
Webgraph - system do analizy i syntezy struktur serwisów www

Dmitrij Żatuchin

Politechnika Wrocławska
Wyb. Wyspiańskiego 27.
50-370 Wrocław
Dmitrij.Zatuchin@pwr.wroc.pl

Abstrakt

Optymalizacja struktury serwisów internetowych jest aktualnym naukowym i praktycznym problemem. Architektura informacji wykorzystuje strukturę jako podstawę w procesie konstruowania projektu stron internetowych. Stąd bardzo ważne jest dostarczenie takiej struktury, która będzie użytkowania przez użytkowników intuicyjnie. Wykorzystanie technik analizy odwiedzania serwisów internetowych uważane jest za przyszłość dziedziny badań użyteczności i personalizacji. Istnieją także odmienne podejścia do rozwiązania problemu optymalizacji architektury informacji (ang. IOP – Information Optimization Problem). Celem niniejszej pracy jest przedstawienie problemu oceny struktury strony internetowej oraz jej optymalizacja. Zostanie przedstawione środowisko Webgraph, które wykorzystuje zaproponowane procedury analizy struktury stron oraz metodę syntezy. Omówione zostaną wybrane eksperymenty przeprowadzone z wykorzystaniem środowiska Webgraph. W praktyce opracowane narzędzie wspomaga proces analizy złożonych struktur serwisów internetowych (np. sklepy internetowe, portale informacyjne), tym samym skracając czas i liczbę zasobów potrzebną na przeprojektowanie. Webgraph jest narzędziem dla architektów informacji i badaczy użyteczności, które pomagają im oceniać jakość struktury serwisów internetowych oraz automatycznie

generować propozycję alternatywnej struktury w oparciu o dane użytkownika serwisów w przeszłości.

Słowa kluczowe

Struktura serwisu internetowego, graf użytkownika, model interfejsu, dane użytkownika, analiza jakości.

Wprowadzenie

Jednym z wyzwań projektowania interfejsu na potrzeby Internetu jest fakt, że użytkownicy mogą przyjść do serwisu internetowego na wiele sposobów – nie tylko poprzez stronę główną, przy czym cele i zadania tych użytkowników różnią się od siebie. Nawigacja powinna wspierać te różnice, poprzez różne typy funkcjonalności nawigacji: hierarchiczną organizację, nawigację opartą o zadania, alfabetyczną, chronologiczną czy bazującą na popularności organizację informacji [1].

Wraz z gwałtownym wzrostem zawartości witryn, zarówno użytkownicy serwisu jak i dysponenty oczekują wysokiej jakości usług dostarczanych przez serwisy internetowe. Celem użytkowników jest otrzymanie informacji wtedy, kiedy potrzebują i takiej, jak oczekują. Celem dysponentów – przyciągnięcie większej liczby użytkowników do swoich serwisów, zatrzymanie ich w serwisie internetowym oraz skorzystanie z oferowanych tam usług. Korzystanie ze strony przez użytkowników jest czynnością polegającą na wybieraniu odpowiednich odnośników ze struktury strony. Schemat nawigacji oraz funkcjonalności powinny pozwolić użytkownikom odnaleźć i uzyskać dostęp do informacji efektywnie i skutecznie, poprzez zestaw rozwiązań nawigacyjnych takich jak: różne rodzaje menu, okruszki nawigacyjne (ang. breadcrumbs), krótkie strony z wieloma odnośnikami, odpowiedni projekt i położenie zakładek, liczba elementów w menu i kategoriach, a także różnicowanie

i grupowanie elementów nawigacyjnych umieszczając najważniejsze w górnej części oraz lewym panelu (lub prawym dla serwisów internetowych w języku arabskim czy jidysz), aby użytkownicy nie zostali uwięzieni w osieroconych lub ślepych podstronach serwisu. [2]. Na stronie statycznej lub z ograniczoną liczbą elementów dynamicznych struktura odnośników ma duży wpływ na jakość dostarczanej treści.

Nawigacja jest bardzo ważną składową użyteczności serwisów internetowych, co wielokrotnie było podkreślane przez takich badaczy i ekspertów jak [3], [4], [5], [6], [7], [8], [9], [10] czy [11].

Jeszcze w 2001 roku Donahue [12] wskazał, że trudna do nawigacji strona internetowa z ograniczoną elastycznością stanowi główny problem użyteczności. Zgodnie z National Institute of Standards and Technology [13] łatwość nawigacji jest niezbędna dla użytkowników na wszystkich poziomach zaawansowania w używaniu komputerów, do sprawnego poruszania się i uzyskiwania poszukiwanych informacji. Dlatego ewaluacja i poprawa struktury serwisów internetowych staje się zadaniem kluczowym.

Optymalizacja struktury poprzez ulepszenie nawigacji

Optymalizacja struktury strony jest problemem aktualnym i ostatnimi laty przykuwa uwagę naukowców. Optymalizacja była i wciąż jest wykonywana ręcznie przez projektantów i architektów informacji, którzy stosują różne wytyczne zgodnie z własną intuicją lub wyniki prowadzonych badań za pomocą metody sortowania kart czy dyskusji z grupą fokusową celem uzyskania nowej bądź modyfikacji struktury nawigacji serwisu internetowego. Projektowanie nawigacji dotyczy także problemu liczby elementów, która ma być przedstawiana użytkownikowi naraz (co jest postrzegane

także jako problem organizacji treści) oraz zagadnienia rozmieszczenia w menu pionowym lub dynamicznym. Wczesne badania wskazywały, że liczba elementów głównego menu, czyli pierwszego rzędu węzłów w grafie, powinna być ograniczona Regułą Siedmiu plus-minus dwa elementy (zgodnie z pracą Millera w dziedzinie psychologii, badania nad możliwościami umysłowymi człowieka) [14], aczkolwiek najnowsze wyniki badań wskazują, że szerokie i płytkie układy menu nawigacyjnego są aktualnym upodobaniem użytkowników. Istnieją zatem sprzeczne porady dotyczące liczby „7+-2” elementów, chociaż dla systemów złożonych, gdzie strukturę nawigacji stanowią dziesiątki lub setki elementów, jest to jeden z niewielu zorganizowanych sposobów na redukcję złożoności struktury.

Optymalizacja nawigacji – pierwsze próby

Pierwsze próby w automatyzacji optymalizacji nawigacji podjął Perkowitz [15]. Celem badań była zmiana położenia odnośników w ramach pojedynczej strony. Następnie Perkowitz zaproponował ideę Adaptacyjnej Strony, która doskonaliłaby się poprzez uczenie się ścieżek dostępu użytkowników do strony. W 1999 Garofalakis [16] zaproponował optymalizację serwisu internetowego w oparciu o popularność każdej strony. Analizowano logi serwera, zaś administrator ustalał wartości parametrów oraz kryteria, które decydowały w jaki sposób miała być generowana nowa struktura serwisu internetowego. Popularność w definicji Garofalakisa brała pod uwagę głębokość strony w strukturze, liczbę stron na tym samym poziomie oraz liczbę odnośników wewnętrznych (czyli należących do mapy stron serwisu internetowego) kierujących na badaną stronę serwisu internetowego. To był krok do przodu i aktualizacja idei Perkowitza, dlatego, że zagadnienie optymalizacji potraktowano w kontekście analizy kilku stron serwisu internetowego, a nie tylko jednej. Strona została

zamodelowana jako drzewo celem uproszczenia działania algorytmu optymalizującego strukturę.

Analiza danych użytkownika a optymalizacja struktury

Dziedzina analizy danych użytkownika serwisów internetowych i wykorzystania wyników analizy do poprawiania i dostosowywania interfejsów do wymagań użytkowników jest określana jako przyszłość badań użyteczności. [17]. Zachowania użytkowników są analizowane poprzez odkrywanie wzorców, które są rozpoznawane z otwartych sesji, logów serwera, ciasteczek (ang. cookies) lub statystyk [18], [19]. Wykorzystanie odkrytych wzorców nawigacji użytkowników do przebudowy struktury serwisu internetowego w podejściu całościowym można spotkać w pracy [20], gdzie zdefiniowano funkcję kosztu składającą się z czasu dotarcia do celu oraz łącznego czasu ładowania stron od pierwszej do ostatniej ($E = Ed + Ec$), przy czym podstrony były zliczane na podstawie danych zapisywanych w sesji użytkownika. Niestety takie podejście jest obciążone wieloma wadami. Pierwsza, to reguła zmiany położenia, która twierdzi, że liście można podpiąć pod dowolną stronę, druga – liście można przesuwac w dół lub w górę. Po każdej reorganizacji przeliczana jest na nowo liczba sesji dla każdej strony indeksowej, i wiele stron staje się liśćmi w strukturze serwisu. Trzecia wada, to czas trwania sesji oraz sama koncepcja wykorzystania danych z sesji do odkrywania zachowania użytkowników.

Następne ważne badania zostały poczynione przez Yen i innych [21], gdzie zdefiniowano środowisko składające się z trzech warstw do ewaluacji i rozszerzania projektów serwisów internetowych z wykorzystaniem grafu jako sposobu modelowania strony. Yang w 2008 roku zdefiniował model konceptualny dla struktury serwisu [22] stosując tym samym koncepcję porządkowania znaną z ontologii danych. Serwisy

internetowe modelowano także w probabilistycznym modelu łańcuchów Markova, gdzie następnie wykorzystywano algorytm skracania ścieżek z jednej strony do drugiej w ramach serwisu internetowego [23]. W tym podejściu proponowano algorytm Odnajdywania Zwrotnych Stron (Finding Backtrack Pages), którego skutkiem jest co najwyżej struktura serwisu uzupełniona o dodatkowe odnośniki w miejscach, w których użytkownicy mogliby mieć potrzebę wrócić poziom wyżej. Także można spotkać prace inspirowane algorytmem PageRank [24], znowuż z wykorzystaniem technik analizy odwiedzania stron internetowych (ang. webmining) celem optymalizacji struktury serwisu [25]. Stosuje się również reguły asocjacyjne wraz z algorytmem grafowym optymalizującym istniejącą topologię strony [26]. Niestety podejście takie jest mało praktyczne, gdyż nie nakłada reguł ograniczających na możliwe struktury serwisów internetowych, co skutkuje możliwością uzyskania nielogicznego wyniku, który nie będzie można wdrożyć. Powodem tego jest możliwość wymieszania konceptów z różnych kategorii, należących formalnie do różnych węzłów-rodziców (np. podstrony z działu *Motoryzacja* z działem *Polityka* itp.).

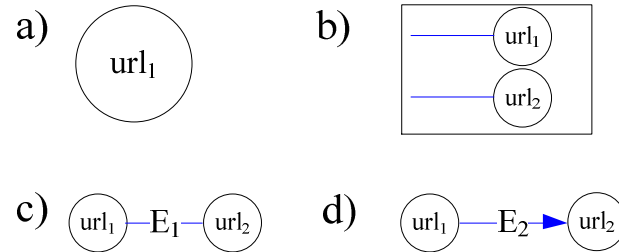
Szczegółowiej na temat optymalizacji struktury serwisu internetowego, ścieżek nawigacji użytkowników a także adaptacji interfejsu można znaleźć w najnowszych pracach: Lin [27], Chou [28], Wang [29], oraz Saremi [30], na którą warto zwrócić uwagę. Praca Saremi traktuje o rozwiązaniu problemu umieszczenia podstron w strukturze serwisu internetowego. Problem ten sformułowano jako kwadratowy problem przypisania stron do istniejących pozycji w serwisie i zaproponowano jego rozwiązanie metodą heurystyczną bazującą na algorytmie optymalizacji kolonii mrówek. Jednak takie podejście ma kilka wad:

1. Nie uwzględnia faktu, że droga, którą podążają użytkownicy, jest najczęściej nie optymalna.
2. Nie eliminuje węzłów tranzytowych, czyli takich, które służą jedynie jako przejście z jednego węzła do docelowego. Im więcej ich jest na drodze użytkownika, tym więcej czasu zajmuje użytkownikowi osiągnięcie celu.
3. Nie uwzględnia głębokości, na której znajduje się podstrona, tym samym najmniej ważna może znaleźć się na samej górze.
4. Jest nieelastycznym rozwiązaniem, które pozwala dokonać analizy struktury z godziny-na-godzinę, tym samym dynamicznie dostosować strukturę serwisu do potrzeb użytkowników czy zmieniających się grup użytkowników korzystających z serwisu internetowego.

Model struktury serwisu internetowego

Do zamodelowania struktury strony wykorzystano teorię grafów. Jeszcze w modelu HAM [31], który jest prekursorem modelowania interfejsów multimedialnych, wykorzystywano strukturę węzła-odnośnika korzystając z teorii grafów.

Struktura serwisu internetowego jest zestawem stron o unikalnej treści, do których można dotrzeć poprzez elementy nawigacji. Struktura reprezentuje całą stronę internetową, gdyż wykorzystuje podstawowe elementy: stronę i połączenie. Nawigacja składa się ze zbioru połączeń i stron, do których kieruje łączenie (**Rysunek 1**). Ponieważ podstawowym zadaniem strony internetowej jest dostarczenie treści i funkcjonalności, które są umieszczone w podstronach, głównym zadaniem nawigacji jest efektywna i skuteczna obsługa żądań użytkowników.



Rysunek 1. Elementy struktury strony internetowej: a) podstrona-węzeł, b) element nawigacji, c) łączenie między dwoma podstronami, d) łączenie skierowane

Model interfejsu w pracach Diaza i Aedo [32] [33] jest zorientowany na działania użytkownika. Mając definicję struktury serwisu internetowego można zdefiniować model interfejsu zorientowany na strony i połączenia między nimi, wtedy model interfejsu przyjmie postać:

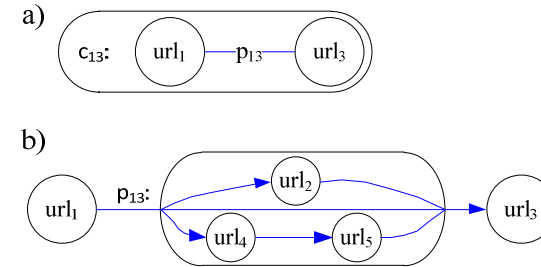
$$I_l = (GI, Us, C, A, Ev, Aset) \quad (1)$$

Gdzie GI – to struktura serwisu internetowego l definiowana jako graf o N podstronach i M łączeniach:

$$GI_l = (URL_l, E_l, P_l, Co_l) \quad (2)$$

Podstrona (Rysunek 1a) jest pojedynczą stroną zawierającą zawartość, elementy nawigacji oraz połączenia do innych stron. Jedną z wszystkich stron jest stroną główną serwisu internetowego, która jest domyślnie wyświetlana po wpisaniu bezpośredniego adresu serwera – nazwy domeny lub IP – do przeglądarki. Element nawigacji (Rysunek 1b) może wystąpić na jednej lub wielu podstronach, przy czym za każdym razem może zawierać inny zestaw podstron i łączeń do nich kierujących. Łączenia mogą być nieskierowane lub

skierowane (Rysunek 1cd). Podstawowe elementy nawigacji modelują ścieżki P oraz połączenia C (**Rysunek 2**).



Rysunek 2. Modelowanie: a) ścieżki, b) połączenia.

Ścieżka jest zbiorem możliwych przejść z jednego węzła do drugiego. Zbiór ścieżek P_l jest zdefiniowany jako:

$$P_l = \{p_{ij}, i, j \in [1, N] \mid \{e_{li} < e_{lj} \mid \exists p_{liz} \wedge \exists p_{lzz}, z \in [1, N]\}\} \quad (3)$$

Połączenie pomiędzy dwoma węzłami istnieje wtedy, jeżeli z jednego węzła do drugiego istnieje możliwość przejścia. Zbiór C_l dla grafu (1) to:

$$Co_l = \{co_{ij}, i, j \in [1, N] \mid \forall co_{ij} \exists p_{lij} \rightarrow url_{li} < url_{lj}\} \quad (4)$$

Ścieżki mogą być bezpośrednie – wtedy między węzłami jest jedna krawędź, lub tranzytowe – gdzie występują na drodze między węzłami inne węzły pośredniczące. Długością połączenia c_{ij} między węzłami url_{li} a url_{lj} będzie średnia długość wszystkich ścieżek między tymi węzłami:

$$Dc(co_{ij}) = average(p_{lij}) \quad (5)$$

Zadanie szacowania jakości struktury strony

Dla węzłów i krawędzi opisano charakterystyki, które wykorzystują informacje o strukturze grafu oraz wskaźniki, które wykorzystują dane użytkownika i są wykorzystywane w zadaniu analizy jakości struktury strony [34] [35]. Tak zdefiniowany model struktury strony pozwala sformułować zadanie szacowania jakości struktury strony WWW, to jest zadanie wyznaczenie wartości energii grafu będącego modelem użytkowanego interfejsu strony WWW.

Dane:

1. Graf struktury strony GI_i ;
2. Dane o użytkowaniu struktury strony w okresie czasu τ_k ;

Wyznaczyć:

1. Jakość odkrytej struktury strony WWW uwzględniając dane użytkownika.

Zauważmy, że dla różnych okresów czasu τ wartość wskaźnika będzie inna, gdyż jest ona zależna od danych użytkownika grafu w określonym okresie. Przy założeniu, że położenie każdego węzła w grafie oraz sposób ich łączenia wyznacza sposób, w jaki użytkownicy korzystają ze strony internetowej, wskaźnik jakości struktury również powinien być uzależniony od charakterystyk węzłów i krawędzi.

Zdefiniujmy zatem kryteria jakości:

1. Liczba użyć węzła;
2. Liczba użyć krawędzi;
3. Położenie węzła względem struktury;
4. Znaczenie węzła w strukturze strony;
5. Znaczenie krawędzi w strukturze strony.

Dane użytkownika bezpośrednio określają czy zaprojektowana pierwotnie struktura jest odpowiednia do celów, które użytkownicy próbują spełniać korzystając z niej.

Na ich podstawie proponuje się miarę jakości struktury serwisu internetowego – Energię Grafu (En), która wykorzystuje zdefiniowane własności modelu struktury:

$$E = En_k(GI_i) = \sum_{n=1}^N ImpU_k(url_{in}) * \sum_{m=1}^M CD_k(e_{lm}) \quad (6)$$

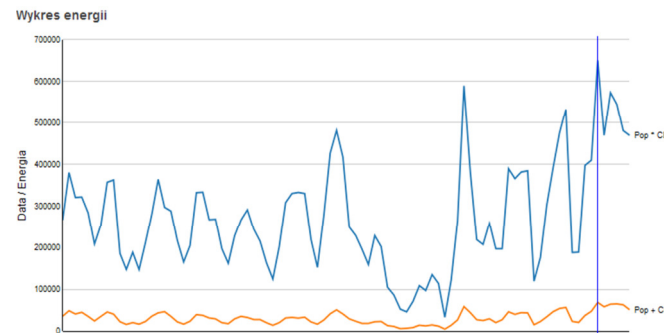
Energia grafu jest wskaźnikiem jakości struktury serwisu, który jako estymator obiektywny wzmacnia metody subiektywne badania użyteczności serwisów internetowych. Obserwacja zmian jakości struktury pozwala wykryć momenty, w których są odstępstwa od określonej wartości En , np. od średniej o wartość odchylenia standardowego σ lub inne wskaźniki wykrywające zmiany w czasie na podstawie serii danych pomiarowych (np. zwiększenie średniej ze stałą wariancją, zwiększenie wariancji przy stałej średniej itp.).

$$E = En_k(GI_i) = \sum_{n=1}^N ImpU_k(url_{in}) + \sum_{m=1}^M CD_k(e_{lm}) \quad (7)$$

W proponowanej postaci (6) wskaźnik jest wrażliwy na zmiany w sposobie użytkowania węzłów. Porównując postać wskaźnika z inną, możliwą postacią (7), na poniższym wykresie widać, że postać (7) nie pozwala na zarejestrowanie dynamiki zmian zachodzących w serwisie internetowym (**Rysunek 3**). Z drugiej strony, zastosowanie w monitoringu struktury serwisu wskaźnika w postaci (7) pozwala wykryć bardzo znaczące zmiany i dopiero wtedy poczynienie kroków w kierunku zmiany struktury serwisu internetowego. Obserwując obydwie krzywe specjalista ma możliwość według uznania poczynić zmianę struktury bazując na krótkoterminowych wahaniami bądź też dopiero, gdy

w obydwu przypadkach będzie potwierdzony znaczący spadek energii grafu.

Poprzez zastosowanie metod wykrywających zmiany w ciągach danych, z szczególnym uwzględnieniem nagłych zmian, możliwe jest zdecydowanie, czy zmiana zachodząca w charakterystyce obserwowanego obiektu, jakim jest struktura serwisu internetowego, zaszła z powodu zdarzenia, na które należy zareagować i zmienić strukturę, czy też z powodu jedynie większego natężenia ruchu w Internecie. Jest to jeden z dalszych kierunków badań.



Rysunek 3. Porównanie energii grafu dla struktury serwisu internetowego Edukacja.CL - wersja wskaźnika liczona jako iloczyn sum Pop i CD (niebieski kolor) oraz jako suma sum Pop i CD (pomarańczowy kolor).

Własności miary energii grafu

Z definicji (6) En jest proporcjonalną do liczby użyć węzłów i krawędzi, liczby osiągalnych węzłów poprzez krawędzie oraz liczbę wchodzących i wychodzących krawędzi do poszczególnych węzłów, zaś odwrotnie proporcjonalna do liczby węzłów N , funkcji odległości dla węzłów, liczby ścieżek

w grafie oraz liczbie krawędzi wychodzących z węzłów. Kolejne własności to:

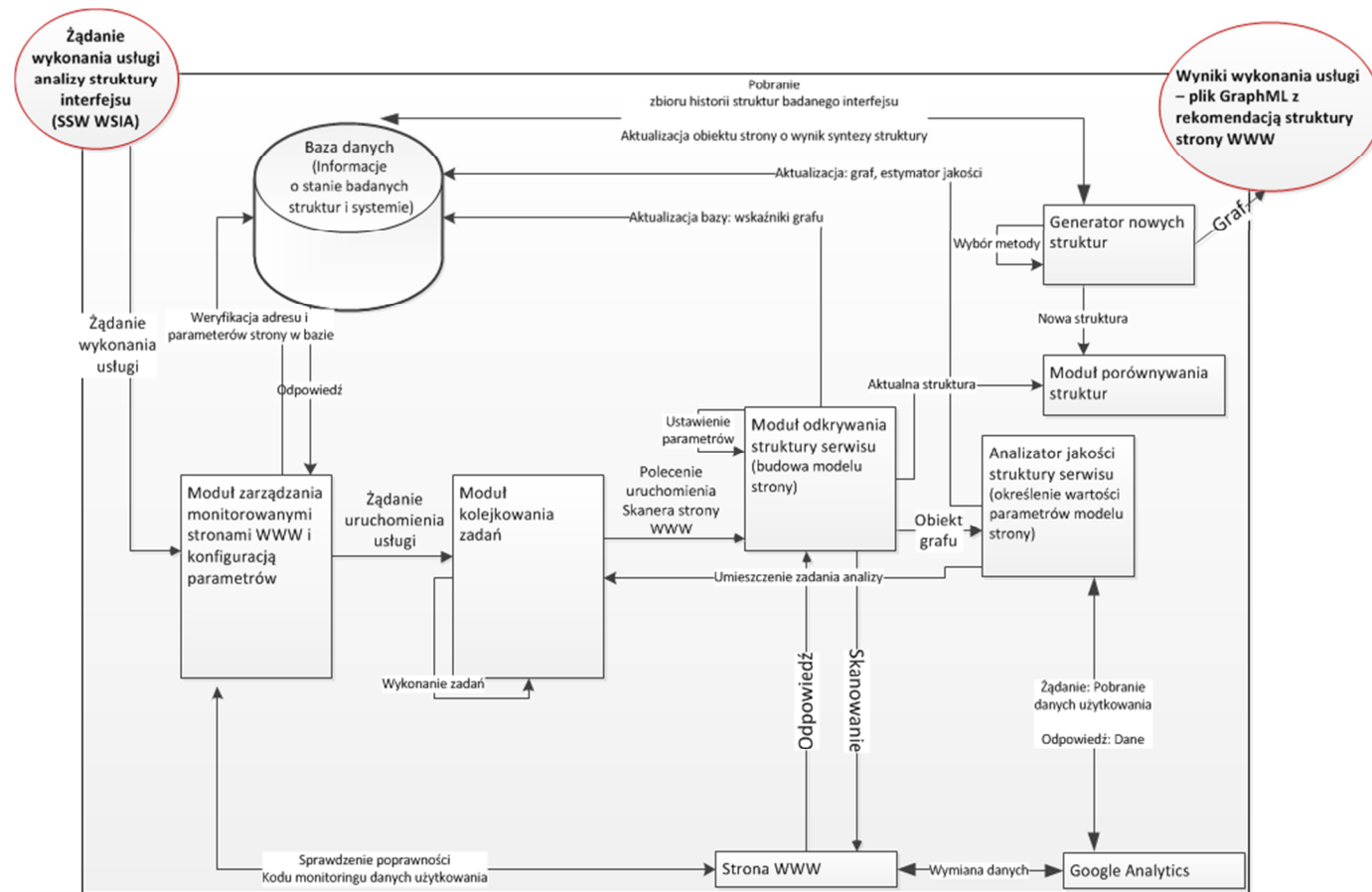
- Wraz ze wzrostem liczby węzłów i krawędzi, przy zachowaniu ruchu (liczby użyć krawędzi i węzłów) – energia maleje;
- Dla takiej samej liczby węzłów i krawędzi, oraz takiej samej sumy użyć krawędzi i węzłów, ale różnemu rozkładowi tych użyć krawędzi, energia jest taka sama;
- Wraz z równomiernym wzrostem liczby użyć węzłów rośnie wartość energii;
- Wraz z równomiernym wzrostem liczby użyć krawędzi rośnie wartość energii;
- Wraz ze wzrostem liczby węzłów, i zachowaniem tej samej liczby krawędzi, energia maleje; [osamotnione węzły]
- Wraz ze wzrostem liczby krawędzi, i zachowaniem tej samej liczby węzłów, oraz wzrostem liczby ścieżek, którym podążają użytkownicy, energia rośnie;
- Wraz z usunięciem krawędzi, która spowoduje osamotnienie węzła, energia maleje;
- Wraz ze wzrostem liczby krawędzi wychodzących z węzła do innych węzłów, rośnie $ImpU$ dla tego węzła;
- Wraz ze wzrostem liczby krawędzi wchodzących do danego węzła, rośnie $ImpU$ tego węzła;
- Wraz ze wzrostem liczby węzłów, do których można dotrzeć poprzez jedną krawędź, rośnie CD dla tej krawędzi;
- Usunięcie węzła o małej liczbie krawędzi wychodzących i osiągalnych za pomocą ich węzłów powoduje wzrost energii;
- Wpływ węzłów, które znajdują się niższych poziomach w grafie, jest mniejszy na strukturę niż tych z wyższych poziomów. Przy użyciu węzła na niższym poziomie większym d razy niż na poziomie wyżej, $ImpU$ dla tych węzłów jest zbliżone.

Środowisko Webgraph

Na potrzeby badań opracowano narzędzie z zaimplementowanymi metodami analizy i syntezy struktur serwisów internetowych. Narzędzie jest dostępne w środowisku internetowym. Komputer-serwer uruchamia cyklicznie skrypty do aktualizowania wyników przetwarzania danych o użytkowaniu i zapisywania ich do modelu danych w postaci zaktualizowanego grafu badanego serwisu. Moduł analizatora przetwarza grafy stosując miarę Energii grafu jako wskaźnik jakości. Narzędzie umożliwia zarządzanie badanymi stronami, dokładną analizę wyników oceny jakości struktury serwisów internetowych poprzez warstwę prezentującą wskaźniki i cechy grafu, a także wykresy. Moduł syntezy zawiera dwie metody generowania struktur stron. Całkowity schemat jest przedstawiony na Rysunku 4.

Badania eksperymentalne

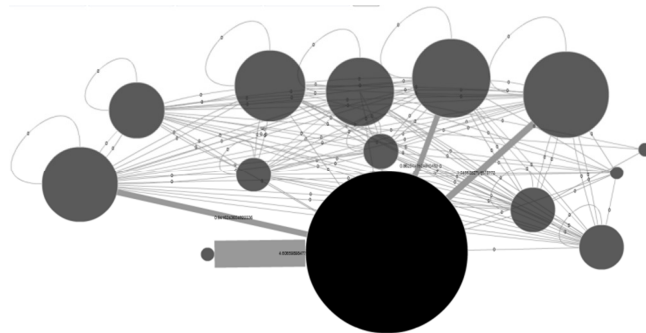
Przeprowadzono serię eksperymentów z wykorzystaniem opracowanego narzędzia celem weryfikacji opracowanych metod analizy i syntezy struktury serwisów internetowych. Badania przeprowadzono na wybranych istniejących w rzeczywistości serwisach internetowych: strona firmowa www.dook.pl, strona konferencji www.icss.pwr.wroc.pl, strona klienta webowego JSOS Edukacja www.edukacja.pwr.wroc.pl, strona firmowa salonu fryzjerskiego www.saloncesare.pl oraz strona portalu do nauki języków obcych www.edustation.pl. We wszystkich przypadkach narzędzie oraz metody zwróciły zadowalający rezultat w postaci struktury o większym wskaźniku En dla symulowanych danych użytkowania na podstawie danych historycznych. Dalej zostaną przedstawione wybrane eksperymenty celem ilustracji działania narzędzia oraz opracowanych metod.



Rysunek 4. Webgraph – schemat postępowania - od żądania wykonania analizy i syntezy do wyniku.

Strona konferencyjna

Zbadano stronę informacyjną o konferencji ICSS pod adresem <http://icss.pwr.wroc.pl>. Strona była monitorowana w okresie 6.03.2010-8.03.2011. Pozyskana struktura zawiera 14 węzłów i 125 krawędzi. To w praktyce świadczy o tym, że prawie z każdej podstrony można przejść na inną w ramach tego serwisu. Energia grafu dla tego okresu wyniosła 1587,15. Na podstawie scalenia pozyskanych struktur uzyskano ostateczną postać struktury (Rysunek 5).



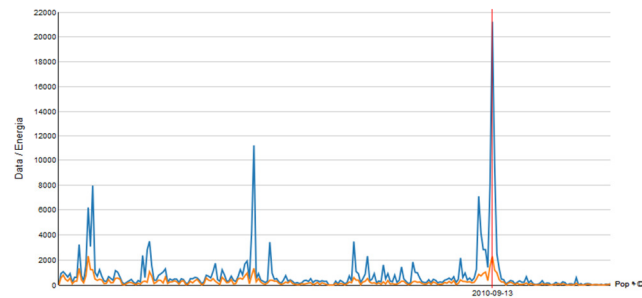
Rysunek 5. Struktura strony z danymi użytkowania.

Kolejność w rozkładzie prawdopodobieństwa użycia węzłów (Tabela 1) różni się od kolejności pozycji w menu interfejsu. Jest to bezpośrednia wskazówka dla projektanta interfejsu jak ułożyć menu.

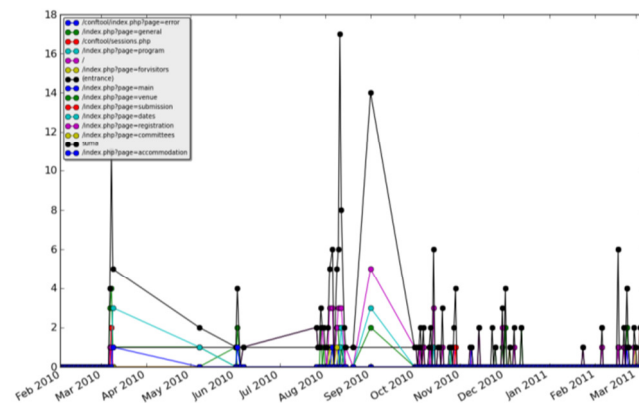
Tabela 1. Rozkład prawdopodobieństwa użycia węzłów icss.pwr.wroc.pl

Węzły	Prawdopodobieństwo użycia
<code>/index.php?page=submission</code>	0,11887469
<code>/index.php?page=dates</code>	0,10646318
<code>/index.php?page=program</code>	0,10278569
<code>/index.php?page=main</code>	0,09405167
<code>/index.php?page=general</code>	0,09019031
<code>/index.php?page=registration</code>	0,06950446
<code>/index.php?page=accommodation</code>	0,05130091
<code>/index.php?page=committees</code>	0,05130091
<code>/index.php?page=venue</code>	0,03557966
<code>/index.php?page=forvisitors</code>	0,03530385

Rysunek 7 prezentuje zmianę ruchu do różnych węzłów w czasie.



Rysunek 6. Wykres energii strony icss.pwr.wroc.pl

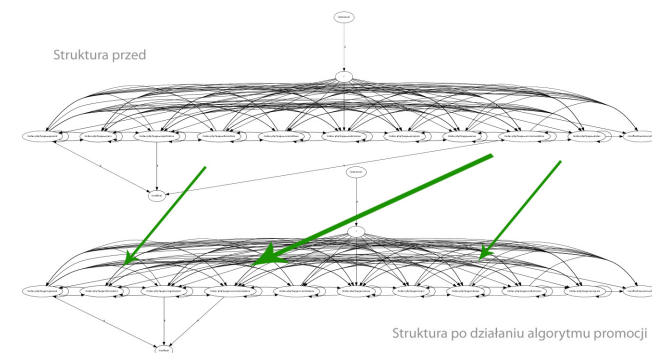


Rysunek 7. Zmiana ruchu do węzła Program.

Obliczono energię grafu struktury serwisu (Rysunek 6). Widać trzy okresy, w których wartość energii jest bardzo wysoka. Zbiega się to z terminami zgłoszenia abstraktu, artykułu oraz okresem organizacji konferencji. W związku z tym ruch do poszczególnych podstron zmienia się w zależności od terminarza konferencji. Badając na przykład węzeł Program (Rysunek 7) obserwuje się wyraźny wzrost w czasie

zgłaszania abstraktu (marzec 2010) oraz przed i w czasie konferencji (sierpień, wrzesień 2010).

Zastosowanie metody promocji nie zmieniło liczby węzłów ani krawędzi. Zasugerowana została zmiana kolejności występowania węzłów w interfejsie (**Rysunek 8**) proponując wypromowanie następujących węzłów: /index.php?page=submission, /index.php?page=general, /index.php?page=program, /index.php?page=main, /index.php?page=dates. Żadnego węzła nie ukarano. Energia nowego grafu wyniosła: 2430,43 czyli o 53% więcej.



Rysunek 8. Zmiana położenia węzłów nowym grafie.

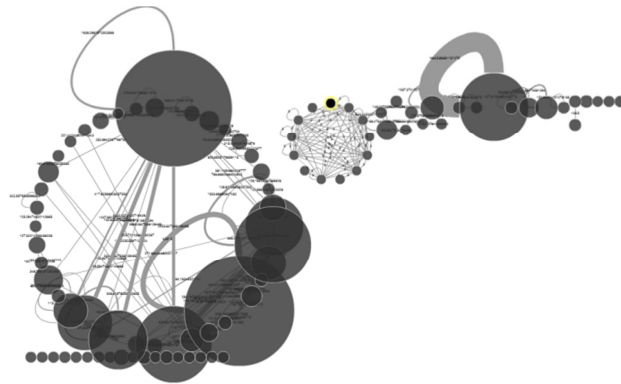
Jednolity System Obsługi Studentów Edukacja.CL

JSOS jest systemem obsługującym studentów, pracowników oraz wspomagającym pracę dziekanatów w Politechnice Wrocławskiej. Dla części odpowiedzialnej za obsługę studentów istnieje portal studencki dostępny w środowisku webowym pod adresem <http://edukacja.pwr.wroc.pl>. W okresie 01.08.2010-05.05.2011 zebrane zostały dane użytkownika, odkryta została struktura serwisu, którą

następnie przebadano w celu pozyskania danych potrzebnych do zadania zmiany interfejsu webowego portalu studenckiego.

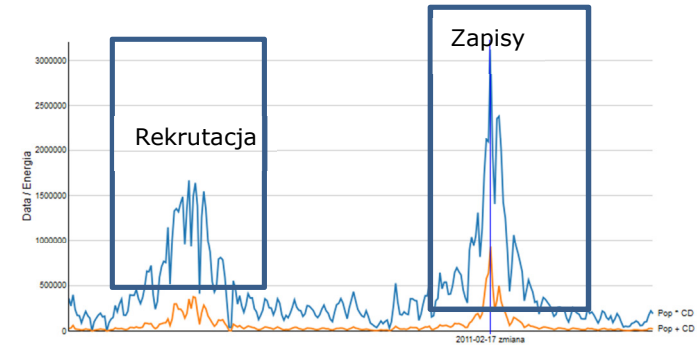
Struktura serwisu zawiera 92 węzły i 186 krawędzi (**Rysunek 9**), energia grafu wynosi 394 470,97.

Na Rysunku 9 także wyraźnie widać, że portal studenta podzielony jest na dwie części: informacyjną (po prawej) oraz funkcjonalną – dostępną po zalogowaniu (po lewej).

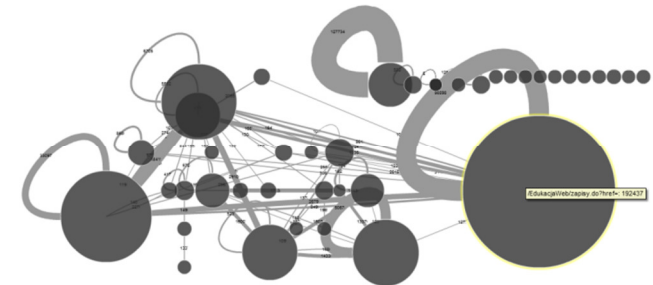


Rysunek 9. Struktura serwisu internetowego portalu studenta.

Środowisko Webgraph zostało wykorzystane w procesie badania klienta webowego JSOS w postaci portalu dla studenta oraz na etapie syntezy nowych prototypów. Szczególnie przydatny okazał się moduł analizy ruchu do poszczególnych węzłów struktury, co pozwoliło na selekcję takich węzłów do głównego menu nawigacyjnego prototypu, jakie miały największe użytkowanie w badanym okresie.



Rysunek 10. Zmiana energii grafu dla systemu JSOS.



Rysunek 11. Graf struktury JSOS z danymi użytkowania z dnia 17 lutego 2011 r. Podświetlony węzeł odpowiada za funkcjonalność zapisów.

Na **Rysunek 10** zaznaczona prostokątami sytuacja, w której należało zmienić strukturę serwisu internetowego w związku z intensyfikacją użyciu wybranych obszarów interfejsu. W pierwszym przypadku było to związane z rekrutacją studentów w JSOS, w drugim – z zapisami na przedmioty semestru letniego.

Na podstawie **Rysunek 11** odnotowano, że największą popularnością wśród użytkowników cechowały się 6 węzłów:

Zapisy, Wybór Słuchacza, Strona startowa, Funkcjonalność „Zapisz” oraz „Wybór Kryteriów” w dziale zapisów.

Tabela 2. Rozkład prawdopodobieństwa użycia węzłów portalu studenta JSOS.

Węzły	Prawdopodobieństwo użycia
/EdukacjaWeb/logInUser.do	0,118719
/EdukacjaWeb/zapisy.do?href=	0,1110173
/EdukacjaWeb/zapisy.do	0,0740466
/EdukacjaWeb/zapisy.do?event=WyborSluchacza	0,0733211
(other)	0,0653803
/EdukacjaWeb/studia.do	0,0542948
/EdukacjaWeb/logoutUser.do	0,0531204
/EdukacjaWeb/	0,0496178
/EdukacjaWeb/zapisy.do?event=ZapiszFiltr&event=wyborKryterium&href=	0,0279914
/EdukacjaWeb/start.do	0,0271183
/EdukacjaWeb/indeks.do?event=WyborSluchacza	0,021664

/EdukacjaWeb/rejestrSpraw.do?event=EntryIssues	0,0185717
/EdukacjaWeb/dorobekSluchacza.do?href=	0,015420
/EdukacjaWeb/rozkladyZajec.do?event=zapiszFiltr	0,0150615
/EdukacjaWeb/lov.do	0,014435
/EdukacjaWeb/rozkladyZajec.do?event=WyborSluchacza	0,0126583
/EdukacjaWeb/dorobekSluchacza.do	0,0125358
/EdukacjaWeb/zapisy.do?event=ZapiszFiltr&event=submitForm&href=	0,0120332
/EdukacjaWeb/oplaty.do?event=WyborSluchacza	0,011134
/EdukacjaWeb/prowGrupyZajeciowe.do	0,0111158
/EdukacjaWeb/changePassword.do	0,0101321
/EdukacjaWeb/zapisyMojeRozkladyZajec.do	0,0090114
/EdukacjaWeb/nowePodanie.do?event=RefreshIssues	0,0083122
/EdukacjaWeb/ocenySluchacza.do?event=WyborSluchacza	0,0078205

Po zastosowaniu algorytmu promocji wypromowane zostają węzły: /EdukacjaWeb/opłaty.do/, /EdukacjaWeb/changePassword.do, /EdukacjaWeb/logInUser.do, /EdukacjaWeb/zapisy.do, /EdukacjaWeb/logOutUser.do, /EdukacjaWeb/start.do /EdukacjaWeb/naliczoneOplaty.do, /EdukacjaWeb/rejestrSpraw.do, /EdukacjaWeb/rozkladyZajec.do, /EdukacjaWeb/ocenySluchacza.do, /EdukacjaWeb/zapisyMojeRozkladyZajec.do.

Ukarane węzły to: /EdukacjaWeb/dorobekSluchacza.do, /kontaktRekrutacje.do, /uczelnia.do, /pokazInformatoryRekrutacji.do, /ankiety.do, /studia.do, /EdukacjaWeb/lov.do, /EdukacjaWeb/prowGrupyZajeciowe.do, /EdukacjaWeb/kursSprawy.do, /EdukacjaWeb/zapisyNaRekrutacje.do, /EdukacjaWeb/nowePodanie.do, /EdukacjaWeb/remindPassword.do, /EdukacjaWeb/rozhchunki.do

Podsumowanie

Zaproponowane w pracy podejście do analizy jakości struktur serwisów internetowych z uwzględnieniem danych użytkownika ma charakter ogólny, umożliwia przy tym obiektywną ocenę i porównanie struktury serwisu internetowego z innymi oraz ciągły jej monitoring. Zaprojektowane algorytmy szacowania jakości struktury i syntezy alternatywnej struktury umożliwiają dostosowanie badanej struktury serwisu internetowego do sposobu użytkowania jej przez użytkowników, tym samym zmniejszają koszt ponoszony przez nich w celu osiągnięcia swoich żądań. Opracowana narzędzie pozwala na zwiększenie skuteczności metodyki projektowania serwisów internetowych

oraz cechuje się skutecznością sprawdzoną eksperymentalnie na prawdziwych serwisach internetowych. Narzędzie będzie testowane na większej liczbie serwisów internetowych i w ostateczności upublicznione w postaci usługi w modelu SaaS.

Podziękowania

Badania, których wyniki są przedstawione w tym artykule, były częściowo sfinansowane przez:

1. Ministerstwo Nauki i Szkolnictwa Wyższego z Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Innowacyjna Gospodarka projekt nr POIG.01.03.01-00-008/08;
2. Stypendium „Młoda Kadra” współfinansowanego przez Unię Europejską w ramach Europejskiego Funduszu Społecznego.

Literatura

- [1] Douglas K. Van Duyne, James A. Landay, and Jason I. Hong, *The Design of Sites*. Upper Saddle River, NJ: Pearson Education, Inc., 2007.
- [2] Sanjay J. Koyani, Robert W. Bailey, and Janice R. Nal, *The Research-Based Web Design & Usability Guidelines*.: U.S. Department of Health and Human Services, 2006.
- [3] Dave Gehrke and Efraim Turban, "Determinants of Successful Website Design: Relative Importance and Recommendations for Effectiveness," in *Thirty-second Annual Hawaii International Conference on System Sciences-Volume 5*, Maui, Hawaii, 1999, p. 5042.
- [4] Jakob Nielsen, *Designing Web usability: The practice of simplicity*. Indianapolis: New Riders Publishing,

- 1999.
- [5] Ping Zhang and Gisela M. von Dran, "User Expectations and Rankings of Quality Factors," *International Journal of Electronic Commerce Vol.6 No.2*, pp. 9-33, 2002.
- [6] Raquel Benbunan-Fich, "Using protocol analysis to evaluate the usability of a commercial web site," *Journal Information and Management Vol. 39 Issue 2*, 2001.
- [7] Janna B. Arney and Paul J. Lazarony, "An Inclusive Guide To Assessing Web Site Effectiveness," *Journal of College Teaching & Learning Vol.2, Number 1*, pp. 27-36, 2005.
- [8] Jinwoo Kim, Jungwon Lee, Kwanghee Han, and Moonkyu Lee, "Businesses as Buildings: Metrics for the Architectural Quality of Internet Businesses," *Information Systems Research Vol. 13 No.3*, pp. 239-254, 2002.
- [9] Johnathan W. Palmer, "Web site usability, design, and performance metrics," *Information Systems Research Vol.13 No.2*, pp. 151-168, June 2002.
- [10] Layla Hasan and Emad Abuelrub, "Assessing the Quality of Web Sites," *INFOCOMP Journal of Computer Science Vol.7 No.4*, 2008.
- [11] Horton S. Lynch P.J., *Web style guide: basic design principles for creating Web sites*. NJ: Yale University Press, 2009. [Online].
<http://info.med.yale.edu/caim/manual>
- [12] George M. Donahue, "Usability and the Bottom Line," *IEEE Software 18 Issue 1*, pp. 31-37, 2001.
- [13] National Institute of Standards&Technology. (2002, May) WebSAT Evaluation Rules. [Online].
http://zing.ncsl.nist.gov/WebTools/WebSAT/websat_rules.html
- [14] George A. Miller. (1956) The Music Animation Machine. [Online].
<http://www.musanim.com/miller1956/>
- [15] Mike Perkowitz and Oren Entzoni, "Adaptive Sites: Automatically Learning from User Access Patterns," Washington, Technical report UW-CSE-97-03-01 1997.
- [16] John Garofalakis, Panagiotis Kappos, and Dimitris Mourloukos, "Web Site Optimization Using Page Popularity," *Web Software*, pp. 22-29, July-August 1999.
- [17] S. Rosenbaum, "The Future of Usability Evaluation," in *Maturing Usability.*: Springer-Verlag London Limited, 2008, pp. 344-378.
- [18] Mohammad El-Ramly and Eleni Stroulia, "Analysis of Web-usage behavior for focused Web sites: a case study," *Journal of Software Maintenance and Evolution: Research and Practice*, pp. 129-150, 2004.
- [19] Fabi Paternò and Carmen Santoro, "Remote Usability Evaluation," in *Maturing Usability.*: Springer-Verlag London Limited, 2008, pp. 197-221.
- [20] Gong Chibing and Mats Nordahl, "Building an Adaptive Website Based on User Access Patterns," in *Proceedings of the 2005 International Conference on Cyberworlds (CW'05)*, Singapore, 2005.
- [21] Benjamin Yen, Paul Hu, and May Wang, "Toward an analytical approach for effective Web site design: A framework for modeling, evaluation and enhancement," *Electronic Commerce Research and Applications 6*, pp. 159-170, 2007.

- [22] Sheng-Yuan Yang, "An ontological website models-supported search agent for web services," *Expert Systems with Applications* 35, pp. 2056–2073, 2008.
- [23] Nicoleta David and Liviu Stelian Begu, "A Website Structure Optimization Model," in *ACS'10 Proceedings of the 10th WSEAS international conference on Applied computer science*, Iwate, 2010, pp. 426-429.
- [24] Serge Brin and Larry Page, "The anatomy of a large-scale hypertextual Web search engine," in *Proceedings of the VII International World Wide Web Conference, in: Computer Networks and ISDN Systems vol. 30*, 1998, pp. 107-117.
- [25] Jonathan Jeffrey, Peter Karski, Björn Lohrmann, Keivan Kianmehr, and Reda Alhajj, "Optimizing Web Structures Using Web Mining Techniques," in *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, vol. 4881, Birmingham, 2007, pp. 653-662.
- [26] Edmond H. Wu and Michael K. Ng, "A Graph-Based Optimization Algorithm for Website Topology Using Interesting Association Rules," in *Advances in Knowledge Discovery and Data Mining, 7th Pacific-Asia Conference, PAKDD 2003, Seoul, Korea*, vol. 2637, Berlin, 2003, pp. 178-190.
- [27] Wen-long Lin and Ye-zheng Liu, "A Novel Website Structure Optimization Model for More Effective Web," in *Workshop on Knowledge Discovery and Data Mining, Adelaide*, 2008, pp. 36-41.
- [28] Pao-Hua Chou, Pi-Hsiang Li, Kuang-Ku Chen, and Menq-Jiun Wu, "Integrating web mining and neural network for personalized e-commerce," *Expert Systems with Applications Vol. 37*, pp. 2898-2910, 2010.
- [29] Youwei Wang, Weihui Dai, and Yufei Yuan, "Website browsing aid: A navigation graph-based recommendation system," *Decision Support Systems Vol. 45*, pp. 387–400, 2008.
- [30] Hamed Q. Saremi, Babak Abedin, and Amirhosein M. Kermani, "Website structure improvement: Quadratic assignment problem approach and ant colony meta-heuristic technique," *Applied Mathematics and Computation Vol. 195*, pp. 285-298, 2008.
- [31] Brad Campbell and Joseph M. Goodman, "HAM: a general purpose hypertext abstract machine," *Communications of the ACM*, vol. 31, no. 7, pp. 856-861, 1988.
- [32] Paloma Díaz, Ignacio Aedo, and Fivos Panetsos, "Labyrinth, an abstract model for hypermedia applications," *Information Systems*, vol. 22, no. 8, pp. 447-464, 1997.
- [33] Paloma Díaz, Susana Montero, and Ignacio Aedo, "Modelling hypermedia and webapplications: the Ariadne Development Method," *Information Systems*, vol. 30, pp. 649-673, 2005.
- [34] Dmitrij Żatuchin, "Metoda przebudowy interfejsu serwisu internetowego oparta na historii użytkownika," in *Interfejs użytkownika : Kansei w praktyce*, Krzysztof Marasek and Marcin Sikorski, Eds. Warszawa: Wydawnictwo Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych, 2010, pp. 98-105.
- [35] Dmitrij Żatuchin and Adam Grzech, "Evaluation of website interface quality.," in *Advances in systems science*, Adam Grzech, Paweł Świątek, and Jarosław Drapała, Eds., 2010, pp. 271-280.