

抓好索引 作振兴中华的技术先锋——上海交通大学 计算机科学与工程系教授王永成在中国索引学会成 立十周年庆典上的报告

[作者] 王永成

[单位] 上海交通大学计算机科学与工程系

[摘要] 抓好索引 作振兴中华的技术先锋——上海交通大学计算机科学与工程系教授王永成在中国索引学会成立十周年庆典上的报告。

[关键词] 索引，索引学会

1、振兴中华是历史赋予我们的光荣

如果说，19 世纪特别在鸦片战争之后是中国人血流流泪的世纪，20 世纪是中国人浴血奋斗的世纪，并建立了独立自主的新中国，毛泽东在天安门上宣布：“中国人民从此站起来了！”但是，光有政治上的独立，还不能使中国人在海外，在国际事务中不受人欺侮，不受人歧视。唯有不仅在政治上，而且在科学技术上发展起来，在经济上富裕起来，国防上强大起来，中国人才能真正地扬眉吐气。那么，21 世纪能否成为中国人扬眉吐气的世纪？这有待我炎黄子孙团结奋斗！

振兴中华，要抓什么？除继续狠抓政治改革之外，在技术上我认为就应当狠抓信息库的建设和利用。

当今时代乃是信息时代，信息时代的特征至少有三：

信息与知识的价值飙升，知识经济社会正在到来。

所谓信息，可用枚举式定义如下：

所谓知识：世界经合组织把人类的知识分为四类：事实知识、原理知识、技巧知识、用人知识。我们则更倾向分为三类：战略知识、战役知识、战术知识。

信息与知识价值的飙升，比尔·盖茨、王选、杨致远的成功都是突出的例证。

所谓知识经济：即以只是作为经济活动中起决定性因素的经济，它至少已表现出下述特性：脑力资源比体力资源更受人青睐；信息业比制造业有更大的效益；信息网比传统网发展更快；知识资本比金融资本更能增值；无形资产比有形资产更受重视。

信息爆炸般地产生，如海潮般地向人们涌来。

据 chinainfobank 总裁讲：每天网上的中文新闻信息就有 3 万条。世界上估计已有网页近 400 亿页；世界著名网站 Google 已索引了 20 多亿网页，Openfind 已索引了 35 亿网页，它们已为天下人所注目；现在，若用新浪搜索“MBA”一查就是 15 万 7 千条。

信息通过 Internet 等迅速传遍全球。

亿万人口利用 internet 传送讯息，万里距离，瞬间即可完成。真使地球村变得越来越小。

众所周知：随着信息时代的来临，以电脑为核心的信息科学、信息技术与信息产业，正在成为 21 世纪带头的科学、带头的技术、带头的产业。

抓好信息科学、信息技术与信息产业的核心是建好信息库。

美国的强盛,首先在于它掌握有世界上最多最新最尖端的信息,拥有世界上最大的信息库群,拥有世界上最好的将信息转化成生产力的工具—软件。

振兴中华,首先就要加强中文信息库与软件的开发。建设信息库的关键就在于信息的搜集、信息的索引、信息的加工、管理与信息的检索利用等等。

建设信息库的瓶颈在于索引。

美国 YAHOO 的成功,主要靠的是手工索引。现在,国内外都在大抓数字图书馆的建设。而建数字图书馆的瓶颈在于:信息的数字化、标引、存贮与检索方法的研发。

索引的媒体不同,索引的内容就不同。对文本来说,主要抓主题、特征项的索引;对视频来说,主要抓关键帧、特征项的索引;对音频来说,主要抓声、调、频的索引。

没有经过索引的信息集合很难高效地利用。它们就等如一山未曾挖掘出的宝藏。

由于我们都是战斗在索引战线的新老战士,因此,历史已把振兴中华技术先锋的重任放到了我们的身上。我们应为此而奋力,生命不息,冲锋不止!

2、中国索引学会应当作好动员和组织振兴中华技术先锋的作用

要为振兴中华作出贡献,要作好先锋,目前的困难是:

战斗的人群被分隔在各自的单位,没有很好地组织起来,形成不了气候。

有些同志有很好的积极性,但因种种思想与实际的困难,很难有所作为。

针对这二点,我提出如下一些看法:

当中国还没有任何组织能帮我们紧密组织起来之时,我们已有一个全国性的中国索引学会,我们应把它建设成一个振兴中华技术先锋的战斗堡垒!因此,我们应当充分利用建会十周年之际不仅抓好新老交替,更应从战斗的需要,来加强学会的领导。一定要把所有同志振兴中华的积极性最大限度地调动起来,一定要把大家的能量尽可能地集中起来;一定要使大家在振兴中华的过程中最大限度地使每个人的才智与财富得到发展与增加!WTO 的加入是中华民族历史的转折点。此时此刻,我们索引学会一定要认真思考,抓好改革,狠抓创新,才能适应时代的需求。为什么我们不充分地发挥它的作用?

无所作为的思想要不得,各自为政的小农思想难有大发展。

为了说明这点,我先让大家看一看我们上海交大的一些科技成果:

中英文自动摘要(网络版本表演)

该系统有如下一些功能:(1)能快速自动摘要:能对绝大部分网上单主题信息在以秒计的时间内作出能基本反映主题、文字流畅的摘要来。(2)可处理多文种:既可处理中文(繁体或简体),也可处理英文。(3)可处理多格式信息:本系统不仅可处理网上信息,而且也可以 word 文档。(4)摘要长度可变:摘要长度默认值为原文的 30%,但只要拖动长度比例光标,即可在瞬间方便地调整。另外还可以具体指定字数,摘要长度误差不超过最后一句话。(5)可批处理:除可对单篇文章做摘要,也可以对批量的文献进行摘要。(6)可中英对译:对于中文或者英文进行即时翻译。(7)允许文字编辑;摘要结果允许人们方便地进行手工修饰与编辑,这可使摘要更接近人们的需要。(8)字体大小可变:摘要结果可以放大或缩小,这特别适合年长者的需要。(9)保存摘要:摘要保存功能将帮您迅速建成自己的数据库或精萃文库。(10)打印摘要:摘要结果可方便地打印。有人已准备用它编辑各式各样文摘(如网上新闻文摘、网上娱乐文摘、网上书评文摘、网络聊天文摘等等)。

声控自动摘要。其流程图如下:(演示)

不少人看了此成果都很感叹,都认为它非常智能,国家八六三组则评价它们为国际领先。其实,它的具体做法,说起来它并不复杂!它基本思想就是“仿人!”

就是由年青人把你们在座的专家长期工作的经验放入电脑。这就成功了!

我们还能用它作偏重。要作到概念的偏重,就要有人做概念的分类与评价。

要机器做主题标引,就要有人对人工标引做总结。

要机器做自动分类,就要把人分类的经验总结出来。所有的标引,索引,全要由人把专家的经验总结起来,当前的标引有特征、主题、分类号、目次(表演)、摘要、图等等。索引,就是检索的引导。图书索引、门牌号码也是索引,……,索引工作多不胜数。

因此,索引工作者大有作为!

因为,我们上海交大的成果的基本方法是“仿人!”,这就要靠专家的指导。不仅方法的提出要靠专家,而且方法的改进也要靠大家试用与指导。

我始终认为:钱不应重点放在选购硬件上,而应重点放在软件研发与信息库的建设上,应坚持“以人为本”。

有人耽心这里用不上多少人。但是,根据我们的实践体会;这里即使组织成千上万人也不嫌多!

关键在于:这里需要经费,尤其需要明白的领导支持!

为了抓好索引工作,我们已组织了一个民营公司。该公司目前仍以师生义务劳动为主力。师生义务劳动故价廉物美。但长期价廉物就难美。我们必须按市场机制运转。

我们还准备在产生较大利润时在内部集资。目的是将大家进一步地组织起来。打破无所作为的思想,坚持老中青三结合,抓住机遇,团结奋斗,我们认为:不管是老人,还是中年人,青年人,都一定会大有作为,一定会为中华民族的振兴作出杰出的贡献。

3、抓住机遇,团结奋斗,夺取胜利

3.1 因为我们已经加入了 WTO

WTO 有四大宗旨:提高国民生活水平、可持续发展、帮助发展中国家与后进国家一起前进、全球一体化。

WTO 还有九大原则:这又分为四大类

非歧视性类:(1)最优惠普适原则;(2)国民待遇原则;(3)互惠互利原则;

自由化类:(4)市场准入原则;(5)关税减让原则(2015 年关税基本减到零);(6)取消数量限制原则;(7)公平竞争(反倾销、反补贴);

透明性类:(8)禁止暗箱操作与无赖手段;

照顾后进类:(9)允许例外与差别待遇。

因此,人们关于 WTO 普遍有十大预测:

政府的作用至关重要。入世首先是政府入世,它能斩断第三只手(看不见的手、政府调控的手)。三农(农业、农村与农民)问题已成了首要问题。国企要过关斩将,否则不堪设想。消费者将全面获益,产品质量将普遍提高,价格有可能还有所下调。显然:
a. 高质量的人力资本看涨, b. 有关虚荣心的服务会涨。弱势群体要想法自救,政府难以周到地照顾到成亿人群。权力结构将有所分化。虽然权力将继续发挥作用,但将受到越来越多的限制。金钱中的人民币最好进行积极投资。知识将显示越来越大的价值,问题是你会不会用。发展趋势:知识的作用将大大地高于金钱的力量与权力。社会将会大幅度分化。高等教育责任重大,这将在很大的程序上决定着国家的前途。美、欧、中三大势

力将称雄世界(2020 年左右)。 抓好信息处理+英文+ “ 游戏规则 ” 的学习将使您成新贵。

3.2 我们的中国索引学会 ,新的十年正在开始 ,因此一切都可以重新考虑。

我们的中国索引学会 ,已成立了十年 ,十年间已取得了很大成绩(最近将创办一本《中国索引》杂志 ,更是不简单) ,但是我们总感到形势逼人。新的十年正在开始 ,我们可以进行大规模地改革 ,让索引学会 ,在未来更充满活力 ,充满活力 ,更能发挥大家的积极性。正如有些同志所讲 ,对我们中国索引学会 ,“ 要么振兴 ,要么消亡 !”

我们要改革运行机制、开辟经济来源、增加监督机制。更严密地组织起来 ,在抓好理论和技术的同时 ,还要发展自己的经济实体 ,以保证学会今后有新的更大发展。