Content-based Recommender Systems for Heritage: Developing a Personalised Museum Tour

Olga Loboda[1], Julianne Nyhan[2], Simon Mahony[3], Daniela M. Romano[4], Melissa Terras[5]

[1,2,3,4]Department of Information Studies, University College London,
Gower Street, London, UK

[5]College of Arts, Humanities and Social Sciences, University of Edinburgh,
George Square, Edinburgh, UK

[1]*olga.loboda.13@ucl.ac.uk*, [2]*j.nyhan@ucl.ac.uk*, [3]*s.mahony@ucl.ac.uk,*

[4]*d.romano@ucl.ac.uk,* [5]*m.terras@ed.ac.uk*

[1,2,3,4]*www.ucl.ac.uk,* [5]*www.ed.ac.uk*

ABSTRACT

How will a content-based recommender system be perceived by museum visitors? How will it transform visitor experience, and how can we adapt recommender systems to meet the needs of users in the museum domain? In this paper, we demonstrate the implementation of a content-based recommender system to generate personalised museum tours within the UCL Grant Museum of Zoology, London, UK. We also outline pilot usability tests that were carried out to collect initial feedback on the system performance in the wild. The findings help detect critical issues before the system is tested with museum visitors to explore the potential transformation in visitor experience that occurs with content-based recommender systems in physical museums.

1. INTRODUCTION

Museum recommender systems (RSs) have the potential to enhance visitor experience (VX) by providing a more personalised way to engage with museum collections. By focusing visitors' attention on a selection of exhibits, RSs could mitigate information overload associated with the overwhelming amounts of information that visitors have to process in a physical museum (Huang *et al.*, 2012). By tailoring recommendations to individual interests and needs, RSs can build a more personal connection between visitors and objects. By engaging visitors with a collection, RSs might also encourage exploration and stimulate learning and reflection (Kontiza *et al.*, 2018). In addition, the benefits of RSs can vary depending on user characteristics, as, for instance, they can assist visitors who are not familiar with the collection to identify their points of interest in the museum (Wang *et al.*, 2007; Fournier *et al.*, 2014; Bartolini *et al.*, 2016). However, little research has been done that provides solid evidence of enhanced VX with the help of RSs.

Over the past few decades, researchers have been testing different approaches to generate museum recommendations: recent studies include Keller and Viennet, 2015; Rossi *et al.*, 2016; Cardoso *et al.*, 2017; Hashemi and Kamps, 2018; Kontiza *et al.*, 2018; Pavlidis, 2019. The studies often tend to be limited to the offline evaluations of filtering methods that may not be able to reveal the effectiveness of RSs in a real-world setting. The algorithms may theoretically be accurate, but the RSs may not meet visitor needs because of many external or situated factors, such as a poorly designed interface and the position of points of interest in the exhibition (Hashemi and Kamps, 2018; Naudet and Deladiennée, 2019). Hence, it is necessary to carry out system evaluations in the wild. For instance, MyMuseum, a mobile guide at the National Arts Gallery of the Philippines, aimed to make gallery visits more informative and personalised by providing artwork recommendations based on personal information, art preferences, user location and item ratings (Alabastro *et al.*, 2010). By conducting online evaluations and collecting direct user experience (UX) feedback, the MyMuseum study revealed the difference in perception of the effectiveness of content-based and collaborative filtering approaches with and without contextual data. Their online evaluations indicated that the cumulative score of both accuracy and coverage was the highest for collaborative filtering without contextual information. From the user acceptance tests, the contextual recommendation approaches received more positive feedback, because the users felt more comfortable when the RS provided location-based suggestions. At the same time, it is also not enough to evaluate RSs with UX-related studies. Considering that the goal of the RSs is to enhance VX, it is necessary to take a step further and to learn about the system by analysing the RS-augmented interaction between the visitor and museum collection (Loboda *et al.*, 2018).

It is also important to acknowledge the idiosyncrasies of developing museum-specific RSs. For instance, Buder and Schwind (2012) suggested that, unlike in e-commerce, RSs in education should not strictly follow user preferences but rather challenge learners with opposite viewpoints. Moreover, the recommendations need to be assessed through an "item consumption" approach: while the evidence of RS's effectiveness in e-commerce is measured by purchasing rates, the same threshold may not be applicable to educational and, similarly, museum RSs (Buder and Schwind, 2012; Loepp and Ziegler, 2019). With a more defined scope of the requirements for museum RSs following their observed impact on VX, it is necessary to revisit the utility of complex RSs in order to make them suitable for the real-world museums. The development of museum RSs depends on data from the diverse museum databases which may not be appropriate to be used in recommenders, and this needs to be explicitly discussed to identify how to approach and transform the available datasets. Moreover, a user-item matrix with ratings for a collaborative filtering method, or behavioural data and interaction logs for implicit recommendations, or facilities for context-awareness may not be practicable to acquire and test in small-scale RS studies with limited museum and user data as well as restricted financial resources to build RSs with no commercial benefits. This may prevent researchers from replicating existing studies in other museum environments and enriching available knowledge about the efficiency of a specific RS approach across different museums.

This research aims to demonstrate the impact of museum RSs through a series of UX and VX-related field studies. We anticipate that the studies will reveal a set of requirements for the museum RSs that can enhance VX and be replicable at the same time. In this paper, we present reGrant, an RS with a traditional content-based filtering method, developed for the UCL Grant Museum of Zoology and Comparative Anatomy. We also discuss the pilot usability test conducted to collect external feedback about the developed RS and to update the system before the main study that aims to explore the impact of RSs on VX.

2. SYSTEM OVERVIEW

Initially, we carried out front-end field evaluations in the Grant Museum to identify the possible areas for improvement for their visitors and to explore how an RS could meet visitors' needs. The Grant Museum was selected because of its role as an experimental museum space for research within a university (Macdonald and Nelson, 2012) where an RS could be tested with real-world visitors. The collected feedback indicated that visitors required more information about the displayed objects and a more accessible way of browsing the collection, while some also enquired about a museum map or a tour. With these aspects included, an RS could be used to generate a personalised tour with more information about the specimens relevant to the individual interests of the visitors and to highlight objects that they may otherwise have missed.

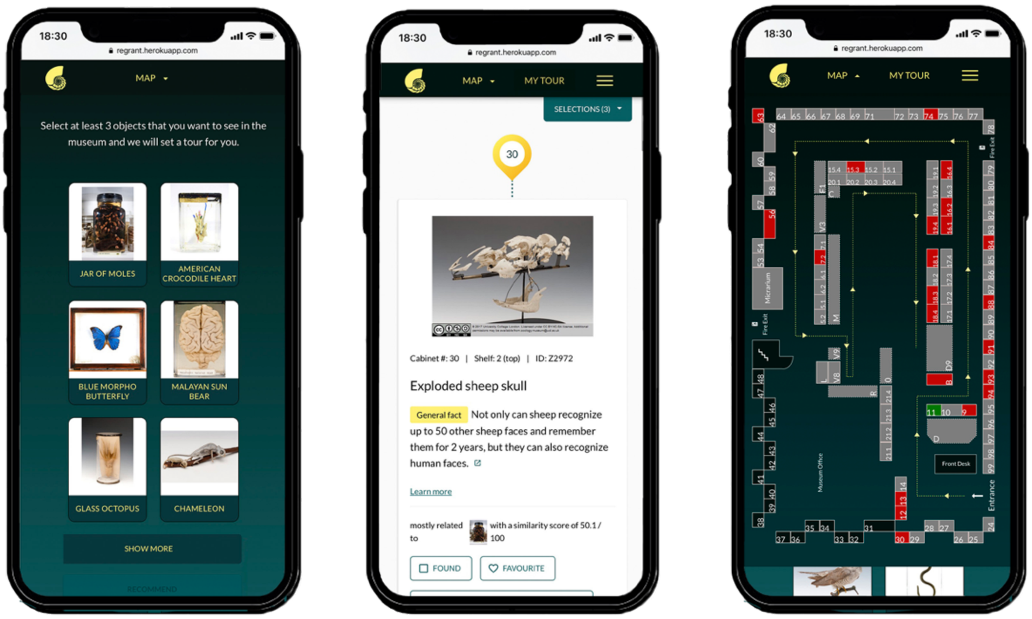

Figure 1. *Recommender system interface – home page with a user preference form (left), tour page with recommended objects (middle), map popup with recommended cabinets (right)*

Our RS is being developed using Angular and Python Flask frameworks and has been deployed to a cloud service, Heroku, at regrant.herokuapp.com. Please note that a live version of our RS may differ from the system overview outlined in this paper due to periodic system updates that reflect our research findings and ongoing experimental studies. The first version, which was developed during April-May, 2019 (see Figure 1), was based on the following scenario: when the user launches the application, they first have to select at least three specimens of interest from a list of 100 featured museum objects. Based on user selections, the system generates a tour with 50 objects. On a tour page, the user can explore some interesting facts about recommended specimens, and they can learn more about the specimens by visiting the corresponding object pages. They can also use the map, where recommended cabinets are highlighted, to navigate in the museum. By tapping on a cabinet icon on the map, the user can see a list of objects showcased in the selected cabinet, and manually add preferred specimens to their tour. As their tour proceeds, the user can rate recommendations as well as mark specimens as favourite and/or found. At the end of the visit, the system provides a visit summary with some information about the user's tour, their preferences and 21 new recommendations for the next visit based on user input gathered during the tour.

3. RECOMMENDER ALGORITHM

In our RS, we employ a traditional content-based filtering method that relies on the similarities between items and does not require any substantial prior user data to generate recommendations. Our filtering method explores the similarities between ten object features, such as species, distribution, conservation status and object type. Figure 2 briefly illustrates the executed recommendation process. More specifically, the data pre-processing step involves data cleaning to fill in missing values and to convert values into lowercase strings, data aggregation to gather all cleaned strings in one column, tokenisation to break up strings into lists of words (tokens) and vectorisation to convert lists into vectors. For our dataset with 500 museum objects, we use CountVectorizer which builds a vocabulary of all tokens and counts how many times a token occurred in the description of each specimen. The returned term-document matrix is then used to calculate the similarity scores between all objects in the dataset using the cosine similarity metric. The metric evaluates the similarity between two vectors and is represented as a sum of multiplied components, A_i and B_i, from vectors A and B divided by the product of the two vectors' lengths:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2}\sqrt{\sum\limits_{i=1}^{n} B_i^2}} \quad (1)$$

where A_i is a component from the vector A; B_i is a component from the vector B; n is a number of all distinct tokens.

When the system receives a recommendation request, it constructs a subset of relevant objects based on the scores from the similarity matrix to then either display recommendations on individual object pages, list them on a tour page in a particular order, or provide new recommendations for the next visit on a visit summary page.

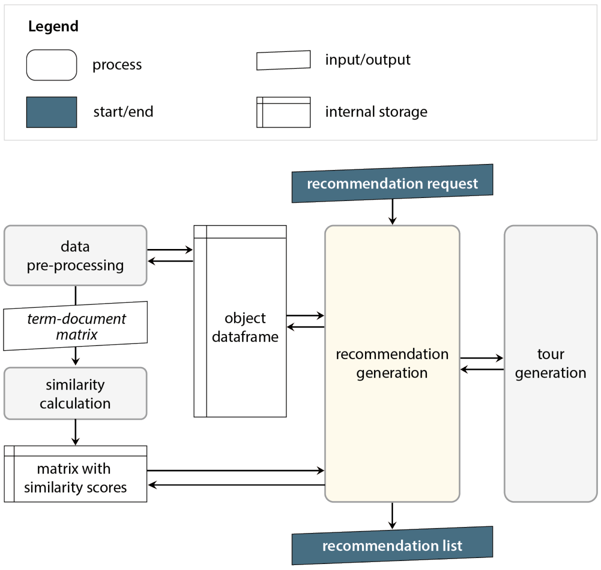

Figure 2. *Recommendation component diagram*

The Alan Turing Institute

4. PILOT USABILITY TESTING

4.1 Usability study procedure

Pilot usability testing was conducted at the Grant Museum and involved both quantitative and qualitative evaluations to collect some external feedback and identify critical issues with the system performance before the RS was used by the Grant Museum visitors in a large-scale study. Over the period of two weeks, 12 current and former UCL staff and students were recruited to take part in our usability testing (9 students and 3 academics; 5 male and 7 female respondents aged between 22 and 44; 10 solo visitors and 1 group visit; 8 first-time and 4 returning visitors). Due to space constraints, in this paper we provide a brief summary of the study, but it will also be addressed in the following publications and the lead author's PhD thesis.

At the museum the participants were debriefed about the goal of the usability testing and the reason why the RS was developed. However, we did not demonstrate how the system worked in order not to affect the perception of the system's ease of use, but rather to imitate a real-world scenario in which the visitors would not be able to participate in a demonstration session. The participants were then asked to explore the museum with the RS for around 20 minutes and then to return back for the post-test evaluations. The questionnaires were used for an overall evaluation of perceived system performance and encompassed 26 items (5-point Likert scales where 1 = disagree, 5 = agree) that were mostly adapted from the Pu *et al.*'s (2011) ResQue survey for measuring the RS qualities. The quantitative study with questionnaires was supported with semi-structured interviews to gather more in-depth feedback about encountered issues and potential system improvements.

Before reporting the findings, it should be noted that two participants had to be excluded from the quantitative study with questionnaires. One participant did not accurately follow the usability protocol to be able to answer all survey questions, but they reflected on their RS-augmented experience during the interview. The other participant encountered an error on their device which prevented them from testing the RS properly, but since they were with a companion, they followed their companion's tour and thus provided some feedback in the semi-structured interview.

4.2 Usability study results

Overall, the participants were satisfied with the recommender ($M = 4.4$, $SD = 0.516$, $n = 10$) and found the system easy to use and easy to learn ("The recommender was easy to use": $M = 4.5$, $SD = 0.707$, $n = 10$; "I became familiar with how to use the recommender very quickly": $M = 4.3$, $SD = 0.675$, $n = 10$). During the interviews, 8 out of 12 participants mentioned potential educational benefits of the system. The respondents reported that the RS made their visits more structured, informative and helped them find interesting objects that they would otherwise have missed. One participant particularly liked the visit summary because it helped them learn not only about the collection but also about their museum preferences. Moreover, the users liked the idea of integrating recommenders into the museum environment ("I would use this recommender again": $M = 4.4$, $SD = 0.516$, $n = 10$; "I would like to use similar systems in other museums": $M = 4.6$, $SD = 0.699$, $n = 10$).

From the quantitative findings, the recommendations received positive feedback ("The recommender gave me good suggestions": $M = 4.1$, $SD = 0.316$, $n = 10$), while the perceived accuracy score was adequate ("The recommended objects matched my interests": $M = 4.0$, $SD = 0.471$, $n = 10$). The participants, however, were not satisfied with the recommendation diversity ("The recommended objects were diverse": $M = 3.5$, $SD = 0.527$, $n = 10$). During the interviews, 3 system users suggested adding random objects to their tours to see how they would respond to less relevant content, while 2 participants were particularly curious to know what their least relevant specimens were. In this regard, the participants were also curious to find out what criteria were used to generate the recommendations as the first RS version did not provide explicit explanations, apart from the similarity scores related to the selected specimens.

It is also important to mention that the participants struggled to locate recommended objects in the museum ("It was easy to find the recommended objects in the museum": $M = 3.5$, $SD = 0.845$, $n = 10$). 5 respondents indicated that the cabinet numbers in the physical museum were confusing because they don't follow a sequential order and they were difficult to see because of their location and colour contrast. Moreover, since the RS was developed for a zoological museum where many objects can be found in the same cabinet and on the same shelf, it was challenging to locate individual objects in a cabinet. As a result, several interviewees reported that their attention decreased as their tour proceeded because 50 suggested specimens to find in one visit was overwhelming, and they skipped some objects if too many were recommended for the same museum cabinet.

4.3 Usability study discussion and limitations

Because the study was limited to the university students and staff, the sample is not necessarily representative of the Grant Museum's wider and more heterogeneous audience to the extent that would allow the findings to be generalised. In addition, the system was tested with only 500 objects and thus more unfavourable diversity results might have been observed by making use of content-based filtering method with a larger object dataset. Nevertheless, the pilot usability evaluation helped identify the areas for improvement before the main study with the Grant Museum visitors. The participants' enquiry about the least relevant and random objects revealed a substantial divergence from recommending only those objects that may match visitor's preferences. This suggests that diversity in museum RSs might be favoured over accuracy. Moreover, the feedback extended beyond the recommendation quality and indicated that an RS might also become a trigger of information overload and museum fatigue, unless the retrieval process of the recommended items in the museum is facilitated with custom tour lengths and/or some adjustments in the physical space.

Following the user feedback, we released the second version of our recommender where, for example, we updated the system to extract n_1 most similar objects as well as randomly pick n_2 least similar objects for the tour, where n_1 and n_2 are set dynamically in a ratio of 4:1 depending on a preferred tour length. More details about all system updates will be addressed in the next publications and will be comprehensively discussed in the PhD thesis.

5. CONCLUSION

In this paper, we demonstrate how we implemented the first version of our content-based RS. The suggested system is anticipated to be replicable in order to encourage other researchers to conduct similar evaluations in other museums, to aggregate and compare the findings and to elicit robust requirements for museum RSs. We also emphasise the importance of field studies to collect real-world feedback about system performance and its effectiveness, and thus we begin to analyse the UX with an early version of our RS to learn how it should be improved before the main study where we explore the transformations of VX with the content-based RS in the physical museum.

Acknowledgements: We would like to thank the UCL Grant Museum of Zoology and Comparative Anatomy for providing information about their collection as well as allowing us to carry out our evaluations in the real-world museum environment. We would also like to thank all current and former UCL staff and students who participated in our usability study for their invaluable feedback.

6. REFERENCES

Alabastro, P., Ang, M., DeGuzman, R., Muhi, M. and Suarez, M. (2010). 'MyMuseum: Integrating personalized recommendation and multimedia for enriched human-system interaction', in *DC '10: Proceedings of the 6th International Conference on Digital Content, Multimedia Technology and Its Applications*. IEEE, pp. 421–426.

Bartolini, I., Moscato, V., Pensa, R. G., Penta, A., Picariello, A., Sansone, C. and Sapino, M. L. (2016). 'Recommending multimedia visiting paths in cultural heritage applications', *Multimedia Tools and Applications*, Vol. 75, No. 7, pp. 3813–3842.

Buder, J. and Schwind, C. (2012). 'Learning with personalized recommender systems: A psychological view', *Computers in Human Behavior*, Vol. 28, No. 1, pp. 207–216.

Cardoso, P. J. S., Rodrigues, J. M. F., Pereira, J. A. R. and Sardo, J. D. P. (2017). 'An Object Visit Recommender Supported in Multiple Visitors and Museums', *International Conference on Universal Access in Human-Computer Interaction*. Edited by M. Antona and C. Stephanidis. Springer International Publishing, pp. 301–312.

Fournier, R., Viennet, E., Sean, S., Soulié Fogelman, F. and Bénaïche, M. (2014). 'AMMICO: social recommendation for museums', in *Proceedings of DI '14: Digital Intelligence – Intelligences Numériques*. Nantes, France.

Hashemi, S. H. and Kamps, J. (2018). 'Exploiting behavioral user models for point of interest recommendation in smart museums', *New Review of Hypermedia and Multimedia*. Taylor & Francis, Vol. 24, No. 3, pp. 228–261.

Huang, Y. M., Liu, C. H., Lee, C. Y. and Huang, Y. M. (2012). 'Designing a personalized guide recommendation system to mitigate information overload in museum learning', *Educational Technology and Society*, Vol. 15, No. 4, pp. 150–166.

Keller, I. and Viennet, E. (2015). 'Recommender Systems for Museums: Evaluation on a Real Dataset', in *IMMM '15: The Fifth International Conference on Advances in Information Mining and Management Recommender*, pp. 65–71.

Kontiza, K., Loboda, O., Deladiennée, L., Castagnos, S. and Naudet, Y. (2018). 'A museum app to trigger users' reflection', in *Mobile CH '18: Proceedings of the 2nd Workshop on Mobile Access to Cultural Heritage*. CEUR Workshop Proceedings.

Loboda, O., Nyhan, J., Mahony, S. and Romano, D. M. (2018). 'Towards evaluating the impact of recommender systems on visitor experience in physical museums', in *Mobile CH '18: Proceedings of the 2nd Workshop on Mobile Access to Cultural Heritage*. CEUR Workshop Proceedings.

Loepp, B. and Ziegler, J. (2019). 'Measuring the Impact of Recommender Systems – A Position Paper on Item Consumption in User Studies', in *ImpactRS '19: Proceedings of the 1st Workshop on Impact of Recommender Systems*.

Macdonald, S. and Nelson, T. (2012). 'A Space for Innovation and Experimentation: University Museums as Test Beds for New Digital Technologies', *A Handbook for Academic Museums: Beyond Exhibitions and Education.*, pp. 418–444.

Naudet, Y. and Deladiennée, L. (2019). 'Towards technology-mediated CH experiences: some open challenges', in *CI '19: Proceedings of the Workshop on Cultural Informatics*. CEUR Workshop Proceedings.

Pavlidis, G. (2018). 'Towards a Novel User Satisfaction Modelling for Museum Visit Recommender Systems', in *VR Technologies in Cultural Heritage*. Springer International Publishing, pp. 60–75.

Pu, P., Chen, L. and Hu, R. (2011). 'A User-Centric Evaluation Framework for Recommender Systems', in *RecSys '11: Proceedings of the Fifth ACM Conference on Recommender Systems*, pp. 157–164.

Rossi, S., Barile, F., Improta, D. and Russo, L. (2016). 'Towards a Collaborative Filtering Framework for Recommendation in Museums: From Preference Elicitation to Group's Visits', *Procedia Computer Science*. Elsevier Masson SAS, Vol. 98, pp. 431–436.

Wang, Y., Aroyo, L. M., Stash, N. and Rutledge, L. (2007). 'Interactive user modeling for personalized access to museum collections: The Rijksmuseum case study', in Conati, C., McCoy, K., and Paliouras, G. (eds) *International Conference on User Modeling*. Springer Berlin Heidelberg, pp. 385–389.