2.2　网络数据挖掘对智能检索的支持

无论是个性化信息检索,还是基于内容的检索,乃至知识检索,网络信息检索系统智能化的关键是知道用户需求什么、提供给用户的东西具有高质量(内容上相关、知识含量高)。数据挖掘对网络智能检索的支持正体现在对用户需求信息和网络源信息的深层分析,以提供智能检索必需的关键知识。

2.2.1　用户知识的挖掘

虽然个体信息用户的信息需求具有特定性,但从用户群整体来看,用户的信息需求又是随机的,这为一般的用户需求信息分析带来很大的困难。数据挖掘从全局出发,以丰富、动态的联机查询和分析来了解用户的信息需求。通过在线提问、调查表等方式,系统可以获取关于用户的用户名、用户访问 IP 地址、用户的职业、年龄、爱好等原始信息。然后,采取一定的挖掘规则(如关联规则、联机分析处理等),对这些数据进行融合分析,其结果是为每个用户建立一个信息需求模型。而且全方位的用户需求信息挖掘,可以将同类信息需求的用户联系起来,从而实施"以一对众"的检索方案。目前用户知识挖掘已步入实用阶段,如 IBM 公司新推出的 DB2 UDB 7.1 就是一种比较理想的用户知识挖掘工具。

2.2.2　网络知识的挖掘

网络知识的挖掘就是要在具有极度不确定性的海量数据中找出信息分布的规律,挖掘隐藏的信息并形成模型,从而发现具有规律性的知识。网络信息分布的规律性就是网络信息内在的关联性。对网络信息这种关联性的挖掘主要体现在两个方面:一是网络内容挖掘。网络内容挖掘是一个从文本、图像、音频、视频、元数据等形式的网络源信息中采用分类、聚类等形式的挖掘方法,发现有用信息,并将这些信息按满足某种检索方式的形式加以组织的过程。二是网络结构挖掘。网络结构挖掘是通过分析一个网页链接和被链接数量以及对象来建立 Web 自身的链接结构模式。这种模式可以用于网页归类,并且由此可以获得有关不同网页间相似度及关联度的信息。这种链接结构模式有利于实现智能导航。

3　结束语

将数据挖掘技术引入到网络资源的开发中来,能加快智能检索的发展。数据挖掘的结果是实现智能检索的基础,智能检索的结果可为数据挖掘提供指南和线索。目前,以数据挖掘技术为主导的网络数据挖掘产品已得到应用。例如,Net Perceotion 公司开发的 Net percerptions,能够挖掘用户信息,从而为实现个性化信息服务打下基础。在开发无尽的网络数据资源时,若能结合机器学习、模式识别等其他人工智能技术,网络数据挖掘技术将在实际应用中更加完善。

参考文献

1　毕强,闫凤英等. Web 信息发布的特征与完全信息结构. 情报学报,2000(6)
2　李爱红. 网络搜索引擎的比较研究. 中国信息导报,1999(1)
3　王军,杨冬青等. 数字图书馆的检索技术. 计算机世界,2000-02-21
4　邹涛,戚广智等. 网络信息挖掘系统 IDGS 的实现. 南京大学学报(自然科学版),2000(2)
5　朱建秋,周皓峰等. 一个基于关联规则的数据采掘工具的设计和实现. http://dmgroup.myetang.com/lw3.htm
6　Hearst. Distinguishing between Web Data Mining and Information Access. Presentation for Web Data Mining Panel at KDD'97
7　Dan R. Greening. Data mining on the web. Web Techniques, 2000(1)

晏创业　武汉大学信息管理学院。通讯地址:湖北武汉。邮编 430072。
张玉峰　武汉大学信息管理学院教授。通讯地址同上。
(来稿时间:2001-10-18)

《永乐大典》编纂 600 周年国际研讨会暨仿真影印出版首发式举行

4 月 17 日,由国家图书馆主办的《永乐大典》编纂 600 周年国际研讨会暨《永乐大典》仿真影印出版首发式在京举行。全国政协副主席罗豪才到会表示祝贺。来自中国、美国、英国、法国、澳大利亚、俄罗斯、日本、韩国等海内外 50 余个研究机构和收藏单位的 90 余位专家学者参加了会议。

《永乐大典》是明成祖朱棣于明永乐元年(公元 1403 年)命太子少师姚广孝和翰林学士解缙主持、3000 多文臣历时 4 年纂修而成。共辑录图书 8000 种,上自先秦,下迄明初,天文地理,人事名物,无所不包。整部典籍共 22877 卷,外加目录等 60 卷,装成 11095 巨册,全部用毛笔工楷书写,是世界上最早、最宏伟的百科全书。目前,存世的《永乐大典》副本零册约 400 册左右,约 800 余卷,不到原书的 4%,分散在 8 个国家和地区。《永乐大典》是中国国家图书馆的四大珍藏之一,从 1912 年第一批《永乐大典》入藏到现在,已拥有 221 册,超过全球藏量的半数,居各处收藏的首位。

在《永乐大典》编纂 600 周年国际研讨会上,来自世界各地的专家、学者就各国收藏、保护、研究《永乐大典》的状况,《永乐大典》的修复、保护、数字化和出版情况进行了广泛而深入的讨论。相信本次会议必将对《永乐大典》的收藏、保护、研究、出版、数字化起到极大的推动作用。

近年来,《永乐大典》在世界各地又续有发现,而这些新发现的《永乐大典》也急需刊布,以嘉惠学林,使其得到充分利用,因此,中国国家图书馆委托北京图书馆出版社将现存我馆的 161 册连同现存国内其他馆收藏的 2 册《永乐大典》率先仿真影印出版。

该社从 2001 年 12 月开始,用特制宣纸,套色印刷,原大仿真分批出版现存于世的《永乐大典》。拟用一年半时间先首批出版收藏于中国大陆的 163 册的大陆珍藏版。待首批出完,再用一年半的时间陆续出版现藏于海外的 200 余册,使之为学界、为大众所共享。