Plan van aanpak I-Pass Project

Risk-Sniffer

Project voor Defensie, LeerDock ICT Woensdrecht

Door Mick Luuring, student HU en studiebeurs Defensie.

Het probleem en de opdracht:

Bij het LeerDock ICT op vliegbasis Woensdrecht maken veel VEVA studenten gebruik van pc's en het netwerk van defensie. Maar helaas is, zoals op elke school, het geval dat studenten af en toe andere dingen doen tijdens de lessen dan dat bedoeld is. Hiervoor moet ik software maken die hun netwerkverkeer logged en analyseert. Hiermee zou de software dan een risico, voor het netwerk van defensie kunnen proberen in te schatten, denk hierbij aan bijvoorbeeld een gekke torrent website, of waar "blacklisted" woorden op te vinden zijn.

Planning:

Ik ga door een semi iteratieve aanpak te hanteren mijn project aanpakken. Dit betekent dat ik een planning op ga stellen met (zeer waarschijnlijk) goedhaalbare deadlines, waar ik een extra marge voor neem mocht iets langer uitpakken dan gedacht. Dit zorgt ervoor dat ik eerst een deel af kan maken, voordat ik door kan met het volgende deel. Ik doe dit omdat ik vaak een bepaalde functie, of feature werkend moet hebben voordat ik door kan met het volgende onderdeel van dit project.

MIJN SCHEMA:

Ik ben van plan om een weekje eerder te beginnen met de voorbereiding van dit project, ik ga hier nog niet volledig aan de slag. Deze week is eigenlijk een week vakantie. Maar ik zelf vindt deze extra tijd een fijne bijkomstigheid in dit project. Hierdoor houd ik een 4 week durende planning aan in plaats van de 3 weken die er vanuit school voor gepland staan.

WEEK 0 (OPSTART + VOORBEREIDING):

- Onderzoek doen naar de mogelijkheid van een transparant proxy op meerdere systemen. (Windows, Linux)
- Sniffing/proxy framework testen en werkend krijgen op mijn 2 test laptops.
- Automatiseren van de installatie van deze framework voor makkelijk gebruik opdrachtgever.
- Netwerk schema uitdenken voor visualisatie van mijn gebruikte aanpak.
- Programmeren van de data lees/log functie die websites moet loggen van http en hopelijk ook https.

WEEK 1 (SNIFFING TOOL AF EN DATA LOGGEN):

Voorwaarden:

Ik heb de proxy framework werkend, en snap hoe ik deze bedien met python, ook moet ik een duidelijk inzicht hebben op hoe ik het netwerk op ga bouwen.

- Back log vorige week afmaken
- Documentatie week 0 net opstellen

- Log/leesfunctie optimaliseren zodat deze minder systeem resources gebruikt, zodat dit in de praktijk ook daadwerkelijk gebruikt kan worden.
- Programmeren van de textmining / webmining functie

WEEK 2 (DATA LOGGING AF EN RISICOANALYSE PROGRAMMA MAKEN)

Voorwaarden:

Ik heb de data logging/realtime analyse volledig werkend, waardoor ik minimaal de url van bezochte websites kan ontvangen, dit zorgt ervoor dat mijn programma zelf met de website kan verbinden en deze kan scrapen om deze te website te kunnen analyseren.

- Backlog vorige week afmaken.
- Documentatie week 1 net opstellen
- Programmeren commandline interface (via ssh?) of in tijdsnood een config.txt file.
- Mits alles is afgerond en werkend kritiek vragen aan mijn opdrachtgever bij defensie.
- Locatie bezichtiging bij vliegbasis Woensdrecht

WEEK 3 (PROGRAMMA AF, KRITIEK VERWERKEN)

Voorwaarden:

Ik heb de programmatuur werkend (minimaal op basis niveau) en kan deze

- Functies testen op fouten/ bugs
- Kritiek verwerken
- Installatie instructies maken
- Documentatie volledig
- Rapport afmaken

Punten van aandacht:

Ik moet de risicoanalyse die ik heb gedaan goed in acht nemen, hier heb ik rekening mee gehouden in mijn planning (Denk aan het toevoegen van tijd voor mijn backlog, of het vergeten van documentatie hier is dus tijd voor in ge plannend.) Ook houd ik een Trello taskboard bij voor de visualisatie van mijn planning en taken die nog moeten gebeuren. Verder houd ik contact met mijn Ipass begeleider en mijn opdrachtgever bij defensie. Deze analyse is niet deel van dit PvA maar zal bijgevoegd worden aan mijn totale documentatie.

Mocht u vraag hebben naar deze analyse kunt u contact met mij opennemen, maar dit zijn voornamelijk risico’s waar ik persoonlijk goed op moet blijven letten.

Bronnen die ik waarschijnlijk ga gebruiken:

Bibliography[i]

Terminologie voor textmining en basis informatie over textmining:

Jan H. Kroeze, M. C. (2003, September). *Differentiating data- and text-mining terminology*. Opgehaald van ACM DIGITAL LIBRARY: https://dl-acm-org.hu.idm.oclc.org/doi/abs/10.5555/954014.954024

Framework voor internet traffic sniffen:

MITM PROXY. (sd). *introduction*. Opgehaald van MITM PROXY: https://docs.mitmproxy.org/stable/

Intressante aanpak of intressante algoritmes:

Qureshi, M. a. (2016, January). *Utilising Wikipedia for Text Mining Applications*. Opgehaald van ACM Digital Library: https://dl-acm-org.hu.idm.oclc.org/doi/abs/10.1145/2888422.2888449

(Bovenstaande lijkt me erg leuk om te proberen, maar ik ben bang dat dit misschien te ambitieus is. Ik zal deze toevoegen als ik denk dat het waarschijnlijk is dat ik tijd over ga hebben.)

Rafal Rzepka. Wenhan Shi, M. P. (2009, February). *Serious processing for frivolous purpose: a chatbot using web-mining supported affect analysis and pun generation*. Opgehaald van ACM Digital Library: https://dl-acm-org.hu.idm.oclc.org/doi/abs/10.1145/1502650.1502728

Theeramunkong, T. (2000, November). *Passage-based Web text mining (poster session)*. Opgehaald van ACM Digital Library: https://dl-acm-org.hu.idm.oclc.org/doi/abs/10.1145/355214.355246

(Bovenstaande aanpak lijkt me het meest voor de hand liggend dat ik iets soort gelijks zal proberen te maken.)

[i] Niet al deze informatie wil ik gaan gebruiken, of zal worden gebruikt. Dit is meer een indicatie voor welke richting ik denk op te gaan of informatie die ik denk misschien nodig te gaan hebben denk hierbij aan terminologie, of interessante algoritmes en tools.