Using process mining to classify habits through the analysis of daily activities

David
Lütke-Sunderhaus
University of Twente
PO Box 217, 7500 AE
Enschede
the Netherlands

d.j.luetke-sunderhaus@student.utwe
nte.nl

ABSTRACT
Routines and habits are an essential part of our daily life. Everyone has them and inevitably develops them. They are the human mind's way of automizing behaviour and let us work more efficiently. However, this behaviour might not always be desired and not everyone is happy with their routines and habits. Some want to change their habits because they see them as disadvantageous, or they are actively harmful. Behaviour support technology can assist them with changing their routines. However, for a behaviour support agent to assist with the change, a clear definition of the routine that needs changing is needed first. Many people do not know the exact sequence of events that form their routine. In this paper, we want to categorize routines and habits and find out how we can identify them. First, we will perform a literature study, to get an understanding of the underlying matter as well as find categories to place habits and routines in. To verify our results from the literature, we will use process mining as a means to analyse datasets [16] containing daily activity logs.

Keywords
Routines, habits, process mining, classification, Behaviour support agents, Daily routines, routine classification, habit classification

1. INTRODUCTION
Habits and routines are an integral part of human behaviour. They largely dictate what we do in any given day and how we perform these activities. While a routine is not exactly the same as a habit, they share many similarities. Both are descriptive terms of frequently repeated behavioural patterns [17].

35th Twente Student Conference on IT, July. 2nd, 2021, Enschede, The Netherlands.

Most of the time, acting habitual is the easier and often the desired way of performing an activity since it speeds up the activity and relieves the individual of having to make the same decision every time they perform this activity, thus easing mental stress [13].

However, habits can also have a negative impact, as many bad behaviours are expressed through habits or routines, such that we do not have to make the conscious decision of expressing the undesirable behaviour every time it is performed. For example, some eating disorders such as binge-eating manifest as eating habits [3, 6]. Other habits some describe as non-beneficial may include checking your phone after waking up, or buying things one might not need [13].

There are desired and actual behaviours people have, that can manifest as a habit [17]. If there is a discrepancy between the two behaviours, the individual will often try to change the actual behaviour to fit more closely with the desired behaviour. Changing a habit however is hard, as habits are ingrained into our everyday lives, and we execute them unconsciously, without intent. They are normal behaviour, and it is difficult to change your normal behaviour [13].

Behaviour support agents provide this needed help. They are aimed at assisting people in organizing or changing their routines [11]. In order for the behaviour support agent to properly function, the person using it needs to specify what their current and desired behaviours are and how they want to change them. The problem with this approach is that habits have stimuli and the person might not be aware of it. We can combat this problem by analysing an activity log and find out which routines exist in the person's day. The dataset we have [16] consists of activity logs from three people over the course of 59 days. While the raw data is very descriptive of what the three individuals do, it is not very useful in its current state. Because we do not know which of the logged activities constitute a routine, we can not analyse these routines to create a better behaviour support agent. Since the behaviour support agent is unable to function on raw data and needs to know a habit before it can help the user change or improve on it, we need to classify the activities and behaviours from the datasets [16].

In the following research paper, we want to achieve the categorization and identification of types of habits, as well as find examples for these habit categories. We will use the data analysis method of process mining to find these examples in the datasets [16]. This dataset is an activity log provided by Sztyler,

T. and Carmona, J. and published by the University of Mannheim. It contains daily activities of three individuals over a period of 59 days. We will analyse these datasets [16] to confirm the habit categories we found in the literature study.

2. BACKGROUND

2.1 Routines and Habits

The Oxford Dictionary defines habits as "a thing that you often do and almost without thinking, especially something that is hard to stop doing"[1]. Many scholars agree with this definition, where a habit is defined as "behaviour that is being performed often (daily) in stable contexts" [13]. Routines on the other hand, although similar to habits, have two important distinctions.

Routines are also a sequence of activities or behaviours, like a habit, where the addition of a stimulus, or trigger, and a reward transform the routine into a habit [17]. Therefore, the biggest difference between the two is that habits are sequences of behaviours, usually in response to a stimulus and with some form of reward, whereas routines are sequences of activities automatically performed when we carry out a known task [17]. Washing your hands, as seen in Figure 1, is a good example of a habit, since it is an automated response to a stimulus. Your hands being dirty is the stimulus, washing them is the routine, with the reward being clean hands.

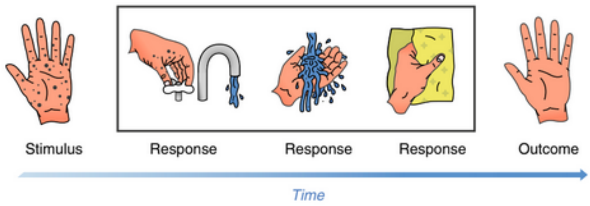


Figure 1: Visual representation of a habit [17]

Although habits and routines have a minor but important distinction, the two words are often used interchangeably. Because of the nature of our data, only being activity logs without the addition of environmental or mental data to identify stimuli, we will use the two words interchangeably as well. However, in the greater picture the distinction is rather important, as changing a habit often includes changing the stimulus or the reward and seldom the routine itself, as we do not intentionally or cognitively perform the routine [13]. Therefore, for a behaviour support agent to assist in changing a routine or habit, it first needs to know which it is and what the stimulus and rewards are in case it is a habit. There are other forms of motivation for a habit to form, which we will describe in detail in the results section, but for the most part, a habit is triggered by the stimulus.

2.2 Process Mining

Process mining is a subdomain of data mining. Its primary use is to extract knowledge from action or event-based data. Usually, process mining is used on business processes, visualizing them in process diagrams and analysing those diagrams to find flaws and bottlenecks in the business process. These flaws are then analysed to optimize the business process [12,18]. This is usually achieved through visualization of the process. When a business process is planned, it is designed to follow a specific set of steps in a specific order. In reality, business processes often look decisively different from the planned straight line through the steps (see Figure 2) [18]. There are workarounds, some steps might not always be applicable or skipped for simplicity. The same can be said about a routine. The idea of a routine is relatively rigid. If it had to be visualized out of memory, it would be just like the ideal process in Figure 2, a straight line. However, this is not the reality, as can be seen in Appendix A. Appendix A is a process map of a single day from the dataset [16] hh_110_labour. As can be clearly seen, it is not a straight line and many activities are repeated frequently.

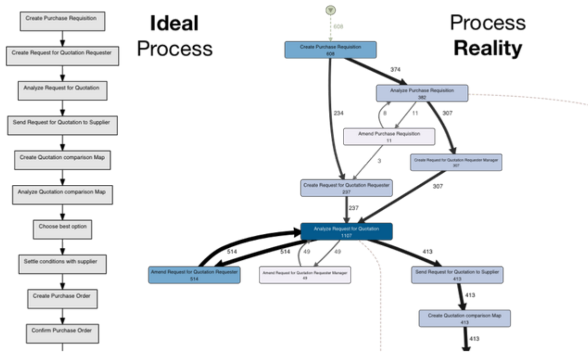


Figure 2: Ideal process vs. Process reality [18]

The use of process mining for this analysis is a natural fit. Although we are not investigating business processes, the data we have is still based on activity logs and is an excellent basis for an analysis through process mining. After all, habits are just frequently recurring behavioural patterns, just like business processes.

2.2.1 Process diagrams

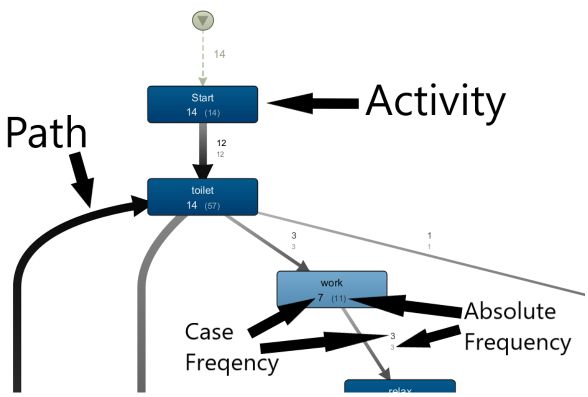


Figure 3: Explanation of a process diagram (see Appendix A for full image)

To understand some of the figures that follow and the ones above, an explanation on how to read process diagrams is necessary. The diagram depicted in Appendix A is such a process diagram. They represent the flow of activities in the dataset. Like many diagrams, process diagrams have nodes and edges as well as some form of description. The nodes are activities, the data subject recorded. The edges are called paths and will be referred to as such from here on out. These represent the time between activities. Additionally, they point from one activity to another, indicating the sequence of activities [14].

[1] https://www.oxfordlearnersdictionaries.com/definition/english/habit?q=habit

The numbers associated with the activities and paths have been chosen to represent the case frequency and absolute frequency. The larger number represents the case frequency, the smaller meaning absolute frequency. Case frequency is describing how many of the cases contain this activity or path. One case here being equal to a day of recoded activities. Absolute frequency, referring to how often the activity or path occurs in the entire dataset. Lastly, the more frequent an activity is, the darker the node responding to it. The same applies to the thickness of the path, the more frequent a path is, the thicker it is [14].

The use of graphs to model habits is not a new idea. Besides process mining, there is a similar way of visualizing and analysing habits. Kließ et al. [8] proposed feature diagrams with expanding detail and occurrence rates attached to the activity. As the tree expanded, the activities got more detailed, with occurrence rates being added.

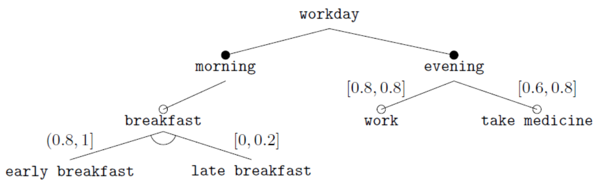


Figure 4: Probabilistic Feature Diagrams as proposed by Kließ et al. [8]

While this is a valid way of visualizing the activities in a day, it is very limited in its usability. Especially when investigating a large number of activities in a day, these representations quickly become far too confusing to read. This is mainly why we decided to utilize process diagrams.

3. METHODS

Since the goal of the research is to find habit categories and confirm the results through the analysis of already existing data, the research was undertaken in two phases. The first phase focused on a literature review. During this phase, we found a definition of a habit as well as finding categories habits could be placed in. The second phase revolved around analysing secondary data using process mining. Here we investigated whether we could confirm our findings from phase one. These phases aim to answer the following research question:

How can habits and routines be categorized?

However, for easier analysis and better understanding of the topic at hand, we decided to separate the research question (RQ) into three research questions and fitting each into the two phases described earlier. These were the resulting RQs:

RQ1: How do we distinguish a routine or a habit from a recurring action?

RQ2: Which categories of habits and routines have been distinguished in literature?

RQ3: Which examples of the categories of routines defined in RQ2 can we observe in the datasets [16]?

3.1 Phase 1: Literature research

The literature research was performed from 03.05.2021 to 04.06.2021. We used the University's library service FINDUT, Scopus and Google Scholar to find the relevant literature. The aim here was to build a foundation upon which we can analyse the data. The literature research led to a basic understanding of habits and their categories, which we could build upon.

In order to find relevant literature, the following search terms were used to search in title, abstract and keywords: "habit definition", "routine definition", "habit classification", "routine classification", "habit categorization" and "routine categorization"

All papers that were not written in English were excluded, as well as papers that were not accessible online.

When determining the relevance of the papers found through the above-mentioned methods, we first read the abstract. If the abstract seemed relevant to the research questions, we included the paper in the literature review process. If, after reading it, it did not in fact contribute to the answering of RQ1 or RQ2, we excluded it again, moving on to the next paper. This resulted in the current list of papers, which can be found in the references. Not all papers we included were fully relevant but only partially, if this was the case, we only considered the, to us relevant, parts of the paper.

In order to get a rough understanding of what routines might look like in a practical application instead of the largely theoretical understanding gained from the literature study, we decided to observe other people who already defined their routines. For this, we took to YouTube. There are many content creators having made a video on the topic of their own, personal morning routines.

3.2 Phase 2: Process mining

The process mining and thus data analysis part of the paper has been performed from 07.06.2021 to 18.06.2021. Here we confirmed some of our findings from the first phase. We were fortunate enough to not have to modify the data much for it to work with our chosen tool, so the only preparation we had to do before we could start analysing the data was to import it into the tool and learn how to use it. We used the knowledge gained through the literature study to search for noticeable patterns in the process diagrams.

3.2.1 The Data

The datasets [16] used to conduct this research was created in October 2010 by Sztyler, T. and Carmona, J. and published by the University of Mannheim. It contains activity logs of three individuals that detail their daily activities.

The data used consists of activity logs of three individuals over the course of several weeks. The data of each data subject has already been transformed into variations, where each variation consists of one day of activities. Additionally, the recordings of each individual have been separated into sets of labour days and weekend days. The data can therefore be viewed as individual days, or as a whole. For data to be usable in process mining, each event requires three attributes: a case ID, an activity and a timestamp. All of these apply to our data.

The aforementioned datasets are the following:

edited_hh_102_labour and edited_hh_102_weekends

edited_hh_104_labour and edited_hh_104_weekends

edited_hh_110_labour and edited_hh_110_weekends

When referring to them in other sections of the paper, we will leave the "edited_" away for simplicity.

3.2.2 Tools

The two most frequently used tools for process mining are Fluxicon Disco[2] and ProM[3]. Fluxicon Disco is a commercial tool with fewer functionalities than ProM, whereas ProM is an open source tool which can be expanded upon using plugins.

For the analysis, we used Fluxicon Disco. The data was created with this process mining tool, as can be read in the README

[2] https://fluxicon.com/disco/

[3] http://www.promtools.org/doku.php

file included with the data, so it is natural to analyse the data with this tool as well. Fluxicon Disco has several tutorials and a user guide, which we used to learn how to use the tool. Additionally, we used Wil van der Aalst's book *Process Mining: Data Science in Action* and his publicly available lectures to learn the basics of process mining.

Fluxion Disco allows us to filter the data as well as unclutter the diagrams through sliders controlling the complexity. The sliders are especially helpful as their purpose is to transform the diagrams into a more readable format. Depending on the chosen metrics, the sliders change the level of detail displayed in the diagram. For our purpose, all the activities have been selected to be shown, but only the most frequent paths will be included in the diagrams, as only those are interesting to us. Additionally, Disco has the ability to animate the processes, which allows us to observe the proceedings of each day as they took place. For this, a ball symbolizing one variation, which is equal to one day in our data, moves along the edges and nodes of the diagram from the Start node to the End node.

3.2.3 The process of process mining

In order to find patterns in the data, we used the filters and detail sliders. We mainly used the filters for "attribute" and "follower". The attribute filter lets us view only selected activities and its repetitions form the dataset. This was especially useful when searching for temporal and duration habits, as we can focus on the selected activities and their properties. The follower filter was used to investigate stable systems and stable temporal systems. This filter allows us to view only those cases that have a specific sequence of activities in them. Apart from using those filters, Fluxicon Discos ability to animate the process was quite helpful as well. We were able to use it to detect habits, as they come up as the same activity or sequence of activities frequently.

For one habit category we used a different approach as we were sure to find examples of them in the datasets [16] but in fact were not able to find any. The first step was to investigate the consistency of an activity. If the consistency was higher than 80%, we examined the mean and median durations as well as the minimum and maximum durations. These metrics can be shown in Fluxicon Disco. For a duration to be deemed consistent, we decided to focus on mean and median durations which, if close together, suggest a normal distribution. Afterwards, we exported the data into a CSV file to calculate standard deviation. If the standard deviation was larger than 30% of the mean duration, we deemed the activity unfit for use as an example of a duration habit.

4. RESULTS

As already outlined in the methodology section, the research questions have two distinct phases. A literature phase focusing on RQ1 and RQ2 and a data analysis phase using process mining to answer RQ3. Sections 4.1 and 4.2 will cover RQ1 and RQ2 respectively, thus encompassing the results from literature. 4.3 will be answering RQ3 and show examples of the habits categories from section 4.2.

4.1 RQ1: Difference between routines or habit and recurring action

The rough definition of habit and routine have already been presented in the introduction. The most important distinction is that a habit is a routine with a stimulus and a reward added.

In this paragraph, we want to focus on the distinction between a habit or routine and some repeated behaviour. As already mentioned when we first covered the definition of a habit, a good indicator is a daily repetition of the same behaviour. However, this level of frequency and consistency is unrealistic when laying the focus on such large activities as we are. On a smaller scale, this does apply, however with the larger habits, as the ones we are observing in the datasets [16], the daily level of frequency and consistency is rarely observable.

Frequency and consistency will be our main measures for identifying habits. In order to further explore the concepts of frequency and consistency, we need to distinguish between micro and macro habits. Micro habits are, as the name suggests, small habits. These are the habits small enough that daily occurrence is normal. A great example of a micro habit is the morning coffee or tea many people have. It takes less than two minutes from start to finish and is, unless for some unforeseen circumstance, prepared every day. Micro-habits are not represented in the datasets [16]. Macro-habits on the other hand are. These are behavioural patterns that take more time to execute and have multiple steps to them. Due to their long nature, they are also less frequent and might not happen every day.

Frequency and consistency both describe how often a particular activity or chain of activities occur. Frequency, relating to how often an activity or chain of activities occurs in the dataset, akin to the absolute frequency described earlier. Consistency, describing in how many of the recorded days the activity or chain of activities occurs, like case frequency.

When examining the process diagrams, it became immediately apparent that some daily activities are repeating. As outlined previously, however, not all repeated behaviour is a habit or routine. Therefore, a threshold would be needed to distinguish habits and routines from loosely repeated activities. We have therefore worked out two distinct levels of frequency and consistency. There is the routine and the rule of thumb. The threshold for a routine was set at 80% consistency. As previously mentioned, we could not reasonably assume that the larger routines we can observe in the datasets [16] are executed every day, or with 100% consistency. However, we have found, that at 80% consistency, these routines still occur often enough to be considered routine behaviour. In most of the cases, 80% consistency describes only a few days of the behaviour not occurring. Kließ et al. [8] proposed to use compliance as a measurement of consistency. These two terms can be used interchangeably, as they refer to the same concept. Additionally, They suggested asking the user of the behaviour support agent how compliant he should be before intervention from the agent. We can safely assume that at a level of consistency above 80%, the activity has either been habituated or intentionally become customary in the data subject's activities. Unfortunately, we have no way of distinguishing these two, so we have to treat them as the same.

Secondly, we have the rule of thumb. A rule of thumb is still a form of routine, but looser defined. They are decision rules that yield a "first approximation" [10] of an automated process and still require "low levels of information processing" [10]. The lower bound of consistency for a rule of thumb was set at 70%. This consistency threshold is purposefully set lower than the threshold for a routine. Because of their looser definition, rules of thumb are less rigid in nature, therefore more difficult to identify. Rules of thumb are not a routine yet, but rather guidelines for behaviour. They might transform into a routine when performed often enough to become unconscious behaviour.

4.2 RQ2: Routine and Habit categories from the literature

4.2.1 Categories from literature

Macro-habits, as explored in the previous section, can be broken down even further. Through our literature review, many subcategories of Macro-Habits have been identified.

Firstly, goal-directed habitual behaviour [1]. Although habits have a distinct stimulus leading to the execution of the routine, the habit consists of, they also have a reward [15, 17]. This reward usually serves as the reinforcing agent, telling the person executing the habit that what they are doing is worth repeating [15, 17]. However, habits can also form out of a desire to change, a goal. In his book Tiny Habits, BJ Fogg [5] describes how these goal related changes can be realized. He points out that small changes in behaviour combined with already existing habits and constant positive reinforcement can help produce large changes in the future. Fogg describes that these changes grow over time and begin forming habits, with the old behaviour serving as the stimulus for the new habit [9]. A *goal-directed habit* can also arise from the search for a set of rules. These rules can then be applied to daily behaviour and eventually lead to the adoption of a new habit.

Systematic habits are similar in execution, but differ in motivation. In Atomic Habits, James Clear [2] describes a very similar system to change one's habits as BJ Fogg does in Tiny Habits [5]. Both focus on the gradual change of behaviour, and the core statement relies on a slow but steady improvement. Clear describes this as "1% better every day" (Atomic habits, p.18). Over one year this would lead to an increase of 3678%. The distinction between Fogg's and Clear's method is that Clear relies on changing the system of learning and intentionally not setting goals for oneself.

The two habit categories mentioned above mainly focus on intended change, but as we explored earlier, habits are mostly unintentional. Since the two methods proposed by Fogg [5] and Clear [2] are so similar in execution, they could be summed up into *conscious habits* or *habit building exercises*. Similarly, due to the focus on habit formation, in the early stages, they fall under rules of thumb. As they evolve into habits, they move out of that category with higher frequency, and consistency.

Another category which, depending on the stage of observation, could be categorized as a conscious habit is a *Skilled learning habit*. Skilled learning is the intentional development of one's abilities [17]. Similar to goal-directed habits, the intention to change one's behaviour is key to identifying this category. However, instead of starting small and improving gradually, a foundation of the skill is already established. T.W. Robbins and Rui M. Costa argue that, through extensive training, the routine sequence embedded in the habit might be even stronger. By intentionally performing the motor sequence related to the skill, one reinforces the habit and the routine. A great example of Skilled learning is playing an instrument like the piano. The more the player practices pressing the keys in the right order at the right time, the better they get at it, proving the reinforcing tendencies of repeating the activity [17]. Additionally, with the added training, cognitive stress is lowered, and the piano player finds it easier to play the song or composition [6].

Developed habits often come from circumstance or environment, as the main purpose of a habit is to automate activities and relieve the individual of cognitive stress related to the execution of the activity [6].

In Requirements for a Temporal Logic of Daily Activities for Supportive Technology, Malte S. Kließ and M. Birna van Riemsdijk define five key temporal dimensions for behaviour support agents. The proposed dimensions are:

1. **Clocktime**. A behaviour support agent, can track deadlines and interpret the time of day. Clocktime can also be used as a measure of consistency by measuring at what time and how consistent an activity is performed at that time.
2. **Ordering**, refers to the order in which the activities are performed
3. **Coherence**, referring to an ordered activity being performed in a similar (coherent) way every time it is performed
4. **Duration**, referees to the duration of this activity it is roughly the same each time it is performed
5. **Repetition** is used to deal with regularly scheduled activities and the inevitable repetition due to the nature of habits.

Each of these dimensions separately can also double as a habit category. Clocktime can be used to categorize behaviour that repeats at a specific time of day. We will refer to habits of the Clocktime dimension as *Temporal habits*. *Duration habits* are relatively self-explaining. An activity that has roughly the same duration every time it is performed, will be referred to as a *Duration habit*. Ordering and Coherence will be grouped as *Stable Systems*. These describe a sequence of activities being executed without interruption. Sequence, hereby referring to two or more activities directly following each other uninterrupted. Lastly, Repetition is a core concept of habits, as, by definition, behaviour becomes a habit with frequent repetition.

4.2.2 Categories from theorization

Stable Temporal systems are, as the name suggests, a combination of stable systems and temporal habits. Chains of activities triggered by the colloquial inner clock. They are executed at roughly the same time of day, each time they are executed.

4.3 RQ3: observed Examples of the categories

In the following sections, we will present the examples we have been able to find of the aforementioned habit categories. We will be answering RQ3 in this section.

4.3.1 Temporal habits

Temporal habits are very common in the datasets [16] and have many examples. Often these are activities we do on a daily basis, so the habit has become strong enough to use the proverbial inner clock as a stimulus. Some examples might include an evening walk or meal preparation when not hungry, but anticipating hunger, around lunchtime.

In the datasets [16], we are unfortunately bound by the limitation the activity names laid upon us.

In hh_102_labour, the temporal habit takes the form of a nightly walk to the toilet. Every night between 2:40 and 4:40 the data subject has to interrupt their sleep to relieve themselves. Interestingly, this habit does not carry over into the weekend version of this data subject's recordings. Here, the nightly walk to the toilet is still very prevalent, but not consistent to any specific time and generally much later.

Similarly, in hh_104_weekends the data subject always works in the evening, past 18:00.

Hh_110_labour and hh_110_weekend also have examples of temporal habits. In both datasets, the subject takes medication every morning (around 9:00), afternoon (around 14:00), and night (around 21:00) in 85.71% of cases. Additionally, in hh_110_labour, the activity of outside occurs twice a day with a consistency of 80.9%. The first relatively reliably around 12:00 with a consistency of 90.4%, with the second time being between 17:00 and 19:00 with a consistency of 80.9%.

4.3.2 Stable systems

In Figure 5 we can observe two almost perfect representations of a stable system. In 2.1.1 we have covered how to read process diagrams, thus we will not go into the details of this process diagram. Hh_110_ weekend has a total of 6 recorded cases. All the activities in the excerpt from the process diagram representing hh_110_weekend have a consistency of 83.3% or higher.

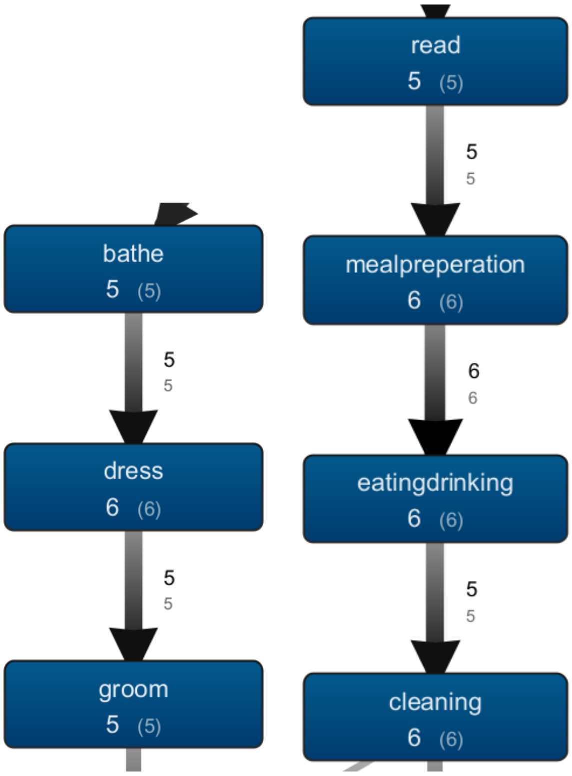


Figure 5: hh_110_weekend stable systems

Furthermore, there are two examples of stable systems in hh_102_labour. With 83.33% consistency, there is a direct transition from personalhygiene to sleep. Additionally, with the same consistency, there is a direct transition from relax to personalhygiene. While it might seem logical, that these three activities form a stable system as well, this is not the case, as they do not pass the threshold for a routine we set earlier.

Moreover, in hh_104_weekends (see Figure 6), with a consistency of 94%, the activities mealpreperation and eatingdrinking form a stable system. While this might be expected, it is still a notable example of a stable system, especially with such a high consistency.

4.3.3 Stable Temporal systems

The stable system depicted on the left-hand side in Figure 4 is also a very good example of a stable temporal system. This stable system begins with 100% consistency between 7:46 and 8:48.

Figure 6: hh_104_weekend stable system

4.3.4 Rule of thumb

Rules of thumb are a less consistent form of the other habit categories presented earlier, So they can fall into those categories but might be more loosely formed.

One great example of this is the stable temporal system, that can be found in hh_102_labour as well. With a 72.22% consistency is the stable temporal system consisting of the activities personalhygiene and sleep as described in the stable systems section performed between 20:43 and 21:39.

4.4 Findings we could not confirm

We were unfortunately unable to find examples of goal-directed habits, systematic habits and skilled learning. The reason for this mainly being that in the datasets [16], no intentions are recorded. We have tried to search for increasing levels of frequency and consistency as these are indicators for improvement in the habit formation or improvement process, but were unable to find such examples.

Additionally, we have not been able to find examples of micro-habits due to the datasets [16] focusing on activities much larger than the aforementioned habit categories.

We have also not been able to find duration habits. We can not quite explain why this is the case, but none of the activities were consistent enough in duration, so that they could have been considered a habit. Investigating the durations of habits was much easier than other habit types, although having to transform the data. We explained this process in detail in section 3.2.3. One might want to argue that we did not take such an extensive approach to other habits types, as we for example gave a range of time the habit would be able to fall into. This was a conscious decision, as the duration of an automated sequence or habit should be rather rigid. A habit is so automated that the execution is entirely unconscious, thus leading to a rigid duration and our decision to set the tolerance comparatively low.

5. LIMITATIONS

As previously mentioned, there were certain restrictions put upon our research due to the datasets [16]. It is very likely that not all activities performed by the data subjects have been recorded. We believe, that some activities might be interpreted as indecent, when recorded in a datasets [16]. Similarly, there might be activities, the data subjects are ashamed of or do not feel comfortable when recorded. Furthermore, the datasets [16] lays its focus on a rather large scope, when considering the activities present in it. Some tasks are simply too small or insignificant to record them efficiently, or might slip under the radar of the person recording the activities.

6. CONCLUSION & FUTURE WORK

Our research focused on finding a definition for habits or routines, then finding categories for these habits or routines, and lastly confirming our findings through the use of process mining on daily activity data.

The first research question focussed on the definition of habits and routines and the difference to recurring behaviour. In order to identify a habit, three main requirements need to be fulfilled. A habit needs to be performed often, without thinking, and have a stimulus and reward [17,8,10]. Routines on the other hand only have the first two requirements. They are often embedded within a habit, where the addition of a stimulus and reward transforms the routine into a habit [17].

Through our further literature research, we were able to answer RQ2 and find the following habit categories. The goal-directed habit, is a habit not controlled by the stimulus, but by the person's desired outcome. They are defined through their intentions and a gradual increase in consistency. Similarly, Systematic habits also lean on the gradual increase of one's abilities, but are intentionally not goal-related. Instead, the system of learning and improving it is the main focus. A skilled learning habit also focuses on the intentional improvement of a person’s abilities, but does not start at a low level. Skilled learning is used to improve on abilities one already has. A great example is practising a song to play on the piano. The longer a person intentionally practices the keystrokes, the less he has to think about the actual movement, thus reinforcing the habitual behaviour. These three habits types focus on intentional change, but this is not the main focus of a habit. That would be the automation of frequently performed activities. Temporal habits, Duration habits and Stable Systems are such habits types. Temporal habits have a consistent start time, duration habits a consistent duration. Stable systems are a sequence of two or more activities directly following each other with high consistency. Stable temporal systems are a combination of stable systems and temporal habits. They are a consistent sequence of activities starting at a consistent time.

Consistency is one of the main traits of a habit, but there is also a habit type, with lower consistency. The rule of thumb is a loosely defined habit, that is less consistent.

We have been able to find examples of the Temporal habits, Stable Systems, Stable temporal systems and Rule of thumb, which you can find in the results section. Due to limitations placed upon us by the datasets [16], we were not able to confirm any of the goal-related or systematic habits. Furthermore, duration habits have also not been confirmed, since we did not find examples of them in the datasets [16].

In future work, it would be beneficial to look at data that is more detailed. One example of this would be the activity of outdoors. What does outdoors mean? The most logical conclusion in our opinion is that it means to take a walk, but we can not be entirely sure about this. Outdoors could also describe working in the garden or be various activities put together. This is also where the next limitation of these datasets [16] lies. Some activities were summed up into one. Personal grooming, cleaning, meal preparation or relax are all activities that could be explored in more detail as a lot of habits and routines lie at a much smaller and deeper level than what we can reasonably find in these datasets [16]. Additionally, the use of activity names that are very similar, or could describe the same activity, is hindering the analysability of the data. The activities bathe, groom and personalhygiene could potentially mean the same thing. They all describe the act of cleaning oneself. Similarly, relaxing can also be synonymous with the activities watchTV and reading, as they are also forms of relaxation. Because of those reasons, creating a more detailed dataset and using our analysis method on it would yield much better results.

Lastly, using manual analysis of the data can only uncover so much. A proper analysis using machine learning and examining whether a machine can reliably identify habits would not just benefit the behaviour support agent project, but also uncover knowledge we were unable to find.

7. REFERENCES

[1] Balleine BW, O'Doherty JP. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology. 2010;35(1):48-69.

[2] Clear, M. (2021). *Atomic habits: let's change your atomic habits!: a full simple guide to break your bad routines and learn new good ones*. CreateSpace.

[3] Eve M. Wolf, Janis H. Crowther, Personality and eating habit variables as predictors of severity of binge eating and weight, Addictive Behaviors, Volume 8, Issue 4, 1983, Pages 335-344, ISSN 0306-4603, https://doi.org/10.1016/0306-4603(83)90033-3. (https://www.sciencedirect.com/science/article/pii/0306460383900333)

[4] Filippo Maria Bianchi, Antonello Rizzi, Alireza Sadeghian, Corrado Moiso, Identifying user habits through data mining on call data records, Engineering Applications of Artificial Intelligence, Volume 54, 2016, Pages 49-61, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2016.05.007

[5] Fogg, B. J. (2020). *Tiny habits: + the small changes that change everything*. Mariner Books.

[6] Jager, W. (2003) Breaking 'bad habits': a dynamical perspective on habit formation and change. in: L. Hendrickx, W. Jager, L. Steg, (Eds.) Human Decision Making and Environmental Perception. Understanding and Assisting Human Decision Making in Real-life Settings. Liber Amicorum for Charles Vlek. Groningen: University of Groningen

[7] Kließ, M. S., & van Riemsdijk, M. B. (2017). Requirements for a Temporal Logic of Daily Activities for Supportive Technology. In *CARe-MAS@ PRIMA* (pp. 43-51).

[8] Kließ, M. S., Stoelinga, M., & van Riemsdijk, M. B. (2019). From Good Intentions to Behaviour Change: Probabilistic Feature Diagrams for Behaviour Support Agents. In M. Baldoni, M. Dastani, B. Liao, Y. Sakurai, & R. Zalila-Wenkstern (Eds.), *PRIMA 2019: Principles and Practice of Multi-Agent Systems: 22nd International Conference, Turin, Italy, October 28-31, 2019, Proceedings* (pp. 354-369). (Lecture Notes in Computer Science; Vol. 11873). Springer. https://doi.org/10.1007/978-3-030-33792-6_22

[9] M. B. van Riemsdijk, C. M. Jonker, and V. Lesser. Creating socially adaptive electronic partners: Interaction, reasoning and ethical challenges. In Proceedings of the fourteenth international joint conference on autonomous agents and multi agent systems (AAMAS'15), pages 1201–1206. IFAAMAS, 2015.

[10] MICHAEL D. COHEN, ROGER BURKHART, GIOVANNI DOSI, MASSIMO EGIDI, LUIGI MARENGO, MASSIMO WARGLIEN, SIDNEY WINTER, Routines and Other Recurring Action Patterns of Organizations: Contemporary Research Issues, *Industrial and Corporate Change*, Volume 5, Issue 3, 1996, Pages 653–698, https://doi.org/10.1093/icc/5.3.653

[11] Oinas-Kukkonen, H. Behavior Change Support Systems: A Research Model and Agenda. In Persuasive Technology: 5th International Conference, PERSUASIVE 2010, Copenhagen, Denmark, June 7-10, 2010. Proceedings (2010), T. Ploug, P. Hasle, and H. Oinas-Kukkonen, Eds., Springer, pp. 4–14.

[12] P., van der A. W. M. (2018). *Process Mining Data Science in Action*. Springer Berlin.

[13] Pasotti, P., van Riemsdijk, M. B., and Jonker, C. M. Representing human habits: Towards a habit support agent. In Proceedings of the 10th International Workshop on Normative Multiagent Systems (NorMAS'16)(2016), LNCS, Springer

[14] Rozinat, A. 2021. Managing Complexity in Process Mining Part I: Quick Simplification Methods — Flux Capacitor. *Fluxicon.com*. https://fluxicon.com/blog/2015/03/managing-complexity-in-process-mining-quick-simplification-methods/

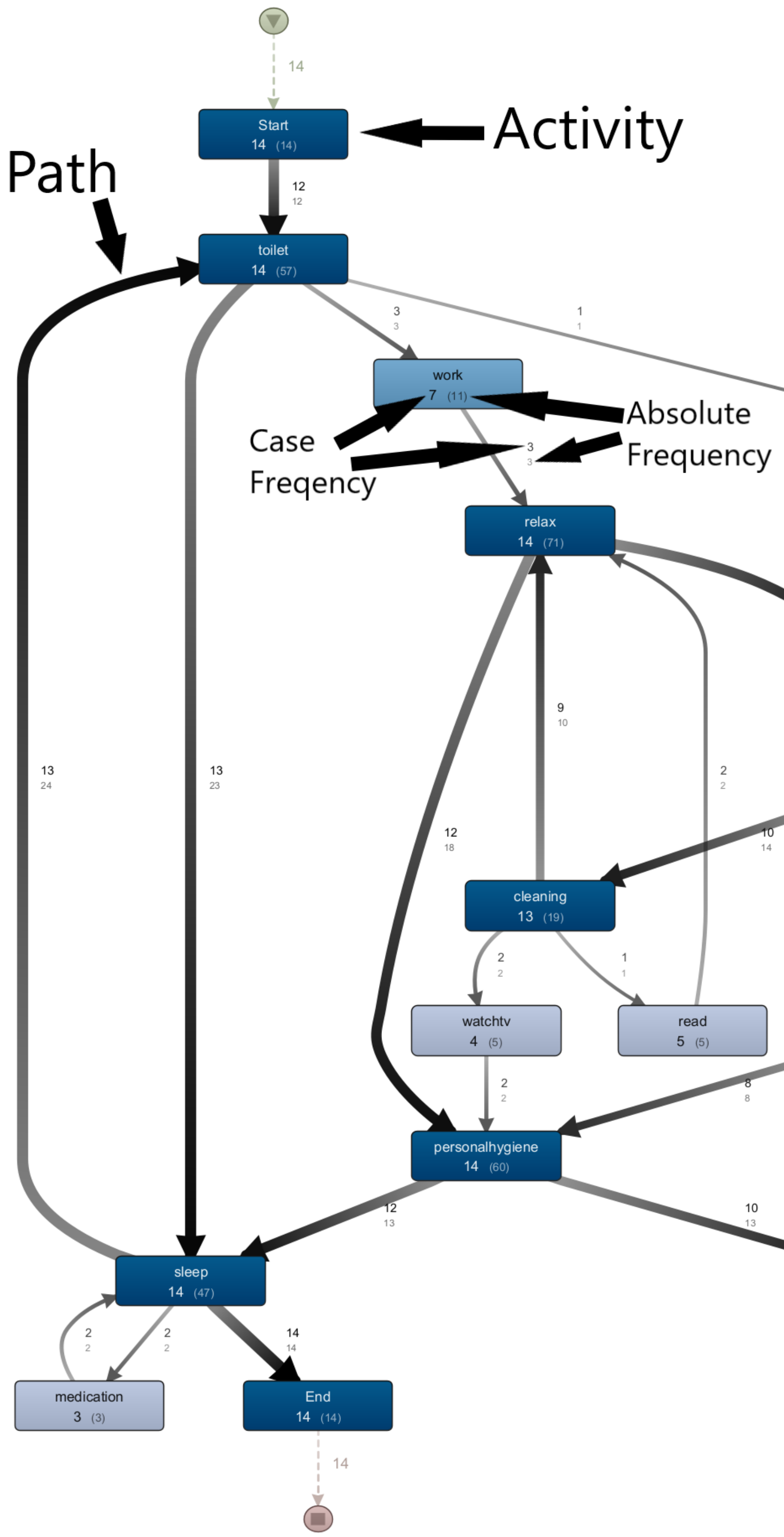
[15] Smith, K. S., & Graybiel, A. M. (2016). Habit formation. *Dialogues in Clinical Neuroscience*, *18*(1), 33–43. https://doi.org/10.31887/dcns.2016.18.1/ksmith

[16] Sztyler, Timo; Carmona, J. (Josep) (2015): Activities of daily living of several individuals. 4TU. ResearchData. Dataset. https://doi.org/10.4121/uuid:01eaba9f-d3ed-4e04-9945-b8b302764176

[17] T.W. Robbins, Rui M. Costa, Habits, Current Biology, Volume 27, Issue 22, 2017, Pages R1200-R1206, ISSN 0960-9822, https://doi.org/10.1016/j.cub.2017.09.060. (https://www.sciencedirect.com/science/article/pii/S0960982217312587)

[18] What Is Process Mining? — Process Mining Book 2.5. 2021. *Fluxicon.com*. https://fluxicon.com/book/read/intro

APPENDIX A



Activity
Path
Case
Freqency
Absolute
Frequency
Start
14 (14)
toilet
14 (57)
work
7 (11)
bathe
6 (6)
relax
14 (71)
mealpreperation
14 (30)
eatingdrinking
14 (27)
cleaning
13 (19)
watchtv
4 (5)
read
5 (5)
dress
10 (11)
groom
8 (9)
personalhygiene
14 (60)
outdoors
13 (32)
sleep
14 (47)
medication
3 (3)
End
14 (14)
snack
13 (33)