

Justyna Wilk

Uniwersytet Ekonomiczny we Wrocławiu

SEGMENTACJA INTERNAUTÓW Z WYKORZYSTANIEM DANYCH SYMBOLICZNYCH I METOD KLASYFIKACJI

1. Wstęp

Internet jest miejscem tworzenia, promowania i sprzedaży różnorodnych produktów i usług. Rynek internetowy ze względu na „wirtualny” charakter dokonywanych transakcji jest rynkiem specyficznym, dlatego należy zwrócić szczególną uwagę na możliwości zaspokajania potrzeb jego segmentów. Badaniem, na podstawie którego firmy formułują strategie marketingowe, jest segmentacja rynku. Proces segmentacji składa się z wyboru kryteriów segmentacji, klasyfikacji obiektów rynkowych i profilowania wyodrębnionych segmentów.

Segmentację przeprowadza się na podstawie danych, zgromadzonych zwykle w wyniku badań ankietowych. Złożoność uzyskiwanych informacji sprawia, że dane w ujęciu klasycznym stają się niewystarczające do opisu konsumentów. Możliwość znacznie pełniejszego opisu dają dane symboliczne. Wśród technik segmentacji rynku szczególną rolę odgrywają metody klasyfikacji. Mimo wielu rodzajów i algorytmów nie zostały one przystosowane do analizy danych symbolicznych. Złożony charakter i struktura danych symbolicznych wymagają stosowania osobnych, adresowanych do nich rozwiązań metodologicznych.

W literaturze polskiej nie ma opracowań podejmujących zagadnienia segmentacji na podstawie danych symbolicznych z wykorzystaniem metod klasyfikacji. W piśmiennictwie obcojęzycznym problematyka ta poruszana jest w nielicznych pracach. Celem artykułu jest prezentacja specyfiki danych symbolicznych i wskazanie ich źródeł w badaniach segmentacyjnych, propozycja procedury klasyfikacji obiektów opisanych zmiennymi symbolicznymi i segmentacja internautów na podstawie danych symbolicznych i metod klasyfikacji.

2. Specyfika danych symbolicznych

Bock, Diday i in. [*Analysis...* 2000, s. 2] wśród podstawowych typów zmiennych symbolicznych wymieniają zmienne „klasyczne” (mierzone na mocnych lub słabych skalach pomiaru) oraz zmienne symboliczne *sensu stricto*, tj. przedział liczbowy, listę wariantów, listę wariantów z wagami i zmienną strukturalną. Realizacją zmiennej w postaci przedziału liczbowego jest przedział (zbiór) wartości o wariantach rozłącznych (np. dochód respondenta) lub nierozłącznych (np. miesięczna wielkość wydatków na produkt w zł: [10, 40], [15, 60], [5, 30], [50, 70]). Szczególnym przypadkiem zmiennej jest „klasyczna” zmienna metryczna.

Zmienna w postaci listy wariantów dopuszcza występowanie wielu kategorii (lub wartości) dla pojedynczego obiektu, np. znajomość języków obcych, posiadane kategorie prawa jazdy. Jeśli dla każdego obiektu realizacją zmiennej jest pojedyncza kategoria, to zmienna ma charakter „klasycznej” zmiennej niemetrycznej. Podobna do niej jest lista wariantów z wagami, z tym że wariantom nadane są stopnie ważności, prawdopodobieństwa, częstości lub udziały, np. struktura wydatków konsumenta na zakupy w hipermarketach, takich jak Auchan (40%), Real (30%), Tesco (20%) i Carrefour (10%). Interpretacja zmiennej jest identyczna jak w przypadku listy wariantów, gdy wszystkie warianty mają takie same wagi.

Tabela 1. Forma pytania ankietowego a charakter zmiennej

Rodzaj pytań ankietowych	Skala pomiaru zmiennej klasycznej				Typ zmiennej symbolicznej <i>sensu stricto</i>			
	nominalna	porządkowa	przedziałowa	ilorazowa	przedział liczbowy	lista wariantów	lista wariantów z wagami	zmienna strukturalna
Pytania alternatywne	+	+						
Kafeteria dysjunktywna	+							+
Kafeteria koniunktywna						+		+
Skala wartościująca [#]			+	+	+		+	
Skala semantyczna		+						
Skala Likerta		+				+		
Skala pozycyjna [#]		+	+	+	+			
Skala rangowa		+						
Skala stałych sum* [#]					+		+	
Pytania powiązane (np. filtrujące)								+
Pytanie otwarte	+	+	+	+	+	+	+	+

* W klasycznym rozumieniu pomiaru dokonuje się na skali porządkowej lub ilorazowej.

[#] Zmienna symboliczna umożliwia określenie przedziału liczbowego zamiast konkretnej wartości.

Źródło: opracowanie własne.

Zmienną strukturalną można zaprezentować w formie struktury drzewiastej o wielu poziomach. Poziomy zmiennej mogą mieć dowolny charakter (np. wartości, kategorie). Różnice między zmiennymi strukturalnymi różnego typu wynikają z rodzaju relacji, jaka zachodzi między jej poziomami. Systematyka zmiennej taksonomicznej jest ogólnie przyjęta (znana) lub ustalana *a priori* przez badacza, np.

samochód można charakteryzować za pomocą trzech poziomów, takich jak: marka, wersja i pojemność silnika (np. Honda Civic 4D 1.8, Renault Laguna GT 2.0). W zmiennej hierarchicznej wariant z poziomu wyższego determinuje warianty z poziomu niższego. Zależność logiczna występuje wtedy, gdy wartość jednego poziomu zależy logicznie lub funkcyjnie od wartości drugiego poziomu.

Obiekt opisany zmiennymi symbolicznymi określany jest obiektem symbolicznym, a zbiór obiektów symbolicznych zapisywany jest w tablicy danych symbolicznych.

Dane symboliczne w badaniach segmentacyjnych wynikają przede wszystkim z:

- pierwotnych cech zjawiska, m.in. gdy realizacja zmiennej dla obiektu zwykle nie jest pojedynczą wartością liczbową (np. wartość jest trudna do oszacowania) ani pojedynczą kategorią (np. odzwierciedla kilka aspektów zjawiska), opisuje zagadnienia złożone,
- braku dostępu do szczegółowych danych, np. w wyniku ich celowego utajnienia,
- intencji badacza, gdy nie interesują go dokładne wartości zmiennej i przyjęta została reguła zbierania danych w zagregowanej formie (np. udziałów procentowych),
- interpretacji badacza, polegającej na analizowaniu danych w zagregowanej formie.

Tabela 1 prezentuje formy pytań ankietowych i odpowiadające im rodzaje zmiennych.

3. Metody klasyfikacji danych symbolicznych

Gatnar [1998], uwzględniając charakter danych empirycznych, wyróżnia metody taksonomii numerycznej i taksonomii symbolicznej. Metody taksonomii numerycznej zostały opracowane dla obiektów opisanych zmiennymi „klasycznymi”. Złożoność danych symbolicznych uniemożliwia stosowanie tych metod w sposób bezpośredni. Metody bazujące na macierzy danych wymagają przekodowania danych symbolicznych na dane „klasyczne”. Wiąże się to jednak z utratą informacji. Z kolei metody bazujące na macierzy odległości wymagają zastosowania miar podobieństwa właściwych dla obiektów symbolicznych (zob. [Analysis... 2000, s. 165-185]). Takie podejście stosuje Gordon [1999, s. 136-142].

Metody taksonomii symbolicznej umożliwiają analizowanie obiektów opisanych zmiennymi symbolicznymi (zob. [Gatnar 1998; Analysis... 2000; Verde 2004]). Metody podziałowe sprowadzają się do metod optymalizacji iteracyjnej i powstały na gruncie klasyfikacji dynamicznej (m.in. metoda klasyfikacji dynamicznej dla danych symbolicznych (SCLUST) [Verde 2004] oraz klasyfikacji pojęciowej, np. CLUSTER) [Gatnar 1998, s. 81-89]. Wśród metod hierarchicznych można wyróżnić metody aglomeracyjne (m.in. metodę hierarchiczną Brito), deglomeracyjne, jak np. klasyfikacja podziałowa oparta na kryteriach (zob. [Analysis... 2000, s. 299-323]), i sekwencyjne, dokonujące zarówno aglomeracji, jak i deglomeracji zbioru obiektów, np. COBWEB [Gatnar 1998, s. 108-114].

4. Charakterystyka badania

W celu pozyskania danych o internautach i ich preferencjach przeprowadzono badanie ankietowe. Badanie jest prowadzone w Internecie od maja 2008 r. Kwestionariusz ankiety składa się z trzech części. Część pierwsza dotyczy użytkownika Internetu. Druga grupa pytań podejmuje kwestię zakupów przez Internet. Część trzecia zawiera metryczkę. Prezentowane wyniki dotyczą danych częściowych, pobranych dla okresu maj-czerwiec. Spośród respondentów wybrano osoby aktywne w Internecie. W tym kontekście warunek losowości badanej zbiorowości nie był najistotniejszy. Badaniem objęto grupę 128 osób z województwa dolnośląskiego (wyniki badań m.in. GUS [2008] wskazują, że aktywność internautów jest zróżnicowana regionalnie). Strukturę respondentów prezentuje tab. 2.

Tabela 2. Struktura respondentów

Zmienna	Struktura zmiennej
Płeć	kobieta (54%), mężczyzna (46%)
Wiek	poniżej 20 (1%), 20-25 (31%), 26-30 (36%), 31-35 (12%), 36-40 (4%), 41-45 (4%), 46-50 (3%), 51-55 (4%), 56-60 (3%), powyżej 60 lat (2%)
Wykształcenie	podstawowe (0%), gimnazjalne (1%), zasadnicze zawodowe (3%), niepełne średnie ogólnokształcące (0%), niepełne średnie zawodowe (0%), średnie ogólnokształcące (8%), średnie zawodowe (12%), pomaturalne (7%), policealne (2%), licencjackie (5%), wyższe inżynierskie (13%), wyższe magisterskie (49%)
Stan cywilny	żonaty/mężatka (33%), partner/partnerka (46%), wdowiec/wdowa (0%), samotny/samotna (21%)
Miejsce zamieszkania	wieś (11%), miasto: do 50 tys. (18%), 50-100 tys. (51%), powyżej 100 tys. mieszkańców (20%)
Wykonywany zawód*	dyrektor/kierownik (3%), specjalista (23%), prywatny przedsiębiorca (7%), pracownik administracyjno-biurowy (14%), pracownik usług/handlu (17%), nauczyciel/wykładowca (3%), robotnik wykwalifikowany (3%), robotnik niewykwalifikowany (0%), rolnik (2%), wolny zawód (0%), gospodyni domowa (5%), uczeń/student (23%), emeryt/rencista (2%), bezrobotny (3%), inny (5%)
Miesięczny dochód netto	brak dochodu (6%), dochód: poniżej 1000 zł (17%), 1000-1500 zł (22%), 1501-2000 zł (24%), 2001-2500 zł (12%), 2501-3000 zł (9%), 3001-4000 zł (6%), powyżej 4000 zł (4%)
Źródło dochodu*	praca zarobkowa (81%), stypendium (11%), kieszonkowe (14%), renta (2%), emerytura (6%), inne (3%)

* Respondent mógł wskazać kilka wariantów odpowiedzi.

Źródło: opracowanie własne.

Badanie miało charakter pilotażowo-poznawczy. Głównym celem było zbadanie motywów dokonywania (lub niedokonywania) zakupów w Internecie, preferencji i zachowań konsumentów na „wirtualnym” rynku. Proces zakupu na rynku internetowym jest specyficzny. Z jednej strony wiąże się ze zwiększonym ryzykiem finansowo-rzeczowym. Z drugiej strony Internet oferuje dostęp do wielu sprzedawców, szeroki asortyment, wysoką dostępność produktów i możliwość szybkiego porówna-

nia ofert. O wyborze produktu mogą zatem decydować inne czynniki niż w przypadku rynku „rzeczywistego”. Ze względu na specyficzny charakter nie jest możliwe swobodne przeniesienie i aplikacja rozwiązań stosowanych na rynku „tradycyjnym”.

5. Systematyka konsumentów

Konsumentów można podzielić na osoby niekorzystające z Internetu oraz użytkowników Internetu. Przedmiotem zainteresowania w niniejszym badaniu byli internauci, tj. osoby, które przynajmniej raz w miesiącu korzystają z Internetu¹. Segmentacja internautów została przeprowadzona z wykorzystaniem podejścia hybrydowego. Za kryterium podziału w pierwszym kroku (podejście *a priori*) przyjęto zmienną „rola Internetu w procesie zakupu”. Na jej podstawie internautów podzielono na osoby niedokonujące zakupów w Internecie (37,75% respondentów) oraz osoby dokonujące zakupów przez Internet (62,25%).

Osoby niedokonujące zakupów za pośrednictwem Internetu jako główne tego powody wskazywały wysokie ryzyko zakupów i przyzwyczajenie do zakupów w sposób tradycyjny. Można ich podzielić na dwie grupy. Pierwszą grupę (46%) stanowią osoby, które wykorzystują Internet m.in. do poszukiwania informacji o produktach i usługach oraz deklarują dokonywanie w przyszłości zakupów przez Internet. Należą tu głównie gospodynie domowe z wykształceniem średnim, mieszkające w małych miastach lub na wsi. Ich średni miesięczny dochód netto nie przekracza 1000 zł.

Drugi segment (54%) skupia osoby, które nie wyraziły zamiaru zakupów w Internecie, mimo że połowa z nich czerpie z Internetu informacje o produktach i usługach. Są to przede wszystkim mężczyźni w wieku 51-60 lat, z wykształceniem przynajmniej średnim, najczęściej robotnicy, renciści lub emeryci, zamieszkujący w mieście średniej wielkości (50-100 tys. mieszkańców), o dochodach netto nieprzekraczających 1500 zł miesięcznie.

Drugą grupę internautów stanowią osoby dokonujące zakupów za pośrednictwem Internetu, tzw. e-konsumenci. Z przeprowadzonego badania wynika, że są to osoby aktywne w Internecie, korzystające z sieci od ponad pięciu lat (zwykle codziennie). Internet wykorzystują przede wszystkim w celu korzystania z usług bankowości elektronicznej, zakupów, komunikowania się oraz poszukiwania informacji. Największym zainteresowaniem w badanej grupie cieszą się wśród produktów – sprzęt AGD/RTV, a wśród płatnych usług – usługi bankowości elektronicznej, natomiast najmniejszym – produkty spożywcze. Ogólna (zagregowana ze względu na poziom konkurencyjności, ryzyka i zadowolenia) ocena zakupów przez Internet jest wysoka (powyżej oceny średniej). Przeprowadzono segmentację e-konsumentów z wykorzystaniem podejścia *post hoc*.

¹ Zgodnie z definicją opracowaną przez Internet Architecture Board (IAB) [www.wikipedia.pl].

5.1. Procedura klasyfikacji e-konsumentów

Procedurę segmentacji poprzedzono badaniem poziomu jednorodności obiektów symbolicznych za pomocą techniki *Symbolic Bi-Dimensional Mapping* [User manual... 2004, s. 103-104]. Wyniki ujawniły heterogeniczność zbioru obiektów, ale na ich podstawie trudno było wyodrębnić wyraźne klasy. Zdecydowano się więc na wykorzystanie metod klasyfikacji. E-konsumentów poddano klasyfikacji ze względu na kryteria zawarte w tab. 3.

Tabela 3. Zbiór zmiennych grupujących

Zmienna	Rodzaj zmiennej	Zbiór realizacji zmiennej
Rola respondenta w procesie zakupu*	lista kategorii	{decydent, nabywca, płacący, użytkownik}
Przyczyny dokonywania zakupów drogą internetową*	lista kategorii	{oszczędność czasu, wysoka dostępność produktów, szeroki asortyment produktów, możliwość szybkiego porównania ofert, atrakcyjne ceny, dostawa „do domu”, dyskrecja, inne}
Poziom ważności czynników* mających wpływ na dokonywane zakupy (%)	lista kategorii z wagami	{marka, cena, opis produktu, promocja, szybkość dostawy, koszt przesyłki, rekomendacje innych}
Rodzaj nabywanych produktów*	lista kategorii	{sprawdzone, nieużytkowane wcześniej, tradycyjne, nowoczesne}
Produkty kupowane najczęściej <i>on-line</i> *	lista kategorii	{książki/ filmy/muzyka/gry komputerowe, telefony/akcesoria GSM, odzież/ obuwie, sprzęt komputerowy/oprogramowanie, sprzęt fotograficzny, kosmetyki pielęgnacyjne/perfumy, sprzęt sportowy, sprzęt AGD/RTV, artykuły dziecięce/zabawki, produkty spożywcze, inne}
Miejsce i częstotliwość dokonywania zakupów (sklep internetowy, aukcja)	struktura hierarchiczna	nie kupuję (<i>NA</i>), kupuję (kilka razy w tygodniu, raz w tygodniu, kilka razy w miesiącu, raz w miesiącu, kilka razy w roku, raz w roku, rzadziej)
Nabywane odpłatne usługi <i>on-line</i> *	struktura hierarchiczna	nie korzystam (<i>NA</i>), korzystam (porady prawne, serwisy ekonomiczne, tłumaczenia obcojęzyczne, edukacja <i>on-line</i> , transakcje bankowe, inne)
Miesięczne wydatki na produkty i usługi zakupione w Internecie (zł)	przedział liczbowy nierozłączny	[0, 3000]
Ocena poziomu konkurencyjności ofert	porządkowa	1 (wysoki poziom), 2, 3, 4, 5 (niski poziom)
Ocena poziomu zadowolenia z zakupu	porządkowa	1 (wysoki poziom), 2, 3, 4, 5 (niski poziom)
Ocena poziomu ryzyka transakcji	porządkowa	1 (niski poziom), 2, 3, 4, 5 (wysoki poziom)

* Respondent mógł wybrać maksymalnie 4 warianty odpowiedzi.

Źródło: opracowanie własne.

W pierwszej kolejności ustalono liczbę klas. Najpierw wyznaczono odległości obiektów. Zmienne w postaci listy wariantów z wagami wymagają osobnych miar odległości od pozostałych zmiennych symbolicznych. W pomiarze odległości wykorzystano metrykę Minkowskiego dla listy kategorii z wagami, a dla pozostałych zmiennych – miarę Ichino-Yaguchiego ($\lambda = 0,3$). Następnie wyznaczono zagregowane odległości obiektów i dokonano klasyfikacji metodami hierarchicznymi – metodą Warda i kompletnego połączenia. Wyznaczono wstępną liczbę klas (3, 4, 7) i przeprowadzono ocenę jakości klasyfikacji za pomocą indeksu sylwetkowego. Indeks osiągnął najwyższą wartość dla podziału zbioru obiektów na 3 klasy. Następnie dokonano klasyfikacji obiektów metodą klasyfikacji dynamicznej dla obiektów symbolicznych. Do opisu klas obiektów symbolicznych można wykorzystać technikę CLINT [*User manual...* 2004, s. 207-214], a do profilowania – drzewa klasyfikacyjne dla obiektów symbolicznych (TREE) [*Analysis...* 2000, s. 266-293]. Jako zmienne profilowe przyjęto zmienne z pierwszej i trzeciej części ankiety.

5.2. Segmenty e-konsumentów

Wyodrębnione segmenty są zróżnicowane pod względem badanych zmiennych. Segment 1 (31% respondentów kupujących w Internecie) cechuje się najniższym potencjałem zakupowym – niska częstotliwość zakupów (kilka razy w roku) i relatywnie niskie wydatki na produkty i usługi zakupione w Internecie (do 400 zł miesięcznie). Najważniejszą przyczyną dokonywania zakupów w Internecie są konkurencyjne, w stosunku do zakupów w sposób tradycyjny, ceny. Produkty nabywane przez tę grupę konsumentów są zróżnicowane. Najczęściej są to produkty nieużytkowane wcześniej. Przy wyborze produktów i usług najczęściej wskazywanym kryterium jest szybkość dostawy, ale największą wagę przypisano opisowi produktu.

Ponad 80% osób korzysta z internetowego konta bankowego. Ten segment wysoko ocenia zakupy przez Internet (szczególnie poziom konkurencyjności ofert i zadowolenia z zakupu). Spędza w Internecie relatywnie najmniej czasu (do 4 godzin dziennie), głównie w celu poszukiwania informacji tematycznych. Można tu wyróżnić:

- grupę 1 (26% segmentu), tj. kobiety poniżej 20 roku życia, uczące się lub studiujące, mieszkające na wsi lub w małym mieście. Ich miesięczne dochody netto nie przekraczają 1000 zł i pochodzą głównie z kieszonkowego lub stypendium. Internet służy im wyłącznie do celów prywatnych, m.in. do zakupów i komunikowania się. Korzystają z internetowych kont pocztowych oraz komunikatorów internetowych (np. Gadu-Gadu, Skype),
- grupę 2 (74%) – należą tu mężczyźni powyżej 35 lat, z wykształceniem licencjackim bądź wyższym, pracujący na stanowisku specjalisty bądź jako nauczyciel lub wykładowca akademicki, żyjący w związku, mieszkający w mieście 50-100 tys. mieszkańców. Utrzymują się z pracy zarobkowej, z której otrzymują dochody w granicach 2000-3000 zł netto miesięcznie. Wykorzystują Internet

do celów zarówno prywatnych, jak i służbowych. Korzystają z internetowych kont pocztowych, a ponad 60% korzysta również z komunikatorów.

Najwyższy potencjał zakupowy charakteryzuje segment 2 (34% kupujących). Osoby z tego segmentu dokonują zakupów z relatywnie dużą częstotliwością, ich miesięczne wydatki są bardzo wysokie (do 1600 zł miesięcznie). Dokonują zakupów w Internecie przede wszystkim ze względu na oszczędność czasu oraz atrakcyjne ceny. Na wybór produktów ma wpływ wiele czynników, jednak największą wagę ma cena. Kupują głównie w sklepach internetowych. Częściej niż pozostałe segmenty nabywają produkty nowoczesne, szczególnie sprzęt komputerowy i AGD/RTV. Zwykle nie użytkują zakupionych produktów. Ponad połowa segmentu nie korzysta z żadnych płatnych usług dostępnych *on-line*.

Segment 2 reprezentują osoby z wykształceniem wyższym, mieszkające w mieście i żyjące w związku. Spędzają relatywnie najwięcej czasu w Internecie. Wykorzystują Internet do celów zarówno prywatnych, jak i służbowych. Można tu wskazać trzy grupy osób:

- grupa 1 (19%) – mężczyźni (pracujący studenci). Utrzymują się z pracy zarobkowej i stypendium, uzyskują dochód w wysokości 1500 zł miesięcznie. Internet służy im do poszukiwania informacji. Oprócz kont e-mail i komunikatorów korzystają z forów dyskusyjnych (50% grupy),
- grupa 2 (52%) to kobiety w wieku 26-30 lat, pracownice administracyjno-biurowe lub handlu i usług, o dochodach 1000-2500 zł netto miesięcznie. W Internecie dokonują zakupów, poszukują informacji bieżących i korzystają z konta bankowego,
- grupa 3 (29%) skupia mężczyzn w wieku 41-50 lat, prywatnych przedsiębiorców lub osoby na stanowisku dyrektora lub kierownika, o relatywnie wysokich dochodach (powyżej 3000 zł miesięcznie). Internet wykorzystują oni do poszukiwania informacji na konkretny temat i w celach zakupowych. Spędzają w nim najwięcej czasu – ponad 8 godzin dziennie.

Segment 3 (35%) to nabywcy zarówno produktów (książki, filmy, muzyka, gry komputerowe oraz sprzęt AGD/RTV), jak i odpłatnych usług (serwisy ekonomiczne, edukacja *on-line*). Jako podstawowe przyczyny robienia zakupów w Internecie wskazują oszczędność czasu oraz szeroki asortyment produktów. Dokonują zakupów częściej na aukcjach *on-line* niż w sklepach internetowych. Preferują produkty sprawdzone (użytkowane wcześniej). Najniżej oceniają zakupy w Internecie, szczególnie poziom konkurencyjności ofert oraz poziom ryzyka. Segment 3 składa się z dwóch grup, tj.:

- grupy 1 (43%) – studujący mężczyźni (prywatni przedsiębiorcy lub pracownicy usług i handlu), w stanie wolnym, mieszkający w mieście, utrzymujący się z pracy zarobkowej, o dochodach 1500-2000 zł miesięcznie. Spędzają w Internecie kilka godzin dziennie w celach prywatnych. Rzadziej niż osoby z pozostałych segmentów korzystają z komunikacji *on-line*.
- grupy 2 (57%) – osoby powyżej 26 roku życia, zatrudnione na stanowisku specjalisty, pracownicy administracyjno-biurowi, o relatywnie wysokich dochodach (2000-4000 zł miesięcznie), żyjący w wolnym związku, mieszkający w

mieście powyżej 50 tys. mieszkańców. Spędzają w Internecie do 12 godzin dziennie. Korzystają z forów dyskusyjnych.

6. Podsumowanie

Zmienne symboliczne umożliwiają odkrycie i analizowanie preferencji internautów w sposób zgodny z naturalnym charakterem zjawisk. W procesie segmentacji wymagają zastosowania specyficznych rozwiązań. Wyniki badania wskazują, że zakupy przez Internet cieszą się dużą popularnością, a wśród e-konsumentów można wyróżnić trzy segmenty o zróżnicowanych preferencjach: segment 1 – „Odkrywcy”, 2 – „Pośrednicy”, 3 – „Sceptycy”. Ponadto duży odsetek osób niekupujących stanowią potencjalni e-konsumenci.

Podobne analizy mogą posłużyć do konstruowania hipotez do dalszych badań oraz sformułowania strategii marketingowych firm działających na rynku internetowym lub wchodzących na niego. Przewiduje się powtórzenie badania na próbie reprezentatywnej.

Literatura

- Analysis of symbolic data. Exploratory methods for extracting statistical information from complex data* (2000), red. H.H. Bock, E. Diday, Springer, Heidelberg.
- Gatnar E. (1998), *Symboliczne metody klasyfikacji danych*, PWN, Warszawa.
- Gordon A.D. (1999), *Classification*, Chapman and Hall/CRC, London.
- User manual for SODAS 2 software* (2004), red. M. Noirhomme-Fraiture, Software Report, Analysis System of Symbolic Official Data, Project Number IST-2000-25161.
- Spoleczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2004-2006* (2008), „Informacje i opracowania statystyczne”, GUS, Warszawa.
- Verde R. (2004), *Clustering methods in symbolic data analysis*, [w:] *Classification, clustering and data mining applications*, red. D. Banks, L. House, E.R. McMorris, P. Arabie, W. Gaul, Springer-Verlag, Heidelberg, 299-317.

MARKET SEGMENTATION OF INTERNET USERS BASED ON SYMBOLIC DATA AND CLUSTER ANALYSIS

Summary

The aim of the paper is to suggest market segmentation procedure based on symbolic data and cluster analysis. The first part of the paper focuses on symbolic data and their sources in market segmentation research. Furthermore, the taxonomy of symbolic data clustering methods was presented. In the second part, the results of market segmentation of Internet users based on symbolic data from questionnaire survey were discussed. The hybrid approach was used to classify Internet users. First of all the consumers were divided into non-buyers and buyers via Internet. A special attention was paid to the second ones. The clustering procedure of symbolic objects was suggested. In the classification results, three segments of buyers via Internet were revealed: “Discoverers”, “Agents”, “Sceptics”.