

PIOTR SZEFLIŃSKI, SEBASTIAN KRZEPKOWSKI
Politechnika Łódzka
Zakład Akademickiej Sieci Bibliotecznej
Centrum Komputerowe

SILOS.

Źródło wiedzy o przeszłości dla przyszłości

Doświadczenia zdobyte przy eksploatacji systemu bibliotecznego Horizon w Łódzkiej Akademickiej Sieci Bibliotecznej wskazują, że standardowe statystyki oferowane przez system nie w pełni zaspokajają potrzeby bibliotekarzy. Dostępne dane są w pewien sposób zagregowane i nawet ich codzienne gromadzenie nie zapewnia uzyskania odpowiedzi na każde pytanie.

*Dlatego w Zakładzie Akademickiej Sieci Bibliotecznej powstał system SILOS, który wychodzi naprzeciw zapotrzebowaniu bibliotek na dane statystyczne **dzisiaj i jutro**. Jest to hurtownia danych, która codziennie archiwizuje stan systemu i wypożyczeń w formie umożliwiającej ich dalszą analizę.*

Zapewnia bibliotekarzom teraz i w przyszłości możliwość stawiania pytań statystycznych bez ograniczeń, w dowolnych przekrojach czasowych i tematycznych; również pytań, co do których dopiero w następnych latach będą wiedzieli, że powinni je postawić.

SILOS dostarcza dane, które mogą służyć do oceny przemian, jakie zachodzą w wykorzystaniu systemu bibliotecznego na przestrzeni czasu, a tym samym może pomóc prognozować przyszłe preferencje czytelników zarówno merytoryczne jak i organizacyjne.

1. Wstęp

Polscy bibliotekarze najczęściej pracują w systemach „Aleph”, „Horizon”, „Prolib” i „Virtua”. Na potrzeby swojej codziennej działalności oraz do okresowych sprawozdań zbierają różnego rodzaju statystyki.

W Łódzkiej Akademickiej Sieci Bibliotecznej, zrzeszającej biblioteki Akademii Muzycznej, Akademii Sztuk Pięknych, Centrum Badań Molekularnych i Makromolekularnych Polskiej Akademii Nauk, Instytutu Medycyny Pracy, Państwowej Wyższej Szkoły Filmowej Telewizyjnej i Teatralnej, Politechniki Łódzkiej, Uniwersytetu Łódzkiego, Uniwersytetu Medycznego, Wojewódzkiej i Miejskiej Biblioteki Publicznej im. Marszałka J. Piłsudskiego oraz Wyższego Seminarium Duchownego, używany jest system „Horizon”.

2. Zarys problemu

„Horizon” daje szczegółowe statystyki odnośnie haseł, rekordów bibliograficznych, egzemplarzy, czytelników, wypożyczeń, płatności i innych. Dostarcza on procedurę o na-

zwie „Dayend”, która przelicza i gromadzi statystyki za okres od ostatniego wykonania procedury i przechowuje je przez ustalony przez bibliotekę czas.

Wiele statystyk można pozyskać bezpośrednio z „Horizona” – niektóre wprost z aplikacji systemowej, niektóre dzięki powstałemu w Zakładzie Akademickiej Sieci Bibliotecznej systemowi zapytań SQL do baz konsorcjum.

System zapytań SQL pozwala zliczać między innymi hasła wzorcowe, hasła bibliograficzne i egzemplarze z podziałem na kategorie, a także wykonać niestandardowe zapytanie do baz „Horizona”.

W trakcie wdrażania tego systemu pojawiły się problemy, których źródłem jest specyfika systemu „Horizon” w wersji 5.1. Nie przechowuje on historii wypożyczeń i przedłużeń poszczególnych egzemplarzy, chociaż statystyki systemowe (o ile są regularnie wykonywane) dostarczają ich liczby. Oznacza to, że system przechowuje dane statystyczne w postaci już zagregowanej, a dane źródłowe, na podstawie których ta agregacja jest wykonywana – są usuwane. Jeśli więc biblioteki potrzebne były statystyki, których nie zapewnia procedura „Dayend”, to nie było możliwości uzyskania takich danych, bo system ich nie gromadzi.

W związku z tym pojawiła się potrzeba opracowania rozwiązania, dzięki któremu bibliotekarze będą mogli mimo wszystko uzyskać potrzebne statystyki w pewnym sensie niezależnie od systemu i jego specyfikacji.

3. Rozwiązanie

Rozwiązaniem prezentowanym przez ZASB na konferencji jest system „SILOS”.

„SILOS” jest systemem kopiowania i gromadzenia wybranych obszarów struktury danych z baz „Horizona” w takiej formie jak w oryginalnej bazie.

Powstał w celu:

1. uzyskania statystyk niedostępnych z systemu Horizon, dzięki:
2. uniezależnieniu się od zakresu czasowego statystyk i od bieżącego stanu bazy, aby:
3. uzyskać odpowiedzi na pytania i potrzeby, które mogą pojawić się w przyszłości, a których w obecnej chwili nie można przewidzieć.

Ogromna ilość gromadzonych danych – średnio 50 MB dziennie/bazę – powoduje, że czas oczekiwania na odpowiedź to około 5 minut w przypadku zapytania za okres 2 miesięcy.

W związku z tym, w drugim etapie projektu, ZASB zaproponował rozwiązanie, które pogodziło potrzebę zapewnienia wszelkich możliwych danych do potencjalnych statystyk z oczywistą chęcią uzyskania natychmiastowej odpowiedzi na zadane pytanie.

Tym rozwiązaniem jest „Elewator”.

„Elewator” jest modułem „SILOSa” zliczającym najczęściej požądane statystyki na podstawie zadanej listy zapytań i przechowującym cząstkowe wyniki w osobnej bazie. Dokładność pomiarów pozostaje niezmienna a jednocześnie długotrwały proces przeglądania danych zostaje skrócony do minimum, ponieważ najbardziej żmudne obliczenia wykonywane są z awczasu.

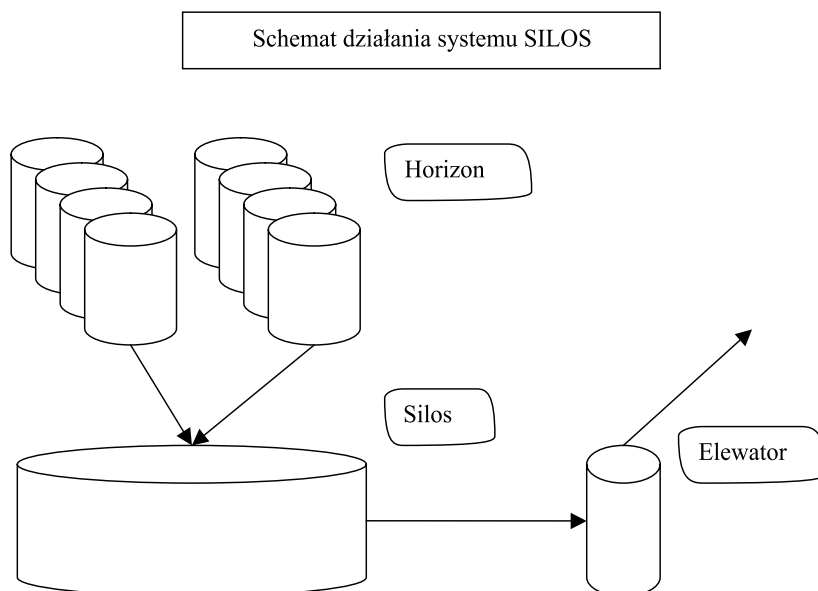
Do jego sprawnego działania zostały przyjęte dwa założenia odnośnie zapytań:

1. Muszą być one sformułowane w ten sposób, żeby wartość obliczeń z całego zadanego okresu czasu była równa sumie wartości z poszczególnych dni (zachowanie dokładności wyników)
2. Jedyną zmienną w zapytaniu może być czas.

Rozwiązanie to z natury rzeczy ogranicza możliwości zadawania dowolnych pytań. Jednak nasze doświadczenia z systemem SQL wskazują, że liczba najbardziej potrzebnych bibliotekom danych statystycznych jest skończona i nie tak bardzo rozbudowana.

Jeżeli jednak zaistnieje potrzeba np. rozróżnienia czytelników na kategorie (pracownicy, studenci, doktoranci itd.), wtedy konieczne będzie utworzenie dla każdej z nich po jednym zapytaniu i przygotowanie wstępnie opracowanych wyników.

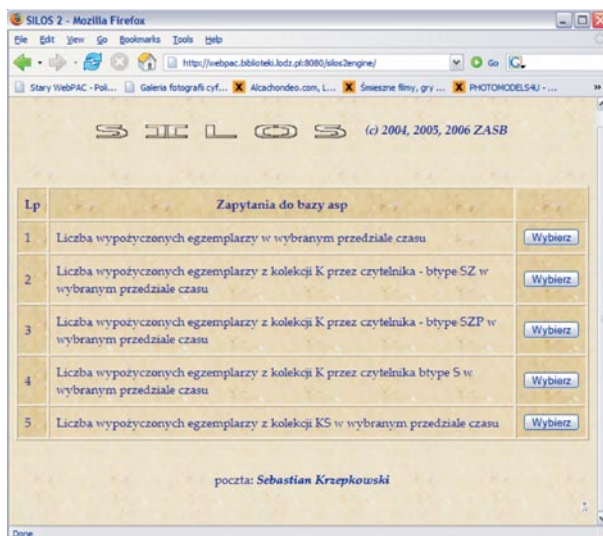
Po wprowadzeniu modułu „Elewator” czas oczekiwania na każdą odpowiedź skrócił się do kilku sekund.



Codziennie wieczorem rozpoczyna się procedura kopiowania wybranej zawartości wszystkich baz bibliotecznych do bazy „Silos” i trwa zwykle godzinę. Następnie na podstawie tych danych liczone są statystyki zbiorcze, co obecnie trwa średnio 3 minuty, i trafiają do oddzielnej bazy „Elewator”.

Bibliotekarze mają dostęp do danych z poziomu serwisu internetowego.

Po zalogowaniu do systemu wybierają rodzaj statystyki, podają przedział czasu i w ciągu kilku sekund otrzymują odpowiedź.



Bazy oraz systemy „Silos” i „Elewator” działają na serwerze Łódzkiej Akademickiej Sieci Bibliotecznej, a dane archiwalne zabezpieczone są na macierzy dyskowej zarządzanej przez Centrum Komputerowe PŁ.

4. Podsumowanie

Każda nowa wersja systemu bibliotecznego z reguły rozszerza zakres funkcjonalności, w tym prezentowanych statystyk, ale wyniki za okres przed wprowadzeniem nowej wersji często nie są już dostępne.

„SILOS”, dzięki niezależnej metodzie gromadzenia danych, oprócz swobody co do terminu opracowania zapytania, umożliwia również sporządzenie przekrojowych statystyk dla różnych wersji systemu. W przyszłości będzie więc możliwe uzyskanie wiarygodnych danych o funkcjonowaniu biblioteki na przestrzeni lat.

SILOS – the source of knowledge about the past for the future purposes

The experience gained during operation of the Horizon Integrated Library System within the Academic Library Network of Łódź shows that the standard statistics offered by the system do not fully satisfy the expectations of the librarians. The available data are aggregated in some way and even though they are collected every day, they do not provide answer for each question.

For this reason, in the Institute of the Academic Library Network SILOS system came into existence, which meets the demand of the librarians for statistical data today and tomorrow. This is the data warehouse, which daily archives the current status of the system and borrowings in the form allowing for further analysis.

It provides the librarians with the possibility of asking statistical questions without any limitations as for the time and thematic field, as well as the questions they will be aware of in next years.

SILOS provides the data that can be used for evaluation of the changes in regard to the operation of the library system within a period of time, and in this way it may be useful in forecasting the future preferences of the readers, including substantial and organisational ones.