

mgr inż. **Marcin Michał MIROŃCZUK**¹
dr hab. inż. **Tadeusz MACIAK**²

PROPOZYCJA MIESZANEGO PRZETWARZANIA PÓLSTRUKTURALNEGO MODELU OPISU ZDARZEŃ Z AKCJI RATOWNICZO-GAŚNICZYCH PAŃSTWOWEJ STRAŻY POŻARNEJ PSP³

Proposition of hybrid process model semi structured description of event from fire services rescues operation

Streszczenie

W opracowaniu przedstawiono aktualnie rozwijane reprezentacje wiedzy i sposoby opisów zdarzeń, dla systemu wnioskowania na podstawie przypadków zdarzeń służb ratowniczych Państwowej Straży Pożarnej PSP. W artykule zaproponowano sposób ich przetwarzania. Przedstawiony sposób bazuje na klasyfikacji i wyszukiwaniu opisów zdarzeń.

Summary

This paper describes a review of actual developed knowledge representation and case representation for fire services cases based reasoning system. The article also describes a method of processing the cases of events. This processing method based on classification and information retrieval.

Słowa kluczowe: klasyfikator Bayesa, naiwny klasyfikator Bayesa, eksploracja tekstu, reprezentacja tekstu, reprezentacja meldunków, wnioskowanie na podstawie przypadków, reprezentacja przypadków zdarzeń, ontologia służb ratowniczych;
Keywords: Bayes Classifier, Naive Bayes Classifier, text mining, text representation, representation of reports, case-based reasoning, ontology for rescue service;

1. Wprowadzenie

W Państwowej Straży Pożarnej PSP forma powstających raportów sporządzanych po każdej akcji ratowniczo-gaśniczej jest regulowana przez Rozporządzenie Ministra Spraw Wewnętrznych [1]. Na podstawie tego rozporządzenia utworzona została w formie papierowej karta *Informacji ze zdarzenia*. Stanowi ona raport ze zdarzenia z akcji ratowniczo-gaśniczej i jest częściowo ustrukturyzowana. Częściowa strukturyzacja polega na tym, że istnieje możliwość wprowadzenia i sprawdzenia informacji o takich elementach akcji ratowniczo-gaśniczej, jak np.: czas zdarzenia, czas działań ratowniczych, rodzaj prowadzonych działań, rodzaj użytego sprzętu, miejsce prowadzonych działań, dane o budynku lub pomieszczeniu, w którym powstało zdarzenie, etc. Kierujący Działaniami Ratowniczymi (KDR),

po każdej akcji wypełnia taki raport, umieszczając w nim odpowiednie informacje dotyczące podjętych działań. KDR ma także możliwość wprowadzenia dodatkowych informacji dotyczących zdarzenia, które nie zostały uwzględnione w karcie, do sekcji zatytułowanej – *Dane opisowe do informacji ze zdarzenia*. Sekcja ta podzielona jest na sześć podpunktów: opis przebiegu działań ratowniczych (zagrożenia i utrudnienia, zużyty i uszkodzony sprzęt), opis jednostek przybyłych na miejsce zdarzenia, opis tego, co uległo zniszczeniu lub spaleni, warunki atmosferyczne, wnioski i uwagi wynikające z przebiegu działań ratowniczych oraz inne uwagi dotyczące danych z pierwszej części formularza. Ze względu na to, że zawartość poszczególnych podpunktów tej sekcji jest wyrażona za pomocą języka naturalnego w postaci zdań, na które składają się słowa oraz frazy, została ona nazwana *częścią półustrukturyzowaną*.

Na bazie poszczególnych przypadków w Komendach Wojewódzkich PSP wykonywane są analizy wybranych zdarzeń i składowane w postaci pa-

¹ Instytut Podstaw Informatyki PAN, Zespół Podstaw Sztucznej Inteligencji

² Politechnika Białostocka, Wydział Informatyki

³ Wkład procentowy autorów w powstanie artykułu wyniósł 80% – M. Mirończuk i 20% – T. Maciak.

pierowej. W Komendzie Głównej PSP specjaliści analizują meldunki pod kątem określonych strategicznych zapytań. Przechowywane w PSP opisy przypadków oraz analizy zdarzeń w informacyjnym systemie ewidencji zdarzeń EWID [2-4] i w komendach wojewódzkich, są też dokumentami tylko częściowo ustrukturyzowanymi i nie nadają się bezpośrednio do przetwarzania komputerowego. Częściowa strukturyzacja wynika z tego, iż sekcje oraz pola z *Karty informacji ze zdarzenia* są mapowane i przedstawiane w postaci relacji oraz odpowiednich typów danych. Jednak w dalszym ciągu sekcja *Dane opisowe do informacji ze zdarzenia* jest reprezentowana za pomocą tekstu opisanego językiem naturalnym. Numeryczną analizę tej części raportu utrudnia fakt, że sześć wcześniej wymienionych podpunktów – składających się na tą sekcję w wersji papierowej – w systemie informacyjnym ewidencji zdarzeń zostaje przedstawionych jako pojedynczy rekord danych bez zachowania należytego podziału. Z tego względu ta cyfrowa sekcja stanowi *część nieustrukturyzowaną*. Ewentualne pozyskanie z niej informacji dla KDR jest kłopotliwe, a samo przekształcenie jej do użytecznych przypadków zdarzeń systemu wnioskowania na podstawie zdarzeń (*ang. case-based reasoning – CBR*) [5] wymaga zastosowania wielu zabiegów semantycznych. Termin *użyteczne przypadki zdarzeń* określa taki zbiór przypadków zdarzeń, które w sposób czytelny i klarowny dostarczą KDR niezbędnej wiedzy o zaistniałym zdarzeniu. Wiedza ta ma dotyczyć problemów i zagrożeń, jakie niesie ze sobą powstałe zdarzenie, wskazówek na co należy uważać, prowadząc działania, oraz możliwych rezultatów powstałych na skutek wyboru danej strategii likwidacji zagrożenia i realizacji działań ratowniczych. Wiedza ta ma być pozyskiwana na podstawie analiz podobnych przypadków z przeszłości zawartych w bazie wiedzy systemu CBR.

W wyniku tego, że analizy zdarzeń wykonywane są przez różne osoby, które definiują i opisują zdarzenie według własnego postrzegania i za pomocą innego słownictwa, powstaje pewnego rodzaju problem semantyczny. Powoduje to, iż do określenia tych samych zdarzeń stosowane są różne nazwy. Badania wykazują, że przy opisywaniu jednego zagadnienia jedynie 20% badanych posługuje się tym samym słownictwem [6]. Zależność ta nie zmienia się znacząco bez względu na to, czy badanymi są eksperci w danej dziedzinie, czy też mniej doświadczone osoby.

PSP nie stworzyła do tej pory standardowego, ujednoliconego, szerokiego słownika zawierającego pojęcia z zakresu ratownictwa, który definiowałby zachodzące między nimi relacje oraz stanowiłby ontologię dla służb ratowniczych. Słownik taki posłużyłby do utworzenia precyzyjniejszej, homogenicznej komunikacji i wymiany wiedzy na temat zdarzeń z zakresu ratownictwa w obrębie PSP. Ponadto róż-

norodna interpretacja i opis podobnych wypadków powodują, iż pozyskiwanie informacji z tego typu dokumentów tj. sekcji opisowej systemu EWID i transformacja ich bezpośrednio do ustandaryzowanego, użytecznego opisu przypadków zdarzeń w sensie systemu CBR, nie jest do końca możliwa i wymaga zastosowania technik z zakresu komputerowej analizy tekstu.

Aktualnie prowadzone są badania zmierzające do usystematyzowania wiedzy w obrębie działań ratowniczo-gaśniczych przeprowadzanych przez służby ratownicze PSP. Pierwszy kierunek tych badań stanowią projekty nad zastosowaniem rozproszonego wnioskowania przy użyciu systemu wnioskowania na podstawie przypadków zdarzeń [5, 7, 8] czy też bardziej kompleksowe rozwiązania takie, jak hybrydowy system wspomagania decyzji HSWD [9, 10]. Skupiają się one bardziej zarówno na architekturze i komponentach samego systemu, procesach oraz metodach jego projektowania, jak i badaniu jego wydajności. Drugi nurt badań dotyczy sposobu pozyskiwania, budowania przetwarzania wiedzy w samym systemie. Dotyczą więc one zagadnień związanych z warstwą nośną tj. z reprezentacją i sposobem wykorzystania wiedzy. Dodatkowo w tym obszarze można wyróżnić dwa trendy. Pierwszy z nich zajmuje się modelowaniem i tworzeniem ontologii dla badanej dziedziny. Drugi natomiast dotyczy analiz nieustrukturyzowanych raportów z sekcji opisowej systemu informacyjnego EWID w celu dostarczenia dodatkowych informacji do budowy ontologii, jak również budowy samych raportów – *użytecznych przypadków zdarzeń* systemu CBR. Transformacja raportów z akcji ratowniczo-gaśniczych zawartych w systemie informacyjnym EWID – które stanowią nieprzetworzony w żaden sposób ciąg zdań bez podziału na ww. sekcje, w półstrukturalne oraz strukturalne przypadki zdarzeń ma odbywać się w sposób automatyczny w celu otrzymywania półstruktur oraz półautomatyczny do otrzymywania pełnych struktur. *Przypadek zdarzenia półstrukturalny* definiowany jest przez autorów jako częściowo ustrukturyzowana informacja z podziałem na sekcje uzyskane w procesie klasyfikacji. Jego półstrukturalność polega na tym, że jest on wyrażony w postaci zdań języka naturalnego zorganizowanego w sekcje. Organizacja taka określona została przez autorów jako wiedza, z tego względu, że stosując opis informacji wyrażony w postaci np. trójki *<opis, zasoby, strateg>* na temat zagrożenia z ontologii akcji ratowniczo-gaśniczych, KDR uzyskuje potrzebne oraz wartościowe dane i wskazówki na temat tego, na co należy uważać i co w przeszłości sprawiało problemy, jak je likwidowano za pomocą dostępnych sił i środków oraz jakie zostaną poniesione straty/koszty w wyniku wybranej strategii. *Przypadek zdarzenia w pełni ustrukturyzowany* definiowany jest przez autorów jako informacja w postaci hierarchii klas

w notacji obiektowej, uzyskana w procesie formalnej analizy pojęć (*ang. formal concept analysis, FCA*) przeprowadzonej na sekcji półstrukturalnej. W pewnym sensie jest to prosta odmiana ontologii, przez co może służyć do rozszerzania i uzupełniania obecnie tworzonej ontologii o nowe elementy istotne z punktu widzenia dziedziny, dla której jest tworzona – opisu akcji ratowniczych PSP.

Ogólnie badania nad reprezentacją wiedzy można podzielić na te, które skupiają się na tworzeniu modelu opisu pojedynczego przypadku zdarzenia oraz na te, które tworzą całościowy model opisu zdarzeń akcji ratowniczo-gaśniczych. Pierwszy model tożsamy jest z przypadkiem zdarzenia lub reprezentacją przypadku zdarzenia i dotyczy on tworzenia szablonu (reprezentacji) tego zdarzenia. Drugi model składa się z prostego modelu hierarchicznego lub ontologicznego *Akcje*, zawierającego model pojedynczego przypadku zdarzenia. Zawiera on kolekcję przypadków, umieszczonych w odpowiednich węzłach ontologii. Użycie w tytule artykułu terminu *półstrukturalny model opisu zdarzeń z akcji ratowniczo-gaśniczych* wynika z tego, że do dostępnego prostego modelu hierarchicznego interwencji PSP lub bardziej złożonego wyrażonego w postaci ontologii można dołączyć półstrukturalny przypadek zdarzenia. Istotną kwestią w tak mieszanej reprezentacji staje się zagadnienie związane z klasyfikacją nowego przypadku zdarzenia oraz wyszukiwaniem informacji na temat podobnych przypadków zdarzeń.

W niniejszym artykule opisano hybrydową metodę opierającą się na zastosowaniu klasyfikatora Bayesa do klasyfikacji raportów ze zdarzeń do odpowiedniego węzła ontologii oraz zastosowaniu binarnych miar do wyszukiwania przypadków zdarzeń. W punkcie 2, 2.1 artykułu zostały przedstawione i omówione istniejące i rozwijane hierarchie oraz ontologie dla służb ratowniczych PSP jako warstwy nośne wiedzy w systemie CBR. W podpunkcie 2.2 zaproponowano i zaprezentowano półstrukturalną reprezentację opisu przypadku zdarzenia za pomocą rozszerzalnego języka znaczników (*ang. extensible markup language – XML*). Następnie w podpunkcie 2.3 pokazano, w jaki sposób

można połączyć ontologię *Akcje*, opisującą akcje ratowniczo-gaśnicze PSP, z półstrukturalną reprezentacją przypadku zdarzenia oraz pokazano sposób etykietowania przypadków zdarzeń do klas (węzłów) ontologii. W punkcie 3. omówiono podstawy teoretyczne klasyfikacji. W dalszej kolejności w punkcie 4. przedstawiono proces klasyfikacji nowego przypadku zdarzenia do *liścia* ontologii w postaci klasy niemającej więcej rozgałęzień. W punkcie 5. ze względu na zastosowanie binarnej wersji Bayesa oraz binarnego indeksowania przypadków zdarzeń przedstawiono miary binarne służące do wyszukiwania binarnego. W punkcie 6. opisano propozycję mieszanej metody do klasyfikowania i przeszukiwania tekstowej bazy przypadków ze zdarzeń (akcji ratowniczo-gaśniczych) dołączonych do ontologii *Akcje*. Na zakończenie w punkcie 7. przedstawiono kierunki rozwoju prowadzonych badań oraz wnioski z opisanych w artykule zagadnień.

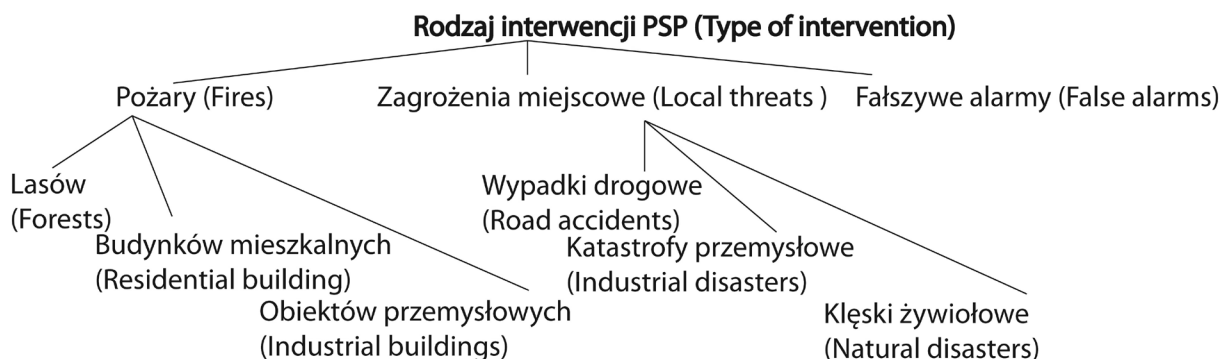
2. Ontologia oraz przypadki zdarzeń – reprezentacja i opis

W podpunkcie 2.1 niniejszego punktu opisano reprezentację podziału akcji ratowniczo gaśniczych w postaci ontologii *Akcje*. W podpunkcie 2.2 przedstawiono półstrukturalny przypadek zdarzenia. Natomiast w ostatnim podpunkcie 2.3 zaprezentowano możliwość dołączania takiego przypadku zdarzenia do wybranej gałęzi ontologii *Akcje*.

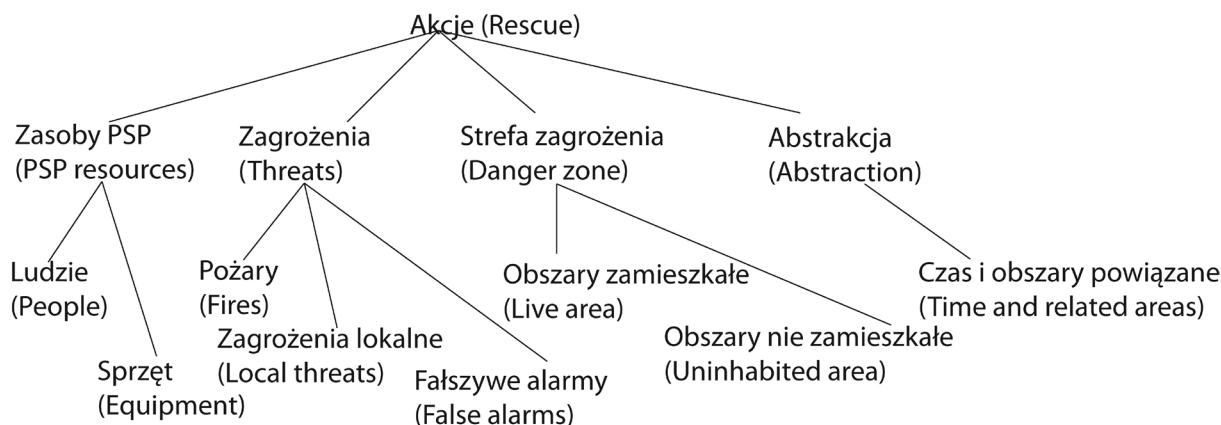
2.1. Reprezentacja i opis podziału akcji ratowniczo-gaśniczych – ontologia *Akcje*

Aktualnie na potrzeby obsługi zdarzeń, w postaci akcji ratowniczo-gaśniczych przeprowadzanych przez Państwową Straż Pożarną, powstała koncepcja, aby opisy zdarzeń utrzymywać w rozproszonym systemie CBR, który stanowi podsystem do wnioskowania w HSWD. Warstwa nośna danych i modelowanie opisów akcji ratowniczo-gaśniczych w systemie CBR ma mieć prostą postać hierarchiczną lub złożoną ontologiczną [11]. Obie postacie prezentują kolejno Ryc.1 i Ryc. 2.

Ryc. 1 prezentująca prostą postać hierarchiczną zawiera jedynie podział interwencji dokonywa-



Ryc. 1. Fragment podziału interwencji PSP – prosta postać hierarchiczna [5]
Fig. 1. Fragment of the division PSP interventions – simple hierarchical form [5]



Ryc. 2. Ontologia Akcje [11]

Fig. 2. Rescue ontology [11]

nych przez służby ratownicze PSP według aktualnego rozporządzenia [1]. Rozszerzoną jego wersją jest reprezentacja ontologiczna *Akcje*, która zawiera, rozszerza i modyfikuje ww. reprezentację hierarchiczną.

Ryc. 2 prezentuje propozycję ontologii *Akcji* zawierającą w sobie klasę *Zagrożenia* z dotychczasowym podziałem interwencji PSP. Dodatkowymi klasami, które pojawiają się w reprezentacji *Akcji*, są: *Zasoby PSP*, *Strefa zagrożenia* oraz *Abstrakcja*. Bliższe szczegóły na temat ontologii i jej tworzenia dla służb ratowniczych można znaleźć w opracowaniach [11].

2.2. Reprezentacja i opis pojedynczego półstrukturalnego przypadku zdarzenia

Pojedynczy półstrukturalny przypadek zdarzenia z akcji ratowniczo-gaśniczej można zaprezentować za pomocą hierarchii, np. wykorzystując do tego rozszerzalny język znaczników, a poszczególne wybrane węzły hierarchii opisać językiem naturalnym (tekstem). Przykład szablonu do opisu przypadku zdarzenia z akcji ratowniczo-gaśniczej przedstawia przypadek zdarzenia 1 zaprezentowany poniżej.

Przypadek zdarzenia 1 Szablon opisu przypadku zdarzenia z akcji ratowniczo-gaśniczej

The case of event 1 Template description of the rescue events

```

<?xml version="1.0" encoding="UTF-8"?>
<przypadekZdarzenia id="identyfikator przypadku">
  <sekcja>
    <opisowa>
      <ogólna>
        <opis>[opis]</opis>
        <zagrożenia><opis>[opis]</opis></zagrożenia>
        <utrudnienia><opis>[opis]</opis></utrudnienia>
      </ogólna>
      <przyczyny>
        <opis>[opis]</opis>
      </przyczyny>
      <szkody>
        <opis>[opis]</opis>
      </szkody>
    </opisowa>
  </sekcja>
</przypadekZdarzenia>
  
```

```

</szkody>
</wskazówki>
  <opis>[opis]</opis>
  <punktyCzerpaniaWody>
    <opis>[opis]</opis>
  </punktyCzerpaniaWody>
</wskazówki>
<działania>
  <opis>[opis]</opis>
</działania>
<zasoby>
  <opis>[opis]</opis>
  <zużyte><opis>[opis]</opis></uszkodzone>
  <uszkodzone><opis>[opis]</opis></uszkodzone>
</zasoby>
<warunkiAtmosferyczne>
  <opis>[opis]</opis>
</warunkiAtmosferyczne>
</opisowa>
</sekcja>
</przypadekZdarzenia>
  
```

Szablon przypadku zdarzenia z akcji ratowniczo-gaśniczej, który można przechowywać w systemie informacyjnym, powstał na bazie karty *Informacji ze zdarzenia*, której format regulowany jest przez rozporządzenia [1]. Autorzy proponują zmienioną formę oryginału z rozporządzenia [1], a same zmiany polegałyby na tym, że:

- sekcja opisu przebiegu działań ratowniczych (zagrożenia i utrudnienia, zużyty i uszkodzony sprzęt) z karty została zaprezentowana jako sekcja *ogólna* i dodatkowo rozbita na trzy podsekcje: *opis*, *zagrożenia*, *utrudnienia*;
- sekcja opisu jednostek przybyłych na miejsce zdarzenia z karty została zaprezentowana jako sekcja *zasoby* i dodatkowo rozbita na trzy podsekcje: *opis*, *zużyte* i *uszkodzone*;
- sekcja opisu tego, co uległo zniszczeniu lub spaleni, została zaprezentowana jako sekcja *szkody*;
- sekcja wnioski i uwagi wynikające z przebiegu działań ratowniczych oraz inne uwagi dotyczące danych wypełnianych w formularzu odnośnie zdarzenia z karty została zaprezentowana jako

sekcja *wskazówki* i dodatkowo rozbita na dwie podsekcje *opis* i *punkty Czerpania Wody*.

Sekcja opisu warunków atmosferycznych z karty *Informacja ze zdarzenia* pozostała bez zmian i w szablonie reprezentowana jest jako węzeł – *warunki atmosferyczne*. Zabiegi te znacznie poprawiają czytelność reportu i możliwość jego przechowywania w postaci cyfrowej w systemie informacyjnym bez utraty informacji o sekcjach, tak jak to jest aktualnie w systemie EWID [2, 4]. Jedną dotychczasową sekcją opisową meldunku ze zdarzenia systemu EWID, w której KDR opisywali zdarzenie, zastępowana jest przez odpowiednie sekcje. Reprezentacja meldunku za pomocą tych wydzielonych, powiązanych sekcji (ryc. 3) do opisu przypadku zdarzenia daje możliwość łatwego przetwarzania i wyszukiwania raportów przez system informatyczny w poszukiwaniu konkretnych rozwiązań na podstawie zadanego pytania. Przyjęta reprezentacja logicznie oddziela od siebie części raportu, które są związane z różnymi aspektami działań ratowniczych, przez co łatwiej można też tworzyć precyzyjniejsze zapytania i ekstrahować niezbędną wiedzę.

2.3. Dołączanie przypadku zdarzenia do wybranej klasy ontologii *Akcje*

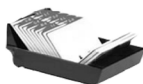
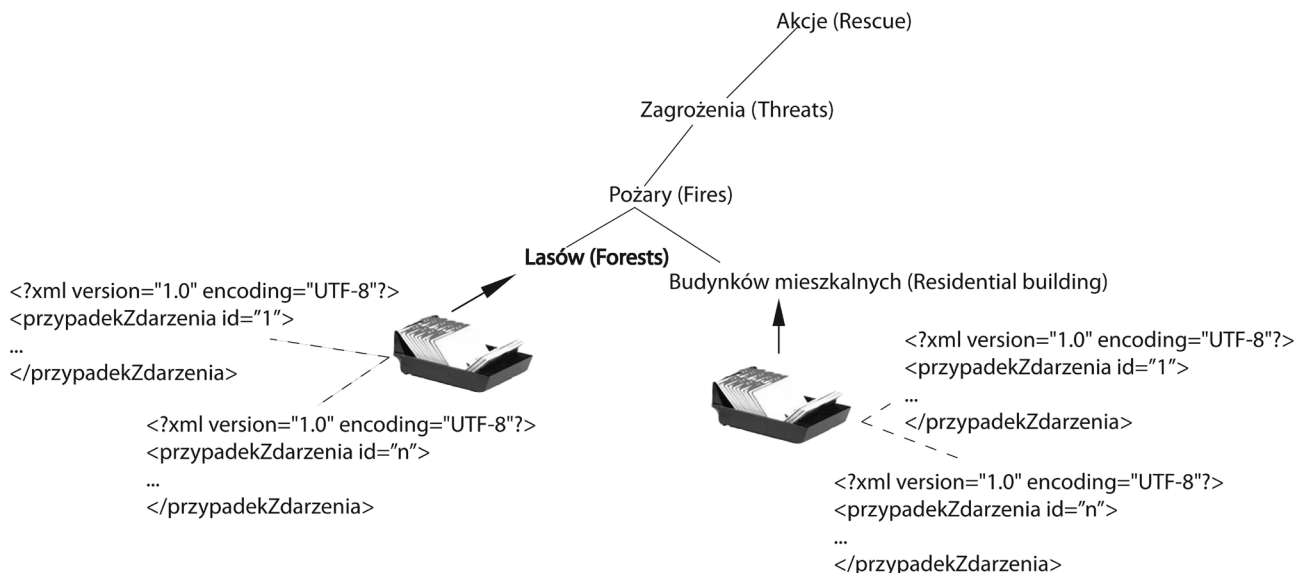
Przypadki zdarzeń opisane za pomocą szablonu zdarzeń, mogą zostać dołączone w ogólnym rozwiązaniu do wybranego węzła hierarchii lub klasy ontologii (kolekcji dokumentów należących do tej samej klasy). W szczególnych sytuacjach przypadki zda-

rzeń mogą być dołączane do *liści* hierarchii lub klas ontologii, które nie posiadają rozgałęzień. W niniejszym opracowaniu przedstawiono przypadek szczególny dla ontologii *Akcje*. Do dalszych rozważań i prezentacji metod klasyfikacji oraz przeszukiwania przyjęto, że do dyspozycji dana jest gałąź ontologii *Akcje-Zagrożenia-Požary* i że będą rozpatrywane pożary *Lasów* oraz *Budynków Mieszkalnych*. Wycinek rozpatrywanej ontologii przedstawia ryc. 3.

Ryc. 3 prezentuje sposób organizacji przypadków zdarzeń opisanych za pomocą ogólnego szablonu przypadku zdarzenia (przypadek zdarzenia 1) w ontologii *Akcje*. Do liści ontologii stanowiących klasy bez rozgałęzień dołączane są konkretne przypadki zdarzeń tj. przypadki zdarzeń będące instancją (stanowiące realizację), wypełnionego ogólnego szablonu przypadku zdarzenia. Każdy z liści może zawierać kolekcję – n takich konkretnych przypadków. Przy takiej reprezentacji, etykiety (nazwy klas liści) stają się automatycznie klasami, w metodach klasyfikacji, do których należy zaklasyfikować pojawiające się, nowe konkretne przypadki zdarzeń o nieznannej klasie docelowej.

3. Klasyfikacja tekstowych półstrukturalnych przypadków zdarzeń

Klasyfikacja, nazywana także kategoryzacją, dokumentów tekstowych (*ang. text document categorization* lub *text document classification*) [12-16] polega na określeniu do jakiej grupy dokumentów można zaliczyć wybrany tekst lub fragment tekstu (zadany za pomocą tzw. wzorca zapytania Q) w przypadku



Symbol graficzny reprezentujący kartotekę np. natywną bazę do przechowywania dokumentów zapisanych w rozszerzalnym języku znaczników XML (*ang. Extensible Markup Language*) (*Graphic symbol representing a directory such as Extensible Markup Language - XML native database*)

Ryc. 3. Organizacja przypadków zdarzeń w liściach ontologii *Akcje* [opracowanie własne]
Fig. 3. The organization of cases in the leaves of the Rescue ontology [own work]

wyszukiwania dokumentów. Celem klasyfikacji jest odnalezienie klasyfikatora, który będzie dokonywał przyporządkowania dokumentów do jednej lub kilku z uprzednio zdefiniowanych klas. Klasy te nie są definiowane wprost, lecz poprzez zbiór trenujący, który stanowi grupa dokumentów już odpowiednio zaklasyfikowana ręcznie np. przez ekspertów. W większości przypadków klasy nie są zagnieżdżane, natomiast przyjmuje się, iż jeden dokument może należeć do więcej niż jednej klasy. Do kategoryzacji dokumentów tekstowych używane są takie techniki, jak: drzewa decyzyjne (*ang. decision tree*), reguły decyzyjne, algorytmy najbliższych sąsiadów, klasyfikator bayesowski, sieci neuronowe, metody regresywne czy też techniki z zakresu maszyn wektorów wspierających (*ang. support vector machines – SVM*), oraz metody odnajdywania wspólnych podgrafów w przypadku zastosowania modelu wektorowego dokumentów [17]. W niniejszym opracowaniu zostanie przedstawiona koncepcja klasyfikacji meldunków za pomocą naiwnego klasyfikatora Bayesa. Klasyfikator Bayesa został zaproponowany do klasyfikacji meldunków ze względu na jego prostą interpretację oraz realizację aplikacyjną. Został on zaproponowany również dlatego, że przy wykorzystaniu reprezentacji przestrzenno-wektorowej dokumentu i jego kodowaniu Boolowskim, daje on dobre rezultaty tj. dobrze klasyfikuje nieznaną, nowe dokumenty do wydzielonych klas [16]. Klasyfikator ten nadaje się w szczególności do problemów o bardzo wielu wymiarach na wejściu. Mimo prostoty tej metody, często działa ona lepiej od innych, bardziej skomplikowanych metod klasyfikujących [18]. Zadaniem klasyfikatora Bayesa jest przyporządkowanie nowego przypadku do jednej z klas decyzyjnych, przy czym zbiór klas decyzyjnych musi być skończony i zdefiniowany *a priori*. Tak więc klasyfikacja polega na wykorzystaniu zbioru dotychczasowych obserwacji w celu ustalenia prawdziwości nowych hipotez [19] – jest to prawdopodobieństwo *a posteriori*. Twierdzenie Bayesa wyrażone jest następującym wzorem [20]:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Gdzie:

- $P(A|B)$ – prawdopodobieństwo warunkowe zajścia zdarzenia A pod warunkiem zajścia zdarzenia B,
- $P(B|A)$ – prawdopodobieństwo warunkowe zajścia zdarzenia B pod warunkiem zajścia zdarzenia A,
- $P(A)$ – prawdopodobieństwo zdarzenia A,
- $P(B)$ – prawdopodobieństwo zdarzenia B,

Na potrzeby niniejszego punktu, rozpatrującego użycie naiwnego klasyfikatora Bayesa do klasyfikacji dokumentów tekstowych (nowych przypadków

zdarzeń), przyjmowane są następujące oznaczenia i założenia:

- w_{ij} – wartość wagi j -tego wyrażenia w i -tym dokumencie,
- rozpatrywany jest przypadek reprezentacji przestrzenno-wektorowej tekstu oraz wagi w_{ij} wyrażen t dla poszczególnych dokumentów z korpusu dokumentów D , przyjmują wartość 0 lub 1 (została przyjęta reprezentacja Boolowska dokumentu),
- t_1, t_2, t_j – zbiór atrybutów warunkowych, które stanowią wyrażenia. Wyrażenia $t_j \in T$, gdzie T jest to zbiór wszystkich wyrażen opisujących przypadki zdarzenia,
- c_1, c_2, c_k – zbiór wartości atrybutów decyzyjnych, należących do klas decyzyjnych, do których klasyfikowany jest nowy przypadek opisujący zdarzenie. Wartości atrybutów decyzyjnych dla klas $c_k \in C$, gdzie C jest to zbiór klas decyzyjnych,
- Q – nowy klasyfikowany opis przypadku zdarzenia, wyrażony w postaci wektora wag wyrażen. Zapis $Q = [t_1 = q_1, \dots, t_j = q_j]$ lub krócej $Q = [q_1, \dots, q_j]$, oznacza przypisanie wagi $q_j = 0$ lub $q_j = 1$ (reprezentacja Boolowska) dla wyrażen t_j wektora Q tj. dane wyrażenie $t_j \in T$ występuje bądź nie w klasyfikowanym przypadku opisanym za pomocą Q .

Posługując się teorią Bayesa, można wykazać, że najbardziej prawdopodobną klasą, do której zostanie zaklasyfikowany nowy przypadek zdarzenia wyrażony w postaci wektora Q , jest klasa c_k , która maksymalizuje prawdopodobieństwo warunkowe $P(c_k | q_1, \dots, q_j)$. Klasa ta oznaczona jest jako a_{MAP} (*ang. maximum a posteriori*) i wyznaczana jest za pomocą wzoru 2.

Wzór 2 określa wybór klasy decyzyjnej c_k dla zadanego przypadku określonego w postaci wektora wag wyrażen Q . Z ostatniej części wzoru usunięto mianownik z tego względu, iż prawdopodobieństwo $P(q_1, \dots, q_j)$ ma wartość stałą, niezależną od klasy decyzyjnej c_k , więc nie ma ono wpływu na wybór klasy a_{MAP} .

$$a_{MAP} = \arg \max_{c_k \in C} P(c_k | q_1, \dots, q_j)$$

$$a_{MAP} = \arg \max_{c_k \in C} \frac{P(q_1, \dots, q_j | c_k) P(c_k)}{P(q_1, \dots, q_j)} \quad (2)$$

$$a_{MAP} = \arg \max_{c_k \in C} P(q_1, \dots, q_j | c_k) P(c_k)$$

Maksymalne prawdopodobieństwo *a posteriori* (a_{MAP}), ze zbioru tych prawdopodobieństw, składa się z iloczynu dwóch czynników:

- prawdopodobieństwa *a priori* określonego jako $P(c_k)$. Prawdopodobieństwo to można oszacować jako iloraz liczby przykładów uczących n_k należących do klasy c_k do liczby wszystkich przykładów uczących n zbioru klas C :

$$P(c_k) = \frac{n_k}{n} \quad (3)$$

Gdzie:

- n_k – liczba przykładów uczących z k -tej klasy
 - $n = n_1 + \dots + n_k$ - liczba wszystkich przykładów uczących

• szansy określanej jako $P(q_1, \dots, q_j | c_k)$. Szansę można oszacować na dwa sposoby. Pierwszy sposób polega na tym, iż $P(q_1, \dots, q_j | c_k)$ szacuje się, dla dużych zbiorów danych treningowych (uczących), jako stosunek liczby przykładów uczących opisanych wartościami atrybutów warunkowych $[q_1, \dots, q_j]$ i należących do klasy c_k do liczby wszystkich przykładów uczących z klasy c_k . Drugi sposób polega na tym, iż do oszacowania szansy $P(q_1, \dots, q_j | c_k)$, w naiwnym klasyfikatorze Bayesa, wprowadza się założenie o warunkowej niezależności wartości atrybutów przy ustalonej klasie decyzyjnej. W kontekście analizy tekstu założenie powyższe oznacza, że wystąpienie w dokumencie tekstowym jednego wyrażenia nie dostarcza żadnych informacji o prawdopodobieństwie wystąpienia drugiego. Dla tekstów zazwyczaj naiwny warunek niezależności nie jest spełniony z tego względu, że na podstawie jednego wyrażenia można przewidzieć następne. Mankament ten można minimalizować za pomocą analizy do wydobywania słów kluczowych, opierając się na badaniu statystycznej współzależności wyrażen [21] lub n -gramowych modeli języka stanowiących aproksymację Markowa [22]. Niemniej założenie to w praktyce nie pogarsza i nie obniża skuteczności klasyfikatora. Po przyjęciu naiwnego założenia o niezależności atrybutów, szansę można zapisać jako:

$$P(q_1, \dots, q_j | c_k) = \prod_{j=1}^n P(q_j | c_k) \quad (4)$$

Gdzie:

- $P(q_j | c_k)$ – prawdopodobieństwo warunkowe, które można oszacować jako iloraz liczby przykładów uczących z klasy c_k , dla których wartość wag d_{ij} atrybutów (wyrażen) t_j równe są wartościom wag q_j wektora wyrażen Q , do liczby wszystkich przykładów uczących z klasy c_k .

Po uwzględnieniu wyżej przyjętego założenia, do klasyfikacji nowego przykładu (przypadku zdarzenia) wybrana zostaje klasa a_{NB} (Naïve Bayes), dla której zachodzi wzór 5.

$$a_{NB} = \arg \max_{c_k \in C} P(c_k) \prod_{j=1}^n p(q_j | c_k) \quad (5)$$

Etap estymacji prawdopodobieństw *a priori* odpowiadają w innych metodach uczenia maszynowe-

go etapowi uczenia się. Naiwny model Bayesa wykorzystywany jest w praktyce ze względu na swoją prostotę, zarówno pod względem zrozumienia jego działania, jak i implementacji algorytmicznej, oraz wydajność obliczeniową. Dla zmiennych o wartościach dyskretnych oraz dla parametrycznych jednowymiarowych modeli gęstości zmiennych o wartościach rzeczywistych wystarczy dokonać jednokrotnego przeglądu danych, aby zbudować naiwny klasyfikator Bayesa [23]. Przechodzenie kilkukrotne w celu budowy klasyfikatora może odbywać się dla bardziej złożonych modeli gęstości, takich jak modele mieszane. Wynika to z ich iteracyjnej natury dopasowania funkcji gęstości. Wszystkie te elementy odgrywają kluczową rolę w budowie i odświeżaniu indeksu binarnego dokumentów w celu ich wyszukiwania przedstawionego w punkcie 6. W szczególności kiedy dodawany jest do klasy nowy dokument i zachodzi potrzeba przebudowy indeksu oraz wyliczenia na nowo wartości poszczególnych prawdopodobieństw dla wszystkich klas.

4. Klasyfikacja półstrukturalnych przypadków zdarzeń przy użyciu naiwnego klasyfikatora Bayesa

– przykład zastosowania

W podpunkcie tym został przybliżony proces klasyfikacji nowego przypadku zdarzenia, zgodnie z teorią i założeniami omówionymi w punkcie 3.

Na podstawie przyjętych założeń możliwe jest zbudowanie ogólnej macierzy łączącej zbiór dokumentów ze zbiorem wyrażen i klasami. Tabela 1 prezentuje ogólny, symboliczny zapis takiej macierzy.

Tabela 1.

Symboliczny zapis macierzowy reprezentacji meldunków [Źródło: opracowanie własne]

Table 1.

The term matrix representation of the report

[Source: own work]

	T – zbiór wyrażen $t_j (t_j \in T)$ (T – set of terms t_j)				Klasa c_k (Class c_k) ($c_k \in C$)	
D – zbiór dokumentów $d_i (d_i \in D)$ (D – set of documents d_i)		t_1	t_2	t_3	t_4	
	d_1	w_{11}	w_{12}	w_{13}	w_{14}	c_1
	d_2	w_{21}	w_{22}	w_{23}	w_{24}	c_1
	d_3	w_{31}	w_{32}	w_{33}	w_{34}	c_2
	d_4	w_{41}	w_{42}	w_{43}	w_{44}	c_2

Do dalszych rozważań założono, że istnieją dwie klasy decyzyjne, do których można zaklasyfikować nowy przypadek zdarzenia. Pierwszą klasę stanowią przypadki opisujące pożary budynków (PB), drugą klasę stanowią pożary lasów (PL). Zatem zbiór klas $C = \{c_1 = PB, c_2 = PL\}$. Przykładowy zbiór uczący składa się z 4 dokumentów $D = \{d_1, d_2,$

d_3, d_4 oraz 4 wyrażen indeksujacych te dokumenty $T = \{t_1, t_2, t_3, t_4\}$. Wyrazenia moga zostac pozyskane z przypadkow zdarzen. Do wybranych w sposob *ad hoc*, przykladowych wyrazen dla celow demonstracyjnych naleza $T = \{t_1 = \text{osmalony}, t_2 = \text{zadymiony}, t_3 = \text{czad}, t_4 = \text{wysoki plomien}\}$. W szczegolnosci wyrazenia pozyskuje sie w procesie wstepnego przetwarzania tekstow (*ang. pre-processing*) i ekstrakcji cech (*ang. feature extraction*) [12, 22].

Wagi dla poszczegolnych wyrazen w dokumentach przyjmuja reprezentacje Boolowska tj. $w_{ij} = 0$ lub $w_{ij} = 1$ ($1 \leq i \leq 4$ oraz $1 \leq j \leq 4$). Dla tak skonstruowanych zalozen i zbioru danych mozliwe jest zbudowanie macierzy, laczonej zbior dokumentow ze zbiorem wyrazen i klasami. Tabela 2 prezentuje konkretny zapis takiej macierzy.

Tabela 2.

Realizacja zapisu macierzowego dla meldunkow.

Źródło: [opracowanie własne]

Table 2.

Realization of the term matrix representation of the report. Source: [own work]

D – zbiór doku- mentów $d_i (d_i \in D)$ (D – set of documents d_i)	T – zbiór wyrazen $t_j (t_j \in T)$ (T – set of terms t_j)				Klasa c_k (Class c_k) ($c_k \in C$)
	$q_1 =$ osma- lony	$q_2 =$ za- dymio- ny	$q_3 =$ czad	$q_4 =$ wysoki plomien	
d_1	1	1	1	0	PB
d_2	0	1	0	1	PB
d_3	0	0	0	1	PL
d_4	0	1	1	1	PL

Tabela 2 reprezentuje zapis informacji, który można interpretować w następujący sposób: grupa ekspertów zidentyfikowała dwie klasy pożarów (klasy decyzyjne) – pożar budynków i pożar lasów, do których przydzielono *a priori* na podstawie analizy treści po dwa dokumenty opisujące zdarzenie (stanowiące zbiór uczący).

Następnie w kolejce do klasyfikacji pojawia się nowy, tym razem niezatetykowany opis zdarzenia. Jego klasyfikacja opiera się na znajdowaniu odpowiednich wyrazen i wnioskowaniu Bayesa w celu określenia dla niego odpowiedniej klasy. Założono, że z przykladowego, nowego, niesklasyfikowanego przypadku zdarzenia wyekstrahowano następujące wyrażenia: *zadymiony*, *czad*. Wektor wag wyrazen Q nowo sklasyfikowanego przypadku można zapisać w następujący sposób $Q = [q_1 = \text{osmalony} = 0, q_2 = \text{zadymiony} = 1, q_3 = \text{czad} = 1, q_4 = \text{wysoki plomien} = 0]$. W celu określenia, do jakiej klasy zostanie przydzielony nowy przypadek na podstawie jego Q przy wykorzystaniu klasyfikatora Bayesa, należy przeprowadzić obliczenia dotyczące:

- oszacowania prawdopodobienstwa *a priori* dla klasy PB i PL (wzór 3). Prawdopodobienstwa te wynoszą $P(c_{PB}) = P(c_{PL}) = \frac{2}{4} = \frac{1}{2}$,

- oszacowania szansy, że dany wektor Q nalezy do klasy PB i klasy PL (wzór 4). Szacowanie tego, że Q będzie nalezeć do klasy PB jest równe $P(0,1,1,0 | c_{PB}) = \frac{1}{2} \cdot \frac{2}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$. Szacowanie tego, że Q będzie nalezeć do klasy PL jest równe $P(0,1,1,0 | c_{PL}) = \frac{2}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4}$. Problema-

tyczne w szacowaniu szansy PL może być to, iż skladowa 4 wektora Q przyjmuje wartość 0 ($q_4 = 0$), zaś w zbiorze danych dla klasy PL żaden przypadek nie ma zerowej tej skladowej. Aby uniknąć mnozenia przez zero przyjmuje się w takim przypadku wartość 1 w dalszym mnozeniu,

- wyliczenia prawdopodobienstwa *a posteriori* tego, że Q zostanie przydzielone do klasy PB i klasy PL (wzór 5.). Prawdopodobienstwo tego, że Q będzie przydzielone do klasy PB wynosi

$$a_{NB1} = \frac{1}{2} \cdot \frac{1}{8} = \frac{1}{16}$$

natomiast tego, że zostanie przydzielony do klasy PL wynosi $a_{NB2} = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$.

Wzór 5. określa także, który rezultat szacowania prawdopodobienstwa *a posteriori* wybrać. W rozpatrywanym przypadku największą wartość klasyfikatora Bayesa ma argument a_{NB2} . Tak więc rozpatrywany dokument wyrażony poprzez zbudowany wektor wyrazen Q zostanie zaklasyfikowany do grupy dokumentow zawierajacych opisy pożarow lasow.

5. Binarne miary podobienstwa półstrukturalnych przypadkow zdarzen

W punkcie 3. założono, że dokumenty są indeksowane binarnie tj. do dyspozycji jest binarny wektor wyrazen, zawierający wagi wyrazen opisujacych dany dokument d . Waga wyrazenia w tej reprezentacji przyjmuje wartość 0 – wyrażenie t nie występuje w dokumencie d lub 1 – wyrażenie t występuje w dokumencie d . W celu odnalezienia dystansu lub podobienstwa pomiędzy samymi przypadkami zdarzen, jak również pomiędzy przypadkami zdarzen a zapytaniem Q , należy posłużyć się jedną z 76 dostępnych miar binarnych opisanych m.in. w pracach [24-26]. Miara binarna jest to miara podobienstwa lub odległości pomiędzy wektorami binarnymi [24]. Zapytanie Q traktowane jest jako zredukowany opis przypadku do kilku znaczących wyrazen t podawanych przez KDR w celu wyszukania na ich podstawie najlepiej pasujacych przypadkow z bazy wiedzy systemu CBR. Miary binarne powstaly na początku XX wieku, a ich rozwój zaczal się od zaproponowa-

nej w 1901 roku przez Jacarda miary podobieństwa gatunków [24]. Miary te są szczególnym przypadkiem miar stosowanych do pomiaru dystansu, podobieństwa pomiędzy wektorami niebinarnymi, czyli takimi, których wagi mogą przybierać wartości nie tylko 0 lub 1. Przegląd i zastosowania miar nie binarnych można odnaleźć m.in. w pracach [27, 28].

6. Mieszane przetwarzanie półstrukturalnego modelu opisu zdarzeń

Mając do dyspozycji przedstawiony w punkcie 2. półstrukturalny model opisu zdarzeń, na który składa się ontologia *Akcja* zawierająca w liściach półstrukturalne przypadki zdarzeń, autorzy zaproponowali i opisali w niniejszym punkcie sposób jego przetwarzania. Przetwarzanie to łączy w sobie elementy klasyfikacji, opisane w punkcie 4. oraz elementy binarnych miar podobieństwa, opisanych w punkcie 5. Wszystkie te składniki po połączeniu ze sobą służą jako wyszukiwarka archiwalnych przypadków zdarzeń w bazie wiedzy systemu CBR. Schemat ideowy funkcjonowania niniejszej wyszukiwarki prezentuje ryc. 4.

Ryc. 4 przedstawia schemat działania wyszukiwarki archiwalnych przypadków zdarzeń zawartych w bazie wiedzy systemu CBR, a dokładniej w określonych liściach ontologii *Akcje*. Proces przeszukiwania rozpoczyna się od podania przez KDR zapytania Q zbudowanego z wyrażen (słów, zdań) opisującego powstałe zagrożenie. Zapytanie to jest następnie przetwarzane, usuwane są z niego zbędne wyrażenia na podstawie *stop-listy*. Stop-lista definiuje wyrażenia nieniosące żadnej wartościowej informacji oraz które należy odfiltrować, do takich wyrażen należą np. „i”, „czy”, etc. W kolejnym kroku wyrażenia, które przeszły filtrowanie, poddawane są procesowi lematyzacji, polegającemu na sprowadzaniu wyrażen do ich formy podstawowej [22]. Po tym procesie wyrażeniom przypisywana jest waga binarna równa 1. Tak zbudowany i uzupełniony wektor jest następnie klasyfikowany, np. za pomocą naiwnego klasyfikatora Bayesa opisanego w punkcie 3. oraz 4., do klas – liści ontologii. Po otrzymaniu wartości prawdopodobieństwa *a posteriori* przynależności wektora do wybranych klas, są one szeregowane od największej do najmniejszej według wartości otrzymanego prawdopodobieństwa. Następnie wybieranych jest n (użytkownik ten parametr może określić sam) najlepiej dopasowanych klas. Klasy niespełniające kryterium wyrażonego w postaci stopnia przynależności (określonego prawdopodobieństwa *a posteriori*), mogą zostać odrzucone. Do dalszej analizy wyszukiwania przechodzi n wybranych klas. Każda wyselekcjonowana klasa zawiera kolekcję przypadków zdarzeń. W związku z tym, że przypadki zostały zaindeksowane binarnie do procesu klasyfikacji oraz że wagi zapytania Q także są binarne, następuje dla każdej kla-

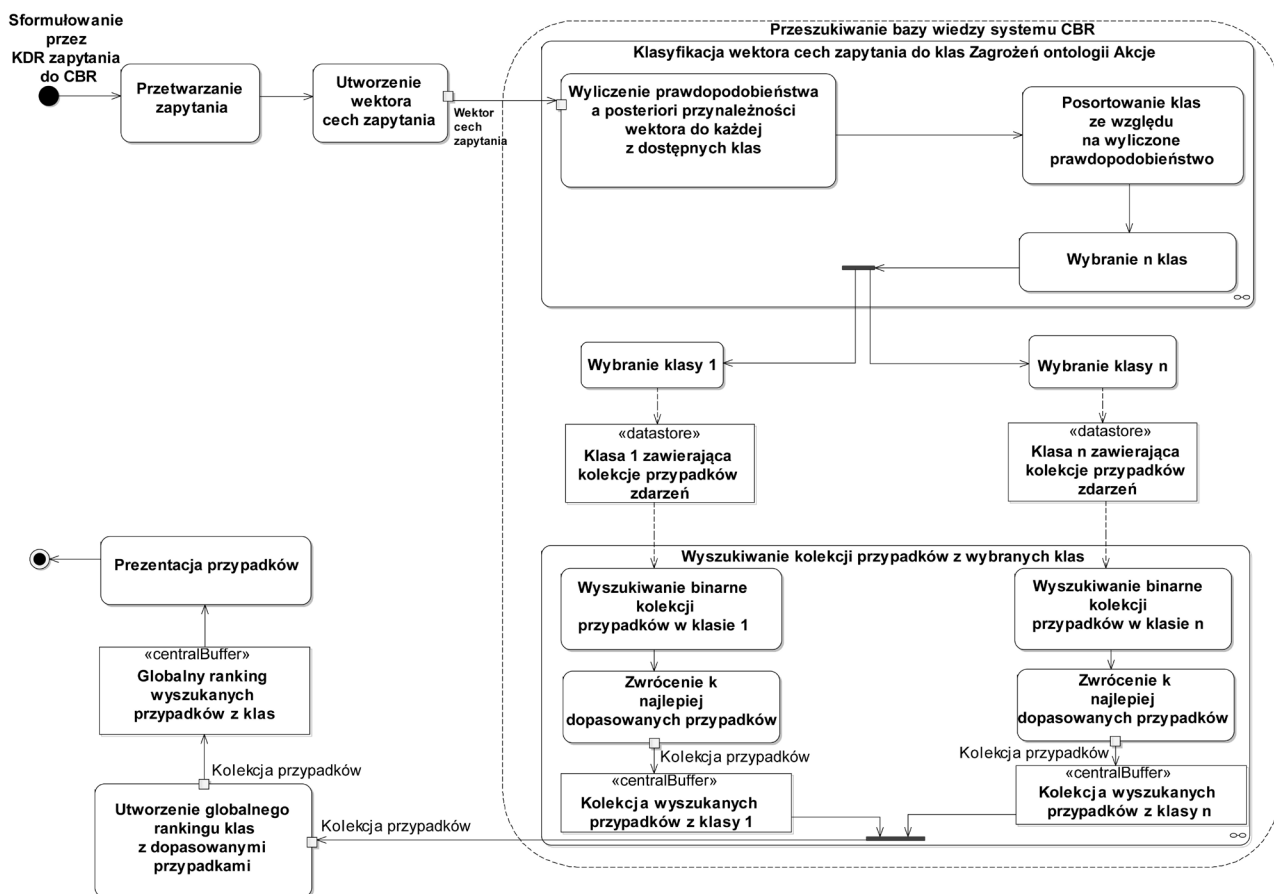
sy z osobna binarny proces wyszukiwania za pomocą miar podobieństwa zdefiniowanych w punkcie 5. Wyszukiwanie polega na odnalezieniu przypadku najbardziej podobnego do wektora zapytania Q (algorytm k -najbliższych sąsiadów). Liczba zwracanych przypadków w klasie może być ograniczana za pomocą parametru k wspólnego dla wszystkich klas (użytkownik ten parametr może określić sam). Wyszukiwanie kończy się otrzymaniem rankingu zawierającego kolekcję $p = n \cdot k$ przypadków. Utworzony ranking zawierający archiwalne przypadki prezentowany jest następnie dla KDR, który wybiera najlepiej pasujący przypadek zdarzenia i adaptuje go w celu rozwiązania zaistniałego zagrożenia.

Podsumowanie

Opracowany, prezentowany na rycinie 4. schemat ideowy wyszukiwarki przypadków zdarzeń, jak również niniejszy artykuł stanowią opis podjętych przez autorów badań z zakresu analizy meldunków z systemu EWID. Aktualna propozycja wyszukiwania będzie podlegała badaniom. Wyniki natomiast ukażą się w następnych publikacjach autorów. Zawarte w niniejszym opracowaniu treści dają nowy i ciekawy pogląd na możliwość łączenia ontologii z tekstowymi półstrukturalnymi przypadkami zdarzeń oraz ich przetwarzania (klasyfikacji czy też wyszukiwania). Jak dotychczas autorzy nie spotkali się z takim podejściem w systemach dla służb ratowniczych czy też w warstwie reprezentacji i przetwarzania wiedzy systemów CBR. Aktualnie w tych systemach dominują rozwiązania polegające na reprezentacji wiedzy albo w pełni strukturalnej albo w pełni niestrukturalnej [29-32]. Z powyższych względów postanowili zaprezentować swoje rozwiązania, z wyprzedzeniem w stosunku do aktualnego harmonogramu badań. Aktualnie wysiłki badawcze autorów koncentrują się na analizie znaczeniowej zdań oraz przywracaniu sekcji z raportów z systemu EWID w celu ich dalszych analiz i strukturalizacji. Podstawą tej analizy jest zaklasyfikowanie całego zdania do wybranej klasy na podstawie analizy znajdujących się w nim wyrażen.

Do obiecujących dalszych kierunków rozwoju opisywanego w artykule tematu, należy klasyfikacja przypadku nie tylko do *liścia* ontologii, ale także do węzła. Przypadek ogólny zakłada, że w węzłach hierarchii (klasach) mogą znajdować się przypadki niesklasyfikowane poprawnie tj. których klasyfikator nie jest pewien. W tym przypadku wyznaczony musi zostać stopień pewności dla każdego węzła z klas, jeśli dany przypadek go nie przekracza to klasyfikacja odbywa się do węzła rodzica lub wyżej i tak aż do korzenia ontologii.

Dalszymi kierunkami rozwoju mogą być modyfikacje opisu półstrukturalnego przypadku zdarzenia polegające na dodaniu wiązań do innej sekcji ontologii, przechowującej np. oznaczony rejestr zasob-



Ryc. 4. Schemat ideowy wyszukiwarki przypadków zdarzeń, zawartych w liściach ontologii Akcje, które stanowią bazę wiedzy systemu CBR [Źródło: opracowanie własne]

Fig. 4. Schematic diagram of the CBR cases search engine [Source: own work]

bów. Wiązania te mogą odbywać się poprzez użycie referencji w *opisie* lub poprzez dodanie nowej sekcji *powiązania* do półstrukturalnych przypadków zdarzeń. W przypadku referencji w opisach, rozwiązanie to ideowo zbliżone jest do hiperlinków sieci ogólnoswiatowej.

Ważny kierunek rozwoju stanowi też budowanie bazy aktywnej, dynamicznie uzupełniającej przypadek zdarzenia o potrzebne informacje na podstawie np. bazy wiedzy i reguł. W celu zobrazowania tego działania autorzy posłużyli się następującym przykładem – z systemu pasywnego, jaki stanowi CBR ze względu na to, że przechowuje „obraz” przeszłości, pozyskany zostaje przypadek zdarzenia. System odkrywa jednak, że aktualnie nie dysponuje takimi zasobami, wówczas wykorzystuje reguły z bazy wiedzy i rejestru aktualnych zasobów, aby uzupełnić raport o potrzebne parametry. Ogólnie, kierunki takich łączonych technik aktualnie są już podejmowane, przykład takiego podejścia można odnaleźć w literaturze [35].

Ostatnią kwestią w proponowanych badaniach pozostaje dobór i badanie klasyfikatorów np. drzew decyzyjnych, reguł decyzyjnych oraz miar binarnych w odniesieniu do tradycyjnych, powszechnie przyjętych miar stosowanych dla tekstów, w repre-

zentacji przestrzenno-wektorowej, do których należą m.in. miary [22]: Jacarda, Dicea, Kosinusów czy też Euklidesa.

Ontologia, jako warstwa nośna informacji o akcjach i działaniach ratowniczo-gaśniczych w połączeniu z pół-strukturalną reprezentacją przypadków zdarzeń dla systemu CBR, daje elastyczny i użyteczny model reprezentacji wiedzy o dziedzinie oraz wnioskowania. Na podstawie tak zaprezentowanej wiedzy o dziedzinie zawierającej opis archiwalnych zdarzeń, istnieje możliwość dostarczania najpotrzebniejszych informacji Kierującemu Działaniami Ratowniczymi. Informacja ta wyrażona jest w postaci opisu zagrożeń, wskazówek etc., które KDR może wziąć pod uwagę podczas przeprowadzanej akcji ratowniczo-gaśniczej. Ponadto ontologia doskonale nadaje się do modelowania zależności pomiędzy elementami dziedziny np. występującymi w opisach zdarzeń zasobami użytymi do likwidacji zagrożenia a samymi zasobami z gałęzi ontologii *Akcje*. Taka prosta referencja daje systemowi wiedzę o tym, jakie zasoby są alokowane do jakich zdarzeń.

Zastosowanie naiwnego klasyfikatora Bayesa może służyć jako element rankingujący klasy w procesie wyszukiwania. Zabieg ten daje możliwość zrównoleglenia obliczeń dotyczących wyszukiwania

grupy binarnie opisanych przypadków, najlepiej pasujących do zadanego przez KDR pytania. Dla każdej z wylosowanych klas możliwe staje się utworzenie oddzielnego procesu, w którym następuje przeszukanie za pomocą przedstawionych miar binarnych i dopasowanie wektora zapytania do znajdujących się w klasie przypadków zdarzeń. Na koniec wyszukiwania procesy te są synchronizowane a rezultat dostarczany Kierującemu Działaniami Ratowniczymi.

Literatura

1. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 29 grudnia 1999 r. w sprawie szczegółowych zasad organizacji krajowego systemu ratowniczo-gaśniczego. Dz.U. 99. 111.1311 § 34 pkt. 5 i 6.
2. Abakus: System EWID99. [on-line] [dostęp: 1 maja 2009] Dostępny w Internecie: http://www.ewid.pl/?set=roz_w_ewid&gr=roz.
3. Abakus: System EWIDSTAT. [on-line] [dostęp: 1 maja 2009] Dostępny w Internecie: <http://www.ewid.pl/?set=ewidstat&gr=prod>.
4. Strona firmy abakus. [on-line] [dostęp: 1 marca 2009] Dostępny w Internecie: <http://www.ewid.pl/?set=main&gr=aba>.
5. Krasuski A., Maciak T. Wykorzystanie rozproszonej bazy danych oraz wnioskowania na podstawie przypadków w procesach decyzyjnych Państwowej Straży Pożarnej. „Zeszyty Naukowe SGSP”, No 36, 2008, s. 17-35.
6. Kozłowski J., Neuman Ł. Wspomaganie wyszukiwania dokumentów mapami samoorganizującymi. [Wrocław]: III Krajowa Konferencja MISSI 2002, 19-20 września - „Multimedialne i Sieciowe Systemy Informacyjne”, 2002. [dostęp: 10 czerwca 2009] Dostępny w Internecie: <http://www.zsi.pwr.wroc.pl/zsi/missi2002/pdf/s507.pdf>.
7. Krasuski A., Krenski K. Building a DSS for Fire Service using jCOLIBRI. Analele Universitatii Bucuresti, Informatica LVI, 2008.
8. 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, No 11, 2008.
9. Mirończuk M., Maciak T. Problematyka projektowania modelu hybrydowego systemu wspomagania decyzji dla Państwowej Straży Pożarnej. „Zeszyty Naukowe SGSP”, No 39, 2009.
10. Mirończuk M., Karol K. Koncepcja systemu ekspertowego do wspomaganie decyzji w Państwowej Straży Pożarnej. In: Grzech A., Juszczyk K., Kwaśnicka H. and Nguyen N.T., editors. Inżynieria Wiedzy i Systemy Ekspertowe. Warszawa: Akademicka Oficyna Wydawnicza EXIT, 2009.
11. 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.
12. Borycki Ł., Sołdacki P. Automatyczna klasyfikacja tekstów. [Wrocław]: III Krajowa Konferencja MISSI 2002, 19-20 września - „Multimedialne i Sieciowe Systemy Informacyjne”, 2002. [dostęp: 10 czerwca 2009] Dostępny w Internecie: <http://www.zsi.pwr.wroc.pl/zsi/missi2002/pdf/s504.pdf>.
13. Song F., Liu S., Yang J. A comparative study on text representation schemes in text categorization. Pattern Analysis & Applications, No 8, 2005, s. 199 - 209
14. Weigend A. S., Wiener E. D., Pedersen J. O. Exploiting Hierarchy in Text Categorization. Information Retrieval, No 1, 1999.
15. Yang Y., Liu X. A re-examination of text categorization methods. [New York]: ACM SIGIR Conference of Research and Development in Information Retrieval, 1998.
16. Łażewski Ł., Pikuła M., Siemion A., Szklarzewski M. Klasyfikacja dokumentów tekstowych. Warszawa: PJWSTK 2005. Dostępny w Internecie: <http://www.scribd.com/doc/2242106/Klasyfikacja-dokumentow-tekstowych>.
17. Schenker A., Kandel A., Bunke H., Last M. Graph-Theoretic Techniques for Web Content Mining. World Scientific Publishing Co, 2005.
18. StatSoft. Naiwny klasyfikator Bayesa. [dostęp: 10 stycznia 2010] Dostępny w Internecie: http://www.statsoft.pl/textbook/stathome_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstnaiveb.html.
19. Aas K., Eikvil L. Text Categorisation: A Survey. Technical Report, Norwegian Computing Center, 1999.
20. Yoshimasa, Tsujii T. J. Training a Naive Bayes Classifier via the EM Algorithm with a Class Distribution Constraint. In: Proceedings of the 7th Conference on Natural Language Learning: Morgan Kaufmann, 2003. s. 127-134.
21. Matsuo Y., Ishizuka M. Keyword Extraction From A Single Document Using Word Co-Occurrence Statistical Information. International Journal on Artificial Intelligence Tools, No 13, 2004, s. 157-169.
22. Mykowiecka A. Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym. Warszawa: PJWSTK, 2007.
23. Hand D., Mannila H., Smith P. Eksploracja danych. Wydanie 1. Warszawa: Wydawnictwo Naukowo-Techniczne, 2005.
24. Choi S. S., Cha S. H., Tappert C. C. A Survey of Binary Similarity and Distance Measures Systemics, Cybernetics and Informatics, No 8, 2010, s. 43-48.
25. Veal B. Similarity Coefficients for Binary Data. Department of Mathematics. London: London School of Economics, 2008.
26. Lourenço F., Lobo V., Bação F. Binary-based similarity measures for categorical data and their application in Self-Organizing Maps. JOCLAD, 2004. p. 1-18.

27. Cha S. H. *Comprehensive Survey on Distance/ Similarity Measures between Probability Density Functions*. International journal of mathematical models and methods in applied sciences, 2007.
28. Kim M. C., Choi K. S. *A comparison of collocation-based similarity measures in query expansion*. Information Processing and Management: an International Journal, No 35, 1999, s. 19-30
29. Kempa A. *Zastosowanie rozszerzonej metodologii wnioskowania na podstawie przypadków - textual cbr w pracy z dokumentami tekstowymi*. Katowice: Systemy Wspomagania Organizacji/, 2005. [dostęp: 1 stycznia 2008] Dostępny w Internecie: <http://www.swo.ae.katowice.pl/content/view/221/32/>.
30. Krupka J., Kasparova M., Jirava P. *Case-Based Reasoning Model in Process of Emergency Management*. Man-Machine Interactions Advances in Soft Computing, 2009. p. 77-84.
31. Jing L. *Case-Based Reasoning Intelligent Decision Approach for Firefighting Tactics*. [Tianjin]: Intelligent Networks and Intelligent Systems, 2009 ICINIS '09 Second International Conference on 1-3 Nov 2009, 2009.
32. Bergmann R., Wilke W., Vollrath I., Wess S. *Integrating General Knowledge with Object-Oriented Case Representation and Reasoning*. 1996.
33. Gliński W. *Języki i narzędzia do tworzenia i wyszukiwania ontologii w kontekście semantycznego weba*. Instytut Informacji Naukowej i Studiów Bibliologicznych UW. [dostęp: 10 sierpnia 2010] Dostępny w Internecie: <http://bbc.uw.edu.pl/Content/20/14.pdf>.
34. 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>.
35. Shimin D., Shen H., Liu H. *Research on Case-Based Reasoning Combined with Rule-Based Reasoning for Emergency*. [Philadelphia, PA, USA]: Service Operations and Logistics, and Informatics, 2007 SOLI 2007 IEEE International Conference on 27-29 Aug 2007, 2007.

Praca naukowa współfinansowana ze środków Europejskiego Funduszu Społecznego, środków Budżetu Państwa oraz ze Środków Budżetu Województwa Podlaskiego w ramach projektu „Podlaska Strategia Innowacji – budowa systemu wdrażania”



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



mgr inż. Marcin Michał Mirończuk, absolwent Wydziału Elektrycznego Politechniki Białostockiej, na którym także ukończył studia doktoranckie. Aktualnie ma wszczęty przewód doktorski na Wydziale Informatyki Politechniki Białostockiej i pracuje w Instytucie Podstaw Informatyki PAN w Warszawie.

dr hab. inż. Tadeusz Maciak, profesor Szkoły Głównej Służby Pożarniczej oraz kierownik Zakładu Informatyki i Łączności w tej szkole. Objął stanowisko adiunkta w Katedrze Mediów Cyfrowych i Grafiki Komputerowej Wydziału Informatyki Politechniki Białostockiej PB.