

基于 AHP 与 TOPSIS 方法的中文搜索引擎评价体系

方曦¹, 李娜², 葛月凤¹

1. 上海应用技术学院, 上海 201418

2. 上海交通大学工业工程系, 上海 200240

摘要 随着搜索引擎在信息化生活中的指导作用越来越强, 建立完整的评价体系已经成为搜索引擎优化的基础。本文从搜索引擎评价体系的现状入手, 结合搜索引擎特点, 采用定性和定量相结合的分析方法, 排除相关性高的指标, 确定最终索引数据库性能、检索效果、用户便利度三大类 13 个评价指标。在此基础上, 运用基于比例标度的层次分析法 (Analytic Hierarchy Process, AHP) 构建层级结构, 借鉴专家意见确定评价指标的权重; 结合 TOPSIS 法 (Technique for Order Preference by Similarity to an Ideal Solution) 进行评价指标数值的规范化处理并排序计算, 构建了较为完整的评价体系。选择四大热门中文搜索引擎作为评价对象, 进行实证分析, 结论表明, 本文所述评价指标体系和方法思路可行, 有较强可操作性和指导作用。

关键词 搜索引擎评价体系; AHP; TOPSIS

中图分类号 F272

文献标识码 A

doi 10.3981/j.issn.1000-7857.2012.14.006

Evaluation System for Chinese Search Engines Based on the AHP-TOPSIS

FANG Xi¹, LI Na², GE Yuefeng¹

1. Shanghai Institute of Technology, Shanghai 201418, China

2. Department of Industrial Engineering of Shanghai Jiaotong University, Shanghai 200240, China

Abstract As search engines are playing a more and more important role in information activities, the related evaluation methods are an urgent issue. According to the features of search engines, based on a wide review of related literature both at home and abroad, the paper proposes to use the qualitative and quantitative analysis method for filtering the established indexes, and 13 evaluation criteria are established in 3 categories which are the index database, the retrieval effects, and the user friendliness. The weights of these evaluation criteria are calculated by Analytic Hierarchy Process based on the index value of TOPSIS. A complete evaluation system is built and then several Chinese search engines are selected as the objects to use the Technique for Order Preference by Similarity to obtain an Ideal Solution for the final evaluation results. The paper discusses what should be done to improve the search engines in the future.

Keywords evaluation criteria system for search engines; AHP; TOPSIS

0 引言

随着互联网的发展, 作为重要的网络信息检索工具之一的搜索引擎发展非常迅猛, 对其评价研究的方法一直是国内外关注重点^[1]。虽然在以前国外就有了搜索引擎相关的评价标准, 但因为国情不同, 中文搜索引擎从使用习惯、公众依赖

度、被约束情况都跟国外搜索引擎不同, 所以对其评价问题一直吸引众多国内学者的关注。

1 中文搜索引擎评价指标体系的构建

国外很多学者提出过各种角度的、完整的评价指标体

收稿日期: 2011-07-05; 修回日期: 2012-03-20

基金项目: 教育部人文社科基金项目 (12YJC630044); 中国国家博士点基金项目 (20100073120080)

作者简介: 方曦, 讲师, 研究方向为企业管理, 电子邮箱: fangxi@sit.edu.cn

系,但是基于语言的不同,搜索引擎的开发方式和功能都不同,对搜索引擎的评价也均有各自的特点,中文搜索引擎评价指标^[2]不能完全照搬国外的指标。虽然国内众多学者在这方面做了很多研究,但由于其发展时间短、速度快、标准不同

等特点,目前国内还没有权威的评价搜索引擎标准的指标。

1.1 指标库的建立

笔者收集了国内外文献中出现频率较高的指标,整理了30个细节化指标,见表1。

表1 搜索引擎评价指标库

Table 1 Evaluation index system of search engines

细化指标				
标引数量	标引范围	网页覆盖率	标引深度	标引的文件种类
更新频率	链接广度	查全率	查准率	死链接率
检索时间	相关性排序	内容显示	过滤功能	搜索结果的满意度
布尔检索	词组检索	字段检索	模糊检索	多语种检索
自然语言检索	多媒体检索	概念检索	目录式浏览检索	交互功能
界面设计	搜索帮助	个性服务	安全性	特色功能

为了构建科学合理、贡献度高的评价指标体系,笔者参阅文献后对指标库以及指标含义修改如:(1)增加搜索请求量指标,用来衡量搜索引擎受欢迎程度;(2)将布尔检索、词组检索、字段检索归类入基本检索功能实现度;模糊检索、多语种检索、自然语言检索、多媒体检索、概念检索、目录式浏览检索归类入高级检索功能实现度;(3)将内容显示指标改为内容显示完整度指标,以免概念混淆。

建立最终指标库,共有指标13个。

结合中文搜索引擎用户习惯,并将定义重复的指标予以删除,整理为如下13个典型细节化指标建立最终指标库。

(1)索引数据库性能,包括:标引数量、标引广度、更新频率、标引深度;(2)检索效果,包括:查全率、查准率、死链接率、检索时间、相关性显示、过滤功能;(3)用户便利度,包括:搜索请求量、交互功能、特色功能。

1.2 指标相关性分析

为了得到这13个指标的标准化数据,进而用AHP-TOPSIS算法对中文搜索引擎进行评价,需要对它们进行定量分析,这里选取百度、谷歌(香港)、必应(bing.com)、搜搜这4大具有代表性中文搜索引擎数据对指标进行定量分析。

(1) 定性指标分析

按照相关性分析的要求,需要将定性指标量化。故针对标引广度、标引深度、链接广度、相关性排序、结果显示完整度、过滤功能、基本检索功能实现、高级检索功能实现、交互程度、特色功能制定量化标准:采用返回搜索结果第1页前20个相关条目,针对指标进行数量判断;基本搜索和高级搜索功能的实现,交互程度和特色功能的量化则用被评价搜索引擎相关功能数量除以所有搜索引擎相关功能,计算比率,得出评分,见表2。

表2 定性指标得分值及量化标准

Table 2 Qualitative indicator's score value and quantitative criteria

评价指标	百度	谷歌(香港)	必应	搜搜	量化标准
标引广度	8	12	4	8	前20条目中非纯网页的数量
标引深度	19	19	12	17	前20条目中的深度
链接广度	18	16	12	16	前20条目中范围
相关性排序	12	15	15	14	前20条目中结果相关度排序正确的
结果显示完整度	17	16	13	13	前20条目中结果显示完整的
过滤功能	6	2	3	4	前20条目中不良网站和有害信息结果的数量
基本检索功能实现	0.909	0.818	0.636	0.727	11项基本搜索功能包含比率/%
高级检索功能实现	0.889	0.889	0.778	0.667	9项高级搜索功能的包含比率/%
交互程度	0.75	0.667	0.583	0.667	12项交互功能的实现比率/%
特色功能	0.791	0.837	0.279	0.349	43项特色功能中的实现比率/%

(2) 定量指标分析

定量指标为标引数量、更新频率、查全率、查准率、死链接率、检索时间、搜索请求量。选择合适的评价角度,根据搜

引擎官方公布数值与文献查阅数值,列表如表3。

(3) 数据标准化处理

将表2和表3数据进行标准化处理,所得数据见表4。



表 3 定量指标数值
Table 3 Indicators of quantitative values

评价指标	百度	谷歌(香港)	必应	搜搜	数据单位
标引数量	1	0.82	0.78	0.92	各搜索引擎数据库总量之比
更新频率	7	9	3	4	平均大规模更新次数/(次·(15d) ⁻¹)
查全率	0.37	0.45	0.23	0.28	搜索结果占总相关条目比率/%
查准率	0.37	32	38	38	准确条目占搜索结果比率/%
死链接率	0.12	0.19	0.18	0.24	搜索结果无法显示比率/%
检索时间	0.001	0.1	0.05	0.009	搜索返回时间平均值/s
搜索请求量	76	18.9	0.4	3.2	搜索请求次数占总次数比率/%

表 4 指标数据标准化
Table 4 Index data standardization

综合指标	评价指标	百度	谷歌(香港)	必应	搜搜
索引数据	标引数量	1.000	0.820	0.780	0.920
	标引广度	0.667	1.000	0.333	0.667
	库性能	1.000	1.000	0.632	0.895
	更新频率	0.778	1.000	0.333	0.444
检索效果	查全率	0.822	1.000	0.511	0.622
	查准率	0.925	0.800	0.950	0.850
	死链接率	0.444	0.704	0.667	0.889
	检索时间	0.010	1.000	0.500	0.090
	相关性排序	0.800	1.000	1.000	0.933
	过滤功能	1.000	0.333	0.500	0.667
用户 便利度	搜索请求量	1.000	0.249	0.005	0.042
	交互功能	1.000	0.889	0.778	0.889
	特色功能	0.944	1.000	0.333	0.417

2 AHP-TOPSIS 算法

2.1 基于 AHP-TOPSIS 算法的综合评价方法原理

TOPSIS 法通过理想解和负理想解的确定,将实际样本中的优值和劣值引入评价模型中,从而使评价结果能够充分体现评价对象的集团特性,使评价结果与实际更加接近。但是 TOPSIS 法由于评价指标权重确定问题,在层次性多元分析的应用中受到限制。如果能寻找到适合层次性多元分析的权重赋值方法,TOPSIS 法可在层次性多元分析中得到广泛应用。

而层次分析法(AHP)正是多层次多因素权重确定的有效工具。层次分析法通过判断矩阵导出排序权值,使权值在定性的基础上得以量化,结果更加客观。但层次分析法对评价指标值的排序处理,过于简单化,难以体现指标特性和集团特征。

将 TOPSIS 法和层次分析法在层次性多元分析评价中进行综合应用,可以吸收 2 种方法的优点,互补不足,使评价方法的应用领域更加广泛,同时可提高评价方法的易操作性和评价结果的客观性、准确性^[3]。

2.2 AHP-TOPSIS 算法综合评价步骤

(1) 建立层次结构模型

在深入分析实际问题的基础上,将有关的各个因素按照不同属性自上而下地分解成若干层次,同一层的诸因素从属于上一层的因素或对上层因素有影响,同时又支配下一层的因素或受到下层因素的作用。最上层为目标层,通常只有 1 个因素,最下层通常为方案或对象层,中间可以有 1 个或几个层次,通常为准则或指标层。当准则过多时(譬如多于 9 个)应进一步分解出子准则层。

(2) 构造成对比较阵

从层次结构模型的第 2 层开始,对于从属于(或影响)上一层每个因素的同一层诸因素,用成对比较法和 1—9 比较尺度构成成对比较阵,直到最下层。

(3) 计算权向量并做一致性检验

对于每一个成对比较阵计算最大特征根及对应特征向量,利用一致性指标 $CI=\frac{\lambda_{\max}-n}{n-1}$ 、随机一致性指标 RI (见表 5)和一致性比率 $CR=CI/RI$ 做一致性检验。

表 5 随机一致性指标 RI 数值表
Table 5 RI numerical tables of random consistency index

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

若 $CR<0.1$ 时,认为判断矩阵具有满意的一致性,特征向量(归一化后)即为权向量,若不通过,需重新构成比较阵。如果满足要求,那么所获得的 w_i 即为准则层对目标层的相对权重。

(4) 各准则层对总指标权重计算

利用各层次的评价指标权重向量,计算出评价指标的总权重向量 w_i 并对总权重向量 w_j 进行一致性检验,形成对角权重矩阵 W 。

(5) 目标评估值矩阵确定

利用评价对象集 U 中的 n 个元素的评价指标值构造决

策矩阵 A ;若指标中即有效益型指标,又有成本指标、佳值指标,均化为效益指标。其中,为第 i 个评价对象的第 j 个评价指标的指标值。

$$V = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}$$

对矩阵 V 进行规范化处理,形成规范化决策矩阵 V 。

(6) 构造规范化矩阵,对指标数据规范化处理得特征矩阵 D ,利用 AHP 法确定的指标的权重向量 R ,得矩阵加权判断矩阵 $Z, Z=RD$ 。

(7) 根据加权判断矩阵获取评估目标的正负理想解。

正理想解:

$$f_j^* = \begin{cases} \max(f_{ij}), j \in J' \\ \min(f_{ij}), j \in J'' \end{cases} \quad j=1, 2, \dots, n$$

负理想解:

$$f_j' = \begin{cases} \min(f_{ij}), j \in J' \\ \max(f_{ij}), j \in J'' \end{cases} \quad j=1, 2, \dots, n$$

其中, J' 为效益型指标, J'' 为成本型指标。

(8) 计算各目标值与理想值之间的欧氏距离:

$$S_i^* = \sqrt{\sum_{j=1}^n (f_{ij} - f_j^*)^2} \quad j=1, 2, \dots, n$$

$$S_i' = \sqrt{\sum_{j=1}^n (f_{ij} - f_j')^2} \quad j=1, 2, \dots, n$$

(9) 计算各个目标的相对贴近度

$$C_i^* = S_i' / (S_i^* + S_i') \quad i=1, 2, \dots, m.$$

(10) 依照相对贴近度的大小对目标进行排序,形成评价结果。

3 实证研究

选取百度、谷歌(香港)、搜搜、必应这 4 个搜索引擎为例子,将指标体系综合利用 AHP 和 TOPSIS 方法进行应用。

3.1 建立层次结构图

根据前文得出的评价指标,建立 4 层层次分析模型,如图 1。最高层即目标层搜索引擎综合评价(G)。第 2、3 层为准则层,第 2 层准则层(H)包括:索引数据库性能(H_1)、检索效果(H_2)、用户便利度(H_3)。第 3 层准则层(P)包括 13 个准则:标引数量、标引广度、标引深度、更新频率、查全率、查准率、死链接率、检索时间、相关性显示、过滤功能;搜索请求量、交互功能、特色功能。最底层是方案层(M),选择百度、谷歌、搜搜以及近期热门的新近搜索必应。

3.2 确定指标权重

根据 1—9 相对重要性比例标度,邀请专家对指标进行两两比较,见表 6。

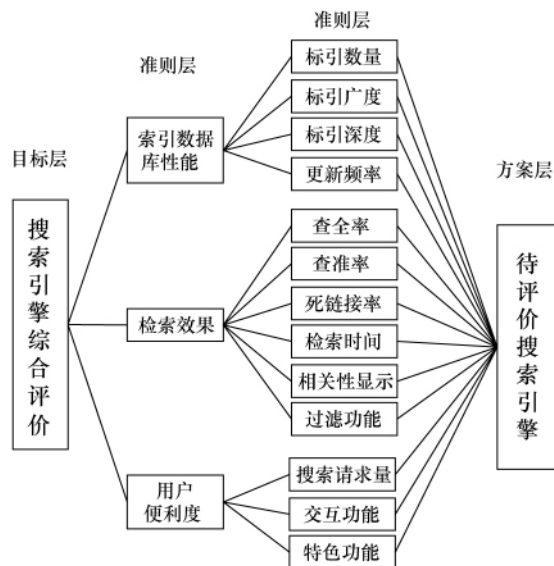


图 1 搜索引擎综合评价层次模型

Fig. 1 Hierarchical model of search engine evaluation

表 6 G 层下各 H 层指标相对重要性比较

Table 6 Indicators of the relative importance of the H layer under G layer

G	索引数据库性能	检索效果	用户负担
索引数据库性能	1	1/5	3
检索效果	5	1	7
用户便利度	1/3	1/7	1

根据表 6 数据,得到 $G-H$ 比较矩阵

$$G = \begin{bmatrix} 1 & 1/5 & 3 \\ 5 & 1 & 7 \\ 1/3 & 1/7 & 1 \end{bmatrix}$$

索引数据库性能下三级指标的比较矩阵

$$H_1 = \begin{bmatrix} 1 & 3 & 2 & 7 \\ 1/3 & 1 & 1/3 & 5 \\ 1/2 & 3 & 1 & 6 \\ 1/7 & 1/5 & 1/6 & 1 \end{bmatrix}$$

检索效果下三级指标的比较矩阵

$$H_2 = \begin{bmatrix} 1 & 1/2 & 3 & 6 & 3 & 5 \\ 2 & 1 & 5 & 7 & 5 & 3 \\ 1/3 & 1/5 & 1 & 3 & 1/3 & 3 \\ 1/6 & 1/7 & 1/3 & 1 & 1/3 & 1 \\ 1/3 & 1/5 & 3 & 3 & 1 & 3 \\ 1/5 & 1/3 & 1/3 & 1 & 1/3 & 1 \end{bmatrix}$$

用户负担下三级指标的比较矩阵

$$H_3 = \begin{bmatrix} 1 & 1/7 & 1/4 \\ 7 & 1 & 5 \\ 4 & 1/5 & 1 \end{bmatrix}$$

表 8 三级指标对二级指标的单排序和一致性检验表
Table 8 Third indicators to the overall goal of a single sort

表 7 二级指标对总目标的单排序和一致性检验表
Table 7 Second indicators to the overall goal of
a single sort and consistency checklist

权重 w_i	特征值 λ	CI	CR
0.1884			
0.7306	3.1155	0.05775	0.09957
0.0810			

上述一致性检验表中,所有 CR 均小于 0.1,矩阵不需要调整。所有三级指标权重对目标层的权重整理如表 9。

3.3 数据规范化处理

根据表 4 中各搜索引擎的数据用线性比例法做标准化

P	权重 w_i	特征值 λ	CI	CR
索引数据库性能	0.4680	4.1387	0.04623	0.0514
	0.1614			
	0.3219			
	0.0487			
检索效果	0.2699	6.2921	0.05842	0.0471
	0.4045			
	0.1115			
	0.0444			
	0.1115			
	0.0581			
用户便利度	0.0727	3.1237	0.04185	0.0721
	0.7223			
	0.2050			

表 9 三级指标对目标层权重

Table 9 Third indicators to the weights of target layer

指标	标引数量	标引广度	更新频率	标引深度	查全率	查准率	死链接率
权重	0.0882	0.0404	0.0706	0.0092	0.1772	0.1955	0.0815
指标	检索时间	相关性显示	过滤功能	搜索请求量	交互功能	特色功能	—
权重	0.0424	0.0815	0.0424	0.0259	0.0585	0.0366	

处理,并进行规划处理:

$$p_{ij} = \frac{y_{ij}}{\sqrt{\sum_{j=1}^m (y_{ij})^2}} \quad (1 \leq i \leq m, 1 \leq j \leq n)$$

得到矩阵

$$D = \begin{bmatrix} 0.57 & 0.47 & 0.56 & 0.56 & 0.54 & 0.52 & 0.32 & 0.01 & 0.51 & 0.74 & 0.97 & 0.56 & 0.64 \\ 0.46 & 0.71 & 0.72 & 0.56 & 0.66 & 0.45 & 0.51 & 0.89 & 0.53 & 0.25 & 0.24 & 0.50 & 0.68 \\ 0.52 & 0.47 & 0.50 & 0.41 & 0.48 & 0.48 & 0.64 & 0.08 & 0.47 & 0.50 & 0.04 & 0.50 & 0.28 \\ 0.44 & 0.24 & 0.35 & 0.34 & 0.54 & 0.54 & 0.48 & 0.45 & 0.49 & 0.37 & 0.00 & 0.44 & 0.23 \end{bmatrix}$$

由于打分标准一致,均为 1—10,故数据也不再需要进行规范化、归一化处理。将特征矩阵 D 与由表得到的权重矩阵 W 相乘,得到加权判断矩阵

$$W = \begin{bmatrix} 0.0882 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0.0366 \end{bmatrix}$$

$Z=DW=$	0.0499	0.0143	0.0341	0.0163	0.1063	0.1076	0.0261	0.0003	0.0414	0.0316	0.0542	0.0328	0.0234
	0.0409	0.0215	0.0438	0.0163	0.1294	0.0931	0.0413	0.0289	0.0433	0.0105	0.0135	0.0291	0.0248
	0.0459	0.0143	0.0195	0.0146	0.0805	0.0989	0.0522	0.0026	0.0384	0.0210	0.0023	0.0291	0.0103
	0.0389	0.0072	0.0146	0.0103	0.0661	0.1105	0.0392	0.0144	0.0398	0.0158	0.0003	0.0255	0.0083

根据加权判断矩阵获取评估目标的正理想解为:

$$f_j^*=(0.0499, 0.0215, 0.0438, 0.0163, 0.1294, 0.1105, 0.0261, 0.0003, 0.0433, 0.0316, 0.0542, 0.0328, 0.0248)$$

负理想解为:

$$f'_j = (0.0389, 0.0072, 0.0146, 0.0103, 0.0661, 0.0931, 0.0522, 0.0289, 0.0384, 0.0105, 0.0003, 0.0255, 0.0083)$$

计算各目标值与理想值之间的欧氏距离:

$$S_i^*=(0.02629,0.05957,0.08326,0.09534)$$

$$S_i'=(0.08887,0.05957,0.08326,0.09534)$$

各个目标的相对贴近度:

$$RK=(0.698687214,0.617094976,0.536170791,0.516798857)$$

可得出,排序为百度>谷歌(香港)>搜搜>必应。

从指标的权重上来看,最重要的指标仍然是查全率与查准率,其次是指标数量与更新频率,这也是中文搜索引擎仍然需要完善的地方。从各搜索的数值来说,综合排名前两位的是百度和谷歌,从数据上看,两者性能相差不大。虽然在朱庆华等^[4]、毛小燕^[5]、张军华等^[6]很多学者得出的结论中谷歌的排名在百度之前,但是谷歌服务器转移至香港之后搜索时间降低很多,中文索引库中的标引范围也变小了,所以得到这样的结果是在意料之中的。从排序也可以看出,搜搜作为腾讯一直推广的搜索引擎,依靠着掌握大规模用户群的 QQ 等工具,已经具备了一定的竞争力;而新近热门的必应,因为有微软的技术依托和特色搜索功能和服务,未来也会在中文搜索市场上分一杯羹的。

4 结论

本文结合中文搜索引擎的特点,确定了评价指标,并利用基于比例标度的层次分析法确定了各项指标的权重,运用接近理想点法进行排序,建立了完善科学的综合评价指标体

系,并选择具有代表性的例子进行实证研究,结果表明,本文所采用的指标和方法是行之有效的。

参考文献 (References)

- [1] 曾民族. 网络信息检索现状和性能评价[J]. 情报学报, 2007, 16(2): 90-99.
Zeng Minzu. *Journal of the China Society for Scientific and Technical Information*, 2007, 16(2): 90-99.
- [2] 宛凌, 杨秀丹, 杜晓静. 试析中文搜索引擎的评价标准[J]. 情报科学, 2000, 18(1): 28-31.
Wan Ling, Yang Xiudan, Du Xiaojing. *Information Science*, 2000, 18(1): 28-31.
- [3] 朱强, 阎子刚. 基于 AHP 与 TOPSIS 算法的供应商选择决策方法[J]. 交通标准化, 2008(02/03): 197-201.
Zhu Qiang, Yan Zigang. *Transport Standardization*, 2008 (02/03): 197-201.
- [4] 朱庆华, 杜佳. 搜索引擎评价指标体系的建立与应用 [J]. 情报学报, 2007, 26(5): 684-690.
Zhu Qinghua, Du Jia. *Journal of the China Society for Scientific and Technical Information*, 2007, 26(5): 684-690.
- [5] 毛小燕. 搜索引擎用户满意度研究的实证分析——以百度和 Google 中国为例[J]. 图书馆杂志, 2008, 27(3): 40-45.
Mao Xiaoyan. *Library Journal*, 2008, 27(3): 40-45.
- [6] 张军华, 韩全会. 中文五大综合搜索引擎主要性能测评 [J]. 情报科学, 2008, 26(9): 1352-1356.
Zhang Junhua, Han Quanhui. *Information Science*, 2008, 26 (9): 1352-1356.

(责任编辑 张军, 岳臣)

· 学术动态 ·

“2012 中国计算机大会”征文



由中国计算机学会主办的“2012 中国计算机大会”拟于 2012 年 10 月 18—20 日在大连市召开。

征文范围:(1) 近代流动显示技术的发展;(2) 经典流动显示技术及其应用;(3) 粒子图像测速技术及其应用;(4) 激光诱导荧光流动显示技术及其应用;(5) 光学干涉流动显示技术及其应用;6. 激光-超声流动显示技术及其应用;(7) 光学层析技术及其应用;(8) 热线、热膜测量技术及其应用;(9) 红外成像技术及其应用;(10) 压敏/温敏涂料压力测量技术及其应用;(11) 流动显示结果的处理技术及其应用;(12) 流动显示结果的理论分析技术及其应用;(13) 数值流动显示技术及其应用;(14) 流动显示技术的工程应用;(15) 其他与流动显示有关的内容。

论文截稿日期:2012 年 6 月 20 日。

通信地址:北京中关村科学院南路 6 号 (100190)。

联系电话:010-62601339。

电子邮箱:cncc@ccf.org.cn。

大会网站: <http://cncc.ccf.org.cn>。