U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

MESTRADO EM CIÊNCIA DA INFORMAÇÃO

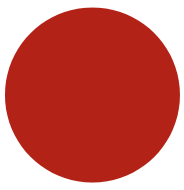
Analysis of web information-seeking behavior of users with different levels of health literacy

Mariana Cláudia Medeiros de Henriques

M
2020

UNIDADES ORGÂNICAS ENVOLVIDAS

FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS



Analysis of web information-seeking behavior of users with different levels of health literacy

Mariana Cláudia Medeiros de Henriques

Dissertation carried out under the master's degree in Information Science, supervised by Professor Carla Teixeira Lopes

Members of the Jury

Chair: Professor Gabriel de Sousa Torcato David

Faculty of Engineering - University of Porto

External Examiner: Professor Ana Margarida Pisco Almeida

University of Aveiro

Supervisor: Professor Carla Teixeira Lopes

Faculty of Engineering - University of Porto

Acknowledgments

I want to express my gratitude to everyone who contributed to the achievement of this dissertation.

First, I want to thank Professor Carla Teixeira Lopes for all the monitoring and advice given from the beginning to the end. Throughout this journey, you were an excellent advisor, taught me, motivated, and inspired me a lot. Always presented me with challenges and taught me to go beyond. You stimulated my interest in this area, especially in the field of research.

To my mother and my brother for always supporting me in my choices, believing in me, and for giving me the best words of comfort in the least positive and most anxious moments.

To nurse Sílvia Costa from Centro Hospitalar São João for being so helpful in my initial topic. Everything has become easier with your help, so I thank you for your valuable support.

To nurse Sofia Marques for her important contribution to this study. I honestly didn't have any health professional in mind who was available in Portugal to help me with the terminology.

To my master's colleagues, Jéssica Barbosa, Mariana Carvalho, and Susana Silva, who were my companions on this journey, for all the support. I hope we keep in touch.

To all the teachers who, during this academic journey, shared all their knowledge and stimulated my interest.

A special thank you to all.

Resumo

Literacia em Saúde é definida como "o nível pelo qual os indivíduos podem obter, processar, compreender e comunicar informação relacionada com saúde necessária para tomar decisões de saúde informadas". Os utilizadores com um baixo nível de literacia em saúde têm menos conhecimentos das suas condições médicas, maior dificuldade em seguir as instruções e compreender a informação dada pelos médicos. Cada vez mais, as pessoas recorrem à web para pesquisar sobre informação de saúde. As dificuldades que os utilizadores de baixa literacia têm no mundo real provavelmente persistem no mundo virtual. O principal objetivo deste estudo é analisar o comportamento de pesquisa de utilizadores com diferentes níveis de literacia em saúde. Pretende-se identificar diferenças entre pessoas com baixa e alta literacia de saúde que depois possam ser utilizadas para a melhoria dos sistemas de recuperação e contribuir, entre outros, para facilitar o acesso à informação e educação das pessoas com baixa literacia. Este estudo surge na sequência de um trabalho prévio que incluiu a anotação dos registos de vídeo de uma experiência com utilizadores realizada anteriormente. Com base na versão preliminar de análise do trabalho anterior, foi proposto um esquema de classificação de eventos que engloba tipos de interação relativos ao navegador, motor de pesquisa e páginas web. Cada tipo de interação é composto por eventos que, por sua vez estão associados a variáveis de análise. Dentro deste esquema, foram construídos módulos para analisar as interrogações de pesquisa submetidas. Com base neste esquema, foi revista a anotação dos vídeos e foi realizada a análise de dados de forma descritiva e inferencial. Os principais resultados demonstram que o grupo de baixa literacia em saúde utilizou mais o botão de voltar atrás; fez mais cliques esquerdos e passou mais tempo a interagir com as páginas de resultados, nomeadamente no scrolling. Por outro lado, o grupo de alta literacia em saúde utilizou mais a barra de endereço para introduzir novas interrogações de pesquisar e selecionar mais vezes o texto do URL. Na página de resultados do motor de pesquisa este grupo fez mais cliques com o botão direito. A nível de reformulação de interrogações, que ocorrem no contexto da mesma necessidade de informação, os utilizadores com baixa literacia em saúde usaram mais as reformulações "totalmente novas", ou seja, sem termos em comum com a interrogação anterior. Por sua vez, o grupo de alta literacia em saúde fez mais reformulações nas quais adicionou palavras, substituiu termos por sinónimos e mudou do plural para o singular.

Palavras-chave: Comportamento de Pesquisa de Informação, Pesquisa na Web, Literacia em Saúde, Interação, Reformulação de Interrogação de Pesquisa, Esquema De Classificação

Abstract

Health Literacy is defined as "the level by which individuals can obtain, process, understand and communicate health-related information necessary to make informed health decisions". Users with a low level of health literacy are less aware of their medical conditions, more difficult to follow instructions and understand the information given by doctors. Increasingly, people turn to the web to search for health information. The difficulties that low literacy users have in the real world are likely to persist in the virtual world. The main objective of this study is to analyze the search behavior of users with different levels of health literacy. It is intended to identify differences between people with low and high health literacy that can then be used to improve recovery systems and contribute, among others, to facilitate access to information and education for people with low literacy. This study follows a previous work that included the annotation of video records of an experience with users previously carried out. Based on the preliminary analysis version of the previous work, an event classification scheme was proposed that includes types of interactions related to the browser, search engine, and web pages. Each type of interaction is composed of events that, in turn, are associated with analysis variables. Within this scheme, modules were built to analyze the search queries submitted. Based on this scheme, the annotation of the videos was reviewed, and the data analysis was performed in a descriptive and inferential manner. The main results demonstrate that the low health literacy group used more the back button; made more left clicks and spent more time interacting with the results pages, namely scrolling. On the other hand, the high health literacy group made more use of the address bar to introduce new search queries and select the URL text more often. On the search engine results page, this group made more right-clicks. At the level of reformulation of questions, which occur in the context of the same need for information, users with low health literacy used more “totally new” reformulations, that is, without terms in common with the previous question. In turn, the high health literacy group did more reformulations in which they added words, substituted terms by their synonym, and changed from the plural to the singular.

Keywords: Information-Seeking Behavior, Web Search, Health Literacy, Interaction, Query Reformulation, Classification Scheme

List of Figures

List of Tables

List of Abbreviations and Acronyms

Ad	Advertisement
AMA	American Medical Association
DAP	Direct Access Page
HISB	Health Information Seeking Behavior
HHL	High Health Literacy
IM	Institute of Medicine
LHL	Low Health Literacy
MeSH	Medical Subject Headings
METER	Medical Term Recognition Test
NLM	National Library of Medicine
NVS	Newest Vital Sign
REALM	Rapid Estimate of Adult Literacy in Medicine
RP	Results Page
SAHLPA	Short Assessment of Health Literacy for Portuguese-speaking Adults
SAHLSA	Short Assessment of Health Literacy for Spanish-speaking Adults
SE	Search Engine
SEBQ	Search Engine Box Query
SEHP	Search Engine Home Page
SERP	Search Engine Results Page
S-TOFHLA	Short Test of Functional Health Literacy in Adults
TOFHLA	Test of Functional Health Literacy in Adults
WHO	World Health Organization

Table of Contents

1. Introduction

The present dissertation entitled "Analysis of web information-seeking behavior of users with different levels of health literacy" has as its first chapter the introduction. In this first chapter, the following topics are covered: context, motivation, objectives, contributions, and structure of the dissertation.

1.1. Context

Health Literacy is defined as “the degree to which individuals can obtain, process, understand and communicate health-related information necessary to make informed health decisions” (Berkman, Davis, and McCormack 2010). There is also reference to the concept of “eHealth Literacy”, which is characterized by the individual ability to use and understand the content of interactive information technologies (Efthymiou et al. 2017).

Online health information search continues to increase, being conducted by a broad diversity of users (Eurostat 2020; Susannah Fox 2014). However, it is necessary to understand whether the information made available by search engines and health portals is accessible to the entire population.

Despite the growing increase in online health searches, some users are frustrated during search (Yilma et al. 2019). One of the factors that have the potential to affect the success of health searches is health literacy. Studies show that users with a low level of health literacy are less successful in web search (Lopes & Ribeiro, 2013) and suggest personalization of search systems based on user health literacy. For that, search engines should detect user-health literacy and retrieve documents appropriate to their level (Lopes & Ribeiro, 2015).

This dissertation aims to analyze and compare the health information-seeking behavior of users with different levels of health literacy. Finding behavioral differences between low and high health literacy users can contribute to improving search engines. Among other refinements, search engines can be personalized to users’ health literacy contributing, for example, to facilitate how users with low health literacy access information. The personalization can also help users with a health literacy degree to find relevant and quality information. This analysis will allow the identification of differences between the two groups which may be an important step towards this personalization.

To accomplish these goals, this work will be built upon previous works. The data collected in a previously conducted user experiment will be used (Sousa 2016), along with the data that resulted from a preliminary analysis of the video recordings of the search sessions.

This preliminary analysis involved the proposal of a classification scheme interaction analysis and its use to annotate each video recording. Based on this material, we proposed a final classification scheme to analyze interactions and proposed modules focused on query reformulation behavior. Part of the video annotation was revised, and the data was analyzed to compare the search behavior in low and high health literacy groups.

1.2. Motivation and Goals

Based on the data collected in a user experiment, we intend to study the search behavior of people with different health literacy levels. As mentioned, the continued growth of online health information search continues to increase. However, all users have very different characteristics and it is necessary to adapt the information search systems to the specifics of each one to reduce the difficulties and inconveniences faced by them. The main factor of analysis is the degree of health literacy of these users that certainly influences their search behavior.

Figure 1 presents the objectives that are intended to be achieved in this project. To accomplish our main goal of "identify differences in the information-seeking behavior of users with low and high health literacy", we have three main objectives: first, to become familiar with the previous study; second, to review annotation and present a proposal for the classification scheme; and third, to analyze the data. To become familiar with the previous study includes analyze previous works by other authors, become familiar with the user experience previously carried out, and becoming familiar with the classification scheme and annotation of videos previously defined. Then, to review the annotation and present proposal it's necessary to review the annotation and proposal of the classification scheme and modules.

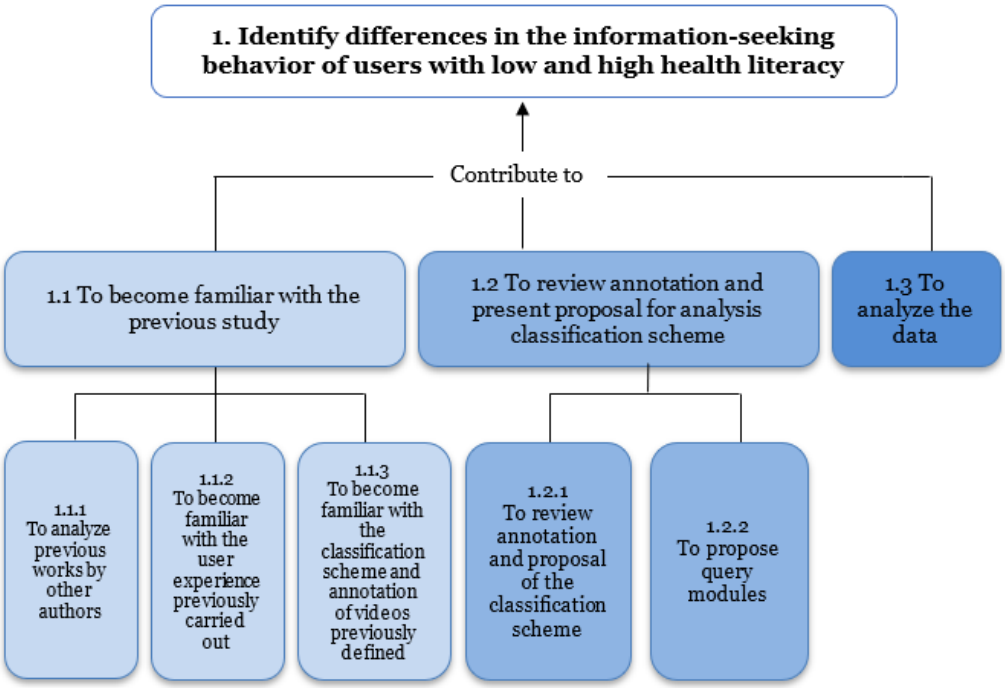


Figure 1 - Goals Tree

The work plan proposed for this project only started in late March 2020 as we changed the dissertation's topic due to the impossibility of continuing with the initial project due to the COVID-19 situation. If you are interested in the initial topic, we presented a summary at ehSemi 2020 - 1st Seminar of students in digital technologies and health/wellness that took place at the University of Aveiro[1].

1.3. Contributions

Studies focused on the analysis of information-seeking behavior show a large diversity of methods and terminology that, in most cases, are specifically designed for that particular study.

In our opinion, our proposed classification scheme has important contributions to the analysis of search behavior. It is generic, which means it can be applied to disregard the used search engines. This classification scheme can be used in manual analysis studies and in (semi) automatic analysis studies in which part of the events can be determined automatically. Furthermore, in the context of health, one of the great contributions is to have a form of analysis by health literacy. About modules, we believe that they can be reusable by other people who are interested in studying search and interaction behavior in various types. It can serve for future researchers to take advantage of some types used.

[1] http://ehealthseminar.web.ua.pt/docs/livro_resumos.pdf

Based on these contributions, we decided to write two scientific articles. The first scientific article aims to present and share the proposed classification scheme, their events, analytic measures, and guidelines for analysis. This first scientific article will be submitted to the Journal of the Association for Information Science and Technology (JASIST). The second scientific article aims to compare the results obtained in the two health literacy groups. The latter will be submitted to the sixth ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR) 2021. Unfortunately, due to the pandemic situation, it was impossible to submit them earlier to have already a decision.

1.4. Dissertation Structure

In addition to the introduction, the dissertation contains five main sections. We begin by addressing health literacy and reviewing the existing literature on Web Information-Seeking Behavior. Subsequently, the methodology and its main phases are presented. The fourth chapter addresses the proposal for the classification scheme. In the fifth chapter, we describe the analysis of information-seeking behavior by the level of health literacy with the application of the classification scheme, the results obtained, and its discussion. In the last chapter, the conclusions, limitations of the study, and perspectives of future work are presented.

2. Literature Review

The literature review consists of two main topics: Health Literacy and Web Information Seeking Behavior. In the section on Health Literacy, several definitions of this concept are presented from the perspective of several authors and in a chronological way. Subsequently, the consequences of the low level of health literacy, the promotion of health literacy, and, finally, the methods of assessing the level of health literacy are mentioned. The second section that addresses the Web Information Seeking Behavior is divided into three parts: the first presents the overall structure of Search Engine Result Pages, the second analyses studies of web seeking behavior in general and the third focuses on the health information-seeking behavior.

The literature review was done narratively. About its process, several queries related to health literacy and web information seeking behavior were formulated in generic and domain-specific databases. Queries were introduced in Google Scholar, PubMed, JASIST (Journal of the Association for Information Science and Technology), ACM (Association for Computing Machinery). There were no restrictions regarding the time interval. The articles had to be original, peer-reviewed and written in English.

2.1. Health Literacy

"Literacy" is a term known to most people as the ability to read and write. However, it is a concept that can have multiple definitions and can be defined in a more complex way.

This section covers definitions of health literacy, the impact of low health literacy, statistics, promotion of health literacy, and finally, methods for assessing the health literacy level.

2.1.1. Definition

According to the US National Literacy Act of 1991, literacy is the "ability of an individual to read, write and speak, in addition to calculating and solving problems at the levels of proficiency required to work and work in society, achieving their goals and develop your knowledge and potential " (Irwin 1991). Literacy can be applied to information in general or specific information topics such as health.

The definition of health literacy is not consensual, being a concept that continues to evolve both in its definition and in how it is measured. Although there are several definitions by different authors, the choice of the definition of health literacy must be related to the objective of the study to be carried out (Berkman, Davis, and McCormack 2010). According to

systematic reviews on health literacy, 250 different definitions were found (Malloy-Weir et al. 2016) the most cited are those of the World Health Organization (WHO), Institute of Medicine, American Medical Association (Sørensen et al. 2012). These definitions are presented in the following paragraphs. However, it is necessary to make a chronological analysis to better understand the origin of this concept.

In the 70s, Simonds referred to the need for health education in the context of social policy (1974). Later, in the Health Promotion Glossary of the WHO emerge a reference to the concept (Kickbusch 1997; Nutbeam 1998). The purpose of this WHO glossary was to facilitate understanding, communication, and cooperation in the area of health promotion at local, regional, and global levels. The first edition was published in 1986 and the second in 1998. Later, it was updated to include ten new concepts (World Health Organization 2020).

In 1999, the concept changed and the ability to include not only reading but also numerical tasks was included: "constellation of capacities, including the ability to perform basic reading and numerical tasks necessary to work in the health care environment, such as the ability to read and understand prescription labels, consultation guides, and other health-related materials" (American Medical Association 1999). After one year, the following definition is published by Nutbeam (2000): "Health literacy represents the cognitive and social capacities that determine the motivation and ability of individuals to gain access, understand and use information in a way that promotes and maintains good health ". In the same year, the National Library of Medicine (Selden et al. 2000) presents a definition that is accepted by the Institute of Medicine in 2004 and used by Healthy People in 2010 (Nielsen-Bohlman, Panzer, and David 2005): "the individual's ability to obtain, process and understand basic health information and services necessary to make appropriate health decisions". Healthy People 2010 was a health promotion and disease prevention agenda launched by the Department of Health and Human Services of the United States of America. This agenda contains a list of 467 objectives to improve the health of the population of the United States during the first decade of the 21st century. This list of objectives is launched every ten years, the first of which appeared in 2000 and is currently in preparation for the 2030 agenda (Centers for Disease Control and Prevention 2020).

Berkman, Davis, and McCormack (2010) propose the inclusion of communication skills in the definition: "the individual's ability to obtain, process, understand and communicate health-related information necessary for making informed health decisions".

Health literacy is represented through a conceptual model by Mancuso (Figure 2). The author affirms the presence of the necessary background to obtain skills in health literacy. In

this model, six dimensions of competences are identified: operational (ability to use tools, procedures, and techniques), interactive (collaboration with one another), autonomous (empowering at a personal level), informational (ability to determine the authority of information health), contextual (the domain of the environment) and cultural (interpretation of the meaning of the system of social practices). The attributes of health literacy are ability, understanding, and communication. These are preceded by the skills to achieve health literacy. The results of health literacy have the potential to influence individuals and society (Mancuso 2008).

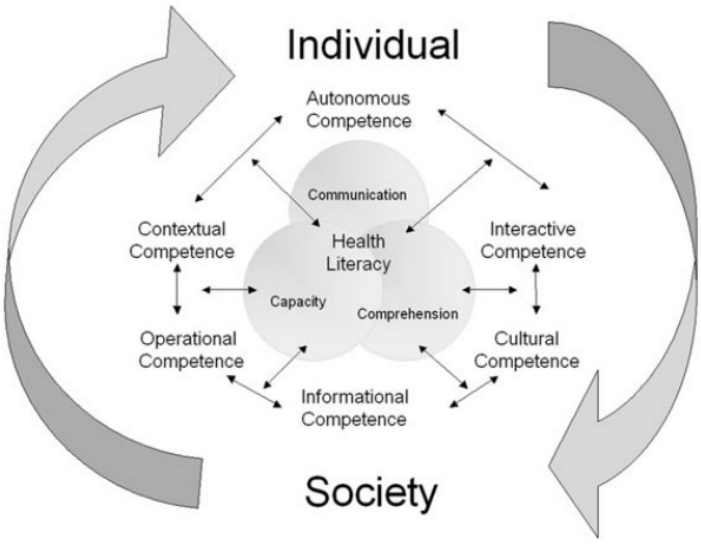


Figure 2 - Conceptual Model of Health Literacy

Source - Mancuso (2008)

Pleasant (2014) draws attention to a broad effort by professionals, researchers, and public members to create an updated definition of health literacy. This effort may be the basis for a new measure of health literacy. Health literacy is not simply the use of skill in the health context, but what people do with the skills they have.

In addition to the notion of health literacy, it is worth mentioning that there is also reference to the concept of "eHealth Literacy", which is characterized by the individual ability to use and understand the content of interactive information technologies. An eHealth literate should be able to read and write, interpret, and communicate health information relevant to him in a given situation and cultural context. It must also be able to apply the tools of accessible technologies (Efthymiou et al. 2017).

2.1.2. Impact of Low Health Literacy Levels

Low health literacy has two types of costs: economic costs to society and the health system costs related to disease's human burden (Nielsen-Bohlman, Panzer, and David 2005).

People with low levels of health literacy tend to have more health problems, recur more to health services, and have higher health care costs (American Medical Association 1999). This situation is even worse in patients with chronic diseases such as diabetes and asthma, where it is essential to follow instructions on a more permanent and continuous basis (Safeer and Keenan 2005).

2.1.3. Statistics on Health Literacy

According to the National Assessment of Adult Literacy 2003, 22% of the American population has a basic knowledge of health literacy and 14% knowledge below the basic level. Most of these individuals obtain health information through television or radio. Regarding age, the percentage of individuals with 65 years or more have a lower average level of health literacy than the younger age groups. In terms of education, health literacy increased with the education level of adults (Kutner et al. 2006).

In the European panorama, data from the European Health Literacy Survey revealed that almost half of the participating population revealed to have a level of health literacy considered inadequate (12%) and problematic (35%). Some groups are more vulnerable, namely those with lower social status, worse health, or relatively old age. Between European countries, there is a great diversity in the levels of health literacy since there are differences in the minimum and maximum in each level (Kickbusch et al. 2013).

In the context of Portugal, the report "Health Literacy in Portugal" prepared by the Center for Research and Studies in Sociology/IUL indicated that about 38% of the Portuguese population has a level of health literacy considered problematic. In the European panorama, Portugal presented slightly lower values compared to the average values of the countries participating in the HLS-EU. It also demonstrated that the lowest levels of health literacy are found mainly in the elderly population (age equal to or over 66 years), low education levels, minimum income, patients with prolonged illnesses, lack of self-perceived health, high frequency in primary health care and limitations due to chronic diseases (Espanha, Ávila, and Mendes 2016).

2.1.4. Promotion of Health Literacy

In the American context, Healthy People is a health promotion and disease prevention agenda launched by the Department of Health and Human Services. It is launched every ten years and the first one appeared in 2000 and is currently in the preparation phase of version 2030. Healthy People 2020 had four main objectives: to provide a higher and higher quality of life, to achieve equity in the health of all groups, creating social and physical environments, and promoting quality of life, healthy development, and healthy behaviors at all stages of life (Centers for Disease Control and Prevention 2020).

The promotion of Health Literacy is seen as an important public health opportunity and challenge. In Portugal, improving health literacy levels is a current challenge (Direção-Geral da Saúde 2019). The general objectives of the plan proposed by Direção Geral de Saúde are the adoption of healthy lifestyles, training for the proper use of the health system, promotion of well-being especially in chronic illness, and, ALS, promoting knowledge and enhancing search. This action plan calls for investment in the area of Health Literacy, namely in promoting new strategies, initiatives, projects, and activities to improve the quality and well-being of the population.

Information and communication technologies appear as an alternative to disseminate health information (Espanha, Ávila, and Mendes 2016). Increasingly, people turn to the web to search for health information. Last year, about 45% of the population of Portugal aged between 16 and 74 years used the web to search for health-related information (Eurostat 2018). The Pew Research Center has been studying the social life of health information since 2000. In one of the last studies, it was confirmed that about seven in ten American adults sought information online about health problems (Susannah Fox 2014). According to the same statistic, searches were performed on search engines, especially on Google (Susannah Fox 2013).

2.1.5. Assessment of Health Literacy

There are several instruments to measure health literacy with different capacities and administration times. The most used methods are Rapid Estimate of Literacy in Medicine (REALM), Test of Functional Health Literacy in Adults (TOFHLA), S-TOFHLA (short version of TOFHLA), Newest Vital Sign (NVS), Short Assessment of Health Literacy for Spanish-speaking Adults (SAHLSA), and Medical Term Recognition Test (METER).

REALM, a screening tool that assesses the ability of an adult to be able to pronounce common medical words and human or disease-related terms, was created. This instrument

was developed to be used in the creation of educational materials appropriate to the patient's level. It is a quick test lasting only two to three minutes. With this instrument, the patient receives a list consisting of sixty-six terms and has to read it aloud. If the patient takes more than five seconds to say the word, it should move on to the next. The result varies according to the number of correct words spoken by the patient. During the test phase, this instrument obtained significant correlations with three others and proved to be a useful tool for assessing the patient's reading ability (Davis et al. 1993).

TOFHLA emerged in 1995 to measure functional health literacy in patients (Parker et al. 1995). This test consists of two main parts: reading comprehension (fifty items) and numeracy (seventeen items). In reading comprehension, the user chooses one of the four options available and only one is correct, the other three are grammatically or contextually incorrect. In the numeracy part, samples of hospital forms and prescription labels are used. This test requires 22 minutes to administer. The test was applied to a total of 505 patients, 256 of whom were English and 203 Spanish together with REALM and the results showed a strong correlation (0.84) (Parker et al. 1995). Due to the long administration time of TOFHLA, S-TOFHLA appeared in 1999, a simpler version of the original with a duration of just 12 minutes. The S-TOFHLA consists of 36 items from the reading comprehension section and 4 items from the numeracy section. In this specific case, it is applied to 211 patients and compared with the REALM in which it obtains a good correlation (0.80) (Baker et al. 1999).

NVS appears in 2005 as a rapid screening test (three minutes) for the assessment of health literacy available in English and Spanish (Weiss et al. 2005). This instrument consists of six questions related to the information present on an ice cream nutrition label. Two of the questions involve the comprehension of information on the label and the other four require numerical calculation. It is a test that assesses comprehension, information seeking, and numeracy. The result varies according to the number of correct answers, that is, between zero and six (Mackert et al. 2013). This test was adapted and validated for the Portuguese version in which it revealed high reliability (Paiva et al. 2017).

In 2010, a health literacy test is developed for populations fluent in English and Spanish, called the Short Assessment of Health Literacy for Spanish-speaking Adults (SAHLSA). This test consists of fifty items and assesses the comprehension and pronunciation of medical terms. It is based on REALM, known as the easiest tool, to assess health literacy in English. It is a test that only requires four to five minutes and has good reliability in the Spanish language (Lee et al. 2006). This model was validated for the Portuguese language, called Short Assessment of Health Literacy in Portuguese-speaking Adults (SAHLPA) (Paiva et al. 2019).

Four years later, METER is developed, which consists of forty medical words and thirty intrusive words that sound like medical terms. For two minutes, the patient only has to select words that correspond to real medical terms. This method was tested in conjunction with REALM, which proved to be a practical test for measuring health literacy (Rawson et al. 2010). This test was adapted and validated as an instrument for assessing health literacy in the Portuguese population (Paiva et al. 2014).

Table 1 presents a comparative analysis of the methods concerning the assessed capacities, the mode of administration, the time of administration, language, the reference to the validated version in Portuguese (if exist), and the example of a question in English.

	REALM (Davis et al. 1993)	**TOFHLA** (Parker et al. 1995)	**S-TOFHLA** (Baker et al. 1999)	**NVS** (Weiss et al. 2005)	**SAHLSA-50** (Lee et al. 2006)	**METER** (Rawson et al. 2010)
Capacities Assessed	Pronunciation	Comprehension, Information Seeking (document), Numeracy	Comprehension, Information Seeking (document), Numeracy	Comprehension, Information Seeking (document), Numeracy	Comprehension Pronunciation	Comprehension
Mode	Verbal	Verbal Written	Verbal Written	Verbal	Verbal Written	Verbal Written
Duration (min)	2-3	22	12	3	4-5	2
Language of validated version	English	English	English	English	Spanish	English
Version validated in PT	-	-	-	NVS Version ((Paiva et al. 2017))	SAHLPA Version (Paiva et al. 2019)	METER Version (Paiva et al. 2014)
Example in English	How many of these words can you read aloud and pronounce correctly, each within five seconds? Start with the first column, reading down. Skip those you cannot read. Fat Flu Pill Dose ...	I want you to figure out which of those 4 words should go in the blank, which word makes the sentence make sense. Your doctor has sent you to have a _____ X-ray. a) stomach b) diabetes c) stitches d) germs	I want you to figure out which of those 4 words should go in the blank, which word makes the sentence make sense. Your doctor has sent you to have a _____ X-ray. e) stomach f) diabetes g) stitches germs	This information is on the back of a container of a pint of ice cream. If you eat the entire container, how many calories will you eat?	Now please read the word above aloud. Stem: Prostate Which of the two words is more similar to the word above? If you don't know the answer, please say, 'I don't know'. Key or Distracter: gland or circulation.	As you read through the list, put an "X" next to the items that you know are real words. _ Irrity _ Arthritis _ Obesity _ Flu _ Behaviose

Table 1 - Comparative analysis of Health Literacy assessment methods

2.2. Web Information Seeking Behavior

This section consists of three subsections related to the structure of Search Engine Result Pages (SERP) and studies on web information-seeking behavior. The first section deals with the structure of SERP. In the second part, existing studies focused on web information-seeking behavior in general are mentioned. In the last one, studies related to the search for health information on the web are addressed.

2.2.1. Structure of Search Engine Result Pages

Disregarding the search engine, the current SERP has various elements (Moz 2020). In this section, we present the main SERP elements and consistent terminology that will be used in the rest of the work.

The main element in SERP are the organic results, that is, the results that emerge naturally from the Web. Overall, they consist of a link, an URL, and a snippet (Figure 3).

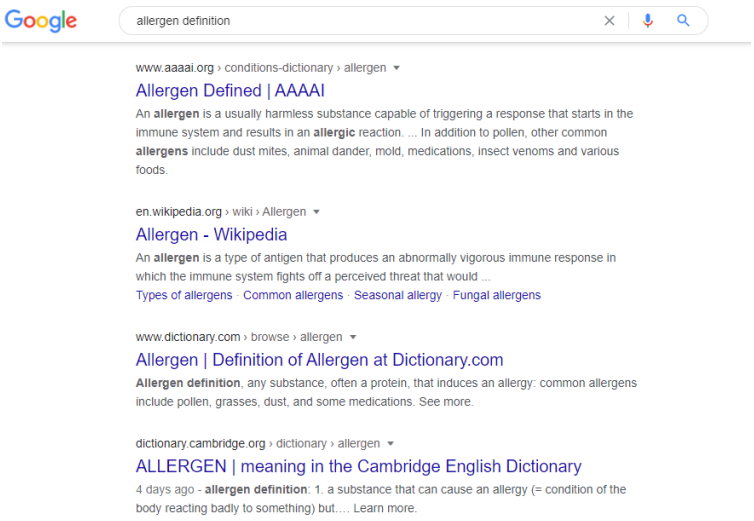


Figure 3 - Organic results (Google Search Engine)

SERP also have paid results that can have images and reviews (Figure 4) or be presented as text results with an Ad tag (Figure 5).

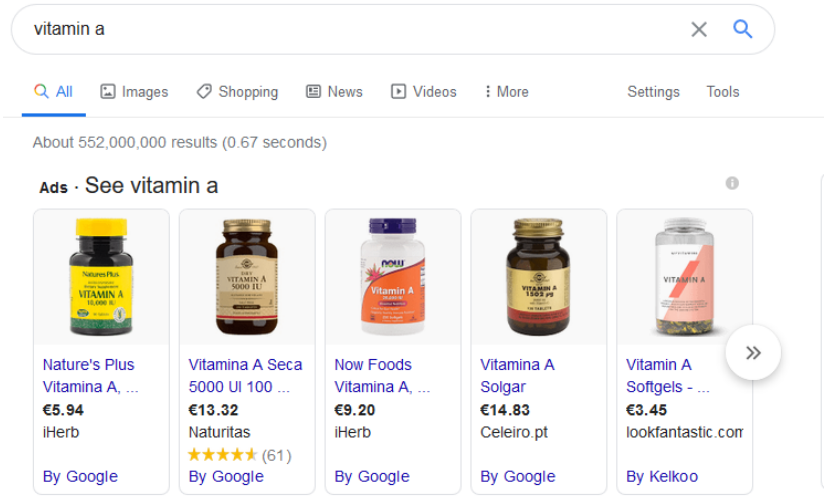


Figure 4 - Paid results with images and reviews (Mozilla Firefox search engine)

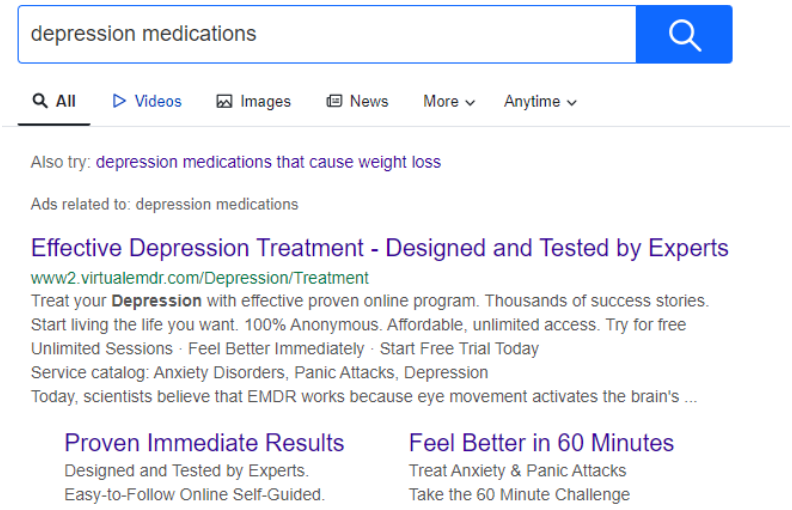


Figure 5 - Advertisement (Yahoo search engine)

Besides, SERP have results that fit in a category of Universal Results. This category includes results that consist of images, videos, featured snippets. Featured snippets are excerpts from website content that summarize a response to the submitted query (Moran and Goray 2020). In this sense, the answer box is an example of a featured snippet (Figure 6). People Also Ask feature is a list of questions related to the query. Each question unveils a featured snippet along with additional questions. For this reason, collections of feature snippets are considered for related queries (Moran and Goray 2020). So, we included the People Also Ask feature (Figure 7). We also consider Google Scholar articles that may appear at the top of the SERP to be part of this category.

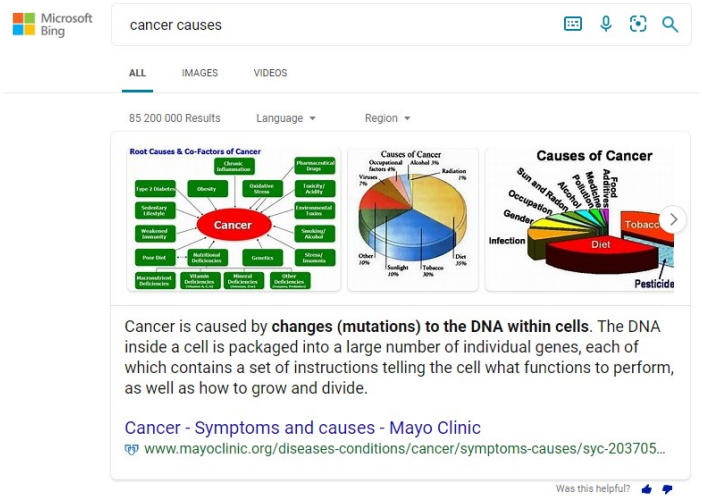


Figure 6 - Answer box (Microsoft Bing search engine)

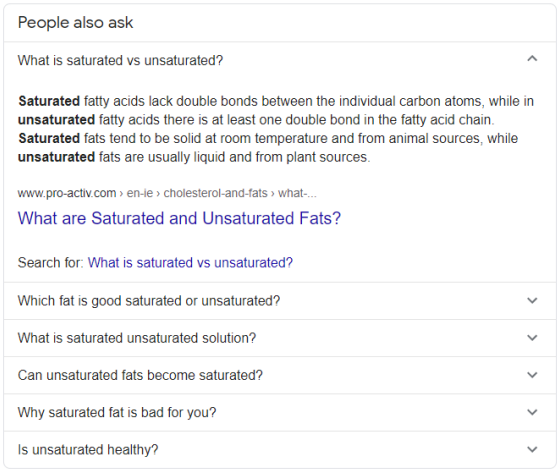


Figure 7 - People Also Ask feature (Google search engine)

Besides results, SERP has other features that enrich the simple list of results. Snippets can be enriched by visual elements as review stars for product ratings (Figure 8).

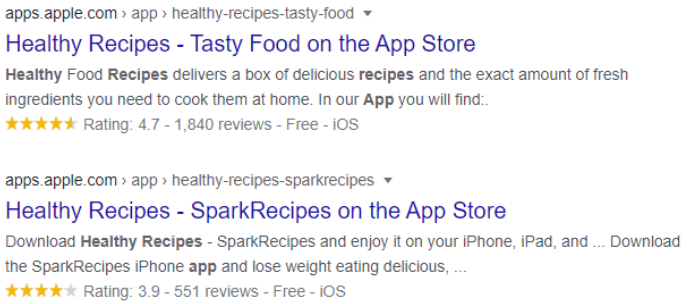


Figure 8 - Enriched Snippets (Google search engine)

The knowledge graph is an information source that provides the data that appears as panels or boxes such as the weather or a celebrity knowledge panel (Figure 9).

Figure 9 - Knowledge Data (Google search engine)

Finally, some tools can be embedded in SERP such as the calculator, converter, and translator (Figure 10).

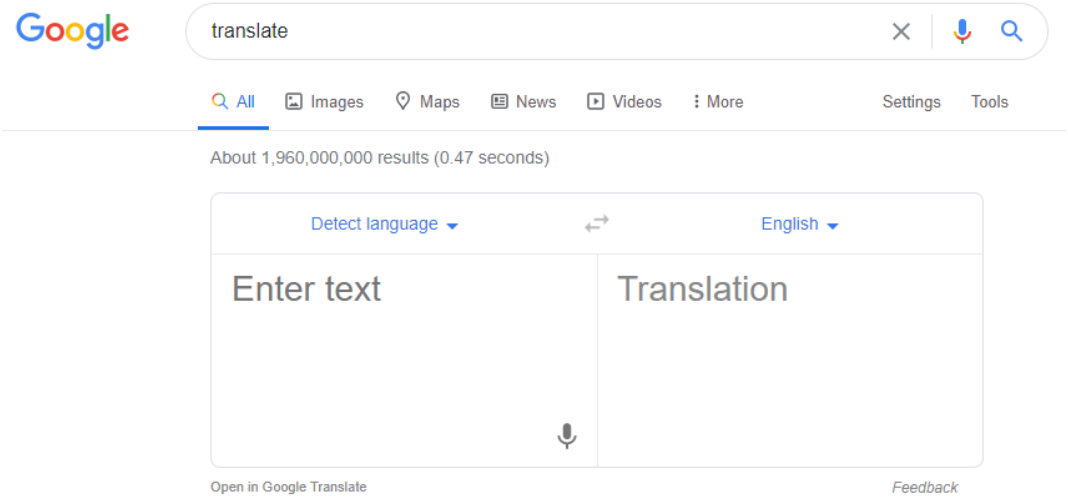


Figure 10 - Example of the translator tool (Google search engine)

2.2.2. Studies on Seeking Behavior in General

In this section, we focus on studies that analyze web information-seeking behavior in general, and studies that focus on query (re)formulation behavior in particular.

Works focused on web information seeking behavior concentrate on a variety of goals and approaches.

According to the studies mentioned, it is possible to divide them into three groups: studies about user characteristics, studies about user characteristics related to the search process, and studies related to the influence of aspects of search.

Concerning user characteristics, some studies investigate the web expertise (Hölscher and Strube 2000; White and Morris 2007), the domain expertise (White, Dumais, and Teevan 2009), the topic familiarity (Diane Kelly and Cool 2002), and the visually impaired and sighted searchers (Sahib, Tombros, and Stockman 2012).

Regarding user characteristics related to the search process, studies address the relationship between the search query and information goals (Downey et al. 2008), the users' information search strategies (Zhu, Liu, and Song 2018), the variability in user intent (Teevan, Dumais, and Liebling 2008), the user satisfaction (Steve Fox et al. 2005; J. Liu and Han 2020) and preference (D. Kelly and Belkin 2001).

Others investigate aspects related to the search process like the influence of tasks (K. Huang et al. 2020; Li and Belkin 2010), the effects of individual and task differences (Buscher et al. 2012), the influence of result summary length (Maxwell, Azzopardi, and Moshfeghi 2017) and cursor behavior (J. Huang, White, and Dumais 2011).

Some studies focus on applying predictive models to interpreting search behavior (Q. Guo and Agichtein 2010), useful documents in different tasks (C. Liu 2012), user click (Chen et al. 2012), search satisfaction (Hassan 2012; Hassan and White 2013; Steve Fox et al. 2005), expertise (White, Dumais, and Teevan 2009; X. Zhang et al. 2015), query performance (Q. Guo, White, and Dumais 2010), user preference (Agitchtein et al. 2006), search engine switching (White et al. 2008; White and Dumais 2009) and to identify queries that can benefit from personalization (Teevan, Dumais, and Liebling 2008).

The studies were analyzed to identify measures of analysis to inspire us for the proposed classification scheme.

Table 2 shows the most used measures in these studies divided into categories: session, browser, search engine results page, query, and webpage. The terminology, the form of analysis, and the grouping of the measures vary from study to study. The criterion for grouping the

different measures was based on the category where the measure or event can take place.

	Measure	Article																								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Session	End Action Type					x																					
	Total Time			x		x			x	x			x				x					x	x	x		x	x
	# Sessions																					x	x				
	# Actions																	x								x	
Browser	# Backward or Forward					x					x		x									x		x			
SERP	Result Position	x		x		x		x				x						x			x	x		x		x	
	Deepest or Lowest Rank																	x									
	Time to Click {First or Last}		x			x	x		x	x							x	x			x			x			
	Time to First Scroll					x																					
	# Scrolling		x										x														
	# Clicks	x	x	x	x		x	x	x	x	x	x	x					x	x		x	x		x			
	# Accessed						x						x			x	x		x	x		x	x			x	x
	# No Click, One-Click, Or Multiple Clicks		x	x				x		x		x										x					
	Time Spent		x					x					x				x	x									x
Query	# Queries			x	x	x		x	x	x	x	x	x			x	x		x	x		x	x	x	x	x	x
	# Tokens, Characters, and (Non-Stop or Stop) Terms				x		x	x		x	x		x			x		x		x	x	x	x	x	x	x	
	# Query Reformulation				x		x				x									x							
	# Queries with Technical Terms																						x				
Webpage	Time Spent	x				x	x			x	x		x	x	x		x	x	x				x	x		x	x
	# Scrolling					x	x						x	x													
	Exit Type					x																					
	# Visited				x	x	x				x		x		x	x	x	x		x			x	x		x	x

1 (Agitchtein et al. 2006); 2 (Buscher et al. 2012); 3 (Chen et al. 2012); 4 (Downey et al. 2008); 5 (Steve Fox et al. 2005); 6 (Q. Guo and Agichtein 2010); 7 (Q. Guo, White, and Dumais 2010); 8 (Hassan 2012); 9 (Hassan and White 2013); 10 (Hölscher and Strube 2000); 11 (J. Huang, White, and Dumais 2011); 12 (K. Huang et al. 2020); 13 (D. Kelly and Belkin 2001); 14 (Diane Kelly and Cool 2002); 15 (Li and Belkin 2010); 16 (C. Liu 2012); 17 (J. Liu and Han 2020); 18 (Maxwell, Azzopardi, and Moshfeghi 2017); 19 (Sahib, Tombros, and Stockman 2012); 20 (Teevan, Dumais, and Liebling 2008); 21 (White and Dumais 2009); 22 (White, Dumais, and Teevan 2009); 23 (White and Morris 2007); 24 (White et al. 2008) ; 25 (X. Zhang et al. 2015); 26 (Zhu, Liu, and Song 2018).

Table 2 - Overview of measures of general behavior studies

In the session category, the total time is the most used measure. In the Browser category, there are some measures related to backward and forward buttons or operations. Concerning SERP, most of the measures are related to the number of clicks. Regarding this measure, some studies specify clicks on results and others measure the number of clicks in general, whether in a result or not. The types of clicks that are mentioned are clicks on organic results, sponsored or advertisements results, and clicks on other parts of the SERP such as pagination (first page, previous and previous page), related searches, spelling suggestions. About the query, the measure relative to the number of terms per query varies widely from study to study. Some studies analyze the number of characters, tokens, terms, and non-stop or stop words. The most used measures are the number of terms and the number of queries. On the web pages, there is a greater occurrence of measures number of pages visited and time spent on a webpage.

Regarding query formulation and reformulation behavior, there are several studies focused on this specific behavior. The query formulation stage is considered a crucial stage in the information retrieval process. This step involves a transformation of the concepts resulting from the step of articulating the need in terms of search, selecting the correct terms, and combining them with Boolean operators. Always taking into account the specific interrogation syntax of the search system (Vanopstal et al. 2012). The formulation is the initial stage in which the search strategy is built. It is normal that for a given information need to be satisfied, more than one query is submitted. The reformulation is the manual modification or with the aid of the initial stage system (Mastora, Monopoli, and Kapidakis 2008) but which addresses the same need for information (J. Huang and Efthimiadis 2009).

Several studies analyze query formulation and reformulation behavior. We present the description of the types of reformulation (Table 3) and an overview of the reformulation types considered in each article (Table 4).

Type	**Description**
Acronym (expand or form)	Form occurs when the reformulated query is an acronym formed from the initial query's words. Expand occurs when the initial query is an acronym and the reformulation is a query consisting of the words that form the acronym.
Addition	This occurs when the initial query is a subset of a reformulated query, that is, all the terms in the initial query are present in the reformulated query. The reformulated query contains more terms than the initial query.
Capitalization	This occurs when there is a capitalization change from the initial query to the reformulated query.
Interruption	This occurs when a query on a topic searched on earlier by a user that has been interrupted by a search on another topic.

Type	Description
Miscellaneous	This occurs when didn't fit in other types or does not contain a query to be allocated to any of the search patterns.
New	This occurs when the initial query and the reformulated query don't contain any common or synonymous terms.
Operator Usage	This occurs when the reformulated query contains Boolean operators, plus sign, minus sign, or quotation marks for a phrase.
Parallel movement	This occurs when the initial query and the next query have partial overlap in meaning, or two queries are dealing with somewhat different aspects of one concept.
Remove	This occurs when the initial query and the reformulated query contain at least one term in common, but the reformulated query contains fewer terms than the initial query.
Reorder	This occurs when the words in the initial query are reordered but unchanged otherwise, producing the reformulated query.
Repeat	This occurs when the initial query and reformulated query contain the same terms, but the format of these terms may be different.
Replace with Synonym	This occurs when the initial query terms are replaced with words that share similar meanings in the reformulated query.
Spelling Correction	This occurs when the term has an error in the initial query and is corrected in the reformulated query. It can be detected automatically using the Levenshtein distance function.
Stemming	This occurs when changing the word stems in the first query. Porter's stemming algorithm can be used.
Sub/Superstring	Substring occurs when the reformulated query is a strict prefix or suffix of the initial query. Superstring occurs when the reformulated query contains the initial query as a prefix or suffix.
Switching Topic	This occurs when coming up with a new concept that did not overlap with the concepts in the previous query.
Term Variation	This occurs when there is a variation in the term. Variations include spelling out the abbreviation, adding a preposition, changing from the singular to the plural, or vice versa.
Types of Resource	This occurs when changing the query terms, making changes in a resource (e.g., articles, pictures) while keeping the same meaning of the query itself.
URL Stripping	This occurs when changing the query to locate the Web site domain.
Whitespace or Punctuation	This occurs when the initial query is a whitespace and punctuation reformulation of the first query if only whitespace and punctuation are altered in the reformulation.
Word Substitution	Occurs when the two queries contain at least one term in common; the Reformulated query has the same length as the initial query but contains some terms that are not in the initial query

Table 3 - Description of the types of (re)formulation

Type	Article																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Acronym (expand or form)	x			x			x									x				
Addition	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Capitalization			x													x				
Interruption											x									
Miscellaneous			x		x															
New					x		x	x	x	x	x	x	x				x			
Operator Usage		x		x										x			x	x	x	x
Parallel movement															x				x	
Remove		x	x		x	x	x				x	x	x	x		x		x	x	
Reorder	x						x									x				
Repeat			x			x	x			x		x	x			x	x	x		
Replacement with Synonym	x											x		x		x	x			
Spelling Correction	x		x	x			x							x		x	x	x		
Stemming		x	x	x			x									x	x			
Sub/Superstring							x													
Switching Topic	x							x	x										x	x
Term Variation	x		x			x	x							x		x			x	
Types of Resource														x						
URL Stripping							x							x		x				
Whitespace or Punctuation			x	x			x									x				
Word substitution	x		x				x			x		x	x			x	x	x	x	

1 (Anick 2003); 2 (Aula 2003); 3 (Bruza and Dennis 1997); 4 (J. Guo et al. 2008); 5 (He, Göker, and Harper 2002); 6 (Hembrooke et al. 2005); 7 (J. Huang and Efthimiadis 2009); 8 (Jansen, Booth, and Spink 2009); 9 (Jansen et al. 2007); 10 (Kinley et al. 2012); 11 (Lau and Horvitz 1999); 12 (C. Liu, Gwizdka, and Belkin 2010); 13 (C. Liu et al. 2010); 14 (Rieh and Xie 2006); 15 (Spink, Jansen, and Ozmultu 2000); 16 (Teevan et al. 2007); 17 (Whittle et al. 2007); 18 (Wildemuth 2004); 19 (Yamin and Ramayah 2011); 20 (Y. Zhang et al. 2012).

Table 4 - Overview of the types of query (re)formulation used per article

In the type of reformulation acronym, Anick (2003), Huang and Efthimiadis (2009) considered their formation and expansion. Guo, Xu, Li, and Cheng (2008) consider their expansion. Teevan, Adar, Jones, and Potts (2007) refer to abbreviations.

The addition is one of the most used types in studies (Bruza and Dennis 1997; J. Huang and Efthimiadis 2009; Wildemuth 2004; Kinley et al. 2012; Spink, Jansen, and Ozmultu 2000; Teevan et al. 2007; Y. Zhang et al. 2012). There are many similarities with this type such as head and modifier, precise query, specialization, elaboration, and refining, specification, a conjoint modification, which extends query retained it as a sub phrase and depth query.

Capitalization is mentioned and also referred to as case changing. Lau and Horvitz (1999) included the interruption that happens when the query on a topic searched on earlier by a user that has been interrupted by a search on another topic.

In cases where no type fits or cannot be allocated to any type of search pattern, miscellaneous type and other types are considered. The new queries, those that do not have terms in common or are on another topic are also counted. Repeat queries, redundancy, same queries, and duplicates of words are other types examined.

In the use of operators, changes in the use of Boolean operators (AND, OR, NOT), plus and minus sign (+,-) and quotation marks for a phrase (“”) are considered. Parallel movement is another type mentioned. It happens when the previous queries and the follow-up queries have partial overlap in meaning, or two queries are dealing with somewhat different aspects of one concept.

Remove words is also common. Others refer to generalization, breadth query, broad query, and broadening. Substitution and replace by synonyms are also evaluated. Similar to these types are alternative, hyponym and change, word swap and synonymous, reformulation, replace and substitution, replacement with synonyms. Spelling correction was also studied. The reorder of words is also analyzed, Anick (2003) refers to a syntactic variation. Stemming is also a type used but there are other related transformations like morphological variants, derived, truncation, and modification on changed word endings.

Huang and Efthitimiadis (2009) also used types of substring and superstring. URL stripping, domain suffix are other examples of types used. Rieh and Xie (2006) used types of resources. Term variation includes abbreviation, prepositions, changing singular to plural, and vice versa. Finally, there are mentioned the types of switching topic, punctuation and whitespace, non-alphanumeric, extra whitespace, word merging/splitting, and joining.

2.2.3. Studies on Health Information Seeking Behavior

This section covers studies that focus on health information-seeking behavior (HISB).

Regarding scientific literature reviews, there are some studies. In the area of health, search behavior leads us to the concept of Health Information Seeking Behavior (HISB). This concept can be defined, at least, in eleven different ways. HISB focuses on how users obtain health information, including the actions and strategies they use. It can be defined as how people obtain information about health, health promotion, diseases, and health risks (Lambert and Loiselle 2007). A more recent concept analysis shows that there is a lack of theoretical inclusion in the HISB studies, suggesting the need for inclusion of theoretical frameworks, particularly information-seeking models (Zimmerman and Shaw 2020). Cutilli (2010) states that the definitions of this concept are not explicit and that there is no dominant unification. In a literature review on health consumer information-seeking behavior, it was found that the search for health information is associated with several factors and, therefore, the search behavior varies according to the type of information sought, the search reasons, and the levels of experience. Searching for information online is seen as a supplement but not a substitute for guidance from healthcare professionals (Higgins et al. 2011).

Regarding health information search experiences among consumers in general, there are several studies. Eysenbach and Köhler (2002) investigated the retrieval and assessment techniques used by consumers when searching for health information on the internet. The participants used search engines as a starting point. Regarding the search strategy, few search queries contained more than one search term. Most of the participants only consulted the first page of results, few were those who consulted the following pages. For each question, it took about 6 minutes to find the answer. Toms and Latter (2007) conducted four health search tasks on the google search engine. Participants create about 1.3 queries for each search task. Each query contained about 4 keywords, 3 of which stopped words. Participants spend about 4.5 to 9 minutes per task and accessed 2.6 web pages. Participants spent a lot of time interpreting the list of results while understanding the information presented. The formulation of interrogations was carried out quickly and in trial and error.

Mukherjee and Bawden (2012) found that users search to obtain information about a health problem or data about a health problem that is affecting someone close to them. They also found that the greatest difficulties encountered in the search for health information on the web are the excess of information, the lack of indication of the reliability of the content, the lack of monitoring of the websites, and the existence of barriers to access. It is important to understand how the sources complementary to the information provided by the doctor are

used to what extent, the factors age, race, education, health literacy alter search behaviors (Cutilli 2010). Jacobs, Amuta, and Jeon (2017) found that there are still people who prefer to seek and obtain information through traditional means (books, leaflets, brochures, and magazines) or from health professionals as the primary source of health information. The factors that lead to this preference are trust and access to the information source.

A study related to the search for healthy food information investigated the search behavior of users in five European countries through a questionnaire. They concluded that health information behavior and literacy are correlated with the individual and with environmental characteristics. A predominant part of the interviewees uses the Google search engine to find nutritional information on the internet, few are those who go directly to online nutrition services. Consumers with less education and socio-economic power are less familiar with the sources of nutritional information and rarely try to look for it (Niedźwiedzka et al. 2014). In another study conducted by a qualitative approach, it aimed to explore the perceptions of online health information search. Most participants confirmed the use of the internet to access health information and described a common experience of searching, filtering, and comparing results to obtain relevant information according to the search need. The main reasons for completing the survey are information saturation and fatigue (Fiksdal et al. 2014).

Gutierrez et al. (2014) conducted a study to measure health literacy in two clinics and determine associations between health literacy, access to information, and internet usage. There were no significant differences between patients with a limited health literacy level and their primary information source. One of the studies presented focused on the impact of contextual features such as age, sex, mother tongue, health literacy, the level of experience on the internet, and the frequency of health information seeking on health information searching. The authors concluded that the frequent search for health information leads to longer queries and that this frequency, the mother tongue, and the health status influence the formulation of queries (Yilma et al. 2019). Yom (2016) evaluated the effects of low health literacy on search behavior, in particular how people use the internet to present diabetes. Dwell times for users with low health literacy are higher as they spend more time reading and concentrating on the page. However, it was found that on the pages that required more reading time, people with adequate health literacy spent more time reading than people with limited health literacy.

In the analysis of health information queries, Hong (2002) analyzed search questions sent directly to a health website or via external search engines. Most users searched using search engines instead of searching directly on the health page. The queries surveyed on the health website were short and simple, about one to two terms. McCray and Tse (2003) studied

the success factor of search questions in two health information systems. They analyzed the search questions, developing a coding scheme. In the end, they built a taxonomy to check if the search failures were due to a content problem, formulation, or system limitations. Most of the problems were related to the misuse of supported operators (or not), spelling errors, terminology different from the terms present in the system, or outside the scope. Spink et al. (2004) also analyzed health search data on two search engines and a question and answer platform. On search engines, the average terms per query were 2.3 and the medical pages of the 10 sites viewed per session were 1.5. About 24% of the questions in the question and answer platform were about health. In large part, users do not reformulate their searches and in terms of length, they are very short.

Concerning the age factor, Miller and Bell (2012) concluded that the use of the internet to search for information was negatively associated with age. Age was negatively associated with confidence and belief that searching for information is easy. Manafo and Wong (2012) found that the participants refer to their flow of information online as incomprehensible or information overload. Participants reflected on the challenges of filtering large amounts of information as well as its quality. In a more recent study, a controlled experiment was carried out in which the elderly completed three search tasks (Wu and Li 2016). Regarding the construction of interrogations, they used an average of 1.5 per task. The average length was 8.4 characters (Chinese) and the average time was 122 seconds. Most users copied the search task description into the search box. Participants clicked on about 3 web pages per task, most of which were mostly on the first page of search engine results. Lam (2012) found that elderly people with a medium level of education and greater health literacy tend to be more frequent users in the search for health information. The greater the health literacy, the greater the frequency, and involvement in search behavior.

The influence of search tasks was also studied. Zhang (2012) also observed the search behavior of users, conducting three surveys on the health site MedlinePlus. The sessions took about 20 minutes and the queries were short, using only about 2 to 3 terms. The most frequent patterns in the formulation of queries were addition, elimination, substitution, and repetition of concepts. Two types of actions related to forms were also evident: changing a form of a term and correcting spelling errors.

About the preference factor, Zhang (2013) found that the high preference group searched with shorter questions and made more parallel moves in the reformulation than the low preference group. High preference users were more likely to use more general searches when looking for specific factual information and were more demanding with the search system. In another study, health literacy was a significant predictor of information seeking

preference in older people (Kim and Utz 2018).

Regarding health vocabularies for consumers, patients use expressions to describe health concepts that differ from users by professionals. In this sense, Zeng and Tse (2006) explored the relationship between the expressions of professionals and consumers and developed a consumer health vocabulary. About health literacy and terminology, it was demonstrated how the level of health literacy, familiarity with the topic, and the terminology used in search questions affect the behavior of users' health search. Regarding the level of health literacy, it was revealed that users with an inadequate level tend to be less successful in web search as well as showing more difficulties in formulating search questions. These users and those unfamiliar with the topic, resort less to medical-scientific terminology compared to users with higher levels of literacy and familiarity with the topic (Lopes and Ribeiro 2015). Lopes and Ribeiro (2016) developed an interrogation suggestion system that provides alternative interrogations combining the users' native language and English with lay and medical scientific terminology. Participants mostly resort to suggestions at the beginning of the sessions. In interactions with suggestions, participants use only about 1.34 terms. At higher levels of health literacy, medical scientific terms are preferred over lay terms. The most surprising result was the preference for suggestions in English rather than Portuguese.

One study examined the effects of familiarity with the topic and search capabilities in reformulating queries on health information search behavior (Hu, Lu, and Joo 2013). To analyze the reformulation actions of queries, they used the scheme proposed by Rieh and Xie (2006). Participants with greater familiarity with the topic tend to make fewer mistakes and use specific terms. Participants with greater search skills both generalize and specify their queries and make fewer mistakes. Participants more familiar with the topic completed their search with little reformulation effort (Hu, Lu, and Joo 2013).

A more recent study aims to investigate the relationship between health search behavior and digital health literacy (Chang 2020). Users will conduct search tasks and three types of behavior will be analyzed: the formulation of questions, the evaluation of information, and the extraction of information. For that, it will be used the taxonomy of Huang and Efhitimiadis (2009). Health literacy will be measured using the Digital Health Literacy tool.

The effect of the expertise was also explored in the medical field in the search for information on the web and found differences in the web pages visited, in the vocabulary, and the search behavior. Domain experts visited mainly technical websites while newbies visited consumer and health-oriented websites. Experts issued longer interrogations used more technical terms than non-experts and spent more time on search than novices (White, Dumais,

and Teevan 2008). In another study, experience in the medical field was also assessed by clarifying search queries on a search engine. Although many documents on the web use lay and medical terms, the results of the study show that this effort is still insufficient. The implicit clarification of the question is useful and does not require the user to know the correct medical terminology (Soldaini et al. 2016).

Table 5 shows the methods used for each study mentioned based on the proposed objectives.

Objective	**Methods**	**Articles**
Investigate scientific literature	Type of studies, inclusion criteria, terms used, and analysis of the studies	(Lambert and Loiselle 2007; Zimmerman and Shaw 2020)
	Models and theories, sources of health information, and factors associated	(Cutilli 2010)
	Number of databases and types of articles used, search terms, inclusion criteria	(Higgins et al. 2011)
Consumers' search for health information	Focus Groups, Usability Tests, and in-depth interviews	(Eysenbach and Köhler 2002)
	Questionnaires, transaction logs, video screenshots and audio recordings, search tasks	(Toms and Latter 2007)
	Empirical questionnaire survey and semi-structured interviews	(Mukherjee and Bawden 2012)
	Information seeking survey in five European countries and computer-assisted online web interviewing	(Niedźwiedzka et al. 2014)
	Focus groups and qualitative data analysis	(Fiksdal et al. 2014)
	Assessment of health literacy (NVS and filling medical forms), health information access, and internet usage	(Gutierrez et al. 2014)
	A cross-sectional survey, user characteristics, logistic regressions	(Jacobs, Amuta, and Jeon 2017)
Context features on health information	An online survey about user features, search tasks, and session	(Yilma et al. 2019)
Impact of Low Health Literacy	Log files, extraction of queries, scoring pages, community-based health literacy, categories of pages	(Yom-Tov et al. 2016)
Investigate queries related to health information	Log files, query data collection, and analysis	(Hong et al. 2002; Spink et al. 2004)
Investigate the success of queries	Codification scheme, queries, taxonomy to verify fails	(McCray and Tse 2003)
Explore age in health information-seeking	In-depth interview, a coding framework to analyze interview transcripts	(Manafo and Wong 2012)
	Health Information National Trends Survey (Hints) sample, internet use, logistic regressions	(Miller and Bell 2012)
	Controller user experiment, search tasks, factors between health information seeking and user characteristics	(Wu and Li 2016)
	Search data, health literacy measure using a specific scale to measure proficiency and internet usage	(Lam and Lam 2012)

Objective	Methods	Articles
The influence of search tasks	Search tasks, demographic information, video recording of a search session, interviews, search behavior measures	(Y. Zhang et al. 2012)
Influencer of health literacy on information seeking preference in older people	Face-to-face surveys, assessment of preference	(Kim and Utz 2018)
The effects of preference in search behavior	Scales to assess consumer preference and search tasks	(Y. Zhang 2013)
Impact of terminology	Different tasks, health literacy (SAHLSA), topic familiarity (Likert scale), and medical accuracy	(Lopes and Ribeiro 2015)
	English proficiency (instrument developed by the European Council), health literacy (meter), familiarity with the topic (Likert scale), question suggestion system	(Lopes and Ribeiro 2016)
Exploring and developing consumer health vocabularies	Bottom-up approach for exploring the relationship between consumer and professional expressions, consumer health vocabulary	(Zeng and Tse 2006)
Influence of e-health literacy and search behavior	Search tasks, types of behavior, eye tracker, a taxonomy of queries, perspective think aloud, eye tracker measures, assessment of health literacy (digital health literacy)	(Chang 2020)
Effects of Topic Familiarity on query reformulation	Design experimental Health IR System, medical articles, familiarity with the topic, usability test, classification of query reformulation	(Hu, Lu, and Joo 2013)
Impact of medical expertise in search behavior	Synonym mappings for clarifying search queries, classifier capable of selecting question clarification	(Soldaini et al. 2016)
	Search logs, division of experts and newbies according to the visit to the PubMed search engine	(White, Dumais, and Teevan 2008)

Table 5 - Studies on Health Information Seeking Behavior

3. Methodology

The methodology consists of five major phases that are described in Figure 11. Each phase and methodological step will be described next.

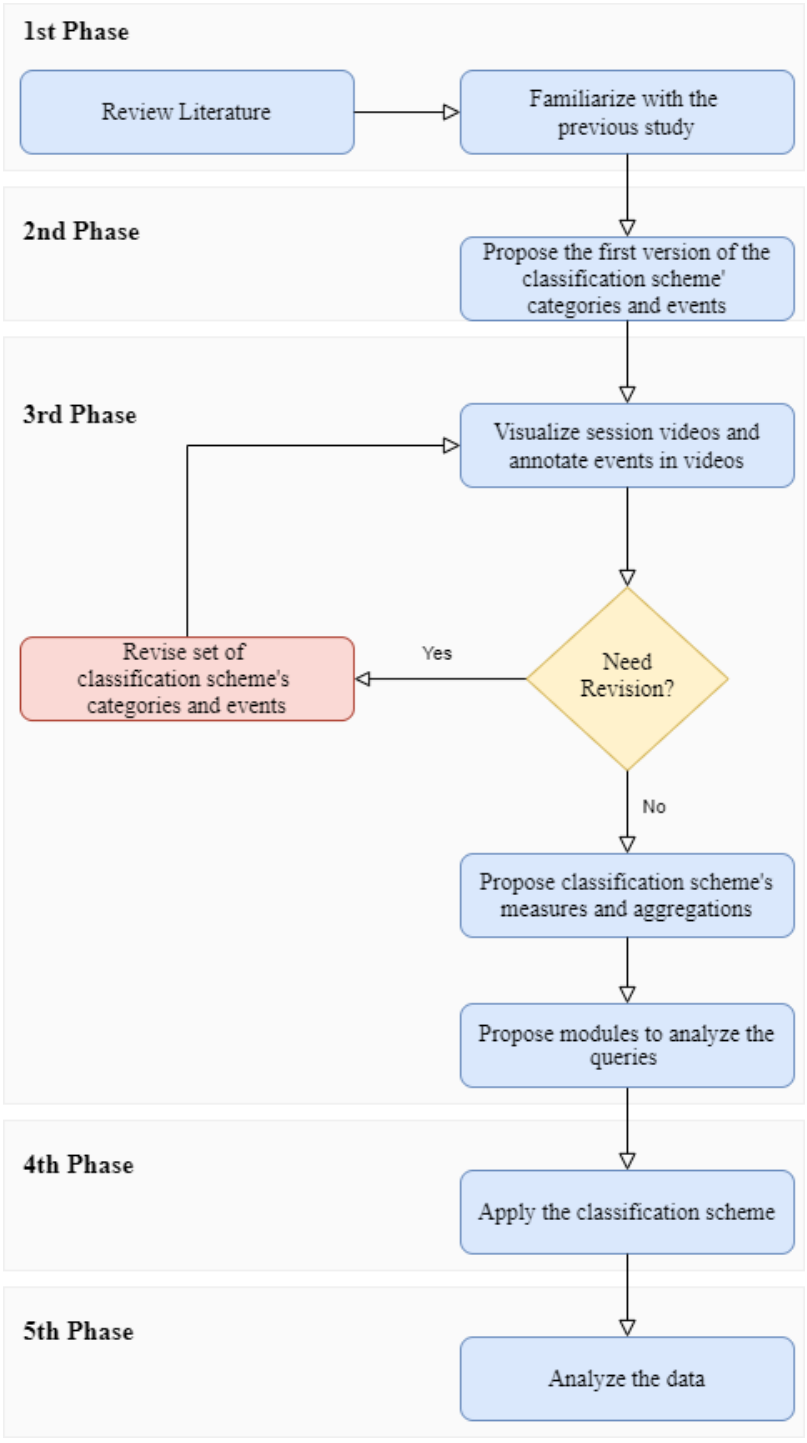


Figure 11 - Diagram of methodological steps

3.1. Literature and Familiarization

In the first phase, we reviewed the existing literature to identify features used by other authors to analyze information-seeking behavior. Subsequently, we familiarized ourselves with the user experiment that was previously conducted.

In the study, users had to conduct online searches to complete ten tasks, five regarding asthma, and five related to nutrition topics (Appendix 1). They could use any search engine, and there was no time limit for finishing the tasks. According to Jansen, Booth, and Spink (2009), a session is "a series of queries submitted by a user and related interactions during an episode of interaction between the user and the Web search engine around a single topic". In this sense, the study consists of ten search sessions per user. Although there is a part of asthma and a part of nutrition, there are very specific questions in which there is not a direct relationship with the previous question. Furthermore, it would not be possible to divide it into parts since we do not have the complete session of two users.

The experiment counted with the participation of 20 users aged between 21 and 35 years. At the time, 95% were completing a bachelor's or master's degree. The remaining part (5%) was a doctorate. Most of the participants were experienced in the fields of Computer Science and the others studied Multimedia and Design. They had the option to choose between the test being in Portuguese or English. Regarding their experience of searching for health information on the web, 45% stated that they rarely search and 55% confirmed that they did it sometimes (Sousa 2016).

Before starting the experience, the degree of health literacy was measured using the Portuguese validated version of the METER instrument (Paiva et al. 2014). This instrument consists of forty medical words and thirty non-words, made up, that sound like medical terms (Appendix 2). The study sample of twenty people revealed that eight (40%) have a degree of literacy considered inadequate. Based on this assessment, users were divided into two groups: low health literacy (LHL) and high health literacy (HHL).

The individual experiences of users when searching for information on the web were all recorded in video format. However, there is no information about the user with ID 2, so only 19 participants were considered. We also use the logs of the queries submitted by users during the search sessions.

3.2. The first version of the classification scheme

Based on the knowledge acquired from the 1st phase, we proposed an initial set of interactions categories and their events. We decide to define the categories by the location

where the interactions took place. The events corresponded to concrete actions that could happen in these places.

The initial part of the annotation was made by Samuel Serafim. In this way, he made the choice and data recording of the various types of events and visualized the videos. He chose to annotate the videos by starting to build a set of types and respective events based on user interactions. The initial version of the classification scheme consisted of seven categories (Appendix 3) and the annotation made contained the following information (Appendix 4). After that, we started to review the set of categories and events that he considered.

3.3. Revision of the classification scheme

After defined the initial set of categories and events, we began to visualize each video and started logging each event that was included in the classification scheme and occurred in the session. Whenever we found an interaction in a video that was not considered by the actual version of the scheme, we updated it. This could mean the simple addition of new events or a reorganization of events. As an example, initially, we have not included the Direct Access Page (DAP) category and its events that will be addressed later. We were not expecting visualizations without a SERP as a referrer. Yet, some users opened a new tab to access the Google Translator website and obtain a translation to use in query formulation. This behavior alerted us to the fact that, during a search session, this is a plausible action and has to be considered.

At all times, we aimed to cover all possible interactions, disregarding of appearing or not in the videos that were being analyzed. Updates in the classification scheme triggered the verification of past video analysis and the continuation of the logging process. Although most of the visualized search sessions were conducted on Google, we assured that the events included were as generic as possible.

This was a manual, meticulous, and time-consuming step of the overall process. Each user session consisted of 20 to 40 minutes per user, but its analysis required much more.

After the set of events reached a stable version, we proposed the first level and second level measures. First-level measures are related to the number and duration (in seconds) of each event. Based on first level measures, we also have second level measures through aggregation.

Afterward, given the specificity of query (re)formulation, we built two additional modules to analyze the queries submitted by users and the types of reformulations carried out. This was made based on the literature review conducted in the 1st phase and the analysis of the user sessions. To evaluate the set of queries and the types of reformulation, it was necessary

to determine when a query is a reformulation. For that, we considered the definition used by the authors Huang and Efhtimiadis (2009) which defines as “modification to a search query that addresses the same information need”. In our case, the need for information is the answer to the question, so the reformulation occurs when there is a modification of the previous search query, but on the same topic. In this sense, we think that the set of queries introduced per question aims to satisfy the same information need, that is, to find the answer to the question. The first query for each question is not a reformulation because it is about a new topic.

This global proposal forms a classification scheme for events that occur during web information-seeking.

3.4. Application of the Classification Scheme

To validate the usefulness of this classification scheme, we applied it to the previously mentioned experiment. This allowed us to validate the events, measures, and use guidelines of the classification scheme.

Each participant of the user study is associated with a video recording, a query log that includes the formulated queries by chronological order, and his health literacy (low or high). To assess the type terminology of the queries we counted with the collaboration of a health professional. Regarding the types of terminology, the medico-scientific and lay types were considered. We consider that a query is medico-scientific if it includes at least one medico-scientific term. All other queries are considered lay.

3.5. Statistical Analysis Strategy

Data analysis was performed using the SPSS software. We have computed descriptive statistics and applied hypothesis tests to analyze the significance of the observed differences.

Overall, for each event, we have compared the low and high health literacy groups in terms of the number of occurrences and, when possible, to compute, their duration measured in seconds. These are considered our first level measures. Based on first level measures, we also calculated second level measures. This analysis makes it possible to compare the selected values between two health literacy groups and various measures of the types of interaction through the average and number of specific cases, the median.

After obtaining descriptive statistics, we intend to analyze the data distribution and apply the parametric or non-parametric hypothesis tests. Given that the assessment of the assumption of normality of the data must always be considered when using parametric tests (Ghasemi and Zahediasl 2012). Normality tests were performed to assess the distribution of

data. In both tests, the confidence level applied was 95%. Thus, the distribution was normal if the sigma value was greater than 0.05. Regarding the normality tests, the Shapiro-Wilk test was chosen because is based on the correlation between the data and the corresponding normal scores and provides a better power (Ghasemi and Zahediasl 2012). Also, non-normality is less likely to be detected, but the Shapiro-Wilk test should be preferred as it is generally more sensitive (Samuels, Marshall, and Voake-Jones 2017). In this sense, the sigma value of the Shapiro-Wilk test was considered to be advisable in samples smaller than 50.

The sigma value if it was greater than 0.05 means that the distribution is normal and, therefore, parametric tests of independent samples of T-Test must be performed. When the distribution is not said to be normal (sig. <0.05), non-parametric tests of independent samples of the Mann-Whitney U test are performed for 2 samples.

The analysis will always aim to compare the values between the two health literacy groups (low and high health literacy). Despite being a small-scale study, hypothesis tests will be applied to verify whether the differences between the two groups are significant. Based on previous studies, we can also verify whether our results corroborate these and, if not, identify the possible causes of these variations.

4. The Classification Scheme

In this section, we will describe our proposal for the classification. As seen in Figure 12, the classification scheme consists of three main categories: Browser, Search Engine, and Web Pages, each corresponding to the area where the event takes place. The Search Engine Home Page (SEHP) and Search Engine Results Page (SERP) categories are part of the search engine and encompass the Search Engine Box Query (SEBQ) category. The Web Pages category has two categories: Result Page (RP) which refers to pages accessed through search engines, and Direct Access Page (DAP) which are pages that are accessed without the use of a search engine.

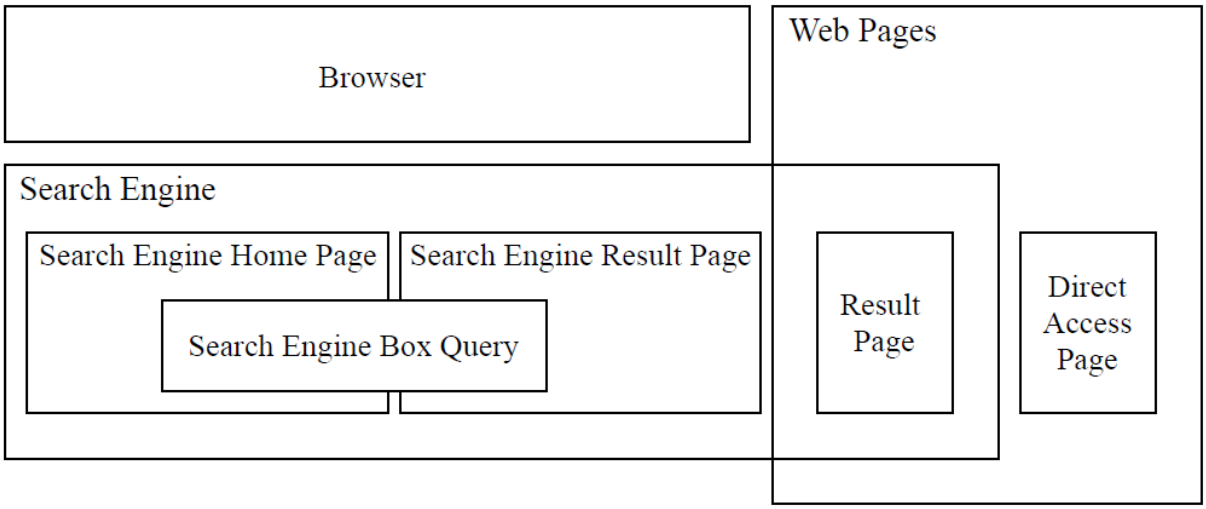


Figure 12 - The Classification Scheme

Each category has specific events as described next. Some of the events are related to query formulation. In these cases, it is possible to characterize the query based on the proposed set of descriptors (Table 11) and, if it is a query reformulation, describe the reformulation based on the descriptors in Table 12.

This section consists of two parts. The first part presents events associated with each category and the query reformulation modules. The second part deals with the analysis measures.

4.1. Event categories and Modules

4.1.1. Browser Events

This category includes general browser events that are commonly used during search sessions and events linked to query formulation in the address bar. All the included events are shown in Table 6.

Event	Description
OpenBrowser	Open the browser
CloseBrowser	Close the browser
OpenNewTab	Open a new tab in the browser
InteractAddressBarQuery	Interact with the address bar without submitting a query
NewAddressBarQuery[Q]	Enter a new query in the address bar
IncrementalSearch[Q]	Select an option from the "Incremental Search" list
RepeatQuery[Q]	Repeat the previous query
SelectTab	Change tab by selecting another one
CloseTab	Close a tab in the browser
ReopenClosedTab	Reopen a tab that was previously closed
ZoomIn	Zoom In on a page
ZoomOut	Zoom Out on a page
TypeURL	Type an URL in the address bar
Backward	Click on the back button
Forward	Click on the forward button
RefreshPage	Refresh the page
CopyText	Copy text that was selected
PasteText	Past text that was copied
SelectTextURL	Select text in the URL
RightClick	Right-click (for extra options)
DoubleClick	Double click on an item
FindBar	Use the Find Bar tool
CloseFindBar	Close the Find Bar tool
NextFindBarItem	Select the next item in the Find Bar list
Home	Click on the Home button
Settings	Access the Settings

[Q]Event related to query formulation.

Table 6 - Browser Events

4.1.2. Search Engine Events

Interaction with the search engine includes interactions with the home page and the results page. Therefore, this section includes the categories Search Engine Home Page (SEHP) and Search Engine Results Page (SERP). The SEHP category comprises two events in which the last one presented is also possible when interacting with SERP (Table 7).

Event Name	Description
ImFeelingLucky	Click on the button "I'm Feeling Lucky"
PrivacyReminder	Interact with the "Privacy Reminder" toolbox

Table 7 - Search Engine Home Page Events

About SERP interactions, Table 8 shows the considered events. In the proposal of these events, we considered the categories mentioned in the Structure of Search Engine Result Pages section: organic results, paid results, enriched snippets, universal results, knowledge data. We also consider the interaction with the SERP tools (e.g., translator).

The event “Interact[Set 2 result]” includes interactions that do not imply leaving the SERP to access a result page. For example, in People Also Ask, a universal result, it is possible to click on the multiple queries for more information without leaving the SERP. Also, in a carousel of images, it is possible to scroll through the various images without clicking on one. This event can occur with all types of results except the organic one, which can only be clicked.

Event Name	Description
ShowingResultsFor^{Q}	Click on the "Showing results for" link
DidYouMean^{Q}	Click on the "Did you mean" link
LeftClick[Set1 result]	Left-click on a result of set1 and is redirected from the SERP to the link
MiddleClick[Set1 result]	Click on a result of set1, and a new tab opens in the browser
RightClick[Set1 result]	Right-click on a result of set1
Interaction[Set2 result]	Interact with a result or tools of set 2
Page2SERP	Go to the 2nd SERP page of a search
OpenLinkNewTab	Click on the open link in the new tab option after right-clicking on a link
AnotherSERP	Go to another SERP
AllSERP	Click on the All tab
ImagesSERP	Click on the Images tab
VideosSERP	Click on the Videos tab
BooksSERP	Click on the Books tab
NewsSERP	Click on the News tab
ToolsSERP	Click on the Tools tab
AdvancedSearch	Click on the Advanced Search option
RelatedSearches	Click on Related Searches option
SelectTextSERP	Select text
PrivacyReminder	Interact with the "Privacy Reminder" toolbox
ScrollDownStart (SDS)	Start scrolling down

Event Name	Description
ScrollDownFinish (SDF)	Finish scrolling down
ScrollUpStart (SUS)	Start scrolling up
ScrollUpFinish (SUF)	Finish scrolling up

[Q]Event related to query formulation. [Set1 result] should be replaced by an element of: {OrganicResult, PaidResult, UniversalResult, EnrichedSnippet, KnowledgeData}. [Set2 result] should be replaced by an element of {PaidResult, UniversalResult, EnrichedSnippet, KnowledgeData, Tools}.

Table 8 - Search Engine Result Page Events

4.1.3. Web Pages Events

We considered two categories of web page interactions: with result pages and with direct access pages. The category DAP was added for cases in which those users access a direct page but not from a SERP click. Both categories share the same events. The existence of two categories allows the identification of the place where the event occurred. Table 9 shows all the considered events.

Event	Description
Select2ndLevelDepthLink	Click on a hyperlink inside a web page
SelectText	Select text on the web page
SelectAnchorLink	Click on the link that redirects him/her to the same web page
SelectImage	Select an image of the web page
CloseAdvert	Close an advert
QueryInternalSearchEngine	Type a query in the internal search engine box of a web page
OpenLinkNewTab	Click on the open link in a new tab option after right-clicking on a link on the web page
ScrollDownStart (SDS)	Start scrolling down a web page
ScrollDownFinish (SDF)	Finish scrolling down a web page
ScrollUpStart (SUS)	Start scrolling up a web page
ScrollUpFinish (SUF)	Finish scrolling up a web page
InputData	Enter data on an input box a web form

Table 9 - Result Page and Direct Access Page Events

4.1.4. Search Engine Box Query Events

The search engine box query category consists of eight events possible to be performed in this box (Table 10).

Event Name	Description
InteractSearchEngineBoxQuery	Interact with the search box without submitting a query
NewSearchEngineBoxQuery[Q]	Enter a new query in the search box
PartialChange[Q]	Modify the previous query for the same question
RepeatQuery[Q]	Repeat the previous query
SelectTextQuery	Select text previously written in the query
SearchByVoice[Q]	Search using "Search by Voice"
IncrementalSearch[Q]	Select an option from the "Incremental Search" list

[Q]Event related to query formulation.

Table 10 - Search Engine Box Query Events

4.1.5. Query (Re)Formulation Modules

After creating the classification scheme, we consider that it could be useful to analyze the set of queries submitted by users, both in terms of their characterization and the types of reformulation carried out. The use of these modules is optional, so they can be applied together or not with the classification scheme.

Some events of the classification scheme are related to query formulation. In these cases, it is possible to characterize the query based on the proposed set of descriptors (Table 11) and, if it is a query reformulation, describe the reformulation based on the descriptors in Table 12. We consider that every query except the first of a search session is a reformulation.

Table 11 presents the characteristics of the queries and Table 12 shows the types of reformulation considered. Qi represents the initial query and Qi + 1 the reformulation of the previous query.

Name	Description
Length	Number of terms per query
Language	Language of the query
Terminology	Lay or Technical
Typographical Errors	Check if the query has typographical errors, that is, errors are caused by mechanical or finger lapses
Orthographic Errors	Check if the query has orthographic errors, that is, errors in the correct form of writing

Table 11 - Characteristics of the queries

Reformulation	Description	Example
TypographicalError Correction	Qi contains at least one term in common with Qi + 1. Q has a typographical error and, in Qi + 1, this has been corrected.	asthma7 → asthma
OrthographicError Correction	Qi contains at least one term in common with Qi + 1 but in Qi, it has a spelling mistake and in Qi + 1, this error has been corrected.	simptoms → symptoms
TotalNew	Qi and Qi + 1 belong to the same information need but Qi + 1 has no terms in common with Qi. The only types of reformulations that can simultaneously occur are the terminology and language changes.	bronchodilator → allergen
LaytoTech or TechtoLay	Qi is a query considered lay and Qi + 1 is a technical query or vice versa.	fiber insertion → fiber intake
Lang1toLang2 or Lang2toLang1	Qi is a query in language 1 and Qi + 1 is a query in language 2 and vice versa. This reformulation implies only the removal and addition of terms common to the previous query. Lang1 and Lang2 must be replaced by the official language acronym according to ISO 3166-1 in the two-letter country code set (alpha-2).	asma → asthma
RepeatQuery	Qi and Qi + 1 must contain the same terms in the same order, without adding or removing words.	heart disease → heart disease
AddWords	Qi is a subset of Qi + 1, that is, all terms present in Qi are in Qi + 1. Qi + 1 contains more terms than Qi.	cancer fiber → cancer eat more fiber
RemoveWords	Qi is a super subset of Qi+1, that is, all terms present in Qi are in Qi+1. Qi+1 contains fewer terms than Qi.	allergen asthma → allergen
SubstitutionWords	Qi and Qi + 1 must contain at least one term in common. Qi + 1 must be the same length as Qi, but it can contain terms that were not in Qi. This substitution must correspond to the same language.	vitamin b → vitamin d
ExpandTotalAcronym	Qi contains an acronym and Qi + 1 expanded all letters composed by that acronym.	PEFR → peak expiratory flow rate
ExpandPartialAcronym	Qi contains an acronym and Qi + 1 has expanded at least 2 letters of that acronym.	PEFR → peak expiratory
[Introduce, Remove] BooleanOperators	Qi + 1 contains at least one Boolean operator (AND, OR, NOT) concerning Qi or vice versa.	PEFR → PEFR and sinusitis
SynonymousSubstitution	Qi + 1 may contain terms in common with Qi but at least one term is replaced by a synonym or something similar. This substitution by synonym must correspond to the same language.	tiredness occurs in menopause → tiredness happens in menopause
Stemming	Qi + 1 has a term in common with Qi but Qi + 1 the word has been reduced to its root. This reformulation must correspond to the output of a stemming algorithm. Otherwise, it is just the removal and addition of words or characters and not stemming.	sinusitis → sinus
VerbalFlexion	Qi has a verb in which Qi + 1 suffered an inflection of the verb.	foods reduce cancer → fiber reduces cancer
[Introduce, Remove] Accentuation	Qi has a term in common with Qi + 1 but in Qi + 1 the accentuation rules were applied to that term or vice versa	frisee lettuce → frisée lettuce
AdjectiveforName or NameforAdjective	Qi contains a term in common with Qi + 1 but in Qi, it is an adjective and in Qi + 1 it is a name or vice versa.	asthmatic child cares → asthma

Reformulation	Description	Example
VerbforName or NameforVerb	Qi contains a term in common with Qi + 1 but in Qi, it is a verb and in Qi + 1 it is a name or vice versa.	fiber reduces cancer → fiber cancer reduction
VerbforAdjective or AdjectiveforVerb	Qi contains a common term with Qi + 1 but in Qi, it is a verb and in Qi + 1 it is an adjective or vice versa.	protect skin → skin protective ways
PluraltoSingular or SingulartoPlural	Qi has at least one term in common with Qi + 1 where Qi that term is written in the plural and Qi + 1 is written in the singular or vice versa. This reformulation can occur from one language to another.	antioxidants → antioxidant

Table 12 - Types of Query Reformulation

4.2. Analysis Measures

Concerning the analysis, we consider two types of measures: the first level measures and the second level measures.

The first level measures are the measures related to the events, in terms of their occurrence and duration. For each event previously presented in the tables, the number of occurrences (if any) was calculated and when it was useful, also its duration in seconds. The form of analysis will be different depending on the type of interaction. It should be performed by SERP or RP whenever we are in the presence of an event of these categories. In events of other categories, the aggregation is done by session. In either case, we always end up aggregating per user. Some examples of these measures are the number of middle clicks per SERP, the number of open new tabs per session, and the number of close adverts per RP.

Based on first level measures, we computed second level measures. The second level measures are those that arise from the first through the aggregation of occurrences and durations of events. We considered the number and duration by category and in total (Table 13), scrolling measures for SERP, RP, and DAP (Table 14), clicks, pages, time to first RP click measures (Table 15), and queries-related measures (aggregation of the address bar with search engine box events shown in Table 16).

The general measures of the study are shown in Table 13 for each group.

Measure – Total #Events and Duration (in seconds)	
Events' Group	**Aggregation**
Browser	Session
SEHP	Session
SEBQ	Session
DAP	Session

Measure – Total #Events and Duration (in seconds)	
Events' Group	**Aggregation**
SERP	Session and SERP
RP	Session and RP
Total	Session

Table 13 - General Measures by Event's Group

Subsequently, Table 14 shows scrolling measures in the SERP, RP, and DAP groups.

Measure – #Events and Duration (in seconds) of Scrolling Up (sus+suf), Scrolling Down (sds+sdf), and Scrolling (up+down)	
Events' Group	**Aggregation**
DAP	Session
SERP	SERP
RP	RP

Table 14 - Scrolling Measures for SERP, RP, and DAP* should be included only right clicks with open new tab after.

Table 15 presents the measures related to clicks and pages viewed in the SERP, RP, and DAP groups.

Event's group	**Measure**	**Aggregation**
SERP	#SERPviewed	Session
	#ResultClick1	Session and SERP
	#SERPwith/without clicks	SERP
	#Timeto1stRPClicksinceSERPLoad	
	#RankClick	
	#Organic, Universal, Paid results, Enriched Snippet, Knowledge Data Clicks	
RP	#RPviewed	Session
	#RPnotviewed	Session and RP
DAP	#DAPviewed	Session

* should be included only right clicks with open new tab after.

Table 15 - Clicks and Pages viewed Measures for SERP, RP, and DAP

Finally, Table 16 presents measures based on the aggregation of the address bar with search engine box events ("Interact", "New", "RepeatQuery" and "IncrementalSearch"). We designate query events measures.

Measure	Aggregation
#TotalInteracts	Session
#TotalNewQueries	
#TotalIncrementalSearch	
#TotalRepeatQuery	
#TotalQueryEvents	
#TimeQueryEvents	

Table 16 - Query Events Measures

Regarding the Query (Re)formulation Modules, the number of queries and reformulations could be analyzed per session. The measures related to the characteristics of the queries could be analyzed per query (Table 17).

Measures	Analysis
#Queries	Session
#Reformulations	
#Terms	Query
#OrthographicErrors	
#TypographicalErrors	
#LayQueries	
#MedSciQueries	
#PTQueries	
#ENQueries	

Table 17 - Query Analysis Measures

In reformulation types, all measures should be analyzed by reformulation.

5. Analysis of information-seeking behavior by the level of health literacy

This section consists of three parts: the first concerns the application of the classification scheme to the user experiment mentioned before, the second shows the results obtained from data analysis by the level of health literacy and in the third part, the results obtained are discussed.

5.1. Classification Scheme Application

Since our objective was to analyze the information-seeking behavior of users with different levels of health literacy, after proposing a classification scheme related to web search interactions, we have used it to log the interactions that occurred in the video recordings from the user study.

Each participant of the user study is associated with a video recording, a query log that includes the formulated queries by chronological order, and his health literacy (low or high).

In the classification scheme, each sheet needed the annotation of individual records, that is, per user, and a sheet in general with the annotation of all users. For each user, we created a file in which we need to log information about the user, task, event, SERP, RP (Table 18).

About	Fields
User	User ID
Task	Task ID
Event	Timestamp, category ID, event ID
SERP	SERP ID[1], click rank
RP	RP ID[1], RP view

[1]Automatically computed based on the event id.

Table 18 - Logged information in the Classification Scheme

To identify users and tasks, we have used sequentially assigned numbers. The event identifiers are the ones proposed in the classification scheme.

In SERP events, we also automatically log the identifier of the SERP. This identifier allows us to aggregate events based on individual SERP. If the previous event is related to query formulation, a new SERP ID is assigned to the event being logged. If not, the SERP ID is the same as before, that is, it does not increase. Note that, with this assignment algorithm, two

visualizations of a SERP with the same URL generate different SERP ID, what, for this work, is not a problem. If the SERP event is a click (left-click, middle-click, or right-click with open new tab) on a enriched snippet, organic result, or universal result, we also log the clicked result's rank position. Note that the computation of rank position ignores paid results and some universal results, such as results pointing to Google Scholar articles.

In RP events, we register the page's identifier for the same reason we do it in SERP events. This identifier is also assigned automatically, in sequential order, whenever the previous event was simultaneously a SERP event and a click one (any of the three types mentioned above). For result pages, we also recorded if the page was viewed or not. We decided to do so because some of the clicked results ended up not being viewed. There were situations in which users opened the results in different tabs and never viewed those tabs or clicked on a result but went back to the SERP before the result page loaded.

The above information was later enriched with two additional elements, the duration of each event and time to the first SERP click. Both elements were computed automatically. The first is simply the time difference, in seconds, between the current event and the following one. The time to first SERP click is the time difference, in seconds, between the query formulation event that generates a new SERP and the first click event (left-click, middle-click or open a new tab after a right-click).

For the query (re)formulation analysis, we logged information in a different file. We logged information about user, task, and query (Table 19).

About	**Fields**
User	User ID
Task	Task ID
Query	Query ID, writing the original query, number of terms, language, terminology, errors, and types of reformulation

Table 19 - Logged information in Modules

About the user and task, the annotation form was similar to the one described in Table 1. In addition to the query itself, we registered its ID, sequentially assigned, its number of terms, language (English or Portuguese), and terminology (lay or medico-scientific). For this work, a query is medico-scientific if it includes at least one medico-scientific term. All other queries are considered lay. To assess the terminology of the queries we counted on the collaboration of a health professional who helped us by giving her opinion on the type of terminology associated with each query.

5.2. Results

This section presents the results of our study by type of interaction.

Regarding health literacy groups, 11 users are part of the high health literacy (HHL) group and 8 users belong to the low health literacy (LHL) group. In terms of the number of sessions, 10 sessions per user were evaluated, except for two users in the HHL group. In these two users, recording began after the start of the experiment. In one case, only 7 sessions were considered, and in the other case, 9 sessions. Since they were the ones that we had access to through the videos. We decided to include these experiences and consider 10 sessions instead of 2 since it would not be possible to divide between part of asthma and part of nutrition. In the analysis, the form of calculation was always considering the number of specific sessions in these two cases.

Comparisons are always between the two groups. Therefore, when one group is mentioned, the comparison with the other group is implicit. In the averages, we bolded the highest value by measure. In the application of parametric and non-parametric tests, the unilateral values will be shown. Regarding decimal places, we chose to use only two. When reporting results, we use * to indicate results significant at p=0.1, ** at p=0.05, and *** at p=0.01. The exploratory nature of this work justifies the consideration of the 0.1 p-level.

There were no search engine interactions and, therefore, there will not be a section for this group of interactions.

As shown in Table 20, the LHL group recorded a greater number of events and interactions' time at a general level. In RP interaction, the duration was longer in this group. More detailed information is present in Appendix 5.

Category	**Measure**	**Average per Session**		**p-value**
		HHL	**LHL**	
Browser	#Events	**7.17**	6.14	0.27
	Duration	**25.29**	23.80	0.40
Search Engine Result Page	#Events	**6.68**	5.98	0.33
	Duration	18.09	**20.91**	0.27
Result Page	#Events	8.96	**10.71**	0.14
	Duration	32.84	**45.36**	0.06*
Search Engine Box Query	#Events	3.49	**3.94**	0.25
	Duration	15.73	**21.48**	0.16
Direct Access Page	#Events	0.02	**0.08**	0.27

Category	Measure	Average per Session		p-value
		HHL	LHL	
	Duration	0.06	**0.64**	0.27
Total	#Events	26.16	**26.84**	0.36
	Duration	91.98	**110.56**	0.16

Table 20 - Global Aggregated Results

5.2.1. Browser Interaction

In the browser interaction, the HHL group closed tabs and selected text from the URL more often. They also introduced more new queries in the address bar. In contrast, the LHL group had more backward events (Table 21). Regarding the results obtained, six measures did not have any type of occurrence (Appendix 6).

Measure	Average per Session		p-value
	HHL	LHL	
#OpenNewTab	**0.25**	0.25	0.36
#SelectTab	**2.02**	1.40	0.22
#CloseTab	**1.27**	0.56	0.08*
#ReopenClosedTab	0.00	**0.03**	0.33
#InteractAddressBarQuery	0.02	**0.04**	0.39
#NewAddressBarQuery	**0.92**	0.58	0.10*
#IncrementalSearchAddressBarQuery	**0.11**	0.04	0.37
#RepeatQueryAddressBarQuery	**0.02**	0.01	0.42
#ZoomIn	0.00	**0.01**	0.45
#TypeURL	0.02	**0.06**	0.36
#Backward	0.54	**1.30**	0.05**
#Forward	0.00	**0.01**	0.33
#RefreshPage	**0.03**	0.00	0.39
#KeyboardShortcut	**0.05**	0.03	0.33
#CopyText	**0.06**	0.05	0.25
#PasteText	**0.07**	0.04	0.22
#SelectTextURL	**0.71**	0.36	0.03**
#RightClick	0.02	**0.03**	0.45
#DoubleClick	**0.01**	0.00	0.39
#FindBar	**0.63**	0.63	0.36

Measure	Average per Session		p-value
	HHL	LHL	
#CloseFindBar	**0.03**	0.03	0.50
#NextFindBarItem	0.38	**0.70**	0.27

Table 21 - Browser Interaction Results

5.2.2. SERP Interaction

In the interaction with the SERP, there are 54 measures, but not all had occurrences (Appendix 7). SERP measures were aggregated by the SERP ID. Besides, we also analyzed the number of SERP and the number of clicks on results by session.

The LHL group made more left-click in general and in organic results. They also had more SERP with clicks and took longer to first RP click. The HHL group had more result clicks. They made more left clicks in universal results, middle clicks in general, and in organic results. They had more SERP without clicks (Table 22).

Measure	Average per SERP		p-value
	HHL	LHL	
#ShowingResultsFor	0.03	**0.04**	0.48
#DidYouMean	**0.01**	0.00	0.39
#LeftClickOrganicResult	0.40	**0.75**	0.03**
#LeftClickUniversalResult	**0.03**	0.01	0.05**
#RightClickOrganicResult	0.07	**0.15**	0.30
#RightClickUniversalResult	0.01	**0.02**	0.45
#MiddleClickOrganicResult	**0.46**	0.11	0.05**
#MiddleClickUniversalResult	**0.01**	0.00	0.27
#InteractionUniversalResult	**0.01**	0.00	0.18
#ToolsInteraction	**0.00**	0.00	0.39
#Page2SERP	0.00	**0.00**	0.33
#OpenLinkNewTab	0.07	**0.17**	0.30
#AnotherSERP	**0.01**	0.00	0.39
#AllSERP	**0.00**	0.00	0.39
#ImagesSERP	0.00	**0.00**	0.45
#ToolsSERP	0.00	**0.04**	0.33
#SelectTextSERP	**0.04**	0.01	0.20
#sds	**0.58**	0.57	0.42

Measure	Average per SERP		p-value
	HHL	**LHL**	
#sdf	**0.58**	0.57	0.42
#sus	0.20	**0.22**	0.36
#suf	0.20	**0.22**	0.36
#EventsTypeSERPinteraction	2.70	**2.90**	0.40
SecondsSERPInteraction	7.26	**9.68**	0.12
#ScrollingUP	0.40	**0.45**	0.36
#ScrollingDown	**1.16**	1.15	0.42
#Scrolling	1.56	**1.60**	0.33
SecondsScrollingUP	0.67	**1.20**	0.39
SecondsScrollingDown	**2.99**	2.73	0.45
SecondsScrolling	3.66	**3.93**	0.45
#SERP[1]	2.55	**2.36**	0.29
#SERPSwithClicks	0.67	**0.76**	0.08*
#SERPSwithoutClicks	**0.33**	0.24	0.08*
Timeto1stRPClicksinceSERPLoad	5.80	**8.30**	0.05**
MedRankClicked[2]	1.50	1.50	0.42
#ResultClick	0.96	**1.03**	0.28
#ResultClick[1]	**2.39**	2.34	0.44
#LeftClicks	0.42	**0.75**	0.04**
#MiddleClicks	**0.47**	0.11	0.03**
#RighClicks (with OpenNewTab)	0.07	**0.17**	0.30
#ClicksOrganicResults	0.92	**1.01**	0.24
#ClicksUniversalResults	**0.04**	0.02	0.33

[1]The value in this line is the average per session. [2]The value in this line is the median per SERP.

Table 22 - SERP Interaction Results

5.2.3. RP Interaction

In the interaction with the results pages, a total of 22 measures were analyzed (Appendix 8). The LHL group had more scrolling down events and scrolling up. They also had more scrolling occurrences and spent more time in this interaction. The aggregate measures of scrolling up, scrolling down, and scrolling in general events also showed more occurrences in the LHL group. The LHL group spent more seconds doing scrolling up, scrolling down, and scrolling in general (Table 23).

Measure	Average per RP viewed		p-value
	HHL	**LHL**	
#Select2ndLevelDepthLink	0.07	**0.12**	0.12
#SelectTextRP	**0.30**	0.18	0.45
#SelectAnchorLink	0.00	**0.01**	0.33
#CloseAdvert	0.07	**0.10**	0.25
#QueryInternalSearchEngine	0.00	**0.00**	0.33
#OpenLinkNewTab	0.00	**0.01**	0.33
#sds	1.57	**1.93**	0.05**
#sdf	1.57	**1.93**	0.05**
#sus	0.19	**0.37**	0.04**
#suf	0.19	**0.37**	0.04**
#InputData	0.00	**0.00**	0.33
#EventsTypeRPinteraction	3.98	**5.02**	0.05**
SecondsTypeRPInteraction	14.27	**21.12**	0.01***
#ScrollingUP	0.38	**0.73**	0.04**
#ScrollingDown	3.15	**3.86**	0.05**
#Scrolling	3.53	**4.59**	0.02**
SecondsScrollingUP	0.92	**2.45**	0.02**
SecondsScrollingDown	11.88	**16.69**	0.06*
SecondsScrolling	12.80	**19.14**	0.02**
#RP[1]	**2.27**	2.21	0.43
#RPnotViewed[1]	0.10	**0.14**	0.33
#RPnotViewed	0.04	**0.06**	0.45

[1]The value in this line is the average per session.

Table 23 - RP Interaction Results

5.2.4. DAP Interaction

In this type of interaction, we chose to analyze per session given the data set (Appendix 9). The low number of DAP in the search sessions led us to analyze these measures by session instead of analyzing them by page, as we have done in SERP (Table 24).

There were 3 cases of users who visited at least one page of direct access without going to the search engines before. we can see that the LHL group made more use of these pages and that they entered data. In these cases, the two users went directly to the google translator website. The HHL group visited only one page and interacted through the scrolling measures.

In this case, the user in question directly accessed a Wikipedia page. None of these measures showed significant differences.

Measure	Average per Session		p-value
	HHL	**LHL**	
#InputData	0.00	**0.08**	0.20
#Sus	**0.01**	0.00	0.39
#Suf	**0.01**	0.00	0.39
#ScrollingUP	**0.02**	0.00	0.39
SecondsScrollingUP	**0.06**	0.00	0.39
#DAP	0.01	**0.03**	0.30

Table 24 - DAP Interaction Results

5.2.5. Query (Re)Formulation Behavior

Results related to query formulation interactions in the SEBQ are presented in Table 25. More information can be seen in Appendix 10.

The LHL group demonstrated greater interaction with the search engine box query.

Measure	Average per Session		p-value
	HHL	**LHL**	
#InteractSearchEngineBoxQuery	0.20	**0.39**	0.06*
#NewSearchEngineBoxQuery	0.79	**1.08**	0.11
#PartialChange	**0.84**	0.71	0.22
#RepeatQuery	0.00	**0.03**	0.20
#SelectTextQuery	0.97	**1.19**	0.15
#IncrementalSearch	**0.69**	0.54	0.26
#SearchEngineBoxQueries	1.63	**1.79**	0.30

Table 25 - SEBQ Results

We analyze the measures related to the queries entered in the search engine box and the browser address bar (Appendix 11). Table 26 shows the results in both the SEBQ and browser address. As shown, the LHL group demonstrated a greater number of query formulation interactions.

Measure	Average per Session		p-value
	HHL	**LHL**	
#TotalInteracts	0.22	**0.43**	0.06*
#TotalNewQueries	**1.71**	1.65	0.41
#TotalIncrementalSearch	**0.80**	0.58	0.16
#TotalRepeatQuery	0.02	**0.05**	0.25
#TotalQueryEvents	4.43	**4.55**	0.43
#TimeQueryEvents	22.62	**26.33**	0.15

Table 26 - Query events results

Concerning the query reformulation modules, three types of measures were analyzed: general measures per session, the characteristics of the queries in the total of queries, and the types of reformulations in the total of reformulations (Appendix 12).

Regarding the measures per session, the HHL group formulated more queries and reformulations (Table 27).

Measure	Average per Session		p-value
	HHL	**LHL**	
#Queries	**2.69**	2.51	0.28
#Reformulations	**1.69**	1.51	0.28

Table 27 - Query (re)formulation per session

In the general characteristics of the queries, the HHL group had more typographical errors (Table 28). The queries in which it was not possible to identify the language because it contained half of the terms in English and half in Portuguese were designated "NALanguageQueries".

Measure	Average per Query		p-value
	HHL	**LHL**	
#TermsperQuery	3.20	**3.43**	0.27
#OrthographicErrors	**0.072**	0.069	0.46
#TypographicalErrors	**0.02**	0.00	0.10*
#LayQueries	**0.42**	0.41	0.46
#MedSciQueries	0.58	**0.59**	0.46
#PTQueries	0.61	**0.76**	0.33
#ENQueries	**0.38**	0.23	0.39
#NALanguageQueries	**0.011**	0.008	0.42
#Reformulations	**0.61**	0.55	0.17

Table 28 - Query characteristics

About reformulations, Table 29 presents the results obtained. More detailed information on Appendix 12.

The HHL group does more word addition, substitutes terms by their synonym, and changes plural to singular in reformulations more often. The LHL group used more totally new reformulations (Table 29).

Measure	Average per Reformulation		p-value
	HHL	**LHL**	
#OrthographicErrorCorrection	0.05	**0.09**	0.20
#TypographicalErrorCorrection	**0.02**	0.00	0.39
#TotalNew	0.07	**0.15**	0.03**
#LaytoMedSci	0.09	**0.13**	0.12
#MedScitoLay	**0.13**	0.10	0.27
#PTtoEN	0.05	**0.06**	0.42
#ENtoPT	0.03	**0.04**	0.33
#RepeatQuery	0.01	**0.07**	0.25
#SubstitutionWords	**0.13**	0.10	0.28
#SynonymousSubstitution	**0.05**	0.01	0.08*
#AddWords	**0.51**	0.39	0.06*
#RemoveWords	**0.37**	0.30	0.15
#ExpandTotalAcronym	0.03	**0.06**	0.36
#ExpandPartialAcronym	**0.03**	0.00	0.27
#FormAcronym	0.00	**0.01**	0.33

Measure	Average per Reformulation		p-value
	HHL	**LHL**	
#IntroduceBooleanOperators	0.00	**0.01**	0.33
#RemoveBooleanOperators	0.00	**0.01**	0.33
#IntroduceAccentuation	**0.01**	0.00	0.39
#Stemming	**0.00**	0.00	0.39
#VerbalFlexion	0.00	**0.01**	0.33
#AdjectiveforName	**0.02**	0.01	0.33
#NameforAdjective	**0.00**	0.00	0.45
#VerbforName	0.01	**0.01**	0.48
#PluraltoSingular	**0.06**	0.04	0.10*
#SingulartoPlural	**0.03**	0.02	0.30

Table 29 - Query reformulation types

5.3. Discussion of Results

In this discussion of results, importance will be given to the measures that revealed statistically significant differences. In this way, these measures are summarized (Table 30) and therefore, possible explanations for these differences will be given, maintaining the comparison between the two health literacy groups. Based on previous studies, we will verify if the results obtained in this study corroborate these, and if not, the possible causes of these variations are identified.

Category	Measure	Group
General	RPDuration	LHL>HHL
Browser	#CloseTab	HHL>LHL
	#NewAddressBarQuery	
	#SelectTextURL	
	#Backward	LHL>HHL
SERP	#LeftClickUniversalResult	HHL>LHL
	#MiddleClickOrganicResult	
	#MiddleClicks	
	#SERPSwithoutClicks	
	#LeftClickOrganicResult	LHL>HHL
	#SERPSwithClicks	
	Timeto1stRPClicksinceSERPLoad	
	#LeftClicks	
RP	#sds	LHL>HHL
	#sdf	
	#sus	

Category	Measure	Group
	#suf	
	#EventsTypeRPinteraction	
	SecondsTypeRPInteraction	
	#ScrollingUP	
	#ScrollingDown	
	#Scrolling	
	SecondsScrollingUP	
	SecondsScrollingDown	
	SecondsScrolling	
SEBQ	#InteractSearchEngineBoxQuery	LHL>HHL
Query Behavior	#TotalnteractS	LHL>HHL
	#TypographicalErrors	HHL>LHL
	#SynonymousSubstitution	
	#AddWords	
	#PluraltoSingular	

Table 30 - Significant Differences

In general, the LHL group spent more time interacting with RP, which is in line with previous studies (Yom-Tov et al. 2016; Duggan and Payne 2008; Diane Kelly and Cool 2002).

About browser interaction, the number of occurrences of the backward button was greater in the LHL group. During the search tasks, it was visible that some users of the LHL group tend to have only one page open and do not open the results in new tabs. Having only one tab open, normally, they often use the back button to return to the SERP. On other hand, the HHL group selected more often text from the URL and closed more tabs. They also made more use of the address bar and introduced more new queries there. Probably, the LHL user does not understand that it is possible to introduce a query in the address as in the search engine box. That may be the reason why they used the search engine box query more often. There is a strong association between health literacy, internet access, and use as it is already known. Individuals with adequate health literacy levels are more likely to access the internet and use it to look for health information (Vida Estacio, Whittle, and Protheroe 2019). Probably the LHL group uses the internet less as a preferred means of accessing information.

In SERP, The LHL group did more left clicks in general and in organic results. This explanation ends up confirming the assumption that was mentioned earlier. When the user left-clicks on an organic result, is redirected from the SERP page to the link. As said before, this group tends to have only one page open, not opening results in new tabs and, therefore, making more left clicks. The HHL group made more middle clicks in general. When the user right-clicks on an organic result, a new tab opens in the browser. In this sense, it is normal for

this group to have several tabs open simultaneously. They also tend to have a higher number of SERP without clicks, maybe because they do more reformulations or write the answer to the question based on the SERP page. Our tendency also showed that the LHL group takes longer to click on the first result.

Regarding RP, the LHL group interacted more with the results pages, had a higher occurrence of events, and longer duration. They had more occurrences and duration of scrolling up, scrolling down, and general scrolling. This group is usually associated with a longer dwell time on web pages because they take more time to read and focus on the page (Yom-Tov et al. 2016). These results are interesting because although the LHL group interacted more at the level of events and seconds with the results pages, they visited fewer pages. It suggests that they view fewer pages, but they spend more time on each one than the HHL group.

In DAP interaction, we noticed that users tend to use the translator tool to seek the English translation of search terms.

In query behavior, the LHL group tends to interact more in both the address bar and SEBQ but mainly in the latter. Still, in formulation behavior, this group submit fewer queries but interacted more. About queries characteristics, both groups preferred queries in the Portuguese language. Queries in English were more frequent in the HHL group. This group had more typographical errors in queries. Perhaps these users type quicker, they made more mistakes in the search. This contradicts the results of Wildemuth (2004), who found that low domain knowledge is associated with a higher frequency of errors. Regarding query reformulation, the LHL group used more the total new reformulation type. This type of reformulation happens when the previous query has no common terms with the reformulated query. In the same search task, LHL users chose to completely reformulate the query with new terms instead of making a partial change. Perhaps the explanation is because they are not being successful in search and therefore opt to reformulate the query completely. A study report that novices use more strategies like making the word plural or singular, repeat search terms, reuse prior search terms, and retains the same basic structure (Hembrooke et al. 2005). The HHL group made more reformulation in which they add words, substitute terms by their synonym, and change plural to singular. Maybe the LHL did less because of the difficulties associated with the formulation search (Lopes and Ribeiro 2015). Regarding terminology, the HHL group had more queries of the lay type, but the most recurrent type of reformulation in this group was the change from a medico-scientific query to a lay one. Lopes and Ribeiro (2016) refer that a higher level of health literacy as preferred medical-scientific.

6. Conclusions

In this work, we analyze the web information-seeking behavior of users with different levels of health literacy. For this purpose, a classification scheme was presented to systematically analyze information-seeking behavior. The scheme consists of three main categories: Browser, Search Engine, and Web Pages. The Search Engine Home Page (SEHP) and Search Engine Results Page (SERP) categories are part of the search engine and encompass the Search Engine Box Query (SEBQ) category. The Web Pages category has two categories: Result Page (RP) and Direct Access Pages (DAP). Through the construction of this scheme, it was possible to analyze the behavior of users in the search for health information.

Regarding the results obtained between the two health literacy groups, it was found that the low health literacy group made more use of the search engine box, made more left clicks, used more backward functionality, and spent more time interacting with the results pages, namely scrolling. On the other hand, the high health literacy group made more use of the address bar, made more middle clicks, used the select text URL functionality more, and made more reformulations.

The objectives initially proposed were fulfilled since it was possible to investigate behavior at the level of interaction and reformulation of search. This study served as a first step in helping to identify ways to make the information retrieval processes on the web more efficient and personalized based on the behaviors evidenced by the two groups.

6.1. Limitations

In this study, it is possible to list some limitations that may have had an impact on the results obtained. First, the annotation of the videos was done manually. Samuel Serafim had already made the annotation, but new information was annotated, and the other was changed. As far as we are concerned, it was a very thorough job of analyzing all the events that users did in seconds and recording from the video viewing, it is necessary to be very specific and some events may have failed. To help in this process, the application of automatic analysis can be very useful.

Regarding the query terminology, we asked a health professional, to give their general opinion on queries, whether she considered a lay query or a scientific medical query. However, it was the general opinion regarding the query and not considering each term. It would have been interesting to analyze all the existing terms of the queries and perhaps make an average and thus determine the type of terminology. However, it was a job that took longer, and it would probably take more than a health professional to have more than one opinion. We could

see that this process is not as easy and objective as it seems. Future works can use resources like the Consumer Health Vocabulary in automatic processes, complemented, later on, by a health professional validation.

The classification scheme has undergone a lot of changes, whenever we analyzed them, we discovered specific cases (some doubtful), new events, new measures of analysis. It was not an easy task to portray all the changes because there were multiple redefinitions until the final version, and this also resulted in a repeated analysis of the data. However, I was always in contact with my supervisor and so I always had great advice. But we think it would be interesting to have had feedback from other researchers about this classification scheme. We are fully aware that this classification scheme could undergo further changes. Just as have suffered since we started working on this topic.

The classification scheme is generic, which means it can be applied to disregard the used search engines. It can be used in manual analysis studies and in (semi) automatic analysis studies in which part of the events can be determined automatically.

The nature of the classification scheme makes it dependable on the evolution of search engines' interfaces. Therefore, it is probable that, in the future, this classification scheme will need to be updated. Changes can also be triggered by the experience of other researchers using this classification scheme.

6.2. Future Work

Regarding future work, here are some suggestions for developing this theme.

In the first phase, this classification scheme could be enriched with events that occur in specific-content SERP such as images SERP. Furthermore, we think it would be interesting to discriminate against the type of results seen in SERP, for example, images, articles, videos, among others. Another aspect that would be interesting to include is the number of views of the same SERP or RP. Through viewing videos, it would not be feasible, and it was not easy to see unless it was followed. But it would be interesting to understand if low health literacy users are re-opening the same RP in different SERP.

Regarding query terminology, we could see that this process is not as easy and objective as it seems. Future works can use resources like the Consumer Health Vocabulary in automatic processes, complemented, later on, by a health professional validation.

It would be interesting to complement this study with an eye-tracking analysis that can give insights into where users look before clicking on a SERP result or an RP. We noticed that some users did not click on the results because they read the information presented on SERP.

This information can be viewed in the titles, snippets, and answer boxes displayed. Visible examples were “fiber consumption reduces the risk of testicular cancer”, “a diet rich in fiber is considered to be protective against cancer”, “fruits and vegetables are cancer-fighting foods” in titles and snippets. In the answer box, “Asthma has no cure” and the difference between saturated and polyunsaturated fats was also visible. In several cases, we could notice that this information was sufficient to answer the question behind the task. This leads us to conclude that people looked and read some information on SERP that led them to not click on any results because they immediately wrote the answer to the question and went on to the next search task. Gaze behavior could help us understand better what factors are associated with task completion.

As a future study, it would be important to consider the role of the user. If he is a caregiver, professional, patient, and in what state he is in relation to the disease.

Finally, we present possible research hypotheses for future work. It would be useful to apply this model to other studies to analyze its applicability more systematically. We could see if our significant differences are observed in other studies and if the other measures can prove to be significant.

References

Agitchtein, Eugene, Eric Brill, Susan Dumais, and Robert Ragno. 2006. “Learning User Interaction Models for Predicting Web Search Result Preferences.” In *SIGIR’06*, 3–10. Seattle, Washington, USA: ACM.

American Medical Association. 1999. “Health Literacy: Report of the Council on Scientific Affairs.” *Journal of the American Medical Association*. Vol. 281. https://doi.org/10.1001/jama.281.6.552.

Anick, Peter. 2003. “Using Terminological Feedback for Web Search Refinement - A Log-Based Study.” *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, no. SPEC. ISS.: 88–95. https://doi.org/10.1145/860435.860453.

Aula, Anne. 2003. “Query Formulation in Web Information Search.” *Proceedings IADIS International Conference WWW/Internet 2003* 1: 403–10. https://doi.org/10.1002/meet.2009.1450460131.

Baker, David W, Mark V Williams, Ruth M Parker, Julie A Gazmararian, and Joanne Nurss. 1999. “Development of a Brief Test to Measure Functional Health Literacy.” *Patient Education and Counseling* 38: 33–42. https://doi.org/10.1016/S0738-3991(98)00116-5.

Berkman, Nancy D., Terry C. Davis, and Lauren McCormack. 2010. “Health Literacy: What Is It?” *Journal of Health Communication* 15 (SUPPL. 2): 9–19. https://doi.org/10.1080/10810730.2010.499985.

Bruza, P.D., and S. Dennis. 1997. “Query Reformulation on the Internet: Empirical Data and the Hyperindex Search Engine.” In *RIAO’97*, 488–99. http://www.workingweb.com.au/training/RIAO97.pdf.

Buscher, Georg, Ryen W. White, Susan T. Dumais, and Jeff Huang. 2012. “Large-Scale Analysis of Individual and Task Differences in Search Result Page Examination Strategies.” *WSDM 2012 - Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, 373–82. https://doi.org/10.1145/2124295.2124341.

Centers for Disease Control and Prevention. 2020. “Healthy People 2020.” 2020.

Chang, Yung Sheng. 2020. “Investigating the Relationship between Ehealth Literacy and Information Search Behaviors.” *CHIIR 2020 - Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, 499–502. https://doi.org/10.1145/3343413.3377944.

Chen, Weizhu, Zheng Chen, Dong Wang, Adish Singla, Yuchen Zhang, and Qiang Yang. 2012. "A Noise-Aware Click Model for Web Search Categories and Subject Descriptors." In *WSDM 2012 - Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, 313–22. Seattle, Washington, USA: ACM. https://doi.org/10.1145/2872427.2883033.

Cutilli, Carolyn Crane. 2010. "Seeking Health Information Online: What Sources Do Your Patients Use?" *Orthopaedic Nursing* 29 (3). https://doi.org/10.1001/jamapediatrics.2016.3109.

Davis, T. C., S. W. Long, R. H. Jackson, E. J. Mayeaux, R. B. George, P. W. Murphy, and M. A. Crouch. 1993. "Rapid Estimate of Adult Literacy in Medicine: A Shortened Screening Instrument." *Family Medicine* 25 (6): 391–95.

Direção-Geral da Saúde. 2019. "Plano de Ação Para a Literacia Em Saúde." https://www.dgs.pt/documentos-e-publicacoes/plano-de-acao-para-a-literacia-em-saude-2019-2021.aspx.

Downey, Doug, Susan Dumais, Dan Liebling, and Eric Horvitz. 2008. "Understanding the Relationship between Searchers' Queries and Information Goals." *International Conference on Information and Knowledge Management, Proceedings*, 449–58. https://doi.org/10.1145/1458082.1458143.

Duggan, Geoffrey B., and Stephen J. Payne. 2008. "Knowledge in the Head and on the Web: Using Topic Expertise to Aid Search." *Conference on Human Factors in Computing Systems - Proceedings*, no. January 2008: 39–48. https://doi.org/10.1145/1357054.1357062.

Efthymiou, Areti, Nicos Middleton, Andreas Charalambous, and Evridiki Papastavrou. 2017. "Toward a Comprehensive Model of EHealth Literacy." *JMIR Research Protocols* 6 (11): e221. https://doi.org/10.2196/resprot.8080.

Espanha, Rita, Patrícia Ávila, and Rita Veloso Mendes. 2016. "Literacia Em Saúde Em Portugal." Lisboa. https://content.gulbenkian.pt/wp-content/uploads/2016/05/29203225/PGISVersCurtaFCB_FINAL2016.pdf.

Eurostat. 2018. "Individuals Using the Internet for Seeking Health-Related Information." 2018. https://ec.europa.eu/eurostat/tgm/graph.do?tab=graph&plugin=1&pcode=tin00101&language=en&toolbox=data.

———. 2020. "53% of EU Citizens Sought Health Information Online." European Commision.

2020. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20200327-1.

Eysenbach, Gunther, and Christian Köhler. 2002. “How Do Consumers Search for and Appraise Health Information on the World Wide Web? Qualitative Study Using Focus Groups, Usability Tests, and in-Depth Interviews.” *British Medical Journal* 324 (7337): 573–77. https://doi.org/10.1136/bmj.324.7337.573.

Fiksdal, Alexander S., Ashok Kumbamu, Ashutosh S. Jadhav, Cristian Cocos, Laurie A. Nelsen, Jyotishman Pathak, and Jennifer B. McCormick. 2014. “Evaluating the Process of Online Health Information Searching: A Qualitative Approach to Exploring Consumer Perspectives.” *Journal of Medical Internet Research* 16 (10): e224. https://doi.org/10.2196/jmir.3341.

Fox, Steve, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. 2005. “Evaluating Implicit Measures to Improve Web Search.” *ACM Transactions on Information Systems* 23 (2): 147–68. https://doi.org/10.1145/1059981.1059982.

Fox, Susannah. 2013. “What Ails America? Dr. Google Can Tell You | Pew Research Center.” Pew Research Center. 2013. https://www.pewresearch.org/fact-tank/2013/12/17/what-ails-america-dr-google-can-tell-you/.

———. 2014. “The Social Life of Health Information.” Pew Research Center. 2014. https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/.

Guo, Jiafeng, Gu Xu, Hang Li, and Xueqi Cheng. 2008. “A Unified and Discriminative Model for Query Refinement Categories and Subject Descriptors.” *Search*, 379–86.

Guo, Qi, and Eugene Agichtein. 2010. “Ready to Buy or Just Browsing? Detecting Web Searcher Goals from Interaction Data.” *SIGIR 2010 Proceedings - 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 130–37. https://doi.org/10.1145/1835449.1835473.

Guo, Qi, RW White, and ST Dumais. 2010. “Predicting Query Performance Using Query, Result, and User Interaction Features.” *RIAO ’10 Adaptivity, Personalization and Fusion of Heterogeneous Information*, 198–201. http://dl.acm.org/citation.cfm?id=1937104.

Gutierrez, Natalia, Tiffany B. Kindratt, Patti Pagels, Barbara Foster, and Nora E. Gimpel. 2014. “Health Literacy, Health Information Seeking Behaviors and Internet Use among Patients Attending a Private and Public Clinic in the Same Geographic Area.” *Journal of Community Health* 39 (1): 83–89. https://doi.org/10.1007/s10900-013-9742-5.

Hassan, Ahmed. 2012. “A Semi-Supervised Approach to Modeling Web Search Satisfaction.” *SIGIR’12 - Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, 275–84. https://doi.org/10.1145/2348283.2348323.

Hassan, Ahmed, and Ryen W. White. 2013. “Personalized Models of Search Satisfaction.” *International Conference on Information and Knowledge Management, Proceedings*, 2009–18. https://doi.org/10.1145/2505515.2505681.

He, Daqing, Ayşe Göker, and David J. Harper. 2002. “Combining Evidence for Automatic Web Session Identification.” *Information Processing and Management* 38 (5): 727–42. https://doi.org/10.1016/S0306-4573(01)00060-7.

Hembrooke, Helene A., Laura A. Granka, Geraldine K. Gay, and Elizabeth D. Liddy. 2005. “The Effects of Expertise and Feedback on Search Term Selection and Subsequent Learning.” *Journal of the American Society for Information Science and Technology* 56 (8): 861–71. https://doi.org/10.1002/asi.20180.

Higgins, Orla, Jane Sixsmith, Margaret M. Barry, and Christine Domegan. 2011. “A Literature Review on Health Information Seeking Behaviour on the Web: A Health Consumer and Health Professional Perspective.” Stockholm. https://doi.org/10.2900/5788.

Hölscher, Christoph, and Gerhard Strube. 2000. “Web Search Behavior of Internet Experts and Newbies.” *Computer Networks* 33 (1): 337–46. https://doi.org/10.1016/S1389-1286(00)00031-1.

Hong, Yi, Norberto de la Cruz, Gary Barnas, Eileen Early, and Rick Gillis. 2002. “A Query Analysis of Consumer Health Information Retrieval.” *Proceedings of the AMIA Symposium*, 1046.

Hu, Rong, Kun Lu, and Soohyung Joo. 2013. “Effects of Topic Familiarity and Search Skills on Query Reformulation Behavior.” *Proceedings of the ASIST Annual Meeting* 50 (1). https://doi.org/10.1002/meet.14505001062.

Huang, Jeff, and Efthimis N. Efthimiadis. 2009. “Analyzing and Evaluating Query Reformulation Strategies in Web Search Logs.” *International Conference on Information and Knowledge Management, Proceedings*, 77–86. https://doi.org/10.1145/1645953.1645966.

Huang, Jeff, Ryen W. White, and Susan Dumais. 2011. “No Clicks, No Problem: Using Cursor Movements to Understand and Improve Search.” In *CHI 2011*, 1225–34. Vancouver, BC, Canada: ACM. https://doi.org/10.1111/j.1550-7408.1981.tb02855.x.

Huang, Kun, Jiaqi Chen, Chang Liu, and Lulu Zhang. 2020. “A Comparative Study of the Relationship between the Subjective Difficulty, Objective Difficulty of Search Tasks and Search Behaviors,” 421–24. https://doi.org/10.1145/3383583.3398614.

Irwin, P. M. 1991. “National Literacy Act of 1991: Major Provisions of P.L. 102-73.” Washington, D.C. http://files.eric.ed.gov/fulltext/ED341851.pdf.

Jacobs, Wura, Ann O. Amuta, and Kwon Chan Jeon. 2017. “Health Information Seeking in the Digital Age: An Analysis of Health Information Seeking Behavior among US Adults.” *Cogent Social Sciences* 3 (1): 1–11. https://doi.org/10.1080/23311886.2017.1302785.

Jansen, Bernard J., Danielle L. Booth, and Amanda Spink. 2009. “Patterns of Query Reformulation During Web Searching.” *Journal of the American Society for Information Science and Technology* 60 (7): 1358–71. https://doi.org/10.1002/asi.21071.

Jansen, Bernard J., Amanda Spink, Chris Blakely, and Sherry Koshman. 2007. “Defining a Session on Web Search Engines.” *Journal of the American Society for Information Science and Technology* 58 (6): 862–71. https://doi.org/10.1002/asi.20564.

Kelly, D., and N. J. Belkin. 2001. “Reading Time, Scrolling and Interaction: Exploring Implicit Sources of User Preferences for Relevance Feedback.” *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, no. 558: 408–9.

Kelly, Diane, and Colleen Cool. 2002. “The Effects of Topic Familiarity on Information Search Behavior.” *Proceedings of the ACM International Conference on Digital Libraries*, 74–75. https://doi.org/10.1145/544229.544232.

Kickbusch, Ilona. 1997. “Think Health: What Makes the Difference?” *Health Promotion International* 12 (4): 265–72. https://doi.org/10.1093/heapro/12.4.265.

Kickbusch, Ilona, Jürgen M. Pelikan, Franklin Apfel, and Agis D. Tsouros. 2013. “Health Literacy: The Solid Facts.” *The Regional Office for Europe of the World Health Organization*. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Health+literacy+The+solid+facts#1%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Health+literacy:+the+solid+facts#1.

Kim, Su Hyun, and Sonja Utz. 2018. “Association of Health Literacy with Health Information-Seeking Preference in Older People: A Correlational, Descriptive Study.” *Nursing and Health Sciences* 20 (3): 355–60. https://doi.org/10.1111/nhs.12413.

Kinley, Khamsum, Dian Tjondronegoro, Helen Partridge, and Sylvia Edwards. 2012. "Human-Computer Interaction: The Impact of Users' Cognitive Styles on Query Reformulation Behaviour during Web Searching." *Proceedings of the 24th Australian Computer-Human Interaction Conference, OzCHI 2012*, no. 2007: 299–307. https://doi.org/10.1145/2414536.2414586.

Kutner, Mark, Elizabeth Greenberg, Ying Jin, and Christine Paulsen. 2006. "The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy (NCES 2006–483)." Washington, D.C. https://nces.ed.gov/pubs2006/2006483.pdf.

Lam, Mary K., and Lawrence T. Lam. 2012. "Health Information-Seeking Behaviour on the Internet and Health Literacy among Older Australians." *Electronic Journal of Health Informatics* 7 (2).

Lambert, Sylvie D., and Carmen G Loiselle. 2007. "Health Research Health Information – Seeking Behavior." *Qualitative Health Research* 17 (8): 1006–19. https://doi.org/10.1177/1049732307305199.

Lau, Tessa, and Eric Horvitz. 1999. "Patterns of Search: Analyzing and Modeling Web Query Refinement." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 407: 119–28. https://doi.org/10.1007/978-3-7091-2490-1_12.

Lee, Shoou Yih D., Deborah E. Bender, Rafael E. Ruiz, and Ik Cho Young. 2006. "Development of an Easy-to-Use Spanish Health Literacy Test." *Health Services Research* 41 (4 I): 1392–1412. https://doi.org/10.1111/j.1475-6773.2006.00532.x.

Li, Yuelin, and Nicholas J. Belkin. 2010. "An Exploration of the Relationships between Work Task and Interactive Information Search Behavior." *Journal of the American Society for Information Science and Technology* 61 (9): 1771–89. https://doi.org/10.1002/asi.21359.

Liu, Chang. 2012. "Personalizing Information Retrieval Using Interaction Behaviors in Search Sessions in Different Types of Tasks." Graduate School-New Brunswick Rutgers, The State University of New Jersey.

Liu, Chang, Jacek Gwizdka, and Nicholas J Belkin. 2010. "Analysis of Query Reformulation Types on Different Search Tasks." *A Poster to Be Presented at IConference 2010, Urbana-Champaign, IL, February 3–6, 2010.*, 1–5.

Liu, Chang, Jacek Gwizdka, Jingjing Liu, Tao Xu, and Nicholas J. Belkin. 2010. "Analysis and

Evaluation of Query Reformulations in Different Task Types." *Proceedings of the ASIST Annual Meeting* 47. https://doi.org/10.1002/meet.14504701214.

Liu, Jiqun, and Fangyuan Han. 2020. "Investigating Reference Dependence Effects on User Search Interaction and Satisfaction," 1141–50. https://doi.org/10.1145/3397271.3401085.

Lopes, Carla Teixeira, and Cristina Ribeiro. 2015. "Effects of Terminology on Health Queries: An Analysis by User's Health Literacy and Topic Familiarity." In *Current Issues in Libraries, Information Science and Related Fields (Advances in Librarianship)*, edited by Anne Woodsworht and David W. Penniman, 39:145–84. Emerald Group Publishing Limited. https://doi.org/10.1108/S0065-283020150000039013.

———. 2016. "Effects of Language and Terminology on the Usage of Health Query Suggestions." In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9822 LNCS:83–95. Springer Verlag. https://doi.org/10.1007/978-3-319-44564-9_7.

Mackert, Michael, Sara E. Champlin, Keryn E. Pasch, and Barry D. Weiss. 2013. "Understanding Health Literacy Measurement through Eye Tracking." *Journal of Health Communication* 18 (SUPPL. 1): 185–96. https://doi.org/10.1080/10810730.2013.825666.

Malloy-Weir, Leslie, Cathy Charles, Amiram Gafni, and Vikki Entwisle. 2016. "A Review of Health Literacy: Definitions, Interpretations, and Implications for Policy Initiatives." *J Public Health Pol* 37: 334–52. https://doi.org/10.1057/jphp.2016.18.

Manafo, Elizabeth, and Sharon Wong. 2012. "Exploring Older Adults' Health Information Seeking Behaviors." *Journal of Nutrition Education and Behavior* 44 (1): 85–89. https://doi.org/10.1016/j.jneb.2011.05.018.

Mancuso, Josephine M. 2008. "Health Literacy: A Concept/Dimensional Analysis." *Nursing and Health Sciences* 10 (3): 248–55. https://doi.org/10.1111/j.1442-2018.2008.00394.x.

Mastora, Anna, Maria Monopoli, and Sarantos Kapidakis. 2008. "Exploring Query Formulation and Reformulation: A Preliminary Study to Map Users' Search Behaviour." In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5173 LNCS:427–30. https://doi.org/10.1007/978-3-540-87599-4_54.

Maxwell, David, Leif Azzopardi, and Yashar Moshfeghi. 2017. "A Study of Snippet Length and

Informativeness Behaviour, Performance and User Experience." *SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 135–44. https://doi.org/10.1145/3077136.3080824.

McCray, Alexa T., and Tony Tse. 2003. "Understanding Search Failures in Consumer Health Information Systems." *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium*, 430–34.

Miller, Lisa M.Soederberg, and Robert A. Bell. 2012. "Online Health Information Seeking: The Influence of Age, Information Trustworthiness, and Search Challenges." *Journal of Aging and Health* 24 (3): 525–41. https://doi.org/10.1177/0898264311428167.

Moran, Kate, and Cami Goray. 2020. "Three Key SERP Features: Featured Snippets, People Also Ask, and Knowledge Panels." Nielsen Norman Group. April 26, 2020. https://www.nngroup.com/articles/key-serp-features/.

Moz. 2020. "Google SERP Features [2020 SEO] - Moz." 2020. https://moz.com/learn/seo/serp-features.

Mukherjee, Abir, and David Bawden. 2012. "Health Information Seeking in the Information Society." *Health Information and Libraries Journal* 29 (3): 242–46. https://doi.org/10.1111/j.1471-1842.2012.00997.x.

Niedźwiedzka, Barbara, Mario Mazzocchi, Jessica Aschemann-Witzel, Laura Gennaro, Wim Verbeke, and W. Bruce Traill. 2014. "Determinants of Information Behaviour and Information Literacy Related to Healthy Eating among Internet Users in Five European Countries." *Information Research* 19 (3).

Nielsen-Bohlman, Lynn, Allison M Panzer, and A David. 2005. *Health Literacy: A Prescription to End Confusion. Choice Reviews Online*. Vol. 42. https://doi.org/10.5860/choice.42-4059.

Nutbeam, Don. 1998. "Health Promotion Glossary." *Health Promotion International* 13 (4): 349–64. https://doi.org/10.1093/heapro/13.4.349.

———. 2000. "Health Literacy as a Public Health Goal: A Challenge for Contemporary Health Education and Communication Strategies into the 21st Century." *Health Promotion International* 15 (3): 259–67. https://doi.org/10.1093/heapro/15.3.259.

Paiva, Dagmara, Susana Silva, Milton Severo, Pedro Ferreira, Osvaldo Santos, Nuno Lunet, and Ana Azevedo. 2014. "Cross-Cultural Adaptation and Validation of the Health

Literacy Assessment Tool METER in the Portuguese Adult Population." *Patient Education and Counseling* 97 (2): 269–75. https://doi.org/10.1016/j.pec.2014.07.024.

Paiva, Dagmara, Susana Silva, Milton Severo, Pedro Moura-Ferreira, Nuno Lunet, and Ana Azevedo. 2017. "Limited Health Literacy in Portugal Assessed with the Newest Vital Sign Prevalência de Literacia Em Saúde Inadequada Em Portugal Medida Com o Newest Vital Sign." *Acta Medica Portuguesa* 30 (12): 861–69. https://doi.org/10.20344/amp.9135.

———. 2019. "Validation of the Short Assessment of Health Literacy in Portuguese-Speaking Adults in Portugal." *Gaceta Sanitaria*. https://doi.org/10.1016/j.gaceta.2019.03.005.

Parker, Ruth M., David W. Baker, Mark V. Willia, and Joanne R. Nurss. 1995. "The Test of Functional Health Literacy in Adults: A New Instrument for Measuring Patients' Literacy Skills." *Journal of General Internal Medicine* 10 (10): 537–41. https://doi.org/10.1007/BF02640361.

Pleasant, Andrew. 2014. "Advancing Health Literacy Measurement: A Pathway to Better Health and Health System Performance." *Journal of Health Communication* 19 (12): 1481–96. https://doi.org/10.1080/10810730.2014.954083.

Rawson, Katherine A., John Gunstad, Joel Hughes, Mary Beth Spitznagel, Vanessa Potter, Donna Waechter, and James Rosneck. 2010. "The METER: A Brief, Self-Administered Measure of Health Literacy." *Journal of General Internal Medicine* 25 (1): 67–71. https://doi.org/10.1007/s11606-009-1158-7.

Rieh, Soo Young, and Hong Xie. 2006. "Analysis of Multiple Query Reformulations on the Web: The Interactive Information Retrieval Context." *Information Processing and Management* 42 (3): 751–68. https://doi.org/10.1016/j.ipm.2005.05.005.

Safeer, Richard S., and Jann Keenan. 2005. "Health Literacy: The Gap between Physicians and Patients." *American Family Physician* 72 (3): 463–68. https://doi.org/10.1007/s11606-009-1158-7.

Sahib, Nuzhah Gooda, Anastasios Tombros, and Tony Stockman. 2012. "A Comparative Analysis of the Information-Seeking Behavior of Visually Impaired and Sighted Searchers." *Journal of the American Society for Information Science and Technology* 63 (2): 377–91. https://doi.org/10.1002/asi.21696.

Selden, Catherine R., Marcia Zorn, Scott C Ratzan, Ruth M Parker, and Compilers. 2000. "National Library of Medicine Current Bibliographies in Medicine: Health Literacy." *National Library of Medicine Current Bibliographies in Medicine: Health Literacy*.,

no. January 2000: 33 p; 617 citations. https://doi.org/10.1017/CBO9781107415324.004.

Simonds, Scott K. 1974. “Health Education as Social Policy.” *Health Education Monographs* 2 (1_suppl): 1–10. https://doi.org/10.1177/10901981740020s102.

Soldaini, Luca, Andrew Yates, Elad Yom-Tov, Ophir Frieder, and Nazli Goharian. 2016. “Enhancing Web Search in the Medical Domain via Query Clarification.” *Information Retrieval* 19 (1–2): 149–73. https://doi.org/10.1007/s10791-015-9258-y.

Sørensen, Kristine, Stephan Van Den Broucke, James Fullam, Gerardine Doyle, Jürgen Pelikan, Zofia Slonska, and Helmut Brand. 2012. “Health Literacy and Public Health: A Systematic Review and Integration of Definitions and Models.” *BMC Public Health* 12 (1): 80. https://doi.org/10.1186/1471-2458-12-80.

Sousa, Hugo Miguel Ribeiro de. 2016. “HealthTranslator: Automatic Annotation of Web Documents in Order to Assist Health Consumerapos;s Searches.” Universidade do Porto.

Spink, Amanda, Bernard J. Jansen, and H. Cenk Ozmultu. 2000. “Use of Query Reformulation and Relevance Feedback by Excite Users.” *Internet Research* 10 (4): 317–28. https://doi.org/10.1108/10662240010342621.

Spink, Amanda, Yin Yang, Jim Jansen, Pirrko Nykanen, Daniel P. Lorence, Seda Ozmutlu, and H. Cenk Ozmutlu. 2004. “A Study of Medical and Health Queries to Web Search Engines.” *Health Information & Libraries Journal* 21 (1): 44–51. https://doi.org/10.1111/j.1471-1842.2004.00481.x.

Teevan, Jaime, Eytan Adar, Rosie Jones, and Michael A. S. Potts. 2007. “Information Re-Retrieval: Repeat Queries in Yahoo’s Logs.” In *SIGIR’07*, 151. https://doi.org/10.1145/1277741.1277770.

Teevan, Jaime, Susan T. Dumais, and Daniel J. Liebling. 2008. “To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent.” *ACM SIGIR 2008 - 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Proceedings*, 163–70. https://doi.org/10.1145/1390334.1390364.

Toms, Elaine G., and Celeste Latter. 2007. “How Consumers Search for Health Information.” *Health Informatics Journal* 13 (3): 223–35. https://doi.org/10.1177/1460458207079901.

Vanopstal, Klaar, Robert Vander Stichele, Godelieve Laureys, and Joost Buysschaert. 2012.

"PubMed Searches by Dutch-Speaking Nursing Students: The Impact of Language and System Experience." *Journal of the American Society for Information Science and Technologyn Science and Technology* 63 (8): 1538–52. https://doi.org/10.1002/asi.

Vida Estacio, Emee, Rebecca Whittle, and Joanne Protheroe. 2019. "The Digital Divide: Examining Socio-Demographic Factors Associated with Health Literacy, Access and Use of Internet to Seek Health Information." *Journal of Health Psychology* 24 (12): 1668–75. https://doi.org/10.1177/1359105317695429.

Weiss, Barry D., Mary Z. Mays, William Martz, Kelley Merriam Castro, Darren A. DeWalt, Michael P. Pignone, Joy Mockbee, and Frank A. Hale. 2005. "Quick Assessment of Literacy in Primary Care: The Newest Vital Sign." *Annals of Family Medicine* 3 (6): 514–22. https://doi.org/10.1370/afm.405.College.

White, Ryen W., and Susan T. Dumais. 2009. "Characterizing and Predicting Search Engine Switching Behavior." *International Conference on Information and Knowledge Management, Proceedings*, 87–96. https://doi.org/10.1145/1645953.1645967.

White, Ryen W., Susan T. Dumais, and Jaime Teevan. 2009. "Characterizing the Influence of Domain Expertise on Web Search Behavior." *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining, WSDM'09*, 132–41. https://doi.org/10.1145/1498759.1498819.

White, Ryen W., Susan Dumais, and Jaime Teevan. 2008. "How Medical Expertise Influences Web Search Interaction." *ACM SIGIR 2008 - 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Proceedings*, 791–92. https://doi.org/10.1145/1390334.1390506.

White, Ryen W., and Dan Morris. 2007. "Investigating the Querying and Browsing Behavior of Advanced Search Engine Users." *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07*, 255–62. https://doi.org/10.1145/1277741.1277787.

White, Ryen W., Matthew Richardson, Mikhail Bilenko, and Allison P. Heath. 2008. "Enhancing Web Search by Promoting Multiple Search Engine Use." *ACM SIGIR 2008 - 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Proceedings*, 43–50. https://doi.org/10.1145/1390334.1390344.

Whittle, Martin, Barry Eaglestone, Nigel Ford, Valerie J Gillet, and Andrew Madden. 2007. "Data Mining of Search Engine Logs." *Journal of the American Society for Information Science and Technology* 58 (14): 2382–2400. https://doi.org/10.1002/asi.20733.

Wildemuth, Barbara M. 2004. “The Effects of Domain Knowledge on Search Tactic Formulation.” *Journal of the American Society for Information Science and Technology* 55 (3): 246–58. https://doi.org/10.1002/asi.10367.

World Health Organization. 2020. “The WHO Health Promotion Glossary.” 2020. https://www.who.int/healthpromotion/about/HPG/en/.

Wu, Dan, and Yizhe Li. 2016. “Online Health Information Seeking Behaviors among Chinese Elderly.” *Library and Information Science Research* 38 (3): 272–79. https://doi.org/10.1016/j.lisr.2016.08.011.

Yamin, Fadhilah Mat, and T. Ramayah. 2011. “User Web Search Behavior on Query Formulation.” *2011 International Conference on Semantic Technology and Information Retrieval, STAIR 2011*, no. June: 182–88. https://doi.org/10.1109/STAIR.2011.5995786.

Yilma, Tesfahun Melese, Anushia Inthiran, Daniel D. Reidpath, and Sylvester Olubolu Orimaye. 2019. “Context-Based Interactive Health Information Searching.” *Information Research* 24 (2).

Yom-Tov, Elad, Barbara Marino, Jennifer Pai, Dawn Harris, and Michael Wolf. 2016. “The Effect of Limited Health Literacy on How Internet Users Learn About Diabetes.” *Journal of Health Communication* 21 (10): 1107–14. https://doi.org/10.1080/10810730.2016.1222033.

Zeng, Qing T., and Tony Tse. 2006. “Exploring and Developing Consumer Health Vocabularies.” *Journal of the American Medical Informatics Association* 13 (1): 24–29. https://doi.org/10.1197/jamia.M1761.

Zhang, Xiangmin, Jingjing Liu, Michael Cole, and Nicholas Belkin. 2015. “Predicting Users’ Domain Knowledge in Information Retrieval Using Multiple Regression Analysis of Search Behaviors.” *Journal of the American Society for Information Science and Technology* 66 (5): 980–1000. https://doi.org/10.1002/asi.

Zhang, Yan. 2013. “The Effects of Preference for Information on Consumers’ Online Health Information Search Behavior.” *Journal of Medical Internet Research* 15 (11). https://doi.org/10.2196/jmir.2783.

Zhang, Yan, Peiling Wang, Amy Heaton, and Heidi Winkler. 2012. “Health Information Searching Behavior in MedlinePlus and the Impact of Tasks.” *IHI’12 - Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 641–50. https://doi.org/10.1145/2110363.2110434.

Zhu, Han, Chang Liu, and Xiaoxuan Song. 2018. “Identification of Information Search Strategies and Their Impact on Learning Outcome in Learning-Related Tasks.” *Proceedings of the Association for Information Science and Technology* 55 (1): 960–62. https://doi.org/10.1002/pra2.2018.14505501194.

Zimmerman, Margaret S., and George Shaw. 2020. “Health Information Seeking Behaviour: A Concept Analysis.” *Health Information and Libraries Journal* 37 (3): 173–91. https://doi.org/10.1111/hir.12287.

Appendices

Appendix 1 - Questionnaire used in the experience

Parte I (Asma)

Indique, para cada item, se é verdadeiro ou falso.

Usa-se um inalador de alívio (broncodilatador ou "bomba") para reduzir a inflamação dos pulmões.	
Usa-se um registo de DEMI (Débito Expiratório Máximo Instantâneo) para ter a certeza que os seios perinasais estão abertos (teste do sopro para detetar se há sinusite).	
Um alergénio é o anticorpo que falta aos asmáticos.	
A maioria das crianças asmáticas tem de ir ao hospital quando tem um ataque de asma.	
Uma criança deixa de ser asmática se, durante vários anos, deixar de ter sintomas como o aperto no peito ou pieira.	

Parte II (Nutrição)

Que doenças ou problemas de saúde estão relacionados com a baixa ingestão de fibra?

(Indique 3 problemas)

_____

Pensas que estes comportamentos ajudam a reduzir a probabilidade de vir a ter certos tipos de cancro?

	Sim	Não
Comer mais fibra		
Comer mais frutas e legumes		

Acreditas que estes comportamentos ajudam a prevenir doenças do coração?

	Sim	Não
Comer menos gordura saturada		
Comer menos sal		
Comer mais frutas e legumes		

Qual destes nutrientes mais contribui para aumentar os níveis de colesterol do sangue das pessoas? *(Escolha uma opção)*

Antioxidantes	
Gorduras polinsaturadas	
Gorduras saturadas	
Colesterol da dieta	

Quais destas vitaminas acreditas que são antioxidantes?

	Sim	Não
Vitamina A		
Vitaminas do complexo B		
Vitamina C		
Vitamina D		
Vitamina E		
Vitamina K		

Source - Sousa (2016)

Appendix 2 – METER-PT Health Literacy Instrument

A lista seguinte inclui alguns termos que existem na linguagem médica. Alguns desses termos estão relacionados com partes ou funções do corpo, com tipos de doenças ou com coisas que podem melhorar ou piorar a saúde. A lista também contém algumas palavras que podem parecer ou soar como termos reais, mas que não existem.

À medida que for lendo esta lista, coloque uma cruz "X" ao lado das palavras que são termos reais. Não tente adivinhar. Coloque uma cruz "X" ao lado das palavras só quando tiver a certeza que existem mesmo.

_____ Imígdala	_____ Jezum
_____ **Artrite**	_____ Súrgico
_____ **Obesidade**	_____ Malorias
_____ **Gripe**	_____ **Cancro**
_____ Nervosite	_____ Alcoolidade
_____ **Sífilis**	_____ **Antibióticos**
_____ **Potássio**	_____ Antideprimido
_____ **Hormonas**	_____ **Colite**
_____ **Nervos**	_____ **Diabetes**
_____ Anquia	_____ Otorringologista
_____ Cástula	_____ Nósea

_____ Ingesto	_____ **Impetigo**
_____ Intestigo	_____ **Menstrual**
_____ **Exercício**	_____ Gatarral
_____ **Pústula**	_____ **Convulsão**
_____ Cerpes	_____ **Apêndice**
_____ **Rim**	_____ Abdominável
_____ **Urgência**	_____ Enxuteca
_____ Xirope	_____ **Dose**
_____ **Menopausa**	_____ **Hemorroidas**
_____ **Diagnóstico**	_____ **Testículo**
_____ Candíase	_____ **Olho**
_____ **Icterícia**	_____ Obstérico
_____ **Bexiga**	_____ Sonambulação
_____ **Aborto**	_____ Drenação
_____ **Hepatite**	_____ **Sexualmente**
_____ Enatoma	_____ Purisia
_____ Unhal	_____ Fibrómico
_____ **Asma**	_____ **Medicação**
_____ **Inflamatório**	_____ **Micróbios**
_____ **Anemia**	_____ **Gonorreia**
_____ Linsoma	_____ Estómico
_____ Ceresiana	_____ **Fadiga**
_____ **Stress**	_____ **Osteoporose**
_____ Algérico	_____ **Obstipação**

Source – Paiva et al. (2014)

Appendix 3 - Initial version of the classification scheme

EventType	Browser interaction	QueryReFormulation	SERPinteraction	RPinteraction	ExtensionInteraction	Questionnaire	SEinteraction	Others
BrowserInteraction	OpenBrowser	AddressBarQuery	ShowingResultsFor	Select2ndLevelDepthLink	Activate Extension	ReadQuestion	ImFeelingLucky	GoogleTranslate
QueryReFormulation	CloseBrowser	SearchEngineBoxQuery	DidYouMean	SelectTextRP	Deactivate Extension	WriteAnswer		
SERPinteraction	OpenNewTab	ModifyQuery	LeftClickOrganicResult	SelectAnchorLink	HoverAnItem			
RPinteraction	SelectTab	ContinuationQueryFormulation	LeftClickSponsoredResult	SelectImage	ExtensionMoreInformation			
ExtensionInteraction	CloseTab	RepeatQuery	RightClickOrganicResult	CloseAdvert	CloseInformationWindow			
Questionnaire	ReopenClosedTab	SelectTextQuery	RightClickSponsoredResult	QueryInternalSearchEngine	SelectTextExtension			
SEinteraction	ZoomIn	SearchByVoice	MiddleClickOrganicResult	OpenLinkNewTab	ExternalReferences			
	ZoomOut	IncrementalSearch	MiddleClickSponsoredResu	sds	FoundIn			
	TypeURL		Page2SERP	sdf	AssociatedMorphology			
	Backward		OpenLinkNewTab	sus	CausativeAgentOf			
	Forward		SERPGoogleTranslate	suf	CanAlsoBe			
	RefreshPage		AnswerBox		CauseOf			
	KeyboardShortcut		AnotherSERP		SameAs			
	CopyText		KnowledgeGraph		OtherRelations			
	PasteText		PeopleAlsoAsk		sds			
	SearchEnginePlugin		AllSERP		sdf			
	SelectTextURL		ImagesSERP		sus			
	RightClick		VideosSERP		suf			
	DoubleClick		BooksSERP					
	FindBar		NewsSERP					
	CloseFindBar		ToolsSERP					
	NextFindBarItem		AdvancedSearch					
	Home		RelatedSearches					
	Settings		SelectTextSERP					
			PrivacyReminder					
			sds					
			sdf					
			sus					
			suf					

Appendix 4 – Initial version of the annotation

User	Literacy	Extension	Question	Time	Duration	Instant	EventType	Name	ID	IDSerp	SERPRankClick	Observations
1	1	1	1	3	3	3,47E-05	BrowserInteraction	OpenNewTab	0	0		
1	1	1	1	6	4	6,94E-05	Questionnaire	ReadQuestion	0	0		
1	1	1	1	10	3	0,000116	QueryReFormulation	AddressBarQuery	1	1		
1	1	1	1	13	7	0,00015	QueryReFormulation	IncrementalSearch	0	1		
1	1	1	1	20	6	0,000231	SERPinteraction	LeftClickOrganicResult	0	1	2	
1	1	1	1	26	2	0,000301	RPinteraction	sds	0	1		
1	1	1	1	28	3	0,000324	RPinteraction	sdf	0	1		
1	1	1	1	31	5	0,000359	ExtensionInteraction	HoverAnItem	0	1		
1	1	1	1	36	3	0,000417	Questionnaire	ReadQuestion	0	1		
1	1	1	1	39	12	0,000451	ExtensionInteraction	ExtensionMoreInformation	0	1		
1	1	1	1	51	1	0,00059	ExtensionInteraction	OtherRelations	0	1		
1	1	1	1	52	2	0,000602	ExtensionInteraction	CanAlsoBe	0	1		
1	1	1	1	54	5	0,000625	ExtensionInteraction	CloseInformationWindow	0	1		
1	1	1	1	59	6	0,000683	ExtensionInteraction	HoverAnItem	0	1		
1	1	1	1	65	4	0,000752	ExtensionInteraction	HoverAnItem	0	1		
1	1	1	1	69	4	0,000799	Questionnaire	WriteAnswer	0	1		
1	1	1	2	73	5	0,000845	Questionnaire	ReadQuestion	0	1		
1	1	1	2	78	2	0,000903	BrowserInteraction	SelectTextURL	0	1		

Appendix 5 – General Interaction

Measure	Average per Session		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#TotalEvents	26.16	**26.84**	0.01	0.94		0.36	95%	No statistical evidence	
TotalSeconds	91.98	**110.56**	0.03	0.37		0.16	95%	No statistical evidence	
#EventsTypeBrowserInteraction	**7.17**	6.14	0.50	0.14	0.27		95%	No statistical evidence	
SecondsBrowserInteraction	**25.28**	23.80	0.40	0.59	0.40		95%	No statistical evidence	
#EventsTypeSERPinteraction	**6.68**	5.98	0.01	0.85	0.33		95%	No statistical evidence	
SecondsSERPInteraction	18.09	**20.91**	0.00	0.57		0.27	95%	No statistical evidence	
#EventsTypeRPinteraction	8.96	**10.71**	0.26	0.31	0.14		95%	No statistical evidence	
SecondsRPInteraction	32.84	**45.36**	0.08	0.57	0.06*		95%	Significant differences	
#EventsTypeSearchEngineBoxQuery	3.49	**3.94**	0.29	0.72	0.25		95%	No statistical evidence	
SecondsSearchEngineBoxQuery	15.73	**21.48**	0.13	0.28	0.16		95%	No statistical evidence	
#EventsTypeDAPinteraction	0.02	**0.08**	0.00	0.00		0.27	95%	No statistical evidence	
SecondsDAPinteraction	0.06	**0.64**	0.00	0.00		0.27	95%	No statistical evidence	

Appendix 6 –Interaction with Browser

Measure	Average per Session		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#OpenBrowser									No data
#CloseBrowser									No data
#OpenNewTab	**0.25**	0.25	0.00	0.11		0.36	95%	No statistical evidence	
#SelectTab	**2.02**	1.40	0.14	0.02		0.22	95%	No statistical evidence	
#CloseTab	**1.27**	0.56	0.34	0.01		0.08*	95%	Significant differences	
#ReopenClosedTab	0.00	**0.03**	.	0.00		0.33	95%	No statistical evidence	
#InteractAddressBarQuery	0.02	**0.04**	0.00	0.00		0.39	95%	No statistical evidence	
#NewAddressBarQuery	**0.92**	0.58	0.83	0.09	0.10		95%	No statistical evidence	
#IncrementalSearch	**0.11**	0.04	0.00	0.00		0.37	95%	No statistical evidence	
#RepeatQuery	**0.02**	0.01	0.00	0.00		0.42	95%	No statistical evidence	
#ZoomIn	0.01	**0.01**	0.00	0.00		0.45	95%	No statistical evidence	
#ZoomOut									No data
#TypeURL	0.02	**0.06**	0.00	0.00		0.36	95%	No statistical evidence	
#Backward	0.54	**1.30**	0.00	0.32		0.05**	95%	Significant differences	
#Forward	0.00	**0.01**	.	0.00		0.33	95%	No statistical evidence	
#RefreshPage	**0.03**	0.00	0.00	.		0.39	95%	No statistical evidence	
#KeyboardShortcut	**0.05**	0.03	0.00	0.00		0.33	95%	No statistical evidence	

Measure	Average per Session		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#CopyText	**0.06**	0.05	0.00	0.00		0.25	95%	No statistical evidence	
#PasteText	**0.07**	0.04	0.00	0.00		0.22	95%	No statistical evidence	
#SelectTextURL	**0.71**	0.36	0.86	0.00		0.03**	95%	Significant differences	
#RightClick	0.02	**0.03**	0.00	0.00		0.45	95%	No statistical evidence	
#DoubleClick	**0.01**	0.00	0.00	.		0.39	95%	No statistical evidence	
#FindBar	**0.63**	0.63	0.09	0.02		0.36	95%	No statistical evidence	
#CloseFindBar	**0.03**	0.03	0.00	0.00		0.50	95%	No statistical evidence	
#NextFindBarItem	0.38	**0.70**	0.00	0.00		0.27	95%	No statistical evidence	
#Home									No data
#Settings									No data

Appendix 7 - Interaction with Search Engine Results Page

Measure	Average by SERP		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#ShowingResultsFor	0.03	**0.04**	0.00	0.00		0.48	95%	No statistical evidence	
#DidYouMean	**0.01**	0.00	0.00	.		0.39	95%	No statistical evidence	
#LeftClickOrganicResult	0.40	**0.75**	0.15	0.95	0.03**		95%	Significant differences	
#LeftClickPaidResult									No data
#LeftClickUniversalResult	**0.03**	0.01	0.00	0.00		0.05**	95%	Significant differences	
#LeftClickEnrichedSnippet									No data
#LeftClickKnowledgeData									No data
#RightClickOrganicResult	0.07	**0.15**	0.00	0.00		0.30	95%	No statistical evidence	
#RightClickPaidResult									No data
#RightClickUniversalResult	0.01	**0.02**	0.00	0.00		0.45	95%	No statistical evidence	
#RightClickEnrichedSnippet									No data
#RightClickKnowledgeData									No data
#MiddleClickOrganicResult	**0.46**	0.11	0.03	0.00		0.05**	95%	Significant differences	
#MiddleClickPaidResult									No data
#MiddleClickUniversalResult	**0.01**	0.00	0.00	0.00		0.27	95%	No statistical evidence	
#MiddleClickEnrichedSnippet									No data
#MiddleClickKnowledgeData									No data

Measure	Average by SERP		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#InteractionPaidResult									No data
#InteractionUniversalResult	**0.01**	0.00	0.00	.		0.18	95%	No statistical evidence	
#InteractionEnrichedSnippet									No data
#InteractionKnowledgeData									No data
#ToolsInteraction	**0.00**	0.00	0.00	.		0.39	95%	No statistical evidence	
#Page2SERP	0.00	**0.00**	.	0.00		0.33	95%	No statistical evidence	
#OpenLinkNewTab	0.07	**0.17**	0.00	0.00		0.30	95%	No statistical evidence	
#AnotherSERP	**0.01**	0.00	0.00	.		0.39	95%	No statistical evidence	
#AllSERP	**0.00**	0.00	0.00	.		0.39	95%	No statistical evidence	
#ImagesSERP	0.00	**0.00**	0.00	0.00		0.45	95%	No statistical evidence	
#VideosSERP									No data
#BooksSERP									No data
#NewsSERP									No data
#ToolsSERP	0.00	**0.04**	.	0.00		0.33	95%	No statistical evidence	
#AdvancedSearch									No data
#RelatedSearch									No data
#SelectTextSERP	**0.04**	0.01	0.00	0.00		0.20	95%	No statistical evidence	
#PrivacyReminder									No data

Measure	Average by SERP		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#sds	**0.58**	0.57	0.01	0.08		0.42	95%	No statistical evidence	
#sdf	**0.58**	0.57	0.01	0.08		0.42	95%	No statistical evidence	
#sus	0.20	**0.22**	0.23	0.00		0.36	95%	No statistical evidence	
#suf	0.20	**0.22**	0.23	0.00		0.36	95%	No statistical evidence	
#EventsTypeSERPinteraction	2.70	**2.90**	0.08	0.16	0.40		95%	No statistical evidence	
SecondsSERPInteraction	7.26	**9.68**	0.00	0.10		0.12	95%	No statistical evidence	
#ScrollingUP (sus+suf)	0.40	**0.45**	0.23	0.00		0.36	95%	No statistical evidence	
#ScrollingDown (sds+sdf)	**1.16**	1.15	0.01	0.08		0.42	95%	No statistical evidence	
#Scrolling (ScrollingUp+ScrollingDown)	1.56	**1.60**	0.03	0.04		0.33	95%	No statistical evidence	
SecondsScrollingUP	0.67	**1.20**	0.04	0.00		0.39	95%	No statistical evidence	
SecondsScrollingDown	**2.99**	2.73	0.00	0.20		0.45	95%	No statistical evidence	
SecondsScrolling	3.66	**3.93**	0.00	0.05		0.45	95%	No statistical evidence	
#SERP[1]	2.55	**2.36**	0.45	0.21	0.29		95%	No statistical evidence	
#SERPSwithClicks	0.67	**0.76**	0.80	0.73	0.08*		95%	Significant differences	
#SERPSwithoutClicks	**0.33**	0.24	0.80	0.73	0.08*		95%	Significant differences	
Timeto1stRPClicksinceSERPLoad	5.80	**8.30**	0.03	0.85		0.05**	95%	Significant differences	
MedRankClickedbySERP[2]	1.50	1.50	0.07	0.18		0.42	95%	No statistical evidence	
#ResultClick	0.96	**1.03**	0.63	0.97	0.28		95%	No statistical evidence	

Measure	Average by SERP		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#ResultClick[1]	**2.39**	2.34	0.73	0.72	0.44		95%	No statistical evidence	
#LeftClicks	0.42	**0.75**	0.17	0.93	0.04**		95%	Significant differences	
#MiddleClicks	**0.47**	0.11	0.03	0.00		0.03**	95%	Significant differences	
#RighClicks (with Open New Tab)	0.07	**0.17**	0.00	0.00		0.30	95%	No statistical evidence	
#ClicksOrganicResults	0.92	**1.01**	0.99	0.78		0.24	95%	No statistical evidence	
#ClicksUniversalResults	**0.04**	0.02	0.00	0.00		0.33	95%	No statistical evidence	

[1]The value in this line is the average per session. [2]The value in this line is the median per SERP.

Appendix 8 - Interaction with Results Page

Measure	Average by RP (viewed)		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#Select2ndLevelDepthLink	0.07	**0.12**	0.01	0.11	0.12		95%	No statistical evidence	
#SelectTextRP	**0.30**	0.18	0.00	0.01		0.45	95%	No statistical evidence	
#SelectAnchorLink	0.00	**0.01**	.	0.00		0.33	95%	No statistical evidence	
#SelectImage									No data
#CloseAdvert	0.07	**0.10**	0.03	0.10		0.25	95%	No statistical evidence	
#QueryInternalSearchEngine	0.00	**0.00**	.	0.00		0.33	95%	No statistical evidence	
#OpenLinkNewTab	0.00	**0.01**	.	0.00		0.33	95%	No statistical evidence	
#sds	1.57	**1.93**	0.35	0.93	0.05**		95%	Significant differences	
#sdf	1.57	**1.93**	0.35	0.93	0.05**		95%	Significant differences	
#sus	0.19	**0.37**	0.08	0.77	0.04**		95%	Significant differences	
#suf	0.19	**0.37**	0.08	0.77	0.04**		95%	Significant differences	
#InputData	0.00	**0.00**	.	0.00		0.33	95%	No statistical evidence	
#EventsTypeRPinteraction	3.98	**5.02**	0.61	0.79	0.05**		95%	Significant differences	
SecondsTypeRPInteraction	14.27	**21.12**	0.69	0.11	0.01***		95%	Significant differences	
#ScrollingUP (sus+suf)	0.38	**0.73**	0.08	0.77	0.04**		95%	Significant differences	
#ScrollingDown (sds+sdf)	3.15	**3.86**	0.35	0.93	0.05**		95%	Significant differences	
#Scrolling (ScrollingUp+ScrollingDown)	3.53	**4.59**	0.54	0.92	0.02**		95%	Significant differences	

Measure	Average by RP (viewed)		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
SecondsScrollingUP	0.92	**2.45**	0.02	0.59		0.02**	95%	Significant differences	
SecondsScrollingDown	11.88	**16.69**	0.47	0.03		0.06	95%	No statistical evidence	
SecondsScrolling	12.80	**19.14**	0.62	0.46	0.02**		95%	Significant differences	
#RP[1]	**2.27**	2.21	0.84	0.54	0.43		95%	No statistical evidence	
#RPnotViewed[1]	0.10	**0.14**	0.01	0.04		0.33	95%	No statistical evidence	
#RPnotViewed	0.04	**0.06**	0.03	0.01		0.45	95%	No statistical evidence	

[1]The value in this line is the average per session.

Appendix 9 - Interaction with Direct Access Page

Measure	Average by Session		Shapiro-Wilk Normality Test Sig.		Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#InputData	0.00	**0.08**	.	0.00	0.20	95%	No statistical evidence	
#Sus	**0.01**	0.00	0.00	.	0.39	95%	No statistical evidence	
#Suf	**0.01**	0.00	0.00	.	0.39	95%	No statistical evidence	
#ScrollingUP	**0.02**	0.00	0.00	.	0.39	95%	No statistical evidence	
SecondsScrollingUP	**0.06**	0.00	0.00	.	0.39	95%	No statistical evidence	
#DAP	0.01	**0.03**	0.00	0.00	0.30	95%	No statistical evidence	

Appendix 10 - Interaction with Search Engine Box Query

Measure	Average per Session		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
Event Name	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#InteractSearchEngineBoxQuery	0.20	**0.39**	0.00	0.01		0.06*	95%	Significant differences	
#NewSearchEngineBoxQuery	0.79	**1.08**	0.98	0.83	0.11		95%	No statistical evidence	
#PartialChange	**0.84**	0.71	0.88	0.90	0.22		95%	No statistical evidence	
#RepeatQuery	0.00	**0.04**	0.00	0.00		0.20	95%	No statistical evidence	
#SelectTextQuery	0.97	**1.19**	0.42	0.24	0.31		95%	No statistical evidence	
#SearchByVoice									No data
#IncrementalSearch	**0.69**	0,54	0.06	0.23	0.26		95%	No statistical evidence	
#SearchEngineBoxQueries (New + PartialChange)	1.63	**1.79**	0.42	0.36	0.30		95%	No statistical evidence	

Appendix 11 - Query Analysis based on event aggregation

Measure	Average per Session		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
Event Name	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#TotalInteracts	0.22	**0.43**	0.00	0.04		0.06*	95%	Significant differences	
#TotalNewQueries	**1.71**	1.65	0.56	0.27	0.81		95%	No statistical evidence	
#TotalIncrementalSearch	**0.80**	0.58	0.15	0.11	0.16		95%	No statistical evidence	
#TotalRepeatQuery	0.02	**0.05**	0.00	0.00		0.25	95%	No statistical evidence	
#TotalQueryEvents	4.43	**4.55**	0.81	0.71	0.43		95%	No statistical evidence	
#TimeQueryEvents	22.62	**26.33**	0.65	0.60	0.15		95%	No statistical evidence	

Appendix 12 - Analysis of Reformulation Modules

Measure	Average		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
Query (re)formulation per Session									
#Queries	**2.69**	2.51	0.32	0.16	0.28		95%	No statistical evidence	
#Reformulations	**1.69**	1.51	0.32	0.16	0.28		95%	No statistical evidence	
Query characteristics per Query									
#AveTermsperQuery	3.20	**3.43**	0.91	0.81	0.27		95%	No statistical evidence	
#OrthographicErrors	**0.07**	0.07	0.21	0.28	0.46		95%	No statistical evidence	
#TypographicalErrors	**0.02**	0.00	0.00	.		0.10*	95%	Significant differences	
#LayQueries	**0.42**	0.41	0.26	0.35	0.46		95%	No statistical evidence	
#MedSciQueries	0.58	**0.59**	0.26	0.35	0.46		95%	No statistical evidence	
#PTQueries	0.61	**0.76**	0.01	0.01		0.33	95%	No statistical evidence	
#ENQueries	**0.38**	0.23	0.01	0.01		0.39	95%	No statistical evidence	
#NALanguageQueries	**0.01**	0.01	0.00	0.00		0.42	95%	No statistical evidence	
#Reformulations	**0.61**	0.55	0.40	0.03	0.17		95%	No statistical evidence	
Query Reformulation types per Reformulation									
#OrthographicErrorCorrection	0.05	**0.09**	0.00	0.06		0.20	95%	No statistical evidence	
#TypographicalErrorCorrection	**0.02**	0.00	0.00	.		0.39	95%	No statistical evidence	
#TotalNew	0.07	**0.15**	0.01	0.50		0.03**	95%	Significant differences	

Measure	Average		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#LaytoMedSci	0.09	**0.13**	0.79	0.37	0.12		95%	No statistical evidence	
#MedScitoLay	**0.13**	0.10	0.85	0.17	0.27		95%	No statistical evidence	
#PTtoEN	0.05	**0.06**	0.07	0.00		0.42	95%	No statistical evidence	
#ENtoPT	0.03	**0.04**	0.00	0.03		0.33	95%	No statistical evidence	
#RepeatQuery	0.01	**0.07**	0.00	0.00		0.25	95%	No statistical evidence	
#SubstitutionWords	**0.13**	0.10	0.14	0.35	0.28		95%	No statistical evidence	
#SynonymiusSubstitution	**0.05**	0.01	0.01	0.00		0.08*	95%	Significant differences	
#AddWords	**0.51**	0.39	0.76	0.15	0.06*		95%	Significant differences	
#RemoveWords	**0.37**	0.30	0.96	0.28	0.15		95%	No statistical evidence	
#ExpandTotalAcronym	0.03	**0.06**	0.01	0.00		0.36	95%	No statistical evidence	
#ExpandPartialAcronym	**0.03**	0.00	0.00	.		0.27	95%	No statistical evidence	
#FormAcronym	0.00	**0.01**	.	0.00		0.33	95%	No statistical evidence	
#IntroduceBooleanOperators	0.00	**0.01**	.	0.00		0.33	95%	No statistical evidence	
#RemoveBooleanOperators	0.00	**0.01**	.	0.00		0.33	95%	No statistical evidence	
#IntroduceAccentuation	**0.01**	0.00	0.00	.		0.39	95%	No statistical evidence	
#RemoveAccentuation									No data
#Stemming	**0.00**	0.00	0.00	.		0.39	95%	No statistical evidence	
#VerbalFlexion	0.00	**0.01**	.	0.00		0.33	95%	No statistical evidence	
#AdjectiveforName	**0.02**	0.01	0.00	0.00		0.33	95%	No statistical evidence	

Measure	Average		Shapiro-Wilk Normality Test Sig.		Parametric Test of independent T-Test samples Sig. unilateral	Nonparametric Test Sig. unilateral	Confidence Test	Conclusion	Observations
	HHL	LHL	HHL	LHL	Assumed Equal Variances	Independent Mann-Whitney U Test Samples	Level	At a significance level of 5%	
#NameforAdjective	**0.01**	0.01	0.00	0.00		0.45	95%	No statistical evidence	
#VerbforName	0.01	**0.01**	0.00	0.00		0.48	95%	No statistical evidence	
#NameforVerb									No data
#VerbforAdjective									No data
#AdjectiveforVerb									No data
#PluraltoSingular	**0.06**	0.04	0.75	0.00		0.10*	95%	Significant differences	
#SingulartoPlural	**0.03**	0.02	0.00	0.00		0.30	95%	No statistical evidence	