PHILIP TETLOW

ROUTLEDGE

UNDERSTANDING INFORMATION AND COMPUTATION

From Einstein to Web Science

A **Gower** Book

Understanding Information and Computation

For Jimmy, Pat and Don.
Special thanks to Keith and Ian.
You are all dearly missed.

Nature's grand book which stands continually open to our gaze, is written in the language of mathematics. Its characters are triangles, circles and other geometrical figures, without which it is humanly impossible to understand a single word of it; without these, one is wandering around in a dark labyrinth.

Galileo Galilei, 1623

Understanding Information and Computation

From Einstein to Web Science

PHILIP TETLOW

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2012 by Gower publishing

Published 2016 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN
711 Third Avenue, New York, NY 10017, USA

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing in Publication Data
Tetlow, Philip.
Understanding information and computation : from Einstein
to Web science.
1. World Wide Web. 2. Mathematics--History. 3. Physics--
History. 4. Information theory. 5. Information retrieval.
I. Title
004.6'78-dc23

ISBN: 978-1-4094-4039-0 (hbk)

Library of Congress Cataloging-in-Publication Data
Tetlow, Philip.
Understanding information and computation : from Einstein to Web science /
by Philip Tetlow.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4094-4039-0 (hardback)
1. Computational complexity. 2. Machine theory. 3. World Wide
Web. I. Title.
QA267.7.T47 2011
003'.54--dc23

2011052792

ISBN: 978-1-4094-4039-0 (hbk)

Contents

List of Figures

This page has been left blank intentionally

List of Tables

This page has been left blank intentionally

About the Author

Philip Tetlow is an Executive IT Architect in IBM's UK Business Analytics and Optimisation practice. He holds the world's first PhD in Web Science and was the author of the award-winning book *The Web's Awake* in 2007. He is a Chartered Engineer, a Fellow of the Institute of Engineering and Technology and a one-time member of the World Wide Web Consortium, having actively helped with the take-up of the Semantic Web.

This page has been left blank intentionally

Acknowledgements

First I must credit the unquestionably insightful work of Professor Stuart Kauffman and particularly thank him for the encouragement and inspiration with which he has followed my more recent work. I must next credit Professors Roger Penrose, Seth Lloyd, David Deutch, Lee Smolin, Ian Stewart, Robert Laughlin and Brian Green, all physicists and mathematicians of one kind or another. I deeply admire their work and their particular aptitude for lateral thinking. Large parts of this book are based on the insights they have brought to the world. I must also pay tribute to Google's Dominic Widdows, for if it were not for his brilliant book *Geometry and Meaning*, this book would still be three-quarters complete. Next I must thank Dr John Tait, Dr Chris Bowerman and Professor 'Keith' van Rijsbergen, for it is they who have taken time out to bring me back down to earth when the temptation was to linger amongst the clouds. Profound thanks are also a due to my long-suffering friends and colleagues Richard Hopkins, Kevin Jenkins, Brad Jones, Ian Nussey, Rick Robinson, Ian Charters, Mark Salthouse, Ian Devo, Karen Smith, Dav Bisessar, Tommy Giovannelli and D.J. McCloskey of IBM, Professors Wendy Hall and Nigel Shadbolt, Professor Yorick Wilks, Craig Gannon, Chris Brown, Alistair Miles, Tom Croucher, L.J. Rich, Dr David Hazelton, Bill Joynes and Graham Klyne, for they chose to stick around and suffer 'the speech' way past the point where reasonable individuals would have made their excuses and left. Next I must thank the various reviewers who tirelessly worked to polish and shine my final drafts. They include Alastair McCullough, Angus Cameron, Jim Zalles, Jonathan Moshinsky, Tess Moffett, Peter Lancett, Sukhvinder Phalora and Roman Podolczuk. Finally and deliberately I save the greatest debt of gratitude until last. It is with true humility that I must thank two inseparable brothers. They are my uncle, Dr Stanley Tetlow, and my dear father Dr Alan Tetlow. They have passed on more to me than I could ever pass on to others.

This page has been left blank intentionally

Foreword

By Yorick Wilks

The next time you crave to tweet that all-consuming idea or poke that newest of acquaintances, spare a thought for what you are about to do. The snippets of information you choose to send might be exactly the same as those used in a letter to a friend perhaps thirty years ago, but today things are quite different. Then your words would almost certainly have been received only by those you wrote to and would have turned into nothing more than cherished memories, even though good paper can last four hundred years, as computer people sometimes forget! But today those words have the potential to reach vast numbers of people and could easily remain in their original form and accessible to anyone well beyond your natural life span. Some people understand this and use it to good effect, but most do not. Andy Warhol's world of fifteen minutes of fame for all is very close to where we are today, yet still we have little idea what that means for us as individuals, communities or society as a whole.

This issue is not just about the changing ways in which we communicate. It is more about our hunger for information and our increasing ability to get it, to process and to share it. It is also about how recent creations, like the World Wide Web, have changed our way of life and how we apply such information constantly in that life. These changes have done so much for those privileged enough to have access to its supporting technologies – and that now means everyone with a high-spec mobile phone – but has the nature of information itself changed as a result? Indeed, have we ever understood information well enough even to ask this question, and should we care?

Information is absolutely central to everything we do and understand, and so, yes, we should care and care deeply. Science has tried many times and under many descriptions to get at the fundamental nature of information, yet still we only have a fragmented picture as that slippery word 'information' shifts from computing to linguistics to quantum physics and so on … . But this is not a matter for pessimism, just because there is no single view of what information is. We now understand that some of the most advanced fields of science rely heavily on the notion of information for their existence and that the tools they have developed to advance scientific thinking can be turned back on information itself, and I am thinking here of quantum information and the possibility that researches at the very smallest level of entities in physics may reveal a great deal about the nature of information. Think how Einstein's speculations about time in physics came to tell us so much about time itself in his theory of relativity. Some of the algorithms in use when we search the Web were first intended to explain the innards of the atom. All this suggests that we have to be brave and radical when thinking about information and not look only in the obvious places.

Enter Phil Tetlow. I was introduced to Phil some time ago through a mutual friend who had believed we might have something in common, and he was right. Since then our paths have crossed many times and we have shared many interesting discussions. This is what academics like me expect to happen when we meet new people, but Phil is not an academic and does not want to be one. He works for a large technology company and makes his living as a consultant. More precisely he is an IT Architect who designs and implements some of the largest and most complex information systems we have today. In essence he lives inside the field itself and is motivated quite differently from those who do pure research. What drives him is not the desire to find results in support of an idea, but rather the search for ideas and reasons behind the day-to-day occurrences he sees in large-scale systems. This often leads him to think in ways that would make most academics nervous. He is not so much interested in meticulous proof, but rather in new ways to unlock the doors in front of him. This is more an engineering perspective than that of a scientist and it has led him to be bold and outspoken. He spots a problem, speaks to those he knows to be experts about such things and then shares his thoughts with those around him in the hope that more ideas will be sparked off. This is the case here with his second book. In its writing I know he has tried out these ideas on world experts; and in many of these areas he is himself an expert. Those who initially created the Web and who are its guardians today are included on that list, but it does not stop there. He has gone on to seek out senior physicists and mathematicians and this has taken him well out of his comfort zone. Speak to him directly and he will be the first to tell you that he is neither a physicist nor a mathematician, and he will openly admit that what you are about to read could ultimately fail rigorous challenge in both areas, but that's not the point. What Phil has recognised, and recognised rightly, is that things are changing at great speed. If new ideas are not forthcoming now, we may well lose control of the Web. If his thinking is correct, then this may well have been a predictable conclusion all along. What you will find in these pages, therefore, is a courageous attempt to reshape the way we look at information and computing, a huge effort to take a look at it all from a fresh perspective. It is creative, novel and in many ways captures the thoughts that his contemporaries share, but have not had the time or conviction to write down. I respect him very much for that and I like this book for the same reason, and am eager to see it widely read.

Yorick Wilks
Professor of Artificial Intelligence (Emeritus) at the University of Sheffield, Senior Research Fellow at the Oxford Internet Institute, Senior Scientist at the Florida Institute for Human and Machine Cognition and British Computer Society Lovelace Medal winner 2010

This page has been left blank intentionally

Foreword

By L.J. Rich

'That's Napoleon's Clock on stage behind you,' they said, as I contemplated how on earth I was going to talk to a gathered audience of two hundred thought-leaders and business pioneers at IBM's 'Smarter Analytics' sustainability summit in London, June 2010. This was a tough crowd, easily as tough as some of the less savoury venues I'd performed at in one of my past lives as a gigging piano player.

I stood in the annexe of the Royal Palace the summit was being held at and hoped Phil was right. When I'd met him a few months earlier, I told him my idea for the keynote speech. I wanted to talk about the history of communication, statistically analyse the Eurovision Song Contest and play fake Bach to the gathered business dignitaries.

His eyebrows danced a little in the way that they do, and I imagined his brain as an elegant but eccentric machine, whirring and clicking – a sound not dissimilar to Bletchley Park's World War II code-breaking machines. Of course Phil loved the premise; it's in his mischievous nature to put a musical cat amongst the business pigeons and watch the resultant pecking and/or scratching. Of course, words like 'ecology, economy and society' sounded much more like elements of a keynote speech than my proposals, yet Phil could see that using a liberal arts view as a starting point would show that everything is connected to how we use, receive and translate information.

Take the Eurovision Song Contest – most people would agree it's as much about politics and geography as it is about music. But computer-generated authentic and emotional-sounding 'Bach' music, although obscure, was an extremely effective way of introducing the power of analytics and the issue of ethics – it demonstrated how we can be fooled by information – or enriched by it.

As it happened, the talk went down a storm – the fake Bach was a real hit, and the seemingly impossible task of trying something as obscure as music analysis to issues that big companies have to deal with on a global basis was indeed accomplished with aplomb and admiration. I was relieved when I took my bow and tried not to look too pleased with myself as I sauntered off stage. And I was euphoric! I had made an impact on influential and powerful people by grasping and shaping the eccentric and obscure, crystallizing those thoughts, and connecting them to real life in the hope that others could see my world. Of course, this was the sort of thing Phil has been doing for years – so it's no wonder we became friends.

At first, and, okay, at second and third glance, Phil Tetlow is easily as complex and as simple as some of the concepts he attempts to unravel in this book. His wry smile, infectious enthusiasm and improbable Yorkshire/Geordie brogue sit at odds with an undeniably planet-sized brain. His ideas and thoughts, like all the best ones, seem (to an amateur like me) to be finely but precariously balanced on that well-documented sword-edge of genius. Here, I thought, was someone who wasn't afraid of speaking his mind, even if not everyone could receive on his frequency.

Fast forward to summer 2011 – and, during one of our many lengthy and gloriously surreal conversations Phil asked me to give him my take on the Web, and integrate it into a foreword to his next book. I like to imagine the way I felt may compare somewhat to the feeling I'd get if Bach popped back from the eighteenth century and asked me to write him a concerto – I felt dis-concerted. Why (I asked) did he want me – a non-Web-science person – writing one of the two forewords to this, a highly specialised and Web-sciency book?

Phil would say that my credentials are sound: I was an early adopter, Webwise, having taken to email in 1993. Later, I had immersed myself in the social media scene during the 'Birth of Twitter'. I've also spent time as a TV producer for BBC Click, one of the globe's most widely distributed technology programmes, and was largely responsible for the growth behind its Twitter account, which had 1.85 million followers at the time of writing.

Phil also proclaimed that I'm 'relevant'. When asked to elaborate, he gruffly commented on my ability to spot next-generation trends and memes ahead of the curve – 'gravitational wells' as he called them. Hang on, I said, the Web has gravity? Turns out Phil's written another book all about that. At the time, I remember thinking 'Ah, of course he has,' as if anything else was preposterous.

A few weeks later, I felt ready to tell Phil how I saw the Web. Now, fully prepared for the eyebrow thing, I made sure he was sitting down and holding a pint of beer to steady himself. I took a breath and shared with him my viewpoint: the Web – as a mirror. A huge and flecked antique of a thing, scratched in places, shiny or even magnified in others. Its huge and uneven gaze held up to the offline society, a strange reflection of at-once darkest and purest moments in humanity. A reflection of the banal and the extraordinary. The minutiae of our everyday lives and lunches, searchable in the same breath as global crises and countless taxonomies. Vast swathes of information gathering momentum, our online avatars enriched by interactions with other mirrored people. I told him I thought the Web isn't good or bad, it just is. Like People, Music, the Sky, the Universe!

You can imagine where the eyebrows were at this point. And now the beer glass was empty. But I could see Phil's brain was full – there he was, processing my grand and excitable unscientific extrapolation with a few more internal clicks and whirrs, filtering and sifting the information to present a clear and relevant result like a good natured organic search engine. It's all information after all, I would imagine him saying, as zeros flipped over to ones and back again with pleasing wooden clunks. There's a place for that information, I just need to find where it fits, he would say, trying to find and hold the shape of the Web for us all to see. I like to think that Web science, or 'What Phil Does', is an attempt to chronicle and present these shapes to a wider society, so that we have a hope of understanding more about the Web, and perhaps even about ourselves.

So, to the book. For those intending to at least read the first few chapters, a friendly caveat: anyone who's flicked through the pages and felt a little uneasy at the generous sprinkling of 'science words' should relax. Read the whole paragraph through, and you'll absorb more than the words on the page – to my mind, a rich landscape of theoretical thought experiments appear – rather like those odd 'magic eye' pictures that don't make sense unless you're looking at them correctly.

Although the book touches upon gravitation, dark matter, quantum mechanics and geometry, it's the story of those areas, not the science. Strange?

Certainly. But to me, it does make some kind of sense. The unconventional re-purposing of words from the fields of physics and mathematics to attempt to tell the story of Web science sounds like a perfect way for Phil to crystallize those eccentric thoughts and connect them to our own way of seeing the world. Those of us who don't have the Tetlow CPU installed can at least access some of the files on our slightly obsolete but still functioning wetware.

So, make some tea, and prepare to explore the deep brain workings of someone who truly makes the Web work. Allow your mind to wander as it contemplates the eccentric insights contained within. These wild-eyed proto-visions of Web science in its infancy could yet lead to new theories and advancement in a field that is still being defined. Or this book could also be the utterly incomprehensible, yet fabulous, ramblings of a genius.

Indeed, quantum theory would dictate it is, er, both (at least until you read it). Either way, it's an unconventional and entertaining book that 'reads you back' – and, I think, a welcome and refreshing take on a subject that seems destined to be explored in great detail as we crash on through this century. We humans have created something unique in the Web, which is already used for extraordinary things. The book gives us a chance to look at this creation from another perspective. And if our imagination is up to the task, we could unlock even more of the incredible potential that the Web holds.

So, it merely remains for me to encourage you, dear reader, to go! Go forth into the unexplored, good luck and click wisely. Any further clunking and whirring you may hear will be your own.

L.J. Rich, London, September 2011
Manager, Perfect Pitch Productions Ltd.

Preface

I start where the last man left off.

Thomas A. Edison

When I started to write this book I singularly wanted to write about the World Wide Web. I really did. The Web is a safe zone for me and I knew that much still needed to be said about this fascinating sociotechnical system. Nevertheless the path of my investigations refused to run straight. The more deeply I read and the more I studied, the more I realised that the things I was interested in simply covered a much broader spectrum. Like a taut piece of elastic, the more I tried to pull the subject matter back towards a Web focus, the more resistance I felt. In the end I gave up and followed the path of least resistance.

In truth I was, and still am, interested in only one thing, and surprisingly that's not the Web at all. Rather I have an innate fascination for 'information' and all that it entails. Information is easily the most underestimated and poorly understood of things, for it plays an absolutely central role in everything we do and perceive. In many ways it literally is our world. Yet although it has been treasured and traded since the very dawn of our consciousness, only recently has science begun to realise its true worth. Because of such things I advise you to proceed with care as you read. I say this not to add dramatic effect, but merely to point out the nature of some of the material to come. Whilst working on this book I had known for some time just how poorly understood information is and also knew that many valuable insights lay outside the areas where most would normally look. Although much can be gleaned from fields such as information technology and computer science, in many ways they only scratch the surface of information's essence. To get truly close one must look further afield and probe in areas such as developmental biology, theoretical physics and advanced mathematics. This causes a significant problem, as such fields use and abuse their own definitions of information in inward-looking ways, rarely peering over the edge of their respective 'boxes' to share the best bits of their insight. At best this implies a lack of communal understanding and at worst it serves to underline that we simply do not have a sensible grasp on the subject matter yet.

In 2007 I was lucky enough to have my first book published. In *The Web's Awake* I made some rather far-reaching and provocative claims about the Web and the information held within it. For months after the book's launch I worried. I worried about being slated by the critics, about being ridiculed by the academics and about not getting the message across. In the end none of that came to pass, but that year of waiting and worrying taught me a lot, primarily showing that I had been overly nervous in writing the book. With hindsight I wish I had been bolder and this book is, in part, an attempt to address that. This time around I have tried to connect a number of ideas that might appear extreme at first sight and I hope you will accept the risk associated with this approach. Whenever and wherever within my reach I have tried hard to qualify the associations made and the conclusions reached, but in some cases there is simply no direct evidence available today to support the ideas presented. To some, I fully understand that this will be unsettling, but the aim of this book is very much to challenge convention. It is hence a book almost entirely about theory, framed by someone with many years' experience of large-scale IT architectures.

Although we might be intimately familiar with the information around us on a daily basis, that actually holds us back in many ways. Each morning we might rise and look in the mirror at the familiar features of our face. We might leave our homes through familiar doorways and walk along familiar paths to work, school or wherever. And all of this is obviously assisted by the presentation of information to us and our ability to absorb and understand it. Disappointingly though, such experience adds little to our understanding of what information actually is. Rather, such data-loaded interactions merely relate to the conveyance of individual information fragments, each linked into the complex mesh of knowledge and understanding that helps keep us in step with our own particular interpretation of reality – the onward march of our everyday life stories. We rarely stop to think about the nature of the information itself. But this we can indeed do, by reaching out to various abstract schemes of description, as Michael Schneider so eloquently describes:

> *Looking closely at nature, the first insight we obtain is that, behind the apparent proliferation of natural objects, there is a far lesser number of apparently fixed types. We see, for example, that through every generation cats are cats and are programmed for catlike behaviour. In the same way, every rose has the unique characteristics of a rose and every oak leaf is definitely an oak leaf. No two specimens of these are ever exactly the same, but each one is clearly the product of its formative*

type. If it were not so, if animals and plants simply inherited their progenitors' characteristics, the order of nature would soon dissolve into an infinite variety of creatures, undifferentiated by species and kinship.

This observation, of one type with innumerable products, gives rise to the old philosophical problem of the One and the Many. The problem is that, whereas the Many are visible and tangible and can be examined at leisure, the One is never seen or sensed, and its very existence is only inferred through the evident effect it has upon its products, the Many. Yet paradoxically, the One is more truly real than the Many in the visible world of nature all is flux. Everything is either being born or dying or moving between the two processes. Nothing ever achieves the goal of perfection or the state of equilibrium that would allow it to be described in essence. The phenomenon of nature said Plato, was always 'becoming,' never actually 'are.' Our five senses tell us that they are real, but the intellect judges differently, reasoning that the One, which is constant, creative, and ever the same, is more entitled to be called real than its ever-fluctuating products.[18]

In my personal search to fathom information's truths I have taken in some rather interesting and eclectic material. At first I was drawn to the well-established worlds of data processing and classical computer science. Then, in more recent times, I have drifted in and out of the Web world, spending time assisting with organisations like IBM and the W3C.[1] All of this has proved hugely rewarding, but, alas, none of it ultimately answered my innermost questions on information. Hence I gradually turned to the natural sciences for assistance and started to focus on how the very universe itself goes about the business of information processing.

Once one realises that the universe is nothing more than the biggest computer in existence – taking in information in the form of matter and transforming it over time – it soon becomes clear that it has to be the ultimate source of information's definition. Because of this I have tried hard to look across the full spectrum of natural information processing systems for answers. I have studied the way in which atoms interact and just how the weird world of the subatomic contributes; I have investigated how biological systems evolve, and have, in due course, become fascinated by how information can be considered in terms of Einstein's ideas on relativity, gravity and light. I have

1 World Wide Web Consortium.

even stopped off to contemplate just what insights entities like black holes, strings and superstrings might add, and, not surprisingly, the value-add has been significant. Indeed this is so much so that the works of physicists such as Stephen Hawking, Michael Green and Lee Smolin have changed my whole perspective on information's definition. Today I no longer think of information in terms of cold, abstract concepts. Today, for me at least, information is vibrant and alive in a very real sense. Now I actually have a problem making the distinction between physics and computing as a direct result of them both depending on information for their very existence. In the simplest of terms, physics is computing and computing is physics. What is important is that information is the bridge that links them both and in such regard its definition can, in many ways, be physicalised too.

Speak to any truly hard-core information engineers and they will talk about information as having dimensions and scale, levels and abstractions, viewpoints and perspectives. To them information is real, as real as you might be to me. It has form and subtlety, movement and strength, grace and poise – almost an energy and mass of its own. But most importantly of all, and without the need to reach out to analogy for explanation, it might even be intimately linked to the physical laws of the universe. Just as the atoms in our bodies are glued to our planet's surface by gravitational attraction, similar attractors also appear to influence information at large scales. The question is now not so much if such laws are relevant, but rather if they might be served better in terms of information and computation than the various equational variants served up by science thus far. Perhaps the universe is not physical at all? Perhaps it is just all just information? Perhaps movies like *The Matrix* might have been right all along and we are just living out our lives in one huge and magnificent dream, a moving hologram of some grander being's creation, shining in from the edge of the universe? Fiction surely? Fantasy certainly? Actually no, there is some very strong evidence to suggest that information is absolutely central to the universe's being.

Some things we know for certain already. We know, for instance, that quantum mechanics provides an extremely strong theoretic model on which to build our understandings of computation and information. The algorithm at the heart of Google's phenomenal success and unbelievable speed, for example, owes much more to the world of subatomic physics than might be expected. Furthermore there are some computations that just cannot be done efficiently without the leap of faith needed to allow randomisation as a key ingredient – the quintessential differentiating property that makes quantum mechanics systems stand out from their atomic-level cousins. We also know that to derive meaning from information we need an appreciation of those who are to use it and the uses to which it will be put. Thus not only is there a need to understand the information in flow itself, but we must also be sympathetic to the observation point of the user or consumer. Two individuals might interpret the same piece of information in totally different ways simply because of their own immediate contexts. This makes certain aspects of information relative, in much the same way that Einstein described our universe in the early years of the twentieth century. What is more intriguing, though, is that certain aspects of information are also invariant, making their presentation identical to all those who choose to investigate them. Intuitively this appears at odds with Einstein's ideas, but in fact it is totally consistent.

Such insights may appear radical to the point of being unbelievable, but that in itself is important. Our understandings of information need a good shake up. Over the past half-century we have made unbelievable advances in computing, but still we have not really moved forward in the way we think about the very stuff over which we compute. Progress was made, however, by Claude Shannon in his seminal work *A Mathematical Theory of Communication*[17] in 1948. This introduced information theory as a branch of applied mathematics and engineering and concentrated on the quantification of information. Shannon's information theory was developed to find fundamental limits on the compression and reliability of communicated data and since its publication it has been broadened to find applications in other areas such as statistical inference, networks other than communication networks as in neurobiology, the evolution and function of molecular codes, model selection in ecology, thermal physics, quantum computing, plagiarism detection and other forms of data analysis. The insights of Shannon and those who would follow have indeed proved pivotal, but there is an important point to remember here in that although information theory, as currently formulated, describes how to handle information – how to process or compute with it – it provides little in the way of a descriptive framework by which to understand its fundamental nature. Shannon never told us what information is. His work was only concerned with how much of it can be moved along a channel like a copper wire. The 'meaning' of the information concerned, or its semantics, was left out of his account[25].

There are others who have added valuable pieces to the jigsaw nonetheless, and in many cases they have not seen Shannon-like fame for their sizable contribution. Keith van Rijsbergen, for instance, has done some superb work on information retrieval and David Deutsch, Seth Lloyd and others have simply blazed a trail when it comes to understanding how information ties to fundamental physical models of the universe. But again, the picture is still not complete. Mathematicians like Andrei Kolmogorov have added further pieces by suggesting that the information in any given symbol string can be equated to the shortest program that can produce it in a 'universal computer'[25], but still the puzzle has bits missing. Some pieces may be within sight now but still the whole has not yet been made from its parts. Hence the need for a book like this. For right or wrong it attempts to piece together a whole picture, and, even if wrong, it still outlines the very strong need for others to attempt to do it better.

By the very nature of the material covered, this book is highly theoretical and academic. That said, I think it is important to mention that I am by no means an academic in the truest sense of the word. Rather I choose to make my living as a consultant, hopefully practising what I preach. I think that's important. I like to keep my sleeves rolled up and my hands dirty. But this pragmatic streak does have its setbacks. For instance, my preferred writing style is much less formal than that favoured by academics and I can see why such a style might be considered shallow by those who seek precision. Furthermore those who just want the basics may also be disappointed, as the engineer in me dearly wants to explain everything down to the last nut and bolt. I know this is a conflict, but I hope it is one you will allow. If I have done this right, I should have both enlightened and entertained by the time you have read the last chapter. Large parts of this book also cover some very deep physics and rather complex mathematical concepts by necessity, so I must also point out that I am neither physicist nor mathematician by training or practice. Rather I am an IT Architect who has spent most of his career working on very large IT systems, typically with very large data content. Even so I hope I have brought together physics, mathematics and information theory in an accurate and valuable way.

Because of the complex nature of the material and ideas covered, you may also notice some repetition. When I have read texts of similar length, depth and complexity I have been acutely worried that the detail would drag me away from the main themes involved. For that reason I reiterate several points at times. The intention behind this is solely to signpost the way clearly.

There is one big question that I must address before we can move on to the main content of the book; why concentrate on Einstein and Web Science in the subtitle? The answer is actually obvious. As you might expect with any reference to Einstein in a title, this text will, at least in part, discuss fundamental physics and explicitly look to Einstein's work on relativity. But rather than this discussion being based on observations and ideas centred on the physical world, as in Einstein's original works, we will focus on the semi-synthetic domain of the World Wide Web instead. This is surprisingly easy repositioning, as the Web is a hugely complex and highly connected dynamic information space, much like the physical world that Einstein pondered over so intently. There is a historical twist too, as when Sir Tim Berners-Lee was in the process of inventing the Web, he worked at the CERN[2] labs in Geneva, first in 1980 as a software consultant and then from 1984 to 1991 as part of a fellowship programme in the data acquisition and control group there. CERN is where the world's finest particle physicists congregate to do their best work on the most fundamental laws of the universe. It is also the place where the world's largest particle accelerator can be found, and it is this device that accelerates atoms to close to the speed of light only to smash them to smithereens through vicious and wilful acts of collision. The physicists do this not for fun but to investigate the tiny fragments of reality that such collisions spew out, and it is through their work that we are able to search for proof of some of the most abstract and wonderful ideas humankind has ever had. This prompts talk of laws, properties and particles with weird names and even stranger behaviours. To the physicists such particles spin, vibrate, attract and repel as if they have a life of their own, but to us they simply sound like imaginative thinking straight from the pages of a *Star Trek* script. But then that's the point. These scientists quite literally are on the edge of science fiction. They are the ones who point the way to our future. They are the select few who stand at the absolute limit of human understanding, and Tim Berners-Lee was right amongst them when he invented the system that would eventually become the World Wide Web.

During his time at CERN Berners-Lee was simply immersed in physics. He almost bathed in it. He would assist with ongoing experiments by hunting down various bits of information and collating them for use by his colleagues. He would sit at coffee tables and talk about weird and wonderful topics like string theory and standard models of this and that. So, for him, the world of cutting-edge science was very much part of his everyday life. A more important part of his life was concerned with applying the computers of the day to the task of modelling scientific ideas, and as a consequence he became frustrated with the way that the various electronic documents at CERN were scattered

2 The name CERN derives from the name of the international council (Counseil Européen la Recherche Nucléaire) which originally started the facility. The council no longer exists, and 'Nuclear' no longer describes the physics done there, so, while the name CERN has stuck, it is not regarded as an acronym.[19]

across the systems in use. This prompted him to think of ways in which this morass of information might be better structured and it was through this work that his genius was helped by those immediately around him. He already knew, for instance, that standard types of index, like the one toward the back of this book, are particularly ineffective when referencing information scattered across numerous disparate locations, but it was the quantum physicists who most probably reminded him that two things can be intimately connected regardless of the 'distance' between them. The physicists referred to this as quantum entanglement, and merely saw it an exquisite side effect of quantum mechanics. Berners-Lee saw it as something much more tangible and applied the concept to great practical effect.

As far back as the early 1980s he mused:

> *Suppose all information stored on computers everywhere were linked. Suppose I could program my computer to create a space in which everything could be linked to everything else.*[19]

And later that decade he finally crystallised his thoughts in formal proposal to CERN for what would eventually become the World Wide Web. In this he wrote:

> *An intriguing possibility given the large hypertext database with typed links, is that it allows some degree of automatic analysis … Imagine making a large three-dimensional model, with people represented by little spheres, and strings between people who have something in common at work.*
>
> *Now imagine picking up the structure and shaking it, until you make sense of the tangle: perhaps you see tightly knit groups in one some places, and in some places weak areas of communication spanned by only a few people. Perhaps a linked in formation will allow us to see the real structure of the organisation in which we work.*

Forget about the Web's eventual birth and the great brush fire of technological take-up that would follow. These two paragraphs are perhaps Berners-Lee's greatest contribution to science and to all of human history. Replace 'three dimensions' with 'any number of dimensions' and 'people' with 'subatomic particles', 'atoms', 'planets' or even 'entire galaxies' and another comparable model emerges. Replace the 'organisation in which we work' with 'the reality within which we all exist' and we get a far grander aspiration than the one ultimately realised in the World Wide Web. What we have actually done is summarise the mission statement of all of physics. We have defined the basic criteria for modelling the entire universe.

Perhaps this is telling us something? Perhaps the Web has more than passed the test as a model for all huge, highly connected information spaces? Perhaps the universe too is just one such mind-blowingly large information space? So, just as Berners-Lee might have been influenced by physics for his inspiration in the Web's design, might the physicists now also look to the World Wide Web as a model to help explain some of the more fundamental characteristics of the universe? The ties certainly appear strong and little research has yet been undertaken to either prove or disprove the proposition. Hence the book before you, and the fascinating set of possibilities it contains. It is not about the World Wide Web per se, but rather something far greater and grander. It also, hopefully, represents a story structured in such a way that will lead on without too steep a climb. Chapter 1 first tries to construct a contemporary view of the Web and explain how it might be analogous to some of the oldest and most stable theories in all of science. Next, in Chapter 2, we examine the problems of choosing a descriptive framework and go on to choose geometry as the language with which we will try to explain the ideas to follow. Chapter 3 then steps in to welcome the science that will accompany these ideas and leans on a historical perspective to assist with the process of introduction. Chapter 4 explains several weaknesses of classical geometry, again from a historical perspective. This gives way to Chapters 5 and 6, which construct a contemporary geometric framework better suited to our needs. Chapter 7 continues by specifically bringing physics centre stage and Chapter 8 stays with this theme by extending the geometric patterns first introduced in Chapters 5 and 6. Chapter 9 then underpins the discussion by explaining why such extensions might well be relevant in informational or computational systems of high scale and complexity. This next directs us into a brief run of chapters that look at reformulations of modern physics with a decidedly informational and computational flavour. That gives Chapters 12 and 13 their purpose. Chapter 14 again introduces more ideas based on physics, and Chapter 15 attempts to bring everything together into a single model. Chapter 16 duly tries to provide evidence in support and, lastly, Chapter 17 adds conclusions and final commentary.

Philip Tetlow, 19 September 2011

This page has been left blank intentionally

1

Introduction

> *If you want to build a ship, don't drum up the men to gather wood, divide the work, and give orders. Instead, teach them to yearn for the vast and endless sea.*
>
> *Antoine de Saint-Exupery*

One balmy summer's night, when the sky is clear and the smell of freshly cut grass hangs heavy in the warm still air, take time out, find a calm open space and gaze up at the stars. What do you see? As you take in the vista of shimmering dots that glisten in the blackness of space, surely you see beauty? Surely you see the majesty of the universe before you? If you do, then you are not alone. Ever since man could look up at the stars, others have marvelled at such beauty and some have even dared to question why it should be so. From such curiosity came the very essence of science itself.

The attention of physics is entirely focused on the universe and all it contains, as it tries to understand the very most fundamental workings of all we perceive as real. But physics is not an island. It is not self-sufficient when it comes to the tools needed to explain what it seeks to describe. In particular, physics carries a high reliance on mathematics and uses it as the predominant language through which it chooses to speak. So tight is this relationship that we can consider mathematics to be the bedrock on which physics is placed, the very foundation from which our deepest understandings of the universe are built. But there are still deeper foundations. Mathematics itself is based on the concept of individuality and the ability to group such individualities together into more useful concepts. From this the familiar concepts of numbers, arithmetic, geometry and algebra are created and today we put them to work with a high degree of success.

For most who care to consider such matters, that's it, dig down to the very base of mathematics and the roots go no further. Once it is understood where the basic building blocks come from, there is nothing below. But what if there was something below, something holding up numbers, all other mathematical concepts and all the various fields of science above that too? If that were the case then surely that something must be the real stuff from which the universe is made? Surely that something must represent the very signature of everything itself?

In recent times the ideas behind information and computation have seen their profile rise and with the advent of technologies like the Internet and the World Wide Web it's not hard to see why. Yet most still see information technology and its near relation computer science as resting on top of physics and mathematics. We even have fundamental models of computation that clearly line up with fundamental physical models such as quantum mechanics. But what if we chose, for some rebellious reason, to turn the whole model on its head? What if we chose to suggest that physics and mathematics were founded on the notions of computation and information; numbers, arithmetic, algebra, quantum mechanics … everything?

Most contemporary scientists would wince at the proposition of mathematics not being fundamental, but there are a growing few who would not. Ask your everyday scientist where information and computing should sit in the stack of scientific disciplines and they will most likely respond with confusion. Perhaps above sociology and economics, they might suggest, going on to say that entities like the Web are clearly propped up by such things. But there are those who think differently. To them the universe is just one gigantic computer system feeding on its own information and changing in ways we choose to consider as the reality around us.

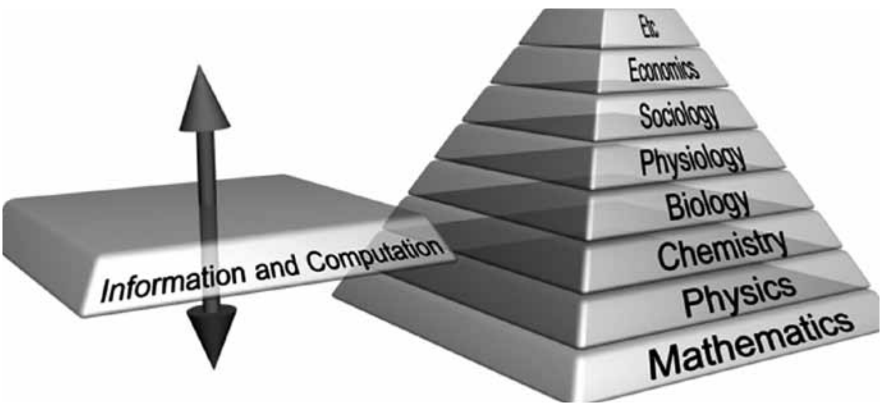

Figure 1.1 Where in science should information and computation sit?

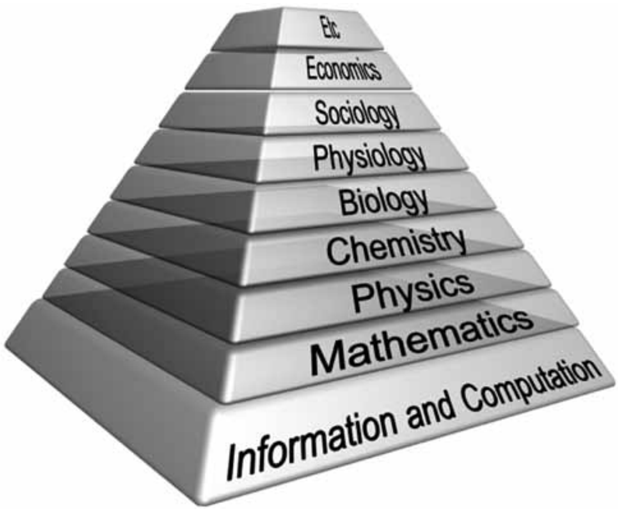

Figure 1.2 Information and computation as the bedrock of all other sciences

The World Wide Web is a truly remarkable innovation. For large sections of this planet's population it now touches our lives through a veritable explosion of change. Some influences are obvious, like the personal knowledge gained from basic Web browsing, but many are not so apparent. For instance we now see extreme cases far removed from the interactions we might traditionally consider as normal within our global society's fabric. These predominantly relate to the vast collection of autonomous Web software now chattering away in the background of our existence. Many refer to the components of this intertwined mesh as collaborating Web services, but this is really a generalisation that has become quickly outmoded. What is profoundly relevant, however, is that the world around these components has changed in recent times and a tipping point has been reached beyond which can be found an automated and intelligent environment hitherto beyond mankind's reach.

Why should this be so? The answer cannot so much be found with the software itself, or not at least if we are happy to consider such software one instance at a time. Rather it comes from the increasingly complex mesh of software-upon-software, computation-upon-computation and information-upon-information into which each individual component is now being placed. From this diverse mixture of connectivity an emergent property may now be starting to rise. This is the swarm intelligence of the Web; the common interpretation of its emergence is predominantly technical. But the Web is not wholly technical. The intention behind its inception may well have been so, but today it has evolved into a complex sociotechnical machine that is radically different. To characterise the modern Web as anything other than a global fusion of society, computation and information would be to do it an injustice. It is simply the largest human information construct in history. Furthermore the emphasis must be on 'machine' here, as evidence exists in support of the Web as a computational device in its own right, independent of the skeletal support donated by its underlying Internet. This changes the game for Web-based software as it acts like molecules in an overall system of much richer, more natural, design.

This string of analogies in connection with the Web is used for deliberate reason, as current research points to the relevance of thinking taken from the physical sciences. In particular the areas of quantum mechanics and relativity stand out as holding particular promise. This implies a number of unfamiliar consequences for those who wish to understand the next generation of the Web-like systems. It also offers great promise for those who work in the classical sciences. Not only is the Web the largest synthetic system humankind has every created, but it also provides the largest sample set of data in existence, outside the informational mass of the very universe itself. If this could be, or more likely when it is, analysed across its full breadth and depth, the chances are high that new types of complex geometries, patterns and trends will be found. The search will then be on to investigate if fundamental laws are at play in their formation and how these might relate to other fundamental laws already known.

It is already established fact, for instance, that the quantum model of computation has greatly strengthened our very understanding of what computation is. So it is plausible to suggest that thinking from physics' other great school of thought – that of the relativists – might also contribute in a similarly profound way. In fact, both physics and computing have already embraced the essence of relativity as a general underlying principle in many of their most fundamental models, the physicists commonly referring to it as 'background-independence'[14] and computer scientists favouring the term 'context-free'.[15] What Albert Einstein taught us was that at larger scales the differences between observable phenomena are not intrinsic to the phenomena but are due entirely to the necessity of describing the phenomena from the viewpoint of the observer.[6] Furthermore in the 1960s a different explanation of relativity was proposed, positing that the differences between unified phenomena were contingent, but not because of the viewpoint of a particular