International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 4227-4231

Towards Increasing and Personalizing of User Experience in the Digital Culture Ecosystem

Desislava I. Paneva-Marinova[1], Alexander I. Iliev[1,2], Radoslav D. Pavlov[1], Lubomir Ch. Zlatkov[1]

[1] *Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 8, Akad. G. Bontchev Str., Sofia, 1113, Bulgaria.*

[2] *University of Wisconsin, Stevens Point, 2001 Fourth Ave., B241 Science Building, Stevens Point, Wisconsin, WI 54481, USA.*

Abstract

In this paper we address the issue of rapidly increasing data by introducing the emergence of novel technological advancements leading to better organization and structuring of data in digital ecosystems. These systems claim to better manage big data and massive of heterogeneous information, offering innovative services for intelligent self-management, self-configuration and self- maintenance. The paper focuses on the problems of providing an improved user experience in the digital culture ecosystem through new methods for personalized "real time" content integration and recommendation We suggest an alternative approach of finding, personalizing and recommending their digital content to users. Furthermore, we propose a method that is using speech and sentiment in particular, which is collected non-intrusively through speech recording and processing. This methodology is based on previously established cues extracted from speech that may lead to the production of advanced practical personalization and recommendation systems for broader usage, rather than only pertaining to specific ecosystems. We used the current ecosystem approach to verify that non-intrusive collection of biometrical cues is a necessary advancement for any system dealing with large datasets that directly interacts with users. Biometric data varies in the way it is collected. For the problem at hand it was important to only focus on ways that were not invasive or disturbing in any way. This was important because user behavior should not be influenced in any way when it comes to working in real world environment, hence speech recordings of voice were considered appropriate for the implementation.

Keywords: Digital Cultural Ecosystem, Visitor Experience, Personalization, Recommendation, Speech Sentiment, Emotion Recognition, Biometric Voice Data

INTRODUCTION

The technological growth and the exponentially increasing usage of digital components have lead to the emergence of new intelligent environments, namely the "digital ecosystems". These environments, as part of the global network, offer access to a number of independent formations (services, software platforms, applications, individuals, organizations), which share common goals and focus on the interactions between them. The digital ecosystems manage big data and systems of heterogeneous information, offering innovative services on their own intelligent platforms. With the aid of high-tech computational intelligence, the digital ecosystems exhibit new characteristics like self-management, self-configuration and self-maintenance, through recurring intelligent combination and evolution of "digital components". Furthermore, the resources, available from each component are stored, managed and used, while flexibly adapted and personalized towards the needs of individual users.

The application of the digital ecosystems for presentation and access to cultural heritage is a modern and dynamically developing field. It appears to respond to the growing willingness to share the wealth of cultural resources to new audiences in novel ways.

Due to the multidisciplinary characteristics of the digital cultural ecosystems (DCEs), their design and development is especially challenging. This may lead to poor understanding of the potential high added value, made possible by the good management of the resources. Moreover, one of the main challenges for DCE is to provide effective assess to its content and personalize the user's experience to fit his/her current goals, interests and needs in the best possible way. From a general user's perspective, this would mean accessing the DCE objects in a way that best fit his/her cognitive needs and preferences. From a content provider's perspective, this would mean to be able to present or to "transform" the DCE content to develop meaningful and effective user experiences for different contexts, and needs. These scenarios however are not efficiently supported in current digital cultural ecosystems.

This paper focuses on the problems of providing an improved user experience in the digital culture ecosystem through innovative methods for personalized "real time" content integration and recommendation. Section 2 presents digital culture ecosystem paradigm and the processes of content aggregation, observation, and their study, as well as the users' roles and activities in the mentioned context. An approach for "real-time" DCE content integration is presented in section 3. Section 4 discusses a possible implementation of an automatic recommendation system based on personalized cues extracted from speech.

DIGITAL CULTURAL ECOSYSTEM PARADIGM

The paradigm of ecosystems for digital cultural assets (also called digital cultural ecosystems, DCEs) appears to respond to the growing willingness to share the wealth of cultural resources and continuous research and study of cultural treasures. These systems virtually assemble various digital collections, archives, virtual museums, digital libraries and cultural heritage sites in order to facilitate the access to their resources, bringing cultural content to new audiences in novel ways. Digital cultural ecosystems demonstrate wide range of applicable services and tools for re-using and repurposing digital assets (or objects, DCOs), paving the way for wider exploitation of cultural resources and boosting innovation. Figure 1 and 2 depict the DCE main content units, user's activities for manipulation, and the content flow.

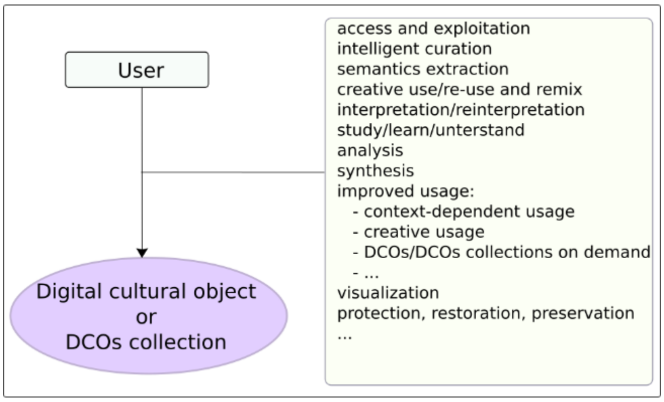

Figure 1. DCE content units and activities for their manipulation.

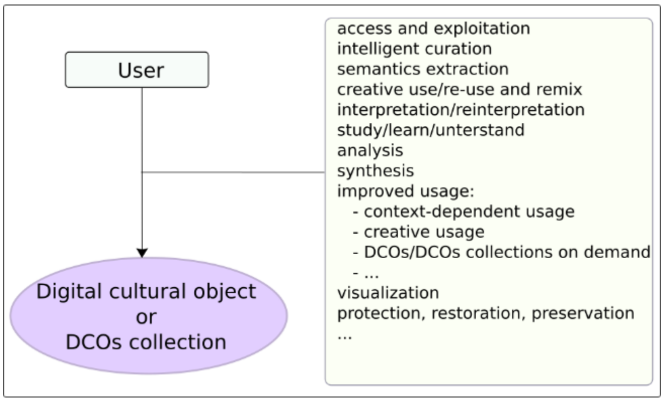

Figure 2. DCE content units and activities for their manipulation. DCE content flow.

Digital cultural objects are the smallest content units in the DCE. Context-based grouping of DCOs creates a collection with a wide variety of usage. The DCOs could be selected according to their type and area, author, style, school, location, date, subject, origin, context of usage, etc. The collections are saved in catalogues for further inclusion and display in exhibitions.

In general, the user's manipulations with DCOs and DCOs collections are related to: access and exploitation, curating, semantic extraction, use/re-use and remix, analysis, study, etc. [1].

The main content units and the activities for their manipulation can be extended according to the concrete DCE' aims, marketing and advertising strategies, target groups, etc. In this study, we concentrated on models and visions for improved use, research and delivery of digital cultural resources in DCE.

AN APPROACH FOR "REAL-TIME" DCE CONTENT INTEGRATION

"Real-time" integration of the DCE content from different DCE sub-platforms following the user's needs could be defined as an automatic content aggregation, based on the available DCE objects and the user's specific requirements [2,3]. Special DCE services track the content repositories and match the requirements of the users. Their needs, interests, goals and preferences have to be carefully studied and the results need to become the starting point for this DCE functionality. This "real-time" content integration could be defined also as a type of personalization. In general, the DCE functionality designers need to know: What are the users' motivations for exploring the target DCE? What are the factors driving the users to engage in observational activities? What impact did the users' prior experiences have? How do the users define their needs? What are the users' actions in the DLE? How is something done or accomplished? The DCE platform increases and personalizes the user experience through:

- Selection and recommendation of information resources according to the personal characteristics of the user on one hand and according to the user's behavior in the environment on the other;
- Adaption of the navigation means, the display format and the ways of providing information resources" to the user.

The formal adaptive logic linking certain user's actions (or user profile) with a DCE system response (single or series of responses), are delivering selected digital objects [2,4], as specified bellow.

Let

$A = \bigcup_{i=1}^{l} A_i$ be the taxonomy of the adaptation aims (for example, content adaptation, interface adaptation, *etc.*), where l is the overall number of adaptation aims;

$B = \bigcup_{i=1}^{s} B_i$ be the taxonomy of the adaptation ways (for example, showing explanations for object/s, showing a selection of objects/parts of the objects, showing groups of objects according to some criteria, *etc.*), where s is the overall number of adaptation ways;

$X = \bigcup_{i=1}^{n} X_i$ be the taxonomy of user model characteristics, where n is the overall number of user model characteristics;

$Y = \bigcup_{i=1}^{k} Y_i$ be the taxonomy of the DCE objects, where k is the overall number of the DCE objects;

$Z = \bigcup_{i=1}^{m} Z_i$ be the taxonomy of the user's actions in the DCE environment (reviewing, choosing, *etc.*), where m is the overall number of the user's actions;

$D = \bigcup_{i=1}^{p} D_i$ be the taxonomy of the devices used by the user to access the DCE environment (PDAs, PCs, smart phones, *etc.*), where p is the overall number of the devices.

Then:

$$\forall p, k, j, r, l, m$$
$$\forall a_{ip} \in A_p, b_{ik} \in B_k, x_{ij} \in X_j, y_{ir} \in Y_r, z_{il} \in Z_l, d_{im} \in D_m$$

$$\exists g_i, g_i \in G : a_{ip} \cup b_{ik} \cup x_{ij} \cup y_{ir} \cup z_{il} \cup d_{im} \to g_i$$

i.e. the association of elements of A, B, X, Y, Z and D the following axiom may be used:

IF ($A_1 = a_{1p}, A_2 = a_{2p}, ..., A_i = a_{ip}, ...$)

AND ($B_1 = b_{1k}, B_2 = b_{2k}, ..., B_i = b_{ik}, ...$)

AND ($X_1 = x_{1j}, X_2 = x_{2j}, ..., X_{ij}, ...$)

AND ($Y_1 = y_{1r}, Y_2 = y_{2r}, ..., Y_i = y_{ir}, ...$)

AND ($Z_1 = z_{1l}, Z_2 = z_{2l}, ..., Z_i = z_{il}, ...$)

AND ($D_1 = d_{1m}, D_2 = d_{2m}, ..., D_i = d_{im}, ...$)

THEN ($G = g_1, g_2, ..., g_n$)

Where G is a selection of sets of objects (and/or part of them) from the DCE knowledge domain ($G \supseteq Y$) that are adapted according to concrete adaptation aims and ways and are consistent with the user model, the user's actions, and access devices.

For example:

IF (A_1 =Adaptation aim: content adaptation)

AND (B_1 =Adaptation way: display of definite descriptive characteristics of the DCO, descriptive characteristics: $h_1, h_2, ..., h_n$)

AND ($X_1 =$ Knowledge level in the DCE domain: Advanced, $X_2 =$ Interest: Artwork)

AND ($Y_1 =$ Art objects: Sculptures, $Y_2 =$ School/Style: Renaissance)

AND ($Z_1 =$ User action: Reviewing DCOs)

THEN ($G =$ Selection of DCOs: Renaissance sculptures (artworks) with descriptive characteristics $h_1, h_2, ..., h_n$, presented for advanced level review).

We could also consider cases that are more complex such as: 1) a DCE engine that extracts the user’s model characteristics automatically, *i.e.* via dynamically generated user models, or 2) the user requests more than one action at once and the system maintains various adaptations in addition.

Since the approach relies on both user provided information and user activity analysis, the lack of user-defined model (i.e. not fully completed profile information or anonymous users) may cause underperformance. A possible solution would be to automatically generate the missing data, based on the user's activity. Furthermore, the automatic generation could be supported by information gathered from other users' profiles matching the available model characteristics. The DCE specification also presupposes an interoperability support on user level and transfer of the user activity data between the DCE ingredient systems (some solutions are presented in [5,6]). Thus, the DCE dynamically improves its self-management, self-configuration and self-maintenance. The DCE developers pursue the support of full semantic interconnection between DCOs as well as user’s activities and the semantic inferring needs become the core issue (a demonstration is made in [7,8]).

INTEGRATION OF SENTIMENT BASED PERSONALIZATION

As true for any novel methodology, its development and implementation in a fully functional system is a necessary requirement that leads to verification with acceptance or rejection of the hypothesis as hand. In our case we tried to show that a specifically selected biometric informational when collected in a non-intrusive fashion, such as recording of voice, can lead essentially to personalization of digital content. Moreover, as a result of a successful personal data extraction, based on mood or sentiment while collected over given time, a recommendation can be made through the use of any such personal approach. Specifically, when speech is recorded it can be preprocessed in several different stages. For example:

- As a first necessary step, it can be determined who the users are. This is necessary in order to keep different users separate from one another while keeping their portion of their biometric voice data separate from the rest. In this way we lay the foundation for further digital content personalization since from that point any information that comes from this particular user will be stored and treated specifically for them.
- A second, and very important step will be to determine how each user spoke. In other words we try to determine the intention with which the speakers talk to the system (directly or indirectly) or through the way they interact with one another so that emotion recognition algorithms can be put in place and further developed the system of non-intrusive personalized recommendation.
- Another important step can deal with spoken words that can be extracted from speech using any conventional speech recognition systems such as Google Speech API. In this way we can create an

order model of specific words through the collection of which we can further narrow the search and discovery of content of interest for the specific user, while working within the given digital ecosystem.

In our method, in order to keep simplicity a factor, as well as be right on target to what our original goals were, it was found pertinent that only the first two stages can be implemented and used to confirm the goals we had in mind. When speaking of automated services provided through speech recognition of any kind, it is generally expected that speakers talk to the system directly. Such is the case when someone is trying to generate a text by speaking to a smart system of any kind. This rule can certainly apply to the case we are trying to build. However, this is a bit more intrusive environment as it requires the user to intentionally interact with such system hence they might be skewing the sentiment in their voice. The other possible way is to speak to the system indirectly, in which case the voice is monitored automatically by the system at all times. The true sentiment and intention from a system of such kind can lead to more natural and unforced results, hence providing a stealthy way of personalizing and recommending content to the user. This can lead to not only more tailored approach on the matter, but also to providing faster service, because at the time when the system is asked to offer choices to the user, these choices are already pre-selected.

The detection and extraction of emotional states from speech signals is explained in details in [9,10,11,12]. For the task we had, it was important to determine what emotions we can consider for our case. Considering the application, it was found relevant that the following three emotional states will be a good fit for the task we had, namely: *happy*, *sad* and *neutral*. They were part of the six basic emotions, also known as the "big-six" as previously suggested by Cowie and Cornelius [13,14]. These are: *anger*, *happiness*, *sadness*, *fear*, *surprise*, and *disgust*. This choice brought not only simplicity to your job, but also it was closer to the targeted effect. To be more specific, we argue that content recommendation is usually made in cases when either [*happier*, *joyful*] emotional state is sought or when one wants to relax and is seeking [*calming*, *melancholic* or *saddening*] mood. Another reason for the choices we made was that the chosen three emotional states have a very distinct footprint when plotted, as can be seen in Figure 4. Furthermore, these sentiments are consistent with the assumption that most users are not angry or fearful when they observe artwork, so *joyful*, *melancholic*, or *neutral* moods matched that criteria pretty well.

In speech, emotions are conveyed through the voiced regions of the signal [15]. Hence, they were determined by using inverse filtering techniques in order to extract the glottal signal that is present in the voiced parts of speech alone. In Figure 3 the basic model of the glottal pulse is depicted. Each pulse has an opening and closing section. So in order to make this extraction gender and speaker independent, the ratio between closing and opening of the glottis was used, also know as Glottal Symmetry (GS).

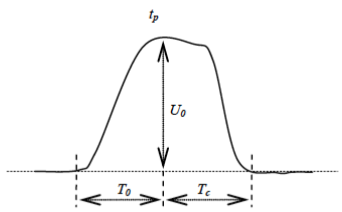

Figure 3. The shape of the basic glottal model defined by Fant 1979 [15].

Glottal Symmetry (*GS*), is represented as follows:

$$GS = \left[\frac{T_c}{T_o}\right]$$

where, *Tc* represents the closing phase of the glottis, *To* depicts the opening phase, *Uo* shows the peak volume velocity, which occurs at the time instance depicted by t_p.

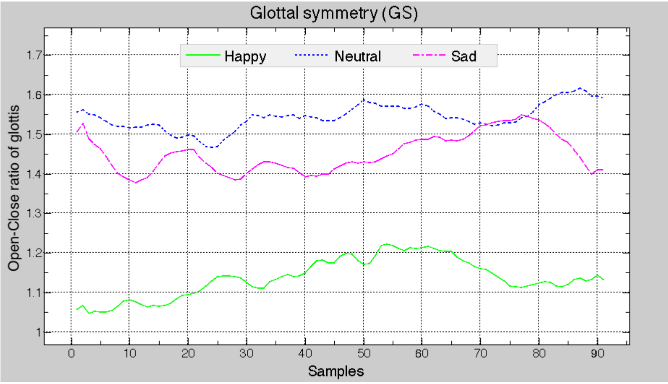

Figure 4. Glottal symmetry extracted from speech for three emotions.

CONCLUSIONS

The key contribution in our current research is the improved usage and delivery of content and knowledge from collections related to cultural and historical heritage. This was established through the implementation of a novel approach related to sentiment extraction from speech signals. Furthermore, the increasing and personalizing of user experience can help addressing some of the problems with handling large volumes of digital cultural data and objects, as well as their dynamic and personalized delivery by the system. By using our approach, several problems can be addressed and solved, such as data loss due to lack of uniform structures; lack of uniform interpretation; insufficient attractiveness of presentation; poorly adaptive and customisable presentation of objects; difficulties in context-based use; lack of personalization, etc. In implementing a clear methodology of the later, it was pointed out that through speech cues, personalization could be further developed to digital content recommendation. It was shown that sentiment could play a key role when it comes to content discovery and delivery to any specific person.

ACKNOWLEDGEMENTS

This work is partly funded by the Bulgarian National Scientific Fund under the research project №DN02/06/15.12.2016 "Concepts and Models for Innovation Ecosystems of Digital Cultural Assets" (web site: http://cultecosys.math.bas.bg), WP2 - Creating models and tools for improved use, research and delivery of digital cultural resources, WP3 - Designing a model of a multifunctional digital culture ecosystem.

REFERENCES

[1] Stanchev, P. L., Paneva-Marinova, D., and Iliev, A., 2017, “Enhanced User Experience and Behavioral Patterns for Digital Cultural Ecosystems,” Proc. 9th International Conference on Management of Digital EcoSystems, ACM, Bangkok, Thailand, pp. 288-293.

[2] Paneva-Marinova, D., Pavlov, R., and Kotuzov, N., 2017, “Approach for Analysis and Improved Usage of Digital Cultural Assets for Learning Purposes,” Int. Journal Cybernetics and Information Technologies, 17(3), pp. 140-151.

[3] Paneva-Marinova, D., Pavlova-Draganova, L., Draganov, L., and Georgiev, V., 2012, “Ontological Presentation of Analysis Method for Technology-enhanced Learning,” Proc. International Conference on Computer Systems and Technologies ComSysTech’12, ACM New York, NY, USA, pp. 384-390.

[4] Paneva-Marinova, D., 2008, “A Semantic-Oriented Architecture of a Functional Module for Personalized and Adaptive Access to the Knowledge in a Multimedia Digital Library,” Serdica Journal of Computing, 2, pp. 403-424.

[5] Paneva-Marinova D., Pavlov R., and Goynov, M., 2012, “Two Integrated Digital Libraries for Knowledge and Iconography of Orthodox Saints,” Proc. 4th International Conference, EuroMed2012, Ioannides M., Fritsch D., Leissner J., Davies R., Remondino F., Caffo R. (eds) Progress in Cultural Heritage Preservation. Lecture Notes in Computer Science, vol 7616. Springer, Berlin, Heidelberg, pp. 684-691.

[6] Paneva-Marinova, D., Goynov, M., Luchev, D., Pavlov, R., 2015, “Solution for content interoperability among digital libraries for orthodox artefacts and knowledge,” Proc. 16th International Conference on Computer Systems and Technologies, ACM New York, NY, USA, pp. 168-175.

[7] Bogdanova, G., Todorov, T., and Noev, N., 2017, “Creating and Representing Semantic Knowledge of Bell Objects,” International Journal of Applied Engineering Research, 12(19), pp. 8986-8994.

[8] Bogdanova G., Todorov T., and Noev N., 2010, “Digitalization and security of "Bulgarian Folklore Heritage" archive,” Proc. International Conference on Computer Systems and Technologies, ACM New York, NY, USA, pp. 335-340.

[9] Iliev, A.I., 2016, “Emotion Recognition in Speech using Inter-Sentence Time-Domain Statistics,” IJIRSET International Journal of Innovative Research in Science, Engineering and Technology, 5(3), pp. 3245-3254.

[10] Iliev, A.I., 2016, “Feature vectors for emotion recognition in speech,” Proc. National Informatics Conference, Sofia, Bulgaria, pp. 225-238.

[11] Iliev, A.I., and Scordilis, M.S., 2008, “Emotion Recognition in Speech using Inter-Sentence Glottal Statistics,” Proc. 15th International Conference on Systems, Signals and Image Processing, IEEE-IWSSIP 2008, Bratislava, Slovakia, pp. 465-468.

[12] Iliev, A.I., Zhang, Y., and Scordilis, M.S., 2007, “Spoken Emotion Classification Using ToBI Features and GMM,” Proc. 14th International Workshop on Signals and Image Processing 2007 and the 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. IEEE-IWSSIP 2007, Maribor, Slovenia, pp. 495-498.

[13] Cornelius, R., 1996, The Science of Emotion. Research and Tradition in the Psychology of Emotion, Upper Saddle River, NJ: Prentice-Hall, pp. 260.

[14] Cowie, R., and Cornelius, R., 2003, “Describing the Emotional States that are Expressed in Speech,” Speech Communication, 40, pp. 5-32.

[15] Iliev, A., 2012, Emotion Recognition from Speech: Using In-depth Analysis of Glottal and Prosodic Features, Various Classifiers, Speech Corpora and Noise Conditions, Lambert Academic Publishing, pp. 168.