

# Beyond Camera-Based Document Analysis and Recognition: Towards Camera-Based Human Document Interactions

Koichi Kise

Osaka Prefecture University, Japan



# Self-Introduction

- Koichi Kise
  - Professor, Osaka Prefecture Univ., Japan
  - Editor-in-Chief, IJDAR
  - Research
    - layout analysis
    - camera-based document retrieval
    - human document interaction
    - intelligence augmentation
  - Started CBDAR in 2005





# Acknowledgment



Experiential  
Supplements

LeCycl  
Learning Cyclotron



Human-Machine  
**Harmonious**  
Collaboration

DFG-ANR-JST  
Joint AI project



Japan Science and Technology Agency

科研費  
KAKENHI



JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE

日本学術振興会



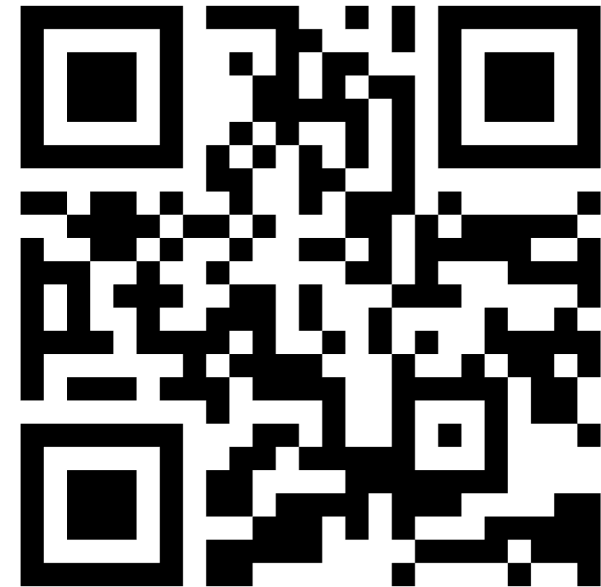
Osaka Pref. Univ.



Osaka  
Metropolitan  
University

# Goal of This Talk

- Not to give a comprehensive summary of SOTA
- Short overview of what we have done
- Think together future topics
- On-line anonymous questions:  
**slido.com #777649**



# CBDAR

Document Analysis goes to the field.



September 6, 2021

CBDAR2021 Keynote

# We are doing CBDAR

---

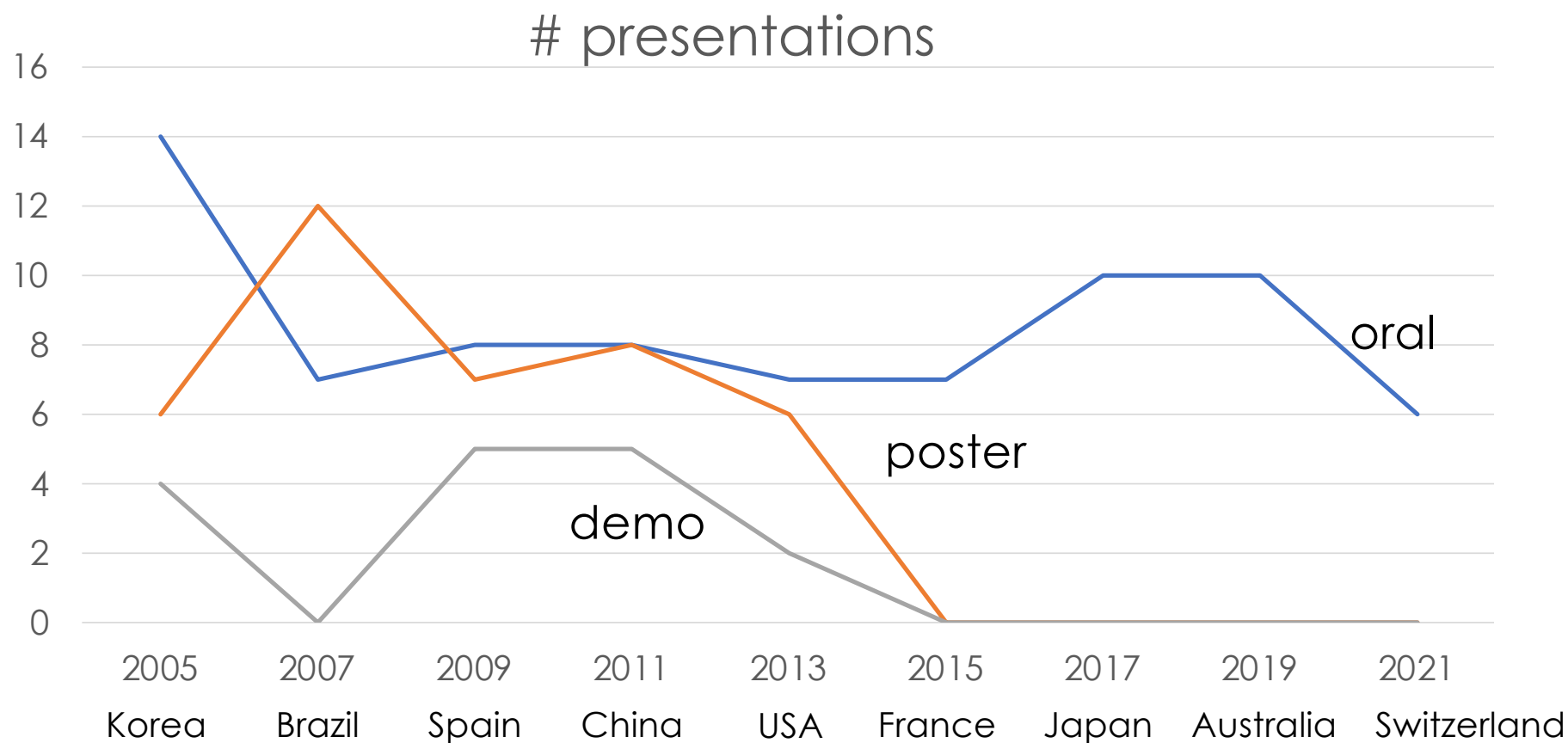
by using our eyes as a  
camera

September 6, 2021

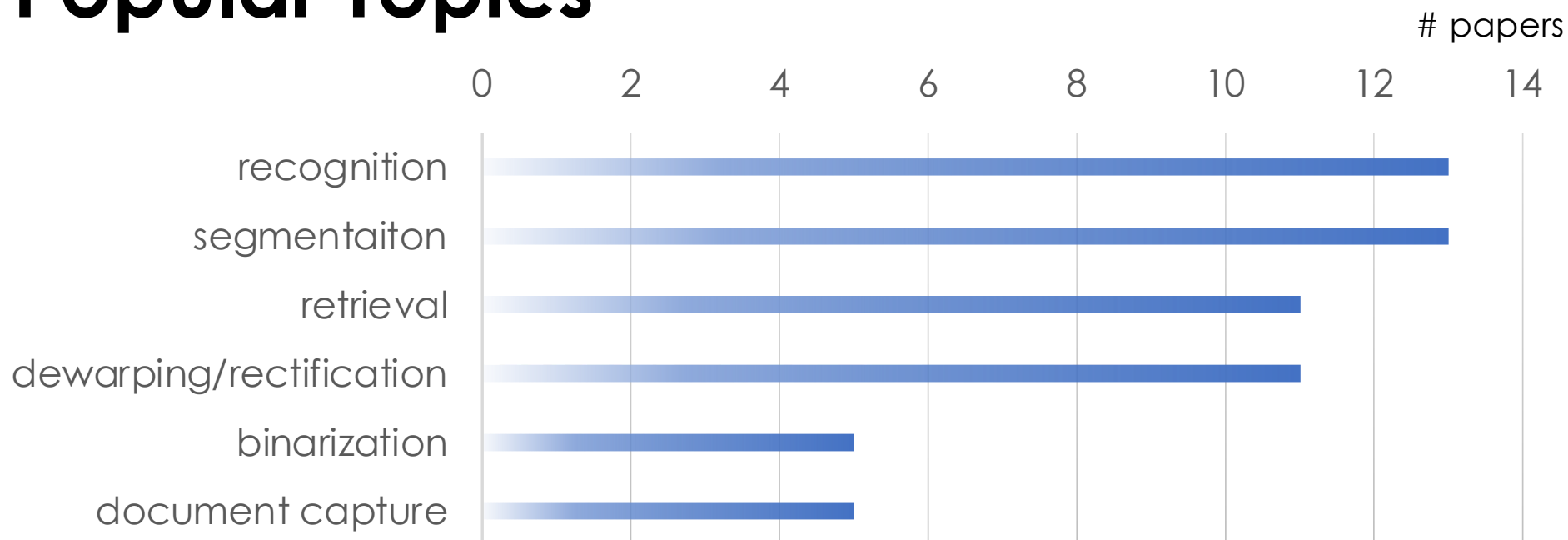


CBDAR2021 Keynote

# History & Stats



# Popular Topics



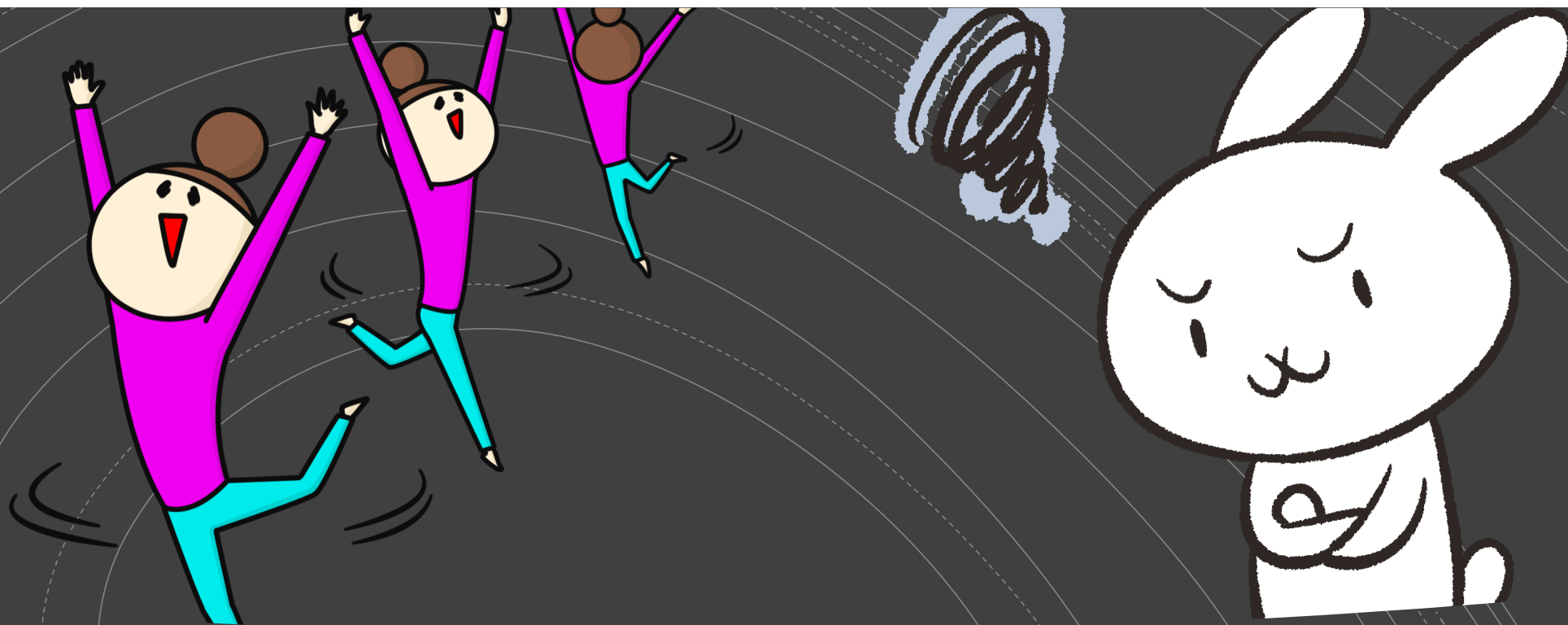
DNN based methods started to appear from 2017

No more low-level processing and retrieval in recent years

# Basic Questions

- Are you satisfied by current research topics and status?
  - end-to-end
  - accuracy continuously improved
- We have started the FDAR workshop since 2017 @Kyoto
  - Important question given by Prof. Uchida





**What if you achieve 100% accuracy?**



This has been  
the ultimate  
goal.

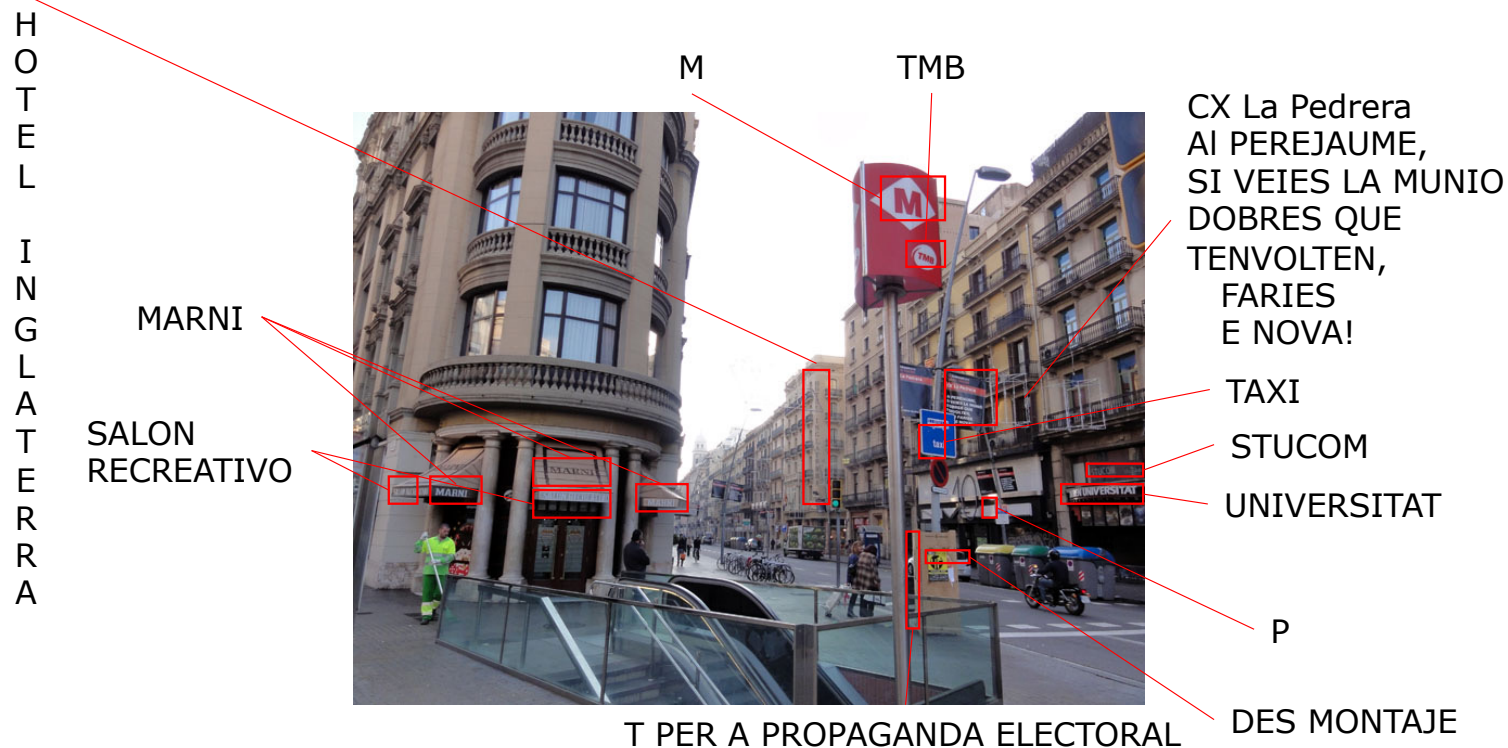
All problems  
can be solved.





Nothing  
to do  
anymore.

# What happens?



# Just telling the text around you...

Especially for the case of blind people



hotel ingratella, M, TMB, CX La Pedrera AI PEREJAUME,...



Masa Iwamura and his students  
have been working on this topic.

# What is necessary?

Goal



I am looking for a hotel.

But now I am feeling thirsty.

Can I buy something to drink?

What to recognize depends on:

Internal states of the user

Environment

Time

# What we have done / What we should do

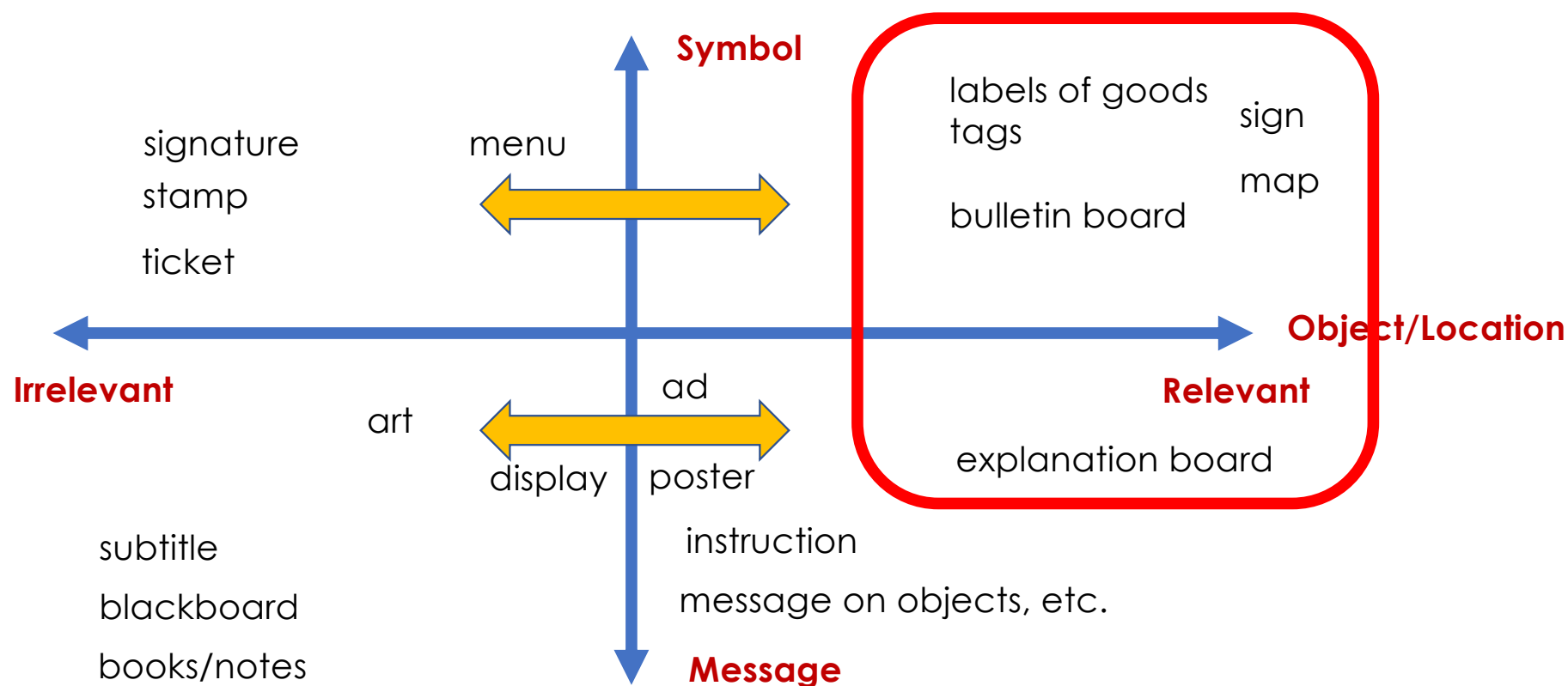
- What we have done:
  - General purpose “recognition”
  - Without thinking the use of results
- What we should do:
  - Focusing on use cases
  - Thinking more about the user

# Various Roles of Scene Texts

- Before discussing what to do, let's take a look at various roles of scene texts
- Their role is not just labeling



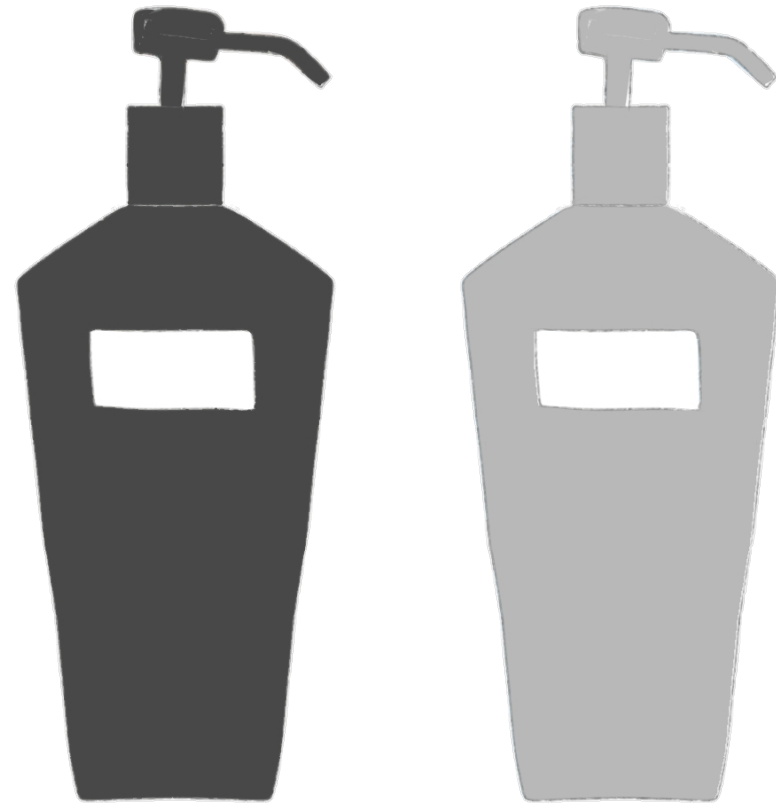
# Spectrum





# Labels of goods, tags

distinction



# Labels of goods, tags





# Labels of goods, tags: exceptional cases

Labels detached from objects



Incorrect labels





# Explanation Boards



September 6, 2021



CBDAR2021 Keynote



22



## 大阪の渡船

海外大衆向けには、その性質や考えを伝える活動があり、1人1人の輸出戦略も必要だと言っただけでなく、人々の理解を得ています。最も多くの人々が海外へ、海外へと伸びた大企業には、人々の心と文化への理解が必要になります。当然のことによって実現された活動は、約20年ほど前に大企業で「海外営業戦略」を定めた「海外展開計画」を行っていたように、約20年ほど前には、世界市場及び国際市場の主要市場については世界地図としてその位置とそれが実現することになりました。

又昭和4年、明治政府の政策により通関税は無料となり、昭和7・8年にはそれまでの保護税を撤廃、ほとんど自由貿易政策の元となりました。そして昭和18年には通関税31の税、通入税額68億（昭和前期）に達、半導体は107億、電機材料は10億のものが国内に輸入、通関税率の44.4%を輸入しました。

しかし、その後の戦後の年間にさほど急激な増加は現れておらず、逆に減少しています。昭和60年に比べて昭和70年にはおよそその半分に減りました。昭和83年には1日あたり消費されましたが、戦後復興とともに消費を促しとめてアスファルトが普及された。セー  
レーションの進展もあって道路の利用率は大幅に減少し、昭和83年には道路幅員13.9m、内車道幅員10.0万人に  
達して以来、約17万人になっています。

今後も、できるだけ多くの皆様に参加していただき、安全通関士として責任を担うのであ  
ることをさせていただきますようお願いいたします。



- ① 天保山泥岩
- ② 天保山砂岩
- ③ 千歲山砂岩
- ④ 聯合上泥岩
- ⑤ 聯合下泥岩
- ⑥ 千本莊泥岩
- ⑦ 船町泥岩
- ⑧ 木津川泥岩

大坂市建設局  
河川・施設管理課様

could be location specific

[illegible]

左のポスター: 携帯電話ご利用のお願い  
 授業中は、携帯電話の電源を切って、静かに授業を受けてください。  
 授業中に携帯電話の電源を切らないで、授業中に携帯電話の音で迷惑をかける場合があります。  
 授業中に携帯電話の電源を切らないで、授業中に携帯電話の音で迷惑をかける場合があります。

右のポスター: 携帯電話の持ち出し禁止  
 携帯電話の持ち出し禁止の理由  
 1. 授業中に携帯電話の電源を切らないで、授業中に携帯電話の音で迷惑をかける場合があります。  
 2. 授業中に携帯電話の電源を切らないで、授業中に携帯電話の音で迷惑をかける場合があります。  
 3. 授業中に携帯電話の電源を切らないで、授業中に携帯電話の音で迷惑をかける場合があります。  
 4. 授業中に携帯電話の電源を切らないで、授業中に携帯電話の音で迷惑をかける場合があります。  
 5. 授業中に携帯電話の電源を切らないで、授業中に携帯電話の音で迷惑をかける場合があります。

September 6, 2021

2021 渡船場ご利用の皆様へ  
事故や故障でのダイヤの乱れが  
生じた時は、下記の要領でご乗

## 21 Keynote



# Signs and Maps



September 6, 2021

CBDAR2021 Keynote

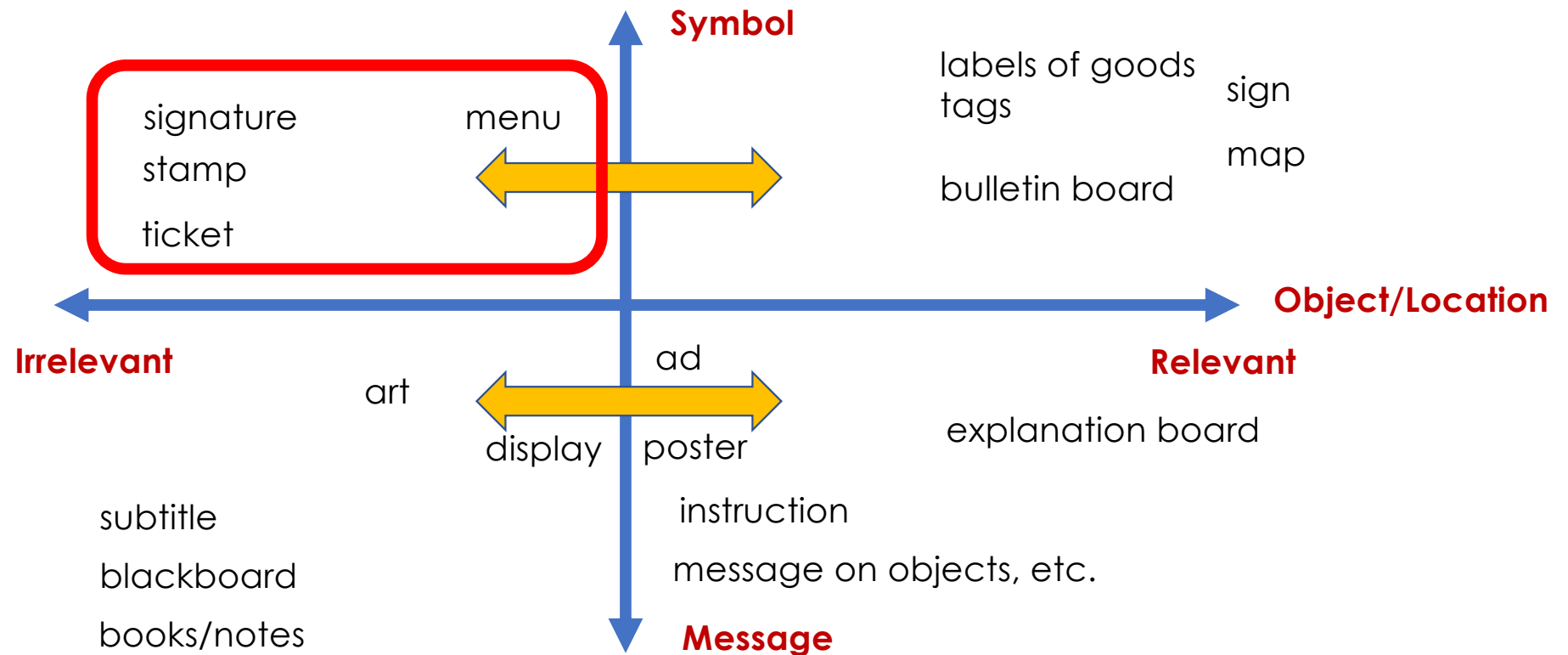
24







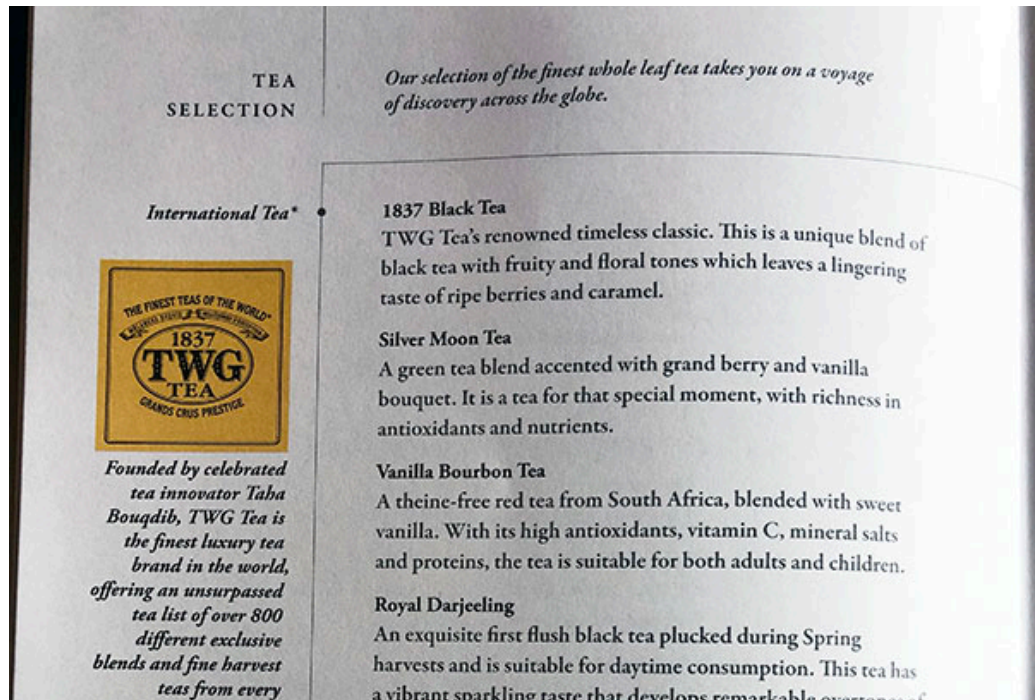
# Spectrum





# Menu

Location irrelevant



September 6, 2021

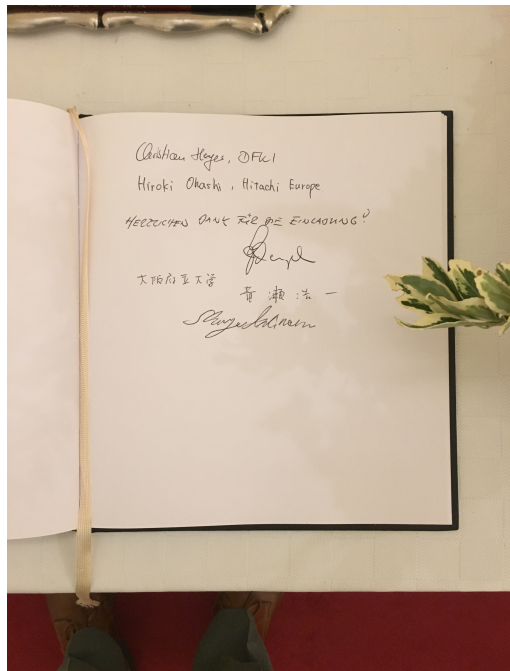
Location relevant



CBDAR2021 Keynote

27

# Signatures, Stamps, and Tickets



September 6, 2021



CBDAR2021 Keynote



Mobile ticket Timetable

10/6/18 Kaiserslautern Hbf - F-  
RPWKRV Flughafen Fernbf.

minianrt: Kaiserslautern+City -  
Frankfurt(M)Flugh., mit ICE

Über: VIA: NW\*SIF\*LU\*  
(WO\*MZ/MA\*BENS\*DA)

Gilt nur mit einem Personalausweis, Reisepass,  
elektronischen Aufenthaltstitel oder BüMA.

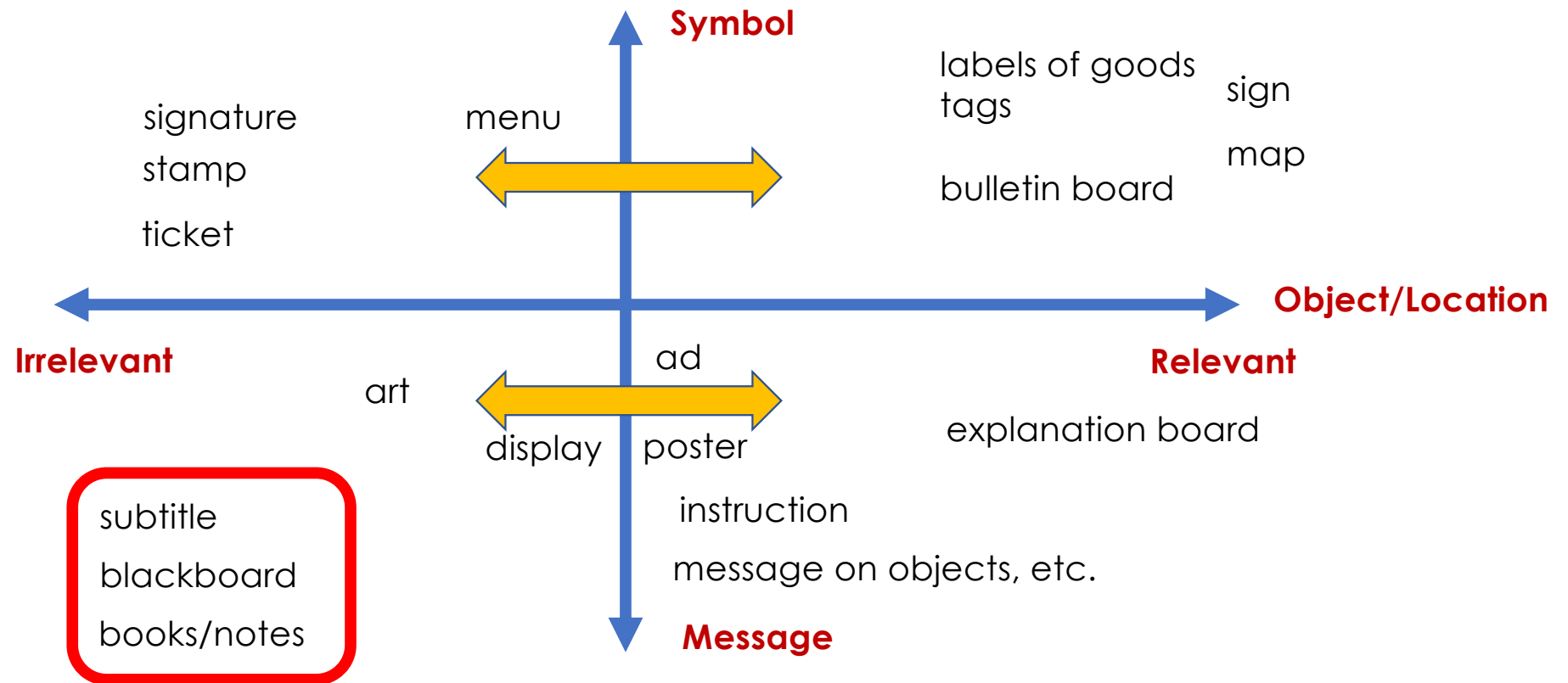
Storno kostenfrei bis 1 Tag vor 1. Geltungstag.

Auftrags-Nr: RPWKRV  
Gesamtpreis: 37,00 EUR  
Gebucht am 06.10.2018 um 12:52 Uhr



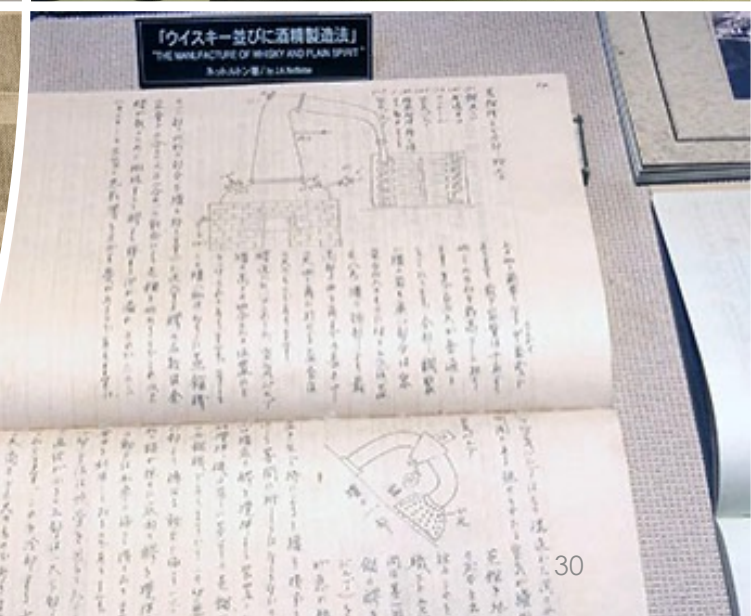
28

# Spectrum

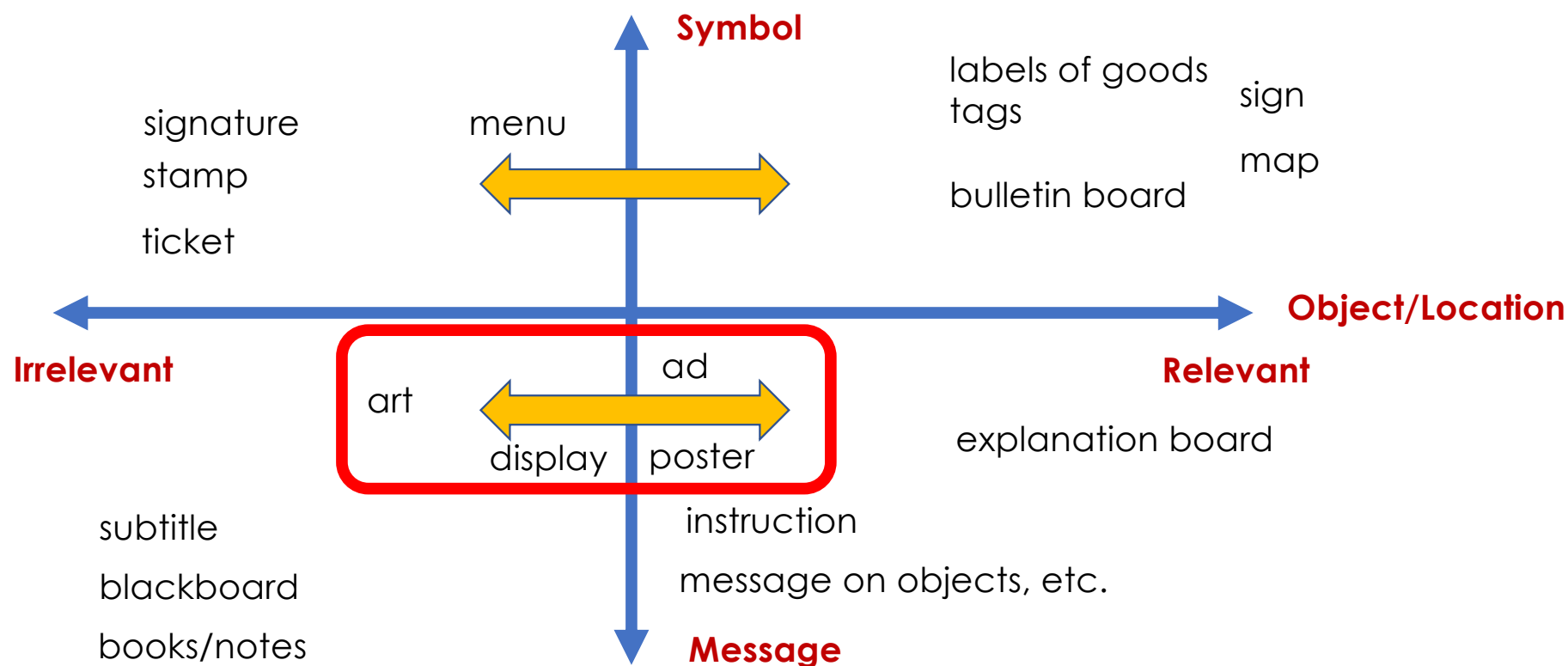




- These have been objects for recognition before CBDAR



# Spectrum







# Ads and Posters

September 6, 2021

CBDAR2021 Keynote

32



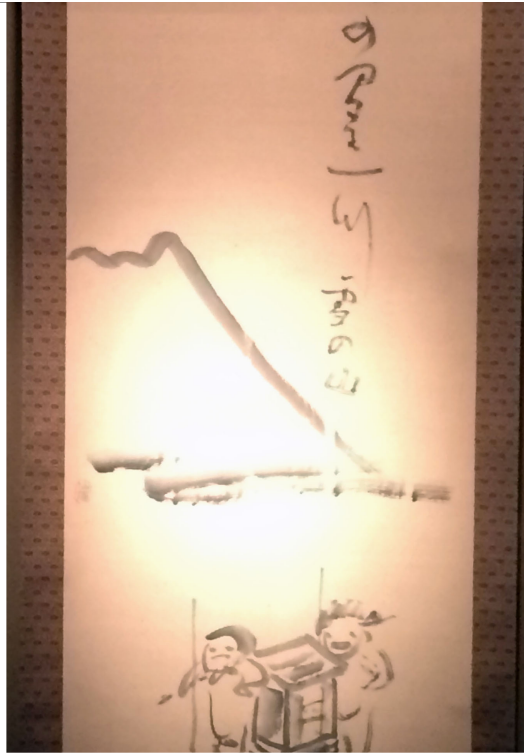


## Display

- Location relevant and irrelevant







# Arts

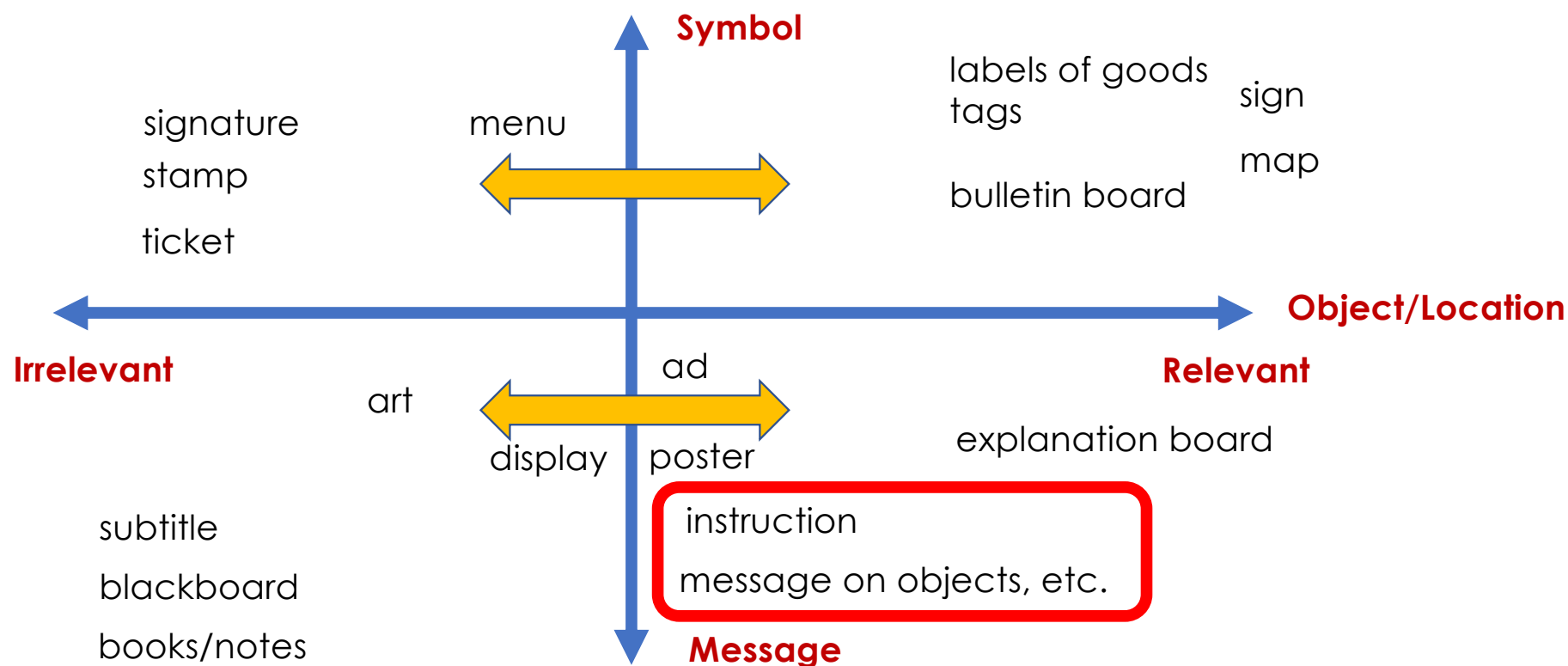
September 6, 2021

CBDAR2021 Keynote

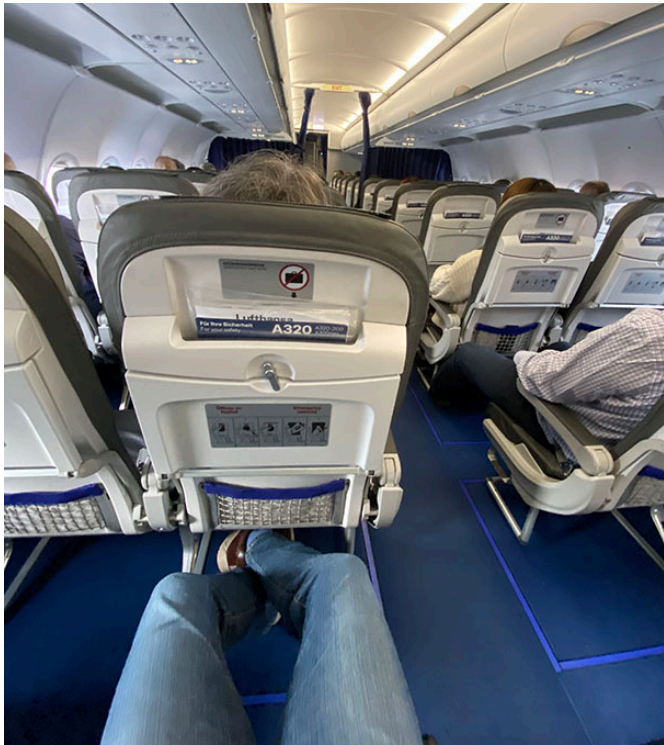
34



# Spectrum



# Instructions



September 6, 2021

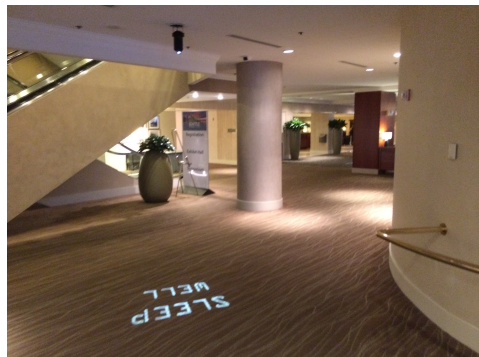


CBDAR2021 Keynote



36





# Messages on Objects

September 6, 2021

CBDAR2021 Keynote

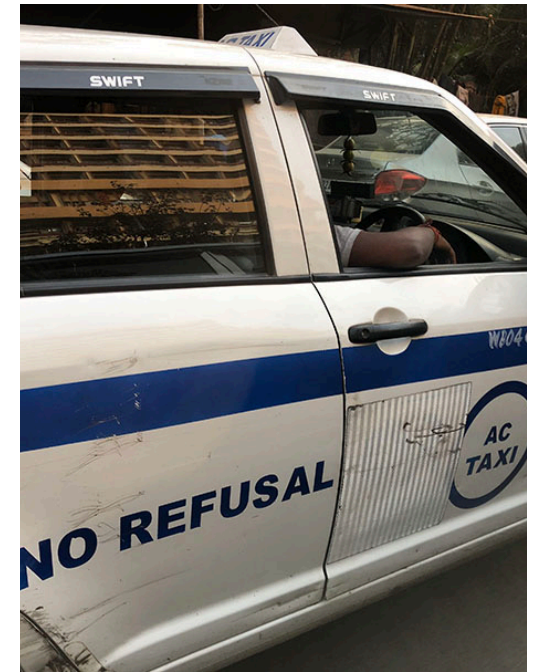
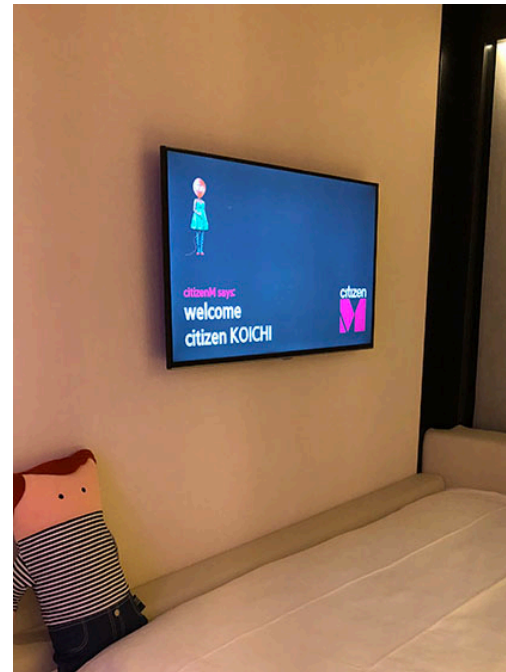
# Intentions of Messages



September 6, 2021



CBDAR2021 Keynote



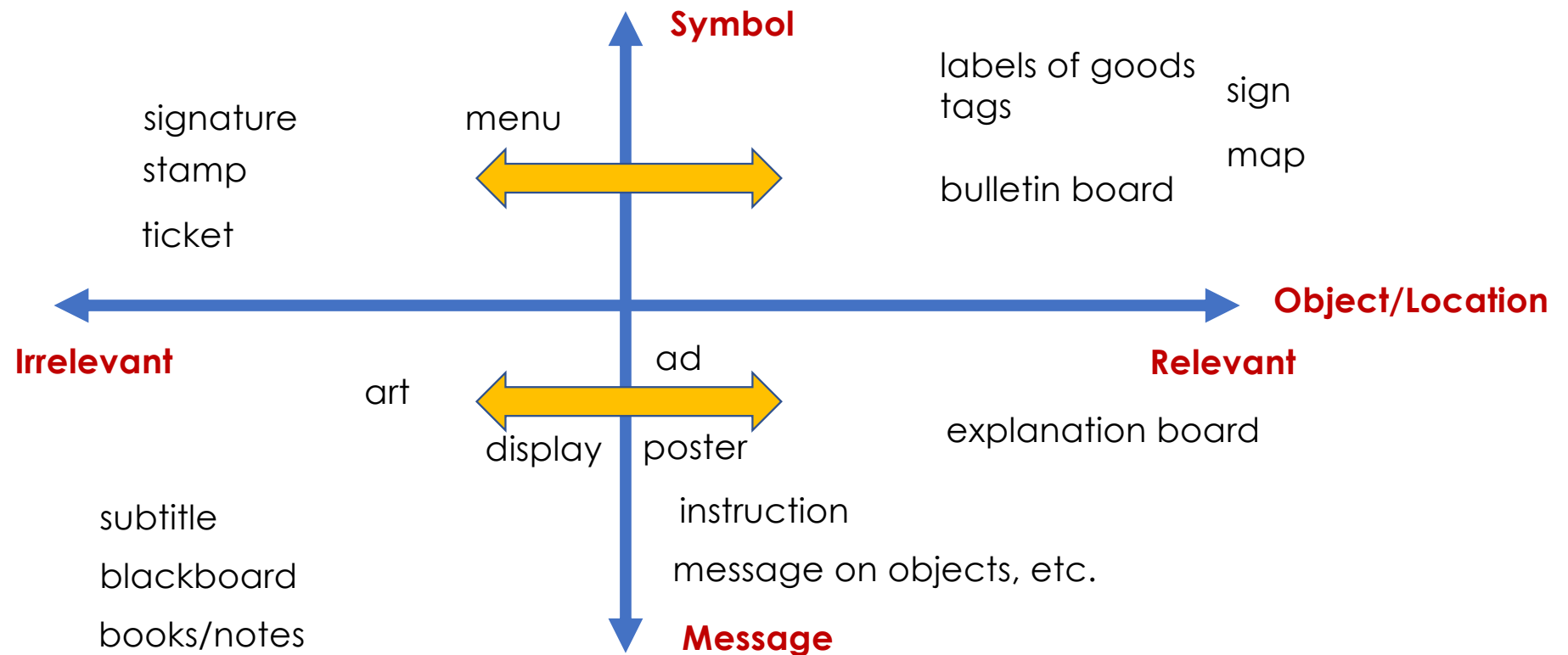


# Religious Messages



September 6, 2021

# Spectrum



# Points to Consider

- object/location relevant?
  - yes: label, tag, explanation of an object, or a location
  - no: books, notes, signatures, stamps, tickets, etc.
  - depends: bulletin board, menu, instruction, etc.
  - exceptional cases: detached label
- role of the scene text
  - typical roles: label, explanation, message
  - exceptional cases: design, art, ad, religion

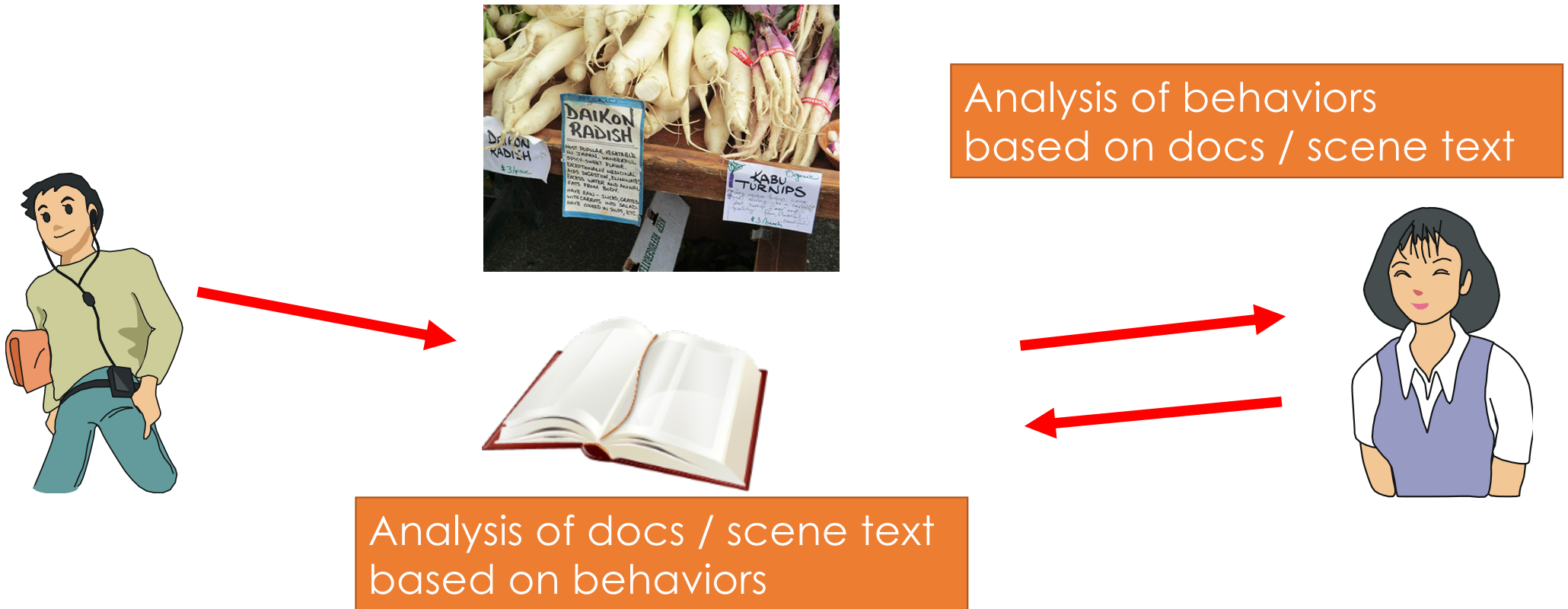
# What we have done / What we should do

- What we have done:
  - General purpose “recognition”
  - Without thinking the use of results
- What we should do:
  - Focusing on use cases
  - Thinking more about the user

## Human-Document Interaction



# Human-Document Interaction



# Explanation of Objects

Analysis of **scene text** based on behaviors

- Audience rating
- Effectiveness

Analysis of **behaviors** based on scene text

- Buying with/without reading



# Acquiring knowledge

Analysis of **docs**  
based on behaviors

- Difficult parts to understand
- Interests

Analysis of **behaviors**  
based on docs

- what have been learnt
- level of understanding



# What we know thru behaviors

- role of scene texts, intention
- property of scene texts (or objects)
  - interest
  - preference
  - difficulty
  - audience rating
- effectiveness to behavior change

# What we know thru documents/scene texts

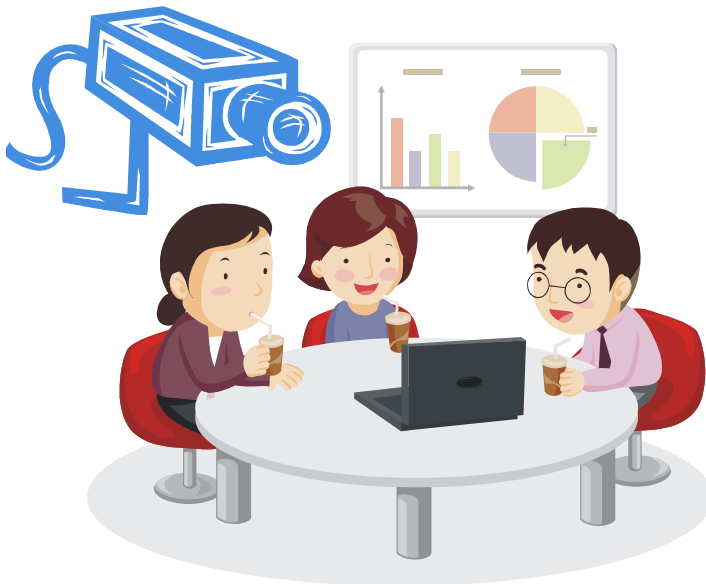
- property of humans
  - interest
  - preference
  - level of knowledge
- influence on his/her behavior

# How to Implement

- Timing
  - real-time?
- Viewpoint
  - where to put the camera?



# Where to put the camera?



3<sup>rd</sup> person



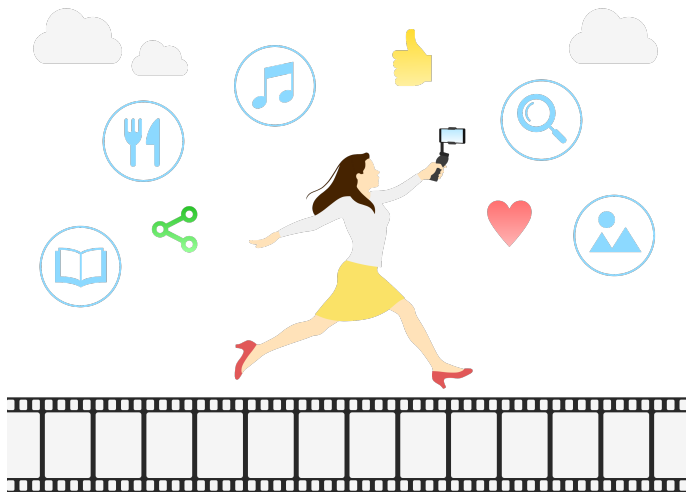
2<sup>nd</sup> person



1<sup>st</sup> person

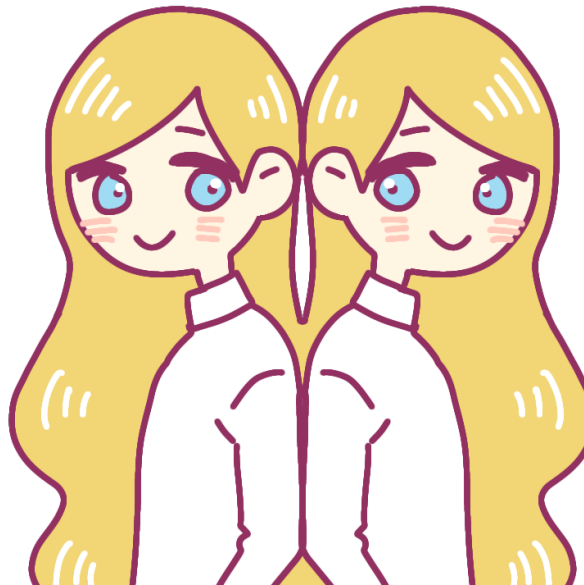
# Applications of CBDAR

Personal Life Record



September 6, 2021

Digital Twin



CBDAR2021 Keynote

Digital Transformation



50



# Personal Life Record

- Can be partially done on the digital world
- Full coverage requires interactions with the physical world
- CBDAR to capture such interactions



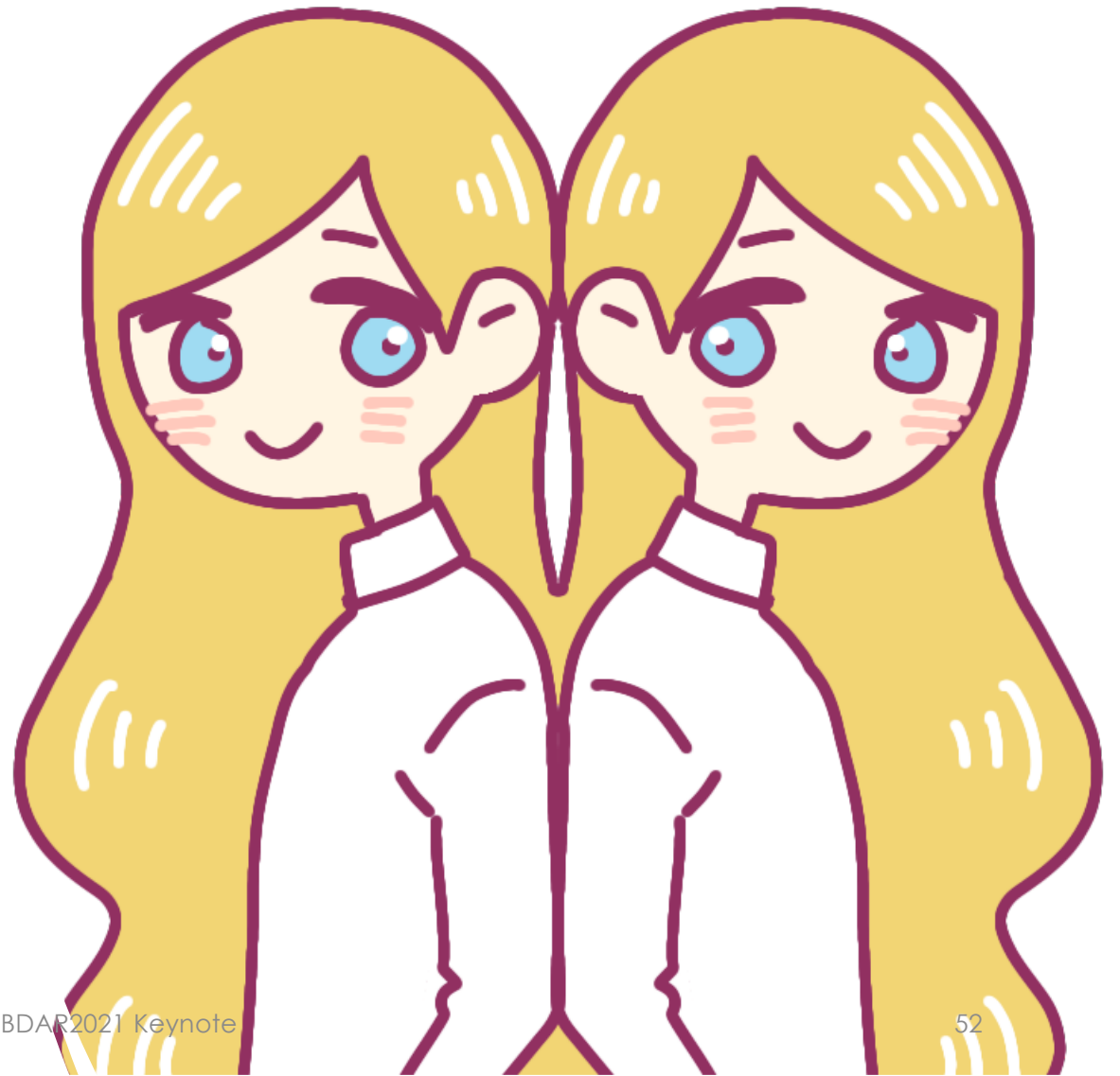
# Digital Twin

---

- Human digital twin can be built upon the recognition of human properties
- CBDAR to capture such properties

September 6, 2021

CBDAR2021 Keynote



# Digital Transformation

---

- DX requires deeper analysis of business processes
- Human-document interactions gives us insights



# Summary

- Current Status of CBDAR
  - Recognition of scene text
  - With a 3<sup>rd</sup> person camera
  - Without purposes
- Future CBDAR
  - Based on deeper understanding of the role of scene texts
  - With a camera on the appropriate viewpoint (1<sup>st</sup> – 3<sup>rd</sup>)
- Important Applications are waiting for your research