STUDIA INFORMATICA 2000
Volume 21 Number 1 (39)

Przemysław GŁOMB
Instytut Informatyki Teoretycznej i Stosowanej PAN

PROJEKT MECHANIZMU ANALIZY I INDEKSOWANIA INFORMACJI MULTIMEDIALNEJ DLA ŚRODOWISKA ORACLE VIDEO SERVER

Streszczenie. Praca prezentuje projekt rozszerzenia środowiska sieciowego OVS, umożliwiającego automatyczne generowanie opisu dołączanych do bazy danych filmów. Przedstawiono jego ideę działania i schemat funkcjonalny, oraz używane w nim metody analizy i klasyfikacji strumieni wideo.

DESIGN OF A UNIT FOR ORACLE VIDEO SERVER SYSTEM ANALYSING AND INDEXING MULTIMEDIA INFORMATION

Summary. This work presents a project of extending the Oracle Video Server with a module that performs automatic generation of a description for films added to the server database. A description of the module is presented, together with overview of the analysis and classification methods that form the base of it.

1. Wstęp

Od niedawna jesteśmy świadkami gwałtownego wzrostu parametrów systemów komputerowych. Na "polu walki" coraz częściej pojawiają się technologie ATM, oferujące znacznie większe niż do tej pory, niesamowite dla przeciętnego użytkownika, szybkości transmisji danych[1].

W miarę wzrostu przepustowości sieci zmienia się struktura ruchu w niej występującego. Tradycyjne obiekty, takie jak np. dokumenty tekstowe, mają już bardzo mały udział w

[1] Obecnie przecież opracowywana jest specyfikacja transmisji przy szybkości 10 Gbps(!), a szybkość 622 Mbps jest niemalże "naturalna".

przesyłanych megabajtach danych. Ich miejsce zajmują obrazy, cyfrowo nagrany dźwięk, filmy wideo.

Innym, powiązanym z powyższym, zjawiskiem, jest tendencja do konwergencji sieci teletransmisyjnej – łączenie sieci telekomunikacyjnych i komputerowych, co już się niemal dokonało w postaci ISDN.

Następnym polem, które podbije wiecznie głodna informacji sieć, będzie transmitowanie telewizji i wideo. Z wykorzystaniem współczesnych technologii kompresji i transmisji, jakość usług cyfrowych w niczym nie odbiega od tradycyjnych, analogowych, a z racji możliwości samych komputerów ma nad nią kilka przewag – łatwiejsze składowanie danych, możliwość zastosowania technologii Wideo na Żądanie (VoD – Video on Demand).

Tendencja ta rośnie i będzie rosnąć tak długo, jak przeciętny zjadacz chleba będzie odczuwał wymierne korzyści z używania technologii komputerowej zamiast tradycyjnej.

Wspomniana wyżej technologia VoD jest bardzo atrakcyjną alternatywą obecnej telewizji. Jej ideą przewodnią jest agregacja filmów na serwerze dużej mocy, i, na żądanie użytkownika, przesyłanie wskazanego przez niego filmu przez sieć. Komputer użytkownika dekoduje i prezentuje wideo w czasie rzeczywistym - odpada konieczność składowania. Dzięki temu ta technologia nie wymaga od strony klienta maszyn o dobrych parametrach - wystarczy tzw. przystawka telewizyjna (ang. *set-top box*). Najważniejsze z punktu widzenia widza jest to, że to on decyduje, jaki film, i kiedy chce obejrzeć. Nie musi przesiadywać przed telewizorem kilku godzin, aby obejrzeć ulubiony film, albo czekać całe dnie na niego.

Serwery oferujące filmy staną przed kilkoma problemami.

Podstawowymi będą problemy czysto techniczne, związane ze złożonością - administracją i przechowywaniem dużej ilości dużych objętościowo danych. Jest problem znany z obszaru baz danych, i można się spodziewać, że wypracowane tam rozwiązania okażą się skuteczne.

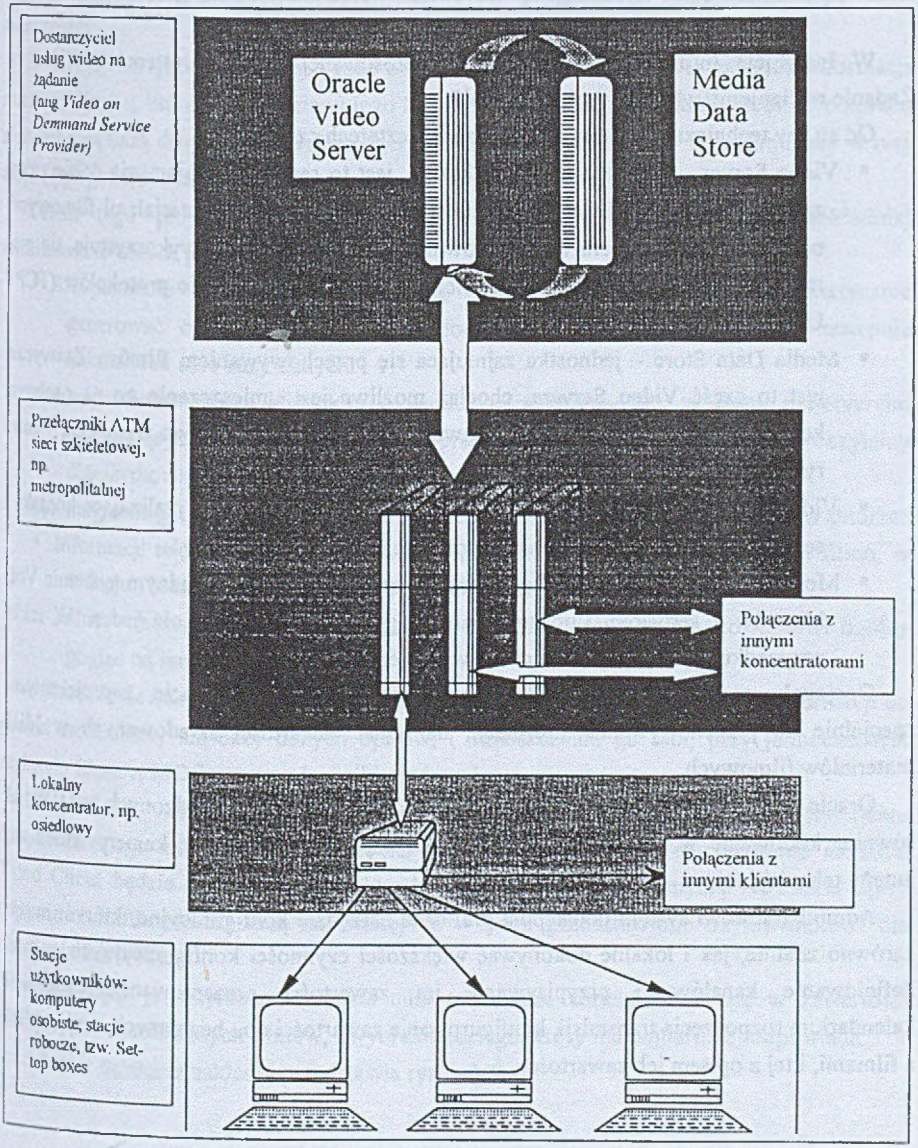
Jednak, gdy się z nimi uporamy – pojawi się następna przeszkoda. Jej istotą jest fakt, że stale wzrasta liczba filmów puszczanych w obieg. W samym Hollywood co tydzień pojawia się kilkanaście premier filmów. Codziennie też pojawiają się wiadomości telewizyjne – w tej chwili jest to w Polsce zaledwie kilka programów dziennie, ale dość szybko możemy się spodziewać wzrostu tej liczby.

Operatorzy serwisów VoD staną przed poważnym problemem szybkiego opisywania / klasyfikacji filmów, które będą chcieli dołączyć do swojej kolekcji.

Ten projekt przedstawia rozszerzenie środowiska VoD Oracle Video Server / Oracle Video Client, które usuwa jedno z największych ograniczeń jego działania, jakim jest konieczność osobistego nadzoru nad modyfikacją filmowej bazy danych. Opisywana tutaj

nakładka jest w stanie w dużym stopniu zautomatyzować proces takiego działania, co oszczędza czas i zwiększa możliwości działania administratora takiego systemu.

Ma to wyjątkowe znaczenie w poważnych zastosowaniach, przykładowo w roli dostarczyciela (ang *provider*) programów telewizyjnych na całe miasto (patrz rysunek 1). W



Rys. 1. System VoD na przykładzie Oracle Video Server
Fig. 1. A VoD system (Oracle Video Server) example

takim przypadku ręczna indeksacja dołączanych do bazy materiałów filmowych byłaby trudna i pracochłonna, jeżeli w ogóle możliwa.

2. Oracle Video Server

W Instytucie Informatyki Teoretycznej i Stosowanej badania systemów Wideo na Żądanie realizujemy właśnie w systemie OVS.

Od strony technicznej, system ten składa się z czterech części[1]:

- Video Server – jak sama nazwa wskazuje, jest to serwer zajmujący się odbieraniem zgłoszeń od stacji klienckich, administracją bazy danych z materiałami filmowymi, oraz ich dostarczaniem do użytkownika. Dla dostarczenia wykorzystuje się sieć szerokopasmową, używając dowolnego z popularnych obecnie protokołów (TCP / UDP, ATM, DVB).
- Media Data Store – jednostka zajmująca się przechowywaniem filmów. Zazwyczaj jest to część Video Servera, chociaż możliwe jest umieszczenie go na osobnym komputerze (np. na macierzy dyskowej), jeżeli zastosowania tego wymagają – patrz rysunek 1.
- Video Client – oprogramowanie klienckie dla stacji roboczych, realizujące kontakt z serwerem, odbiór i wyświetlanie w czasie rzeczywistym filmów.
- Media Net – jest to środowisko umożliwiające komunikację w jednym systemie VoD klientów z serwerem (lub serwerami – gdy jest ich więcej niż jeden. W takim przypadku zajmuje się także równoważeniem obciążenia sieci).

Opcjonalnym komponentem dodatkowym jest baza danych Oracle, zaprojektowana specjalnie do przechowywania informacji, na temat zawartości składowanych w MDS materiałów filmowych.

Oracle Video Serwer umożliwia transmisję zarówno danych zgromadzonych w MDS, jak również kodowanie w czasie rzeczywistym i przesyłanie sygnałów z kamery wideo czy anteny telewizyjnej.

Administratorowi system udostępnia graficzne narzędzie konfiguracyjne, którym można zarówno zdalnie, jak i lokalne dokonywać większości czynności konfiguracyjnych, w tym: definiowanie kanałów, i przypisywanie im zawartości prezentowanej, ustawianie kalendarium rozpoczęcia transmisji, konfigurowanie zawartości obu baz danych – tej w MDS z filmami, i tej z opisem ich zawartości.

3. Projekt nakładki

Oracle Video Server jest dobrym środowiskiem do tworzenia systemów VoD. Jedną z jego największych zalet jest rozszerzalność, łatwość dołączania do niego nowych aplikacji. Sam system udostępnia wszechstronne biblioteki programisty API, które dają duże możliwości.

W swojej wersji podstawowej Video Server nie przechowuje prawie żadnych informacji na temat filmów, które znajdują się w jego MDS. Wspomniano, że umożliwia on współpracę z dodatkową bazą danych, zawierającą opisy zawartości, tym niemniej wszelkie opisy w niej muszą być generowane ręcznie.

Celem tego projektu jest usunięcie tego ostatniego ograniczenia, a dokładniej zrealizowanie dwóch podstawowych zadań:

- Uzupełnienie Video Servera o moduł, który będzie analizować i automatycznie generować opis strumienia wideo dodawanego do bazy danych, a następnie uaktualniać obie bazy danych.
- Uzupełnienie części klienckiej o moduł kontaktujący się w sieci z Serverem, odczytujący tak stworzony opis, oraz prezentujący te informacje w sposób czytelny dla użytkownika.

Automatyczna generacja opisu polegać będzie na stworzeniu dwóch informacji o zbiorze:

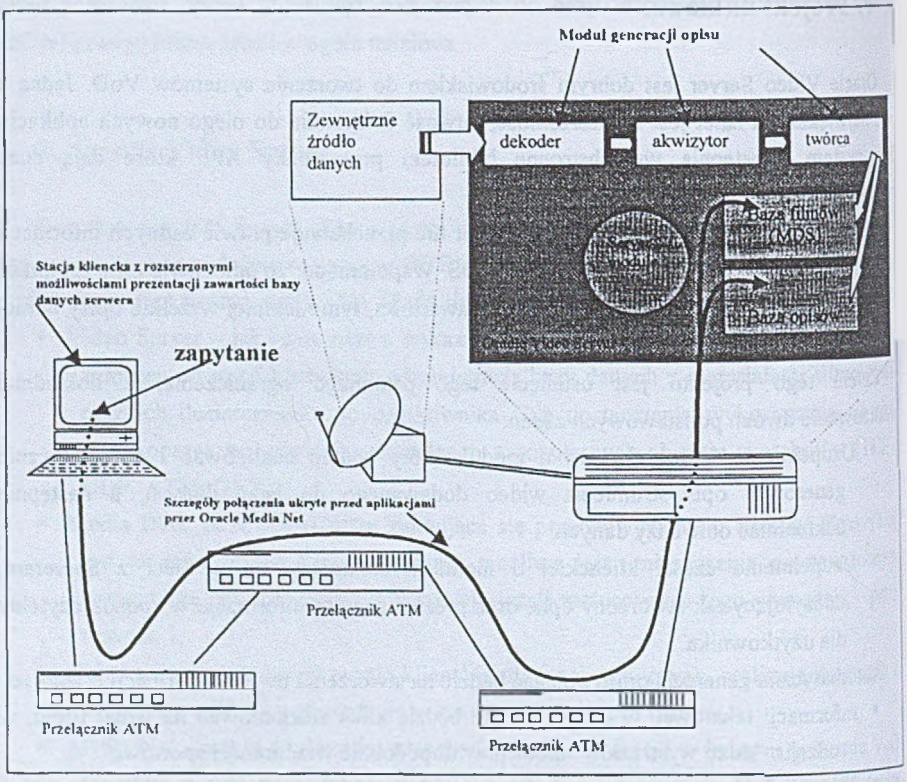
- Informacji tekstowej, w której zawarte będzie kilka sformułowań na temat filmu, w rodzaju: "dużo wystrzałów" albo "prawdopodobnie wiadomości sportowe".
- Jedno- lub dwuminutowego skrótu (ang *abstract*), który da użytkownikowi ogólny pogląd na temat treści.

Uaktualnienie bazy danych będzie polegać na sformatowaniu tak uzyskanej informacji do postaci zrozumiałej dla bazy danych opisów, i umieszczenie go tam, przy jednoczesnym zapisaniu filmu w MDS wraz z odnośnikiem do opisu.

Uzupełnienie części klienckiej będzie polegać na poszerzeniu fragmentu odpowiedzialnego za odbieranie z Servera listy filmów. Po zaimplementowaniu nakładki Video Client będzie dodatkowo odwoływał się do zdalnej bazy danych, i odbierał opisy odpowiednich filmów. Zostaną one następnie w pełni przedstawione użytkownikowi dla ułatwienia mu wyboru.

Przeglądając je, użytkownik będzie miał możliwość zorientowania się w zawartości istniejących w bazie danych filmów, i wybrania takiego, który mu najbardziej odpowiada.

Schemat działania nakładki przedstawia rysunek 2.



Rys. 2. Umiejscowienie projektowanej nakładki w OVS i realizacja zapytania
Fig. 2. Emplacement of the module in the OVS system, and realization of user's query

4. Mechanizmy analizy i generowania opisu strumienia wideo

4.1. Mechanizmy akwizycji informacji ze strumienia wideo

Część od strony serwera składa się z trzech modułów, zwanych odpowiednio: "dekoder", "akwizytor", i "twórca".

"Dekoder" ma za zadanie rozkodowania strumienia mpeg, w celu wyłonienia z niego poszczególnych klatek, odkładanych następnie do bufora.

Moduł "akwizytor" bada klatki w buforze, tworząc za pomocą niżej przedstawionych metod wektor cech opisujących strumień wideo:

- bada histogramy rozkładu kolorów w poszczególnych klatkach,

- wykrywa sceny; w tym celu odejmuje histogramy kolorów sąsiednich klatek. Na granicy scen różnice będą dużo większe niż w obrębie sceny[6]. Odpowiednia lista scen, oraz histogram ich długości są tworzone podczas tej operacji,
- bada natężenie ruchu w danej klatce. Wykrywanie polega na porównaniu, ile w badanej klatce wystąpiło bloków obrazu (ang *macroblocks*) takich samych jak we wcześniejszej, ale w innych miejscach[6]. Interpretowane jest to jako przesuwanie się obiektów po ekranie, i daje pewne pojęcie o dynamice tego odcinka filmu,
- bada ścieżkę dźwiękową. Sporządzane są dwie funkcje: natężenia i częstotliwości podstawowej dźwięku,
- wykrywa kolor dominujący (lub kolory dominujące), oraz ilość jego wystąpień w poszczególnych klatkach. Dane te otrzymuje się po przeanalizowaniu otrzymanych w punkcie pierwszym histogramów kolorów. Kolor dominujący ma znaczenie dla filmów dziejących się w określonej scenerii, np. "Waterworld",
- wykrywa zdarzenia, występujące w filmie. "Zdarzenia" – to charakterystyczne sekwencje dźwiękowe, przydatne w konstruowaniu opisu. Wybrane sekwencje dźwiękowe to: odgłosy wystrzałów i wybuchów. Wybrano takie ze względu na ich specyfikę i przypisanie do pewnych kategorii filmów. Wykrywanie ich polega na porównywaniu ścieżki dźwiękowej ze wzorcem, charakterystycznym dla danego zdarzenia[4]. Ta metoda umożliwia też wykrywanie innych zdarzeń, np. krzyk, śmiech, pisk hamulców, aczkolwiek w tej chwili "akwizytor" będzie wyszukiwał jedynie dwa pierwsze zdarzenia.

Klatki, które już przebadano, dla zaoszczędzenia miejsca usuwa się z bufora. Z chwilą gdy cały film zostanie zdekodowany, "dekoder" i "akwizytor" kończą swoje działanie, umieszczając w buforze otrzymany wektor cech.

4.2. Analiza wektora cech i generacja opisu filmu

Po zakończeniu procesu akwizycji rozpoczyna działanie moduł "twórca". W oparciu o otrzymane w poprzednim punkcie dane realizuje on dwa zadania:

4.2.1. Stworzenie opisu tekstowego

Opis ten ma postać kilku słów, dających ogólne wyobrażenie o zawartości filmu.

Rozpoczyna się od próby klasyfikacji filmu do jednej z kilku kategorii[5]:

- wiadomości: charakteryzują się powtarzającymi się scenami i charakterystycznym wzorem ścieżki dźwiękowej,
- transmisje sportowe zawierają dużo ruchu, krótkie, często zmieniające się sceny, również charakterystyczny wzór ścieżki dźwiękowej,
- programy komediowe: duża ilość śmiechu,

- kreskówki: charakterystyczny rozkład kolorów, dłuższe sceny, momenty, w których natężenie dźwięku spada do zera.

Dobór takich kategorii jest uzasadniony występującą, także u ludzi, trudnością w wyłonieniu jednoznacznego podziału informacji zawartej w filmach. Szansa, że film będzie do któreś z wyżej wymienionych kategorii należał, jest mała (obejmie, w zależności od rodzaju bazy danych, prawdopodobnie kilkanaście procent obiektów), ale za to ich znaczenie nie budzi wątpliwości. Te obiekty filmowe, dla których nie zostanie stwierdzona przynależność do żadnej z tych grup, zostaną opatrzone stwierdzeniami klasy "dużo akcji" (jeżeli natężenie ruchu w filmie będzie odpowiednio duże), lub "dużo wystrzałów / wybuchów" (gdy odpowiednia ilość takich zdarzeń zostanie wychwycona).

"Odpowiednia ilość" jest wielkością zdefiniowaną arbitralnie, określoną w oparciu o analizę typowych dla każdego obiektu filmów.

4.2.2. Stworzenie skrótu

Informacja słowna, otrzymana w punkcie pierwszym, będzie bardzo często niewielka albo nie będzie jej wcale. Stąd dużo ważniejsze jest stworzenie kilkunastosekundowego skrótu, dającego wgląd w treść filmu.

Tworzy się go przy następujących założeniach:

- jako materiał wybiera się około 80% filmu – unika się dzięki temu zdradzania zakończenia,
- z każdej sceny wybiera się jeden fragment o stałej długości,
- klasyfikuje się fragmenty; o tym, które z nich trafią do skrótu, decydują sformułowane niżej funkcje oceny, oraz rozsypanie po całym przebiegu strumienia (aby uniknąć sytuacji, w której skrót został utworzony na podstawie np. wyłącznie pierwszej ćwiartki filmu).

Rozpoczyna się od utworzenia listy scen.

Następnie wybiera się te, w których wystąpiły wybuchy i wystrzały. Wybiera się z każdej sceny fragment. Ocenia się je pod względem zgodności ze wzorcem danego zdarzenia.

Te sceny, w których wychwycono wspomniane wyżej fragmenty, usuwa się z listy scen.

Następnie z pozostałych scen wybiera się po jednym fragmencie. Miejsce wyboru, i zarazem bezwzględną przydatność fragmentu, określa funkcja oceny:

$$f = v + k + d(f) \tag{1}$$

gdzie:

v – wartość funkcji natężenia ruchu w tej klatce,

k – natężenie koloru(ów) dominującego(ych) w tej klatce w stosunku do reszty,

$d(f)$ – pochodna zmian częstotliwości podstawowej dźwięku.

Dzięki temu otrzymujemy fragment zawierający akcję, o tle kolorystycznym zbliżonym do tła filmu, i odpowiednio dynamicznym fragmencie dźwiękowym [4].

Ostatnim krokiem jest wybranie z listy fragmentów ocenionych powyższą funkcją tych, które dostały najmniej punktów. Z odpowiadających im scen wybiera się nowe fragmenty, używając tym razem innej funkcji:

$$f = (Vmax - v) + (Dmax - d(f)) \tag{2}$$

gdzie:

$v, d(f)$ jak w przypadku drugim,

$Vmax$ to maksymalna wartość, którą dla tego filmu osiąga funkcja v,

$Dmax$ to maksymalna wartość, którą osiąga wartość $d(f)$.

Celem jest w tym przypadku wyłonienie tych kawałków filmu, które są spokojne (małe natężenie ruchu, odpowiedni dźwięk). Istnieje podejrzenie, że odpowiadające im sceny w filmie mają za zadanie budowanie napięcia, i przekazywanie ładunku emocjonalnego.

Mając już tak określone trzy listy fragmentów, uporządkowane odpowiednimi funkcjami oceny, trzeba podjąć decyzję, które z nich zostaną wybrane do skrótu (będzie ich bardzo dużo – czas trwania scen liczona jest zwykle w sekundach, a długość filmu w godzinach).

Zadanie jest o tyle istotne, że nie powinno się tworzyć skrótu pełnego akcji np. dla melodramatu, albo wybierać zbyt dużo scen wolnych i spokojnych dla filmu sensacyjnego.

W tym celu porównuje się ze sobą trzy pierwsze oceny z każdej z list. Jeżeli oceny jakości któreś z grup będą przeważać, do skrótu zostanie włączona większa ilość fragmentów z danej listy. Oceny są bezwzględne, więc mogą być porównywane ze sobą bezpośrednio.

Funkcja, dokonująca porównania, ma następującą postać:

$$g(v11, v12, v13, v21, v22, v23, v31, v32, v33) = [A\ B\ C] \tag{3}$$

vnm – ocena m-tego fragmentu z listy n.

A: ilość scen zakwalifikowanych z grupy pierwszej - będzie miała wartości 0-3 (są to sceny wystrzałów i wybuchów – powinny, jako dynamiczne i przykuwające uwagę, znaleźć się w skrócie, ale w rozsądniej ilości);

B : 2-10 (sceny akcji);

C : 0-5 (sceny spokojne, nawet w filmie, w którym nie ma akcji, nie powinny zajmować całego skrótu – będzie on wtedy statyczny, utraci siłę przekazu obrazu ruchomego [4]).

Zgodnie z założeniami, $A+B+C = 10$.

Po wybraniu w powyższy sposób dziesięciu fragmentów następuje ich zakodowanie, w losowej kolejności, do zbioru w formacie mpeg.

Można zauważyć, że otrzymany w ten sposób skrót wykazuje podobieństwa do tzw. zapowiedzi filmowych (and *trailer*). Jest to podobieństwo zamierzone – taki sposób prezentacji treści przekazu audiowizualnego autor uznał za najodpowiedniejszy (najbardziej sugestywny).

4.3. Zapis opisu do bazy danych

Etap ten polega na zapisaniu badanego filmu oraz utworzonego skrótu w odpowiednim miejscu MDS, dopisanie opisu do zewnętrznej bazy danych, oraz uaktualnienie konfiguracji OVS.

5. Rozszerzenie modułu klienta

Oracle Video Client w swojej podstawowej postaci ma dość ubogie możliwości: potrafi tylko wyświetlać wybrany film, z możliwością przewijania do konkretnego miejsca, zatrzymywania itp.

Gdy użytkownik pyta o listę dostępnych filmów, w odpowiedzi dostaje jedynie nazwy zbiorów.

Aby zawartość bazy danych Video Servera była bardziej czytelna, projekt przewiduje w tym momencie wyświetlanie wraz z kursorem myszy opisu tekstowego z bazy danych i dodatkowo wyświetlanie w oknie odpowiadającego pozycji skrótu.

Te informacje będą sprowadzane w momencie, gdy oryginalny klient zgłosi zapytanie o listę filmów przechowywaną w MDS.

Taka forma prezentacji jest o wiele bardziej przejrzysta niż standardowa, pozwala "jednym rzutem oka" objąć zawartość bazy danych systemu VoD.

6. Podsumowanie i wnioski

Duża ilość danych, której magazynowanie i przesyłanie umożliwiają współczesne technologie komputerowe i telekomunikacyjne, woła o narzędzia nowego typu, radzące sobie sprawniej niż dotychczasowe z ich przetwarzaniem i zarządzaniem nimi. Narzędzia inteligentne, ukierunkowane na zawartość, stają się coraz bardziej potrzebne.

W miarę upowszechniania się systemów wideo na żądanie (co już niedługo nastąpi, biorąc pod uwagę dynamiczny rozwój architektury sieciowej) systemy te będą musiały zostać

rozbudowane o takie właśnie elementy – w przeciwnym razie ich obsługa i utrzymanie będzie wymagało wiele pracy.

Zaproponowane rozszerzenie jest według mnie przykładem takiego właśnie narzędzia nowej generacji. Dołączenie go do środowiska Oracle Video Servera zwalnia administratora systemu z obowiązku ręcznego dokładania materiałów filmowych. Wprawdzie część informacji będzie on musiał uzupełnić osobiście (na przykład nazwę filmu), ale zostanie mu dużo mniej pracy.

A to co zostanie, będzie w stanie wykonać łatwiej – choćby przeglądając skrót, zamiast poświęcać godziny na obejrzenie całego filmu.

LITERATURA

1. Oracle Video Server Documentation: Administration guide and command reference. Oracle corp. Release 3.0, 1998.
2. Oracle Video Server Documentation: Content administration guide. Oracle corp. Release 3.0, 1998.
3. A. Pope, R. Kumar, H. Sawhney, Ch. Wan: Video Abstraction: ... Proc. Thirty-Second Asilomar Conference on Signals, Systems & Computers, 1998, s. 915–919.
4. R. Lienhart, S. Pfeiffer, W. Effelsberg: Video Abstracting. Communications of ACM, 1997.
5. S.Fisher, R. Lienhart, W.Effelsberg: Automatic Recognition of Film Genres. Reihe informatik 6/95, 1995.
6. P. Wu, S. R. Kattamuri(Raja), I. Grini, G. L. Larsen: Detecting Start of Play Scenes in MPEG2 Streams of American Football. University of California, Santa Barbara, 1998.
7. Ch. Tyfonas: Mpeg-2 transport over ATM networks. University of California, Santa Cruz, 1996.
8. J. Peltoniemi: Video on Demmand Overview. Tampere University of Technology, Telecommunications Laboratory. Tampere 1995.
9. Norma ISO / IEC 13818: Coding of Moving Pictures and Associated Audio. International Organization for Standarization.

Recenzent: Dr inż. Katarzyna Stąpor

Wpłynęło do Redakcji 29 marca 2000 r.

Abstract

As soon as the Video on Demand systems enter commercial use, their administrators will meet a problem of coping with large amount of data, that needs to be added to the server database.

This work presents a project of a module for Oracle Video Server system, which performs automatic generation of a film description. This description consists of a group of words, that give an idea of the film's nature, and an abstract, composed of selected fragments of the original movie.

Word description is based on a classification of the movie to a group of intuitive types like sports, news, cartoon, or on an amount of certain events noticed in a film.

Abstract is designed by evaluating scenes in the movie for the amount of interesting events (explosions and gunshots), and amount of action. Depending on the value of the fragments (computed by evaluation functions), a group of ten of them is selected.

This way generated description is inserted into the database.

An extension of the Oracle Video Client is designed, taking advantage of the information generated, to efficiently present to the user the content of the Video Server Database.

An overview of the Oracle Video Server system is also presented.