INFORMATION SYSTEMS, WEB AND PERVASIVE COMPUTING SERIES

ADVANCES IN INFORMATION SYSTEMS SET

Volume 8

Artificial Intelligence and Big Data

The Birth of a New Intelligence

Fernando Iafrate

iSTE

WILEY

Copyrighted material

Advances in Information Systems Set

coordinated by
Camille Rosenthal-Sabroux

Volume 8

Artificial Intelligence and Big Data

The Birth of a New Intelligence

Fernando Iafrate

iSTE

WILEY

Copyrighted material

First published 2018 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

© ISTE Ltd 2018
The rights of Fernando Iafrate to be identified as the authors of this work have been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Control Number: 2017961949

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-78630-083-6

Copyrighted material

Contents

Copyrighted material

Copyrighted material

Copyrighted material

Copyrighted material

List of Figures

Copyrighted material

Copyrighted material

Copyrighted material

Copyrighted material

Preface

This book follows on from a previous book, *From Big Data to Smart Data* [IAF 15], for which the original French title contained a subtitle: "For a connected world". Today, we could add "without latency" to this title, as time has become the key word; it all revolves around acting faster and better than competitors in the digital environment, where information travels through the Internet at light speed.

Today more than ever before, time represents an "immaterial asset" with such a high added value (high-frequency trading operated by banks is an obvious example, I invite you to read Michael Lewis' book, Flash Boys: A Wall Street Revolt[1] [LEW 14]). It seems obvious that a large part of our decisions and subsequent actions (personal or professional) are dependent on the digital world (which mixes information and algorithms for processing this information); imagine spending a day without your laptop, smartphone or tablet, and you will see the extent to which we have organized our lives around this "Digital Intelligence". Although it does render us many services and

1 This book by Michael Lewis looks at the ins and outs of high-frequency trading (HFT): its history, means used, the stakes involved and so on.

Copyrighted material

increases our autonomy, it also accentuates our dependence and even addiction to these technologies (what a paradox!). This "new" world is structured around the Internet and requires companies to make decisions and act in a highly competitive environment, managing complex data in a matter of milliseconds (or less).

We live in a world where "customer experience" has become the key and our demand as consumers (for all types of goods, services or content: messaging, products, offers, information) is only growing. We demand to be "processed" in a relevant way, even as we navigate in this digital world "anonymously" (without formerly having used a personal authenticated account), which implies that other mechanisms must be in place to allow this "traceability". Who was it who said that "the habit does not make the monk"? I fear that in this digitized world, our clothes in the Internet are the traces we leave (navigation, cookies, IP address, download history, etc.), voluntarily or not, allowing a digital identity to be built without our knowledge and therefore being one that we barely or do not have any control over!

All this information is interconnected, joined together as they are being generated, following the "keyring" principle (see Figure 1). They are then exploited by targeting, segmenting and through recommendation engine solutions, which have been implemented over the last decade or so and are based on software agents backed by rule engines (recommendation engines). In order to meet a contact's expectation of "relevance", "a company does not own a customer but merely the time that he chooses to devote". During this time, which becomes the "grail" for companies to unveil vaults of imaginative ideas (but also much spending in terms of finances) to attract customers to their channels (website, call center, shops, etc.), they must be as "relevant" as possible.

Copyrighted material

The solutions currently in place (rule/recommendation engine) are not very interactive with their environment (they are predefined models based on a limited number of descriptive variables for the situation), they do not exhibit much self-learning (updating of models after analytical processing, which is often very arduous) and the result is that the same causes (identified by a few variables) trigger the same effects. These solutions do not or take very little account of context variations in real time (how a user arrived on a web page, what content they saw just before, what the nature of their search is, etc.), or do they consider results from previous decisions and actions. Last but not least, they barely or do not allow all contextual data to be exploited (navigation behavior, what was previously proposed in terms of content, the resulting actions, etc.).

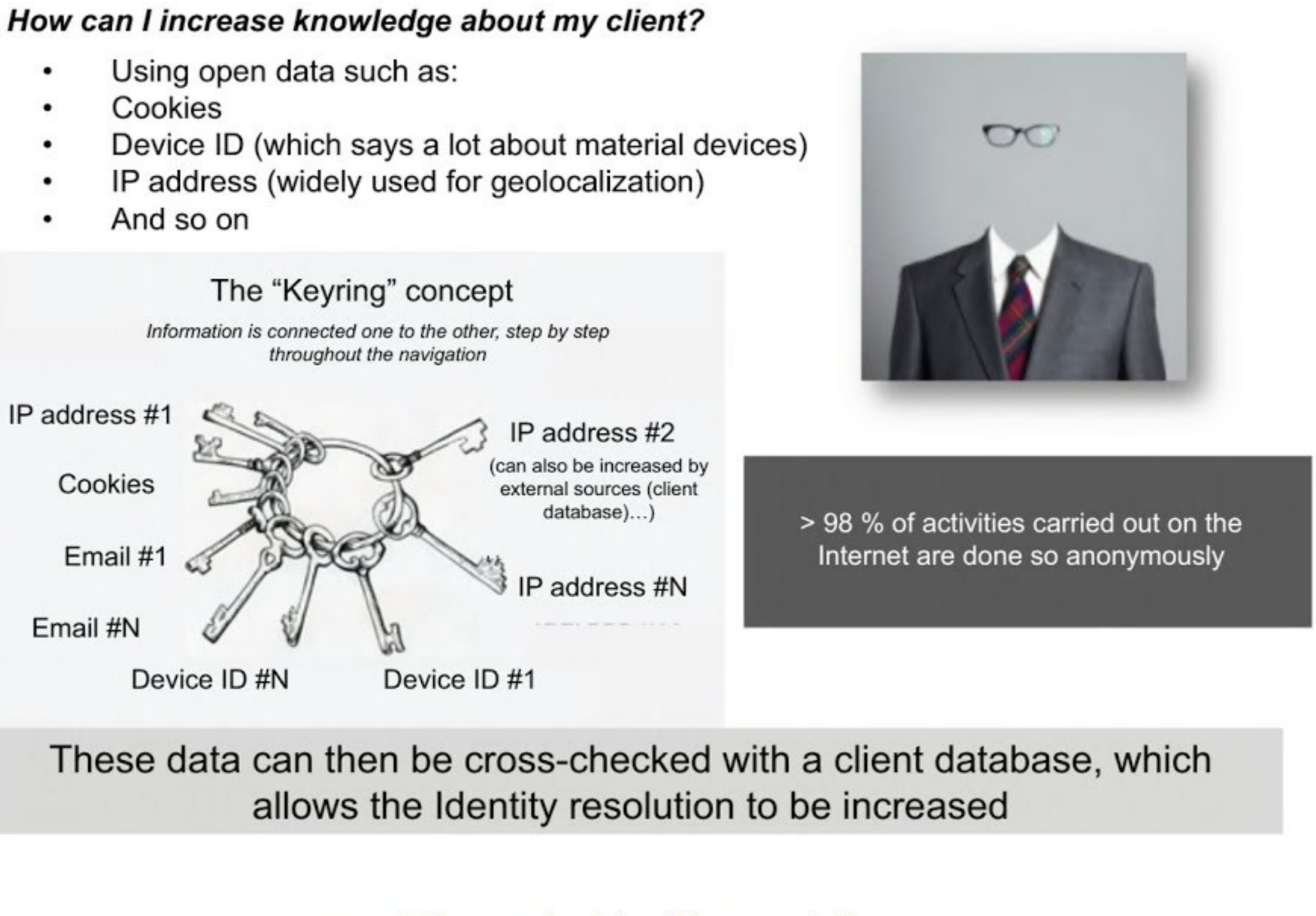

Figure 1. *Identity resolution*

This need to act and react in real time in a complex environment has been the case for years, and the advent of

Copyrighted material

scientists do not see any limits to one day being able to achieve machine intelligence (or an equivalent material element) in theory, a machine with a certain consciousness, one that could have emotions. This topic, as you may have read just before, is the subject of much debate. If today we do not yet have computers or robots that are as intelligent as humans, it is not due to a hardware problem but rather a problem of design. Therefore, we can consider that there is no functional limitation. In order to determine whether a machine can be considered as having strong AI or not, it must pass the Turing test.[3]

Weak AI consists of implementing increasingly autonomous, self-learning systems with algorithms that can solve problems of a certain class. But in this case, the machine acts as if it were intelligent, but it is more of a "simulation" of human intelligence based on learning (supervised or not). We can teach machines to recognize sounds and images from a database that represents the type of learning expected (such as recognizing a car in a batch of images, for example) – this is supervised learning. The machine can discover by itself the elements it analyzes, up to naming them. In the example of an image of a car, the machine analyzes the images that are proposed to it, and bit by bit (deep learning via neural networks) will learn by itself

---

3 From Wikipedia: "To demonstrate this approach, Turing proposes a test inspired by a party game, known as the "Imitation Game", in which a man and a woman go into separate rooms and guests try to tell them apart by writing a series of questions and reading the typewritten answers sent back. In this game, both the man and the woman aim to convince the guests that they are the other. […] Turing described his new version of the game as follows: We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"".

Copyrighted material

to associate the concept of car to the analyzed images and, when one of the associated images is labeled as a car, it will know how to "verbalize" it – this is non-supervised learning.

Fernando IAFRATE
December 2017

Copyrighted material

Copyrighted material