Preface

Information Retrieval (IR) is the complex of activities performed by a computer system so as to retrieve from a collection of documents all and only the documents which contain information relevant to the user's information need. The peculiar difficulty of IR is the fact that relevance cannot be precisely and exhaustively described using data; for example, a relevant document cannot be precisely and exhaustively described using text even if this text includes or considers all elements or aspects of a topic; a user's information need cannot be precisely and exhaustively described using a query even if this query is a fully comprehensive description of the need. An unbridgeable gap between relevance and document content and between information need and user's request exists, such that what can be said about the document's relevance or the user's information need can only be inferred by the document content and the user's request (or all the other sources of evidence available to an IR system).

Information retrieval researchers are constantly searching for theoretical frameworks which help them make a major breakthrough in the overall effectiveness of retrieval systems. Since its advent, information retrieval has been marked by a series of significant theoretical advances. It is customary to correspond these advances with the logical models, the vector space models, and the probabilistic models. More recently, the machine learning models were added to this series of advances. All these models are now not only the main pillars of modern systems and applications, they are also the building blocks of the formulation of new models and techniques, thanks to the mathematical frameworks of logic, geometry, and probability on which they rely. Recently, a new perspective from theoretical physics was added to this series of advances.

Newton's laws provide a correct description of physical systems and in particular macroscopic systems (e.g., sand grains, balls, and planets) in terms of, for example, position, size, or distance. However, these laws must be revised when position, size, or distance is observed at the microscopic level, thus making the prediction provided by these laws uncertain. In the nineteenth century, the gap between observation and prediction gave rise to Quantum Mechanics (QM), which deals with the mathematical description of the motion and interaction of subatomic particles.

QM established the impossibility of measuring physical systems at the microscopic level with arbitrary precision, thus legitimizing the axiom that physical systems at the microscopic level cannot be precisely and exhaustively observed using any device and that any observation can always be subjected to a probability measure of the degree to which the observed value is real. Therefore, there is an unbridgeable gap other than the gap mentioned as regards IR; it lies between the unknowable world of subatomic particles and the outcomes produced by the devices used for describing this world.

The similarity between the barrier across the space separating relevance from content, or information need from request, and the barrier across the space separating the values observed through a device from the reality of subatomic particle was the reason why some researchers investigated the utilization in IR of the quantum mechanical framework. The fundamental idea underlying this utilization was the potential offered by the quantum mechanical framework to predict the values which can only be observed in conditions of uncertainty. However, this uncertainty is not only caused by the existence of different observable values (e.g., the number of spots on a dice), but it is also caused by the ignorance of the internal structure of the system under observation. The ignorance of the internal structure of the user's background, context, and plans of future actions causes an uncertainty different from the uncertainty caused by the variety of possible values that can be observed from the user. The first uncertainty is caused by the interaction between the measurement of the observable variables that describe the user's background, context, and plans and the user's background, context, and plans themselves. When these aspects are measured, not only some variable values are observed, but these values come to existence and evolve because of the measurement devices. The second uncertainty is "only" caused by the fact that there is more than one possible value that can be observed, and the actual outcome of a measurement depends on the randomness intrinsic in the measurement device; for example, the outcome of the draw of a dice depends on the randomness of the draw, and the uncertainty is caused by the existence of six distinct values.

In the first decade of this century, some studies investigated whether the quantum mechanical framework could be applied to research areas such as human cognition, natural language processing, and information retrieval other than theoretical physics. These studies and the results thereof can be regarded as an approach to investigating these research areas that might be named as "quantum inspired" or "quantum like." The idea behind the quantum-like approach to disciplines other than physics is that, although the quantum properties exhibited by particles such as photons cannot be exhibited by macroscopic objects, some phenomena can be described by the language or have some characteristics of the phenomena (e.g., superposition or entanglement) described by the quantum mechanical framework in physics. The other idea is that because modeling information processing may be a very daunting task, the designer of an information system such as an information retrieval system needs to be provided with theoretical frameworks that go beyond the traditional logic, geometric, or probabilistic frameworks.

The main thrust of this book is to illustrate how the quantum mechanical framework has been and may be applied to IR. The book is placed at the intersection between IR and QM which leverages the similarity between the barrier across the space separating relevance from content, or information need from request, and the barrier across the space separating the values observed through a device from the reality of subatomic particles. The book is not another proposal of retrieval model; in contrast, it aims to highlight the correspondences between the physical phenomena observed through the quantum mechanical framework (e.g., entanglement, superposition, and interference) and the phenomena which might be encountered during the retrieval process. The book is not about quantum phenomena in IR; in contrast, it aims to propose the use of the mathematical language of the quantum mechanical framework for describing the mode of operation of a retrieval system. The book cannot be an exhaustive description of the potential offered by the quantum mechanical framework; however, it aims to motivate researchers to delve into this framework and find further correspondences between IR and QM and design effective retrieval and indexing procedures which leverage the potential of the framework. To this end, the book is organized in four chapters.

Chapter 1 illustrates the main modeling approaches to information retrieval. This chapter is not an exhaustive illustration of the retrieval models proposed in decades of research; in contrast, it aims to highlight the main concepts which will be further addressed in the other two chapters of the book. In particular, the chapter illustrates Boolean logic, vector spaces, and two probabilistic models because these three concepts occur and are integrated in the quantum mechanical framework; it also briefly describes the machine learning-based approach to information retrieval since it has been a significant approach and may open further research directions within the intersection with the quantum mechanical framework. Although the first chapter is about information retrieval, it also introduces some notions of the quantum mechanical framework to build the bridge between the two fields.

Chapter 2 briefly explains the main concepts of the quantum mechanical framework; the literature is immense both because of the intrinsic complexity and the long time that has passed since the advent of the subject; therefore, we selected the concepts that may be linked to information retrieval; for example, superposition is explained in this chapter because it was investigated in some research reported in the third chapter. Some topics—entanglement is a glaring example—may appear rather difficult and sometimes really arduous. The book introduces them as gently as possible and suggests some further and more technical readings.

Chapter 3 surveys the research conducted in the intersection between information retrieval and the quantum mechanical framework. The chapter is organized in topics, and each topic is often ascribable to one research group or a few research groups. The literature on the intersection between information retrieval and the quantum mechanical framework has often been reported in journal or conference papers; only a small part has been reported as a book. Each chapter ends with a section suggesting the most relevant and interesting readings and listing the books and papers by which this book has been inspired and prepared.

Chapter 4 concludes with some suggestions for future research. Some research directions that are considered essential topics are briefly outlined in the hope that the quantum mechanical framework will be fully leveraged to achieve effective and efficient information retrieval systems.

The approach taken by this book is to organize the presentation of the main notions of the quantum mechanical framework (e.g., incompatibility, interference, superposition, and entanglement) in terms of observables and probability as it is customarily done in information retrieval in order to explain how they can inspire a new view and give rise to a new potential of information retrieval. Indeed, information retrieval systems are based on probability and observables in a similar way as incompatibility, interference, superposition, and entanglement are. The illustration based on probability and observables is not constrictive, since a large part of the literature of the quantum mechanical framework has this emphasis. On the contrary, this illustration leverages the long tradition in probability and statistics and naturally fits with many foundational problems of information retrieval.

This book is intended to be accessible to computer scientists in general and in particular to researchers working in information retrieval, database systems, and machine learning. The reader is expected to be a postgraduate, a PhD student, or a researcher who wants to have a clear picture in a short time of the potential of the quantum mechanical framework to pursue his own research interests. The mention of some crucial topics of information retrieval, the introduction to the main notions of the quantum mechanical framework, and the illustration of the way the quantum mechanical framework has inspired information retrieval are indeed intended to link the reader's background in information retrieval with the new notions he is acquiring while reading this book. The introduction of some crucial topics of information retrieval also aims to make this book accessible to computer scientists without a strong background in information retrieval and willing to address the use of the quantum mechanical framework in fields other than physics from the standpoint of an information scientist. Finally, the book may be of interest to noncomputer scientists who have utilized the quantum mechanical framework as an inspiration for their research carried out in domains other than physics.

The introduction of the quantum mechanical framework inevitably utilizes mathematical instruments; these instruments are based on the complex vector spaces which are the theoretical infrastructure that support the probability measures of the uncertainty occurring in QM and the logic underlying the observables applied to the physical world. The use of mathematical instruments may be problematic since they make the comprehension of the quantum mechanical framework harder, and, importantly, they might be used to build theoretical descriptions with no correspondence to the physical world; indeed, Polkinghorne (2002) and Zeilinger (2010) highlighted that some mathematical dissertations can lead to misleading or erroneous descriptions of the physical world. Such a mistake can be avoided only if the real world of IR is carefully examined.

As the topic described in this book has a relatively short history, the notation and the definitions used in the publications surveyed in this book may be inconsistent. Therefore, a common notation and glossary for notations and definitions are used in this book to make the understanding of the notions of the quantum mechanical framework less demanding than would otherwise be the case if these notions were studied using different publications.

Although almost every researcher acknowledges the importance of theory in information retrieval, they are reluctant to adopt further mathematical formalisms without clear evidence that this adoption brings about some actual experimental improvements. Because the use of the quantum mechanical framework in information retrieval is based on the mathematical formalism of the Hilbert spaces and has been debated for some years without reaching a consensus, the main expected benefit gained from this book is the clarification of whether, why, and when the use of the mathematical formalism and of the notions of the quantum mechanical framework can effectively be adopted for addressing some foundational issues of information retrieval.

Padova, Italy Massimo Melucci
June 2015

Springer

http://www.springer.com/978-3-662-48312-1

Introduction to Information Retrieval and Quantum Mechanics
Melucci, M.
2015, XVIII, 232 p., Hardcover
ISBN: 978-3-662-48312-1