studyportals
taking you further

RECOMMENDER

SYSTEM

Project Plan

ABSTRACT

"Because good personalized recommendations can add another dimension to the user experience."

Bilger Yahov
Intern at StudyPortals

PROJECT PLAN

FONTYS UNIVERSITY OF APPLIED SCIENCES

HBO-ICT: English Stream

Data student:	
Family name , initials:	**Yahov, B.Y.**
Student number:	**2442450**
project period: (from – till)	**1st February – 30th June**
Data company:	
Name company/institution:	**StudyPortals B.V.**
Department:	**Engineering**
Address:	**Klokgebouw 280** **5617 AC Eindhoven**
Company tutor:	
Family name, initials:	**Janssen, R.J.M.M.**
Position:	**Senior Software Engineer**
University tutor:	
Family name , initials:	**Kuprys, A.K.**
Project plan:	
Title:	**Recommender System**
Version:	**3.0**
Date:	**27.02.2017**

Approved and signed by the company tutor:

Date:
Signature:

Approved and signed by the university tutor:

Date:
Signature:

Agreed and signed by the student:

Date:
Signature:

Table of Contents

1. Formal Client

StudyPortals is a company based in Eindhoven, the Netherlands. It provides an online study choice platforms listing more than 100 000 undergraduate and postgraduate programs worldwide, as well as other resources. Started out as MastersPortal.eu at the end of 2007, nowadays the team consists of over 120 people with more than 30 different nationalities.

StudyPortals' mission is to find the best fitting study for a student. Their mission is: "Empowering the world to choose education." How? By making study choice transparent globally. They help universities with easier and more effective international marketing and recruitment solutions.

People at StudyPortals are driven to realize a positive contribution to the world with their platforms. Beyond that, the StudyPortals Foundation realizes charitable projects related to their mission or team, mostly with the UNICEF Schools for Africa campaign.

2. The Intern

Bilger Yahov is an excellent student at Fontys University of Applied Sciences in Eindhoven, the Netherlands. The course followed by him is Information and Communication Technology with specialization in Software Engineering. He starts working at StudyPortals in mid-January 2016 as a part-time Web-based Software Engineer. By the time his internship starts, he decides that the best choice for a company would be StudyPortals. His choice is based on the knowledge that he has gained there for the time, which he has worked as a part-time engineer. Continuing his full-time internship at StudyPortals would open him even more opportunities than he has had before.

3. Current Situation

Students can define their preferences explicitly (filling in the information in their profile) or implicitly (their behavior on the website). At the moment, recommendations are based on the explicit preferences - discipline and country of interest. This is not sufficient to make a best fitting recommendation, so there is a need for more information from the student.

There is also a rich set of data that currently goes underused: the implicit behavior of students on the website. If it is known that a student searches mostly for studies that are in a certain budget range or region, a good prediction can already be made, of what they're going to be interested in; one just has to follow that trail.

StudyPortals as a company has been focusing mostly on the exploration. That is, students search and filter to get studies that may interest them. They are now shifting towards recommendation, which requires analysis of patterns of user interest in studies, disciplines, organizations. Because good personalized recommendations can add another dimension to the user experience, StudyPortals wishes to make such a recommender system a salient part of their websites.

4. Project Justification and Expected Product

At the moment StudyPortals' system to recommend studies to students is underdeveloped. As a result of this the company cannot offer the best possible user experience. The management team has decided that there is a need for such a system.

The task is to develop a recommender system, which eventually will get integrated in the products of the company. The newly built system will make much more accurate recommendations to users and will improve the user experience offered.

The intern has to build a system which will work with two sources of data. One of the sources, being information from the database, which is gathered through user's profile. The second source of information underlies in the user behavior. By user's behavior it is meant: visited study programs, visited university pages, time spent on a university page, number of programs viewed in a certain discipline and much more. Therefore the crucial point is to build an engine, which will track and analyze user's behavior. Based on that, the data will be processed using algorithms. As a result the best possible recommendations will be offered to the particular user.

The recommender system, will represent a self-learning engine, which will run silently on StudyPortals' websites.

5. Project Deliverables and Non-Deliverables

As the internship topic already reveals, the intern will be working on creating a recommender system for StudyPortals' web sites. Throughout this process there are a couple of important deliverables which will be provided by the student to the company. Both sides agree on deliverables for the project being:

- Project Plan

  Document, which will be act as a contract between the intern and the company, will contain the project justification and expected product as well as a time schedule for each planned milestone.

- Communication Plan

Plan in which deadlines and hand-in methods for each milestone are specified.

- Recommender System as working software

  Self-learning engine, which will run silently on StudyPortals' websites and will provide the users with best possible recommendations.

- Documentation about the software implemented

  All the source code written will be well-documented. All the algorithms used for mathematical processing will be explicitly explained as documentation.

- Graphical representation of the system's infrastructure – flow charts document

  Illustrations, which will help understanding the underlying architecture of the engine.

- Project Report

  Report describing the whole work process of during the internship, including the research that has been done, important key points, difficulties throughout the process and a self-reflection part.

- Project Presentation

  Final presentation of the project in front of the Company tutor and the University tutor.

As every project, the current one also has important keys, which need to be explicitly specified as non-deliverables. Non-deliverables are all the documents, software implementations and research topics which will not be handled during the internship process, therefore will not be delivered by the intern to the company. They can be found listed below:

- Any software implementations which will deal with something different than offering recommendations

  The internship project will be limited only to delivering working recommender system, supported by the needed documentation.

6. Project Risks

Risk identification is the process of determining risks that could potentially prevent the program, enterprise, or investment from achieving its objectives. Therefore the current

project contains important factors which can be potentially determined as risks. Below is given a table with a brief description and chances to occur of each:

Risk	Chances to occur	Prevention	Impact
Intern not having enough knowledge to implement the Mathematical algorithms for data processing	High	The whole process of work will be supported by a data scientist, who will be continuously helping the student with Mathematical problems	High
Implemented recommender system does not provide the most accurate recommendations	Medium	If problem is related with the algorithms used, there will be a moment of inspection what could have gone wrong and based on the results, ways to improve will be searched for	Medium

7. Project Phasing and Timing

The internship will last roughly 22 to 23 weeks. During this period there are several important deadlines. At each deadline there is a deliverable to provide. Below there is a list of project deadlines. Date and deliverable is specified for each.

In order to be sure about the dates, which each week covers, please do not forget to look at the Excel sheet attached.

1. Draft Project Plan
    a. Deadline: By the end of week 2
    b. Deliverable(s):
        i. First draft version of Project Plan
        ii. First Project Planning Schedule
        iii. Invitation for University tutor visit
2. University tutor visit
    a. Deadline: by the end of week 3
    b. Expected result(s): Feedback on Project Plan

3. Final Project Plan
    a. Deadline: By the end of week 4
    b. Deliverable(s):
        i. Final version of Project Plan
4. Iterations over the project and writing the supporting documentation
    a. Deadline: By the end of week 17
    b. Deliverable(s):
        i. Finalize working on project and documentation
5. Final Project and Internship report
    a. Deadline: By the end of week 18
    b. Deliverable(s):
        i. Final version project
        ii. All supporting documentation including the Internship report
        iii. Arranging the University tutor visit day and presentation day
6. University tutor visit
    a. Deadline: By the end of week 19
    b. Expected result(s): Feedback on internship process
7. Final presentation
    a. Deadline: By the end of week 20
    b. Deliverable(s):
        i. Final presentation in front of Company tutor and University tutor

8. Quality

Throughout the work process there will be many quality inspections. They will mainly assure that:

- Documentation written is well-descriptive
- Software implemented is well-structured, adhering to good-practices, well-documented and last but not least – well-functioning
- Software Design concept images are well-descriptive and they contain all the key parts of the system

The points described above will be kept on a good level, by making sure that there are regular inspection sessions. Those sessions will be conducted by the Intern. It will be his responsibility taking care about the good quality of the project. There might be cases when need from company tutor is required. Those cases will be mainly when code written is complex and contains structures that might be difficult to comprehend.

Furthermore in order to assure the good quality of the system, tests will be written which will eventually point out potential problems and flaws.

9. Skills

In order to successfully finalize the project some very important skills will be needed. First to mention is the fact that the internship is a part of a Software Engineering study. Therefore the most important skills needed and to be gained also, during this project are related with the actual study field. All the programming, infrastructure design, object-orientation, database principles and knowledge gained at Fontys for the last 2 and a half years will be needed during this project.

The two main programming languages used during the implementation phase will be PHP and JavaScript. This will give the intern the opportunity to enrich his competence in those two Web-based software engineering languages.

Apart from the programming skills required, a very big part of the project will also be the Mathematics and logic involved. Since the project is about implementing a smart, self-learning recommender system, this clearly reveals that there will be used quite a lot of mathematical formulas, algorithms and processes. For this reason there will be a Data Scientist assisting the Intern during this project.
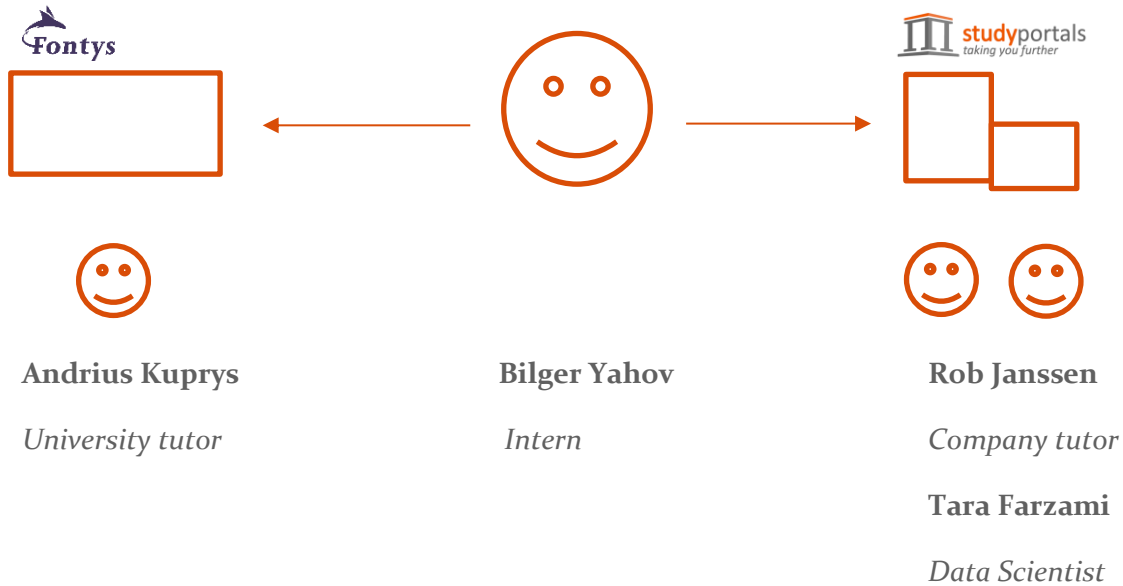
10. Approach to reaching the end goal

Since StudyPortals is a company working with the Agile methodology and the Engineering team working within the Scrum framework, the project will run in Sprints. This would mean that every two weeks there will be meetings between the Intern and the Stakeholders of the project – being the Scrum team in which the Intern is placed. During those meetings the state of the project will be discussed, feedback will be given on the product and future plans will be shaped.

During each Sprint the Intern is going to base his work on the "DOT" framework. The research topic will be based on techniques, best to use when creating a recommender system and algorithms for data processing.

The intern will make a field research on the topic, regarding tracking user behavior and analyzing the gathered data. There will be a research done for similar recommender systems, the ways that they operate and the ways that they offer the best user experience. Due to the fact that there will be Mathematical algorithms used, also research will be done in the sense of how to best optimize them.

11. Organization

Andrius Kuprys

University tutor

Bilger Yahov

Intern

Rob Janssen

Company tutor

Tara Farzami

Data Scientist

Bilger Yahov, being the Intern, will be mainly responsible for his project. He will be communicating with both sides directly.

Andrius Kuprys, being the University tutor, will be kept regularly informed about the project state.

Rob Janssen, being the Company tutor, will be the first-contact person, when any problems occur. His responsibility is to give regular feedback to the Intern and provide help when a situation requires.

Tara Farzami, being the Data Scientist, will be mainly helping the Intern with the complex Mathematical data processing algorithms.

12. Communication plan with University tutor

Communication between the Intern and the University tutor will be done via e-mail messages. Both parties will use their Fontys University e-mail addresses, being:

Bilger Yahov: b.yahov@student.fontys.nl

Andrius Kuprys: a.kuprys@fontys.nl

There will be meetings scheduled between the Intern and the University tutor only when a situation requires so. This will be done via e-mail messaging at least one week before the meeting takes place.