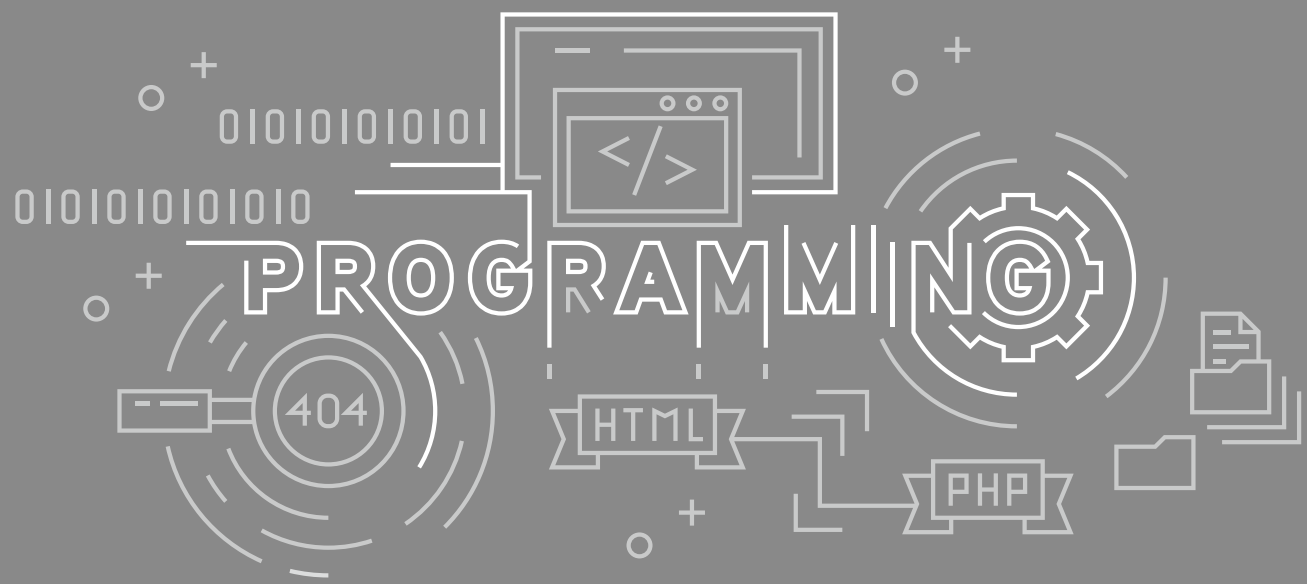


第一章

导论

本章导言

“译者”（Translator）和“编程”（Programming）这两个词放在一起兴许会让翻译专业学生觉得很陌生，因为绝大部分正在学翻译和正在做翻译的人都不曾想过有一天自己会去写代码编程序。本书将展示如何开发简易的程序来解决翻译过程中遇到的技术问题。不论你是否曾被他人或自己打上“文科生”的标签，不论你是否曾笃定自己此生与编程无缘，相信本书内容会让你直观感受到技术给翻译服务带来的便利。

本章内容将重点介绍在今天的翻译行业发展背景下译者为什么要学习编程，以及如何学习编程。

1.1 翻译简史

“语言的起源”是科学界的一道难题[1]，语言学家、人类学家、考古学家等尝试通过各种方法来探究其根源，但始终没有达成共识。据估计[2]，在我们今天所处的星球上有多达六七千种语言，我们习惯于将两种语言之间的转换称为“翻译”，而翻译又可分为“口译”“笔译”和“手语翻译”。

翻译职业是世界上最古老的职业之一，如果在中国历史中去追寻翻译职业的痕迹，可以在许多史料中看到相关的记载。

如记录周朝（前 1046 年—前 256 年）职官制度和相关礼制的《周礼》一书中写道：“象胥，掌蛮、夷、闽、貉、戎、狄之国使，掌传王之言而谕说焉，以和亲之。若以时入宾，则协其礼与其言辞传之。”

儒家经典《礼记》中有这样的描述：“五方之民，言语不通，嗜欲不同，达其志，通其欲。东方曰寄，南方曰象，西方曰狄鞮，北方曰译。”唐朝经学家孔颖达这样解释[3]：“其通传东方之语官谓之曰寄，言传寄外内言语。通传南方语官谓之曰象者，言放（仿）象外内之言。其通传西方谓之狄鞮者，鞮，知也；谓通传夷狄之语与中国相知。其通传北方语官谓之曰译者，译，陈也；谓陈说外内之言。”

简单而言，在我国古代，“寄”“象”“狄鞮”“译”都是翻译人员的称谓，区别在于他们的工作语言不同。“胥”是指官府中有才智的小吏[4]，所以“象胥”是一种官名。

---

1 Christiansen, M. H. and S. Kirby.（2003）. Language evolution: the hardest problem in science? In M. H. Christiansen and S. Kirby (eds.), *Language Evolution*（pp. 1-15）. Oxford: Oxford University Press.

2 http://tac-online.org.cn/ch/tran/2010-07/02/content_8255857.htm

3 http://m.guoxuedashi.com/diangu/32763x/

4 http://agri-history.ihns.ac.cn/scholars/yxl/yxl135.htm

不过，中国历史上第一次大规模的翻译活动源自两汉时期佛经的翻译，历时 1000 余年。“翻译”一词也在这一时期出现，唐代史志目录《隋书·经籍志》[1] 写道：“章帝时，楚王英以崇敬佛法闻，西域沙门，赍佛经而至者甚众。永平中，法兰又译《十住经》。其馀传译，多未能通。至桓帝时，有安息国沙门安静，赍经至洛，翻译最为通解。”

自此之后，在中国的历史上，“翻译”一词便替代了前述的“寄”“象”“狄鞮”“译”等词，沿用至今。

1.2 翻译行业发展现状

许多翻译爱好者都会因翻译职业的悠久历史以及他们对语言本身的热爱而下定决心做一名翻译，但又因不理解当今翻译行业发展的现状而对翻译职业产生了误解。

中国今天的翻译行业又称“语言服务业”或“翻译服务业”，是伴随改革开放诞生的新兴服务行业。改革开放作为中国的基本国策，在正式推行后开放了外商直接投资，这种类型的投资目的在于利用中国廉价的劳动力、原材料，来生成产品，并将产品销售到中国以外的地区。在这个过程中，翻译服务并未产生太大的价值，直到跨国公司开始进行市场寻求型投资，将其产品销售到中国，为了适应中国的语言、文化、政策、法律法规而对产品进行适应性改变，由此催生了专业化的翻译行业。20 世纪 70 年代开始升温的跨国公司市场寻求型投资是语言服务业兴起的原动力[2]。但此时的语言服务业最重要的业务领域还是传统的笔译和口译业务，与计算机技术还没有太直接的联系。

谈起计算机技术，1946 年，世界上第一台电子计算机“ENIAC”（Electronic Numerical Integrator And Computer，译为“电子数字积分器和计算机”）在美国宾夕法尼亚大学问世。彼时的计算机体积庞大、运算速度低、成本高，主要用于科学计算，寻常百姓尚不能负担这样的计算机价格。到了 20 世纪 70 年代，不同类型的个人计算机接连问世，美国肯巴克（Kenbak）公司在 1971 年推出了一台名为“Kenbak-1”的计算机，号称是世界上第一款个人计算机。这时的计算机内部已经借助大规模集成电路极大提升了运行速度，从 20 世纪 40 年代的每秒几千次到几万次提升至当前的每秒上千万次到亿次。

当个人计算机开始进入企业、学校和家庭的时候，人们也开始进一步尝试将计算机与自己的日常工作结合在一起。1980 年，荷兰人 Jaap van der Meer 和 Simon Andriessen 在阿姆斯特丹创建了一家名为“INK”的翻译公司，他们在 1982 年获得了

1 https://ctext.org/wiki.pl?if=gb&res=386407&searchu=%E7%BF%BB%E8%AF%91&remap=gb

2 王传英 . 2014. 语言服务业发展与启示 . 中国翻译，35（02）：78-82.

IBM DisplayWrite System 的文档翻译合同。为了能够快速完成大批量、多语言的翻译项目，他们开发了“INK TextTools”[1] 工具。开发者认为，翻译质量受到诸多因素影响，其中最重要的是术语的一致性和准确性。“INK TextTools”包括多个组件，其中“Texan”用于创建和更新翻译文本中的术语，“LookUp”嵌入在当时的字处理工具 WordPerfect 和 WordStar 中，用于查询“Texan”中管理起来的术语。因此，本质上来说“INK TextTools”是一款术语管理工具。

可惜，“INK TextTools”因为种种原因没能最终成功商业化，但是它却成为了今天全世界的译者都在广泛使用的计算机辅助翻译工具塔多思（SDL Trados）的基础。

目光转至同时期的中国，从 1989 年开始，中国经济从过热进入低谷，宏观经济政策从抑制过热转向了启动经济。世界上少数国家对中国经济的封锁和改革的停滞，使中国的经济发展出现了停滞[2]。但在 1992 年，邓小平同志南巡，中国的改革开放进入“新时代”，中国的语言服务行业也迎来了巨变。1992 年前后，Oracle（甲骨文公司）、Microsoft（微软公司）、IBM（国际商业机器股份有限公司）等世界著名计算机技术巨头争先在中国成立办事处或中国分公司，中国的计算机市场进入快速发展期。1993 年微软发布视窗 3.1（Windows 3.1）简体中文版，标志着微软正式进军中国软件市场。为了降低软件国际化过程中语言翻译技术和人力资源等成本，国际大型软件开发商倾向于将软件翻译的业务交给专业的软件翻译公司，这类公司不仅要高质量完成翻译工作，还需要有足够的技术实力与软件开发商对话——而这是传统的翻译公司无法做到的。这类公司也因此演化为软件本地化（Software Localization）公司，他们所用的技术手段称为软件本地化技术，软件本地化技术的发展促成了专业语言服务商从语言翻译服务到软件本地化服务的转型[3]。

1.3 译者培养现状

20 世纪 90 年代软件本地化技术的广泛应用给语言服务人才培养带去新的挑战，爱尔兰有欧洲“硅谷”之称，IBM、微软、甲骨文等世界级企业的软件本地化工作均在爱尔兰完成，因此爱尔兰利默里克大学（University of Limerick）早在 1997 年就开始设立“本地化技术”专业[4]，在两学期的课程中要求学生学习“本地化导论”“计算机编程导论”“语言工程基础”“项目管理导论”“国际化最佳实践”“本地化工程”“质量与本地化”和“本

1 http://www.mt-archive.info/70/LangTech-1988-Olsen.pdf

2 http://news.ifeng.com/special/30economy/

3 杨颖波，王华伟，崔启亮 . 2011. 本地化与翻译导论 . 北京：北京大学出版社 .

4 https://multilingual.com/multilingual-newsletter/?nl_id=267

地化工具与技术”等课程。今天，这个专业已经更名为“多语言计算和本地化”，课程内容更加丰富，涉及“本地化项目管理”“高级语言工程”“本地化中的主要问题”“本地化标准和最佳实践”“本地化流程自动化”“翻译技术”“国际化最佳实践”等专业课程[1]。

由于20世纪90年代“本地化”概念才在中国开始萌芽，我国的翻译学科发展和翻译人才培养历程中，“本地化”很晚才出现，如今的本地化专业人才培养还要得益于“中国入世”。

2001年12月11日，中国正式加入世界贸易组织（WTO），中国的语言服务行业迎来了新的机遇。

中国入世后不久，上海外国语大学、广东外语外贸大学和北京外国语大学分别于2004、2006、2008年在外国语言文学一级学科内自主设置了翻译学学位点，培养翻译学的博士生和硕士生[2]。

2006年，教育部正式批准复旦大学、广东外语外贸大学与河北师范大学设立翻译本科专业（BTI, Bachelor of Translation and Interpreting）。同年，北京大学软件与微电子学院和北京大学计算语言所共同创建语言信息工程系，2007年9月第一批统招计算机辅助翻译专业（CAT, Computer-Aided Translation）学生入学，学生专业背景为计算机专业或语言专业，毕业后授予学位为“工程硕士”。

“在我国对外文化交流和社会发展的新形势下，社会对翻译的高级专门人才在数量和质量上的要求急剧上升，旨在让中国文化走出去的汉译外问题甚至提到了国务院办公机构的议事日程上。为适应这一形势，2007年1月，国务院学位委员会第23次会议以全票通过设置‘翻译硕士专业学位’（MTI，Master of Translation and Interpreting）。”[3]

截至2018年2月，全国翻译硕士培养单位246所，本科培养单位252所，但在翻译硕士和翻译专业本科的课程内容中，“本地化”依然是个陌生词，与利默里克大学的“本地化技术”专业相似的仅有北京大学的计算机辅助翻译专业，而此专业培养的是“工程硕士”，并非“翻译专业硕士”。

在很长一段时间内，由于中国传统的翻译公司和培养翻译人才的高校所处的“翻译行业”与“本地化行业”是独立发展的，“本地化”并没有引起太多重视，翻译行业相关的企业和学校内均没有太多专业本地化人才参与翻译业务和翻译人才培养，反而是“技术替代论”的盛行使得一大批从事翻译工作的人排斥技术，拒绝使用技术。

---

1 https://www.ul.ie/international/sites/default/files/Postgrad_Prospectus.pdf

2 http://www.china.com.cn/culture/zhuanti/zgyxd6/2009-11/09/content_18853197.htm

3 杨晓荣 . 2008. 翻译专业指什么：正名过程及正名之后 . 中国翻译，(03)：31-34.

实际上，彼时的“本地化技术”完全是机助人译的计算机辅助翻译技术，更多是通过工具的应用帮助译者提升翻译效率，流程自动化工具辅助译者完成简单的翻译质量控制，即便那时的机器翻译技术发展迅猛，也还达不到替代译者的地步。

直到 2016 年 9 月末，这样几条媒体报道在互联网上不胫而走，《谷歌发布神经机器翻译，翻译质量接近笔译人员》《谷歌神经机器翻译再突破：实现高质量多语言翻译和 zero-shot 翻译》《AlphaGo 神经网络应用于 Google 翻译，将接近人类水平》《谷歌再掀机器翻译革命，人工翻译何去何从？》，这些新闻震动了翻译行业从业人员和高校翻译专业师生，许多人发自内心的惊呼“要被替代了！”

2016 年 9 月 27 日“谷歌大脑小组”的科学家发表了一篇网文：《一个产品规模的用于机器翻译的神经网络》，提及谷歌最新的神经机器翻译系统 GNMT（Google Neural Machine Translation System，谷歌神经翻译系统）取得重大突破，文中一些描述“与经典的基于短语的统计机器翻译模型相比，GNMT 将若干关键语言对之间的翻译错误率显著降低了 55% 到 85%”“法英、英西语言对的机器翻译质量已非常接近人工翻译质量”，经过媒体解读后，谷歌机器翻译随即变身成翻译从业人员的“终结者”，并与此前战胜人类顶尖围棋选手的人工智能系统 AlphaGo 相提并论。

人们本以为这样的新闻不过是昙花一现，关于机器翻译质量超过人工翻译质量的新闻很快就会从公众视野中消失，但没想到神经机器翻译的技术自此飞速演进，依靠新硬件、新技术和高质量的大数据，在越来越多的专业领域媲美人工翻译。

在这样的背景下，传统的翻译专业应当如何发展，新时代的译者应当具备怎样的能力，翻译教育过程应当进行怎样的升级改良，等等，这些都是亟待解决的新问题。

1.4 翻译专业（本地化方向）人才培养现状

2015 年 4 月，北京语言大学高级翻译学院召开了本科翻译专业（本地化方向）论证会。中国翻译协会本地化服务委员会秘书长崔启亮、北京大学软件与微电子学院语言信息工程系副主任俞敬松、原北京大学翻译硕士教育中心王华树、北京语言大学信息科学学院语言监测与社会计算研究所刘鹏远、北京语言大学信息科学学院大数据与语言教育研究所于东、资深本地化专家师建胜等 6 位在本地化、自然语言处理、机器翻译、翻译教育领域的校内外知名专家与高级翻译学院荣誉院长刘和平、副院长许明、本科专业负责人刘丹、笔译系主任梁爽，以及卢宁、韩林涛、校教学督导组成员沈素琴共同参与此次论证会。

此次论证会后，北京语言大学高级翻译学院建立起了包含语言课程、翻译课程和技术课程三个主要模块的专业课程体系，学生在入学后先后学习相应难度的语言类课程、

翻译类课程和信息技术类课程，并充分利用翻译实践、上机实践、企业实践等机会与社区内的互联网公司、语言服务企业以及本校信科学院的计算机专业学生共同探索网站建设、移动应用开发、游戏开发过程中的多语言本地化和国际化解决方案，为中国的信息技术产品、互联网产品走出去和国际产品引入中国做好专业知识储备。

2015 年 9 月，北京语言大学高级翻译学院翻译专业（本地化方向）第一届学生正式入学。该专业方向学生所学课程由五大细分模块构成：语言基础课程、计算机类课程、翻译类课程、本地化课程和其他课程。

语言基础课程包括英语读写、英语听说、英语国家概况、科技英语等；计算机类课程包括数学综合、初级编程、计算导论、数据库原理、网站设计与开发、数据结构、语言计算与信息检索、计算机辅助翻译等；翻译类课程包括翻译理论与实务、基础笔译、非文学翻译、经贸翻译等；本地化课程包括翻译与本地化实践、本地化概论、本地化实务等；其他课程包括思想道德修养、中国近现代史纲要、心理健康教育、军事理论等由学校面向全体学生开设的公共选修课程。

若按培养阶段来分：

本科一年级课程以语言基础课程和计算机基础课程为主，主要是帮助学生夯实语言基础、计算机基础，比如加强听说读写训练，了解数学基础知识、计算机基础知识、编程基础知识等。语言类课程由高级翻译学院教师承担，计算机基础课程由信息科学学院教师承担；

本科二年级开始，学生将会继续夯实语言基础和计算机基础，并开始培养较为基础的翻译能力；

本科三年级开始，侧重翻译教学和计算机教学的结合，并且开始引入本地化概论、本地化实践等真正与本地化行业所需专业知识相关的课程；

本科四年级则继续加强学生的专业学习，引导学生根据能力和兴趣接触行业，为学生的未来发展提供更多支持。

截至 2018 年，该方向同学连续三年成功申请四项国家级大学生创新创业训练计划项目（其中三项已评为国家级优秀项目），如：

“互联网 +”时代多语言电子杂志制作和发布技术探究，2016 年

内容：“互联网 +”时代的新媒体技术迅速发展，信息传播的媒介层出不穷，新媒体平台以全新的方式改变了人们沟通和交流的方式。在“互联网 +”时代，如何借助现代信息技术的力量讲好中国故事、推动中国文化走出去是当下的热点话题之一。其中的一步重要举措是信息的多语化。多语言信息的载体之一——多语言电子杂志的制作和发布是一项值得探究的课题。

中国文化微视频外译，2017 年

内容：适逢“互联网 +”时代的新媒体技术迅速发展，如何借助现代信息技术和新媒体平台推动中国文化走出去是当下的热点话题之一。目前，国内已有许多制作精良、内容翔实的中国文化系列节目（如央视出品的《文化中国》），而海外社交媒体平台上类似系列视频较少，因而输出本地化后的国产优质文化视频将会是中国文化“走出去”的有效途径。作为北语本地化专业的学生，我们通过一年多的“语言 + 技术”学习，已熟练掌握汉英翻译及视频本地化的相关技能，在这样的时代背景和专业背景下，我们希望能充分运用专业特长，学以致用，组建“中国文化微视频外译”项目组，将与中国文化相关的系列视频本地化，发布到国内外社交平台上以供更多人学习研究，在为传播中国文化贡献力量的同时提高专业翻译能力。

中外学生语言技术与本地化体验营，2017 年

内容：在全球化席卷世界的今天，如何借助现代信息技术的力量展示出当代语言技术发展的成果，促进中外信息技术成果的交流是当下的热点话题之一。而建立中外语言技术体验营，如何以“教学 + 体验”的方式，将中国当代语言技术发展中最有特色的成果展示给外国留学生亦是一项值得探究的课题。作为国内本科阶段第一批学习该专业的学生，我们计划结合北语留学生众多的环境优势，与北语中外语言服务人才培养基地合作，以“教学 + 体验”的方式，初期介绍本地化的基本概念、本地化流程、中国本地化发展的现状和成果，中期组织留学生体验本地化过程中运用的语言处理技术，后期带领留学生体验并实践一个完整的本地化项目，达到将中国当代语言技术发展的特色成果展示给外国留学生的目的。

机器翻译辅助英语文本阅读平台，2018 年

内容：英语作为一门国际通用语，其重要性不言而喻。学习并掌握英语的听说读写能力，已经成为衡量大学生文化素养的重要标准。提升英语读写能力的一个有效途径就是阅读英语文章，但阅读时会受到很多因素的限制，比如词汇量、阅读速度、文化背景知识储备等。随着人工智能技术的发展，机器翻译的质量越来越高。在英语文章的阅读中，机器翻译可以承担一个重要的角色，以帮助英语学习者提高阅读效率与阅读质量。本项目计划基于机器翻译搭建一个在线阅读平台，面向有英语学习需求的用户。机器翻译辅助英语文本阅读平台，是基于网页的一个在线文本阅读与词汇管理系统。

这些项目共同的特点是：把语言（翻译）和技术结合在一起。

2019 年 7 月，该专业本科学生顺利毕业，就业率 100%，位居当年全校就业率排名榜首。

1.5 译者为什么要学习编程

作为北京语言大学翻译专业（本地化方向）的负责人，笔者一直对翻译专业本科生同时学习翻译和计算机技术抱有极大的热情和希望，因为笔者在四年的英语专业学习和三年的计算机辅助翻译专业学习后受益颇多，所以也希望该方向同学也能同样享受“语言 + 技术”带来的红利。

今天我们在谈及“翻译行业”和“本地化行业”时，已经用“语言服务行业”来统称我们所处的这个行业，全球语言服务需求的快速增长和新技术的不断应用推动了传统翻译行业与本地化行业的深度融合，语言服务行业的发展也受到了诸多因素的制约，核心因素当属技术和人才。

1) 技术因素

目前对语言服务行业冲击最大的是机器翻译技术的发展。然而，现在无论是本地化行业的资深人士，还是翻译行业的资深人士，真正理解目前最新的“神经机器翻译技术”的人不多，因为要想理解神经机器翻译技术的原理，需要有一定的数学基础、一定的计算机技术基础、一定的语言基础等。如果这个行业里最优秀的人都不懂对自己所在行业影响最大的技术，这个行业的前景就岌岌可危了。

2) 人才因素

如今，越来越多的外语院校和翻译学院开始教学生怎么去使用计算机辅助翻译软件，这些软件专门服务于企业的复杂业务流程开发，每个都需要花费至少 10 个小时，才可以初步掌握 20% 的主要功能，对于文科生和文科背景的教师而言，无疑是有难度的。如果高校培养的语言服务人才无法适应语言服务行业技术升级，不具备掌握新技术应用的能力，语言服务行业的发展就会受到阻碍。

在北京语言大学高级翻译学院，我们尝试将翻译专业（本地化方向）当作“胚胎干细胞”来培养，学生掌握了一定的语言能力和计算机技术应用能力后，根据自己的兴趣来决定未来如何借助“语言”和“技术”这两项工具来发展。

本教材并非鼓励全部翻译专业学生学习编程，而是针对那些想学编程、有时间学编程的翻译专业同学和翻译爱好者。2018 年 7 月 20 日，国务院印发《新一代人工智能发展规划》[1]，其中明确指出人工智能（AI, Artificial Intelligence）成为国际竞争的新焦点，

1 http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

应“支持开展形式多样的人工智能科普活动，鼓励广大科技工作者投身人工智能的科普与推广，全面提高全社会对人工智能的整体认知和应用水平。实施全民智能教育项目，在中小学阶段设置人工智能相关课程，逐步推广编程教育，鼓励社会力量参与寓教于乐的编程教学软件、游戏的开发和推广。”由此可预见，随着中小学不断推广编程教育，几年后入学的翻译专业本科生将具备一定的编程基础，如若不尽早在今天的翻译专业本科阶段推动编程教育，那么这些本科生一入学便会进入“编程教育断层”，错失在高等教育阶段深入学习人工智能技术的机遇。

由于国内目前还没有专门针对翻译专业学生的计算机技术教材，因此，经过四年的教学实践，笔者撰写了这本《译者编程入门指南》，试图通过翻译专业学生可以接受的方式介绍一些简易的、与翻译可以结合到一起的编程知识，以期帮助更多翻译专业学生提升信息技术素养，更好地适应未来的语言服务行业。

1.6 译者如何学习编程

对于没有任何计算机基础的翻译专业同学而言，阅读本书学习编程的目的并非是复写本书的每一行代码。本书涉及的代码将尽量简化，重点在于展现代码与“翻译”的关系，让读者直观感受到许多看似复杂的功能其实不难实现。本书的根本目的在于“译者编程入门”，而非对计算机技术和编程技能进行系统性讲解。在入门过程中，译者需要靠自身努力去写自己的代码，实现自己想实现的功能，解决自己在翻译学习、实践和研究中遇到的问题；而在入门完成后，如果对编程产生了浓厚的兴趣，希望继续深入学习，仍应该根据自己的兴趣方向系统性学习计算机技术的原理和编程技能。

作为一名翻译专业学生，编程技能仅是翻译技能的辅助。目前国内开设翻译专业的学校中，所有学校均是以培养学生的翻译能力为根本目的。在完成编程入门后，译者应将所学编程技能与翻译学习、翻译实践、翻译研究等结合起来，一方面可以尝试将本书中所讲解的案例应用于实际问题的解决，另一方面可以尝试与计算机专业背景的教师和同学合作，开展跨学科研究，发现新问题，寻找新方法，提供新的解决方案。

我们今天培养的是“未来的译者”（Future Translator），而未来的优秀译者绝不是“技术盲”，他们在未来之所以优秀，是与他们还在上大学时就学了编程是密不可分的。未来的语言服务行业将会更加技术化，对于众多学习翻译的同学而言，面对这样的行业变化，大家通过编程学习可以：

1) 把自己变成翻译水平很高的人；

2) 把自己变成管理能力很高的人；

3) 把自己变成技术思维很强的人。

这样一来，未来的译者可以利用技术手段提高自己的学习效率和学习质量，利用技术管理好自己的学习、生活和工作。翻译专业的学生不需要自己去开发多么高深的软件，但至少应当知道什么技术对自己是有用的，善于应用技术解决实际问题。

当大家都已经认为外语是一种工具的时候，大家不妨也把“翻译”“技术”和“管理”也置入自己的工具包内。翻译专业学生毕业后不一定必须去翻译公司做翻译，或者去本地化公司做本地化，而是可以依靠自己的“翻译”“技术”和“管理”技能去做更多的事情。比如，可以去一家农产品公司引进国外先进的农业公司管理经验和技术，推动农业的“互联网 +”，比如可以去一家中国建筑公司推动中国先进的建筑行业管理经验和技术在国外开花发芽，推动中国先进技术走出去。

从以上角度来思考，你会发现未来的语言服务行业不会是一个孤立的行业，而是一个与其他任何一个前沿领域都可以有机整合的行业，这个行业可以通过“语言”和“技术”这两块引擎成为未来核心的服务业。

1.7 如何使用本书

编程学习不能仅靠读书，而应以操练为主。本书在编写过程中完全从初学者零基础学习编程的角度设置章节内容，各章节之间以案例形式循序渐进，讲解编程基础知识，详细解释每段代码的编写方法和目的，将系统的编程基础知识点贯穿在与翻译实践密切相关的案例中。

本书从第二章开始正式讲解编程基础知识，正文部分共分为八章：

第二章：从万维网的起源开始讲起，学习如何制作简单的双语网页。

第三章：在读者对网页制作有初步认识后，为学习者介绍如何在满足学习要求的计算机上搭建编程环境。

第四章：双语术语是绝大部分外语学习者和翻译学习者接触过的概念，本章着重介绍如何开发一个简易的在线双语术语库，完成本章学习后读者将可以在自己的计算机上开发一个属于自己的“简明双语词典”，摆脱 Excel 表格的束缚。

第五章：一个仅能用于查询的双语术语库还无法满足译者的查词需求，本章将从术语表的增、删、改、查、登录、退出等几个方面详细介绍如何开发一个有登录退出功能的在线术语管理工具。

第六章：译者在做翻译时不仅需要查双语术语，还需要查双语的例句，在翻译实践中存储双语句子的工具一般称为“翻译记忆库”。在本章的学习中，读者将学习如何开发简易的在线翻译记忆库，学成之后将可以在自己的计算机上建立小规模的双语平行句库。

第七章：有了术语库和翻译记忆库这两大“法宝”后，译者便可以开启翻译实践之路。在本章中读者将学习翻译实践过程中字数统计的基本原理，了解如何开发一个简易的字数统计工具。

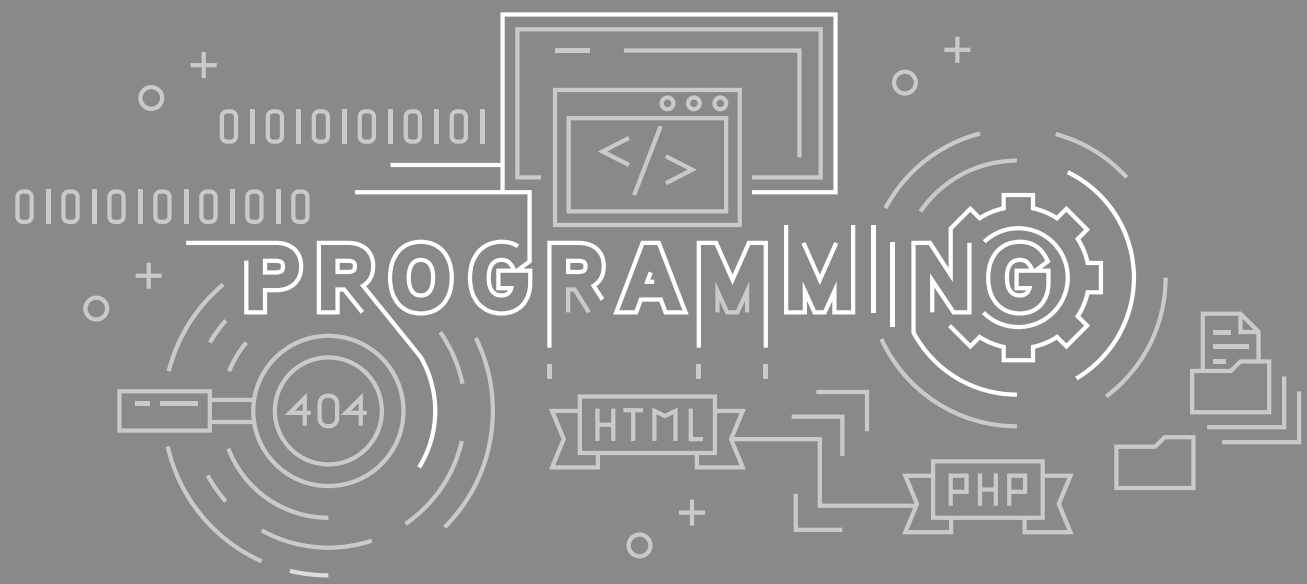
第八章：知道要翻译的文本有多少字了，译者便可以开始做翻译了。但是在翻译实践过程中，译者处理的文本内有许多特定类型的字符串，经常是标点符号、中文、英文等不同的字符混合在一起，难以处理。在本章中，读者将学习如何在译前准备中使用正则表达式来处理单语文本和双语文本，学成之后便可以轻松处理中英混杂的双语数据。

第九章：在今天这个谈“机译”即色变的时代，对于外语学习者和翻译学习者而言，机器翻译其实是“工具”，而非“敌人”，在本书的最后一章中，读者将学习如何开发一个“一文多译”工具来同时使用“百度翻译”和“有道翻译”这两款国内领先的机器翻译引擎，学成之后大家就不必在做翻译时每次都去这两个工具的官网上查看机器翻译结果，而是能在自己开发的“一文多译”网站上一键获取多个机器翻译引擎的译文。

以上就是本书正文八个章节的重点内容，读者在学习本书时切勿“跳章学习”，而应按照章节安排逐章推进，确保每章内容均理解清楚后再向后续章节推进。本书每章内容的学习都建立在前序章节的学习基础之上，为方便读者回顾每章的重要代码知识，在本书的附录 B 中读者可以通过本书“核心代码功能速查表”了解每章核心代码的主要功能。

由于教材的主要对象是零基础入门学习的译者，基于笔者在北京语言大学高级翻译学院的编程教学实践，本书并未参照传统编程教材的章节顺序编写内容，许多编程基础知识也未在本书中涉及，目的是让文科背景的初学者能够直观体验编程给翻译实践带来的益处，了解常见计算机辅助翻译技术的基本原理。为了让读者更好地理解计算机专业术语，本书还在内容撰写过程中专门给出了所涉及计算机专业术语的英文全称及部分术语的来源，并在本书附录 C 中一一列举，读者也许会从中感受到译者学习编程的优势。

敬望各位读者在使用本书时能与作者和其他读者一同精进前行！



第二章

入门

本章导言

如果你看完第一章后没有合上此书，那么祝贺你，你已经迈出了勇敢的第一步。从本章开始，我们将基于一个个案例由浅入深带你开启编程之路。由于本书涉及的所有编程案例都是与翻译相关的“在线工具”，“在线”意味着你可以在自己计算机上的浏览器中打开并使用，所以在本章中我们将学习如何制作简单的网页。

2.1 万维网的起源

2.1.1 何为“超文本”？

在我们的计算机[1]上有一个程序叫“记事本”（Notepad），通过这个程序创建的文件后缀名一般为“.txt”，打开文件后在其中输入的文字没有任何修饰，你无法对其加粗、加下画线。我们常称这类文件为“纯文本格式文件”，将其中的内容称为“纯文本”（Plain Text）。

作为一个学语言的人，你是否想过“text”这个英文单词是如何产生的？

很早很早以前，外国人的祖先认为我们人的思想就像线（Thread）一样，而缠线（Yarn）的纺织机（Spinner）则是话痨（Raconteur），只会傻傻地一层层缠线。

真正有思想、会讲故事的人像纺织工人（Weaver），能把不同的线编织在一起，织出有美丽纹理（Texture）的布来。会写字的人，把真正的思想记录下来，变成了内涵丰富的文字，取名为“Textus”，意为“布”（Cloth），后来演变为“Text”[2]。

在计算机出现之前，写在纸上的字叫作“Text”，计算机出现后，显示在屏幕上的字也叫作“Text”。但是，计算机屏幕不光可以显示字，还可以显示其他内容，比如：表格、图片等。这些内容叫作“媒体”（Media），也可以称为“资源”（Resource）。

当我们在计算机键盘上打字时，比如键入：“请参见《蒙娜丽莎》这幅作品。” 并且希望读者用鼠标点击“蒙娜丽莎”这四个字就能看到图片，或者跳转到包含这幅图片的网址，则需要在这四个字上添加一个“链接”（Link），把文字与对应的资源连接在一起。

那么当“文字”和“链接”合在一起，叫什么呢？美国信息技术先驱者、哲学家、

1 此处指 Windows 操作系统的计算机。

2 http://www.etymonline.com/index.php?term=Text

社会学家泰德·尼尔森（Ted Nelson）在他 26 岁的时候自己造了一个词：“HyperText”，译为“超文本”，用来指称有链接的文本；而加在超文本上的链接叫作“HyperLink”，译为“超链接”。

我们不妨在自己的计算机上试一下。

→第一步：

在计算机桌面上创建一个空白的文件夹，命名为“Test”。

→第二步：

在文件夹中创建一个名为“index.txt”的纯文本文件，并下载一张名为“MonaLisa.jpg”的图片，与其放在一起，如图 2-1 所示。

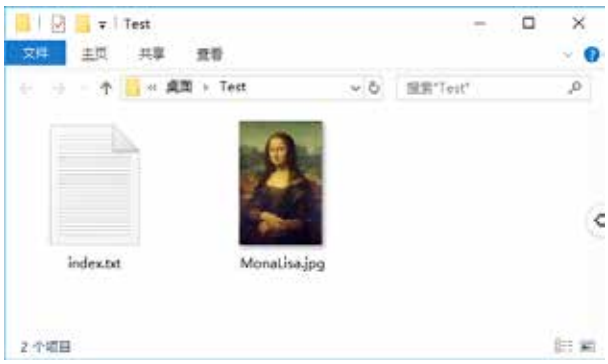


图 2-1　准备纯文本文件和图片文件

注：如果你在计算机上创建的记事本文件看不到后缀名，如图 2-2 所示。

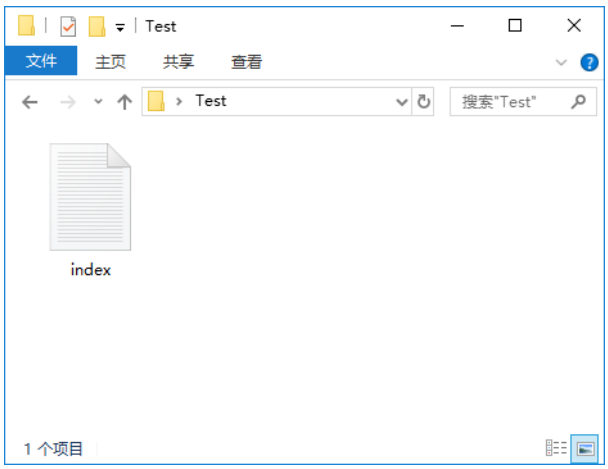


图 2-2　部分操作系统中文件后缀名会隐藏

点击顶部菜单栏中“查看”，并勾选“文件扩展名”，这样即可看到文件后缀名，如图 2-3 所示。

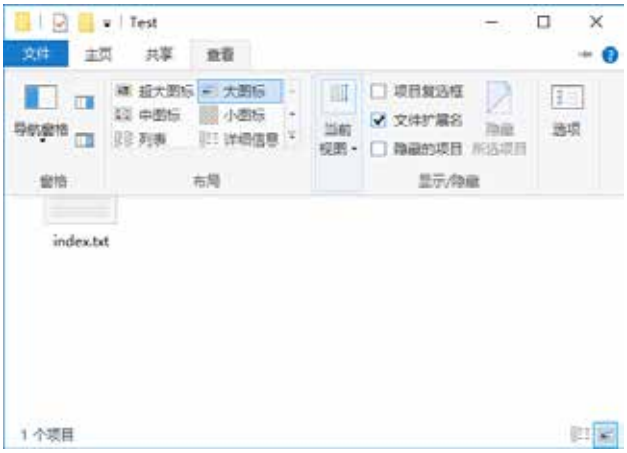


图 2-3　显示“index.txt”文件后缀名

→ **第三步：**

双击打开“index.txt”，并在其中键入以下内容：

```
1. <a href="MonaLisa.jpg"> 蒙娜丽莎 </a>
```

如图 2-4 所示。

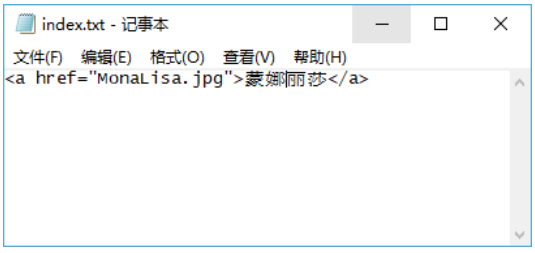


图 2-4　在“index.txt”文件中键入 HTML 代码

注：为了更好地显示本书所采用的代码，本书所有代码均会在每一行最前方显示行号，在自主练习时可以不必键入行号，而是直接键入行号后的代码。

→ **第四步：**

将“index.txt”的后缀名从“.txt”修改为“.html”，如图 2-5 所示。

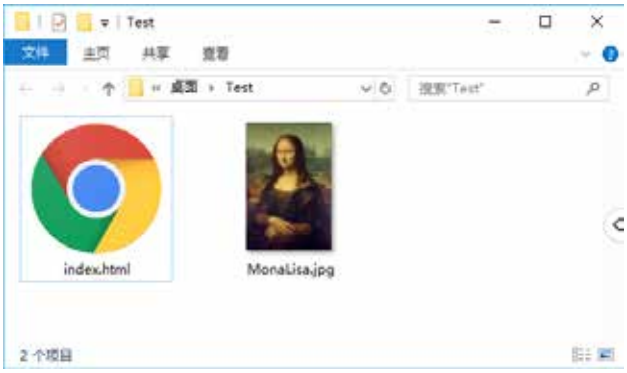


图 2-5 将“index.txt”文件的后缀名改为“.html”

→ **第五步：**

双击打开“index.html”文件，在浏览器中查看该文件，如图 2-6 所示。

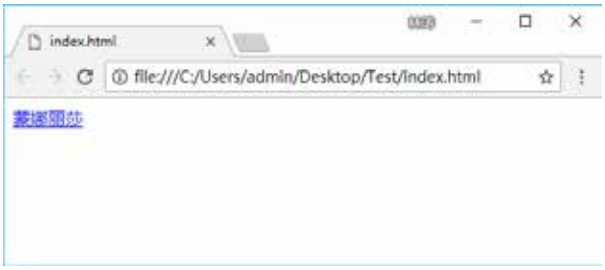


图 2-6 在浏览器中查看“index.html”文件

此时，会看到“蒙娜丽莎”四个字为蓝色且有下画线。点击后会呈现一幅“蒙娜丽莎”的图片，如图 2-7 所示。

图 2-7 在浏览器中点击超链接查看图片

由此，大家可以直观理解何为“超文本”（HyperText）。

2.1.2 万维网（World Wide Web）的背景

在“HyperText”诞生21年后，即1984年，29岁的蒂姆·伯纳斯·李（Tim Berners-Lee）来到位于日内瓦的欧洲原子核研究会粒子实验室工作，他的工作是开发一款软件，把分布在欧洲各个国家实验室的重要资料汇聚在一起，共同分享。

5年后，他使用超文本技术开发的软件把各个单位的资料连接在了一起，通过点击文本上的链接，他的同事可以访问欧洲原子核研究会粒子实验室位于不同国家实验室计算机上的资料。

蒂姆·伯纳斯·李当时购买了一台史蒂夫·乔布斯（Steve Jobs）所在公司研发的NeXT计算机，1985年乔布斯被迫离开了他自己创办的苹果公司（Apple），同年创立了NeXT公司，1989年3月30日乔布斯于旧金山公开展示了NeXT计算机。

乔布斯也是1955年出生的，与他同龄的蒂姆·伯纳斯·李在1989年的夏天，成功开发了世界上第一台网站服务器（Web Server）。他在这台服务器上存储了他所在实验室每位研究人员的电话号码，允许用户登录到服务器中查询电话号码。

这个功能虽然很简单，但却真正实现了他最初的想法。他把“资源”放到一台计算机上，也就是服务器（Server）上，允许用户（User）在自己的计算机上，也就是客户端（Client）访问服务器中的资源。

蒂姆·伯纳斯·李把人们所访问的网络命名为World Wide Web，简称“WWW”，译为“万维网”。

又过了两年，他把“万维网”接入到了“互联网”，也就是在1960年就诞生的“Internet”。

要想访问他的服务器上的资源，需要安装一个软件，名为“网页浏览器”（Web Browser），在浏览器中需要输入自己想找的资源的名字，而且这个名字必须是全球唯一的，这个“名字”叫作“Universal Resource Locator”，后来改名叫作“Uniform Resource Locator”，中文一般译为“统一资源定位符”，这就跟你出生后有身份证、入学后有学号一样，国家或学校统一给你分配一个只有你才有的识别序号，网站也需要一个全球统一分配的识别方式。它的简称是“URL”，小名叫“Web Address”或“Website Address”，中文称之为“网址”或“网站地址”。

现在大家都知道，如果我们上网查资料时，“数据库”（Database）连接不上，原因是：网站（Website）打不开了，服务器（Server）连不上了，浏览器（Browser）不支持，没有权限（Authorization）或用户名密码，等等。

当你去商店买计算机时，你的身份是顾客（Customer）或客户（Client），而卖给你计算机的人是老板或提供服务（Service）的服务人员（Server）。当你买到计算机后准备上网（Web）时，需要申请网络服务，获得上网账号，然后才能连接到互联网（Internet）上。

现在你看到这些名词后，也许能更理解它们的功能。

2.1.3 “在线”的意义

当你拥有了一台计算机，把你想分享的资源放到这台计算机上，给它取一个全球唯一的名字，然后将这台计算机连接到互联网，这时你的“资源”就真正“在线”了。而当你可以为他人提供在线服务的时候，你所提供的信息会变得更有价值。

2.2 HTML 入门

2.2.1 如何给文字“穿衣服”——标记语言

在上一节中，我们通过一行代码实现了这样一个功能：在浏览器中点击“蒙娜丽莎”四个字后看到蒙娜丽莎的图片。之所以能实现这个功能是因为我们将“蒙娜丽莎”从纯文本变成了超文本。

下面，我们再做一个演示。

→第一步：

打开桌面上的“Test”文件夹，将“index.html”的后缀名改回“.txt”。

→第二步：

在代码中“蒙娜丽莎”四个字的两侧分别加上以下标记： 和 ，如代码 2-1 所示。

```
1. <a href="MonaLisa.jpg"><b> 蒙娜丽莎 </b></a>
```

代码 2-1 在 <a> 元素中添加 和

修改后的“index.txt”如图 2-8 所示。

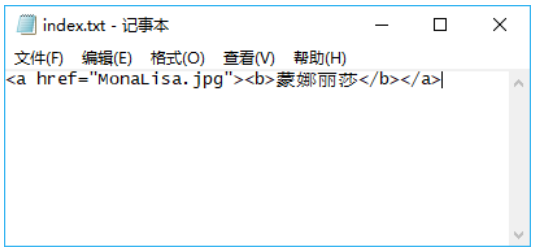


图 2-8 在“index.txt 记事本”中添加 和

→第三步：

将“index.txt”的后缀名“.txt”修改为“.html”，双击打开“index.html”文件，在浏览器中查看该文件，如图 2-9 所示。

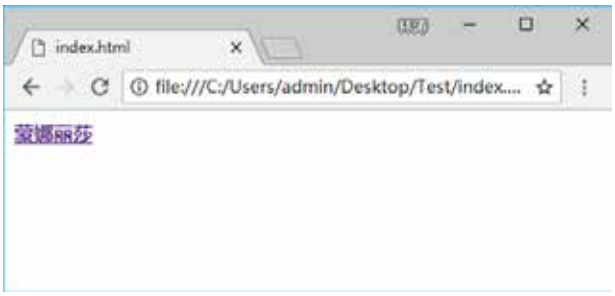


图 2-9　在浏览器中查看修改后的“index.html”文件

我们再将之前的截图和这张截图放在一起比较，如图 2-10 所示。

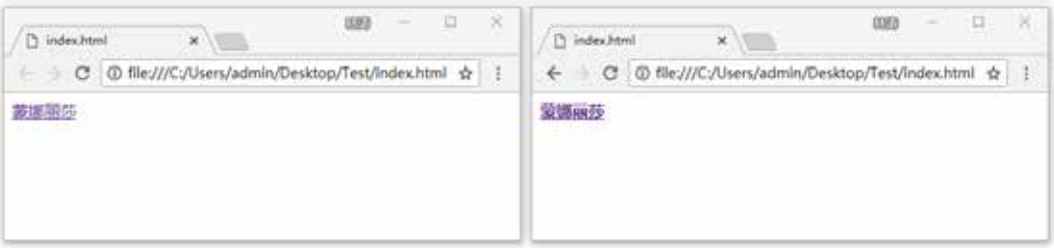


图 2-10　对比添加 和 前后的“index.html”文件显示效果

从上图可以看出，在代码中添加了 和 两个标记后，“蒙娜丽莎”四个字变粗了。如果我们将 和 替换成 和 ，重复上述步骤，如代码 2-2 所示。

```
1.  <a href="MonaLisa.jpg"><i>蒙娜丽莎</i></a>
```

代码 2-2　在 <a> 元素中将 和 替换成 和

在浏览器中显示的效果如图 2-11 所示。

图 2-11　对比添加 和 前后的“index.html”文件显示效果