

《搜尋引擎沒告訴你的事》

图书基本信息

书名：《搜尋引擎沒告訴你的事》

13位ISBN编号：9789866723636

10位ISBN编号：9866723631

出版时间：2012-3-1

出版社：人智-人文化

作者：伊萊．帕理澤

页数：294

译者：宋瑛堂

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《搜尋引擎沒告訴你的事》

內容概要

Google新版隱私權條款上路

Facebook動態時報公測中

你，是否準備好迎接訊息全面個人化的時代？

和非洲死了一群人的新聞比較起來，你家門前死了一隻松鼠的消息更能引起你當下的注意。

—馬克·扎克伯格，臉書創辦人

我們曾經以為「在網路上沒人知道你是一隻狗」，

但現在網站不但知道你是狗，

知道你最愛吃哪一牌狗食，

還知道怎樣的廣告最能撩撥你的心。

我們曾經以為Web2.0終於讓每個人都能接觸不同的想法，

但是網站紛紛使用精心設計的演算法將訊息個人化，

你可能點閱許多文章，但卻全是同一個主題，

永遠只能聽到自己的回音。

我們的網路經驗正在劇烈改變，我們造訪的網站愈來愈積極調查你的個人喜好，並且利用這些喜好來調整網站餵給你的訊息，好讓你對網站愛不釋手，把裡面的文章一篇接一篇點下去。

許多人天天都會利用免費的超大容量的電子郵件、跟朋友保持連繫的社交網站，然而要使用這些方便的服務，就得交出你的個人資訊。其中也不少人非但不介意，還十分以此為樂，凡走過哪個景點，必留下打卡的痕跡。

而如果不找一個朋友坐在身邊上網，你很難察覺原來你螢幕上的Google或Yahoo新聞和別人居然有如此大的差異。個人化消蝕了共同討論的基礎，把這個問題搬上檯面來談，是刻不容緩的要務。

《搜尋引擎沒告訴你的事》

作者簡介

伊萊．帕理澤

網路民運組織MoveOn.org的前執行長，現任理事長，會員五百萬人，是網路政治的先鋒。他目前是羅斯福伉儷學會的資深研究員，也與人合創Avaaz.org，是全球規模最大的公民社團之一。他執筆的評論曾刊登在《華盛頓郵報》、《洛杉磯時報》、《華爾街日報》的民意論壇版。他為本書所作的演講可見於TED talk網

《搜尋引擎沒告訴你的事》

書籍目錄

第1章 衝高關聯度

約翰．厄文問題 / 點擊訊號 / 處處有臉書 / 資料市場

第2章 使用者即數位內容

閱聽大眾的興衰 / 新的中間人 / 大板 / 蘋果與阿富汗

第3章 注意力太集中的社會

微妙的平衡 / 注意力太集中的社會 / 探索新大陸 / 加州大島論

第4章 老是撞見你自己

亂猜你一通 / 直鑽你的弱點 / 深而窄的途徑 / 偶發事件與尋奇探險

第5章 公共事務靠邊站

雲端之主 / 友善世界症候群 / 看不見的選戰 / 區隔化 / 對話與民主

第6章 哈囉，世界！

機智王國 / 新型建築師 / 兩手策略 / 五百億美元的沙堡 / 你玩的是什麼遊戲？

第7章 你要的東西，全硬塞給你

能偵測同志傾向的機器人 / 未來早已來臨 / 理論的末日 / 天下沒有白吃的午餐 / 隨人蛻變的世界 / 喪失控制權

第8章 逃出小圈子

次文化的集錦 / 個人的對策 / 企業界能盡的心力 / 政府與民眾能盡的心力

章节摘录

第一章 衝高關聯度 一九九四年春天，尼克拉斯·尼葛洛龐帝（Nicholas Negroponte）坐在麻省理工的媒體實驗室裡，邊寫字邊思考著。他是媒體實驗室的創辦人，集結了年輕的晶片設計師、虛擬實境美術師、機器人設計師，一同潛心創建未來的電子產品與工具。然而，尼葛洛龐帝思索的問題比較單純，是廣大民眾天天自問的一句話：該看哪一臺的節目？ 在一九九〇年代中期，電視即時播放的頻道有幾百臺，全年無休，但多數節目不是慘不忍睹，就是無聊透頂：廚房用品的購物節目、當紅一曲歌王合唱團的MV、卡通、影劇新聞。對任何一位視聽人而言，有意思的資訊可能僅佔極小的比例。 隨著頻道數目增加，常用的瀏覽頻道方式是愈來愈難以負荷。只有五臺的時候，找節目很輕鬆，但頻道增加到五百臺時，挑節目就不容易了。頻道數激增到五千的時候，搜尋五臺的老方法根本不管用。 但是尼葛洛龐帝並不擔心。人類還沒有被頻道擊垮，應變之道即將出爐。「電視未來的關鍵在於，」他寫道，「不應該再把電視當成電視看待，」而是開始將電視視為有內建智慧的器具。消費者需要的是能自行操控的遙控器。這種自動化的助手有頭腦，能從視聽人收看的內容當中學習，能捕捉與視聽人相關的節目。「今天的電視機能讓你控制亮度、音量、頻道，」尼葛洛龐帝敲著鍵盤。「明天應該能讓你調整鹹濕度和政治傾向。」 只調整電視，不夠吧？在尼葛洛龐帝憧憬的未來世界裡，這一種智慧型的仲介是到處都有，能替人類解決電視頻道太多之類的難題。智慧型仲介猶如個人專屬的總管，守在門口，只讓你最愛看的節目和主題過關。「想像一下，」尼葛洛龐帝寫道，「未來你的介面仲介能閱讀全地球的通訊社、報紙的新聞，能為你捕捉所有廣電節目，進而為你訂做一份摘要。這一類的報紙只印一份...可稱為《我的日報》。」 他愈深入思考這問題，愈覺得有道理。數位時代的資訊氾濫成災，對應之道是聰明、個人化、內建的編輯程式。事實上，這種智慧型仲介不一定局限在電視——他向新科技雜誌《連線》總編建議，「電腦運算界將來明確的走向是智慧型仲介。」 在舊金山，傑容·藍尼爾（Jaron Lanier）以惶恐的態度回應他的論點。藍尼爾是虛擬實境的始祖之一，從一九八〇年代起，他一直致力於拉近電腦與人類之間的距離。在他看來，智慧型仲介的言論是荒誕不經。「你們的腦筋全秀斗了嗎？」他在個人網站上寫道，對象是《連線》雜誌型的讀者群。「『智慧型仲介』的概念既離譜又邪惡.....。智慧型仲介恐將決定（網際網路）會比電視好幾倍，或是爛到不行。」 藍尼爾深信，由於智慧型仲介終究不是真人，這種裝置會強迫人類以彆扭、失真的方式和仲介互動。「仲介會把你有興趣的事物歸納成一個近似卡通的模式，你透過仲介的眼睛看見的世界將會是一個卡通版的世界，」他寫道。 另外存在的一個問題是，仲介如果要做得盡善盡美，一定非過濾掉所有廣告不可。然而，由於網路商業的活水是廣告，這些公司似乎不可能推出這種殘害利潤的仲介來自戕。藍尼爾寫道，比較有可能的推論是，這種仲介效忠的對象不只一個，是可以買通的仲介。「它們效勞的對象並不明確。」 藍尼爾的論點明確而振聾發聵。可惜的是，這話雖然在網路新聞群組引發一些回響，網路初期的軟體大公司卻不為所動。這些公司信服的是尼葛洛龐帝的邏輯：能設法從數位乾草堆裡過濾出金塊的公司，一定能在未來勝出。這些公司看得出，隨著每人接觸到的資訊種類趨近於無限多，注意力當機的時代即將來臨。想賺錢的公司必須叫用戶專心。而在注意力貧乏的世界，讓消費者專心的上策是提供精準的資訊，最好能條條切中個人的癖好、慾望、需求。在矽谷公司的走廊和資料中心裡，當時的熱門關鍵詞是：關聯度（relevance）。 當時大家急著推出「智慧型」產品。在位於瑞德蒙（Redmond）的總部，微軟公佈了一種作業系統Bob，依據的全是智慧型仲介的概念，而搞笑的logo竟然神似比爾·蓋茲，令人不寒而慄。在庫博蒂諾（Cupertino），幾乎在蘋果推出iPhone整整十年之前，蘋果的牛頓全新上市，標榜的是「個人桌面助理」，主打的賣點是：智慧型仲介正潛伏在米黃色的機殼下為你賣命。 然而，智慧型新產品上市之後卻慘敗。網路聊天群組和電郵通訊團中，有一群人以嘲諷Bob為嗜好。消費者受不了Bob。《PC世界》雜誌將Bob列為史上最爛科技產品的前二十五名。蘋果的牛頓也好不到哪裡去：雖然蘋果耗資一億餘美元開發牛頓，上市半年的銷售量卻吃癩。用戶只需和一九九〇年代中期的智慧型仲介互動一下，就能認清真相：智慧型產品沒有那麼聰明。 如今，事隔十年的演變，智慧型仲介依然不見蹤影，令人覺得尼葛洛龐帝倡導的智慧型仲介革命失敗了。每天起床後，我們並沒有吩咐電子管家代為關照今日行程和喜好。 但這並不表示智慧型仲介不存在。它們只是被藏起來了。個人智慧型仲介潛藏在我們點閱的每一個網站底下。它們的智能是逐日增長，功能也愈來愈強，能針對我們的特質和興趣累積更多的資訊。誠如藍尼爾的預測，智慧型仲介不只效忠我們，也為Google等軟體大公司效勞，在提供資訊的同時散佈廣告。這些仲介雖然缺乏Bob那張卡通臉，我們的網路活動受它們操控的比

《搜尋引擎沒告訴你的事》

例卻愈來愈高。一九九五年，提供個人關聯度的競賽才剛開跑。把網際網路塑造成今日這種面貌的因素，或許正是各家競逐關聯度的趨勢。約翰·厄文（John Irving）問題 只要善用關聯度的效力，財源便能滾滾而來，最早領悟這道理的人之一是亞馬遜執行長傑夫·貝佐斯（Jeff Bezos）。從一九九四年起，他的願景是讓網路書店「回歸小書商的時代，店員對顧客的認識很深，會向顧客建議：『我知道你喜歡約翰·厄文的書。最近有個新作者出書，我覺得他的風格和約翰·厄文非常接近，』」貝佐斯對傳記作家說。說起來很容易，顧客一大票湧現時，店員怎麼一個個推薦？貝佐斯認為，亞馬遜需要成為「一種擅長人工智慧的小公司，」而公司運用的演算法能即時讓顧客和圖書配對。在一九九四年，資訊新秀貝佐斯在華爾街上班，一家創投公司想在蓬勃發展的網路尋求商機，請貝佐斯探尋做生意的點子。貝佐斯做事講求方法，為創投團隊列舉二十種理論上能在網路上賣的商品，例如音樂、服飾、電子產品，也逐一鑽研這些產業的奧秘。一開始，圖書事業在他的名單上吊車尾，但他最後彙整資料時卻發現，開書店是二十行業當中最適合上網的一種。書籍合適在網路上銷售的原因有以下幾個。第一，出版業的權力不集中，最大的書商蘭登只掌握一成的市場。如果有哪一家出版社不願和他合作，出版社多的是，他另找願意合作的對象並不難。原因之二是，傳統書店以外的圖書銷售量已經佔大多數。和其他商品比較起來，顧客上網買書的適應期比較短。此外，不同於服飾的是，買書不需要試穿試戴。然而，網路書店的點子之所以誘人，最主要的原因是書太多了，一九九四年市面流通的書籍高達三百萬種，讓為數三十萬的CD瞠乎其後。這麼多的書，單一的實體書店絕對無法胃納，網路書店卻有辦法全部上架。貝佐斯向上司提出這項研究心得時，創投卻興趣缺缺，認為賣書在資訊時代顯得好落伍。但這份商機一直在貝佐斯腦中縈繞。如果上網開一家書店，能上架的書籍是無限多，而且能提供顧客更親近、更貼心的經驗，這些優點讓邊境（Borders）或邦諾（Barnes & Noble）等連鎖大書店難以望其項背。貝佐斯把亞馬遜的目標訂在強化發掘佳作的過程。他想創造的是一間個人化的書店，能協助讀者發現好書，能將好書介紹給讀者。問題是，如何達成這個目標？貝佐斯開始從機器學習（machine learning）的角度來思考。機器學習是艱深的學問，但從一九五〇年代開始，麻省理工和加大柏克萊等研究機構的工程師和科學家已開始鑽研。他們將這領域的學問稱為「控制論」（cybernetics）。這單字的發明人是柏拉圖，原意是一個能自我制衡、調節的體系，民主便是一例。對早期的控制論學家而言，最振奮人心的事莫過於創造能根據意見來自自我調整的系統。接下來數十年間，控制論學家研究不少出演算法和理論，為亞馬遜的成長奠定基礎。

一九九〇年，全錄的帕洛亞托研究中心（PARC）的一組研究員活用控制論思考，用來解決一種新問題。PARC的創舉常被外界推廣上市，例如圖形使用者介面（GUI）和滑鼠就是PARC的創舉。和當年走在時代尖端的科技人一樣，PARC的研究員也是最早使用電郵的一群人，收發的電郵數以百計。電郵很實用，沒錯，但缺點很快就浮現了。只寫一封信，可以向無限多的對象發送，而且費用是零，不用過多久，無用的訊息會如潮水湧現，會把用戶淹得無法呼吸。為避免被上漲的洪流淹沒，PARC團隊開始研發一種名為協同過濾（collaborative filtering）的程序。他們將這種程序命名為織錦（Tapestry）。織錦程式能觀察用戶應付大批電郵的行為，看用戶打開哪些電郵，回覆哪幾封，刪除的電郵有哪些，然後運用這些資訊來調整收件匣裡的電郵順序。用戶時常點閱的電郵會被挪到上面，經常被刪除或不點閱的電郵則被壓到底下。本質上，協同過濾能節省時間，用戶不必在成堆的電郵中選閱，可以依賴程式在收件匣事先篩選。織錦程式這麼好用，當然不只能應用在電郵上。開發織錦的研究員寫道：「本程式能應付友方傳來的任何一種電子文件。電子郵件只是一例。通訊社的新聞和網際網路新聞也屬於這一類。」織錦程式將協同過濾介紹給全世界，但在一九九〇年，世人對它不太感興趣。當年的網民只有幾百萬人，生態體系仍小，資訊沒有多到無法分類，頻寬也沒有大到無所不下載的程度。因此，多年來，協同過濾的使用者只有軟體研究人員和閒著沒事做的大學生。在一九九四年，如果你列出幾張你愛聽的專輯，寄給ringo@media.mit.edu，系統會回信建議相關作品和樂評。根據該網站的說法，「每隔一小時，伺服器會處理所有來信，必要時發出回信。」這系統是串流音樂網Pandora理念的始祖，是窄頻時代的個人化音樂服務。一九九五年亞馬遜開張之後，隨之而起的是一片新氣象。打從一開始，亞馬遜就將個人化服務內建於書店，一面觀察用戶購買哪些書，一面利用PARC發明的協同過濾法，即時提出建議書單。（「你想買《無師自通西洋劍指南》？要不要搭配一本《醒來全盲：眼傷訴訟大全》？」）追蹤顧客購買行為一段時間後，亞馬遜可以歸納出哪些用戶的品味相近。（「書性和你相似的讀者買了本週新書《西洋劍就位！》」）向亞馬遜買書的人愈多，個人化的段數就愈高。一九九七年，亞馬遜書店的顧客總數突破一百萬。六個月之後，顧客數目成長至兩百萬。到了二〇〇一年，亞馬遜首度發表淨營收季報，成為最早證明網路金礦豐沛的公

《搜尋引擎沒告訴你的事》

司之一。亞馬遜能否複製街角小書店的人情味，見仁見智，但個人化程式卻相當有效。究竟個人化為亞馬遜帶來多少營收，亞馬遜的主管是守口如瓶，但主管在說明亞馬遜的成長過程時，經常提及個人化程式是成功的一大要素。亞馬遜對用戶資料的追求是不遺餘力：你用Kindle讀書時，被你劃線的句子、被你翻閱的篇章、是否從頭讀到尾或略讀，這些資料全數自動傳回到亞馬遜的伺服器，可以用來向你推薦下一本書。你在沙灘上用Kindle閱讀電子書，隔天登入亞馬遜的網站，亞馬遜能針對你剛讀過的電子書，以不著痕跡的方式調整網站內容。如果你花了很多時間閱讀詹姆斯·派特森（James Patterson）最新作品，卻只瀏覽一下那本新上市的節食指南，亞馬遜網站呈現給你的讀物會以大眾驚悚小說居多，養生保健的書籍會比較少。亞馬遜的用戶對個人化已經太熟悉了，該網站現在使用逆向操作的手法來多賺一點錢。在實體的書店裡，出版社可以付費讓書籍陳列在吸睛的位置，卻無法收買店員的意見。但是，一如藍尼爾的預測，收買演算法是易如反掌的事：只要塞給亞馬遜的鈔票夠多，你的書就能獲得亞馬遜軟體的推薦——冒充成「客觀」推薦，呈現給讀者。多數顧客根本無法辨別「客觀」的真偽。亞馬遜證明了關聯度能幫助公司稱霸商場，但在網路資訊汪洋中如何淘金？兩位史丹福研究生應用機器學習的原理，後來回答了這個問題。點擊訊號 在貝佐斯的新公司逐漸起飛之際，Google創辦人賴瑞·佩吉和謝爾蓋·布林（Sergey Brin）仍在史丹福大學忙著攻讀博士學位。他們知道亞馬遜有多成功——在一九九七年，網路泡沫鬧得沸沸揚揚，亞馬遜至少就市值而言具有數十億的身價。佩吉和布林是數學天才，佩吉對人工智慧的學問特別執著。但他們研究的主軸是：演算法可以提昇行銷效率，但在篩選多如牛毛的網站時，能不能也用演算法來代勞？佩吉寫出一套奇特的演算法，而身為宅男的他喜歡一語雙關，所以把自己的姓加進這套演算法的名稱：PageRank。當年多數搜尋引擎根據關鍵字來分類網站，面對用戶鍵入的一個字，很難理解哪一網頁對用戶的關聯度最大。在一九九七年提出的一份論文中，布林和佩吉把苗頭對準四大搜尋引擎，嘲諷其中三家連自己的網站也搜尋不到。「由於稍微相關的文件可能有幾十萬條，」他們寫道，「我們希望『關聯度』的概念只涵蓋好上加好的文件。」佩吉理解到，網際網路具有連結的架構，多數搜尋引擎只運用到表層的資料，其實可供開採的資料還有很多。A網頁如果連結到B網頁，等於是對B網頁「投一票」。在史丹福大學，佩吉見到教授計算論文被引用的次數，可用來概略衡量論文的重要性。他理解到，網頁好比學術論文，被其他網頁引用的次數愈多——例如雅虎的首頁——「重要性」應該比較大，被其他網頁投一票的網頁應該比較重要。佩吉強調，此一過程「善用了網路架構中的民主特質。」在當時，Google活在google.stanford.edu的網站上，布林和佩吉深信，Google應該屬於非營利事業，拒登廣告。「我們認為，搜尋引擎如果有廠商資助，本質上會對廠商偏心，會偏離消費者的需求，」他們寫道。「搜尋引擎愈精良，所需的廣告就愈少，消費者不需要廣告就能找到他們想要的東西……我們相信，廣告引來的誘因太龐雜，搜尋引擎如果要做得更加精進，必須保持透明，必須駐足在學術領域。」但當他們釋出試用版以後，網站流量是垂直上昇。Google真好用——甫一推出，立即成為網際網路的最佳搜尋網站。不久後，轉型為營利事業的誘惑排山倒海而來，令兩個二十幾歲的創辦人招架不住。在Google的神話故事當中，讓Google稱霸全球的功臣首推PageRank。我猜，Google喜歡這種說法，是因為聽起來簡單明瞭：故事結局是搜尋引擎轟動全世界，起源則是創辦人之一的一項巧思創舉。但從一開始，PageRank只是Google工程的一小部份。布林和佩吉當時已經理解出這一點：關聯度的關鍵——整理網路上大批資料的解決之道——是……蒐集更多資料。布林和佩吉注重的不僅是網頁之間連結。一個連結出現在網頁上的位置、連結的大小、網頁存在的時間長短，全是他們關心的重點。這幾年來，Google開始將這些附在資料上的線索稱為[訊號]。從一開始，佩吉和布林知道，在最重要的訊號當中，有些訊號會由搜尋引擎的使用者提供。如果用戶搜尋「賴瑞·佩吉」，然後點選第二條結果，這也算投了一票：此舉暗示，對這位用戶而言，第二條結果的關聯度大於第一條。他們將這種訊號稱為[點擊訊號]。佩吉和布林寫道，「在最有意思的研究中，有一部份將涉及運用現代網路系統現存的大批使用資料……這一類資料非常難以取得，主因是這種資料被公認是商機。」不久後，他們坐擁的資料庫之大，在全球是數一數二。Google對資料的胃口很大。一碰到資料，布林和佩吉決心照單保存。搜尋引擎找過的每一網頁，每一用戶的每一次點擊，Google全記錄下來。不久，Google的伺服器裡無所不包，收集了網路疆土大部份的即時頁庫存檔。Google從這些資料當中東挑西撿，自信能從中發掘更多線索、更多訊號，可用來進一步改良搜尋結果。Google的搜尋品質部門具有一種軍方機密行動的味道，常規是儘量不見來賓，嚴守秘密。「想把搜尋引擎的效力發揮到極致的話，」佩吉喜歡說，「必須教搜尋引擎明白使用者的語義，讓搜尋引擎精準找出使用者的目標。」Google不想列出幾千頁搜尋結果，只願列出一條，也就是用戶最想要

《搜尋引擎沒告訴你的事》

的答案。但是，某一條答案對用戶A或許是完美的答案，對用戶B就不一定是十全十美。如果我輸入「panthers」，我大概想找「美洲豹」的資料，假如搜尋者換成是美式足球迷，他想找的可能是卡羅萊納黑豹隊。搜尋引擎若想提供完美的關聯度，必須先瞭解每位用戶的興趣，必須先知道我對美式足球是一竅不通，必須先認識我這個人。認識用戶的難題在於，如果取得的資料不夠充份，也無法理解個別用戶關心的是什麼樣的事物。瞭解他人的語義是一門高深的學問——想徹底明瞭，必須先摸清對方長時間以來的行為。用什麼方式來認識用戶呢？

二〇〇四年，Google想出一套創新的策略，開始推出另一種功能，提供用戶非登入不可的服務。廣受網民愛用的電郵Gmail，就屬於Google最早推出的這一類服務。媒體的焦點放在Gmail印在側邊欄的廣告，但Google開放這項服務的動機不可能只想賺廣告費。由於電郵用戶非登入不可，Google可以自由探勘廣大的資料，也就是用戶每天收發的億萬份電郵。這麼一來，Google左手握著用戶的電郵和瀏覽行為，右手握著用戶點選的搜尋結果，更能交互比對出用戶的意向。Google也推出一套應用程式，包含了線上文書處理、試算表等功能，既能蠶食死對頭微軟的市場，又能讓用戶捨不得登出，繼續回傳點擊訊號，對Google可以說是一箭雙雕。資料這麼多，加快了Google為個別用戶建立身份論的程序，知道個別用戶的取向何在，瞭解個別用戶點擊什麼樣的連結。

至二〇〇八年十一月，Google已申請到多項個人化演算法的專利，能以程式歸類個別用戶屬於什麼團體，能根據這團體的好惡來調整個別用戶的搜尋結果。Google設想的團體類別定義相當狹隘。在申請專利時，Google舉例的類別是「有興趣收集古代鯊魚牙齒的所有人」和「沒興趣收集古代鯊魚牙齒的所有人」。屬於前一類的用戶如果搜尋「大白鯊前齒」等字詞，得到的結果有別於第二類的用戶。現在，只要是Google能接觸到的點擊訊號，Google是無所不監視。這種資料的威力不容低估：如果Google發現我先在紐約登入，然後在舊金山登入，接著又在紐約登入，Google會知道我是東西岸的空中飛人，會依照我的這項特點調整搜尋結果。Google也會觀察我使用什麼瀏覽器，從中推估我的年齡，甚至可能連我的政黨傾向也猜得出來。輸入搜尋字詞，然後點選結果，這兩個動作之間的秒數也能幫助Google分析用戶的個性。不用說，用戶輸入的搜尋字更能大曝個人興趣。即使你不登入，Google照樣能為你調整搜尋結果。Google知道你登入的大致方位——甚至明確到哪一個鄰里，而所在地能曝露個人的身份和興趣。如果搜尋「Sox」的用戶位於華爾街，這人想找的大概是「薩班斯—奧克斯利法案」（Sarbanes Oxley）的縮寫。如果用戶位於上灣（Upper Bay）對岸的斯塔頓島（Staten Island），「Sox」指的八成是「白襪隊」。

「大家都以為，我們在『搜尋』領域的研究已經停了，」佩吉曾於二〇〇九年說。「其實是大錯特錯。在搜尋方面，我們才完成百分之五而已。我們想創造的是能理解所有事物的終極搜尋引擎……有些人可能會把這種東西叫做人工智慧。」

在二〇〇六年，Google舉辦名為「Google記者日」的活動，執行長施密特在會中揭露Google的五年計劃。他指出，有朝一日，Google將能回答「我該唸哪一所大學？」之類的問題。他也說，「想至少回答這類問題的一部分，還需要幾年的努力，不過Google的終極目標是……回答較為假設性的問題。」

處處有臉書（Facebook Everywhere）

Google演算法的效力無與倫比，卻難以誘導用戶自曝品味和興趣。二〇〇四年二月，馬克·扎克伯格在哈佛宿舍裡想出一種比較容易的方法。想瞭解用戶關心什麼，與其從點擊訊號當中抽絲剝繭，不如直接開口問用戶。他開創的臉書就是依循這項原則。從大一開始，扎克伯格就對他所稱的「社交圖譜」（social graph）萌生興趣。社交圖譜象徵個人的人際關係脈絡。如果對電腦輸入社交圖譜的資料，電腦能開始分析出耐人尋味的實用資訊，例如得知朋友的近況、所在地、關注的焦點。社交圖譜也影響到新聞：在最早期的版本，臉書只限哈佛人使用，當時如果有學生的姓名出現在《哈佛紅報》（Crimson）上，臉書會自動在用戶帳號上附加校刊的連結。

以社交網站而言，臉書的起步相當晚。在扎克伯格熬夜架設臉書的那段期間，有一個版面繁雜的音樂網站紅遍半邊天——MySpace。在MySpace傲笑網路山海之前，Friendster曾受到科技圈的短暫矚目。但扎克伯格構想的網站不同。他的網站不會像Friendster，不是悶騷型的交友網站。他的構想更不會走MySpace的路線，因為MySpace鼓勵生熟人一家親。臉書想搭的便車是真實世界中既有的人際關係。和前輩比較起來，臉書簡潔俐落：重點在於資訊，而非醒目的圖形或流行文化。「我們做的是公用事業（utility），」扎克伯格後來表示。臉書不像夜店，比較近似電話公司，是一座通訊和合作的中立平臺。即使是最原始的版本，臉書一推出就如野火般延燒。陸續對少數幾所常春藤名校開放之後，扎克伯格的信箱被其他大學的學生灌爆，大家乞求他趕快把臉書的大門打開。到了二〇〇五年五月，全美八百餘所大專院校的學生都能上臉書。然而，把臉書推上更高境界的並非用戶人數，而是翌年九月問世的「動態消息」（News Feed）。在Friendster和MySpace，如果你想瞭解朋友的近況，只能去他們個人的網頁查看。啟動動態消息的演算法之後，臉書從廣大的資料庫裡

《搜尋引擎沒告訴你的事》

抽取所有人的近況，集中在一個地方，擺在首頁，用戶一登入就看得見。一夕之間，原本是頁頁相連成網的臉書，搖身一變，成了個人化的報紙，記者和主角統統是你的朋友。好友和你息息相關，關聯度之泉的水質多精純，對手哪比得上？這股泉水是一發不可收拾。在二〇〇六年，臉書用戶公佈了數十億條近況，如哲理名言、約會對象的瑣事、早餐吃了什麼。扎克伯格和工作團隊鼓勵大家多多公佈近況，因為用戶交給臉書的資料愈多，臉書會愈來愈好用，用戶再上臉書的意願也跟著提昇。臉書很早就增加了上載相片的功能，如今收集到的相片總數居全球之冠。臉書鼓勵用戶發佈其他網站的網址，用戶也踴躍支持。到了二〇〇七年，扎克伯格吹噓說，「我們每天為一千九百萬用戶產生的新聞，比任何一家媒體成立至今發佈的新聞總數還多。」起初，朋友在臉書上的一舉一動，動態消息幾乎是全部刊登。但隨著文章和朋友數增加，動態消息變得難以閱讀，令人無法消受。即使你只有一百個朋友，動態消息也多到讀不完。臉書的解決之道是EdgeRank，以這種演算法來整理預設的帳號首頁版面，亦即頭條動態消息。EdgeRank的作用是為臉書上的每一種互動排名，演算法很複雜，但基本概念相當簡單，依據三種因素來計算。第一種因素是親密度：你在某人的帳號上逗留時間愈長，表示你和這人的親密度愈大，臉書愈有可能對你發佈這人的近況。第二種因素是資訊類型的比重：以「感情狀態」為例，加權的比重非常高，因為大家都想知道誰正在和誰交往。（許多人猜測，這種加權因人而異，因為大家關心的訊息種類不盡然相同。）第三種因素是時間：近期張貼的事項比舊事重要。EdgeRank彰顯了關聯度競賽的問題核心。為了提供關聯度，個人化演算法需要資料。但資料愈多，過濾器就需要被改良得更精密，否則無法整理資料。環環相扣之下的過程是永無止境。到了二〇〇九年，臉書的用戶突破三億，之後每個月增加一千萬。年方二十五歲的扎克伯格成了帳面上的億萬富翁。但臉書有更遠大的雄心壯志。動態消息整理的是社交資訊，扎克伯格想以同樣的方式來處理所有資訊。他雖然不明講，目標卻顯而易見：運用社交圖譜和臉書族提供的大批資訊，扎克伯格希望讓臉書的動態消息演算法成為網際網路的軸心。即使是如此，在二〇一〇年四月二十一日，讀者造訪《華盛頓郵報》網站時仍然大吃一驚，因為他們發現，朋友竟然上了報社的網站。編輯都知道，讀者視線的第一個落點在右上角。那天的讀者在右上角看見的是一個明顯的框框，名為「人脈新聞」，裡面是一條條新聞的連結，每個上網的讀者看到的標題不同，全是朋友在臉書上分享的新聞。新聞網站最值錢的土地是頭版，《華盛頓郵報》卻讓臉書去編輯。《紐約時報》不久也跟進。這項新功能只是一小部份，臉書逐步推出的全盤大計稱為「處處是臉書」，最早在年度大會f8（發音是「fate」（命運））宣佈。蘋果的賈伯斯在推出新產品時，口頭禪是「炫到抓狂」（insanely great），科技人有樣學樣，從此浮誇之詞成為矽谷傳統。但在二〇一〇年四月二十一日，當扎克伯格步上講臺，他的話令人覺得似乎不無道理。「我們為網路做過的所有事情當中，以這項工程最具革新力，」他宣佈。處處是臉書的目標很簡單：讓整個網際網路「社交」起來，將臉書型的個人化散佈到目前缺乏個人化的數百萬網站。想知道臉書朋友之間流行聽什麼歌？Pandora串流音樂網現在能告訴你。想知道朋友喜歡哪間餐廳？現在餐飲點評網Yelp能答覆你。從網路的《赫芬頓郵報》（Huffington Post）到《華盛頓郵報》，時下許多新聞網站都已走入個人化時代。臉書把「讚」的按鈕傳播到網際網路各地，方便大家喜歡就按。這項新功能推出才二十四小時，「讚」就被按了十億次——所有的資料回流至臉書的伺服器。臉書的平臺工程組長布列特·泰勒（Bret Taylor）曾宣佈，用戶每月分享的事物高達兩百五十億。一度在關聯度競賽公認是一馬當先的Google，似乎開始擔心近在幾英哩外的這家競爭對手。兩大網路巨人如今已進入肉搏戰：臉書向Google的主管挖角，Google則努力編撰類似臉書的社交程式。然而，這兩個新媒體巨擘在爭什麼，多數人是看得一頭霧水。畢竟，Google以回答問題為職志，臉書的核心宗旨則是幫助大家和朋友搭上線。話說回來，兩間公司的利潤來源是同一種：目標集中、高關聯度的廣告。Google唯一的收入是文字廣告，放在搜尋結果的旁邊，也放在別家的網頁上。雖然臉書不公開財務狀況，知情人士表示過，廣告營收是該公司的主要收入。Google和臉書的起步雖然不相同，策略也互異——前者的起點是在資訊之間尋找關係，後者的起點是靠人與人之間的關係——但到頭來，兩者競逐的同樣是廣告金主。在Google或臉書上打廣告，哪個比較好？從網路廣告廠商的觀點來看，這問題很簡單。哪一家公司的投資報酬率最高？這又要扯到關聯度了。臉書和Google累積的大批資料有兩種用途。對使用者而言，這些資料讓公司更能提供和用戶個人攸戚相關的新聞與搜尋結果。對廣告主而言，這些資料是尋覓潛在顧客的關鍵。Google也好，臉書也好，只要資料最多，而且能發揮資料的最大效益，就能獲得廣告主的青睞。我們來看一看「鎖住效應」（lock-in）。使用者如果對某種科技投入太多時間或心血，即使競爭對手提出更好用的服務，使用者也認為不值得換新，這就是鎖住效應。如果你是臉書迷，發現另一個更棒的社交網站，即使這網站的

《搜尋引擎沒告訴你的事》

功能比臉書好用幾倍，你可能不會考慮轉移陣地，因為換網站的代價太高了——不僅要重新創造一個基本資料頁，上傳一大堆相片，更要苦哈哈地加入所有朋友，全是乏味到極點的程序。你等於是被鎖進臉書世界了。同理，Gmail、Gchat、Google Voice電話、Google文件、以及眾多其他產品，也是Google的佈局，目的是鎖住用戶。Google和臉書的對抗賽中，最關鍵的一役在於誰能鎖住最多使用者。鎖住效應可由梅特卡夫定律（Metcalfe's law）的角度來詮釋。網際網路上的電腦靠乙太網路協定來相互溝通，梅特卡夫正是乙太網路的發明人。梅特卡夫定律指的是，在一套網路裡，網路的實用程度是用戶人數的平方。假如你認識的人當中，只有你有傳真機，你能傳真給誰？反過來說，假如你合作的對象是人人都有一臺傳真機，唯獨你沒有，你就虧大了。鎖住效應是梅特卡夫定律的黑暗面：臉書好用，最大的原因是很多人都迷臉書。管理上出現大毛病，臉書才可能鎖不住用戶。你被鎖得愈緊，網站愈容易慫恿你登入——只要你經常登入，這些公司就能追蹤你的資料，即使你離開它們的網站，資料照常會回傳給公司。如果你登入Gmail，然後來到一個使用Google的DoubleClick廣告服務的網站，這個舉動會被記錄在你的Google帳戶上。無論是臉書或Google，它們會在你的電腦上安裝追蹤cookie，能根據你在第三方網站上的個人訊息對你打廣告。整個網際網路有可能變成Google或臉書的平臺。

然而，爭天下的並不是只有Google和臉書。Google和臉書的地盤戰每天都有財經記者和部落格報導，另一道戰線正悄悄開打卻鮮為人知。這些公司雖然行事低調，最後卻可能定奪個人化的未來。

資料市場 九一一事件之後，美國政府追緝共犯的範圍之廣在史上罕見。攻擊事件爆發之初，案情並不明朗。是否仍有劫機犯還沒被揪出來？恐怖份子的集團多大？事發三天，C I A、F B I以及其他頭文字的機構夜以繼日查緝共犯，下令全美飛機停飛，關閉機場。救兵來時，卻來自一個令人意想不到的地方。九月十四日，F B I公佈劫機者姓名，呼籲——懇請——認識歹徒的人勇於舉報。宣佈的同一天，F B I接獲一通電話，來電者是麥克·麥科拉提（Mack McLarty），曾任職於白宮，目前是民間公司的董事。這家公司雖然名不見經傳，獲利卻奇高，名稱是安客誠（Acxiom）。劫機者的姓名一公佈，安客誠立即搜尋公司內部的大資料庫。該公司的資料中心位於阿肯色州小鎮康維（Conway），佔地五英畝。安客誠過濾出歹徒一些耐人尋味的資料。具體而言，在十九個劫機犯當中，安客誠對其中十一人掌握到的資料比整個美國政府還多，不但能查出他們的新舊住址，更知道同一屋簷下住了什麼人。安客誠提供政府的資料是什麼，我們可能永遠也無法知道（不過，安客誠的一名主管向記者表示，公司提供的線索最後導致幾人遭驅逐出境、起訴。）但是，安客誠的資料涵蓋全美百分之九十六的家庭以及國外五億人，掌握的資訊包括：新舊地址、繳信用卡款的頻率、有無貓狗（品種也知道）、慣用左手或右手、服用什麼藥物（取自藥房紀錄）...單人的資料總數高達一千五百筆。儘管安客誠的作風低調——英文名字取得幾乎難以發音，或許正有此意——客戶卻包含美國多數的大公司：十大信用卡公司裡，有九家是安客誠的客戶，消費品牌從微軟到百事達，也和安客誠打交道。「如果把（安客誠）比喻為自動化工廠，」一位工程師曾告訴記者，「我們製造的產品就是資料。」安客誠對未來有什麼樣的憧憬？看看旅遊搜尋網站Travelocity和Kayak就知道。這些

網站靠什麼來賺錢，你可曾想過？Kayak的財源有兩條。第一條財源很單純，延續自旅行社時代：Kayak提供你航空公司的連結，你去買機票，然後航空公司會對Kayak賞以小額介紹費。第二條財源就複雜多了。你搜尋班機時，Kayak會在你的電腦上安裝cookie。這種小檔案好比一張便條貼，上面寫著「我想搜尋專飛東西岸的低價機票」，貼在用戶的額頭上。接著，Kayak能把這筆資料賣給安客誠，或者賣給其對手BlueKai等公司，這些公司再把同一筆資料轉賣給出價最高的商家。以機票為例，得標者可能是聯合航空之類的大公司。聯合航空一旦獲知你有遠行的意向，可以對你打廣告，讓你知道相關的航班——廣告出現的地方不只在Kayak的網站上，而是你造訪的幾乎任何一個網站。從收集你的資料到資料轉手給聯合航空，全程不到一秒。鼓吹這種做法的人士稱之為「精準行銷」（譯註：behavioral retargeting，直譯為「行為再瞄準」）。網路零售商發現，逛網路商店的網友當中，有百分之九十八的人空手離開。有了精準行銷，商家再也不必因為顧客走掉就認輸。以慢跑鞋為例，假使你上某個網站瀏覽一雙慢跑鞋，沒有衝動買下，如果這家鞋店網站採用精準行銷，同一家網店的廣告會在網路上跟蹤你，或許會顯示你剛考慮過的同一雙鞋子的圖片，擺在昨晚球賽比數的旁邊，或者張貼在你最愛看的部落格上。假使你終於舉白旗，買下那一雙慢跑鞋，那家鞋店會把這份資訊賣給BlueKai，讓BlueKai去物色買主。如果得標者是一家運動服飾店，不久後你每次上網，左逛右逛，看見的全是吸汗襪的廣告。這種緊追不捨的個人化廣告不只發生在你的電腦上，連手機也成為戰場。以Loopt和Foursquare之類的適地應用程式為例，都會從你的手機廣播所在地，替廣告廠商製造打廣告的機會，即使用戶不掛在網上，廠商照樣能以針對性的廣告接觸消費者。Loopt正在研究一套廣告系

《搜尋引擎沒告訴你的事》

統，能方便實體商店對老顧客提供優惠價和促銷品，顧客前腳一踏進店門，手機立刻顯示促銷廣告。如果你坐上西南航空的班機，前座椅背的小螢幕出現的影片可能和鄰座不同。西南航空畢竟知道你的姓名和基本資料，和類似安客誠公司的資料庫交相比對之後，對你的瞭解是加倍透徹。既然如此，何不顯示一些針對你準備的廣告呢？何不更上一層樓，放一支針對你播放的節目，讓你更捨不得關機？

處理這種資訊的新公司之一是TargusInfo，號稱能「每年提供超過六百二十億筆即時屬性資料（real-time attributes）。」換言之，TargusInfo擁有六百二十億筆資料，知道顧客的基本資料、目前在做什麼、想要什麼。另外有一家公司是盧比孔計劃（Rubicon Project），名稱同樣令人心裡毛毛的，聲稱資料庫裡囊括五億以上網友的資料。現階段，使用精準行銷的主要是廣告業者，但我們沒有理由不認為出版業和內容提供者不會跟進。再怎麼說，如果《洛杉磯時報》知道你是八卦王子帕瑞茲·希爾頓（Perez Hilton）的粉絲，今天正好專訪到他，報社可以把這則專訪挪到頭版，讓你更有可能駐足網站，多點擊一些連結。這些做法意味著，你的行為如今已成為一種商品，成了一小塊市場，為網路個人化提供平臺。我們習慣將網路互動視為一對一的關係：你和雅虎之間的互動關係是一回事，你和心愛的部落格之間的關係又是另外一回事。然而，在檯面底下，網際網路愈來愈融合。營利公司逐漸明瞭，分享資料能提高獲利。在安客誠和資料市場的推波助瀾之下，網站能將關聯度最高的產品挪到最前線，同時躲在你背後交頭接耳。關聯度競賽促成今日網路巨人的崛起，也促使公司行號對用戶蒐集更多資料，然後根據這些資料在背地裡調整用戶的網路經驗，逐漸改變網際網路的本質。新聞閱讀、投票、甚至腦袋裡的想法，個人化時代將對這些範疇產生更顯著的影響，其後果容我繼續詳述。

《搜尋引擎沒告訴你的事》

媒体关注与评论

在我認識的人當中，對數位科技與民主程序交互作用有所研究的人裡面，伊萊．帕理澤不僅是見解最精闢的一個，他的歷練也無人能出其右。《搜尋引擎的秘密》說明電腦程式如何限制網友的眼界，如何更能預測網友的反應。只要你關心人類在數位世界何去何從，你就應該閱讀本書——亞馬遜對你推薦的書單如果漏掉這本書，你就更應該詳讀。」——《大腦操縱：行銷不能說的秘密》作者／道格拉斯．洛西可夫「十餘年來，具反省能力的人士一直在擔心一件事：個人化發揮到極致，會產生什麼後果。伊萊．帕理澤的剖析是最鞭辟入裡，最發

《搜尋引擎沒告訴你的事》

精彩短評

- 1、推荐系统的存在可以说是现代互联网的基石，而被滥用的推荐系统可能会导致一系列的问题，可能人们的视野更狭窄了，互联网开始把人们分成小圈子.....
- 2、梁文道在《开卷八分钟》里介绍了这本书。视频在这里：bit.ly/IEb7Vo 简而言之，我认为这是一本讲智能搜索引擎时代的趋和效应的书。老话说得好：兼听则明偏听则暗。
- 3、好书总是一下就读完。
- 4、当真震撼到了，作为电子商务学生，竟然没有察觉到书中所提的未来。感觉观点太过极端，偶尔有忽悠人，内容有点冗余。不过可怕之处在于，书中所言之极端现象太有可能发生了。而且再次印证世界就由强大者来控制。。。。
- 5、很适合刚接触传播的人读。书中的观点并不新鲜，作者面面俱到反而失去了深挖的力度。Google多次躺枪
- 6、在开卷八分钟里听到的
- 7、说科技是中立的，显然不正确。每个人都有一套搜索引擎，大家都得到不同的搜索结果。一方面不断强化某一部分认知，一方面让自己变得越来越狭隘。可能科技在带领大家走向新的无知时代
- 8、“过于个性化的算法，会导致你看到的都是你希望/你想看到的，导致更多「不知的未知」的产生。
-师北辰
- ”
- 9、过于个性化的算法，会导致你看到的都是你希望/你想看到的，导致更多「不知的未知」的产生。
- 10、The Filter Bubble, 过滤泡泡，探讨的是整个互联网的信息传播模式和个人作为信息节点在互联网世界中的焦虑。奇怪的书名翻译，只说搜索引擎，听起来像本工具书。
- 11、we shape our tools,and thereafter our tools shape us

《搜尋引擎沒告訴你的事》

精彩书评

1、看完全書，驚覺現今之世，人的存在價值已經越來越與其人之消費能力掛鉤，商家爲了把恰到好處的廣告送到人們面前，已經無所不用其極，不計後果。消費生產已經成爲社會的絕對價值，只要一股腦兒向着那個方向前進，幸福就會自然而然到來，無須猶疑，不用討論。多年來，這種價值觀已經成功破壞了我們的文化，Dumping Down的淺薄文化充斥市場，然後我們破壞環境，耗盡子孫的資源只爲填滿自助餐桌，接下來，我們要破壞的是人性，谷歌面書的投其所好，讓人們活在各自的小天地，不管世事，而世事卻慢慢潛進家裏，管轄你的一切，監視你的一舉一動，而大家沉默着接受，仍然覺得No big deal。書中帶出來的未來景象其實相當恐怖，但又有多少人會爲此而做點什麼呢？你會從此不再用微博QQ百度嗎？你會從今天開始不做低頭族即時把手機裏的app全部刪除嗎？我會，我會停用google，不玩facebook，也不看youtube，真的，尤其當我在中國大陸的時候。

《搜尋引擎沒告訴你的事》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu000.com