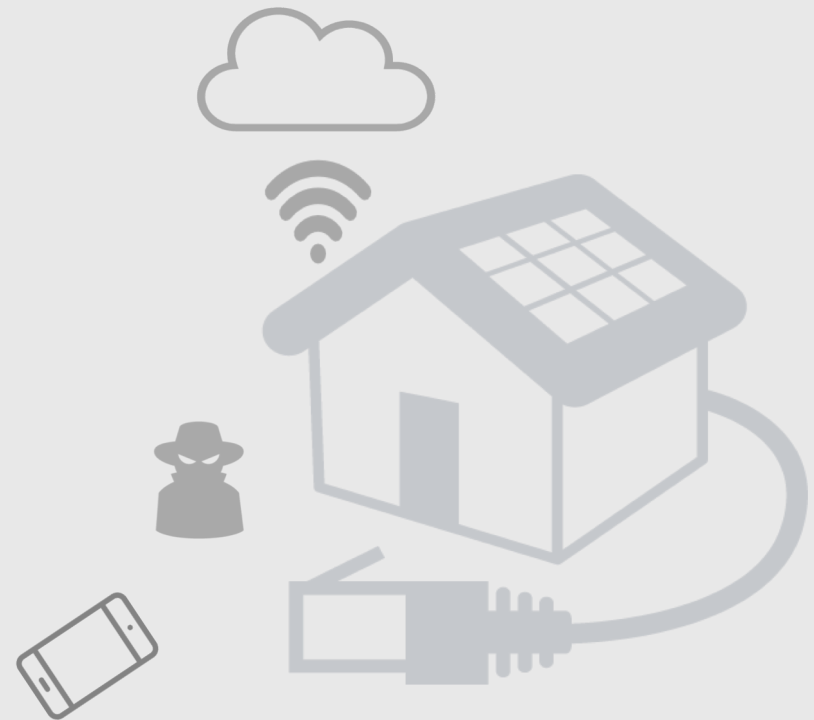# An Empirical Analysis of Smart Connected Home Data

Joseph Bugeja

Andreas Jacobsson

Paul Davidsson

MALMÖ UNIVERSITY
INTERNET OF THINGS AND PEOPLE RESEARCH CENTER

KK·stiftelsen

verisure
ALARMS WITH IQ

AGENDA

THE INTERNET OF THINGS

- In 2017, there was an estimated 8.4 billion IoT devices

- Recent surveys estimate the number of IoT devices to exceed 20 billion by 2020

- Consumer applications, e.g., the smart connected home, represent the largest user base
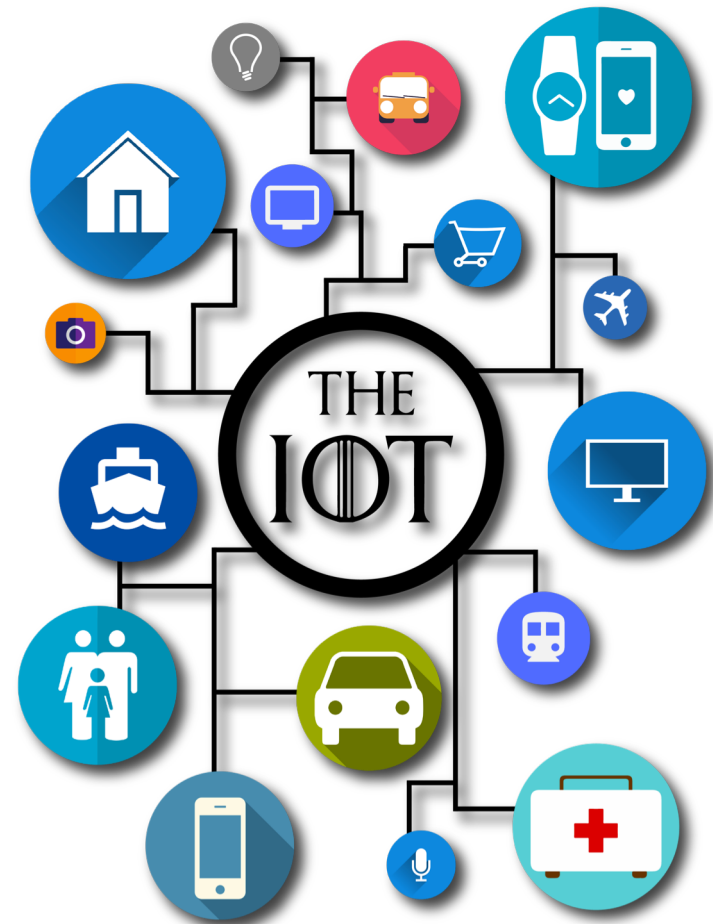
THE SMART CONNECTED HOME

- A smart connected home leverages IoT technology to improve the quality and efficiency of life to the residents

SMART CONNECTED HOME DEVICES

What are some examples of popular devices inside a smart connected home?

Plugs

> Energy and resource management

Audio speakers

> Entertainment systems

Cloud cameras

> Security and Safety

Scales

> Health and Wellness

Remote controls

> Human-machine interface

Gateways/Hubs

> Networking and utilities

Vacuum cleaners

> Household appliances and kitchen aids

Door sensors

> Sensors

MAIN RESEARCH QUESTION

What is the main research question being studied?

What type of data smart home devices collect?

PRIMARY MOTIVATION

What is the main motivation of this work?

Identifying data is key for understanding what is at stake and as a precursor especially for privacy studies

SOME LIMITATIONS OF LAB EXPERIMENTS

What are some of the limitations with the majority of lab experiments?

- **Small number of inspected devices**

    - Commonly < 10

- **Focus on a subset of devices**

    - E.g., web cameras

- **Target some data states**

    - E.g., concerning stored data
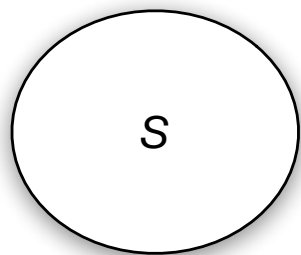
PRIVACY POLICY

What is a privacy policy and why it is important?

- Governing document serving as a 'contract' between users and their service providers

- Aims to answer questions related to data practices of a company

- In the US, regulators e.g., the FTC, leverage policies to enforce compliance and accountability

- The European GDPR increased the importance of privacy policies to meet individual rights to restrict data processing
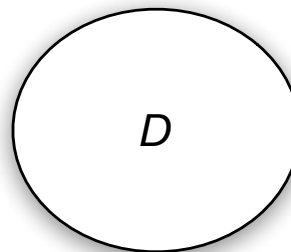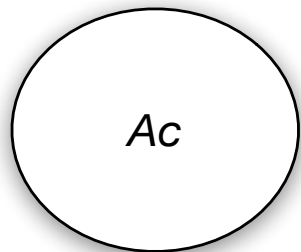
PRIVACY POLICY STRUCTURE
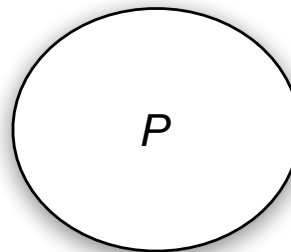
What is the basic structure of a privacy policy?
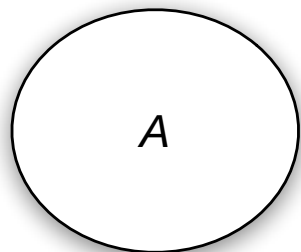
S **Scope**

D **Data Elements**

Ac **Actors**

P **Purposes**

A **Actions**

PRIVACY POLICY STRUCTURE

What is the basic structure of a privacy policy?

- **Scope**
    - *{device, website, service, ...}*

- **Actors**
    - *{service provider, third-party, affiliate, ...}*

- **Actions**
    - *{interaction, collection, dissemination, ...}*

- **Data Elements**
    - *{ip address, username, password, ...}*

- **Purposes**
    - *{ads, tech support, payment processing, ...}*

PRIVACY POLICY STRUCTURE

What components of a privacy policy were used for answering the posed research question?

- **Scope**
    - *{device & website & service}*

- **Actors**
    - *{service provider, third-party, affiliate, ...}*

- **Actions**
    - *{interaction, collection, dissemination, ...}*

- **Data Elements**
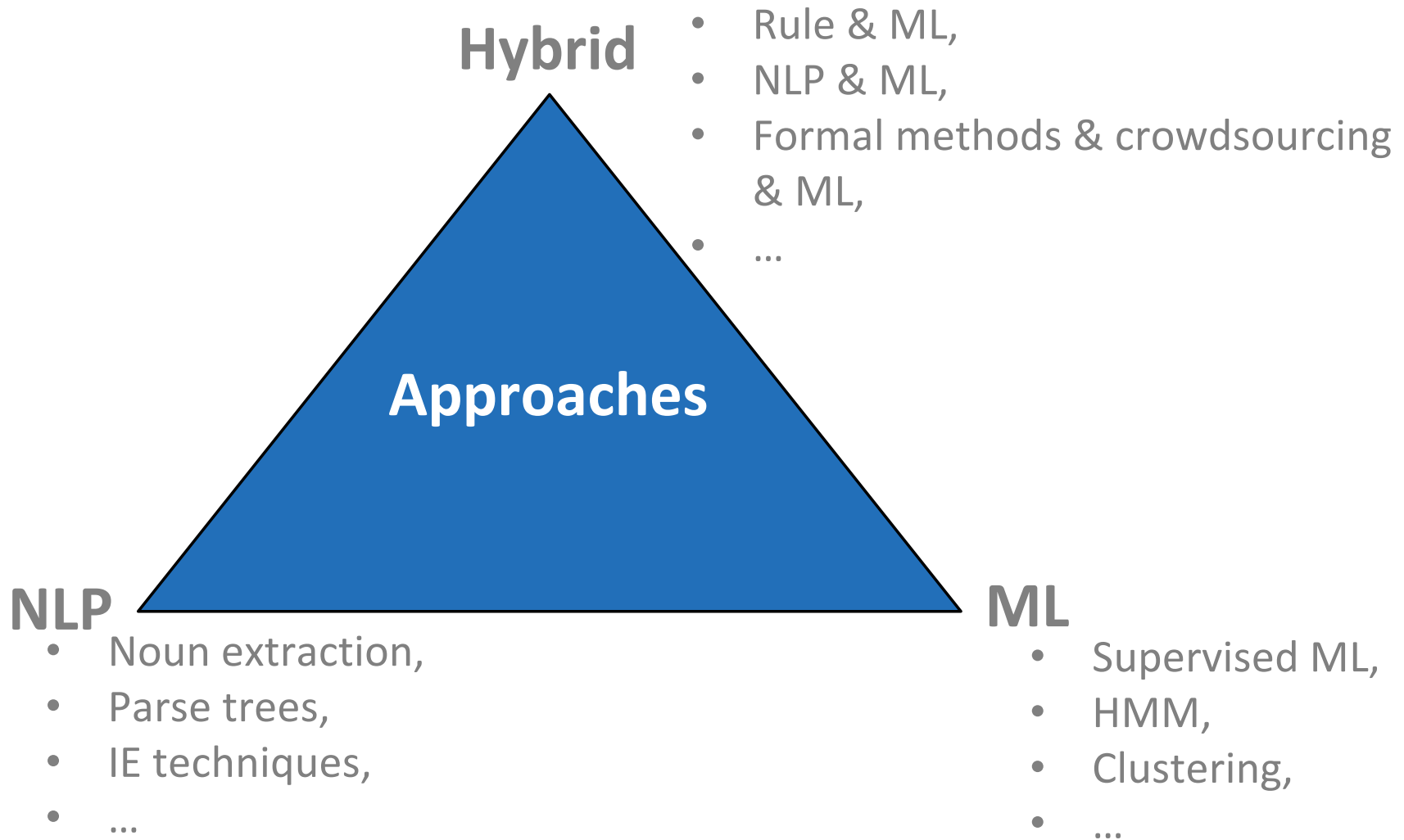   - *{ip address, username, password, ...}*

- **Purposes**
    - *{ads, tech support, payment processing, ...}*

EXAMINING PRIVACY POLICIES

What are some of the existing approaches used for analyzing privacy policies?

- Privacy policies tend to be unstructured and are in general written by and intended for humans

Hybrid

- Rule & ML,
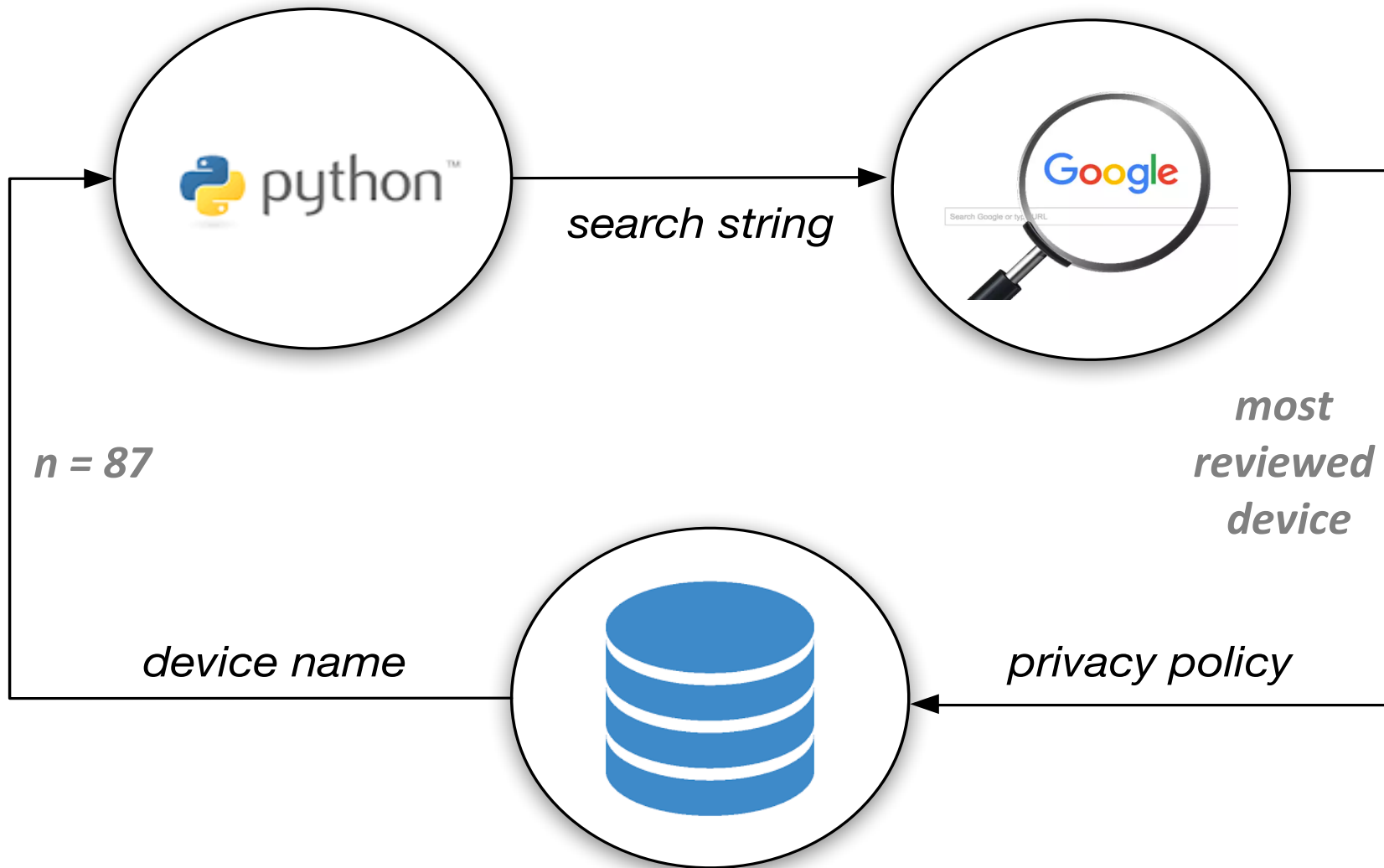- NLP & ML,
- Formal methods & crowdsourcing & ML,
- …

Approaches

NLP

- Noun extraction,
- Parse trees,
- IE techniques,
- …

ML

- Supervised ML,
- HMM,
- Clustering,
- …

SMART HOME DEVICE DATABASE

What devices were examined in this study?

Name	Category	Number of Reviews
Amazon Fire TV (2nd Gen)	MediaPlayer	34372
Ring Video Doorbell	DoorBell	27862
Ring Chime	AudioSpeaker	26199
Nest Learning Thermostat (3rd Gen)	Thermostat	18614
Arlo 2 HD Camera Security System	CloudCamera	16446
Fitbit Aria Wi-Fi Smart Scale	Scale	9402
Apple TV (4th Gen, 32GB)	Gateway/Hub:OpenEcosystem	9417
TP-LINK Wi-Fi Smart Plug	Plug	8682
TP-Link Smart Light Switch	LightSwitch	8828
Ooma Telo	Gateway/Hub:ClosedEcosystem	4138
Google Home	VoiceCommandDevice	5562
eero Home WiFi System	WirelessSignalExtender	3034
GE In-Wall Add-On Switch	Switch	2545
GE Wall Outlet	PowerOutlet	2528
iRobot Roomba 880	VacuumCleaner	2491
SONOS CONNECT	MusicPlayer	2560
Chamberlain MyQ Garage	GarageDoorController	2105
iRobot Braava	FloorMopper	1904
TP-Link Smart Bulb LB110 (White Light)	LightBulb:WhiteLight	2376
TP-Link Smart Bulb LB130 (Color)	LightBulb:Color	2375
Samsung UN55KU6300 55-Inch 4K Ultra HD Smart TV	TV	1504
Rachio Smart Sprinkler Controller, 8 Zones (2nd Gen)	IrrigationController	1952
Tagg GPS Pet Tracker	Tracker	1299
Withings Wireless Blood Pressure Monitor	BloodPressureMonitor	1302
Dolphin Nautilus Robotic Pool Cleaner	PoolCleaner	1253
Nest S2001 Protect	SmokeDetector	1004
Philips Hue Dimmer Switch	RemoteControl	1148

DATA COLLECTION

How were smart home policies retrieved?

search string

most reviewed device

n = 87

device name

privacy policy

DATA PROCESSING

How were the gathered privacy policies analyzed?

- Manual sorting of policies according to their scope
- Translated non-English docs to English

- Removed sections from policies not pertaining to data collection
- Preprocessed text using standard data cleaning functions

N-grams {gps, ip addr, …}

- Transformed document into a TDM
- Retrieved uni/bigrams (possible data types)
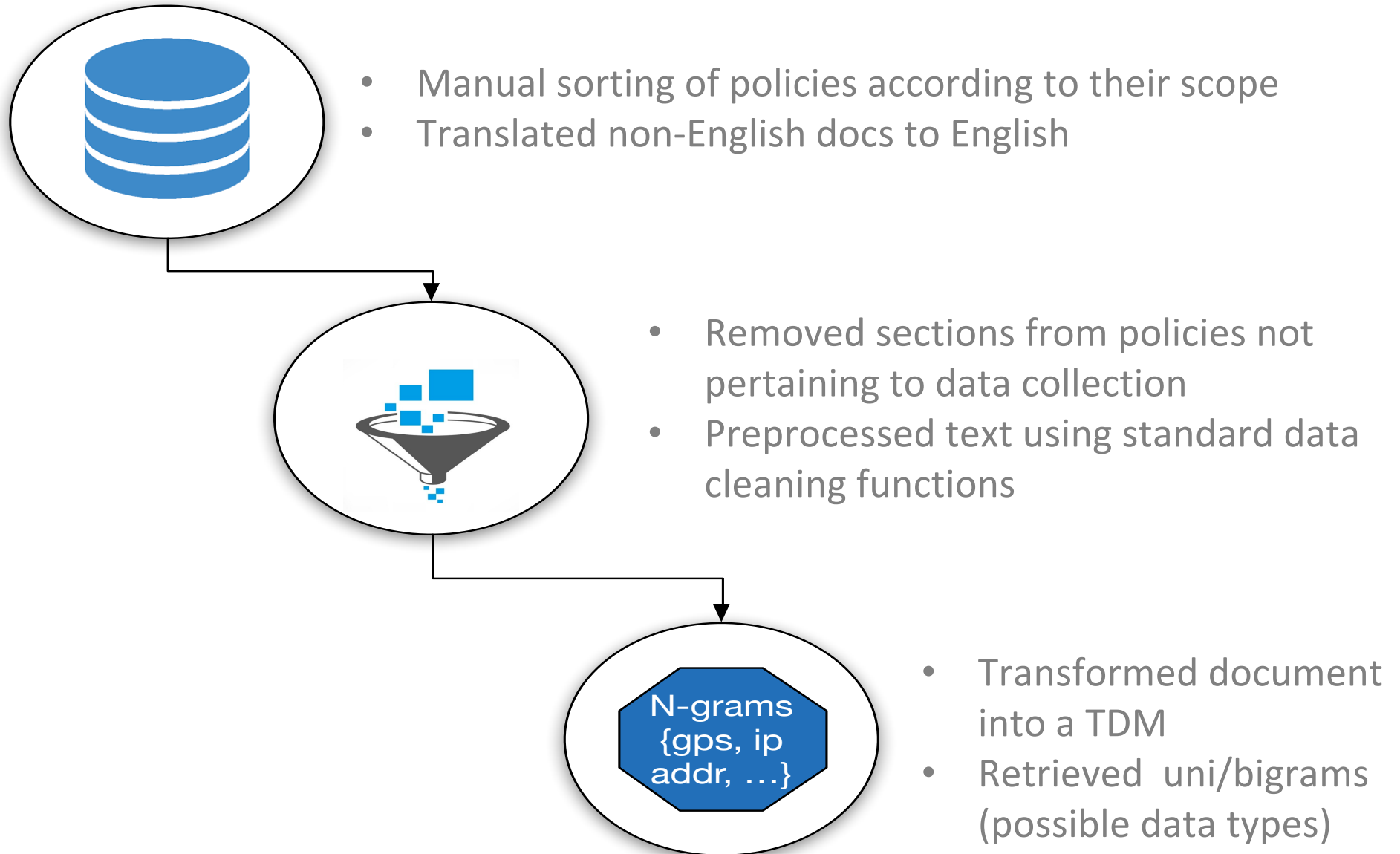
IDENTIFYING DATA TYPES

How were the data types identified and categorized?

- Hybrid approach combining manual & automated data extraction techniques

Data type Identification

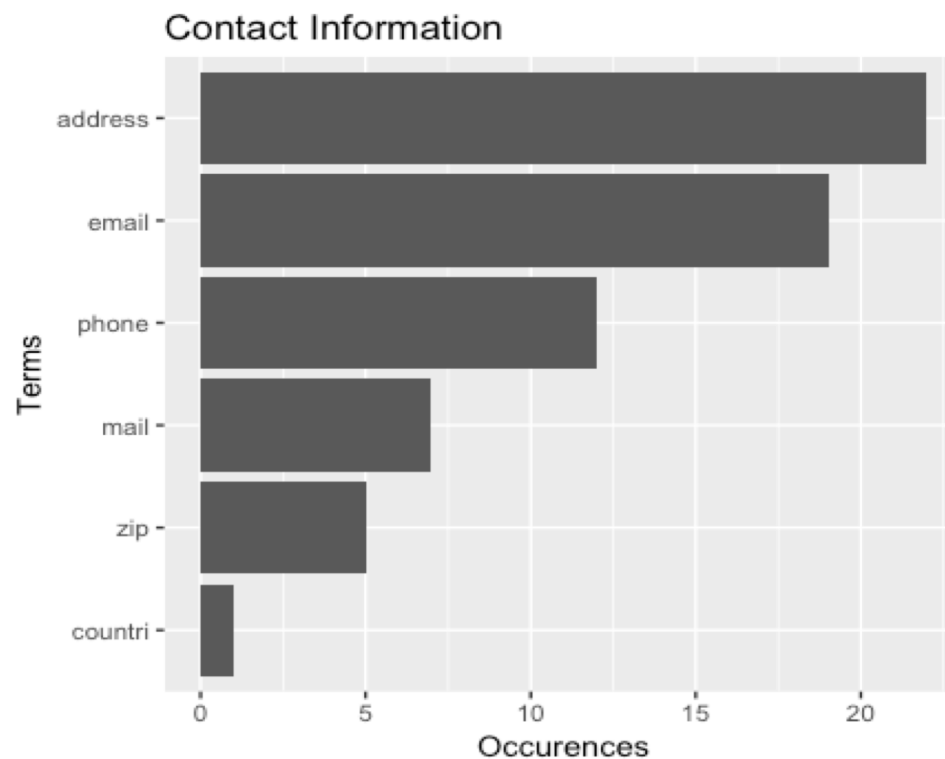| word	freq
person data	11
prefer	11
profil inform	11
signal	11
technic inform	11
temperatur	11

Grouping data types

SMART HOME DATA COLLECTION ENTITIES

What are the main subjects of data collection by a smart connected home system?

Entities

User

- Residents
- Guests
- Maintenance Staff
- ...

Device

- Sensor
- Home Appliance
- End-User Device
- ...
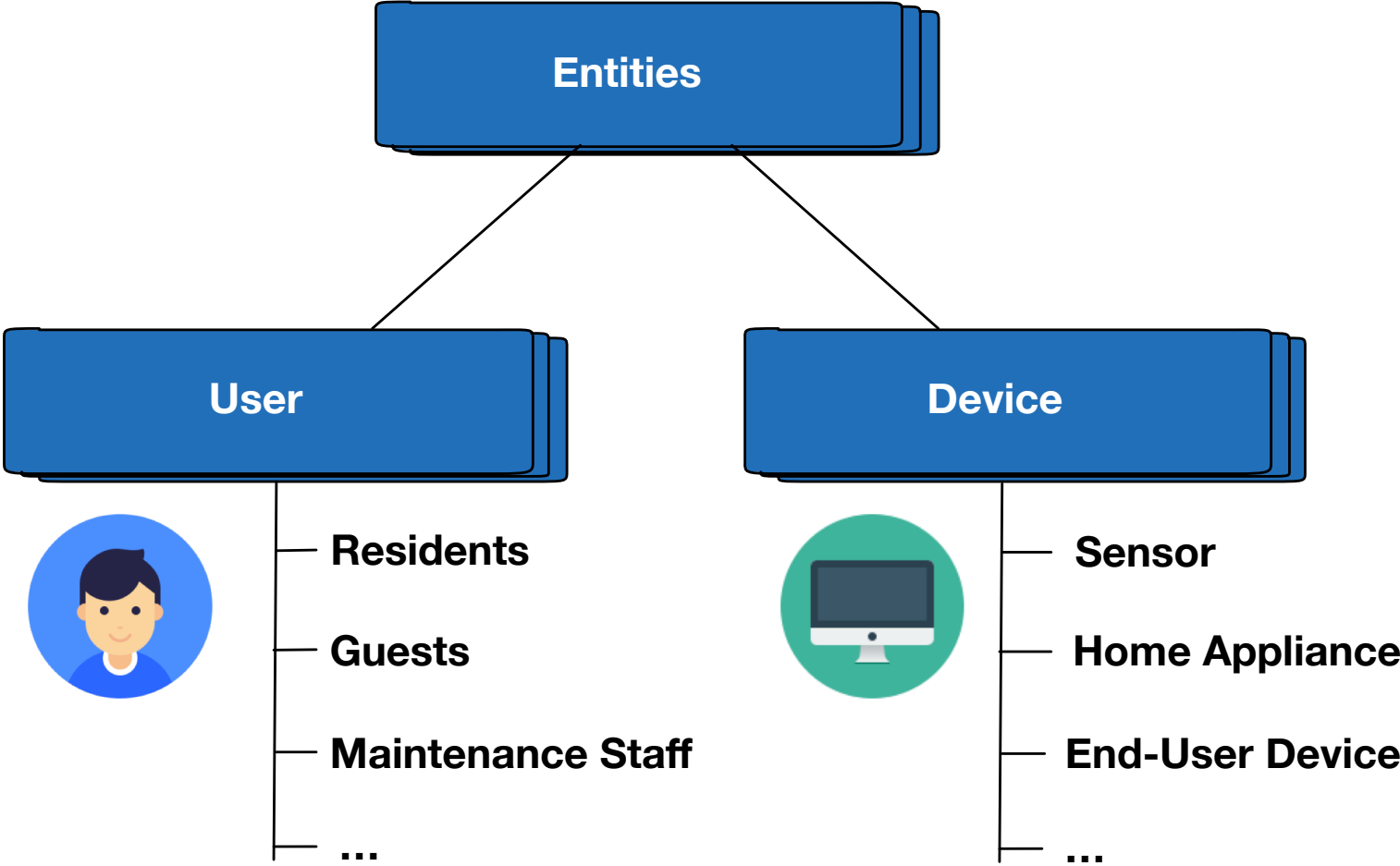
USER DATA

What type of data pertaining to the user are collected by a smart home system?

Contact Information

User-generated Content

Personal and Account Details

Configuration Settings

Offline Data

User Activity Data

Financial Information

Location Information
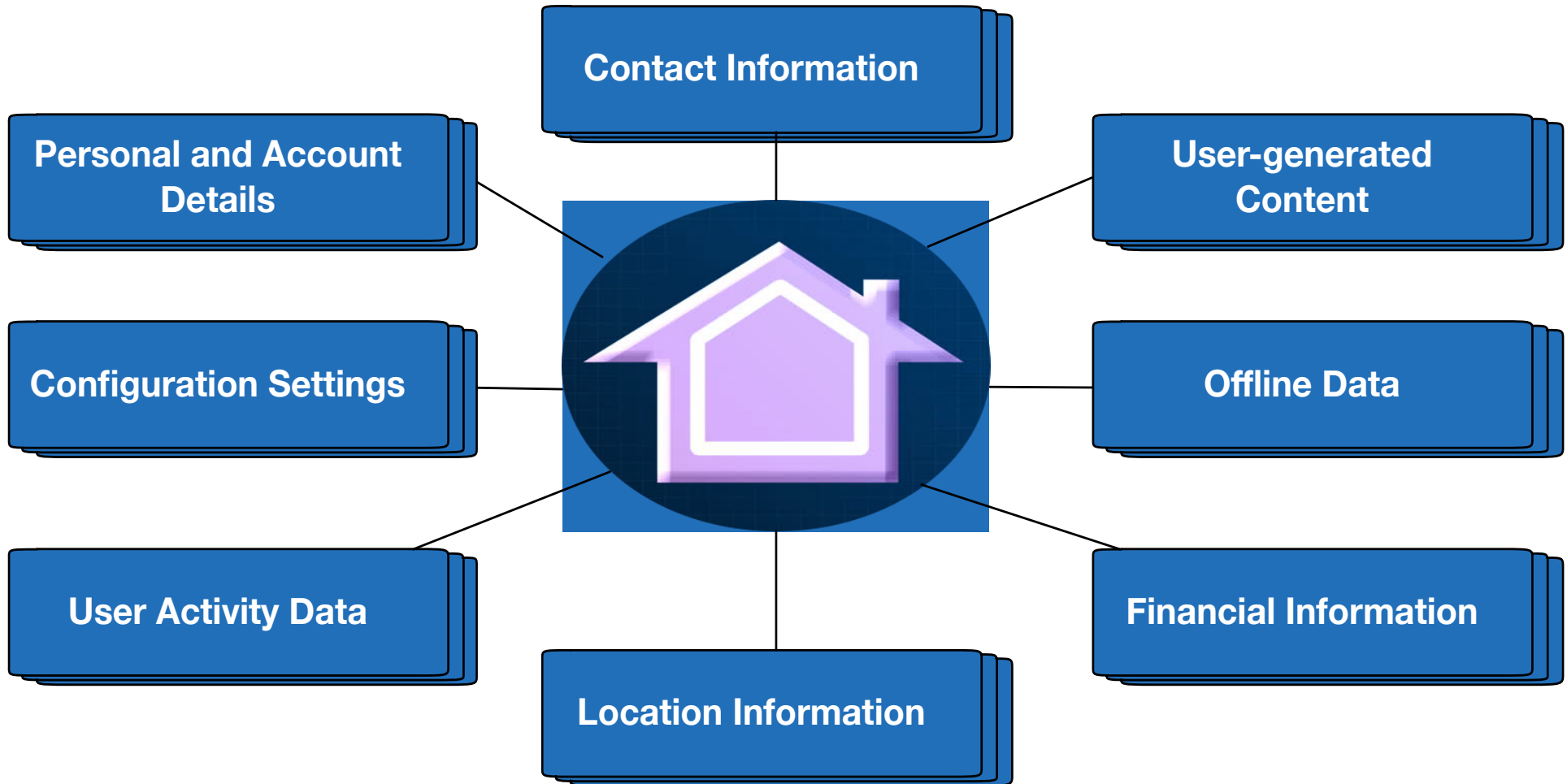
DEVICE DATA

What type of data pertaining to the device are collected by a smart home system?

Device Information

— Device identifiers

— Performance related

— Network/connectivity

— Firmware and software version

— Sensor status

— …

Environmental Data

— motion

— humidity

— ambient light

— CO2 concentration

— rain level

— …
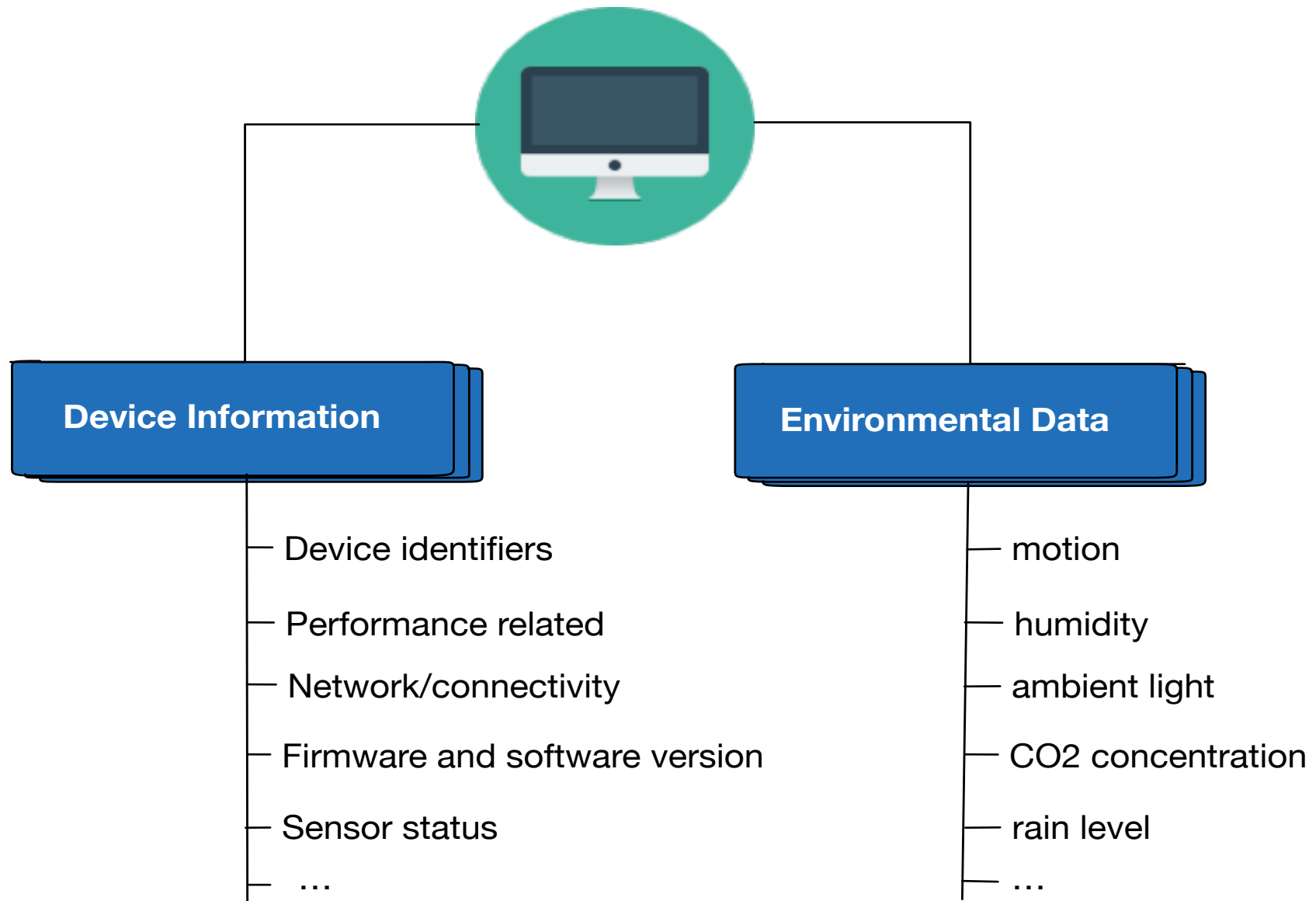
SOME OBSERVATIONS

What are some interesting observations from the identified data?

- **All devices tend to capture:**
    - *Contact info, device info, personal & account details, and user activity data**

- **2/39 device types collect all identified data categories:**
    - *Gateways/hubs and media player*

- **Least captured data categories:**
    - *Configuration settings and offline data*

*- Except, accelerometer sensor and shower head water meter

SOME IMPLICATIONS

What are some of the interesting implications of this study?

- The continuous collection and monitoring by sensor technologies inside the home carries serious privacy implications

- Indeed, all surveyed manufacturers collect instances of personal data (can be very sensitive!)

- Some categories of data are collected automatically and cannot be opted-out (easily) without adversely affecting functionality

CLOSING REMARKS

- Analyzed the most reviewed smart home devices in terms of their collected data

- Identified 10 different data categories including their data source, collection method, and process (more details in paper)

- First contribution in this area that targets such a broad range of devices and using actual manufacturer privacy policies as a medium

FUTURE WORK

What are some possible avenues for future work?

- Identify privacy practices of service providers

- Conduct lab experiment as a complementary method

- Develop controls that allow users to be notified about data collection

Thank you for your attention!

joseph.bugeja@mau.se

https://www.bugejajoseph.com