

Report on the CHIIR 2021 Third Workshop on Evaluation of Personalisation in Information Retrieval (WEPIR 2021)

Gareth J. F. Jones
Dublin City University, Ireland
gareth.jones@dcu.ie

Nicholas J. Belkin
Rutgers University, USA
belkin@rutgers.edu

Noriko Kando
National Institute of Informatics,
Japan
kando@nii.ac.jp

Gabriella Pasi
University of Milano-Bicocca,
Italy
gabriella.pasi@unimib.it

Abstract

The Third Workshop on Evaluation of Personalisation in Information Retrieval (WEPIR 2021) was held in conjunction with the ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR 2021) in Canberra, Australia, as a virtual event. WEPIR 2021 followed on from the first and second WEPIRs held at CHIIR 2018 and 2019. The purpose of the workshop was again to bring together researchers from different backgrounds, interested in advancing the evaluation of personalisation in information retrieval. The workshop focused on further development of a common understanding of the challenges, requirements and practical limitations of personalisation in information retrieval and its evaluation.

1 Introduction

The main aim of an Information retrieval (IR) system is to propose useful items to a user, based on a formal representation of her/his information needs. It has been long advocated that IR is an interactive activity, where the user plays a central role, given that the user's information need is personal, and grounded on the user's context and expertise.

In last two decades, the issue of detecting and exploiting user context and preferences in search has raised increasing attention both in the research community and in companies; as a matter of fact, commercial search engines profile their users by observing various kinds of interactions with the systems, and also based on the user's search history.

Besides the above mentioned implicit signals of user's preferences, there is nowadays a considerable amount of information that can be fruitfully exploited to gather evidence of the user context, such as user generated content in social media.

The ultimate aim of user modelling in a search context is to provide a search outcome that is most likely compatible with the personal expectations of the user. To personalise the outcome of a search process, the information formally represented in a user model must be injected at some stage in the process. This can be done in various ways, including personal query expansion and results re-ranking, to create a process of personalised information retrieval (PIR).

In order to determine how effectiveness of an IR application is, an evaluation strategy is required. Evaluation of personalised search is not simple, as it requires involving real users in the process.

The first WEPIR was held in conjunction with CHIIR 2018 [Jones et al., 2018a], while the second meeting WEPIR 2019 was organised at CHIIR 2019 in Glasgow, U.K. [Jones et al., 2018b]. This paper reports on the third WEPIR¹, held in conjunction with CHIIR 2021.

2 Background

As outlined in the Introduction, the first WEPIR Workshop was held in 2018, with the main motivation to analyse and contribute to the issue of the evaluation of the personalised search task. This issue raised the interest of the research community well before this first workshop, and it is still ongoing; however, despite various attempts, there is still the need to find a suitable and reliable evaluation protocol for evaluation of personalisation.

Among the most relevant activities aimed at exploring this issue by recognising the central role of the user is the Interactive Track at the TREC conferences, which ran for twelve years [Dumais and Belkin, 2005]. This initiative is of relevance to this workshop for several reasons. One is that it developed methods for evaluating various aspects of system performance over entire search sessions, which we deem a crucial aspect in the evaluation of personalisation. Another important reason is that one of the main findings of this track was the difficulty of applying the general TREC/Cranfield evaluation model to the dynamic situation of interactive information retrieval, in which the user plays a key role, which is a key aspect of the personalised search task.

More recently, from 2010 to 2014, the TREC Session Track has aimed to provide both test collections and evaluation measures for analysing the search process over user sessions with multiple stages of query reformulation rather than one-time queries. While introducing some modified evaluation metrics for session based search [Kanoulas et al., 2011; Carterette et al., 2015], this track had the limitation that the information need was assumed to remain static for a query across the session.

Scientific initiatives carried out in the last decade, such as the 2013 Dagstuhl Seminar on Evaluation Methodologies in Information [Agosti et al., 2014], and the 2012 NII-Shonan Seminar on Whole-Session Evaluation of Interactive Information Retrieval Systems [Belkin et al., 2016], addressed some evaluation issues that are relevant to the WEPIR workshops, by encompassing the definition of evaluation measures for entire search sessions, and the inclusion of user modeling in the evaluation. However they stopped short of the problem of evaluation of personalisation of information retrieval.

Introduced at CLEF 2017, the Personalised Information Retrieval (PIR-CLEF) task sought to develop a framework for the repeatable evaluation of algorithms for personalized search, and for

¹<https://wepir2021.disco.unimib.it/>

the evaluation of user models [Pasi et al., 2017, 2018, 2019]. PIR-CLEF 2017, PIR-CLEF 2018 and PIR-CLEF 2019 focused on a benchmark web search task that provided user data gathered during a single search session. These data relate to various activities undertaken during their search session by each participant, including details of relevant documents as marked by the searchers [Sanvitto et al., 2016]. PIR-CLEF 2019 expanded the scope of this investigation by the inclusion of a task exploring the evaluation of personalisation in a medical search task. Unlike the information retrieval research community, the User Modeling research community has traditionally not had a significant focus on comparative evaluation or shared evaluation tasks. However, this situation is changing with the emergence of the Eval-UMAP workshop series exploring the evaluation of user modeling, adaptation and personalization, which began at the UMAP 2016 conference [Conlan et al., 2016], and is currently being held on an annual basis.

The WEPIR 2018 and WEPIR 2019 workshops featured invited keynotes, a small number of short paper presentations, and extended sessions of breakout group discussions. At WEPIR 2018 these breakout groups focused on the topics of measurement, understanding and context [Jones et al., 2018a], and at WEPIR 2019 on the topics of Personalisation for Individuals, Personalisation for Groups, Presentation of Personalisation to Users and Evaluating the Presentation protocol for this search task [Jones et al., 2018b]. In this report we summarise the breakout discussions from WEPIR 2021 which focused on the topics of Personalisation of Museum Visitor’s Experience and Personalised Medical Search.

3 Keynotes

WEPIR 2021 included two invited Keynote Presentations: by Professor Hideo Joho of the University of Tsukuba, Japan; and, by Professor Evangelos Kanoulas of the University of Amsterdam, The Netherlands.

Professor Hideo Joho: *Making a Difference in User Search Experience*

In this presentation Professor Joho focused on the potential for the exploitation of personal and public information sources within emerging artificial intelligence (AI) driven applications, and more generally the merging of physical and cyber information spaces. Investigation of this topic involves contributions from a number of areas of examination and technologies, Hideo’s talk focused on consideration of six areas for examination, in particular considering the role of AI and personalisation. **Complex (human) problem solving** which is concerned with how humans use information to address complex problems in their daily lives. Information retrieval is a relatively mature technology which is widely used by individuals as a tool to assist them in addressing complex tasks, but there is scope to consider how AI might be used to extend the role of technology in easing the completion of such tasks. **Knowledge-Information-Data (KID) Space** which is concerned with the established relationships between data, information and knowledge sources, but also incorporates the use of these sources with the more complex human concept of wisdom. Hideo illustrated this with examples of the combination of personal and public information spaces within applications, where a particularly interesting examples of a personal information space is lifelog archives. **Advanced Human-Agent Interaction** where there is much scope for the extension of agent technologies to include elements of personalisation to enable them to provide

richer and more effective engagement with human users. **Enabling Technologies**, relating to pure and applied AI technologies, including machine learning, natural language processing, computer vision, multimedia, human computer interaction, robotics, and dialogue systems. The last of these is particularly important for the development of advanced tools for engagement with information, such as conversational search. **Smart Devices** are emerging with many uses in information capture, processing and presentation. These are little explored in current work on personalised information processing. Hideo suggested that wider exploration of their use within information spaces could be a fruitful area for future research. **FAccT / FACTS** is concerned with other ethical and societal dimensions of technologies in human-centred society. Topics of importance here include: fairness (and bias), accountability, confidentiality, transparency and security. For more details, the reader is referred to Professor Joho's presentation, available on the WEPIR 2021 website.

Professor Evangelos Kanoulas: *From Sessions to Conversations and Back Again*

In his presentation Professor Kanoulas related the issue of evaluation of personalisation to consideration of support for information seeking as a conversation between an agent and an information seeker, within the context of a search session. He began by pointing out the increasingly positive effect of personalisation based on the current search session, as the session continues, vis-à-vis the static effect of personalisation based on the information seeker's previous history. He then specified, described, and enumerated research efforts toward, and evaluation problems of four necessary functions (expectations) of a conversational search agent: 1) Profile the user; 2) Account for session history; 3) Ask clarifying questions; and 4) Decide next action. For each function, Evangelos provided a detailed explanation of what that function entailed, an extensive overview of research, much of it conducted by his research groups, into accomplishing and evaluating success of that function, and a description of the challenges yet remaining to their accomplishment and evaluation. He concluded his presentation by suggesting that evaluation of session-based personalisation of support for information seeking could be accomplished by developing test collections which simulate mixed-initiative interaction between information seeker and search agent, instantiating these four functions. Within such a framework, costs and gains of means of invoking the functions should be the basis for determining measures to evaluate their success, both individually, and with respect to the support of the search session as a whole. This was an inspiring presentation, setting the scene for both subsequent presentations, and for the discussions which took place in the two working groups. For details, the reader is referred to Professor Kanoulas's presentation, available on the WEPIR 2021 website.

4 Contributed Papers

In this section we provide brief overviews of the three contributed papers accepted for publication at the workshop.

In the position paper entitled *Information Retrieval Evaluation in Knowledge Acquisition Tasks*, Yasin Ghafourian, Petr Knoth and Allan Hanbury reflected on the limitations of the Cranfield Paradigm in the context of knowledge acquisition (or learning) tasks, and proposed several alternatives based on the notion of evaluating a session [Ghafourian et al., 2021]. The authors pro-

posed that the interaction of a user with IR systems in knowledge acquisition tasks is personalised with knowledge and determined by the interest of a user and their prior knowledge.

The notion of novel interactive museum user guides that are able to support a user's personalized discovery and learning was addressed in the paper entitled *Quiz Generation on the Electronic Guide Application for Improving Learning Experience in the Museum*, by Masaki Ueta, Tomoya Hashiguchi, Hun-Long Pham, Yoshiyuki Shoji, Noriko Kando, Yusuke Yamamoto, Takehiro Yamamoto and Hiroaki Ohshima. In their series of investigations [Ueta et al., 2021], the personalized approach proposed integrates the user's active exploratory search-and-browsing in both the virtual and the physical museum space to recommend artefacts. On top of this, the paper proposed a method to automatically generate quizzes on each of the artefacts in front of each visitor and examined the effects to enhance the museum experience for each visitor.

The paper entitled *Location-based Reminder for Memorizing What Visitors Learn at a Museum* by Yoshiyuki Shoji, Kenro Aihara, Martin J. Dürst, Noriko Kando, Takuya Nakaya, Hiroaki Ohshima, Takehiro Yamamoto and Yusuke Yamamoto was inspired by the Context Model of Falk and Dierking [2018] and addressed a challenging issue to make the museum experience more unforgettable to each visitor, and deeply connected and integrated into the personal context along time through the social life of each visitor after their museum visits [Shoji et al., 2021]. The proposed approach sends reminders to the visitor's iPhone when the visitor passes nearby something related to the favorite artefacts of each visitor.

The papers can be obtained from CEUR Workshop Proceedings² and the presentation slides from the WEPIR 2021 workshop website.

5 Breakout Working Groups

Following on from the high level of engagement by workshop participants at the earlier WEPIR 2018 and WEPIR 2019 meetings, one of the key objectives of WEPIR 2021 was again to encourage attendees to engage in discussion of topics relating to the evaluation of PIR. The distribution of the participants online across multiple international timezones made this more challenging at WEPIR 2021 than at the previous meetings, nevertheless participants engaged actively in discussions at the end of the keynotes and the paper presentations, and in particular in the breakout working groups.

Discussions at the previous WEPIR meetings focused on examination of the concepts relating to PIR and its evaluation. For WEPIR 2021, the breakout working groups focused on two specific use cases for which personalisation has the potential to improve the effectiveness of search for individual users. The goal of the groups was to explore the search requirements of the use case, look at the scope for personalisation in more effectively meeting these requirements, and to move towards a strategy for evaluation of the utility of personalisation in these settings.

The selected use case areas for examination were personalisation of a museum visitor's experience and personalisation in medical search.

²<http://ceur-ws.org/Vol-2863/>

5.1 Group 1: Personalisation of Museum Visitor's Experience

One important domain where personalisation may have a significant impact on user experience is that of the museum visitor. The interests and knowledge with vary greatly between visitors, in association with the context context of a visit and museum, meaning that personalisation of the information provided to them during their visit may result in a more informative, enjoyable and memorable visit. For example, recommending an artefact relevant to the visitor's interests or generating a personalised explanation of that artefact could make the visitor's experience more attractive and memorable. This group had an exploratory discussion about what and how personalisation of museum experience should be done.

5.1.1 The Goals of Personalisation of the Museum Experience

The group considered the variety of goals of personalisation of the museum experience. These included the following:

- To help the visitor discover artefacts of interest.
- To help the visitor better learn about artefacts.
- To encourage the visitor to re-visit the museum again.
- To make the visitor's experience more memorable.

5.1.2 How Museum Personalisation Could be Accomplished

The group identified the following factors according to which personalisation of the museum experience could be accomplished:

- Recommending artefacts/routes/themes based on the visitor's knowledge and interest.
- Generating personal description of the artefact.
- Generating personalized souvenirs.
- Recalling what the visitor experienced after leaving the museum.

5.1.3 What Data Could be Used for Museum Personalisation

The group considered what data are necessary and could be used for accomplishing personalisation:

- The visitor's current behavior in the museum.
- The visitor's past behaviors in the museum.
- The visitor's personal data stored in SNS.
- The visitor's personal data stored in their mobile device.

5.1.4 Evaluation of Personalisation of the Museum Experience

The group identified the following factors which should be measured for determining success of personalisation of the museum experience:

- Knowledge gain about the artefact(s).
- Quality of the experience of the visit.
- Long-term effect of how the experience affects the visitor's life.

At the closing session, several other issues were identified from the discussions among the workshop participants. Personalisation of the visitor's experience at a museum based on their past behaviors in other museums would be an interesting topic. The privacy issue of the data to be used for personalisation matters. How and who will manage the personal data would be an important topic in personalisation in a museum.

5.2 Group 2: Personalised Medical Search

This group focused on the evaluation of PIR systems that are designed to support medical related search, while exploring medical knowledge among medical records and documents. The outcomes of this group focused on the following topics: medical related search scenarios, relevant personalisation factors, evaluation considerations and the challenges of personalisation in medical search.

5.2.1 Use cases for Medical Search

We approached the personalisation evaluation task while considering the different actors and use cases for medical search. Medical professionals such as nurses and doctors may use search engines looking for materials for patient education and program implementation, e.g., collecting facts to support a clinical intervention. They also use search engines for health care decision making for diagnosis, testing and treatments for clinical cases such as precision medicine. Health data consumers are also engaged in self-diagnosis and disease and treatment search before and after consulting a healthcare professional. Personalisation of an IR system should therefore be tailored to the needs of the user that will most frequently use it and the use case that it is going to be used for.

5.2.2 Personalisation Factors

The question of whether personalised IR systems can be beneficial to medical knowledge searchers, as opposed to conventional IR systems, can be answered by considering the level of personalisation implemented in a personalised IR system and the factors that are used to implement personalisation.

- **User understandability:** The first factor that can be taken into account in a personalised IR system while answering a user's query is the user's domain knowledge of the field that

he/she is searching in. Similar to adjusting a presentation to make it suitable for a specific audience, a system that has access to knowledge of its user in terms of the depth of familiarity with the field, can provide more relevant documents concerning varying levels of detail with regard to the inquiry to the user. Another factor of user understandability to take into account which can also be a factor affecting the user's satisfaction of working with the IR system is his/her knowledge of the language in which the search is done. The idea is that if a system is aware that the user is trying to find a novel piece of information in a language with less familiarity, it will put first the documents that are highly relevant to that piece of information the user is looking for but are also in a language easier for the user to understand (the user's first language or another language in which the user has shown more proficiency by some means throughout time).

- **The current health state of the user:** The current health state of the user not only lends weight to the criticality of access to credible information presented to the user, but also differentiates between the information needs when the input query is the same for different users. A case in point could be where users in different stages of a treatment are seeking information about a drug which is used in a specific stage of treatment. One user could be looking for the side effects and another user might want to find about drug Interaction of this specific drug with other drugs. Another example is from a use case which is also a popular trend in current social media is when a user is searching for a specific COVID-19 vaccine. The results presented to the users could be diversified based on varying granularity of their current health status from general health information such as their age, underlying diseases, whether or not they have already received a dose of vaccine, etc., to finer details such as whether the vaccine they have already had matches the one they are seeking information about, details of their underlying condition etc.
- **User context:** Other user contexts such as location of the user, task type, task urgency could also have determining impacts on the results that users expect from a search engine. As an example, concerning the location of a user, considering details like whether the user is in a private address or at a hospital or considering more general location info contexts like the city or the country of the user, could tell us about the health status of that location and access to medical support and facilities.

5.2.3 Personalisation Evaluation

The degree to which incorporating personalisation factors improve user experience while searching the medical information space can be measured using user studies designed to measure the following user-related aspects:

- **Relevance dimensions** such as user **understandability** of the presented medical information, which is particularly critical in the medical domain given its complexity with higher chances of misunderstanding. **Credibility** is also a relevant relevance measure that should be maintained. While attempting to match user cognitive abilities with the retrieved medical information, it is essential to ensure directing users to high-quality credible health resources.

-
- **User engagement** such as the time spent searching, clickthrough rate and the volume of issued queries. Such engagement measures should be carefully interpreted as they may lead to different conclusions depending on the search task, e.g., exploratory search tasks tend to encourage more user engagement while other time-critical tasks need to be completed with minimum interaction.
 - **User satisfaction** is also important to capture user overall experience while using the search engine. That can be achieved explicitly using user feedback/surveys or implicitly using some user interaction measures.

5.2.4 Challenges in Medical Search Personalisation

We also identified a number of challenges that need to be addressed when working on personalised medical search.

- When to personalise? Sometimes it might not be suitable to personalise; for example, a nurse searching patient's records for relevant data must always retrieve the same information about the patient's health state.
- How to measure the personalisation quality? In many scenarios, it is also tough to measure the quality of the decisions made by the doctors.
- How to personalise the search to preserve explainability? Explainability should allow the users to notice how much the search was personalised (e.g. if the information was kept back because the system made assumptions on a user profile).
- How to convince health practitioners to use electronic search more often? Many doctors and nurses still prefer to use colleagues or consult a print source as their first source of information.
- How can the user control the level of personalisation? If a user could adapt the way the system does the personalisation, it could improve the quality of the search and increase the trust in the system.

5.3 Breakout Review

The discussions of the breakout groups moved examination of the issues and evaluation of PIR for these user cases forward, with useful exploration of the issues which would need to be addressed to develop an evaluation strategy in these settings. There is though still further work required to operationalize the ideas developed here for actual evaluation plans. Looking to the future there remains much further work required to develop a general framework for evaluation of PIR which can be applied in different search settings.

6 Concluding Remarks

Organising an online workshop with widely geographically distributed participants made it challenging to develop a programme accessible across different territories, particularly for the breakout sessions. The workshop timetable was spread across an extended time period for the days of the workshop. Given the challenging setting of the workshop, we were very pleased with the level of participation throughout the workshop, particularly with how the breakout groups organised their sessions and the energy that they brought to their discussions. While WEPIR 2021 made progress in the development of evaluation in PIR, it is clear that much work remains to be done to develop a better understanding of the many issues raised, and there is certainly scope for further meetings examining evaluation of PIR.

Acknowledgements

This workshop was partially supported by Science Foundation Ireland as part of the ADAPT Centre³ (Grant No. 13/RC/2106), and by the US National Science Foundation under grant IIS-1423239. The workshop chairs greatly appreciate the enthusiasm with which the participants approached this workshop and their contributions in particular to the reporting of the outcomes of the breakout groups.

References

- Maristella Agosti, Norbert Fuhr, Elaine Toms, and Perti Vakkari. Evaluation methodologies in information retrieval. Dagstuhl Seminar 13441, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Schloss Dagstuhl, Germany, 2014.
- Nicholas J. Belkin, Susan Dumais, Noriko Kando, and Mark Sanderson. Whole-session evaluation of interactive information retrieval systems. NII Shonan Meeting Report 2012-7, National Institute of Informatics, Japan, Tokyo, Japan, 2016.
- Ben Carterette, Evangelos Kanouros, Mark Hall, and Peter Clough. Overview of the trec 2014 session track. In Ellen M. Voorheen, editor, *Proceedings of the 23rd Text REtrieval Conference, TREC 2014*, Gaithersburg, MD, 2015. National Institute of Standards and Technology. URL <http://trec.nist.gov/pubs/trec23/papers/overview-session.pdf>.
- Owen Conlan, Liadh Kelly, Kevin Koidl, Seamus Lawless, Killian Levacher, and Athanasios Staikopoulos, editors. *EvalUMAP2016: Towards Comparative Evaluation in the User Modelling, Adaptation and Personalization*, Halifax, Canada, 2016.
- Susan Dumais and Nicholas J. Belkin. The TREC interactive tracks: Putting the user into search. In Ellen M. Voorhees and Donna K. Harman, editors, *TREC. Experiment and evaluation in information retrieval*, pages 123 – 152. MIT Press, Cambridge, MA, 2005.

³www.adaptcentre.ie

John H Falk and Lynn D Dierking. *Learning from museums*. Rowman & Littlefield, 2018.

Yasin Ghafourian, Petr Knoth, and Allan Hanbury. Information retrieval evaluation in knowledge acquisition tasks. In *Proceedings of the Third Workshop on Evaluation of Personalisation in Information Retrieval (WEPIR 2021)*, pages 88 – 95, Canberra, Australia, Virtual event, 2021.

Gareth J. F. Jones, Nicholas J. Belkin, Seamus Lawless, and Gabriella Pasi. Report on the CHIIR 2018 workshop on evaluation of personalisation in information retrieval (WEPIR 2018). *SIGIR Forum*, 52(1):129–134, 2018a.

Gareth J. F. Jones, Nicholas J. Belkin, Seamus Lawless, and Gabriella Pasi. Report on the CHIIR 2018 workshop on evaluation of personalisation in information retrieval (WEPIR 2019). *SIGIR Forum*, 53(1):29–37, 2018b.

Evangelos Kanoulas, Ben Carterette, Paul D. Clough, and Mark Sanderson. Evaluating multi-query sessions. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR 2011, pages 1053–1062, Beijing, China, 2011. ACM.

Gabriella Pasi, Gareth J. F. Jones, Stefania Marrara, Camilla Sanvitto, Debasis Ganguly, and Prochzta Sen. Overview of the CLEF 2017 personalised information retrieval pilot lab (PIR-CLEF 2017). In *Proceedings of CLEF 2017*, Dublin, Ireland, 2017. Springer.

Gabriella Pasi, Gareth J. F. Jones, Keith Curtis, Stefania Marrara, Camilla Sanvitto, Debasis Ganguly, and Prochzta Sen. Overview of the CLEF 2018 personalised information retrieval pilot lab (PIR-CLEF 2018). In *Proceedings of CLEF 2018*, Avignon, France, 2018. Springer.

Gabriella Pasi, Gareth J. F. Jones, Lorraine Goeuriot, Liadh Kelly, Stefania Marrara, and Camilla Sanvitto. Overview of the CLEF 2019 personalised information retrieval lab (2019). In *Proceedings of CLEF 2019*, pages 417–424, Lugano, Switzerland, 2019. Springer.

Camilla Sanvitto, Debasis Ganguly, Gareth J. F. Jones, and Gabriella Pasi. A laboratory-based method for the evaluation of personalised search. In Emine Yilmaz and Charles L. A. Clarke, editors, *Proceedings of The Seventh International Workshop on Evaluating Information Access - A Satellite Workshop of NTCIR-12*, pages 13 – 16, Tokyo, Japan, 2016.

Yoshiyuki Shoji, Kenro Aihara, Martin J. Dürst, Noriko Kando, Takuya Nakaya, Hiroaki Ohshima, Takehiro Yamamoto, and Yusuke Yamamoto. Location-based reminder for memorizing what visitors learn at a museum. In *Proceedings of the Third Workshop on Evaluation of Personalisation in Information Retrieval (WEPIR 2021)*, pages 79 – 87, Canberra, Australia, Virtual event, 2021.

Masaki Ueta, Tomoya Hashiguchi, Hun-Long Pham, Yoshiyuki Shoji, Noriko Kando, Yusuke Yamamoto, Takehiro Yamamoto, and Hiroaki Ohshima. Quiz generation on the electronic guide application for improving learning experience in the museum. In *Proceedings of the Third Workshop on Evaluation of Personalisation in Information Retrieval (WEPIR 2021)*, pages 96 – 104, Canberra, Australia, Virtual event, 2021.