



StatSoft®

business intelligence • data mining • sterowanie jakością • analizy w środowisku internetowym

The *STATISTICA* Enterprise-wide SPC System (*SEWSS*):
Narzędzia analizy danych
dla systemów zbierania i przechowywania danych

Część 2: Interaktywne filtrowanie danych

Praktyczny materiał informacyjny StatSoft

StatSoft Polska Sp. z o.o. • ul. Kraszewskiego 36 • 30-110Kraków • tel. (12) 4284300 • Fax: (12) 4284301 • info@statsoft.pl • www.statsoft.pl

Australia: StatSoft Pacific Pty Ltd.
Brazil: StatSoft Brazil Ltda.
Czech Republic: StatSoft Czech Rep. s.r.o.
France: StatSoft France

Germany: StatSoft GmbH
Hungary: StatSoft Hungary Ltd.
Israel: StatSoft Israel Ltd.
Italy: StatSoft Italia srl

Japan: StatSoft Japan Inc.
Korea: StatSoft Korea
Netherlands: StatSoft Benelux BV
USA: **StatSoft, Inc.**

Portugal: StatSoft Iberica Ltda.
Russia: StatSoft Russia
Singapore: StatSoft Singapore
S. Africa: StatSoft S. Africa (Pty) Ltd.

Spain: StatSoft Espana
Sweden: StatSoft Scandinavia AB
Taiwan: StatSoft Taiwan
UK: StatSoft Ltd.

Wstęp

Firma StatSoft, producent zestawu programów **STATISTICA**, współpracuje ze swoimi klientami projektując i dostarczając im rozwiązania najwyższej jakości i niezawodności. **STATISTICA Enterprise-wide SPC System (SEWSS)** jest najbardziej zaawansowanym systemem analizy danych dla działów badań i rozwoju (R&D) jak i zapewnienia jakości. Z doświadczeń StatSoft, zebranych przy współpracy z klientami wynika, że głównymi warunkami stosowana analizy danych są:

- **Łatwe korzystanie z systemu.** Bez względu na to jak wielkie są możliwości narzędzia analizy danych, nie będzie ono używane przez użytkowników, jeśli nie jest intuicyjne w obsłudze.
- **Korzystanie z różnych źródeł danych.** Pracownicy, którzy chcą analizować dane i wiedzą jakie informacje są im potrzebne, często muszą poświęcić dużo czasu na współpracę z osobami, które dokładnie znają struktury baz danych i sposoby pobierania danych. System powinien zminimalizować czas i wysiłek potrzebny na pobranie danych, poprzez pełną integrację i przezroczystą współpracę z repozytoriami danych.
- **Współpraca.** Zadania działów badania i rozwoju (R&D) i zapewnienia jakości rzadko realizuje jedna osoba działająca samodzielnie. Najczęściej projekty takie wymagają wspólnej pracy wielu osób.

Biorąc pod uwagę powyższe fakty, firma StatSoft zaprojektowała system **SEWSS**, który wprowadzono do sprzedaży w 1997 r. Od tego czasu system jest ciągle rozwijany i udoskonalany.

SEWSS stanowi rozbudowany zestaw narzędzi data mining, analiz i wizualizacji danych, przyczym wszystkie one zintegrowane są w jeden system. Użytkownicy **SEWSS** korzystają ze środowisk dostosowanych do przypisanych im zadań (np. inżynier jakości). W systemie można definiować szablony analiz umożliwiające automatyzację rutynowych zadań. W ramach systemu, wielu użytkowników może wspólnie korzystać z danych, raportów i wyników analiz przy zachowaniu bezpieczeństwa informacji. Każdy pracownik, technik i analityk może potrzebować dostępu do danych i narzędzi analitycznych. **SEWSS** jest rozwiązaniem zaspokajającym te potrzeby.

Niniejszy artykuł jest drugim z serii trzech artykułów wprowadzających do kluczowych właściwości systemu **SEWSS**, ze szczególnym uwzględnieniem narzędzi współpracy z istniejącymi źródłami danych. Jeśli czytelnik zaznajomił się z treścią pierwszego artykułu “Część 1: Współpraca z bazami danych” zaleca się pominięcie rozdziału **Przeznaczenie SEWSS** i przejście bezpośrednio do rozdziału zawierającego przykład interaktywnego filtrowania danych.

Interaktywne filtrowanie pozwala obsłużyć wiele zadań analitycznych w obrębie jednego szablonu analizy, przez umożliwienie analizującemu określenia, które dane go interesują w momencie uruchamiania analizy. Pola w bazie danych, które mogą odnosić się do przedziału daty/czasu, numeru próbki czy też partii są wybierane ściśle dla konkretnych analiz.

Korzyści, jakie daje filtrowanie, to przede wszystkim zapewnienie organizacji możliwości pełnego i ciągłego wykorzystania wartościowych danych.

Przeznaczenie SEWSS

W wielu organizacjach bieżące wykorzystanie bogactwa zbieranych i przechowywanych danych nie jest optymalne. Powszechnie jest przeświadczenie, że potrzebny jest zintegrowany, kompleksowy system analizy danych i raportowania. Przeświadczeniu temu towarzyszy obawa, że opracowanie i wdrożenie takiego systemu byłoby kosztowne, a potencjalni użytkownicy musieliby zostać przeszkoleni, aby z niego skorzystać.

Wiele organizacji dysponuje bogactwem danych, a ich pracownicy mogliby korzystać z tych danych przy podejmowaniu decyzji. Niestety brak odpowiednich narzędzi (spełniających przedstawionej powyżej wymogi) oraz ograniczenia czasowe, uniemożliwiają czerpanie korzyści z analizy danych. Potrzebny jest zintegrowany, kompleksowy system dysponujący narzędziami budowy zapytań i pobierania danych, analizy i wizualizacji, a wszystko to w bezpiecznym, wieloużytkownikowym środowisku pracy zespołowej.

Wyobraźmy sobie system, w którym do uzyskania wyników analizy aktualnych danych wystarczy:

1. - być uprawnionym do korzystania z danej analizy
2. - kliknąć jej nazwę na hierarchicznie uporządkowanej liście dostępnych analiz.

Użytkownik nie musi wiedzieć, gdzie i w jaki sposób przechowywane są dane. Nie musi on także wiedzieć, że np. dane pochodzą z dwóch różnych źródeł, że specyfikacje produktu zostały automatycznie odczytane z odpowiedniej bazy danych, w momencie wybrania konkretnego produktu i specyfikacji. Wszystko, co powinien wiedzieć użytkownik, to jak nazwana została się analiza, którą chce wykonać.

Właśnie tak działa system SEWSS!

Wprowadzenie do SEWSS

SEWSS to wieloużytkownikowe oprogramowanie umożliwiające korzystanie z wielu różnorodnych funkcji. Najważniejsze elementy systemu **SEWSS** to:

- **Połączenia ze źródłami danych.** **SEWSS** konfigurowany jest tak, aby mógł korzystać z repozytoriów danych w celu analizy i wizualizacji danych. Każdemu szablonowi odpowiada źródło danych, a wymagane dane pobierane są automatycznie.
- **Szablony analizy.** W systemie **SEWSS** szablony (wzorce) analiz są przechowywane centralnie i mogą być uruchamiane automatycznie lub na żądanie użytkownika. Szablony analizy mogą mieć ustalone ustawione wszystkie parametry, albo użytkownik może częściej z nich określać przy uruchamianiu analizy (np. wybierać podzbiór analizowanych danych lub konkretną wersję analizy).
- **Interaktywne filtrowanie:** **SEWSS** zawiera niezawodne narzędzia filtrowania danych. Dzięki temu użytkownik nie musi poznawać złożonych struktur danych i języka SQL. Zamiast tego, użytkownik wybiera tylko odpowiednie wartości (np. nazwę produktu, zakres daty i czasu) i zadaje kryteria dla danych.
- **Raporty.** System **SEWSS** umożliwia korzystanie z modyfikowalnych szablonów raportów, w których możemy określić układ i format strony dla regularnie tworzonego raportu. W skład **SEWSS** wchodzi aplikacja automatycznie tworząca raporty, w zadanych odstępach czasu.
- **Bezpieczeństwo i uprawnienia.** W systemie wieloużytkownikowym nie powinno się każdego użytkownika traktować tak samo. Zadania i zakres odpowiedzialności są różne dla różnych osób. W systemie **SEWSS** zapisywane są uprawnienia do przeglądania i modyfikowania szablonów analiz oraz do korzystania z funkcji systemu. Łatwość korzystania z systemu zwiększa to, że użytkownik nie widzi obiektów, które nie są mu potrzebne.

Dzięki powyższym funkcjom system **SEWSS** dodaje do repozytoriów danych praktyczne możliwości analizy. **SEWSS** jest gotowym produktem i zawiera ogólne obiekty, które są konfigurowane za pomocą wbudowanego w system interfejsu administratora. Najważniejsze obiekty systemu opisano poniżej. Zapoznanie się z podstawowymi terminami **SEWSS** daje podstawę do zrozumienia przykładów przedstawionych niżej.

- **Połączenia.** W połączeniach zawarte są określenia źródeł danych i metod dostępu do nich.
- **Profile.** Profile to zbiory zapytań, które określają dane podlegające analizie. W profilach zapisywane są również meta-dane, umożliwiające odpowiednią interpretację danych podczas analizy (np. to, że w pewnej kolumnie zapisano oznaczenie czasu, a w innej nazwę produkty, względem której filtrujemy dane).
- **Monitory.** Monitory to szablony analiz, określające rodzaj analizy i jej parametry.
- **STATISTICA Visual Basic:** **SEWSS** jest zorientowaną obiektowo aplikacją wyposażoną w interfejs API realizowany przez język Visual Basic for Applications (VBA). Działaniem **SEWSS** możemy sterować nie tylko interakcyjnie, ale również za pośrednictwem ponad 11 000 funkcji modelu obiektowego **STATISTICA**.

Przykład: Interaktywne filtrowanie danych

W niniejszym rozdziale przedstawiamy przykład wykorzystania systemu **SEWSS** jako narzędzia analitycznego połączonego z repozytorium danych. Poniższy przykład został oparty na przykładzie zaprezentowanym w pierwszym artykule z serii, zatytułowanym: “Współpraca z bazami danych”. Zaleca się zapoznanie z tym przykładem, ponieważ zawiera on elementy które nie są zamieszczone poniżej.

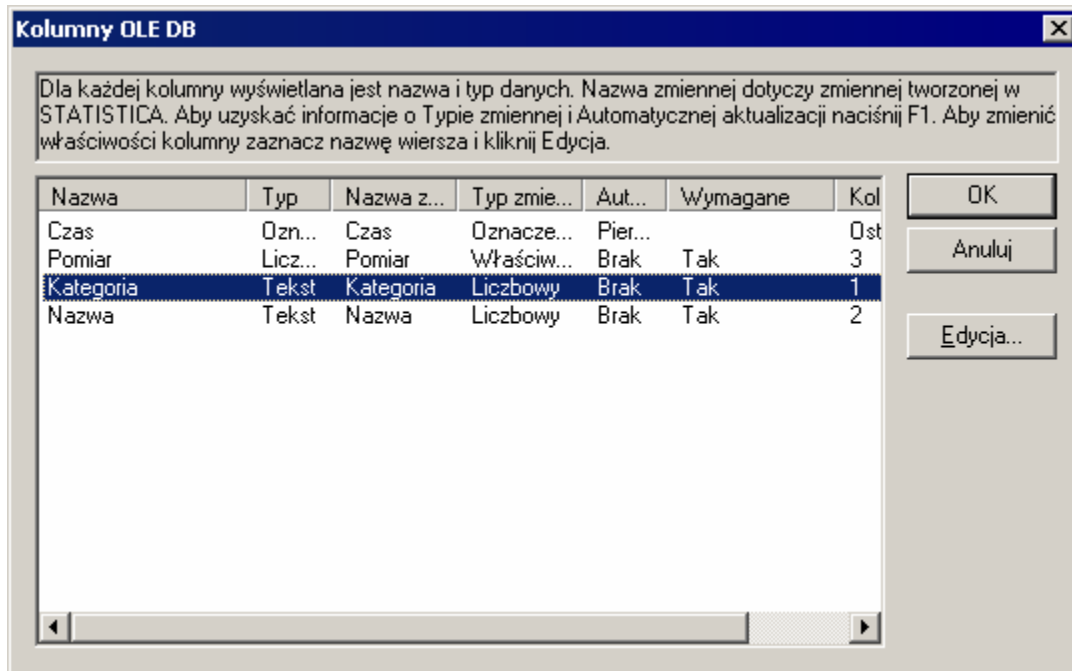
SEWSS w sposób elastyczny umożliwia użytkownikom systemu na interaktywne wybieranie danych istotnych dla konkretnych analiz. W trakcie tworzenia profili, administrator systemu decyduje, względem których pól będzie można filtrować dane. Dodatkowo udostępnionych jest kilka opcji odnoszących się do własności pól filtrujących.

Administrator systemu decyduje czy filtrowanie według zawartości danego pola jest wymagane czy opcjonalne. Dobrym przykładem użyteczności jest opcji obowiązkowego filtrowania, jest przypadek, gdy zapytanie profilu nie ogranicza zakresu czasu dla wyszukiwanych danych. W tym przypadku, ustawienie pola daty/czasu jako pola obowiązkowo filtrowanego uniemożliwia użytkownikowi wykonania pomyłkowego zapytania dotyczącego całości danych w bazie.

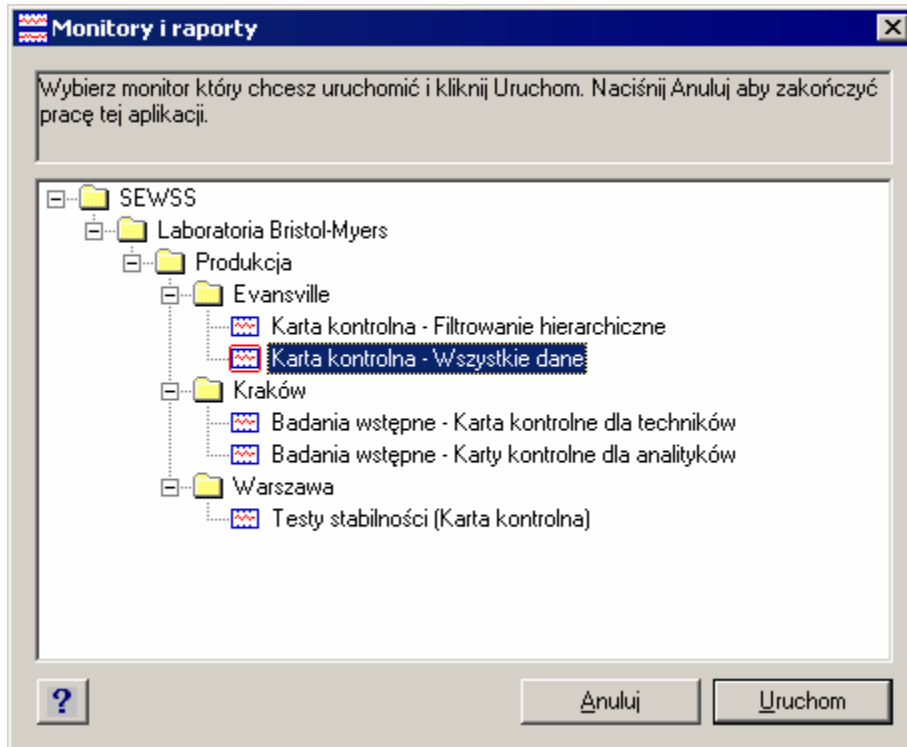
Dodatkową opcją jest możliwość ustalenia filtrowania jako “równoległego” lub “hierarchicznego”. Filtrowanie równoległe występuje w przypadku, gdy pola dopuszczające filtrowanie wyświetlone są łącznie, umożliwiając wybór interesujących wartości. W takim wypadku użytkownik ma możliwość wybrania na przykład: “Urządzenie=Urządzenie A” oraz “Technik=Grzegorz”. Filtrowanie równoległe ma zastosowanie, gdy pola dopuszczające filtrowanie nie mają żadnej naturalnej hierarchii, bo na przykład każdy technik może pracować na każdym urządzeniu.

W przypadku, gdy pola dopuszczające filtrowanie są zależne, administrator **SEWSS** może zdefiniować filtrowanie hierarchiczne. Przy filtrowaniu hierarchicznym pola dopuszczające filtrowanie są wyświetlane w kolejności zdefiniowanej w trakcie tworzenia profilu, na którym będzie oparty szablonu analizy. Późniejsze wartości pól dopuszczających filtrowanie są ograniczone wcześniejszymi wyborami. Efekty uzyskane w wyniku podejścia hierarchicznego najlepiej zilustrować przykładem. Organizacja, która produkuje wiele produktów, ma zdefiniowane ich kategorie i określone badania dla określonych produktów. W tym przypadku, filtrowanie hierarchiczne powinno pozwolić użytkownikowi wybrać najpierw kategorię produktu, następnie określony produkt tej kategorii i dopiero po tym typ badania.

Poniższy przykład pokazuje logikę działania i możliwości **SEWSS**, dla przypadku poruszonego powyżej. W *Kreatorze profilu*, gdy tylko zostanie zdefiniowane zapytanie, administrator systemu ma możliwość zdefiniowania metadanych, które będą przechowywane w systemie, opisujących właściwości tego zapytania. W **SEWSS** właściwości kolumn można ustalać w oknie *Edycja właściwości zmiennej*.

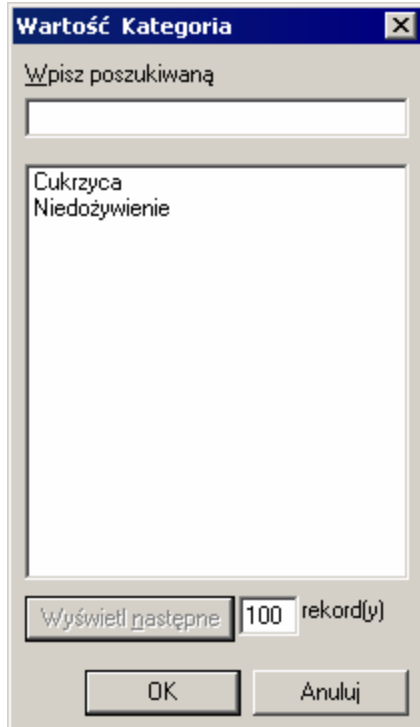


W powyższym interfejsie, najistotniejszymi właściwościami są *Wymagane* oraz *Kolejność*. *Wymagane* oznacza, że użytkownik analizy bazującej na tym profilu musi wybrać przynajmniej jedną wartość z oznaczonego w ten sposób pola. *Kolejność* jest właściwością specyficzną dla filtrowania hierarchicznego, definiuje ona kolejność z jaką wyświetlane są prośby wyboru kolejnych elementów w hierarchii. Wszystkie elementy widoczne powyżej są polami wymaganymi i będą prezentowane w kolejności określonej w kolumnie *Kolejność*. Użytkownik wykonujący analizę proszony będzie o dokonywanie wyborów w ustalonej kolejności.



Na początku użytkownik wybiera żadaną analizę. W tym wypadku, zaznaczony monitor pozwala na wybranie wszystkich danych testowych, dla wszystkich produktów. Wybór danych testowych, rodzaju produktu i przedziału czasowego zależy od potrzeb indywidualnego użytkownika. Taka możliwość sprawia, że system jest bardzo elastyczny; pojedynczy szablon analizy może obsłużyć wiele różnych potrzeb analitycznych.

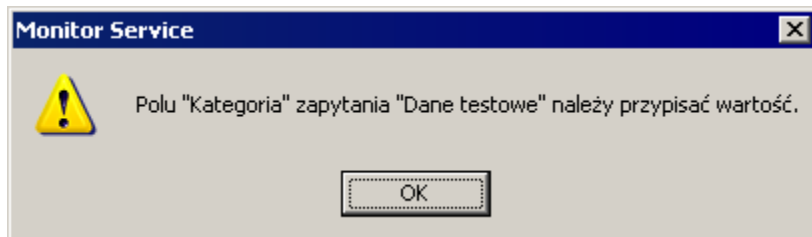
Zaraz po uruchomieniu analizy, wyświetlona zostaje opcja wyboru kategorii produktu, tak jak to przedstawiono poniżej:



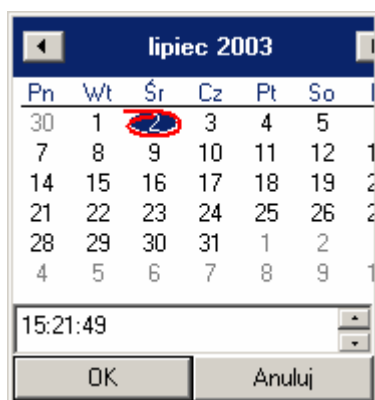
Po wybraniu kategorii, użytkownik ma do wyboru należące do niej produkty.



Należy zauważyć, że niektóre pola w profilu zostały zdefiniowane jako wymagane podczas filtrowania. Jeśli użytkownik próbuje uruchomić analizę bez wcześniejszego wyboru wartości dla wszystkich pól, dla których filtrowanie jest wymagane, wtedy program wyświetli komunikat:



Dla wyboru daty/czasu, wyświetlany jest łatwy w użyciu interfejs.



Po wybraniu zakresu daty i czasu, zostaje uruchomiona analiza i otrzymujemy wyniki. Hierarchiczna selekcja zapobiega wyborowi nieodpowiednich kombinacji. Wymagane filtrowanie zapobiega omyłkowemu uruchamianiu analizy dla całego zbioru danych (tj. bez określenia zakresu czasowego) lub w poprzek niejednorodnego zbioru danych (np. łącznie dla zupełnie różnych produktów).

Podsumowanie

SEWSS dostarcza potężnej, elastycznej i łatwej w użyciu platformy programowej dla potrzeb data mining, analizy i prezentacji wyników dla całej organizacji. Ten artykuł zawarł wstępne przedstawienie cech i zalet **SEWSS**, na przykładzie jego zdolności do interakcyjnego filtrowania. Interakcyjne filtrowanie pozwala obsłużyć wiele zadań analitycznych w obrębie jednego szablonu analizy, przez umożliwienie analizującemu definiowania interesujących danych w momencie uruchomienia analizy. Dzięki łatwym w użyciu interfejsom, nie wymagającym od użytkownika znajomości języka SQL, **SEWSS** dostarcza wielu użytecznych możliwości filtrowania.

W celu bardziej szczegółowego zapoznania się z możliwościami **SEWSS** lub urządzenia indywidualnej prezentacji systemu, dostosowanej do konkretnych potrzeb, prosimy o kontakt ze StatSoft Polska (tel. 012-4284300; e-mail: info@statsoft.pl).

Niniejszy dokument: “Część 2: Interaktywne filtrowanie danych“, jest drugim z grupy dokumentów prezentujących możliwości SEWSS dotyczące współpracy z bazami danych. Kolejnych informacji dostarczają powiązane z niniejszym dokumenty zatytułowane: “Część 1: Współpraca z bazami danych” oraz “Część 3: Przypisywanie specyfikacji”.