Piotr Stolarski, Krzysztof Węcel, Bartosz Perkowski, Agata Filipowska, Witold Abramowicz

Katedra Informatyki Ekonomicznej
Uniwersytet Ekonomiczny w Poznaniu

Jacek Piotrowski

Sygnity SA

EWOLUCJA MODELI DANYCH NA PRZYKŁADZIE AKTUALIZACJI PROFILI ZAGROŻEŃ W SYSTEMIE SEMANTYCZNY MONITORING CYBERPRZESTRZENI

1. Wprowadzenie

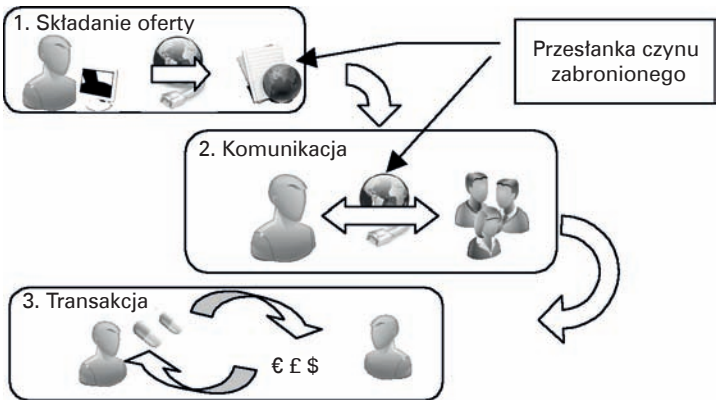
Projekt Semantyczny Monitoring Cyberprzestrzeni (SMC)[1] jest projektem badawczo-rozwojowym, którego celem jest opracowanie prototypu wraz z architekturą referencyjną systemu wspierającego nadzór nad publiczną przestrzenią wymiany informacji w wybranych mediach elektronicznych, w celu wykrywania potencjalnych działań przestępczych.

Prace prowadzone w ramach projektu rozpoczęły się od wyboru modelowego scenariusza biznesowego[2]. Wybrany scenariusz, zgodnie z założeniami projektu,

[1] http://smc.kie.ue.poznan.pl.

[2] „Scenariusz biznesowy" jest pojęciem odnoszonym do metody projektowania systemu informatycznego, nie ma natomiast związku z dziedziną wykorzystania tego systemu.

dotyczy naruszenia prawa poza domeną medium elektronicznego, choć naruszenie to ma związek z treścią przesyłanej informacji. Upubliczniona informacja stanowi istotną przesłankę zaistnienia penalizowanej sytuacji, co przedstawiono na rysunku 1.



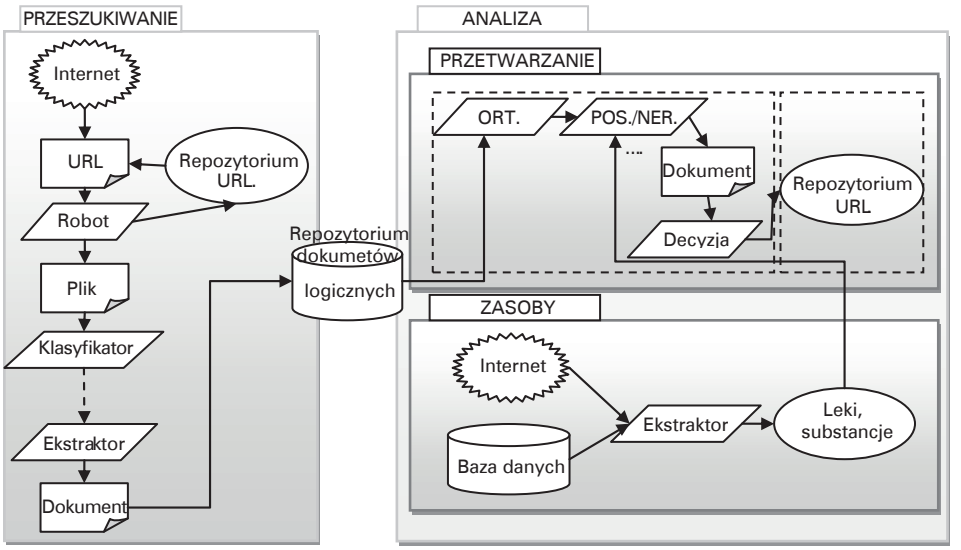
Rysunek 1. Realizowany scenariusz biznesowy

Źródło: opracowanie własne.

W rozważanym scenariuszu zakłada się monitoring różnorodnych źródeł informacji w sieci WWW dla identyfikowania przypadków nielegalnych transakcji dotyczących niezgodnego z prawem obrotu lekami i receptami uprawniającymi do zakupu leków. Scenariusz ten jest szczególnie predestynowany do prowadzenia prac badawczych i implementacji prototypu pozwalającego udowodnić tezę przydatności podobnych narzędzi do wybranej klasy problemów – śledzenia poszlak nielegalnych działań poprzez analizę dokumentów w Internecie, przede wszystkim ze względu na: jasność i względną prostotę wykładni przepisów, dużą częstotliwość przypadków handlu lekami w polskim Internecie oraz ważny aspekt społeczny tego typu zjawisk.

Na podstawie szczegółowej analizy klasy problemów reprezentowanych przez wybrany scenariusz biznesowy, a także po przeprowadzeniu analizy wymagań dla systemu został opracowany model architektury systemu SMC. Ze względu na złożoność realizowanego przez system zadania model ten ma budowę modułową, a komunikacja systemu z końcowymi użytkownikami odbywa się za pomocą graficznego interfejsu aplikacji internetowej. Rysunek 2 przedstawia architekturę prototypu systemu SMC.

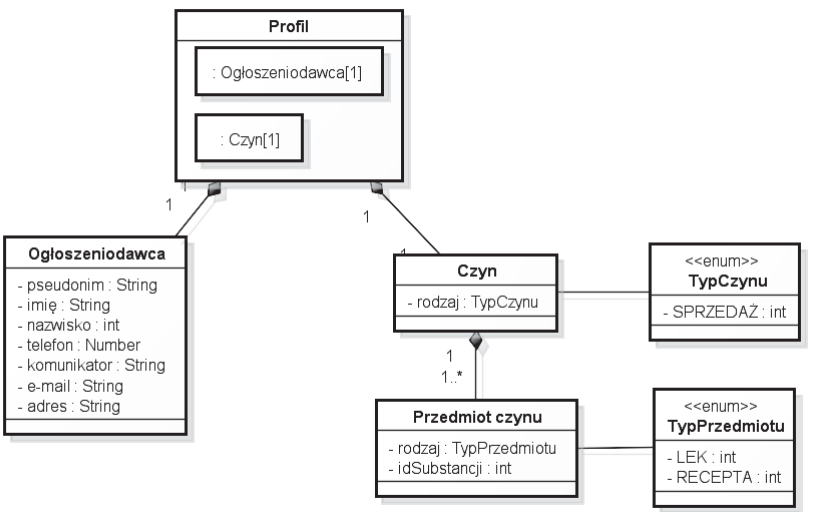
Centralnym elementem proponowanego modelu jest repozytorium dokumentów logicznych. Dokumentom tym odpowiadają instancje opracowanego profilu zagrożenia. Dokument logiczny w systemie SMC określa byt informacyjny, będący wynikiem przetworzenia dokumentu pochodzącego z monitorowanego źródła, na podstawie którego dokonuje się w kolejnych krokach uzupełniania przyjętego profilu zagrożenia.



ORT., POS., NER. – elementy modułu ekstrakcji leksykalnej

Rysunek 2. Architektura prototypu systemu SMC dla wybranego scenariusza biznesowego

Źródło: opracowanie własne.



Rysunek 3. Diagram UML reprezentujący schemat profilu zagrożenia dla przyjętego scenariusza

Źródło: opracowanie własne.

Profil zagrożenia zaprezentowany na rysunku 3 ma charakter wzorca zbudowanego z atrybutów zagrożenia. Konkretyzacja wzorca oznacza wypełnienie atrybutów wartościami odzwierciedlającymi domenę monitorowanych zagrożeń. W przypadku profilu implementowanego prototypu atrybutami zagrożenia są: **ogłoszeniodawca**, **czyn** oraz **przedmiot czynu**, których domeną są odpowiednio: osoby fizyczne, kupno lub sprzedaż oraz lek lub recepta.

Profil jest uzależniony od rozważanego scenariusza biznesowego. W rezultacie zmiana scenariusza biznesowego powoduje konieczność zmiany dziedziny oraz definicji na poziomie konkretyzacji profilu. Zmienność scenariusza biznesowego nie jest jednak jedynym możliwym powodem konieczności aktualizacji profilu zagrożenia w rozważanej architekturze. Do innych czynników wymuszających potrzebę aktualizacji profilu zagrożenia należą następujące sytuacje:

- zmiany struktury źródeł informacji, wymuszające modyfikację reguł ekstrakcji strukturalnej,
- zmiany zawartości źródeł informacji, wymuszające modyfikację dziedziny atrybutów profilu oraz reguł normalizacji i ekstrakcji leksykalnej,
- zmiany wartości oceny instancji profili zagrożenia w kontekście nowej informacji pochodzącej z monitorowanych źródeł lub potrzeb analitycznych użytkowników.

Każdą z powyższych sytuacji wymuszających ewolucję profilu zagrożenia określono jako odrębną perspektywę ewolucji. W części 2 niniejszej pracy omówiono literaturę związaną z zagadnieniami zarządzania zmianą i modelowania profili w innych projektach badawczych. Z kolei w części 3 przedstawiono szczegółowo każdą z perspektyw ewolucji profilu, a w 4 określono, jaki wpływ mają poszczególne perspektywy na wzorzec architektury systemu SMC.

2. Ewolucja – powiązane prace

Fundamentalną cechą informacji oraz baz danych jest ich zmienność. Informacje charakteryzują się tym, że ewoluują w czasie, jednakże sposób ewolucji jest zależny od mechanizmów odpowiedzialnych za ewolucję danych. Autorzy artykułu *Evolution and Change in Data Management – Issues and Directions*[3] prezentują czynniki, przez które dane ulegają zmianom, oraz sposoby, w jakie zmiany są przeprowadzane. Jednym z kierunków wytyczanych przez autorów jest podejście temporalne i przestrzenne dotyczące rozpoznawania danych.

[3] J.F. Roddick et al., *Evolution and Change in Data Management – Issues and Directions*, seria „ACM SIGMOD Record" 2000, vol. 29/1, March.

Z kolei w pracy *A Framework for Diagnosing Changes in Evolving Data Streams*[4] wskazuje się na techniki diagnozowania tendencji w ewoluujących strumieniach danych poprzez tworzenie profili mierzących tempo zmian koncentracji danych w określonej lokalizacji przestrzennej w zdefiniowanym przez użytkownika horyzoncie czasowym. Profile takie obrazują, jak przedstawia się gromadzenie danych w przekroju czasowym i geograficznym. Estymacja pozwala na stworzenie temporalnego profilu prędkości oraz przestrzennego profilu prędkości, których zadaniem jest umożliwienie przewidzenia trzech rodzajów ewolucji danych: rozpadu, koagulacji oraz zmiany. Temporalny profil prędkości pozwala na badanie częstotliwości zmian gęstości w stałej lokalizacji przestrzennej, natomiast przestrzenny profil prędkości obrazuje zmienność danych oraz umożliwia przegląd zmian gęstości danych w różnych momentach czasowych. Wspomniany artykuł prezentuje nowatorskie techniki diagnozowania trendów w ewoluujących strumieniach danych, a dzięki generowanym dwóm rodzajom profili dostarcza różnych punktów widzenia na naturę zmian w trakcie analizy charakterystyk danych.

Ewolucja danych w znacznym stopniu dotyczy baz danych. Jest procesem polegającym na aktualizacji schematu bazy lub hurtowni danych oraz na ewolucji danych wskutek aktualizowania schematu bazy. W kolejnym artykule, *Efficient and Scalable Data Evolution with Column Oriented Databases*[5], zostały wymienione przyczyny ewolucji baz danych. Jedną z nich jest pozyskiwanie informacji o danych, które wymagają aktualizacji, dodania lub usunięcia atrybutów w bazie danych, a nawet modyfikacji struktury tabel. Drugą z przyczyn jest natomiast napływ informacji o obciążeniu bazy, co wynika z różnych wzorców dostępu oraz wymagań dotyczących schematów służących do zoptymalizowania obciążenia. Autorzy proponują rozwiązanie, które zamiast łączyć kolumny w celu uzyskiwania rezultatów zapytań SQL, a następnie kompresować i przechowywać oddzielnie każdą kolumnę, pozwala na ewolucję danych bezpośrednio z i do skompresowanych źródeł. Wyniki prowadzonych badań dowodzą, że takie podejście pozwala zredukować czas i lepiej skalować dane niż poprzez ewolucję zapytań do bazy danych.

Kolejnym podejściem do ewolucji, zaprezentowanym w artykule *Dynamic Ontologies on the Web*, jest rozwój ontologii, której zmiana może być rozwiązana na kilka sposobów[6]. Przykładowym problemem przy usuwaniu pojęcia jest postępowanie z jego podpojęciami. Kompleksowa zmiana ontologii powinna umożliwiać zmianę tych podpojęć oraz ich usuwanie, zapewniając równocześnie ich modyfikowalność.

---

[4] Ch.C. Aggarwal, *A Framework for Diagnosing Changes in Evolving Data Streams*, w: *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, ACM, New York 2003.

[5] Z. Liu et al., *Efficient and Scalable Data Evolution with Column Oriented Databases*, w: *Proceedings of the 14th International Conference on Extending Database Technology*, ACM, New York 2011.

[6] J. Helfin, J.A.Hendler, *Dynamic Ontologies on the Web,* w: *Proceedings of the 17th National Conference on Artificial Intelligence*, AA Press, Cambridge 2000.

Wzrost złożoności możliwych zmian powoduje wzrost stopnia rozbudowania rozwiązań służących ewolucji ontologii. Z kolei autorzy artykułu *Ontology Evolution as Reconfiguration-Design Problem Solving*[7] skupiają się na dwóch istotnych czynnikach dotyczących ewolucji ontologii: sposobu, w jaki użytkownik specyfikuje żądanie zmian ontologii, oraz tego, jak dane żądanie jest realizowane. Problemem związanym ze specyfikacją zmian jest elastyczność, w ramach której użytkownik ma ograniczoną możliwość przeprowadzania zmian. W przywołanym artykule przedstawiono rozwiązanie problemu ewolucji ontologii poprzez modyfikowanie schematów przy zastosowaniu grafów. Podejście to poprzez wprowadzanie heurystyk związanych z szeregowaniem sprzyja łatwiejszemu rozbudowywaniu i efektywniejszemu utrzymaniu systemu związanego z ewolucją ontologii. Szeregowanie służy temu, aby dla każdej zmiany ontologii było wybierane najlepiej dopasowane rozwiązanie.

Zmienność ontologii wynika z rozwoju badań. Systemy często ignorują ontologie, jako „żyjące byty" i podmioty, które mogą w sposób ciągły ulegać zmianom. Wynika to z szybkiego tempa rozwoju badań[8], w skutek którego ontologie ewoluują, a związane z nimi odwołania stają się nieprawidłowe. Zmiana ontologii jest związana z odwzorowaniami, które przestają być aktualne, i wówczas konieczna jest ich modyfikacja oraz adaptacja. W innym tekście[9] zaproponowano nowy system integracji danych poprzez wskazanie różnic w proponowanej architekturze w stosunku do tradycyjnych podejść dotyczących ewolucji danych bazujących na ontologiach, zastosowanie niskopoziomowych operatorów oraz wykorzystanie logów w celu modyfikacji zapytań użytkownika odnoszących się do różnych wersji ontologii. Prezentowane rozwiązanie nie pozwala na automatyczne odwzorowanie. Zmiany są przechowywane w logach, które następnie dzięki zapytaniom użytkowników zostają zmodyfikowane. Autorzy wykazują, że utrzymanie odpowiednich mapowań w ontologiach jest dużo bardziej skomplikowane niż w przypadku baz danych. Różnica ta wynika z formalizmów reprezentujących ontologie, które zawierają pojęcie ważności, co oznacza, że konkretne kombinacje aksjomatów ontologii nie muszą być obowiązujące. W bazach danych natomiast każdy zbiór pokrywający się ze schematem jest dopuszczalny. Wizja ewolucji ontologii jest coraz bardziej realna, natomiast obecnie żaden system nie wspiera jej w pełni.

Prezentowane prace wyszczególniają zróżnicowane podejścia do ewolucji informacji oraz baz danych. Istotny wpływ na architekturę systemu ma baza danych oraz przechowywane w niej informacje. Pobierane informacje wpływają na schemat profilu zagrożenia, dlatego istotne jest dokonanie analizy danych oraz ich zawartości w celu prawidłowej rozbudowy bazy danych. Kolejnym kluczowym czynnikiem jest

---

[7] L. Stojanovic et al., *Ontology Evolution as Reconfiguration-Design Problem Solving*, w: *Proceedings of the 2nd international conference on Knowledge capture*, ACM, New York 2003.

[8] Przede wszystkim chodzi tutaj o poziom zaawansowania w zakresie modelowania domeny ontologii.

[9] H. Kondylakis, D. Plexousakis, Enabling *Ontology Evolution in Data Integration*, w: *Proceedings of the 2010 EDBT/ICDT Workshops*, ACM, New York 2010.

rodzaj przechowywanych informacji, co wiąże się z podejściem do ewolucji ontologii. Definiowanie reguł ekstrakcji jest podobne do problematyki rozwoju pojęć stosowanych w ontologii. Oba przypadki są zależne od stopnia rozbudowania rozwiązań, dlatego ich modyfikowanie może zostać ułatwione dzięki zastosowaniu grafów. Ostatni rodzaj zmian może wynikać z różnego przedziału czasowego w podejściu do analizowanych instancji profilu zagrożenia. Ma to wpływ na ocenę istotności konkretnej instancji. Uwzględnione publikacje wpływają na działania związane z ewolucją profilu zagrożenia, wytyczając kierunki, którymi ewolucja ta może podążać.

3. Perspektywy ewolucji profilu zagrożenia w systemie SMC

3.1. Metoda budowy profilu zagrożenia

Metoda budowy profilu zagrożenia pokazana jest na rysunku 4. Została stworzona na potrzeby opracowywania profilu zagrożenia dla prototypowej implementacji systemu SMC. Przy opracowaniu poszczególnych etapów tej metody zespół projektowy korzystał z doświadczenia uzyskanego w ramach innych projektów[10], a także z kompetencji projektantów z zakresu inżynierii systemów, inżynierii wiedzy czy projektowania schematów baz danych. Dodatkowo wzięto pod uwagę doświadczenia przedstawione w literaturze przedmiotu, m.in. związanej z tematyką zarządzania zmianami w systemach informacyjnych. Powstała metoda nawiązuje do metod znanych z dziedzin: modelowania danych – również za pomocą schematów ER, modelowania zorientowanego obiektowo oraz inżynierii wiedzy z konstruowaniem ontologii[11], ale zawiera także aspekty, które wymagały dodatkowego uwzględnienia[12]. Są to przede wszystkim: powiązanie modelu z uregulowaniami prawnymi monitorowanej dziedziny oraz kwestie dostępności i organizacji informacji, którą system może pozyskać ze źródeł informacji o zagrożeniach.

Przedstawiona metoda ma powtarzalny charakter w dwóch zakresach. Fragmentaryczne zastosowanie pierwotnej metody ma sens w przypadku konieczności modyfikacji profilu w trakcie eksploatacji systemu. Sytuacja taka może zajść np. w razie istotnych zmian w regulacjach prawnych związanych z dziedziną profilu. Może także

[10] W. Abramowicz et al., *Reprezentacja wiedzy w systemie wyszukiwania ekspertów eXtraSpec*, w: *Systemy informacyjne w zarządzaniu: księga jubileuszowa z okazji 70-lecia urodzin Profesora Adama Nowickiego*, red. J. Korczak, I. Chomiak-Orsa, H. Sroka, Wrocław 2010, s. 149–160.

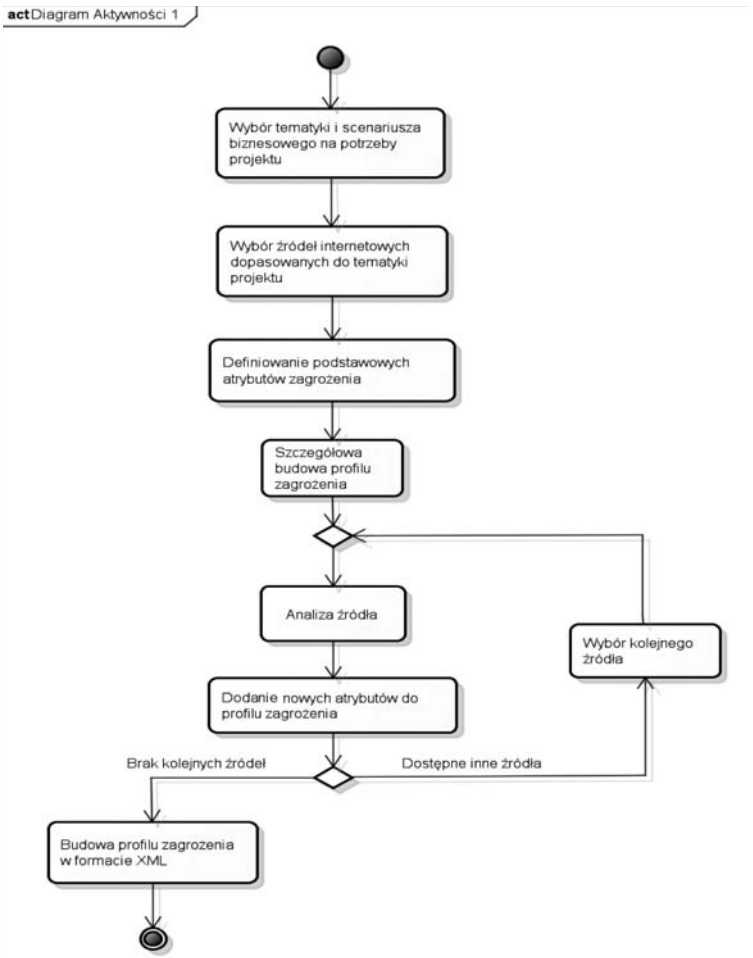
[11] M. Afsharchi, B.H. Far, *Automated ontology evolution in a multi-agent system*, w: *Proceedings of the 1st international conference on Scalable information systems*, ACM, New York 2006.

[12] D.J. Ernst et al., *Hybird and custom data structures: evolution of the data structures course*, w: *Proceedings of the 14th annual ACM SIGCSE conference on Innovation and technology in computer science education*, ACM, New York 2009.

wynikać z potrzeby dostosowania systemu do innych lub zmienionych źródeł informacji o zagrożeniach. Drugi zakres zastosowania dotyczy samej metody. Zastosowanie takie będzie mieć miejsce w przypadku potrzeby przystosowania systemu SMC do tej zmiany scenariusza biznesowego, która prowadzi do monitorowania innych zasobów lub powiązania systemu z innymi normami prawnymi.

3.2. Adaptatywność zasobów i reguł systemu

Definicja profilu zagrożenia dla danego scenariusza biznesowego nie sprowadza się wyłącznie do schematu opisującego strukturę takiego profilu. Równie istotnymi elementami są sposoby, w jaki instancje profilu wypełniane są informacją.



Rysunek 4. Metoda budowy profilu zagrożenia

Źródło: opracowanie własne.

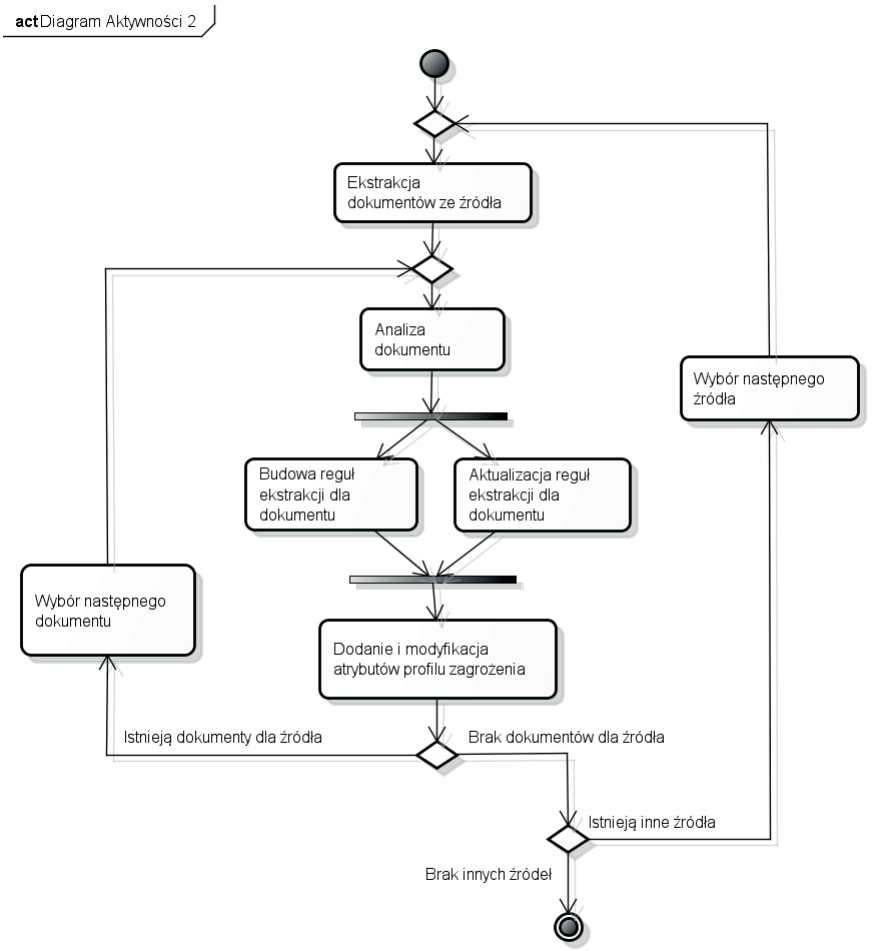
Elementami tymi w systemie SMC są przede wszystkim:

- referencyjne dane pochodzące z głębokiego Internetu;
- referencyjne dane ze źródeł informacji o zagrożeniach;
- reguły ekstrakcji strukturalnej;
- reguły ekstrakcji leksykalnej;
- metody normalizacji wartości atrybutów zagrożenia.

Przyjęto, że profil zagrożenia związany jest ze źródłami pochodzącymi z głębokiego Internetu. Tego typu źródła wymagają częstej adaptacji, ale są istotnie związane z domeną wybranego scenariusza biznesowego. Korzystanie z takich źródeł dotyczy jednak szerokiej klasy senariuszy biznesowych, dlatego badania w tym zakresie są uzasadnione. W realizowanym prototypie są nimi bazy leków ze szczegółową informacją dotyczącą nazewnictwa, substancji aktywnych, producentów etc. Odzwierciedlają one sytuację na zmiennym rynku farmaceutycznym, zatem konieczne jest ich regularne aktualizowanie. Istotnym aspektem wymagającym zaimplementowania jest z jednej strony niezależność systemu przy dostępie do tego typy źródeł, z drugiej – możliwość regularnego odświeżania agregowanej informacji.

Rozpatrując kategorię referencyjnych danych ze źródeł informacji o zagrożeniach, uwzględniamy przede wszystkim fakt, że źródła te zawierają informację i wiedzę przechowywaną w języku naturalnym. W efekcie analiza takiej informacji jest utrudniona ze względu na fakt występowania różnorakich zjawisk lingwistycznych. Zjawiska te wykazują się dużą różnorodnością co najmniej w trzech wymiarach: czasowym, geograficznym oraz społecznościowym. Z tym ostatnim wymiarem związane są dodatkowo takie zjawiska, jak: slang, specyficzne nazewnictwo, skróty i akronimy. W rezultacie analiza treści wymaga gromadzenia danych referencyjnych, np. w postaci odpowiednich słowników. Słowniki takie są częściowo zależne od konkretnego źródła oraz pojawiających się mód.

Prototyp systemu SMC agreguje informację zapisaną w języku naturalnym pochodzącą ze źródeł internetowych. Źródła takie cechują się dobrze określoną strukturą na poziomie całego dokumentu. Poszczególne elementy dokumentów, będąc informacją zapisaną w języku naturalnym, mają z kolei tendencję do braku jakiejkolwiek struktury. Stąd potrzeba dwóch rodzajów przetwarzania informacji z postaci źródłowej do postaci wynikowej, czyli instancji profilu zagrożenia. W terminologii systemu SMC tymi metodami przetwarzania są: ekstrakcja strukturalna oraz ekstrakcja leksykalna. Dla każdej z nich potrzebne są specyficzne zasoby: reguły ekstrakcji strukturalnej i reguły ekstrakcji leksykalnej (rysunek 5).

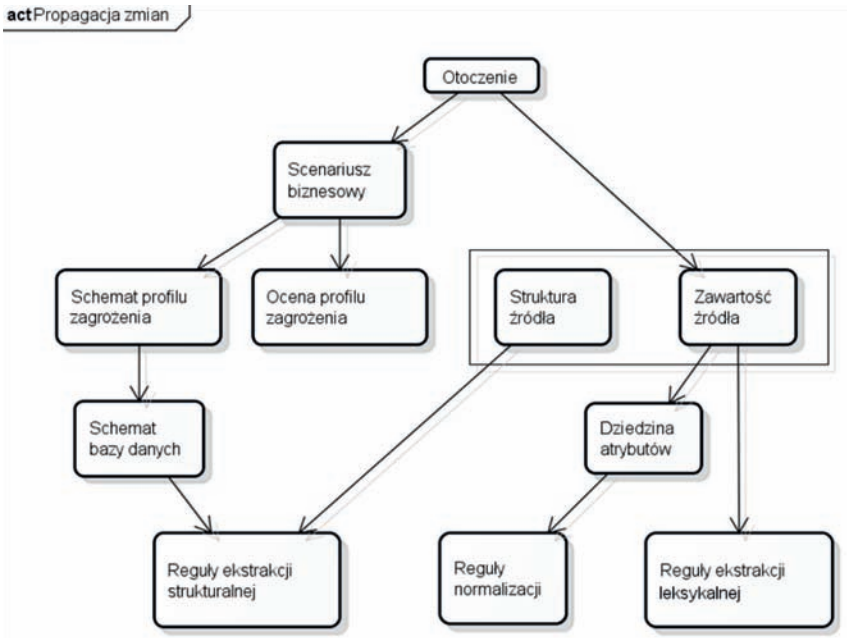


Rysunek 5. Adaptatywność zasobów i reguł systemu do środowiska internetowego
Źródło: opracowanie własne.

Ekstrakcja leksykalna jest związana dodatkowo z koniecznością normalizacji wartości atrybutów zagrożenia. Instancje profilu zagrożenia są następnie automatycznie oceniane i wraz z rekomendacjami w postaci oceny przedstawiane użytkownikom końcowym.

Każdy z powyższych elementów systemu może podlegać zmianie, co będzie prowadziło do zmian w przechowywanych profilach zagrożenia. Opisane wyżej elementy są wzajemnie powiązane, co powoduje, że bardzo trudno uniknąć złożonych modyfikacji wynikających z pojedynczej zmiany (rysunek 6). Zjawisku temu ma zapobiegać przyjęty wzorzec architektury, ale w ramach rozwoju prototypu taka sytuacja

występowała często. Dodatkowo, zmiana elementów innych perspektyw ewolucji może wymuszać zmiany w powyższych elementach.



Rysunek 6. Propagacja zmian w architekturze systemu SMC

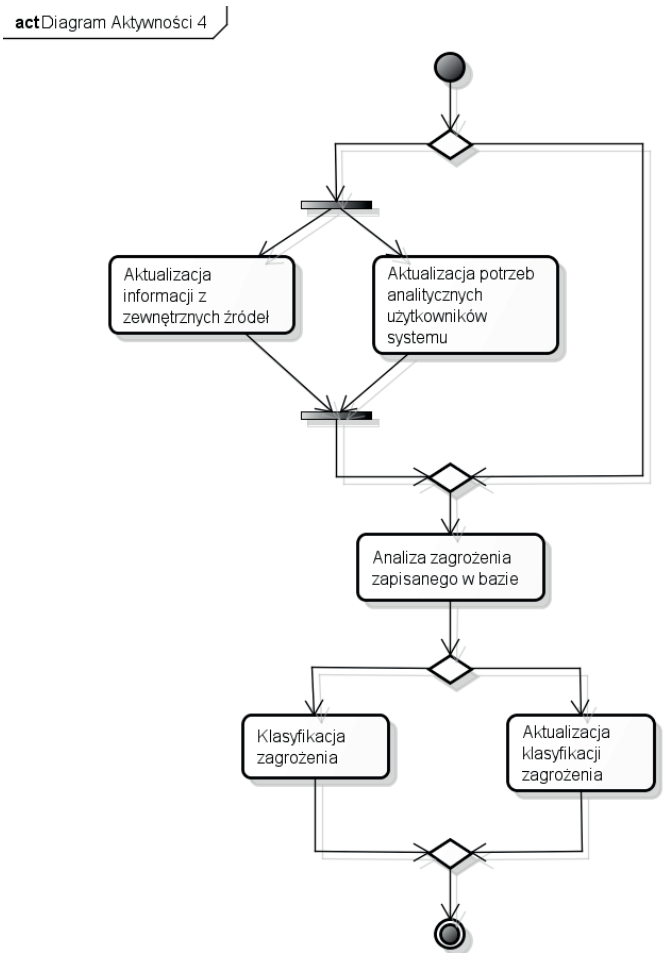
Źródło: opracowanie własne.

3.3. Ewolucja oceny instancji profilu zagrożenia

Ostatnią omawianą perspektywą ewolucji jest ewolucja ocen instancji profilu zagrożenia. System SMC ocenia (klasyfikuje) poszczególne zagrożenia według ustalonych kryteriów. W zależności od przyjętego algorytmu oceny może on wykazywać tendencję do niestabilności. Powodem takiego stanu rzeczy jest zmienność środowiska informacyjnego systemu.

Zmienność środowiska informacyjnego wynika z powiększania bazy dokumentów, ale także jest rozumiana jako zmiany w informacji reprezentującej zbiory wartości referencyjnych dla poszczególnych atrybutów. Jeżeli w algorytmie oceny bierze się pod uwagę dane historyczne, to wraz z upływem czasu pojawi się konieczność zmiany dotychczas określonych ocen, a więc klasyfikacji wcześniej utworzonych instancji profili zagrożeń. Aby uniknąć tego zjawiska, w projekcie przyjęto, że w przypadku rozwijanego prototypu raz dokonana ocena nie podlega modyfikacjom, natomiast proponowany algorytm oceny uwzględnia wszystkie dotychczas utworzone instancje profilu zagrożenia.

Zmiana informacji w zakresie kryterium oceny może wywołać zmianę oceny danej instancji profilu zagrożenia. Dzieje się tak w następujących przypadkach szczególnych: pojawienia się nowych dokumentów źródłowych powiązanych z wcześniej ocenionym lub zmiany parametrów algorytmu klasyfikującego dokonanej przez użytkownika.



Rysunek 7. Ewolucja oceny instancji profilu zagrożenia

Źródło: opracowanie własne.

4. Wpływ poszczególnych perspektyw ewolucji profilu zagrożenia na prototyp systemu SMC

Podsumowując powyższą analizę, należy wskazać na perspektywy ewolucji profilu poprzez określenie ich zakresu wpływu na całokształt własności funkcjonalnych i pozafunkcjonalnych projektowanego systemu.

Perspektywa pierwsza: metoda budowy profilu zagrożenia. Perspektywa ta nie nakłada szczególnych wymogów w zakresie ewolucji przypadku dla istniejącej implementacji systemu przy założeniu niezmienności realizowanego scenariusza biznesowego. Użycie metody do innego scenariusza wymaga implementacji systemu wspierającej: zmienność zestawów reguł ekstrakcji strukturalnej i reguł ekstrakcji leksykalnej, zmienność zestawów reguł normalizacji. Dodatkowymi wymaganiami mogą być: modułowość implementacji z dobrze ustrukturyzowanym kodem oraz przystosowanie architektury systemu w stosunku do architektury wzorcowej.

Perspektywy druga: adaptatywność zasobów i reguł systemu. Wymogi nałożone na właściwości systemu obejmują: utworzenie mechanizmu oceny skuteczności reguł ekstrakcji strukturalnej i ekstrakcji leksykalnej; zaimplementowanie narzędzia wspierającego zmianę reguł ekstrakcji strukturalnej oraz reguł ekstrakcji leksykalnej; wsparcie dla pozyskiwania domenowej informacji ze źródeł zewnętrznych (np. bazy leków) oraz implementację metody włączania wiedzy o „anomaliach" językowych (np. slang na forach).

Perspektywa trzecia: ewolucja oceny instancji profilu zagrożenia. Ostatnia z analizowanych perspektyw ewolucji profilu, związana ze zmiennością ocen instancji profili, wymaga od systemu uwzględnienia: odpowiedniej implementacji algorytmu klasyfikacji instancji profili; zaprojektowania dla użytkownika interfejsu pozwalającego zachować przejrzystość zasad nawigacji i prezentacji informacji, pomimo zmian oceny profili; stworzenia mechanizmu rozwiązującego kwestię zachowywania wyników analizy instancji profili przez użytkownika.

Literatura

1. Abramowicz W., Stolarski P., Węcel K., Wieloch K., *Reprezentacja wiedzy w systemie wyszukiwania ekspertów eXtraSpec*, w: *Systemy informacyjne w zarządzaniu: księga jubileuszowa z okazji 70-lecia urodzin Profesora Adama Nowickiego*, red. J. Korczak, I. Chomiak-Orsa, H. Sroka, Wrocław 2010.
2. Afsharchi M., Far B.H., *Automated ontology evolution in a multi-agent system,* w: *Proceedings of the 1st international conference on Scalable information systems (InfoScale '06)*, ACM, New York 2006.

3. Aggarwal Ch., *A Framework for Diagnosing Changes in Evolving Data Streams*, w: *Proceedings of the 2003 ACM SIGMOD international conference on Management of data (SIGMOD '03)*, ACM, New York 2003, s. 575–586.
4. Ernst D.J., Stevenson D.E., Wagner P.J., *Hybrid and custom data structures: evolution of the data structures course*, w: *Proceedings of the 14th annual ACM SIGCSE conference on Innovation and technology in computer science education (ITiCSE '09)*, ACM, New York 2009, s. 213–217.
5. Heflin J., Hendler J.A., *Dynamic Ontologies on the Web*, w: *Proceedings of the 17th National Conference on Artificial Intelligence*, AAAI Press, Cambridge 2000, s. 443–449.
6. Kondylakis H., Plexousakis D., *Enabling Ontology Evolution in Data Integration*, w: *Proceedings of the 2010 EDBT/ICDT Workshops (EDBT '10)*, ACM, New York 2010.
7. Liu Z., He B., Hsiao H., Chen Y., *Efficient and Scalable Data Evolution with Column Oriented Databases*, w: *Proceedings of the 14th International Conference on Extending Database Technology*, ACM, New York 2011, s. 105–116.
8. Roddick J. et al., *Evolution and Change in Data Management – Issues and Directions*, seria „ACM SIGMOD Record" 2000, vol. 29/1, March.
9. Stojanovic L., Maedche A., Stojanovic N., Studer R., *Ontology Evolution as Reconfiguration-Design Problem Solving*, w: *Proceedings of the 2nd international conference on Knowledge capture (K-CAP '03)*, ACM, New York 2003, s. 162–171.

Summary

Transformation of data models on an example of a threat profile evolution in the Semantic Monitoring of the Cyberspace system

The Semantic Monitoring of the Cyberspace (SMC) project aims at developing a prototype and a reference architecture of a system that is to support the public information exchange space surveillance in order to detect potential criminal activities. In the paper, we introduce the architecture of the solution implemented for detection of a selected type of actions i.e. illegal pharmaceutical trade. Then, we discuss the generalization of the data models and algorithms utilized for an easy adaptation of the system to other domains. The text uncovers three different aspects of such a generalization. We conclude by enlisting the functional and non-functional requirements for each of the generalization perspectives.