

# TREC 评测项目概述

何四维

(中南大学湘雅医学院 湖南长沙 410013)

**摘要:**介绍了文本检索会议(TREC)的研究目标和意义,对历年进行的评测项目进行了介绍,并对1992~2011年TREC发展状况进行了阐述。

**关键词:**文本检索会议 信息检索 评测

**中图分类号:**G354.4

**文献标识码:**A

**doi:**10.3969/j.issn.1005-8095.2012.01.014

## 1 TREC 的历史

文本检索会议(TREC)是由美国国家标准技术协会(NIST)和美国国防部共同主办的,文本检索领域最为权威的评测会议。目前,TREC所涵盖的内容早已超越了它的名称(文本检索)。1992年,作为TIPSTER检索项目的一部分,TREC成功举办了第一届会议<sup>[1]</sup>。TREC每年举办一次,会议为参与者提供一个标准的文档库、统一的评测标准,对各种系统和方法进行公平的评测。一般在2月份组织报名参加当年的评测会议;5~6月份会议发布统一标准的语料库、问题集以及评测方法;7~9月份会议主办方收集参与者的检索结果,并对检索结果进行评测;10~11月份主办方发布评测结果和会议论文集。直至2010年,TREC已成功开展了19次评测会议。

### 1 TREC 的研究目标和意义

TREC的目的是为大规模的文本检索评价提供必要的基础条件,以支持信息检索领域的研究。TREC主要有以下4个研究目标:

(1)鼓励进行大规模的文本检索研究。为了反映

现实系统的主题多样性,必须保证有足够的实验语料集,TREC的文献集合一般在2GB左右,包括50万~100万篇文献<sup>[2]</sup>。(2)通过建立开放式的论坛,加强企业、学术机构的研究思想的交流。(3)示范信息检索理论在解决实际问题方面的重大进步,加速实验室技术向商业应用的转化。如果某种技术表现出良好的实验效果,那么它就可能比较迅速地得到商业化的应用。(4)为工业界和学术机构提高评测技术的可用性,并开发新的评价技术,使之更适用于当前的系统。

TREC使用的评估软件是开放的,这使得任何组织和个人都可以利用TREC提供的软件对自己的系统进行评测。TREC体现的是先进、实用的检索技术和系统,无实用性的系统和技术将会遭到淘汰。目前,TREC已成功地达到了促进最新信息检索技术的转化的价值目标。在TREC开展的第6年,检索系统效能大约增加了一倍,目前所使用的大多数商业搜索引擎和技术都最早出现在TREC<sup>[3]</sup>。

### 2 TREC 的评测项目简介

绿色信息通道,提供优于一般读者的信息服务等。

总之,图书荐购是借助读者的知识与能力提高馆藏质量及馆藏利用率的有效途径。消除图书荐购中一切不利于读者的因素,采取相应的激励措施,有助于吸引更多的读者参与荐购图书,有助于实现图书馆馆藏建设和读者利用图书馆藏书的双赢。

#### 参考文献

- [1] 范春梅. 高校图书馆资源荐购服务探讨 [J]. 图书馆建设, 2010(5): 11-13, 17
- [2] 罗毅. 高校图书馆荐购系统现状与问题研究 [J]. 图书馆学研究, 2010(10): 46-49, 26
- [3] 姜广飏, 等. 利用移动学习模式构建大学公共外语学习的新平台 [J]. 科技信息, 2009(23): 31
- [4] 蒋建国. 我国图书出版品种 总印数居世界第一 [EB/ON]. [2011-04-28]. [http://book.ifeng.com/culture/gundong/detail\\_2010\\_08/19/1987842\\_0.shtml](http://book.ifeng.com/culture/gundong/detail_2010_08/19/1987842_0.shtml)
- [5] 刘振喜. 发挥优势 继承传统 贴近科研 完善服务 [N]. 中国社会科学院院报, 2006-01-05(2)

质量上乘的图书不能被荐购。为了提高馆藏的多样性及与读者需求的吻合度,提高读者荐购图书的自由度,必须把荐购范围扩大到一切正规的出版源(包括合法的内部出版源)及图书以外的电子资源、音像制品等新媒体信息产品。

#### 4.4 及时处理荐购要求并回复

及时处理读者的荐购,既是对读者所付出的劳动的尊重,也是做好图书荐购工作的保证。图书馆应安排专人负责处理读者的荐购要求,及时将是否接受荐购的结果和理由通知读者,并向读者表示感谢。

#### 4.5 完善激励机制

读者向图书馆荐购图书的行为是值得鼓励的公益行为,图书馆应该制订相应的奖励制度。目前普遍的奖励措施是成功荐购图书者可优先借阅所荐购的图书<sup>[5]</sup>。笔者认为还可以采取以下的措施给予奖励:对用短信荐购数量多的读者适当补贴话费;每年评选一次荐购图书先进个人;为荐购图书的读者开通

TREC 评测项目的更新非常快,每年都会引进新的评测项目,并对以前的项目进行总结淘汰,直至

2011年,共有26个评测项目出现过,表1对这些项目进行简要的介绍。

表1 TREC 评测项目简要介绍

项目名称	简要介绍
众包	为2011年最新项目,该任务的基本思想是利用最新出现的“众包”的思想,即将工作分配给众多团体或个人,为检索的评价方法寻找新的途径和形式,来优化检索评价过程,并开发新型的自动群体检索系统,以节省人力物力
电子病历检索	为2011年最新项目,电子病历作为患者就医的医疗记录档案,不仅是临床进行正确诊断、选择治疗方案和护理的科学依据,也是卫生统计机关资料来源的基础。TREC的电子病历检索项目是在这个背景下提出的,对基于文本免费的电子病历进行检索和评测
微博检索	为2011年最新项目,随着各种多媒体交流平台的产生,微博以它的便捷性、易获得性、交互性逐渐在网络用户中占据重要的位置,给信息检索领域带来了新的挑战和机遇。TREC的微博检索任务就是在此环境下的检索行为的检索任务和评价方法进行的评测
Session 检索	强调用户的交互作用,通过一系列的检索来帮助评测信息检索系统
Chemical IR 检索	此项目通过构建大型化学实验平台来评价最新的化学信息检索技术和工具。主要任务包括检索化学方面的专利文献、生物-化学领域的相关技术,或者给出化合物的图片描述文件,检索得到该化合物的结构描述文件
实体检索	在万维网环境下评价针对实体的检索任务。此项目更好地执行需要返回具体的实体对象,而非文本信息
新 Web 检索	新的 Web 项目,主要是执行具体的 Web 检索任务,通过检索超过 10 亿个网页,评价检索的多样性和有效性
Million 检索	对基于多而完备的预设信息需求的测试集优于传统 TREC 集这一假设进行检验
博客检索	对博客领域的信息查找行为进行探讨
法律检索	开发适合律师的搜索技术,帮助他们在数字文献集中有效检索信息
企业检索	对企业搜索进行研究,满足用户为了完成任务而对某组织进行有针对性的信息搜索需求
SPAM 检索	对当前垃圾邮件过滤方法进行评估,为一般的邮件过滤和检索评估奠定基础
Terabyte 检索	探讨信息检索领域优于当前 TREC 所用的、对大型文献集基于传统信息检索测试集的评估
Robust 检索	包括传统的 Ad hoc 检索任务,但是与之不同的是它更加注重个人用户需求检索的有效性,而非平均有效性
基因	研究特定领域——基因学领域的检索(不仅仅包括基因序列,而且包括研究文献、实验室报告等)
新颖性检索	研究检索系统定位新信息的能力
视频信息检索	研究包括数字媒体的自动分段、索引和基于内容的检索
相关反馈检索	为影响相关性反馈结果的不同因素提供一个框架
HARD 检索	通过利用检索者和检索背景等额外信息,采用诸如片段检索或与检索者有关联的交互等技术,提高文献检索的准确性
问答系统	针对信息进行的检索,而非针对文献的检索
Web 检索	根据文献集的搜索任务的特征,对万维网进行概化。此项目于 2004 年就结束了
跨语言检索	在跨语言环境中研究检索系统查找文献的能力
过滤检索	此任务的用户的的需求是稳定的,并且相关文献都是已知的,但是还是有一些新的文献。对于这些文献,系统必须根据用户的需求以确定他们与用户需求的相关性。
交互	是一项研究文本检索系统的用户交互的项目。参与团队制定统一的实验协议,研究普通人群作为使用者的查询行为,探讨交互技术对检索效果的评价。此项目于 2003 年并入 Web 项目
Routing	定题检索实验,信息需求相对固定,语料处于变动中
Ad Hoc	传统的批处理检索实验,类似于传统图书馆查找资料的过程,语料集已知,目标任务未知

### 3 TREC 的发展状况

#### 3.1 TREC 参与者队伍逐渐壮大

TREC 的参加者是来自各个国家 IT 界一流的学术研究机构、企业和政府机构,如麻省理工大学、斯坦福大学、日本东京大学、微软研究院、Google、IBM 研究院等。近几年,我国一些学术机构也陆续加入了 TREC 的队伍,如 2010 年,中国参与 TREC 项目的机构有:北京大学、清华大学、北京理工大学、北京邮电大学、中国科学院、大连理工大学、复旦大学、黑龙江理工大学等。从图 1 可以看出,参与 TREC 项目的机构总体呈上升趋势,这也从侧面反映出 TREC 在信息检索评测方面的权威性和影响力。

#### 3.2 TREC 评测项目多样化

检索平台的发展、多媒体信息的出现、用户的需求的增加,传统的仅针对文本信息进行的评测已经不能满足信息检索的发展,经过多年的发展和众多

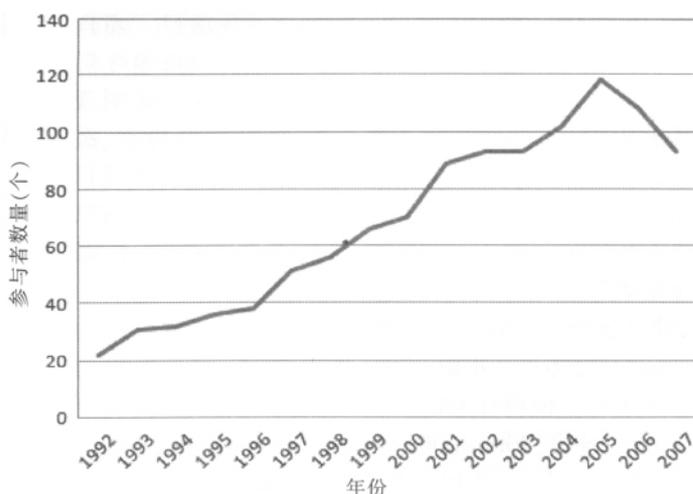


图1 TREC 参与者数量

参与者的实践,从表2我们可以看出TREC的评测项目和评测标准是在不断发展和完善的。TREC最初3次会议的任务主要是进行批处理检索试验(Ad

表2 1992~2011年TREC项目参与者数量

(单位:个)

项目名称	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
众包																				√
电子病历检索																				√
微博检索																				√
Session 检索																			√	√
Chemical IR																		8	√	√
实体检索																		13	√	√
新 Web 检索																		26	√	√
Million 检索																11	7	8		
博客检索															16	24	20	11	√	
法律检索															6	14	10	11	√	√
企业检索															23	25	20			
SPAM 检索															13	9	12			
Terabyte 检索													17	19	21					
Robust 检索												16	14	17						
基因												29	33	41	30	25				
新颖性检索												13	14	14						
视频信息检索										12	19									
问答系统								20	28	36	34	33	28	33	31	28				
Web 检索									17	23	30	23	27	18						
跨语言检索							13	9	13	16	10	9								
过滤检索				4	7	10	12	14	15	19	21									
交互			3	11	2	9	8	7	6	6	6									
Routing	16	25	25	15	16	21														
Ad Hoc	18	24	26	23	28	31	42	41												

注:2010~2011年数据尚未公布。

Hoc)和定题检索实验(Routing)。1992年,NIST与其他参与组织一起制定了50个检索提问(Topic),并要求参与者对每个检索提问进行检索,并列出了前200个文档。TREC1对这次检索进行了总结,并对遇到的问题进行了讨论。例如,检索提问过长,平均有107个词;得到的相关结果太多而不准确,每个检索提问平均得到277个相关的检索结果<sup>[4]</sup>。这是对TREC1的总结,也是对TREC2的启示。在第2次会议中,要求参与者总共列出1000个文档。

每一年的TREC都会引进新的评测项目,而且在会议总结时会总结每个项目的发展,以决定今后的发展。直至2011年TREC-20,共有26个评测项目出现过<sup>[5]</sup>。新的项目不断出现,例如微博检索、医疗记录检索等新的研究项目的产生;过时或达到目标的项目则退出历史舞台或者并入其他项目,如以前曾经出现过的视频检索项目、机器人检索项目、博客检索项目等多个项目已经完成其任务,退出了TREC的舞台;2002年人机交互项目完成阶段性的目标,从2003年开始,并入Web项目,作为其中的一个子项。值得注意的是,TREC的研究项目具有更替性。例如,2004~2006年开展的Terabyte检索项目,在2007年由Million检索项目所代替。而2006~2010年开展的博客检索项目,在2011年由微博检索项目所替代。这种替代是对时代的反映,更是对检索系统和技术的挑战。

总之,经过多年的发展,TREC在信息检索领域

已经建立了其权威的地位,其国际化的影响,吸引了越来越多来自世界各地的高水平的参加者,中国也有越来越多的学术机构承担了TREC的项目。但是综观国内多年来在信息检索和系统评价方面的现状,我们不得不承认与国际上的差距。目前,网络信息检索的评价正在如火如荼地进行,我们不仅需要了解国内外的发展趋势,与之共同发展,共同提高,也要结合中国特色的信息资源,如中文搜索引擎,积极承担TREC的项目,以促进检索系统的发展,信息检索效率的提高和我国信息检索评价的发展。

## 参考文献

- [1] R. Merchant. TIPSTER phase I: Program overview [C]. In The Proceedings of the TIPSTER Text Program - Phase I, 1994: 1-2
- [2] 张秀坤,赵丹群. TREC概况及其最新发展研究[J]. 情报理论与实践, 2004(5): 537-540
- [3] National Institute of Standards and Technology. Overview[EB/OL]. [2010-08-10]. <http://trec.nist.gov/overview.html>
- [4] Donna Harman. The Development and Evolution of TREC and DUC[R]. The Third NTCIR Workshop, 2001-2002
- [5] National Institute of Standards and Technology. 2011 TREC Tracks[EB/OL]. [2011-03-29]. <http://trec.nist.gov/tracks.html>

收稿日期:2011-05-12

作者简介:何四维(1988—),女,2010级医药信息系情报学硕士研究生。