**37. Cele i metody pomiarów Internetu. Narzędzia i usługi pomiarowe. System MWING.**

Przydatność metod pomiarowych wynika z następujących powodów: obciążenie ruchem sieci IP jest trudne do przewidzenia, wynikające z różnorodności aplikacji, z których korzystają użytkownicy i braku kontroli dopuszczenia ruchu do sieci, obecne sieci IP są strukturami coraz bardziej rozbudowanymi, opartymi na różnych, wzajemnie się przenikających technologiach; dostępne i stosowane dotychczas modele analityczne nie umożliwiają dostarczenia wystarczającej wiedzy o stanie sieci; wprowadzenie jakości obsługi do sieci pakietowych wymaga wiedzy o stanie obciążenia ruchowego w sieci, co biorąc pod uwagę implikuje, iż systemy monitorowania i pomiarów w sieciach pakietowych mogą w istotny sposób wspierać pozyskiwanie wiedzy o stanie sieci niezbędnej dla potrzeb operatora, użytkownika, jak również dla mechanizmów sieciowych i aplikacji.

Metody pomiaru są dwie: Aktywna oraz pasywna

**Metoda pomiaru aktywnego** pozwala zmierzyć wartości metryk QoS (opóźnienie, zmienność opóźnienia, poziom strat pakietów, przepływności) poprzez wysyłanie specjalnych pakietów pomiarowych (ang. „probing packets”) w ramach monitorowanego strumienia ruchu. Metoda aktywna zakłada, że pakiety pomiarowe są przesyłane tą samą drogą i obsługiwane dokładnie w ten sam sposób, co pakiety użytkowników. Dzięki temu można uznać, że wartość metryk zmierzonych dla ruchu pomiarowego jest przybliżeniem wartości odpowiednich metryk dla pakietów użytkowników. Jednakże należy zwrócić uwagę, iż pomiar metodą aktywną pozwala jedynie estymować parametry populacji generalnej na podstawie wartości elementów próby (kolejnych wyników pomiaru).

Metoda aktywna zakłada, że nadajnik nadaje wysyłanym pakietom znacznik czasowy, który jest zapisywany w polu danych. Drugi znacznik czasowy jest nadawany przez odbiornik natychmiast po odebraniu pakietu pomiarowego. Porównanie wartości tych dwóch znaczników pozwala obliczyć wartości metryk związanych z opóźnieniem przesłania pakietu w sieci. Należy zwrócić uwagę na to, że kluczowym warunkiem poprawności pomiaru jest zapewnienie wspólnej podstawy czasu w nadajniku i odbiorniku, co można uzyskać np. synchronizując zegary korzystając z systemu GPS (Global Positioning System) lub protokołu NTP (Network Time Protocol).

Głównym problemem podczas stosowania metody aktywnej jest odpowiedni dobór wielkości ruchu pomiarowego oraz jego profilu. Ruch pomiarowy powinien być tak dobrany, aby można było uzyskać informacje na temat mierzonej ścieżki. Z drugiej strony wprowadzony ruch pomiarowy powinien być na tyle mały, aby wprowadzone dodatkowe obciążenie sieci nie zniekształcało wartości mierzonej metryki. Oprócz odpowiedniego doboru wielkości ruchu pomiarowego istotny jest dobór odpowiedniego profilu ruchu pomiarowego. Profil ten powinien być zbliżony do profilu ruchu mierzonego. W szczególności w praktyce stosuje się ruch pomiarowy o stałej szybkości bitowej lub generowany w sposób losowy. Istotną kwestią jest również dobranie odpowiedniej długości pakietów pomiarowych.

Ze względu na stosunkowo łatwą implementację i zarządzanie pomiarami, metoda aktywna jest wykorzystywana przez większość znanych narzędzi i systemów pomiaru parametrów QoS w komercyjnych i badawczych sieciach IP.

**Pasywna metoda pomiaru** polega na obserwacji pakietów w danym punkcie pomiarowym (lub w wielu punktach) i odpowiedniej analizie zarejestrowanego zapisu ruchu. W przypadku obserwacji w jednym tylko punkcie pomiar może dotyczyć charakterystyki przesyłanego ruchu, jak np. średniej szybkości bitowej ruchu na danym łączu. W przypadku, kiedy rozważamy dwa punkty pomiarowe, metoda pasywna może też być wykorzystana do pomiaru wartości metryk jakości przekazu pomiędzy tymi punktami.

Pasywna metoda pomiaru parametrów QoS wymaga rejestrowania (ang. „trace”) zaobserwowanych pakietów i odpowiadających im znaczników czasowych w dwóch punktach pomiarowych. Następnie, zbiory zebrane w poszczególnych punktach pomiarowych są przesyłane do serwera zarządzającego pomiarami, gdzie są analizowane. Porównując znaczniki czasowe nadane danemu pakietowi w dwóch różnych punktach pomiarowych można obliczyć czas przesłania tego pakietu pomiędzy rozważanymi punktami. Oczywiście, zapewnienie synchronizacji zegarów w poszczególnych punktach pomiarowych jest kluczowe dla uzyskania poprawnego wyniku.

Metoda pasywna, w odróżnieniu od metody aktywnej, nie wprowadza dodatkowego ruchu pomiarowego. Oczywiście, przesłanie zbioru z zapisem ruchu zaobserwowanego w punkcie pomiarowym do serwera zarządzającego także wiąże się z dodatkowym ruchem, ale nie obciąża on bezpośrednio monitorowanej ścieżki i może być przesyłany np. z niższym priorytetem obsługi niż ruch użytkowników. Należy zwrócić uwagę na to, że metoda pasywna pozwala na bezpośredni pomiar jakości przekazu uzyskiwanej przez pakiety użytkowników. Z drugiej strony, jej podstawową wadą jest trudność implementacji, co jest związane z koniecznością rejestracji całego ruchu w danym punkcie pomiarowym i stosunkowo skomplikowanym przetwarzaniem uzyskanych w ten sposób zbiorów.

<http://tnt.tele.pw.edu.pl/include/didactics/swus_lab4_pomiary.pdf>

Narzędzia do aktywnego pomiaru:

* Ping
* Traceroute
* Scriptroute

Narzędzia do pasywnego pomiaru:

* TCP dump
* Route table dumps
* Route views

Środowisko **Measurement Lab (M-Lab)** stanowi otwartą i rozproszoną platformę serwerową oferującą użytkownikom, w tym również naukowcom i badaczom, narzędzia do testowania szerokopasmowego dostępu do Internetu. Celem, dla którego ww. narzędzia powstały jest rozwój badań dotyczących sieci, a przede wszystkim udostępnienie jej użytkownikom użytecznej i wiarygodnej informacji na temat wartości parametrów opisujących szerokopasmowy dostęp do Internetu, z którego oni korzystają. Składa się z 45 serwerach, umieszczonych w 15 krajach.

**SpeedTest**- stosunkowo proste narzędzie utworzone w technologii Adobe Flash , które umożliwia szybkie sprawdzenie łącza poprzez badanie szybkości pobierania oraz szybkości wysyłania danych.

**Iperf**- jest to narzędzie za pomocą można robić pomiary związane z wydajnością łącza. Używa strumieni TCP.

I wiele innych tj. Argus, cFosSpeed, NetLimiter, OmniPeek, Wireshark, Sniffer

Szybkość transmisji/ przepływność- mierzona jest w bitach na sekundę

Miary dystansu:

* RTT(Round Trip Time)
* Długość ścieżki IP
* Długość ścieżki AS
* Dystans geograficzny

Pomiary Internetu można podzielić na trzy kategorie. Każdy z nich posiada następujące cechy, które można badać

* Wydajność
  + Opóźnienie
  + Przepustowość
* Strukturę
  + Topologie
  + Relacje
* Zachowanie
  + Stabilność

Celem takich pomiarów jest uzależniony od rzeczą, które chcemy zmierzyć. Przykładowe cele pomiarów Internetu to:

* Zmierzyć wydajność
* Zbadania protokołów
* Zidentyfikowania problemów, a co wiąże się z tym- zaproponowania rozwiązań.
* Powstania wzorców ruchu
* Uzyskania wiedzy o stanie sieci oraz zjawiskach w niej zachodzących

<http://www.cs.rutgers.edu/~badri/552dir/notes/w8meas-four.pdf>

MWING jest następca Winga, miał on spełniać następujące Cechy:

* Stabilność i dokładność pomiarów
* Skrócenie czasu pojedynczego pomiaru
* Zaawansowane możliwości harmonogramowania

badań

* Możliwość przeprowadzania równoległych

pomiarów

* Możliwość wykonywania badań w oparciu o

rozproszoną architekturę agentową

* Przetwarzanie i analiza danych offline
* Wsparcie dla analiz wykorzystujących metody

eksploracji danych (data mining)

System **MWING**(Multi-agent WING) jest wieloagentową platformę do zarządzania i wykonywania pomiaraów aktywnych i biernych w Internecie. Jest to system o konstrukcji systemu wieloagentowego, w którym agenty prowadzące pomiary mogą być zlokalizowane na dowolnym serwerze internetowym z adresem publicznym IP. System umożliwia badania dowolnych serwerów webowych o znanych publicznych adresach internetowych. System zarządza przygotowaniem środowiska sieciowego do pomiarów, w tym umożliwia instalację, konfigurację i zarządzanie poszczególnymi agentami na zdalnych węzłach sieciowych z lokalnymi bazami danych oraz zarządzanie węzłem centralnym z centralną bazą danych.

System obsługuje dowolne narzędzia pomiarowe użyte przez agenty. W szczególności mogą to być pomiary dokonane z wykorzystaniem agentów typu Wing. System zapewnia wspólną funkcjonalność wszystkim agentom pomiarowym, utrzymuje centralną bazę wyników pomiarów i synchronizuje aktywności agentów działających w sieci Internet.

Ma na cel pobieranie stron internetowych z serwera do użytkownika. MWING się składa ze czterech komponentów: Web Aplikacja(odpowiada za komunikacją między użytkownikami), Conroller, baza danych i zbiór agentów. Badanie polegało na pobierania pliku rfc1945 co kilka godzin.

Jego agenci znajdowały się w Las Vegas, w Gdańsku, W Gliwicach, Server harvard.edu, server cgsecurity.com, server uib.no

Opis przykładowego eksperymentu

• typ eksperymentu:

– aktywny,

• plan eksperymentu:

– losowanie próbki rozproszonych po całym świecie serwerów WWW,

– z każdego serwera transmisja zasobu o jednakowej wielkości,

– transmisja powtarzana 10 razy na dobę przez 20 tygodni,

przedmiot pomiaru:

• – czas pomiędzy momentami nadejścia pierwszego i ostatniego

pakietu zawierającego zasób,

• kryteria doboru rozmiaru zasobu:

– czas transmisji musi być znacząco większy w stosunku do

popełnianego błędu pomiarowego,

– transmisja nie może zbytnio obciążać sieci i serwerów.

**Identyfikacja topologii Internetu. Tomografia internetowa.**

Topologia sieci jest przedstawienie powiązań między bezpośrednio związanych peer-ów w sieci. W fizycznej topologii sieci, peer-y są porty urządzeń podłączonych przez fizyczne łącze transmisji. Topologia fizyczna odpowiada wielu logicznych topologii, każdy na innym poziomie abstrakcji. Na przykład, na poziomie IP, peer są gospodarze lub routerów na jeden IP hop od siebie. Zadanie topologie jest znalezienie struktury Internetu. Jest to bardzo trudno zadanie, ponieważ siec Internet cały czas rośnie.

**ISP( Internet Service Provider)-** Dostawca usługi internetowej. Organizacja non-profit lub

biznesowa udostępniająca usługę dostępu do sieci Internet

**POP( Points of Presence)**- Punkt dostępowy do sieci (Internetu) – fizyczny ośrodek liczący jeden lub kilka routerów (także innych urządzeń jak np. przełączników) dostawcy ISP, z którymi mogą łączyć się routery znajdujące się w sieciach jego klientów lub sieciach innych dostawców.

ISP zazwyczaj mają wiele POP

**Peering-** Wzajemna usługa dostarczania komunikacji do własnych klientów (nabywców tranzytowych)

**Tranzyt(** Transit) - ISP zapewnia dostęp do wszystkich znanych sobie innych części sieci (Internetu)

**IXP(Internet eXchange Point)** - Punkt wymiany ruchu internetowego – infrastruktura dla wymiany ruchu internetowego między sieciami dostawców ISP

**Sieci typu Tier-1**

* Sieci najwyższego rzędu - ogólnokrajowe lub ogólnoświatowe sieci szkieletowe (ang. *backbone*)
* Sieć użytkująca zasoby Internetu jedynie poprzez bezpłatny peering
  + Oznacza to peering z każdą inną siecią Tier-1
  + W rzeczywistości nieznane są umowy (płatna wymiana czy też nie)
* Ruch pomiędzy dwoma sieciami Tier-1 zależy od wzajemnej umowy
  + Zerwanie prowadzi do niemożności komunikacji pomiędzy częściami umowy
    - Zdarzało się kilka razy
    - Odtworzenie: powrót do peeringu lub opłacenie tranzytu
    - Nie uderza to zazwyczaj w ISP niższego rzędu

**Siec typu Tier- 2**

* Sieć ISP zaangażowanego głównie w peering z innymi sieciami, ale także korzystająca z tranzytu sieci Tier-1 (opłacająca tranzyt)
* Zazwyczaj sieci regionalne
* Najpowszechniejsze sieci w Internecie
  + Bywają większe niż sieci Tier-1
  + W regionie mogą pełnić rolę sieci Tier-1 (szkielet Internetu)
  + Mające wiele umów peeringowych są często najbliższe użytkownikom końcowym

**Siec typu Tier-3**

* Lokalni dostawcy Internetu
* Sieci które dostęp do Internetu zapewniają klientom wyłącznie poprzez zakup tranzytu od innych ISP
  + Zazwyczaj single-homed

**Hierarchiczna organizacja Internetu**

* *End system* (ES) – element sieci bez funkcji przekazywania ruchu
* *Intermediate system* (IS) – element sieci realizujący trasowanie (ruting) i inne funkcje przekazywania ruchu
  + *Intradomain IS* – komunikacja w ramach jednego fragmentu intersieci, tzw. systemu autonomicznego – ang. Autonomous System (AS)
  + *Interdomain IS* – komunikacja wewnątrz i poza danym systemem autonomicznym

**Internet jest zbiorem systemów autonomicznych (AS) określających zasady organizacji ruchu w całej sieci**

* ASy są dzielone na obszary (ang. *areas*)
* AS nazywany bywa domeną (ang. *domain*)

**System autonomiczny jest zbiorem prefiksów IP (adresó wsieci) pod kontrolą administracyjną jednego lub więcej operatora, w którym utrzymywany jest spójny JEDEN JASNO OKREŚLONY schemat trasowania (ang. *Routing policy*)**

Tomografia sieci jest to nauka oparta na informacji dostarczonych od końcowych punktów sieci. Według niej jest możliwe określić ścieżkę, które przechodzą dane w Internet poprzez analizy danych z węzłów brzegowych- urządzenie, z którego te dane pochodzą oraz to które je zażądało.

Informacja ta jest przydatna w celu opracowywania bardziej wydajnych sieci komputerowych. Wykorzystane są również do podniesienia jakości usług poprzez ograniczenie utraconych pakietów

Tomografia można podzielić na dwa rodzaje: Stratna oraz opóźnienie. Stratna ma na cel znalezienia stratnych linków w sieci poprzez wysyłania aktywnych próbek, z wielu różnych punktów w sieci(Internetu). Tomografia opóźnienia ma na cel znalezienia opóźnień łączach za pomocą próbek end-to-end. Dzięki temu można ograniczyć łączności z dużych opóźnień z powodów wąskich gardeł.

Inne zastosowanie tomografii jest znalezienia linków, które są wspólne dla wielu węzłów, co może spowodować powstania wąskiego gardła w przyszłości.

Źródło: <http://en.wikipedia.org/wiki/Network_tomography>

Uważa się pierwszą osobę, które zaczęła pracę nad tym zagadnieniem to Y. Vardi

**Efektywne i niezawodne pozyskiwanie zasobów w Internecie.**

W celu zdobycia informacji w sieci możemy korzystać z wielu metod i narzędzi. Większość z nich udostępniają różnego rodzaju portale internetowe. Za ich pośrednictwem uzyskamy dostęp do serwisów informacyjnych, katalogów i wyszukiwarek. Do dyspozycji mamy zarówno portale polskie jak i zagraniczne.

W celu zwiększenia prawdopodobieństwa znalezienia wyników można korzystać z cudzysłów(wtedy wyszukiwarka traktuje zawartość między „” jako całość i szukamy właśnie taki ciąg słów). Użycie operatorów + i – (+ oznacza że słowo musi wystąpić na stronie, a – że nie może, np. Republika, +serwis –kraj wtedy znajdziemy serwisy o nazwie Republika,) operatory jak link:, serwer:,site:, domain:, url: title: ,object: , filetype: i wielu innych

* Dobór słów ma znaczenia
* Każde słowo jest ważne
* Kolejność słów ma znaczenie

Zdecydowanie największy problem jest znalezienia multimedialnych zasobach w Internecie. Roboty zazwyczaj analizują strony pod kątem rozszerzenia plików. Wtedy są indeksowany nazwy plików, metadane, tekst dalszy i ewentualnie wielkość zdjęcia, rozmiar pliku, czas trwania nagrania. Są specjalne opcje w standardowych wyszukiwarek

**WYSZUKIWARKI**

Wyszukiwarki stanowią połączenie dwóch elementów wpływających na ich funkcjonalność. Po stronie użytkownika mamy program, który po wpisaniu słowa lub wyrażenia do wyszukania, wyświetli zapamiętane w bazie danych informacje o stronach, na których takie wyrażenie występuje. Uzyskujemy w ten sposób bardzo często dużą liczbę stron, które nie koniecznie związane są z poszukiwanymi przez nas informacjami, chociaż w większości przypadków szansa na uzyskanie właściwych odpowiedzi jest bardzo duża. Druga strona wyszukiwarki to tzw. robot. Robot samodzielnie wędruje po zasobach sieci i stara się zbierać informacje o napotkanych stronach, zapamiętując z nich słowa kluczowe, opisy oraz jak najwięcej informacji, które mogą być przydatne do określenia zawartości strony. Na tej podstawie tworzony jest tzw. index, który pozwala na szybkie wyświetlenie spisu stron, na których występują wyrażenia wprowadzone w pole wyszukiwarki. Opisy stron wyświetlone przez wyszukiwarkę są tym, co umieścił na stronie jej autor. Mogą zatem być one niepełne, nieobiektywne lub w przypadku gdy autor zapomniał umieścić opisu swojej strony, robot zapamięta początkowe wyrazy znajdujące się na stronie. Może to spowodować, że między treściami przez nas oczekiwanymi znajdą się także treści niepożądane.

**SERWISY INFORMACYJNE i BAZY DANYCH**

Szukając informacji na temat aktualności z świata gospodarki, polityki, wydarzeń kulturalnych czy po prostu pogody na kolejny dzień sięgnijmy do jednego z wielu serwisów portalu internetowego. Po wyszukaniu odpowiedniego odsyłacza po kilku kliknięciach myszą uzyskujemy potrzebne informacje. Skróty do najbardziej "gorących wiadomości" znajdziemy w najbardziej widocznym miejscu portalu. Duże portale prowadzą wiele specjalistycznych baz danych. Są one z reguły ograniczone do węższej tematyki, zawierają bardzo szczegółowe informacje i precyzyjnie opisują dane zagadnienie.

**KATALOGI**

Katalogi tworzone są przez zespoły redaktorów lub przez użytkowników sieci, dodających informacje o swoich stronach do katalogu. Katalog jest podzielony na kategorie, a strony związane z określoną grupą informacji umieszczane są w jednej z nich. Jeżeli tylko w katalogu odnajdziemy dziedzinę, która nas interesuje, to nie musimy martwić się o to czy zadaliśmy właściwe zapytanie (tak jak ma to miejsce przy posługiwaniu się wyszukiwarkami), bo od razu uzyskujemy informacje o wszystkich zarejestrowanych w katalogu stronach na dany temat. Przeglądanie katalogu można by porównać do przechodzenia między regałami biblioteki, na których umieszczono opisy znajdujących się tam książek. Gdy wybierzemy się do biblioteki w poszukiwaniu lektury „Szatan z siódmej klasy” to: odszukamy regał z napisem "lektury", potem spojrzymy na półkę z napisem "lektury dla szkoły podstawowej". Dbaniem o to, aby na właściwej półce biblioteki znalazła się właściwa książka zajmuje się bibliotekarz. Porządkowaniem katalogów stron internetowych zajmują się ich redaktorzy. Jedyną wadą katalogów internetowych jest to, że nie jest możliwe umieszczenie w nim na raz wszystkich stron WWW. Ponieważ każda umieszczana tam strona jest przeglądana i klasyfikowana przez redaktorów, to czynność ta musi jakiś czas potrwać. Jeżeli chcemy własną stronę umieścić w katalogu, powinniśmy zgłosić ją do redaktorów katalogu, aby została dodana we właściwe (łatwe do odnalezienia) miejsce.

**Multiwyszukiwarki i programy wyszukujące**

Multiwyszukiwarki to serwisy, przy pomocy których możemy wysłać zapytania do kilku wyszukiwarek jednocześnie. Program multiwyszukwarki zbierze odpowiedzi innych wyszukiwarek i zaprezentuje je nam w formie jednej listy.

Podobne zadanie jak multiwyszukiwarki spełniają zewnętrzne programy wyszukujące. W swojej pracy korzystają z informacji dostarczonych przez kilka sieciowych systemów wyszukiwawczych, a za pomocą dodatkowych mechanizmów wykonują porządkowanie, usuwanie duplikatów i sortowanie otrzymanych wyników. Z reguły program taki pozwala na stosowanie operatorów logicznych oraz określenie dodatkowych warunków wyszukiwania. Mogą być nimi np. ograniczanie liczby pożądanych rezultatów, weryfikacja znalezionych adresów, zachowywanie wyników wyszukiwania, stosowana wersja językowa, zmienna liczby wykorzystywanych mechanizmów wyszukujących, przeszukiwanie FTP, grup dyskusyjnych, adresów e-mail i wiele innych

**Wyszukiwarki specjalistyczne**

Niektóre wyszukiwarki zajmują się gromadzeniem informacji na temat określonych zasobów, np. plików programów, muzyki, zdjęć, filmów. Jeżeli potrzebujemy tego typu konkretnych informacji skorzystajmy z usług właśnie takich serwisów i wyszukiwarek.

Pamiętajmy także o tym, że większość "dużych" systemów wyszukiwawczych omówionych wcześniej, również oferuje możliwości wyszukiwania konkretnych zasobów. Mogą być nimi np. różnego rodzaju multimedia lub programy. Korzystając z wyszukiwarek musimy pamiętać o tym, że nie wszystkie z nich stosują identyczną składnie zapytać. Dlatego też jeżeli decydujemy się na wykorzystanie podstawowych mechanizmów wyszukiwawczych powinniśmy wpisywać zapytania we właściwych formularzach, a chcąc uzyskać bardziej szczegółowe wyniki zapytać musimy posługiwać się dodatkowymi operatorami opisywanymi w systemach pomocy poszczególnych wyszukiwarek lub formularzami wyszukiwania zaawansowanego.

Źródło: <https://www.google.pl/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCsQFjAA&url=http%3A%2F%2Fszkolazklasa20.nq.pl%2Findex.php%3Fsec%3Dpobierz_plik_szk%26id%3D3419&ei=uqfEUZy7LoTptQb2l4GIAQ&usg=AFQjCNF890GPwfmRc6eCd0XhKrpMcpXyew&sig2=kQrNUgv0BWMi_dbUUyd9NA&cad=rja>

Semantic Web

**Semantic Web** – projekt, który ma przyczynić się do utworzenia i rozpowszechnienia [standardów](http://pl.wikipedia.org/wiki/Standard) opisywania treści w [Internecie](http://pl.wikipedia.org/wiki/Internet) w sposób, który umożliwi maszynom i [programom](http://pl.wikipedia.org/wiki/Program_komputerowy) (np. tzw. [agentom](http://pl.wikipedia.org/wiki/Agent_%28programowanie%29)) przetwarzanie informacji w sposób odpowiedni do ich znaczenia. Wśród standardów Semantic Web znajdują się m.in. [OWL](http://pl.wikipedia.org/wiki/Web_Ontology_Language), [RDF](http://pl.wikipedia.org/w/index.php?title=Refused_Derived_Fuel&action=edit&redlink=1), [RDF Schema](http://pl.wikipedia.org/wiki/RDF_Schema) (inaczej RDFS). Znaczenia zasobów informacyjnych określa się za pomocą tzw. [ontologii](http://pl.wikipedia.org/wiki/Ontologia_%28informatyka%29).

Semantic Web jest wizją [Tima Bernersa-Lee](http://pl.wikipedia.org/wiki/Tim_Berners-Lee) (twórcy standardu [WWW](http://pl.wikipedia.org/wiki/World_Wide_Web) i pierwszej przeglądarki internetowej, a także przewodniczącego [W3C](http://pl.wikipedia.org/wiki/World_Wide_Web_Consortium)). W swoich założeniach Semantic Web ma korzystać z istniejącego protokołu komunikacyjnego, na którym bazuje dzisiejszy [Internet](http://pl.wikipedia.org/wiki/Internet). Różnica miałaby polegać na tym, że przesyłane dane mogłyby być 'rozumiane' także przez maszyny. Owo 'rozumienie' polegałoby na tym, że dane przekazywane byłyby w postaci, w której można by powiązać ich znaczenia między sobą a także w ramach odpowiedniego kontekstu.

Informacje przekazywane w ramach Semantic Web wymagałyby nie tylko samych [[dane}danych]], ale także informacji o tychże (tzw. meta-danych). To właśnie meta-dane zawierałyby sformułowania dotyczące relacji między danymi oraz prawa logiki, które można do nich zastosować. Dzięki temu można by:

* powiązać różne dane znajdujące się w [Internecie](http://pl.wikipedia.org/wiki/Internet) w ramach wspólnych jednostek znaczeniowych (np. strony dotyczące filmów, dziedzin nauki, kuchni francuskiej, etc.)
* rozróżnić dane, które dla maszyn są w tej chwili nierozróżnialne ze względu na identyczny zapis tekstowy (np. zamek - urządzenie do zamykania drzwi; urządzenie do łączenia w ustalonym położeniu elementów ubrania; okazała budowla mieszkalno-obronna)
* przeprowadzać na tychże danych wnioskowania, tzn. otrzymywać informacje na ich temat, które nie są zawarte [*explicite*](http://pl.wikipedia.org/w/index.php?title=Explicite&action=edit&redlink=1) (np. na podstawie danej "Ewa jest żoną Adama", możemy też dowiedzieć się, że Ewa jest kobietą, Adam mężczyzną, Adam jest mężem Ewy, żaden inny mężczyzna nie jest mężem Ewy, etc.)

Semantic Web zbudowany ma być na bazie już istniejących, wykorzystywanych i sprawdzonych [standardów internetowych](http://pl.wikipedia.org/w/index.php?title=Standard_internetowy&action=edit&redlink=1), nadbudowanych przez kilka kolejnych standardów. Ponieważ każdy kolejny standard nakłada się na kolejny, dotycząc innego poziomu abstrakcji, toteż ochrzczone zostały one mianem 'warstwowego placka' (ang. layer cake). Inne popularne ich określenie to 'semantyczny stos' (ang. semantic stack).

Kolejne warstwy placka prezentują się następująco (od dołu):

* [Unicode](http://pl.wikipedia.org/wiki/Unicode)
* [URI](http://pl.wikipedia.org/wiki/Uniform_Resource_Identifier)
* [XML](http://pl.wikipedia.org/wiki/XML) i [XML Schema](http://pl.wikipedia.org/wiki/XML_Schema)
* [RDF](http://pl.wikipedia.org/wiki/Resource_Description_Framework) i [RDF Schema](http://pl.wikipedia.org/wiki/RDF_Schema)
* [OWL](http://pl.wikipedia.org/wiki/Web_Ontology_Language)
* mechanizmy wnioskowania
* mechanizmy certyfikacji i zaufania

Dzięki standaryzacji możliwa stanie się swobodna wymiana danych oraz formalizacja i unifikacja dotychczasowej już zelektronizowanej wiedzy. Dzięki dobrze określonym strukturom reprezentacji [komputer](http://pl.wikipedia.org/wiki/Komputer) będzie mógł w łatwiejszy oraz bardziej trafny sposób wyszukiwać informacji lub nawet wnioskować w poszukiwaniu nowych faktów i powiązań.

Wyobraźmy sobie, że szukamy w Internecie informacji o jakiejś konkretnej grupie ludzi. Standardowym zachowaniem jest wpisanie w wyszukiwarkę internetową nazwiska tych osób oraz analiza wyników wyszukiwania pod kątem oceniania czy znaleźliśmy rzeczywiście szukane osoby. Jeśli jednak wyszukujemy w informacjach opisanych siecią semantyczną możemy pytać o osoby wskazując jakiś konkretny, bardzo szczegółowy fakt, np. "dyrektorzy przychodni lekarskich" i w wyniku otrzymamy kompletną listę spełniającą zapytanie.