Integrating Interactive TV Services and the Web through Semantics

Vassileios Tsetsos, University of Athens, Greece

Antonis Papadimitriou, University of Athens, Greece

Christos Anagnostopoulos, University of Athens, Greece

Stathes Hadjiefthymiades, University of Athens, Greece

ABSTRACT

Interactive TV has started to penetrate broadcasting markets, providing a new user experience through novel services to subscribers and new revenue opportunities for companies. Personalization and intelligent behavior, such as proactive content delivery are considered key features for the services of the future TV. However, most of the work in this area is limited to personalization of electronic program guides and advanced program recommendation. In this article, the authors adopt a more horizontal approach and describe the application of concepts, practices and modern Web trends to the TV domain in the context of the POLYSEMA platform. A key characteristic of this approach is the formal modeling of multimedia and user semantics that enables novel TV services. Specifically, Semantic Web methodologies are employed (e.g., ontologies and rules) while compatibility with the MPEG-7 standard is also pursued. The paper describes the overall architecture of the platform, provides implementation details and investigates business issues.

Keywords: *Business Models, Interactive TV, MPEG-7, Personalization, Semantic Web*

INTRODUCTION

During the last two decades, we witnessed many radical changes in the computing paradigms and environments. The recent advances in many areas of Information Technology (IT) seem to converge to a new global computing infrastructure that will be based on an enhanced World Wide Web (WWW). This enhanced WWW will be the corner-stone for human communication, interoperable computing systems, intelligent ultra-wide scale applications and, eventually, improved used experience. Among the aforementioned advances and trends we can distinguish: a) the technologies and social networking principles of Web 2.0, b) the huge online collections of multimedia content, c) the Service Oriented Architectures (SOA), d) the networked multimedia applications (e.g., IPTV, interactive TV) and, finally, e) the explicit rep-

DOI: 10.4018/jswis.2010010101

resentation of content and application semantics through structured metadata. This semantically enhanced application framework is, generally, referred to as Web 3.0. These new technologies can alter the current human-computer interaction practices and facilitate the consolidation of diverse IT solutions into really innovative products and services.

To better illustrate the type of services we refer to, let us consider a sample service that continuously displays information relevant to a currently playing documentary in a small window in a corner of the TV screen. We assume that each scene of the film is annotated with special metadata describing the entities presented (e.g., place, objects, time, persons, activities). These are correlated with the profile of the user as set prior to using the service so that the system decides what information to display. The actual information may be retrieved by different sources in the Web, such as Wikipedia, Flickr, IMDb, etc. Hence, we eventually have a kind of a "Web mashup" in the residential environment created through representation and reasoning of semantics. Another example service can be an advanced parental control service, where the rules for characterizing the content are based on its metadata. Such service would enable the characterization of specific scenes instead of the entire film. We will describe in more detail such service in a following section.

However, despite the maturity of most of the aforementioned technologies, their integration into platforms able to deliver services like those just described has not made much progress. One reason for this is that most researchers focused on specific technology integration tasks. For example, some researchers addressed the integration of TV and Web content while others focused on personalizing specific TV services, usually the EPG (Electronic Program Guide). On the other hand, most of the works ignore the semantics of the content or users. The only metadata they exploit are simple tags or keywords. One reason for this is that rich and formal semantic metadata are hardly available. There are no easy-to-use tools and the publicly available metadata collections do not usually refer to audiovisual content. Unfortunately, this lack of semantics data results in greater effort for developing sophisticated algorithms. Our thesis is that by exploiting the semantics metadata of TV content, we can rely on declarative ways of personalizing services that simplify a lot the development of such systems.

In this article, we describe the application of the Web 3.0 paradigm to a residential infotainment environment that is developed over interactive digital TV (iTV). The discussed approach, implemented in the context of the POLYSEMA project (POLYSEMA, 2009), employs Web technologies and services, multimedia semantics and an extensible residential infrastructure that can support a flexible service lifecycle. Specifically, the main contributions of our solution are:

1. Semantic annotation of audiovisual content through Semantic Web ontologies (e.g., MPEG-7). This results in formal metadata descriptions and advanced support for multimedia content indexing and retrieval. Based on an existing ontology we adapted it and we built a tool for annotating video files (Valkanas et al., 2007).
2. Declarative representation of user profiles and preferences with (Semantic Web-compliant) rules.
3. A testbed for standardized industry technologies (e.g., DVB-T, DVB-MHP, OSGi) and research techniques and tools (context-aware Web content retrieval, rule-based personalization).
4. An open platform that exploits the Web 3.0 framework in the home infotainment domain. The design principles of this platform facilitate the introduction of residential SOA architectures and the introduction of more flexible business models.

In this article, the main architectural orientations of this platform are discussed, as well as core implementation technologies. The requirements that drove the design and development of the platform were collected by a Digital TV provider, member of the POLYSEMA consortium. The findings of the requirements analysis indicated a strong demand for a flexible iTV solution that would:

- Be capable of delivering well known iTV services (e.g., advanced Electronic Program Guide - EPG, video on demand, TV of yesterday)
- Be based on standardized broadcasting technologies such as DVB and be compatible with existing broadcasting platforms and residential gateways (thus reducing the time-to-market)
- Be able to integrate with other data communication channels such as the Web
- Support flexible business models through its support for novel infotainment services.

In the rest of this article we describe how the identified requirements were materialized in the developed system. The flexibility and openness of the solution is demonstrated through the description of a sample parental control service and new business models that can be supported.

RELATED WORK ON SEMANTICS-ENABLED INTERACTIVE TV AND TV-WEB INTEGRATION

Since the appearance of digital TV, both academia and industrial stakeholders have focused on the integration of the Web with the world of television broadcasting. A strong indication of this is the fact that the DVB-MHP (Digital Video Broadcast – Multimedia Home Platform) specification includes a profile which supports Web browsing and e-mail client functionality. Web browsing is realized via the introduction of the DVB-HTML markup language. Unfortunately, DVB-HTML has not been widely adopted yet by STB manufacturers, because it is a quite complex specification, thus leading to higher device costs and interoperability/transcoding issues. An alternative to DVB-HTML was recently proposed in (Ferretti et al., 2007). The authors developed a method for transcoding Web pages to MHP-compliant visual components that can be displayed in TV screens by DVB-enabled set-top-boxes. In this way, they try to integrate the Web with a TV environment. Moreover, the authors proposed an integrated solution for intelligent Web content browsing/retrieval through mobile agents that search the Web for user-relevant content. Towards the same direction, the author of (Tanaka, 2007) has proposed an approach for augmenting TV content with Web content, by exploiting intelligent information retrieval techniques applied to closed captions.

However, in our view, convergence of WWW with iTV does not only refer to browsing the Web from a television screen using a remote control, which was the main topic of the aforementioned works. It also signifies that a broad scope of diverse applications can migrate from the Web to the TV broadcasting world. Two major research efforts towards this direction are the AVATAR (Blanco Fernandez et al., 2006) and the MediaNet (MediaNET, 2009) projects. AVATAR is a project which shares many techniques and technologies with POLYSEMA, such as annotation of DVB video streams, ontology-based modeling, multimedia metadata, user profiling and personalization through semantic reasoning. However, the project's main objective is to create a personalized digital TV program recommender and not transferring Web content in the residential infotainment domain in a personalized way. The system accomplishes the task by taking into account the user's profile and TV-Anytime meta-information of each TV programme. The MediaNET project is divided in five subprojects, each of which covers a significant area of the multimedia content lifecycle: creation, service providers, network operators, etc. As far as interactive services are concerned, it delivers the AmigoTV platform and a PVR. Moreover, MediaNET offers integrated iTV/Web services such as e-voting and e-shopping. AmigoTV is a social TV platform that delivers services over triple-play networks. Although the service makes significant progress towards importing popular social networking applications into the iTV domain (e.g., messaging, voice communications, games), it does not investigate the benefits from using metadata during the various phases of the multimedia content and service lifecycles. Hence, all actions are user-initiated and the system does not provide means for intelligent automation of service executions or (Web) content retrieval.

Another project that exploits semantic metadata in order to provide personalized access to TV content is the SenSee framework (Aroyo et al., 2007). This framework harvests TV metadata from various sources (e.g., BBC, IMDb) and provides a personalized TV guide. Several ontologies are used for integrating the harvested metadata and to model the current user context. All this information is used for inferring content of interest to the user.

The use of multimedia semantics in different phases of the multimedia content lifecycle is of course not a new approach. Several researchers have performed significant contributions through designing and demonstrating the use of multimedia ontologies (Arndt et al., 2007; Tsinaraki et al., 2004; W3C Multimedia Semantics Incubator Group, 2009). However, in most of the related works, semantics are used mainly for content classification and for (meta-) data integration purposes and not for actual reasoning. Moreover, most efforts deal with still images and less dynamic types of media. In this work we also exploit an MPEG-7 ontology to describe video files. Moreover, we model all other information involved in the service provisioning though ontologies and eventually apply rules on the resulting knowledge base.

Finally, in terms of middleware infrastructure a very similar approach that integrates the MHP and OSGi platforms is presented in (Vilas et al., 2006). The architecture presented in (Vilas et al., 2006) relies on a hybrid MHP/OSGi component called XBundLET. This component serves as a proxy between the two platforms, enabling bidirectional service discovery and invocation. In our case, since we did not require so tight integration between the two platforms, we adopted a more lightweight approach relying on communication over HTTP, as described in section "The POLYSEMA Platform". Apart from this technological similarity the two approaches do not have much in common, since our middleware is enhanced by the Semantic Component and constitutes an integrated iTV service platform.

From the survey just presented, it becomes apparent that all these related works do not try to develop a general-purpose framework that can provide intelligent and personalized services to the users, but rather focus on specific parts of such system (e.g., Web-TV content integration, personalized EPGs). In the following sections we present the POLYSEMA platform that adopts a more horizontal approach.

THE POLYSEMA PLATFORM

The POLYSEMA platform is one of the first prototypes to deliver next generation services to TV viewers. It constitutes a proof-of-concept system for developing and assessing personalized iTV services. It covers the whole service lifecycle by providing a residential gateway and some basic server-side functionality. Specifically, the platform provides: content preparation (segmentation), metadata creation and delivery, Web content retrieval and presentation, personalization based on rule-based inference. One challenge addressed by such platform was the introduction of rich metadata (based on Semantic Web practices) in the entire workflow of service provisioning. All metadata are aligned to domain models that represent shared vocabularies for the development of services. This way many interoperability problems are handled. The main difficulty in this task was to build the models (i.e., ontologies) in a suitable form and populate them with the actual data. For example, the MPEG-7 ontology should enable standard reasoning tasks (such as concept classification) but not be too strict and demanding. Towards this end several ontology elements were devised accordingly or totally removed.

Architecture Overview

Components

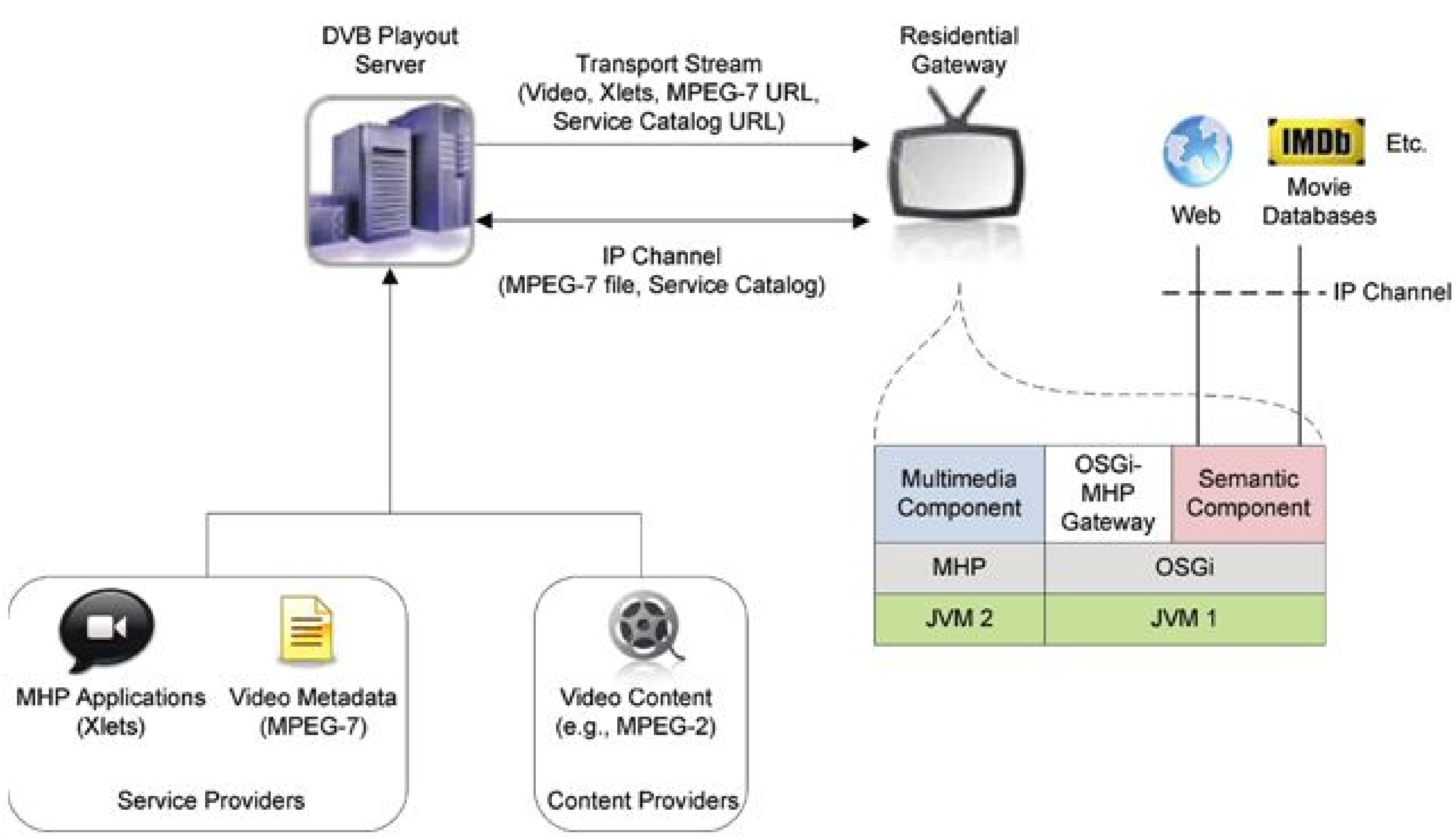
Figure 1 presents a bird's eye view of the end-to-end architecture. The architectural components can be divided to those residing on the server side, responsible for delivering content and applications to the iTV subscribers and those deployed on the client side. The main server-side component is the Video Playout Server, which performs streaming of the audiovisual content. Moreover, several other (Web) servers may be available for providing other types of content. The client is implemented as a residential gateway, which handles service execution, user interaction, service personalization, and content presentation. The core components inside this gateway are the Multimedia Component and the Semantic Component. The first is responsible for decoding/displaying the broadcast video stream, executing iTV applications, receiving content-relevant notifications from the broadcaster and handling the user interaction issues of the platform. The Semantic Component is responsible for deciding and executing additional personalized actions that should be taken during the play-out period of the audiovisual content. Such actions are decided (in real-time) based on the service, user and content semantics. Typical actions are service invocations (e.g., video recording, parental control) and retrieval of metadata and related content from the Web. Evidently, coordination of the Multimedia and Semantic Components is necessary for delivering the POLYSEMA Services. Such coordination is achieved through a gateway between the two components/platforms.

Content

Until now we referred to content in a rather abstract way. In the context of POLYSEMA, content can be categorized to three types:

a) *Audiovisual content*. This is the traditional TV content transmitted by the broadcaster (e.g., movies, advertisements). Such content is not further processed by the platform.
b) *Metadata*. This is meta-information about the TV content, which describes its structure and captures other semantic aspects (e.g., movie cast, scene descriptions). This metadata is used for personalizing the various iTV services. It can be provided by the broadcaster or other providers.
c) *Web content*. This type of content is retrieved when required by the services. It may be related to the audiovisual content

Figure 1. Overall POLYSEMA architecture



(e.g., movie reviews) or to the user preferences (e.g., news feeds). It is retrieved from Web servers, typically not owned by the broadcaster.

Services

Before delving into the details of the platform we should clarify that a POLYSEMA Service is a value added service (VAS) exploiting TV content semantics in order to provide additional relevant content (typically retrieved from the Web in a context-aware manner) and intelligent behavior. A POLYSEMA Service consists of three parts: a) the Multimedia part, b) the Semantic part (it may be the business logic of a service or a wrapper for an external system, such as Personal Video Recorder - PVR) and c) a Service Description with details about service functionality and operational semantics. The "Multimedia part" is used by the Multimedia Component for presentation and human-computer interaction purposes (e.g., for displaying text messages and multimedia content, switching channels, or adjusting sound volume). The "Semantic part" contains the implementation of the core service logic. Finally, the "Service Description", is used for service discovery and configuration purposes. In addition, it is used by the Semantic Component so that the correct method calls to the Semantic or Multimedia service parts can be made. The set of all service descriptions that are available *for a specific subscriber* is referred to as Service Catalog in Figure 1.

Technologies

Regarding the content, the technology used for video streaming is DVB-T, a widely adopted standard for digital TV. For video broadcasting the IRT DVB Playout Server was used (IRT, 2009). The content metadata is represented through MPEG-7–based video annotation. The residential gateway is an open service platform, namely Open Service Gateway Initiative (OSGi, 2009). OSGi enables, among other things, advanced networking and provides service lifecycle management capabilities. Specifically, OSGi is used for implementing the execution environment for the business logic of the POLYSEMA Services (i.e., Semantic Component). Moreover, we achieve seamless service upgrade, through the OSGi framework. We exploit the OSGi mechanisms in order to install new services and allow for future integration of the system with home networking and automation applications.

The Multimedia Component of the residential gateway, on the other hand, relies on the Multimedia Home Platform (MHP) middleware and Execution Environment (MHP, 2009). MHP is a Java-based execution platform (i.e., it runs on top of a Java Virtual Machine - JVM) for MHP applications, called Xlets. Xlets are to the digital TV what applets are to the Web. Such execution environment fully controls the Xlet lifecycle, including initialization, activation, and destruction. MHP contains several code libraries for the development of Xlets. Many of these libraries rely on other specifications such as Home Video Audio Interoperability (HAVi, 2009) and Java TV (JavaTV, 2009).

OSGi and MHP environments have different properties as they adopt different design principles (Vilas et al., 2006). In order to retain full compatibility with industry standards such as MHP and DVB, and still exploit the service openness and flexibility of OSGi, we decided to build a proxy bundle in the OSGi platform that conceals the nature of MHP applications from the rest of the system (MHP-OSGi Gateway). The MHP-based Multimedia Component communicates with this proxy bundle via the IP return channel of the receiver. Similarly, the OSGi-based components of the Semantic Component interact with the Multimedia Component by accessing the respective MHP-delegate OSGi service.

Finally, the POLYSEMA services are implemented as a mixture of technologies. The Multimedia part is implemented as an MHP application (i.e., Xlet). The Semantic part is implemented as an OSGi bundle (i.e., a packaged Java application). The Service Description is represented through XML.

Multimedia Component

The Multimedia Component consists of MHP applications that can be downloaded and executed by any DVB-MHP receiver. The only requirement is that the MHP middleware running at the Set-Top-Box (STB) supports Return Channel (RC) communication, which is based on TCP/IP. This is necessary for the interaction between the Multimedia Component and the Semantic Component (see OSGi-MHP Gateway in Figure 1).

The operation of the Multimedia Component relies on a basic MHP application running at the receiver's end. This Xlet (called "POLYSEMA Xlet") includes the Multimedia parts of all the services supported by the broadcaster. Another task assigned to this Xlet is to read the broadcast file system (Digital Storage Media Command and Control or DSM-CC or 'Carousel', in MHP terminology) and retrieve the URL that will provide the MPEG-7 metadata. Such metadata describe the audiovisual content and the Service Catalog for the subscribed POLYSEMA Services. Both these URLs are, subsequently, fed to the Semantic Component which is responsible for fetching, locally storing and processing the respective files.

Finally, the POLYSEMA Xlet subscribes and listens to incoming stream events, inserted into the DVB Transport Stream (TS) by the broadcaster. Typically, these events indicate the beginning of a new video scene (a.k.a., segment) and are automatically generated at the playout server from the corresponding MPEG-7 temporal decomposition of the video, as provided by the service provider. Upon the reception of a stream event, the POLYSEMA Xlet inquires the Semantic Component for services that should be invoked for the specific video scene. These services are specified in a plan, as described in the following section. Hence, through these events the streaming video can be synchronized with the VAS.

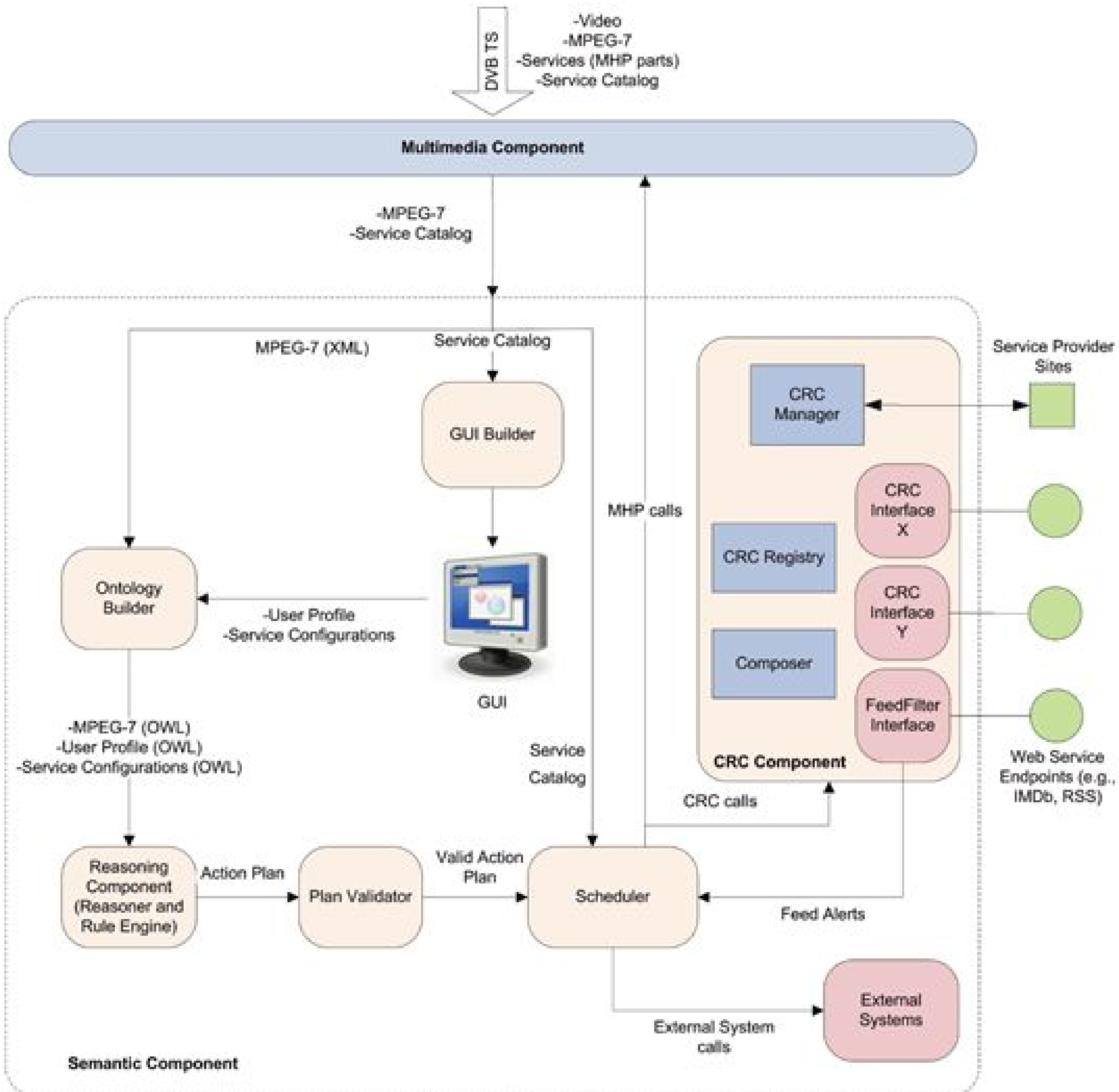
Semantic Component

At the heart of the platform lies the Semantic Component. It is responsible for the coordination of the VAS offered by the platform as well as the inference processes that drive the personalized service provisioning. In a few words, one could describe this component as the knowledge- and rule-based infrastructure that provides personalized 'mashups' of content inside a TV receiver. A diagram with the basic elements of its architecture and their main interactions is shown in Figure 2.

Upon retrieval of the Service Catalog file, a graphical user interface (GUI) for managing the service and user profiles is automatically generated using Java Server Pages. These profiles, once defined by the user, are automatically translated into ontological instances. These profiles are static, in the sense that they are not dynamically adjusted by the system. Once a user defines her profile, it remains unchanged until the next update. The MPEG-7 data are also translated to instances of a MPEG-7 ontology (see also section "Semantics-based Personalization of iTV Services"). We should remind that ontology is an "explicit specification of a conceptualization" according to (Gruber, 1993). In our case it provides a hierarchical representation of concepts along with their relationships that describes the specific multimedia domain. Hence, according to the classifications in (Mizoguchi, 2004), the MPEG-7 ontology is a type of light-weight ontology: it is mainly used for data integration reasons and not for the formal definition of the domain concepts.

The instantiation of this ontology is performed by the Ontology Builder component and is necessary so that all data are represented in a common format, compatible with the inference engine (Jang & Sohn, 2004) of the Reasoning Component. The Ontology Builder consists of a specially designed parser for valid MPEG-7 XML files. It exploits naming conventions between the XML elements of the MPEG-7 standard and the class names in the MPEG-7 ontology in order to populate the latter.

Figure 2. POLYSEMA semantic component



The rules that are applied to these ontological data are specified by the individual third party service providers in the Service Catalog. Once the rules are executed, a first execution plan is in place, which describes the actions that should be scheduled for the current user and the current TV program. Before submitting this action plan to the Scheduler, who is responsible for the plan execution, the Plan Validator is invoked in order to produce a more refined version of the plan and resolve possible conflicts among actions. Such conflicts may be forced by device specifications, the semantics of individual services or the temporal relationships between the scheduled actions. For instance, the Plan Validator does not allow the execution of two Personal Video Recorder (PVR) services with temporal overlap, as the Set-Top-Box (STB) may not support concurrent recording of two different TV programs. Moreover, a Web information retrieval service should not be executed when a parental control service is already running, since the additional information may also be inappropriate for underage viewers. Another validation rule example is the following: if two services should present Web content in the same region of the screen concurrently, one of them is removed from the plan. In general, the rules for performing plan validation are based on heuristics and common sense, and their aim is to ensure "smooth" invocation of the services and presentation of their results. These rules are applied once the actions for each video segment have been decided by the Reasoning Component. Hence, conflict resolution is not performed in the typical way that rule engines do, but on the inferred action set. Currently, we have implemented validation rules like the aforementioned, but many more should be probably needed in an actual deployment. Moreover, more sophisticated and formal ways of handling conflicting actions should be investigated in the future.

Finally, the Scheduler component invokes the methods that correspond to the inferred actions. For these mappings between actions and service methods, it consults with the Service Catalog. Such methods may be MHP calls, calls to external Web servers or calls to external systems (PVR, networked home devices, etc.).

Interfaces with Web and Other External Content/Service Providers

One of the objectives of the POLYSEMA platform is to constitute a residential implementation of the Web 2.0 that will introduce rich internet applications to the home environment. Through its infrastructure, open and proprietary APIs can be accessed in order to bridge the gap between Web resources and home infotainment. The data fetched from the WWW can be either in textual (movie information, encyclopedic knowledge, sport statistics, translation services etc.), visual (e.g., images) or any other format that can be "consumed" by the STB middleware.

The integration of a home infotainment TV environment with the modern Web technologies and trends is mainly implemented through the Content Retrieval & Composition Component (CRC). Once the Reasoning Component has preprocessed the semantic video description and has "inferred" that certain Web information should be fetched and displayed to the viewer, the Scheduler instructs the CRC component to start downloading the relevant data.

CRC consists of several interfaces (implemented as OSGi bundles) that provide access to external Web resources. For instance, in the prototype system we have developed bundles which access content from the Internet Movie Database (IMDb) and Wikipedia. The respective classes follow a common abstract design, which involves management of created objects, caching of retrieved data till the end of the corresponding TV program and basic formatting of the requested Web information. Through this modular design data heterogeneity is handled too, since content retrieval is performed by these specialized bundles. Each bundle handles content retrieval from specific sources and, thus, "knows" how to parse it and what to present to the user.

Surely, the various CRC interfaces may differ with regard to the actual information they can retrieve. Hence, each installed bundle updates a registry, called CRC Registry, which is used by the Scheduler to map user requested information to CRC interface method calls. Typical information stored in this registry includes the specific methods and parameters that correspond to types of available information as well as the format of this information (image, text, etc.). For example, if we can retrieve the director of a given movie from IMDb, there is a corresponding option in the relevant Service Description, which can be selected by the user through the GUI and further mapped to a CRC method call. This mapping is performed in the CRC Registry. We should note that a CRC interface is actually the Semantic part of a POLYSEMA Service (for this example, the IMDb Web content retrieval service). Hence, a service creator/provider is the sole responsible for deciding what info each service can retrieve. Of course, most CRC methods take some input parameters. For example, a method retrieving an actor's biography takes its name as input. The values for these parameters are extracted from the MPEG-7 metadata. Such metadata is expressed through ontologies, as it will be described in section "Semantics-based Personalization of iTV Services". Hence, a MPEG-7 document that supports the IMDb service should annotate the actors based on a special purpose IMDb ontology containing all valid actor names (a sample ontology can be found in (POLYSEMA, 2009)).

The additional content retrieved by the CRC interfaces is returned to the Scheduler, which then stores it in text or image files at a local lightweight HTTP Server. This server is running within the Semantic Component and is the core component of the OSGi-MHP Gateway. Subsequently, the Multimedia Component (after a Scheduler's call) accesses the HTTP Server in order to retrieve the Web content and display it to the user (this workflow is shown in Figure 3). A special component in the CRC design is the FeedFilter interface. A user can register feeds of interest and define the way she wants to receive notifications about feed updates (currently only RSS feeds are supported).

Finally, the CRC component has its own updating mechanism, which is responsible for communicating with the service provider sites and installing new CRC interfaces or upgrading existing ones. Such updating is totally handled by the OSGi framework.

POLYSEMA SERVICES

Several services have been implemented in our prototype:

- A parental control service
- A service that retrieves and displays Web content from IMDb. Such content can be actor-specific, and is, thus, displayed in every segment that contains some new actor, or movie-specific (e.g., rating, reviews, director details) and is displayed during the beginning/end of a movie
- A service that retrieves and displays Web content from Wikipedia for segments that contain specific objects
- A PVR service that records segments that contain specific activities, objects, places or actors (these are specified by the user and adhere to the terminology of the LSCOM ontology (Naphade, 2006)
- A notification service bound to user-specified RSS feeds.

As an example, the parental control service performs some action (e.g., mutes sound, changes channel, toggles off video, displays alert) for video segments that are not appropriate for the current user, judging from her age. This service differs from typical parental control services, in that it performs a more fine-grained control over content display, i.e., on a per video segment basis and not on the video as a whole. Figure 4a shows the code that is executed in the MHP environment upon service invocation (Multimedia part); while Figure 4b shows the service description that defines the service invocation rule templates and other service metadata[1]. Rule templates are expressed in the rule language described in (Jang & Sohn, 2004). Each rule template defines when a service should be invoked and with which input parameters. The actual values of these parameters can be obtained from the MPEG-7 file and the current status of the system. In our example, the only parameter is *currentUser* the value of which is set upon user logon.

Knowledge Engineering Details

The exploitation of user, service and content semantics is the cornerstone of our system. Specifically, all the video content is assumed to be annotated according to the MPEG-7 Multimedia Description Scheme (MDS). However, we limited the MDS expressiveness only to Description Schemes (DS) and Descriptors

Figure 3. UML sequence diagram describing the content retrieval process

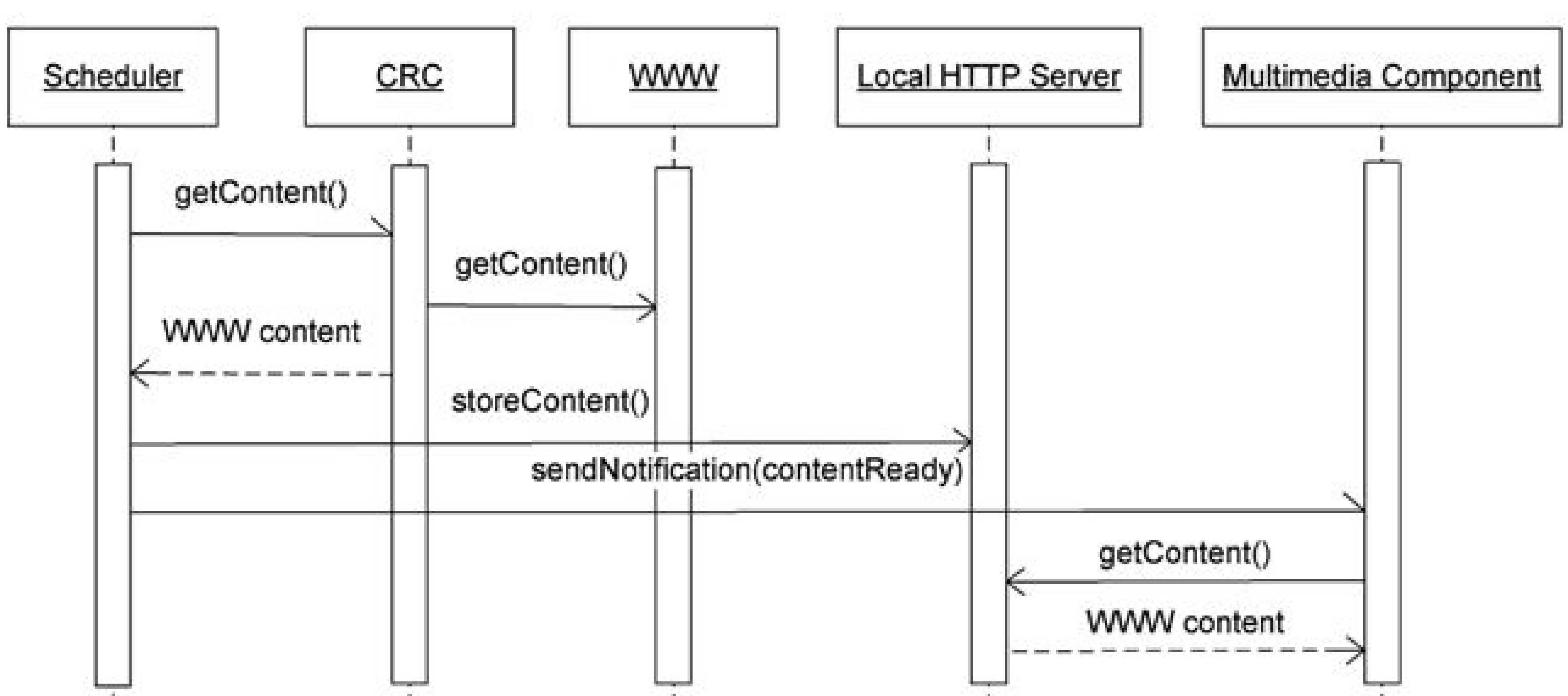


Figure 4. An example parental control service a) Multimedia part, b) Service description

```
public class ParentalControl{
  //we overide the paint method of the HContainer..
  public void paint(Graphics g) {
    DVBGraphics dvbG = (DVBGraphics) g;
    DVBAlphaComposite compositingRule;
    // When alpha = 1, the display gets black
    compositingRule = DVBAlphaComposite.getInstance(DVBAlphaComposite.DST_OVER, (float) this.alpha);
    dvbG.setDVBComposite(compositingRule);
    super.paint(dvbG);
}
```

(a)

```
<service id="2" name="parentalControlService" version="1">
  <GUIName>Parental Control</GUIName>
  <description>Takes actions upon detection of inappropriate content</description>
  <API_method>NULL</API_method>
  <ruleTemplate>if TVUO:Under7(?u) and [?u = TVUO:*currentUser*] and mds:Why(?sat, MPAA:directed_to_older_children) and
   mds:TextAnnotation(?vst, ?tat) and mds:VideoSegmentType(?vst) and mds:StructuredAnnotation(?tat, ?sat) then iTVServices:action(?vst,
   iTVServices:parentalControlService)</ruleTemplate>
  <params>
    <param id="1">
      <name>action</name>
      <GUIName>Action</GUIName>
      <description>Determines the action for parental control</description>
      <type primitive="false">exclusive-list</type>
      <exclusive-list>
        <item>
          <value>toggle.video</value>
          <GUIName>Toggle video</GUIName>
        </item>
        <item>
          <value>alert</value>
          <GUIName>Alert Viewer</GUIName>
        </item>
      </exclusive-list>
    </param>
  </params>
</service>
```

(b)

relevant to "high-level semantics" of video files (e.g., those describing persons, activities, places, time, objects). A significant differentiation with respect to the MPEG-7 standard is that we did not adopt its term definition and term reference mechanisms (e.g., TermUseType and ControlledTermUseType elements, classification schemes) and used only domain ontologies for describing video semantics, instead. Hence, the MPEG-7 ontology we used, although retains most of the vocabulary of the MPEG-7, it allows extension of the annotation terms through third-party ontology resources.

Moreover, we did not adopt the MPEG-7 Semantic DS elements (e.g., Semantic, SemanticBase, SemanticType, Time, Event), which allows for very expressive high level semantic descriptions, for the following reasons:

- The effort required by the annotator for such descriptions is too high. An indication of this is the fact that none of the available (commercial or not) video annotation tools does support such descriptions. One exception to this is the COMM (Core Ontology on Multimedia) API (Arndt et al., 2007).
- There is no sound and accepted methodology that defines how such annotation should be performed or how the applications should exploit such information. The lack of such methodology causes ambiguity during metadata interpretation. This problem is also reported in (Rahman et al., 2006), where an ontology for the semantics of MPEG-7 Semantic DS is proposed as a potential solution.

Since the intelligence of the system is provided by a rule-based system, it was necessary that all data related to the user, TV program, and services are represented in a common knowledge representation formalism. This formalism was decided to be the Web Ontology Language (OWL) due to its popularity and its adoption by the World Wide Web Consortium

(W3C). For the ontology-based representation of the metadata we elaborated the model of Tsinaraki et al. (2004), implemented in OWL, too. In (W3C Multimedia Semantics Incubator Group, 2009), some other candidate models are also listed but they are either too abstract or too detailed for our needs. We performed several changes to the original MPEG-7 ontology in order to render it more flexible (e.g., omitted certain concepts and properties). Flexibility mainly refers to the ability to: a) automatically populate such ontology from original MPEG-7 XML files, b) perform typical reasoning tasks such as classification, and c) allow for linking to external domain ontologies for describing video segments. In general, we kept only MPEG-7 elements relevant to the provision of iTV services. The ontology resulted in consisting of 53 classes, 53 datatype properties and 63 object properties. The corresponding OWL file can be found in the POLYSEMA site (POLYSEMA, 2009). Indicative modifications are:

1. The classes TermUseType and ControlledTermUseType were removed, as already mentioned at the beginning of the section, since the role of such vocabularies is substituted by domain ontologies. Moreover, the Term class (or some subclass) was removed from the range of all properties.
2. The elements for relative time reference were removed (e.g., MediaRelTimePointType, RelIncrTimePointType) as well as the elements for "incremental" timing (e.g., MediaIncrDuration, IncrDuration).
3. The elememts relevant to Classification Schemes were removed (e.g., ClassificationSchemeBaseType, ClassificationSchemeDescriptionType).
4. All elements relevant to the Semantic DS were removed, as already explained.
5. All elements describing graphs were removed (e.g., GraphType, Node).
6. Several object properties that in practice will assume primitive data types were converted to datatype properties (e.g., FreeTextAnnotation, Language, GivenName).

We should note that the performed modifications can surely affect the interoperability between our ontology and the original one. However, interoperability was not one of our requirements. The original ontology was used just as a starting point for the design of a more "pragmatic" ontology that could be directly used in annotation tools and in reasoning tasks.

All actual data (instances) in this MPEG-7 ontology were expressed through domain ontologies. Some of these ontologies derive from the MPEG-7 and TV-Anytime classification schemes (e.g., those related to parental rating and video categorization). Others are custom ontologies that restrict the vocabulary that can be used by certain services. For example, as already mentioned, we have designed an IMDb ontology that contains valid actor names. These ontologies are used both during video annotation (Valkanas et al., 2007) and service execution. Since these ontologies are used for rule execution, some singleton instances were also created besides the class hierarchies. As a convention, each singleton instance has the name of the class it belongs (and a prefix). For instance, in MPEG7RoleCS ontology that described movie roles we have the following instance assertion: <Art_Director rdf:ID="art_Director"/>.

Similarly to the translation of MPEG-7 XML data to OWL data, each service configuration is transformed to a proprietary service ontology (iTVServices) and each user profile to a TV user ontology (TVUO). The TVUO class hierarchy is shown in Figure 5. As one can see, this model can be instantiated with some demographic information about the user, and her preferences regarding the content delivered and the services provided (i.e., which services should be executed and with what parameters). All the ontologies mentioned in this section are available in (POLYSEMA, 2009).

The adoption and engineering of formal ontologies for structuring the metadata (i.e., MPEG-7, TVUO) and their instantiation through domain ontologies (and not just literal values), makes the platform a pure knowledge-based system, capable of inferencing and reasoning over explicit and implicit knowledge.

In Figure 6, some sample data are presented so that the reader can understand how the system performs reasoning and computation of the action plan. In Figure 6a, a rule (of a parental control service) is depicted. In natural language, it states that "if the current user is under 7 years old and the current video segment is directed to older children, according to the MPAA (Motion Picture Association of America) classification scheme, then the service should be triggered for that segment" (we assume that parental ratings are expressed in the "Why" element of MPEG-7). This is a generic rule applied to the actual metadata of the video and the user (i.e., the variables, starting with "?", will be evaluated with some instances of the video MPEG-7 ontology and the ontological instances of the other ontologies). In case the user is described in a user ontology as presented in Figure 6b and 6c (represented in Description Logics notation), and the video metadata are that of Figure 6d, then the service will be added to the plan for all video segments that satisfy the rule conditions. The specific action taken will depend on the value of the "action" service parameter (see Figure 4). The "mds" namespace refers to the MPEG-7 ontology, the "TVUO" and the "iTVServices" to the respective profiles (i.e., the description of Figure 4b is transformed to iTVServices instances) and MPAA to a domain ontology that defines parental rating classes.

New Business Models for Interactive TV

The openness and extensibility of the POLYSEMA implementation technologies along with the innovative exploitation of semantics allows for new business models in the iTV domain. An indicative business model that is supported by the platform is illustrated in Figure 7. In this figure, one can see the revenue (Ri, i=1,2,…) and information (Ij, j=1,2,…) flows between the core players and roles of this model. The innovation of this model is the introduction of the Metadata Provider role and the Service Aggregator role (the last will be typically assumed by the Broadcaster). Another new role is that

Figure 5. TVUO class hierarchy

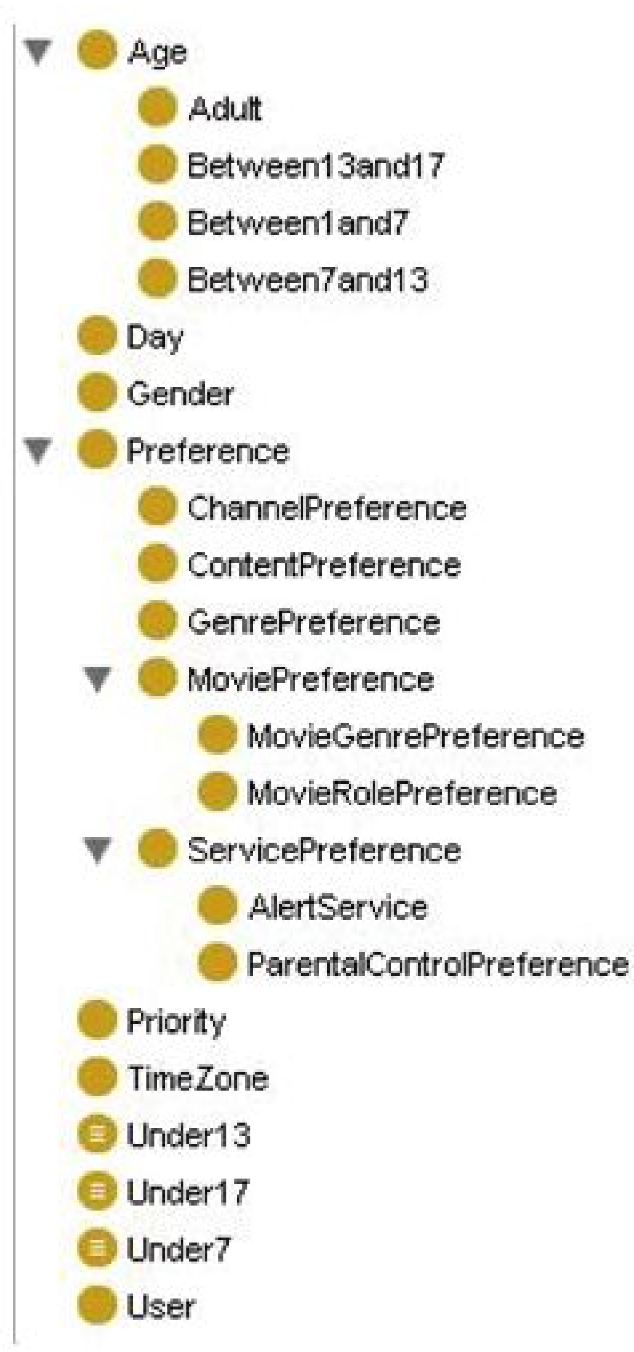


Figure 6. Example data for rule-based service personalization

a) service rule

```
rule parentalControl is if
TVUO:CurrentUser(?u) and TVUO:Under7(?u) and mds:Why(?sat, MPAA:directed_to_older_children) and
mds:TextAnnotation(?vst, ?tat) and mds:VideoSegmentType(?vst) and mds:StructuredAnnotation(?tat, ?sat)
then
iTVServices:action(?vst, iTVServices:parentalControlService)
```

b) user profile

```
<TVUO:User rdf:ID="bill">
  <TVUO:hasUsername rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Bill</TVUO:hasUsername>
  <TVUO:hasGender rdf:resource="http://polysema.di.uoa.gr/ont/TVUO.owl#male"/>
  <TVUO:hasAge>
    <TVUO:Between1and7 rdf:ID="ageOfBill"/>
  </TVUO:hasAge>
</TVUO:User>
```

c) a user profile definition

$$\text{Under7} \equiv \text{User} \sqcap \exists \text{hasAge.Between1and7}$$

d) extract of the MPEG-7 instances

```
<mds:VideoType rdf:ID="MultimediaContent3">
   <mds:Video>
   ......
      <mds:VideoSegmentType rdf:ID="segment7">
        <mds:MediaTime>
        ......
        </mds:MediaTime>
        <mds:TextAnnotation>
          <mds:TextAnnotationType rdf:ID="TextAnnotation20">
            <mds:StructuredAnnotation>
              <mds:StructuredAnnotationType rdf:ID="StructuredAnnotation21">
                <mds:Why rdf:resource="&MPAA;#directed_to_older_children"/>
              </mds:StructuredAnnotationType>
            </mds:StructuredAnnotation>
          </mds:TextAnnotationType>
        </mds:TextAnnotation>
      </mds:VideoSegmentType>
            ......
```

of a Service & Metadata Provider. Although, separate Service Providers and Metadata Providers may exist, a new role that combines both of their responsibilities is also supported by this model. This new role may even be necessary so that the service logic (rules) can be paired with meaningful metadata. For example, if one Service Provider provides localized Web information retrieval to augment documentaries, then it should make sure that such geospatial information is also available in the documentary metadata, or else, the rules of its service may never "fire".

The need for the Service Aggregator is attributed to the fact that each user will be typically registered for several POLYSEMA services. Hence, the Service Catalog has to be dynamically composed. If, for a certain video file, there are numerous metadata from different providers, these should be aggregated too. Such metadata aggregation is a very challenging issue, unless the same annotation tools and conventions are used by all Metadata Providers. Similar issues are also met in other areas, such as metadata harvesting in digital libraries. The aggregation process may encounter difficulties at the syntax level (e.g., different syntax describing the same semantics) or at the semantics level (e.g., inconsistency of descriptions). However, this topic is out of the scope of this work and will not be further discussed.

A more detailed description of the information flows between the various players of the business model is summarized in Table 1. Finally, regarding the revenue flows we should note that various payment models and subscription types can be supported. For example, a user may subscribe for a fixed set of services or for unlimited use of services (flow R1). Alternatively, she can pay for some basic set of services and be additionally charged depending on her service usage. The flow R2 can be also implemented through several compensation methods. The 'pay-per-install' method gives income to the Service Providers depending on the installations of their services by the users. Another, Internet-oriented, method is 'pay-per-impression'. According to this model, each provider receives commission for several hundreds or thousands of views (it is similar to the 'pay-per-view' model).

CHALLENGES AND FUTURE WORK

In this paper, we presented a semantically enriched platform for Interactive TV. Our

Figure 7. Business model roles and information/revenue flows between them

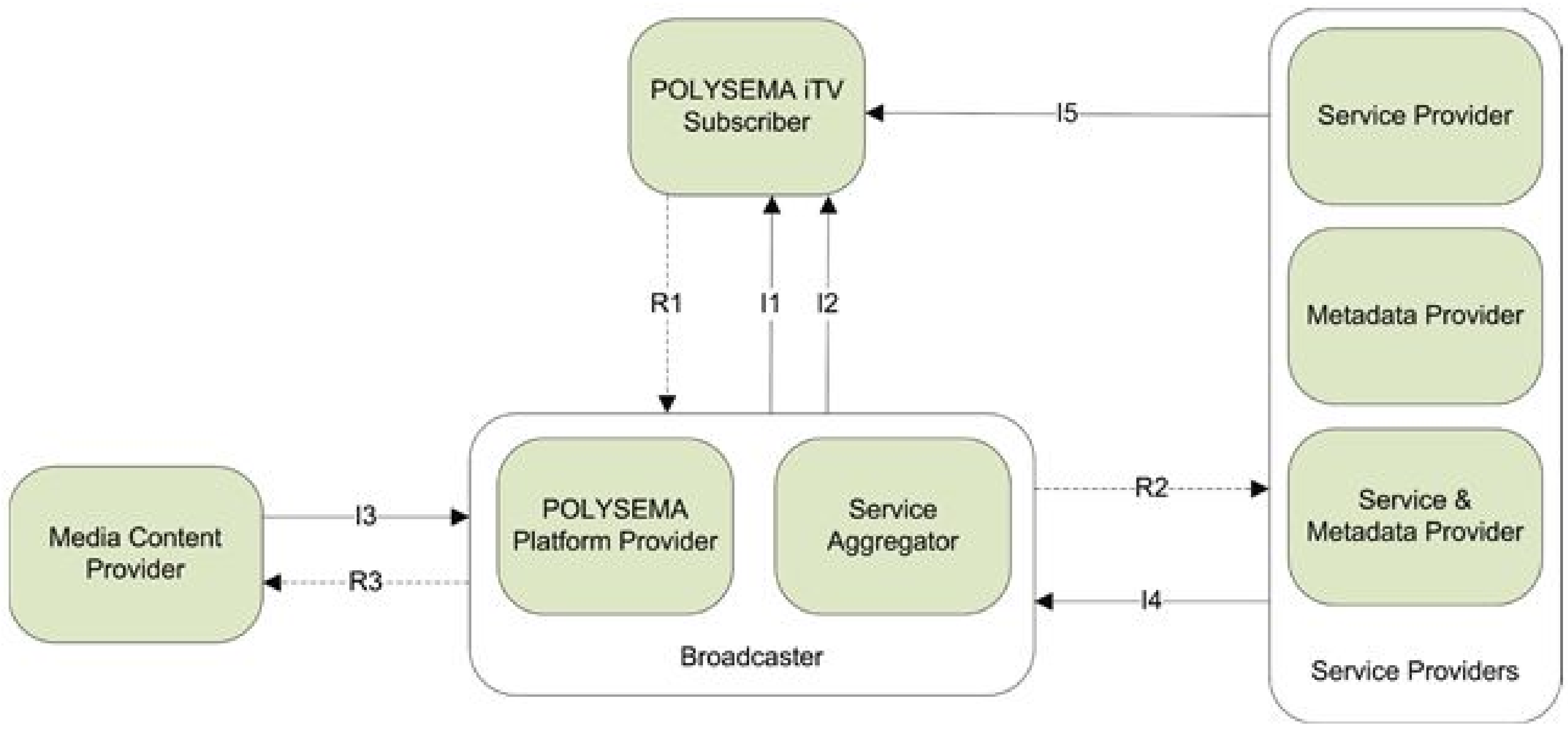


Table 1. Description of information flows

Flow ID	Description
I1	This flow represents the Transport Stream that also contains the DVB Carousel. It consists of a) the video content, b) typical broadcast information, c) the Multimedia service part and d) a link to the Service Aggregator service endpoint
I2	This flow represents the information sent to the subscriber upon login to the platform. It consists of the Service Catalog that corresponds to its subscription, and b) the MPEG-7 metadata file that may also depend on its subscription (in case some services come with their own MPEG-7 annotations)
I3	The video content delivered through the broadcast channel
I4	a) service data and b) MPEG-7 metadata for the broadcast program
I5	a) Semantic service part (e.g., CRC bundles), b) service-specific domain ontologies

platform, called POLYSEMA, is based on open technologies and emerging standards (e.g., MPEG-7, OSGi, OWL). We provided an extensive overview of the POLYSEMA architecture emphasizing on the issues of knowledge engineering, Web-iTV integration and the business models supported through the particular features of the platform.

Our future plans for extensions/improvements include a thorough study of the adopted metadata formalisms. Specifically, in our approach we have used the MPEG-7 standard for capturing multimedia content semantics. However, there is growing skepticism among the multimedia metadata community on the value of MPEG-7. Its high complexity has played critical role for its low adoption and tool support. In fact, many researchers try to develop formalisms that substitute (parts of) the standard. From our experience, both in metadata production (Valkanas et al., 2007) and metadata consumption, MPEG-7 is indeed an inefficient standard. Our conclusions are that sophisticated and user-friendly semi-automatic video annotation tools are necessary for the standard to survive the "metadata war" as well as a further modularization and layering of the specification. Modularization is merely supported by the MPEG-7 structure (Description Schemes), but should be extended and better documented. Layering refers to the definition of various species of the standard in a way analogous to the Web Ontology Language (Lite, DL and Full). Moreover, such layering would facilitate its comprehension by the developers and users.

From our experience we noticed that the MHP is a heavyweight middleware platform for Web-enabled interactive TV. On the other hand, more open and lightweight solutions (possibly implemented through IPTV and Web-based technologies), such as (Opera, 2009), are not widely adopted and tested, yet. Hence, choosing an iTV platform for building value added service platforms like POLYSEMA is an issue that deserves further study.

Finally, what is not addressed to a satisfactory degree in our platform yet, is the support for social networking and advanced recommendation services. These are key elements of the Web 2.0 ecosystem. Therefore, we are currently working on the proof-of-concept implementation of similar services in POLYSEMA.

REFERENCES

W3C Multimedia Semantics Incubator Group. (2009). *Multimedia Vocabularies on the Semantic Web*. Retrieved July 25, 2009 from http://www.w3.org/2005/Incubator/mmsem/XGR-vocabularies/

Anisur Rahman, M., Hossain, A., Kiringa, I., & El Saddik, A. (2006, November). *Towards an Ontology for MPEG-7 Semantic Descriptions*. Paper presented at the Intelligent Interactive Learning Object Repositories (I2LOR) Conference, Montreal, QC, Canada.

Arndt, R., Troncy, R., Staab, S., Hardman, L., & Vacura, M. (2007). COMM: Designing a Well-Founded Multimedia Ontology for the Web. In *Proceedings of the 6th International Semantic Web Conference: The Semantic Web,* Busan, Korea (LNCS 4825, pp. 30-43). Berlin: Springer.

Aroyo, L., Bellekens, P., Bjorkman, M., Houben, G. J., Akkermans, P., & Kaptein, A. (2007). SenSee Framework for Personalized Access to TV Content. In *Proceedings of the Interactive TV: a Shared Experience,* Amsterdam, The Netherlands (LNCS 4471, pp. 156-165). Berlin: Springer.

Blanco Fernandez, Y., Pazos Arias, J. J., Lopez Nores, M., Gil Solla, A., & Ramos Cabrer, M. (2006). AVATAR: An Improved Solution for Personalized TV based on Semantic Inference. *IEEE Transactions on Consumer Electronics*, *52*(1), 223–231. doi:10.1109/TCE.2006.1605051

Ferretti, S., Roccetti, M., & Palazzi, C. (2007). Web Content Search and Adaptation for IDTV: One Step Forward in the Mediamorphosis Process toward Personal-TV. *Advances in Multimedia*, *2007*, 1–13. doi:10.1155/2007/16296

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, *5*(2), 199–220. doi:10.1006/knac.1993.1008

HAVi. (2009). *Home Audio / Video Interoperability*. Retrieved July 25, 2009 from http://www.havi.org

IRT. (2009). *DVB Playout Server*. Retrieved July 25, 2009 from http://www.irt.de/en/products/digital-television/dvb-playout-server.html

Jang, M., & Sohn, J. (2004). Bossam: An Extended Rule Engine for OWL Inferencing. In *Proceedings of the Rules and Rule Markup Languages for the Semantic Web: RuleML,* Hiroshima, Japan (LNCS 3323, pp. 128-138). Berlin: Springer.

Java, T. V. (2009). *JavaTV Specification*. Retrieved July 25, 2009 from http://java.sun.com/products/javatv/

Media, N. E. T. (2009). *The MediaNET Project*. Retrieved July 25, 2009 from http://www.ip-medianet.org

MHP. (2009). *DVB – Multimedia Home Platform Specification 1.1.1, ETSI TS 102 812 V1.2.2*. Retrieved July 25, 2009 from http://www.mhp.org/fullspeclist.htm

Mizoguchi, R. (2004). Tutorial on ontological engineering, Part 3: Advanced course of ontological engineering. *New Generation Computing*, *22*(2), 198–220. doi:10.1007/BF03040960

Naphade, M., Smith, J. R., Tesic, J., Chang, S. F., Hsu, W., & Kennedy, L. (2006). Large-Scale Concept Ontology for Multimedia. *IEEE MultiMedia*, *13*(3), 86–91. doi:10.1109/MMUL.2006.63

Opera. (2009). *Opera for Devices iTV SDK*. Retrieved July 25, 2009 from http://www.opera.com/products/devices/

OSGi. (2009). *Open Service Gateway Initiative Alliance*. Retrieved July 25, 2009 from http://www.osgi.org

POLYSEMA. (2009). *The POLYSEMA Project*. Retrieved July 25, 2009 from http://polysema.di.uoa.gr

Tanaka, K. (2007). Towards New Content Services by Fusion of Web and Broadcasting Contents. In *Proceedings of the 20th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems: New Trends in Applied Artificial Intelligence*, Kyoto, Japan (LNCS 4570, pp. 1-11). Berlin: Springer.

Tsinaraki, C., Polydoros, P., & Christodoulakis, S. (2004). Interoperability support for Ontology-based Video Retrieval Applications. In *Proceedings of the 3rd International Conference on Image and Video Retrieval (CIVR 2004)*, Dublin, Ireland (LNCS 3115, pp. 582-591). Berlin: Springer.

Valkanas, G., Tsetsos, V., & Hadjiefthymiades, S. (2007). The POLYSEMA MPEG-7 Video Annotator. In *Proceedings of the Poster and Demo Conference on Semantic and Digital Media Technologies*, Genova, Italy.

Vilas, A. F., Díaz Redondo, R., Ramos Cabrer, M., Pazos Arias, J., Gil Solla, A., & García Duque, J. (2006). MHP-OSGi convergence: a new model for open residential gateways. *Software, Practice & Experience*, *36*(13), 1421–1442. doi:10.1002/spe.727

ENDNOTE

[1] Note that the specific service does not have a Semantic part

Vassileios Tsetsos received his BSc in Informatics from the Department of Informatics & Telecommunications at the University of Athens, Greece in 2003 and his M.Sc. in "Communication Systems and Data Networks" from the same Department in 2005. Nowadays, he is a Ph.D. candidate in the department. He is a member of the Communication Networks Laboratory (CNL) of the University of Athens and the Pervasive Computing Research Group (p-comp). He has participated in the PoLoS (Integrated Platform for Location Based Services) project, funded by the EU IST programme, and in several national research projects. His research interests are in the areas of pervasive and mobile computing, Semantic Web technologies and middleware for context-aware and sensor-based services.

Antonis Papadimitriou holds a MSc degree from the National & Kapodistrian University of Athens and belongs to the Pervasive Computing Research Group of the same institution. He is also a former member of the ASAP Research Group of the French National Institute for Research in Computer Science and Control (INRIA). His research interests focus on engineering secure and scalable applications in distributed computing environments (P2P Systems, WSNs, Ad-hoc Networks).

Christos B. Anagnostopoulos has received his BSc in Computer Science from the Department of Informatics and Telecommunications at the National & Kapodistrian University of Athens (NKUA), Greece, in 2001 and his MSc in Computer Science, Advanced Information Systems from the same department in 2003. He holds a PhD in Modeling Mobile Computing Systems (2008) from NKUA. His research interest is focused on Mobile Computing, Distributed Systems, Context- aware Computing, Approximate Reasoning (Fuzzy Logic), Semantic Web and Ontological Engineering. He had also participated in projects realized in the context of European Union Programs.

Stathes Hadjiefthymiades received the BSc, MSc and PhD in Informatics & Telecommunciations from the University of Athens, Athens, Greece. He also holds an Joint MSc in engineering-economics

awarded by the National Technical University of Athens. In 1992, he joined the Greek consulting firm Advanced Services Group, Ltd., as an analyst/developer of telematic systems. In 1995, he joined the Communication Networks Laboratory, University of Athens. From September 2001 to July 2002, he was a visiting assistant professor at the Department of Information and Communication Systems Engineering, University of Aegean. In July 2002, he joined the faculty of the Hellenic Open University, Patras, Greece, as an assistant professor. Since December 2003, he has been with the faculty of the Department of Informatics and Telecommunications, University of Athens, where he is currently an assistant professor. He has participated in numerous projects realized in the context of EU program and national initiatives. His research interests include the areas of mobile, pervasive computing, web systems engineering, and networked multimedia applications. He is the author of more than 150 publications in these areas.