

·科技纵横·

[编者按] 2015年5月,在广州举行的第17届中国科协年会“国际科学大师论坛”邀请数位诺贝尔奖获得者作了精彩的学术报告。这些报告或涉及某一重要科学问题的发展、或涉及科学家的某一段科研心路历程,也不乏科学家的研究心得。《科技导报》精心整理了其中的4篇报告,自21期开始在“科技纵横”栏目刊载,每期一篇,以飨读者。

报告人/温顿·瑟夫(Vinton G. Cerf)

数字化“羊皮纸”:为时代保存数字化内容



温顿·瑟夫,2004年图灵奖获得者,谷歌副总裁

历史中,我们的先祖曾使用石块或者石碑来记录信息。很多年后,虽然石块上的符号仍然存在,但是我们已经不知道怎样去理解上面的文字了。生活中有很多信息的存储介质,如图1中的羊皮、纸张等。很多信息存储介质都具有一定的物理形态,也有一些存储介质并不具备物理形态。

20世纪,桃丽丝·基恩斯·古德温所著的《林肯与劲敌幕僚》(A Team of Rivals)中就涉及到了信息存储。这本书非常有意思的是,能够重现林肯与他的劲敌幕僚以及其他人员之间的对话。作者曾经去美国不同的博物馆和档案馆查阅资料,发现了她所写的故事中那些人物间的来往信件,然后了解到这些人物是通过怎样的方式彼此沟通的,这就成为了她写作的依据。不妨设想一下,到了22世纪,人们怎样了解21世纪早期人类

的生活方式呢?比如,怎样寻找一个被储存在某个地方(存储介质)的电子邮件、微博或微信的记录呢?

人们的确有很多信息的存储介质。20世纪70年代,已经出现软盘。虽然有些人现在还保存着这种软盘,但是已经没有

相应的设备来读取它了。因为最新的软件往往不能读取老格式文件,我们不能要求每个软件都向下兼容。目前正在使

用的存储介质还有硬盘、光盘等等,总有一天,其中存储的信息也会像软盘一样不能被读取出来。我们利用存储介质进行存储,但是并不知道这些存储介质的有效期。从最开始用的大软盘到小软盘、再到CD,存储介质不断改变,并且每种介质所存储的信息量也在不断增加。能否找到一种方式,使得信息可以储存一千年、两千年甚至更久,仍然能被读取出来呢?

如果你平时使用电子书阅读器(如Kindle)或者电脑,就经常会看到图2中的标志。这些标志代表了常见的互联网上的软件,能够帮助我们很好地理解比特所承载的信息,所以我们要能够使用互联网软件的设备。

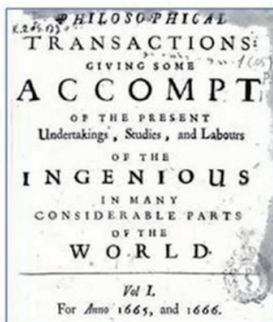


图1 信息的存储介质



图2 互联网上的软件

另一个问题是信息的可持续性,这个问题更复杂。文字、音频或者视频是一些静态的内容,但是,对于可以与人互动的内容(如网络游戏、工作表),它们的价值是什么呢?很多年前,我用过一个非常好的应用软件来进行文档写作,直到现在还保存着这些文档。可是,现在已经没有相应的硬件来识别这些文档了。就算我可以执行这个软件,可以了解这些文档的内容,但其中的信息对我以外的人来说已经没有任何用处了。不管我们多么小心地去保存信息,把它们从一种物理形态的介质转存到另一种物理形态的介质,类似问题依然会出现。

我们有计算机博物馆,但是还没有软件博物馆。我发现目前美国有一个类似的博物馆,已经开始保存软件了。如何去保存软件是一项挑战,因为硬件会按照软件的要求执行,我们需要软件来帮助解读比特信息。如果我们一直保存这些软件,100年以后,可能再也找不到相应的硬件和操作系统了。知识产权也

是一个重要问题,在什么样的情况下可以一直使用这个软件?此外,还有一些问题,比如,某个软件的开发商由于转型而导致这个软件不再支持某一种应用程序了,等等。所以,在技术上,我们必须要有好的方式去存储数字

形式的信息。我们要思考,如何保存合适的硬件,使得管理硬件的操作系统能够很好地运行,进而操作软件来识别目标文件中的信息。

现在,已经有了解决办法,就是Olive软件。我的好朋友Mahadev参与到这个软件的开发过程中。他发明了一种保存可运行软件的方式,使得人们在遥远的未来也可以运行很久以前的软件。这一点非常重要!首先必须有能够准确重复运行的硬件,能够完整描述文件运行的类型,并运行这个软件的操作系统。

虚拟环境对虚拟机是至关重要的,而虚拟机可以使我们很好地去运行这些软件。通过建立虚拟环境,就可以进行所需的操作,比方说现在正使用一个操作系统,如运行一个编辑文档的应用软件。如果我们用一个数字X光机来看一下硬件指令的流程,就可以获得一个数字化版本。这个关于操作系统的数字化版本是可以运行的,把它保存在稍后可

以获取的一个地方。我们会发现,建立一个虚拟机本身就是一个大问题,因为它会模拟一整套硬件,这就意味着要很大的存储空间来解决硬件模拟的问题。我的朋友成功地解决了这个问题。他是如何做的呢?以典型的X86计算机所用的操作系统为例,在这个操作系统之上建立一个虚拟机,虚拟机会连接到互联网上去发现它所要的软件,并且从云里面获得想要运行的内容。接下来是硬件模拟器,如果是原来的硬件或者X86的系统,就不需要模拟器了。旧操作系统和旧应用程序随时可以复制,人们只需要有这个虚拟机就可以了,不再需要更新硬件。进一步地,如果可以不断地去更新虚拟机,使其从一种操作系统更新到另一种操作系统,人们就不需要去担心旧的操作系统的的问题。这就是所谓的“小把戏”,通过虚拟机来做一个虚拟的系统,或者说一个虚拟环境,虚拟机的镜像可以在不同的云服务器上运行,虚拟机之上再运行客户的软件。这样,所有的应用就可以长期保存并随时运行。

虽然这种方法不能解决所有的问题,但是却解决了某些非常棘手的问题,让这些数字文件可以有效保存很长时间。这是基础性的工作,对于我们这一代在未来变得“可见”是非常重要的。

(北京大学信息科学技术学院教授黄罡、北京航空航天大学计算机学院教授许可、国防科学技术大学计算机学院教授卢凯帮助审阅、完善本文,特此致谢。)

(整理/王丽娜)