Integrating Eye Tracking, Feature Use, and Emotional Valence: A Multimodal Approach to Evaluating Search Interfaces

ABBAS PIRMORADI, University of Regina, Canada
ORLAND HOEBER, University of Regina, Canada
MORGAN HARVEY, University of Sheffield, UK
MILAD MOMENI, University of Regina, Canada
DAVID GLEESON, Northumbria University, UK

Interactive Information Retrieval (IIR) interfaces are typically evaluated using questionnaires that gather post-task subjective measures such as ease of use, usefulness, satisfaction, and user engagement, along with in-task objective measures derived from log analysis. However, a comprehensive evaluation requires a deeper understanding of user behaviour beyond such traditional measures. Integrating eye tracking data with logged feature use and emotional valence provides a multimodal approach to evaluating a search interface at the feature level. To validate this approach, we examined three search interfaces in a controlled laboratory study focused on exploratory search within the context of digital humanities archives. A key benefit of this multimodal approach is that it allows us to evaluate both traditional interaction with the search interface (looking at a feature, using it, and experiencing an emotional response) as well as passive interaction with the search interface (looking at a feature, choosing not to use it but possibly getting information from it, and experiencing an emotional response). Using this approach, we were able to identify specific features of the interfaces that generated positive and negative emotional valence responses when used, as well as features that generated such emotional valence responses when viewed but not used. Such feature-level assessments would be difficult to capture using other means, providing insight into the nature of the searchers' experiences using the search interfaces.

CCS Concepts: • **Human-centered computing** → **User studies**; **User interface design**; **Interaction design**; **Affective computing**; *Sensor-based interaction*; • **Information systems** → *Search interfaces*; *Digital libraries and archives*; *Interaction data logging*.

Additional Key Words and Phrases: Search interface user studies, controlled laboratory study, facial emotional detection, emotional valence responses, eye tracking, keyword/result linking, exploratory search, digital humanities

ACM Reference Format:
Abbas Pirmoradi, Orland Hoeber, Morgan Harvey, Milad Momeni, and David Gleeson. 2025. Integrating Eye Tracking, Feature Use, and Emotional Valence: A Multimodal Approach to Evaluating Search Interfaces. 1, 1 (January 2025), 22 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The evaluation of novel search interfaces requires both rigorous scientific methodologies and a realistic approach that considers end users' perspectives [35]. In interactive information retrieval research, evaluation relies heavily on measurement, with some studies focusing specifically on developing and assessing measures for evaluating search interfaces [11, 53, 76]. There are four main categories of such measures: contextual, interaction, performance, and usability. Particularly, the usability category includes evaluative feedback collected from subjective responses, which explore users' perspectives, attitudes, and experiences [35]. These subjective responses are generally collected at the end of a search task, which may not accurately reflect the nature of the interaction that occurred in the midst of the search task [57]

As the academic community has developed a greater understanding of how people search for information, how search is supported in a search user interface has become an increasingly important research area [23]. The design of the search interface influences how users search for, evaluate, and interact with information [68]. There is also an emotional aspect to searching, as documented within stages of Kuhlthau's information seeking process [39]. Such emotional factors play a crucial role as they can greatly affect search behaviour, performance, and evaluation of information relevance [1, 46]. This implies

Authors' Contact Information: Abbas Pirmoradi, Department of Computer Science, University of Regina, Regina, Canada, abbaspirmorady@uregina.ca; Orland Hoeber, Department of Computer Science, University of Regina, Regina, Canada, orland.hoeber@uregina.ca; Morgan Harvey, Information School, University of Sheffield, Sheffield, UK, m.harvey@sheffield.ac.uk; Milad Momeni, Department of Computer Science, University of Regina, Regina, Canada, miladmomeni@uregina.ca; David Gleeson, Department of Humanities, Northumbria University, Newcastle, UK, david.gleeson@northumbria.ac.uk.

that emotional factors should be considered when evaluating search systems [21]. While studies of emotional responses to search interfaces are typically done at a system level [7, 45], linking these emotions to feature-level use will provide more granular insight into what aspects of the interface caused the emotional responses.

This study introduces a novel method that builds upon our previous work to develop a framework for linking emotional valence (measured via facial emotional expressions classified as positive and negative) to feature use (measured via user interaction logs) [57]. By incorporating eye-tracking data into this framework, we provide a multimodal view of user behaviour throughout a search session and a basis for enhancing search interface evaluation methods. A key novelty of this work is the ability to map emotional valence not only to features that are actively used (typical interaction) but also to those that are viewed but not interacted with (passive interaction). The eye tracking data for these passively viewed features becomes crucial, as they are often missed in traditional log-emotion analysis [1, 46], yet may convey important information especially when the search interface is enhanced with visualization techniques [26, 27]. This extension enables us to use these combined data sources to evaluate feature-level interaction of search interfaces, supplementing typical post-task evaluation data.

In order to validate this approach, we have used it to study an set of exploratory search interfaces designed to support searching within a digital humanities archive (Europeana [17]). An evaluation of these interfaces using typical research methods found that providing searchers with mechanisms to interactively show the relationships between resources was valuable both when evaluating resources returned from a query, and when considering what has been saved in the workspace [28]. This paper extends that work by incorporating eye-tracking data with user interaction logs and emotional valence data, providing deeper insight into the value of specific features of the interfaces (both from a typical interaction perspective and that of passive interaction) than what is possible with typical user study methods. To this end, this research was guided by the following research questions:

- RQ1: How does the integration of eye tracking, feature use, and emotional valence enable the evaluation of search interface features used via typical interaction?
- RQ2: How does the integration of eye tracking, feature use, and emotional valence enable the evaluation of search interface features used via passive interaction?

The structure of this paper is as follows: Section 2 reviews the existing literature that serves as the foundation for this research. Section 3 describes the three digital humanities archives search interfaces used in this study. Section 4 details the procedures for integrating eye tracking, feature use, and emotion valence data. Section 5 describes the study design, tasks, measures, participants, and procedures employed to gather data to address the research questions. Section 7 reports this study's findings. Section 8 discusses the implications of the results. Finally, section 9 summarizes the key contributions and outlines directions for future research.

2 Background and Related Work

2.1 Eye Tracking

The primary purpose of an eye tracking system is to provide constant real-time information about where a user's attention is focused and insight into their mental state [43]. Attention allocation can be studied by observing the duration, number, and amount of eye fixations (re-fixing a particular area or object multiple times), which may reveal patterns of how they focus their attention in a particular area [43]. Users' mental state can be indicated by their fixation positions, fixation durations, and saccade lengths (the distance th gaze moves between fixations) [43]. Eye tracking is becoming an increasingly common method for collecting data in usability studies. It is typically used in these studies in combination with supplementary data collection methods, such as post-task questionnaires [38], think-aloud protocols [15, 41], and emotion detection tools [62].

Eye tracking has been widely used in library and information science studies to explore user behaviour in information retrieval processes and identify usability issues within digital systems [4, 8, 44]. A study of user interactions with image search result lists examined the relationship between relevance judgments, fixation durations, and object-clicking patterns, which found that users examine more relevant results for longer periods of time [75]. Bibliographic information systems have also been studied with eye tracking to analyze specific interface components, such as faceted navigation [36, 41], and to better understand user search behaviour [13, 70]. Similarly, eye tracking has been used to assess how cognitive factors such as topical literacy influence attention when seeking information online [67].

This approach has also been used in the realm of digital humanities archives, where the method has been useful for evaluating user interactions, usability issues, and search behaviour within the Europeana collection [38, 65]. The primary advantage of eye tracking studies is their ability to pinpoint specific areas of user interface challenges, thereby providing insights that extend beyond a general user experience evaluation.

2.2 Feature Use

User interaction log data serves as the foundation for identifying feature use in a search system. At the most fundamental level, interaction logs can be used to identify queries issued and search results accessed [31]. Client-side logging enables the collection of detailed user interactions, including mouse movements, clicks, and keyboard inputs, directly within the browser without requiring server access [3]. A more advanced tool like LogUI, which captures timestamps and contextual information along with detailed sequences of user actions, offers researchers a rich dataset for analyzing feature use and navigation patterns over time [49]. Further, browser extensions and plugins have been developed to continuously log user interactions across sessions, allowing researchers to capture user interactions in various web-based applications [14]. Event-based logging systems enhance the granularity of logged data by triggering specific events when interactions occur (e.g., clicking a button or submitting a query) by capturing the action and its temporal context [69]. Researchers can use these tools to analyze user behaviour over time and gain valuable insights into how users interact with search systems.

In interactive information retrieval studies, understanding user actions such as click patterns and feature use is key to understanding how users interact with search systems and improving the user experience. A wide range of user activities can be captured in interaction logs, including query submissions, document clicks, and navigation between search results, which enables researchers to determine how users utilize different search interface features [42]. Analyzing interaction logs can provide a deeper understanding of user preferences and search strategies, allowing for a more granular evaluation of interface features [51]. Such analysis may also help to identify usability challenges, such as overlooked or misunderstood features, and suggesting improvements to interface design and evaluation [5].

2.3 Emotional Valence

Emotional valence data collection typically relies on observational or physiological methods such as facial expression analysis [1, 46, 58], skin conductance and heart rate [50], or combined facial, voice, and brain wave evidence [2]. Facial expressions are one of the most important ways to express emotions [16, 55]. Facial expression analysis provides a non-intrusive, real-time method to assess emotional states without interrupting user tasks or requiring additional devices. In our prior work [57], emotional responses were captured in real-time using facial emotion recognition technology based on Ekman's Facial Action Coding System (FACS) [16], which identifies emotional states based on facial muscle movements; we translate these into emotional valence by grouping the positive and negative emotion states.

Emotional responses and emotional valence have increasingly been studied in the context of interactive information retrieval to understand how emotions influence search behaviour and interface interactions [19, 46, 47]. Information behaviour is influenced by affective factors, with emotions being fundamental to how users find and use information [52]. A comparison of Kuhlthau [39]'s and O'Brien [52]'s models shows that cognitive and affective factors are both important to understanding information behaviour [59]. An extensive review of studies across Library and Information Science, Information Retrieval, and Human-Computer Interaction emphasizes the need to integrate emotional intelligence into search interfaces [47]. An evaluation of specific design strategies on emotions was conducted to improve the design of search interfaces and their evaluation [45]. There have been studies that classify interface attributes, including information presentation, navigation/orientation, text, and visual elements, examining their potential emotional impacts [73, 79]. Additionally, a psychological framework has been used to explore how user interfaces in mobile libraries affect emotions [78]. This highlights the importance of emotional factors when designing search interfaces. Emotional design methodologies emphasize the importance of designing search interfaces that consider users' psychological and emotional needs [19].

2.4 Integrating Approaches

Integrating eye tracking and feature use data, researchers have explored how users interact with search interfaces, tracking both their actions and visual attention. This combination allows the study of the relationships between visual attention and click behaviour, revealing insights into how users assess search result relevance, and which parts of the SERP capture attention but may not be clicked [20]. Similar studies have been conducted on query auto-completion, capturing where searchers focus their attention and how they interact with suggestions [30]. Integrating eye tracking data with interaction logs can also improve the evaluation of interactive visualization systems [9]. Combining these methods provides insight into user attention and interaction patterns, enabling a better understanding of system usability.

Research that integrates emotional valence with eye tracking data is limited, but holds the promise of providing insight into how visual attention on search interface features correlates with emotional responses. In the context of the digital humanities, one study found that linking emotional valence with eye-tracking data can help identify elements of search interfaces that engage and attract users [38]. Other work investigated information-seeking behaviour in public digital libraries using eye-tracking data and emotional valence to analyze the interaction between different age groups with an e-book wall search interface Wu and Huang [74].

Studies have also been conducted examining the relationship between emotional responses and feature use, through logging user activities. Emotional responses and feature use data were collected during an exploratory user study on the role of emotions in the information-seeking process [1]. Using a digital library search interface, another study combined emotional responses and feature use with a digital library search interface to understand the relationship between primary and secondary emotions, mood, and the online search process [46]. Our prior study integrated emotional valence with feature use data to identify specific interactive features used during the experience of positive and negative facial expression of emotion [57]

While previous literature has examined emotional valence, eye tracking, and feature use data individually or in combinations of pairs of these methods, the comprehensive integration of all three methods in the context of evaluating search interfaces is novel.

3 Europeana Search Interface Design

Europeana is a digital multimedia repository that offers access to a multitude of cultural artifacts, including art, historical documents, and multimedia resources [17]. Researchers, professionals, and the general public can access this collection through a typical search interface [25]. However, such typical search interfaces may not provide adequate support for undertaking complex search tasks and employing exploratory search strategies.

In previous work, we studied two new approaches that allow searchers of the Europeana collection to interactively link related search results through their common use of keywords using visualization methods, both in the search results list and among the saved resources in the workspace [29]. These approaches were designed to support undertaking complex search scenarios using exploratory search processes, and were compared to a baseline interface using typical interactive information retrieval research methods. As the current study is an extension of this prior work, we provide a brief description of the interfaces for the sake of completeness.

3.1 Baseline (B) Interface

The Baseline search interface was developed using Europeana's existing interface as a template and the Europeana Search API as a data source [18]. A few style adjustments were made, such as separating metadata from actions for each search result (e.g., relocating the like and save buttons to the bottom-left corner of the search result card) and removing irrelevant features (e.g., collection suggestions). In order to ensure consistency, a caching mechanism was implemented so that all participants received identical search results. Screenshots of the Baseline search interface and workspace are shown in Figure 1.

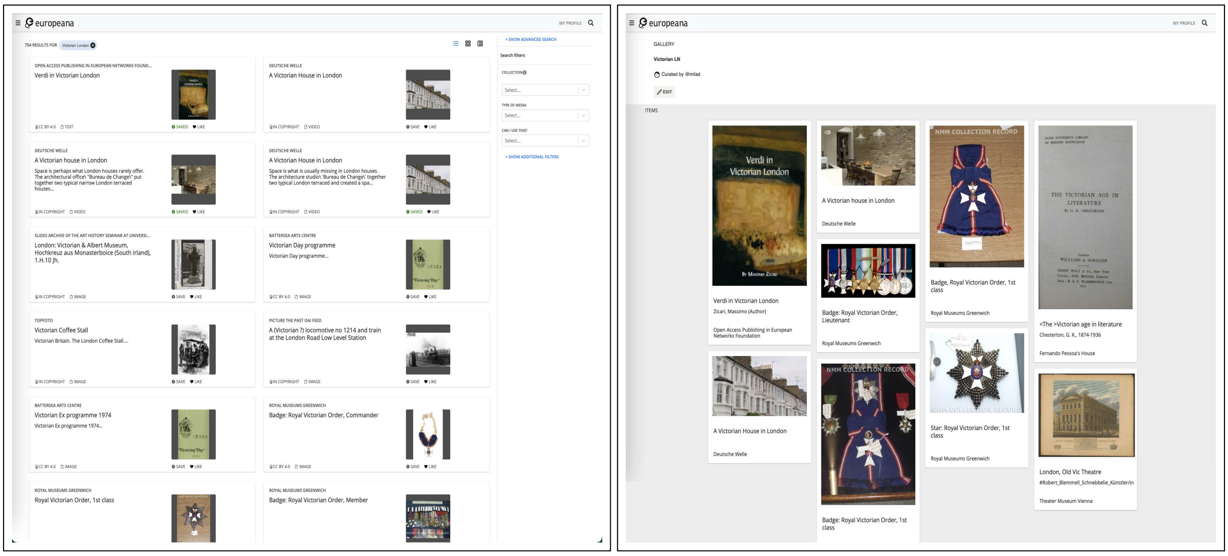
3.2 Result-Focused (RF) Interface

This interface builds on the basic search functionality by incorporating visual keyword highlighting, a feature that allows users to select keywords of interest which are then highlighted across all the search results [29]. The keyword selected by the user is colour-coded, and all occurrences of the keyword throughout the results are highlighted in the same colour. As a result of this visual connection, users can see patterns and relationships between different results based on shared keywords, facilitating an exploratory browsing experience. As the keywords remain connected to the results, we label this approach as Result-Focused.

With a slight modification to the layout, this approach is replicated in the workspace, which makes keywords associated with search results interactive for uncovering relationships. Screenshots of the Result-Focused interface can be seen in Figure 2, which shows both the search results page and workspace with a subset of selected keywords.

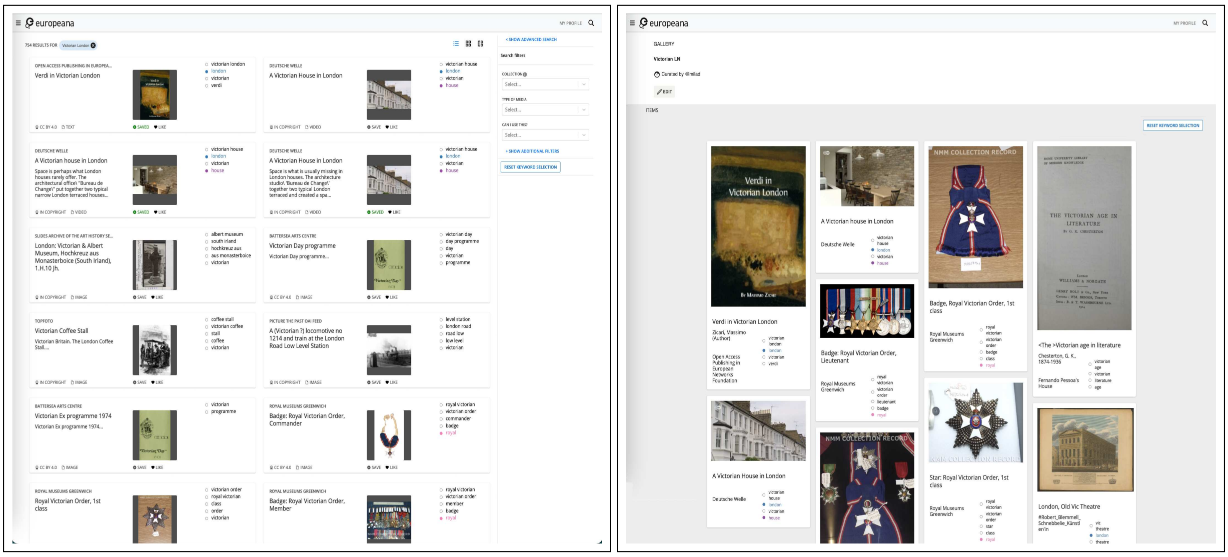
3.3 SERP-Focused (SF) Interface

This approach aggregates the keywords for all search results on the current search engine results page (SERP), displaying them alongside the search results set. These keywords provide an overview of the current search results, allowing searchers to quickly scan this information without having to scroll through individual results. The list allows users to highlight specific



(a) Search Interface (b) Workspace

Fig. 1. The Baseline search interface and workspace are designed to closely resemble the current Europeana search interface.



(a) Search Interface (b) Workspace

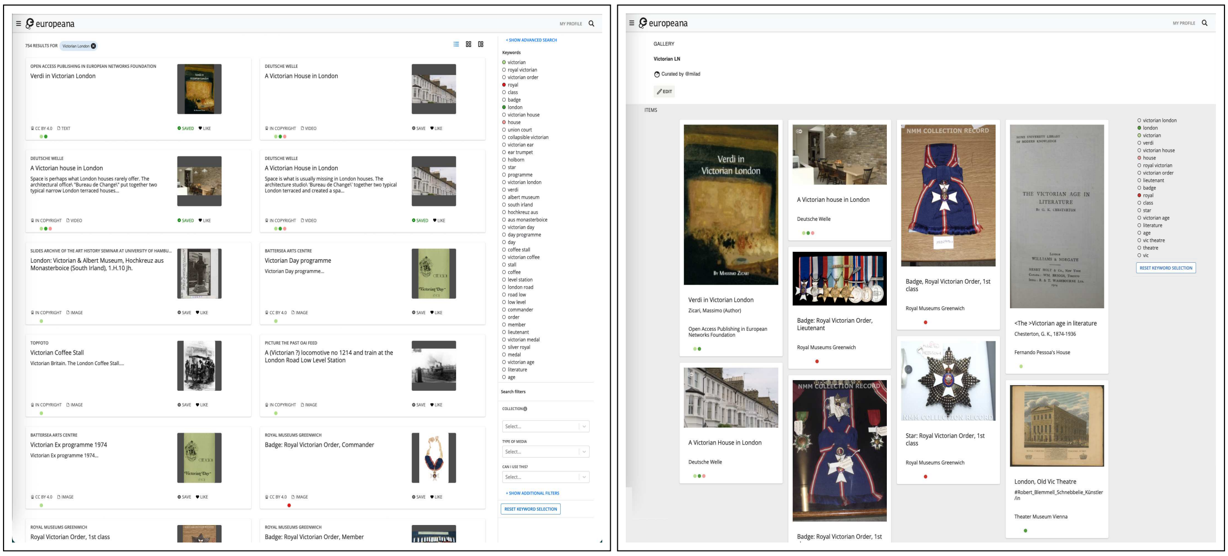
Fig. 2. In the Result-Focused interface, keywords are displayed as a list next to each search result. Colour-encoded discs beside each keyword highlight the same keywords in other search results after the user selects keywords.

keywords, and these keywords are colour-coded with their corresponding search results, making it possible to identify related content [29]. As the keywords are provided for the entire SERP, we call this approach SERP-Focused.

In the workspace, this SERP-focused method is also used, with the keyword list derived from the saved search results. In Figure 3, screenshots of the SERP-focused interface are shown, along with a small list of keywords selected for the search results page and workspace.

4 Integration of Eye Tracking, Feature Use, and Emotional Valence

Within the scope of this research, our approach focuses on three measures used in prior information-seeking research: eye tracking, feature use, and emotional valence [1, 38, 46]. While eye tracking and feature use data can be considered interaction data as they reflect user behaviour and engagement with the interface, emotional valence data are reactions to users' interactions with the system. Assessing facial emotion expressions classified as emotional valence, and their connection to eye tracking data and feature use, is the primary focus of this study.



(a) Search Interface (b) Workspace

Fig. 3. The SERP-Focused interface provides a unified list of keywords for the currently shown search results. It allows searchers to choose keywords to highlight their source result using colour-coded discs beneath the corresponding search result.

In our prior work, we demonstrated the feasibility of matching emotional valence to feature use [57]. The novel contribution of this work is the integration of eye tracking data with feature use and emotional valence data, enabling the evaluation of both features that are interacted with in the typical way (e.g., click, select) and features that are interacted with in a passive way (e.g., visual attention without a corresponding typical interaction).

4.1 Eye Tracking

Pupil Labs' eye tracking glasses were used in this study [34]. The screen resolution was 1920x1080 pixels. We recorded gaze samples at a frequency of 120Hz. A pre-processing analysis was performed using Pupil Player software from Pupil Labs. Following that, all raw data was exported as CSV files for further analysis. In cases where the eye could not be detected, either through tracking errors or blinking, we excluded such gaze samples. All gaze points associated with a specific fixation were aggregated into one fixation point positioned at the center of the associated gaze points. We defined areas of interest (AOIs) for specific interface features of the interfaces under investigation (e.g., search box, keywords, buttons), identifying when the fixation point was within these feature-based AOIs. We considered the following measurements: the fixations on each AOI, the dwell time on the AOI (i.e., the total time spent fixated at an AOI), and the time stamp of the start of the fixation.

4.2 Logging Search Activities

Participants' interactions with the search interfaces were recorded using the LogUI JavaScript library [49]. We integrated this library into all three search interfaces. To track user activities via the library, we set up the search interface features in the configuration object. In order to facilitate the analysis of the feature use, we classified each one according to Wilson's taxonomy of search interface features [73]. In Table 1, the complete list of interface features available in all three interfaces is provided, organized according to Wilson's taxonomy.

4.3 Measuring Facial Emotion Expressions

Front-facing cameras on participants' computers were used to capture video frames. The search interface collects these frames using a custom JavaScript library developed for our previous work [57]. Upon loading the search interface, the front-facing camera is activated. Web Real-Time Communication (WebRTC) [56] establishes a peer-to-peer connection between the web browser and our secure server. Video frames are transmitted from the clients' browsers to the server using HTTPS, without requiring plugins or additional software on users' devices. To balance transfer speeds and storage requirements, a frame rate of 5 frames per second was chosen.

The DeepFace framework is an effective tool for identifying faces in video frames and extracting emotional responses [61]. This study utilized the guided back propagation method embedded in the HyperExtended LightFace (commonly known as

Table 1. Classification of search interface features according to Wilson's taxonomy [73]. The last column indicates which of the interfaces has each of the features: Baseline (B), Result-Focused (RF), SERP-Focused (SF).

Interaction Type	Feature	Search Action	Search Interface
Input	Query input box	Submit query	B, RF, SF
Control	Search results navigation Search filters	Click on pagination icon Select value from search filter	B, RF, SF
Informational	View search result View search result in Workspace	Click on search result Click on search result in workspace	B, RF, SF
Personalization	Like result	Click like result	B, RF, SF
	Unlike result	Click unlike result	B, RF, SF
	Save result	Click save result	B, RF, SF
	Go to workspace	Click "My Profile"	B, RF, SF
	Change view	Click change layout	B, RF, SF
	Create gallery	Click create gallery	B, RF, SF
	Add result to gallery	Click add result to gallery	B, RF, SF
	Like result in workspace	Click like results in workspace	B, RF, SF
	Remove liked result in workspace	Click unlike results in workspace	B, RF, SF
	Remove result from gallery	Click on remove result in workspace	B, RF, SF
	Delete gallery	Click delete gallery in workspace	B, RF, SF
	Highlight keyword	Click highlight keyword	RF, SF
	Unhighlight keyword	Click unhighlight keyword	RF, SF
	Reset selection	Click reset button	RF, SF
	Highlight keyword in workspace	Click highlight keyword in workspace	RF, SF
	Unhighlight keyword in workspace	Click unhighlight keyword in workspace	RF, SF
	Reset selection in workspace	Click reset in workspace	RF, SF

DeepFace) implementation [61] for emotion recognition. This open-source, offline tool allows for the discrete analysis of emotions displayed in facial video frames without requiring additional training. A combination of 60% accuracy in emotion detection and 73% accuracy in happiness and surprise [61] makes DeepFace one of the most accurate facial emotion detection approaches available. This package incorporates well-studied models such as VGG-Face [54], FaceNet [60], OpenFace [6], DeepFace [66], DeepID [63, 64], and Dlib [37]. The facial attribute analysis module classifies video frames in seven emotion categories: anger, fear, neutral, sadness, disgust, happiness, and surprise. Several studies in the wider literature have employed DeepFace to integrate emotional data into visual question answering models [12], to anonymize faces [24], analyze learners' emotions [77], and identify emotional states through facial expressions in photo and video materials containing human faces [33]. DeepFace's widespread adoption highlights its flexibility and effectiveness.

Each frame is assigned to one of the seven emotional categories in the following way. DeepFace's emotion recognition module generates a vector consisting of eight scores, each ranging from 0 to 100. Seven of these scores indicate the confidence level for each emotion category, while the eighth score represents the confidence level for face detection in each frame. To determine whether a frame should be classified based on its highest emotion score, a threshold of 90 is used ("emotion confidence threshold"). Similarly, a threshold of 90 is applied to face confidence ("face detection confidence threshold"). The cut-off values are based on benchmarks established in similar emotion detection research [1, 46]. The frame is categorized as neutral if the emotion score falls below the "emotion confidence threshold" or if the highest emotion score is neutral. The analysis also excludes frames that did not meet the "face detection confidence threshold". Time stamps are also included for each frame to record the exact time it was captured. Of note, in this study only 176 frames were excluded due to unmet thresholds.

To focus shift the focus to emotional valence, all positive emotions (happiness and surprise) were aggregated into a single positive valence category, and all negative emotions (anger, fear, sadness, and disgust) were grouped into a single negative valence category. This approach enables the analysis of broader emotional valence while minimizing the potential impact of misclassified extreme emotions.

4.4 Matching Data

In prior work, we developed a method to capture real-time emotional valence during search tasks and synchronize this with user interactions [57]. Building on this work, we now incorporate eye tracking data to enhance the analysis.

First, we measured the emotional valence associated with features that participants both viewed and interacted with (typical interaction). Second, we assessed emotional valence related to features that participants viewed but chose not to use (passive interaction). Such passive interactions are gaze-only interactions where participants look at a feature but choose not to use it. These two approaches are summarized in Figure 4 using set theory notation to highlight which data is combined and which data is removed, isolating typical interactions from the passive interactions. The details of these two approaches are explained in more detail below.

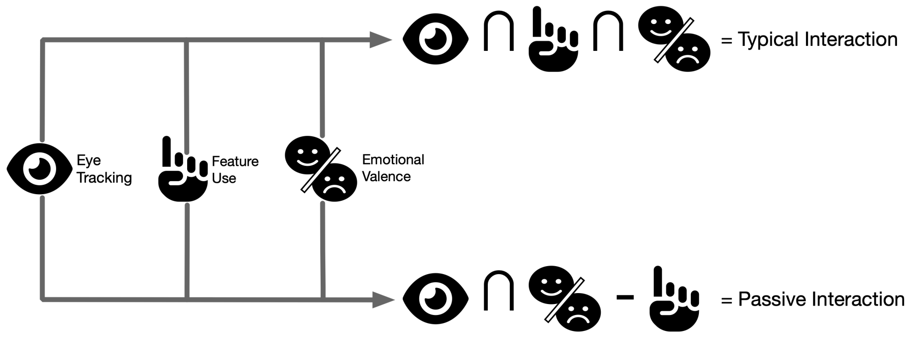


Fig. 4. The approach for integrating eye tracking, feature use, and emotional response data to measure typical interaction and passive interaction.

For the first approach, we aligned emotional responses with LogUI interaction data to measure how emotional states change before and after each feature use [57]. We calculated the probabilities for each emotion category within five 3-second intervals (total of 15 seconds) prior to and after the feature use [46]. We focused on aggregated positive valence and negative valence, and deliberately excluded the neutral category, since it was not relevant to understanding the emotional impacts of search interactions. We aligned the emotional valence data with the time stamp of each AOI fixation to include data where participants looked at the interactive element. By aligning emotional responses with features that participants looked at and used, this method provides insight into how visual attention and subsequent feature use may generate emotional responses.

For the second approach, we synchronized emotional valence with time stamps on each AOI fixation to measure how emotional valence changes when searchers look at interface elements. The LogUI interaction data was used to filter out all instances that included interactive use (which are captured in the first approach described above), leaving those features that were viewed but not interacted with. Using the timestamp and the midpoint of the duration of each AOI fixation as a reference point, we measured emotional valence before and after each fixation. At five 3-second intervals (total of 15 seconds) before and after each fixation [57], we calculated the probabilities for aggregated positive valence and negative valence, excluding the neutral category. This approach captures passive interactions, when a user looks at a feature, possibly extracts information from it, but chooses to not explicitly interacting with it. For instance, a searcher might scan a search result, confirm some information they already know and have a positive emotional response to doing so, and then choose not to view the result in detail. With typical approaches to log analysis of feature use, such passive interaction is completely missed.

5 User Study Methodology

5.1 Study Design

In our initial evaluation of these exploratory search user interfaces for the Europeana collection, we focused on typical usability measures and performance metrics [29]. In this study, we extend that work by analyzing interactions (typical and passive) through the integration of eye tracking, feature use, and emotional valence data.

In a controlled laboratory study, we analyzed the interactions to consider how changing the independent variable (interface type) affects dependent variables related to feature use during the search process. We examined three interfaces: Baseline, Result-Focused, and SERP-Focused. A within-subjects design was used in the study, which allowed participants to experience all three types of search interfaces while performing three different search tasks. In order to reduce order effects, search interfaces and tasks were assigned using a 3x9 Graeco-Latin square. The study was conducted at three different universities (two in the United Kingdom and one in Canada). After receiving approval from the Research Ethics Board at the first authors' institution, the study was submitted to the corresponding ethics committees at the other two institutions and subsequently approved there as well.

5.2 Simulated Search Tasks

We designed three similar exploratory search tasks in consultation with a historian familiar with the Europeana collection, presenting these as simulated scenarios with contextual details and task instructions as per Borland's simulated work task structure [10]. These tasks follow the method described by Kules and Capra [40], designed to elicit uncertainty and curiosity while being engaging yet novel. The three search tasks were described similarly in order to ensure consistency in context and motivation. We confirmed in advance that the tasks had a similar level of coverage within Europeana's digital archive. In each task, the exploratory search process was described as a combination of an exploratory browsing style of searching and the use of a workspace for reviewing and assessing the resources that have been found. Table 2 provides the specific instructions used and the details of each of the three tasks.

Table 2. The general format of the simulated search task, and the details for the three specific tasks.

Task Template (SERP) Suppose you are starting a research project on {task topic}. Your goal is to explore among the Europeana collection for a set of resources that can help you to represent the breadth of this topic. You do not need to perform a deep or comprehensive search right now; instead, your goal is to save a diverse set of resources that will serve as the basis for a more focused search that you will perform later.
Task Template (Workspace) Our goal now is to examine and evaluate the resources you have saved, removing those that are no longer relevant or useful for your task.
Task 1 {search topic} = the range of Roman artefacts found in England.
Task 2 {search topic} = how depictions of the New World in maps have changed over time.
Task 3 {search topic} = how museum collections depict female athletes competing in the Olympics games.

5.3 Participants

Participants were recruited from among digital humanities historians across the three universities. We used a snowball sampling approach, sending emails to faculty members known for using digital humanities methods. In the recruitment email, faculty were invited to participate in the study and asked to forward it to their research staff and graduate students.

The 18 study participants included one senior undergraduate, five doctoral candidates, three postdoctoral researchers, three Lecturers/Assistant Professors, four Senior Lecturers/Associate Professors, and two Readers/Professors. Six participants identified as females and twelve as males. None reported that they had colour vision deficiencies. The participants reported that they used digital humanities collections at least once a month. Two participants used Europeana frequently; the others used it irregularly or not at all. One participant's self-reported search skills were "extremely good", nine were "somewhat good", and eight were "neither good nor bad". Even though most of the study participants were male and had limited prior experience with Europeana, the sample was nonetheless representative of a broad spectrum of digital humanities historians.

5.4 Procedures

The study began with collecting informed consent, followed by asking the participants to complete an initial questionnaire to collect demographic data, academic standing, and experience searching among digital humanities collections. The next step was to have the participants perform three search tasks with the three different search interfaces. For each, we provided a training video and allowed the participants to perform a training task to familiarize themselves with the assigned search interface. Pre-task questionnaires were administered prior to undertaking the search tasks. Then, participants searched for relevant resources related to the search task topic, and assessed what they found in the workspace. Following the completion of the search tasks, post-task questionnaires were administered and a five-minute break was provided to reduce cognitive fatigue between the search tasks. Throughout the search sessions, we collected eye tracking data, feature use log data, and emotional response data from the front-facing camera on the computer. Each study session took approximately two hours. Those in the UK received £40 in compensation, while those in Canada received the equivalent of $70.

6 Data Analysis

To analyze the emotional responses, we categorized them into three broad valence categories: positive, negative, and neutral. For each participant, the total time spent in each emotional valence was converted into frequency counts, reflecting the relative time spent in each category for each interface. These frequency counts were then analyzed using Pearson's chi-square test to compare the distribution of emotional responses before and after search interaction. Statistical significance was determined at a p-value threshold of 0.05.

7 Results

7.1 RQ1: Typical Interaction Feature Use

We examined the change in emotional responses before and after each interaction with the search interface features. To determine if there were any correlations between emotional responses and using the features, we considered the change in positive valence and the change in negative valence. If the positive valence went up and the negative valence went down at statistically significant levels, we considered this a *positive reaction*; if the opposite happened, we considered this as a *negative reaction*. The results of these analyses are provided in the following three tables for each of the three interfaces: Tables 3, 4, and 5. Within these tables, the features that resulted in positive reactions are marked as ⊕ and the features that resulted in negative reactions are marked as ⊖.

7.1.1 Input Features. Across all three interfaces, the variation in emotional valence around the input features did not significantly change before and after using these features. This is not unexpected given that the mechanisms for entering and submitting a query in all three interfaces followed a design pattern that is commonly used and provides a predictable outcome in the interface.

7.1.2 Control Features. While the use of the pagination feature did not elicit significant changes in emotional valence, the use of the search filter did. Across all three interfaces, participants had positive reactions to using the search filter. This suggests that the outcome of using the search filters was appreciated, validating the importance of providing searchers with mechanisms for interactively refining their queries based on attributes of the collection.

7.1.3 Informational Features. The two informational features provided across all three interfaces were the ability to click on the search results when viewing the SERP and when viewing the workspace. When using the Baseline interface, clicking on search results resulted in negative reactions. However, when using the two visual keyword/result linking approaches (the Result-Focused and SERP-Focused interfaces), clicking on search results resulted in positive reactions. This pattern was present both when accessing the search results from the SERP and when accessing them from the workspace. Clearly, participants using the Baseline were having difficulty finding search results they found valuable or useful. This was not the case with the other two interfaces; allowing the searchers to highlight related search results resulted in positive reactions when the results were clicked to view the details. We attribute this to the interactive features that allow searchers to reveal relationships between search results through keyword selection.

7.1.4 Personalization Features. Among the personalization features that were present in all the interfaces, many did not elicit significant changes in emotional valence. Many of these were to control aspects of the interface that were supplemental to the search tasks assigned.

Among those features that did elicit emotional reactions, there was one feature that generated a positive reaction across all interfaces: the ability to save a result to the workspace. This finding confirms the value of providing workspaces within exploratory search interfaces.Searchers are likely to experience positive emotions when they identify a relevant search result that they want to save for later use.

Although the ability to create and add search results to a gallery exists in all three interfaces, participants only had positive reactions to this feature when using the visual keyword/result linking interfaces. Since the gallery provides a mechanism for creating named sub-workspaces, perhaps the searchers found this more useful when combined with the interactive features provided by the Result-Focused and SERP-Focused interfaces, compared to serving as just another place to save resources in the Baseline.

There were a set of workspace features for which the participants had a negative reaction when used, which was present across all interfaces: the features to unlike results in the workspace, remove results from the workspace, and delete galleries. The common aspect of these features is that they undo previous work. The reaction to undoing such work was negative,

Table 3. Typical interaction with the Baseline; statistically significant changes in the positive and negative valence are highlighted in bold.

Interaction	Search action	Mean (SD) positive valence 15–0s before the event	Mean (SD) positive valence 0–15s after the event	Mean Δ	Mean (SD) negative valence 15–0s before the event	Mean (SD) negative valence 0–15s after the event	Mean Δ	Chi-square test result on positive valence	Chi-square test result on negative valence
Input	Submit a query using the search button/clicking the enter key	0.25 (0.05)	0.27 (0.04)	+0.02	0.30 (0.04)	0.27 (0.03)	-0.03	χ^2(1, N=114)=0.08, p=0.78	χ^2(1, N=114)=0.04, p=0.83
Control	Click on pagination icon	0.17 (0.03)	0.18 (0.03)	+0.01	0.30 (0.03)	0.31 (0.04)	+0.01	χ^2(1, N=45)=0.01, p=0.96	χ^2(1, N=45)=0.83, p=0.36
	⊕ Select value from search filter	0.19 (0.04)	0.30 (0.04)	+0.11	0.24 (0.03)	0.15 (0.02)	-0.09	χ^2**(1, N=87)=12.14, p<0.001**	χ^2**(1, N=87)=10.63, p<0.01**
Informational	⊖ Click on search result	0.19 (0.03)	0.12 (0.04)	-0.07	0.24 (0.05)	0.30 (0.05)	+0.06	χ^2**(1, N=223)=12.97, p<0.001**	χ^2**(1, N=223)=4.34, p<0.05**
	⊖ Click on search result in workspace	0.16 (0.03)	0.07 (0.03)	-0.09	0.34 (0.03)	0.44 (0.04)	+0.10	χ^2**(1, N=231)=8.81, p<0.01**	χ^2**(1, N=231)=11.58, p<0.001**
Personalization	Click like result	0.15 (0.03)	0.10 (0.05)	-0.05	0.29 (0.02)	0.26 (0.03)	-0.03	χ^2(1, N=88)=0.46, p=0.50	χ^2(1, N=88)=0.05, p=0.82
	Click unlike result	0.23 (0.03)	0.17 (0.04)	-0.06	0.17 (0.04)	0.19 (0.03)	+0.02	χ^2(1, N=40)=0.03, p=0.87	χ^2(1, N=40)=0.01, p=0.96
	⊕ Click save result	0.28 (0.03)	0.33 (0.04)	+0.05	0.35 (0.04)	0.24 (0.05)	-0.11	χ^2**(1, N=157)=6.04, p<0.05**	χ^2**(1, N=157)=5.61, p<0.05**
	Click "My Profile"	0.14 (0.03)	0.09 (0.04)	-0.05	0.35 (0.05)	0.37 (0.03)	+0.02	χ^2(1, N=93)=0.27, p=0.61	χ^2(1, N=93)=0.04, p=0.85
	Click change layout	0.21 (0.05)	0.17 (0.04)	-0.04	0.32 (0.03)	0.33 (0.04)	+0.01	χ^2(1, N=41)=0.12, p=0.73	χ^2(1, N=41)=2.08, p=0.15
	Click create gallery	0.17 (0.03)	0.09 (0.04)	-0.08	0.31 (0.02)	0.27 (0.04)	-0.04	χ^2(1, N=95)=0.88, p=0.35	χ^2(1, N=95)=0.24, p=0.62
	Click add result to gallery	0.16 (0.04)	0.12 (0.04)	-0.04	0.18 (0.05)	0.29 (0.04)	+0.11	χ^2(1, N=138)=0.21, p=0.65	χ^2(1, N=138)=0.31, p=0.58
	Click like results in workspace	0.12 (0.04)	0.18 (0.04)	+0.06	0.23 (0.05)	0.21 (0.04)	-0.02	χ^2(1, N=47)=0.20, p=0.66	χ^2(1, N=47)=1.06, p=0.30
	⊖ Click unlike results in workspace	0.18 (0.03)	0.07 (0.03)	-0.11	0.23 (0.03)	0.33 (0.02)	+0.10	χ^2**(1, N=33)=8.10, p<0.01**	χ^2**(1, N=33)=13.44, p<0.001**
	⊖ Click on remove result in workspace	0.21 (0.02)	0.13 (0.03)	-0.08	0.22 (0.04)	0.31 (0.02)	+0.09	χ^2**(1, N=48)=10.70, p<0.01**	χ^2**(1, N=48)=7.56, p<0.01**
	⊖ Click delete gallery in workspace	0.23 (0.03)	0.12 (0.02)	-0.11	0.22 (0.06)	0.37 (0.05)	+0.15	χ^2**(1, N=43)=9.91, p<0.01**	χ^2**(1, N=43)=9.72, p<0.01**

Table 4. Typical interaction with Result-Focused; statistically significant changes in the positive and negative valence are highlighted in bold.

Interaction	Search action	Mean (SD) positive valence 15–0s before the event	Mean (SD) positive valence 0–15s after the event	Mean Δ	Mean (SD) negative valence 15–0s before the event	Mean (SD) negative valence 0–15s after the event	Mean Δ	Chi-square test result on positive valence	Chi-square test result on negative valence
Input	Submit a query using the search button/clicking the enter key	0.30 (0.04)	0.35 (0.05)	+0.05	0.16 (0.03)	0.12 (0.04)	-0.04	χ^2(1, N=138)=0.01, p=0.96	χ^2(1, N=138)=1.72, p=0.19
Control	Click on pagination icon	0.15 (0.03)	0.111 (0.02)	-0.04	0.18 (0.05)	0.21 (0.03)	+0.03	χ^2(1, N=77)=0.13, p=0.72	χ^2(1, N=77)=0.37, p=0.54
	⊕ Select value from search filter	0.18 (0.03)	0.28 (0.05)	+0.10	0.37 (0.03)	0.31 (0.02)	-0.06	χ^2**(1, N=115)=14.48, p<0.001**	χ^2**(1, N=115)=7.82, p<0.01**
Informational	⊕ Click on search result	0.31 (0.05)	0.40 (0.04)	+0.09	0.26 (0.04)	0.22 (0.03)	-0.04	χ^2**(1, N=201)=5.71, p<0.05**	χ^2**(1, N=201)=5.53, p<0.05**
	⊕ Click on search result in workspace	0.35 (0.05)	0.41 (0.05)	+0.06	0.15 (0.02)	0.10 (0.04)	-0.05	χ^2**(1, N=151)=8.72, p<0.01**	χ^2**(1, N=151)=6.14, p<0.05**
Personalization	Click like result	0.23 (0.02)	0.28 (0.03)	+0.05	0.18 (0.04)	0.14 (0.03)	-0.04	χ^2(1, N=103)=0.39, p=0.53	χ^2(1, N=103)=0.52, p=0.47
	Click unlike result	0.24 (0.03)	0.17 (0.03)	-0.07	0.24 (0.04)	0.28 (0.03)	+0.04	χ^2(1, N=53)=0.01, p=0.96	χ^2(1, N=53)=0.03, p=0.87
	⊕ Click save result	0.44 (0.03)	0.54 (0.07)	+0.10	0.20 (0.04)	0.12 (0.03)	-0.08	χ^2**(1, N=199)=9.00, p<0.01**	χ^2**(1, N=199)=5.72, p<0.05**
	Click "My Profile"	0.25 (0.05)	0.30 (0.03)	+0.05	0.20 (0.02)	0.12 (0.04)	-0.08	χ^2(1, N=101)=0.01, p=0.96	χ^2(1, N=101)=1.49, p=0.22
	Click change layout	0.26 (0.04)	0.19 (0.03)	-0.07	0.31 (0.05)	0.43 (0.03)	+0.12	χ^2(1, N=24)=0.51, p=0.47	χ^2(1, N=24)=0.37, p=0.54
	Click create gallery	0.25 (0.03)	0.31 (0.05)	+0.06	0.12 (0.02)	0.11 (0.03)	-0.01	χ^2(1, N=118)=0.19, p=0.66	χ^2(1, N=118)=0.15, p=0.70
	⊕ Click add result to gallery	0.42 (0.02)	0.48 (0.03)	+0.06	0.21 (0.03)	0.14 (0.04)	-0.07	χ^2**(1, N=188)=9.81, p<0.01**	χ^2**(1, N=188)=8.10, p<0.01**
	Click like results in workspace	0.28 (0.03)	0.31 (0.04)	+0.03	0.18 (0.05)	0.14 (0.03)	-0.04	χ^2(1, N=26)=2.12, p=0.15	χ^2(1, N=26)=2.21, p=0.14
	⊖ Click unlike results in workspace	0.41 (0.04)	0.30 (0.03)	-0.11	0.17 (0.03)	0.31 (0.03)	+0.14	χ^2**(1, N=56)=13.19, p<0.001**	χ^2**(1, N=56)=12.69, p<0.001**
	⊖ Click on remove result in workspace	0.41 (0.04)	0.30 (0.02)	-0.11	0.21 (0.04)	0.30 (0.05)	+0.09	χ^2**(1, N=43)=8.12, p<0.01**	χ^2**(1, N=43)=6.80, p<0.01**
	⊖ Click delete gallery in workspace	0.34 (0.03)	0.22 (0.05)	-0.12	0.20 (0.03)	0.28 (0.03)	+0.08	χ^2**(1, N=35)=13.88, p<0.001**	χ^2**(1, N=35)=9.96, p<0.01**
	⊕ Click highlight keyword	0.47 (0.02)	0.57 (0.03)	+0.10	0.31 (0.04)	0.17 (0.03)	-0.14	χ^2**(1, N=261)=8.91, p<0.01**	χ^2**(1, N=261)=11.29, p<0.001**
	⊖ Click unhighlight keyword	0.30 (0.04)	0.22 (0.03)	-0.08	0.20 (0.04)	0.39 (0.03)	+0.19	χ^2**(1, N=80)=14.41, p<0.001**	χ^2**(1, N=80)=15.28, p<0.0001**
	⊖ Click reset button	0.27 (0.04)	0.16 (0.02)	-0.11	0.18 (0.02)	0.33 (0.03)	+0.15	χ^2**(1, N=57)=13.32, p<0.001**	χ^2**(1, N=57)=15.03, p<0.001**
	⊕ Click highlight keyword in workspace	0.42 (0.05)	0.53 (0.02)	+0.11	0.24 (0.04)	0.18 (0.04)	-0.06	χ^2**(1, N=287)=12.35, p<0.001**	χ^2**(1, N=287)=5.29, p<0.05**
	⊖ Click unhighlight keyword in workspace	0.47 (0.02)	0.33 (0.04)	-0.14	0.15 (0.04)	0.25 (0.03)	+0.10	χ^2**(1, N=96)=14.59, p<0.001**	χ^2**(1, N=96)=7.24, p<0.01**
	⊖ Click reset in workspace	0.29 (0.04)	0.23 (0.03)	-0.06	0.21 (0.05)	0.30 (0.05)	+0.09	χ^2**(1, N=47)=5.90, p<0.05**	χ^2**(1, N=47)=10.64, p<0.01**

Table 5. Typical interaction with SERP-Focused; statistically significant changes in the positive and negative valence are highlighted in bold.

Interaction	Search action	Mean (SD) positive valence 15–0s before the event	Mean (SD) positive valence 0–15s after the event	Mean Δ	Mean (SD) negative valence 15–0s before the event	Mean (SD) negative valence 0–15s after the event	Mean Δ	Chi-square test result on positive valence	Chi-square test result on negative valence
Input	Submit a query using the search button/clicking the enter key	0.40 (0.04)	0.43 (0.04)	+0.03	0.28 (0.03)	0.20 (0.03)	-0.08	χ^2(1, N=134)=0.41, p=0.52	χ^2(1, N=134)=0.70, p=0.40
Control	Click on pagination icon	0.22 (0.03)	0.25 (0.03)	+0.03	0.17 (0.05)	0.20 (0.01)	+0.03	χ^2(1, N=86)=0.01, p=0.92	χ^2(1, N=86)=0.39, p=0.53
	⊕ Select value from search filter	0.22 (0.04)	0.38 (0.02)	+0.16	0.31 (0.03)	0.24 (0.05)	-0.07	χ^2**(1, N=118)=15.63, p<0.0001**	χ^2**(1, N=118)=9.33, p<0.01**
Informational	⊕ Click on search result	0.27 (0.04)	0.45 (0.05)	+0.18	0.22 (0.03)	0.17 (0.05)	-0.05	χ^2**(1, N=254)=7.03, p<0.01**	χ^2**(1, N=254)=5.50, p<0.05**
	⊕ Click on search result in workspace	0.40 (0.04)	0.45 (0.01)	+0.05	0.23 (0.05)	0.13 (0.02)	-0.10	χ^2**(1, N=150)=5.36, p<0.05**	χ^2**(1, N=150)=7.23, p<0.01**
Personalization	Click like result	0.30 (0.03)	0.33 (0.04)	+0.03	0.18 (0.03)	0.16 (0.04)	-0.02	χ^2(1, N=123)=0.20, p=0.66	χ^2(1, N=123)=0.01, p=0.98
	Click unlike result	0.28 (0.04)	0.24 (0.04)	-0.04	0.33 (0.03)	0.39 (0.03)	+0.05	χ^2(1, N=58)=0.01, p=0.92	χ^2**(1, N=58)=3.99, p<0.05**
	⊕ Click save result	0.38 (0.04)	0.56 (0.04)	+0.18	0.27 (0.05)	0.13 (0.02)	-0.14	χ^2**(1, N=196)=12.89, p<0.001**	χ^2**(1, N=196)=14.83, p<0.001**
	Click "My Profile"	0.25 (0.05)	0.27 (0.03)	+0.02	0.15 (0.03)	0.12 (0.04)	-0.03	χ^2(1, N=88)=0.09, p=0.77	χ^2(1, N=88)=0.84, p=0.36
	Click change layout	0.23 (0.04)	0.20 (0.02)	-0.03	0.30 (0.05)	0.33 (0.06)	+0.03	χ^2(1, N=19)=0.07, p=0.78	χ^2(1, N=19)=0.01, p=0.96
	Click create gallery	0.33 (0.02)	0.40 (0.04)	+0.07	0.21 (0.04)	0.16 (0.07)	-0.05	χ^2(1, N=129)=0.04, p=0.83	χ^2(1, N=129)=0.01, p=0.93
	⊕ Click add result to gallery	0.37 (0.05)	0.45 (0.02)	+0.08	0.23 (0.04)	0.11 (0.02)	-0.12	χ^2**(1, N=188)=6.28, p<0.05**	χ^2**(1, N=188)=11.03, p<0.001**
	Click like results in workspace	0.30 (0.04)	0.35 (0.04)	+0.05	0.17 (0.04)	0.12 (0.03)	-0.05	χ^2(1, N=48)=1.69, p=0.19	χ^2**(1, N=48)=0.02, p<0.89**
	⊖ Click unlike results in workspace	0.28 (0.03)	0.17 (0.04)	-0.11	0.22 (0.05)	0.30 (0.02)	+0.08	χ^2**(1, N=25)=12.39, p<0.001**	χ^2**(1, N=25)=9.61, p<0.01**
	⊖ Click on remove result in workspace	0.30 (0.02)	0.20 (0.03)	-0.10	0.22 (0.05)	0.34 (0.04)	+0.12	χ^2**(1, N=47)=6.74, p<0.01**	χ^2**(1, N=47)=9.19, p<0.01**
	⊖ Click delete gallery in workspace	0.25 (0.03)	0.17 (0.04)	-0.08	0.22 (0.03)	0.30 (0.05)	+0.08	χ^2**(1, N=36)=6.39, p<0.01**	χ^2**(1, N=36)=7.19, p<0.01**
	⊕ Click highlight keyword	0.35 (0.05)	0.47 (0.03)	+0.12	0.20 (0.05)	0.11 (0.04)	-0.09	χ^2**(1, N=251)=10.70, p<0.01**	χ^2**(1, N=251)=10.00, p<0.01**
	⊖Click unhighlight keyword	0.30 (0.04)	0.20 (0.05)	-0.10	0.18 (0.04)	0.25 (0.02)	+0.07	χ^2**(1, N=72)=12.56, p<0.001**	χ^2**(1, N=72)=9.85, p<0.01**
	⊖ Click reset button	0.27 (0.04)	0.19 (0.05)	-0.08	0.21 (0.04)	0.35 (0.03)	+0.14	χ^2**(1, N=42)=12.14, p<0.001**	χ^2**(1, N=42)=16.06, p<0.0001**
	⊕ Click highlight keyword in workspace	0.35 (0.04)	0.43 (0.06)	+0.08	0.17 (0.04)	0.13 (0.05)	-0.04	χ^2**(1, N=263)=6.55, p<0.05**	χ^2**(1, N=263)=4.49, p<0.05**
	⊖ Click unhighlight keyword in workspace	0.46 (0.05)	0.34 (0.04)	-0.12	0.23 (0.03)	0.31 (0.05)	+0.08	χ^2**(1, N=76)=9.46, p<0.01**	χ^2**(1, N=76)=14.15, p<0.001**
	⊖ Click reset in workspace	0.35 (0.04)	0.30 (0.03)	-0.05	0.24 (0.04)	0.34 (0.05)	+0.10	χ^2**(1, N=34)=5.61, p<0.05**	χ^2**(1, N=34)=7.46, p<0.01**

Table 6. Summary of typical interactions (Mean and Standard Deviation) across the Baseline (B), Result-focused (RF), and SERP-focused (SF) interfaces (* indicates statistically significant differences versus the Baseline).

Feature Type	Search Action	Interactions - B	Interactions - RF	Interactions - SF	ANOVA
Input	Submit a query using the search button	6.34 (3.14)	7.67 (3.27)	7.45 (3.25)	F(2,51)=1.18, p=0.31
Control	Click on pagination icon	4.50 (1.75)	4.28 (1.93)	4.78 (1.98)	F(2,51)=1.99, p=0.15
	Select value from search filter	4.84 (2.99)	6.39 (3.14)	6.56 (2.16)	F(2,51)=0.86, p=0.43
Informational	Click on search result	12.39 (3.74)	11.17 (3.62)	14.12 (3.02)	F(2,51)=0.05, p=0.95
	Click on search result in workspace	12.84 (2.79)	8.39* (3.34)	8.34* (1.94)	**F(2,51)=4.82, p<0.05**
Personalization	Click like result	4.89 (2.02)	5.73* (3.18)	6.84* (1.19)	**F(2,51)=4.67, p<0.05**
	Click unlike result	2.23 (1.73)	2.95 (2.08)	3.23 (2.83)	F(2,51)=0.74, p=0.48
	Click save result	8.72 (3.54)	11.06* (3.01)	10.89* (2.59)	**F(2,51)=6.14, p<0.01**
	Click "My Profile"	5.17 (3.54)	5.62 (2.07)	4.89 (3.99)	F(2,51)=1.27, p=0.29
	Click change layout	2.28 (1.73)	1.34 (0.64)	1.06 (1.02)	F(2,51)=0.69, p=0.51
	Click create gallery	5.28 (3.03)	6.56 (2.16)	7.17 (3.22)	F(2,51)=0.48, p=0.62
	Click add result to gallery	7.66 (4.38)	10.41* (3.16)	10.45* (2.55)	**F(2,51)=5.81, p<0.01**
	Click like results in workspace	2.62 (1.77)	1.45 (1.05)	2.67 (1.17)	F(2,51)=0.98, p=0.38
	Click unlike results in workspace	1.84 (0.69)	3.12 (2.82)	1.39 (0.64)	F(2,51)=0.87, p=0.42
	Click remove result in workspace	2.67 (1.77)	2.39 (2.00)	2.62 (2.17)	F(2,51)=0.23, p=0.80
	Click delete gallery in workspace	2.39 (1.74)	1.95 (0.53)	2.00 (1.70)	F(2,51)=2.16, p=0.13
	Click highlight keyword	-	14.5 (3.95)	13.95 (2.56)	F(1,34)=2.55, p=0.12
	Click unhighlight keyword	-	4.45 (2.95)	4.00 (1.78)	F(1,34)=0.67, p=0.42
	Click reset button	-	3.17 (1.82)	2.34 (2.04)	F(1,34)=0.53, p=0.47
	Click highlight keyword in workspace	-	15.95 (3.12)	14.62 (3.97)	F(1,34)=0.30, p=0.59
	Click unhighlight keyword in workspace	-	5.34 (3.04)	4.23 (1.93)	F(1,34)=1.40, p=0.25
	Click reset in workspace	-	2.62 (1.77)	1.89 (0.69)	F(1,34)=2.43, p=0.13

which highlights the frustration searchers may experience when finding something they think might be relevant to only later decide that it is not.

A set of interactive features related to keywords were present in the two interfaces that provided visual keyword/result linking. These features represented the key difference with the Baseline, and were represented in two different formats in the Result-Focused and SERP-Focused interfaces. These different formats did not have an impact on the participants' reactions to using the features. When keywords were clicked to highlight search results, participants had a positive reaction. This was present both when using this feature in SERP and in the Workspace, highlighting the value of allowing the searchers to interactively surface relationships between search results using interactive keyword selection and visualization. When the selection of a particular keyword was no longer of value in evaluating the search results or workspace, choosing to unhighlight it or unhighlight all via the reset button resulted in a negative reaction. Here again, when searchers found themselves needing to undo previous work, they reacted negatively.

7.1.5 Frequency of Interaction. The frequency of interaction with each of the features across all three interfaces is provided in Table 6. Repeated measures ANOVA was conducted to compare this data across the interface conditions. In cases where statistical significance was found, a post-hoc Tukey HSD test was performed to identify across which pairs of interfaces statistical significance was present. Overall, the participants in this study engaged with the search process and used the core features of the interfaces at a similar level, with three exceptions.

When using the Baseline interface, participants clicked on the search results in the workspace more frequently than when they used the two visual keyword/result linking interfaces. This combined with the earlier observation of negative reactions to clicking on the results in the workspace when using the Baseline provides further evidence of the difficulties the searchers had with finding relevant information for the assigned task. That is, the participants needed to access the full resource records to verify that what they had was relevant, as opposed to using the keywords surfaced in the other two interfaces as a mechanism to support this activity.

When using the Baseline, participants also used the like feature less frequently, and added fewer results to the gallery. While there was no difference in the emotional reactions to using the like feature between the interfaces, those who used the two visual keyword/result linking interfaces had positive reactions to adding results to the gallery. Perhaps by providing extra interactive features within these new interfaces, participants were also willing to try out other features that they were not familiar with. They may not have noticed that the gallery was present in the Baseline interface, as the familiar design may not have motivated an investigation or discovery of its features.

7.1.6 Summary. As a result of integrating eye tracking, feature use, and emotional valence, we were able to identify patterns in the interaction with the features of the search interfaces. Participants had positive reactions across all three interfaces with respect to using the filters and being able to save results to the workspace. Participants had negative reactions with respect to using features that had a result of undoing previous work. Being able to interact with keywords to surface relationships among the search results generated positive reactions, with little difference between the Result-Focused and SERP-Focused approaches. This fundamental difference versus the Baseline had the effect of participants having negative reactions when clicking on search results to show more details with the Baseline, but positive reactions for the same interaction when using the visual keyword/result linking interfaces. This depth of analysis provides a richer view of the participants' experiences when using the interfaces than our previous analysis was able to do following typical IIR evaluation methods [57].

7.2 RQ2: Passive Interaction Feature Use

We examined the changes in emotional valence before and after passive interaction with search interface features (i.e., when searchers viewed a feature but chose not to use it). As with the previous analysis, positive reactions were identified by a statistically significant increase in positive valence and a statistically significant decrease in negative valence; negative reactions were identified in the opposite situation. The results of these analyses are provided in the following three tables for each of the three interfaces: Tables 7, 8, and 9.

7.2.1 Input Features. The passive interaction with the single input feature - viewing the query input box but choosing not to change the query - did not generate a statistically significant change emotional valence for any of the interfaces. The layout and operation of the query feature was the same across all interfaces, and the interactive features provided in the visual keyword/result linking interfaces did not provide explicit mechanisms for query refinement. As such, the action of viewing the query without interacting with it would typically occur when using the query box as a memory aid - when the searcher wishes to remind themselves of the query they issued.

7.2.2 Control Features. Across all three interfaces, there was a common pattern of viewing the search filters but not finding something worth selecting, resulting in a negative reaction. In the evaluation of the typical interaction with this feature in the previous section, we found that those who used the search filters had positive reactions; yet here we found the opposite reaction. While we do not have data to explain the circumstances under which the search filters were used versus viewed but not used, what we can infer is that in situations in which the searcher wanted to use the filter but could not make a selection, they had a negative reaction. Perhaps this was due to a lack of understanding of the options provided in the filter, highlighting the importance of ensuring that such filters present information in a format that is understandable by searchers.

7.2.3 Informational Features. When passively interacting with individual search results (looking at them but not interacting with them), there was a negative reaction. This is what one would expect, as viewing a search result and choosing not to click on it or save it suggests that it was not considered relevant by the participants. This negative reaction was consistent across all three interfaces.

When passively interacting with the search results saved in the workspace, the results were different between the interfaces. When participants used the Baseline, there was a negative reaction, which suggests that the participants were not satisfied with the search results they had previously saved. The opposite effect was present in the visual keyword/result linking interfaces, where viewing a search result in the workspace generated a positive reaction among the participants. This suggests that the participants had different impressions of the quality of what they had saved in the workspace between the Baseline and the other two interfaces, which we can attribute to the value of the interactive keyword/result highlighting features provided in the other two interfaces.

7.2.4 Personalization Features. When considering the personalization features and choosing not to use them, some features did not elicit much change in emotional valence. Among those where there was a statistically significant change in the valence, a few elicited mixed emotional reactions between the interfaces, and others elicited consistent emotional reactions across the interfaces.

Participants who passively interacted with the like result feature using the SERP-focused interface had a positive reaction. Participants who passively interacted with the save result feature using the Baseline interface had a negative reaction. As both of these features were present across all three interfaces and were not influenced by the differences between the interfaces, we do not have an explanation for why participants had these emotional reactions to viewing these features. Perhaps there were subtle differences in the layouts of these features in relation to other interactive elements in the interfaces that drew the searchers' attention to them, but did not motivate the searcher to use the features.

When viewing the delete gallery feature and choosing not to use it, participants had a negative reaction across all three interfaces. We attribute this to the uncertainty of the outcome of using this feature, and how it would affect the work the participants had done thus far. For example, because the gallery exists within the workspace, it may have been unclear to participants whether deleting the gallery would delete all saved resources in the gallery as well, or whether these would then be moved to the general workspace. This result highlights the need for ensuring that any features that delete or undo prior work have clear and obvious outcomes.

Passive interaction with the keywords present in the Result-Focused and SERP-Focused interfaces resulted in positive reactions, both when they were presented in the SERP and when they were presented in the workspace. Whether the keywords remained connected to the search results (in the Result-Focused approach) or were aggregated and provided along side the search results (in the SERP-focused approach) had little difference. This highlights the overall value of surfacing keywords in support of exploratory search - even when these keywords are not used explicitly they provide searchers with value, the result of which is measurable with our approach.

When participants viewed the reset button (associated with the keyword selection) and chose not to use it, they had a positive reaction. This suggests that they were considering removing all previously selected keywords and starting over, but in choosing not to, they reaffirmed their positive reaction to the keyword choices they had made.

7.2.5 Frequency of Fixation and Length of Dwell Time. We calculated the frequency of fixations and dwell times per participant on each AOI for each of the three interfaces, along with statistical analysis (repeated measures ANOVA and post-hoc Tukey HSD) to identify which features are being passively interacted with more frequently in which interfaces. These results are provided in Table 10.

Table 7. Passive interaction with the Baseline; statistically significant changes in the positive and negative valence are highlighted in bold.

Feature Type	AOI	Mean (SD) positive valence 15–0s before the event	Mean (SD) positive valence 0–15s after the event	Mean Δ	Mean (SD) negative valence 15–0s before the event	Mean (SD) negative valence 0–15s after the event	Mean Δ	Chi-square test result on positive valence	Chi-square test result on negative valence
Input	Query input box	0.30 (0.03)	0.40 (0.03)	+0.10	0.17 (0.04)	0.13 (0.03)	-0.04	χ^2(1, N=89)=2.94, p=0.09	χ^2(1, N=89)=1.55, p=0.21
Control	Search results navigation	0.37 (0.02)	0.37 (0.03)	0.00	0.13 (0.04)	0.20 (0.04)	+0.07	χ^2(1, N=48)=0.13, p=0.72	χ^2(1, N=48)=2.62, p=0.11
	⊖ Search filters	0.29 (0.03)	0.17 (0.03)	-0.12	0.24 (0.03)	0.38 (0.03)	+0.14	χ^2**(1, N=78)=8.79, p<0.01**	χ^2**(1, N=78)=5.04, p<0.05**
Informational	⊖ Search result	0.27 (0.04)	0.19 (0.02)	-0.08	0.18 (0.04)	0.24 (0.06)	+0.06	χ^2**(1, N=74)=9.16, p<0.01**	χ^2**(1, N=74)=4.35, p<0.05**
	⊖ Search result in workspace	0.29 (0.04)	0.22 (0.03)	-0.07	0.22 (0.02)	0.32 (0.04)	+0.10	χ^2**(1, N=76)=4.58, p<0.05**	χ^2**(1, N=76)=7.60, p<0.01**
Personalization	Like result	0.17 (0.04)	0.13 (0.04)	-0.04	0.47 (0.04)	0.47 (0.04)	0.00	χ^2(1, N=36)=0.86, p=0.35	χ^2(1, N=36)=2.61, p=0.11
	⊖ Save result	0.17 (0.01)	0.09 (0.03)	-0.08	0.26 (0.04)	0.39 (0.03)	+0.13	χ^2**(1, N=98)=9.12, p<0.01**	χ^2**(1, N=98)=8.44, p<0.01**
	Go to workspace	0.17 (0.03)	0.17 (0.03)	0.00	0.29 (0.03)	0.31 (0.04)	+0.02	χ^2(1, N=39)=0.05, p=0.83	χ^2(1, N=39)=1.15, p=0.28
	Change layout	0.28 (0.03)	0.27(0.04)	-0.01	0.29 (0.03)	0.37 (0.04)	+0.08	χ^2(1, N=28)=0.12, p=0.73	χ^2(1, N=28)=1.14, p=0.29
	Create gallery	0.20 (0.03)	0.14 (0.04)	-0.06	0.30 (0.05)	0.36 (0.03)	+0.06	χ^2(1, N=20)=1.06, p=0.30	χ^2(1, N=20)=0.53, p=0.47
	Add result to gallery	0.25 (0.05)	0.26 (0.05)	+0.01	0.31 (0.03)	0.17 (0.04)	-0.14	χ^2(1, N=43)=0.17, p=0.68	χ^2(1, N=43)=1.40, p=0.24
	Like result in workspace	0.23 (0.04)	0.24 (0.05)	+0.01	0.16 (0.03)	0.17 (0.02)	+0.01	χ^2(1, N=46)=0.13, p=0.72	χ^2(1, N=46)=1.08, p=0.29
	⊖ Delete gallery	0.29 (0.02)	0.18 (0.02)	-0.11	0.31 (0.03)	0.38 (0.03)	+0.07	χ^2**(1, N=26)=4.79, p<0.05**	χ^2**(1, N=26)=12.99, p<0.001**

Table 8. Passive interaction with Result-Focused; statistically significant changes in the positive and negative valence are highlighted in bold.

Feature Type	AOI	Mean (SD) positive valence 15–0s before the event	Mean (SD) positive valence 0–15s after the event	Mean Δ	Mean (SD) negative valence 15–0s before the event	Mean (SD) negative valence 0–15s after the event	Mean Δ	Chi-square test result on positive valence	Chi-square test result on negative valence
Input	Query input box	0.34 (0.04)	0.39 (0.04)	+0.05	0.21 (0.03)	0.16 (0.03)	-0.05	χ^2(1, N=85)=1.55, p=0.21	χ^2(1, N=85)=0.16, p=0.69
Control	Search results navigation	0.13 (0.03)	0.13 (0.03)	0.00	0.17 (0.04)	0.16 (0.04)	-0.01	χ^2(1, N=22)=0.01, p=0.96	χ^2(1, N=22)=0.17, p=0.68
	⊖ Search filters	0.33 (0.05)	0.19 (0.03)	-0.14	0.23 (0.03)	0.36 (0.03)	+0.13	χ^2**(1, N=101)=8.42, p<0.01**	χ^2**(1, N=101)=6.15, p<0.05**
Informational	⊖ Search result	0.30 (0.04)	0.21 (0.05)	-0.09	0.17 (0.02)	0.27 (0.03)	+0.10	χ^2**(1, N=53)=5.94, p<0.05**	χ^2**(1, N=53)=8.06, p<0.01**
	⊕ Search result in workspace	0.24 (0.04)	0.31 (0.06)	+0.07	0.27 (0.02)	0.18 (0.04)	-0.09	χ^2**(1, N=79)=7.83, p<0.01**	χ^2**(1, N=79)=10.15, p<0.01**
Personalization	Like result	0.23 (0.04)	0.30 (0.04)	+0.07	0.20 (0.03)	0.17 (0.03)	-0.03	χ^2(1, N=63)=0.03, p=0.88	χ^2(1, N=63)=0.25, p=0.62
	Save result	0.34 (0.04)	0.44 (0.03)	+0.10	0.23 (0.03)	0.10 (0.02)	-0.13	χ^2(1, N=45)=3.69, p=0.05	χ^2(1, N=45)=0.63, p=0.43
	Go to workspace	0.17 (0.03)	0.17 (0.03)	0.00	0.16 (0.04)	0.13 (0.04)	-0.03	χ^2(1, N=38)=1.89, p=0.16	χ^2(1, N=38)=0.13, p=0.72
	Change layout	0.30 (0.04)	0.29(0.03)	-0.01	0.29 (0.02)	0.30 (0.03)	+0.01	χ^2(1, N=23)=0.51, p=0.47	χ^2(1, N=23)=0.29, p=0.59
	Create gallery	0.23 (0.04)	0.20 (0.05)	-0.03	0.12 (0.03)	0.26 (0.04)	+0.14	χ^2(1, N=33)=0.95, p=0.32	χ^2(1, N=33)=0.96, p=0.33
	Add result to gallery	0.50 (0.03)	0.54 (0.03)	+0.04	0.24 (0.04)	0.10 (0.03)	-0.14	χ^2(1, N=38)=2.35, p=0.13	χ^2(1, N=38)=2.62, p=0.11
	Like result in workspace	0.24 (0.04)	0.24 (0.05)	0.00	0.27 (0.03)	0.23 (0.02)	-0.04	χ^2(1, N=21)=1.26, p=0.26	χ^2(1, N=21)=1.84, p=0.17
	⊖ Delete gallery	0.35 (0.01)	0.26 (0.02)	-0.09	0.28 (0.03)	0.33 (0.03)	+0.05	χ^2**(1, N=17)=4.78, p<0.05**	χ^2**(1, N=17)=4.56, p<0.05**
	⊕ Keyword	0.35 (0.03)	0.57 (0.04)	+0.22	0.17 (0.03)	0.11 (0.02)	-0.06	χ^2**(1, N=34)=7.11, p<0.01**	χ^2**(1, N=34)=4.67, p<0.05**
	⊕ Reset button	0.24 (0.02)	0.32 (0.01)	+0.08	0.22 (0.03)	0.18 (0.02)	-0.04	χ^2**(1, N=15)=4.09, p<0.05**	χ^2**(1, N=15)=5.97, p<0.05**
	⊕ Keyword in workspace	0.45 (0.04)	0.63 (0.05)	+0.18	0.23 (0.03)	0.17 (0.03)	-0.06	χ^2**(1, N=42)=6.06, p<0.05**	χ^2**(1, N=42)=4.16, p<0.05**
	⊕ Reset button in workspace	0.34 (0.01)	0.43 (0.01)	+0.09	0.24 (0.01)	0.14 (0.02)	-0.10	χ^2**(1, N=18)=4.27, p<0.05**	χ^2**(1, N=18)=8.53, p<0.01**

Table 9. Passive interaction with SERP-Focused; statistically significant changes in the positive and negative valence are highlighted in bold.

Feature Type	AOI	Mean (SD) positive valence 15–0s before the event	Mean (SD) positive valence 0–15s after the event	Mean Δ	Mean (SD) negative valence 15–0s before the event	Mean (SD) negative valence 0–15s after the event	Mean Δ	Chi-square test result on positive valence	Chi-square test result on negative valence
Input	Query input box	0.20 (0.04)	0.26 (0.04)	+0.06	0.21 (0.03)	0.15 (0.03)	-0.06	χ^2(1, N=86)=0.01, p=0.95	χ^2(1, N=86)=2.81, p=0.09
Control	Search results navigation	0.26 (0.03)	0.20 (0.04)	-0.06	0.30 (0.04)	0.30 (0.04)	0.00	χ^2(1, N=27)=0.02, p=0.86	χ^2(1, N=27)=0.09, p=0.76
	⊖ Search filters	0.37 (0.03)	0.19 (0.03)	-0.18	0.25 (0.03)	0.38 (0.02)	+0.13	**χ^2(1, N=83)=5.17, p<0.05**	**χ^2(1, N=83)=8.18, p<0.01**
Informational	⊖ Search result	0.37 (0.02)	0.26 (0.04)	-0.11	0.19 (0.03)	0.26 (0.03)	+0.07	**χ^2(1, N=46)=7.46, p<0.01**	**χ^2(1, N=46)=4.71, p<0.05**
	⊕ Search result in workspace	0.27 (0.04)	0.35 (0.03)	+0.08	0.25 (0.05)	0.14 (0.06)	-0.11	**χ^2(1, N=87)=7.84, p<0.01**	**χ^2(1, N=87)=10.15, p<0.01**
Personalization	⊕ Like result	0.17 (0.03)	0.25 (0.03)	+0.08	0.17 (0.03)	0.14 (0.02)	-0.03	**χ^2(1, N=38)=6.41, p<0.05**	**χ^2(1, N=38)=4.75, p<0.05**
	Save result	0.34 (0.04)	0.44 (0.03)	+0.10	0.23 (0.03)	0.10 (0.02)	-0.13	χ^2(1, N=42)=3.69, p=0.05	χ^2(1, N=42)=0.63, p=0.43
	Go to workspace	0.20 (0.04)	0.21 (0.03)	+0.01	0.12 (0.04)	0.11 (0.03)	-0.01	χ^2(1, N=28)=0.22, p=0.64	χ^2(1, N=28)=1.89, p=0.18
	Change layout	0.26 (0.06)	0.20(0.03)	-0.06	0.29 (0.02)	0.32 (0.04)	+0.03	χ^2(1, N=17)=1.73, p=0.21	χ^2(1, N=17)=0.12, p=0.74
	Create gallery	0.19 (0.03)	0.17 (0.03)	-0.02	0.14 (0.03)	0.23 (0.04)	+0.09	χ^2(1, N=35)=3.80, p=0.06	χ^2(1, N=35)=0.001, p=0.97
	Add result to gallery	0.42 (0.05)	0.46 (0.03)	+0.04	0.13 (0.03)	0.06 (0.02)	-0.07	χ^2(1, N=19)=0.82, p=0.38	χ^2(1, N=19)=1.32, p=0.27
	Like result in workspace	0.21 (0.04)	0.24 (0.05)	+0.03	0.17 (0.03)	0.10 (0.02)	-0.07	χ^2(1, N=17)=0.02, p=0.88	χ^2(1, N=17)=0.53, p=0.48
	⊖ Delete gallery	0.34 (0.02)	0.24 (0.02)	-0.10	0.29 (0.03)	0.40 (0.03)	+0.11	**χ^2(1, N=21)=13.12, p<0.01**	**χ^2(1, N=21)=21.02, p<0.001**
	⊕ Keyword	0.40 (0.02)	0.56 (0.03)	+0.16	0.19 (0.03)	0.15 (0.02)	-0.04	**χ^2(1, N=40)=6.61, p<0.05**	**χ^2(1, N=40)=4.57, p<0.05**
	⊕ Reset button	0.20 (0.03)	0.35 (0.04)	+0.15	0.25 (0.02)	0.09 (0.04)	-0.16	**χ^2(1, N=17)=9.73, p<0.01**	**χ^2(1, N=17)=6.13, p<0.05**
	⊕ Keyword in workspace	0.36 (0.04)	0.53 (0.04)	+0.17	0.13 (0.02)	0.09 (0.03)	-0.04	**χ^2(1, N=52)=6.79, p<0.01**	**χ^2(1, N=52)=5.67, p<0.05**
	⊕ Reset button in workspace	0.28 (0.03)	0.45 (0.04)	+0.17	0.29 (0.05)	0.11 (0.02)	-0.17	**χ^2(1, N=22)=6.80, p<0.01**	**χ^2(1, N=22)=5.64, p<0.05**

Table 10. Summary of fixations and dwell time (mean and standard deviation) for AOIs across the Baseline (B), Result-focused (RF), and SERP-focused (SF) interfaces (* indicates statistically significant differences versus the Baseline).

Feature Type	AOI	Fixations - B	Fixations - RF	Fixations - SF	ANOVA	Dwell Time - B	Dwell Time - RF	Dwell Time - SF	ANOVA
Input	Query input box	4.94 (1.68)	4.72 (2.21)	4.78 (2.36)	F(2,51)=3.80, p=0.06	1.35 (0.54)	1.14 (0.67)	1.21 (0.83)	F(2,51)=0.13, p=0.72
Control	Search results navigation	2.67 (1.48)	1.22 (0.80)	1.50 (1.15)	F(2,51)=1.74, p=0.20	2.05 (0.68)	1.75 (0.55)	1.61 (0.62)	F(2,51)=1.89, p=0.18
	Search filters	4.33 (2.88)	5.61 (1.95)	4.61 (2.23)	F(2,51)=0.73, p=0.40	4.50 (0.70)	4.80 (0.60)	5.01 (0.61)	F(2,51)=2.88, p=0.09
Informational	Search result	4.11 (1.05)	2.94 (1.35)	5.56 (1.91)	F(2,51)=1.85, p=0.18	3.17 (0.52)	2.38 (0.43)	1.97 (0.78)	F(2,51)=1.76, p=0.20
	Search result in workspace	4.22 (1.99)	4.39* (2.84)	4.83* (1.25)	**F(2,51)=12.21, p<0.01**	3.33 (0.53)	2.05* (0.65)	1.53* (0.50)	**F(2,51)=5.85, p<0.05**
Personalization	Like result	2.00 (1.33)	3.50 (1.57)	2.11 (0.56)	F(2,51)=2.35, p=0.13	1.52 (0.70)	1.33 (0.73)	1.73 (0.94)	F(2,51)=0.56, p=0.46
	Save result	5.44 (1.34)	2.50* (1.28)	2.33* (1.06)	**F(2,51)=10.27, p<0.01**	4.31 (0.57)	1.71* (0.62)	1.57* (0.60)	**F(2,51)=16.32, p<0.001**
	"My Profile"	2.17 (1.66)	2.11 (1.14)	1.56 (0.78)	F(2,51)=1.86, p=0.19	1.06 (0.75)	1.43 (0.80)	1.44 (0.77)	F(2,51)=0.83, p=0.37
	Change layout	1.56 (1.03)	1.28 (0.82)	0.94 (0.40)	F(2,51)=1.74, p=0.20	1.34 (0.43)	1.77 (0.41)	2.05 (0.46)	F(2,51)=0.53, p=0.48
	Create gallery	1.11 (0.77)	1.83 (1.19)	1.94 (1.35)	F(2,51)=0.26, p=0.61	1.24 (0.46)	2.34 (0.71)	2.56 (0.62)	F(2,51)=0.06, p=0.81
	Add result to gallery	2.39 (1.37)	2.11 (0.99)	1.06 (0.51)	F(2,51)=1.55, p=0.23	1.87 (0.70)	1.43 (0.96)	1.93 (0.44)	F(2,51)=2.04, p=0.16
	Like result in workspace	2.56 (0.36)	1.17 (0.56)	0.94 (0.48)	F(2,51)=0.09, p=0.77	1.24 (0.72)	1.52 (0.91)	1.30 (0.98)	F(2,51)=0.04, p=0.84
	Delete gallery	1.44 (1.02)	0.94 (0.52)	1.17 (0.53)	F(2,51)=0.58, p=0.45	2.97 (0.62)	1.53 (0.77)	2.14 (0.61)	F(2,51)=1.10, p=0.30
	Keyword	-	1.88 (1.37)	2.22 (2.14)	F(1,34)=1.34, p=0.25	-	1.33 (0.55)	1.01 (0.43)	F(1,34)=0.78, p=0.38
	Reset button	-	0.83 (0.90)	0.94 (1.23)	F(1,34)=0.31, p=0.58	-	1.73 (0.53)	1.47 (0.54)	F(1,34)=1.66, p=0.21
	Keyword in workspace	-	2.33 (1.56)	2.88 (2.42)	F(1,34)=0.72, p=0.40	-	0.93 (0.43)	1.15 (0.43)	F(1,34)=1.86, p=0.18
	Reset button in workspace	-	1.00 (0.96)	1.22 (0.83)	F(1,34)=0.59, p=0.45	-	1.33 (0.83)	1.19 (0.50)	F(1,34)=0.10, p=0.75

In addition to providing context to the results reported above, we identified two search interface features where there were differences between the Baseline interface and the two visual keyword/result linking interfaces. Participants had fewer fixations on search results in the workspace, but had longer dwell times on them. As noted above, this resulted in negative reactions, which highlights again the challenges the participants experienced using the Baseline interface to evaluate the search results in the workspace.

Participants also had more frequent fixations on the save result feature in the Baseline, and had longer dwell times on this before deciding not to use the feature. Connecting this back to the emotional responses, these same participants had negative reactions to this passive interaction. As the same search results were provided for all interfaces, this is evidence of dissatisfaction with their ability to decide relevance, which was not present when they could interactively use keywords to show relationships among the search results.

7.2.6 Summary. As a result of integrating eye tracking with emotional valence and then removing the interaction instances, we were able to isolate passive interaction with the search interfaces features and identify patterns from this data.

In some of the cases identified in this study, viewing an interface feature and choosing not to use it generated a negative reaction across all interfaces (search filters, delete gallery). When searchers are unsure how to use a feature or are uncertain about what it will do to their previous work, they have a negative reaction. Further, negative reactions were found when viewing search results but choosing not to interact with them, which is an expected outcome for irrelevant search results.

A noteworthy difference between the Baseline interface and the two visual keyword/result linking interfaces was found with respect to viewing saved search results in the workspace. Doing so gave participants a negative reaction in the Baseline and a positive reaction in the other interfaces. There was also a difference in the frequency of fixations and the dwell time: participants viewed such search results less frequently, but when doing so spent more time considering them. This showcases the lack of satisfaction in what was found and saved when using the Baseline, and on the flip-side, the benefit that the visual keyword/result linking interfaces provided in helping the participants to find and easily confirm that what was found was beneficial.

There was also a noteworthy difference with the Baseline when considering whether to save a result and choosing not to. Participants who used the Baseline had a negative reaction to this. There were more fixations on this feature than with the other two interfaces, and more time was spent considering whether to use the feature and ultimately choosing not to. Conversely, when participants used the visual keyword/result linking interfaces, they were able to use the keyword highlighting feature to surface relationships among the search results, which meant that they could leverage a decision made about relevance of one search result to more easily make a decision about other related ones. While this did not generate an emotional reaction, fewer fixations were needed and the consideration and choice not to save a result occurred more quickly.

The novel features of the Result-focused and SERP-focused interfaces are the surfacing of the keywords to support the search process. Even when these were used passively, there were positive reactions to them. The passive interaction with the reset button resulting in a positive reaction is an example of how not using reset or undo features can highlight the value of the interaction with the features they reset or undo.

8 Discussion

The research questions were motivated by an opportunity to connecting eye tracking data with feature use and measurements of emotional valence. This study examined participants' emotional reactions to the use of features of the search interfaces through two approaches: integrating this data to assess the typical interaction, and isolating the features that were viewed but not interacted with to assess the passive (gaze-only) interaction. We designed this dual approach to investigate the nature of these two types of interactions with respect to the features of three search interfaces.

In analyzing the typical interaction, we were able to find valuable aspects of all three interfaces whose use resulted in positive reactions. We were also able to identify features whose use resulted in different emotional responses depending on which interface was being used by the participants. The key findings are summarized in Figure 5. The feature-level granularity of these results provides insight into the specific aspects of the interfaces that were viewed positively and negatively while being used.

Data to support such findings are not available with multimodal studies that use system-level survey instruments such as the technology acceptance model [71] or the user engagement scale [53]. While similar data could be collected via post-task questionnaires that ask questions about each individual feature in an interface, these may result in survey fatigue and be subject to recency effects [22, 57]. It may also be possible to collect similar data using retrospective think-aloud protocols [48],

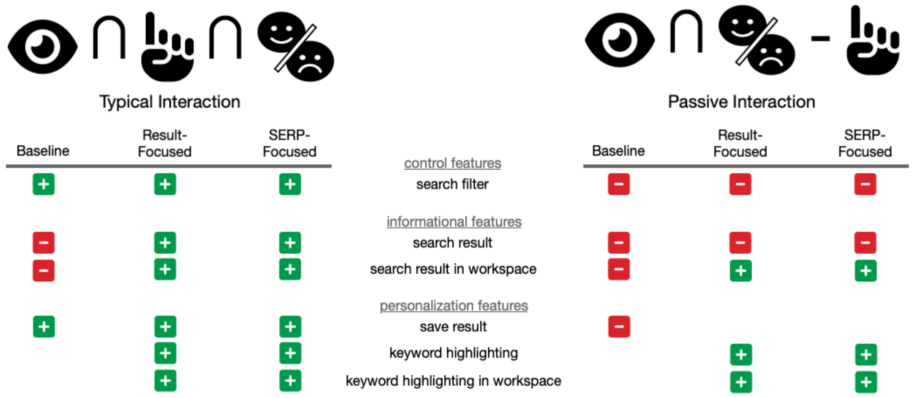


Fig. 5. A summary of the key features whose interaction (typical and passive) resulted in positive or negative emotions.

but in addition to being influenced by participants' memory, they too are subject to recency effects. We were able to collect such data unobtrusively, allowing the study to proceed with minimal delay between tasks for collecting subjective opinions.

In analyzing the passive (gaze-only) interaction, we were able to identify features whose observation but subsequent use elected negative reactions, and others that elicited positive reactions. While this approach does not provide data to explain why the searchers had such reactions, we are able to make inferences from the findings based on the nature of the features and the differences between the interfaces.

As with the typical interactions, these passive interactions were identified in an unobtrusive manner, which contributes to measurement validity. To measure these in other manners, such as using post-task questionnaires, would be difficult. While we can expect a participant in a study to form an opinion about the features they explicitly use, they may have difficulty forming such opinions based on the passive interaction with features.

In previous studies of visual keyword/result linking search interfaces [28, 29], typical IIR evaluation methods were used and inferences were made based on the high-level differences between the interfaces. With the method presented in this paper, and the study conducted using this method, we were able to conduct a feature-level study in an unobtrusive way so as to not interfere with the flow of performing the searches. More importantly, we were able to isolate not only the typical interaction with these features, but also the passive interaction. From this, we identified when such interactions resulted in positive and negative reactions, giving us the ability to make inferences regarding the benefits or challenges the participants faced when using these features.

Relating the findings of this study to a prior evaluation of these interfaces (which utilized a standard user study framework with post-task subjective measures such as usefulness, ease of use, satisfaction, and user engagement)[28] provides feature-level explanations for those findings. That is, we can now attribute the statistically significant differences in the subjective measures specifically to the positive emotional reaction to when using the visual keyword/result linking methods to reveal relationships among the search results both within the SERP and the workspace. Such positive reactions were also present even when viewing the keywords and choosing not to use them. Further, in the earlier study we found higher a higher precision among the saved resources compared to the baseline; in this study we found positive reactions to both viewing and accessing these saved resources when using the new interfaces, providing corroborating evidence for the support the interfaces provide to helping the searchers find and saved relevant resources. For the tasks used in this study, there remains little difference between whether they keywords are presented in a result-focused or SERP-focused manner.

In addition, the observed emotional responses provide insight into utility and alignment of search interface features with search activity requirements. Positive emotional reactions such as those elicited when saving results in the workspace, indicate satisfaction with the relevance of saved search results. In contrast, negative emotional reactions, such as those observed during passive interactions with the search filter, may reflect frustration due to its complexity or uncertainty in how it might support fulfilling the search task objectives. These findings suggest that emotional reactions are closely tied to the perceived usefulness and ease of use of a feature, reinforcing the importance of intuitive and task-aligned interface design. By linking emotional reactions to the practical utility of the features, we highlight the need for search interface designs that reduce cognitive load and align with searchers' mental models.

Findings of our study align with those of scholars [7, 47] who have discussed broader trends in interface design and their relationship to emotion. By exploring how particular features of search interfaces elicit emotional reactions from users during exploratory search tasks, our study contributes to this ongoing discussion. It is also consistent with the emphasis on emotional aspects of interface design [19, 78].

A limitation of this study is the possibility that the facial expressions captured may reflect task difficulty (workload) rather than emotional responses directly related to the interface features. To address this concern, perceived task difficulty was assessed both before and after each task using participant surveys. The collected responses did not reveal any significant outliers, suggesting workload perceptions were consistent. This consistency reduces the likelihood that variations in facial expressions were primarily driven by task difficulty rather than interface design.

Another limitation of this study lies in the reliance on machine learning-based emotion classification methods. While such methods are robust, they are subject to occasional inaccuracies, introducing a moderate risk of mis-classification. Factors such as the quality of video frames from front-facing cameras, variability in lighting conditions, unusual facial expressions (e.g., talking, chewing, or drinking), and rigid head movements further impact the accuracy of emotional response detection. To mitigate these potential inaccuracies, we employed strict data inclusion thresholds, discarding low-confidence frames and ensuring a consistent recording setup across participants. While this means that we cannot claim that this approach can replace other established methods for evaluating IIR systems, we have shown how it can be used to identify both positive and negative feature-level interactions.

Although the relatively small sample size of 18 participants limits the statistical power, more importantly it limited our ability to reveal differences between the two different approaches for visual keyword/result linking. A larger study in which we can classify participants based on their preferences for local information (result-focused) versus global information (SERP-focused) may provide further insight into how best to support exploratory search with visual keyword/result linking approaches.

This study is also limited by the variability of environmental conditions across three data collection locations. However, all data was collected by the same members of our research team to ensure consistency. The sessions were conducted in private meeting rooms with minimal visual distractions; these locations had comparable setups, including even lighting and standardized seating arrangements.

9 Conclusion

This work makes three contributions to the body of knowledge on evaluating interactive information retrieval systems: (1) the development of an innovative method integrating eye tracking, feature use, and emotional valence to isolate both typical interaction and passive interaction with search interface features; (2) the validation of the feasibility of using this method in the context of a controlled laboratory experiment; and (3) feature-level evidence of the value of visual keyword/result linking approaches to support exploratory search within digital humanities archives.

The value this method brings to conducting user studies is the ability to pinpoint specific features that participants found beneficial or problematic, both through typical interaction and when the feature is the focus of visual attention but not explicitly used. Such insight is difficult to capture otherwise, and may be used to further refine existing search interfaces or inspire new ways of presenting information or providing interactive features, and considering whether they add value through typical interaction, passive interaction, or both.

In the context of our study of visual keyword/result linking approaches to support exploratory search within digital humanities archives, this method has allowed us to more precisely consider the impact of the feature-level differences between the interfaces under investigation. Surfacing keyword information and making it interactive resulted in positive reactions, even when the interactive aspect was not used. The similar reaction to this information regardless of how it was presented in the interfaces (Result-Focused versus SERP-Focused) suggests that the differences may come down to individual preferences. The evidence of positive emotions when using the keywords (typically and passively) provides motivation for the continued study of visual methods for revealing relationships among search results within exploratory search settings.

As front-facing cameras improve, emotional detection approaches become faster, more accurate, and ethical, and eye tracking is integrated within user interface hardware, there may be an opportunity to use this approach to create adaptive exploratory search interfaces that are responsive to searcher interaction (typical and passive) based on emotional reactions. Such adaptation could identify when useful information is found and automatically highlight related content, detect when frustration with a feature occurs and suggest how it can be used more effectively, or adjust the simplicity/complexity of the overall interface depending on which features are being used versus those that are being considered and not used. Further, there may be the possibility of detecting the fluctuating uncertainty that is an inherent aspect of exploratory search [72], tuning the features of the search interface as the searcher transitions from exploratory browsing to focused searching, and back to exploratory browsing when new information is found that causes an uncertainty rebound [32]. Such an approach would enhance the user experience by making the search process more intuitive when undertaking complex tasks and pursuing exploratory search processes, ensuring that both typical interaction and passive behaviours contribute to more responsive and user-centered interfaces.

Our more immediate future work is to continue to focus on how eye tracking and emotional responses to search interface feature use can enhance the evaluation of such interfaces. Streamlining the approach will make it more accessible to other researchers. Using it to study novel visual and interactive features added to search interfaces will help to validate the individual contributions of those features at both the typical and passive interaction level. There are also opportunities to generalize this approach beyond the context of search interface evaluation, using it to study user interfaces designed for performing knowledge-centred tasks at the feature-level.

Acknowledgments

The authors acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), through the Discovery Grants (RGPIN-2017-06446 and RGPIN-2023-04070) held by the second author.

The authors acknowledge the Woi wurrung and Boon wurrung language groups of the eastern Kulin Nation on whose unceded lands ACM SIGIR CHIIR 2025 was hosted. We pay our respect to their Elders past and present and extend that respect to all Aboriginal and Torres Strait Islander peoples today and their connections to land, sea, sky, and community.

References

[1] Ioannis Arapakis, Joemon M. Jose, and Philip D. Gray. 2008. Affective feedback: An investigation into the role of emotions in the information seeking process. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 395–402. https://doi.org/10.1145/1390334.1390403

[2] Ioannis Arapakis, Ioannis Konstas, Joemon M. Jose, and Ioannis Kompatsiaris. 2009. Modeling facial expressions and peripheral physiological signals to predict topical relevance. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 728–729. https://doi.org/10.1145/1571941.1572099

[3] Ioannis Arapakis and Luis A. Leiva. 2020. Learning efficient representations of mouse movements to predict user attention. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 1309–1318. https://doi.org/10.1145/3397271.3401031

[4] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. 2016. Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology* 67, 11 (2016), 2635–2651. https://doi.org/10.1002/asi.23617

[5] Tamara Babaian, Wendy Lucas, and Heikki Topi. 2006. Making memories: Applying user input logs to interface design and evaluation. In *Proceedings of the CHI Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, 496–501. https://doi.org/10.1145/1125451.1125559

[6] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. OpenFace: An open source facial behavior analysis toolkit. In *Proceedings of the IEEE Conference on Applications of Computer Vision*. 1–10. https://doi.org/10.1109/WACV.2016.7477553

[7] Maram Barifah and Monica Landoni. 2020. Emotions associated with failed searches in a digital library. In *Proceedings of ISIC, the Information Behaviour Conference*, Vol. 25. University of Borås, Sweden. https://doi.org/10.47989/irisic2027

[8] Nilavra Bhattacharya and Jacek Gwizdka. 2019. Measuring learning during search: Differences in interactions, eye-gaze, and semantic similarity to expert knowledge. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, New York, NY, 63–71. https://doi.org/10.1145/3295750.3298926

[9] Tanja Blascheck and Thomas Ertl. 2014. Towards analyzing eye tracking data for evaluating interactive visualization systems. In *Proceedings of the Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. ACM, New York, NY, 70–77. https://doi.org/10.1145/2669557.2669569

[10] Pia Borlund. 2003. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research* 8, 3 (2003), 3–8. http://informationr.net/ir/8-3/paper152.html

[11] B.R. Boyce, C.T. Meadow, and D.H. Kraft. 1994. *Measurement in Information Science*. Academic Press, San Diego, California. https://books.google.ca/books?id=JQXhAAAAMAAJ

[12] Jin Cai and Guoyong Cai. 2023. Multimodal visual question answering model enhanced with image emotional information. In *Proceedings of the International Conference on Natural Language Processing*. IEEE, New York, NY, 268–273. https://doi.org/10.1109/ICNLP58431.2023.00056

[13] Zeljko Carevic, Maria Lusky, Wilko van Hoek, and Philipp Mayr. 2018. Investigating exploratory search activities based on the stratagem level in digital libraries. *International Journal on Digital Libraries* 19, 2 (2018), 231–251. https://doi.org/10.1007/s00799-017-0226-6

[14] Vincenzo Deufemia, Massimiliano Giordano, Giuseppe Polese, and Genoveffa Tortora. 2013. Capturing user's interest from human-computer interaction logging. In *Proceedings of the Web Information Systems and Technologies*, José Cordeiro and Karl-Heinz Krempels (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 312–327. https://doi.org/10.1007/978-3-642-36608-6_20

[15] Joseph Dumas and Jean Fox. 2007. *Chapter 57: Usability Testing: Current Practice and Future Directions*. CRC Press, Florida, USA, 1136–1143. https://doi.org/10.1201/9781410615862

[16] Paul Ekman. 2003. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Times Books/Henry Holt and Co, New York, NY.

[17] Europeana. 2024. Discover Europe's digital cultural heritage. https://www.europeana.eu/en

[18] Europeana. 2024. Documentation on The Europeana Search API. https://pro.europeana.eu/page/search

[19] Zhimin Gao and Jiaxi Huang. 2022. Human-computer interaction emotional design and innovative cultural and creative product design. *Frontiers in Psychology* 13 (2022), 982303. https://doi.org/10.3389/fpsyg.2022.982303

[20] Laura A. Granka, Thorsten Joachims, and Geri Gay. 2004. Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 478–479. https://doi.org/10.1145/1008992.1009079

[21] Shirley Gregor, Aleck C. H. Lin, Tom Gedeon, Amir Riaz, and Dingyun Zhu. 2014. Neuroscience and a nomological network for the understanding and assessment of emotions in information systems research. *Journal of Management Information Systems* 30, 4 (2014), 13–48. https://doi.org/10.2753/MIS0742-1222300402

[22] Marc Hassenzahl and Nina Sandweg. 2004. From mental effort to perceived usability: Transforming experiences into summary assessments. In *Proceedings of the CHI Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, 1283–1286. https://doi.org/10.1145/985921.986044

[23] Marti Hearst. 2009. *The Design of Search User Interfaces*. Cambridge University Press, Cambridge, UK, 1–28. https://doi.org/10.1017/CBO9781139644082

[24] Fabio Hellmann, Silvan Mertes, Mohamed Benouis, Alexander Hustinx, Tzung-Chien Hsieh, Cristina Conati, Peter Krawitz, and Elisabeth André. 2024. GANonymization: A gan-based face anonymization framework for preserving emotional expressions. *ACM Transactions on Multimedia Computing, Communications, and Applications* 21, 1 (2024). https://doi.org/10.1145/3641107

[25] Timothy Hill, Valentine Charles, Juliane Stiller, and Antoine Isaac. 2016. "Searching for inspiration": User needs and search architecture in Europeana collections. *Proceedings of the Association for Information Science and Technology* 53, 1 (2016), 1–7. https://doi.org/10.1002/pra2.2016.14505301043

[26] Orland Hoeber. 2018. Information visualization for interactive information retrieval. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, New York, NY, 371–374. https://doi.org/10.1145/3176349.3176898

[27] Orland Hoeber. 2025. Principles of exploratory search interface design. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery, New York, NY, USA.

[28] Orland Hoeber, Morgan Harvey, Milad Momeni, Abbas Pirmoradi, and David Gleeson. 2024. Exploratory search in digital humanities: A study of visual keyword/result linking. *Proceedings of the Association for Information Science and Technology* 61, 1 (2024), 161–171. https://doi.org/10.1002/pra2.1017

[29] Orland Hoeber, Abbas Pirmoradi, Sebastian Gomes, Baran Erfani, Zakiyyah Noorally, and Yug Shah. 2024. Visually keyword/result linking: Using interaction to dynamically reveal relationshiops. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, New York, NY, 66–76. https://doi.org/10.1145/3627508.3638307

[30] Kajta Hofmann, Bhaskar Mitra, Filip Radlinski, and Milad Shokouhi. 2014. An eye-tracking study of user interactions with query auto completion. In *Proceedings of the ACM International Conference on Conference on Information and Knowledge Management*. ACM, New York, NY, 549–558. https://doi.org/10.1145/2661829.2661922

[31] Bernard J. Jansen. 2006. Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research* 28, 3 (2006), 407–432. https://doi.org/10.1016/j.lisr.2006.06.005

[32] Tingting Jiang. 2014. *Exploratory search: A critical analysis of the theoretical foundations, system features, and research trends*. Springer, Berlin, Heidelberg, 79–103. https://doi.org/10.1007/978-3-642-54812-3_7

[33] Oleg Kalyta, Olexander Barmak, Pavlo Radiuk, and Iurii Krak. 2023. Facial emotion recognition for photo and video surveillance based on machine learning and visual analytics. *Applied Sciences* 13, 17 (2023). https://doi.org/10.3390/app13179890

[34] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, New York, NY, 1151–1160. https://doi.org/10.1145/2638728.2641695

[35] Diane Kelly. 2009. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval* 3, 1–2 (2009), 1–224. https://doi.org/10.1561/1500000012

[36] Max Kemman, Martijn Kleppe, and Jim Maarseveen. 2013. Eye Tracking the use of a collapsible facets panel in a search interface. In *Research and Advanced Technology for Digital Libraries*, Trond Aalberg, Christos Papatheodorou, Milena Dobreva, Giannis Tsakonas, and Charles J. Farrugia (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 405–408. https://doi.org/10.1007/978-3-642-40501-3_47

[37] Davis E. King. 2009. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research* 10 (2009), 1755–1758. https://api.semanticscholar.org/CorpusID:6155330

[38] Maja Kuhar and Tanja Merčun. 2022. Exploring user experience in digital libraries through questionnaire and eye-tracking data. *Library & Information Science Research* 44, 3 (2022), 101175. https://doi.org/10.1016/j.lisr.2022.101175

[39] Carol Collier Kuhlthau. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science* 42, 5 (1991), 361–371. https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#

[40] Bill Kules and Robert Capra. 2009. Designing exploratory search tasks for user studies of information seeking support systems. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries*. ACM, New York, NY, 419–420. https://doi.org/10.1145/1555400.1555492

[41] Bill Kules and Robert Capra. 2012. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. *Journal of the American Society for Information Science and Technology* 63, 1 (2012), 114–138. https://doi.org/10.1002/asi.21647

[42] Xiangsheng Li, Maarten de Rijke, Yiqun Liu, Jiaxin Mao, Weizhi Ma, Min Zhang, and Shaoping Ma. 2021. Investigating session search behavior with knowledge graphs. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 1708–1712. https://doi.org/10.1145/3404835.3463107

[43] Han-Chin Liu and Hsueh-Hua Chuang. 2011. An examination of cognitive processing of multimedia information based on viewers' eye movements. *Interactive Learning Environments* 19, 5 (2011), 503–517. https://doi.org/10.1080/10494820903520123

[44] Ying-Hsang Liu, Paul Thomas, Marijana Bacic, Tom Gedeon, and Xindi Li. 2017. Natural Search user interfaces for complex biomedical search: An eye tracking study. *Journal of the Australian Library and Information Association* 66, 4 (2017), 364–381. https://doi.org/10.1080/24750158.2017.1357915

[45] Damien Lockner, Nathalie Bonnardel, Carole Bouchard, and Vincent Rieuf. 2014. Emotion and interface design. In *Proceedings of the Ergonomie et Informatique Avancée Conference - Design, Ergonomie et IHM: Quelle Articulation Pour La Co-Conception de l'interaction*. ACM, New York, NY, 33–40. https://doi.org/10.1145/2671470.2671475

[46] Irene Lopatovska. 2014. Toward a model of emotions and mood in the online information search process. *Journal of the Association for Information Science and Technology* 65, 9 (2014), 1775–1793. https://doi.org/10.1002/asi.23078

[47] Irene Lopatovska and Ioannis Arapakis. 2011. Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. *Information Processing & Management* 47, 4 (2011), 575–592. https://doi.org/10.1016/j.ipm.2010.09.001

[48] Menno De Jong Maaike van den Haak and Peter Jan Schellens. 2003. Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology* 22, 5 (2003), 339–351. https://doi.org/10.1080/0044929031000

[49] David Maxwell and Claudia Hauff. 2021. LogUI: Contemporary logging infrastructure for web-based experiments. In *Proceedings of the European Conference on Information Retrieval*. Springer, Cham, Switzerland, 525–530. https://doi.org/10.1007/978-3-030-72240-1_59

[50] Mickael Ménard, Paul Richard, Hamza Hamdi, Bruno Dauce, and Takehiko Yamaguchi. 2015. Emotion recognition based on heart rate and skin conductance. In *Proceedings of the International Conference on Physiological Computing System*. Science and Technology, Setúbal, Portugal, 26–32. https://doi.org/10.5220/0005241100260032

[51] Gheorghe Muresan and Bing Bai. 2007. Exploring interactive information retrieval: An integrated approach to interface design and interaction analysis. In *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, Paris, 712–718. https://dl.acm.org/doi/10.5555/1931390.1931458

[52] Heather L. O'Brien. 2009. D. Nahl, D. Bilal (eds.): Information and emotion: The emergent affective paradigm in information behaviour research and theory. *Information Retrieval* 12, 5 (2009), 605–608. https://doi.org/10.1007/s10791-009-9095-y

[53] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

[54] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. 2015. Deep face recognition. In *Proceedings of the British Machine Vision Conference*, Xianghua Xie, Mark W. Jones, and Gary K. L. Tam (Eds.). BMVA Press, 41.1–41.12. https://doi.org/10.5244/C.29.41

[55] Nazil Perveen, Debaditya Roy, and Krishna Mohan Chalavadi. 2020. Facial expression recognition in videos using dynamic kernels. *IEEE Transactions on Image Processing* 29 (2020), 8316–8325. https://doi.org/10.1109/TIP.2020.3011846

[56] Manop Phankokkruad and Phichaya Jaturawat. 2015. An evaluation of technical study and performance for real-time face detection using web real-time communication. In *Proceedings of the Conference on Computer, Communications, and Control Technology*. IEEE, New York, NY, 162–166. https://doi.org/10.1109/I4CT.2015.7219558

[57] Abbas Pirmoradi and Orland Hoeber. 2025. Bridging in-task emotional responses with post-task evaluations in digital library search interface user studies. *Information Processing & Management* (2025).

[58] Wataru Sato and Sakiko Yoshikawa. 2023. Influence of stimulus manipulation on conscious awareness of emotional facial expressions in the match-to-sample paradigm. *Scientific Reports* 13, 1 (2023), 20727. https://doi.org/10.1038/s41598-023-47995-9

[59] Reijo Savolainen. 2015. The interplay of affective and cognitive factors in information seeking and use. *Journal of Documentation* 71, 1 (2015), 175–197. https://doi.org/10.1108/JD-10-2013-0134

[60] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 815–823. https://doi.org/10.1109/CVPR.2015.7298682

[61] Sefik Ilkin Serengil and Alper Ozpinar. 2021. HyperExtended LightFace: A facial attribute analysis framework. In *Proceedings of the International Conference on Engineering and Emerging Technologies*. IEEE, New York, NY, 1–4. https://doi.org/10.1109/ICEET53442.2021.9659697

[62] Prabin Sharma, Shubham Joshi, Subash Gautam, Sneha Maharjan, Salik Ram Khanal, Manuel Cabral Reis, João Barroso, and Vítor Manuel de Jesus Filipe. 2022. Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. In *Proceedings of the Technology and Innovation in Learning, Teaching and Education*, Arsénio Reis, João Barroso, Paulo Martins, Athanassios Jimoyiannis, Ray Yueh-Min Huang, and Roberto Henriques (Eds.), Vol. 1720. Springer Nature, Cham, Switzerland, 52–68. https://doi.org/10.1007/978-3-031-22918-3_5

[63] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. 2014. Deep learning face representation by joint identification-verification. In *Proceedings of the International Conference on Neural Information Processing Systems - Volume 2*. MIT Press, Cambridge, MA, USA, 1988–1996. https://dl.acm.org/doi/10.5555/2969033.2969049

[64] Y. Sun, X. Wang, and X. Tang. 2014. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 1891–1898. https://doi.org/10.1109/CVPR.2014.244

[65] Jonathan Sykes, Milena Dobreva, Duncan Birrell, Emma McCulloch, Ian Ruthven, Yurdagül Ünal, and Pierluigi Feliciati. 2010. A new focus on end users: Eye-tracking analysis for digital libraries. In *Research and Advanced Technology for Digital Libraries*, Mounia Lalmas, Joemon Jose, Andreas Rauber, Fabrizio Sebastiani, and Ingo Frommholz (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 510–513. https://doi.org/10.1007/978-3-642-15464-5_69

[66] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. 2014. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1701–1708. https://doi.org/10.1109/CVPR.2014.220

[67] Carla Teixeira Lopes and Edgar Ramos. 2020. Studying how health literacy influences attention during online information seeking. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, New York, NY, 283–291. https://doi.org/10.1145/3343413.3377966

[68] Giannis Tsakonas and Christos Papatheodorou. 2007. Critical constructs of digital library interaction. In *Proceedings of the Panhellenic Conference on Informatics*. 57–67. https://api.semanticscholar.org/CorpusID:55030751

[69] Robin Turkington, Maurice Mulvenna, Raymond Bond, S. O'Neill, and C. Armour. 2018. The application of user event log data for mental health and wellbeing analysis. In *Proceedings of the British Human Computer Interaction Conference*. BCS Learning and Development Ltd, 1–14. https://dx.doi.org/10.14236/EWIC/HCI2018.4

[70] Pertti Vakkari, Arto Luoma, and Janna Pöntinen. 2014. Books' interest grading and dwell time in metadata in selecting fiction. In *Proceedings of the Information Interaction in Context Symposium*. ACM, New York, NY, 28–37. https://doi.org/10.1145/2637002.2637008

[71] Viswanath Venkatesh and Fred Davis. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science* 46, 2 (2000), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

[72] Ryen White and Resa Roth. 2009. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan & Claypool Publishers, San Rafael, CA. https://doi.org/10.2200/S00174ED1V01Y200901ICR003

[73] Max Wilson. 2012. *Search User Interface Design*. Morgan & Claypool Publishers, San Rafael, CA. https://doi.org/10.2200/S00371ED1V01Y201111ICR020

[74] Kochiu Wu and Yi-Hsieh Huang. 2018. Emotions and eye-tracking of differing age groups searching on e-book wall. *Aslib Journal of Information Management* 70, 4 (2018), 434–454. https://doi.org/10.1108/AJIM-01-2018-0017

[75] Xiaohui Xie, Yiqun Liu, Xiaochuan Wang, Meng Wang, Zhijing Wu, Yingying Wu, Min Zhang, and Shaoping Ma. 2017. Investigating examination behavior of image search users. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 275–284. https://doi.org/10.1145/3077136.3080799

[76] Weijing Yuan and Charles T. Meadow. 1999. A study of the use of variables in information retrieval user studies. *Journal of the American Society for Information Science* 50, 2 (1999), 140–150. https://doi.org/10.1002/(SICI)1097-4571(1999)50:2<140::AID-ASI5>3.0.CO;2-P

[77] Hae Seon Yun, Heiko Hübert, Johann Chevalère, Niels Pinkwart, Verena V. Hafner, and Rebecca Lazarides. 2023. Analyzing learners' emotion from an HRI experiment using facial expression recognition systems. In *Proceedings of the Learning and Collaboration Technologies*, Panayiotis Zaphiris and Andri Ioannou (Eds.). Springer Nature Switzerland, Cham, Switzerland, 396–407. https://doi.org/10.1007/978-3-031-34550-0_29

[78] Yang Zhao, Dan Xie, Ruoxin Zhou, Ning Wang, and Bin Yang. 2022. Evaluating users' emotional experience in mobile libraries: An emotional model based on the pleasure-arousal-dominance emotion model and the five factor model. *Frontiers in Psychology* 13 (2022), 942198. https://doi.org/10.3389/fpsyg.2022.942198

[79] Jean Éthier, Pierre Hadaya, Jean Talbot, and Jean Cadieux. 2008. Interface design and emotions experienced on B2C Web sites: Empirical testing of a research model. *Computers in Human Behavior* 24, 6 (2008), 2771–2791. https://doi.org/10.1016/j.chb.2008.04.004