

008 | 曾经辉煌的雅虎研究院

2017-10-20 洪亮劫

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 10:41 大小 4.89M



雅虎是最早成功的互联网公司之一，也是最早意识到需要把基础研究，特别是机器学习以及人工智能研究，应用到实际产品中的公司。雅虎从很早就开始招聘和培养研究型人才，雅虎研究院就是在这个过程中应运而生的。

今天我就来说一说雅虎研究院的历史，以及过去十多年间取得的成就，聊一聊如何通过引进高级人才，迅速构建起一支世界级的研发团队。当然，也会聊一聊研究院的衰落。高级研发机构对于企业而言往往是锦上添花的事情，在整个公司产品 and 视野都欠缺的情况下，也往往避免不了最后衰败的结局。

雅虎研究院的创立

雅虎研究院的故事要从一个叫乌萨马·菲亚德（Usama Fayyad）的人说起。乌萨马出生在北非突尼斯的迦太基（Carthage），早年在突尼斯以及其他地中海沿岸国家度过，包括中东、非洲以及南欧的一些国家。高中时期在约旦的安曼生活，后来在美国密歇根大学度过了他的本科（1984 年）、硕士（1986 年）以及博士（1991 年）生涯。毕业之后，乌萨马来到了美国加州南部隶属于美国国家航空航天局（NASA）的喷气推进实验室（Jet Propulsion Laboratory）工作，一直到 1996 年。之后加入微软研究院，从事数据挖掘的研究工作。

早在 1994 年，乌萨马就和拉马萨米（Ramasamy Uthurusamy）一起组织了最后一届 KDD 研讨班，然后在 1995 年，他们把这个研讨班升级成了会议，并在加拿大蒙特利尔举办了第一届 KDD 大会（First International Conference on Knowledge Discovery in Data）。从此，KDD 大会成了数据挖掘、数据科学以及应用机器学习的顶级会议。

1996 年，乌萨马又创办了一本叫《数据挖掘和知识发现》（Data Mining and Knowledge Discovery）的学术期刊，并亲自担任主编。这本期刊也渐渐成了数据挖掘领域主要的学术期刊之一。乌萨马本人可以说在 20 世纪 90 年代中期，就已经开始成为数据挖掘领域重要的领军人物。

进入 21 世纪，乌萨马先是在 2000 年创立了一家叫 Audience Science 的数据挖掘公司并担任 CEO，然后又在 2003 年创立了一家叫 DMX Group 的数据挖掘咨询公司，后者于 2004 年被雅虎收购。不久后，他成为雅虎的执行副总裁以及首席数据官（Chief Data Officer），这也是互联网历史上的第一位首席数据官。

因为雅虎在搜索以及广告业务上的扩展，乌萨马意识到应该成立一个类似于微软研究院，但更偏向于互联网业务的研究组织，这个想法得到了公司 CEO 杨致远的支持。乌萨马当时的首要任务是为研究院物色一位院长。

经过一段时间的寻找，他成功邀请到普拉巴卡·拉加万（Prabhakar Raghavan）来担纲。今天回头来看，普拉巴卡无疑成功地引领了雅虎研究院，并让其一度成为人人向往的互联网研究机构。当然，这跟普拉巴卡本人的经历也密切相关。

首先，他本人就是知名的学者，参与撰写的经典教科书《随机算法》（Randomized Algorithms）和《信息检索导论》（Introduction to Information Retrieval）在学术界享有盛誉。他还是 ACM、IEEE 的院士，也是美国工程院院士，这为他招纳学术界权威人士和

博士生提供了便捷。加入雅虎之前，他已经在 IBM 研究院以及 Verity 任职多年，IBM 的从业经历更是让他对企业文化和工业界的研究机构有了很深的了解。

2005 年 7 月，雅虎研究院正式成立，普拉巴卡担任研究院负责人，向乌萨马汇报。2008 年，雅虎研究院与之前就在搜索与广告事业部存在的应用科学部门合并。在卡罗尔·巴茨 (Carol Bartz) 任职 CEO 期间，普拉巴卡直接给她汇报，并且普拉巴卡还曾担任首席战略官。

雅虎研究院的蓬勃发展和辉煌

雅虎研究院组建之后，首要任务当然就是吸引工业界和学术界的知名学者，从而能够组建一个有效的团队。普拉巴卡利用他个人和乌萨马的声望，很快就做到了这点。

比如，之前和普拉巴卡在 IBM 共事的安德鲁·汤姆金斯 (Andrew Tomkins) 加入团队，担任负责搜索的首席科学家以及搜索方面的副总裁 (安德鲁后于 2009 年之后加入谷歌担任工程总监)。再比如，曾和普拉巴卡在 IBM 共事的安德烈·布罗德 (Andrei Broder) 2005 年加入团队，担任负责计算广告方面的副总裁。

安德烈本人大有名头。他在斯坦福大学攻读博士期间师从图灵奖得主高德纳 (Donald Knuth)，然后在曾经名噪一时的第一代搜索引擎公司 AltaVista 担任首席科学家，之后加入位于纽约的 IBM 研究院组建企业级搜索平台。和普拉巴卡一样，安德烈也是 ACM 和 IEEE 的双料院士。2012 年安德烈加入谷歌，担任杰出科学家 (Distinguished Scientist)。

我们这里简单列举一些曾经在雅虎研究院工作过的知名学者，我们便可一览其盛况：

Ronald J. Brachman：哈佛大学计算机科学博士，加入雅虎研究院之前长期于贝尔实验室工作，曾担任贝尔实验室人工智能研究部的负责人。1996 年之后担任 AT&T 实验室通信服务研究中心副总裁。2005 年加入雅虎研究院协助普拉巴卡进行管理，并于 2012 年到 2016 年间担任雅虎研究院首席科学家以及负责人。Ronald 曾任 AAAI 主席。2016 年之后担任纽约康奈尔科技大学的 Jacobs Technion-Cornell 研究院院长。

Yoelle Maarek：以色列理工大学计算机科学博士，加入雅虎研究院之前曾任 IBM 研究院的杰出工程师和谷歌的工程总监。历任雅虎研究院以色列分部的负责人、高级研究总监，并在 2016 年 Ronald 离开之后任雅虎研究院的负责人。

Jan Pedersen：斯坦福大学统计学博士。2002 年加入 AltaVista 担任首席科学家（在安德烈之后）。2003 年加入雅虎研究院担任搜索和广告方面的首席科学家（在安德鲁·汤姆金斯之前）。2009 年加入微软，担任 Bing 核心搜索部门（Core Search）的首席科学家。2017 年加入 Twitter，担任数据科学副总裁。

Ben Shahshahani：普渡大学电气工程博士。曾在 Nuance Communications 担任工程总监。2005 年加入雅虎研究院，之后历任负责搜索广告的高级总监以及搜索与媒体科学组的副总裁。2012 年加入谷歌任工程总监。2014 年回到雅虎，任广告科学方面副总裁。

Ricardo Baeza-Yates：滑铁卢大学计算机科学博士，ACM 和 IEEE 双料院士，信息检索和搜索方面的权威，著有《现代信息检索》（Modern Information Retrieval）一书。他在雅虎研究院担任拉美和欧洲分部的副总裁直至 2016 年，也是智利科学院以及工程院的院士。

Ravi Kumar：康奈尔计算机科学博士，加入雅虎研究院之前在 IBM 研究院从事数据挖掘算法的研究。2005 年加入研究院之后担任首席研究科学家。2012 年加入谷歌担任高级主任研究科学家（Senior Staff Research Scientist）。他的论文引用数达 3 万次以上。

Evgeniy Gabrilovich：以色列理工大学博士，在雅虎研究院担任首席研究科学家，并且担任自然语言处理方向研究的负责人。2012 年加入谷歌担任高级主任研究科学家。2012 年当选 ACM 杰出科学家（ACM Distinguished Scientist）。

Deepak Agarwal：康涅狄格大学（University of Connecticut）统计学博士，加入雅虎研究院之前在 AT&T 担任高级研究科学家一职。2006 年加入雅虎研究院担任首席研究科学家，主要研究推荐系统相关的内容。2012 年加入 LinkedIn，担任人工智能和机器学习方面的副总裁。

Alexander Smola：柏林理工大学计算机科学博士，加入雅虎研究院之前任澳大利亚国立大学教授。2008 年加入雅虎研究院后任首席研究科学家（Principal Research Scientist）。2013 年加入卡内基梅隆大学任教授一职。2016 年加入亚马逊担任机器学习方面的总监。他的论文引用数达 8 万次以上。

Jianchang (JC) Mao：密歇根州立大学计算机科学博士，加入雅虎研究院之前曾在 IBM 研究院任职，还曾担任 Verity 的首席软件架构师。2004 年加入雅虎之后任广告科学方面副总裁。2012 年加入微软之后，先后担任 Bing 的多个职务并于 2016 年被提升为公司副总裁。他的论文引用数达 1 万次以上。

Raghu Ramakrishnan：德克萨斯大学奥斯汀分校计算机科学博士，加入雅虎研究院之前担任威斯康星大学教授。2006 年加入雅虎研究院之后任云计算方面的副总裁。2012 年加入微软之后一直担任 CTO，负责云计算领域。他的论文引用数达 3 万次以上。

当然，在雅虎研究院工作过的知名人士还有很多，这里无法一一列举。不过我们可以看出，不少人在离开雅虎之后，依然在业界发挥着不小的作用。

除了招揽到一批优秀人才，雅虎研究院也发表了一系列有价值的研究成果，在很短的时间内建立了学术研究上的威望。在 10 年间，据不完全统计，雅虎研究院的学者获得过两次信息检索顶级会议 ACM SIGIR 的最佳论文、3 次数据科学和数据挖掘顶级会议 ACM KDD 的最佳论文、两次机器学习顶级会议 ICML 的最佳论文、两次推荐系统顶级会议 ACM RecSys 的最佳论文、两次信息检索以及网络信息挖掘的权威会议 ACM WSDM 的最佳论文、两次信息检索和数据库领域顶级论文 ACM CIKM 最佳论文以及一系列有影响力的最佳论文奖项，涵盖了搜索、广告、推荐系统、数据挖掘、机器学习、人机交互等诸多方面，为互联网研究和发展做出了重大贡献。

可以说在非常短的时间内，雅虎研究院就用卓越的研究成果向世人证明了这个团队和组织的实力。曾经在某一段时期内，世界各国的优秀研究人员和博士毕业生都希望跻身雅虎研究院的研发队列。

雅虎研究院逐渐成为历史

2012 年是雅虎历史上格外动荡的一年。先是公司 CEO 卡罗尔·巴茨在上一年 9 月份被董事会解雇；然后经历了短暂的临时 CEO——蒂姆·莫尔斯（Tim Morse）；之后新 CEO 斯科特·汤普森（Scott Thompson）在 1 月上任，5 月份就因学历造假丑闻离职；罗斯·莱文索恩（Ross Levinsohn）之后担任公司临时 CEO 直至 7 月。然后，玛丽莎·梅耶尔（Marissa Mayer）加入公司担任 CEO。短短不到一年的时间里，共有 5 个人担当了 CEO 的职位。

在这个过程中，普拉巴卡离职并加入谷歌，很多之前追随他的人也先后加入谷歌。普拉巴卡离开后，罗纳德·布拉赫曼（Ronald J. Brachman）接过了研究院领导人的位置，并在 2012 到 2016 的 4 年间为玛丽莎重新招募了超过 100 名博士科学家。

2016 年 2 月，雅虎宣布研究院不再作为一个独立实体而存在，罗纳德离职，所有研究人员被分散到各个工程部门，依然保留雅虎研究院的对外旗号，耶艾尔·玛瑞克（Yoelle Maarek）担任负责人。2017 年雅虎和 Verizon 合并，雅虎作为一个独立的公司成为历史。

雅虎研究院逐渐淡出历史舞台，这固然有公司高层频繁更换的原因，也有一些更加深层次的原因。研究院的成果往往都需要一定时间才能直接在产品中体现出来，因此，虽然在技术上

研究院能够帮助公司提升水平，但是实际产品的效果未必就一定能够受到用户的青睐。

例如，研究院曾经投入了大量人力物力，利用机器学习来提高搜索引擎的搜索品质。可以说，雅虎是最早将人工智能和机器学习技术大规模应用在搜索引擎上的公司。但是搜索引擎的好坏很多时候是一个产品、技术、设计的综合体现，雅虎研究院研发的算法并没有在产品的综合表现中挣得额外加分。

相似的例子还包括雅虎研究院在早期就投入了很多力量研发广告平台，甚至包括安德烈·布罗德本人到斯坦福大学开设了世界历史上第一门计算广告学的课程。然而，雅虎整个平台的产品都在下滑，因此广告平台受到了额外的压力。虽然研究院的科学家们在算法和模型上做出了很多创新，也在一定时间内带来了不小的收益，但都无法改变整个公司产品线运营不佳的情况。于是，**雅虎研究院的成果在雅虎整体业绩不理想、公司产品缺乏想象力的大背景下显得杯水车薪，并不能从整体上扭转公司的颓势。**在公司进入动荡之后，研究院对于高层领导来说，往往也就不是公司的重点发展对象了，研究院的瓦解也就成了必然。

小结

今天我为你分析了雅虎研究院的兴衰。一起回顾下要点：首先，雅虎研究院曾通过引进高级人才的方式，迅速构建起了一支世界级的研发团队，并发表了一系列有价值的研究成果，建立起在学术研究上的威望，创造了研究院曾经的辉煌；其次，因高层变动以及一些深层次的问题，雅虎研究院没有摆脱最后衰落的结局，一切辉煌终成历史。

最后，给你留一个思考题：到底什么样的**企业环境**能够最好地发挥研究院的成果，又是怎样的**研发流程**能够使研究院成为公司新动力的源泉？雅虎研究院在当年并没有找到答案。不知道随着最近一批互联网新贵纷纷成立人工智能研发团队的契机，大家是否能够找到更好的研究院运作模式。

欢迎你给我留言，和我一起讨论。


AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 007 | LDA模型的前世今生

下一篇 009 | 数据科学家高阶能力之分析产品

精选留言 (10)

 写留言



沉 置顶

2017-10-20

 2

我想也许是雅虎的企业文化太缺乏推动力，商业团队的内心里可能从未认真想过能快速有效的将技术转化为商业



范深

2017-10-21

 4

产品和商业的创新和升级，并不能完全通过算法或技术来弥补。

展开 ∨





王远方
2017-10-20

👍 3

由雅虎开发的 Hadoop 对大数据的发展有很大推动，仅凭这点雅虎就值得尊重。



兔子ORZ
2018-04-14

👍 2

企业环境：没有盈利压力，长时间的投入一项研究，更能出成果。
研发流程：研发目标最好是和公司现有或未来目标一致，能发挥想象力开发新战略的更好。
展开 ▾



潜行
2018-08-28

👍 1

人工智能不能靠发paper为主，学术界和工业届还是有差距的，比如阿里达摩院成立的时候马云就强调研究成果要保证能落地。



JIA
2017-10-20

👍 1

看了这篇，不得不想到阿里刚刚成立了达摩院，也是要人有人，要钱有钱啊。从雅虎研究院的发展来看钱和大牛好像都不能解决问题，也不是保证啊，不过很期待。
展开 ▾



帅帅
2018-10-20

👍

这个团队太过厉害~~~~~

一群特种兵混在市井之间的感觉~~~
展开 ▾



Lynn
2018-05-26

👍

到底什么样的企业环境能够最好地发挥研究院的成果？
研究结果可以很好的落地到项目中，才会使研究成果变现成为可能。
让研发成果的嵌入像系统核心模块一样存在，并且完善A/B Test机制，让研究成果可感知。

...

展开 ▾



安乐天

2018-01-19



雅虎,施乐,IBM,AT&T这些工业界的研究院随着企业的兴衰而boom and bust.时也势也, IT世界迭代太快, 企业如此, 研究院也如此, 人也如此



Xuan

2017-10-24



研究的成果带来的利益能cover住研究本身的开销才能长久的存在吧

展开 ▾