



## ORGANIZACJA GROMADZENIA I ANALIZY DANYCH ORAZ TWORZENIA RAPORTÓW DLA POTRZEB ADMINISTRACJI

*Tomasz Demski  
StatSoft Polska Sp. z o.o.*

### **Gromadzenie danych i ich analiza – podstawowe wymagania**

Działalność organów administracji rządowej i samorządowej wiąże się z potrzebą gromadzenia i przetwarzania danych. Wynika to z wymogów prawne, konieczności informowania obywateli (np. o tym w jaki sposób wydawane są ich pieniądze), a także z potrzeby tworzenia strategii rozwoju i monitorowania realizacji zadań. Można powiedzieć, że warunkiem dobrego działania demokratycznego państwa jest z jednej strony gromadzenie przez administrację publiczną informacji o obywatelach (jakie są ich potrzeby, bolączki itd.), a z drugiej strony informowanie opinii publicznej o działaniu administracji publicznej. Również, jeżeli jakkolwiek organizacja (także organ administracji publicznej) pragnie poprawiać jakość swojego działania (np. poprzez wdrożenie TQM lub norm ISO 9000) musi przynajmniej wiedzieć w jakim stanie jest obecnie i do jakiego stanu powinna dążyć: wiedzy tej nie można uzyskać bez regularnego zbierania i przechowywania danych oraz ich analizy.

Przy obecnym poziomie informatyki samo gromadzenie danych (nawet bardzo dużych ilości) w zasadzie nie stanowi problemu. Największe kłopoty pojawiają się na etapie wyciągania wniosków ze zgromadzonej informacji – wystarczy powiedzieć, że wg badań IBM tylko 7% danych jest w jakikolwiek sposób analizowane. Na pierwszy rzut oka wydaje się to wręcz niemożliwe: wydawane są ogromne pieniądze na zdobywanie i przechowywanie danych, z których prawie się nie korzysta. Jeżeli jednak przyjrzymy się bliżej całej sprawie, to dostrzeżemy kilka zasadniczych przyczyn utrudniających przetworzenie zgromadzonych danych na użyteczną wiedzę.

- ◆ dotarcie do potrzebnych danych jest utrudnione ze względu na ich rozproszenie i dużą ilość,
- ◆ dane są niejednolite (np. w różnych działach urzędu stosuje się różny sposób zapisu tych samych informacji),
- ◆ dane przechowywane są w różnych formatach i przed ich analizą najpierw należy je odczytać, co wymaga odpowiedniej wiedzy i czasami jest skomplikowane,
- ◆ do uzyskania wartościowych wyników często trzeba stosować złożone metody analityczne,
- ◆ poszczególne etapy wnioskowania (tj. pobranie danych, wykonanie analizy, przygotowanie raportu i udostępnienie go zainteresowanym) realizują zazwyczaj różne aplikacje dostarczane przez różnych producentów (co często oznacza, że ich współpraca też układa się „różnie”).

Warto zwrócić uwagę, że to co pozornie może wydawać się najtrudniejsze, tzn. samo wykonanie obliczeń i graficzna prezentacja wyników najczęściej nie nastęrcza już trudności – tym bardziej oczywiste i dotkliwe stają się przedstawione wyżej problemy.



Rozwiązaniem części problemów jest zbudowanie hurtowni danych, tj. składnicy danych, w której przechowywane są ujednolicone, sprawdzone i często wstępnie przetworzone dane, potrzebne we wspomaganie podejmowania decyzji. Bardzo ważną korzyścią płynącą z zastosowania hurtowni danych jest oddzielenie analitycznej bazy danych od bazy operacyjnej.

Hurtownia danych zawiera potrzebne dane, ale pozostaje problem ich analizy. W związku z tym pojawia się wymaganie, aby narzędzia analityczne i służące do komunikacji z hurtownią danych dobrze ze sobą współpracowały, najlepiej tak aby, osoba analizująca dane w ogóle nie musiała wiedzieć skąd one są pobierane i w jaki sposób.

Po wykonaniu analizy pojawia się kolejny problem: jak udostępnić jej wyniki wszystkim zainteresowanym? Można oczywiście rozesłać raport zainteresowanym osobom, ale jeżeli takich osób jest więcej to może się okazać, że analityk przede wszystkim zajmuje się rozsyłaniem raportów, a nie wykonywaniem analiz. Poza tym żyjemy przecież w dobie Internetu i Intranetu: jeżeli za pomocą sieci można zamówić pizzę, to dlaczego by nie można zamówić prognozy przychodów na najbliższe 3 miesiące? Warto zauważyć, że przynajmniej część informacji i raportów uzyskiwanych w wyniku analiz powinna być udostępniana dosłownie każdemu obywatelowi i tutaj doskonałym medium przekazu informacji będzie Internet. Oczywistym wymaganiem jest więc łatwe, a najlepiej w pełni automatyczne, publikowanie wyników w Internecie.

Odbiorcy raportów w postaci elektronicznej często chcieliby mieć możliwość ich modyfikacji: od stosunkowo niewielkich (np. zmiana skalowania wykresu) do bardzo znaczących (np. uzyskanie przygotowanego wcześniej raportu, ale tylko dla wybranej grupy gmin).

Zmora wielu analityków są rutynowe analizy, które wykonuje się regularnie, czasami ze stosunkowo niewielkimi modyfikacjami. Zrozumiałe jest wymaganie użytkowników systemu analitycznego, aby takie rutynowe raporty były wykonywane w możliwie jak najszerszym stopniu automatycznie.

Kolejnym problemem, który warto wskazać (a który powinien zostać rozwiązany przez twórców systemu analitycznego) jest zapisywanie wyników analiz i ich definicji w taki sposób, żeby można je było łatwo odnaleźć i stosować. Mówiąc bardziej technicznie, system powinien przechowywać nie tylko same dane przeznaczone do analizy, ale również dokładne dane samych analiz i inne dane o systemie (nazywane metadanymi).

Biorąc pod uwagę powyższe postulaty StatSoft przygotował zintegrowany system analizy danych i rozpowszechniania wyników analiz *SENS*.

## **Ogólne własności systemu analitycznego *SENS***

Podstawową ideą zintegrowanego systemu analitycznego *SENS* jest zasada, że dane powinny być centralnie przechowywane i udostępniane wszystkim zainteresowanym. Dotyczy to nie tylko danych będących przedmiotem analiz, ale również wszelkich danych o systemie, takich jak dane o danych (np. jak interpretować daną liczbę), szablony analiz, definicje dostępu do danych, informacje w jaki sposób dane mają być rozpowszechniane, prawa dostępu do obiektów systemu itp. Takie dane określamy terminem metadane.

Centralne przechowywanie danych umożliwia zorganizowanie pracy grupowej oraz współdzielenia wiedzy. Dzięki temu, że użytkownicy systemu mogą wzajemnie wykorzystywać wyniki swojej pracy i nie ma potrzeby wielokrotnego wykonywania tych samych czynności (np. zapytanie zdefiniowane



przez Kowalskiego może być punktem wyjścia do analizy wykonywanej przez Nowaka), uzyskujemy wzrost efektywności pracy wszystkich użytkowników systemu.

Zintegrowany system analityczny powinien być kompletnym rozwiązaniem. Innymi słowy w skład systemu muszą wchodzić narzędzia potrzebne na wszystkich etapach wnioskowania: od pobrania danych, przez wykonanie obliczeń i wykresów, a skończywszy na przygotowaniu raportu.

*SENS* to kompleksowy system analizy danych. W jego skład wchodzi narzędzia analityczne *STATISTICA*, mechanizmy dostępu do zewnętrznych baz danych, generator raportów oraz narzędzia pracy grupowej. Dzięki przejrzystemu, graficznemu środowisku *SENS* umożliwia łatwe tworzenie zapytań do nawet bardzo złożonych, niejednorodnych baz danych i nie wymaga od użytkownika wiedzy informatycznej. Jako narzędzie analityczne *SENS* wykorzystuje program *STATISTICA*, wielokrotnie uznawany za najlepszy program statystyczny.

## Korzystanie z systemu *SENS*

Jedną z ważniejszych zalet systemu *SENS* jest możliwość dostosowanie go do potrzeb, uprawnień i umiejętności użytkownika. Innymi słowy użytkownicy systemu mogą pracować w takim środowisku jaki im najbardziej odpowiada. Poniżej przedstawimy przykłady korzystania z systemu przez użytkowników różnego typu.

### Odbiorcy raportów

Największą grupą korzystającą z pracy systemu *SENS* są osoby będące wyłącznie odbiorcami wyników analiz, zestawień i raportów. System można skonfigurować tak, aby tacy użytkownicy korzystali z wyników pracy *SENS* w doskonale im znanym, prostym środowisku – w szczególności mogą oni w ogóle nie wiedzieć, że używają zaawansowanego technologicznie narzędzia analitycznego.

Przykładowo *SENS* może w pełni automatycznie generować raport, umieszczać go na stronie Internetowej i udostępniać go w ten sposób zainteresowanym i uprawnionym osobom. Automatycznie można również generować raporty odczytywalne przez praktycznie wszystkie edytory tekstu dla Windows (np. MS Word). Dzięki temu odbiorcy wyników analiz w ogóle nie muszą go poznawać systemu.

### Analitycy

Pierwszym etapem analizy jest odczytanie potrzebnych danych np. z bazy danych. *SENS* może współpracować z bardzo różnorodnymi źródłami danych, w tym z praktycznie wszystkimi nowoczesnymi systemami zarządzania bazą danych np. MS Access, Oracle, SQL Server, Sybase, Informix. Można oczywiście korzystać także z plików danych w formatach popularnych aplikacji takich jak MS Excel, dBase, Lotus oraz plików tekstowych.

Pobieranie danych z zewnętrznych źródeł danych w systemie *SENS* jest maksymalnie ułatwione. Wykonywanie wszystkich czynności niezbędnych do pobrania danych jest wspomagane przez *Kreatory* i odbywa się tylko raz. Po zdefiniowaniu szablonu dostępu do danych można z nich swobodnie korzystać bez znajomości technologii baz danych, struktur danych, języków programowania itp. Szablon dostępu do danych nazywany jest w systemie *profilem*. W profilu zapisywane są informacje umożliwiające pobranie danych (np. z której bazy danych skorzystać i jaki mechanizm dostępu jest dla niej odpowiedni) oraz ich interpretację (np. że pewną liczbę należy odczytać jako procent bezrobotnych).



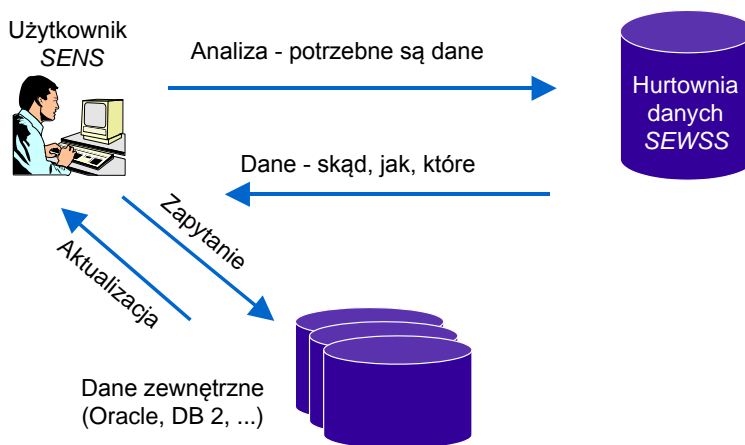
Osobom wykonującym analizy *SENS* daje do dyspozycji potężne narzędzia analityczne z rodziny *STATISTICA*, które wielokrotnie były uznawana w niezależnych porównaniach za najlepsze programy do analizy danych.

Po przeprowadzeniu analizy w trybie interakcyjnym możemy zapisać jej wyniki w bazie danych *SENS* jako aktywny szablon analizy (tzw. monitor) lub jako statyczny raport – tak aby mogli z niej korzystać wszyscy zainteresowani użytkownicy, pod warunkiem że mają odpowiednie uprawnienia. Podobnie jak w przypadku definiowania dostępu do danych, do budowy raportów i publikowania ich np. w Internecie nie wymagana jest wiedza informatyczna, znajomość języka HTML itp.

## Przebieg analizy

W momencie rozpoczęcia pracy w *SENS* użytkownik musi podać swój identyfikator i hasło. System sprawdza wtedy jego uprawnienia i udostępnia odpowiednie dla niego narzędzia oraz szablony szablonów analiz i dostępu do danych.

Przebieg analizy przedstawia poniższy schemat:



Po wybraniu analizy z listy analiz dostępnych dla użytkownika, system sprawdza jakich danych ona dotyczy, skąd je pobrać i w jaki sposób. Następnie do bazy danych wysyłane jest zapytanie, w odpowiedzi na które na stację roboczą użytkownika przesyłane są dane. Następnie automatycznie wykonywana jest analiza, a jej wyniki prezentowane są użytkownikowi.

## Podsumowanie: główne zalety systemu *SENS*

### Zgodność danych i raportów

Jednym z najważniejszych wymagań stawianych zintegrowanemu systemowi analitycznemu jest zapewnienie zgodności uzyskiwanych wyników. W systemie *SENS* wszystkie specyfikacje dotyczące analizy przechowywane są w jednej bazie danych, tak aby wszyscy użytkownicy systemu uzyskiwali identyczne wyniki analiz.

### Łatwy i szybki dostęp do danych

Użytkownicy nie znający języka SQL i zaawansowanej technologii informatycznej potrzebują łatwego dostępu do danych z różnych baz korporacji. System umożliwia im to poprzez



przechowywanie gotowych zapytań, które można wykorzystać do pobrania danych. Zapytania do baz danych można budować w intuicyjny sposób za pomocą graficznego środowiska użytkownika.

## **Praca grupowa i współdzielenie wiedzy**

*SENS* umożliwia wspólną pracę zespołu ludzi, dzięki czemu zapewnia efektywność pracy swoich użytkowników. Obiekty systemu (analizy, raporty, definicje dostępów do danych) stworzone przez jednego użytkownika mogą być udostępnione innym użytkownikom. Wiele osób może korzystać z pracy wykonanej przez jedną i nie ma potrzeby wielokrotnego powtarzania tych samych czynności przez różnych użytkowników.

## **Dostosowanie do użytkownika**

Użytkownicy mają dostęp do tych informacji, których potrzebują. Wygląd systemu określany jest w momencie zalogowania użytkownika i uzależniony jest od uprawnień użytkownika oraz grup, do których on należy. Dzięki takiemu rozwiązaniu łatwiej znaleźć odpowiednie analizy - nie są widoczne te elementy systemu, które danemu użytkownikowi nie są potrzebne, a jedynie zaciemniałyby obraz.

## **Polska wersja językowa**

*SENS* dostępny jest w polskiej wersji językowej – co na pewno ułatwia korzystanie z niego.

## **Skalowalność**

System jest w pełni skalowalny i bezproblemowo rośnie wraz z potrzebami organizacji. Można zacząć od niewielkiej instalacji i przejść do skali całego przedsiębiorstwa, bez konieczności wymiany systemu czy zmiany jego architektury.

## **Bogaty zestaw narzędzi analitycznych**

Częścią *SENS* jest *STATISTICA* - pakiet wysokiej jakości narzędzi wizualizacji i analizy danych, który od momentu wejścia na rynek wygrywa wszystkie porównania z innymi programami tego typu. Warto zwrócić uwagę na dostępne w systemie wydajne narzędzia *data mining* (zgłębianie danych), dzięki którym system może automatycznie generować reguły decyzyjne, wybierać cechy w istotny sposób wpływające na siebie, prognozować i budować modele zależności.

## **Organizacja dostępu do wyników analiz**

Ponieważ wszystkie szablony analiz i raporty przechowywane są w systemie *SENS* w logicznie uporządkowanym schemacie systemu łatwiej jest odnaleźć potrzebną analizę.

## **Niski koszt wdrożenia i korzystania z systemu**

*SENS* jest systemem z półki. Wymagane jest zwykle pewne przystosowanie go do konkretnych potrzeb (wstępna konfiguracja), jednak jest to proces prosty i szybki. System jest rozwiązaniem sprawdzonym i unika się wielu pułapek związanych z aplikacjami tworzonymi na zamówienie (długie czasy wdrożenia, problemy z utrzymaniem się w harmonogramie, odkrywanie błędów konstrukcyjnych w trakcie pracy z systemem i ciągle ich usuwanie). Narzędzia administracyjne pozwalają łatwo rozbudowywać system o nowe analizy, raporty, dodawać nowych użytkowników.

## **Centralne zarządzanie i łatwość administracji**

Większość operacji administracyjnych w systemie (takich jak na przykład definiowanie nowego profilu dostępu do danych) można wykonać za pomocą wygodnych kreatorów. Zmiany zapisane



w *SENS* (wprowadzone przez administratora) są od razu uwzględniane na stacjach roboczych użytkowników.

### **Niewielkie wymagania sprzętowe**

System *SENS* nie ma dużych wymagań sprzętowych – do korzystania z systemu potrzebne są:

- ◆ sieć komputerowa
- ◆ stacje robocze – komputery z systemem Windows 95, 98, NT 4.0 lub Windows 2000
- ◆ komputer pełniący rolę serwera - zazwyczaj można wykorzystać istniejący serwer (wymagane parametry serwera zależą od wielu różnych czynników, m.in. od innych aplikacji korzystających z serwera).