

**Justyna Wilk**

Uniwersytet Ekonomiczny we Wrocławiu

---

## **METODY SEGMENTACJI RYNKU NA PODSTAWIE DANYCH SYMBOLICZNYCH**

---

**Streszczenie:** Stosowanie przez przedsiębiorstwo strategii segmentacji rynku wymaga precyzyjnej informacji o konsumentach. Opis konsumentów i segmentów w sposób pełniejszy umożliwiają dane symboliczne. Struktura danych symbolicznych uniemożliwia bezpośrednią aplikację metod opracowanych dla danych klasycznych (liczby rzeczywiste lub kategorie).

Problematyka metod segmentacji dla danych klasycznych jest dość dobrze opisana w literaturze. Natomiast tematyka segmentacji rynku na podstawie danych symbolicznych jest podejmowana w niewielu pracach. Celem artykułu jest zaprezentowanie podejść i metod analizy, jakie można stosować w segmentacji rynku na podstawie danych symbolicznych.

### **1. Wstęp**

Stosowanie przez przedsiębiorstwo strategii segmentacji rynku wymaga precyzyjnej informacji o konsumentach. Opis konsumentów i segmentów w sposób pełniejszy umożliwiają dane symboliczne. Struktura danych symbolicznych uniemożliwia bezpośrednią aplikację metod opracowanych dla danych klasycznych (liczby rzeczywiste lub kategorie).

O ile problematyka metod segmentacji dla danych klasycznych jest dość dobrze opisana w literaturze (zob. np. [Wedel, Kamakura 1998]), o tyle w przypadku danych symbolicznych brakuje publikacji na ten temat. Celem artykułu jest zaprezentowanie podejść i metod analizy, jakie można zastosować w segmentacji rynku na podstawie danych symbolicznych.

W pierwszej części artykułu przedstawiona zostanie specyfika i źródła danych symbolicznych w badaniach segmentacyjnych. W drugiej części artykułu zaprezentowane zostaną podejścia i metody segmentacji rynku na podstawie danych symbolicznych.

### **2. Specyfika danych symbolicznych**

W analizie danych symbolicznych rozróżnia się obiekty i zmienne symboliczne. Różnice między zmiennymi symbolicznymi a zmiennymi klasycznymi prezentuje tab. 1.

**Tabela 1.** Zmienne symboliczne a zmienne klasyczne

Lp.	Wyszczególnienie	Zmienne klasyczne	Zmienne symboliczne
1	Rodzaje zmiennych	metryczne (ilorazowe, przedziałowe), niometryczne (porządkowe, nominalne)	zmienne metryczne i niometryczne, listy wariantów, listy wariantów z wagami, przedziały liczbowe, zmienne strukturalne (hierarchiczne, taksonomiczne, logiczne)
2	Realizacja zmiennej	liczba rzeczywista, kategoria	liczba rzeczywista, kategoria, przedział liczbowy, zbiór wariantów (kategorii, wartości), zbiór wariantów z wagami, struktura drzewiasta
3	Liczba wariantów zmiennej dla obiektu	jeden wariant	jeden lub wiele wariantów
4	Warianty zmiennej	rozłączne	rozłączne lub nierozłączne
5	Wagi wariantów zmiennej dla obiektu	jednakowe	jednakowe lub zróżnicowane
6	Powiązania zmiennych relacjami	dopuszczalne, ale nie uwzględniane	dopuszczalne i uwzględniane

Źródło: opracowanie własne na podstawie pracy [Bock, Diday 2000].

Wyróżnia się trzy rodzaje obiektów symbolicznych (zob. [Bock, Diday 2000]):

- obiekt symboliczny I-go rzędu – obiekt rozumiany w sensie klasycznym (np. konsument, produkt, przedsiębiorstwo) opisany zmiennymi symbolicznymi,
- obiekt symboliczny II-go rzędu – jako wynik agregacji zbioru obiektów symbolicznych I-go rzędu, np. grupa konsumentów preferująca określoną markę produktu, rynek geograficzny złożony z podregionów,
- obiekt syntetyczny – reprezentant klasy obiektów symbolicznych I-go lub II-go rzędu.

W ujęciu klasycznym wielowymiarowa obserwacja zapisywana jest w formie wektora realizacji zmiennych, natomiast w analizie danych symbolicznych jako koniunkcja cech. Zbiór obiektów i zmiennych klasycznych prezentowany jest w macierzy danych, natomiast zbiór obiektów i zmiennych symbolicznych – w tablicy danych symbolicznych.

### 3. Źródła danych symbolicznych w zagadnieniach segmentacyjnych

Dane symboliczne można uzyskać w kilku etapach procedury badania segmentacyjnego – na etapie projektowania badania i gromadzenia danych, analizy danych oraz interpretacji wyników. Wśród źródeł danych symbolicznych można wymienić następujące:

1. Specyfika zjawiska, np. dane wielowariantowe (np. składniki produktu), nieprecyzyjne (np. struktura wydatków konsumenta), strukturalne (np. marka i model samochodu).

## 2. Metoda gromadzenia danych:

a) badania ankietowe. Dane symboliczne wynikają z konstrukcji kwestionariusza ankiety i różnych form pytań ankietowych, m.in. kafeterii koniunkcyjnej, skali stałych sum, pytań o kwestie poufne, np. dochód konsumenta [Wilk 2009c],

b) bazy danych. Narzędziem umożliwiającym wyszukiwanie danych w bazie danych są tzw. kwerendy, których wyniki mają postać tablicy danych klasycznych. Jednak w tej formie są trudne w interpretacji. Dokonuje się agregacji danych klasycznych, w rezultacie której otrzymywane są dane symboliczne (zob. [Bock, Diday 2000]).

## 3. Agregacja danych:

a) agregacja obiektów symbolicznych I rzędu, w wyniku której otrzymuje się obiekty symboliczne II rzędu [Bock, Diday 2000]. Agregacja obiektów znajduje zastosowanie w przypadku bardzo dużych zbiorów danych marketingowych,

b) agregacja zmiennych. Konsumentów zwykle opisano wieloma zmiennymi. Propozycją autorki jest metoda redukcji opisu polegająca na agregacji zmiennych klasycznych, w wyniku której uzyskuje się zmienne symboliczne [Wilk 2009d].

4. Opis klas za pomocą techniki CLINT [Lechevallier 2001, s. 31-33; Wilk 2007].

## 4. Podejścia w segmentacji rynku na podstawie danych symbolicznych

Złożona i nietypowa struktura danych symbolicznych powoduje, że w procesie segmentacji rynku należy zastosować specyficzne rozwiązania metodologiczne. Można w tym celu zaproponować następujące podejścia:

1. Przekształcenie danych symbolicznych na dane klasyczne.
2. Modyfikacja dotychczasowych (znanych) procedur.
3. Zastosowanie metod analizy danych symbolicznych.

Sposób pierwszy łączy się z utratą części informacji zawartych w zmiennych symbolicznych. Stosowalność podejścia drugiego jest z kolei determinowana możliwościami modyfikacji procedur. Na przykład w analizie danych symbolicznych można zastosować metody statystycznej analizy wielowymiarowej (SAW) opracowane dla danych klasycznych, bazujące na macierzy odległości, np. hierarchiczne metody klasyfikacji. Do pomiaru odległości należy jednak wykorzystać miary opracowane dla danych symbolicznych, np. miary Ichino-Yaguchiego lub de Carvalha (zob. [Bock, Diday 2000; Wilk 2005; 2006]).

Trzecim podejściem jest zastosowanie metod analizy danych symbolicznych (SDA), która jest rozszerzeniem metod SAW na dane symboliczne. Adaptacja metod SAW polegała na wprowadzeniu nowych algorytmów i procedur (np. metody skalowania wielowymiarowego, drzew decyzyjnych i większości algorytmów analizy skupień) lub modyfikacji znanych algorytmów (dla danych klasycznych), np. sztucznych sieci neuronowych.

## 5. Metody analizy danych symbolicznych w badaniach segmentacyjnych

Metody segmentacji klasyfikuje się ze względu na charakter metody i podejście w segmentacji – tab. 2.

**Tabela 2.** Metody stosowane w segmentacji rynku

Charakter metody	Podejście w segmentacji rynku	
	<i>a priori</i>	<i>post hoc</i>
Opisowa	tablice kontyngencji modele log-liniowe	metody klasyfikacji* skalowanie wielowymiarowe* analiza korespondencji analiza klas ukrytych sztuczne sieci neuronowe* analiza czynnikowa*
Predyktywna	tablice korelacyjne analiza regresji* analiza logitowa analiza dyskryminacyjna* regresja logistyczna	drzewa klasyfikacyjne* sztuczne sieci neuronowe* analiza <i>conjoint</i> regresja skupieniowa regresja klas ukrytych modele mieszanek rozkładów analiza kanoniczna*

\* – metody zaadaptowane na grunt SDA.

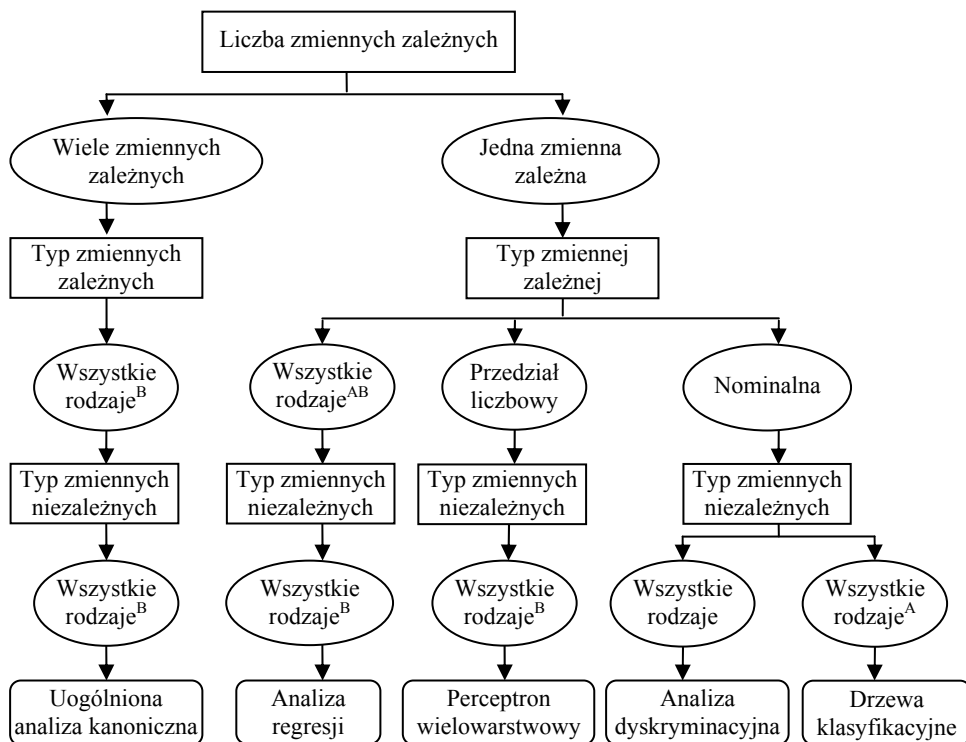
Źródło: opracowanie z wykorzystaniem [Wedel i Kamakura 1998; Walesiak 2000; Wilk 2009b].

Większość metod stosowanych w segmentacji rynku należy do grupy metod statystycznej analizy wielowymiarowej. Rozróżnia się metody badania zależności i współwystępowania. W metodach badania zależności analizuje się wpływ zmiennych niezależnych na zmienną (zmiennie) zależną (zależne). Z kolei celem metod badania współwystępowania jest ustalenie wzajemnego usytuowania obiektów lub zmiennych w wielowymiarowej przestrzeni.

Podstawowe metody badania zależności zaadaptowane na grunt SDA prezentuje rys. 1.

Metody badania współwystępowania opracowane w ramach SDA prezentuje rys. 2.

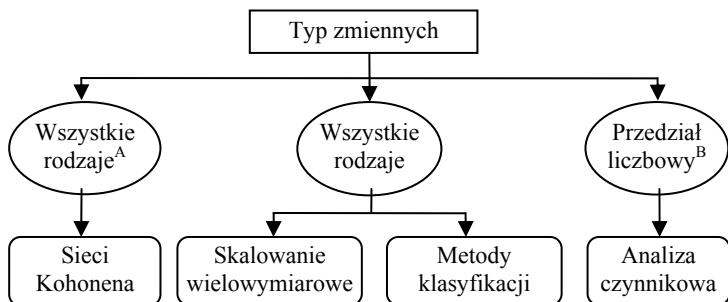
Na wybór metody SDA wpływa rodzaj zmiennych symbolicznych opisujących zbiór obiektów. Niektóre z metod nie uwzględniają wszystkich typów zmiennych symbolicznych (np. zmiennych strukturalnych) lub ich zastosowanie łączy się z koniecznością przekodowania zmiennych symbolicznych (większość metod badania zależności).



Oznaczenia: A – oprócz zmiennych strukturalnych, B – konieczne przekształcenie zmiennych symbolicznych.

Rys. 1. Metody badania zależności w analizie danych symbolicznych

Źródło: opracowanie własne na podstawie [Bock, Diday 2000; Diday, Noirhomme-Fraiture 2008].



Oznaczenia: A – oprócz zmiennych strukturalnych, B – konieczne przekształcenie zmiennych symbolicznych.

Rys. 2. Metody badania współwystępowania w analizie danych symbolicznych

Źródło: opracowanie własne na podstawie prac [Bock, Diday 2000; Lechevallier 2001; Noirhomme-Fraiture 2004; Diday, Noirhomme-Fraiture 2008].

## 6. Charakterystyka wybranych metod analizy danych symbolicznych

Stosowanie podejścia *a priori* w segmentacji rynku na podstawie danych symbolicznych jest utrudnione, ponieważ niewiele metod z tej grupy zostało zaadaptowanych na grunt analizy danych symbolicznych. Szczególnie użyteczne w badaniach segmentacyjnych jest podejście *post hoc*. Wśród podstawowych metod analizy danych symbolicznych, które można zastosować w tym podejściu, należy wymienić skalowanie wielowymiarowe, metody klasyfikacji i drzewa klasyfikacyjne.

### 6.1. Skalowanie wielowymiarowe dla danych symbolicznych

Celem skalowania wielowymiarowego jest przedstawienie w zredukowanej przestrzeni relacji zachodzących między obiektami wielowymiarowymi. Punktem wyjściowym w skalowaniu wielowymiarowym dla danych symbolicznych jest wyznaczenie:

a) macierzy odległości – z wykorzystaniem miar odległości danych symbolicznych,

b) macierzy odległości minimalnych i maksymalnych, którą można uzyskać:

- dla obiektów opisanych zmiennymi symbolicznymi w postaci przedziałów liczbowych, stosując odpowiednie formuły (zob. [Lechevallier 2001, s. 57]),
- na podstawie opinii ekspertów, którzy oceniają podobieństwo obiektów. Ustalenie odległości może być dokonane np. przez przyjęcie kwartyła pierwszego ocen za odległość minimalną, a kwartyła trzeciego za odległość maksymalną.

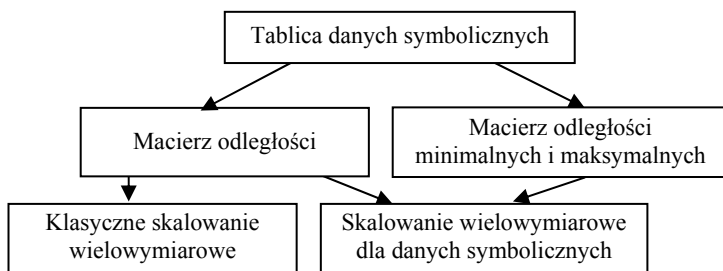
W ramach SDA zaproponowano dwie grupy algorytmów skalowania wielowymiarowego zróżnicowane ze względu na rodzaj macierzy wejściowej (tab. 3).

**Tabela 3.** Metody skalowania wielowymiarowego dla danych symbolicznych

Lp.	Algorytmy	Dane wejściowe	Typ zmiennych symbolicznych	Reprezentacja obiektów
1	<i>Bi-dimensional-mapping</i>	tablica danych symbolicznych => macierz odległości macierz odległości	dowolne	punkty w przestrzeni 2-wymiarowej ( $s = 2$ )
2	<i>InterScal, SymScal, I-Scal</i>	tablica danych symbolicznych => macierz odległości minimalnych i maksymalnych	przedziały liczbowe	hiperprostokąty, np. dla $s = 2$ – prostokąt,
		macierz odległości minimalnych i maksymalnych	dowolne	$s = 3$ – prostokąty

Źródło: opracowanie własne na podstawie prac [Noirhomme-Fraiture 2004; Lechevallier 2001].

W analizie danych symbolicznych jest ponadto możliwość stosowania procedur skalowania wielowymiarowego opracowanych dla danych klasycznych bazujących na macierzy odległości, tzn. klasycznego skalowania wielowymiarowego (rys. 3).



Rys. 3. Podejścia w skalowaniu wielowymiarowym danych symbolicznych

Źródło: opracowanie własne.

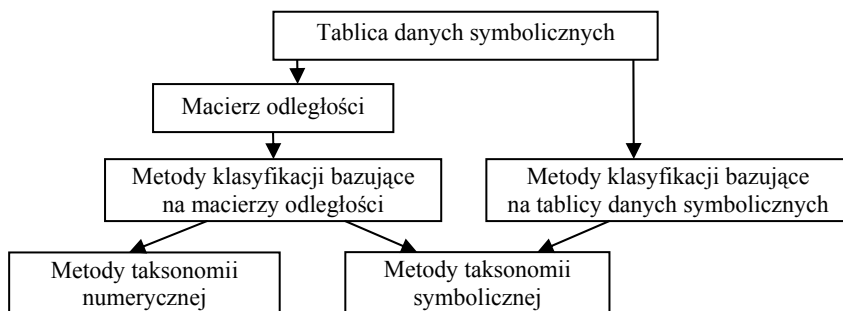
## 6.2. Metody analizy skupień dla danych symbolicznych

Celem analizy skupień jest zbadanie podobieństwa obiektów i na tej podstawie podział obiektów na względnie jednorodne klasy. Gatnar [1998], ze względu na charakter danych empirycznych, podzielił metody klasyfikacji na metody taksonomii numerycznej (adresowane do danych klasycznych) oraz metody taksonomii symbolicznej (powstałe w ramach SDA).

Metody taksonomii symbolicznej są najliczniejszą grupą metod SDA. Wyróżnia się ([Bock, Diday 2000; Diday, Noirhomme-Fraiture 2008; Wilk 2009a]):

- a) metody hierarchiczne, w tym:
  - aglomeracyjne, np. klasyfikacja Gowdy i Didaya, metoda hierarchiczna Brito,
  - deglomeracyjne, np. klasyfikacja podziałowa oparta na kryteriach;
- b) metody niehierarchiczne (optymalizacyjno-iteracyjne), bazujące na:
  - tablicy danych symbolicznych, np. SCLUST,  $k$ -średnich, Verde,
  - macierzy odległości, np. DCLUST.

W analizie danych symbolicznych jest ponadto możliwość stosowania metod taksonomii numerycznej bazujących na macierzy odległości, np. metod hierarchicznych, metody  $k$ -medoidów (rys. 4).



Rys. 4. Podejścia w klasyfikacji danych symbolicznych

Źródło: opracowanie własne.

### 6.3. Drzewa klasyfikacyjne dla danych symbolicznych

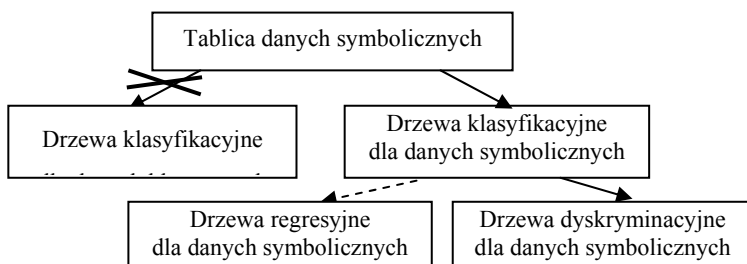
Drzewo klasyfikacyjne reprezentuje podział zbioru obiektów na jednorodne podzbiory na podstawie zmiennych niezależnych, biorąc pod uwagę w kolejnych krokach podziału tę zmienną, która wykazuje największy wpływ na zmienną zależną. W ramach SDA zaproponowano dotąd tylko algorytmy z grupy drzew dyskryminacyjnych (tab. 4).

Tabela 4. Drzewa klasyfikacyjne danych symbolicznych

Algorytmy	Sposób podziału	Rodzaj klasyfikacji	Typ symbolicznej zmiennej zależnej	Typ symbolicznych zmiennych niezależnych
<i>TREE</i>	binarny	rozłączna lub nierozłączna	nominalna lub lista wariantów	przedział liczbowy, lista wariantów, lista wariantów z wagami
<i>SDT</i>	binarny lub niebinarny	rozłączna	nominalna	lista wariantów z wagami
<i>SBTREE</i>	binarny	rozłączna	nominalna dwustanowa	przedział liczbowy

Źródło: opracowanie własne na podstawie prac [Bock, Diday 2000; Noirhomme-Fraiture 2004].

W analizie danych symbolicznych nie ma możliwości stosowania drzew klasyfikacyjnych opracowanych dla danych klasycznych (rys. 5).



Rys. 5. Podejścia w analizie danych symbolicznych z wykorzystaniem drzew klasyfikacyjnych

Źródło: opracowanie własne.

## 7. Podsumowanie

Metody SDA są stosowane przede wszystkim na gruncie chemii, medycyny i socjologii. W zakresie ekonomii znalazły zastosowanie głównie w badaniach regionalnych. Tematyka segmentacji rynku na podstawie danych symbolicznych jest



podejmowana w niewielu pracach (zob. [Bravo 2001; Wilk 2009c]). W artykule zaprezentowano podejścia i metody statystycznej analizy wielowymiarowej, jakie można zastosować w segmentacji rynku na podstawie danych symbolicznych.

## Literatura

- Bock H.H., Diday E. (red.), *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, Berlin 2000.
- Bravo M.C., *Análisis de segmentación en el análisis de datos simbólicos*, Universidad Complutense de Madrid (rozprawa doktorska), 2001.
- Diday E., Noirhomme-Fraiture M. (red.), *Symbolic Data Analysis and the Sodas Software*, John Wiley & Sons, Chichester 2008.
- Gatnar E., *Symboliczne metody klasyfikacji danych*, PWN, Warszawa 1998.
- Lechevallier Y. (red.), *Scientific Report for Unsupervised Classification, Validation And Cluster Representation*, ASSO Scientific Report for WP6, 2001.
- Noirhomme-Fraiture M. (red.), *User Manual for SODAS 2 Software*, Software Report, ASSO Scientific Report for WP3, 2004.
- Walesiak M., *Segmentacja rynku. Kryteria i metody*, [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, A. Zeliaś (red.), AE, Kraków 2000.
- Wedel M., Kamakura W.A., *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publisher, 1998.
- Wilk J., *Miary odległości obiektów opisanych zmiennymi symbolicznymi z wagami*, [w:] *Klasyfikacja i analiza danych – teoria i zastosowania*, Taksonomia 13, K. Jajuga, M. Walesiak (red.), Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr , AE, Wrocław 2005.
- Wilk J., *Problemy klasyfikacji obiektów symbolicznych. Symboliczne miary odległości*, [w:] *Ilościowe i jakościowe metody badania rynku. Pomiar i jego skuteczność*, J. Garczarczyk (red.), ZN AE w Poznaniu nr 71, Poznań 2006.
- Wilk J., *Interpretacja wyników klasyfikacji obiektów symbolicznych*, [w:] *Zastosowania metod ilościowych*, Ekonometria 18, J. Dziechciarz (red.), Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1151, AE, Wrocław 2007.
- Wilk J., *Cluster Analysis Methods in Symbolic Data Analysis*, III Ogólnopolska Konferencja Naukowa im. Prof. Aleksandra Zeliasia nt. „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych”, Zakopane, 5-8 maja 2009 r. (w druku), 2009a.
- Wilk J., *Przegląd metod wielowymiarowej analizy statystycznej wykorzystywanych w badaniach segmentacyjnych*, [w:] *Zastosowania metod ilościowych*, Ekonometria 23, J. Dziechciarz (red.), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu (w druku), 2009b.
- Wilk J., *Segmentacja internautów z wykorzystaniem danych symbolicznych i metod klasyfikacji*, [w:] *Klasyfikacja i analiza danych – teoria i zastosowania*, Taksonomia 16, K. Jajuga, M. Walesiak (red.), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 47, UE, Wrocław 2009c.
- Wilk J., *Problemy segmentacji rynku z wykorzystaniem metod klasyfikacji i danych symbolicznych*, rozprawa doktorska, 2009d.

## **MARKET SEGMENTATION METHODS ON THE BASIS OF SYMBOLIC DATA**

**Summary:** The application of market segmentation strategy requires collecting precise information about customers. Symbolic data enable to characterize customers and market segments in an explanatory way. However, symbolic data complexity prevents direct application of methods intended for classical data situation (real numbers or categories).

Segmentation methodology for classical data analysis has been widely discussed in the literature. Whereas the application of symbolic data analysis in market segmentation research is taken in few articles. The aim of the paper is to suggest approaches and introduce methods suitable for symbolic data analysis in market segmentation research.