Thesis Work – Purging Sensitive Data in Logs using AI

1 Introduction

This paper describes a thesis work that is to be carried out at Telia Company in Uppsala, Strandbodgatan 1. We are a unit of about 200 people responsible for Telia Company’s contact center offerings to enterprise customers. The thesis work should be performed by 1-2 students from an engineering, information systems or computer sciences program (2xB.Sc or 1xM.Sc.).

2 Background scenario

In an enterprise, customer service is often handled by a special organizational unit called the contact center. The contact center handles a company’s customer interactions, such as telephone calls, email, chat, etc. A contact center platform is an application framework to handle customer interaction flows and directing each interaction to the most suitable agent. Agents working in the contact center log in to the framework which is configured with the agents’ skills, departments, and other information which is used by the framework to correctly route interactions. Agents are provided with interaction information when the interaction is delivered, and may also have that information sent to backend systems automatically to provide the agent with the background information required to meet customer needs.

The different applications in a contact center platform create large amount of log files, necessary for debugging purposes. As Telia Company is a vendor and supplier of these contact center platforms, customers must occasionally send us log files containing sensitive information. However, with the new General Data Protection Regulation (GDPR, Regulation [EU] 2016/679) forthcoming in May 2018 concerning personal information, the log files must be purged of this sensitive data.

3 Purpose

The purpose of this thesis work is to purge sensitive data from different kinds of text-based files. Log files created by a contact center platform take many shapes. It is therefore difficult to generalize an automated solution that can effectively purge sensitive data. The general idea is thus to use AI (such as machine learning) to train a program to perform these duties.

The work will be based on a contact center platform from Genesys. As a minimum, an intelligent purging tool shall be built at Telia Company’s lab. The following aspects/questions should be taken into account in the report, where the first three are the most important:

1. Use Cases: Define use cases for the tool. What are the common sensitive data in the log files and how can it be formalized to purge data?

2. Design solution: Define and prioritize between possible solutions based on the use cases. Pros and cons? Which machine learning method is best suited for the tool?

3. Build: Create an application that analyzes and purges sensitive data from log files from the different Genesys applications.
4. Advanced functionality: Extend the tool with functions that assist when investigating the log files for erratic behavior – masked identifiers must still be unique.

5. Tests: Benchmark the intelligent purging tool against different competitive solutions. What are the metrics to use for comparison, and what is the accuracy of the tool?

4 Suggested time plan

A preliminary start date would be April/May 2017. For M.Sc. programs it would progress for 20 weeks. For B.Sc programs it would progress for 10 weeks for two students.

Following is a rough estimate of the parts the thesis work should consist of and how many weeks each part needs.

|   | 
|---|---|
| 1-2 | Learn about the Contact Center platform and its different modules. |
| 3-14 (M.Sc) 3-8 (B.Sc) | Define different use cases and implement the tool. |
| 15-17 (M.Sc) 8-9 (B.Sc) | Consider the above listed aspects and prepare a summary with conclusions to be included in the report. |
| 18-20 (M.Sc) 9-10 (B.Sc) | Report writing. |

5 Applications

We look for students from the engineering, information systems or computer sciences programs (B.Sc or M.Sc.) with an interest in IT and telecommunications. The candidates should have good experience in C/C++, Java, Python, .NET, or any equivalent functional, imperative, or scripting programming language and be fluent in Swedish and English. Good communications skills and working well in a team are also important qualifications.

Your application, consisting of a personal letter and an attended courses register printout, can be sent to henrik.z.thalin@teliacompany.com and/or jonas.wulff@teliacompany.com.