# Project: Deep Activity Tracking on Android Smartphones

Smartphones are supporting us in our everyday lives, e.g. by providing tailored information on demand anywhere and anytime. However, the smartphone mostly acts as a "passive servant", meaning that users have to actively request help or information. Smartphones can only rarely act proactively themselves yet.

The basis for proactive support is intent prediction: The smartphone needs to know what the user might want without the user explicitly stating it. This might be impossible to achieve for any arbitrary situation, however for context dependent, repetitive tasks smartphones could learn from past behaviors to predict user desires in the future.

Therefore a broad picture of the users activities (e.g. Which apps are used? What is the user doing in the app?), state (e.g. cognitive load, emotion, stress level), and context (e.g. Where is the user? What is he currently doing in the real world?) is needed. While context aware applications are meanwhile supported by device APIs [e.g. the Google Snapshot API[1]] and assessing user's (mental) state is an active research topic with many proposed solutions, work on collecting smartphone usage is superficial: I.e. most work satisfies with knowledge on what apps are used, but do not regard what users are actually doing inside an app. However "deep app tracking" could empower intent prediction systems to not just prompt the next used app, but instead offer a deep link that directly performs a user action.

Example: Tom commutes to his office daily by public transport. He has multiple connection options to choose of, depending on when exactly he is leaving from home and whether there are any service issues on one of the connections. Thus he checks the public transport company's app every morning, querying for the fastest connection from home to work by entering his home- and office location. Current intelligent devices would learn that Tom uses the public transport company app every morning, and thus place its icon prominently on the screen on mornings. However with the ability of deep app tracking, the device could learn that Tom is querying the fastest connection from his home to office every morning. The device could offer a deep link, i.e. an icon that directly performs the desired action.

Existing approaches for deep app tracking evaluate screenshots [e.g. the Screenomics project[2]]. Although that approach is comprehensive, it is very privacy invasive as raw screen contents are transferred to a server. Furthermore it's resulting data is hard to use, as features describing higher level actions have to be extracted with computer vision algorithms. A different approach is enabled by Android's accessibility service[3]: An app can retrieve the window's XML structure and subscribe to interaction events. However this approach does not work for all views, as for example the content of web views and prerendered content is not represented in the XML structure. Furthermore this vast amount of unstructured data is hard to gain valuable information from.

[1] https://developers.google.com/awareness/android-api/snapshot-api-overview
[2] http://clgiles.ist.psu.edu/pubs/HCI2019.pdf
[3] https://developer.android.com/reference/android/accessibilityservice/AccessibilityService#retrieving-window-content

I can currently imagine the following theses in this project. Due to the project complexity and advanced technologies I would recommend them for a Master thesis, however if you are looking for a Bachelor thesis and can argue why your experience is sufficient for this project feel free to contact me as well. The topics would also be suitable with some adaptions as Einzelpraktikum for Master students.

## Transforming UI tree data to vectors

In order to work with user actions that can be retrieved from UI trees, they have to be converted to some uniform structure. Related approaches have yet been investigated for given interaction trace datasets[4], screenshot sequences[5] , HTML trees[6], and source code snippets[7]. The goal of your thesis would be to develop a procedure that transforms sequences of UI events and trees, that are fetched from Android's accessibility services, into a vector representation. The resulting data should be usable to a.o. work on action prediction models, or conduct behavioral analyses on smartphone usage patterns in HCI and psychology. A central aspect should thereby be the user privacy aspect. The data basis, UI tree sequences and events, are highly sensitive user data. A transformation into vector data not only reduces complexity, but also can be used to remove technically unnecessary but privacy invasive details. Depending on the use case, different aspects of the data can or may not be abstracted. Therefore in your thesis, you should come up with a way to parameterize your vectorization approach, so that the resulting data's granularity and level of detail are a good fit and minimalistic regarding the desired use case.

## On-Device Sequence Prediction

This thesis should explore how intent prediction can be applied in the wild. This encompasses (1) using sequence prediction models on-device (i.e. on the smartphone), and (2) creating UX concepts that create a benefit for their users.

The machine learning part (i.e. step 1) could be implemented e.g. using tensor flow (see Ghods et al.[8] for a similar project). Assuming that a continuous stream of smartphone usage data is given, you should explore how this can be used to predict future usage sequences in practice. Do we need a pretrained model? Is it better to train a personalized model on-device? Or a mixture of both, e.g. transfer or federated learning[9]?

In step 2 of the thesis you will create UX concepts, on how the obtained predictions can be used to create benefit for the user. A prototypical implementation should be created, and evaluated in a user study. Your findings should inform the development of future interfaces that use predictions on the user's intents and next actions.

[4] Jia-Jun Li et al. 2021. Screen2Vec: Semantic Embedding of GUI Screens and GUI Components
[5] Reeves et al. 2021. Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways Technology Shapes Them
[6] Wu et al. 2022. Distributed Representations of HTML Page
[7] Alon et al. 2019 code2vec: Learning distributed representations of code
[8] Ghods et al. 2019. Activity2Vec: Learning ADL Embeddings from Sensor Data with a Sequence-to-Sequence Model
[9] https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Each thesis consists of…

- A comprehensive literature research in your domain of research. Your concept and design decisions have to be based on related work where possible. You have to work out a research gap that your project fits into, to avoid reinventing the wheel.
- Coming up with an approach on how you plan to tackle your research question, including steps and timeplan
- Writing up your project in a thesis document
- Present your project in the disputation seminar two times, “Antrittsvortrag” and final presentation

You are a good fit for this topic if you…

- Have advanced programming skills and experience in the relevant technologies
- Love to dive deep into and work with large amounts of data
- Have a creative and proactive mindset to come up with ideas and approaches to tackle problems
- Are in your third (or higher) semester of your Master studies in media informatics, computer science, or a related field
- You are the project leader of your thesis. This requires an active working style and self-management, including making decisions to stay on track towards meeting your thesis’ goal.

If you are interested, please drop me a message including some information why you feel you are a good fit for this topic and a transcript of records to florian.bemmann@ifi.lmu.de !