AXMEDIS 2008

Workshops on Industrial Applications

Organised by

DISIT-DSI, University of Florence (Italy)

Coordinated by

Prof. Paolo Nesi, University of Florence (Italy)

Nicola Mitolo, AxMediaTech s.r.l. (Italy)

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

dContentWare Project:

Combining semantic web and multimedia distribution technologies to realize innovative business models for the provision of digital contents.

Giuseppe Bux
Graphiservice S.r.l.
Bari, Italy
giuseppe_bux@alice.it

Federica Dentamaro
Gius. Laterza & Figli S.p.A
Bari, Italy
dentamaro@laterza.it

Luigi Intonti, Mina Ligorio
Software Design S.r.l.
Bari, Italy
lintonti@softwaredesign.it
mligorio@softwaredesign.it

Ivana Malatesta, Lorella Lamacchia
AI2 – Applicazioni di Ingegneria Informatica S.r.l.
Bari, Italy
malatesta@ai2.it
lorella.lamacchia@gmail.com

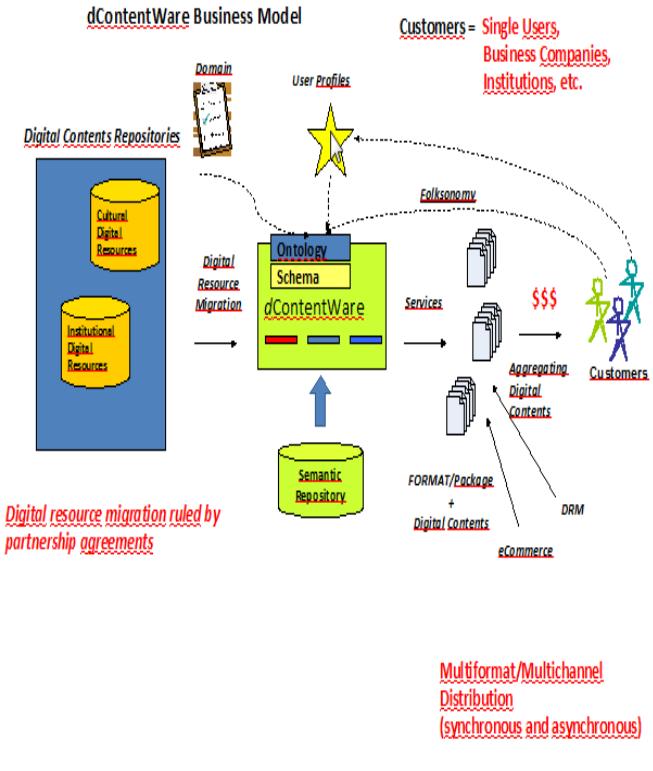
Abstract— **Research on semantic web and multimedia distribution technologies have gained meaningful advances in the recent years; it is now time to experiment, on top of them, innovative business models for the worldwide provision of digital contents: this is the primary goal of the dContentWare project. This paper describes the experience being done in the dContentWare project[1] to reach its primary goal by customizing and combining services provided by JeromeDL, a digital library with semantics and AxMedis, a technological framework for multiformat aggregation and multichannel distribution of digital contents. The proposed business model is being experimented on a concrete use case aiming at deploying a process of semantic migration of bibliographic resources, from the historical archive of Gius. Laterza & Figli publisher into JeromeDL, and their subsequent multiformat aggregation and multichannel distribution through AxMedis.**

Keywords- JeromeDL, Axmedis, digital library, db2rdf, semantic query, social bookmarking, multiformat composition, multichannell delivery, DRM, user profiling.

I. THE DCONTENTWARE BUSINESS MODEL AND THE SET UP OF ITS TECHNOLOGICAL INFRASTRUCTURE

The here proposed dContentWare business model (Figure 1) develops on three deployment frameworks:

1. migration of digital resources from conventional repositories to the dContentWare semantic repository: the migration occurs according to partnership's agreements between the dContentWare consortium and digital contents providers;
2. social bookmarking and semantic searching of digital contents, driven by user profiles;
3. bridging semantic search results towards multiformat and multichannel distribution of digital contents.



[1] The project is co-founded by the government of the Apulia region and is being deployed by an enterprise consortium between the following companies: Gius. Laterza & Figli S.p.a (http://www.laterza.it), Graphiservice Srl.(http://www.graphiservice.it), AI2 Srl., Software Design Srl (http://www.softwaredesign.it)

FIGURE 1. DCONTENTWARE BUSINESS MODEL

The dContentWare infrastructure is built on top of two basic technological platforms:

- JeromeDL[2] : social semantic digital library that makes use of Semantic Web and Social networking technologies to enhance knowledge sharing about digital contents;
- AxMedis[3]: software platform providing services for: authoring multi-format composition models of digital contents; offering configurable research schema of digital contents; aggregating digital contents in predefined combination models; providing multichannel distribution of digital contents by managing Digital Right (DRM) protections.

In the dContentWare business model perspective, JeromeDL acts as the repository of digital contents, while AxMedis acts as the digital contents composer and distributor.

A dContentWare portal (http://www.dcontentware.it), set up in the project, addresses the services provided by JeromeDL and AxMedis, and provides a web service that makes interoperable the JeromeDL repository and the AxMedis composition services.

- A concrete use case of the dContenrtWare business model and its underlying technological framework is currently being deployed and experimented. It is based on the semantic migration of a selected sample of digital resources from the historical archive of Gius. Laterza & Figli S.p.a., one of the leading Italian publishers of humanistic literature, partner of the dContentWare consortium.
- The following sections of this paper describe the customization/implementation process which has been carried out on top of each adopted technological framework, in order to realize and demonstrate the dContentWare business model.

II. DCONTENTWARE: MIGRATING, SOCIALIZING AND SEARCHING DIGITAL RESOURCES IN THE JEROMEDL SEMANTIC REPOSITORY

A. *Migrating*

Digital resource migration is the first process context of the dContentWare business model. It occurs in the frame of a partnership agreement between the owner of an archive of cultural digital resources and the dContentWare consortium.

Usually cultural digital resources are available and managed in conventional repositories, generally in relational databases. Such storage status of digital resources is the first issue to be managed for the concrete succeeding of semantic web technologies. Digital resources need to be extracted from databases available in institutions and made them understandable to semantic web technologies by translating their descriptions in a semantically understandable language, according to the language architecture of the semantic web paradigm [1].

In the dContentWare demonstration use case, the conventional repository of cultural digital resources is constituted by the Historical Archive of Laterza publisher.

The dContentWare migration process takes place in five steps:

1. Analysis and normalization of the logical schema of the original repository: a selection is done of entities and properties involved in the migration process and a resulting entity-relationship model is produced;
2. Specification of a JeromeDL based use case ontology: an ontology model is represented by enriching the JeromeDL ontology model [2] with new properties resulting both from the Analysis step and from emerging design decisions;
3. Design and Building of a transition repository : a normalized transition repository is designed and implemented in a MySql database, by combining the entity-relationship model resulting from the Analysis step, and the new entities/properties emerging from the design of the use case ontology; then a PhP script, with inside MySql queries, extracts resource descriptions from the original archive and moves them into the transition repository;
4. Semantic mapping Db-->JeromeDLOntology: a rdb2rdf mapping specification is produced in D2RQ language [3], in order to map the rdb logical model, of the being migrated archive, to the target JeromeDL rdf based ontology model; the running of the mapping specification, on the D2RQ engine [3], produces a RDF/XML description of the being migrated digital contents;
5. Importing digital contents into JeromeDL: the rdf/xml description of the historical archive is imported directly in the Sesame [4] repository underlying the JeromeDL digital library.

Obviously, JeromeDL provides its own on-line feature to upload resources. In the dContentWare migration process such a feature was used to upload contents components of already migrated resource descriptions, as well to describe and upload single new resources and to edit updating to already migrated/uploaded resources.

---

[2] The JeromeDL platform (http://www.jeromedl.org) results from a research project managed by Main Library of Gdansk University of Technology [http://www.bg.pg.gda.pl/]and DERI.International [http://www.deri.org/].

[3] The AXMEDIS (http://www.axmedis.org) platform results from a research project co-funded by the European Commission in the frame of IST FP6 and involves about 35 partner as: University of Florence, TISCALI, AFI, SEJER, ILABS, EUTELSAT, HP, Telecom Italia, Telecom Lituania, Telecom Estonia, EPFL, FHGIGD, ACIT, Technical University of Catalonia, University of Leeds, etc.

B. Socializing

The original semantic web paradigm [1] was conceived to make machine understandable the web contents. On top of this technological base, a new web is emerging: the collaborative web or web 2.0. Such an evolution is particularly relevant for an innovative business model of digital contents, such as the dContentWare one. It aims at evolving the digital library portals, traditionally conceived as repositories and providers of nude digital contents, towards a new concept of promoters of collaborative generation and sharing of user knowledge about the provided digital contents. Such a new concept mainly characterizes the JeromeDL digital library, so its suitability to support the realization of the dContentWare business model.

The dContentWare users, when acting in the dContentWare/JeromeDL instance, can specify their own private virtual bookshelves (Fig. 2) and their preferred dContentWare digital resources. Each bookshelf is built in terms of a hierarchy of folders of preferred digital resources. Each folder is tagged with concepts extracted from controlled cultural taxonomies originally provided by JeromeDL, as well as from dContentWare own taxonomies resulting from the migration process, as above described. So each folder, with its classification tags, becomes a bookmarking framework of the user preferred digital resources.

Each JeromeDL user is characterized by a FOAFRealm user profile [5], where the friendship relationships of the user with other JeromeDL users are specified. So the user private bookshelf of bookmarked digital contents, can include also the private bookshelves of his friends, so sharing with them their respective preferred digital resources, all tagged by using common taxonomies provided by JeromeDL. This collaborative filtering process is well specified in the SSCF model [6] underlying the bookmarking process in JeromeDL. Other than bookmarking digital resources, the JeromeDL users can comment them and reply to comments of other users in a weblog like feature provided by JeromeDL. An additional JeromeDL feature, on user annotations, concerns their internal representation in RDF, according to the SIOC ontology [7]. It provides a machine understandable representation of user annotations, that can be exported in external more extended weblog environments, such as the WordPress platform, which was adopted by dContentWare in order to animate on-line communities on themes and topics concerned with its provided digital contents.

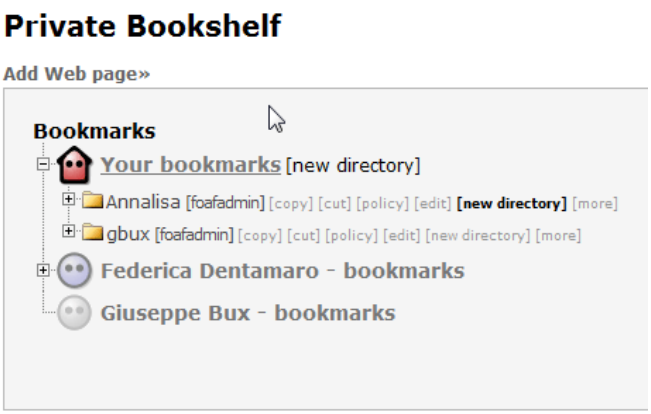


FIGURE 2. USER BOOKSHELF OF BOOKMARKS

C. Searching

The semantic web paradigm [1] introduced semantics in resource description. What characterizes and distinguishes JeromeDL from conventional digital libraries is the capability to manage resource searching on a semantic base. Such a searching capability, in JeromeDL, has five foundations:

1. Rdf [9] description of resources: it is based on metadata, whose semantics is unambiguously defined in the JeromeDL ontology [2] and in its ancillary external ontologies; it generates an internal representation of a resource description as an oriented graph;
2. Semantic vocabularies: they include Jonto [10] ontologies of Wordnet vocabulary and worldwide adopted cultural taxonomies, respectively supporting keyword and domain classification of resources;
3. SeRQL: a powerful RDF query language, provided by Sesame [3], enabling the specification of fine grain criteria in semantic queries, so allowing to rapidly reach resources matching the user search objectives;
4. Regular Expressions: foundation of computer language theories; they are adopted in JeromeDL to specify patterns of queries expressed in Natural Language (NL);
5. Templates of queries: predefined templates of SeRQL and NL queries coded in the system; they allow an user to select the NL query more concerned with his search objective and to instantiate it with search criteria. On the base of the selected and instantiated NL query template, the system in turn selects, instantiates and makes running, on the Sesame [4] repository, the proper SeRQL query template, so extracting the resources matching the search criteria specified in the NL query template.

On the above foundations, it is possible to specify, in the system, effective and user friendly query templates, which enable users to dynamically compose and launch very powerful semantic queries acting on semantic metadata of the whole oriented graph that characterizes a RDF repository.

Obviously, as in the conventional digital libraries, JeromeDL provides string based search mechanisms, based on full text indexing engines, but, as commonly known, the limitation of such an approach is the large grain of the generated search results, because of lack of semantic meaning of research criteria.

The semantic query capability of JeromeDL is particularly suitable to realize the dContentWare Business Model. It in facts allows the dContentWare query designer to customize JeromeDL with templates of rdf semantic queries,

based on the resource aggregation criteria of the AxMedis composition services.

III. dCONTENTWARE: MULTIFORMAT COMPOSITION AND MULTICHANNEL DISTRIBUTION OF DIGITAL CONTENTS

A. Preliminary remarks

A critical issue of the dContentWare Business Model is the identification of appropriate categories of services, providing added value to the final users.

In order to define and experiment presentation forms of such added value services, a fundamental aspect is the capability to aggregate digital resources on semantic base and to compose them in predefined presentation formats, as well as the capability to distribute them on different distribution channels, by assuring DRM protections.

In the specific framework of the dContentWare Business Model, an added value service is characterized by:

- contents: i.e. the digital resources the service works on;
- distribution: i.e. format, channel, scheduling implemented by the service;
- administration : i.e management of fee, rights,... of digital resources.

Taking into account these preliminary remarks, the choice of Axmedis platform has turned out to be extremely functional for the purpose of the dContentWare Business Model. It allows to test the different functionalities of an added value service, according to its envisaged concept in the dContentWare Business Model.

B. Working in dContentWare/AxMedis

The dContentWare/Axmedis users , entering in the system (Fig. 3), can subscribe a specific service by selecting it in the list of the current active ones. Moreover the user can define specific settings related to selection and fruition of digital contents. Each dContentWare/Axmedis user describes in the system his own profile, which is functional to the system for the proper presentation of services to the user itself.

In order to test the features of the different services, being configured for the dContentWare purpose, and thei related user subscriptions, all information generated in the services are stored in a MySQL Database, that is accessible from the Axmedis Content Processing component (AXCP).

The service setting process includes parameters such as:

- scheduling: daily, weekly, etc.
- distribution channel: email, podcast, etc.
- reference domain/category: literature, history, ...
- subject: ancient history, modern history, ...
- source: Laterza publisher archive, Jeromedl repository,
- license: “pay to play,” subscription, ...
- admitted user rights: Modify, copy, ...

In the dContentWare Business Model perspective, AxMedis does not manage its own repository of digital contents, but it gets resources from the JeromeDL semantic repository.

According to the predefined schedule of the service , the AXCP component executes a procedure for its deployment.

The first step of the AXCP procedure manages the acquisition of the digital resources from JeromeDL, via the “Transducer” web service, provided by the dContentWare portal (see next section).

A second step of the procedure, running as from its predefined schedule, composes the resources, get from JeromeDL, into new Axmedis objects. The composition is ruled by predefined logics and formats, specified for the service by filled in SMIL (Synchronized Multimedia Integration Language) [10] templates or by customized html pages. The filled in SMIL templates constitute the requirements scenarios for semantic searching of digital contents in the JeromeDL repository. Dublin Core [11] metadata are adopted to describe the AxMedis object description.

Moreover, based on the user subscriptions to each service, the AXCP component creates a DRM license for each Axmedis object, where digital rights are specified, related to the user, the composer and the adopted distribution device (both on the web and on mobile systems).

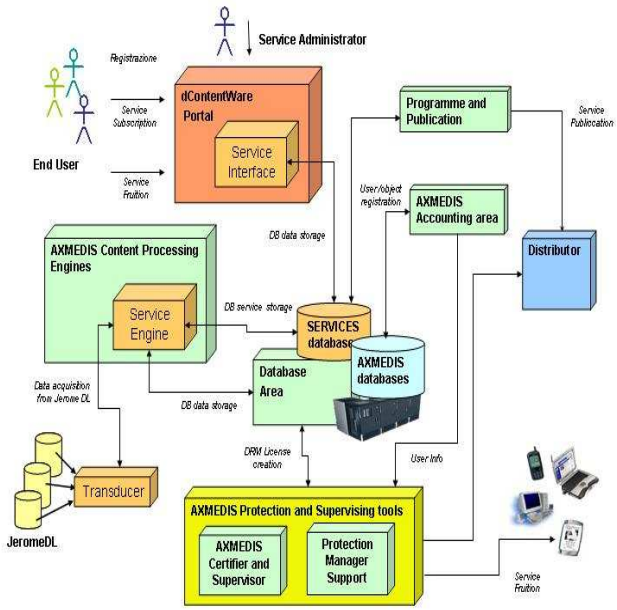


FIGURE 3. INTERACTION BETWEEN AXMEDIS AND DCONTENTWARE

C. The use case of dContentWare/AxMedis services

The current use case of dContentWare/AxMedis services is being deployed on bibliographic resources migrated from the Historical Archive of Laterza publisher into the dContentWare/JeromeDL semantic repository (see previous section).

The AxMedis service configuration has been defined according to the rules defined in the partnership agreement between the dContentWare consortium and the owner/provider of the digital contents repository.

Such rules concern the kinds of provided digital resources and the related distribution licenses. The established kind of digital resources is driving the specification of resource aggregation criteria and composition formats, and in turn semantic search criteria in the JeromeDL repository.

Examples of semantic search/aggregation criteria are:

- resources written/published/edited by a given author/publisher/editor in a given temporal period;
- resources within a being required cultural domain and with a specific key subject;
- resources whose contents refers to a cultural domain and related to a given historical period;
- resources bookmarked by dContentWare/JeromeDL users within a given cultural domain and/or with a given key subject;
- resources belonging to a given resource collection;
- etc..

The being initiated life cycle of the dContentWare/Axmedis services will suggest additional retrieval/aggregation criteria.

Such retrieval/aggregation criteria drive the specification of templates of semantic queries in JeromeDL. These are instantiated at deployment time of an Axmedis service, by enacting the "Transducer" web service provided by the dContentWare portal (see next section).

IV. dContentWare: bridging digital contents from JeromeDL semantic repository to AxMedis multiformat composition services.

As until now described, the dContentWare Business Model is supported by JeromeDL and AxMedis, respectively a digital library and a resource composition and distribution framework. In order to assure interoperability between such environments, two additional technological components were developed in the project:

- the dContentWare Portal, which allows the interactive access to the whole dContentWare environment;
- a "Transducer" of JeromeDL resources versus AxMedis handling services.

A. The dContentWare Portal

The dContentWare portal (www.dcontentware.it) is the main front-end for the user. It introduces the dContentWare Business Model concepts and provides direct access to the dContentWare services, available through JeromeDL and AxMedis, as well as trough the dContentWare ancillary Wiki and Weblog environments.

The JeromeDL services are exposed within an own Jeromedl portal, whose link (www.dcontentware.it/jeromedl) is provided to the user by the dContentWare portal.

The Axmedis services, differently from JeromeDL, are not exposed within an Axmedis portal. The underlying functionalities are presented in an user "Subscription" page provided by the dContentWare portal. Here all the AxMedis composition and distribution procedures of digital contents are listed.

The user can define some settings for each procedure, and then he can subscribe the related AxMedis services, offered him according to specific DRM licenses.

The settings of an AxMedis procedure include different inputs:

- user settings: they are provided at the subscription stage of the procedure and include information such as profile user descriptions, delivery time of the services results, etc.
- AxMedis editor settings, regarding contents aggregation models.

On the base of such settings, the system launches semantic queries in JeromeDL, by instantiating therelated RDF query templates.

A web service application manages the communication between the dContentWare portal and AxMedis, by exhanging data coded in XML Such XML datainclude the list of the available AxMedis procedures and the list of the procedures subscribed by a specific user.

The portal, moreover, provides a web page that, on the base of parameters provided by AxMedis, is able to dynamically produce forms that allows the setting of contents aggregation and distribution parameters. These forms are returned back to AxMedis by an additional web service.

B. The JeromeDL-AxMedis Transducer

Digital contents being composed and distributed through AxMedis are available and retrieved in the JeromeDL digital library. JeromeDL supports different kind of queries (namely: simple, advanced, semantic), where the more suitable for the AxMedis purpose are the semantic queries. JeromeDL query results are returned in RDF [8] format, representation that is not supported by AxMedis. The current 2.1 release of JeromeDL, adopted in dContentWare, provides its query framework via user interface, as well as it provides mechanisms for distributed queries on a network of JeromeDL instances, by using the HyperCup [10] communication protocol. AxMedis in turn provides a

technological framework to enact its services via web services applications. In order to overcome the different technological protocols of JeromeDL and Axmedis, a web service was implemented to make them interoperable. This is the "Transducer" web services (Figure 4): it is interfaced with AxMedis and it allows to launch queries on JeromeDL and to give back results at Axmedis, without using the HyperCup communication protocol adopted in JeromeDL.

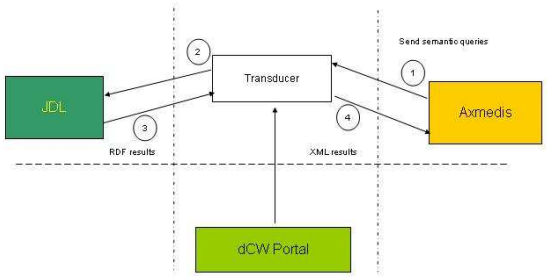


Figure 4. dContentWare Transducer web service

The "Transducer" web service then, by launching queries in JeromeDL and by capturing the query results, in programmatic way, provides the composition and fruition procedures of AxMedis with the required resources.

Relevant feature of the "Transducer" web service is the adoption of the same JeromeDL mechanism to launch a query, as well as the making understandable, to AxMedis, the query results from JeromeDL. The web service methods, in facts, launch a query by generating the related JeromeDL URI, as it would be generated in the JeromeDL user interface context. JeromeDL gives back query results by providing features for representing them in RDF compatible code (XML, n3, ntriple, etc.). The "Transducer" enacts the feature to represent RDF in XML code: this provides a verbose description of the JeromeDL history about a resource. The "Transducer" intercepts the RDF/XML stream, then it captures the pure information about a resource, such as "title", creator", abstract, and URIs of its contents components, and it composes it in a simplified XML document submitted to AxMedis.

Moreover, it is possible to properly configure the "Transducer" in order to allow it to include, in the simplified XML description of a JeromeDL resource, additional rdf tags , without modifying the "Transducer" code.

REFERENCES

[1] Tim Berners-Lee, James Hendler and Ora Lassila: The Semantic Web - http://www.sciam.com/article.cfm?id=the-semantic-web

[2] Sebastian Ryszard Kruk: JeromeDL Ontology 1.0 Specification - http://www.jeromedl.org/structure/spec/index.html

[3] Chris Bizer: The D2RQ Plattform - Treating Non-RDF Databases as Virtual RDF Graphs - http://www4.wiwiss.fu-berlin.de/bizer/d2rq/

[4] Open RDF.org: Sesame RDF Schema Querying and Storage - http://www.openrdf.org/

[5] Slawomir Grzonkowski , Sebastian Kruk : D-FOAF: Role-based Access Control Standard - http://www.foafrealm.org/documentation/AccessControl/

[6] Adam Gzella: Social Semantic Collaborative Filtering - http://library.deri.ie/resource/FUxGypZk

[7] Uldis Bojars, Jhon G. Breslin: SIOC Core Ontology Specification - http://rdfs.org/sioc/spec/

[8] W3C: Resource Description Framework (RDF) - http://www.w3.org/RDF/

[9] Jonto project: Java bindings for ontologies - http://sourceforge.net/projects/jonto

[10] SMIL: Syncronized Multimedia Integration Language - http://www.w3schools.com/smil/default.asp

[11] Dublin Core: Dublin Core Metadata Initiative - http://dublincore.org/

[12] Slawomir Grzonkowski , Sebastian Kruk : HyperCup Lightweight Implementation - http://library.deri.ie/resource/AlZpcTQx.