编码:隐匿在计算机软硬件背后的语言

作者：查尔斯•佩措尔德 (Charles Petzold)

第1章　电 筒 密 谈

假若你才 10岁，你的好朋友与你临街而住，而且你们卧室的窗户面对着面 。每天晚上， 当父母像平常一样很早催你上床睡觉时，你可能还想与好朋友交流思想、发现、小秘密、传 闻、笑话和梦想，没有人可以责备你，毕竟 ,渴望交流是大多数人的天性。

当你们卧室还亮着灯时，你和你的好朋友可以临窗舞动手臂、打手势或以身体语言来交 流思想，但复杂一些的交流就有些困难了。而且一旦父母宣布“熄灯”，交流也就无法继续进 行了。

如何联系呢？用电话吗？ 10岁的小孩子屋里有电话吗？即使有，你们的谈话可能被偷听。 如果家里的电脑通过电话线联了网，它可能会提供无声的帮助，不过很不幸，它也不会在你 的房间里。

你和朋友采用的方法是用手电筒。所有的人都知道手电筒是为孩子们藏在被窝里看书而 发明的，它也适合在黑暗中用来交流。它无声无息，且光的方向性很好，不会从卧室的门缝 中泄露而使家人起疑。

用手电筒的光可以交谈吗？这值得一试。一年级你就学过在纸上写字母和单词，把这种 方法运用到手电筒上看起来也合情合理。你所需做的就是临窗而站，用光画出字母。画字母 ‘O’，就打开电筒，在空中画个圈，然后关上开关；字母‘ I’则是画竖直的一笔。但是你很 快发现这种方法行不通，当你注视来去飞舞的光柱时，会发现在脑海中将它们组合起来不是 件容易的事，这些光划成的圈圈杠杠太不准确了。

也许你曾经看过一部电影，影片中两个水手隔海用闪烁的光传递消息。在另一部电影中， 一个间谍用镜子反射阳光向一间屋子中被俘获的同伙发送讯息。这就给了你启发，你起先设 计一种简单的交流方法，使字母表中的每个字母与一定数目的闪烁相对应。 A闪一下， B闪两 下，C闪三下，如此递推， Z就闪烁 26下。BAD这个词由字母间有间隔的两闪、一闪、四闪组 成，这样你不会误以为它是闪七下的字母 G了。词间的停顿则比字母间的停顿时间稍长一些。

这看起来很有希望，采用这种方法的优点是你不需要在空中挥舞手电筒，只需对准方向 按开关就行了；缺点是你试图发送的第一个消息（“How are you?” ）就需要 131次闪烁，更 糟的是，你忘了定义标点符号，所以无法表示句尾的问号了。

这离问题的解决已经很近了，你想别人以前肯定也遇到过类似的问题，你解决它的思想 一定是正确的。为了解决问题，白天的图书馆之行使你发现了神奇的摩尔斯电码（morse code）， 这正是你想要的，即使你不得不重新学习如何“写”字母表中的字母。

以下就是区别：在你发明的体系中，每个字母是一定数目的闪烁，从闪烁一下的 A到闪烁 26的Z；而在摩尔斯电码中，有长短两种闪烁，当然，这会使摩尔斯电码更为复杂，但它在实 际应用中却被证实是更有效的。那句“ How are you? ”现在仅需 32次而不是 131次闪烁，而 且这还包含了问号。

在讨论摩尔斯电码的工作原理时，人们并不说“长闪烁”、“短闪烁”，他们使用“点（dot）”和“划（ dash）”，因为这样易于在印刷品上表示。在摩尔斯电码中，字母表中的每一

个字母与一个点划序列相对应，正如在下表中你所看到的：

尽管摩尔斯电码与计算机毫不相关，但熟悉它的本质却对深入了解计算机内部语言和软 硬件的内部结构有很大的帮助。

在本书中，编码或代码（ code）通常指一种在人和机器之间进行信息转换的系统（体系）。 换句话说，编码便是交流。有时我们将编码看成是密码 (机密)，其实大多数编码并不是的。大 多数的编码都需要被很好地理解，因为它们是人类交流的基础。

在《百年孤独》的一书的开篇，马尔克斯回忆了一个时代，那时“世界一片混沌，许多 事物没有名字。为了加以区别才给事物各个命名。”这些名字都是随意的，没有什么原因说明 为什么不把猫称为狗或不把狗称为猫。可以说英语词汇就是一种编码。

我们用嘴发出声音组成单词，这些词可以为那些听得到我们声音，理解我们所用语言的 人所听懂，我们称这种编码为“口头语言”或“语音”。对写在纸上（或凿在石头上、刻在木 头上或通过比划写在空气中）的词，还有一种编码方式，那就是我们在印刷的报刊，杂志和 书籍上看到的字符，称之为“书面语言”或“文本”。在许多语言中，语音和文本间有很强的 联系。例如在英语中，字母或一组字母与一定的读音相对应。

手势语言的发明帮助了聋哑人进行面对面的交流。这是一种用手和胳膊的动作组合来表 达词语中的单个字母、整个词及其基本概念的语言。对盲人来说，他们可以使用布莱叶盲文

（Braille）。这种文字使用凸起的点代表字母，字母串和单词。当谈话内容要被迅速地记录下 来时，缩写和速记是很有用的。

人们在相互沟通时使用了各种不同的编码，因为在不同的应用场合，其中的一些较其他 的更为简便。例如，语言不能在纸上存储，所以使用了文字；语言、文字不适合用来在黑夜 中安静地传递消息，故摩尔斯电码是一个方便的替代品。只要一种编码可以适用于其他编码 所不能适用的场合，它就是一种有用的编码。

以后将看到，计算机中使用了不同的编码来传递和存储数字、声音、音乐、图像和视频

（电影）。计算机不能直接处理人类世界的编码，因为它不能模拟人类的眼睛、鼻子、嘴和手 指来接收信息。尽管这些年来计算机的发展趋势使我们的桌上电脑具有捕获、存储、处理和 提供人类交流中所使用的各种信息的能，而且不论这些信息是视觉的（文字和图片）、听觉的

（语言、声音及音乐）还是两者的混合（动画和电影）。所有这些信息都要求使用它们自己的 编码方式，正如交谈需要使用人的某些器官（嘴和耳朵），而书写和阅读则需要使用另外一些

器官（手和眼睛）一样。 用手电筒发送摩尔斯电码时，电筒的开关快速地合开代表一个点，让电筒照亮稍长的时

间则代表一个划。举例来说，发送字母 A，要先快速地合开开关，然后再稍慢些合开。在发送 下一个字母前要有短暂的停顿。约定划的时间大约是点的 3倍。例如，如果点的照亮时间为 1 秒，那么划就是 3秒。（实际上，摩尔斯电码的传递速度要快得多。）接收者看到了短闪和长闪 就知道是 A。

摩尔斯电码中点划之间的间隔是极为关键的。例如，发送字母 A时，点划之间的间隔应与 一个点的时间大致相同（如果点的时间是 1秒，那么间隔的时间也是 1秒）。同一个词中字母间 间隔稍长，约为划的持续时间（或者 3秒，如果那是划的持续时间的话）。下面是单词“ hello”对应的摩尔斯电码，图中示意了字母间的间隔（隙）：

单词之间相隔大约 2倍于划的时间（如果划是 3秒，那么间隔即为 6秒）。下面是“ hi there”对应的编码（码字）：

手电筒开和关的时间长度并没有限定，这取决于点的时间长度，点长又由手电筒开关触 发的速度和摩尔斯电码发送者记忆电码的熟练程度来决定，熟练发送者的划也许与生手的点 等长。这个小问题会使接收电码有些困难，但在一两个字母之后，接收者通常就可以辨认出 哪个是点，哪个是划了。

粗看起来，摩尔斯电码的定义 — 这里所谓的定义是指与字母表中的字母相对应的各种 点划序列 — 与打字机字母的排列一样是随意的。但仔细观察后你会发现不完全如此，简短 的码字分配给了使用频率较高的字母，例如 E和T，爱赌博的人和“财富之轮”爱好者可能一 下就注意到了这个问题；不常用的字母如 Q和Z（它们在赌局中是 10点）则分配以较长的码 字。

几乎所有人都知道一点儿摩尔斯电码，国际遇险信号 SOS的摩尔斯电码为“三点三划三 点”。SOS并非缩写，选择它仅仅因为它有一个易记的摩尔斯电码序列。第二次世界大战中， 英国广播公司选用贝多芬第五交响曲中的片段作为节目前奏 — B A H 、 B A H 、 B A H 、 BAHMMMMM，听起来颇像摩尔斯电码中 V（代表Victory）的码字。

摩尔斯电码的一个缺点是它没有对大小写字母进行区分。除表示字母外，摩尔斯电码还 用5位长的码字来表示数字：

这些数字的码字看起来还有些规律 (相对于字母对应的码字而言)。大多数标点符号的码字 采用5位、6位或7位的码长：

对欧洲一些语言中的重音字母以及一些有特殊用途的缩写定义了特别的码字， SOS就是 这样一个缩写：发送时每个字母的码字之间仅有一点的时间间隔。

如果有特制的用于发送摩尔斯电码的手电筒，你和朋友之间的交流就方便多了。这种手 电筒除了常有的开关，还有一个按钮，按压按钮就可以控制电筒的亮灭。经过练习后，你们 每分钟可以发送和接收 5～10个单词。虽然仍比交谈慢（大概每分钟 100个词左右）但已足够 用了。

当你和朋友最终熟记了摩尔斯电码时（这是唯一精通发送接收的方法），你也可以用它代 替日常用的语言。为了达到最高的速度，可以发“滴（ dih）”音代表点、“嗒(dah)”音代表划。 摩尔斯电码同样也可将文字简化为用点和划两个符号表示。

以上的关键在于“两”这个词 —“滴、嗒”两个声音，“点、划”两种方式。实际上任 何两种不同的东西经过一定的组合都可以代表任何种类的信息。

第2章　编码与组合

摩尔斯电码由萨缪尔·摩尔斯（ 1791—1872）发明，本书后面会在多处提到他。摩尔斯 电码是随着电报机的发明而产生的，电报机我们以后也还要做详尽的说明。正如摩尔斯电码 很好地说明了编码的本质一样，电报机也提供了理解计算机硬件的良好途径。

大多数人认为摩尔斯电码的发送易于接收，即使你没有记住摩尔斯电码，也可以方便地 借助下面这张按字母顺序排列的表发送：

接收摩尔斯电码并将其翻译回单词比发送费时费力多了，因为译码者必须反向地将已编 码的“滴-嗒”序列与字母对应。例如，在确定接收到的字母是“ Y”之前，必须按字母逐个 地对照编码表。

问题是我们仅有一张提供“字母→摩尔斯电码”的编码表，而没有一张可供逆向查找的 “摩尔斯电码→字母”译码表。在学习摩尔斯电码的初级阶段，这张译码表肯定会提供很大的 便利。然而，如何构造译码表却毫无头绪，因为我们似乎无法找出这些按字母顺序排列的 “滴-嗒”序列的规律。

那么忘记那些字母序列吧，也许按照码字中“滴”“嗒”的个数来排列会是个更好的尝试。 例如，仅含一个“滴”或“嗒”的摩尔斯电码序列只可能代表 E或T这两个字母之一：

两个“滴”或“嗒”的组合则代表了 4个字母I、A、N、M： 三个“滴”或“嗒”的序列代表了 8个字母：

最后（如果不考虑数字和标点符号的摩尔斯电码），四个“滴”或“嗒”的序列则共代表 了16个字母：

四张表共包括2＋4＋8＋16 ＝ 30个编码，可与30个字母相对应，比拉丁字母所需的26个字 母还多了4个。出于这个原因，在最后一张表中，你可能注意到有 4个编码与重音字母相对应。 在翻译别人发送的摩尔斯电码时，上面 4张表提供了极大的便利。当你接收到一个代表特

定字母的码字时，按其中含有的“滴”“嗒”个数，至少可以跳到其对应的那张表中去查找。 每张表中，全“滴”的字母排在左上角，全“嗒”的字母排在右下角。

你注意到 4张表大小的规律了吗？每张表都恰好是其前一张表的两倍大小。这其中包含的 意义是：前一张表的码字后加一个“滴”或加一个“嗒“，即构成了后一张表。

可以按下面的方式总结这个有趣的规律：

点划数	码字数
1	2
2	4
3.	8
4	16

四张表中每张码字数都是前一张的两倍，那么如果第一张表含 2个码字，第二张表则含2×2

个码字，第三张表 2×2×2个码字。以下是另一种表达方式：

点划数	码字数
1	2
2	2×2
3	2×2×2
4	2×2×2×2

当然，如果遇到数的自乘，可以用幂表示，例如 2×2×2×2可以写成 24。数字 2、4、8、 16分别是 2的1、2、3、4次幂，因为可以用依次乘 2的方法将它们计算出来。由此我们的总结 还可以写成下面的方式：

点划数	码字数
1	21
2	22
3	23
4	24

这张表简单明了，码字数是 2的次方，次方数目与码字中含有的“滴”“嗒”数目相同。 我们可以把表总结为一个简单的公式：

码字数＝2“滴”与“嗒”的数目 很多编码中都用到 2的幂，在下一章中我们会看到另一个例子。 为了使译码的过程更为简便，可以画出如下一张树形图：

这张表表示出了由“滴”与“嗒”的连续序列得出的字母。译码时，按箭头所指从左到 右进行。例如，你想知道电码“滴-嗒-滴”代表的字母，那么从最左边开始选择点，沿箭头 向右选择划，接着又是点，得出对应的字母是 R，它写在最后一个点的旁边。

如果认真考虑，会发现事先建立这样一张表是定义摩尔斯电码所必需的。首先，它保证 了你不会犯给不同的字母相同码字的错误！其次，它保证你使用了全部的可用码字，而没有 使“滴”与“嗒”的序列毫无必要的冗长。

我们可以加长码字至 5位或更长， 5位长的码字又提供了额外的 32（2×2×2×2×2或25）

个码字。一般而言，这就足够 10个数字和 16个标点符号使用。实际上，摩尔斯电码中的数字 确实是5位的，但在许多其他编码方式中， 5位码字常用于重音字母而不是标点符号。

为了包含所有的标点符号，系统必须扩充至 6位表示，提供 64个附加编码，此时系统可表 示2+4+8+16+32+64共126个字符。这对摩尔斯电码而言太多了，以至于留下许多“未定义”的码字。此处“未定义”指不代表任何意义的码字，如果在你接收的摩尔斯电码中有未定义 的码字，就可以肯定发送方出了差错。

由于推出了下面这条公式：

码字数＝2“滴”与“嗒”的数目 我们就可以继续导出更长的码字位数所代表的码字数目。很幸运，我们不必为确定码字数目 而写出所有可能的码字，我们所要做的不过是不断地乘 2而已：

点划数　　码字数

1　　21 = 2

2　　22 = 4

3　　23 = 8

4　　24 = 16

5　　25 = 32

6　　26 = 64

7　　27 = 128

8　　28 = 256

9　　29 = 512

10　　210 = 1024

摩尔斯电码被称为 二元码 （binary code ），因为编码中仅含“滴”和“嗒”。这与一个硬 币很相似，硬币着地时只可能是正面或反面。二元事物（例如硬币）、二元编码（例如摩尔斯 电码）常常用 2的乘方来描述。

上面所做的对二元编码的分析在数学上的一个分支—组合学或组合分析 里只能算是一个 简单的练习。传统上，由于组合分析能够用来确定事件出现的几率，例如硬币或骰子组合的 数目，所以它常用于概率统计，但它也同样有助于我们理解编码的合成与分解。

第3章　布莱叶盲文与二元编码

摩尔斯不是第一个成功地将书写语言中的字母翻译成可解释代码的人，他也不是第一个 因为其编码而受到人们纪念的人，享有这个荣誉的是一个晚摩尔斯 18年出生的早慧的法国失 明少年。虽然人们对他的生平所知甚少，但就是所知的这一些却足以给后人留下深刻印象。

路易斯·布莱叶 1809年出生于法国的 Coupvray，他的家 乡在巴黎以东 25英里，父亲以打造马具为生。 3岁时，在这 个本不该在父亲作坊里玩耍的年龄，小布莱叶意外地被尖头 的工具戳中了眼睛。由于伤口发炎，感染了另一只眼，他从 此双目失明。布莱叶原本注定在贫困潦倒中度过一生（正如 那时大多数盲人一样），但他的聪明才智和求知欲不久即显 露了出来。在本地牧师和一位学校老师的帮助下，布莱叶和 其他孩子一道上了学， 10岁那年又前往巴黎的皇家盲人青年 学院学习。

盲人教育的一大障碍就是他们无法阅读印刷书籍。

Valentin Haüy(1745—1822)，巴黎学校的创始人，发明了一 种将字母凸印以供触摸阅读的方法。但这种方法使用起来较为困难，并且只有很少的书籍用 这种方法“制造”。

视力正常的 Haüy陷入了一种误区。对他而言，字母 A就是A，它看起来（或感觉起来）也 必须像是个 A。（如果给他手电筒作为交流工具，他也会试图在空气中画出字母的形状，而我 们已经知道这种方法并不有效。）Haüy也许没有意识到一种与印刷字母完全不同的编码会更适 于盲人使用。

另一种可选的编码有一个出人意料的起源。法国陆军上尉 Charles Barbier在1819年发明了 一种他自称为 écriture nocturne的书写体系，这种体系也被称为“夜间文字。他使用厚纸板上 有规律凸起的点划来供士兵们在夜间无声地传递口信（便条），士兵们使用尖锥状的铁笔在纸 的背面刺点和划，凸起的点可以用手指感觉阅读。

Barbier体系的问题是其过于复杂。 Barbier没有用凸起的点来代表字母表中的字母，而是 用其代表声音。这样的系统中一个单词通常需要许多码字表达。这种方法在野外传递短小消 息还算有效，但对长一些的文章而言则有明显不足，更不要说是整本的书籍了。

布莱叶在 12岁时就熟悉 Barbier方法了，他喜欢使用这些凸点，不仅因为它们易于用手指 阅读，更因为它们易于书写。教室里拿着铁笔和纸板的学生可以记笔记供课后阅读。布莱叶 勤奋地工作试图改进这种编码系统。不出 3年（在他 15岁时），他创建了自己的系统，其原理 直到今天还在使用。布莱叶系统有很长时间仅局限在他所在的学校使用，后来它逐渐扩散到 世界各地。 1835年，布莱叶染上了结核病。 1852年，在他 43岁生日过后不久，他便去世了。

时至今日，布莱叶系统的改进版本甚至可以与有声录音带竞争，它为盲人提供了与书写 世界联系的途径。布莱叶方法仍是适于既聋又盲的人阅读的唯一方法。近来年，随着电梯和

自动语言机的普及，布莱叶系统更加广为人知。 本章将剖析布莱叶编码的编码方法及其工作原理，不过不必真正学习布莱叶编码或记住

任何东西，我们只要大概了解一下编码的本质就行了。 布莱叶编码中，普通书写语言的每个字符—具体而言如数字、字母和标点符号—都被

编码成局限在 2×3小格中一个或多个凸起的点。这些小格一般被标记为 1～6：

在当今实际使用中，特殊的打字机或刻印机可以在纸上打出布莱叶编码中的小点。 由于在书中夹印几页布莱叶编码极其昂贵，我们使用了在通常印刷品中常用的布莱叶码

的表示方法。在这种表示方法中，小格中的 6个点全部印刷出来，大点代表小格中的凸起点， 小点则代表平滑的点。例如下图中的布莱叶字母中，点 1、3、5是凸起的，点 2、4、6则没有：

在这里吸引我们的问题是：点是二元的。一个特定的点不是凸起的就是平滑的，那么 6个 点的组合数目就是 2×2×2×2×2×2，或64(26)。

因此，布莱叶编码系统可以代表 64个不同的码字。以下就是所有的 64个码字：

如果我们发现布莱叶编码只用了 64个码字中的一部分，我们会疑问为什么 64个码字中有

一些不被使用；如果发现布莱叶编码使用了多于 64个的码字，则又会让人怀疑我们是否神志 清醒或数字计算的真实性， 2乘2是等于4吗？

分析布莱叶编码，还是从基本的小写字母开始：