SMART AUTHORING AND SHARING OF MULTIMEDIA CONTENT IN PERSONAL AREA NETWORKS BASED ON SUBJECT OF INTEREST

Belal Abu-Naim | *Wolfgang Klas*

University of Vienna,
Multimedia Information Systems
a0600380@unet.univie.ac.at,

University of Vienna,
Multimedia Information Systems
wolfgang.klas@univie.ac.at

ABSTRACT

The evolution of smart phones' hardware and operating systems, users tendency to join social networks and to share multimedia content and daily life events, well-established methods and technologies of Semantic web, and the increasing establishment of Linked Open Data (LOD) APIs, motivate us to introduce a new approach in multimedia content composition and sharing in personal area networks that automatically analyzes, selects, composes, and shares the authored content. The capabilities of social network applications and the applications that address multimedia document composition, retrieval and presentation, and multimedia content sharing, do not go beyond allowing the users to share text, pictures, or other types of media content in social networks, performing manual or semi-automatic multimedia document composition, retrieving a list of pre-composed multimedia documents that eventually include datasets retrieved from DBpedia based on the geographic location. There is a lack of applications that are capable to automatically analyze the multimedia content on the devices of the users, compose multimedia documents about the Subject of Interest (SOI), retrieve and use additional data from LOD sources, and achieve a cross-multimedia document models authoring. In this paper we introduce our innovative approach of automatic analysis, composition, and sharing of multimedia content driven by a user's subject of interest (SOI). Our new approach enables us to achieve a smart multimedia authoring and sharing by incorporating new phases within the authoring process, which have not yet been applied by other applications.

Index Terms – Mobile Multimedia Authoring System

1. INTRODUCTION

Many applications and approaches address different aspects in the domain of mobile multimedia content management, including content retrieval, adaptation, personalization, viewing, authoring, and sharing. Those activities could be initiated by different types of triggers, such as, multimedia content could be retrieved based on the geographical location as proposed in [1], automatic photo album generation could be achieved by analyzing the high and low level features of the images as in [2], authoring of multimedia content guided by an application that provides a user interface to make the multimedia content authoring on mobile devices more achievable as in [3] and [4].

In this paper we draw the complete picture of an innovative multimedia authoring approach that targets personal area networks (PANs). PANs are a rich environment for authoring and sharing of multimedia content. Our approach is motivated by four aspects, the evolution of smart phones, the rapid growth of social network communities, the well-established technologies of semantic web, and the increasing number of linked open data access points. Unlike other approaches and applications, the driving key element of the authoring process is what we define as the Subject of Interest (SOI), which we infer and use to initiate the process of multimedia content authoring and sharing.

Smart phones have been evolved rapidly, their hardware capabilities as well as their operating systems. Thus, the user behavior have been accordingly influenced by this evolution. Users became more depended on their devices in managing different aspects of their daily life and their social activities. Multimedia content, more often self-produced or acquired from external sources, became available on smart phones. In parallel social network platforms such as Facebook, Google+, and LinkedIn are expanding rapidly. Mobile device users constitute a considerable part in those platforms, which indicates that smart phones and tablets can play a vital role in social activities and accordingly in multimedia content authoring and sharing.

The development of semantic web and linked open data led to interesting aspects with respect to standardization: The semantic web led to a standardized way of interrelating information and of reasoning about relationships between information items to derive implicit information. The linked open data approach led to a standardized way of making data available for applications via W3C standards and web services. Both technologies can be integrated and used in the process of multimedia content authoring and sharing, providing new opportunities for the automatic composition of multimedia content.

2. RELATED WORK

In the domain of multimedia authoring and presentation the efforts of the scientific community have yielded a significant amount of work. We briefly sketch the approaches and applications which we consider to be relevant to our work, we classified them into two groups, context-driven multimedia retrieval and location-based applications, and multimedia authoring and editing tools.

2.1. Context-driven multimedia retrieval and location-based applications

MobiDENK [1] is a location-based application that retrieves a HTML multimedia presentation that contains text and images about nearby historic sites. In [5] a set of requirements for developing cultural applications on PDAs and mobile devices is presented. The authors reviewed the state of the art technologies for developing mobile applications for visiting cultural sites as well as tourists traveling, and made case studies for three different applications, an authoring tool for the development of a cultural multimedia application on PDAs, a news reader for mobile phones, and an electronic tourist guide on mobile phones. In VGP [6] the user can download an application which grants access to a service point that allows to view content by means of navigating through a menu system. In [7] a personalized multimedia application is presented which is compiled for a specific user based on her profile. Users can read the documents with a special multimedia viewer. Again, the user can view the included multimedia content only by navigational means through a menu system. In [8] a metadata ontology is used to ease retrieving multimedia presentations from the web. Annotations which are used for retrieval of presentations are represented in RDF and embedded within the multimedia documents that support RDF, like SMIL and SVG. In [9] a platform for sharing content produced by the users using mobile devices is presented. To enable a more enhanced retrieval of user-generated multimedia content and to overcome the limitations of keyword-based search - after uploading the content to the server - additional information such as GPS data, timestamps, user defined tags etc. are associated with the content in triples. Such kind of enrichment enables the usage of more powerful querying making use of annotations and semantics.

2.2. Multimedia authoring and editing tools

xSmart [10] is a semi-automatic multimedia content composition tool that supports the author throughout the steps of the process of multimedia content composition. It offers a very basic set of functionalities of multimedia composition such as the selection of multimedia content and the creation of a simple layout in a page-oriented fashion. InAuthor [11] incorporates a visual approach to activity creation. The authoring process supports drag and drop of content elements such as text and images into the workspace, arranging their position, and tagging with a geographic location. In [4] a manual multimedia content composition approach is presented, which provides an editor interface that assists the user to compose multimedia content on the mobile device by systematically following a user-centered design approach. The editor interface consists of four views: presentation view, player view, edit presentation view, and presentation preview. mProducer [3] is a system for generating personal experience content on mobile devices which has four components: a Storage Constrained Uploading (SCU) algorithm, a sensor to detect and remove the blurry frames, a map-based content management interface, and a key frame-based editing tool. The process of authoring comprises two phases: During the capturing phase data are captured and saved either on the device or on the server. This phase includes preparing and finding key frames and filtering blurry frames. During the editing phase the user selects content to be edited by choosing a point on the map. Then a list of content elements which are organized based on capturing location is displayed. iPhotobook [2] is an approach to create photo books on mobile devices which integrates algorithms for image selection, cropping, pagination, page layout, and background scaling. Users utilize a GUI that provide touch and motion gestures for arranging the photo album. In the more conceptual model of an Ecosystem for Semantics [12] underlying an authoring tool for an extended photo album consisting of not only photos, but also audio and video, the authoring process uses some rules represented in Object Constraint Language (OCL), which have already been extracted from the semantics of analyzed photo albums. The target of the approach presented in [13] is to provide a multi-channel multimedia presentation generation by adopting an abstract document model that covers the central characteristics of multimedia presentation formats and which could be translated to a concrete multimedia presentation based on individual device configuration. In [14] a canonical process covering the general creation chain for authoring personalized and semantically rich multimedia presentations is defined. The general creation chain consists of four phases: selection, assembling, transformation, and presentation of content. These four phases cover eight canonical processes for media production: creation of media assets and querying are part of the first phase. Annotation, constructing messages, and organizing content are part of the second phase. Packaging and publishing are part of the third phase. Distributing for presentation is part of the last phase. The defined processes were implemented in SemanticMM4U. A framework for ontology-driven multimedia analysis and composition including two complementary approaches were introduced in [15]: a semantic framework for multimedia retrieval, which automates the generation of annotations of media elements from high level semantics and low level attributes, and a semantic framework for multimedia composition, which is based on reasoning over domain knowledge. This results in an adaptive and automated multimedia generation process.

3. DCMC: DYNAMIC CROSS-MODEL COMPOSITION

With DCMC we aim at the composition of multimedia content based on knowledge explored and extracted from content individual users have and are willing to share and which is encoded in different kinds of multimedia models. The authoring process in DCMC is guided by means of the semantics extracted from the original sources. Let us illustrate this by means of the following scenario: Two colleagues visited a Zoo during the vacation. In the Zoo they downloaded some multimedia content about the animals hosted. The two colleagues share the same subject of interest (SOI), and since both of them are socially active and usually tend to share multimedia content and live experiences on social network platforms, they also use DCMC to share multimedia content. As they share the office room, their smartphones have automatically established a personal area network and can exchange data. The authoring process of DCMC has inferred the SOI and as a part of the process the newly composed document was enriched with content retrieved via the LOD APIs of DBpedia. After the process has completed, both colleagues were notified about the new multimedia content. Each colleague has opened the new document, and found an interesting multimedia content about "Mammals" retrieved from LOD sources, accompanied with images they have taken and seen before.

3.1. Requirements of Multimedia Content Authoring in personal area networks

We have defined two groups of requirements for authoring and sharing multimedia media content in personal area network (PAN). (1) Requirements for personal area network establishment and data exchange. In this work, we have opted to use Bluetooth for establishing the PAN. Amongst the many protocols of Bluetooth, we are particularly interested in Personal Area Networking profile [16] and Serial Port Protocol profile (SPP) [17]. (2) Requirements for the multimedia authoring process. The system should be able to find and analyze the multimedia documents serving as potential sources for the composition. The result of the analysis is to be encoded as RDF-triples in order to be able to process the semantics. This includes all low-level features as well as the semantics related to the source documents, and the individual media elements included in these documents. This allows then (i) to infer the shared SOI within the group of users in the PAN, (ii) to enrich the original content during the authoring process with additional information, (iii) to compose, and (iv) to share the final document.

3.2. Authoring Process

The core of multimedia authoring process in DCMC follows the generic process chain of multimedia composition as presented in [18]. DCMC extends that process and adds additional phases, which yield an authoring process consisting of eight phases: 1. *Analysis*: the multimedia documents the users share are analyzed and the results are encoded in RDF. 2. *PAN establishment*: a personal area network is established between two or more devices. 3. *SOI inference*: one or more SOIs are inferred from the multimedia content. 4. *Enrichment*: additional information about the SOI is retrieved from LOD sources. 5. *Selection*: media elements related to the SOI are selected for inclusion in the authoring process. 6. *Composition*: infer the positioning of the selected media elements with respect to the temporal and spatial spaces and encode the results in RDF 7. *Transformation*: transform the composed multimedia document that is described in terms of RDF triples to the target multimedia document format. 8. *Presentation*: the newly authored document is transferred and presented on the devices.

3.3. System Design and Architecture

Having defined the requirements and the authoring process, we present the architecture of DCMC. It consists of seven logically separated components: 1) *Networking Framework*, which includes all services required for establishing a PAN, HTTP connections, and transferring data between devices. 2) *Framework for Multimedia Documents and RDF (FMMDR)*, which is a framework for mapping different models of multimedia documents such as SMIL, SVG, or HTML5 to RDF, and mapping RDF-encoded documents back into a multimedia document models. To support a certain type of multimedia document model, for example SMIL documents, an implementation of the APIs of the framework for that document model should be provided. 3) *Analyzer*: this component enables the system to work with RDF documents and to apply the rules to infer the SOI and other data required during the authoring process. 4) *Selector*: this component selects the media content that mostly relates to the SOI. 5) *Composer*: it merges and composes the selected media content, considering the temporal and special relationships. 6) *LOD API Framework*: it is a framework for communicating with LOD sources. Similar to FMMDR, to support communication with a certain LOD API, for example DBpedia, an implementation for the particular LOD API should be provided. The framework provides four modules to communicate with LOD APIs, a) the Query API for constructing the query, whereas the query could be either a SPARQL query or a GET/POST call to a RESTfull web service, b) Parser API, which allows to parse a LOD source document into objects, c) *Mapping API*, which allows to map the parsed objects to the internal data model of the DCMC, d) *HTTP API* providing the facility to establish connections over HTTP in order to send and receive data. 7) *Inference Engine*, which is invoked by other components of the system during the life cycle of the authoring process.

4. IMPLEMENTAION, EXPERIMENTS AND EXPLANATION OF RESULTS

4.1. Networking Framework

We have opted to use Bluetooth technology and implemented the Networking Interfaces accordingly. The Networking Framework offers three main interfaces: *IPANService* which offers all the required services such as connect(), send(), receive(). *IPANMessage* which represents the messages to be exchanged between devices, and *IPANListener* which provides callback functions which are called by *IPANService*. Callback functions include onConnected(), onMessageSent(), onMessageReceived(), onDisconnected().

4.2. Framework for Multimedia Documents and RDF (FMMDR)

For illustration purposes, we have opted to provide an implementation for SMIL documents. The framework has several modules and utilities: The local storage scanning service is a File I/O service that searches for .smil or .sml in the shared documents' folder and captures a list of available SMIL documents. The SMIL2RDF service is the implementation of the *Multimedia Document to RDF* service for SMIL documents. The SMIL2RDF has three processes: the conversion of the XML that represents the SMIL document to its equivalent Document Object Model (DOM), the conversion of the SMIL-DOM to RDF, and the analysis of the individual media elements. We have implemented the classes of the model that represents a SMIL document in Java. For creating and manipulating the RDF document we have used the Jena framework [19] for Android devices. For text analysis we used Lucene framework [20]. For reading EXIF data of the images, we have implemented a class in Java.

SMIL2RDF can be further described in terms of input, output, and process:
Input: a set of SMIL multimedia documents.
Output: RDF graph that represents multimedia documents and captures all the required information for the authoring process.
Process: Each SMIL document is deserialized into a DOM representation. The root subject of the RDF graph is the document itself. All attributes of the root document and its sub-elements are represented as RDF triples. The individual media (e.g., Text, Image) elements found in the SMIL documents are also analyzed and added to the RDF graph. Analyzing the text results in identifying keywords according to the subjects covered by the SMIL document. The EXIF data and other metadata of the images are also incorporated in the analysis.

RDF2SMIL is the reverse process to SMIL2RDF, where the RDF document can be converted back to a SMIL document. The component can be described again in terms of Input, Output, and Process as follows:
Input: RDF document.
Output: SMIL document.
Process: The RDF triples are converted to SMIL elements. Table 1 shows a small snippet of a SMIL document, the <root-layout> element, in its RDF and SMIL representations.

Table 1: RDF2SMIL - Transformation of RDF to SMIL

```
RDF:
<rdf:Description
rdf:about="http://univie.ac.at/dcmc/Lion.smil/L1">
   <dcmc:id>L1</dcmc:id>
   <dcmc:width>800</dcmc:width>
   <dcmc:height>600</dcmc:height>
   <dcmc:type>root-layout</dcmc:type>
</rdf:Description>

SMIL
<root-layout id="L1" height="600" width="800"  />
```

4.3. Analyzer

The Analyzer is responsible for detecting a set of SOI common to all or some of the shared documents. Again, the component can be described in terms of Input, Output, and Process as follows:
Input: RDF documents, inference rules, and SPARQL statements.
Output: set of SOI.
Process: We defined a set of rules and SPARQL statements to identify the set of SOI. In this work we analyze two parts of the RDF documents: the keywords part, which is the result of performing text analysis on text content, and the metadata part, which is extracted from the metadata tags of the multimedia documents and its media elements. The set of SOI is then the result of applying the rules and SPARQL statements to all the input documents.

4.4. Selector

The selector is responsible for selecting the media elements required for the composition according to the SOI. The description is as follows:
Input: a set of SOI and inference rules.
Output: a set of media elements.
Process: The first step in the selection process is to find the media elements that are related to the set of SOI. The second step is to select the most relevant media content to each SOI, which will be included in the composed document. In order to determine which media elements are more relevant with respect to the set of SOI, we calculate the distances between the feature vectors and the SOI [21]. Media elements found to have a distance within a specific threshold will be selected. In this work we have opted to consider for selection only the media elements of type Image and Text. Quite an obvious rule

to provide new information in addition to the original documents, original text elements will be selected only if no new text elements were retrieved from LOD resources during the *Enrichment* phase.

4.5. LOD API Framework

The LOD component realizes the retrieval of data from LOD sources. It can be described as follows:
Input: set of SOI.
Output: Media elements retrieved from LOD.
Process: This component provides an implementation for the interfaces provided by the *LOD API Framework.* We have opted to implement the interfaces for DBpedia.

4.6. Composer

The Composer realizes the composition of the final document and can be described as follows:
Input: the set of selected media elements, the RDF representations of the original documents, composition rules, and optionally a set of multimedia document composition templates.
Output: RDF document that represents the final composed multimedia document.
Process: Find the subgraph of each selected media element in the RDF representations. On the one hand we use the concise bounded description (CBD) [22] to calculate these subgraphs. Since CBD finds all the statements where the subject is a starting node, the resulting subgraph lacks the statements where the media element represents the Object of the RDF triple. On the other hand, since we need to find the location of the element in the original multimedia document, we have to find the statements that describe the composition features of the Subject media element. For example, in a SMIL document the media element is placed in a region. The region can have a parent region, and all regions are located under the root layout, thus we need to recursively trace the regions in the RDF tree and include each node found in the subgraph until we reach the root-layout. Figure 1 depicts the calculated subgraph of the media element (Loin.jpg). The composition templates are predefined multimedia document structures. In a particular use case the *Composer* may be pre-configured to always make use of a given composition template. In another use case the *Composer* may use a template as an alternative or fall back because of lacking useful information on the composition of the elements.

4.7. Inference Engine

The inference engine is built on top of Jena framework [19] and is invoked during the composition process by other components of DCMC. The input are RDF statements and a set of rules and SPARQL statements, the output, e.g., set of SOI or media elements, differs according to the invoking component.

Our illustrating experiment uses a pair of two devices and four multimedia documents. Table 2 summarizes the input and output of each phase of the authoring process for a demonstration of the prototype.

Figure 1: the calculated subgraph of the media element (Lion.jpg)

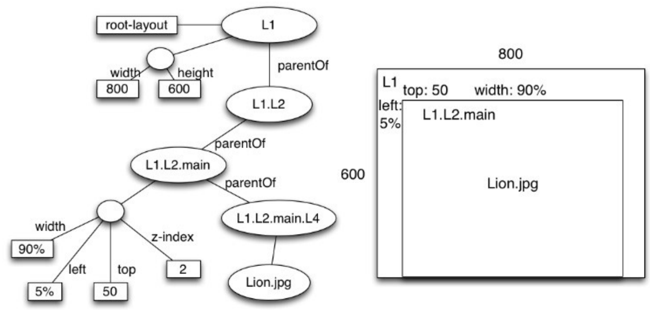

Table 2

Phase	*Input*	*Output*
PAN Establishment	Smartphone (S1), Tablet (T2)	PAN
Analysis	Lion.smil, Koala.smil, Panda.smil, Tiger.smil	S1-RDF, T1-RDF
SOI Inference	S1-RDF, T1-RDF, SPARQL, Rules	Set of SOI = {Mammal + Wild}
Enrichment	Set of SOI = {Mammal + Wild}	1 Article about Mammals, and ambiguous search result "Wild"
Selection	Inference rules Text: Mammals, Lion, Koala, Panda, Tiger Images: Lion, Koala, Panda, Tiger	Text Mammals Images: Lion, Panda
Composition	Text: Mammals, Images: Lion, Panda, Rules, S1-RDF, T1-RDF	New Multimedia content. Subgraph: Mammals.RDF
Transformation	Subgraph: Mammals.RDF	Document Mammals.smil

5. CONCLUSION AND FUTURE WORK:

We have presented a new, innovative multimedia authoring approach, called DCMC. Unlike other multimedia authoring approaches and applications we discussed in section 2, the authoring process in DCMC is supported by and makes use of Semantic Web technologies in its different phases. It is driven by what we have presented as the Subject of Interest (SOI) of users, inferred from content shared by various users within a PAN. By converting the multimedia source documents to RDF statements, DCMC allows for the enrichment of content by interlinking with sources from the

LOD-cloud. Inference rules guide the selection of additional content elements relevant to the SOI. The authoring process results in newly composed and expanded multimedia content according to individual composition rules and/or pre-specified composition templates. The new content is finally transformed from RDF into the document format needed for distribution and presentation. The approach has been implemented as a flexible framework and - for the purpose of illustration – it has been demonstrated using SMIL documents as input and output format and DBpedia as a source within the LOD-cloud.

6. REFERENCES

[1] J. Krösche, J. Baldzer, and S. Boll, “MobiDENK-Mobile Multimedia in Monument Conservation,” *IEEE MultiMedia*, vol. 11, no. 2, pp. 72–77, 2004.

[2] J. Xiao, N. Lyons, C. B. Atkins, Y. Gao, H. Chao, and X. Zhang, “iPhotobook: Creating Photo Books on Mobile Devices,” in *Proceedings of the International Conference on Multimedia*, New York, NY, USA: ACM, 2010, pp. 1551–1554.

[3] C.-M. Teng, C.-I. Wu, Y.-C. Chen, H.-h. Chu, and J. Y.-j. Hsu, “Design and Evaluation of mProducer: A Mobile Authoring Tool for Personal Experience Computing,” in *Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia*, New York, NY, USA: ACM, 2004, pp. 141–148.

[4] T. Jokela, J. T. Lehikoinen, and H. Korhonen, “Mobile Multimedia Presentation Editor: Enabling Creation of Audio-visual Stories on Mobile Devices,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, 2008, pp. 63–72.

[5] D. Economou, D. Gavalas, M. Kenteris, and G. E. Tsekouras, “Cultural Applications for Mobile Devices: Issues and Requirements for Authoring Tools and Development Platforms,” *SIGMOBILE Mob. Comput. Commun. Rev,* vol. 12, no. 3, pp. 18–33, 2008.

[6] M. D’Souza, A. Postula, N. Bergmann, and M. Ros, “A Multimedia Guidebook Implementation Using a Bluetooth Wireless Information Point Network,” in *Proceedings of the 3rd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, New York, NY, USA: ACM, 2005, pp. 33–38.

[7] M. Kenteris, D. Gavalas, and D. Economou, “An innovative mobile electronic tourist guide application,” *Personal and Ubiquitous Computing*, vol. 13, no. 2, pp. 103–118, 2009.

[8] C. Saathoff and A. Scherp, “Unlocking the Semantics of Multimedia Presentations in the Web with the Multimedia Metadata Ontology,” in *Proceedings of the 19th International Conference on World Wide Web*, New York, NY, USA: ACM, 2010, pp. 831–840.

[9] O. R. Rocha, C. Criminisi, F. Mondin, and L.-W. Goix, “LODifying Personal Content Sharing,” in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, New York, NY, USA: ACM, 2012, pp. 124–130.

[10] A. Scherp and S. Boll, “Context-driven Smart Authoring of Multimedia Content with xSMART,” in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, New York, NY, USA: ACM, 2005, pp. 802–803.

[11] F. Barrenho, T. Romão, T. Martins, and N. Correia, “InAuthoring Environment: Interfaces for Creating Spatial Stories and Gaming Activities,” in *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, New York, NY, USA: ACM, 2006.

[12] A. Scherp and R. Jain, “An Ecosystem for Semantics,” *IEEE MultiMedia*, vol. 16, no. 2, pp. 18–25, 2009.

[13] A. Scherp and S. Boll, “Paving the Last Mile for Multi-Channel Multimedia Presentation Generation,” in *Proceedings of the 11th International Multimedia Modelling Conference*, Washington, DC, USA: IEEE Computer Society, 2005, pp. 190–197.

[14] A. Scherp, “Canonical processes for creating personalized semantically rich multimedia presentations,” *Multimedia Systems*, vol. 14, no. 6, pp. 415–425, 2008.

[15] S. Dasiopoulou, I. Kompatsiaris, and Nixon, Lyndon J. B, “A Framework for Ontology-driven Multimedia Analysis and Composition,” in *Proceedings of the Eight International Workshop on Image Analysis for Multimedia Interactive Services*, Washington, DC, USA: IEEE Computer Society, 2007, pp. 47-.

[16] *Bluetooth Personal Area Networking Profile* (2014, Jan. 25).

[17] *Bluetooth Serial Port Protocol* (2014, Jan. 25).

[18] A. Scherp and S. Boll, “MM4U: A Framework for Creating Personalized Multimedia Content,” in *Managing Multimedia Semantics*, U. Srinivasan and S. Nepal, Ed.: IGI Global, 2005, pp. 246–287.

[19] *JENA Framework.* Available: http://jena.apache.org (25.01.2104).

[20] *Lucene Framework.* Available: http://lucene.apache.org (2014, Jan. 25).

[21] G. Grimnes, P. Edwards, and A. Preece, “Instance Based Clustering of Semantic Web Resources,” in *Lecture Notes in Computer Science, The Semantic Web: Research and Applications*, S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, Eds.: Springer Berlin Heidelberg, 2008, pp. 303–317.

[22] *Concise Bounding Description.* Available: http://www.w3.org/Submission/CBD/ (25.01.2104).