

Supervisor Meeting 3

Date:	18 Jan 2017
Time:	7.15pm – 8.20pm
Venue:	Prakash's Office
Attendees:	Lim Yu Xiang Bendexter Tan Jun Rong Wang Jing Xuan Professor Prakash
Absentees:	
Agenda:	Proposal run-through & sharing and discussion with Prof

No.	Details	Action By	Due Date
1	<p><u>Proposal Run-Through</u></p> <p>Context: This is our first meeting with Prof Prakash after Prof Kam introduced him as a supervisor, thus there is a need to re-align expectations with Prof Prakash and bring him up to date with what have been done thus far.</p> <p><u>Detailed Comments from Prof on adjusting tidying up the Proposal (In Proposal Format)</u></p> <p>Given 2 extra pieces of literacy reviews to browse through and reference if necessary</p> <ul style="list-style-type: none"> • Cite numbers according to the SMU website • Extensiveness? How do you say this? • Effectiveness: what do we define it as? • Resource don't have <u>capabilities</u>, people have! • <i>Don't know what to do with the data from ezproxy</i>: this is a warehousing problem, not an analysis problem. Not our problem and we don't have to address this. • Must have executive summary • Gravity of the problem: transaction volume, hourly, weekly queries at aggregate levels. (The gravity of the problem, when well-defined, provides weightage to the project and answers the question: is it worth our time or not? If they don't have data, then we can't solve the problem as there would be no backing • Define the problem clearly in points • What are they currently doing, how are they looking into it • Standardize the commas in numbers, eg. 80,000 	Jun Rong & Bendexter	27 Jan 2017

	<ul style="list-style-type: none"> • Project objectives: What sort of user-behaviour? Is it gender based, age based? What timing? Faculty level analysis? Student level analysis? External IP addresses? • This project is a Research-based undertaking→ where we would present findings only. This needs to be communicated to the sponsor, presenting audience and Professors so as to set expectations right at the front of the project presentation • Dataset description: frequency count and name of resource being looked up, click stream in sequential analysis. • Tool: include where it is from, link, main uses of this tool, open-source or not • Dataset record must also include the other dataset given to us (card tap dataset into and out of library) • Why looking for these data points? Eg. Dates of holidays? Show your thought process, how it can impact your end goal? Surge in searches 2 weeks prior to group presentation would be good <p>Questionnaire: let the library send out our questionnaire, architect it and finalize it to send it.</p> <ul style="list-style-type: none"> • Whenever you have something you are unclear, mention the research you have done and you may or may not use the tool. Look for solutions on how to manage the dataset of 2gb+ • Probably techniques, matrix to show u have certain mindset of an exposure of what the end goal might be, the tools we might use. • Descriptive analysis of what? • Risk and limitations to put in table format and how to mitigate, what is your back-up plan? If implemented successfully, how do you test it and present to the clients. Accuracies? Solution may be a plug-in, may completely replace their solution. Needs to have a testing strategy, eg. Clickstream analysis→ looking for SAS book, we can perhaps start recommending SAS tools related search kind of thing. • May test against our questionnaire results too • Possible matrix to test effectiveness of solution • Edit supervisor to Prakash • User behaviour in an online catalogue: what aspects are you looking into? How students are interacting with it 		
--	---	--	--

	<p>in terms of timings, spikes, split by demographics, how this particular group leads to the spike. Use visualizations, filter it.</p> <ul style="list-style-type: none"> • Possible user groups: • Students: basic books? Go back and forth from online to hardcopy resources? • Faculty: advanced research? • Gantt chart: needs to have a split between group members • Number the references, do footnote style 		
2	<p><u>Dataset Run-Through</u></p> <p><u>Questions on the data analysis</u></p> <p>Data dictionary: starting and ending points, patterns. Hash table: then this can be used for everything.</p> <ul style="list-style-type: none"> - There is a key and value. Key-value pair. - Remove white spaces so that it will be a combined string. White spaces are not accounted for <ul style="list-style-type: none"> o Because they are text with semi-structures. o Go through stuff which are not useful or required. o Use framework that can handle spark. Try 'Google big table' o Start with 2 data sources first, build strategy around that one first. Then extend it to fit other data sources. <p>Write complexity of the project: what we don't know</p> <ul style="list-style-type: none"> - Eg. We don't even know how many data source <p>Considerations of the project: write that you signed NDA with library</p> <p>Can talk to LARC smu: they live this issue daily.</p> <p>Interim expectations: work with 2 databases for now, no direction at the moment but look up this thing and present findings to meet deadline. By right data preparation must be done already.</p> <ul style="list-style-type: none"> - Someone should take data mining, data cleaning part of it, whatever. So, it looks good on the presentation. 	<p>Jing Xuan (identifying Key-Value Pairs)</p> <p>Jun Rong (Documenting and Literature Read-up on what can be done for this step)</p> <p>Bendexter (Assist Jing Xuan & Literature Read-up)</p>	28 Jan 2017

	Agreed on deadline of proposal: 27 Jan 2017		
--	---	--	--

The meeting was adjourned at 8.20pm. These minutes will be circulated and adopted if there are no amendments reported in the next three days.

Agenda for next meeting

1. Prof to run-through submitted Proposal and Cleaned Dataset

Prepared by,
Tan Jun Rong

Vetted and edited by,
Bendexter Lim