоліком застосування підходу на основі умовних випадкових полів є складність обчислювальних процесів під час аналізу навчальної вибірки. Це утруднює постійне оновлення моделі у разі появи нових навчальних даних, але за умови існування єдиної відомчої бази існує технічна можливість автоматично оновлювати модель, наприклад, щомісяця.

З огляду на три зазначені вище проблеми актуальною також є потреба у створенні засобів автоматизації для віднаходження подібностей у різномірних даних та ситуаціях, описаних цими даними. Відповідним оптимальним рішенням є машинне навчання на основі слабоструктурованих даних, здобутих з текстів документів. Під час створення компонент машинного навчання потрібно враховувати особливості методів та форму отриманих результатів. Є низка моделей, що дають змогу встановити наявність сутностей і/або зв'язків між сутностями, але при цьому відсутня можливість виділення дійсних чинників (що, наприклад, характерно для багатьох архітектур нейронних мереж), а значить, і можливість чітко вказати перелік факторів, що зумовили той чи інший результат оцінки даних комп'ютером. До того ж, потрібно забезпечити можливість перенавчання моделі. У тому разі, коли вона буде залежність з урахуванням насправді неважливих факторів, потім на етапі отримання даних, де ці неважливі фактори проявилися інакше порівняно з навчальною вибіркою, модель дає неправильний прогноз. Обидва зазначені аспекти можуть суттєво вплинути на результативність роботи підрозділу, який використовує відповідні аналітичні засоби, найімовірніше, негативно. Це зумовлює виникнення низки питань практичного, а, можливо, й процесуального та юридичного характеру, якщо неправильні висновки спричинили значні негативні наслідки.

Як основну архітектуру запропоновано застосувати інструмент глибокого навчання — так звану довгу короткострокову пам'ять (LSTM) [13]. Це зумовлено тим, що значна частка даних утворює хронологічні послідовності. Більш того, саме впорядкування за часом дає змогу встановити, яка з подій була причиною, а яка — наслідком.

LSTM-мережа — це рекурентна нейронна мережа, що містить спеціалізовані LSTM-модулі замість інших складових або на додаток до них. LSTM-модуль — це рекурентний модуль, що складається з комірки пам'яті, вхідного вентиля, вихідного вентиля і вентиля забування. На рівні концепту вхідний вентиль контролює міру входження нового значення в пам'ять, вихідний вентиль визначає міру впливу значення, отриманого в модулі, на результат, вентиль забування визначає міру швид-

кості, з якою забувається значення, що зберігається в модулі. У разі простого навчання рекурсивних нейронних мереж у міру поширення помилки в часі, розмір помилки, що обчислюється «назад» у часі, може або спадати до нуля або неконтрольно наростати, залежно від початкових значень вагових коефіцієнтів. У першому випадку це може призвести до коливань вагових коефіцієнтів, у другому — до неконтрольного їх зростання, що, своєю чергою, практично унеможливує встановлення взаємозв'язку між подіями, рознесеними в часі.

Опис функціонування LSTM-мережі в матричній формі має такий вигляд:

$$\begin{aligned} f_t &= \delta_g(W_f x_t + U_f h_{t-1} + b_f), \\ i_t &= \delta_g(W_i x_t + U_i h_{t-1} + b_i), \\ o_t &= \delta_g(W_o x_t + U_o h_{t-1} + b_o), \\ c_t &= f_t \circ c_{t-1} + i_t \circ \delta_c(W_c x_t + U_c h_{t-1} + b_c), \\ h_t &= o_t \circ \delta_h(c_t), \end{aligned}$$

де  $t$  — момент часу (крок),  $x_t \in R^l$  — вхідний вектор LSTM-модуля,  $f_t \in R^k$  — вектор активації вентиля забування,  $i_t \in R^k$  — вектор активації вхідного вентиля,  $o_t \in R^k$  — вектор активації вихідного вентиля,  $h_t \in R^k$  — вихідний вектор LSTM-модуля,  $c_t \in R^k$  — вектор станів комірок пам'яті,  $W \in R^{k \times l}$ ,  $U \in R^{k \times k}$  — матриці вагових коефіцієнтів, що підлягають вивченню,  $b \in R^k$  — вектор зсувів, що підлягають вивченню,  $\delta_g$  — сигмоїдальна функція активації,  $\delta_c$  — гіперболічний тангенс,  $\delta_h$  — гіперболічний тангенс (для спеціальних мереж — функція тотожності), а операція  $\circ$  є позначенням поелементного добутку. На першому кроці стани комірок та вихідні значення дорівнюють нулю.

Головною перевагою LSTM-мереж є саме здатність мережі до налаштування на успішне ігнорування непотрібних вхідних даних і завчання лише потрібних.

Суттєвою проблемою, яка виникає у всіх нейронних мереж, є потреба у перенавчанні. Для LSTM-мереж ефективним рішенням є зменшення розміру вектору параметрів, які повинна вивчити мережа. Отже, так само, як і у звичайній практиці, добре діє принцип «бритва Оккама», коли додавання нових потенційних параметрів має сенс лише тоді, коли попередні кроки показали невідповідність результатів навчання.

Таким чином, сучасний розвиток ЕОМ та математики надає потужні інструменти для дослідження наборів послідовних подій (фактів). Отже, дуже бажано реалізувати та впровадити відповідну систему, яка забезпечить можливість для віднаходження схожих обставин, оскільки це має сприяти пошуку злочинців, які характеризуються більш-менш сталим способом дії.

Як основний елемент для оброблення запропоновано структуру такого виду: «Іменована сутність», «Де», «Коли», «Дія», «Об'єкт дії», «Інструмент», «Примітки». Теоретично, можна обійтися набором більш простих структур, об'єднаних у керувальний простір [14] як показано в [15], шляхом розділення складних структур на послідовність пар («об'єкт»–«властивість») та трійок («суб'єкт»–«предикат»–«об'єкт»). Проте, наявність деталізованих даних суттєво полегшує аналіз, дає змогу скоротити кількість хибних гіпотез та, насамперед, оптимізувати обчислення, оскільки можна застосувати упорядкування подій у часі як основну евристику. І справді, хоча в загальному випадку не можна стверджувати, що «якщо після того, то у наслідок», але можна стверджувати, що «якщо до того, то не у наслідок».

Під час використання зазначених вище засобів у розглянутій архітектурі добре проявляються сильні сторони відповідних засобів. У разі визначення іме-

нованих сутностей з використанням умовних випадкових полів перевагою є відсутність потреби розмічати всі наявні у текстах іменовані сутності, при цьому, якщо якась сутність у загальній базі відсутня, то систему можна довести на її розпізнавання та внести у бази. Для системи визначення послідовностей важливим є те, що завдяки евристичній оптимізації можна суттєво прискорити роботу системи, а завдяки перевагам LSTM-мережі віднайти зв'язки між подіями, навіть якщо вони рознесені у часі та мають між собою велику кількість інших подій.

У підсумку зазначимо, що кримінальний аналіз є специфічним видом інформаційно-аналітичної діяльності правоохоронних органів, який полягає у перевірці та оцінці інформації, її інтерпретації, встановленні зв'язків між даними, отриманими у процесі виявлення, припинення та розслідування злочинів, які мають значення для оперативно-розшукової діяльності та кримінального провадження, з метою їхнього використання правоохоронними органами і судом, а також подальшого проведення оперативного, тактичного і стратегічного аналізу.

На нашу думку, успішна реалізація та впровадження нових методів кримінального аналізу дадуть змогу в майбутньому поширити їх на всю систему Національної поліції та активно використовувати сучасні аналітичні методи і прийоми, завдяки яким можна створити передумови для більш ефективного виконання оперативними і слідчими підрозділами відповідних завдань, що, своєю чергою, сприятиме підвищенню ефективності протидії злочинності в цілому.

#### СПИСОК ЛІТЕРАТУРИ

1. Правоохоронна діяльність, керована аналітикою: передова методика сучасної правоохоронної діяльності. URL: <http://euam.php7.postbox.kiev.ua/ua/news/opinion/intelligence-led-policing-the-cutting-edge-of-modern-law-enforcement>.
2. Carter J.G., Phillips S.W., Gayadeen S.M.. Implementing Intelligence-led policing: an application of loose-coupling theory. *Journal of Criminal Justice*. 2014. N 42. P. 433–442.
3. Кримінальний аналіз — це ефективна робота поліції та безпека громадян. URL: [http://mvs.gov.ua/ua/news/10309\\_Kriminalniy\\_analiz\\_ce\\_efektivna\\_robota\\_policii\\_ta\\_bezpeka\\_gromadyan\\_FOTO.htm](http://mvs.gov.ua/ua/news/10309_Kriminalniy_analiz_ce_efektivna_robota_policii_ta_bezpeka_gromadyan_FOTO.htm).
4. Возможности использования аналитических программ в борьбе с организованной преступностью. URL: <https://articlekz.com/article/11838>.
5. Работа полиции. Системы полицейской информации и разведки : пособие по оценке систем уголовного правосудия. Нью-Йорк: Управление Организации Объединенных Наций по наркотикам и преступности. Вена, 2010. 38 с.
6. IBM i2 TextChart Release Notes. URL: <http://www-01.ibm.com/support/docview.wss?uid=swg27027043>.
7. О компании Anacapa Sciences Inc. URL: <http://www.spi2.ru/about/partners/anacapa>.
8. ИПС «Позшук». URL: <https://wanted.mvs.gov.ua>.
9. Yuta Tsuboi, Hisashi Kashima, Shinsuke Mori, Hiroki Oda, Yuji Matsumoto. Training conditional random fields using incomplete annotations. *Proceedings of the 22nd International Conference on Computational Linguistics — Coling 2008* (Manchester, Aug. 2008), 2008. P. 897–904. URL: <http://www.aclweb.org/anthology/C08-1113>.
10. Sutton C., McCallum A. An introduction to conditional random fields for relational learning. URL: [http://odjt9j2ec.bkt.clouddn.com/hmm\\_sub\\_crf-crf-tutorial.pdf](http://odjt9j2ec.bkt.clouddn.com/hmm_sub_crf-crf-tutorial.pdf).
11. McCallum A., Freitag D., Pereira F. Maximum entropy Markov models for information extraction and segmentation. URL: <http://www.ai.mit.edu/courses/6.891-nlp/READINGS/maxent.pdf>.
12. Lafferty J., McCallum A., Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the 18th International Conference on Machine Learning* (Williamstown, Massachusetts), 2001. P. 282–289.
13. Hochreiter S., Schmidhuber J. Long short term memory. *Neural Computation*. 1997. Vol. 9, N 8. P. 1735–1780. URL: <http://www.bioinf.jku.at/publications/older/2604.pdf>.



14. Анисимов А.В. Управляющее пространство синтаксических структур естественного языка. *Кибернетика*. 1990. № 3. С. 11–17.
15. Anisimov A., Marchenko O., Taranukha V., Vozniuk T. Semantic and syntactic model of natural language based on non-negative matrix and tensor factorization. *Proceedings of the 9th International Conference on NLP, PolTAL 2014* (Warsaw, Poland, Sept. 17–19, 2014). *Lecture Notes in Artificial Intelligence*. 2014. Vol. 8686. P. 177–184.

Надійшла до редакції 20.12.2017

**А.В. Мовчан, В.Ю. Тарануха**

**ПОСТРОЕНИЕ СИСТЕМЫ АВТОМАТИЗАЦИИ ДЛЯ ВНЕДРЕНИЯ МОДЕЛИ ПОЛИЦЕЙСКОЙ ДЕЯТЕЛЬНОСТИ, УПРАВЛЯЕМОЙ АНАЛИТИКОЙ, В ОРГАНАХ НАЦИОНАЛЬНОЙ ПОЛИЦИИ УКРАИНЫ**

**Аннотация.** Проанализированы актуальные проблемы внедрения модели полицейской деятельности, управляемой аналитикой, в органах Национальной полиции Украины. На основании результатов этого анализа предложена архитектура аналитической системы и рассмотрен математический аппарат построения элементов будущей системы. Успешное внедрение и применение новых методов криминального анализа позволит в дальнейшем распространить его на всю систему Национальной полиции Украины, а также активно использовать аналитические методы и технологии, обеспечивающие эффективное расследование уголовных правонарушений.

**Ключевые слова:** информационно-аналитическая работа, криминальный анализ, модель полицейской деятельности, управляемой аналитикой, средства автоматизации, компьютерная лингвистика, условные случайные поля, глубокое обучение.

**A.V. Movchan, V.Yu. Taranukha**

**CONSTRUCTING AN AUTOMATION SYSTEM TO IMPLEMENT THE INTELLIGENCE-LED POLICING INTO THE NATIONAL POLICE OF UKRAINE**

**Abstract.** The paper analyzes implementation of the Intelligence-Led Policing into the National Police of Ukraine. The architecture of analytical system is proposed based on the analysis of state-of-the-art in the National Police of Ukraine. The mathematical apparatus is considered for the future components of the analytical system. Successful implementation and application of new methods of criminal analysis will allow expanding it to the entire system of the National Police of Ukraine in the future and to actively use analytical methods and technologies that provide efficient investigation of criminal offenses.

**Keywords:** information and analytical work, criminal analysis, Intelligence-Led Policing, automation systems, computational linguistics, conditional random fields, deep learning.

**Мовчан Анатолій Васильович,**

доктор юр. наук, старший науковий співробітник, професор кафедри Львівського державного університету внутрішніх справ, e-mail: movchan.anatol@gmail.com.

**Тарануха Володимир Юрійович,**

кандидат фіз.-мат. наук, асистент кафедри Київського національного університету імені Тараса Шевченка, e-mail: taranukha@ukr.net.