Using Affective Signals as Implicit Indicators of Information Relevance and Information Processing Strategies

Roberto González-Ibáñez
Departamento de Ingeniería Informática
Universidad de Santiago de Chile
Santiago, Chile
roberto.gonzalez.i@usach.cl

Chirag Shah
School of Communication and Information (SC&I)
Rutgers University
New Brunswick, NJ, 08901
chirags@rutgers.edu

ABSTRACT

Search engines have become increasingly better at providing information to users. However, they still face major challenges, such as determining how searchers process information, how they make relevance judgments, and how their cognitive or emotional states affect their search progress. We address these challenges by exploring searchers' affective dimension. In particular, we investigate how feelings, facial expressions, and electrodermal activity (EDA) could help to understand information relevance, search progress, and information processing strategies (IPS). To meet this goal, we designed an experiment in which 45 participants were exposed to affective stimuli prior to solving a fact-finding search task. Results indicate that initial affective dimensions are linked to IPSs, search progress, and task completion. However, further analyses suggest that affective-related features alone have limited utility in the binary classification of relevance using machine learning techniques.

Keywords

Information Retrieval; Information Processing Strategies; Affective Signals.

INTRODUCTION

In the last fifteen years, search engines have evolved progressively from keyword-based indexing methods to the development of sophisticated ranking algorithms that use relevance feedback (Lv & Zhai, 2010), query expansion (Xu & Croft, 1996), and various forms of user modeling to provide personalization (Liu & Belkin, 2010) and recommendation. Some of the recent approaches have also started incorporating social signals (e.g. Carmel et al. 2010). As part of this evolutionary process, the interactions between users and search engines have also changed through the incorporation of different methods that searchers can use to represent their information needs (e.g. text, audio, images) and visualize the results retrieved by search engines. From a system perspective, it is expected that such improvements are able to fulfill the information needs of their users and, at the same time, improve their experience. Yet, systems fail in determining how their users process information and what they feel during their search processes. In the absence of explicit relevance judgments, search engines rely on implicit measures, such as dwell time on Web pages (Fox et al., 2005) and click-through behaviors (Hassan et al., 2010), to infer what pages are likely to be relevant to users, which in turn can help systems with the ranking process.

To fill this gap in the interaction between users and search engines, there have been attempts to explore hidden cognitive and affective aspects aiming to model how searchers interact with systems and the information provided by them (e.g. Gwizdka and Cole (2013); and Moshfeghi et al. (2013)). Inspired by particular findings in psychology, this article explores the affective dimension of searchers in order to identify elements that could be used to enhance the interaction between users and search engines. In particular, research in psychology has suggested that affective states may shape the way people face problems, make decisions, and process information (Isen et al., 1983; Forgas, 1991).

In this research, as an extension of our preliminary results (González-Ibáñez & Shah, 2015), we focused in the following two research questions:

- *(RQ1) Are there affective patterns derived from initial affective states that can be used as implicit indicators of IPS and information relevance?*
- *(RQ2) To what extent, if any, could affective-related features be used for automatic classification of information relevance?*

On one hand, to address RQ1, we conducted an experiment with two groups of participants exposed to affective in

ASIST 2016, October 14-18, 2016, Copenhagen, Denmark.
Copyright held by the authors.

order to evaluate their effects on information processing strategies and relevance judgments. On the other hand, to address RQ2 we used affective-related features from the data collected in the experiment in order to train and test classification models for information relevance.

In the next section we provide an overview of related work and theoretical background on this research topic. Following, we describe our methodological approach. Then we present the results of the study to conclude with the discussion and implications of our findings.

BACKGROUND

Research in information science has acknowledged the intrinsic role of affective processes in information search (Kuhlthau, 1991; Nahl, 2009). Recent studies have also shown that, with the support of technologies, people express different affective states while searching for information online. For example Arapakis et al. (2008), González-Ibáñez et al. (2011), and Lopatovska (2011) investigated facial expressions (FE) of users while working on a search task. Others have worked on predicting searchers' frustration levels using expressive, physiological, and behavioral signals, aiming to assist users in the search process (Feild et al., 2010). Some studies have shown that information search can lead searchers to experience positive or negative affective states (Poddar, 2010). Arapakis et al. (2008) and, more recently, Moshfeghi & Jose (2013), have shown that affective signals could be used in tandem with behavioral signals for implicit relevance feedback. While most research in this domain has focused on affective processes as intrinsic factors to the search process, few have considered the role of the affective dimension as an extrinsic component to the search process (e.g. How do affective states change the way people formulate queries and assess information?).

As part of information seeking and searching behaviors, information processing plays a central role. To this end, research in psychology has suggested that affective processes may have direct implications in the strategies that people employ to process information. For example, Sinclair and Mark (1995) showed through experimental studies (using emotion elicitation procedures) that people in positive affective states (with particular reference to happy participants) employed information processing strategies that are "relatively passive or nonsystematic, [and] less detailed [(accurate)]" (p. 417). Conversely, people in negative affective states (with particular reference to unhappy participants) were found to be "more active or systematic, [and] detailed" (p. 417). Along the same lines, Isen et al. (1983) found that decision making in a group of participants with induced positive affective states was faster than it was in a control group, which in turn contributed to the former group's higher efficiency. With regard to information processing strategies, the authors noted that the participants in positive affective states typically did not revisit information already seen. According to the Affect Infusion Model (AIM) (Forgas, 1991; Forgas & George, 2001), differences in the levels of influence of mood in information processing could be attributed to additional factors, such as familiarity with the situation, target complexity, specific motivations, cognitive capacity, and situational pragmatics.

In this context, a study conducted by Lopatovska (2009) reported that participants' positive mood prior to the start of a search task affected some search behaviors; however, the study relied mainly on participants' subjective evaluations through self-assessments to determine their initial affective states. Few studies have focused on different levels of affective processes and their relationship with the phenomenon under study. For instance, Palmero et al. (2006) distinguish four levels of affective processes, namely: feelings (subjective, self-awareness), emotions (objective, expressive, categorical, short duration), mood and affect (objective, internal, medium- to long-term durations, described in terms of dimensions such as valence and arousal). This distinction of affective processes has practical implications in the way studies are designed and also in the way affective processes are measured. Note that affective processes, affective states, and affective dimension are used in this article interchangeably to broadly refer to these four levels or categories.

To investigate affective processes, different techniques and methodological approaches are reported in the literature. In experimental settings, emotion elicitation techniques are commonly used by psychologists (Coan & Allen, 2007) to study the effects of specific affective processes in areas such as perception, decision making, and information processing. Besides methodological approaches, it is also necessary to establish adequate measurement approaches of affective processes. In this sense, different instruments have been developed and validated as reported in the literature. For example, questionnaires such as the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994) and PANAS (Watson, 1988) are used to measure self-reported affective experiences. On the other hand, facial expressions can be used to infer basic and complex emotions (Ekman, 1972; Izard, 1977). At the neurophysiological level, electrodermal activity (EDA), electroencephalogram (EEG), and functional magnetic resonance imaging (fMRI), to name a few, have been used to investigate internal changes linked to affective processes, such as mood and affect. In particular EDA is a response of the human body when the individual "becomes mentally, emotionally, or physically aroused" (Strauss et al., 2005, p. 701). Such response, as part of the sympathetic nervous system, has been linked to affective processes expressed by the sweat glands in the skin.

At the technological level, different devices and software can be used to monitor and measure affective processes objectively. For instance, FE can be automatically analyzed and classified into basic emotions and mental states by specialized software (Sebe et al., 2004; Küblbeck & Ernst,

2006). Likewise, to measure EDA, devices such as the *Affectiva Q Sensor* and *Bitalino* provide access to physiological changes almost unobtrusively.

METHODS

In order to address RQ1, we designed and conducted a controlled experiment. This section provides a detailed description of the study.

Experimental Design

We designed a multiple-group study involving two experimental groups and a control group. On one hand, the two experimental groups were linked to participants in positive ($C1^{+}$) and negative ($C2^{-}$) affective states. On the other hand, the control group ($C3^{ctrl}$) served as a baseline to perform comparisons and interpretation of results.

As shown in Table 1, the experimental design consisted of two major stages: (1) affective induction and (2) evaluation of prolonged effects (PE). In the first stage, the participants were randomly assigned to the two experimental groups and the control group. Participants in the experimental groups were treated with affective stimuli to elicit positive (X^{+}) and negative (X^{-}) affective states while performing a search task. More details about the affective stimuli stage and the search task are provided below. In the second stage (evaluation of PE), participants in the three groups performed the main search task (MT) (i.e. a series of search challenges with observations in between), which was similar to that in the first stage, but in the absence of affective stimuli. The PE stage was intended to evaluate how long stimuli effects lasted.

	Stage 1: Affective Induction							Stage 2: PE		
$C1^{+}$	R	O_1	PreS	O_2	X^{+}	O_n	PostS	O_{n+1}	MT	O_{n+m}
$C2^{-}$	R	O_1	PreS	O_2	X^{-}	O_n	PostS	O_{n+1}	MT	O_{n+m}
$C3^{ctrl}$	R	O_1	PreS	O_2		O_n	PostS	O_{n+1}	MT	O_{n+m}

Table 1. Experimental design summary. (R): Random placement, (PreS): Pre-Stimuli, (PostS): Post-Stimuli, (O_i): observations, (X): treatment/stimuli, and (MT): main task.

The affective induction stage, as depicted in Table 1, was designed as a pretest-posttest design (O X O), which was structured in three parts: (1) Pre-stimuli evaluation (PreS), in which the participants performed a search challenge before being exposed to affective stimuli; (2) stimuli exposure (X), in which the participants worked on a set of search challenges for 10 minutes as they received affective stimuli; and (3) post-stimuli evaluation (PostS), which aimed to evaluate the efficacy of affective stimuli in another search challenge. Before and after each search challenge, observations (O) were made through different instruments (details are provided below). Overall, the affective induction stage lasted 20 minutes, whereas the PE stage lasted 25 minutes.

Task Description

To study affective processes in a short span of time, we designed a precision-oriented search task based on multiple-step fact finding. The task was comprised of a set of search challenges (questions) that were presented sequentially to the participants. Search challenges were independent from each other and the participants were given a maximum of five minutes to find the responses. The goal for participants was to respond as many question as possible within the time given in each stage. Moreover, as they searched for answers, participants were instructed to save passages (snippets) from Web pages that contained clues to find the answers or the answers themselves.

The questions used in the study were obtained from A Google a Day[1]; these are each known for having a unique answer that can be found through different paths. The set of questions used in this study corresponds to those posted between April 7th and August 31st of 2011. Questions were collected along with answers and suggested search paths. The level of complexity of these questions was evaluated in terms of the number of steps or queries suggested to find the answer. Results from a pilot study showed that level-2 questions (two steps required) that only required textual information were most effective in terms of perceived difficulty, response precision, topic familiarity, and response time. Therefore, a random set of level-2 questions was used in the affective induction stage and in the evaluation of PE. Below we provide an example of a level-2 question, its search path, and its answer according to A Google a Day:

***Question**: How long is the river bordering the two countries that once were home to the Hamangia?*

***How to find the answer**: Search [Hamangia]. You will find that the Hamangia culture is a late Neolithic culture that once existed in what is now Romania and Bulgaria. Search [river bordering Romania and Bulgaria]. You will find the answer is the Danube River, which is about 1771 miles long.*

***Answer**: 1771 miles.*

Sample

The study was carried out with a sample of convenience that consisted of 45 undergraduate students (15 per group) from Rutgers University. Recruitment was conducted through open calls (e.g. announcements posted on campuses' bus stops, facilities, and email lists). Participants were compensated with $10 in cash for a one-hour session. As extra motivation, they were offered the possibility to win cash prices based on performance ranking at the end of the study (i.e. $50 first place, $25 second place, or $15 third place). Ages among the participants ranged between 18 and 27 years (M=20.44; SD=1.64). In this sample, 57.77% of the participants were women. Participants' search skills

[1] http://www.agoogleaday.com/

were reported as intermediate to high. All participants were native English speakers.

System

To support the study design and collect data during participants' sessions, a system based on Coagmento (González-Ibáñez & Shah, 2011) was implemented. This new system was designed to support experiment protocols, experimental designs involving affective treatments, timed tasks, multiple-stages, multiple sessions, and enhanced logging capabilities. As depicted in Figure 1, the system comprises three main components: (1) a toolbar, (2) a sidebar, and (3) a server-side Web application. The toolbar and the sidebar were implemented as a Firefox add-on.

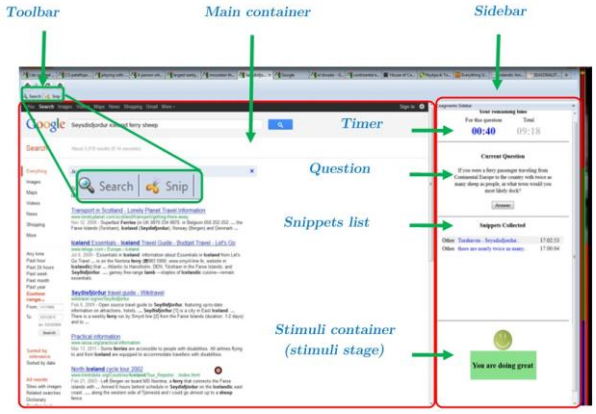

Figure 1. Snapshot of the system during an actual session.

The toolbar consists of two buttons and logging capabilities. First, the search button provided access to Google's home page (the only search engine enabled during the experiment). Note that this button also set Google parameters to restrict searches to pages indexed before April 7th of 2011. This ensured that participants did not have access to pages that could contain the questions used in this study and their answers posted by other people. Second, the snip button allowed the participants to save snippets of text from Web pages along with their sources.

In terms of logging capabilities, the toolbar was configured to capture browsing activity and different users' actions within the browser (e.g. clicks, queries, and pages), which were submitted to the server-side application through Web services. Note that data were logged along with local timestamps in order to facilitate synchronization with other data sources (e.g. keystrokes, webcam, EDA).

The sidebar contained different elements depending upon the stage of the task. First, on top of the sidebar, the remaining time for each question and the remaining time for the corresponding stages were displayed. Second, the current search challenge and a button to jump to the answer form were provided. Third, saved snippets for the corresponding search challenge were listed. Fourth, and only for the affective induction stage, stimuli were displayed on the bottom part of the sidebar.

Elicitation of Positive and Negative Affective States

The first stage of the experimental design depicted in Table 1 aimed to set the initial affective states of the participants in $C1^{+}$ and $C2^{-}$ for the evaluation of prolonged effects (PE). Based on results from a pilot study (González-Ibáñez & Shah, 2012) we used game feedback (or false feedback) (Martin, 1990) to elicit affective states in the context of the search task used in the experimental evaluation. This technique consists of providing either positive (e.g. "You are doing great!") or negative (e.g. "Wrong. That was disappointing") feedback to participants regardless of their actual performance when working on a given task, in order to elicit the desired affective state. Side effects of this approach include frustration, disinterest in performing the task, and overconfidence. To counteract such effects, balanced feedback was provided to the participants in each condition. More specifically, the participants in $C1^{+}$ received a semi-negative stimulus (e.g. "So so. You can do it better next time") for every three positive stimuli. Conversely, those in $C2^{-}$ received one semi-positive ("Not bad this time") stimulus for every three negative stimuli. Semi-negative and semi-positive stimuli were used instead of completely negative and positive feedback in order to avoid nulling effects on the intended affective state of each condition. Note that participants were told before starting the session that they would receive explicit feedback about their performance during the first part of the session.

As shown in Figure 2, stimuli consisted of a color box containing (1) a predefined text message and (2) a blinking emoticon (smiley, frowning, or neutral face) on top of the box. First, text messages were composed of words from the positive and negative categories listed in the LIWC (Pennebaker et al., 2001). Second, the blinking effect was implemented to grab the attention of participants who were working on the task. Third, boxes and emoticons were presented in three different colors (i.e., green for positive stimuli, red for negative stimuli, and yellow for semi-negative and semi-positive stimuli). Color boxes with corresponding messages and emoticons were presented to participants during the stimuli stage on the sidebar panel of the experimental system (Figure 1) for 15 seconds in intervals of 30 seconds. Finally, stimuli were also presented at the moment of submitting the answers to each question.

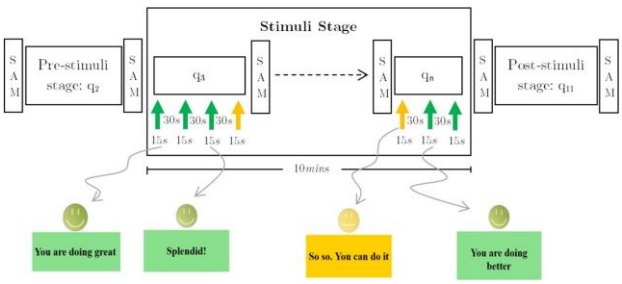

Figure 2. A closer look at the stimuli stage of positive condition ($C1^{+}$).

Note that the efficacy of stimuli relied on the assumption that they were able to elicit internal affective changes (i.e.

mood or affect) regardless of the participants' subjective experiences (i.e. feelings). That is to say, if a given participant received negative stimuli and they reported feeling positive affective states through self-reports, it was still assumed that their internal affective states moved toward a negative valence.

Laboratory Setup

The study was conducted in an interaction laboratory with one room for the participants and one for the study supervisor. Rooms were isolated, however, the supervisor monitored the study remotely and assisted participants if necessary. Room conditions (i.e. temperature and humidity) were monitored and regulated. The participants worked on a desktop computer equipped with a 19" display, full size keyboard, mouse, and headset.

Session Workflow

Each session was conducted following a research protocol devised to ensure consistency in all sessions and provide better documentation for possible replications of the study. As part of this protocol, first, participants were briefly introduced to the study. Second, participants signed consent forms (note that these two initial stages were part of the resting period prior to attaching and calibrating sensors to participants). Third instruments were calibrated. Fourth, participants performed a warm-up task (several repetitions of stand up and sit down for 30 to 60 seconds) before starting actual measurements, which allowed to establish a baseline for the particular case of electrodermal activity (EDA). Fifth, participants filled out a demographic questionnaire. Sixth, participants watched a brief tutorial that explains how to use the system. Seventh, participants familiarized themselves with the system while working on a practice search challenge. Eighth, affective elicitation stage was carried out. Ninth, evaluation of prolonged effect (PE) was performed. Finally, end-session questionnaires were distributed, followed by a brief interview. Overall sessions lasted approximately 60 minutes. This session workflow was automatically guided by the system described above.

Instruments

Data collection was performed using devices, instruments, and software. Resources are organized into three groups: hardware, questionnaires, and software. Following, brief descriptions of the instruments are presented.

Hardware

First, to record EDA, an *Affectiva Q Sensors 2.0 Curve* was used. Participants wore this sensor on their non-dominant hand (or the one that was not controlling the mouse) during the entire session. The sensor was set to capture data at 32 Hz and it was synchronized with the local time of the computer used by the participants in order to facilitate later event-related analyses with other data sources.

Second, to capture participants' FE, a Logitech C910 webcam was placed on top of the computer screen to capture participants' frontal faces. High definition video was captured with this camera at 15fps with a resolution of 920x720.

Questionnaire

To further investigate the participants' experiences, questionnaires and a semi-structured interview were used.

Measures of the participants' perceived affective dimensions were performed with the SAM questionnaire (Bradley & Lang, 1994). Using SAM, the participants reported how they felt immediately before and after working in each search challenge.

Two other questionnaires were designed to evaluate the users' experience with regard to topic familiarity, topic complexity, and level of confidence before and after each search challenge. Answers to these questionnaires were provided on a 5-point Likert scale.

In addition, cognitive workload was measured at the end of sessions using a simplified version of the NASA TLX (Task Load Index), also known as Raw TLX or RTLX, which omits the weighting subscales (Hart, 2006). Finally, a brief semi-structured interview was conducted at the end of each session.

Software

The final group of data collection resources consists of software that helped the researcher to keep track of browsing activity, record interviews, and perform observations. The list of software include the laboratory system introduced in the previous section, *Morae Recorder*, *Morae Observer*, *NCH Debut*, and *Affectiva Q Live*.

RESULTS

This section provides a description of the quantitative analyses carried out to address the following two research questions: (RQ1) are there affective patterns derived from initial affective states that can be used as implicit indicators of IPS and information relevance? And (RQ2) to what extent, if any, could affective-related features be used for automatic classification of information relevance? First, we focus on RQ1 through the evaluation of prolonged effects (PE). Second, we present results for RQ2 based on experiments using machine learning techniques over a set of affective-related features for the classification of information relevance.

RQ1: Affective Patterns

As explained in the previous section, participants in $C1^{+}$ and $C2^{-}$ were exposed to affective stimuli aiming to elicit positive and negative affective states, respectively, prior to starting the main search task. Based on self-reports in the context of the pretest-posttest design, the affective stimuli applied were 71.67% effective in the elicitation of positive affective states in $C1^{+}$ and 65% effective in the elicitation of negative affective states in $C2^{-}$. Note that these results are only referential to illustrate the participants' subjective

experiences after the affective induction stage. Nevertheless, as noted above, it was assumed that affective stimuli were able to elicit the desired internal affective states (i.e. mood and affect) regardless of the participants' awareness of such variations or that they were not able to express them accurately through the SAM. Measuring actual internal affective changes as a result of affective stimuli was beyond the scope of this work.

Moreover, as part of the pretest-posttest design, the Wilcoxon signed-rank tests showed that particular search behaviors presented significant variations in either $C1^{+}$ or $C2^{-}$ (but not in both) after stimuli exposure (e.g. higher relevant coverage in $C1^{+}$ ($W=204$, $p<.01$, $r=-.17$), lower recall in $C2^{-}$ ($W=932$, $p<.05$, $r=-.14$), and higher average time to collect snippets from relevant pages in $C2^{-}$ ($W=256$, $p<.01$, $r=-0.24$), to name a few), which illustrates that affective stimuli had different effects on both groups of participants. More importantly, no significant variations were reported on the control group ($C3^{ctrl}$).

Following, we present results of the evaluation of PE with special emphasis on the results for (1) FE analyses, (2) EDA, (3) self-reported experiences, and (4) performance.

Facial Expressions

To perform facial expression (FE) analyses, three independent software were used: *eMotion* (Sebe et al., 2004), *FaceDetect* (Küblbeck, 2006), and *BMERS* (González-Ibáñez, 2006). Unlike previous studies (Arapakis et al., 2008; Lopatovska, 2011, Moshfeghi & Jose, 2013), we designed an approach to validate the classification of FE by comparing the results provided by the above-mentioned software. Our analyses focused on a set of overlapping FE associated with the following basic emotions: surprise, sadness, anger, and happiness. We used the three tools to process the FE of the 45 participants. Results showed consistency in the detection of smiles (expression of happiness) in two out of the three tools. Results for the other FE were inconsistent. Smiles accounted for 9.78% of the FE in $C1^{+}$, 11.52% in $C2^{-}$, and 12.75% in $C3^{ctrl}$. We also verified the detection of smiles and other FE with manual inspections by picking up random samples from where the detections were made. We found that smiles were detected even when there were subtle signs of their presence. Conversely, for other FE, there were several false positives. Similar to past studies that used *eMotion* to process FE (Arapakis et al., 2008; Lopatovska, 2011), surprise accounted for the majority of the FE detected in the three experimental conditions (up to 45%). Similar results were found with FaceDetect, however, the most common FE detected were those associated with anger (up to 80%). Such results would suggest that the participants were surprised or angry most of the time; however, based on self-reports, observations, and interviews, this was not the case. Although Sebe et al., (2004), Küblbeck (2006), and González-Ibáñez (2006) report high accuracy rates in the overlapping set of FE, such results are derived from tests performed on specific face databases in which FE for each basic emotion are clearly distinguishable.

Based on these results we constrained the rest of our FE analyses to smiles. To map FE on the participants' search processes, we normalized the time that the participants spent in all search challenges (the A Google a Day questions) from 0 to 1 and produced ten segments. Then, we aggregated the number of smiles along the ten segments and normalized the corresponding values by expressing them in percentages. As shown in Figure 3, higher concentrations of smiles were typically found in the later stages of the participants' search processes. According to this analysis, only $C1^{+}$ presented a higher concentration of smiles (10.98%) in an early stage (segment 0.3). The fact that smiles were typically concentrated in later stages of the search processes associated with search challenges can be attributed to feelings of satisfaction, relief, and success experienced by the participants after answering each question. Feelings such as satisfaction or success can be related to the levels of confidence that participants reported at the moment of providing their answers.

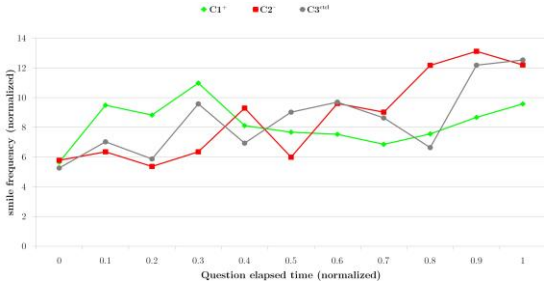

Figure 3. Normalized number of smiles expressed in percentages at different stages during the development of the questions (search challenges).

A closer examination was carried out in order to study the expression of smiles in relevant and non-relevant webpages. To define "relevance," we focused on users' perceived relevance, which in this case corresponded to the set of webpages from which snippets were saved. Analyses in this respect focused on comparing smiles and other FE (grouped into one single category) in the two types of webpages. For the case of relevant pages, results in all three conditions showed that the majority of smiles on such pages lasted on average less than 1.36 seconds; that is to say, less than 3% of the average dwell time spent on relevant pages ($M=45.22s$, $SD=48.03$). The density of smile duration on non-relevant pages was similar to that found on relevant pages; however, since the average dwell time on non-relevant pages was 17.50 seconds ($SD=18.73$), smiles on such pages typically lasted a fraction of a second. On the other hand, the duration of other FE account for up to 30% of the time participants spent on both relevant and non-relevant pages.

While smiles' duration in relevant and non-relevant pages were found to be proportionally similar with respect to dwell time on such webpages, the actual difference

(estimated in one extra second on relevant pages) could serve as an indicator of searchers' satisfaction.

We also investigated when the participants smiled during the exposure to the content of relevant and non-relevant webpages. This examination was carried out by aggregating smiles with respect to the elapsed time (normalized) of webpages.

Results for this analysis are depicted in Figure 4. As shown in this figure, smiles in $C2^{-}$ were mostly concentrated in later segments of the exposure to relevant pages. The absence of the first peak in this group, which was found to be a common aspect in the other two conditions, could be attributed to an influence of negative affective states in the information processing strategies used by the participants in $C2^{-}$ (e.g. affect infusion (Forgas, 1991; Forgas & George, 2001)). Negative affective states induced in the previous stage could have influenced the evaluative criteria and perception of the participants, thus making the participants in this group more reluctant to consider a page potentially relevant in the first seconds of exposure to the content of such pages. High concentrations of smiles in the last seconds of exposure to relevant pages could be related to the adoption of information processing strategies that made the participants in $C2^{-}$ more systematic, critical, and meticulous in the evaluation of the content of pages. Therefore, they only expressed smiles after carefully reviewing the content of webpages and finding the information they were looking for.

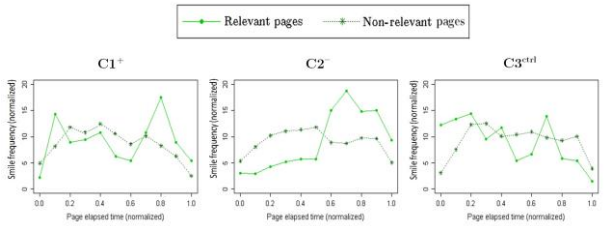

Figure 4. Normalized number of smiles expressed in percentages during the exposure to the content of relevant and non-relevant webpages.

On the other hand, concentration of smiles in initial segments, which is the case of $C1^{+}$ and $C3^{ctrl}$, could indicate a positive first impression of pages from the participants' point of view. For instance, at the moment of a page visit, particular features (e.g. title, images, and highlights, among others) may have indicated that searchers were in the right place, thus resulting in early smiles. Smiles detected in later segments could be related to the moment in which the participants found information that helped them to find the answers, thus denoting satisfaction or success.

Another analysis was carried out to explore whether or not the participants smiled when saving snippets from a page, thus marking a page relevant. Results for this analysis showed that the participants smiled more than 80% of the times at the exact moment in which snippets were saved. This percentage increased slightly when the analyses were conducted with a window of 10 seconds around the snip action (five seconds before and after).

Electrodermal Activity

The *Affectiva Q Analytics* was used to extract peaks and related measures from the EDA signals of all the participants. EDA analyses presented here focus on the occurrence of peaks as indicators of participants' responses to particular events, such as finding relevant pages or finding the answers to the questions. These particular events correspond to stimuli that are intrinsic to the development of the search challenges. Following a procedure similar to that performed with FE, the factor of interest (in this case EDA peaks) was aggregated in each segment of time of the normalized duration of the search challenges in which they were detected. Figure 5 depicts the cumulative frequencies of EDA peaks (expressed in percentages) in each unit of time during the development of search challenges. As seen in this figure, the three conditions displayed progressive positive trends of aggregated peaks toward the last minutes of the development of the A Google a Day questions. In this figure, it is possible to observe that $C2^{-}$ presents an abrupt change with salient points in the last two segments, which account for more than 40% of the total number of EDA peaks detected in this group. An explanation for this prominent concentration of peaks could be related to the participants' feelings of satisfaction or distress as a result of finding or not finding the necessary information to complete the search challenges.

The progression of cumulative peaks toward the end of the search process can be interpreted as an indicator of engagement and success (if accompanied by smiles) or frustration in the development of information-search tasks.

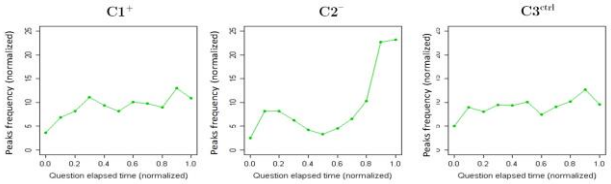

Figure 5. Aggregated number of EDA peaks (normalized) detected at different stages of search challenges.

Further analyses of EDA peaks in the context of relevant and non-relevant pages did not show meaningful patterns. Moreover, when analyzing whether the snip action was preceded (five seconds before), accompanied (at the exact moment), or followed (five seconds after) by peaks in EDA, our results showed that peaks in EDA occurred very few times (less than 2%) around the snip action. This finding indicates that peaks in EDA were not necessarily attributed to local findings of relevant content, which were explicitly indicated through the snip action.

Self-reports

Self-reported feelings were studied with regard to valence (i.e. positive – negative scale), activation (i.e. excited – calm scale), and dominance (i.e. controlled – in control scale) reported through SAM. Between-group comparisons were conducted with the Kruskal-Wallis test and post-hoc analyses were carried out with the Wilcoxon rank-sum test.

Results showed that valence was significantly different only at the moment of starting the evaluation of prolonged effects (PE). In particular, it was found that valence was significantly higher in $C1^{+}$ (Mdn=6) than in the other two conditions ($C1^{+}$ vs $C2^{-}$: W=877.5, $p<0.1$ | $C1^{+}$ vs $C3^{ctrl}$: W=900, $p<0.1$), which means that the participants in this condition felt more positive (happy as expressed in SAM). Valence in $C3^{ctrl}$ was found to be significantly higher (Mdn=5) than in $C2^{-}$ (Mdn=4) (W=877.5, $p<.01$). This result places the participants in $C3^{ctrl}$ in-between those in $C1^{+}$ and $C2^{-}$, thus confirming the efficacy of the stimuli applied during the affective induction stage.

Between-group comparisons with respect to activation before the first question showed that this was significantly higher in $C2^{-}$ (Mdn=5) than in $C1^{+}$ (Mdn=5) (W=832.5, $p<.01$). In turn, activation in $C1^{+}$ was significantly higher than in $C3^{ctrl}$ (Mdn=4) (W=802.5, $p<.01$). In other words, $C2^{-} > C1^{+} > C3^{ctrl}$.

The fact that significant differences in these three dimensions did not persist during the evaluation of PE can be attributed to a convergence and stabilization of the participants' affective experiences. Three possible explanations for this change are: (1) the absence of stimuli, (2) participants got used to the task, or (3) participants got used to SAM or became tired of having to answer it after each question. In any case, these could explain that, at some point, the participants' answers were placed in the center of the scales of valence, activation, and dominance.

The overall frustration level of the participants was measured at the end of the evaluation of PE. This measure was obtained from the responses to the sixth question of the RTLX questionnaire (i.e. "How insecure, discouraged, irritated, stressed, and annoyed were you?"). Results showed that frustration levels were significantly higher in $C2^{-}$ (Mdn=11) than in $C3^{ctrl}$ (Mdn=11) (W=855, $p<.01$). In turn, levels of frustration in the latter group were found to be significantly higher than in $C1^{+}$ (Mdn=9) (W=847.5, $p<.01$). In other words, $C2^{-} > C3^{ctrl} > C1^{+}$. These findings also support the overall effectiveness of the affective stimuli from the perspective of participants. That is to say, participants who received positive stimuli felt less frustrated than those who were exposed to negative stimuli.

With respect to the other dimensions measured through RTLX (i.e. mental effort, physical effort, time pressure, performance, and task difficulty) and the total score, no significant differences were reported by the Kruskal-Wallis test.

Finally, in terms of affective load (AL), as operationalized by (Nahl, 2009), this was found to be significantly higher in $C2^{-}$ (Mdn=165) than in $C3^{ctrl}$ (Mdn=105) (W=900, $p<.01$). In turn, AL in $C3^{ctrl}$ was found to be significantly higher than in $C1^{+}$ (Mdn=80) (W=900, $p<.01$). In other words, $C2^{-} > C3^{ctrl} > C1^{+}$.

Performance

We evaluated performance in terms of response precision. This measure was defined as the ratio between the number of correct answers over the total number of search challenges addressed. Between-group comparisons using both measures were conducted with the Kruskal-Wallis test, and post-hoc analyses were carried out with the Wilcoxon rank-sum test.

Results showed that response precision in $C1^{+}$ (Mdn=0.6) was significantly lower than in $C2^{-}$ (Mdn=0.63) (W=7.5, $p<.01$). In turn, response precisions in these two groups were found to be significantly lower than in $C3^{ctrl}$ (Mdn=0.67) ($C1^{+}$ vs $C3^{ctrl}$: W=15, $p<0.1$ | $C2^{-}$ vs $C3^{ctrl}$: W=7.5, $p<0.1$). Thus, $C1^{+} < C2^{-} < C3^{ctrl}$.

Differences in response precision could be attributed in part to the different levels of participants' rigor when addressing the search challenges. As explained through facial expression (FE) analyses, it is possible that the information processing strategies used by the participants in $C1^{+}$ were less systematic than those used in $C2^{-}$, thus affecting the quality of their answers to the search challenges.

RQ2: Classification Experiments

In this section we present results from classification experiments using machine learning techniques over a set of affective-related features to classify relevant and non-relevant pages. Specifically, we used normalized smile frequencies and average height of EDA peaks to create vectors of features associated with relevant and non-relevant pages as perceived by the participants. Then, we used these vectors to train and test univariate and bivariate classification models using Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR) with 10-fold cross validation.

	Feature	Accuracy	Precision	Recall	F-Measure
NB	Smiles	**51.88%**	0.59	0.51	0.39
	EDA	50.31%	0.50	0.50	0.44
LR	Smiles	**53.14%**	0.55	0.53	0.48
	EDA	52.20%	0.55	0.52	0.44
SVM	Smiles	**52.83%**	0.56	0.52	0.45
	EDA	51.57%	0.52	0.51	0.46

Table 2. Classification results.

Table 2 presents a summary of the best classification models built with these features. Note that all these models are univariate. Bivariate models as a result of combination of individual features did not outperform univariate models.

Despite affective patterns presented in the previous section, relevance classification based on smiles (which were highly

prominent around the snip action) and EDA peaks indicate that these individual features have limited practical usage – at least when they are not combined with other features (e.g. Arapakis et al. (2008)) or when used in this type of search task – to applications such as relevance feedback.

DISCUSSION AND CONCLUSION

In this article, we studied the role of initial affective states in information search. In particular, we investigated their influence, if any, in (1) information processing strategies (IPS), (2) relevance judgments, and (3) progress in the completion of search tasks. To address these problems, we focused on the following research questions: (RQ1) are there affective patterns derived from initial affective states that can be used as implicit indicators of IPS and information relevance? and (RQ2) To what extent, if any, could affective-related features be used for automatic classification of information relevance? To address RQ1 we designed an experiment to contrast the influence of positive and negative initial affective states. Then we performed analyses to study facial expressions (FE), electrodermal activity (EDA), self-reported feelings, and performance. On the other hand, to address RQ2 we used affective-related features derived from our analyses for RQ1 to train and test classification models for information relevance.

Unlike previous work on information seeking that focused on affective changes as a result of search processes (Poddar & Ruthven, 2010), this study focused on affective states as inputs to the search process in order to investigate how they lead to changes, patterns, or related phenomena during and after the search process.

Results suggest that initial affective states effect the way that information is processed, as well as the affective reactions during the exposure to information. Our analyses revealed that initial affective states would play a central role in the definition or selection of information processing strategies. Specifically, initial positive affective states would lead searchers to employ less systematic and less detailed information processing strategies than those used by searchers who commenced their search processes in negative affective states. Although this particular finding seems to be counterintuitive with what literature on positive psychology suggests regarding positive and negative affective processes, it is consistent with some psychology studies on information processing (Isen et al., 1983; Forgas, 1991; Sinclair & Mark, 1995; Forgas & George, 2001).

Results derived from FE analyses suggest that smiles could serve as a local indicator of information relevance. Yet, when and how searchers smile could be determined by their initial affective states, which may shape the way information is processed when performing relevance judgments.

Likewise, at the physiological level, analyses suggest that EDA could serve as a global indicator of IPS. Specifically, EDA peaks may signal levels of engagement at different stages of the search process, and also may signal success (if accompanied by smiles) or frustration in the development of information search tasks.

Despite these findings, results from our classification experiments, which applied machine learning techniques to smiles and EDA features, indicate limited practical usage – at least when used in isolation from other modalities of data (e.g. behavioral, attentional) – to applications such as relevance feedback.

In the long run, initial affective states seem to have implications in the quality of the work performed. In the study reported here, this aspect was expressed through a performance measure (response precision). According to our results, initial negative affective states result in higher-quality work than initial positive affective states. These particular results relate to the use of systematic and non-systematic IPS, respectively.

While the results of this study suggest that initial affective states would influence the way people search, process, evaluate, and use information, it is necessary to consider underlying limitations, such as sample demographics, stimuli type, search task, and assumptions. For instance, in terms of search task, findings from this study could only apply to precision-oriented search tasks, but not necessarily to recall-oriented ones such as exploratory search. Likewise, the affective stimuli used in the study only focus on inducing positive and negative affective states under a dimensional approach. However, particular emotions from a discrete approach (e.g. happiness, sadness, anger) may have different effects. Although the limitations of this study provide high internal validity, they limit the generalization (external validity) of the results.

At the theoretical level, the major implication of the results presented in this article is that initial affective states may be determining factors for the way search processes are carried out. This elaborates upon what traditional models in information science have suggested in the past: that affective processes are typically investigated as intrinsic factors to the search process.

The results and theoretical implications of this study could have different practical implications. By way of example, the results of this work can be related to research issues such as relevance feedback and the design of systems to aid search processes. There have been considerable efforts to bridge the gap between searchers and IR systems; however, such attempts have failed in incorporating the affective dimension. According to the results presented in this article, affective processes could help to gain access to hidden cognitive aspects (e.g. information processing strategies). For example, a determination of what information processing strategies are likely to be used based on initial affective states would allow systems to properly deliver and present personalized information to their users.

Although identifying affective states is as difficult as determining information processing strategies, to date different disciplines have contributed to the development of methods and techniques that enable technology to capture, process, and interpret affective signals (e.g. Picard, 1997). Different technologies with such capabilities have been developed in the last decade. Moreover, they are becoming increasingly popular and within the reach of regular people. Thus, it is not difficult to imagine that within a few years, technology will progress in such a way that it will be possible to accurately determine the affective states of people at any time. More importantly, this information will be available to IR systems as an implicit signal, serving as relevance feedback and helping the user meet their information needs more effectively.

ACKNOWLEDGMENTS

The work described in this article was partially supported by Proyecto DICYT, Código 061519GI, Vicerrectoría de Investigación, Desarrollo e Innovación, Universidad de Santiago de Chile; and the Institute of Museum and Library Services (IMLS) Early Career Development grant RE-04-12-0105-12.

REFERENCES

Arapakis, I., Jose, J. M., & Gray, P. D. (2008). Affective feedback: An investigation into the role of emotions in the information seeking process. *Proc. 31st ACM SIGIR* (pp. 395–402). New York, NY, USA.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy & Experimental Psychiatry*, 25 (1), (pp. 49-59).

Carmel, D., Roitman, H., & Yom-Tov, E. (2010, December). Social bookmark weighting for search and recommendation. *The VLDB Journal*, 19 (6), (pp. 761–775).

Coan, J. A., & Allen, J. J. B. (2007). Handbook of emotion elicitation and assessment. Oxford university press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). Emotion in the Human Face. Oxford University Press. Paperback.

Feild, H. A., Allan, J., & Jones, R. (2010). Predicting searcher frustration. *Proc. of the 33rd ACM SIGIR* (pp. 34–41). New York, NY, USA: ACM.

Forgas, J. P. (1991). Affective influences on partner choice: role of mood in social decisions. *Journal of personality and social psychology*, 61, (pp. 708-720).

Forgas, J. P., & George, J. M. (2001). Affective influences on judgments and behavior in organizations: An information processing perspective. *Organizational Behavior and Human Decision Processes*, 86 (1), (pp. 3 – 34).

Fox, S., Karnawat, K., Mydland, M., Dumais, S., & White, T. (2005). Evaluating implicit measures to improve web search. *ACM Trans. Inf. Syst.*, 23 (2), (pp. 147–168).

González-Ibáñez, R. (2006). Evaluación de la integración del darse-cuenta emocional en una aplicación colaborativa (Unpublished master's thesis). Universidad de Santiago de Chile.

González-Ibáñez, R., Shah, C., & Córdova-Rubio, N. (2011). Smile! studying expressivity of happiness as a synergic factor in collaborative information seeking. *Proc. of ASIST'11*, 48 (1), (pp. 1–10). New Orleans, LA, USA.

González-Ibáñez, R., & Shah, C. (2011). Coagmento: A system for supporting collaborative information seeking. *Proc. of ASIST'11*, 48 (1), (pp. 1–4). New Orleans, LA, USA.

González-Ibáñez, R., & Shah, C. (2012). Investigating positive and negative affects in collaborative information seeking: A pilot study report. *Proc. of ASIST'12*, 49(1), (pp. 1-4).

Gonzalez-Ibañez, R. I., & Shah, C. (2015). Affective Signals as Implicit Indicators of Information Relevancy and Information Processing Strategies. *Proc. of iConference*, Newport Beach, CA.

Gwizdka, J., & Cole, M. (2013). Does interactive search results overview help?: An eye tracking study. *Proc. of CHI EA'13* (pp. 1869–1874). New York, NY, USA: ACM.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proc. of the human factors and ergonomics society annual meeting*, 50(9), (pp. 904–908).

Hassan, A., Jones, R., & Klinkner, K. L. (2010). Beyond DCG: User behavior as a predictor of a successful search. *Proc. of WSDM'10* (pp. 221-230). New York, NY, USA: ACM.

Isen, A., Means, B., Patrick, R., & Nowicki, G. (1983). The influence of positive affect on decision making strategy. *Social Cognition*, 2(1), (pp. 18–31).

Izard, E. (1977). Human emotions. Plenum Press.

Küblbeck, C., & Ernst, A. (2006). Face detection and tracking in video sequences using the modified census transformation. *Image Vision Computing*, 24 (6), (pp. 564–572).

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *JASIST*, 42 (5), (pp. 361–371).

Liu, J., & Belkin, N. J. (2010). Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. *Proceedings of the 33rd ACM SIGIR* (pp. 26–33). New York, NY, USA: ACM.

Lopatovska, I. (2009). Does the mood matter? In Affective computing and intelligent interaction and workshops, 2009. Proc. of *ACII'2009*. (pp. 1-4).

Lopatovska, I. (2011). Emotional correlates of information retrieval behaviors. *Proc. of WACI 2011*, IEEE Workshop, (pp. 1-7).

Lv, Y., & Zhai, C. (2010). Positional relevance model for pseudo-relevance feedback. *Proc. of the 33rd ACM SIGIR* (pp. 579–586). New York, NY, USA: ACM.

Martin, M. (1990). On the induction of mood. *Clinical Psychology Review*, 10 (6), (pp. 669 – 697).

Moshfeghi, Y., & Jose, J. M. (2013). An effective implicit relevance feedback technique using affective, physiological and behavioural features. *Proc. of the 36th ACM SIGIR* (pp. 133–142). New York, NY, USA: ACM.

Nahl, D. (2009). The centrality of affective in information behavior. In D. N. D. Bilal (Ed.), (pp. 3-37). Information Today, Inc.

Palmero, F., Guerrero, C., Gómez, C., & Carpi, A. (2006). Certezas y controversia en el estudio de la emoción. *Revista electrónica de motivación y emoción (R.E.M.E)*, 9(23-24), 1.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count. Mahwah, NJ: Lawerence Erlbaum Associates.

Picard, R. W. (1997). Affective computing. Cambridge, MA, USA: MIT Press.

Poddar, A., & Ruthven, I. (2010). The emotional impact of search tasks. *Proc. of the 3rd IIix'2010*.

Sebe, N., Lew, M., Cohen, I., Sun, Y., Gevers, T., & Huang, T. (2004). Authentic facial expression analysis. *Proc. of the 6th IEEE FG*, (pp. 517-522). Seoul, South Korea: IEEE.

Sinclair, R. C., & Mark, M. M. (1995). The effects of mood state on judgemental accuracy: Processing strategy as a mechanism. *Cognition & Emotion*, 9 (5), (pp. 417-438).

Strauss, M., Reynolds, C., Hughes, S., Park, K., Mcdarby, G., & Picard, R. W. (2005). The handwave bluetooth skin conductance sensor. *Proc. of ACII'2005*, (pp. 699-706). Springer.

Watson, D., Clark, L. A., & Tellegen, A. (1988, June). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54 (6), (pp. 1063–1070).

Xu, J., & Croft, W. B. (1996). Query expansion using local and global document analysis. *Proc. of the 19th ACM SIGIR*, (pp. 4–11). New York, NY, USA: ACM.