**Pytania ITI**

**23. Architektura Internetu i usług internetowych. Systemy webowe i P2P**

**Architektura Internetu** – jest to struktura jaka istnieje w obecnie istniejącym Internecie. Zdefiniowana jest poprzez systemy autonomiczne, czyli zbiór IP pod kontrolą administracyjną jednego lub wielu operatorów.

Ze względu na AS i ich zbiorów Internet możemy podzielić na następujące typy sieci:

* Tier-1 – ogólnokrajowe lub ogólnoświatowe, komunikują się z innymi Tier-1 oraz dają możliwość wykupienia transferu dla sieci Tier-2
* Tier-2 – Zazwyczaj sieci regionalne, są one najpopularniejsze w sieci i zazwyczaj oni są dostawcami końcowymi Internetu, posiadają wiele umów.
* Tier-3 – Sieci lokalne, dostęp do sieci zapewniają wyłącznie poprzez dostęp do transfer z innymi Tier.

**Usługa internetowa** – usługa świadczona przez sieć telekomunikacyjną, a w tym sieć komputerową, w szczególności Internet.

Architektura usług internetowych - przykład:

SOAP – protokół zdalnego dostępu do obiektów z wykorzystaniem XML do kodowania wywołania.

**<soap:Envelope**

xmlns:soap="http://www.w3.org/2001/12/soap-envelope"

soap:encodingStyle="http://www.w3.org/2001/12/soap-encoding"**>**

**<soap:Body** xmlns:m="http://www.example.org/stock"**>**

**<m:GetStockPrice>**

**<m:StockName>**IBM**</m:StockName>**

**</m:GetStockPrice>**

**</soap:Body>**

**</soap:Envelope>**

**Systemy webowe** – strona bądź serwis internetowy, który wraz z użyciem sieci udostępnianie dynamicznie treść, pozwalają także na interakcje z użytkownikiem.

Najważniejszym aspektem aplikacji webowej jest jej fizyczna instalacja serwerowa, na której wykonywane są wszelkie obliczenia i operacje.

**P2P** – są to systemy, w których realizacja usług poprzez system rozproszony oparta jest o wiele pojedynczych węzłów, z których każdy zdolny jest do realizacji części bądź wszystkich usług. Często role węzłów pełnią komputery domowe. Systemy takie są o wiele bardziej skalowalne niż systemy klient-serwer oraz często są wykorzystywane do wymiany plików – np. torrenty.

**24. Protokoły HTTP i P2P. Charakterystyka ruchu internetowego. Prawo potęgowe w Internecie.**

**HTTP** – Jest to protokół bezstanowy uwzględniający specyfikację systemu WWW. Komunikacja odbywa się według modelu bezpołączeniowego – żądanie klienta o zasób, serwer przekazuje żądane zasoby, zamknięcie połączenia.

Protokół ten określa format żądań i odpowiedzi, gdzie przykładem może być:

GET nazwa\_pliku.jpg http/1.0 [pusty wiersz]

Składa się on z następujących komend:

* PUT – umieszczenie zasobu na serwerze
* GET – pobranie zasobu z serwera
* POST – przekazanie danych, które mają zmodyfikować stan istniejącego zasobu
* DELETE – usuwa zasób
* HEAD – służy do sprawdzenia dostępności zasobu

**P2P** – są to systemy, w których realizacja usług poprzez system rozproszony oparta jest o wiele pojedynczych węzłów, z których każdy zdolny jest do realizacji części bądź wszystkich usług. Często role węzłów pełnią komputery domowe. Systemy takie są o wiele bardziej skalowalne niż systemy klient-serwer oraz często są wykorzystywane do wymiany plików – np. torrenty.

**Ruchem teleinformatycznym** nazywamy przepływ zgłoszeń, połączeń i wiadomości w sieciach pakietowych. W przypadku sieci jednousługowej ruch ma charakter homogeniczny, natomiast w przypadku sieci wielousługowej jest to ruch heterogeniczny.

Elementami ruchu teleinformatycznego są:

* Ruch generowany przez transmisję danych masowych. Długi czas połączenia, przesył długich wiadomości, duże odstępy pomiędzy połączeniami danego użytkownika
* Ruch generowany przez transmisję poczty elektronicznej. Krótkie wiadomości głownie tekstowe, czasem zawierające załączniki w postaci plików, przesyłane w bardzo długich odstępach czasu.
* Ruch generowany przez usługę zdalnego terminala, np. telnet. Krótkie wiadomości, przesyłane w krótkich odstępach czasu
* Ruch generowany przez usługe WWW. Krótkie wiadomości, niewielkie pliki multimedialne, przesyłane w bardzo krótkich odstępach czasu, po których następują długie okresy braku aktywności
* Ruch generowany przez system multimedialny. Przesyłanie bardzo krótkich wiadomości lub bardzo długich wiadomości wideo, głosu i obrazu lub ich kombinacji. Cechą charakterystyczną jest generowanie ruchu w stałych odstępach czasu.

**Prawo potęgowe w Internecie**

W wielu dziedzinach praktycznych obserwuje się, że zależność pomiędzy dwiema wielkościami empirycznymi spełnia tzw. prawo potęgowe, tzn. że przyjmuje postać:

**Y=xw**

Prostym "szkolnym" przykładem tej zależności jest funkcja kwadratowa (w=2), ale w ekonomii i zarządzaniu szczególne znaczenie mają prawa potęgowe z wykładnikiem ujemnym (funkcje malejące), dotyczące wielkości opisujących rozmiar bądź skalę zjawisk.

W kwestii Internetu odnosi się to do szybkości rozrostu np. sieci społecznościowych.

**25. Systemy Gridowe i klastrowe. Problemy eksploatacyjne i rozwojowe**

**Klaster** – grupa połączonych jednostek komputerowych, które współpracują ze sobą w celu udostępnienia zintegrowanego środowiska pracy. Komputery wchodzące w skład klastra (będące członkami klastra) nazywamy węzłami (ang. node).

Zalety

* Skalowalność – łatwo rozbudować dodając nową jednostkę (węzeł)
* Dostępność – awaria jednego węzła nie powinna wpłynąć na działanie jego pracy
* Anonimowość węzłów

Problemy

* Licencje na oprogramowanie – zazwyczaj oznacza potrzebę kupna licencji na każdy węzeł osobno
* Brak pojedynczego systemu – jest czynnikiem utrudniającym zarządzającego klastrem
* Niedostosowane oprogramowanie – bardzo często oprogramowanie klastrowe wymaga specjalnego przerobienia oprogramowania do możliwości uruchomienia na wielu węzłach.

**Grid** - System, który integruje i zarządza zasobami będącymi pod kontrolą różnych i połączonymi siecią komputerową, używa standardowych, otwartych protokołów i interfejsów ogólnego przeznaczenia oraz dostarcza usług odpowiedniej jakości. Celem technologii gridowej jest stworzenie prostego, lecz mimo to wielkiego i potężnego, wirtualnego komputera z ogromnej ilości połączonych, niejednorodnych systemów współdzielących różnego rodzaju zasoby.

Zalety

* Równomierne wykorzystanie zasobów
* Możliwość spożytkowania wolnych mocy obliczeniowych połączonych maszyn

Problemy

* Synchronizacja – potrzeba jednoczesnego dostępu między innymi do baz danych oraz łączenia wyników gdzie opóźnieniem może być szybkość przesyły danych przez Internet może być kłopotliwe
* Skalowalność zadania – Nie wszystkie zadania da się podzielić na obliczanie na wielu jednostkach

**26. Definicje i szacowanie dystansu w Internecie.**

Miary dystansu w internecie:

* RRT - minimalny czas na przesłanie sygnału w obu kierunkachod nadawcy do odbiorcy i od odbiorcy do nadawcy
* dystans geograficzny
* długość ścierzki IP
* AS - system autonomiczny to zbiór adresów IP pod wspólną administracyjną kontrolą.

Szacowanie dystansu:

* wskażnik sąsiedstwa AS
* wskażnik sąsiedstwa geo serwerów
* hybryda wskaźników

**27. Podstawowe problemy i metody odkrywania Internetu** (moje domysły, lepiej coś niż nic nie mówić)

Problemem odkrywania Internetu nazywamy między innymi próbę zidentyfikowania zależności pomiędzy zasobami i strukturą sieci. Gdzie ich identyfikacja może odbywać się poprzez analizę sieci AS.

Podstawione problemy

* Konfiguracja sieci aby nie odpowiadała na pingi itp.
* Serwery PROXY i im podobne ukrywają za sobą prawdziwą strukturę Internetu.

Przykłady metody

* Ping
* Traceroute

**28. Identyfikacja topologii Internetu. Tomografia internetowa**

**Topologia Internetu** – jest to struktura powiązań między węzłami w sieci. Zadaniem topologii jest znalezienie struktury Internetu, jest to zadanie trudne bo sieć cały czas rośnie. Strukturę Internetu możemy nazwać inaczej architekturą Internetu.

**Architektura Internetu** – jest to struktura jaka istnieje w obecnie istniejącym Internecie. Zdefiniowana jest poprzez systemy autonomiczne, czyli zbiór IP pod kontrolą administracyjną jednego lub wielu operatorów.

Ze względu na AS i ich zbiorów Internet możemy podzielić na następujące typy sieci:

* Tier-1 – ogólnokrajowe lub ogólnoświatowe, komunikują się z innymi Tier-1 oraz dają możliwość wykupienia transferu dla sieci Tier-2
* Tier-2 – Zazwyczaj sieci regionalne, są one najpopularniejsze w sieci i zazwyczaj oni są dostawcami końcowymi Internetu, posiadają wiele umów.
* Tier-3 – Sieci lokalne, dostęp do sieci zapewniają wyłącznie poprzez dostęp do transfer z innymi Tier.

Najczęściej w celu identyfikacji topologii korzysta się z usługi traceroute.

**Tomografia internetowa** – jest to nauka oparta na informacji dostarczonych z końcowych punktów sieci. Według niej możliwe jest określenie ścieżek, poprzez analizę danych węzłów brzegowych – tego który posiada zasób i tego który go żąda.

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

Wyznaczyć możemy dwie strategie

* Stratna – ma na celu znalezieniu stratnych linków w sieci poprzez wysyłanie aktywnych próbek z wielu różnych punktów w sieci.
* Opóźnienia – ma na celu znalezieniu opóźnień łączących za pomocą próbek.

**29. Pomiar, szacowanie i predykcja czasu transmisji danych w Internecie**

Miary dystansu w internecie:

* RRT - minimalny czas na przesłanie sygnału w obu kierunkachod nadawcy do odbiorcy i od odbiorcy do nadawcy
* dystans geograficzny
* długość ścierzki IP
* AS - system autonomiczny to zbiór adresów IP pod wspólną administracyjną kontrolą.

Szacowanie dystansu:

* wskażnik sąsiedstwa AS
* wskażnik sąsiedstwa geo serwerów
* hybryda wskaźników

**30. Zapewnienie jakości usług w Internecie. Metody i rozwiązania.**

Jakość usług transmisji (QoS) oraz jakoś usługi WWW (QoWS) możemy podzielić na:

* Wydajność
* Niezawodność
* Dostępność
* Bezpieczeństwo

Sposoby zapewniania jakości usług WWW

* zwiększenie pamięci i liczby procesorów serwera
* zwiększenie liczby serwerów w skali lokalnej
* zwiększenie liczby serwerów w skali globalnej
* Wykorzystanie rozproszonych i klastrowych systemów webowych
* Stosowanie Proxy, cache

Przyspieszenie działania przeglądarek

* Należy wysyłać żądania tak szybko jak to tylko jest możliwe
* Stosować strumieniowanie – przeglądarki nie są agresywne w tym względzie, a powinny
* Należy wykorzystywać otwarte już połączenia

Mechanizmy poprawiające:

* Regulatory ruchu webowego - kształtowanie ruchu i aktywna kontrola nad ruchem „niepożądanym”
* Urządzenia buforujące dane webowe - oszczędzanie pasma przez zaniechanie
* transmisji
* Globalnie rozmieszczone serwery z zawartością internetową - np. sieć serwerów Akamai - fizyczne zbliżenie zawartości Internetu do użytkowników
* kompresja HTML zmniejsza o 24% ruch HTTP - oszczędność czasu i pieniędzy
  + Serwer WWW ma standardowo wyłączoną opcję kompresji danych
  + Nagłówki HTTP nie mogą być kompresowane - należy używać jak najmniejszych nagłówków
  + Strony WWW powinny być małe
  + Duża liczba małych grafik zabija wydajność - każda grafika to nagłówek

**31. Transakcje webowe i ocena ich wydajności**

**Transakcją webową** nazywamy przepływ zgłoszeń, połączeń i wiadomości pomiędzy witryną sieci web a przeglądarką.

Transakcje webowe możemy podzielić na

* Transakcje statyczne – HTTP, Obrazki, HTML
* Transakcje dynamiczne – PHP, JSP, HTTP-POST
* Transakcje bazodanowe – SQL, Cookies
* Transakcje aplikacji – JavaScript, DOM, CSS, AJAX

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

Obecnie najlepszym sposobem pozyskiwania zasobów jest wyszukiwanie informacji z użyciem przeglądarki Google. Za wyjątkiem standardowej daje ona nam możliwość wyszukiwania Obrazków, map, Dokumentów naukowych i wielu innych. Niestety czasami wyniki nie dają pożądanych rezultatów, tak więc aby zwiększyć trafność wyników, a tym samym efektywność pozyskiwania zasobów możemy użyć:

* Operatorów – „”, +, -, link:, serwer: , site:, url:, title:, filetype: i wielu innych
* Rozpoczynać wyszukiwane frazy od bardziej istotnych słów oraz pomijać popularne słowa

**33. Model serwera WWW. Algorytmy dostępu i szeregowania żądań HTTP w serwerze WWW**

**Serwer www** – program działający na serwerze internetowym, posługujący żądania protokołu komunikacyjnego HTTP. Z serwerem WWW łączy się, poprzez sieć komputerową, przeglądarka internetowa, będąca jego klientem, aby pobrać wskazaną stronę WWW. Modelem serwera WWW jest model Klient-Serwer.

**Klient-Serwer** – architektura systemu komputerowego, umożliwiająca podział żądań. Polega na ustaleniu, że serwer zapewnia usługi dla klientów, zgłaszających do serwera żądania. W architekturze klient-serwer zarządzanie zasobami jest po stronie serwera.

**Podział pracy**

* Klient – jest aktywny, wysyła żądania do serwera, oczekuje na odpowiedź serwera
* Serwer – jest pasywny, czeka na żądania od klientów, przetwarza je, a następnie wysyła odpowiedź.

**HTTP** – Jest to protokół bezstanowy uwzględniający specyfikację systemu WWW. Komunikacja odbywa się według modelu bezpołączeniowego – żądanie klienta o zasób, serwer przekazuje żądane zasoby, zamknięcie połączenia.

Szeregowanie żądań http następuje poprzez sieć kolejkową, czyli układ złożony z węzłów, w których znajdują się pojedyncze lub zrównoleglone systemu obsługi masowej. Sieć kolejowa może być

* Otwarta – nowe zadania mogą zarówno napływać jak i obecne opuszczać
* Zamknięta – zadania cyrkulują w sieci bez możliwości jej opuszczenia

Ponadto możemy wyróżnić następujące typy sieci:

* M/M/1 – kolejka z pojedynczym stanowiskiem obsługi i maksymalną M pojemnością kolejki, a blokada następuje dla M+1 zadań
* M/M/N – kolejka z N stanowiskami obsługi i maksymalną pojemnością M kolejki, a blokada następuje przy M\*N+1 zadań

**34. Metody i algorytmy dystrybucji żądań HTTP w klastrach i rozporoszonych systemach webowych.**

**Klaster** – grupa połączonych jednostek komputerowych, które współpracują ze sobą w celu udostępnienia zintegrowanego środowiska pracy. Komputery wchodzące w skład klastra (będące członkami klastra) nazywamy węzłami (ang. node).

**System webowy** jest to strona bądź serwis internetowy, który wraz z użyciem sieci udostępnianie dynamicznie treść, pozwalają także na interakcje z użytkownikiem.

**Rozproszony System webowe** to zbiór niezależnych urządzeń technicznych połączonych w jedną, spójną logicznie całość, która tworzy system webowy.

W skład klastrów webowych i rozproszonych systemów webowych wchodzą najczęściej: przełącznik webowy, który odbiera od klientów żądania HTTP, serwery WWW i inne serwery zaplecza takie jak serwery aplikacji i bazodanowe. Zadaniem przełącznika webowego, jest odebranie żądania HTTP od klienta, wybranie z pomocą algorytmu dystrybucji żądań serwera do obsługi żądania i przekierowanie do niego żądania. Po otrzymaniu z serwera odpowiedzi przełącznik przesyła odpowiedź do klienta. Od rodzaju zastosowanego algorytmu dystrybucji żądań i mechanizmu przekierowania w dużej mierze zależy wydajność całego serwisu webowego.

Metody dystrybucji żądań http można podzielić na

* Statyczne
  + Random
  + Round-robin
  + Static weighted RR
* Dynamiczne
  + weighted RR
  + Client affinity – na początku odsyła do najmniej obciążonego serwera, następnie utrzymuje połączenie
  + LFNRD - wybranie spośród S serwerów WWW pracujących w klastrze tego serwera, który oferuje najkrótszy czas odpowiedzi dla nadchodzącego żądania http. Decyzja podejmowana jest w oparciu o wiedzę o aktualnych obciążeniach serwerów oraz o przeszłych czasach odpowiedzi na żądania.

**35. Metody i algorytmy buforowania treści**

**Buforowanie obiektów WWW** jest jedną z najsilniej rozwijających się technik poprawy efektywności aplikacji internetowych. Zadaniem buforowania jest przechowywanie kopii obiektów poza miejscem ich publikowania w celu minimalizacji drogi pokonywanej przez żądanych zasób.

Buforowanie możemy podzielić na metody

* Pasywne – w buforze zapisywany jest dokument którego jawnie zażądał klient
* Aktywne – w buforze zapisywany jest dokument którego jawnie zażądał klient oraz dokumenty powiązane z żądanym zasobem, dzięki czemu bufor przygotowuje się już do obsługi przewidywanego następnego kroku klienta.

Bufor może być zlokalizowany po stronie Klienta – dane są pobierane i zapisywane przez przeglądarkę, serwera WWW lub Serwera PROXY. Ponadto bufor może być scentralizowany lub rozporoszony między wiele węzłów. Popularnymi metodami buforowania są ICP oraz CARP.

**Protokół ICP** pozwala serwerowi otrzymującemu żądanie od użytkownika na przeszukaniu zawartości buforów okolicznych serwerów PROXY. Jeśli żądany zasób będzie znaleziony wtedy będzie on dostarczony klientowi. Protokół ICP cechuje się skalowalnością oraz tendencją do wprowadzania redundancji.

**Protokół CARP** – umożliwia zorganizowanie wielu serwerów PROXY w jeden duży bufor WWW. Gdy użytkownik wysyła żądanie dostarczeni zasobu, wtedy przy pomocy funkcji haszującej dokonuje się wyboru serwera PROXY, który będzie odpowiedzialny za realizację żądania. Protokół CARP cechuje się skalowalnością oraz brakiem wprowadzania redundancji.

**36. Organizacja sieci dostarczania usług CDN**

**Content delivery network** (ang. CDN) – jest to duży rozproszony system dostarczania treści do wielu centrów danych i punktów wymiany ruchu w Internecie. Celem CDN jest udostępnianie zawartości o wysokiej dostępności i wydajności końcowym użytkownikom.

CDN zakłada, że

* Podstawową barierą, która blokuje polepszanie wydajności jest niedoskonałość sieci Internet.
* Ponieważ sieć stanowi najsłabsze ogniwo, powinno się zminimalizować jej wpływ na tę usługę.
* Polepszenie wydajności usług można rozwiązać poprzez inteligentne buforowanie zasobów.

W praktyce oznacza to, że CDN buforuje treść obsługiwanych serwerów na globalnie rozproszonych serwerach WWW, minimalizując tym samym drogę potrzebną do pokonania między serwerem a klientem końcowym. Internet w takim założeniu służy wówczas do napełniania i aktualizacji buforów.

Newralgicznym punktem CDN są mechanizmy i algorytmy wyboru najbliższego klientowi serwera treści, bardzo popularnym rozwiązaniem jest pomiar RTT, gdzie klientowi przydzielany jest serwer z najmniejszym wynikiem RTT.

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

**Pomiary Internetu** wykonuje się ponieważ przydatność metod pomiarowych wynika z następujących powodów

* Obciążenie ruchem IP jest trudne do przewidzenia
* Obecne sieci IP są strukturami coraz bardziej rozbudowanymi
* Modele analityczne uniemożliwiają uzyskanie wiarygodnych wyników
* Zwiększenie jakości obsługi do sieci pakietowych wymaga wiedzy o aktualnym ruchu w sieci

Metody pomiaru internetowego są dwie, aktywna i pasywna.

**Metoda aktywna** polega na zmierzeniu metryk opóźnienia, poziomu strat pakietów i przepływności poprzez wysłanie specjalnych pakietów pomiarowych. Metoda aktywna zakłada, że pakiety pomiarowe są przesyłane tą samą i obsługiwane w ten sam sposób co pakiety użytkowników. Niewątpliwą zaletą metody aktywnej jest łatwość obróbki i interpretacji danych, jednak znaczącą wadą metody aktywnej jest wpływanie samych narzędzi pomiarowych na wyniki pomiarów.

Narzędzia do aktywnego pomiaru to

* Ping
* Traceroute
* Scriptroute

**Metoda pasywna** polega na obserwacji pakietów w danym punkcie pomiarowym i odpowiedniej analizie zarejestrowanego zapisu. Pasywna metoda pomiaru wymaga rejestrowania zaobserwowanych pakietów i odpowiadającym im znacznikom czasowym w dwóch punktach pomiarowych. Następnie, dane zebrane w poszczególnych punktach przesyłane są do serwera zarządzającego pomiarami, gdzie są analizowane. Metoda pasywna w przeciwieństwie do aktywnej nie generuje dodatkowego ruchu, ale sporą wagą tego rozwiązania jest skomplikowane przetwarzanie uzyskanych danych.

Narzędzia do pasywnego pomiaru to

* TCP dump, w tym wireshark
* Route table dumps
* Route views

**System MWING** jest wieloagentową platformą do zarządzania i wykonywania pomiarów aktywnych. Polega on na badaniu danych związanych z przesyłaniem identycznego zasobu pobieranego w równych odstępach czasu z różnych miejsc w sieci. System MWING zakłada, że transmisja nie może zbytnio obciążać sieci i serwerów. Przykładem wykorzystania systemu MWING było badanie pobierania rfc1945 co kilka godzin z kilku agentów rozmieszczonych na świecie.

**38. Badania i pomiary Internetu w zakresie topologii, ruchu, stanu i aplikacji**

Pomiary Internetu można wyrazić poprzez

* Opóźnienie (latency) [ms]
* Utrata pakietów (packet loss) [%]
* Szybkość transmisji (transfer rate) [b/s]
* Wykorzystanie łącza (link utilization) [%]

**Pomiary Internetu** mogą być aktywne i pasywne.

**Metoda aktywna** polega na zmierzeniu metryk opóźnienia, poziomu strat pakietów i przepływności poprzez wysłanie specjalnych pakietów pomiarowych. Metoda aktywna zakłada, że pakiety pomiarowe są przesyłane tą samą i obsługiwane w ten sam sposób co pakiety użytkowników. Niewątpliwą zaletą metody aktywnej jest łatwość obróbki i interpretacji danych, jednak znaczącą wadą metody aktywnej jest wpływanie samych narzędzi pomiarowych na wyniki pomiarów.

Narzędzia do aktywnego pomiaru to

* Ping
* Traceroute
* Scriptroute

**Metoda pasywna** polega na obserwacji pakietów w danym punkcie pomiarowym i odpowiedniej analizie zarejestrowanego zapisu. Pasywna metoda pomiaru wymaga rejestrowania zaobserwowanych pakietów i odpowiadającym im znacznikom czasowym w dwóch punktach pomiarowych. Następnie, dane zebrane w poszczególnych punktach przesyłane są do serwera zarządzającego pomiarami, gdzie są analizowane. Metoda pasywna w przeciwieństwie do aktywnej nie generuje dodatkowego ruchu, ale sporą wagą tego rozwiązania jest skomplikowane przetwarzanie uzyskanych danych.

Narzędzia do pasywnego pomiaru to

* TCP dump, w tym wireshark
* Route table dumps
* Route views

**39. Wykorzystanie eksploracji danych w analizie logów serwerów WWW.**

**Log serwerów WWW** jest to chronologiczny zapis zawierający informację o zdarzeniach i działaniach dotyczących systemu WWW. Log tworzony jest automatycznie przez dany program komputerowy, a sama czynność zapisywania do logu nazywana jest też logowaniem. Logi zawierają informację co dzieje się na serwerze w każdej chwili.

**Eksploracja danych** jest to jeden z etapów odkrywania wiedzy z baz danych. Idea eksploracji polega danych polega na wykorzystaniu szybkości komputera do znajdowania prawidłowości w zgromadzonych danych.

Analiza zachowania internautów jest bardzo atrakcyjną czynnością z punktu widzenia badań marketingowych. Eksploracja użytecznych reguł odnoszących się do rzeczywistych działań użytkowników podczas wizyty na stronie internetowej pozwala zwiększyć użyteczność i funkcjonalność witryn, a także zwiększyć dochody proponując klientowi usługi które mogą go zainteresować.

**40. Wykorzystanie eksploracji danych w analizie wydajności systemów webowych**

**Systemy webowe** jest to strona bądź serwis internetowy, który wraz z użyciem sieci udostępnianie dynamicznie treść, pozwalają także na interakcje z użytkownikiem.

**Eksploracja danych** jest to jeden z etapów odkrywania wiedzy z baz danych. Idea eksploracji polega danych polega na wykorzystaniu szybkości komputera do znajdowania prawidłowości w zgromadzonych danych.

Analiza zachowania internautów jest bardzo atrakcyjną czynnością z punktu widzenia badań marketingowych. Eksploracja użytecznych reguł odnoszących się do rzeczywistych działań użytkowników podczas wizyty na stronie internetowej pozwala zwiększyć użyteczność i funkcjonalność witryn.

**41. Semantic Web, ontologie**

**Semantic web** jest to projekt mający przyczynić się do utworzenia i rozpowszechnienia standardów opisywania treści w Internecie, który umożliwi maszynom i programom przetwarzanie informacji w sposób odpowiedni do ich znaczenia.

W swoich założeniach Semantic web ma korzystać z istniejącego protokołu komunikacyjnego, na którym bazuje obecnie funkcjonujący Internet. Różnica miała by polegać na tym, że przesyłane dany mogły by być rozumiane także poprzez maszyny. Uzyskane było by to poprzez odpowiednie reguły oraz opis obiektów poprzez meta dane. Dzięki temu, można było by

* Powiązać różne dane znajdujące się w Internecie w ramach wspólnych jednostek znaczeniowych (filmy x, kuchnia y itp.)
* Rozróżnić dane, które dla maszyn są w tej chwili nierozróżnialne ze względu na identyczny zapis tekstowy, bądź będące identycznie brzmiące. Np. zamek (budynek i ten od spodni)
* Prowadzić na tych danych wnioskowania. Np. Ewa jest żoną Adama – Ewa jest kobietą, Adam jest mężem Ewy itd.

**Ontologia** w sensie informatycznym jest to formalna reprezentacja pewnej dziedziny wiedzy, na którą składa się zapis zbiorów pojęć i relacji między nimi. Zapis ten tworzy schemat pojęciowa, który będąc opisem danej dziedziny wiedzy, może służyć jednocześnie jako podstawa do wnioskowania o własności opisywanych ontologią pojęć.

Ze względu na stopień formalizacji możemy wyróżnić następujące rodzaje ontologii

* Nieformalne
  + Predefiniowane słownictwo
  + Słowniki
  + Tezaurusy
  + Taksonomie
* Formalne
  + Ontologie oparte n danych
  + Ontologie oparte na logice

**42. Technologie informacyjne i komunikacyjne (ICT) w nowoczesnym społeczeństwie**

**43. Systemu autonomiczne, proaktywne systemy i sieci wszechogarniające oraz inne nowatorskie** **zastosowania ICT.**

**44. Perspektywy rozwoju ICT oraz internetu**