Public Cloud Computing

IQSS WORKSHOP

Outline

- Motivations
- Methodology
- Contents
- Experiment: "*Presidential Commercial Campaigns Dataset*"
- Next steps
- Future work

Motivations for Public Cloud

- Embed advance computing services
- Shift from one to another service easily
- Not provided / expensive to provide in house
- Hot topic / many investments

Overall cheaper and easier than learning and implementing your own algorithm and infrastructure

Methodology

- How social scientist can benefits from cloud?

- Is it for data collection?

- Is it for data analysis?

- Is it for exposure to other fields advancements?

Iterative process: question > data > analysis > update question

Cloud in nutshell:

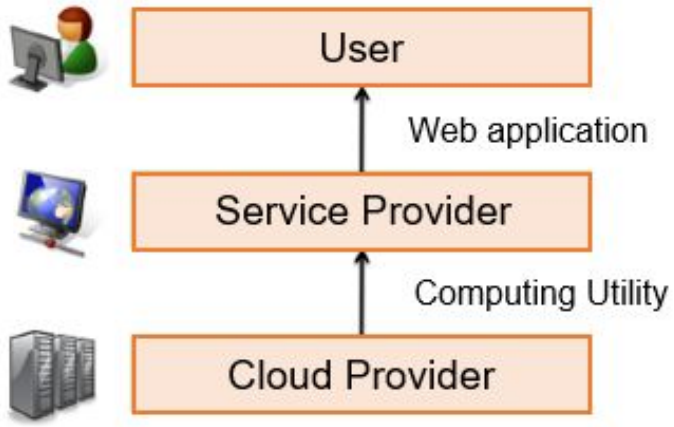


Image adapted from Armbrust et al, 2009. Above the Clouds: A Berkeley View of Cloud Computing

Why Turning to Public Cloud?

1) Eliminates/reduces costs

a) buying hardware and software
b) electricity for power and cooling
c) IT experts for managing the infrastructure

2) Vast amounts of computing resources

3) Provides right amount of IT resources

a) computing power
b) storage
c) bandwidth
d) geographic location

4) Removes the need for on-site datacenters

5) Data Centers are regularly upgraded

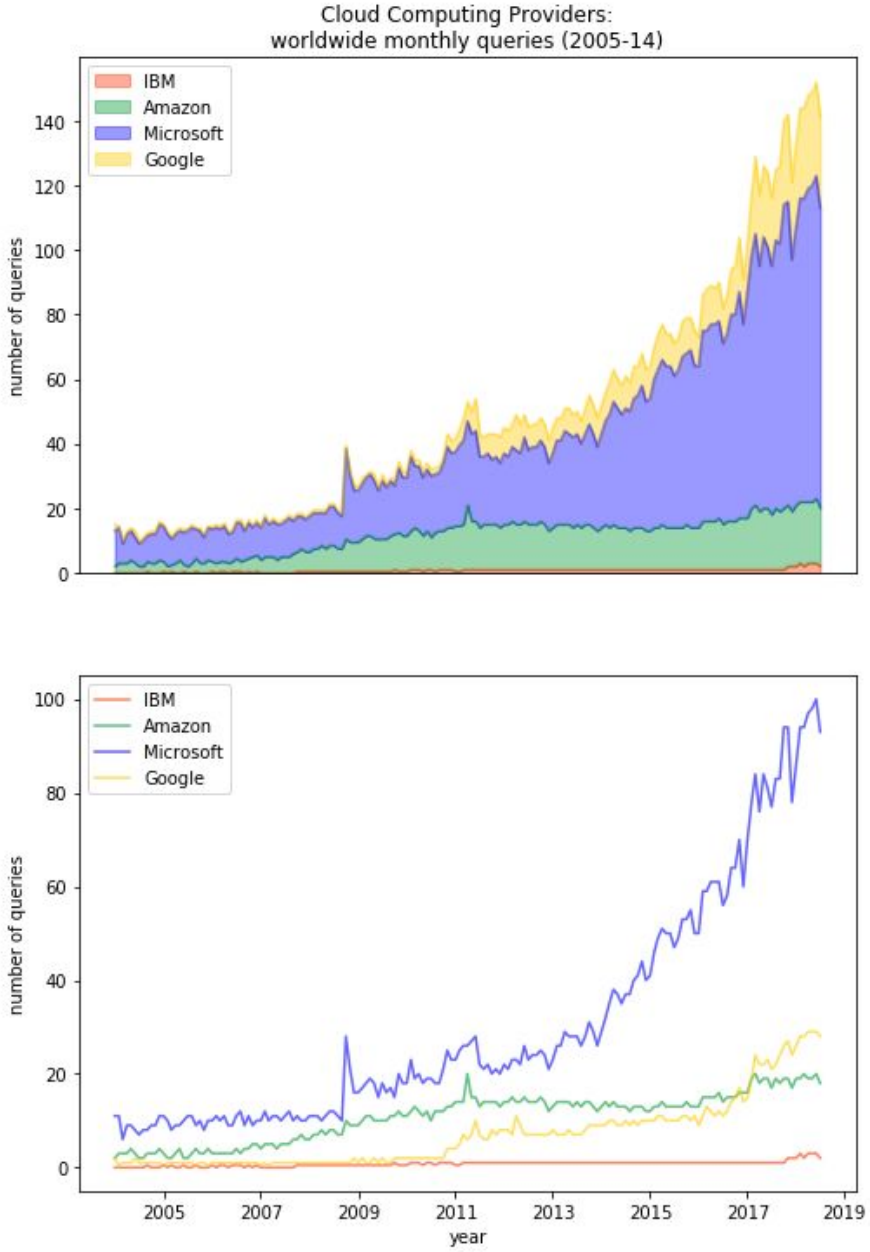
6) Data can be mirrored at multiple redundant sites

Cloud in nutshell:

Disadvantages	Advantages
Security in the Cloud	Almost Unlimited Storage
Technical Issues	Quick Development
Non-Interoperability	Cost Efficient
Dependency and vendor lock-in	Automatic Software Integration
Internet Required	Shared Resources
Less Reliability	Easy Access to Information
Less management	Mobility
Raised Vulnerability	Better Hardware Management
Prone to Attack	Backup and Recovery

Image adapted from Naeem et al, 2016.
Cluster Computing vs Cloud Computing: a comparison and overview

Cloud in nutshell:



Data from Google Trends: providers name and cloud computing words associated (elaborated by the authors - Accessed July 2018)

Cloud in nutshell:

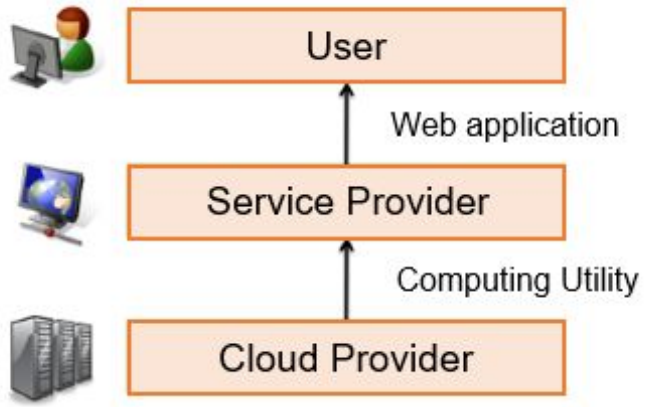
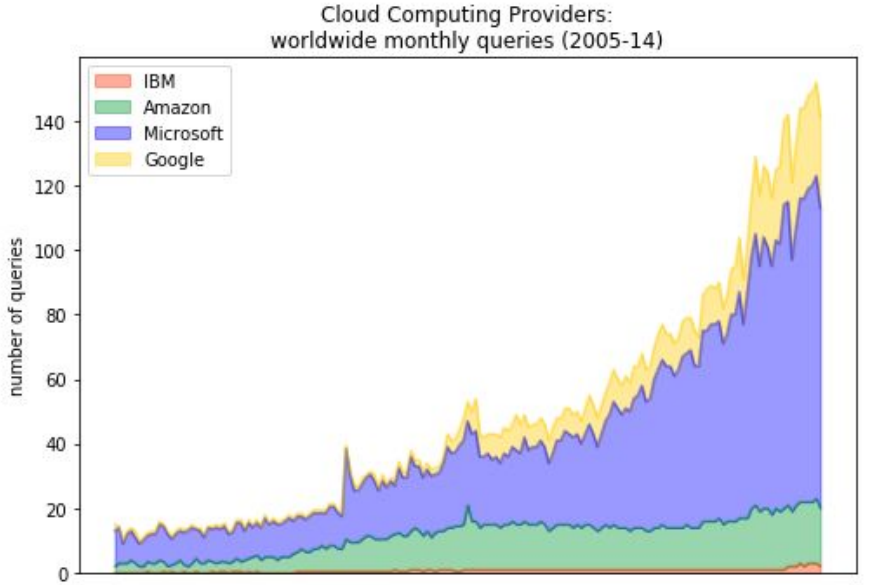
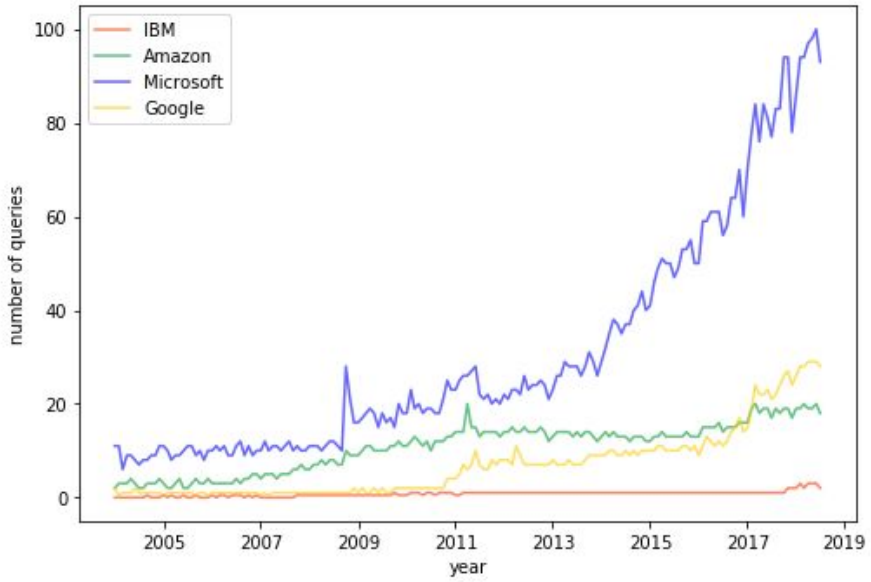


Image adapted from Armbrust et al, 2009. Above the Clouds: A Berkeley View of Cloud Computing

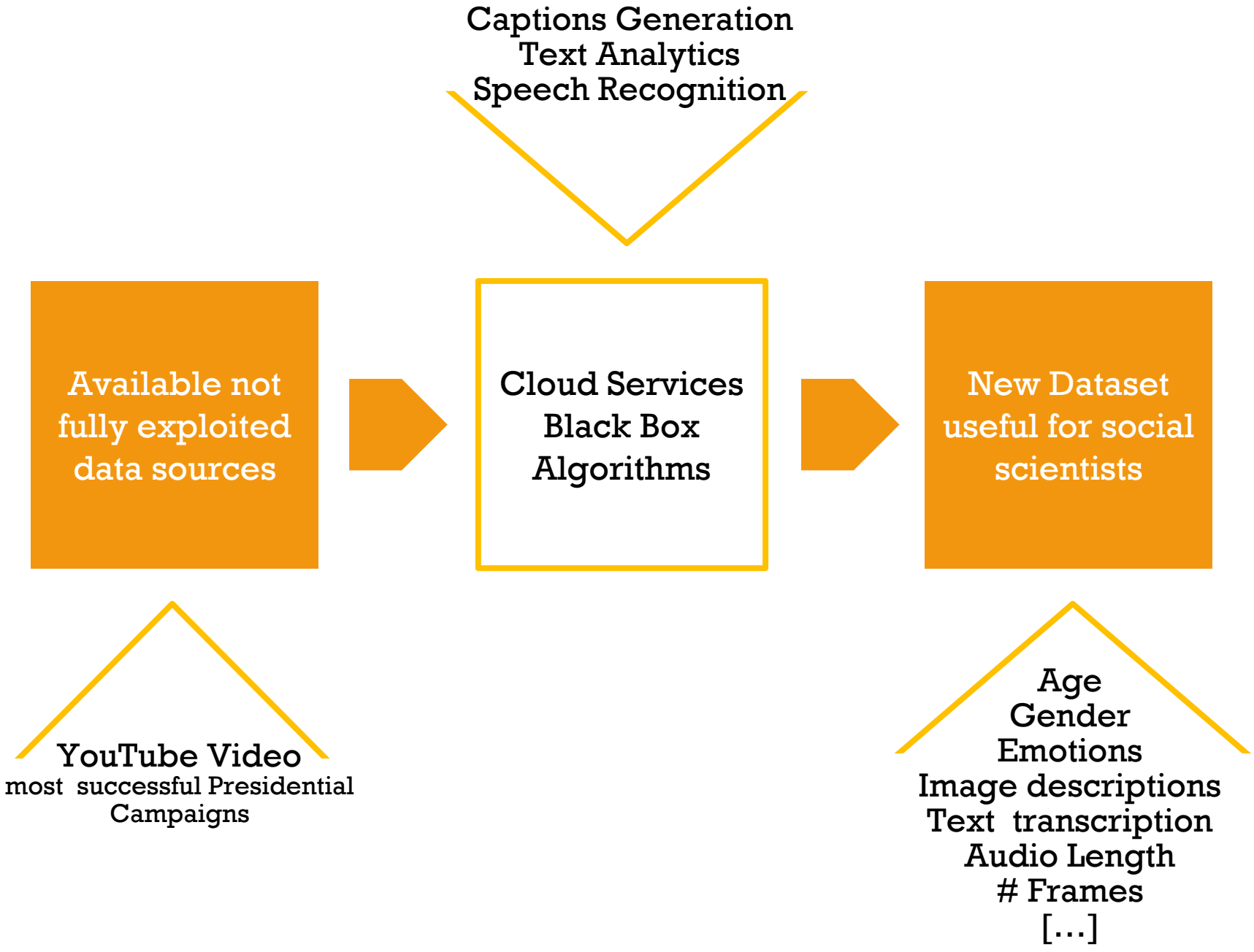
Disadvantages	Advantages
Security in the Cloud	Almost Unlimited Storage
Technical Issues	Quick Development
Non-Interoperability	Cost Efficient
Dependency and vendor lock-in	Automatic Software Integration
Internet Required	Shared Resources
Less Reliability	Easy Access to Information
Less management	Mobility
Raised Vulnerability	Better Hardware Management
Prone to Attack	Backup and Recovery

Image adapted from Naeem et al, 2016. Cluster Computing vs Cloud Computing: a comparison and overview





Data from Google Trends: providers name and cloud computing words associated (elaborated by the authors - Accessed July 2018)


Captions Generation
Text Analytics
Speech Recognition
Available not fully exploited data sources
Cloud Services Black Box Algorithms
New Dataset useful for social scientists
YouTube Video
most successful Presidential Campaigns
Age
Gender
Emotions
Image descriptions
Text transcription
Audio Length
Frames
[…]

Experiment:

Experiment: Phase I

Phase I: Data Collection

Available not fully exploited data sources

YouTube Video
most successful Presidential Campaigns

Captions Generation
Text Analytics
Speech Recognition

Cloud Services
Black Box
Algorithms

New Dataset useful for social scientists

Age
Gender
Emotions
Image descriptions
Text transcription
Audio Length
Frames
[…]

Phase II: Features Extraction

Experiment: Phase II

Phase I: Data Collection

Phase II: Features Extraction

Available not fully exploited data sources

YouTube Video
most successful Presidential Campaigns

Cloud Services
Black Box Algorithms

Captions Generation
Text Analytics
Speech Recognition

New Dataset useful for social scientists

Age
Gender
Emotions
Image descriptions
Text transcription
Audio Length
Frames
[…]

Contents

Survey

Guides

Experiment

IQSS Workshop Materials

About us Survey GitHub repository

Public Cloud Computing

Cloud Services for Social Scientists

This tutorial is an endeavor by IQSS to expand training opportunities for social scientists. It shows how to use some of the most interesting public cloud services available and it provides justifications for their use by social scientists.

by Andrea Porelli and Ista Zahn

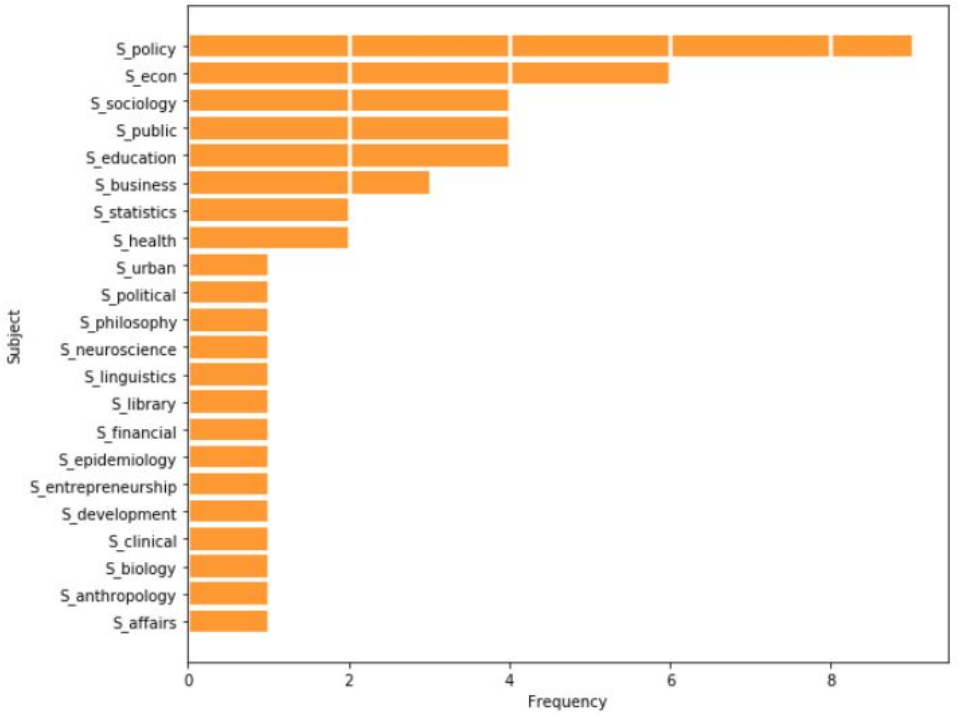
Contents

Links to guides and experiment

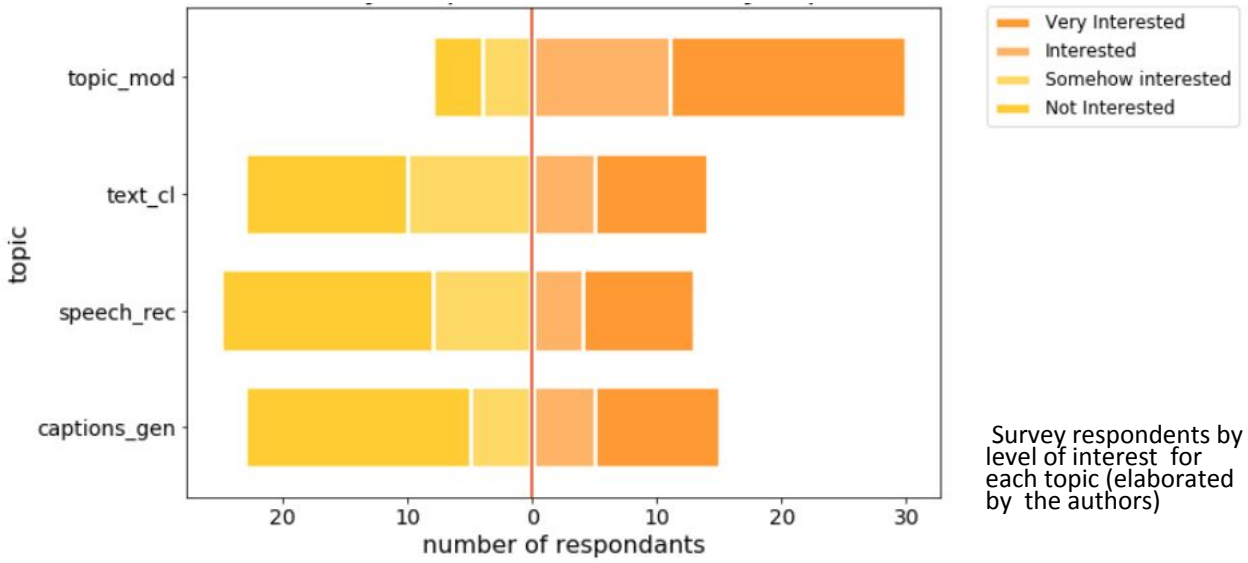
Web site:

Link to website

Contents: Survey

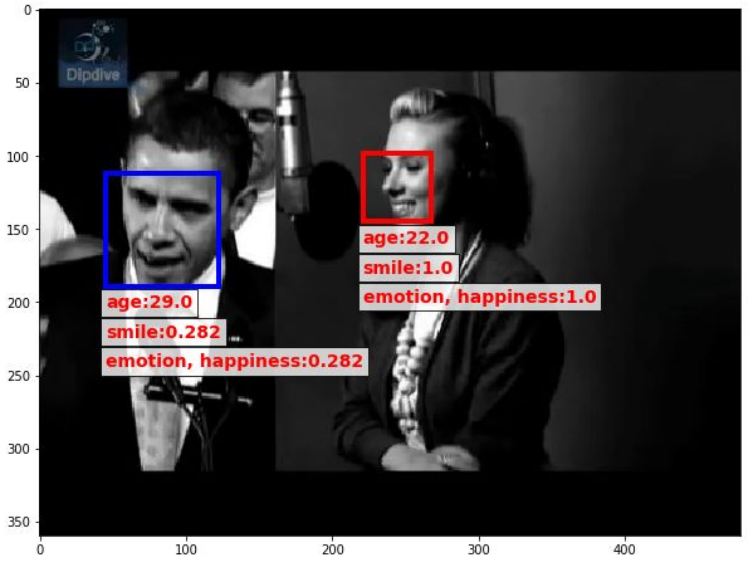


Survey respondents by field of study
(elaborated by the authors)



Survey respondents by level of interest for each topic (elaborated by the authors)

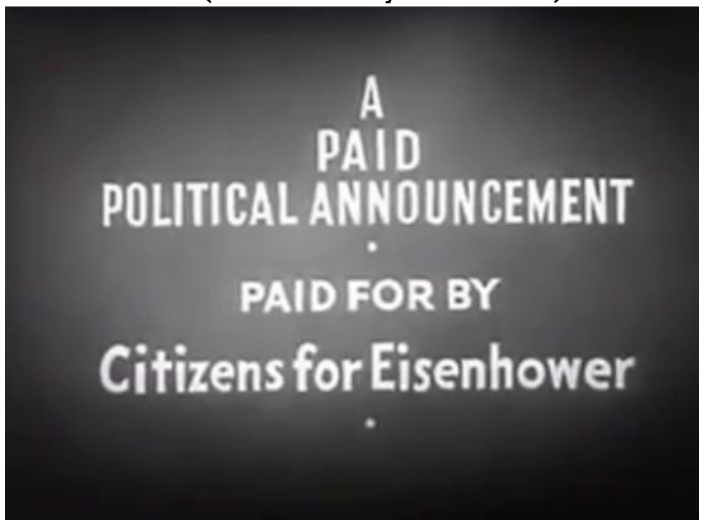
Contents: Guides



FACE API service applied to Yes We Can campaign in support of Barack Obama in 2008 (elaborated by the authors)

"i was born a little town called hope arkansas in the grandparents very limited income it was e boys nation program an IRA member just think thing would be given the opportunity to meet t work my way through law school with part time lot of money i just wanna go home and see if i and we real progress now it's exhilarating tom tter and bring hope back to the american dream

SPEECH RECOGNITION API service applied to Hope campaign in support of Bill Clinton in 1992 (elaborated by the authors)




```
Optical Character Recognition Summary Results:
-----------------------------------------------------------------
The words that appear in the image are:
 ['PAID', 'POLITICALANNOUNCEMENT', 'PAID', 'FOR', 'BY', 'Citizens',
  'for', 'Eisenhower']
```

OCRAPI service applied to Eisenhower For President campaign in 1952 (elaborated by the authors)

```
Analyze Image Summary Results:
-----------------------------------
The image is within the building category
The image is a large white building
```

ANALYZE IMAGE API service applied to It's Morning in America Again in support of Ronald Reagan campaign in 1988 (elaborated by the authors)

Contents: Experiment

video_title	video_length[sec]	video_frames[n]	frame_sec[n/sec]	frame_extracted[n]	year
eisenhower_for_president_1952	62.09	1859	29.940409	19	1952
kennedy_for_me_campaign_jingle_jfk_1960	60.23	1789	29.702806	18	1960
high_quality_famous_daisy_attack_ad_from_1964_...	66.90	2003	29.940209	20	1964
humphrey_laughing_at_spiro_agnew_1968_politica...	19.25	575	29.870130	6	1968
mcgovern_defense_plan_ad_nixon_1972_presidenti...	60.05	1798	29.941715	18	1972
ronald_reagan_tv_ad_its_morning_in_america_again	59.95	1793	29.908257	18	1984
1988_george_bush_sr_revolving_door_attack_ad_c...	29.88	894	29.919679	9	1988
bill_clinton_hope_ad_1992	60.26	1802	29.903750	18	1992
historical_campaign_ad_windsurfing_bushcheney_04	30.09	900	29.910269	9	2004
yes_we_can__barack_obama_music_video	270.21	3781	13.992820	38	2008

Extract from Data Collection tutorial: video description (elaborated by the authors)

title	stt_text	stt_words_count	sentiment_text	key_phrases	sentiment_key_phrases
1964	Nan play hello by standing ben Nan please are ...	50	0.856453	[stakes, god, president johnson, s children, v...	0.762500
barack_obama	Nan how Nan Nan how what Nan what people of th...	33	0.243644	[people, nation false hope]	0.815000
bill_clinton	i was born a little town called hope arkansas ...	173	0.5	[cause i, present i, graduated i, lot of money...	0.571613
bushcheney	i'm george W bush and i approve this message i...	78	0.5	[john kerry lead carey, john kerry whichever w...	0.677273
eisenhower	i for president for president i like my comput...	31	0.918349	[president i, computer billy graham washingto...	0.700000
george_bush_sr	who is governor michael dukakis vitov mandator...	63	0.0557907	[massachusetts america, governor michael dukak...	0.529167
humphrey	Nan Nan	2	Nan	Nan	0.400000
jfk	Nan do you wanna man for president who season ...	17	0.5	[wanna man, president]	0.380000
nixon	the mcgovern defense plan he would cut the mar...	84	0.5	[cut navy personnel, navy fleet, mcgovern defe...	0.602667
ronald_reagan	it's morning again in america today more men a...	112	0.5	[short years, half, young men, rates, leadersh...	0.590588

Extract from Features with Text Analytics API: audio to text and text analysis (elaborated by the author)

Dataset: *"Presidential Commercial Campaigns Ads"*

Data type and data structure list

field	data type	data description
index	**string**	presidential campaign title title
video_url	**string**	YouTube video url
video_length[sec]	**float**	YouTube video lenght
video_frames[n]	**integer**	YouTube video number of frames
frame_sec[n/sec]	**float**	ratio number of frames in the video and video lenght
frame_extracted[n]	**integer**	number of frames extracted by the orignal video (i.e. 100th frame and multiples)
year	**integer**	presidential campaign year
candidate_name	**string**	presidential candidate name
party	**string**	candidate party
stt_text	**string**	text trascription of the video (i.e. script)
stt_words_count	**integer**	number of words in the script
ta_sentiment_text	**float**	scores close to 1 indicate positive sentiment in the script of the video, while scores close to 0 indicate negative sentiment
ta_key_phrases	**string**	list of strings denoting the key talking points in the script of the video
ta_sentiment_key_phrases	**float**	scores close to 1 indicate positive sentiment in the key talking point, while scores close to 0 indicate negative sentiment
caption	**string**	sentence describing image contents
caption_score	**float**	probability of the captions being accurate
tag	**string**	key points in the image
tag_description	**string**	list of subjects the image
tag_score	**float**	probability of the tag being accurate
people_gender	**string**	gender of people in the image
people_age	**integer**	age of people in the image
people_celebrities	**string**	name of celebrities in the image
image_text	**string**	character recognition in the image
category	**string**	main subject in the image
category_score	**float**	probability of the category being accurate

iqss_workshop\workshops\public_cloud_computing\data\dataset\presidential_campaigns_ads_dataset.csv

Next Steps

- Workshop
- Evaluation
- Close the feedbacks loop
- Invite Microsoft to conduct Cloud training for staff
- Advocating to increase funding on public cloud services in the university
  - Bottlenecks: not fully documented and not easy to access

Future Work

- Enlarge Dataset using Museum of Presidential Campaign (Link)
- Apply for funds to expand workshops in public cloud computing (Microsoft Azure)
- Add *Presidential Campaigns Ads Dataset* to DataVerse
- Advance modelling using Azure ML Studio