RAWDATA

Portfolio Project 1
Introduction

Henrik Bulskov & Troels Andreasen

The Project Portfolio: Four subprojects

- **Project Portfolio**
  - an independent activity
  - closely linked with this course
  - consists of four subprojects with problems and challenges related to issues covered in the four corresponding sections of the course

- **Portfolio project 1: Database**
  - design and implement databases as well as stored procedures to access data in these

- **Portfolio project 2: Web Service (+IR)**
  - design and implement web services to access and manipulate data in databases implemented in Portfolio project 1.

- **Portfolio project 3: Information Retrieval**
  - improve search and retrieval functionality implemented in Portfolio project 1 and 2 and consider data presentation and visualization aspects preparing project 4.

- **Portfolio project 4: Responsive applications**
  - develop responsive front-end applications that build on services and functionality developed in previous projects.

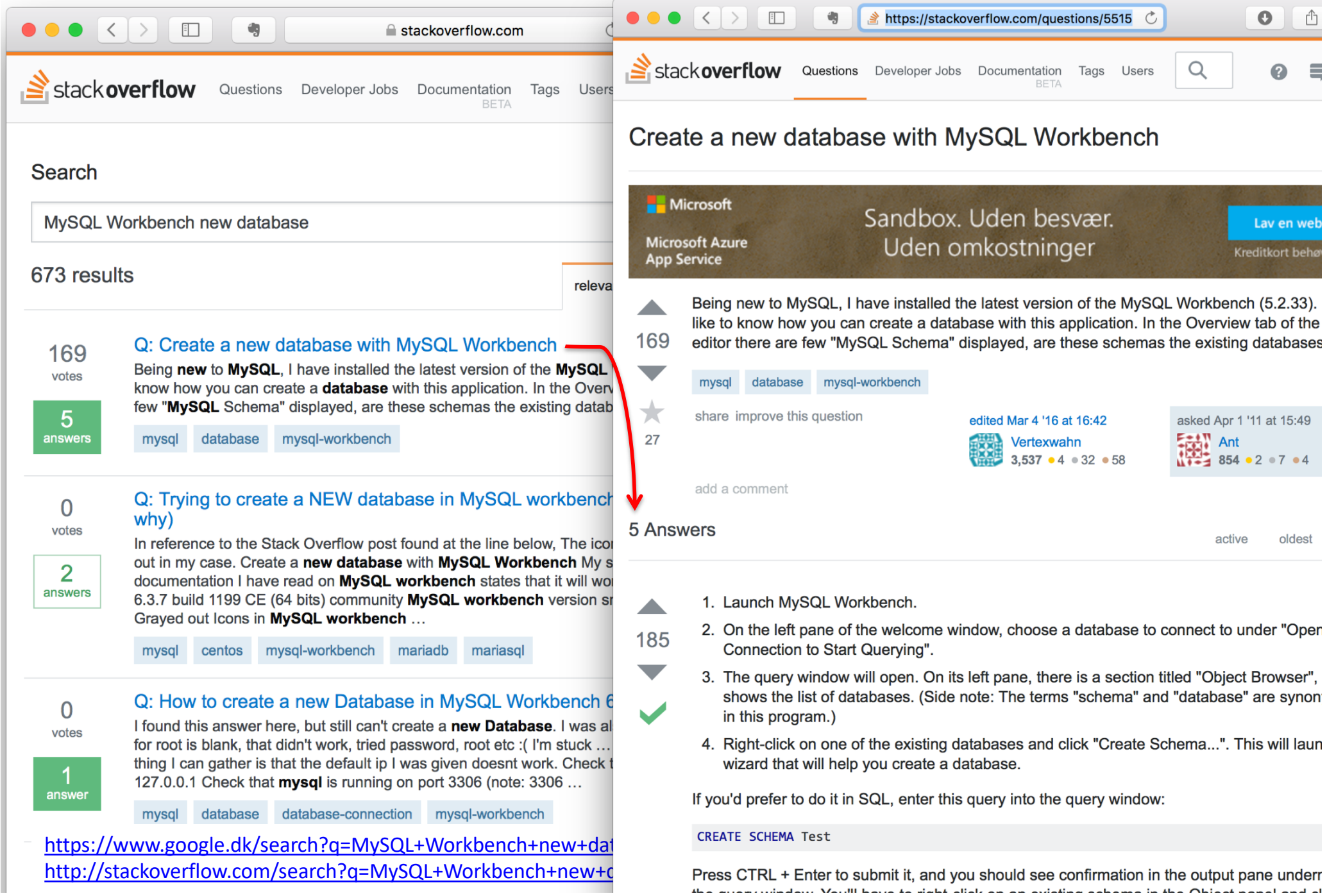
The Project Portfolio

- **Project Portfolio – Problem & Domain**
  - provide a tool to **help computer programmers** develop skills while they are working
  - two complementary functions
    - a keyword-based **search** for answers to questions related to computer programming and
    - a **history and marking** function that keeps track of what's already retrieved (search history) and what parts were the most interesting

- **Key source of data**
  - The Q&A site: Stack Overflow
  - We will use a public available dump of data from this site
    - and we will, to begin with, consider only a very small excerpt of this data

- **What is Stack Overflow?**
  - A question and answer (Q&A) site for programmers,
  - you can search the knowledge captured in the answers and comments to the close to 15.4 million questions
  - answers are ranked and generally of high quality

What is Stack Overflow?



The Project Portfolio goal: SOVA application

- **Our goal: a Stack Overflow Viewer Application (SOVA)**

- **Basic requirements**
  - Search for posts and comments in Stack Overflow.
  - Present search results as (ranked) lists of posts
  - Visualize search results by most frequent words using ranked lists or word clouds.
  - Keep track of search history.
  - Provide a marking option for posts of special interest among posts presented in the search result and allow optional annotation to marked posts.

- **Open-ended set of additional features**
  - Provide statistics and visualize frequent Stack Overflow topics
  - Similar words search
  - Phrase search
  - Browse topics of interest.
  - Build and visualize networks of associated words and or topics
  - Provide alternative visualizations of marked/annotated posts, such as word graphs showing significant words and their relations
  - …
  - Plenty of room for your own ideas here
  - …

however

The Project Portfolio goal: SOVA application

- **Our goal: a Stack Overflow Viewer Application (SOVA)**

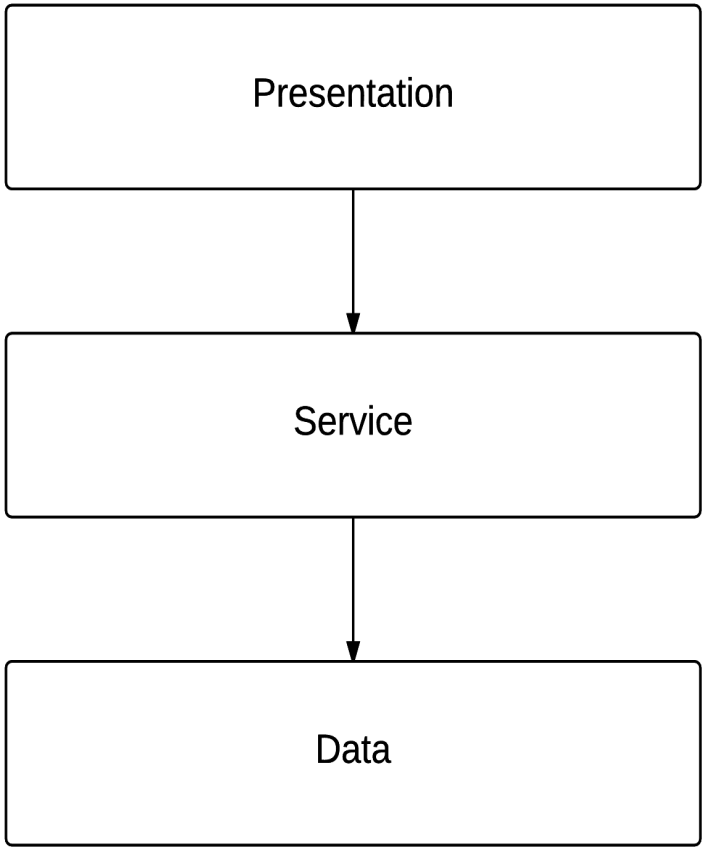
- **Challenge**
  - to decide on a **small** but well chosen **set of features**

- **we are NOT aiming for**
  - the full functionality of StackOverflow

- **we are aiming for**
  - a system with a static database
  - an application that may be single-user

A multilayer architecture

- **The presentation layer**
  - A web based responsive application.
- **The service layer**
  - defines the application logic.
  - provide an interface to the presentation layer through web services.
- **The data layer**
  - encapsulates storage and retrieval of data
  - expose basic functionality related to this



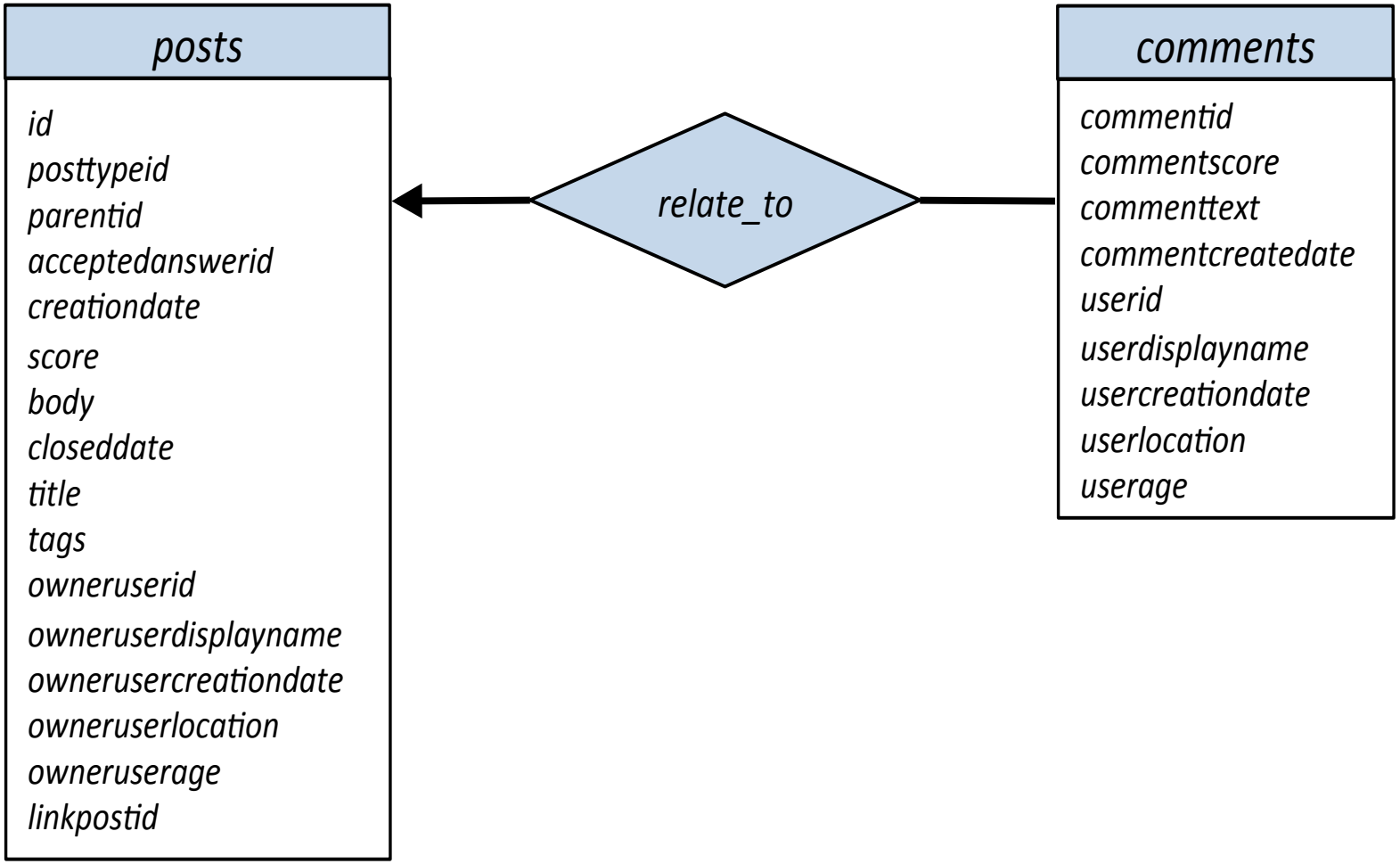
Portfolio project 1

- **The goal**
  - to provide a database for the SOVA application and to prepare the key functionality of the application.
  - two independent data models
    - a QA-data model and
    - a history and marking model
  - provide support for
    - browsing and search for answers and
    - personal activity history and marking of special interest posts.

Starting point

- **Data provided in a two-table database including a *comments* and a *posts* table, on moodle:**
  - stackoverflow_sample_universal.sql

Model for this simple database:



What to do

- **A. Application design**
  - Sketch a preliminary design of your application
- **B. The QA-model**
  - Develop a good design so that
    - all data from the source database (stackoverflow_sample_universal.sql) can be represented
    - your own preferences regarding search functionality are met
- **C. History and Marking model**
  - Design a (complementary) model to meet
    - the basic requirements to the history and marking function of your application
- **D. Functionality**
  - Design and implement key functions to be exposed from the data to the service layer
- **E. Testing**
  - Demonstrate by examples that the results of D work as intended. (More elaborate testing later)

The project report

- **Work in groups**
  - each group:
    - one Portfolio Project 1 report and
    - a product, including the database and the implemented functionality,
- **The report**
  - around 8 normal-pages (8*2400 characters) excluding appendices
  - your Portfolio project 1 report is not supposed to be revised later
  - your product may be subject to revision later if documented in report 2, 3 or 4
- **The submission deadline**
  - for the report as well as the product is 5/3-2018.