

mgr inż. Marcin MIRONCZUK
Instytut Podstaw Informatyki PAN
Zakład Sztucznej Inteligencji
dr hab. inż. Tadeusz MACIAK, profesor SGSP
Zakład Informatyki i Łączności, SGSP

Propozycja komponentu wyszukiwania systemu CBR dla PSP opartego o ontologię dziedzinową

W artykule przedstawiono propozycję projektową komponentu wyszukiwania systemu wnioskowania na podstawie przypadków zdarzeń (ang. *case based reasoning*). Komponent ten bazuje na opracowanej ontologii dziedzinowej wspierającej proces wyszukiwania przypadków zdarzeń. Opracowana ontologia jest wynikiem m.in. przeprowadzonej przez autora analizy dokumentacji opisujących akcje ratowniczo-gaśnicze.

This paper describes the problems of designing the search module of case based reasoning system. The author proposes on ontology layer to support the searching process of case in this module. This ontology layer is a result of the conducted, by the author's analysis of the documentation describing the fire rescue operations.

Słowa kluczowe: ontologia, systemy wyszukiwania informacji, systemy CBR, system wnioskowania na podstawie przypadków, projektowanie warstwy wyszukiwania, wyszukiwanie oparte o ontologię.

Keywords: ontologies, rescue fire service ontology, retrieval systems, case-based reasoning systems, designing search layer, ontology-based search, designing search layer case-based reasoning.

1. Wstęp

W Państwowej Straży Pożarnej istnieje system ewidencjonowania zdarzeń EWIDSTAT, który służy do przeprowadzenia analiz w oparciu o informacje zgromadzone w systemie EWID99 od 1993 r. do chwili obecnej (aktualnie System Wspomagania Decyzji – ST) [1–3]. Platforma ta stanowi logiczną rozbudowę systemu EWID99 i pozwala użytkownikom na przeprowadzenie analiz statystycznych i tabel zestawieniowych, które zostały opracowane w oparciu o wytyczne Komendy Głównej PSP i zgodnie z rozporządzeniem Ministra Spraw Wewnętrznych i Administracji [4]. Ponieważ baza danych systemu zawiera informacje przydatne z punktu widzenia Kierującego Działaniami Ratowniczymi (KDR), powstała idea zaadoptowania go jako bazy wiedzy rozproszonego, katalogowego systemu wnioskowania na podstawie przypadków zdarzeń (ang. *case based reasoning system* – CBR) [5, 6].

Systemy CBR realizują, za pomocą platform informacyjnych, metodę decydowania na podstawie rozpoznania (ang. *recognition primed decision* – RPD), która jest domeną badań psychologicznych nad modelowaniem sposobu podejmowania szybkich decyzji przez człowieka w złożonych sytuacjach decyzyjnych. Metoda RPD polega na rozpoznaniu danej sytuacji, znalezieniu w pamięci podobnego przypadku, przystosowania jego rozwiązania do warunków bieżących oraz zastosowania jej. W przypadku niepowodzenia przyjętego działania zbadanie jego przyczyn, zdobycie nowej wiedzy, modyfikacja metody i jej powtórne zastosowanie [5]. W odróżnieniu od RPD, systemy CBR zamiast ludzi we wnioskowaniu wykorzystują maszyny i systemy informacyjne. System CBR składa się z takich elementów, jak [7]:

- baza danych, w której przechowywane są przypadki zdarzeń oraz metody ich rozwiązywania,
- narzędzia i rozwiązania do poszukiwania, modyfikacji (adaptacji), oceny i korygowania oraz zapisywania przypadków w bazie danych.

Cykl wnioskowania, który realizuje system CBR jest następujący [5]: zdarzenie, które należy rozwiązać, parametryzowane jest przez zbiór atrybutów. W przypadku poszukiwania sprawnych hydrantów po drodze do miejsca zdarzenia, np. pożaru lasu, mogą to być: nazwa ulicy oraz sprawność określająca, czy hydrant działa, czy też nie etc. Atrybuty te porównywane są z opisem zdarzeń przechowywanych w bazie. Następnie system wyszukuje najbardziej podobne przypadki zdarzeń w postaci np. opisów miejsc tankowania wody ze sprawnych hydrantów. Ze znalezionych przypadków ekstrahowane jest rozwiązanie w postaci nazwy ulicy, na której znajduje się sprawny hydrant. Na podstawie otrzymanego przypadku można opracować nowe rozwiązanie, np. opisać to, że hydrant nie działa. W razie takiej sytuacji, poprzez odpowiednie algorytmy adaptacji, wyznacza się rozwiązanie dla zaistniałego zdarzenia – np. wyszukanie

rozwiązania tego, jak dokonać odblokowania hydrantu lub odnalezienie w pobliżu innego. Następnie 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.

CBR mogą służyć w PSP jako systemy wspomaganie decyzji SWD dla KDR. Powstało kilka propozycji prototypów rozwiązań w postaci platform informacyjnych realizujących wnioskowanie oparte o CBR, m.in. hybrydowy system wspomaganie decyzji I i II generacji (ang. *hybrid decision support system*) czy też rozproszone usługi katalogowe dla służb ratowniczo-gaśniczych PSP [5, 6, 8–11]. Analizy przeprowadzone w trakcie ich opracowywania oraz wyniki analiz dostępnej dokumentacji z systemu EWIDSTAT wykazały, że baza tego systemu nie może być bezpośrednio zaadaptowana i traktowana jako baza przypadków systemu CBR [12]. W szczególności część nieustrukturalizowana (zmapowane pole *Dane opisowe do informacji ze zdarzenia* papierowego formularza *Informacji ze zdarzenia* [4] do systemu ewidencji) nie nadaje się do tego celu, ponieważ podczas jej przeszukiwania KDR może otrzymywać nieoczekiwane rezultaty [12]. Objawia się to na przykład tym, że zamiast listy hydrantów przy ulicy *Powstańców KDR* otrzyma listę raportów, które opisują nie tylko hydranty, ale również wszelkie inne działania ratowniczo-gaśnicze odbywające się przy tej ulicy, iż dzieje się tak, ponieważ część nieustrukturalizowana wyrażona jest niejednoznacznym językiem naturalnym. Niemniej sekcja ta jest ważna, gdyż to w niej znajduje się większość cennych informacji dla KDR, m.in. o sposobach i wskazówkach zwalczania powstałego zagrożenia czy też użytego do tego celu sprzętu. Badania przeprowadzone przez autora wykazały, że na 12 753 przebadanych segmentów (zdań) z raportów np. 37% opisuje, w jaki sposób rozwiązano powstałe zagrożenie, natomiast 17% z nich zawiera informacje o używanym sprzęcie. Z tych względów dokumentacja ta powinna stanowić podstawę przyszłego systemu CBR. W celu próby jej adaptacji autor artykułu opracował proces do strukturalizacji danych tekstowych i ekstrakcji z nich informacji [13]. Dzięki przeprowadzonym analizom i strukturalizacji sekcji *Dane opisowe do informacji ze zdarzenia* powstało też kilka odpowiednich prostych ontologii dziedzinowych oraz propozycji w oparciu o nie systemu wyszukiwania informacji w systemie CBR [14, 15].

Ontologia definiowana jest jako dziedzina filozofii zajmująca się naturą bytu, mniej ogólnie bada strukturę rzeczywistości, docieka natury wszystkiego co istnieje lub może zaistnieć, zarówno w sferze realnej (przedmioty, zdarzenia, fakty), jak i abstrakcyjnej (pojęcia, kategorie, terminy). Analizuje właściwości tych elementów i związki między nimi, a także sugeruje, jak je grupować, tworzyć hierarchie oraz dzielić według podobieństw i różnic [7, 16, 17]. W artykule przedstawiono i opisano możliwość zastosowania ontologii dziedzinowej dla PSP po-

wstałej w wyniku analizy i strukturalizacji dostępnych opisów akcji ratowniczo-gaśniczych z pola *Dane opisowe do informacji ze zdarzenia* oraz połączenia kilku innych propozycji [14]. Ontologia ta została zaadaptowana na rzecz komponentu wyszukiwania systemu CBR, który został omówiony w punkcie 2 artykułu. W podpunkcie 2.1 omówiono warstwę ontologii oraz rejestrów, natomiast w podpunkcie 2.2 przedstawiono sposób jej wykorzystania w wyszukiwaniu informacji poprzez formułowanie zapytań do odpowiednich rejestrów (także rozproszonych) i dobierania z nich odpowiedzi.

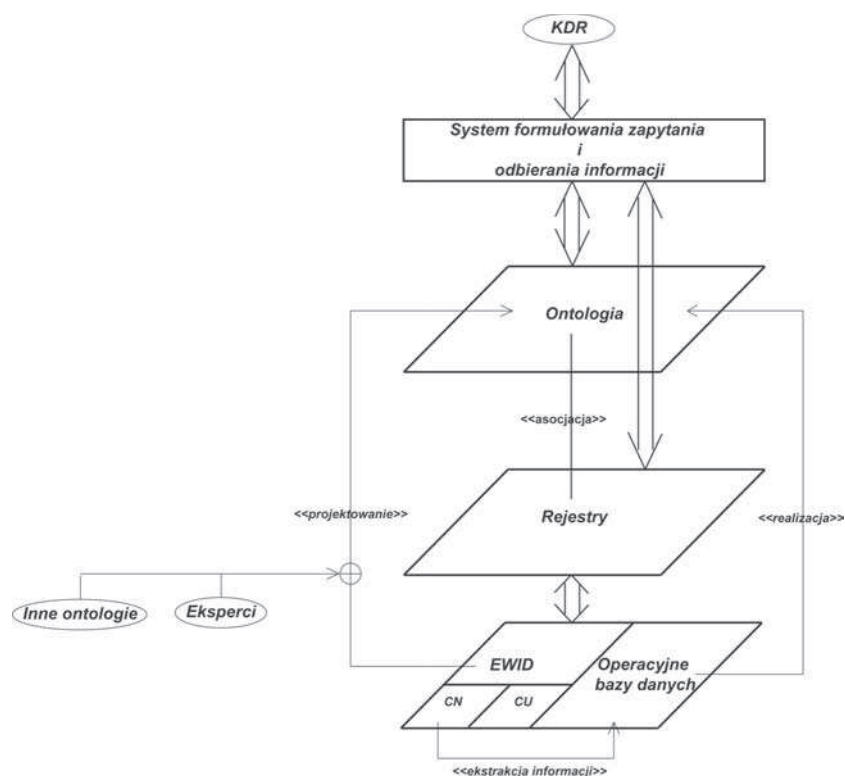
2. System CBR – projekt komponentu wyszukiwania

W niniejszym punkcie przedstawiono projekt komponentu wyszukiwania systemu CBR dla PSP. Komponent ten adaptuje warstwę ontologii, która wspomaga proces wyszukiwania oraz dotarcia do odpowiednich informacji niezbędnych dla KDR podczas prowadzonej przez niego akcji ratowniczo-gaśniczej. Informacja ta została zebrana w rejestrach w postaci operacyjnych baz danych [4]. Ontologia ta jest także mapą lokalizacji zasobów informacyjnych i ich opisem, które mogą powstać i działać w obrębie PSP. W łatwy sposób może być adoptowana do systemu rozproszonego w celu połączenia różnych źródeł informacji i zintegrowania ich w jeden logiczny system. Dla KDR system będzie widoczny jako jedna heterogeniczna baza danych. Tymczasem potrzebne mu informacje mogą być rozproszone po różnych bazach danych [5, 18–22]. Dane o odpowiednich zasobach, prawach dostępu oraz rodzaju przechowywanej w nich informacji etc., kodowane są w odpowiednich gałęziach ontologii. Z warstwy ontologii oraz ustalonych na jej podstawie rejestrów korzysta system *formułowania zapytania i odbierania informacji*, który jest pośrednikiem między KDR, a rejestrami, w których zebrana została informacja niezbędna do wspierania działań ratowniczo-gaśniczych. Projekt proponowanego przez autora komponentu wyszukiwania systemu CBR przedstawiono na rys. 1.

Projektowany komponent wyszukiwania informacji systemu CBR, który zobrazowano na rys. 1, składa się z trzech głównych warstw:

- a) rejestrów,
- b) ontologii,
- c) systemu formułowania zapytania i odbierania informacji.

Omówienie wymienionych warstw, które wchodzi w skład modułu wyszukiwania systemu CBR dla PSP, rozpoczęte zostało od najniższej warstwy, czyli rejestrów oraz ontologii (podpunkt 2.1). Następnie przedstawiono warstwę formułowania zapytania i odbierania informacji (podpunkt 2.2).



Rys. 1. Komponent wyszukiwania informacji systemu CBR wspierany ontologią dziedzinową

Źródło: opracowanie własne.

2.1. System CBR – warstwa ontologii oraz rejestrów komponentu wyszukiwania

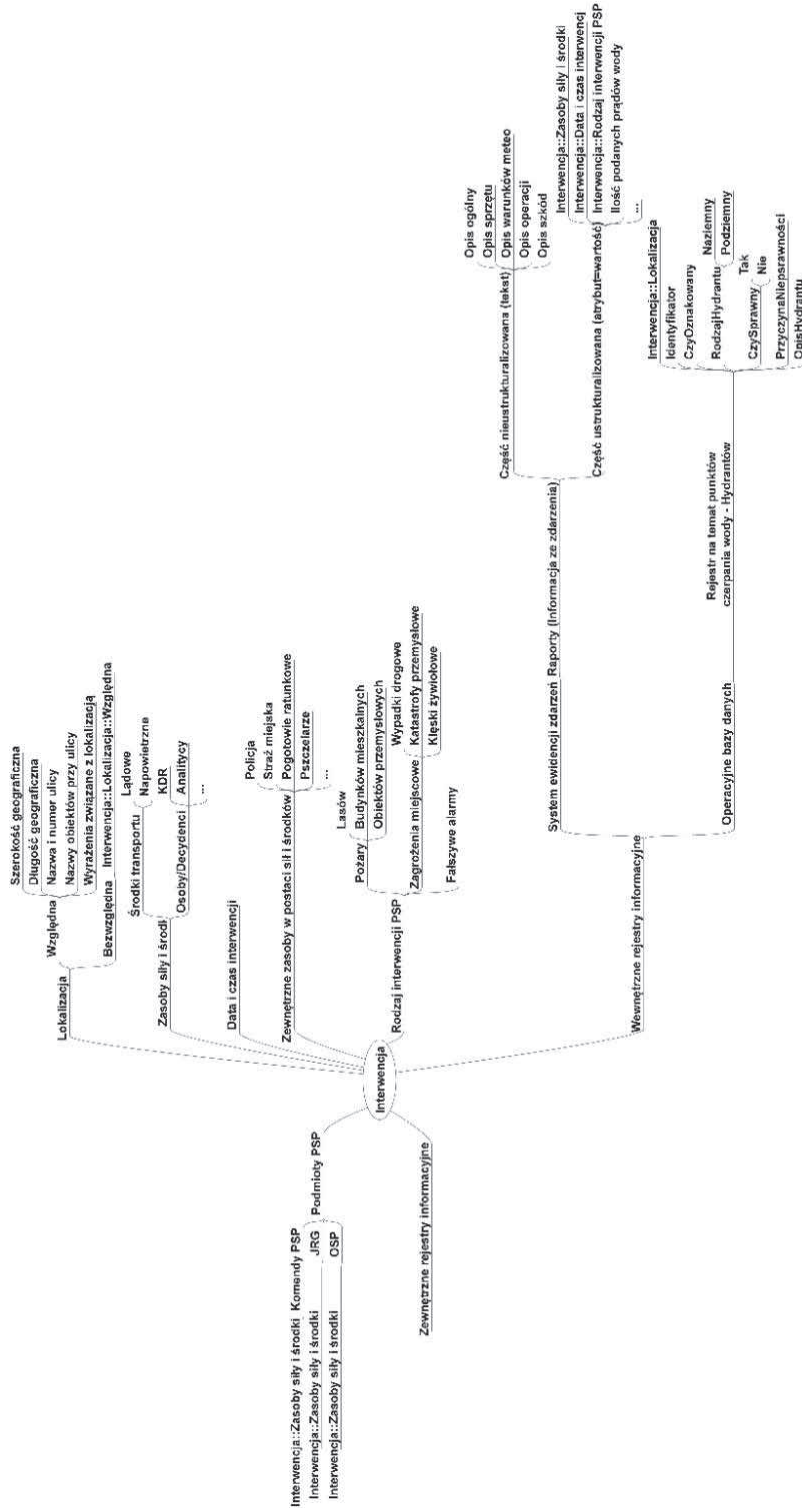
Aktualnie w PSP, jak wspomniano na wstępie, istnieje system ewidencji zdarzeń EWID. KDR po każdej akcji ratowniczo-gaśniczej sporządza w nim dokumentację ze zdarzenia w postaci formularza *Informacji ze zdarzenia* – zgodnie z formą regulowaną przez rozporządzenie [4]. System ten można podzielić na część ustrukturalizowaną i nieustrukturalizowaną. Część ustrukturalizowana (oznaczona jako CU na rys. 1) przechowuje informację w postaci atrybut-wartość, np. *data i czas interwencji = 20-10-2011 19:20*. Część nieustrukturalizowana (oznaczona na rys. 1 jako CN) w postaci pola nr 33 pt. *Dane opisowe do informacji ze zdarzenia* przechowuje informacje wyrażone za pomocą języka naturalnego (tekstu). W sekcji tej znajduje się wiele cennych informacji dotyczących m.in. tego, jak neutralizowano powstałe zagrożenie podczas akcji ratowniczo-gaśniczej, z jakich punktów czerpania wody korzystano i czy były one sprawne etc. Wykorzystanie jej bezpośrednio w systemie CBR do wspierania działań KDR jest jednak mocno ograniczone. W wyniku bezpośredniego jej zastosowania,

tj. przeszukiwania pełnotekstowego podczas akcji ratowniczo-gaśniczej przez KDR, może on dostać, jak to podkreślono we wstępie, niespodziewane i nieoczekiwane wyniki [12]. Spowodowane jest to przez niejednoznaczności wynikające z reprezentacji informacji – tekst wyrażony językiem naturalnym oraz przez to, że w sekcji tej nie zostały odwzorowane odpowiednie podpunkty z wersji papierowej formularza *Informacji ze zdarzenia*, m.in. opis działań, uszkodzony sprzęt etc. [4]. W pierwszej kolejności należało więc przeprowadzić proces strukturalizacji dostępnej dokumentacji elektronicznej, tj. jej części nieustrukturalizowanej. Podczas jego konstruowania oraz na podstawie analizy ilościowo-jakościowej tej części, w której mogą brać udział zewnątrzni eksperci, np. KDR, dołączających istniejące już ontologie (rys. 1), powstała wstępna ogólna ontologia dziedziny pt. *Interwencja*. Skonstruowana wstępna ontologia była realizacją analizy sekcji *Dane opisowe do informacji ze zdarzenia*. Przykładową skonstruowaną ontologię przedstawiono na rys. 2.

Ontologia, którą przedstawiono na rys. 2, stanowi opis podstawowych, wyselekcjonowanych pojęć oraz relacji, które między nimi zachodzą. Elementy te zostały określone na podstawie analizy jakościowej dostępnych raportów, jak i ich zawartości w postaci opisów akcji ratowniczo-gaśniczych dostępnych w sekcji *Dane opisowe do informacji ze zdarzenia*. W szczególności dotyczy to gałęzi ontologii pt. *Wewnętrzne rejestry informacyjne*, która powstała z obserwacji faktu, że w PSP obok systemu ewidencji zdarzeń EWID mogą funkcjonować inne operacyjne bazy danych, wspierające KDR w działaniach ratowniczo-gaśniczych [4]. W celu zwiększenia czytelności ontologii, którą przedstawiono na rys. 2, zastosowano notację w postaci „:”. Oznacza ona dziedziczenie wartości odpowiednich jej węzłów. Na przykład zapis *Interwencja::Lokalizacja* w węźle *Rejestr punktów czerpania wody – Hydranty* oznacza, że węzeł potomny tego węzła zawiera wszystkie pojęcia (węzły) z węzła *Lokalizacja*.

Na podstawie analiz ilościowych wykorzystujących metody i techniki sztucznej inteligencji poddano strukturalizacji dostępną dokumentację. Najpierw wydzielono podstawowe elementy semantyczne raportów, segmenty przydzielone do klas semantycznych, a następnie, za pomocą odpowiednich narzędzi, dokonano strukturalizacji wybranego kontekstu [13]. W celu zademonstrowania pierwszego etapu strukturalizacji, ekstrakcji informacji semantycznej na temat znaczenia segmentów (zdań) budujących opisy zdarzeń, założono, że dostępny jest raport w następującej postaci:

„Po przybyciu na miejsce zdarzenia stwierdzono, iż na balkonie 3 kondygnacji otwartym ogniem palą się szafki, koszyki wiklinowe, szmaty oraz okna i przylegająca elewacja. Działania polegały na podaniu dwóch prądów wody w natarciu: 1 z ziemi na balkon, 2 – prowadzony klatką schodową do mieszkania. Zniszczeniu uległy drzwi wejściowe podczas wyważania. Pomieszczenie oddymiono, miejsce zdarzenia przekazano właścicielowi ---- Samochód zatankowano przy ul Łabiszyńskiej nr 1673 – sprawny.”



Rys. 2. Przykładowa część utworzonej ontologii dziedzinnowej pt. *Interwencja powstałej w wyniku analizy formularza Informacji ze zdarzenia oraz strukturalizacji sekcji. Dane opisowe do informacji ze zdarzenia*

Źródło: opracowanie własne, na podstawie [23, 24].

Na podstawie analizy jakościowej raportów autor ustalił, że można w nich wyróżnić pięć typów klas, do których mogą należeć znajdujące się w nich segmenty. Tymi klasami były klasa *operacje*, *sprzęt*, *szkody*, *meteo* i *ogólna* [12, 14]. Informacja o klasach została zawarta w ontologii w gałęzi *Interwencja* → *Wewnętrzne rejestry informacyjne* → *System ewidencji zdarzeń* → *Część nieustrukturalizowana*. Po procesie segmentacji raportu oraz zaklasyfikowania jego poszczególnych segmentów do ww. klas otrzymywany jest półustrukturalizowany użyteczny przypadek zdarzenia [13]. Przykład takiego przypadku przedstawiono w tabeli 1.

Tabela 1. Przykładowy półustrukturalizowany raport z zaklasyfikowanymi segmentami do odpowiednich klas

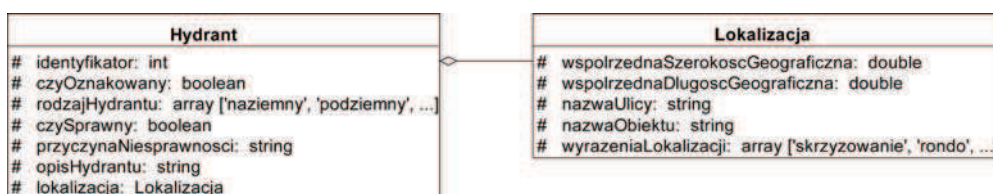
Segment	Klasa semantyczna
Po przybyciu na miejsce zdarzenia stwierdzono, iż na balkonie 3. kondygnacji otwartym ogniem palą się szafki, koszyki wiklinowe, szmaty oraz okna i przylegająca elewacja.	opis
Działania polegały na podaniu dwóch prądów wody w natarciu: 1 z ziemi na balkon, 2 – prowadzony klatką schodową do mieszkania.	operacje
Zniszczeniu uległy drzwi wejściowe (podczas wyważania). Pomieszczenie oddymiono, miejsce zdarzenia przekazano właścicielowi. -----	zniszczenia
Pomieszczenie oddymiono, miejsce zdarzenia przekazano właścicielowi. -----	operacje
Samochód zatankowano przy ul Łabiszyńskiej nr 1673 – sprawny	sprzęt

Źródło: opracowanie własne.

Drugi etap przetwarzania dokumentacji, po sklasyfikowaniu segmentów do poszczególnych klas, polegał na dalszym modelowaniu informacji, poprzez strukturalizowanie wybranej klasy semantycznej. W celu prezentacji tego etapu autor wybrał klasę *sprzęt* i opisy odnoszące się do hydrantów. Na podstawie dalszych analiz autor ustalił dla tych opisów ustrukturalizowany model, do którego w dalszej kolejności ekstrahowana była informacja na temat *Hydrantów* [25]. System ten był realizacją, w postaci rejestru, wybranej gałęzi ontologii (*Interwencja* → *Wewnętrzne rejestry informacyjne* → *Operacyjne bazy danych* → *Rejestr na temat punktów czerpania wody – Hydrantów*). Model ten, interfejs rejestru, wyrażony w notacji obiektowej w postaci klas przedstawiono na rys. 3.

Rysunek 3 przedstawia finalną wersję interfejsu modelu reprezentującego koncept *Punkt czerpania wody – Hydrant*. Interfejs ten składa się z dwóch klas *Lokalizacja* oraz *Hydrant*. Klasa *Lokalizacja* składa się z pięciu atrybutów:

- *wspolrzednaSzerokoscGeograficzna* – przechowuje informacje o szerokości geograficznej, na której znajduje się hydrant,
- *wspolrzednaDlugoscGeograficzna* – przechowuje informacje o długości geograficznej, na której znajduje się hydrant,
- *nazwaUlicy* – przechowuje informacje o nazwie ulicy przy której znajduje się hydrant (nazwa ulicy może pochodzić z konceptu *ulica*),
- *nazwaObiektu* – przechowuje informacje o obiektach znajdujących się przy danej ulicy (nazwy obiektów mogą pochodzić z konceptu *lokalizacja* z *nazwą obiektu*),
- *wyrazeniaLokalizacji* – przechowuje informacje o dodatkowych wyrażeniach związanych z lokalizacją występujących m.in. w konceptie *lokalizacja* opisana przez dodatkowy zbiór wyrażeń.



Rys. 3. Finalna wersja utworzonego interfejsu modelu konceptu:

Punkt czerpania wody – Hydrant

Źródło: opracowanie własne.

Należy zauważyć, że atrybuty *wspolrzednaSzerokoscGeograficzna* oraz *wspolrzednaDlugoscGeograficzna* nie pochodzą bezpośrednio z analizowanych tekstów. Pośrednio jednak można je uzyskać z procesu translacji nazwy ulicy, atrybutu *nazwaUlicy* na współrzędne geograficzne. Proces taki umożliwiają interfejsy programowania aplikacji (ang. *application programming interface* – *API*) dostarczane przez niektóre firmy zajmujące się mapami cyfrowymi [26]. Należy także zwrócić uwagę, że położenie hydrantu pomimo translacji będzie względne, tj. będzie odnosiło się do jakiegoś obiektu w przestrzeni, np. numeru bloku, przy którym umiejscowiony jest dany hydrant, a nie do położenia bezpośredniego samego hydrantu. Niemniej sytuacja taka nie powinna pogarszać zlokalizowania hydrantu na miejscu lub w pobliżu interwencji.

Utworzona podczas modelowania klasa *Hydrant* zawiera natomiast siedem atrybutów:

- *identyfikator* – przechowuje informacje na temat numerycznego identyfikatora hydrantu,
- *czyOznakowany* – przechowuje informacje o tym, czy hydrant jest oznakowany, a więc czy ma identyfikator,
- *rodzajHydrantu* – przechowuje informacje o rodzaju hydrantu,
- *czySprawny* – przechowuje informacje o sprawności hydrantu, a więc, czy możliwe jest z niego korzystanie,

- *przyczynaNiesprawności* – przechowuje informacje o przyczynach, z powodu których nie można było skorzystać z hydrantu,
- *opisHydrantu* przechowuje informacje o opisach hydrantu,
- *lokalizacja* – stanowi uchwyt do klasy *Lokalizacja*, która została omówiona wcześniej.

Utworzony model opisany powyższym interfejsem może być utrwalany w dowolnej bazie danych. W celu utrzymania spójności z wcześniejszymi propozycjami i badaniami innych autorów nad ich wykorzystaniem w PSP [5, 18–21, 27] preferowanym rozwiązaniem jest katalogowa baza danych. Niemniej dla rozważań autora sposób utrwalania, czyli w jaki sposób dane będą utrwalane i gdzie nie odgrywa znaczącej roli.

Dodatkowo autor zaprojektował rozwiązanie w postaci systemu ekstrakcji informacji SEI do automatycznego uzupełniania rejestru odpowiednią informacją [13, 25]. Na rys. 1 proces ten został oznaczony jako *ekstrakcja informacji*. Podczas jego realizacji z dokumentacji tekstowej wydobywana jest odpowiednia informacja i zapisywana w ustrukturalizowanym, utworzonym i opisanym powyżej rejestrze. Przykładowo niech będzie dostępny segment pochodzący z klasy *sprzet* opisujący obiekt (hydrant) w następującej postaci:

Sprawny hydrant nr 34922 ul. Szaserów 99

Z wyżej przytoczonego zdania można wyekstrahować do rozpatrywanego modelu następujące dane (atrybut, wartość, opis):

- atrybut używalność przyjmuje wartość prawdy logicznej (ang. *true*), wyrażenie sprawny w danym zdaniu sugeruje, że hydrant działa,
- atrybut numer identyfikacyjny przyjmuje wartość 34922, który występuje w danym zdaniu,
- atrybut położenie względne może przyjąć wartość strukturalną, złożoną np. z dwóch dodatkowych atrybutów w postaci nazwy ulicy o wartości *Szaserów* i jej numeru 99. Obie wartości zostały także pozyskane z prezentowanego zdania.

Jak zademonstrowano na powyższym przykładzie, ekstrakcja encji polega na rozpoznawaniu i klasyfikowaniu wykrytych wyrażen z tekstu, takich jak: *nazwy ulicy, identyfikatory, lokalizacja* etc. do utworzonego modelu.

2.2. System CBR – warstwa formułowania zapytania i odbierania informacji komponentu wyszukiwania

Ostatni element do omówienia to warstwa *formułowania zapytania i odbierania informacji* systemu wyszukiwania, która jest pośrednikiem między KDR, ontologią a rejestrami. System ten działa dwustopniowo. Na podstawie zapytania i jego analizy pod kątem występujących w nim wyrażen, przeszukiwana jest ontologia w przypadku rejestrów nieustrukturalizowanych. W przypadku rejestrów ustruk-

turalizowanych mających postać atrybut-wartość istnieje możliwość tworzenia odpowiednich graficznych interfejsów użytkownika do komunikowania się i pobierania z nich informacji [12]. Niemniej w przypadku braku takich interfejsów istnieje dalsza możliwość komunikacji z nimi poprzez niżej opisane zapytania formułowane w języku naturalnym przez KDR.

Przed przystąpieniem do dalszej analizy tej warstwy należy omówić jeden z aspektów zbudowanej i omówionej w poprzednim podpunkcie ontologii *Interwencja*. Mianowicie wybrane węzły ontologii przechowują informacje na temat rejestrów m.in. ich lokalizacji, zawartości czy też uprawnień. Informacje te mogą być ewentualnie zapisane na oddzielnej warstwie. Opis taki, zakodowany bezpośrednio w ontologii czy też na oddzielnej warstwie informacji, np. bazie XML czy katalogowej, wiąże ontologię z rejestrami. Stanowi ich mapę, dzięki której można odnaleźć potrzebne rejestry (zasoby), w których przechowywana jest odpowiednia informacja. Wybrane węzły ontologii lub warstwy pośredniej przechowują informacje na temat rejestrów w postaci klucz-wartość. Klucze i wartości opisane, są następującymi przykładowymi parametrami:

- model rejestru: [katalogowy, relacyjny, obiektowy, ...]
- rodzaj rejestru: [mysql, postgresql, mongoDB, dbo4, ...]
- użytkownik rejestru (nazwa użytkownika): [Anonim, ...]
- hasło użytkownika rejestru (hasło użytkownika): [Anonim, ...]
- adres rejestru (IP): [111.111.111.111, ...]
- połączenie do rejestru (uchwyt): [użytkownik::hasło@adres_rejestru/baza_danych, użytkownik::hasło@adres_rejestru/web_serwis, ...]
- nazwa bazy danych: [ewid, punkty czerpania wody – Hydranty, ...]
- nazwa tabeli: [tabela1, tabela2, ...]
- nazwa atrybutów: [atrybut1, atrybut2, ...]
- nazwa serwisu: [serwis hydrantów, ...]
- metody serwisu: [metoda1, metoda2, ...]

Zgodnie z powyższą listą atrybuty opisują takie aspekty rejestrów, jak ich model i rodzaj, sposób połączenia i autoryzacji czy też schematu i interfejsu, jaki udostępniają do wyszukiwania w nich danych po ustanowieniu z nimi połączenia.

Aktualnie zostanie zaprezentowane działanie warstwy *formułowania zapytania i odbierania informacji*. Po wprowadzeniu przez KDR zapytania w języku naturalnym składającego się z wyrażen opisujących kontekst wyszukiwania Q , zostaje ono dopasowane do nazw węzłów ontologii. Następnie, na podstawie danych pozyskanych z ontologii, przeszukiwane są odpowiednie rejestry. W celu zdemontowania procesu wyszukiwania rozważono dwa scenariusze. Pierwszy prosty dotyczący wyszukiwania w bazach ustrukturalizowanych, które obok przedstawianego interfejsu tekstowego, służącego do komunikacji z rejestrami, mogą zawierać interfejs graficzny (ang. *graphical user interface – GUI*). Drugi bardziej złożony pokazujący działanie wyszukiwania w sekcjach nieustrukturalizowanych.

Przykład pierwszy dotyczy wyszukiwania informacji na temat sprawnych hydrantów umiejscowionych przy ulicy Szaserów. Zapytanie z_1 wyrażone może zostać za pomocą języka naturalnego w następującej postaci:

z_1 : *sprawne hydranty ulica Szaserów*

Zapytanie z_1 składa się z czterech wyrażen $W = \{\textit{sprawne, hydranty, ulica, szaserów}\}$. W pierwszej kolejności wyrażenia z W zostaną dopasowane do węzłów ontologii. Ustalony zostanie węzeł główny, na podstawie wyrażenia hydrant z zapytania z_1 , w postaci *Punkty czerpania wody – Hydranty* pochodzący z gałęzi *Interwencja* → *Wewnętrzne rejestry informacyjne* → *operacyjne bazy danych* → *Punkty czerpania wody – Hydranty* ontologii opisaney w poprzednim punkcie 2.1. Następnie z tego węzła zostaną pobrane informacje na temat rejestru i opisujących go atrybutów, które są w tym przypadku równoważne z atrybutami, nazwami węzłów potomnych dla tego węzła ontologii tj. *lokalizacja, czyOznakowany, rodzajHydrantu, czySprawny* itd. W dalszej kolejności następuje dopasowanie pozostałych wyrażen z zapytania z_1 do pozyskanych atrybutów. Widać, że na podstawie wyrażen W można ustalić dwa atrybuty, jeden jawnie wyrażony w zapytaniu i drugi niejawni. Jawnie występującym atrybutem jest ulica (*nazwaUlity*), związana z lokalizacją. Drugi atrybut niejawni związany jest ze sprawnością hydrantu, tj. atrybutem *czySprawny*. Atrybut ten można ustalić na podstawie występującego w zapytaniu z_1 wyrażenia *sprawny*. Dodatkowo z aplikacji może pochodzić drugi niejawni atrybut związany z lokalizacją określający miasto, z którego pochodzi zapytanie w celu ograniczenia kontekstu wyszukiwania Q , np. tylko do miasta *Warszawa*. Po ustaleniu atrybutów następuje przypisanie do nich wartości. I tak atrybut *czySprawny* przyjmie wartość prawdy logicznej, wyrażenie *sprawny* oznacza, że hydrant był sprawny. Atrybut *nazwaUlity* przyjmie natomiast wartość *Szaserów*. Kolejnym krokiem jest wysłanie zapytania w postaci sparametryzowanej za pomocą ustalonych wcześniej atrybutów i ich wartości poprzez kwerendę strukturalnego języka zapytań (ang. *structural query language – sql*) lub przy wykorzystaniu metody, z np. web serwisu do rejestru/rejestrów z którymi ustanowiono połączenie. Rejestr jest przeszukiwany i wyniki tego wyszukiwania zwracane są następnie KDR. Prezentacja wyników zależy od rodzaju pozyskiwanej informacji, w omawianym przypadku prezentacja może być dokonana np. za pomocą tabeli, w której zebrane zostaną zarówno *opisy, identyfikatory*, jak i *lokalizacje* (długość i szerokość geograficzna czy też ulica z jej numerem) działających hydrantów przy ulicy *Szaserów* w *Warszawie*. Ewentualnie w tym przypadku informacja może zostać zwizualizowana bezpośrednio na mapie cyfrowej [12].

Drugi przykład prezentuje możliwości przeszukiwania rozproszonych przestrzennie rejestrów i ich sekcji nieustrukturalizowanych. Zgodnie z jedną z propozycji opisujących zastosowanie rozproszonych baz danych dla PSP, poszczególne opisy działań ratowniczo-gaśniczych, w zależności od rodzaju interwencji PSP, lokowane są w różnych rozproszonych bazach danych [22, 27]. Tak więc rodzaj interwencji PSP determinuje to, gdzie raporty zostaną zapisane.

I tak pożary lasów mogą być zapisywane w Poznaniu i Krakowie, fałszywe alarmy w Warszawie etc. Taka przykładowa strategia zapisu informacji może stanowić dobrą podstawę do zaprezentowania możliwości utworzonej przez autora ontologii oraz zaprojektowanego w oparciu o nią prezentowanego systemu wyszukiwania. W celu omówienia tego przypadku posłużono się przykładem wyszukiwania rozwiązań w postaci podobnych przypadków zdarzeń opisujących operacje oraz użyty sprzęt do gaszenia pożaru lasu. Zapytanie z_2 sformułowane przed KDR ma następującą postać:

z_2 : pożar lasu warunki meteo padający deszcz wiatr południowy

Proces przetwarzania zapytania z_2 przebiega podobnie jak przetwarzanie zapytania z_1 . Na początku, na podstawie zapytania z_2 i znajdującego się w nim wyrażenia *pożar lasu*, ustalony zostaje węzeł główny ontologii *Lasów z gałęzi Interwencja* → *Rodzaj interwencji PSP* → *Pożar*. Węzeł ten przechowuje informacje o dwóch rejestrach znajdujących się w Poznaniu oraz Krakowie zawierających dokumentację na temat pożarów lasów. Następnie z tymi rejestrami zostaje zestawione połączenie. Na podstawie wyrażenia *pożar lasu* oraz *meteo* zostają także określone rodzaje atrybutów opisujących sekcje dokumentacji półustrukturalizowanej, które zostaną przeszukane. Sekcje te znajdują się w gałęzi *Interwencja* → *Wewnętrzne rejestry informacyjne* → *System ewidencji zdarzeń* → *Część nieustrukturalizowana i noszą nazwę opis ogólny i opis warunków meteo*. Atrybutom tym zostaną przypisane wartości, *opis ogólny* przyjmie wartość *pożar lasu* natomiast *opis warunki meteo* będzie miał wartość *padający deszcz wiatr południowy*. Po przypisaniu atrybutom wartości zostaną przeszukane dwa rejestry przypadków zdarzeń w Poznaniu i Krakowie, które w swoich opisach, w polu *opis ogólny*, zawierają wyrażenia *pożar lasu* oraz w polu *opis warunków meteo* wyrażenia *deszcz wiatr południowy*. W wyniku przeszukania rejestru w Poznaniu zostanie sformułowana odpowiedź O' natomiast w przypadku rejestru znajdującego się w Krakowie odpowiedź będzie O'' . Obie odpowiedzi będą zawierać *opis sprzętu* oraz *opis operacji* wykonanych przy podobnym zdarzeniu i warunkach meteorologicznych. Odpowiedź prezentowana KDR stanowi natomiast sumę poszczególnych odpowiedzi $O = O' + O''$. Z odebranych opisów podobnych przypadków zdarzeń KDR wybiera najlepiej pasujący i go adaptuje do zaistniałej sytuacji. Wyszukiwanie to może być w razie potrzeby dodatkowo ograniczone, np. przez automatyczne ograniczenie przez system i wybór tylko tych przypadków, które zawierają odpowiedni sprzęt, tj. osiągalny dla KDR. System może sprawdzić rejestr związany z gałęzią ontologii *Interwencja* → *Siły i środki*, które są dostępne dla KDR i wybrać rozwiązania pokrywające się ze zwróconymi w odpowiedzi O . Tak więc z odpowiedzi można odrzucić opisy przypadków, które nie spełniają pewnych warunków, np. dany sprzęt jest nieosiągalny na miejscu zdarzenia.

3. Wnioski

Przedstawiona w artykule część ontologii dla służb ratowniczych PSP rozszerza dyskusję na temat możliwości zastosowania tego rodzaju rozwiązań w wybranej dziedzinie (interwencje PSP) i zaadaptowania ich w systemach rozproszonych [15]. Zaprojektowana ontologia usprawnia proces przeszukiwania nieustrukturalizowanej, częściowo strukturalizowanej, jak i w pełni ustrukturalizowanej rozproszonej informacji. Informacja ta jest niezbędna dla KDR podczas prowadzonych przez niego działań w postaci np. opisów podjętych podobnych działań na miejscu zdarzenia czy też rejestrów informacyjnych w postaci *lokalizacji punktów czerpania wody – Hydrantów*. System ten sprawdza się wtedy, gdy dziedzina oraz zachodzące w niej procesy nie podlegają pełnej strukturalizacji w postaci atrybut-wartość. Prezentowany system daje możliwość przeszukiwania za pomocą interfejsu tekstowego, zatem zapytania do systemu formułowane są w języku naturalnym. Takie podejście daje możliwość szybkiego wprowadzania informacji na temat poszukiwanego rozwiązania przypadku. W sytuacji próby tworzenia graficznego interfejsu użytkownika zawierającego pola do ograniczania wyszukiwania może dojść do sytuacji, że stanie się on nadmiarowy, nieskuteczny i nieużyteczny podczas akcji ratowniczo-gaśniczej. Przykładowo w przypadku próby zmapowania wszystkich pól z formularza *Informacji ze zdarzenia* jako atrybutów ograniczających wyszukiwanie przypadków zdarzeń może spaść użyteczność takiego rozwiązania. Mapowanie takie doprowadzi do rozbudowania i nadmiarowych funkcji. Wymieniona krata posiada ponad 33 różnego rodzaju pól. Próba wypełnienia takiego interfejsu podczas akcji ratowniczo-gaśniczej przez KDR byłaby bezcelowa i nieakceptowalna. Ewentualnym problemem w przypadku zastosowania interfejsu tekstowego i formułowania zapytania w postaci języka naturalnego może być związany z odnalezieniem kontekstu wyszukiwania oraz przetwarzaniem samego zapytania. Sprawy te są związane z implementacją analizatora składniowego (parsera) zapytania, który jest zawarty w *systemie formułowania zapytania i odbierania informacji*. W zapytaniu należy odpowiednio wykrywać atrybuty i dopasowywać je do węzłów z ontologii oraz dowiązywać do nich wartości. Problemem może okazać się np. odmiana wyrażen z zapytania i translacja ich na nazwy węzłów ontologii. KDR może chcieć np. wyszukać *hydranty przy ulicy Mickiewicza*. Widać, że w ontologii nie ma węzła o nazwie *hydranty* czy *ulicy* a więc odpowiednie węzły nie zostaną odnalezione. Problem ten jednak można rozwiązać za pomocą wstępnego przetwarzania zapytania poprzez zastosowanie procesu lematyzacji i zabiegów semantycznych, sprowadzających wyrażenia do pierwszej osoby liczby pojedynczej czy też usuwania zbędnych wyrażen ze stop listy [28]. Po takich zabiegach przykładowe zapytanie przyjmie następującą postać: *hydrant ulica Mickiewicza*. Po takim przetworzeniu zapytania nie będzie problemu z odnalezieniem odpowiednich węzłów ontologii i z powiązaniem atrybutów z ich wartościami.

Przedstawione rozwiązanie rozszerza także omówienie aspektów związanych z zastosowaniem i zarządzaniem rozproszonymi bazami danych. Nie ogranicza się ono tylko do rozwiązań katalogowych, wprowadza także możliwość korzystania z rejestrów które dostarczają odpowiednie interfejsy zewnętrzne dostępne poprzez np. web serwisy. Proponowane rozwiązanie nie jest związane bezpośrednio ze stosowaną technologią przechowywania i zarządzania danymi w wybranych węzłach [29, 30].

Opisana część wyszukiwania klasuje propozycję systemu CBR opartego o ontologię w rozwiązaniach mieszanych [31, 32], tj. w takich, w których nie wszystkie aspekty dziedziny da się sparametryzować, wyrazić w sposób strukturalny lub aktualnie nie są one sparametryzowane przez zapis np. atrybut-wartość budujących ramy, klasy etc. [7, 16, 33]. Część, a nawet większość informacji, wiedzy dostępna jest nadal w postaci tekstowej w systemie ewidencji zdarzeń. Prezentowana ontologia wiąże obie te reprezentacje, które od strony użytkownika KDR widziane są jako jeden logiczny, monolityczny system.

Na zakończenie przedstawiono propozycję uzupełniania i rozszerzania aktualnie skonstruowanej ontologii. Autor rozważył i opracował w swych dotychczasowych badaniach możliwość zastosowania formalnej analizy pojęć (ang. *formal concept analysis – FCA*), na podstawie której zaprojektował prezentowane w artykule rozwiązanie służące do strukturalizacji dokumentacji [25]. Niemniej istnieje także alternatywny sposób uzupełniania ontologii, który może stanowić nowy kierunek badań. Propozycja rozbudowy ontologii systemu CBR dla PSP miałaby się opierać na analizie częstotliwości lub powiązania ze sobą wyrażen występujących w wyselekcjonowanych sekcjach. Analizy takie tworzą najczęściej unigramy, bigramy, trigramy a ogólnie n -gramy lub grupują dokumenty ze względu na podobieństwo semantyczne za pomocą np. ukrytej analizy semantycznej (ang. *latent semantic analysis – LSA*) [28, 34–36]. W taki sposób mogą zostać wykryte np. częste wyrażenia w postaci *helikopter, wóz bojowy, śmigłowiec, do karetki* lub mogą one stanowić składowe o dużej wadze rozróżniające grupę segmentów opisujących *sprzęt* od innych grup [37]. Obserwacje te mogą prowadzić do powstania nowej, rozbudowania aktualnej gałęzi ontologii w postaci *środku transportu* lub uzupełnienie jej o nowe nieznanne atrybuty. W ten sposób mogą powstawać różne gałęzie ontologii dziedzinowej lub nawet ich całe katalogi, które będzie można adaptować w zależności od rozwiązywanego zadania.

Piśmiennictwo

- [1] Abakus: System EWID99 [on-line] [dostęp: 1 maja 2009]. Dostępny w Internecie: http://www.ewid.pl/?set=rozw_ewid&gr=roz.
- [2] Abakus: System EWIDSTAT [on-line] [dostęp: 1 maja 2009]. Dostępny w Internecie: <http://www.ewid.pl/?set=ewidstat&gr=prod>.
- [3] Strona firmy abakus [on-line] [dostęp: 1 marca 2009]. Dostępny w Internecie: <http://www.ewid.pl/?set=main&gr=aba>.
- [4] Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z 29 grudnia 1999 r. w sprawie szczegółowych zasad organizacji krajowego systemu ratowniczo-gaśniczego. DzU 99.111.1311 § 34 pkt. 5 i 6.
- [5] Krasuski A., Maciak T.: Wykorzystanie rozproszonej bazy danych oraz wnioskowania na podstawie przypadków w procesach decyzyjnych Państwowej Straży Pożarnej. *Zeszyty Naukowe SGSP* 2008, nr 36, s. 17–35.
- [6] Krasuski A., Maciak T., Kreński K.: Decision Support System for Fire Service based on Distributed Database and Case-based Reasoning. *Studies of logic grammar and rethoric* 2008, nr 11.
- [7] Traczyk W.: Wiedza strukturalna. Inżynieria Wiedzy. Akademicka Oficyna Wydawnicza EXIT, Warszawa 2010.
- [8] Mirończuk M., Kreński K.: Koncepcja systemu ekspertowego do wspomaganie decyzji w Państwowej Straży Pożarnej. w: A. Grzech, K. Juszczyń, H. Kwaśnicka, N.T. Nguyen (red.) Inżynieria Wiedzy i Systemy Ekspertowe. Akademicka Oficyna Wydawnicza EXIT, Warszawa 2009.
- [9] Mirończuk M., Maciak T.: Problematyka projektowania modelu hybrydowego systemu wspomaganie decyzji dla Państwowej Straży Pożarnej. *Zeszyty Naukowe SGSP* 2009, nr 39.
- [10] Mirończuk M., Maciak T.: Wykorzystanie komponentów platformy ERP do budowy drugiej generacji hybrydowego systemu wspomaganie decyzji dla PSP. *Metody Informatyki Stosowanej* 2011, nr 3.
- [11] Krasuski A., Kreński K.: Building a DSS for Fire Service using COLIBRI. *Analele Universitatii Bucuresti, Informatica LVI*, 2008.
- [12] Mirończuk M.: Zmodyfikowana analiza FMEA z elementami SFTA w projektowaniu systemu wyszukiwania informacji na temat obiektów hydrotechnicznych w nierelacyjnym katalogowym rejestrze. *Studia Informatica* 2011, nr 2.
- [13] Mirończuk M., Maciak T.: Proces i metody eksploracji danych tekstowych do przetwarzania raportów z akcji ratowniczo-gaśniczych. *Metody Informatyki Stosowanej* 2011, nr 4.
- [14] Mirończuk M., Maciak T.: Propozycja mieszanego przetwarzania półstrukturalnego modelu opisu zdarzeń z akcji ratowniczo-gaśniczych Państwowej Straży Pożarnej PSP. *CNBOP Bezpieczeństwo i Technika Pożarnicza* (w cyklu recenzyjnym).

- [15] Kreński K., Krasuski A.: The foundations for an ontology-based knowledge representation layer for a CBR system in fire service. *Analele Universitatii Bucuresti, Informatica LVI*, 2008.
- [16] Goczyła K.: *Ontologie w systemach informatycznych*. Akademicka Oficyna Wydawnicza EXIT, Warszawa 2011.
- [17] Gliński W.: *Ontologie. Próba uporządkowania terminologicznego chaosu*. Instytut Informacji Naukowej i Studiów Bibliologicznych UW [dostęp: 10 sierpnia 2010]. Dostępny w Internecie: <http://bbc.uw.edu.pl/Content/20/13.pdf>.
- [18] Krasuski A.: Rozproszona baza danych – możliwości wykorzystania w PSP. *Przegląd Pożarniczy* 2006, nr 5, s. 30–33.
- [19] Krasuski A.: Usługi katalogowe w architekturze rozproszonej bazy danych w procesie wspomaganie decyzji w Państwowej Straży Pożarnej. *Seminarium Zakładu Logiki Matematycznej*, 2009.
- [20] Krasuski A.: Możliwość wykorzystania usługi katalogowej w architekturze rozproszonej bazy danych jako podstawy systemu treningu i wspomaganie decyzji w Państwowej Straży Pożarnej. *Wydział Informatyki. Politechnika Białostocka, Białystok* 2010.
- [21] Krasuski A., Maciak T.: Rozproszone bazy danych – architektura funkcjonalna. *Bezpieczeństwo i Technika Pożarnicza* 2007, nr 1.
- [22] Krasuski A., Maciak T.: Rozproszone bazy danych, możliwości ich wykorzystania w Państwowej Straży Pożarnej. *Zeszyty Naukowe SGSP* 2006, nr 34, s. 23–42.
- [23] FreeMind – free mind mapping software [dostęp: 12 lipca 2011]. Dostępny w Internecie: http://freemind.sourceforge.net/wiki/index.php/Main_Page.
- [24] Foltin C.: FreeMind. *sourceforge.net*, 2000.
- [25] Mirończuk M.: Wykorzystanie FCA do konstruowania reprezentacji wiedzy w postaci reguł wnioskowania dla systemu ekstrakcji informacji inteligentnego społecznego systemu informacyjnego. *Biuletyn WAT* 2011 (w cyklu recenzyjnym).
- [26] Google Maps API Family [dostęp: 20 września 2011]. Dostępny w Internecie: <http://code.google.com/intl/pl-PL/apis/maps/index.html>.
- [27] Krasuski A., Maciak T.: Rozproszone bazy danych w Państwowej Straży Pożarnej – model systemu. w: Kozielski T. (red.). *Bazy danych, technologie, narzędzia*. WKŁ, Warszawa 2005, s. 135-142.
- [28] Mirończuk M.: Przegląd metod i technik eksploracji danych tekstowych. *Studia i Materiały Informatyki Stosowanej SIMIS* 2011, nr 2 (w cyklu recenzyjnym).
- [29] Krasuski A., Maciak T.: Historia rozwoju systemów zarządzania bazami danych. *CNBOP Bezpieczeństwo i Technika Pożarnicza* 2006, s. 213–226.
- [30] Mirończuk M.: Systemy zarządzania bazą danych i architektura agentowa w służbach ratowniczych Państwowej Straży Pożarnej. *CNBOP Bezpieczeństwo i Technika Pożarnicza* 2011, nr 1.

- [31] Kempa A.: Zastosowanie rozszerzonej metodologii wnioskowania na podstawie przypadków – textual CBR w pracy z dokumentami tekstowymi. Systemy Wspomagania Organizacji, Katowice 2005 [dostęp: 1 stycznia 2008]. Dostępny w Internecie: <http://www.swo.ae.katowice.pl/content/view/221/32/>.
- [32] Weber R., Aha D.W., Sandhu N., Munoz-Avila H.: A Textual Case-Based Reasoning Framework for Knowledge Management Applications In proceedings of the ninth german workshop on case-based reasoning shaker verlag. Shaker Verlag, Aachen 2001, s. 244–253.
- [33] Meyer B.: Programowanie zorientowane obiektowo. Warszawa 2005.
- [34] Mykowiecka A.: Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym. PJWSTK, Warszawa 2007.
- [35] Deerwester S., Dumais T., Furnas G.W., Landauer T.K., Harshman R.: Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 1990, nr 41, s. 391–407.
- [36] Matsuo Y., Ishizuka M.: Keyword Extraction From a Single Document Using Word Co-Occurrence Statistical Information. *International Journal on Artificial Intelligence Tools* 2004, nr 13, s. 157–169.
- [37] Kreński K., Krasuski A., Łazowy S.: Data mining and shallow text analysis for the data of State Fire Service. Pułtusk, Poland: Concurrency, Specification and Programming XXth International Workshop, CS&P, September 28–30.

Summary

Marcin MIRONCZUK
Tadeusz MACIAK

Proposed Searches Component of the CBR for the State Fire Service Based on the Domain Ontology

This paper describes problems of designing the search module of case based reasoning system. In the first part of this article the author describes a review of solutions available in fire service such as decision support system with implements reasoning solution – case based reasoning CBR. Second part of this article describes a search component of this CBR system. The author proposes on ontology layer to support the searching process of case in this module. This ontology layer is a result of the conducted, by the author's analysis of the documentation describing the rescue actions. In the last section the author summarizes proposed project of this component and presents, a new developed way to construct, refactorize and extend the proposed ontology.