

Media Orchestration 에서의 공간적 Orchestration 기술 방안

이의상, 강동진, 강전호 김규현
경희대학교

euisang91@khu.ac.kr, cpffh0729@khu.ac.kr, gaonam@khu.ac.kr,
kyuheonkim@khu.ac.kr

Method of Spatial Orchestration in Media Orchestration system

Euisang Lee, Dongjin Kang, Jeonho Kang, Kyuheon Kim
Kyung Hee Univ.

요 약

미디어의 생성 및 소비를 위한 기기의 다양화 및 소형화와 이로 인한 미디어 생성 및 소비의 패러다임이 변화함으로 인해, 이와 같은 시장의 요구사항에 부합하기 위한 다양한 기술의 개발 및 표준화가 가속화되고 있다. 이러한 시대의 변화에 발맞추어 국제 표준화 기구인 MPEG 은 네트워크 상에서의 다양한 미디어 간 시공간적 관계를 파악하고 관리하기 위한 표준인 Media Orchestration 표준화를 진행하였다. 이에 본 논문에서는 Media Orchestration 에서 제시한 미디어 간 공간적 Orchestration 구조를 기반으로 다양한 Metadata 를 이용하여 단말 및 미디어 간의 공간적 관계의 기술, 전달 및 소비 방안을 제시하였다.

1. 서론

전통적인 미디어의 생성 및 소비 방식은 전문적인 장비를 보유한 방송사가 제작한 미디어를 방송을 통해 일방적으로 소비자에게 전달되는 전형적인 PULL 서비스 방식이었다. 하지만, 미디어를 제작할 수 있는 기기가 소형화 및 다양화되고, 일반인들에게도 널리 보급되고, 또한 네트워크 기술의 빠른 발전에 따라 미디어의 생성 및 소비 방식은 매우 다양화되었으며, 미디어 생성과 소비의 주체의 경계가 모호해지고 있다. 이러한 미디어 생성 및 소비 형태의 변화는 미디어 시장의 수요와 공급의 변화를 가져왔으며, 이에 따라 다양한 기업, 표준화 단체 등에서 다양한 서비스를 제공하기 위한 기술의 개발에 힘쓰고 있다.

국제 표준 기구 ISO 산하의 동영상의 압축 및 전송, 소비 등의 표준을 제정하는 MPEG(Moving Picture Expert Group)은 이와 같은 시장의 변화에 발 맞추어 2015 년 네트워크 상에서 다양한 형태의 미디어들의 시공간적 관계를 파악하고 이를 관리하기 위한 표준인 MPEG-B Part 13 Media Orchestration[1]의 표준화를 시작하여 2017 년 7 월에 Draft of International Standard(DIS) [2]를 발행하였다. 해당 표준은 다양한 기기 간의 동기화, 콘텐츠 인식 등에 관한 표준인 DVB(Digital Video Broadcasting)-CSS(Companion Screens & Stream)[3] 및 MPEG 의 미디어 저장 표준인 ISO Base Media File Format(ISOBMFF) [4], 미디어의 전송 및 저장 프로토콜인 MPEG-2 Transport Stream(TS) [5], 다양한 메타데이터에 관한 표준인 MPEG-V[6] 등을 참조 및 확장하는 동시에 네트워크 상에서의 미디어의 전송 및 관리를 위한 Control Architecture, Spatial Orchestration 등을 정의하고 있다.

본 논문에서는 이와 같은 Media Orchestration 내 Spatial Orchestration 의 구조를 기반으로 다양한 단말의 Metadata 를 활용하여 미디어 간 Spatial Orchestration 을 기술하는 Orchestration 데이터를 생성하고 이를 전달하는 것으로 Orchestrated Media 를 소비하는 방안을 제안한다.

2 장에서는 본 논문에서 제안하는 방안을 이해하기 위한 배경 기술로써 Media Orchestration 에 대하여 설명하고, 3 장에서는 본 논문에서 가정하는 서비스 시나리오와 이와 같은 서비스를 제공하기 위한 방안을 제시하고, 4 장에서 실험결과를 보인 후, 5 장의 결론으로 본 논문의 결론을 맺는다.

2. Media Orchestration

본 장에서는 본 논문에서 제안하는 기술의 바탕이 되는 Media Orchestration 에 대하여 자세하게 기술한다. 앞서 소개한 바와 같이, 2015 년 표준화를 시작한 Media Orchestration 은 현재 DIS 가 발행된 상태이며, 해당 DIS 에 대한 Ballot 을 진행중이다. 본 장의 1 절에서는 Media Orchestration 의 기본이 되는 구조에 대하여 설명하고, 2 절에서는 Media Orchestration 에서 활용하는 다양한 메타데이터의 소개 및 그들의 저장 및 전송 방안에 대하여 설명한다.

1. Media Orchestration 기본 구조

Media Orchestration 은 네트워크 상에서 다양한 미디어 간의 관계를 파악하고, 이들을 관리하기 위한 표준이다. 이와 같은 네트워크 상에서의 다양한 미디어의 생성과 소비 및 미디어 들의 관리, 단말 간의 관계 등을 정리하기 위해, Media Orchestration 표준은 4 개의 기본 구성요소(element)를

정의하였다.

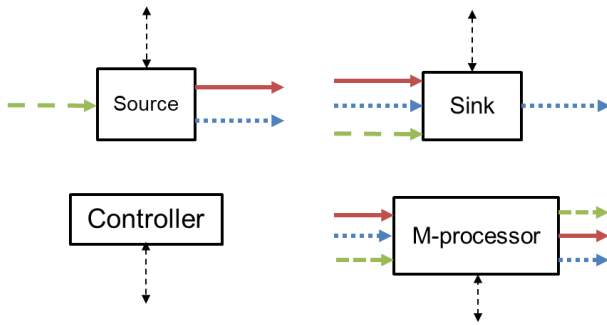


그림 1. Media Orchestration 구성요소

표 1 Media Orchestration 데이터 타입.

	Media data (timed or non-timed, unidirectional)
	Metadata (timed or non-timed, unidirectional)
	Orchestration data (timed or non-timed, unidirectional)
	Control data for messaging and control (non-timed, possibly bi-directional)

그림 1 은 앞서 소개한 Media Orchestration 에서 정의한 4 개의 기본 구성 요소를 도식화한 그림이다. 그림 1 에서 보이는 바와 같이, Media Orchestration 에서는 4 가지의 구성요소(element)를 정의하고 있으며, 각 요소의 이름은 Source, Sink, M-processor, Controller 이다. 또한, 표 1 은 그림 1 에서 화살표로 표시된 다양한 데이터들의 타입을 정리한 표이다. Media Orchestration 에서는 데이터의 타입을 4 가지로 나누어 관리하며, 각 데이터들의 명칭은 각각 미디어 데이터, 메타데이터, Orchestration 데이터, 컨트롤 데이터이다.

먼저, 그림 1 의 ‘Source’ (이하 소스)는 미디어를 생성(Capture)하는 구성요소이다. 소스는 Orchestration 데이터 및 컨트롤 데이터를 입력으로 받을 수 있으며, 미디어 데이터와 메타데이터를 생성하고 이들을 출력으로 내보내는 역할을 한다. 두 번째로, 그림 1 의 ‘Sink’ (이하 싱크)는 미디어를 소비(Consumption)하는 구성요소이다. 싱크는 Orchestration 데이터, 미디어 데이터, 메타데이터를 입력으로 받아 유저가 미디어를 소비할 수 있게 하는 구성요소이며, 다양한 Orchestration 을 위하여 메타데이터를 출력으로 내보낼 수 있다. 세 번째로, 그림 1 의 ‘M-processor’ 는 미디어 및 메타데이터를 활용한 프로세스들을 처리하는 구성요소이다. M-processor 는 미디어 데이터, 메타데이터, Orchestration 데이터 모두를 입력으로 한다. 클라우드 서버 등은 물론 스마트폰, PC 와 같은 단말도 M-processor 가 될 수 있다. 마지막으로, 그림 1 의 Controller 는 하나 이상의 구성요소들을 관리하는 역할을 하며, 각 구성요소와 컨트롤 데이터를 이용하여 통신할 수 있다. 컨트롤러는 각 요소들을 관리하는 역할만 하며, 다른 입출력 데이터는 존재하지 않는다. 컨트롤러는 각 구성요소들의 발견, 등록, 단말 간의 연결 등을 관리하는 역할을 한다.

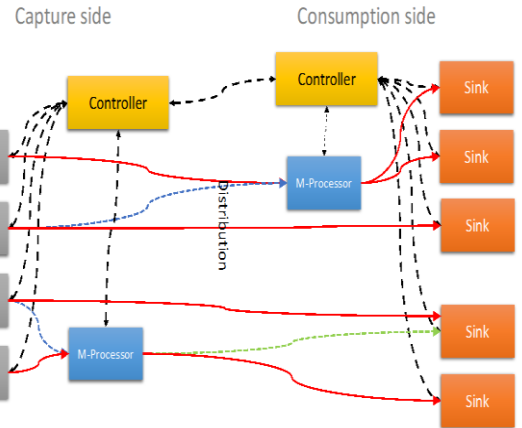


그림 2. Media Orchestration 세션 내 데이터 흐름 예시

그림 2 는 Media Orchestration 세션 내 각 요소들 간 데이터의 흐름의 예시를 나타낸 그림이다. Media Orchestration 세션이란 미디어의 생성(Capture)부터 미디어 프로세싱 (Processing) 및 미디어 렌더링/소비 (Consumption)을 포함하는 일련의 연관된 응용 계층의 컨텍스트를 의미한다. 그림 2 에서 보이는 바와 같이, Media Orchestration 세션은 2-1 절에서 소개한 4 가지 구성요소로 이루어진다. 그림 2 에서 보이는 바와 같이, 미디어 및 메타데이터는 소스로부터 싱크로 직접적으로 전달될 수도 있으며, M-processor 를 거쳐 미디어 프로세싱 이후 전달될 수도 있다. 다만, Orchestration 데이터는 M-processor 에서만 생성될 수 있으며, 이와 같은 Orchestration 데이터는 소스 혹은 싱크로 전달되어 다양한 역할을 수행할 수 있다.

2. Media Orchestration 내 메타데이터

Media Orchestration 표준은 다양한 미디어의 관계를 파악하고 이를 관리하기 위하여 각 미디어와 연관된 다양한 메타데이터를 소개 혹은 정의하고, 이들의 저장 및 전송 방식을 제안하고 있다. Media Orchestration 표준에서 이야기하는 메타데이터란, 독립적으로 렌더링 될 수 없지만, 다양한 방식으로 다른 미디어 데이터의 렌더링, 프로세싱, Orchestration 에 영향을 줄 수 있는 데이터를 의미한다. Media Orchestration 표준에서는 메타데이터를 ISOBMFF [4]와 MPEG-2 TS [5]의 두 가지 방식으로 저장하는 방안을 정의하고 있으며, MPEG Media Transport [7] 및 MPEG-DASH [8]를 이용한 전송 방안을 제공하고 있다.

Media Orchestration 표준에서 명기되어 있는 메타데이터는 GPS 와 같은 위치 정보에 대한 메타데이터인 ‘Position Metadata’, 기기의 기울임 정도 등에 관한 메타데이터인 ‘Orientation Metadata’, 영상 내 특정 관심 영역에 대한 메타데이터인 ‘Regions of Interest’ [9], 다양한 미디어 스트림 중 특정 스트림을 선택하기 위한 메타데이터인 ‘Quality Metadata’, DVB-CSS 의 Correlation Timestamp [3]을 전달하기 위한 메타데이터인 ‘Stream Monitor’, 특정 프레임 내 특정 부위를 지정하는 데 사용되는 메타데이터인 ‘Capture Mask’, 초점거리, 셔터속도 등 다양한 카메라 고유의 정보를 전달하기 위한 메타데이터인 ‘Camera Metadata’, 영상의 깊이정보를 포함하는 메타데이터인 ‘Depth Metadata’ 의 8 가지이다.

3. Spatial Orchestration 방안 제안

본 장에서는 이와 같은 Spatial Orchestration 을 진행하는 데 있어 기본이 되는 구조를 제안하고, 이에 합당하는 Spatial Orchestration Description 를 제안하고자 한다. 본 장에서 제안하는 구조는 Media Orchestration 그룹에 기술 기고를 마쳤고 그룹의 합의를 통해 DIS(Draft of International Standard)에 반영되었으며, Spatial Orchestration Description 는 TuC(Technology under Consideration)로 채택되어 추후 기고 및 논의를 통해 Amendment 로써 표준에 반영되는 것으로 결정된 바 있다.

1. Spatial Orchestration 구조

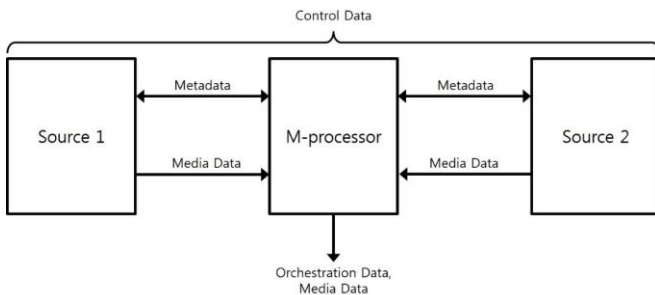


그림 3. M-processor 단에서의 Spatial Orchestration

그림 3 은 M-processor 단에서의 Spatial Orchestration 을 단순화하여 도식화한 그림이다. M-processor 는 하나 이상의 소스로부터 미디어데이터를 수신함과 동시에 다양한 메타데이터를 이용하여 소스와 통신하며, 미디어데이터와 이에 관련된 Orchestration 데이터를 출력으로 내보낸다. M-processor 의 출력은 메타데이터도 포함할 수 있지만, 해당 구조에서 이는 소스와의 통신을 위한 것으로 한정한다. 이와 같은 모든 동작은 Controller 가 관리 하에 진행된다.

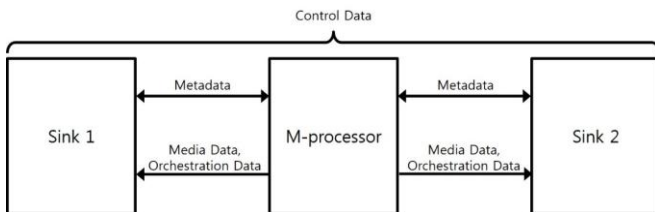


그림 4. 싱크 단에서의 Spatial Orchestration

그림 4 는 싱크 단에서의 Spatial Orchestration 을 단순화하여 도식화한 그림이다. 싱크는 M-processor 로부터 미디어 데이터, 메타데이터 및 Orchestration 데이터를 수신하여 렌더링하는 역할을 한다. 이 과정에서, 싱크는 메타데이터를 이용하여 M-processor 와 통신하며, 이와 같은 모든 동작은 Controller 가 관리 하에 진행된다.

2. Spatial Orchestration Description

본 절에서는 앞서 설명한 Spatial Orchestration에 사용되는 Orchestration 데이터로 Spatial Orchestration Description(이하 SOD)의 구조를 제안한다. Spatial Orchestration을 위한 Orchestration 데이터는 M-processor가 파악한 다양한 미디어 데이터, 메타데이터와 같은 데이터와 싱크, 소스 등의 Media Orchestration 구성 요소 간의 관계를 기술할 수 있어야 한다.

이에 본 절에서 제시하는 SOD는 트리 구조를 이용하여 각 요소들의 관계를 정의하고, XML [?] 문법을 이용하여 이를 기술하여 제공한다.

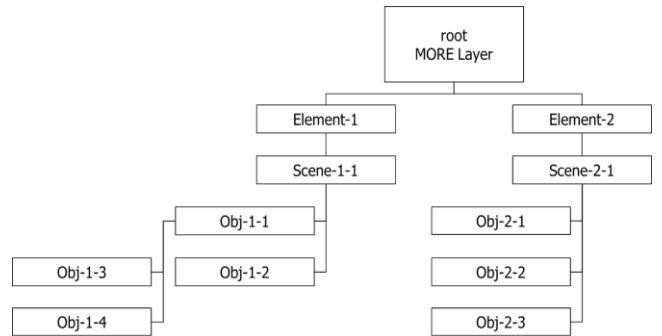


그림 5. Media Orchestration 의 Spatial Orchestration 트리 구조

그림 5 는 Spatial Orchestration 을 수행함에 있어, 각 데이터 및 구성 요소들의 관계를 나타내기 위한 트리 구조를 도식화한 그림이다. 해당 구조는 Element, Scene, Object 의 가지 노드를 가지는 계층적 구조를 지니고 있으며, MORE Layer 는 Media Orchestration 세션을 의미한다. 이와 같은 Spatial Orchestration 트리는 M-processor 에 의해 생성되어야 하며, 다양한 미디어 데이터와 이와 연관된 메타데이터, Media Orchestration 의 구성요소로 구성된다.

Spatial Orchestration 트리 구조의 Element 노드는 Media Orchestration 의 싱크, 소스와 같은 구성 요소를 나타내며, Scene 은 다양한 미디어 데이터와 메타데이터 등으로 구성되는 하나의 서비스를 의미한다. 마지막으로, Object 는 비디오, 오디오 등과 같은 미디어 데이터 및 GPS, 기울기 센서 정보 등 다양한 메타데이터를 의미한다. 표 2 는 이와 같은 계층 구조를 XML [10]으로 기술하기 위한 문법의 예시를 나타낸다. 각 노드는 모두 고유의 id 로 구분 될 수 있으며, Element 노드와 Object 노드는 Type 을 가지는 것으로 포함하는 구성 요소 혹은 데이터 타입을 구분할 수 있게 하였다. Scene 의 'SplitSceneFlag' 어트리뷰트는 하나의 서비스가 다양한 미디어로 구성되는 경우를 상정하여 정의하였다.

표 2 Spatial Orchestration Description 문법 예시

Element or Attribute Name	Use
Element	
@id	M
@ElementType	M
Scene	M
@id	M
@SplitSceneFlag	OD
Object	1...N
@id	M
@ContentType	M
Object	0...N

4. 실험 결과

본 장에서는 앞서 소개한 Media Orchestration 세션에서의 공간적 Orchestration 을 제공하기 위한 SOD 를 생성하고 이를 파싱하여 미디어를 재생하는 것으로 제안 방안을 검증한다.

본 장에서 보이는 실험 결과를 위한 개발은 모두 Windows 10 환경에서 이루어졌으며, Visual Studio 2015 를 활용하여 개발되었다. 본 실험에서 사용된 콘텐츠는 동일한 이벤트에 대한 다시점 영상의 재생을 위한 콘텐츠로 모든 콘텐츠는 나고야 대학의 Tanimoto Lab 에서 촬영되었으며 학술적 목적만으로서 허가 사용을 한 콘텐츠임을 명시한다[11]. 나고야 대학의 Tanimoto Lab 에서 제공하는 콘텐츠 중 본 실험에 사용한 콘텐츠는 “Dog” 시퀀스 중 10 번, 20 번, 30 번 카메라로 촬영된 영상이다.

```

<Spatial_Orchestration_Description>
  <Element type="Element_type" id="Element_id">
    <Scene id="Scene_1" SplitSceneFlag="1">
      <Object id="0" file_path="010.yuv" h="960" w="1280" y="0" x="0" type="Video"/>
    </Scene>
    <Scene id="Scene_2" SplitSceneFlag="1">
      <Object id="1" file_path="020.yuv" h="960" w="1280" y="0" x="100" type="Video"/>
    </Scene>
    <Scene id="Scene_3" SplitSceneFlag="1">
      <Object id="2" file_path="030.yuv" h="960" w="1280" y="0" x="200" type="Video"/>
    </Scene>
  </Element>
</Spatial_Orchestration_Description>
    
```

그림 6. SOD 생성 예시

그림 6 은 SOD 를 생성한 예시를 나타낸 그림이다. Element 및 각 Scene 과 Object 들은 모두 고유 식별자를 가지며, Object 의 경우 비디오 데이터를 포함하므로 미디어의 위치, 넓이와 높이, 렌더링 위치 등을 추가하였다. 여러 Scene 을 합쳐 하나의 서비스를 제공하도록 하기 위해 “SplitSceneFlag” 어트리뷰트는 1 로 설정하여, 복수 개의 Scene 이 하나의 서비스를 제공함을 표현하였다.

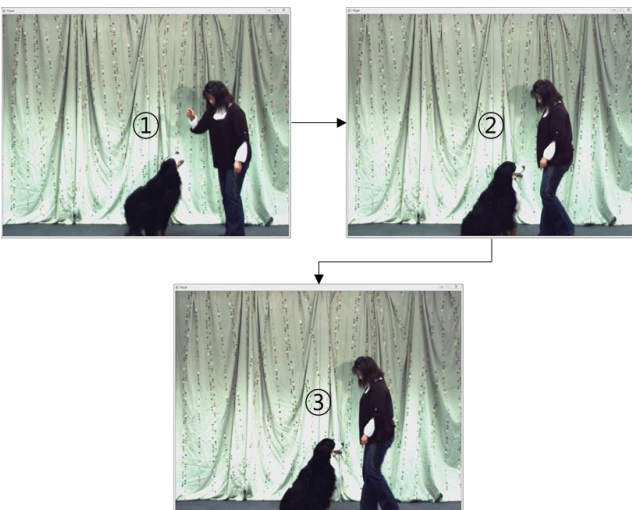


그림 7. SOD 미디어 재생 화면

그림 7 은 SOD 를 기반으로 한 미디어의 재생을 나타내는 그림이다. 그림 7 에서 표시된 ①~③은 각각 1~3 시점의 미디어를 나타낸다. 각 시점의 미디어에 대한 정보는 SOD 로부터 받아들일 수 있다. 플레이어는 SOD 에 따라, 미디어의 위치에서 미디어를 읽어 미디어의 재생을 시작한다. 먼저 제일 첫번째 미디어인 1 시점의 미디어를 재생하다가, 사용자가 보고 싶은 시점을 선택하면 다른 시점의 미디어를 재생하도록 하였다.

5. 결론

네트워크 상에서의 다양한 미디어를 관리하고 서비스하기 위하여 Media Orchestration 은 2015 년 이후 미디어의 동기화를 포함한 시공간적 연관성 관리에 대한 연구를 지속해왔다. 이에 본 논문에서는 미디어 간 공간성 연관성을 제공할 수 있는 공간적 Orchestration 데이터로 SOD 를 제안하였다.

이와 같은 미디어 간 공간적 Orchestration 기술은 다수의 입력을 기반으로 수많은 사용자에게 서비스를 제공하는 데 있어 큰 역할을 할 수 있을 것으로 보인다. 특히, 사용자가 하나의 기기가 아닌 다양한 기기로 미디어를 소비하는 환경에서 보다 높은 효과를 볼 수 있을 것으로 보인다.

하지만, 현재 Media Orchestration 표준에서 공간적 Orchestration 에 관한 논의가 계속되고 있고, 통일된 표준이 아직 발행되지 않았으므로, 표준화의 동향을 파악하는 동시에, 지속적인 연구개발이 필요할 것으로 보인다.

* 이 논문은 2017 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. (No.2015-0-00231, 퍼즐형 Ultra-wide viewing 공간 미디어 생성 및 소비 기술 개발)

6. 참고 문헌

- [1]ISO/IEC 23001-13, Information Technology — MPEG Systems Technologies — Part 13: Media Orchestration, Committee Draft, 2016
- [2]ISO/IEC 23001-13, Information Technology — MPEG Systems Technologies — Part 13: Media Orchestration, Draft International Standard, 2017
- [3]DVB Companion Screens and Streams; Part 2: Content Identification and Media Synchronisation, ETSI, July 2014.
- [4]ISO/IEC 14496-12:2008, “Information technology Coding of audiovisual objects Part 12: ISO Base Media File Format”
- [5]ISO/IEC 13818-1, Information technology Generic coding of moving pictures and associated audio information: Systems, Third Edition, 2007
- [6] ISO/IEC 23005-2, “Information Technology – Media Context and Control – Part 2: Control Information”, 4th edition
- [7] ISO/IEC 23008-1, Information Technology – High efficiency coding and media delivery in heterogeneous environments – Part 1: MPEG Media Transport(MMT), 2014
- [8]ISO/IEC 23009-1:2014 (Second edition), Information technology Dynamic adaptive streaming over HTTP (DASH) Part 1: Media presentation description and segment formats, 2014
- [9] MPEG-B SYSTEMS, "MPEG systems technologies — Part 10: Carriage of timed metadata metrics of media in ISO base media file format / Amd 1 Carriage of spatial information", ISO/IEC 23001-10 Amd1, ISO/IEC JTC1/SC29/WG11/W16191, MPEG 115, Geneva, June 2016
- [10] W3C XML WG, <https://www.w3.org/standards/xml/>
- [11] <http://www.tanimoto.nuee.nagoya-u.ac.jp/>