

Karol Łopaciński

Uniwersytet Ekonomiczny we Wrocławiu

KONCEPCJA TEMATYCZNEJ HURTOWNI DANYCH NA POTRZEBY BADANIA SKUTECZNOŚCI PROMOCJI W INTERECIE

Streszczenie: Prowadzenie skutecznych kampanii promocyjnych w Internecie wymaga systematycznego badania efektów realizowanych działań. Szacowanie skuteczności e-promocji związane jest z gromadzeniem, przetwarzaniem i analizą danych pochodzących z różnych źródeł. Szerokie możliwości integrowania rozproszonych zasobów informacyjnych oraz ich analizowania dają obecnie systemy klasy *Business Intelligence* oparte na hurtowniach danych. W artykule zaprezentowany zostanie ogólny projekt tematycznej hurtowni danych, której zadaniem ma być gromadzenie zasobów informacyjnych na temat przebiegu kampanii promocyjnych na stronach internetowych. Na podstawie zebranych w ten sposób danych możliwe będzie przeprowadzanie analiz wielowymiarowych OLAP, których zadaniem będzie określanie skuteczności prowadzonych przez przedsiębiorstwo działań promocyjnych w Internecie.

Słowa kluczowe: promocja internetowa, pomiar skuteczności promocji w Internecie, tematyczna hurtownia danych, przetwarzanie analityczne *online* (OLAP).

1. Wstęp

Coraz szerszy zakres wykorzystania Internetu przez przedsiębiorstwa w ich działalności promocyjnej powoduje konieczność badania jej wyników i efektów. Promocja internetowa wiąże się z coraz wyższymi nakładami, przez co skłania menedżerów do trafnego jej planowania i optymalizacji tychże działań.

Analiza i szacowanie efektów promocyjnych to niewątpliwie trudny etap procesu zarządzania promocją internetową, co wynika z konieczności rozwiązania wielu problemów metodologicznych dotyczących stosowanych miar, wskaźników oraz budowy odpowiednich modeli obserwowanych zjawisk. Wypełnienie wynikającej stąd ewidentnej luki w literaturze skłania do proponowania nowych ujęć uwzględniających złożoną naturę badanego problemu oraz duże rozproszenie źródeł danych.

Szerokie możliwości integrowania rozproszonych zasobów informacyjnych oraz analizowania danych dają obecnie systemy klasy *Business Intelligence*. Zawierają one szeroki zespół technologii i narzędzi opartych na hurtowni danych. Szczególnie użyteczne wydają się narzędzia OLAP (przetwarzania analitycznego *online*)

służące do przeprowadzania analiz wielowymiarowych na danych zgromadzonych w hurtowni. Wyniki analiz są prezentowane użytkownikom końcowym w formie kompleksowych zestawień, raportów i wizualizacji, które w czytelny i przejrzysty sposób obrazują przebieg i efekty dokonanych przez przedsiębiorstwo działań.

Wykorzystanie tego potencjału do pomiaru efektów promocji internetowej jest jednak niewystarczające. W praktyce gospodarczej badanie skuteczności promocji internetowej często jest prowadzone w ograniczonym zakresie i niejednokrotnie sprowadza się do analiz wykorzystujących tylko niektóre z istniejących metod, narzędzi i technologii. Istnieje zatem potrzeba zaproponowania i opracowania nowych podejść do badania i pomiaru skuteczności promocji w Internecie.

Powiązanie wykorzystywanych do tej pory metod szacowania skuteczności e-promocji z możliwościami rozwiązań *Business Intelligence* daje szansę na stworzenie narzędzia, które istotnie wspomagałoby proces zarządzania promocją w Internecie.

W niniejszym artykule zaprezentowany zostanie ogólny projekt tematycznej hurtowni danych, której zadaniem ma być gromadzenie zasobów informacyjnych na temat przebiegu kampanii promocyjnych na stronach internetowych. Na podstawie zebranych w ten sposób danych możliwe będzie przeprowadzanie analiz wielowymiarowych OLAP, których zadaniem będzie określanie skuteczności prowadzonych przez przedsiębiorstwo działań promocyjnych w Internecie.

2. Istota i cechy przetwarzania analitycznego OLAP

Ze względu na coraz bardziej złożony charakter danych gromadzonych i przetwarzanych przez współczesne przedsiębiorstwa miejsce prostych baz danych zaczęły zajmować hurtownie danych zawierające ogromne zbiory zasobów informacyjnych o bardzo złożonej strukturze. Hurtownie danych, będące podstawą i niezbędnym elementem systemów *Business Intelligence*, są popularnym sposobem gromadzenia i przechowywania danych w systemach informatycznych przedsiębiorstw i instytucji. W.H. Inmon, będący twórcą zagadnienia hurtowni danych (*data warehousing*), definiuje ją jako: „centralne repozytorium wszystkich istotnych danych, które są gromadzone przez poszczególne systemy biznesowe przedsiębiorstwa” [Inmon 1992]. Hurtownie danych są niejako odpowiedzią na problemy związane z funkcjonowaniem w przedsiębiorstwach wielu rozproszonych i niepowiązanych ze sobą baz danych. Przetwarzanie i analiza danych zgromadzonych w wielu miejscach i będących w różnych postaciach jest procesem wymagającym dużych nakładów pracy, czasu i wiążącym się z wysokimi kosztami. Hurtownie danych eliminują tego typu problemy, ponieważ gromadzą dane o różnych formatach, pochodzące z odrębnych źródeł i obszarów danej organizacji, i umieszczają je w jednolitym, zunifikowanym, nieredundantnym środowisku. Możliwe jest dzięki temu wykorzystanie zgromadzonych danych przez aplikacje analityczne, jak i zapytania interfunkcjonalne oraz przekrojowe użytkowników. Podstawowym celem hurtowni danych jest natomiast

odkrywanie, porównywanie lub identyfikacja nowych wzorców danych, stanowiących bazę informacyjną określonych problemów decyzyjnych.

W tej sytuacji analizy danych prowadzone przy użyciu arkuszy kalkulacyjnych oraz zapytań w języku SQL stały się niewystarczające, firmy bowiem zaczęły zgłaszać zapotrzebowanie na bardziej złożone analizy wielowymiarowe. Odpowiedzią na te potrzeby są rozwiązania analityczne OLAP, które w pełni współpracują z hurtownią danych oraz dają duże możliwości w zakresie złożonej analizy zawartych w niej danych.

Technologię OLAP, czyli przetwarzanie analityczne *online*, można zdefiniować jako wielowymiarową analizę danych na potrzeby użytkownika biznesowego, obejmującą złożone mechanizmy raportowania, analizy i wizualizacji danych [Surma 2009]. Aplikacje OLAP umożliwiają użytkownikom, tj. osobom zgłaszającym zapotrzebowanie na informacje zarządcze, intuicyjny oraz szybki dostęp do obszernych zasobów danych w celu ułatwienia wniosków opartych na rozumowaniu analitycznym [<http://www.microstrategy.pl...>]. OLAP ma za zadanie wsparcie procesów decyzyjnych poprzez dostarczenie narzędzi i technik umożliwiających analizę danych w wielu wymiarach określonych przez użytkowników [Miłosz, Muryjas 2004].

Do podstawowych cech technologii przetwarzania analitycznego *online* możemy zaliczyć:

- wysoką szybkość przetwarzania danych,
- możliwość dokonywania analiz danych w wielu wymiarach według kryteriów wyszukiwania na różnych poziomach szczegółowości,
- intuicyjny i multimedialny interfejs, który umożliwia użytkownikowi bezpośrednio oddziaływanie na budowę tabel i redefiniowanie obiektów tymczasowych [Olszak, Ziemia 2007],
- możliwość generowania złożonych raportów bez wymogu znajomości języków programowania [Olszak, Ziemia 2007],
- szybkie generowanie odpowiedzi na skomplikowane zapytania zadawane *ad hoc*.

OLAP skupia się na dostarczaniu zestawu cech danych z baz zorganizowanych według pewnych kryteriów, np. produktu, czasu lub regionu sprzedaży [Dyche 2002]. Umożliwia to wielowymiarowy wgląd w dane, identyfikację i syntezę nowych, niewidocznych wcześniej zależności między danymi oraz sporządzanie prognoz dotyczących okresów przyszłych na podstawie danych wtórnych.

3. Pomiar aktywności użytkowników na witrynach internetowych

Prowadzenie kampanii promocyjnych przez współczesne przedsiębiorstwa niesie ze sobą potrzebę monitorowania ich przebiegu i wyników. W tradycyjnym marketingu najczęściej wymaga to złożonej procedury badawczej, w której wykorzystywane są rozmaite metody ilościowe i jakościowe. W przypadku promocji internetowej sytuacja jest korzystniejsza ze względu na specyfikę i możliwości Internetu w zakre-

sie zapisywania i archiwizowania szerokiego spektrum danych na temat poczynań internautów na stronach WWW [Collin 2002]. Dlatego też najczęściej stosowaną metodą pomiaru skuteczności kampanii promocyjnych w Internecie jest monitoring stron WWW, na których prowadzone są owe działania promocyjne. Monitorowanie serwisów internetowych polega na wykorzystaniu narzędzi umożliwiających analizę ruchu odbywającego się na stronie.

Pomiar ruchu na stronach WWW w trakcie trwania kampanii promocyjnych prowadzony jest na poszczególnych witrynach internetowych przy wykorzystaniu zainstalowanego na nich oprogramowania trackingowego. Rejestruje ono aktywność wszystkich użytkowników w obszarze całej witryny. Są to przede wszystkim dane dotyczące: wejść na witrynę, wyjść z witryny, słów lub fraz, po których wpisaniu w wyszukiwarce nastąpiło wejście na stronę, kliknięć na elementy znajdujące się na witrynie oraz przejścia na podstrony, dokonywania różnego rodzaju akcji (np. zakup produktu w sklepie internetowym, pobieranie plików, umieszczanie komentarzy itp.), zakresu korzystania z aplikacji interaktywnych umieszczonych na stronie. Działania te opisane są w czasie z dokładnością co do sekundy, dzięki czemu możliwe jest szacowanie długości wizyt, obserwowanie czasu przebywania w poszczególnych sekcjach strony czy też określanie okresów o największym natężeniu ruchu na stronie.

Dane dotyczące aktywności użytkownika na stronie uzyskiwane są dzięki tzw. skryptom trackingowym, które program monitorujący umieszcza w kodzie źródłowym strony. Skrypty zliczają dane o ruchach i działaniach użytkowników witryny oraz na bieżąco przesyłają je na odpowiedni serwer witryny, gdzie są zapisywane oraz archiwizowane w formie logów serwera, czyli plików log files. Logi, zwane również plikami dziennika lub rejestrem zdarzeń [[http://pl.wikipedia.org/wiki/Log_\(informatyka\)](http://pl.wikipedia.org/wiki/Log_(informatyka))], stanowią szczegółowy oraz chronologicznie uporządkowany zapis wszystkich działań i sytuacji, do których doszło na danej witrynie. Na podstawie logów możliwe jest więc zgromadzenie danych dotyczących ruchu na stronie. Mogą one stanowić podstawę sporządzania przydatnych statystyk i zestawień. Zebrane dane można bowiem odpowiednio przetworzyć poprzez ich uporządkowanie, pogrupowanie oraz dokonanie pożądaných podsumowań lub obliczeń innego typu.

Programy trackingowe wykorzystują technologię *cookies*, która opiera się na rozpoznawaniu użytkowników monitorowanych stron bez naruszania ich prywatności. *Cookies*, czyli tzw. ciasteczka, to małe informacje tekstowe, które serwer sieciowy określonej strony WWW przesyła do przeglądarki internetowej danego użytkownika. W momencie wejścia przez internautę na daną witrynę jej serwer zwraca się z zapytaniem o numer identyfikacyjny przeglądarki [Kaushik 2007]. W sytuacji gdy przeglądarka nie wysyła serwerowi takiego numeru, zostaje on automatycznie wygenerowany i przypisany jej przez serwer. Informacje o numerze identyfikacyjnym zostają ostatecznie zapisane przez przeglądarkę w niewielkim pliku tekstowym. Należy podkreślić, że do każdego wygenerowanego numeru ID przypisane są ogólne cechy użytkownika korzystającego z przeglądarki, tj.: miejsce, z którego użytkow-

nik połączył się ze stroną (kraj, województwo, miejscowość itp.), rodzaj i wersja stosowanej przeglądarki internetowej, rodzaj i wersja systemu operacyjnego, rozdzielczość ekranu monitora.

Przy kolejnych wizytach użytkownika na tej samej stronie WWW jej serwer otrzyma od przeglądarki dane zawarte w pliku *cookie*. Gdy plik zostaje skasowany przez użytkownika podczas następnego wejścia na stronę, jej serwer przesyła nowy *cookie*. Celem omówionej operacji jest więc oznaczenie przeglądarki unikalnym identyfikatorem, dzięki czemu możliwe jest późniejsze rozpoznawanie danego użytkownika, nazywanego wówczas użytkownikiem unikalnym (*unique user*). Trzeba ponadto zaznaczyć, że domyślne parametry „ciasteczek” pozwalają na odczytanie informacji w nich zawartych jedynie serwerowi, który je utworzył.

Programy trackingowe odpowiedzialne za gromadzenie danych działają w sposób ciągły oraz zbierają dane o aktywności każdego internauty na stronie bez jego wiedzy. Zgoda internautów na ten proces nie jest wymagana, ponieważ nie podają oni żadnych informacji osobowych, które dałyby szansę ich identyfikacji. Ich prywatność nie jest więc naruszana. Wyjątkiem jest możliwość lokalizacji geograficznej użytkownika po numerze IP komputera, jednakże jej dokładność jest ograniczona jedynie do miasta. Pomiar aktywności użytkowników strony WWW umożliwia uzyskanie całościowego obrazu tego, co dzieje się na stronie, pozwala na wyciąganie odpowiednich wniosków oraz podejmowanie na ich podstawie takich decyzji, jak: zmiany wyglądu i ergonomii strony WWW, ulepszenie interfejsu czy też modyfikacje dotyczące umiejscowienia, wyglądu i treści form promocyjnych umieszczanych w obrębie witryny.

4. Model konceptualny bazy danych OLAP na potrzeby badania skuteczności promocji internetowej

Dane zebrane w procesie monitoringu witryn internetowych mogą być bardzo cennym źródłem zasilającym hurtownię danych lub tematyczną hurtownię danych, tzw. *data mart*¹. Zorganizowane w ten sposób zasoby informacyjne mogą następnie stanowić podstawę do przeprowadzania różnorodnych analiz dotyczących skuteczności kampanii e-promocyjnych, np. przy wykorzystaniu narzędzi OLAP.

Zgromadzeniem, integracją i organizacją danych o przebiegu akcji e-promocyjnych oraz skupieniem ich w jednym środowisku, jakim jest hurtownia danych, są zainteresowane przede wszystkim osoby odpowiedzialne za zarządzanie obszarem promocji internetowej przedsiębiorstwa. Analiza i interpretacja zgromadzonych w hurtowni zasobów informacyjnych powinny więc pozwalać na udzielanie szybkich i trafnych odpowiedzi na pytania dotyczące operacyjnych, taktycznych, a nawet

¹ *Data mart* – tematyczna hurtownia danych tworzona w celu gromadzenia i organizacji danych odnoszących się do konkretnego obszaru biznesowego firmy, na potrzeby decyzyjne osób nim zarządzających; zazwyczaj wyodrębniona z głównej hurtowni danych przedsiębiorstwa.

strategicznych kierunków działań w obszarze promocji internetowej. Dyrektorzy, menedżerowie lub analitycy mogą zgłaszać do hurtowni danych pytania, takie jak np.:

- Które formy reklamowe umieszczone na stronach WWW wzbudzały największe zainteresowanie użytkowników odwiedzających serwis internetowy?
- Jak było zainteresowanie kampanią promocyjną na poszczególnych witrynach internetowych?
- Jak kształtowała się liczba osób, które zareagowały na reklamę internetową w poszczególnych województwach w trakcie trwania kampanii na witrynie?
- Które działania e-promocyjne należy zintensyfikować, aby osiągnąć założone cele?
- Która forma reklamowa wykorzystana w kampanii przyniosła najlepsze wyniki w stosunku do ceny zakupu miejsca na witrynie, na której została umieszczona?

Uzyskanie odpowiedzi na powyższe pytania jest możliwe dzięki instrumentom analitycznym OLAP, które korzystając z danych zawartych w hurtowni, udostępniają użytkownikowi końcowemu odpowiednio przetworzone wielowymiarowe informacje. Narzędzia OLAP mogą być więc wykorzystywane do tworzenia kompleksowych zestawień na temat dokonanych działań promocyjnych w Internecie w ujęciach dostosowanych do potrzeb użytkownika. Analizowane dane są zazwyczaj zintegrowane na najwyższym poziomie ogólności, lecz w momencie wystąpienia konkretnej potrzeby informacyjnej mogą być rozwijane do dowolnego poziomu szczegółowości. Wielowymiarowość danych w OLAP wiąże się z możliwością prezentacji wyników zapytań w formie tabel przestawnych, których osie mogą być dowolnie zmieniane. Niesie to ze sobą wiele korzyści, które wiążą się z analizą badanego zjawiska pod różnymi kątami i w różnych perspektywach.

Prowadzenie analiz OLAP na danych zgromadzonych w trakcie trwania akcji promocyjnych wymaga na wstępie zaprojektowania odpowiedniego repozytorium danych. Ponieważ pozyskane dane dotyczą bardzo wąskiego obszaru biznesowego, tj. monitorowania kampanii promocyjnych w Internecie, zasadny wydaje się wybór tematycznej hurtowni danych (*data mart*) jako miejsca gromadzenia danych. Hurtownie tematyczne często nazywane są również hurtowniami OLAP, ponieważ celem ich budowy jest organizacja i przetwarzanie danych *stricto* na potrzeby analiz i raportów OLAP.

Podstawową kwestią przy określaniu postaci logicznego (konceptualnego) modelu wspomnianej tematycznej hurtowni danych jest zaprojektowanie struktury wszystkich tabel, tj.

- określenie nazw poszczególnych tabel zawierających dane,
- wskazanie nazw pól (kolumn),
- sformułowanie typów danych, które będą się znajdować w określonych polach,
- zdefiniowanie kluczy w poszczególnych tabelach,
- określenie charakteru relacji między tabelami.

Projekt i sposób modelowania danych w hurtowni powinien wynikać z przewidywanego zakresu informacji, jakich ma ona dostarczyć na potrzeby decyzyjne w przedsiębiorstwie. Decydując się na konkretny model hurtowni, warto wcześniej wyspecyfikować wskaźniki biznesowe, których analiza, w różnych przekrojach i perspektywach, będzie prezentowana w formie raportów. Wskaźnikiem tym w omawianym przypadku będzie zainteresowanie daną kampanią e-promocyjną, które wyraża się oczekiwaną aktywnością poszczególnych internautów na stronie WWW podczas trwania akcji reklamowej. Oczekiwana aktywność polega na klikaniu na umieszczone na stronie formy reklamowe, które przenoszą zainteresowaną osobę na stronę zleceniodawcy reklamy.

Z powyższego wynika, że zainteresowanie kampanią będzie obiektem analizy, co oznacza, że będzie ono traktowane jako zbiór tzw. faktów. Fakt bowiem opisuje pojedyncze zdarzenie i jest daną ilościową, która reprezentuje aktywność biznesową [Nycz 2009]. Fakty natomiast opisane są przez miary i wymiary. Miarą nazywamy wartość liczbową przyporządkowaną do określonego faktu. Na przykład dla faktu sprzedaż miarą może być wartość sprzedaży lub liczba sprzedanych sztuk. W naszym przypadku fakt – zainteresowanie kampanią e-promocyjną – może być opisany przez takie miary, jak:

- liczba reakcji (kliknięć) na formę reklamową,
- liczba użytkowników, którzy zareagowali (kliknęli) na reklamę,
- liczba użytkowników (*cookies*) mających kontakt z reklamą (tj. użytkowników, którzy weszli na stronę, na której umieszczona była forma reklamowa),
- liczba kontaktów użytkowników z reklamą.

Wartość każdej miary uzależniona jest natomiast od zbioru wymiarów, który określa kontekst miary [Nycz 2009]. Wymiary są więc cechami opisującymi dany fakt. Dlatego też zainteresowanie kampanią promocyjną na witrynie WWW można rozpatrywać według takich wymiarów, jak:

- **czas** – odnosić się będzie do okresu trwania kampanii reklamowej na witrynie,
- **obszar** – dotyczy regionu geograficznego, z którego pochodzi internauta mający kontakt z reklamą; identyfikacja będzie następować poprzez analizę adresu IP komputera,
- **forma reklamowa** – obiekty reklamowe zawierające komunikaty promocyjne umieszczane na witrynach internetowych w celu przyciągnięcia uwagi internautów.

Zarówno fakty, jak i wymiary należy zawrzeć w odpowiednio zaprojektowanych tabelach, co umożliwi wyspecyfikowanie danych niezbędnych do późniejszej analizy. Poniżej przedstawione zostały układy poszczególnych tabel (tab. 1-4).

W każdej z tabel został zaznaczony poprzez podkreślenie tzw. klucz podstawowy. W danej tabeli jest nim pole lub pola, które jednoznacznie identyfikują wszystkie rekordy (wiersze) występujące w tej tabeli. Dane zawarte w polu (lub zestawie pól) oznaczonym/ych kluczem podstawowym są więc unikalne dla danej tabeli, czyli nie ma możliwości, aby się w niej powtórzyły. Dlatego też w przypadku

Tabela 1. Fragment tabeli faktów Zainteresowanie

Kod_Czasu	Kod_Obszaru	Kod_Formy_Rekl	Liczba_Reakcji	Liczba_Kontaktów	Liczba_Uzytkow_Reakcja	Liczba_Uzytkow_Kontakt
...
C_681	O_126	F_3	2	12	2	10
C_689	O_126	F_3	4	22	3	15
C_890	O_126	F_3	1	9	1	9
C_891	O_126	F_3	3	21	2	20
C_900	O_126	F_3	2	11	2	8
...

Źródło: opracowanie własne.

Tabela 2. Fragment tabeli dla wymiaru Czas

Kod_Czasu	Rok	Miesiąc	Dzień_tygodnia	Dzień	Godzina	Min_Sec
C_1	2010	6	czw	17	0	00:01
C_2	2010	6	czw	17	0	00:02
C_3	2010	6	czw	17	0	00:03
C_4	2010	6	czw	17	0	00:04
C_5	2010	6	czw	17	0	00:05
...

Źródło: opracowanie własne.

Tabela 3. Fragment tabeli dla wymiaru Obszar

Kod_Obszaru	Nazwa_Województwa	Nazwa_Powiatu	Nazwa_Gminy	Nazwa_Miejscowości
...
O_126	dolnośląskie	wrocławski	Sobótka	Sobótka
O_127	dolnośląskie	wrocławski	Sobótka	Strachów
O_128	dolnośląskie	wrocławski	Sobótka	Strzegomiany
O_129	dolnośląskie	wrocławski	Sobótka	Sulistrowice
O_130	dolnośląskie	wrocławski	Sobótka	Sulistrowiczki
...

Źródło: opracowanie własne.

Tabela 4. Fragment tabeli dla wymiaru Forma_Reklamowa

Kod_Formy_Rekl	Rodzaj_Formy_Rekl	Kod_Witryny	Nazwa_Witryny	Rodzaj_Witryny
F_1	Baner	W_2	Onet.pl	portal informacyjny
F_2	Billboard	W_2	Onet.pl	portal informacyjny
F_3	Brandmark	W_66	Podroze.pl	wortal tematyczny
F_4	Pop-up	W_2	Onet.pl	portal informacyjny
F_5	Pop-under	W_178	http:// otowakacje. ceneo.pl/	porównywarka cen
...	

Źródło: opracowanie własne.

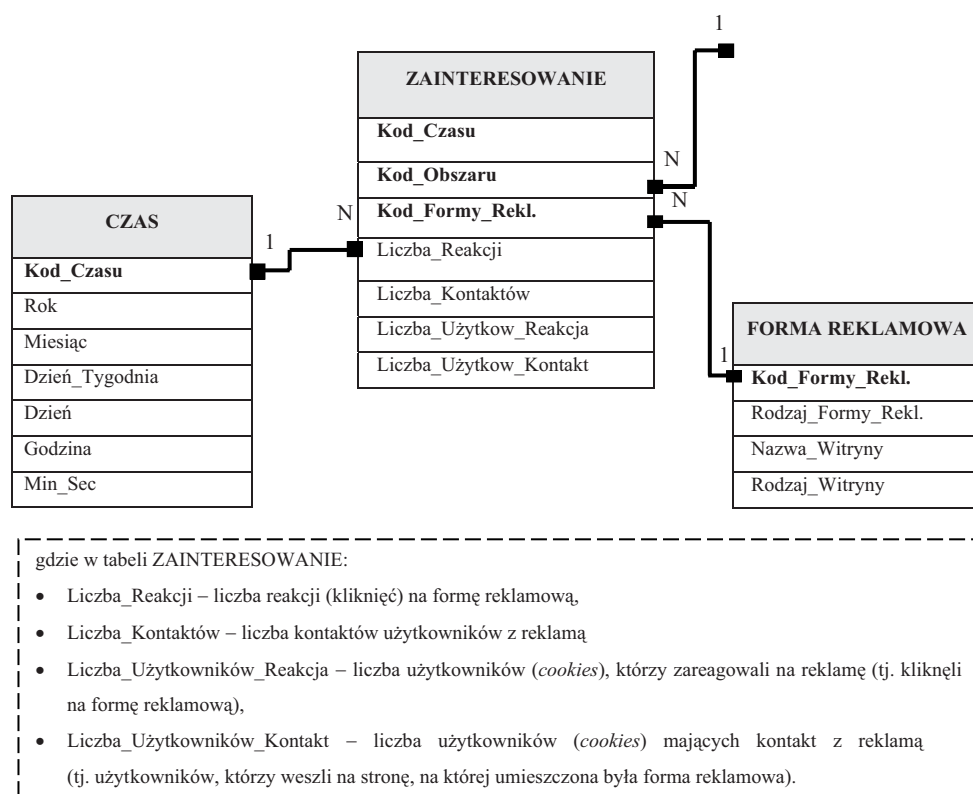
tabeli Czas kluczem podstawowym jest Kod_czasu, w tabeli Obszar Kod_obszaru, natomiast w tabeli Forma_Reklamowa – Kod_Formy_Rekl. Tabela faktów Zainteresowanie jest specyficzna, gdyż posiada tzw. klucz zagregowany. Tworzą go wspólnie trzy pola: Kod_Czasu, Kod_Obszaru i Kod_Formy_Rekl. Tabela faktów połączona jest relacjami z pozostałymi trzema tabelami wymiarów w następujący sposób:

- tabela Zainteresowanie z tabelą Czas łączy się poprzez pole Kod_Czasu,
- tabela Zainteresowanie z tabelą Obszar łączy się poprzez pole Kod_Obszaru,
- tabela Zainteresowanie z tabelą Forma_Reklamowa łączy się poprzez pole Kod_Formy_Rekl.

Warto zaznaczyć że, pola Kod_Czasu, Kod_Obszaru oraz Kod_Formy_Rekl w tabeli Zainteresowanie są tzw. kluczami obcymi. Pola te są, jak wspomniano wcześniej, kluczami podstawowymi w tabelach Czas, Obszar i Forma Reklamowa. Oparte jest to na właściwości, że klucz obcy jest polem w jednej tabeli, której wartości są określone przez klucz podstawowy z innej tabeli [Poe i in. 2000]. Klucze podstawowe z tabel wymiarów łączą się więc poprzez relację z kluczami obcymi z tabeli faktów, co powoduje, że np. jednemu rekordowi z tabeli Obszar może odpowiadać wiele rekordów z tabeli Zainteresowanie. W tabeli Obszar każdy rekord ma bowiem unikalny identyfikator, do którego przypisany jest określony obszar z dokładnością do nazwy miejscowości. Dlatego też każdy identyfikator w tabeli Obszar może wystąpić tylko raz. Jednakże w tabeli Zainteresowanie może wystąpić wielokrotnie ze względu np. na zainteresowanie internautów z danej miejscowości różnymi formami reklamowymi w różnych okresach. Analogicznie jest w przypadku relacji tabel Obszar – Zainteresowanie i Forma_Reklamowa – Zainteresowanie. W modelu tym pojedyncza tabela faktów powiązana jest poprzez relacje typu jeden do wielu ze znajdującymi się wokół niej tabelami wymiarów. Taki typ wielowymiarowej struktury hurtowni danych jest charakterystyczny dla schematu gwiazdy.

Na rys. 1 przedstawiono układ tabel i relacji, który jest propozycją tzw. modelu logicznego (konceptualnego) hurtowni, w której będą gromadzone i analizowane dane dotyczące monitorowania kampanii promocyjnych w Internecie. Taki sposób pre-

zentacji układu bazy danych nazywamy diagramem związków encji [Surma 2009]. Wszystkie cztery tabele są według tego standardu nazewnictwa zbiorami encji, co oznacza, że pojedynczy rekord każdej z tabel stanowi jedną encję. Wyszczególnione nazwy pól są zatem opisami atrybutów, które charakteryzują dany zbiór encji. Pola zaznaczone na diagramie pogrubioną czcionką oznaczają klucz encji w zbiorze encji. Zobrazowany na rysunku diagram związków encji jest więc graficzną postacią modelu konceptualnego (logicznego) tematycznej hurtowni danych, w której będą gromadzone i analizowane dane na temat monitorowania kampanii promocyjnych w Internecie. Realizacja wspomnianego modelu w określonej bazie danych będzie tzw. modelem fizycznym [Surma 2009].



Rys. 1. Diagram związków encji dla modelu konceptualnego tematycznej hurtowni danych na potrzeby monitorowania kampanii promocyjnych w Internecie

Źródło: opracowanie własne.

Z diagramu wynika, że zainteresowanie kampanią będzie obiektem analizy, co oznacza, że będzie ono traktowane jako zbiór faktów zawartych w tabeli Zainteresowanie. Fakty natomiast opisane są przez miary i wymiary.

Na rysunku widać również, że zainteresowanie kampanią promocyjną na witrynie WWW można rozpatrywać według takich wymiarów, jak: czas (odnosić się będzie do okresu trwania kampanii reklamowej na witrynie), obszar, tj. region geograficzny, z którego pochodzi internauta mający kontakt z reklamą, oraz forma reklamowa, tzn. obiekty reklamowe zawierające komunikaty promocyjne umieszczone na witrynach internetowych w celu przyciągnięcia uwagi internautów.

Wyróżnione wymiary są niejako zbiorami uporządkowanych wartości określonego typu. Przede wszystkim jednak są one perspektywami, z których można dokonywać analizy zgromadzonego zbioru danych dotyczących czynnika, jakim jest zainteresowanie konkretnymi działaniami promocyjnymi w sieci.

Hurtownia danych o opisanej powyżej strukturze gwiazdy zawiera informacje na temat zainteresowania komunikatami promocyjnymi zamieszczonymi za pośrednictwem obiektów reklamowych na stronach WWW. Informacje te są zagregowane ze względu na czas, obszar oraz formę reklamową. Jeżeli przedsiębiorstwo korzystające z tego repozytorium danych przykładą dużą wagę do kampanii promocyjnych w Internecie, można założyć, że prowadzi działalność promocyjną w sieci od dłuższego czasu oraz z dużą częstotliwością. W takim przypadku przedsiębiorstwo musi się liczyć z istnieniem gigantycznej ilości danych z okresu kilku lat, dotyczących wszystkich akcji promocyjnych prowadzonych za pośrednictwem wielu form reklamowych, na wielu witrynach, we wszystkich możliwych regionach. Decydenci lub analitycy zainteresowani są jednak pobraniem z tak pojemnej bazy konkretnych informacji w formie zwartych raportów zestawień czy też wizualizacji danych. Odpowiedzią na te oczekiwania są możliwości wielowymiarowych analiz OLAP. Odpowiednio zdefiniowane zapytania do hurtowni danych dają szansę na analizę danych poprzez kolejne poziomy szczegółowości [Jarke i in. 2003]. Dzięki hierarchicznej strukturze tabel wymiarów określone zakresy danych o efektach działań promocyjnych mogą być bowiem zagregowane na wyższych poziomach ogólności, a następnie, w razie potrzeby, rozwijane do poziomów bardziej szczegółowych.

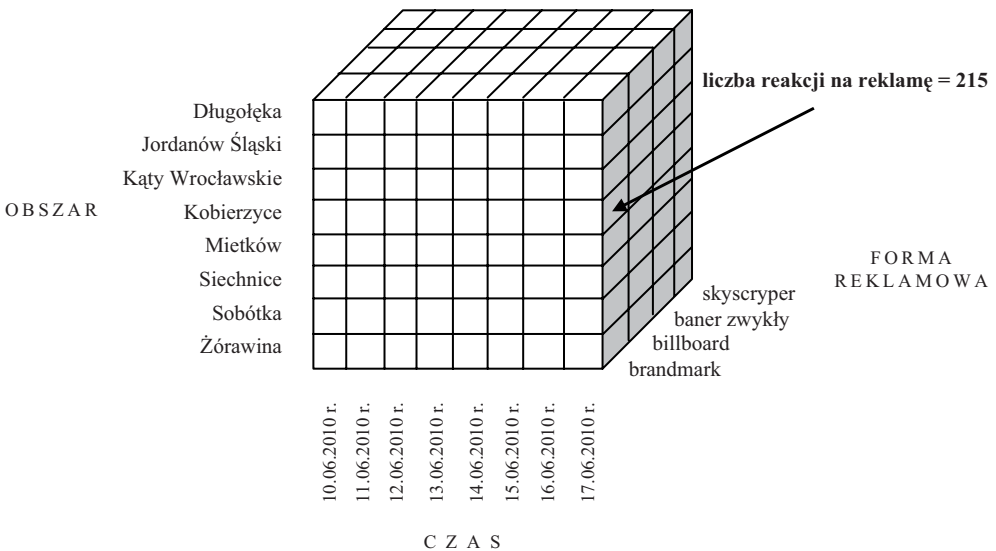
5. Określenie struktury wielowymiarowej kostki OLAP na potrzeby analizy skuteczności akcji promocyjnych w Internecie

Korzystanie z możliwości, jakie daje aplikacja OLAP, wymaga na wstępie odpowiedniego zaprojektowania tzw. kostek OLAP, które obrazują wielowymiarową strukturę danych dotyczących określonego obszaru działalności e-promocyjnej. Kostki mogą mieć nawet kilkadziesiąt wymiarów oraz zawierać dane numeryczne i tekstowe pochodzące z tabel. Kostki OLAP są więc wielowymiarowymi tabelami, których liczba oraz budowa (tj. rodzaj, liczba wymiarów itp.) zależą od indywidualnego zapotrzebowania na informację zgłaszanego przez użytkownika. W omawianym przypadku dane na temat efektów kampanii promocyjnych będziemy analizować przez pryzmat trzech wymiarów, tj. czasu, obszaru i formy reklamowej. Wymiary w mniejszym

lub większym stopniu charakteryzują się hierarchiczną strukturą, tzn. składają się z wynikających z siebie poziomów szczegółowości. Na przykład wymiar Obszar posiada hierarchię województwo – powiat – gmina – miejscowość. Przede wszystkim jednak wymiary będą stanowić osie współrzędnych dla wybranego faktu, jakim będzie liczba reakcji na formę promocyjną. Dane dotyczące faktów zawartych w obszarze kostki wielowymiarowej odpowiadają tym polom (kolumnom) relacyjnej bazy danych, których wartości zależą od innych pól. Dlatego też wartości zawarte w polu Liczba_Reakcji tabeli faktów Zainteresowanie (zob. rys. 1) będą odpowiadać wartościom miar tabeli wielowymiarowej, tzn. będą się znajdować w określonych lokalizacjach powstałej przestrzeni wielowymiarowej [Jarke i in. 2003]. Aby poznać wartość konkretnej miary znajdującej się wewnątrz kostki OLAP, użytkownik musi sprecyzować wielkości odpowiednich wymiarów istniejących na jej osiach. W analizowanym przypadku, aby poznać wartość faktu, którym jest liczba reakcji na formę reklamową, należy podać interesujący nas obszar, zakres czasu oraz wybrać określoną formę reklamową. Na przykład dla następujących wartości wymiarów:

- Czas – 17 czerwca 2010 r.,
- Obszar – gmina Długołęka,
- Forma reklamowa – obiekt reklamowy typu brandmark umieszczony na portalu informacyjnym Onet.pl,

wartość faktu liczba reakcji na formę reklamową wynosi 215. Zostało to zobrazowane na rys. 2.



Rys. 2. Kostka wielowymiarowa dla faktu liczba reakcji na formę reklamową przez pryzmat wymiarów czas, obszar i forma reklamowa

Źródło: opracowanie własne.

Zdefiniowanie wartości wszystkich trzech wymiarów spowodowało więc zlokalizowanie pojedynczej komórki w tabeli wielowymiarowej. Należy zauważyć, że uzyskana wartość faktu równa 215 reakcji na reklamę odnosi się do okresu jednego dnia. Ponieważ rejestracja ruchu na witrynie odbywa się co 1 sekundę, uzyskany wynik dla pojedynczego dnia wynika z podsumowania pojedynczych faktów znajdujących się na najniższym poziomie szczegółowości.

Odpowiednio zaprojektowana kostka OLAP gromadzi więc informacje na temat faktów określających cechy ilościowe oraz wymiarów, które wprowadzają hierarchiczną strukturę danych i formułują zakres ich analizy [Olszak, Ziemba 2007].

Użytkownicy końcowi, jak np. decydenci, analitycy, chcący w pełni korzystać z zasobów informacyjnych zawartych w kostce OLAP mają do dyspozycji szereg operacji analizy wielowymiarowej, jak [Jarke i in. 2003; Surma 2009]:

- 1) rozwinięcie, drażnienie (*roll-down, drill-down*),
- 2) zwinięcie, agregacja (*drill-up, roll-up*),
- 3) selekcja, wycinanie (*slice and dice*),
- 4) obracanie, zmiana orientacji (*pivot, rotate*),
- 5) działania na wskaźnikach [Surma 2009].

Operacje te umożliwiają sprawne poruszanie się wewnątrz kostki, jak również w znacznym stopniu wspomagają proces wyszukiwania pożądaných informacji, tworzenia wizualizacji i zestawień danych, a także ich interpretacji oraz wyciągania właściwych wniosków.

6. Zakończenie

Zaprezentowana w artykule koncepcja tematycznej hurtowni danych jest rozwiązaniem, które może znacznie wspomóc badanie efektów prowadzone przez przedsiębiorstwo kampanii promocyjnych w Internecie. Zgromadzenie w jednym środowisku danych pozyskiwanych na stronach internetowych przez systemy trackingowe niesie ze sobą duże możliwości. Właściwie uporządkowane w formie tabel zasoby informacyjne mogą być przede wszystkim podstawą dla analiz OLAP, które służą do przeprowadzania analiz wielowymiarowych na danych zgromadzonych w hurtowni oraz obejmują złożone mechanizmy raportowania, analizy i wizualizacji danych [Surma 2009]. Uzyskane w ten sposób wysoce przetworzone informacje na temat przebiegu oraz skuteczności poszczególnych akcji promocyjnych na stronach WWW pozwolą kadrze zarządzającej podejmować trafniejsze decyzje dotyczące kierunków dalszych działań promocyjnych w Internecie.

W dzisiejszym dynamicznie zmieniającym się otoczeniu firmy menedżerowie ds. marketingu i promocji powinni podejmować decyzje, bazując na znajomości użytecznej i możliwie pełnej informacji dostarczonej im w przejrzystej i zagregowanej formie. Często powinno się wówczas pojawiać pytanie „dlaczego?” – w odniesieniu do przeszłości i przyszłości. Funkcjonalność i możliwości narzędzi analizy wielowymiarowej OLAP mogą w istotnym stopniu wspomagać odpowiedzi na tego typu pytania.

Literatura

- Collin S., *Marketing w sieci*, Wydawnictwo Felberg, Warszawa 2002.
- Dyba A., *Cookies, czyli permanentna inwigilacja*, „Modern Marketing” 1999, nr 5.
- Dyche J., *CRM – Relacje z klientami*, Helion, Gliwice 2002.
- Inmon W., *Building the Data Warehouse*, John Wiley & Sons, New York 1992.
- Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P., *Fundamentals of Data Warehouses*, Springer-Verlag, Berlin-Heidelberg 2003.
- Kaushik A., *Web Analytics: An Hour a Day*, Wiley Publishing, Indiana 2007.
- Miłosz M., Muryjas P., *Współczesne technologie informatyczne – Obiektowość w informatyce*, Wydawnictwo Mikom, Warszawa 2004.
- Nycz M., *Techniki pozyskiwania wiedzy w hurtowni danych*, [w:] R. Knosala (red.), *Komputerowo zintegrowane zarządzanie*, t. II, Oficyna Wydawnicza PTZP, Opole 2009.
- Olszak C., Ziemia E., *Strategie i modele gospodarki elektronicznej*, Wydawnictwo Naukowe PWN, Warszawa 2007.
- Płoski Z., *Słownik encyklopedyczny – Informatyka*, Wydawnictwo Europa, Wrocław 1999.
- Poe V., Klauer P., Brobst S., *Tworzenie hurtowni danych*, Wydawnictwa Naukowo-Techniczne, Warszawa 2000.
- Surma J., *Business Intelligence. Systemy wspomaganie decyzji biznesowych*, Wydawnictwo Naukowe PWN, Warszawa 2009.

Źródła internetowe

- [http://pl.wikipedia.org/wiki/Log_\(informatyka\)](http://pl.wikipedia.org/wiki/Log_(informatyka)).
- <http://www.microstrategy.pl/OLAP/>.

DATA MART CONCEPTION TO RESEARCH THE EFFECTIVENESS OF INTERNET PROMOTION

Summary: Executing an effective promotional campaign on the Internet requires systematic research of the effects of promotional actions. Estimating effectiveness of the company's e-promotion means a constant need of collecting, processing and analysing data coming from various sources. Business Intelligence tools based on Data Warehouse offer wide opportunities of integrating, analysing and processing information resources. The article presents a general data mart project, which serves as a tool to retain the information resources concerning the course of promotional campaigns on websites. This method of data collection will allow to make multi-dimensional OLAP analyses, which task is to measure the effectiveness of e-promotion.

Key words: Internet promotion, measuring the effectiveness of promotion in the Internet, Data Warehouse OLAP.