Marek Kisiel-Dorohinicki[1], Rafał Dreżewski[1],
Łukasz Hernik, Tomasz Miłoś[1], Kamil Piętak[1], Aleksander Pohl[1]

2. Zintegrowane środowisko wspomagania analizy kryminalnej

Proces analizy kryminalnej jest pracochłonny, bardzo złożony i często opiera się na ogromnej ilości danych. Praca analityka kryminalnego ma na celu pozyskanie informacji, które nie są dostępne wprost, lecz pojawiają się dopiero po przeprowadzeniu wnikliwego oglądu dostępnych danych pod różnymi kątami. Co więcej, próba przejścia od surowych danych do wniosków powinna często posiadać wartość dowodową. Wykonana ekspertyza musi zatem zostać solidnie udokumentowana, aby nie mogła być podważona w postępowaniu sądowym.

Z uwagi na dużą pracochłonność czynności analitycznych i ograniczone możliwości człowieka, zastosowanie narzędzi informatycznych może znacząco ułatwić pracę analityka kryminalnego. Systemy informatyczne doskonale nadają się do przetwarzania dużych ilości danych, a zastosowanie zautomatyzowanego podejścia algorytmicznego może pozwolić na wydobycie ze zbiorów danych wielu cennych informacji, które wymagałby wielu godzin żmudnej i podatnej na pomyłki pracy człowieka. Oczywiście nie sposób wyobrazić sobie, że w tak złożonej materii zastosowanie rozwiązań informatycznych wyeliminuje udział człowieka w procesie analizy. Spodziewać się raczej należy, że analityk, który będzie mógł pracować na materiale dobrze przygotowanym, korzystając z wielu wariantów wizualizacji, śledząc dane z zastosowaniem filtrów i wbudowanych algorytmów, będzie mógł realizować swoje zadania znacznie efektywniej, koncentrując się na ich najważniejszych aspektach.

Ogólna charakterystyka zadań analizy kryminalnej widzianych z perspektywy dzisiejszej praktyki polskiej Policji, a także zakres możliwości ich wspomagania przez systemy informatyczne stanowi wprowadzenie do niniejszego rozdziału. Prezentowane w dalszej części rozdziału zintegrowane środowisko wspomagania analizy kryminalnej LINK jest realizacją pomysłu na zunifikowane narzędzie, które można wykorzystać zarówno do automatyzacji podstawowych czynności, takich jak wstępne przetwarzanie danych pochodzących z różnych źródeł, czy ich wizualizacja, jak również w zakresie bardziej zaawanasowanych technik pozwalających szybko ocenić dostępne dane i wyciągnąć wstępne wnioski pod kątem dalszej ich analizy. Szczególną cechą środowiska LINK, projektowanego we współpracy z polskimi analitykami kryminalnymi, jest fakt, iż w odróżnieniu od innych systemów tego typu dostępnych na rynku, dostarcza ono narzędzi dostosowanych do potrzeb polskich służb mundurowych. Dzięki wykorzystaniu architektury opartej o technologię Eclipse Rich Client Platform, LINK jest platformą rozszerzalną. Dodanie nowej funkcjonalności odbywa się przez

[1] Katedra Informatyki, Wydział EAIiE, Akademia Górniczo-Hutnicza w Krakowie.

dostarczenie nowego modułu (ang. *plugin*). Nowe rozszerzenia jak i aktualizacje istniejących modułów mogą być pobierane z Internetu dzięki mechanizmowi zdalnej aktualizacji.

2.1. Wprowadzenie w zagadnienia analizy kryminalnej

Niewątpliwie początki analizy kryminalnej[2] należy wiązać ze Stanami Zjednoczonymi, gdy w latach 60. ubiegłego wieku Prezydencka Komisja ds. Przestępczości, w raporcie opisała zagrożenie społeczeństwa amerykańskiego przestępczością zorganizowaną. Przedstawiony, w oparciu o doświadczenia organów ścigania, kształt analizy kryminalnej w znacznym zakresie jest zbliżony do tego, z jakim obecnie można zetknąć się w Policji. Czy można jednoznacznie stwierdzić, od kiedy zaczęto stosować analizę kryminalną w polskiej Policji? Wydaje się, że przypisywanie sztywnych dat, które wskazywałyby na powstanie analizy kryminalnej jest nieporozumieniem. O ile można mówić o pewnych metodach analitycznych czy też wskazywać zakresy dat, w których rozpoczęto korzystanie z zaawansowanych metod i narzędzi analizy kryminalnej (np. początki stosowania oprogramowania analitycznego można datować na połowę lat 90.), to stosowanie analizy miało miejsce dużo wcześniej, a może od zawsze – czy policjant prowadzący postępowanie przygotowawcze, czy też prokurator w ramach śledztwa nie wykonują analiz kryminalnych? Podobne pytanie można postawić również w odniesieniu do innych służb tzw. porządku prawnego.

Warto więc wyjaśnić czym jest analiza kryminalna. Trafnym wydaje się posługiwanie definicją opracowaną przez Międzynarodową Organizację Policji Kryminalnych „Interpol", gdzie przez analizę kryminalną rozumie się metodę pracy policji, która polega na konsekwentnym i zorganizowanym wyszukiwaniu i wykazywaniu związków danych dotyczących przestępstwa z innymi możliwymi do wyróżnienia informacjami, które będą stanowiły podstawę do przygotowania wniosków wspomagających procesy decyzyjne – często formułowane w odpowiednich postanowieniach. Bardzo zbliżona definicja przyjęta w Policji określa analizę kryminalną jako czynność związaną z poszukiwaniem i identyfikacją powiązań między informacjami dotyczącymi przestępstwa lub przestępcy oraz wszelkimi innymi danymi uzyskanymi z różnych źródeł i wykorzystanie ich do celów operacyjnych i procesowych.

Propagowana przez MOPK „Interpol" w krajach członkowskich wiedza o metodologii analizy kryminalnej zaowocowała, przy finansowym wsparciu ze strony UE w ramach funduszu „PHARE", utworzeniem w polskiej Policji pierwszych zalążków komórek analizy kryminalnej wyposażonych w sprzęt wraz z oprogramowaniem analitycznym (Analyst's Notebook i iBase). Oczywiście bardzo cennym okazało się wykorzystanie doświadczeń amerykańskich oraz krajów Unii Europejskiej. W Komendzie Wojewódzkiej Policji w Krakowie w roku 2001 powstał pierwszy w polskiej Policji dziesięcioosobowy Wydział Analizy Kryminalnej, który od początku stał się cennym elementem wykorzystywanym w prowadzonych sprawach zarówno procesowych, jak i operacyjnych. Sposób funkcjonowania wspomnianego wydziału

[2] Uznając merytoryczną wartość artykułu Stefana Czarneckiego pt. *Analiza kryminalna – narzędzie pracy Policji* opublikowanego w czasopiśmie PROKURATOR 1(29)/2007 potwierdzoną ponadto w licznych rozmowach ze specjalistami w dziedzinie, autorzy niniejszego rozdziału zdecydowali się wykorzystać treść tego artykułu jak następuje: podrozdział 2.1 stanowi przedruk części początkowej oryginalnej publikacji, zaś podrozdział 2.2 jest istotnie wzorowany na jej dalszych sekcjach. Uzyskano zgodę redakcji pisma PROKURATOR na wykorzystanie tekstu w ten sposób. Niestety z powodu śmierci Autora powyższe postępowanie mogło być jedyną formą współpracy z Nim jako prekursorem i popularyzatorem analizy kryminalnej w Polsce.

w kontekście korzyści jakie były widoczne gołym okiem dostarczył niezbędnej wiedzy, na podstawie której zostało przesądzone powstanie w Policji kolejnych wyspecjalizowanych komórek analizy kryminalnej.

2.1.1. Kiedy stosować analizę kryminalną?

Realizacja poszczególnych celów dochodzeń czy śledztw (ustalenie sprawcy, uzyskanie niezbędnych dowodów popełnienia przestępstwa, odzyskanie mienia) realizowane jest przez proces przetwarzania informacji oparty na logicznym i kreatywnym sposobie myślenia, jakim jest analiza, ocena i interpretacja informacji.

Szczególnie ważnym elementem jest stosowanie ujednoliconych technik stawiania hipotez, rekonstrukcji przebiegu poszczególnych przestępstw kryminalnych, identyfikacji kolejnych przestępstw, określania struktury grup i związków przestępczych oraz analizowania zakresu i sposobu prowadzenia działalności przestępczej.

Analiza kryminalna jest szczególnie przydatna, a czasami wręcz niezbędna, potwierdzając swoje zastosowanie:

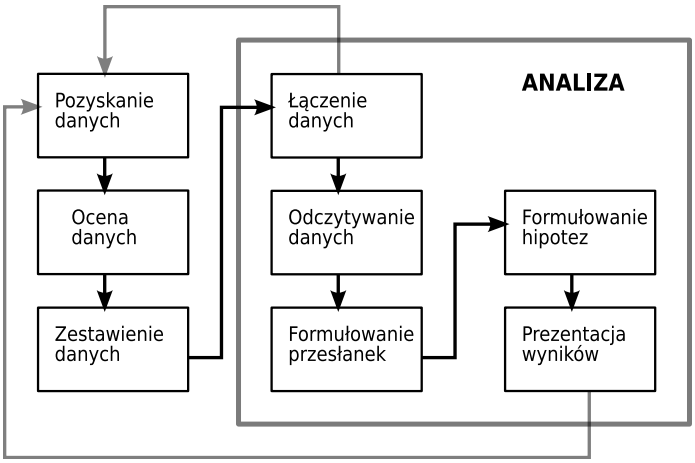
- w sprawach wielowątkowych,
- obejmujących duży zasięg terytorialny, w których występuje skomplikowana i rozbudowana struktura powiązań przestępczych połączona ze znaczną ilością informacji.

Kolejne tomy gromadzonych informacji, wkrótce powodują sytuację, gdy ogarnięcie kolejnych faktów i ustaleń staje się niemożliwe, a wychwycenie jakichkolwiek nieścisłości w gromadzonym materiale graniczy z cudem.

Obecnie rozwój technologiczny, jakiego doświadczamy na co dzień, przekłada się na wzrost ilości informacji. Gromadzone w sprawach tomy uzyskanych wykazów połączeń telefonicznych, transakcji finansowych czy też przepływów towarów wymagają niestandardowego podejścia i umiejętności, jakimi dysponuje właśnie analityk kryminalny. Co ważne, tradycyjnie stosowane metody pozwalają na wychwycenie tych informacji, które spodziewamy się znaleźć, jednak takie postępowanie powoduje, że pozostałe informacje istotne dla sprawy „uciekają" nam lub też trafiamy na nie w sposób przypadkowy. Z jednej strony jesteśmy zalewani informacją, a z drugiej ciągle informacji tej poszukujemy. Pamiętać jednak musimy, że analiza kryminalna nie funkcjonuje jako oderwany element, ale jest jednym z trzech podstawowych filarów określanych łącznie jako wywiad kryminalny, gdzie podstawowy proces oparty jest na pozyskaniu, ocenie, gromadzeniu i analizie informacji – ze szczególnym naciskiem położonym na analizę (rys. 2.1). Przedstawiony sposób podejścia do informacji pozwala na ukierunkowane i efektywne pozyskiwanie informacji, a tym samym efektywne prowadzenie spraw. Kiedy natomiast sięganie po instrumenty analizy kryminalnej jest najmniej efektywne? Wtedy gdy korzysta się z tego narzędzia jedynie w końcowym etapie prowadzonej sprawy, gdzie często prowadzący postępowanie stoi przed przysłowiową ścianą, nie mogąc poradzić sobie z ogromem zebranych informacji.

2.1.2. Rodzaje analizy kryminalnej

Mając na uwadze cel, zakres zastosowania oraz oczekiwany efekt możemy podzielić analizę kryminalną na następujące dwa rodzaje: **analizę operacyjną** oraz **analizę strategiczną**.



Rys. 2.1. Etapy procesu wywiadu

Często jako nieporozumienie i błąd jest traktowanie analizy operacyjnej jako narzędzia stosowanego do prowadzonych przez policjantów spraw operacyjnych. Metody i techniki stosowane przy analizie operacyjnej są tożsame dla obu obszarów (procesowy, operacyjny) a jedyne, co je rozróżnia, to zakres materiału, jaki w danej analizie może zostać użyty.

Analiza operacyjna – służy osiągnięciu w krótkim czasie zamierzonego przez organy ścigania celu w postaci aresztowania sprawcy, zajęcia przedmiotu przestępstwa lub jego konfiskaty [3, s. 17]. Podejmowane próby wyodrębniania, dodatkowego podziału analizy kryminalnej na analizę ekonomiczną, finansową wydaje się niecelowe i bezzasadne. Aczkolwiek bardzo często w ramach uczestniczenia w różnego rodzaju warsztatach z takim pojęciami się zetkniemy. Podkreślanie dodatkowego podziału bardziej związane jest z zakresem analizowanych informacji niż ze stosowaniem poszczególnych metod analizy kryminalnej. Cel, jaki chcemy osiągnąć, powoduje, że właśnie z tą analizą najczęściej się spotkamy.

Analiza strategiczna – jej przedmiotem są problemy i cele długoterminowe, ustalenie priorytetów i strategii zwalczania przestępczości kryminalnej na podstawie dogłębnych badań oraz prognozowanie jej rozwoju [3, s. 17]. Jest niezbędnym produktem do wsparcia procesów decyzyjnych. Najprawdopodobniej już wkrótce powstanie jako jedyna taka komórką w polskiej Policji, Wydział Wywiadu Strategicznego w ramach Biura Wywiadu Kryminalnego KGP, który będzie realizował wspomniany zakres analiz dostarczając decydentom niezbędnych informacji do podejmowania najbardziej trafnych decyzji do opracowywania koncepcji zwalczania nie tylko samej przestępczości, ale przede wszystkim wskazywania rozwiązań do niwelowania czynników je powodujących.

2.1.3. Formy analizy kryminalnej operacyjnej

W ramach tej analizy rozróżnić możemy pięć podstawowych form [4, s. 15 i nast.]:

1. **Analiza konkretnej sprawy (analiza przestępstwa)** – jest to próba chronologicznej rekonstrukcji przebiegu przestępstwa polegająca na ustaleniu kolejności składających

się na nie zdarzeń, w celu zalecenia dalszego kierunku pracy oraz stwierdzenia nieścisłości w informacjach pochodzących z różnych źródeł. Pozwoli odpowiedzieć nam na pytania o przebieg zdarzenia, wskazać na sprzeczności czy też luki informacji w zgromadzonym materiale, ale również często ukazuje przełożenie zdarzenia na jego skutek. Najczęściej przyjmie postać rekonstrukcji zdarzenia.

2. **Analiza porównawcza spraw** – jest to wyszukiwanie (według ustalonego wzorca oraz w oparciu o wybrane kryteria) i kojarzenie informacji o podobnych zdarzeniach przestępczych w celu ustalenia, które z nich mogły być popełnione lub zorganizowane przez te same osoby (przestępstwa seryjne). Pomoże wskazać nam, jakie czyny i w jakim zakresie są podobne do siebie, okoliczności wskazujące, że zostały popełnione przez tego samego sprawcę/sprawców, a często wyodrębnić ślady przemawiające za tym. Bardzo często pozwala, do już toczącego się postępowania, w którym ustalono sprawcę/sprawców, dołączyć kolejne przestępstwa dotychczas niewykryte.

3. **Analiza grup przestępczych** – jest to uporządkowanie dostępnych informacji o znanej (lub przypuszczalnej) grupie przestępczej, w celu określenia jej struktury, zakresu działalności oraz roli każdego z jej członków lub instytucji z nią związanych. Jest to bardzo często zlecana analitykom forma analizy operacyjnej, gdzie w ramach wyniku przedstawiana jest organizacja grupy, zaplecze logistyczne lub źródła finansowania, hierarchia i podział ról. Wielokrotnie to właśnie analityk kryminalny przedstawia informacje, które wskazują na osoby „pozornie" nie związane z grupą jako istotny element tej właśnie grupy.

4. **Analiza profilu szczególnego** – jest to próba określenia na podstawie opisu przestępstwa cech charakterologicznych i fizycznych osoby, która je popełniła. Nie jest to jednak analiza wykonywana samodzielnie przez analityka kryminalnego, ale z uwagi na złożoność problematyki zmuszony jest on korzystać ze specjalistycznej wiedzy i doświadczenia osób posiadających odpowiednie przygotowanie (najczęściej psychologów, psychiatrów), którzy pomagają podkreślić szczególne umiejętności, wiedzę lub zdolności sprawcy, wskazać ilość sprawców. Najczęściej opracowanie w tym zakresie oparte jest na informacjach pozostawionych na miejscu zdarzenia.

5. **Analiza prowadzenia sprawy** – jest to kompleksowa ocena działań i czynności zrealizowanych w danej sprawie w celu określenia dalszego jej przebiegu oraz wpływu określonych czynności na skuteczność i efektywność działań wykrywczych. Niestety bardzo często utożsamiana z kontrolą prowadzonej sprawy. A w jej wynikach wskazuje się na zaniedbania, efektywność wykorzystania śladów dowodowych, poprawność formułowanych wniosków na podstawie posiadanych przesłanek. Głównym celem jednak nigdy nie powinno stać się dążenie do pociągnięcia do odpowiedzialności prowadzącego daną sprawę, ale wypracowania mechanizmów na podstawie których możliwe będzie w przyszłości eliminowanie stwierdzonych błędów.

2.2. Narzędzia informatyczne wspomagające analizę kryminalną

Niniejszy podrozdział będzie próbą spojrzenia na całościowy proces analizy kryminalnej z punktu widzenia osoby prowadzącej dochodzenie. Dla policjanta prowadzącego postępowanie przygotowawcze mniej istotne jest zagadnienie związane z techniczną stroną podejścia do problemu analizy, za to cenne jest omówienie **korzyści**, jakie analiza kryminalna może przynieść w codziennej pracy policjanta, dzięki zastosowaniu narzędzi informatycznych.

2.2.1. Analizowane materiały

Proces analizy jest wynikiem zapotrzebowania przez prowadzącego sprawę, który określa swoje oczekiwania co do efektów, jakie chce uzyskać. Wniosek poddania analizie kryminalnej określa odbiorcę (zleceniodawcę) efektów prac analitycznych, oraz szkic celu, jak i zakresu analizy jakim będą poddane zebrane dane. Niezbędna jest komunikacja pomiędzy zleceniodawcą a samym analitykiem, ich wzajemne ustalenia pozwolą w przyszłości na trafne określenie zakresu prac, jakie będą wykonywane przez analityka, i które będą spełniały oczekiwania za strony zleceniodawcy.

Analiza materiałów w obszarze postępowania przygotowawczego może zostać zlecona przez:

- policjanta pionu dochodzeniowego, prowadzącego postępowanie,
- prokuratora, który bezpośrednio występuje o powołanie biegłego oraz wykonanie analizy kryminalnej.

Każdy z tych wariantów jest dobry, jeśli w sposób jednoznaczny zostanie przedstawiony cel analizy kryminalnej.

Analityk w przypadku stwierdzenia niekompletności materiału lub braku odpowiednich informacji w dostarczonych do analizy danych, jest zobowiązany w ramach współpracy ze zleceniodawcą o zwrócenie się z prośbą uzupełnienia informacji. W sytuacji nie otrzymania brakującego materiału, który z różnych względów może być niedostępny, zmieniany jest obszar podlegający analizie lub następuje zwrot materiałów do zleceniodawcy. Podstawową zasadą, o której należy pamiętać, jest fakt, że praca analityka kryminalnego służy zleceniodawcy a nie na odwrót. Końcowy efekt postępowania jest bezwzględnie uzależniony od wyników, jakie dostarczy analityk. Niedopuszczalne, a wręcz szkodliwe, wprowadzające duże zagrożenie dla efektów prowadzonego przez analityka postępowania, jest zawężenie dostępu do materiałów jedynie w takim obszarze, jaki zleceniodawca uważa za wystarczający. Przecież analityk nie stanowi konkurencji dla prowadzącego i na pewno nie pozbawi zaszczytu, sukcesu, jakim może być „rozwiązanie" sprawy. Jedynie w niektórych przypadkach, ze względu na specyficzny zakres uzyskiwanych materiałów, analityk skupia się nie na całości materiału, lecz na ich części, aby np. graficznie przedstawić rozbudowaną i skomplikowaną strukturę grupy przestępczej. Z częściową analizą spraw bardzo często mamy do czynienie w przypadkach analizy dziesiątek kart wykazów historii przelewów bankowych, połączeń telefonicznych, czy faktur, które obecnie stanowią około 90–95 procent wszystkich wykonywanych analiz kryminalnych.

2.2.2. Przygotowanie danych

Zastosowanie oprogramowania analitycznego jest niepodważalnie doceniane, gdy analityk zmuszony jest przetwarzać tzw. dane masowe (np. tysiące stron bilingów telefonicznych). Jako dane źródłowe trafiają do analityka typowe zestawienia bilingów, które to dane operator udostępnia jako materiał potrzebny analitykowi podczas prowadzonych spraw. W identyczny sposób pozyskiwane są dane z systemów finansowych, które bank przekazuje do analizy w postaci zestawień transakcji na rachunku za konkretny okres rozliczeniowy.

Zaskakujące i godne podziwu jest, gdy tak obszerne (wielostronicowe) dane, poddawane są „ręcznej obróbce" i gdy pojedyncze numery zaznaczane są różnymi kolorami. Takie podejście jest mało efektywne i np. przy kolejnym zaznaczonym numerze telefonu zniechęcona osoba może się poddać, bo stosując tak pracochłonną metodę, z góry skazana jest na porażkę. Dopiero po uświadomieniu sobie złożoności analizy przekazuje materiał do właściwej osoby czyli analityka kryminalnego, lecz nie jest to już oryginalny dokument a brudnopis z naniesionymi notatkami (kolorowymi zakreśleniami). Gdy proces analizy musi rozpocząć się od zamiany formy papierowej na postać elektroniczną, wprowadzone wcześniej zapiski, podkreślenia przez osobę próbującą podjąć wyzwanie ręcznej analizy, utrudniają lub często wykluczają możliwość automatycznego rozpoznania tekstu po procesie skanowania, co w znacznym stopniu komplikuje i tym samym wydłuża cały proces analizy.

Na stopień skomplikowania i czas wykonywanej analizy w znacznym stopniu wpływa proces sprowadzania danych do postaci elektronicznej bazy danych. Baza elektroniczna będzie podstawą w dalszej analizie i czym szybciej uda się ją wygenerować, tym szybciej będzie można rozpocząć proces właściwej analizy. Jednoznacznie nasuwa się wniosek, że uzyskanie jak największej ilości materiałów w formie elektronicznej pozwala znacznie przyśpieszyć cały proces, dlatego istotne jest, aby na etapie gromadzenia danych od podmiotów, które administrują tego rodzaju danymi (np. operatorzy telekomunikacyjni, instytucje finansowe) na samym początku wskazać im w jakim formacie powinny być otrzymane pozyskiwane dane, tak aby zminimalizować ryzyko ręcznego przepisywania z papierowych wydruków do formatu elektronicznego. Przy wielotysięcznych transakcjach może okazać się to wręcz niemożliwe ze względu na pracochłonność takiej transformacji i w efekcie może zawrócić toczące się postępowanie do ponownego etapu uzyskania danych.

Jednym z poważniejszych problemów, na jaki napotykają analitycy kryminalni na etapie tworzenia elektronicznej bazy danych, jest różnorodność formatów danych dostarczanych do analizy. Na przykład billingi telefoniczne, mimo że zazwyczaj zawierają podobne dane, często różnią się rozkładem kolumn, reprezentacją lub formatem danych. Przykładowo w niektórych billingach kierunek połączenia telefonicznego zapisywany jest w osobnej kolumnie (wartości typu: „przychodzące", „wychodzące"), natomiast w innych, kierunek bazuje na kolejności występowania numerów telefonów w kolumnach. Co więcej formaty danych wejściowych ulegają nieustannym zmianom co uniemożliwia jednokrotne stworzenie zestawu reguł pozwalających na wczytanie danych pochodzących z konkretnych źródeł.

2.2.3. Analiza

Pozyskane dane zostają następnie poddane właściwej analizie. Ciekawą możliwością, którą może dostarczać oprogramowanie przeznaczone dla analityka, jest funkcja sklejania

kilku wykazów połączeń w jeden zbiór danych w ramach prowadzonego projektu – sprawy. Korzystając z tej funkcjonalności, możemy np. z trzech bilingów połączeń telefonicznych przedstawić jako część wspólną charakterystyczne połączenia dla wybranych numerów telefonicznych. Dodatkowo linie łączące poszczególne numery opatrzone mogą być wartościami ilości nawiązanych czy odebranych połączeń oraz możemy z tego poziomu dotrzeć do daty tych zdarzeń. Ta sama funkcjonalność doskonale sprawdza się podczas analizy przelewów bankowych. Istotną kwestią jest fakt że diagram jest odwzorowaniem celu jakiemu ma służyć. Niezastąpiony jest więc przy prezentowaniu pozyskanej wiedzy np. o przepływach na rachunkach bankowych. Dobrze wykonany diagram prezentujący zakres informacji o historii rachunków bankowych dodatkowo zestawiony z diagramem przepływu towarów czy usług pomiędzy osobami lub firmami, do których konta należą, sprawią że stosy dokumentów przestaną być straszne i nic nie mówiące, a staną się czytelne i zrozumiałe. Zastosowanie kolorowych diagramów pozwala dostrzec mechanizmy, jakie wykorzystywali przestępcy.

Aplikacja wspierająca pracę analityka powinna pozwalać na przedstawienie uzyskanej wiedzy w wielu różnorodnych wariantach graficznej prezentacji, a dobór i ich użycie powinno zależeć od pomysłowości analityka i rzecz jasna oczekiwań zleceniodawcy. Przykładem może być próba odtworzenia fizycznej drogi przemieszczania się podejrzanych osób na podstawie uzyskanych bilingów rozmów telefonicznych. Takie informacje można uzyskać dzięki śledzeniu pozycji geograficznej BTS-u (stacji bazowej, przekaźnikowej sieci komórkowej), z której korzystała podejrzana osoba.

Jednym z ciekawszych diagramów jest diagram czasowy, który obrazuje między innymi połączenia telefoniczne w układzie chronologicznym. Połączenia przedstawione są jako poszczególne linie łączące wskazane numery telefonów, które zostały przedstawione na osi czasu. Diagram ten jest przydatny, gdy na oś czasu nałożymy wydarzenia, które są przedmiotem prowadzonej sprawy. Funkcje generowania tego rodzaju diagramów można wykorzystać również przy wykonywaniu analizy przepływów towarów, usług, czy środków finansowych. Uogólniając, diagram może przedstawiać dwa połączone ze sobą punkty, podsumowanie ilości połączeń oraz ich kierunek i to „coś”, co wędruje pomiędzy dwoma punktami, a może to być środek finansowy, informacja w postaci np. rozmowy telefonicznej, towar lub usługa.

2.2.4. Raport końcowy

Oprócz diagramów oraz wykresów sporządzanych dzięki odpowiedniemu oprogramowaniu, analityk jako końcowy efekt swojej pracy sporządza „raport” bądź też „analizę”. W niektórych przypadkach analiza kryminalna może przybrać swoją brzegową postać, w której analityk stara się za wszelką cenę zgromadzone informacje tekstowe zaprezentować w formie graficznej. Takie wymuszone podejście do prezentacji danych może doprowadzić do stanu, że w wyniku dalej będziemy mieli do czynienie z nieuporządkowaną informacją, dlatego oprócz funkcji programu niezbędna jest wiedza i doświadczenie analityka. Graficzna prezentacja (diagramy, wykresy, schematy) to dalej tylko materiał, który wyróżnia i pozwala lepiej zrozumieć i wysnuć prawidłowe wnioski, jakie zostaną przedstawione w sporządzonym przez analityka dokumencie raportu z przeprowadzonego procesu analizy kryminalnej. Na diagramach nie może znaleźć się nic poza tym, co zostało zawarte w raporcie, analizie. Niespójność raportu z diagramami jest to często spotykany błąd w wykonywanych analizach. Najczęstszym sposobem przedstawienia wyników procesu analizy to forma pisemna zwana

jako „raport" lub „analiza". Tak sporządzony dokument oczywiście może być dodatkowo przekazywany w formie interaktywnej prezentacji. Jedną z form przedstawienia wyników analizy jest omówienie jej wyników w formie ustnej. W przypadku gdy materiał ten ma stanowić kluczowy element dowodowy w prowadzonej sprawie, dużo lepszym sposobem będzie zwrócenie się o pomoc do analityka kryminalnego. Analityk kryminalny nie będzie miał trudności z przeprowadzeniem prezentacji, jak i w udokumentowaniu wyników swojej analizy, gdyż będzie mógł się posłużyć diagramami, jak również wydrukiem dodatkowych materiałów źródłowych (logów) z całego procesu, które pozwolą na potwierdzenie uzyskanych wniosków. Analityk kryminalny jako osoba bezstronna będzie mógł ocenić materiał obiektywnie, a ze względu na swoją pozycję i doświadczenie ma znacznie większe szanse na przeprowadzenie poprawnej analizy i tym samym do wykrycia ewentualnych niespójności, braków czy nawet błędów.

2.3. Funkcjonalności dostarczane przez środowisko LINK

Środowisko LINK stanowi kompleksowe rozwiązanie informatyczne dedykowane dla potrzeb wspierania pracy analityków kryminalnych. Podstawowa wersja systemu udostępnia zestaw narzędzi do integracji, wstępnego przetwarzania oraz wizualizacji danych pochodzących z różnych źródeł (m.in. billingów telefonicznych, zestawień operacji bankowych oraz innych transakcji elektronicznych). LINK stanowi także platformę integracji różnych metod (pół)automatycznej analizy danych, które mogą zostać dołączone do systemu w postaci niezależnie wytworzonych komponentów. W ten sposób uzyskujemy praktycznie nieograniczone możliwości rozszerzania funkcjonalności systemu i dostosowywania jego możliwości do potrzeb konkretnych jednostek prowadzących prace z zakresu analizy kryminalnej.

2.3.1. Wizualizacja danych na diagramach

Jedną z najważniejszych funkcjonalności dostarczanych użytkownikowi przez środowisko LINK należy graficzna wizualizacja danych w postaci różnorodnych diagramów. Dane mogą być prezentowane w postaci grafów ukazujących strukturę ich zależności, gdzie węzły reprezentują obiekty (np. numery telefoniczne lub rachunki bankowe), a połączenia między nimi symbolizują powiązania (np. połączenia telefoniczne lub transakcje bankowe). Taki sposób wizualizacji podkreśla zależności występujące pomiędzy obiektami będącymi przedmiotem analizy (np. pokazując sposób, w jaki komunikują się podejrzane osoby), co upraszcza interpretację danych oraz pozwala na stawianie i weryfikację hipotez.

W celu zwiększenia przejrzystości i czytelności diagramów, ich elementy rysowane są za pomocą ikon odzwierciedlających znaczenie (np. ikony przedstawiające człowieka, telefon, itd.). Automatyczne porządkowanie elementów według wybranego algorytmu rozkładu, może również znacząco poprawić jakość prezentacji danych, a także w znacznym stopniu ułatwić ich dalszą interpretację. W zależności od struktury zależności danych, przydatne mogą się okazać różne rozkłady, stąd szeroka paleta możliwości dostarczanych w tym względzie przez aplikację:

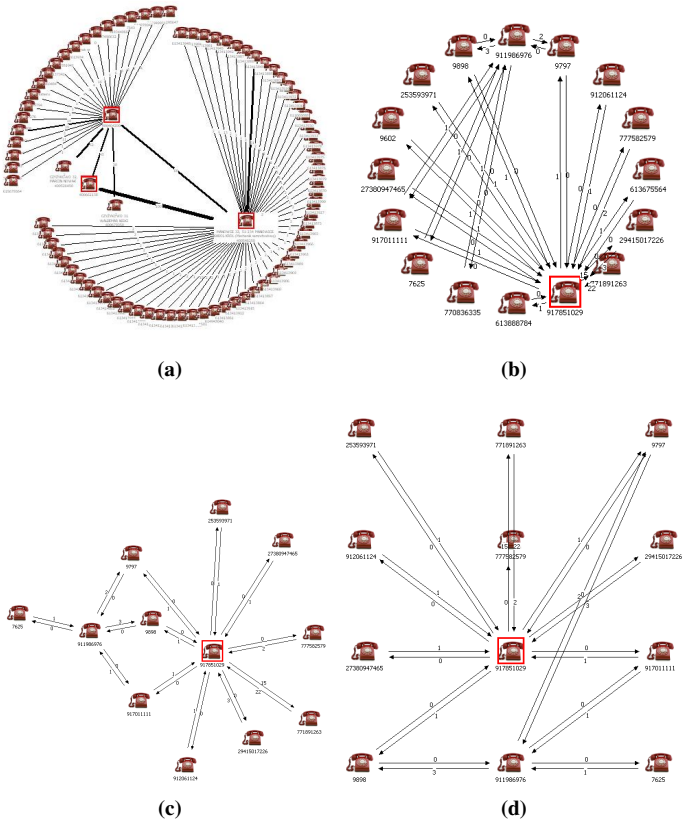
– rozkład grupowy (rys. 2.2a) – prezentuje węzły główne na kole, a pozostałe grupy dookoła nich na zewnątrz koła,

– rozkład kołowy (rys. 2.2b) – służy rozmieszczeniu wszystkich elementów na kole, przy czym główne elementy są wierzchołkami wielokąta foremnego wpisanego w to koło,

– rozkład „pawi ogon” (rys. 2.2c) – służy rozmieszczeniu węzłów głównych w centrum, a elementów z nimi połączonych dookoła danego elementu głównego – tak że w efekcie przypomina to pawi ogon,

– rozkład siatkowy (rys. 2.2d) – rozkłada węzły na równomiernej siatce,

– rozkład losowy – rozmieszcza węzły na diagramie w losowo wybranych miejscach.



Rys. 2.2. Rozkłady elementów na diagramie schematycznym – rozkład grupowy (a), rozkład kołowy (b), rozkład „pawi ogon” (c), rozkład siatkowy (d)

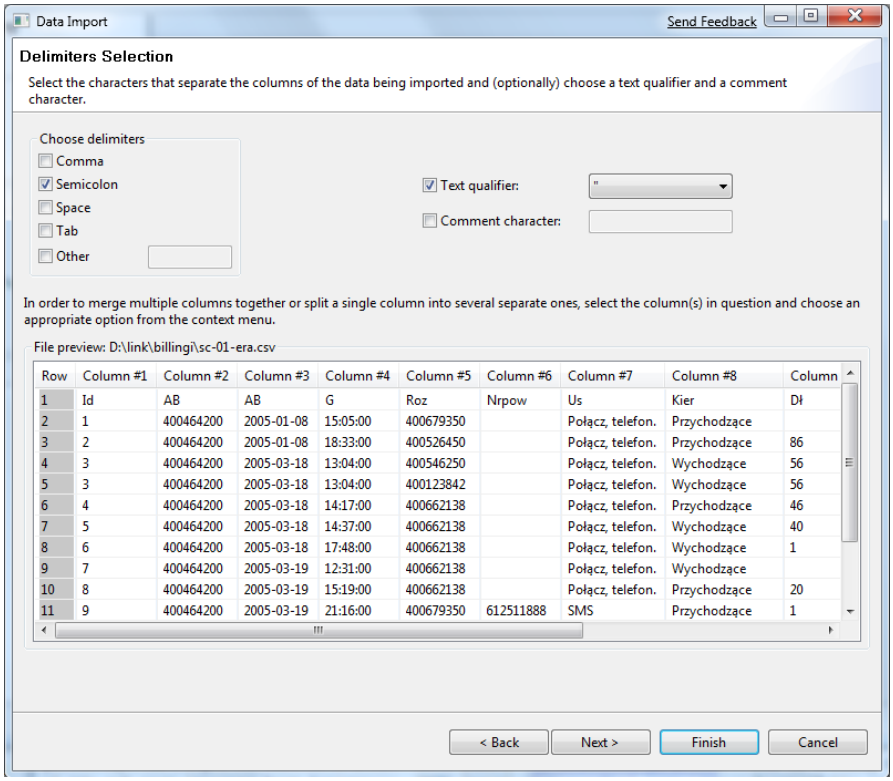
Środowisko LINK wyposażono także w edytor, pozwalający na „ręczne” tworzenie i modyfikację diagramów. Edytor dostarcza zestawu elementów, które mogą zostać dodane do

diagramu poprzez „upuszczenie” na wybrany fragment diagramu. Pozwala to na uzupełnienie zależności pozyskanych z analizowanych danych o informacje dostępne z innych źródeł. Edytor umożliwia również nadawanie ról elementom przez dodanie wyróżnienia, a także zmianę etykiet, co pozwala na dodawanie dodatkowych informacji o poszczególnych elementach diagramów. Zestaw narzędzi do wyrównywania elementów, zmiany skali diagramu, a także podglądu całego diagramu w znaczny sposób upraszcza pracę diagramami.

Integracja wyników wizualizacji z innymi aplikacjami (np. w celu stworzenia prezentacji dotyczącej materiału dowodowego) możliwa jest dzięki funkcjonalności zapisu diagramów w postaci plików graficznych. Dostępne są różne popularne formaty rastrowe (takie jak JPEG, PNG, itp.), które mogą być wykorzystane np. w pakietach biurowych ogólnego przeznaczenia.

2.3.2. Pozyskiwanie i przetwarzanie danych

Środowisko LINK zawiera rozbudowany kreator importu danych (rys. 2.3), za pomocą którego w intuicyjny sposób można wczytać do systemu różnego typu dane zapisane w plikach o różnych formatach.



Rys. 2.3. Kreator importu danych

Kreator importu dokonuje walidacji danych, dzięki czemu automatycznie wykrywana jest część błędnych danych (np. brakujący numer telefonu, czy nieprawidłowy numer konta). Podczas importu użytkownik widzi podgląd danych oraz ma możliwość wykonania niezbędnych modyfikacji (np. poprawienie formatu daty, ustawienie domyślnych wartości, itp.). Wszystkie modyfikacje są zapisywane i mogą zostać wyświetlone po zakończonym imporcie, dzięki czemu użytkownik ma pełną kontrolę nad danymi i ich pochodzeniem.

Jedną z głównych zalet zrealizowanego mechanizmu importu danych jest jego elastyczność. Oparty został on o reguły odpowiedzialne za konwersję i interpretację pojedynczych fragmentów danych (takich jak np. źródłowy numer telefonu, data rozpoczęcia i zakończenia połączenia), co pozwala na każdorazowe dostosowanie procesu importu do formatu wejściowego. Z drugiej strony, dzięki zastosowaniu szablonów importu, środowisko LINK umożliwia łatwe i szybkie importowanie danych w takim samym formacie, co znacznie skraca proces wczytywania danych. Szablony, zapisane w formacie XML, mogą być również przenoszone pomiędzy instalacjami aplikacji, co stwarza możliwość łatwej wymiany szablonów miedzy analitykami (np. za pomocą poczty elektronicznej lub stron WWW).

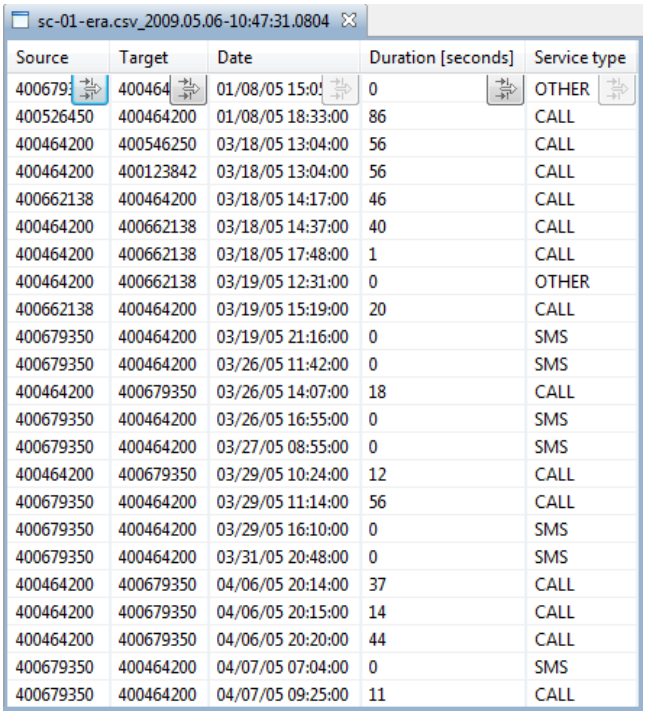
Zestaw dedykowanych narzędzi analitycznych wbudowanych w podstawową wersję systemu LINK umożliwia wyekstrahowanie z wielkiej ilości danych informacji najważniejszych z punktu widzenia analizy. Filtrowanie danych według typu, zakresu wartości liczbowych, przedziału czasowego lub według wzorców, pozwala na ukrycie danych nieistotnych (rys. 2.4). Użytkownik ma również możliwość uzyskania raportów z przetwarzanych danych w postaci statystyk, które ukazują ich najważniejsze cechy, jak np. liczba zależności danego typu, liczba powiązań pomiędzy wybranymi elementami, itp. Statystyki dostarczają ogólnego spojrzenia na duże zbiory danych, które zostały wstępnie przetworzone, co pozwala np. na wykrycie numerów telefonów, które kontaktowały się ze sobą najczęściej.

2.4. System LINK w analizie bilingów telefonicznych

Aktualna wersja dostosowana jest głównie do potrzeb analizy danych dostarczanych przez operatorów sieci telefonicznych. Nowa sprawa analizy bilingów telefonicznych rozpoczyna się od utworzenia w programie LINK indywidualnego projektu, który pełni rolę wirtualnego katalogu, skoroszytu. Każdy projekt posiada swoją nazwę pozwalającą na jego identyfikacje, dla lepszego zobrazowania możemy nazwać przykładowy projekt jako „Sprawa Kowalskiego”. W ramach projektu zapisywane są dane pochodzące z bilingów telefonicznych, powstałe diagramy i dodatkowe informacje (np. pliki aktywności) pojawiające się podczas pracy nad „Sprawą Kowalskiego”.

2.4.1. Podstawowy schemat pracy z aplikacją

Każda z prowadzonych spraw opiera się na informacjach zawartych w bilingach telefonicznych otrzymanych od operatora telefonii np. firmy Era, Orange, Plus, Play. Aby móc dokonać analizy otrzymanego materiału przy użyciu środowiska LINK, należy dokonać importu danych. Mapowanie pomiędzy danymi źródłowymi a modelem danych w systemie LINK dokonywane jest na podstawie utworzonego szablonu. Importowany materiał poddawany jest procedurze walidacji, podczas której wskazywane są wszelkie nieprawidłowości.



sc-01-era.csv_2009.05.06-10:47:31.0804

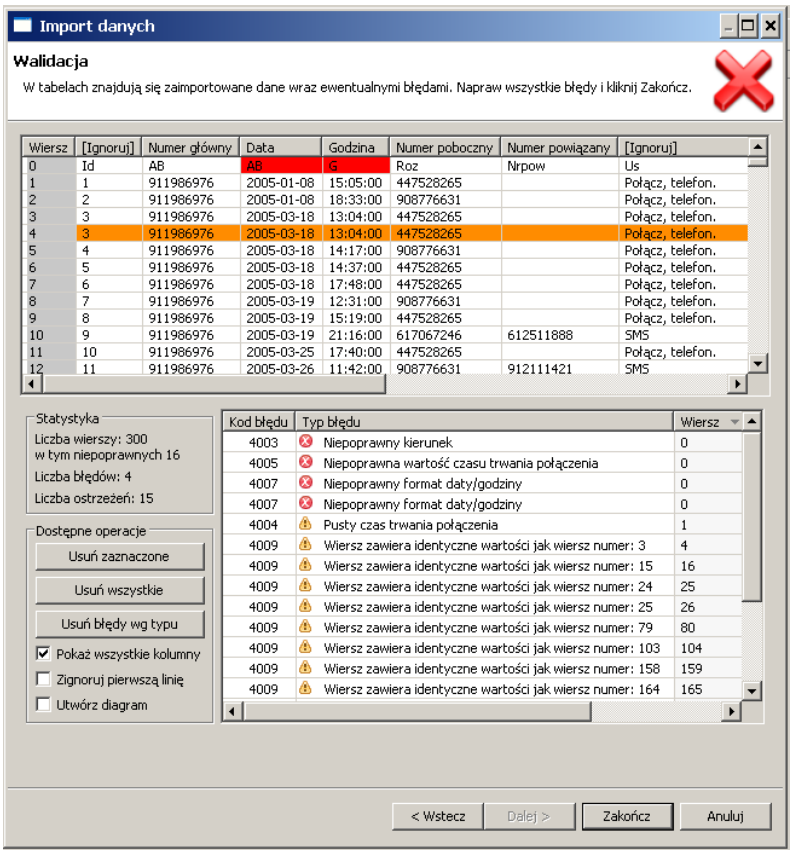
Source	Target	Date	Duration [seconds]	Service type
400679	400464	01/08/05 15:0	0	OTHER
400526450	400464200	01/08/05 18:33:00	86	CALL
400464200	400546250	03/18/05 13:04:00	56	CALL
400464200	400123842	03/18/05 13:04:00	56	CALL
400662138	400464200	03/18/05 14:17:00	46	CALL
400464200	400662138	03/18/05 14:37:00	40	CALL
400464200	400662138	03/18/05 17:48:00	1	CALL
400464200	400662138	03/19/05 12:31:00	0	OTHER
400662138	400464200	03/19/05 15:19:00	20	CALL
400679350	400464200	03/19/05 21:16:00	0	SMS
400679350	400464200	03/26/05 11:42:00	0	SMS
400464200	400679350	03/26/05 14:07:00	18	CALL
400679350	400464200	03/26/05 16:55:00	0	SMS
400679350	400464200	03/27/05 08:55:00	0	SMS
400464200	400679350	03/29/05 10:24:00	12	CALL
400679350	400464200	03/29/05 11:14:00	56	CALL
400679350	400464200	03/29/05 16:10:00	0	SMS
400679350	400464200	03/31/05 20:48:00	0	SMS
400464200	400679350	04/06/05 20:14:00	37	CALL
400464200	400679350	04/06/05 20:15:00	14	CALL
400464200	400679350	04/06/05 20:20:00	44	CALL
400679350	400464200	04/07/05 07:04:00	0	SMS
400679350	400464200	04/07/05 09:25:00	11	CALL

Rys. 2.4. Tabelaryczny widok danych zapewniający ich zaawansowane filtrowanie

Przykładowo sprawdzane są typy zgodności danych źródłowych z oczekiwanymi przez system. Kolorami wyróżniane są braki w rekordach, a w oddzielnej tabeli znajduje się lista nieprawidłowości wraz z opisem (rys. 2.5).

W trakcie importu analityk może wprowadzić pewne korekty lub usunąć błędne rekordy. Po poprawnym zakończeniu procesu importu, dane z bilingu trafią do katalogu prowadzonej „Sprawy Kowalskiego", na podstawie których uaktywnia się możliwość utworzenia diagramów rozkładu połączeń telefonicznych. Diagram prezentuje graficzny obraz wygenerowany na podstawie zaimportowanych tekstowych danych bilingów.

Aby utworzyć nowy diagram należy z menu głównego lub kontekstowego wybrać pozycję z nazwą odpowiedniego diagramu. W zależności od sposobu prezentacji danych wśród możliwych typów diagramów wyróżnić można trzy główne grupy: schematyczne, hierarchiczne i czasowe. Diagram schematyczny dostępny w aktualnej wersji systemu LINK (przykładowe diagramy tego typu zostały omówione w rozdziale 2.3), pokazuje połączenia oraz ich liczbę pomiędzy numerami telefonów. Połączenia pomiędzy abonentami mogą być prezentowane jako podwójne, wtedy połączenia przychodzące oraz wychodzące traktowane są oddzielnie, można także agregować wszystkie połączenia pomiędzy węzłami. Innym rodzajem połączeń są połączenia zmiennej grubości, gdzie grubość linii uzależniona jest od ilości połączeń, co w sposób obrazowy pozwala określić intensywność kontaktów pomiędzy rozmówcami.



Rys. 2.5. Walidacja, weryfikacja danych

Przy pomocy diagramu schematycznego możemy dostrzec relacje, jakie zachodzą pomiędzy abonentami w dostarczonych bilingach. Różne rozkłady węzłów pozwalają szybko dostrzec abonentów, do których schodzą się połączenia lub odkryć zależności pomiędzy podejrzanymi osobami w prowadzonym dochodzeniu. Istnieje również funkcja nanoszenia poprawek na diagramach. Możemy usuwać zbyteczne numery abonentów, które np. wykluczamy z prowadzonej sprawy, gdyż są to numery infolinii, dodawać nowe, łączyć je z innymi diagramami. Dodatkowo edycji poddawane są właściwości – atrybuty węzłów.

2.4.2. Funkcje zaawansowane

W celu zidentyfikowania zależności zdarzeń oraz powiązań podejrzanych występujących w zgromadzonych danych, analitycy kryminalni niejednokrotnie muszą przeanalizować duże ilości danych. Często sprowadza się to do znalezienia tej samej informacji w danych pochodzących z różnych źródeł lub wręcz wykrycia, że dwie pozornie różne informacje są

tożsame, co poprzez oglądanie diagramów może być trudne lub wręcz niemożliwe. System LINK wspiera analityków w tym obszarze poprzez dostarczenie narzędzi służących do automatyzacji ich typowych działań. Pozwala im to skoncentrować się na charakterystycznych aspektach prowadzonej analizy znacznie redukując poświęcony jej czas.

W przypadku analizy bilingów rozmów telefonicznych środowisko LINK jest w stanie szybko przedstawić zestawienia numerów telefonów, między którymi występowały połączenia telefoniczne określonego typu – głosowe lub tekstowe/multimedialne (SMS, MMS). Na podstawie zgromadzonych danych, można także sprawdzić jakie numery abonenckie oraz numery IMEI (reprezentujące aparaty telefoniczne) współpracowały ze sobą, co pozwala na wykrycie sytuacji, gdy podejrzany korzysta z różnych kart SIM do połączeń z określonymi osobami. Możliwe jest również powiązanie danych z billingów z danymi z książek adresowych pochodzących z pozyskanych aparatów telefonicznych osób zamieszanych w sprawę, co w rezultacie może doprowadzić do identyfikacji użytkownika danego numeru lub wykazać zbieżność nazw kontaktów dla różnych numerów telefonicznych. W danych dostarczonych przez operatorów komórkowych mogą znajdować się informacje o stacjach bazowych (BTS), które były użyte podczas połączenia – na tej podstawie analityk może określić miejsce przebywania podejrzanego w danej chwili lub w określonych porach, a także określić jego trasę przemieszczania się. W billingach stacje mogą być określone nic nie znaczącymi (z punktu widzenia analizy) identyfikatorami, jednak system LINK umożliwia automatyczne połączenie tych identyfikatorów z istniejącą bazą informacji o stacjach.

Istotnym elementem systemu LINK jest wyszukiwanie fragmentu tekstu w atrybutach abonentów. Efektem tego działania jest wyróżnienie abonentów spełniających dane warunki wyszukiwania. Dodatkowo do wyszukiwania możemy zaznaczyć opcje:

- dopasuj całe słowa – wyszukiwanie elementów tożsamych z wprowadzonym tekstem,
- uwzględnij wielkość liter – wielkość liter jest rozróżniana,
- według wzorca – w polu tekstowym możemy używać dodatkowych znaków „*” i „?”, które oznaczają odpowiednio dowolna ilość znaków i dowolny znak pojedynczy.

Funkcja wyszukiwania może zostać użyta w odszukiwaniu np. abonentów po unikalnym numerze telefonu IMEI, numerze nadajnika, aby odnaleźć osoby korzystających z tego samego BTS-u.

LINK pozwala na eksportowanie przygotowanego diagramu do pliku graficznego, aby wykorzystać go do raportu ze sporządzanej analizy bilingu. Poprzez panel „Projekty” zaznaczamy ten diagram, który chcemy wyeksportować, możemy również dokonać eksportu fragmentu bieżącego diagramu. Przed zapisem pliku z eksportem pojawia się podgląd zapisywanego obrazu. Tak przygotowany materiał może być traktowany jako jeden z dowodów w prowadzonej sprawie analizy kryminalnej.

2.4.3. Scenariusz przykładowego użycia

Załóżmy, że analityk, nazwijmy go Policjant Kowalski, prowadzący sprawę kradzieży pojazdów luksusowych, przeglądając internetowe ogłoszenia sprzedaży pojazdów zwrócił uwagę na ofertę sprzedaży samochodu, który najprawdopodobniej został skradziony. Jak

wynikało z ogłoszenia, sprzedający kontaktował się z kupującymi poprzez telefon komórkowy o numerze 400464200. Od operatora sieci komórkowej pozyskano wykaz połączeń dla numeru 400464200. Dokonano importu danych do LINK i wykreślono linie połączeń. Jak wynikało ze statystyki połączeń powyższy numer kontaktował się wielokrotnie z 5 numerami:

– 400679350,

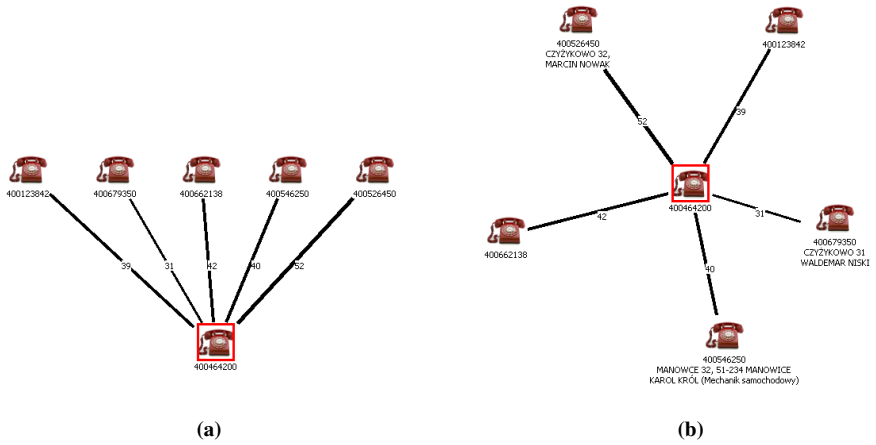
– 400526450,

– 400546250,

– 400123842,

– 400662138.

Oprócz tego było wiele pojedynczych połączeń przychodzących (rys. 2.6a). Dla w/w numerów zostały pobrane od operatorów dane abonenckie (rys. 2.6b).

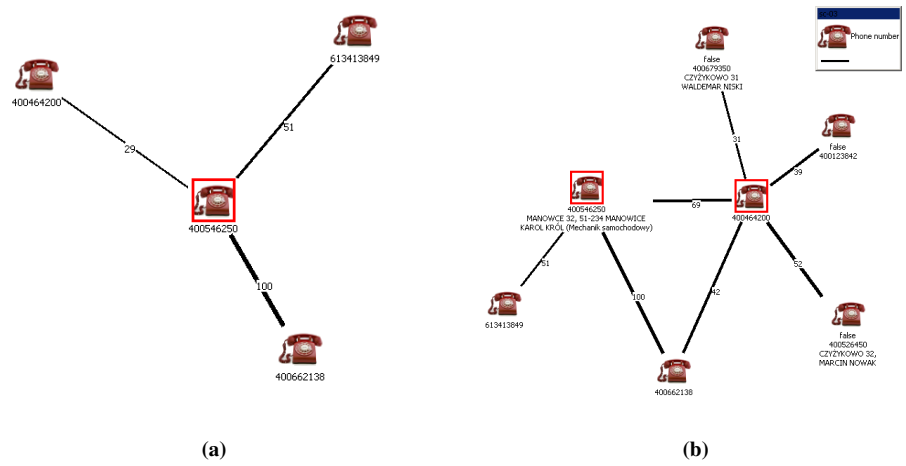


(a) **(b)**

Rys. 2.6. Wykaz połączeń dla numeru 400464200 (a), wykaz połączeń dla numerów łączących się z numerem 400464200 (b)

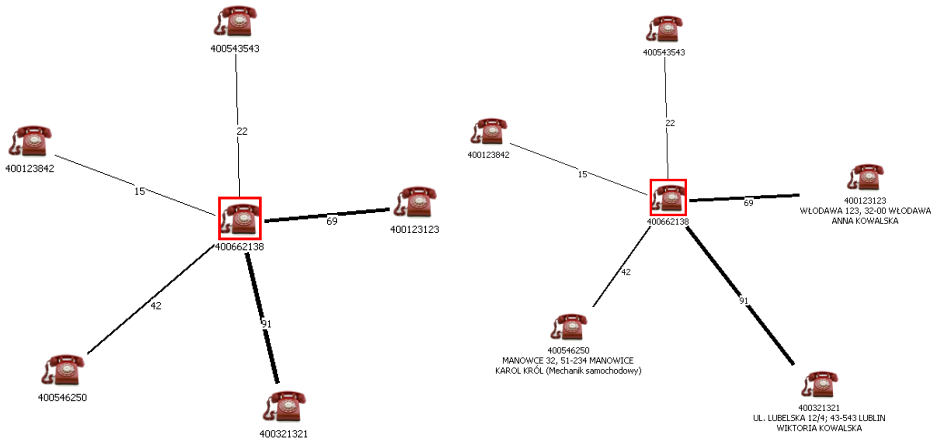
Po ustaleniu abonentów tych numerów okazało się, że jednym z podejrzanych (numer 400526250) jest znany policji mechanik wcześniej karany za udział w kradzieżach pojazdów. Pozyskano wykaz połączeń tego numeru. Po zaimportowaniu danych i wykreśleniu grafu określono najczęściej wybierane numery (rys. 2.7a).

Po połączeniu z poprzednim diagramem (rys. 2.7b) okazało się, że powyższe numery (400464200 oraz 400526250) kilkakrotnie kontaktowały się z numerem 400662138. Pozyskano od operatora wykaz połączeń nr 400662138 i zaimportowano dane do LINK.



(a) **(b)**

Rys. 2.7. Najczęściej wybierane numery (a), widok po połączeniu z diagramem przedstawionym na rys. 2.6b (b)



(a) **(b)**

Rys. 2.8. Najczęstsze kontakty badanego numeru (a), identyfikacja poszukiwanego przestępcy (b)

Wykreślono diagram (rys. 2.8a) i na podstawie statystyki danych zawartych w wykazie połączeń ustalono do kogo rozmówca wysyła SMS-y i z kim najczęściej rozmawia (numery 400123123 i 400321321).

Pozyskano dane właścicieli tych numerów i zaimportowano je do LINK (rys. 2.8b). Na podstawie tych danych oraz innych informacji ustalono, że osoba posługująca się numerem telefonu 400662138 jest złodziejem samochodów, od dłuższego czasu poszukiwanym przez policję.

Użycie programu LINK podczas analizy bilingów pozwoliło Policjantowi Kowalskiemu w szybki i trafny sposób dotrzeć do podejrzanego złodzieja samochodów. Policjant zaoszczędził sporo żmudnej i podatnej na pomyłki pracy związanej z przeglądaniem, przeszukiwaniem wielostronicowych wykazów połączeń oraz wykonywania obliczeń statystycznych. Graficzna wizualizacja bilingu na różnorodnych diagramach umożliwiła Policjantowi natychmiastowe podjęcie dalszych kroków postępowania, oraz na trafną ocenę sytuacji i bezbłędne wskazanie zależności pomiędzy numerami telefonów podejrzanych osób.

2.5. Wybrane aspekty implementacji środowiska LINK

Biorąc pod uwagę różnorodność przewidywanych wariantów zastosowań, rozwiązania technologiczne środowiska LINK zostały dobrane pod kątem możliwie największej przenośności, otwartości i elastyczności, w szczególności w zakresie dodawania nowych funkcjonalności. Służy temu wykorzystanie języka Java, modularna architektura oparta o *Eclipse RCP*, oraz zunifikowany interfejs zarządzania danymi o różnej strukturze i znaczeniu.

2.5.1. Architektura środowiska LINK

Rdzeń systemu jest oparty o wtyczkową architekturę Eclipse (ang. *Eclipse Plugin Architecture*) bazującą na koncepcji rozszerzeń (ang. *Extension*) i punktów rozszerzeń (ang. *Extension Point*). Rozszerzenia i punkty rozszerzeń stanowią mechanizm pozwalający uniknąć ścisłych powiązań pomiędzy poszczególnymi komponentami w systemie, co ułatwia modyfikowanie i rozszerzanie funkcjonalności aplikacji. *Extension* jest modułem, który rozszerza system, natomiast *Extension Point* jest miejscem połączenia rozszerzenia z pozostałą jego częścią. Dzięki temu, że konfiguracja zależności między modułami realizowana jest poza kodem źródłowym, można dodawać nowe funkcje bez ingerencji w istniejący kod aplikacji.

Rysunek 2.9 przedstawia strukturę głównych komponentów systemu LINK. Jako baza systemu (na dole diagramu) znajdują się komponenty należące do platformy *Eclipse RCP* (ang. *Rich Client Platform*) [5], które są odpowiedzialne za zarządzanie aplikacją oraz dostarczają takich funkcji, jak: pomoc, automatyczne aktualizacje czy zarządzanie sprawami. Ponadto moduły te, wspierają opisany wcześniej mechanizm rozszerzeń.

Do wizualizacji danych na diagramach LINK wykorzystuje bibliotekę *Graphical Editing Framework (GEF)* [1] dostarczoną jako dodatek do środowiska Eclipse RCP. W górnej części diagramu znajdują się moduły LINK, których rolą jest:

– *Core* – rdzeń systemu adaptujący technologię Eclipse RCP do potrzeb systemu,

– *Data Model* – definiuje modele danych (opisane poniżej) wykorzystywane w systemie,

– *Data Provider* – moduł zapisu danych, może być rozszerzany w celu dostosowania zapisu danych do wybranych potrzeb,

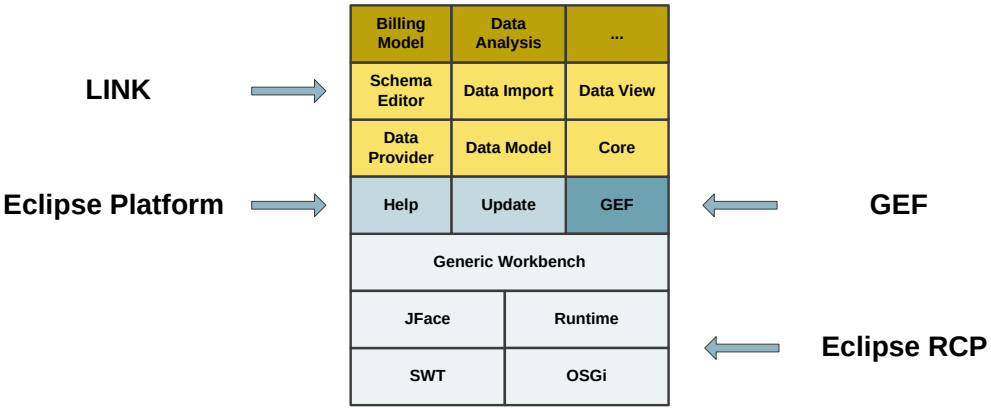
– *Data Import* – moduł służący do importu danych pochodzących z różnych źródeł,

– *Data View* – definiuje tabelaryczny widok danych wraz z zestawem zaawansowanych filtrów,

– *Schema Editor* – definiuje edytor diagramów schematycznych, przedstawiający dane w postaci grafów zależności,

– *Data Analysis* – reprezentuje wszystkie moduły wspomagające analizę danych; może to być np. wyszukiwanie ścieżek pomiędzy dwoma obiektami w grafie, wyszukiwanie zależności pomiędzy numerami IMEI i numerami telefonów, itp.

– *Billing Model* – definiuje model opisujący dziedzinę billingów telefonicznych oraz rozszerzenia importera danych, pozwalające na import billingów.

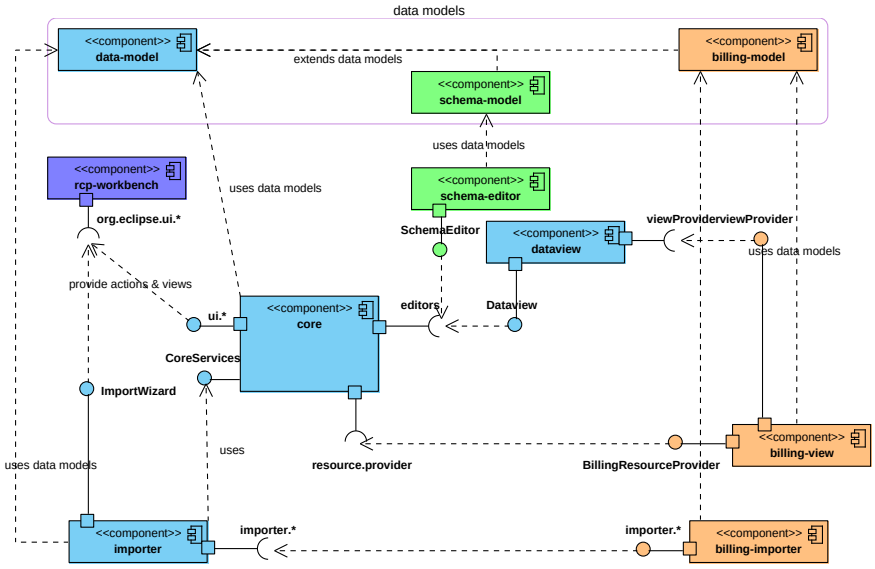


Rys. 2.9. Architektura systemu LINK

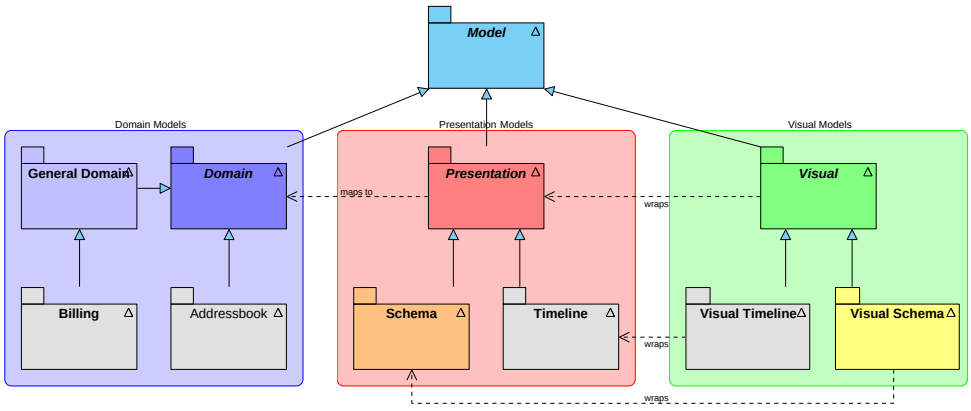
Diagram 2.10 przedstawia zależności implementacyjne pomiędzy istniejącymi komponentami środowiska LINK. Istnieje możliwość rozszerzania środowiska o kolejne moduły, przy pomocy których można zdefiniować zupełnie nowe metody wizualizacji danych (np. edytor czasowy), nowe metody analizy danych czy też dziedziny analiz (np. analiza przepływów finansowych czy plików aktywności wybranych usług).

2.5.2. Model danych

Dane w systemie LINK identyfikowane i przechowywane są w *zbiorach* (ang. *data sets*) zdefiniowanych przez modele danych. Na rysunku 2.11 zaprezentowano hierarchię modeli, które zostały zdefiniowane wewnątrz rdzenia platformy LINK.



Rys. 2.10. Zależności pomiędzy komponentami systemu LINK



Rys. 2.11. Modele danych systemu LINK

Wyodrębniono 3 rozłączne grupy modeli, odgrywające specyficzne role w systemie.

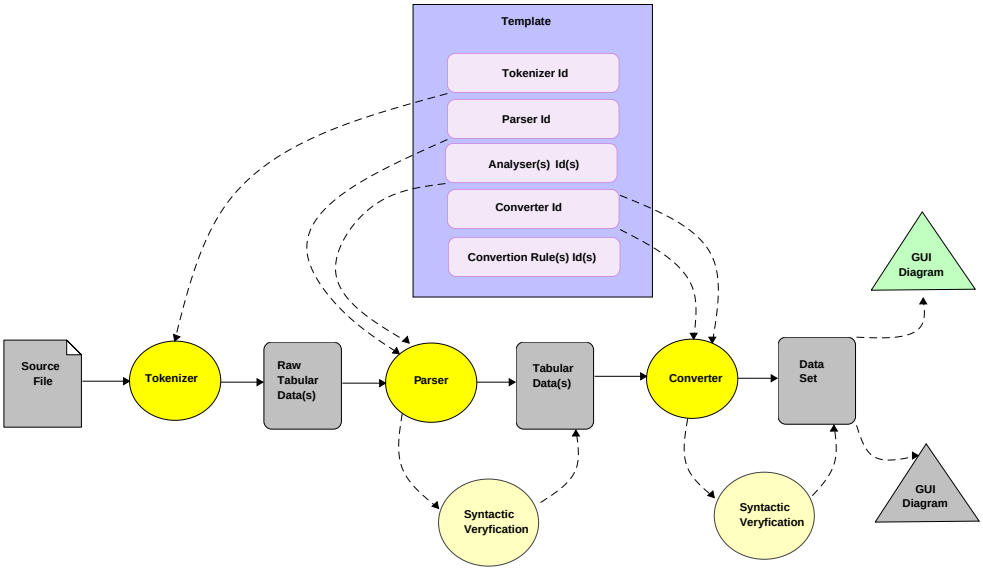
Modele dziedzinowe – (*Domain Models*) służą do reprezentowania danych podlegających analizie, z uwzględnieniem dziedziny życia, z której pochodzą. Przykładowo są to model billingów telefonicznych lub model książki adresowej.

Modele prezentacyjne – (*Presentation Models*) służą do przedstawiania danych (na ogół pochodzących z modeli dziedzinowych) w formie dostosowanej do specyficznego narzędzia lub formy analizy. Przykładowo jest to model schematyczny, który zawiera dane potrzebne do wyświetlenia diagramu schematycznego.

Modele wizualizacyjne – (*Visual Models*) służą do opisu sposobu wizualizacji konkretnego modelu prezentacyjnego. Modele wizualizacyjne definiują dane typu: położenie węzła, kolor połączenie, itp.

2.5.3. Import danych

Importowanie danych do wspólnego formatu jest jednym z kluczowych elementów systemu LINK. Dane źródłowe mogą być plikami tekstowymi (*.txt, *.ack) lub arkuszami danych (*.csv, *.xls). Schemat ideowy przepływu danych[3] importowanych do systemu został przedstawiony na rys. 2.12.



Rys. 2.12. Schemat przepływu danych podczas importu

Dekompozycja procesu importu pozwala uniezależnić import od konkretnych modeli danych, formatów plików oraz sytuacji szczególnych występujących w danych dostarczanych z zewnętrznych źródeł. Proces importu składa się z następujących etapów:

[3] Dokumentacja systemu CAST, http://caribou.iisg.agh.edu.pl/proj/cast/.

1. Etap wczytania danych: dane wejściowe wczytywane są z pliku wejściowego (*Source file*) zapisanego w jednym z obsługiwanych formatów za pomocą *Tokenizera*.

2. Etap sprawdzania poprawności danych dokonywany przez *Parser*: sprawdzane są wartości umieszczone w pojedynczych komórkach (np. data połączenia, numer telefonu, czas trwania połączenia telefonicznego, itp.). Należy zwrócić uwagę, że etap weryfikacji jest bardzo istotny, gdyż często podczas importu pojawia się konieczność poprawy danych, które nie odpowiadają ustalonym szablonom. Aplikacja LINK ma wbudowane wykrywanie takich sytuacji i umożliwia korektę bądź opuszczenie niepoprawnych rekordów, aby nie zaburzały dalszego procesu analizy. Wszystkie zmiany danych wejściowych są zapisywane, dzięki czemu analityk zachowuje pełną kontrolę nad danymi.

3. Etap konwersji wykonywany przez *Converter*: zweryfikowane dane są konwertowane do zbiorów danych – wewnętrznego formatu LINK. Wykrywane są również niepoprawne zależności pomiędzy danymi, np. połączenie bez numeru telefonu, itp.

Wszystkie opisane elementy, tj. *Tokenizer*, *Parser* oraz *Converter*, są dostarczane do importera w postaci rozszerzeń, dzięki czemu wzbogacenie importera o interpretację nowych modeli danych, odczyt innych formatów plików, bardziej zaawansowane techniki weryfikacja danych, odbywa się przez dostarczenie i zarejestrowanie nowych komponentów. Proces aktualizacji nie wymaga żadnych zmian w kodzie źródłowym aplikacji i może odbywać się za pomocą wbudowanego mechanizmu aktualizacji.

W celu uproszczenia i przyspieszenia importu podobnych plików źródłowych (np. billingów telefonicznych pochodzących od tego samego operatora), wprowadzone zostały tzw. szablony importu (*templates*). Szablony te zawierają pełną informacje o procesie importu, dzięki czemu import kolejnych danych pochodzących z tego samego źródła można wykonać za pomocą kilku kliknięć.

Pojedynczy szablon opisuje sposób importu danych z jednego typu pliku do jednego modelu dziedzinowego. Każdy szablon identyfikowany jest po nazwie, dodatkowo posiada opis, który zawiera informację o zastosowaniu szablonu. Szablony pogrupowane są w grupy o podobnym zastosowaniu (np. szablony dotyczące billingów). Ponadto eksport i import szablonów do plików XML pozwala na prostą wymianę szablonów importu pomiędzy użytkownikami systemu.

2.5.4. Model wizualizacyjny diagramów schematycznych

Warstwa wizualizacji oparta jest na technologii GEF (ang. *Graphical Editing Framework*). GEF stanowi zrąb (ang. *framework*) wykorzystujący wzorzec Model-Widok-Kontroler, służący do budowy graficznego edytorów zintegrowanych z technologią Eclipse RCP. Rysunek 2.13 przedstawia najważniejsze elementy łączące technologie GEF oraz Eclipse RCP.

GEF dostarcza szeroki wachlarz funkcjonalności, które zwiększają ergonomię pracy ze środowiskiem. Do najważniejszych możemy zaliczyć:

– przesuwanie, tworzenie, łączenie elementów,

– kasowanie, operacje Cofnij/Ponów, bezpośrednia edycja elementów (tzw. edycja *inline*),

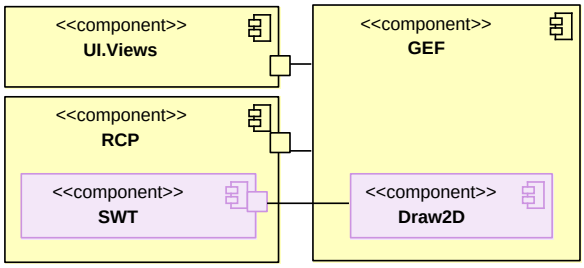
– podgląd diagramu w postaci miniatury,

– skalowanie diagramu,

– konfigurowalną paletę,

– dostęp do funkcji poprzez skróty klawiaturowe.

W środowisku LINK podstawą (modelem) diagramu jest prezentacyjny zbiór danych, zawierający elementy do wyświetlenia w edytorze. Prezentacyjny zbiór danych jest dekorowany przez odpowiedni zbiór danych wizualizacyjnych opisujący wizualne właściwości elementów, tj. położenie, kolor linii, itp. Na podstawie tych zbiorów tworzony jest widok i kontroler (elementy GEFa), które są odpowiedzialne za wyświetlenie diagramu oraz zarządzanie interakcją pomiędzy użytkownikiem a edytorem.



Rys. 2.13. Model GEF w połączeniu z RCP [2]

Każdy model prezentacyjny i wizualizacyjny odpowiada jednemu rodzajowi edytora. Obecnie zdefiniowany jest model schematyczny (prezentacyjny i wizualizacyjny) opisujący dane dla diagramów schematycznych. W przyszłości planuje się rozszerzenie modeli o kolejne rodzaje dla innych edytorów, takich jak np. edytor czasowy.

Aby wyświetlić zaimportowane dane w wybranym edytorze, należy skonwertować dane ze zbioru danych dziedzinowych (np. billingów) do prezentacyjnego zbioru danych. W tym celu został stworzony mechanizm konwerterów, który pozwala na wykonanie odpowiedniej translacji jednego zbioru danych w drugi.

W tym miejscu należy zwrócić uwagę na siłę takiego rozwiązania. Dane dziedzinowe mogą być zwizualizowane na diagramie na wiele różnych sposobów (np. w przypadku billingów telefonicznych mamy: połączenia pomiędzy numerami telefonów, numery telefonów połączone z numerami IMEI aparatów, z których były wykonywane połączenia). Każdy typ interpretacji danych jest realizowany przez pojedynczy konwerter. Zważywszy na fakt, że konwertery mogą być dostarczane do środowiska w postaci osobnych komponentów, umożliwia to bardzo proste rozszerzanie LINK o nowe możliwości interpretacji danych. Co więcej separacja kodu (programista piszący nowy konwerter skupia się tylko na sposobie interpretacji danych, pomijając szczegóły technologiczne) oraz zdalne aktualizacje, pozwalają na szybkie dodawanie nowych mechanizmów interpretacji danych.

2.5.5. Warstwa persystencji danych

Ważnym elementem środowiska LINK jest warstwa zapisu danych. Warstwa ta powinna zapewniać wydajne operacje na dużych ilościach danych o różnej strukturze (billingi telefoniczne, książka adresowa, przelewy bankowe, diagramy i inne). Ze względu na charakter pracy, dane dotyczące jednej sprawy powinny być odseparowane od innych. Ponadto warstwa persystencji powinna zapewniać również możliwość całkowitego wyczyszczenia danych sprawy oraz łatwej migracji danych pomiędzy różnymi instalacjami systemu LINK.

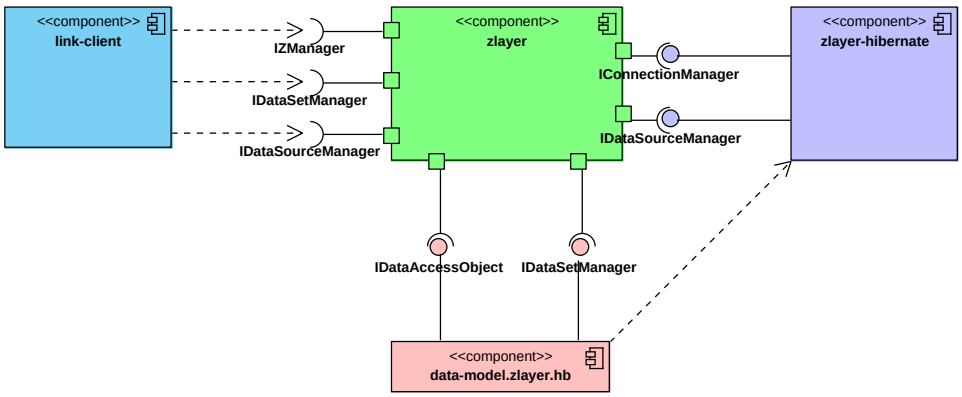
Dla potrzeb zapisu danych zastosowano dedykowane rozwiązanie oparte o plikową bazę danych SQLite z wykorzystaniem technologii Hibernate. Rysunek 2.14 przedstawia architekturę warstwy persystencji. Do jej głównych elementów należą:

– *zlayer* – kluczowy komponent warstwy zapisu danych (w skrócie WZ); zawiera wszystkie interfejsy zapewniające wymaganą funkcjonalność; część interfejsów jest realizowana bezpośrednio w tym komponencie, niektóre wymagają odrębnych komponentów dedykowanych do danego typu zbioru danych lub rodzaju źródła danych.

– *zlayer-hibernate* – komponent realizujący podstawowe interfejsy odpowiedzialne za połączenie ze źródłem danych, którym w obecnej implementacji jest baza danych SQLite przesłonięta warstwą abstrakcji zapewnianą przez framework Hibernate; zapewnia również bazowe klasy dla komponentów implementujących persystencję już dla konkretnych zbiorów danych w oparciu o tę technologię,

– *data-model.zlayer.hb* – komponent implementujący interfejsy niezbędne do uzyskania dostępu do wybranego typu zbioru danych.

– *link-client* – jest abstrakcją systemu LINK jako klienta warstwy persystencji.



Rys. 2.14. Architektura warstwy persystencji danych

Dzięki zastosowaniu wzorca Strategia, możliwe jest wykorzystanie różnych implementacji warstwy persystencji, w zależności od potrzeb przetwarzania danych.

2.6. Podsumowanie

W ramach niniejszego rozdziału, na tle krótkiej charakterystyki potrzeb związanych z praktyką analizy kryminalnej, została opisana koncepcja oraz wybrane funkcjonalności zintegrowanego środowiska wspomagania analizy kryminalnej LINK. Dzięki współpracy z Komendą Wojewódzką Policji w Krakowie możliwe było ustalenie szczegółowych wymagań dotyczących sposobu wykorzystania aplikacji i zrealizowanie w pełni funkcjonalnej wersji, która została pozytywnie zweryfikowana w praktyce.

Niewątpliwym atutem aplikacji jest łatwość obsługi, dzięki czemu nie wymaga ona specjalistycznego przygotowania użytkownika od strony informatycznej. Wychodząc na przeciw problemom związanym z różnorodnością danych dostarczanych do analizy, system LINK dostarcza dedykowany importer danych dostosowany do polskich realiów. Importer pozwala nie tylko na wczytanie i interpretacje zewnętrznych danych do formatu aplikacji, ale również umożliwia ich wstępną weryfikację, np. wykrycie pustych wartości, niepoprawnego formatu danych itp. Dzięki podglądowi danych oraz zapisywaniu zmian wprowadzanych podczas procesu importu, analityk ma pełną kontrolę nad wczytywaniem i interpretacją danych. Z drugiej strony zautomatyzowany proces wczytywania danych na podstawie szablonów znacząco przyspiesza pracę z danymi pozyskanymi z różnych źródeł. Zastosowanie komponentowej budowy importera, daje możliwość łatwego rozszerzania jego funkcjonalności, tak aby sprostać nowym przypadkom występującym w danych wejściowych.

Raz wczytane dane przechowywane są w systemie pomiędzy jego kolejnymi uruchomieniami, co pozwala na ich wielokrotne wykorzystanie np. do utworzenia wielu różnych diagramów. Graficzny interfejs użytkownika z bogatą paletą możliwości wizualizacyjnych pozwala na prezentacje danych na różne sposoby, umożliwiając przejrzyste zobrazowanie struktury ich zależności. Diagramy prezentujące wybrane zależności w danych źródłowych mogą podlegać dalszej edycji, a w ostatniej fazie analizy mogą zostać wyeksportowane w postaci obrazów do plików zewnętrznych i załączone jako materiał dowodowy do prowadzonych spraw.

Warto podkreślić, iż konstrukcja systemu pozwala na dołączanie nowych funkcjonalności, które mogą być dostarczone w postaci niezależnie wytworzonych komponentów. Pozwala to na rozszerzanie funkcjonalności aplikacji bez konieczności modyfikowania jej kodu źródłowego, co znacząco ułatwia jej rozwój w przyszłości. Odbywa się to za pomocą wbudowanego mechanizmu aktualizacji oprogramowania, które można pobrać zarówno ze stron aktualizacji umieszczonych w Internecie, jak również z plików dostarczonych na nośnikach typu CD, pamięć flash, itp.

System jest nieustannie rozwijany i w niedługim czasie spodziewać się należy wersji o nowych możliwościach prezentacji danych, w szczególności w zakresie wizualizacji na diagramach z uwzględnieniem zależności czasowych oraz filtrowania danych na podstawie zadanych kryteriów. W kolejnych rozdziałach opisano prototypowe realizacje modułów dostarczających różnorodnych inteligentnych metod analizy danych, pozwalających np. na (pół)automatyczne wykrywanie wzorców, klasyfikację danych, czy rozpoznawanie schematów wybranych typów przestępstw.

Bibliografia

[1] *Create an Eclipse-based application using the Graphical Editing Framework*, http://www.ibm.com/developerworks/opensource/library/os-gef/.

[2] *Eclipse: Tutorial 23 GEF In Depth*, http://www.eclipse.org.

[3] *Analiza Kryminalna cz. I*, zeszyt 24 serii Biblioteka Doskonalenia Zawodowego, Szczytno, 2002.

[4] W. Ignaczak, *Wybrane zagadnienia analizy kryminalnej*, Wyższa Szkoła Policji, Szczytno, 2005.

[5] L. Vogel, *Eclipse RCP – Tutorial (Eclipse 3.5)*, http://www.vogella.de/.