



SMC

Semantyczny Monitoring Cyberprzestrzeni

Semantyczny Monitoring Cyberprzestrzeni

**Katedra Informatyki Ekonomicznej
Uniwersytet Ekonomiczny w Poznaniu**

Partnerzy projektu:



Katedra
Informatyki
Ekonomicznej

Sygnity
competence on



Zarys problemu

- Źródło internetowe jako zasób **użytecznych** informacji o konkretnej tematyce
- Rodzaje zasobów, a ich zawartość (płytki a głęboki Internet)
- Monitorowanie danych i informacji pojawiających się w źródłach internetowych na przykładzie projektu:
 - **Semantyczny Monitoring Cyberprzestrzeni**

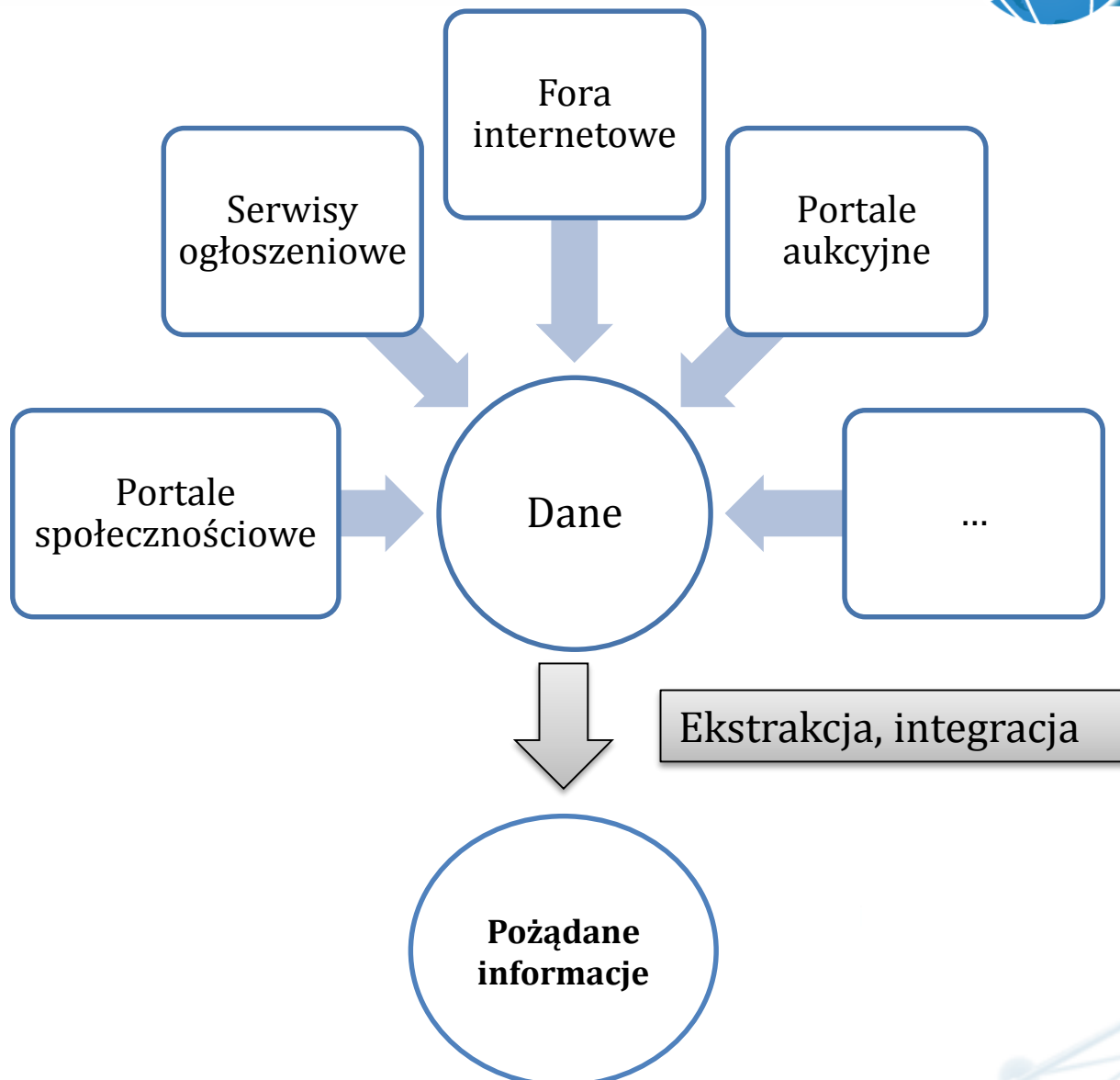




Koncepcja

System pozwalający na automatyczne pozyskiwanie pożądaných treści ze źródeł internetowych wskazanych (wyszukanych przez system i zatwierdzonych) przez użytkownika





WIDOCZNE KOLUMNY

- Flaga △ ▾ ✕
- Lek △ ▾ ✕
- Data ogłoszenia △ ▾ ✕
- Kategoria △ ▾ ✕
- Ogłoszeniodawca △ ▾ ✕
- Etykieta △ ▾ ✕

- dodaj nową kolumnę -

Zastosuj

LISTA

Ustaw priorytet Dodaj etykietę Usuń etykietę Oznacz jako: Zmień widok Połącz Podobni nadawcy

<input type="checkbox"/>	Flaga	Lek	Data ogłoszenia	Kategoria	Ogłoszeniodawca	Etykieta
<input type="checkbox"/>		null, Mefedron	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>		Relanium5mg	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>		DEPRALIN 10 mg	2012-01-30 00:00:00	1	Kaminka	-
<input type="checkbox"/>		YASMINELE, Yasminelle	2012-01-30 00:00:00	1	koala	testowa
<input type="checkbox"/>		YAZ	2012-01-30 00:00:00	1	-	aaaa

Zagrozenie:

ID	191
Ranking	0
Adres URL	smc%20project/data/owi/owi_001.htm
Adres IP	

Ogłoszeniodawca:

Lokalizacja	Wrocław
Telefon	697605210

Leki:

Lek	Arthrotecu
Postać	tabletek
Ilość	10 i mniej tabletek

Lek	Arthrotecu
-----	------------

FILTR

HISTORIA

Liczba nowych zagrożeń: 167 Liczba zagrożeń ogółem: 197

Cel działania systemu:

- wykrycie w określonych źródłach treści, które mogą świadczyć o wystąpieniu określonych typów zagrożenia



Monitorowany typ zagrożenia:

- nielegalna sprzedaż leków



Możliwość rozszerzenia na inne typy zagrożeń
(sprzedaż broni, narkotyków, pedofilia itp.)



Model danych

Przykład: profil zagrożenia



- Pozyskanie informacji m.in. dotyczących:
 - ogłoszeniodawcy (nr telefonu, e-mail, adres)
 - oferowanego towaru i jego parametrów (nazwa, ilość, cena)
 - typu akcji (kupno, sprzedaż)

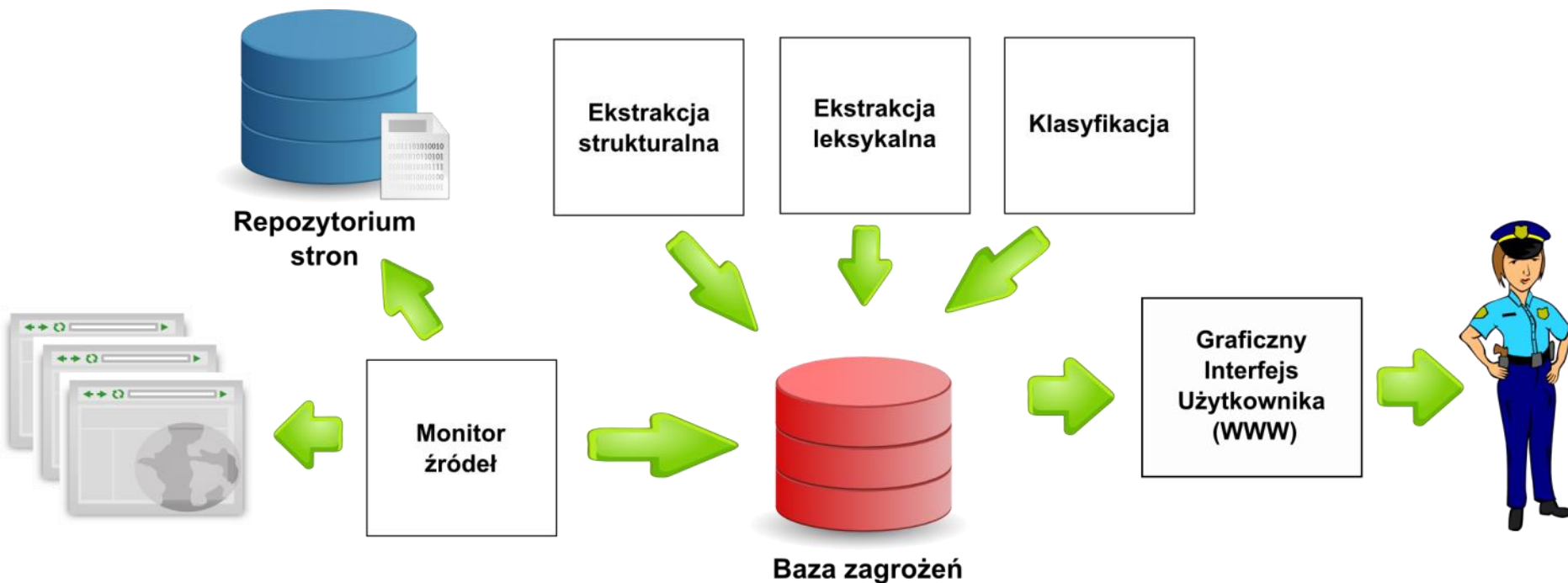


Aktualizacja profili zagrożeń

- Rozwiązanie uniwersalne – możliwość zastosowania do wykrywania różnych zagrożeń
- Adaptacja profili zagrożeń do zmian w otoczeniu, w tym modyfikacji:
 - bazy leków
 - adresów i struktury źródeł internetowych
 - określeń w języku potocznym



Architektura systemu



Rezultaty

Mechanizm monitorowania źródeł głębokiego i płytkiego Internetu

- równoczesna i ciągła obserwacja wielu źródeł
- pozyskiwanie elementów struktury dokumentu z drzewa DOM (tzw. dokumentów logicznych)





Rezultaty

Metody pozwalające na wykrycie zagrożeń z pozyskanych danych



- ekstrakcja informacji z tekstu o nieokreślonej strukturze
- automatyczne wykrywanie zagrożeń
- łączenie informacji pozyskanych z różnych źródeł
- klasyfikacja zagrożeń



Rezultaty

Prototyp systemu

- Platforma łącząca wszystkie komponenty od pozyskania danych ze źródeł do prezentacji użytkownikowi wykrytych zagrożeń
- Graficzny Interfejs Użytkownika
 - lista znalezionych zagrożeń
 - szczegółowe dane
 - analiza zagrożeń
 - eksport danych



GUI

WIDOCZNE KOLUMNY

- Flaga
- Lek
- Data ogłoszenia
- Kategoria
- Ogłoszeniodawca
- Etykieta

Panel narzędzi do zarządzania zagrożeniami

Ustaw priorytet Oznacz jako:

<input type="checkbox"/>	Flaga	Lek	Data ogłoszenia	Kategoria	Ogłoszeniodawca	Etykieta
<input type="checkbox"/>		null, Mefedron	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>		Relanium5mg	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>		DEPRALIN 10 mg	2012-01-30 00:00:00	1	Kaminka	-
<input type="checkbox"/>		YASMINELLE, Yasminelle	2012-01-30 00:00:00	1	koala	testowa
<input type="checkbox"/>		YAZ	2012-01-30 00:00:00	1	-	aaaa

Zagrozenie:

ID	191
Ranking	0
Adres URL	smc%20project/data/owi/owi_001.htm
Adres IP	

Ogłoszeniodawca:

Lokalizacja	Wrocław
Telefon	697605210

Leki:

Lek	Arthrotecu
Postać	tabletek
Ilość	10 i mniej tabletek

Liczba nowych zagrożeń: 167 Liczba zagrożeń ogółem: 197

FILTR

HISTORIA

Narzędzia do zarządzania wyglądem listy zagrożeń

Lista zagrożeń

Szczegóły dot. zagrożenia

Podsumowanie zagrożeń



Korzyści

- Pozyskane informacje
 - aktualne
 - wykryte automatycznie
 - zgodne z wymaganiami wyrażonymi w profilu
 - zintegrowane
 - uporządkowane, sklasyfikowane
 - zaprezentowane w przystępny dla użytkownika sposób **wraz z rankingiem**



Podsumowanie: możliwości

- Możliwość zastosowania w różnych dziedzinach (podobny sposób działania)
- „Wydobycie” informacji z tekstu o nieokreślonej strukturze
- Łączenie informacji pozyskanych z różnych źródeł
- Automatyczne wykrywanie zdefiniowanych wzorców w pozyskanych informacjach





<http://smc.kie.ue.poznan.pl>

Katedra Informatyki Ekonomicznej
Uniwersytet Ekonomiczny w Poznaniu

tel: +48 61 8543381

fax: +48 61 8543633

e-mail: sekretariat@kie.ue.poznan.pl

