Polonizacja SAS

Zespół Wsparcia Technicznego, SAS Institute Polska grudzień 2014

Szanowni Państwo !

Oddajemy do rąk Państwa kolejną wersję naszego dokumentu opisującego zagadnienia związane z korzystaniem z oprogramowania SAS w polskich warunkach czyli aspekty lokalizacyjne SAS-a. Wśród omawianych tematów znajdują się m.in.:

- Sortowanie danych
- Formatowanie liczb, dat oraz walut
- Kodowanie znaków
- Konfiguracja aplikacji klienckich

Dokument jest polecany wszystkim, stykającym się z zagadnieniami związanymi z polskimi znakami podczas pracy z SAS. Lektura może oszczędzić wysiłku w czasie tworzenia i uruchamiania aplikacji.

Jednocześnie zaznaczamy, że dokument ten jest uzupełnieniem, a nie zastąpieniem dokumentacji. W szczególności polecamy publikację **„SAS 9.4 National Language Support (NLS): Reference Guide”** dostępną dla wszystkich naszych Klientów w wersji PDF oraz HTML.

Najnowsze wersje dokumentu i inne informacje związane z tematem będą publikowane na naszych stronach WWW pod adresem http://www.sas.com/poland/services

W razie dodatkowych pytań prosimy o kontakt z naszym Zespołem, telefonicznie na numer **(22) 560 4666** lub elektronicznie pod adresem support@spl.sas.com.

Prosimy o przesyłanie wszelkich uwagi i sugestii. Pomogą one w przygotowaniu następnych wersji tego dokumentu.

Zawsze służąc Państwu pomocą,
Zespół Pomocy Technicznej
Centrum Technologii SAS Polska

Spis treści

1. Instalacja SAS ... 3
1.1 Wersje językowe aplikacji klienckich SAS ... 3
1.1.1 Aplikacje SAS dostępne w języku polskim ... 4
1.2 Ustawienia regionalne ... 5
2 Wersje językowe aplikacji klienckich ... 6
2.1 SAS Foundation ... 6
2.1.1 Czcionki ... 7
2.1.2 Drukowanie ... 7
2.2 SAS Enterprise Guide ... 7
2.3 SAS Add-in for Microsoft Office ... 8
2.4 Aplikacje JAVA-owe ... 8
2.4.1 SAS Enterprise Miner ... 10
2.5 Aplikacje WEB-owe ... 11
2.5.1 SAS Preferences ... 11
2.5.2 Język przeglądarki ... 11
3 Przetwarzanie danych ... 12
3.1 Ustawienia regionalne ... 12
3.1.1 Liczby ... 12
3.1.2 Waluty ... 13
3.1.3 Data i czas ... 15
3.1.4 Modyfikacja ustawień reginalnych (SAS 9.4) ... 17
3.2 Kodowanie znaków ... 19
3.2.1 Unicode (UTF-8) ... 20
3.2.2 Postać uniwersalna (SAS 9.4) ... 21
3.3 Sortowanie ... 21
3.3.1 Ustawianie kolejności sortowania ... 21
3.3.2 Własny porządek sortowania ... 22
3.3.3 Sortowanie lingwistyczne (sortseq=linguistic) ... 23
3.3.4 Znane problemy ... 24
3.4 Nazwy tabel i kolumn ... 24
3.5 Przenoszenie danych ... 25
3.5.1 Procedura COPY ... 25
3.5.2 Procerury CPORT oraz CIMPORT ... 26
3.5.3 Ręczne przekodowanie danych ... 27
3.5.4 Przenoszenie tabel z dużą ilością obserwacji (SAS 9.4) ... 28
4 Wykresy i raporty ... 29
4.1 Czcionki systemowe (HARDWARE FONTS) ... 29
4.2 Czcionki SAS/GRAPH (SOFTWARE FONTS) ... 31
4.3 Mapa Polski ... 32
5 Zewnętrzne źródła danych ... 33
6 Konfiguracja serwerów obliczeniowych SAS ... 34

1. Instalacja SAS

Pierwsze decyzje związane z polonizacją SAS podejmowane są już w trakcie instalacji. W jej trakcie dokonujemy dwóch podstawowych wyborów:

1. Jakie wersje językowe aplikacji są instalowane?
2. Z jakich ustawień regionalnych ma korzystać SAS?

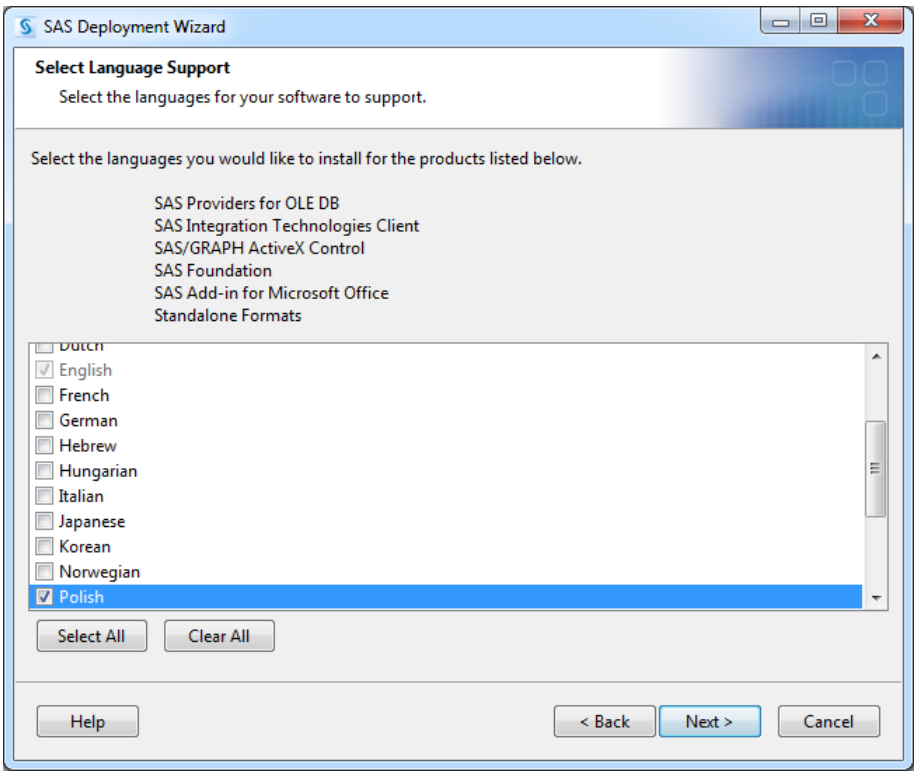
W dalszej części tego rozdziału przyjrzymy się dokładniej w jaki sposób wymienione decyzje są podejmowane oraz jaki jest ich wpływ na pracę z SAS-em.

1.1 Wersje językowe aplikacji klienckich SAS

Wersja językowa aplikacji jest to język interfejsu użytkownika – pozycje menu, przyciski, opisy. Nie ma ona wpływu na przetwarzanie danych*.

**** Istnieją wyjątki od tej reguły. Szczegółowe informacje znajdują się w rozdziałach poświęconych poszczególnym aplikacjom.***

Wybór wersji językowych aplikacji klienckich SAS jest podejmowany w ramach kroku **„Select Language Support”**. Ekran wyboru powinien wyglądać podobnie do przedstawionego poniżej:



W ramach tego kroku wybieramy w jakich wersjach mają być dostępne aplikacje wymienione u góry okna. Domyślnie instalowana jest wyłącznie wersja angielska oraz unicode. Lista dostępnych aplikacji może się różnić w zależności od konkretnego zestawu instalowanych produktów.

W przypadku wymienionych programów dodanie dodatkowego języka wymaga ponownego uruchomienia procesu instalacji.

UWAGA. Jeżeli aplikacja, którą instalujemy, nie znajduje się na liście, oznacza to iż jest ona instalowana ze wszystkimi dostępnymi wersjami językowymi. Dotyczy to w szczególności aplikacji web-owych (SAS Visual Analytics, SAS Web Report Studio, SAS Information Delivery Portal) oraz java-owych (SAS Enterprise Miner, SAS Forecast Server).

1.1.1 Aplikacje SAS dostępne w języku polskim

W momencie tworzenia tego dokumentu w języku polskim dostępne były następujące aplikacje:

- SAS Add-In for Microsoft Office
- SAS Asset Performance Analytics
- SAS BI Dashboard
- SAS Business Data Network
- SAS Custom Page Builder
- SAS Enterprise Case Management
- SAS Enterprise GRC (dawny SAS OpRisk Monitor)
- SAS Enterprise Guide
- SAS Enterprise Miner
- SAS Financial Crimes Monitor
- SAS Financial Management
- SAS Forecast Server
- SAS Foundation (dawny SAS Base)
- SAS High Performance Analytics
- SAS High Performance Management Console
- SAS Information Delivery Portal
- SAS Lineage Mid-Tier
- SAS Master Data Management
- SAS MEMEX
- SAS OpRisk Var
- SAS Quality Analytic Suite
- SAS Search Interface to SAS Content
- SAS Social Network Analysis Server
- SAS Strategy Management
- SAS Studio
- SAS Text Miner
- SAS Time Series Studio
- SAS Universal Viewer
- SAS University Edition Information Center
- SAS Visual Analytics
- SAS Visual Scenario Designer
- SAS Visual Statistics
- SAS Web Parts for MS Sharepoint
- SAS Web Report Studio
- SAS Workflow Studio

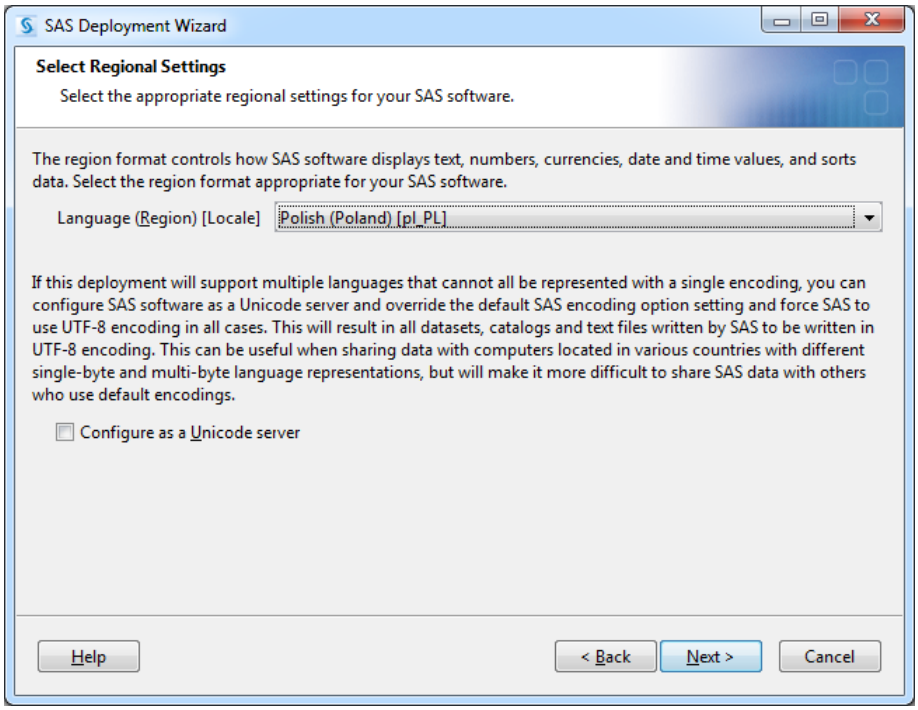
Na życzenie klienta jest również dostępna polska wersja dla poniższych produktów:

- SAS Data Integration Studio
- SAS Information Map Studio
- SAS Management Console
- SAS OLAP Cube Studio

1.2 Ustawienia regionalne

Ustawienia regionalne decydują o wielu aspektach związanych z przetwarzaniem danych, takich jak sposób zapisu tekstów, liczb, dat czy walut.

Wybór domyślnych ustawień regionalnych SAS jest podejmowany w ramach kroku **„Select Regional Settings”**.



W ramach tego kroku wskazujemy kraj lub region, którego ustawienia regionalne mają być użyte. Dla języka polskiego należy wybrać pozycję „**Polish (Poland) [pl_PL]**”. Oprócz tego mamy możliwość wyboru czy chcemy korzystać z kodowania Unicode. Pozwala ono m.in. na przechowywanie w jednym zbiorze tekstów zawierających litery z różnych języków (nie mieszczących się w obrębie jednego formatu kodowania).

UWAGA. Przetwarzanie tekstów w Unicode wymaga zastosowania specjalnych technik oraz funkcji. Szczegółowe informacje na ten temat znajdują się w rozdziale 3.2.1.

2 Wersje językowe aplikacji klienckich

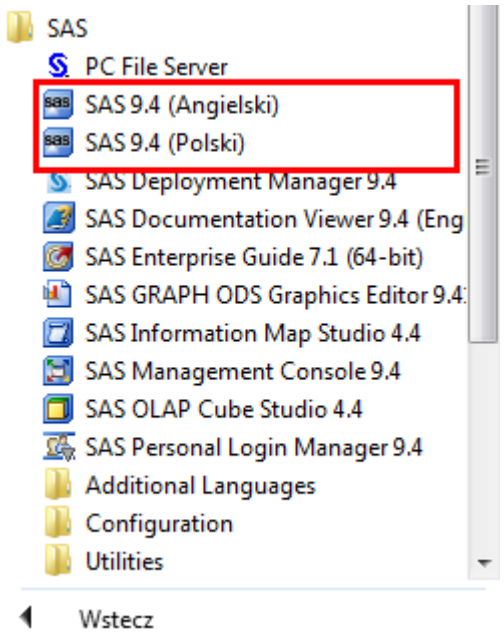
Domyślnie aplikacje SAS działają w języku angielskim. W tym rozdziale przedstawiamy metody pozwalające na uzyskanie interfejsu użytkownika w języku polskim.

UWAGA. W większości przypadków język aplikacji wpływa na ustawienia regionalne sesji SAS. W pewnych przypadkach może to również wpływać na kodowanie znaków i powodować błędy podczas przetwarzania danych. Szczegółowy opis tego zachowania oraz proponowane rozwiązania znajdują się na stronie http://support.sas.com/kb/35/644.html.

2.1 SAS Foundation

- Windows

Jeżeli instalujemy SAS Foundation na platformie Windows, w menu start tworzone są skróty do wszystkich wersji językowych wybranych podczas instalacji. Obrazuje to poniższy zrzut ekranu.



Aby uruchomić SAS w odpowiednim języku należy wybrać odpowiadającą mu pozycję w menu start. Jeżeli szukanej wersji nie ma w folderze SAS, należy sprawdzić zawartość folderu ***Additional Languages***.

W przypadku uruchomienia w trybie wsadowym, wersja językowa jest ustalana na podstawie pliku ***sasv9.cfg*** wskazanego w ramach opcji ***–config***. Domyślnie SAS jest uruchamiany zgodnie z ustawieniami zapisanymi w pliku ***<SASHome>\SASFoundation\9.4\sasv9.cfg***. Aby wskazać inny język dla takiej sesji (w tym przypadku język polski), należy uruchomić SAS z opcją ***–config <SASHome>\SASFoundation\9.4\nls\pl\sasv9.cfg***.

- Linux / UNIX

W systemach z rodziny Linux / UNIX wybór wersji językowej wymaga modyfikacji pliku **<SASHOME>/SASFoundation/9.4/sas**. Plik ten jest łączem symbolicznym do pliku sas_<język> znajdującego się w katalogu**<SASHOME>/SASFoundation/9.4/bin**. Aby ustawić wersję polską należy zmodyfikować plik aby wskazywał na **sas_pl**.

Przykładowy kod wykonujący taką modyfikację:

```
cd /opt/SASHome/SASFoundation/9.4
ln -s bin/sas_pl sas
```

2.1.1 Czcionki

Po wybraniu polskiej wersji SAS wszystkie okna i menu powinny być wyświetlane w języku polskim. Jeżeli występuje problem z wyświetlaniem polskich znaków należy upewnić się że wybrana czcionka obsługuje oraz ma ustawiony skrypt 'Środkowoeuropejski' albo 'Europa Środkowa'.

Ustawienia czcionki modyfikujemy w następujący sposób:

- edytor programów – z menu ***Narzędzia*** wybieramy ***Opcje → Enhanced Editor → Wygląd***
- log i wynik – z menu ***Narzędzia*** wybierany ***Opcje → Czcionki…***

Aby wprowadzone zmiany zostały zachowane po zamknięciu programu, należy wybrać ***Narzędzia → Opcje → Preferencje → Ogólne*** oraz zaznaczyć pole „**Zapisuj ustawienia końcowe**”.

UWAGA. Jeżeli w menu Narzędzia → Opcje nie ma szukanej pozycji należy kliknąć element, który chcemy zmodyfikować (np. okno logu) i spróbować ponownie.

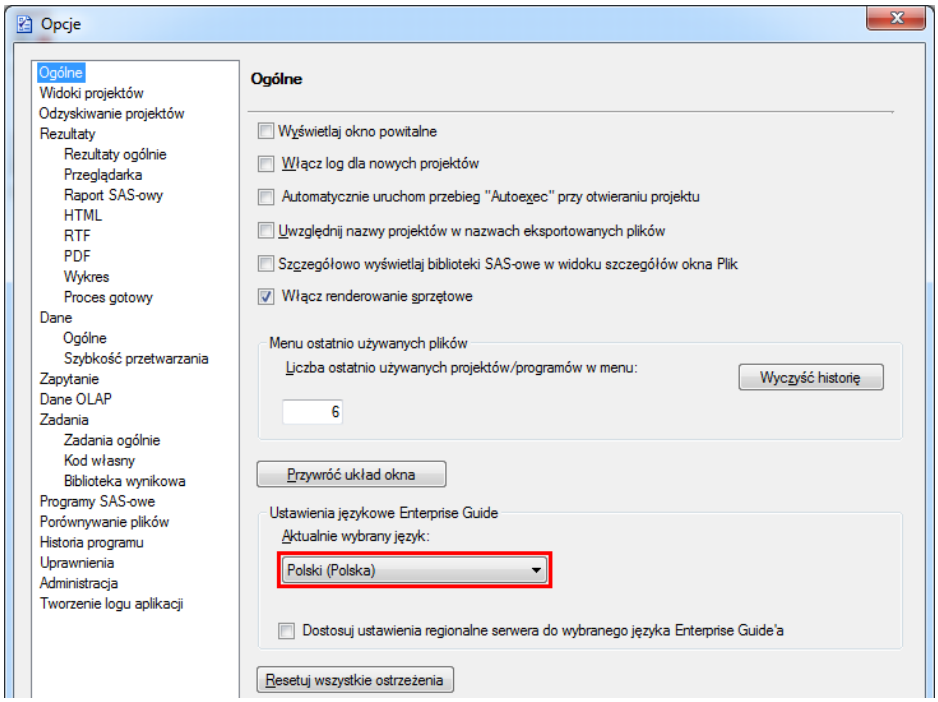
2.1.2 Drukowanie

Jeżeli na wydruku nie ma polskich liter, najprawdopodobniej jest to problem użytej czcionki. Aby zmienić czcionkę używaną do wydruku należy wejść w ustawienia drukarki (***Plik → Ustawienia drukarki***) i kliknąć przycisk ***Czcionka***. W wyświetlonym oknie wybieramy czcionkę zawierającą polskie znaki. Należy pamiętać, aby po wyborze ustawić skrypt na Środkowoeuropejski.

Alternatywną metodą jest ustawienie opcji ***sysprintfont*** z poziomu kodu SAS-owego.

2.2 SAS Enterprise Guide

Wersję językową programu SAS Enterprise Guide ustawiamy w opcjach programu. W tym celu z menu ***Narzędzia*** wybieramy pozycję ***Opcje…*** i ustawiamy żądany język w sekcji **Ustawienie językowe Enterprise Guide.**



Zmiana wersji językowej następuje po ponownym uruchomieniu aplikacji.

UWAGA. Od wersji 7.1 jest dostępna opcja wyłączenia zależności ustawień regionalnych od języka aplikacji. W tym celu zaleźy odznaczyć pozycję „Dostosuj ustawienia regionalne serwera do wybranego języka Enterprise Guide'a" w opcjach programu.

2.3 SAS Add-in for Microsoft Office

Domyślnie SAS Add-In do Microsoft Office jest wyświetlany w tym samym języku, co pakiet Office. Aby zmienić język aplikacji bez zmiany języka pakietu Office należy wykonać następujące kroki:

- W folderze instalacji programu (np. *<SAS_HOME>\x86\SASAddinforMicrosoftOffice\7.1*) tworzymy plik ***SAS.OfficeAddin.dll.config***.
- W pliku SAS.OfficeAddin.dll.config umieszczamy poniższy fragment:

```
<configuration>
    <assemblySettings>
        <add key="CultureOverride" value="pl-PL"/>
    </assemblySettings>
</configuration>
```

Powyższy kod powoduje, że SAS Add-In do Microsoft Office będzie zawsze wyświetlany w polskiej wersji językowej. Aby uzyskać wersję angielską w polu CultureOverride zamiast ***pl-PL*** wpisujemy ***en-US.***

2.4 Aplikacje JAVA-owe

Język aplikacji JAVA-owych jest ustawiany na postawie wartości parametrów zdefiniowanych w pliku .ini, znajdującym się w katalogu instalacyjnym aplikacji. Parametry te można edytować ręcznie bądź użyć narzędzia **SAS Deployment Manager** (SAS 9.4) lub **SAS Locale Setup Manager** (SAS 9.2 i SAS 9.3).

- **SAS Deployment Manager**

Zmiana języka aplikacji przy pomocy **SAS Deployment Manager** została wprowadzona w SAS 9.4 i przebiega następująco:

- Po uruchomieniu programu z listy dostępnych akcji wybieramy pozycję ***Locale Setup Manager***, znajdującą się w sekcji ***Administration Tasks***.
- Wybieramy docelowy język z listy rozwijanej oraz wskazujemy aplikacje, które chcemy zmodyfikować.

Wybrane aplikacje są modyfikowane po wciśnięciu przycisku ***Next***. Nowa wersja językowa jest dostępna po ponownym uruchomieniu aplikacji.

- **SAS Locale Setup Manager**

Aplikacja **SAS Locale Setup Manager** jest dostępna w SAS 9.2 oraz SAS 9.3. Aby zmienić wersję językową aplikacji należy wykonać następujące kroki:

- Jeżeli szukanej aplikacji nie ma na liście produktów musimy ją ręcznie dodać. W tym celu wybieramy przycisk ***Dodaj*** i wskazujemy plik exe aplikacji którą chcemy zmodyfikować.
- Po dodaniu aplikacji klikamy na niej, wybieramy opcję ***Modyfikuj***, wybieramy nowe ustawienia regionalne i potwierdzamy zmiany.
- Aby zmiany zostały wprowadzone wybieramy opcję ***Aktualizuj ustawienia regionalne produktu*** i potwierdzamy chęć wprowadzenia zmian.

UWAGA. Zmiany są wprowadzany dopiero w po wybraniu opcji Aktualizuj ustawienia regionalne produktu.

- **Ręczna edycja plików ini.**

Ręczna edycja wersji językowej odbywa się w następujący sposób:

- W katalogu instalacji programu znajdujemy plik z rozszerzeniem ini i otwieramy go w dowolnym edytorze tekstów.
- Ustawiamy wartość opcji ***–Duser.language*** oraz ***–Duser.country*** na **pl**.
- Jeżeli w pliku nie ma ustawionych opcji ***–Duser.language*** oraz ***–Duser.country***, należy je dopisać na końcu bloku linii zaczynających się od JavaArgs. Poniżej przedstawione zostały przykładowe wartości opcji dla języka polskiego.

```
JavaArgs_15=-Duser.language=pl
JavaArgs_16=-Duser.country=PL
```

UWAGA. Przy dopisywaniu opcji do pliku należy pamiętać aby zachować dostosować się do numeru opcji rozpoczynających się od „JavaArgs_”.

Niektóre aplikacje klienckie wymagają dodatkowych akcji w celu pełnej zmiany języka. Szczegółowe informacje znajdują się w poniższych rozdziałach.

2.4.1 SAS Enterprise Miner

Od wersji 13.1 zmianę języka Enterprise Miner można zrealizować przez zdefiniowanie zmiennych środowiskowych ***emclient.user.language*** oraz ***emclient.user.country***. Zmienne te kontrolują zarówno ustawienia standardowego klienta jak również wpływają na język klienta instalowanego mechanizmem Java Web Start.

Jeżeli wyniki przetwarzań są wyświetlane w innym języku niż reszta aplikacji, oznacza to jest konieczna modyfikacja ustawień regionalnych (opcja ***locale***). Zalecaną metodą zmiany ustawień jest wymuszenie właściwej wartości opcji ***locale*** w ramach kodu wykonywanego przy starcie projektu (**Project start-up code**). Kod ten definiuje się we właściwościach projektu.

UWAGA. Aby zmiana ustawień regionalnych została uwzględniona, konieczne jest ponowne przeliczenie diagramu. W tym celu w opcjach zbioru danych (Data Source) należy ustawić opcję Rerun = Yes i ponownie uruchomić projekt. Istnieje również możliwość zdefiniowania Project start-up code dla wszystkich projektów. Aby to zrobić należy uruchomić program SAS Management Console, a następnie wybrać zakładkę Plugins → Application Management → Enterprise Miner → Projects → SASApp – Logical Workspace Server i wejść we właściwości. W zakładce Options w polu Initialization Code można wskazać plik zawierający kod, który ma być uruchamiany przy starcie projektu.

UWAGA. W przypadku posiadania programu SAS Enterprise Miner w wersji Workstation możliwości konfiguracji ustawień regionalnych ograniczają się do modyfikacji ustawień SAS Foundation.

2.5 Aplikacje WEB-owe

Język aplikacji web-owych SAS można ustawić za pomocą aplikacji ***SAS Preferences***. Domyślnie aplikacja korzysta z domyślnego języka wyświetlania zdefiniowanego w przeglądarce internetowej.

2.5.1 SAS Preferences

Aplikacja ***SAS Preferences*** jest dostępna pod adresem **http://<serwer>:<port>/SASPreferences**. Aby zmienić język aplikacji web-owych należy przejść do sekcji ***Regionalne (Regional)*** i wybrać odpowiedni język z menu ***Ustawienia regionalne użytkownika (User locale)***. Następnie zatwierdzamy zmiany klikając przycisk OK. Nowe ustawienia dotyczą wszystkich aplikacji i są uwzględniane po ponownym zalogowaniu.

UWAGA. Zmiana ustawień językowych jest również dostępna z poziomu niektórych aplikacji. Pozwalają na to m.in. SAS Information Delivery Portal oraz SAS Visual Analytics. Tak jak w przypadku SAS Preferences, wprowadzone zmiany dotyczą wszystkich aplikacji.

2.5.2 Język przeglądarki

Domyślnie język aplikacji web-owych zależy od domyślnego języka ustawionego w przeglądarce internetowej. Poniżej przedstawione zostały sposoby zmiany domyślnego języka dla najpopularniejszych przeglądarek:

- Internet Explorer
  - W menu ***Narzędzia*** wybieramy ***Opcje internetowe***.
  - W zakładce ***Ogólne*** w sekcji ***Wygląd*** klikamy przycisk ***Języki***.
  - Kolejność języków ustawiamy na liście w sekcji ***Preferencje językowe***.
  - Zatwierdzamy zmiany.
- Mozilla Firefox
  - W menu ***Narzędzia*** wybieramy ***Opcje***.
  - Przechodzimy na zakładkę ***Treść*** i klikamy przycisk ***Wybierz…*** w sekcji ***Języki***.
  - Ustawiamy kolejność języków na wyświetlonej liście.
  - Zatwierdzamy zmiany.
- Google Chrome
  - Z menu ***Dostosowywanie i kontrolowanie Google Chrome*** wybieramy pozycję ***Ustawienia***.
  - Klikamy link ***Pokaż ustawienia zaawansowane***.
  - Znajdujemy sekcję języki i klikamy przycisk ***Ustawienia języka i wprowadzania treści…***
  - Ustawiamy kolejność języków na wyświetlonej liście.
  - Zatwierdzamy zmiany.

Tak jak w przypadku zmiany języka przy użyciu SAS Preferences, zmiana języka następuje po ponownym zalogowaniu do aplikacji.

3 Przetwarzanie danych

W rozdziale tym skupimy się na aspektach polonizacyjnych związanych z przetwarzaniem danych w systemie SAS.

3.1 Ustawienia regionalne

Ustawienia regionalne określają elementy prezentacji danych specyficzne dla danego kraju bądź regionu. Decydują one m.in. o następujących elementach:

- Domyślne kodowanie znaków (wlatin2 czy wlatin1)
- Sposób zapisu daty (01 stycznia 2015 czy January 01, 2015)
- Sposób zapisu liczb (12 345,67 czy 12,345.67)
- Sposób zapisu walut (12 345,67 zł czy $12,345.67)
- Domyślny format wydruku (A4 czy letter)

Wartości domyślne wymieniowych elementów są ustalane na podstawie wartości zmiennej ***locale***. Wartością odpowiadającą Polsce jest ***PL_PL*** lub ***POLISH***. Oprócz tego każdy element może być konfigurowany przy użyciu dedykowanej opcji. Najważniejsze opcje to:

Opcja	Opis
DATESTYLE	Sposób zapisu daty, tzn. kolejność takich elementów jak dzień, miesiąc i rok.
DFLANG	Określa język dla międzynarodowych formatów zapisu daty.
PAPERSIZE	Rozmiar papieru używanego podczas drukowania.
TIMEZONE	Strefa czasowa
ENCODING	Domyślne kodowanie znaków

UWAGA. Wartość opcji ENCODING jest ustawiana podczas startu sesji SAS i nie może być zmieniona w jej trakcie.

Pełną listę opcji kontrolujących ustawienia regionalne można uzyskać wykonując poniższy kod:

```
proc options group=languagecontrol;
run;
```

3.1.1 Liczby

W większości przypadków sposób wyświetlania wartości liczbowych w SAS jest niezależny od aktualnych ustawień regionalnych. Wyjątek stanowi format ***nlnum*****,** który może używać różnych separatorów tysięcy oraz separatora dziesiętnego. W przypadku polskich ustawień regionalnych (***locale=polish***) używane są odpowiednio spacja oraz przecinek (np. -1 234,56). Dla ustawień amerykańskich jest to przecinek i kropka (np. -1,234.56). Obrazuje to poniższy przykład.

```
option locale=english;
data _null_;
    l = -1234.56;
    put l nlnum12.2;
run;

option locale=polish;
data _null_;
    l = -1234.56;
    put l nlnum12.2;
run;

Wynik:
   -1,234.56
   -1 234,56
```

Poniżej wymienione zostały przykładowe formaty zapisu liczb niezależne od regionu.

```
data _null_;
    l=-1234.56;
    put l;              *bez formatu;
    put l 12.2;         *format domyślny;
    put l z12.2;        *zera wiodące;
    put l numx12.2;     *przecinek separatorem dziesiętnym;
    put l comma12.2;    *przecinek pomiędzy tysiącami;
    put l commax12.2;   *kropka pomiędzy tysiącami;
run;
Wynik:
-3384.7
    -3384.70
-00003384.70
    -3384,70
   -3,384.70
   -3.384,70
```

3.1.2 Waluty

W przypadku walut, ustawienia regionalne (opcja ***dflang***) określają domyślną walutę, pozycję symbolu waluty (przed czy za kwotą) oraz znak separatora dziesiętnego (przecinek czy kropka).

```
options locale=english;
data _null_;
    kwota = 1.5;
    put kwota nlmny10.2;
    put kwota nlmnlpln10.2;
run;

Wynik:

    $1.50
   zł1.50

options locale=polish;
data _null_;
    kwota = 1.5;
    put kwota nlmny10.2;
    put kwota nlmnlusd10.2;
run;

Wynik:

   1,50 zł
   1,50 zł
```

Przeliczanie walut

SAS umożliwia przeliczanie walut przy pomocy formatów. Formaty te mają postać *<waluta_1><kierunek zmiany><waluta_2>,* np. *eurtoplz* (przeliczanie euro na złotówki) lub usdfreur (przelicznie euro na dolary). Przeliczanie to następuje w oparciu o zawartość tabeli wskazywanej przez **filename eurfrtbl,** bądź , jeżeli taka tabela nie została wskazana, w oparciu o wartości domyśle.

Działanie przeliczania walut przy użyciu formatów pokazuje poniższy przykład.

```
data _null_;
    kwota = 10;
    put '10 euro to ' kwota eurtoplz10.2 'zł';
    put '10 zł to ' kwota eurfrplz10.2 'euro';
run;

Wynik:

10 euro to      42.00zł
10 zł to      €2,38euro

*zmiana przelicznika złotówek na euro

filename eurfrtbl 'C:\kursy.txt';
data _null_;
    input;
    file eurfrtbl;
    put _infile_;
    cards;
    eurfrplz=4.0;  *wartość przykładowa
run;

data _null_;
    kwota = 10;
    put '10 euro to ' kwota eurtoplz10.2 ' zł';
    put '10 zł to ' kwota eurfrplz10.2 ' euro';
    *to samo można osiągnąć przy użyciu funkcji;
    nowa_kwota = eurocurr(10, 'plz', 'eur');
    put '10 zł to ' nowa_kwota 10.2 ' euro';
run;

Wynik:

10 euro to      40.00 zł
10 zł to      €2,50 euro
10 zł to       2.50 euro
```

3.1.3 Data i czas

Data

Podobnie jak w przypadku liczb, ustawienia regionalne wpływają jedynie na wybrane formaty. W przypadku dat jest to format ***nldate*** oraz formaty z rodziny ***eurdf****. Poniżej zamieszczone zostały przykładowe formaty. Można je uzyskać uruchamiając poniższy kod.

```
data _null_;
    data='1jan2015'd;
    put "date     " data date.;
    put "nldate   " data nldate.;
    put "poldfde  " data poldfde.;
    put "poldfdt  " data poldfdt.;
    put "poldfdd  " data poldfdd.;
    put "poldfdwn " data poldfdwn.;
    put "poldfmn  " data poldfmn.;
    put "poldfmy  " data poldfmy.;
    put "poldfwkx " data poldfwkx.;
    put "poldfdn  " data poldfdn.;
    put "poldfwdx " data poldfwdx.;
run;

Wynik:

date     01JAN15
nldate   01 stycznia 2015
poldfde  01sty15
poldfdt  01sty60:05:34:49
poldfdd  01-01-15
poldfdwn     czwartek
poldfmn       styczeń
poldfmy  sty15
poldfwkx           czwartek, 1 styczeń 2015
poldfdn  4
poldfwdx       1 styczeń 2015
```

W SAS 9.4 w formacie nldate zmieniony został przypadek, w którym podawana nazwa miesiąca. Obecnie jest to dopełniacz (np. 1 stycznia 2015) zamiast dotychczasowego mianownika (np. 1 styczeń 2015).

Data i czas

W przypadku danych czasowych dostępne są następujące formaty:

```
data _null_;
    czas='1jan2015 00:00:00'dt;
    put "time     " czas time.;
    put "datetime " czas datetime20.;
    put "nldatm   " czas nldatm40.;
    put "nltime   " czas nltime40.;
    put "nltimap  " czas nltimap40.;
run;

Wynik:

time       482136
datetime   01JAN2015:00:00:00
nldatm   01 stycznia 2015 00:00:00
nltime   00:00:00
nltimap  12:00:00 AM
```

Strefy czasowe (SAS 9.4)

Od wersji 9.4 SAS posiada wsparcie dla stref czasowych. Obejmuje ono możliwość definiowania aktualnej strefy czasowej, doboru strefy czasowej na podstawie lokalizacji geograficznej.

Informacje o aktualnej strefie czasowej przechowywane są w opcji **TIMEZONE**. Wartością odpowiadającą polskiej strefie czasowej jest **Europe/Warsaw** lub **CET** (czas zimowy) bądź **CEST** (czas letni).

Do pobrania informacji o aktualnie zdefiniowanej służą następujące funkcje:

- TZONEID() – zwraca identyfikator aktualnej strefy czasowej
- TZONENAME() – zwraca nazwę aktualnej strefy czasowej

Dodatkowo możliwa jest konwersja wartości czasu pomiędzy czasem uniwersalnym (UTC) a aktualną strefą czasową. Służą do tego funkcje:

- TZONES2U – konwertuje wartość czasową do czasu uniwersalnego (UTC).
- TZONEU2S – konwertuje wartość czasową z czasu uniwersalnego (UTC) do aktualnej strefy czasowej.

```
options timezone=cest;    /*ustawiamy czas środkowoeuropejski letni*/

data _null_;
    dt   = datetime();    /*aktualna data i czas*/
    tzid = tzoneid();     /*identyfikator strefy czasowej*/
    tzname = tzonename(); /*nazwa strefy czasowej*/
    udt = tzones2u(dt);   /*konwersja aktualnego czasu do UTC*/
    dt2 = tzoneu2s(udt);  /*i z powrotem*/

    put tzid ;
    put tzname;
    put dt  datetime.;
    put udt datetime.;
    put dt2 datetime.;
run;

Wynik:

ETC/GMT-2
CEST
18JUL13:09:34:55
18JUL13:07:34:55
18JUL13:09:34:55
```

Do wypisywania wartości czasowej razem z informacją o strefie można posłużyć się poniższymi formatami:

- B8601LX – format dla wartości datetime; wypisuje aktualną datę i czas bez wprowadzania zmian
- B8601DX – format dla wartości datetime; zakłada, że data jest w formacie UTC i konwertuje ją do aktualnej strefy czasowej
- B8601TX – format dla wartości time; zakłada, że godzina jest w formacie UTC i konwertuje ją do aktualnej strefy czasowej

Przykład użycia:

```
data _null_;
    dt = datetime();
    put dt datetime.;
    put dt e8601lx.;
    put dt e8601dx.;

    t = time();
    put t time.;
    put t e8601tx.;
run;

Wynik:

18JUL13:11:21:58
 2013-07-18T11:21:58+02:00
 2013-07-18T13:21:58+02:00
11:21:58
13:21:58+02:00
```

UWAGA. Pełna lista formatów wspierających strefy czasowe znajduje się w dokumencie SAS(R) 9.4 Formats and Informats: Reference.

3.1.4 Modyfikacja ustawień reginalnych (SAS 9.4)

W SAS 9.4 wprowadzona została możliwość modyfikacji ustawień regionalnych dla danego regionu. Służy do tego procedura ***LOCALEDATA***. Pozwala ona modyfikować m.in. takie aspekty jak:

- Nazwy miesięcy
- Nazwy dni tygodnia
- Formaty zapisu daty i czasu
- Formaty zapisu waluty
- Domyślne formaty papieru

Przykład użycia:

```
proc localedata;

  load saslocale;     /* ładowanie ustawień regionalnych */
                      /* modyfikacja ustawień */
  modify key=MON1 value='styczeń';
  modify key=MON2 value='luty';
  modify key=MON3 value='marzec';
  modify key=MON4 value='kwiecien';
...
  modify key=MON12 value='grudzień';

/* zapisanie wyników - nadpisywana jest poprzednia definicja */
  save registry / _all_ syntax=sas;

  /* podgląd nowych ustawień */
  contents _all_;

quit;
```

Powyższy kod pozwala m.in. na przywrócenie starego sposobu zapisu miesiąca w formacie NLDATE. Po jego wykonaniu format ten zwróci następujący wynik:

```
data _null_;
    data='1jan2015'd;
    put data nldate.;
run;

Wynik:

01 styczeń 2015
```

UWAGA. Aby dokonana modyfikacja została zaczytana, należy uruchomić SAS z parametrem -LOCALEDATA REGISTRY. Dodatkowo, po wykonaniu procedury LOCALEDATA, wymagany jest restart sesji.

Dokładny opis działania procedury ***LOCALEDATA*** znajduje się w pomocy SAS-a oraz w dokumencie SAS(R) 9.4 National Language Support (NLS): Reference Guide.

3.2 Kodowanie znaków

Równie ważnym aspektem związanym z przetwarzaniem danych w języku polskim jest sposób zapisu znaków, zwany też kodowaniem znaków (ang. encoding). Decyduje o tym opcja ***encoding***.

UWAGA. Wartość opcji encoding jest ustalana przy powoływaniu sesji SAS i nie może być zmieniona w jej trakcie. Jeżeli opcja nie jest zdefiniowana w żadnym pliku konfiguracyjnym, jej wartość jest ustawiana w oparciu o ustawienia regionalne (opcja locale).

Wszystkie informacje przetwarzane przez komputer, mają postać liczby. Dotyczy to również danych tekstowych. Proces zamiany tekstu na jego reprezentację liczbową (i z powrotem) odbywa się przy pomocy tzw. stron kodowych (ang. codepage). Strony te zawierają zazwyczaj 256 pozycji, z których pierwsza połowa (pierwsze 128 pozycji) jest wspólna dla wszystkich stron, natomiast różnią się zawartością drugiej połowy. Kodowanie określa która strona kodowa zostanie użyta do odczytu tekstu.
Polskie znaki dostępne są w kodowaniach ISO 8859-2 (latin2) oraz Windows 1250 (wlatin2). Pierwszy format jest używany domyślnie na komputerach z systemem operacyjnym Linux / UNIX, natomiast drugi jest wykorzystywany na systemie Windows. Oba kodowania zawierają wszystkie polskie znaki, jednakże nie są one ze sobą kompatybilne (kody większości polskich znaków nie pokrywają się).

Opcja ***encoding*** wskazuje domyślne kodowanie znaków używanie przez SAS. Informacje o kodowaniu są zapisywane w nagłówku zbioru. Aby je odczytać należy użyć procedury **PROC DATASETS**. Przykładowy kod oraz wynik pokazuje poniższy przykład.

```
proc datasets library=sashelp;
    contents data=class;
quit;
```

Procedura DATASETS

Nazwa zbioru	SASHELP.CLASS	**Obserwacje**	19
Typ elementu	DATA	**Zmienne**	5
Mechanizm	V9	**Indeksy**	0
Utworzono	20.06.2013 06:29:59	**Długość obserwacji**	40
Ostatnio zmodyfikowano	20.06.2013 06:29:59	**Usuniętych obserwacji**	0
Zabezpieczenie		**Skompresowany**	Nie
Typ zbioru		**Posortowany**	Nie
Etykieta	Student Data		
Reprezentacja danych	WINDOWS_64		
Kodowanie	us-ascii ASCII (ANSI)		

Ustawienie globalnej opcji ***encoding*** można nadpisać, podając ją jawnie w instrukcji filename (opcja **encoding=**) jako opcję biblioteki (opcje **inencoding=** oraz **outencoding=**) bądź pojedynczego zbioru (opcja **encoding=**). Dzięki temu można czytać i zapisywać dane używając innego kodowania niż aktualne kodowanie sesji SAS.

SAS, gdy wykryje, że odczytywane dane pochodzą z innego systemu operacyjnego, bądź zostały utworzone przy użyciu innego kodowanie, dokonuje automatycznego przekodowania (ang. transcoding) danych. Mechanizm ten nazywa się **CEDA** (Cross Environment Data Exchange). Jego użycie powoduje niewielkie zwiększenie zużycia procesora podczas odczytu i zapisu danych. Dodatkowo jest ono sygnalizowane poniższym komunikatem w logu:

```
NOTE: Plik danych TEST.CLASS.DATA ma format innego systemu operacyjnego lub jego kodowanie
nie odpowiada kodowaniu sesji. Użyty będzie Cross Environment Data Access, co może wymagać
dodatkowych zasobów procesora i spowolnić pracę.
```

Oprócz tego, mechanizm CEDA ma następujące ograniczenia:

- Dane są przekodowywane w locie. Oznacza to że w przypadku wielokrotnego przetwarzania zbioru, musi on zostać przekodowany za każdym razem.
- Brak wsparcia dla widoków SAS, widoków SAS/ACCESS innych niż Oracle bądź Sybase.
- Brak wsparcia dla katalogów SAS-owych oraz skompilowanych DATA step-ów.
- Brak wsparcia dla wszystkich plików SAS-owych pochodzących z wersji starszych niż SAS 7.
- CEDA nie może być użyta do modyfikacji zbioru.
- Brak wsparcia dla indeksów.

Więcej informacji o mechanizmie CEDA znajduje się w dokumencie SAS®9.4 Language Reference: Concepts.

3.2.1 Unicode (UTF-8)

Jak wspomnieliśmy w poprzednim rozdziale, większość stron kodowych zawiera 256 znaków. Odpowiada to liczbie wartości, możliwych do przedstawienia przy użyciu 1 bajta. Ilość ta jest jednak niewystarczająca aby przedstawić wszystkie możliwe znaki. Powoduje to, że np. w jednej tabeli nie mogą znajdować się jednocześnie polskie i niemieckie litery.

Rozwiązaniem tego problemu jest kodowanie Unicode, które w założeniu ma zawierać wszystkie możliwe znaki ze wszystkich języków. SAS pozwala na jego użycie, jednak wymaga spełnienia szeregu warunków:

- SAS musi zostać uruchomiony w specjalnym trybie obsługi Unicode. Tryb ten jest dostępny z poziomy Menu Start w folderze SAS -> Additional Languages -> SAS 9.4 (Obsługa Unicode).
- Podczas przetwarzania tekstów należy używać specjalnych funkcji obsługujących Unicode. Wynika to z faktu, że standardowe funkcje zakładają że każdy znak jest reprezentowany przez 1 bajt danych, co nie jest prawdą w przypadku Unicode. Funkcje obsługujące Unicode wyróżnione zostały przez dodanie litery 'K' przed nazwą funkcji, którą zastępują. Przykładowe dostępne funkcje to:
  - KLENGTH
  - KLOWCASE
- Długości zmiennych tekstowych muszą uwzględniać fakt, że pojedyncza litera może zajmować więcej niż 1 bajt (do 4 bajtów). Przykładowo słowo **gżegżółka** w kodowaniu Unicode wymaga pola o długości co najmniej 13, mimo iż zawiera tylko 9 liter.

Więcej informacji na temat przetwarzania wielojęzycznych danych w SAS-ie w artykule Processing Multilingual Data with the SAS® 9.2 Unicode Server. Dokument został stworzony dla SAS 9.2, jednakże treści w nim zawarte są w dalszym ciągu aktualne.

3.2.2 Postać uniwersalna (SAS 9.4)

W SAS 9.4 dodana została procedura, pozwalająca na przekształcenie tekstu do postaci uniwersalnej. Postać uniwersalna charakteryzuję się tym, że tekst zapisany jest jedynie przy użyciu podstawowych znaków alfabetu łacińskiego (A-Z, a-z, 0-9, znaki specjalne). Konwersja dokonywana jest w następujący sposób:

- Usuwane są wszystkie akcenty z liter (np. 'ó' jest zamieniane na 'o')
- Litery, których nie można przedstawić za pomocą podstawowych znaków są reprezentowane ich kodem w UTF-8. Kody mogą być zapisane w jednym z trzech formatów:
  - ESC – Unicode Escape; np. "\u00cd"
  - NCR – Numeric Character Representation; np. "Å"
  - PAREN – Unicode Parenthesis Escape; np. "<u00c5>"

Poniżej przedstawiony został sposób, w jaki funkcja BASECHAR obsługuje polskie znaki:

```
data cities;length name$24 name_ascii$24;
  input name;
  name_ascii = basechar(name, "ESC"); put name_ascii=;
cards;
Żółć
Łódź
Gęś
Jaźń
run;

Wynik:

name_ascii=Zo\u0142c
name_ascii=\u0141odz
name_ascii=Ges
name_ascii=Jazn
```

3.3 Sortowanie

Za kolejność sortowania w SAS odpowiada opcja ***sortseq***. Domyślne sortowanie odbywa się według wartości liczbowych poszczególnych znaków. W rezultacie otrzymujemy następujący porządek sortowania - 0123…ABC…XYZabc…xyz. Metoda ta nie obsługuje jednak poprawnie polskich znaków. Aby to zmienić należy ustawić wartość opcji ***sortseq*** na ***polish*** bądź ***linguistic***.

3.3.1 Ustawianie kolejności sortowania

Kolejność sortowania można ustawić w następujących miejscach:

- Parametr uruchomienia / plik konfiguracyjny – zmienia domyślną kolejność sortowania dla wszystkich sesji SAS. Wartość ustawia się za pomocą opcji `-sortseq Polish`.
- Opcja sortseq sesji SAS – zmienia kolejność sortowania dla wszystkich operacji sortowania od momentu ustawienia do kolejnej zmiany wartości. Wartość opcji zmieniamy za pomocą polecenia `options sortseq=Polish;`.

- Opcja sortseq procedury ***sort*** – ustawia kolejność sortowania dla pojedynczej operacji. Przykładowy kod:

```
proc sort data=input out=input_sorted sortseq=polish;
    by nazwisko, imie;
run;
```

Jeżeli opcja ***sortseq*** została zdefiniowana w na różnych poziomach, priorytet jest następujący:

1. Opcja procedury SORT
2. Opcja sesji SAS
3. Domyślna wartość opcji.

Przykładowy wynik sortowania:

```
options sortseq=Polish;

data t;
    length dane $20;
    input dane @@;
    cards;
    zbir żuk Monika ą źdźbło maszyna Lew łopata Ą Ala analiza lwica Łódź Adam
;
run;

proc sort data=t;
    by dane;
    title1 'Zbiór po posortowaniu słownikowym';
run;

proc print;
run;

Wynik:
                                        Obs    dane

                                          1    analiza
                                          2    Adam
                                          3    Ala
                                          4    ą
                                          5    Ą
                                          6    lwica
                                          7    Lew
                                          8    łopata
                                          9    Łódź
                                         10    maszyna
                                         11    Monika
                                         12    zbir
                                         13    źdźbło
                                         14    żuk
```

Jak widać na powyższym przykładzie, wstawienie opcji ***sortseq*** na wartość ***Polish*** wprowadza następujący porządek sortowania: 0123…aAąĄbBcCćĆ…zZżŻźŹ.

3.3.2 Własny porządek sortowania

W przypadku konieczności zdefiniowania innej niż domyślna kolejności sortowania, można zastosować procedurę ***TRANTAB.*** Opis działania tej procedury wraz z przykładami użycia można znaleźć w dokumentacji programu SAS oraz w dokumencie SAS® 9.4 National Language Support (NLS): Reference Guide.

W celu ułatwienia tworzenia własnych kolejności sortowania stworzone zostało makro ***maketbl***. Makro to jest dostępne w postaci kodu źródłowego, do pobrania pod tym adresem. Po pobraniu pliku z makrem można z niego korzystać w następujący sposób:

```
%include "C:\maketbl.sas";    *załączamy definicję makra;

%maketbl(AaĄąBbCcĆćDdEeĘęFfGgHhIiJjKkLlŁłMmNnŃńOoÓóPpQqRrSsŚśTtUuVvWwXxYyZzŹźŻż, pldict);
*definiujemy nową tabelę translacji;

proc sort data=work.t out=t_sorted_2 sortseq=pldict; *przykład użycia nowej tabeli
translacji;
    BY dane;
run;
```

UWAGA. Jeśli chcemy, aby makro było dostępne automatycznie po uruchomieniu sesji SAS-a, należy plik z makrem umieścić w katalogu <SASHome>\SASFoundation\9.4\core\sasmacro.

Jeżeli chcemy udostępnić wszystkim użytkownikom własną tablicę sortowania (***trantab***), to należy ją umieścić w jednym z katalogów: ***SASHELP.HOST*** bądź ***SASHELP.LOCALE***. Domyślnie tablica translacji umieszczana jest w bibliotece SASUSER w katalogu Profile, który jest tworzony indywidualnie dla każdego użytkownika.

UWAGA. Przedstawiona metoda nie działa w przypadku sesji SAS pracujących w kodowaniu UTF-8.

3.3.3 Sortowanie lingwistyczne (sortseq=linguistic)

W SAS 9.2 wprowadzona została opcja sortowania lingwistycznego (***sortseq=linguistic***). Opcja ta, w odróżnieniu od standardowych metod sortowania (***sortseq=polish***) pozwala na modyfikację szeregu parametrów procesu sortowania. Najważniejsze z nich zostały wypisane poniżej.

- **sortowanie słownikowe** – w przypadku występowania kilku wariantów jednego słowa, różniących się wielkością liter, słowa te będą stanowiły jedną grupę w zbiorze posortowanym. Przykładowo, zbiór „Kat, kat, kot” po posortowaniu słownikowym będzie miał kolejność „kat, Kat, kot”, zamiast „kat, kot, Kat”, którą uzyskujemy przy sortowaniu z ***sortseq=Polish***.
- **sortowanie numeryczne (numeric_collation)** – proces sortowania jest w stanie wykryć sytuację, w której mamy do czynienia ze słowami zawierającymi w sobie numer, np. nazwy ulic: „Informatyczna 201”, „Informatyczna 21”, „Informatyczna 22”. Sortowanie LINGUISTIC może być ustawione w taki sposób, że takie słowa będą sortowane według kolejności numerycznej. Powyższy zbiór zamiast kolejności „Informatyczna_201, Informatyczna_21, Informatyczna_22” może więc po posortowaniu mieć kolejność „Informatyczna_21, Informatyczna_22, Informatyczna_201”. Aby włączyć sortowanie numeryczne, należy użyć instrukcji **sortseq=linguistic(numeric_collation=on).**
- **zamiana kolejności sortowania dużych i małych liter (case_first)**– standardowo kolejność sortowania ustawia małe litery przed ich wielkimi odpowiednikami (aAąĄbBcCćĆ…). Kolejność ta może zostać odwrócona przez użycie opcji **sortseq=linguistic(case_first=upper).**
- **ustalenie traktowania znaków innych niż litery i cyfry (alternate_handling)** – domyślnie różnice pomiędzy symbolami nie będącymi literami i cyframi (znaki przestankowe, spacje, symbole) są tak samo istotne jak różnice między literami. **sortseq=linguistic(alternate_handling=shifted)** powoduje, że różnice między tymi znakami są brane pod uwagę tylko wtedy, gdy nie ma różnic między literami i cyframi.

Szczegółowe informacje na temat sortowania słownikowego można znaleźć w dokumencie Linguistic Collation: Everyone Can Get What They Expect oraz w dokumentacji SAS.

UWAGA. Opcja sortseq=linguistic może być ustawiona jedynie na poziomie procedury SORT.

3.3.4 Znane problemy

Łączenie zbiorów posortowanych wg. różnych porządków:

W przypadku łączenia zbiorów o różnych porządkach sortowania może wystąpić następujący komunikat:

```
ERROR: Input data sets cannot be combined because they have different collating sequences
       (SORTSEQ).
```

Rozwiązaniem tego jest ponowne posortowanie łączonych zbiorów, tak aby wszystkie były posortowane według tego samego porządku. Gdy zmiana porządku sortowania zbioru jest niemożliwa, można zignorować błąd poprzez ustawienie opcji ***nobysorted*** dla sesji bądź użycie wartości ***notsorted*** w klauzuli ***by***. Metoda ta nie jest dostępna dla operacji ***MERGE*** oraz ***UPDATE***.

Sortowanie w kodowaniu UTF-8

Sesje SAS działające w kodowaniu Unicode nie odsługują wartości ***sortseq=polish*** ani kolejności zdefiniowanych przez użytkownika.

W rezultacie jedyną opcją sortowania polskiego jest ustawianie opcji ***sortseq=linguistic*** przy każdym uruchomieniu procedury **SORT**.

3.4 Nazwy tabel i kolumn

Nazwy tabel i kolumn w SAS muszą spełniać następujące warunki:

- Maksymalna długość nazwy to 32 znaki,
- Nazwa musi zawierać co najmniej 1 literę,
- Nazwa nie może zaczynać się od pustego znaku bądź kropki,
- Wielkie i małe litery są dopuszczalne, lecz nie są rozróżniane

Oprócz tego domyślnie nazwa może zawierać jedynie litery alfabetu łacińskiego (A-Z, a-z), cyfry oraz podkreślenia. Zestaw ten można jednak rozszerzyć przy użyciu opcji ***validmemname=extend*** (nazwy tabel i widoków) oraz ***validvarname=any*** (nazwy kolumn).

Po ustawieniu wymienionych opcji nazwy kolumn i tabel mogą zawierać polskie znaki, spacje oraz inne znaki specjalne. Nazwy zawierające takie znaki muszą być zapisane w postaci ***'<nazwa_zmiennej>'n***. Przykładowe użycie zostało przedstawione poniżej.

```
options validvarname=any;
options validmemname=extend;

data 'Śląskie'n;
    'Dąbrowa Górnicza'n = 124701;
run;
```

3.5 Przenoszenie danych

W poniższym rozdziale skupimy się na sposobach przenoszenia danych między różnymi środowiskami SAS. Najczęstsze źródła problemów to:

- Przenoszenie danych między różnymi systemami operacyjnymi (np. Linux <-> Windows).
- Przenoszenie danych w kodowaniu Unicode.
- Przenoszenie danych między różnymi wersjami SAS.

W zależności od różnic między środowiskiem źródłowym a docelowym, konieczne może być użycie różnych metod przenoszenia danych. Dostępne sposoby opisane zostały w poniższych podrozdziałach.

3.5.1 Procedura COPY

Jeżeli środowisko docelowe jest dostępne z poziomu środowiska źródłowego (np. znajdują się w tej samej podsieci), najlepszą metodą przenoszenia danych jest procedura ***COPY***. Pozwala ona na przekopiowanie zawartości biblioteki z jednej lokalizacji do drugiej. Dodatkowo w trakcie wykonywania procedury dokonuje ona reorganizacji zbiorów i katalogów. Dzięki temu rozmiar danych na środowisku docelowym jest często mniejszy od rozmiaru oryginału.

Jedynym ograniczeniem w stosowaniu tej metody jest fakt, że w chwili wykonania procedury zarówno lokalizacja źródłowa jak i docelowa muszą być widoczne i dostępne dla sesji SAS-owej.

```
proc copy in=source out=dest;
run;
```

Domyślnie procedura ***COPY*** zachowuje większość parametrów tabel źródłowych. Są to m.in.:

- Rozmiar buforów wejściowych i wyjściowych
- Kompresja tabeli
- Kodowanie znaków
- Informacja o systemie operacyjnym

W pewnych przypadkach (np. przy przenoszeniu danych z systemu Windows na Linux/Unix) może to powodować, że po przeniesieniu danych może być używany mechanizm CEDA. Aby tego uniknąć, należy uruchomić procedurę COPY z opcją ***NOCLONE***.

UWAGA. Przenoszenie danych przy pomocy PROC COPY powoduje przebudowę wszystkich indeksów utworzonych na przenoszonych tabelach.

3.5.2 Procedury CPORT oraz CIMPORT

Jeżeli nie ma możliwości zapewnienia jednoczesnego dostępu do danych na obu środowiskach, do przenoszenia danych można użyć procedur ***CPORT*** oraz ***CIMPORT***. Proces przenoszenia danych w tym przypadku przebiega następująco:

1. Na środowisku źródłowym dokonujemy eksportu zawartości biblioteki do pliku transportowego. W tym celu wykonujemy procedurę ***CPORT*** na wybranej bibliotece.

```
filename tranfile "C:\sas_transport.txt";
libname export "C:\SAS\data\biblioteka_01";

proc cport library=export file=tranfile memtype=all;
run;
```

2. Wygenerowany plik transportowy przenosimy na środowisko docelowe.
3. Na środowisku docelowym, przy pomocy procedury CIMPORT importujemy zawartość pliku transportowego do biblioteki.

```
filename tranfile "C:\sas_transport.txt";
libname inport"/sasdata/biblioteka_01";

proc cimport library=inport file=tranfile memtype=all;
run;
```

Główną zaletą procedur CPORT oraz CIMPORT jest możliwość przenoszenia danych pomiędzy różnymi wersjami SAS-a oraz pomiędzy różnymi systemami operacyjnymi. Procedury te mają jednak kilka ograniczeń. Do najważniejszych należą:

- Utrata precyzji w przypadku przenoszenia bardzo dużych liczb bądź liczb z dużą ilości pozycji po przecinku. Przyczyny takiego zachowania zostały opisany tutaj.
- Brak wsparcia dla skompilowanych makroprogramów.
- Brak możliwości przenoszenia danych między kodowaniem DBCS (np. UTF-8) a SBCS (np. wlatin2). Próba przeniesienia danych kończy się w takim przypadku poniższym błędem o niekompatybilności pliku transportowego i aktualnego kodowania sesji. Przykładowy komunikat błędu wygląda następująco:

```
ERROR: The transport file was created in the encoding wlatin2.  This session encoding,
utf-8, can read and create transport files in the encoding, utf-8.  Data not imported.
```

- Przy przenoszeniu danych pomiędzy sesjami SBCS o różnych kodowaniach, część znaków może nie zostać przeniesiona poprawnie. Wynika to zazwyczaj z faktu że dany znak w kodowaniu źródłowym nie ma swojego odpowiednika w kodowaniu docelowym. Wystąpienie takiej sytuacji jest sygnalizowanie ostrzeżeniem w logu. Przykładowy komunikat wygląda nastęująco:

```
WARNING: The transport file was created in the encoding wlatin2.  This session encoding,
         Wlatin1, can read and create transport files in the encoding wlatin1. Data
         might not be imported correctly.
```

UWAGA. Jeżeli przenosimy dane z środowiska z SAS 9.1 lub starszego, ostrzeżenia o różnicach w kodowaniu nie są generowane. Wynika to z faktu, że informacja o kodowaniu sesji źródłowej jest zapisywana dopiero od wersji 9.2. Jeżeli plik transportowy został utworzony w sesji o kodowaniu UTF-8, podczas importu należy użyć procedury CIMPORT z opcją ISFILEUTF8=yes.

3.5.3 Ręczne przekodowanie danych

Najtrudniejszy przypadek przenoszenia danych SAS-owych ma miejsce gdy jedno środowisko działa w sesji SBCS (np. wlatin2) a drugie w sesji DBCS (UTF-8). Sytuacja taka nie jest obsługiwana ani przez procedurę COPY ani przez procedury CPORT oraz CIMPORT.

Rozwiązaniem jest ręczne przekodowanie danych. W tym celu należy użyć poniższego makra:

```
%macro transcode(input_table, output_table, inenc=utf-8, outenc=wlatin2, vars=_character_,
verify=yes);
  proc delete data=&output_table;
  run;

  data &output_table ;
    set &input_table (encoding=binary)  ;
    array _c_ $ &vars ;
    length __temp __temp2  $ 32767;
    do __i = 1 to dim(_c_);
      __temp2 = kcvt(_c_[__i], "&inenc.", "&outenc.") ;

      if __temp2 ne _c_[__i] then do ;
        _c_[__i] = __temp2 ;
      end;

    end ;
    drop __temp __temp2 __i ;
  run;

  %if %upcase(&verify) EQ %str(YES) %then %do ;
    proc compare base=&dsn(encoding=binary) compare=&out ;
    run;
  %end ;
%mend;
```

Parametry wywołania powyższego makra są następujące:

- input_table – tabela, którą chcemy przekodować
- output_table – wyjściowa tabela
- inenc – kodowanie tabeli wejściowej
- outenc – docelowe kodowanie tabeli
- vars – kolumny, które chcemy przekodować (domyślnie wszystkie tekstowe)
- verify – włącza bądź wyłącza sprawdzanie różnic po przekodowaniu

Jeżeli dany znak nie ma swojego odpowiednika w kodowaniu docelowym, jego wartość po przekodowaniu może być inna niż w oryginale. W tym przypadku konieczne może się okazać dodatkowa ręczna korekcja wartości znaku, np. przy pomocy funkcji transtrn.

Zbiory po przekodowaniu można przenieść stosując jedną z metod opisanych wcześniej.

3.5.4 Przenoszenie tabel z dużą ilością obserwacji (SAS 9.4)

W SAS 9.3 wprowadzono obsługę rozszerzonego licznika obserwacji (opcja ***EXTENDOBSCOUNTER*** albo ***EOS***). Jego użycie pozwala na obsługę tabel zawierających więcej niż 2 miliardy obserwacji. W SAS 9.4 opcja ta jest domyślnie włączona. Powoduje to błąd przy próbie przeniesienia danych z SAS 9.4 do SAS 9.2.

Aby tego uniknąć, przed przenoszeniem danych należy wyłączyć wspomnianą opcję i utworzyć kopię zbioru(ów), który będzie przenoszony.

4 Wykresy i raporty

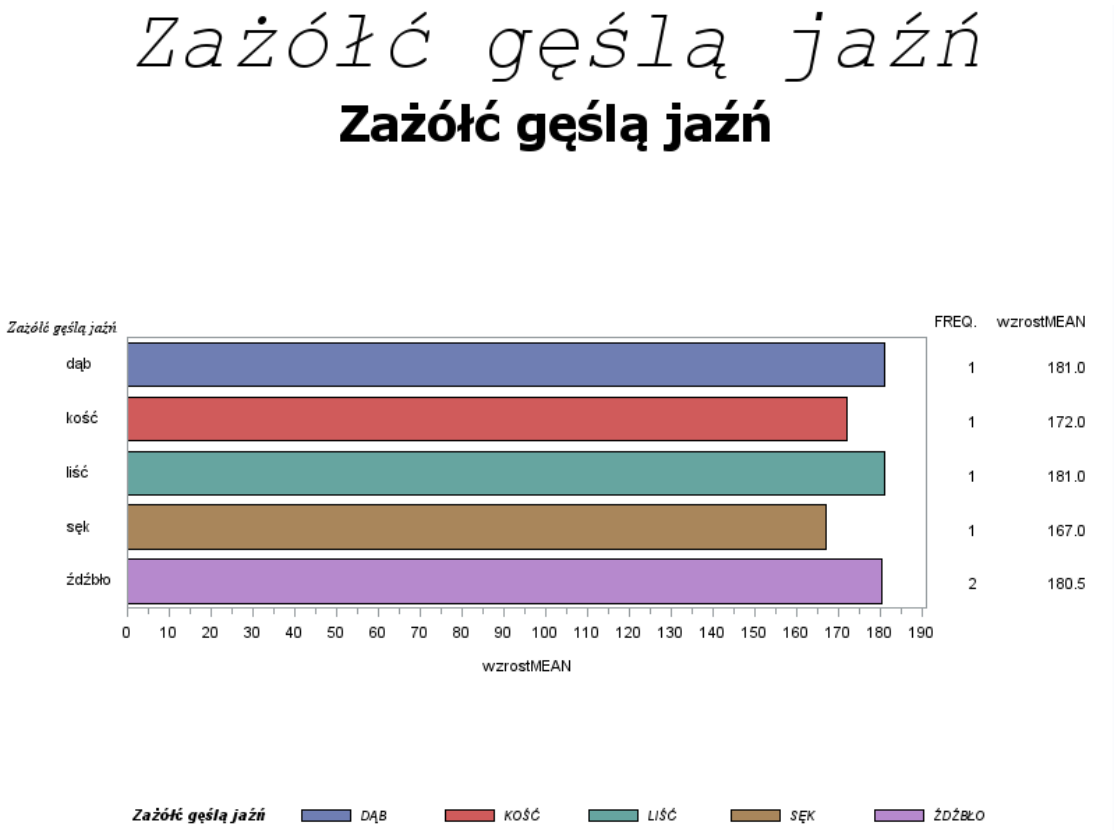
4.1 Czcionki systemowe (HARDWARE FONTS)

Przy generowaniu grafiki można korzystać z czcionek TrueType dostępnych w systemie operacyjnym Windows. Jeżeli korzystamy z polskiej wersji Windows wystarczy używać nazw czcionek jako wartości odpowiednich opcji. Dodając do nazwy czcionki /bo, uzyskamy pogrubienie, a /it - czcionkę pochyłą.

Przykładowy program demonstrujący użycie czcionek:

```
data klasa;
  length wzrost 8 kategoria $ 10;
  input  wzrost kategoria $;
  kategoria_u=upcase(kategoria);  kategoria_l=lowcase(kategoria);
  label kategoria_u='Kat. ąęćłńóśźż'   kategoria_l='Kat. ĄĘĆŁŃÓŚŹŻ';
cards;
181     Liść
167     Sęk
180     Źdźbło
181     Źdźbło
172     Kość
181     Dąb
;
run;

ods pdf file='C:\report_sas.pdf';

goptions reset=all;
title1 f='Courier New/it' h=10 pct 'Zażółć gęślą jaźń';
title2 f='Tahoma/bo' h=6 pct 'Zażółć gęślą jaźń';
legend1 label=(font='Arial/bo/it' 'Zażółć gęślą jaźń')
        value=(font='Arial/it/' height=1.5 pct);
axis1   label=(font='Times New Roman/bo/it' 'Zażółć gęślą jaźń');

goptions ftext='Arial';

proc gchart data= klasa;
  hbar kategoria_l /type=mean sumvar=wzrost
        maxis=axis1
        subgroup=kategoria_u legend=legend1;
  run;
quit;

ods pdf close;
```

W wyniku wykonania powyższego kodu generowany plik pdf zawierający poniższy wykres.



W analogiczny sposób, zmieniając jedynie format w instrukcji **ods**, możemy uzyskać wynik w postaci strony html, pliku rtf, e-book-a (format epub) czy prezentacji w PowerPoint. Szczegółowe informacje na ten temat można znaleźć w dokumencie SAS® 9.4 Output Delivery System: User's Guide.

4.2 Czcionki SAS/GRAPH (SOFTWARE FONTS)

Tworząc grafikę za pomocą SAS/GRAPH zamiast czcionek systemowych można wykorzystać czcionki dostarczane razem z tym modułem. Ich podstawową zaletą jest fakt, że są to wewnętrzne czcionki SAS, dzięki czemu nawet po przeniesieniu na inny system operacyjny wszystkie raporty będą wyglądać tak samo. Czcionki zawierające polskie znaki to **CENTX**, **SWISS** oraz **ZAPF**.

Warunkiem poprawnej obsługi polskich znaków przez te czcionki jest zdefiniowanie odpowiedniej tablicy przekodowań (keymapy) ze strony systemowej (Windows 1250) na wewnętrzną stronę kodową SAS/GRAPH. Tablicę przekodowań tworzymy wykonując poniższy kod.

```
/* generowanie keymapy dla WINDOWS */
libname gfont0 (sashelp);

proc gkeymap name=winansi out=temp keymap;
run;
data mod;
    retain tolen 1;
    from="d3"x; TO="08"x; output;  /* Ó */
    from="f3"x; TO="a2"x; output;  /* ó */
    from="f1"x; TO="c0"x; output;  /* ń */
    from="d1"x; TO="c1"x; output;  /* Ń */
    from="9c"x; TO="c2"x; output;  /* ś */
    from="8c"x; TO="c3"x; output;  /* Ś */
    from="9f"x; TO="c4"x; output;  /* ź */
    from="8f"x; TO="c5"x; output;  /* Ź */
    from="bf"x; TO="c6"x; output;  /* ż */
    from="af"x; TO="c7"x; output;  /* Ż */
    from="b3"x; TO="c8"x; output;  /* ł */
    from="a3"x; TO="c9"x; output;  /* Ł */
    from="b9"x; TO="ca"x; output;  /* ą */
    from="a5"x; TO="cb"x; output;  /* Ą */
    from="ea"x; TO="cc"x; output;  /* ę */
    from="ca"x; TO="cd"x; output;  /* Ę */
    from="c6"x; TO="f4"x; output;  /* Ć */
    from="e6"x; TO="f9"x; output;  /* ć */
run;
proc sort data=mod; by from;
run;
data temp;
  update temp mod; by from;
run;
proc gkeymap name=win1250 data=temp keymap;
run;
```

Dla wygody użytkowania, SAS dostarcza gotowe makro, które wykonuje powyższe operacje - *%lsgraph(wlt2)*.

Czcionek SAS/GRAPH używany w identyczny sposób jak czcionek systemowych. Poniżej zamieszony został fragment obrazujący przykładowe użycie wymienionych czcionek.

```
goptions reset=all;
libname gfont0 (sashelp);
goptions keymap=win1250;
proc gslide;
title1 f=swiss h=10 pct 'Zażółć gęślą jaźń';
```

4.3 Mapa Polski

Wśród standardowych zbiorów dostarczanych przez SAS w bibliotece MAPS znajduje się zbiór Poland, przedstawiający Polskę w podziale na powiaty. Oprócz tego zbioru w bibliotece MAPS znajduje się zbiór Poland2, zawierający powiązania powiatów z województwami. Aby otrzymać mapę Polski w podziale na województwa, należy najpierw połączyć zbiory Poland i Poland2, a następnie użyć procedury GREMOVE w celu usunięcia granic powiatów. Przedstawia to poniższy kod:

```
data poland_wojew;
    merge maps.poland maps.poland2;
    by id;
run;

proc sort data=poland_wojew;
    by wojid id;
run;

proc gremove data=poland_wojew out=poland_wojew;
    id id;
    by wojid;
run;
```

Aby wyświetlić mapę z województwami w SAS-ie można m.in. wykonać poniższy kod:

```
*zbior zawierajacy opisy poszczegolnych wojewodztw;
data dane;
    do i=1 to 16;
        wojid=i;
        value=i;
        output;
    end;
    drop i;
run;

goptions reset=all;

*wyswietlenie mapy w podziale na wojewodztwa;
proc gmap data=dane map=poland_wojew;
    id wojid;
    choro value / levels=16;
run;
quit;
```

Od wersji 9.3M2 SAS zawiera 3 domyślne biblioteki przechowujące mapy: MAPS, MAPSGFK oraz MAPSSAS.
Biblioteka MAPSGFK zawiera mapy w postaci cyfrowej dostarczone przez GfK GeoMarketing GmbH. MAPSSAS zawiera aktualne wersje tradycyjnych map SAS-owych.
Biblioteka MAPS może być tożsama z MAPSGFK bądź MAPSSAS. Domyślnie jest to MAPSSAS.

5 Zewnętrzne źródła danych

Jedną z najważniejszych funkcjonalności SAS jest możliwość importu i eksportu danych z wielu źródeł. W zależności od posiadanych modułów mogą to być:

- Pliki tekstowe
- Microsoft Excel - pliki xls oraz xlsx
- Microsoft Access – pliki mdb oraz accdb
- Bazy danych – Oracle, SQL Server, Teradata, ODBC i wiele innych.

Konfiguracja dostępu do zewnętrznych źródeł danych wymaga często wykonania dodatkowych kroków podczas instalacji SAS. Szczegółowe instrukcje na temat konfiguracji dostępu do danych z SAS znajdują się w dokumencie SAS i zewnętrzne źródła danych.

6 Konfiguracja serwerów obliczeniowych SAS

Konfiguracja serwerów obliczeniowych SAS przebiega w następujący sposób:

1. Zaczytywana jest konfiguracja domyślna SAS Foundation.
2. Aplikowane są opcje zdefiniowane na poziomie serwera aplikacji (SASApp). Opcje te zapisane są w plikach **sasv9.cfg** oraz **sasv9_usermods.cfg** znajdujące się w katalogu serwera aplikacji (np. **C:\SAS\Lev1\SASApp**).
3. Aplikowane są opcje zdefiniowane na poziomie serwera obliczeniowego (WorkspaceServer, StoredProcessServer, PooledWorkspaceServer, OLAPServer, ConnectServer itd.). Opcje te są zapisane w plikach **sasv9.cfg** oraz **sasv9_usermods.cfg** w katalogu serwera obliczeniowego (np. **C:\SAS\Lev1\SASApp\Workspace Server**).
4. Aplikowane są opcje podawane w skryptach bat (Windows) / sh (Linux/Unix) startujących serwer.
5. Aplikowane są opcje zdefiniowane we właściwościach serwera w metadanych.

Po wykonaniu powyższych czynności sesja obliczeniowa jest uruchamiana i rozpoczyna się wykonywanie kodów startowych. Domyślnie proces ten przebiega następująco:

1. Wykonywany jest kod zapisany w pliku **autoexec.sas** serwera obliczeniowego (np. **C:\SAS\Lev1\SASApp\WorkspaceServer\autoexec.sas**).
2. Wykonywany jest kod zapisany w plikach **appserver_autoexec.sas** oraz **appserver_autoexec_usermods.sas** znajdujących się w katalogu serwera aplikacji (np. **C:\SAS\Lev1\SASApp**).
3. Wykonywany jest kod zapisany w pliku **autoexec_usermods.sas** serwera obliczeniowego (np. **C:\SAS\Lev1\SASApp\WorkspaceServer\autoexec_usermods.sas**).
4. Wykonywany jest kod zapisany w pliku wskazanym jako **Initialization program** dla serwera. Plik ten można ustalić w SAS Management Console w opcjach serwera logicznego (np. **SASApp – Logical Workspace Server → Properties → Options → Set Server Properties → Initialization program**). Domyślnie opcja ta jest pusta.

UWAGA. Zmiana ustawień serwerów obliczeniowych SAS może spowodować, że niektóre obiekty, takie jak kostki OLAP czy mapy informacyjne, przestaną działać poprawnie. Aby tego uniknąć zaleca się ustawienie kodowania znaków zaraz po zakończeniu instalacji SAS-a, zanim jakiekolwiek obiekty zostaną utworzone. Dotyczy to w szczególności kodowania znaków (opcja encoding).

Miejsce, w którym dana opcja została zdefiniowana można sprawdzić procedurę ***OPTIONS*** z opcją ***value***. Poniżej przedstawione zostało przykładowe uruchomienie tej procedury.

```
proc options option=locale value;
run;

Wynik:

Option Value Information For SAS Option LOCALE
    Value: PL_PL
    Scope: SAS Session
    How option value set: Config File
    Config file name:
            C:\Program Files\SASHome\SASFoundation\9.4\nls\pl\sasv9.cfg
```