Course Finder for Executive Education

Kashif Kudalkar
kkudalkar3@gatech.edu

***Abstract*—**For the M Assignments I will be investigating & designing an interface which will help users:
* explore *Executive Education Courses* offered on timestsw.com.
* filter the courses basis various relevant parameters.
* shortlist course(s) most suited to their profiles and/or needs.
TimesTSW is a prominent operator in India offering Executive Education course from premier Indian institutions. There has been a significant surge in demand amongst working professionals to up skill. Hence the number of courses being offered is steadily increasing, thereby overwhelming users with excessive options. A *"Course Finder"* functionality will benefit users & enhance the appeal of the website.

1 PROBLEM SPACE

1.1 Rise in demand led to Increased Supply

The demand for Executive education in India is on the rise (Malik, 2021). It is driven by multiple factors. Rapid changes in the world of business & technology is making existing know-how outdated (Murray, 2019). Organizations require a workforce that is up-to-date (Murray, 2019). Working professionals have realized that up-skilling leads to better career prospects. (Buddhapriya, 2021).
The surge in online searches endorses the increased interest in this regard (Own research conducted using Google Search Trends). The Executive Education sector is now in the focus of many organizations — from large conglomerates to entrepreneurial start-ups.

1.2 Availability of a Multitude of Options & Information

The increase in focus in the sector is bringing in fresh investments & has drastically increased the number of courses being offered by existing & new players. (Arakali, 2020). Courses being offered are not limited to Management or Technology, but now address domains like Supply Chain, Healthcare etc. as well.
Learners also have choices of several learning modes (On-line, Self-paced, In-Campus etc.) as well as the options of different providers and institutes.

The growth in options comes along with a torrent of information associated with the courses as each provider attempts to attract prospects to get enrolments.

1.3 Growing use of Mobile Phones & reduced Screen Real Estate

India is seeing an unprecedented increase in the use of mobile phones as well as mobile internet (EconomicTimes, 2019) This reflects in the web analytical reports for websites where mobile phones now contribute over 70% of the traffic (Own research using Google Analytics implemented on www.timestsw.com).
Users have access to vast volumes of information via mobile internet but are limited by how much they can consume at a time due to the smaller screen.

1.4 Leading to Complexity in Decision Making

For Users, more course options, extensive information, easy access to it at any-place or anytime, under different contexts (e.g. while focused or when distracted) makes decision making difficult (Schwartz, 2006 Lee and Lee, 2004).
Thus, the idea behind the proposed interface is to **be a guide & help users choose the course best suited to their profile and/or requirements**. Having such a functionality will not only benefit users but also enhance the appeal of the TimesTSW brand and website.

2 USER TYPES

At a broad level, the target audience of the interface being designed will be all possible prospects for an Executive Education Course. Hence the User Types will be defined and dictated by the Course requirements. The diversity can be explained using demographic details like:
Geography: Individuals residing in India (citizens as well as expats).
Gender: There is no restriction or limitation.
Age Group: 24 years to about 55 years old.
Education: At least Graduates. Can be Post-Graduates or PhDs as well.
Work Profile: Individuals with white-collared jobs.
Work Experience: Minimum 3 years to about 25 years.
Expertise: Junior to Mid to Senior Management. Fluent with English.
Income: Minimum $ 30,000 (Annual)
Motivation: Likely to have high motivation to identify the best course.

Motivation for engaging in the task will depend on another aspect of looking at

the User Type i.e. focused user vs distracted user. Use of the interface will **not always be at the same level of motivation or focus due to the user being in a different context** (e.g. browsing during a commute vs browsing from a study desk or visit triggered by an ad vs search based visit).

3 NEEDFINDING PLAN 1: ANALYSIS OF EXISTING DATA LOGS

3.1 Where will you find the Data Logs?:

Search Engines are the first & frequent port to call for users nowadays (Statistia, 2020). Google commands a lion's share amongst Search Engines in India (Statistia, 2020 and Statista, 2020). A query entered into a Search Engine is the only instance of **Explicit Expression of Intent** by users. By analyzing Search Queries for Executive Education in India we can:

- Identify the user's intent (*what is the user looking for?*)
- Group the intent into relevant clusters & thus,
- Understand the user's requirements better (*what exactly does the user's goals & what do they need?*)

Since Google's share amongst Search Engines is over 97% in India, we limit the logs to Search Queries entered in Google (Statistia, 2020). We looked back into queries executed over the last 12 months for queries linked only to Executive Education in India. We will extract the queries directly from the Google Search Console (Google, 2021). We will also use a plugin to get the data (Aperghis, 2020). Lastly to be more thorough, we will use the subscription version of the Ahref Tool to get queries related to competition & hence cover the domain comprehensively (Ahref, 2021).

3.2 How will you analyse them Systematically?:

The following course of action will be followed:
Step 1 - Data Pre-Processing:
This will include: converting the queries to lowercase; removing non-letter, non-number & non-space characters; removing duplicates.

Step 2 - Data Preparation:

- *Tokenization* of the words will be done using NLTK (NLTK Project, 2021).
- *Stopwords* will be removed using the collection of English stopwords from Natural Language Tool Kit (NLTK).

• Between *Stemming* & *Lemmatization* for *Text Normalization*, I will use *Lemmatization* as it gives meaningful words. *Stemming* can produce non-meaningful words which are not present in the dictionary. We want to identify the Users Search Intent for which we need words that make intuitive sense.
• The queries will be *reconstructed'* as we want to cluster queries as whole.
• Outliers and/or irrelevant queries will be removed.
• Any duplicates will be also be removed.

Step 3 - Data Analysis:
• Search Query logs do not have labels, so we are dealing with an Unsupervised Learning Problem. And since we want to group similar queries together, we are looking at a Clustering Analysis.
• To prepare the text (queries) for clustering, I will use *Word2Vec* (which uses Neural Networks). It will provide similar vectors of similar words which we can aggregate over a query.
• The high-dimensional data will be reduced to a manageable structure using *Principal Component Analysis.*
• To get the best result, I will explore *Soft Clustering* and *Hard Clustering* techniques.
• For *Soft Clustering* approach, wherein we get a probability of cluster association. I will use *Gaussian Mixture Models (GMM)*.
• For *Hard Clustering*, where cluster associations are absolute, I will explore *K-Means*, *Spectral Clustering* & *DBSCAN*.

3.3 Needfinding Plan 1 & Data Inventory:

Analysis of Existing Data Logs should address the following from the data inventory:
- *What are their goals?*
- *What do they need?*

3.4 Influence of Potential Biases & Associated Precautions:

There is a chance of *Confirmation Bias* creeping in during the final step when the resultant clusters are to be identified & named. To avoid this bias, a neutral person will be employed to validate the cluster nomenclature.

3.5 Recurring Search Query Analysis:

We will re-run the Search Query Log analysis (outlined above) at a monthly frequency to keep track of changes in the Intent behind the Search Queries.

4 NEEDFINDING PLAN 2: THINK-ALOUD

The participants for this task will be from the target group of the courses being offered.
The objective here is to understand the process users follow to accomplish the task. Hence we will not be restricting the activity to any interfaces i.e. neither our own website (timestsw.com, for which we are working on creating the new interface), nor any competition website or any other specific website(s).

4.1 What will participants do?

The participants will be asked to:
• Assume that they are looking for an Executive Education Course for themselves.
• They can use the device of their choice.
• They should follow a process that comes naturally to them, &
• Find a suitable course which they would consider enrolling to.

4.2 What will you ask them while or after they do it?

Prior to Starting: At the outset the participants would be informed that we are not evaluating them. Rather we are trying to build a better system which does not exist. (To avoid Social Desirability Bias).

From this Needfinding exercise, we are attempting to understand each of the steps users take to fulfill the task provided to them. This includes:
• How do users initiate the task.
• The sequence of steps they take to achieve the task.
• Their reason for taking each particular step.
• What outcome were they expecting when they took the step.
• Was the outcome as they expected.
• Does the process vary across devices and if yes, how does it vary.

The participants will be asked to 'Think-Aloud' regarding the above mentioned points as they work through the task.

During the Attempt: While they are attempting the task, we will avoid interrupting the participant. We will prompt the participants to comment on a particular aspect if they miss out on it.

Post the Attempt: We will be observing & taking notes of the users comments during the 'Think-Aloud Session'. If we come across any specific participant action or comment during the session which warrants further investigation, we will not interrupt the user. We will follow-up with further questions once the session has been completed. The follow up questions will typically around understanding *what did the participant mean by the comment* and *why did the participant think/feel that way*).

4.3 Needfinding Plan 2 & Data Inventory:

This *Think-Aloud* exercise should help answer the following questions of the data inventory:
- *What are their goals?*
- *What do they need?*
- *What are their tasks?*
- *What are their sub-tasks?*

4.4 Influence of Potential Biases & Associated Precautions:

This Needfinding exercise is likely to be affected by the *Observer Bias* wherein I would subconsciously influence the participants as well as *Social Desirability Bias* where participants might tweak their comments in a manner they feel will be viewed favorably by me.
To curb Observer Bias, I would take care of certain aspects like: I will use a predefined script to explain the process & for any further interaction to ensure I do not effect the participants. I will not interrupt the participant until the flow ends.
To curb Social Desirability Bias, I would take care to explain to the participant that we are exploring the process & neither me nor they are being tested.
Also refraining from giving a feedback or commenting on users opinions will be helpful to address both biases.

4.5 'Think-Aloud' for the New Interface:

A similar 'Think-Aloud' approach will be deployed at a later stage of the design process to evaluate the efficacy of the new interface. The only difference will be

that we will restrict the user to perform the task on the newly designed interface.

5 NEEDFINDING PLAN 3: SURVEYS

We will employ surveys to complement the other Needfinding methods listed earlier. Using the above mentioned methods & approaches we will get a fair idea about the goals, needs, tasks & sub-tasks associated with users.
The gap that needs to be filled is to understand the different contexts surrounding the users. This includes the various circumstances under which the user is likely to encounter the interface. This is where a *Survey* fits into the scheme of things. In addition the survey will help validate the description & motivations outlined in *User Type*.

5.1 What will you ask?:

The survey questions will revolve around understanding the following:
• Where are the users?
The users are in which parts of the country (cities, towns etc). This will also include where are they physically (at home, in office, in commute etc.) when they execute the task.
• What is the context of the task?
Building up on their actual location when executing the task, we will get an idea regarding what else competes for their attention.
• What are their goals?
Understand the driving force behind users' actions & their motivations.
• What are their tasks?
To be able to complete their tasks, what sources of information do they seek to increase knowledge/understanding.
• What do they need?
Here we will try to understand the elements of the information they seek.
• Who are the users?
Finally, we'll have questions exploring about the participants demographic characteristics like age, education, work experience, designation, industry/domain, income group & marital status.

5.2 Who will you send the survey to?:

The survey will be sent out to a target audience who fits the description listed down in *Section 2: User Types*. Since it will be difficult to include all the parame-

ters defined in the User Type in the Survey targeting, we will only consider:
Geography: Individuals residing in India (citizens as well as expats).
Age Group: 24 years to about 55 years old.
Education: At least Graduates. Can be Post-Graduates or PhDs as well.
Work Experience: Minimum 3 years to about 25 years.

5.3 Needfinding Plan 3 & Data Inventory:

Surveys should throw light on the following aspects of the data inventory:
- *Who are the users?*
- *Where are the users?*
- *What is the context of the task?*
- *What are their goals?*
- *What do they need?*
- *What are their tasks?*

5.4 Influence of Potential Biases & Associated Precautions:

We will need to watch out for Non-response Bias if a certain section of the prospective participants do not respond. Response Bias, leading to less-than-truthful responses, can be an issue. Confirmation bias might creep in at the time of analysis. We will also need to be aware of Leading Questions, Loaded Questions, Double-Barreled Questions, Absolute Questions & Unclear Questions.
Some of the methods to steer clear of biases that we will employ are using a multiple-choice questionnaire, use of a Likert Scale rather than a yes-no scale, include “Prefer Not to Answer"for questions which may not be applicable to some participants, Include all options or a catchall “other” option, Ending with Demographic and Personal Questions, Resurveying Non-Respondents etc. Also we will test out the questions before rolling out the final survey.

6 CONCLUSION

The focus of this Needfinding is to answer the question *What problem are we trying to solve?*. We want to keep the evaluation more user & task based and avoid drifting towards concerns regarding the interface. Since we see a surge in demand in this domain, this study will help address user needs via the proposed new interface.

7 APPENDICES

Data points which the survey questions will look at addressing are as follows.
Note - All questions are related to the task of *trying to find a suitable Executive Education Course for themselves*

1. Where are the users likely to be when doing the task?

Possible Locations	**Very Unlikely**	**Unlikely**	**Neither Unlikely nor Likely**	**Likely**	**Very Likely**
At Home:					
In Office:					
During Commute:					
Others (Please mention)					

2. What are the possible triggers that can initiate the task?

Possible Triggers	**Very Unlikely**	**Unlikely**	**Neither Unlikely nor Likely**	**Likely**	**Very Likely**
Television Ads:					
Newspaper Ads or Article:					
Magazine Ads or Article:					
Outdoor Ads:					
Online Ads or Article:					
Recommendation from an acquaintance:					
Others (Please mention)					

3. What are the top 3 reasons for getting started with the task?
   This will be an Open Ended Question.

4. How often do you use the following sources of information for the task?

Possible Sources	Never	Rarely	Occasionally	Frequently	Very Frequently
Search Engines:					
Social Media:					
Friends & Family:					
Professional Connections:					
Newspapers:					
Magazines:					
Specific Websites:					
Others (Please mention)					

5. If Specific Websites marked as 'Likely' or 'Very Likely' in the question above, please list the website names.
   Open Ended Question.

6. Which course related information is important for selection?

Course Information	Not Important	Low Importance	Average Importance	Very Important	Absolutely Essential
Course Category:					
Course Type (Certificate | Diploma | Degree):					
Duration & Family:					
Eligibility Criteria:					
Fees:					
Admission Process:					
Institute:					
Delivery Type (Online | On-Campus | Hybrid):					
Class Schedule:					
Faculty:					
Alumni Status:					
Others (Please mention)					

7. Who are the users & where are they? Demographic Information -
   Age Group:
   Education:
   Work Experience:
   Current Designation:
   Current Industry/Domain:
   Income Group:
   Marital Status:
   City/Town of Residence:

REFERENCES

[1] Ahref (2021). *Ahref*. URL: http://ahref.com/.

[2] Aperghis, MA (2020). *Search Analytics for Sheets*. URL: https://searchanalyticsforsheets.com/.

[3] Arakali, HA (Sept. 2020). *Fresh investments in edtech sector to address urgent need*. URL: https://www.forbesindia.com/article/leaderboard/fresh-investments-in-edtech-sector-to-address-urgent-need/62911/1.

[4] Buddhapriya, SB (May 2021). *How executive education or courses for employees can upskill you and boost your career*. URL: https://www.indiatoday.in/education-today/featurephilia/story/how-executive-education-or-courses-for-employees-can-upskill-you-and-boost-your-career-1806479-2021-05-24.

[5] EconomicTimes, ET (Apr. 2019). *Internet users in India to rise by 40%, smartphones to double by 2023: McKinsey*. URL: https://economictimes.indiatimes.com/tech/internet/internet-users-in-india-to-rise-by-40-smartphones-to-double-by-2023-mckinsey/articleshow/69040395.cms?from=mdr.

[6] Google (2021). *Google Search Console*. URL: https://search.google.com/search-console/.

[7] Lee, Byung-Kwan and Lee, Wei-Na (Mar. 2004). "The Effect of Information Overload on Consumer Choice Quality in an On-Line Environment". In: *Psychology and Marketing* 21, pp. 159–183. DOI: 10.1002/mar.20000.

[8] Malik, AM (Jan. 2021). *Trends that will give edtech a quantum leap in India in 2021*. URL: https://www.indiatoday.in/education-today/featurephilia/story/trends-that-will-give-edtech-a-quantum-leap-in-india-in-2021-1758985-2021-01-14.

[9] Murray, SM (June 2019). *What Next For Executive Education?* URL: https://www.businessbecause.com/news/emba/6094/what-next-executive-education.

[10] NLTK Project (May 2021). *Natural Language ToolKit*. URL: https://www.nltk.org/.

[11] Schwartz, BS (June 2006). *More Isn't Always Better*. URL: https://hbr.org/2006/06/more-isnt-always-better.

[12] Statista (June 2020). *Share of Google in the desktop search market across India from June 2019 to June 2020*. URL: https://www.statista.com/statistics/938796/india-google-share-in-desktop-search-market/.

[13] Statistia (Dec. 2020a). *Leading websites in India as of December 2020, by traffic*. URL: https://www.statista.com/statistics/1108779/india-websites-ranking-by-traffic/.

[14] Statistia (May 2020b). *Share of Google in the mobile search engine market across India from April 2019 to May 2020*. URL: https://www.statista.com/statistics/938846/india-google-share-in-mobile-search-market/.