Enhancing Immersive Sensemaking with Rich Semantic Interaction

Ibrahim Tahmid

Committee Members

Doug A. Bowman

Chris North

Kirsten Whitley

Brendan David-John

John Wenskovitch

Prelim

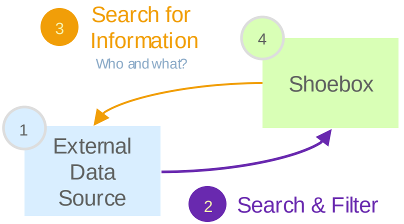
April 1, 2024

How to think like a detective

Suspicious activity in North America from January 2024

Find out what's going to happen

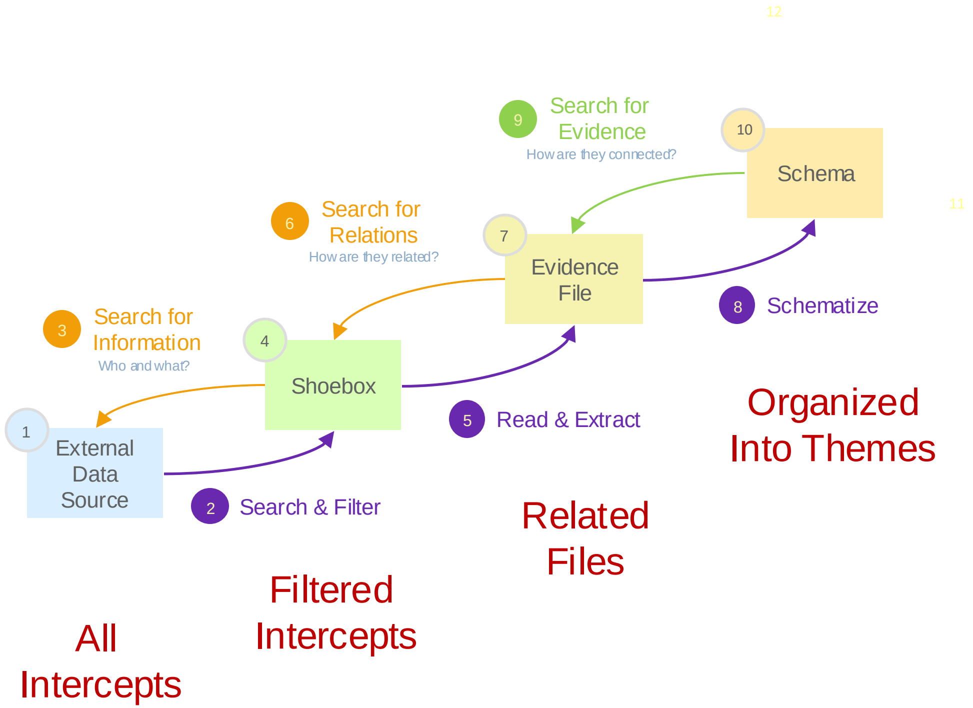
Filter with "North America, Date"



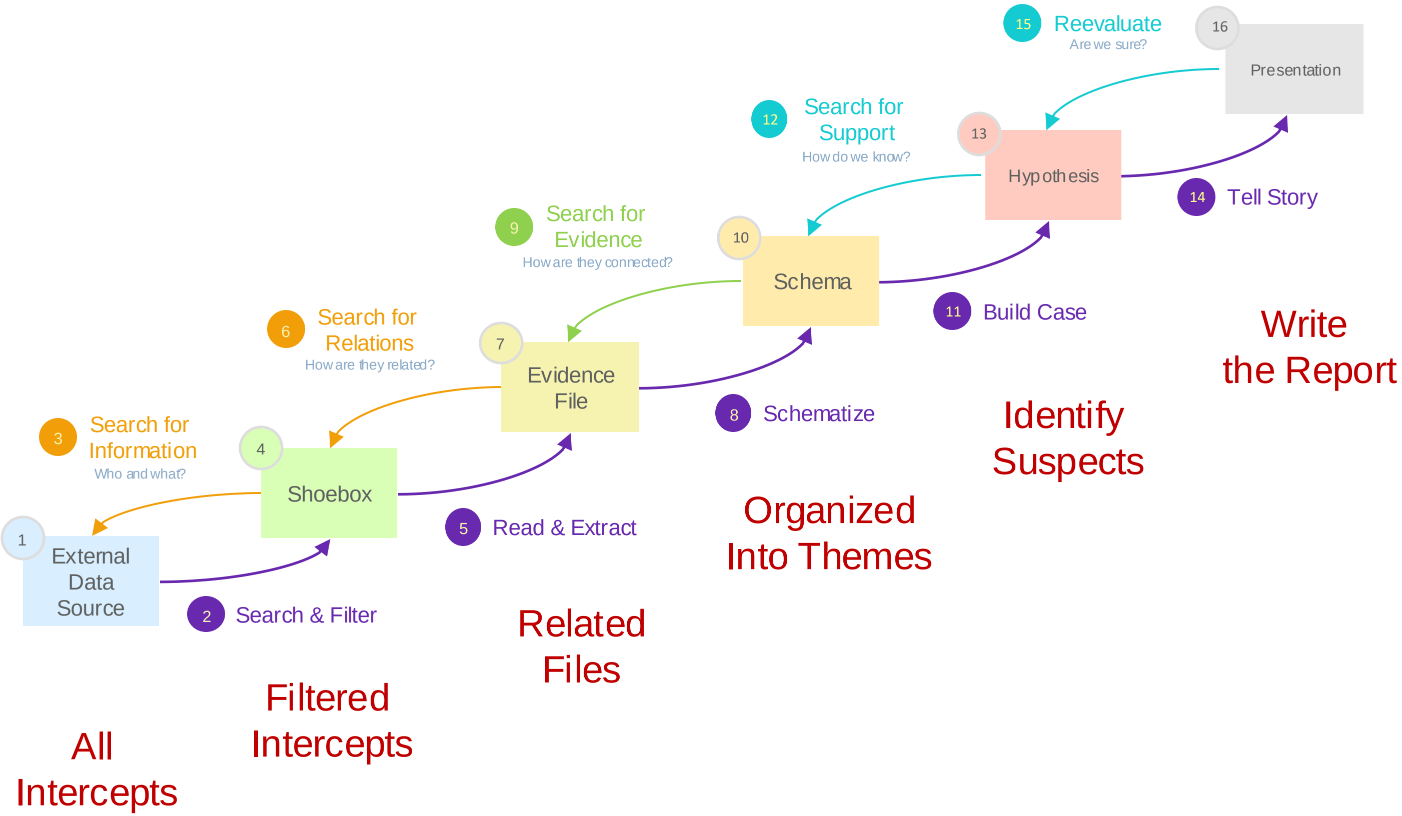
Filtered
Intercepts

All
Intercepts

Read and Extract

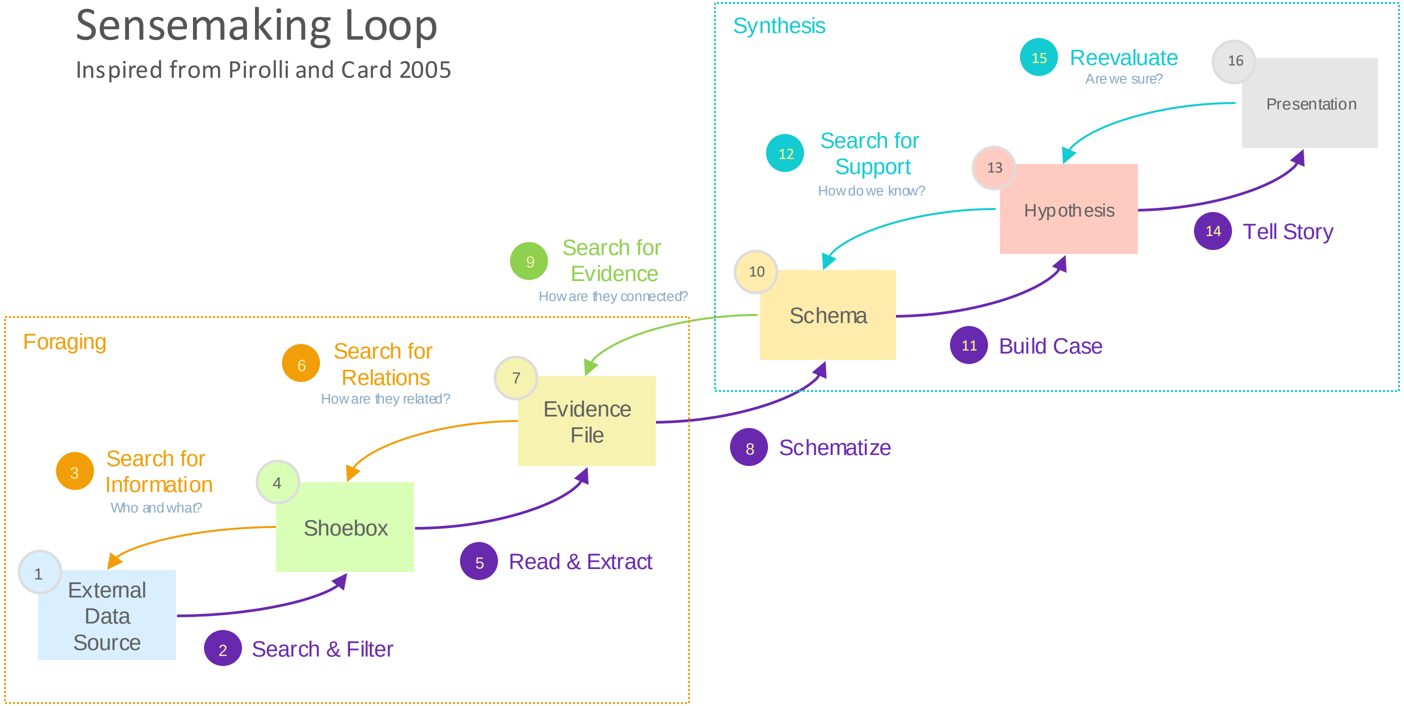


"Connect the Dots"



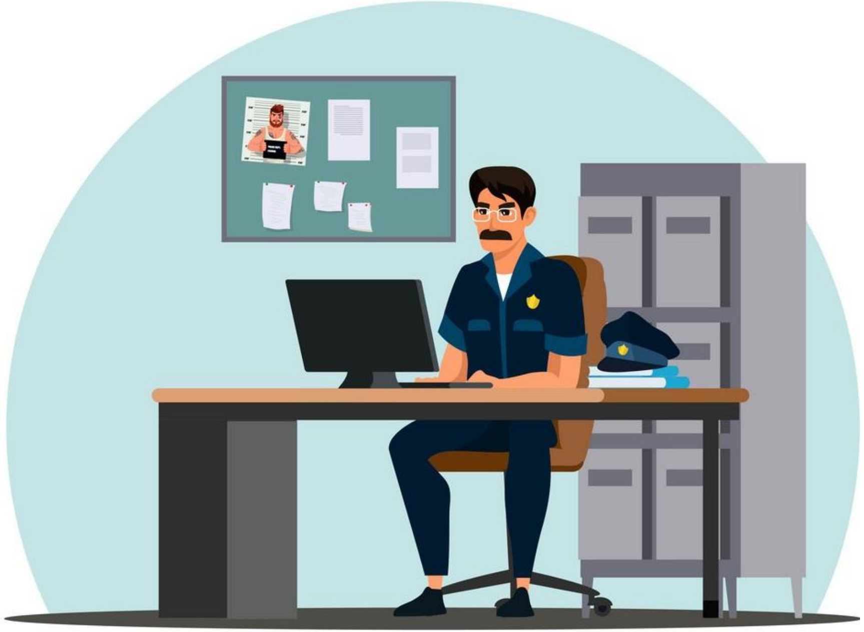
Sensemaking Loop

Inspired from Pirolli and Card 2005



Tools for Sensemaking

Evidence Board



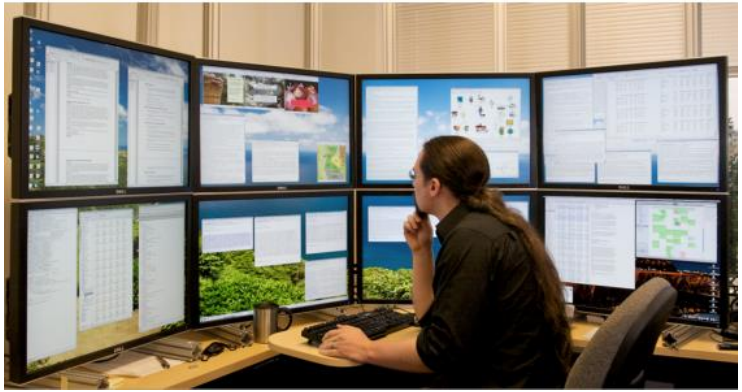
Personal Computer

Challenges in Sensemaking

Too Many
Inter-Connected Documents

Exhaustive
Browsing

Solutions for Too Many Documents



Space to Think Andrews et al. 2010

- 4x2 grid of 30’’ LCD panels
- 10240x3200 resolution
- Mouse
- Keyboard

Immersive Space to Think (IST) Lisle et al. 2020

- 3D Environment (AR/VR)
- Physical navigation e.g., walking
- 6-DOF controllers / Hand gestures
- Keyboard on a rolling cart

Takeaways

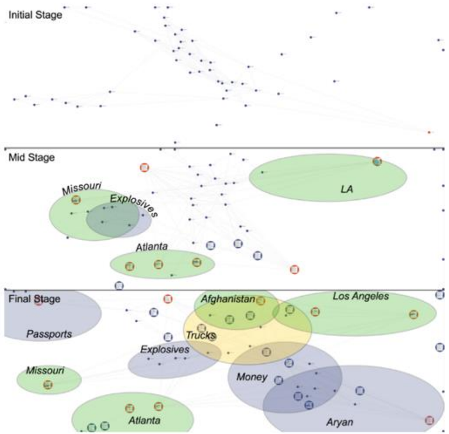
Spatial memory associated with dataset

Better **externalization** of mental hypothesis

Innovative spatial layouts

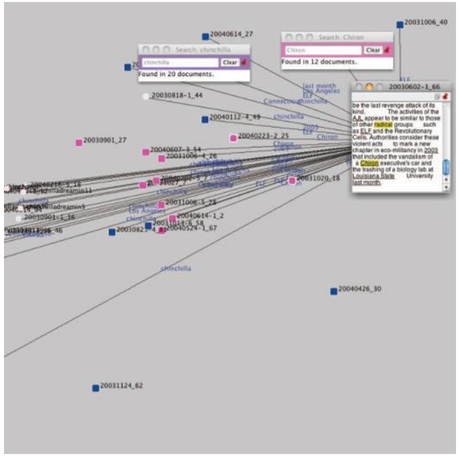
More space creates room for better sensemaking

Solutions for Exhaustive Browsing



Force-SPIRE Endert et al. 2012

- Spatial layout updates based on user action
- **Semantic Interaction**
- Example: Searching “LA” increases it’s weight and brings documents with the word “LA” closer to each other



Star-SPIRE Bradel et al. 2014

- Force-SPIRE + **document retrieval** based on user-perceived relevance

Takeaways

Allows **offloading** organization to an automated system

Brings humans into the equation

Give humans more time to focus on **analytic reasoning**

Semantic Interaction unifies the sensemaking loop by sharing tasks with an automated system

Solutions for Enhancing Sensemaking

Immersive Space

Semantic Interaction

How do the **semantic interactions** transfer to the **immersive space**?

Semantic Interaction in IST

	Space to Think	IST
Update user-perceived relevance based on user action	✓	✓
Update spatial organization based on user action	✓	✗

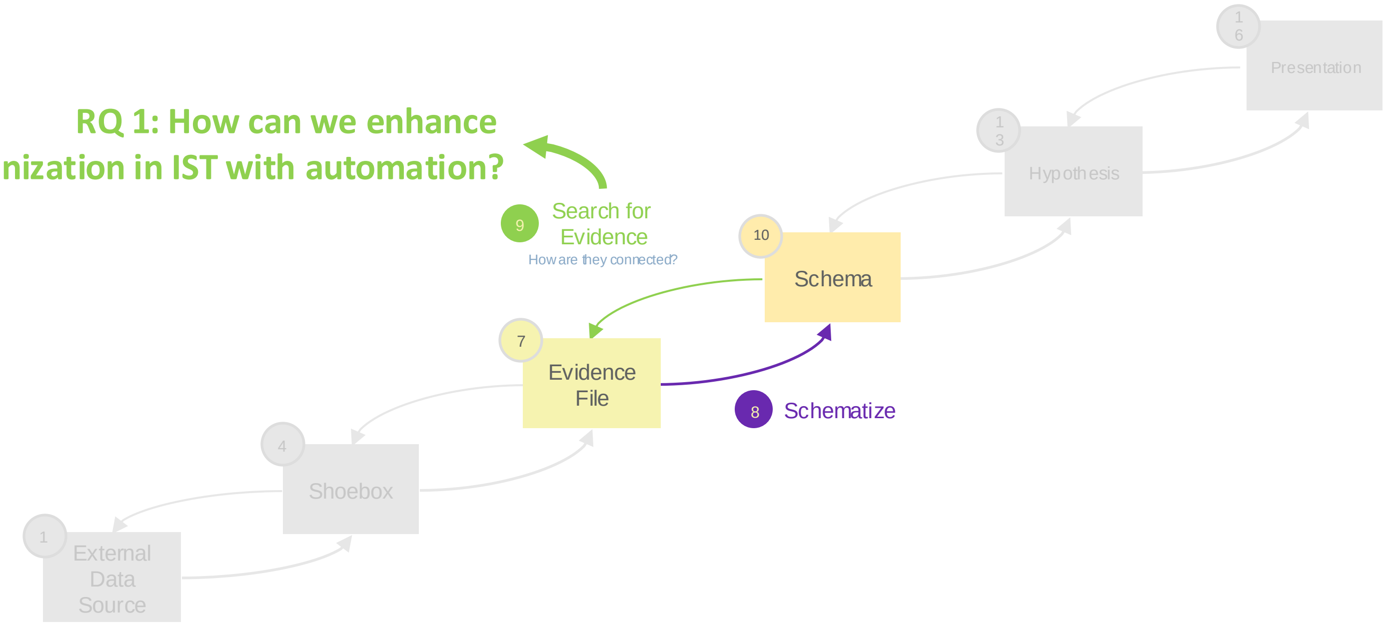
In an immersive space, spatial layout updates can be **out of sight**

Possibilities of

- Spatial memory loss
- Losing control over layout

RQ 1: Enhancing Organization

RQ 1: How can we enhance organization in IST with automation?



What does IST bring to the table?

A multi-sensory experience

Gaze

Speech

Physical Navigation

Neural Signals

“

Rich Semantic Interaction

is a mode of user-system interaction where the system can predict the user’s intents from a wide range of natural human interactions in the immersive space, such as motion, speech, eye gaze, and even brain signals that are engaged in their analytic process.

”

Gaze reflects the user’s cognitive process

Accessible and non-invasive

Gaze

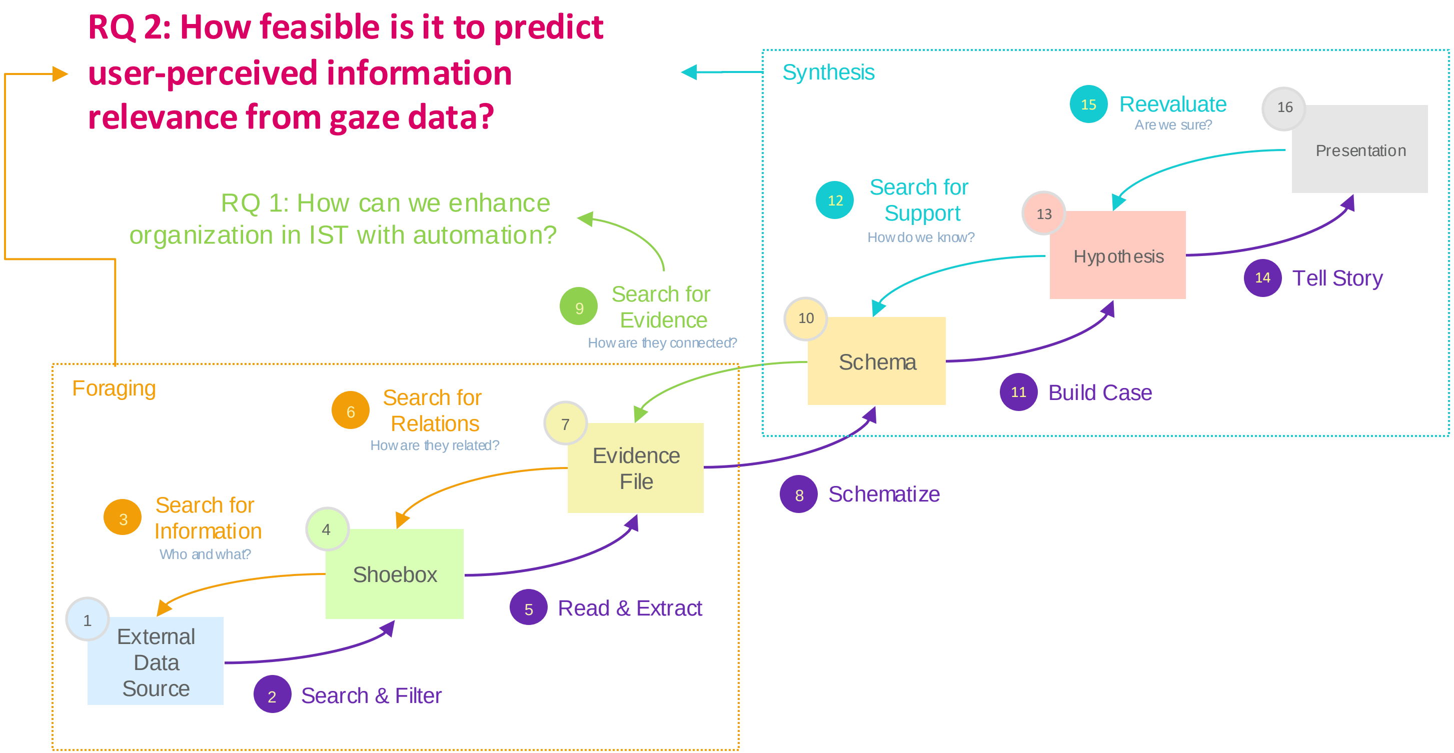
- Accessible
- Non-invasive
- **Informative**

Speech

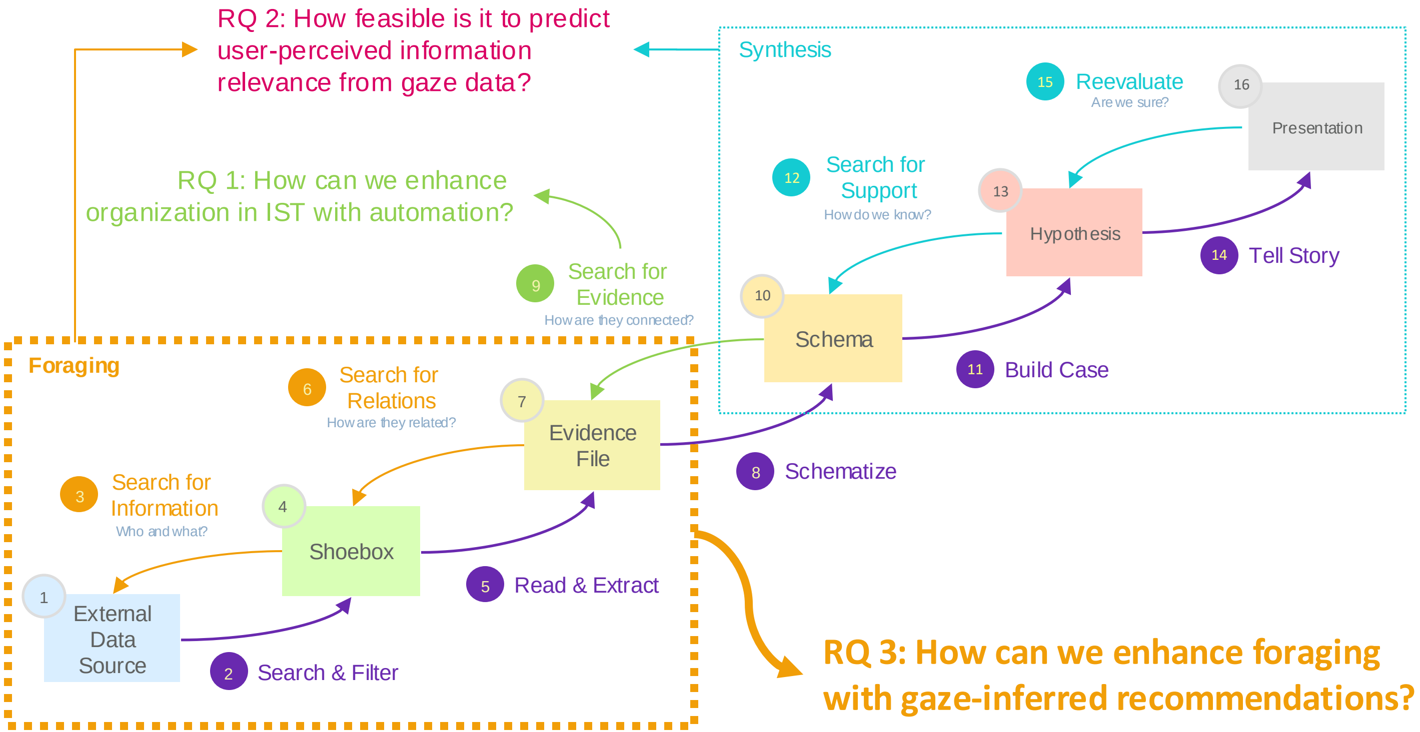
Physical Navigation

Neural Signals

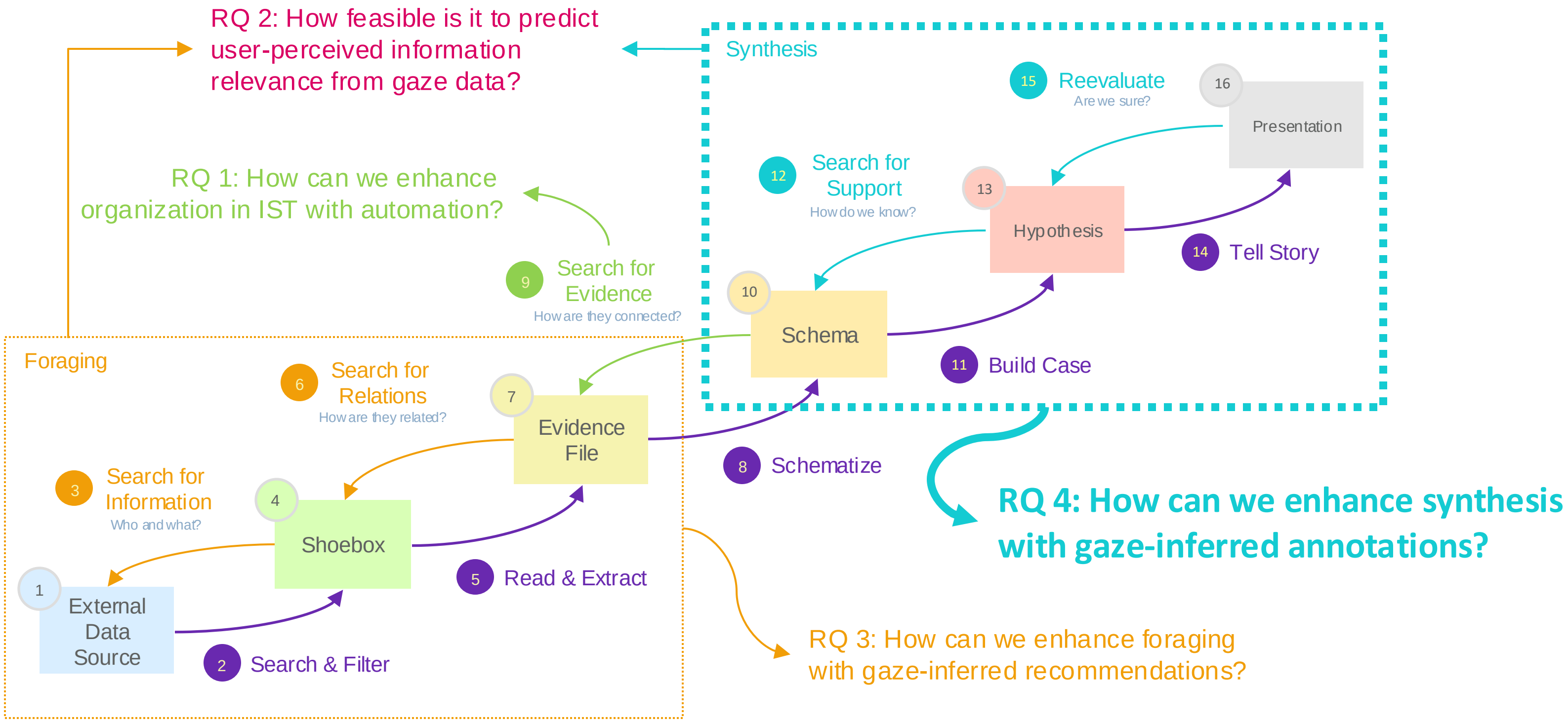
RQ 2: User-Perceived Relevance from Gaze



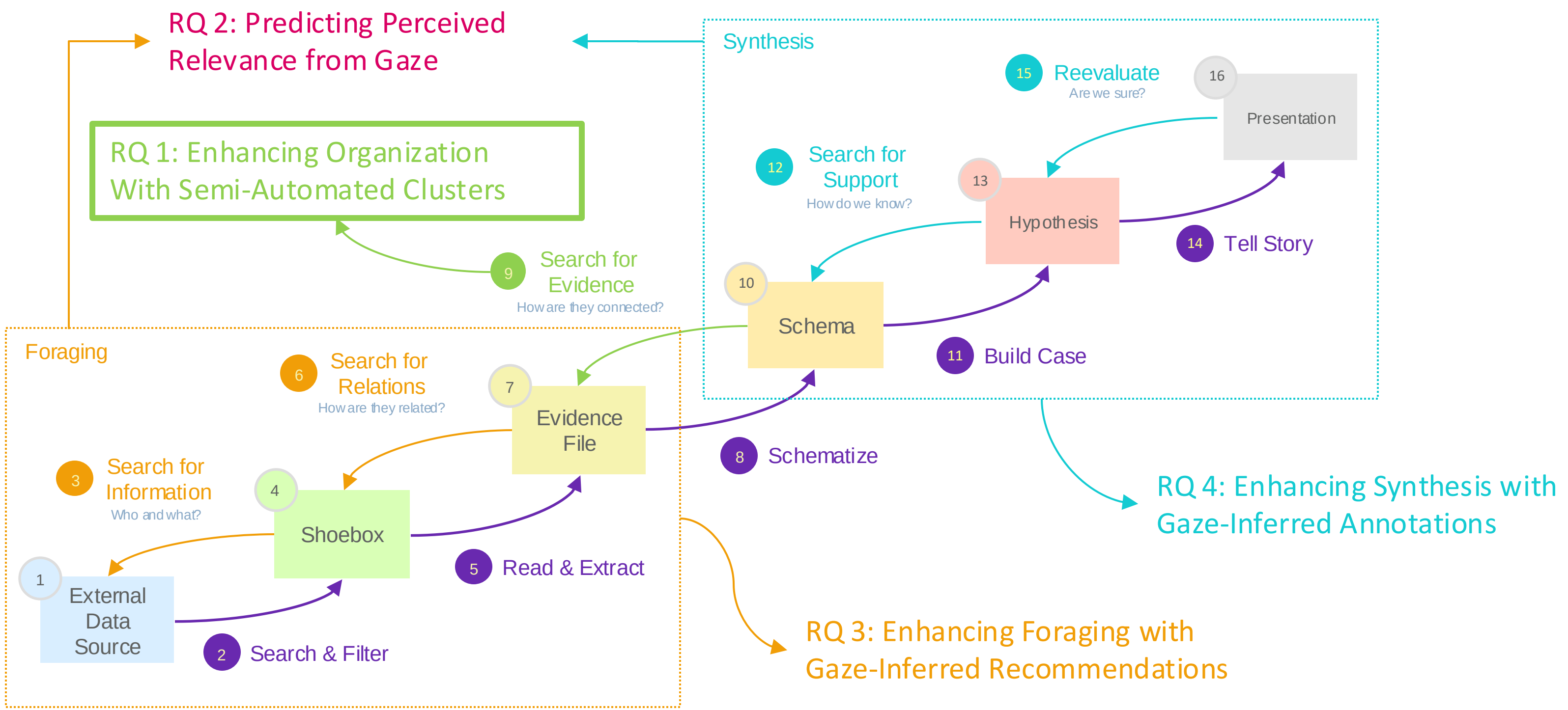
RQ 3: Foraging with Gaze-Inferred Recommendations



RQ 4: Synthesis with Gaze-Inferred Annotations



Roadmap



RQ 2: Predicting Perceived Relevance from Gaze
RQ 1: Enhancing Organization With Semi-Automated Clusters
Synthesis
15 Reevaluate
Are we sure?
16 Presentation
12 Search for Support
How do we know?
13 Hypothesis
14 Tell Story
9 Search for Evidence
How are they connected?
10 Schema
11 Build Case
Foraging
6 Search for Relations
How are they related?
7 Evidence File
8 Schematize
RQ 4: Enhancing Synthesis with Gaze-Inferred Annotations
3 Search for Information
Who and what?
4 Shoebox
5 Read & Extract
1 External Data Source
2 Search & Filter
RQ 3: Enhancing Foraging with Gaze-Inferred Recommendations

COMPLETED

RQ 1

Enhancing Organization

How can we enhance organization with automation?

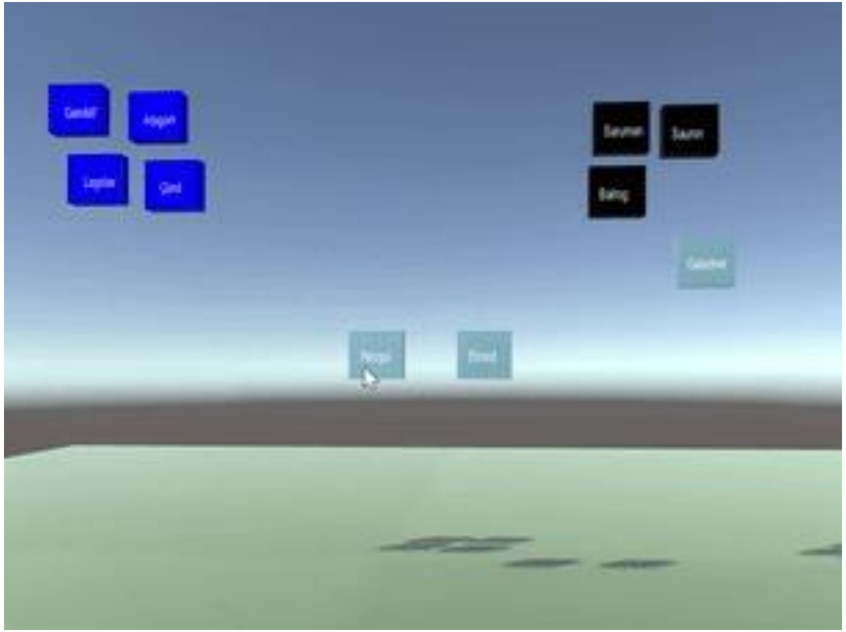
RQ 1.1

What is an appropriate level of automation for clustering in IST?

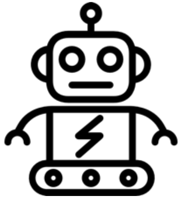
Exploring a Fully Automated System

After each interaction, the system applies the clustering algorithm to the whole layout

- Creates new cluster(s)
- Expands/shrinks prior cluster(s)



Takeaways



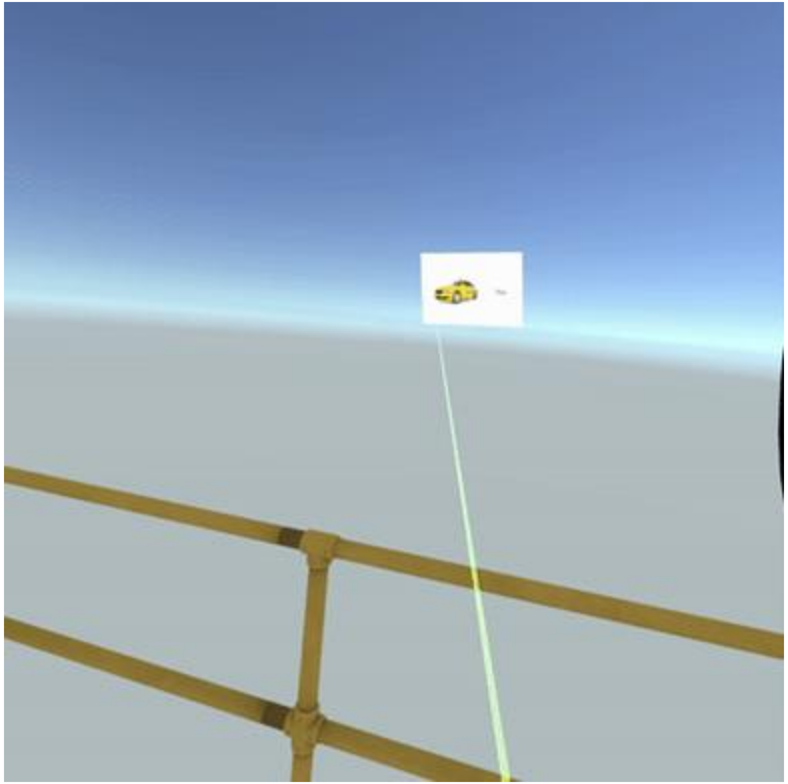
System output does **not match user intent** always

Users are left confused, frustrated, and **disoriented**

Spends more time on fixing **unwanted spatial structures**

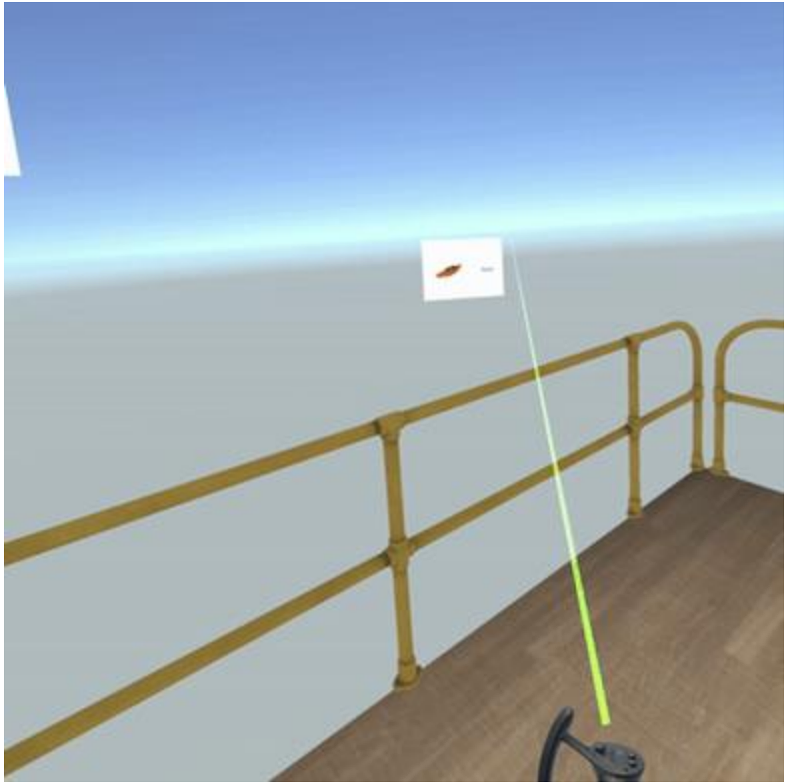
Semi-automated cluster

- System assists in creating the clusters
- User controls expansion/shrinking



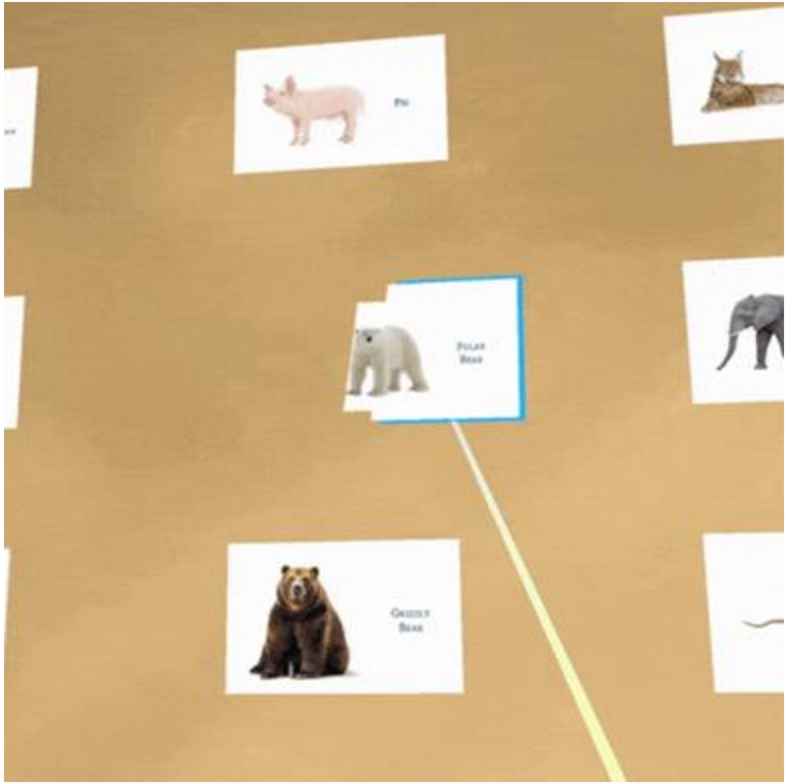
PROXIMITY

System creates clusters
with nearby documents using
Dirichlet Process Mixture Model



OVERLAP

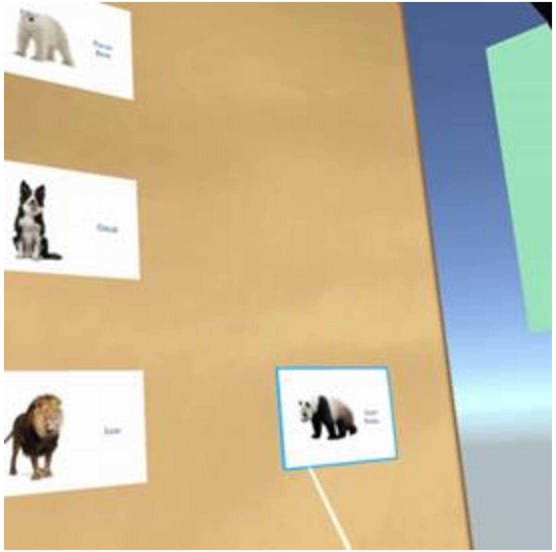
User creates clusters
by overlapping two documents



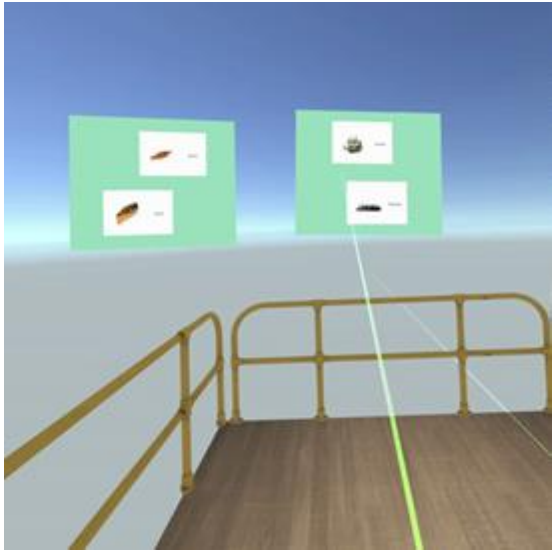
FREESTYLE

No explicit clusters

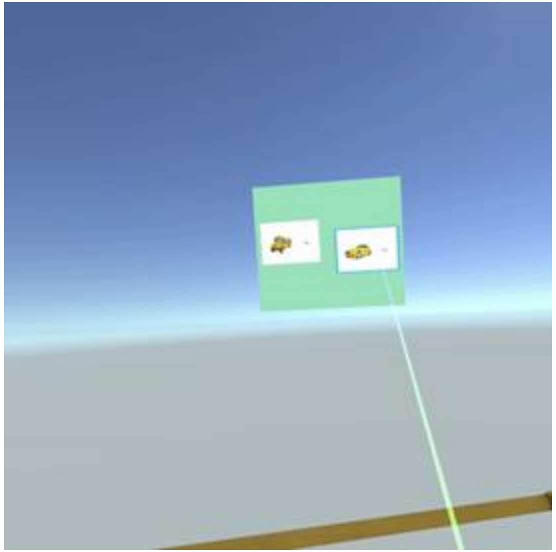
Cluster Interactions



EXPAND
existing clusters



MERGE
two or more clusters



REMOVE
lone clusters

PARTICIPANTS

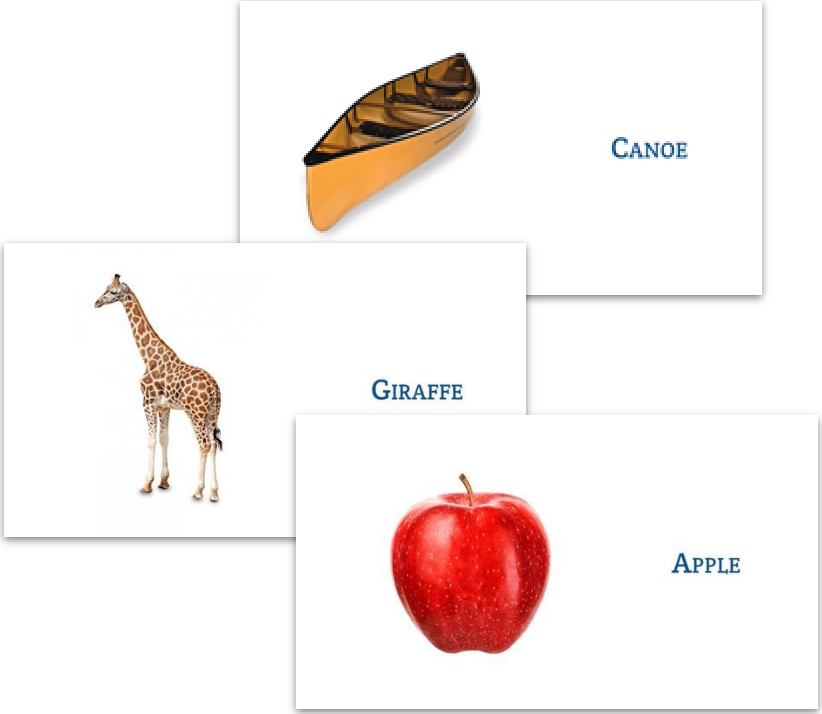
Within-subject
27 participants (10F)
6 with no prior VR experience

DATASET

3 sets of 30 images
(Foods, Animals, Vehicles)

TASK

Organize an exhibition space
with 3-8 clusters in 10 minutes



Latin Square Design

P1: Freestyle, Overlap, Proximity
P2: Overlap, Proximity, Freestyle
P3: Proximity, Freestyle, Overlap
P4: …

CONDITION 1

Pre-Study Questionnaire

Training

15 cards from **Set 1**

15 cards from **Set 1**

NASA TLX

SUS

CONDITION 2

Training

15 cards from **Set 2**

15 cards from **Set 2**

NASA TLX

SUS

CONDITION 3

Training

15 cards from **Set 3**

15 cards from **Set 3**

NASA TLX

SUS

Ranking 3 conditions

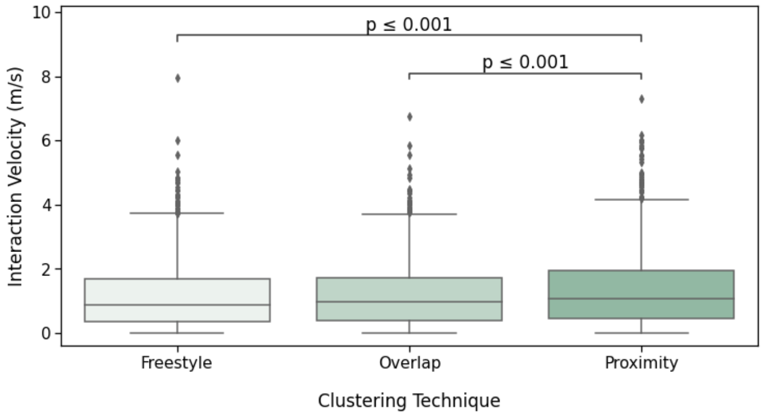
Semi-structured interview

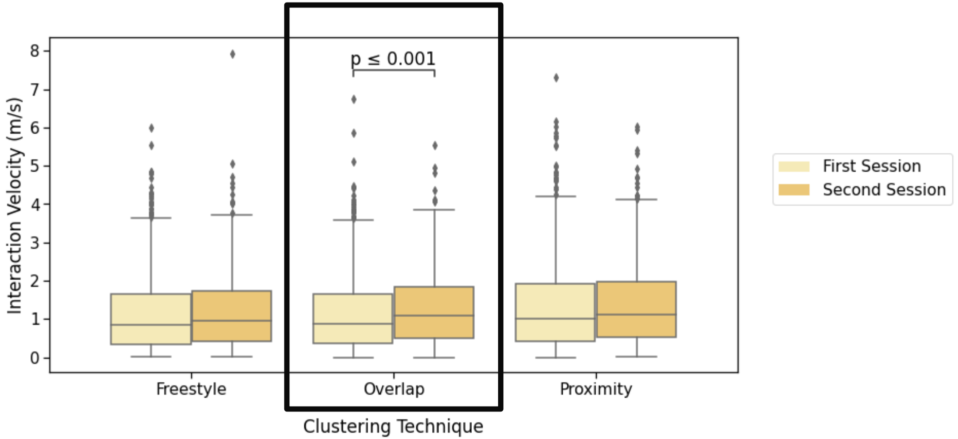
RQ 1.2

How do explicit clusters help analysts organize in IST?

Hypothesis: H1.2a

Explicit clusters would make analysts faster (Partially supported)



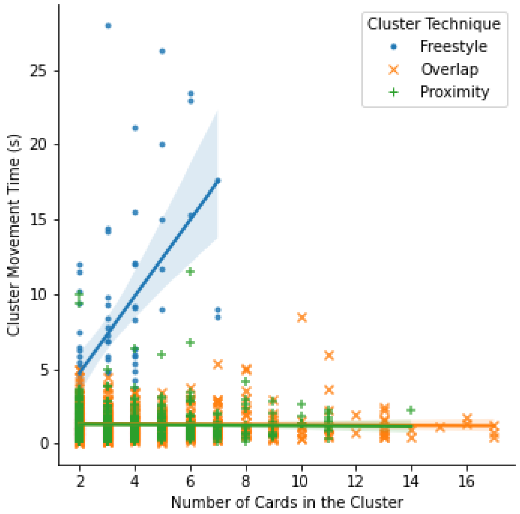


Proximity made participants faster

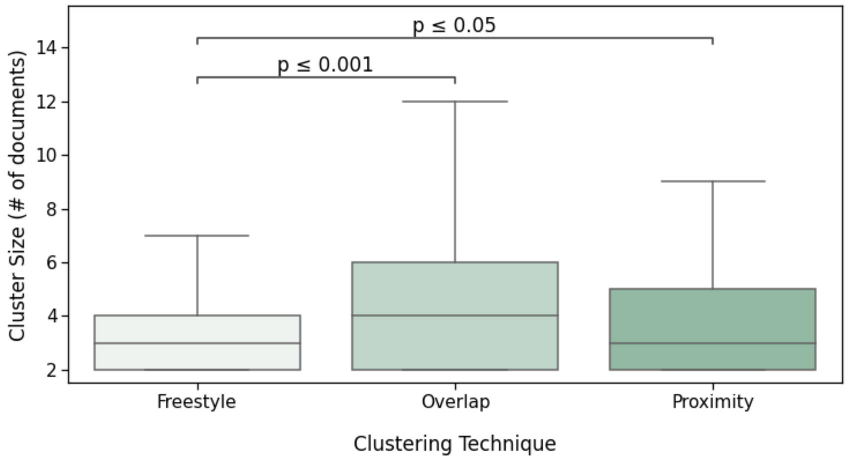
Overlap needed some time getting used to

Hypothesis: H1.2b

Explicit clusters would speed up the process of reorganizing workspace (Supported)



Cluster movement time is constant for Overlap and Proximity



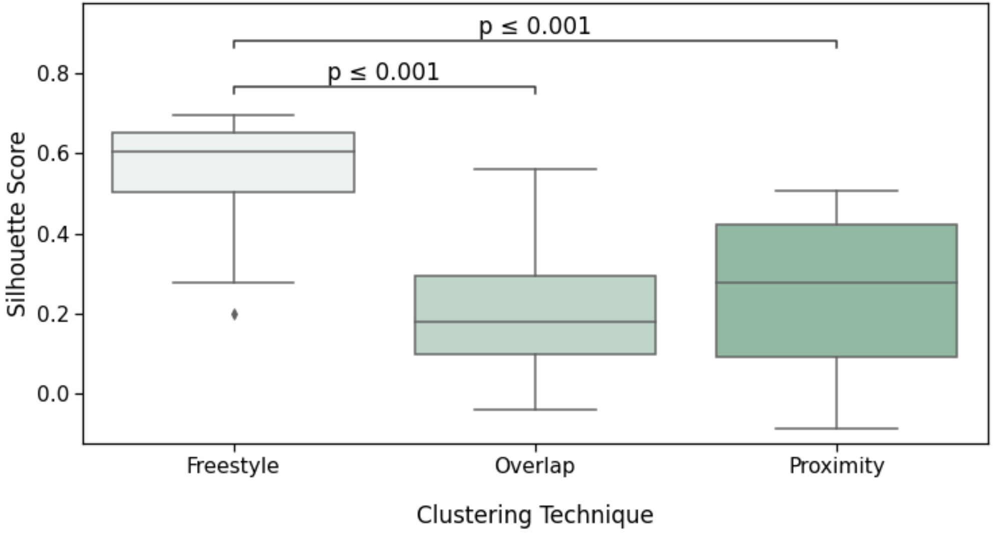
Participants tend to create bigger clusters with Overlap and Proximity

Hypothesis: H1.2c

Explicit clusters would make the final layout less ambiguous

(Partially Supported)

Tighter
Clusters

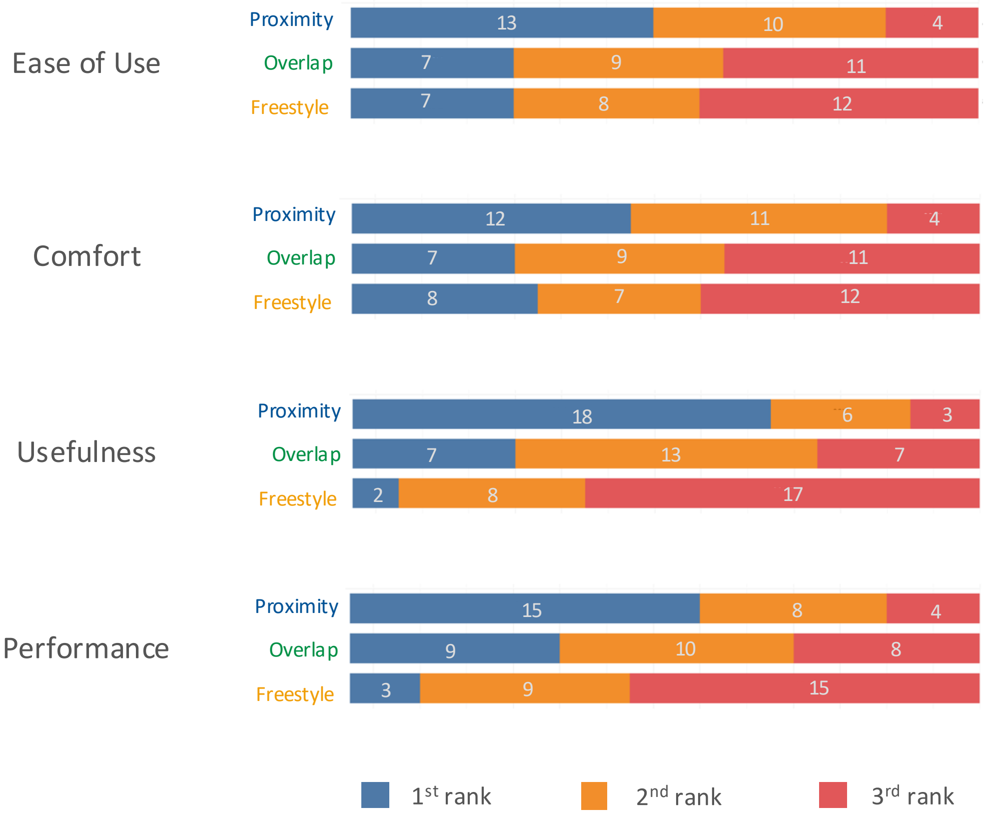


Participants used a tighter space in Overlap and Proximity

RQ 1.3

What are the benefits and challenges of semi-automated clusters in IST?

Benefits of **Proximity**



"

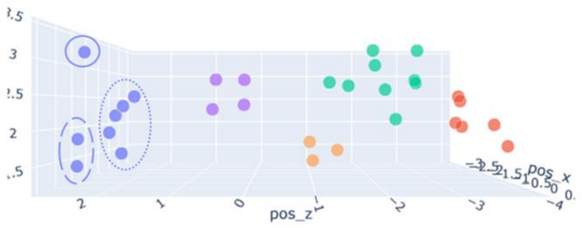
It [Proximity] was as easy as Freestyle, with the added benefits of the explicit clusters

Challenges of Proximity

Three participants were frustrated with Proximity because of **losing control**, deviating from user intent

Nine participants preferred Overlap as that gave them **full control** over their workspace

Three participants chose Freestyle because of the **creativity** it offered

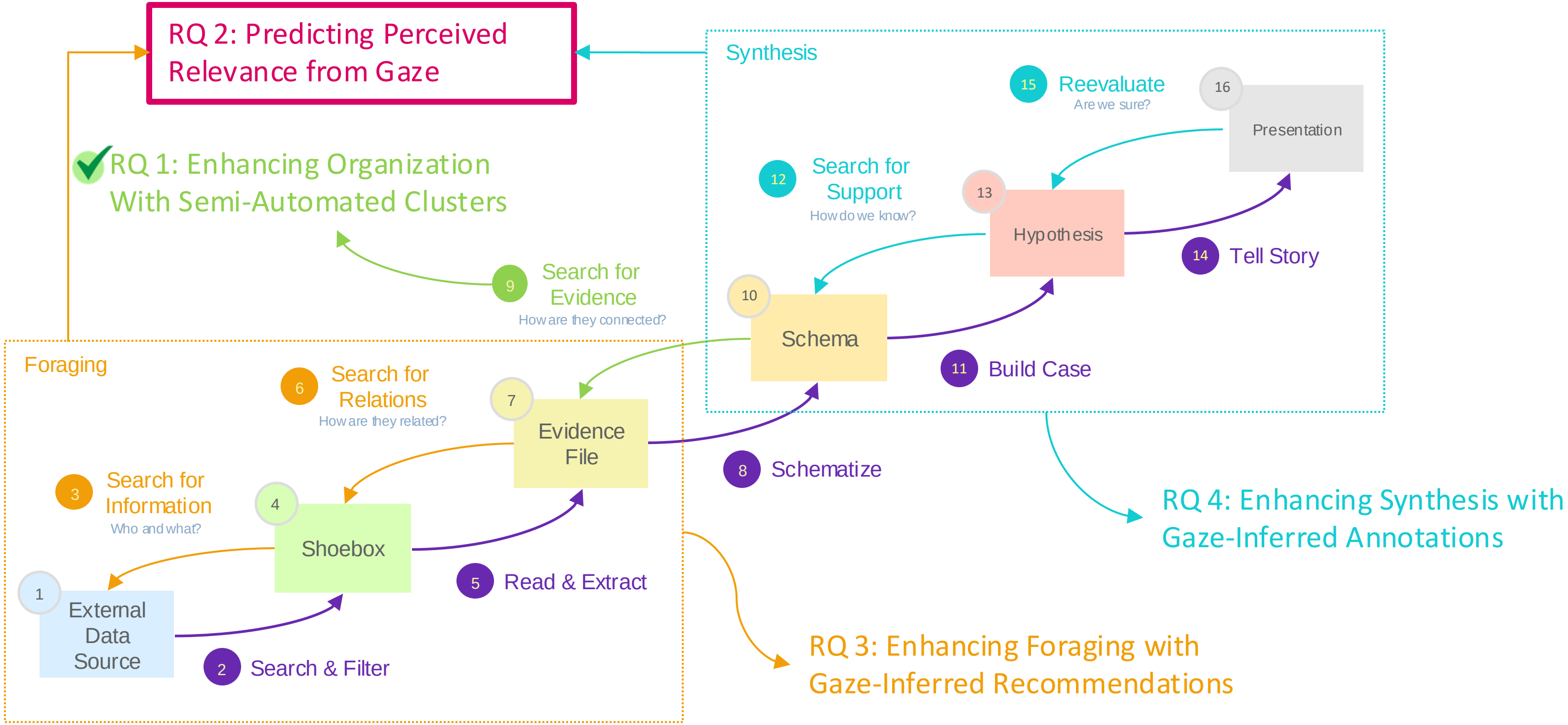


Cluster within cluster

Takeaway

Automation
made the organization step easier
but users need **more control** over the process

Roadmap


RQ 2: Predicting Perceived Relevance from Gaze
RQ 1: Enhancing Organization With Semi-Automated Clusters
Synthesis
15 Reevaluate
Are we sure?
16 Presentation
12 Search for Support
How do we know?
13 Hypothesis
14 Tell Story
9 Search for Evidence
How are they connected?
10 Schema
11 Build Case
Foraging
6 Search for Relations
How are they related?
7 Evidence File
8 Schematize
RQ 4: Enhancing Synthesis with Gaze-Inferred Annotations
3 Search for Information
Who and what?
4 Shoebox
5 Read & Extract
1 External Data Source
2 Search & Filter
RQ 3: Enhancing Foraging with Gaze-Inferred Recommendations

COMPLETED

RQ 2

Predicting Perceived Relevance from Gaze

RQ 2: How feasible is it to predict user-perceived information relevance from gaze data?

RQ 2.1

How can we design a gaze-based metric that can predict the user-perceived relevance during sensemaking?

Gaze Measures for Reading

Gaze Duration (GD)

Amount of time spent on a document or a word

Unique Dwell Count (UD)

Number of times the reader shifts their attention to a document or a word

Walkthrough

Predicting relevance from GD or UD

Relevant documents and words receive more attention from readers [Davari et al., White et al.]

Report Date 1 April, 2003. FBI: ----- Mark Davis is the owner of the Select Gourmet Foods shop in Springfield Mall, Springfield, VA. [Phone number 703-659-2317].

Words ranked by GD

- Davis
- 703-659-2317
- Springfield
- Date

- M. Davari, D. Hienert, D. Kern, and S. Dietze. The role of wordeye-fixations for query term prediction. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, pp. 422–426, 2020.
- R. W. White, J. M. Jose, and I. Ruthven. An implicit feedback approach for interactive information retrieval. Information processing & management, 42(1):166–190, 2006

Walkthrough

Predicting relevance from GD or UD

Relevant documents and words receive more attention from readers [Davari et al., White et al.]

Report Date 1 April, 2003. FBI: ----- Mark Davis is the owner of the Select Gourmet Foods shop in Springfield Mall, Springfield, VA. [Phone number 703-659-2317].

Report Date 5 April, 2003. FBI: ----- Passport control at Dulles Airport in Wash DC records that Mark Davis. holder of US passport# 177183634

Words ranked by GD

Davis

703-659-2317

Date

Springfield

- M. Davari, D. Hienert, D. Kern, and S. Dietze. The role of wordeye-fixations for query term prediction. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, pp. 422–426, 2020.
- R. W. White, J. M. Jose, and I. Ruthven. An implicit feedback approach for interactive information retrieval. Information processing & management, 42(1):166–190, 2006

Walkthrough

Predicting relevance from GD or UD

Relevant documents and words receive more attention from readers [Davari et al., White et al.]

Report Date 1 April, 2003. FBI: ----- Mark Davis is the owner of the Select Gourmet Foods shop in Springfield Mall, Springfield, VA. [Phone number 703-659-2317].

Report Date 5 April, 2003. FBI: ----- Passport control at Dulles Airport in Wash DC records that Mark Davis. holder of US passport# 177183634

Report Date 20 April, 2003: FBI: ----- Clark Webster has an account at the Virginia National Bank in Charlottesville, VA. He has deposited $13,000 in the last three months, drawn on an account held by Mark Davis.

Report Date 21 April, 2003:- ---- Frequent recent phone calls from John Smith to the following numbers: 804-774-8920 [Charlottesville, VA]; 718-352-8479 [Queens, NYC].

Words ranked by GD

- Date
- Davis
- 703-659-2317
- Springfield

- M. Davari, D. Hienert, D. Kern, and S. Dietze. The role of wordeye-fixations for query term prediction. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, pp. 422–426, 2020.
- R. W. White, J. M. Jose, and I. Ruthven. An implicit feedback approach for interactive information retrieval. Information processing & management, 42(1):166–190, 2006

Frequency Bias

Multiple inter-connected documents introduces frequency bias

Some words get **more attention for high frequency** rather than their perceived relevance

Gaze Duration or Unique Dwell alone **cannot address** the bias

User Reported	Sorted by GD
Mark Davis	Name
Foysal Goba	2003
6302	Address
Texas	Date
Virginia	Bank
Passport	Phone
Laundering	List
Bank	Virginia
Deposit	Check
Myrtle	6302

Length Bias

Longer documents get **more attention for length** rather than their perceived relevance

Gaze Duration alone cannot address the length bias

$GD_a < GD_b$

GD_a	GD_b
Report Date 1 April, 2003. FBI: ----- Mark Davis is the owner of the Select Gourmet Foods shop in Springfield Mall, Springfield, VA. [Phone number 703-659-2317].	Report Date 25 April, 2003. FBI: ----- A report from AMTRAK reveals a reservation, paid in cash in Charlottesville, and made by Faysal Goba on 23 April, 2003. Reservation is for three one-way first class tickets and one sleeping compartment from Charlottesville, VA to Atlanta, GA on 29 April, 2003. Reservation is on AMTRAK Train # 19, which runs between Penn Station NYC and New Orleans, LA. Reservations are in the names: Faysal Goba, Mukhtar Galab and Yasein Mosed.

Introducing Gaze Score (GS)

$$Z_{UD_i} = \frac{x_{UD_i} - mean}{Std.Dev}$$

$$Z_{GD_i} = \frac{x_{GD_i} - mean}{Std.Dev}$$

Normalize to remove **length bias**

$$GazeScore_i = \boxed{Z_i} * \boxed{IDF_i}$$

Weight factor to address **frequency bias**

$IDF_{\text{rarest word}} = 1$

$IDF_{\text{most common word}} = 0$

$IDF_{\text{document}} = 1$

Reflects **combined effect** of GD and UD

$$Z_i = \frac{Z_{GD_i} + Z_{UD_i}}{2}$$

Introducing Gaze Rank (GR)

People have varying reading patterns

Arbitrary values for Gaze Score

Prevents comparison

GazeRank

Index on sorted Gaze Score

Fixed max and min value

Allows **comparison**

RQ 2.2

To what extent does the gaze-based metric predict the user-perceived information relevance?

Study Details

12 Participants

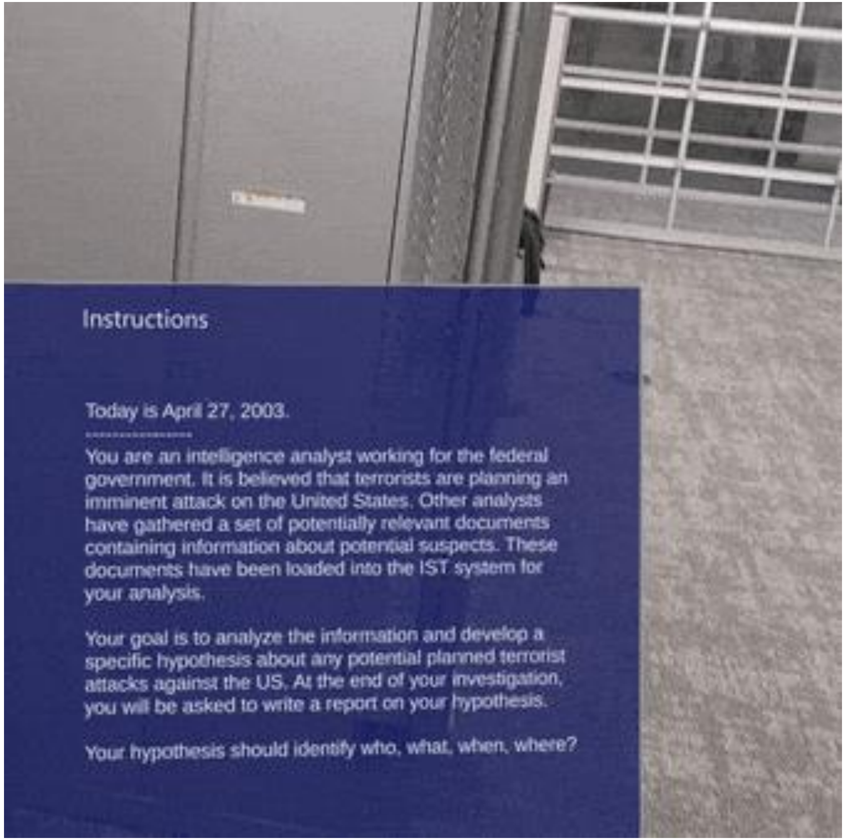
3F, 1NB

Sign of the Crescent

24 Documents, 2 Terrorist plots
4 Distractors

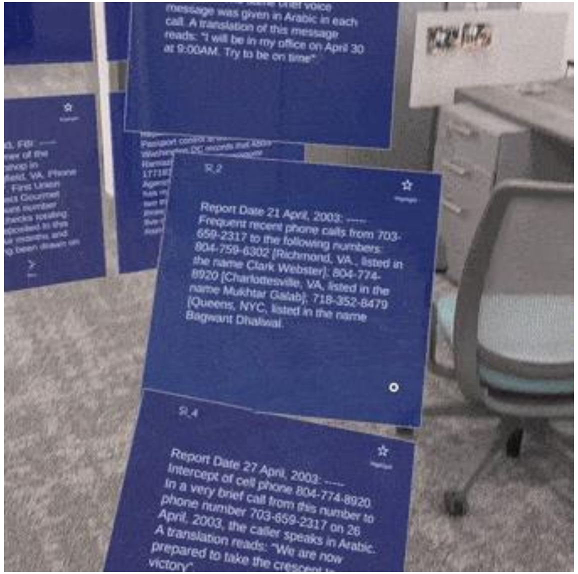
1 Notepad

Make labels
Search keywords

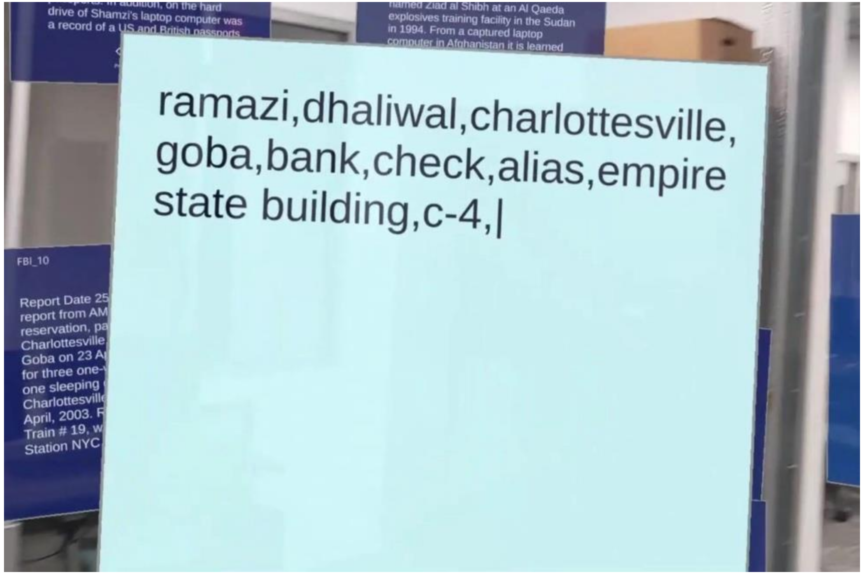


Free Response

Participants report 4 relevant documents
and 10 relevant words with **no prior knowledge about Gaze Rank**



Highlight Document

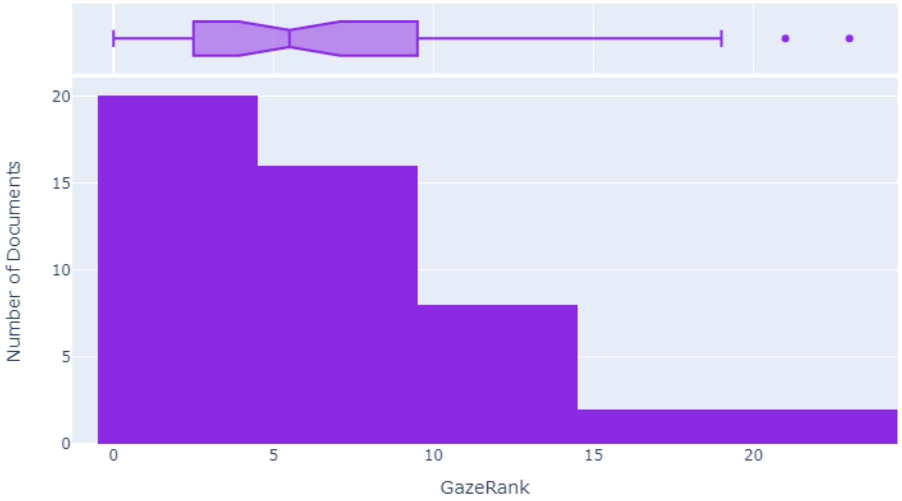


Write Down Keywords

We analyze the Gaze Ranks for these documents and words

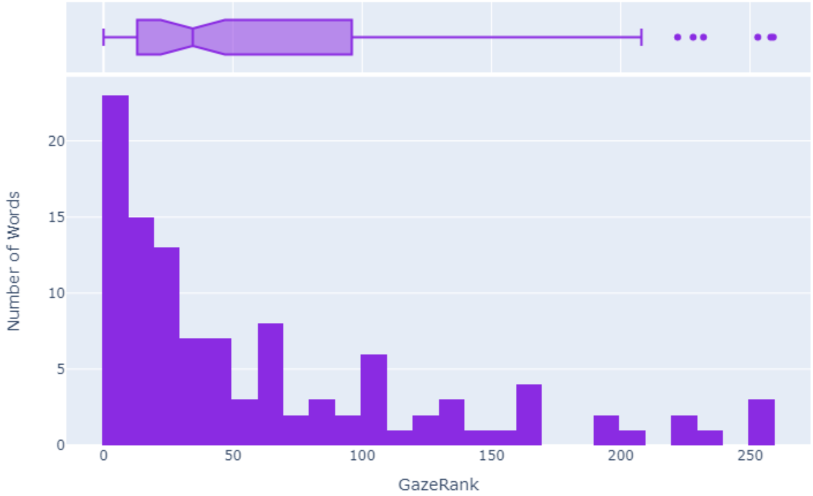
Free Response

Participants report 4 relevant documents
and 10 relevant words with **no prior knowledge about Gaze Rank**



Downward slope of GR with median at 5.5

38% documents are in top 4
(random chance for being in top 4: 16.67%)

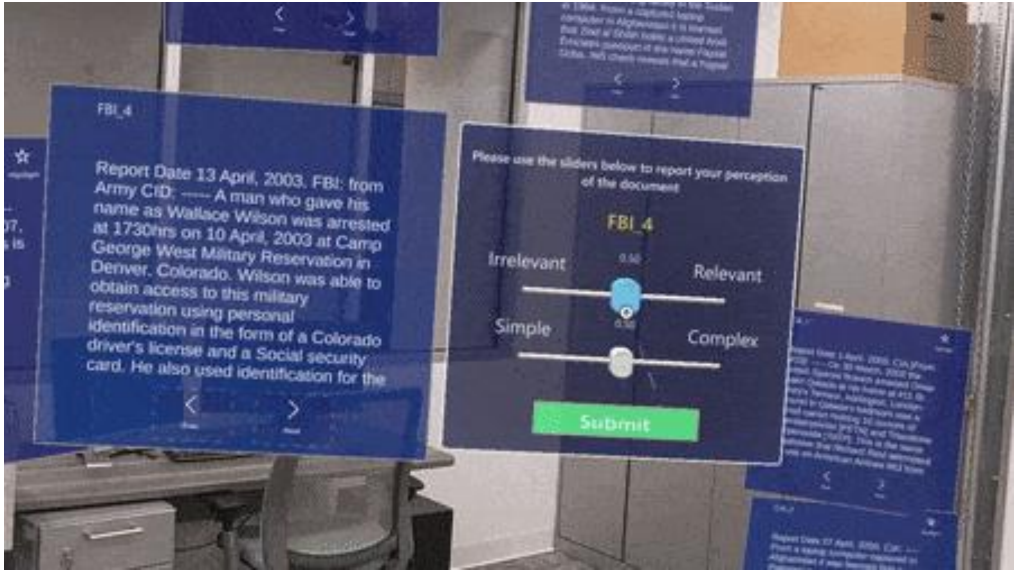


Downward slope of GR with median at 34.5

19.17% words are in top 10
(random chance for being in top 10: 1.9%)

Rated Response

Documents



4 documents with **high** Gaze Rank

4 documents with **medium** Gaze Rank

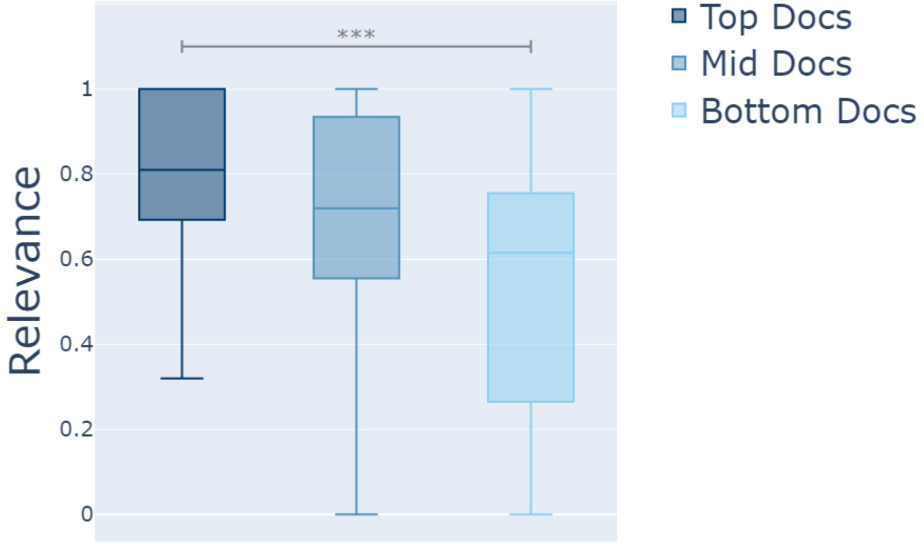
4 documents with **low** Gaze Rank

12 documents are randomized

Rated on relevance and complexity

Rated Response

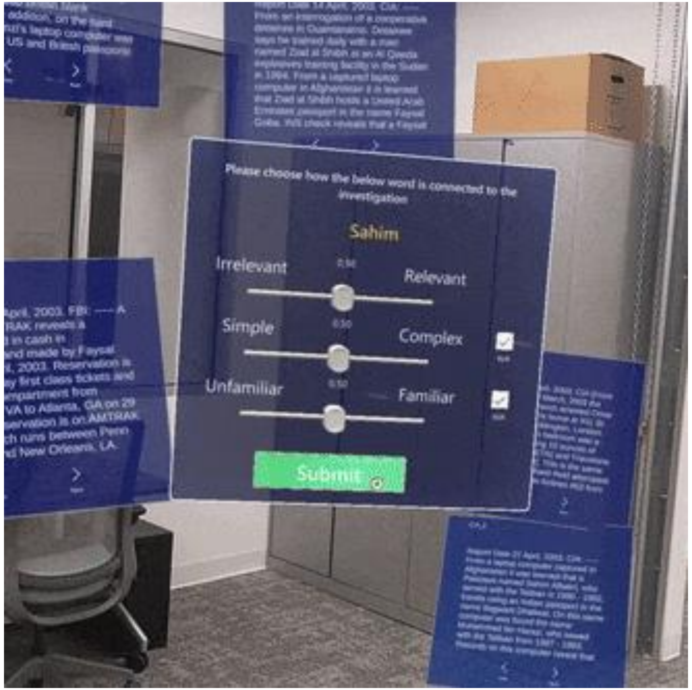
Documents



Documents with **high Gaze Ranks are rated as more relevant** than documents with low Gaze Ranks

Rated Response

Words



10 words with **high** Gaze Rank

10 words with **medium** Gaze Rank

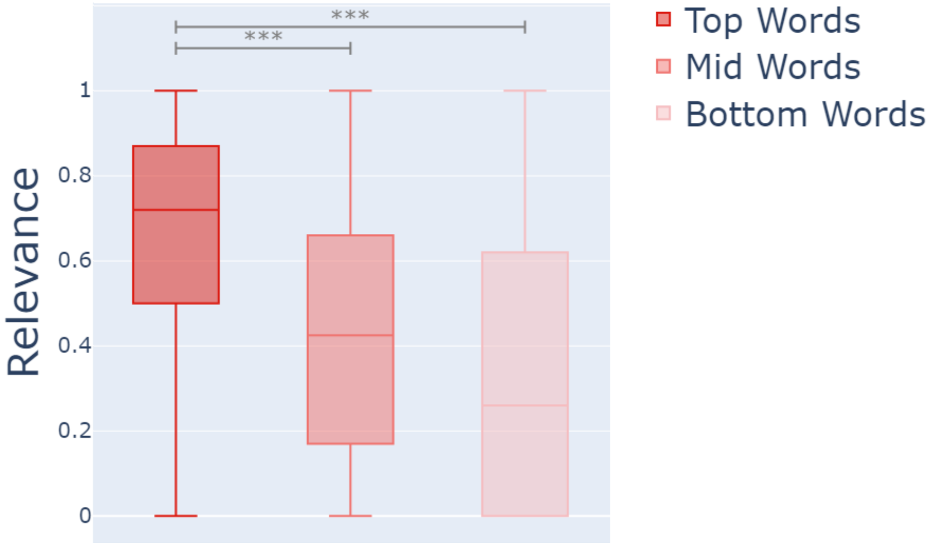
10 words with **low** Gaze Rank

30 words are randomized

Rated on relevance, complexity and familiarity

Rated Response

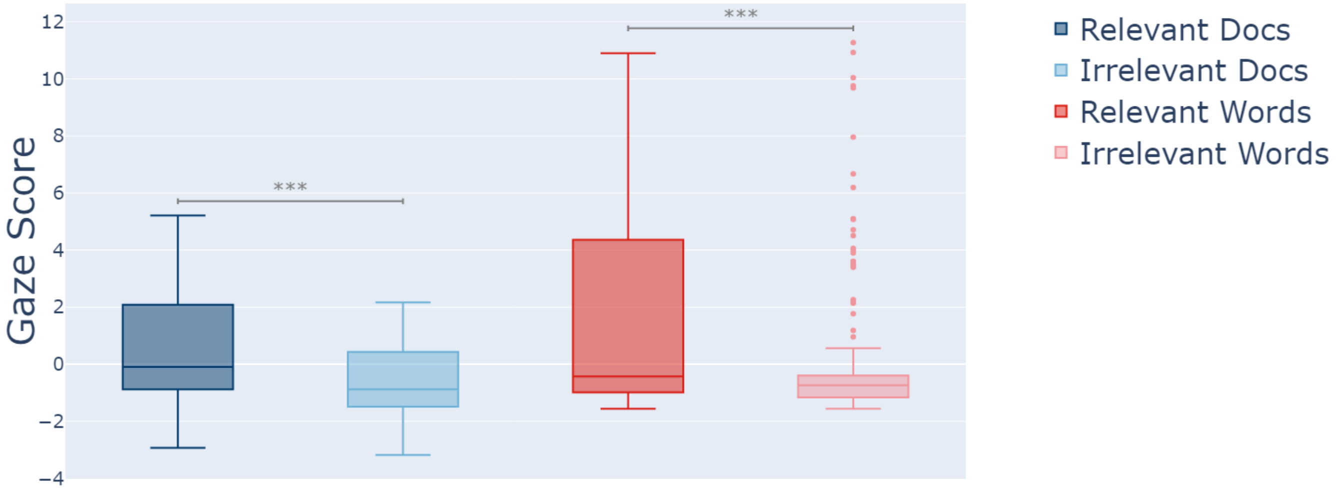
Words



Words with **high Gaze Ranks are rated as more relevant** than other words

Gaze Score

Performance in inferring user-perceived relevance



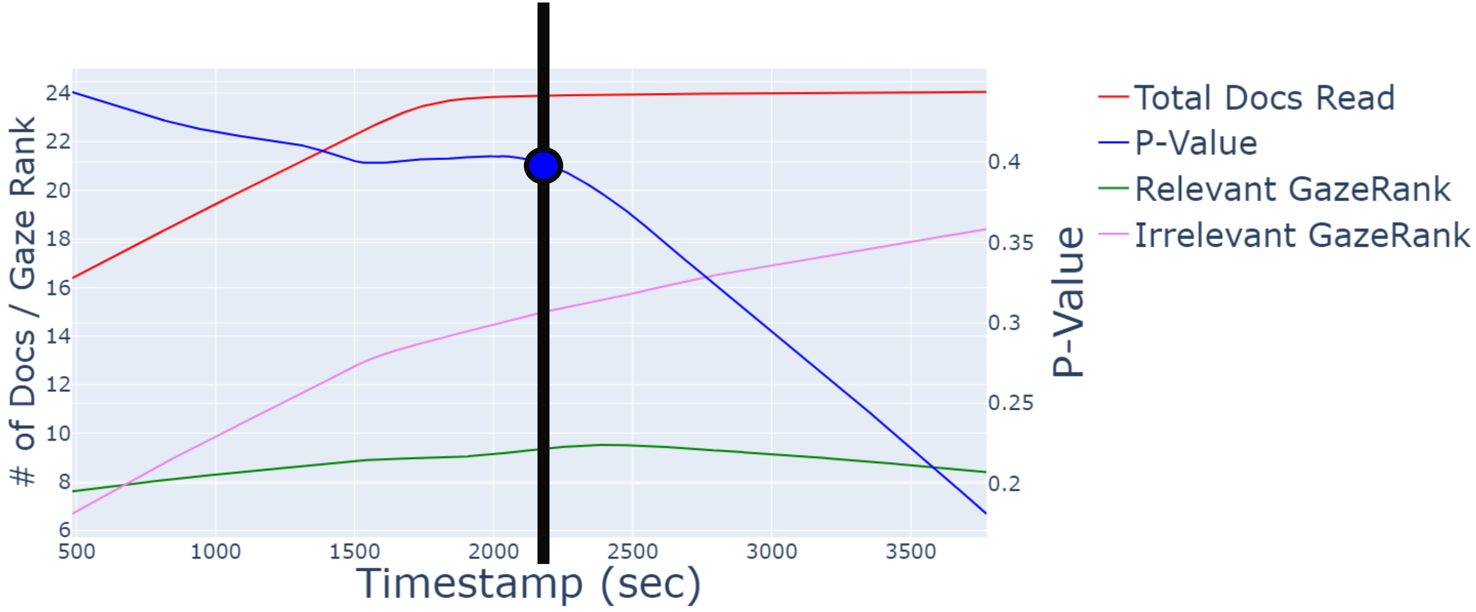
Relevant information has **higher Gaze Score** than irrelevant information

Gaze Rank Timeline

Documents

Difference between relevant and irrelevant documents start increasing **after all are read**

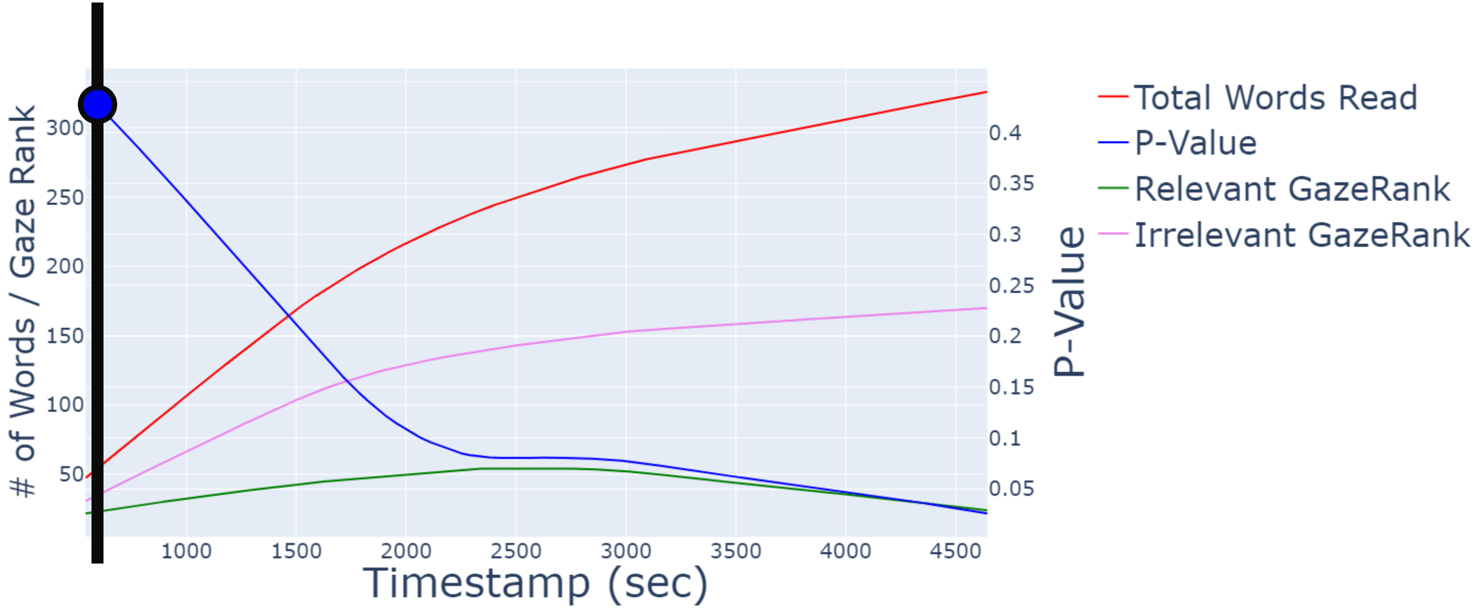
Effective for **larger datasets**



Gaze Rank Timeline

Words

Difference between relevant and irrelevant words start increasing **from the beginning**
More effective than documents if we want real-time assistance

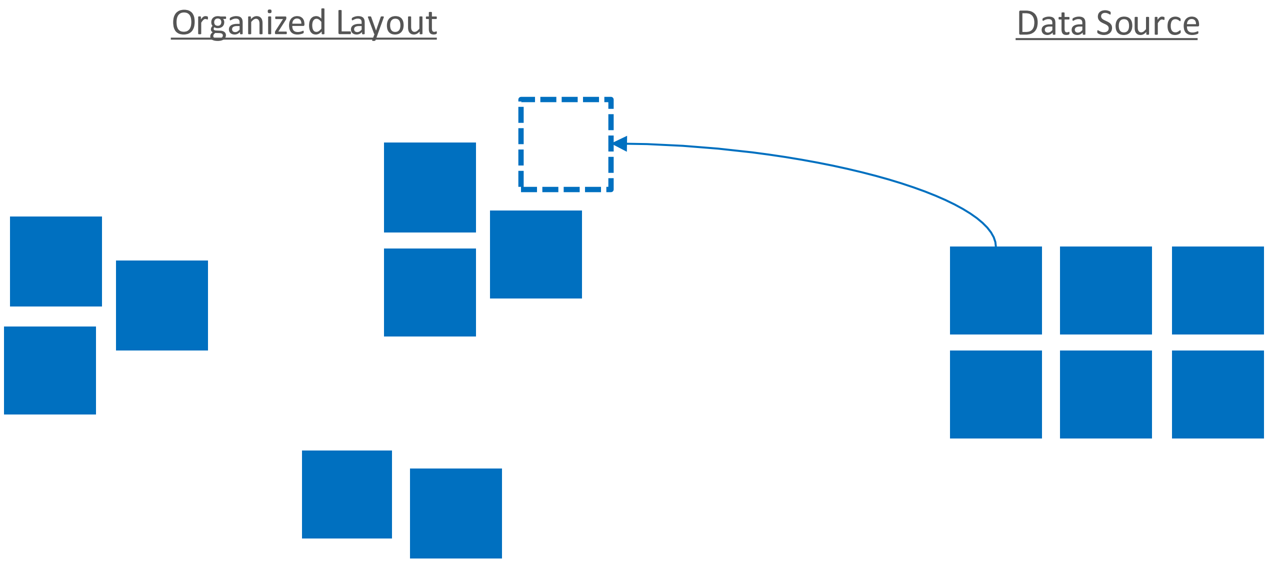


RQ 2.3

How can we use Gaze Score in a real-time sensemaking task?

Application of Gaze Score

Personal Recommendations

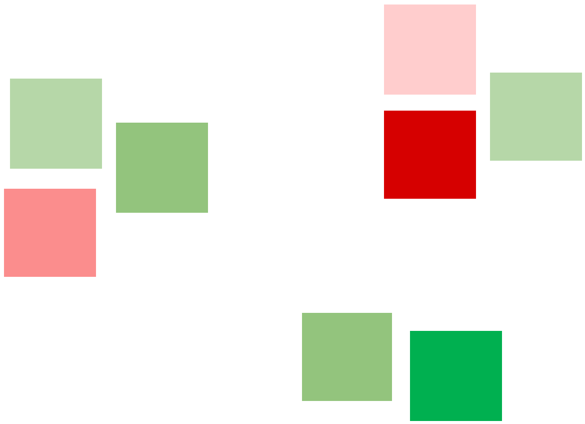


A gaze-driven recommender is aware of the analyst’s interest and helps them expand their knowledge with additional information

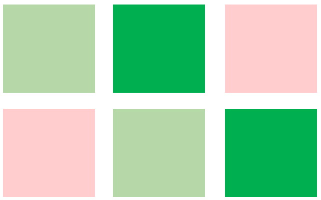
Application of Gaze Score

Externalization

Organized Layout



Data Source



Gaze-driven externalization is a non-invasive, implicit way for an overview of the analyst's mental model, helping in synthesizing information

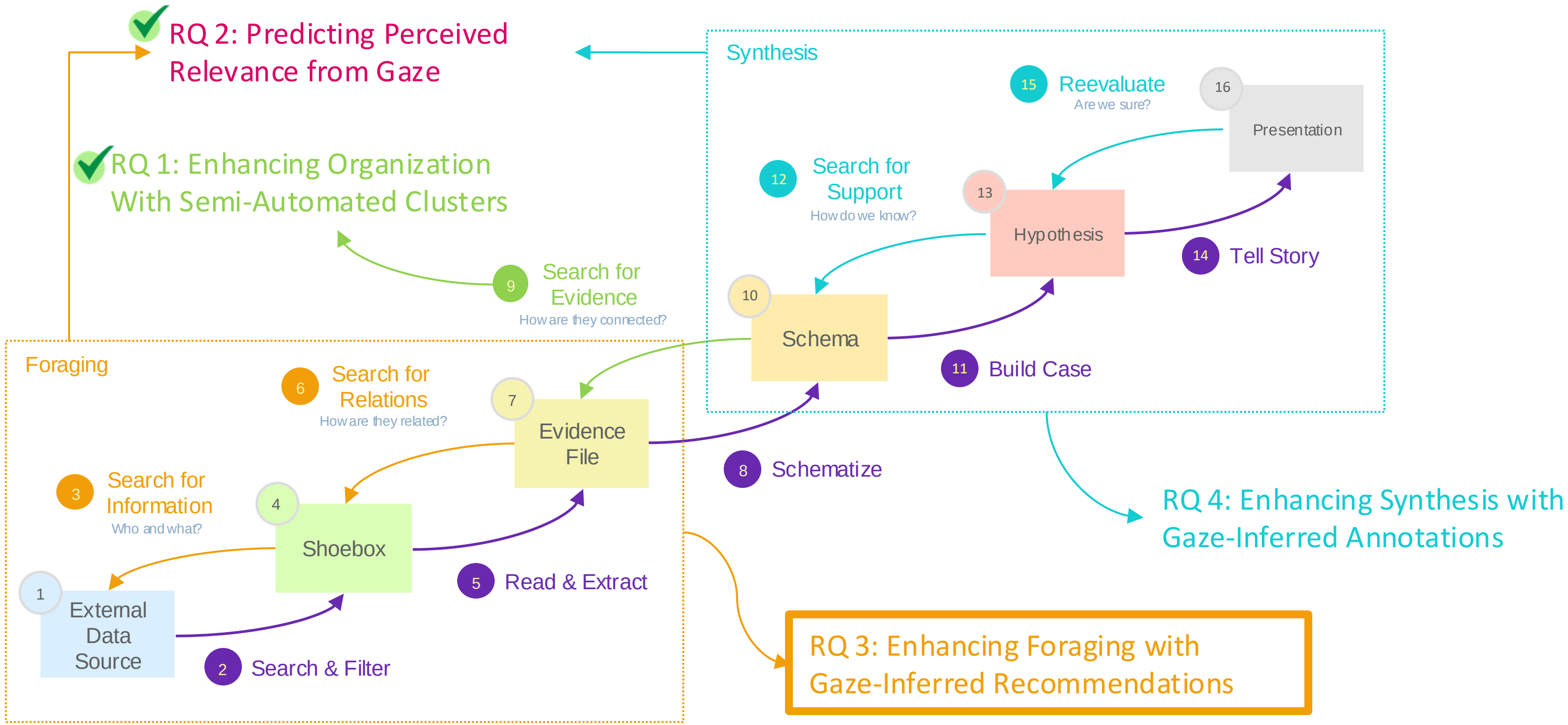
Takeaways

We can **infer user-perceived information relevance** from gaze for sensemaking with multiple documents

Gaze-inferred relevance is **more effective for words**, even for small datasets

Gaze-inferred relevance can be useful for **enhancing foraging and synthesis**

Roadmap



RQ 2: Predicting Perceived Relevance from Gaze
RQ 1: Enhancing Organization With Semi-Automated Clusters
Synthesis
15 Reevaluate
Are we sure?
16 Presentation
12 Search for Support
How do we know?
13 Hypothesis
14 Tell Story
9 Search for Evidence
How are they connected?
10 Schema
11 Build Case
Foraging
6 Search for Relations
How are they related?
7 Evidence File
8 Schematize
3 Search for Information
Who and what?
4 Shoebox
5 Read & Extract
1 External Data Source
2 Search & Filter
RQ 4: Enhancing Synthesis with Gaze-Inferred Annotations
RQ 3: Enhancing Foraging with Gaze-Inferred Recommendations

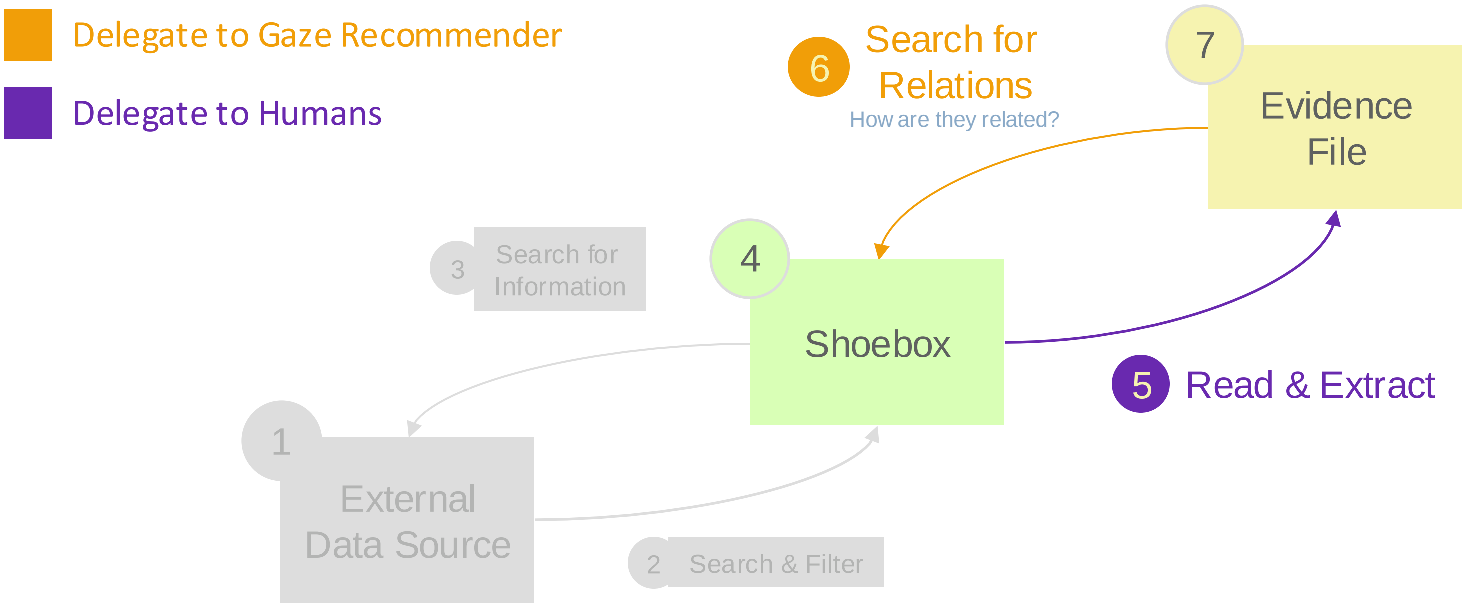
ONGOING
PROPOSED

RQ 3

Gaze-Inferred Recommendations

RQ 3: How can we enhance foraging with gaze-inferred recommendations?

Foraging Loop



In a controlled study environment, the dataset is already filtered and curated by the experimenter

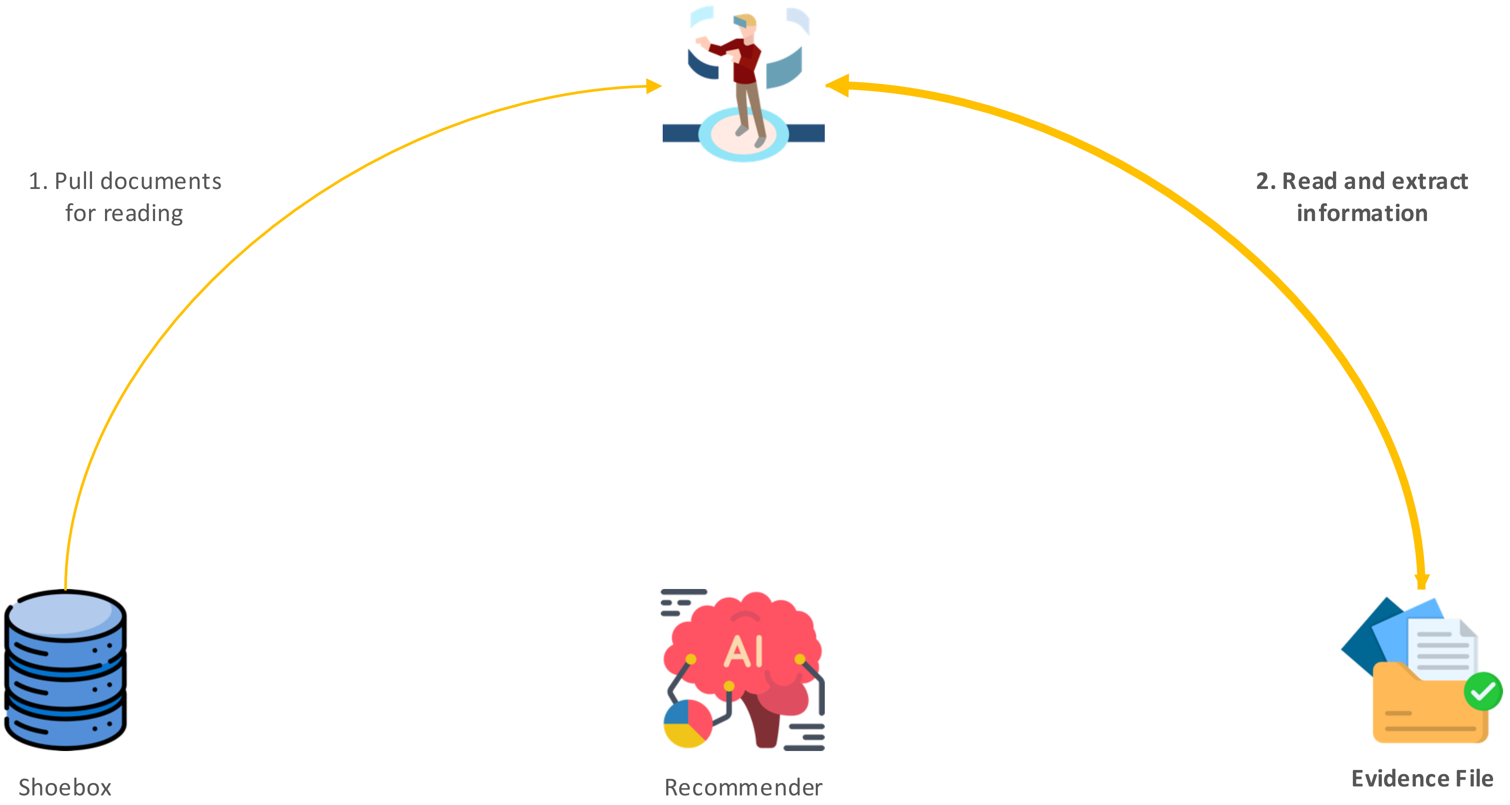
Gaze Recommendation Model

Shoebox

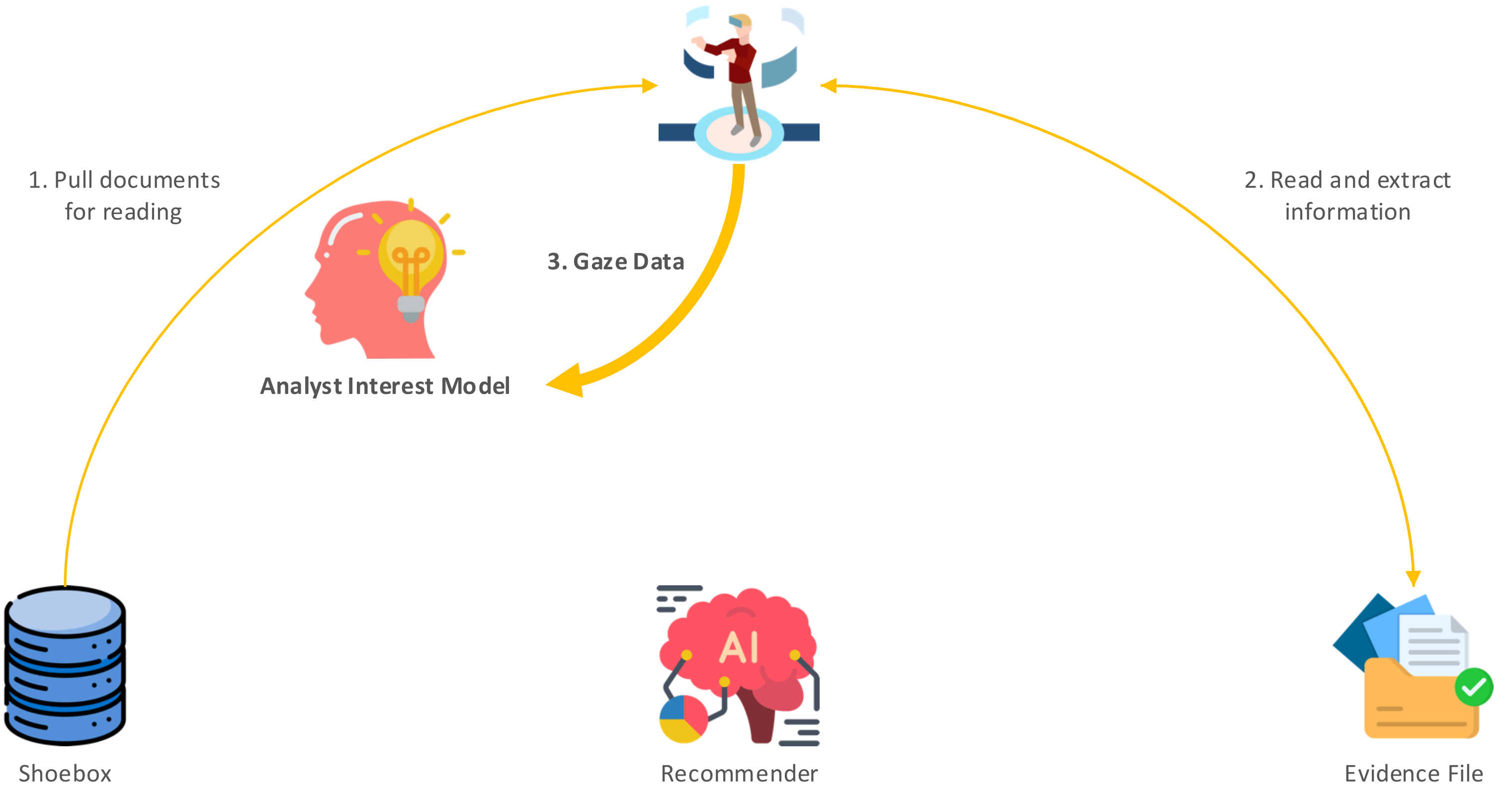
Recommender

Evidence File

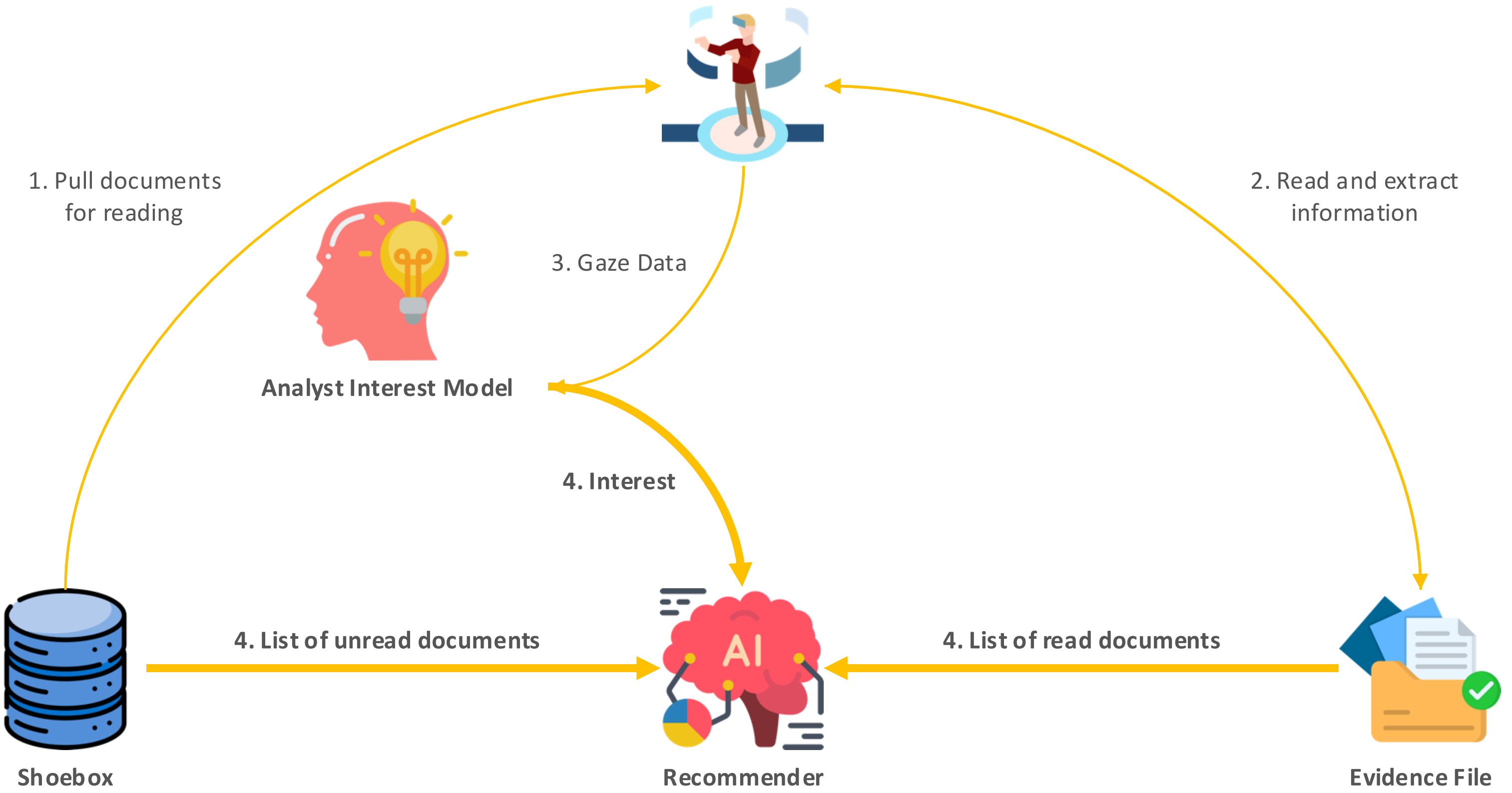
Gaze Recommendation Model



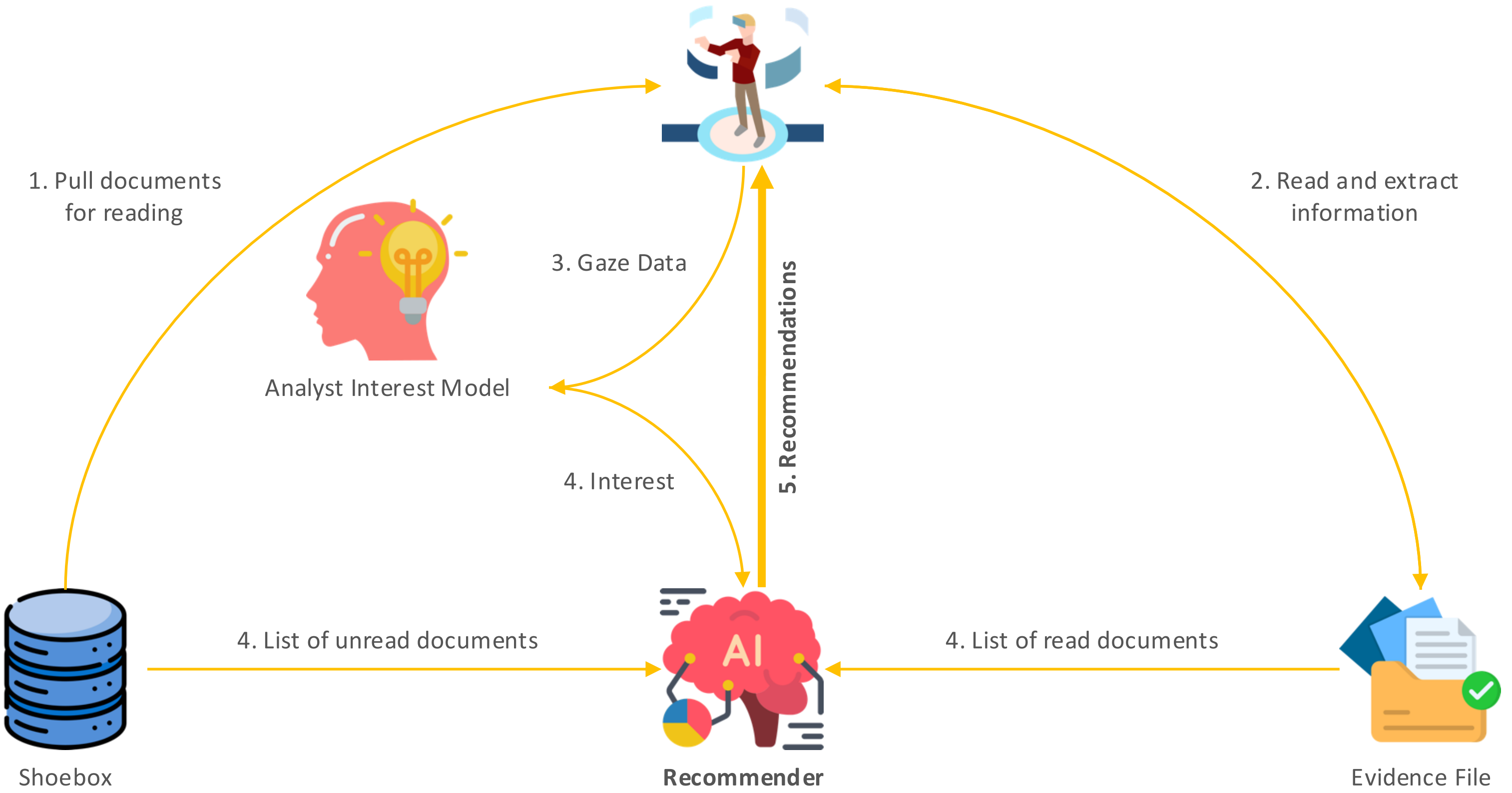
Gaze Recommendation Model



Gaze Recommendation Model



Gaze Recommendation Model



Analyst Interest Model

Local Interest (LI)

How relevant is a word within **a single document**?

Analyst Interest Model

Local Interest (LI)

How relevant is a word within **a single document**?

$$GS_x = \frac{\frac{GD_x - \mu_{GD}}{\sigma_{GD}} + \frac{UD_x - \mu_{UD}}{\sigma_{UD}}}{2} * IDF_x$$

GS = LI
GD: **local** Gaze Duration
UD: **local** Unique Dwell Count
IDF = 1

Analyst Interest Model

Local Interest (LI)

How relevant is a word within **a single document**?

$$GS_x = \frac{\frac{GD_x - \mu_{GD}}{\sigma_{GD}} + \frac{UD_x - \mu_{UD}}{\sigma_{UD}}}{2} * IDF_x$$

GS = LI
GD: **local** Gaze Duration
UD: **local** Unique Dwell Count
IDF = 1

Local Interest vector for document A

$$LI_A = \{LI_{w_1}, LI_{w_2}, LI_{w_3}, \cdots, LI_{w_N}\}$$

Analyst Interest Model

Global Interest (GI)

How relevant is a word within the **whole dataset**?

$$GS_x = \frac{\frac{GD_x - \mu_{GD}}{\sigma_{GD}} + \frac{UD_x - \mu_{UD}}{\sigma_{UD}}}{2} * IDF_x$$

GS = GI
GD: **global** Gaze Duration
UD: **global** Unique Dwell Count
IDF = log (D/d)
D = number of total documents
d = number of documents with the word

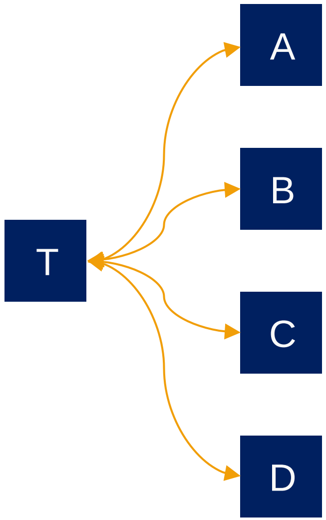
Global Interest vector

$$GI = \{GI_{w_1}, GI_{w_2}, GI_{w_3}, \cdots, GI_{w_N}\}$$

Recommender

$$Similarity(A, B) = cosine(LI_A, LI_B)$$

Let's find more documents **similar** to this document (T)

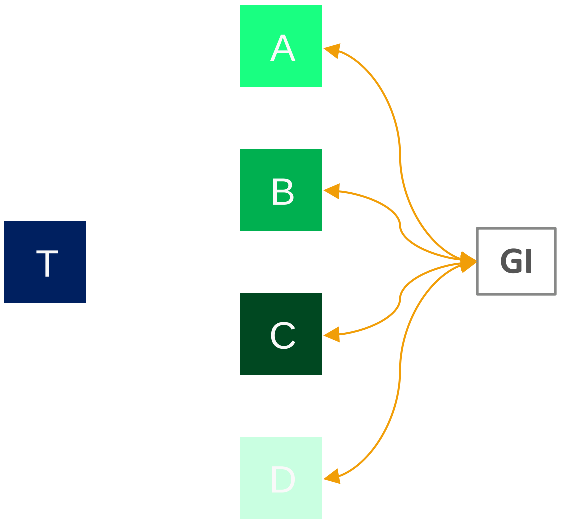


Recommender

$$Relevance(A) = cosine(LI_A, GI)$$

Let’s find more documents **similar** to this document (T)

Let’s find documents **relevant to my global interest**



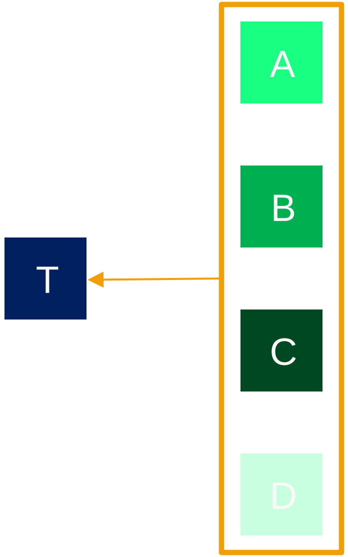
Recommender

$$RecommendationScore_A = \frac{Similarity(A, target) + Relevance(A)}{2}$$

Let’s find more documents **similar** to this document (T)

Let’s find documents **relevant to user’s global interest**

Let’s **consider both** for final list of recommendations



Recommender Design

- Show recommendations
- Give rationale
- Preserve the spatial layout
- Allow user interaction

Recommender Design

- Show recommendations
- Give rationale
- Preserve the spatial layout
- Allow user interaction

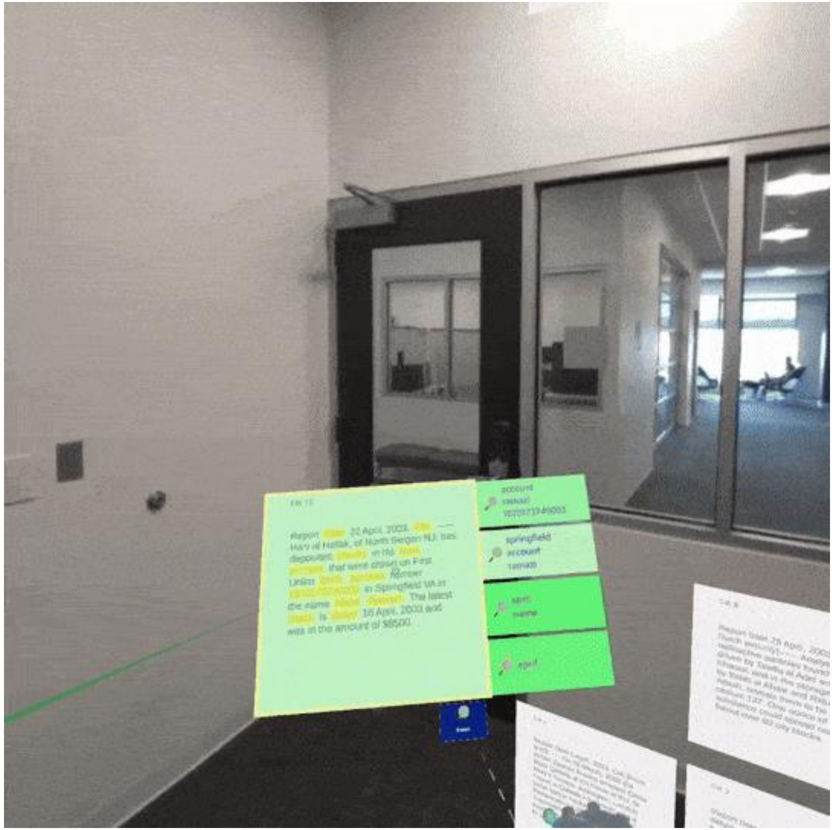
Recommender Design

Show recommendations

Give rationale

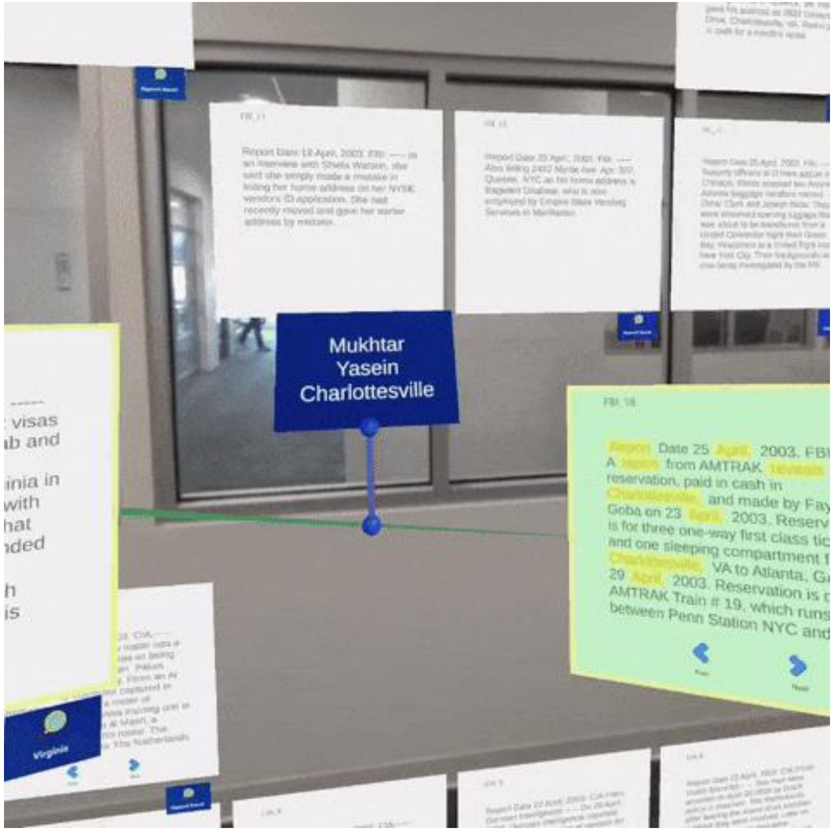
Preserve the spatial layout

Allow user interaction

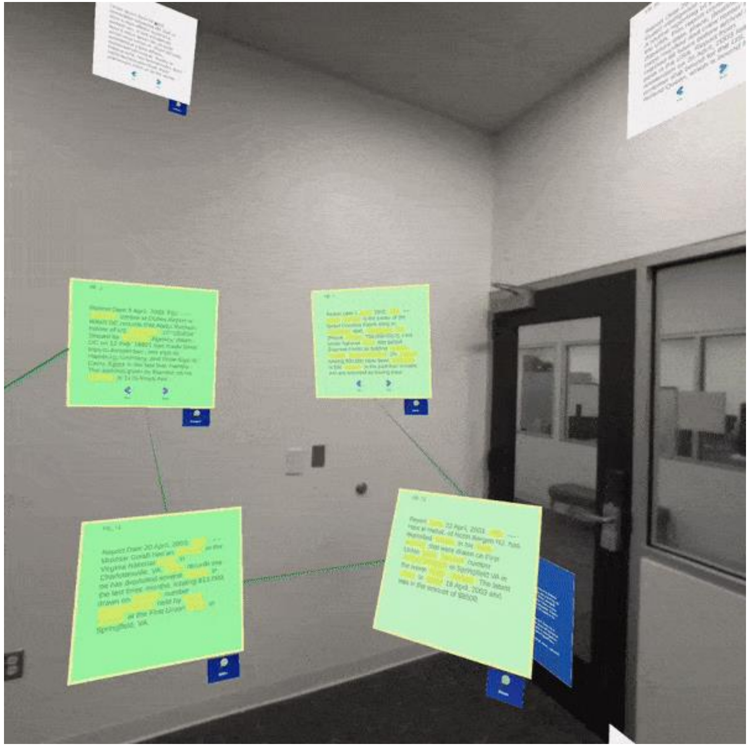


Recommender Design

- Show recommendations
- Give rationale
- Preserve the spatial layout
- Allow user interaction



Recommender Demo



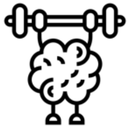
Shoebox and Evidence File

Research Questions

How do gaze–inferred recommendations affect …

Task
Performance

Sensemaking
Strategy



Mental
Effort

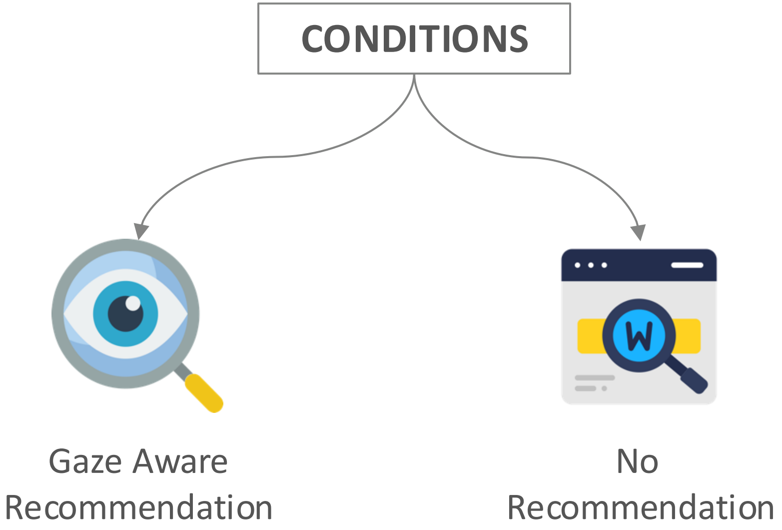
User
Experience

Experiment Design

26 Participants*
Between subjects

Blue Iguanodon
100 Documents*
33 with ground truth

Create Notes
Make labels
Search keywords



* correction from the prelim document

Procedure

1. Introduction

- Consent Form
- Pre-study Questionnaire
- Eye Calibration

2. Tutorial

- Intro to IST features
- Complete a dummy task

3. Main Study

- **4 documents as starting point**
- **2 distractors**
- 45 minutes to investigate

4. Post-Study

- NASA TLX, UEQ
- Semi-structured interview

Hypotheses

With gaze-inferred recommendations, participants will

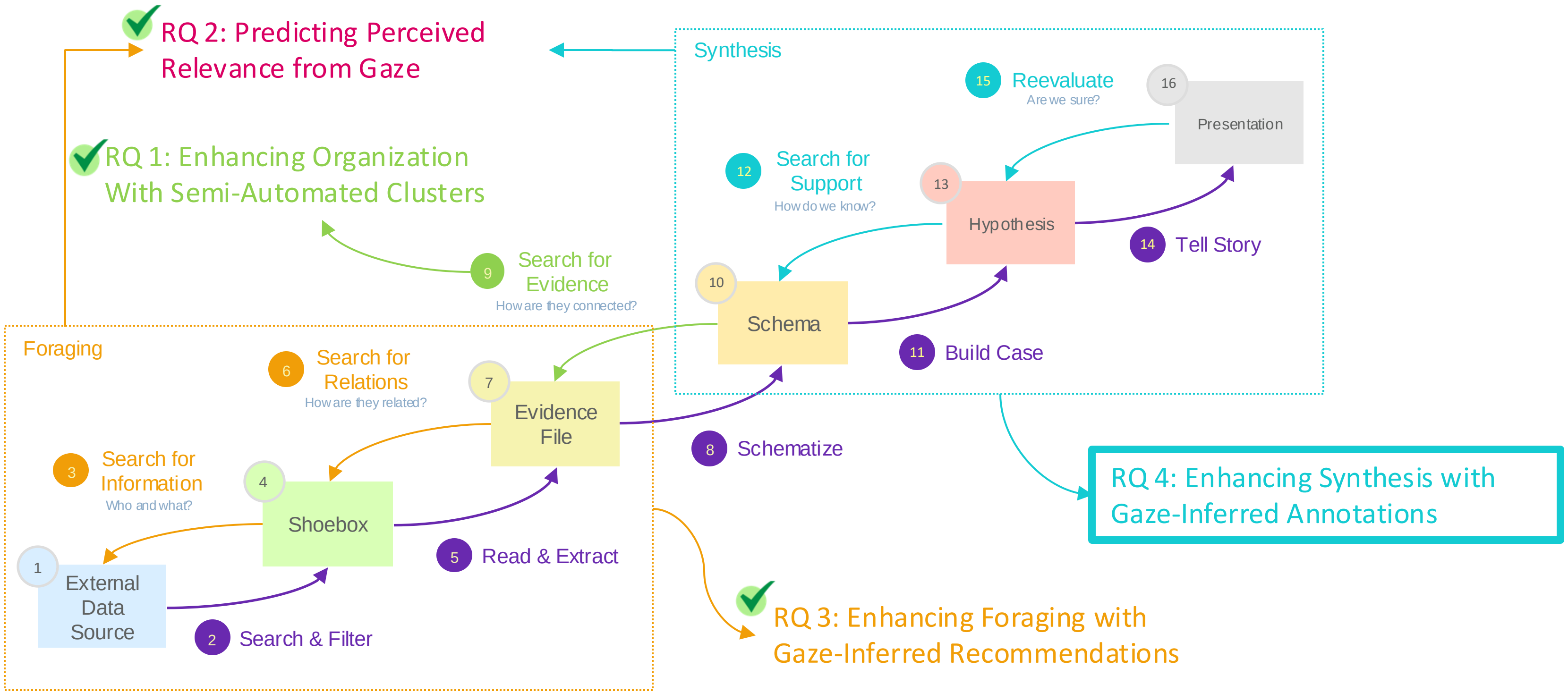
- Find the first relevant document in **less time**
- Find the first relevant word in **less time**
- Find more clues from **less documents**
- Report **less mental effort**
- Not score higher

Expected Outcome

- Understanding the **workflow of an analyst** during sensemaking with a personalized recommender system

- Understanding how the analyst’s **foraging and synthesis strategies** change while working with a personalized recommender

Roadmap


RQ 2: Predicting Perceived Relevance from Gaze
RQ 1: Enhancing Organization With Semi-Automated Clusters
Synthesis
15 Reevaluate
Are we sure?
16
Presentation
12 Search for Support
How do we know?
13
Hypothesis
14 Tell Story
9 Search for Evidence
How are they connected?
10
Schema
11 Build Case
Foraging
6 Search for Relations
How are they related?
7
Evidence File
8 Schematize
RQ 4: Enhancing Synthesis with Gaze-Inferred Annotations
3 Search for Information
Who and what?
4
Shoebox
1
External Data Source
5 Read & Extract
2 Search & Filter
RQ 3: Enhancing Foraging with Gaze-Inferred Recommendations

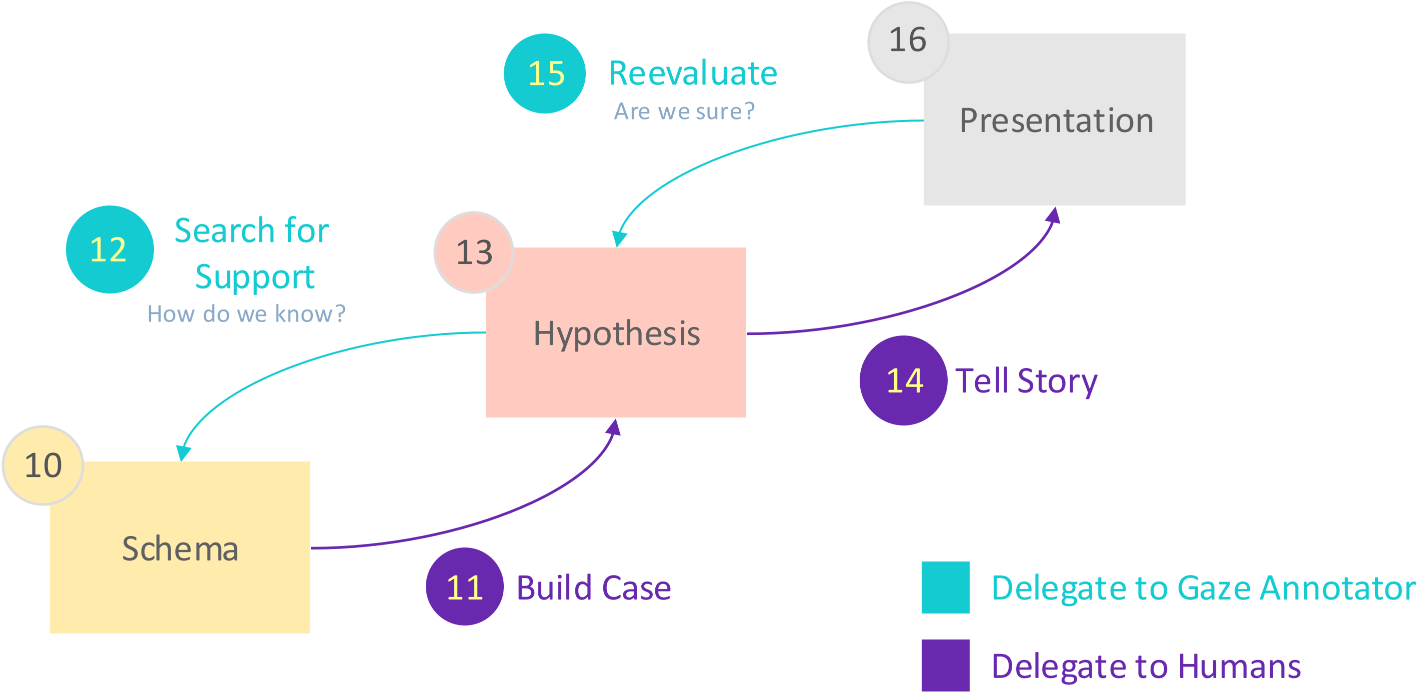
PROPOSED

RQ 4

Gaze-Inferred Annotations

RQ 4: How can we enhance synthesis with gaze-inferred annotations?

Synthesis Loop

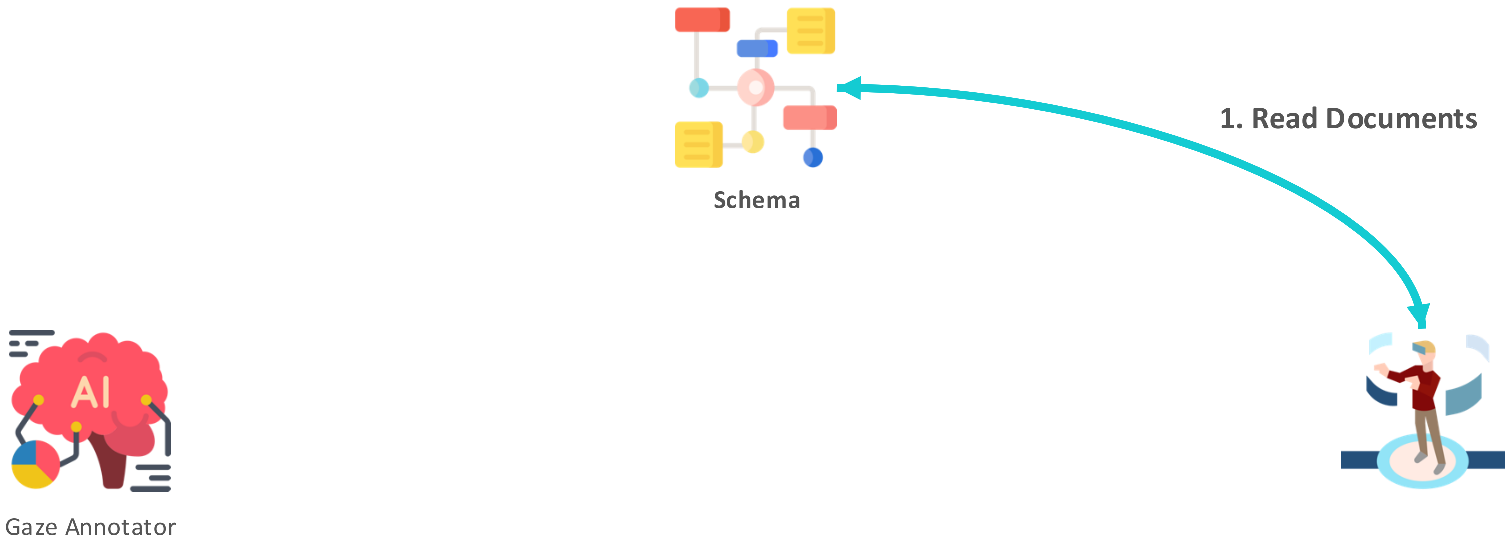


Gaze Annotation Model

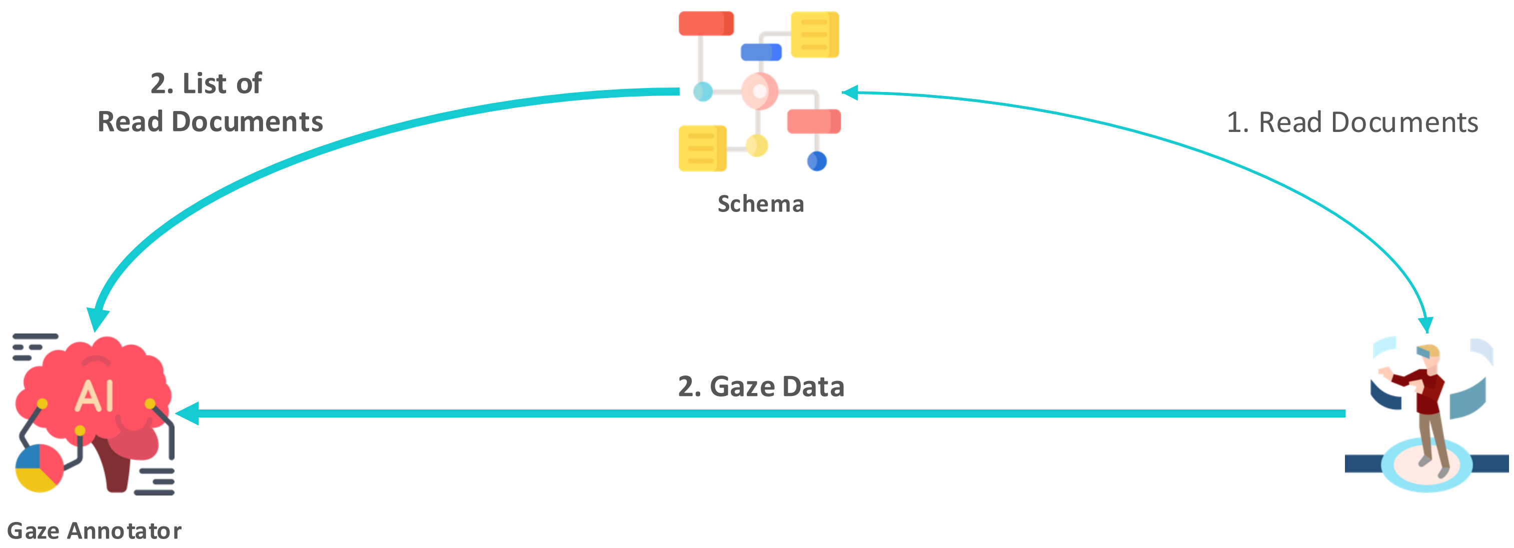
Schema

Gaze Annotator

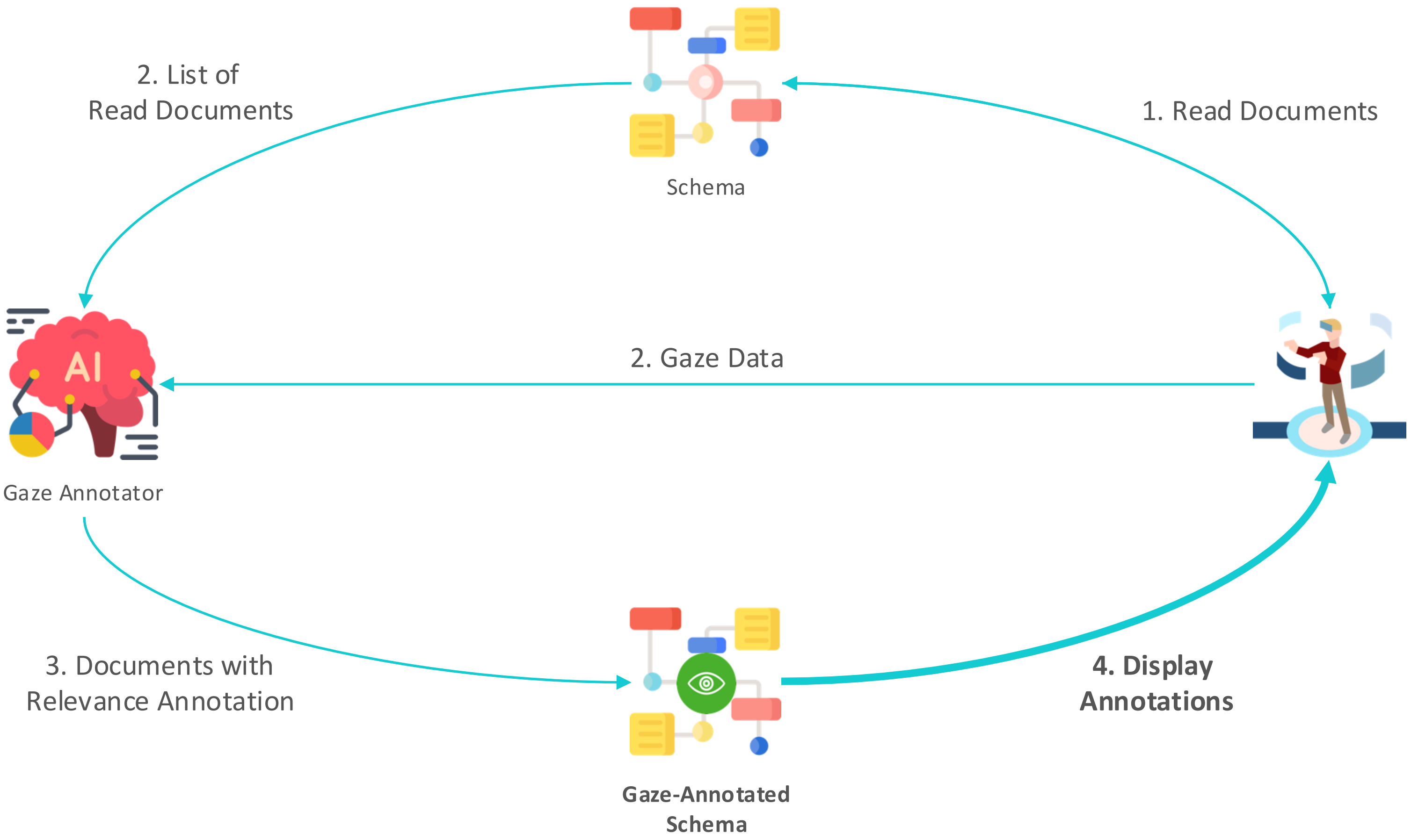
Gaze Annotation Model



Gaze Annotation Model



Gaze Annotation Model



RQ 4.1

How should we design the annotations in a meaningful way?

Annotation Design

Document Relevance Visualization

Word Relevance Visualization

Document Relevance Visualization

Evaluating Techniques

Criteria

- **Detecting** relevant documents
- **Quantifying** the relevance
- **Readability of relevant** documents
- **Readability of irrelevant** documents
- **Spatial memory** preservation

Visualization Techniques*

- Size
- Background Color
- Border Color
- Orientation
- Animation
- Depth

* Refer to the prelim document for detailed evaluation

Word Relevance Visualization

Evaluating Techniques

Criteria	Visualization Techniques*
Noticeability	**bold**
Readability	*italic*
Quantifiability	underline
Visual Overload	highlight
	CAPITALIZE

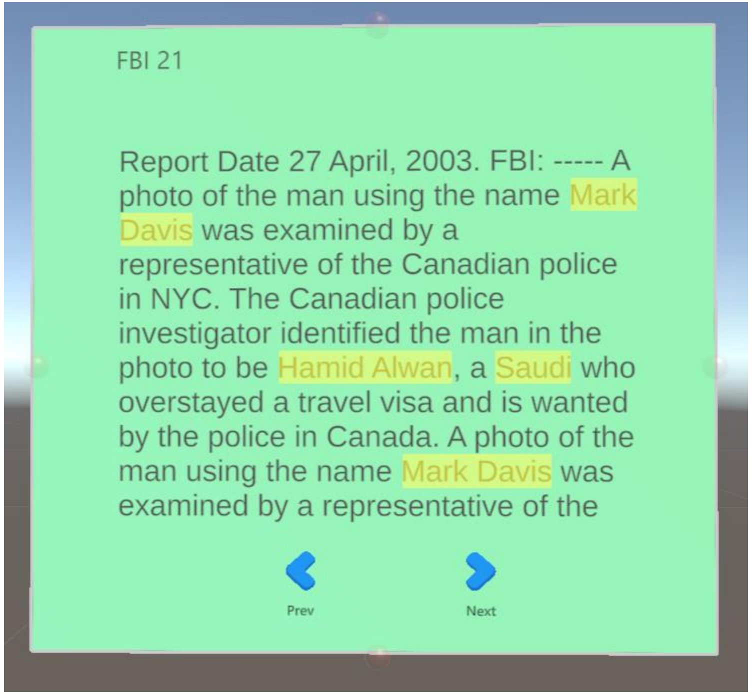
* Refer to the prelim document for detailed evaluation

Annotation Design

Final Output

Background Color with shades of green attracts attention without compromising readability

Highlighting attracts attention maintaining appropriate contrast with the background color



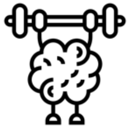
Final visualization technique for gaze-inferred annotation

Research Questions

How do gaze-annotated documents affect …

Synthesis
Strategy

Task
Performance



Mental
Effort

User
Experience

Research Questions

Challenges with evaluating synthesis

Synthesis Strategy

- Mostly in **second-half** of sensemaking
- Data from a potentially **fatigued** user
- Gaze annotation is only as good as the user's sensemaking **skill**

Task Performance

Mental Effort

User Experience

Experiment Design

Collecting Gaze Annotation

3 Professional Analysts

Full Sensemaking Process

Evaluating Gaze Annotation

26 Participants

Synthesis from analyst's layout

Collecting Gaze Annotation

3 Professional Analysts

Dept. of Defense

Sign of the Crescent

41 Documents, 3 Terrorist plots

Create notes
Make labels
Search keywords

Post-Session

- Collect the gaze data during session
- **Save the spatial layout**

- Grade the reports
- **Choose the analyst with the highest score**
- Manually curate the notes and labels to avoid obvious hints

- **Annotate** the documents with gaze-inferred annotations

Evaluating Gaze Annotation

26 Participants

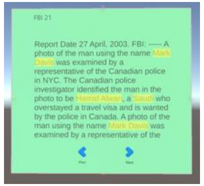
Between subjects

Sign of the Crescent

Analyst's spatial layout
Analyst's notes
Analyst's labels

Create notes
Make labels
Search keywords

CONDITIONS



Gaze Annotated Documents

Plain Documents

Hypotheses

With gaze annotation, participants will

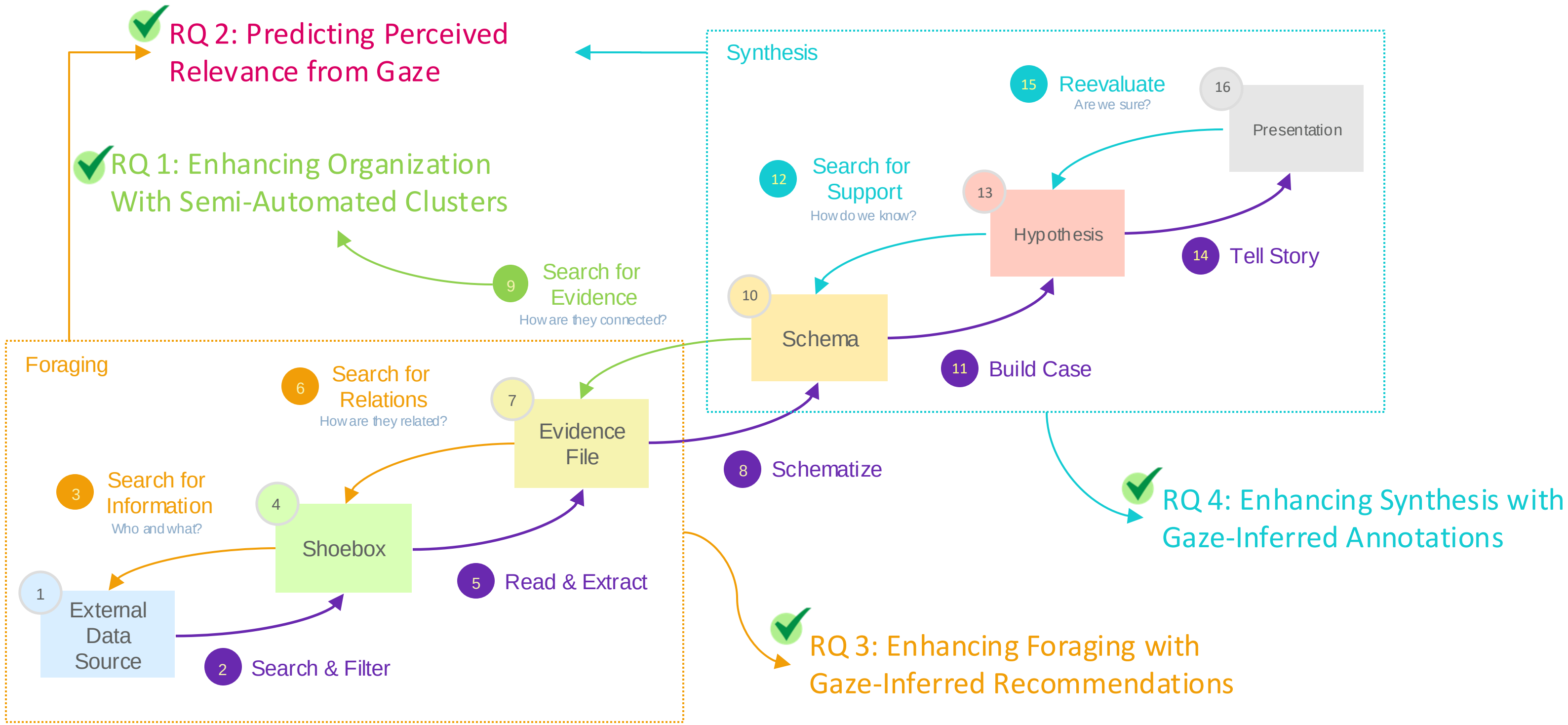
- find the first relevant document in **less time**
- find the first relevant word in **less time**
- spend **more time on relevant** information
- find **more relevant** information
- will make **less mistakes**
- not score higher

Expected Outcome

Understanding the benefits and challenges of **designing gaze annotations** for enhancing synthesis

Understanding how the user's **synthesis strategies** are affected by implicit gaze-derived externalizations

Roadmap

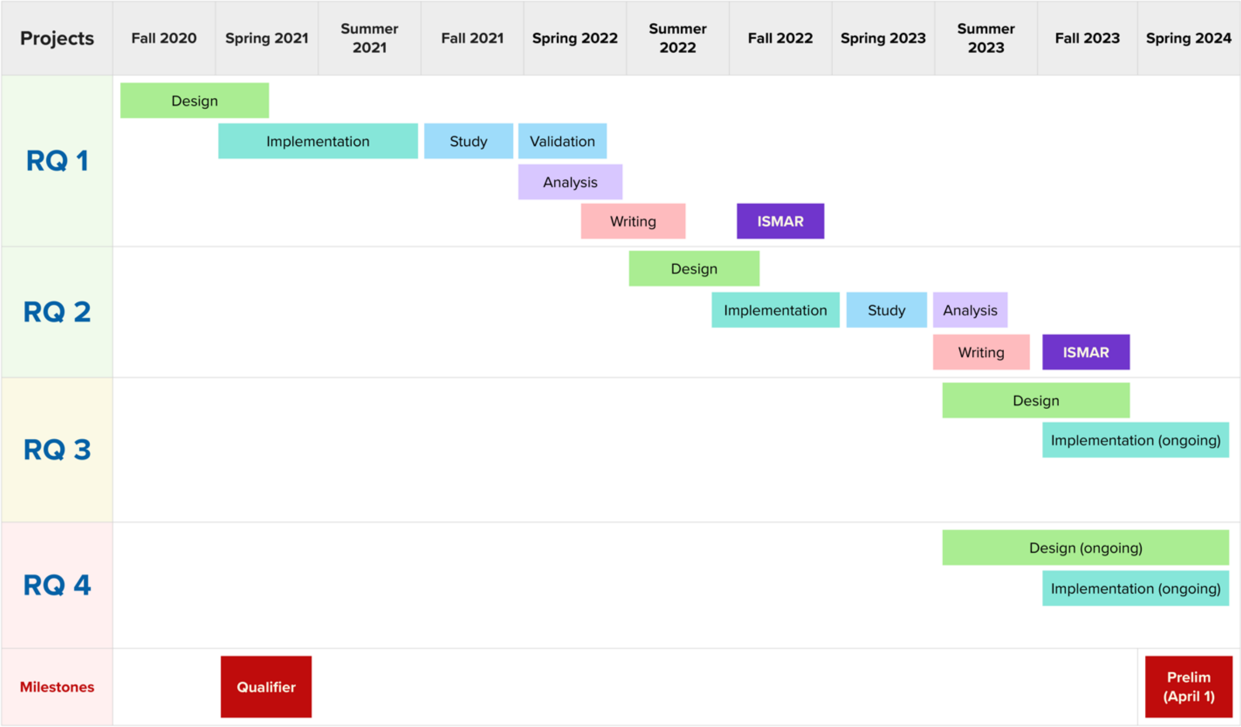


RQ 2: Predicting Perceived Relevance from Gaze
RQ 1: Enhancing Organization With Semi-Automated Clusters
Synthesis
15 Reevaluate
Are we sure?
16 Presentation
12 Search for Support
How do we know?
13 Hypothesis
14 Tell Story
9 Search for Evidence
How are they connected?
10 Schema
11 Build Case
Foraging
6 Search for Relations
How are they related?
7 Evidence File
8 Schematize
RQ 4: Enhancing Synthesis with Gaze-Inferred Annotations
3 Search for Information
Who and what?
4 Shoebox
5 Read & Extract
1 External Data Source
2 Search & Filter
RQ 3: Enhancing Foraging with Gaze-Inferred Recommendations

Timeline

Until Prelim

RQ 1: Enhancing Organization with Cluster
RQ 2: Evaluating Prediction with Eye Gaze
RQ 3: Enhancing Foraging with Gaze-Inferred Recommendations
RQ 4: Enhancing Synthesis with Gaze-Annotated Documents

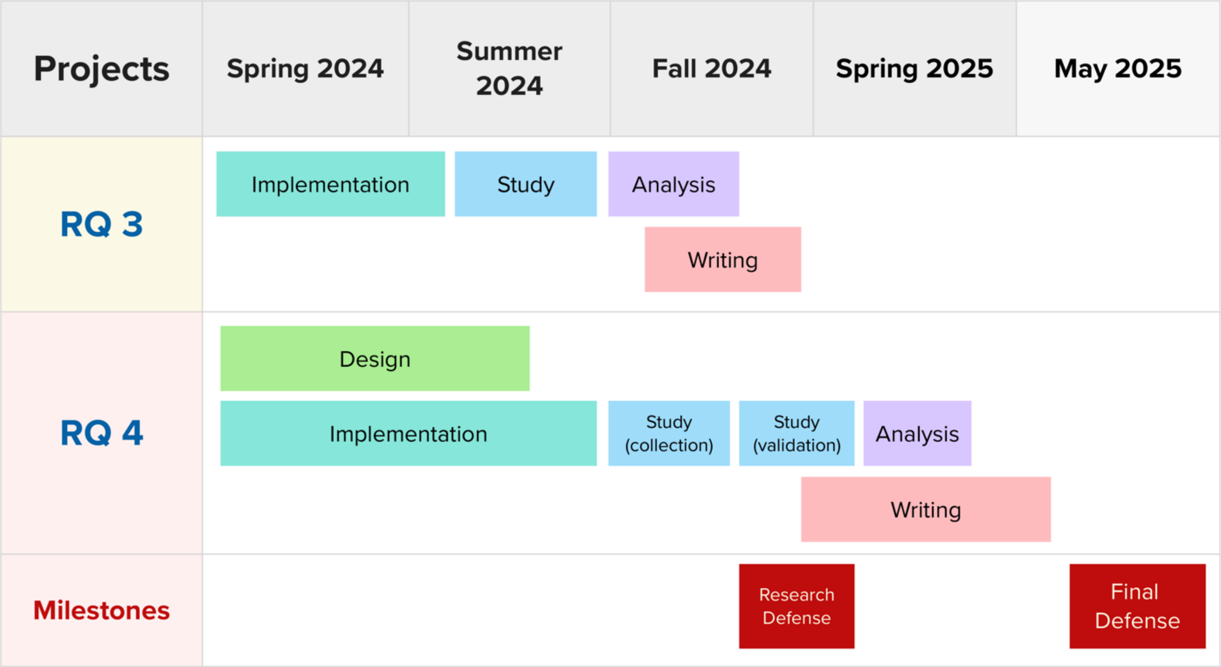


Timeline

After Prelim

RQ 3: Enhancing Foraging with Gaze-Inferred Recommendations

RQ 4: Enhancing Synthesis with Gaze-Annotated Documents



Publication

Published

RQ 1

[1] **Tahmid, I. A.,** Lisle, L., Davidson, K., North, C., & Bowman, D. A. (2022, October). Evaluating the benefits of explicit and semi-automated clusters for immersive sensemaking. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 479-488). IEEE.

RQ 2

[2] **Tahmid, I. A.,** Lisle, L., Davidson, K., Whitley, K., North, C., & Bowman, D. A. (2023, October). Evaluating the Feasibility of Predicting Information Relevance During Sensemaking with Eye Gaze Data. In 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 713-722). IEEE.

Publication

Planned

RQ 3

[*] **Tahmid, I. A.,** Davidson, K., Whitley, K., North, C., & Bowman, D. A.; Exploring the Effects of Smart Recommendations based on User's Eye Gaze in Immersive Sensemaking.
Target: IEEE VR 2025. Submission: October, 2024

RQ 4

[*] **Tahmid, I. A.,** Davidson, K., Whitley, K., North, C., & Bowman, D. A.; Evaluating the Benefits and Challenges of Gaze-Annotated Documents for Immersive Sensemaking.
Target: ISMAR 2025. Submission: May, 2025

Publication

Other IST Aspects

[3] **Tahmid, I. A.,** Rodrigues, F., Giovannelli, A., Lisle, L., Thomas, J., & Bowman, D. A. (2023, October). CoLT: Enhancing Collaborative Literature Review Tasks with Synchronous and Asynchronous Awareness Across the Reality-Virtuality Continuum. In 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 831-836). IEEE.

[4] Davidson, K., Lisle, L., **Tahmid, I. A.,** Whitley, K., North, C., & Bowman, D. A. (2023, October). Uncovering Best Practices in Immersive Space to Think. In 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 1094-1103). IEEE.

[5] Lisle, L., Davidson, K., Pavanatto, L., **Tahmid, I. A.,** North, C., & Bowman, D. A. (2023, October). Spaces to Think: A Comparison of Small, Large, and Immersive Displays for the Sensemaking Process. In 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 1084-1093). IEEE.

Publication

Unrelated

[6] Li, Y., **Tahmid, I. A.,** Lu, F., & Bowman, D. A. (2022). Evaluation of pointing ray techniques for distant object referencing in model-free outdoor collaborative augmented reality. IEEE Transactions on Visualization and Computer Graphics, 28(11), 3896-3906.

[7] Azizi, A., **Tahmid, I. A.,** Waheed, A., Mangaokar, N., Pu, J., Javed, M., ... & Viswanath, B. (2021). {T-Miner}: A generative approach to defend against trojan attacks on {DNN-based} text classification. In 30th USENIX Security Symposium (USENIX Security 21) (pp. 2255-2272).

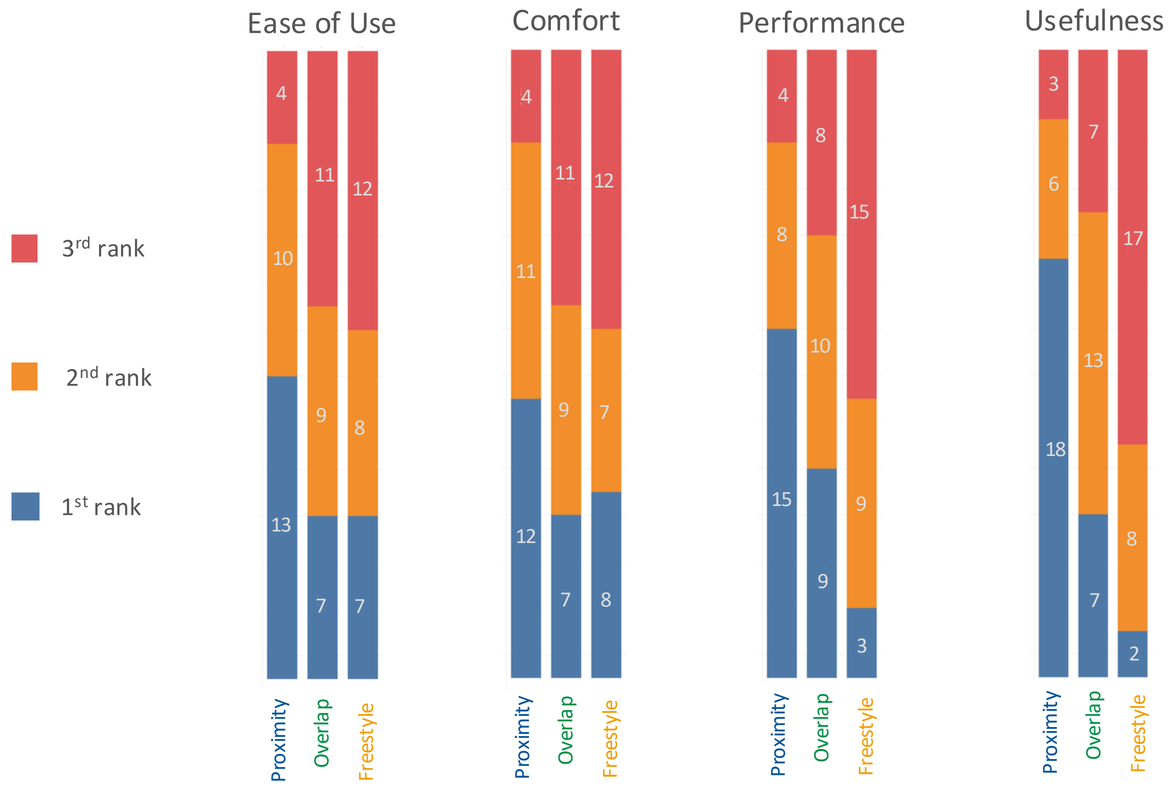
[8] Zhang, L., Lu, F., **Tahmid, I. A.,** Davari, S., Lisle, L., Gutkowski, N., ... & Bowman, D. A. (2021, March). Fantastic voyage 2021: Using interactive VR storytelling to explain targeted COVID-19 vaccine delivery to antigen-presenting cells. In 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 695-696). IEEE.

[9] Lisle, L., Lu, F., Davari, S., **Tahmid, I. A.,** Giovannelli, A., Llo, C., ... & Bowman, D. A. (2022, March). Clean the ocean: An immersive vr experience proposing new modifications to go-go and wim techniques. In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 920-921). IEEE.

[10] Giovannelli, A., Rodrigues, F., Davari, S., **Tahmid, I. A.,** Lane, L., Connor, C., ... & Bowman, D. A. (2023, March). CLUE HOG: An Immersive Competitive Lock-Unlock Experience using Hook On Go-Go Technique for Authentication in the Metaverse. In 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 945-946). IEEE.

THANK YOU

Benefits of **Proximity**



> *It [Proximity] was as easy as Freestyle, with the added benefits of the explicit clusters*

RQ1 Qualitative Analysis

24 participants (89%) preferred having cluster interactions over Freestyle

18 participants (67%) thought Proximity had **better performance** and was **more useful** than Overlap

19 participants (70%) found Proximity **easier to use** and required **less mental workload** than Overlap

24 participant20 participants (74%) would choose Proximity over Overlap given the same tasks (89%) preferred having cluster interactions over Freestyle

RQ1 Qualitative Analysis

Proximity

Faster
Easy as the Freestyle
Confused about the merging constraint
Challenging for bigger dataset

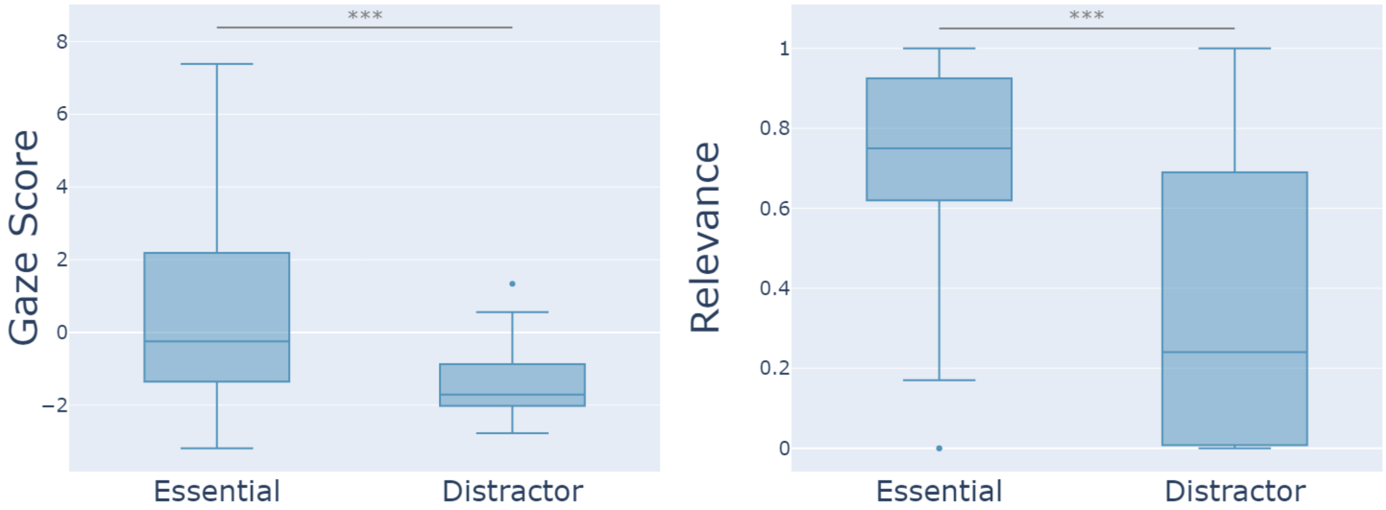
Overlap

Intuitive
Natural
Visual feedback gave control
Required conscious effort to overlap

Freestyle

Allows creativity
Quick to adapt
Useful for rare cases (one doc -> multiple clusters)
Extra cognitive load to keep clusters separate

Essential vs Distractors

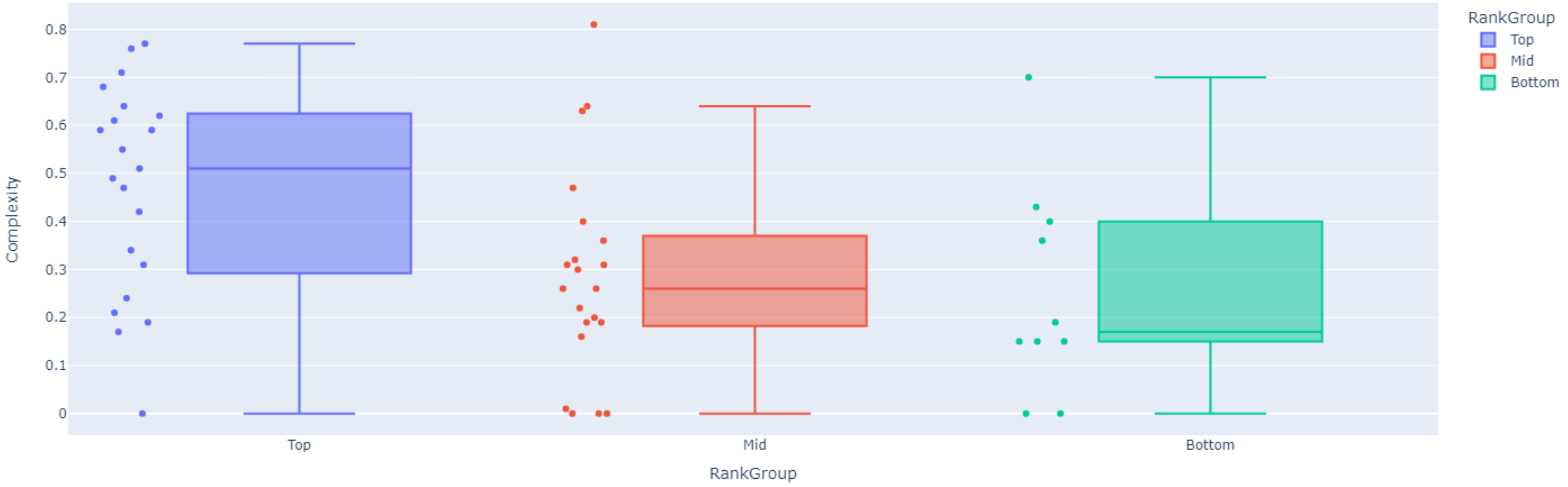


Gaze Score vs **Gaze Duration** vs **Unique Dwell**

- GD and UD both perform **on par with the GS**
- BUT **rank matters**: 5th word and 15th word both fall within top 3%, but while recommending which one should we choose?
- Let's compare the ranks for documents and words that are found relevant by the user. We consider a metric better if it has a higher rank for the relevant information.

- GS outperforms GD for 50% of documents
- GS outperforms UD for 54% of documents
- GS outperforms GD for 75% of words
- **UD outperforms GS for 63% of words**

Gaze Score vs Gaze Duration vs Unique Dwell



- **Complex words have higher UD** => possibility of mislead
- Does not address frequency bias => challenging for bigger dataset

Feedback on Recommender

Evaluation from one experimenter

1. The common words help **build confidence** on the system
2. The color on the tab **does not influence click-for-detail** feature
3. The color on the document background **influences in reviewing information**
4. **MUST read some documents** before the recommendations start getting good
5. After a while, for some documents, three or all four recommendations are from the evidence file. Suggestions for **a way to always look for new documents**.

Study 3: Interview Questions

1. How well do you think you performed today?
2. What strategy did you follow to complete the task?
3. (for the Gaze Aware condition) What was your perception of the recommendations? Did you feel confident with the suggestions?
4. (for the Gaze Aware condition) Did the recommendations help in your task? Why or why not?
5. Did you feel in control of your layout? Why or why not?
6. Do you have any additional comments or suggestions to improve the system?

Study 4: Interview Questions

1. How well do you think you performed today?
2. What strategy did you follow to complete the task?
3. Did the annotations help you in finding the solution? Why or why not?
4. Did the annotations interfere with your thought process during the synthesis process? Why or why not?
5. (for Gaze Aware condition) Did you prefer any one annotation (gaze or explicit) over the other? Why or why not?
6. Do you have any additional comments or suggestions to improve the system?

Sensemaking