CSIS0801 Final Year Project 2014 – 2015

FYP14003 A Portable and Intelligent Student Interview System

Project Plan

Supervisor: Dr. Cheng Reynold

Cheng Man Fung Kevin  3035042423
Fung Chin Pan       3035044641
Lau Hiu Tsun        3035042423
Tso Hei Lok         3035043738
Table of Contents

Executive Summary ........................................... 3
Section 1: Introduction ........................................ 4
  1.1 Background ........................................... 4
  1.2 Previous Work Analysis ............................. 4
Section 2: Objective ........................................... 5
  2.1 Develop an offline version for the system ........ 5
  2.2 Integrate the system with video conferencing and recording features ............... 5
  2.3 Provide intelligence for smoothening the interview process ............ 5
Section 3: Problem Setting .................................... 6
  3.1 Technical Process ..................................... 6
  3.2 Possible Difficulties .................................. 7
  3.3 Possible Solutions .................................... 7
Section 4: Team Responsibility Allocation ............... 8
Section 5: Project Environment Specification .......... 8
Section 6: Project Schedule ................................ 8
Section 7: Budget .............................................. 9
References ....................................................... 9
Executive Summary

For years, the Computer Science Department of The University of Hong Kong has established an early recruitment scheme for graduate studies. The scheme targets outstanding undergraduate students and master students studying in premier universities, no matter in mainland China or other places. The scheme offers different research areas of which the applicants of the scheme can choose, say, programming language, bioinformatics and algorithms, systems and networking, etc. As part of the recruitment process, an interview has to be conducted. In this project, we will create a student interview system for the better management for the interview process.

As the team continuing the project this year, we would like to have varies enhancements to the developed system. Due to the unavailability of internet connection for some of the interview venue, we would like to develop an offline system to cater the use of the professors. Integration of the interview process onto the system by video conferencing and recording feature is another point to note. Also, we would like to use data mining technique on the pre-interview and post-interview step to provide intelligent interview suggestions and statistics.
Section 1: Introduction

1.1 Background

To recruit excellent students for graduate studies, the Department of Computer Science, HKU has established a scheme offering different research areas for the applicants to choose. To enhance the management of the interview process, an interview management system is developed to reduce the cost and save time. Also, to handle the possible bad network environment, we will develop an offline version for the system for managing the interview system at anywhere and anytime. To further develop the system, there will be new features to be added, such as video conferencing and recording. Providing intelligent suggestions and statistics by applying data mining techniques will also one of the main aspects in this project.

1.2 Previous Work Analysis

An existing system has been developed last year for the interview process. The system was well developed in terms of application form, reviewer management and coordination. However, there is room for improvement for the previous system:

Integration of the interviewing process

Though the system aimed at managing interview process of the recruitment scheme for graduate studies, the face to face interviewing process, which is the core part of the flow of the recruitment scheme, is not included in the system. Video conferencing and recording function is not provided or maintained by the system and it is relying on third party equipment / software.

Internet connection concern

The system previously developed is based on Python Django, which is a web framework. Network capability is the prerequisite for the system to be accessible. It is a huge problem to use it in Mainland China, where the interview is usually held with limited network capability.

Limited Analysis

Analysis based on the students’ information and reviewer’s comments are limited. More data mining can be done to help the interview process and also post interview analysis.

This year, we will try to redevelop the system to achieve more than what the previous work could do and satisfy the user requirement in real case scenario.
Section 2: Objective

2.1 Develop an offline version for the system

To handle the possible bad network environment by providing offline support, which smoothen the whole interview process

2.2 Integrate the system with video conferencing and recording features

To provide an all-in-one system for users not to depend on third-party video conferencing and recording software

2.3 Provide intelligence for smoothening the interview process

To improve the existing interview process by providing managing tools and automated intelligence with the aid of data-mining techniques.
Section 3: Problem Setting

3.1 Technical Process

Video Conferencing and Recording

To build a system with video conferencing so that the professor can watch the live interviews overseas, we decided to use WebRTC to achieve it. WebRTC can achieve video conferencing in Firefox and Chrome without third party plug-in and it is free.

WebRTC will create channels between clients. These channels are created dynamically to make sure that SDP/ICE is exchanged among relevant users. These ensure the data is transmitted security between clients because only the relevant clients can access the data. The video conferencing feature is also planned to have recording function by capturing media stream as per the online demos found.

When there is interview to be conducted, an interface shall be provided to perform recording and a room will be opened for video conferencing on demand. When another user joins the room the video conferencing will start.

Data Mining

Our work would mainly focus on these few categories. First, there will be measurement on degree of positive or negative of comment. Second, extraction of the university names and locations from best students shall be done. Finally, comparison of students’ previous work to his specified research interest in the application shall be done in order to find the consistency of the items mentioned.

Data mining consists of several parts. First step is data preparation. Since data is provided from department and supervisor, most of the data is useful. So, we would extract related information for different process. For the first purpose, we would use Naïve Bayesian Classification, after feeding the training data, to classify the words used in the comments of the reviewer to determine its degree of positive and degree of negative. For the second purpose, we would extract the universities of recruited students for statistical analysis and future uses. For the final purpose, since the applicants had provide the information on research interest and their CV or research proposal on an online system, we can know whether their previous work from CV matches their interest specified by using decision tree. We would present the result as whether they are match or not match.
3.2 Possible Difficulties

WebRTC Recording API unimplemented

WebRTC currently has their recording API unimplemented. Other methods have to be found to capture video/audio form the media stream.

Limited Data Source

For the data mining to be performed, training data is needed to allow the Naïve Bayes Classifier (NBC) to perform correct analysis on the real data being fed into the classifier. Each year, there are only around 200 applicants for the scheme and only around 30 applicants will conduct the interview process (which is the source of comments of reviewers) after screening. The less training data we have, the less accurate the NBC will be.

Security

Our system handles students’ personal data. If data are to be transmitted over the internet in form of plain text, then there can be possible leak of personal data. Measures must be taken to handle security threat.

3.3 Possible Solution

WebRTC Recording API unimplemented

As told by the developer, although WebRTC does not have recording API now, the developer suggests that there is effort on capturing the audio and video, such as RecordRTC.

Limited Data Source

By gathering more information from the previous years, the data source will then be analyzed and being the training data for Naïve Bayes Classifier (NBC).

Security

Data encryption technology will be used to prevent the leak of sensitive data. Encrypted data packets will be sent and be decrypted on the receiver side. Also, code implementation will be done in a secure manner in order to prevent security attack, such as SQL injection. Testing on the system will also be carried in order to eliminate possible failures of the system or exposure of the sensitive data.
Section 4: Team Responsibility Allocation

There are several main parts of our projects: developing offline version of the system, integrating video conferencing and recording features, data mining and infrastructure setting. For these 4 main parts, each of us will be responsible for one of the parts. Captain of our team will also take the role to manage the project progress. Meanwhile, we will support and communicate with each other when needed.

Section 5: Project Environment Specification

- Ubuntu 12.04 Server with MySQL, Node js, Apache, PHP, Java, ffmpeg
- Chrome / Firefox / Opera Stable Version for web version client
- Internet connection required for WebRTC video conferencing feature

Section 6: Project Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 Sep 2014</td>
<td>Submit detailed project plan &amp; finish project web page</td>
</tr>
<tr>
<td>Oct 2014 – Jan 2015</td>
<td>Preliminary implementation</td>
</tr>
<tr>
<td>Jan 2015</td>
<td>Prepare first presentation</td>
</tr>
<tr>
<td>12-16 Jan 2015</td>
<td>First presentation</td>
</tr>
<tr>
<td>25 Jan 2015</td>
<td>Submit detailed interim report</td>
</tr>
<tr>
<td>Jan 2015 – Apr 2015</td>
<td>Finalize tested implementation</td>
</tr>
<tr>
<td>Mar 2015 – Apr 2015</td>
<td>Document final report</td>
</tr>
<tr>
<td>19 Apr 2015</td>
<td>Submit final report</td>
</tr>
<tr>
<td>20-24 Apr 2015</td>
<td>Final presentation</td>
</tr>
<tr>
<td>4 May 2015</td>
<td>Project exhibition</td>
</tr>
</tbody>
</table>
Section 7: Budget

<table>
<thead>
<tr>
<th>Particulars</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budget allocated (4 members X $1000 per member)</td>
<td>4000</td>
</tr>
<tr>
<td>Virtual Server</td>
<td>(700)</td>
</tr>
<tr>
<td>Balance</td>
<td>3300</td>
</tr>
</tbody>
</table>

References

Sushil Kulkarni, Jan 25, 2009, Classification algorithms used in Data Mining:  
http://www.scribd.com/doc/11352378/Classification-algorithms-used-in-Data-Mining-This-is-a-lecture-given-to-Msc-students

Karina Giberta, Miquel Sànchez-Marrèa, Víctor Codinaa, 2010, Choosing the Right Data Mining Technique: Classification of Methods and Intelligent Recommendation:  

Xindong Wu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang · Hiroshi Motoda · Geoffrey J. McLachlan · Angus Ng · Bing Liu · Philip S. Yu · Zhi-Hua Zhou · Michael Steinbach · David J. Hand · Dan Steinberg, 4 December 2007, Top 10 algorithms in data mining:  

Data Mining:  

Data mining algorithms: Classification:  
http://www.cs.ccsu.edu/~markov/ccsu_courses/DataMining-7.html

Imran Khan, Data Mining Basics:  
http://mis343dw.wikispaces.com/file/view/IntroductionToDataMining.pdf

Data Mining: Concepts and Techniques, 2nd Ed.  
The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor  
Morgan Kaufmann Publishers, March 2006

WebRTC Signaling Concepts ® Muaz Khan:  
https://www.webrtc-experiment.com/docs/WebRTC-Signaling-Concepts.html

RecordRTC® Muaz Khan:  
https://github.com/muaz-khan/WebRTC-Experiment/tree/master/RecordRTC