Gebruikersdictionary: een woordenboek vol bruikbare gebruikerssporen?

AFSTUDEERVERSLAG

Naam student	Nabila Agni
Studentnummer	S1114455
Stageperiode	07/02/2022 - 14/07/2022
Afstudeerorganisatie	Nederlands Forensisch Instituut (NFI)
Onderwijsinstelling	Hogeschool Leiden – opleiding Informatica, specialisatie Forensisch ICT (FICT)

Versienummer	1.0
Versietype	Definitieve versie
Versiedatum	14/06/2022

I. Voorwoord

Voor u ligt het afstudeerverslag “Gebruikersdictionary: een woordenboek vol bruikbare gebruikerssporen?”. Het onderzoek voor dit afstudeerverslag is uitgevoerd bij het Nederlands Forensisch Instituut en ook in opdracht van het Nederlands Forensisch Instituut. Ik heb mij de afgelopen vijf maanden bezig gehouden met dit onderzoek. Het afstudeerverslag is geschreven in kader van het afstuderen aan de opleiding Forensisch-ICT aan de Hogeschool Leiden.

De onderzoeksvraag van mijn onderzoek is samen met mijn bedrijfsbegeleider, Ruud Schamp, opgesteld en luidt als volgt:

“Op welke manier kan er aan de hand van de inhoud van de gebruikerswoordenboek bepaald worden of een bericht dat op een toestel staat geschreven is door de gebruiker of dat het bericht op een andere manier op het toestel is gekomen?”

Tijdens het onderzoek heb ik veel nieuwe kennis opgedaan. Ik heb veel collega’s gesproken en kennis van hen opgestoken. Ik heb veel aan de input van mijn collega’s gehad en daar wil ik ze graag voor bedanken.

Ik wil beginnen met het bedanken van mijn afstudeerdocent, Jurrien Bijhold, en mijn bedrijfsbegeleider Ruud Schamp voor hun betrokkenheid, getoonde interesse en begeleiding. Ik heb mijn afstudeer docent misschien niet vaak gesproken maar ik vond elk gesprek dat ik met hem heb gehad nuttig. Mijn bedrijfsbegeleider heb ik wat vaker gesproken. Ik heb veel nieuwe kennis van hem opgestoken en heb gezien dat er ook nog zoveel meer te leren is. Ook elk gesprek dat ik met hem heb gehad was erg nuttig. Ook al was hij soms erg druk, was hij toch erg betrokken en kon ik hem altijd bereiken of zijn werkkamer binnenstormen wanneer dat nodig was. En nee Ruud, je bent geen boomer. Geen zorgen.

Ik wil ook alle collega’s die hun kennis met mij hebben gedeeld bedanken voor hun tijd en moeite. Het is fijn dat iedereen zo behulpzaam is. Ik zie jullie ook als collega’s, medestagiaires. Jullie wil ik ook bedanken voor de gezelligheid op de kamer en in de pauzes. Het is fijn om te zien hoe iedereen zo geïnteresseerd is in elkaars opdracht en altijd klaar staat om een ander te helpen.

Tot slot wil ik mijn familie en vrienden bedanken voor hun steun, hulp en getoonde interesse. Ik heb een aantal van jullie best wel wat werk gegeven om na te lezen en jullie hebben dat aandachtig gedaan en fijne en behulpzame feedback geleverd. Bedankt.

Nabila Agni

Veghel, 14 juni 2022

II. Documentinformatie

Versienummer	Versiedatum	Samenvatting aanpassing	Documentversie
0.1	08-02-2022	Document aangemaakt.	Conceptversie
0.2	21-04-2022	Document gedeeld met een medestudent, mijn bedrijfsbegeleider en afstudeerbegeleider voor feedback.	Conceptversie
0.3	29-05-2022	Document gedeeld met mijn bedrijfsbegeleider en afstudeerbegeleider voor feedback.	Conceptversie
1.0	14-06-2022	Definitieve versie. Gedeeld met mijn bedrijfsbegeleider en ingeleverd in het afstudeerdossier	Definitieve versie

III. Samenvatting

In dit verslag bespreek ik mijn afstudeerproces en reflecteer ik op mijn afstudeerproces, het product en de competenties. Mijn afstudeerorganisatie is het Nederlands Forensisch Instituut (NFI). Ik heb mijn afstudeeropdracht uitgevoerd binnen de divisie Digitale en Biometrische Sporen (DBS), op de afdeling Digitale Technologie (DT), binnen het subteam Data analyse. Tijdens de uitvoering van mijn opdracht heb ik mij beziggehouden met het gebruikerswoordenboek op een Samsung toestel met Androidversie 11.0.

Mijn afstudeeropdracht is ontstaan omdat verdachten in strafrechtelijke onderzoeken vaak beweren niet de auteur te zijn van een bericht dat op zijn/haar toestel is aangetroffen. Met deze opdracht gaan we proberen aan te tonen hoe een bericht op een toestel terecht is gekomen. Met behulp van de gebruikersdictionary en woordvolgorden op een toestel zou er aangetoond kunnen worden of een bericht op het toestel zelf is getypt of niet. Een collega wilde dit al een tijdje gaan onderzoeken maar heeft hier nog niet de tijd voor gehad. Het is dus geen zelf bedachte opdracht.

Het is de bedoeling dat mijn woordenboekparser, een stukje code dat de nodige data zal verzamelen, die ik heb gebouwd geïmplementeerd zal worden in Hansken. Hier zijn collega's werkzaam die zich bezighouden met de ontwikkeling van forensische tool Hanksen. Mijn onderzoek heeft een bijdrage aan de onafhankelijke waarheidsbevinding van het NFI. Een bericht kan op de volgende manieren op een toestel zijn gekomen:

- Door het bericht te typen.
- Door het op een ander apparaat te typen en naar het toestel te synchroniseren (Whatsapp Web).
- Door het bericht door te sturen.
- Door het bericht te kopiëren.

Na het uitgevoerd te hebben van diverse experimenten is er achterhaald dat de toetsenbord data terug te vinden is het in een bestand genaamd dynamic.lm. Door dit bestand te openen met een Hex Editor, kan het bestand opgedeeld worden in zes delen. In het eerste deel is de file header signature (flue) samen met een stukje tekst over dat het bestand is gecreëerd in Dynamic Ngam Term Model terug te vinden. Een N-gram model is een type Language Model dat gaat over het inschatten van kansen bij woordreeksen. In het tweede deel is een gecomprimeerd stuk te zien. De hoge entropie wijst op een compressie. Het is niet gelukt om te achterhalen welk compressie algoritme er is gebruikt. De woorden uit de gebruikersdictionary zijn hoogstwaarschijnlijk op deze plek vastgelegd. Hiernaast is het totaal aantal woorden en zijn het aantal karakters waaruit elk woord in de dictionary bestaat in dit deel terug te vinden. De betekenis van het derde en vierde deel zijn nog onbekend. In het vijfde deel zijn woordreeksen terug te vinden. De reeksen staan opgeslagen in een Trie structuur. In het zesde deel is de end-of-file marker terug te vinden.

Er is geen statistisch model opgesteld omdat hier twee databases voor nodig zijn. Om een uitspraak te doen over de betrouwbaarheid van de parser is er op zijn minst één referentie dataset nodig. Er kan betekenis aan de uitkomst van een berekening worden gegeven wanneer deze vergeleken kan worden met een soortgelijke waarde. Omdat het bepalen of een bericht wel of niet door een gebruiker getypt kan zijn een vereiste was van de parser heb ik een aantal vervolgonderzoeken beschreven waar dit op een nauwkeurigere manier mee gedaan kan worden.

Op basis van mijn bevindingen en de werking van mijn woordenboekparser heb ik de volgende adviezen opgesteld:

1. De parser kan gebruikt worden om woordreeksen die opgeslagen staan in de toetsenborddata van een toestel in kaart te brengen.
2. De parser kan gebruikt worden om de woorden uit een bericht te vergelijken met de woorden die in de gebruikersdictionary staan.
    a. Uit het percentage overeenkomende woorden kan geen geldige conclusie worden getrokken omdat ik heb waargenomen dat niet alle woorden die ik heb getypt opgeslagen zijn in de gebruikersdictionary en omdat er nog geen rekening wordt gehouden met de woordreeksen.
3. Het uitvoeren van vervolgonderzoeken, bijvoorbeeld:
    a. Achterhalen of de toetsenborddata op andere Samsung toestellen op dezelfde manier opgeslagen wordt.
    b. Reverse engineering toepassen voor het achterhalen van een compressie algoritme.
    c. Met behulp van twee datasets (een dataset van verdachte 1 en een dataset van verdachte 2) een statistisch model opzetten om te bepalen voor welke verdachte het waarschijnlijker is geweest dat diegene het bericht heeft getypt.
    d. Met behulp van de woordreeksen de op een toestel getypte zinnen reproduceren om te achterhalen welke zinnen er op een toestel zijn getypt.

Adviezen zullen mondeling gedeeld worden tijdens een presentatie. Er is op dit moment nog geen datum gepland voor het plaatsvinden van de presentatie, maar dit zal binnenkort gebeuren.

Inhoudsopgave

1. Inleiding

In dit verslag bespreek ik mijn afstudeerproces en reflecteer ik op mijn afstudeerproces, het product en de competenties. Mijn afstudeerorganisatie is het Nederlands Forensisch Instituut (NFI). Ik heb mijn afstudeeropdracht uitgevoerd binnen de divisie Digitale en Biometrische Sporen (DBS), op de afdeling Digitale Technologie (DT), binnen het subteam Data analyse. Tijdens de uitvoering van mijn opdracht heb ik mij beziggehouden met het gebruikerswoordenboek op een Samsung toestel met Androidversie 11.0.

In de subhoofdstukken van dit hoofdstuk is informatie over het NFI, mijn positie binnen het NFI, de afstudeeropdracht en de stakeholders van mijn opdracht terug te vinden. In *Hoofdstuk 2. Het proces* is er een samenvatting van mijn Plan van aanpak, zijn mijn uitgevoerde onderzoeken en experimenten terug te vinden en is de aansluiting van mijn uitgevoerde werkzaamheden op de competenties beschreven. De onderzoeken en experimenten zijn opgesplitst in zeven subhoofdstukken. De competenties zijn ook opgesplitst, zodat er per competentie beschreven kon worden wat ik heb gedaan om deze competentie aan te tonen. In *Hoofstuk 3. Reflectie* zijn mijn reflecties terug te vinden. Er is gereflecteerd op vijf onderdelen. Elk onderdeel staat los beschreven. De reflectie is ook meteen het nawoord.

1.1 Het Nederlands Forensisch Instituut (NFI)

Het NFI is een verzelfstandigde organisatie (agentschap) van het ministerie van Justitie en Veiligheid (JenV) en valt onder het Directoraat-Generaal Rechtshandhaving (NFI, 2021) (Rijksoverheid, 2021). Het NFI levert forensische producten en diensten, met als doel om met onafhankelijk forensisch onderzoek de waarheidsbevinding in een strafrechtelijk onderzoek te bevorderen. Daarmee is het NFI een belangrijke schakel in de strafrechtketen. Ze voorzien nationale en internationale organisaties en ketenpartners die zich inzetten voor vrede, recht en veiligheid van betrouwbare informatie uit bronnen. De ketenpartners zijn onder andere de politie, het Openbaar Ministerie (OM) en de Zittende Magistratuur (ZM) (NFI, 2021).

Agentschappen zijn officieel onderdeel van een ministerie en leveren daardoor tegenover een betaling producten en/of diensten aan organisaties binnen het Rijk. Agentschappen werken volgens een 'resultaatgericht sturingsmodel'. Binnen dit model werken drie partijen samen: de opdrachtgever, de opdrachtnemer en de eigenaar. Deze drie partijen maken vooraf afspraken over de prestaties, kwaliteit, kosten en risico's. Agentschappen moeten kun kosten dekken met de inkomsten die zij halen uit het leveren van producten en/of diensten aan de ministeries (Rijksoverheid, 2022). Het NFI wordt dus gefinancierd door het ministerie van Justitie en Veiligheid.

Het NFI streeft ernaar om de meest innovatieve en klantgerichte leverancier van forensische producten en diensten te zijn. Dit doen ze onder andere door te investeren in kennis en innovatie. Hierdoor kan het NFI inspelen op actuele maatschappelijke, technologische en wetenschappelijke ontwikkelingen. Internationaal gezien loopt het NFI hiermee voorop. Hierdoor staat het NFI internationaal hoog aangeschreven (NFI, 2021).

Binnen het NFI zijn er vier verschillende divisies die (sporen)onderzoek verrichten:

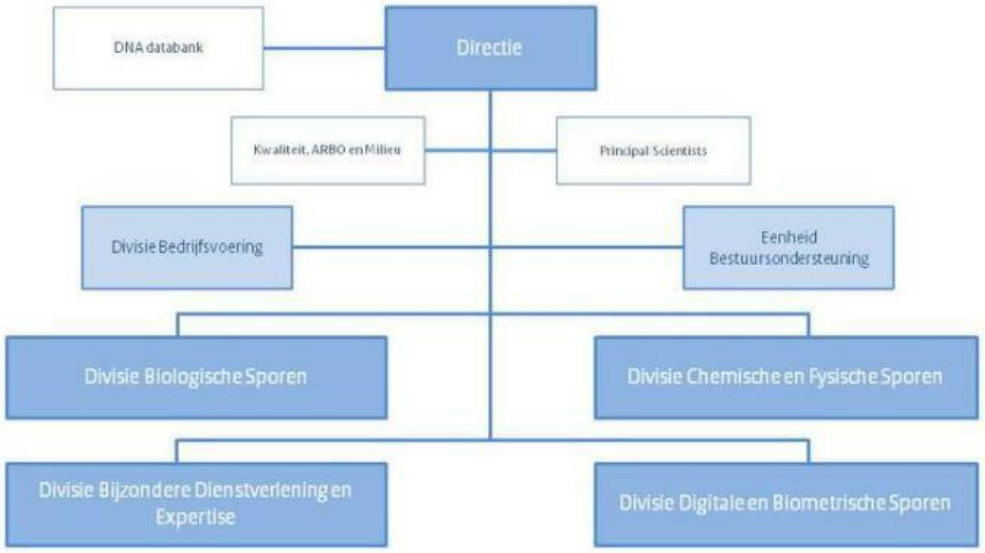
1. Divisie Biologische Sporen (BiS)
2. Divisie Chemische en Fysische Sporen (CFS)
3. Divisie Bijzondere Dienstverlening en Expertise (BDE)
4. Divisie Digitale en Biometrische Sporen (DBS)

Zoals ik eerder heb benoemd heb ik mijn afstudeeropdracht binnen de divisie Digitale en Biometrische Sporen (DBS), binnen het team Digitale Technologie (DT) en subteam Data analyse, uitgevoerd. Mijn bedrijfsbegeleider is ook werkzaam binnen dit subteam. De divisie DBS bestaat uit zes teams en het team DT bestaat uit vier subteams. De teams binnen DBS en subteams binnen DT zijn als volgt:

1. Forensische Biometrie, Spraak/Beeld/Vingersporen (FBS)
2. Forensische Software Engineering A (FSE-A)
3. Forensische Software Engineering LP (FSE-LP)
4. Forensische Big Data Analyse/Forensische Statistiek (FBDA/FSM)
5. Forensische Digitale Technologie (FDT/DT)
   a. Hardware
   b. Data analyse
   c. Crypto
   d. Verkeer

1.2 Mijn positie

Zoals ik eerder heb vermeld voer ik mijn afstudeeropdracht uit binnen het subteam Data analyse van de afdeling DT, wat onder de divisie DBS valt.



Figuur 1: Organogram Nederlands Forensisch Instituut. Bron: (NFI, 2021)

1.3 De afstudeeropdracht en stakeholders

Mijn afstudeeropdracht is ontstaan omdat verdachten in strafrechtelijke onderzoeken vaak beweren niet de auteur te zijn van een bericht dat op zijn/haar toestel is aangetroffen. Met deze opdracht gaan we proberen aan te tonen hoe een bericht op een toestel terecht is gekomen. Met behulp van de gebruikersdictionary en woordvolgorden op een toestel zou er aangetoond kunnen worden of een bericht op het toestel zelf is getypt of niet. Een collega wilde dit al een tijdje gaan onderzoeken maar heeft hier nog niet de tijd voor gehad. Het is dus geen zelf bedachte opdracht. Mijn woordenboekparser, een stukje code dat de nodige data zal verzamelen, die ik ga maken zal geïmplementeerd worden bij de collega's die zich bezighouden met de ontwikkeling van forensische tools. Deze collega's zijn werkzaam binnen de teams FSE-A en FSE-LP.

De volgende stakeholders zijn direct en indirect betrokken geweest bij mijn opdracht:

Direct betrokken

- NFI ontwikkelaars voor forensische tools
  Zij houden zich bezig met de ontwikkeling van forensische tools. Mijn woordenboekparser zou toegevoegd kunnen worden als een tool.

Indirecte betrokken

- Digitale deskundigen NFI
  Zij voeren onderzoeken uit en worden mogelijk opgeroepen om bij een rechtszaak uitleg te geven over hun bevindingen. Hiermee zou er mogelijk bewezen kunnen worden dat een bericht wel of niet op een toestel is getypt – wat het aannemelijker zou kunnen maken dat de eigenaar van het toestel het heeft getypt.
- Het Openbaar Ministerie
  Zorgt ervoor dat een persoon die ervan verdacht wordt een strafbaar feit te hebben gepleegd voor de rechter komen. Ze zorgen ervoor dat strafbare feiten worden opgespoord en vervolgd.

2. Het proces

In dit hoofdstuk wordt het afstudeerproces besproken. Er is een korte samenvatting van het Plan van Aanpak weergeven. Dit wordt gevolgd door een opsomming van alle uitgevoerde experimenten en de genomen stappen. Hierop volgt er wat diepgang in de aan te tonen competenties. De drie onderwerpen zijn onderverdeeld in drie subhoofdstukken.

2.1 Plan van aanpak

Het doel van mijn onderzoek was het achterhalen of het mogelijk was om te bepalen of een bericht op een toestel geschreven is of dat het bericht op een andere manier op het toestel zou zijn gekomen. Dit kan van belang zijn bij strafrechtelijke onderzoeken waarin verdachten beweren dat zij een bericht op hun toestel niet zelf hebben getypt. Deze beweringen zouden, wanneer ze het wel zelf hebben getypt, ontkracht kunnen worden. Mijn onderzoek heeft een bijdrage aan de onafhankelijke waarheidsbevinding van het NFI. Een bericht kan op de volgende manieren op een toestel zijn gekomen:

- Door het bericht te typen.
- Door het op een ander apparaat te typen en naar het toestel te synchroniseren (Whatsapp Web).
- Door het bericht door te sturen.
- Door het bericht te kopiëren.

De hoofdvraag van onderzoek luidt als volgt:
"*Op welke manier kan er aan de hand van de inhoud van de gebruikerswoordenboek bepaald worden of een bericht dat op een toestel staat geschreven is door de gebruiker of dat het bericht op een andere manier op het toestel is gekomen?"*

Voor mijn onderzoek wilde ik eerst vooronderzoek gaan verrichten om mezelf te kunnen verdiepen in het onderwerp. Dit wilde ik doen door eerst onderzoeksvragen op te stellen en deze vervolgens te beantwoorden. De opgestelde onderzoeksvragen zijn als volgt:

1. Van welke Samsung toestellen is het, ten behoeve van de uitvoering van mijn onderzoek, bekend dat ze makkelijk te rooten zijn?
2. Wat zijn de meest gebruikte toetsenbordapplicaties voor Android toestellen?
    a. Wat voor toetsenbordapplicatie draait er op de Samsung toestellen die makkelijk te rooten zijn?
3. Hoe zit het besturingssysteem van een Samsung (Android) toestel in elkaar?
    a. Waar is de toetsenbordapplicatie te vinden?
    b. Hoe wordt de data van de toetsenbordapplicatie opgeslagen?
    c. Waar zijn de woordvolgorden terug te vinden?
    d. Hoe worden de woordvolgorden opgeslagen?
4. Welke woorden staan er in een 'kaal' Samsung (Android) woordenboek opgeslagen?
    a. Kunnen de standaard woorden verwijderd of overschreven worden?
        i. Zo ja, hoe? Zo nee, waarom niet?

Na het verrichten van vooronderzoek zou ik weten waar de data van de toetsenbordapplicatie zich zou bevinden en zou ik kunnen uitzoeken hoe ik de data op een zo efficiënt mogelijke manier van het toestel af kan krijgen. Hierna zou ik kunnen beginnen met het bouwen van mijn woordenboekparser om de data te verzamelen en leesbaar te maken. Als mijn woordenboekparser gebouwd is kan er een statistisch model bedacht en opgezet worden om de betrouwbaarheid van mijn woordenboekparser te kunnen meten. Op basis van onder andere de betrouwbaarheidsmeting zou er een advies gevormd

kunnen worden over mijn woordenboekparser. Dit advies zou mondeling, waarschijnlijk in de vorm van een presentatie, toegelicht worden.

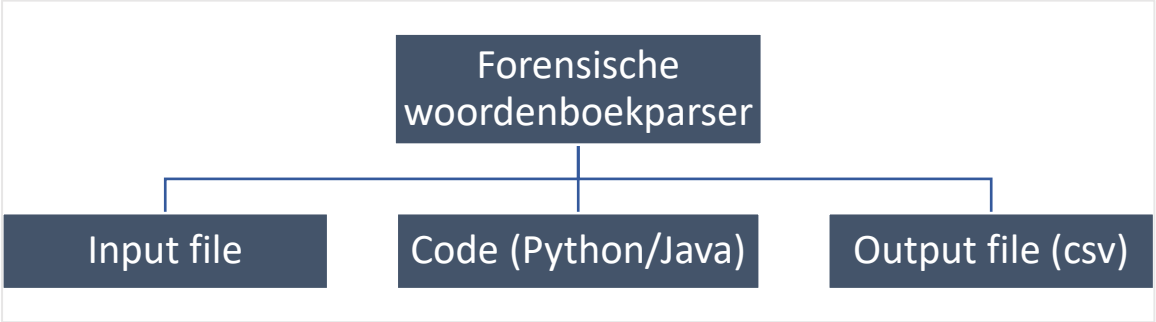
De werkmethode die ik gedurende mijn onderzoek aan wilde houden was de watervalmethode, een methode waarmee een project opgesplitst wordt in opeenvolgende fasen (Toolshero, 2022). Ik heb voor deze methode gekozen omdat ik het een fijne, gestructureerde manier van werken vindt. Als we mijn beschreven projectfasen uittekenen op basis van de watervalmethode ziet mijn planning er als volgt uit:

- Vooronderzoek verrichten
    - Het bestand met de toetsenborddata onderzoeken
        - Woordenboekparser bouwen
            1. Statistisch model opzetten
                a. Betrouwbaarheid woordenboekparser meten
                    i. Advies vormen + mondeling toelichten

Ik zou de volgende eindproducten opleveren aan het NFI:

1. Een forensische woordenboekparser (Python/Java)
2. Een statistisch model.
3. Een eindverslag.

Voor mijn woordenboekparser had ik een (korte) Product Breakdown Structure (PBS) opgesteld. In de PSB is te zien waar de woordenboekparser uit zal bestaan. De PSB ziet er als volgt uit:



Figuur 2 - Product Breakdown Structure van één van mijn eindproducten: de woordenboekparser

Hiernaast heb ik een planning opgesteld. De oranje kolommen staan voor conceptdeadlines en de rode kolommen staan voor officiële deadlines. Mijn planning is terug te vinden onder *Hoofstuk 6. Interne Bijlagen* als *Bijlage Int-II Planning*.

Tijdens mijn afstudeerstage wilde ik de volgende beroepstaken (A- en B-competenties) aantonen:
A-Competenties:
- Onderzoek
- Leren
- Professioneel werken
- Innovatie

B-Competenties:
- Infrastructuur analyseren
- Software analyseren
- Software adviseren

2.2 Uitgevoerde onderzoeken en experimenten

Ik ben op basis van mijn Plan van Aanpak en planning begonnen met mijn onderzoek. Het werd al snel duidelijk dat ik een aantal zaken in mijn planning verkeerd had ingeschat: vooronderzoek had meer tijd nodig en tussen het vooronderzoek en bouwen van de parser moest ik een tussenonderzoek verrichten. Het vinden van het bestand waar de data van de toetsenbordapplicatie in opgeslagen staat was niet zo makkelijk te vinden als verwacht. Althans, mijn bedrijfsbegeleider en ik hebben het bestand al snel op het oog gehad maar ik kon nergens de bevestiging vinden dat het daadwerkelijk dat bestand zou moeten zijn. De inhoud van het bestand dat wij op het oog hadden bestaat namelijk uit diverse en voor ons niet leesbare/begrijpbare karakters.

Omdat het dus nog niet met zekerheid bekend was welk bestand daadwerkelijk de data bevatte, was het nodig om extra onderzoek te verrichten. Om erachter te komen welk bestand het moest zijn heb ik diverse onderzoeken en experimenten uitgevoerd. De uitgevoerde onderzoeken, experimenten en genomen stappen worden hieronder per fase toegelicht. Tijdens het uitvoeren van de experimenten is er data (hierna: toetsenborddata) gegenereerd met het toetsenbord van het onderzoekstoestel. Er is bijgehouden op welk tijdstip welk woord is getypt zodat dit altijd terug te vinden is.

2.2.1 Vooronderzoek – Openbronnenonderzoek

Ik ben begonnen met het uitvoeren van vooronderzoek om zoveel mogelijk te weten te komen over mijn onderwerp. Mijn vooronderzoek bestond uit drie delen:

1. Een openbronnenonderzoek.
2. Een onderzoek naar de toetsenbordapplicatie.
3. Een onderzoek naar het bestand dat de data van de toetsenbordapplicatie bevat.

Voor het opdelen van mijn vooronderzoek heb ik gebruik gemaakt van de watervalmethode (Toolshero, 2022). Ik wilde eerst informatie opdoen over de onderwerpen waar ik wist dat ik mee te maken zou gaan krijgen, namelijk: Android, Samsung, toetsenbordapplicaties, Android woordenboeken, woordvolgordes en rooten. Vervolgens was het de bedoeling om de toetsenbordapplicatie op mijn onderzoekstoestel in kaart te brengen om hier vervolgens de opgeslagen data van te vinden.

Gedurende mijn afstudeerperiode heb ik voor de uitvoering van mijn afstudeeropdracht gebruik gemaakt van een onderzoekstoestel. Het onderzoekstoestel was een Samsung Galaxy S21 met Androidversie 11.0. Er is gebruik gemaakt van een Samsung toestel omdat het vanuit de afstudeerorganisatie gewenst was om specifiek naar de toetsenbordapplicatie van Samsung te kijken. Ik ben aan de slag gegaan met een Samsung Galaxy S21.

Om volledige toegang te verkrijgen tot het toestel moet het toestel geroot worden, oftewel: het besturingssysteem moest ontgrendeld worden zodat ik als gebruiker het toestel kan aanpassen. Het rooten van een toestel is legaal, maar kan er wel er wel voor zorgen dat de garantie van een toestel verloren gaat (Hildenbrand, Jerry, 2020). Mijn collega's hebben mij geadviseerd om niet met de nieuwste Androidversie aan het werk te gaan omdat kwetsbaarheden van de oudere Androidversie in de nieuwe verholpen kunnen zijn, wat het rooten van een toestel lastiger kan maken. Om deze reden ben ik met een Android 11.0 aan de slag gegaan.

Android is een besturingssysteem dat is bedacht door het bedrijf Android Inc., een bedrijf dat in oktober 2003 is opgericht. Het doel van Android was om slimmere mobiele toestellen zich bewust te maken van de locatie en de voorkeuren van de gebruiker. In 2005 heeft Google het bedrijf Android Inc. overgenomen. Google heeft in 2007 de eerste Android SDK Beta gelanceerd. Het eerste toestel dat op Android 1.0 draaide was de in september 2008 door T-Mobile uitgebrachte HTC Dream G1. Nu, ongeveer 3,5 jaar later, is de laatst uitgebrachte Androidversie Android 12.0. (van 't Klaphek, Michel, 2022).

Samsung is de naam van de grootste Android fabrikant ter wereld. Samsung is opgericht in Zuid-Korea en maakt naast telefoons ook andere producten zoals smartwatches, oortjes en huishoudelijke apparatuur. Mijn onderzoekstoestel behoort tot de productielijn 'Samsung Galaxy S', de high-end productielijn van Samsung. Sinds Android 9.0 draait de Samsung One UI (User Interface) op de Samsung toestellen. De laatste Samsung One UI versie heet Samsung One UI 4 (AndroidPlanet, 2022).

Volgens een aantal websites is de Samsung keyboard, de toetsenbordapplicatie die ontworpen is door Samsung, één van de beste Android toetsenbordapplicaties (Myrick & Wagoner, 2021) (Vyas, 2021) (Arici, 2021). Er is in de openbronnen echter niks te vinden over de opslaglocatie van de data van de Samsung Keyboard. Ik heb in mijn zoekopdrachten voornamelijk gebruik gemaakt de Booleaanse operator "AND".

Van bijvoorbeeld het Google toetsenbord, Gboard, is in de openbronnen wel de opslaglocatie van de toetsenborddata terug te vinden. De data wordt opgeslagen in de User Dictionary. De User Dictionary is op de volgende locatie terug te vinden: '/data/data/com.android.providers. userdictionary/ databases/user_dict.db' (zie de rode pijl in *Figuur 8*) (SomeGuyOnAComputer, 2019). De locatie heb ik zelf kunnen bevestigen door in Android Studio een virtueel Google Pixel toestel aan te maken en door het bestandssysteem heen te gaan met behulp van de Device File Explorer.

Mijn volledige vooronderzoek is uitgewerkt in een extern document, welke terug te vinden is onder *Hoofdstuk 5. Externe Bijlagen* als *Bijlage Ext-I Vooronderzoek*.

2.2.2 Vooronderzoek – Onderzoek naar de toetsenbordapplicatie

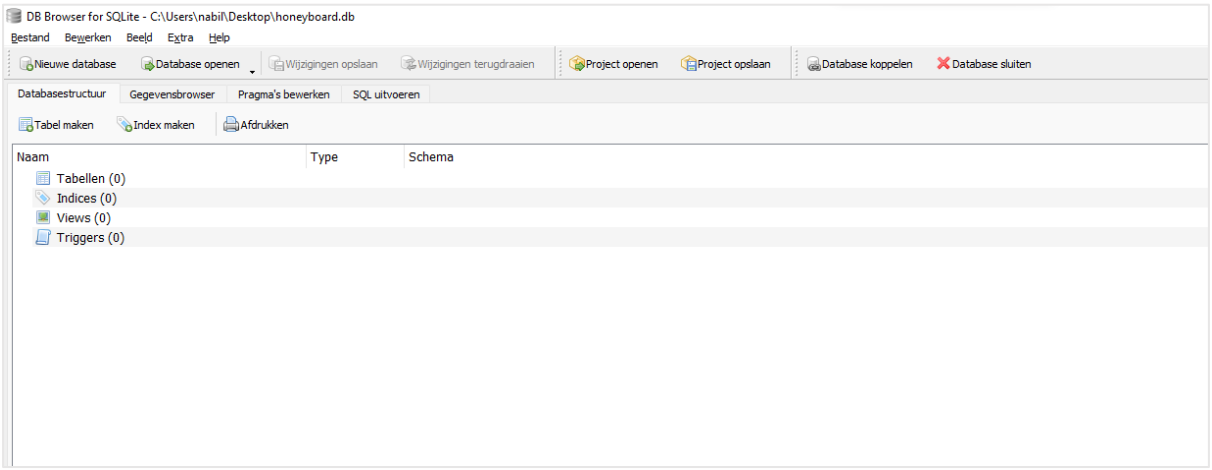
Tijdens het tweede deel van mijn vooronderzoek ben ik met behulp van zoektermen in back-up van mijn onderzoekstoestel, de Samsung Galaxy S21 met Androidversie 11.0, op zoek gegaan naar de Samsung toetsenbordapplicatie die op het toestel draait. In de openbronnen was er geen informatie over de opslaglocatie van de data van de Samsung Keyboard te vinden. Door met behulp van zoektermen op zoek te gaan naar de applicatiemap die bij de Samsung Keyboard applicatie zou horen, werd het zoekgebied naar het bestand met de data uit het toetsenbord verkleind.

Ik heb met behulp van ADB een back-up gemaakt van de "Data" map van het toestel. ADB staat voor Android Debug Bridge. ADB is een command-line tool waarmee er met een apparaat gecommuniceerd kan worden. Het zoekterm "board" heeft geleid tot een applicatie met de naam 'Honeyboard'. De applicatie package (APK) van Honeyboard heet 'com.samsung.android.honeyboard' (Samsung, 2020). Het pad dat naar deze APK leidt is '/data/data/com.samsung.android.honeyboard'.

Door het opzoeken van deze applicatie in openbronnen heb ik kunnen achterhalen dat dit de naam van een Samsung toetsenbord is (Thakur, 2022). Opvallend was dat in de data map van de Honeyboard-applicatie een map genaamd "Swiftkey" zat. Swiftkey is de naam van het toetsenbord dat ontworpen is door Microsoft. Dit toetsenbord stond echter wel standaard geïnstalleerd op mijn Samsung onderzoekstoestel. Onder de data map van Honeyboard was ook een bestand genaamd "honeyboard.db" te vinden. Omdat de gebruikersdictionary bij Gboard ook in een database bestandsformaat (.db) opgeslagen is, leek het waarschijnlijk dat dit bij Honeyboard ook het geval zou zijn.

2.2.3 Vooronderzoek – Onderzoek naar het bestand dat de toetsenbordapplicatie data bevat

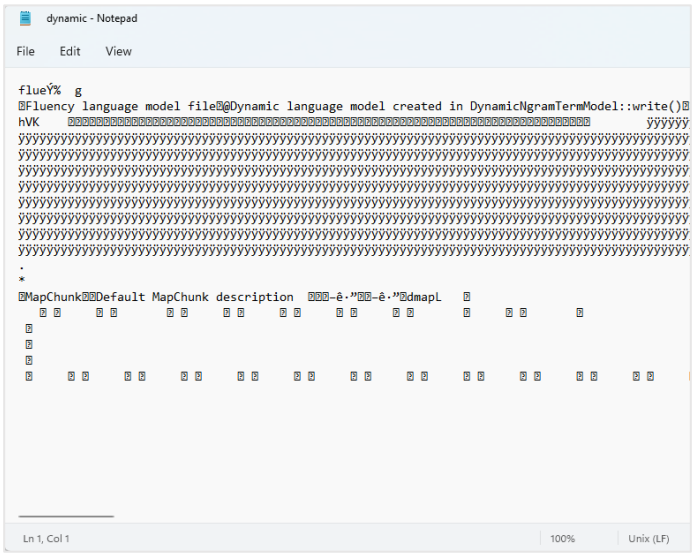
Tijdens het derde deel van mijn vooronderzoek heb ik het “honeyboard.db” bestand, dat tijdens het tweede deel van mijn vooronderzoek naar voren was gekomen, geopend met de applicatie DB Browser (SQLite). Dit is een tool waarmee bestanden met een database bestandsformaat geopend kunnen worden. De database bevatte echter niet de verwachtte data – de database was leeg. Na het genereren van meer toetsenborddata was de database nog steeds leeg (zie *Figuur 9*). Er is toetsenborddata gegenereerd door op het onderzoekstoestel woorden te blijven typen tot ze voorspeld kunnen worden zodat de woorden zeker ergens op zouden moeten zijn geslagen.



Figuur 3 - Inhoud "honeyboard.db" bestand voor en na het genereren van toetsenborddata

In de Swiftkey map, die in de data map van de Honeyboard- applicatie zit, zit een map genaamd ‘user’. In deze map zitten twee bestanden die opvielen: learned.js en dynamic.lm. Van beide bestanden is er een back-up te vinden in een ‘backup’ map. Het learned.js bestand bevat leesbare woorden. Er lijkt een telling bijgehouden te worden van opgeslagen woorden (zie figuur 9). Het dynamic.lm bestand bevat een aantal leesbare woorden en een hoop diverse karakters die niet leesbaar/begrijpbaar zijn (zie *Figuur 10*). Beide bestanden zijn terug te vinden onder *Hoofstuk 5. Externe Bijlagen* als *Bijlage Ext-V Woordenboekparser*.

Figuur 5 - Schermafbeelding van de inhoud van het learned.json bestand, geopend in de Windows applicatie Notepad



Figuur 4 - Schermafbeelding van een deel van de inhoud van het Dynamic.lm bestand, geopend in de Windows applicatie Notepad

Interne tool NFI die wijzigingen in bestanden detecteert

Binnen het NFI is er een tool ontwikkeld die snapshots van een toestel kan maken en de snapshots met elkaar kan vergelijken. De tool laat zien bij welke bestanden er een wijziging heeft plaatsgevonden. Een wijziging kan zijn: verwijdering, aanpassing of toevoeging.

Samen met een collega, een van de ontwikkelaars van de tool, heb ik een aantal kleine experimenten uitgevoerd om te achterhalen waar de toetsenborddata op wordt geslagen. Dit proces verliep als volgt: er werd een snapshot gemaakt van het toestel (een soort nulpunt), vervolgens typten wij een aantal woorden in die makkelijk te herkennen zouden zijn en na het typen van de woorden werd er nog een snapshot van het toestel gemaakt. De tool vergeleek de snapshots met elkaar en weergaf na het vergelijken een lijst met bestanden waarin wijzigingen plaats hadden gevonden.

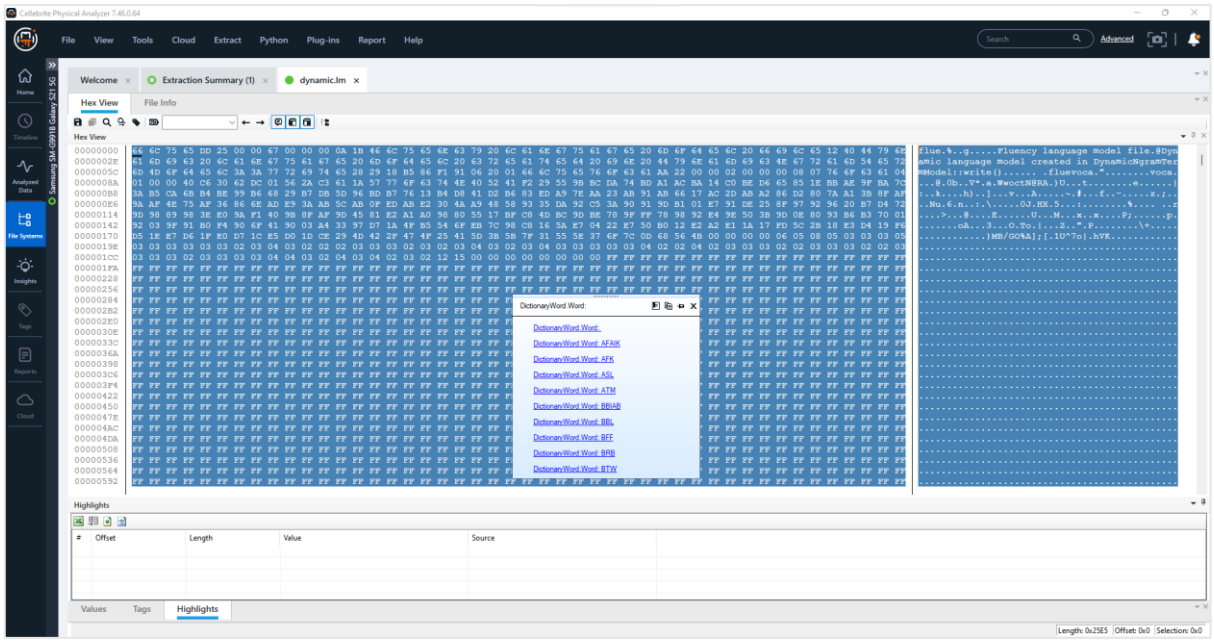
Naast in de applicatiedata bestanden van de applicaties waar wij de woorden in hadden getypt, WhatsApp en Samsung Notes, hebben wij de woorden niet expliciet terug kunnen vinden. Wat wel opviel is dat de dynamic.lm en learned.json bestanden, twee bestanden ik eerder heb bekeken, elke keer wijzigingen bevatte. Deze twee bestanden zijn ook de enige bestanden met wijzigingen die relevant leken te zijn voor het onderzoek. De andere bestanden waren voornamelijk bestanden van de applicaties waar de woorden in getypt waren. Het honeyboard.db bestand is geen enkele keer naar voren gekomen.

Universal Forensics Extraction Device (UFED) Touch/4PC/Physical Analyzer (PA)

Een collega kwam met het idee om te kijken of UFED Physical Analyzer (PA) de woorden er misschien niet zelf uit het toestel zou kunnen halen. UFED is een product van het bedrijf Cellebrite, een bedrijf dat zich bezighoudt met Digital Intelligence (het verzamelen, bekijken, analyseren en beheren van data) bij onderzoeken (Cellebrite, 2022). Op aanraden van de collega heb ik geprobeerd om met de UFED Touch en met UFED 4PC een fysieke kopie van het toestel te maken om de data te kunnen bekijken met de UFED PA. Het is met beide tools niet gelukt om een fysieke kopie van het toestel te maken, waarschijnlijk doordat het onderzoekstoestel een vrij nieuw toestel is.

In UFED PA kunnen ook back-up bestanden in een gecomprimeerd formaat ingeladen worden om de data te bekijken. Ik heb met behulp van Adb een back-up van het onderzoekstoestel gemaakt. De gemaakte kopie was in .tar formaat, een gecomprimeerd bestandsformaat (vergelijkbaar met zip) waarin meerdere bestanden in één archiefbestand verzameld kunnen worden (WinZip, 2022). Ik heb de fysieke kopie als Android Back-up ingeladen in UFED PA.

Via de filesystem heb ik het dynamic.lm bestand geopend omdat dit bestand bij elk experiment met de interne tool van het NFI naar voren is gekomen. Tijdens het openen van het bestand gaf UFED PA aan dat er een woordenlijst bij het bestand hoorde (zie *Figuur 11*).



Figuur 6 - Dynamic.lm in UFED Physical Analyzer - UFED PA haalt woordenboek woorden uit het bestand

Onderaan in de woordenlijst waren de woorden die ik op eerdere momenten heb getypt terug te vinden. In het bestand staat er bij elk woord ook een frequentie vermeld. Merendeels van de woorden die in de woordenlijst staan heb ik zelf niet getypt. Het lijkt erop dat die er al standaard in staat. UFED PA geeft in het dynamic.lm bestand niet expliciet aan waar de woorden vermeld staan. In dynamic.lm staat ".lm" voor het bestandsformaat 'Language Model', een formaat dat gebruikt wordt bij Microsoft Windows en Microsoft applicaties. Een Language Model bestand bevat statistische data die gebruikt wordt om voorspellende teksten te weergeven in Windows en Microsoft applicaties (FileInfo, 2022).

Nulpunt creëren

Na het vinden van de woordenlijst in UFED PA heb ik een nulpunt gecreëerd voor mijn onderzoek om zo op een meer gecontroleerde manier woorden te typen, zinnen te vormen en om te zien hoe de woordenlijst er "leeg" uitziet. Het nulpunt creëren heb ik gedaan door de persoonlijke voorspellingen op het onderzoekstoestel te verwijderen. Om te testen of de woorden echt verwijderd waren, heb ik de woorden die ik eerder getypt had door het toetsenbord geprobeerd te laten voorspellen. De woorden werden niet meer voorspeld.

Na het testen van de voorspellingen heb ik een nieuwe kopie gemaakt van het systeem om in te laden in UFED PA. Op deze manier heb ik een woordenlijst verkregen met woorden die al standaard in de lijst terug te vinden zijn. Er staan al standaard 75 woorden in het woordenboek (zie *Tabel 1*). Alle woorden, op twee na, hebben ook een standaard frequentie van 5. Hier heb ik tijdens mijn onderzoek rekening mee kunnen houden.

#	Woord	Freq. *	#	Woord	Freq. *	#	Woord	Freq. *
1	(lege ruimte)	0	26	hw	5	51	POS	5
2	Donald	0	27	IAC	5	52	ppl	5
3	Trump	5	28	IC	5	53	qt	5
4	Sterling	5	29	IDK	5	54	re	5
5	AFAIK	5	30	IIRC	5	55	SMH	5
6	AFK	5	31	IKR	5	56	sry	5
7	ASL	5	32	IM	5	57	TBA	5
8	ATM	5	33	IMO	5	58	TBC	5
9	BBIAB	5	34	IMHO	5	59	TC	5
10	BBL	5	35	IRL	5	60	thx	5
11	BFF	5	36	LMK	5	61	TIA	5
12	BRB	5	37	LOL	5	62	TLC	5
13	BTW	5	38	MMB	5	63	TMI	5
14	CTN	5	39	msg	5	64	TTFN	5
15	CYE	5	40	MYOB	5	65	TTYL	5
16	dl	5	41	NC	5	66	txt	5
17	ETA	5	42	NM	5	67	TY	5
18	FWIW	5	43	noob	5	68	XOXO	5
19	FYI	5	44	NP	5	69	Ynt	5
20	GG	5	45	NTN	5	70	YOLO	5
21	GJ	5	46	OMG	5	71	YW	5
22	GL	5	47	OMW	5	72	ZZZ	5
23	GTG	5	48	OT	5	73	DM	5
24	GMV	5	49	PC	5	74	ㅁㅗㅇㅏㅋㅣ	5
25	HTH	5	50	pls	5	75	ㅌㅏㅇㅣㅈㅖㄴ	5

Tabel 1 - Standaard woordenlijst Samsung Galaxy S21, Android 11.0

** freq. = frequentie*

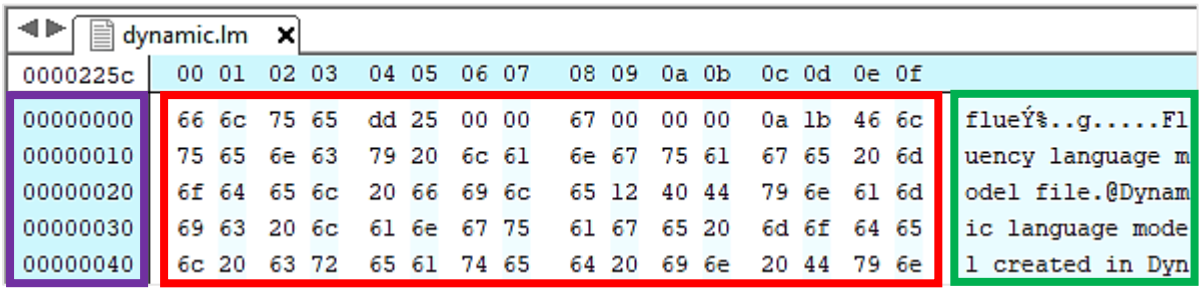
In de tabel zijn er 3 namen, 2 Koreaanse woorden, 1 lege ruimte en 70 Engelse (veel in berichten gebruikte) afkortingen te vinden. Een aantal voorbeelden van de Engelse afkortingen luiden als volgt (SlickText, 2020):

- BRB = Be Right Back
- BTW = By The Way
- IMO = In My Opinion
- IMHO = In My Humble Opinion
- NP = No Problem
- LOL = Laugh Out Loud
- TBA = To Be Announced
- TTYL = Talk To You Later

2.2.4 Data in dynamic.lm inzichtelijk maken

Zoals in het voorgaande hoofdstuk duidelijk is geworden, is de tekst in het dynamic.lm bestand niet leesbaar/begrijpbaar bij het openen van het bestand met het programma Notepad. Elk programma interpreteert een bestand bij openen op zijn eigen manier. Sommige karakters kunnen bijvoorbeeld onzichtbaar zijn omdat het programma deze niet kan interpreteren. Door een bestand met een Hex Editor te openen kan een bestand byte voor byte (karakter voor karakter) weergeven worden. Een Hex Editor is een speciaal type editor die de inhoud van elk bestandsformaat kan weergeven, waardoor ook de karakters die in andere programma's onzichtbaar zijn zichtbaar gemaakt kunnen worden. "Hex" staat voor hexadecimaal, wat een getallensysteem met het grondtal 16 is. Het grondtal van een decimaal getallensysteem is 10 (Mead, Ian, 2018) (Novell.com, n.d.).

Door het dynamic.lm bestand met een Hex Editor te openen werden een aantal delen uit het bestand leesbaar. De Hex Editor die ik heb gebruikt, Hex Editor Neo, geef bestanden op de volgende manier weer:



Figuur 7 - Het dynamic.lm bestand geopend met een Hex Viewer, op te delen in drie kolommen (geopend met Hex Viewer Neo, markeringen zelf toegevoegd).

De weergave van Hex Editor Neo is op te delen in drie kolommen (zie *Figuur 13*:

1. De offset (paarse markering)
2. Hexadecimale waarden (rode markering)
3. ASCII waarden (groene markering)

In de eerste kolom wordt een zogenoemde "offset" weergeven. De offset geeft aan wat de hexadecimale positie van het begin van de regel is. Het eerste hexadecimale getal in het bestand is "66" en heeft de offset "00000000". Het tweede getal is "6c" en heeft de offset "00000001". Het laatste getal op de eerste regel is ook "6c", maar heeft de offset "0000000f". Een offset van een hexadecimaal getal geeft dus de positie van dat getal aan.

In de tweede kolom zijn hexadecimale waarden terug te vinden. Een hexadecimaal getal kan de cijfers 0 t/m 9 en letters A t/m F bevatten. In de derde kolom zijn de hexadecimale waarden uit de tweede kolom omgezet naar ASCII (American Standard Code for Information Interchange). ASCII is een numerieke representatie zoals een "a" of "@", een actie of iets soortgelijks. In *Figuur 14* is een ASCII tabel weergeven waarin onder andere de hexadecimale getallen met bijbehorende ASCII representaties zichtbaar zijn (AsciiTable, n.d.):

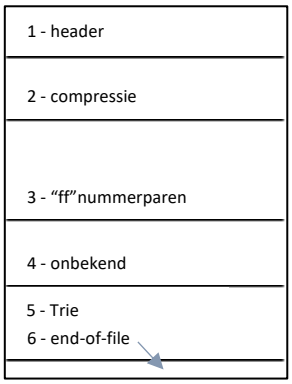
Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040		Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174			|
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177	&#127;	DEL

Figuur 8 - ASCII tabel (bron: https://www.asciitable.com/)

In bovenstaande tabel is onder andere te zien dat het hexadecimaal getal “2A” in ASCII geïnterpreteerd wordt als een plus (+), dat het hexadecimaal getal “70” geïnterpreteerd wordt als een kleine letter ‘p’ en dat het hexadecimaal getal “50” geïnterpreteerd wordt als een grote letter ‘P’.

Het dynamic.lm bestand lijkt opgedeeld te zijn in zes delen:

1. De file header en een aantal leesbare woorden.
2. Een deel met een hoge entropie (hoge maat van wanorde (Jaspers, Arnout, 2021)), gevolgd door een klein stukje met een lage entropie.
3. Een deel met alleen “ff” nummerparen.
4. Een deel waarin de nummerparen “bf 24 8e 64” een aantal keer herhaald worden, met een aantal leesbare woorden.
5. Een oplopende hexadecimale nummering.
6. Het einde van het bestand.



Figuur 9 - Opdeling dynamic.lm bestand

Op de volgende pagina’s worden de waarnemingen en bevindingen met betrekking tot de zes bestandsdelen los van elkaar behandeld. Bij elk deel is ter illustratie een schermafbeelding toegevoegd. De schermafbeeldingen zijn genomen tijdens het bekijken van het dynamic.lm bestand dat afkomstig is uit de kopie dat als “nulpunt” wordt beschouwd. Het nulpunt is hoe het bestand er standaard uitziet op het onderzoekstoestel, zonder dat de data gewijzigd is.

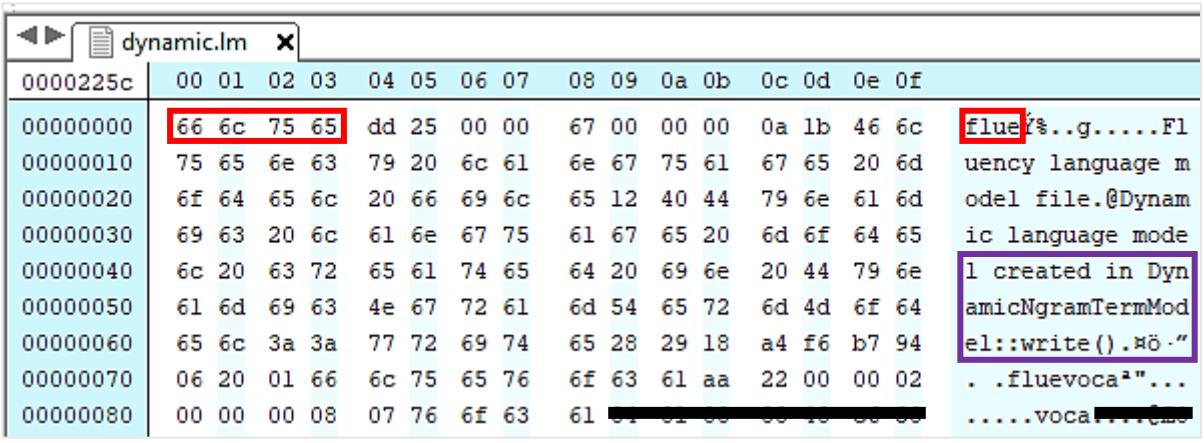
Deel 1: de file header met een aantal leesbare woorden

Zoals de titel het al aangeeft, zijn in het eerste deel van het dynamic.lm bestand de file header en een aantal leesbare woorden terug te vinden. Elk bestand bevat een file header die als een signatuur voor het bestandstype dient. Door het signatuur weet een programma of een besturingssysteem om wat voor type bestand het gaat en wat er met de inhoud moet gebeuren/hoe het bestand geopend moet worden. Een drietal voorbeelden van file headers van bekende bestandsformaten zijn (Kessler, Gary, 2022):

- “50 4b 03 04 14 00 06 00” (in ASCII: PK.....) voor DOCX, PPTX en XLSX bestanden.
- “ff d8” (in ASCII: ÿØ) voor een JPE, JPEG en JPG bestanden.
- “14 40 44 46” (in ASCII: %PDF) voor PDF, FDF en AI bestanden.

Voor het dynamic.lm bestand is de file header “66 6c 75 65”, wat in ASCII voor “flue” staat (zie rode markeringen in *Figuur 13*). “Flue” lijkt hierbij de afkorting te zijn voor “Fluency”, wat verder in de ASCII interpretaties te lezen is. Hiernaast is er ook te lezen dat de Dynamic Language Model gecreëerd is in “DynamicNgramTermModel” (zie paarse markering in *Figuur 16*). Een N-gram model is een model waarmee voorspeld kan worden wat het hoogstwaarschijnlijke, eerstopvolgende woord in een reeks kan zijn. Een N-gram model wordt gebouwd op basis van de frequentie waarin bepaalde woord- of letterreeksen voorkomen in een corpus (een verzameling van teksten/uitspraken) en door hier vervolgens de kansen van in te schatten. Een N-gram model is een type Language Model dat gaat over het inschatten van kansen bij woordreeksen (arvindpdmn, 2021).

Dat er in het dynamic.lm bestand vermeld staat dat het bestand in een N-gram Model gecreëerd is laat zien dat dit bestand betrekking heeft op de woordreeksen die voorspeld worden wanneer een gebruiker op zijn/haar telefoontoetsenbord typt. In het bestand is echter niet te zien hoe de kansen worden voorspeld en waar dit opgeslagen staat.

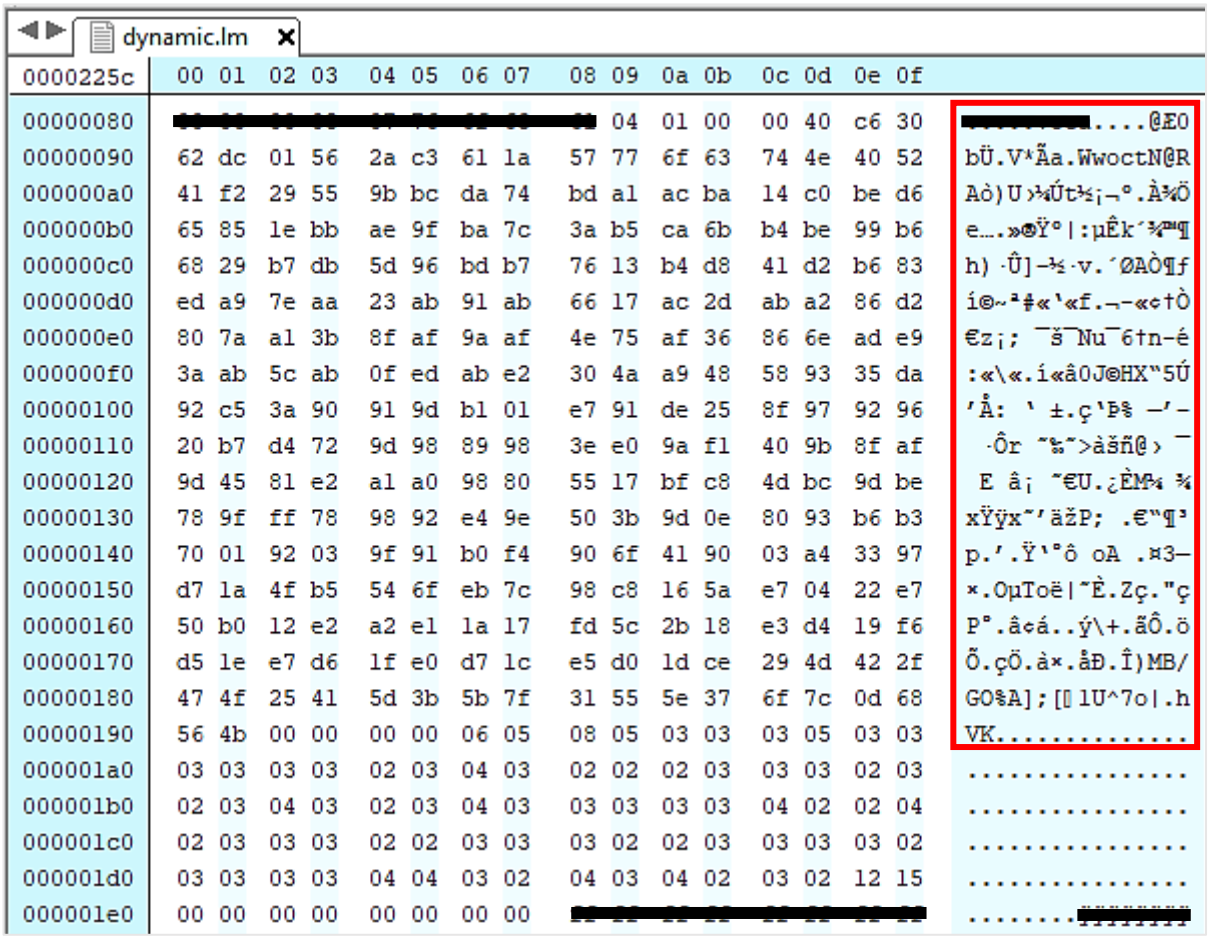


Figuur 10 - Deel 1 van het dynamic.lm bestand: de header met een aantal leesbare woorden (geopend met Hex Editor Neo, markeringen en zwarte strepen zelf toegevoegd).

Deel 2: het deel met de hoge entropie, gevolgd door een stukje met een lage entropie

In het tweede deel van het bestand is een hoge entropie te zien, wat wil zeggen dat er grote maat van wanorde is (zie rode markering in *Figuur 17*). Er zijn veel karakters te zien waar - net als in het eerste deel - niets leesbaars uit te halen is. Mijn begeleider heeft mij erop gewezen dat een hoge entropie kan wijzen op een compressie; een samendrukking van data. Bij het comprimeren van data wordt de data met behulp van een algoritme verkleind zonder dat er data verloren gaat (Encyclo, 2007).

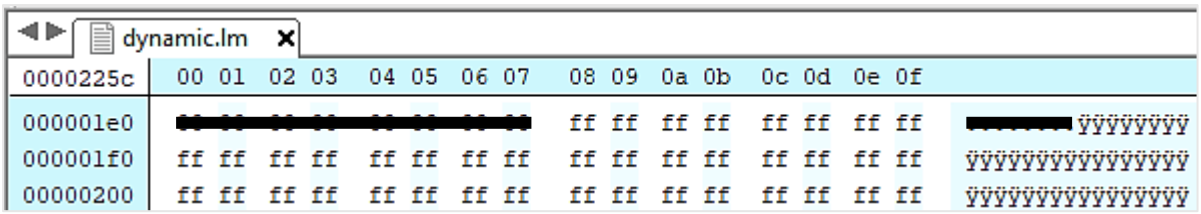
In het geval van het dynamic.lm bestand zou de hoge entropie er dus op kunnen wijzen dat er op deze plek in het bestand veel data staat, maar dat deze samengedrukt is om bijvoorbeeld ruimte te besparen. Er is niet direct te zien welk algoritme er gebruikt zou zijn voor de compressie. De betekenis van de waarden in het stukje met de lage entropie is (nog) onbekend.



Figuur 11 - Deel 2 van het dynamic.lm bestand: het deel met de hoge entropie (geopend met Hex Editor Neo, markering en zwarte strepen zelf toegevoegd).

Deel 3: het deel met de "ff" nummerparen

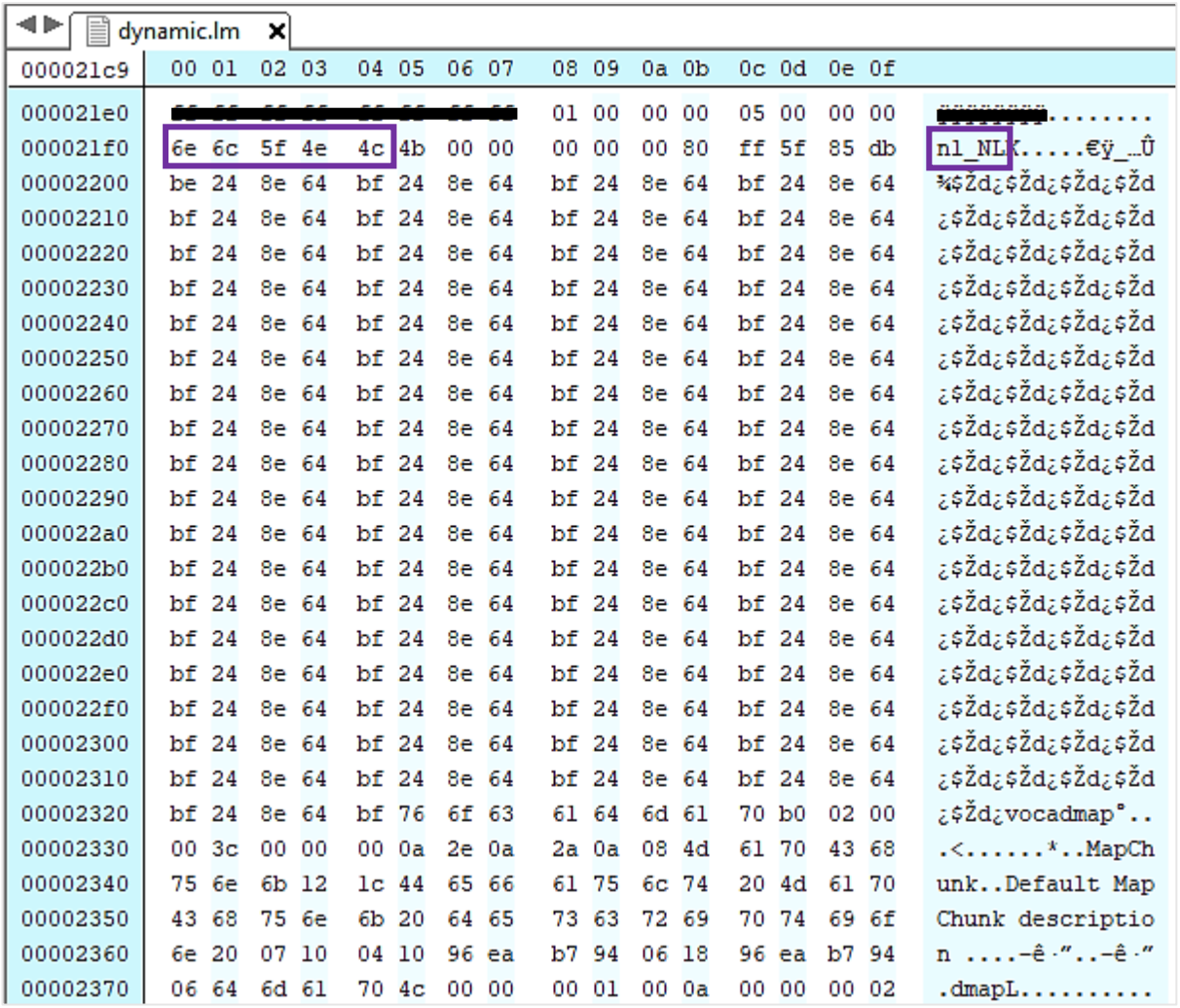
In dit deel is een groot aantal (zeven kantjes vol) "ff" nummerparen achter elkaar terug te vinden. Vanwege het grote aantal is ervoor gekozen om maar een aantal regels van dit deel in onderstaand figuur weer te geven, aangezien de andere delen er hetzelfde uitzien. Het is vooralsnog onbekend wat de grote hoeveelheid "ff" nummerparen zouden kunnen betekenen.



Figuur 12 - Deel 3 van het dynamic.lm bestand: een stukje van het deel met de "ff' nummerparen (geopend met Hex Editor Neo, zwarte strepen zelf toegevoegd).

Deel 4: het deel waarin "bf 24 8e 64" herhaald wordt, met een aantal leesbare woorden

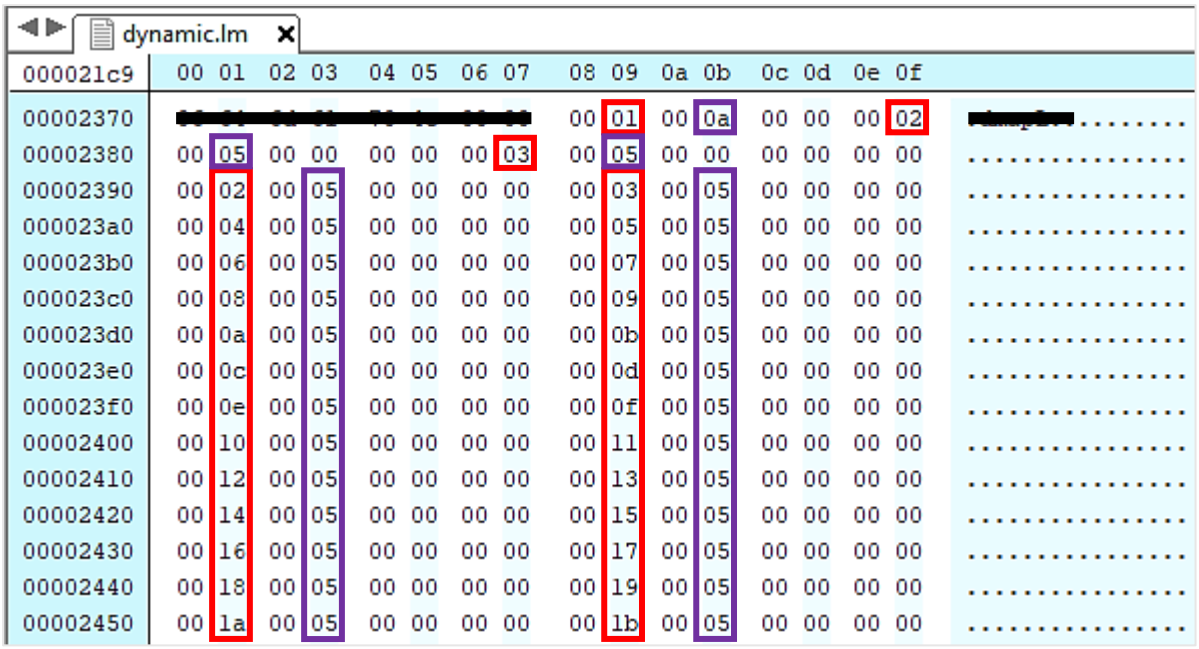
In dit deel van het bestand is "nl_NL" (zie paarse markering in figuur 19), samen met een herhaling van de nummerparen "bf 24 8e 64" en een aantal woorden te zien. "nl_NL" is een identiteitswaarde die verwijst naar "Nederlands (Nederland)" (LocalePlanet, n.d.). Er wordt specifiek "Nederland" aangegeven omdat er ook een identiteitswaarde voor Nederlands in België is (nl_BE). Het Nederlands (Nederland) lijkt een verwijzing te zijn naar de taal die ingesteld is op mijn onderzoekstoestel. Het is vooralsnog onbekend wat de nummerparen zouden kunnen betekenen.



Figuur 13 - Deel 4 van het dynamic.lm bestand: het deel waarin de nummerparen "bf 24 8e 64" een aantal keer herhaald worden, met een aantal leesbare woorden (geopend met Hex Editor Neo, zwarte strepen zelf toegevoegd).

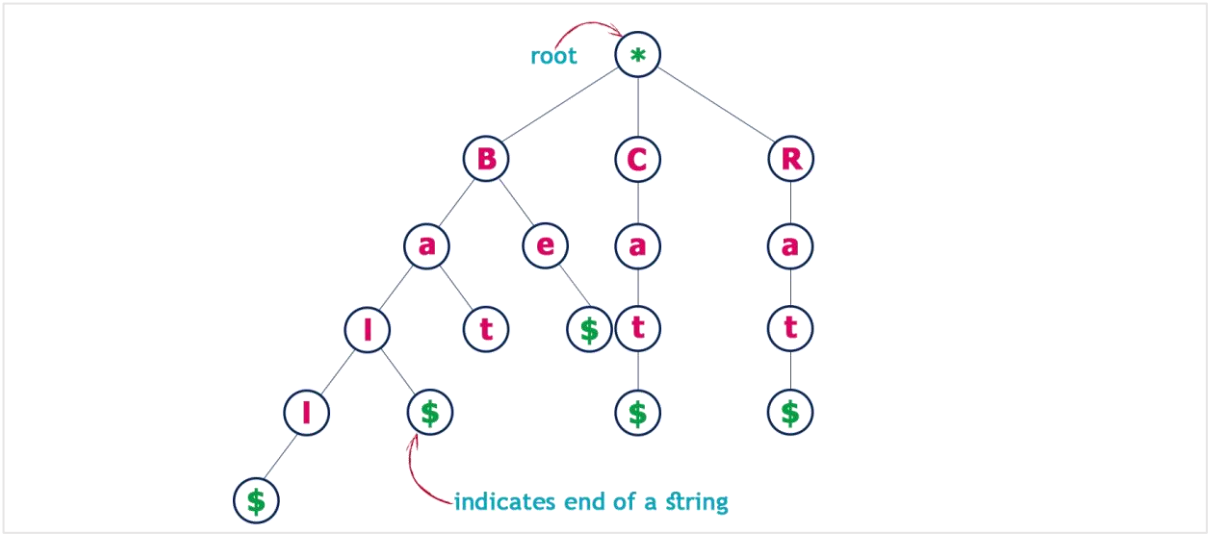
Deel 5: het deel met een oplopende hexadecimale nummering

Dit is het op een na laatste deel van het bestand. In dit deel is er een oplopende, hexadecimale nummering waargenomen (zie rode markeringen in *Figuur 20*). De oplopende nummering zou kunnen wijzen op indexnummers. Elk hexadecimaal getal, behalve “02”, komt maar één keer voor in de nummering. Aan elk oplopend nummer lijkt nog een hexadecimaal getal gekoppeld te zijn. Dit andere hexadecimale getal is, op één keer na, elke keer “05” (zie paarse markeringen in *Figuur 20*). Omdat dit deel een redelijk lang stuk is en de nummering op dezelfde manier doorloopt, is ervoor gekozen om niet het gehele deel te weergeven in onderstaand figuur.



Figuur 14 - Deel 5 van het dynamic.lm bestand: een stukje het deel waarin de oplopende, hexadecimale nummering te zien is (geopend met Hex Editor Neo, markeringen en zwarte strepen zelf toegevoegd).

Een collega constateerde dat een ‘Trie’ datastructuur een voor de hand liggende structuur is bij bovenstaande data omdat deze datastructuur gebruikt wordt voor het opslaan van een verzameling van woorden of strings. De naam ‘Trie’ is afkomstig van het woord ‘retrieval’ (terugzoeken). Een Trie wordt ook wel Prefix Tree of Digital Tree genoemd. Een Trie maakt het terugvinden van een string in een verzameling van strings makkelijker (BSC, n.d.) (Açıl, Sıddık, 2020).



Figuur 15 - Visueel voorbeeld van een Trie datastructuur (Açıl, Sıddık, 2020)

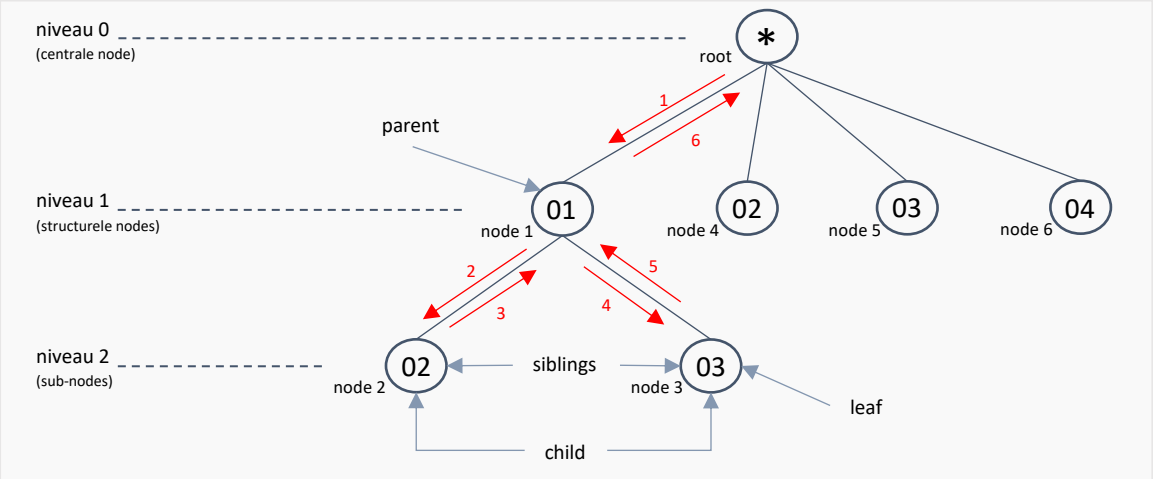
Een Trie is een type tree datastructuur. Een tree (boom) is een niet-lineaire, hiërarchische datastructuur. Elk type tree bestaat uit een verzameling van nodes (knooppunten) waar de tree data in opslaat. Een tree beschikt over een centrale node, structurele nodes en sub-nodes. De centrale node wordt de root-node genoemd. De root node is de bovenste node en bevindt zich op niveau 0. Dit is het begin van een tree. Op niveau 1, de laag onder de root-node, bevinden de structurele nodes zich. Elke structurele node is uniek – er kunnen geen duplicaten op dit niveau zitten. De sub-nodes bevinden zich vanaf niveau 2.

Twee nodes kunnen met behulp van een edge (rand) met elkaar verbonden worden. Wanneer een sub-node en een structurele node met behulp van een edge zijn verbonden, worde deze nodes child- en parent-nodes genoemd. De structurele node is dan een parent van de sub-node. Een parent-node kan meerdere child-nodes hebben, maar een child-node kan maar één parent-node hebben. Twee sub-nodes die dezelfde parent-node hebben zijn siblings van elkaar. Een node die geen child-node heeft wordt een leaf-node worden genoemd. De root-node en de structurele nodes zijn hierbij een uitzondering. De structurele nodes worden geen child-nodes van de root-node genoemd, zijn geen siblings en worden ook geen leaf-node genoemd als zij geen child-node hebben (GeeksforGeeks, 2022) (TutorialAndExamples, 2020):

Hieronder zijn de benoemde termen met betrekking tot de een tree-structuur kort samengevat:

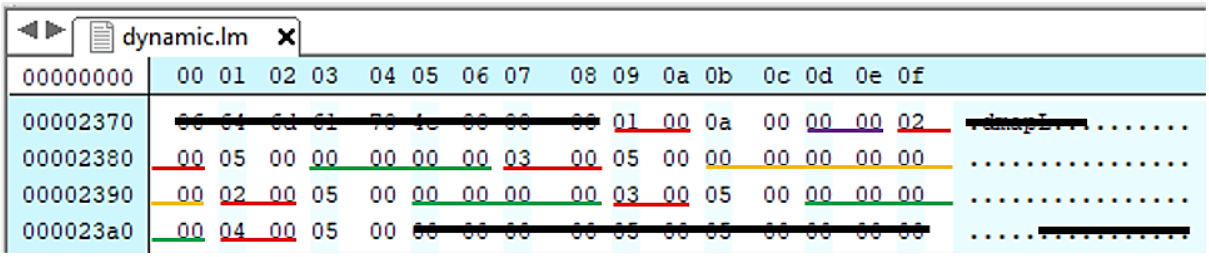
- Een node is een knooppunt waarin data wordt opgeslagen.
- De centrale en bovenste node heet de root-node.
- Een parent-node is een node met een sub-node.
    - Uitzondering: de root-node wordt niet beschouwd als parent-node van de structurele nodes.
- Een child-node is sub-node van één andere node.
    - Uitzondering: de structurele nodes worden niet als child-nodes van de root-node beschouwd.
- Siblings zijn child-nodes die dezelfde parent node hebben.
    - Uitzondering: de structurele nodes worden niet als siblings van elkaar beschouwd.
- Een edge is een verbinding tussen twee nodes.

Wanneer het eerste stukje van de oplopende hexadecimale nummering uit het dynamic.lm bestand (zie *Figuur 20*) in een Trie structuur getekend wordt, ziet dit er als volgt uit:



Figuur 16 - Een deel van de oplopende hexadecimale nummering uit het dynamic.lm bestand, getekend in een Trie datastructuur met zichtbare niveaus en aangewezen voorbeelden van child-, parent-, sibling- en leaf-nodes. De rode pijlen geven de volgorde waarop de Trie getekend wordt aan.

Het deel van het dynamic.lm bestand waar de getekende nodes zich in bevinden ziet er als volgt uit:



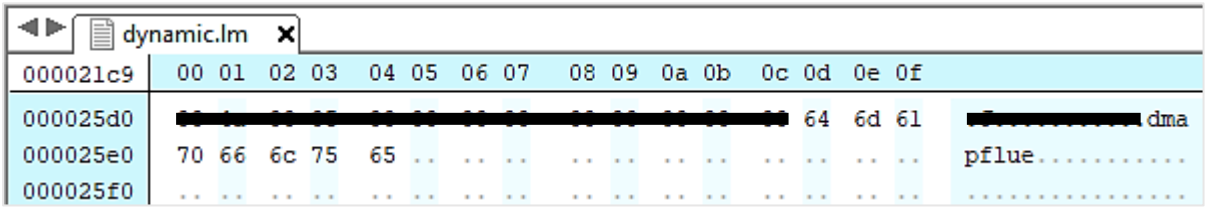
Figuur 17 - Een deel van de hex van de oplopende hexadecimale nummering uit het dynamic.lm bestand (markeringen toegevoegd. Rood = indexnummer, paars = twee "00" nummerparen, groen = vier "00" nummerparen, oranje = zes "00" nummerparen). De gekleurde markeringen komen overeen met de gekleurde woorden in de onderstaande genummerde lijst.

Tussen de waarde van de verschillende nodes staan, zoals in bovenstaand figuur te zien is, diverse hoeveelheden "00" nummerparen. Het aantal "00" nummerparen zou aangeven op welk niveau een node zich, ten opzichte van de voorgaande node, bevindt. De "00" nummerparen uit figuur 23 zouden de volgende betekenissen hebben (de namen van de nodes zijn afkomstig uit figuur 22):

1. Wanneer er twee "00" nummerparen tussen de waarden van twee nodes staan, betekent het dat de tweede waarde een child van de eerste waarde is. Tussen node 1 en node 2 staan twee "00" nummerparen. Node 1 is de parent van node 2 en node 2 is de child van node 1.
2. Wanneer er vier "00" nummerparen tussen de waarden van twee nodes staan, betekent het dat de nodes op hetzelfde niveau zitten. Het kan zijn dat deze twee nodes siblings zijn en dus dezelfde parent-node delen of dat deze twee nodes structurele nodes zijn en dus beiden verbonden zijn met de root-node. Tussen node 2 en node 3 staan vier "00" nummerparen. Omdat node 2 een child van node 1 is, wordt node 3 ook een child van node 1. Node 2 en node 3 zijn siblings.
3. Wanneer er zes "00" nummerparen tussen de waarden van twee nodes staan, betekent het dat de tweede waarde een edge wordt van een waarde die twee niveaus boven de eerste waarde zit. Het kan zijn dat de tweede waarde een child van de grandparent (de parent van de parent) van de eerste waarde wordt of dat de tweede waarde een structurele node wordt en dus verboden wordt met de root-node. Tussen node 3 en node 4 staan zes "00" nummerparen. Omdat de grandparent van node 3 de root zou zijn, wordt node 4 een edge van de root-node.
4. De betekenis van vier "00" nummerparen is terug te vinden onder punt 2. Tussen node 4 en 5 staan vier "00" nummerparen. Omdat node 3 een structurele node is, wordt node 5 ook een structurele node. Voor de vier "00" nummerparen tussen node 5 en 6 geldt hetzelfde.

Deel 6: het einde van het bestand

De meeste bestanden hebben geen "end-of-file" marker nodig (Ellis, Scott R., 2013). Het lijkt er echter op dat het bij een Language Model bestand misschien wel mogelijk is. Het einde van het bestand wordt duidelijk aangegeven met de ASCII tekst "dmapflue".



Figuur 18 - Deel 6 van het dynamic.lm bestand: waarin de end-of-file marker te zien is (geopend met Hex Editor Neo, zwarte strepen zelf toegevoegd).

2.2.5 Data in dynamic.lm analyseren en ontcijferen

Tijdens en na het vastleggen van waarnemingen met betrekking tot het dynamic.lm bestand ben ik het bestand gaan analyseren en ben ik de inhoud gaan proberen te ontcijferen om inzichtelijk te maken wat voor data er in het bestand staat.

Ik heb ervoor gekozen om de inhoud van het bestand te analyseren door twee verschillende versies van het bestand naast elkaar te leggen: het nulpunt en een versie waarin door een aantal, door mij op het onderzoekstoestel getypte, woorden terug te vinden zijn. Er zal op de volgende manier naar deze versies van het dynamic.lm bestand verwezen worden in de tekst:

1. Dynamic V0 = het nulpunt
2. Dynamic V1 = een versie waarin door mij getypte woorden terug te vinden zijn

Ik heb beide bestanden geopend in de Hex Editor 'Hex Editor Neo', de hex van beide bestanden uitgeprint en naast elkaar gelegd. Ik heb gebruikt gemaakt van de commandline tool 'vbindiff' om de overeenkomsten tussen beide bestanden weg te strepen. Vbindiff is. De tool kan echter niet alle overeenkomsten in kaart brengen omdat Dynamic V1 op sommige plaatsen in het bestand extra karakters bevat. Door Dynamic V0 met Dynamic V1 te vergelijken heb ik het volgende kunnen constateren:

- Het begin en het einde van het bestand zijn hetzelfde in Dynamic V0 en Dynamic V1.
- In het middenstuk bevat Dynamic V1 meer karakters dan Dynamic V0.
- De hexadecimale nummering die te zien was in Dynamic V0 (zie figuur 24) is terug te vinden in Dynamic V1. Er zijn achteraan de nummering extra karakters toegevoegd.

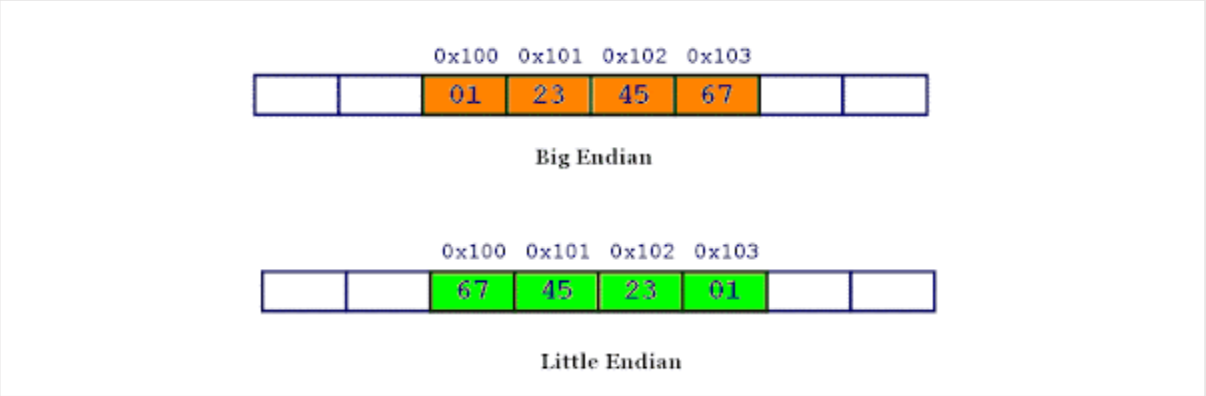
Om bovenstaande constateringen te analyseren en tot nieuwe bevindingen te komen, heb ik meer data op het onderzoekstoestel gegenereerd. Dit heb ik gedaan door woorden en zinnen op het onderzoekstoestel te typen. Hierdoor kon ik meerdere, verschillende versies van het dynamic.lm bestand vergelijken met Dynamic V0 en Dynamic V1. Ik heb in totaal 35 verschillende versies van het dynamic.lm bestand gegenereerd. Door meerdere versies van het dynamic.lm bestand met elkaar te vergelijken heb ik het volgende kunnen constateren:

1. Het eerste deel van het bestand, de header, verandert niet. Voor het stukje met leesbare tekst staat een hexadecimaal getal dat overeenkomt met de lengte deze tekst.
2. Het tweede deel van het bestand, het deel met de compressie, wordt bij elke nieuw toegevoegde woord een stukje langer. Vóór de compressie staat een hexadecimaal getal dat overeenkomt met de lengte van het gecomprimeerde deel. Na de compressie staat een hexadecimaal getal dat overeenkomt met het aantal woorden. Na dit getal staat er voor elk woord één hexadecimaal getal vermeld. Dit getal geeft aan uit hoeveel letters het woord bestaat.
3. Het derde en vierde deel van het bestand worden ook bij elk nieuw toegevoegde woord een stukje langer. Het is niet duidelijk geworden wat voor betekenis deze delen van het bestand hebben.
4. Het vijfde deel van het bestand, het deel met de oplopende hexadecimale nummering, is ook bij elk nieuw toegevoegde woord een stukje langer geworden. De nummering loopt steeds meer op en een aantal getallen worden vaker herhaald.
5. Het laatste deel van het bestand, de end-of-file marker, verandert niet.

De aangegeven lengtes en de Little- en Big Endian telmethodes

In de eerste twee delen van het dynamic.lm bestand zijn hexadecimale getallen waargenomen die lengtes aan lijken te geven. In beiden gevallen gaat het om hexadecimale getallen die met behulp van de Little Endian telmethode vastgelegd zijn, wat betekent dat het getal begint met de laatste byte van het getal. Bij Big Endian eindigt een getal met de laatste byte van het getal (Bakker, Jasper, 2015) (GeeksforGeeks, 2022).

De Big Endian en Little Endian telmethodes zien er als volgt uit:



Figuur 19 - Voorbeelden van Big- en Little Endian telmethodes. Bron: https://www.geeksforgeeks.org/little-and-big-endian-mystery/

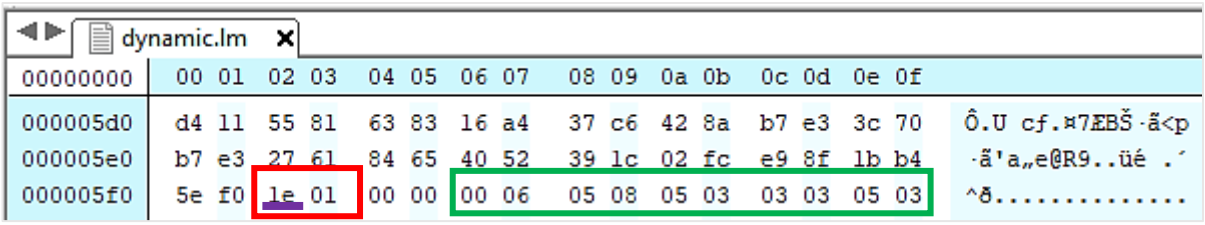
Het getal dat in mijn laatste versie van het dynamic.lm bestand voor de compressie staat is “65 05 00 00”. Zoals met behulp van bovenstaande figuur waar te nemen is, is de telmethode Little Endian gebruikt. Voor het berekenen van het einde van het gecomprimeerde deel van het bestand is het makkelijker om het getal om te zetten naar Big Endian. Het getal “65 05 00 00” wordt dan “00 00 05 65”. De positie na dit getal, waar de compressie lijkt te beginnen, is “00 00 00 8d”. De positie van het einde van het gecomprimeerde deel kan berekend worden door de lengte op te tellen bij de beginpositie, dus:

eindpositie = beginpositie + lengte
eindpositie = 0x8d (141) + 0x565 (1381)
eindpositie = 0x5f2 (1522)

Bij elke versie, behalve de laatste versie, van het dynamic.lm bestand komt de eindpositie overeen met het einde van het deel met de hoge entropie. Bij de laatste versie lijkt de eindpositie op 1 na niet overeen te komen. Hier staat namelijk het getal “01”, wat bij elke andere versie “00” was.

Het hexadecimaal getal op de eindpositie kwam ook bij elke versie, behalve bij de laatste versie, overeen met het totaal aantal woorden dat UFED PA uit het dynamic.lm bestand haalde. Zo is het laatste hexadecimale getal van het gecomprimeerd deel bij het nulpunt “4b”, wat in decimale notatie ‘75’ is. UFED PA heeft bij het nulpunt een lijst van 75 woorden gegenereerd. Het getal op de eindpositie van de laatste versie is “1e" (zie paarse markering in figuur26, wat in decimale notatie 30 is.

UFED PA heeft bij de laatste versie een lijst van 286 woorden gegenereerd. Wanneer het getal na de eindpositie als één gezien wordt met het getal op de eindpositie, ontstaat het getal “1e 01” (zie rode markering in figuur 26), wat wel overeenkomt met het aantal woorden dat UFED PA gegenereerd heeft. In de Trie komt het laatste getal namelijk op één na niet overeen met het aantal woorden in de lijst. Het is onbekend waarom het laatste getal, de “01”, niet meegenomen wordt in de aangegeven lengte.



Figuur 20 - Laatste stukje van het gecomprimeerd deel in de laatste versie van het dynamic.lm bestand (markeringen zelf toegevoegd).

Er is geprobeerd om het deel met de compressie te decomprimeren met behulp van de vijf populairste compressie algoritmes (LZ77, LZR, LZSS, deflate en LZMA). Het is hiermee niet gelukt om het gecomprimeerde deel te decomprimeren. Er is een sterk vermoeden dat de woorden uit de gebruikersdictionary in dit gecomprimeerde deel opgeslagen staan, aangezien het UFED PA lukt om woorden uit het dynamic.lm bestand weet te extraheren. Dat UFED PA de woordenlijst uit het bestand kan halen betekent dat de woorden ergens moeten staan. Door woorden gecomprimeerd op te slaan nemen ze minder plaats in beslag in het bestand, zonder dat er data verloren gaat. Er is in overleg met mijn bedrijfsbegeleider besloten dat het ontcijferen van de compressie een lagere prioriteit heeft gekregen omdat het met behulp van UFED PA mogelijk is om de woorden uit het dynamic.lm bestand te halen.

De hexadecimale waarden die overeenkomt met het totaal aantal woorden in de woordenlijst worden gevolgd door een deel met lage hexadecimale waarden en een lage entropie. Er is waargenomen dat deze waarden overeenkomen met het aantal karakters dat de woorden uit de woordenlijst bevaten. Het eerste stukje van dit deel is te zien in *Figuur 26* en is met groen gemarkeerd. De eerste woorden die in de woordenlijst staan zijn als volgt:

Woord uit de woordenlijst	Aantal karakters in het woord	Hexadecimale waarde
<leeg vak>	0	00
Donald	6	06
Trump	5	05
Sterling	8	08
AFAIK	5	05
AFK	3	03
ASL	3	03
ATM	3	03
BBIAB	5	05
BBL	3	03

Tabel 2 – Woorden uit de woordenlijst, het aantal karakters van deze woorden en de bijbehorende hexadecimale waarden.

De aantallen komen volledig overeen met alle woorden in de woordenlijst. Na de hexadecimale waarden die het aantal karakters van de woorden aangeven wordt de entropie ook weer iets hoger. Het is onbekend wat de betekenis van die waarden is.

De oplopende hexadecimale nummering en de woordenlijst van UFED PA

Het vijfde deel van het dynamic.lm bestand werd bij elk nieuw toegevoegde woord een stukje langer. Het hoogst voorkomende hexadecimale getal in de nummering kwam bij elke versie net niet overeen met het aantal woorden dat UFED PA uit het bestand wist te halen. Het aantal woorden is elke keer 1 hoger dan het hoogst voorkomende getal in de nummering. Het getal dat aan het einde van de compressie staat is, wat in bovenstaande tekst besproken is, is dus ook elke keer 1 hoger dan het hoogst voorkomende getal in de nummering. Bij het nulpunt is het getal aan het einde van de compressie “4b”, is het aantal woorden dat UFED PA heeft gegenereerd 75 en is het hoogst voorkomende getal in de nummering “4a”, wat in decimale notatie ‘76’ is.

Het woord met indexnummer 1, in de woordenlijst die UFED PA heeft gegenereerd, is leeg. Elk nieuw toegevoegde woord wordt ook altijd aan het einde van woordenlijst toegevoegd. Mijn vermoeden was de hexadecimale getallen uit de oplopende nummering een indexnummer zijn voor de woorden die in de woordenlijst staan. Om dit vermoeden te bevestigen ben ik de hexadecimale getallen gaan vertalen naar de woorden uit de woordenlijst van UFED PA. Het koppelen van het hexadecimale getal aan hetzelfde getal in de UFED PA woordenlijst leek niet goed te kloppen. Door de getallen te vertalen ontstonden er een aantal woordreeksen. De woordreeksen kwamen niet volledig overeen met de volgordes van de woorden in de zinnen die ik op het onderzoekstoestel had getypt.

da 00 02 00 00 00 **d6** 00 02 00 00 00 **72** 00 02 00 00 00 **db** 00 02 00
verkozen 02 00 00 00 **niks** 02 00 00 00 **aan** 02 00 00 00 **woord** 02 00

Figuur 21 - Stukje uit het deel met de oplopende hexadecimale nummering uit de laatste versie van het dynamic.lm bestand. Hexadecimale indexnummers vertaald naar de woorden uit de woordenlijst van UFED PA met hetzelfde indexnummer.

Hierdoor kreeg ik het vermoeden dat het eerste woord uit de woordenlijst van UFED PA misschien niet mee wordt genomen in de hexadecimale nummering. Door het indexnummer 'hexadecimaal getal + 1' te gebruiken, zouden de woorden uit figuur 27 er vertaald als volgt uitzien:

da 00 02 00 00 00 **d6** 00 02 00 00 00 **72** 00 02 00 00 00 **db** 00 02 00
woord 02 00 00 00 **van** 02 00 00 00 **het** 02 00 00 00 **jaar** 02 00

Figuur 22 - - Stukje uit het deel met de oplopende hexadecimale nummering uit de laatste versie van het dynamic.lm bestand. Hexadecimale indexnummers vertaald naar de woorden uit de woordenlijst van UFED PA met hetzelfde indexnummer +1.

Dit komt overeen met één van de zinnen die ik op het onderzoekstoestel heb getypt: "Verkozen woord van het jaar 2021: prikspijt". De waarden achter de hexadecimale indexnummers komen overeen met de frequenties van de woorden. De frequenties van de woorden zijn te zien in de woordenlijst van UFED PA. Na het proberen te valideren van deze waarneming heb ik kunnen concluderen dat de frequentie achter een woord (hexadecimaal indexnummer) niet altijd overeenkomt met de frequentie die het woord in de woordenlijst van UFED PA heeft. De frequentie van het woord "het", welke in figuur 28 te zien is, is als child van het woord 'van' 2. In de woordenlijst van UFED PA heeft dit woord een frequentie van '21'. De frequentie van een node in een reeks is afhankelijk van de andere nodes in de reeks.

De frequenties in de woordenlijst geven aan hoe vaak een woord getypt is. Dit heb ik kunnen waarnemen door verschillende versies van dynamic.lm te analyseren. Wanneer ik een woord opnieuw had gebruikt, was de frequentie hoger geworden. In het nulpunt staat achter elk getal, op het eerste getal na, in de hexadecimale nummering het getal "05 00" vermeld. In de woordenlijst van UFED PA is bij het nulpunt ook te zien dat elk woord, op het eerste woord na, een frequentie van 5 heeft. Het eerste woord zou in de hexadecimale nummering een frequentie van "0a" (decimaal '10') hebben. In de woordenlijst heeft dit woord een frequentie van 0.

	#	Word	Locale	Frequen	Usage patte	Suggested wo	Typed te	Deleted
5	2	Donald		0	Unknown			
6	2(1)	Donald		0	Unknown			
7	3	Trump		5	Unknown			
	3(1)	Trump		5	Unknown			

Figuur 23 - Schermafbeelding van een stukje van de woordenlijst van UFED PA. Indexnummers, woorden en frequenties zijn gezien. De woorden staan er dubbel in omdat UFED PA de woorden uit het originele dynamic.lm bestand en uit de back-up haalt.

Het lijkt erop dat de frequenties in de hexadecimale nummering niet overal de frequentie van hoe vaak het woord in totaal getypt is aangeeft, maar dat de frequentie aangeeft hoe vaak een woord in combinatie met de andere woorden in de woordreeks gebruikt is.

De getallen van de oplopende hexadecimale nummering en de frequenties worden ook vastgelegd met behulp van de Little Endian telmethode. Vanaf het hexadecimale getal '100' is er te zien hoe deze genoteerd wordt, omdat dit getal het eerste getal is dat gebruik moet maken van twee nummerparen. Het hexadecimale getal '100' wordt dus als "00 01", '101' als "01 01", et cetera. Het lijkt er hierdoor wel op dat er een limiet is voor het aantal woorden en de hoogte van de frequentie die opgeslagen kunnen worden. Het limiet zou dan "ff" zijn, wat in decimaal '65535' is.

De woordenlijst die UFED PA uit de laatste versie van het dynamic.lm bestand heeft gegenereerd is terug te vinden als externe Ext-V Woordenboekparser & bijbehorende bestanden. Hierin zijn de woorden die in de user dictionary van het onderzoekstoestel staan terug te vinden.

2.2.6 Parser bouwen

De woordenboekparser die ik heb gebouwd is een stukje code dat informatie extraheert uit de gebruikersdictionary van mijn onderzoekstoestel. Met deze informatie creeert de parser een tekstbestand waarin woordreeksen terug te vinden zijn en kan de parser de woorden uit een bericht vergelijken met de woorden die voorkomen in de gebruikersdictionary van het toestel.
De werking van de parser kan opgedeeld worden in 6 delen:

1. Gebruikersinput ontvangen.
2. Logbestand aanmaken.
3. Data uit het Excelbestand (één van de drie inputbestanden) extraheren.
4. Het begin van de Trie zoeken.
5. De Trie tekenen.
6. Woorden uit het tekstbericht vergelijken en het percentage overeenkomende woorden berekenen.

1. Gebruikersinput ontvangen

De parser heeft gebruikersinput nodig om zijn werk te kunnen doen. De gebruiker moet het volgende invoeren:

1. De locatie en naam van de map waar de gebruiker de outputbestanden in wil hebben.
   Voorbeeld: C:\Users\user\Desktop
2. De locatie en naam van het language model (.lm) bestand dat afkomstig is van het toestel.
   Voorbeeld: C:\Users\nabil\Downloads\dynamic.lm
3. De locatie en naam van het Excel (.xlsx) bestand dat de door UFED PA gegenereerde woordenlijst bevat.
   Voorbeeld: C:\Users\nabil\Downloads\Report.xlsx
4. De locatie en naam van het tekstbestand (.txt) waarvan de gebruiker de inhoud wil vergelijken met de woordenlijst.
   Voorbeeld: C:\Users\nabil\Desktop\Testbericht.txt

De parser controleert elke invoer op mogelijke fouten. Fouten die voor kunnen komen zijn:

- Het opgegeven pad bestaat niet.
- Het opgegeven bestand bestaat niet.
- Het opgegeven bestand beschikt niet over de juiste bestandsextensie.
- De map kan niet aangemaakt worden op de opgegeven locatie.
- De mapnaam bevat karakters die verboden zijn in de naamgeving.

2. Logbestand aanmaken

Nadat de map voor de outputbestanden is aangemaakt, wordt er in deze map een tekstbestand voor de logging aangemaakt. Alles wat de gebruiker vanaf dat moment invoert, wordt gelogd. Ook de foutieve invoeren. Bij alles wat er wordt gelogd staat er een datum (yyyy-mm-dd), tijdstip (hh:mm:ss,000), een naam van een functie en een regelnummer vermeld.

3. Data uit het Excelbestand extraheren

De parser creëert hierna twee dictionaries op basis van het ingevoerde Excel bestand:

1. Een dictionary met als key-value paren een indexnummer en een woord.
2. Een dictionary met als key-value paren een woord, met de twee values indexnummer en frequentie.

Dictionary 1 wordt gebruikt voor het vertalen van de woordreeksen en dictionary 2 voor het vergelijken van de woorden in het tekstbericht. Tijdens het creëren van de dictionaries worden niet alle regels uit het Excel bestand meegenomen omdat er duplicaten in zitten. De duplicaten ontstaan doordat UFED PA de gebruikersdictionaries van het originele dynamic.lm bestand en de back-up in één bestand verwerkt. Wanneer een woord dubbel voorkomt, heeft dit woord een '(1)' in het indexnummer staan. Alle regels waarbij het indexnummer de symbolen '(' en ')' bevat, worden niet meegenomen in de dictionaries. Er is gekozen in plaats van voor '(1)' voor '(' en ')' gekozen zodat eventuele extra duplicaten ook niet meegenomen gaan worden in de dictionaries. Ik heb echter zelf nog niet ervaren dat een woord maar één keer voorkomt of dat eenzelfde woord meer dan twee keer voorkomt. In het logbestand wordt gelogd dat de twee dictionaries aan zijn gemaakt.

Zoals in hoofdstuk 2.2.5 beschreven staat komen de indexnummers in de Trie niet overeen met de indexnummers in de woordenlijst van UFED PA. Wanneer het eerste woord in de woordenlijst niet mee wordt gerekend, lijken de woordreeksen in de Trie beter te kloppen. Het laatste woord in de woordenlijst zou anders ook niet gebruikt worden. Het eerste woord is een lege regel. Om de woordreeksen te laten kloppen, wordt er bij het toevoegen van de indexnummers aan de dictionaries voor gezorgd dat de nummers kloppend zijn voor het genereren van de woordreeksen. Om ze kloppend te maken wordt er bij elk indexnummer 1 afgetrokken (indexnummer 2 wordt 1, 3 wordt 2, et cetera).

Wanneer de gebruiker bij de gebruikersinput heeft verwezen naar een Excelbestand waar niet de juiste data in staat, stopt het programma. De dictionaries kunnen dan niet aangemaakt worden. In het logbestand wordt gelogd dat het niet mogelijk is om data uit het Excelbestand te extraheren en dat het programma is gestopt.

4. Het begin van de Trie zoeken

Als het is gelukt om de dictionaries te creëren, gaat de parser data uit het Language Model bestand extraheren om de Trie te kunnen tekenen. Hiervoor gaat de parser eerst opzoek naar de offset van de eerste waarde van de Trie ("01") omdat de Trie vanaf daar getekend zal gaan worden. In het dynamic.lm bestand staan veel "01" nummerparen dus het zoeken naar dit nummerpaar kan een onjuiste offset opleveren.

Zoals eerder besproken in *Hoofdstuk 2.2.5 Data in dynamic.lm analyseren en ontcijferen*, heb ik tijdens mijn onderzoek waargenomen dat het deel vóór de oplopende hexadecimale nummering bij elke versie van het dynamic.lm bestand eindigt met de nummerparen "06 64 6d 61 70". Tussen deze nummerparen en het begin van de nummering staan nog vier nummerparen, maar deze komen niet in elke versie overeen. "06 64 6d 61 70" is een set opvolgende nummerparen die niet op een andere plek in het bestand is voorgekomen. Ik heb ervoor gekozen om deze set te gebruiken voor het bepalen van de offset van het begin van de Trie omdat deze set opvolgende nummerparen tot nu toe elke keer uniek is geweest in de verschillende dynamic.lm versies.

Omdat er ook in elke versie van het dynamic.lm bestand vier nummerparen tussen het einde van de set en het begin van de Trie stonden, is dit gebruikt voor het bepalen van de offset van het begin van de Trie. Tussen het einde van de set en de waarde voor het begin van de Trie zitten dus drie nummerparen. Zoeken naar de string "06646d6170" levert de positie van het eerste getal van het eerste nummerpaar, dus van "0", op. Door hier 18 (9 keer 2 omdat een nummerpaar uit 2 karakters bestaat) bij op te tellen, kom je bij het eerste getal van de waarde van het begin van de Trie uit. Om hier een beeld bij te krijgen:

Hexadecimale string:	**06**	**64**	**6d**	**61**	**70**	**4c**	**00**	**00**	**00**	**01**
Posities:	0	1	2	3	4	5	6	7	8	9

Tabel 3 - Visueel beeld bij het bepalen van de positie van de eerste waarde van de Trie, op basis van een vaste set nummerparen. Het eerste getal uit de set is geel/blauw gemarkeerd en het eerste getal van de eerste waarde in de Trie is geel/rood gemarkeerd

Met behulp van de positie van het eerste getal van de waarde van het begin van de Trie kan de offset bepaald worden. De offset wordt gelogd in het logbestand. Wanneer de gebruiker bij de gebruikersinput heeft verwezen naar een Language Model bestand waar niet de juiste data in staat, stopt het programma. De offset van de eerste waarde van de Trie kan dan niet bepaald worden. In het logbestand wordt gelogd dat het Language Model bestand niet de juiste data bevat om voor het programma te gebruiken en dat het programma is gestopt.

5. De Trie tekenen

Wanneer de offset van de eerste waarde van de Trie is bepaald, wordt de Trie getekend. Zoals in hoofdstuk 2.2.4 en hoofdstuk 2.2.5 is beschreven, geven de "00" paren tussen de verschillende Trie-waarden aan op welk niveau een woord zit. De "00" paren geven aan op welke plek in de Trie een waarde zich bevindt, ten opzichte van de voorgaande waarde. Dit gaat als volgt in zijn werking:

- Twee "00" nummerparen indiceren dat waarde B een niveau onder waarde A zit.
  Wanneer waarde A de root is, wordt waarde B een structurele node die met een edge verbonden wordt aan de root. Wanneer waarde A geen root is, wordt waarde B een child van waarde A. Waarde A is dan de parent van waarde B.

- Vier "00" nummerparen indiceren dat waarde B op hetzelfde niveau als waarde A zit.
  Wanneer waarde A een structurele node is, wordt waarde B ook een structurele node. Wanneer waarde A een sub-node is, wordt waarde B een sibling van waarde A. Waarde A en B delen van dezelfde parent.

- Zes "00" nummerparen indiceren dat waarde B één niveau hoger dan waarde A zit.
  Wanneer waarde A een structurele parent-node heeft, wordt waarde B een structurele node. Wanneer waarde A een grandparent (parent van parent) heeft, wordt waarde B een sibling van de parent-node van waarde A. Waarde B is dan een uncle (sibling van de parent) van waarde A.

- Acht "00" nummerparen indiceren dat waarde B twee niveaus hoger dan waarde A zit.
  Dit komt alleen voor op plaatsen waarbij waarde A een grandparent heeft. Waarde B wordt een structurele node.

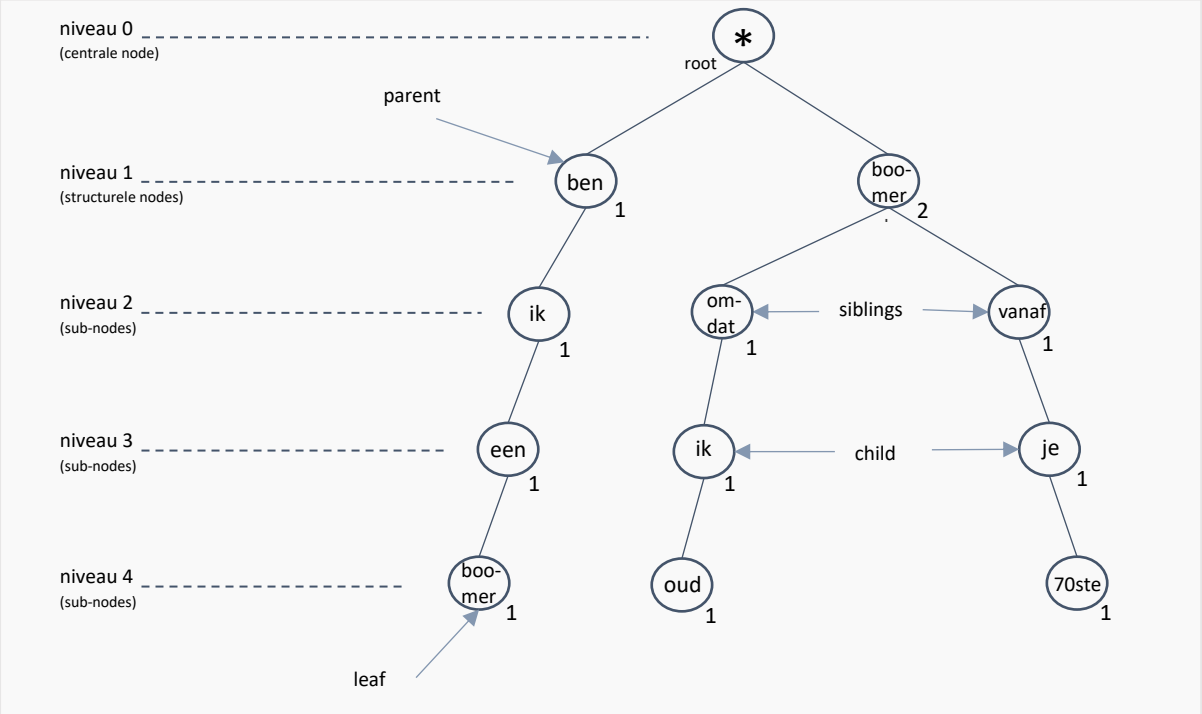
Het is (nog) niet voorgekomen dat er meer dan acht "00" nummerparen achter elkaar staan. De maximale diepte die bereikt wordt is 4 (niveau 0 t/m niveau 4).

Zoals ook in hoofdstuk 2.2.5 beschreven staat kunnen de waarden van de nodes in de Trie vertaald worden naar woorden uit de woordenlijst van UFED PA. De waarden van de nodes zijn indexnummers. Sommige nummers komen meerdere keren voor, vermoedelijk omdat een woord in meerdere woordreeksen voor kan komen. Achter elk indexnummer staat nog een waarde. Deze waarde is de frequentie van het woord waar het indexnummer naar vertaald kan worden. De frequentie komt niet altijd overeen met de frequentie die in de woordenlijst van UFED PA te zien is. Dit komt vermoedelijk doordat de frequentie gebaseerd wordt op de frequentie waarin een woord in een reeks, op die volgorde, met de andere woorden in de reeks gebruikt is. Er is bijvoorbeeld is een woordreeks die er, met de frequenties in die reeks tussen haakjes, als volgt uitziet:

root -> Ben(1) -> ik(1) -> een(1) -> boomer(1)

Het woord "ik" heeft een frequentie van 1 in deze reeks, terwijl de frequentie van dit woord in de woordenlijst van UFED PA een frequentie van 17 heeft. De frequentie van de woorden in de verschillende reeksen is dus niet een onbelangrijk detail – het zegt iets over het gebruik van deze woorden.

Wanneer er een deel van de Trie uit de laatste versie van het dynamic.lm bestand wordt getekend en de indexnummers om worden gezet in woorden (met bijbehorende frequenties), ziet dit er als volgt uit:



Figuur 25 - Een deel van de Trie uit het dynamic.lm bestand, getekend met zichtbare niveaus en aangewezen voorbeelden van child-, parent-, sibling- en leaf-nodes. De hexadecimale indexnummers zijn vertaald naar woorden met de woordenlijst van UFED PA. Bij elk woord staat een frequentie vermeld.

dynamic.lm

000031a3	00 01	02 03	04 05	06 07	08 09	0a 0b	0c 0d	0e 0f	
00005950	00 00	00 00	00 f0	00 01	00 00	00 6c	00 01	00 00	.....ð.....l....
00005960	00 6f	00 01	00 00	00 f1	00 01	00 00	00 00	00 00	.o.....ñ........
00005970	00 00	00 00	00 f1	00 02	00 00	00 f2	00 01	00 00	.....ñ.....ò....
00005980	00 6c	00 01	00 00	00 f3	00 01	00 00	00 00	00 00	.l.....ó........
00005990	00 00	00 f5	00 01	00 00	00 62	00 01	00 00	00 f6	...õ.....b.....ö
000059a0	00 01	00 00	00 00	00 00	00 00	00 00	00 f2	00 01	.............ò..

Figuur 24 - Deel van het dynamic.lm bestand waar de waarden uit de Trie uit figuur 31 in terug te vinden zijn (rood = ben ik een boomer, paars = boomer omdat ik oud, groen = boomer vanaf je 70ste)..

De parser tekent de Trie met behulp van het Language Model bestand en dictionary 1 die met behulp van het Excelbestand is gemaakt. De woordreeksen uit de Trie worden weggeschreven naar een tekstbestand genaamd 'result_predict.txt'. Aan de bestandsnaam worden de datum en tijd waarop het bestand aangemaakt wordt toegevoegd. Het tekstbestand wordt aangemaakt in de map die gecreëerd is voor de outputbestanden. In het logbestand wordt gelogd hoe het outputbestand heet, waar deze opgeslagen is, dat de parser de Trie aan het maken is en dat de woordreeksen weggeschreven worden naar het tekstbestand. Een woordreeks wordt als volgt, met bijbehorende frequenties, weggeschreven:

root -> Ben(1) -> ik(1) -> een(1) -> boomer(1)
root -> boomer(2) -> omdat(1) -> ik(1) -> oud(1)

Als een indexnummer uit de Trie niet voorkomt in de dictionary en daardoor niet vertaald kan worden naar een woord, wordt het indexnummer in het outputbestand vermeld.

6. Woorden uit het tekstbericht vergelijken & percentage overeenkomende woorden berekenen

Als laatste worden de woorden uit het door de gebruiker aangewezen tekstbestand vergeleken met de woorden in dictionary 2, die met behulp van het Excelbestand is gemaakt. Er wordt een teller bijgehouden met het aantal woorden uit het tekstbericht dat overeenkomt met woorden uit de dictionary. Wanneer alle woorden uit het tekstbericht zijn vergeleken, wordt er berekend hoeveel procent van de woorden uit het tekst overeenkomt met de woorden uit de dictionary. De berekening ziet er als volgt uit:

% overeenkomende woorden = aantal overeenkomende woorden/ totaal aantal worden * 100

Als er vier van de acht woorden uit het tekstbericht overeenkomen met woorden uit de dictionary, wordt er een percentage van 50% overeenkomende woorden gegeven. Het berekende percentage wordt samen met de overeenkomende woorden weggeschreven naar een outputbestand genaamd 'result_message.txt'. Aan de bestandsnaam worden de datum en tijd waarop het bestand aangemaakt wordt toegevoegd. Het tekstbestand wordt aangemaakt in de map die gecreëerd is voor de outputbestanden. Bij elk overeenkomend woord staat het indexnummer en de totale frequentie van het woord vermeld. Deze data wordt uit de dictionary gehaald.

In het logbestand wordt gelogd hoe het outputbestand heet, waar deze opgeslagen is, wanneer de woorden weggeschreven worden, wanneer het percentage overeenkomende woorden wordt berekend en wat het percentage is. Wanneer de gebruiker bij de gebruikersinput heeft verwezen naar een tekstbestand dat leeg is, kunnen er geen woorden worden vergeleken en kan er geen percentage worden berekend. In het logbestand wordt gelogd dat het tekstbestand leeg is en dat het niet mogelijk is om woorden te vergelijken. De zin "This file was empty. Unable to compare words if there are none." wordt in het lege tekstbestand geschreven.

Omdat dit het einde van het programma is, wordt er aangegeven dat de parser klaar is en staat er vermeld naar welke map de outputbestanden weg zijn geschreven.

2.2.7 Overige waarnemingen met betrekking tot het dynamic.lm bestand en mogelijkheden voor vervolgonderzoeken

Tijdens het genereren van de data heb ik opgemerkt dat er een hoop waarden niet mee werden genomen. Ik heb een aantal korte experimenten uitgevoerd om te kijken wanneer de woorden wel en niet mee werden genomen. Hieruit heb ik het volgende kunnen concluderen:

- Van de drie in Samsung Notes getypte notities is van geen enkele notitie de inhoud terug te vinden in de gebruikersdictionary. De titels van alle drie de notities zijn wel terug te vinden.
- Niet alle woorden van alle WhatsApp berichten die ik heb gestuurd zijn terug te vinden in de gebruikersdictionary.
- Als een woord alleen uit cijfers bestaat wordt deze niet opgeslagen. Het woord een combinatie van letters en cijfers is wordt het woord wel opgeslagen.
  “70ste" is bijvoorbeeld wel opgeslagen maar “2021” niet.
- Er wordt onderscheid gemaakt tussen hoofdletters en kleine letters. Wanneer op het toestel automatisch hoofdletters worden gebruikt aan het begin van een zin, wordt dit woord niet gezien als woord met hoofdletter.
  Bijvoorbeeld: als ik op Whatsapp de zin “Hallo, ik ben Nabila.” type, worden de woorden uit deze zin opgeslagen als ”hallo , ik ben Nabila .“
- Er wordt geen onderscheid gemaakt tussen woorden die met en zonder diakritisch teken (een teken dat boven, onder of door een letter wordt gezet) geschreven zijn. Het woord “Oekraïne” wordt bijvoorbeeld opgeslagen als “Oekraine”.
- De gebruikersnaam van het gecreëerde Snapchat-account, het e-mailadres van mijn aangemaakte Gmail account en de ingevoerde voor- en achternaam zijn terug te vinden in de gebruikersdictionary.

Ik beveel de volgende vier vervolgonderzoeken aan:

Vervolgonderzoek 1: Achterhalen of de toetsenborddata op andere Samsung toestellen op dezelfde manier opgeslagen wordt (wordt aanbevolen om als eerste uit te laten voeren)

Door hier eerst onderzoek naar te verrichten kan er op basis van de resultaten besloten worden of het verrichten van verdere onderzoeken de tijd en moeite waard is. Als het S21 model het enige model is dat de toetsenborddata op deze manier opslaat, zou het betekenen dat er waarschijnlijk voor elk nieuwe model een nieuw onderzoek verricht moet worden naar de manier waarop de data opgeslagen wordt en naar hoe de data geëxtraheerd kan worden. Mocht het bij meerdere Samsung modellen op dezelfde manier werken, kan het verrichten van extra onderzoek de tijd en moeite meer waard zijn.

Vervolgonderzoek 2: reverse engineering voor het achterhalen van een compressie algoritme

In het dynamic.lm bestand staat een gecomprimeerd deel, waar hoogstwaarschijnlijk de opgeslagen in terug te vinden zijn, waarvan nog niet achterhaald is welk compressie algoritme hierop is toegepast. Een mogelijkheid om de compressie te achterhalen is door hier reverse engineering voor te gebruiken. Reverse engineering, ook wel back engineering genoemd, is het ontleden van een product om hier informatie uit te halen (Hess, Brian, 2019). Er wordt dus vanuit het eindproduct terug gewerkt om erachter te komen hoe het product tot stand is gekomen. Omdat er met UFED Physical Analyzer (PA) achterhaald kan worden welke woorden er terug te vinden moeten zijn, kan dit gebruikt worden voor de reverse engineering. Reverse engineering kan veel tijd vereisten. Tijdens mijn onderzoek heb ik, in overleg met mijn bedrijfsbegeleider, besloten dat ik hier niet aan zou gaan beginnen omdat UFED PA ook een lijst van de woorden op kan leveren en hierdoor het bouwen van de parser een hogere prioriteit had dan het achterhalen van het compressie algoritme.

Vervolgonderzoek 3: met behulp van twee datasets een statistisch model opzetten

Wanneer er twee datasets zijn, kan berekend worden wat de kans is dat de gebruiker van dataset 1 en de gebruiker van dataset 2 een bericht zouden hebben getypt. De kans zou berekend kunnen worden met behulp van de woordreeksen. De uitkomst van de berekening op dataset 1 kan zo vergeleken worden met de uitkomst op dataset 2. Op die manier kan er een conclusie getrokken worden over de waarschijnlijkheid dat de gebruiker van dataset 1 of de gebruiker van dataset 2 het bericht getypt zou hebben.

Vervolgonderzoek 4: met behulp van de woordreeksen de op een toestel getypte zinnen reproduceren

Ook is het mogelijk om te proberen om getypte zinnen te reproduceren op basis van de woordreeksen. Zoals in *Hoofdstuk* ***Fout! Verwijzingsbron niet gevonden.*** te lezen was, wordt er gebruik gemaakt van de reguliere expressie '^' om het begin van een woordreeks aan te geven. Hieronder volgt een voorbeeld van een zin die ik zou kunnen reproduceren met behulp van de woordreeksen:

Eén van de zinnen die ik op het onderzoekstoestel heb getypt is:

> "Misschien helpt het als ik kortere zinnen type."

Door de automatische hoofdletter is het woord "misschien" met alleen kleine letters terug te vinden. In de woordreeksen (die ik met behulp van mijn parser heb gecreëerd) is bovenstaande zin als volgt terug te vinden:

root -> ^(53) -> misschien(1) -> helpt(1) -> het(1)
root -> misschien(2) -> helpt(1) -> het(1) -> als(1)
root -> helpt(2) -> het(1) -> als(1) -> ik(1)
root -> het(21) -> als(1) -> ik(1) -> kortere(1)
root -> als(5) -> ik(1) -> kortere(1) -> zinnen(1)
root -> ik(17) -> kortere(1) -> zinnen(1) -> type(1)
root -> kortere(1) -> zinnen(1) -> type(1) -> .(1)
root -> zinnen(1) -> type(1) -> .(1)
root -> type(1) -> .(1)

Zoals te zien is schuiven de woorden elke keer één plaats op en is er een maximum van vier woorden in een reeks. Op deze manier zouden mogelijk alle getypte zinnen gereproduceerd kunnen worden. Om te kunnen testen of dit goed werkt en tunnelvisie te voorkomen, zou de persoon die de zinnen probeert te reproduceren niet de persoon moeten zijn die de zinnen getypt heeft.

2.3 Aansluiting op de competenties

Eén van de belangrijkste, dan wel niet het belangrijkste onderdeel van het afstudeerproces is het aantonen van de beroepstaken: de A- en B-competenties. Tijdens het proces was het de bedoeling om vier A-competenties en drie, zelf gekozen, B-competenties aan te tonen. Hieronder licht ik per competentie toe welke werkzaamheden ervoor hebben gezorgd dat ik de competentie heb aangetoond:

2.3.1 A-Competentie: Onderzoek

Voorafgaand het uitvoeren van mijn vooronderzoek heb ik in mijn afstudeerplan het probleem en de doelstelling van mijn onderzoek gedefinieerd. Ik had voorafgaand mijn onderzoek weinig kennis over hetgeen waar ik mij mee bezig zou gaan houden: de toetsenbordapplicatie van Samsung. Omdat ik zelf al jaren gebruik maak van Samsung toestellen wist ik wel al hoe het voorspellen van woorden er op het toetsenbord uitziet, maar ik wist niet op welke manier de woorden en volgordes opgeslagen werden. Om te achterhalen op welke manier de woorden opgeslagen worden, was het nodig om het bestand te vinden waarin de data van de toetsenbordapplicatie opgeslagen wordt.

De doelen van het vooronderzoek waren als volgt:

- Meer kennis op te doen over Android, Samsung, Android toetsenbordapplicaties, Android woordenboeken, woordvolgordes en rooten
- Informatie verzamelen over bovenstaande onderwerpen.
- Achterhalen hoe de data van de toetsenbordapplicatie op Samsung toestellen opgeslagen wordt.

Ik heb meer kennis opgedaan, informatie verzameld en geprobeerd te achterhalen hoe en waar de data van de toetsenbordapplicatie opgeslagen wordt door in openbronnen te zoeken naar antwoorden op mijn vooraf opgestelde onderzoeksvragen. Mijn hoofdvraag, onderzoeksvragen, het probleem en de doelstelling waren al eerder te lezen onder *Hoofdstuk 2.1 Plan van aanpak*. De hoofdvraag en (hoofd)onderzoeksvragen luidde als volgt:

"Op welke manier kan er, binnen een periode van 6 maanden, aan de hand van de inhoud van de gebruikerswoordenboek bepaald worden of een bericht dat op een toestel staat geschreven is door de gebruiker of dat het bericht op een andere manier op het toestel is gekomen?"

(Hoofd)onderzoeksvragen:

1. Van welke Samsung toestellen is het, ten behoeve van de uitvoering van mijn onderzoek, bekend dat ze makkelijk te rooten zijn?
2. Wat zijn de meest gebruikte toetsenbordapplicaties voor Android toestellen?
3. Hoe zit het besturingssysteem van een Samsung (Android) toestel in elkaar?
4. Welke woorden staan er in een 'kaal' Samsung (Android) woordenboek opgeslagen?

De resultaten van mijn onderzoek waren al eerder te lezen onder *Hoofstuk 2.2 Uitgevoerde onderzoeken en experimenten.* Hier zijn verzamelen van informatie en het toepassen van geschikte methodes/kennis/inzichten/theorieën ook aan bod gekomen. Met behulp van mijn onderzoeksvragen heb ik relevante informatie voor mijn onderzoek kunnen verzamelen. Een aantal voorbeelden van methoden die ik tijdens mijn gehele onderzoek heb gebruikt zijn als volgt:

- Watervalmethode (project opsplitsen in opeenvolgende fasen)
- Sneeuwbalmethode (vanuit één publicatie zoeken naar literatuur over eenzelfde onderwerp)
- MoSCoW methode (methode om requirements in kaart te brengen)

Alle informatie voor mijn vooronderzoek is verwerkt in één document, welke terug te vinden is als externe bijlage *Ext-I Vooronderzoek*. De laatst benoemde methode, de MoSCoW methode, heb ik gebruikt tijdens het verrichten van onderzoek naar de software Hansken en het in kaart brengen van de

requirements voor de implementatie van mijn woordenboekparser. Over de B-competenties is op een later moment in dit verslag meer te lezen.

2.3.2 A-Competentie: Leren

Door vooronderzoek te verrichten naar de Samsung toetsenbordapplicatie voelde ik mij meer bekwaam om de afstudeeropdracht uit te kunnen voeren. Van tevoren wist ik niet waar ik (naast het vooronderzoek) zou moeten beginnen met kijken. Na het zelfstandig uitvoeren van het vooronderzoek ben ik meer te weten gekomen over het onderwerp. Hierna heb ik ook zelfstandig aan de andere fases van mijn onderzoek kunnen beginnen. Ik ben hier erg gemotiveerd mee bezig omdat ik benieuwd ben naar de werking en uitkomst – het zou een mooie toevoeging op de software kunnen zijn. Dat het een bijdrage kan gaan leveren is ook een erg motiverende factor geweest voor mij.

Naast mezelf te blijven motiveren probeer ik dit bij de andere stagiaires ook te doen. Door regelmatig te vragen hoe het gaat en waar ze mee bezig zijn delen ze vol enthousiasme de voortgang van hun opdrachten. Als er iemand vastloopt kijk ik ook graag mee om diegene te proberen te helpen en om ze te blijven motiveren. Dit gebeurt andersom ook. De stagiaires moedigen mij ook aan om door te blijven gaan, ook al lukt iets niet altijd. Het is fijn om te zien hoe alle stagiaires elkaar aanmoedigen.

2.3.3 A-Competentie: Professioneel werken

Er is contact gelegd met de stakeholders voor het uitvoeren van de infrastructuur- en software analyse. Dit is gedaan om alle relevante onderdelen met betrekking tot de ICT infrastructuur in kaart te brengen, om de requirements met betrekking tot het implementeren van software in Hansken in kaart te brengen en om rekening te kunnen houden met de wensen/eisen met betrekking tot mijn woordenboekparser.

Om zo gestructureerd mogelijk te blijven werken heb ik voorafgaand het onderzoek een werkwijze, planning en een Product Breakdown Structure (PBS) opgesteld. De werkwijze is met behulp van de watervalmethode opgesteld. De watervalmethode zorgt voor gestructureerde, opeenvolgende onderzoeksfasen. In de planning zijn alle deliverables en bijbehorende activiteiten terug te vinden. In de PBS is te zien hoe de woordenboekparser in zijn werking gaat. Met behulp van de watervalmethode, planning en PBS kon ik methodisch en planmatig te werk gaan. De werkwijze, planning en PSB zijn terug te vinden onder *Hoofdstuk 2.1 Plan van aanpak*.

Aan het eind van de afstudeerperiode zal ik een presentatie verzorgen voor onder andere de stakeholders. Tijdens deze presentatie zal ik een advies uitbrengen over de door mij gebouwde software, de woordenboekparser. Er is gekeken naar de betrouwbaarheid van de parser, welke gemeten is met een statistisch model. Het statistisch model is vastgelegd in een extern document welke terug te vinden is onder *Hoofdstuk 5. Externe Bijlagen* als *Ext-IV Adviesraportage Woordenboekparser*. Er is meer over dit hoofdstuk te lezen in *Hoofdstuk 2.3.7 B-Competentie: Software adviseren*. De rapportage heeft een professionele opzet, is buiten de beoordeling van de afstudeerstage ook bruikbaar, bevat zo min mogelijk fouten en bevat een zakelijke en helder taalgebruik.

Ik heb gereflecteerd op mijn afstudeerproces, het product en de competenties. De reflecties zijn te lezen onder *Hoofdstuk 3. Reflectie*

2.3.4 A-Competentie: Innovatie

Het doel van mijn opdracht was het bepalen of mogelijk is om met behulp van de gebruikersdictionary aan te tonen of een gebruiker van een toestel wel of niet een bericht kan hebben getypt. Ik merkte dat ik veel tijd verloor in het proberen te achterhalen welke compressie er in het bestand zit. Zoals eerder benoemd is in *Hoofdstuk 2. Het proces* is er een sterk vermoeden dat de woorden die opgeslagen zijn in de gebruikersdictionary gecomprimeerd opgeslagen in het dynamic.lm bestand. Hierdoor ben ik op zoek gegaan naar andere oplossingen zodat mijn woordenboekparser een bijdrage kan geven aan de doelstelling.

Voordat ik op zoek ben gegaan naar een andere oplossing heb ik mijn probleem met meerdere collega's en mede-stagiaires besproken om ideeën van anderen uit te proberen en heb ik openbronnen geraadpleegd. Zo heb ik bijvoorbeeld, zoals eerder benoemd is in *Hoofdstuk 2.2 Uitgevoerde onderzoeken en experimenten*, met de commandline analyse tool Binwalk geprobeerd te achterhalen welke compressie er toegepast was op het bestand. Geen van de ideeën/gebruikte tools heeft geleid tot het achterhalen van de compressie.

In plaats van de opgeslagen woorden direct uit het dynamic.lm te halen, is het ook mogelijk om een kopie van een deel van het toestel in te laden in UFED Physical Analyzer (PA). UFED PA kan, zoals ook eerder benoemd is in *Hoofdstuk 2.2 Uitgevoerde onderzoeken en experimenten*, de gebruikersdictionary uit het dynamic.lm bestand halen. De gebruikersdictionary kan geëxporteerd worden naar verschillende bestandsformaten, waaronder .xlsx (een bestandsformaat dat onder andere geopend kan worden met behulp van Microsoft Excel).

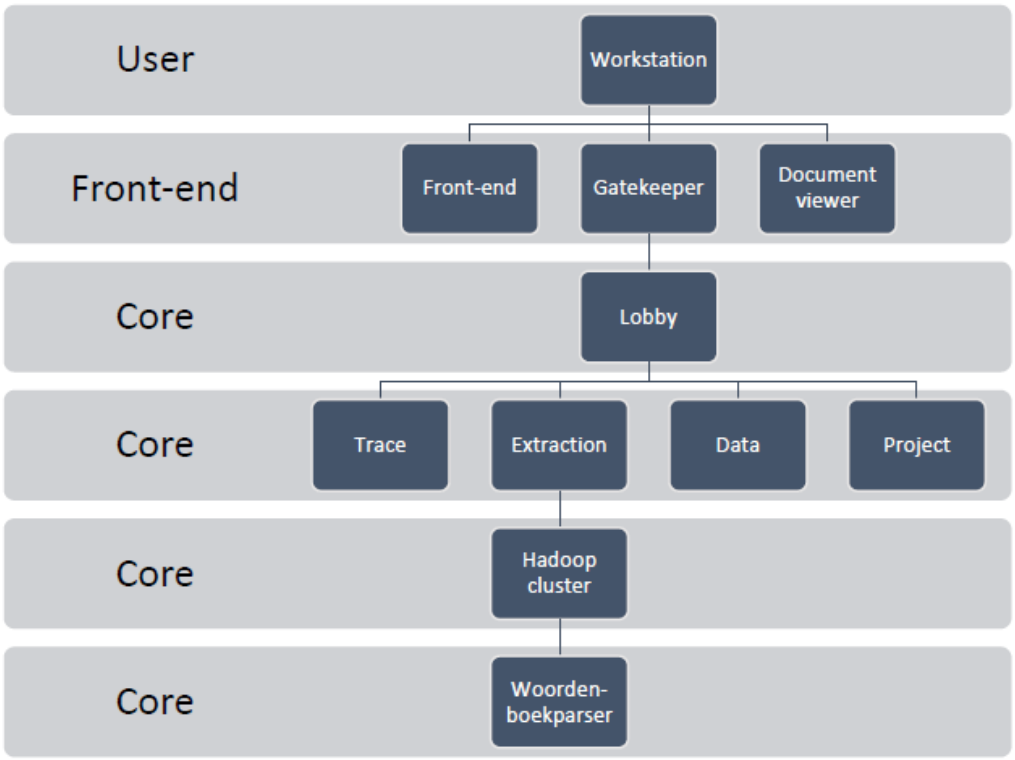
Voor mijn woordenboekparser maak ik hiermee gebruik van drie inputbestanden in plaats van twee. Met behulp van de met UFED PA geëxporteerde woordenlijst is het alsnog mogelijk om aan te tonen in welke volgordes de woorden gebruikt zijn. Toen ik mijn idee aan het uitwerken was heb ik mijn bedrijfsbegeleider verteld over mijn idee om het probleem mee aan te pakken.

2.3.5 B-Competentie: Infrastructuur analyseren

Zoals eerder vermeld is, is het de bedoeling dat mijn woordenboekparser uiteindelijk geïmplementeerd gaat worden in Hansken. Ik heb een requirementsanalyse opgesteld om de ICT infrastructuur van Hansken te analyseren en om nieuwe ontwikkelingen met betrekking tot ICT infrastructuur in kaart te brengen. Ik heb hierbij eerst gekeken naar de infrastructuur van Hansken en waar mijn woordenboekparser geimplementeerd zou gaan worden. Vervolgens heb ik een deskresearch uitgevoerd en heb ik hierbij zes ontwikkelingen en trends die gericht zijn op een infrastructuur van de toekomst in kaart gebracht. Tot slot heb ik de zes trends en ontwikkelingen die ik tijdens mijn deskresearch in kaart heb gebracht vergeleken met hoe deze trends/ontwikkelingen wel of niet terug te vinden zijn binnen Hansken.

De infrastructuur van een Hansken workstation kan opgedeeld worden in drie lagen: user, front-end en core. In de “user” laag zit het werkstation. Dit is de plek waar de gebruiker inlogt om gebruik te kunnen maken van Hansken. Het werkstation bevat drie onderdelen die onder de “front-end” vallen: een front-end, gatekeeper en document viewer. De front-end is de voorkant van de applicatie, de gatekeeper is een module die communiceert met de buitenwereld en ervoor zorgt dat gebruikers niet in de applicatie kunnen komen zonder authenticatie en de document viewer zorgt ervoor dat documenten bekeken kunnen worden. Aan de gatekeeper hangt de lobby. De lobby valt onder de “core” laag en is verantwoordelijk voor het aanroepen van de functies in de juiste volgorde.

De lobby wordt gevolgd door vier onderdelen die ook onder de “core” vallen: trace, extraction, data en project. Trace is verantwoordelijk voor de bewaring en het ontvangen van sporen, extraction is verantwoordelijk voor het analyseren van de data en het extraheren van sporen uit de data, data is verantwoordelijk voor het verkrijgen van data uit images (één op één kopieën van opslagmedia, bijvoorbeeld een laptop) en project is verantwoordelijk voor de bewaring van informatie met betrekking tot de images en de zaken (ook wel projecten genoemd) (van Beek, et al., 2015) (Bakker & Creeten, 2022). Bovenstaande onderdelen van de infrastructuur zien er uitgetekend als volgt uit:



Figuur 26 - ICT infrastructuur Hansken + implementatie van mijn woordenboekparser (het figuur toont alleen de onderdelen die in de lijn liggen van de infrastructuur waarin mijn woordenboekparser terecht komt. Het figuur weergeeft dus niet de volledige infrastructuur van Hansken)

De trends en ontwikkelingen die ik in kaart heb gebracht voor een ICT infrastructuur van de toekomst luiden als volgt:

1. Remote wordt de standaard
2. Meer snelheid
3. Groeiende businesswaarde van applicaties
4. Zelflerende netwerken
5. Security wordt steeds belangrijker
6. Flexibele IT'ers

Voor het opstellen van een requirementsanalyse kan er gebruik gemaakt worden van de MoSCoW methode, een methode waarmee vereisten opgedeeld kunnen worden in vier categorieën: "must haves", "should haves", "could haves" en "won't/would haves". Bovenstaande trends en ontwikkelingen zijn belangrijk om mee te nemen in de toekomst, maar zijn geen vereisten. Wanneer je een ICT infrastructuur van de toekomst op wil stellen, zouden bovenstaande trends en ontwikkelingen als "should haves" beschouwd moeten worden. Dit betekent dat het wel gewenst is om deze zaken te hebben, maar dat het niet vereist is om deze zaken te hebben om iets te laten werken.

Na het bekijken of Hansken meegaat in bovenstaande trends en ontwikkelingen, heb ik kunnen concluderen dat het Hansken team rekening houdt met alle genoemde trends en ontwikkelingen. Sinds de thuiswerk coronamaatregel is het mogelijk om Hansken remote te gebruiken. Of het voor een gebruiker gewenst is om Hansken vanaf bijvoorbeeld de thuislocatie te gebruiken, kan per gebruiker verschillen. Zaakdata is zeer vertrouwelijk en gevoelig en kan data bevatten die je thuis liever niet ziet.

Aan meer snelheid is gedacht bij het maken van een verbeterde versie van de voorganger van Hansken, Xiraf. Voor de gebruikers is het van belang om de data zo snel mogelijk te kunnen verwerken en analyseren. Hansken is ten opzichte van zijn voorganger goed vooruit gegaan op snelheid. Ook kan Hansken eigen gemaakt worden door de gebruiker doordat de gebruikers zelfs implementaties kunnen toepassen. Dit draagt bij aan de groeiende businesswaarde van applicaties waarin het gewenst is dat een applicatie op maat gemaakt kan worden voor een gebruiker en/of de organisatie. Bij Hansken kan de gebruiker hier zelf voor zorgen. Binnen Hansken zou het mogelijk zijn om met behulp van een zelflerend netwerk de herkenning van bepaalde voorwerpen/objecten in afbeeldingen te verbeteren, echter is dit niet toegestaan. Er is geen toestemming om afbeeldingen uit zaken te gebruiken voor een betere herkenning van voorwerpen/objecten.

Binnen Hansken is de nummer 1 prioriteit de beveiliging. Alle hoeken van de CIA (Confidentiality, Integrity en Availability) driehoek worden omvat in de beveiliging van Hansken. Hiermee voldoet Hansken ook aan ISO 27001, de ISO standaard voor informatiebeveiliging (van Beek, et al., 2015). Het is echter niet effectief om met flexibele IT'ers in het Hansken team te werk te gaan. Het Hansken team bestaat uit forensische softwareontwikkelaars die in subteams te werk gaan. Op die manier kunnen ze Hansken zo goed mogelijk blijven ontwikkelen. De collega's kunnen elkaar altijd raadplegen.

De volledige infrastructuur analyse is terug te vinden in een extern document, onder *Hoofdstuk 5. Externe Bijlagen* als *Ext-II Requirementanalyse Infrastructuur Hansken*.

2.3.6 B-Competentie: Software analyseren

Naast een requirementsanalyse voor de ICT infrastructuur heb ik ook een requirementsanalyse voor het implementeren van plug-ins in Hansken opgesteld. Mijn woordenboekparser, de woordenboekparser, zou als plug-in geimplementeerd worden in Hansken. Tijdens de analyse heb ik de kwaliteitsnormen en de vereisten met betrekking tot het implementeren van een Python plug-in in Hansken in kaart gebracht. Hiernaast heb ik ook de vereisten met betrekking tot de werking van mijn woordenboekparser in kaart gebracht. Alle vereisten zijn in kaart gebracht met behulp van de MoSCoW methode. Ook is er een acceptatiecriteria gedefinieerd op basis van een risicoanalyse, zijn de gewenste en huidige situatie met betrekking tot het implementeren van plug-ins in kaart gebracht en is er gekeken naar de forensische werking van Hansken.

Gebruikers kunnen zelf plug-ins implementeren in Hanksen. Hierdoor kan de applicatie, zoals eerder benoemd in *Hoofdstuk 2.3.5 B-Competentie: Infrastructuur analyseren* eigen gemaakt worden voor een gebruiker en/of een organisatie. Het implementeren van een in Python geschreven plug-in is momenteel alleen nog maar mogelijk in de bètaversie van Hansken. Een plug-in kan als extractie plug-in geimplementeerd worden of kan een plug-in gebruikt worden met Hansken.py, een Python client van Hansken. Het verschil tussen beiden is als volgt:

- Wanneer een plug-in als extractie plug-in geimplementeerd wordt, wordt deze toegevoegd aan de Hansken tools-lijst en kan de plug-in door jou en andere gebruikers (wanneer zij al gebruik maken van de bètaversie) die hem inschakelen gebruikt worden tijdens extracties.
- Wanneer een plug-in gebruikt wordt met Hansken.py, kan de gebruiker een plug-in uitvoeren op een project dat al geëxtraheerd is.

Het gebruik van een extractie plug-in die niet door het Hansken team is geschreven wordt beschouwd als third-party extractie plug-in. Het gebruik hiervan is dan ook geheel op eigen risico. Het gebruiken van een plug-in met Hansken.py is een goede manier om een plug-in tijdens de ontwikkelfase te testen. Het ontwikkelproces verloopt op die manier sneller dan door de plug-in elke keer te moeten implementeren wanneer je deze wil testen. Het is echter alleen nuttig om tijdens de ontwikkelfase van de plug-in gebruik te maken van Hansken.py. Het uitvoeren van een plug-in tijdens de Hansken extractie verloopt sneller dan het uitvoeren van een plug-in op een al bestaande extractie.

Omdat ik mijn woordenboekparser in Python heb geschreven, heb ik rekening moeten houden met de PEP-8 codeerstijl. PEP-8 is een codeerstijl die gehandhaafd wordt voor scripts die in Python worden geschreven. Het is een stijlgids die opgesteld is door Guido van Rossum, een Nederlandse informaticus die de programmeertaal Python heeft ontworpen (van Rosusm, Guido, n.d.). De stijlgids bevat richtlijnen voor het schrijven van code in Python. De richtlijnen zorgen voor een betere leesbaarheid van de code en zorgt voor wereldwijze consistentie. Er zijn wel een aantal goede redenen om één of meerdere richtlijnen te negeren:

- Het toepassen van de richtlijn maakt de code minder goed leesbaar, ook voor personen die gewend zijn om code met deze richtlijnen te lezen.
- Om consistent te blijven met omliggende code. De omliggende code zou echter ook aangepast kunnen worden om de codeerstijl de richtlijnen te handhaven.
- Omdat de code vóór de codeerstijl is opgesteld en er geen andere reden is om wijzigingen aan te brengen aan de code.
- Wanneer de code compatibel moet blijven met oudere Pythonversies die door de stijlgids aanbevolen functies niet ondersteunen.

In de stijlgids staan een hoop richtlijnen. De volledige stijlgids is terug te vinden onder *Hoofdstuk 6. Interne Bijlagen* als *Ext-VI PEP-8 Style Guide*. Hieronder zijn de meest relevante richtlijnen voor mijn woordenboekparser samengevat terug te vinden:

1. Layout: laat een regel maximaal 79 karakters bevatten.
2. Layout: gebruik voor elke import een eigen regel, voeg ze niet samen en zet ze bovenaan.
3. Quotes: enkele en dubbele quotes zijn beiden toegestaan. Hier is geen aanbeveling voor. Houd het zelf consistent.
4. Witte ruimtes: vermijd ze zoveel mogelijk (bijvoorbeeld voor en tussen alle vormen van haakjes en voor komma's, dubbele punten en puntkomma's).
5. Haakjes: het gebruik hiervan is vaak optioneel maar wordt wel aanbevolen.
6. Opmerkingen: een opmerking op een regel, na een stukje code is onnodig en afleidend. Het kan soms handig zijn om dit wel ergens te doen, maar niet overal.
7. Naamgeving: probeer de volgende letters te vermijden als variabelen om verwarring te voorkomen: I (L), O (o) en I (i). Gebruik liever hoofdletter 'L', kleine letter 'o' en kleine letter 'i'.
8. Naamgeving: package- en modulenamen zouden kort en in kleine letters geschreven moeten zijn. Er zou een '_' gebruikt kunnen worden voor de leesbaarheid, maar liever niet.
9. Naamgeving: namen van klassen zouden er als volgt uit moeten zien: klasseNaam.
10. Naamgeving: namen van variabelen zouden kort moeten zijn en er als volgt uit moeten zien: VarNaam.
11. Programmeren: de code zou niet nadelig moeten worden in andere implementaties van python (bijvoorbeeld PyPy, Jython of IronPython). Gebruik in plaats van 'None' 'is' of 'is not' en gebruik in plaats van 'not .. is' 'is not'.

De vereisten kunnen met behulp van de MoSCoW methode opgedeeld worden in vier categorieën:

1. Must haves: de zaken die essentieel zijn voor het eindproduct.
2. Should haves: de aanvullende en zeer gewenste eisen.
3. Could-haves: de aanvullende, optionele eisen voor als er tijd over is.
4. Won't/Would-haves: de wensen die vaak niet mogelijk zijn/te veel tijd vergen.

Hieronder zijn de vereisten voor het implementeren van een stukje software als plug-in in de Bètaversie van Hansken en de vereisten met betrekking tot de werking van mijn woordenboekparser per categorie terug te vinden:

	Implementatie Hansken	**Werking woordenboekparser**
Must-haves	- Pythonversie 3.6, 3.7 of 3.8. - Java 11 (voor het uitvoeren van het test-framework, welke geimplementeerd zijn in Java) - Hansken extractie plug-in (Python Extractie Plug-in SDK)	- Op basis van de woordenlijst en een bericht kan er bepaald worden of het bericht wel of niet door een gebruiker zou zijn getypt
Should-haves	- Docker (nodig voor testen). - Gzip (uitpakken van testdata). - Tox (voor het bouwen en testen van een Python plug-in code)	- De code kan aangeven op welke manier bepaald is of een bericht wel of niet door een gebruiker zou zijn getypt
Could-haves	- Niet van toepassing	- Decompressie van de woorden in het dynamic.lm bestand
Won't/would-haves	- Niet van toepassing	- Klaar zijn om geimplementeerd te worden in Hansken

Tabel 4 - Requirements met betrekking tot software

Voor het definiëren van de acceptatiecriteria voor de implementatie van mijn woordenboekparser als plug-in in Hansken is er een risicoanalyse uitgevoerd. Elk risico heeft een eigen risiconummer, beschrijving een preventie en/of oplossingsmogelijkheid en twee getallen: één getal geeft aan hoe groot het gevolg van het risico is (weging risico) en één getal geeft aan hoe hoog de kans op het risico is (weging kans). Het gevolg en de kans kunnen een nummering van 1 t/m 5 hebben, waarbij 1 het kleinste/laagste en 5 het grootste/hoogste is. De risico's en de bijbehorende risicomatrix zijn hieronder terug te vinden:

Risico #	Risicobeschrijving	Preventie/ oplossingsmogelijkheden	Weging risico	Weging kans
R001	De plug-in van de woordenboekparser werkt niet meer naar behoren in een nieuwe Python versie.	Niet te voorkomen. Wijzigingen aanbrengen in de code van de woordenboekparser.	4	3
R002	Door een verandering in de manier waarop de toetsenborddata opgeslagen wordt is de woordenboekparser niet meer te gebruiken.	Niet te voorkomen. Het nieuwe databestand onderzoeken, code hergebruiken als er gelijkenissen zijn.	5	3
R003	De plug-in van de woordenboekparser heeft weinig toegevoegde waarde aan Hansken.	De woordenboekparser is gericht op een Samsung toetsenbord. De woordenboekparser uitbreiden en compatibel maken met meerdere toetsenborden maakt de parser waardevoller.	2	1
R004	Een gebruiker met een ander systeem dan Linux probeert gebruik te maken van de woordenboekparser als plug-in.	Het gebruik van Python plug-ins is alleen getest op Linux en zou op een ander systeem misschien nog niet naar behoren werken. De plug-in kan wel gebruikt worden met behulp van Hansken.py.	1	5
R005	Een gebruiker met de officiele versie van Hansken probeert gebruik te maken van mijn woordenboekparser als extractie plug-in.	In de officiele versie van Hansken is momenteel alleen mogelijk in de bètaversie van Hansken. De plug-in kan wel gebruikt worden met behulp van Hansken.py.	1	4

Tabel 5 - Risico's risicomatrix (volledige uitwerking van de preventie en oplossingsmogelijkheden is terug te vinden in de volledige requirementsanalyse, welke toegevoegd is als externe bijlage Ext-III Requirementanalyse Hansken & Woordenboekparser).

Op basis van bovenstaande in kaart gebrachte risico's is de risicomatrix opgesteld. In de risicomatrix zijn vijf verschillende kleuren terug te vinden die aangeven hoe hoog een risico is. De legenda van de kleuren in de risicomatrix en de risicomatrix zelf zijn hieronder terug te vinden:

verwaarloosbaar	laag	gemiddeld	hoog	extreem hoog

Tabel 6 - Legenda risicomatrix

Gevolg (klein <- naar -> groot)

Tabel 7 – Risicomatrix (opgesteld op basis van de Risicomatrix van scribbr.nl (Benders, Lou, 2020))

De risico's waar ik geen controle op heb wegen het zwaarste. Voorkomen is misschien niet mogelijk, maar de problemen kunnen wel opgelost worden door de code van de plug-in aan te passen wanneer dit nodig is. Momenteel is het in de huidige, officiële versie nog niet mogelijk om Python plug-ins te gebruiken en/of te implementeren. Dit is momenteel alleen mogelijk met de bètaversie van Hansken en is alleen nog maar getest op Linux systemen. Gebruikers die geen gebruik maken van de bètaversie en op een andere machine dan Linux werken zouden mijn woordenboekparser (mogelijk) nog niet als extractie plug-in kunnen gebruiken. Zij kunnen de plug-in wel gebruiken met behulp van Hansken.py en uitvoeren op een al uitgevoerde extractie.

Het is gewenst dat het uiteindelijk mogelijk is om in een officiële versie van Hansken gebruik te kunnen maken van plug-ins die in diverse programmeertalen geschreven zijn. Voor nu is het voor in ieder geval elke gebruiker van Hansken mogelijk om plug-ins die in de programmeertaal Java zijn geschreven te implementeren en te gebruiken.

Voor de forensische werking van Hansken heb ik gekeken naar uitspraken die in de rechtspraak over Hansken zijn gedaan en heb ik de software vergeleken met de tool UFED Physical Analyzer (PA), een tool waarin het ook mogelijk is om plug-ins te implementeren. Zoekterm "Hansken" komt voor in 54 gepubliceerde rechterlijke uitspraken. De rechterlijke uitspraken laten zien dat Hansken volgens deskundigen geen onbetrouwbare gegevens produceert en dat de gegevens die onderzocht worden niet worden aantast. De grootste verschillen tussen Hansken en UFED PA zijn als volgt:

	Hansken (officiële versie)	Hansken (bètaversie)	Cellebrite 's UFED Physical Analyzer
Mogelijkheid om Python plug-in uit te voeren tijdens een extractie		X	
Mogelijkheid om een eigen Java plug-in te implementeren	X	X	
Snelle verwerking van grote hoeveelheden data	X	X	
Te gebruiken voor iedereen (ook buiten opsporingsdiensten en wetenschappelijke instituten)			X

Tabel 8 - Gedeelte van de vergelijking van Hansken & UFED Physical Analyzer

De volledige software requirementsanalyse is terug te vinden in een extern document, onder *Hoofdstuk 5. Externe Bijlagen* als *Ext-III Requirementanalyse Hansken &* Woordenboekparser .

2.3.7 B-Competentie: Software adviseren

Voor het uitbrengen van een advies over mijn woordenboekparser heb ik een adviesrapport opgesteld. In dit rapport staan onder andere de werking van mijn parser, de forensisch correcte werking van mijn parser, de inhoud van het bestand waar mijn parser informatie uit haalt, een aantal oplossingsrichtingen/mogelijkheden en een advies beschreven.

De inhoud van het bestand waar mijn parser informatie uit haalt is het bestand genaamd 'dynamic.lm'. Dit bestand bevat de toetsenborddata van mijn onderzoekstoestel. De inhoud en waarnemingen met betrekking tot dit bestand staan beschreven in *Hoofdstuk 2.2.4 Data in dynamic.lm inzichtelijk maken* en *2.2.5 Data in dynamic.lm analyseren en ontcijferen*. De werking van mijn parser staat beschreven in *Hoofdstuk 2.2.6 Parser bouwen*.

Tijdens mijn Software Analyse heb ik onder andere beschreven waar mijn woordenboekparser aan moet voldoen (de requirements) voor een goede werking, waar mijn woordenboekparser aan moet voldoen om hem te kunnen implementeren in Hansken en wat de huidige en gewenste situatie van Hansken met betrekking tot het implementeren van scripts is. Dit is ook terug te lezen onder *Hoofdstuk 2.3.6 B-Competentie: Software analyseren*.

Bij het bouwen van mijn parser heb ik rekening gehouden met de requirements die ik had opgesteld in mijn Software Analyse. Voor de implementatie van de parser in de bètaversie van Hansken is het een Must dat de parser in Pythonversie 3.6, 3.7 of 3.8 werkt en dat de Hansken extractie plug-in geïmplementeerd is. Voor de werking van de parser is het vereist dat deze op basis van een bericht kan bepalen of het bericht wel of niet door een gebruiker getypt zou zijn. Er is geprobeerd om rekening te houden met de opgestelde risico's die voor kunnen vallen bij de implementatie en het gebruik van mijn parser. Het voorvallen van vier van de vijf risico's ligt buiten mijn controle. De risico waar ik controle op heb is risico R003. Dit gaat over de toegevoegde waarde van mijn parser. Ik heb geprobeerd de parser zo waardevol mogelijk te maken door zoveel mogelijk informatie uit het 'dynamic.lm' bestand te extraheren, zodat de parser in de toekomst ook nog waardevol kan zijn. De toekomstige ontwikkeling van de officiële versie van Hansken zal zijn dat het ook mogelijk gaat worden om gebruik te maken van Python extractie plug-ins. Aangezien mijn parser in Python is geschreven, zal de parser in de toekomst ook nog gewoon gebruikt kunnen worden.

De parser kan nog waardevoller worden door de betrouwbaarheid van de parser te testen met behulp van een statistisch model. Het was de bedoeling dat ik een statistisch model op zou zetten voor mijn parser, maar dit heb ik uiteindelijk niet kunnen doen omdat ik maar één dataset had om mee te werken. Wanneer er maar met één dataset en statistische berekening wordt gedaan, heeft de waarde die uit de berekening komt geen betekenis. Er kan betekenis aan een waarde worden gegeven wanneer deze vergeleken kan worden met een soortgelijke waarde. Het opzetten van een statistisch model is één van mijn adviezen voor vervolgonderzoek.

Het enige wat mijn woordenboekparser op dit moment berekent is een percentage van het aantal woorden uit het bericht dat voorkomt in de gebruikersdictionary. De woorden in het bericht worden vergeleken met de woorden in de gebruikersdictionary. Wanneer een gebruiker een bericht op zijn/haar toestel heeft getypt, zou je een percentage van 100% overeenkomende woorden verwachten omdat elk getypte woord opgeslagen wordt in de gebruikersdictionary. Omdat ik tijdens mijn onderzoek heb waargenomen dat niet alle door mij getypte zinnen terug te vinden zijn in het dynamic.lm bestand, kan er op basis van dit percentage geen geldige conclusie worden getrokken. Dat het percentage lager dan 100% is hoeft dus niet te betekenen dat het bericht niet getypt kan zijn door de gebruiker, aangezien niet altijd alles opgeslagen wordt in de gebruikersdictionary. Het percentage dat mijn parser berekend wordt net als de woordreeksen en de logging weggeschreven naar een tekstbestand.

Voor het controleren van mijn tool op forensische werking heb ik gekeken naar zaken die zorgen voor transparantie en integriteit bij het werken met een tool. De volgende zaken zorgen voor transparantie en integriteit (INFOSEC, 2021) (NFI, 2022)

1. Data zou niet gemanipuleerd moeten kunnen worden.
2. Sporen zouden herleidbaar moeten zijn.
3. Het zou duidelijk moeten zijn welke handelingen er zijn uitgevoerd.

Door logging toe te voegen aan mijn woordenboekparser werkt mijn parser volgens bovenstaande punten forensisch correct. De parser brengt geen wijzigingen aan in de bestanden die als input worden gebruikt, houdt een logbestand bij waarin de ingevoerde bestandsnamen en de handelingen die de parser uitvoert worden gelogd, beschikt over transparante code en maakt gebruik van tijdstempels. Bij elk item dat gelogd wordt, staat een tijdstempel vermeld. Er kan dan terug gelezen worden op welk moment een handeling is uitgevoerd en er kan bekeken worden hoe lang het uitvoeren van een handeling heeft geduurd. De outputbestanden die mijn parser aanmaakt beschikken ook allemaal over een tijdstempel. De tijdstempel is bij de outputbestanden terug te vinden in de naam van het bestand.

Met behulp van mijn verslaglegging over het dynamic.lm bestand, kan een gebruiker zien welke data zich op welke plaats in het dynamic.lm bestand bevindt. Mijn parser bepaalt op basis van een bekende, consistente string vanaf welke positie de Trie getekend kan worden. De string kan als consistent worden beschouwd omdat de string in de 35 verschillende versies die ik van het dynamic.lm bestand heb gemaakt elke keer hetzelfde was. Omdat er op basis van het dynamic.lm bestand alleen een Trie wordt getekend, wordt er maar met één deel van het gehele bestand gewerkt. De andere delen worden niet gebruikt omdat sommige delen nog verder onderzoek vereisen en omdat ik sommige delen niet nodig had voor mijn doel, het op basis van een gebruikersdictionary bepalen of een bericht wel of niet door een gebruiker getypt kan zijn.

x

3. Reflectie

Momenteel zit ik in week 17 van het afstudeerproces, de week waarin het stagedossier opgeleverd dient te worden. Onder dit hoofdstuk reflecteer ik op de volgende zaken die betrekking hebben tot mijn afstudeerproces:

- Het gehele afstudeerproces.
- Het product dat ik heb opgeleverd.
- De begeleiding die ik vanuit mijn afstudeerorganisatie en de hogeschool heb gehad.
- De aangetoonde A- en B-competenties.
- Op mijzelf, een persoonlijke reflectie.

3.1 Afstudeerproces

Het gehele proces is naar mijn idee prima gegaan. Ik had misschien wat meer tijd in mijn reflectie moeten steken zodat ik goed zou kunnen reflecteren op onder andere het gehele proces. Ik heb veel nieuwe kennis opgestoken tijdens mijn afstudeerproces, maar heb ook gemerkt dat ik nog veel meer kan leren.

3.2 Het product

Op dit moment ben ik bezig met het bouwen van mijn parser. Ik loop ongeveer twee weken achter op de planning die ik had gemaakt. Het begin van de woordenboekparser is af, maar er moet nog een hoop mee gebeuren. Ik heb vertraging opgelopen in mijn planning omdat het vinden en het begrijpen van het te parsen bestand, het bestand waar de woorden uit het toetsenbord in opgeslagen worden, complexer was dan verwacht. Zo complex dat we bijna verder wilden gaan met een ander merk Android toestel, Motorola. Het bestand is gevonden en deels te begrijpen – over het andere deel zit waarschijnlijk een (nog) onbekende compressie.

Als ik hierop terugkijk heb ik het beginproces te makkelijk ingeschat. Naar mijn idee zou ik meer tijd kwijt zijn aan het bouwen van de woordenboekparser dan aan het vinden en begrijpen van het te parsen bestand. Het bouwen van de parser heeft uiteindelijk minder tijd gekost dan in eerst instantie ingeschat was.

3.3 Begeleiding

Mijn begeleiding is vanaf de eerste week al goed geweest, met mijn bedrijfsbegeleider en met mijn afstudeerdocent. Omdat mijn bedrijfsbegeleider in mijn eerste week in quarantaine zat, heb ik hem online gesproken en heeft hij ervoor gezorgd dat de andere stagiaire en ik toch een dag naar het NFI konden gaan, onder andere om het thuiswerkstation in te richten. Hij heeft geregeld dat wij netjes opgevangen zouden worden en dat alles wat moest gebeuren gedaan zou worden. Er was ook al een fysieke pc geregeld voor het werken op locatie. Het was fijn dat alles zo goed geregeld werd.

Ik heb mijn begeleider minstens één keer per week gesproken. Ik probeer hem zo goed mogelijk op de hoogste te houden van de stand van zaken zodat hij weet waar ik mee bezig ben. Als ik ergens tegenaan loop kan ik ook altijd binnenlopen of bellen. Ondanks dat hij druk is, is hij erg betrokken bij mijn opdracht. Ik heb niks op de begeleiding van mijn bedrijfsbegeleider aan te merken.

Mijn afstudeerdocent heb ik minder gesproken – wat ook de bedoeling is tijdens het afstudeerproces. Ik heb wel al gemerkt dat hij altijd bereikbaar is via de mail of via Teams. Hij reageert snel, wat heel prettig is. De laatste soort vorm van begeleiding die ik heb gekregen was feedback op mijn afstudeer-plan. Ik heb nuttige en leerzame feedback gehad van mijn afstudeerdocent. Zo kan ik mijn verslagen nóg beter maken. Ik heb gemerkt dat hij ook erg geïnteresseerd is in mijn opdracht en hij is ook erg

behulpzaam. Voor een eerste keer als begeleider van afstudeerders vind ik dat het, in ieder geval bij mij, heel goed gaat. Ik heb niks op de begeleiding van mijn afstudeerdocent aan te merken.

3.4 Competenties

Er zijn in totaal vier A-competenties en drie B-competenties die ik aantoon tijdens mijn afstudeerproces. De vier A-competenties heb ik in de eerste helft van de afstudeerperiode al aangetikt. Ik heb vooral veel onderzoek uitgevoerd, een hoop nieuwe dingen geleerd, ben professioneel aan het werk geweest en geïnnoveerd. Ik heb vooronderzoek uitgevoerd en onderzoek uitgevoerd naar het bestand waar de toetsenbord data in opgeslagen staat. Door het bestand te analyseren, kleine experimenten uit te voeren en verschillende versies van het bestand te vergelijken is een deel van het bestand ontcijferd. Het andere deel bevat waarschijnlijk een compressie. Dit moet nog verder onderzocht en achterhaald worden.

Tijdens de uitgevoerde onderzoeken heb ik gebruik gemaakt van diverse hulpbronnen: het internet, mijn collega's, mijn medestagiaire en mijn creativiteit. Door kleine, zelfbedachte experimenten op te stellen is het gelukt om een deel van het bestand te ontcijferen. Ik heb het professioneel werken aan kunnen tikken door methodisch onderzoek uit te voeren naar de organisatiestructuur en cultuur – welke terug te vinden zijn in mijn afstudeerplan. Tijdens het maken van mijn woordenboekparser zal ik vaker feedback vragen aan collega's om zo'n goed en net mogelijke code te schrijven.

3.5 Zelfreflectie

Ik vind dat ik redelijk goed bezig ben geweest. Ik vind dat het redelijk is omdat ik het gevoel heb dat ik meer had kunnen doen. In de eerste week had ik een planning opgesteld – welke ik op dat moment erg reëel vond. In de vijfde stageweek heeft het coronavirus mij te pakken gekregen waardoor ik er een week uit heb gelegen. Dit heeft invloed gehad op mijn planning. Ik was bang dat ik te weinig tijd zou hebben voor het schrijven van de code maar nu blijkt het dat ik te veel tijd in had gepland voor het schrijven van de code. Voor het vooronderzoek en het vinden van het juiste bestand was meer tijd nodig dan ik had verwacht. Het bestand was lastig te lokaliseren, mede omdat er op het internet bijna geen informatie over te vinden is. Mijn begeleider en ik hadden al snel een bestand op het oog, maar ik kon dit niet snel bevestigd krijgen.

Uiteindelijk bleek dat we toch goed zaten en dat voelde als een opluchting. Op naar de volgende stap. Soms vraag ik mij af of ik wel hard genoeg werk omdat ik weinig positieve resultaten zie en heb gezien. Hier wil ik verandering in aanbrengen. Ik moet ook accepteren dat de zogenaamde 'negatieve' resultaten ook resultaten zijn. Dingen uitsluiten hoort er ook bij en het zou ook niet leuk zijn om iets te doen wat veel te makkelijk is. Ik ben blij met de uitdaging. Ik moet er alleen misschien wat minder hard zijn voor mezelf.

4. Bibliografie

Abhijeet, M. (2020, 04 16). *It's official: Samsung Cloud will no longer sync your keyboard data*. Retrieved 03 02, 2022, from Sammobile: https://www.sammobile.com/news/its-official-samsung-cloud-will-no-longer-sync-keyboard-data/#:~:text=Samsung%20has%20officially%20announced%20that,2020%20running%20One%20UI%202.1.

Açıl, Sıddık. (2020, 04 06). *A Simple Trie Implementation*. Retrieved 06 09, 2022, from Medium: https://medium.com/@sddkal/a-simple-trie-implementation-e03ce49fe861

Android. (2022, 02 10). *Android 13 Developer Preview1*. Retrieved 03 01, 2022, from Developer Android: https://developer.android.com/about/versions/13

AndroidPlanet. (2022). *Android rooten: tips en trucs*. Retrieved 03 01, 2022, from Android Planet: https://www.androidplanet.nl/thema/rooten/

AndroidPlanet. (2022). *Samsung*. Retrieved 03 14, 2022, from Android Planet: https://www.androidplanet.nl/samsung/

AP. (2022). *Persoonsgegevens*. Retrieved 02 16, 2022, from Autoriteit Persoonsgegevens: https://autoriteitpersoonsgegevens.nl/nl/over-privacy/persoonsgegevens

Arici, A. (2021, 02 26). *7 of the Best Gboard Alternatives for Android Users*. Retrieved 03 02, 2022, from Make Tech Easier: https://www.maketecheasier.com/best-gboard-alternatives-android/

arvindpdmn. (2021, 03 29). *N-Gram Model*. Retrieved 06 08, 2022, from Devopedia: https://devopedia.org/n-gram-model

AsciiTable. (n.d.). *ASCII Table*. Retrieved 06 08, 2022, from Ascii Table: https://www.asciitable.com/

Bakker, C., & Creeten, C. (2022). *Hoe bouw je een zoekmachine voor criminele data?* Retrieved 05 12, 2022, from Werken voor Nederland: https://www.werkenvoornederland.nl/organisaties/ministerie-van-justitie-en-veiligheid/nederlands-forensisch-instituut/hoe-bouw-je-een-zoekmachine-voor-criminele-data

Bakker, Jasper. (2015, 08 05). *'De endian-strijd is voorbij'*. Retrieved 06 11, 2022, from Comptable: https://www.computable.nl/artikel/opinie/discussie/5442795/5213713/de-endian-strijd-is-voorbij.html#:~:text=Big%2Dendian%20is%20simpel%20gezegd,bijvoorbeeld%20%C3%A9%C3%A9n%20en%20dan%20oplopend.

Begam, Or. (2019, 08 05). *Customize and Automate Your Data Analysis with Python Code in Cellebrite Physical Analyzer*. Retrieved 05 31, 2022, from Cellebrite: https://cellebrite.com/en/customize-and-automate-your-data-analysis-with-python-code/

Benders, Lou. (2020, 05 22). *Een perfecte risicoanalyse voor jouw opdrachtgever*. Retrieved 05 29, 2022, from Scribbr: https://www.scribbr.nl/scriptie-structuur/risicoanalyse/

Broida, Rick. (2016, 06 03). *How to easily root an Android device*. Retrieved 03 01, 2022, from CNET: https://www.cnet.com/tech/services-and-software/how-to-easily-root-an-android-device/

BSC. (n.d.). *Data Structures - Tries*. Retrieved 06 11, 2022, from Btech Smart Class: http://btechsmartclass.com/data_structures/tries.html

Cellebrite. (2021). *Cellebrite*. Retrieved 12 22, 2021, from Cellebrite: https://www.cellebrite.com/en/physical-analyzer/

Cellebrite. (2022). *About Cellebrite*. Retrieved 05 28, 2022, from Cellebrite: https://cellebrite.com/en/about/

Cerato. (2021, 01 14). *Samsung Galaxy S21 128GB 5G Grijs*. Retrieved 05 27, 2022, from Cerato: https://www.cerato.be/telefonie/mobiele-telefonie/smartphones/samsung-galaxy-s21-128gb-5g-grijs/?srsltid=AWLEVJzlItIBZKSykMthmgoY-ZYKRgoMGEO9OQ3DYqkruefSaqfa3wnH-q8

Dingemans, Bas. (2021, 07 14). *Wat is een API?* Retrieved 01 08, 2022, from ProgrammeerPlaats: https://programmeerplaats.nl/wat-is-een-api/

Ellis, Scott R. (2013). *Signature File*. Retrieved 06 08, 2022, from ScienceDirect: https://www.sciencedirect.com/topics/computer-science/signature-file

Encyclo. (2007, 03 10). *Compressie*. Retrieved 06 08, 2022, from Encyclo: https://www.encyclo.nl/begrip/compressie

Encyclo. (2022, 05 23). *Statischtisch model definitie*. Retrieved from Encyclo.nl - Nederlandse Encyclopedie: https://www.encyclo.nl/begrip/statistisch_model

Felton. (2021). *Wie zijn wij?* Retrieved 04 29, 2022, from Felton: https://felton.nl/over-ons/wie-wij-zijn/

Felton. (2021). *Zes trends en ontwikkelingen met betrekking tot IT-infrastructuren*. Retrieved 04 29, 2022, from Felton: https://felton.nl/artikelen/hoe-ziet-de-ict-infrastructuur-van-de-toekomst-eruit/

FileInfo. (2022). *.LM File Extension*. Retrieved 05 28, 2022, from File Info: https://fileinfo.com/extension/lm

GeeksforGeeks. (2022, 03 18). *Introduction to Tree Data Structure*. Retrieved 06 11, 2022, from Geeks for Geeks: https://www.geeksforgeeks.org/introduction-to-tree-data-structure/

GeeksforGeeks. (2022, 03 01). *Little and Big Endian Mystery*. Retrieved 06 11, 2022, from Geeks for Geeks: https://www.geeksforgeeks.org/little-and-big-endian-mystery/

GoogleDevelopers. (2022). *Android Studio Downloads*. Retrieved 02 07, 2022, from Developer Android: https://developer.android.com/studio#downloads

GvB. (n.d.). *Forensisch feitenonderzoek & toedrachtonderzoek*. Retrieved 06 13, 2022, from GvB Integrity Services: https://www.integrity-services.nl/index.php/competenties/fraude/forensisch-feitenonderzoek-toedrachtonderzoek

Hadoop. (2017). In A. Pauk, N. Chilamkurti, S. Rho, & A. Daniel, *Intelligent Vehicular Networks and Communications* (p. Chapter 7). Elsevier. Retrieved from https://www.sciencedirect.com/topics/computer-science/hadoop

Hansken. (2022). *Hansken Academy*. Retrieved 05 31, 2022, from Hansken: https://www.hansken.nl/hansken-academy

Hansken. (2022). *Hansken Community*. Retrieved 05 31, 2022, from Hansken: https://www.hansken.nl/hansken-community

Hess, Brian. (2019, 09 09). *What Is Reverse Engineering and How Does It Work?* Retrieved 06 14, 2022, from Astro Machine Works: https://astromachineworks.com/what-is-reverse-engineering/#:~:text=Reverse%20engineering%2C%20sometimes%20called%20back,individual%20components%20of%20larger%20products.

Hildenbrand, Jerry. (2020, 02 17). *Root Your Android Phone: What is Root & How To*. Retrieved 02 23, 2022, from Android Central: https://www.androidcentral.com/root

HoC. (2022). *Organisatiestructuren Mintzberg*. Retrieved 04 05, 2022, from House of Control: https://www.house-of-control.nl/organisatiestructuren-mintzberg-configuratie-coordinatiemechansimen-organisatietypen.html

Hoffman, Chris. (2021, 08 12). *What Is an API, and How Do Developers Use Them?* Retrieved 05 17, 2022, from How to geek: https://www.howtogeek.com/343877/what-is-an-api/

Hofstede. (2022). *Country Comparison - Netherlands*. Retrieved 04 05, 2022, from Hofstede Insights: https://www.hofstede-insights.com/country-comparison/the-netherlands/

Hofstede, G. (2022). *Culture*. Retrieved 04 05, 2022, from Geert Hofstede: https://geerthofstede.com/culture-geert-hofstede-gert-jan-hofstede/6d-model-of-national-culture/

HSL. (2021, 05 12). *NFI leidt met Hogeschool Leiden studenten op om met zoekmachine Hansken te werken*. Retrieved 06 01, 2022, from Hogeschool Leiden: https://www.hsleiden.nl/actueel/nieuws/digital-forensics-en-e-discovery/nfi-leidt-met-hogeschool-leiden-studenten-op-om-met-zoekmachine-hansken-te-werken.html

Huijbregts, Julian. (2015, 10 13). *Nieuwe forensische zoekmachine van NFI is 48 keer zo snel als voorganger*. Retrieved 05 13, 2022, from Tweakers: https://tweakers.net/nieuws/105755/nieuwe-forensische-zoekmachine-van-nfi-is-48-keer-zo-snel-als-voorganger.html

INFOSEC. (2021, 01 06). *7 best computer forensics tools [updated 2021]*. Retrieved 06 13, 2022, from Resources Infosecinsitute: https://resources.infosecinstitute.com/topic/7-best-computer-forensics-tools/

InspectElement. (2022). *Wat is web scraping?* Retrieved 01 08, 2022, from Inspect Elemnt: https://www.inspectelement.nl/wat-is-web-scraping/

Jaspers, Arnout. (2021, 01 13). *Wat is entropie?* Retrieved 06 08, 2022, from NPO Kennis: https://npokennis.nl/longread/7995/wat-is-entropie#id-5127

JetBrains. (2022). *PyCharm - The Python IDE for Professional Developers*. Retrieved 01 22, 2022, from Jet Brains: https://www.jetbrains.com/pycharm/

Kessler, Gary. (2022, 05 04). *GCK'S FILE SIGNATURES TABLE*. Retrieved 06 08, 2022, from Garry Kessler: https://www.garykessler.net/library/file_sigs.html

LocalePlanet. (n.d.). *ICU Locale "Nederlands (Nederland)" (nl_NL)*. Retrieved 06 08, 2022, from LocalePlanet: https://www.localeplanet.com/icu/nl-NL/index.html

Mead, Ian. (2018, 04 24). *What is a Hex Editor, and Why Might You Use One?* Retrieved 06 08, 2022, from UltraEdit: https://www.ultraedit.com/company/blog/community/what-is-a-hex-editor-why-use-one.html

Myrick, A., & Wagoner, A. (2021, 12 15). *Best keyboard apps for Android 2022*. Retrieved 03 02, 2022, from Android Central: https://www.androidcentral.com/best-keyboard-android

NFI. (2017, 05 10). *Rapport Samen bouwen aan de toekomst van het NFI*. Retrieved 04 04, 2022, from NFI: https://www.forensischinstituut.nl/publicaties/publicaties/2017/05/10/rapport-samen-bouwen-aan-de-toekomst-van-het-nfi

NFI. (2021). *Organisatiestructuur*. Retrieved 12 08, 2021, from NFI: https://www.forensischinstituut.nl/over-het-nfi/organisatie/organisatiestructuur

NFI. (2022). *Hansken*. Retrieved 05 02, 2022, from Forensisch Instituut: https://www.forensischinstituut.nl/wetenschap-innovatie/laatste-forensische-innovaties/hansken

NFI. (2022). *Hansken.* Retrieved 05 09, 2022, from NFI: https://www.forensischinstituut.nl/wetenschap-innovatie/laatste-forensische-innovaties/hansken

NFI. (2022). Hansken extraction plugin SDK documentation for plugin developers. Nederlands Forensisch Instituut. Retrieved 05 14, 2022

NFI. (2022). *Kennis- en onderzoeksagenda Digitaal Forensisch Onderzoek*. Retrieved 06 13, 2022, from Forensisch Instituut: https://www.forensischinstituut.nl/wetenschap-innovatie/disciplines/digitaal-biometrisch/uitgebreid

Nijsen, Wouter. (2022, 01 20). *Overzicht: deze toestellen krijgen (waarschijnlijk) de Android 12-update*. Retrieved 03 01, 2022, from Android Planet: https://www.androidplanet.nl/nieuws/android-12-update-overzicht/

Novell.com. (n.d.). *Hex Editor*. Retrieved 06 08, 2022, from Novell: https://www.novell.com/documentation/ndsv8/usnds/c1help/novell_common/hexeditor.html#:~:text=cursor%20in%20decimal.-,The%20offset%20is%20the%20number%20of%20bytes%20from%20the%20beginning,the%20beginning%20of%20the%20string.

Overheid. (2012, 05 19). *Regeling taken NFI*. Retrieved 02 15, 2022, from Wetten Overheid: https://wetten.overheid.nl/BWBR0031558/2012-05-19/

Rechtspraak. (2018, 04 19). *ECLI:NL:RBAMS:2018:2504*. Retrieved 06 02, 2022, from Uitspraken Rechtspraak:

https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI%3ANL%3ARBAMS%3A2018%3A2504

Rechtspraak. (2021, 12 23). *ECLI:NL:GHARL:2021:11610*. Retrieved 06 02, 2022, from Uitspraken Rechtspraak: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHARL:2021:11610&showbutton=true&keyword=hansken

Rechtspraak. (2021, 12 23). *ECLI:NL:GHARL:2021:11736*. Retrieved 06 02, 2022, from Uitspraken Rechtspraak: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:GHARL:2021:11736&showbutton=true&keyword=hansken

Rechtspraak. (2021, 09 21). *ECLI:NL:RBROT:2021:9085*. Retrieved 06 02, 2022, from Uitspraken Rechtspraak: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBROT:2021:9085&showbutton=true&keyword=hansken

Rijksoverheid. (2021). *Agentschappen*. Retrieved 12 08, 2021, from Rijksoverheid: https://www.rijksoverheid.nl/onderwerpen/rijksoverheid/agentschappen

Rijksoverheid. (2021, 09 30). *September 2021: Aantal besmettingen blijft dalen, loslaten verplichte 1,5 meter maatregel*. Retrieved from Rijksoverheid: https://www.rijksoverheid.nl/onderwerpen/coronavirus-tijdlijn/september-2021-september-2021-aantal-besmettingen-blijft-dalen-loslaten-verplichte-15-meter-maatregel

Rijksoverheid. (2022). *Agentschappen Rijksoverheid*. Retrieved 04 05, 2022, from Rijksoverheid: https://www.rijksoverheid.nl/onderwerpen/rijksoverheid/agentschappen

Rijsoverheid. (2022, 03 21). *Pak de kansen van hybride werken*. Retrieved 04 29, 2022, from Rijsoverheid: https://www.rijksoverheid.nl/actueel/nieuws/2022/03/21/pak-de-kansen-van-hybride-werken

Samsung. (2020, 03 01). *Galaxy S21 5G(G991B) Application List*. Retrieved 04 12, 2022, from Samsung Knox: https://docs.samsungknox.com/CCMode/G991B_5G_R.pdf

Samsung. (2022, 11 08). *android.uid.honeyboard on Samsung Galaxy S9*. Retrieved 03 15, 2022, from Samsung: https://us.community.samsung.com/t5/Galaxy-S-Phones/android-uid-honeyboard-on-Samsung-Galaxy-S9/td-p/1366281

SamsungCommunity. (2020, 10 23). *Topic: Adding Words to Personal Dictionary in S20 Ultra*. Retrieved 03 03, 2022, from EU Community Samsung: https://eu.community.samsung.com/t5/galaxy-s20-series/adding-words-to-personal-dictionary-in-s20-ultra/td-p/2180859/page/2

Scribbr. (2022). *Statistiek voor Beginners | 5 Stappen & Voorbeelden*. Retrieved 05 23, 2022, from Scribbr: https://www.scribbr.nl/category/statistiek/

Simplilearn. (2021, 07 22). *What is requirement analysis article*. Retrieved 12 27, 2021, from Simpli Learn: https://www.simplilearn.com/what-is-requirement-analysis-article

SlickText. (2020, 04 27). *50 Text Abbreviations and How to Use Them [UPDATED]*. Retrieved 05 28, 2022, from SlickText: https://www.slicktext.com/blog/2019/02/text-abbreviations-guide/

SomeGuyOnAComputer. (2019, 02 03). *Where's the stock keyboard's predictive dictionary located?* Retrieved 03 02, 2022, from Android Stack Exchange: https://android.stackexchange.com/questions/22456/wheres-the-stock-keyboards-predictive-dictionary-located

Tew, Sarah; Dolcourt, Jessica; Lang, Angela. (2018, 10 19). *T-Mobile G1: The first Android phone never looked so good*. Retrieved 03 01, 2022, from CNET: https://www.cnet.com/pictures/t-mobile-first-g1-android-phone-never-looked-so-good/

Thakur, P. (2022). *com.samsung.android.honeyboard | com samsung android honeyboard*. Retrieved 03 15, 2022, from Gossip Funda: https://gossipfunda.com/com-samsung-android-honeyboard/

TheDaniel. (2018, 07 20). *Does anyone know how to clear "learned words"*. Retrieved 03 03, 2022, from US Community Samsung: https://us.community.samsung.com/t5/Other-Mobile-Devices/Does-anyone-know-how-to-clear-quot-learned-words-quot/td-p/362176

Toolshero. (2022, 02 25). *Configuraties van Mintzberg*. Retrieved 04 05, 2022, from Toolshero: https://www.toolshero.nl/verandermanagement/configuraties-van-mintzberg/

Toolshero. (2022, 01 10). *MoSCoW Methode*. Retrieved 05 14, 2022, from Toolshero: https://www.toolshero.nl/project-management/moscow-methode/

Toolshero. (2022, 01 31). *Watervalmethode*. Retrieved 05 29, 2022, from Toolshero: https://www.toolshero.nl/informatie-technologie/watervalmethode/

TutorialAndExamples. (2020, 10 06). *Tree in Data Structure*. Retrieved 06 11, 2022, from Tutorial and Example: https://www.tutorialandexample.com/tree-in-ds

Urma, Raoul-Gabriel. (2014, 07). *Alternative Languages for the JVM*. Retrieved 17 05, 2022, from Oracle: https://www.oracle.com/technical-resources/articles/java/architect-languages.html

van Beek, H., van Eijk, E., van Baar, R., Ugen, M., Bodde, J., & A.J., S. (2015, 12 08). *Digital forensics as a service: Game on.* Retrieved 05 12, 2022, from ScienceDirect: https://www.sciencedirect.com/science/article/pii/S1742287615000857

van Rossum, G., Warsaw, B., & Coghlan, N. (2001, 07 05). *PEP 8 - Style Guide for Python Code*. Retrieved 05 14, 2022, from Peps Python: https://peps.python.org/pep-0008/

van Rosusm, Guido. (n.d.). *Guido van Rossum - Personal Home Page*. Retrieved 05 14, 2022, from GitHub: https://gvanrossum.github.io/

van 't Klaphek, Michel. (2022). *Android 13: Alles over dé Android-update van 2022*. Retrieved 05 27, 2022, from Android Planet: https://www.androidplanet.nl/thema/android-13/

Vyas, K. (2021, 06 03). *The Best Android Keyboard Apps: Gboard, Swiftkey, Chrooma, and more!* Retrieved 03 02, 2022, from XDA Developers: https://www.xda-developers.com/best-android-keyboard/

Weijers, F., & Kerssenberg, D. (2022, 02 25). *Samsung Galaxy S22 Ultra Review*. Retrieved 03 14, 2022, from Tweakers: https://tweakers.net/reviews/9838/samsung-galaxy-s22-ultra-een-vermomde-galaxy-note.html

WinZip. (2022). *U wilt een TAR-bestand openen?* Retrieved 05 28, 2022, from WinZip: https://www.winzip.com/nl/learn/file-formats/tar/

5. Externe Bijlagen

Ext-I Vooronderzoek
Vooronderzoek Nabila Agni.pdf

Ext-II Requirementanalyse Infrastructuur Hansken
Requirementanalyse Infrastructuur Hansken Nabila Agni

Ext-III Requirementanalyse Hansken & Woordenboekparser
Requirementanalyse Hansken en Woordenboekparser Software Nabila Agni

Ext-IV Adviesraportage Woordenboekparser
Adviesrapportage Woordenboekparser Software Nabila Agni

Ext-V Woordenboekparser & bijbehorende bestanden
Woordenboekparser.zip

Ext-VI PEP-8 Style Guide
PEP 8 - Style Guide for Python Code (peps.python.org).pdf

6. Interne Bijlagen

Int-I Afstudeervoorstel

Gebruikersdictionary: een woordenboek vol bruikbare gebruikerssporen?

IWLAB - AFSTUDEREN

Naam student	Nabila Agni
Studentnummer	S1114455
Stageperiode	07/02/2022 - 14/07/2022
Afstudeerorganisatie	Nederlands Forensisch Instituut (NFI)
Onderwijsinstelling	Hogeschool Leiden – opleiding Informatica, specialisatie Forensisch ICT (FICT)

Versienummer	2.0
Versietype	Definitieve versie
Versiedatum	17/12/2021

Inhoudsopgave

1. Organisatorische aspecten

Afstudeerder

Naam	Nabila Agni
Studentnummer	S1114455
Adres	De Jonghstraat 10
Postcode en woonplaats	5461HD, Veghel
E-mailadres (school + privé)	S1114455@student.hsleiden.nl nabila.agni@hotmail.com
Telefoonnummer	0636480525

Afstudeerbedrijf

Naam	Nederlands Forensisch Instituut (NFI)
Adres	Laan van Ypenburg 6
Postcode en woonplaats	2497 GB Den Haag
Divisie en afdeling	Digitale en Biometrische Sporen (DBS), Digitale Technologie (DT)

Bedrijfsbegeleider

Naam	Ruud Schramp
E-mailadres	schramp@holmes.nl
Telefoonnummer	070-8886447

Studieloopbaanbegeleider

Naam	Jaap Haasnoot
E-mailadres	haasnoot.j@hsleiden.nl
Telefoonnummer	06-55429073

2. Context/achtergrond van het afstudeerbedrijf

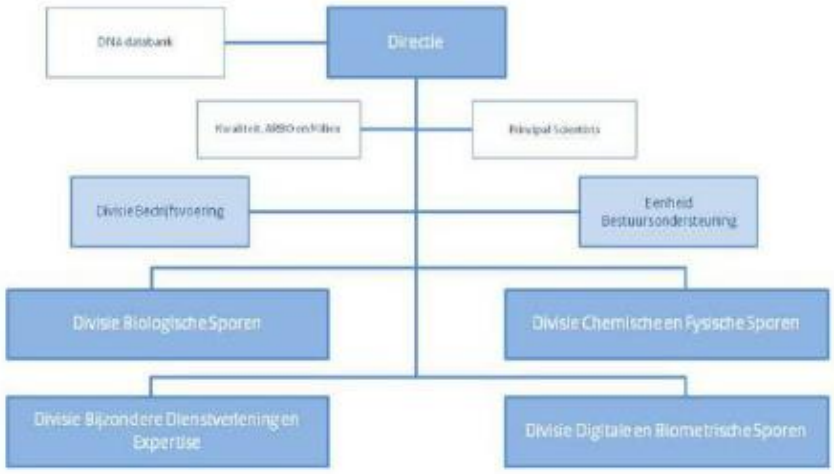
Het Nederlands Forensisch Instituut (NFI) is een verzelfstandigde organisatie (agentschap) van het ministerie van Justitie en Veiligheid (JenV) en valt onder het Directoraat-Generaal Rechtshandhaving (NFI, Organisatiestructuur, n.d.) (Rijksoverheid, n.d.). Het NFI levert forensische producten en diensten, met als doel om met onafhankelijk forensisch onderzoek de waarheidsbevinding in een strafrechtelijk onderzoek te bevorderen. Ze voorzien nationale en internationale organisaties die zich inzetten voor vrede, recht en veiligheid van betrouwbare informatie uit bronnen.

Het NFI streeft ernaar om de meest innovatieve en klantgerichte leverancier van forensische producten en diensten te zijn. Dit doen ze onder andere door te investeren in kennis en innovatie. Hierdoor kan het NFI inspelen op actuele maatschappelijke, technologische en wetenschappelijke ontwikkelingen. Internationaal gezien loopt het NFI hiermee voorop (NFI, Over het NFI, n.d.).

Voor sporenonderzoek heeft het NFI drie verschillende divisies:

1. Divisie Biologische Sporen
2. Divisie Chemische en Fysische Sporen
3. Divisie Digitale en Biometrische Sporen

In totaal bestaat het NFI uit vijf divisies en de eenheid bestuursondersteuning. Deze zijn in onderstaande organogram terug te zien (NFI, Organisatiestructuur, n.d.):



Figuur 1: Organogram Nederland Forensisch Instituut (NFI)

Mijn afstudeeropdracht voer ik binnen de divisie Digitale en Biometrische Sporen (DBS) uit, op de afdeling Digitale Technologie (DT). Mijn bedrijfsbegeleider is ook werkzaam op deze afdeling.

3. Probleem- of kans-beschrijving

In strafrechtelijke onderzoeken wordt er door verdachten vaak beweerd dat zij niet de auteur van een bericht zijn dat op zijn/haar toestel is gevonden. Al is er een bericht op iemands toestel gevonden, betekent dit niet meteen dat de eigenaar van het toestel het bericht ook daadwerkelijk getypt heeft. Het bericht kan door meerdere scenario's tot stand zijn gekomen:

- Het bericht is getypt.
- Het bericht is doorgestuurd.
- Het bericht is gekopieerd.

Bij het typen op een toestel ontstaan er sporen in de gebruikersdictionary van het toestel. Het toestel leert woorden die specifiek zijn voor de berichten die op het toestel zijn getypt en woordvolgorden die op het toestel worden gebruikt. De woorden die het toestel leert komen niet voor in het algemene woordenboek en de woordvolgorden kunnen karakteristiek zijn voor de berichten die op het toestel geschreven worden.

4. Bedrijfsdoelstelling

Het doel van het onderzoek is om te achterhalen of het mogelijk is om te bepalen of een bericht op een toestel geschreven is, of dat het bericht op een andere manier op het toestel is gekomen. Wanneer er achterhaald kan worden of een bericht wel of niet op een toestel geschreven is, kan dit gebruikt worden in de strafrechtelijke onderzoeken. Zo kunnen ook eventuele beweringen van verdachten die aangeven het bericht niet zelf hebben getypt maar het wel zelf hebben gedaan ontkracht worden. Het onderzoek draagt bij aan de onafhankelijke waarheidsbevinding.

5. Concrete opdrachtomschrijving

Om te kunnen bepalen of een bericht op een toestel geschreven is zal er een parser worden gemaakt. Er is vanuit de afstudeerorganisatie een voorkeur om voor de Proof of Concept van de parser gebruik te maken van de programmeertaal Python3. Om de parser te kunnen maken zal er eerst naar onder andere de volgende zaken vooronderzoek worden verricht:

- Hoe de woordenboeken door onder andere de toetsenbord applicaties van Samsung en Motorola toestellen worden opgeslagen.
- Hoe woordvolgorden worden opgeslagen.
- Of er een soortgelijke parser bestaat (eventueel in een andere programmeertaal).

Met de bovenstaande informatie zal er gekeken gaan worden naar op welke manier het mogelijk is om te bepalen of een bericht op een toestel geschreven is en hoe betrouwbaar het is. De betrouwbaarheid zal met behulp van een statisch model worden gemeten.

6. Planning

Onderstaand zijn er twee overzichten te vinden: één overzicht waarin de belangrijkste activiteiten die ik uit ga voeren terug te vinden zijn en één overzicht waarin belangrijke deadlines te vinden zijn. De activiteiten zijn gebaseerd op de deliverables en de planning is gebaseerd op de activiteiten en deadlines.

Fase	Deliverables	Activiteiten
Voorbereiding	D1 - Afstudeerplan	D1.1 - Een opbouw voor het afstudeerplan maken. D1.2 - Het uiteenzetten van het plan voor het onderzoek.
	D2 - Vooronderzoek	D2.1 - Een opbouw voor het vooronderzoek maken. D2.2 - Vooronderzoek verrichten.
	D3 - Requirementsanalyses	D3.1 - Een opbouw voor een requirementsanalyse maken.
	D4 - Risicoanalyse	D4.1 - Een opbouw voor een risicoanalyse maken.
	D5 - Adviesrapportage	D5.1 - Een opbouw voor een adviesrapportage maken.
	D7 - Statisch model	D7.1 - Een opbouw voor een statisch model maken.
	D8 - Afstudeerverslag	D8.1 - Een opbouw voor het afstudeerverslag maken.
Uitvoering opdracht	D3 - Requirementsanalyses	D3A.1 - Het maken van de requirementsanalyse voor een infrastructuur. D3B.1 - Het maken van de requirementsanalyse voor een softwaresysteem.
	D4 - Risicoanalyse	D4.2 - Het maken van de risicoanalyse voor een softwaresysteem.
	D5 - Adviesrapportage	D5.2 - Het maken van een adviesrapport voor een softwarearchitectuur.
	D6 - Forensische parser	D6.1 - Het bouwen van de parser. D6.2 - Het testen van de parser.
	D7 - Statisch model	D7.2 - Het maken van het statisch model.
	D8 - Afstudeerverslag	D8.2 - Het maken van het afstudeerverslag
Afronding	D7 - Presentatie	D7.1 Voorbereiding voor de presentatie.

Tabel 1: Activiteiten

Datum	Deadline
14-02-2022 - 25-02-2022	Eerste bedrijfsbezoek afstudeerdocent. Goedkeuring afstudeerplan door afstudeerdocent.
18-02-2022	Inleveren afstudeerplan.
18-04-2022 - 13-05-2022	Uiterlijke datum tussentijdse evaluatie bedrijfsbegeleider. Uiterlijke datum advies voortgang afstudeerdocent.
03-06-2022	Inleveren pre-final versie afstudeerverslag bij afstudeerdocent (t.b.v. feedback).
17-06-2022	Inleveren definitieve versie afstudeerverslag en producten (via de DLO).
03-07-2022 - 07-07-2022	Afstudeerzitting.

Tabel 2: Deadlines

De projectplanning is terug te vinden onder het hoofdstuk 'Bijlagen' als 'Bijlage I – Projectplanning Nabila Agni'.

7. Deliverables (producten)

D1 – Afstudeerplan
In het afstudeerplan wordt de afstudeeropdracht beschreven. Hierin worden de achtergrond van de opdracht, de probleemstelling, de doelstelling, de aanpak, planning, eventuele beperkingen en risico's meegenomen.

D2 – Vooronderzoek
In het vooronderzoek wordt er informatie met betrekking tot het onderwerp en de opdracht beschreven.

D3 – Requirementsanalyses
Voor het behalen van de B-competenties "Infrastructuur analyseren" en "Software analyseren" is het nodig om een requirementsanalyse te maken. In een requirementsanalyse worden vereisten van een product gedefinieerd.

- **D3A – Requirementsanalyse Infrastructuur**
- **D3B – Requirementsanalyse Software**

D4 – Risicoanalyse
Voor het behalen van de B-competentie "Software analyseren" is het nodig om een risicoanalyse te maken. In een risicoanalyse worden interne en externe risicofactoren in kaart gebracht. Het is een inventarisatie van gebeurtenissen die kunnen optreden.

D5 – Adviesrapportage
Voor het behalen van de B-competentie "Software adviseren" is het nodig om een advies op te stellen. Dit zal ik in de vorm van een adviesrapportage doen.

D6 – Forensische parser
De te maken parser voor het bepalen of een bericht op een toestel is getypt (in Python3).

D7 – Statisch model
Het model waarmee de betrouwbaarheid van de parser gemeten zal worden.

D8 – Afstudeerverslag
Het afstudeerverslag is het uiteindelijke verslag over de uitgevoerde afstudeeropdracht. Hierin worden de opdracht en het uiteindelijke resultaat beschreven.

D9 – Presentatie
Na het uitvoeren van de afstudeeropdracht en het afronden van de documentatie wordt er een presentatie verzorgd. In de presentatie zullen de afstudeeropdracht, gemaakte parser, resultaten en conclusies besproken worden.

7. Deliverables (producten)

D1 – Afstudeerplan
In het afstudeerplan wordt de afstudeeropdracht beschreven. Hierin worden de achtergrond van de opdracht, de probleemstelling, de doelstelling, de aanpak, planning, eventuele beperkingen en risico's meegenomen.

D2 – Vooronderzoek
In het vooronderzoek wordt er informatie met betrekking tot het onderwerp en de opdracht beschreven.

D3 – Requirementsanalyses
Voor het behalen van de B-competenties "Infrastructuur analyseren" en "Software analyseren" is het nodig om een requirementsanalyse te maken. In een requirementsanalyse worden vereisten van een product gedefinieerd.

- **D3A – Requirementsanalyse Infrastructuur**
- **D3B – Requirementsanalyse Software**

D4 – Risicoanalyse
Voor het behalen van de B-competentie "Software analyseren" is het nodig om een risicoanalyse te maken. In een risicoanalyse worden interne en externe risicofactoren in kaart gebracht. Het is een inventarisatie van gebeurtenissen die kunnen optreden.

D5 – Adviesrapportage
Voor het behalen van de B-competentie "Software adviseren" is het nodig om een advies op te stellen. Dit zal ik in de vorm van een adviesrapportage doen.

D6 – Forensische parser
De te maken parser voor het bepalen of een bericht op een toestel is getypt (in Python3).

D7 – Statisch model
Het model waarmee de betrouwbaarheid van de parser gemeten zal worden.

D8 – Afstudeerverslag
Het afstudeerverslag is het uiteindelijke verslag over de uitgevoerde afstudeeropdracht. Hierin worden de opdracht en het uiteindelijke resultaat beschreven.

D9 – Presentatie
Na het uitvoeren van de afstudeeropdracht en het afronden van de documentatie wordt er een presentatie verzorgd. In de presentatie zullen de afstudeeropdracht, gemaakte parser, resultaten en conclusies besproken worden.

8. Verdediging A-Competenties

De te behalen A-Competenties zijn als volgt:

- Onderzoek
- Leren
- Professioneel werken
- Innovatie

Hieronder beschrijf ik per A-Competentie op welke manier mijn afstudeeropdracht de mogelijkheid biedt om aan de competentie te werken.

8.1 Onderzoek

Om deze competentie te behalen moet ik:

- Een SMART hoofdvraag met meerdere deelvragen op basis van het probleem definiëren.
- Gebruik maken van passende methoden en technieken.
- Informatie met betrekking tot het onderzoek verzamelen en deze verwerken in een document.
- Een onderbouwde analyse uitvoeren en hierbij de deelvragen beantwoorden.
- Conclusies trekken, SMART aanbevelingen formuleren en ruimte maken voor discussie.

Er is een probleem. Om het probleem op te kunnen lossen wordt er een hoofdvraag gedefinieerd en om de hoofdvraag te kunnen beantwoorden worden er meerdere deelvragen gedefinieerd. Met de beantwoording van de deelvragen verzamel ik de benodigde kennis en informatie voor het maken van de parser. Met de parser kan ik analyses uitvoeren om onder andere de betrouwbaarheid van de parser te meten. Na het analyseren kan ik op basis van de analyses conclusies trekken en kan ik op basis van de conclusies aanbevelingen opstellen voor het NFI.

8.2 Leren

Om deze competentie te behalen moet ik:

- Een positief zelfbeeld creëren en reflecteren op mijn eigen werk.
- Zelfstandig hulpbronnen verzamelen en inzetten.
- Reflecteren op mijn eigen leerproces.
- Mezelf en anderen motiveren door een open en leergierige houding aannemen.

Omdat ik de afstudeeropdracht zelfstandig uitvoer is het voor mijzelf belangrijk om het idee te hebben dat ik de opdracht ook daadwerkelijk uit kan voeren. Omdat ik weinig kennis over de verschillende telefoontoetsenbord applicaties heb, ga ik ervoor zorgen dat ik mij hierin verdiep. Ik ga zelfstandig hulpbronnen verzamelen en deze toepassen. Ik vind de opdracht erg interessant en ik ben zelf ook heel erg benieuwd naar de uitkomst. Dit is mijn motivatie en deze wil ik uiten door initiatief en een goede inzet te tonen voor mijn onderzoek. Hiernaast ga ik een open en leergierige houding aannemen om mezelf te blijven motiveren en anderen ook te kunnen motiveren. Na het uitvoeren van de opdracht ga ik kritisch reflecteren op mijzelf en mijn werk zodat ik kan leren van de dingen die ik goed heb gedaan en van de verbeterpunten.

8.3 Professioneel werken

Om deze competentie te behalen moet ik:

- Het probleem analyseren.
- Planmatig werken.
- Methodisch werken.
- Communiceren binnen de omgeving en/of richting stakeholders.
- Proactieve houding tonen als het gaat om het krijgen en verwerken van feedback.
- Reflecteren op mijn eigen handelen.

Er is een probleem dat ik kan analyseren. De manier waarop ik te werk ga zijn vrij voor invulling. Hierbij moet ik een realistische planning voor mezelf opzetten en kiezen voor een logische werkmethode. De planning is niet statisch maar dynamisch. Ik ga voordat ik aan de opdracht begin alle stakeholders in kaart te brengen zodat ik met hen kan communiceren wanneer nodig. Ik vind het fijn om feedback te ontvangen en hier zal ik ook naar vragen waar mogelijk. De feedback zal ik meenemen en verwerken zodat ik van de goede en de verbeterpunten kan leren en zodat ik deze punten mee kan nemen in toekomstige opdrachten.

8.4 Innovatie

Om deze competentie te behalen moet ik:

- Creativiteit gebruiken.
- Gebruik maken van verschillende perspectieven.
- Gecalculeerde risico's nemen en collega's aanmoedigen om met nieuwe ideeën te komen.

Door zelf ideeën en oplossingen te bedenken voor het probleem kan ik mijn creativiteit toepassen. Voordat ik ook maar iets uitvoer, zal ik mijn plannen en ideeën bespreken met het NFI. Collega's zouden hun eventuele ideeën met mij kunnen delen, waar ik ze ook actief naar zal vragen, en ik ga actuele ontwikkelingen met betrekking tot het onderwerp van mijn opdracht in de gaten houden. Op deze manier kan ik gebruik maken van verschillende perspectieven. Ook zal ik, wanneer nodig en nuttig kan zijn, in overleg met het NFI voorberekende risico's nemen.

9. Aan te tonen B-Competenties

Tijdens het uitvoeren van mijn afstudeeropdracht wil ik de volgende drie FICT B-Competenties aantonen:

- Infrastructuur analyseren
- Software analyseren
- Software adviseren

Hieronder beschrijf ik per B-Competentie op welke manier mijn afstudeeropdracht de mogelijkheid biedt om aan de competentie aan te tonen.

9.1 Infrastructuur analyseren

Om deze competentie aan te kunnen tonen moet ik onderzoek doen naar trends op het gebied van ICT-infrastructuur op basis van technologische, economische en maatschappelijke ontwikkelingen en innovaties. Hierbij wordt er verwacht dat ik een requirementsanalyse uitvoer voor een bedrijfsinfrastructuur om functionele en niet-functionele eisen in kaart te brengen.

Voor mijn opdracht ga ik een parser schrijven omdat er nog geen methode bestaat voor het bepalen of een bericht op een toestel getypt is. Hierbij zal er gebruik worden gemaakt van de Android woordenboek applicatie. Het verrichten van vooronderzoek biedt mij de mogelijkheid om deze competentie aan te tonen. Tijdens het vooronderzoek ga ik mij verdiepen in de infrastructuur van Android toestellen omdat ik mij hier tijdens mijn opdracht mee bezig zal gaan houden. Hierbij ga ik eerst op een breed niveau bekijken hoe een Android toestel qua applicaties in elkaar zit. Zo kan ik inzoomen tot ik de woordenboek applicatie heb bereikt.

9.2 Software analyseren

Om deze competentie aan te kunnen tonen moet ik een requirementsanalyse uitvoeren voor een softwaresysteem met verschillende belanghebbenden in een context van bestaande systemen. Hierbij wordt er verwacht dat dat ik de integratie en migratie problematiek in kaart worden breng en moet er een acceptatiecriteria gedefinieerd worden aan de hand van kwaliteit eigenschappen en een uitgevoerde risicoanalyse.

Tijdens mijn vooronderzoek breng ik onder andere in kaart waar de woordenboek applicatie op een Android toestel zit. Om te onderzoeken hoe er data in het woordenboek terecht komt, zal ik zelf data gaan genereren op toestellen en zal ik deze data gaan onderzoeken. Om de parser zo goed mogelijk te kunnen implementeren in het huidige softwaresysteem is van belang om dit systeem te leren kennen. Hierbij ga ik kijken naar welke eventuele problemen er kunnen ontstaan bij het implementeren van de parser in het huidige softwaresysteem en ga ik een risicoanalyse uitvoeren. Op basis van de risicoanalyse zal er een acceptatiecriteria voor de implementatie van de nieuwe software gedefinieerd worden.

9.3 Software adviseren

Om deze competentie aan te kunnen tonen moet ik een advies geven met betrekking tot de keuze voor softwarearchitectuur of software frameworks, waarbij kostenaspecten en kwaliteitskenmerken zoals beschikbaarheid een rol spelen. Het advies gaat over de inrichting van een softwareontwikkelproces, waaronder het testproces.

Op basis van de uitgevoerde software analyse kan ik een advies opstellen over het implementeren van de door mij gemaakte parser in de bestaande software framework. Hierbij ga ik rekening houden met diverse belangrijke zaken, waaronder kostenaspecten en kwaliteitskenmerken zoals beschikbaarheid. Voor de parser wordt er een statisch model gemaakt om deze te testen op betrouwbaarheid. De testuitslagen zullen meegenomen worden bij het opstellen van het advies.

10. Literatuurlijst

NFI. (sd). *Organisatiestructuur*. Opgeroepen op 12 08, 2021, van Forensisch Instituut: https://www.forensischinstituut.nl/over-het-nfi/organisatie/organisatiestructuur

NFI. (sd). *Over het NFI*. Opgeroepen op 12 08, 2021, van Forensisch Instituut: https://www.forensischinstituut.nl/over-het-nfi

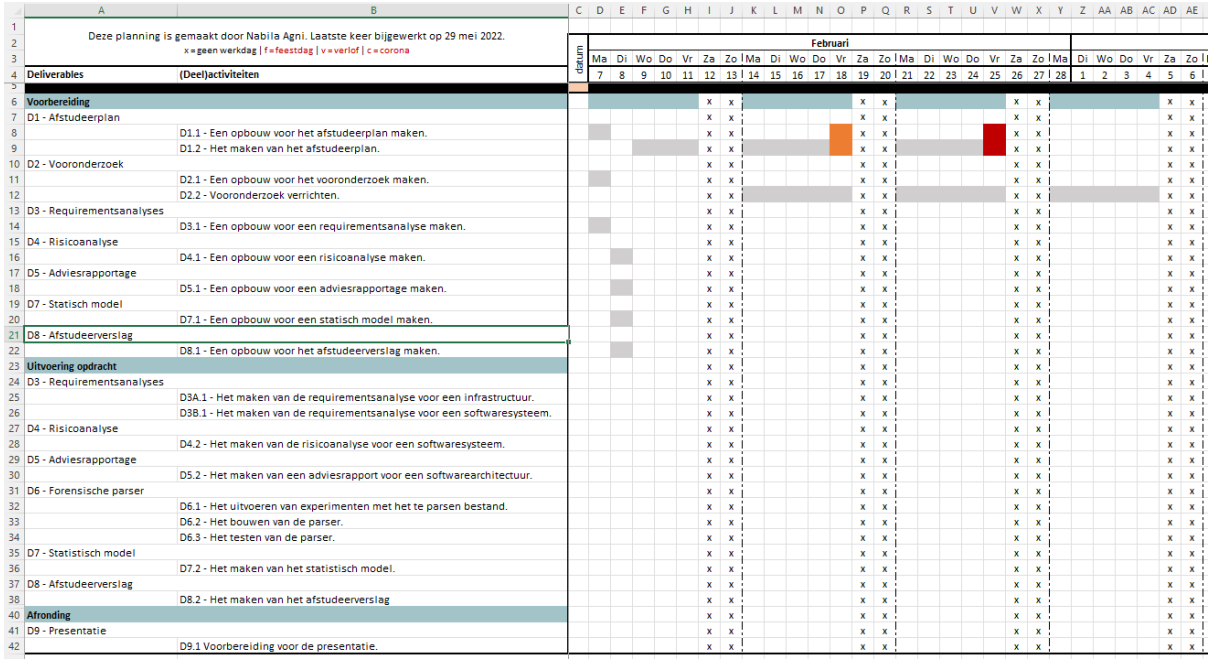
Rijksoverheid. (sd). *Agentschappen Rijksoverheid*. Opgeroepen op 12 08, 2021, van Rijksoverheid: https://www.rijksoverheid.nl/onderwerpen/rijksoverheid/agentschappen

11. Bijlagen

Bijlage I – Projectplanning

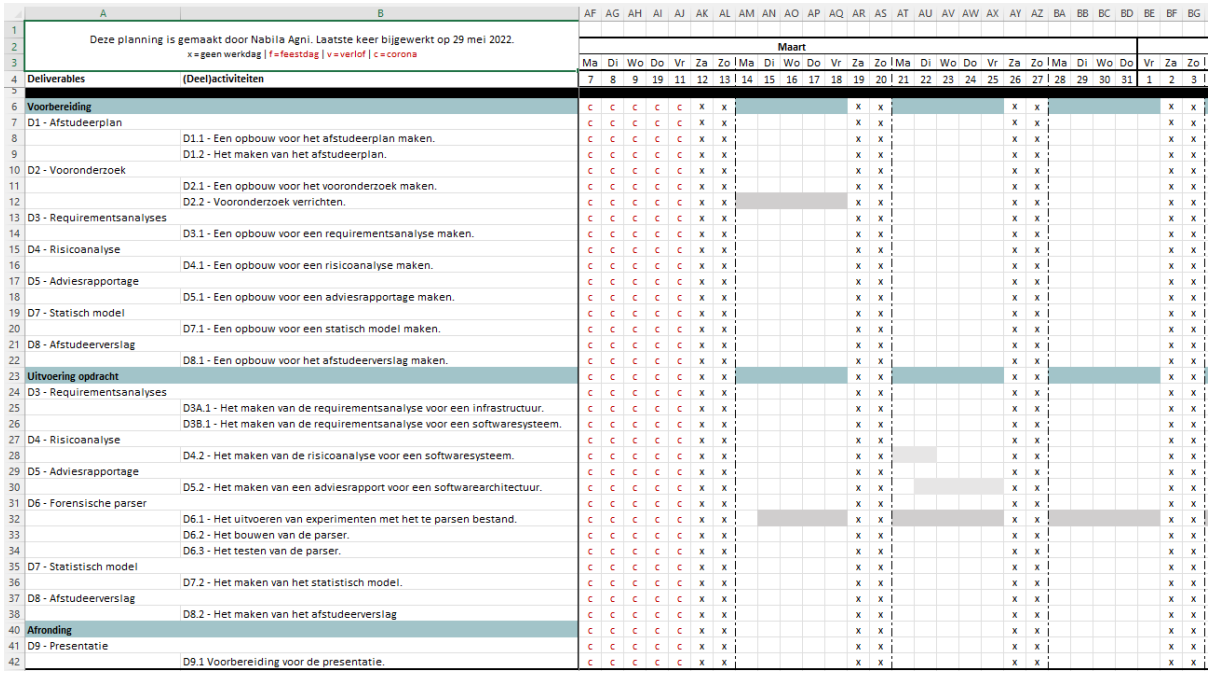
Projectplanning Nabila Agni.xlsx

Int-II Planning



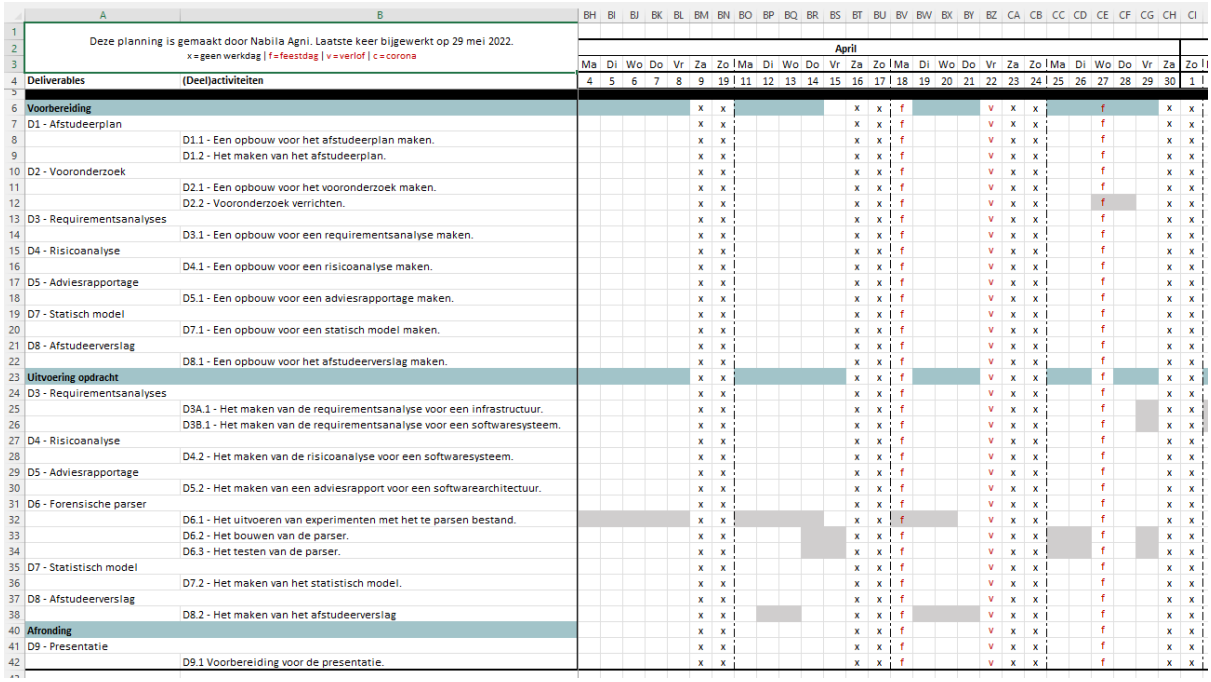
Deze planning is gemaakt door Nabila Agni. Laatste keer bijgewerkt op 29 mei 2022.
x = geen werkdag | f = feestdag | v = verlof | c = corona

Deliverables	(Deel)activiteiten	datum	Februari																																
			Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo
			7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	1	2	3	4	5	6							
Voorbereiding								x	x						x	x						x	x						x	x				x	x
D1 - Afstudeerplan								x	x						x	x						x	x						x	x				x	x
	D1.1 - Een opbouw voor het afstudeerplan maken.							x	x						x	x						x	x						x	x				x	x
	D1.2 - Het maken van het afstudeerplan.							x	x						x	x						x	x						x	x				x	x
D2 - Vooronderzoek								x	x						x	x						x	x						x	x				x	x
	D2.1 - Een opbouw voor het vooronderzoek maken.							x	x						x	x						x	x						x	x				x	x
	D2.2 - Vooronderzoek verrichten.							x	x						x	x						x	x						x	x				x	x
D3 - Requirementsanalyses								x	x						x	x						x	x						x	x				x	x
	D3.1 - Een opbouw voor een requirementsanalyse maken.							x	x						x	x						x	x						x	x				x	x
D4 - Risicoanalyse								x	x						x	x						x	x						x	x				x	x
	D4.1 - Een opbouw voor een risicoanalyse maken.							x	x						x	x						x	x						x	x				x	x
D5 - Adviesrapportage								x	x						x	x						x	x						x	x				x	x
	D5.1 - Een opbouw voor een adviesrapportage maken.							x	x						x	x						x	x						x	x				x	x
D7 - Statisch model								x	x						x	x						x	x						x	x				x	x
	D7.1 - Een opbouw voor een statisch model maken.							x	x						x	x						x	x						x	x				x	x
D8 - Afstudeerverslag								x	x						x	x						x	x						x	x				x	x
	D8.1 - Een opbouw voor het afstudeerverslag maken.							x	x						x	x						x	x						x	x				x	x
Uitvoering opdracht								x	x						x	x						x	x						x	x				x	x
D3 - Requirementsanalyses								x	x						x	x						x	x						x	x				x	x
	D3A.1 - Het maken van de requirementsanalyse voor een infrastructuur.							x	x						x	x						x	x						x	x				x	x
	D3B.1 - Het maken van de requirementsanalyse voor een softwaresysteem.							x	x						x	x						x	x						x	x				x	x
D4 - Risicoanalyse								x	x						x	x						x	x						x	x				x	x
	D4.2 - Het maken van de risicoanalyse voor een softwaresysteem.							x	x						x	x						x	x						x	x				x	x
D5 - Adviesrapportage								x	x						x	x						x	x						x	x				x	x
	D5.2 - Het maken van een adviesrapport voor een softwarearchitectuur.							x	x						x	x						x	x						x	x				x	x
D6 - Forensische parser								x	x						x	x						x	x						x	x				x	x
	D6.1 - Het uitvoeren van experimenten met het te parsen bestand.							x	x						x	x						x	x						x	x				x	x
	D6.2 - Het bouwen van de parser.							x	x						x	x						x	x						x	x				x	x
	D6.3 - Het testen van de parser.							x	x						x	x						x	x						x	x				x	x
D7 - Statistisch model								x	x						x	x						x	x						x	x				x	x
	D7.2 - Het maken van het statistisch model.							x	x						x	x						x	x						x	x				x	x
D8 - Afstudeerverslag								x	x						x	x						x	x						x	x				x	x
	D8.2 - Het maken van het afstudeerverslag							x	x						x	x						x	x						x	x				x	x
Afronding								x	x						x	x						x	x						x	x				x	x
D9 - Presentatie								x	x						x	x						x	x						x	x				x	x
	D9.1 Voorbereiding voor de presentatie.							x	x						x	x						x	x						x	x				x	x



Deze planning is gemaakt door Nabila Agni. Laatste keer bijgewerkt op 29 mei 2022.
x = geen werkdag | f = feestdag | v = verlof | c = corona

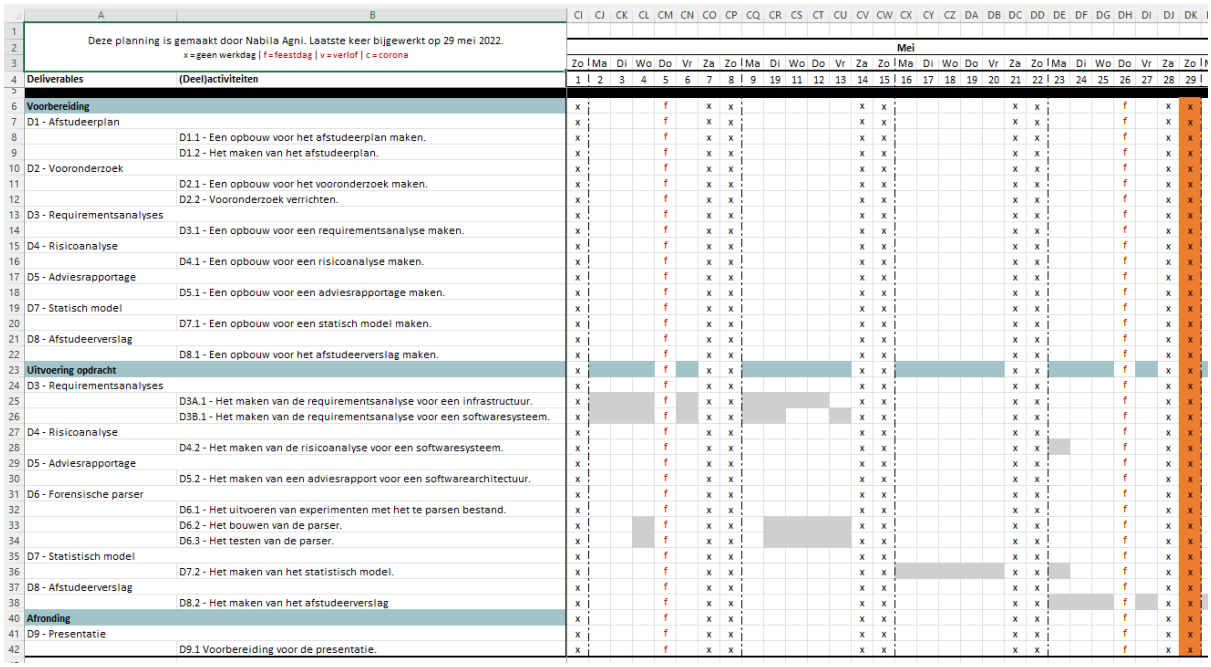
Deliverables	(Deel)activiteiten	Maart																																	
		Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo
		7	8	9	19	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3							
Voorbereiding		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D1 - Afstudeerplan		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D1.1 - Een opbouw voor het afstudeerplan maken.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D1.2 - Het maken van het afstudeerplan.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D2 - Vooronderzoek		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D2.1 - Een opbouw voor het vooronderzoek maken.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D2.2 - Vooronderzoek verrichten.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D3 - Requirementsanalyses		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D3.1 - Een opbouw voor een requirementsanalyse maken.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D4 - Risicoanalyse		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D4.1 - Een opbouw voor een risicoanalyse maken.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D5 - Adviesrapportage		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D5.1 - Een opbouw voor een adviesrapportage maken.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D7 - Statisch model		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D7.1 - Een opbouw voor een statisch model maken.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D8 - Afstudeerverslag		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D8.1 - Een opbouw voor het afstudeerverslag maken.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
Uitvoering opdracht		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D3 - Requirementsanalyses		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D3A.1 - Het maken van de requirementsanalyse voor een infrastructuur.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D3B.1 - Het maken van de requirementsanalyse voor een softwaresysteem.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D4 - Risicoanalyse		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D4.2 - Het maken van de risicoanalyse voor een softwaresysteem.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D5 - Adviesrapportage		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D5.2 - Het maken van een adviesrapport voor een softwarearchitectuur.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D6 - Forensische parser		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D6.1 - Het uitvoeren van experimenten met het te parsen bestand.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D6.2 - Het bouwen van de parser.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D6.3 - Het testen van de parser.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D7 - Statistisch model		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D7.2 - Het maken van het statistisch model.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D8 - Afstudeerverslag		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D8.2 - Het maken van het afstudeerverslag	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
Afronding		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
D9 - Presentatie		c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		
	D9.1 Voorbereiding voor de presentatie.	c	c	c	c	c	x	x						x	x						x	x						x	x				x	x		



Deze planning is gemaakt door Nabila Agni. Laatste keer bijgewerkt op 29 mei 2022.
x = geen werkdag | f = feestdag | v = verlof | c = corona

April

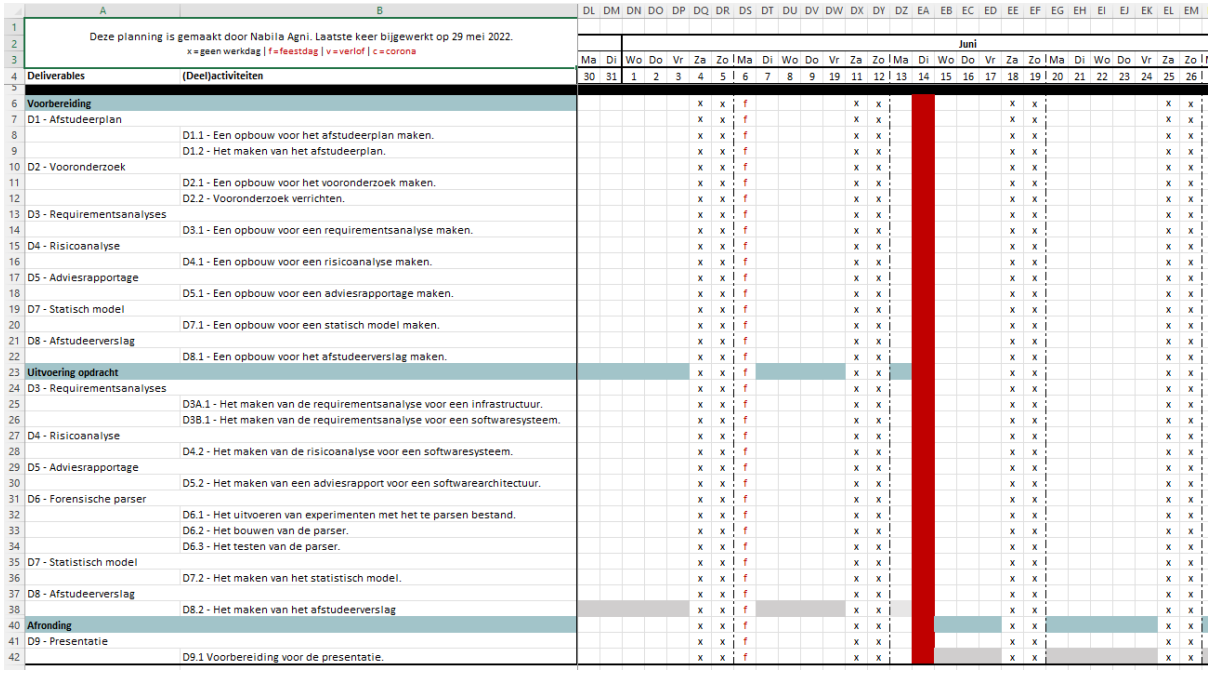
Deliverables	(Deel)activiteiten	Ma 4	Di 5	Wo 6	Do 7	Vr 8	Za 9	Zo 10	Ma 11	Di 12	Wo 13	Do 14	Vr 15	Za 16	Zo 17	Ma 18	Di 19	Wo 20	Do 21	Vr 22	Za 23	Zo 24	Ma 25	Di 26	Wo 27	Do 28	Vr 29	Za 30	Zo 1
Voorbereiding							x	x					x	x	f				v	x	x			f			x	x
D1 - Afstudeerplan							x	x					x	x	f				v	x	x			f			x	x
	D1.1 - Een opbouw voor het afstudeerplan maken.						x	x					x	x	f				v	x	x			f			x	x
	D1.2 - Het maken van het afstudeerplan.						x	x					x	x	f				v	x	x			f			x	x
D2 - Vooronderzoek							x	x					x	x	f				v	x	x			f			x	x
	D2.1 - Een opbouw voor het vooronderzoek maken.						x	x					x	x	f				v	x	x			f			x	x
	D2.2 - Vooronderzoek verrichten.						x	x					x	x	f				v	x	x			f			x	x
D3 - Requirementsanalyses							x	x					x	x	f				v	x	x			f			x	x
	D3.1 - Een opbouw voor een requirementsanalyse maken.						x	x					x	x	f				v	x	x			f			x	x
D4 - Risicoanalyse							x	x					x	x	f				v	x	x			f			x	x
	D4.1 - Een opbouw voor een risicoanalyse maken.						x	x					x	x	f				v	x	x			f			x	x
D5 - Adviesrapportage							x	x					x	x	f				v	x	x			f			x	x
	D5.1 - Een opbouw voor een adviesrapportage maken.						x	x					x	x	f				v	x	x			f			x	x
D7 - Statisch model							x	x					x	x	f				v	x	x			f			x	x
	D7.1 - Een opbouw voor een statisch model maken.						x	x					x	x	f				v	x	x			f			x	x
D8 - Afstudeerverslag							x	x					x	x	f				v	x	x			f			x	x
	D8.1 - Een opbouw voor het afstudeerverslag maken.						x	x					x	x	f				v	x	x			f			x	x
Uitvoering opdracht							x	x					x	x	f				v	x	x			f			x	x
D3 - Requirementsanalyses							x	x					x	x	f				v	x	x			f			x	x
	D3A.1 - Het maken van de requirementsanalyse voor een infrastructuur.						x	x					x	x	f				v	x	x			f			x	x
	D3B.1 - Het maken van de requirementsanalyse voor een softwaresysteem.						x	x					x	x	f				v	x	x			f			x	x
D4 - Risicoanalyse							x	x					x	x	f				v	x	x			f			x	x
	D4.2 - Het maken van de risicoanalyse voor een softwaresysteem.						x	x					x	x	f				v	x	x			f			x	x
D5 - Adviesrapportage							x	x					x	x	f				v	x	x			f			x	x
	D5.2 - Het maken van een adviesrapport voor een softwarearchitectuur.						x	x					x	x	f				v	x	x			f			x	x
D6 - Forensische parser							x	x					x	x	f				v	x	x			f			x	x
	D6.1 - Het uitvoeren van experimenten met het te parsen bestand.						x	x					x	x	f				v	x	x			f			x	x
	D6.2 - Het bouwen van de parser.						x	x					x	x	f				v	x	x			f			x	x
	D6.3 - Het testen van de parser.						x	x					x	x	f				v	x	x			f			x	x
D7 - Statistisch model							x	x					x	x	f				v	x	x			f			x	x
	D7.2 - Het maken van het statistisch model.						x	x					x	x	f				v	x	x			f			x	x
D8 - Afstudeerverslag							x	x					x	x	f				v	x	x			f			x	x
	D8.2 - Het maken van het afstudeerverslag						x	x					x	x	f				v	x	x			f			x	x
Afronding							x	x					x	x	f				v	x	x			f			x	x
D9 - Presentatie							x	x					x	x	f				v	x	x			f			x	x
	D9.1 Voorbereiding voor de presentatie.						x	x					x	x	f				v	x	x			f			x	x



Deze planning is gemaakt door Nabila Agni. Laatste keer bijgewerkt op 29 mei 2022.
x = geen werkdag | f = feestdag | v = verlof | c = corona

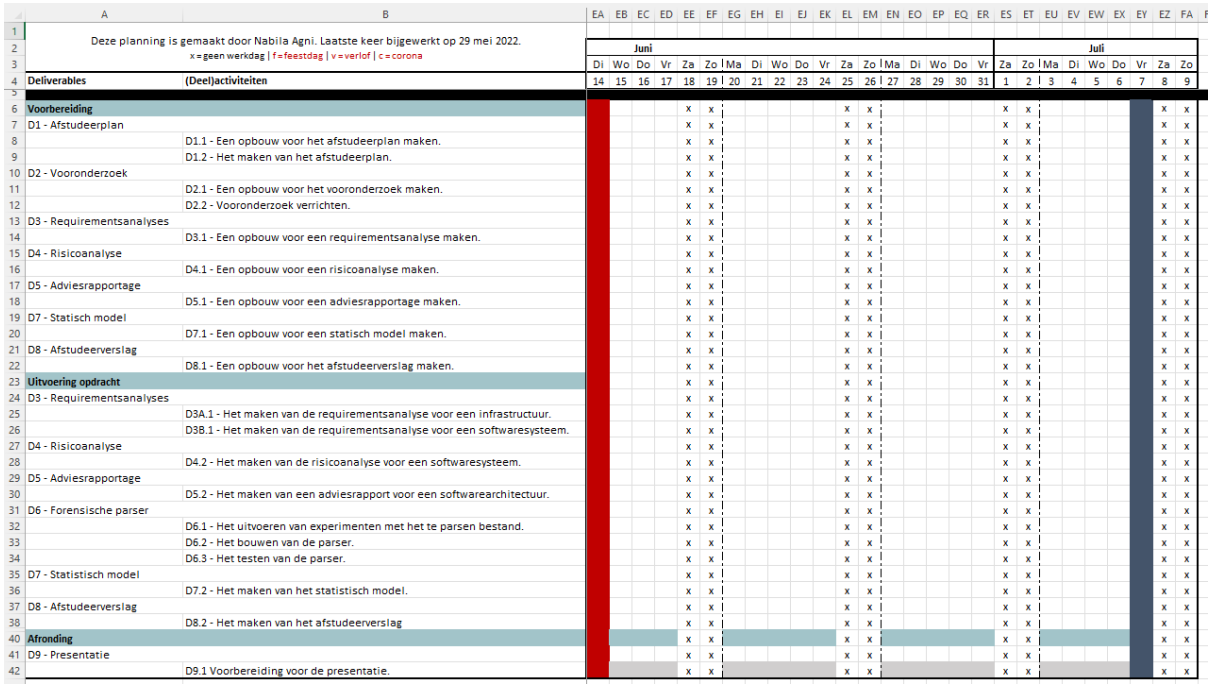
Mei

Deliverables	(Deel)activiteiten	Zo 1	Ma 2	Di 3	Wo 4	Do 5	Vr 6	Za 7	Zo 8	Ma 9	Di 19	Wo 11	Do 12	Vr 13	Za 14	Zo 15	Ma 16	Di 17	Wo 18	Do 19	Vr 20	Za 21	Zo 22	Ma 23	Di 24	Wo 25	Do 26	Vr 27	Za 28	Zo 29
Voorbereiding		x				f		x	x					x	x							x	x			f		x	x
D1 - Afstudeerplan		x				f		x	x					x	x							x	x			f		x	x
	D1.1 - Een opbouw voor het afstudeerplan maken.	x				f		x	x					x	x							x	x			f		x	x
	D1.2 - Het maken van het afstudeerplan.	x				f		x	x					x	x							x	x			f		x	x
D2 - Vooronderzoek		x				f		x	x					x	x							x	x			f		x	x
	D2.1 - Een opbouw voor het vooronderzoek maken.	x				f		x	x					x	x							x	x			f		x	x
	D2.2 - Vooronderzoek verrichten.	x				f		x	x					x	x							x	x			f		x	x
D3 - Requirementsanalyses		x				f		x	x					x	x							x	x			f		x	x
	D3.1 - Een opbouw voor een requirementsanalyse maken.	x				f		x	x					x	x							x	x			f		x	x
D4 - Risicoanalyse		x				f		x	x					x	x							x	x			f		x	x
	D4.1 - Een opbouw voor een risicoanalyse maken.	x				f		x	x					x	x							x	x			f		x	x
D5 - Adviesrapportage		x				f		x	x					x	x							x	x			f		x	x
	D5.1 - Een opbouw voor een adviesrapportage maken.	x				f		x	x					x	x							x	x			f		x	x
D7 - Statisch model		x				f		x	x					x	x							x	x			f		x	x
	D7.1 - Een opbouw voor een statisch model maken.	x				f		x	x					x	x							x	x			f		x	x
D8 - Afstudeerverslag		x				f		x	x					x	x							x	x			f		x	x
	D8.1 - Een opbouw voor het afstudeerverslag maken.	x				f		x	x					x	x							x	x			f		x	x
Uitvoering opdracht		x				f		x	x					x	x							x	x			f		x	x
D3 - Requirementsanalyses		x				f		x	x					x	x							x	x			f		x	x
	D3A.1 - Het maken van de requirementsanalyse voor een infrastructuur.	x				f		x	x					x	x							x	x			f		x	x
	D3B.1 - Het maken van de requirementsanalyse voor een softwaresysteem.	x				f		x	x					x	x							x	x			f		x	x
D4 - Risicoanalyse		x				f		x	x					x	x							x	x			f		x	x
	D4.2 - Het maken van de risicoanalyse voor een softwaresysteem.	x				f		x	x					x	x							x	x			f		x	x
D5 - Adviesrapportage		x				f		x	x					x	x							x	x			f		x	x
	D5.2 - Het maken van een adviesrapport voor een softwarearchitectuur.	x				f		x	x					x	x							x	x			f		x	x
D6 - Forensische parser		x				f		x	x					x	x							x	x			f		x	x
	D6.1 - Het uitvoeren van experimenten met het te parsen bestand.	x				f		x	x					x	x							x	x			f		x	x
	D6.2 - Het bouwen van de parser.	x				f		x	x					x	x							x	x			f		x	x
	D6.3 - Het testen van de parser.	x				f		x	x					x	x							x	x			f		x	x
D7 - Statistisch model		x				f		x	x					x	x							x	x			f		x	x
	D7.2 - Het maken van het statistisch model.	x				f		x	x					x	x							x	x			f		x	x
D8 - Afstudeerverslag		x				f		x	x					x	x							x	x			f		x	x
	D8.2 - Het maken van het afstudeerverslag	x				f		x	x					x	x							x	x			f		x	x
Afronding		x				f		x	x					x	x							x	x			f		x	x
D9 - Presentatie		x				f		x	x					x	x							x	x			f		x	x
	D9.1 Voorbereiding voor de presentatie.	x				f		x	x					x	x							x	x			f		x	x



Deze planning is gemaakt door Nabila Agni. Laatste keer bijgewerkt op 29 mei 2022.
x = geen werkdag | f = feestdag | v = verlof | c = corona

Deliverables	(Deel)activiteiten	Juni																																		
		Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo
		30	31	1	2	3	4	5	6	7	8	9	19	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26							
Voorbereiding							x	x	f					x	x						x	x						x	x							
D1 - Afstudeerplan							x	x	f					x	x						x	x						x	x							
	D1.1 - Een opbouw voor het afstudeerplan maken.						x	x	f					x	x						x	x						x	x							
	D1.2 - Het maken van het afstudeerplan.						x	x	f					x	x						x	x						x	x							
D2 - Vooronderzoek							x	x	f					x	x						x	x						x	x							
	D2.1 - Een opbouw voor het vooronderzoek maken.						x	x	f					x	x						x	x						x	x							
	D2.2 - Vooronderzoek verrichten.						x	x	f					x	x						x	x						x	x							
D3 - Requirementsanalyses							x	x	f					x	x						x	x						x	x							
	D3.1 - Een opbouw voor een requirementsanalyse maken.						x	x	f					x	x						x	x						x	x							
D4 - Risicoanalyse							x	x	f					x	x						x	x						x	x							
	D4.1 - Een opbouw voor een risicoanalyse maken.						x	x	f					x	x						x	x						x	x							
D5 - Adviesrapportage							x	x	f					x	x						x	x						x	x							
	D5.1 - Een opbouw voor een adviesrapportage maken.						x	x	f					x	x						x	x						x	x							
D7 - Statisch model							x	x	f					x	x						x	x						x	x							
	D7.1 - Een opbouw voor een statisch model maken.						x	x	f					x	x						x	x						x	x							
D8 - Afstudeerverslag							x	x	f					x	x						x	x						x	x							
	D8.1 - Een opbouw voor het afstudeerverslag maken.						x	x	f					x	x						x	x						x	x							
Uitvoering opdracht							x	x	f					x	x						x	x						x	x							
D3 - Requirementsanalyses							x	x	f					x	x						x	x						x	x							
	D3A.1 - Het maken van de requirementsanalyse voor een infrastructuur.						x	x	f					x	x						x	x						x	x							
	D3B.1 - Het maken van de requirementsanalyse voor een softwaresysteem.						x	x	f					x	x						x	x						x	x							
D4 - Risicoanalyse							x	x	f					x	x						x	x						x	x							
	D4.2 - Het maken van de risicoanalyse voor een softwaresysteem.						x	x	f					x	x						x	x						x	x							
D5 - Adviesrapportage							x	x	f					x	x						x	x						x	x							
	D5.2 - Het maken van een adviesrapport voor een softwarearchitectuur.						x	x	f					x	x						x	x						x	x							
D6 - Forensische parser							x	x	f					x	x						x	x						x	x							
	D6.1 - Het uitvoeren van experimenten met het te parsen bestand.						x	x	f					x	x						x	x						x	x							
	D6.2 - Het bouwen van de parser.						x	x	f					x	x						x	x						x	x							
	D6.3 - Het testen van de parser.						x	x	f					x	x						x	x						x	x							
D7 - Statistisch model							x	x	f					x	x						x	x						x	x							
	D7.2 - Het maken van het statistisch model.						x	x	f					x	x						x	x						x	x							
D8 - Afstudeerverslag							x	x	f					x	x						x	x						x	x							
	D8.2 - Het maken van het afstudeerverslag						x	x	f					x	x						x	x						x	x							
Afronding							x	x	f					x	x						x	x						x	x							
D9 - Presentatie							x	x	f					x	x						x	x						x	x							
	D9.1 Voorbereiding voor de presentatie.						x	x	f					x	x						x	x						x	x							



Deze planning is gemaakt door Nabila Agni. Laatste keer bijgewerkt op 29 mei 2022.
x = geen werkdag | f = feestdag | v = verlof | c = corona

Deliverables	(Deel)activiteiten	Juni																					Juli													
		Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo	Ma	Di	Wo	Do	Vr	Za	Zo
		14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3	4	5	6	7	8	9										
Voorbereiding						x	x						x	x						x	x						x	x									
D1 - Afstudeerplan						x	x						x	x						x	x						x	x									
	D1.1 - Een opbouw voor het afstudeerplan maken.					x	x						x	x						x	x						x	x									
	D1.2 - Het maken van het afstudeerplan.					x	x						x	x						x	x						x	x									
D2 - Vooronderzoek						x	x						x	x						x	x						x	x									
	D2.1 - Een opbouw voor het vooronderzoek maken.					x	x						x	x						x	x						x	x									
	D2.2 - Vooronderzoek verrichten.					x	x						x	x						x	x						x	x									
D3 - Requirementsanalyses						x	x						x	x						x	x						x	x									
	D3.1 - Een opbouw voor een requirementsanalyse maken.					x	x						x	x						x	x						x	x									
D4 - Risicoanalyse						x	x						x	x						x	x						x	x									
	D4.1 - Een opbouw voor een risicoanalyse maken.					x	x						x	x						x	x						x	x									
D5 - Adviesrapportage						x	x						x	x						x	x						x	x									
	D5.1 - Een opbouw voor een adviesrapportage maken.					x	x						x	x						x	x						x	x									
D7 - Statisch model						x	x						x	x						x	x						x	x									
	D7.1 - Een opbouw voor een statisch model maken.					x	x						x	x						x	x						x	x									
D8 - Afstudeerverslag						x	x						x	x						x	x						x	x									
	D8.1 - Een opbouw voor het afstudeerverslag maken.					x	x						x	x						x	x						x	x									
Uitvoering opdracht						x	x						x	x						x	x						x	x									
D3 - Requirementsanalyses						x	x						x	x						x	x						x	x									
	D3A.1 - Het maken van de requirementsanalyse voor een infrastructuur.					x	x						x	x						x	x						x	x									
	D3B.1 - Het maken van de requirementsanalyse voor een softwaresysteem.					x	x						x	x						x	x						x	x									
D4 - Risicoanalyse						x	x						x	x						x	x						x	x									
	D4.2 - Het maken van de risicoanalyse voor een softwaresysteem.					x	x						x	x						x	x						x	x									
D5 - Adviesrapportage						x	x						x	x						x	x						x	x									
	D5.2 - Het maken van een adviesrapport voor een softwarearchitectuur.					x	x						x	x						x	x						x	x									
D6 - Forensische parser						x	x						x	x						x	x						x	x									
	D6.1 - Het uitvoeren van experimenten met het te parsen bestand.					x	x						x	x						x	x						x	x									
	D6.2 - Het bouwen van de parser.					x	x						x	x						x	x						x	x									
	D6.3 - Het testen van de parser.					x	x						x	x						x	x						x	x									
D7 - Statistisch model						x	x						x	x						x	x						x	x									
	D7.2 - Het maken van het statistisch model.					x	x						x	x						x	x						x	x									
D8 - Afstudeerverslag						x	x						x	x						x	x						x	x									
	D8.2 - Het maken van het afstudeerverslag					x	x						x	x						x	x						x	x									
Afronding						x	x						x	x						x	x						x	x									
D9 - Presentatie						x	x						x	x						x	x						x	x									
	D9.1 Voorbereiding voor de presentatie.					x	x						x	x						x	x						x	x									