

Research Methodology

Method of Data Collection



Contents

- 1) Data Sources
- 2) Types of Data
- 3) Data Collection Approaches
- 4) Examples

1) Data Sources

Why need data?

- In a research, **data** is needed for
 - **Discovering** new knowledge
 - **Testing** a proposed solution, such as algorithm, tool, model, etc.
 - **Comparing** purpose
 - Etc.

Where do data come from?

- Data can come from
 - Database, exist **corpus**
 - On **Internet** (by crawling)
 - Survey, **questionnaires**, interview
 - Experimentation
- Existing data (most) is required to
 - Pass through a **pre-processing step** (before can be used)
 - Cleaning, filtering, correcting
 - Translating (meaning or languages)

An Important Question to ask

- What you **must ask yourself**
 - Will the data **answer** my research question?
- To answer that
 - You **must first decide**
 - What your research question(s) is/are?
 - Then, you **need to decide**
 - What data/variables are needed to scientifically answer the question?

Data Collection Choice

- If we are starting from scratch, how do we collect/find data?
 - **Primary data**: data you collect
 - **Secondary data**: data someone else has collected
- If **data exist in secondary form**
 - Then use them to the **extent** as you can (keeping in mind limitations)
- But if **it does not**, and you are able to find primary data
 - Then you choose a data collection method

2) Types of Data

Data Types

- In general, the data can be
 - Primary data
 - Secondary data
- In **IT** field (mostly in the machine learning model research work)
 - Learning/Training data
 - Testing data

Primary Data – Limitations

- Do you have the time and money for
 - Designing your **collection instrument**?
 - Selecting your **population** or **sample**?
 - Pretesting/piloting the instrument to work out **sources of bias**?
 - Administration of the instrument?
 - Entry/collation of data?

Examples of Sources of Primary Data

- These are examples of sources of primary data
 - Surveys
 - Focus groups
 - Questionnaires
 - Personal interviews
 - Observational study
 - Experiments
 - Ground truth process

Secondary Data – Limitations

- When was it collected? For how long?
 - May be **out of date** for
 - What you want to analyze
 - May not have been collected **long enough** for
 - Detecting **trends** (a general direction – developing or changing)
- Is the **dataset** complete?
 - There may be **missing information** on some observations
 - Unless such missing info. is **caught** and **corrected**
 - Or analysis will be **biased**

Secondary Data – Limitations

- Are the data **consistent/reliable**?
 - Did variables **drop out** over time?
 - Did variables **change** in definition over time?
- Is the information exactly what you need?
 - In some cases, may have to use “**proxy variables**”
 - Variables that **may approximate something** you really wanted to measure

Secondary Data – Advantages

- **No need to reinvent the wheel**
 - If someone has already found the data, take advantage of it
- **It will save you money**
 - Even if you have to pay for access
 - It is often cheaper in terms of money than collecting your own data
- **It will save you time**
 - Primary data collection is very time consuming

Secondary Data – Advantages

- **It may be very accurate**
 - When especially a government agency has collected the data
 - Incredible amounts of time & money went into it (probably highly accurate)
- **It has great exploratory value**
 - Exploring research questions and formulating hypothesis to test

Examples of Sources of Secondary Data

- These are examples of sources of secondary data
 - Hospital, clinic, school nurse records
 - Vital statistics – birth, death certificates
 - Private and foundation databases
 - Surveillance data from state government programs/organization
 - Etc.

3) Data Collection Approaches

Rules for Collecting Data

- Use **multiple** data collection methods
- Use **available** data, but need to know
 - How did the **measures** are defined?
 - How did the data are **collected** and **cleaned**?
 - How **accurate** the data is?
 - Does it **require** cleaning?

Rules for Collecting Data

- It must collect **original** data
 - Be **sensitive** to burden (trouble) on others
 - Establish **procedures** and follow them (protocol)
 - Maintain **accurate records** of definitions and coding
 - Verify **accuracy of coding, data input**

Data Collection Methods

Obtrusive

- Directly **obtain information** from those being evaluated
 - e.g. interviews, surveys, focus groups

Unobtrusive

- Do **not collect information** directly from evaluation
 - e.g. document analysis, Google Earth, observation at a distance

- There are **2** types of data collection approach
 - **Structured Approach**
 - **Semi-structured Approach**

Structured Approach

- All data collected in the **same** way
- Important for **multi-site** and cluster evaluations (can compare)
- Important to **make comparisons** with alternate sites
- When does structured approach use?
 - Need to make comparisons across **different sites** or interventions
 - Know what needs to be **measured**
 - Need to address **extent questions**
 - Have a **large** sample or population
 - Need to show **results** numerically

Semi-Structured Approach

- Not collected in the **same way** every time
- More **open** and **fluid** (no fixed shape)
- Not follow a **rigid** (fixed) **script**
 - may ask for **more detail**
 - people can tell what they **want** in their own way
- When does **semi-structured approach** use?
 - Conduct **exploratory work**
 - Seek to understand **unexpected results, themes, and/or issues**
 - Need **narratives** or **stories**
 - Want in-depth, rich, “**backstage**” information

Characteristics of Good Measures

- 4 **components** to express characteristics of good measures
 - Is the measure **relevance**?
 - Is the measure **credibility**?
 - Is the measure **validity**?
 - Is the measure **reliability**?

Relevance and Credibility

- **Relevance**

- Does the measure capture what **matters**?
- Do not measure what is **easy** instead of what is needed

- **Credibility**

- Is the measure **believable**?
- Will it be viewed as a **reasonable** and **appropriate way** to capture the information sought?

Validity and Reliability

- **Internal Validity**

- How well does the measure capture what it is **supposed to**?

- **Reliability**

- A measure's **precision** and **stability**
 - Extent to which the **same result** would be obtained with **repeated trials**

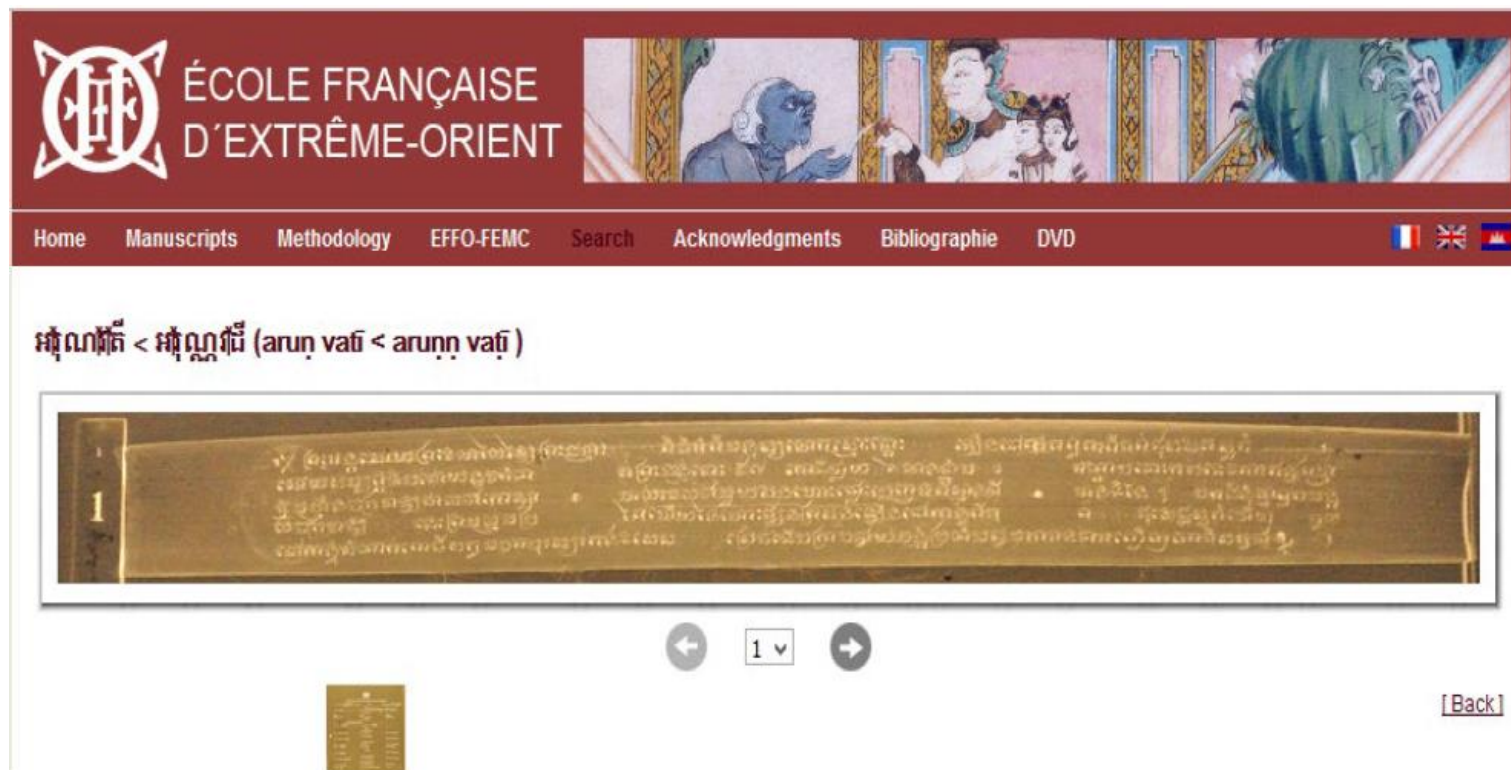
Triangulation to Increase Accuracy of Data

- **Triangulation of methods**
 - Collection of same information using **different methods**
- **Triangulation of sources**
 - Collection of same information from a **variety of sources**
- **Triangulation of evaluators**
 - Collection of same information from more than **one evaluator**

4) Examples

Digitization and Indexation of Khmer Ancient Documents – Palm Leaf

- Existing Digitized Data: **EFEO**



Digitization and Indexation of Khmer Ancient Documents – Palm Leaf

- Existing Digitized Data: **Buddhist Institute**



Digitization and Indexation of Khmer Ancient Documents – Palm Leaf

- **Digitization Campaign**



Romanization of Khmer: Automatic Latin-To-Khmer-Based Text Conversion

- **Training Data:** Student list, Facebook comments, Dictionary

Nº	ID	Name in English	Name in Khme
1	e20140009	ANG SOPHEAK	អាង សុភ័ក្រ
2	e20140014	BLANG VIBOL	ប្លង់ វិបុល
3	e20140048	CHEATH SIEVPHYGN	ជាតិ សៀវភិញ
4	e20140049	CHEM BUN HONG	ចែម ប៊ុនហុង
5	e20130061	CHHAT BUNSREYLY	ឆាត ប៊ុនស្រីលី
6	e20140058	CHHEANG VANNY RATANAK	ឈាង វ៉ាន់រតនៈ
7	e20140062	CHHENG RATANA	ឆេង រតនា
8	e20140066	CHHIENG MENGKEA	ឈឿង ម៉េងកា
9	e20140070	CHHIM SOVAN RITH	ឈឹម សុវណ្ណរិទ្ធ
10	e20140078	CHHORN BUNCHHUONG	ឆន ប៊ុនឈួង
11	e20140096	CHIM SOPHEAKTRA	ជឹម សុភត្រ័ត្រា
12	e20140099	CHIV NENG	ជីវ ណេង
13	e20140074	CHOEB VANNARY	ជីប វណ្ណារី
14	e20130119	CHOU SEAKNY	ជូ សៀកនី
15	e20151107	DANH THI HONG HOA	យ៉ញ ធី ហុង ហ្វា

Romanian-English Machine Translation

- Data from News website

Ultima actualizare: 22 Noiembrie, 2005 - Published 15:49 GMT

Trmite un prieten Versiune pentru tiprire

Ucraina: un an de la Revoluția Portocalie

Ucraina aniversează un an de la protestele pro-democrație de după alegerile de anul trecut, cunoscute acum drept Revoluția Portocalie. Demonstrațiile au izbucnit în urma anunțării rezultatelor alegerilor prezidențiale și au durat trei săptămâni, determinând în cele din urmă venirea la putere a liderului opoziției, Viktor Iuschenko.

Aniversarea de azi este marcată printr-un discurs al președintelui Iușchenko și un concert în aer liber.

Piața Independenței din Kiev e împodobită din nou în portocaliu. Aici au început protestele în masă de acum un an.

Sondajele de opinie arată că mulți ucrainieni sunt dezamăgiți de lipsa de succes a noii puteri. Promisiunile nu au fost tinute.

Printre cei aflați în Piața Independenței din Kiev se numără și câțiva care sunt dispuși să fie răbdători:

"Nu mă așteptam să se întâmple prea multe într-un an și îmi dau seama că președintele Iușchenko are nevoie de mai mult timp ca să își pună ideile în aplicare".

De altfel, vorbind înaintea aniversării, președintele Iușchenko a

Last Updated: Tuesday, 22 November 2005, 15:26 GMT

E-mail this to a friend Printable version

Ukraine marks Orange Revolution

Tens of thousands of Ukrainians are gathering in Kiev to mark the first anniversary of the mass street protests known as the Orange Revolution.

The demonstrations were in support of Viktor Yushchenko, the losing candidate in a rigged presidential election.

Mr Yushchenko - now president - is due to address the nation on a stage in Independence Square.

Analysts say many expectations in 2004 were unrealistically high and some key pledges have not been realised.

But President Yushchenko has urged people to focus on the past year's achievements, which, he says, include greater democracy, reports the BBC's Helen Fawkes in Kiev.


Day of Freedom

Independence Square - known in Ukrainian as Maidan - has turned orange once again, as people gather in the square - many of them wearing scarves or waving flags in the bright colour of the revolution.



Just like last year it is a freezing cold day in the capital, our correspondent says.

Source: Machine Translation, KAING Hour

Travel Recommendation System

 **tripadvisor**
NEW ZEALAND

Angkor Wat, Siem Reap

  Review


Siem Reap ▾ Hotels ▾ Flights Holiday Homes Restaurants Things to Do Forum Best of 2017 More ▾


Find: Things to Do

Near: Siem Reap, Cambodia

Asia ▸ Cambodia ▸ Siem Reap Province ▸ Siem Reap ▸ Things to do in Siem Reap ▸ Angkor Wat

Angkor Wat

 32,259 reviews | **#1** of 179 things to do in Siem Reap

Ancient Ruins, Sights & Landmarks |  As featured in 3 days in Siem Reap and 10 other guides

Overview


Tours & Tickets

Reviews (32,260)

Q&A (166)

Location

Travel Recommendation System




Nancy C
Level 2 Contributor
5 reviews
5 attraction reviews

"What amazing Angkor Wat" NEW

★★★★★ Reviewed yesterday

i visited Angkor Wat and it was so amazing for me and also this is my first time in Cambodia. I saw the sunset raised up in the middle of Angkor Wat Tower. I can't believe that. Pretty cool High recommend

Helpful? Thank Nancy C Report











Daniel_Biel
Biel, Bern, Switzerland
Level 3 Contributor
13 reviews
5 attraction reviews
4 helpful votes

"Beutiful and a must to see" NEW


★★★★☆ Reviewed yesterday

It is a historical nice place, but a lot of people they like to see too. Even early in the morning on the sunrise have a lot people. But the Wat is really nice and beautiful, take your minimum 4 nights there it have a lot different Places zu see around.

[See all photos](#)

Helpful? Thank Daniel_Biel Report



Blair P
Level 5 Contributor

"Must visit temples" NEW

★★★★☆ Reviewed yesterday via mobile

The tours tend to last a longggg time and are so expensive compared to everything else in Cambodia. But it's a must do as the temples are so unique

Right Job, Right Person Application



[Login](#) | [Register](#) | [Change Language](#)

May consider buying data



Popular Search: Full-time | Part-time | Internship | Urgent | New Jobs | Managerial Jobs |
NGO | Fresh Graduated | 中文招聘专区

Browse Jobs

	Function	Industry	Location	Salary
Accounting (432)	Administration (215)	Architecture/Engineering (137)		
Assistant/Secretary (115)	Audit/Taxation (27)	Bank/Insurance (170)		
Cashier/Receptionist (110)	Catering/Restaurant (74)	Consultancy (17)		
Cook/Cleaner/Maid (65)	Customer Service (195)	Design (131)		
Education/Training (154)	Finance (69)	Freight/Shipping/Delivery/War... (92)		
Hotel/Hospitality (180)	Human Resource (141)	IT (182)		
Lawyer/Legal Service (12)	Management (254)	Manufacturing (101)		
Marketing (324)	Media/Advertising (21)	Medical/Health/Nursing (36)		
Merchandising/Purchasing (59)	Operations (95)	Others (740)		
Project Management (37)	QC/QA (29)	Resort/Casino (9)		
Sales (836)	Security/Driver (102)	Technician (157)		
Telecommunication (3)	Translation/Interpretation (69)	Travel Agent/Ticket Sales (42)		

Urgent Jobs

[More >](#)

Operation Executive-Chinese Speaking Only ...

Driver(400\$)Up 司机薪水(400\$)以上 - Sino Gre...

General Manager - MING HOUR GROUP Co.,L...

Accountant - MOL LOGISTICS CAMBODIA

ISO Documents Controller - Diadent (Cambodi...

Sale Supervisor/Sale Representative - Vipar A...

Admin&HR Supervisor - Vipar Auto (Cambodia...

Sales - K8 Capital Co., Ltd.

Sales Representative PETFOOD - Neovia

Security/Driver(ម្ចាស់រថភ្លើង/សម្លាប់) - PARK...

Front Desk Receptionist 2post / 接待員2名 - P...

Cashier 1post / 會計出納一名 - PARKWAY INV...

Right Job, Right Person Application

The screenshot displays the 'Lite Version' job portal interface. At the top, a dark blue navigation bar contains the 'Lite VERSION' logo and links for 'Job List', 'Job Category', 'Company List', and 'Job Location'. On the right side of this bar are buttons for 'CLASSIFIED ADS' and 'Others'. Below the navigation bar is a large banner image featuring a cup of coffee on a saucer. Overlaid on this banner is a search section with the heading 'Search your favourite jobs'. It includes a text input field labeled 'Input your keywords or BTDC-ID', a dropdown menu for 'All Job Categories', a blue 'Search' button, and a filter icon. To the left of the main content area is a sidebar with 'Quick Links' (Part-Time Jobs, Short-Term Jobs, Training/Workshops, Invitation for Bids) and 'Job Category' (Execute. / Management with 232 jobs, Business Administration with 100 jobs). The main content area is titled 'Latest Jobs' with a count of 399. It lists three job opportunities: 1) 'Various Position' at 'Learning Jungle School' (ID 61983, 3 days old, posted 26-May-2017); 2) 'Senior Credit Officer, Junior Credit Officer, and IT Executive' at 'Microfinance Amatak Capital Plc.' (ID 61894, 8 days old, posted 21-May-2017); and 3) 'IT Officer, and Sales Officer' at 'TPT Investment Co., Ltd'.

Lite VERSION Job List Job Category Company List Job Location CLASSIFIED ADS Others

Search your favourite jobs

Input your keywords or BTDC-ID All Job Categories. Search

Quick Links

- Part-Time Jobs
- Short-Term Jobs
- Training/Workshops
- Invitation for Bids

Job Category

- Execute. / Management 232
- Business Administration 100

Latest Jobs 399

- Various Position**
Learning Jungle School
ID: 61983 | 3 days | 26-May-2017
- Senior Credit Officer, Junior Credit Officer, and IT Executive**
Microfinance Amatak Capital Plc.
ID: 61894 | 8 days | 21-May-2017
- IT Officer, and Sales Officer**
TPT Investment Co., Ltd