

Koncepcja szkoleniowej bazy danych „Dynamiczna populacja” – projekt realizowany przez koło naukowe

Andrzej Ptasznik*, Przemysław Jeżewski**, Arkadiusz Redzik***

Warszawska Wyższa Szkoła Informatyki

Streszczenie

W artykule opisano koncepcję realizowanej przez koło naukowe szkoleniowej bazy danych. Wskazano genezę projektu oraz jego główne założenia. Przedstawiona została realizacja podstawowych etapów projektu oraz wizja dalszego rozwoju systemu.

Słowa kluczowe – Business Intelligence, szkoleniowa baza danych, populacja.

1. Wprowadzenie

W ramach prac koła naukowego WWSI.Net szukano tematu i wyzwania, dającego szansę realizacji ciekawego projektu, którego efekt byłby trwały i nadający się do szerokiego wykorzystania w procesie dydaktycznym Uczelni. Istotnym celem było też zapewnienie ciągłości pracy koła naukowego. Kolejne grupy studentów powinny zapoznawać się z aktualnym stanem projektu i przystępować do realizacji kolejnych jego etapów. Wokół tych podstawowych założeń powstała grupa studentów, członków koła naukowego WWSI.Net, która w październiku 2012 roku przystąpiła do realizacji projektu „Dynamiczna populacja”.

* E-mail: aptaszni@wwsi.edu.pl

** E-mail: p.jezewskii@poczta.wwsi.edu.pl

*** E-mail: a_redzik@poczta.wwsi.edu.pl

Założono że powstanie baza danych, która będzie mogła być wykorzystywana jako pomoc naukowa do zajęć wymagających operowania na dużych zestawach danych, w ramach różnych przedmiotów specjalistycznych. Wiodącym motywem podjętych działań był brak szkoleniowych baz danych, nadających się do specjalistycznych ćwiczeń z zakresu analizy danych oraz także fakt, iż zadanie takie wydawało się bardzo ciekawym wyzwaniem badawczym.

2. Przygotowanie wyjściowego obrazu populacji

Projekt „Dynamiczna populacja” ma na celu odzwierciedlenie statycznego stanu populacji Polski, jak i modelowanie dynamicznych procesów w niej zachodzących. Projekt jest budowany w oparciu o technologię MS SQL Server.

W pierwszym kroku ustalono główne zadania stawiane w procesie realizacji projektu:

- utworzenie bazy danych będącej obraz populacji ludności w Polsce,
- opracowanie symulatora zmian populacji i wybranych zdarzeń zachodzących w Polsce,
- opracowanie scenariuszy wykorzystania systemu.

Realizacja pierwszego zadania, wymagała przygotowania odpowiednich statystyk, które miały posłużyć do wygenerowania wyjściowego obrazu populacji.

Do procesu generowania rozkładu populacji wykorzystano dwie statystyki.

Na rysunku 1 pokazano tabelę w której zamieszone zostały dane o liczbie osób zamieszkujących gminy z podziałem na płeć i lokalizację miasto-wieś. Dane zostały przygotowane na podstawie udostępnionego przez Główny Urząd Statystyczny zestawienia *Ludność w gminach według stanu w dniu 31.12.2011 r. bilans opracowany w oparciu o wyniki NSP'2011* [1].

Pokazane na rysunku 1 dane narzucały ilościową charakterystykę rozkładu populacji bez podziału na wiek osób. Konieczne było skorzystanie z kolejnej statystyki, pokazanej na rysunku 2, która zawierała procentowy rozkład populacji w powiatach według płci wieku i lokalizacji miasto-wieś.

KodTeryt	Rok	WiekOd	WiekDO	KobMiasto	Mmiasto	KobWies	Mwies	Id
0201011	2011	0	150	21134	18985	0	0	1
0201022	2011	0	150	0	0	6890	6824	2
0201032	2011	0	150	0	0	2749	2742	3
0201043	2011	0	150	2196	2052	5583	5456	4
0201052	2011	0	150	0	0	3742	3805	5
0201062	2011	0	150	0	0	4194	4176	6
0202011	2011	0	150	16879	14704	0	0	7
0202021	2011	0	150	18497	16341	0	0	8
0202031	2011	0	150	5031	4657	0	0	9
0202041	2011	0	150	3529	3247	0	0	10
0202052	2011	0	150	0	0	4735	4678	11
0202062	2011	0	150	0	0	3902	3694	12
0202073	2011	0	150	1609	1535	1369	1355	13
0203011	2011	0	150	35763	33496	0	0	14
0203022	2011	0	150	0	0	3195	3168	15
0203032	2011	0	150	0	0	2157	2162	16
0203042	2011	0	150	0	0	2177	2164	17
0203052	2011	0	150	0	0	1155	1204	18

Rysunek 1. Liczba osób w gminach wg płci i podziału miasto/wieś

KodTeryt	Wiek	PKobWies	PMMiasto	PMWies	PKobMiasto
0201	0	0.01175283953865	0.01025361600951	0.01226506661960	0.00835727343954
0201	1	0.01122659299215	0.01030315038636	0.01239742345363	0.00861843823452
0201	2	0.01122659299215	0.01020408163265	0.01230918556428	0.00844432837120
0201	3	0.01091961584002	0.01010501287893	0.01191211506220	0.00853138330286
0201	4	0.01109503135552	0.01015454725579	0.01116209300273	0.00783494384956
0201	5	0.01026180765688	0.01040221914008	0.01072090355598	0.00796552624706
0201	6	0.01026180765688	0.00995640974836	0.00913262154769	0.00944546008531
0201	7	0.00894619129061	0.01000594412522	0.01107385511338	0.00753025158875
0201	8	0.00969170723150	0.00956013473350	0.01067678461131	0.00813963611038
0201	9	0.00947243783712	0.00886665345749	0.01098561722403	0.00744319665709
0201	10	0.00991097662588	0.00960966911036	0.01191211506220	0.00822669104204
0201	11	0.01113888523439	0.01010501287893	0.01252978028765	0.00770436145207
0201	12	0.01096346971889	0.00926292847235	0.01173563928350	0.00744319665709
0201	13	0.01105117747664	0.01045175351694	0.01261801817700	0.00813963611038
0201	14	0.01232293996403	0.01223499108381	0.01292685078972	0.00792199878123
0201	15	0.01223523220628	0.01005547850208	0.01305920762375	0.00809610864455
0201	16	0.01337543305705	0.01035268476322	0.01367687284920	0.01014189953860
0201	17	0.01442792615006	0.01129383792351	0.01632400952969	0.00996778967528

Rysunek 2. Procentowy rozkład populacji w powiatach według wieku, płci i lokalizacji miasto-wieś

Dodatkowym zadaniem, koniecznym do wykonania przed przystąpieniem do procesu generowania wyjściowego obrazu populacji, było przygotowanie danych opisujących lokalizację. Zostało to zrobione w oparciu o rejestr TERYT, udostępniany przez główny Urząd Statystyczny. Rejestr TERYT udostępnia dane wszystkich lokalizacji w Polsce do poziomu ulicy, a dodatkowym zadaniem było uzupełnienie tych danych o współrzędne geograficzne co w przyszłości ma umożliwić realizację zobrażeń przestrzennych. Na rysunku 3 pokazano fragment tabeli zawierającej przygotowane i uzupełnione o współrzędne geograficzne dane z rejestru TERYT.

IdTeryt	Name	Hid	typeID	geo	SHid	Gmina	Powiat	Wojewodztwo	KodTeryt	CzyMiasto	SIMC
1	Dzierżonów	04585AB0	1	04E610000010C825D485623A8304017833CD68C5D4940	/1/2/2/1/	Dzierżonów	dzierżonowski	DOLNOŚLĄSKIE	0202021	1	0883988
2	Dziewa	046D39CA	1	04E610000010C80D2504390783240518138846D5D4440	/2/3/1/6/	Dąbrowa Białupia	nowocelawski	KUJAWSKO-POMORSKIE	0407022	0	0083859
3	Dziewiągła	04B73BE1	1	04E610000010C39844E01D534354089F4CAAB29594940	/13/7/2/4/	Iwaniska	opatowski	ŚWIĘTOKRZYSKIE	2606022	0	0793609
4	Dziewule	049F6FD01	1	04E610000010C1D12088526736408962F20698064440	/7/30/13/8/	Złuczyn	średzki	MAZOWIECKIE	1426132	0	0696361
5	Dziegielec	04AF331D5	1	04E610000010C16E8C9E877E93140FD288F13CB3F4840	/11/20/5/10/	Leczyce	węgrowski	POMORSKIE	2215082	0	0167214
6	Dziegiele Oleckie	048B7D85C0	1	04E610000010C38C76CF878364048878FFFD8F64440	/14/15/2/11/	Olecko	olecki	WARMIŃSKO-MAZURSKIE	2813043	0	0763650
7	Dzikowice	048DCB440	1	04E610000010CF8167BD4043A2F40D3F38C5059FCB4940	/4/13/6/3/	Sopotawa	zagajski	LUBUSKIE	0810073	0	0914600
8	Dzikowec	045D3A7440	1	04E610000010CA88FC01F7E923040F19C34CE10494940	/1/3/5/3/	Nowa Rudka	klodzki	DOLNOŚLĄSKIE	0208112	0	0854140
9	Dzikowo	0485F0D9	1	04E610000010CF880C07C8D683D4089F34E0579F84940	/4/3/4/12/	Gubin	kozłowski	LUBUSKIE	0802052	0	0909609
10	Dzikówko	04C118289C	1	04E610000010C584E958F14F2E40FC3FE282657E4440	/16/12/1/7/	Baliniek	myliński	ZACHODNIOPOMORSKIE	3210013	0	0177603
11	Dzunin	048595A80	1	04E610000010CAEE988CE488534403A158A5C0EDC4940	/7/6/7/10/	Mogielnica	grójcki	MAZOWIECKIE	1406073	0	0629270
12	Dzmitrzów	045CED78	1	04E610000010C8B5E5756992F406C5494096784940	/1/7/2/3/	Jeziw Sudecki	jeleniogórski	DOLNOŚLĄSKIE	0206062	0	0189760
13	Dzwońgra	04C11C178D80	1	04E610000010C8B5A718BE1DA30407948E78287E4440	/16/19/2/13/	Policzyn-Ząbki	świdwiński	ZACHODNIOPOMORSKIE	3216033	0	0310002
14	Dzwin	0492CDF18	1	04E610000010CC24AA48E2CF3340D874EFD0804440	/7/41/3/5/	Lubowidz	zamojski	MAZOWIECKIE	1437032	0	0118523
15	Dzwoniek	048F8EC30	1	04E610000010C5C47064086835409860581EFF784440	/7/15/3/12/	Czerwini	ostrołęcki	MAZOWIECKIE	1415032	0	0508201
16	Dzwonkowiec	048F8E78C	1	04E610000010C17E41D996C4E344023FF929323F1440	/5/19/7/5/	Nowy Kawęczyn	skarżewski	ŁÓDZKIE	1015072	0	0734222
17	Dzubece Cerkiewne	048AC3080	1	04E610000010C52B4F42F497137408541DC8B1E53440	/10/6/4/4/	Dzubece Cerkiewne	hajnrowski	PODLASKIE	2005052	0	0027915
18	Dulany	04A49EC30	1	04E610000010C363960F8E237408198B48263864440	/10/2/7/12/	Mchalowo	balotocki	PODLASKIE	2002073	0	0034252
19	Duchów	049EC8E80	1	04E610000010C30CF44D737BA35406259FC7E78274440	/7/12/3/8/	Dobre	miński	MAZOWIECKIE	1412062	0	0670290

Rysunek 3. Fragment tabeli zawierającej zmodyfikowane dane rejestru TERYT

W oparciu o statystyki Głównego Urzędu Statystycznego, ze szczególnym uwzględnieniem wyników Narodowego Spisu Powszechnego Ludności i Mieszkań z roku 2011, została utworzona wirtualna populacja o następujących cechach:

- jest tak samo liczna jak populacja wykazana w spisie powszechnym,
- jej rozmieszczenie terytorialne do poziomu gmin jest takie samo, jak to wykazane w spisie powszechnym,
- na podstawie wyników spisu powszechnego został przybliżony rozkład występowania określonych imion i nazwisk. Imiona i nazwiska w wirtualnej populacji zostały wylosowane na podstawie tego przybliżonego rozkładu,
- rozkłady płci i wieku w każdej gminie są przybliżeniem stanu prawdziwej populacji opisanym w wynikach spisu powszechnego,
- każdej osobie w wirtualnej populacji został nadany 14-cyfrowy zmodyfikowany PESEL, na zasadach podobnych do generacji prawdziwego numeru PESEL (11-cyfrowego),
- w systemie są uwzględnione wszystkie jednostki terytorialne z TERYT – urzędowego rejestru podziału terytorialnego prowadzonego przez Główny Urząd Statystyczny,

g) każdej wirtualnej osobie przypisano najmniejszą możliwą jednostką terytorialną z systemu TERYT (ulica w przypadku miast, wieś w przypadku wsi bez ulic).

h) każdej minimalnej jednostce terytorialnej zostały przyporządkowane współrzędne geograficzne pobrane z Google Maps API (Interfejs programistyczny aplikacji, ang. *Application Programming Interface*). Tym samym, każdej wirtualnej osobie także zostały przypisane współrzędne geograficzne miejsca zamieszkania.

Na rysunku 4 pokazano fragment tabeli zawierającej dane wygenerowanego obrazu populacji (stan na 3.12.2015).

IdOsoby	Nazwisko	Imie	DataUrodzenia	DPPesel	IdTeryt	CzyZyje	CzyKobieta	IdStanuCywilnego	InneDane	IdWykształcenia	RokUr	Wiek
-2000000000	Ozmbłowska	Karolina	2011-06-26	2011062699075	271360	1	1	1	<Dane />	1	2011	2
-1999999999	Dranka	Karolina	2011-01-12	20110112999056	71745	1	1	1	<Dane />	1	2011	2
-1999999998	Czulok	Jagoda	2011-07-28	20110728999000	214996	1	1	1	<Dane />	1	2011	2
-1999999997	Domkow	Aleksandra	2011-06-10	20110610999050	203418	1	1	1	<Dane />	1	2011	2
-1999999996	Kowalczyk	Olena	2011-09-13	20110913999059	25687	1	1	1	<Dane />	1	2011	2
-1999999995	Piestrak	Oliwia	2011-10-25	20111025999077	264851	1	1	1	<Dane />	1	2011	2
-1999999994	Przemieniecki	Amelia	2011-04-18	20110418999074	167512	1	1	1	<Dane />	1	2011	2
-1999999993	Przystaj	Małgorzata	2011-10-03	20111003999057	162644	1	1	1	<Dane />	1	2011	2
-1999999992	Pozienkiewicz	Maria	2011-05-14	20110514999062	49479	1	1	1	<Dane />	1	2011	2
-1999999991	Gienek	Izabela	2011-11-23	20111123999071	90299	1	1	1	<Dane />	1	2011	2
-1999999990	Grawirska	Maja	2011-05-20	20110520999057	117247	1	1	1	<Dane />	1	2011	2
-1999999989	Wlutke	Pola	2011-01-06	20110106999061	197152	1	1	1	<Dane />	1	2011	2
-1999999988	Sapiewska	Marika	2011-09-01	20110901999046	37826	1	1	1	<Dane />	1	2011	2
-1999999986	Kęmbłowska	Wiktona	2011-05-31	20110531999067	197152	1	1	1	<Dane />	1	2011	2
-1999999985	Kowalska	Hanna	2011-11-10	20111110999055	167512	1	1	1	<Dane />	1	2011	2
-1999999984	Pogoda	Antonina	2011-02-23	20110223999066	220029	1	1	1	<Dane />	1	2011	2
-1999999983	Taparek	Maja	2011-12-18	20111218999079	170016	1	1	1	<Dane />	1	2011	2
-1999999982	Smaudzik	Hanna	2011-06-10	20110610999050	256493	1	1	1	<Dane />	1	2011	2
-1999999981	Prześlakiewicz	Maja	2011-07-29	20110729999003	225740	1	1	1	<Dane />	1	2011	2
-1999999980	Gmośirska	Pola	2011-07-07	20110707999064	54148	1	1	1	<Dane />	1	2011	2
-1999999979	Osialkowska	Zuzanna	2011-09-05	20110905999058	224944	1	1	1	<Dane />	1	2011	2
-1999999978	Górska	Zuzanna	2011-06-27	20110627999078	256493	1	1	1	<Dane />	1	2011	2
-1999999977	Brzutek	Maja	2011-10-08	20111008999072	256493	1	1	1	<Dane />	1	2011	2
-1999999976	Artus	Izabela	2011-11-05	20111105999063	38898	1	1	1	<Dane />	1	2011	2
-1999999975	Machtycz	Małgorzata	2011-06-15	20110615999065	140173	1	1	1	<Dane />	1	2011	2
-1999999974	Loesche	Julia	2011-09-26	20110926999075	101464	1	1	1	<Dane />	1	2011	2
-1999999973	Organista	Marta	2011-07-27	20110727999078	35549	1	1	1	<Dane />	1	2011	2
-1999999972	Panaczek	Julia	2011-12-29	20111229999008	257309	1	1	1	<Dane />	1	2011	2

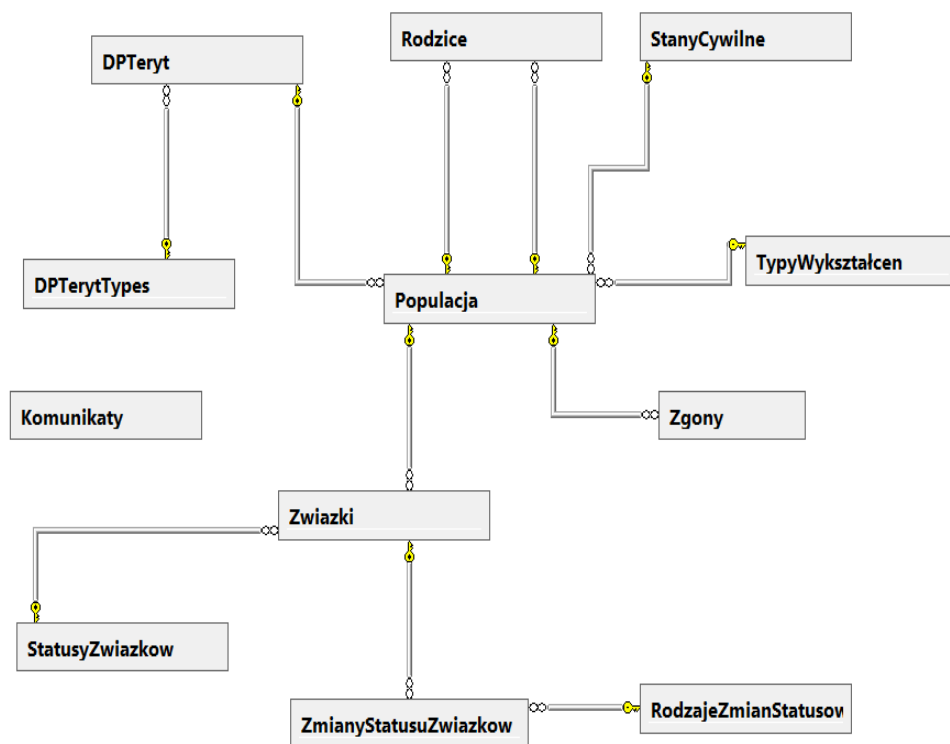
Rysunek 4. Fragment tabeli zawierającej wygenerowane dane

Każdy wiersz w tabeli, pokazanej na rysunku 4, opisuje jedną wirtualną osobę. Wygenerowane dane zapewniają, dla stanu z 31 grudnia 2015 roku, następujące cechy:

- rozkład terytorialny według płci, wieku i lokalizacji miasto-wieś zgony ze stanem rzeczywistym,
- rozkład wykształcenia według typu wykształcenia płci i lokalizacji miasto-wieś,
- rozkład popularnych nazwisk i imion zgodny z dostępnymi statystykami.

W opisie wirtualnej osoby wykorzystano zmodyfikowany numer Pesel, który odzwierciedla datę urodzenia, płeć oraz sumę kontrolną, został jednak zmodyfikowany w celu odróżnienia go od danych rzeczywistych.

Tabela zawierająca wygenerowane dane zawiera ok. 38 milionów wierszy i jest podstawową tabelą bazy danych, której schemat pokazano na rysunku 5.



Rysunek 5. Schemat bazy danych „Dynamiczna populacja”

3. Generatory zdarzeń

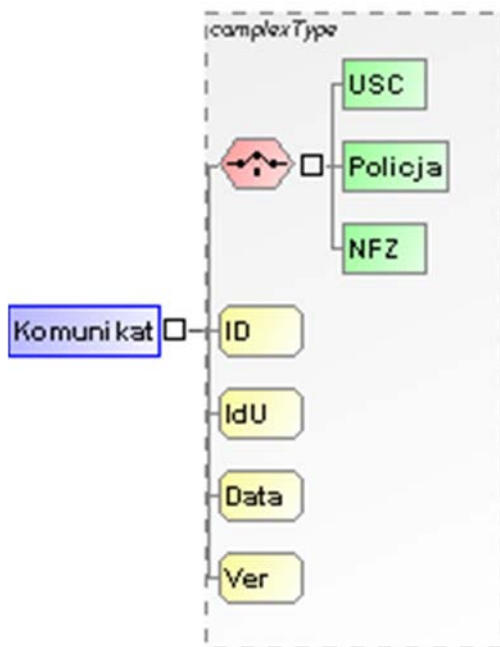
Podczas kolejnego etapu należało przygotować generatory symulujące zmiany stanu wyjściowego populacji. Mechanizm generowania danych opisujących zmiany populacji działa według schematu:

- codziennie, w kontekście każdej gminy uruchamiany jest proces, który ma przygotować komunikat o zmianach populacji (urodzenia, zgony i zmiany stanu cywilnego),

- generowanie dziennych danych na poziomie gminy oparte jest o dostępne statystyki,
- komunikaty zapisywane są w bazie danych,
- po zapisaniu komunikatu oddzielny proces zapisuje dane z komunikatu w odpowiednich tabelach.

Na potrzeby procesu generatora zdarzeń został opracowany schemat dokumentu XML, który opisywał postaci komunikatów, które będą tworzone dla każdej gminy.

Na rysunku 6 pokazano schemat budowy komunikatu.



Rysunek 6. Schemat komunikatu generatora zdarzeń

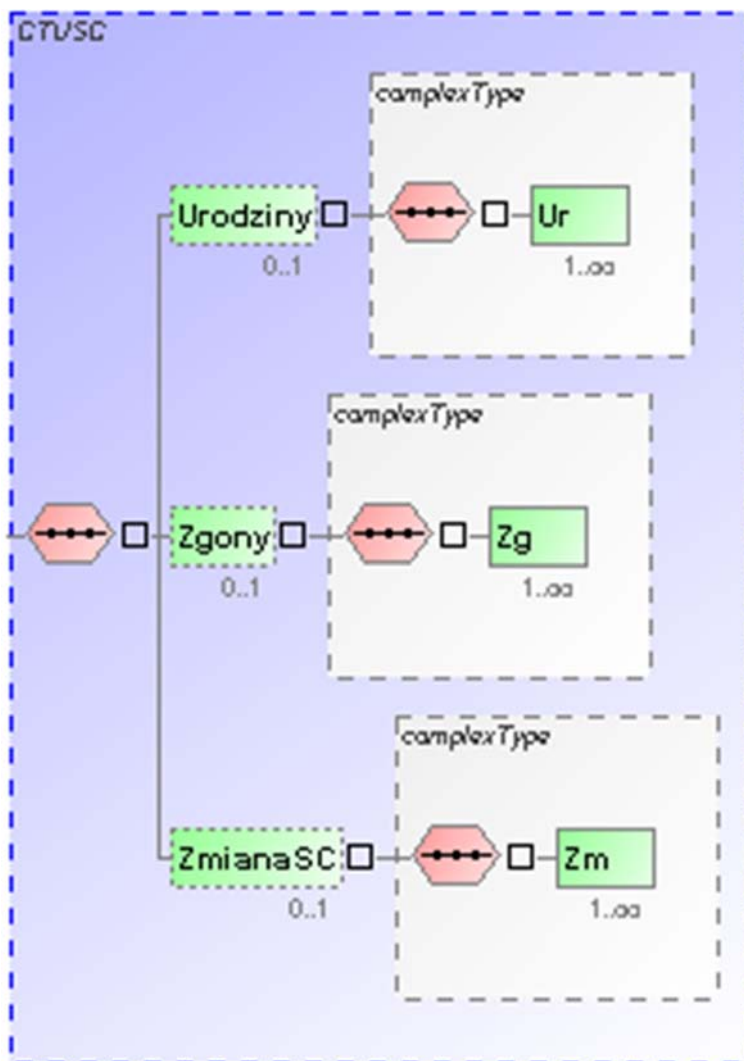
W schemacie przewidziano trzy główne sekcje:

- USC,
- Policja,
- NFZ.

Sekcje Policja i NFZ zostały zaproponowane do dalszych prac rozwijających system o moduł rejestracji informacji o wypadkach drogowych oraz o problemy epidemio-

logiczne i na obecnym etapie nie były jeszcze wykorzystywane. Pierwszy etap realizacji generatorów w systemie „Dynamiczna populacja” zakładał generowanie komunikatów sekcji USC.

Na rysunku 7 pokazano schemat dokumentu XML dla sekcji USC.



Rysunek 7. Schemat komunikatu XML sekcji USC

Komunikat w sekcji USC powinien zawierać dane dotyczące urodzeń, zgonów oraz zmian stanu cywilnego w konkretnej gminie, danego dnia.

Na rysunku 8 pokazano przykładową postać gotowego komunikatu.

```
<Komunikat Id="1" IdU="18781" Data="20130318" Ver="1">
  <USC>
    <Urodziny>
      <Ur Pesel="" Imie="Jacek" Nazwisko="Kowalski" PeselO="19940724899909" PeselM="19940202904812" Data="20130315" Plec="M"/>
      <Ur Pesel="" Imie="Olga" Nazwisko="Masłowska" PeselO="19820918864961" PeselM="19820210870844" Data="20130316" Plec="K"/>
    </Urodziny>
    <Zgony>
      <Zg Pesel="19390803914876" Data="20130311" ICD="Q01"/>
      <Zg Pesel="19350912927875" Data="20130314" ICD="Y85"/>
      <Zg Pesel="19340504959911" Data="20130314" ICD="T51"/>
    </Zgony>
    <ZmianySC>
      <Zm Pes1="19851125860873" Naz1Po="Iżewicz-Friess" Pes2="19790221861961" Data="20130314" Typ="M"/>
      <Zm Pes1="19830927864868" Naz1Po="Wermajster-Hilber" Pes2="19701113905901" Naz2Po="Hilber-Wermajster" Data="20130314" Typ="M"/>
    </ZmianySC>
  </USC>
</Komunikat>
```

Rysunek 8. Przykłady komunikat sekcji USC

Na potrzeby generowania informacji o zgonach przygotowana została tabela zawierająca kody ICD10.

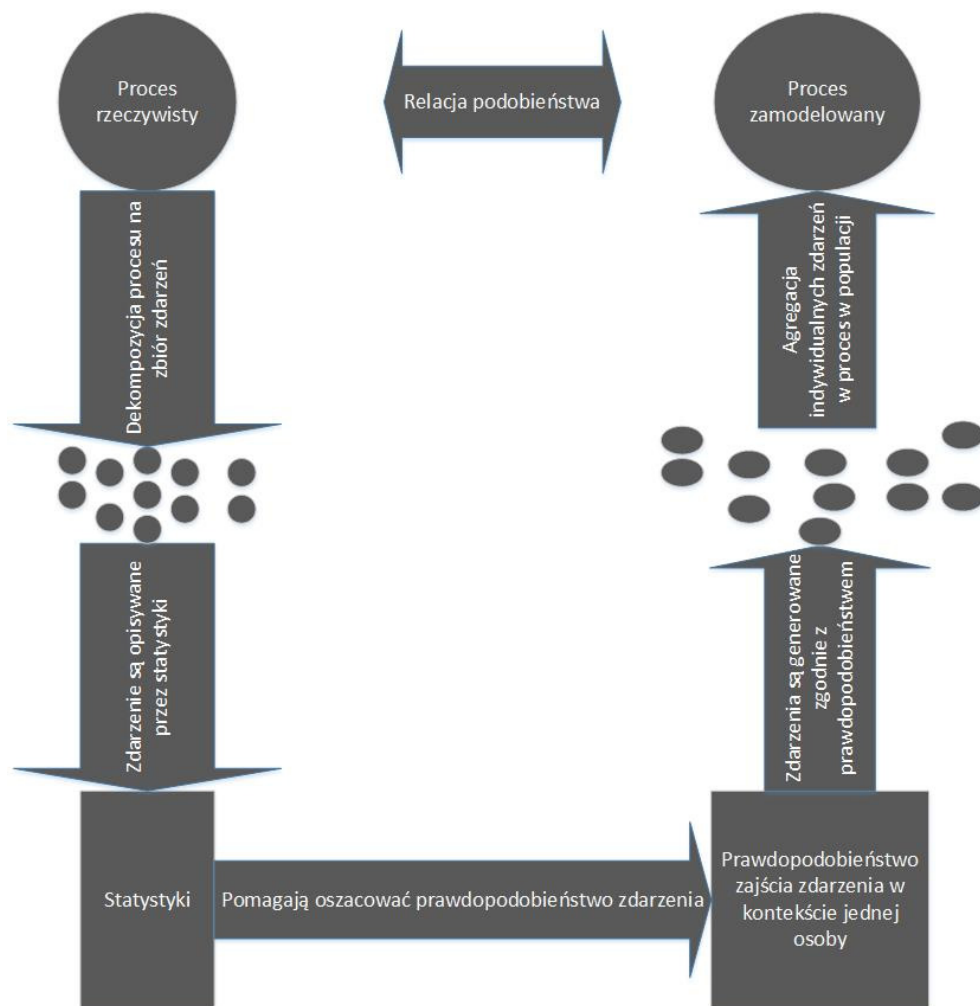
Na rysunku 9 pokazano fragment tabeli zawierającej dane kodów ICD10.

id	name	parent
D34	Niezłośliwy nowotwór tarczycy	D10-D36
D35	Niezłośliwy nowotwór innych i nieokreślonych gruczołów wydzielania wewnętrznego	D10-D36
D50	Niedokrwistości z niedoboru żelaza	D50-D53
D50-D53	Niedokrwistości z niedoborów pokarmowych	D50-D89
D50-D89	Choroby krwi i narządów krwiotwórczych oraz niektóre choroby przebiegające z udziałem mechanizmów autoimmunologicznych	
D51	Niedokrwistości z niedoboru witaminy B12	D50-D53
D52	Niedokrwistość z niedoboru kwasu foliowego	D50-D53
D55	Niedokrwistość spowodowana zaburzeniami enzymatycznymi	D55-D59
D55-D59	Niedokrwistości hemolityczne	D50-D89
D56	Talasemia	D55-D59
D57	Zaburzenia połączone z sierpowatokrwinkowocią	D55-D59
D58	Inne dziedziczne niedokrwistości hemolityczne	D55-D59
D60	Nabyta aplazja czysto czerwonekrwinkowa (erythroblastopenia)	D60-D64
D60-D64	Niedokrwistości aplastyczne i inne	D50-D89
D61	Inne niedokrwistości aplastyczne	D60-D64
D62	Ostra niedokrwistość pokrwotoczna	D60-D64
D63	Niedokrwistość w przebiegu chorób przewlekłych sklasyfikowanych gdzie indziej	D60-D64
D65	Rozlane wykrzepianie wewnątrznaczyniowe	D65-D69
D65-D69	Zaburzenia krzepnięcia, płamice oraz inne skazy krwotoczne	D50-D89

Rysunek 9. Fragment tabeli ICD10

Po przygotowaniu opisanych elementów w systemie „Dynamiczna Populacja” można było przystąpić do modelowania rzeczywistych globalnych procesów społecznych.

Na rysunku 10 pokazano schemat działania procesów generatora zdarzeń,



Rysunek 10. Modelowanie i symulacja procesów zachodzących w populacji

W oparciu o stan początkowy populacji, jak i statystyki opisujące procesy zachodzące w populacji rzeczywistej, podjęta została próba symulowania tych procesów w populacji wirtualnej. W pierwszej kolejności do symulacji wybrano procesy:

- Zgony – typowanie wirtualnych osób tak, by rozkład zgonów według płci, wieku i miejsca zamieszkania był zgodny ze statystykami. Wskazywanie przyczyny zgonu, tak by rozkład przyczyn zgonów był zgodny również ze statystykami GUS. Do klasyfikacji przyczyn zgonów użyto Międzynarodowej Statystycznej Klasyfikacji Chorób i Problemów Zdrowotnych ICD-10 [2].

- Urodzenia – typowanie matki i ojca dla nowej wirtualnej osoby i przypisywanie jej wszystkich wymaganych cech. Rozkład urodzeń w zależności od wieku matki i ojca jest inny dla każdego województwa i zgodny ze statystykami GUS za rok 2011.

4. Posumowanie

Prace koła naukowego w ramach projektu „Dynamiczna populacja” przewidują jeszcze następujące etapy:

- Uruchomienie testowego generowania danych sekcji USC – aktualnie działa generowanie zgonów i urodzeń,
- Przeprowadzenie analizy statystycznej wyników symulacji – porównanie wyników symulacji z rzeczywistymi danymi dla kolejnych lat,
- Uruchomienie pełnej wersji komunikatu USC – implementacja generowania zmian stanu cywilnego jest w trakcie opracowania,
- Uruchomienie pełnej symulacji – dla dodatkowych sekcji komunikatu: Policja i NFZ,
- Budowa systemów analitycznych,
- Tworzenie alertów i scenariuszy wykorzystania,

Baza danych systemu „Dynamiczna populacja” jest wykorzystywana w pracach koła naukowego a także w trakcie zajęć specjalizacyjnych kierunku Inżynieria Baz Danych. Aktualnie prowadzone jest badanie porównujące działanie generatora w latach 2012-2014 z rzeczywistymi danymi udostępnionymi przez Główny Urząd Statystyczny [3].

Bibliografia

- [1] Główny Urząd Statystyczny, *Ludność w gminach według stanu w dniu 31. 12. 2011 r. bilans opracowany w oparciu o wyniki NSP’2011*, <http://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/ludnosc-w-gminach-wedlug-stanu-w-dniu-3112-2011-r-bilans-opracowany-w-oparciu-o-wyniki-nsp-2011,2,1.html>
- [2] Międzynarodowa Statystyczna Klasyfikacja Chorób i Problemów Zdrowotnych ICD-10, <http://www.icd10.pl/>
- [3] <http://demografia.stat.gov.pl/bazademo>

**A concept of a “Dynamic Population” Training Database.
Project implemented by Students Scientific Association**

Abstract

The paper presents a concept of a Training Database, which is being implemented by Students Scientific Association. It characterizes the genesis of the project, along with its main principles. Realization of the fundamental stages of the project and the vision of the further development of the system has been shown.

Keywords – Business Intelligence, Training Database, Dynamic Population