

RAWDATA Project portfolio – General Introduction

This is a brief description of the **RAWDATA Project portfolio**. We indicate the domain and the problem to work with, but leave out specific details. These details will be revealed in requirement specifications to be published later during the semester. **You are not supposed to start working** on the project now. **Wait until the Portfolio project 1 requirements are published** (around mid September).

The common goal of the project portfolio is to provide a tool to help computer programmers develop skills while they are working. The tool should support two complementary functions – a keyword-based search and a search history that keeps track of what's already retrieved and what parts were the most interesting.

Project portfolio = 4 subprojects + 1 Reflexive Synopsis

The project portfolio includes four subprojects that will relate to areas covered in four corresponding sections of the RAWDATA course. The four subprojects are each supposed to be submitted before the next will be started. In addition, to finish the project portfolio a so-called *reflective synopsis* has to be submitted with deadline by the end of the semester. The reflective synopsis summarizes what is covered in the four subprojects, discuss and present reflections that relate to these and to the combined project as a whole. The synopsis should also provide an overall conclusion and comments on future work.

In brief the four subprojects will aim at building a data repository, developing services to access the repository, refining the retrieval to improve the functionality of the services and finally developing one or more applications utilizing the provided services. A key to a successful result is obviously to draw on a source of useful information within the domain – programming. For this purpose, we have chosen Stack Overflow.

What is Stack Overflow?

Stack Overflow (<http://stackoverflow.com/>) is a question and answer (QA) site for professional and enthusiast programmers, working together to build a library of detailed answers to, in principle, every question about programming. The main goal of the site is all about getting answers. It is not a discussion forum and there is no chitchat. If you register, you will be able to ask, answer, comment and vote on relevance on both answers and questions. As a guest, you can search the knowledge captured in the answers and comment to the more than 16 million questions. Furthermore, answers can be graded, so a best answer can be identified. Due to the success of Stack Overflow it has now become a valuable resource of information for programmers seeking for answers.

In this project portfolio we will build a local data repository and load it with a limited selection of data from publicly available dumps from Stack Overflow.

What are we aiming at?

The task in the project portfolio is to create a Stack Overflow Viewer application (SOVA). The primary aim is to provide functionality for searching information in the data extracted from Stack Overflow. This involves keyword based search to access information, as well as some means of tracking of personal use, like search history, marking of found solutions, etc.