

关于专题数据库价值判断的 几个问题

人民出版社 沈水荣

问题提出：

专题图书数据库越来越多地进入各类图书馆采购，究竟如何选择专题数据库？

传统出版社的数据库（举例）

知识产权出版社的知识产权文书数据库

人民医学出版社的医学类图书数据库

交通出版社的交通类图书数据库

械工业出版社的机械类图书数据库

建筑工业出版社的建筑类图书数据库

外研社的外语教学类图书数据库

社科文献出版社的皮书数据库

法律出版社的法律类图书数据库

人民法院出版社的审判案例数据库

大百科全书出版社的大百科全书数据库

其他数据库（举例）

同方知网期刊数据库

超星图书数据库

人民日报数据库

人民大学报刊复印资料数据库

人民出版知识资源数据库系列产品

党政图书馆

马列经典库

领袖著作库

中央文件库

毛泽东思想库

中国特色社会主义理论体系库

史党建资料库

国际共运资料库

重大专题库

法律法规库

马克思主义中国化90年电子书展览

革命时期图书库

新华月报库

史知识库

学知识库

经济知识库

文化百科库

选择的正确方法：

要把握专题数据库的特点，不能用传统纸质图书的采购标准来判断专题数据库的优劣。

从数据库生产方的角度，谈谈数据库中的秘密。

讨论三个问题：

- 仅以单本图书相加衡量数据库的价钱是否可行？
- 按短时间点击量评价数据库是否科学？
- 数据库一定要装在自家的服务器里才算是本单位资产吗？

答案：否定

讨论问题一：

- 仅以单本图书相加衡量数据库价钱是否可行？
-

误区：

有的用户把数据库仅仅看作是一个装载电子书的容器，采购一个数据库，看划算不划算，是看花了多少钱，买到了些什么样的书，有多少册数、字数、**G**数。这是一个误区。

一个有一定质量图书数据库，其中的图书只是原材料，它是采用一定数量图书创作而成的、能够实现更多使用价值的新作品，而不是简单存放这些图书的容器。。

因此，判断一个数据库的价值如何，主要得看这个新作品所能满足的需求，而装载图书的数量质量仅仅是一个基础性的重要方面。

假如有一批同样数量质量的图书，不同开发商做出来的数据库，起码还可以从四个维度上比较优劣：

- 图书系统化程度
 - 图书结构化程度
 - 图书数字化深度
 - 图书检索准确度
-

一是图书系统化程度

系统化就是在某一知识领域内把图书种类收入得比较齐全，形成比较系统的知识覆盖面，使用户查阅时感觉到会应有尽有。

一个数据库要想囊括天下所有图书室是做不到的。

举例：

当今中国有一家非常著名数字出版公司开发的图书数字库，有一位图书馆采购安装之后，该馆馆长告诉我们，人民出版社的马克思主义经典著作他那里全有。然后我对比了一下，发现他那里擅自收录的此类著作占我们“党政图书馆”的5%都不到。

有的数据库只求“多”不求“全”，虽然里面的图书数量很多，但不系统。

为了回避这一弱点，数据库往往自己随意弄出一个分类。假如点开网页用中图法表来对，发现有的种类大片大片是空白，而有的种类大量重复。

这样的数据库，其价值就会大打折扣。而且只有用户使用时才会体验到，图书馆馆长也不一定能知道。

正确的做法：

坚持“数量”与“系统”相统一。

以“中图法”和相关知识体系为基础，把某一学科的知识类型分得很细，然后每一个细目下面都收入若干具有代表性、质量较优的图书，使数据库尽可能覆盖该领域所有知识点。

“党政图书馆” 收入图书做到了“五个全覆盖”

覆盖我国出版的
所有马列经典著作

覆盖党和国家主要
领导人所有著作

覆盖公开发表的所
有中文文件文献

覆盖国家所有法律
法规

覆盖党的思想理论
领域所有知识点

可见：

一个数据库不可能囊括天下所有图书，
但一个数据库可以覆盖某一学科领域有代表
性的重点图书。

我们的理念：

“内容为王， 王在系统”。

二是图书结构化程度。

就是数据库中的图书的存放要体现相互关联性，要按照知识的内在联系、用户的使用方式进行有序存放，使之形成一定逻辑结构和应用模式。

而不能把图书当作萝卜、白菜那样往里装，根本不考虑图书与图书之间的知识联系。

结构化的重要性：

结构化程度越高，越能实现知识元素相互之间的关联，越有利于实现阅读方式的智能化。随着结构化程度的提高，图书资源的使用价值能够以几何级数增长。

举例：

人民出版社开发的“马克思主义中国化90年电子图书展”，将建党以来历年出版的15000多本图书，完全按马克思主义中国化过程中每一个理论观点初次提出、演变、发展的过程进行排列展现。

举例：中国化马克思主义探索的源头

关于资产阶级和农民问题的探讨：

毛泽东1925年12月《中国社会各阶级的分析》、1927年3月《湖南农民运动考察报告》（编入《毛泽东选集》第一卷），1926年1月《中国农民中各阶级的分析及其对革命的态度》

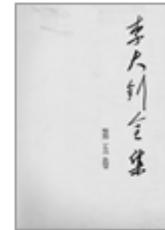
陈独秀1926年5月《第二次和第三次劳工大会之间的中国劳动运动》（编入《陈独秀文章选编》）

李大钊1925年12月至1926年2月《土地与农民》编入《李大钊全集》（第五卷）

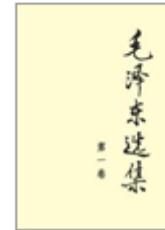
1926年5月，党领导召开的广东省第二次农民代表大会作出《农民运动在国民革命中之地位决议案》



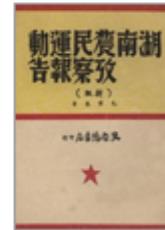
《陈独秀文章选编》（下），三联书店编，三联书店1984年6月出版。



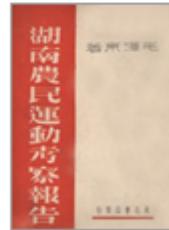
《李大钊全集》第五卷，中国李大钊研究会编注，人民出版社2006年3月出版。



《毛泽东选集》第一卷，人民出版社1991年6月出版。



《湖南农民运动考察报告》（新版），毛泽东著，冀鲁豫书店出版。



《湖南农民运动考察报告》，东北书店1947年12月出版。



《湖南农民运动考察报告》，毛泽东著，东北书店1947年12月出版。



《湖南农民运动考察报告》，毛泽东著，太岳新华书店1948年6月出版。



《湖南农民运动考察报告》，毛泽东著，东北书店1948年10月出版。



更多

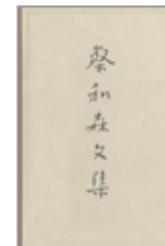
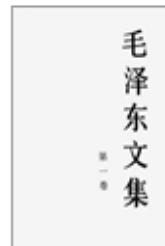
关于中国革命性质和前途问题的探讨：

蔡和森1923年5月《中国革命运动与国际之关系》（编入《蔡和森文集》）

毛泽东1925年冬《国民党右派分离的原因及其对于革命前途的影响》（编入《毛泽东文集》第一卷）

瞿秋白1926年3月《国民革命运中之阶级分化——国民党右派与国家主义派之分析》、同年5月《北京屠杀与国民革命运之前途》、1925年《中国国民革命运与戴季陶主义》
（编入《瞿秋白选集》）

党的四大《对于民族革命运之决议案》（编入《六大以前——党的历史材料》）



《瞿秋白选集》，人民出版社1985
年6月出版。

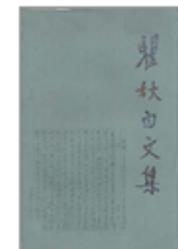
《毛泽东文集》第一卷，中共中央
文献研究室编，人民出版社1993年
12月出版。

《蔡和森文集》，人民出版社1980
年3月出版。

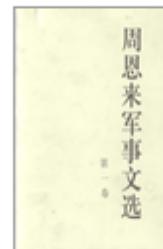
关于武装斗争问题的探讨：

周恩来1925年6月《军队的性质和组织》（编入《周恩来军事文选》第一卷）

瞿秋白1927年2月《中国革命之武装斗争问题》（编入《瞿秋白文集》政治理论编第四卷）



《瞿秋白文集》政治理论编第四
卷，人民出版社1993年7月出版。



《周恩来军事文选》第一卷，人民
出版社1997年11月出版。



举例：关于科学发展观的形成过程

第八部分 科学发展观的形成与全面建设小康社会

（一）科学发展观的形成和对马克思主义发展理论的丰富

进入新世纪、新阶段，以胡锦涛为总书记的党中央立足社会主义初级阶段基本国情，科学分析我国发展过程中的新情况新问题，总结我国发展实践，借鉴国外发展经验，围绕“实现什么样的发展、怎样发展”的重大问题，适时提出了科学发展观。科学发展观，第一要义是发展，核心是以人为本，基本要求是全面协调可持续，根本方法是统筹兼顾。胡锦涛所作的党的十七大报告系统阐述了科学发展观，指出：科学发展观，是对党的三代中央领导集体关于发展的重要思想的继承和发展，是马克思主义关于发展的世界观和方法论的集中体现，是同马克思列宁主义、毛泽东思想、邓小平理论和“三个代表”重要思想既一脉相承又与时俱进的科学理论，是我国经济社会发展的重要指导方针，是发展中国特色社会主义必须坚持和贯彻的重大战略思想。在新的发展阶段继续全面建设小康社会、发展中国特色社会主义，必须坚持以邓小平理论和“三个代表”重要思想为指导，深入贯彻落实科学发展观。

党的十七大之后，全党通过广泛深入开展学习实践科学发展观的活动，推动了经济社会发展各个方面的理论创新和科学发展。



相关文件文献

2003年10月，党的十六届三中全会通过《中共中央关于完善社会主义市场经济体制若干问题的决定》，提出“坚持以人为本，树立全面、协调、可持续的发展观”，对需要解决的重要体制问题提出了改革目标和任务。

2003年底，在中央召开的经济工作会议上，胡锦涛再次明确指出：牢固树立和认真落实以人为本，全面、协调、可持续的发展观。

2004年2月，党中央举办省部级主要领导干部“树立和落实科学发展观”专题研讨班，正式和公开使用了“科学发展观”的概念。

2004年3月10日，胡锦涛在中央人口资源环境工作座谈会上的讲话，集中概括了科学发展观的内涵。

2004年9月，党的十六届四中全会通过的《中共中央关于加强党的执政能力建设的决定》，提出把推进全面协调可持续发展作为党的执政能力的主要方面。

2005年10月党的十六届五中全会指出，科学发展观是指导发展问题的世界观和方法论的集中体现，全会提出我国区域发展总体战略，强调要继续推进西部大开发、振兴东北地区等老工业基地，形成合理的区域发展格局。

2007年10月15日胡锦涛在中国共产党第十七次全国代表大会上的报告，全面系统阐述科学发展观，指出科学发展观是我国经济社会发展的重要指导方针，是发展中国特色社会主义必须坚持和贯彻的重大战略思想。

2008年9月19日胡锦涛在全党深入学习实践科学发展观动员大会暨省部级主要领导干部专题研讨班上的讲话，提出努力把贯彻落实科学发展观提高到新水平。

2008年9月《中共中央关于在全党开展深入学习实践科学发展观活动的意见》。

2010年2月3日，胡锦涛在省部级主要领导干部深入贯彻落实科学发展观加快经济发展方式转变专题研讨班上发表重要讲话强调，紧紧抓住历史机遇承担起历史使命，毫不动摇地加快经济发展方式转变。

2010年4月胡锦涛在全党深入学习实践科学发展观活动总结大会上的讲话，系统总结学习实践活动的丰富经验。

2010年10月18日《中国共产党第十七届中央委员会第五次全体会议通过中共中央关于制定国民经济和社会发展第十二个五年规划的建议》。

2010年10月，党的十七届五中全会提出，做好“十二五”时期经济社会发展工作，必须以科学发展为主题，把加快转变经济发展方式作为主线。



这个数据库，利用历史上的出版物展现了一部活生生地马克思主义中国化历史；每一部作品都不仅仅是一个独立的个体，而是被放到完整的理论体系、历史过程中来展现。

三是图书数字化深度

图书数字化深度，就是对图书内容的编码程度。

人民出版社的重要图书数字化，从6个层面上编码：
每个文字编码、每个语段编码、每个章节编码、每页
作品编码、每件作品编码、每各作品体系编码。

误区：

有人以为，只要是电子书都是一样的。如有位朋友在10多年前买过一套光盘版的“党政图书馆”，他说这些书我都有了，不需要买人民出版社的“党政图书馆”了。

两者根本不是一码事，只是名字相同而已。别的不说，电子书做法也根本不一样。

电子书制作深度往往决定数据库的使用功能，
数字化做得越深，数据库就越好用。

如，每字数字化就可以按每个字检索，每个作
品体系数字化就可以按每个作品体系检索。

数字化加工深度不一样，花的成本也大不一样，同样一本图书，可以花几十元，也可以花几万、几十万元。

所以，采购数据库，一定要了解电子书的加工深度。

误区：

一些单位采购数据库，一般要问“有多少个**G**”的图书数据。这作为一个参考数据是可以的，但必须联系其他因素综合来看。

数据“G”数大的数据库不一定
图书资源就多。

如，一本扫描图做成的电子书，一页书
的数据量可以顶一本纯文本电子书。

□ 扫描图电子书

遇到对手的情况，进行投机，把各国人民革命斗争的成果說成是他們的核讹詐政策的胜利。一旦遇到对手，他們就牺牲革命人民的利益，在对方的核讹詐面前屈膝投降，从冒险主义演到投降主义。当他們进行投机的时候，他們把自己手中核武器的作用吹得天花乱鑿。当他們实行投降的时候，他們就把核战争說成是一触即发，而他們的投降是服从“时代威严的命令”，“拯救了世界和平”，如果誰要是批评他們，誰就是“想发动世界热核战争”。

我们认为，社会主义国家手中的核武器應該永远是防御的武器，任何社会主义国家在任何情况下都不容許首先使用核武器。社会主义国家和所有爱好和平的国家反对帝国主义侵略的最可靠的力量是自己的国防力量和人民力量，而不是其他国家的核武器。现代修正主义者威胁說要用火箭核武器去防止帝国主义对这个或那个国家的侵略，是吹牛的，在政治上是完全錯誤的。（暴風雨般的鼓掌和欢呼）如果他們竟然首先使用核武器挑起一場世界核战争，那就必将受到全世界人民最严厉的譴責。（雷动般的掌声和欢呼）

□ 纯文本电子书

然而思想落后于实际的事是常有的，这是因为人的认识受了许多社会条件的限制的缘故。我们反对革命队伍中的顽固派，他们的思想不能随变化了的客观情况而前进，在历史上表现为右倾机会主义。这些人看不出矛盾的斗争已将客观过程推向前进了，而他们的认识仍然停止在旧阶段。一切顽固党的思想都有这样的特征。他们的思想离开了社会的实践，他们不能站在社会车轮的前头充任向导的工作，他们只知跟在车子后面怨恨车子走得太快了，企图把它向后拉，开倒车。

我们也反对“左”翼空谈主义。他们的思想超过客观过程的一定发展阶段，有些把幻想看作真理，有些则把仅在将来有现实可能性的理想，勉强地放在现时来做，离开了当前大多数人的实践，离开了当前的现实性，在行动上表现为冒险主义。

唯心论和机械唯物论，机会主义和冒险主义，都是以主观和客观相分裂，以认识和实践相脱离为特征的。以科学的社会实践为特征的马克思列宁主义的认识论，不能不坚决反对这些错误思想。马克思主义者承认，在绝对的总的宇宙发展过程中，各个具体过程的发展都是相对的，因而在绝对真理的长河中，人们对于每一个一定发展阶段上的具体过程的认识

扫描图电子书的文字

遇到对手后
果真成了一

果真成了一

纯文本电子书的文字

告，宣布双
政府又采用

四是图书检索准确度

数据库价值远远不是承载图书的数量很多，它还应该具备许许多多的功能，如检索功能。

市场上不同数据库的检索功能大不一样。

- 有的只有一般的关键词检索，有的采用了语义检索技术；
 - 有的只能检索到书目信息，有的还能检索到章节，检索出每一句有完整意思的话。
-

“党政图书馆”的检索功能

语义查询 引文比对

概念关联 模糊找句

-
- ◆ 可查询知识点近亿个
 - ◆ 达到了语义检索的水平，查询平均准确率达到75—95%，使图书检索从传统纸质书的章节检索，发展到了以语段为单位的知识点检索。
 - ◆ 获得三项自主知识产权
 - ◆ 获得王选新闻科技一等奖
-

我们的检索功能之所以优于其它一般数据库，
是因为采用了机器标引与人工标引相结合的做法，
在社内外有一支几十人的专家标引队伍，加入了大
量人工标引，取得了事半功倍的效果。

讨论问题二：

□按短时间点击量评价数据库是不是科学？

误区：

不少单位采购数据库，先把数据库挂到自己网上试用，并把点击率高低作为是否采购的主要依据。这也不完全可靠。

我的观点：

短时间的点击率高低只能作为一个重要参考因素，但不能作为主要依据。

为什么？

因为短时间点击量与数据库优劣不一定成正比。

举例说明：

- 有的数据库检索功能较差，用户越是查得不准，越会带来点击次数的增多；
 - 有的数据库结构化水平低，想要的资料来回翻也找不着，这也会增加点击率；
 - 有的数据库收入图书过多过滥，干扰读者查询阅读，会造成用户停留时间延长；
 - 有的数据库目录做得很系统，但是实际上里面没有电子书，根本打不开，系统还提示你哪个技术问题没解决，由此造成停留时间延长。
-

举例：

在“党政图书馆”中查找党和国家主要领导人关于党的群众路线的论述

□ 群众路线论述148条“一网打尽”



范围词：群众路线 相似度： 60 检索 高级检索

定点词(事物性状)： 意义作用 状态情况 时地数序 概念内容 本质要义 依从由来
定点词(文章表达)： 评价判别 法律法规 意见要求 主张声明 预见展望 经典名言 事例典故 特定表述 篇目章节

二次检索

时间 全部 机构 全部 作者 全部 书目 请输入 检索 使用说明 | 定点词含义 | 留言

语义检索结果 共148条

1、范围词：群众路线

“群众是我们力量的源泉，群众路线和群众观点是我们的传家宝。党的组织、党员和党的干部，必须同群众打成一片，绝对不能同群众相对立。如果哪个党组织严重脱离群众而不能坚决改正，那就丧失了力量的源泉，就一定要失败，就会被人民抛弃。”（邓小平语）（引自：胡锦涛《在邓小平同志诞辰100周年纪念大会上的讲话》，第8页，北京：人民出版社，2004年版）反馈

2、范围词：群众路线

邓小平同志一生都在身体力行“凡是于人民有利的事情，无不尽力提倡与实行”的诺言。邓小平同志心中始终装着人民，时刻惦记着人民的安危冷暖，强调贫穷不是社会主义，社会主义要消灭贫穷；坚持走社会主义道路，根本目标是实现全体人民共同富裕。邓小平同志始终相信人民、依靠人民，总是把人民拥护不拥护、人民赞成不赞成、人民高兴不高兴、人民答

（引自：胡锦涛《在邓小平同志诞辰100周年纪念大会上的讲话》，第8页，北

检索类型： 全部 语义检索 全

全文检索结果 共628条

1

八、群众路线问题。群众路线有没有？有多少？

（引自：《毛泽东文集》第八卷，第80页，北京：人民出版社，1999年版）反馈

2

所谓群众路线，包括调查研究。（引自：《邓小平文选》第一卷，第290页，

京：人民出版社，1994年版）反馈

3

我们在肃反工作中的路线是群众肃反的路线。（引自：《毛泽东文集》第一卷，第218页，北京：人民出版社，1999年版）反馈

4

群众运动是实行群众路线的一种形式。（引自：《邓小平文选》第一卷，第10页，北京：人民出版社，1994年版）反馈

□ 一般的同类数据库中，从这些图书找出148条论述，恐怕得用几百小时，而且不见得找那么准确。

可见，简单认为“点击量大，数据库就好”的看法也是不可靠的。

讨论问题三：

- 数据库一定要装在自家的服务器里才算是本单位资产吗？
-

误区：

- 有的单位采购数据库，一般不愿意从互联网上买“帐号”，而愿意采购“镜像版”安装到自己内网上，认为前者只是买“服务”，不算自家的资产；后者才是买“产品”，才算是自家的资产。
- 还有的采购数据库时，要很在乎是“买断”还是“租用”。宁愿多花点钱一次性“买断”，而不愿意少花点钱“租用”。

□ 我的观点

这些考虑是多余的，是用采购传统纸质图书的观念来采购数据库。

□ 为什么？

因为数字化技术更新换代很快，产品使用周期都很短。

所以，买“产品”还是“服务”、“买断”还是“租用”几乎都是一样的。不像一个实体用具，买了可以永久拥有，放时间长了还可以成为古董。而数字化电子产品不一样。

□ 举例：

一台电脑，你采购的时候是一次性出钱“买断”的，但用了七八年以后技术被淘汰，当废品卖才几十元。你说，这与买“服务”、与“租用”有区别吗？不同的是多了一个废品。

□ 举例：

有人10年前买了一套光盘版的所谓“世界名著图书馆”，放到今天的电脑上连打都打不开，没有哪个机器能用它，只有你心理上存在，其实没有了。你说，这与买“服务”、与租用”有区别吗？也没有。

同样道理，今天采购的数据库，即便装到您自家的服务器里，五年、十年之后，技术肯定落后了，展现、传播、阅读、查询的方式肯定会发生很大变化。如果没有技术更新，就会变成一堆看不见、摸不着的废物。

所以也不存在买“服务”与买“产品”、“买断”与“租用”的区别。

有人会问，这样难道就无法建设自家的数字化图书馆了吗？

完全可以，那就需要服务商的服务永久陪伴你。

有的图书馆指望若干年内一次性把各类专题图书数据库买足了，作为一笔数字资产存放起来，然后摆脱服务商的服务，关起门来慢慢享用，那是做不到的。

而且你也没有这些图书的著作权，你那时即使有技术也无法将这些电子书做成一个新的数据库产品。

总之，建议各单位，采购数据库一定要着眼于当前好用，不要设想购买积累“永久的数字资产”。因为数字资产不可能是永久的。

以上意见供讨论， 谢谢大家！
