Cortical Activity of Relevance

Zuzana Pinkosova

Computer and Information Sciences

University of Strathclyde

Thesis submitted for the degree of *Doctor of Philosophy*

January 2023

Glasgow

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date: 20 January 2023

Abstract

Despite decades of research, relevance remains a central focus of Information Retrieval (IR) research. Many theoretical approaches in IR assume that relevance is based on the mutual interaction of the system and user [1, 2]. Past studies have mainly focused on the system side, while user-centred studies are more recent and arguably more challenging to conduct due to no universally accepted research methodology nor established relevance definition [3, 4]. Despite many competing theories, researchers in general agree that relevance is an internal and subjective process. Therefore, experimental approaches investigating relevance should consider the underlying physiological, psychological and behavioural mechanisms involved [5]. With the development of brain imaging, a new multidisciplinary research direction (termed NeuraSearch [6]) has begun to investigate user relevance by analysing brain activity. The combination of information science, cognitive science, psychology and neuroscience has provided a unique insight into relevance phenomena and established the foundation for brain imaging research methodology within the IR field.

Therefore, this thesis builds upon the successful NeuraSearch framework to gain a better understanding of relevance phenomena from a neuro-cognitive point of view, to test existing relevance theories, and to gain in-depth insight into mental processes that underpin relevance evaluation. To do so, we conducted a user study, during which participants provided relevance assessments in the context of the Question-Answering (Q/A) Task, during assessment with an electroencephalogram (EEG). Collected neurophysiological data were analysed using a data-driven approach, which offers a comprehensive overview of all the neurocognitive elements that play an essential role during relevance assessment.

In this thesis, we investigated relevance as a binary (i.e. relevant vs. non relevant) and graded (e.g. highly relevant, low relevant, non relevant) variable. Additionally, we have explored the role of users' cognitive context (namely the self-perceived knowledge (SPK)) on relevance assessment formation. Using a data-driven approach within the NeuraSearch experimental framework, we present the following research contributions:

- By re-visiting binary relevance using a data-driven approach, we have not only confirmed the findings of previous studies but also shed light on previously not reported Event-Related Potential (ERP) component - P100. The data-driven approach has been proven effective in discovering novel ERP phenomenon, which have been shown to modulate early attention allocation [7] (see Chapter 4).

- Relevance is a complex and context-dependent. Thus, this research investigated the impact of users' SPK on binary relevance assessment. The results indicate that the SPK within the relevance context is associated with significant differences in cognitive processing related to attention, semantic integration and categorisation, memory and decision making (see Chapter 5).

- So far, brain imaging studies have mainly considered relevance as a binary variable. The research presented in this thesis is the first to investigate relevance granularity. We observed significant differences in ERPs in response to words processed in the context of high-relevance, low-relevance and no-relevance. It is possible that differences in attentional engagement, semantic mismatch (between the question and answer) and memory processing may underpin the electrophysiological responses to the relevance assessment. The results support the concept of graded relevance and knowledge of the electrophysiological modulation to each type of stimulus may help to improve the search system design (see Chapter 6).

Overall, presented findings may help to better understand the cognitive levels of individuals and recommend content based on their cognitive abilities, which would lead to an increase in search success. A better understanding of relevance is an important step toward improving personalisation in the IR process.

Contents

List of Figures

List of Tables

Preface/Acknowledgements

Being able to apply and pursue this PhD would have not been possible without the input and support of many amazing people over many years. I would be forever grateful for all the kindness, love, friendship and for all those wonderful memories. While I cannot possibly thank everyone, there are a few people who deserve particular praise. My sincere gratitude goes to my two PhD supervisors, Dr Yashar Moshfeghi and Dr William McGeown, for their support, wisdom, encouragement and guidance. I would like to thank all the members of NeuraSearch laboratory, for their feedback, help and kind words. Dominika, thank you for motivating me, for being a fantastic friend, and for being a little ray of sunshine during my cloudy days. Amine, thank you for showing me the best food and coffee places around, for many walks and laughs that I have enjoyed so much. Francesco, grazie mille for all the memes, laughs, help, for always having my back and cheering me up every time I needed it the most. Words can not explain how fortunate I am to have met you!

Another big thank you goes to the "Humnans of CIS" group. I loved our chats and socials. Sylvain - merci beaucoup for all the food (the sausages!), all the brilliant memories (Metallica!), tons of laughter and for being such a great friend. Mateusz, I always admired your work and you were often my inspiration - thank you for your help, wise words and feedback. Olivia, I am so glad I have gained such an amazing friend like yourself. My PhD would have not been the same without your stories, memes and laughs. I am forever grateful that I was able to gain such wonderful friends from this experience. I will cherish our memories forever and I am looking forward to more.

Another big THANK YOU goes to all the participants who took part in the study, for their enthusiastic participation, and candour. I also want to mention wonderful

SIGIR, CHIIR and ACAIN communities - I am grateful for lovely memories, feedback and friendships.

Next, I want to thank Graham, Cassie, Ellie, Cat and Riley, Adam, Sofia, Iris, Nilavra, Maja, Stuart and Teo for their support and many years of friendship! It was great to know that I can always count on you and rely on you. Graham, Iris, and Sofia, you have helped me tremendously during this journey. Maja, thank you for all those years of friendship and wonderful memories not only during the PhD.

I dedicate this thesis to my family and my partner, who have been a source of strength throughout this whole process. I thank them for their love, support and constant belief in me. Thank you mum and dad for being here for me all the time and for being so supportive of my ambitions - mami, tati, veľmi vás ľúbim a ďakujem vám za všetko! Marian, my brother, thank you for being the one who pays for our family Netflix subscription :). Seriously - you are the bestest brother in the entire universe and without you, I would have never achieved this. Alex, thank you for always looking after us, for your love and support. Eva Mae and Lucas - I love you to the moon and back. Spending time with you always helped me to realise what is important in life. Little baby bump, my niece Olivia, I can not wait to meet you. Karl, thank you for your support and cosy evenings in when I needed them the most. Rastislav, thank you for your love, (loads of) patience, and continuous encouragement in completing this journey - I am already looking forward to our future destinations together.

I am very grateful for the Engineering and Physical Sciences Research Council (EPSRC) funding (without which my PhD would not be possible), Scottish Informatics and Computer Science Alliance (SICSA) Research Grant, Annual Conference of the Association for Computing Machinery Special Interest Group in Information Retrieval (ACM SIGIR) Travel Grant and ACM Student Scholarship. Another big thanks goes to the University of Strathclyde - to all the staff that made a difference and helped me to achieve my dream. My time at Strathclyde has been character-building, and a wonderful experience I will cherish forever.

List of Abbreviations and Symbols

ASK Anomalous State of Knowledge. 16

BCI Brain-Computer Interface. 45, 55, 62, 134, 138

CPP Centro-parietal Positivity. 106, 114, 115, 118, 120, 121, 126, 128–133

DV Dependent Variable. 71, 96, 104, 113

EEG Electroencephalograph. ii, 39–45, 54, 57, 58, 60–62, 64, 65, 69–72, 75–77, 81, 82, 84, 85, 90, 93–96, 104, 112, 113, 118, 134–136, 138

ERP Event-Related Potential. iii, ix–xi, 6, 8, 44, 45, 58, 61, 62, 64, 65, 78, 85, 94, 95, 98–101, 103, 105–109, 115, 117–121, 124–126, 128, 130, 132, 133, 135, 137

fMRI Functional Magnetic Resonance. 39, 40, 45, 54–57, 59

HCI Human Computer-Interaction. 19, 23

HIGHR High Relevance. x–xii, 71, 79, 81, 83, 113–119, 130

HII Human Information-Interaction. 19, 54, 55

Hz Hertz. 43, 44, 71, 75, 82, 84

ICA Independent Component Analysis. 82, 84, 96

IN Information Need. 3, 4, 6, 13–19, 26, 27, 31, 55–60, 62–64, 66, 78, 86, 95, 96, 124, 125, 130, 132–134

IR Information Retrieval. ii, iii, ix, 3, 4, 6–8, 11–16, 18–21, 23–31, 36, 37, 39, 44, 46, 54–56, 58, 59, 62–65, 67, 71, 72, 78, 94–96, 99, 103, 111, 112, 123–126, 129, 133–138

ISP Information Search Process. 20–22

IV Independent Variable. 70, 71, 96, 104, 113

know Knowledgeable. x, 71, 79, 80, 83, 104–109

LOWR Low Relevance. xi, xii, 71, 79, 80, 83, 113, 114, 117–121, 130

LPC Late Positive Component / Complex. 6, 44, 98, 99, 106, 108, 116, 118, 121, 126–133

MEG Magnetoencephalogram. 39, 40, 45, 54, 55, 60

ms Millisecond. ix–xi, 44, 60, 61, 78, 79, 82–85, 97–100, 105–109, 114–121, 126, 129, 131, 132

NONR No-Relevance. x, xi, 71, 79, 80, 83, 113–121, 130

notknow Not Knowledgeable. x, 71, 79, 80, 83, 104–109

nr Non-relevant. ix, x, xii, 71, 79, 83, 95–100, 103–105, 107, 125, 127

Q/A Question-Answering. ii, 60, 65, 67, 71, 73, 80, 95, 103, 104, 112, 113, 125

rel Relevant. ix, x, xii, 71, 79, 83, 95–100, 103–106, 108, 109, 125–127

ROI Region of Interest. 45, 85

SD Standard Deviation. xii, 69, 74, 81, 86–89, 96, 97, 103, 104, 113, 114

SPK Self-Perceived Knowledge. iii, xii, 5–8, 10, 64–66, 71, 80, 93, 102–106, 108, 110, 111, 124, 125, 127–130, 136, 137

List of Abbreviations and Symbols

Part I

Theoretical Contributions

Chapter 1

Introduction

This chapter introduces the core of the thesis topic, explores the general context of the conducted research, and then continues with a discussion of the motivations and research aims, leading to the thesis statement. Next, the chapter presents the main contributions, published work and closes with a summary of the thesis layout.

1.1 Motivation

Despite decades of research and scientific advances, relevance remains a central, timeless, and fundamental topic given its critical importance to the field of IR. The IR relevance research relies on methods that are able to distinguish relevant from irrelevant information [8, 9]. These methods are based on obtaining relevance assessments from users (i.e. user relevance) when they are examining specific information items retrieved by the system (i.e. system relevance) [10]. Traditionally, relevance has been mainly considered from a system side but more recent user-centred approaches suggest that relevance is not just a match between the query and information but rather a complex and important aspect of human–information interaction that occurs within a specific Information Need (IN) context [11]. Studying relevance is, therefore, challenging as the construct is considered to be personal, multidimensional and dynamic which depends on the specific problem at hand [1, 10, 12–16]. Relevance terminology is often inconsistent and instead of a universal definition, there are many competing theories [3, 17]. Furthermore, despite relevance being considered a measurable phenomenon, there are no universally accepted research methodologies and frameworks to assess user relevance [18–20]. For instance, the debates surrounding relevance granularity are still ongoing [21]. While the binary approach (i.e. information assessed as relevant or non-relevant) is prevalent, seminal theories have proposed relevance as a graded concept (i.e. information assessed to be relevant to a different degree) [22–24]. As a result, despite the vast amount of available literature, there are still significant gaps in our understanding of users' relevance perception. Thus, understanding factors affecting relevance decisions and the associated cognitive processes continues to be important [25].

Generally, research investigating users' relevance relies on explicit, implicit or the combination of both feedback approaches [26]. The explicit approach requires users to explicitly assess content relevance and, therefore, it can be cognitively demanding [27]. The implicit approach relies on unobtrusive relevance assessments collected using behavioural or/and physiological signals [5,26]. However, implicit feedback can be noisy and, as a result, less accurate [5].

The last decade of relevance research was associated with a significant advancement attributed to the neuro-cognitive empirical perspective, which enabled researchers for the first time to access complex mental phenomena that underpin information relevance evaluation. These studies employed a wide variety of brain imaging technologies while investigating relevance assessment within the context of different stimuli modalities. These neuro-cognitive studies can be categorised into two groups:

- Studies that considered relevance as a part of IR, therefore, taking into account users' INs (termed NeuraSearch [6]).
- Studies that considered relevance in terms of word associations, without taking account of users' INs [18].

It is possible to argue that relevance should be investigated as an integral part of IR. Therefore, users' INs should always be considered because they provide essential context determining the problem situation, such as one's awareness of available information, affective or emotional states, expectations, time constraints and information about one's gaps in knowledge [28].

Thus, the research presented in this thesis follows a NeuraSearch approach with the aim to gain an in-depth understanding of neurophysiological characteristics related to the cognitive processes that underlie the relevance assessment of textual information (i.e. the most common form of information consumption [29]). We revisit the concept of binary relevance as so far prior studies have mainly investigated this construct using a component-driven approach. However, the component-driven approach might not allow for the discovery of novel neurophysiological components and mental phenomena. On the other hand, we propose to investigate binary relevance by employing a data-driven approach, which has been proven effective in discovering novel neuro-cognitive components [30]. The data-driven approach might be especially useful in gaining an in-depth, holistic understanding of complex cognitive constructs which lacks a comprehensive theoretical and empirical overview, such as relevance [19, 30].

Furthermore, relevance assessment involves complex interactions among various factors including but not restricted to users' cognitive, affective and social aspects [10]. The work presented in this thesis aims to gain a better understanding of the relationship between users' cognitive states and relevance assessment. In particular, we investigate users' SPK states, which play a central role in the evaluation of information. Therefore, it is important to examine the relevance assessment process while considering various degrees of users' SPK, which would enable us to better understand how humans interact with information while considering their cognitive context.

Relevance is frequently investigated as a binary concept but there are many theories and behavioural studies proposing relevance as a graded concept [22–24]. However, so far it is not clear whether there are neurological signatures associated with the processing of the information relevant to a different degree in the brain. Therefore, in this thesis, we aim to address this gap in research by investigating neurophysiological signature differences associated with the processing of each relevance grade using a data-driven approach.

The technological progress in the field of Information Science and Retrieval (IS&R)

and the continuous formation and development of information society lead to the demand for efficient and accessible information search by the end users [31]. As a result, there is a continuous need for the improvement of communication between the system and the users [32]. To do so, it is important to adequately assist the users with their INs by understanding how users interact with information. Over the last decade, studies examining users' complex mental states and cognitive functioning have significantly benefited the field of IR by uncovering important relationships between users' internal processes and information evaluation using brain imaging. This thesis aims to continue in this direction by exploring the above-mentioned research gaps, which may lead to the improvement of existing IR systems.

1.2 Contribution to Knowledge

In general, this thesis explores the application of the data-driven approach used to potentially uncover previously not reported cognitive phenomena that underpin the formation of relevance assessment. The key theoretical, methodological and empirical contributions are summarised as follows:

- Binary Relevance
  - The findings of the first experimental chapter re-visiting textual binary relevance assessment confirm the result of the previous studies (e.g. [33]) as the data-driven approach revealed statistically significant differences associated with N400 and Late Positive Component (LPC) ERP components.
  - Furthermore, through the application of a data-driven approach we were able to uncover the P100 ERP component that has not been previously reported and explored within the context of binary relevance assessment using other data analysis methods.
- SPK
  - Given the importance of the users' internal cognitive context for relevance assessment, the thesis explored the effect of the users' SPK states on textual

binary relevance assessment. The results of the data-driven analysis suggest that user's SPK plays an important role in relevance assessment and it is associated with differences in cognitive processes related to attention, memory retrieval and learning.

  - We explained how the users' SPK modulates the formation of relevant and non-relevant assessments.
  - The findings contribute towards empirical validation of Ingwersen's Cognitive Theory of IR [2].

- Graded Relevance
  - The findings of the last experimental chapter provide neuroscientific support for graded relevance. Past studies examining users' graded relevance assessments have mainly relied on explicit self-assessments provided by the participants (e.g. [24]) and it was not clear whether the concept of graded relevance has a neural origin (i.e. whether the perception of different grades is associated with distinct cognitive processes).
  - Additionally, the chapter provides an in-depth overview of cognitive processes that contribute to graded relevance formation and explains the differences between each grade of relevance.
  - Lastly, neurological differences associated with the processing of each relevance grade provide empirical support for theories considering relevance as a continuous rather than binary variable.

1.3 Thesis Statement

Given the above, the main statement of this thesis is that "*by investigating textual relevance as a subjective notion while considering the users' cognitive states will strengthen its theoretical foundations and unravelling novel neurological phenomena involved using the neuro-cognitive approach will contribute toward more realistic modelling of IR*". Relevance is a complex human notion which should be investigated as a subjective perception of the relationship between certain information and the problem at hand, so within the appropriate IR context [10, 34]. Furthermore, the user's relevance processing always happens as a result of interaction between the information property and the user's mental state [10]. However, users' mental states within IR research only began to receive attention in recent years. Therefore, there are still many gaps in understanding of this complex relationship. The research presented in this thesis aims to unravel parts of this interaction by capturing and examining the user's subjective internal states manifested through the neurophysiological signals.

Thesis findings can aid researchers in improving the current state of IR systems through the consideration of unobtrusively collected signals accurately reflecting users' states during the information interaction process. Enabling automated information recommendations based on implicit data might significantly improve user satisfaction.

1.4 Research Objectives

The multi-level user study constructed within the NeuraSearch framework was designed to address the following objectives:

- **RQ1:** "Does a data-driven approach reveal additional previously not reported ERP components associated with binary relevance phenomena?";
- **RQ2:** "Are findings of the data-driven approach aligned with findings of previous studies examining neurological signatures of binary relevance assessment?";
- **RQ3:** "Are there clear and detectable neural manifestations associated with distinct users' SPK states during binary relevance assessment?";
- **RQ4:** "How do the neural mechanisms associated with different SPK states drive the cognitive processes underpinning the binary relevance assessment?";
- **RQ5:** "Are there clear, detectable, physical manifestations of graded relevance in human brains?";

- **RQ6:** "Do such manifestations differ when a user perceives different degrees of relevance (i.e., when searchers assess a document as highly relevant, low relevant or non-relevant)?";

- **RQ7:** "What is the nature of graded relevance from a cognitive neuroscience perspective?"

1.5 Publications Resulting from this Thesis

The research presented in this thesis and completed throughout the duration of the author's PhD programme has been submitted and published at the peer-reviewed venues listed below:

- Pinkosova, Z., McGeown, W. and Moshfeghi, Y., 2022, September. Revisiting neurological aspects of relevance: an EEG study. *In Advanced Online & Onsite Course & Symposium on Artificial Intelligence & Neuroscience.* Certosa di Pontignano, Italy.

- Pinkosova, Z., McGeown, W.J. and Moshfeghi, Y., 2020, July. The cortical activity of graded relevance. *In Proceedings of the 43rd international acm sigir conference on research and development in information retrieval* (pp. 299-308) Xi'an, China. https://doi.org/10.1145/3397271.3401106

- Pinkosova, Z. and Moshfeghi, Y., 2019, July. Cortical activity of relevance. *In CEUR Workshop Proceedings* (Vol. 2537, pp. 10-15). Milan, Italy. ISSN 1613-0073

Submitted for Review:

- Pinkosova, Z., McGeown, W. and Moshfeghi, Y., 2022, Moderating Effects of Self-perceived Knowledge in a Relevance Assessment Task: an EEG Study. *Computers in Human Behavior Reports*, Submitted.

1.6 Thesis Layout

This thesis is divided into two main parts which contain seven chapters in total.

- Part I: Theoretical Contributions
  - Chapter 1: Introduction. The first chapter briefly introduces the context of this thesis as well as research aims and contributions to the field.
  - Chapter 2: Background. Drawing on key theories, this chapter begins with a comprehensive overview of the terminology and relevant background information about recent IS&R and Neuroscience advances to explain the context of the research and consequently identifies research gaps which lead to the Research Motivation described in Section 2.7.
  - Chapter 3: Methodology. The chapter elicits the employed paradigm, data acquisition, experiment protocol and experimental set-up for the recording, pre-processing, feature extraction and data-driven analysis of the neurophysiological signal.
- Part II: Empirical Contributions
  - Chapter 4: The Cortical Activity of Binary Relevance. The first empirical chapter re-visits the neuro-cognitive aspects of textual relevance processing by using a data-driven analysis method not previously employed to investigate this complex construct.
  - Chapter 5: SPK in a Relevance Assessment Task. In this chapter, the user's cognitive context (namely the SPK) is considered within the context of binary relevance assessment using the data-driven method to obtain an in-depth understanding of the involved complex cognitive phenomena.
  - Chapter 6: The Cortical Activity of Graded Relevance. Given the ongoing debate surrounding relevance granularity, this thesis also considers relevance as a graded variable using a data-driven method to test whether there are any significant differences associated with the processing of information relevant to a different degree.

  - Chapter 7: Discussion and Conclusions. This final chapter concludes the thesis by highlighting achieved objectives, discussing innovations, and new insights along with the various research findings.

- The Appendix contains supplementary information for Chapter 3 and Part II of the thesis.

The overall purpose of the experimental work presented in this thesis is to better understand the neurophysiological aspects of relevance assessment within the context of Information Retrieval (IR).

Chapter 2

Background and Motivation

This chapter introduces the key terminology and extensive review of the literature required to define key concepts relevant to the contributions made in this thesis. Furthermore, the chapter provides an overview of the research approaches and application context within which this research is situated.

The chapter opens with an overview of IR (Section 2.1) and discusses influential theories in the field (Section 2.2). Next, Section 2.4 presents relevance as the central notion in IR and introduces theoretical and empirical work which significantly shaped the conceptualisation of the topic. Section 2.5 provides an introduction to brain imaging techniques. Section 2.6 explains how brain imaging was applied to study complex IR phenomena, namely relevance (Section 2.6.2), which is the main thesis focus. Finally, Section 2.7 at the end of this chapter gathers the gaps in the existing scientific literature that were explored in the Part II of this thesis.

2.1 Information Retrieval

The history of IR is inseparably connected with the early period of computer usage. IR, as a research discipline, emerged in the 1950s as a response to the growing volume of machine-stored information, collected and maintained by library systems. With the growing number of information, manual information search and maintenance was getting more complicated. As a result, early research directions aiming to automatise existing manual indexing and document searching strategies accelerated the development of IR. The term IR was first introduced and defined by Calvin Mooers in 1948 (published in 1952) as: "*the problem of directing a user to stored information, some of which may be unknown to him, is the problem of IR*" [35].

In the 1950s and 1960s, the field of IR focused on two main areas: how to index individual documents and how to retrieve them as easily and effectively as possible [36]. The fundamental hallmark in the IR research was the empirical evaluation of search system performance [37]. The evaluation of retrieval systems was based on the influential Cranfield tests, which have set up the prevalent methodology for IR assessing how well a system meets the needs of the user which involves three components: (i) a test collection consisting of a set of documents, (ii) a set of topics, and (iii) a set of relevance assessments involving human assessors [38]. A topic refers to a description of the information being sought. Relevance assessment (typically binary responses) specifies the documents that should be retrieved in response to the topic. The methodology set up by Cranfield tests is still one of the most used for evaluating retrieval systems [39].

In the 1970s, the first conferences covering topics related to IR began to take place, one of them being the SIGIR (International Conference on Information Storage and Retrieval) in 1971. The TREC (Text REtrieval) Conference, which contributed greatly to the development of the field of information research, was held for the first time in 1992. The IR conferences have influenced research in the field of evaluation of retrieval systems in particular. It is important to note that in the very beginnings of the IR research, all attention was mainly focused on information systems and the notion of a user was hardly discussed [40].

Recent developments in IR research were fundamentally influenced by cognitive, relevance, and interactive revolutions in the field [41]. The cognitive revolution has opened up the possibility of analysing INs and their subsequent development processes, which can change over time. It highlights the cognitive processes of the user associated with information search and aims to narrow the gap between how the IN is understood vs. how it is subsequently interpreted by the user. Furthermore, the revolution has emphasised that IN is the user's personal and individual perception of the information requirement [41]. In the relevance revolution, there is an increasing interest in the information requirement, which the user enters into the search system in the form of a query. Retrieved information should be assessed against the individual's IN situation, not the submitted query or search request. Therefore, information relevance should be assessed in relation to the IN or a problem at hand situation experienced by individual users [11,41]. The advent of the Web in the 1990s and increased use of retrieval systems highlighted the issues related to IR interaction between the user and the system [36]. In this period, the focus begins to shift toward interactive information retrieval and empirical work starts to focus more on users' search behaviour, mainly on query formulation and reformulation. This revolution highlighted the fact that retrieval systems cannot be evaluated without considering user's interaction [41] and modern research aims to accurately capture user search preferences.

The modern field of IR aims to study and understand information storage, access and search in order to design, build and test search systems [42] that would assist human information seekers to find information items containing answer sources to what the seekers are looking for. This is because, within the context of IR, users often interact with the system in order to resolve a problematic situation, caused by an information gap. An outcome of such interaction is the desired change in the information state. The such outcome often occurs by virtue of engagement with the information material (usually textual documents) in the context of the system, through searching within a document collection for particular information that resolves the users' need(s) or helps them to achieve the search task goal [43]. Relevance is a central notion in this process, indicating system effectiveness and retrieval performance [10, 34]. Therefore, the main aim is to retrieve the most relevant documents to the query from the collection of documents [44]. A perfect IR system should ideally retrieve only relevant documents. A schematic representation of IR interaction process is depicted in Figure 2.1.

2.1.1 IN

IN remains one of the most essential concepts in information science [45] and it is undoubtedly related to the information search. However, the concept is challenging to

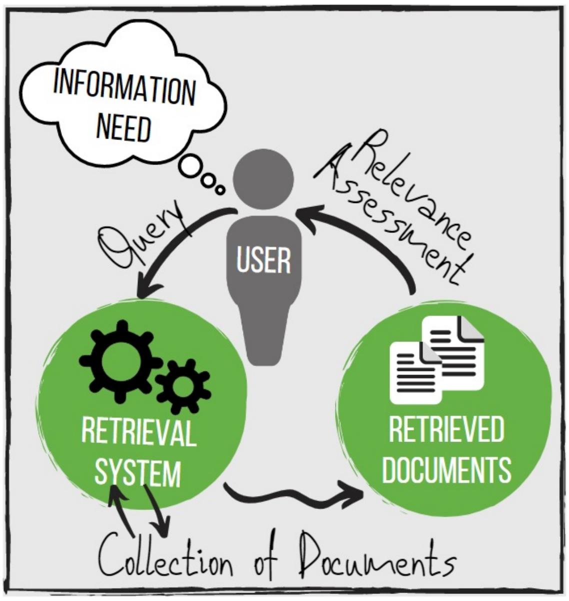

Figure 2.1: Schematic Diagram of IR.

Users experiencing IN formulate a request in the form of a query. The IR system then responds by retrieving documents from a collection of documents about the required information. The user then evaluates retrieved documents with relation to their IN satisfaction.

clarify and up to this date, there is no universally accepted definition [46]. IN can be expressed as a state in which individuals find that their own knowledge is insufficient to accomplish certain goals or tasks. Thus, the concept can be understood as the difference between the existing knowledge about a problem or topic and the knowledge that the user needs to have in order to solve the problem at hand. To satisfy the knowledge gap, users engage in information search which aims to identify desired information. The reasons for INs are varied and generally are defined based on three basic motivations: searching for answers to questions, a reduction of uncertainty, or a search for meaning.

One of the earliest IN theories, Taylor's Classic Model [47] developed in the 1960s, was based on research examining communication between libraries and library users. According to this Classic Model, the IN can be characterised through four levels:

- Physical (visceral) need: a conscious or unconscious IN state that is verbally inexpressible; it manifests itself as a vague feeling of dissatisfaction.

- Conscious need: refers to conscious but ambiguous IN formulation and expression.

- Formalised need: at this stage, the user is able to formulate an intelligible statement about their IN. However, it is not certain, whether this IN will be answered, either by humans or information system.
- Compromised need: refers to the last stage in which the IN is expressed by a query, which is submitted to a librarian or search engine.

Thomas D. Wilson [48, 49] considers the IN as secondary, based on a person's primary needs - physiological, cognitive, and affective. The determinants and contexts of the IN are the personal characteristics of the individual as well as various stimuli from the environment; the same set of contexts also provide the obstacles (barriers) that hinder subsequent information seeking.

Another influential theory, Belkin's Anomalous States of Knowledge (ASK) for IR (1980) [46], considered IN from a cognitive viewpoint. According to Nicholas Belkin, information is linked with uncertainty and the main incentive for information seeking is the ASK. Such a state occurs when an individual begins to become aware of anomaly in their knowledge that motivates them in a given situation to resolve this uncertainty by trying to find the necessary information. Having found the information, the individual then evaluates whether the anomaly has been resolved - if it hasn't, the ASK may reappear or the motivation to continue in this state may disappear.

A different way of approaching the IN concept is proposed by Brenda Dervin, who developed the theory of Sense-Making (the search for sense or/and meaning) [50, 51]. Dervin's theory describes the cognitive gap that a person consciously perceives and as a result of which they develop an IN. This cognitive imbalance must be bridged by acquiring new information or knowledge through the understanding of the situation. Dervin's model consists of four basic elements [49–51]:

- The current situation in time and space, which defines the context of the information problem;
- The information gap, which expresses the difference between the context situation and the desired situation;

- The outcome is the desired situation, such as the results of the sense-making process, or/and the solution to the information problem;

- Finally, the bridge, which bridges the gap between the actual and desired situation.

Charles Cole explains IN in a similar way - as a gap in understanding, that allows information to enter a person's cognitive system. It is a gap between identified problem and the problem solution, where IN is the input phase of the problem identification and problem-solving process. IN is an integral part of the adaptive mechanism that enables a person to adapt to changes in the physical/social environment. It is a mismatch between the perceived environmental stimulus (the bottom-up principle) and the response of the individual's cognitive system to that stimulus (the top-down principle). Cole's IN theory is the proposition that the user's IN itself does not evolve, but only aspects of the topic evolve or shift in the course of information seeking [45].

Ingwersen lists three basic types of INs related to search in information systems [2]:

- Verificative INs: the user needs to verify certain structured information (e.g. bibliographic data);

- Conscious topical INs: the user needs to clarify, assess or find certain unstructured topical information (terms, concepts, etc.);

- Muddled topical INs: the user is looking for new or needs to clarify unfamiliar concepts.

It is important to mention that INs do not exist in a vacuum, but are based on basic human needs and are described as secondary to those needs. Among the basic human needs include physiological (e.g. food, shelter), psychological (e.g. security) and cognitive (e.g. the need to learn skills) needs [46].

INs can also be classified as unrecognised (the user may not be aware that they have an IN until they encounter relevant information), unexpressed (people are aware of their INs but do nothing about them because they either can not or do not want to), wants (what the user would like to have) and requirements (a demand for information that the users believe they need). Some obstacles to meeting INs might be time, access to information sources, information overload (i.e. a cognitive state associated with an inability to process information efficiently and adequately [52]), personality, and the availability of information resources [53].

IN user states are difficult to research, as they are described as subjective states of mind, cannot be directly observed, and can be unconscious [54]. INs also often arise only when a person is confronted with a particular problem or is exposed to a particular situation. Furthermore, INs tend to change during the search process and over time in general. However, in information science, neither the "change" nor "over time" are clearly defined and researchers are inconsistent in interpreting these concepts [55].

To fully understand INs, it is necessary to examine the broader context of motivations for the use of retrieval systems that are increasingly becoming a part of our everyday life [54]. The IN consists of complex neurocognitive processes such as user's knowledge states and their feelings of knowing (i.e. one's own assessment of their extent of knowledge that at present they cannot recall [56]) and it is important for future research to recognise, address and explore these processes. An effort in this direction has been made by the recent work of Michalkova et al. [57,58], who explored the drivers of IN through investigation of the user's cognitive context.

2.1.2 Query Formulation

Another important concept closely related to IN is the query formulation, which refers to a user's attempt to express their IN. A query is therefore a formalisation of IN, a conceptualised difference between the user's actual knowledge about the problem and the knowledge they need to have to solve it. Formulated queries can, however, often suffer from ambiguity and bias which is an important challenge in IR [59].

2.1.3 Information Behaviour

Information behaviour is the sum of human behaviour in relation to information sources and information channels, including active and passive information seeking and usage [60]. Generally, information behaviour can be understood as all the human activities in the information environment (such as seeking, acquisition, processing, and retrieval), which manifest themselves through information interests, needs, and demands.

Interaction with information and resources can be active or passive. Active information interaction is concerned with the direct search and retrieval of information on the basis of a pre-formulated IN in the form of a query using keywords. Passive information interaction represents the reception of information by the user without any effort in the form of its retrieval, subsequent retrieval and processing [60].

Information behaviour also includes information avoidance, which refers to a deliberate rejection of available information that might happen for example as a result of information overload [52]. Additionally, people may avoid information that is in direct conflict with their views or beliefs and rather seek out information that supports them. This behaviour can also be referred to as confirmation bias [61].

Information Interaction. From the information science point of view, information interaction represents a multilateral exchange of information between IR units. Information interactions represent the mutual functioning of information and human beings in the information environment. It is the relationship between people and information in various forms and purposes, which can be divided into the area of human information-interaction (HII) and the area of human computer-interaction (HCI).

HCI is considered to be a multidisciplinary field focused on the design of computing technology and especially on the interaction between the user and the computer. The main focus of HCI is the exploration of user interfaces with a user-centric orientation with the aim of defining principles and ways of presenting information that facilitate effective human interaction with information [62]. It is a discipline concerned with the design, evaluation, and implementation of interactive computing systems for the study of the phenomena that surround them [63]. HII examines how people interact with information. It brings together different disciplines ranging from human-computer interaction to cooperative computing, human factors, computer and information science [64]. HII deals with how and why people use, search, consume and work with information to solve their information problems, as well as decision-making, learning and planning, and performing other tasks and activities [63].

2.2 Models of IR

Modern theories and models of IR have been heavily influenced by the cognitive approach to information science, which emerged around the mid-1980s [42]. A few examples of such prominent works that have shaped the foundations of IR through focusing on the user's cognition in the process of interaction with the system are:

- Ingwersen's Cognitive Approach [42]
- Kuhlthau's Information Seeking Process (ISP) model [65]
- Wilson's Nested Model of Information Behaviour [48, 49, 60]
- Saracevic's Stratified Model of IR [66]

The cognitive IR approaches, inspired by cognitive psychology, are seeking to bridge the gap between the user and the system by no longer viewing the users as a homogeneous group, but rather highlighting the existence of individual and contextual differences in search [67]. In general, these cognitive approaches usually depict the relationships of knowledge structures involved in the information transfer and IR interaction with varying degrees of detail [42]. The cognitive approach to IR recognises the the importance of user's cognitive states involved in the formation of mental representations and conceptual structures that underpin information processing. These cognitive states of users include perception, concentration, readiness to perceive information, motivation, the emotional component, fatigue and others.

Understanding cognitive aspects can significantly contribute to better user interaction with the system and increase the likelihood of users finding relevant information. Information systems should be designed to be consistent with human needs, abilities and characteristics. To design such systems, it is necessary to have a good knowledge of users, the reasons for their information behaviour and the variables, in which these processes take place [68]. Without addressing the cognitive variables influencing the information behaviour, it is not possible to fully understand and predict what information people will need and how best to help them in solving specific problems [69].

2.2.1 Ingwersen's Cognitive Approach

The penetration of cognitive science into information science has marked a fundamental change in the approach to IR. One of the first information scientists to address this cognitive shift was Peter Ingwersen, who developed several models based on cognitive user states [42]. According to Ingwersen, the application of the "cognitive perspective" to information science emphasises the complementarity between the social dimension of cognition and the individually oriented cognitive processes. This complementarity may provide unprecedented insight into the interaction of IR and information transfer. Ingwersen's models were based on Piaget's concept of cognitive structures, which can be characterised as patterns of mental or physical actions that underlie specific acts of intelligence. Ingwersen introduced the concept of "world knowledge" consisting of cognitive and knowledge structures of an individual [42]. According to Ingwersen, the cognitive structure is part of the interaction of the user's mental states and the mental models which are weaved into semantic (classifying information), and episodic memory (creating information of an event) [42].

2.2.2 Kuhlthau's ISP Model

The ISP Model [65,70] proposed by Kuhlthau captures the search process with emphasis on the user's cognitive and emotional states. This model began to take shape in the mid-1980s and captures generalised information seeking. While it was predominately applied in a library environment, it is also applicable in digital environments. The ISP model describes a user's experience in IR as a series of thoughts, feelings, and actions. Thoughts, initially uncertain, vague and ambiguous, become clearer, more focused and specific as the search process takes place. Feelings of anxiety and doubt turn into feelings of confidence or certainty [65].

The model consists of six phases:

- Initiation: occurs when an individual discovers a lack of knowledge or understanding, which commonly causes feelings of uncertainty and anxiety.

- Selection: in this phase, the area, topic or problem is identified and the initial uncertainty experienced by the individual is followed by a momentary feeling of optimism and readiness to start searching.
- Exploration: in this phase, inconsistent, incompatible information begins to appear and uncertainty, confusion and doubt often increase, while self-confidence decreases.
- Formulation: a focused perspective is created and uncertainty decreases with increasing confidence.
- Collection: as the user gathers more information and insight about the problem, their uncertainty decreases.
- Presentation: during this phase, the search is complete, there is a new understanding that allows the individual information seeker to gain insight about the problem, or there is a disappointment if the search fails.
- In the last, assessment phase (which is not one of the direct phases of the ISP, but is mentioned in the model) there is increased self-awareness and self-accomplishment of the searcher.

2.2.3 Wilson's Nested Model of Information Behaviour

According to Wilson's Nested Model of Information Behaviour, information seeking is a broader domain - an activity of deliberate search for information as a result of a need to satisfy a particular goal. Wilson divides human information activities and their search as follows:

- Information behaviour is generally human behaviour in relation to sources and channels of information, including active and passive information seeking and information use. Therefore, it includes direct communication with other people as well as the passive reception of information (e.g. people watch TV commercials without the intention of acting on the information presented).

- Information-seeking behaviour is the purposeful search for information as a result of the need to complete a certain task. During the search, the individual can interact with manual information systems (i.e. systems whereby individuals are required to perform all the tasks manually, such as libraries) or automatic systems (computer search systems) [71].

- Information search behaviour is a micro level of user behaviour during interaction with information systems of all kinds. It includes all interactions with the system, whether at the level of HCI (e.g. using a mouse and clicking a link) or at the intellectual level (e.g. learning a Boolean search strategy or determining criteria for deciding which of two books sitting next to each other on a library shelf is the most useful), which also includes mental activities such as evaluating the relevance of retrieved information [49, 60].

The hierarchy of individual modes of behaviour is illustrated by Wilson's model [60] in Figure 2.2. This hierarchy is captured by the different use of the terms seeking and search. The term "information seeking" draws attention to the context of the solved problem and the cognitive state of the user. Searching expresses the purposeful activity of the user and, according to Marchionini [72], "is closer to answering a question or learning". IR is only part of the search process. Therefore, some researchers use the term seeking for user interaction with the search system (e.g. Spink - information seeking and retrieving [73]). The term search refers more to the purpose for which information is sought to the solved problem.

2.2.4 Saracevic Stratified Model of IR

According to Saracevic's Stratified Model [66], the IR is seen as an interaction between two main elements: a user and a system. Both the user and system side consists of several mutually interactive levels or strata. Each stratum/level involves different elements and/or specific processes. On the human side, processes may be physiological, psychological, affective, and cognitive. On the computer side, they may be physical and symbolic. The interface provides for an interaction on the surface level in which:

Figure 2.2: A nested model of the Information Seeking and Information Searching research areas developed by Wilson.

1. Users carry out a dialogue with the computer through an interface by making utterances (e.g. commands) to receive responses (computer utterances). During this process, users not only engage in searching and matching (as proposed by the traditional IR model) but also in other processes such as understanding and eliciting the attributes of a given computer component, or information resource; browsing; navigating within and among information resources; determining the state of a given process; visualising displays and results; obtaining and providing various types of feedback; making relevance assessments; and so on.

2. Computers interact with users by providing responses in this dialogue.

The main purpose of user-system interaction is to affect the cognitive state of the user through the effective use of relevant information in connection with the problem at hand situated in the specific contextual environment. The dialogue can be reiterative, incorporating among other things, various feedback types, and can exhibit a number of patterns. Saracevic's Stratified Model of IR is depicted in Figure 2.3 [66].

2.3 Current Trends in IR Research

The research methodology examining user behaviour when searching uses a whole range of quantitative and qualitative techniques. Quantitative techniques consist, for example, in recording all performed operations down to the level of keystrokes or mouse move-

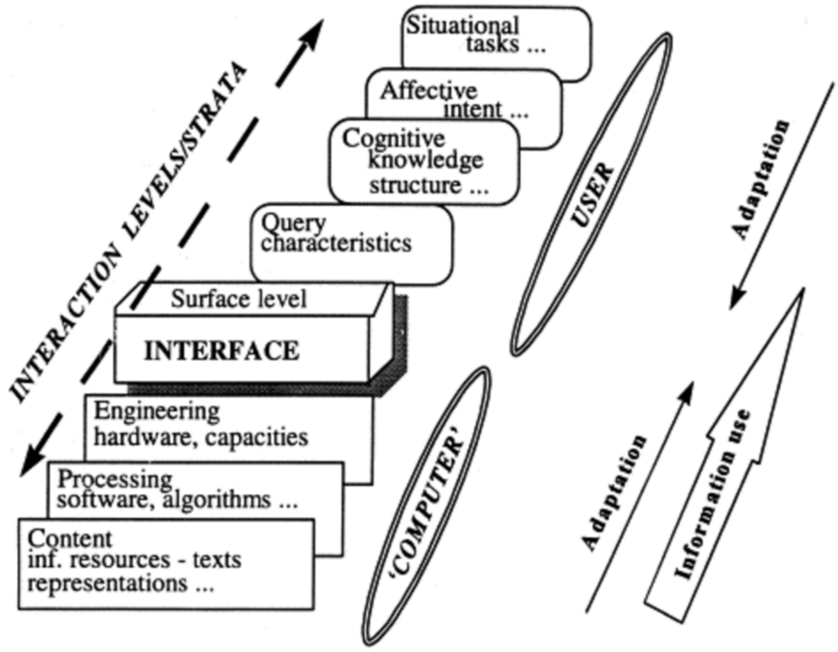

Figure 2.3: Saracevic's Stratified Model of IR.

ments (the so-called execution protocol), examining the number of records searched and evaluating the effectiveness of the search according to the criteria of accuracy and completeness, monitoring the elapsed time, the number of viewed and saved documents, creating statistics of the use of individual functions of the system, etc. Qualitative methods may include filling out questionnaires before and after the search session, monitoring and subsequent analysis of the user's interaction with the information worker, asking the user to explain all his steps in the system - the so-called "thinking aloud" protocol, examining the consistency of the relevance records evaluation during the search and after its completion with printed records in hand, taking video recordings of users interacting with the search system and the intermediary (if present) and their subsequent analysis, and many others.

Recent research in the field of IR focuses mainly on users. The main focus of user-oriented studies is to isolate and track individual aspects of the search process [74]. It is important to note that everyone searches in a different way, uses different search techniques, and the overall way of working with retrieval systems is different. This means that individual users might come across different search results [74].

2.4 Relevance

The main goal of the IR systems is to retrieve relevant information or information units that would help users to satisfy their INs and to achieve the search task goal [43]. Therefore, relevance is commonly referred to as the fundamental and timeless concept within the IS&R [10,14,34,75]. It constitutes a major research area in IS&R as it plays a crucial role in the user-system interaction and it is a substantial indicator of system retrieval performance, representing a relation and a measure [10, 76].

Relevance has been considered in information science as a multidimensional [10,12, 14], dynamic and complex process [1,13,15,16], which is difficult to quantify and which depends on users' perception of information relating to the specific IN situation at a certain time point [10, 77–79].

People usually understand the meaning of the term relevance intuitively. However, available definitions in information science often differ from one another. In general, the concept can be understood as the degree of utility that exists between text or document and the user's request for query or information [42]. Relevance can be measured on the basis of units of measurement: precision and recall. Precision measures the extent to which a retrieval system finds only relevant documents. The recall is a measure of completeness or quantity, as it measures the success of the retrieval system in finding all relevant documents that can be found [54].

Within the context of modern information science, relevance is no longer just a black and white category (i.e. relevant vs. non-relevant), but a complex contextual and socio-cognitive relationship often supported by the functionality of technologies and connections in the electronic environment. Therefore, when talking about relevance, we do not ask whether the information is relevant, but from what perspective and in what context it is relevant.

For some users, the evaluation of relevance ends with an assessment of the formal characteristics and basic content matching of the information resource to the information requirement; for others, a more detailed understanding of the content is part of the process. In this respect, we can also speak of 'primary' and 'secondary' relevance. While in the assessment of 'primary' relevance, users largely decide whether to classify an information resource (such as textual document) for detailed study at all, the assessment of 'secondary' relevance can already be considered to some extent as part of the study or research process.

2.4.1 Relevance Theories

The information science literature discusses relevance from two dominant perspectives. The first view is characterised as systemic. Here, relevance is referred to as objective and the usefulness of a resource is based on the quality of its inherent attributes [80]. The second view is characteristic of studies focusing on users from a cognitive perspective. These studies are mainly concerned with the subjective aspects of relevance. Subjective relevance is an umbrella term for subjective thematic and situational relevance (or usefulness) [12, 81, 82]. The extent of information usefulness is determined by how well it can be applied to the specific context (situation) of the information search [12].

Relevance is only meaningful in relation to goals or tasks, and this relation can only be assessed by the human being. Thus, relevance according to Hjørland and Christensen [83] can be understood as something serving as a tool to a goal. Here, a tool is viewed in a broad sense and includes resources, information, stimuli, ideas, things, etc. The evaluation of relevance depends on the user's clarification, what they want to achieve and what alternative views are represented in the information resources [83].

Cognitive relevance is defined as a relation between the information object and a cognitive state of the user's knowledge and is inferred by the criteria such as informativeness, information quality or/and novelty [34]. Understanding of cognitive relevance might lead to a better understanding of the user's engagement with the document [84].

Saracevic's Stratified Model of Relevance.

As a part of the IR process, the user usually expresses the subjective IN through the formulation of the query. The query is submitted to the system, which then presents the user with retrieved information (system relevance). The user then interprets and relates the retrieved information to the problem at hand, cognitive state, and other aspects (user relevance). According to Saracevic [10], relevance is, therefore, one of the most important concepts in the field of IR. Relevance can be understood as a measure of the appropriateness of a given answer to a query. This implies that information relevance is never absolute, but is always based on its relationship to another entity (e.g. a query or - more generally - a context).

Relevance can be understood in two ways: as the objective relevance of a document to the query topic or as the relevance of the document to the user. Relevant information can then be referred to as information, that is relevant to a given problem. Saracevic [85], in his analysis of approaches to relevance research, also states that relevance implies a certain relationship, with the type of this relationship defined by the involved entities. Based on this, the theory distinguishes between five relevance types:

- System relevance is defined by the relationship between the search request/query and the information unit.
- Topical (subject) relevance is the relationship between the topic of the information requested and the topic of the information unit.
- Cognitive relevance (pertinence) is the relationship between the user's knowledge state of the subject and the information unit.
- Situational relevance (utility) is the relationship between the current situation, task, problem and the information unit.
- Motivational (affective) relevance is the relationship between intentions, goals, motivation of a user and the information unit [85].

Information or information units are selected as relevant (or expressed on some continuum of relevance) from a number of available existing, or even competing information sources. The selection process involves a series of information interactions of various kinds. Relevance has a set of general attributes that are rooted in human cognition:

- Relation: Relevance arises when expressing a relation along certain properties, frequently in communicative exchanges that involve people as well as information or information objects.

- Intention: The relation in the expression of relevance involves intention(s) — objectives, roles, and expectations. Motivation is involved.

- Context: The intention in the expression of relevance always comes from a context and is directed toward that context. Relevance cannot be considered without context. It is possible to distinguish between internal context (involves cognitive and affective states of a user) and external context (directed toward a situation, tasks, and problem-at-hand). Additionally, social and cultural components may be involved as well.

- Inference: Relevance involves the assessment about a relation, and it is, therefore, created or derived on that basis. Inference may also involve a selection from competing sources geared toward the maximisation of results and/or minimisation of effort in dealing with results.

- Interaction: Inference is accomplished as a dynamic, interacting process, in which an interpretation of other attributes may change, as context changes.

- Measurement: Relevance involves a graduated assessment of the effectiveness or degree of maximisation of a given relation, such as an assessment of some information sought, for an intention geared toward a context.

It is important to note that one of the main limitations of Saracevic's stratified relevance model is that the model is not detailed enough for experimentation and verification [10]. However, Weigl and Guastavino [17] discussed the model's potential application and usefulness in user-centred music IR research. White confines Saracevic's Relevance Theory to the application of the cognitive effects and processing effort. In his theory, White states that these are the components that can be used as predictive mechanisms for the operational assessment of relevance [86].

2.4.2 Relevance Granularity

Relevance has always been an equivocal concept in IR and thus it has given rise to many studies and scientific discussions [87]. Up to this date, there are still debates in IR over the granularity level of relevance assessment that should be collected from users [21]. Relevance can be considered on a binary (i.e. relevant or non-relevant) or graded (i.e. highly, partially or non-relevant) level. The predictive accuracy of graded relevance has outperformed the approaches relying on a binary scale, which has proven its efficiency [87, 88]. Past research has been mainly dedicated to examining user's relevance in binary terms [10]. However, the utilisation of binary scale is only one of the options for information categorisation. Moreover, recent findings support the idea of categorical thinking [24], which suggests that users usually divide retrieved results into 3-5 categories based on results' relevance [79]. Employing graded relevance, in comparison to the binary one, has also been shown to improve ranking functions [87,88]. However, users' perception of relevance continuity has not been yet examined in depth and our understanding of how users perceive a different degree of information relevance [88] is limited.

It is crucial to understand what each grade of relevance actually means. The value of evaluating information based on graded relevance has begun to receive attention in recent years both from system [88, 89] and user [5, 33, 90] point of views. This is particularly important since the granularity of relevance judgements in previous studies have been based on investigating this phenomenon indirectly, via some sort of mediator [22,24]. Past theoretical concepts [10,43] have proposed to subdivide relevance judgement into regions of high, middle and low relevance assessments, as relevance seem to be bi-modal (having high peaks at endpoints of the range) [10]. Current research examining system performance employing graded relevance supports Saracevic's theory, suggesting that graded relevance improves the document retrieval effectiveness [23,91].

2.4.3 Challenges Associated with Relevance

Relevance is difficult to define [92] and the terminology has not been consistent. Different researchers assigned different meanings, with diverse components and criteria that underline this concept [93], aiming to develop an ideal and widely accepted relevance model [3, 17]. Such conceptual inconsistencies might be related to the fact that despite a rich theoretical background, empirical research examining relevance is still relatively recent. Furthermore, relevance feedback submitted by individuals is frequently associated with a large variability that is comparable to individual differences in other cognitive processes involving information processing, such as indexing, classifying, searching, feedback, and so on. Individual differences are one of the most prominent features and factors in relevance inferences.

2.4.4 Capturing Relevance Assessments

The concept of relevance has a long theoretical background, which emphasises the complementary relationship between the user and the system. While system-oriented research is well-established, the investigation of the users' internal processes happening during relevance assessment is still relatively recent [26, 94]. However, the role of the user in IR is critical as they play an active and integral role during the evaluation of retrieved results, which is vital to the functioning of the system [10, 75, 95]. The user interacts with the system through relevance feedback [96,97], which has been developed to improve the representation of user's IN [76]. Relevance feedback is a complex iterative cyclic process involving a series of user-system interactions aiming to reduce the semantic gap between the user's IN and formulated queries [10, 26, 98, 99]. The feedback cycle aims to progressively and interactively determine the user's desired output based on the user's evaluations, which are used to automatically modify the retrieval process [100]. The utilisation of the relevance feedback cycle significantly speeds up the search process and improves retrieval performance [95, 98] as the system uses the user's feedback information as directions to retrieve topically similar documents [99]. Given the importance of the user side of relevance, IR systems have been employing mechanisms to capture this phenomenon to maximise the relevance of retrieved results. Past research investigating users' perceived relevance has introduced a number of feedback techniques, which vary from explicit, implicit, and psycho-physiological signals. These feedback techniques determine document relevance with respect to the cognitive, situational, and psycho-physiological levels of the interactive dialogue occurring between the user and the retrieval system [27]. Features extracted from users' explicit, implicit, and psycho-physiological feedback can be used to build models, which are able to automatically predict the relevance of information [101].

Explicit Feedback

Explicit feedback is the most common and the most robust practice used to directly annotate content within the user-system interaction with the aim to improve retrieval effectiveness [29, 102]. This traditional relevance feedback technique requires users to manually submit feedback on the information content using mechanisms such as rating, tagging and bookmarking [29, 103, 104], which frequently leads to a significant improvement of the search result rankings quality for a given query [96]. Furthermore, explicit feedback is easy to use and has low uncertainty due to a user's overt control [105]. However, obtaining explicit feedback is frequently referred to as challenging due to the cognitive burden and manual effort associated with direct user interaction [26, 29], which can be physically and mentally demanding. The user is required to explicitly state whether presented content is subjectively perceived as relevant or non-relevant [106], which increases the task complexity and the cognitive resources required from the user [27]. Additionally, explicit feedback suffers from a trade-off in terms of the user's willingness to devote time (especially without a clear incentive) to explicitly evaluate retrieved information, despite the fact that they are aware that doing so will improve the search performance [29,107]. Explicit feedback mechanisms may also suffer from biases and as they do not allow for continuous monitoring they are not always applicable [26, 108].

Implicit Feedback

Implicit feedback, in turn, refers to techniques that attempt to automatically and indirectly infer the relevance assessments by monitoring the users in an unobtrusive manner (i.e. without users' additional explicit input) [102, 109]. The motivation for using implicit feedback is in relieving the user from the cognitive burden associated with the laborious task of providing document rating and relevance assessments [27] by passively observing users-system natural interactions and taking advantage of users' behaviour. Therefore, implicit feedback aims at analysing user context or interpreting the user's natural interactions with the search interface in order to generate relevance annotations of information items [29, 110, 111]. Most of the research in this area relied on the use of surrogate interactional measures based on behavioural features (e.g. document retention, search results click-through, mouse movements, reading time, scrolling, saving, printing, text selection) [104, 112–114] and psycho-physiological signals (e.g. eye-movements, galvanic skin response, facial expressions) [26, 27, 115–117] or a combination of these recorded from users [26, 118]. Nevertheless, implicit feedback is often considered to be less accurate due to the noise associated with it [5]. Additionally, implicit research is often limited to tasks for which implicit behaviour can be observed and often require data recorded from a large participant sample. Although the estimated relevance scores are typically not perfect characterisations of the user's needs, they can still be used for guiding the search [110] as they offer an excellent possibility of obtaining annotations for a large number of information items without imposing additional mental load, which users find intriguing [119].

Behavioural Features. Early research employing the implicit feedback method was mainly focused on the use of dwell time (i.e. document viewing time), as an indicator of relevance [103, 120, 121], because of its applicability for real-time systems [122]. The traditional approach of using dwell time as a potential factor for predicting document relevance was based on premise that users spend more time previewing relevant documents compared to the non-relevant documents [120]. Nonetheless, past studies [114, 123, 124] investigating relevance within more complex and naturalistic settings suggest that dwell time alone is not a reliable behavioural implicit relevance signal measure when considered on its own, as it can vary significantly according to a specific task and individual differences [113]. As a result, later research attempted to combine dwell time with the task information, which has improved retrieval performance [125]. However, as information about the search task is not always available, dwell time has also been combined with additional behavioural data obtained by monitoring the user information-interaction, such as scrolling [120], mouse movements [126–128], search results click-through [104, 112], text selection actions [129] and exit types [103]. Analysis of user's behaviour during document evaluation is a good indicator of relevance, leading to improvement in relevance prediction, which can significantly improve search [26]. It is important to note that relevance feedback relying on behavioural features has been shown to have mixed effectiveness because the selected measures of users' interests and preferences are often affected by many factors. For instance, users' clicks are noisy, and frequently susceptible to position and trust bias [130]. Therefore, inferences drawn solely from behavioural interactions might not always be valid [131].

Psycho-Physiological Features. Although implicit feedback is typically obtained by analysing behavioural actions (tracking mouse movements, scrolling, link clicks, etc.), the more relevant perspective for this work is provided by studies inferring user feedback from physiological signals. The psycho-physiological feedback has been proposed in addition to behavioural implicit feedback in order to better capture variable cognitive states during relevance operations [26, 27]. The idea of psycho-physiological feedback is to track eyes [132], capture physiological signals (such as galvanic skin response, skin temperature, and heart rate) [115], facial expressions [26, 118] and use them as implicit relevance assessment.

Eye-tracking. Eye-tracking is the most widely used unobtrusive research approach, where pupil dilation, fixations, gaze points, and eye movements are the basic output measures of interest. The essential functioning of the eye-tracker is based on the "eye-mind hypothesis" suggesting a link between the direction of the human gaze and the focus of attention [133]. The fixation length of a given area is linked to the amount of time required to process the information. Longer processing time indicates higher cognitive effort and vice versa [133]. Therefore, eye-tracking seems to be well suited to provide valid and valuable information about the user's perceived relevance of evaluated information. Overall, the findings of past eye-tracking studies established a relationship between several EYE features and text passage relevance [110] as well as improved classification of processing states on three simulated search tasks (subjective interest, question-answer, and word search) [134]. Furthermore, the findings indicate that text relevance influences reading behaviour and visual processing of text [135]. Furthermore, eye-tracking data can be effectively used for binary relevance prediction, with an accuracy of 70 - 75% [136].

Measuring pupil dilation has roots in early cognitive psychology research. Pupil dilation is mediated by the Autonomic Nervous System and early research has associated this physiological response with a number of cognitive functions, such as surprise [137], decision making [138], interest [139, 140] and mental workload [141–143]. In general, pupil size variations are linked to attention [144–146]. Therefore, it is reasonable to expect that document relevance will be linked to attention or mental workload and consequently, pupil diameter [135]. However, there are only a few published studies examining the relationship between information relevance and pupil dilation. For instance, the findings of Oliviera et al. [147] suggest that pupil dilation is linked to the higher relevance of text and image web search results. Additionally, Gwizdka et al. [135] investigated the relevance of short text documents and Web pages [136] and showed significant pupil dilation on relevant documents, related to cognitive effort. The pupil dilation was the most significant in the one-two second period preceding relevance decision [136]. They also showed significant differences in pupil dilation on fixations on relevant words and on relevance decisions [132].

Another commonly assessed eye-tracking aspect used to interfere with cognitive processes or mental states are gaze-fixations. Fixation is considered as a cluster of eye-gaze coordinates within a specified range in time and space when our eyes hold the vision in place so that the visual system can uptake the visual information of interest [148]. Past research examining the association of eye fixations and relevance processing found that relevant information is associated with a higher number of fixations [149, 150]. However, in terms of fixation length, previous studies yield conflicting results. While Balatsoukas and Ruthven [151] have shown that users made more frequent and longer fixations on non-relevant document surrogates, Gwizdka [135] and Villa & Halvey [152] found that non-relevant documents impose the lowest mental workload. Furthermore, Buscher [110] found that fixation duration was not an effective discriminator between relevant and not relevant text. Despite fixation on a word being usually interpreted as user interest or word relevance, eye fixation data are often noisy and can have different causes depending on the user's cognitive states. Therefore, it is not clear yet to what extent fixation duration relates to relevance [110]. Additionally, simple gaze-based measures like fixation usually rely on the analysis of single terms in a text. However, the information and relations in a text document are mainly based on a specific combination of words in sentences and paragraphs rather than single isolated terms [110]. Taylor [153] found that first fixations and regressions are the most useful EYE features for the selection of additional query terms in an implicit relevance feedback system.

Using eye movements is considered to be more appropriate in examining the relevance of not only textual [154] but also visual [155, 156] information. The findings of previous research suggest that eye-movements features are an important indicator of implicit relevance feedback [110] with a potential to significantly improve IR system's performance [157]. Marcos, Gavin, and Arapakis [158] examined the eye and mouse movement behaviours of web users who interact with SERP snippets incorporating images, multimedia, and text. They developed measures of noticeability and interest using fixations data, and conversion using click-through to better understand the features of a well-designed, engaging, attractive and aggregated SERP [158]. Moe et al. [159] found that the amount of reading behaviour is informative to the relevance of the reading text. It is important to note that, generally, eye-tracking data are relatively tightly coupled with cognitive processes.

However, eye movements are often subconscious and there is a large number of internal and external unknown factors influencing them. Hence, obtained data is usually very noisy and require careful interpretation and manipulation. Additionally, it is important to mention that there are reading-related individual differences among readers as well as document-induced changes for a single reader. For example, saccade sizes can range from 1 to 15 characters, while fixation duration can vary between 100 - 500ms for the same reader. The variability is further influenced by a variety of characteristics, including the reader's reading style, background knowledge, word predictability, and reading difficulty [160].

Emotional Features. Growing scientific evidence suggests that users naturally express emotions during their interactions with search systems and information items [161–163]. Emotions influence users' interests [164], motivation [165], and play an important role in the process of IR as they can be used as an effective source of implicit feedback, to personalise search [115, 116]. Tkalcic et al. [166] used the user's emotional metadata in combination with generic metadata in image recommender systems for improving image recommendation results. Overall, their findings indicated that using the user's emotional information improved the image recommender system's effectiveness compared to using only generic metadata. Emotional features have been shown to enhance recommendation effectiveness not only for images but also video content [167, 168]. Later work by Moshfeghi et al. [169] has shown that emotional features can be combined with physiological signals to model relevance and predict task types.

2.4.5 The Role of Brain Imaging

Relevance assessments are underpinned by a series of complex cognitive phenomena and existing implicit feedback techniques consider relevance with respect to the cognitive and situational levels of interaction [109]. However, these methods can only help researchers to understand the concept of relevance to a certain degree, and the effectiveness of some of these techniques is limited [90]. IR is one of the fields that could significantly benefit from the use of a direct neuroscientific approach used to access complex mental processes in the brain [105]. Mental processes can reveal information about information relevance thereby providing an effective way to implicitly collect relevance feedback with great efficiency. Brain activity can be used to automatically annotate information items for future use and collaborative filtering [29]. Furthermore, understanding cognitive processes can provide important information that can potentially improve information presentation to the user, considering their cognitive workload, awareness, and other mental states [105]. Brain signals can be unobtrusively recorded in the background in real-time, as the user interacts with the system, which in the future has the potential to augment standard input devices (e.g. computer keyboard and mouse) for interaction between the user and the machine [170].

2.5 Introduction to Brain Imaging

The human brain is the most complex organ with diverse functions. Through the processing of different types of information inputs from different parts of the body, it gives rise to elaborate molecular, cellular, and neuronal phenomena which consequently form the physical and biological basis of cognition. The brain is organised into anatomically or functionally defined cortical regions constituting the foundation of cognitive functions that are optimally adaptable to environmental perturbations [171].

The past decades have shown revolutionary development, improvement and modernisation of our ability to non-invasively image human brain activity, with current spatial (on the order of millimetres) and temporal (on the order of milliseconds) resolutions meeting standards previously reserved for invasive methods in animal models. The employment of non-invasive brain imaging technologies enables the exploration of functional brain areas and structures which has significantly enriched experimental research and contributed to our understanding of information processing in the brain. The brain imaging research is based on the following empirically tested facts:

- A nerve cell called a neuron is the fundamental functional unit of the brain and nervous system.
- Communication between neurons generates the electrical and magnetic signal. The signal, conducted from the brain source, travels within the brain but also up through other brain areas, the blood-brain barrier, through the skull and the scalp, resulting in microvolt (μV) signals reaching the scalp.
- The brain is divided into areas that have different functions. These brain functions are approximately the same across individuals.
- The human brain consists of very complex webs of multiple interconnected structures, creating neural networks.

- Distinct brain areas are interdependent and do not work in isolation.

Brain imaging refers to the use of quantitative (computational) techniques that employ an interaction between brain tissue and various forms of physiological energy (e.g. electromagnetic or particle radiation) to capture positional data about the structure and function of the brain. Such data are used to generate corresponding brain maps [172], which aim to provide a detailed picture of brain connectivity and organisation. With recent advances in the field of neuroscience, there are many brain imaging tools available, such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and EEG, that have been used in IR.

fMRI. fMRI is an effective non-invasive imaging technique based on the examination of changes in brain blood flow (i.e. hemodynamic response). At the site of brain activity, there is an increased influx of oxygenated blood which supplies activated neurons with energy. The fMRI method uses a magnetic field to detect the subtle changes in the blood oxygenation levels as oxygenation impacts blood magnetic properties [173]. The measurement of neurological responses based on differentiated blood oxygenation is called BOLD (Blood Oxygenation Level Dependent). Although the BOLD signal is an indirect measure of brain activity and is susceptible to influence by many physiological activities of the body, past empirical findings demonstrated a strong correlation between brain tissue activity and mental processes [174]. It is important to note that the hemodynamic responses are relatively slow, noisy and weak. Therefore, low temporal resolution is often considered as one of the major limitations of fMRI. The typical BOLD hemodynamic post-stimulus response begins to rise after 1-2 seconds, peaks at 4-6 seconds and returns to baseline after 12-16 seconds. Furthermore, evoked fMRI signals in short time scale events might overlap, which increases the difficulty of determining individual events [175]. The fMRI is associated with additional limitations, such as high scanning cost, long acquisition time, presence of a strong magnetic field and participant sample selection. In terms of experimental design, participants are required to lay down still to avoid image distortion, which can cause fatigue and discomfort, consequently affecting participants' cognitive performance. Therefore, careful consideration must be taken with regard to study design and implementation [176]. However, fMRI is one of the most important methods for examining brain activity as high spatial resolution enables a comprehensive coverage of the whole brain [177]. The fMRI imaging can provide important information about the functional brain architecture and the organisation of neural interactions [178].

MEG. MEG is another non-invasive technique for investigating human brain activity which uses superconducting sensors (SQUIDs) to detect very subtle changes in the magnetic field generated by electric currents in the brain. Unlike fMRI, MEG can record neural changes with millisecond precision, and thus, it has a high temporal resolution. Additionally, it has a good spatial resolution; sources can be localised with millimetre precision [179]. MEG is an extremely sensitive method as magnetic fields are not attenuated by the skin, scalp, and skull, in contrast to electrical potentials [180–182]. During the MEG assessment participant is in a sitting position, which allows for conducting cognitive experiments that more closely resemble real-life scenarios. MEG does not require exposure to strong magnetic fields or loud noises, which is an advantage when compared to fMRI acquisition sequences. As the MEG acquisition is silent, the method can be used to study neural responses to sound stimuli [183]. Therefore, MEG is a low-risk assessment that can be repeated in participants of all ages as often as needed. The subject preparation time for the assessment is reduced in MEG compared with EEG as there is no need to connect all conductive electrodes on the scalp (however, modern EEG systems offer dry or quick electrode application alternatives) [184]. The main disadvantage of this method is the need for specialised shielding to eliminate the magnetic interference found in a typical urban environment. MEG is also more expensive and not as good as fMRI at localising precise brain activity locations. Therefore, obtained MEG data are often combined with fMRI data. Furthermore, for overall signal quality, it is important that the participant's head is as close to the MEG sensor arrays as possible and head movements are restricted [185], which makes it hard for participants to interact naturally [186]. Compared to EEG, the traditional SQUID-based MEG is less portable. However, with the recent development, new MEG devices (i.e.

OPMs – optically pumped magnetometers) are becoming more portable, comfortable (permitting head movement) and cost-efficient (no need for expensive supercoolant). Therefore, there is an opportunity to record brain activity while participants perform realistic actions and movements.

EEG. EEG is a non-hazardous and non-invasive brain imaging technique that allows measurement of the electrical potential of brain activity on the scalp surface that is generated by the activation of neurons, while the brain is at rest or performs different cognitive and behavioural tasks. The EEG method allows monitoring complex cognitive activity with a near real-time precision due to its high temporal resolution [187]. Changes in neural electric currents reach the surface of the head, where they are detected by a set of electrodes (also called sensors or channels) attached to specific standardised scalp locations. Each electrode measures the voltage at its location and enables the transfer of electrical activity from the scalp surface to the EEG input of the device. While recorded electrical activity differences are typically evident, the raw activity measured on any given site does not necessarily reflect the unique activity in the region since a strong signal from a distant brain region could dominate a local signal in the vicinity of the electrode site. Furthermore, the electrodes are separated from current sources in the brain by cerebrospinal fluid (CSF), the skull, and the scalp which might considerably infiltrate the EEG potentials [188, 189].

For the purpose of the data collection in this work, we used 128-channel Geodesic Sensor Net. Figure 2.4 shows the electrode placement scheme used for data recording. Following the approach of Bian et al. [190] the scalp is divided into five regions and two hemispheres (left (LH) and right (RH)), namely: frontal (F), right temporal (RT), central (C), left temporal (LT), and posterior (P) regions. Figure 2.5 depicts a colour-coded map of the electrodes in the 128-channel layout, with each colour being assigned to a different scalp region. The EEG channels are arranged in a cap which is placed on the participant's head, aligning the Cz electrode with the top-centre point of the head.

The EEG captures electric potential fluctuation changes between the channels and a reference point (an electrode whose potential is stable and frequently equals zero).

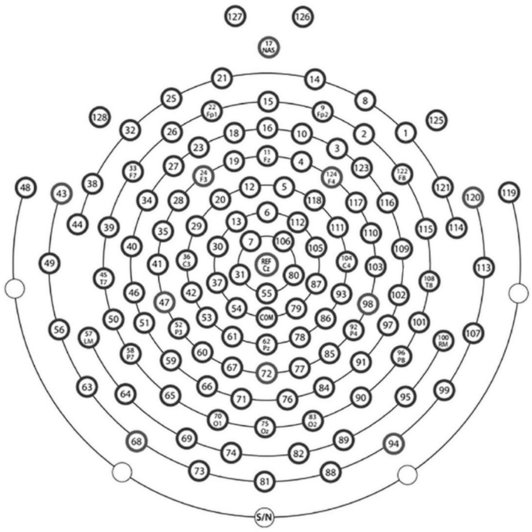

Figure 2.4: 128 channel HydroCel Geodesic Sensor Net (HCGSN).

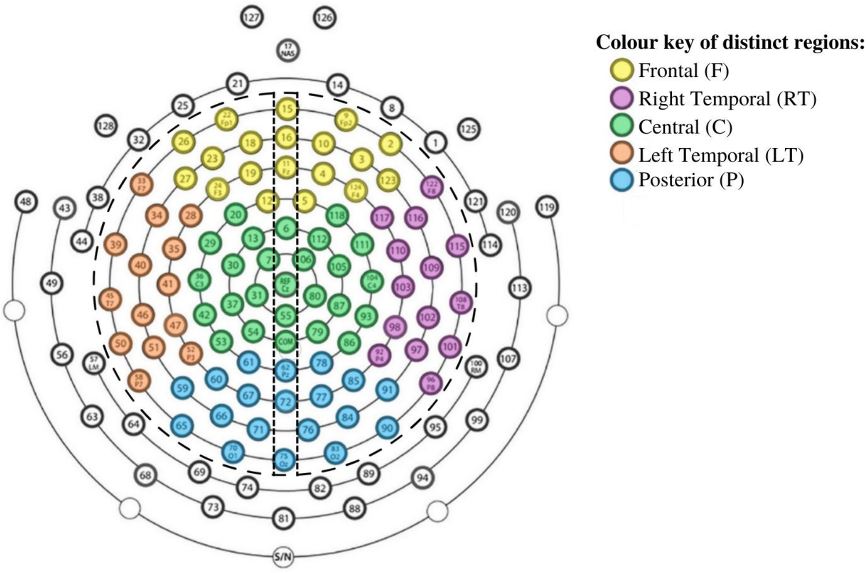

Figure 2.5: Colour-coded electrode map for 128 channel layout.

Each colour represents a particular region. Electrodes of interest are within the black dotted circular region. A vertical dotted box divides the net into left and right hemispheres.

The information recorded from the EEG takes the form of the sum of voltage changes of neurons that are detected by different electrode types. The result is a wave representing the course of potential difference changes in time. Amplitude refers to the height of a waveform or the strength of the pattern in terms of μV of the EEG signal. The recorded signal consists of many waves with different characteristics. The rate at which the waveform data is sampled in order to convert it into a continuous digital format is known as the sampling rate (measured in Hertz (Hz)).

Voltage amplitude changes range from 0.5 to 100μV [191], while the conventional bandwidth of EEG studies focuses on the waveform analysis ranging from 0.5Hz to 70Hz [192]. The signal of each recording electrode is contrasted with a reference electrode, which in turn influences the amplitude at each EEG channel and time point [193]. The synchronisation between the behavioural responses of the participant and their brain signals is facilitated via the amplifier. Obtained neurological data may contain interfering elements at different frequencies with extracerebral origin (e.g. eye movements, muscle contractions or/and ambient electrical noise). Additionally, high-density EEG recordings are commonly associated with bad channels, which are common phenomena that arise due to various technical reasons, such as a bad connection between the electrode and the scalp. To account for interfering elements, acquired (raw) data usually undergo a series of pre-processing steps (see Section 3.3.1) which aim to maximise signal-to-noise ratio.

The brain waves' shape, location and character are directly dependent on the actual activity of the brain. For the most accurate interpretation of brain activity, it is necessary to analyse the recorded neurological signal. The brain waves can, therefore, be divided into bandwidths to describe their functions delineating slow, moderate, and fast waves:

- Delta waves (0.5 to 4Hz) have the largest amplitudes and are commonly referred to as slow-wave activity. Normally, the delta band brain waves are associated with the deep sleep stages over frontal scalp locations [194].

- Theta waves (4 to 8Hz) are related to the number of cognitive tasks, especially involving working memory, executive control, goal-oriented behaviour and short-term memory load [195–197].

- Alpha waves (8 to 14Hz) reflect visual stimulation, attention, target discrimination and aid overall mental coordination [195, 198, 199].

- Beta waves (13 to 30Hz) are high-frequency and low-amplitude waves, the most prominent in the frontal and central scalp locations [192]. Beta brain waves are commonly observed in an awakened state and reflect the brain's active engagement in cognitive processes such as assessment of subjectively-relevant stimuli [200].

- Gamma waves (30Hz and upwards) are considered to be the smallest and fastest brain activity signatures [201] responsible for cognitive functioning, learning, memory, and information processing [202].

Another commonly used approach to analyse multichannel data between conditions is to quantify the difference of the topography in a given EEG segment or a time window of interest (i.e. epoch) and to test it for significance. This approach is not only applicable for the analysis of continuous EEG signal, but also for the analysis of ERPs. ERPs are scalp-recorded long latency voltage fluctuations that measure neural response time-locked to an onset (start) of a specific event or stimulus which reflects cognitive processing [203]. The ERPs provide unique insight into neurological processes with unrivalled time resolution. The ERP component represents a deflection from the baseline of EEG activity which correlates with cognitive processes. The ERP waveforms consist of a series of positive and negative amplitude fluctuations. Although some ERP components are denoted by acronyms (e.g. LPC), most components are denoted by a letter (N/P) indicating polarity (negative/ positive), followed by a number specifying either the latency (delay) in milliseconds or the position of the component in the wave. For example, the negative peak of the curve, which is the first significant peak in the wave and often occurs about 100ms after the stimulus, is usually called N100 (with a latency of 100ms after the stimulus and negative polarity) or N1 (indicating that it is the first peak and is negative). The N100/N1 component is frequently followed by a positive peak, usually called P200 or P2. Latencies for ERP components are often quite variable [204]. The P300, N400 and P600/ LPC ERP components have been the most commonly studied within the context of IR [5, 33, 101, 205, 206]. The component-driven

ERP analysis relies on a priori analysis decisions made based on previous research (e.g. selecting regions of interest (ROIs)[1] or the time measurement window). On the other hand, in the data-driven ERP analysis, researchers often use statistical tests to identify ROIs. The data-driven ERP identification compared to component-driven analysis avoids the analytical biases introduced by apriori implication of known ERP components [30].

The EEG is by far the most widespread brain activity recording modality because of its advantages such as the possibility of repeated examination, non-invasiveness and the possibility of a longer duration of the examination, which increases the application in practice. Compared to fMRI and MEG, EEG equipment is cheap, portable, easy to use, set up and more readily available [208,209]. Given the EEG benefits, the technique is becoming increasingly used in commercial Brain-Computer Interfaces (BCIs). Nevertheless, EEG has several limitations. EEG is only sensitive to post-synaptic potentials generated in the superficial layers of several cortical regions that are often detected simultaneously. Therefore, with the low spatial resolution, it can be difficult to precisely localise the exact area/region of activity. EEG is not sensitive to neuronal responses from structures that are deep in the brain, such as the hippocampus [177]. Additionally, the EEG signal is very sensitive to noise and artifacts whose origins are not cerebral. They may arise from the participant (i.e. muscle movements, skin resistance) or/and from electrical interference with a power line or surrounding electrical apparatus [208]. Furthermore, it is almost impossible to reconstruct a unique intracranial current source distribution for a given EEG signal, although substantial recent progress has been made in this area [177].

Brain Activity Modeling. Brain imaging technologies enable researchers to perform analyses resulting in assumptions about the nature of the observed neurological signals. These assumptions can be encoded in powerful computational models that bridge descriptive analyses and neuroscientific theory in a mutually explanatory manner. Computational brain activity models are able to elucidate how sensory information

[1]ROI refers to a selected region of neighbouring electrodes that jointly and significantly contribute toward neurophysiological phenomena of interest [207].

is represented in the brain and predict brain activity in response to certain stimuli [210]. The ultimate purpose of these models is to explain behavioural, neurophysiological, and neuroanatomical data in a manner that provides a comprehensive description of processes ranging from high-level brain function and behaviour down to the single-neuron level [211]. While the existing models present exciting possibilities for understanding the brain, the behavioural repertoire of current models is still limited. Therefore, the effort to expand such models' functional capabilities and improve their behavioural flexibility is still ongoing.

2.5.1 Brain, Mind and Behaviour Relationship

An in-depth understanding of users' information processing and evaluation requires a multidisciplinary approach, combining recent research advances in IR and cognitive science. From a cognitive perspective, it is important to understand how mental processes underpin user information interaction. In a short period of time, the brain is able to integrate sensory signals and produce a cognitive representation of the internal and external environment. Sensory inputs activate brain systems that allow the emergence of cognitive processes, which in turn influence human behaviour. This section focuses on defining and conceptualising frequently reported cognitive processes within the IR literature, such as interrelated basic (perception [11], attention [212], motor cognition [90], memory [213]), higher-order (problem-solving [214], learning [215], language [216], and executive functioning) and metacognitive abilities (i.e. organisation and evaluation of one's thought processes which relate to learning and problem-solving) [217]. The human brain, mind and behaviour are closely interconnected components and research methods studying neural mechanisms that underpin cognitive states aim to provide a holistic approach to understanding human behaviour.

Perception. Perception refers to the ability to detect, capture, process, interpret and integrate information about our internal and external percepts received from our sensory receptors [218]. Perception is, therefore, not only the passive recipient of sensory signals but instead, an active process, influenced by an individual's memory, learning, expectations and attention, involving the transformation of low-level sensory information to higher-level cognition [219, 220].

Perception occurs in three stages: selection, organisation, and interpretation. Selection is the first stage, during which we focus our attention on certain incoming sensory information. We tend to pay attention to information that is salient. Salience is the degree to which the information attracts our attention in a particular context. Selected information is then organised by the means of subjectively meaningful categorisation which is based on innate and learned cognitive patterns. Interpretation, the final perception phase, refers to the process of attaching meaning to the selected stimuli using mental structures known as schemata. Schemata are like databases of stored, related information that we use to interpret new experiences [221].

Generally, the process of perception can be understood with two fundamental approaches to perceptual processing: bottom-up and top-down [222]. Bottom-up processing is an explanation for perceptions that start with sensory input and work upwards through the uninterrupted cascade of transformations until a mental representation of the perceived information is obtained. This process suggests that our perceptual experience is based entirely on the data available from our senses that can independently create increasingly complex representations [223]. On the other hand, top-down processing can be understood as the influence of our inner goals on stimulus selection [224]. Therefore, perception is developed through contextual information - the process begins with the most general known information (that has already been brought in by the senses) and moves toward the more specific (the interpretation of finer details) [225].

Attention. Attention is the means by which we actively, selectively and consciously process a limited amount of internal and external stimuli [226]. The selective aspect of attention has an important adaptive function, protecting individuals from being overwhelmed by the abundance of information and allowing them to focus primarily on currently relevant information. Thus, attention acts as a selective filter mechanism that allows us to focus our cognitive resources on relevant information.

According to Sohlberg & Mateer's model [227–229], which is widely used and operationalised, hierarchically orders attention [230] into five levels: (i) focused (withholding irrelevant information while focusing on relevant information), (ii) sustained (maintained continuous response to the presented stimulus), (iii) selective (focusing on one relevant stimulus at the time), (iv) alternating (ability to control attentional allocations in order to switch between dissimilar cognitive tasks), (v) divided (ability to respond simultaneously to multiple task demands while maintaining accuracy and speed). Attention can be divided as either voluntary or involuntary [231], which serve different functions and are controlled by distinct mechanisms [232]. Voluntary (also referred to as endogenous, or top-down) attention is a sustained goal-driven process, which involves higher mental effort due to being instructed to orient attention to a particular location. On the other hand, involuntary (also referred to as exogenous, or bottom-up) attention is a passive, transient, automatic, stimulus-driven process, during which automatically captured signals propagate from lower sensory areas to higher cognitive processing areas [233]. Both voluntary and involuntary attention are interrelated and simultaneously affect a certain proportion of psychological activity.

The attention control, intensity and shifting can be affected by internal and external determining factors. Internal factors depend on individual differences and some examples are motivation, interest, the effort required by the task, the individual's physical state, thought processes, and personal or social significance. Internally oriented, goal-driven attention is also referred to as top-down or endogenous attention [234]. External factors are environment-dependent and usually governed by the characteristics of the stimuli. These external factors could be related to the stimuli novelty and familiarity, intensity, environment contrast, repetition, movement, size etc. Stimulus-driven attention is referred to as bottom-up or exogenous attention [234].

Motor Cognition. Motor cognition (i.e., cognitive processes that underlie complex motor output) encompasses the mental processes involved in the planning, preparation, and production movements with the aim to satisfy a specific motor goal, as well as the higher cognitive processes involved in anticipating, predicting, recognising, mimicking, understanding and interpreting events in the physical and social environments. The fundamental aspect of the motor cognition paradigm is the perception–action cycle, involving the transformation of perceived patterns of intended movement into coordinated patterns of actual movement. Both cognitive and motor functions are controlled by brain areas such as the cerebellum, basal ganglia and frontal lobes that govern the executive function and intentional movements requiring anticipation, planning and prediction [235].

Memory. Memory is a complex, three-stage process involving the encoding, storage and retrieval of past experience [236, 237]. Any successful act of remembering requires all three stages to be intact. Encoding is the process of converting physical stimuli into a form that the brain's memory system can interpret and use. During encoding sensory input (acoustic, elaborative, visual and semantic) is detected by sensory receptors which send the signal into the somatosensory centre in the brain. Storage refers to the process of keeping memories intact in the brain's memory system over time. The brain can store episodic, procedural and semantic memories. Retrieval is the process of accessing specific memories in storage and bringing them into consciousness. There are two types of memory retrieval - recall (i.e. unaided retrieval of memories) and recognition (retrieval with the help of hints) [238].

Retrieved memories can be categorised as explicit (conscious storage and subsequent processing of memories) or implicit (unintentional recollection). Explicit memory can be divided into semantic and episodic. Semantic memory is focused on storing concepts, facts, data and terms. Episodic memory, on the other hand, captures events, time, place, percepts, thoughts, and emotional information [239].

Many traditional theories distinguish between short-term and long-term memory. which differ based on their capacity and duration. While long-term memory is thought to have unlimited capacity and memories can be stored for up to several decades, short-term memory can only hold a few items for a short time duration. Past research suggests that people can store between five to nine items [240] or approximately four chunks or information pieces [241]. The term short-term memory is often used interchangeably with working memory. However, it is important to note that working memory refers to the processes used to temporarily store, organise, and manipulate information, while short-term memory, refers only to temporary information storage. Working memory plays an important part in the integration of short-term and long-term memory information and guides reasoning, decision-making and behaviour [242, 243]. The distinction between short-term and long-term memory was central to the multi-store model of Atkinson and Shiffrin [244]. However, recent theories have proposed unitary-store models, in which this distinction is not so clear-cut as the approaches focus on similarities between short-term and long-term memory, rather than their differences. Both of these approaches have considerable strengths and limitations. It is important to note that memory research is still ongoing and novel discoveries call for existing memory theories to be revised. For instance, recent research suggests that memories can be stored in the brain even though they could not be retrieved through natural recall cues. Memory storage does not rely on the strengthening of neural connections, as it has been originally thought. Therefore, long-term memory storage is possible even without cellular protein synthesis as it is stored as a specific type of connectivity between neurons [245].

Problem-Solving and Decision-Making. The terms problem-solving and decision-making are frequently used interchangeably, but they are not the same [246]. Problem-solving refers to a structured analytical process of investigating the given information and finding all possible solutions to the situation at hand [247].

It is possible to distinguish between five strategies for finding new solutions to solve problems: trial and error, memory retrieval, algorithms, heuristics, and insight. The trial and error problem-solving approach involves multiple activity patterns for evaluating solution ideas. If the evaluation result is negative, a different activity pattern needs to be generated and evaluated again until success or a solution is reached [248]. Memory retrieval in the context of problem-solving is based on recognition and refers to the retrieval of previously acquired knowledge [249]. The use of algorithms refers to the methodological development of a step-by-step method to solve a problem [250]. Heuristics are not a formal problem-solving models as such. They are defined as mental shortcuts (sometimes based on information stored in memory), which help with the thinking processes in a problem-solving situation which do not require an optimal solution [251]. Means/ends analysis is an example of a heuristic tactic, which requires the recognition of discrepancies existing between the current and goal situation and the identification of intermediate steps necessary to reduce the difference. Another heuristic tactic is decomposition, which requires the overarching problem to be broken into smaller pieces [252]. In some cases, the solution to a problem can appear as a sudden insight, which is an unexpected emergence of a solution strategy. Generally, in contrast with other problem-solving approaches, insight is not deliberate and conscious processing that advances step by step and problem-solvers are often not able to consciously explain how they generated a solution in a sequential manner [251].

However, problem-solving is not always a flawless process and there are numerous barriers that can interfere with an individual's problem-solving ability. Commonly, problem-solving barriers refer to mental constructs impeding the ability to correctly and efficiently solve problems. Some examples of problem-solving impediments include functional fixedness, confirmation bias and mental sets. Functional fixedness is a cognitive bias involving a tendency to view an activity or an object as only working in a particular way, which, therefore, impacts an individual's creative abilities. Confirmation bias is the tendency to search for, interpret, focus, favour, and recall information that confirms or supports one's existing beliefs, views or values and to neglect evidence that disconfirms it [253]. A mental set is a form of rigidity to approach problems in a certain fashion due to prior experience which can lead people to make problem-solving assumptions without considering all the available information [254].

Traditionally, it is argued that problem-solving is a step toward decision-making, so that the information gathered in that process may be used for decision-making. Therefore, decision-making refers to an action based on either an intuitive or reasoned process (or a combination of the two) derived during the problem-solving, which is based on assumptions of values, preferences and beliefs of the decision-maker [255]. Intuitive decision-making (or 'gut feeling') relies on non-sequential, rapid, non-conscious information processing and recognition of patterns and associations to derive affectively charged judgements [256]. On the other hand, reasoning is a more complicated decision-making process and tends to require a more formal, structured approach, consisting of a sequence of linear and logical steps designed to rationally develop a desired solution [257]. Both problem-solving and decision-making involve critical thinking, which is a process of discovering, conceptualising, analysing, applying, synthesising, and/or evaluating information (gathered or generated via reflection, reasoning, communication, or/and observation) with the goal of finding the best possible solution to the problem [258].

Learning. Learning can be understood as an enrichment of individual experiences. It is defined as either temporary or permanent behavioural change obtained via purposeful, planned and systematic acquisition of skills, knowledge and habits which occur as a result of practice. The definition does not include behavioural changes caused by maturation or temporary organism state (e.g. due to substance abuse) [259].

In a broader sense, it is possible to distinguish between four types of learning behaviours: habituation, classical conditioning, operant conditioning and cognitive learning. Habituation, the simplest form of learning, is a decline in responding to a repeated or prolonged stimulus presentation that is not caused by fatigue or adaptation [260]. The concept of classical conditioning (also known as Pavlovian or respondent conditioning) is a type of learning that happens unconsciously through the association of a neutral stimulus paired with a biologically potent stimulus (i.e. reflexive response). Classical conditioning was first demonstrated by Pavlov, who found that dogs began to salivate as a response to a bell sound prior to being fed [261]. Operant conditioning, also known as instrumental learning or conditioning, occurs when behaviour changes as a function of its consequences called reinforcement schedules, i.e. by using behaviour to 'operate' with the environment. A reinforcement schedule is any procedure that presents (or removes) a reinforcer (or punishers) to an organism as a result of specified behaviour following precise rules [262]. Cognitive learning is an active, constructive, and long-lasting process of change in knowledge attributable to experience by which the learner takes in, interprets, stores, and retrieves information [263].

Language and the System of Thinking. Language and thoughts play central roles in human cognition. Although interrelated, language and thinking are not identical. Thinking is a higher cognitive process in which the mind employs various structures and processes of conscious awareness such as mental imagery, episodic memories or explicit anticipations to produce thoughts. Thoughts can be expressed and shared via language [264].

Language (spoken, signed, or written) is a communicative, structured, arbitrary, dynamic, and generative system that enables humans to perceive, store and share information [265, 266]. Language, as a structured hierarchical communication system, involves systematic grammatical rules to organise lexical information transmission from one individual to another. Lexicon refers to the vocabulary of a given language. Thus, the lexicon is a language's vocabulary. Grammar refers to the wide set of structural rules that are used to convey the construction of sentences, clauses, phrases and words in a language through the use of the lexicon. The syntax is a part of grammar which refers to the way words are organised into sentences [267] and it plays an important role in the interpretation of the sentence meaning (i.e. semantics of the sentence).

Executive Functioning. Executive functioning includes high-level mental control processes that underlie effective planning and organisation of goal-directed behaviour, especially in the novel, unstructured, and non-routine situations that require some degree of judgement. Executive functioning refers to the human ability to drive the attention in the desired direction and move away from unwanted stimuli, anticipate future events, self-control, formulate realistic goals and set priorities. Thanks to the executive functions, we are able to multi-task, recognise errors and learn from them and adapt to (unexpected) changes [268, 269].

Brain Region Networks. Studies of functional connectivity patterns (based on contexts or correlations) between brain regions have shown that brain regions do not work in isolation. Instead, human functioning is attributed to the neural functioning of multiple brain areas working together in predictable sequences and forming large-scale neural networks that subserve different cognitive functions [270].

2.6 NeuraSearch Science

In recent years, novel NeuraSearch research direction (an innovative user-centred approach bridging brain imaging methods and IR) represents a growing interest in HII and IR research [6]. This research line was able to bring new knowledge and understanding of IR phenomena as the field profits from direct access to neural signatures associated with user's mental processes [6, 271–274] such as attention, cognitive workload and memory. The main benefit of the application of the neuroscientific approach is that the internal mental states of an individual can reveal information about IR process in response to particular information items, which overcomes the self-referential nature of direct and obtrusive methods. Therefore, the neuroscientific approach reduces cognitive load and enables data collection in real-time without disrupting the user's search process, while also tackling measurability issues [105].

The most frequently used neuroimaging methods in the NeuraSearch research have been fMRI [90, 102, 275–277], MEG [278] and EEG [5, 94, 105, 170, 205, 216, 273, 279–283]. These neuroimaging techniques have been classified by their spatial vs. temporal resolution and portability. Spatial resolution refers to the capacity a technique has to allow discrimination between different active brain regions with high accuracy. On the other hand, temporal resolution refers to the accuracy of the scanner to capture brain activation in relation to time.

Opportunities and Challenges in NeuraSearch

The NeuraSearch experiments employing brain imaging methods must be carefully designed and conducted, considering the strengths and limitations of available methods. Commonly used brain imaging methods restrict participants' movements and, therefore, can not be used with uncooperative participants. In addition, functional imaging is not merely correlational. Brain imaging studies provide mainly information about the casual relationship between manipulated variables (such as stimuli and task instructions) and their influence on brain activity. Therefore, inferring cognitive process engagement from the activated brain sites should be done with great caution [284]. The use of neuroscientific equipment to examine IR phenomena requires multidisciplinary expertise and skills from fields such as neuroscience, cognitive psychology and information science. Equipment cost and accessibility are other important factors to be considered as in particular fMRI and MEG are very expensive and require access to neuroimaging centres [285].

On the other hand, brain imaging has improved our understanding of cognitive processes underpinning HII and helped to explain fundamental IR phenomena. Brain imaging data provide an objective measure of information-seeking behaviours while considering unobservable (e.g. mental workload, attention) and observable (click behaviour, information selection) variables [285]. Brain imaging can be used to verify existing theoretical models. Furthermore, brain signals can be potentially decoded using BCI [286,287] which enables the communication between the brain and an external device (e.g. computer) [288].

2.6.1 NeuraSearch Research

Novel NeuraSearch research has begun to unravel the brain activities of users interacting with information systems when conducting a problem-solving task, which has fundamentally changed the foundations of modern IR research [6]. Users frequently engage in IR to find a solution to a subjectively experienced problem-solving scenario, motivated by the state of uncertainty stemming from their lack of knowledge (coined as IN). To overcome the state of uncertainty, users usually formulate the query (that best represents their IN) which is submitted to the system. The system then retrieves documents for the user to evaluate. Therefore, the user judges retrieved documents (coined as relevance judgement) based on their relevance to the experienced IN until they are able to solve the problem at hand and consequently satisfy their IN. Information search and retrieval are, therefore, complex processes consisting of many interconnected steps and involving a variety of intricate cognitive mechanisms. NeuraSearch research direction attempts to better understand neurological signatures underlying IR and search to improve users' experience in such tasks. This section presents existing research in this direction.

Recent studies have employed brain imaging techniques to gain a better understanding of the IR&S process as a whole [276, 289] and also to examine its interactive parts such as IN (e.g. [275]), query formulation [290], relevance (e.g. [278]) and search satisfaction [276, 291].

Information Search. Information search has become the most performed activity on the web, forming an important part of everyday life and helping humans to find a wide range of information. Seminal approaches and theories in the IR have subdivided the search process into subprocesses which can be analysed and evaluated. The influential work of Moshfeghi and colleagues [276] investigated the search process using a cognitive approach suggesting that the search process can be divided into different mutually interactive cognitive processes. Using fMRI, the researchers discovered and mapped large-scale functional neural networks associated with functional brain activity changes between different stages of the search. These stages included IN realisation, query formulation and query expression, relevance assessment and search satisfaction. Therefore, the analysis of neural activation enables the identification of distinct search process parts as the user moves through them [276].

Furthermore, the study of Paisalnan et al. [289] was able to identify and compare functional brain activity differences between IN, relevance and satisfaction leading to a better understanding of the involvement of several cognitive functions and their relationship within the search process.

Information Need. One particular area of emphasis for NeuraSearch research has been to examine the IN process [102, 275]. This is due to challenges associated with IN understanding, capturing, and in turn, satisfying [277]. IN realisation is underpinned by complex high-level cognitive processes involving memory retrieval, decision-making, attention and executive functioning. Therefore, traditional user-based research relying on implicit or/and explicit user assessment could provide only limited insight into how exactly such a phenomenon emerges. Moshfeghi, Triantafillou and Pollick [275] used fMRI to example neural activation associated during the complex and cognitively demanding process of IN realisation. The authors found detectable neural correlates of

INs that can be identified at an early stage of the information seeking and retrieval process, as well as distributed network of specific brain regions where INs manifest themselves. The fMRI analysis has revealed higher activation in the caudate, thalamus and right inferior frontal gyrus when participants did not experience IN. In contrast, the brain activation in the region known as the dorsal posterior cingulate cortex was associated with the realisation of IN. Monitoring the activity in the posterior cingulate cortex might, therefore, provide useful signals necessary for the detection and identification of IN [275].

Additionally, it has been found that the prediction of the IN state experienced by the user is possible using brain signals [102]. Moshfeghi and colleagues [102] also examined personalised (unique to the individual) or generalised (same across participants) brain activity patterns associated with the IN realisation and found that personalised IN prediction is more accurate than the generalised approach. This finding suggests that despite some general similarities associated with brain activation (mostly concentrated in frontal and occipital brain regions), individual differences are an important factor to consider within the context of IN realisation. The work, therefore, highlights possibilities for the prediction of complex cognitive processes using brain activation recorded using fMRI [102].

The findings of fMRI research have led to the development of a novel IN model incorporating the user's neurophysiological responses to the presented information [277]. The model has been proposed by Moshfeghi and Pollick [277] and suggests, based on increased activation of certain neural networks, that the realisation of IN is underpinned by the three interrelated components: (a) successful memory retrieval, (b) an information flow regulation, and (c) a high-level perception.

To understand cognitive processes underpinning the realisation of IN, Michalkova et al. [292] conducted an EEG experiment comparing the IN states with the non-IN scenarios. The authors found differences in the early stages of the IN realisation attributed to the N1-P2 complex (related to early stimulus recognition), which suggest that the realisation of IN is manifested in the user's brain before they are consciously aware of it.

Query Formulation. Searcher's ability to formulate a query is a crucial part of the search process as it directly influences the search system's performance. Despite the critical role query formulation plays in the search process, our understanding of neuro-cognitive processes associated with the recall and reorganisation of query terms specific to a particular document was until recently very limited. The study of Kangassalo et al. [290] investigated neuro-cognitive mechanisms involved in the estimation of query terms goodness using EEG. The authors found that the term specificity was associated with the amplitudes of P200, P300, N400 and P600 ERP components, which differed significantly between query-specific and non-specific terms. Therefore, the authors have demonstrated that the human ability to detect specific terms has neural origins [290].

Search Satisfaction. While the above-mentioned study of Moshfeghi and Pollick [276] identified search satisfaction as one of the stages of the search process, it did not focus on satisfaction on its own. The user's experience of satisfaction of their IN plays an important role during IR as it signals that the user has gathered adequate information to answer their need. The realisation of search satisfaction frequently leads to the stop of the search process. To better understand the user's internal states associated with IN satisfaction, Paisalnan et al. [291] examined the involvement of brain regions during this subjective and complex process. The findings suggest that user's satisfaction arises from the actions of inter-related neural regions associated with both cognitive and emotional control. Paisalnan et al. [293] found that IN satisfaction is not only underpinned by cognitive but also affective processes that differ significantly when the IN is vs. is not satisfied.

2.6.2 Neuroscience & Relevance

The neuroscientific approach to relevance has brought novel knowledge and understanding in terms of unravelling neural processes happening in the brain [90]. The efforts in this direction might be categorised in two ways based on the experimental design used to measure relevance. The first line of brain-imaging research (NeuraSearch [6]) has position relevance within the context of IR task while considering the user's IN [206].

Participants experienced the IN through the creation of a problem-solving scenario. In order to satisfy their IN, participants engaged in the relevance assessment tasks, while their brain activity was monitored. Another direction of research examined relevance in the context of information associations, where participants did not experience IN and were mainly instructed to judge information association to the topic. Therefore, it is possible to argue relevance was investigated in loose terms, without considering its relationship with the IN of a user [18, 206]. Overall, the results of these studies, employing brain imaging, suggest that overall, neurological signals differ significantly during the processing of relevant content vs. non-relevant content across individuals.

Relevance and NeuraSearch Research

The first line of brain-imaging research has positioned relevance within the IR task [90]. The tasks of these studies were designed to include simulated IN situations through the creation of a problem-solving scenario. To satisfy the IN, participants were presented with images [5, 90, 95, 170, 278], videos [26, 205, 294, 295] or textual information [94, 105, 216, 276, 280, 296] and instructed to judge the relevance of the presented content with relation to the problem at hand, while their brain activity was monitored. Many different types of search tasks have been used to operationalise relevance, ranging from known-item identification [5, 90], information lookup [94, 276], exploratory search [216], and browsing [105].

Moshfeghi and colleagues [90] conducted the first brain imaging study examining neural signatures associated with relevance by employing the fMRI technique. The study aimed to answer the fundamental question "how does relevance happen in the brain" through the localisation of brain activity differences in cortical regions during the processing of relevant vs. non-relevant images. Contrasting relevant and non-relevant images was associated with significant differences in the right brain hemisphere, including the superior frontal gyrus, the inferior parietal lobe and the posterior area of the inferior temporal gyrus. The three cortical regions have shown greater activation associated with judging relevant compared to non-relevant images. A later fMRI study conducted by Moshfeghi and Pollick [276] identified two brain networks showing functional connectivity during relevance assessments: frontoparietal task control and the default-mode network. While the frontoparietal network is essential for our ability to coordinate behaviour in a goal-driven, rapid, accurate, and flexible manner, the default-mode network might be related to self-referential or self-initiated thought [276]. These study results have provided an important insight into how the activity of large-scale brain networks that subserve fundamental cognitive functions underpin relevance assessments.

Kauppi and colleagues [278] employed MEG to show that the frequency content of the MEG signal, along with eye movement data can be used to decode the perceived relevance of viewed images. The study constitutes an important step toward the development of brain-activity and eye-movements-based interactive image retrieval systems [278].

Temporal pattern of brain activity related to relevance assessment phenomena has also been studied using EEG alone or in combination with pupillometry or/and eye-tracking devices [94, 170]. For instance, Behneman et al. [296] demonstrated that changes in the EEG signals (theta and alpha bands) can be used to differentiate between relevant and irrelevant sentences to a given IN. Gwizdka and colleagues [94] investigated the dynamics of text relevance decision-making in a Q/A task. The findings suggest differences in cognitive processes (reflected via EEG-measured alpha frequency band) used to evaluate texts of varying relevance levels, as well as provide evidence for the ability to detect these differences in information search sessions using direct measurement of eye movement via eye-tracking and EEG [94]. Later, Gwizdka [216] found significant differences in EEG-measured power of alpha frequency band and in EEG-detected attention levels associated with relevant and non-relevant judgements during information search tasks conducted on the open web. Ye and colleagues [297] found that relevance assessment of textual stimuli happens around 800ms. Furthermore, monitoring the neural activity associated with post-relevance assessment revealed differences in the processing of non-relevant and relevant words, that persisted for approximately 260ms to 320ms for relevant words and 500ms to 530ms for irrelevant words [280].

Using EEG, relevance has been inferred not only for textual stimuli but also for images [5, 170, 271] and videos [205, 294, 295]. In terms of image stimuli, Allegreti et al. [5] examined neural activity related to the processing of relevant vs. non-relevant images to identify when relevance assessment happens in the brain. The researchers identified time intervals and significant differences in brain activity shifting during relevant and non-relevant image processing. The findings suggest that such processing happens as follows: 180 - 300ms - an early process of implicit assessment of relevance (frontal areas - F1; AF4) and stimuli processing. At this stage, there is no relevance assessment. Between the 300 - 500ms time interval, the differences in the brain activity elicited by relevant and non-relevant images start to shift towards central areas C2 and CP2. Finally, the most significant differences are associated with the 500 – 800ms time interval in the electrodes located around Cz and C1 over the central area of the scalp. Hence, the human brain requires around 800ms to assess relevance of visual stimuli [5]. Past research demonstrated [170, 271] that EEG and eye-tracking data can be used to estimate the user's subjectively perceived relevance from a user's image search. Additionally, studies [170, 271] demonstrated that implicit data constitute an important input necessary to resolve possible image search result ambiguities when users query for an ambiguous search term. Mohedano and colleagues [95] explored relevance feedback mechanisms for object detection image retrieval. Researchers compared the performance of brain signals to mouse-based interfaces and found that when users have limited time to judge the relevance of the image, both interfaces are comparable in performance. However, when using data only from the best-performing users, researchers found that EEG-based relevance feedback outperformed mouse-based feedback [95].

In the context of video stimuli, Kim and Kim [294] found that the central, frontal, and pre-frontal positions of the left hemisphere were linked to the most significant neural activity differences during the topical relevance detection of video skims. The neural differences were the most apparent during the 300$\sim$500ms and 500$\sim$700ms time range, corresponding to the N400 and P600 ERP components, which have been shown to be indicators of relevant and non-relevant judgements [205, 294]. The N400 ERP component was associated with greater amplitude negativity in topic-irrelevant shots than in topic-relevant shots. The P600 ERP component elicited greater positivity in

ERP amplitude for topic-relevant compared to topic-irrelevant shots [295]. Both ERP components showed fronto-central scalp distribution pattern [295].

Above mentioned studies have provided an important empirical grounding for relevance through the understanding of neuro-cognitive processes, which is an important step towards the development of neuro-adaptive IR systems. This has been demonstrated by a recent study by Jacucci and colleagues [105], showing that relevance can be predicted in real-time from EEG brain signals and eye movements while the user engages with the system and IR task.

Relevance in the Word Associations Context

Another stream of relevance research has considered relevance in the context of information associations, employing EEG either in isolation [33, 101] or in combination with eye gaze [283]. In these scenarios, participants did not experience IN, but they engaged in judging information associations to the given topic. For instance, in the study of Eugster et al. [101], participants were asked to assess the relevance of the term stimulus to a predetermined topic. A similar approach was employed in a study designed by Wenzel and colleagues [283], where participants were instructed to find words (e.g. goat, tractor, mouse) belonging to one out of five semantic categories (e.g. category: transportation). Overall, study findings suggest that information associations can be traced to the EEG signal as well as ERPs [101], which is reflected by the corresponding P300 [101, 283, 298], N400 [33] and P600 components [290]. Mentioned P300, N400 and P600 ERP components have also been reported in the studies positioning relevance within the context of the IR task. EEG signal, as well as ERPs, can be used to define a set of features that enable decoding and inference of human judgements directly from the human cognitive states without any direct explicit physical intervention of the user [33, 283, 290, 299] not only for textual but also picture stimuli [288, 298]. The Pz EEG electrode location was associated with the highest significant difference between category relevant and irrelevant information, regardless of stimulus modality [33, 290, 298]. Based on the latest methodological and technical advances in BCI, there is increasing interest in using neurological signatures trained to detect relevant information for the automatic generation of relevant content that matches given task [288, 298]. For instance, brainsourcing, a recently developed paradigm, allows direct mapping of neurally measured implicit reactions within recognition task to predict targets of interest to appear [119].

2.7 Research Motivation

Relevance assessment remains a major study area in the field of IR. Recent findings employing brain imaging to investigate relevance have shown that human mental experiences during the IR process can be understood and decoded using non-invasive measurements of brain activity. Hence, the recent application of the neuroscientific approach has brought valid and valuable insight into a better understanding of relevance. Relevance has been studied within the context of multiple stimuli modalities (e.g. videos, images) but the most information consumption in IR happens in textual format [29]. Assessing the relevance of textual documents, given IN, involves several cognitive processes including reading comprehension. Therefore, it is one of the most complex cognitive activities in IR [300].

Additionally, since relevance is a complex process, it is important to highlight the benefit of combining multiple data collection tools, which has become very popular in recent years. As Kelly and Belkin suggested [124] tools such as questionnaires enable researchers to explore participant views of a task and topic familiarity, which influence relevance perception. Furthermore, the researchers highlighted the importance of the naturalistic approach, which optimises ecological validity [124]. It is essential to design the task, which will closely model real-life user-system interaction and place relevance assessment within the context of IR. The experimental design needs to consider that subjective relevance assessment made by user [78] is preceded by the IN and that encountering a relevant document should result in a change of user's knowledge state [42]. Additionally, physical, cognitive or/and affective aspects of information should also be considered [76]. Topical relevance, in analogy with text comprehension, can be examined through the relevance of individual words, sentences or/and the main ideas of the document to the query [301].

In this thesis, we aim to investigate cognitive processes that underpin textual relevance assessment employing the NeuraSearch framework. In the task, we evoke artificial IN and engage participants in relevance assessment. We will aim to capture the time-course of the relevance assessment during the visual presentation of text. The controlled experimental task is designed to mimic real-life human-information interaction while keeping the limitations associated with the EEG in mind to avoid signal contamination. In addition to obtaining EEG signals, we gather survey and behavioural data to better understand participants' subjective experiences during the task. The study is built on the previous literature investigating relevance through the comparison of neural signals associated with the content of different relevance [5, 33, 101]. In addition, to provide a better insight into the cognitive stratum of relevance, the task is be designed to explore participants' SPK.

2.7.1 Binary Relevance

Relevance assessment is a holistic cognitive response with underlying neuropsychological mechanisms that form more basic perceptional and cognitive features of some sort. In terms of relevance granularity, the binary division has been the prevalent approach in the IR field, keeping the assessment cost low while maximising the number of relevant documents per topic, guaranteeing measure stability [10, 302]. Therefore, the majority of EEG studies using the neuroscientific approach have investigated relevance as a binary variable. However, it is important to note that these studies have either misplaced relevance within the context of word-relatedness (i.e. IN was not considered) [33] or (when considering IN) predominantly focused on ERP component-driven analysis [206].

While past ERP component-driven analysis and theory-driven approaches have contributed to the understanding of binary relevance assessment formation, these approaches can constrain knowledge by potentially overlooking all the possible features or dimensions that synthesise complex phenomena such as relevance assessment. On the other hand, a data-driven approach is a useful tool when it comes to making sense of behavioural responses during complex tasks. Despite its advantages, the data-driven analysis might be challenging because the EEG data frequently exhibit spatial heterogeneity, have non-stationary and multiscale dynamics, and are typified by substantial individual variability [303]. Many scientific observations of brain activity may be scope limited, constrained by opportunistic participant sampling, and have reduced reproducible controls. Relying on varied background assumptions while employing a data-driven exploratory approach might help to overcome the above-mentioned limitations and provide significant benefits associated with the potential discovery of novel, previously not reported cognitive phenomena.

This research aims to explore previously not reported potential components that could arise from binary relevance phenomena, while avoiding the potential analytical bias introduced by the restriction to distinct ERPs reported by previous studies [30]. We aim to re-visit textual binary relevance assessment, which would enable us to compare experimental results obtained using a data-driven approach with previously reported results associated with textual binary relevance assessment formation.

This is the first NeuraSearch EEG study investigating binary relevance assessment using a data-driven approach to gain a holistic view of ERP components underpinning complex cognitive processes associated with relevance assessment. To do so, we capture the users' binary relevance assessment in real-time as they engage in the Q/A task. The findings can help to further explore cortical differences associated with the two types of textual relevance assessment and to validate the results of past studies. The outlined research motivation to investigate binary relevance using a data-driven approach is addressed in Chapter 6.

2.7.2 Moderating Effect of SPK in Relevance Assessment

Relevance assessment is often investigated in a context-independent manner [22, 304]. Nonetheless, relevance assessment strongly depends on the user's cognitive states, knowledge, and perception [25, 305], which provide psychological context determining the problem and situation at hand [10, 12, 306, 307]. This work, therefore, aims to better understand the role of users' SPK within the relevance assessment context.

Past IR studies have mainly focused on topical knowledge referring to the relationship between one's prior knowledge and the conceptual aspects of the topic they engage in [308]. These studies have found that topical knowledge influences users' relevance criteria and evaluation process [305, 309, 310], as users rely on their knowledge to discriminate between relevant and non-relevant information. Furthermore, topical knowledge has also been shown to help users assess information credibility with higher accuracy [22]. Although topical knowledge plays an important role in information processing, users are often unaware of their knowledge anomalies [311] which significantly impact information search motivation and decision-making [312]. This paper focuses on SPK, which refers to self-assessment of knowledge that one believes to hold irrespective of what they actually know [313]. We follow a common approach using post-task assessment to evaluate participants' SPK states, which allows participants to be more cautious with the estimation of their SPK through the recognition of their anomalous knowledge states [311, 314]. However, both relevance assessment and SPK are dynamic, complex and subjective phenomena, which are difficult to quantify [90]. Thus, the present study takes the neuroscience approach which addresses the aforementioned challenges by offering the unique possibility of investigating these complex cognitive phenomena directly through the understanding of neurophysiological correlates of cognitive processes [90]. The outlined research motivation to explore users' SPK within relevance assessment using a data-driven approach is addressed in Chapter 5.

2.7.3 Graded Relevance

Past research has been mainly dedicated to examining users' relevance in binary terms, which is only one of the options for information categorisation. Binary relevance implies a direct, fixed and unchanging topical relationship between the user and IN, which might not reflect the true nature of relevance mental representations [10, 43]. Thus, to account for such circumstances, the relevance of information should be inferred on a continuum and comparatively.

On the other hand, the importance of information evaluation based on graded relevance has begun to receive attention in recent years mainly from a system point of view. Within the system side, graded relevance (in comparison to the binary one) has been shown to improve ranking functions [87, 88]. Within the user side, recent research supports the idea of categorical thinking [24], suggesting that users divide retrieved results into 3-5 categories based on relevance [79]. However, levels of granularity were decided based on a self-report mechanism, without clear evidence that those levels have different physical manifestations in the brain. In this paper, we aim to provide evidence for different grades of relevance from a neuroscience perspective.

It is crucial to understand what each grade of relevance actually means. The value of evaluating information based on graded relevance has begun to receive attention in recent years both from system [88, 89] and user [5, 33, 90] point of views. This is particularly important since the granularity of relevance assessment in previous studies has been based on investigating this phenomenon indirectly, via some sort of mediator [22, 24]. Therefore, our understanding of how searchers perceive different degrees of information relevance is limited [88]. This paper aims to investigate the neural underpinnings of graded relevance directly. In particular, we focus on discovering and mapping the brain mechanisms of graded relevance, within an IR process performed by humans engaged in a Q/A retrieval task. Examining differential judgement perception and execution from the user's point of view could provide a simple extension to the traditional relevance research [315]. Additionally, understanding graded relevance at the visceral level could lead to a better understanding of automatic graded relevance prediction. The outlined research motivation to investigate graded relevance using a data-driven approach is addressed in Chapter 6.

2.8 Conclusion

This chapter aimed to establish a basic theoretical background for the thesis in light of recent developments. Relevance, a key concept in IR, is an intrinsically difficult phenomenon to capture and understand effectively. A multidisciplinary approach provides significant advantages by introducing novel techniques and methods to study this complex phenomenon. As relevance is a human notion, it is essential to examine the neurological, psychological, and physiological mechanisms involved, which is the main scope of this work. This chapter introduces the NeuraSearch research branch, providing the main motivation for the theoretical and empirical design of the present study. As a result of identified research gaps in this chapter, we argue that:

- Our understanding of cognitive mechanisms that contribute to the formation of relevance is still not complete.
- It is not clear how the user's cognitive states, which provide psychological context, influence the formation of relevance assessment.
- While past studies have shown that brain activity differs significantly during the processing of relevant vs. non-relevant information, so far our understanding of relevance as a graded variable from a cognitive neuroscience standpoint is limited.

The following chapters explain how the research is carried out and introduce the contributions that the work undertaken within this thesis provides.

Chapter 3

Research Methodology

This Chapter outlines detailed settings of the experimental methodology utilised in the user study, investigating neurological processes during the relevance assessment. Section 3.1 focuses on the user study setup. Then, Section 3.2, presents the procedural model of the study and the EEG data acquisition process. Section 3.3 describes how the collected physiological data were pre-processed and statistically analysed. Finally, Section 3.4 reports the results of the Questionnaires used to assess participants' subjective task perception.

3.1 Experimental Setup

3.1.1 Participants

For the purpose of this project, we have recruited a participant sample consisting of forty-two individuals (16 males (38.10%) and 26 females (61.90%)) between 18 to 40 years old and with a mean age of 24.88 and a standard deviation (SD) 5.71 years. All participants were recruited either via the SONA System (the School of Psychological Sciences and Health Participant Pool), social media advertisements or flyers posted at the University of Strathclyde campus using opportunistic sampling. There was no monetary payment associated with the study participation, but eligible participants (enrolled in Psychology or Psychology combined degree at the University of Strathclyde), received two academic credits. Participants reported themselves to be neurologically and physically healthy with normal or corrected-to-normal vision. Over half of the participants were students (61.90%), and the rest were either employed in skilled jobs (33.33%) or unemployed (4.48%). Three participants reported being left-handed. All participants have completed part of their education in English and either had high English proficiency (n = 22) or were native English speakers (n = 20). The majority of non-native speakers were originally Romanian (3) and the rest of the participants were from Germany (2), India (2), Poland (2), Slovakia (2), Spain (2), China (1), Czechia (1), Ghana (1), Greece (1), Hungary (1), Iran (1), Lybia (1), Malta (1) and Syria (1). On average, participants had an experience of 16.83 ($\pm$3.26) years of formal education. Recruited participant sample was not equally balanced based on demographic information (i.e. sex, nationality, age), as these factors were independent of the variables of interest influencing the effects under study, which in turn reduces potential sampling bias [316].

Participants sub-selection

A priori power analysis was performed, which identified that a minimum sample of 27 participants was required to achieve a medium effect size with 0.80 power for each experiment presented in this thesis. Prior to every main experimental analysis, the collected data were visually inspected in order to detect EEG signal disturbances. Participants with a high number of artefacts (caused for example by excessive movement, insufficient number of trials or/and poor contact between the conductive electrode and a scalp) present in the data were manually excluded based on visual inspection of pre-processed data channels and topographic plots of averaged epochs (detailed pre-processing steps are described in Section 3.3.1).

3.1.2 Study Design

This controlled user-centred work followed a within-subject experimental design. Participants were exposed to every level of an independent variable (IV) presented in randomised order to measure neurophysiological changes resulting from exposure to different experimental conditions using EEG. In this experiment, participants engaged in the Q/A task, which is a well-established and frequently employed paradigm to investigate IR phenomena (see e.g. [275]). During the task, participants were presented with a question and either relevant or non-relevant answer, presented word by word. To investigate neural correlates of binary relevance assessments, participants were instructed to make a binary relevance assessment once they gathered enough information to determine the answer relevance in relation to the presented question. The binary relevance assessment had 2 levels: relevant ('rel') and not relevant ('nr'). Next, to investigate the effect of SPK on relevance assessment, participants were asked to provide an assessment of their SPK of the question answer. The SPK assessment had 2 levels: knowledgeable ('know') and not knowledgeable ('notknow'). Lastly, in the graded relevance paradigm, participants were presented with the same question and answer they initially saw and instructed to provide graded relevance judgement at each point of the answer as it was unfolding to them. The graded relevance judgement had 3 levels (non-relevant, low relevant and high relevant). The dependent variable (DV) was the EEG signal. The overview of IVs, DVs and their comparison is presented in Table 3.1.

Table 3.1: The overview of all experimental variables investigated in this thesis.

<table>
<tr><td colspan="6">Experimental Variables</td></tr>
<tr><td rowspan="4">IV</td><td>Binary Relevance</td><td colspan="2">rel</td><td colspan="2">nr</td></tr>
<tr><td rowspan="2">SPK</td><td colspan="2">know</td><td colspan="2">notknow</td></tr>
<tr><td colspan="2">know_rel vs. notknow_rel</td><td colspan="2">know_nr vs. notknow_nr</td></tr>
<tr><td>Graded Relevance</td><td>HIGHR</td><td colspan="2">LOWR</td><td>NONR</td></tr>
<tr><td>DV</td><td colspan="5">EEG Signal</td></tr>
</table>

3.1.3 Stimulus Presentation

The experimental stimuli were presented on a 22-inch colour Mitsubishi Diamond Pro 2040u NF CRT monitor (with a resolution of 2048 x 1536 and refresh rate of 75 Hz) using E-Prime 2.0. Participants were seated approximately 60cm from the computer screen, and response keys were located on a standard QWERTY keyboard. All text events were presented in Arial font, size 16 with grey background to reduce screen luminance. A soft light was used to eliminate external visual distractions.

3.1.4 Questionnaires

Throughout the experiment, participants were asked to fill in the Entry, Pre-Task, Post-Task and Exit Questionnaires. The Pre-Task, Post-Task and Exit Questionnaires were administered using Qualtrics online survey tool, which automatically creates and stores the log file. Participants engaged with the questionnaires using either their personal smartphone or the experimenter's tablet. On the other hand, the Entry questionnaire (designed to gather participants' demographics and medical history information) was administered in a printed format, so that the experimenter could check more efficiently whether participants meet the pre-defined inclusion criteria. A copy of the Entry Questionnaire is available in Appendix C.1. Inclusion criteria included individuals between 18 to 55 years of age, without any pre-existing neurological or psychiatric condition, and not under influence of drugs or medication that might impact the EEG signal recordings. Furthermore, as prior sleep has been shown to significantly affect the EEG signals [317], we required participants to feel rested. There were no selection criteria based on handedness. Prior to participating in the task, participants filled in the Pre-Task questionnaire (see Appendix C.2), with the instruction to self-rate their general knowledge of 6 disciplines (i.e. sport, history, politics, science, medicine, history). After completing the task, participants filled in the Post-Task questionnaire (see Appendix C.3), assessing their task perception. At the end of the experiment, participants completed the Exit Questionnaire (see Appendix C.4), designed to examine the participants' perception of their overall performance.

3.1.5 Q/A Data Set

The data set employed for the studies included in this thesis was initially developed and used by Moshfeghi, Triantafillou, and Pollick [275]. We have chosen this data set as it has been proven effective in investigating IR phenomena from a neuroscience standpoint [275, 276]. The original data set developed by Moshfeghi, Triantafillou, and Pollick [275] contained 138 general knowledge questions with either relevant or non-relevant answers and difficulty assessments. For the main experimental part (described in Section 3.2.4, we kept 55 original questions with the difficulty ratings, which were further expanded by the inclusion of 73 Q/As from TREC-8 and TREC-2001 [1] following an example from the original data set. Additional three questions were selected for the trial part of the experiment (see Appendix D.1). The TREC-8 and TREC-2001 Tracks were selected as they (i) cover a wide range of disciplines, (ii) they are independent of one another, and (iii) they provide a correct answer to the question as well as relevance assessment. We ensured that the selected Q/As were accurate and not time-dependent or ambiguous by manually validating each answer using a search engine. The created data set was then split into two parts (Data Set A (see Appendix D.2) and Data Set B (see Appendix D.3)), each containing 64 questions, answers and relevance assessments in total. The decision of splitting the data set into two balanced parts was made during the pilot study after observing the length of the experiment and to reduce the fatigue of the participants. We ensured that Data Set A and B have similar characteristics (shown in Table 3.2) to avoid introducing any bias in our results. For example, we balanced both data sets to contain apriori 50% relevant and 50% no-relevant answers presented to the participant. Selected questions covered a wide range of topics (such as sport, geography, politics, science, history, medicine etc.) to reduce any potential bias that might occur from the emphasis of one particular topic area.

To address our research questions, we have further adapted and balanced the answer length (long vs. short) [2], answer presentation (grammatically complete sentence vs. snippet), question difficulty (easy vs. difficult)[3] and readability (readable vs. not readable) [4]. This was done to reduce any potential bias that might occur from the emphasis of one particular question/answer type. An example of an easy question presented to the participants was “What is epilepsy?”, which was followed by the short, relevant answer “Epilepsy is a brain disorder characterised by seizures”. The order of

[1] https://trec.nist.gov/data/qamain.html

[2] The answer length was measured by the number of words the answer consisted of. The answer length ranged from 8 to 25 words

[3] To assess the difficulty level, two independent annotators separately judged question difficulty on a two-point scale - i.e. difficult vs. easy. The overall inter-annotator agreement was reasonably high (Cohen’s kappa, $\kappa = 0.72$)

[4] The answer readability was calculated using Flesch Score. Following the approach of Gargoum and Keeffe, answers with Flesch Schore ¿ 60 were classified as readable, whereas answers with Flesch Score ¡ 60 were classified as not readable [318]. Answer readability was calculated using the *textstat_readability* function [319] implemented in RStudio with R 3.6.1

the questions was randomised for each participant. This randomisation ensured that the recorded signals and effects were related to the users' subjective relevance assessment of the presented stimuli, and not related to the stimulus presentation frequency, potentially causing an oddball effect [33]. We ensured that the questions from both data sets were approximately the same length, based on the mean number of characters per question category. We have further modified some words with American English spelling to suit British readers and corrected some minor grammatical mistakes contained within the TREC-8 and TREC-2001 collections. This was done to minimise any potential artefacts present in the neurological data resulting from the presentation of grammatical violations. Furthermore, we have modified the length of some of the answers to fit the answer word limit. Participants were randomly assigned to one of the two data sets.

Table 3.2: The Mean length and of the answer word-count based on category for Data Set A and Data Set B.

	Data Sets			
	A		B	
Answer Length	Mean_A	SD_A	Mean_B	SD_B
Total	14.88	6.24	15.02	6.31
Relevant	15.00	6.24	14.97	6.34
No-Relevant	14.75	6.33	15.06	6.38
Difficult	15.19	6.42	14.69	6.02
Easy	14.56	6.13	15.34	6.67
Long	20.84	2.17	21.09	1.99
Short	8.91	0.89	8.94	0.80
Complete	14.63	6.25	14.94	5.96
Incomplete	15.13	6.31	15.09	6.74
Readable	14.20	6.53	14.57	6.27
Not Readable	15.38	6.10	15.39	6.38

3.1.6 EEG Recordings

The experiment took place at a neuropsychology laboratory at the University of Strathclyde in a semi-darkened room, free of any visual or auditory distractions to avoid interference from deviant neurophysiological activity. Brain signals were acquired using the

128-channel HydroCel Geodesic Sensor Net (Electrical Geodesics Inc., Oregon Eugene, USA) and recorded within the standard EGI package Net Station 4.3.1 software. The location of the 128 electrode placement (E1 to E128) has been shown in Section 2.5, Figure 2.4. A Net Amps 200 amplifier was used for the recording and to facilitate the synchronisation between the behavioural response of the participant and their brain signals. To set the system for recording, we followed Electrical Geodesic Inc guidelines. We aimed to keep the electrode impedances below 50 kΩ, according to the recommended system value. Raw EEG data were recorded at a sampling rate of 1000 Hz and referenced to the vertex electrode (Cz). Before fitting the sensor net over the scalp, the electrodes were soaked in potassium chloride (KCl) electrolyte solution to facilitate conductivity between the skin and electrodes. Each participant's head was measured to determine the correct EEG net size, and the net was positioned using standardised procedures, ensuring that the vertex (located at the central intersection of the sagittal and coronal planes) is halfway between the inion and nasion and halfway between both bilateral preauricular points [320]. The EEG net Cz electrode was positioned at the marked vertex.

3.2 Experimental Procedure

3.2.1 Ethics

Ethical approval (no. 948) was obtained from the Computer and Information Science Ethics Committee at the University of Strathclyde and the experiment was carried out in accordance with the ethical guidelines. The current study did not involve any invasive procedure for data collection. All participants were assigned a unique ID number to ensure participation anonymity. Anonymised data were securely stored on a personal, password-protected database within the university network with authorised access to ensure a high level of data protection. Collected data were processed with strict adherence to the Code of practice and with the General Data Protection Regulations. The Data Management Plan (containing information about data collection, data documentation, metadata, ethics, data storage and data preservation) was submitted as a part of the ethics application.

3.2.2 Procedure Outline

The user study was carried out in the following manner. The formal meeting with the participants took place in the laboratory setting. The investigator was present and available for the entire duration of the experiment in order to monitor the experiment, answer questions and provide clarification if necessary.

At the beginning of the session, all the participants were briefed as to the procedure and the purpose of the experiment through the information sheet (see Appendix A.1 for reference). All participants received the same experimental instructions and the experiment only began once participants fully understood the task and felt confident to perform it. Then, participants were asked to provide informed consent (see Appendix A.2), confirming their willingness to voluntarily participate in the experiment. All participants were notified about their right to withdraw at any time during the study, without giving a reason and without any consequences. After that, they filled in an Entry Questionnaire. Next, participants completed a Pre-Task Questionnaire.

Once the sensor net has been placed on the participant's scalp and impedances were brought to an acceptable level, all the channels were checked for any flat activity. Participants were instructed to avoid eye movements, eye blinks, jaw clenching, muscle tensing and any head movements during the data acquisition. For this purpose, we have provided a short demonstration during which participants were instructed to perform certain movements while they observed the raw EEG activity displayed in real-time on the screen. This process helped participants understand the negative effect of physiological artefacts on the data quality.

Prior to the main experimental trials, participants underwent a number of three training trials, which resembled the main experimental task. Participants were able to repeat the training until they confirmed to have a good understanding of the procedure. The data obtained during the training procedure were not included in the main analysis. After completing the training procedure, participants proceeded to the main experimental task. In total, every participant completed 64 main experimental trials.

To avoid fatigue, the trails were split into two equally long blocks separated by a break. After completing the main experimental task, participants were instructed to fill out the Post-Task and Exit Questionnaires. A debriefing sheet was provided at the end of the experimental session (see Appendix A.3).

3.2.3 Synchronisation of EEG signal and Behavioural responses

The EEG signal and behavioural responses were recorded throughout the entire duration of the main experimental task, during which participants interacted with presented stimuli as their brain activity was monitored. The synchronisation between the real-time raw EEG signal and behavioural responses was maintained via Network Time Protocol (NTP). The NTP is a standard and commonly used mechanism for communicating clock and timing information between devices and software so that events in E-Prime are synchronised with the EEG data collected by EGI's amplifier. E-Prime is, therefore, able to send the event marker with a unique ID and a timestamp representing the event of interest (i.e. stimulus presentation, button press) to the EEG recording system. Each experimental session generated two complementary data files:

- Behavioural data .edat output file containing information about stimuli, stimulus presentation sequence, participant details, participants' reaction times and responses linked to the chronological presentation of every trial.

- Real-time raw EEG signal (i.e. voltage fluctuations measured at each electrode) containing event markers.

3.2.4 Experimental Task

The schematic task representation is depicted in Figure 3.1. At the beginning of the task, participants were presented with instructions. Next, they viewed a randomly selected question from the data set, followed by the fixation cross, which indicated the location of the answer presentation. Eye movements introduce large artefacts to the EEG signal and in standard EEG paradigms trials contaminated by eye movements or blinks are often discarded [321]. Therefore, to control free-viewing and minimise the presence of any confounding artefacts (i.e. saccades), the answer was presented in the middle of the screen word by word. Words were presented to the participant using a rapid serial visual presentation (RSVP) with a duration of 950ms for each word stimulus. The presentation duration of 950ms has been tested and determined by the pilot study outcomes. Furthermore, the duration has been deemed sufficient to model fluent reading and to avoid the overlapping effect of two consecutive words on the ERPs [33]. Furthermore, the engagement of cognitive processes underpinning information processing and decision making during word RSVP is frequently captured within the 100 to 900ms time interval [322–324]. Rather than using a naturalistic reading approach (i.e. self-paced reading [325]), we employed RSVP as it reduces the effect of time differences associated with inter-individual stimulus processing variability. The ERP components were, therefore, time-locked to the word presentation. The RSVP approach has been commonly applied to examine neurological signatures of reading in the ERP studies (e.g. [326]) as well as in the IR paradigms investigating IN [292] and query construction [290].

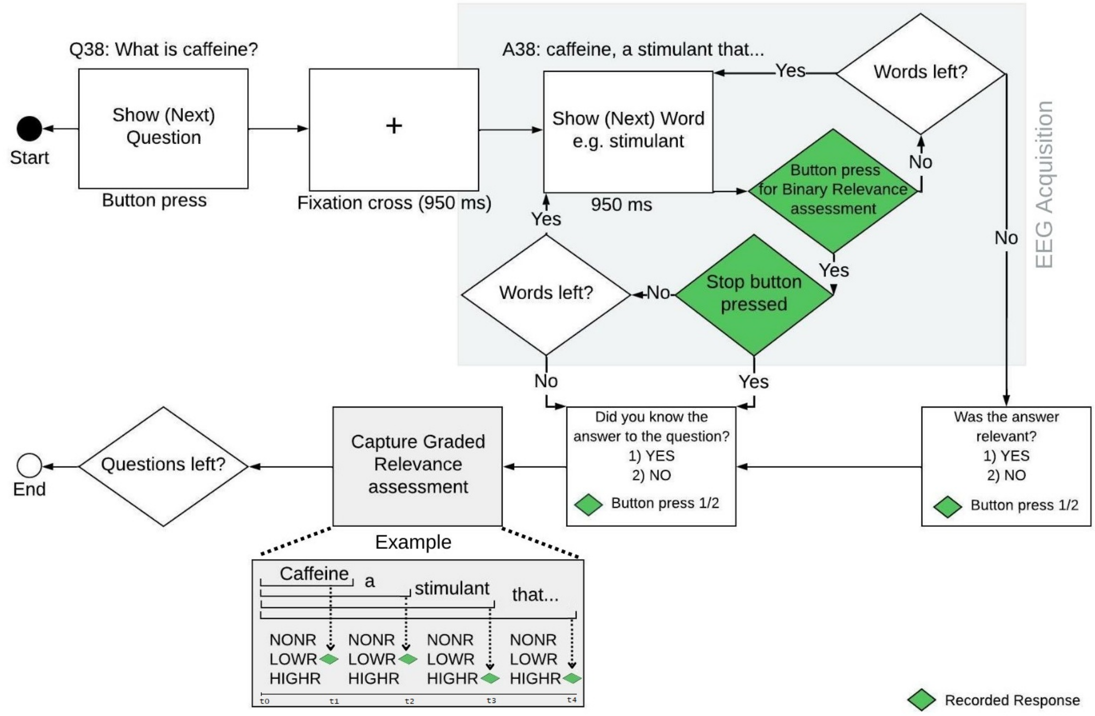

Figure 3.1: The flow diagram of the experimental task.

The figure shows the structure of a task. From the left (START), the question is presented in a randomised order. Once ready, the participant presses a button on the keyboard to start. Firstly, a fixation cross was presented for 950ms. Then, an answer is presented word by word. Each word is shown for 950ms. The participant makes a binary relevance assessment (i.e. 'rel' vs. 'nr') once enough information is gathered and he/she has the option to stop the word presentation once the assessment has been made. If the participant does not make a relevance assessment during the word presentation, they are prompted to do so in the next step, without the time limit. Next, they are instructed to respond to the question assessing their prior knowledge and indicate whether they already knew the answer to the question (i.e. 'know' vs. 'notknow'). After that, participants proceed with the graded relevance judgement ('NONR', 'LOWR', 'HIGHR'), with no time restrictions. Within this step, the participant assesses relevance based on the subjectively perceived information accumulation process. Hence, while the answers were presented word-by-word again, participants were asked to submit a subjective graded relevance assessment for the information segment presented to them from the first stimulus (i.e. word) up to and including the current stimulus. The process is repeated for all 64 questions (END).

Participants were instructed to carefully read individual words that would form either relevant or non-relevant answers. Once participants gathered enough information, they had an option to terminate the word presentation sequence (and to continue to the next step), or to view the sequence in full. In general, we controlled the number of relevant and non-relevant answers presented to the participant, but we did not control the number of words each participant saw. This allowed us to simulate an information search and retrieval, as participants were not required to read through the whole answer. Instead, they were able to terminate the answer presentation once the relevance assessment was made. As brain activity was recorded during the reading, to avoid the possibility of confounding hemispheric effects (due to motor planning or execution), counter-balancing was used, and participants were instructed to interact with the keyboard using either their left or right hand. Participants were then asked whether they already knew the answer to the presented question (i.e. SPK assessment - 'know' vs. 'notknow'). The SPK evaluation was performed after completing each trial (after seeing the answer to the question). This allowed participants to make a more informed judgement about their knowledge state, as opposed to asking participants about their knowledge state prior to seeing the answer. This is because participants may not be completely aware of whether they know the answer as there might be difficulties in distinguishing whether someone actually knows something or is instead simply familiar with it and whether they can recall or only recognise the information they believe to have knowledge of. Additionally, levels of confidence and criterion levels for judgements of this nature can vary across participants [311, 314]. In other words, asking the question after the participant sees the answer, can make participants aware of any anomalies in their knowledge [311]. It is important to note that in general, there is a positive relationship between overclaiming non-existent knowledge and self-perceived knowledge driven by impression-management concerns [327]. Therefore, we were not interested in the participants' knowledge accuracy, but rather in their feelings of certainty and confidence associated with Q/A interaction and processing resulting from self-perceived domain expertise [328].

After that, participants were again presented with the same answer appearing on the screen in the same order (up to the point of presentation abandonment). In this stage of the trial, the answers were presented word by word as continuous text. Participants were instructed to assign graded relevance assessments (NONR, LOWR, and

HIGHR) for each word segment of the answer while taking into account information accumulation, rather than assessing words in isolation with relationship to the question. The graded relevance assessments were then assigned retrospectively, enabling this detailed information to be applied to the corresponding EEG segments of interest. The interpretation of binary and graded relevance assessment categories depended on each participant's subjectively perceived information accumulation process, which enabled capturing the subjective nature of relevance assessment [10].

3.2.5 Pilot Studies

Before commencing the main user study, we performed a pilot study with four participants whose data were not included in the final analysis. The sample of pilot participants consisted of 3 females and 1 male between 23 to 38 years of age, with a mean age of 29.75 years (SD = 6.24). Based on the participants' experience and detailed feedback, we adjusted the study design and presentation by increasing the font size, decreasing screen luminance and clarifying small parts of the instructions. Additionally, we optimised background colour to reduce eye-blinks [329, 330]. After the pilot study, it was determined that the participants were able to complete the user study without problems, including having adequate time to comfortably read and respond to presented stimuli, and that the system was correctly logging participants' behavioural responses and neurological signals.

3.3 Data Pre-processing and Analysis

3.3.1 Pre-processing Steps

The brain activity was recorded from participants as they engaged with relevant and non-relevant content, up to the point where the participant submitted binary relevance judgement. Because the recorded EEG signal is often weak and noisy, raw data typically requires a series of preparation steps. To prepare data for analysis, an automated pre-processing pipeline was built through the adaptation of the EEGLAB tools [331] performed in MATLAB (The MathWorks, Inc., Natick, Massachusetts, US). The EEG data pre-processing steps were based on Makoto's Pre-processing Pipeline[5]. The order of pre-processing steps with brief descriptions is available in table 3.3.1. All collected neurological raw data were first loaded to the EEGLAB and visually inspected. Then a low-pass filter of 30Hz was applied. We down-sampled the data from 1000Hz to 250Hz. Downsampling, a commonly applied procedure, is used to reduce file size for easier data manipulation. Then a high pass filter of 0.3Hz was applied. Filtering is another common procedure used to attenuate frequencies associated with noise rather than a signal of interest. In other words, filtering efficiently increases the signal-to-noise ratio by removing frequency bands of non-neural origin. We then automatically rejected bad channels EEG sensors that were not functioning properly during the data acquisition and that were high in noise throughout the task). The re-referencing to average (across all electrodes) was subsequently performed (to provide an approximation of zero µV for the reference at each time point). The CleanLine EEGLAB plugin was used to filter line noise. All epochs (the time windows of interest) were then extracted from 200ms before stimulus presentation to 950ms afterwards. To detect and remove components associated with ocular, cardiac and muscular artefacts based on their power spectrum and time-course, we performed Independent Component Analysis (ICA) and rejected artefacts using ADJUST [332]. Both ICA and ADJUST are capable of extracting relevant information within the noisy signal, allowing the separation of measured signal into fundamental underlying components [333]. Bad channels were interpolated using a spherical interpolation method. The spherical interpolation method refers to a common data-repair method that computes interpolated data in the bad channels based on the good channel values [334]. Next, we removed the two outermost belts of electrodes of the sensor net. We removed 38 peripheral channels: E1, E8, E14, E17, E21, E25, E32, E38, E43, E44, E48, E49, E56, E57, E63, E64, E68, E69, E73, E74, E81, E82, E88, E89, E94, E95, E99, E100, E107, E113, E114, E119, E120, E121, E125, E126, E127, E128 which are prone to show muscular artefacts, following the approaches of Bian et al. [190] and Calbi et al. [335]. Epochs were then extracted again from 100ms before stimulus presentation to 950ms afterwards based on the stimulus labels for every

[5] `https://sccn.ucsd.edu/wiki/Makoto's_preprocessing_pipeline`

condition of interest as follows:

- Binary Relevance: 'rel' and 'nr'
- Graded Relevance: 'HIGHR', 'LOWR' and 'NONR'
- Self Perceived Knowledge: 'know', 'notknow', 'know_rel', 'notknow_rel', 'know_nr' and 'notknow_nr

We used automatic epoch rejection based on thresholding (i.e. rejecting epochs by detecting outlier values greater than $\pm 100\mu$V due to the signal contamination). All epochs were baseline corrected. After pre-processing the data, epochs of interest were grand averaged.

3.3.2 Statistical Analysis of EEG data

To test for statistically significant differences in the neurological processing associated with the condition of interest, we employed a data-driven approach, which is particularly effective in whole-brain analysis of complex mental phenomena as it minimises the upfront assumptions and allows for the contribution of many distinct areas at different time points [30]. To identify significant cortical differences, we compared the values for 109 electrode pairs at every time point (every 4ms, 237-time points in total) over the 100 - 950ms time window. The initial time interval (0 - 100ms) was excluded from the main analysis as we were not interested in the initial sensory processing of stimulus features [336]. The data-driven approach applied a non-parametric permutation-based paired t-test (1000 permutations) using the *statcond* function implemented in the EEGLAB [331]. The *statcond* function produces a matrix of p values per each time interval of interest and per compared condition pair. Instead of employing formal corrections for multiple comparisons, we have followed the recommendations of Rothman [337] and the approach of Laganaro [338] outlined in Subsection 3.3.3. As a result, the probability of false positives was reduced by employing rigorous clustering methods. In all statistical analyses, a significance level of 0.05 was used. This means that only p values which were less than or equal to 0.05 were considered statistically significant.

Table 3.3: An overview of data pre-processing pipeline steps.

Pre-Processing Step	Description
1. Channel Locations Import	Modification of default channel names which do not correspond to those in the montage.
2. Low-pass Filter	Cutting off frequencies higher than 30Hz.
3. Data Re-sampling	Re-sampling the data to the rate of 250Hz.
4. High-pass Filter	Cutting off frequencies lower than 0.3Hz.
5. Clean_rawdata	Automated identification of bad channels.
6. Re-reference to Average	Creating a common reference for the data using the vertex electrode (Cz).
7. Cleanline	Automatic adaptive estimation and removal of sinusoidal noise.
8. Epoching	Optional step to reduce the data set size and to improve the outcomes of ICA. Epoching at this stage removes noisy periods between the epochs of interest. As a result, specific time windows of interest (-200 to 950ms) are extracted.
9. ICA & ADJUST	ADJUST is based on ICA and used for the identification and removal of artefacted ICA components.
10. Interpolation	Automated transformation of channel values identified through automatic rejection as bad on the basis of channel values identified as good.
11. Electrode Removal	Optional step of removing two outermost belts of electrodes, prone to contain muscular artefacts.
12. Epoching	Extracting specific time windows for every condition of interest (-200 to 950ms) from the continuous EEG signal.
13. Baseline Correction	Automated transformation and correction of the baseline for every epoch segment over the -100ms to 0ms time window.
14. Epochs Auto-rejection	Automated artefact rejection based on thresholding (i.e. the function finds values exceeding threshold values of -100 to 100μV).

3.3.3 Identifying ROIs

As the present study utilises a data-driven approach, for optimal detection of effects, the ROIs were determined based on statistically significant differences between compared conditions of interest. Therefore, we used the features of the data under analysis to position the ROIs. We were not interested in isolated electrodes where a test statistic might happen to be large. Instead, we applied the method utilised by Laganaro and colleagues [338]. To identify potential ROIs, we only considered clusters with at least five electrodes next to each other extending over at least 20ms and retained with an alpha criterion of 0.05.

3.3.4 Identifying ERP Components

Obtained pre-processed signal reflects the average neural activity per condition time-locked to an event of interest, recorded at each of the scalp electrodes. The high temporal resolution of the order of the ms makes EEG an excellent tool for studying the time course of cognitive and neural processes. The paired t-test described in Subsection 3.3.2 was able to identify ROIs and time intervals reflecting significant changes in the neural activity between conditions of interest. Identified significant ROI clusters and time intervals were used to visualise topographic plots and grand averaged ERP waveform for every condition of interest. To determine the ERP component, topographic plots, the ROI distribution and grand-averaged ERP waveforms were visually inspected and compared with existing literature based on their latency, location on the scalp, functional sensitivity and amplitude polarity. A similar approach was used by Kaganovich, Schumaker and Rowland [339] and by Tacikowski, Cygan and Nowicka [340].

3.4 Questionnaire Analysis

This section presents the analysis of the quantitative and qualitative feedback obtained from forty-two participants using questionnaires to understand participants' overall background and subjective experiences. The questionnaire analysis was done prior to the main experimental result analysis with the aim to better understanding factors that might affect participants' engagement with the main experimental task. The questionnaires were distributed to the participants before (Entry and Pre-Task Questionnaires) and after the main experimental task (Post-Task and Exit Questionnaires) and there was no time limit associated with the completion of each questionnaire. The outcomes of the Entry Questionnaire, gathering mainly participants' demographic and background information such as gender or age were reported earlier in Section 3.1.1. The Pre-Task questionnaire was designed to assess participants' use of search engines and general knowledge across different disciplines. The Post-Task Questionnaire was designed to shed light on a participant's specific actions and responses during the task and the Exit Questionnaire allowed participants to assess their own performance. In general, questions included in the Pre-Task and Post-Task Questionnaires were a structured forced-choice type, while Exit Questionnaire included an unstructured open-ended question.

3.4.1 Pre-Task Questionnaire

The main goal of the Pre-Task Questionnaire was to collect information on the participants' engagement with search engines and to understand participants' overall prior knowledge with respect to the topics included in the main experimental task. We have decided to exclude the first question from the analysis as many participants found it ambiguous. The next two questions briefly assessing the frequency at which participants engage with search engines used a 5-point Likert-type agreement response format (answers: 1: "I do not use search engines", 2: "Less than several times a month", 3: "Several times a month", 4: "Several times a week", 5: "Several times a day"). Overall, results suggest that all participants use search engines on average several times a day (M = 4.88, SD = 0.40) and as well as submit their query in the form of a question to answer their IN (M = 4.66, SD = 0.62).

Additionally, we asked participants to what extent they are familiar with the following topics: history, sport, science, geography, medicine and politics using a 5-point Likert response scale (answers: 1: "Not knowledgeable at all", 2: "Slightly knowledgeable", 3: "Moderately knowledgeable", 4: "Very knowledgeable", 5: "Extremely

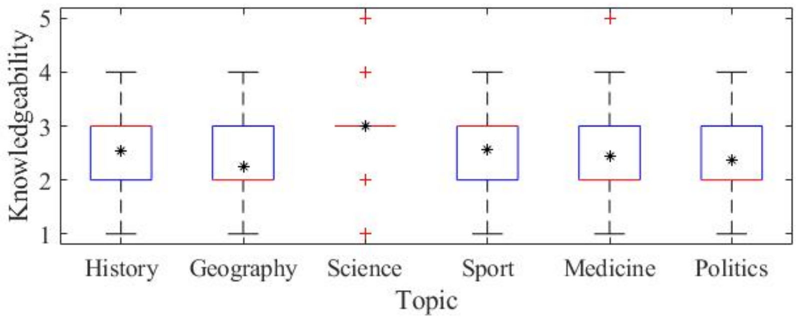

Figure 3.2: Graphical representation of Pre-Task Questionnaire results.
Box plot of the participants' knowledgeability across different disciplines (History, Sport, Science, Geography, Medicine, Politics). The asterisk (*) represents the mean value, while the cross (+) represents the outlier value.

knowledgeable"). The results shown in Figure 3.2 indicate that participants were the most knowledgeable in the area of Science (M = 3.00, SD = 0.83) and least knowledgeable in the area of Sport (M = 2.24, SD = 1.01).

3.4.2 Post-Task Questionnaire

The Post-Task Questionnaire was developed by expanding and adapting questions and answers published in the study of Moshfeghi and Jose [169]. The main aim of the Post-Task Questionnaire was to capture participants' subjective experiences and impressions of performing the main experimental task. Participants rated their perception of the task, presented questions, familiarity with questions and comfort using a 7-point Likert Scale (answers: 1: "Strongly Disagree", 2: "Disagree", 3: "Somewhat Disagree", 4: "Neither Agree nor Disagree", 5: "Somewhat Agree", 6: "Agree", 7: "Strongly Agree").

The results shown in Figure 3.3 indicate that participants found the task (M = 6.13, SD = 1.18), questions (M = 5.97, SD = 0.90) and selected question topics (M = 5.95, SD = 1.04) rather interesting. Perceived difficulty of the task (M = 4.35, SD = 1.64), questions (M = 4.18, SD = 1.58) and selected question topics (M = 4.30, SD = 1.59) was rated as moderate. Presented questions (M = 5.95, SD = 1.18) and task in general (M = 6.00, SD = 1.11) were overall considered as readable. Additionally, both, questions (M = 5.88, SD = 1.26) and task (M = 5.78, SD = 1.12) were also considered

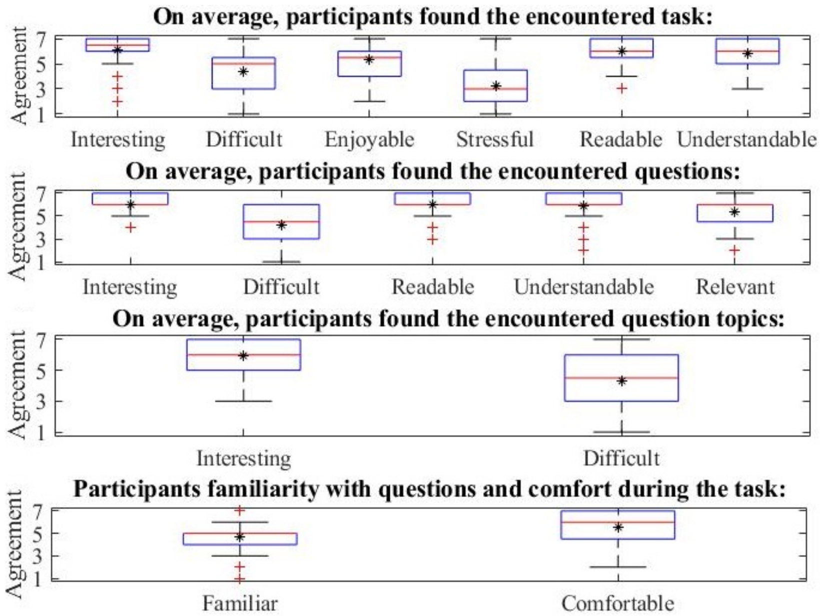

Figure 3.3: Graphical representation of Post-Task Questionnaire results.
Box plot of the participants' task-related perceptions. The asterisk (*) represents the mean value, while the cross (+) represents the outlier value.

as understandable. Overall, participants indicated that they have somewhat enjoyed the task (M = 5.33, SD = 1.35). On average, participants felt moderate physical comfort (M = 5.53, SD = 1.40) and task was not rated as too stressful (M = 3.28, SD = 1.78). Questions selected for the experiment were perceived by participants as moderately familiar (M = 4.70, SD = 1.32) and relevant to them (M = 5.35, SD = 1.39). In general, the results of the Post-Task Questionnaire indicate that participants did not perceive any difficulties with the experimental design that might have made caused them discomfort and impacted their engagement.

3.4.3 Exit Questionnaire

Using the Exit Questionnaire, we examined participants' experiment-related impressions (Part A), perceptions of their own overall performance (Part B), as well as general comments for the user study. To record participants' responses, we used a 7-point Likert Scale (answers: 1: "Strongly Disagree", 2: "Disagree", 3: "Somewhat Disagree", 4: "Neither Agree nor Disagree", 5: "Somewhat Agree", 6: "Agree", 7: "Strongly

Agree") and one open-ended question, which gave participants an opportunity to provide a qualitative description of their experience during the experiment.

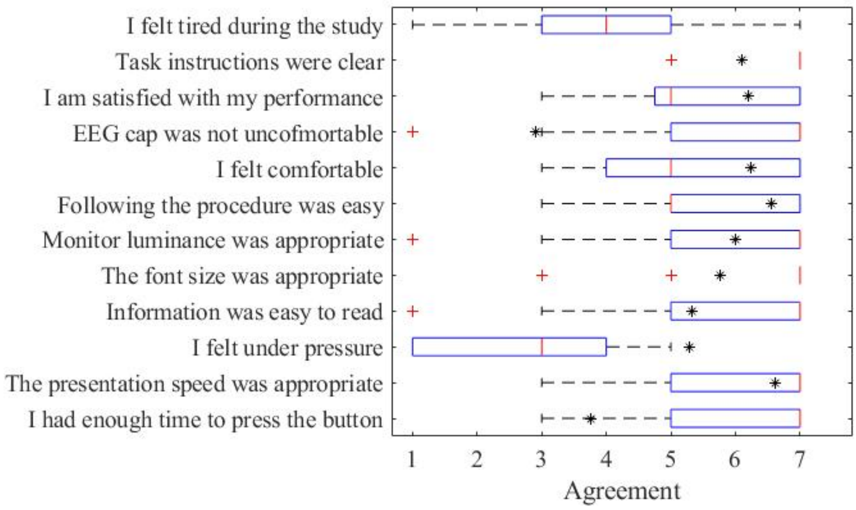

Figure 3.4: Graphical representation of Exit Questionnaire results.
Box plot of the participants' experiment-related impressions and their performance. The asterisk (*) represents the mean value, while the cross (+) represents the outlier value.

Overall, the results (presented in Figure 3.4) indicate that participants felt that they had enough time to press a button to terminate the answer presentation (M = 6.10, SD = 1.28). Additionally, they found the speed of the word presentation (M = 6.20, SD = 1.12) to be appropriate for reading. The font size (M = 6.56, SD = 0.95) and monitor luminance (M = 6.00, SD = 1.58) were also rated to be task appropriate. Most of the participants felt moderately comfortable (M = 5.32, SD = 1.65) during the task and the EEG cap was not causing them significant discomfort (M = 6.00, SD = 1.50). Participants found following the procedure to be somewhat easy (M = 5.76, SD = 1.26), with easy to read (M = 6.24, SD = 1.36) information and clear instructions (M = 6.61, SD = 0.80) and they were somewhat satisfied with their performance (M = 5.29, SD = 1.33). Participants indicated that they have answered (M = 6.80, SD = 0.60) all the questions honestly (M = 5.27, SD = 1.55), somewhat correctly and to their best abilities (M = 6.59, SD = 0.87). On average, participants did not feel under significant pressure (M = 2.90, SD = 1.53) and they did not feel tired (M = 3.76, SD =

1.83). The majority of participants rated their effort (37) and their performance (31) to be constant throughout the study. The rest of the participants felt that their effort (3) and performance (6) was greater during the first half of the main experimental task (i.e. before the break). Out of 42 participants, 13 participants took the opportunity to provide qualitative feedback, which opened up the opportunity to gain a better insight into the participant's experience. The majority of qualitative responses (8) were rather positive. Participants found the experiment interesting (3), enjoyable (3), intellectually stimulating (1), fun (1), amazing (1) and they enjoyed having a new experience (1). In terms of negative feedback, participants found the task slightly uncomfortable (2) and tiring (1) and would appreciate more breaks (1). One participant would prefer a more narrow definition of relevance categories (however, our aim was to investigate subjectively perceived relevance). One participant made a suggestion to space relevance response buttons further apart.

3.5 Chapter Summary

The current Chapter has provided a detailed overview of the methodological and analytical framework used in the current thesis. In particular, the chapter outlines:

- Considerations and requirements behind the study design and implementation.
- Participant recruitment, participant exclusion.
- Study design, explaining variables of interest used in the thesis to address our Research Goals outlined in Section 2.7.
- Overview of the procedural steps necessary to conduct the experimental study as well as ethical considerations including the informed consent process and data privacy measures.
- Detailed steps involved in the EEG data pre-processing and preparation for the statistical analysis.
- Detailed steps involved in statistical analysis of obtained neurological signals.

- Questionnaire results describing participants' subjective perceptions of their task engagement.

Part II

Empirical Contributions

Since relevance is a complex mental phenomenon, it is essential to also consider the underlying perceptual and cognitive processes. To do so, as mentioned in the thesis motivation (Section 2.7), this work aims to gather behavioural and neuro-physiological EEG data in order to better understand participant's experience during relevance judgement tasks. Chapter 4 investigates neurological phenomena that underpin binary relevance assessment using a data-driven approach. Chapter 5 considers contextual aspects of relevance assessment formation, namely participants' SPK. Chapter 6 examines relevance as a graded variable and explore potential neurological and cognitive differences associated with each grade of relevance. The approaches used in the empirical part were designed to gain a better understanding of this complex construct and provide empirical support for relevance theories, but also existing studies.

Chapter 4

The Cortical Activity of Binary Relevance

This Chapter describes the EEG investigation of binary relevance assessment and provides the means to address the first thesis motivation outlined in Section 2.7.1. The main contribution of this chapter is the use of a data-driven approach, which has not been previously used to investigate textual binary relevance assessment. In comparison to traditionally used methods (e.g. theory-driven ERP analysis), the data-driven approach used in this study highlights important and previously not reported neurological differences associated with the processing of relevant vs. not relevant content. The chapter is organised as follows. Section 4.1 presents theoretical background evaluating relevant literature and explaining the assumptions that guide the present study. The Chapter then continues with the outline of the experimental setup (Section 4.2), Section 4.3 where the results are presented and discussed before conclusions are drawn in Section 4.4.

4.1 Background

After Roccio's introduction of the binary relevance feedback in 1965, relevance became a central active concept, vital for the functioning of the IR systems [75, 97]. Up to this date, binary relevance remains the most common standardised evaluation method of text documents [341], as well as the most used methodological tool in experimental research [302]. Traditionally, binary relevance evaluation reflects the relationship between the content of a retrieved document and the user's IN expressed as a query [342] (i.e. the content is either classified as relevant or non-relevant).

Past user-centred relevance assessment experiments rely on the user's subjective notion of relevance, which makes users an integral part of the IR. As users' subjective perceptions can be challenging to collect using traditional IR explicit feedback methods, novel research started to focus on the application of brain imaging, such as the EEG. Past EEG studies have found that processing of relevant vs. non-relevant information elicits significantly different neural responses. However, these studies have mainly utilised component-driven approaches to investigate binary relevance phenomena.

The component-driven approach to researching relevance assessment offers invaluable insights, but it is only partially capable of identifying and quantifying any previously unknown and unreported ERP components that could arise from relevance phenomena. As a result, a component-driven analysis may overlook key cognitive aspects that contribute significantly to unexplored and complex cognitive processes during relevance assessment formation. To minimise the potential analytical bias induced by previous research's restriction to particular ERPs, this study adopts a data-driven methodology instead [30].

This chapter aims to study the relevance assessment phenomena using a data-driven approach. The current experiment was carried out to investigate binary relevance assessment using a data-driven approach to gain a better understanding of complex cognitive processes underpinning relevance assessment formation. In particular, the user's neural signals associated with relevant and not relevant relevance assessments were recorded in real-time during a Q/A relevance assessment task. Using a data-driven approach we compared information assessed as relevant ('rel') to that which was non-relevant ('nr').

4.2 Experimental Setup

In the first study, we aim to explore aspects of binary relevance assessment through the examination of the user's physiological and behavioural signals. These signals are obtained through naturalistic tasks designed for this purpose (described in Section 3.2.4), placing relevance assessment within the context of the IR process and aiming to incorporate all its aspects, such as user's IN. The study is built on the previous literature investigating relevance assessment through the comparison of signals associated with relevant and non-relevant information.

4.2.1 Participants

Twelve participants were excluded from the final study analysis due to the high number of physiological artefacts present in the EEG data. The 30 remaining participants (18 females and 12 males) were between 19 to 40 years old and with a mean age of 24.53 and a SD of 5.74 years. Fourteen participants were randomly assigned to Data Set A and sixteen participants were assigned to Data Set B.

4.2.2 Data Preparation

By design, participants were instructed to make explicit binary relevance assessments once they acquired enough information to determine answer relevance. Post-relevance assessment events were not considered for further analysis as it is not the scope of the thesis. The IV was the user's binary relevance assessment (with two levels: "Non-Relevant" ('nr') and "Relevant" ('rel')). The DV was the EEG signal gathered during the word-by-word answer presentation.

During data pre-processing we have removed on average 17.17 ($\pm$9.82) bad channels and ADJUST (based on ICA) isolated and removed a mean number of 16.33 ($\pm$10.03) component artifacts. After the data pre-processing, 330.03 (50.57%) of accepted trials were marked as 'rel' and 322.63 (49.43%) as 'nr' (descriptive statistics are displayed in Table 4.1).

4.2.3 Data Analysis

We used a data-driven approach (described in Section 3.3.2) to test whether there are statistically significant differences in the neurological processing associated with the judgement of 'rel' vs. 'nr' information, which is particularly effective in whole-brain analysis of complex mental phenomena because it minimises upfront assumptions and allows for the contribution of many distinct areas at different time points. [343, 344].

Table 4.1: The Mean number and SD of accepted and rejected epochs for 'rel' and 'nr' condition.

	Accepted Epochs		Rejected Epochs	
Condition	Mean	SD	Mean	SD
rel	330.03	130.27	75.87	86.43
nr	322.63	112.14	72.53	64.97

4.2.4 Statistical Analysis of Button Responses

Participants' behavioural responses were compared to the relevance assessment set provided with the TREC-8 and TREC-2001. In general. The overall accuracy of the participants' binary judgements was 81.09%. Although participants' accuracy indicates how well the task was performed, we were primarily interested in the subjectively perceived relevance of each answer. Hence, these results were excluded from further analysis.

On average, participants were presented with 801.07 words (±168.49) and the main experimental task lasted approximately 49.68 minutes (+/- 8.43). Overall, the mean button-press response time for binary relevance assessment from the point that the stimulus was presented was 501.14ms (±259.48ms). In particular, the average response time for relevant judgement from stimulus onset was 500.78ms (±259.26ms) and for non-relevant judgement, it was 500.41ms (±259.04ms)

4.3 Results

100 - 200ms. The first observed a statistically significant difference in neural activity between the 'rel' and 'nr' conditions emerged in the 100 - 200ms interval. When compared to the 'nr' condition, the 'rel' condition was associated with significantly greater positivity in the right posterio-temporal region and significantly greater negativity in the left fronto-temporal region. Figure 4.1 (row I) shows significant electrode clusters (5 electrodes: E90, E91, E92, E98, E101), time intervals, and ERP waveforms, as well as topographic plots. Given the topographies and waveform peaks, the differences are most likely due to P100 ERP component variability (similar distributions reported, e.g. by [345]). The P100 ERP component is a positive waveform reflecting initial visual field activation, and increased P100 amplitude observed during relevant information processing could indicate early selective attention allocation, with more early attention allocated to relevant stimuli [204]. This early P100 selective stimulus encoding may have an impact on later ERP components associated with working-memory [346], such as the LPC which has been commonly reported in relevance assessment studies [206].

450 - 600ms. The processing of 'nr' content versus 'rel' content was associated with lower amplitude in an electrode cluster (16 electrodes: E6, E7, E13, E55, E78, E79, E80, E85, E86, E87, E92, E93, E98, E103, E104, E106) that bridged the right centro-parietal negativity within the 450 - 600ms time interval, as shown in 4.1, row II. The significant differences were related to the increased centro-parietal negativity associated with the 'nr' condition versus the 'rel' condition. Previous studies have found that observed anterior negativity and co-occurring posterior positivity reflect the N400 ERP component, with similar topographic distributions [347,348]. Lower N400 amplitudes in response to 'nr' content may be associated with a higher semantic incongruity between the question context and the provided answer. The fact that 'nr' content is associated with less positivity during this time period is consistent with the findings of e.g. [205], who discovered that irrelevant content produced more negative N400 responses.

600 - 750ms. It is possible to trace the significant differences in ERP positivity recorded over the right centro-posterio-temporal cluster (7 electrodes: E78, E79, E85, E86, E92, E93, E98) shown in the topographic plots between 600 and 750ms time frame (see Figure 4.1, row III) to the LPC[1] (e.g. [33, 101]). The LPC component is a positive-going deflection that appears 600ms after the stimulus and is typically strongest over the posterio-medial brain areas [349]. In comparison to the processing of the 'nr' content, the positivity was noticeably higher for 'rel' content. Higher LPC amplitudes are connected to decision-making and memory processing, which may indicate an effort to retain pertinent information during cumulative information exposure [205, 350]. In addition, larger LPC amplitude deflection has been observed in response to task-related stimuli in the past studies [351], which is consistent with our results.

4.4 Conclusion

The data-driven approach was found to be an effective tool to explore novel, previously not reported neurological signatures associated with the subjective perception of binary relevance assessment. Our findings add to our understanding of the concept of relevance and provide evidence to support its theoretical foundations. Overall, the study supports the empirical findings of previous studies that investigated textual relevance processing associated with the N400 and LPC components. Furthermore, the data-driven approach revealed previously not reported neural differences in an early P100 component, which provide novel insight into cognitive mechanisms that contribute to the formation of relevance assessment. Attention-related P100 might trigger early stimulus processing, such as input registration and classification. Finally, we believe our findings are a significant step toward understanding the nature of relevance assessment in terms of electrophysiological modulation and operationalising it for the IR process.

[1]LPC and P600 are frequently used interchangeably. The P600 ERP component has frequently been related to relevance assessment (e.g. [33]). However, in language studies, the P600 component is mostly related to "syntactic re-analyses." As a result, given that the LPC has been connected to memory and recognition processes, the term LPC might be more appropriate to utilise while concentrating on relevance assessment.

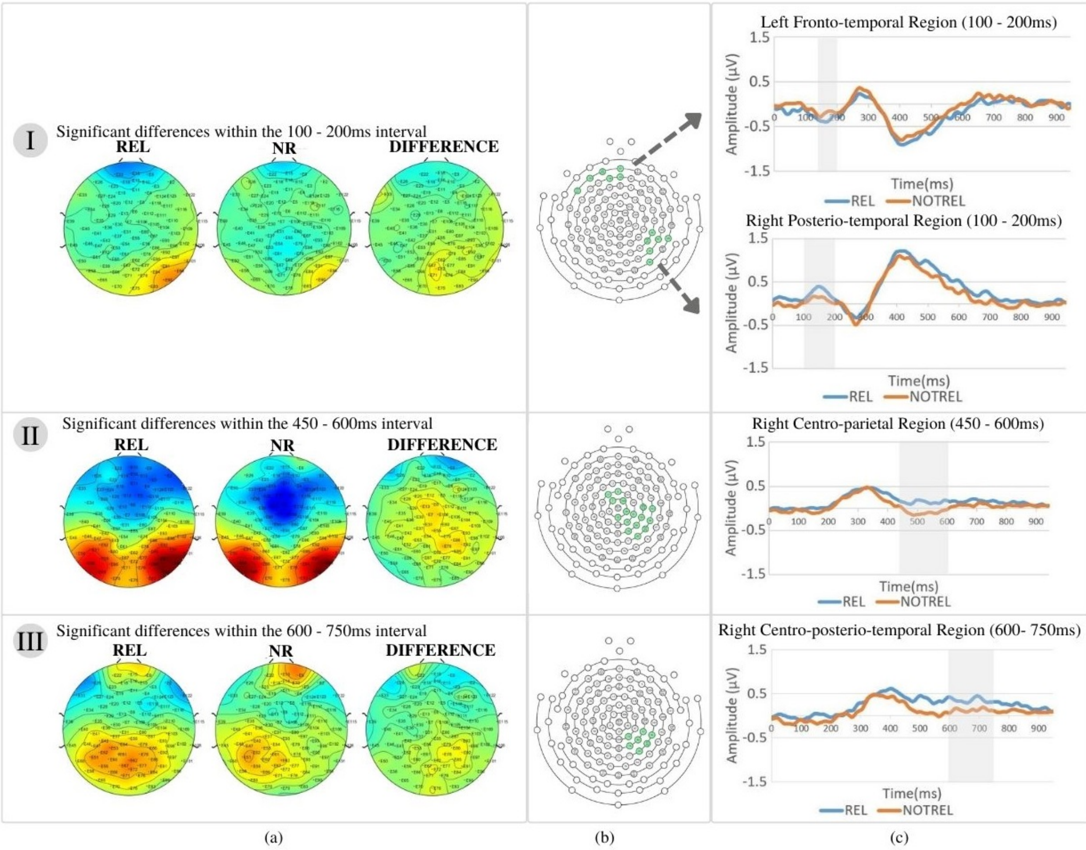

Figure 4.1: (a) Topographic plots for 'rel' vs. 'nr' conditions, including a mean difference plot for the 100 - 200ms (I), 450 - 600ms (II) and 600 - 750msms (III) time windows. Reddish colours of the scalp topography indicate positive ERP values, whereas bluish colours indicate negative ERP values. (b) The 128-channel net graph with highlighted statistically significant electrode sites for each significant time interval. (c) The comparison of grand averaged ERP waveforms for 'rel' (blue) vs. 'nr' (orange) condition. Significant time intervals are highlighted in grey for each significant time period.

4.5 Chapter Summary

This chapter revisited neural aspects of binary relevance assessment. The findings of this study are consistent with previous literature examining users' subjective binary relevance perception. Additionally, we have observed significant differences in neural activity related to the P100 ERP component, related to early visual processing, which has not been reported by previous studies.

Chapter 5

Self-perceived Knowledge in a Relevance Assessment Task

The previous chapter presented findings from the data-driven binary relevance assessment analysis and outlined distinct cognitive processes that contribute to the binary relevance assessment formation. As relevance assessment is significantly influenced by the user's internal context, this chapter will examine the effect of context (namely the SPK variability) on binary relevance assessment formation using a neuroscientific approach.

5.1 Background

The SPK plays an integral role in shaping cognition and influencing decision-making within information processing [312] as it impacts subjectively perceived information importance value [352]. Despite the potential construct importance, SPK has not been investigated within the context of relevance assessment, and past studies have predominantly focused on topical knowledge [305]. However, the users' SPK is a better predictor of information-interaction behaviour than their actual topical knowledge [312] as users' are often unable to accurately assess their actual knowledge [311]. The main aim of this work is to investigate complex cognitive processes that underpin SPK within the relevance assessment context from a neuroscience perspective.

Taking the neuroscience approach offers the unique possibility of exploring dynamic and complex SPK states while addressing measurement and subjectivity challenges frequently associated with relevance [90].

This is the first study investigating user's SPK as a contextual aspect of relevance assessment during real-time information processing employing electrophysiological measurement. We capture the user's SPK, binary relevance assessments and associated brain activity in relation to the Q/A task. The data-driven approach employed in this study provides the benefit of avoiding potential analytical bias introduced by the restriction to distinct ERPs [30]. Understanding brain activity associated with the user's cognitive states related to SPK could lead to innovative IR techniques improving retrieval performance and satisfying searchers' needs more effectively through the adaptation to individual differences.

5.2 Experimental Setup

The main aim of this study was to explore the effect of user's SPK on the binary relevance assessment formation from a neuro-cognitive perspective. Participants' behavioural and neurological signals were recorded as they engaged in the IR task (described in Section 3.2.4). We compared the signal associated with distinct SPK states for each binary relevance assessment (i.e. 'rel' vs. 'nr').

5.2.1 Participants

Out of a total number of forty-two recruited participants, we have excluded nineteen due to the high amount of physiological artefacts present in their data, which was in some cases related to the low number of trials. The low number of trials was caused through further data sub-sectioning given the variables of interest. The study included a sample of 23 participants, out of which 9 were males (39.13%) and 14 (60.87%) were females with a mean age of 24.57 years (SD = 5.84). As the sample size is smaller than required by power analysis it is important to keep in mind lower statistical power (75% instead of pre-defined 80%). Data Set A was randomly assigned to 12 participants, whereas Data

Set B was assigned to 11 participants. On average, participants were presented with 810.17 words (±145.76) and the main experimental task lasted approximately 49.77 minutes (±9.13).

5.2.2 Data Preparation

In this study, participants were instructed to provide explicit binary relevance assessments and their perceived SPK states for every Q/A trial. The IVs were user's SPK states (with two levels: "Knowledgeable" ('know'), "Not Knowledgeable" ('notknow')), and relevance assessments (with two levels: "Non-Relevant" ('nr') and "Relevant" ('rel')). The DV was the EEG signal gathered during the Q/A task. Following the approach of prior studies, the SPK assessment, as a metacognitive evaluation, was performed after completing each trial [311, 353, 354]. This is because participants may not be completely aware of whether they know the answer[1] [311, 314]. After the data pre-processing (described in Section 3.3.1) a mean number of 17.61 (±11.18) components were removed and on average 15.87 bad channels (±10.13) were automatically identified for removal. The mean number and SDs of accepted vs. rejected epochs for every condition of interest in this study are presented in Table 5.1.

Table 5.1: The Mean number and SD of accepted and rejected epochs for every SPK condition of interest within binary relevance assessment context.

Condition		Rejected Epochs		Accepted Epochs	
		Mean	SD	Mean	SD
SPK					
know		58.70	61.02	260.04	118.34
notknow		103.65	112.42	387.78	126.19
know	rel	31.83	33.80	148.43	89.99
	nr	26.87	29.50	111.61	57.57
notknow	rel	50.65	66.84	179.17	78.91
	nr	53.00	49.39	208.61	80.31

[1]There are difficulties in distinguishing whether someone actually knows something or is instead simply familiar with it, or whether they can recall or only recognise the information they believe in having the knowledge of. Also, confidence and criterion levels for assessments of this nature can vary across individuals [311, 314].

5.3 Results

Effects of SPK. The data-driven comparison of 'know' and 'notknow' conditions (irrespective of relevance assessment) revealed no statistically significant differences in brain activity. On the other hand, the comparisons of 'know_rel' vs. 'notknow_rel' and 'know_nr' vs. 'notknow_nr' conditions were associated with significant brain signal differences within multiple time intervals and with wide scalp distributions. The SPK, therefore, has an effect on relevance assessment and can modulate this process at the neural level.

Non-Relevant Assessments: There were no significant differences between the 'know' vs. 'notknow' conditions for non-relevant information in early 100 - 350ms time interval.

The comparison of 'know_nr' and 'notknow_nr' conditions revealed significant differences in the right centro-parietal cluster within the 350 - 450ms time interval (cluster of 9 electrodes: E55, E78, E79, E80, E85, E86, E87, E93, E104), as displayed in Figure 5.3, row I. Similar activity patterns are observed in a later time interval of 500 - 550ms with a more posterior cluster of 5 electrodes (E85, E90, E91, E96, E97) (Figure 5.3, row II). The significant differences were driven by the higher centro-parietal positivity associated with 'know_nr' compared to the 'notknow_nr' condition. Centro-posterior positivity of this nature has been shown to co-occur with the N400 [347]. If interpreting this difference in the N400 context, the greater positivity may indicate that SPK attenuates semantic incongruity (e.g. perhaps through a process where SPK informs the participant that the information is not relevant, and they, therefore, do not focus as intently on the relationship and the incongruence between the answer and question, as someone who does not have SPK).

The 'know_nr' condition was associated with a significantly greater right posterio-temporal positivity compared to the 'notknow_nr' condition. Topographic plots and ERP waveforms (displayed in Figure 5.3, row III and IV) indicate a transition between two components as observed bilateral positivity (row III) is becoming more centralised (row IV). Given the topographies and waveform distributions within the 600 - 750ms time interval, the differences are likely to reflect the transition from N400 ERP component (row III, a significant cluster of 7 electrodes: E78, E85, E90, E91, E92, E96, E97) the LPC component (row IV, a significant cluster of 8 electrodes: E84, E90, E91, E92, E96, E97, E98, E101). Both N400 and LPC components are related and sensitive to semantic violations. The LPC component is a positive-going deflection, emerging around 600ms post-stimulus usually largest over the medial posterior brain areas [355, 356]. The LPC amplitudes are proportional to the effort invested in working memory maintenance. Similar to the N400 response, but with a different polarity, more positive LPC amplitudes are elicited in response to incongruent stimuli [357]. Visual inspections of scalp topographies within 700 - 750ms time interval suggests that the small significant cluster of 6 electrodes (E3, E4, E11, E15, E16, E18) in the frontal region could be likely attributed to the noise associated with excessive eye-movements. Therefore, the differences in the frontal region are excluded from further interpretation.

Relevant Assessments: The earliest neural activity differences for information assessed as relevant emerged within the 250 - 350ms time interval over the left fronto-centro-temporal region (significant cluster of 6 electrodes: E27, E28, E29, E35, E40, E41). The 'know_rel' condition was associated with a greater positivity compared to the notknow_rel' condition. Significant electrode clusters, time intervals and ERP waveforms, as well as topographic plots, are displayed in Figure 5.3, row I. Given the topographies and waveform peaks at around 300ms post-stimulus, the differences are likely to reflect variability in the P300/ centro-parietal positivity (CPP) (similar distributions are reported, e.g. by [358, 359]). The CPP is commensurate with the P300 family of ERPs and similarly to P300, it increases proportionally with the strength of the exogenously presented evidence (i.e. stimulus intensity) over time [360]. The higher amplitude observed when people indicated to have SPK might suggest processing ease associated with reduced cognitive load (see e.g. [361]).

Next, the comparison of 'know_rel' and 'notknow_rel' conditions was associated with statistically significant differences in the right posterio-temporal regions, which reflected greater positivity in the no knowledge condition. Significant electrode clusters (8 electrodes: E75, E78, E83, E84, E85, E90, E91, E97), time intervals and ERP

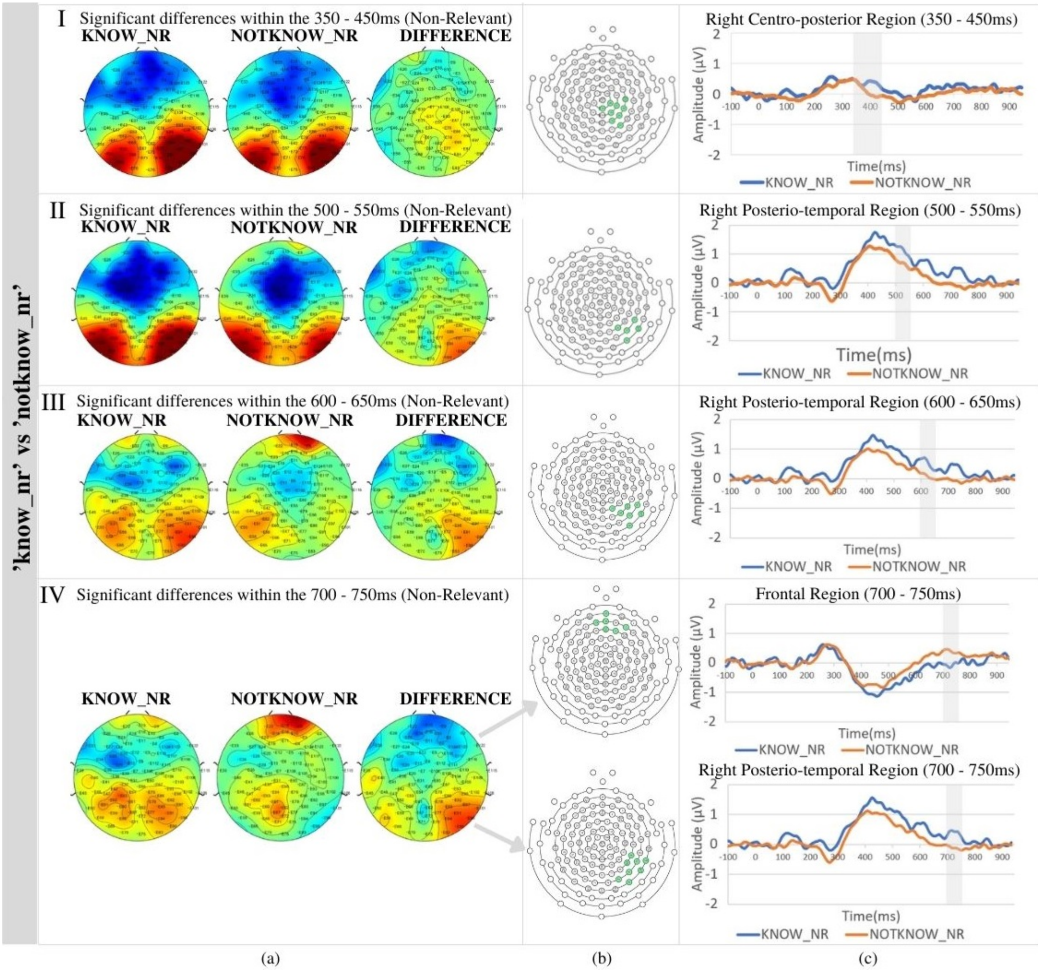

Figure 5.1: (a) Topographic plots for 'know_nr' vs. 'notknow_nr' conditions, including a mean difference plot for the 350 - 450ms (I), 500 - 550ms (II), 600 - 650ms (III), and 700 - 750ms (IV) time windows. Reddish colours of the scalp topography indicate positive ERP values, whereas bluish colours indicate negative ERP values. (b) The 128-channel net graph with highlighted statistically significant electrode sites for each significant time interval. (c) The comparison of grand averaged ERP waveforms for 'know_nr' (blue) vs. 'notknow_nr' (orange) condition. Significant time intervals are highlighted in grey.

waveforms, as well as topographic plots, are displayed in Figure 5.3, row II. Observed significance might be related to the N400 ERP component, as N400 scalp distribution shows anterior negativity and a posterior positivity [362].

The comparison of 'know_rel' and 'notknow_rel' conditions revealed significant differences in the bilateral fronto-central region within the 350 - 450ms time interval (see Figure 5.3, row V). Greater frontal negativity was observed in the cluster of 6 electrodes (E5, E11, E12, E13, E19, E20) for 'notknow_rel' compared to the 'know_rel' condition. The negativity reflects the N400 component, which has been previously described [347, 348, 363]. The decreased N400 amplitude, when judging information to be relevant and aligned with the question, appears to indicate that SPK helps to decrease semantic incongruity and to integrate the words into context [364].

Significant differences between 'know_rel' vs. 'notknow_rel' conditions were observed within the 600 to 700ms time-window over the right centro-posterior region of 6 electrodes (E78, E84, E85, E86, E90, E91; see Figure 5.3, row VI). The differences were associated with higher positive-going ERP amplitudes associated with the processing of 'notknow_rel' compared to 'know_rel' information. The topographic distribution with a characteristic posterior positivity can be attributed to the LPC component. Greater LPC amplitudes have been associated with information accumulation and decision-making processes [365]. Additionally, the LPC reflects the information learning process [366] through codification and strengthening of episodic memory [367]. Greater posterior positivity across 'notknow_rel' compared to the 'know_rel' condition might therefore reflect the enhanced episodic memory activation, enabling the lexico-semantic facilitation of learning novel information.

5.4 Conclusion

The data-driven comparison of 'know' and 'notknow' conditions (irrespective of relevance assessment) revealed no statistically significant differences in brain activity. On the other hand, the comparisons of 'know_rel' vs. 'notknow_rel' and 'know_nr' vs. 'notknow_nr' conditions were associated with significant brain signal differences within multiple time intervals and with wide scalp distributions. SPK, therefore, has an effect

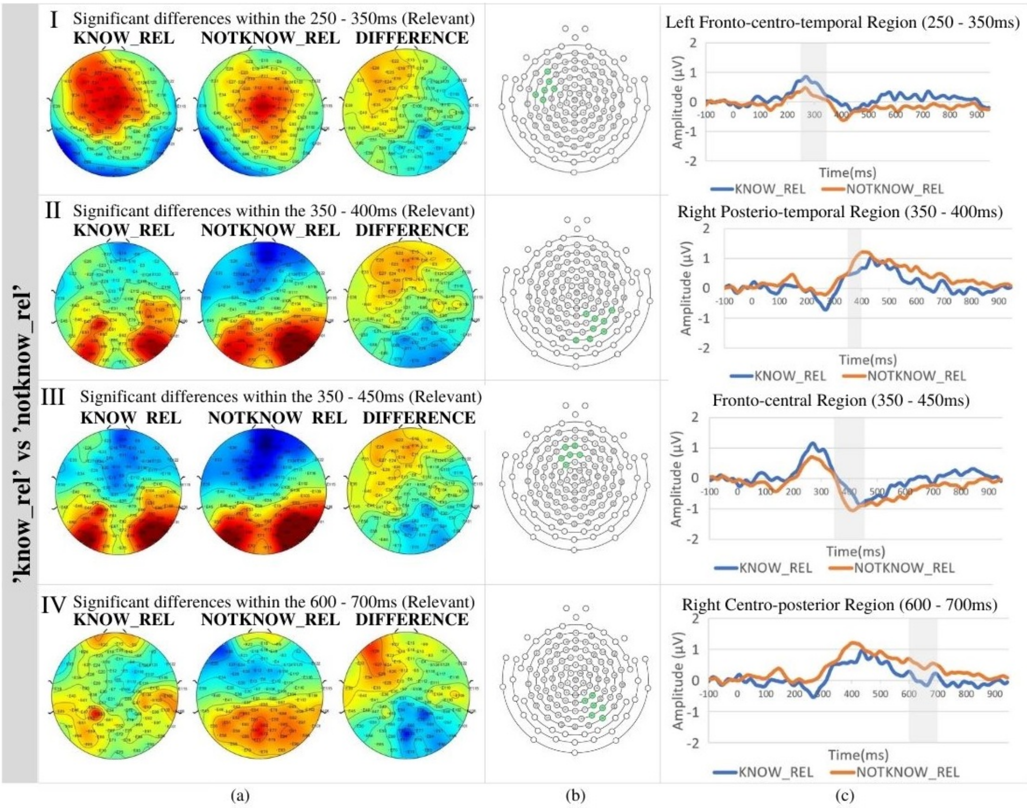

Figure 5.2: (a) Topographic plots for 'know_rel' vs. 'notknow_rel' conditions, including a mean difference plot for the 250 - 350ms (I), 350 - 400ms (II), 350 - 450ms (III), and 600 - 700ms (IV) time windows. Reddish colours of the scalp topography indicate positive ERP values, whereas bluish colours indicate negative ERP values. (b) The 128-channel net graph with highlighted statistically significant electrode sites for each significant time interval. (c) The comparison of grand averaged ERP waveforms for 'know_rel' (blue) vs. 'notknow_rel' (orange) condition. Significant time intervals are highlighted in grey.

on binary relevance assessment and can modulate this process at the neural level. The current research results establish an important step towards understanding the distinct cognitive and neural mechanisms involved in relevance assessments within the context of SPK. As the first attempt to study relevance assessment using brain imaging in the context of SPK, our work opens up an array of interesting future directions and future studies might want to consider the importance and implications of user's SPK.

5.5 Chapter Summary

The current chapter investigated the contextual aspects of relevance assessment formation. The main aim was to gain a better understanding of neural correlates associated with user's SPK states during binary relevance assessment. The results of this research highlight the importance of contextual aspects affecting relevance assessment formation.

Chapter 6

The Cortical Activity of Graded Relevance

The current chapter presents findings from the final part of the empirical research. Previous chapters described neurophysiological signatures of binary relevance alone and within the context of user's SPK. However, in the field of IR, debates surrounding the concept of relevance granularity are still ongoing. Therefore, this chapter aims to investigate whether the processing of distinct relevance grades is also associated with neural differences in the brain. To do so, we employ a data-driven approach, which has been proven effective in discovering novel, previously non-reported neurophysiological phenomena. This would allow us to better understand potential differences and relationships between each relevance grade.

6.1 Background

While a binary relevance division is prevalent in IR, seminal theories have proposed relevance as a graded variable; i.e. having different degrees. Empirical studies investigating relevance as a graded construct mainly focused on explicit user ratings and graded relevance was often investigated indirectly [24].

Therefore, our understanding of cognitive processes that underpin each relevance grade is still limited [88]. To unravel this complex phenomenon, we utilise an experimental design that enables the investigation of graded relevance within the context of NeuraSearch paradigm and in real-time. The NeuraSearch approach constitutes a complementary and promising area that can enhance the understanding of complex relevance assessment through the employment of multidisciplinary knowledge. This study is the first to incorporate relevance theory and a cognitive neuroscience approach to investigate the neural correlates of graded relevance assessment employing a data-driven approach.

Our central aims are to identify: (i) the brain activity associated with distinct graded relevance assessment across time from stimulus onset (ii) test whether there are neural manifestations of cognitive activity underlying each grade of relevance assessment and (iii) test whether processing distinct grades of relevance is associated with significantly different neural signatures.

The findings might also potentially increase our understanding of the neurological properties underlying this process. In addition, the findings can lead to an improvement of user-system interaction, which might result in greater search success. This is because the system might be able to recommend information to the user which is relevant to different degrees, potentially increasing the interaction effectiveness between the user and a system [368]. This is because capturing and decoding brain processes can provide enrichment for information recommendation through the development of novel softwares and personalisation techniques paired with wireless and portable EEG devices, enabling everyday unobtrusive signal acquisition.

This chapter focuses on discovering and mapping the brain mechanisms of graded relevance, within an IR process performed by humans engaged in a Q/A retrieval task. This investigation will undoubtedly further our understanding of the concept of relevance and will provide the evidence needed to strengthen the theoretical foundations.

6.2 Experimental Setup

The paradigm developed for this study enabled the assessment of relevance to be investigated in a graded fashion, and for the corresponding neural activity to be recorded in real-time. The brain activity obtained in the first part of the experimental trial (described in Section 3.2.4) was matched with the graded relevance assessment for each word from the second part of the trial. That way the brain activity for each word was labelled with an appropriate graded relevance assessment. Specifically, participants reflected on sentences that they had seen in response to a question and reported after each word of the sentence what their perception of relevance was at that time. Participants were, therefore, processing each word of the sentence within the context of whether they subjectively perceived the information segment at that time to be relevant to the question.

6.2.1 Participants

After the participant exclusion, this study was carried out with a sample of fourteen remaining participants. The sample constituted of 7 females (50%) and 7 males (50%) with a mean age of 24.93 (SD = 6.27) years. The smaller sample size was a result of data sub-selection (resulting in a low number of trials). However, despite the small sample size, the study can provide a reliable indication for the direction of further research and explain existing behavioural studies examining graded relevance. Six participants were randomly assigned Data Set A and eight participants were randomly assigned Data Set B. Participants were presented with a mean number of 870.57 words (SD = 127.62) and the main experimental task lasted approximately 46.19 minutes (SD = 9.26).

6.2.2 Data Preparation

The data-driven EEG analysis relied upon a participant's graded relevance assessment to the presented information. We used a within-subject experimental design, where IV was graded relevance assessment (with three levels: "No-Relevance" (NONR), "Low Relevance" (LOWR), and "High Relevance" (HIGHR)). The DV was the EEG brain signal, gathered from the users during the Q/A task. The acquired continuous EEG signals were pre-processed using steps described in Section 3.3.1. During pre-processing steps, a mean number of 15.86 ($\pm$9.52) bad channels and 15.86 ($\pm$10.34) components were automatically removed. The mean number and SDs of accepted vs. rejected epochs for every condition of interest in this experiment are presented in Table 6.1.

Table 6.1: The Mean number and SD of accepted and rejected epochs across HIGHR, LOWR and NONR conditions.

Condition	Rejected Epochs		Accepted Epochs	
	Mean	SD	Mean	SD
Graded Relevance				
HIGHR	19.93	17.97	201.36	58.14
LOWR	16.43	17.60	148.29	66.18
NONR	47.29	43.41	381.71	135.03

6.3 Results

Overall, our experimental results show that significant differences exist in brain activity when assessing information as having high-relevance, low-relevance, or no-relevance. Significant differences associated with the mutual comparison of each relevance grade suggest the presence of distinct cognitive processes that are underpinning the formation of graded relevance assessment.

6.3.1 HIGHR vs. NONR

200 - 350ms: The first significant time interval associated with the comparison of HIGHR vs. NONR condition within the 200 - 350ms was associated with 2 significant electrode clusters and higher widespread central positivity associated with HIGHR condition. The first, earlier significance, within the 200 - 300ms over the centro-frontal scalp region (cluster of 13 electrodes: E2, E4, E5, E10, E12, E13, E18, E20, E23, E24, E118, E123, E124) was associated with higher positive amplitudes for HIGHR condition (Figure 6.1, row I). The second, later significant electrode cluster of 5 electrodes (E70, E72, E75, E76, E77), emerged within 250 - 350ms time interval over the posterior scalp locations (Figure 6.1, row II). Based on previous studies reporting similar component distributions [361, 369], observed significant differences might be related to the P300/CPP component, as visual P300/CPP is associated with anterior positivity and posterior negativity. Greater P300/CPP amplitudes for HIGHR condition when compared to NONR condition suggests that during this early stage of implicit relevance assessment, users' selective attention is allocated towards highly relevant stimuli, which are also easier to process in terms of cognitive load [361, 370]. These findings are consistent with the previous literature, showing that the degree of subjectively perceived information relevance is proportional to the P300/CPP component amplitude [5, 371].

300 - 400ms: The next significant cluster of 29 electrodes (E3, E4, E5, E6, E7, E10, E11, E12, E13, E15, E16, E18, E19, E20, E22, E23, E24, E28, E29, E30, E34, E35, E36, E41, E80, E106, E111, E118, E124) within 300 - 400ms time interval emerged within the large fronto-centro-temporal region, more prominent over the left hemisphere (see Figure 6.1, row III). Topographic plots, as well as ERP waveforms, suggest a possible transition between the P300/CPP and N400 ERP components. Centro-frontal positivity observed within the earlier significant time interval (200 - 350ms) appears to be shifting towards bilateral posterior regions. Such activity shifting is also reflected in the ERP waveforms. The P300/CPP component, peaking around 300ms post-stimulus becomes negative, with amplitudes peaking around 400ms post-stimulus. The P300/CPP and N400 share similarities in some respect as they reflect unexpected events, and thus they might share common resources [372]. Both ERP components are sensitive to predictability during linguistic categorisation but represent different cognitive processes. According to Alday and Kretzschmar [373], while the P300/CPP is sensitive to the dynamics of the stimulus categorisation process itself, the N400 component indexes the processing of stimulus properties relevant for categorisation and there is a degree of overlap usually exists between these components.

400 - 500ms: The statistical comparison of HIGHR and NONR conditions was significant within the 400 - 500ms time interval of 34 electrodes (E5, E6, E7, E11, E12, E13, E19, E20, E28, E29, E30, E35, E36, E37, E47, E52, E53, E54, E55, E60, E61, E78, E79, E80, E85, E86, E87, E93, E104, E105, E106, E111, E112, E118), as displayed in Figure 6.1, row IV. Observed bilateral posterior positivity, frontal negativity and significant differences within the centro-posterior cluster were associated with negative-going amplitude deflection, which was the most prominent for the NONR condition, which is consistent with previous literature [33, 205, 283]. The context within this experiment has been provided through the question, and hence it is possible to assume that higher

N400 amplitudes elicited during NONR condition signalise contextual violation [372] while highly relevant information reduces the N400 amplitude [374].

550 - 750ms: The last significant time-interval associated with the comparison of HIGHR and NONR conditions emerged within the 550 - 750ms time interval over the left centro-posterio-temporal cluster of 31 electrodes (E5, E6, E7, E11, E12, E13, E20, E30, E31, E36, E37, E42, E52, E53, E54, E55, E60, E61, E62, E67, E78, E79, E80, E85, E86, E87, E93, E104, E105, E106, E112). Our findings are in alignment with previous binary relevance studies, suggesting that processing of NONR information is associated with reduced LPC amplitudes [33, 205]. As the LPC amplitude is higher for the HIGHR condition, this may reflect that the amount of information carried by the processed term is higher in comparison to the NONR condition [290].

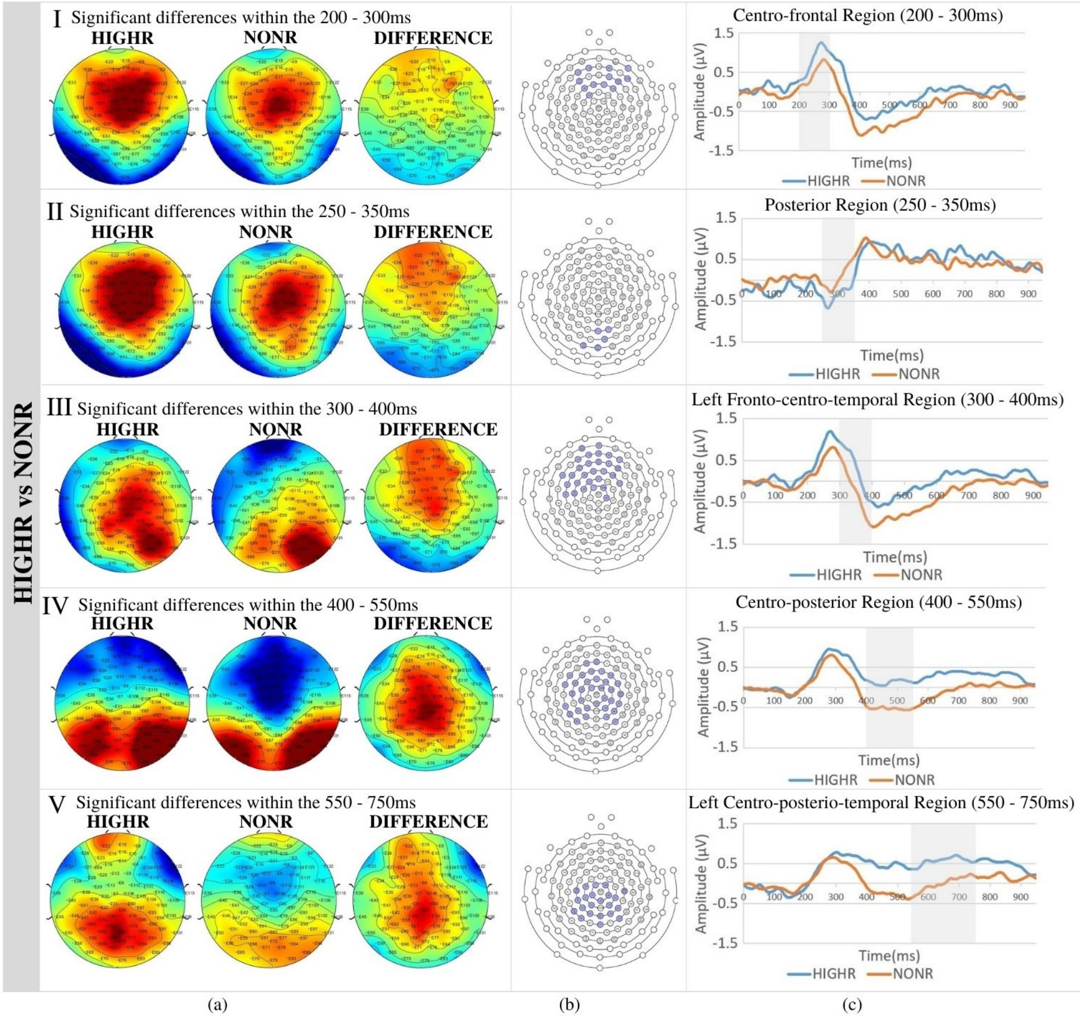

Figure 6.1: (a) Topographic plots for HIGHR vs. LOWR conditions, including a mean difference plot for the 200 - 300ms (I), 250 - 350ms (II), 300 - 400msms (III), 400 - 450ms (IV), and 550 - 750ms (V) time windows. Reddish colours of the scalp topography indicate positive ERP values, whereas bluish colours indicate negative ERP values. (b) The 128-channel net graph with highlighted statistically significant electrode sites for each significant time interval. (c) The comparison of grand averaged ERP waveforms for HIGHR (blue) vs. NONR (orange) condition. Significant time intervals are highlighted in grey for each significant time period.

6.3.2 HIGHR vs. LOWR

300 - 350ms: The data-driven comparison of HIGHR vs. LOWR condition revealed a statistically significant differences within the 300 - 350ms time interval, over the left posterio-temporal cluster of 8 electrodes (E50, E58, E65, E66, E67, E70, E71, E75), as displayed in Figure 6.2 (row I). The posterior negativity differences were associated with negative amplitude deflections (higher for HIGHR condition), which were likely related to the P300/CPP distributions. These findings and distributions are very similar to the previously reported HIGHR vs. NONR comparison, Figure 6.1 (row II). Therefore, the results suggest that HIGHR information is still perceived by the users as significantly more relevant compared to the LOWR information [5, 371].

300 - 400ms: Second significant differences in the EEG signal emerged within the 300 - 400ms time interval. The differences were significant over the frontal cluster of 6 electrodes (E5, E10, E11, E12, E15, E19), as displayed in Figure 6.2 (row II). Similarly to previously reported significant differences associated with the comparison of HIGHR vs.NONR conditions within the 300 - 400ms time interval (6.1 (row III), the topographic plots as well as ERP waveforms associated with the HIGHR and LOWR condition suggests transition between the P300/CPP and N400 ERP components.

350 - 550ms: Significant differences within the 350 - 550ms time interval were recorded over the central cluster of 8 electrodes: E7, E31, E36, E37, E53, E54, E55, E80 (Figure 6.1, row III). The differences were associated with negative amplitude deflections, higher for LOWR condition compared to the HIGHR condition. Scalp distributions and ERP waveforms suggest that the differences are driven by the N400 ERP component. The findings, therefore, suggest that LOWR content still causes some degree of semantic violation which amplifies the N400 component. Furthermore, the processing of LOWR content might be associated with uncertainty in relation to semantic expectancy, which is linked to negative voltage deflection [375] when compared to the HIGHR condition.

550 - 750ms: The last significant time interval resulting from a comparison of HIGHR and LOWR condition emerged within the 550 - 750ms time interval over the centro-posterior cluster of 7 electrodes (E53, E54, E55, E61, E78, E79, E80). Significant electrodes, topo plots and ERP waveforms are displayed in Figure 6.2, row IV. Significant differences related to centro-posterior positivity were likely driven by the LPC component which is amplified by context-relevant information [376]. Similarly to the comparison of HIGHR vs. NONR condition, context-relevant words carry higher amount of information and therefore amplify the HIGHR but not LOWR amplitudes.

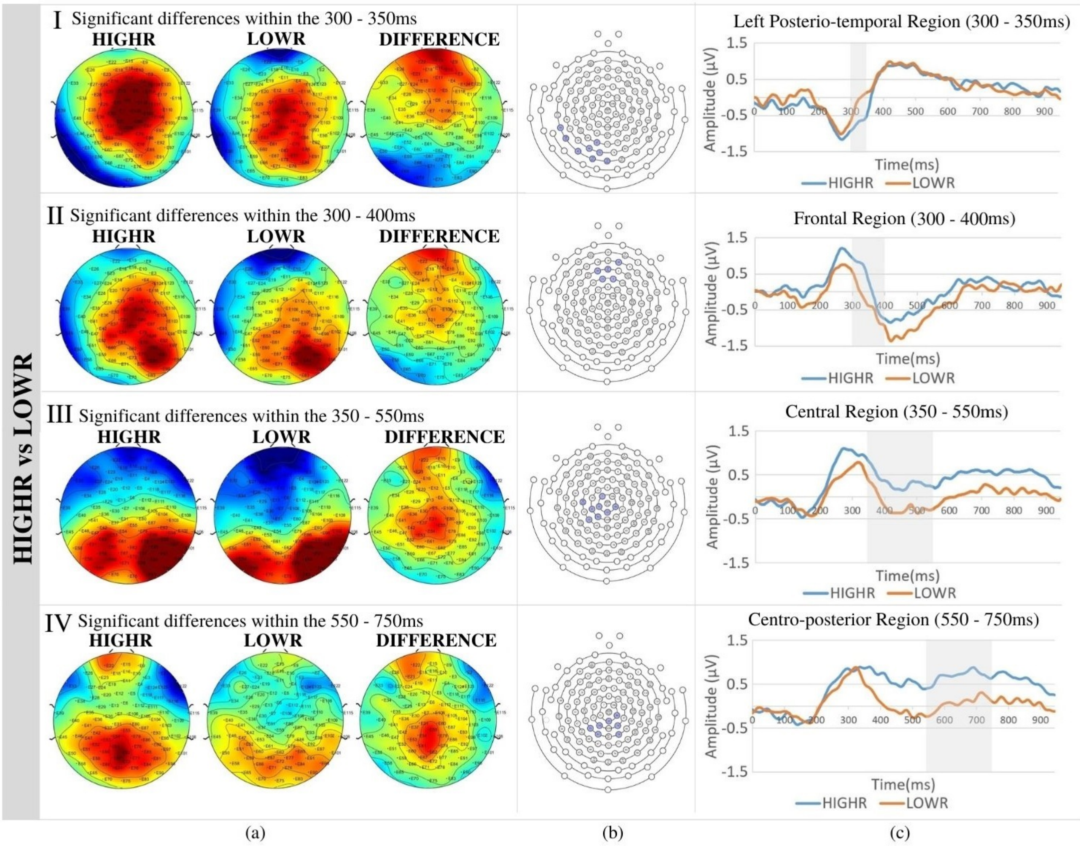

Figure 6.2: (a) Topographic plots for HIGHR vs. LOWR conditions, including a mean difference plot for the 300 - 350ms (I), 300 - 400ms (II) and 350 - 550ms (III), and 550 - 750ms (IV) time windows. Reddish colours of the scalp topography indicate positive ERP values, whereas bluish colours indicate negative ERP values. (b) The 128-channel net graph with highlighted statistically significant electrode sites for each significant time interval. (c) The comparison of grand averaged ERP waveforms for HIGHR (blue) vs. LOWR (orange) condition. Significant time intervals are highlighted in grey for each significant time period.

6.3.3 LOWR vs. NONR

250 - 300ms: The earliest time interval associated with the comparison of LOWR vs. NONR condition was associated with the 250 - 300ms time interval over the left centro-temporal cluster of 5 electrodes (E36, E41, E42, E46, E51), displayed in Figure 6.3, row I. The differences were associated with a small positive going amplitude deflection, higher for the NONR condition. The observed effect might be due to the fact that the P300/CPP amplitudes are sensitive to the degree of certainty such that more certain probabilities elicit higher P300/CPP amplitudes [377, 378]. It is possible that when perceiving information rated as NONR, participants might feel more confident to submit their assessments when compared to the LOWR content.

350 - 600ms: Next significant differences were associated with the two time intervals over the fronto-central cluster. The first significant time interval emerged within the 350 - 450ms interval (cluster of 8 electrodes: E4, E5, E6, E13, E30, E111, E118, E124) displayed in Figure 6.3, row II; and the second time interval emerged within the 500 - 600ms interval (cluster of 11 electrodes: E4, E5, E6, E7, E12, E13, E20, E24, E106, E118, E124) displayed in Figure 6.3, row III. Both of these significant time intervals were associated with bilateral posterior positivity and frontally distributed negativity, characteristic for the N400 component. The differences in negative-going ERP amplitude deflections, which was the most prominent for NONR condition. Our findings are consistent with previous literature [33, 205, 283] and suggest that NONR stimuli require significantly greater cognitive effort when compared to LOWR to process and integrate within the given context [379].

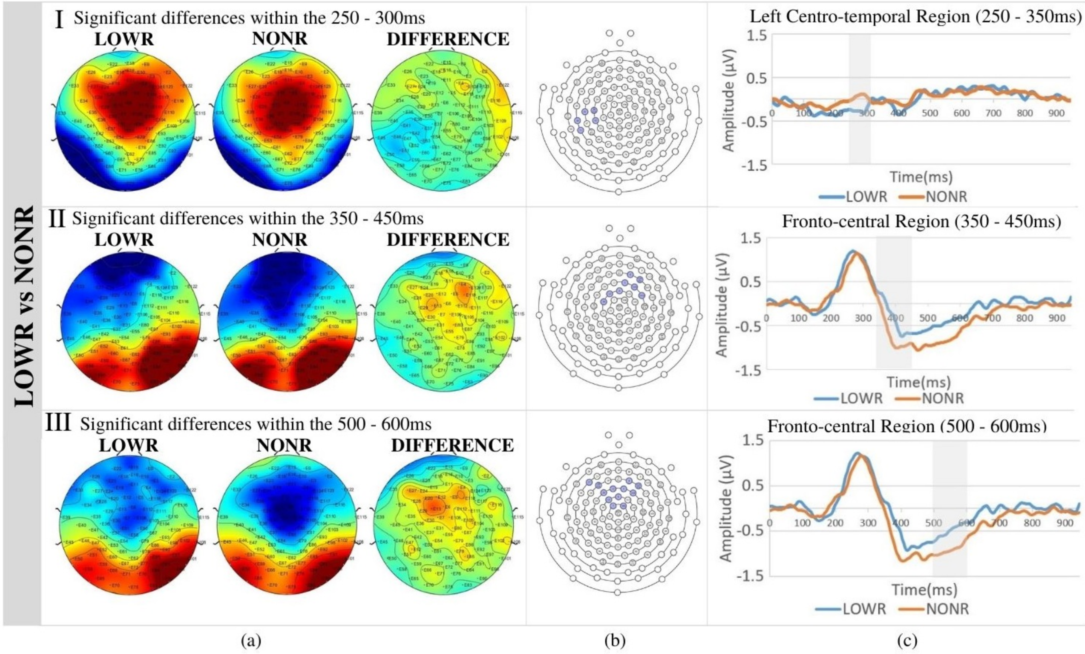

Figure 6.3: (a) Topographic plots for LOWR vs. NONR conditions, including a mean difference plot for the 250 - 350ms (I), 350 - 450ms (II), and 500 - 600ms (III) time windows. Reddish colours of the scalp topography indicate positive ERP values, whereas bluish colours indicate negative ERP values. (b) The 128-channel net graph with highlighted statistically significant electrode sites for each significant time interval. (c) The comparison of grand averaged ERP waveforms for LOWR (blue) vs. NONR (orange) condition. Significant time intervals are highlighted in grey for each significant time period.

6.4 Conclusion

In conclusion, our findings provide support for the concept of graded relevance, given the clear differences in neural activity when information segments are perceived as having high relevance, low relevance or no-relevance. The P300/CPP, N400 and P600/LPC all differed due to the perceived relevance of the answer. Being able to detect relevance in graded manner inputs to information systems, which in turn could lead to improved retrieval effectiveness and greater searcher's satisfaction. Despite a number of ERP components being identified that relate in different ways to the perceived relevance level, it will be important to understand how robust/reliable these differences are and how alterations in the questions (e.g. the difficulty level) or in the answer (e.g. the length of the response) may interact with these features. Finally, we believe our conclusions constitute an important step in unravelling the nature of graded relevance and knowledge of the electrophysiological modulation to each grade of relevance.

6.5 Chapter Summary

This chapter explored graded relevance assessment from the neurocognitive point of view. Our results suggest that there are distinct cognitive processes that underpin each relevance grade. Therefore, we believe our conclusions constitute an important step in unravelling the nature of graded relevance and knowledge of the electrophysiological modulation to each grade of relevance.

Chapter 7

Conclusions

After introducing theoretical background (Section 2) outlining key concept in this thesis and identifying important research gaps, which were then empirically investigated and results presented (Sections 3 and Part II), the thesis comes to a conclusion in this final part. This final chapter concludes the main achievements and contributions presented in this thesis, then discusses the limitations, present its implications for academia and industry and suggests ideas for further research opportunities in the area of IR.

7.1 Thesis Summary

Relevance plays a significant role in IR but up to this date, there is not a widely accepted theory and understanding of relevance. Thus, there is a great need to further investigate this key IR concept, which is often labelled as challenging and difficult to examine mainly due to its complex and subjective nature. This empirical work addresses challenges associated with relevance subjectivity and measurability by employing the NeuraSearch approach. The approach benefits from a direct real-time recording of neuropsychological aspects that contribute to relevance assessment while considering the user's perspective. Theoretically, no other features can reflect the user's subjective relevance perception better than the signals collected directly from the users themselves [130].

As indicated in Chapter 2.7, the main aim of this thesis was to investigate the user's subjective perception of relevance assessment in terms of the cognitive context and assessment granularity. In particular, we conducted and reported on the user study examining neurological differences associated with binary and graded relevance assessment. Furthermore, we have explored the role of the user's SPK in the formation of binary relevance assessment. The empirical work outlined in this thesis has benefited from the employment of a data-driven approach, which enabled us to explore neurophysiological phenomena associated with relevance in detail while trying to minimise literature-driven ERP component selection bias.

7.2 Findings and Contributions

In this thesis, we have presented significant advancements in our understanding of human neurophysiological signals and cognitive mechanisms that drive binary and graded relevance assessment within the context of IR research. The presented work is based on the following premises: (a) understanding user's neural correlates during relevance assessment can provide an in-depth and realistic understanding of relevance processing; (b) as our understanding of relevance is still incomplete, focusing on data-driven rather than literature-driven analysis can facilitate exploration of previously unreported phenomena; (c) by addressing these neurocognitive mechanisms contributing to information processing, it is possible to improve current IR systems which would lead to more efficient search and user's satisfaction. The results of a data-driven examination of relevance revealed a number of novel, interesting areas of discussion. We first present the overall contributions guided by our research aims which are then discussed in more detail.

Overall, this research has reported the following advancements:

- Chapter 2 summarised current views of relevance within the context of IR, identified critical gaps in existing research and emphasised the importance of cognitive aspects that play a crucial role during information processing and categorisation.
- This work proposed to investigate relevance assessment using a data-driven approach while considering the user's IN. The data-driven framework (in comparison

to the literature-driven one) has helped to reveal previously not reported cognitive phenomena that significantly contribute to relevance assessment formation.

- Our results strengthen existing theoretical and empirical relevance foundations in the following ways: (a) the data-driven analysis of binary relevance yielded comparable results to the ones reported by previous studies; (b) neurophysiological differences associated with the processing of graded relevance provide support for the concept of relevance granularity as previously suggested by e.g. [24, 34]; (c) the SPK as a cognitive contextual variable seem to modulate neural activity during relevance assessment, which supports Ingwersen's Cognitive Theory [380].

7.2.1 Binary Relevance

From empirical evidence presented in Chapter 4, it was shown that neurophysiological signals associated with the processing of relevant vs. non-relevant textual information significantly differ (e.g. [33,101]). However, so far relevance has either been investigated within the context of word associations (where the relationship between the user's IN and relevance assessment was not considered, e.g. [33, 101]) or by the employment of component-driven approach (with focus on pre-defined ERP components [206]). Although these approaches provide valid empirical insights, for an in-depth understanding of this phenomenon it is important to investigate relevance as a part of IR while considering the entire time-scale of relevance assessment build-up and not just its parts. In order to gain an in-depth understanding of the intricate cognitive processes that underlie the formation of binary relevance assessments, the current experiment employed a data-driven approach in contrast to the above-mentioned component-driven approaches. In particular, we recorded the user's neural signals associated with the processing of subjectively assessed 'rel' and 'nr' textual content in real-time in response to the Q/A relevance assessment task. The data-driven approach was used to compare information assessed as relevant to that which was assessed as non-relevant. The data-driven analysis revealed significant differences in neurophysiological signal associated with the user's subjectively perceived relevance in distinct significant time intervals linked to differences in topographic plots and ERP waveform distributions (addressing the **RQ1** outlined in Section 1.4). Along with previously reported N400 and LPC ERP components in studies investigating binary relevance (e.g. [206, 299]) (addressing the **RQ2** outlined in Section 1.4), our results revealed significant differences in neural activity in the early time interval associated with the P100 ERP component (addressing the **RQ3** outlined in Section 1.4). The neurological variations linked to the P100 component could be early indicators of selective attention allocation, showing an increased focus on 'rel' information during early sensory facilitation, which is later transferred to higher levels of cognitive processing. The past IR studies provide mixed results in terms of the P300 component. While some studies have found reported differences related to P300/CPP component within the context of textual relevance assessment between relevant and irrelevant words (e.g. [94, 280]), others did not - which is consistent with our results (e.g. [33]). Further studies are needed to provide clarification as the P300 component is complex and frequently elicited by different combinations of experimental variables [381]. The different contributions of neural sub-processes to the overall P300 amplitude at varying time points should be considered to better address the component's functional significance.

P100. The 100 - 200ms time interval seems to be an early time point associated with the P100 ERP component, reflecting the early activation of primary visual areas linked to the participant's selective attention modulation associated with the processing of relevant information. The P100 component is not sensitive to the stimulus task-relevance [7], but rather the P100 amplitude enhancement is linked to attentional relevance coupled with enhanced neural excitability of the visual cortex [7]. Such enhanced visual excitability might reflect the pre-activation of sensory networks in response to effect anticipation [382]. The P100 may indicate participants' initial capacity and processing effort to recognise relevant stimuli during relevance assessment construction, according to previous studies that have shown a direct correlation between the P100 and working memory performance [346]. Further research could explore the relationship between the P100 and LPC ERP components, which are related to attention during relevance assessment formation and provide clarification on their mutual interaction.

N400. The N400 topographies and time-window differences between the "rel" and "nr" conditions were significantly different, with "rel" condition being associated with significantly higher amplitude than the "nr" condition (see Figure 4.1, line II). The N400 component has been extensively researched in the sense of semantic processing and the findings have shown that N400 represents many aspects of semantic knowledge integration and retrieval. The processing of semantic mismatch increases N400 negativity (see e.g. [383]). Accordingly, less negative amplitude deflections for the "rel" condition may signify a higher degree of answer relevance to the query, as was previously observed by [206], who discovered a lower N400 for highly relevant information.

LPC. The significant differences within the LPC topographies and waveforms (see Figure 4.1, line III) demonstrate that the amplitudes associated with the LPC component are much higher for information assessed as relevant compared to non-relevant. The LPC is frequently reported to follow the N400 component and is connected to cumulative evidence exposure during decision-dependent tasks [350]. Thus, when the memory judgement at hand requires consideration of the relevance dimension in search tasks, the LPC amplitudes appear to be influenced by the participant's response to a stimulus. Additionally, higher amplitudes observed under the relevant condition might indicate that the process of categorising words is less cognitively demanding for participants.

7.2.2 SPK

The second empirical contribution of this thesis focused on the investigation of the role of SPK as a cognitive contextual variable during binary relevance assessment (see Chapter 5). The main finding, which addresses **RQ4**, outlined in Section 1.4, is that there are significant differences in neural activity associated with the user's SPK when they perceive information as relevant or as non-relevant. Data-driven analyses revealed distinct significant time intervals and cortical differences driven by the self-perceived level of knowledge the user had about the question during relevance assessment. The differences in neural activity suggest that a user's SPK affects a variety of cognitive processes, which underpin relevance assessment formation, such as attentional engagement, perception of semantic relatedness and working memory engagement (addressing

RQ5, outlined in Section 1.4).

SPK & Non-Relevance Assessment. When judging information to be non-relevant, SPK was associated with key differences that emerged within two significant time intervals with similar topographic activity patterns. The differences in both time intervals were driven by centro-parietal positivity during a time period corresponding with the N400. The SPK might facilitate the cognitive expectancy process and potentially help with information integration. If the positivity is taken to reflect the same processes as the N400 (given the bipolar representation across the scalp), then the greater amplitude in relation to SPK might reflect a greater degree of perceived semantic congruency (e.g. the answer is not relevant and the participant is aware of that). Users with SPK might experience reduced uncertainty levels and make more accurate information relevance predictions [22].

Next significant neurophysiological differences were associated with the LPC amplitudes, which were higher across knowledgeable condition within the context of no-relevance. Significant amplitude differences related to LPC ERP component with differences recorded over the right posterio-temporal region were higher when participants subjectively perceived some degree of knowledge of the information presented. The LPC amplitudes are higher for previously seen stimuli, especially those classified as "old," than for stimuli classified as "new" [384]. This is assumed to be the index for recollection — recognition accompanied by accurate source memory [349,385–387]. The LPC amplitudes correlate with the item memory strength (confidence) [388]. The left hemisphere supports the recollection of word associations, which results in more positive LPC amplitudes for task-related stimuli compared to unrelated ones. On the other hand, the right hemisphere processes categorically related words (as indicated by a tendency to observe a difference between categorically related and unrelated words) [351]. Higher amplitudes recorded over the right hemisphere across all the know conditions might suggest that prior knowledge facilitates the word categorisation process during the presentation with non-relevant content.

SPK & Relevance Assessment. The significant time interval (within the left fronto-centro-temporal region), was associated with significantly higher P300/CPP amplitudes for content assessed as knowledgeable. This may be influenced by the amount of cognitive control [336], referring to high-level executive functions such as attention, salience detection, working memory and task management. The P300/CPP may be related to a process such as recognition of previously encountered information [389].

A reduction of the N400 within the time interval of $350 - 450$ms, more prominent for subjectively perceived known relevant information, might be related to semantic information retrieval [364]. The N400 amplitude positively correlates with the ease of semantic processing [390], and information recognition during the presentation of self-important information [347]. This may suggest that SPK decreases cognitive effort when processing information within a subjectively relevant context [379]. It is possible that the P300/CPP component and N400 deflections associated with the processing of subjectively perceived known relevant information are interdependently modulating relevance assessments [372], as both of these components have been frequently linked to relevance processing (e.g. [206]). However, further research is required to provide clarification.

Another important difference was seen in the LPC which is commonly reported to follow the N400 [391] and it is a key component that relates to memory-based decisions [392]. No SPK conditions might, therefore, require higher memory effort during decision-making tasks that require relevance considerations [350]. Also, past studies have reported that learning is correlated with an increase in LPC amplitude [366] which supports Ingwersen's Cognitive Theory, suggesting that IR facilitates information transfer into knowledge and novel cognition [380].

The differences observed within the LPC component might suggest that content relevance has a higher impact on brain activity modulations than the user's SPK. Our results suggest that for non-relevant information items, the LPC amplitudes were higher for SPK conditions but for subjectively relevant items the amplitudes were higher across no SPK conditions. It seems that the LPC component amplitudes are strongly modulated with the evidence accumulation that contributes to the overall relevance assessment decision rather than the user's SPK. Therefore, it might be that the user's SPK plays an important role during early information processing, but as the user becomes more consciously aware of the nature of semantic information the role of SPK as a cognitive context recedes into the background. However, further studies are required to provide clarification.

7.2.3 Graded Relevance

The paradigm developed for this study enabled assessments of relevance to be judged in a graded fashion, and for the corresponding neural activity to be recorded. Specifically, participants reflected on sentences that they had seen in response to a question and reported after each word of the sentences what their perception of relevance was at that time (HIGHR, LOWR or NONR). Participants were, therefore, processing each word of the sentence within the context of whether they subjectively perceived the information segment at that time to be relevant to the question and hence, to their simulated IN.

The key findings which emerged from the study (presented in Chapter 6 are that levels of neural activity across time are dependent on whether the person perceived the sentence as of high relevance, low relevance or no-relevance to the question (this finding addresses **RQ6**, outlined in Section 1.4). A data-driven approach used in this study has detected three ERP components (P300/CPP, N400 and LPC) associated with the information processed in the context of subjectively perceived high relevance, low relevance or no-relevance. The differences in neural activity suggest that during the assessment of relevance, a variety of cognitive processes are relied upon to different degrees. For example, higher relevance might be linked to greater attentional engagement, higher perception of semantic relatedness (the lower the semantic incongruency between the context of the question and the answer), and a greater requirement for engagement of memory (relevant information might be deemed more important to encode and recall than irrelevant information) - this discussion provides an early step towards answering **RQ7**, outlined in Section 1.4.

High Relevance: The results suggest that the greater attentional resources are allocated to highly relevant stimuli in comparison to stimuli of no and low-relevance, as indicated through the differences observed within the P300/CPP component. The P300/CPP amplitude has been shown to be proportional to attentional engagement

[393]. Greater P300/CPP amplitude has also been suggested to reflect the quantity of information transmitted [394], the quantity of useful information [395], relevance to the self [371], or relevance to the task [396,397], or to assessment of relevance specifically [5]. Our results also suggest significant differences in the time interval during which there seemed to be a transition between the P300 and N400 components. It is possible that our P300/CPP measure is influenced by the N400 deflections, but the P300/CPP and N400 may both be modulated during the assessment of relevance. The underlying processes might therefore be related [372], but future research is necessary to provide clarification.

There was a clear reduction of the N400 associated with the time interval of approximately 300 – 550ms when comparing highly relevant content to non- and low-relevant one. Typically, in studies of language, the N400 provides an index of semantic relatedness; it is larger when there is a semantic mismatch than semantic congruency (see e.g. [383]). Given the task requirements of the current study and the differences we observed in the N400 in response to graded relevance, it seems that in this case the component is modulated not necessarily in response to the meaning of the word, but to the relatedness of the sentence to the question. Words processed in the context of high relevance are semantically aligned to the question, which likely explains the attenuated N400 response. The N400 results fit closely with the study by Eugster and colleagues [33], who found a reduced N400 for relevant words, compared to irrelevant words.

In our study, words linked to sentences of high relevance were associated with the highest LPC amplitudes compared to the words of non and low relevant sentences. Similarly, Eugster et al. [33] found in their study that relevant words elicited larger P600 components than irrelevant words. The link between higher relevance and the P600/LPC amplitude is not completely clear. In terms of P600 linked to syntactic processing, later research has flagged a semantic-thematic role (however, larger P600 amplitudes are found due to violations – see e.g. [398]). A more likely reason, or at least a partial explanation, may be that the late stage positivity is instead linked to memory processing (e.g. through a process such as recognising that the answer is relevant and is linked to the question) and instead of P600 component, it is the LPC component that should be emphasised within the context of relevance processing. A LPC has been observed during memory recognition, it is higher for old versus new stimuli, occurs at around 600ms after a stimulus and also has a central posterior topography [350].

No-Relevance: Words processed without any perceived relevance to the question had the lowest P300/CPP and the greatest N400 amplitudes across all the comparison conditions. Additionally, the non-relevant content was associated with lower LPC amplitudes during the comparison with highly relevant content. Conversely to the words processed in the high relevance context, the words processed that are not relevant to the question may have a low P300 due to cognitive factors such as low attentional engagement, a large N400 due to a mismatch between the semantic material offered in the answer given the context of the question (larger semantic incongruity), and a lower LPC due to reduced memory processing given that the answer is not relevant to the question (e.g. information to be retained results in larger LPC amplitudes than information to be forgotten - see e.g. [399]. No significant differences observed during the comparison of non-relevant and low-relevant content might suggest that participants are not intentionally storing this type of content in their memory for further processing as it does not provide an answer to their IN.

Low Relevance: A crucial question relates to the manner in which words are processed in the low relevance context. Specifically, is the processing of these words more similar to words viewed in the highly relevant context or the non-relevant context? The ERP component amplitudes for the words processed in the context of low relevance fell somewhere between those processed in the context of high relevance and those processed in the context of no relevance (significant differences were seen for P300/CPP and N400 components). However, during the later information processing, there were no significant differences during the comparison of non-relevant and low-relevant content.

7.3 Study Implications

The findings presented in this thesis have a number of implications for both theory and practice. The major contribution of this thesis to the literature is an in-depth, interdisciplinary, data-driven exploration of neurophysiological phenomena that contribute to binary and graded relevance formation while considering the user's cognitive context. The field of IR can benefit from objective and proactive detection of user's cognitive states during relevance assessment, which can be implemented to improve the current system design [400, 401].

Further understanding of neurological properties of relevance might provide valuable insight into personalisation within IR [124], leading to a significant improvement in addressing user's INs [3]. This is because annotating content with some measure of relevance might be especially useful to filter or/and personalise the content to users. The more accurate information the system possesses about the user, the higher the efficacy in assisting the user by providing them with relevant content. For instance, if the content is too complex for the user to comprehend, the user might be unable to effectively interact with the retrieved information. As a result, the problem-solving may fail to occur [368]. In this context, gathering relevance feedback efficiently, unobtrusively and in real-time on a set of results allows better iteration of results [29], leading to effective IN satisfaction while reducing information overload through the reduction of searcher's effort.

In terms of graded relevance feedback, previous research has not yet specified clear criteria and the borderline between distinct relevance grades was unclear [302] despite the fact that graded assessments reflect the user's subjective relevance perception more thoroughly when compared to the traditional binary scale [402]. The current study was able to provide valuable insight into cognitive processes that contribute to the formation of distinct relevance assessment grades. We provide empirical support for the theory of graded relevance and suggest that the human brain processes distinct grades of relevance significantly different. Emerged significant differences in neural processing manifested themselves in the form of P300/CPP, N400 and LPC ERP components.

Generally, both the user and the system would significantly benefit from further cognitively focused IR research as human-information interaction undoubtedly involves a series of complex subjective mental representations and cognitive processes. Better understanding and support of such interactions between human users and elements of the system might improve the accessibility of information environments. Our research highlights cognitive processes that contribute to relevance assessment formation and can be used as an objective measure to inform the system about the distinct mental perception of a user. The NeuraSearch approach and in particular the employment of EEG is a highly informative tool which enables to capture the triggers and drivers of subjective mental phenomena with excellent temporal precision. Furthermore, the extraction, classification and automatic prediction of the key EEG features can be integrated to create effective BCI based IR systems. On the other hand, despite significant advances in the field and promising impact, the use of EEG technology to capture user's implicit feedback as a natural interaction modality is associated with technological and economical challenges that have so far prevented mass adoption of this approach. At the same time, researchers should continue to explore the possibilities of everyday applications of the EEG technology to improve IR due to its proven significant advantages.

7.4 Study Limitations and Further Research Avenues

The study design of this thesis was based on an improved version of previous text-based relevance studies (e.g. [33, 101]). Such improvement was achieved by positioning relevance in the context of IR through incorporating IN rather than focusing on the judgement of word-relatedness. However, it is possible to argue that the IN was introduced as an external and artificial factor through the question presentation. Therefore, as IN is an important aspect that within the IR context precedes relevance assessment, future studies should consider incorporating more naturalistic representations of the user's actual INs, such as user's own query submissions while considering their true INs. This work mainly focused on one of the parts of IR process. Investigating relevance while considering all components of IR process while incorporating the user's actual rather than simulated INs might further improve the research validity. However, it is important to consider the cognitive demands placed on the user while completing the full IR task. Furthermore, it is also important to account for motor and eye movements resulting from participants' interaction with the system while submitting the query and searching through the results.

Another general limitation across all the work in this thesis is that participants were presented with a question answer presented word by word rather than as a continuous text. Although this is a common approach in many EEG studies examining textual processing as it minimises eye-movement-related artifacts [326], presenting participants with continuous text would be better suited to simulate naturalistic information interaction.

For the purposes of our experiment, we collected data from a total number of forty-two participants which is more than the requirement of power analysis. However, we had to exclude a large number of participants due to the high number of artefacts present in their data. Subsetting existing data have further amplified the issue as some participants had a low number of trials during certain conditions and therefore had to be excluded. Smaller participant samples might decrease the statistical power of the test, which might in turn reduce the reliability and generalisation of our findings. While this is a valid concern, it is important to note that our results are mainly consistent with previous NeuraSearch and neurocognitive studies examining cognitive processes during text comprehension and decision-making.

The experimentation in this thesis examining user relevance assessment was conducted using textual stimuli. However, it is important to note that ERP components as the neural bases of cognitive processes are sensitive to stimulus modalities [403]. While most of the information consumption online happens in textual format, gaining a better understanding of binary and graded relevance assessments in response to image, audio and video retrieval might significantly enrich the IR field.

It is important to mention that so far there is no general EEG methodological protocol to investigate textual relevance phenomena and tasks employed across past experiments differ in their approaches. Usually, the studies would present the participant with a question to simulate the query (within IR context) or with a term or topic. Then, in the next step, the participant would express their subjectively perceived relevance of the offered content. However, the way the content is presented differs across relevance experiments. For instance, Gwizdka et al. [94] presented participants with a short continuous paragraph. On the other hand, in the study of Eugster et al. [101] participants were instructed to submit their relevance judgement after each term they have been presented with. Our approach can be seen as a combination of the aforementioned study designs. For binary relevance, participants submitted their relevance assessment after they acquired enough information. For graded relevance, participants submitted their assessments word-by-word in response to information accumulation. Such differences in approaches to evaluating relevance might be considered a general limitation within the field and should be addressed by future research.

Further questions relate to determining the neural thresholds that must be exceeded for relevance assessment to be made, and to identifying electrophysiological features with the greatest power to predict relevance assessment and users' cognitive context both for binary and graded assessments. Being able to automatically detect relevance assessment while considering users' cognitive states gives more sophisticated ways to detect users' reactions to information and therefore gives better inputs to information systems. This is because capturing and decoding brain processes can provide enrichment for information recommendation through the development of novel softwares and personalisation techniques paired with wireless and portable EEG devices, enabling everyday unobtrusive signal acquisition [101, 170, 283]. This could lead to a reduction or elimination of explicit relevance assessments, which might help to make IR systems much more user-friendly as minimising user effort is critical. Additionally, future studies should take into account the proliferation of mobile and tablet devices. This is because Ong and colleagues [404] demonstrated that search behaviours do differ between individuals using desktop computers and smartphones. Therefore, the process of relevance assessment formation should also be considered within this context.

The work in this thesis mainly focused on one aspect of the user's cognitive states - their SPK of the content topic. However, users' internal states involve additional complex factors such as their emotions, expectations, perceived time pressure and stress, mental capacity etc. Many of these internal states and their influence on relevance assessment remain unexplored and therefore this is the area where future work could be done.

7.5 Final Reflections

The thesis provided valuable insight into relevance phenomena and can serve as an important basis for further research. Relevance is a difficult phenomenon to understand. There is a wide range of different factors that influence this internal decision-making process that depends on the user's subjective perception of information which occurs within the context of IR. Despite the inherently difficult task that understanding of relevance represents we believe that the potential benefits of further exploration in this area will undoubtedly aid the searchers and researchers of future IR systems.

The work presented in this thesis was an effort in this direction. We have re-visited binary relevance assessment from a neuroscientific point of view and furthered our existing understanding of this phenomenon by employing a data-driven approach. The findings were consistent with previous studies focusing on binary relevance (e.g. [33]), but also revealed additional important ERP component - P100 which was previously undetected using component-driven approaches. Furthermore, we have explored the user's SPK state and its effect on relevance assessment formation. Lastly, we explored relevance as a graded variable. Existing user studies using brain imaging considered relevance only in binary terms which treats relevance grades as equally important. This premise is clearly not true, as users consider the importance of information to different degrees.

Moreover, the efficiency of graded relevance in recommender systems has been proven, as its predictive accuracy has outperformed the approaches relying on a binary scale [23]. Examining differential judgement perception and execution from a user point of view provided a simple extension to the traditional relevance research [315]. This is important as exploring manifestations of graded relevance at the visceral level could lead to a better understanding of automatic relevance prediction. If researchers really start using graded relevance metrics, this enables them to build IR systems that can return highly relevant documents on top of partially relevant ones [405].

It is to be expected that future IR systems will routinely use additional unobtrusive information sources (such as brain imaging) as soon as the necessary measurement techniques are widely available; they have already been demonstrated to provide useful information that should not be ignored. It is unlikely that more complex behaviour could be accurately decoded without measurements of brain activity [278]. Offline neurological signal processing is often a necessary preliminary step preceding the development of online BCI research [406]. Future NeuraSearch research within the context of IR could focus on bridging the gap between EEG data classification and their application in online BCIs.

7.6 Chapter Summary

This chapter contrasted the research findings and objectives, discussed the implications of the findings for academia and industry, unveiled the limitations of the study, and recommended future research avenues. The empirical work presented in this thesis concludes, that neurophysiological signals play an important role in furthering our understanding of relevance phenomena, as they provide an important insight into users' neurocognitive processes and should be considered in future works.

Bibliography

[1] C. Cool, N. Belkin, O. Frieder, and P. Kantor, "Characteristics of text affecting relevance judgments," in *National online meeting*, vol. 14. LEARNED INFORMATION (EUROPE) LTD, 1993, pp. 77–77.

[2] P. Ingwersen, "Cognitive analysis and the role of the intermediary in information retrieval," *Intelligent information systems*, pp. 206–237, 1986.

[3] I. Ruthven and D. Kelly, *Interactive Information Seeking, Behaviour and Retrieval*, Ian Ruthven and Diane Kelly, Eds. London: Facet Publishing, 2011.

[4] D. M. Weigl and C. Guastavino, "Applying the stratified model of relevance interactions to music information retrieval," in *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*. American Society for Information Science, 2013, p. 136.

[5] M. Allegretti, Y. Moshfeghi, M. Hadjigeorgieva, F. E. Pollick, J. M. Jose, and G. Pasi, "When relevance judgement is happening? an eeg-based study," in *Proceedings of the 38th international acm sigir conference on research and development in information retrieval*, 2015, pp. 719–722.

[6] Y. Moshfeghi, "Neurasearch: Neuroscience and information retrieval," in *CEUR Workshop Proceedings*, vol. 2950, 2021, pp. 193–194.

[7] C. Popovich and W. R. Staines, "The attentional-relevance and temporal dynamics of visual-tactile crossmodal interactions differentially influence early stages of somatosensory processing," *Brain and Behavior*, vol. 4, no. 2, pp. 247–260, 2014.

[8] W. S. Cooper, "A definition of relevance for information retrieval," *Information storage and retrieval*, vol. 7, no. 1, pp. 19–37, 1971.

[9] E. Riloff and L. Hollaar, "Text databases and information retrieval," *ACM Computing Surveys (CSUR)*, vol. 28, no. 1, pp. 133–135, 1996.

[10] T. Saracevic, "Relevance: A review of the literature and a framework for thinking on the notion in information science. part iii: Behavior and effects of relevance," *Journal of the American Society for information Science and Technology*, vol. 58, no. 13, pp. 2126–2144, 2007.

[11] P. Ingwersen, "Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory," *Journal of documentation*, 1996.

[12] E. Cosijn and P. Ingwersen, "Dimensions of relevance," *IP&M*, vol. 36, no. 4, pp. 533–550, 2000.

[13] T. J. Froehlich, "Relevance reconsidered—towards an agenda for the 21st century: Introduction to special topic issue on relevance research," *Journal of the American Society for Information Science*, vol. 45, no. 3, pp. 124–134, 1994.

[14] S. Mizzaro, "Relevance: The whole history," *Journal of the American society for information science*, vol. 48, no. 9, pp. 810–832, 1997.

[15] ——, "How many relevances in information retrieval?" *Interacting with computers*, vol. 10, no. 3, pp. 303–320, 1998.

[16] L. Schamber, M. B. Eisenberg, and M. S. Nilan, "A re-examination of relevance: toward a dynamic, situational definition," *Information processing & management*, vol. 26, no. 6, pp. 755–776, 1990.

[17] D. M. Weigl and C. Guastavino, "Applying the stratified model of relevance interactions to music information retrieval," *Proceedings of the ASIST Annual Meeting*, vol. 50, no. 1, 2013.

[18] Z. Pinkosova and Y. Moshfeghi, "Cortical activity of relevance," in *CEUR Workshop Proceedings*, vol. 2537, 2019, pp. 10–15.

[19] Z. Pinkosova, W. McGeown, and Y. Moshfeghi, "Revisiting neurological aspects of relevance: an eeg study," in *Advanced Online & Onsite Course & Symposium on Artificial Intelligence & Neuroscience*, 2022.

[20] T. Saracevic, "Effects of inconsistent relevance judgments on information retrieval test results," in *The Notion of Relevance in Information Science.* Springer, 2017, pp. 81–87.

[21] T. McDonnell, M. Lease, M. Kutlu, and T. Elsayed, "Why is that relevant? collecting annotator rationales for relevance judgments," in *Fourth AAAI Conference on Human Computation and Crowdsourcing*, 2016.

[22] J. Jiang, D. He, D. Kelly, and J. Allan, "Understanding ephemeral state of relevance," in *CHIIR '17.* NY, USA: ACM, 2017, p. 137–146.

[23] J. Kekäläinen and K. Järvelin, "Using graded relevance assessments in ir evaluation," *Journal of the American Society for Information Science and Technology*, vol. 53, no. 13, pp. 1120–1129, 2002.

[24] M. Zhitomirsky-Geffet, J. Bar-Ilan, and M. Levene, "How and why do users change their assessment of search results over time?" *ASIST*, vol. 52, no. 1, pp. 1–4, 2015.

[25] C. L. Barry, "User-defined relevance criteria: An exploratory study," *JASIST*, vol. 45, no. 3, pp. 149–159, 1994.

[26] Y. Moshfeghi and J. M. Jose, "An effective implicit relevance feedback technique using affective, physiological and behavioural features," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 2013, pp. 133–142.

[27] I. Arapakis, J. M. Jose, and P. D. Gray, "Affective feedback: An investigation into the role of emotions in the information seeking process," in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '08. New York, NY, USA:

Association for Computing Machinery, 2008, p. 395–402. [Online]. Available: https://doi.org/10.1145/1390334.1390403

[28] N. J. Belkin, "Anomalous states of knowledge as a basis for information retrieval," *Canadian journal of information science*, vol. 5, no. 1, pp. 133–143, 1980.

[29] O. Barral, I. Kosunen, T. Ruotsalo, M. M. Spapé, M. J. Eugster, N. Ravaja, S. Kaski, and G. Jacucci, "Extracting relevance and affect information from physiological text annotation," *User Modeling and User-Adapted Interaction*, vol. 26, no. 5, pp. 493–520, 2016.

[30] L. Schmüser, A. Sebastian, A. Mobascher, K. Lieb, O. Tüscher, and B. Feige, "Data-driven analysis of simultaneous eeg/fmri using an ica approach," *Frontiers in neuroscience*, vol. 8, p. 175, 2014.

[31] T. D. Wilson, "Information needs and uses: fifty years of progress," *Fifty years of information progress: a Journal of Documentation review*, vol. 28, no. 1, pp. 15–51, 1994.

[32] W. B. Croft, "The importance of interaction for information retrieval." in *SIGIR*, vol. 19, 2019, pp. 1–2.

[33] M. J. Eugster, T. Ruotsalo, M. M. Spapé, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski, "Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals," *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[34] T. Saracevic, "The notion of relevance in information science: Everybody knows what relevance is. but, what is it really?" *Synthesis Lectures on Information Concepts, Retrieval, and Services*, vol. 8, no. 3, pp. i–109, 2016.

[35] C. N. Mooers, "Application of random codes to the gathering of statistical information," Ph.D. dissertation, Massachusetts Institute of Technology, 1948.

[36] M. Sanderson and W. B. Croft, "The history of information retrieval research," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1444–1451, 2012.

[37] E. M. Voorhees, "The evolution of cranfield," in *Information retrieval evaluation in a changing world.* Springer, 2019, pp. 45–69.

[38] C. Buckley and E. M. Voorhees, "Retrieval evaluation with incomplete information," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, 2004, pp. 25–32.

[39] G. Amati, "Information retrieval," in *Encyclopedia of Database Systems, Second Edition*, L. Liu and M. T. Özsu, Eds. Springer, 2018. [Online]. Available: https://doi.org/10.1007/978-1-4614-8265-9_915

[40] P. Ingwersen, "Users in context," in *European Summer School on Information Retrieval.* Springer, 2000, pp. 157–178.

[41] P. Borlund, "Interactive information retrieval: An introduction," 2013.

[42] P. Ingwersen, *Information retrieval interaction.* Taylor Graham London, 1992, vol. 246.

[43] Y. Kagolovsky and J. R. Mohr, "A new approach to the concept of" relevance" in information retrieval (ir)," in *MedInfo.* Citeseer, 2001, pp. 348–352.

[44] C. D. Manning, *Introduction to information retrieval.* Syngress Publishing,, 2008.

[45] C. Cole, "A theory of information need for information retrieval that connects information to knowledge," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 7, pp. 1216–1231, 2011.

[46] D. O. Case and L. M. Given, "Looking for information: A survey of research on information seeking, needs, and behavior," 2016.

[47] R. S. Taylor, "Question-negotiation an information-seeking in libraries." LEHIGH UNIV BETHLEHEM PA CENTER FOR INFORMATION SCIENCE, Tech. Rep., 1967.

[48] T. D. Wilson, "On user studies and information needs," *Journal of documentation*, 1981.

[49] ——, "Models in information behaviour research," *Journal of documentation*, vol. 55, no. 3, pp. 249–270, 1999.

[50] B. Dervin, "An overview of sense-making research: concepts, methods and results to date." International Communications Association Annual Meeting, 1983.

[51] B. Dervin, L. Foreman-Wernet, and E. Lauterbach, *Sense-making methodology reader: Selected writings of Brenda Dervin.* Hampton Press, 2003.

[52] R. Golman, D. Hagmann, and G. Loewenstein, "Information avoidance," *Journal of economic literature*, vol. 55, no. 1, pp. 96–135, 2017.

[53] D. Nicholas, *Assessing information needs: tools, techniques and concepts for the internet age.* Routledge, 2003.

[54] D. Bawden and L. Robinson, *Introduction to information science.* Facet Publishing, 2015.

[55] T. J. Bothma and H. Bergenholtz, "" information needs changing over time": a critical discussion," *South African Journal of Libraries and Information Science*, vol. 79, no. 1, pp. 22–34, 2013.

[56] J. T. Hart, "Memory and the feeling-of-knowing experience." *Journal of educational psychology*, vol. 56, no. 4, p. 208, 1965.

[57] D. Michalkova, M. Parra Rodriguez, and Y. Moshfeghi, "Drivers of information needs: A behavioural study–exploring searcher's feeling-of-knowing," in *Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval*, 2022, pp. 171–181.

[58] ——, "Confidence perceptions as part of searcher's cognitive context," in *Advanced Online & Onsite Course & Symposium on Artificial Intelligence & Neuroscience*, 2022.

[59] L. Rossetto, C. Tănase, and H. Schuldt, "Dealing with ambiguous queries in multimodal video retrieval," in *International Conference on Multimedia Modeling.* Springer, 2016, pp. 898–909.

[60] T. D. Wilson, "Human information behavior," *Informing science*, vol. 3, p. 49, 2000.

[61] J. Klayman and Y.-w. Ha, "Hypothesis testing in rule discovery: Strategy, structure, and content." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 15, no. 4, p. 596, 1989.

[62] R. Fidel, *Human information interaction: An ecological approach to information behavior.* Mit Press, 2012.

[63] S. Russell, I. S. Moskowitz, and A. Raglin, "Human information interaction, artificial intelligence, and errors," in *Autonomy and Artificial Intelligence: A Threat or Savior?* Springer, 2017, pp. 71–101.

[64] M. J. Albers, "Human–information interaction with complex information for decision-making," in *Informatics*, vol. 2, no. 2. MDPI, 2015, pp. 4–19.

[65] C. C. Kuhlthau, "Inside the search process: Information seeking from the user's perspective," *Journal of the American society for information science*, vol. 42, no. 5, pp. 361–371, 1991.

[66] T. Saracevic, "The stratified model of information retrieval interaction: Extension and applications," *Proceedings of the Annual Meeting-American Society for Information Science*, vol. 34, pp. 313–327, 1997.

[67] K. Järvelin, "An analysis of two approaches in information retrieval: From frameworks to study designs," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 971–986, 2007.

[68] R. Fidel, A. Mark Pejtersen, B. Cleal, and H. Bruce, "A multidimensional approach to the study of human-information interaction: A case study of collaborative information retrieval," *Journal of the American Society for information Science and Technology*, vol. 55, no. 11, pp. 939–953, 2004.

[69] S. Talja, H. Keso, and T. Pietiläinen, "The production of 'context'in information seeking research: a metatheoretical view," *Information Processing & Management*, vol. 35, no. 6, pp. 751–763, 1999.

[70] C. C. Kuhlthau, "Developing a model of the library search process: Cognitive and affective aspects," *Rq*, pp. 232–242, 1988.

[71] C. Allinson, "The process of audit and control–a comparison of manual and electronic information systems," *Policing: An International Journal of Police Strategies & Management*, 2004.

[72] G. Marchionini, *Information seeking in electronic environments.* Cambridge university press, 1997, no. 9.

[73] A. Spink, "Information science: A third feedback framework," *Journal of the American Society for Information Science*, vol. 48, no. 8, pp. 728–740, 1997.

[74] D. Kelly *et al.*, "Methods for evaluating interactive information retrieval systems with users," *Foundations and Trends® in Information Retrieval*, vol. 3, no. 1–2, pp. 1–224, 2009.

[75] L. Schamber and M. Eisenberg, "Relevance: The search for a definition." 1988.

[76] B. J. Jansen and S. Y. Rieh, "The seventeen theoretical constructs of information searching and information retrieval," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 8, pp. 1517–1534, 2010.

[77] P. Borlund, "The concept of relevance in ir," *Journal of the American Society for information Science and Technology*, vol. 54, no. 10, pp. 913–925, 2003.

[78] J. Mao, Y. Liu, K. Zhou, J.-Y. Nie, J. Song, M. Zhang, S. Ma, J. Sun, and H. Luo, "When does relevance mean usefulness and user satisfaction in web search?" in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval.* ACM, 2016, pp. 463–472.

[79] M. Levene, J. Bar-Ilan, and M. Zhitomirsky-Geffet, "Categorical relevance judgment," *Journal of the Association for Information Science and Technology*, 2018.

[80] G. Tsakonas and C. Papatheodorou, "Analysing and evaluating usefulness and usability in electronic information services," *Journal of information science*, vol. 32, no. 5, pp. 400–419, 2006.

[81] B. Hjørland, "The foundation of the concept of relevance," *Journal of the american society for information science and technology*, vol. 61, no. 2, pp. 217–237, 2010.

[82] Y. Xu and Z. Chen, "Relevance judgment: What do information users consider beyond topicality?" *Journal of the American Society for Information Science and Technology*, vol. 57, no. 7, pp. 961–973, 2006.

[83] B. Hjørland and F. S. Christensen, "Work tasks and socio-cognitive relevance: A specific example," *Journal of the American Society for Information Science and Technology*, vol. 53, no. 11, pp. 960–965, 2002.

[84] L. Kane, J. Carthy, and J. Dunnion, "Readability applied to information retrieval," in *European Conference on Information Retrieval.* Springer, 2006, pp. 523–526.

[85] T. Saracevic, "Relevance reconsidered," in *Proceedings of the second conference on conceptions of library and information science (CoLIS 2)*, 1996, pp. 201–218.

[86] H. D. White, "Combining bibliometrics, information retrieval, and relevance theory, part 1: First examples of a synthesis," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 4, pp. 536–559, 2007.

[87] J. Kekäläinen, "Binary and graded relevance in ir evaluations—comparison of the effects on ranking of ir systems," *Information processing & management*, vol. 41, no. 5, pp. 1019–1033, 2005.

[88] S. E. Robertson, E. Kanoulas, and E. Yilmaz, "Extending average precision to graded relevance judgments," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.* ACM, 2010, pp. 603–610.

[89] L. Lerche and D. Jannach, "Using graded implicit feedback for bayesian personalized ranking," in *Proceedings of the 8th ACM Conference on Recommender Systems*, ser. RecSys '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 353–356. [Online]. Available: https://doi.org/10.1145/2645710.2645759

[90] Y. Moshfeghi, L. R. Pinto, F. E. Pollick, and J. M. Jose, "Understanding relevance: An fmri study," in *European conference on information retrieval.* Springer, 2013, pp. 14–25.

[91] W. Qiying, M. Halvey, and R. Villa, "Video test collection with graded relevance assessments," in *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval.* ACM, 2016, pp. 309–312.

[92] L. Schamber, "Relevance and Information Behavior," *Annual review of information science and technology (ARIST)*, vol. 29, pp. 3–48, 1994.

[93] S. Mizzaro, "How many relevances in information retrieval?" *Interacting with Computers*, vol. 10, no. 3, pp. 303–320, 1998.

[94] J. Gwizdka, R. Hosseini, M. Cole, and S. Wang, "Temporal dynamics of eyetracking and eeg during reading and relevance decisions," *Journal of the Association for Information Science and Technology*, vol. 68, no. 10, pp. 2299–2312, 2017.

[95] E. Mohedano, K. McGuinness, G. Healy, N. E. O'Connor, A. F. Smeaton, A. Salvador, S. Porta, and X. Giró-i Nieto, "Exploring eeg for object detection and retrieval," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 2015, pp. 591–594.

[96] J. Rocchio, "Relevance feedback in information retrieval," *The Smart retrieval system-experiments in automatic document processing*, pp. 313–323, 1971.

[97] F. Wissbrock, "Information need assessment in information retrieval; beyond lists and queries," in *27th German Conference on Artificial Intelligence, KI2004, University of Ulm, Germany*, 2004.

[98] N. J. Belkin *et al.*, "Interaction with texts: Information retrieval as information seeking behavior," *Information retrieval*, vol. 93, pp. 55–66, 1993.

[99] I. Ruthven and M. Lalmas, "A survey on the use of relevance feedback for information access systems," *The Knowledge Engineering Review*, vol. 18, no. 2, pp. 95–145, 2003.

[100] R. R. Korfhage, *Information storage and retrieval.* John Wiley & Sons, Inc., 1997.

[101] M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski, "Predicting term-relevance from brain signals," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 425–434.

[102] Y. Moshfeghi, P. Triantafillou, and F. Pollick, "Towards predicting a realisation of an information need based on brain signals," in *The World Wide Web Conference*, 2019, pp. 1300–1309.

[103] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White, "Evaluating implicit measures to improve web search," *ACM Transactions on Information Systems (TOIS)*, vol. 23, no. 2, pp. 147–168, 2005.

[104] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *ACM SIGIR Forum*, vol. 51, no. 1. Acm New York, NY, USA, 2017, pp. 4–11.

[105] G. Jacucci, O. Barral, P. Daee, M. Wenzel, B. Serim, T. Ruotsalo, P. Pluchino, J. Freeman, L. Gamberini, S. Kaski *et al.*, "Integrating neurophysiologic relevance feedback in intent modeling for information retrieval," *Journal of the Association for Information Science and Technology*, vol. 70, no. 9, pp. 917–930, 2019.

[106] R. W. White, I. Ruthven, and J. M. Jose, "The use of implicit evidence for relevance feedback in web retrieval," in *European Conference on Information Retrieval*. Springer, 2002, pp. 93–109.

[107] D. Kelly and X. Fu, "Elicitation of term relevance feedback: an investigation of term source and context," in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 2006, pp. 453–460.

[108] O. Barral, I. Kosunen, T. Ruotsalo, M. M. Spapé, M. J. Eugster, N. Ravaja, S. Kaski, and G. Jacucci, "Bci for physiological text annotation," in *Proceedings of the 2017 ACM Workshop on An Application-oriented Approach to BCI out of the laboratory*, 2017, pp. 9–13.

[109] I. Arapakis, J. M. Jose, and P. D. Gray, "Affective feedback: an investigation into the role of emotions in the information seeking process," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 395–402.

[110] G. Buscher, A. Dengel, R. Biedert, and L. V. Elst, "Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 1, no. 2, pp. 1–30, 2012.

[111] D. Kelly and J. Teevan, "Implicit feedback for inferring user preference: a bibliography," in *Acm Sigir Forum*, vol. 37, no. 2. ACM New York, NY, USA, 2003, pp. 18–28.

[112] G. Dupret and C. Liao, “A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine,” in *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 181–190.

[113] M. Kellar, C. Watters, J. Duffy, and M. Shepherd, “Effect of task on time spent reading as an implicit measure of interest,” *Proceedings of the American Society for Information Science and Technology*, vol. 41, no. 1, pp. 168–175, 2004.

[114] J. Liu and N. J. Belkin, “Personalizing information retrieval for multi-session tasks: The roles of task stage and task type,” in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR ’10. New York, NY, USA: Association for Computing Machinery, 2010, p. 26–33. [Online]. Available: https://doi.org/10.1145/1835449.1835457

[115] I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose, “Enriching user profiling with affective features for the improvement of a multimodal recommender system,” in *Proceedings of the ACM international conference on image and video retrieval*, 2009, pp. 1–8.

[116] I. Arapakis, K. Athanasakos, and J. M. Jose, “A comparison of general vs personalised affective models for the prediction of topical relevance,” in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 2010, pp. 371–378.

[117] L. Lorigo, M. Haridasan, H. Brynjarsdóttir, L. Xia, T. Joachims, G. Gay, L. Granka, F. Pellacini, and B. Pan, “Eye tracking and online search: Lessons learned and challenges ahead,” *Journal of the American Society for Information Science and Technology*, vol. 59, no. 7, pp. 1041–1052, 2008.

[118] I. Arapakis, I. Konstas, and J. M. Jose, “Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance,” in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 461–470.

[119] K. M. Davis, L. Kangassalo, M. Spapé, and T. Ruotsalo, "Brainsourcing: Crowdsourcing recognition tasks via collaborative brain-computer interfacing," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–14. [Online]. Available: https://doi.org/10.1145/3313831.3376288

[120] M. Claypool, P. Le, M. Wased, and D. Brown, "Implicit interest indicators," in *Proceedings of the 6th international conference on Intelligent user interfaces*, 2001, pp. 33–40.

[121] M. Morita and Y. Shinoda, "Information filtering based on user behavior analysis and best match text retrieval," in *SIGIR'94*. Springer, 1994, pp. 272–281.

[122] J. Gwizdka, C. Liu, N. J. Belkin, and J. Liu, "Predicting task difficulty for different task types," in *In Proceedings of ASIS&T'10 (pnp*. Citeseer, 2010.

[123] D. Kelly and N. J. Belkin, "Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001, pp. 408–409.

[124] D. Kelly and N. J. Belkin, "Display time as implicit feedback: Understanding task effects," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 377–384. [Online]. Available: https://doi.org/10.1145/1008992.1009057

[125] R. W. White and D. Kelly, "A study on the effects of personalization and task information on implicit feedback performance," in *Proceedings of the 15th ACM international conference on Information and knowledge management*, 2006, pp. 297–306.

[126] L. Cooke, "Is the mouse a" poor man's eye tracker"?" in *Annual Conference-Society for Technical Communication*, vol. 53, 2006, p. 252.

[127] Q. Guo and E. Agichtein, "Towards predicting web searcher gaze position from mouse movements," in *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 3601–3606. [Online]. Available: https://doi.org/10.1145/1753846.1754025

[128] M. D. Smucker, X. S. Guo, and A. Toulis, "Mouse movement during relevance judging: Implications for determining user attention," in *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, ser. SIGIR '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 979–982. [Online]. Available: https://doi.org/10.1145/2600428.2609489

[129] R. W. White and G. Buscher, "Text selections as implicit relevance feedback," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 1151–1152.

[130] T. Yang, C. Luo, H. Lu, P. Gupta, B. Yin, and Q. Ai, "Can clicks be both labels and features? unbiased behavior feature collection and uncertainty-aware learning to rank," 2022.

[131] O. Barral, M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, N. Ravaja, S. Kaski, and G. Jacucci, "Exploring peripheral physiology as a predictor of perceived relevance in information retrieval," in *Proceedings of the 20th international conference on intelligent user interfaces*, 2015, pp. 389–399.

[132] J. Gwizdka, "Differences in reading between word search and information relevance decisions: evidence from eye-tracking," in *Information Systems and Neuroscience*. Springer, 2017, pp. 141–147.

[133] M. A. Just and P. A. Carpenter, "A theory of reading: From eye fixations to comprehension." *Psychological review*, vol. 87, no. 4, p. 329, 1980.

[134] J. Simola, J. Salojärvi, and I. Kojo, "Using hidden markov model to uncover processing states from eye movements in information search tasks," *Cognitive systems research*, vol. 9, no. 4, pp. 237–251, 2008.

[135] J. Gwizdka, "Characterizing relevance with eye-tracking measures," in *Proceedings of the 5th information interaction in context symposium*, 2014, pp. 58–67.

[136] J. Gwizdka and Y. Zhang, "Differences in eye-tracking measures between visits and revisits to relevant and irrelevant web pages," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015, pp. 811–814.

[137] K. Preuschoff, B. M. t Hart, and W. Einhauser, "Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making," *Frontiers in neuroscience*, vol. 5, p. 115, 2011.

[138] W. Einhauser, C. Koch, and O. Carter, "Pupil dilation betrays the timing of decisions," *Frontiers in human neuroscience*, vol. 4, p. 18, 2010.

[139] E. H. Hess and J. M. Polt, "Pupil size as related to interest value of visual stimuli," *Science*, vol. 132, no. 3423, pp. 349–350, 1960.

[140] H. E. Krugman, "Some applications of pupil measurement," *Journal of Marketing Research*, vol. 1, no. 4, pp. 15–19, 1964.

[141] D. Kahneman and J. Beatty, "Pupil diameter and load on memory," *Science*, vol. 154, no. 3756, pp. 1583–1585, 1966.

[142] D. Kahneman, B. Tursky, D. Shapiro, and A. Crider, "Pupillary, heart rate, and skin resistance changes during a mental task." *Journal of experimental psychology*, vol. 79, no. 1p1, p. 164, 1969.

[143] W. S. Peavler, "Pupil size, information overload, and performance differences," *Psychophysiology*, vol. 11, no. 5, pp. 559–566, 1974.

[144] B. Hoeks and W. J. Levelt, "Pupillary dilation as a measure of attention: A quantitative system analysis," *Behavior Research Methods, Instruments, & Computers*, vol. 25, no. 1, pp. 16–26, 1993.

[145] F. Onorati, R. Barbieri, M. Mauri, V. Russo, and L. Mainardi, "Characterization of affective states by pupillary dynamics and autonomic correlates," *Frontiers in neuroengineering*, vol. 6, p. 9, 2013.

[146] S. M. Wierda, H. van Rijn, N. A. Taatgen, and S. Martens, "Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution," *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8456–8460, 2012.

[147] F. T. Oliveira, A. Aula, and D. M. Russell, "Discriminating the relevance of web search results with measures of pupil size," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 2209–2212.

[148] R. S. Hessels, D. C. Niehorster, M. Nyström, R. Andersson, and I. T. Hooge, "Is the eye-movement field confused about fixations and saccades? a survey among 124 researchers," *Royal Society open science*, vol. 5, no. 8, p. 180502, 2018.

[149] P. Brooks, K. Y. Phang, R. Bradley, D. Oard, R. White, and F. Guimbretire, "Measuring the utility of gaze detection for task modeling: A preliminary study," in *Workshop on Intelligent Interfaces for Intelligent Analysis*, 2006.

[150] S. Xu, H. Jiang, and F. C. Lau, "User-oriented document summarization through vision-based eye-tracking," in *Proceedings of the 14th international conference on Intelligent user interfaces*, 2009, pp. 7–16.

[151] P. Balatsoukas and I. Ruthven, "An eye-tracking approach to the analysis of relevance judgments on the web: The case of google search engine," *Journal of the American Society for Information Science and technology*, vol. 63, no. 9, pp. 1728–1746, 2012.

[152] R. Villa and M. Halvey, "Is relevance hard work? evaluating the effort of making relevant assessments," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 2013, pp. 765–768.

[153] A. Taylor, "User relevance criteria choices and the information search process," *Information Processing & Management*, vol. 48, no. 1, pp. 136–153, 2012.

[154] K. Puolamäki, J. Salojärvi, E. Savia, J. Simola, and S. Kaski, "Combining eye movements and collaborative filtering for proactive information retrieval," in *Proceedings of the 28th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, 2005, pp. 146–153.

[155] M. Kandemir and S. Kaski, "Learning relevance from natural eye movements in pervasive interfaces," in *Proceedings of the 14th ACM international conference on Multimodal interaction*, 2012, pp. 85–92.

[156] A. Klami, C. Saunders, T. E. de Campos, and S. Kaski, "Can relevance of images be inferred from eye movements?" in *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, 2008, pp. 134–140.

[157] A. Ajanki, D. R. Hardoon, S. Kaski, K. Puolamäki, and J. Shawe-Taylor, "Can eyes reveal interest? implicit queries from gaze patterns," *User Modeling and User-Adapted Interaction*, vol. 19, no. 4, pp. 307–339, 2009.

[158] M.-C. Marcos, F. Gavin, and I. Arapakis, "Effect of snippets on user experience in web search," in *Proceedings of the XVI International Conference on Human Computer Interaction*, 2015, pp. 1–8.

[159] K. K. Moe, J. M. Jensen, and B. Larsen, "A qualitative look at eye-tracking for implicit relevance feedback," in *Proceedings of the Workshop on Context-Based Information Retrieval*, vol. 326, 2007, pp. 36–47.

[160] K. Rayner, "Eye movements in reading and information processing: 20 years of research." *Psychological bulletin*, vol. 124, no. 3, p. 372, 1998.

[161] R. W. Picard, A. Wexelblat, and C. I. N. I. Clifford I. Nass, "Future interfaces: social and emotional," in *CHI'02 extended abstracts on Human factors in computing systems*, 2002, pp. 698–699.

[162] B. Reeves and C. Nass, "The media equation: How people treat computers, television, and new media like real people," *Cambridge, UK*, vol. 10, p. 236605, 1996.

[163] A. Vinciarelli, N. Suditu, and M. Pantic, "Implicit human-centered tagging," in *2009 IEEE International Conference on Multimedia and Expo.* IEEE, 2009, pp. 1428–1431.

[164] Y. Moshfeghi, "Affective adaptive retrieval: study of emotion in adaptive retrieval," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 2009, pp. 852–852.

[165] Y. Moshfeghi and J. M. Jose, "Role of emotion in information retrieval for entertainment (position paper)," in *Searching4FUN Workshop in ECIR*, 2012.

[166] M. Tkalčič, U. Burnik, and A. Košir, "Using affective parameters in a content-based recommender system for images," *User Modeling and User-Adapted Interaction*, vol. 20, no. 4, pp. 279–311, 2010.

[167] I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose, "Integrating facial expressions into user profiling for the improvement of a multimodal recommender system," in *2009 IEEE International Conference on Multimedia and Expo.* IEEE, 2009, pp. 1440–1443.

[168] Y. Moshfeghi and J. M. Jose, "Role of emotional features in collaborative recommendation," in *European Conference on Information Retrieval.* Springer, 2011, pp. 738–742.

[169] ——, "On cognition, emotion, and interaction aspects of search tasks with different search intentions," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 931–942.

[170] J.-E. Golenia, M. A. Wenzel, M. Bogojeski, and B. Blankertz, “Implicit relevance feedback from electroencephalography and eye tracking in image search,” *Journal of neural engineering*, vol. 15, no. 2, p. 026002, 2018.

[171] D. S. Bassett and M. S. Gazzaniga, “Understanding complexity in the human brain,” *Trends in cognitive sciences*, vol. 15, no. 5, pp. 200–209, 2011.

[172] A. Lenartowicz and R. Poldrack, “Brain imaging,” 2017.

[173] S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank, “Brain magnetic resonance imaging with contrast dependent on blood oxygenation,” *proceedings of the National Academy of Sciences*, vol. 87, no. 24, pp. 9868–9872, 1990.

[174] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, “Neurophysiological investigation of the basis of the fmri signal,” *nature*, vol. 412, no. 6843, pp. 150–157, 2001.

[175] S. A. Huettel, A. W. Song, G. McCarthy *et al.*, *Functional magnetic resonance imaging*. Sinauer Associates Sunderland, MA, 2004, vol. 1.

[176] D. J. Greene, K. J. Black, and B. L. Schlaggar, “Considerations for mri study design and implementation in pediatric and clinical populations,” *Developmental cognitive neuroscience*, vol. 18, pp. 101–112, 2016.

[177] X. Gui, C. Chuansheng, L. Zhong-Lin, and D. Qi, “Brain imaging techniques and their applications in decision-making research,” *Xin li xue bao. Acta psychologica Sinica*, vol. 42, no. 1, p. 120, 2010.

[178] N. K. Logothetis, “What we can do and what we cannot do with fmri,” *Nature*, vol. 453, no. 7197, pp. 869–878, 2008.

[179] S. P. Singh, “Magnetoencephalography: basic principles,” *Annals of Indian Academy of Neurology*, vol. 17, no. Suppl 1, p. S107, 2014.

[180] M. Antonakakis, S. Schrader, Ü. Aydin, A. Khan, J. Gross, M. Zervakis, S. Rampp, and C. H. Wolters, “Inter-subject variability of skull conductivity and

thickness in calibrated realistic head models," *Neuroimage*, vol. 223, p. 117353, 2020.

[181] S. Vallaghé and M. Clerc, "A global sensitivity analysis of three-and four-layer eeg conductivity models," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 988–995, 2008.

[182] J. Vorwerk, Ü. Aydin, C. H. Wolters, and C. R. Butson, "Influence of head tissue conductivity uncertainties on eeg dipole reconstruction," *Frontiers in neuroscience*, vol. 13, p. 531, 2019.

[183] A. Gutschalk, "Meg auditory research," *Magnetoencephalography: from signals to dynamic cortical networks*, pp. 907–941, 2019.

[184] J. C. Mosher and M. Funke, "Towards best practices in clinical magnetoencephalography: patient preparation and data acquisition," *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, vol. 37, no. 6, p. 498, 2020.

[185] S. P. Ahlfors and M. Mody, "Overview of meg," *Organizational research methods*, vol. 22, no. 1, pp. 95–115, 2019.

[186] R. M. Hill, E. Boto, N. Holmes, C. Hartley, Z. A. Seedat, J. Leggett, G. Roberts, V. Shah, T. M. Tierney, M. W. Woolrich *et al.*, "A tool for functional brain imaging with lifespan compliance," *Nature communications*, vol. 10, no. 1, pp. 1–11, 2019.

[187] D. A. Kaiser, "Basic principles of quantitative eeg," *Journal of Adult Development*, vol. 12, no. 2, pp. 99–104, 2005.

[188] M. Lehtinen, K. Forsmart, J. Malmivuo, and H. Eskola, "Effect of skull and scalp thickness on eeg," *MEDICAL AND BIOLOGICAL ENGINEERING AND COMPUTING*, vol. 34, pp. 263–264, 1996.

[189] R. Srinivasan, "Methods to improve the spatial resolution of eeg," *International journal of bioelectromagnetism*, vol. 1, no. 1, pp. 102–111, 1999.

[190] Z. Bian, Q. Li, L. Wang, C. Lu, S. Yin, and X. Li, "Relative power and coherence of eeg series are related to amnestic mild cognitive impairment in diabetes," *Frontiers in aging neuroscience*, vol. 6, p. 11, 2014.

[191] M. Teplan, "Fundamental of eeg measurement‖ measurement science review," *Measurement Science Review*, vol. 2, 2002.

[192] C. S. Nayak and A. C. Anilkumar, "Eeg normal waveforms," 2019.

[193] X. Lei and K. Liao, "Understanding the influences of eeg reference: a large-scale brain network perspective," *Frontiers in neuroscience*, vol. 11, p. 205, 2017.

[194] F. Amzica and M. Steriade, "Electrophysiological correlates of sleep delta waves," *Electroencephalography and clinical neurophysiology*, vol. 107, no. 2, pp. 69–83, 1998.

[195] N. A. Busch, J. Dubois, and R. VanRullen, "The phase of ongoing eeg oscillations predicts visual perception," *Journal of neuroscience*, vol. 29, no. 24, pp. 7869–7876, 2009.

[196] P. Mussel, N. Ulrich, J. J. Allen, R. Osinsky, and J. Hewig, "Patterns of theta oscillation reflect the neural basis of individual differences in epistemic motivation," *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[197] M. Z. Zakrzewska and A. Brzezicka, "Working memory capacity as a moderator of load-related frontal midline theta variability in sternberg task," *Frontiers in human neuroscience*, vol. 8, p. 399, 2014.

[198] L. Dugué and R. VanRullen, "The dynamics of attentional sampling during visual search revealed by fourier analysis of periodic noise interference," *Journal of vision*, vol. 14, no. 2, pp. 11–11, 2014.

[199] S. Hanslmayr, W. Klimesch, P. Sauseng, W. Gruber, M. Doppelmayr, R. Freunberger, and T. Pecherstorfer, "Visual discrimination performance is related to decreased alpha amplitude but increased phase locking," *Neuroscience letters*, vol. 375, no. 1, pp. 64–68, 2005.

[200] J. Gross, F. Schmitz, I. Schnitzler, K. Kessler, K. Shapiro, B. Hommel, and A. Schnitzler, "Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans," *Proceedings of the national Academy of Sciences*, vol. 101, no. 35, pp. 13 050–13 055, 2004.

[201] M. Kalaivani, V. Kalaivani, and V. A. Devi, "Analysis of eeg signal for the detection of brain abnormalities," *International Journal of Computer Applications (ijca)*, vol. 1, no. 2, pp. 1–6, 2014.

[202] A. Seal, P. P. N. Reddy, P. Chaithanya, A. Meghana, K. Jahnavi, O. Krejcar, and R. Hudak, "An eeg database and its initial benchmark emotion classification performance," *Computational and mathematical methods in medicine*, vol. 2020, 2020.

[203] C. S. Herrmann and R. T. Knight, "Mechanisms of human attention: event-related potentials and oscillations," *Neuroscience & Biobehavioral Reviews*, vol. 25, no. 6, pp. 465–476, 2001.

[204] S. J. Luck, *An introduction to the event-related potential technique.* MIT press, 2014.

[205] H. H. Kim and Y. H. Kim, "Erp/mmr algorithm for classifying topic-relevant and topic-irrelevant visual shots of documentary videos," *Journal of the Association for Information Science and Technology*, vol. 70, no. 9, pp. 931–941, 2019.

[206] Z. Pinkosova, W. J. McGeown, and Y. Moshfeghi, "The cortical activity of graded relevance," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 299–308. [Online]. Available: https://doi.org/10.1145/3397271.3401106

[207] J. L. Brooks, A. Zoumpoulaki, and H. Bowman, "Data-driven region-of-interest selection without inflating type i error rate," *Psychophysiology*, vol. 54, no. 1, pp. 100–113, 2017.

[208] T. Castermans, M. Duvinage, G. Cheron, and T. Dutoit, "Towards effective non-invasive brain-computer interfaces dedicated to gait rehabilitation systems," *Brain sciences*, vol. 4, no. 1, pp. 1–48, 2013.

[209] M. Van Steen and G. Kristo, "Contribution to roadmap," 2015.

[210] G. Felsen and Y. Dan, "A natural approach to studying vision," *Nature neuroscience*, vol. 8, no. 12, pp. 1643–1646, 2005.

[211] D. Rasmussen and C. Eliasmith, "Modeling brain function: current developments and future prospects," *JAMA neurology*, vol. 70, no. 10, pp. 1325–1329, 2013.

[212] M. D. Smucker, X. S. Guo, and A. Toulis, "Mouse movement during relevance judging: implications for determining user attention," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 979–982.

[213] D. A. Michel, "What is used during cognitive processing in information retrieval and library searching? eleven sources of search information," *Journal of the American Society for Information Science*, vol. 45, no. 7, pp. 498–514, 1994.

[214] P. Ingwersen and P. Willett, "An introduction to algorithmic and cognitive approaches for information retrieval," 1995.

[215] H. Hyman, *Learning and Relevance in Information Retrieval: A Study in the Application of Exploration and User Knowledge to Enhance Performance.* University of South Florida, 2012.

[216] J. Gwizdka, "Inferring web page relevance using pupillometry and single channel eeg," in *Information Systems and Neuroscience.* Springer, 2018, pp. 175–183.

[217] D. Marschalek, "A review of basic cognitive processes and their relevance to understanding responses to works of art," *Visual Arts Research*, vol. 9, no. 1 (17, pp. 23–33, 1983.

[218] Y. Wang, "On the cognitive processes of human perception with emotions, motivations, and attitudes," *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, vol. 1, no. 4, pp. 1–13, 2007.

[219] D. Bernstein, *Essentials of psychology.* Cengage learning, 2018.

[220] R. Gregory, "Perception in gregory," 1987.

[221] O. Qiong, "A brief introduction to perception," *Studies in literature and language*, vol. 15, no. 4, pp. 18–28, 2017.

[222] R. A. Kinchla and J. M. Wolfe, "The order of visual processing:"top-down," "bottom-up," or "middle-out"," *Perception & psychophysics*, vol. 25, no. 3, pp. 225–231, 1979.

[223] N. Shea, "Distinguishing top-down from bottom-up effects," *Perception and its modalities*, pp. 73–91, 2015.

[224] L. Melloni, S. van Leeuwen, A. Alink, and N. G. Müller, "Interaction between bottom-up saliency and top-down control: how saliency maps are created in the human brain," *Cerebral cortex*, vol. 22, no. 12, pp. 2943–2952, 2012.

[225] R. L. Gregory, *Concepts and mechanisms of perception.* Charles Scribner's Sons, 1974.

[226] R. J. Sternberg, *Cognitive psychology.* Wadsworth-Thomson learning, 2003.

[227] M. M. Sohlberg and C. A. Mateer, "Effectiveness of an attention-training program," *Journal of clinical and experimental neuropsychology*, vol. 9, no. 2, pp. 117–130, 1987.

[228] M. M. Sohlberg and C. Mateer, *Cognitive rehabilitation: An integrative neuropsychological approach.* Guilford Press, 2001.

[229] S. Zickefoose, K. Hux, J. Brown, and K. Wulf, "Let the games begin: A preliminary study using attention process training-3 and lumosity™ brain games

to remediate attention deficits following traumatic brain injury," *Brain injury*, vol. 27, no. 6, pp. 707–716, 2013.

[230] C. Shen, F. C. Popescu, E. Hahn, T. T. Ta, M. Dettling, and A. H. Neuhaus, "Neurocognitive pattern analysis reveals classificatory hierarchy of attention deficits in schizophrenia," *Schizophrenia bulletin*, vol. 40, no. 4, pp. 878–885, 2014.

[231] W. Wundt, "Outlines of psychology," in *Wilhelm Wundt and the making of a scientific psychology*. Springer, 1980, pp. 179–195.

[232] W. Prinzmetal, C. McCool, and S. Park, "Attention: reaction time and accuracy reveal different mechanisms." *Journal of Experimental Psychology: General*, vol. 134, no. 1, p. 73, 2005.

[233] K. N. Nguyen, T. Watanabe, and G. J. Andersen, "Role of endogenous and exogenous attention in task-relevant visual perceptual learning," *Plos one*, vol. 15, no. 8, p. e0237912, 2020.

[234] M. I. Posner and S. J. Boies, "Components of attention." *Psychological review*, vol. 78, no. 5, p. 391, 1971.

[235] G. Leisman, A. A. Moustafa, and T. Shafir, "Thinking, walking, talking: integratory motor and cognitive brain function," *Frontiers in public health*, p. 94, 2016.

[236] A. W. Melton, "Implications of short-term memory for a general theory of memory," *Journal of verbal Learning and verbal Behavior*, vol. 2, no. 1, pp. 1–21, 1963.

[237] D. Tromp, A. Dufour, S. Lithfous, T. Pebayle, and O. Després, "Episodic memory in normal aging and alzheimer disease: Insights from imaging and behavioral studies," *Ageing research reviews*, vol. 24, pp. 232–262, 2015.

[238] K. B. McDermott and H. L. Roediger, "Memory (encoding, storage, retrieval)," *General Psychology FA2018. Noba Project: Milwaukie, OR*, pp. 117–153, 2018.

[239] M. D. Kopelman, B. Wilson, and A. D. Baddeley, "The autobiographical memory interview: a new assessment of autobiographical and personal semantic memory in amnesic patients," *Journal of clinical and experimental neuropsychology*, vol. 11, no. 5, pp. 724–744, 1989.

[240] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological review*, vol. 63, no. 2, p. 81, 1956.

[241] N. Cowan, "What are the differences between long-term, short-term, and working memory?" *Progress in brain research*, vol. 169, pp. 323–338, 2008.

[242] A. Diamond, "Executive functions," *Annual review of psychology*, vol. 64, pp. 135–168, 2013.

[243] R. Malenka, E. Nestler, and S. Hyman, "Higher cognitive function and behavioral control," *Molecular neuropharmacology: A foundation for clinical neuroscience*, pp. 313–321, 2009.

[244] R. C. Atkinson and R. M. Shiffrin, "Human memory: A proposed system and its control processes," in *Psychology of learning and motivation*. Elsevier, 1968, vol. 2, pp. 89–195.

[245] D. S. Roy, S. Muralidhar, L. M. Smith, and S. Tonegawa, "Silent memory engrams as the basis for retrograde amnesia," *Proceedings of the National Academy of Sciences*, vol. 114, no. 46, pp. E9972–E9979, 2017.

[246] E. B. Goldsmith, *Social influence and sustainable consumption*. Springer, 2015.

[247] R. E. Mayer, "Problem solving," *Encyclopedia of creativity*, vol. 2, pp. 437–447, 1999.

[248] K.-C. Tönnsen, "The relevance of trial-and-error: Can trial-and-error be a sufficient learning method in technical problem-solving-contexts?" *Techne serien-Forskning i slöjdpedagogik och slöjdvetenskap*, vol. 28, no. 2, pp. 303–312, 2021.

[249] F. Huang, S. Tang, and Z. Hu, "Unconditional perseveration of the short-term mental set in chunk decomposition," *Frontiers in Psychology*, p. 2568, 2018.

[250] J. M. Lang, J. D. Ford, and M. M. Fitzgerald, "An algorithm for determining use of trauma-focused cognitive–behavioral therapy." *Psychotherapy: Theory, Research, Practice, Training*, vol. 47, no. 4, p. 554, 2010.

[251] V. Sarathy, "Real world problem-solving," *Frontiers in human neuroscience*, vol. 12, p. 261, 2018.

[252] J. Bendor and T. H. Hammond, "Rethinking allison's models," *American Political Science Review*, vol. 86, no. 2, pp. 301–322, 1992.

[253] A. D. Galinsky and G. B. Moskowitz, "Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives," *Journal of Experimental Social Psychology*, vol. 36, no. 4, pp. 384–409, 2000.

[254] Y. Zhao, S. Tu, M. Lei, J. Qiu, O. Ybarra, and Q. Zhang, "The neural basis of breaking mental set: an event-related potential study," *Experimental brain research*, vol. 208, no. 2, pp. 181–187, 2011.

[255] H. A. Simon, "The new science of management decision." 1960.

[256] E. Dane and M. G. Pratt, "Exploring intuition and its role in managerial decision making," *Academy of management review*, vol. 32, no. 1, pp. 33–54, 2007.

[257] G. Calabretta, G. Gemser, and N. M. Wijnberg, "The interplay between intuition and rationality in strategic decision making: A paradox perspective," *Organization Studies*, vol. 38, no. 3-4, pp. 365–401, 2017.

[258] A. Malmir and S. Shoorcheh, "An investigation of the impact of teaching critical thinking on the iranian efl learners' speaking skill," *Journal of Language Teaching and Research*, vol. 3, no. 4, pp. 608–617, 2012.

[259] R. Atkinson, S. Nolen-Hoeksema, B. Fredrickson, G. Loftus, and C. Lutz, "Atkinson & hilgard's introduction to psychology," *Translated by Arjmand, M. & Rafiee, H. Tehran: Arjmand*, 2003.

[260] S. Schmid, D. A. Wilson, and C. H. Rankin, "Habituation mechanisms and their importance for cognitive function," *Frontiers in integrative neuroscience*, vol. 8, p. 97, 2015.

[261] I. Rehman, N. Mahabadi, T. Sanvictores, and C. I. Rehman, "Classical conditioning," 2017.

[262] E. Segers, T. Beckers, H. Geurts, L. Claes, M. Danckaerts, and S. Van der Oord, "Working memory and reinforcement schedule jointly determine reinforcement learning in children: Potential implications for behavioral parent training," *Frontiers in Psychology*, vol. 9, p. 394, 2018.

[263] R. E. Mayer, *Cognitive Learning.* Boston, MA: Springer US, 2012, pp. 594–596. [Online]. Available: https://doi.org/10.1007/978-1-4419-1428-6_390

[264] J. Zlatev and J. Blomberg, "Language may indeed influence thought," *Frontiers in psychology*, vol. 6, p. 1631, 2015.

[265] T. M. Holtgraves, *Language as social action: Social psychology and language use.* Psychology Press, 2013.

[266] D. T. Willingham and C. Riener, *Cognition: The thinking animal.* Cambridge University Press, 2019.

[267] E. M. Fernández, H. S. Cairns, and J. Wiley, *The handbook of psycholinguistics.* Wiley Online Library, 2018.

[268] M. T. Banich, "Executive function: The search for an integrated account," *Current directions in psychological science*, vol. 18, no. 2, pp. 89–94, 2009.

[269] S. J. Gilbert and P. W. Burgess, "Executive function," *Current biology*, vol. 18, no. 3, pp. R110–R114, 2008.

[270] J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar *et al.*, "Functional network organization of the human brain," *Neuron*, vol. 72, no. 4, pp. 665–678, 2011.

[271] J.-E. Golenia, M. Wenzel, and B. Blankertz, "Live demonstrator of eeg and eye-tracking input for disambiguation of image search results," in *International Workshop on Symbiotic Interaction*. Springer, 2015, pp. 81–86.

[272] J. Gwizdka and J. Mostafa, "Neuroiir: Challenges in bringing neuroscience to research in human-information interaction," in *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, ser. CHIIR '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 437–438. [Online]. Available: https://doi.org/10.1145/3020165.3022165

[273] K. Kingphai and Y. Moshfeghi, "Mental workload prediction level from eeg signals using deep learning models," 2021.

[274] ——, "On time series cross-validation for deep learning classification model of mental workload levels based on eeg signals," in *Advanced Online & Onsite Course & Symposium on Artificial Intelligence & Neuroscience*, 2022.

[275] Y. Moshfeghi, P. Triantafillou, and F. E. Pollick, "Understanding information need: An fmri study," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 335–344.

[276] Y. Moshfeghi and F. E. Pollick, "Search process as transitions between neural states," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 1683–1692.

[277] ——, "Neuropsychological model of the realization of information need," *J. Assoc. Inf. Sci. Technol.*, vol. 70, no. 9, p. 954–967, aug 2019. [Online]. Available: https://doi.org/10.1002/asi.24242

[278] J.-P. Kauppi, M. Kandemir, V.-M. Saarinen, L. Hirvenkari, L. Parkkonen, A. Klami, R. Hari, and S. Kaski, "Towards brain-activity-controlled information retrieval: Decoding image relevance from meg signals," *NeuroImage*, vol. 112, pp. 288–298, 2015.

[279] O. Barral *et al.*, "Implicit interaction with textual information using physiological signals," 2018.

[280] A. Frey, G. Ionescu, B. Lemaire, F. López-Orozco, T. Baccino, and A. Guérin-Dugué, "Decision-making in information seeking on texts: an eye-fixation-related potentials investigation," *Frontiers in systems neuroscience*, vol. 7, p. 39, 2013.

[281] K. Kingphai and Y. Moshfeghi, "On eeg preprocessing role in deep learning effectiveness for mental workload classification," in *International Symposium on Human Mental Workload: Models and Applications*. Springer, 2021, pp. 81–98.

[282] G. Slanzi, J. A. Balazs, and J. D. Velsquez, "Combining eye tracking, pupil dilation and eeg analysis for predicting web users click intention," *Inf. Fusion*, vol. 35, no. C, p. 51–57, may 2017. [Online]. Available: https://doi.org/10.1016/j.inffus.2016.09.003

[283] M. A. Wenzel, M. Bogojeski, and B. Blankertz, "Real-time inference of word relevance from electroencephalogram and eye gaze," *Journal of neural engineering*, vol. 14, no. 5, p. 056007, 2017.

[284] M. J. Weber and S. L. Thompson-Schill, "Functional neuroimaging can support causal claims about brain function," *Journal of cognitive neuroscience*, vol. 22, no. 11, pp. 2415–2416, 2010.

[285] J. Gwizdka, J. Mostafa, Y. Moshfeghi, O. Bergman, and F. E. Pollick, "Applications of neuroimaging in information science: Challenges and opportunities," *Proceedings of the American Society for Information Science and Technology*, vol. 50, no. 1, pp. 1–4, 2013.

[286] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio *et al.*, "The berlin brain–computer interface: non-medical uses of bci technology," *Frontiers in neuroscience*, p. 198, 2010.

[287] K.-R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz, "Machine learning for real-time single-trial eeg-analysis: from brain–computer interfacing to mental state monitoring," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 82–90, 2008.

[288] L. Kangassalo, M. Spapé, and T. Ruotsalo, "Neuroadaptive modelling for generating images matching perceptual categories," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.

[289] S. Paisalnan, F. Pollick, and Y. Moshfeghi, "Towards understanding neuroscience of realisation of information need in light of relevance and satisfaction judgement," in *International Conference on Machine Learning, Optimization, and Data Science.* Springer, 2021, pp. 41–56.

[290] L. Kangassalo, M. Spapé, G. Jacucci, and T. Ruotsalo, "Why do users issue good queries? neural correlates of term specificity," in *Proceedings of the 42nd international acm sigir conference on research and development in information retrieval*, 2019, pp. 375–384.

[291] S. Paisalnan, Y. Moshfeghi, and F. Pollick, "Neural correlates of realisation of satisfaction in a successful search process," *Proceedings of the Association for Information Science and Technology*, vol. 58, no. 1, pp. 282–291, 2021.

[292] D. Michalkova, M. P. Rodriguez, and Y. Moshfeghi, "Information need awareness: an eeg study," in *Special Interest Group on Information Retrieval (SIGIR) 2022*, 2022.

[293] S. Paisalnan, Y. Moshfeghi, and F. E. Pollick, "Neural correlates of satisfaction of an information need," in *Advanced Online & Onsite Course & Symposium on Artificial Intelligence & Neuroscience*, 2022.

[294] Y. H. Kim and H. H. Kim, "Automatic extraction techniques of topic-relevant visual shots using realtime brainwave responses," *Journal of Korea Multimedia Society*, vol. 19, no. 8, pp. 1260–1274, 2016.

[295] H. H. Kim and Y. H. Kim, "Video summarization using event-related potential responses to shot boundaries in real-time video watching," *Journal of the Association for Information Science and Technology*, vol. 70, no. 2, pp. 164–175, 2019.

[296] A. Behneman, N. Kintz, R. Johnson, C. Berka, K. Hale, S. Fuchs, P. Axelsson, and A. Baskin, "Enhancing text-based analysis using neurophysiological measures," in *International Conference on Foundations of Augmented Cognition.* Springer, 2009, pp. 449–458.

[297] Z. Ye, X. Xie, Y. Liu, Z. Wang, X. Li, J. Li, X. Chen, M. Zhang, and S. Ma, "Why don't you click: Neural correlates of non-click behaviors in web search," *arXiv preprint arXiv:2109.10560*, 2021.

[298] C. de la Torre-Ortiz, M. M. Spapé, L. Kangassalo, and T. Ruotsalo, "Brain relevance feedback for interactive image generation," in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020, pp. 1060–1070.

[299] L. Kangassalo, M. Spapé, N. Ravaja, and T. Ruotsalo, "Information gain modulates brain activity evoked by reading," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.

[300] A. M. Elleman and E. L. Oslund, "Reading comprehension research: Implications for practice and policy," *Policy Insights from the Behavioral and Brain Sciences*, vol. 6, no. 1, pp. 3–11, 2019.

[301] J. Mikk, "Sentence length for revealing the cognitive load reversal effect in text comprehension," *Educational Studies*, vol. 34, no. 2, pp. 119–127, 2008.

[302] E. Sormunen, "Liberal relevance criteria of trec-: Counting on negligible documents?" in *SIGIR '02.* ACM, 2002, pp. 324–330.

[303] S. I. Dimitriadis, C. Salis, I. Tarnanas, and D. E. Linden, "Topological filtering of dynamic functional brain networks unfolds informative chronnectomics: a novel

data-driven thresholding scheme based on orthogonal minimal spanning trees (omsts)," *Frontiers in neuroinformatics*, vol. 11, p. 28, 2017.

[304] P. Wang, "Contextualizing user relevance criteria: A meta-ethnographic approach to user-centered relevance studies," in *IIiX '10*. NY, USA: ACM, 2010, p. 293–298.

[305] I. Ruthven, "Relevance behaviour in trec," *J DOC*, vol. 70, no. 6, pp. 1098–1117, 2014.

[306] P. Ingwersen, "Context in information interaction–revisited 2006," in *Proceedings of the Fourth Biennial DISSAnet Conference. Pretoria: University of Pretoria*, 2006, pp. 13–23.

[307] M. Sanchiz, A. Chevalier, W.-T. Fu, and F. Amadieu, "Relationships between age, domain knowledge and prior knowledge pre-activation on information searching," in *CHIIR '17*. NY, USA: ACM, 2017, p. 289–292.

[308] P. A. Alexander, D. L. Schallert, and V. C. Hare, "Coming to terms: How researchers in learning and literacy talk about knowledge," *Review of educational research*, vol. 61, no. 3, pp. 315–343, 1991.

[309] M. A. Fitzgerald, "Skills for evaluating web-based information," 2005.

[310] P. Vakkari and N. Hakala, "Changes in relevance criteria and problem stages in task performance," *J DOC*, pp. 295–310, 2000.

[311] M. Versteeg and P. Steendijk, "Putting post-decision wagering to the test: a measure of self-perceived knowledge in basic sciences?" *Perspectives on Medical Education*, vol. 8, no. 1, pp. 9–16, 2019.

[312] C. M. Radecki and J. Jaccard, "Perceptions of knowledge, actual knowledge, and information search behavior," *Journal of experimental social psychology*, vol. 31, no. 2, pp. 107–138, 1995.

[313] C.-Y. Park, "News media exposure and self-perceived knowledge: The illusion of knowing." *International Journal of Public Opinion Research*, 2001.

[314] J. Kruger and D. Dunning, "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments." *Journal of personality and social psychology*, vol. 77, no. 6, p. 1121, 1999.

[315] I. Ruthven, "Integration approaches to relevance," in *New directions in cognitive information retrieval.* Springer, 2005, pp. 61–80.

[316] R. H. Riffenburgh, *Statistics in medicine.* Elsevier, 2011.

[317] M. Corsi-Cabrera, J. Ramos, C. Arce, M. Guevara, M. Ponce-de Leon, and I. Lorenzo, "Changes in the waking eeg as a consequence of sleep and sleep deprivation," *Sleep*, vol. 15, no. 6, pp. 550–555, 1992.

[318] F. S. Gargoum and S. T. O'Keeffe, "Readability and content of patient information leaflets for endoscopic procedures," *Irish Journal of Medical Science (1971-)*, vol. 183, no. 3, pp. 429–432, 2014.

[319] K. Benoit, K. Watanabe, H. Wang, J. W. Lua, and J. Kuha, "Package 'quanteda. textstats'," *Research Bulletin*, vol. 27, no. 2, pp. 37–54, 2021.

[320] S. Sanei and J. Chambers, "Eeg signal processing, centre of digital signal processing, cardiff university, uk," *Publicação acadêmica, 1ª Edição, Elsevier*, vol. 16, 2007.

[321] M. Plöchl, J. P. Ossandón, and P. König, "Combining eeg and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data," *Frontiers in human neuroscience*, vol. 6, p. 278, 2012.

[322] J. Dien, "The neurocognitive basis of reading single words as seen through early latency erps: a model of converging pathways," *Biological psychology*, vol. 80, no. 1, pp. 10–22, 2009.

[323] E. Kaan, A. Harris, E. Gibson, and P. Holcomb, "The p600 as an index of syntactic integration difficulty," *Language and cognitive processes*, vol. 15, no. 2, pp. 159–201, 2000.

[324] E. Kaan, "Event-related potentials and language processing: A brief overview," *Language and linguistics compass*, vol. 1, no. 6, pp. 571–591, 2007.

[325] M. A. Just, P. A. Carpenter, and J. D. Woolley, "Paradigms and processes in reading comprehension." *Journal of experimental psychology: General*, vol. 111, no. 2, p. 228, 1982.

[326] O. Dimigen, W. Sommer, A. Hohlfeld, A. M. Jacobs, and R. Kliegl, "Coregistration of eye movements and eeg in natural reading: analyses and review." *Journal of Experimental Psychology: General*, vol. 140, no. 4, p. 552, 2011.

[327] S. Atir, E. Rosenzweig, and D. Dunning, "When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge," *Psychological Science*, vol. 26, no. 8, pp. 1295–1303, 2015.

[328] J. V. Bradley, "Overconfidence in ignorant experts," *Bulletin of the Psychonomic Society*, vol. 17, no. 2, pp. 82–84, 1981.

[329] J. W. C. Medithe and U. R. Nelakuditi, "Study on the impact of light on human physiology and electroencephalogram," in *Journal of Biomimetics, Biomaterials and Biomedical Engineering*, vol. 28. Trans Tech Publ, 2016, pp. 36–43.

[330] K.-K. Shieh and M.-H. Chen, "Effects of display medium and luminance contrast on concept formation and eeg response," *Perceptual and motor skills*, vol. 100, no. 3_suppl, pp. 943–954, 2005.

[331] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[332] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011.

[333] S. Wang and C. J. James, "Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis," *Computational intelligence and neuroscience*, vol. 2007, 2007.

[334] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier, "Spherical splines for scalp potential and current density mapping," *Electroencephalography and clinical neurophysiology*, vol. 72, no. 2, pp. 184–187, 1989.

[335] M. Calbi, F. Siri, K. Heimann, D. Barratt, V. Gallese, A. Kolesnikov, and M. A. Umiltà, "How context influences the interpretation of facial expressions: a source localization high-density eeg study on the "kuleshov effect"," *Scientific reports*, vol. 9, no. 1, pp. 1–16, 2019.

[336] X. Liu, H. Zhou, C. Jiang, Y. Xue, Z. Zhou, and J. Wang, "Cognitive control deficits in alcohol dependence are a trait-and state-dependent biomarker: An erp study," *Frontiers in Psychiatry*, vol. 11, p. 1389, 2020.

[337] K. J. Rothman, "No adjustments are needed for multiple comparisons," *Epidemiology*, pp. 43–46, 1990.

[338] M. Laganaro and C. Perret, "Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects," *Brain Topography*, vol. 24, no. 1, pp. 19–29, 2011.

[339] N. Kaganovich, J. Schumaker, and C. Rowland, "Matching heard and seen speech: an erp study of audiovisual word recognition," *Brain and language*, vol. 157, pp. 14–24, 2016.

[340] P. Tacikowski, H. B. Cygan, and A. Nowicka, "Neural correlates of own and close-other's name recognition: Erp evidence," *Frontiers in human neuroscience*, vol. 8, p. 194, 2014.

[341] V. Della Mea, L. Di Gaspero, and S. Mizzaro, "Evaluating adm on a four-level relevance scale document set from ntcir." in *NTCIR*, 2004.

[342] R. Green and C. A. Bean, "Topical relevance relationships. ii. an exploratory study and preliminary typology," *Journal of the American Society for Information Science*, vol. 46, no. 9, pp. 654–662, 1995.

[343] V. Calhoun, "Data-driven approaches for identifying links between brain structure and function in health and disease," *Dialogues in clinical neuroscience*, vol. 20, no. 2, p. 87, 2018.

[344] A. H. Meghdadi, M. Karić, and C. Berka, "Eeg analytics: benefits and challenges of data driven eeg biomarkers for neurodegenerative diseases," *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 1280–1285, 2019.

[345] P. LePendu, D. Dou, G. A. Frishkoff, and J. Rong, "Ontology database: A new method for semantic modeling and an application to brainwave data," in *International Conference on Scientific and Statistical Database Management*. Springer, 2008, pp. 313–330.

[346] A. M. Rutman, W. C. Clapp, J. Z. Chadick, and A. Gazzaley, "Early top–down control of visual processing predicts working memory performance," *Journal of cognitive neuroscience*, vol. 22, no. 6, pp. 1224–1234, 2010.

[347] A. Savostyanov, A. Bocharov, T. Astakhova, S. Tamozhnikov, A. Saprygin, and G. Knyazev, "The behavioral and erp responses to self-and other-referenced adjectives," *Brain Sciences*, vol. 10, no. 11, p. 782, 2020.

[348] C. Spironelli and A. Angrilli, "Complex time-dependent erp hemispheric asymmetries during word matching in phonological, semantic and orthographical matching judgment tasks," *Symmetry*, vol. 13, no. 1, p. 74, 2021.

[349] T. Curran, "Brain potentials of recollection and familiarity," *Memory & cognition*, vol. 28, no. 6, pp. 923–938, 2000.

[350] H. Yang, G. Laforge, B. Stojanoski, E. S. Nichols, K. McRae, and S. Köhler, "Late positive complex in event-related potentials tracks memory signals when they are decision relevant," *Scientific reports*, vol. 9, no. 1, pp. 1–15, 2019.

[351] S. Bouaffre and F. Faita-Ainseba, "Hemispheric differences in the time-course of semantic priming processes: Evidence from event-related potentials (erps)," *Brain and Cognition*, vol. 63, no. 2, pp. 123–135, 2007.

[352] C. W. Park, M. P. Gardner, and V. K. Thukral, "Self-perceived knowledge: Some effects on information processing for a choice task," *The American Journal of Psychology*, pp. 401–424, 1988.

[353] P. R. Pintrich, "The role of metacognitive knowledge in learning, teaching, and assessing," *Theory into practice*, vol. 41, no. 4, pp. 219–225, 2002.

[354] A. Zohar and S. Barzilai, "A review of research on metacognition in science education: Current and future directions," *Studies in Science education*, vol. 49, no. 2, pp. 121–169, 2013.

[355] T. Curran and J. Dien, "Differentiating amodal familiarity from modality-specific memory processes: An erp study," *Psychophysiology*, vol. 40, no. 6, pp. 979–988, 2003.

[356] D. Friedman and R. Johnson Jr, "Event-related potential (erp) studies of memory encoding and retrieval: A selective review," *Microscopy research and technique*, vol. 51, no. 1, pp. 6–28, 2000.

[357] S. Wang, X. Dong, Y. Ren, and Y. Yang, "The development of semantic priming effect in childhood: an event-related potential study," *NeuroReport*, vol. 20, no. 6, pp. 574–578, 2009.

[358] C. F. Tagliabue, D. Veniero, C. S. Benwell, R. Cecere, S. Savazzi, and G. Thut, "The eeg signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[359] D. M. Twomey, P. R. Murphy, S. P. Kelly, and R. G. O'Connell, "The classic p300 encodes a build-to-threshold decision variable," *European journal of neuroscience*, vol. 42, no. 1, pp. 1636–1643, 2015.

[360] P. L. Smith and R. Ratcliff, "Psychology and neurobiology of simple decisions," *Trends in neurosciences*, vol. 27, no. 3, pp. 161–168, 2004.

[361] J. Polich, "Updating p300: an integrative theory of p3a and p3b," *Clinical neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007.

[362] D. T. Stuss, T. Picton, and A. Cerri, "Searching for the names of pictures: An event-related potential study," *Psychophysiology*, vol. 23, no. 2, pp. 215–223, 1986.

[363] Z. Song, C. Liu, R. Shi, M. Zhang, H. Wang, and Y. Mei, "Neural activities during the evaluation of luxury goods-to-service brand extension: An event-related potentials (erps) study." *Journal of Neuroscience, Psychology, and Economics*, vol. 13, no. 3, p. 127, 2020.

[364] J. Dien, C. A. Michelson, and M. S. Franklin, "Separating the visual sentence n400 effect from the p400 sequential expectancy effect: Cognitive and neuroanatomical implications," *Brain research*, vol. 1355, pp. 126–140, 2010.

[365] C. J. Mueller, C. N. White, and L. Kuchinke, "Electrophysiological correlates of the drift diffusion model in visual word recognition," *Human brain mapping*, vol. 38, no. 11, pp. 5616–5627, 2017.

[366] C. Wachinger, S. Volkmer, K. Bublath, J. Bruder, J. Bartling, and G. Schulte-Körne, "Does the late positive component reflect successful reading acquisition? a longitudinal erp study," *NeuroImage: Clinical*, vol. 17, pp. 232–240, 2018.

[367] B. Bermúdez-Margaretto, D. Beltrán, F. Cuetos, and A. Domínguez, "Novel word learning: event-related brain potentials reflect pure lexical and task-related effects," *Frontiers in human neuroscience*, vol. 13, p. 347, 2019.

[368] J. Back and C. Oppenheim, "A model of cognitive load for ir: implications for user relevance feedback interaction," *Information Research*, vol. 6, no. 2, pp. 6–2, 2001.

[369] S. Halder, E. M. Hammer, S. C. Kleih, M. Bogdan, W. Rosenstiel, N. Birbaumer, and A. Kübler, "Prediction of auditory and visual p300 brain-computer interface aptitude," *PloS one*, vol. 8, no. 2, p. e53513, 2013.

[370] L. Ahmed and J. W. de Fockert, "Working memory load can both improve and impair selective attention: evidence from the navon paradigm," *Attention, Perception, & Psychophysics*, vol. 74, no. 7, pp. 1397–1405, 2012.

[371] H. M. Gray, N. Ambady, W. T. Lowenthal, and P. Deldin, "P300 as an index of attention to self-relevant stimuli," *Journal of experimental social psychology*, vol. 40, no. 2, pp. 216–224, 2004.

[372] Y. Arbel, K. M. Spencer, and E. Donchin, "The n400 and the p300 are not all that independent," *Psychophysiology*, vol. 48, no. 6, pp. 861–875, 2011.

[373] P. M. Alday and F. Kretzschmar, "Speed-accuracy tradeoffs in brain and behavior: testing the independence of p300 and n400 related processes in behavioral responses to sentence categorization," *Frontiers in human neuroscience*, vol. 13, p. 285, 2019.

[374] S. C. Steffensen, A. J. Ohran, D. N. Shipp, K. Hales, S. H. Stobbs, and D. E. Fleming, "Gender-selective effects of the p300 and n400 components of the visual evoked potential," *Vision research*, vol. 48, no. 7, pp. 917–925, 2008.

[375] C. Aurnhammer, F. Delogu, M. Schulz, H. Brouwer, and M. W. Crocker, "Retrieval (n400) and integration (p600) in expectation-based comprehension," *Plos one*, vol. 16, no. 9, p. e0257430, 2021.

[376] A. Schacht, W. Sommer, O. Shmuilovich, P. C. Martíenz, and M. Martín-Loeches, "Differential task effects on n400 and p600 elicited by semantic and syntactic violations," *PloS one*, vol. 9, no. 3, p. e91226, 2014.

[377] C. C. Duncan-Johnson and E. Donchin, "On quantifying surprise: The variation of event-related potentials with subjective probability," *Psychophysiology*, vol. 14, no. 5, pp. 456–467, 1977.

[378] B. Kopp, C. Seer, F. Lange, A. Kluytmans, A. Kolossa, T. Fingscheidt, and H. Hoijtink, "P300 amplitude variations, prior probabilities, and likelihoods: A bayesian erp study," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 16, no. 5, pp. 911–928, 2016.

[379] J. B. Debruille, "The n400 potential could index a semantic inhibition," *Brain research reviews*, vol. 56, no. 2, pp. 472–477, 2007.

[380] P. Ingwersen, "Polyrepresentation of information needs and semantic entities elements of a cognitive theory for information retrieval interaction," in *SIGIR'94*. Springer, 1994, pp. 101–110.

[381] R. Johnson Jr, "On the neural generators of the p300 component of the event-related potential," *Psychophysiology*, vol. 30, no. 1, pp. 90–97, 1993.

[382] N. R. Harrison and M. Ziessler, "Effect anticipation affects perceptual, cognitive, and motor phases of response preparation: evidence from an event-related potential (erp) study," *Frontiers in human neuroscience*, vol. 10, p. 5, 2016.

[383] T. Sitnikova, D. F. Salisbury, G. Kuperberg, and P. J. Holcomb, "Electrophysiological insights into language processing in schizophrenia," *Psychophysiology*, vol. 39, no. 6, pp. 851–860, 2002.

[384] K. L. Vilberg, R. F. Moosavi, and M. D. Rugg, "The relationship between electrophysiological correlates of recollection and amount of information retrieved," *Brain research*, vol. 1122, no. 1, pp. 161–170, 2006.

[385] K. A. Paller and M. Kutas, "Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming," *Journal of cognitive neuroscience*, vol. 4, no. 4, pp. 375–392, 1992.

[386] M. D. Rugg, R. E. Mark, P. Walla, A. M. Schloerscheidt, C. S. Birch, and K. Allan, “Dissociation of the neural correlates of implicit and explicit memory,” *Nature*, vol. 392, no. 6676, pp. 595–598, 1998.

[387] Z. Zheng, J. Li, F. Xiao, W. Ren, and R. He, “Unitization improves source memory in older adults: An event-related potential study,” *Neuropsychologia*, vol. 89, pp. 232–244, 2016.

[388] B. Woroch and B. D. Gonsalves, “Event-related potential correlates of item and source memory strength,” *Brain Research*, vol. 1317, pp. 180–191, 2010.

[389] J. B. Meixner and J. P. Rosenfeld, “Detecting knowledge of incidentally acquired, real-world memories using a p300-based concealed-information test,” *Psychological science*, vol. 25, no. 11, pp. 1994–2005, 2014.

[390] J. L. Voss and K. D. Federmeier, “Fn400 potentials are functionally identical to n400 potentials and reflect semantic processing during recognition testing,” *Psychophysiology*, vol. 48, no. 4, pp. 532–546, 2011.

[391] P. Stróżak, C. W. Bird, K. Corby, G. Frishkoff, and T. Curran, “Fn400 and lpc memory effects for concrete and abstract words,” *Psychophysiology*, vol. 53, no. 11, pp. 1669–1678, 2016.

[392] R. Ratcliff, P. B. Sederberg, T. A. Smith, and R. Childers, “A single trial analysis of eeg in recognition memory: Tracking the neural correlates of memory strength,” *Neuropsychologia*, vol. 93, pp. 128–141, 2016.

[393] R. Johnson Jr, “The amplitude of the p300 component of the event-related potential: Review and synthesis,” *Advances in psychophysiology*, vol. 3, pp. 69–137, 1988.

[394] D. Ruchkin and S. Sutton, “Emitted p300 potentials and temporal unvertainty,” *Electroencephalography and Clinical Neurophysiology*, vol. 45, no. 2, pp. 268–277, 1978.

[395] R. Johnson Jr and E. Donchin, "On how p300 amplitude varies with the utility of the eliciting stimuli," *Electroencephalography and Clinical Neurophysiology*, vol. 44, no. 4, pp. 424–437, 1978.

[396] L. A. Farwell and E. Donchin, "The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials," *Psychophysiology*, vol. 28, no. 5, pp. 531–547, 1991.

[397] K. C. Squires, E. Donchin, R. I. Herning, and G. McCarthy, "On the influence of task relevance and stimulus probability on event-related-potential components," *Electroencephalography and clinical neurophysiology*, vol. 42, no. 1, pp. 1–14, 1977.

[398] G. R. Kuperberg, "Neural mechanisms of language comprehension: Challenges to syntax," *Brain research*, vol. 1146, pp. 23–49, 2007.

[399] X. Xiao, H. D. Lucas, K. A. Paller, J.-h. Ding, and C.-y. Guo, "Retrieval intention modulates the effects of directed forgetting instructions on recollection," *PloS one*, vol. 9, no. 8, p. e104701, 2014.

[400] P. Ingwersen and K. Järvelin, "Information retrieval in context: Irix," in *Acm sigir forum*, vol. 39, no. 2. ACM New York, NY, USA, 2005, pp. 31–39.

[401] ——, *The turn: Integration of information seeking and retrieval in context.* Springer Science & Business Media, 2006, vol. 18.

[402] D. Zellhöfer, "An extensible personal photograph collection for graded relevance assessments and user simulation," in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval.* ACM, 2012, p. 29.

[403] D. Blackwood and W. J. Muir, "Cognitive brain potentials and their application," *The British Journal of Psychiatry*, vol. 157, no. S9, pp. 96–101, 1990.

[404] K. Ong, K. Järvelin, M. Sanderson, and F. Scholer, "Using information scent to understand mobile and desktop web search behavior," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, pp. 295–304.

[405] T. Sakai, "On the reliability of information retrieval metrics based on graded relevance," *Information processing & management*, vol. 43, no. 2, pp. 531–548, 2007.

[406] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.

Appendix A

Ethic Forms

A.1 Information Sheet

Participant Information Sheet

Name of department: Information & Computer Sciences
Title of the study: An EEG examination of cortical activity during relevance processing

Ethics Approval No.: 948

Introduction
My name is Zuzana Pinkosova and I am a PhD candidate at the Department of Computer and Information Sciences at the University of Strathclyde. We are currently conducting research into the Information Retrieval process by recording brain activity with electroencephalography (EEG) while people make relevance judgements. The experiment is conducted at the School of Psychological Sciences and Health laboratories, Graham Hills Building at the University of Strathclyde. You are being invited to take part in this study. Before you decide to participate, it is important for you to understand why the research is being conducted and what will be involved during the procedure. Please take time to read the following information carefully and ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the purpose of this research?
The aim of the present study is to investigate the relevance process during question-answering task using EEG. Through examining the difference between relevant and non-relevant relevance judgements, we would like to understand the neurological processes that underline relevance judgement.

What is EEG?
EEG is a completely non-hazardous and non-invasive brain imaging technique that allows us to measure electrical currents produced by the neurons in the brain, while the brain is at rest or performs different tasks. EEG offers us an insight into the dynamics of neuronal activity and the associated cognitive processes, with very accurate temporal resolution.

Do you have to take part?
You do not have to take part in this study and participation is voluntary. Should you consent to participate in this study you still have the right to withdraw at any time without providing any explanation.

What will you do in the project?
The experiment will take place in the Graham Hills building (Lab 666) at the University of Strathclyde. During the study, you will be seated comfortably in the chair. At the beginning of the experiment, we will attach electrodes for measuring electrical activity on your scalp surface (which is painless) using an EEG cap. We will use a saline solution, to obtain a good signal transmission, as the amplitudes of the signals produced by the brain are very small. The EEG electrodes are then connected to a computer that records your brain activity, while you are performing experimental tasks. In the experiment, we will continuously record EEG. This will give us valuable information on the relevance processes in the brain. There will be sufficient breaks to avoid fatigue. During the experiment, you will be asked to complete four questionnaires. During the first questionnaire, you will be asked

The place of useful learning
The University of Strathclyde is a charitable body, registered in Scotland, number SC015263

to fill-in demographic information and you will be also screened to assess whether you are eligible to take part in the study. You will be asked to provide details about any existing medical condition you might be diagnosed with that might impact the EEG signal. Then, you will be asked to fill-in a Pre-Task Questionnaire, Post-Task Questionnaire, and Exit Questionnaire, which will help us to understand your interests, area of knowledge and also your experience during the experiment. The study will consist of 64 trials. The 64 trials will be split into two identical sessions (32 trials each). After completing the 1st session (32 trials), you will be instructed to take a break. You will have an opportunity to go through the practice trials first until you are confident with the task. Every trial consists of 2 parts. During the first part of the trial, you will be first presented with a question. You will be required to indicate by a button press once you have fully understood the question and when you are ready to proceed with the task. Then a fixation cross will appear for about a second on the screen to indicate where the answer is going to appear. After that, an answer is going to appear on the screen. However, the answer is going to be presented as a set of separate words, presented one after another. You will be instructed to indicate by a button press, as soon as you know whether the presented answer is relevant or not relevant. Once you made your relevant judgement, you can decide whether you would like to see the rest of the sentence or to terminate it. Next, we would like to ask you to indicate by pressing one of two buttons, whether you have already known the answer to the question.

After this, you will proceed to the second part. You will be instructed to again indicate how relevant was the answer to the question. Again, you will be presented with the same question (from the first part of the trial) and a sequence of words will appear on the screen one after another. However, this time, you will be required to reflect back on part one and to indicate by a button press how relevant were the pieces of information as the answer was unfolding to you using the following scale: ''highly relevant'' , ''low relevant'' and ''non-relevant".

You will receive instructions again prior to the experiment and you will have an opportunity to ask questions. The entire duration of the experiment is estimated to be around 2 hours. You will be asked to sit still during the measurements since movements will interfere with getting accurate data. You will be notified in advance when it is required for you to remain still. In addition, we will ask you to keep your gaze still. At the end of the experiment, the conductive solution will leave your hair damp/ slightly wet. However, you will be provided with towels and hairdryer if required. Eligible participants will be compensated with 2 credits required for a Psychology module.

Why have you been invited to take part?
You have been chosen to take part in the experiment because you are a fluent English speaker, over 18 years old and you are computer literate. You are also neurologically healthy.

What are the potential risks to you in taking part?
There are no risks associated with an EEG. The recording is painless and safe. However, to avoid you feeling uncomfortable or tired through the need to sit still throughout the recordings, you will be provided with frequent breaks as required. You may withdraw participation at any time without giving any reason.

The place of useful learning
The University of Strathclyde is a charitable body, registered in Scotland, number SC015263

What information is being collected in the project?
Demographic information, participant's views, and experience will be collected through the questionnaire. Researchers will also collect behavioural and physiological data using EPrime 2 and EEG. Hence, the obtained data will be classified as sensitive information.

What happens to the information collected from the project?
All information and data collected during the experiment will be anonymised to the best possibilities. Your personal details will be stored securely in digital format and will be encrypted. The search logs and survey data we collect will be retained by the below-mentioned researchers (Dr Yashar Moshfeghi, Dr William McGeown, Zuzana Pinkosova) and may be used in future project publications, following similar ethically approved research protocol. Your participation will remain confidential, and your name or any other directly identifiable information will NOT appear in any published documents relating to the research conducted.

All information and data collected during the experiment will be anonymised to the best possibilities. Your personal details will be stored securely in digital format and will be encrypted. The search logs and survey data we collect will be retained by the below-mentioned researchers (Dr Yashar Moshfeghi, Dr William McGeown, Zuzana Pinkosova) and may be used in future project publications, following similar ethically approved research protocol. Your participation will remain confidential, and your name or any other directly identifiable information will NOT appear in any published documents relating to the research conducted.

Who will have access to the information?
Only below-mentioned researchers (Dr Yashar Moshfeghi, Dr William McGeown, Zuzana Pinkosova) will have access to the data. It is possible that the data may be used by the below-mentioned researchers for other similar ethically approved research protocols, where the same standards of confidentiality will apply. Due to the sensitive nature of the data, the data will not be shared (unless approved by the Principal Investigator, Dr Yashar Moshfeghi).

Where will the information be stored and how long will it be kept for?
Only below-mentioned researchers (Dr Yashar Moshfeghi, Dr William McGeown, Zuzana Pinkosova) will have access to the data. It is possible that the data may be used by the below-mentioned researchers for other similar ethically approved research protocols, where the same standards of confidentiality will apply. Due to the sensitive nature of the data, the data will not be shared (unless approved by the Principal Investigator, Dr Yashar Moshfeghi).

What happens next?
If you are happy to be involved in this project, please read and then complete the following consent form and then we can proceed to begin the experiment. Otherwise, we thank you for your time and attention.

Thank you for reading this information. Please feel free to contact the researcher if you are unsure about this experiment**Researcher contact details:**

The place of useful learning
The University of Strathclyde is a charitable body, registered in Scotland, number SC015263

Zuzana Pinkosova	Dr Yashar Moshfeghi	Dr William McGeown
PhD Candidate Department of Computer & Information Sciences University of Strathclyde	Strathclyde Chancellor's Fellow Department of Computer & Information Sciences University of Strathclyde	Senior Lecturer School of Psychological Sciences and Health University of Strathclyde
Livingstone Tower 16 Richmond Street Glasgow, G1 1XQ	Livingstone Tower 16 Richmond Street Glasgow, G1 1XQ	Graham Hills Building 40 George Street Glasgow G1 1QE
zuzana.pinkosova@strath.ac.uk	yashar.moshfeghi@strath.ac.uk	william.mcgeown@strath.ac.uk

Chief Investigator details:

This research was granted ethical approval by the University of Strathclyde Ethics Committee under application number *948*. If you have any questions or concerns, before, during or after the investigation, or wish to contact an independent person to whom any questions may be directed or further information may be sought from, contact details are provided below:

Departmental Ethics Committee
Department of Computer & Information Sciences
University of Strathclyde
Livingstone Tower

26 Richmond Street
Glasgow
G1 1XH
Scotland, United Kingdom

Telephone: 0141 548 3707
Email: ethics@strath.ac.uk

The place of useful learning
The University of Strathclyde is a charitable body, registered in Scotland, number SC015263

A.2 Consent Form

Consent Form

Title of the project: **An EEG examination of cortical activity during relevance processing**

Ethics approval no.: 948

Researcher's name: Zuzana Pinkosova
Researcher's email: zuzana.pinkosova@strath.ac.uk

Please read the following statements and indicate whether you agree by providing your initials in a box for each statement:

I confirm that I have read and understood the information sheet for the experiment. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

I understand that my participation is entirely voluntary and that I am free to withdraw from the experiment at any time, up to the point of completion, without having to give a reason and without any consequences.

I understand that I can withdraw from the study any personal data (i.e. data which may identify me personally) at any time.

I understand that anonymised data (i.e. data which does not identify me personally) cannot be withdrawn once they have been included in the study.

I understand that any information recorded in the study will be treated confidentially and no information that directly identifies me will be made publicly available.

I confirm that I have read and understood the Privacy Notice for Participants in Research Projects and understand how my personal information will be used and what will happen to it.

I understand that anonymised data will be stored in a secured location for as long as it will be required by involved researchers.

I consent to be a participant in this study.

(PRINT NAME)	
Signature of Participant:	Date:

If you would like a copy of the consent form to keep, please ask the researcher. If you have any complaints or concerns about this research, you can direct these to Departmental Ethics Committee, in writing by email at: ethics@cis.strath.ac.uk

A.3 Debriefing Form

University of Strathclyde Science

Debriefing From

Title of the project: **An EEG examination of cortical activity during relevance processing**

Ethics approval no.: 948

Researcher's name: Zuzana Pinkosova
Researcher's email: zuzana.pinkosova@strath.ac.uk

Thank you for taking part in this research. The aim of this research is to investigate the relevance process using electroencephalography (EEG) during a specially designed question-answering task.

If you would like more information about this study once it is completed, please contact the researcher: Zuzana Pinkosova by email: zuzana.pinkosova@strath.ac.uk, or my supervisor: Dr Yashar Moshfeghi by email: yashar.moshfeghi@strath.ac.uk.

If you are interested in this area of research, you may wish to read the following references:

- Allegretti, M., Moshfeghi, Y., Hadjigeorgieva, M., Pollick, F. E., Jose, J. M., & Pasi, G. (2015). When Relevance Judgement is Happening? *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR 15* doi:10.1145/2766462.2767811

- Eugster, M. J., Ruotsalo, T., Spapé, M. M., Kosunen, I., Barral, O., Ravaja, N., Kaski, S. (2014). Predicting term relevance from brain signals.*Proceedings of the 37th International* ACM *SIGIR Conference on Research & Development in Information Retrieval SIGIR 14* doi:10.1145/2600428.2609594

- Eugster, M., Ruotsalo, T., Spapé, M., Barral, O., Ravaja, N., Jacucci, G. and Kaski, S. (2016). Natural brain information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific Reports*, 6(1).

If you would like a copy of the consent form to keep, please ask the researcher. If you have any complaints or concerns about this research, you can direct these to Departmental Ethics Committee, in writing by email at: ethics@cis.strath.ac.uk

Appendix B

Participant Recruitment

I am a PhD student working on an experiment, which aim is to examine brain activity (using EEG) during information retrieval task and I would like to **invite you to take part!**

I am looking for neurologically and psychologically healthy, fluent English speakers to take part. The study will take around 1 hour and 30 mins to complete.

You will be able to see your own brain activity and you will also have an opportunity to learn about EEG. Eligible participants will also receive 2 credits.

Thank you for your help!

Appendix C

Questionnaires

C.1 Entry Questionnaire

University of Strathclyde Science

Entry Questionnaire

Participant's Code: __ __ Date: __ __/__ __/__ __ (D/M/Y)

Investigator name: Zuzana Pinkosova Investigator's contact detail: zuzana.pinkosova@strath.ac.uk

Name: ______________________________ Age: __ __

Gender: Male ☐ Female ☐ Other ☐

Nationality: ______________________ Years of formal education: __ __

What is the highest degree or level of school you have completed? *If currently enrolled, highest degree received.*

____

Handedness: Left ☐ Right ☐

Occupation: ____

English level skills: Native ☐ Fluent ☐ Advanced ☐ Intermediate ☐ Basic ☐

Questionnaire EEG

Have you:

1. Had EEG before?	Yes ☐	No ☐
2. Had an unexplained spell of loss of consciousness?	Yes ☐	No ☐
3. Had any brain-related, neurological injury, psychiatric condition or illness?	Yes ☐	No ☐
4. Recently taken any psycho-active drug or alcohol?	Yes ☐	No ☐
5. Are you feeling rested?	Yes ☐	No ☐
6. Are you taking any medications?	Yes ☐	No ☐
7. Do you suffer from frequent or severe headaches?	Yes ☐	No ☐

As there are some factors that can affect the EEG signal recordings, we would like to ask you to provide detailed information for any « YES » response (apart from Q1 & Q5). We understand that this is a confidential information and we respect your right not to share it with the researchers. However, without detailed information, we would not be able to include you in the study and you have a right to withdraw.

____
____
____

SIGNATURES

Participant: ______________________ Date: __ __/__ __/__ __(D/M/Y)

Investigator: ______________________ Date: __ __/__ __/__ __ (D/M/Y)

C.2 Pre-Task Questionnaire

In this experiment, you will be presented with a series of questions and answers. Your task will be to judge, whether the answer is relevant to the question. You will receive detailed information with regards to the task. From now on, this will be referred to as Question-Answering task.

Have you ever completed a Question-Answering (Q-A) task in English?

- Yes
- No

On average, how often do you...

	Several times a day	Several times a week	Several times a month	Less than several times a month	I do not use search engines
perform a search in internet search engines (i.e. Google, Bing, Yahoo...)?	○	○	○	○	○
put a question into an internet search engine to gain the information you need?	○	○	○	○	○

In general, how knowledgeable do you feel that you are in the area of...

	Extremely knowledgeable	Very knowledgeable	Moderately knowledgeable	Slightly knowledgeable	Not knowledgeable at all
history	○	○	○	○	○
sport	○	○	○	○	○
science	○	○	○	○	○
geography	○	○	○	○	○
medicine	○	○	○	○	○
politics	○	○	○	○	○

Signature

Date

C.3 Post-Task Questionnaire

In this questionnaire, we would like to ask you about your perception of the encountered task. We would like to ask you to indicate your level of agreement with the following statements:

On average, I found the encountered task:

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
Interesting	○	○	○	○	○	○	○
Difficult	○	○	○	○	○	○	○
Enjoyable	○	○	○	○	○	○	○
Stressful	○	○	○	○	○	○	○
Easy to read	○	○	○	○	○	○	○
Easy to understand	○	○	○	○	○	○	○

On average, I found the encountered questions:

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
Interesting	○	○	○	○	○	○	○
Difficult	○	○	○	○	○	○	○
Easy to read	○	○	○	○	○	○	○
Easy to understand	○	○	○	○	○	○	○
Relevant	○	○	○	○	○	○	○

On average, I found the encountered topic categories (i.e. sport, history,science):

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
Interesting	○	○	○	○	○	○	○
Difficult	○	○	○	○	○	○	○

On average, I was:

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
familiar with the answers to the questions	○	○	○	○	○	○	○
comfortable during the task	○	○	○	○	○	○	○

Signature

Date

C.4 Exit Questionnaire

In this questionnaire, consisting of two parts (Part A, Part B), we would like to ask you about your perception of encountered study.These questions will help researchers identify any possible factors that could potentially impact your performance

Part A

In this section of the questionnaire, we would like to ask you to indicate your level of agreement with the following statements:

On average, I feel that:

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
I had enough time to press the button for relevance judgement	○	○	○	○	○
The speed of words presentation was appropriate for reading	○	○	○	○	○
I felt under pressure	○	○	○	○	○
The presented information was easy to read	○	○	○	○	○
The font size was appropriate for reading	○	○	○	○	○
The monitor luminance was appropriate	○	○	○	○	○
Following the procedure was easy	○	○	○	○	○
I was comfortable during the procedure	○	○	○	○	○
EEG cap was not causing me discomfort	○	○	○	○	○
I am satisfied with my performance	○	○	○	○	○
Participant's instructions were clear	○	○	○	○	○
I felt tired during the study	○	○	○	○	○

I have answered all the questions:

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
honestly	○	○	○	○	○
correctly	○	○	○	○	○
to my best abilities	○	○	○	○	○

Part B

The study you have performed consisted of 2 sessions (session 1–before the break and session 2–after the break). We would like to ask you about your effort and performance throughout both parts of the study

My effort was constant throughout both parts of the study.*

Yes

No

*If you answered no, please identify the phase which you think you have put more effort in:

The first part (before the break)

The second part (after the break)

My performance was constant throughout both parts of the study?**

Yes

No

**If you answered no, please identify the phase in which you think you have performed better

The first part (before the break)

The second part (after the break)

Do you have any comments about the task?

Signature

Date

Appendix D

Data Sets

D.1 Data Set - Practice Task

Q no.	TREC	Question	Answer
1	8	What two US biochemists won the Nobel Prize in medicine in 1992?	Two US biochemists Edwin Kerbs and Edmond Fischer, jointly won the 992 Nobel Medicine Prize for work that could advance the search for an anti-cancer drug.
2	8	Who was the first American in space?	who have successfully developed and operated the heave Space Shuttle
3	2001	What is the effect of acid rain?	destroys aquatic life in lakes and streams

Q no.	DOC	TOPIC	Flesch Score	Answer Readability	TREC Relevance	Answer Presentation	Answer Length	Question Difficulty	Q - No. of characters	Q - No. of words	A - No. of words
1	16	FT924-14045	57	0	R	Y	L	D	53	12	26
2	21	FBIS3-23678	36	0	N	Y	S	D	30	7	10
3	1103	AP890922-0109	79	1	R	N	S	E	26	7	7

D.2 Data Set A

Q no.	TREC	Question	Answer
1	8	Who was the first doctor to successfully transplant a liver?	Starzl, who performed the world successful liver transplant at the University of Colorado in 1967, was admitted to the hospital
2	8	When did Nixon die?	April 27, Wednesday, a national day of mourning for former U.S. President Richard Nixon who died on April 22
3	8	When was London's Docklands Light Railway constructed?	Built in the 1980s for Pounds 77m, the Docklands Light Railway is being extended and upgraded at a cost of Pounds 800m to increase capacity
4	8	In what year did Joe DiMaggio compile his 56-game hitting streak?	Joe DiMaggio's record 56-game hitting streak in 1941, and New York officials even put a Yankee cap on Antley's head in the winner's circle
5	8	How much could you rent a Volkswagen bug for in 1966?	news to Arlo Wilson. When Dollar Rent a Car opened in 1966, you could rent a Volkswagen bug for $1 a day.
6	8	Name the first private citizen to fly in space.	within seconds of take-off, snuffing out the lives of its seven astronauts. Among them was Christa McAuliffe, the first private citizen to fly in space.
7	8	Who won two gold medals in skiing in the Olympic Games in Calgary?	that Tomba spent too much time on the party circuit last summer after winning two Olympic gold medals at Calgary
8	8	Who first circumnavigated the globe?	first landed in Brazil in 1500. Sailing under Spanish colours, Magellan began the first circumnavigation of the globe in 1519.
9	2001	What is the length of the coastline of the state of Alaska?	affected only two percent of Alaska's 6,640 miles of coastline.
10	2001	What does Phi Beta Kappa mean?	was a member of Phi Beta Kappa, the academic honour society.
11	2001	Where is the volcano Mauna Loa?	has branches feeding Hawaii's active Kilawea and Mauna Loa volcanoes.
12	2001	What is phenylalanine?	produce phenylalanine, an enzyme used in the artificial sweetener aspartame.
13	2001	What is the oldest city in the United States?	Boston, the oldest city in the United States
14	2001	What is Hawaii's state flower?	Yellow hibiscus is the state flower of Hawaii.
15	2001	How far is Pluto from the sun?	Pluto averages more than 3.7 billion miles from the sun
16	2001	Where did Howard Hughes die?	Howard Hughes died in Houston at age 72.
17	2001	George Bush purchased a small interest in which baseball team?	Bush is a former captain of the Yale baseball team
18	2001	What is amitriptyline?	Prescription drugs must undergo tests that lead to human trials
19	2001	Who invented the slinky?	The company produces other toys, including building blocks
20	2001	When was Algeria colonized?	New Caledonia was colonized by France in 1853.
21	2001	What is the average speed of the horses at the Kentucky Derby?	was beaten by more than 22 lengths in the Kentucky
22	2001	What did Edward Binney and Howard Smith invent in 1903?	rainy days and car trips; a smell of paraffin
23	2001	Where is the Orinoco River?	made it to the West, newspapers reported Saturday.
24	2001	What river flows between Fargo, North	where the Mississippi River was four feet below flood

		Dakota and Moorhead, Minnesota?	
25	2001	What is the average body temperature?	The problem is that when the air temperature gets close to the body's 98.6 degree Fahrenheit normal, heat doesn't dissipate outward as readily.
26	8	Who wrote "Hamlet"?	Coleridge thought he was Hamlet, and managed three plays himself. But none wrote a play in this obscure Shakespearean style.
27	8	What is the name of the highest mountain in Africa?	Yet Kilimanjaro, Africa's highest mountain, remained veiled in ominous mist as we drove towards it out of Kenya.
28	8	Who leads the star ship Enterprise in Star Trek?	The star ship Enterprise in Star Trek, the television series creation of Gene Roddenberry, is led by Captain Kirk
29	8	What is the largest city in Germany?	a sign that such position will be moved forward. Qiao said Berlin, the largest city in Germany, has great influence on the world.
30	2001	What hemisphere is the Philippines in?	said China ordinarily buys its sugar from Thailand, Australia, the Philippines and other exporters in the Eastern Hemisphere.
31	2001	What do bats eat?	bug-eating bats, which have a wingspan of about six inches, were some of the thousands displaced from inside the ceiling
32	2001	What is neuropathy?	cause the most serious bike injuries, but regular biking also can cause neuropathy, a malfunction of the nerves
33	2001	What is sodium chloride?	Table salt is forty percent sodium and sixty percent chloride.
34	2001	Who was Galileo?	In 1642, astronomer Galileo died in Arcetri, Italy.
35	2001	Where is the Louvre?	Mona Lisa, which hangs in the Louvre Museum in Paris
36	2001	What is epilepsy?	Epilepsy is a brain disorder characterized by seizures
37	2001	Who invented the telephone?	telephone by turning down Alexander Graham Bells invention of it
38	2001	What is caffeine?	caffeine, a stimulant that comes from coffee beans, tea leaves
39	2001	What French ruler was defeated at the battle of Waterloo?	defeated Napoleon at the 1815 Battle of Waterloo.
40	2001	What primary colours do you mix to make orange?	orange, from yellow and red; and purple, from blue
41	8	In 1990, what day of the week did Christmas fall on?	the former state of division, which existed before 22 May 1990, after which date Yemen was reunified and the Republic of Yemen proclaimed.
42	8	What cancer is commonly associated with AIDS?	the dosage is considerably smaller, and more targeted, than the radiation commonly administered for treatment of brain tumours or cancers
43	8	Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"?	Britain's first woman PM. Now the Iron Lady is gone from the front benches of the House of Commons
44	8	What is the name of the medical condition in which a baby is born without a brain?	an underdeveloped heart that was replaced by Dr Leonard Bailey with the walnut-sized heart of Baby Gabriel, born with virtually no brain.
45	8	What is the acronym for the rating system for air conditioner efficiency?	The data from fan manufacturers did not provide information on the effect of different fan types on air system efficiency
46	8	Who was President of Costa Rica in 1994?	The following is a compilation of reports on Costa Rican economic developments

			monitored through 29 March Central Bank President Jorge Corrales
47	8	Who won the first general election for President held in Malawi in May 1994?	We have also had cases of the Young Pioneers who have been going disguising themselves as MCP opposition members
48	2001	Who was the first woman killed in the Vietnam War?	A memorial statue of the nurse stands in the courtyard of Aultman Hospital in Canton, where she went to nursing school.
49	2001	What is ozone depletion?	depletion, the nations of the world have agreed
50	2001	What does cc in engines mean?	below 1800 cc to 2000 cc middle displacement.
51	2001	What is the pH scale?	learn about the 14-point pH scale and the difference
52	2001	What is Wimbledon?	has been pressure on Wimbledon to change surface
53	2001	What is a shaman?	Navajo shaman is a pivotal figure in Coyote Waits
54	2001	What type of currency is used in Australia?	He said Australia would follow the United Nations
55	2001	During which season do most thunderstorms occur?	The robbery occurred Sunday night during a thunderstorm
56	2001	What continent is Egypt on?	Bush will formally ask Congress today for the money
57	8	Who played the part of the Godfather in the movie, "The Godfather"?	Pyonsas became so popular, they were treated as celebrities and sometimes influenced the contents of the movies until the profession died
58	2001	What is Teflon?	It's a discovery that almost slipped away. Plunkett, two years out of college, was working with refrigeration gases
59	2001	Material called linen is made from what plant?	The blast apparently occurred in an area where combustible materials are used, especially designed with reinforced walls to minimize damage from an explosion
60	2001	What is the longest bone in the human body?	A ghoulish case in which prosecutors presented body fragments as evidence that an airline pilot murdered his wife
61	2001	What is neurology?	there has been much irreversible damage, Antonio Damasio, head of the neurology department at the University of Iowa, said at a news conference.
62	2001	What is amoxicillin?	expects to analyse more than a thousand samples from hundreds of companies to check potency, rate of dissolution in the body and chemical composition.
63	2001	What are polymers?	an agreement with Advanced Polymer Systems Inc., of Redwood City, California, to develop a chewing gum that encapsulates the flavour oils
64	2001	What are invertebrates?	invertebrate marine animals, and a world authority on hydrozoans, which are small creatures related to sea anemones and corals

Q no.	DOC	TOPIC	Flesch Score	Answer Readability	TREC Relevance	Answer Presentation	Answer Length	Question Difficulty	Q - No. of characters	Q - No. of words	A - no. of words
1	LA071090-0162	53	34	0	R	Y	L	D	50	10	20
2	FBIS4-3661	54	64	1	R	Y	L	D	15	4	19
3	FT931-3210	15	60	1	R	Y	L	D	47	7	25
4	LA042589-0086	71	60	1	R	Y	L	D	54	11	24
5	LA060790-0003	13	88	1	R	N	L	D	42	11	22
6	LA022590-0228	66	67	1	R	N	L	D	38	9	25
7	LA021089-0110	86	55	0	R	N	L	D	53	13	20
8	LA122089-0137	93	53	0	R	N	L	D	31	5	20
9	AP890606-0203	1006	53	0	R	N	S	D	47	12	10
10	AP901010-0153	1042	34	0	R	N	S	D	24	6	10
11	AP901206-0033	1057	27	0	R	N	S	D	25	6	10
12	WSJ910208-0112	1126	-6	0	R	N	S	D	19	3	10
13	FBIS4-11044	954	61	1	R	Y	S	D	36	9	8
14	LA081989-0091	1008	51	0	R	Y	S	D	25	5	8
15	AP880610-0117	1073	78	1	R	Y	S	D	23	7	10
16	AP890327-0178	1082	93	1	R	Y	S	D	23	5	8
17	AP890504-0181	900	87	1	N	Y	S	D	52	10	10
18	AP890817-0059	936	61	1	N	Y	S	D	19	3	10
19	LA122890-0120	1096	30	0	N	Y	S	D	20	4	8
20	AP881009-0092	1047	51	0	N	Y	S	D	23	4	8
21	LA060990-0078	924	87	1	N	N	S	D	50	12	10
22	LA061890-0098	1146	85	1	N	N	S	D	45	10	9
23	AP881029-0112	930	51	0	N	N	S	D	22	5	8
24	AP890408-0106	956	66	1	N	N	S	D	59	10	9
25	AP880625-0220	1101	47	0	R	Y	L	E	31	6	23
26	FT933-6893	196	78	1	R	Y	L	E	15	3	20
27	FT942-13604	35	43	0	R	Y	L	E	41	10	18
28	FT942-456	74	50	0	R	Y	L	E	39	9	19
29	FBIS3-45318	33	70	1	R	N	L	E	29	7	23
30	AP880702-0185	923	24	0	R	N	L	E	32	6	18
31	AP881027-0159	957	69	1	R	N	L	E	13	4	20
32	AP900806-0171	1075	43	0	R	N	L	E	16	3	19

33	AP900208-0237	1110	44	0	R	Y	S	E	20	4	10
34	AP890103-0182	896	8	0	R	Y	S	E	13	3	8
35	AP880715-0037	1070	61	1	R	Y	S	E	16	4	10
36	LA032690-0045	912	19	0	R	Y	S	E	14	3	8
37	WSJ891102-0101	1123	36	0	R	N	S	E	23	4	10
38	FT922-4819	920	78	1	R	N	S	E	14	3	10
39	AP880723-0041	1080	30	0	R	N	S	E	47	10	8
40	SJMN91-06256195	1072	85	1	R	N	S	E	38	9	9
41	FBIS4-59089	36	44	0	N	N	L	D	41	11	23
42	LA010490-0128	75	22	0	N	N	L	D	38	7	20
43	LA112890-0021	1	81	1	N	N	L	D	65	15	19
44	LA061189-0085	113	47	0	N	N	L	D	65	17	23
45	FR940304-1-00096	57	34	0	N	Y	L	D	61	12	20
46	FBIS3-12077	59	12	0	N	Y	L	D	32	8	21
47	FBIS4-48354	121	50	0	N	Y	L	D	62	14	20
48	AP880623-0242	1200	61	1	N	Y	L	D	40	10	21
49	AP890109-0223	1114	72	1	N	N	S	E	20	4	8
50	FBIS4-29960	949	51	0	N	N	S	E	23	6	8
51	SJMN91-06098188	969	78	1	N	N	S	E	16	5	9
52	FT922-1731	1022	72	1	N	N	S	E	15	3	8
53	AP900221-0238	937	38	0	N	Y	S	E	13	4	9
54	FBIS3-29068	970	51	0	N	Y	S	E	35	8	8
55	AP880727-0077	1021	40	0	N	Y	S	E	41	7	8
56	AP900914-0044	1049	66	1	N	Y	S	E	22	5	9
57	LA091290-0040	77	36	0	N	Y	L	E	54	12	21
58	AP880331-0277	935	52	0	N	Y	L	E	14	3	18
59	AP880725-0189	934	18	0	N	Y	L	E	38	8	23
60	AP880716-0049	1039	43	0	N	Y	L	E	34	9	18
61	AP880407-0193	994	14	0	N	N	L	E	15	3	23
62	AP890817-0058	980	27	0	N	N	L	E	17	3	24
63	AP890110-0155	1055	24	0	N	N	L	E	15	3	21
64	AP890106-0184	926	23	0	N	N	L	E	20	3	19

D.3 Data Set B

Q no.	TREC	Question	Answer
1	8	Who won the Nobel Peace Prize in 1991?	Suu Kyi, the 1991 Nobel Peace Prize laureate and the most outspoken critic of the military's heavy-handed policies
2	8	When was China's first nuclear test?	In China's first nuclear test on a tower in October 1964, the explosion time was chosen to be in the afternoon
3	8	Who was the second man to walk on the moon?	Aldrin's permanent status as the second man to walk on the moon has led to endless interview questions and a suggestion in his book
4	2001	When was Ulysses S. Grant born?	In 1822, the 18th President of the United States, Ulysses S. Grant, was born in Point Pleasant, Ohio.
5	8	How did Socrates die?	snake root, which is toxic, and poison hemlock, which for over two thousand years has been famous for curing Socrates of life.
6	2001	What is peyote?	two members of the Native American Church who used peyote, a hallucinogenic drug derived from a cactus, as part of their worship.
7	2001	How old was the youngest president of the United States?	Kennedy, at 43 the youngest president elected, served nearly three years before he was shot and killed in a Dallas motorcade
8	2001	Who founded American Red Cross?	as Red Cross founder Barton did, but she could end up having easily as many battles on her hand
9	2001	When was the first stamp issued?	Since the stamps were first issued in the mid-1800s
10	2001	Where is the Euphrates River?	security concerns over Turkish control of the Euphrates
11	2001	When was the first Wal-Mart store opened?	not until 1962 that he opened his first Wal-Mart
12	2001	What year did Oklahoma become a state?	for Oklahoma to become the 46th state in 1907
13	2001	What type of polymer is used for bulletproof vests?	Fiberglass, graphite or Kevlar, a material used in bulletproof vests.
14	2001	How did Janis Joplin die?	Janis Joplin, who later died of a drug overdose
15	2001	Where is the Shawnee National Forest?	Shawnee National Forest in southern Illinois has given environmentalists hope
16	2001	What is the speed hummingbirds fly?	Hummingbirds' wings speed up to 200 beats per second
17	2001	What year did the Milwaukee Braves become the Atlanta Braves?	The Atlanta Braves made their second major move
18	2001	Who invented the hula hoop?	Some inept hoopers twirled the hoops on their wrists
19	2001	Who was the 23rd president of the United States?	In 1767, the seventh president of the United States, Andrew Jackson

20	2001	How tall is the Sears Building?	Chicago skyscraper might overshadow the world's tallest building.
21	2001	When was Rosa Parks born?	catalysts for the civil rights movement in 1955
22	2001	Who was the first female United States Representative?	signed a treaty in Washington setting aside Antarctica
23	2001	What country did Ponce de Leon come from?	abandoned the country and took asylum in Panama
24	2001	What was the most popular toy in 1957?	one of the most popular TV shows of all time.
25	2001	Which country gave New York the Statue of Liberty?	The Statue of Liberty, a gift from the people of France, was dedicated in New York Harbour by President Cleveland
26	2001	When is the official first day of summer?	Spring, by the calendar method, lasts until the summer solstice, the northernmost point reached by the sun, which comes June 21 this year.
27	2001	When is St. Patrick's Day?	Two and a half weeks ahead of time, Reagan took note of the commemoration of St. Patrick's Day on March 17
28	2001	What is bipolar disorder?	Lewis determined Bundy suffered from a bipolar mood disorder, or a manic-depressive illness that produced violent mood swings in the convicted killer
29	2001	What are coral reefs?	can take 100 years for coral, a hard shelter built up by millions of tiny animals, to grow one yard.
30	2001	What is schizophrenia?	the severe mental illness of schizophrenia, a study says today. Several earlier studies have described what appear to be differences in the brains of schizophrenics.
31	2001	What is vertigo?	suffered vertigo, skewed balance that caused her to inadvertently to veer to the right and hearing loss.
32	2001	What is an ulcer?	digestive ulcers. They are holes in the lining of the stomach or the duodenum, which is the upper part of the small intestine.
33	2001	Who discovered radium?	Pierre and Marie Curie discovered radium on April 20, 1902
34	2001	What is pectin?	Pectin, a carbohydrate, is commonly used as a thickener
35	2001	What is the name of the satellite that the Soviet Union sent into space in 1957?	The 1957 launching of Soviet space satellite Sputnik
36	2001	What is acupuncture?	acupuncture is a technique to relieve pain and stress
37	2001	What is metabolism?	metabolism, the rate at which it uses up calories
38	2001	Who discovered America?	Santa Maria, used by Columbus when he discovered America
39	2001	What mineral helps prevent osteoporosis?	with calcium citrate can reverse the effect of osteoporosis
40	2001	What is cryptography?	cryptography , the science of writing and decipher codes

41	8	What was the monetary value of the Nobel Peace Prize in 1989?	the Nobel Peace Prize, which has honoured in the past some of the greatest champions of justice and models of virtue
42	2001	What date was Dwight D. Eisenhower born?	after the Supreme Court vacated a stay granted by Justice William O. Douglas and President Eisenhower refused to intervene.
43	2001	What was the name of the plane Lindbergh flew solo across the Atlantic?	trying to become the first aviator to fly around the world in a float plane got tangled in some Soviet red tape
44	8	What costume designer decided that Michael Jackson should only wear one glove?	was asked to leave a jewelry store in Simi Valley last spring when employees became suspicious of a customer wearing a wig
45	2001	What are Quaaludes?	The Barretts allegedly administered Valium, Quaaludes, Percodan and laxatives while they stole her retirement and Social Security checks
46	2001	What is Maryland's state bird?	The eagle population, devastated by the use of the pesticide DDT in the 1950s and 1960s, had been restored
47	2001	What is the atomic weight of silver?	Scientists have pushed down temperatures to a record low of only two billionths of a degree above absolute zero
48	2001	What are amphibians?	Wake says amphibians, which absorb large amounts of water through their skins, are highly susceptible to heavy metals and other toxic material
49	2001	What is a biosphere?	biosphere is able to compensate all the man made
50	2001	What does target heart rate mean?	it will cease attacks on Israeli military targets
51	2001	What is supernova?	no supernova has been close enough and bright enough
52	2001	Who developed the Macintosh computer?	megabyte of internal memory, the same as the Macintosh
53	2001	What is autism?	How do you figure out what day somebody was born
54	2001	What is the average weight of a Yellow Labrador?	About 170 people are already working at the site
55	2001	What is cholesterol?	Cholesterol was high, so he made me come back
56	8	What does the Peugeot company manufacture?	industry veteran, past president of the Society of Motor Manufacturers
57	2001	What were Christopher Columbus' three ships?	Japan will build a replica of Christopher Columbus's ship, the Santa Maria, and sail it from Spain to Japan in 1992
58	2001	What are sunspots?	Only when the jet streams reach a solar latitude of 30 or 35 degrees do they produce sunspots, Ulrich said.
59	2001	What imaginary line is halfway between the North and South Poles?	Dmitri Shparo, Anglican priest Laurie Dexter of Canada, and the other skiers planned two days of rest before continuing the Polar Bridge expedition.

60	2001	What is Valentine's Day?	On Friday, Filipino news organizations received anonymous letters that said a coup called Operation Valentine would be staged around Valentine's Day
61	2001	Why does the moon turn orange?	that was the side of Earth facing the moon, or they have begun before moonrise or ended after the moon set, astronomers say.
62	2001	Where are the British crown jewels kept?	sold off many Allied divisions to pay off its debt in that purchase, and analysts have said they expect Campeau will do the same
63	2001	What is a thermometer?	foreign aid circles, and come up with a lot of ideas, from pedal-power fishing craft more efficient than rowboats to artificial mangrove branches
64	2001	What is the active ingredient in baking soda?	the most widely used active ingredient used in expectorants and is contained in the major brands of cough-cold medicines, the FDA said.

Q no.	DOC	TOPIC	Flesch Score	Answer Readability	TREC Relevance	Answer Presentation	Answer Length	Question Difficulty	Q - No. of characters	Q - No. of Words	A - No. of words
1	FBIS3-48061	40	40.61	0	R	Y	L	D	30	8	18
2	FBIS4-66382	107	60.63	0	R	Y	L	D	30	6	21
3	LA071689-0207	128	59.1	0	R	Y	L	D	33	10	24
4	AP880418-0236	1279	66.1	1	R	Y	L	D	25	6	18
5	LA112990-0023	198	61.45	1	R	N	L	D	17	4	22
6	AP900516-0077	1168	38.38	0	R	N	L	D	12	3	22
7	AP881031-0264	1278	36.46	0	R	N	L	D	46	10	21
8	AP901107-0005	1063	71.78	1	R	N	L	D	26	5	19
9	AP890802-0233	1071	103.62	1	R	N	S	D	26	6	9
10	FT931-14451	1076	18.94	0	R	N	S	D	24	5	8
11	FT944-1217	1087	86.71	1	R	N	S	D	34	7	9
12	AP890421-0160	1171	75.5	1	R	N	S	D	31	7	9
13	AP901210-0070	1368	35.95	0	R	Y	S	D	42	9	10
14	AP890113-0031	1163	66.1	1	R	Y	S	D	20	5	9
15	AP891009-0171	1068	2.11	0	R	Y	S	D	31	6	10
16	AP890810-0193	953	84.9	1	R	Y	S	D	29	6	9
17	LA120590-0160	1064	61.24	1	N	Y	S	D	51	10	8
18	AP880325-0172	1091	84.9	1	N	Y	S	D	22	5	9
19	AP890306-0203	1100	49.54	0	N	Y	S	D	39	9	11
20	AP890708-0110	899	18.94	0	N	Y	S	D	25	6	8
21	AP890403-0117	978	61.24	1	N	N	S	D	20	5	8
22	AP891120-0160	981	18.94	0	N	N	S	D	46	8	8
23	FBIS3-32431	983	29.52	0	N	N	S	D	33	8	8
24	AP890418-0204	1092	95.17	1	N	N	S	D	30	8	10
25	AP891016-0194	1373	55.41	0	R	Y	L	E	41	9	20
26	AP900319-0020	1331	58.43	1	R	Y	L	E	33	8	23
27	AP880229-0179	1342	79.35	1	R	Y	L	E	21	5	21

28	AP890125-0064	917	21.65	0	R	Y	L	E	21	4	22
29	AP891113-0107	992	72.33	1	R	N	L	E	17	4	20
30	AP900322-0023	1352	31.72	0	R	N	L	E	19	3	25
31	AP901206-0051	1329	45.26	1	R	N	L	E	13	3	17
32	AP890112-0140	1328	66.42	1	R	N	L	E	13	4	23
33	AP881212-0224	1045	52.87	0	R	Y	S	E	19	3	10
34	WSJ910924-0085	1384	28.5	0	R	Y	S	E	12	3	9
35	LA071589-0002	1094	61.24	0	R	Y	S	E	64	16	8
36	AP890523-0088	1005	56.7	1	R	Y	S	E	17	3	9
37	AP890405-0114	1303	56.7	0	R	N	S	E	16	3	9
38	AP880714-0128	1017	0.3	0	R	N	S	E	20	3	9
39	AP900208-0204	1011	37.9	0	R	N	S	E	35	5	9
40	FT933-9901	1077	40.09	0	R	N	S	E	18	3	8
41	LA102490-0028	2	68.69	1	N	N	L	D	49	12	21
42	AP880613-0221	1218	32.45	0	N	N	L	D	33	7	19
43	AP900522-0071	1316	76.83	1	N	N	L	D	58	13	22
44	LA123189-0065	43	42.22	0	N	N	L	D	66	12	22
45	AP880524-0257	1181	-13.53	0	N	Y	L	D	16	3	18
46	AP880514-0020	1247	58.42	0	N	Y	L	D	25	5	19
47	AP891005-0076	1264	40.61	0	N	Y	L	D	29	7	19
48	AP900607-0196	944	46.07	0	N	Y	L	D	17	3	22
49	FBIS3-22119	915	84.9	1	N	N	S	E	16	3	9
50	AP881219-0188	1026	40.09	0	N	N	S	E	27	6	8
51	AP880316-0070	1067	66.1	1	N	N	S	E	15	3	9
52	AP880223-0267	919	37.9	0	N	N	S	E	32	5	9
53	AP890308-0005	903	61.33	1	N	Y	S	E	12	3	10
54	FT943-14639	906	75.5	1	N	Y	S	E	39	9	9
55	LA011189-0005	918	84.9	1	N	Y	S	E	17	3	9

56	FT922-12252	3	2.11	0	N	Y	S	E	36	6	10
57	AP880714-0128	1041	56.61	0	N	Y	L	E	38	6	21
58	AP880406-0181	1046	63.87	1	N	Y	L	E	15	3	20
59	AP880426-0223	921	36.36	0	N	Y	L	E	55	11	23
60	AP890212-0043	1135	-7.85	0	N	Y	L	E	20	4	21
61	AP890803-0247	902	58.43	1	N	N	L	E	24	6	23
62	AP880407-0276	1326	55.58	1	N	N	L	E	33	7	24
63	AP880328-0015	1370	48.53	0	N	N	L	E	18	4	23
64	AP890227-0155	1346	40.04	0	N	N	L	E	37	8	22