Content Analysis and Data Source

Nith Kosal

Future Forum

February 18, 2021

Content Analysis

- One of numerous research methods used to analyze text data.
- Text data: Primary sources vs Secondary sources

Content Analysis

- One of numerous research methods used to analyze text data.
- Text data: Primary sources vs Secondary sources

Reason of the content analysis:

- Obtain descriptive information
- Analyze observable/interview data
- Test hypothesis
- Check other research findings

Categorising and Coding

- RQ and/or hypothesis
- Define the universe of texts
- Select the texts
- Determine unit of analysis (mutually exclusive and exhaustive)
- Develop coding schedule and coding manual
- Coding data ▸ see for example
- Analyzing data (e.g. descriptive statistics and frequencies)
- Interpreting the data

IMPORTANCE OF CATEGORIZATION

- All studies must convert (code) descriptive information into categories:
  - Researcher determines categories before analysis begins.
  - Researcher becomes familiar with the descriptive information collected and allows the categories to emerge as analysis continues.

IMPORTANCE OF CATEGORIZATION

- All studies must convert (code) descriptive information into categories:
  - Researcher determines categories before analysis begins.
  - Researcher becomes familiar with the descriptive information collected and allows the categories to emerge as analysis continues.

MAJOR CODING CATEGORIES

- **Manifest:** Refers to the specific, clear, surface contents: words, pictures, images, and such that are easily categorized.
- **Latent:** Refers to the meaning underlying what is contained in a communication.
- **Both:** Combination.

Qualitative Analysis

- The materials start out qualitative
- The analysis starts out qualitative
- It can remain primarily qualitative
  - Identifying themes
  - Identifying patterns
  - Describing situations

Qualitative Analysis

- The materials start out qualitative
- The analysis starts out qualitative
- It can remain primarily qualitative
  - Identifying themes
  - Identifying patterns
  - Describing situations

Quantitative Analysis

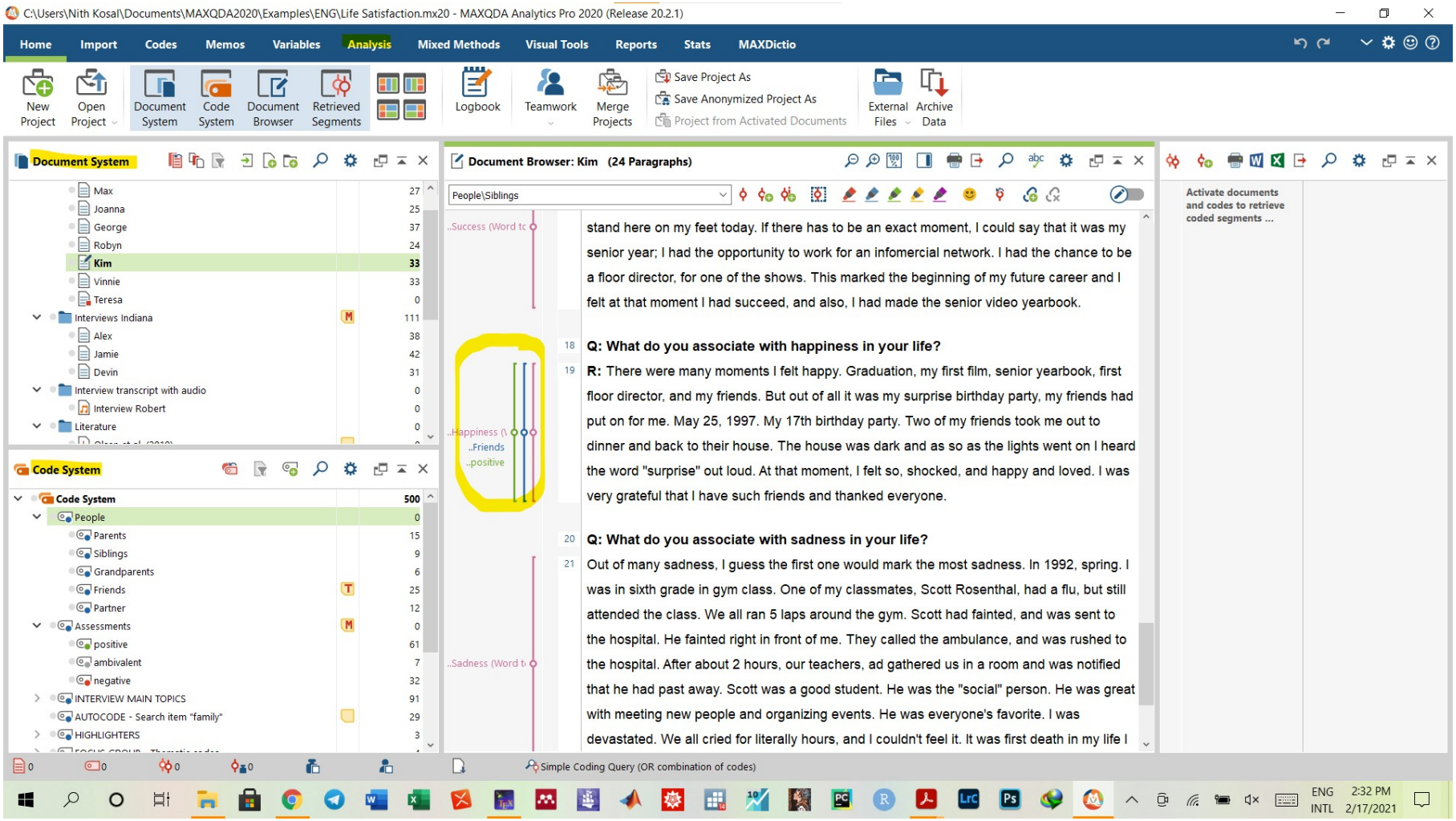
- Creating Codes to define categories
- Counting instances to see frequency
- Coding and Recoding to see the range
- Making comparisons between groups
- Presenting findings in tables and graphs
- Deepening analysis with statistics

Application Software

MAXQDA, NVivo, ATLAS.ti, Excel, Word, Stata, R, MATLAB, and others.

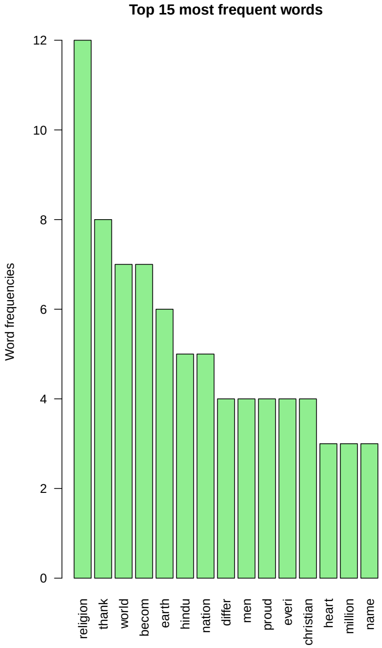
Application Software

MAXQDA, NVivo, ATLAS.ti, Excel, Word, Stata, R, MATLAB, and others.



Text Mining and Sentiment Analysis with R

```
# Install for text mining, text stemming, word-cloud generator, color palettes, sentiment analysis, and plotting graphs
install.packages("tm", "SnowballC", "wordcloud", "RColorBrewer", "syuzhet", "ggplot2")
# Load
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("syuzhet")
library("ggplot2")
URL <- "https://datascienceplus.com/wp-content/uploads/2017/10/chicago.txt"
download.file(URL, destfile = "./data/data.txt", method="curl")
```


Top 15 most frequent words
Word frequencies
0
2
4
6
8
10
12
religion
thank
world
becom
earth
hindu
nation
differ
men
proud
everi
christian
heart
million
name


religion
becom
thank
world
nation
earth
hindu
christian
everi
men
differ
proud
human
million
heart
dream
buddhist
name
toler
harmoni
belong
come
plant
destruct
parliament
will
hope
water
assimil
god
india
shelter
face
fanatic
forbid
uniti
seed
law
reach
despair
violenc
purest
produc
told
wish
now
time
grow
fill
brother
piti
given
air
path
accept
peopl
septemb
line
class
help
brethren
holi
other
exclus
honor
wonder
labor
upon
remnant
may
one
platform
possess
lead
univers
repeat
convent
growth
persecut
shown
taught
banner
horribl
roman
often
whose
ground
far
truth
monk
feel
chicago
enlighten
true
grand
evid
peac
ï»¿sister

Sources of Data

- National Institute of Statistics of Cambodia
  - Macro Indicator Data: Camstat
  - Micro Data: Nada

Sources of Data

- National Institute of Statistics of Cambodia
  - Macro Indicator Data: Camstat
  - Micro Data: Nada
- World Bank: Macro data and Micro data

Sources of Data

- National Institute of Statistics of Cambodia
  - Macro Indicator Data: Camstat
  - Micro Data: Nada
- World Bank: Macro data and Micro data
- Government official websites (MoEF, MoC, NBC), research institutes and other international NGOs (UN, UNCTAD, OECD, UNDP, ADB, IFM, ILO,).

Sources of Data

- National Institute of Statistics of Cambodia
  - Macro Indicator Data: Camstat
  - Micro Data: Nada
- World Bank: Macro data and Micro data
- Government official websites (MoEF, MoC, NBC), research institutes and other international NGOs (UN, UNCTAD, OECD, UNDP, ADB, IFM, ILO,).
- International IPUMS

Sources of Data

- National Institute of Statistics of Cambodia
  - Macro Indicator Data: Camstat
  - Micro Data: Nada
- World Bank: Macro data and Micro data
- Government official websites (MoEF, MoC, NBC), research institutes and other international NGOs (UN, UNCTAD, OECD, UNDP, ADB, IFM, ILO,).
- International IPUMS
- Demographic and Health surveys

Sources of Data

- National Institute of Statistics of Cambodia
  - Macro Indicator Data: Camstat
  - Micro Data: Nada
- World Bank: Macro data and Micro data
- Government official websites (MoEF, MoC, NBC), research institutes and other international NGOs (UN, UNCTAD, OECD, UNDP, ADB, IFM, ILO,).
- International IPUMS
- Demographic and Health surveys
- ICPSR: A data repository with many data sets on lots of subjects.

- Harvard-MIT Data center and Harvard Dataverse where many researchers archive their data.

- Harvard-MIT Data center and Harvard Dataverse where many researchers archive their data.
- The American Economic Association journals require posting of any data used for research: there is lots of data on the AEA website.

- Harvard-MIT Data center and Harvard Dataverse where many researchers archive their data.
- The American Economic Association journals require posting of any data used for research: there is lots of data on the AEA website.
- Researcher webpage, or GitHub profile.

- Harvard-MIT Data center and Harvard Dataverse where many researchers archive their data.
- The American Economic Association journals require posting of any data used for research: there is lots of data on the AEA website.
- Researcher webpage, or GitHub profile.
- Getting help and learning more, Google is still your best friend!

Any Questions

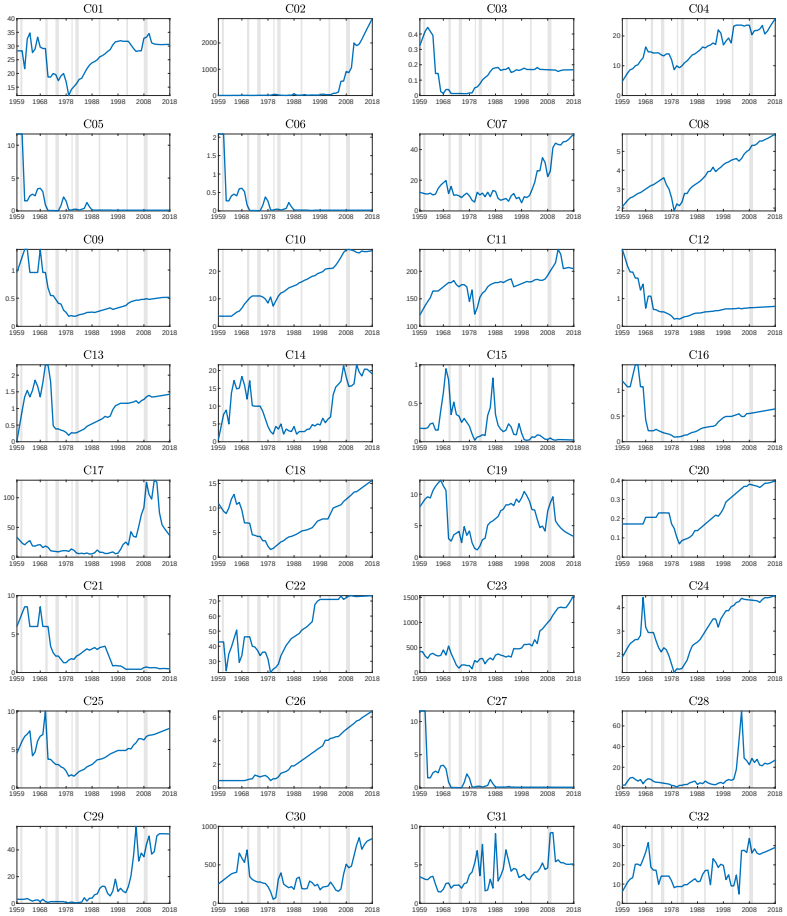
Additional Information

Table A8 continued from previous page

Code	Commodity	1959-1988 (Million US$)				1989-2018 (Million US$)				1959-2018 (Million US$)			
		Min	Mean	Max	St. Dev.	Min	Mean	Max	St. Dev.	Min	Mean	Max	St. Dev.
C23	Paddy rice	74.74	279.89	529.88	110.94	308.56	798.72	1535.96	396.98	74.74	539.31	1535.96	390.19
C24	Pepper and piper spp	1.23	2.30	4.44	0.65	2.64	3.93	4.52	0.55	1.23	3.12	4.52	1.01
C25	Pineapples	1.52	4.07	9.98	2.12	3.35	5.61	7.75	1.35	1.52	4.84	9.98	1.94
C26	Roots and tubers	0.62	0.91	1.86	0.36	2.02	4.29	6.51	1.34	0.62	2.60	6.51	1.95
C27	Seed cotton	0.04	2.12	11.59	3.32	0.10	0.11	0.22	0.02	0.04	1.12	11.59	2.55
C28	Sesame seed	0.83	5.17	10.15	2.43	3.14	17.97	74.41	15.17	0.83	11.57	74.41	12.61
C29	Soybeans	0.32	1.90	4.18	1.14	5.55	27.95	57.63	18.42	0.32	14.93	57.63	18.44
C30	Sugarcane	52.87	316.17	693.89	145.89	156.18	404.02	853.12	241.26	52.87	360.09	853.12	204.14
C31	Sweet potatoes	1.48	3.24	9.05	1.77	2.90	4.82	9.20	1.46	1.48	4.03	9.20	1.80
C32	Unmanufactured tobacco	6.13	14.16	31.59	5.79	4.80	20.84	33.73	7.45	4.80	17.50	33.73	7.46
L01	Buffalo meat	1.15	16.80	30.70	11.43	24.91	32.28	38.84	4.05	1.15	24.54	38.84	11.55
L02	Buffalo meat (indigenous)	4.28	19.59	31.39	8.74	28.85	33.26	38.84	2.95	4.28	26.42	38.84	9.44
L03	Cattle meat	8.39	16.42	25.82	4.67	29.10	55.05	70.65	12.98	8.39	35.73	70.65	21.63
L04	Cattle meat (indigenous)	8.39	16.63	25.84	5.16	29.10	56.55	70.65	13.92	8.39	36.59	70.65	22.55
L05	Chicken meat	7.82	18.49	42.72	7.76	37.65	54.05	70.39	6.97	7.82	36.27	70.39	19.25
L06	Chicken meat (indigenous)	7.82	18.49	42.72	7.76	37.65	53.61	70.39	6.75	7.82	36.05	70.39	19.01
L07	Duck meat	0.91	4.02	7.42	1.68	7.54	13.92	18.41	3.85	0.91	8.97	18.41	5.77
L08	Duck meat (indigenous)	0.91	4.02	7.42	1.68	7.54	13.80	17.68	3.72	0.91	8.91	17.68	5.68
L09	Hen eggs in shell	3.33	5.37	7.87	1.28	8.48	13.49	20.19	3.29	3.33	9.43	20.19	4.76
L10	Other bird eggs in shell	0.55	2.31	3.15	0.83	3.18	4.35	5.31	0.68	0.55	3.33	5.31	1.27
L11	Pig meat	13.05	65.39	149.09	37.67	150.58	253.88	344.16	46.66	13.05	159.63	344.16	103.34
L12	Pig meat (indigenous)	13.05	67.71	149.09	35.33	150.58	235.71	341.69	45.44	13.05	151.71	341.69	93.34
L13	Silkworm coiling cocoons	0.14	0.29	0.34	0.07	0.21	0.33	0.47	0.08	0.14	0.31	0.47	0.08
L14	Whole fresh cow milk	0.59	0.86	1.10	0.13	0.91	8.74	10.75	2.26	0.59	4.80	10.75	4.25

D Additional Graphics

Figure 3: Gross Crop Production in Cambodia Between 1959–2018



Note: This figure presents the gross agricultural production of the crop product in million of US dollars at constant prices in 2004–2006. The author used the FAOSTAT dataset between 1961–2018 on crop production. From 1959–1960 and other missing values in the dataset are used linear interpolation method to predict. The 32 crop products are represented by the identification code from C01–C32. The definition of each crop identification code is shown in the Table A8. According to the dataset, it shows that cassava, dry chilies and peppers, grapefruit and pommelos, green coffee, lemons and limes, mangoes, nuts, paddy rice, pineapples, roots and tubes are rapidly increasing production and incomes.

	A	B	C	D	E	F	G
1	Categorical	Variable	Dependent Variables				
2	Source	Author	FAOSTAT	FAOSTAT	FAOSTAT	FAOSTAT	FAOSTAT
3	Unit	Annually	Millions US	Millions US	Millions US	Millions US	Millions US
4	Descriptio	Time serie	Value of g	Value of g	Value of g	Value of g	Value of g
5	Variable N	Year	Bananas	Cassava	Castor oil	Coconuts	Cotton lin
6	ID Code	Year	C01	C02	C03	C04	C05
7	1	1989	24.3356	15.677	0.18238	15.314	0.12771
8	2	1990	24.9874	14.9305	0.16284	16.3127	0.12771
9	3	1991	26.0738	13.9351	0.16935	15.9798	0.12771
10	4	1992	26.5084	37.3263	0.16973	16.6457	0.12771
11	5	1993	27.1602	12.7636	0.18238	16.9786	0.12133
12	6	1994	28.0294	16.1747	0.14981	17.6444	0.12771
13	7	1995	28.6812	20.3926	0.15632	17.145	0.12771
14	8	1996	30.4195	17.3333	0.16752	22.0555	0.106
15	9	1997	31.5059	19.228	0.16284	20.9735	0.106
16	10	1998	31.7231	16.5564	0.16935	16.9786	0.10856
17	11	1999	31.9404	56.8634	0.17782	18.3102	0.11111
18	12	2000	31.7231	36.7696	0.16935	19.309	0.11111
19	13	2001	31.7231	35.4008	0.16935	17.6384	0.11111
20	14	2002	31.7231	30.3622	0.1686	23.3039	0.11111
21	15	2003	30.4195	82.2793	0.16941	23.6368	0.11111
22	16	2004	29.1158	90.0932	0.18238	23.6368	0.11111
23	17	2005	28.0294	133.285	0.17	23.6368	0.11111
24	18	2006	28.2466	542.984	0.16963	23.3039	0.11111

Data | Production | LPro | MonthlyPrices