

# 디지털 방송에서의 이벤트 추천

김준형, 조상영  
한국외국어대학교 컴퓨터공학과

## Event Recommendation in Digital Broadcasting

Jun-Hyung Kim, Sang-Young Cho  
Dept. of Computer Science, Hankuk University of Foreign Studies  
E-mail : [kimjh@humaxdigital.com](mailto:kimjh@humaxdigital.com), [sycho@hufs.ac.kr](mailto:sycho@hufs.ac.kr)

### 요 약

디지털 방송은 오디오 및 비디오 신호를 압축하여 전달 하기 때문에 같은 대역폭에 아날로그 방송보다 더 많은 방송을 전달할 수 있다. 또한 기존 아날로그 방송에서는 불가능하였던 여러 데이터 정보를 포함하여 방송을 할 수 있다. 디지털 위성 방송의 경우, 사용자가 시청할 수 있는 이벤트의 개수가 많기 때문에 사용자에게 유용한 프로그램의 선택을 도와주는 이벤트 추천 시스템은 디지털 방송에서의 유용한 기능으로 부각되고 있다. 본 논문은 디지털 방송에서 지원하는 SI(Service Information) 데이터 중에 DVB에 정의된 EIT를 정보를 기반으로 하여 경험적으로 사용자의 성향을 판단하고 사용자가 원하는 이벤트를 자동으로 추천해 주는 시스템의 설계 및 구현에 대하여 기술한다.

### 1. 서론

최근 디지털 방송에 관한 관심이 많아지고 있다. 기존의 아날로그 텔레비전 방송은 하나의 전파 대역폭에 하나의 영상 및 음성 정보 밖에 실을 수가 없었다. 그러나 디지털방송의 경우 하나의 전파 대역폭에 복수의 영상이나 음성을 압축하여 실을 수 있으며 아날로그에 비해 잡음에 대한 강인성이 뛰어나기 때문에 영상 및 음성의 품질이 우수하다. 이런 이점 외에도 디지털방송은 오디오 비디오 신호외에 디지털 방송에 사용되는 데이터 정보들을 함께 방송할 수 있다 [6]. MPEG-2의 System Layer에 들어 있는 이 데이터 정보들은 디지털 방송의 이점을 더 높여 주는 역할을 한다. EPG(Electronic Program Guide)는 이러한 데이터 정보들 중의 하나로 서비스

들의 진행 중인 이벤트 또는 진행될 이벤트들에 관한 정보를 담고 있으며 MPEG-2의 SI(System Information) 및 이를 확장한 DVB(Digital Video Broadcasting)-SI(Service Information) 규격에 정의되어 있다.

디지털 방송에 경우, 사용자가 시청할 수 있는 방송의 수가 많기 때문에 시청자가 자신에 유용한 이벤트를 찾아 시청하는데 많은 노력이 필요하다. 실제로, 유럽의 경우를 보면 하나의 위성에서 나오는 디지털 방송 서비스를 모두 수신하면 1000 개 가까이 된다. 각 서비스를 당 하루에 해당하는 이벤트의 개수가 약 20 개 정도 이기 때문에 하루에 20000 개 정도의 이벤트 중에서 자신에게 가장 적합한 방송을 찾아야 한다. 본 논문은 이들 이벤트를 분석하고 사용자의 취향 및 성향을 경험적으로 판단해서 사용자

가 선호하는 이벤트를 자동 추천 해 주는 이벤트 추천 시스템의 설계와 구현에 대하여 기술한다.

2절에서는 MPEG-2의 System Layer 및 이를 확장한 DVB-SI에 관해 소개하고 3절에서는 기존의 STB(Set-top Box)에서의 이벤트의 선택에 관한 내용을 언급한다. 4절에서는 이벤트 추천을 위한 방법 및 구현상의 내용을 기술한다. 5절에서 결론을 맺는다.

## 2. 관련 규격

STB는 기존의 아날로그 방식의 TV를 통해 디지털 방식의 방송을 수신할 수 있도록 도와주는 장치이다. 디지털 방송 정보가 수신 되면 STB에서 수신된 정보를 분석하고 이 정보를 아날로그화 해서 아날로그 텔레비전에 보내주게 된다.

본 논문에서는 여러 가지 디지털 방송의 전송 방식과 데이터 구조 중에서 유럽에서 상용화 되어 있는 DVB-SI를 기반으로 한 이벤트 추천 시스템에 대해 다룬다.

### MPEG System Layer

MPEG-2 시스템에는 PS(Program Stream)과 TS(Transport Stream) 두 종류의 방식이 있다. PS는 MPEG-1과 같이 하나의 프로그램을 하나의 비트열로 구성하는 방식이고 TS는 복수의 프로그램을 하나의 비트열로 구성하는 방식이다. 따라서 TS는 일반 TV 방송에 대응할 수 있고 자유도가 많은 편성이나 스크램블 기능에도 용이하게 적용될 수 있다. 또한 MPEG-2 비디오에 대응하기 위한 확장기능이나 각종 응용을 위한 부가기능을 갖고 있다. 따라서 디지털 방송에서는 TS 방식을 사용하여 서비스를 제공한다.

TS의 복호, 재생에 있어서는 한 비트열의 복수 프로그램 중에서 하나의 서비스를 선택하고 방송 중인 프로그램을 복호, 재생해야 한다. 이를 위해 개별 비트열의 트랜스포트 패킷의 PID와 이 패킷들의 개별 비트열의 파라미터 정보와 연계정보를 알아야 한다. 이와 같은 다단계 동작을 위해 필요한 것이 PSI(Program Specific Information)이다. PSI는

Section이라 불리는 데이터 구조에 의해 전송되는데 이것은 PAT(Program Association Table), PMT(Program Map Table), CAT(Conditional Access Table)로 나누어져 있다[1]. 디지털 방송에서는 이들 세 개의 표 정보를 바탕으로 서비스를 구성하게 된다.

### DVB-SI

사용자를 위한 서비스 및 이벤트를 제공하기 위한 데이터를 지원하기 위해 ETSI(European Telecommunication Standard Institute)에서 PSI에 추가한 정보가 DVB-SI이다. PSI에 있는 PAT, CAT, PMT와는 반대로 이들 정보는 actual multiplex가 해당되어 있는 것의 정보만을 보내준다. 이들은 9가지 테이블로 나누어져 있는데 BAT, SDT, EIT, RST, TOT, ST, SIT, DIT 등이다. 이들 중 대표적으로 많이 사용하는 것은 표 1의 BAT(Bouquet Association Table), SDT(Service Description Table), EIT(Event Information Table), TDT(Time and Date Table)이다[2].

표 1 SI 정보

표	설명
BAT	Bouquet에 관한 정보를 의미하며 Bouquet의 이름 뿐만 아니라 각 Bouquet이 가지고 있는 서비스의 리스트도 제공한다.
SDT	서비스의 이름이나 서비스의 제공자 정보 등을 나타내는 서비스 정보가 들어 있다.
EIT	이벤트의 이름, 이벤트의 시작시간, 이벤트의 끝나는 시간들의 정보들과 같은 이벤트의 정보에 관계된 데이터를 가지고 있다.
TDT	현재 날짜 및 시간에 관한 정보를 가지고 있으며 이 정보들은 모든 이벤트 관리를 위한 기본 정보가 된다

EIT의 내용 중에서 이벤트 추천 시스템에 사용되는 필드는 표 2에 정리하였다.

표 2 추천시스템에서 사용되는 EIT 필드

필드 이름	필드 설명
Event_id	16bit 데이터형으로 각각의 이벤트를 숫자로 구분한 정보.
Start_time	40bit 데이터형으로 MJD (Modified Julian Date)로 날짜 및 시간을 나타낸 정보
Duration_time	24bit 데이터 형으로 BCD 형식으로 방영시간을 나타냄.
Parental-rating	관람 등급
Event_title	이벤트의 제목
Event_text	이벤트의 내용
Extended_event	이벤트의 줄거리

### 3. STB에서의 이벤트 관리

STB는 A/V정보의 재생 외에도 DVB-SI규격에서 정의 하는 여러 다른 정보들을 분석하여 사용자가 사용 가능한 정보로 만들어 주는 기능을 포함하고 있다. 그림 1은 기존의 STB에서 DVB-SI에서 정의 하는 데이터를 구조화 하는 부분을 나타낸 것이다.

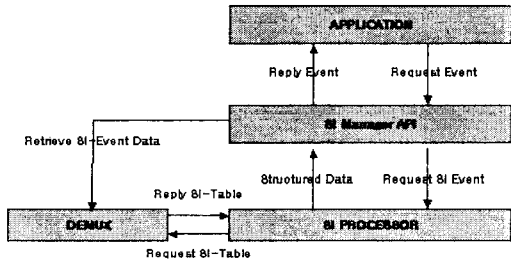


그림 1 기존의 SI 데이터 처리 방식

어플리케이션이 요구 하는 어떤 정보를 받기 위해서는 먼저 SI Manager가 DEMUX에게 SI중에 어플리케이션이 요구하는 정보에 관한 필터링을 지시한다.

DEMUX에서는 해당 정보를 위한 필터링을 시작하고 필터링된 표 형태의 데이터를 SI Processor에 전달한다. SI Processor는 이들 데이터를 다시 구조화된 형태의 데이터로 만들고 SI Manager에서 이 정보를 어플리케이션에서 요구 하는 형태의 정보로 데이터

베이스화 한다. 이후에 SI Manager는 어플리케이션이 요구하는 정보를 전달한다.

### 4. 이벤트 추천 시스템의 구성

#### 4.1. 이벤트 스케줄러의 동작

이벤트 추천을 위해서 기존 STB의 데이터 처리방식에서 그림 2와 같이 Event Scheduler를 추가하였다.

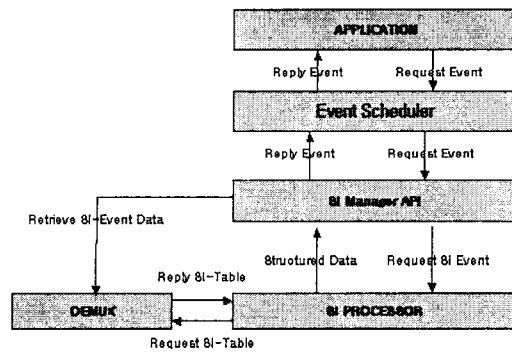


그림 2 이벤트 추천을 위한 SI 데이터 처리 방식

이벤트 스케줄러는 어플리케이션의 요구에 따라 SI Manager로부터 받은 이벤트 정보를 바탕으로 이벤트 데이터베이스를 작성하고 데이터베이스화된 정보를 어플리케이션에 전달하여 이벤트 데이터베이스를 바탕으로 사용자의 성향을 판단하는 역할을 수행한다.

이벤트 스케줄러의 동작은 크게 사용자가 이벤트를 선택했을 때의 처리와 사용자가 이벤트의 추천을 요구할 때의 처리 두 부분으로 나누어진다.

#### 4.1.1 사용자의 이벤트 선택

이벤트 스케줄러는 사용자가 선택한 이벤트를 기준으로 사용자의 성향을 분석한다. 그림 3은 사용자가 이벤트를 선택했을 때 처리 순서를 나타낸 것이다.

사용자가 하나의 이벤트를 선택하면 이벤트 스케줄러는 해당 하는 이벤트의 정보를 SI Manager에게 요구하고 SI Manager에서는 해당하는 이벤트의 정보를 돌려준다. 이벤트 스케줄러는 반환받은 이벤트 정보에서 시간 정보와 Parental Rating 정보를 기반으로

1차적인 정보의 분류를 한다. 그리고 해당 이벤트에서 Title, Text, Extended Event를 기반으로 하여 2차적인 정보 분류를 수행한다.

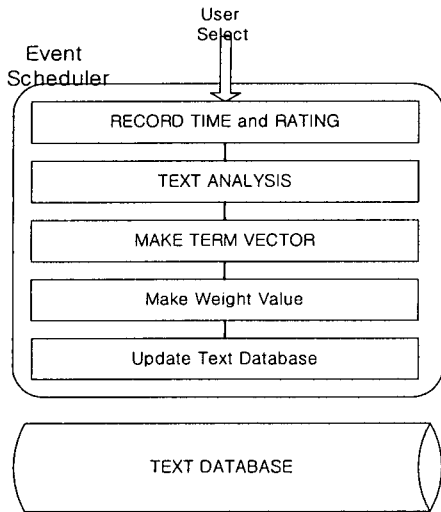


그림 3 이벤트 선택 시의 처리 순서

2차 정보 분류는 Text 자체를 분석하는 방식으로 이루어 지는데 이는 4가지 단계로 이루어 진다. 첫 번째, 이벤트에 있는 내용들을 각각의 유일한 단어로 나눈다. 두 번째, 단어의 글자 수가 4 글자 이내이거나 각각의 이벤트에 대해 공통적으로 많은 빈도가 나타나는 단어를 없앤다. 이들 단어는 보통 식별자이거나 그 단어 자체로는 이벤트의 특성을 확인하는데 필요가 없는 단어이다. 이 과정을 거친 후 남은 단어에 한하여 접미사를 제거해서 이벤트의 특성 파악을 더욱 명료하게 한다[4]. 여기서 걸러지는 단어는 StopList에 추가하여 다음 번 분석 시에도 활용할 수 있게 한다. 세 번째, 전체 단어 리스트에서 남아 있는 단어에 대해서 가중치를 조정한다. 마지막으로 높은 가중치를 갖는 단어를 기준으로 Text 데이터베이스를 갱신한다.

단어 빈도의 가중치를 구하는 방법에는 여러 가지가 있으나 여기서는 더블로그 TF 방식을 사용한다. 더블로그 TF 는 로그 TF 방식의 발전된 형태로서 질의가 소수의 질의어로 구성된 경우에는 로그 TF 로도 tf 가 높은 단어의 지나친 영향력을 낮추는 것이 불

충분하다고 판단하여 Singhal 등이 TREC-7 에서 제안한 공식이다[5].

$$TF = 1 + \log(1 + \log(tf))$$

이렇게 해서 구해진 단어의 가중치를 전체 데이터 베이스에 적용하기 위해서는 전체적인 가중치를 다시 계산 하여야 한다. 이는 현재 까지 본 이벤트들 중에 선택된 단어를 사용하는 이벤트의 개수와 그들 이벤트에서 구해진 TF 를 고려하여 전체적인 가중치를 다시 부과하는 것을 의미 한다. 이때 아래와 같은 공식을 사용하여 가중치를 계산한다.

$$WordFreq = \frac{Total\ Reference * TF}{nDocument}$$

#### 4.1.2 사용자의 이벤트 추천 요구

사용자가 예정된 이벤트들 중에서 자신의 성향에 맞는 이벤트의 추천을 요구하면 과거의 사용자의 이벤트 선택으로 데이터베이스화 되어있는 정보를 바탕으로 추천이 이루어 진다. 그러나 디지털방송의 특성상 하나의 TP(Transponder)에 저장될 수 있는 서비스는 대략 20 정도이고 한번에 받을 수 있는 정보 또한 이들 20 개의 서비스에 해당하는 정보만 받을 수 있다. 따라서 사용자가 요구하는 정보에 보다 근접한 정보를 제공하기 위해서는 사용자가 볼 수 있는 TP 에 있는 정보들을 모두 확인해 볼 필요가 있다. 이를 가능하게 하기 위해서는 두 가지 방법이 있으며 하나는 저전력 모드 기능을 이용하는 방법이고 다른 하나의 특정 시간대에 STB 가 자동으로 모든 TP 를 스캐닝하는 방법이다. 저전력 모드를 이용하는 경우는 STB 의 전원이 꺼졌을 때 전체 전원을 꺼는 것이 아니고 SI 에 관계된 데이터 처리를 위한 부분은 전원을 끄지 않고 지속적으로 SI 정보를 받아서 데이터베이스화 하도록 한다. 두 번째 방법은 STB 가 스탠바이 모드에 있을 경우 사용자가 입력한 특정 시간대에 STB 는 자동으로 Power ON Mode 가 되어 자동으로 TP 을 스캐닝하여 SI 데이터를 처리하고 모든 데이터의 처리 후에 다시 Standby Mode 로 돌아 가는 방식이다. TP 스캐닝 시에 사용자에게 추천할 이벤

트의 정보가 만들어 진다. 두 번째 방법이 첫번째 방법보다 Side-Effect 가 적으므로 여기서는 두 번째 방법을 사용한다.

그림 4 는 이벤트의 추천을 위한 순서를 도식화 한 것이다.

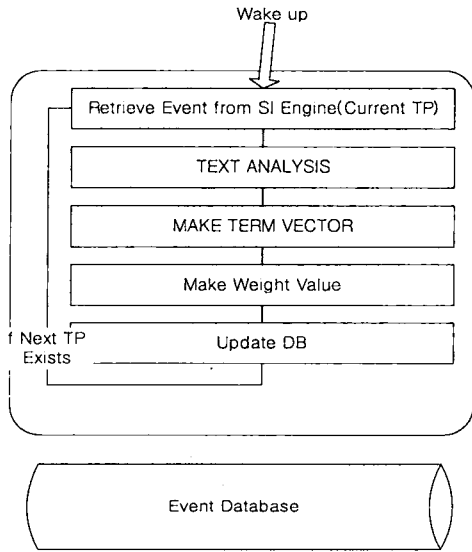


그림 4 이벤트 추천을 위한 TP 스캐닝

추천을 위한 TP 스캐닝 중의 이벤트의 분석은 5 개의 단계를 거친다.

1. 사용자가 정의한 시간대에 STB 는 Wake-up 이 된다.
2. Wake-Up 후의 이벤트 스케줄러는 현재의 TP 에 대한 SI 정보를 요청하고 이에 대한 이벤트 정보를 SI Manager 로부터 반환 받는다.
3. 반환된 이벤트 정보를 바탕으로 Text 를 분석한다.
4. 분석된 단어 단위의 정보를 기반으로 Term Vector 를 작성한다.
5. 각각의 이벤트 별로 단어에 대한 가중치를 부과한다. 이때의 가중치는 Text 데이터베이스에 정의된 단어 단위의 가중치에 기반하여 작성 된다.
6. 이벤트에 대한 단어단위의 가중치와 시간 및 Parental Rating 정보를 이벤트 데이터베이스에 Update 한다.

7. 현재 TP 외에 사용자가 수신 가능한 다른 TP 가 있는지를 보고 만약 있다면 2 번의 작업 부터 다시 반복한다.

이벤트 데이터베이스에는 사용 가능한 이벤트에 대한 이벤트의 ID 와 단어 단위의 분석을 마친 가중치의 값, 그리고 시간 및 Parental Rating 의 정보만 저장된다.

사용자가 이벤트의 추천을 요구할 때 이벤트 스케줄러는 기존에 작성된 이벤트 데이터베이스를 바탕으로 하여 추천 가능한 이벤트리스트를 작성하여 사용자에게 알려준다.

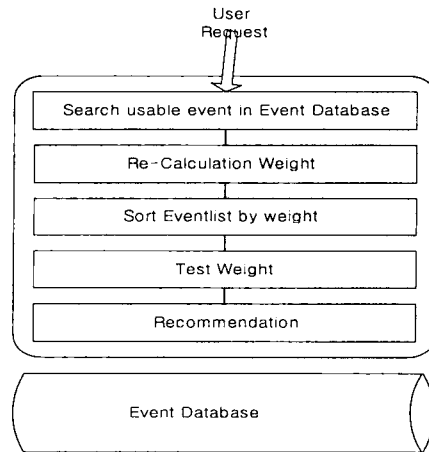


그림 5 사용자의 요청에 의한 이벤트의 추천

그림 5 는 사용자의 요청에 의한 이벤트의 추천과정을 나타낸 것이다. 사용자가 이벤트의 추천을 요구 하면 이벤트 스케줄러는 이벤트 데이터베이스에 있는 정보들 중에서 현재 시간을 기준으로 사용 가능한 이벤트 리스트를 추출해 낸다. 추출된 이벤트 정보를 바탕으로 시간 및 Parental Rating 의 정보를 추가한 새로운 가중치를 재계산하게 되고 재계산된 가중치정보를 기준으로 순차정렬을 한다. 정렬된 이벤트리스트에서 현재 시간에 맞는 이벤트를 추천해 준다.

#### 4.2 Data Sampling

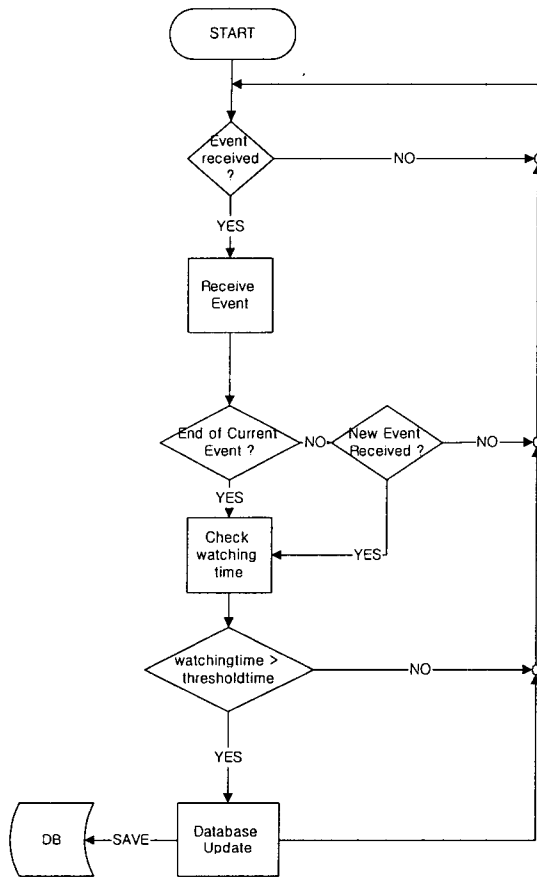


그림 6 이벤트 샘플링

사용자가 시청하는 이벤트를 기준으로 모든 정보들이 구조화 되기 때문에 이벤트에 대해서 유효한 이벤트인지 유효하지 않은 이벤트인지를 판단할 필요가 있다. 예를 들면 시청하는 서비스가 아니고 단지 검색을 위해 서비스를 전환하고 있을 때는 유효하지 않은 정보이다. 그림 6은 이벤트 샘플링 부분을 도식화 한 것이다. 새로운 이벤트가 들어올 경우 이 이벤트가 사용자가 보는 이벤트인지 아니면 단순히 서비스 검색 시에 들어 온 이벤트인지 검사한다. 그리고 마지막으로 본 이벤트에 대한 시청 시간을 임계시간과 비교해서 데이터로서의 가치가 있는지 없는지를 판단한다. 그 후, 이 조건들을 만족할 때 사용자의 이벤트 선택이라고 판단하고 이벤트 스케줄러에게 알려 준다.

STB에서의 이벤트 추천 시스템은 Window 2000-

Professional 환경에서 Visual C++ 6.0을 이용해서 구현하였으며 테스트 데이터는 [www.tvguide.com](http://www.tvguide.com)에서 제공되는 이벤트 정보를 추출하여 사용하고 있으며 윈도우 환경에서 실험 중에 있다.

### 5. 결론

본 논문에서는 디지털 방송에서 사용자에게 이벤트를 추천 하는 시스템에 대한 설계에 대하여 기술하였다. 본 이벤트 추천 시스템은 기존의 STB의 SI 데이터 처리 부분에 이벤트 스케줄러를 삽입하여 사용자의 이벤트 선택에 따른 정보를 저장하고 이를 학습하여 사용자의 취향 또는 성향을 분석한다. 이 데이터를 바탕으로 방송 이벤트 정보에서 사용자에게 유용한 방송을 추천한다. 향후 연구계획으로는 MRU(Most Recently Used)방식 이외에 Neural Network을 이용한 가중치의 적용방법과 Embedded System에서의 제한적인 Resource에 따른 Text Database관리에 관한 것이 있다.

### [참고문헌]

- [1] "GENERIC CODING OF MOVING PICTURES AND ASSOCIATED AUDIO : SYSTEMS", ISO 13818-1.
- [2] "Digital Video Broadcasting (DVB); Specification for Service Information (SI) in DVB systems", EN 300 468 v1.3.1, <http://www.dvb.org>.
- [3] Y. S. CHOI and S. I. YOO, "Text Database Discovery on the web: Neural Net Based Approach.", Journal of Intelligent Information Systems, 16, pp.5-20, 2001.
- [4] I. Khan, D. Blight, R.D. McLeod, and H.C. Card "Categorizing Web Documents using Competitive Learning: An Ingredient of a Personal Adaptive Agent", Proc. of the 1997 Int. Conf. on Neural Networks (ICNN'97), Vol 1, pp.6 -9, 1997.
- [5] A. Singhal, et al., AT&T at TREC-7. Proceedings of the Seventh Text Retrieval Conference (TREC 7): pp.239-251, 1998
- [6] F. Baylin, "The Digital Revolution", Baylin Publications, 1995.