

*kpt. mgr inż. Adam KRASUSKI
prof. dr hab. inż. Tadeusz MACIAK
SGSP, Katedra Techniki Pożarniczej
Zakład Informatyki i Łączności*

WYKORZYSTANIE ROZPROSZONEJ BAZY DANYCH ORAZ WNIOSKOWANIA NA PODSTAWIE PRZYPADKÓW W PROCESACH DECYZYJNYCH PAŃSTWOWEJ STRAŻY POŻARNEJ

This article describes Decision Support System for Fire Service. The system is based on Directory Services and Case-Based Reasoning techniques. Article contains short review of Case-Based Reasoning methodology. Proposal and analyses of data layer for Decision Support System are also included.

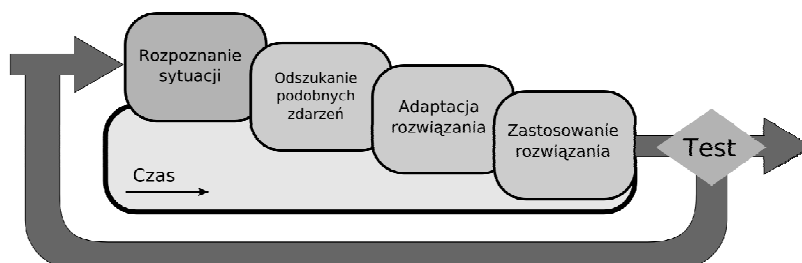
W artykule przedstawiono projekt systemu wspomagania decyzji dla strażaków PSP. System wykorzystuje katalogową bazę danych w architekturze rozproszonej do przechowywania opisu interwencji PSP. Podstawę dla wyszukiwania i porównywania przypadków zapisanych w bazie stanowi metoda CBR.

1. Wprowadzenie

Młodzi dowódcy kierujący akcją ratowniczą korzystają najczęściej z procedur, które przyswoili podczas szkolenia. Z biegiem czasu rośnie ich własne doświadczenie i zmienia się sposób podejmowania decyzji. Biorąc udział w akcjach ratowniczych, gromadzą własne doświadczenia odnośnie metod zwalczania zagrożeń. W przyszłości zdobyta wiedza staje się podstawą do wypracowywania metod likwidacji zagrożeń, w których biorą udział. Przenoszą skuteczne rozwiązania zastosowane w poprzednich akcjach na rozwiązanie nowych problemów. Doświadczenie zdobyte podczas akcji pozwala im również na unikanie zagrożeń.

Przy różnorodności działań prowadzonych przez Państwową Straż Pożarną (PSP) bardzo rzadko zdarzają się dwie identyczne akcje. Jednakże zawsze istnieje kilka elementów wspólnych, pozwalających na przeniesienie części rozwiązań

stosowanych w przeszłości na sytuacje nowe. Dlatego też podejmowanie decyzji w sytuacjach kryzysowych, na podstawie zdobytego doświadczenia jest metodą powszechnie wykorzystywaną przez dowódców.



Rys. 1. Diagram podejmowania decyzji metodą RPD [2]

Badania prowadzone przez Kleina [1, 2] odnośnie podejmowania decyzji w sytuacjach złożonych oraz pod presją czasu pokazały, iż dowódcy z dużym doświadczeniem bardzo często podejmują trafne decyzje prowadzące do rozwiązań problemu już za pierwszym razem. Metoda podejmowania decyzji na podstawie doświadczenia charakteryzuje się dużą szybkością oraz trafnością. Stanowi to o jej dużej przydatności w sytuacjach kryzysowych. Dalsze analizy prowadzone przez Kleina pozwoliły na opracowanie modelu tej metody. Opracowana metoda otrzymała nazwę decydowania na podstawie rozpoznania RPD (*ang. recognition-primed decision*) [1]. RPD jest techniką psychologiczną modelującą sposób podejmowania szybkich decyzji w złożonych sytuacjach. Metoda polega na rozpoznaniu danej sytuacji, znalezieniu w pamięci podobnego przypadku, przystosowania jego rozwiązania do warunków bieżących oraz zastosowanie. W przypadku niepowodzenia przyjętego działania zbadanie jego przyczyn, zdobycie nowej wiedzy, modyfikacja metody i jej powtórne zastosowanie. Schemat 1 ilustruje metodę RPD.

Jak pokazały badania [2], bardzo ważnym czynnikiem w trafności podejmowanych rozwiązań jest doświadczenie podejmującego decyzję. W przypadku strażaka kierującego akcją ratowniczą (KAR) w większości przypadków jest to doświadczenie własne lub kolegów z tej samej jednostki. Przy różnorodności zdarzeń, w których interweniuje PSP doświadczenie jednego człowieka, czy nawet całej jednostki jest niewystarczające do skutecznego i bezpiecznego prowadzenia każdej akcji. Dlatego też bardzo często zdarza się, iż błędy popełnione przez jednych strażaków są potem wielokrotnie popełniane przez innych. Aby skutecznie wykorzystywać metodę RPD, KAR musi mieć dostęp do wiedzy i doświadczenia zdobytego przez strażaków z całego kraju. Umożliwić może to specjalny system komputerowy, którego podstawę stanowić będzie baza wiedzy zawierająca opisy wszystkich akcji prowadzonych na terenie kraju oraz metody ich rozwiązań. Dodatkowo system musi zapewnić metodę pozwalającą na porównywanie i wyszukiwanie przy-

padków przechowywanych w bazie w taki sposób, aby liczba proponowanych rozwiązań była jak najmniejsza.

W artykule przedstawiono założenia systemu wspomaganie decyzji dla strażaków PSP. System wykorzystuje architekturę rozproszonych baz danych jako warstwę przechowywania zdarzeń oraz metodę wnioskowania na podstawie przypadków CBR (*ang. case-based reasoning*) jako technikę porównywania przypadków w bazie danych oraz wyznaczania rozwiązań.

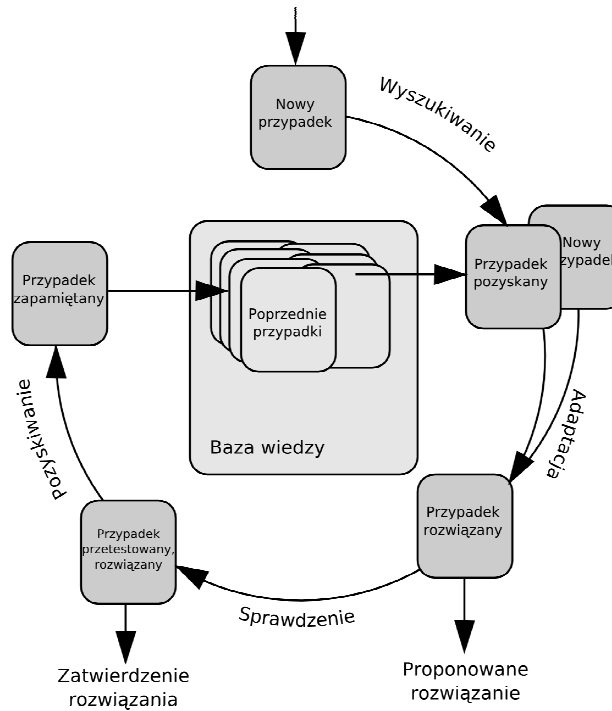
2. Wnioskowanie na podstawie przypadków

We wstępie przedstawiono metodę RPD jako najbardziej optymalną do podejmowania decyzji w sytuacjach kryzysowych. Jednakże RPD jest techniką podejmowania decyzji wykorzystywaną przez ludzi. Badania nad rozwojem sztucznej inteligencji AI (*ang. artificial intelligence*) przyczyniły się do powstania kilku technik symulacji tej metody przez systemy komputerowe. Jedną z najbardziej znanych jest wnioskowanie na podstawie przypadków CBR [3, 4]. Technika ta działa podobnie do używanej przez ludzi metody RPD.

Systemy CBR składają się najczęściej z bazy danych, w której przechowywane są przypadki oraz metody ich rozwiązania. Dodatkowo posiadają narzędzia implementujące algorytmy, które potrafią porównywać przypadki oraz adaptować rozwiązania. Działanie systemów CBR jest następujące [3]: zdarzenie, które należy rozwiązać, opisywane jest przez zbiór atrybutów. W przypadku pożaru mieszkania atrybutami takimi przykładowo mogą być: liczba kondygnacji budynku, rodzaj klatki schodowej itp. Atrybuty te porównywane są z opisem zdarzeń przechowywanych w bazie. System wyszukuje z nich te, które są najbardziej podobne do występującego zdarzenia. Ze znalezionych przypadków pobierane jest rozwiązanie. Stanowi ono podstawę do opracowania nowego rozwiązania. Poprzez odpowiednie algorytmy adaptacji wyznacza się rozwiązanie dla obecnego zdarzenia. Rozwiązanie wdraża się i odczytuje wyniki jego zastosowania. W przypadku niepowodzenia bada się jego przyczyny i modyfikuje rozwiązanie. Działanie powtarza się aż do osiągnięcia zamierzonego celu. Wyznaczone skuteczne rozwiązanie zapisywane jest w bazie wraz z opisem zdarzenia. Na rys. 2. przedstawiono sposób działania systemów wykorzystujących technikę CBR.

Technika CBR znalazła zastosowanie w następujących systemach [4]:

- Diagnostyki i określania awarii urządzeń technicznych;
- Zarządzania wiedzą;
- Wspomagania sprzedaży w sklepach internetowych;
- Projektowania oraz inżynierii;
- Wspomagających diagnozę lekarską;
- W oprogramowaniu prawniczym.



Rys. 2. Cykl wyszukiwania rozwiązania w metodzie CBR [3]

Jakość oraz skuteczność metody CBR zależy od czterech podstawowych czynników:

- sposobu oraz dokładności odwzorowania modelowanej dziedziny;
- jakości dokonywania porównań;
- metod adaptacji rozwiązań;
- dokładności opisywanych przypadków.

3. System wspomaganie decyzji dla PSP

Metoda CBR wykorzystywana została do budowy systemu wspomaganie decyzji dla PSP. Podstawą działania systemu są opisy akcji ratowniczych, nazywanych zdarzeniami lub przypadkami, przechowywane w bazie. Przypadki opisane są przez odpowiednio dobrane zbiory atrybutów w zależności od rodzaju zdarzenia. Dodatkowo, zdarzenia zostały podzielone na kategorie tworzące hierarchiczną strukturę.

Działanie systemu zostanie omówione na przykładzie. Dyspozytor stanowiska kierowania otrzymuje telefoniczne zgłoszenie o pożarze lub innym zagrożeniu.

Dane otrzymane od zgłaszającego wprowadza do formularza umożliwiającego wyszukiwanie przypadków w bazie. Pozwala to na wstępne określenie kategorii zdarzenia. Na przykład kategoria pożary-> lasy. Hierarchiczny zapis zdarzeń w bazie pozwala już we wstępnej fazie zawęzić wyniki poszukiwania. Następnie dyspozytor stanowiska kierowania wysyła zastęp w miejsce zdarzenia. Po dojechaniu na miejsce akcji KAR składa wstępny meldunek. W meldunku tym zawarte są bardziej szczegółowe informacje określające zdarzenie. Na ich podstawie dyspozytor stanowiska kierowania może zdefiniować kolejne atrybuty zdarzenia. Wprowadzone do systemu umożliwiają wyszukanie w bazie podobnych zdarzeń. System za pomocą odpowiednich funkcji wybiera jeden najbardziej podobny przypadek. Następnie przekazuje jego rozwiązanie dyspozytorowi. Dyspozytor drogą radiową przekazuje sugestie kierującemu akcją. KAR na podstawie otrzymanych wskazówek oraz oceny sytuacji dokonuje adaptacji rozwiązania do bieżącego zdarzenia. Rozwiązanie jest wdrażane, a jego skuteczność poddana ocenie.

W trakcie prowadzenia akcji zmieniające się parametry pożaru przekazywane są dyspozytorowi, który wprowadza je do systemu i na bieżąco modyfikuje kryteria wyszukiwania. W wyniku tego zmienić się może proponowane rozwiązanie. W przypadku skutecznego zakończenia akcji parametry zdarzenia oraz zastosowane rozwiązania są automatycznie zapisywane w bazie, tworząc nowy przypadek.

Do porównywania przypadków system wykorzystuje algorytm zaproponowany przez Cortesa [5]. Algorytm ten wyznacza wartość funkcji podobieństwa pomiędzy zapytaniem Q a przypadkiem C z bazy C . Wartość funkcji jest średnią ważoną atrybutów porównywanych przypadków. Sposób wyznaczania wartości funkcji podobieństwa określa zależność (1).

$$\text{SIM}(Q, C) = \frac{\sum_{i=1}^n w_i \text{sim}_i(Q_i, C_i)}{n} \quad (1)$$

gdzie: Q oznacza zapytanie, C – przypadek zapisany w bazie, $\text{sim}(Q, C)$ jest wartością globalnego podobieństwa między zapytaniem a przypadkiem, n jest liczbą porównywanych atrybutów, w_i wagą danego atrybutu, natomiast sim_i wartością podobieństwa atrybutu i -tego.

Do porównania przypadków została użyta średnia ważona, ponieważ poszczególne atrybuty mogą mieć różne znaczenie w zależności od sytuacji. Na przykład w przypadku pożaru budynku mieszkalnego usytuowanego w miejscu, gdzie nie ma sieci hydrantowej, kluczowym zagadnieniem stanowiącym o pomyślnym opamowaniu pożaru jest zorganizowanie zaopatrzenia wodnego. Jeżeli natomiast przy wspomnianym budynku istnieje sieć hydrantowa, czynnik ten ma już drugorzędne znaczenie dla sposobu prowadzenia akcji.

Opisy przypadków przechowywane w bazie mogą być charakteryzowane przez atrybuty o różnych typach. Mogą to być liczby, tekst, wyrażenia boolowskie, data oraz wiele innych. Nawet w przypadku liczb wyznaczenie podobieństwa atrybutów nie jest sprawą prostą. Nie zawsze można przyjąć, że mniejsza wartość bezwzględna różnicy pomiędzy wartościami atrybutów przekłada się na większe podobieństwo. Bardzo często ważny jest również znak tej różnicy.

Przykładem może być przypadek akcji ratowniczej po wycieku substancji wybuchowej. Jednym z parametrów charakteryzujących to zdarzenie może być stężenie substancji. Po wykonaniu pomiaru okazało się, iż stężenie substancji w powietrzu przekracza dolną granicę wybuchowości (DGW). W bazie systemu znajdują się dwa podobne przypadki, w jednym różnica pomiędzy wartością parametrów wyszukiwanych i zapisanych jest bardzo mała, jednakże substancja jest poniżej DGW, w drugim różnica jest większa, jednakże substancja znajduje się powyżej DGW. Na podstawie tych porównań można stwierdzić, iż bardziej podobnym przypadkiem jest ten, w którym różnica jest większa, ale stężenie znajduje się powyżej DGW. Pozwala on lepiej opisać mogące wystąpić zagrożenia.

Trudności z porównywaniem zdarzeń mogą również wystąpić w przypadku tekstu. Na przykład porównywanie parametru określającego kolor dymu. W bazie znajdują się przypadki, w których kolor dymu ma barwę brunatną oraz czerwoną, natomiast w zapytaniu barwa ta określona jest jako ciemnoczerwona. Który przypadek lepiej pasuje do zdarzenia z ciemnoczerwoną barwą dymu?

Sposób dokonywania porównań przypadków wymaga doskonałej znajomości modelowanej dziedziny. Aby system mógł precyzyjnie porównywać przypadki, konieczne są jego testy w warunkach rzeczywistych. W trakcie tych testów możliwe będzie wyznaczenie metod porównywania szczególnych przypadków oraz opracowanie odpowiednich słowników synonimów wyrazów.

Porównywanie przypadków jest kluczową sprawą w systemach wspomaganie dowodzenie bazujących na technice CBR. Liczba wyników otrzymanych z porównania powinna być jak najmniejsza. Podczas akcji ratowniczej nie ma czasu na studiowanie wielu przypadków i ekstrapolowanie rozwiązania ze zbioru zwróconych rozwiązań. Metoda porównywania przypadków musi być na tyle precyzyjna, aby liczba proponowanych rozwiązań ograniczona była maksymalnie do trzech.

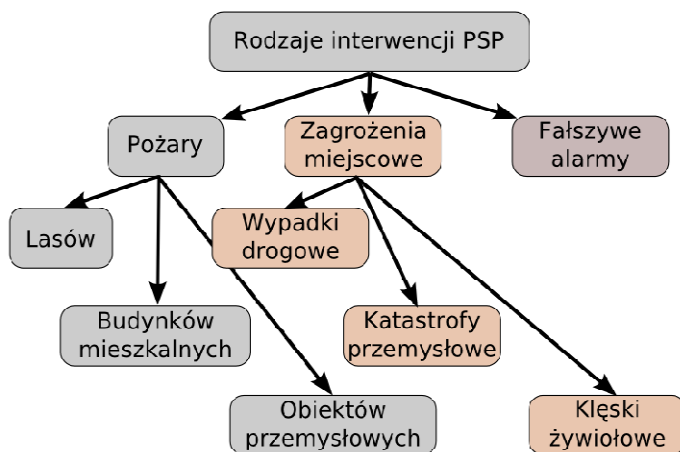
Bardzo rzadko zdarza się, iż przypadki przechowywane w bazie są na tyle podobne do wyszukiwanego, że proponowane rozwiązanie można zastosować bezpośrednio. W większości sytuacji wymagana jest modyfikacja rozwiązania do warunków panujących w obecnym zdarzeniu. Dlatego też systemy wspomaganie decyzji powinny również posiadać narzędzia do adaptacji rozwiązania.

Zakres interwencji PSP jest bardzo szeroki [6]. Wymienić tutaj można gaszenie pożarów, likwidację zagrożeń chemicznych, ekologicznych, klęsk żywiołowych, a także pomoc w wypadkach drogowych. Dla tak szerokiego spektrum

bardzo trudno jest ustalić algorytm adaptacji rozwiązań. Algorytmów takich powinno być raczej więcej w zależności od rodzaju akcji. Jest to bardzo trudne zadanie nawet w dziedzinach mniej skomplikowanych niż interwencje PSP. Dlatego też wyznaczenie algorytmów adaptacji wymaga szerokich badań i testów. Na obecnym etapie pracy założono, iż system nie będzie dokonywał adaptacji rozwiązań. Jego zadaniem będzie znalezienie najbardziej podobnych przypadków i przesłanie ich rozwiązań do prowadzącego działania. Kierujący działaniami na podstawie przesłanych informacji oraz oceny sytuacji dokona adaptacji rozwiązania do danego zdarzenia.

3.1. Architektura bazy danych

Zgodnie z rozporządzeniem [6] oraz specyfikacją opracowaną dla programu ewidencji zdarzeń PSP o nazwie EWID [7], interwencje PSP można podzielić na trzy podstawowe grupy: pożary, miejscowe zagrożenia oraz fałszywe alarmy. Każda z tych kategorii ma swoje podgrupy w zależności od rodzaju budynku lub występującego zdarzenia. Fragment tego podziału przedstawiono na rys. 3.



Rys. 3. Fragment podziału interwencji PSP

Źródło: opracowanie własne.

Jako baza przypadków dla aplikacji CBR wykorzystywane się różne technologie. Wymienić tutaj można: zapis danych w nieustrukturyzowanych plikach tekstowych [8,9] lub wykorzystujących składnie XML [10], CRN (ang. Case Retrieval Net)[11], hurtownie danych [12] oraz relacyjne bazy danych [13]. Jak już wspomniano jednym z ważnych czynników decydującym o powodzeniu działań ratowniczych jest czas. Należy zatem wykluczyć te technologie, które nie posiadają mechanizmów przyspieszających wyszukiwanie informacji. Zaliczyć do nich można

pliki tekstowe oraz CRN. Hurtownie danych należy również wykluczyć jako narzędzie do przechowywania przypadków opisujących akcje ratownicze, ponieważ dziedzina ta nie jest podatna na agregację. Zatem użycie hurtowni danych miałyby się z celem. Najbardziej racjonalnym wydawałoby się wykorzystanie relacyjnych baz danych. Jednakże autorzy proponują pewne nowatorskie rozwiązanie w postaci wykorzystania katalogowych baz danych.

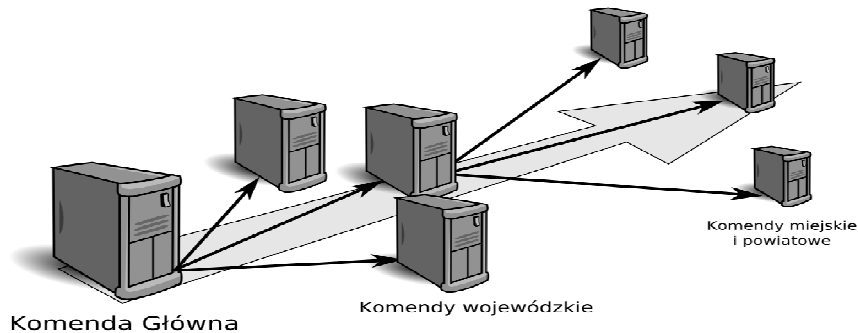
Zaprezentowana struktura ma postać hierarchiczną i w związku z tym bardzo dobrze nadaje się do modelowania w katalogowych bazach. Dodatkowo opis zdarzeń poprzez atrybuty jest sposobem opisu obiektów stosowanych w katalogach. Ponadto katalogowe bazy danych są optymalizowane pod kątem wyszukiwania informacji, co jest istotne z punktu widzenia systemu przeznaczonego do wspomaganie decyzji w czasie akcji ratowniczych. Dlatego też podstawę warstwy danych stanowić będzie technologia LDAP [14].

W korzeniu drzewa informacji katalogowej zostaną zdefiniowane trzy konteksty. Będą one dotyczyły odpowiednio: pożarów, zagrożeń miejscowych oraz fałszywych alarmów. Reprezentację tych obiektów w katalogach stanowić będą kontenery, które podzielą przypadki na kolejne podgrupy. Struktura będzie rozgałęziać się w dół, aż do kompleksowego opisu wszystkich rodzajów interwencji PSP zgodnie z rozporządzeniem. Kontenery najniższego poziomu przechowywać będą obiekty klasy przypadek (*ang. case*). Stanowiąc one będą reprezentację zdarzeń, w których interweniowała PSP.

Architekturę fizyczną dla warstwy danych stanowić będzie rozproszona baza danych. W tym celu wykorzystana zostanie baza danych przedstawiona w pracy [15, 16]. Baza ta fizycznie składa się z komputerów rozmieszczonych w jednostkach PSP na terenie całego kraju. Projekt jej został opracowany w celu wspomaganie treningu i podnoszenia wyszkolenia strażaków PSP. Baza ta przechowuje nieustrukturyzowane informacje w postaci analiz zdarzeń lub innych dokumentów opisujących akcję lub przeprowadzone ćwiczenia. Część z tych dokumentów zostanie również wykorzystana do tworzenia nowych przypadków bazy wspomaganie decyzji. Schemat logiczny bazy danych przedstawiono na rys. 4. Każdy z węzłów przedstawionych na rysunku stanowi serwer LDAP. Dany serwer przechowuje dokumentację akcji prowadzonych na danym terenie. Struktura tworzy całość dzięki wykorzystaniu obiektów typu odnośnik (*ang. referral*) [17].

Na tak zdefiniowaną warstwę fizyczną należy nanieść logiczny schemat bazy przedstawiony na rys. 3. Istnieje kilka sposobów alokacji danych, przedstawionych na schemacie, w rozproszonej architekturze. Jednym z nich jest utworzenie na każdym z węzłów schematu bazy zaprezentowanej na rys. 3. Każdy węzeł przechowywałby opisy zdarzeń, w których interweniowała PSP z terenu danej jednostki, podzielone zgodnie z rozporządzeniem [6]. Taki sposób przechowywania danych procentowałby proporcjonalnym dzieleniem obciążenia przez węzły oraz zapew-

niona byłaby bardzo duża lokalność odwołań. Jednakże dla zapytań o charakterze globalnym przeszukiwane byłyby zawsze wszystkie węzły struktury.



Rys. 4. Struktura rozproszonej bazy danych wykorzystywana do przechowywania informacji generowanych w PSP

Źródło: opracowanie własne.

Wydłużałoby to znacznie czas wyszukiwania informacji, a w przypadku awarii mogłoby narazić na niepotrzebne przerwy. Czas wyszukiwania globalnego przedstawiałby się zgodnie z zależnością (2).

$$t_c = n(t_k + t_p) \quad (2)$$

gdzie:

n – współczynnik przeszukiwania równoległego,

t_k – czas komunikacji klient – serwer,

t_p – czas przetwarzania zapytania przez serwer.

Realizacja zapytania globalnego uruchamiałaby procesy wyszukiwania na wszystkich węzłach. Algorytm podążania za odnośnikiem nie umożliwia w pełni równoległego przetwarzania zapytania. Zatem w przypadku realizacji założenia, iż w każdej komendzie wojewódzkiej oraz powiatowej umiejscowiony zostanie serwer LDAP wartość n będzie bardzo duża.

Łącza komputerowe wykorzystywane w komendach powiatowych mają różną przepustowość. Ponadto jakość części z nich jest bardzo niska, stąd też czas odpowiedzi może być długi. Problemy te będą powodować kolejne opóźnienia w realizacji przez system zapytań globalnych. Należy zatem przyjąć, iż mimo tego, że czas przetwarzania żądania przez węzły byłby stosunkowo krótki, to przy takiej alokacji danych całkowity czas realizacji zapytania globalnego byłby bardzo długi. Jest to niedopuszczalne dla systemu odpowiedzialnego za wspomaganie decyzji

w sytuacjach kryzysowych. Dlatego konieczne jest wykorzystanie innej architektury warstwy danych.

Przeciwnieństwem architektury całkowicie rozproszonej jest system scentralizowany. W architekturze tej cała baza danych wykorzystywana w systemie wspomagania decyzji umieszczona byłaby na jednym komputerze. Wówczas czas realizacji zapytania przedstawiałby się zgodnie z zależnością (3).

$$t_c = t_k + t_p \quad (3)$$

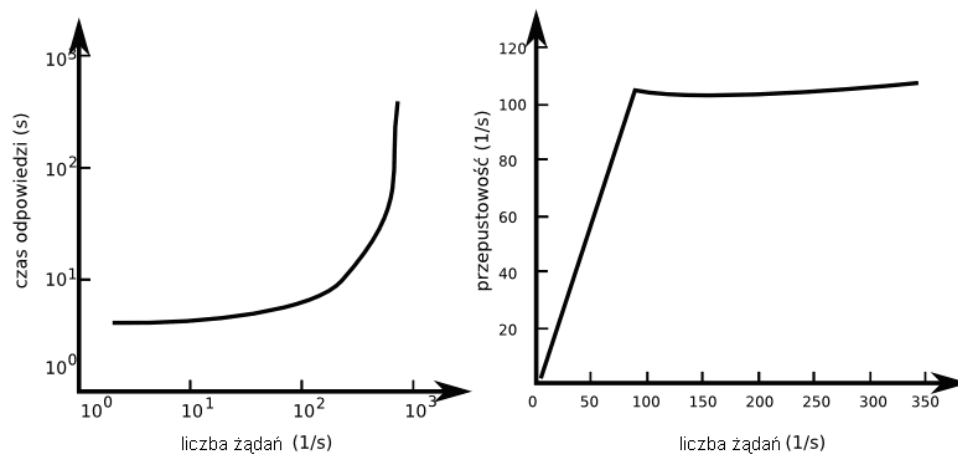
gdzie:

t_k – czas komunikacji klient – serwer,

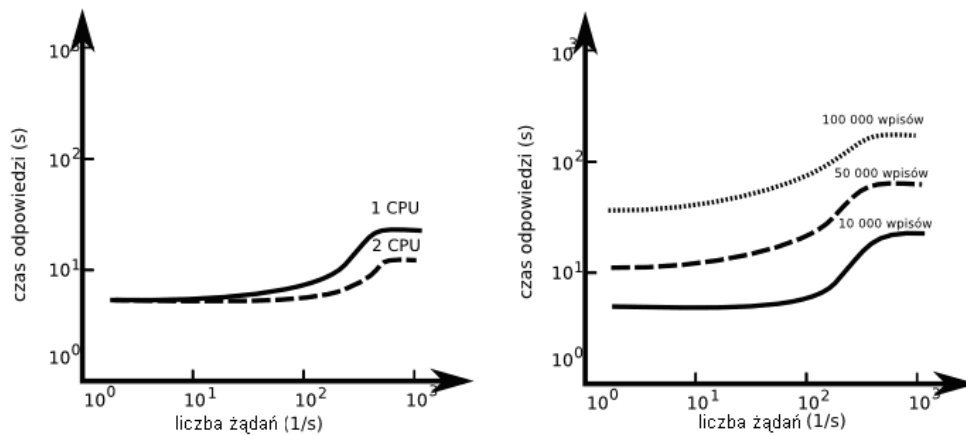
t_p – czas przetwarzania zapytania przez serwer.

W porównaniu z systemem rozproszonym czas przetwarzania zapytania przez serwer byłby bardzo długi. Wynika to między innymi z tego, iż liczba danych przetwarzanych przez system byłaby bardzo duża oraz liczba żądań klienckich w jednostce czasu byłaby również znaczna. Te dwie cechy wpływają na czas przetwarzania żądania przez system. Na rys. 5. zaprezentowano wyniki badań przeprowadzone przez Wanga oraz Schulzrizne [18]. Pokazują one zależność czasu realizacji zapytania oraz przepustowości serwera LDAP od liczby połączeń.

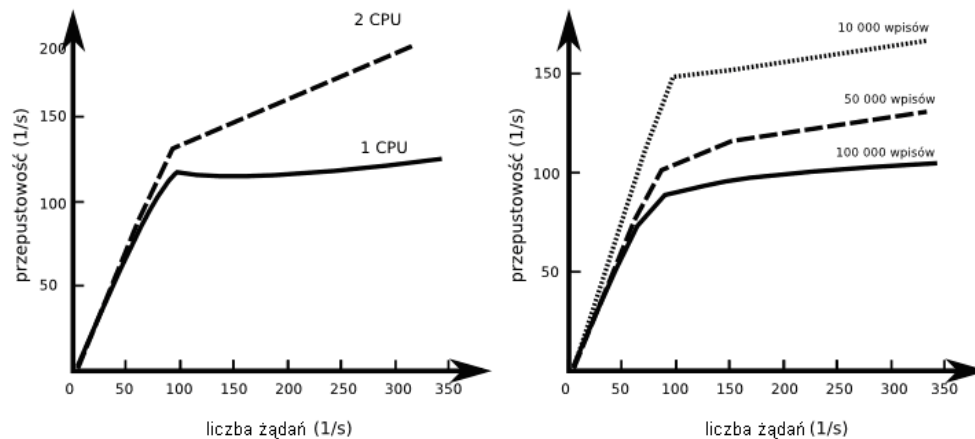
Na rys. 5. wykazano, iż przy pewnej liczbie połączeń w jednostce czasu następuje wyczerpanie potencjału serwera i czas realizacji połączenia gwałtownie rośnie. Kolejne wykresy (rys. 6., 7.) pokazują zależność wydajności serwera w funkcji wielkości bazy danych oraz liczby procesorów.



Rys. 5. Zależność czasu odpowiedzi oraz przepustowości serwera LDAP od liczby żądań [18]



Rys. 6. Zależność czasu odpowiedzi serwera LDAP od liczby procesorów oraz wielkości bazy [18]



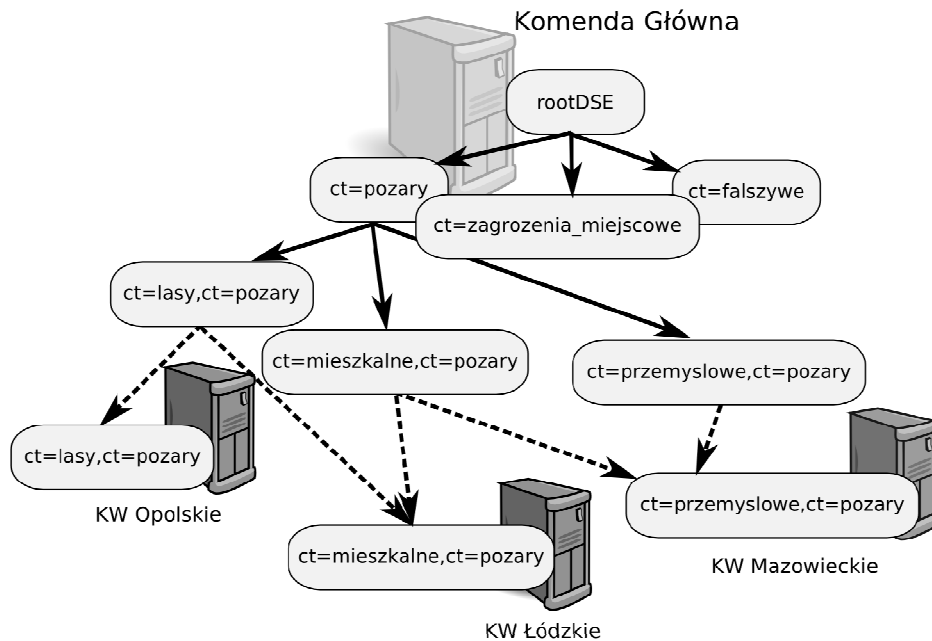
Rys. 7. Zależność przepustowości serwera LDAP od liczby procesorów oraz wielkości bazy [18]

Na przedstawionych wykresach uwidocznione zostało, iż wraz ze wzrostem wielkości bazy danych rośnie czas realizacji zapytania. Spada natomiast liczba przetwarzanych żądań. Parametry te poprawić może zwiększenie liczby procesorów w systemie bazy danych.

Z powyższych rozważań wynika, iż czysty system scentralizowany nie jest najlepszym rozwiązaniem, jeżeli chodzi o szybkość realizacji żądania. Ponadto systemy scentralizowane generują znaczne koszty w ich budowie oraz utrzymaniu.

Dlatego też sposób alokacji danych wykorzystywanych w trybie wspomaganie decyzji powinien bazować na rozwiązaniach kompromisowych, pomiędzy całkowitym rozproszeniem a centralizacją.

Na podstawie tych rozważań w systemie wspomaganie decyzji zaproponowano następujące rozwiązanie. Serwer w Komendzie Głównej (KG) przechowuje korzeń drzewa struktury oraz kontener z obiektami stanowiącymi reprezentację głównych kategorii zdarzeń. Obiekty te są również kontenerami przechowującymi kolejne podgrupy zdarzeń. Co oznacza, że na serwerze w KG przechowywane są dwa poziomy podziału zdarzeń. Drugi poziom stanowią odnośniki do serwerów zajmujących się obsługą tych grup. Serwery te zlokalizowane są w komendach wojewódzkich. Liczba grup drugiego rzędu jest około dwudziestu trzech przy szesnastu serwerach w komendach wojewódzkich (KW). Dlatego też konieczne jest, aby jeden serwer przechowywał kilka grup zdarzeń. Ponadto należy każdy z serwerów dodatkowo replikować przez inny serwer z komendy wojewódzkiej. Zwiększy to znacznie niezawodność systemu. W przypadku awarii któregoś z serwerów jego zadanie przejmuje inny. Schemat poprawionej struktury przedstawiono na rys. 8.



Rys. 8: Struktura bazy danych wykorzystywana w trybie wspomaganie decyzji

Źródło: opracowanie własne.

Zaproponowane rozwiązanie przynosi wymierne korzyści:

- rozdzielenie danych na serwery powoduje, iż pojedynczy serwer ma do przetworzenia mniejszą liczbę danych, przez co czas realizacji żądania się skraca (patrz rys. 6.);
- całość bazy obsługiwana jest przez kilkanaście procesorów, co również zmniejsza czas realizacji żądania;
- komendy wojewódzkie wyposażone są w dobre łącza internetowe, co poprawia czas komunikacji;
- replikowanie fragmentów bazy poprawia niezawodność systemu, a także przyspiesza szybkość opracowania żądania przez rozkładanie obciążenia.

Zalety te potwierdzają, iż rozdzielenie struktury bazy na fragmenty i przechowywanie ich na serwerach w komendach wojewódzkich jest najbardziej optymalnym rozwiązaniem. Dodatkowo replikacja fragmentów bazy w innych węzłach poprawia szybkość działania oraz niezawodność systemu. W przedstawionej architekturze czas realizacji żądania opisuje zależność (4):

$$t_c = t_{kg} + t_{pg} + n(t_{kw} + t_{pw}) \quad (4)$$

gdzie:

t_{kg} – czas komunikacji klient serwer KG,

t_{pg} – czas przetwarzania żądania przez serwer KG,

t_{kw} – czas komunikacji klient serwer KW,

t_{pw} – czas przetwarzania żądania przez serwer KW,

n – liczba grup danych do odczytania.

Przy odpowiednio dobrych łączach komend wojewódzkich można założyć, iż czas komunikacji z serwerem w Komendzie Głównej oraz w komendzie wojewódzkiej jest taki sam. Zatem $t_{kg} = t_{kw}$. Dodatkowo, ze względu na liczbę przechowywanych danych czas przetwarzania żądania przez serwer w KG jest dużo mniejszy od serwera w KW. W przypadku wspomaganie decyzji można przyjąć, iż wyszukiwane będą zdarzenia wewnątrz tylko jednej kategorii, np. pożary lasów. Zatem n przyjmuje wartość 1. Po uwzględnieniu tych założeń czas realizacji żądania przedstawia się zgodnie z zależnością 5.

$$t_c = 2t_k + t_{pw} \quad (5)$$

Otrzymany wzór jest podobny do zależności (3), z tym że liczba przetwarzanych danych jest znacznie mniejsza oraz realizowane jest to na wielu procesorach.

3.2. Opis przypadków przechowywanych w bazie

Na rys. 3. przedstawiono fragment struktury bazy przeznaczonej do przechowywania danych wykorzystywanych w trybie treningu. Na samym szczycie struk-

tury znajdują się konteksty dla trzech podstawowych kategorii zdarzeń: pożarów, miejscowych zagrożeń oraz fałszywych alarmów. Odpowiedniki tych kategorii reprezentują w bazie danych obiekty typu kontener. W systemie wspomaganie decyzji do opisu tych obiektów wykorzystano klasy typu *javaContainer*. Klasa wymaga zdefiniowania jednego parametru o nazwie *commonName*. Parametr ten posłuży do określenia nazwy kategorii. Wewnątrz głównych kontenerów umieszczone są kolejne kontenery do dalszego podziału grup zdarzeń. Grupy zostały wyznaczone zgodnie z rozporządzeniem [6].

Zgodnie z tym źródłem w grupie pożarów wyznaczone są między innymi podgrupy:

- obiekty użyteczności publicznej;
- obiekty mieszkalne;
- środki transportu;
- lasy,
- inne.

Miejscowe zagrożenia dzielone są natomiast między innymi według następujących kategorii:

- chemiczne;
- ekologiczne;
- radiologiczne;
- budowlane;
- w komunikacji drogowej;
- medyczne.

Zaewidencjonowane oraz wykorzystywane będą również fałszywe alarmy. Wprowadzanie ich do bazy będzie miało na celu ich lepsze wykrywanie. Fałszywe alarmy zostaną podzielone na dwie kategorie:

- zgłoszone przez ludzi;
- z instalacji wykrywania pożarów.

Wszystkie kategorie wymienione są w rozporządzeniu [6].

Przedstawione kategorie stanowią będą nazwy kontenerów dla obiektów klasy przypadek. Obiekty te charakteryzowane przez odpowiednie atrybuty będą reprezentowały poszczególne interwencje PSP. Prawidłowo dobrane atrybuty stanowią kluczową rolę w skuteczności systemów wykorzystujących metodę CBR. Dlatego też bardzo ważne jest odpowiednie przeprowadzenie tego procesu. Dobór atrybutów musi być poprzedzony odpowiednią analizą modelowanej dziedziny oraz empirycznie potwierdzony w trakcie testów. Nie da się z poziomu teoretycznego ustalić, czy parametr typu kolor dymu będzie miał istotne znaczenie dla wyszukiwania rozwiązania danej sytuacji. Dlatego też, we wstępnym etapie prac do opisu przypadków w bazie przyjęto strukturę zdefiniowaną w rozporządzeniu [6]. Zostanie ona rozbudowana i uszczegółowiona na podstawie informacji z wywiadów odbytych ze strażakami Jednostki Ratowniczo-Gaśniczej SGSP.

Zgodnie z przyjętymi założeniami zaproponowano ogólną klasę *case* – *przypadek*, charakteryzowaną przez następujące atrybuty:

- rodzaj obiektu – określa rodzaj obiektu, którego dotyczy zdarzenie. W przypadku pożaru obiektami mogą być: hotele, maszyny i urządzenia technologiczne, samochody osobowe itp. Wykaz możliwych obiektów reguluje rozporządzenie [6];
- wielkość zdarzenia – określa skalę występowania danego zdarzenia. Możliwe wartości to: małe, lokalne, średnie, duże, bardzo duże. Sposób klasyfikacji danego zdarzenia reguluje rozporządzenie;
- opis rozwiązania – krótkie streszczenie najważniejszych informacji odnośnie likwidacji zagrożenia. Parametr ten zawiera sedno informacji potrzebnych do prawidłowego działania w trakcie akcji.

Ogólna klasa *przypadek* będzie rozszerzana przez klasy charakteryzujące rodzaj danego zdarzenia. I tak na przykład dla pożarów budynków mieszkalnych utworzono klasę *dziedziczącą* z klasy *przypadek* o nazwie *home_fire_case*. Klasa ta rozszerza klasę *case* o następujące atrybuty:

- kategoria zagrożenia – kategoria zagrożenia ludzi, do jakiej należy dany obiekt;
- zaopatrzenie wodne – rodzaj zaopatrzenia wodnego, w jaki wyposażony jest obiekt. Możliwe wartości to sieć hydrantowa wewnętrzna, tryskaczowa lub zraszaczowi;
- sieć hydrantowa zewnętrzna – możliwe parametry: brak lub odległość sieci od budynku;
- zbiorniki przeciwpożarowe – brak, naturalne lub sztuczne;
- powierzchnia obiektu;
- kubatura obiektu;
- liczba kondygnacji nadziemnych;
- liczba kondygnacji podziemnych;
- liczba wejść do budynku;
- liczba klatek schodowych;
- wolnostojący – true, false;
- strefy pożarowe – true, false;
- największa strefa pożarowa;
- instalacja elektryczna – true, false;
- instalacja gazowa – true, false;
- ogrzewanie parowe – true, false;
- ogrzewanie olejowe – true, false;
- instalacja technologiczna – true, false;
- klimatyzacja centralna – true, false;
- wentylacja mechaniczna – true, false;
- instalacje oddymiające – true, false;
- oświetlenie awaryjne – true, false;

- oświetlenie ewakuacyjne – true, false;
- urządzenia ewakuacyjne – true, false;
- urządzenia ratownicze – true, false;
- urządzenia sygnalizacji pożaru – true, false;
- konstrukcja nośna – żelbetowa, stalowa, murowana (tradycyjna), drewniana, drewnopochodna, mieszana;
- konstrukcja dachu/stropodachu – żelbetowa, stalowa, murowana (tradycyjna), drewniana, drewnopochodna, mieszana;
- pokrycie dachu – palne, niepalne, mieszane;
- konstrukcja ścian wnętrza – żelbetowa, stalowa, murowana (tradycyjna), drewniana, drewnopochodna/gipsowa, mieszana, ściany osłonowe/elewacje;
- klatki schodowe – zabiegowe, spocznikowe;
- dojazd do budynku – brak, utrudniony, normalny.

Podobny sposób rozszerzania klasy podstawowej *case* wykorzystywany będzie do innego typu zdarzeń. W zależności od rodzaju zdarzenia definiowane będą poszczególne subclassy rozszerzające klasę główną i pozwalające na odpowiedni opis danego zdarzenia. Dobór odpowiednich atrybutów w zależności od rodzaju zdarzenia opracowany został w [19].

Pokazany sposób opisu stanowi jedynie założenia nakreślające ideę projektu. Zakłada ona opis każdego zdarzenia na zasadzie atrybut–wartość. Sposób ten umożliwia później wyznaczenie wartości funkcji podobieństwa opisywanej w rozdziale 2.

3.3. Pozyskiwanie przypadków

Wraz z wprowadzeniem systemu do użytku dodawanie nowych przypadków do bazy odbywało się za pomocą specjalnego interfejsu. Zapewnia on od razu odpowiedni poziom strukturyzacji danych potrzebnych do wspomagania decyzji. Jednakże należy pamiętać, że wprowadzenie systemu do użytkowania bez wypełnionej bazy danych mija się z celem. Z pustą bazą danych system nie jest w stanie dostarczyć żadnego rozwiązania prowadzącemu akcję. Dlatego przed jego wdrożeniem baza musi być uzupełniona o obiekty typu zdarzenia. Zostaną one pozyskane z dwóch źródeł: po pierwsze z systemu ewidencji zdarzeń EWID, po drugie z dokumentów tekstowych w postaci analiz zdarzeń przechowywanych w bazie danych trybu treningu [15,16]. System ewidencji zdarzeń zapewni podstawowe dane w postaci określenia typu zdarzenia, jego kategorii, obiektu zdarzenia, a także jego wielkości. Natomiast pozostałe atrybuty, a także opis samego rozwiązania musi być zaimportowany z dokumentów tekstowych, typu analizy zdarzeń. Analizy zdarzenia to dokumenty sporządzane po niektórych akcjach, mające na celu zwięzły i syntetyczny opis akcji oraz prowadzonych działań. Rodzaj zdarzeń, do których wymagane jest wykonanie analizy, reguluje ustawa [6].

Powiązanie odpowiedniego zdarzenia z EWID oraz dokumentu analizy zapewnia unikalny numer zdarzenia występujący w obu obiektach. Na podstawie tej zależności można dane zawarte w systemie ewidencji zdarzeń rozszerzać o informacje występujące w dokumentach tekstowych.

Powiązanie tych dwóch informacji i utworzenie nowego obiektu klasy *przypadek* nie jest rzeczą prostą. Trudności wynikają przede wszystkim z tego, iż analizy zdarzeń są dokumentami nieustrukturyzowanymi. Opisane są luźnym tekstem najczęściej w postaci dokumentów typu .doc. Pozyskanie z nich informacji wymaga wielu zabiegów semantycznych. Ułatwienie strukturyzacji dokumentów może zapewnić fakt, iż ich forma regulowana jest przez odpowiedni akt normatywny [6]. Wyszczególniono w nim poszczególne punkty, jakie muszą być zawarte w analizie.

Analizy zdarzeń sporządzane są przez różne osoby i w związku z tym bardzo często zdarza się, iż stosowane są różne nazwy dla rzeczy o tym samym znaczeniu. Dlatego też pozyskiwanie informacji z tego typu dokumentów wymaga zastosowania zaawansowanych technik analizy tekstu.

Dziedzina techniki zajmująca się przetwarzaniem nieustrukturyzowanych dokumentów tekstowych i wyciąganiem z nich informacji wysokiej jakości nazywa się eksploracją tekstu (*ang. text mining*) [20, 21]. W zależności od potrzeb stosuje się różne metody analizy tekstu. Można tutaj wymienić [22]:

- kategoryzację tekstu (*ang. text categorization*) [23, 24] – ma ona na celu określenie, do jakiej grupy dokumentów dany tekst można zaliczyć. Metoda wykorzystywana między innymi w analizowaniu korespondencji mailowej w celu wykrywania spamu;
- grupowanie tekstu (*ang. text clustering*) [25] – proces ten, wykorzystując analizę statystyczną słów zawartych w tekście określa tematykę dokumentu;
- ekstrakcję encji i modelowanie ich relacji (*ang. concept/entity extraction*) [26] – w procesie tym z dokumentu pozyskiwane są nazwy obiektów lub osób występujących w tekście oraz wyznaczone są związki pomiędzy tymi obiektami. Na przykład można pozyskać z tekstu nazwy miast, imiona i nazwiska osób, kody pocztowe, PESEL itp.;
- analizę odczuciową (*ang. sentiment analysis*) [27] – metoda wykorzystywana do pozyskiwania opinii o danych produktach, osobach lub strategiach. Na podstawie odpowiednich algorytmów można ustalić stosunek autora tekstu do jakiegoś produktu. Metoda ta wyszukuje frazy określające negatywne, pozytywne lub neutralne emocje do danego obiektu. Wykorzystywana jest najczęściej do przeszukiwania grup dyskusyjnych w celu określenia opinii na temat danego produktu;
- streszczanie dokumentów (*ang. document summarization*) [28] – służy do wytwarzania streszczenia z obszernego dokumentu lub grupy dokumentów. Przykładowy algorytm bada powiązania między wyrażeniami. Jeżeli kilka wyrażen odwołuje się do danego wyrażenia, wówczas zwiększają jego ranking.

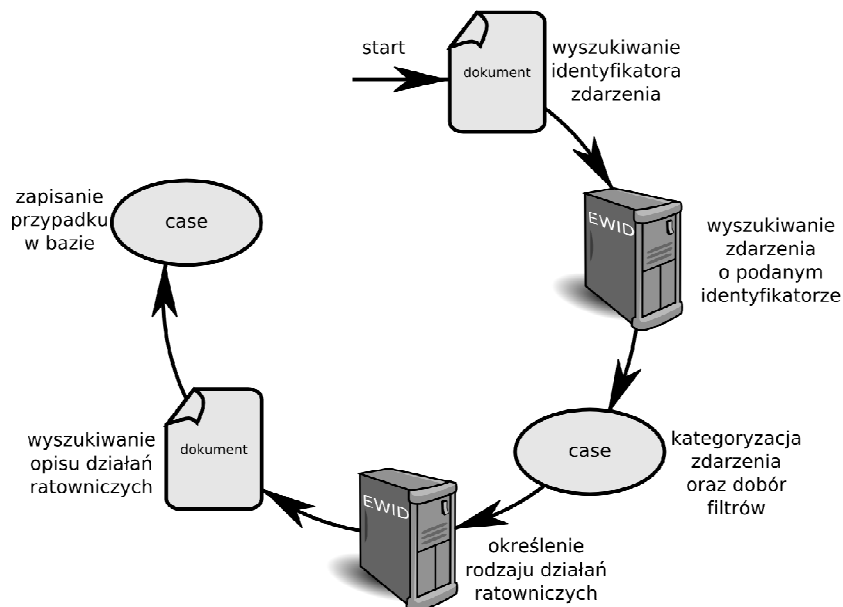
Jako podsumowanie analizy wyświetlane jest n zdań o najwyższym rankingu, tworząc streszczenie.

Bardzo rzadko wykorzystuje się wszystkie metody do analizy danego tekstu. W zależności od potrzeb wybierane są konkretne metody.

Również w przypadku systemu wspomagania decyzji dla PSP wykorzystanie wszystkich algorytmów przetwarzania tekstu nie jest konieczne. Wynika to między innymi z tego, iż część informacji zostanie uzyskana z systemu ewidencji zdarzeń. Określenie kategorii dokumentu, jego tematyki oraz ekstrakcja obiektów oraz encji wykonywana będzie na podstawie informacji zawartych w systemie EWID. Pozostałe informacje uzyskuje się poprzez analizę tekstu. Tworzenie nowego przypadku na podstawie analizy zdarzenia zostanie przeprowadzone zgodnie z metodologią opisaną poniżej.

1. Analiza zdarzenia zostaje pobrana z bazy danych trybu treningu. Uruchamiana zostaje funkcja strukturyzacji dokumentu. Dokument tekstowy przeszukiwany jest za pomocą wyrażenia regularnego pod kątem numeru identyfikacyjnego zdarzenia.
2. Program łączy się z bazą EWID z poleceniem wyszukania zdarzenia o podanym identyfikatorze.
3. Baza EWID zwraca wynik polecenia, w którym zawarte są między innymi kategoria zdarzenia (pożar, miejscowe zagrożenie, fałszywy alarm), rodzaj obiektu (budynek mieszkalny, przedszkole, samochód osobowy) oraz wielkość zdarzenia.
4. Na podstawie parametrów zwróconych z bazy dokonuje się kategoryzacji zdarzenia i umieszcza w odpowiednim miejscu struktury bazy (rys. 11.).
5. Kategoria zdarzenia determinuje wybór filtrów do analizy tekstu. Przykładowo w przypadku wypadku drogowego z udziałem cysterny przewożącej substancję niebezpieczną uruchomione zostaną filtry wyszukujące numer ONZ lub RID substancji, słowa kluczowe dotyczące cieków wodnych, np. strumień, jezioro, rzeka, studzienka kanalizacyjna. Filtry te mają na celu dostarczenie wartości dla atrybutów opisujących dane zdarzenie zgodnie ze schematem przedstawionym w rozdziale 3.2.
6. Za pomocą parametrów dotyczących rodzaju zastosowanych działań pozyskanych z bazy EWID wyszukiwane są zdania opisujące sposób likwidacji zagrożenia. Wykorzystany zostaje algorytm opracowany przez Feldmana [29].
7. Po uzyskaniu wszystkich informacji dotyczących zdarzenia przypadek zostaje zapisany w bazie wiedzy.

Na rys. 9. przedstawiono proces tworzenia nowego przypadku na podstawie danych z EWID oraz analizy tekstu zawartego w dokumencie.



Rys. 9. Proces tworzenia obiektów klasy przypadek

Źródło: opracowanie własne.

Każdy typ zdarzenia będzie posiadał grupę odpowiednich filtrów umożliwiających określenie wartości atrybutów koniecznych do zdefiniowania danego przypadku. Opracowanie odpowiednich filtrów w zależności od sytuacji zależy od definicji atrybutów dla danego przypadku. Jak zaznaczono w punkcie 3.2 wymaga to głębszych analiz i stanowić może odrębny temat badań.

3.4. Ewaluacja jakości rozwiązań przechowywanych w bazie

Pierwsze proponowane rozwiązania dla zdarzeń tworzone będą na podstawie danych zawartych w analizie. Później wraz z wdrożeniem systemu do użytku rozwiązanie wyznaczone będzie na podstawie metod wykorzystanych w trakcie prowadzenia akcji. Jakość rozwiązań, które wyznaczone są w trakcie prowadzenia akcji, w ekstremalnych warunkach, może być różna. Od bardzo dobrych rozwiązań po chybione. Dlatego też konieczna jest ewaluacja rozwiązań zapisanych w bazie. System powinien mieć specjalny panel ekspercki, za pomocą którego można będzie oceniać sposób prowadzenia akcji, a także dokonywać modyfikacji rozwiązań. Serwis ten powinien być zrealizowany przykładowo na wzór encyklopedii internetowej Wikipedia [30]. Każdy z uprawnionych użytkowników będzie mógł komentować rozwiązanie, dyskutować nad jego zasadnością, a także zaproponować

zmiany. Autoryzowane one będą przez osobę prowadzącą akcję. Tylko ona, w związku z tym, iż była na miejscu akcji, może ocenić, czy dane rozwiązanie byłoby możliwe i najlepsze do wykonania w danej sytuacji.

4. Podsumowanie

Metoda RPD, według badań przeprowadzonych przez Kleina [1, 2], jest najbardziej optymalną metodą podejmowania decyzji w sytuacjach kryzysowych. Realizacją metody RPD w systemach komputerowych jest technika CBR. Podstawą działania CBR jest odpowiednia baza przypadków. W sytuacji PSP baza taka przechowywać będzie opisy zdarzeń, w których likwidacji brali udział strażacy. System wspomaganie decyzji wykorzystywał będzie zdarzenia przechowywane w bazie do wyszukiwania rozwiązań dla nowych sytuacji. Dowódcy kierujący akcją będą mieli dostęp do opisów podobnych akcji prowadzonych na terenie całego kraju. Dzięki temu wyposażeni zostaną w odpowiednią wiedzę do trafnego wyboru metody zwalczania zagrożenia.

Opisy akcji przechowywane w bazie zdefiniowane będą przez zbiór atrybutów. Wyszukiwanie najbardziej podobnej akcji, w celu wyznaczenia rozwiązania będzie się odbywać na podstawie porównywania tych atrybutów. Funkcja ta wykorzystywała będzie średnią ważoną.

Do przechowywania obiektów reprezentujących zdarzenia, opisanych przez zmienną ilość atrybutów najbardziej optymalne wydaje się wykorzystanie katalogowych baz danych. Umożliwiają one definiowanie nowych struktur danych oraz optymalizowane są pod kątem wyszukiwania informacji.

W artykule przedstawiono wnioski z początkowej części badań dotyczących wykorzystania katalogowej rozproszonej bazy danych oraz wnioskowania na podstawie przypadków jako podstawy systemu wspomaganie decyzji dla PSP. Zaprezentowane rozwiązanie bazuje na informacjach zawartych w literaturze dotyczących wykorzystanych technologii oraz na wstępnych badaniach. W celu potwierdzenia pełnej przydatności prezentowanego rozwiązania wymagane są dalsze jego testy.

Jednym z priorytetów systemu wspomaganie decyzji dla PSP jest szybkość wyznaczania rozwiązań. Dlatego też kolejnym etapem badań prowadzonych przez autora będzie porównanie szybkości relacyjnej oraz katalogowej bazy danych jako bazy przypadków aplikacji CBR.

SUMMARY

Adam KRASUSKI
Tadeusz MACIAK

USING DISTRIBUTED DIRECTORY SERVICES AND CASE-BASED REASONING IN FIRE SERVICE DECISION PROCESSES

The article describes Decision Support System for Fire Service. The system is based on Distributed Directory Services and Case-Based Reasoning technique. The introduction of this article contains description of the current problems with decision taking in Fire Service. Short revision about currently used solutions in other fields is also described. Introduction also contains motivation and argumentation of the Recognition-Primed Decision (RPD) model choice in Fire Service.

The second part of the article contains principles of Case-Based Reasoning (CBR). CBR is the most similar to the RPD model of Artificial Intelligence technique. Therefore, it was used as a base for the Decision Support System in Fire Service.

The third part of the article is a description of the proposed Decision Support System. It consists of four subsections. The first describes distributed architecture of the Case Base. The pros and cons of using distributed architecture are described.

The second subsection contains methods description concerning representation of Fire Service operations in the Case Base.

The third subsection contains the description of procedures and methods used in collecting and acquisition of cases for the Case Base. The last subsection describes evaluation process of the cases stored in database.

The article is summarized with short conclusion about the proposed system and further works.

PIŚMIENNICTWO

1. Klein G.: Sources of Power: How People Make Decisions. Massachusetts, MIT Press, 1999.
2. Ross K. G. i inni: The Recognition-Primed Decision Model. *Military Review* 2004, Nr 4, s. 6–10.
3. Aamodt A., Plaza W.: Case-Based Reasoning. *IOComm – Artificial Intelligence Communications* 1994, Nr 1, s. 39–59.
4. Shiu S. C. K., Pal S. K.: Case-Base Reasoning: Concepts, Features and Soft Computing. *Applied Intelligence* 2004, Nr 3, s. 233–238.
5. Cortes U. i inni: Deliverable 4.1. Report on current reasonign engine practice and integration strategies. W: Report on current reasonign engine practice and integration strategies. University of Girona, Girona 2003.

6. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 29 grudnia 1999 r. w sprawie szczegółowych zasad organizacji krajowego systemu ratowniczo-gaśniczego. Dz. U. 1999, Nr 111.
7. Praca zbiorowa: Ewidencja zdarzeń – EWID99. [online]. [dostęp: 23.04.2007]. [Dostępne w Internecie: <http://www.ewid.pl/?set=rozw_ewid&gr=roz>.
8. Fang L., Ross S., Hipel K. W: Case-based reasoning support system for conflict modeling. Systems, Man, and Cybernetics, 2000 IEEE International Conference on Volume 1, Issue, 2000, vol. 1, s. 524–530.
9. An L., Yan J., Tong L.: An Integrated Rule-based and Case-based Reasoning System for Customer Service Management. W: *e-Business Engineering* 2005.
10. Chaudhury S., Singh T. Goswami P.S.: Distributed fuzzy case based reasoning. *Applied Soft Computing Journal* 2004, Nr 4, s. 323–343.
11. Lenz M., Burkhard H.D., Bruckner S.: Applying Case Retrieval Nets for Diagnostic Tasks in Technical Domains. W: I. Smith and B. Faltings Advances in Case-Based Reasoning: Third European Workshop, Lausanne Third European Workshop, Ewibr-96, Lausanne, Switzerland, November 14 – 16, 1996.
12. Huang M.J., Chen M.Y., Lee S.C.: Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. *Expert Systems with Application* 2007, Nr 32, s. 856–867.
13. Shimazu H., Kitano H., Shibata A.: Retrieving cases from relational data-bases: Another stride towards corporate-wide case-base systems. W: Proceedings of IJCAI-93, 1993.
14. Arkills B.: LDAP Directories Explained. An Introduction and Analysis. Reading, Addison Wesley, 2004.
15. Krasuski A., Maciak T.: Rozproszone bazy danych w Państwowej Straży Pożarnej – model systemu. W: Kozielski T. i inni: Bazy danych, technologie, narzędzia. Tom 1. WKŁ, Warszawa 2005, s. 135–142.
16. Krasuski A.: Rozproszona baza danych – możliwości wykorzystania w PSP. *Przegląd Pożarniczy* 2006, Nr 5, s. 30–33.
17. Zeilenga K.: Named Subordinate References in LDAP Directories. W: RFC 3296. Reston, The Internet Society 2002.
18. Wang X. i inni: Measurement and Analysis of LDAP Performance. W: Joint International Conference on Measurement and Modeling of Computer Systems. Santa Clara, ACM Press 2000, s. 156-165.
19. Maciak T., Kreński K.: Dobór atrybutów bazy przeciwpożarowej budynków systemu informacji przestrzennej służb ratowniczych. *Zeszyty Naukowe SGSP* 2005, Nr 32, s. 59–65.
20. Feldman R., Dagan I., Hirsh H.: Mining Text Using Keyword Distributions. *Journal of Intelligent Information Systems* 1998, Nr 3, s. 281–300.
21. Witten I. i inni: Text mining in a digital library. *International Journal on Digital Libraries* 2004, Nr 1, s. 56–59.

22. Feldman R., Sanger J.: *The Text Mining Handbook*. Cambridge, Cambridge University Press, 2000.
23. Song F., Liu S., Yang J.: A comparative study on text representation schemes in text categorization Pattern. *Analysis and Applications* 2005, Nr 1-2, s. 199-209.
24. Weigend A. S., Wiener E. D., Pedersen J. O.: Exploiting Hierarchy in Text Categorization Information. Retrieval. *Information Retrieval* 1999, Nr 3, s. 193-216.
25. Domeniconi C., Gunopulos D., Ma S.: Locally adaptive metrics for clustering high dimensional data. *Data Mining and Knowledge Discovery* 2007, Nr 1, s. 63-97.
26. Bikel D. M., Schwartz R., Weischedel R. M.: An Algorithm that Learns What's in a Name. *Machine Learning* 1999, Nr 1-3, s. 211-231.
27. Yi J., Nasukawa T., Bunescu R., Niblack W.: Sentiment Analyzer: Extracting Sentiments About A Given Topic Using Natural Language Processing Techniques. W: The Third IEEE International Conference on Data Mining 2003, s.n. s. 427-34.
28. Stergos A., Vangelis K., Panagiotis S.: Summarization from medical documents. *Artificial Intelligence in Medicine* 2005, Nr 2, s. 157-77.
29. Loh S., Oliveira J. P. M., Gammeiro M. A.: Knowledge Discovery in Texts for Constructing Decision Support System. *Applied Intelligence* 2003, Nr 3, s. 357-66.
30. Praca zbiorowa: *Wikipedia: O Wikipedii*. [online]. Wikipedia. [dostęp: 12.03.2007]. [Dostępne w Internecie: <http://en.wikipedia.org/wiki/Main_Page>.

