

MAREK MELANIUK

Uniwersytet Łódzki

## TABELE PRZESTAWNE NARZĘDZIEM ANALIZY PROCESÓW SPOŁECZNYCH

### Wprowadzenie

W artykule pokazano możliwości wykorzystania tabel przestawnych w procesie odkrywania (pozyskiwania, wydobywania) wiedzy z baz danych społecznych. Analizy wykonano głównie za pomocą pakietu MsExcel w wersji 2007. Jednocześnie zwrócono uwagę na możliwości zastosowania narzędzia Pilot Danych (DataPilot) wchodzącego w skład arkusza kalkulacyjnego Calc z darmowego pakietu OpenOffice.org 3.0.1. Analizowaną bazą danych społecznych zgromadzonych w zasobach systemów administracyjnych jest przykład bazy danych o biorcach otrzymujących dodatki mieszkaniowe w Miejskim Ośrodku Pomocy Społecznej w Zgierzu.

Problematyka eksploracji danych (*Data Mining*) doczekała się już pokaźnej literatury w postaci opracowań zagranicznych i krajowych, na przykład w pracach Akademii Ekonomicznej we Wrocławiu [6; 7]. Moduły eksploracji danych zawarte są również w niektórych pakietach statystycznych (np. Statistica, SPSS). Algorytmy analitycznej eksploracji danych zintegrowane z technikami wizualizacyjnymi dostępne są w tworzonych w tym celu dedykowanych pakietach, na przykład GhostMiner Developer firmy Fujitsu [5]. Są to specjalizowane systemy z ogromnymi możliwościami eksploracyjnymi.

W praktyce do wykonania analiz biznesowych w firmach produkcyjnych, handlowych i usługowych (zwłaszcza sektora MSP), a także w praktyce administracyjnej urzędów, w zupełności wystarczająca jest mniej zaawansowana technologia do uzyskania przynajmniej podstawowych zależności między danymi i wyciągnięcia wniosków niezbędnych do wspomagania podejmowania decyzji.

Z obserwacji i ścisłej współpracy autora z jednostkami administracyjnymi na polu informatycznym wynika, że urzędy dysponują już obecnie istotnymi danymi źródłowymi, które są jednak w minimalnym stopniu wykorzystywane w codziennej pracy decyzyjnej wójta, burmistrza lub prezydenta miasta. Wynika to z następujących powodów:

- a) system informatyczny dostarczony przez firmy zewnętrzne służy głównie do generowania informacji sprawozdawczych dla jednostek zwierzchnich, a jedynie w niewielkim stopniu do bieżącego zarządzania jednostką administracyjną;
- b) użytkownicy z reguły nie potrafią (nie dysponując specjalistyczną wiedzą informatyczną) wyciągnąć wymaganych danych z systemów bazodanowych, a następnie przetworzyć je w sposób analityczny za pomocą posiadanych pakietów biurowych.

Niejednokrotnie się zdarza, że analizy społeczno-ekonomiczne wykonywane są ręcznie na podstawie źródłowych dokumentów papierowych, mimo że dane takie znajdują się w systemach informatycznych.

Obecnie urzędy administracyjne mają już zainstalowane pakiety biurowe w wersji bezpłatnej. Są nimi z reguły systemy OpenOffice.org. Dość często urzędy użytkują przynajmniej jedną komercyjną wersję pakietu Microsoft Office 2003 lub wersję 2007. W arkuszu kalkulacyjnym Excel 2003/2007 wbudowane jest narzędzie o znacznych możliwościach raportujących w postaci tabel przestawnych. Narzędzie tabel przestawnych znajduje się też w pakiecie OpenOffice.org w opcji Pilot Danych (DataPilot).

Badania dowodzą, że bardzo niewielu użytkowników potrafi korzystać z tabel przestawnych, nie wspominając już o możliwości wykorzystania języka VBA do automatyzacji działań oraz stosowania modułów OLAP. Nawet podczas specjalistycznych kursów z Excela lub programu Calc tabele przestawne są traktowane dość pobieżnie, bez pokazywania użytkownikowi, w jaki sposób w ciągu kilkadziesiątu (a nawet kilkunastu) sekund i za pomocą jedynie kilku kliknięć myszką mogą uzyskać dane analityczne na podstawie wielu tysięcy rekordów danych.

Oczywiście, tabele przestawne nie są uniwersalnym narzędziem, spełniającym wszystkie potrzeby analityczne i nie mają możliwości tak dogłębnego odkrywania wiedzy, jak systemy klasy *Data Mining*. Jednak do wykonania większości – nawet zaawansowanych – rozwiązań analitycznych w administracji publicznej

tabele przestawne zawarte w arkuszach kalkulacyjnych są narzędziem w pełni zadowalającym.

## 1. Tabele przestawne

Autorem koncepcji arkusza kalkulacyjnego jest Pito Salas, który przedstawił ją w 1986 roku, pracując na stanowisku *software developer* w Lotus Advanced Technology Group. Nowy produkt został rozwinięty przez firmę Lotus w 1989 roku pod nazwą Lotus Improv. Pomysł ten był stopniowo rozwijany pod nazwą tabele przestawne (*pivot tables*) przez firmę Microsoft, począwszy od produktu Excel 5. Tabele przestawne w ostatniej, aktualnej wersji Excel 2007 mają kilka nowości w stosunku do poprzedniej wersji Excel 2003: wzrosła pojemność arkusza z 16 777 216 komórek do ponad 17 mld (16 384 kolumn razy 1 048 576 wierszy) oraz – co ma znaczne walory analityczne – obecnie jest możliwość zaznaczania dowolnej liczby pól filtrujących w obszarze strony. Jednak nie wszystkim użytkownikom pakietu Office przypadł do gustu pomysł interfejsu wstążki (*Ribbon*) w wersji 2007 w stosunku do interfejsu menu w wersjach wcześniejszych. W pracy [1] autor dość krytycznie podchodzi do zmian dokonanych w wersji 2007 oraz wspomina o swojej początkowej irytacji i swoich kolegów na zmianę interfejsu z menu na wstążki i o spadku wydajności pracy o 25% z powodu długiego szukania opcji w wersji 2007, które były bezbłędnie lokalizowane w starszej wersji Excela.

Odpowiednikiem tabeli przestawnej i wykresu przestawnego Excela jest opcja Pilot danych arkusza Calc w pakiecie OpenOffice.org. Istotną różnicą rozszerzającą możliwości arkusza Calc w stosunku do arkusza Excel 2003 jest zwiększenie dostępnej liczby kolumn z 256 (Excel) do 1024 w arkuszu Calc. Dzięki temu kolumny są adresowane od litery A do liter AMJ (w Excelu 2003 – od A do IV). Przy okazji należy zaznaczyć, że pakiet OpenOffice.org jest przykładem otwartego oprogramowania biurowego (*free software*) i w każdej chwili może być pobrana jego bezpłatna aktualizacja.

Tabele przestawne opierają się na tabelarycznym przedstawieniu danych składających się z rekordów (zapisów), którymi są wiersze tabeli, oraz pól informacyjnych, które są reprezentowane przez kolumny tabeli. Każde pole informacyjne powinno być wcześniej nazwane. Nazwy pól powinny być tworzone na zasadzie mnemonicznej, czyli takiej, że nazwa (skrótowa) pola świadczy o jego zawartości.

## 2. Baza danych społecznych

Baza danych społecznych rozpatrywanych w niniejszym opracowaniu obejmuje 791 rodzin, którym przyznano dodatki mieszkaniowe w grudniu 2008 roku przez Miejski Ośrodek Pomocy Społecznej w Zgierzu.

Ze względu na ochronę danych osobowych w analizie nie uwzględniono danych wrażliwych (nazwiska, imienia, PESEL-u, adresu zamieszkania), a pod uwagę wzięto następujące pola informacyjne:

- ULICA – nazwa ulicy, którą zamieszkuje rodzina otrzymująca dodatek,
- DODATEK – kwota otrzymanego dodatku (w grudniu 2008 roku),
- RYCZAŁT – kwota dodatku otrzymanego w formie ryczałtu z powodu braku w lokalu co najmniej jednego z mediów: centralnego ogrzewania, ciepłej wody, instalacji gazu przewodowego,
- PRZELEW – kwota dodatku przekazywanego zarządcy nieruchomości,
- POW\_UŻ – powierzchnia użytkowa mieszkania (m<sup>2</sup>),
- CO – wyposażenie lokalu w instalację centralnego ogrzewania (T = tak, N = nie),
- CW – wyposażenie lokalu w instalację ciepłej wody (T = tak, N = nie),
- GAZ – wyposażenie lokalu w instalację gazu przewodowego (T = tak, N = nie),
- OSOBY – liczba osób w rodzinie,
- DOCHODY – dochody gospodarstwa domowego za ostatnie 3 miesiące,
- WYDATKI – faktycznie poniesione wydatki z tytułu użytkowania lokalu.

W tabeli 1 pokazano fragment bazy danych. Na przykład: w pierwszym rekordzie znajdują się dane o osobie samotnej (OSOBY = 1) mieszkającej na ul. 3 Maja, której lokal nie ma żadnych mediów, natomiast w szóstym rekordzie (ul. Tuwima) są dane o rodzinie 3-osobowej, której lokal ma wszystkie media (centralne ogrzewanie, ciepłą wodę, doprowadzony gaz) i dlatego rodzina ta nie otrzymuje ryczałtu na przykład na podgrzanie wody, a cały dodatek w kwocie 168 zł 32 gr jest przekazywany przelewem na konto zarządcy. Dzięki temu, rodzina opłaca czynsz za użytkowanie lokalu pomniejszony o kwotę dodatku.

Tabela 1

Fragment bazy danych zawierający pierwszych dziesięć rekordów, którymi są rodziny otrzymujące dodatki mieszkaniowe

ULICA	DODATEK	RYCZAŁT	PRZELEW	POW_UZ	CO	CW	GAZ	OSOBY	DOCHODY	WYDATKI
3 Maja	42,07	15,54	26,53	14,43	N	N	N	1	478,27	71,78
Wiosny Ludów	91,43	55,35	36,08	39,00	N	N	N	3	263,86	72,17
Rembowskiiego	72,94	29,30	43,64	29,49	N	N	N	2	888,00	107,40
Plater	137,74	1,86	135,88	36,41	T	T	N	1	49,26	282,71
Długa	51,82	24,19	27,63	18,80	N	N	N	1	49,66	55,25
Tuwima	168,32	0,00	168,32	37,94	T	T	T	3	1527,42	351,61
Narutowicza	62,15	32,22	29,93	25,32	N	N	N	1	16,15	59,85
Kolejowa	183,27	3,81	179,46	45,50	T	N	T	1	950,93	358,92
Długa	93,56	41,26	52,30	31,75	N	N	N	2	258,27	104,60
Rembowskiiego	80,21	2,67	77,54	47,79	T	N	T	2	1831,55	459,56

Źródło: opracowanie własne.

### 3. Analiza zależności finansowych i społecznych

Zastosowanie tabel przestawnych jako narzędzia analitycznego umożliwia bardzo szybkie wykonanie odpowiednich obliczeń statystycznych i matematycznych według zadanych kryteriów. W tabeli 2 przedstawiono obliczone przeciętne dodatki i liczby rodzin w zależności od powierzchni użytkowej mieszkań. Należy zwrócić uwagę na możliwość grupowania dużej liczby danych na mniejsze kategorie (szeregi rozdzielcze), dzięki czemu bardziej przejrzysta jest ich analiza. W arkuszu Calc grupowanie znajduje się w opcji Dane – Konspekt – Grupuj, natomiast w Excelu 2003 i 2007 grupowanie wybiera się z opcji menu po pokazaniu danego pola. Tabelę wykonano za pomocą tabel przestawnych w Excelu 2007. Pole IDE jest identyfikatorem rekordu (numerem identyfikacyjnym rodziny).

Tabela 2

Przeciętne wartości przyznanych dodatków mieszkaniowych  
i liczba rodzin otrzymujących dodatki (Licznik z IDE)  
w zależności od pogrupowanej powierzchni użytkowej (POW\_UŻ) mieszkań

Dane		
POW_UŻ	Średnia z DODATEK	Licznik z IDE
10–20	44,36	37
20–30	72,97	111
30–40	113,47	329
40–50	146,12	241
50–60	168,66	37
60–70	197,45	32
70–80	204,55	4
Suma końcowa	120,94	791

Źródło: opracowanie własne.

Tabela 3

Tabela wykonana  
za pomocą tabel przestawnych  
w Excelu 2003

POW_UŻ	Dane	Suma
10–20	Średnia z DODATEK	44,36
	Licznik z IDE	37
20–30	Średnia z DODATEK	72,97
	Licznik z IDE	111
30–40	Średnia z DODATEK	113,47
	Licznik z IDE	329
40–50	Średnia z DODATEK	146,12
	Licznik z IDE	241
50–60	Średnia z DODATEK	168,66
	Licznik z IDE	37
60–70	Średnia z DODATEK	197,45
	Licznik z IDE	32
70–80	Średnia z DODATEK	204,55
	Licznik z IDE	4
Średnia z DODATEK, Razem		120,94
Licznik z IDE, Razem		791

Źródło: opracowanie własne.

Tabela 4

Tabela wykonana za pomocą Pilota  
danych w arkuszu Calc  
pakietu OpenOffice.org

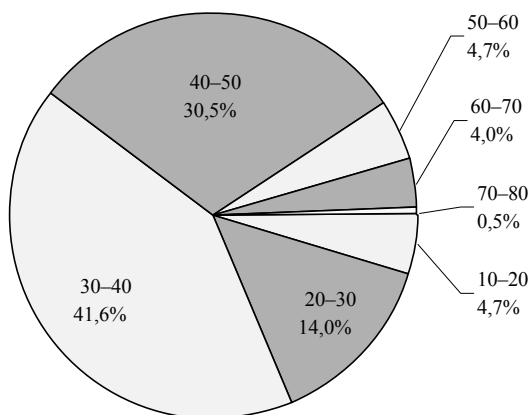
POW_UŻ	Dane	
10–20	Średnia – DODATEK	44,36
	Liczba – IDE	37
20–30	Średnia – DODATEK	72,97
	Liczba – IDE	111
30–40	Średnia – DODATEK	113,47
	Liczba – IDE	329
40–50	Średnia – DODATEK	146,12
	Liczba – IDE	241
50–60	Średnia – DODATEK	168,66
	Liczba – IDE	37
60–70	Średnia – DODATEK	197,45
	Liczba – IDE	32
70–80	Średnia – DODATEK	204,55
	Liczba – IDE	4
<b>Suma Średnia – DODATEK</b>		<b>120,94</b>
<b>Suma Liczba – IDE</b>		<b>791</b>

Źródło: opracowanie własne.

W tabelach 2–4 przedstawiono te same informacje, jednak wykonane są przy użyciu różnych narzędzi. Jak widać z tabeli 2, w programie Excel 2007 wyniki z tabel przestawnych są ułożone w sposób kolumnowy, dzięki czemu ich ogólny wygląd jest bardziej komunikatywny i znacznie łatwiejsza jest analiza danych.

Podobny do siebie wygląd mają wyniki z programów Excel 2003 i Calc (tabele 3 i 4), natomiast w przypadku większej liczby informacji umieszczenie ich w wierszach powoduje, że znacznie się zmniejsza czytelność wyników.

Istotnym elementem analizy rzeczywistości są analizy strukturalne. Wykres kołowy pokazany na rysunku 1 otrzymano na podstawie danych liczbowych znajdujących się w tabeli 2. Pokazuje on strukturę (udział procentowy) rodzin otrzymujących dodatki w zależności od powierzchni użytkowej mieszkania. Z wykresu kołowego wyraźnie wynika dominujący udział rodzin (41,6%) zamieszkujących lokale o powierzchni 30–40 m<sup>2</sup>. Udziały procentowe program wylicza automatycznie na wykresie kołowym, bez konieczności ręcznego wpisywania odpowiednich formuł. Tego typu analizy strukturalne są ważnym elementem analizy danych społecznych.



Rys. 1. Struktura rodzin otrzymujących dodatki w zależności od pogrupowanej powierzchni użytkowej mieszkań w m<sup>2</sup> (na podstawie danych z tabeli 4)

Źródło: opracowanie własne.

W badaniach społecznych ważną odgrywają również analizy strukturalne, w których kryterium analizy jest wielkość rodziny, to znaczy liczby osób w rodzinie. Zależy nam na obliczeniu liczby rodzin i liczby znajdujących się w nich

osób oraz na pokazaniu wartości odpowiednich zmiennych charakteryzujących te rodziny.

Raporty z wykonanych tabel przestawnych zawarto w tabelach 5 i 6. Wartości przeciętnych dodatków mieszkaniowych, dochodów rodzin i wydatków na cele mieszkaniowe to istotne czynniki kształtujące politykę społeczną. Z tabeli 5 między innymi wynika, że już nie wzrastają przeciętne dochody (kolumna 5) oraz wydatki na cele mieszkaniowe (kolumna 6) rodzin o liczbie osób większej niż 4 lub 5. Tego typu wyniki powinny być przedmiotem zainteresowania odpowiednich służb społecznych.

Tabela 5

Liczby rodzin, osób w rodzinach,  
przeciętne wartości udzielonych dodatków mieszkaniowych  
oraz przeciętne dochody rodzin i wydatki na cele mieszkaniowe  
w zależności od liczby osób w rodzinie

Liczba osób w rodzinie	Liczby rodzin	Liczby osób w rodzinach	Średnie dodatki	Średnie dochody	Średnie wydatki
1	2	3	4	5	6
1	330	330	85,06	600,84	186,95
2	188	376	123,15	977,98	267,86
3	136	408	151,64	1409,52	329,49
4	91	364	168,36	1677,99	362,48
5	32	160	187,20	2079,82	355,32
6	13	78	174,89	1982,39	291,74
8	1	8	232,76	1333,69	291,05
Ogółem	791	1724	120,94	1036,90	259,54

Źródło: opracowanie własne.

Analiza wyposażenia lokali mieszkalnych w media (centralne ogrzewanie, ciepła woda, doprowadzony gaz) w zależności od liczebności rodzin jest istotnym czynnikiem w procesie podejmowania decyzji w zakresie mieszkalnictwa na terenie miasta lub gminy (tabela 6).

Wykorzystując tabele przestawne, szybko i łatwo możemy obliczyć liczbę elementów w zbiorze o zadanej właściwości; w przypadku centralnego ogrzewania w tabeli 6 dotyczy to kolumn 2 i 3. Z opcji pół tabeli przestawnej możemy wybrać różne kombinacje wyników procentowych (z wierszy, kolumn, całości lub różnicy); przedstawiają to kolumny 5, 9, 13 w tabeli 6.

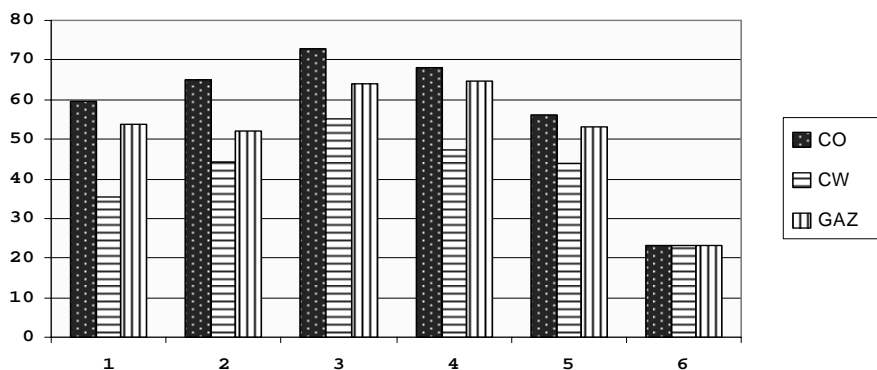


Tabela 6

Liczby rodzin zamieszkujących w lokalach wyposażonych (T) lub niewyposażonych (N) w centralne ogrzewanie (CO), ciepłą wodę (CW) i instalację gazu przewodowego (GAZ) oraz procent rodzin mających te instalacje w zależności od liczby osób w rodzinie

Liczba osób w rodzinie	CO				CW				GAZ			
	N	T	suma	%	N	T	suma	%	N	T	suma	%
1	2	3	4	5	6	7	8	9	10	11	12	13
1	133	197	330	59,7	213	117	330	35,5	152	178	330	53,9
2	66	122	188	64,9	105	83	188	44,1	90	98	188	52,1
3	37	99	136	72,8	61	75	136	55,1	49	87	136	64,0
4	29	62	91	68,1	48	43	91	47,3	32	59	91	64,8
5	14	18	32	56,3	18	14	32	43,8	15	17	32	53,1
6	10	3	13	23,1	10	3	13	23,1	10	3	13	23,1
8	1		1	0,0	1		1	0,0	1		1	0,0
Ogółem	290	501	791	63,3	456	335	791	42,4	349	442	791	55,9

Źródło: opracowanie własne.



Rys. 2. Procent lokali (zamieszkałych przez rodziny otrzymujące dodatki mieszkaniowe) wyposażonych w instalację centralnego ogrzewania (CO), ciepłej wody (CW) i gazu przewodowego (GAZ)

Źródło: opracowanie własne.

W celu przedstawienia istotnych wyników (znajdujących się w raportach tabel przestawnych) w sposób bardziej przejrzysty można zastosować graficzną prezentację danych. Na rysunku 2 przedstawiono dane procentowe z tabeli 6

(kolumny 5, 9, 13). Widać wyraźnie coraz mniejsze wyposażenie lokali mieszkalnych w media, zwłaszcza w rodzinach wieloosobowych, czyli wielodzietnych. Stąd wnioszek, że wyniki jednych analiz świadczą o konieczności wykonania dalszych badań, pokazujących w tym przypadku trudną sytuację mieszkaniową i materialną dzieci.

#### 4. Względne ważności danych

Obliczmy liczby rodzin w zależności od wyposażenia lokali mieszkalnych w media. Odpowiednie wyniki z raportu tabeli przestawnej pokazano w tabeli 7. Wyraźnie z niej wynika, że najliczniejszą grupę, wynoszącą 284 lokale mieszkalne, tworzą te, które nie mają żadnych (!) mediów. Jest to aż 36% wszystkich rodzin korzystających z dodatków mieszkaniowych.

Tabela 7

Liczby rodzin otrzymujących dodatki mieszkaniowe, zamieszkałych w lokalach wyposażonych (T) lub niewyposażonych (N) w centralne ogrzewanie (CO), ciepłą wodę (CW) i instalację gazu przewodowego (GAZ)

GAZ		CW		Suma ogółem
		N	T	
N	N	284	3	287
	T	7	55	62
Suma		291	58	349
T	N	3	0	3
	T	162	277	439
Suma		165	277	442
Suma ogółem		456	335	791

Źródło: opracowanie własne.

W tabeli przestawnej na liście opcji dla pól danych znajduje się nietypowa opcja Indeks. Obliczenie wartości indeksu podaje względną wagę danej komórki w ramach kolumny. Sposób obliczenia wagi jest następujący:

$$\text{Wartość indeksu} = \frac{(\text{liczba w komórce}) \cdot (\text{suma końcowa})}{(\text{suma ogółem w wierszu}) \cdot (\text{suma ogółem w kolumnie})}$$

W tabeli 8 indeks 1,72 jest wynikiem wyrażenia:  $(284 \cdot 791)/(456 \cdot 287)$ . W wielu przypadkach, zwłaszcza dla problemów gospodarczych, obliczanie tego

indeksu ma sens, jednak czasami wartości indeksów są trudne do wytłumaczenia. Dotyczy to na przykład indeksów 1,73 i 0,64 dla lokali bez ciepłej wody.

Należy zauważyć, że gdyby w tabeli 7 zamiast liczby lokali zostały obliczone na przykład przeciętne kwoty wypłaconych dodatków, to wynikające stąd indeksy wag przedstawione w tabeli 8 miałyby całkiem inne wartości.

Tabela 8

Względne ważności (indeksy) przeciętnych wartości dodatków mieszkaniowych dla rodzin zamieszkających w lokalach wyposażonych (T) lub niewyposażonych (N) w centralne ogrzewanie (CO), ciepłą wodę (CW) i instalację gazu przewodowego (GAZ)

GAZ	CO	CW		Indeks końcowy
		N	T	
N	N	1,72	0,02	1,00
	T	0,20	2,09	1,00
Indeks pośredni		1,45	0,39	1,00
T	N	1,73	–	1,00
	T	0,64	1,49	1,00
Indeks pośredni		0,65	1,48	1,00
Indeks końcowy		1,00	1,00	1,00

Źródło: opracowanie własne.

## 5. Analizy przestrzenne

Dla celów zarządzania miastem lub regionem i podejmowania efektywnych decyzji alokacyjnych niezbędne jest wykonanie odpowiednich analiz przestrzennych. Przykład takiej analizy pokazano w tabeli 9. W tabeli tej ulice posortowano według liczby osób w rodzinach korzystających z dodatków mieszkaniowych. Pełniej odzwierciedla to skalę zjawiska niż układ według liczby rodzin. Jaką wiedzę można osiągnąć z takiej tablicy liczbowej i kto powinien wyciągnąć odpowiednie wnioski?

1. Niezerowe kwoty ryczałtów (kolumna 8) wskazują na brak połączeń budynków do sieci centralnego ogrzewania, ciepłej wody lub gazu.

2. Łatwo można obliczyć przeciętną wielkość rodziny według poszczególnych ulic lub wyszukać ulice o znacznej liczbie osób samotnych korzystających z dodatków mieszkaniowych (w Zgierzu 41% rodzin otrzymujących dodatek to osoby samotne i ten odsetek dość szybko rośnie w kolejnych latach).

3. Największą korzyścią z pewnością jest określenie obszarów o największym natężeniu danego zjawiska, czyli w tym przypadku pokazanie enklaw biedy z punktu widzenia dodatków mieszkaniowych.

4. Podobne układy przestrzenne powinny być wykonane dla innych form świadczeń społecznych (pomoc społeczna, świadczenia rodzinne, fundusz alimentacyjny) i dzięki temu można otrzymać pełne mapy biedy w danej miejscowości.

5. Na tak skonstruowane enklawy biedy można nałożyć mapy korzystających z opieki zdrowotnej, mapy jakości stanu budynków (według rodzajów zarządców) itp., i dzięki temu otrzymać analizy korelacji przestrzennych. Oczywiście, w tym przypadku pomocne byłyby numeryczne mapy terenu wykorzystywane w systemach informacji przestrzennej (GIS).

6. Zasygnalizowane materiały analityczne (o różnym stopniu szczegółowości) powinny być jednymi z podstawowych narzędzi zarządzania regionem przez prezydenta miasta, burmistrza, wójta a także cennymi wskazówkami dla radnych na szczeblu samorządowym.

Tabela 9

Układ przestrzenny (według ulic w Zgierzu) przyznanych dodatków mieszkaniowych w ujęciu ilościowym i wartościowym

Nazwy ulic	Liczby rodzin	Liczby osób w rodzinach	Liczby mieszkańców	Procent otrzymujących	Kwoty dodatków	Kwoty przelewów	Kwoty ryczałtów
1	2	3	4	5	6	7	8
Parzęczewska	80	157	4546	3,45	9753,96	9509,23	244,73
Długa	57	132	2509	5,26	6878,92	5325,70	1553,22
Rembowskiego	64	127	2240	5,67	8299,74	7413,75	885,99
Dubois	57	120	2485	4,83	7476,59	7079,93	396,66
Staffa	39	110	4310	2,55	6048,08	6048,08	0,00
Gałczyńskiego	40	100	1562	6,40	5633,93	5633,93	0,00
Narutowicza	36	77	871	8,84	3436,25	2105,38	1330,87
Dąbrowskiego	23	62	682	9,09	2717,70	1895,21	822,49
3 Maja	30	61	1033	5,91	3389,70	2754,35	635,35
Witkacego	31	60	2175	2,76	3780,98	3780,98	0,00
1 Maja	25	59	2123	2,78	2755,54	2324,58	430,96
Boya Żeleńskiego	23	56	2476	2,26	3636,98	3636,98	0,00

1	2	3	4	5	6	7	8
Kolejowa	22	49	611	8,02	3369,19	3210,97	158,22
Baczyńskiego	22	47	786	5,98	2840,90	2840,90	0,00
Pl. Kilińskiego	16	43	278	15,47	2286,11	1575,67	710,44
Słowackiego	15	36	252	14,29	1434,34	880,76	553,58
Śniechowskiego	14	31	499	6,21	1634,33	1342,33	292,00
Mielczarskiego	14	30	338	8,88	1198,17	843,99	354,18
Łęczycka	16	27	741	3,64	1005,07	442,87	562,20
Pozostałe ulice	142	290	25740	1,13	14673,13	11758,07	2915,06
Suma	791	1724	58351	2,95	95662,63	83579,28	12083,35

Ciemniejszym tłem zaznaczono ulice, na których mieszka co najmniej 5% osób otrzymujących dodatki mieszkaniowe w stosunku do ogólnej liczby osób zamieszkujących daną ulicę. W tabeli podano informacje o pierwszych 19 ulicach posortowanych według największej liczby rodzin otrzymujących dodatki mieszkaniowe.

Źródło: opracowanie własne.

## Podsumowanie

Problem eksploracji danych polega na efektywnym znajdowaniu dotychczas nieznanymi zależności i związków między danymi. W tym celu stosuje się zaawansowany aparat matematyczno-statystyczny wykorzystywany w metodach klastrowania, odkrywania asocjacji, wzorców sekwencji lub klasyfikacji. W systemach administracji publicznej opartych obecnie na coraz większych komputerowych bazach danych większość podejmowanych decyzji jest ściśle uzależniona od istniejących ustaw sejmowych, rozporządzeń ministerialnych oraz wynika z zasad kodeksu postępowania administracyjnego. Decyzje administracyjne mogą i powinny być wspomagane bardziej aktywnymi narzędziami dostępnymi w użytkowanych pakietach biurowych. Obecnie wymaga się, aby każdy pracownik biurowy potrafił wykorzystać nie tylko edytor tekstu, ale również arkusz kalkulacyjny, i to w taki sposób, by na podstawie istniejących faktów móc odkryć istniejące zależności i prawidłowości między danymi (praktyka, niestety, w znacznym stopniu rozmija się ze stawianymi wymaganiami).

Reasumując, umiejętność wykorzystania informacji (czyli przetworzonych danych) do podjęcia racjonalnych decyzji jest nadal sztuką, natomiast barierami wykorzystania nie powinny być na obecnym etapie rozwoju technologii informacyjnej narzędzia informatyczne wspomagające pracę decydenta. Takim narzędziem decyzyjnym są właśnie tabele przestawne.

Sposób wykorzystania tabel przestawnych pod nazwą Pilot danych w arkuszu kalkulacyjnym Calc, będącym komponentem otwartego oprogramowania biurowego, jest w zasadzie tak samo przyjazny jak arkusz Excel 2003. Z punktu widzenia jednostek administracyjnych i wielu firm, zwłaszcza sektora MSP, brak konieczności ponoszenia opłat licencyjnych na pewno jest jednym z głównych argumentów za ich zastosowaniem w praktyce.

W celu łatwiejszego przyswojenia sobie wiedzy na temat arkuszy kalkulacyjnych oraz tabel przestawnych bardzo pomocna byłaby pozycja, której autorami są Bill Jelen i Michael Alexander [3]. Bill Jelen prowadzi bardzo ciekawy portal internetowy MrExcel.com, który wart jest odwiedzenia. Drugą, znakomitą pozycją zasługującą na uwagę jest praca Paula McFedriesa [4].

### Literatura

1. Carlberg C., *Microsoft Excel 2007 PL. Analizy biznesowe*, Helion, Warszawa 2009.
2. Fundacja Wolnego i Otwartego Oprogramowania, [www.fwioo.pl/](http://www.fwioo.pl/).
3. Jelen B., Alexander M., *Tabele przestawne*, Mikom 2006.
4. McFedries P., *Excel. Tabele i wykresy przestawne. Przewodnik po tworzeniu dynamicznych arkuszy kalkulacyjnych*, Helion, Warszawa 2007.
5. Pakiet Ghost Miner Developer, [www.fqs.pl/business\\_intelligence/ghostminer](http://www.fqs.pl/business_intelligence/ghostminer).
6. *Pozyskiwanie wiedzy i zarządzanie wiedzą*, red. M. Nycz, M. Owoc, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2006.
7. *Zarządzanie wiedzą w systemach informacyjnych*, red. W. Abramowicz, A. Nowicki, M. Owoc, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2004.

## KNOWLEDGE MINING FROM DATABASES BY MEANS OF OPEN SOURCE SOFTWARE

### Summary

The article describes the use of pivot tables as a tool of knowledge discovery in social databases. The calculations are made by means of pivot table in Excel 2007 and PilotData as a Calc spreadsheet tool in OpenOffice.org package. All of the calculations are made on the residence allowance database from Social Welfare Centre at Zgierz city. In particular there is shown a way of social and financial dependency analysis and structural and space analysis.

*Translated by Marek Melaniuk*