

化柏林, 张新民 (中国科学技术信息研究所, 北京 100038)

从情报学与相关学科的差异性看情报学的核心技术

摘要: 情报学与图书馆学、信息管理、通信科学、智能科学、计算机科学、计算语言学等学科有着密切的关系, 但它们之间又有着本质的不同。通过分析情报学与相关技术学科的差别, 得出情报学的核心技术是分析技术。因此, 应当在充分发展以检索为代表的传统情报技术的基础上, 大力加强分析技术, 特别是文本内容分析技术, 以推动相关理论、技术与实践的发展。

关键词: 情报学; 情报技术; 相关学科

Abstract: Information science is closely linked with library science, information management, communication science, intelligence science, computer science and computational linguistics. However, there are essential differences among them. The analysis of the differences among information science and the other related technical subjects concludes that the core technology of information science is analysis technology. Therefore, we should put stress on analysis technology, especially analysis technology of text content on the basis of vigorously developing the traditional information technology with retrieval technology as its representative so as to promote the development of correlated theory, technology and application.

Keywords: information science; information technology; related subject

情报学的目标就是通过新信息与原有知识的融合, 形成新的认知, 进而成为真正的情报, 用来服务于决策^[1]。情报学需要技术的支撑, 但不能技术至上。技术不是万能的, 但没有技术又是万万不能的, 即“为技术而不惟技术”。如果忽略了技术的开发和利用, 那情报学将会更加软弱; 如果太看重了技术, 就可能会忽略情报学的本质和属性^[2]。所以情报学要根植于技术, 但应高于技术。不能绕过技术, 而是要穿越技术^[2]。

当前情报学涉及的技术非常多, 但究竟哪些技术属于情报学的研究范畴, 哪些技术超出了情报学的研究范畴, 这个问题还没有得到充分重视和详细研究。笔者认为, 可以通过分析情报学与其相关技术学科的差别, 使情报学的核心技术更加清晰。

1 情报学与图书馆学

情报学有着很强的技术特征, 而图书馆学有着浓郁的人文特征。历史传统上, 图书馆学的分类比检索重要, 而在情报学里检索比分类重要。如果说检索是情报学的“看家本领”的话, 那分类就是图书馆学的“镇山之宝”。随着高校图书馆购买电子文献的经费和使用率都超过了传统的印刷型文献, 数字图书馆正在迅速崛起, 图书馆学里的检索渐渐地超越了分类。而随着搜索引擎对情报学的冲击, 检索在情报领域的相对地位也有所下降, 转而关注挖

掘与抽取等深层分析。关注文本数据是图书馆学与情报学的共同点, 对于图书馆而言, 元数据是一件很值得研究的事情, 它有助于迅速找到所需要的资源; 而对于情报学而言, 元数据的研究并不那么重要, 内容本身才是情报关注的焦点。因此可以说数字图书馆关注检索, 现代情报关注抽取与挖掘; 数字图书馆关注加工存储, 现代情报关注分析与理解; 数字图书馆关注元数据, 现代情报关注内容与本体; 数字图书馆关注词法与句法, 现代情报关注语义与语用; 数字图书馆关注用户访问量, 现代情报关注决策反馈效果。

图书馆学的研究是从大到小, 情报学的研究是从小到大。图书馆学的研究是逐步求精, 不断细化的过程。分类就是最好的例子, 从门类到大类到小类再到种次号等, 分类号越来越长, 细化程度越来越高。而情报学首先从问题出发, 然后再研究问题的前因后果, 最后寻求问题的解决方案, 属于牵一发而动全身的研究路径。

情报重视交流, 图书馆重视传播。交流的双向性更强, 传播的双向主要依赖反馈。在实际的图书馆服务中, 这种反馈往往反映在对读者的问卷调查上, 情报服务却少有问卷调查。情报领域的问卷调查往往不是用户; 而图书馆的问卷调查往往是用户。如果情报领域的问卷调查是情报用户本身, 那显然不能为情报用户提供情报, 真正的情报应该是关于他方的^[1]。技术是知识管理的手段, 但改变

不了图书情报机构的本质和属性,图书馆是传播知识的重要阵地,而情报机构更倾向于成为服务决策的智囊团。知识管理与知识服务在图书馆与情报学界都得到了高度认可,发展趋势非常迅猛,但是它们之间存在着质的不同。从文献服务到信息服务,最终走向知识服务是图书馆的发展路径;而情报活动中也涉及文献服务、信息服务、知识服务,但这3种服务在情报活动中并不是递进的关系,而是并列的关系,它们共同作用来支撑情报服务。无论哪种服务,自始至终必须贯穿情报的成分,必须围绕着情报展开,知识服务并不是情报活动的最终目标,情报活动的最终目标是在信息服务与知识服务的基础上,提供更快、更准、更专的情报,以服务于决策。信息化与知识化有力地推动着情报学的发展,但信息化与知识化都不能成为情报学的发展方向。

2 情报学与管理科学

信息管理的核心过程在于管理,核心要素是信息。而情报学的本质在于分析和研究,分析的目标是获得情报,研究的目的是服务决策。信息管理所涉及的过程主要包括组织、存储、分类、排序、标引、转化、共享等,而对信息的分析并不是信息管理所必需的。情报学的核心在于情报研究,“研究”这个动词实际上是一个没有意义的词,因为它的概念范畴太大,而“情报研究”中的“研究”似乎范畴更小一些,更突出分析,因此也有人把情报研究称为信息分析。信息分析是对信息进行分析,情报研究是研究出情报。前者强调过程,后者强调结果,应该是一个概念的两种表述,强调重点有所不同而已。情报研究的对象除了信息外,还可能涉及数据或知识,但仅从单纯的数据或知识中是很难分析出情报的。

信息管理与信息系统专业并不能完全代表情报学的本科专业。尽管它有一部分来源于过去的科技情报专业,但从全国范围来看,很多学校的信息管理与信息系统专业基本失去了情报学的属性和特色。这个专业培养出来的学生和传统意义的情报专业的学生在知识结构、素质和能力上都有很大的不同。信息管理与信息系统专业偏重管理与系统,而情报学应该偏重分析。前者适合做系统分析员、CD、技术总监等,适合于大多数IT企业;传统意义的情报学还是要回归到耳目、尖兵、参谋上来,即使是企业,也只适合企业的情报部门或者以全文数据为主营业务的那些IT企业。信息管理的分析更系统,强调整体;情报的分析更注重附加值,强调对决策的作用。还有计算机技术本科加MBA也特别适合项目经理,它的执行力很强,而情报人员的执行力相对较弱,但对信息更敏感,从信息中分析出情报的能力更强。因此,应尽量避免把情报学硕士

培养成高级程序员、网络管理员、数据库管理员等技术人员,这方面情报学不如计算机专业有优势。

造成信息管理与信息系统学科不是情报学本科专业的主要原因是,信息管理与信息系统全都在高校,而情报机构自身不设本科专业。高校为了拓宽学生知识面,增加就业机会,需要面向社会需求调整培养目标。

3 情报学与计算机科学

情报学关注结果与实现,注重效果;计算机科学技术关注过程与算法,注重效率。情报学关注新的应用,从需求的角度来学习新技术,类似于信息获取中的 Pull;计算机科学技术关注新的技术,先开发出新技术,再看能解决什么问题,类似于信息获取中的 Push。

目前的情报技术比较注重文献内容的分析处理,文本数据更是关注的焦点,对数值、图形、图像、音频、视频等数据的研究目前还不是主流。因此情报学里的技术应当更加针对全文数据,围绕全文数据的一系列活动,包括数据的来龙去脉以及中间分析,是情报技术关注的焦点。

很多情报学硕士点教学单位都为情报学研究生开设了“数据库系统原理”、“网络建设与管理”等课程,然而,50学时的数据库课程却很少提及全文数据库,结果上成了计算机专业的本科课程。从其授课大纲就可以发现,情报学硕士的专业课变成其他专业的本科课程的现象在情报教育界非常普遍。情报学归属于管理学大类,因此无论情报学有多么强的技术特色,都应有所限定,既然属于管理学,那么情报学的技术应该相对较软。也就是说,应更多地关注与利用计算机软件技术,特别是利用计算机来处理文本数据方面的系列技术,而不用去关注计算机本身特别是硬件技术及系统结构方面的技术。所以有些学校在招收情报学硕士生时加考“微机原理”、“微机技术及应用”等偏硬件和偏系统本身的课程似乎有些不妥^[2]。要让计算机技术服务于情报学,情报学里的技术要偏重于计算机技术的应用,而不是去研究计算机技术本身^[2],在这一点上,《现代图书情报技术》杂志的定位就是很好的范例。

4 情报学与通信科学

通信关注变量的变化、变量的影响以及对变量的控制^[3]。情报学侧重交流,而通信科学侧重传递;交流更注重人的参与,从这个角度讲,情报学更接近社会科学;而通信科学注重物理层面,从这个角度讲是纯粹的自然科学。

关于通信的学科名称,目前是把“信息与通信工程”作为一级学科名称。本文之所以把网络与通信放在一起,其原因是,网络是交流的一种形式,从ISO七层协议上来

讲, 顶层的应用是人与人之间的交流, 底层的传输是计算机与计算机之间的通信。因此网络更偏向于通信, 而不是计算机科学技术。网络的重点在协议解析、数据传输; 计算机科学技术重点还是计算。当然, 现在的网络计算也已兴起, 分布式存储更多地反映通信, 当然也有适量计算, 网络里计算特征非常明显的当属网格。

对于情报学而言, 对网络通信技术的关注应该集中在网络内容的分析处理而不是网络本身^[4], 在网络空间里, 计算机专家和图书情报专家的分工可以概括为技术和内容的分工。从技术出发的计算机专家主要考虑怎样提供一种完全智能、自动化、高效率的信息存取机制。图书情报专家更多地考虑知识内容本身的合理逻辑性, 寻求科学的知识建构、组织和控制的途径^[4]。

5 情报学与智能科学

史忠植教授指出, 人类的科学事业正面临着四大问题的挑战, 即物质的本质、宇宙的起源、生命的本质和智力的产生。智能科学作为四大挑战之一, 是生命科学技术的精华、信息科学技术的核心、现代科学技术的前沿和制高点, 涉及自然科学的深层奥秘, 触及哲学的基本命题^[5]。智能科学研究智能的基本理论和实现技术, 是由脑科学、认知科学、人工智能等学科构成的交叉学科^[6]。智能科学的任务是了解人的心智, 模拟思维与智能, 建立人机结合(人机一体化)系统的理论。智能工程的任务则是构建各种实用的智能系统, 研制各种智能系统的开发工具^[7]。

情报不等于信息在情报界已得到大多数人的认可, 而情报与智能的英文都对对应着同一个单词即 Intelligence, 说明它们有太多的相似之处与渊源关系, 那就是都围绕着智能。钟义信教授给出了智能的完整概念。在给定问题、问题的环境以及主体目的这 3 个要素的条件下, 智能就是有针对性地获取关于问题与环境的信息, 恰当地处理这些信息以达到认知, 在此基础上结合主体的目的信息合理地产生解决问题的策略信息, 利用得到的策略信息在给定的环境下成功地解决问题以达到主体目的的能力^[8]。从他对智能的论述中不难看出, 智能与情报是非常相似的。尽管情报学与智能科学如此相近, 但事实上, 它们差别非常大。

智能科学关注计算机如何才能模拟人的思维与判断, 而情报学并不关注模拟的过程, 只关注利用这些模拟技术能为之提供什么样的服务。智能科学的核心在于智能实现的过程与机理, 而情报学的核心在于智能的应用。因此, 情报学的发展应该更多地融合智能成分, 不论这种智能是人工的还是人类的。人工智能的核心是知识获取与搜索。人工智能的搜索与信息的搜索有着很大的不同。人工智能的搜索是带路径的, 搜索的过程是有条件的, 如“深蓝”

下棋时要考虑棋谱之间的关系。而信息搜索不是一次性的, 即一次性判断该文献是否命中, 即使有多个检索词, 那么再用下一个检索词来匹配该文献, 在判断命中的过程中不涉及目标文献之间的操作。

6 情报学与计算语言学

计算语言学通过建立形式化的数学模型来分析、处理自然语言, 并在计算机上用程序实现分析和处理, 从而达到用机器来模拟人的全部或者部分语言能力的目的^[9]。计算语言学是一个横跨语言学、数学和计算机科学的交叉学科。语言学和数学都是有着悠久历史的学科, 语言学历来被视为是典型的人文科学, 数学则被许多人看作是最重要的自然科学^[10]。在学校教学中语文和数学被认为是两门最基础的学科, 它们似乎成了学校教学的两个极点: 一个极点是作为文科代表者的语文; 另一个极点是作为理科代表者的数学^[10]。而计算语言学恰恰是把这两个极点结合起来。

计算语言学是“大文大理”的结合, “大文”是指深入研究人际交流媒介(语言)的各种复杂现象, 具有浓厚的人文色彩; “大理”是指深入研究计算机处理语言时所涉及的算法、过程, 具有明显的技术特征。计算语言学相关会议一般由两部分人组成: 那就是语言学专家与计算机专家, 语言学专家主要关注语言的单个现象, 找出个别现象的特征, 总结其规律, 如专门研究“‘的’字结构”、“‘于’式结构”等; 计算机专家主要针对语言的普遍规律设计解决问题的方案。与计算语言学相比, 情报学只能算是“小文小理”的结合, 因为它对技术的要求不像计算语言学那样高深, 而人文科学研究也无须像计算语言学那样精细, 即“文得不够活, 理得不够深”。

计算语言学的条件是语言具有计算性, 计算过程中突出语言现象的处理; 而情报学是从应用的角度来关注整个应用过程, 分析处理更注重结果。语言学是先关注个性, 再总结出共性, 情报分析是先关注共性, 逐步求精与完善时再关注个性。计算语言学会研究模型, 情报学只关注应用, 并不去研究计算语言学模型本身。

语言学家总结更多的语言学规律, 计算语言学家研究出更好的计算模型, 而情报学家研究出更好的应用实践, 解决具体的实际问题。计算语言学家关注语言学的各个层面, 从处理单位来讲包括词语分析、句子分析、语段分析及篇章分析; 从分析层面来讲包括语音分析、语形分析、语法分析、语义分析、语用分析。情报学家只研究语义和语用(文献[11]提到的第二个突破), 语形和语法不属于研究层面, 但会用到。钟义信教授提到的全信息理论, 包括语法信息、语义信息和语用信息^[12-13], 正好处于计算

语言学与情报学的中间地带。

7 情报学的核心技术

从 1989—2005 年的 17 种中文核心期刊论文（数据来源于重庆维普期刊数据库）中抽取关键词，这些分别以“数据”、“信息”、“知识”、“情报”、“文献”开头、后接动词的高频四字关键词，如表 1 所示。

表 1 图书情报领域核心四字关键词（各取前 10 位）

| | | | | | | | | | |
|------|-----|------|------|------|-----|------|-----|------|-----|
| 数据挖掘 | 146 | 信息服务 | 2103 | 知识管理 | 655 | 情报检索 | 810 | 文献检索 | 274 |
| 数据存储 | 26 | 信息检索 | 580 | 知识创新 | 168 | 情报工作 | 462 | 文献利用 | 93 |
| 数据处理 | 22 | 信息管理 | 359 | 知识服务 | 147 | 情报服务 | 382 | 文献采访 | 67 |
| 数据转换 | 22 | 信息咨询 | 170 | 知识共享 | 67 | 情报研究 | 232 | 文献分类 | 65 |
| 数据管理 | 14 | 信息工作 | 133 | 知识发现 | 62 | 情报教育 | 106 | 文献传递 | 58 |
| 数据分析 | 13 | 信息利用 | 123 | 知识传播 | 31 | 情报分析 | 105 | 文献工作 | 54 |
| 数据压缩 | 11 | 信息开发 | 119 | 知识获取 | 26 | 情报管理 | 36 | 文献计量 | 52 |
| 数据采集 | 11 | 信息交流 | 112 | 知识检索 | 24 | 情报搜集 | 33 | 文献服务 | 48 |
| 数据检索 | 10 | 信息共享 | 101 | 知识转化 | 23 | 情报交流 | 32 | 文献收集 | 44 |
| 数据集成 | 9 | 信息传播 | 80 | 知识挖掘 | 19 | 情报利用 | 26 | 文献老化 | 44 |

从表 1 中可以看出，在图书情报领域，排位最高词都是同列次位词的 n 倍，数据挖掘、信息服务、知识管理、情报检索、文献检索分别是数据存储、信息检索、知识创新、情报工作、文献利用的 5 倍、3.5 倍、4 倍、2 倍、3 倍。也就是说，在四字关键词中都有一个出现频率相对比较高的词，分别为数据挖掘、信息服务、知识管理、情报检索、文献检索，它们在一定程度上代表与反映着研究的相对集中。

情报技术研究历来就是围绕着（情报）信息检索技术与（情报）信息组织技术进行的，这两大研究领域自然也是当前提高情报学技术含量的主攻方向^[14]，检索技术在情报学领域的地位毫无疑问，但组织技术是不是情报学的核心有待论证。当然，近些年的实际情况就是如此，正是因为太看重组织技术而忽略了分析技术，才使得情报学与以数字图书馆为发展方向的图书馆学合流，导致了智能的缺失^[15]。

通过以上对比分析，笔者认为情报学的核心技术应该是分析技术，无论是人工分析还是自动分析，无论是定性分析还是定量分析，无论分析的对象是数据、信息、知识还是情报，都体现着智能的成分。分析过程的目标是实现信息知识与情报之间的转化，这些分析过程都是针对各种资源、运用各种手段、通过各种方法分析得出决策所需要的情报，这就是情报活动的本质。分析只是过程，可以服务于种种应用，如用于智能检索、机器翻译、知识抽取、技术监测、技术预见、非相关文献的知识发现等。因此，情报学的核心技术应当在充分发展以检索为代表的序化技术基础上，大力发展分析技术，特别是文本内容分析技

术，以推动相关理论、技术与实践的发展。

参考文献

- [1] 池建文. 论情报的两个基本问题 [J]. 情报学报, 2006, 26 (S1): 290-293
- [2] 化柏林, 张新民. 情报学学科范畴研究的方法论 [J]. 情报学报, 2007, 26 (5)
- [3] Shannon C E. A mathematical theory of communication [J]. The Bell System Technical Journal, 1948, 27 (7): 379-423
- [4] 吴慰慈, 罗志勇. 新技术革命对图书馆学情报学体系的影响 [J]. 河北大学学报: 哲学社会科学版, 2001, 26 (3): 102-111
- [5] 史忠植. 智能科学 [M]. 北京: 清华大学出版社, 2006: 9-11
- [6] 史忠植. 展望智能科学 [J]. 科学中国人, 2003 (8): 47-49
- [7] 戴汝为. 人一机结合的智能科学和智能工程 [J]. 中国工程科学, 2004, 6 (5): 24-28
- [8] 钟义信. “知识论”基础研究 [J]. 电子学报, 2001, 29 (1): 96-102
- [9] 俞士汶, 柏晓静. 计算语言学与外语教学 [J]. 外语电化教学, 2006 (5): 3-11
- [10] 冯志伟. 我与计算语言学的缘分 [J]. 现代语文 (语言研究版), 2007 (3): 127-128
- [11] 马费成. 情报学的进展与深化 [J]. 情报学报, 1996, 15 (5): 338-344
- [12] 钟义信. 全信息自然语言理解方法论 [G] / 徐波, 孙茂松, 靳光瑾. 中文信息处理若干重要问题. 北京: 科学出版社, 2003: 56-67
- [13] 钟义信. 自然语言理解的全信息方法论 [J]. 北京邮电大学学报, 2004, 27 (4): 1-12
- [14] 肖勇. 情报学研究趋势 [M] / 孙永发, 等编. 情报学进展. 北京: 国防工业出版社, 2003: 1-41
- [15] 包昌火, 李艳. 情报缺失的中国情报学 [J]. 情报学报, 2007, 26 (1): 29-34

作者简介: 化柏林, 男, 1977 年生, 硕士, 助理研究员。

研究方向: 自然语言处理。

张新民, 男, 1970 年生, 博士, 副研究员。

研究方向: 信息构建与知识管理。

收稿日期: 2007 - 07 - 06