

GOOGLE 创世纪

文 / John Battelle 译 / 王博

Larry 认为 Sergey 太自大，而 Sergey 认为 Larry 太讨厌，但是他们又相互被对方所吸引，这可能预示着某件伟大的事业的开始。

事情是从一场争论开始的。他们第一次相遇是在 1995 年的夏天，那时 Sergey Brin 还是斯坦福大学计算机系的二年级研究生。由于天生喜爱社交，Brin 志愿担当了未来的一年级学生的校园向导。这些“未来”的学生已经被斯坦福录取，但是他们自己还没有决定是否来这个学校读书。Brin 的任务就是带领他们参观校园并且游览一下附近的旧金山。Page，他是一个刚刚从密歇根工程专业毕业的学生，正好就在 Brin 带领的那小组里。

这次见面绝对不是所谓的一见钟情。那天在旧金山的山路上，他们之间的争论就没有停息，争论的是诸如哪条去城里的路线更好之类的话题。“Sergey 真是爱交际的家伙，他喜欢与各种人相识。”Page 回忆道，认为自己与他相比确实很沉默寡言。“我认为他实在是太讨厌了。他在很多问题上都有很强的主见，不过，我想我也是这样的。”

“我们都很讨厌对方，”当我告诉他 Page 的反应的时候，Brin 反击道，“但那只是戏谑的说法。我们彼此交谈得很多，这能说明某些事情。我们都喜欢开这种善意的玩笑。”Page 和 Brin 可能发生过冲突，有过争论，但是毫无疑问的

是，他们都彼此相互吸引。

几个月以后 Page 在斯坦福大学出现的时候，他选择了人机交互先锋 Terry Winograd 作为他的导师。不久他就开始寻找他的博士论文题目。Page 的爸爸是密歇根大学的计算机科学的教授。他告诉 Page，博士论文是如此的重要，以至于可以决定将来的学术生涯。所以博士论文题目的选择对于 Page 来说是一个非常重要的决定。在剔除了十个左右都很有趣的想法之后，最终他发现真正吸引

它的结果就要比那时候常用的搜索引擎返回的结果好上许多。而那些搜索引擎主要是以关键字来排列搜索结果。这样，Page 和 Brin 知道他们已经开始了某项伟大的事业。

他的就是正在飞速发展的互联网。

Page 一开始并没有想到要去以一种更好的方法搜索整个互联网。尽管斯坦福的毕业生都因为网络公司而变得越来越富有，但是网络真正吸引 Page 的还是它的数学特性。每个电脑都是一个节点，而网页上的链接是就是节点的联系——这是经典的图（graph）结构。“计算机科学喜爱图，”Page 告诉我。在 Page 的理论里，整个互联网就是有史以来人们创造的最大的图，并且还在以惊人的速度在继续成长。在这张大图的顶点上隐

藏着许多有用的知识，等待着这个好奇的研究生去发现他们。在导师 Winograd 的同意下，Page 开始考虑互联网的链接结构的问题。

引用和 BackRub

Page 的思考是很有成效的。他观察到一个事实：从页面的一个链接点到另外一个页面确实是很平常的事情，但是如果找出某个页面被哪些页面反向链接就不平常了。换种说法，当你浏览一个页面的时候，你并不知道哪些别的页面可以链接到这个你正在浏览的页面。这个问题让 Page 很困扰。他认为如果知道页面被谁反向链接的话，将会是非常有用的。

为什么呢？为了更好地理解这个问题的答案，我们先要首先看看学术出版的世界。对于教授来说，尤其是那些从事像化学和数学这种精确科学（hard science）研究的教授来说，没有什么事情能比让自己的论文发表更重要了。也许只有一件事情例外，那就是引用。

科学家们撰写的论文都是基于引用的基础来构造的：每个论文都要以先前发表的论文作为论据，来支持作者的论点，最后得到一个作者自己的结论。我们如果要评判一个论文的好坏，也不是看论文本身的思想，而是要参考这篇论

文引用的论文的数量,和反向引用这篇论文的别的论文的数量,以及每个引用的重要程度。引用是如此的重要,以至于这个世界上有一个专门研究它们的学科:文献计量学。

很好,但是这和互联网有什么关系呢?是的,就是 Tim Berners-Lee 对这个系统的正向的改进,使得互联网诞生了。同样的,由于 Larry Page 和 Sergey Brin 对 Berners-Lee 的互联网的反向改进,使得 google 诞生了。将这些努力串连起来的那个绳索就是引用——指向别人的研究成果以便得到你自己的研究成果的行为。

Page 的第一个关于反向链接的研究是一个叫 BackRub 的工程。

他推理整个互联网是松散地建立在引用的基础上的——毕竟,链接就是某种意义上的引用吧?如果他可以发明一种方法找出网络上反向链接的数量,并且评估他们的质量,就像 Page 指出的那样“网络将会变得更有价值。”

在 Page 构想他的 BackRub 工程的时候,整个互联网由大概一千万个文件,以及他们之间无法计数的链接组成。爬行如此巨大的互联网所需要的计算资源,远远超出了一个学生项目力所能及的范围。那时 Page 也许并没有准确地意识到他投身的工作究竟是什么,但是他已经开始去设计他自己的网络爬虫。

那个想法的复杂度和规模将 Brin 吸引到了这个工程。这个从未在一个论文题目上钻研的博学者总是从一个项目跳到另外一个项目,他发现这个 BackRub 工程背后的假定十分的有趣。“我和学校中的许多的研究小组都有过接触”, Brin 回忆道,“但是这个却是最激动人心的项目,因为它研究了网络——象征人类知识的网络,还有一个原因就是我喜欢 Larry。”

大胆的评价

1996 年 3 月, Page 将他的网络爬虫指向了他在斯坦福的个人主页,并且放开了这只爬虫的缰绳。于是它从那个页面开始了在整个互联网上的旅行。

爬行整个网络,发现全部的链接是这个工程的主要工作,但是简单的爬行并不是 BackRub 真正创新的地方。Page 对学术出版物的排行有一种天生的敏锐。他指出网络的图结构不仅仅是要发现谁链接到了谁,并且是要评估链接过去的“谁”的重要性。这个重要性的评估是基于这个做这个链接的站点的各种不同的属性的。他发现每个链接需要他们自己的评估和排行。他从引用分析中看到,链接到这个页面的链接数量会是这个页面评估的一个有用的标准。但是这种方法会有一个极其困难的数学上的挑战——你不仅仅要计算一个单独的页面的链接,还需要去计算从别处链接到这个链接的链接。这样的话,工作就会变得十分的困难。

幸运的是, Page 现在和数学天才 Brin 一起工作。Brin 出生在俄罗斯,他的爸爸是 NASA 的科学家,并且是马里兰大学的数学教授。他在六岁的时候就和全家一起移民到了美国。在他上中学之前,他的数学天分就被人发现了。高中还剩下一年,他就直接进入了马里兰大学。当他毕业的时候,立即就被斯坦福大学录取。而他的天才可以让他在哪里轻松地混日子。他告诉我,天气真是好,他上的课程都是那些非学术性的课程——帆船,游泳,潜水。他将他的智力能量主要花在有趣的项目上而不是实际的课程工作上。

Page 和 Brin 一起创立了一个评估的系统。这个系统奖励那些来自重要源头的链接,并且惩罚那些不是来自重要源头的链接。例如,许多站点都有 IBM.com



的链接。这些链接范围从技术工业领域的商业伙伴,到伊利诺州郊区刚刚得到了 ThinkPad 作为圣诞礼物的小孩。从一个观察者的角度,商业伙伴链接的重要性要大一些。但是一个算法要怎么样才能理解这个事实呢。

Page 和 Brin 的突破之处就是创造了一种叫做 PageRank 的算法。这种算法会考虑链接到某个站点的链接,并且同时会考虑那些反向链接所在的站点——也就是也考察链接到那些站点的

链接。这其实就是学术引用计数的一个翻版。这种算法成功了。在上面的例子中,我们假设只有很少的站点链接到了这个小孩的站点。让我们再次假设,链接到这个小孩的站点同样也只得到了很少的外部链接。但是相对的,平均会有数千个链接连向 Intel,或者类似 Intel 的站点。PageRank 就会认为小孩的链接没有 Intel 的链接那么重要,至少是对 IBM 来说。

这仅仅是一个过于简化的例子,事实上,Page 和 Brin 需要去改正任何数学上的死胡同难题。但是不管怎么样,他们做到了这样:受欢迎的站点会在他们注释列表的顶端,而不那么受欢迎的将会沉到底部。

当他们拿着这些结果瞎玩的时候,他们发现他们的数据可能会在互联网的搜索上有所应用。事实上,这个将 BackRub 排列好的页面结果用于搜索的想法是如此的自然,甚至在他们发现这一点之前就是这样了。BackRub 已经像一个搜索引擎那样的工作,你给它一个 URL,它就会给你一个按重要性排列的反向链接的列表。“我们意识到我们拥有了一个查询的好工具,”Page 回忆道,“它会给你一个总体上的页面排名,并且会按顺序排列它们。”

Page 和 Brin 发现他们的 BackRub 的搜索结果要好于那些已经存在的搜索引擎,比如说 AltaVista 和 Excite,这些搜索引擎总是返回不相关的列表。“他们总是只留意页面上的文本,但是忽略了其他的重要的信号。”Page 回忆。这个信号就是现在为世人所熟知的 PageRank。为了测试它在实际的搜索应用中是否会工作良好,Brin 和 Page 一起开发出了一个 BackRub 的搜索工具。它只搜索页面的标题中的文本,并且运用 PageRank 来将搜索结果按相关性排列。仅仅这样,它的结果就要比那时候常用

的搜索引擎返回的结果好得多。而那些搜索引擎主要是以关键字来排列搜索结果。这样,Page 和 Brin 知道他们已经开始了某项伟大的事业。

不仅引擎本身很好,他们还意识到引擎的规模会随着互联网规模的扩大而扩大。因为 PageRank 是以分析链接来工作的,网络越大,引擎就会越好。这个事实使得引擎的发明者将引擎命名为 google。其实最早是 googol,就是 1 后面跟了 100 个零的那个数字。他们在相遇的一年之后,在斯坦福的网站上发布了 google 的第一个版本。

Google 让斯坦福的一小部分知情者眼前一亮。被激励的 Page 和 Brin 开始改进他们的服务,加入全文的搜索,并且将越来越多的页面加入到索引中。他们很快地意识到搜索引擎需要巨大的计算资源。他们没有钱去买新的电脑,他们到处祈求别人的帮助并且借钱,终于让 Google 出世了。他们的努力包括从网络实验室搞到的硬盘,还有系里面闲置的 CPU。Page 的宿舍成了他们的实验室,他们像现代电脑版的 Frankenstein (英国女作家 Mary Wollstonecraft Shelley 所著小说中主人公,一个生理学家,手创一个怪物,但结果自己被怪物所毁。)那样用闲置的零件拼起来服务器,并且将它连到斯坦福的宽带校园网上。在将 Page 的宿舍填满之后,他们开始将 Brin 的宿舍变成一间办公室和编程中心。

在计算机系和校园网管理办公室,他们的项目已经成了某种传奇。那时,BackRub 的爬虫几乎占用了斯坦福大学整个网络带宽的一半。你得知道,斯坦福大学是这个世界上拥有最棒的校园网的学校之一。在 1998 年的秋天,这个项目几乎就要毁了整个斯坦福的网络连接。

“我们很幸运,在斯坦福有很多有远见的人,”Page 回忆道,“他们并没有

因为我们占用过多的网络资源而与我们争吵。”

公司的创立

当 Brin 和 Page 继续他们的 BackRub 工程和 Google 应用的实验的时候,他们在斯坦福的校园内和网络搜索的学术界都引起了注意。

其中一个听说 Page 和 Brin 工作的是康乃尔大学的 Jon Kleinberg 教授,他当时正在 San Jose 的 IBM 的 Almaden 中心研究文献计量学和搜索技术。他的 hubs-and-authorities 搜索技术也许是 PageRank 之后第二著名的搜索技术。在 1997 年的夏天,Kleingerg 来到斯坦福并且与 Page 见面。Kleinberg 那时已经完成了他的论文“可信的源”的初稿。Page 向他展示了 Google 的早期版本。Kleingerg 鼓励 Page 发表关于 PageRank 的学术论文。

Page 告诉了 Kleinberg 他对于发表论文十分担心的原因。“他十分担心别人会偷走他的想法,Page 觉得 PageRank 就是他的秘密配方。”Kleinberg 告诉我。(Page 和 Brin 最终还是发表了论文。)

另一方面,开创公司并且运营毕竟是十分艰辛的事情,他们还没有决定是否要去真正创立自己的公司。Page 在斯坦福的第一年,他的父亲去世了,Page 的朋友回忆当时 Page 在某种程度上将他的博士学位看作对他父亲的祭品。同样生长在学术家庭的 Brin 也不愿意放弃自己的学业。

Brin 记得他的导师在和他谈话的时候,告诉他:“试试看吧,如果 google 成功了,那自然很好。如果没有成功,你可以回到我们的研究生院继续完成你的学位的。”他轻声地笑了起来,然后说:“我当时说道,‘好吧,为什么不呢?我会试试看的。’”