

●王崇德 李 美

论超文本信息系统

ABSTRACT The similarities and differences between traditional text and the hypertext information system are analyzed; the basic principles of hypertext information system are expounded and the strong and weak points of hypertext system are discussed. The authors point out that the two complement and not repel each other. 12 refs.

SUBJECT TERMS Hypertext system—Studies Printed documents—Comperison

CLASS NUMBER G356

1982年美国学者 Jonassen 断言,“不出10年,我们所熟知的书本将要象活字印刷一样过时而被淘汰。”^[1]但14年过去了,印刷型的书本文献并没有被废弃。不过超文本信息系统——作为大有发展前景的新型文献集合却迅速发展起来。现代图书情报学家们普遍认为,传统的文本将读者和作者的构思和联想限定在单纯的线性模式之中,而超文本却不同,它能够“更加自然地反映人类思维进行情报处理的特征,从而解放并维持情报结构的本身。”^[2]本文拟就超文本的基本原理、特点以及优势与不足加以论述。

1 文本与超文本

什么是“超文本”?20多年前,文献中出现了“Hypertext”一词,当时的一般理解是,“超文本是无序的作品,它是充分利用了计算机的随机链接和动态显示信息的能力来存储自然语言的文本。”从今天的观点来看,这种表述和理解不够充分。首先,应

当将其对象“自然语言的文本”扩充至一切可用计算机存储的信息,如文本、图象和声音等。不仅如此,在最初阶段对“Hypertext”一词的用法也不十分严格,有时指窗口系统、电子邮件或电子会议等。时至今日,学术界对超文本的概念也未能达成共识并确定下来^[3]。不过根据超文本信息系统的功能和用途,图书情报界较为一致的看法是,超文本是以计算机支持的加工、存储、检索、咨询、评阅、编辑、交流专题信息的非线性高级文本系统。它与现在已流行的诸如超媒体、超索引、超文献、超网络、超立方体和超空间等一样,“都是对原有的单向线性工作、单值媒体或单值排列的一种扩充和开拓^[4]。曾经在 LITA 国家会议上因创造性建立在超文本基础上的图书馆管理软件包而荣获蓝绶带的 Bevilacqua 女士指出^[5],“超文本是一种组织原则,正如15世纪字母顺序的发明和柏拉图辩证思想的创立一样。”实际上,早在20世纪60年代,杜撰这一词汇的 Nelson 就将“Hyper”释义为

“扩展的、广义的、多维的”^[6]。Heim 在讲到文本的自然语言属性时指出^[7]:“文本一词最初源于拉丁语‘texere’,表示编织或交错编织的材料。”

由此可见,文本的最原始概念特征是线性结构。由此而顾名思义,“超文本”便是指超出文本之上的具有更高级结构的文本。

实际上,人们开发并运用超文本技术的历史已源远流长。恰恰是因为百科全书、目录卡片、引文索引以及文摘等这些“手工超文本”的相互作用和借鉴,才导致人类世界这一纷繁复杂的超文本知识网的创造。电子超文本信息系统极大地克服了传统文本体系的诸多缺陷,因而成为图书情报界的令人瞩目的、改革传统文献单一结构的焦点与突破口。为了讨论超文本,我们似应从它的前身——文本入手。

(1) 文本。文本是记录情报的实体。它与文献大体是同义词,主要包括自然书面语言,也含有图表信息。文本的信息组织是线性的、顺序的。从物理上看,它是以字符、行、页作为基本单位;从逻辑上来看,它是以字、句、段、节、章为单位。由于普通文本多为文字材料,其知识单元按线性顺序排列,因此也叫做线性文本。

用户利用线性文本接收情报时,每当他接收一个情报片段(如读完某个自然段、句子或术语等)后,通常有两种选择:要么往下读,要么倒回去参阅以前读过的内容。这种选择行为自然也是线性的。其实,在传统文本中也含有非线性的结构因素。例如,文后的参考文献指示的就是本文的情报内容与其他文献所含情报内容的参照关系。根据这种指示,用户在进一步的情报需求中就可能进行多向选择。这种表达情报间相互关系的“引用链”就是非线性的和多维网状的。但是,这种文后附录式引用链的一个明显缺陷是,它只能表明文献正文的整体与其它文献之间的联结关系,而无法表达这种文中各

情报片断与其它文献间最直接的联结关系,同时也无法表达这种联结关系的类型,如支持型、反对型、注释型、补充型和约束条件型等。用户在使用这种引用链时需要做大量的查阅工作,极不方便。传统文本中采用的另一种非线性组织因素是“随文脚注”,即页末注。据此,用户在线性地阅读某段内容后,还可进行非线性的选择性扩展阅读。这种引文链虽能直接表达文中某一情报片断内容与另一文献中的情报片断间的直接联结关系,但仍不能显示联结类型,而且这种联结还拘泥在有限的范围之内。

既然人类知识是一个在立体空间中由多种复杂关系链联结起来的、呈网状结构的整体,那末,人们在吸收情报时就自然需要全方位、长镜头、广视野、多层次地进行非线性的继承、借鉴和参考,以期将文本中的相关概念通过追踪阅读重组在一起,形成网状结构。但传统文本在这方面有严重局限,必须代之以更合理的文献信息系统来弥补这一缺陷。

再看一下传统文本与情报交流的障碍与弊端。通常,一项科研成果问世便会出现一篇记录该成果的文献。形成文献情报的一般过程是:人们在进行科学劳动时,首先通过感官了解研究对象的表象,同时记录初期的数据与资料;随着科研活动的加深,对研究对象逐渐形成全面的、新颖的认识,于是产生了新的科研成果;最后,要将这一成果用文字记录下来,也就是体现在文本式文献之中。整个科研行为和利用与加工知识的进程是,从网状到树状,再从树状到线状。而人们阅读文献的过程,则是一种相反的过程。但单纯阅读线性文本,无法形成个人的网状和立体的知识结构。这是人们依循文本吸收情报难以克服的障碍和弊端。

在以上的情报传递过程中,还有两点值得思考:第一,表征科研成果的情报本是立体网状结构,但记录于文献时却不得不硬性

将其压缩在线性的文本之中,这种“硬性”的加工必然要损失情报;第二,形成文献必然要消耗时间,从而使这一成果的情报传播产生了时差,即时间的滞后。这两点制约因素也是人们追求新型的、更为合理的文本的重要原因。

(2) 超文本。超文本信息系统便克服了上述的缺陷与弊端。从传统文本的页末注引用链进一步延伸,经过特定的技术处理和载体替换,就可以实现:①可将情报内容根据需要切割为任意大小而数量不限的情报片断或情报单元;②使人们表述和接受情报具有方便性。这两点正是超文本信息系统的基本要求。而用计算机信息系统便能实现这一要求。一般来说,超文本信息系统需要有专用数据库,以便把知识单元及其相互关系存储于其中。强调并突出这种“关系”是超文本信息系统数据库的最重要特征。结点和链路是超文本数据库的基础构件,结点表示知识单元(概念)、片断或其组合,链路表示结点之间的关系,如引用、反驳、赞同、修正和补充等。一组相关结点相当于一个数据库,一个结点相当于一个记录,但分布在各个层次上,超文本系统比一般数据库的内容要丰富得多,功能也更强大。用户使用超文本时,利用鼠标器或触觉控制,根据咨询的要求,显示屏幕上以单个窗口显示。超文本的特征就是在文本片断之间建立链结关系,具有组织非线性文本的能力。它与普通的窗口系统、纲要处理机、文本格式化系统有所不同^[8]。上述各种装置和系统均具有超文本的某些特点,但都不能算是超文本,因为它们的根本属性与超文本的有所不同。超文本是由计算机支持、协助思维活动和思想交流的一种媒体,它在机器保证下能进行“跟踪参考”。这种功能由链路提供,所以链路是超文本的根本属性。人的思维过程是异常复杂的,尤其是建立一个新思路,有时要在几个不同的层次上进行,时而展开,时而

抛弃,时而发展。对于这种缠绕不清的思路的记录与交流,是我们面临的一大难题。传统线性方式已不能满足要求,必须改用超文本信息系统,亦即借助计算机,以联结的方式提供非线性的概念空间。

2 超文本信息系统的基本原理

超文本的基本原理是实现文本的高速转移。所谓转移,就是根据中间结果来确定程序的应取时间。超文本具有最大结点度和链接力。结点度是指超文本数据库与外界客体的对应程度。链接力是指文本允许用户任意查阅建立或追踪参考项的能力。超文本的结点很多,链路则是各结点之间的“粘剂”,结点与链路的结合和统一,便构成了超文本信息系统的有机体。

(1) 结点。结点是超文本数据库的基础构件之一,是表达信息的单位。在结点中,表达信息的方法可以是文本、图形、图像、音频、视频和动画,甚至是一段计算机程序。大多数超文本以表述一个概念或一个思想的内容作为一个结点,它比传统的文档要小得多。超文本的设计者如何把一篇篇文献模块化,变成分门别类的结点,则是一种技巧。

在某些超文本信息系统中,把结点分成不同类型,这种结点叫做类型化结点,不同类型的结点表达不同的信息,而有的系统却只有一个结点,如KMS中的框架就无类型之分。系统中通过结点的大小、颜色或图示形式的不同来表示不同的结点,使用户对结点的类型一目了然。例如,注释型、决策型、目标型、约束条件型等结点分别用红、兰、黄、紫颜色来区分。

除类型化结点外,还有复合结点。超文本复合结点,是把相关的知识单元聚集在一起合成一个结点。

(2) 链路。链路是超文本信息系统数据库的另一个基础构件,其主要功能是链接与

追踪。在有些系统(如 KMS、Notecards 等)中,链路和结点是在一起的,即链路镶嵌在结点之中。但在有些系统(如 Intermedia 等)中,两者是分开的,即链路和结点单独存储。超文本信息系统的链路主要有 5 种功能:①把文献的参考项链接至文献本身;②把用户的评述和标引链接到被评述和标引的文献;③建立同一文献的两个或多个片断之间的关系;④链接表格、图形及其描述;⑤提供结构化信息,建立书刊目录款目与章节、段落之间的关系。

链路可有名称与类型,例如,支持型、反驳型和补充型等。超文本信息系统设有专门的装置以建立新的链路、删除(或改变)已有链路或列举链路等。一般来讲,超文本信息系统两结点的链接有两种方法:其一是参考法(又称非层次法),用于文本中“点”、“域”之间的链接;其二是组织法(又称层次法),以对具有层次关系的结点进行操作。

超文本信息系统在表述情报时,首先需要将情报切分为情报片断和情报单元——即结点,并将其存储起来。由于超文本中包含有普通的文本信息和图象信息,而这些信息的利用又是以“屏”为单位,所以超文本的 1 个结点往往是 1 屏或少于 1 屏的信息。1 个结点可以有多个“指出”链路以指向多个结点,也可以有多个“指入”链路以被多个结点所指向。这些众多的链路将超文本信息系统联结成为一个立体网状的情报空间。

用户在利用超文本信息系统接收情报时,即当阅读完一个结点后,可以同样方便地选择此后的各种阅读行为。例如,可以根据组织型链路进行本文献内的顺序或跳跃性阅读,也可以根据参考型的链路去查阅与本结点相关的其他文献中的结点,甚至可以把自己的注释、心得、体会、观点等输入计算机,成为一个或多个新结点,并利用相应的链路将这些结点与原文中的相应结点相

联,从而使情报利用与情报生产统一起来。显然,当超文本信息系统在局域性或全球性网络上运行时,用户刚存入的新结点又可以被网络上其他用户接收和利用,从而产生双向或多向的情报交流。这种即时性原始情报的介入,对于合作研究是十分方便和异常有甚至使两者合二为一。

3 超文本信息系统的优越性

超文本信息系统的出现,为追踪和建立参考项提供了新的途径。它的最大的优越性是,用户能够沿着系统提供的或自定义的链路来实现情报单元之间的快速移动和浏览,从而逼近以计算机为媒介进行类似于人的思考和情报传递的模式。同时,它还可以真实地反映客观情报源之间的相互作用的某些规律。超文本信息系统的优势如下^[8,9]:

(1) 易于追踪参考项。超文本设有以电子计算机支持的链路跟踪装置,可以对所有参考项进行向前或向后追踪。

(2) 易于建立新的参考项。用户可以根据自己的意愿,在不破坏原来体系和内容的前提下,方便地增添文本内容、加注评论和建立新的参考项。

(3) 实现情报的结构化。零散的情报可按树型和网型结构组织起来。

(4) 提供全景视图。超文本信息系统以目录表的形式提供全景视图,供用户浏览,也可以提供局部视图。两者并举,易于显示和弄清知识结构的内部关系。

(5) 任意剪裁文献。通过多种途径,把超文本片断按需要集锦在一起,使其具有多种服务功能。

(6) 赋予情报模块化。超文本的片断,如同积木一样可以组合,或者多处被参考,以避免重叠。

(7) 保证情报的连贯性。参考项被嵌入文本后,如果文本的内容发生变化,或与其

他文本合并时,被嵌入的参考项仍保留其链路,可以提取,保持其连贯性。

(8) 实现任务批处理。超文本可将几个不同途径的查询结果进行集中,同时在荧光屏上显示。

(9) 实现合作。多个作者可利用超文本共同合作撰写论文、项目设计,可以起到“电子会议”的同样作用。

就图书情报界而言,超文本信息系统最突出的优点是对情报联机检索的改革与创新。以往联机情报检索是采取检索一步看一步的方法,以便逐步逼近用户的情报提问。如果采用超文本信息系统则可以先利用它编制一个与情报提问有关的“关联图”,这样在检索时就可以解决检索哪些库、如何正确地编写检索提问式等问题。另外,超文本信息系统还有可能使检索提问式通用化,不必因使用不同的数据库而要求修改提问式。超文本信息系统可以知识单元为单位,通过链路将同一文献或不同文献的相关部分联接起来,检索时可深入到知识单元(文献片断);其他情报检索系统以文献为单位,检索结果必然是整篇文献。其他的情报检索系统采用准确匹配的检索方法,检索结果是未经排列的文献,无法区分它们的重要性;而在超文本检索系统中,文献是以结构化形式建立起来的,并非处于同一层次。用户在使用超文本信息系统时可以看到文献间的链路以及两篇文献间的路径或相隔的结点数,并由此确定文献的重要性。同时,还可以根据需要在没有链路的文献间加上链路。在一般的检索中,由于不熟悉检索语言和检索策略会给用户带来极大的不便,加之每一个数据库均会有不同的特征和使用不同的检索语言,检索难度大;超文本信息系统可以通过浏览链路,找到所需情报,避免了检索语言的困扰。当然,超文本信息系统还可以作为一个独特的用户界面,将不同数据库的检索语言一体化^[10,11]。

4 超文本信息系统的缺陷与不足

超文本信息系统存在的问题主要有两个,一是目前使用中暴露的问题;二是超文本自身存在的问题。前者包括查询到的信息显示、打印过程的时间延后、名称和链接属性的限制、浏览途径的缺乏和性能不佳等。下面主要讨论超文本本身存在的问题。

(1) 容易迷失方向。指在超文本环境中,人们迷失空间方向问题。迄今为止,人们一直习惯于通过书本及图书馆来阅读文献。由于目录、索引、标题、脚注及目录卡片不仅能揭示知识之间的相互联系,而且在某种程度上表达了人们试图通过印刷文献摹拟思维过程的愿望,因而得到了不断发展和完善。例如书后所附索引使读者不必通读全书即可很快获得所需段落。除此之外,人们还可借助书签、夹注、下划线及旁注(私人书籍)等来提高阅读质量。然而,对于超文本来说,尽管它具有组织知识网络能力,但却不能向人们提供他们熟悉的上述“路标”式的帮助。对于复杂问题,超文本必须帮助用户解决怎样才能确切地知道自己所处的位置以及如何到达某一位置的问题。当然,也存在着类似的空间障碍问题。因为超文本在向人们提供最大自由度、更多维数的活动空间以及更大的选择余地的同时,也带来了更大的迷失方向的危险(亦即产生定向障碍)。为此,通常采用两种解决办法,一是象 *Intermedia*、*Neptune* 和 *Notecarad* 等系统那样使用图形浏览器来帮助用户定位。其中,最重要和最常见的是全局浏览器,通过它可以在网络四处移动,看到所有结点。另一种办法是设置询问/查检装置。

(2) 认知负载过重。指在同一时刻,由于维持许多并行工作和记忆错综复杂的路径线索而造成的过度努力和精神负载。这是超文本信息系统目前面临的另一难题。对于

一个人来讲,面对一个头绪繁杂的网络,单靠大脑机械记忆,想保持结点的生成、命名以及它们间的联系等诸多因素的清晰图象是非常困难的。即使是计算机在处理超文本信息系统时也会出现“过载”现象。为了解决这一难题,通常的做法是设置快速显示交叉结点,在窗口提供瞬时项目说明,改进接口办法,开发情报过滤技术(即对链路和结点类型或结点扩展深度进行限制等)。

J. Caesar 曾因一直主张毁除亚历山大图书馆模式而受到人们的严厉批评^[12]。但是,当超文本以其逐渐成熟的姿态呈现在人们面前之后,“文献老化”、“情报爆炸”等一系列曾经令人难以排解的问题,不仅一下子荡然无存,而且给我们带来了一个新的亚历山大时代,使得人们不必因自身阅读与记忆的局限而放慢步伐或停滞不前。

还应指出的是,至少在今天,超文本和传统文本是相互补充、相得益彰的,而不是彼此去留和替代的关系。传统文本价格低廉、便于携带、无需其他阅读条件等优点,在未来相当长的时期内是不会因超文本具有很多优越性而被取代的。而且从人文主义角度来说,对包括超文本信息系统在内的高新信息技术还颇多微词,这也是传统文本不会被“废弃”和被“淘汰”的一种不可忽视的社会舆论与思潮。

参考文献

- 1 Andrew D. The Human Factors of Hypertext. *International Forum on Information and Documentation*, 1990, 15 (4): 32~35
 - 2 Beeman W., Anderson K., Gader G., Larkin J., McQuillan M., Shields M. Hypertext and Pluralism: From Lineal to Non - Lineal Thinking. In: *Proceedings of Hypertext 87*. University of North Carolina, 1987, 67~88
 - 3 Hester B., Soims S H. Hypermedia, Multimedia and Hypertext: Definitions and Overview. *The Electronic Library* 1993, 11 (4/5): 259~267
 - 4 王永成,杜锋. 信息处理发展的新动向——超文本系统. *现代图书情报技术*, 1992, (4): 31~35
 - 5 Ann F H. Hypertext: Behind the Hype. *American Libraries*, 1989, 20 (2): 158~1626
 - 6 Nelson T. H. A Conceptual Framework for Man-Machine Everything. In: 1973 National Computer Conference and Exposition. New York: AFIPS Conference Proceedings, 1973, M22~23
 - 7 Heim M. Electric Language: A Philosophical Study of Word Processing. New Haven: Yale University Press, 1987, 140~151
 - 8 Conklin J. A Hypertext: An Introduction and Survey. *IEEE Computer*, 1987, 20 (9): 17~41
 - 9 苗杰. 超文本——情报信息科学的又一新领域. *情报科学技术*, 1989, (6): 23~28
 - 10 Dunlop M D, Van Rusbergen C J. Hypermedia and Free Text Retrieval. *Information Processing & Management*, 1993, 29 (3): 287~298
 - 11 Croft W B, Howard R T. Retrieval Strategies for Hypertext. *Information Processing & Management*, 1993, 29 (3): 313~324
 - 12 Dan E. Hypertext in Myth and Reality: A Technology for the New Alexandrine Age. *Australian Library Journal*, 1991, 40 (4): 289~303
- 王崇德 1961 年大连理工大学化工系毕业。现为天津师大应用文科学院信息产业学系主任、教授。发表论文近 200 篇,出版专著 8 种、译著 3 种。主要成果有《文献计量学教程》和《情报科学原理》等。通讯地址:天津, 邮码 300070。
- 李 美 通讯地址:同上。
- (来稿时间:1996—03—11。编发者:丘 峰。)