

黄丽霞

WWW 环境下信息检索策略实验调查研究

摘要 人们在网上传询某个问题时,采取的第一检索策略是非常重要的。通过实验调查研究,重点探讨了 WWW 环境下对不同类型的问题人们选择的第一检索策略。在实验中将待查询的问题分为 4 个方面:查找具体性问题、查找细节性问题、查找方向性问题、查找可能性问题。在取得实验数据基础上探讨了存在的共性问题。表 6。参考文献 12。

关键词 WWW 信息检索 搜索引擎 检索策略

分类号 G354.2

ABSTRACT The author thinks that the first retrieval strategy is very important when a person searches online information. Therefore, she summarizes first strategies for different issues in WWW search, classifies them into four groups, and analyzes common natures. 6 tabs. 12 refs.

KEY WORDS WWW. Information retrieval. Search engine. Search strategy.

CLASS NUMBER G354.2

1 引言

因特网的出现,从根本上改变了人们获取和交流信息的方式,引发了一场新的信息革命。因特网的发展和普及使人类的信息交流突破了地域和空间限制,拓宽了信息交流的范围,实现了瞬间交流。网络上丰富的信息资源使人们可以得到比以前任何时候都多的信息,但并不意味着人们总能得到自己想要的信息。因此,采取正确的检索策略是非常必要的。

虽然人们在查询某个问题时,通常会采取几种策略,但是,人们采取的第一检索策略(主要集中在直接输入地址、主题指南、搜索引擎和检索词)往往是非常重要的,它是解决一个问题的开端。问题的切入点找准了,就能比较顺利地完成任务。

本文的目的在于通过实验,研究从被查询的问题出发,考察被调查者对不同问题作出的第一反应,即他们采取的第一检索策略。为此,本实验将待查询的问题分为 4 个类型:查找具体性问题。指问题的结果明确、具体,信息来源确定、可预测。查找细节性问题。指问题的结果有一定范围;信息来源不确定,无法预测。查找方向性问题。指问题的结果范围广,具有开放性;信息来源确切,可预测。

查找可能性问题。指问题的结果范围极广,更具开放性;信息来源不确切,不能预测。

针对上述的每一类型问题各设计了 4 个题目,共 16 个题目。这些题目随机出现在调查表中,考察被调查者选择的第一检索策略。

实验调查发现,人们在进行检索策略的选择时,往往喜欢使用输入检索词、搜索引擎、主题指南、直接输入地址这 4 种策略来解决问题,而其他方法用得很少。为此,本实验将全部 16 个问题使用这 4 种策略来考察。通过实验调查,考察每个问题人们采取的第一检索策略。这 16 个问题是:

查找具体性问题:

- (1) WTO 现有多少个成员国?各大洲分布情况。
- (2) 进入世界百强企业的中国企业有哪些?
- (3) Intel 公司现任总裁是谁?
- (4) 据最新统计,全世界目前人口有多少?

查找细节性问题:

(5) 2001 年利用因特网的用户情况,如:国别、年龄、性别、收入等。

(6) 2001 年世界电子书增长最迅速的领域?

(7) 加拿大近 2 年来有哪些针对移民的新政策?

(8) 下一届 IFLA 大会在什么地方召开?讨论的主题是什么?

查找方向性问题:

(9) 查找有关人工智能和电子商务方面的资料。

(10) 在预防艾滋病工作中,世界卫生组织都做了哪些努力?

(11) 如何申请去英国牛津大学留学?

(12) 美国国会图书馆在实施数字图书馆计划中是如何解决版权问题的?

查找可能性问题:

(13) 9.11 事件对美国社会带来了极大影响,具

具体情况如何?

(14)在世界杯足球赛中,引起暴力行为和骚乱的原因是什么?

(15)在 Web 上能够获得哪些北极熊的图片资料?

(16)在保护知识产权方面,欧洲和美洲的做法有何不同?

2 实验的数据来源

本实验的调查对象是黑龙江大学 3、4 年级的学生,他们分别来自计算机科学技术学院和电子工程学院,其中计算机科学技术学院的学生占 80%。被调查的这些学生对网络非常熟悉,经常利用学校的计算机或个人的计算机上网。他们对网络信息的检索能力和对问题的理解能力都很强。在被调查的学生中,有 62 名男生,44 名女生,共 106 人。全部问题以调查表的形式发给每个学生,他们经过认真思考,对每个问题选出自己认为最满意的答案。本实验得出的各种数据正是基于该调查表完成的。

3 WWW 环境下对不同问题采取的相应策略实验调查

本实验提供的 16 个问题是随机出现在调查问卷上的,被访者并不知道题目的类型,调查结果如表 1。

表 1 不同类型问题检索策略统计

序号	输入地址	主题指南	搜索引擎	输入检索词
1	9	30	24	43
2	11	42	24	30
3	38	20	17	30
4	2	38	34	32
5	11	34	40	21
6	5	45	32	24
7	18	43	33	12
8	9	25	29	43
9	1	23	36	44
10	27	35	18	26
11	40	33	25	9
12	31	38	27	11
13	6	48	18	34
14	17	28	22	41
15	20	6	56	24
16	7	31	38	28
合计	252	519	473	452
比例(%)	14.86	30.60	27.89	26.65

从总的检索结果看,选择每种检索策略的数量依次为:主题指南 519 个,占 30.60%;搜索引擎 473

个,占 27.89%;输入检索词 452 个,占 26.65%;直接输入地址 252 个,占 14.86%。但涉及到具体类型的问题时,情况又有所不同。表 2~表 5 总的检索策略数均为 424(即:被调查的学生总数 \times 4 种策略)个,所得到的百分比是与总数 424 之比。

表 2 查找具体问题检索策略统计

序号	输入地址	主题指南	搜索引擎	输入检索词
1	9	30	24	43
2	11	42	24	30
3	38	20	17	30
4	2	38	34	32
合计	60	130	99	135
比例(%)	14.15	30.66	23.35	31.84

从表 2 看出,查找具体性问题采用的检索策略从多到少依次为:输入检索词 135 个,占 31.84%;主题指南 130 个,占 30.66%;搜索引擎 99 个,占 23.35%;直接输入地址 60 个,占 14.15%。

出现上述结果的原因是,查找具体性问题需要找到确切的答案,而输入检索词和主题指南有利于查找明确答案。特别是输入检索词,目标更加明确。使用搜索引擎排在第 3 位是因为搜索引擎涉及的范围较广,不利于作为第一检索策略,而直接输入地址需要被访者事先记住大量的 URL 地址,非常麻烦。

表 3 查找细节问题检索策略统计

序号	输入地址	主题指南	搜索引擎	输入检索词
5	11	34	40	21
6	5	45	32	24
7	18	43	33	12
8	9	25	29	43
合计	43	147	134	100
比例(%)	10.14	34.67	31.60	23.59

从表 3 看出,查找细节性问题采用的检索策略从多到少依次为:主题指南 147 个,占 34.67%;搜索引擎 134 个,占 31.60%;输入检索词 100 个,占 23.59%;直接输入地址 43 个,占 10.14%。

出现上述结果的原因是查找细节性问题需要查找的结果比较明确,有一定范围,但信息来源不确定,无法预测。主题指南和搜索引擎分别有利于查找结果较明确、信息来源不确定的问题,因而,选择主题指南和搜索引擎的人数居多。输入检索词排在第 3 位是因为该项检索策略过于专指,使得选择的人数少于前两种。使用直接输入地址的人数仅有 43 个,占 10.14%,因为这种

检索策略过于繁杂,需要事先知道 URL 地址。

表4 查找方向性问题检索策略统计

序号	输入地址	主题指南	搜索引擎	输入检索词
9	1	23	36	44
10	27	35	18	26
11	40	33	25	9
12	31	38	27	11
合计	99	129	106	90
比例(%)	23.35	30.42	25.00	21.23

从表4看出,查找方向性问题采用的检索策略由多到少依次为:主题指南129个,占30.42%;搜索引擎106个,占25.00%;直接输入地址99个,占23.35%;输入检索词90个,占21.23%。

出现上述结果的原因是,查找方向性问题时,被访者在分析题目时易与细节性问题混淆,在问题的认识上容易产生偏颇。此外,像主题指南这样的检索策略不仅适用于查找结果确切的问题,也适用于结果范围较广的问题。从统计的结果,我们第一次看到直接输入地址突破20%的比例,而且还超过了输入检索词,与其他3种检索策略的差距很小,这在4个问题的调查中是惟一较高比例的一次。

表5 查找可能性问题检索策略统计

序号	输入地址	主题指南	搜索引擎	输入检索词
13	6	48	18	34
14	17	28	22	41
15	20	6	56	24
16	7	31	38	28
合计	50	113	134	127
比例(%)	11.79	26.65	31.60	29.96

从表5看出,查找可能性问题采用的检索策略依次为:搜索引擎134个,占31.60%;输入检索词127个,占29.96%;主题指南113个,占26.65%;直接输入地址50个,占11.79%。搜索引擎第一次超过了主题指南和输入检索词排在第1位,这说明人们在查找可能性问题时趋于使用搜索引擎。但输入检索词的比例与搜索引擎也十分接近,表明被访者青睐使用这两种策略。而其他两种策略中,主题指南居第3位,使用者也较多。只有直接输入地址的人较少,仅占11.32%,与其他检索策略相差较远。

4 结论

通过上述对查找具体问题、细节性问题、方向性问题和可能性问题在WWW环境下检索策略的调查,可以初步形成针对不同问题采取的相应策略,这

些策略是被访者的第一检索策略。表6是针对不同类型问题采取的第一检索策略排序,所有顺序皆按使用频率的多少由高到低排列,结果见表6。

表6 不同类型问题检索策略频率统计

类别	第1位	第2位	第3位	第4位
具体性问题	输入检索词	主题指南	搜索引擎	输入地址
细节性问题	主题指南	搜索引擎	输入检索词	输入地址
方向性问题	主题指南	搜索引擎	输入地址	输入检索词
可能性问题	搜索引擎	输入检索词	主题指南	输入地址

5 存在的问题及今后研究方向

上述结果是通过实验调查得出的结论,它只涉及了问题的第一检索策略,只是对解决问题的一点粗浅认识,还需要今后进一步在理论上和实践上研究探讨。此外,在实验中难免出现这样或那样影响正确结论的因素,如被访者对问题的理解程度、实验数据的采集是否规范等,这些都是在以后的研究工作中应该引起注意的。

今后的研究工作,应继续以下几方面的工作:细致地研究各类问题与检索策略间的关系;探讨每类问题可能涉及的所有检索策略;广泛采集数据,增加实验组数,将实验做得更深更细;扩大被访者层次,不仅仅局限于某一局部领域。

参考文献

- 1 储荷婷等. Internet 网络信息检索. 北京:清华大学出版社,1999
- 2 潘卫. Internet WWW 信息检索策略初探. 情报学报, 1999(增刊)
- 3 宋玲,马军. Internet 信息检索分析与研究. 现代图书情报技术, 2001(1)
- 4 曾民族. 网络信息检索现状与性能评价. 情报学报, 1997(2)
- 5 雷燕. WWW 信息检索技巧. 现代图书情报技术, 1998(6)
- 6 Marilyn Domas Whites, Mirja Iivonen. Questions as a factor in Web search strategy. Information Processing and Management, 2001(37)
- 7 Jacques Savoy, Justin Picard. Retrieval effectiveness on the Web. Information Processing and Management, 2001(37)
- 8 <http://www.Fsearch.com.cn/Fsearch1.html/>
- 9 <http://sanf.3322.net/seek/seekgengduo2.html/>
- 10 <http://lib.yctc.edu.cn/xhm/myfiles/comsearch.html>
- 11 <http://chen1999.363.net/resource/>
- 12 <http://www.filinfo.gov.cn/publicat/gbts/001/11.htm>

黄丽霞 黑龙江大学信息管理学院副院长、副教授。
通讯地址:哈尔滨市。邮编150080。

(来稿时间:2002-10-23)