WildKey: A Privacy-Aware Keyboard Toolkit for Data Collection In-The-Wild

André Rodrigues
afrodrigues@fc.ul.pt
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisbon, Portugal

André Santos
abranco@lasige.di.fc.ul.pt
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisbon, Portugal

Kyle Montague
kyle.montague@northumbria.ac.uk
NorSC, Northumbria University
Newcastle, United Kingdom

Hugo Nicolau
hugo.nicolau@tecnico.ulisboa.pt
ITI/LARSyS, Instituto Superior Técnico, Universidade de Lisboa
Lisbon, Portugal

Tiago Guerreiro
tjvg@di.fc.ul.pt
LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisbon, Portugal

ABSTRACT

Touch data, and in particular text-entry data, has been predominantly collected in laboratory settings, under controlled conditions. While touch and text-entry data has consistently shown its potential for monitoring and detecting a variety of conditions and impairments, its deployment in-the-wild remains a challenge. In this paper, we present WildKey, an Android keyboard toolkit that allows for the usable deployment of in-the-wild user studies. WildKey is able to analyse text-entry behaviours through implicit and explicit text-entry data collection while ensuring user privacy. We detail each of the WildKey's components and features, metrics collected, and discuss the steps taken to ensure user privacy thus promoting compliance.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility**; **Ubiquitous and mobile computing**.

KEYWORDS

Text-Entry; Touch Dynamics; In-the-wild; Smartphones; Data Collection

ACM Reference Format:
André Rodrigues, André Santos, Kyle Montague, Hugo Nicolau, and Tiago Guerreiro. 2021. WildKey: A Privacy-Aware Keyboard Toolkit for Data Collection In-The-Wild. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3460418.3479366

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479366

1 INTRODUCTION

The widespread use of smartphones has led to a massive increase in the generation of personal data, with text input being one of the most common tasks. People type to send messages, write emails, engage in social networks and much more. The data generated from these interactions has remarkable potential as a digital endpoint for disease detection and monitoring, to the quantified self movement, and biometrics. As a digital endpoint, text-entry metrics have been used to distinguish between people with Parkinson's Disease (PD) and control groups showing potential for early disease detection [4, 8, 12], to assess stress [7], fatigue [3], distinguish between patients with multiple sclerosis and controls [15], and even for ubiquitous inebriation assessment [18]. In the field of user authentication, keystrokes dynamics have been used to discriminate among users detecting potential impostors, indicating that typing behaviour is highly personal [14]. Preliminary work also suggests that differences between typing sessions can be associated with users' emotions [11]. Despite all the potential shown by recent work, collecting this type of data, and particularly in-the-wild, raises several challenges in regards to required user effort, compliance, privacy, ease of deployment and study oversight.

There have been multiple approaches to collecting typing data in-the-wild. The most simple method is explicitly prompting users to do a specific task (e.g. transcription tasks) in a custom made application [21]. However, this method requires additional effort from the user and it is not able to collect natural typing behaviour nor provide spontaneous assessments. Overall, one can expect less compliance as more data is requested from users. It has the benefit of mimicking controlled laboratory tasks, which may provide data with less noise. A second approach is to collect everything the user is typing by logging keypresses and touches [19], which faces issues regarding user privacy and overall adherence to the study. A third solution is to only collect typing metrics that are not related with the content written such as flight time and hold times [12]. The approach ensures user privacy at the cost of limiting the type of metrics one can extract from typing sessions. A similar approach is to obfuscate the text written, either by only parsing parts of the data [6] or by analysing the typing session and only storing an abstract representation of the sentences (e.g. "Noun Verb Adjective" or substituting every letter with an "M") [5, 9]. Depending on the method used, a different set of potential features can be extracted. All of the approaches described have advantages and disadvantages to them, and depending on the context, different ones, or a combination of several may be the ideal solution.

The WildKey toolkit was designed to support the collection of data both explicitly with prompted tasks, and implicitly providing the best of both. With the WildKey keyboard one can passively analyse all text written regardless of the application the user is in. Unlike prior approaches, WildKey neither stores raw textual data, nor it obfuscates typing sessions. Instead, for implicit data collecting, we shifted the analyses that required access to the raw text to the device, and only calculated metrics with no potential to reconstruct the original text. The toolkit enables researchers and developers to tailor the WildKey data collecting to what best suits their study. The WildKey keyboard supports an unconstrained text-entry protocol [17] where users can freely to input text, including using suggestions, auto-correct, and cursor changes [23]. The toolkit is able to provide all the traditional text-entry metrics on speed, error rates [17], and touch behaviours [4, 13, 15]. As a toolkit for in-the-wild data collection it provides additional features that are commonly needed and even required to successfully run this type of studies. In addition to the text-entry tasks, one can create and schedule questionnaires and other custom made tasks such as the Alternate Finger Tapping assessment. Lastly, the toolkit has a Study Manager application that enables researchers/developers to easily schedule, deploy, and oversee their active studies.

2 WILDKEY

The WildKey toolkit is composed of a *Keyboard Android app, a Study Management & Analytics app (React), and a NoSQL Database (Firebase).* The toolkit was developed to enable researchers and developers to extend and deploy their own standalone ecosystems. The repository is open source, under the license Attribution 4.0 International (CC BY 4.0) and available at [2].

The WildKey toolkit was designed to be **Privacy-Aware** in order to not only comply with the current standards for data protection, but also promote user compliance and adherence to study protocols. As such, during implicit text-entry collection (i.e. when the user is writing using the keyboard in any and all text fields regardless of application) no raw text is ever recorded on device storage or sent to the cloud, nor any data that would enable anyone to reconstruct the written text. All text entries that require content analysis are done locally on the device, and only processed data is sent to the cloud database. We highlight these and other design choices in the *Privacy-Aware* section.

2.1 Android Keyboard

The WildKey keyboard extends the Android Open Source Project Keyboard [1]. As such, it supports 26 languages, provides auto-correct, suggestions, customization options among many other traditional keyboard features. WildKey keyboard is responsible for collecting, and processing all the data inserted and storing its results on the defined cloud database storage.

2.2 Data Collection

The WildKey Keyboard app is prepared to collect three types of data: text-entry, questionnaires and custom made tasks. For text-entry, the keyboard is able to assess *Implicit Text-entry* data and to prompt users to do text-entry tasks collecting *Explicit Text-Entry* data. The toolkit also enables developers to create *Questionnaires* to be answered within the app that accompanies the keyboard. Lastly, we designed the toolkit to be flexible and support the creation of *Custom-Made* tasks relevant for a specific study protocol. Through the Study Management & Analytics application researchers are able to define and schedule their user studies and associate study tasks/schedules to registered users.

2.2.1 *Explicit Text-Entry Collection.*

There are two types of text-entry tasks supported by WildKey. Transcription tasks used in traditional text-entry studies (i.e. asking the user to transcribe a set of phrases); and Composition tasks (i.e. by asking the user a set of questions). Users can be prompted at scheduled times to complete the tasks. In Composition tasks where we do not have the ability to know for certain the intended sentence the user was trying to write, we estimate it. We relied on the approach presented in [9] where target intent is calculated using a spell checker.

While Trasnscription tasks enable us to calculate error rate related metrics with high confidence, Composition tasks may assess more natural typing behaviours. Both types of explicit collection face similar challenges in its ability to collect data spontaneously, for extended or frequent periods of time.

2.2.2 *Implicit Text-Entry Collection.*

When the keyboard is installed and active, all written text, regardless of where it is written, is analysed to calculate all the metrics described below in Analytics. The exceptions to the rule are when users are writing in password fields or inserting only numbers, where no metrics are calculated. Similar to composition tasks, during implicit data collection, we do not know the user intended target sentence. As such we use the protocol described in the previous section to calculate user intent enabling us to assess error related metrics. To preserve users' privacy and promote compliance, no text or any data that would enable its reconstruction is recorded during implicit collection. All metrics are calculated locally on the device and only metrics with no textual content are sent and stored in the cloud database.

2.2.3 *Questionnaires & Custom Tasks.*

A key feature to conduct in-the-wild studies is the ability to prompt users to answer and fill in short questionnaires and scales. As such, the Wildkey toolkit has the ability to schedule and prompt users to answer custom made questionnaires. Similarly, Wildkey supports the creation of custom made tasks to schedule and deploy to its users. Currently, the toolkit has one example of such a task: the Alternate Finger Tapping test [16].

2.3 Study Management & Analytics

The React app has two key features: study management; and analytics. One can create, check and manage all the existing study schedules, tasks and registered users. Through the study dashboard, researchers/developers can quickly oversee the participants compliance and overall performance. All components of Wildkey were developed to be easily expandable to accommodate other tasks/metrics/visualizations which may be fundamental to the studies one desires to deploy.

All metrics collected are calculated locally on the device before being sent out to the cloud database. By default, all data is stored in a JSON format and is readily available for download to post-process it wherever you desire. We divided the collected and calculated metrics into five types: speed, errors, touch dynamics, action and character counts and other. You can find a complete list in [2]. All questionnaire answers produce data which is also stored in the same database.

2.3.1 Metrics. We consider a text-entry session to be either a single task in explicit collection or begin when the keyboard is opened and finish when the keyboard is hidden. We calculate all metrics for each session (with the ability to discard sessions with fewer than X characters written). **Speed** metrics [17] are calculated taking into account the time from the first entered character to the time of the last character entered. **Error rate** metrics [19, 22] are an approximation, and a characterization of the errors users made while writing. WildKey collects a variety of **touch related metrics** from Flight and Hold time[4] to Touch Offsets [10] and even raw touch points in explicit tasks. To enable us to better characterize the users text-entry behaviours, Wildkey collects a wide variety of **action and character counts** (e.g. Selected Suggestions, Cursor Changes). Lastly, WildKey collects Raw Text in explicit tasks and a variety of device and operating system characteristics (e.g. keyboard language).

2.4 Privacy-Aware

Smartphones are becoming an extension of oneself [20] and data privacy has become of the utmost importance. Approaches that seek to collect text-entry behaviours can expect to be met with resistance by some users and suffer from lower levels of adherence and compliance. Aware of the challenges, WildKey was devised to not store any textual content other than in explicit tasks. Moreover, all metrics that would enable the reconstruction of the text content are not collected during implicit text-entry. Furthermore, when the user is inserting only numbers or in passwords fields no analysis is conducted. To ensure users are in control of the data they are sharing, WildKey has an always available button on the top left of the keyboard to activate an “incognito mode”, which resets every time the keyboard is closed. When active, no data is analysed.

3 CONCLUSIONS

Data collected from typing sessions has continually shown its potential for a variety of different domains. As a complex task that combines motor and cognitive functions, the features one can extract can be explored for disease detection and monitoring, for biometrics, personalized assistive technology, text input researchers, and within the quantified self movement. To support and promote further explorations, we developed and made available to all the WildKey Toolkit with its first release. The toolkit is currently being leveraged in multiple studies with topics from privacy and compliance, to fatigue, to monitoring motor fluctuations in neurodegenerative diseases. The toolkit is an ongoing open source project and we welcome contributions. New features will continue to be added as new requisites appear to support studies with different goals and in different contexts.

ACKNOWLEDGMENTS

We would like to thank our collaborators from OpenLab Newcastle as the first external team to use the keyboard toolkit providing us with valuable feedback and contributions to the project. This project was partially supported by FCT through project mIDR (AAC 02/SAICT/-2017, project 30347, cofunded by COMPETE/FEDER/FNR), LASIGE Research Unit funding, ref. UIDB/00408/2020 and ref. UIDP/00408/2020, LARSyS Research Unit funding, ref. UIDB/50009/2020, and the IDEA-FAST project which has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 853981. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and associated partner.

REFERENCES

[1] 2021. Android Open Source Project. https://source.android.com/. (Accessed on 05/18/2021).

[2] 2021. WildKey. https://techandpeople.github.io/keyboard/. (Accessed on 06/14/2021).

[3] Hilal Al-Libawy, Ali Al-Ataby, Waleed Al-Nuaimy, Majid A. Al-Taee, and Qussay Al-Jubouri. 2016. Fatigue Detection Method Based on Smartphone Text Entry Performance Metrics. In *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*. 40–44. https://doi.org/10.1109/DeSE.2016.9

[4] Teresa Arroyo-Gallego, María Jesus Ledesma-Carbayo, Álvaro Sánchez-Ferro, Ian Butterworth, Carlos S. Mendoza, Michele Matarazzo, Paloma Montero, Roberto López-Blanco, Verónica Puertas-Martín, Rocío Trincado, and Luca Giancardo. 2017. Detection of Motor Impairment in Parkinson's Disease Via Mobile Touchscreen Typing. *IEEE Transactions on Biomedical Engineering* 64, 9 (2017), 1994–2002. https://doi.org/10.1109/TBME.2017.2664802

[5] Florian Bemmann and Daniel Buschek. 2020. LanguageLogger: A Mobile Keyboard Application for Studying Language Use in Everyday Text Communication in the Wild. *Proc. ACM Hum.-Comput. Interact.* 4, EICS, Article 84 (June 2020), 24 pages. https://doi.org/10.1145/3397872

[6] Daniel Buschek, Benjamin Bisinger, and Florian Alt. 2018. ResearchIME: A Mobile Keyboard Application for Studying Free Typing Behaviour in the Wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173829

[7] Matteo Ciman, Katarzyna Wac, and Ombretta Gaggi. 2015. iSensestress: Assessing stress through human-smartphone interaction analysis. In *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. 84–91. https://doi.org/10.4108/icst.pervasivehealth.2015.259280

[8] Neil Dhir, Mathias Edman, Álvaro Sanchez Ferro, Tom Stafford, and Colin Bannard. 2020. Identifying robust markers of Parkinson's disease in typing behaviour using a CNN-LSTM network. In *Proceedings of the 24th Conference on Computational Natural Language Learning*. Association for Computational Linguistics, Online, 578–595. https://doi.org/10.18653/v1/2020.conll-1.47

[9] Abigail Evans and Jacob Wobbrock. 2012. Taming Wild Behavior: The Input Observer for Obtaining Text Entry and Mouse Pointing Measures from Everyday Computer Use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1947–1956. https://doi.org/10.1145/2207676.2208338

[10] João Guerreiro, André Rodrigues, Kyle Montague, Tiago Guerreiro, Hugo Nicolau, and Daniel Gonçalves. 2015. TabLETS Get Physical: Non-Visual Text Entry on Tablet Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 39–42. https://doi.org/10.1145/2702123.2702373

[11] Stelios Hadjidimitriou, Dimitrios Iakovakis, Vasileios Charisis, Sofia B. Dias, José A. Diniz, Julien Mercier, and Leontios J. Hadjileontiadis. 2017. On Capturing Older Adults' Smartphone Keyboard Interaction as a Means for Behavioral Change Under Emotional Stimuli Within i-PROGNOSIS Framework. In *Universal Access in Human–Computer Interaction. Design and Development Approaches and Methods*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, 346–356.

[12] Dimitrios Iakovakis, Stelios Hadjidimitriou, Vasileios Charisis, Sevasti Bostantjopoulou, Zoe Katsarou, Lisa Klingelhoefer, Heinz Reichmann, Sofia B. Dias,

José A. Diniz, Dhaval Trivedi, K. Ray Chaudhuri, and Leontios J. Hadjileontiadis. 2018. Motor Impairment Estimates via Touchscreen Typing Dynamics Toward Parkinson's Disease Detection From Data Harvested In-the-Wild. *Frontiers in ICT* 5 (2018), 28. https://doi.org/10.3389/fict.2018.00028

[13] Dimitrios Iakovakis, Stelios Hadjidimitriou, Vasileios Charisis, Sevasti Bostantzopoulou, Zoe Katsarou, and Leontios J. Hadjileontiadis. 2018. Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage Parkinson's disease. *Scientific Reports* 8, 1 (May 2018). https://doi.org/10.1038/s41598-018-25999-0

[14] Kevin S. Killourhy and Roy A. Maxion. 2009. Comparing anomaly-detection algorithms for keystroke dynamics. In *2009 IEEE/IFIP International Conference on Dependable Systems Networks*. 125–134. https://doi.org/10.1109/DSN.2009.5270346

[15] KH Lam, KA Meijer, FC Loonstra, EME Coerver, J Twose, E Redeman, B Moraal, F Barkhof, V de Groot, BMJ Uitdehaag, and J Killestein. 2020. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis. *Multiple Sclerosis Journal* 0, 0 (2020), 1352458520968797. https://doi.org/10.1177/1352458520968797 arXiv:https://doi.org/10.1177/1352458520968797 PMID: 33150823.

[16] Chae Young Lee, Seong Jun Kang, Sang-Kyoon Hong, Hyeo-Il Ma, Unjoo Lee, and Yun Joong Kim. 2016. A Validation Study of a Smartphone-Based Finger Tapping Application for Quantitative Assessment of Bradykinesia in Parkinson's Disease. *PLOS ONE* 11, 7 (07 2016), 1–11. https://doi.org/10.1371/journal.pone.0158852

[17] I. Scott MacKenzie, Tatu Kauppinen, and Miika Silfverberg. 2001. Accuracy Measures for Evaluating Computer Pointing Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 9–16. https://doi.org/10.1145/365024.365028

[18] Alex Mariakakis, Sayna Parsi, Shwetak N. Patel, and Jacob O. Wobbrock. 2018. Drunk User Interfaces: Determining Blood Alcohol Level Through Everyday Smartphone Tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 234, 13 pages. https://doi.org/10.1145/3173574.3173808

[19] Hugo Nicolau, Kyle Montague, Tiago Guerreiro, André Rodrigues, and Vicki L. Hanson. 2017. Investigating Laboratory and Everyday Typing Performance of Blind Users. *ACM Trans. Access. Comput.* 10, 1, Article 4 (March 2017), 26 pages. https://doi.org/10.1145/3046785

[20] Chang Sup Park and Barbara K. Kaye. 2019. Smartphone and self-extension: Functionally, anthropomorphically, and ontologically extending self via the smartphone. *Mobile Media & Communication* 7, 2 (2019), 215–231. https://doi.org/10.1177/2050157918808327 arXiv:https://doi.org/10.1177/2050157918808327

[21] Shyam Reyal, Shumin Zhai, and Per Ola Kristensson. 2015. Performance and User Experience of Touchscreen and Gesture Keyboards in a Lab Setting and in the Wild. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 679–688. https://doi.org/10.1145/2702123.2702597

[22] Jacob O. Wobbrock and Brad A. Myers. 2006. Analyzing the Input Stream for Character- Level Errors in Unconstrained Text Entry Evaluations. *ACM Trans. Comput.-Hum. Interact.* 13, 4 (Dec. 2006), 458–489. https://doi.org/10.1145/1188816.1188819

[23] Mingrui Ray Zhang and Jacob O. Wobbrock. 2019. Beyond the Input Stream: Making Text Entry Evaluations More Flexible with Transcription Sequences. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 831–842. https://doi.org/10.1145/3332165.3347922