Information retrieval is a central and essential activity. It is indeed difficult to find a human activity that does not need to retrieve information in an environment which is often increasingly digital: moving and navigating, learning, having fun, communicating, informing, making a decision, etc. Most human activities are intimately linked to our ability to search quickly and effectively for relevant information, the stakes are sometimes extremely important: passing an exam, voting, finding a job, remaining autonomous, being socially connected, developing a critical spirit, or simply surviving.

The author of this book presents a summary of work undertaken over several years relative to the behaviors and cognitive processes involved in information retrieval in digital environments. He presents several examples of theoretical models and studies to better understand the difficulties, behaviors and strategies of individuals searching for information in digital environments.

Jérôme Dinet is a Psychologist and Assistant Professor in cognitive psychology and ergonomics at the University of Lorraine in France.
Information Retrieval in Digital Environments
CHAPTER 1. INFORMATION RETRIEVAL IN DIGITAL ENVIRONMENTS: DEBATE AND SCIENTIFIC DIRECTIONS . . . 1

1.1. Information retrieval, current and future challenges ........................................... 1
1.2. What are we talking about? .......................................................... 3
1.3. Interaction and navigation at the heart of information retrieval ......................... 7
1.4. Why should we be interested in information retrieval? ...................................... 9
   1.4.1. Economy: maximize profitability and minimize risks. ......................... 10
   1.4.2. Information technology: mathematical concepts of the relevance of information .......... 12
   1.4.3. Robotics: improving movements and interactions. ..................................... 14

CHAPTER 2. CONCEPTUAL AND METHODOLOGICAL APPROACHES TO INFORMATION RETRIEVAL IN DIGITAL ENVIRONMENTS ........................................ 19

2.1. The approaches of information sciences: the precursors .................................... 19
2.2. The Marchionini sequential iterative model ................................................. 21
2.3. The holistic model of Kuhlthau .......................................................... 23
2.4. The first studies of psychology and cognitive ergonomics ............................... 26
2.5. The cyclic model of David, Song, Hayes and Fredin ................................................. 31
2.6. The skills-centered model of Brand-Gruwel ......................................................... 33
2.7. Kitajima’s predictive model ....................................................................................... 36
2.8. The hyper-specialized model of Sharit, Hernandez, Czaja and Pirolli .................. 39
2.9. The Landscape Model “diversion” by Dinet ............................................................ 42

CHAPTER 3. INFORMATION RETRIEVAL: PSYCHOERGONOMIC APPROACH ................................. 49

3.1. Introduction ............................................................................................................. 49
3.2. Identifying difficulties in modifying interfaces ...................................................... 51
  3.2.1. Hierarchical task analysis .................................................................................. 52
  3.2.2. Analysis of the end users’ behavior ................................................................. 55
  3.2.3. Implications for the (re)design of interfaces .................................................. 61
3.3. Anticipating the needs of users .............................................................................. 63
  3.3.1. “If we built it, they will come” ....................................................................... 64
  3.3.2. The analysis of users’ expectations and behaviors ............................................. 65
  3.3.3. Prospective ergonomics and technological innovation ..................................... 70
  3.3.4. Anticipating and understanding the needs of users: the method of staff made up of
         community experts ................................................................................................. 73
  3.3.5. An example of application of the method of staff made up of community experts ........................................................................ 76
3.4. The motor dimension ............................................................................................. 79
  3.4.1. Motor ability and information retrieval in digital environments .................... 79
  3.4.2. Toward a lexicon of intuitive gestures ............................................................. 85
3.5. The social dimension and collaborative ............................................................... 88
  3.5.1. From individual research to collaborative information retrieval ....................... 89
  3.5.2. Benefits and limitations of collaborative information retrieval ..................... 90
3.6. Impact of emotional ties between collaborators ..................................................... 92
  3.6.1. Ties between collaborators and impact on information retrieval ..................... 94
  3.6.2. “RCI-Web”: software to assist information retrieval ......................................... 97
3.7. The cultural dimension ........................................ 102
  3.7.1. About the importance of the home page ............ 102
  3.7.2. Culture and design of Websites’ home pages: an
        ergonomic inspection .................................... 105
  3.7.3. Information retrieval culture and behavior
        navigation .............................................. 107
3.8. The visual exploration strategies ......................... 109
  3.8.1. Impact of the typographical marking
         (bottom-up approach) ................................ 112
  3.8.2. Impact of the mental model
         (top-down approach) ................................ 117

CONCLUSION .................................................. 123

BIBLIOGRAPHY ............................................... 125

INDEX .......................................................... 165
1.1. Information retrieval, current and future challenges

Information retrieval is a central and essential activity. It is indeed difficult to find a human activity that does not need to retrieve information in an often increasingly digital environment: moving and navigating, learning, having fun, communicating, informing, making a decision, etc. Most human activities are intimately linked with our ability to search quickly and effectively for relevant information, the stakes are sometimes extremely important: passing an exam, voting, finding a job, remaining autonomous, being socially connected, developing a critical spirit or simply surviving.

From the psychological point of view, the activity of information retrieval presents several characteristics that make them both unique, complex and fascinating [DIN 12a]:

– Information retrieval in digital environments is a necessity in a growing number of human activities. Daily, we must search for information (an address, a phone number, a time slot, a name, etc.) in various digital environments of
constant evolution (GPS, mobile phone, Website, electronic terminal, etc.).

– Information retrieval is a composite activity, which typically requires reading, memorizing, writing, taking decisions, etc., these activities being cognitively complex and corresponding to areas of scientific research on their own.

– Information retrieval is a dynamic activity in the sense that the environment in which the activity is carried out evolves independently of the actions of the user (the same query in a search engine on the Web gives different results in an interval of a few seconds).

– Information retrieval is an iterative process, since the behavior and knowledge of the user are constantly changing at the pace of the information returned by the information system.

In 1998, the first Google index had 26 million Web pages. In 2008, Googles engineers counted a trillion Web pages (or a thousand billion). Since then, the number of pages has no longer been counted, and only the number of Internet sites is recognized. In June 2011, there were 346,004,403 Internet sites (source: Netcraft.com).

It is virtually impossible to know with accuracy the number of Internet sites and/or Web pages (sites protected by password, military sites, invisible Web, etc.). However, this is no matter, the only important data are the incredible progression of this mass of information, which everyone can easily access today. But, more than the mass of information, it is especially the relationship between humans and information, that has significantly evolved and is still evolving.
1.2. What are we talking about?

The term “information retrieval” (“recherche d’information” in French) includes both the scientific field and the related human activity. Some professional French-speaking organizations, such as the French association of information and documentation professionals (ADBS 2004) or even the AFNOR (French national organization for standardization 1993), proposed distinguishing information retrieval from information research:

– Information retrieval (recherche d’information in French) is the “totality of methods, procedures and techniques that allows, based on search criteria specific to the user, the selection of information in one or more repository of documents more or less structured” [AFN 93].

– Information research (recherche d’information in French) is the “totality of methods, procedures and techniques that aim at retrieving the relevant information from a document or a set of documents” [AFN 93].

The difference remains subtle and above all too superficial regarding the place of the end user in the activity. Because the English lexicon concerning the areas of computing and new technologies is more extended and more precise, it is logical to find more words (and therefore more accuracy) on topics related to these areas, such as information retrieval. In the English language, several specific terms coexist for mentioning the scientific fields related to information retrieval:

– Information retrieval: generally noted “IR”, these terms refer to the scientific field concerned with everything that falls under the search for documents or information, research in documents and whatever environments are considered (physical environment, off-line digital environment, such as CD-Roms, digital library, Internet, etc.). IR is a multidisciplinary field primarily involving disciplines such as computer science, mathematics,
information science and communication, psychology or even economics. Initially, one of the first objectives of researchers in the field of IR was the creation and development of technical systems and interfaces allowing quick access to relevant information in the increasingly large and more dynamic informational bodies.

– Information behavior: this refers to all of the human behavior related to all sources and channels of information (television, telephone, paper documents, Internet, face-to-face communication, etc.). In this case, human behaviors are not necessarily motivated by a need for information. For example, studies interested in the impact of exposure to advertisements (i.e. the “passive” consumption of information) are part of this scientific sub-domain.

– Information-seeking behavior: this refers to all of the human behaviors explicitly oriented toward obtaining information, regardless of the media (e.g. textbooks paper, magazines and Websites). In this case, a need for information is, therefore, at the basis of the studied behaviors, whether this need is conscious or not.

– Information-searching behavior: this refers to human behavior directly concerned with information retrieval, from a sensorimotor point of view (e.g. using the mouse or the keyboard) or intellectual (e.g. evaluation of the relevance of information).

– Information use behavior: this corresponds to human behavior involved in the processing and the assimilation of information found to knowledge already stored in memory. Here again, the sensorimotor aspects are sometimes studied (e.g. note-taking and annotation of texts) as well as the psychological aspects (e.g. memorization of information found).

If it is difficult to define with precision the exact shape of the different sub-areas related to IR, the task becomes even more difficult in the case of human–machine interactions (HMIs).
In order to define the different scientific sub-fields related to HMI, Wania, Atwood and McCain [WAN 06] have investigated the “genealogy” of different productions (articles and conferences proceedings). More specifically, these authors have studied the citations and co-citations in various publications based on the 64 authors most frequently quoted in HMI. After a review of bibliographic databases for the Institute of Information Science (ISI), Wania, Atwood and McCain [WAN 06] were able to identify the relationships existing between the articles; these relations take the form of citations and co-citations of the authors of reference. In addition, seven distinct groups could be identified in the studies conducted on HMI (Figure 1.1).
The two axes along which the seven sub-areas of HMI are divided and have been translated by [WAN 06] in the following manner:

– The horizontal axis distinguishes the studies according to the main center of interest, from technical systems to the human individual.

– The vertical axis distinguishes the studies according to the degree of involvement of end users in the design and/or evaluation process. For example, the work of Jacob Nielsen is located at the right end of the horizontal axis indicating clearly that the end user is the main factor of interest; however, since his work rarely results in experiments or empirical studies, this author is located in the lower half due to the involvement of individuals in his research.

This classification of studies in HMI according to these two axes allows us to address differently the conceptual and methodological frameworks in HMI. In fact, generally, research about HMI is presented by distinguishing works at the design level (i.e. conception) from works at the assessment level. However, it is difficult to distinguish studies according to this dichotomy, since most of the studies address the two dimensions (to various degrees of course).

Thus, not only does information retrieval cover a multitude of studies, sometimes very different, but works on HMI also cover a plurality of disparate studies. It is appropriate to add a third difficulty in the apprehension of scientific studies interested in human behavior in information retrieval in digital environments: the level of granularity chosen by the researcher. According to Fidel et al., [FID 04], the “size of the magnifying glass” used by the ergonomist researcher or practitioner determines its theoretical and methodological choice. And this “size of magnifying glass” is itself determined by the questions underlying the research or the intervention (Figure 1.2):
Impact of organizational factors? Strategies of visual exploration? Task analysis? etc.

Figure 1.2. The concentric approach of Fidel et al. [FID 04]

1.3. Interaction and navigation at the heart of information retrieval

As we will see in detail in the following chapters, the individual has been quickly (re)placed at the center of interests of scientific studies focusing on the activity of information retrieval (for a summary, see: [DIN 12]).

According to Kelly [KEL 07], most of these studies can be located on a continuum (Figure 1.3). At one end of this continuum, studies are mainly focused on systems and technical devices related to information retrieval; at the other end of this axis are found studies mainly focused on behaviors and human factors. According to the approach
considered, the theoretical frameworks and methodologies used differ.

![Figure 1.3. Scientific research on information retrieval according to the continuum proposed by Kelly [KEL 07]](image)

The significance of Kellys design [KEL 07] is to consider that the balance and/or break between these two extremes is the interaction: indeed, the human dimensions appear from the moment where the interaction becomes the center of interest (“IIR” for “Interaction Information Retrieval”). Some authors also propose a historical reading of the evolution of theoretical frameworks and methodological studies, for example, noting the evolution of topics discussed during important international scientific events [BEA 96, OVE 01].

The links between information retrieval and navigation are so narrow that the terms are often used interchangeably. From a formal point of view, navigation encompasses the whole range of techniques and procedures that allow, on the one hand, finding the position of a moving object in relation to a reference system or a fixed point and, on the other hand, calculating and measuring the route of a given point to another point. The moving object can be a human individual or an object, depending on the case.