A framework for collection of contextual data for AI Assistants

Anantha Chickanayakanahalli, IDC, IIT Bombay

Assistive technologies need to respond to our contexts to be effective while making our lives easy.

AI assistants require a huge database to be more natural in their responses. Not just in using better language but also understanding a human being's context(s). This is essential for the assistant to really make any difference to us. Creating this database is a mammoth of a task. If a computer program has to comprehend a person's context and using this it shall take decisions, there must be a novel way to define these contexts we have in our lives.

Reference Image showing OCR [2]

This proposal aims to investigate if written content or printed characters in the world around us (like home) is sufficient to inform the spatial context for an ai assistant. By analysing and using those keywords or phrases available to be read off of our environment definitions of various contexts can be constructed. A database of such pointers can be created to help a computer program like an AI assistant respond to us in context (spatial).

The Background

Consider a scenario where a designer wants to build an AI assistant which is voice based. He or she has access to this particular type of database that helps a computer program understand human contexts. This AI assistant can be useful for example in a kitchen to help with anything one may want to cook or find in the kitchen. The user of the kitchen may want assistance with a recipe for cooking a certain dish. Or it may be just be a step by step guide to making good coffee. We want the system to not only hear and respond to the user in natural language but also want it to understand the context when user asks questions like 'can you tell me where the salt is?'.

A framework for collection of contextual data for AI Assistants

It is next to impossible making a computer program understand our ever changing contexts and would probably be something that quantum computing may enable for us soon in the years to come. Data of the context has to be more objective for a computer program to be able to make use of it. The subjective part of the data will not be useful for machines.

User Experience Design is concerned with designing for a user and their needs and to create products that are easy for them to use. In user experience design, to avoid designer biases and to make sure all the needs of user are met, there is extensive study investigating the total context of the user. There are journey maps created for understanding the steps a user goes through for establishing a goal. A well positioned design may reduce a lot of pain points being faced by the user. The study for investigating the context is called **Contextual Inquiry** and it includes the following.

1. Talking to user the about how they go about their day.
2. List of all objects used in their tasks.
3. All the analog and digital touch-points where information is exchanged by for user's daily tasks.
4. All the pain-points faced by the user faces before, while getting their work done and after
5. All hand-offs happening if there are more than one person involved in their task(s) and the responsibility of those additional actors.

Contextual data can be collected by observing, investigating a situation, a scene or a subject. To understand what constitutes a context in a human life, one may avoid the exhaustive questionnaires, interviews and coming back with collected information that needs to be manually crunched.

On the look out for shortcuts in contextual inquiry methods for getting data that illustrates a context of a user, the thought came of finding **what is the next best thing of being in a user's shoes?** It is to be the user themselves or rather augmenting the user with a second pair of eyes and ears. The intention is to place a multi-sensory camera (having a microphone) on individuals as they go about their day. There could be many ways a camera and other sensors can be augmented on a human being. It could be a head-mounted Go-Pro helmet or Google Glass.

https://commons.wikimedia.org/wiki/File:Stephen_Balaban_wearing_Google_Glass_in_2013.jpg

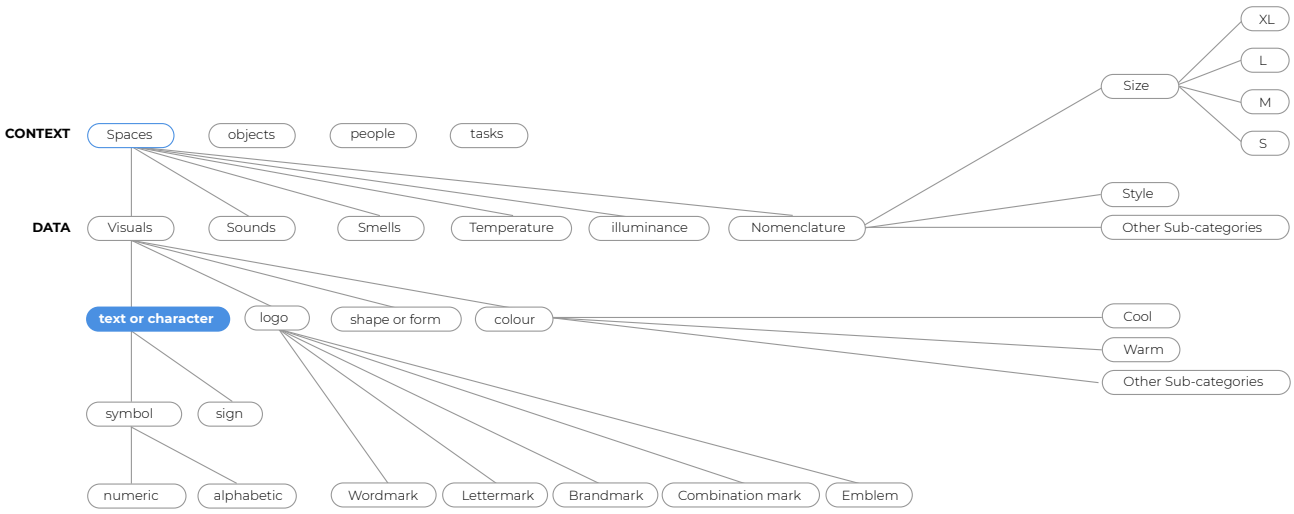
A framework for collection of contextual data for AI Assistants

Capturing all the objects and information appearing in a person's public or social life as they go about their chores, in their home or going to work, etc (without their identity coming into picture) **would give us the contexts they deal with and are a part of.**

One can collect such a data bank of objects, things, furniture, names, brands, signs, sounds, conversations, etc. (everything that is present in a person's environment and the location where each of it is captured from). An analysis would possibly give us a peek into what distinctive features constitute a context for a persona. This could be a useful set of data points not only for anthropologists, sociologists but also user experience designers. It will vary from place to place, and be unique with language, with different cultures. If one collects contextual data from a place like a kitchen, at various times of the day, visually we can expect to identify the space of the kitchen, the fittings, the counters, chimney, utensils, power supply, food material, furnitures, etc. Also one will collect the information of sounds that can be heard in the kitchen at various times of the day. Now this data collected will be analysed for consistent occurrence of visual and auditory samples. The source of sounds could be from water, cutting board, utensils, etc. There may be speech samples relevant to the place or otherwise. There may be readable textual information on the refrigerator or on the storage units, bottles, etc. It should be possible to collect data that is highly relevant only to kitchens. This outcome of analysis can be useful for finding **the invariant set** and obligatory features that defines what the context of a kitchen likely is.

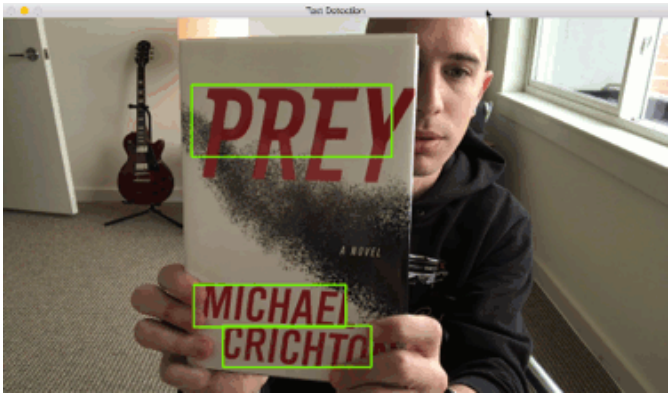
The method

A subset of the above problem area could be taken. **A single dimension of the data, for example, just the text or readable characters may be considered.**



The experiment can be to search for objects with written and printed material in our surroundings, be it in our homes or outside. Each place has cues or information as to which place it may belong to. If we scan for those written materials in our environment, it should be possible to make a program identify the context of the space with these pointers. The crux of it is to establish the relationship of keywords appearing in the

environment and connect them to the context. For the ease of designing the experiment one can ignore public spaces for now. Collecting a variety of data to be mined while keeping some experiment parameters constant like a location or a type of space like a living room or a kitchen.



Reference Image showing OCR [2]

An example

Take the context of a living room. A television sits in the living room, its probably got a brand and model name printed on it, probably has its size tagged on it as well. Xbox will have XYAB buttons. A tv remote control unit has its channels numbers printed on buttons that can be read. Similarly in a kitchen, a mixer will have keywords ‘ON’, ‘OFF’, on it. To make a machine recognise a kitchen we need a large set of objective data. The context of kitchens are usually varied. They are different due to food habits, geography, weather, family structure and culture.

A framework for collection of contextual data for AI Assistants

In this room, one can find the following keywords or phrases from looking around.

room = {'**MacBook Pro**','**M**l','**Regza**','**647978**', '**OFFER**', '**X**', '**Y**', '**A**', '**B**', '**Stop watching tv and do something**', '**Veena Musicals**', '**TIME**'}

What can be taken out from the above set is the data points of "laptop" and 'tv'. So the research will have a testable hypothesis like - for this space to be identified as a "room" it should have objects like tv and laptop. The rest of the keywords may be meaningless by themselves. By getting this set for multiple rooms/spaces, the invariant data point(s) in the sets can be pulled out. Not all rooms will have tv. But most of them may have laptop names that will most likely occur in the rooms studied. The experiment may give some keywords that cannot be used by a computer program to look at and identify a place. The data collected will provide a platform for mining various relationships, also revealing if this approach is sufficient for defining a context. These findings when put together may help in enabling assistive technologies to be responsive to our context.

Null Hypothesis

The words that can be read off any space by an AI assistant (eg. Alexa, Siri) is not sufficient to identify the context of the space it is located in.

The Experiment

Approached youth population (aged 15 - 34 years) for a panorama picture of their living space (could be a room, a hall, etc) who live in Bangalore India. Also collected the written content that is readable by a camera in their living spaces. The 18 data samples collected are below. User data includes demography of Karnataka, Maharashtra, Andra Pradesh, Telangana, Tamil Nadu, West Bengal, Uttar Pradesh, Kerala.

CG Crompton greaves
Speed
Swing
Thomson
Philips
Europa
ISI
Netflix
Source
Mute
I+
Tatasky
Tv
LIC
Ph

A framework for collection of contextual data for AI Assistants

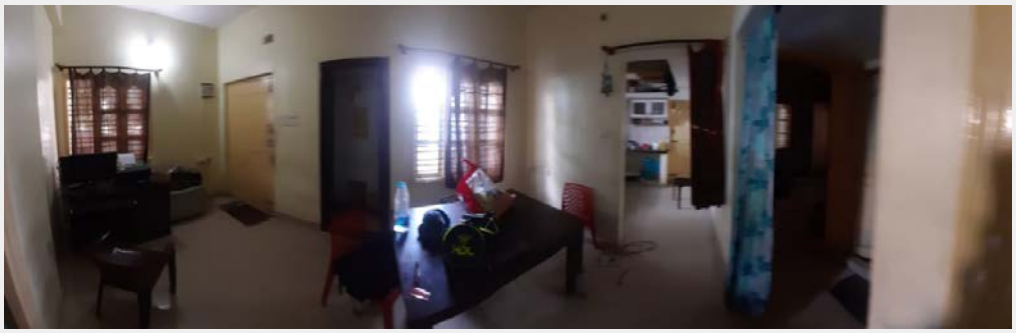
UPS
Dell
Classmate
Ben 10
Krishna
Helmet
Bat
Infinity
LG TV
Tintin
DH
Trophy
Linc

Book chor
Lock the box
Apple
Sonata
Castrol
Steelbird
Vega
Dell
Usha
HP
ARCHIES
MOSERBEAR
LED
GOODS AND SERVICE TAX
SEARCH
FINANCIAL YEAR
ZOOP
D LINK
SONY
0,1,2,3
ATTRACTIVE
DESIGN
LIFT
HEIGHT
GOLD
RICE

Oreva (watch)
&PrivéHD
The Perks of Being. A Wallflower
Basic (mom's tshirt)
tp-link
white owl
sony
Toronto

LG
Tata Sky
July 2020
Hand Sanitizer
Certificate
Frontech
S1mk2 dynamic speaker
Elle 18
Tide
Tupperware
Om Sai Ram

A framework for collection of contextual data for AI Assistants



Acer
Hp
Cremaffin
Gati
Nps
Havells
Nandi aqua
Hdl
Wingreens
Kalajyothi

Dear data
Leading
Notes from a defeatist
Journalism
Retrospective of comics
Images
Palestine
Cyanide n happiness
The frat economist
Pepsi
Train from Pakistan

Sony Blue ray Disk, samsung
Crompton Greaves
Quartz
Mortein
Tata sky
Chandana Live Calender
Keys
Lock
Flowers
Bata
Lisha
Solarizer
Pegeon

Samsung
Colors
ಯುವ ಭಾರತ
CONA
Sine wave UPS
LCD Display
AC IN = 235V
1
0
POWER
VOL
CH
+
-
10
LG
SMART
INVERTER
COMPRESSOR
12345
POWER SAVINGS GUIDE
235 L
LG

Kasturi
Den
ICIP-UVCE-2011
Quartz
Nice
Roman no. On watch
11360
Super line
Magnum
Off 1 2 3 4
New computerized design
ICERECT
Vk
Asus
Mom

A framework for collection of contextual data for AI Assistants

Sony
Monday
Tuesday....
12345..
Marcos
Ok
V+ v-
Vijayavani
Happy

Propel
Spiral pad
The bangalore press
Kempegowda
Siti digital
Best wishes
Graduation
No signal
ICA
Jackpot
15:49
777VR1
5400B

Sony
Ajantha
Tata sky
Run
V.Guard
June
3m
Europa

Switch board: Schneider Electric
20A
Honeywell
Daikin
Inverter
Ac on/off
Temp
Home leave
Power ful
Superfast charger
Sony
Bravia
Tata sky live TV
Autoline open/close curtains
Himalaya sanitizer
Tirumala calender

"IF YOU DONT LET US DREAM, WE WON'T LET YOU S;EEP!" - V For Vendetta
HAPPY BIRTHDAY SANDY
JBL
DON NORMAN The DESIGN of EVERDAY THINGS
WITH MULTISTAGE AIR PURIFICATION
i-PURE TECHNOLOGY
BACTERIA FILTER
SMELL FILTER
PM2.5 WASH FILTER
DUST FILTER
ALLERGY FILTER TIO2
SOMETIMES I THINK THE SUREST THAT INTELLIGENT LIFE EXISTS ELSEWHERE IN THE UNIVERSE IS THAT NONE OF IT HAS TRIED TO CONTACT US
REALITY CONTINUES TO RUIN MY LIFE
2020 FEBRUARY 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

A framework for collection of contextual data for AI Assistants

SAMSUNG
INTEX
AJANTA
LIC
TATA SKY
BEETLE
GM
BESCOM
ROCK LIGHT
HD
Udaya HD

MEE MEE
HARMAN KARDON
SONY
abcdefghijklmnopqrstuvwxyzx
V
K
van

The words collected have been categorised below:-

names	𝍸 𝍸 𝍸 II
devices	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 II
device settings	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 IIII
numbers	II
books	𝍸 𝍸 I
calendar	𝍸 𝍸 I
tv content	𝍸 II
certificates	I
photoframes	𝍸 I
magnet or stickers	𝍸 I
tabloid	𝍸
packaging	III
switch	𝍸
trophy	II
toiletry	𝍸
cutlery	I
clock	𝍸 III
bills	II
misc	𝍸 𝍸 𝍸 𝍸 𝍸 IIII
stationery	I
beverage	I
toys	I

This is still empirical data. We can see that the highest category of words that is readable in the living environments are device names or identifiers. It could be brand names of devices.

MacBook Pro
MI
Regza
SAMSUNG
Sine wave UPS
LG
SMART INVERTER
UPS
Dell
Tata Sky
Frontech
HP
1+
SONY
Diakin
Europa
Moserbaer
tp-link
JBL
HARMAN KARDON
Acer
Havells
Honeywell

And the next bigger list is as follows which includes the terms found on devices for controlling their settings.

X
Y
A
B
AC IN = 235V
1
0
+
-
V+
V-
CH
POWER
OFF 1 2 3 4
Ac on/off
20A
0, 1, 2, 3

Consider the same thing done by an AI Assistant like Alexa, which would have camera(s) to look around and it searches for these keywords (the last two sets) in the place it is in. If Alexa finds any of the above keywords, it can get a clue that it is probably being used in the living room. If we design more robust sets with more analytical data, Alexa can now start placing its replies relevant to the space it is kept. If Alexa identifies that it is in a kitchen, the conversation with it will be a little more specific because Alexa now knows in what context something is being asked if it were to be related to the kitchen.

Has this been already done?

1. Need to be studied
2. Need to be studied
3. Need to be studied

References

[1] Natural scene text detection challenges described by Celine Mancas-Thillou and Bernard Gosselin in their 2017 paper, Natural Scene Text Understanding

[2] https://www.pyimagesearch.com/2018/08/20/opencv-text-detection-east-text-detector/