Preface

We set sail on this new sea because there is knowledge to be gained.

John F. Kennedy (1917–1963)

As a general rule the most successful man in life
is the man who has the best information.

Benjamin Disraeli (1804–1881)

There is an ancient Greek legend about the mystic creature Sphinx who plagued the city of Thebes. The Sphinx would routinely ask the passerby's a riddle and whoever could not answer it would be snatched up and eaten by the beast. The hero Oedipus then went before the Sphinx and asked it to tell him the riddle, upon telling him Oedipus responded with the correct answer to the riddle. Hearing this, the Sphinx disappeared, bringing an end its threat to people.

Now people encounter a new mystic Sphinx that emerges from the technological development. This modern Sphinx asks the question what is information and how it functions. If people do not find a correct answer to this problem, technology threats to destroy humankind because people become more and more dependent on information technology. Really, the importance of information for society grows so fast that our time is called the "information age" (cf., for example, (Giuliano, 1983; Goguen, 1997; Mason, 1986; Crawford and Bray-Crawford, 1995; Stephanidis, 2000)). It is generally acknowledged that we have been living in the "information age", at least, since the middle of the 20th century. Information not only constitutes the very foundation of most industrial sectors, but more significantly has been now transformed into a primary tradable resource or commodity (Macgregor, 2005). In fact, there is no aspect of human experience that lies outside information influence.

For instance, Martin (1995) acknowledges that information functions as the lifeblood of society, writing:

Without an uninterrupted flow of the vital resource [information], society, as we know it, would quickly run into difficulties, with business and industry, education, leisure, travel, and communications, national and international affairs all vulnerable to disruption. In more advanced societies, this vulnerability is heightened by an increasing dependence on the enabling powers of information and communication technologies.

Information is all around us, as well as in us. Our senses collect and our brains filter, organize and process information every second of the day. Information makes our lives possible.

Since creation of the Internet, the volume of information available to people has grown exponentially. The World Wide Web, despite only being generally available since 1995, has thoroughly trampled all existing information media in its path to become one of the primary information delivery mechanisms (Duff, 2003).

An estimated 1-2 exabytes (1 exabyte is equal to 10^{18} bites) of new information is produced (and stored) each year. This includes all media of information storage: books, magazines, documents, the Internet, PCs, photographs, x-rays, TV, radio, music, CDs, DVDs, etc. It makes roughly 250 megabytes for every person on the planet. However, most of information is digital (93%, to be precise).

Printed materials are estimated as 240 terabytes (1 terabyte is equal to 10^{15} bites). This makes up less than a fraction of 1% of the total new information storage. However, there are about 7.5 billion office documents printed each year, as well as almost one million books, 40,000 journals, and 80,000 magazines.

At the same time, deficiencies with information quality impose significant costs on individual companies, organizations, and the whole economy, resulting in the estimated costs only to the US economy at $600 billion per year (cf., (Hill, 2004)).

During the last fifty years or so, the term *information* succeeded in pervading science, technology, humanities, economy, politics, as well as everyday life. Goverments, companies and some individuals spend vast fortunes to acquire necessary information. However, there are many problems with this term. As Goguen writes (1997), "we live in an "Age of Information," but it is an open scandal that there is no theory, nor even definition, of information that is both broad and precise enough to make such an assertion meaningful." Indeed, we are overwhelmed with myriad information from a wide spectrum of information sources, such as the World Wide Web, e-mails, images, speeches, documents, books, journals, etc. Moreover, as Loewenstein (1999) demonstrates, "information flow … is the prime mover of life – molecular information flowing in circles brings forth the organization we call 'organism' and maintains it against the ever-present disorganizing pressures in the physics universe." Each living being on the Earth has the structure determined by information encoded in its DNA.

At the same time, our experience demonstrates that mundane understanding of the notion of information may be very misleading. For instance, many identify information and message. However, receiving a message that consists of a random sequence of letters, a person gets no information. Thus, there are messages without information. Another example that demonstrates that a message and information are distinct essences is a situation when one and same message gives different information to different people. So, as it often happens in science, to overcome limitations of the commonplace image of information, the main problem is to find or to build the right theory.

Looking into information science, we encounter a peculiar situation. On the one hand, it has a lot of theories, a diversity of results, and even a proclaimed success. Scientists have created a bulk of information theories: Shannon's statistical information theory, semantic information theory, dynamic information theory, qualitative information theory, Marschak's economical information theory, utility information theory, Fisher's statistical information theory, algorithmic information theory and so on. Researchers study information ecology (cf., for example, (Davenport, 1997) and information economics (cf., for example, (Marschak, 1959; 1964; Arrow, 1984; Godin, 2008)), created information algebra (cf., for example, (Burgin, 1997b; Kohlas, 2003)), information geometry (cf., for example, (Amari and Nagaoka, 1985)), information logic (van Rijsbergen, 1986; 1989; Demri and Orlowska, 1999), information calculus (van Rijsbergen and Laimas, 1996), physics of information (cf., for example, (Stonier, 1990; Siegfried, 2000; Pattee, 2006)), and philosophy of information (cf., for example, (Herold, 2004)), but still do not know what information is. Each year, dozens of books and thousands of papers are published on problems of information. On the other hand, as it is written in the introduction to the authoritative book "Information Policy and Scientific Research" (1986), "*Our main problem is that we do not really know what information is*."

Moreover, due to the huge diversity of phenomena that are considered under the name "information", some researchers have come to the conclusion that it is impossible to have an all-encompassing definition of information, as well as to build a unified theory of information (cf., for example, (Capuro, Fleissner, and Hofkirchner, 1999) or (Melik-Gaikazyan, 1997)).

The situation of insufficient knowledge on information is really critical, because society more and more relies on information processing. Computers are becoming smaller, cheaper, interconnected and ubiquitous, while the impact of the Internet grows continuously. Computers solve a multitude of problems for people. Embedded devices control cars and planes, ships and spacecrafts. Furth (1994) writes:

"In 1991 companies for the first time spent more on computing and communications gear — the capital goods of the new era — than on industrial, mining, farm, and construction machines. Info tech is now as vital, and often as intangible, as the air we breathe, which is filled with radio waves. In an automobile you notice the $675 worth of steel around you, but not the $782 worth of microelectronics."

The question "What is information?" is singled out as one of the most profound and pervasive problems for computer science. As Wing writes (2008), it reflects the kind of far-reaching issues that drive day-to-day research and researchers toward understanding and expanding the frontiers of computing.

As we see, the concept of information is one of the greatest ideas ever created by the mind of human being, and our task is to make this concept exact in the whole its entirety. For instance, Stonier (1996) studies information as a basic property of the universe.

Information processes dominate society, permeating into every sphere of life, and a tiny amount of information can cause destruction of the whole Earth because pushing a small button can start a nuclear war. Von Baeyer states (2004), *information is the essence of science*, while contemporary society is based on modern technology, which in turn, is based on science. Thus, it is vital to understand information as a phenomenon, to know regularities of information processes, and to use this knowledge to the benefit of humankind. Stonier (1991) and other researchers stress, information science is badly in need of an information theory.

Numerous attempts to achieve this knowledge and to build such a theory are reflected in the literature where information is considered. There are many excellent, good and not so good books on information. Every year dozen of books appear in which some issues of information treated. So, suggesting to a reader one more book on information, it is necessary to explain why this book is different from the diversity of those books on information that have already been published.

The main goal of this book is not to demonstrate a series of sophisticated theorems and not to describe one or two (even the most popular) ways of understanding and modeling the phenomenon of information. Other books will do this. This book is aimed at finding ways for a synthesized understanding the information phenomena and at building a general theory of information encapsulating other existing information theories in one unified theoretical system. With this goal in mind, after presenting the general theory of information, the main attention in the book is paid to the most popular information theories with the solid mathematical core. This allows the reader to understand how the general theory of information unifies existing knowledge in information studies.

From this perspective, the book has a three-fold purpose. The first aspiration is to present the main mathematically based directions in information theory. The second objective is analysis and synthesis of existing directions into a unified system based on a new approach that is called the *general theory of information.* The third aspiration of the book is to explain how this synthesis opens new kinds of possibilities for information technology, information sciences, computer science, knowledge engineering, psychology, linguistics, social sciences, and education.

At the beginning of the book, contemporary situation in information studies and information processing is analyzed. Problems and paradoxes related to theoretical issues of information and information processes are discussed. To develop a valid and efficient theory, we need a thorough analysis of information processes in nature, technology, and society. We provide such an analysis in the first chapter of this book.

Elaborating on this analysis, as well on synthesis of the core ideas in existing information theories, we make a natural step to the general theory of information, foundations of which are presented in the second chapter of this book. This theory provides a base for systematizing other information theories and for unification of our knowledge about information, information systems, and information processes. It makes possible to overcome arguments of those who argue the impossibility of a unified information theory.

According to Arthur Clarke (1962), the only way of discovering the limits of the possible is to venture a little way past them into the impossible. In the general theory of information, this is done by means of a parametric approach to information with a specific system parameter called an infological system. This provides for a better understanding, as well as for more efficient theory development and practical applications. Discoveries of new, unknown before types of information are portrayed in the book. In addition, applications of the general theory of information to different fields are demonstrated in the second chapter. Relations between knowledge and information are found and analyzed from the point of view of knowledge management. The function of information in education is made clear, changing our understanding of educational processes. The role of information in the functioning of an individual is illuminated, demonstrating existence of new kinds of information.

Next chapters contain the most important and developed approaches and special theories of information, which are presented from a general perspective given by the general theory of information. An emphasis is made on mathematically oriented theories. We consider such theories as statistical information theory (including Shannon's theory and its later development), Fisher's approach to information and its application to physics, algorithmic information theory, and different versions of semantic information theory and of qualitative information theory. A wide spectrum of information theories is presented, demonstrating a lot of research that goes beyond the most popular direction called Shannon's information, or communication, theory. The general theory of information does not only allow one to unify this variety of theoretical directions, but also it helps to enlighten applications and development of theoretical results.

The primary audience for the book includes undergraduate and graduate students, information technology professionals, and researchers in information sciences, computer science, knowledge engineering, cognitive sciences, philosophy, artificial intelligence, psychology, linguistics, bioinformatics, social sciences, and education. Experts in information technology will be exposed to tools that information theory suggests for their professional activity: in designing software, building computers, applying computers to solving various problems, and developing global and local networks.

In addition, a reader with an interest in science and technology can find here an explanation of information essence and functioning, as well as answers to the following questions:

- how information is related to knowledge and data,
- how information is modeled by mathematical structures,
- how these models are used to better understand computers and the Internet, cognition and education, communication and computation.

It is necessary to inform potential readers who would prefer to understand the essence without dealing with mathematical formulas that the book gives this opportunity. In spite that it has many formulas, it is possible to skip these formulas and still get a lot of information about the information phenomena. The exposition is constructed in such a way that it allows the reader to achieve several levels of understanding. It is possible to read without formulas and to achieve the first level. Understanding of basic formulas brings the reader to the second level. Mathematical results formulated in the form of propositions and theorems elevate the reader to the third level of understanding.