

SEARCH

How the Data Explosion
Makes Us Smarter

搜索

开启智能时代的新引擎

〔美〕斯特凡·韦茨◎著

(Stefan Weitz)

任颂华◎译

中信出版集团



SEARCH

How the Data Explosion Makes Us Smarter

搜索

开启智能时代的新引擎

【美】斯特凡·韦茨◎著
(Stefan Weitz)

任颂华◎译

图书在版编目 (CIP) 数据

搜索：开启智能时代的新引擎 / (美) 斯特凡·韦

茨著；任颂华译. --北京：中信出版社，2017.5

书名原文：Search: How the Data Explosion Makes

Us Smarter

ISBN 978-7-5086-7322-6

I. ①搜… II. ①斯… ②任… III. ①搜索引擎—应

用—电子信息产业 IV. ①F49

中国版本图书馆CIP数据核字 (2017) 第 041390 号

Search: How the Data Explosion Makes Us Smarter by Stefan Weitz

Copyright © 2014 by Stefan Weitz

First published by Bibliomotion, Inc., Brookline, Massachusetts, USA

This translation is published by arrangement with Bibliomotion, Inc. through Andrew Nurnberg Associates International Limited

Simplified Chinese translation copyright © 2017 by CITIC Press Corporation

ALL RIGHTS RESERVED

本书仅限中国大陆地区发行销售

搜索：开启智能时代的新引擎

著 者：[美]斯特凡·韦茨

译 者：任颂华

出版发行：中信出版集团股份有限公司

(北京市朝阳区惠新东街甲 4 号富盛大厦 2 座 邮编 100029)

承 印 者：北京通州皇家印刷厂

开 本：880mm×1230mm 1/32

版 次：2017 年 5 月第 1 版

京权图字：01-2014-8132

书 号：ISBN 978-7-5086-7322-6

定 价：45.00 元

印 张：8 字 数：159 千字

印 次：2017 年 5 月第 1 次印刷

广告经营许可证：京朝工商广字第 8087 号

版权所有·侵权必究

如有印刷、装订问题，本公司负责调换。

服务热线：400-600-8099

投稿邮箱：author@citicpub.com

这本书重新思考了我们每个人之间的关系以及我们与世间万物之间相互联系的本质。这本书没有讨论搜索引擎或者任何一家公司，而是旨在帮助你理解为什么商店很快就能知道你想要买什么，以及为什么政府即便在不明真相的情况下还能告诉我们事实。

——彼得·H.戴曼迪斯

医学博士，XPRIZE基金会主席，奇点大学创始人，《纽约时报》畅销书《富足：改变人类未来的4大力量》的作者之一

在这本书中，韦茨引领我们走向那令人目眩神迷的未来：“能力网络”提高了人类的智能，搜索变得更强大，而物理世界自身因为网络所诞生的智慧而光彩万分。

——亚当·奇耶

语言控制应用Siri合创者

蓝牙信标、增强现实和预测未来有什么相似之处呢？斯特凡·韦茨编织了一个颇具说服力的童话，向我们展示了他独一无二的才华，将这一纷繁复杂的主题提纲挈领地加以表述，使得我们这些普通人也能够理解。

——安迪·格里尼翁

iPhone 的联合发明者，现为 Eightly 公司的首席执行官

如果斯特凡·韦茨是对的（我也是如此坚信的），那么搜索技术就将引领我们这个社会进入一个全新的世界。如果你希望在未来科技领域出人头地，就读读这本书吧。

——亚力克斯·班纳言

美国风投公司 Alsop Louie Partners 最年轻的风险投资家，即将出版《第三道门：成功人士如何战胜系统并开创自己的职业生涯》

韦茨带领我们在未来的搜索领域进行了一次旋风式的旅程。这次旅行以现代科学为基础，但对我们一旦克服了如今的科技和商业限制后能拥有怎样的革命性结果进行了想象。对于在人机交互领域工作的你来说，本书必读。

——丹尼尔·敦克朗

领英“Query Understanding”（查询理解）技术总监，Endeca 搜索公司创始人

这本书不仅仅对搜索的未来进行了一次不可思议的探索，也让我们眼界大开，了解搜索和未来科技如何增强人类的能力，改变我们所做的一切。如果你希望对未来有所了解，就读读这本书吧。

——本·帕尔

获奖记者、作者、企业家、投资者

韦茨艺术性地描绘了搜索之无所不在的特性，以及其在改变我们生活和工作方式上不断增长的重要性。

——乔·马尔凯塞

True(x)公司创始人，Reserve.com联合创始人

本书抓住了重点，即数字化为什么扰乱了世界上所有的工业，连媒体、通信和信息也被扰乱了。接下去是谁？银行业、制造业、教育业、政府机构？读读这本书，了解一下数字化会怎样预测你的未来。

——蒂姆·德雷珀

德丰杰公司创始人

随着我们从搜索转向服务，世界将会发生怎样的改变？韦茨做出了很好的预测。让我们进入韦茨的思想，去领会搜索和客户行为将走向何处。这不仅很有趣，而且对任何技术专家来说都是其保持竞争力所必需的。

——吉娜·比安基尼

Ning and Mightybell公司创始人

本书对搜索目前走了多远以及未来还要走多远进行了重要的探索。我们如今提出的问题没什么解答不了的。但只有当机器不仅能回答问题而且能提出问题时，真正令人激动的搜索特性才会展现。韦茨这本书不仅让我们进行想象，而且让我们去理解这样的未来：计算机能使用海量的数据做出自主的选择。我个人觉得，我对韦茨描绘的未来感到兴奋莫名。

——戴夫·霍尔尼克

August Capital公司合伙人

韦茨对搜索的展望近在咫尺，几乎就是显而易见的，但也如此振聋发聩——我们需要他的帮助才能看清。如果你不想在工作中受到科技的冲击而措手不及，那就必须读读这本书。

——丹·夏皮罗

谷歌购物前首席执行官，购物网站Sparkbuy创始人，Robot Turtles桌面游戏发明人

本书可以激发你思考未来5年、10年乃至20年间搜索科技能为我们所做的一切。韦茨已经为未来编好了索引，就让我们读读他的成果吧！

——德罗尔·伯曼

Innovation Endeavors公司管理合伙人

一个简单搜索框背后的科技所释放出来的力量对我们的生活有着深刻的影响。它影响了我们如何购物，如何与朋友和同事交互。本书帮助我们理解这一技术冲击的好处以及后果：我们充满希望也面临重重挑战。它着重强调了在不久的将来我们不得不做出的那些重要决断。

——韦斯利·陈

谷歌工具栏产品经理，谷歌风投合伙人

在不远的将来，你的生活将得到运行在云端的、个性化搜索服务的帮助。它将追随你的每次行动，替你挖掘世界上的信息，在你提出问题之前就告知了问题的答案。韦茨像一位大师，描述了实现这一未来场景的挑战和机遇：搜索将扩展并映射物理世界，解释来自传感器和信号的海量数据，给我们的兴趣和环境建模。本书预见性地描绘了被技术辅助的生活，而我迫切地想置身于他描绘的那种生活中。

——帕特·金塞尔

Spindle联合创始人，现为Polaris VC公司的风投合伙人

一切都是从一本书开始的。

曾因深信不疑的假设受到了挑战，就开启了一场不可思议的对话。充满朝气的争论一场接着一场，在一代人中划分着朋友和对手。这一循环的结果要么是革命，要么是启蒙运动。回首望去，总有一本书值得被感谢或者被诘问。

本书要表述的观点需要大家共同来讨论，需要我们对外来的概念加以解构，因为它们与更大也更复杂的问题相关。我们要设计并构建出解决方案来处理现如今和未来的挑战，它们是促使我们进行更深入探索的催化剂，亦能弥合现实 and 理想之间的鸿沟。

... (faint text) ...

SEARCH

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

... (faint text) ...

可见的未来

机器人的想法已经产生多久了？我们又花了多少年等待智能机器为我们自动驾车、倾倒垃圾、照顾孩子，并在日常生活中听取我们的命令？自从人类开始梦想着扩展我们的能力，离开这个星球，甚至摆脱平凡乏味的生活，机器就一直是这一远大旅程中的核心。尽管看起来在地平线之外总还可能有另一个进步等待着我们并为我们带来这一自由，我们却突然之间——也是意料之外地——面对着一系列的创新。另外，这些创新的范围和影响是如此令人难以置信，它们总体所预示的东西也异常深刻：我们不仅会有更好的机器，而且会变成更好的自己。

我们花费了数千年建造机器和系统。它们最初是很原始的，慢慢地越来越复杂，用来帮助我们更清晰地思考，达到更多的成

就，并扩展我们人类的自然感官和能力。从最早的轮子到轮滑杠杆系统，再到石器、铁器，然后到计算尺，我们人类长久以来就能通过创造物件来增强我们的能力。这么做当然有充分的理由。一般的机器，特别是计算系统，在某些认知功能方面比我们强得多。这些方面包括：

机械计算。（本书写成时）世界上最快的超级计算机每秒钟可以处理 33 000 000 000 000 000 次计算。比较而言，里根总统在 20 世纪 80 年代推出“星球大战”计划时，被设计飘浮在空中、用来监视俄罗斯发射的导弹、以每小时 17 000 英里^①的速度跟踪并摧毁这些导弹的那个系统，据说也只需要一台每秒进行 6 000 万次计算的计算机。

重复地遵守复杂的规则和结构而不会有偏差或者失误。机器的优势在于，只要将一系列规则编好程序，它们就能单纯地执行任务。微软的 Windows XP 操作系统有 5 000 万行代码，它基本上就是系统应该如何作为的规则。计算机不会分心，也不会问问题。如果同样的问题重复出现，那么应用——而且是快速地应用——一个已知的可行方案将会事半功倍。

在海量数据中找出模式。我们不仅可以给机器编程使其遵循某一套规则，还能让机器在海量的数据中找到关联和关系。想象一下从数以亿计乃至数以兆计的信息中发

① 1 英里≈1.609 千米。——编者注

现某种趋势会有多难？虽然我们擅长视觉和模拟模式匹配（比如说在一幅画中看到众多相同的颜色），但数字模式不是我们的强项。

几乎瞬时地也几乎没有限制地保存并获取信息。毫无疑问，我们的大脑保存着庞大的数据量，而且我们保存数据的方式非常优雅，能找到机器通常看不到的联系。但是，我们也受制于有限的容量。更糟糕的是，我们回忆信息时，它的真实性会衰减，特别是时间一长或者这些信息很少被访问的情形下。而机器只要没有机械故障，就可以保存几乎无限量的数据，完美地重现这些数据数亿次，而且不论访问这些数据的频率有多么低，它们几乎都能瞬时获取。另外，不像人类需要7—10次的重复才能记住一个事实并将其置入长期记忆，机器只需要被告知一次就可以。

以有效且明晰的方式沟通。人类的沟通方式显示出不同层次的复杂度和含糊性（有人会说这就是人类的魅力所在），但系统所使用的符号语言却像精确的公式一样。我的女朋友也许会对她说她不想出门，但她的言语也许掩盖了她的真实想法。计算机不会出现这种情况。对于计算机来说，上就是上，下就是下，而假就是假。这样一种坚实的语言加上与之对应的真实判定告诉我们与该系统进行交流的另一个系统收到了该信息，这意味着分布式智慧可以（也确实）存在。

永远保持警觉，永远不需要休息。这一点也许是最明

显的了。只要供电良好（而且有效消除电子线路生成的热量），计算机从来就不需要关机。相比之下，这本书的作者现在就需要休息，而且不休息不行。

但这不是说系统已经就绪，可以代替我们这些“行尸走肉”了。还有一些任务和问题我们能处理得更好，而且在未来数十年乃至数百年间将一直如此。这是我们作为物种之一持续存在以及持续进化的积极信号。即使我们能在不久的将来为机器植入“思考”和“感受”的能力，我们相对缓慢、以自我为中心、带有偏见的人类大脑在某些情形下还是至高无上的。与机器相比，我们还是能胜出的。比如：

为我们从未碰到过的问题找到理性化的解决方法。当面对一种情形时，哪怕是我们之前从未遇到过与之完全一致的情形，人类有能力采用思维模型和进行启发式学习来找出方法。这是数百万年进化的结果。在问题中——哪怕它与我们之前见过的东西只有些微小的类似——我们可以找到模式，然后将这些模式应用到当前问题中去。举个简单的例子来说，十字路口的“行/停”标识就可以让一台机器不知所措：因为这台机器中设定的程序是读取文字，而交通标识用的是小红人或小绿人的符号。但对人来说，理解这些符号的含义实在太简单了。

学习但不用被告知如何学，自主学习相关知识或有趣的东西。一个人不需要明显的指导就能随着时间流逝而积

累起自己的知识。拿小孩子来说，他们不用接受专门的教导就能通过试错法学到最基本的东西：走路，认识爸爸妈妈等。

即便身处陌生环境也能理解并掌握该物理环境。你把某人带到一间他从未去过的旅馆房间里，他能很随意地在房内走动，找到浴室，打开咖啡机。这样灵活的操作是当今的机器还不能做到的。

同理心、创造力以及和其他人的关联。人性软弱的一面其实也是我们比机器更优越的那一面。在机器中植入的阅读人类情感、进行独立思考以及模仿人类交互的能力虽然有了发展，但还只是对实际情况的苍白模拟。

但让我激动的是，不要将人类和机器的能力视作非黑即白，而是二者相互弥补。我们要的是一个将机器最好的部分和人类最好的部分相连接的“铰链”。我们需要胼胝体（位于大脑中央的神经纤维束，连接着大脑的两个半球）的一个对应物。除了其他功能之外，胼胝体促进了半球间的交流，使得主管逻辑的左半球和主管情感的右半球协同工作，完成单个半球可能无法独立完成的一些任务。

我认为这一人机之间的“铰链”就是“搜索”，但不是我们如今所知的搜索，甚至也不是那些科技巨擘正在建立的搜索。只有我们不再将搜索看作一个页面操作工具，而更多地将搜索看作一系列的功能——这些功能可以使得我们更聪明、更快乐、更好

地与现实生活关联，这样的搜索才会脱颖而出。

到目前为止，搜索只解决了这一挑战中很少一部分。事实上，更可能的情形是，最早的那批搜索开发者甚至都不知道这个“铰链”才是他们试图解决的问题。他们最关心的是找到一个页面，而这个页面包含了在搜索框里输入的一个或多个词语。这么说不是贬低能基于几个词语在几毫秒内从数千亿文档中找到需要的文档的搜索科技。但是就如我们在后续几章会看到的那样，我们所处的世界日益数字化，这也就意味着所有的一切——从 Crate & Barrel 家具公司的一张吉普森咖啡桌的完整规格，到我在全球旅行的全部历史记录——都被计算机系统捕获并保存。伴随这一进步而来的应用将非常广泛。而我们看到搜索的影响同样深远，特别是当我们开始意识到搜索不是一个万能的工具，而每次查询不只是一个“钉子”时，它更应该是一系列的工具，有助于大幅度地扩展我们人类的潜力。

未来将会怎样？

我的日常工作是微软必应（Bing）的高级搜索总监。有很多员工致力于推广并提升我们的搜索引擎技术，我只是其中一员。我从一开始就在必应工作，也很幸运一直参与在我想象中最繁复的启动工作中。

2013 年 12 月，有人打电话问我是不是愿意写一本关于搜索的书。我的反应是当然可以，为什么不呢？当时我正计划环球旅

行，到 14 个城市去进行已经安排好的谈话和会议。我正在微软启动若干新的产品，也刚刚戒除咖啡因。很显然，我有点儿失衡。

每个设备、我们周围的每个物体、每个人都互相连接，我们也开发出从噪声中识别出模式的系统。面临这样的现实，写下我对此的观点很重要。搜索的力量在于提醒我们那些自己都不知道但其实应该知道的东西，也能在广义上让我们变得更好。我看到数字化的黄金时代正在来临，每个人、每个地方、每样东西都可以用数字的形式描述。一旦那个时代降临，而系统从我们复杂的思想中过滤掉那些知识残渣，那么人类的能力将得以增强，我们就可以做出更好的决定，还能在一念之间就操纵物理世界，而最终也会更加幸福。我在晚宴或者讲座上描述我的工作以及我对搜索未来的愿景时，人们无一例外地对这些可能性感到兴奋不已。当我谈到科技带来的超级链接、超级分析、超级指引到底意味着什么等深入内容时，更多的人才开始深思。

技术产业的圈子其实很小，我就有很多朋友在谷歌工作。而谷歌通常被认为是微软的主要竞争对手。虽然我对工作在山景城的同胞在做什么没有明确的信息，但在用晚餐的时候我会听到他们在思考什么，以及必应和谷歌都在试图解决的问题。

这本书接下来所讲述的并不是必应、微软或者任何其他公司建立这样的未来搜索产品的计划。事实上，书中很多概念是矛盾的，也涉及监管方面的擦边球。在某些情形中，社会就是不能接受一个超级互联的世界以及由此带来的透明度。我们这本书说的是未来的形势。而我每天深度生活在互联网之中，一年要和数

百家公司的工程师和技术专家以及那些推进着技术发展的学者交流。总体来说，我相信技术的力量能建立起一个更自由也更幸福的社会。

当人类与机器融合

我们人类总是想寻找方法捕捉我们在现实世界中看到的东西，从而将其保存、传承并从中学习。岩画、日记、散文、照片、录音、录像……所有这些踪迹使我们能织出一张生动又不朽的挂毯，其描绘的世界超越了我们个体生命的短暂本质，而且我们越来越多地用机械来保存、分析、解释这些踪迹。

与此同时，我们一直在尝试通过一些架构连接人机之间不兼容的界面，从而将两者结合。比如纽约中央火车站是一个汇聚了人类和机器的移动的中心。我们通过结合人类和机器的两个独一无二的特性——人类精细、自主但是缓慢的移动结合机器利用引擎和轨道进行的粗放、限定但快速的移动——让人们走得更快，走得更远。

融合了人类和机器各自的能力后会发生什么？如果我们用机器可以利用的方式捕获现实世界，然后利用机器的能力来大幅提升人类的能力会怎样？这就是搜索的未来，其独一无二的能力是对现实世界进行有意义的索引，并同时作为这些知识的通用接口。搜索就是我们一直想要的“铰链”。

我重申一次，这不是我们今天所认为的搜索，那只是一个