Advanced Information and Knowledge Processing

Alfons Josef Schuster *Editor*

Understanding Information

From the Big Bang to Big Data

Springer

Advanced Information and Knowledge Processing

Information systems and intelligent knowledge processing are playing an increasing role in business, science and technology. Recently, advanced information systems have evolved to facilitate the co-evolution of human and information networks within communities. These advanced information systems use various paradigms including artificial intelligence, knowledge management, and neural science as well as conventional information processing paradigms. The aim of this series is to publish books on new designs and applications of advanced information and knowledge processing paradigms in areas including but not limited to aviation, business, security, education, engineering, health, management, and science. Books in the series should have a strong focus on information processing—preferably combined with, or extended by, new results from adjacent sciences. Proposals for research monographs, reference books, coherently integrated multi-author edited books, and handbooks will be considered for the series and each proposal will be reviewed by the Series Editors, with additional reviews from the editorial board and independent reviewers where appropriate. Titles published within the Advanced Information and Knowledge Processing series are included in Thomson Reuters' Book Citation Index.

More information about this series at http://www.springer.com/series/4738

Alfons Josef Schuster
Editor

Understanding Information

From the Big Bang to Big Data

Springer

Editor
Dr Alfons Josef Schuster
(Assistant Professor in Information Studies)
School of International Liberal Studies
Waseda University
4F, Building 11
1-6-1 Nishi-Waseda, 169-8050
Shinjuku-ku, Tokyo, Japan

ISSN 1610-3947 ISSN 2197-8441 (electronic)
Advanced Information and Knowledge Processing
ISBN 978-3-319-59089-9 ISBN 978-3-319-59090-5 (eBook)
DOI 10.1007/978-3-319-59090-5

Library of Congress Control Number: 2017946619

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To all informational agents – existing in all possible worlds.

Preface

There is an agreement across various domains about the impact of "information" on our modern-day life. Terms such as information, revolution or information society express this realization succinctly.

Simplistically, we may say that we are all involved in a kind of transition or race, in which a postindustrial society passes on its baton to information society. Although there is no single definition for concepts such as information society, it is possible to consider information society as a society where the state of well-being and advancement (on an individual as well as on a collective level) seems to depend on the efficient management of the so-called "life cycle" of information.

Essentially, the life cycle of information considers information as a product and involves fundamental information processes such as the acquisition of information and its storage, manipulation, retrieval, dissemination, or usage.

The relationship between information society and these fundamental information processes is extremely rich and versatile. For instance, the question of how information is generated could be divided into dimensions: material, biological, and mental. The material dimension may consider the processes that are responsible for the existence of our universe, while the biological and the mental dimensions may contemplate the production of biological systems and organisms or the processes of human creativity facilitated by a brain.

The motivation of this edited book is to understand the subject of information from a variety of perspectives. In order to generate this understanding, this book includes contributions ranging from cosmology, quantum physics, biology, neuroscience, computer science, and artificial intelligence to the Internet, big data, information society, and philosophy. Although each chapter provides its own domain-specific treatment of information, this edited book aims to synthesize these individual contributions, in order to generate an understanding that goes beyond the intuitive and often too casual conceptions that exist about one of the most important concepts of modern society and frontier science.

Tokyo, Japan
March 2017

Alfons Josef Schuster

Acknowledgements

I would like to take this opportunity to thank everyone who contributed to this edited book. I am grateful beyond words to the various authors for their wonderful and profound contributions. In times, where time sometimes can be a luxury, I am deeply touched and thankful for experiencing their understanding, dedication, and reliability. The same thankfulness needs to be extended to the various reviewers in this project. Their commitment, thoroughness, and positive feedback always led to an improvement of the quality of this edited book. My gratitude also goes out to the publisher Springer, in particular, to Helen Desmond, for her kind support, advice, and guidance throughout this project. Finally, I simply would like to say many thanks to you all again!

Contents

Contributors

Fred Adams Department of Linguistics and Cognitive Science, University of Delaware Newark, Newark, DE, USA

Fiona Browne School of Computing and Mathematics, Ulster University, Newtownabbey, Co. Antrim, UK

Iztok Fister Jr. Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

Karin Fister Faculty of Medicine, University of Maribor, Maribor, Slovenia

Valeriia Haberland Tungsten Centre for Intelligent Data Analytics, Goldsmiths, University of London, New Cross, London, UK

Satoshi Inaba Faculty of International Research and Education, School of International Liberal Studies, Waseda University, Shinjuku-ku, Tokyo, Japan

Tibor Koltay Faculty of Pedagogy, Institute of Learning Technologies, Eszterházy Károly University, Jászberény, Hungary

Gaye Lightbody School of Computing and Mathematics, Ulster University, Newtownabbey, Co. Antrim, UK

Jana Murovec Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

Katsuhiko Sano Department of Philosophy, Graduate School of Letters, Hokkaido University, Sapporo, Hokkaido, Japan

Naoyuki Sato School of Systems Information Science, Future University Hakodate, Hakodate, Hokkaido, Japan

Dr Alfons Josef Schuster (Assistant Professor in Information Studies), School of International Liberal Studies, Waseda University, Shinjuku-ku, Tokyo, Japan

Mai Sugimoto Faculty of Sociology, Kansai University, Suita, Osaka, Japan

Jürgen Vogel Engineering and Information Technology Department, Bern University of Applied Sciences, Biel, Bern, Switzerland

Andrew Whitaker School of Mathematics and Physics, Queen's University Belfast, Belfast, Northern Ireland, UK

Ko Yamada Faculty of Political Science and Economics, School of Political Science and Economics, Waseda University, Shinjuku-ku, Tokyo, Japan

Part I
Introduction

Chapter 1
From the Tannhäuser Gate to z8_GND_5296: A Day Trip on the Life-Cycle of Information

Alfons Josef Schuster

Abstract Modern-day computer power is a great servant for today's information hungry society. The increasing pervasiveness of such powerful machinery greatly influences fundamental information processes such as, for instance, the acquisition of information, its storage, manipulation, retrieval, dissemination, or its usage. Information society depends on these fundamental information processes in various ways. This chapter investigates the diverse and dynamic relationship between information society and the fundamental information processes just mentioned from a modern technology perspective.

1.1 Introduction

Information is an exciting and challenging subject. At the end of this introductory chapter, we hope that the reader will understand that to cover the subject of information within the space of a single chapter is impossible. The goal of this text, therefore, is to provide the reader with a general introduction to various aspects and concepts related to the field. One objective is to create a sense of understanding among readers that information is a concept that appears in a wide range of contexts, each with its own specific motivations, observations, interpretations, definitions, methods, technologies, and challenges. Forthcoming chapters in this edited book are going to explore many of these contexts and idiosyncrasies in more detail. In order to get a taste for the diversity of the subject, we begin our journey of information in the domain of entertainment – science fiction.

The replicant Roy Batty (played by Rutger Hauer) is a charismatic character in Sir Ridley Scotts 1982 classic science fiction film *Blade Runner*. Replicants are genetically engineered organic robots. To the naked eye, they are indistinguishable from humans. Manufactured with a predetermined lifespan, their engineered traits

Dr A.J. Schuster (✉)
(Assistant Professor in Information Studies), School of International Liberal Studies, Waseda University, 4F, Building 11, 1-6-1 Nishi-Waseda, 169-8050, Shinjuku-ku, Tokyo, Japan
e-mail: a.schuster@aoni.waseda.jp

A.J. Schuster (eds.), *Understanding Information*, Advanced Information and Knowledge Processing, DOI 10.1007/978-3-319-59090-5_1

allow them to be utilized for various, often dangerous tasks in off-world colonies. By law, the use of replicants on the Earth is prohibited. With this background, the film revolves around a group of replicants who, disobeying this law, have returned to the Earth to find out more about their existence and the possibility of extending their lifespans.

Blade runners are agents whose job is to hunt down and destroy such rogue replicants. As the story unfolds, the blade runner Rick Deckard (played by Harrison Ford) has hunted down and eliminated all but one replicant in the group – he is unable to overcome Roy. In the climax of the film Deckard and Roy engage in a battle of life and death. The battle ends with Deckard ultimately facing his death. However, in one of the key moments of the film, Deckard is saved from certain death by the replicant Roy, the very 'machine' Deckard was assigned to destroy. As Deckard's life is allowed to continue, Roy's life, with its preset lifespan approaching its end, begins to run out. In the terminating moments of his life, Roy produces the famous *Tears in Rain* soliloquy, one of the most moving monologues in cinema history. In this monologue, Roy reminiscences some of the powerful experiences he had in his life. He recounts attack ships on Orion, or C-beams glitter in the dark near a myth enshrouded 'Tannhäuser Gate'. In the moment of his dying, he is utterly sad and devastated that all these memories, all this information, all this knowledge, should be lost in time forever, like 'tears in rain'.

We have chosen the Blade Runner film as a backdrop to this chapter, because the theme of this classic science fiction centers around some of the key terms (e.g., information, intelligence, technology, or computerized societies) around which this chapter revolves. Nevertheless, it is time now to depart from Roy and the fictional location of the Tannhäuser Gate in order to, very briefly, touch upon the notion of information itself.

For a start, it is worthwhile to mention that it is surprising and very often overlooked that most investigations about information very quickly lead to a paradox. On the one hand, everybody seems to have a natural, intuitive, and immediate understanding about information. This understanding makes it relatively easy for any of us to appreciate, for instance, the important role that information holds as a driving-force behind our short-term actions and desires, as well as our more long-term ambitions and goals. On the other hand, there is the simple fact that, as soon as we look at information more closely, it turns out that information is a rather elusive concept and that it is extremely difficult to pin down, exactly, what information 'is'. It is important to understand, first of all, that there simply is no single, universally acknowledged definition of information. It may be helpful, however, to mention that the study of information could be divided into the following sections:

- One larger section deals with the philosophical side of information, and may involve discussions about the fundamental nature of information, how information may be generated, how it may become meaningful, or how information relates to moral issues such as responsibility or privacy (Wacks 2010).
- The content of the second larger section considers the more physical and technical side of information. This may include the study of different information

representation formats (e.g., as a molecular structure in a human genome, or as a string of bits transmitted over a telecommunication network). It may also include various aspects of information management (e.g., the processing of information, its storage, and maintenance), or the definition of a mathematical framework for the quantification of information (e.g., as in the mathematical theory of information introduced by Claude Elwood Shannon, who is often regarded to be the father of information theory (Roch 2010)).

- The third larger section relates to how information affects various types of agents in various types of information environments such as humans in an information society,[1] biological organisms in a biosphere, or artificial agents and biological agents inhabiting virtual reality worlds (e.g., Press and Williams 2010 or Kadushin 2011).

It is important here to understand that there is no reason to feel intimidated by the sheer complexity of these various dimensions of information. After all, it is a characteristic feature of our world that it is largely unknown to the human mind. Actually, throughout the course of history, the journey of mankind appears to be a perpetual quest in which we try to acquire a more complete understanding about this world, attempt to keep this acquired understanding safe, and make an effort to pass this understanding on from one generation to another.

In this chapter, we try to get a handle on information, by analyzing various sides of the information life-cycle at a time where powerful technology generates enormous amounts of data for a society that is equally information hungry as it is information dependent. In order to achieve this goal, Sect. 1.2 continues by generating an image about the restlessness of the human mind in its quest for universal understanding. Section 1.3 reminds us that the current interest in information is largely based on digital information. The section also describes information as a type of product that, from its conception to its usage, passes through the so-called life-cycle of information. Section 1.4 elaborates on the distinction between data, information, and knowledge. Section 1.5 returns to the information life-cycle. Section 1.6 philosophizes on information society, and Sect. 1.7 ends the chapter with a summary.

1.2 From Caveman to Spaceman

The death of Icarus, who came too close to the sun with his wings that his father Daedalus constructed from feathers and wax, has been interpreted across the categories of art in various ways. The mythological narrative in Ovid's

[1]There is no general definition for the phenomenon termed 'information society'. This chapter incorporates a view mentioned by Floridi (2010, pp. 3–18), who comments on information society as a society in which human progress and welfare seems to depend on the efficient management of the so-called 'life-cycle' of information. Section 1.3.1 provides further details on this life-cycle.

Metamorphoses (Ovid 2008, pp. 176–178) is one example, while another example could be the beautiful painting *The Lament for Icarus* by the English Classicist painter Herbert James Draper, who captures the tragic ending of the flight of Icarus by showing the dead body of Icarus surrounded by several curiously lamenting nymphs. Actually, even in our modern-day world the famous myth experiences the occasional revival (e.g., as in the 1983 album *Piece of Mind* of the heavy metal band *Iron Maiden*, which includes the song *Flight of Icarus*).

What is interesting in all of these representations is that it is usually the young and reckless figure of Icarus that captivates the reader, viewer, or listener, while the ingenuity and craftsmanship of his father Daedalus is often overlooked in this wondrous work of Greek mythology. There is another interesting point to this story, namely, how willingly we may accept the discrepancy between Icarus wings, true flying, and actually flying to the Sun. One reason for this willingness may come from a deeply rooted confidence in mankind that one has already completed an astonishing journey that took it from relatively primitive beings (in terms of the development and the utilization of technology) to the restless architects of artifacts of astonishing technical knowhow, sophistication, and beauty.

In this section, we would like to stimulate the imagination of the reader by visualizing this astonishing journey via the following three examples:

- A *cave painting* (e.g., the cave paintings in El Castillo Cave, discovered in 1903, in northern Spain (Pike et al. 2012)).
- The *Voyager 1* spacecraft (a space probe launched by NASA on September 5, 1977).[2]
- A so-called *Golden Record* fitted to the frame of Voyager 1.[2]

It is informative to compare these creations of the human mind from the perspectives of time, space, and intent. In terms of time, the distance between these artifacts accounts for a few thousand years. The cave paintings mentioned in Pike et al. (2012) date back[3] approximately 30,000 to 35,000 years ago, whereas the few Golden Records in existence and the Voyager 1 spacecraft have been produced only a few decades ago in the late 1970s. In terms of spacial distance, taking the Earth as a point of origin, the objects just mentioned convey the following information. The cave paintings mentioned in Pike et al. (2012) are in northern Spain, Europe, here on the planet Earth. For the Golden Record, the notion of the planet Earth may only evoke faint memories. One reason for this technological amnesia may

[2] Voyager the Interstellar Mission. http://voyager.jpl.nasa.gov/. Accessed: 2016-11-14.

[3] In terms of timing, it is necessary to highlight that accurate dating of cave paintings is a fairly difficult task. Until very recently, the cave paintings in Chauvet Cave in France have been considered to be the oldest examples of cave art in Europe. Recent discoveries in El Castillo Cave in northern Spain, however, suggest that the earliest examples of European cave art date back to as early as 48,000 years ago. From an information perspective such a finding is crucial. It may suggest that painting caves was not only part of the cultural repertoire of the first modern humans in Europe, but that Neanderthals, perhaps, also engaged in early human symbolic behavior (Pike et al. 2012).

be the fact that the record is fitted to the frame of the Voyager 1 spacecraft – the most distant human-made object in space. For example, at the time of this writing, this silent ambassador of mankind is journeying at a distance of more than 20 billion kilometers from the Earth.[2] Actually, on 12 September 2013, NASA announced that Voyager 1 has entered interstellar space (the space between the stars) to become a truly free interstellar space probe (Gurnett et al. 2013). Finally, regarding the issue of intent, it can only be subjective guesswork to figure out what those who left the cave paintings in El Castillo Cave behind had in mind. In the case of the Voyager 1 spacecraft and the Golden Record, all subjective guesswork can be firmly put aside. They are consciously crafted objects. The Golden Record contains sounds and images[2] that have been carefully selected in order to portray the diversity of life and culture on the Earth, while Voyager 1 stands out as our modern-day Icarus. In a forward-looking gaze, their intended audiences include future descendants of the human race, potential intelligent extraterrestrial life forms, or anybody or anything else who cares enough about crossing their path.

Intriguing as Voyager 1 and the Golden Records may be, they may convey different meanings to different minds. A philosopher, for instance, may see in them the deeply rooted desire of humanity to understand itself and the world at large, and to make itself understood to others and to the world at large, while an anthropologist, perhaps, may regard them simply as two more man-made artifacts created in the course of human evolution. By contrast, the minds of more flamboyant agents such as replicants or science fiction writers may be excited for other reasons. They may foresee journeys that may take mankind beyond places such as *z8_GND_529*, one of the most distant galaxies in our universe today, and a place where 700 million years after the Big Bang stars form more than 100 times faster than our Milky Way does (Finkelstein et al. 2013). Getting there, however, may require the accumulation of more and more data and information. Digital computers are designed to accomplish such a task.

1.3 Digits, Revolutions, and the Information Life-Cycle

The mathematician and philosopher Norbert Wiener (1894–1964) is widely regarded to be the originator of cybernetics. Crudely, cybernetics may be described as the study of feed-back systems. A feed-back system can be envisaged as a system that draws recourses from an environment, does some processing, and generates and feeds back some output into the environment. This output may be responsible for some changes in the environment, and, in doing so, may challenge the environment to some form of response. The system, which takes its input from a possibly changing environment, experiences a similar pressure. It too, needs to act in order to continue its operations. This process of control and communication repeats itself in a cyclic motion.

Control and communication are fundamental concepts in cybernetics. They are also non-trivial affairs. For instance, any feed-back system of sufficient complexity needs to be able to communicate with its environment (e.g., via sensors), but it also needs to be able to perform some kind of self-evaluation or self-assessment, which may require some form of a higher-order processing unit (e.g., a brain, or some other form of intelligent system). In addition, a crucial assumption in cybernetics is its indifference to the type of system. Cybernetics does not distinguish between biological systems (e.g., humans), and artificial systems (e.g., machines or software applications). It assumes that the mechanisms by which these systems operate are fundamentally the same.

Cybernetics also assumes that data, information, and information processing, play a central role in any communication system.[4] It is useful, therefore, to turn our attention to the fundamental processing unit behind modern-day telecommunication systems (and arguably the so-called digital/information revolution) – the binary digit, or simply the 'bit'.

A simple dissection of the term digital revolution reveals two facts. First, that there has been, or still is going on, or maybe is coming towards us, a 'revolution', and second, that there is a 'digital' element involved in this revolution. Even without any concrete examples, we understand the word revolution by meaning that there has been a fundamental structural change (e.g., physical, meta-physical, or behavioral) in a system previously existing in some pre-revolutionary mode of organization or structure. At a closer inspection, the word digital tells us that the digital revolution is not a revolution that is motivated by an ideology (e.g., like the French Revolution of 1789, which was inspired by the ideas of liberty, equality, and fraternity). Rather, the digital revolution is driven by powerful advances in technology. Actually, a more appropriate and encompassing term than technology alone would be the term ICTs, which stands for Information and Communications Technologies.

Please note the difference between the digital revolution and the information revolution. Although the two terms are intertwined, the digital revolution may refer to the rapid progress of computing and communication technology, which began during the late 1950s, while the information revolution involves the impact this development has (and may have) on human society (individually and collectively). In a sense, the digital revolution may be viewed as a facilitator of the information revolution. For instance, computer technology allows the operation of the Internet, while the higher-level services (typically accessible through the World Wide Web [WWW]) the Internet provides constitute various forms (e.g., print, audio, video) of informational content. This content, of course, has the potential to affect, for better or worse, the members of a modern society in various ways (Schuster and Berrar 2011).

[4] In addition, Wiener (1961, p. 132) also clearly comments on the unique character of information in his famous quote that: 'Information is information, not matter or energy'. Note that later in this section, we are going to pick up on this property of information again.

On a time-line the progress in computer technology may be divided into three phases. (Please note that this paragraph is based on Schuster (2016).) Although it is possible to mention many more names for each phase, the list of people who have been instrumental in the first phase may include Alan Mathison Turing (1912–1954), John von Neumann (1903–1957), and Claude Elwood Shannon (1916–2001). In simple terms, Turing specified the ultimate digital machine, the so-called 'universal Turing machine'. Surprisingly, the only difference between such a machine and any ordinary PC is that the former has infinite memory. This subtle difference, however, downgrades any modern-day supercomputer to the status of a mere wannabe.[5] The legacy of John von Neumann includes the description of a computer architecture (often referred to as the von Neumann architecture) that remains to this day a fundamental design feature of any modern computer. Shannon's contributions are similarly lasting. They include a mathematical framework for the treatment of information, as well as an investigation into the question whether the reliable transmission of information through 'any' information channel is possible in principle. The second generation of computing pioneers has been largely responsible for the development of powerful (commercial) computer systems and their underlying hardware and software (e.g., William Henry 'Bill' Gates III [born 1955], or the late Steven Paul 'Steve' Jobs [1955–2011]), as well as for the invention of inventions, the Internet and the WWW (e.g., Vinton Gray Cerf [born 1943], and Sir Timothy John Berners-Lee [born 1955]). The third phase of progress includes the founders of companies like *Google*, *Facebook*, *YouTube*, or *Twitter*, naming only a few. They all pursue their business under fancy-full names such as 'Web 2.0' or 'social web'. The social web is a very powerful abstraction. Not only does the social web urge information society to redefine traditional values, such as ownership (Heaven 2013), or friendship (Brent 2014), it is also a synonym for the seamless integration, augmentation, and infiltration of computing devices (of various degrees of intelligence) and the services they provide into our information hungry society as a new way of life.[6]

Let us now return to Norbert Wiener again. If, as Wiener[4] suggests, information is information, not matter or energy Wiener (1961, p. 132), then what is it? In order to understand this question better, it is useful to get a better idea about what

[5]At the time of this writing, the *Sunway TaihuLight* spearheads the list of the Top 500 Supercomputers at http://www.top500.org/. On this site, the performance of this supercomputer is given as 93 petaflop/s (quadrillions of calculations per second) on the Linpack benchmark.

[6]Quite naturally, this text can only provide a very short introduction to these fascinating topics. A reader with a wider interest in these topics, may find the following resources rewarding. The textbooks by Wright (2007) and Press and Williams (2010) describe the rise of the value of information through the ages, while the philosopher (Floridi 2011), one of the founders of the field of the 'philosophy of information', provides a sound treatment of the topic of information from a meta-physical, physical, and societal point of view. The books by Cohen (1996) or Jones and Jones (2000) may be useful for those readers with an interest in the theory of computing, or the mathematical theory of information, respectively. Our own work also provided various treatments of related concepts such as intelligent computing (Schuster 2007), or artificial intelligence (Schuster and Yamaguchi 2011), for instance.

type of revolution the digital revolution actually is. We already observed that the digital revolution is not ideology-driven, but its dynamic stems from a progress in technology. This observation makes the digital revolution similar to the Industrial Revolution starting in the late eighteenth century. If the steam engine represents the Industrial Revolution then its digital revolutionary twin is the digital computer. And, if the Industrial Revolution was responsible, among other things, for the mass production of goods that were physical, material, and tangible then the digital computer is responsible for the mass production of information, which is often non-physical, non-material, and non-tangible. At this point, some caution is needed. The text already mentioned that there simply is no universally agreed definition of information. If this is the case, how then can we speak of the mass-production of information? Although forthcoming sections are going to elaborate on this point in more detail, for the moment, we may say that the digital revolution mass-produces informational goods such as emails, YouTube videos, or complete virtual reality worlds. Perhaps, from a Platonic point of view, we may say that we are talking about instances of the idea of information, but that we do not really have a complete understanding about the idea of information itself, 'yet'.

For the sake of getting closer to the idea of information, let us put the notion of information aside for the time being, and instead, let us elaborate on the concept of an information life-cycle. We already mentioned that the Industrial Revolution provided an opportunity for the mass-production of tangible, material goods. While this is true, it soon turned out, however, that the production process itself (i.e., the process from product conception to the final product) became the object of careful study and analysis too. A main goal of this study was directed towards the establishment of general guidelines that would help manufacturers to produce any kind of product in an optimized and clearly defined way, taking into account a range of goals and objectives. To put it simply, the subject of 'project management' (including the study of product life-cycles) and its growing importance became widely recognized. As we shall see, the same concerns apply to non-tangible informational goods and the life-cycle of information too.

1.3.1 The Information Life-Cycle

Although the terminology employed differs occasionally, the following list captures the main processes involved in the so-called information life-cycle.

- Creation and production of information.
- Information gathering and collection.
- Organization, recording, and storage of information.
- Processing of information.
- Distribution and dissemination of information.
- Information application and usage.
- Information maintenance and recycling.