

项目 4 本地机的信息检索

【项目介绍】

1. 项目描述

本地机的信息检索是将信息按一定的方式组织和存储起来,并根据信息用户的需要找出有关的信息过程。信息管理专业人员将信息存储在数据库中,信息用户利用数据库进行检索,可以快速找到硬盘上的 MP3、图片、Office 文档、电子邮件等信息。

2. 项目目标

- (1) 理解计算机信息检索的定义和类型;
- (2) 掌握计算机信息检索的基本原理;
- (3) 理解信息检索的检索技术。

【项目实施】

如今,硬盘的容量越来越大,自 2010 年 12 月,日立环球存储科技公司宣布,向全球 OEM 厂商和部分分销合作伙伴推出的硬盘系列容量已经达到 3TB、2TB 和 1.5TB。这样大容量的硬盘中可以存放大量的数据,而当用户不知道数据位于什么位置,或者希望查找某种类型的数据以及某个日期范围内建立的数据文件,这时则可利用 Windows XP 提供的搜索功能来快速确定数据所处的位置。

(1) 单击“开始”→“搜索”菜单项,打开搜索窗口,如图 4-1、4-2 所示。



图 4-1 “开始”→“搜索”



图 4-2 搜索界面

(2) 用户根据要查找的内容的类型来单击相应的选项。下面来解释一下各选项的含义。

- “图片、音乐或视频”。搜索图片、音乐或视频格式的文件。
- “文档(文字处理、电子数据表等)”。搜索文字处理、电子数据表等,如 *.txt、*.doc、*.html 和 *.xls 等。
- “所有文件或文件夹”。搜索所有的文件或文件夹。
- “打印机、计算机或用户”。搜索网络上的计算机或通讯簿中的人。
- “帮助和支持中心的信息”。在帮助和支持中心搜索相关的信息。
- “搜索 Internet”在 Internet 中搜索相关的信息。
- “改变首选项”对搜索助理进行一些相关的设置。

(3) 根据文件名称搜索本地机文件:用户在搜索的过程中可以使用通配符“*”和“?”来帮助搜索。

“*”表示代替文件名中任意的一个字符串,不论有几个字符也不管它是什么字符。比如使用了通配符“*”的“X*”,它可以匹配 X1、XA、X1A 等,只要开头是 X 的所有文件或文件夹。

“?”表示代替一个单个字符,不论是什么字符。比如使用了通配符“?”的“X?”,它可匹配 XA、X1,而不能匹配 X1A,X 后面只能有一个字符。

这里单击最常用的“所有文件或文件夹”选项,进入下一个窗口。在“全部或部分文件名”文本框中输入要搜索的文件名,比如要查找以“电”开头的文件,可以输入“电*”,单击“在这里寻找”右边的下三角按钮,在弹出的列表中选择要搜索的范围,单击 **搜索(R)** 按钮即可开始搜索,搜索结果显示在文件列表框中,如图 4-3 所示。

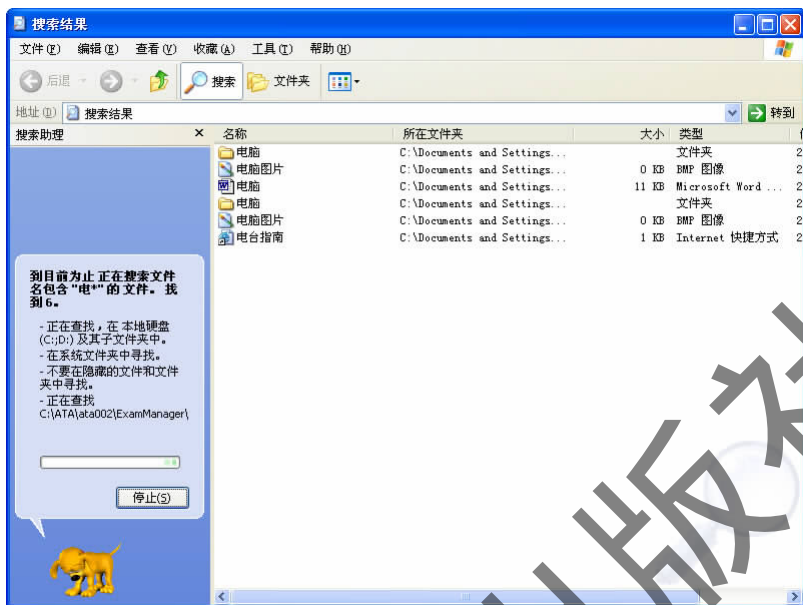



图 4-3 按文件名称搜索结果

(4) 如果用户想更加精确地搜索, 可对搜索条件进一步设定。

在“搜索助理”窗口中, 单击“什么时间修改的”选项右边的  按钮, 弹出详细的“什么时候修改”选项, 用户可以在其中设定详细的修改时间信息, 如图 4-4 所示。

其他详细搜索条件还包含文件大小、精确的文件类型、是否搜索系统文件、是否搜索隐藏文件和文件夹、是否搜索子文件夹、是否区分字母的大小以及是否搜索磁带备份等。

(5) 根据文档内容搜索: 进入如图 4-2 所示搜索窗口, 选择“文档(文字处理、电子数据表等)”命令, 在左侧选择或输入相应的内容, 如在“文档中的一个字或词组”中输入“Flash”, 设置如图 4-5 所示。

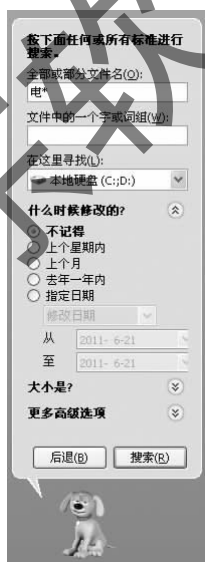


图 4-4 精确搜索界面



图 4-5 “文档(文字处理、电子数据表等)”命令

单击 **搜索(B)** 按钮即可开始搜索,搜索结果显示在文件列表框中,如图 4-6 所示。



图 4-6 按文档内容搜索结果

(6) 搜索局域网内计算机等共享的硬件或文件:如果你所需要的文档或硬件如打印机等在局域网内的某台计算机上,则可以通过 Windows XP 提供的搜索功能来快速找出相应的计算机,方法有两种。

方法一:使用对方计算机机名进行搜索。进入如图 4-1 所示搜索窗口,单击“计算机或人”命令,如图 4-7 所示。

打开窗口如图 4-8 所示,单击“网络上的一个计算机”命令。

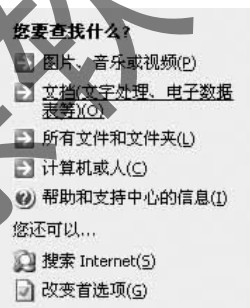


图 4-7 “计算机或人”命令



图 4-8 “网络上的一个计算机”命令

在“计算机名”中输入对方计算机名,如“xx”,单击 **搜索(B)** 按钮,如图 4-9 所示。



图 4-9 输入计算机名

搜索结果如图 4-10 所示。双击计算机名称,通常无需输入用户名和密码即可打开它所共享的软硬件资源。如需输入用户名和密码,则需要从该计算机管理员处获取。



图 4-10 局域网按计算机名称搜索结果

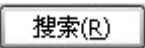
方法二:使用对方计算机 IP 地址进行搜索。打开如图 4-8 所示窗口,单击“网络上的一个计算机”命令;在“计算机名”中输入对方计算机 IP 地址,如“172.18.55.189”,单击  按钮,搜索结果如图 4-11 所示。




图 4-11 局域网按计算机 IP 搜索结果

(7)如在 Office 文档中需要获得一个定义、一个同义词、公司财务信息、一篇百科全书文章或来自网站的其他类型信息,Microsoft Office 2003 中新的信息检索服务可以提供这一切。以文档“社会主义和谐社会”中获取“和谐”的信息为例。

打开随书光盘根目录下的“社会主义和谐社会.doc”文档,单击“工具”→“信息检索”菜单项,如图 4-12 所示。



图 4-12 “工具”→“信息检索”菜单

“信息检索”任务窗格会在文档窗口右侧方便地打开,如图 4-13 所示。在搜索框中输入文字“和谐”,单击  按钮。

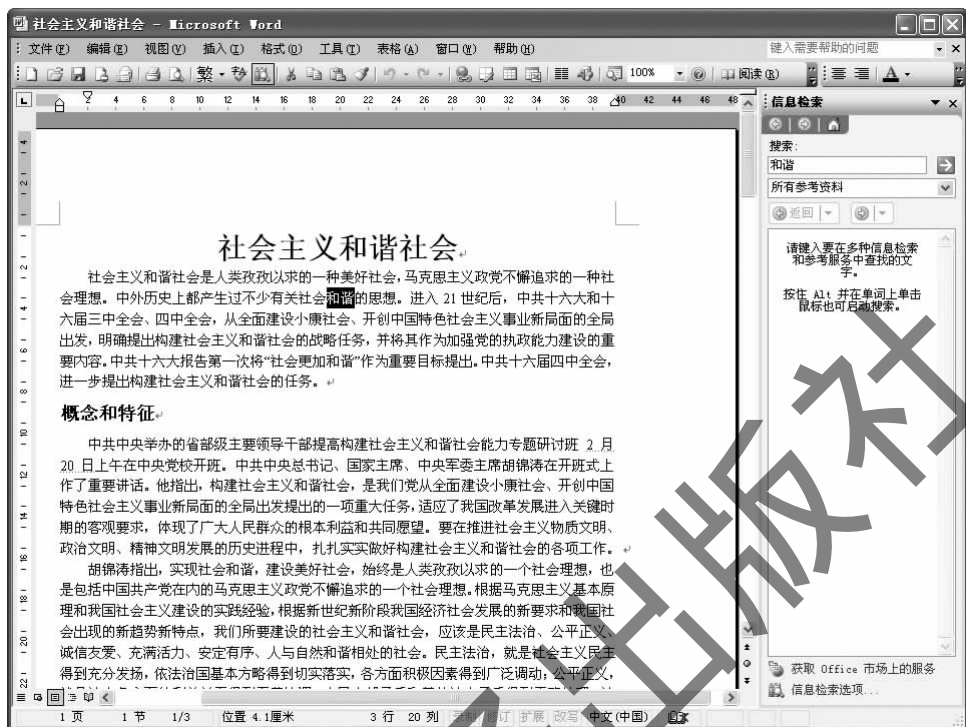


图 4-13 “信息检索”任务窗格

搜索结果,如图 4-14 所示。

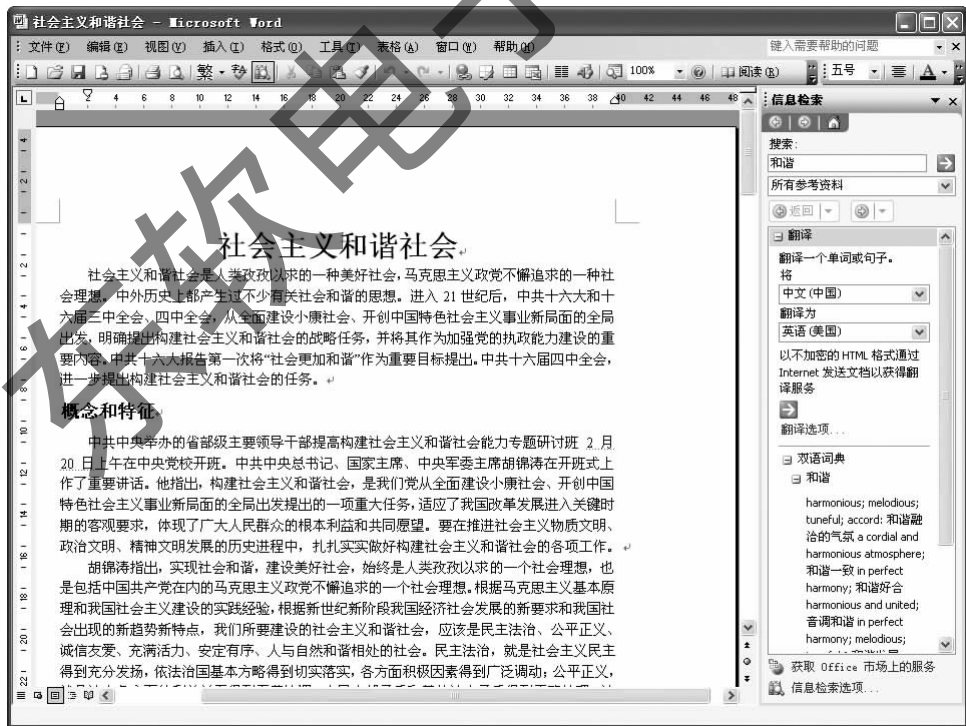


图 4-14 “信息检索”结果

【相关知识】

4.1 计算机信息检索概述

4.1.1 计算机信息检索的定义

计算机信息检索(Computer Information Retrieval, 简写 CIR)是指在计算机或计算机检索网络终端上,使用特定的检索指令、检索词和检索策略,从计算机检索系统的数据库中检索出需要的信息的一种计算机应用技术。它至少包含信息存储和信息检索两个方面。

信息的存储过程是将一次文献进行主题分析、标引和著录,按一定的结构存入计算机并能通过文献的部分特征标识进行检索;信息的检索则是面向用户,由用户对自己所要检索的课题进行分析,确定主题概念,抽取必要的关键词后,用一定的检索语言来表达课题内容,形成检索标识和检索策略,并输入计算机检索系统界面查找。这个过程一般是由计算机自动完成。

计算机信息检索系统(Computer Information Retrieval System, 简写 CIRS)是信息检索所用的硬件资源、系统软件和检索软件的总和。它能存储大量的信息,并对信息条目(有特定逻辑含义的基本信息单位)进行分类、编目或编制索引,它可以根据用户要求从已存储的信息集合中抽取特定的信息,并提供插入、修改和删除某些信息的能力,是实现计算机信息检索的平台。

4.1.2 计算机信息检索的类型

根据检索系统的工作方式,计算机信息检索系统可分为以下几种。

1. 脱机信息检索系统

由用户向计算机操作人员提问,操作人员对提问内容进行主题分析、标引、编写提问式,输入计算机,建立用户提问档,按提问档定期对新到的文献进行批量检索,并将结果及时通知用户。这是上个世纪 50 年代发展应用的检索系统。采用磁带作为存储介质,进行连续的顺序检索方式,适合大批量的定题信息检索,也称为成批检索或定题服务。因为是脱机的,因此,这种检索方式相对封闭,受到时间和空间上的极大限制,现在已基本被淘汰了。

2. 联机信息检索系统

用户利用终端设备,通过通信网络或通信线路与检索系统联机,进行“人机对话”,从检索中心的数据库查找所需要文献信息的过程。这种系统利用公共通信网络,突破了时空的限制,具有分时操作能力,可以使多台独立终端同时进行检索,实现用户与计算机的交互,计算机能对用户的提问及时处理、快速回答,用户还可以浏览相关信息,随时修改提问,直至得到满意的答案。

联机信息检索系统的优点如下。

- (1) 检索速度快,效率高;
- (2) 检索范围广泛、全面;

(3)检索途径多,方便、灵活;检索内容新,实时性强;

(4)检索辅助功能完善(人机对话、检索结果输出方式灵活,输出格式多样等)。

但同时也存在检索费用高,对检索系统及其文档(数据库)的收录、标引、特点等问题较难了解以及检索技术和技巧不易掌握的缺点。

3. 光盘信息检索系统

用光盘作为存储介质的系统,实现了 CD-ROM 资源共享。存储容量大、保存时间长、成本低。它通过塔式光盘驱动器、自动光盘换盘机和局域网,使多个终端用户实时共享上百张的光盘数据信息,也具备联机检索的特点。

4. 网络化信息检索系统

用户使用自己的计算机通过通信网络获取信息的过程。与联机检索系统所不同的是用户和服务器之间是平等关系,而不是主从关系。Internet 可以说是该系统的典型代表,它连接了一百多个国家、几万个信息网络、几百万台主机、几千万个终端用户,并能够跨越时空,进行实时信息检索、资源共享的国际性超级计算机网络。目前,90%的国际联机检索系统都已进入 Internet,世界上许多国家(包括中国)都从 Internet 上获取重要的科技和经济信息资源。

4.1.3 计算机信息检索系统(CIRS)的组成

从功能上说,CIRS 由服务器、数据库、通信网络和检索终端组成。

1. 服务器

通常包括硬件和软件,硬件决定系统的存储容量和检索速度,软件则充分发挥硬件功能,并进行信息的存储、处理、检索及系统管理。服务器硬件和软件的整体性能直接决定系统的检索服务能力,是检索系统的核心部分。

2. 数据库

按一定的方式合理组织并存储在计算机设备上的相互关联的数据集合。数据库是计算机信息检索的重要组成部分,是检索系统中存储信息的信息源。

3. 通信网络

实现联机检索和网络检索的必要条件。

4. 检索终端

面向用户的系统接口,主要有电传终端、数据终端和微机终端等。现在普遍是微机终端。

4.1.4 计算机检索与手工检索的比较

首先应该肯定的是,计算机检索的很多原理和规则都来源于手工检索,但由于查找的直接执行者不同,两者的检索过程又存在差异。

(1)手工检索中,人的思维起决定性作用,检索者可以不断修正检索方法,确认文献信息内容、概念或形式上的一致而无需进行严格的字面匹配,所检索到的结果一般都能符合需要;而计算机作为无思维能力的检索对象,在提问时要求检索者严格设计提问标识以尽量达到字面的完全匹配来进行检索。

(2)手工检索的费用相对较低,但耗时多、效率差、检索入口少。对于多数检索者来说,条件具备时,首先还是应考虑计算机检索系统来查找文献,既节省时间,而且容易达到效果。

(3) 不论手工检索还是计算机检索都需要检索者系统地学习信息检索知识和技术,采用合适的检索方法来达到快速而准确的检索效果。

4.2 计算机信息检索的基本原理和检索技术

4.2.1 计算机信息检索的基本原理

计算机发展史上有一个重要人物图灵,他在上世纪 50 年代曾经提出了一个假想的实验,认为计算机可以具有人类的思维能力,被称为“图灵实验”。他并且预言,在 20 世纪末,具有人工智能的计算机将会出现。

图灵实验——一种确定计算机是否会思考的实验:一个人向计算机发问,另一个不知情的人试图从回答中区分是人还是计算机。如果计算机没有被辨认出,便通过了图灵实验。但时至今日,仍未有任何一台计算机能通过“图灵实验”。

计算机既然如此“无能”,连三岁小孩的智能都不具有,那么为什么人们要在网上查找信息,却总要通过计算机帮忙呢? 计算机检索的奥妙在哪里呢?

原来,计算机实现检索的奥妙,就在于它能把你输入的检索词,与它后台数据库中存储的文件关键词进行比对,如果能够匹配,就认为这条信息是你需要的,立刻输出给你。这就是计算机信息检索的基本原理,即计算机将输入检索系统的用户提问标识(检索词)与已存储在系统中数据库内的文献特征标识(标引词)进行机械性匹配比较,凡符合给定的比较原则和逻辑运算条件的即为命中信息。

4.2.2 计算机信息检索技术

1. 布尔检索(Boolean Search)

布尔逻辑(Boolean Logical)检索是用布尔逻辑算符将检索词、短语或代码进行逻辑组配,指定文献的命中条件和组配次序,凡符合逻辑组配所规定条件的为命中文献,否则为非命中文献。它是机检系统中最常用的一种检索方法。逻辑算符主要有 And(与)、Or(或)、No(非)),分述如下。

(1) 逻辑与,运算符为 AND 或“*”。检索词 A 和检索词 B 用“与”组配,检索式为 A And B 或者 A * B,它表示检出同时含有 A、B 两个检索词的记录。逻辑与检索能增强检索的专指性,使检索范围缩小,此算符适于连接有限定关系或交叉关系的词。

(2) 逻辑或,运算符为 OR 或“+”。检索词 A 和检索词 B 用“或”组配,检索式为 A Or B 或者 A + B,它表示检出所有含有 A 词或者 B 词的记录。逻辑或检索扩大了检索范围,此算符适于连接有同义关系或相关关系的词。

(3) 逻辑非,运算符为 NOT 或“-”。检索词 A 和检索词 B 用非组配,检索式为 A Not B 或者 A - B,它表示检出含有 A 词,但不含 B 词的记录。逻辑非和逻辑与运算的作用类似,可以缩小检索范围,增强检索的准确性。此运算适于排除那些含有某个指定检索词的记录,但如果使用不当,将会排除有用文献,从而导致漏检。

例如:检索式“人工智能 And 机器人 And 算法”,要求检索出的文献中同时包含“人工智能”、“机器人”和“算法”三个词。

检索式“人工智能 Or 机器人 Or 算法”,要求检索出的文献中可以包含“人工智能”、“机器人”和“算法”三个词的一个即可,当然也可同时包含两至三个词。

检索式“人工智能一算法”,要求检索出的文献中可以包含“人工智能”但不能包含“算法”这个词。

2. 截词检索(Truncation Search)

截词(Truncation)检索是指用给定的词干做检索词,查找含有该词干的全部检索词的记录,也称词干检索或字符屏蔽检索。它可以起到扩大检索范围、提高查全率、减少检索词的输入量、节省检索时间、降低检索费用等作用。检索时,若遇到名词的单复数形式、词的不同拼写法、词的前缀或后缀变化,均可采用此方法。

截词的方式有多种,按截断部位可分为右截断、左截断、中间截断、复合截断等,按截断长度可分为有限截断、无限截断。

(1)右截断。截去某个词的后部,是词的前方一致比较,也称前方一致检索。例如:输入“geolog? (? 为截断符号)”,将会把含有 geological、geologic、geologist、geologize、geology 等词的记录检索出来。

若输入“PY=199?”,会把 90 年代的记录全部查出来。

(2)左截断。截去某个词的前部,是词的后方一致比较,也称后方一致检索。例如:输入“? magnetic”,能够检出含有 magnetic、electromagnetic、paramagnetic、thermo-magnetic 等词的记录。

(3)中间截断。截去某个词的中间部分,是词的两边一致比较,也称两边一致检索。例如:输入“organi? ation”,可以检出 organization、organisation。

输入“f??t,”可查出 foot、feet。

(4)复合截断。是指同时采用两种以上的截断方式。例如: ? chemi?,可以检出 chemical、chemist、chemistry、electrochemistry、electrochemical、physicochemical、thermochemistry 等。

(5)有限截断。是指允许截去有限个字符。例如: acid?? 表示截去一个字符,它可检出 acid、acids,但不能检出 acidic、acidity、acidity 等。

又如 compu????,可检出 compute、computer、computers、computing 等词,不能检出 computable、computation、computerize 等。

注意:词干后面连续的问号是截断符,表示允许截去字符的个数,最后一个问号是终止符,它与截断符之间要有一个空格,输入时一定要注意。

(6)无限截断。是指允许截去的字符数量不限,也称开放式截断。上面右截断、左截断所举的例子均属此类型。

由上述可见,任何一种截词检索,都隐含着布尔逻辑检索的“或”运算。采用截词检索时,既要灵活、又要谨慎,截词的部位要适当,如果截得太短(输入的字符不得少于 3 个),将增加检索噪声,影响查准率。另外,不同的机检系统使用的截词符不同,各数据库所支持的截断类型也不同,例如 DIALOG 系统和 STN 系统用“?”、ORBIT 系统用“:”、BRS 系统用“\$”、ESAIRS 系统用“+”等。

3. 限制检索(Limitation Search)

限制检索是通过限制检索范围,达到优化检索结果的方法。限制检索的方式有多种,例如进行字段检索、使用限制符、采用限制检索命令等。

(1)字段检索。是把检索词限定在某个(些)字段中,如果记录的相应字段中含有输入的检索词则为命中记录,否则检索不到。例如,查找微型机和个人计算机方面的文章。要求“微型机”一词出现在叙词字段、标题字段或文摘字段中,“个人计算机”一词出现在标题字段或文摘字段中,检索式可写为 microcomputer?? /de,ti,ab OR personal computer/ti,ab。

又如,查找 wang wei 写的文章,可以输入检索式 au=wang wei。

(2)使用限制符。用表示语种、文献类型、出版国家、出版年代等的字段标识符来限制检索范围。例如要查找 1999 年出版的英文或法文的微型机或个人计算机方面的期刊,检索式为 (microcomputer?? /de,ti,ab OR personal computer/ti,ab) AND PY=1999 AND(LA=EN OR FR)AND DT=Serial。

(3)使用范围符号。如 Less than, Greater than, From to 等。例如,查找 1989 年—1999 年的文献,可表示为 PY=1989:1999 或者 PY=1989 to PY=1999。

又如,查找 2000 年以来的计算机方面的文献,可表示为 computer?? b And Greater than1999。

查找在指定的文摘号范围内有关地震方面的文献,可表示为 earthquake? /1635000—800000。

(4)使用限制指令。限制指令可以分为:一般限制指令(Limit)和全限制指令(Limit all)。一般限制指令对事先生成的检索集合进行限制,全限制指令是在输入检索式之前向系统发出的,它把检索的全过程限制在某些指定的字段内。例如 Limit S5/328000—560000,表示把先先生成的第 5 个检索集合限定在指定的文摘号内。又如 Limit all/de,ti,表示将后续检索限定在叙词和题名字段。

上述几种限制检索方法既可独立使用,也可以混合使用。

4. 位置检索(Proximate Search)

位置(Proximate)检索是在检索词之间使用位置算符,也称邻近算符,来规定算符两边的检索词出现在记录中的位置,从而获得不仅包含有指定检索词而且这些词在记录中的位置也符合特定要求的记录。这种方法能够提高检索的准确性。当检索的概念要用词组表达,或者要求两个词在记录中位置相邻或相连时,可使用该方法。机检系统中常用的位置算符(按限制强度递增顺序排列)如下。

(1)(f)算符(Field)。要求被连接的检索词出现在同一字段中,字段类型和词序均不限。例如,happiness(f)sadness and crying。又如,pollution(f)control/ti,ab。

(2)(s)算符(Sub-field/Sentence)。要求被连接的检索词出现在同一句子(同一子段)中,词序不限。例如,machine(s)Plant

(3)(n)算符(near)。要求被连接的检索词必须紧密相连,词之间除允许有空格、标点、连字符外,不得夹单词或字母,词序不限;(Nn)表示两个检索词之间最多可以夹 N 个词(N 为自然数 1,2,3,⋯,X 且词序任意。例如,information(n)retrieval 可以检出 information retrieval 和 retrieval information,又如,econom???(2n)recovery 可以检出 economic recovery,recovery of

the economy and recovery from economic troubles

(4)(w)算符(With)。要求检索词必须按指定顺序紧密相连,词序不可变,词之间除允许有空格、标点、连字符外,不得夹单词或字母;(NW)表示连接的两个词之间最多可夹入N个词(N为自然数),词序不得颠倒。例如,input(w)output可检出input output;而wear(1w)materials可检出wear materials,wear of materials。

采用位置算符检索时,通常最严谨的算符放在最左面,例如:european(w)economic(w)community(f)patio,又如redwood(3n)deck?(s)(swimming(w)Pool?)

注意:不同的机检系统,位置检索的功能及算符不同,应参看机检系统的说明,上述为DIALOG系统的位置算符。

位置检索可以反映出两个检索词在文献中的邻近关系。常用的表示有:中间可插入几个字、两词可否颠倒位置、紧密相连、在同一句话中等。这种检索技术常用在全文检索中,可以弥补布尔检索的不足。

5. 加权检索 (Weight Search)

加权检索与布尔检索、截词检索、位置检索一样,也是信息检索的一个基本检索手段。不同的是,加权检索的侧重点不在于是否检索到某篇文献,而是对检索出的文献与需求的相关度作评判。因此,加权检索并不是所有信息系统都提供的检索技术。

加权检索中,检索者根据检索词在需求中的重要程度给定一个权值。在检索中,由系统先查找存在这些检索词的文献,并计算它们的权值总和。然后,检索者再给定一个阈值(Threshold)。只有当存在这些检索词的文献的权值之和大于或等于该阈值时,才算命中。加权技术有词加权技术和词频加权技术。前者是由用户在提问式中给定权值,需要人工干预;而权值的权值由词在文献中出现的频率决定,由系统自动赋值,不需要人工干预。

6. 多媒体检索 (Multimedia Search)

随着多媒体计算技术的迅猛发展,各种音频、图像、视频信息开始层出不穷,人们已不再满足于传统的文字检索,提出了对多媒体信息的检索需求,因此,基于内容的多媒体信息检索应运而生。

基于内容的检索是指根据媒体和媒体对象的内容及上下文联系,在大规模多媒体数据库中进行检索。它的目标是提供在没有人类参与的情况下能自动识别或理解声音、图像、视频重要特征的算法。

基于内容的音频检索包括以语音为中心,采用语音识别技术的语音检索;以音乐为中心,采用音符和旋律等音乐特征的音乐检索;以波形声音为对象的音频检索。基于内容的音乐检索系统主要针对音高、音长、音强等音乐特征的提取、识别和检索,包括音乐特征的规范化和提取、用户输入识别及特征提取、音乐特征的匹配检索及输出、相关反馈等。目前已成熟的系统有:MUS-CLEFISH、新西兰数字图书馆研究项目组所开发的MELDEX、卡内基-梅隆大学开发的QPD等。

基于内容的图像信息检索的主要工作集中在识别和描述图像的颜色、纹理、形状、空间关系上。对于视频数据,还有视频分割、关键帧提取、场景变换探测以及故事情节重构等问题。由此可见,这是一门涉及面很广的交叉学科,需要以图像处理、模式识别、计算机视觉、图像理解等领域知识作为基础,还需从认知科学、人工智能、数据库管理系统、人机交互、信息检索等领域引

人新的媒体数据表示和数据模型,从而设计出可靠、有效的检索算法、系统结构以及友好的人机界面。

视频结构的模型化或形式化是解决基于内容视频检索问题的关键技术,需要解决关键帧抽取与镜头分割、视频结构重构等技术。

基于内容的多媒体检索技术的日益成熟不仅将创造出巨大的社会价值,而且将改变人们的生活方式。因为它与传统数据库技术相结合,可以方便地实现海量多媒体数据的存储和管理;与传统 Web 搜索引擎技术相结合,可以用来检索 HTML 网页中丰富的多媒体信息。在可预见的将来,基于内容的多媒体检索技术将会在以下领域中得到广泛应用:多媒体数据库、知识产权保护、数字图书馆、网络多媒体搜索引擎、交互电视、艺术收藏和博物馆管理、遥感和地球资源管理、远程医疗、天气预报以及军事指挥系统等等。

目前,国内外已开发出的基于图像/视频检索的原型系统已有多种。IBM Almaden 研究中心的 QBIC 是基于内容检索系统的典型代表,允许使用例子图像、用户构建的草图和图画、选择的颜色和纹理模式、镜头和目标运动和其他图形信息等,对大型图像和视频数据库进行查询。此外,还有 MIT 媒体实验室开发的 Photobook 系统、新加坡国立大学开发的 CORE 系统、美国哥伦比亚大学图像和高级电视实验室开发的 VisualSEEK 系统等。

7. 超文本检索(Hypertext Search)

超文本是一种信息的组织方法。它把不定长的基本信息单元存放在节点上,这些基本信息单元可以是单个字、句子、章节、文献,甚至是图像、音乐或录像。节点以链路方式链接。链路可以分为层次链、交叉引用链、索引链等,构成网状层次结构。超文本的特点是以联想式的、非线性的、链路的网状层次关系,允许用户在阅读过程中从其认为有意义的地方入口,直接快速地检索到所需要的目标信息。

超文本检索时其内容排列是非线性的,按照知识(信息)单元及其关系建立起知识结构网络,操作时用鼠标去点击相关的知识单元,检索便可追踪下去,进入下面各层菜单。这种检索方式常用在多媒体电子出版物中。这类出版物不仅采用超文本,而且常采用超媒体(Hypermedia),提供文本和图形接口。Internet 上的 WWW 便是典型例子。

思考题

1. 简述布尔逻辑算符 AND、OR 和 NOT 在编制检索提问式中的作用。
2. 布尔检索中,使用布尔算符时要注意什么问题?
3. 常见的位置算符有哪几种? 请举例说明它们各自的含义。
4. 高级检索技术包括哪些检索技术,分别简述之。
5. 请为检索课题“航空或航天发动机的设计和制造”编制检索策略。

上机实践

在当前电脑上搜索以下符合条件的文件。

- (1)所有 jpg 格式文件。
- (2)文件名中包含有文字“图片”的文件。
- (3)文件名的第一个汉字是“中”的文件。
- (4)C 盘文件内容中包含有文字“系统”的文件。
- (5)近一个月来大于 100KB 的 doc 文件。

东软电子出版社