

Gorączka danych – czas pionierów!



Witold@Abramowicz.pl

Copyright Witold Abramowicz

Lilith (1977 – 1984) – ETH

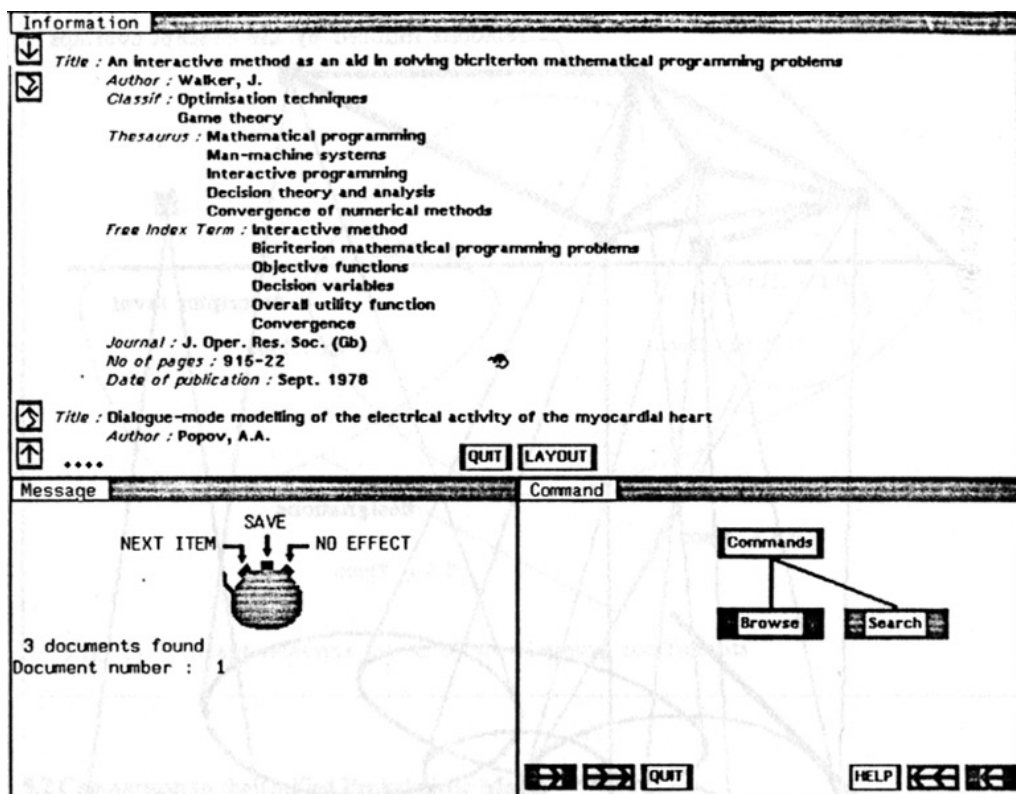


- 128 kB RAM
- 5 MB dysk

Konkurencja

- Apple II max 48 kB (1977)
- IBM PC – premiera 12. 8.1981

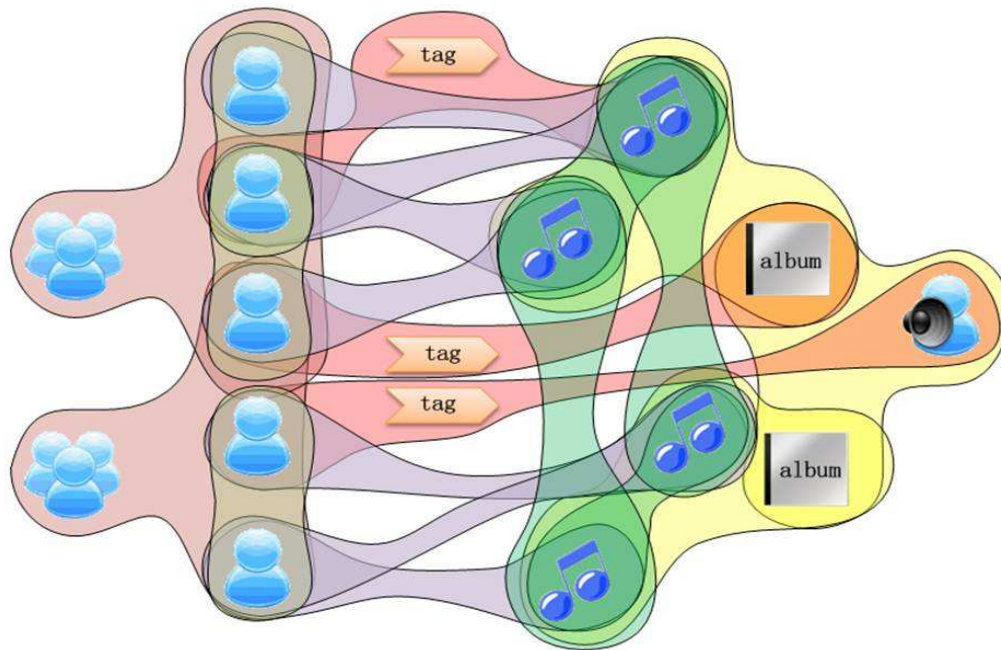
ATHENE – SDI System – ETH



Copyright Witold Abramowicz

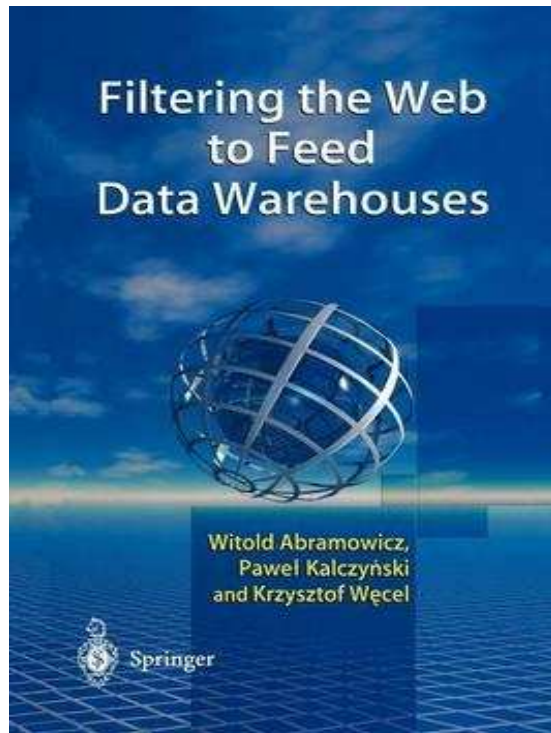
- Filtrowanie informacji na podstawie profili użytkowników wyrażających stałe potrzeby informacyjne
- Cel dostarczanie **opisów** bibliograficznych natychmiast po opublikowaniu
- **Volume 2MB**
 - brak sieci
 - dane na taśmach magnetycznych!
- Konkurencja: drukowanie fiszek z opisami bibliograficznymi

Hipergrafowy model hipertekstu – '90 HUB



- Modelowanie dokumentów i powiązań pomiędzy nimi w sieciach komputerowych – hipergraf
- **Volume** wszystkie dokumenty dostępne w sieci
- **Variety** – niestrukturalizowane teksty
- Konkurencja: 12.11.1991 opublikowanie projektu WWW – prostota!

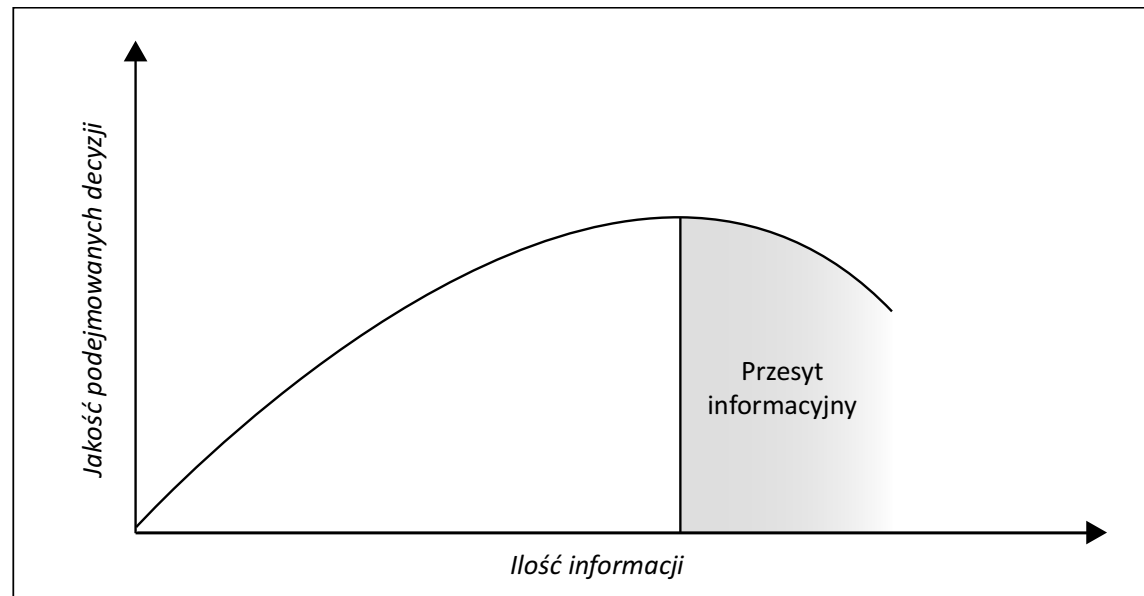
Filtrowanie informacji z Sieci zasilające hurtownie danych – 2002



- System transakcyjny zasilający hurtownie danych – analizy **mikroekonomiczne**
- Profil hurtowni podstawą do filtrowania informacji wzbogacających dane w hurtowni – analizy **makroekonomiczne**
- **Volume** wybrane źródła dostępne w Sieci
- **Variety** – dane strukturalizowane i niestrukturalizowane

Początek wieku – rok 2001

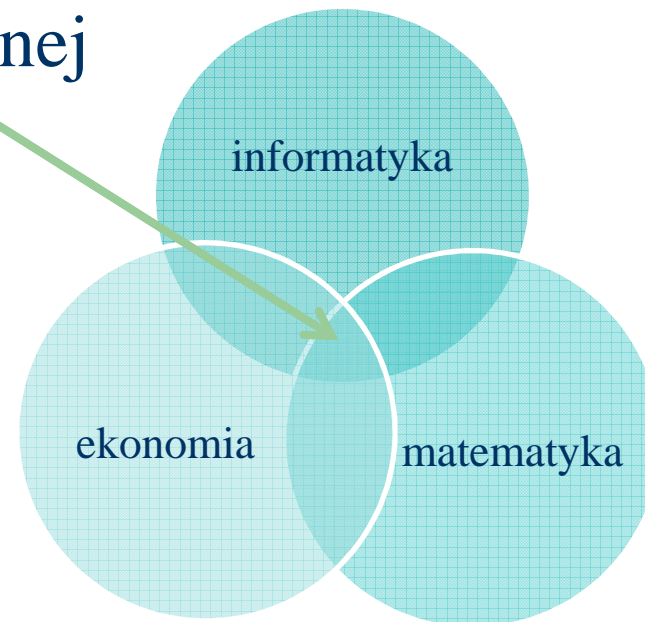
- Nagroda Nobla – Joseph E. Stiglitz – **asymetria informacji**



Początek wieku

Sukces informatyki ekonomicznej

dziedziny badań naukowych, kształcenia akademickiego oraz praktyki gospodarczej, wyodrębnionej z informatyki i poświęconej metodyce i pragmatyce zastosowania w ekonomii środków i narzędzi techniki komputerowej.



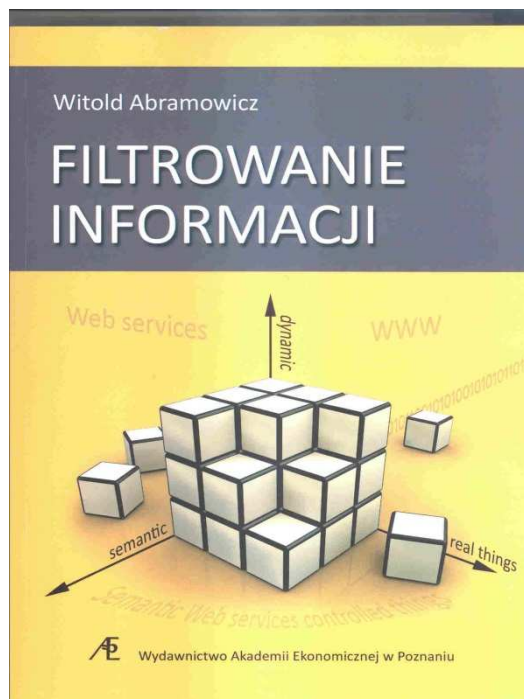
Początek wieku – rok 2003

- **Sukces** informatyki ekonomicznej
- Nagroda Nobla – Daniel Kahneman – ekonomia behawioralna
 - Adam Smith „Teoria uczuć moralnych”
 - Jeremy Bentham (1748-1832) decyzje ekonomiczne wynikają z uwarunkowań psychologicznych: homo oeconomicus – człowiek podejmujący **racjonalne** ekonomiczne decyzje, zawsze mający na względzie **interes własny**
 - John Maynard Keynes i Laszlo Garai (XX wiek) – wpływ **zdarzeń przypadkowych** i **nawyków** konsumentów na ich decyzje
 - Daniel Kahneman i Amos Tversky - „Analiza podejmowania decyzji w sytuacji zagrożenia” (1979)
 - Herbert Simon - „Ograniczona racjonalność” (1980)

Początek wieku

- Prawo Kryder's – Law of Mass Digital Storage – **wolumen** gromadzonych **danych podwaja się co roku** – czynniki sprzyjające
 - Prawo Moore'a
 - Moc obliczeniowa komputerów podwaja się co 18 miesięcy (teraz częściej twierdzi się: co 24 miesiące)
 - Podobna relacja stosunku mocy obliczeniowej (pamięci) do jej kosztu
 - Prawo Metcalfe'a – **użyteczność** systemu teleinformatycznego rośnie **proporcjonalnie do kwadratu liczby podłączonych użytkowników**

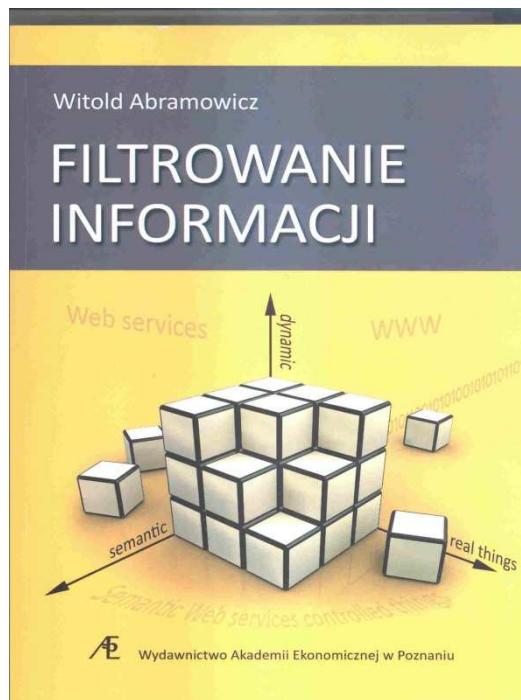
Filtrowanie informacji – rok 2008



Uwarunkowanie

- Gwałtowny wzrost wolumenu informacji, którymi można się wymieniać zwiększając ich wartość
- Podejmowanie decyzji w warunkach asymetrii informacji i jej nadmiaru
- Ekonomia behawioralna warunkowana egoistyczną racjonalnością oraz zależna od przypadkowości i sytuacji wyjątkowych

Filtrowanie informacji – rok 2008



Filtrowanie a wyszukiwanie

- **Volume** – Internet płytki a głęboki
- **Variety** – informacja heterogenna analizowana syntaktycznie w nadziei poprawności semantycznej

Big Data = 1xV: Volume

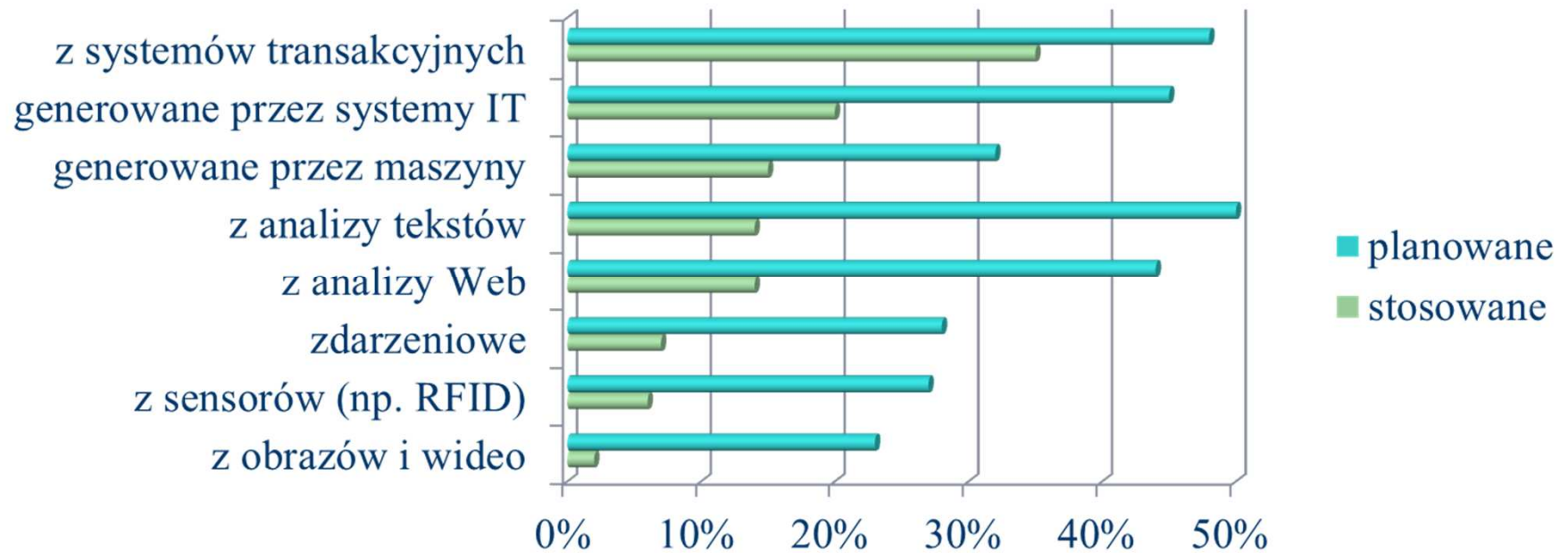
- Niedawno
 - Athene – 2MB
 - Hipergrafy < 100 GB
- Dzisiaj
 - Każdej minuty¹
 - W Google 2 mln. zapytań
 - 30 godzin nowych filmów na YouTube
 - 204 mln. nowych e-mails
 - 10 000 tweets
 - Google dziennie 24 PB
 - CERN 70 PB danych naukowych
- W historii ludzkości wypowiedziano 42 ZB

- 1 megabajt – 10^6 bajtów
- 1 gigabajt – 10^9 bajtów
- 1 terabajt – 10^{12} bajtów
- 1 petabajt – 10^{15} bajtów
- 1 eksabajt – 10^{18} bajtów
- 1 zettabajt – 10^{21} bajtów

UE próbuje określić wielkość Big Data

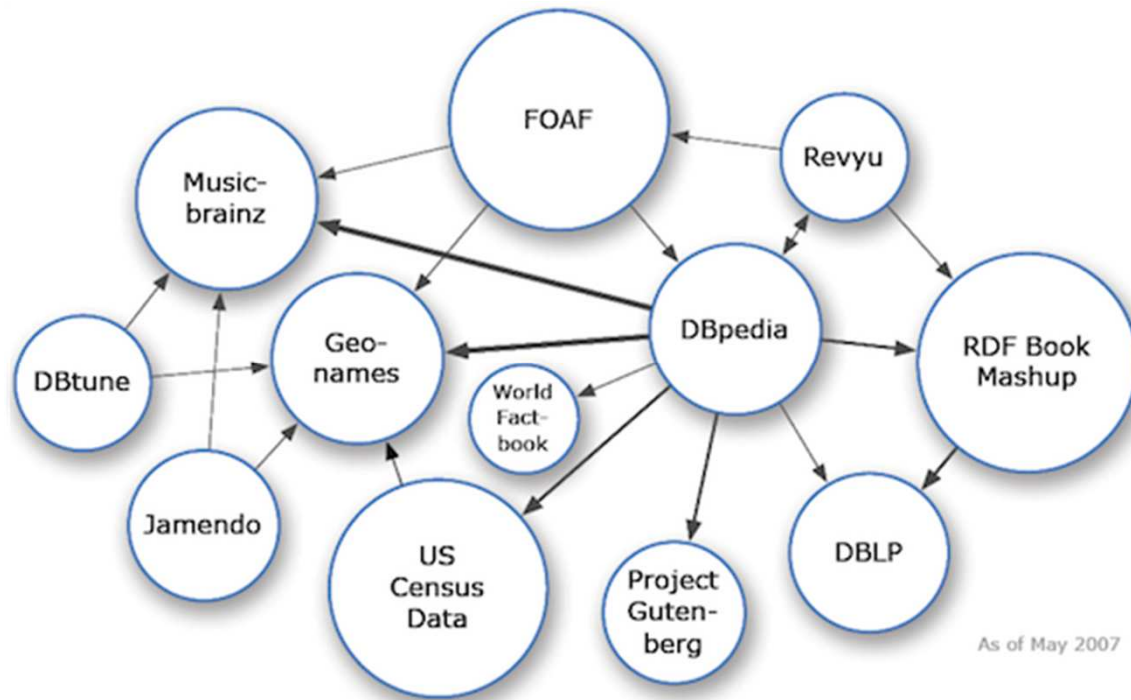
BIG DATA << Internet

Źródła Big Data % firm zainteresowanych

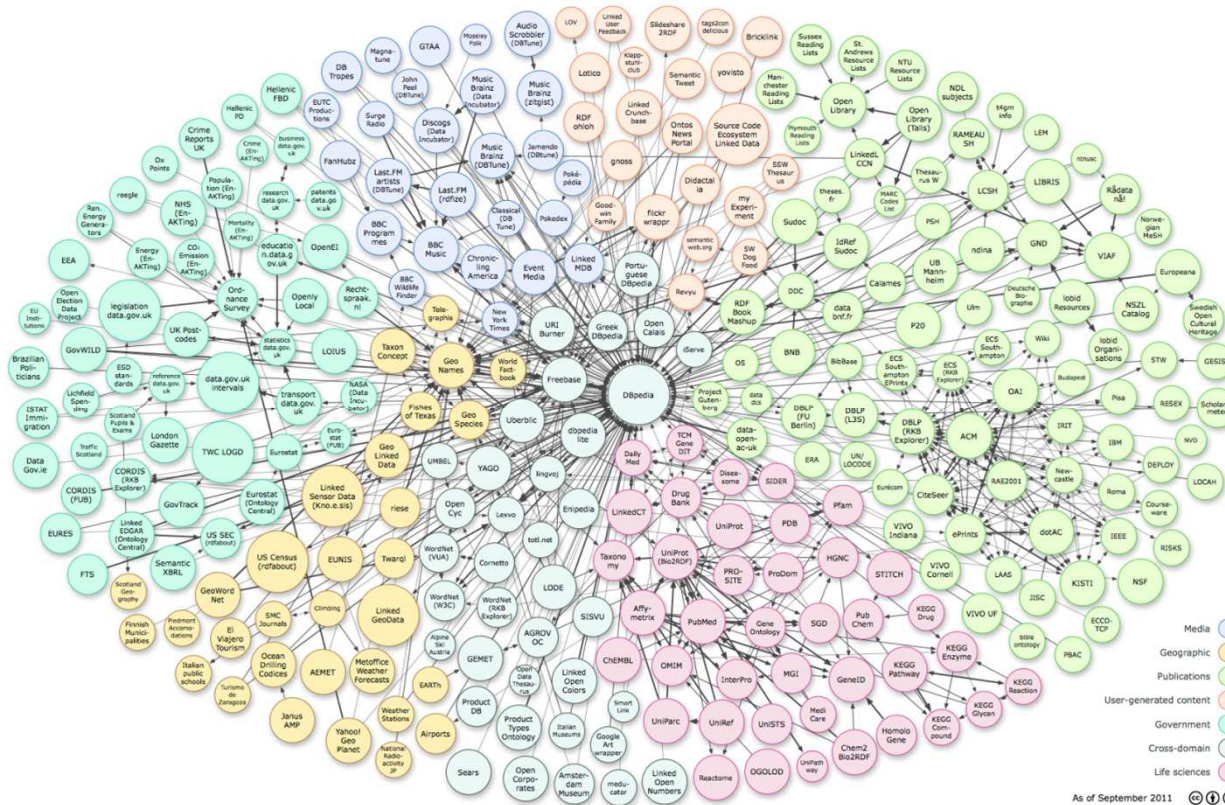


<http://www.silicon.de/41599384/pac-und-barc-dach-cxp-group/>

Źródła Big Data Linked Open Data 2007

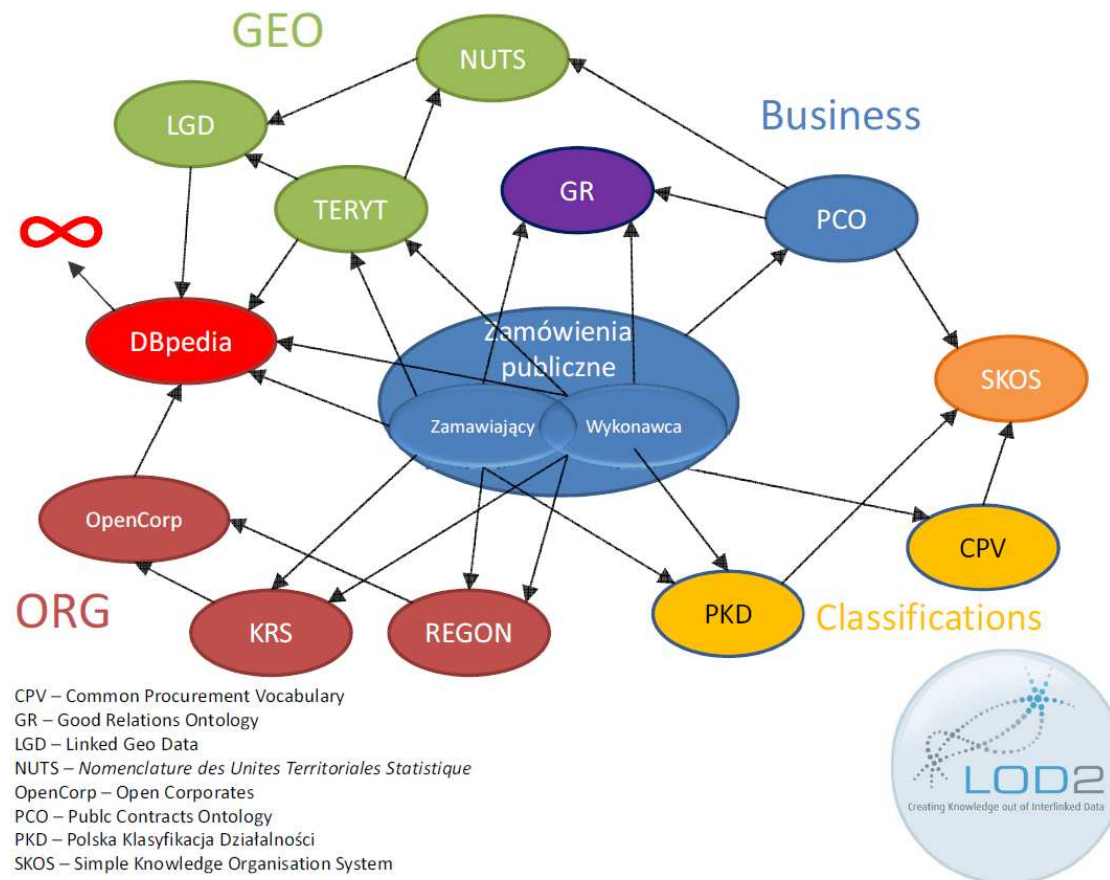


Źródła Big Data Linked Open Data



Copyright Witold Abramowicz

Projekt LOD 2

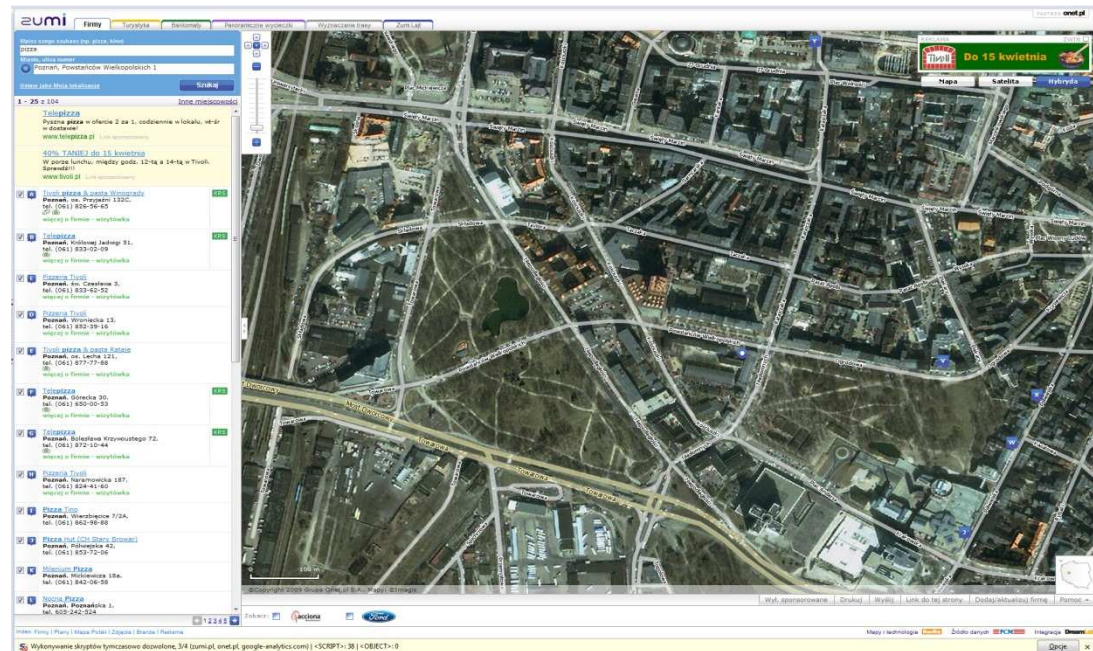


Big Data = 2xV: Variety

- Kryterium struktury
 - Strukturalizowane
 - Semistrukturalizowane
 - Niestrukturalizowane
- Kryterium zasobu
 - Płytki a głęboki Internet
 - Różnorodność źródeł
 - Repozytoria a strumienie

Big Data = 3xV: Veracity – Data Quality

- Spójność meta danych
- Prawdziwość
- Aktualność
- **Wiarygodność**
- Obiektywność a subiektywność
- Użyteczność i relewancja
 - Kognitywna
 - Lokalizacyjna
 - Czasowa
 - Interpersonalna
 - Ekonomiczna
 - Technologiczna
- ...



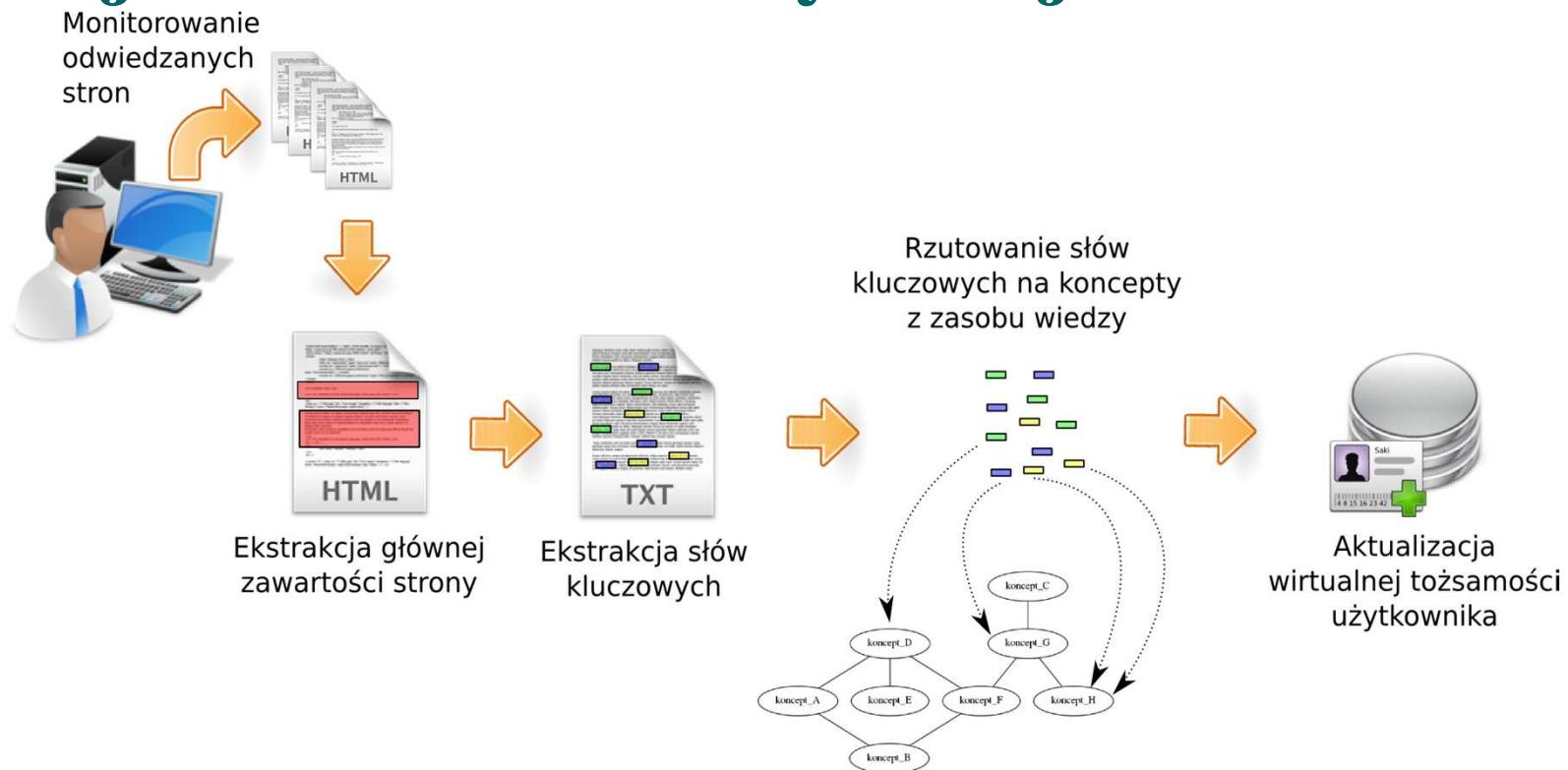
Big Data = 3xV: Veracity – Data Quality

- Spójność meta danych
- Prawdziwość
- Aktualność
- Wiarygodność
- Obiektywność a subiektywność
- Użyteczność i relewancja
 - Kognitywna
 - Lokalizacyjna
 - Czasowa
 - Interpersonalna
 - Ekonomiczna
 - Technologiczna
- ...
- Dane a **powiązania pomiędzy danymi**
 - Atrybuty powiązań podobne do atrybutów danych
- Jakość danych a jakość **źródeł danych**
 - Atrybuty źródeł podobne do atrybutów danych
 - Częściowe dziedziczenie atrybutów danych od atrybutów źródeł
 - Pamiętanie źródeł danych
- Wiedzieć
 - **a móc** wiedzieć
 - **a móc** wykorzystać
- ...

Big Data = 4xV: Velocity

- **Zmienność w czasie atrybutów źródeł, danych i ich powiązań, np. relewancji**
- Strumienie a zdolność do szybkości analizy
- Analiza *prawie* w czasie rzeczywistym
- Analiza w czasie rzeczywistym
- Analiza syntaktyczna a semantyczna
- Dane zasobem stałym
 - ale skalowalnym (Prawo Kryder's – Law of Mass Digital Storage)
 - cel analizy *chyży*
 - zakres analizowanych danych *chyży* – zdolność do koncentracji na analizie danych istotnych

Projekt EGO – identyfikacja tożsamości



Big Data = 4xV: Velocity

- **Zmienność w czasie atrybutów źródeł, danych i ich powiązań, np. relewancji**
- Strumienie a zdolność do szybkości analizy
- Analiza *prawie* w czasie rzeczywistym
- Analiza w czasie rzeczywistym
- Analiza syntaktyczna a semantyczna
- **Dane zasobem stałym**
 - ale skalowalnym (Prawo Kryder's – Law of Mass Digital Storage)
 - cel analizy *chyży*
 - zakres analizowanych danych *chyży* – zdolność do koncentracji na analizie danych istotnych

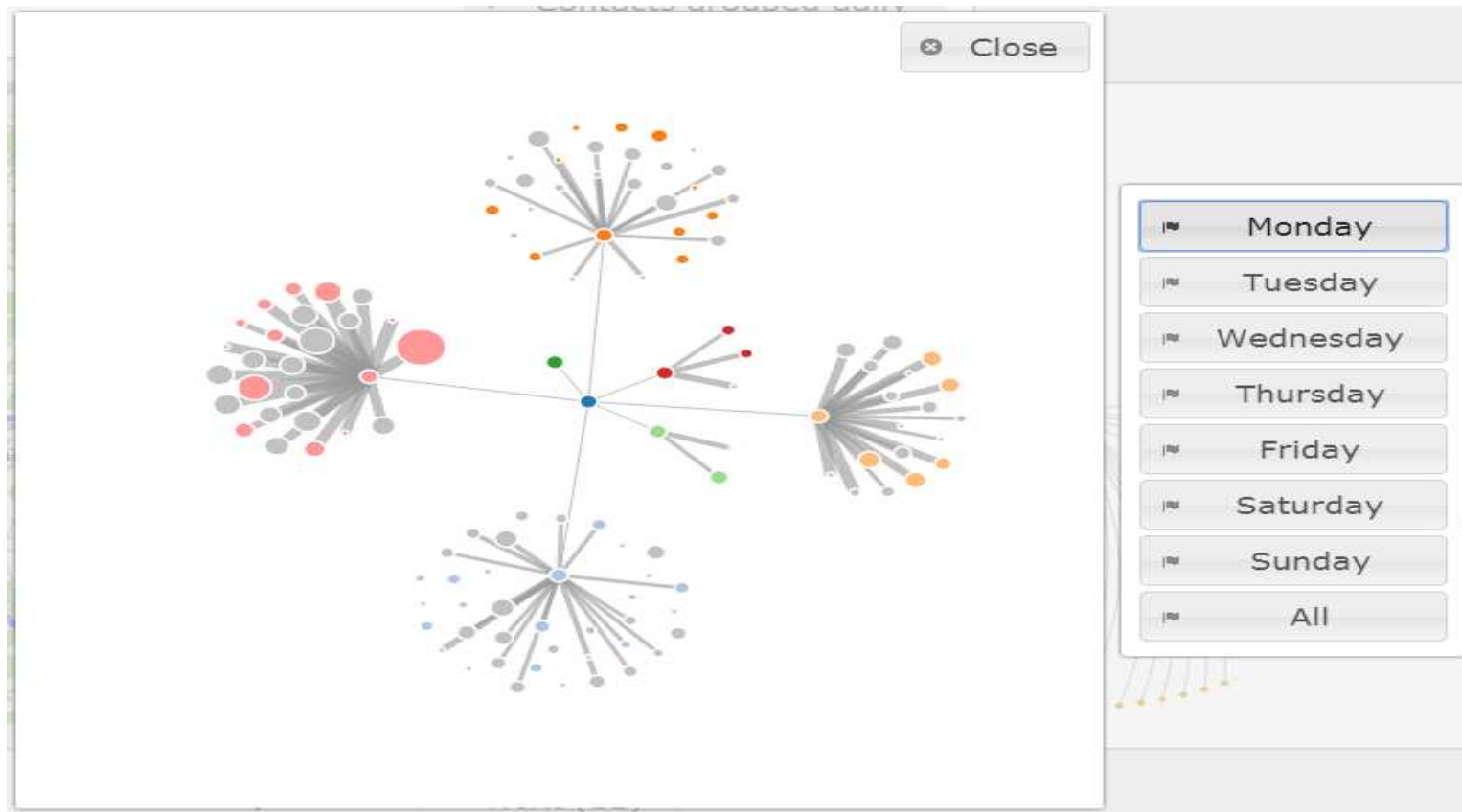
SIMMO System for Intelligent Maritime Monitoring

- Dane
 - o ruchach statków na oceanach i morzach > 100 tys. obiektów
 - wzbogacające
 - z baz danych
 - z Internetu
- Cele zmieniające się w zależności od
 - analizy danych
 - identyfikacja nowych, nieznanych źródeł danych
 - identyfikacja nowych, nieznanych przedmiotów analizy
 - sytuacji gospodarczej i politycznej

Big Data = 4xV: Velocity – wyzwania

- Gromadzenie, np. cloud, i przetwarzanie, np. in-memory computing
 - Danych
 - Powiązań pomiędzy danymi
- **Ochrona danych i prywatności podmiotów, których dotyczą dane**
- Bezpieczeństwo danych właściciela Big Data przetwarzającego np. w cloud
- Udostępnienie i **wizualizacja**
- Wykorzystanie w procesach biznesowych
- Wydzielenie z Big Data Smart Data
- ...

Projekt – usługi masowe – intensywność kontaktów pomiędzy klientami – wykrywanie *influencer*



Big Data = 4xV: Velocity – wyzwania

- Gromadzenie, np. cloud, i przetwarzanie, np. in-memory computing
 - Danych
 - Powiązań pomiędzy danymi
- Ochrona danych i prywatności podmiotów, których dotyczą dane
- Bezpieczeństwo danych właściciela Big Data przetwarzającego np. w cloud
- Udostępnienie i wizualizacja
- Wykorzystanie w procesach biznesowych
- Wydzielenie z Big Data **Smart Data**
- ...

Energetyka – BIG DATA

- Przykłady naszych projektów w HPI Future SOC Lab (Lipsk)
 - *Quasi Real-Time Individual Customer Based Forecasting of Energy Load Demand Using In Memory Computing*
 - *Forecasting of Energy Load Demand and Energy Production from Renewable Sources using In-Memory Computing*
 - *Prototype of an In-Memory Business Intelligence Solution for the Support of Forecasting of Energy Load Demand*
- Infrastruktura HPI Future SOC Lab
 - Sponsoring EMC, Fujitsu, HPI, SAP
 - Np. maszyna z 6 GB 1000 rdzeniowej pamięci
- Europejskie projekty badawcze
 - SAP HANA
 - Jedyne polski partner

Energetyka – wyniki

- Big Data

- Prognozowanie produkcji ze źródeł odnawialnych
 - Energia wiatrowa
 - Energia solarna
- Prognozowanie zachowania prosumentów
- Analiza informacji z danych społecznościowych

- Smart Data → Smart Grid

- Prognozowanie zachowania prosumentów w zakresie
 - Produkcji energii ze źródeł odnawialnych
 - Konsumpcji energii
- Rekomendacje racjonalnych zachowań prosumentów
- Przesłanki do negocjacji dostaw i zakupów energii przez Smart Grid
- Rozproszony Smart Grid, odległe od siebie np. sklepy i punkty usługowe, siedziby banków i TW, stacje benzynowe

Interaktywne określanie profili prosumentów

Menu

- Skrynia odbiorcza
- Punkty poboru
- Urządzenia odbiorcze
- Źródła wytwórcze
- Zużycie energii
- Prognozy
- Oferty

Mój profil


- Salon
- Sypialnia
- Łazienka
- Kuchnia**

Punkt poboru energii: 61-717 Poznań Kościuszki 107

kuchenka

Moc (W)	Liczba	Zużycie dobowe
4000	1	Usun

OK Przelicz



Lista urządzeń, dla których nie zdefiniowano parametrów odbiorczych

Nazwa	Zużycie	Lokalizacja	Rodzaj
kuchenka	Średni	Kuchnia	Odbiorcze Ustaw

Świat jest Smart (Data)

Przykład niemieckiego Federalnego Ministerstwa
Gospodarki i Technologii – obszary

- Przemysł
- Energia
- Mobilność
- Zdrowie

Big Data = 4xV: Velocity – wyzwania

- ...
- Internet Rzeczy
 - komunikacja
 - człowiek ↔ rzeczy
 - **rzeczy ↔ rzeczy**
 - Zastosowania
 - Industry 4.0, np. elastyczne systemy produkcyjne, *zdrowie maszyn*
 - Logistyka, np. RFID
 - ...
 - ...

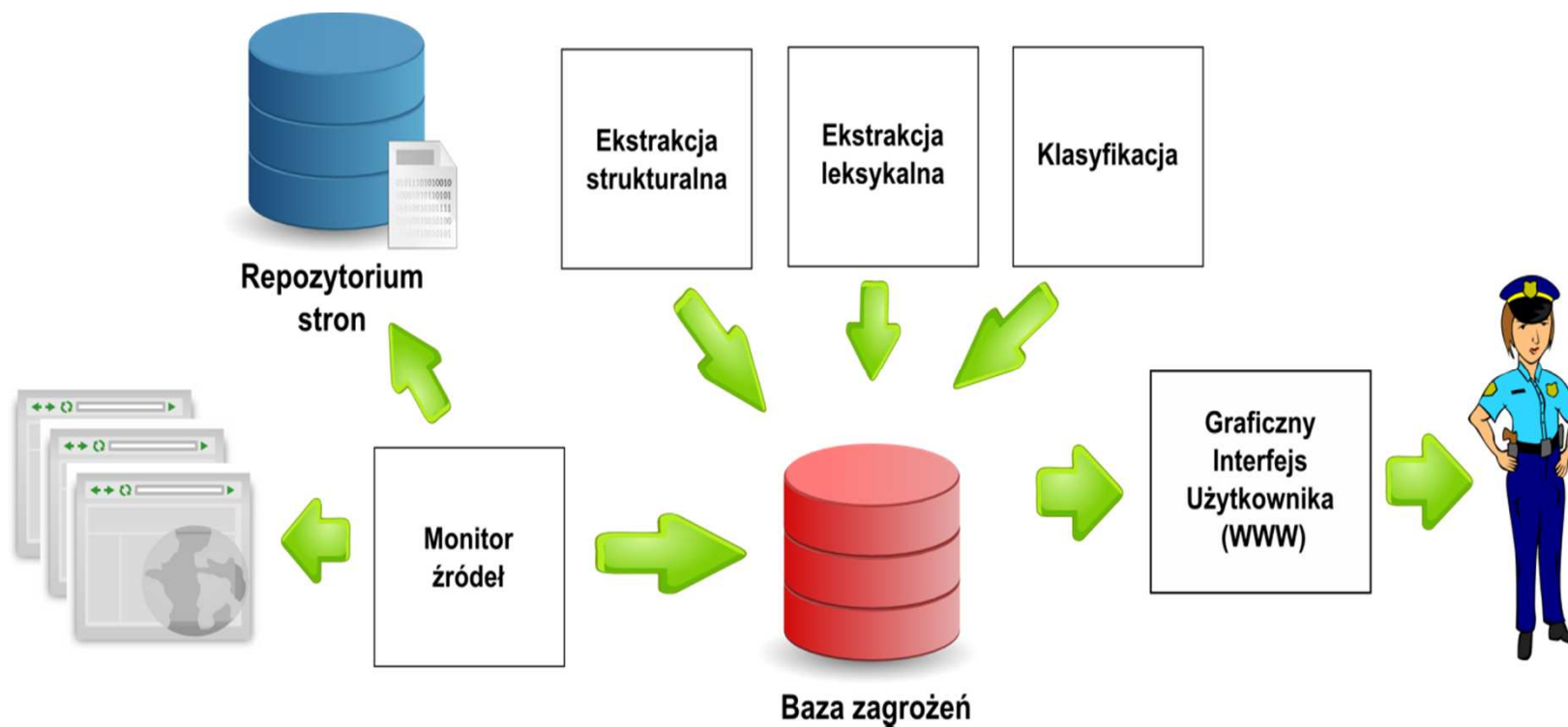
Obiecujące obszary zastosowań

- Energetyka – utilities
- Analiza mediów społecznościowych, np. racjonalizacja mix marketing, identyfikacja relacji interpersonalnych, np. influencer marketing
- Bezpieczeństwo cyberprzestrzeni, np. fraud detection and prevention
- Medycyna – spersonalizowana
- Zarządzanie cyklem życia produktów i usług, np. motoryzacja, ubezpieczenia, energetyka, ale także produkty kupowane ciągle – środki farmaceutyczne, nawozy

Semantyczny Monitoring Cyberprzestrzeni

- Energetyka – utilities
- **Analiza mediów społecznościowych**, np. racjonalizacja mix marketing, identyfikacja relacji interpersonalnych
- **Bezpieczeństwo cyberprzestrzeni**, np. fraud detection and prevention
- Medycyna – spersonalizowana
- Zarządzanie cyklem życia produktów i usług, np. motoryzacja, ubezpieczenia, energetyka, ale także produkty kupowane ciągle – środki farmaceutyczne, nawozy
- Center for Strategic and International Studies - **Cyberprzestępczość kosztuje globalną gospodarkę około o 445 mld USD rocznie¹**

Semantyczny Monitoring Cyberprzestrzeni



Semantyczny Monitoring Cyberprzestrzeni

WIDOCZNE KOLUMNY

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

- dodaj nową kolumnę -

✔ Zastosuj

LISTA

Arthrotecu, Arthrotecu, Ru ✕
misoprost 200mg, ru 486 ✕
Yasminelle, YASMINELLE ✕

Ustaw priorytet
Dodaj etykiety
Usuń etykiety
Oznacz jako:

Zmień widok
Połącz
Podobni nadawcy

☐	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
<input type="checkbox"/>		Mizoprostol, arthrotec	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>	🚩	Arthrotecu, Ru486	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>	🚩	zelixy, SLIMEX 15MG, meridi	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>	🚩	Betaferon	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>	🚩	ru 486, misoprost 200mg	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>		climazon	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>		Ru486, Arthrotecu, Ru 486	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>	🚩	c a l m e p a m 3 mg, C A L M E P A M (bromazepam) 3 mg	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>	🚩	Mife priston- RU 486, Miso prostol- Arth rotec 50mg (200mcg	2012-01-30 00:00:00	1	-	-
<input type="checkbox"/>		CURACNE 20 mg	2012-01-30 00:00:00	1	kupię curacne	-
<input type="checkbox"/>	🚩	izoteku	2012-01-30 00:00:00	1	annako79	-
<input type="checkbox"/>	🚩	Aknenormin	2012-01-30 00:00:00	1	99Aknenormin99	-
<input type="checkbox"/>	🚩	tretinex, aknenormin 20mg, izotek, roaccutan	2012-01-30 00:00:00	1	agata155	-
<input type="checkbox"/>		roaccutan, 20 mg	2012-01-30 00:00:00	1	irene	-
<input type="checkbox"/>		-	2012-01-30 00:00:00	1	ojj nie przesadzajcieee	-

1-20 of 197

FILTR

HISTORIA

Liczba nowych zagrożeń: 167

Liczba zagrożeń ogółem: 197

Gorączka danych – czas pionierów!



Witold@Abramowicz.pl

Copyright Witold Abramowicz