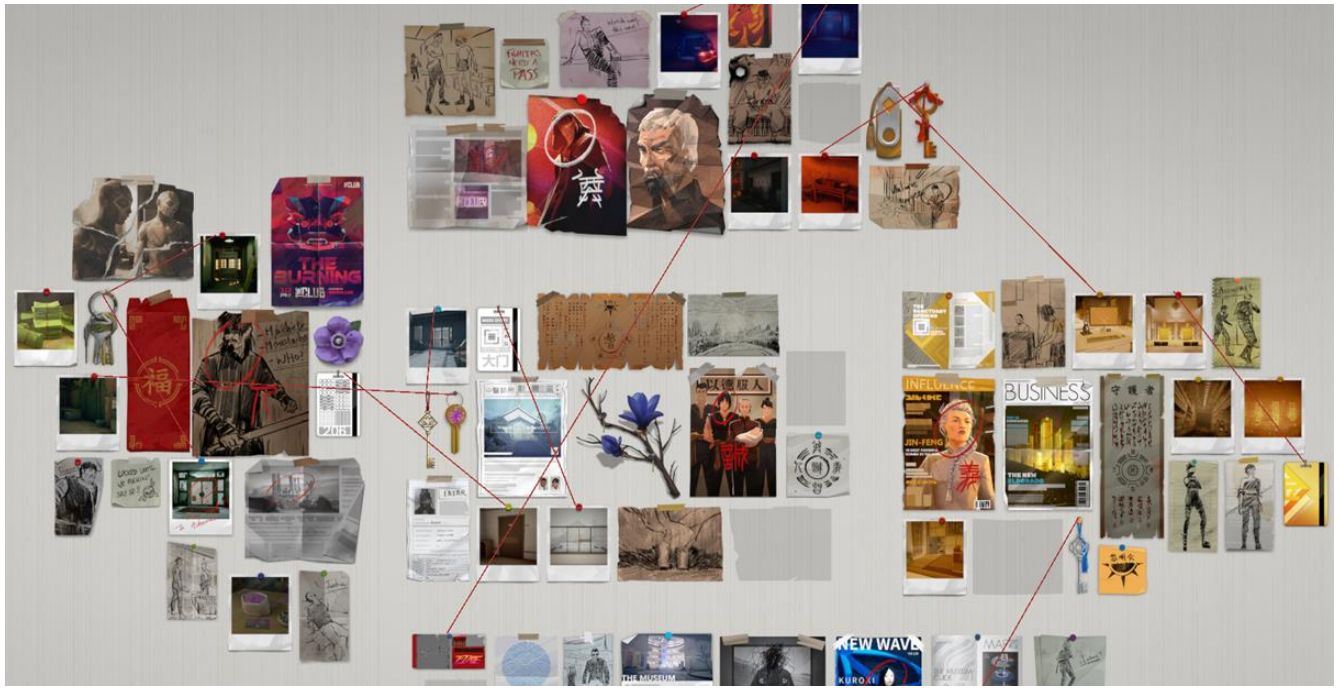
Evaluating the Feasibility of Predicting Information Relevance During Sensemaking with Eye Gaze Data

Ibrahim Tahmid[1],
Lee Lisle[1], Kylie Davidson[1], Kirsten Whitley[2], Chris North[1], and Doug A. Bowman[1]

[1]Virginia Tech
[2]US Department of Defense

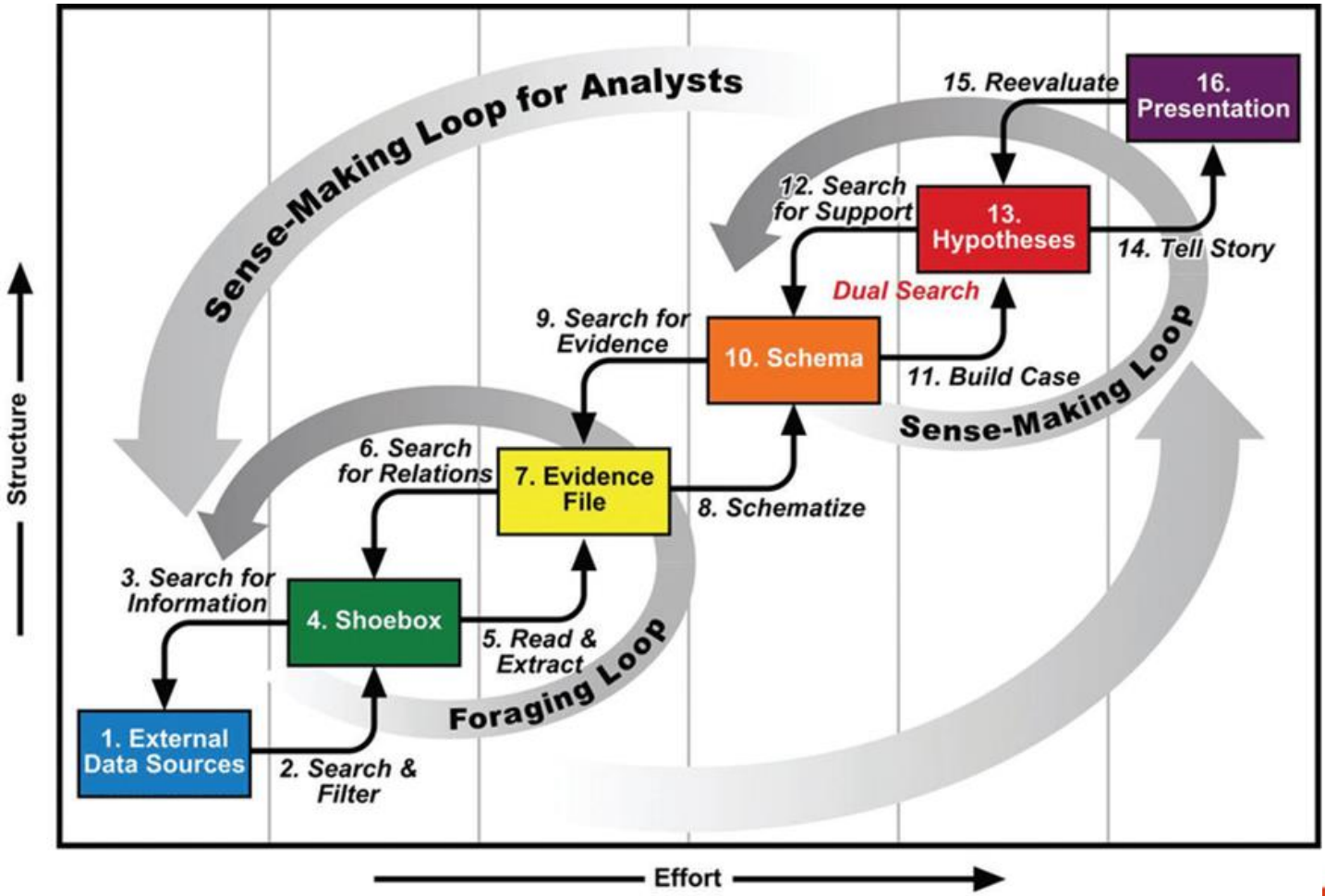
Sensemaking

- **Example:** intelligence analyst working on a case
  - **Gather** information
  - **Extract** insights
  - Build a **schema** with evidence
  - Generate **hypothesis**
  - Tell a **story**



Sensemaking

- Sensemaking loop by Pirolli & Card
- Effort => Structure
- The process is cognitively stressful [1]
- What if the visual analytic system could share some of that stress?



[1] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In Proceedings of international conference on intelligence analysis, vol. 5, pp. 2–4. McLean, VA, USA, 2005.

Challenges

- High cognitive effort
- Multiple inter-connected documents

Proposed Solution

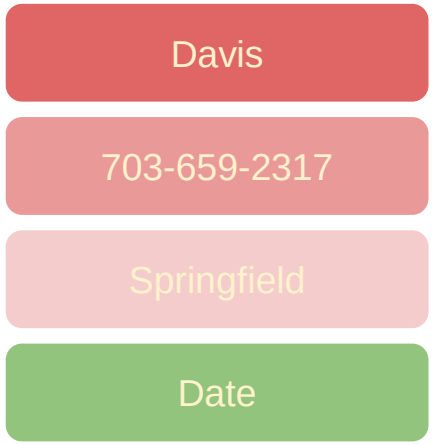
- Intelligent Immersive Analytics System
- Infer the intent of analysts by tracking eye gaze
- Provide related content based on their interest

Role of Eye Gaze

- Gaze Duration
  - Amount of time spend on a document or a word
- Unique Dwell Count
  - Number of times the reader shifted their attention to a document or a word

User Perceived Relevance
from Gaze Duration or Unique Dwell

- Relevant documents and words get more attention from readers [2,3]
- Words by relevance rank



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].

[2] 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.
[3] 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

User Perceived Relevance
from Gaze Duration or Unique Dwell

- Relevant documents and words get more attention from readers [2,3]
- Words by relevance rank

Davis

703-659-2317

Date

Springfield

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

[2] 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.
[3] 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

Multiple Documents
where Gaze Duration and Unique Dwell mislead

- Relevant documents and words get more attention from readers [2,3]
- Words by relevance rank

Date

Davis

703-659-2317

Springfield

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].

[2] 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.
[3] 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 documents introduces frequency bias
- Some words get more attention for high frequency rather than their relevance
- Gaze duration or dwell count alone cannot address the frequency bias

User Reported	Gaze Inferred
Mark Davis	Name
Foysal Goba	2003
6302	Address
Texas	Date
Virginia	Bank
Passport	Phone
Laundering	List
Bank	Virginia
Deposit	Check
Myrtle	805-759-6302

Length Bias

- Longer documents get more attention for large number of words rather than their relevance
- Gaze duration alone cannot address the length bias

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.

Research Questions

How can we design a gaze metric to capture analyst's focus of attention during sensemaking with multiple text documents?

To what extent does the metric predict the relevance of documents and words?

How can we use the metric in a real-time sensemaking task to improve intelligent suggestions to analysts?

Introducing GazeScore

Considerations

- Normalize the gaze data to remove frequency bias for words
- Normalize the gaze data to remove length bias for documents
- Combine Gaze Duration and Unique Dwell to infer analyst's perception of relevance of docs and words

Normalize and Combine

$$GazeScore_i = Z_i * IDF_i$$

Remove Frequency Bias

[IDF (rarest word) = 1]
[IDF (most common word) = 0]
[IDF (doc) = 0]

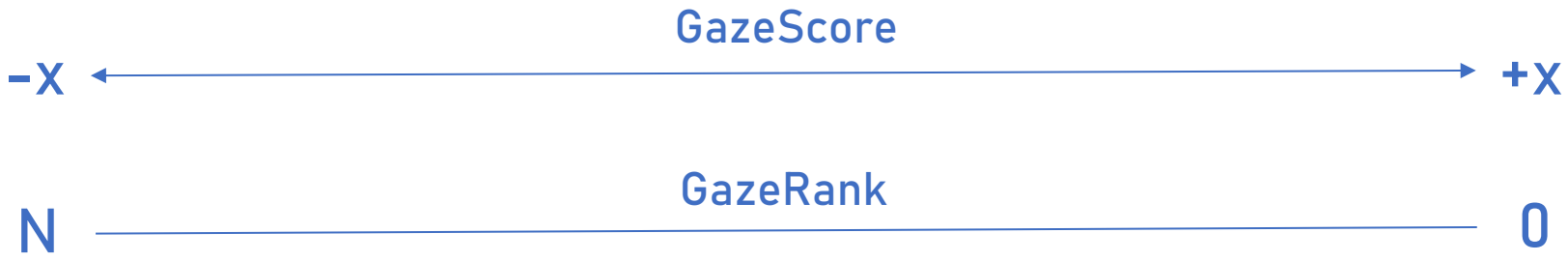
$$Z_{UD_i} = \frac{x_{UD_i} - mean}{Std.Dev}$$

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

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

Introducing GazeRank

- Different people have different reading speeds
  - Arbitrary high and low value for GazeScore
  - Prevents comparison
- **GazeRank:** index on the sorted GazeScores
  - Fixed high and low value
  - Allows comparison between participants



Research Questions

~~How can we design a gaze metric to capture analyst's focus of attention during sensemaking with multiple text documents?~~

To what extent does the metric predict the relevance of documents and words?

How can we use the metric in a real-time sensemaking task to improve intelligent suggestions to analysts?

User Study

3♀ 8♂ 1⚧ 12 PARTICIPANTS

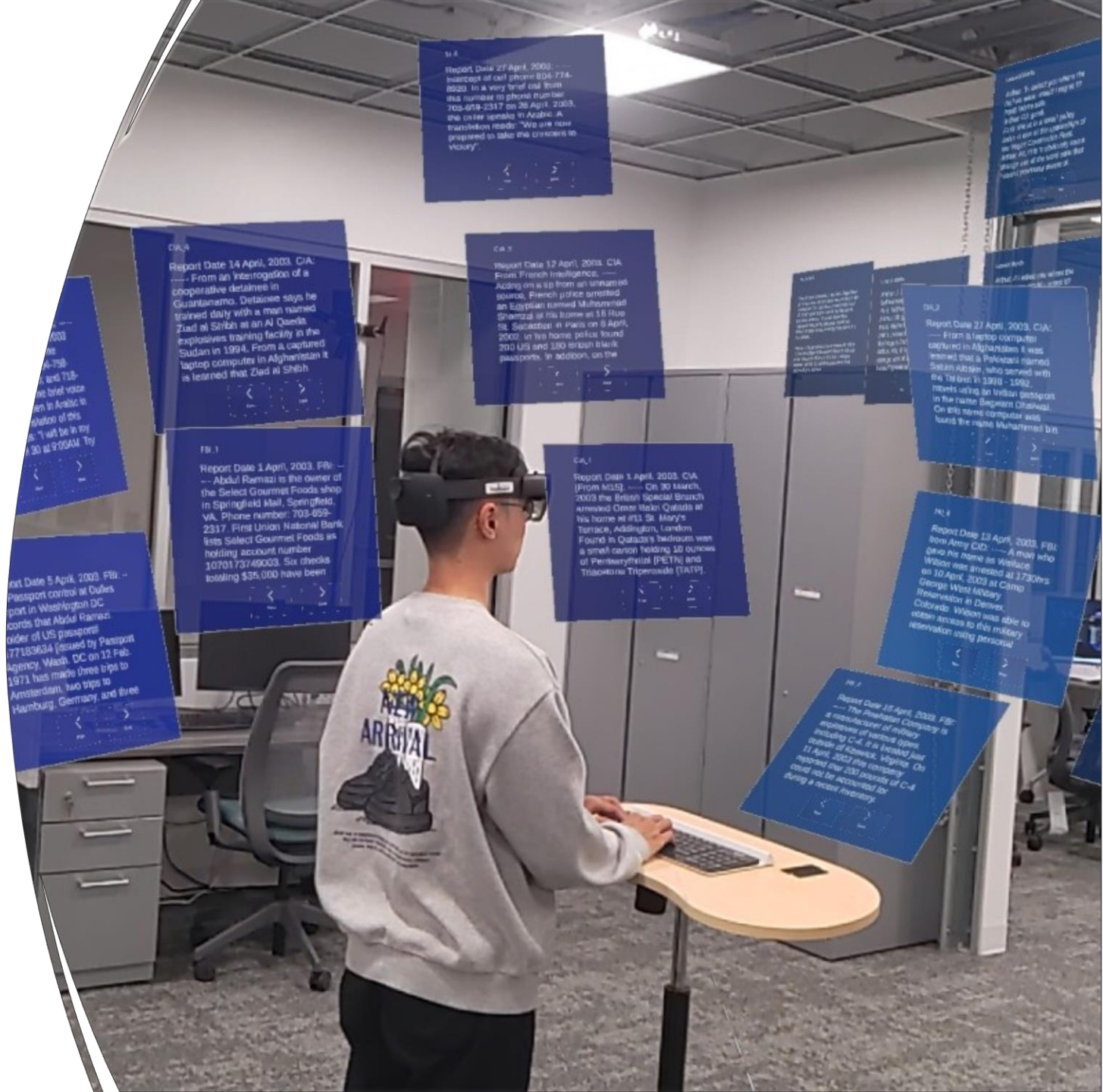
HoloLens 2

Sign of the Crescent
(24 Documents, 2 Terrorist Plots)

Note

Label

Search

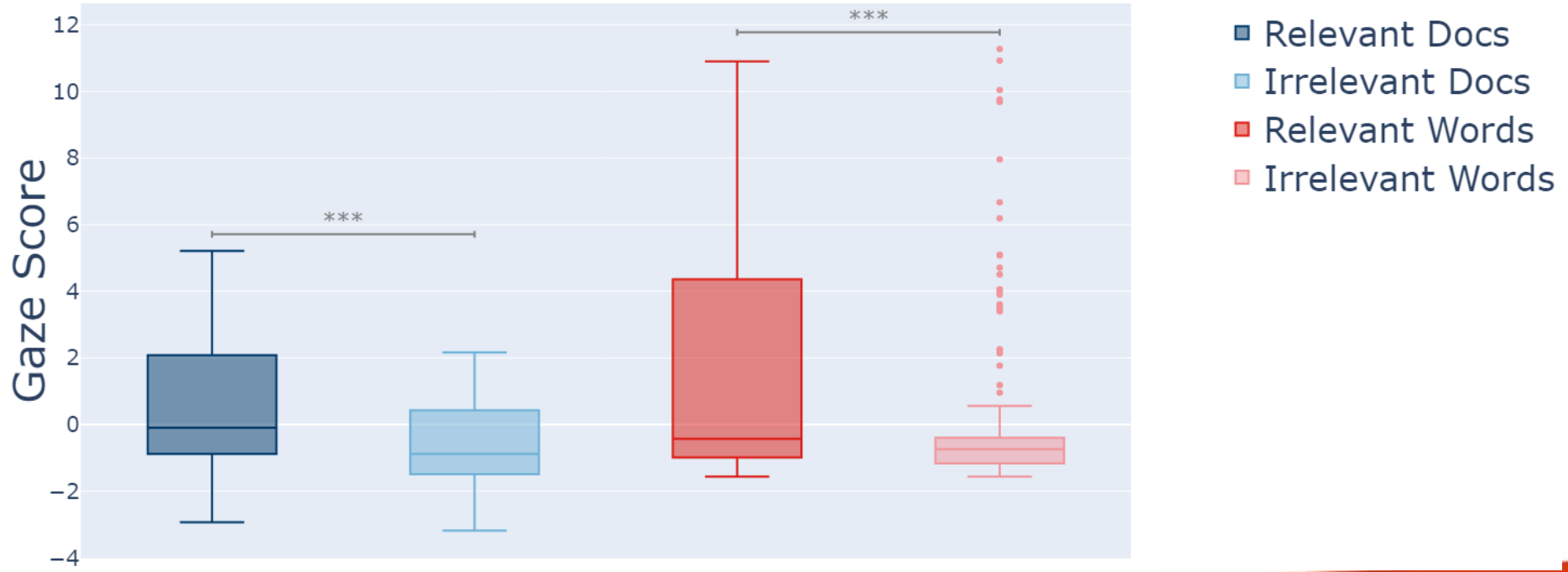


Data Collection

- **Behind the scenes**
  - GazeScores for all documents and words were calculated
  - We sorted them in descending order and derived their GazeRank
- **User Response**
  - **Free Response:** Report relevant documents and words without any knowledge of the GazeRank
  - **Rated Response:** Rate documents and words with different GazeRanks on how relevant they were

Results
GazeScore on Text Relevance

Relevant documents and words have **higher GazeScore** than **irrelevant** documents and words

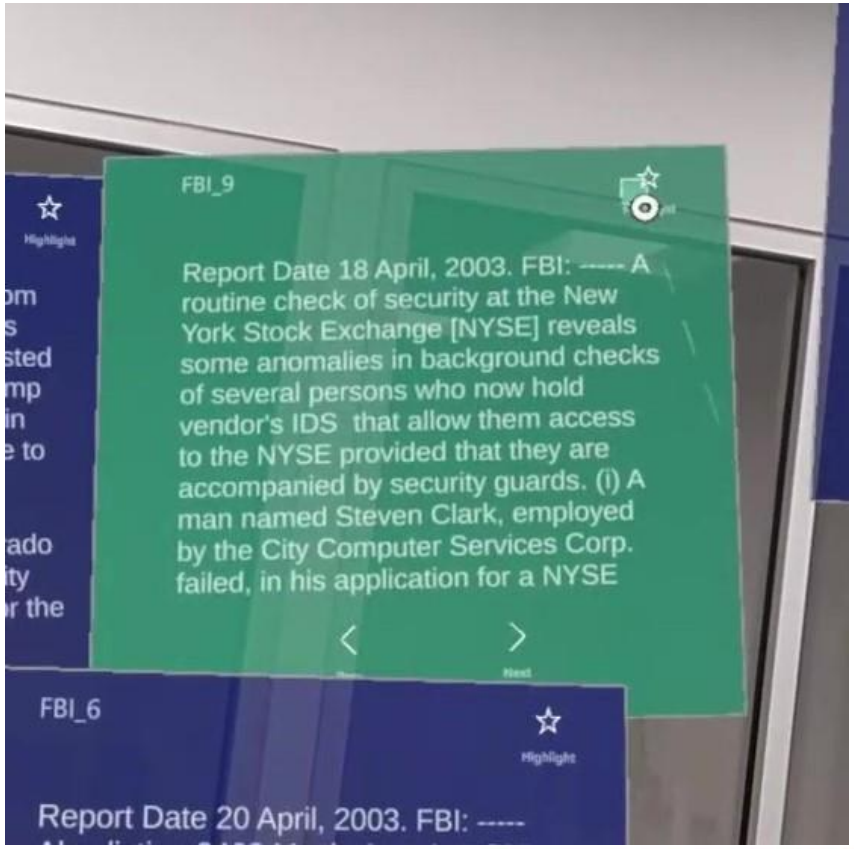


Free Response Documents

Data Collection

Participants highlighted 4 documents to be relevant.

During analysis, we ranked the 4 documents based on their GazeScore.

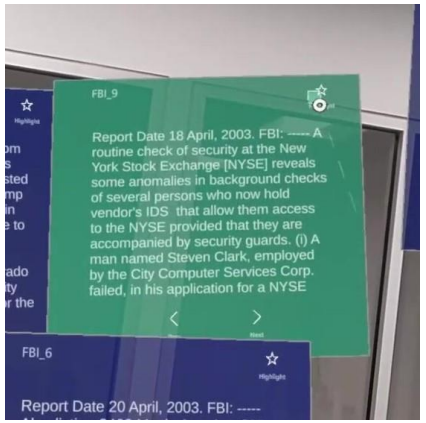


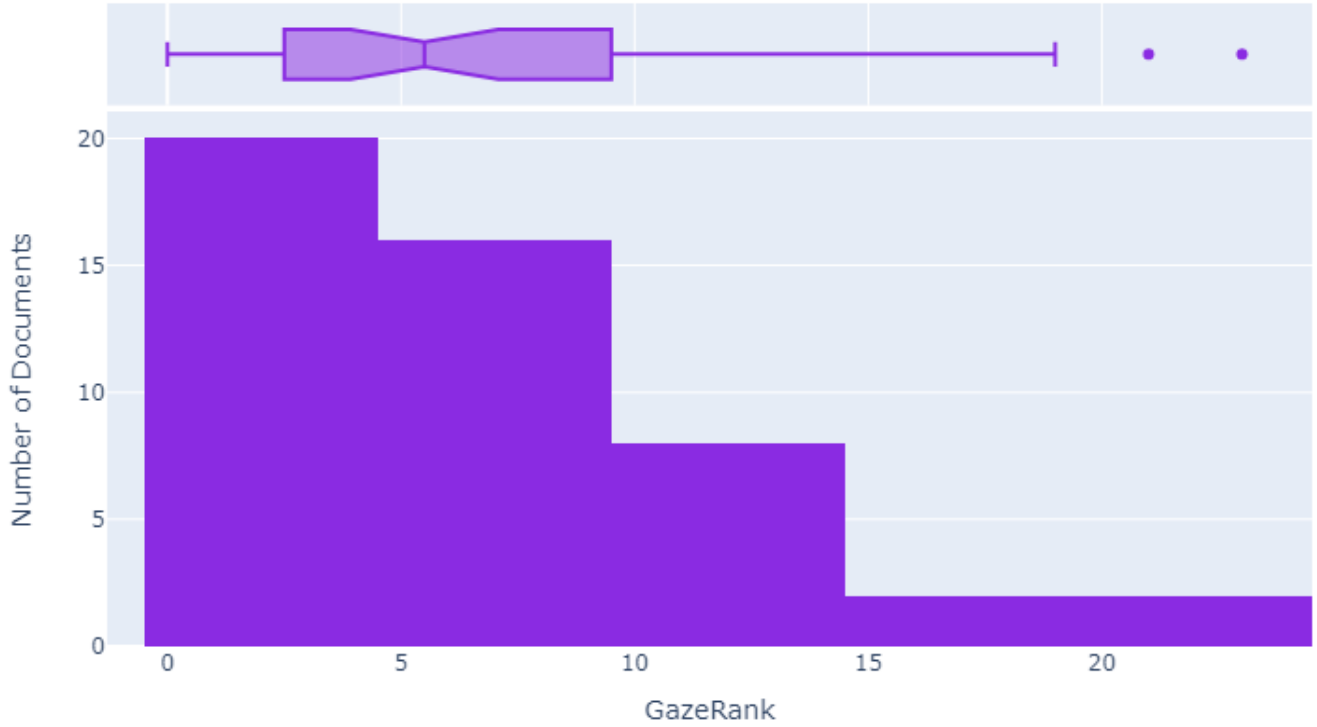
Free Response Documents

Results

Downward slope of GazeRank with median at 5.5.

38% of the documents had ranks 4 or less, compared to 16.67% random chance



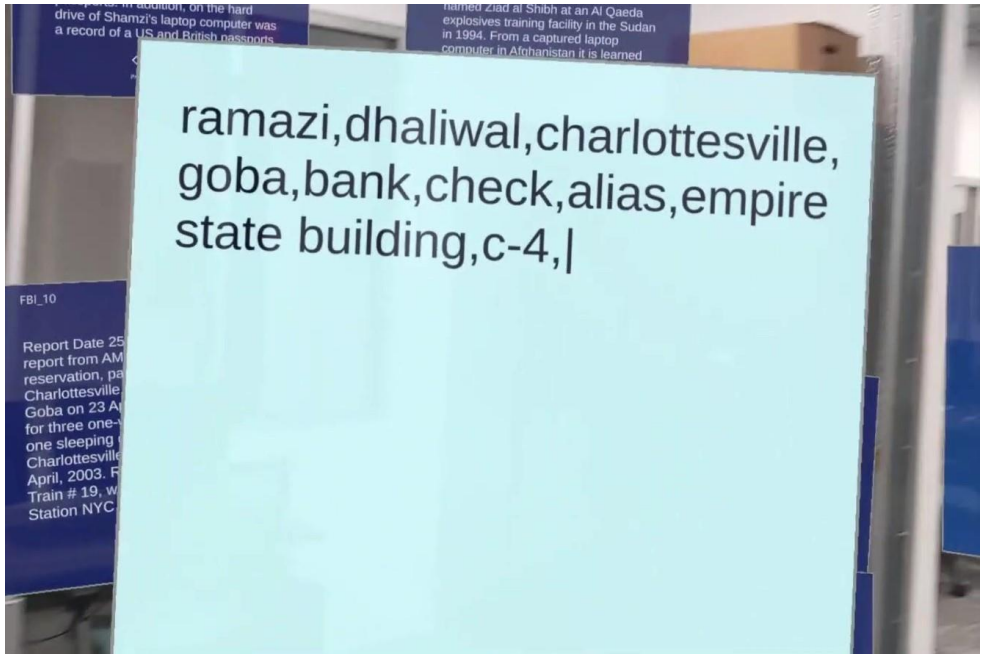


Free Response Words

Data Collection

Participants wrote down 10 keywords as relevant.

During analysis, we ranked the 10 words based on their GazeScore.

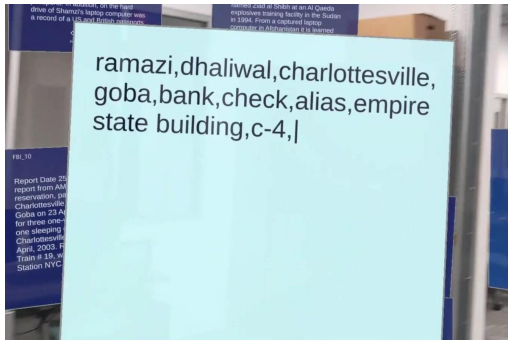


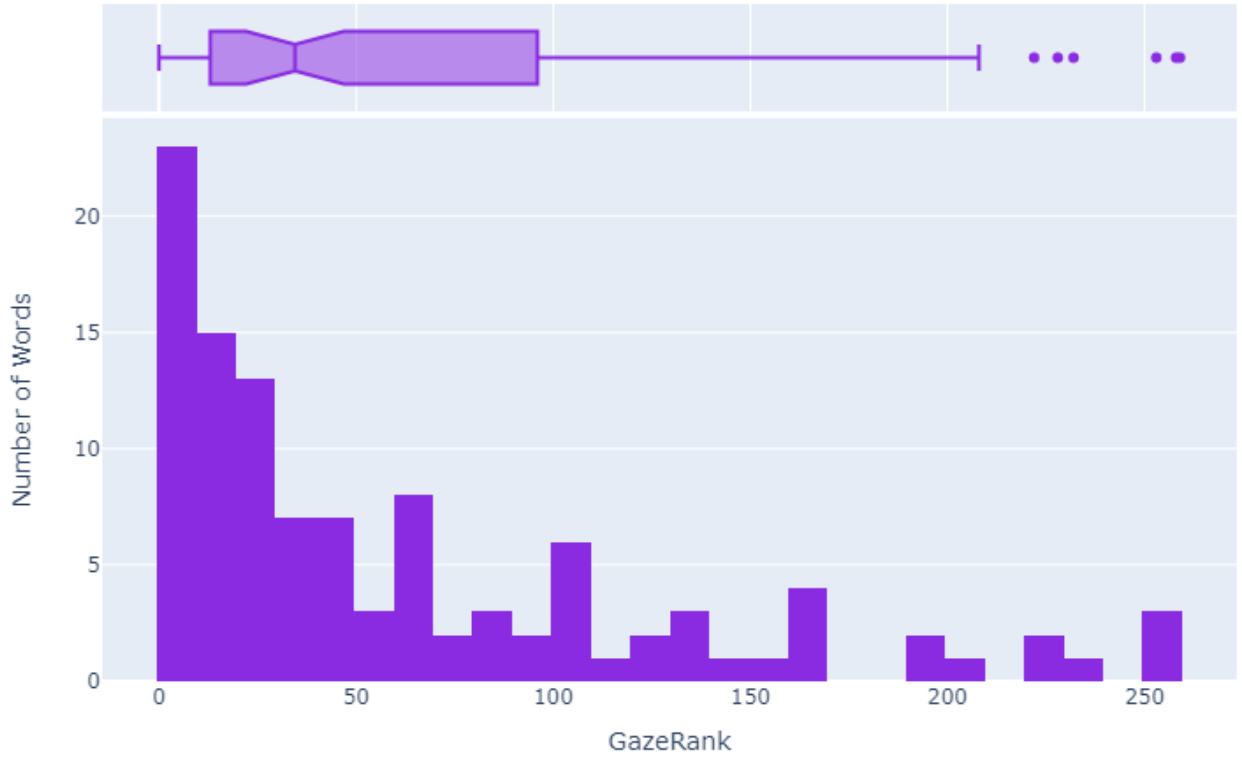
Free Response Words

Results

Downward slope of GazeRank with median at 34.5.

19.17% of the words had ranks 10 or less, compared to 1.9% random chance

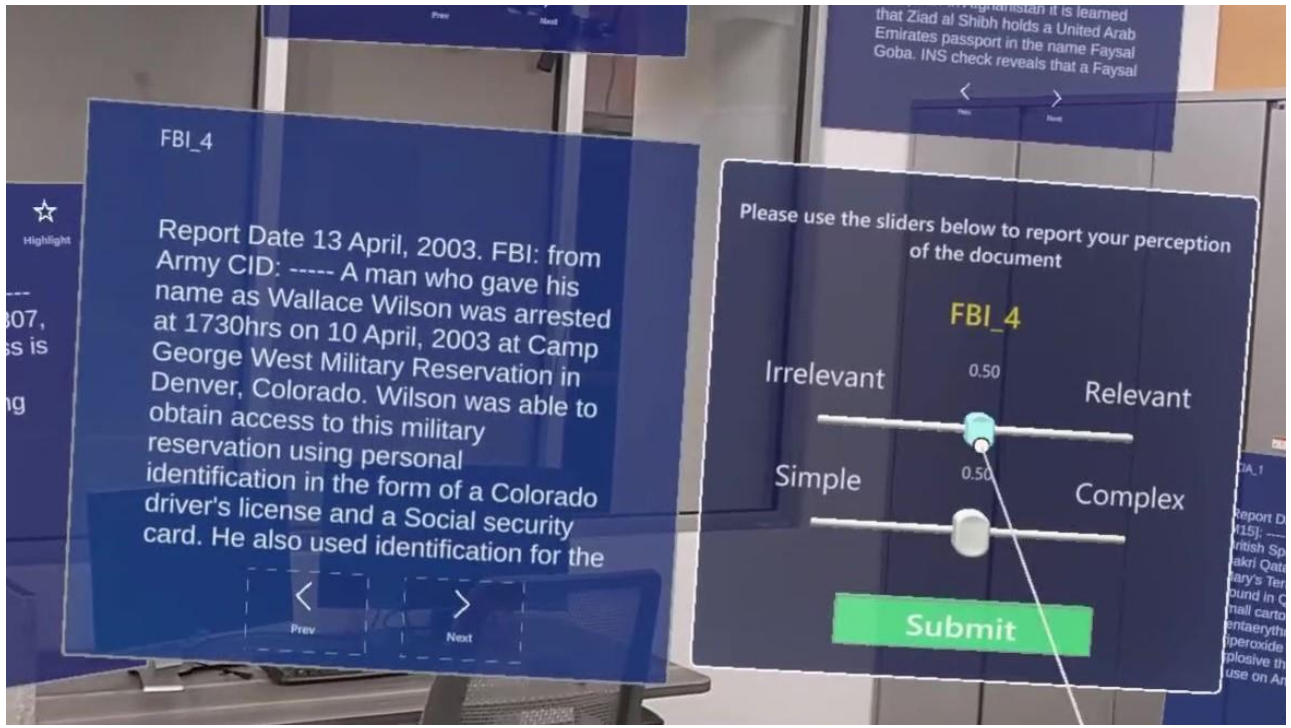




Rated Documents

Data Collection

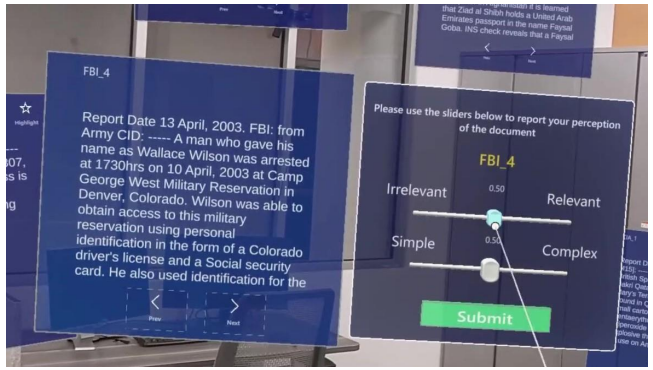
Users rated 12 documents (4 each with high, medium, and low GazeRanks) on their relevance and complexity

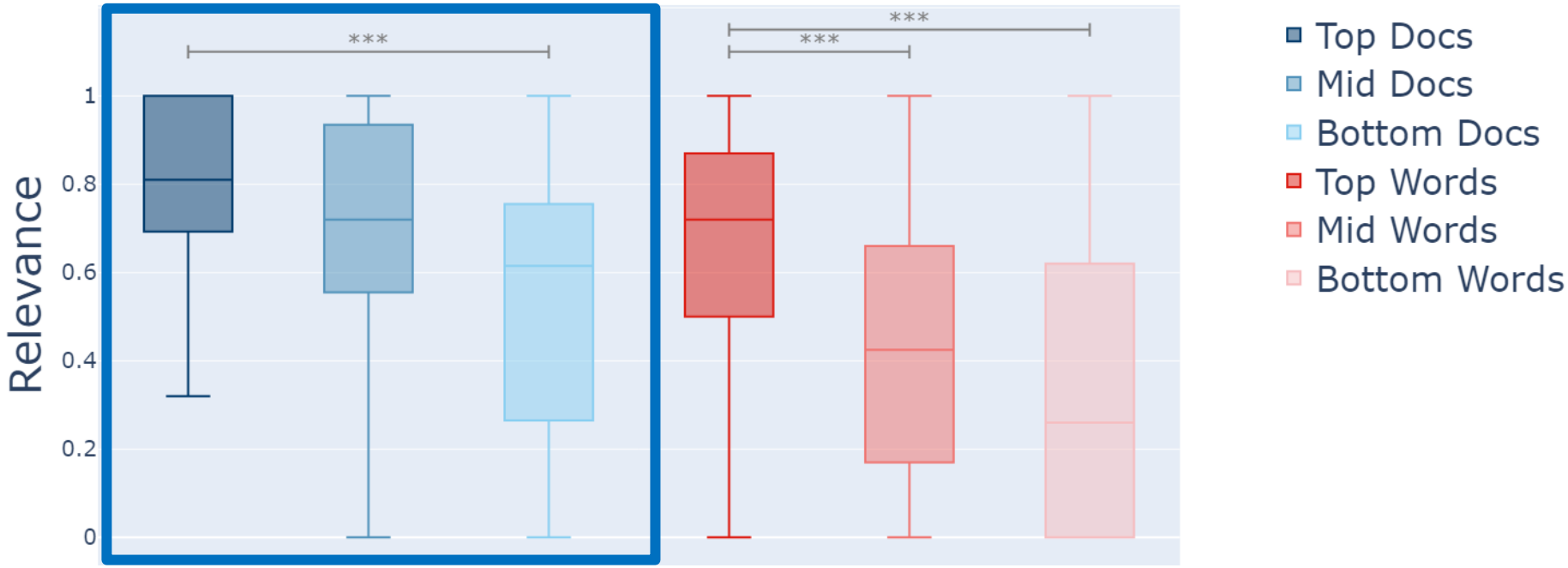


Rated Documents

Results

Documents with high GazeRanks were rated as more relevant by participants than the documents with low GazeRanks

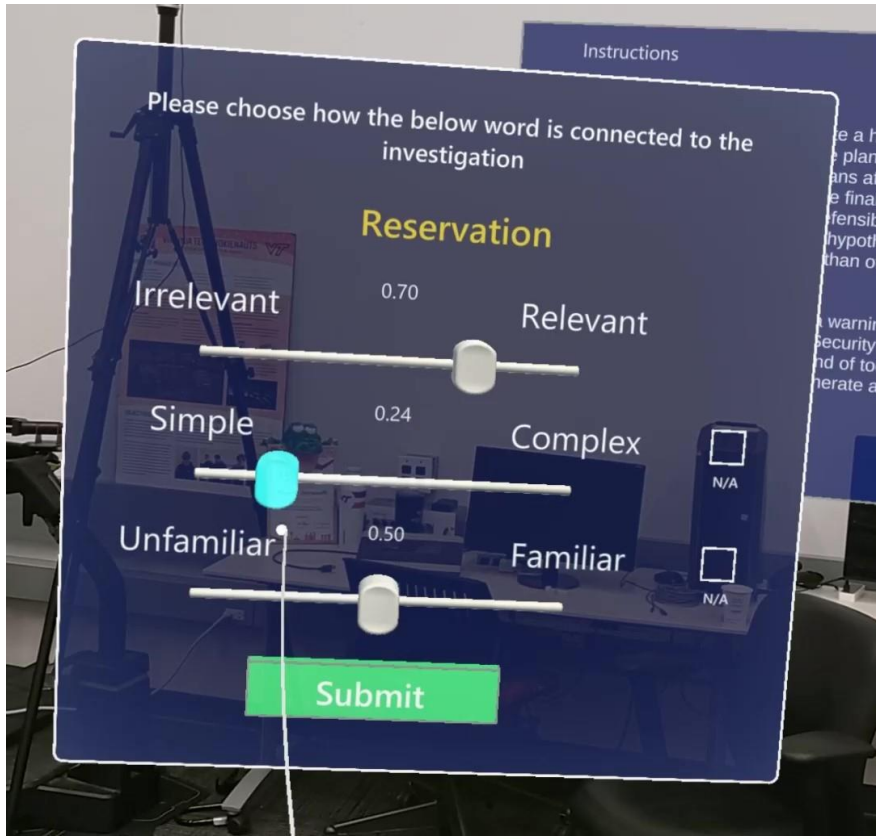




Rated Words

Data Collection

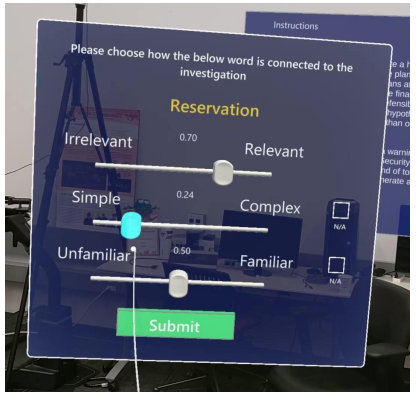
Users rated 30 words (10 each with high, medium, and low GazeRanks) on their familiarity, relevance and complexity

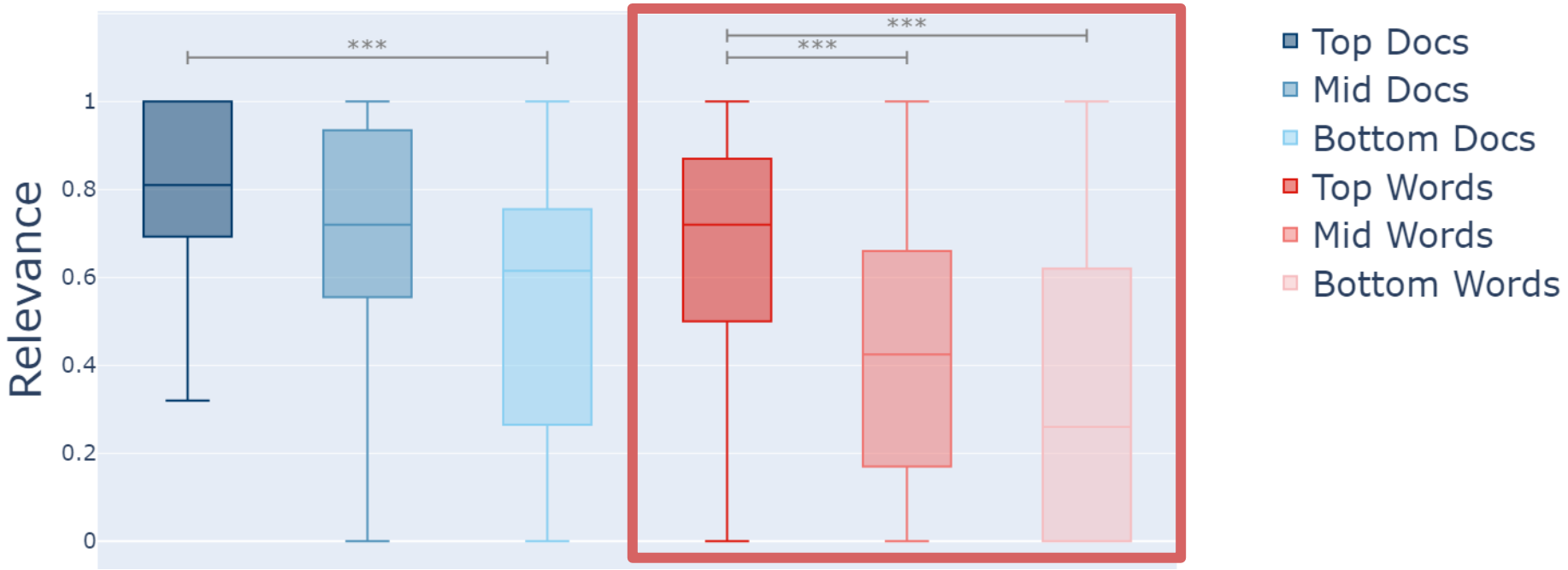


Rated Words

Results

Words with high GazeRanks were rated as more relevant by participants than the documents with medium and low GazeRanks





Research Questions

~~How can we design a gaze metric to capture analyst's focus of attention during sensemaking with multiple text documents?~~

~~To what extent does the metric predict the relevance of documents and words?~~

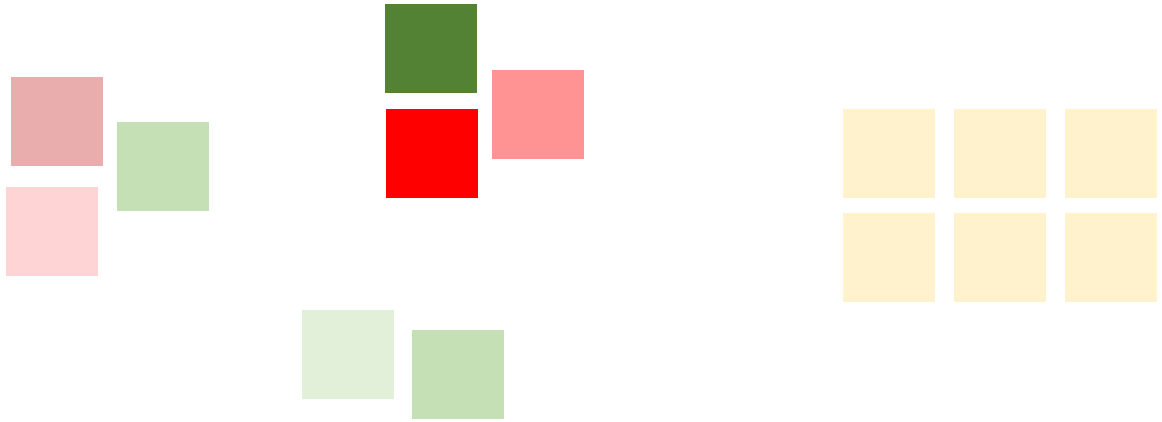
How can we use the metric in a real-time sensemaking task to improve intelligent suggestions to analysts?

Applications of Eye Gaze
Externalization

- Externalization of mental concepts by visualizing attention paid by analysts

Applications of Eye Gaze
Externalization

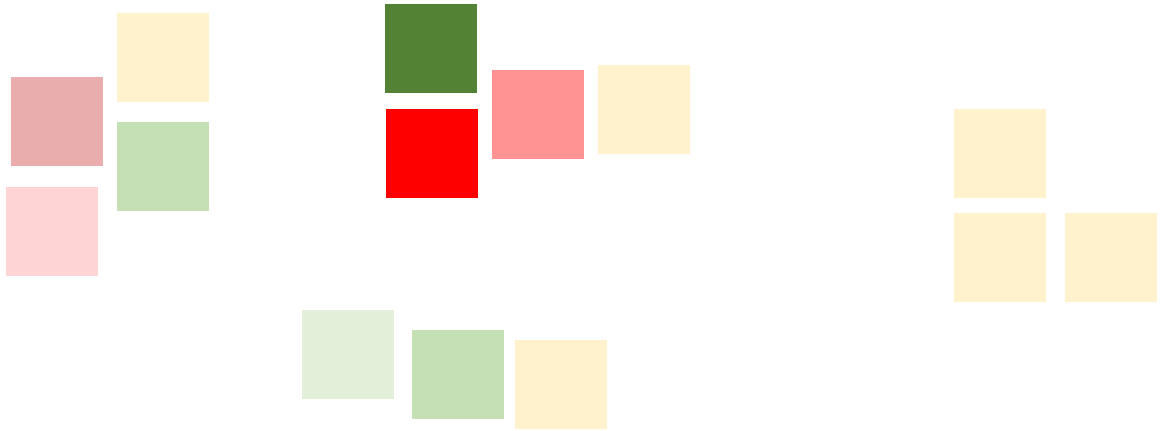
- Externalization of mental concepts by visualizing attention paid by analysts



Applications of Eye Gaze

Relevant Information Retrieval

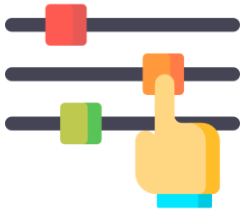
- Externalization of mental concepts by visualizing attention paid by analysts
- Retrieves additional documents of interest to the analyst



Takeaway

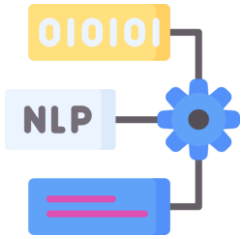
- Eye gaze data can **infer user-perceived relevance** of documents and words during sensemaking with multiple documents
- Gaze data is **more effective** for inferring relevance of **words**, even for small datasets
- Gaze data can be used to develop an **intelligent immersive analytic system** for sensemaking tasks

Future Work



Refinement of the Metric

Dynamic Adaptation of the Model with Human-in-the-Model



Integration with NLP Techniques

Thank You

Scan for Full Paper

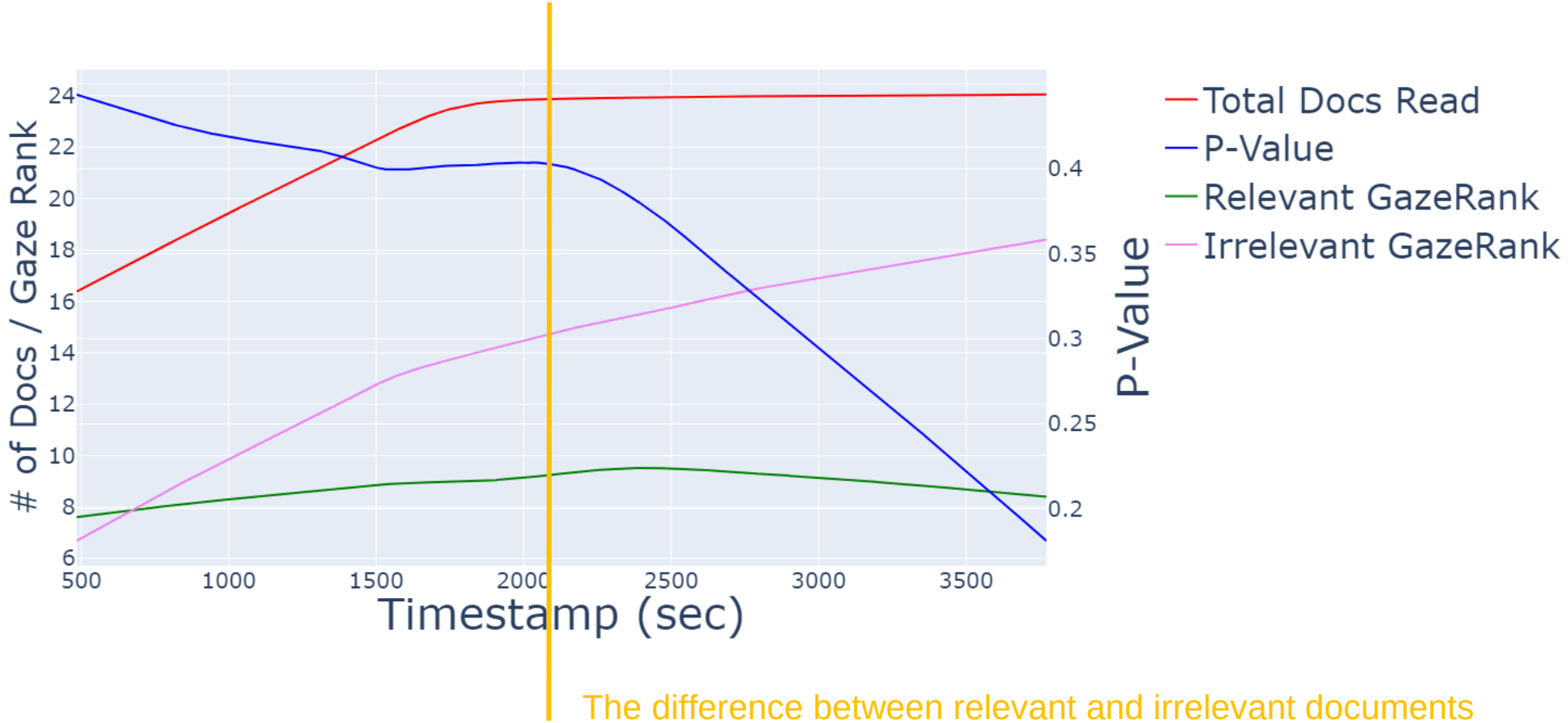
ISMAR 23 SYD
Oct 16 - 20

IEEE

IEEE COMPUTER SOCIETY

IEEE vgtc

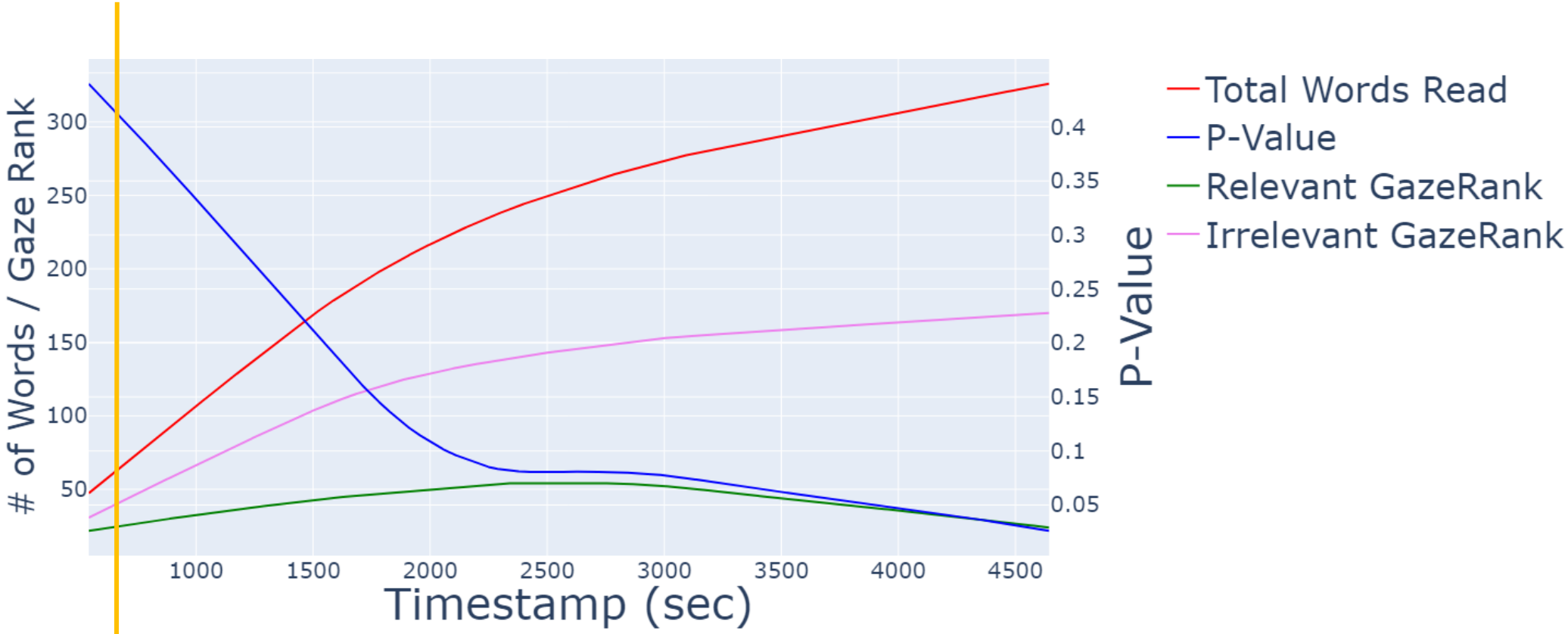
Can Eye Really Read Mind?
Yes and No



The difference between relevant and irrelevant documents start increasing after all the documents are read

Can Eye Really Read Mind?

Yes and No



The difference between relevant and irrelevant words start increasing almost from the beginning

Why GazeScore?

- GazeScore outperforms GazeDurataion for 50% of the documents and 75% of the words
- GazeScore outperforms UniqueDwell for 54% of the documents and 37% of the words
- UniqueDwell does not consider the dataset complexity
- UniqueDwell does not adjust for frequency bias