

水资源学研究的关键词共词聚类分析

王富强,韩宇平,王朋,王静

(华北水利水电大学水利学院,河南郑州 450045)

摘要:以中国知网学术期刊网络出版总库中2003—2012年的水资源学科核心期刊论文为基础,采用Bibexcel统计关键词频次,采用Ucinet和Netdraw绘制共词网络,运用SPSS进行共词聚类分析与多维尺度分析,研究高频关键词之间的内在联系,评述水资源学科的研究热点与研究现状。研究表明,水资源学科在研究过程中形成了水资源可持续利用、社会可持续、水资源系统和水资源管理四大类群体,并且四者之间具有较强的关联性。

关键词:水资源;关键词;共词聚类分析;共词强度;共词矩阵

中图分类号:TV213.4;G353.1 **文献标志码:**A **文章编号:**1006-7647(2014)04-0029-05

Co-word cluster analysis of keywords in water resources science research//WANG Fuqiang, HAN Yuping, WANG Peng, WANG Jing(*North China University of Water Conservancy and Electric Power, Zhengzhou 450045, China*)

Abstract: Cluster analysis and multivariate statistics analysis for the co-words are performed taking into account the keywords in core journal papers on water resources issued in China Academic Journal Network Publishing Database (CAJD) during 2003 to 2012. Further, the keyword frequency counted by Bibexcel, co-word network set up by Ucinet and Netdraw, co-word clustering analysis and multidimensional scaling analysis by SPSS, internal relations among high frequency keywords were studied. Additionally, the research focus and research status were analyzed. The results show that in the course of water resources science research four major communities of water resources research are developed: sustainable utilization, social sustainable development, water resources system and water resources management. Furthermore, these four major communities are greatly related with each other.

Key words: water resources; keyword; co-word cluster analysis; co-word intensity; co-word matrix

随着期刊数量的增长与学科的细化发展,科技文献数量增长迅速,这给学者的研究工作带来了极大便利,相应地对文献的管理与分析利用也有了更高要求^[1]。随着信息技术的发展、应用和开辟,有必要整理与审视某一学科的研究热点与研究现状,以促进该学科的进一步发展,近几年运用文献计量学研究学科的发展已成为一种不可或缺的手段。自2003年以来,水资源的期刊论文数量增加显著,但是运用文献计量学来研究水资源学科的研究热点与研究现状的相关论文还没有见到,本文尝试以中国知网学术期刊网络出版总库在2003—2012年的相关论文为基础,采用共词聚类分析法来探讨水资源学科的研究热点与研究现状,为该学科的发展和建设提供参考。

共词聚类分析法属于一种内容分析法,主要统计一组词中两两同时出现于一篇文献的频次(共词强度),以这种共现频次反映这些词之间的关联程度,两个词的共词强度越高,则它们之间的关联越紧密^[2]。关键词是文献检索标识,是表达文献主题概念的自然语言词汇,能够简单、直接、全面地概括文献的核心研究内容。高频关键词可以反映学科的研究热点,而关键词的变化也可以在一定程度上反映学科的发展趋势^[3]。关键词共现分析法是指两个关键词同时出现在一篇文章中的频次,反映了这两个关键词之间的关联性,关联性强的关键词代表了学科研究的核心问题和热点问题^[4]。本文采用文献计量软件Bibexcel统计关键词频次、使用Ucinet和Netdraw绘制共词网络,同时利用SPSS软件进行共词聚类分析以及多维尺度分析,并与共词网络进行对比,来揭示水资源领域的研究热点和研究现状。

1 高频关键词

在中国知网中输入“水资源”进行主题检索,限定为核心期刊检索,检索时间为2003—2012年,共检索出符合条件的期刊论文15 090篇,论文数量年度分布如表1所示。

表1 水资源学科期刊论文数量年度分布

年份	篇数	年份	篇数
2003	843	2008	1 790
2004	1 148	2009	1 756
2005	1 204	2010	1 730
2006	1 334	2011	1 917
2007	1 525	2012	1 843

由表1可以看出,在2003—2008年间水资源学科的核心期刊论文数量不断增加,在2008年以后,论文数量维持在1800篇左右,比较稳定。对表1中的论文,依据关键词、主题以及文章标题来排除与水资源学科相关性较小的论文。由于在一些论文中关键词的表述不太规范,本文对于部分同义词和近义词进行了合并处理,如将水资源承载力、水资源承载能力、承载能力合并为水资源承载力;将水资源配置、水资源优化配置合并为水资源配置等。使用Bibexcel统计2003—2012年与水资源学科相关性较大的期刊论文3 314篇,其中关键词共4 737个,频次大于10的有174个,频次大于20的有63个。去除频次最高的关键词“水资源”,剩余62个关键词在一定程度上反映了水资源学科研究的核心问题与热点问题,关键词频次的降序排列如表2所示。

从表2可以看出,近10年的水资源研究现状是:①关于水资源承载力、水资源配置的文献较多,一直是研究热点,这与中国的水资源短缺有很大关系;②水资源可持续利用与可持续发展出现也较为频繁,这与社会可持续发展理念具有紧密的关联性,因此研究较多;③气候变化出现了105次,通过查阅

表2 水资源学科高频关键词

关键词	频次	关键词	频次	关键词	频次	关键词	频次	关键词	频次
水资源承载力	280	水资源保护	57	水资源评价	38	水足迹	28	预 测	23
水资源配置	277	黄河流域	52	水资源安全	37	人类活动	28	节水灌溉	23
可持续利用	184	水环境	51	中 国	37	水 质	28	GIS	23
可持续发展	180	黄 河	50	模 型	36	多目标	27	西北地区	23
水资源管理	149	虚拟水	48	水污染	35	水资源开发	27	水循环	23
指标体系	111	石羊河流域	47	遗传算法	34	农业水资源	27	新 疆	22
气候变化	105	节 水	46	水 价	34	降 水	26	水资源开发利用	21
生态环境	85	长江流域	46	供需平衡	34	模糊综合评价	25	干旱区	21
对 策	73	流域	43	管 理	33	水资源短缺	25	水资源价值	20
水资源利用	72	黑河流域	42	合理配置	33	南水北调	24	水资源量	20
评 价	70	层次分析法	41	系统动力学	32	灌 区	24		
综合评价	70	地下水	40	水资源系统	32	径 流	24		
水 权	58	开发利用	38	城市化	30	主成分分析	23		

文献发现,国内关于气候变化对水资源影响的研究起步较晚,但发展较快;④从水权、水价等关键词发现对于水资源经济学的研究也较多;⑤水资源学科的主要研究方法有层次分析法、遗传算法、主成分分析法等,其中越来越多的借助GIS来研究水资源;⑥虚拟水作为水资源一个新兴的研究方向得到了较多的关注;⑦黄河流域和长江流域是水资源研究的典型区域,出现频次较高。

2 共词网络分析

高频关键词在一定程度上代表了水资源学科的研究热点和研究方向,但仅按出现频次对这些词进行线性排列,还不能全面反映它们之间的关系,需要进行共词分析。首先采用Bibexcel构建共词矩阵,对以高频关键词进行两两配对,统计它们在3 314篇文献中共现频次,形成62×62的共词矩阵,主对角线的数据为该词出现的频次,非主对角线上的数据表示两个关键词共同出现在同一篇论文中的频次,部分数据如表3所示。

表3 部分高频关键词共词矩阵

关键词	水资源承载力	水资源配置	可持续利用	可持续发展	水资源管理	指标体系	气候变化	生态环境	对策
水资源承载力	280	3	5	9	2	4	0	2	2
水资源配置	3	277	9	12	7	3	0	3	1
可持续利用	5	9	184	1	5	8	2	2	18
可持续发展	9	12	1	180	9	3	2	7	4
水资源管理	2	7	5	9	149	0	1	1	2
指标体系	4	3	8	3	0	111	0	1	0
气候变化	0	0	2	2	1	0	105	1	4
生态环境	2	3	2	7	1	1	1	85	1
对 策	2	1	18	4	2	0	4	1	73

多元统计分析方法对矩阵的数据结构有不同的要求,为了统计分析方便,本文将共词矩阵转化为相关和相异矩阵。由于关键词的频次悬殊,共词矩阵中数据相差也比较大,为了消除频次悬殊造成的影

响,可以用式(1)计算得 Ochiai 系数,从而得到共词相异矩阵。

$$y = \frac{x_{AB}}{\sqrt{x_A} \sqrt{x_B}} \tag{1}$$

式中: y 为 Ochiai 系数; x_{AB} 为 A、B 两词同时出现的频次; x_A 、 x_B 分别为 A 词、B 词出现的次数。

为了呈现关键词之间的内在相关关系,需要借助共词网络与多元统计分析方法来揭示这些关键词之间的关系以及它们所反映的研究主题。根据高频关键词共词矩阵,利用 Ucinet 将共现矩阵转化为 .h 的文件,用 Netdraw 绘制共词网络图谱,得到高频关键词共词网络可视图,如图 1 所示。

图 1 中节点越大表示其中心度越高,是网络中最重要的节点。节点之间连线的粗细表示两组关键词出现频次的高低,频次越高,关键词之间关系越密切。由图 1 可以看出,水资源配置、可持续利用、可持续发展处于网络的中心位置,其中心度最大,在一定程度上也反映了水资源学科的主要研究内容与研究主题,围绕它们形成了水资源保护、水资源短缺、水资源利用、水污染、气候变化、生态环境等研究热点结构,它们之间具有紧密的关联性。另外,黄河流域、长江流域与其他关键词之间的连线较密且较粗,共现的频次较高,关系较为紧密,它们是近 10 年来水资源研究的主要区域。

3 共词聚类分析

以共词矩阵为基础,采用聚类分析法对高频关

键词之间的连接强度进行分析。聚类分析是根据事物本身的特性研究个体分类的统计方法,它基于数据的相似性分类,将当前最紧密的对象合并为一类,直到所有个体聚为一个大类^[5]。聚类分析的基本思想是认为研究的变量之间存在着程度不同的相似性(亲疏关系),可根据一批研究对象的多个变量指标,具体找出一些能够测度这些变量指标之间相似程度的统计量,以这些统计量为划分依据,把一些相似程度较大的变量聚合为一类,关系密切的聚合到一个小的分类单位,关系疏远的聚合到一个大的分类单位,直到把所有的变量都聚合完毕,把不同的类型一一划出来,形成一个由小到大的分类系统^[6]。在对关键词进行聚类时,首先将每一个关键词看成独立的一类,把最近的两类合并,然后重新计算类与类之间的距离,再把距离最近的两类合并,每一步减少一类,一直持续到所有的关键词归为一类^[7]。以公式(1)构造的相异矩阵为基础,运用 SPSS 进行共词聚类分析,得出共词聚类冰挂图和树形图,如图 2 和图 3 所示。

结合图 2 与图 3,可以将高频关键词划分为 4 个主要的聚类群体,如表 4 所示。高频关键词聚类群体划分的依据是关键词的共词频次,首先将关系密切的关键词进行聚类,最后形成大的聚类群体。

由表 4 可以看出,群体 1 涉及的内容较多,既有虚拟水、气候变化,又有水资源开发利用、水经济和生态环境等,经过文献分析,可以把群体 1 的主要研

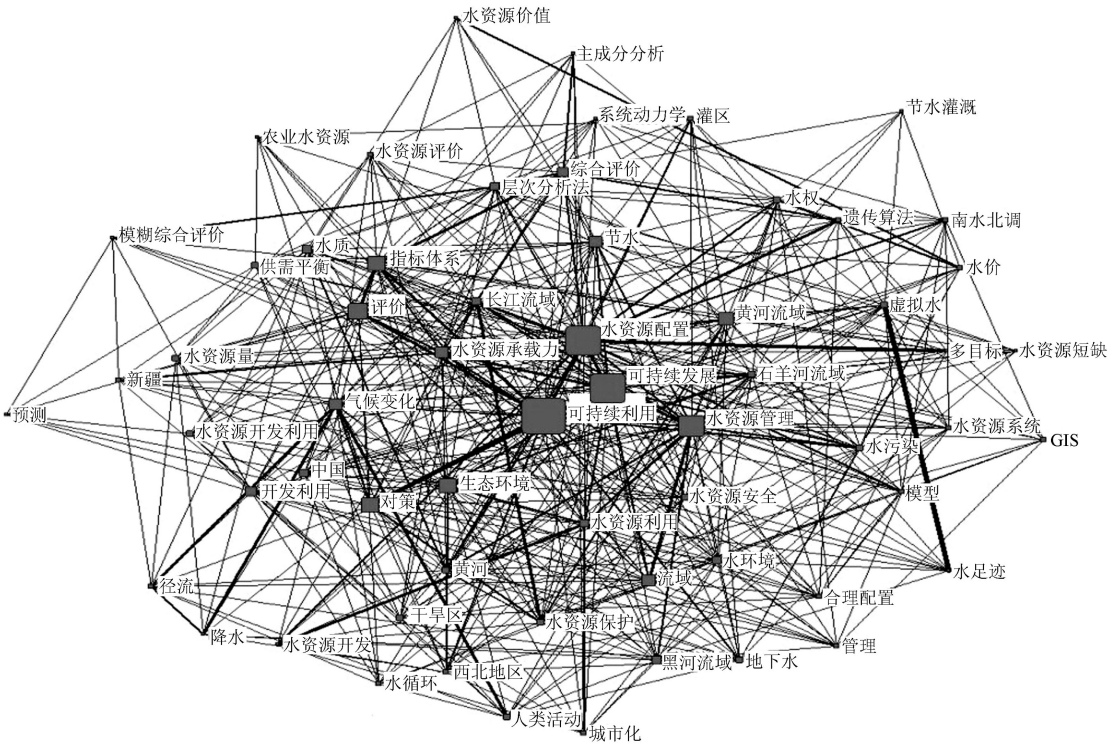


图 1 高频关键词共现网络可视图

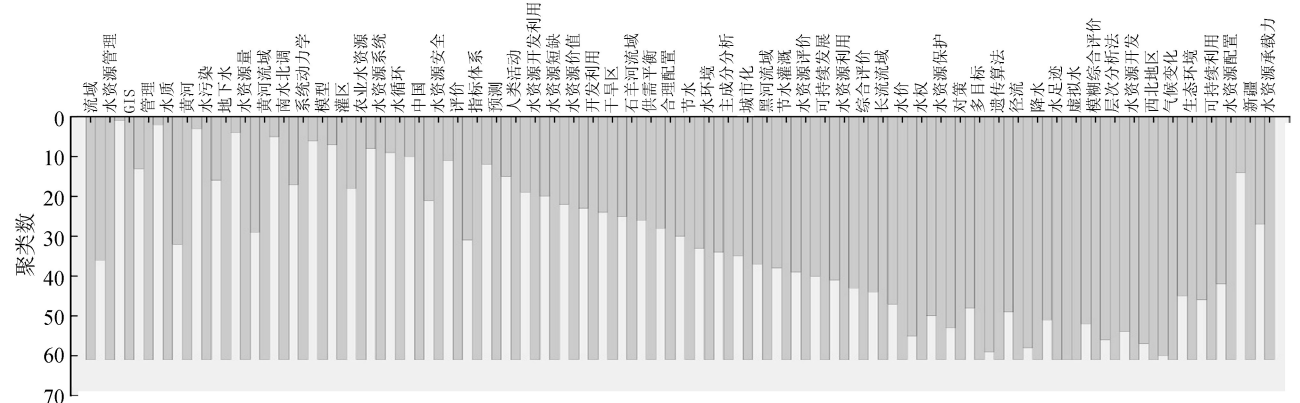


图2 共词聚类分析冰挂图

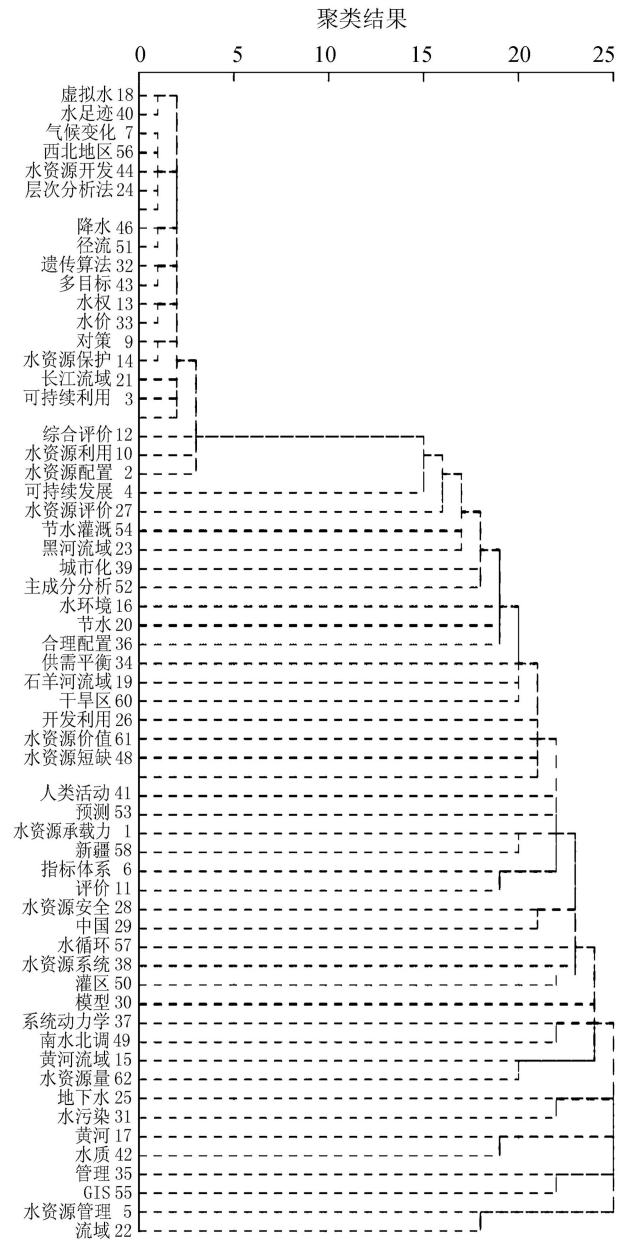


图3 层次聚类分析树状图

究方向归纳为水资源的可持续利用。群体2涉及的内容包含水资源评价、节水灌溉、城市化、水环境、节水等,主要研究水资源的可持续利用支撑经济社会

表4 高频关键词聚类群体划分

群体	关键词
1	虚拟水、水足迹、气候变化、西北地区、水资源开发、层次分析法、降水、径流、遗传算法、多目标、水权、水价、对策、水资源保护、长江流域、可持续利用、生态环境、综合评价、水资源利用
2	可持续发展、水资源评价、节水灌溉、黑河流域、城市化、主成分分析、水环境、节水
3	供需平衡、石羊河流域、干旱区、开发利用、水资源价值、水资源短缺、人类活动、预测、水资源承载力、新疆、指标体系、评价、水资源安全、中国、水循环、水资源系统
4	灌区、模型、系统动力学、南水北调、黄河流域、水资源量、地下水、水污染、黄河、水质、管理、GIS、水资源管理、流域

的可持续发展。群体3涉及水资源供需平衡、水资源价值、水资源短缺、人类活动、水资源承载力等内容,从水资源系统的视角进行水资源研究。群体4涉及灌区、南水北调、黄河流域、水资源量等内容,以流域水资源管理为基础进行研究。

4 多维尺度分析

为了更好的显示关键词的聚类群体,采用 SPSS 中的多维尺度分析功能来构建多维尺度可视化图谱,从多维尺度分析结果可以直观、形象地看出关键词的聚类群体,结果如图4所示。

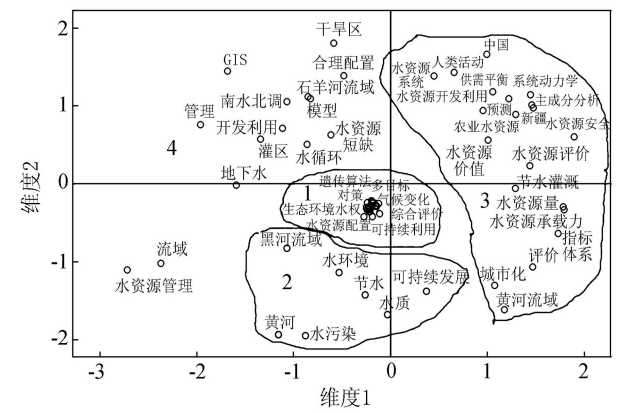


图4 多维尺度分析结果

图4较好地显示了关键词聚类群体结果,与图2、图3的共词聚类结果基本一致,根据每个群体在

坐标中的位置(即欧几里得距离),可以看出各类群体间的紧密程度,群体1、群体2、群体4联系较为紧密,由图2、图3的分析可知,水资源可持续利用、社会可持续发展和水资源管理是紧密相关的,三者是水资源研究的主要目的与方向;群体3则与群体1、群体2存在联系,水资源系统与水资源可持续利用、社会可持续发展互相影响,许多文章是从水资源系统的视角进行水资源研究的。

5 结 论

- a. 水资源学科的主要研究热点集中于水资源配置、水资源承载力、可持续利用、可持续发展、水资源管理,围绕着这几个研究热点形成了水资源学科新的研究视角。
- b. 基于共词网络分析可知,水资源配置、可持续利用、可持续发展、水资源承载力、水资源管理处于核心位置,是最重要的研究热点方向,在其研究基础上形成了水资源保护、水资源短缺、水资源利用、水污染、气候变化、生态环境等研究热点结构,它们之间的关联性较为紧密。
- c. 基于共词聚类分析与多维尺度分析可知,水资源学科在研究过程中形成了水资源可持续利用、社会可持续、水资源系统、水资源管理4大类群体,四者之间显现出一定的相关性,相互存在联系。

参考文献:

[1] 任智军,朱东华,谢菲. 科技文本的可视化分析研究[J]. 北京理工大学学报:社会科学版,2007,9(1):13-17. (REN Zhijun, ZHU Donghua, XIE Fei. The visualization analysis of scientific text[J]. Journal of Beijing Institute of

Technology: Social Sciences Edition, 2007, 9(1):13-17. (in Chinese))

[2] FRY J, TALJA F. The intellectual and social organization of academic fields and the shaping of digital resources[J]. Journal of Information Science, 2007, 33(2):115.

[3] 尹相旭,张更平,李晓菲. 基于关键词统计的情报学研究现状分析[J]. 情报杂志, 2009, 28(11):38-41. (YIN Xiangxu, ZHANG Gengping, LI Xiaofei. A research situation analysis of information sciences based on key words statistics[J]. Journal of Information, 2009, 28(11):38-41. (in Chinese))

[4] 舒琅. 基于社会网络分析的项目管理学科热点和主流知识群的研究[D]. 杭州:浙江大学, 2010.

[5] 李长玲,支岭,纪雪梅,等. 我国情报学研究进展:基于期刊论文关键词的统计分析[J]. 图书情报工作, 2010, 54(24):31-36. (LI Changling, ZHI Ling, JI Xuemei, et al. Trend analysis on information science based on keyword statistics[J]. Library and Information Service, 2010, 54(24):31-36. (in Chinese))

[6] 曹玲,杨静,夏严. 国内竞争情报领域研究论文的共词聚类分析[J]. 情报科学, 2010, 28(6):923-925. (CAO Ling, YANG Jing, XIA Yan. Co-word clustered analysis of competitive intelligence dissertations in China [J]. Information Science, 2010, 28(6):923-925. (in Chinese))

[7] 许振亮,刘则渊,侯海燕,等. 中国技术创新理论研究前沿共词知识图谱分析[J]. 图书情报工作, 2009, 53(6):46-49. (XU Zhenliang, LIU Zeyuan, HOU Haiyan, et al. Knowledge mapping of the theory fronts of China technology innovation: based on the view of author co-cited[J]. Library and Information Service, 2009, 53(6):46-49. (in Chinese))

(收稿日期:2013-06-27 编辑:周红梅)



· 简 介 ·

《2014 年联合国世界水资源开发报告》简介

2014 年 3 月,联合国教科文组织出版了《2014 年联合国世界水资源开发报告》(The United Nations World Water Development Report 2014),由联合国粮食及农业组织、联合国环境规划署、联合国人类住区规划署、联合国工业发展组织和世界银行分别负责论证和起草该报告的不同分支主题,同时该报告还受益于联合国水机制众多正式成员及合作伙伴的相关工作成果。《2014 年联合国世界水资源开发报告》分为两卷:第 1 卷是“水资源和能源”,宏观探讨目前及今后一段时期人类在水资源开发、利用和管理中存在的一些问题及其与能源之间的复杂关系;第 2 卷是“直面挑战”,提供来自全球的 13 个案例,介绍各地在应对水资源和能源问题中的一些典型做法。编者致力于提供一个以基本事实为依据、以均衡中立为价值导向的综合报告,期望其能够代表当前人类对水资源和能源认识的最新水平,呼吁加强水资源和能源的综合管理,弱化经济增长对水资源的依赖,改善经济发展的可持续性。这是联合国推出的第 15 份《联合国世界水资源开发报告》,也是首次以年度报告的形式编撰,同时《联合国世界水资源开发报告》的年度主题将与世界水日的主题相一致,以相应扩大宣传活动的影响力。

(楚行军供稿)