



This is a repository copy of *Rethinking the test collection methodology for personal self-tracking data*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/152523/>

Version: Accepted Version

Proceedings Paper:

Hopfgartner, F. orcid.org/0000-0003-0380-6088, Gurrin, C. and Joho, H. (2019) Rethinking the test collection methodology for personal self-tracking data. In: Cheng, W.-H., Kim, J., Chu, W.-T., Cui, P., Choi, J.-W., Hu, M.-C. and De Neve, W., (eds.) Proceedings of the 26th International Conference on Multimedia Modeling, MMM2020. 26th International Conference on Multimedia Modeling 2020, 05-08 Jan 2020, Daejeon, South Korea. Lecture Notes in Computer Science, 11962 . Springer , pp. 463-474. ISBN 9783030377335

https://doi.org/10.1007/978-3-030-37734-2_38

This is a post-peer-review, pre-copyedit version of an article published in Cheng WH. et al. (eds) MultiMedia Modeling. MMM 2020. Lecture Notes in Computer Science. The final authenticated version is available online at: https://doi.org/10.1007/978-3-030-37734-2_38.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Rethinking the Test Collection Methodology for Personal Self-Tracking Data

Frank Hopfgartner¹, Cathal Gurrin², and Hideo Joho³

¹ University of Sheffield, Sheffield, UK

² Dublin City University, Dublin, Ireland

³ University of Tsukuba, Tsukuba, Japan

Abstract. While vast volumes of personal data are being gathered daily by individuals, the MMM community has not really been tackling the challenge of developing novel retrieval algorithms for this data, due to the challenges of getting access to the data in the first place. While initial efforts have taken place on a small scale, it is our conjecture that a new evaluation paradigm is required in order to make progress in analysing, modeling and retrieving from personal data archives. In this position paper, we propose a new model of Evaluation-as-a-Service that re-imagines the test collection methodology for personal multimedia data in order to address the many challenges of releasing test collections of personal multimedia data.

Keywords: lifelogging · evaluation · self-tracking.

1 Introduction

The Information Retrieval community has a long and rich tradition of creating and using reusable test collections which, if carefully curated, can support valid comparative evaluation of multiple systems and techniques. Within the past two decades, we have seen the emergence of various workshops and conferences that support community driven benchmarking activities. TREC [38], TRECVID [3], ImageCLEF [23], MediaEval [26], etc., all have played a part in supporting the community to generate test collections and coordinate such international comparative benchmarking efforts. In most cases, these activities have focused on the release of test collections that support the retrieval of online content, such as web pages, online or professional videos, blogs, and other crawlable content. In effect these can be considered to be DtA (Data-to-Algorithm) evaluation efforts where the test collection is centrally gathered and organised before being sent to participants to be used to evaluate their algorithms.

However, most of these DtA evaluation efforts have ignored user-generated content such as self-tracking or lifelogging data. This is due to many reasons, such as the difficulty of gathering personal user data, privacy concerns or data governance/legal issues. It is our conjecture that the community is consequently failing to address the very real need of managing modern personal data, such as personal photo archives, email archives, or even health-data archives. As a

consequence, the data created by individuals every day is not being well served by the current generation of test collections and benchmarking activities. While there have been some initial efforts, such as the lifelogging challenge at NT-CIR [16], or the lifelogging tasks at ImageCLEF [7] and MediaEval, even these can be considered to be pseudo-personal data challenges, since the test collections involved (lifelogs) are typically small in nature (from a few individuals). Furthermore, due to the constraints of DtA evaluation, the datasets have, by necessity, been filtered to remove content such as identifiable entities or content that the data gatherers were unwilling to make public. Clearly, if the community is to assist the real problem of individuals managing their own personal data collections, then an alternative approach is required that facilitates researchers to evaluate algorithms while reducing the risks inherent in releasing personal data test collections. Addressing this need, the contribution of this position paper is to propose an alternative approach, based on replacing the DtA evaluation effort with an AtD (Algorithm-to-Data) effort. This contribution is framed by considering the challenges of releasing a rich personal data archive for conventional DtA evaluation. The paper is structured as follows: In Section 2, we provide an overview of related work to set this position paper into perspective. Section 3 discusses steps necessary to enable collaborative research on personal data. The paper concludes with a discussion on future work in Section 4.

2 Background

While many forms of personal data exist, passively captured data, such as self-tracking data [8] or lifelog data [9], are likely to pose the biggest research challenges due to their diverse nature, volume and difficulty of managing. Self-tracking is concerned with individual-driven quantisation of some aspect of human life experience by passively (or actively) capturing data in order to achieve some form of self enhancement or betterment, whereas lifelogging is more focused on the indiscriminate passive capture of a wide range of personal data, from audio or visual data, right up to the total capture proposed in the MyLifeBits project [10].

Focusing on self-tracking data, this section provides a literature survey on the use of such data for research purposes. We first provide an introduction to self-tracking data in Section 2.1. Section 2.2 then describes ongoing efforts of archiving personal data. An overview of prior efforts on the creation of test collections for research is given in Section 2.3.

2.1 Self-Tracking Data

Self-tracking applications are usually developed to “help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge” [27]. In clinical psychology, the act of collecting personal data is also referred to as self-monitoring [25]. Inspired by the work of Gurrin et al. in [13], we list different categories of self-tracking data which can be applied at the time of writing:

- Passive visual capture. Utilising wearable devices such as the Narrative Clip⁴, the Microsoft SenseCam [20], or a first-generation Augmented Reality glasses, will allow for the continuous and passive capture of life activities as a sequence of digital images. Studies suggest that passive visual capture can be employed to help patients to cope with various memory difficulties, such as offsetting the effects of early stage dementia [32].
- Passive audio capture. Audio capture could allow for the identification of events or identification people who were speaking. Audio can easily be captured by smartphones. Audio recordings such as the ICSI Meeting Corpus [24] have played an essential role in the field of speech research.
- Personal biometrics. Sensing devices are becoming more common and widely used by self-trackers as it allow them to monitor their sleep duration, distance traveled, caloric output, and other biometrics data. The popularity of wearable devices such as Fitbit or Garmin suggest that capturing biometrics is a very common activity.
- Mobile device context. This refers to using the phone to continuously and passively capture the user’s context (e.g., location, movement, or acceleration). Combined with smart watches, mobile devices are able to capture a detailed trace of life activity and can be used e.g., to adapt information services [29].
- Communication activities. This refers to the phone or PC passively logging messages, emails, phone calls, or other means of communications. Tracking such activities help us to better understand the impact of technology on our communication behaviour (e.g., [2]).
- Data creation/access activities. Logging data consumed and created on a computer, for example, words typed, web pages visited, videos watched and so on. Collated analysis of such data can provide valuable insights on users’ interests (e.g., [39]).
- Environmental context and media. Other sources such as external sensor data or surveillance cameras could also be employed for self-tracking.
- Manual logging. This refers to the manual or direct logging of activity that is initiated by the user, for example, video recording, personal logs and diary.

2.2 Personal Data Archiving

As evident by the emergence of self-tracking communities such as The Quantified Self movement⁵ or parkrun [36], more and more people now own personal data as described above. As Hopfgartner and Davidson [21] point out though, although self-tracking has become very popular, efforts to archive such data is very limited in scale. They argue that this is mainly due to several challenges that need to be resolved first, including the creation of data, its active use, appraisal & selection, data transfer, storage & preservation, and access & re-use.

First organised efforts by the public sector to address these challenges can be observed in Japan where, in July 2019, the Information Technology Federation

⁴ <http://getnarrative.com>

⁵ <https://quantifiedself.com>

of Japan certified two companies as “information banks” to allow them to collect and use personal data for business purposes. This certification is based on the Japanese government’s guidelines compiled in June 2018 for information banking services to utilize personal information while protecting privacy for individuals⁶. This is an important movement for personal data since, just like a monetary bank, if a user prefers, they can delegate the collection and management of their personal data to a third-party institution, rather than managing by themselves. Although this could facilitate the use of personal data for business purposes, it could impact the development of community-based approaches to personal data.

Another development is the implementation of trading market platforms for personal data. For example, the company EverySense⁷ has introduced a platform to connect a business entity in need for relevant data to develop or enhance their services and products, with another party or individuals who agree to use sensors to collect and sell such data. This type of platform aims to trade various sensor data including personally collected data, in several domains such as healthcare, construction, car manufacturing, sports, or agriculture. This could include blood pressure, step counter, car navigation, smartphone, wearable devices, temperature, noise level, and so forth. Such a platform has a significant implication for the future of personal data since it could directly relate to the monetary benefit at a scale much bigger than what we have now.

Concluding from these developments, we argue that there are opportunities emerging that can result in the creation of centrally managed personal archives. An interesting research challenge of today is to explore how these archives can be used for further research. Following the argumentation of Chang et al. [6], we argue that this can be of particular importance in the field of computational social science where behavioural data is analysed to gain insights on society. Similar thoughts are presented by Sellen and Whittaker [33] who outline the usage domains of personal lifelog and sensor data. In the next section, we point to a number of initial efforts in the space.

2.3 Personal Data Test Collections

For decades, the information retrieval research community has been at the forefront to explore methods to enable reproducible and comparable research. As evidence, we point to success stories such as the creation of shared test collections and the development of widely accepted evaluation metrics. Please refer to Harman [17] for an historic overview of this development.

In the context of the creation of test collections consisting of personal sensor / lifelog data, the first thoroughly annotated dataset was developed for the Lifelog task of the Japanese-based evaluation conference NTCIR-12 [14]. This dataset consists of 79 days of data from three lifeloggers, totalling around 88k filtered and anonymised images with substantial metadata (locations, activities). This

⁶ <https://www.nippon.com/en/news/yjj2019070800965/aeon-unit-sumitomo-mitsui-trust-bank-certified-as-1st-ever-info-banks.html>

⁷ <https://every-sense.com/eng/>

first dataset was followed by another richer dataset for the NTCIR-13 competition [16], this time grouping visual, activity, biometric (e.g., heart rate, galvanic skin response, calorie burn, steps, etc.), location and metadata information. This new dataset consists of 90 days of data from two lifeloggers, totalling around 114k filtered and anonymised images. Subsequently, following the trend in society for richer personal data, a third related collection was released for NTCIR-14 [16], which consisted of 81k filtered and anonymised images from two users, spanning over 43 days, with rich personal activities and biometric data, which was focused on quantising the health, activities and diet of the two individuals.

While such datasets may have addressed the challenges of supporting initial efforts at comparative benchmarking for retrieval and analytics tasks, as outlined in the relevant overview papers from the NTICR competitions [14–16], they are still limited by the focus on a small number of (two or three) individuals and the datasets are anonymised in order to meet ethical, legal and privacy-by-design [5] principles. In fact, anonymisation of such datasets requires that the organisers of benchmarking efforts ignore the significant use-cases of data retrieval that require named individuals, which would be an important aspect of any lifelog retrieval system. While they serve their purpose, to fully address the challenges of personal data analytics and retrieval, larger archives of raw data are required.

We note some efforts at gathering limited or focused datasets from larger numbers of individuals. In most cases, these datasets have gathered egocentric video or lifelog image data. The Nagoya-COI [18] dataset is a collection of multimodal data recorded by one long-term and 18 short-term subjects, containing accelerometer, sound and video information captured by a smartphone and a video-camera. Heruzzo et al. [19] propose the LAP dataset, where they address the problem of analyzing socializing, eating and sedentary lifestyle patterns by recognizing the lifelogger’s activities. In total, they gathered 45k images from four different people in consecutive days and labeled them for diet-related activities. Lidon et al. [28] created a dataset with 10 day lifelogs from 5 different subjects producing a total of 7.1k images. A larger dataset named EgoSum+gaze [40] was generated by five subjects with 21 videos, each lasting between 15 and 90 minutes, adding up to about 15 hours of data. Miyanishi et al. [30] propose a daily living question-answering dataset which contains 20 continuous activities in six different places, each performed ten times, by eight subjects. Aghaei et al. [1] released the EgoSocialStyle dataset of 130k images from eight users, for the purpose of characterising the social patterns of a person relying exclusively on visual data. UT EE [37] gathered and crowd source annotated 27 videos from nine users, covering 14 hours in order to detecting the wearer’s engagement with the surrounding environment. An event segmentation dataset was published by Gupta and Gurrin [12], consisting of 14k images from ten participants. Each participant segmented their day into a set of activities, which was then labeled as the ground truth data for subsequent evaluation.

Where larger numbers of users are involved, in some cases the data released was limited to a feature vector, as opposed to the raw data, or the data was not released at all. For example, the R3 dataset [31] covers 57 users, over a period

of 1,723 days (1.5M images) and was used for event segmentation. However, due to privacy concerns, only the extracted visual features were ever publicly released. In the KidsCam dataset [35], 169 children captured 1.5M egocentric images with the purpose of detecting the presence of food and non-alcoholic beverages marketing and availability. The dataset was manually annotated with an custom vocabulary, but it was never released for community use, due to the sensitive nature of the data. It is the authors conjecture that few of these datasets are of sufficient volume or completeness to fully support the research community. While the NTCIR datasets are among the richest and largest, they are limited to a small number of individual data gatherers. At the same time, the KidsCam dataset was extensive in numbers, but limited in media (image and location) and could not be released. Additionally, all of these datasets are traditional centrally-hosted DtA collections where willing volunteers gather data and release datasets, which pose major challenges for researchers to release for community use, as outlined in the following section.

3 Enabling Collaborative Research on Personal Data

Above efforts illustrate that there is a clear desire in the research community to perform further research on self-tracking data. However, as has been shown, there are several shortcomings that hinder us in exploring open research challenges further. We argue that while it would be preferable to release datasets to the community in a conventional DtA manner, that are rich, non-anonymised and sourced from a large number of people, there are a number of reasons, for this not happening. In this section, we first elaborate further on the challenges for releasing personal test collections in a conventional manner (Section 3.1). In Section 3.2, we then argue to employ Evaluation-as-a-Service as an option to address these issues and challenges. Finally, in Section 3.3, we describe what an evaluation campaign on self-tracking data employing the Evaluation-as-a-Service paradigm would look like.

3.1 The Challenges of Releasing Personal Datasets in a Conventional DtA Manner

Prior efforts at releasing test collections of personal data faced a number of challenges, that ultimately reduced the number of such collections available. In particular, we can identify the following challenges from past work:

- Willing Volunteers. It has always proven difficult to recruit willing volunteers to gather data in a free-living, long-term environment.
- Legal Constraints. Recent data legislation (in Europe and elsewhere) means that the risks for challenge organisers are significantly higher once any form of personal data is released. While efforts at pseudo-anonymisation have been performed, it is likely that any effort to balance anonymisation and usefulness of data will mean that individuals could potentially be identified in the data.

- Maintaining Control of Test Collections. Legal and ethical expectations require that data owners (creators) maintain control of their data, so that it can be deleted after its lifetime, or upon request from a subject in the collection. Once a test collection is released to the public, even with comprehensive data agreements in place, the control of the data is effectively lost by the organisers and owners.
- Data Pre-processing. Data anonymisation efforts, e.g., as presented in [4], can go a long way toward the solving of legal and ethical constraints, however these occur at the expense of test collection usefulness, such efforts are actually very costly in terms of human effort. Cross-checking and ensuring effective anonymisation is still a human process, even if the anonymisation (e.g., blurring) algorithms are automatic. Legal requirements insist that state-of-the-art tools and reasonable effort are required to ensure appropriate levels of anonymisation.
- Ethical Data Release. To release a personal data from an academic or industrial source would usually require the prior-approval of an institutional ethics committee, which may not always be possible, due to the sensitive nature of the personal data.

Of course, there are benefits to the implementation of centrally controlled DtA test collections and evaluation campaigns, such as the ability of organisers to pre-process, filter and manually ensure accuracy and coverage of the data. It is also easier to control the process of creating appropriate research tasks and sourcing carefully curated information needs. Nevertheless, we argue that above barriers have often been too hard to overcome and therefore hampered further efforts to create and release public test collections. In the next section, we discuss how a methodology referred to as Evaluation-as-a-Service can help us to move one step forward.

3.2 Data Banks as Evaluation Platforms

Looking at above challenges, it is clear that the public release of self-tracking data for research is far more problematic than the release of test collections that do not reveal any personal information. We therefore have to explore options on how to allow further research on self-tracking data while at the same time guaranteeing and respecting users' privacy.

Considering recent developments on the establishment of data banks as a central facility to archive personal data, we argue that these archives could be used to gather data from large numbers of people, hence solving the challenge of finding users willing to share their data as outlined above. The potential of gaining access to this data depends on several factors. Most importantly, users have to be willing for their data can be used for research purposes. This is in particular important to address ethical issues as highlighted above. As outlined in Section 2, first services have entered the market now that allow users to sell their data. It is yet to be seen whether such services will be accepted by society and whether self-tracking data from a sufficient number of people could be employed for research purposes.

Addressing the challenge of maintaining control of data in the context of shared evaluation campaigns, Hopfgartner et al. [22] outline a new evaluation paradigm referred to as Evaluation-as-a-Service (EaaS). The basic idea of EaaS is to employ the so-called “Algorithms-to-Data paradigm” (AtD) rather than the established “Data-to-Algorithm” (DtA) approach, that we have considered thus far in this paper. The AtD paradigm would also facilitate comparative evaluation of systems and approaches using the same test collections, but in a less problematic manner. The difference between these two paradigms is as follows: The DtA paradigm is based on the approach of allowing researchers to download test collections and then run their experiments on their own IT infrastructure. Almost all evaluation campaigns that have been organised in the past few decades follow this approach. The AtD paradigm is approaching this from a different perspective. Rather than allowing full download of a test collection, participants are required to deploy their algorithms on a central IT infrastructure. So instead of having to give up control over the spreading of the collection, evaluation campaigns following the EaaS methodology can lock up their data on a central server where it can only be accessed for the purpose outlined in the evaluation protocol. As the authors clarify in [22], the benefits of this approach is manifold and for a detailed discussion, we refer to their paper. In the context of this paper, we consider the ability to maintain control over the data and to protect users’ privacy the main benefits of employing this approach. So instead of having to perform all the privacy-preserving steps outlined by Gurrin et al. [13], EaaS allows us to run algorithms on complete and unmodified datasets. In addition, since access to the full data is restricted to the algorithms, there is no need for tedious data cleaning. Careful configuration of the evaluation service can guarantee that no data is released without the consent of the data owner. It is also feasible then, to gather real-world information needs from the central IT infrastructure, as would be done by a real-world search engine evaluating ranking algorithm variations.

Summarising, in order to address the challenges that hinder the release of test collections containing personal data, we argue for the application of the EaaS paradigm where data banks play a key role in guaranteeing control over the data to protect users privacy. In the next section, we describe how a data challenge that employs this methodology would look like.

3.3 Toward a New Personal Data Challenge

By applying the AtD evaluation, we are able to facilitate the same comparative benchmarking efforts that the research community is accustomed to, while at the same time reducing the burden of creating and sharing a suitable dataset. Assuming that users are willing to keep their data with data banks, the challenge is now how to use these data banks as the backbone of a privacy-preserving EaaS infrastructure.

The first challenge is to employ an entity to serve as data bank. This has to be an entity that is able to attract large numbers of self-trackers, and at the same time gained their trust to handle their data with care.

Next, this data bank needs to run an EaaS platform. With the open source TIRA platform [11], a AtD solution already exists that allows researchers to upload software and data collections to the cloud, where the software can then be executed remotely. Technically, this is done by asking participants to deploy binaries within virtual machines, which are then executed on the platform. Apart from allowing these binaries to operate on a protected data collection, the platform also provides a web front-end to display evaluation results. A similar idea is presented by [34] who propose to rely on the cloud for training privacy-preserving activity recognition models using personal sensor data.

Having both data and infrastructure in place, a more demanding challenge becomes the need for the creation of realistic information needs with relevance judgements if an IR-like task is to be introduced. At prior evaluation campaigns that dealt with personal data (e.g., the NTCIR Lifelog task), the task of identifying information needs was given to the creator of the personal data. The main argument for this was that it is the user who is most familiar with their own data and therefore is most suitable to identify suitable known-item search tasks. Potentially, however, this is a very time consuming task as the creator needs to reflect on their behaviour and go over the dataset to identify suitable tasks, often generating complete relevance judgements in the process, or in some cases, pooled judgements. If a new data challenge uses data shared by users of data banks, the burden of formulating information needs should be kept to a very minimum and real-world user interactions with their archives, could be used as the basis of the queries and relevance judgments. Additionally, an automated analysis of the data collection might help to identify additional challenges and tasks for the research community to tackle.

4 Conclusion and Future Work

In this position paper, we discussed the challenge of enabling research on personal self-tracking data. After providing a detailed overview of prior research on the creation and use of self-tracking data for research, we identify issues that emerge when creating test collections of self-tracking data as commonly used by shared evaluation campaigns. This includes in particular the challenge of finding self-trackers willing to share their data, legal constraints that require expensive data preparation and cleaning before a potential release to the public, as well as ethical considerations. In order to address these challenges, we then argue to employ Evaluation-as-a-Service as a novel evaluation paradigm to enable collaborative research on personal self-tracking data. Evaluation-as-a-Service relies on the idea of a central data infrastructure that guarantees full protection of the data, while at the same time allowing algorithms to operate on this protected data. We further highlight the importance of data banks in this scenario. Finally, we briefly outline technical aspects that would allow setting up a shared evaluation campaign on self-tracking data.

As future work, we intend to explore this idea further by setting up an evaluation campaign on self-tracking data that fully implements the Evaluation-as-a-Service methodology.

References

1. Aghaei, M., Dimiccoli, M., Ferrer, C.C., Radeva, P.: Towards social pattern characterization in egocentric photo-streams. *Computer Vision and Image Understanding* (2018)
2. Alsuhaibani, A., Cox, A., Hopfgartner, F.: Investigating the role of social media during the transition of international students to the UK. In: *iConference 2019 Poster Proceedings. IDEALS* (2019)
3. Awad, G., Butt, A., Curtis, K., Lee, Y., Fiscus, J., Godil, A., Joy, D., Delgado, A., Smeaton, A.F., Graham, Y., Kraaij, W., Qunot, G., Magalhaes, J., Semedo, D., Blasi, S.: *trecvid*
4. Bailer, W.: Face swapping for solving collateral privacy issues in multimedia analytics. In: *MMM'19*. pp. 169–177 (2019)
5. Cavoukian, A.: Privacy by design: The 7 foundational principles. implementation and mapping of fair information practices. Information and Privacy Commissioner of Ontario, Canada (2010)
6. Chang, R.M., Kauffman, R.J., Kwon, Y.: Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems* **63**, 67–80 (2014)
7. Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of ImageCLEFLifelog 2018: Daily Living Understanding and Lifelog Moment Retrieval. In: *CLEF2018 Working Notes*. Avignon, France (2018)
8. Dix, A., Ellis, G.: The Alan Walks Wales Dataset: Quantified Self and Open Data, pp. 56–66. Open data as open educational resources: case studies of emerging practice, The Open University (2015)
9. Dodge, M., Kitchin, R.: 'Outlines of a world coming into existence': Pervasive computing and the ethics of forgetting. *Environment and Planning B: Planning and Design* **34**(3), 431–445 (2007). <https://doi.org/10.1068/b32041t>
10. Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C.: Mylifebits: Fulfilling the memex vision. In: *Proceedings of the Tenth ACM International Conference on Multimedia*. pp. 235–238. MULTIMEDIA '02, ACM, New York, NY, USA (2002)
11. Gollub, T., Stein, B., Burrows, S.: Ousting ivory tower research: towards a web framework for providing experiments as a service. In: *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12*, Portland, OR, USA, August 12-16, 2012. pp. 1125–1126 (2012)
12. Gupta, R., Gurrin, C.: Approaches for event segmentation of visual lifelog data. In: *MultiMedia Modeling*. pp. 581–593. Springer International Publishing, Cham (2018)
13. Gurrin, C., Albatal, R., Joho, H., Ishii, K.: A privacy by design approach to lifelogging. In: *Digital Enlightenment Yearbook 2014*, pp. 49–73. IOS Press (2014)
14. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Albatal, R.: Overview of NTCIR-12 Lifelog Task. In: Kando, N., Kishida, K., Kato, M.P., Yamamoto, S. (eds.) *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies*. pp. 354–360 (2016)

15. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Gupta, R., Albatal, R., Dang-Nguyen, D.T.: Overview of NTCIR-13 Lifelog-2 Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies (2017)
16. Gurrin C and Joho H and Hopfgartner F and Zhou L and Ninh V-T and Le T-K and Albatal R and Dang-Nguyen D-T and Healy G: Overview of the NTCIR-14 Lifelog-3 Task. In: Online Proceedings of the Fourteenth NTCIR conference (NTCIR-14). NII (6 2019)
17. Harman, D.: Information Retrieval: The Early Years. *Foundations and Trends in Information Retrieval* **13**(5) (2019)
18. Hayashi, T., Nishida, M., Kitaoka, N., TODA, T., Takeda, K.: Daily activity recognition with large-scaled real-life recording datasets based on deep neural network using multi-modal signals. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* **E101.A**, 199–210 (01 2018)
19. Herruzo, P., Portell, L., Soto, A., Remeseiro, B.: Analyzing first-person stories based on socializing, eating and sedentary patterns. In: *New Trends in Image Analysis and Processing – ICIAP 2017*. pp. 109–119. Springer International Publishing, Cham (2017)
20. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., Wood, K.: Sensecam: A retrospective memory aid. In: Dourish, P., Friday, A. (eds.) *UbiComp 2006: Ubiquitous Computing*. pp. 177–193. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
21. Hopfgartner, F., Davidson, J.: Digital preservation and curation of self-tracking data: A position paper. In: Proceedings of the 1st Workshop on Knowledge Discovery and User Modelling for Smart Cities co-located with 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, UMCit at KDD 2018, London, United Kingdom, August 20, 2018. pp. 1–5 (2018)
22. Hopfgartner, F., Hanbury, A., Müller, H., Eggel, I., Balog, K., Brodt, T., Cormack, G.V., Lin, J., Kalpathy-Cramer, J., Kando, N., Kato, M.P., Krithara, A., Gollub, T., Potthast, M., Viegas, E., Mercer, S.: Evaluation-as-a-service for the computational sciences: Overview and outlook. *J. Data and Information Quality* **10**(4), 15:1–15:32 (2018)
23. Ionescu, B., Müller, H., Péteri, R., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Cid, Y.D., Liauchuk, V., Kovalev, V., Ben Abacha, A., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Pelka, O., Friedrich, C.M., Chamberlain, J., Clark, A., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasillopoulos, N., Karampidis, K.: Imageclef 2019: Multimedia retrieval in lifelogging, medical, nature, and security applications. In: *Advances in Information Retrieval*. pp. 301–308. Springer International Publishing, Cham (2019)
24. Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., Wooters, C.: The icsi meeting corpus. In: *ICASSP'03* (April 2003)
25. Korotitisch, W.J., Nelson-Gray, R.O.: An overview of self-monitoring research in assessment and treatment. *Psychological Assessment* **11**, 415–425 (1999)
26. Larson, M., Arora, P., Demarty, C., Riegler, M., Bischke, B., Dellandréa, E., Lux, M., Porter, A., Jones, G.J.F. (eds.): *Working Notes Proceedings of the MediaEval 2018 Workshop*, Sophia Antipolis, France, 29-31 October 2018, CEUR Workshop Proceedings, vol. 2283. CEUR-WS.org (2018)
27. Li, I., Dey, A.K., Forlizzi, J.: A stage-based model of personal informatics systems. In: *Proceedings of the 28th International Conference on Human Factors in Com-*

- puting Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010. pp. 557–566 (2010)
28. Lidon, A., Bolaños, M., Dimiccoli, M., Radeva, P., Garolera, M., Giró, X.: Semantic summarization of egocentric photo stream events. In: LTA@MM (2017)
 29. Lorenz, F., Yuan, J., Lommatzsch, A., Mu, M., Race, N.J.P., Hopfgartner, F., Albayrak, S.: Countering contextual bias in TV watching behavior: Introducing social trend as external contextual factor in TV recommenders. In: Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video, Hilversum, The Netherlands, June 14-16, 2017. pp. 21–30 (2017)
 30. Miyanishi, T., Hirayama, J., Kanemura, A., Kawanabe, M.: Answering mixed type questions about daily living episodes. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI-18. pp. 4265–4271 (7 2018)
 31. Garcia del Molino, A., Lim, J.H., Tan, A.H.: Predicting visual context for unsupervised event segmentation in continuous photo-streams. In: ACM Multimedia conference (ACMMM18). pp. 10–17. MM '18, ACM, New York, NY, USA (2018)
 32. Piasek, P.: Case Studies in Therapeutic SenseCam use aimed at Identity Maintenance in Early Stage Dementia. Ph.D. thesis, Dublin City University (2015)
 33. Sellen, A.J., Whittaker, S.: Beyond total capture: A constructive critique of lifelogging. *Commun. ACM* **53**(5), 70–77 (May 2010)
 34. Servia-Rodriguez, S., Wang, L., Zhao, J., Mortier, R., Haddadi, H.: Privacy-preserving personal model training. In: ACM/IEEE International Conference on Internet of Things Design and Implementation. pp. 153–164 (2018)
 35. Smeaton, A.F., McGuinness, K., Gurrin, C., Zhou, J., O'Connor, N.E., Wang, P., Davis, B., Azevedo, L., Freitas, A., Signal, L., Smith, M., Stanley, J., Barr, M., Chambers, T., Mhurchu, C.N.: Semantic indexing of wearable camera images: Kids'cam concepts. In: Proceedings of the 2016 ACM Workshop on Vision and Language Integration Meets Multimedia Fusion (2016)
 36. Stevinson, C., Wiltshire, G., Hickson, M.: Facilitating participation in health-enhancing physical activity: A qualitative study of parkrun. *International Journal of Behavioral Medicine* **22**(2), 170–177 (Apr 2015)
 37. Su, Y.C., Grauman, K.: Detecting engagement in egocentric video. In: Proceedings of the European Conference on Computer Vision (ECCV) (2016)
 38. Voorhees, E.M., Harman, D.K.: TREC: Experiment and Evaluation in Information Retrieval (Digital Libraries and Electronic Publishing). The MIT Press (2005)
 39. Walsh, D., Clough, P., Hall, M., Hopfgartner, F., Foster, J., Konstonatsios, G.: Analysis of transaction logs of national museums liverpool. In: Proceedings of TPD'L'19. pp. 84–98. Springer Verlag (2019)
 40. Xu, J., Mukherjee, L., Li, Y., Warner, J., Rehg, J.M., Singh, V.: Gaze-enabled egocentric video summarization via constrained submodular maximization. In: Proc. CVPR (2015)