

ANALY482 MEETING MINUTES WITH SUPERVISOR(S)

Date:	08 February 2017
Time:	9:30 – 12:00
Venue:	Meeting Rm 4-1 (School of Information Systems)
Attendees:	Prof Kam, Lu Ning, Song Rui, Dina
Absentees:	
Agenda:	Updates about project progress

S/N	Things Discussed/Done	Remark
1.	Updates about project progress	<p><u>Recap</u> Pattern approach – a function to support the proxy approach</p> <p><u>Law databases</u> How much does this constitute to the data that we have? 20 -30%</p> <ul style="list-style-type: none"> • Prof mentions a lot of data loss • Miss out some of the other services which is big enough • Journal is article by article but books is by book • Finish the data cleaning for law databases and ebrary <p><u>csv file for student demographic:</u></p> <ul style="list-style-type: none"> • Do the join: how many search, borrow you've done → analysis • Size: no response • Need to flag out the zero • http call: how many portal calls are we having from these dataset • Prof mentioned need to have a derived column prof mentioned <p>Prof suggests: start analyzing the data and explore and find out some of the useful info we can get from the data</p> <p><u>Text mining</u></p> <ul style="list-style-type: none"> • Mine the topics or categorization from the user input • Show example text! What happens if u don't get anything • What is the expected example text? e.g. Captured in the first • Record what I'm searching for (keyword) • You need to mine the title!

		<p><u>Prof stated an example:</u></p> <ul style="list-style-type: none"> • The most I can know is that I visit this service, I don't know the content of the book • If we don't have enough information, we should tell the sponsor • If I search something and I don't get what I searched, then it's also info found • Proportion of DOI > proportion of query that contains readable info (article title or user search query): problem • Check with sponsor • EDA: Descriptive/ differential (mainly descriptive) • Basic univariate analysis/ cluster analysis • 1 year of data – 3 teaching terms <p><u>Trendline</u></p> <ol style="list-style-type: none"> 1. Aggregate the count by database 2. Do an aggregation before we step into the EDA 3. If the result is not ideal, it's also part of our findings 4. How do we aggregate? 5. By count of hour? We can use Microsoft SQL tool 6. Lu Ning will share the server with Song Rui <p><u>Clean the user data</u></p> <ul style="list-style-type: none"> • Others, exchange students • Get the queries • Tabulate into csv 7. To do k means – Ray needs to install enterprise miner (Dina ask Prakash) 8. Microsoft data tool: merging the data together (MS SQL) Alex query, view, download
--	--	---

<p>Item Due (Team)/Action(s)</p>
<p>Deadline: 24 February 2017 – Midterms 3pm</p> <p><u>Tasks</u></p> <ul style="list-style-type: none"> • To produce a dashboard for the correlation between student attributes DB access on the timeline (Ray) • To support this, we need to clean up student data, aggregate the query data (Alex) To aggregate query data, Alex needs to finish identifying user access in the query (Alex) • Admin: report, slides, wiki, minutes, presentation (Dina) • K means clustering for databases (Dina) <p><u>Schedule for next meeting:</u></p> <ul style="list-style-type: none"> • Friday 9am • This Sun before 2pm 930am