Thesis Work – Understanding unstructured data using Machine Learning

1 Introduction

This paper describes a master 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 one (1) student from an engineering, information systems or computer sciences program.

2 Background scenario

Artificial Intelligence technologies are becoming increasingly important for customer service organisations in order to increase service levels, customer satisfaction and internal efficiency. AI applications for customer service may roughly be divided into three types: (1) conversational solutions, (2) analytics, and (3) information retrieval and knowledge processing. Each of these application types utilize somewhat different tools and technologies and aim at solving different types of challenges, although all of them may work together to solve a particular problem. This thesis focuses on the 3:rd type of applications – information retrieval and knowledge processing.

3 Purpose

In any type of organisation, a lot of information resides in unstructured data sources such as documents, online manuals and social forums. This type of information is written for a human being to read and understand. This however makes it more challenging from a data processing perspective when the information contained in such unstructured sources need to be extracted and used to answer questions or give guidance.

Machine Learning techniques offer a powerful way to interpret such data and is used today to summarize text, do content and sentiment analysis, etc. When combined with Natural Language Understanding we can start to ask questions where the answers reside in unstructured sources rather than being provided a priori.

In this thesis, we want to examine the possibility to train a model that can understand content in a (set of) document(s) and be able to give probable answers to user questions.

The corpus and use case are yet to be defined, but a number of use cases already exist that could provide a baseline for examination.

The methods, tools and data used to train the model(s) will be defined as part of the project.

4 Suggested time plan

A preliminary start date would be April/May 2017 extending for 20 weeks.

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 relevant ML products used at Telia and other relevant tools that may be under consideration for thesis work.
3-14 |  Agree on use cases and find suitable methods and tools for the thesis work. Conduct literature studies and train application.
15-17|  Run tests and evaluations according to suitable methods.
18-20|  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 Artificial Intelligence and Machine Learning. The candidates should have experience in Java and Python and have taken relevant courses in AI and Machine Learning. Candidates should 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 mathias.johansson@telia.com.