Personalized Semantic News: Combining Semantics and Television

Roberto Borgotallo[1], Roberto Del Pero[1], Alberto Messina[1], Fulvio Negro[1], Luca Vignaroli[1], Lora Aroyo[2], Chris van Aart[2], and Alex Conconi[3]

[1] Rai Radiotelevisione Italiana, Italy
[2] VU University Amsterdam, The Netherlands
[3] TXT Polymedia, Italy

Abstract. The integration of semantic technologies and television services is an important innovation in traditional broadcasting in order to improve services delivered to end users in an extended home environment: new methods emerge for getting TV content via the Web and interacting with TV content on end users devices. This paper gives a short description of a Personalized Semantic News scenario in the context of the NoTube project illustrating the use of semantics for personalized filtering and access of news items.

Keywords: NoTube, television services, user-centric, personalized content, device adaptability, semantic annotation, content enrichment, home ambient.

1 Introduction

The TV industry landscape is developing into a highly-interactive distributed environment in which people interact with multiple devices portable devices and home equipment, as well as with multiple applications. People more than ever become early adopters of technology. The Web and these 'new technologies' are steadily transforming this state of the TV industry. New methods emerge for getting TV content via the Web and interacting with TV content on set top boxes. Companies are already attempting to bundle Electronic Program Guides (EPGs) into their software, along with personal recommendation services based on users viewing habits. However, most of those services are still bound to one platform only, e.g. either set top box or Web and stay rather TV-centric. Additionally, users are also increasingly involved in multiple virtual environments (e.g. MySpace, Flickr, YouTube, Amazon, entertainment sites) in each of them with a different identity (e.g. login information, preferences). There is very limited integration and reuse of these user data, or if there exists integration it is not always under the control of the user and there is a lack of transparency in the use of personal data between different applications.

The NoTube[1][1] project aims to overcome these deficiencies and to cover these new requirements using Semantic Web languages and technologies. The ultimate goal

[1] NoTube (*Networks and Ontologies for the Transformation and Unification of Broadcasting and the Internet*) is an EU FP7 Integrated Project (2009-2012).

P. Daras and O. Mayora (Eds.): UCMedia 2009, LNICST 40, pp. 137–140, 2010.

of the project is to develop flexible/adaptive end-to-end architecture, based on semantic technologies, for personalized creation, distribution and consumption of TV content. We take a user-centric approach to investigate fundamental aspects of consumers' content-customization needs, interaction requirements and entertainment wishes, which will shape the future of the "TV" in all its new forms. The project explores three different scenarios: (1) personalized semantic news, (2) narrowcasting advertisement and (3) community based content selection.

In this paper, we focus on the first scenario, i.e. personalized semantic news, provided by RAI Research Center in Torino. We present an overall sketch of the data, services and users involved. We address issues related to the use of semantic tools for the context-aware management of multimedia archive content and its exploitation beyond the creation of traditional TV products. With the advancement of Web technologies and with the convergence of various platforms for the access of multimedia content, new added value services are explored to enable the exploitation of the so-called long tail phenomenon.

2 Personalized Semantic News Scenario

The Personalized Semantic News scenario focus on the creation, distribution and usage of personalized news services that will be able to (1) acquire news items from generic broadcast streams, (2) understand the meaning of video news items, (3) understand the physical context in which news items are shown and (4) apply criteria for matching the user profile with the available news items (see Figure 1 for data flow).

The creation of personalized news services is performed at the *Service Provider Environment*, considering service provider editorial requirements and generic and privacy non-sensitive user profile information; as well as at the *Home Ambient Environment*, considering local context and user information and data enrichment services. The Home Ambient Environment, an extended home environment, consists of two parts:

- *Physical Home Ambient:* a portion of the environment where the user has access to personalized services (context is determined by various sensors input) distributed through the home LAN and enjoyed through different devices.
- *Logic Home Ambient:* virtual space, in which the users, content, metadata and services are identified, filtered and stored. In this space also the content semantics and operational rules are defined.

Three main services are provided at the Home Ambient Environment, e.g. "My News Agency", which automatically generates a personalized local news multimedia channel; "News Alerts", which issues alerts for an incoming News Items relevant to the user's interests; and "News Search" which provides searching capabilities based on semantic filtering of the available News content. Only News Items that match the Home Ambient rules are stored in the home ambient (service level semantic filtering). Each News Item stored in the home ambient has a period of expiration defined either at the provider side or at the user side and it should be locally enriched with metadata

and resources automatically retrieved from local repository or from predefined (following Home Ambient rules) area of the Web (Home Ambient enrichment). Home Ambient Services can be automatically created by grouping News Items (My News Agency, News Search) or based on in-coming events (News Alerts). Following User, Device and Environment rules, dynamic device adaptation is performed on different user devices, e.g. Sofa Television and Hand Television.

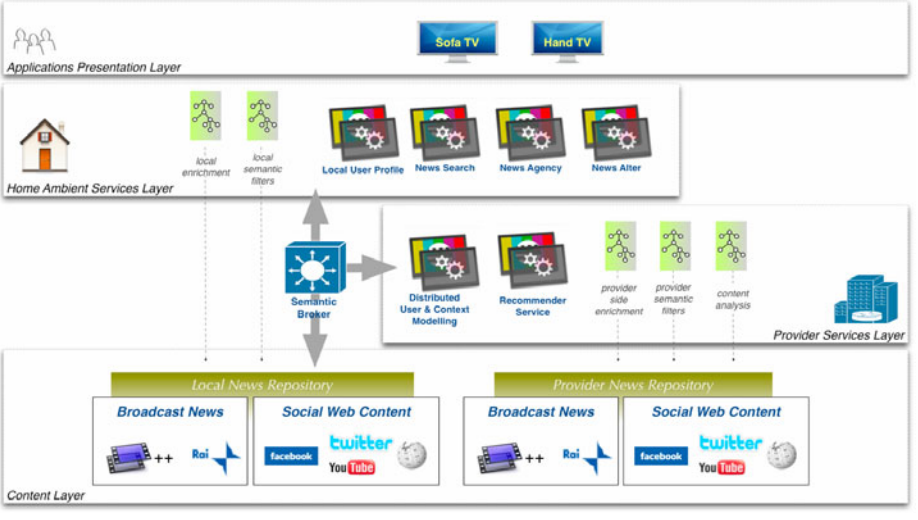

Fig. 1. Basic data flow of the Personalized News scenario

In the *Service Provider Environment* all audio-video segments detected as "News Item" are extracted from the available broadcast content (by means of the Automatic Newscast Transcription System ANTS[2][3]). Each extracted News Item is further enriched with information extracted from internal and external Web resources, as well as with metadata from different related domain vocabularies. Service Provider rules are defined for this enrichment step. The Service Provider rules take in consideration single and group user behavior (in a privacy preserving manner) and apply semantic filtering methods according to the users' (or stereotypes) interests and preferences.

At the Service Provider side we can identify a set components supporting the provider in the preparation of the TV contents, e.g. User and Context services (generic user/context profile categories), Content Annotation and Enrichment services (meta-tagging of TV contents), Metadata Management services (aimed at adapting metadata to professionals/non-professionals) and Model and Semantic services (common background for metadata annotation and enrichment).

The *Semantic Broker* is a core part of the architecture. It discovers the basic internal services organizing them through a specific ontology; it discovers external sources that are semantically related to the application scenario; it exposes meta-services to the upper level (the Application Logic layer) providing an individual entry point for the system, regardless of the physical location of the underlying services; and it composes services in order to perform more complex tasks.

At the Home Ambient side another instance of the Semantic Broker manages internal services, provides semantically related content and services from external sources (typically the Web) and combines locally stored data, e.g. news, advertising, web feeds, etc. with the user and context profiles with the final goal to produce a personalized experience to the end user.

The NoTube infrastructure described in this use case is service-oriented, follows SESA (Semantically Enabled Service Architecture) paradigm and exposes services, both at the Service Provider and at the Home Ambient side, with SOAP, REST APIs, or both. It supports different type of middleware. All the services are organized in four layers: (Internal/External) Contents from Broadcast and the Web, Home Ambient Services and Provider Services layers, and finally the Application and Presentation layer.

3 Expected Added Value

The main added value granted by this kind of service is to give to the user the possibility to get in his home environment programs segments of his own interest represented not only by audio-video, but also by audio-only or text-only or by other metadata (content objects, multi-modal service, dynamic device adaptation) and links to related external resources. The choice of program segments to store in the home environment is automatically done by the local system (user privacy granted) following some user semantic input rules; the user consumes these contents by means of several multi-modal, user adapted, locally created services (semantic local service user/ambient rules, personalized local service).

References

1. The NoTube project web-site, `http://www.notube.tv`
2. Messina, A., Borgotallo, R., Dimino, G., Boch, L., Airola Gnota, D.: An Automatic Indexing System for Television Newscasts. In: Proc. of IEEE ICME 2008 (2008)
3. Messina, A., Montagnuolo, M.: A Generalised Cross-Modal Clustering Method Applied to Multimedia News Semantic Indexing and Retrieval. In: Proc. of WWW 2009 (2009)