DOI:10.13530/j.cnki.jlis.2020016

创意类开放数据竞赛作品评价指标体系构建与测定
——以数字人文项目为例*

赵宇翔　刘周颖　刘　炜　高胜寒　张　磊

摘　要　为激励大众参与公开数据的开发利用,各领域纷纷举办开放数据竞赛活动,并引发了社会各界的广泛关注。近年来数字人文项目的兴起一方面产生了大量的开放数据,另一方面由于其跨领域和学科交叉等特性,也给开放数据竞赛带来了新的契机,特别是针对数字人文项目开展相关的创意类开放数据竞赛,其影响力和社会价值不容小觑。随着数字人文领域创意类开放数据竞赛的推动和发展,对于作品的评价是确保开放数据竞赛公平公正的有效途径,也是项目后期孵化和推广的重要依据。目前国内外在这方面的实践还处于摸索起步阶段,理论研究方面更是鲜有关注,因此有必要对创意类开放数据竞赛中的竞选作品构建一套科学系统的评价指标体系。本文以"全评价"分析框架和 D&M 模型为基础,结合创意类开放数据竞赛的目的和数字人文领域的特色,初步确立了评价指标体系。在此基础上运用德尔菲法征询专家意见,完善指标体系并获取指标重要性得分,选用基于指数标度的层次分析法计算各指标权重;以上海图书馆举办的创意类开放数据竞赛为例,采用逼近理想解排序法和评价指标体系对部分竞选项目进行评估,从而完成指标体系的最终测定;最后对数字人文领域创意类开放数据竞赛的参赛项目评审和数字人文类互联网应用的开发提出若干建议。图 2。表 9。参考文献 55。

关键词　开放数据竞赛　数字人文　"全评价"分析框架　评价指标体系　德尔菲法　层次分析法

分类号　G203

Building the Evaluation Index for Creative Open Data Competitions: A Case Study on the Project of Digital Humanities

ZHAO Yuxiang, LIU Zhouying, LIU Wei, GAO Shenghan & ZHANG Lei

ABSTRACT

In order to encourage the public to participate in the development and utilization of open data, open data competitions have been launched in various fields and have attracted wide attention from the academia and industry. On the one hand, the rise of digital humanities projects have produced a large amount of open data in recent years. On the other hand, due to the cross-disciplinary and interdisciplinary characteristics of digital humanities, some new opportunities have been brought to the open data competitions. The

* 本文系国家社会科学基金重大项目"面向大数据的数字图书馆移动视觉搜索机制及应用研究"(编号:15ZDB126)和国家社会科学基金一般项目"面向数字人文研究的图书馆开放数据体系构建与服务模式设计"(编号:18BTQ027)的研究成果之一。(This article is an outcome of the major project "Exploring the Mobile Visual Search of Digital Library in Big Data Era: Mechanism and Application" (No. 15ZDB126) and the project "Library Open Data System Construction and Service Model Design for Digital Humanities Research" (No. 18BTQ027) supported by National Social Science Foundation of China.)

通信作者:赵宇翔,Email: yxzhao@vip.163.com, ORCID: 0000-0001-9281-3030 (Correspondence should be addressed to ZHAO Yuxiang, Email: yxzhao@vip.163.com, ORCID: 0000-0001-9281-3030)

influence and social value of creative open data competitions supporting digital humanities projects should not be underestimated. With the promotion and development of the creative open data competitions in the field of digital humanities, to evaluate the entries is not only an effective way to ensure the fairness and justice of open data competitions but also an important basis for the incubation and promotion of the projects in the later stage. At present, the practice of creative open data competition in the field of digital humanities at home and abroad is still in the initial stage, and theoretical investigation on this realm is rarely concerned. In this regard, there is a pressing-need to construct a scientific and systematic evaluation index system for the creative open data competition in the field of digital humanities.

To begin with, we review the relevant literature about creative open data competition, the evaluation of internet application in the field of digital humanities, and analytical framework and application of "all-round evaluation". Then, based on the "all-round evaluation" analytical framework and D&M model, this paper preliminarily establishes the evaluation index system by combining the purpose of the creative open data competition and the domain characteristics of digital humanities. Then, we use Delphi method to determine the final evaluation index system which consists of 6 first-level indexes, 18 second-level indexes and 39 third-level indexes. This paper also uses the analytic hierarchy process method (AHP) based on the index scale to calculate the weight of each index. The weight of six first-level indices, namely, the development and utilization of open data, the compliance degree of competition rules, information content, system performance, function design and interface design, gradually decreases. Next, taking the creative open data competition held by Shanghai Library as an example, we use the approximation ideal sorting method and the evaluation index system to evaluate some entries, thereby completing the final determination of the index system. Finally, we put forward some suggestions for the evaluation and development of the creative open data competition in the field of digital humanities.

In terms of the theoretical contribution of this work, the paper constructs a scientific and feasible evaluation index system, enriching our understanding of the important components of the creative open data competition in the field of digital humanities. In terms of practical implications, this study provides important guidance for the creative open data competition organizers and participants, especially for the designing and operation of the open data contest in the digital humanities domain. This study also has some limitations. Firstly, the evaluation system lacks the quantitative indices, which may have some biases in judging the comprehensive scale of the evaluation. Secondly, restricted by realistic conditions, this study cannot fully reflect the supervision mechanism and feedback mechanism that are advocated in the "all-round evaluation" analytical framework. In the future, we will further explore the quantitative data to measure the index and improve the evaluation system by more field methods (e.g., action research). 2 figs. 9 tabs. 55 refs.

KEY WORDS

Open data competition. Digital humanities. All-round evaluation framework. Evaluation index system. Delphi method. Analytic hierarchy process.

0 引言

在大数据时代,各行各业都逐渐意识到数据的重要性。为促进数据的合理利用与深度开发,有些领域的数据和接口被开放出来,公众在利用数据资源的同时也有可能参与数据资源的开发与管护。部分企业和机构则是通过积极开

展开放数据竞赛的形式来激励相关行业从业者和社会大众广泛参与,通过公开征集开放数据开发利用的优秀项目和案例,挖掘其中的数据价值并催生商业孵化和社会创新。

目前,国内外数据科学、教育科学、公共管理和数字人文等领域均陆续举办了若干开放数据竞赛,在学界和业界都产生了较大的影响力和示范效应。根据竞赛目标的不同,笔者将现有的开放数据竞赛归为如下两种类型。第一类是挑战类开放数据竞赛。该类竞赛以征集模型或算法等解决方案或产品原型为主要目的,同时也希望通过竞赛选拔优秀人才。挑战类开放数据竞赛对技术水平和专业知识有较高的要求,对于参赛者有较强的进入壁垒,任务设定和流程设计具有一定的对抗性。譬如,欧盟牵头的旨在促进教育领域开放数据使用和应用推广的 LinkedUp Challenge 系列竞赛①、数据科学平台 Kaggle 举办的基于众包模式的开放数据竞赛②、利用政府和企业大数据解决城市治理问题的上海开放数据创新应用大赛(SODA 大赛)③等都是典型的挑战类开放数据竞赛。第二类为创意类开放数据竞赛。该类竞赛的主要目的是通过竞赛的方式去征集创意,并将这些创意以产品或网站服务等开放数据形式去落地呈现。创意类开放数据竞赛强调作品的趣味性和新颖性,侧重于参赛者的设计能力和创新能力。典型的创意类开放数据竞赛如上海图书馆举办的针对特色馆藏资源和数字人文资源开发利用的竞赛活动④。近年来,数字人文作为跨学科、跨领域的交叉研究方向,不断涌现出一些优秀的项目及大量的开放数据集。针对数字人文项目的开放数据竞赛能够提升人文数据资源的有效利用,促进数字人文实践成果的发展应用,并推动面向人文研究的数据基础设施建设[1]。张磊等构建了包含内容、平台、服务三个组成部分的数字人文开放数据服务框架,并通过上海图书馆开放数据竞赛案例分析,论证了该框架的可行性[2]。

随着技术的发展和设计理念的变迁,现有的互联网应用评价指标也需要不断更新迭代以适应新业务、新模式和新环境。不同互联网应用的评价指标及其内涵各具特色,针对创意类开放数据竞赛而言,各项竞赛的目的、主题、形式、赛制均有所区别,相应地对竞选项目的评判标准也有所差异。然而,由于创意类开放数据竞赛尚处于起步阶段,目前在设置竞选项目评价标准时,大多“摸着石头过河”且缺乏理论支撑。具体到数字人文这一垂直领域,笔者认为有必要对创意类开放数据竞赛中的数字人文作品构建一套科学系统的评价指标体系,既能体现创意类开放数据竞赛的特色,又能兼顾数字人文领域的相关属性,从而助力数字人文领域互联网应用的开发与迭代,提升该领域创意类开放数据竞赛作品的质量,增强竞赛评审规则的科学性和规范性,促进数字人文领域中研究与实践的融合。

1 研究综述

1.1 创意类开放设计竞赛研究进展

创意类开放数据竞赛旨在征集基于开放数据所设计的创意作品,本质上是一种创意类开放设计竞赛,故笔者对创意类设计竞赛的相关文献进行了梳理。目前,该领域的大多数文献主要集中在对参与者的行为特征[3,4]及影响因素[5-10]的研究,而探索创意类设计作品评价机制的文献则较少。Cui 等研究表明,相比于排名

① AIMS. Veni, Vidi, Vici - three developer competitions by LinkedUpchallenge[EB/OL]. [2019-03-10]. http://aims.fao.org/community/blogs/veni-vidi-vici-three-developer-competitions-linkedup-challenge.

② Kaggle. Competitions[EB/OL]. [2019-03-18]. https://www.kaggle.com/competitions.

③ 上海开放数据创新应用大赛. SODA[EB/OL]. [2019-01-12]. http://shanghai.sodachallenges.com/.

④ 上海图书馆. 开放数据应用开发竞赛 2018[EB/OL]. [2019-01-12]. http://opendata.library.sh.cn/.

法,评分法更能有效地遴选出高质量的创意作品[11]。Özaygen 等从社会网络的角度来研究创意评价的影响因素,结果显示参与者的入度、出度和权威值会正向影响创意作品的评价[12]。创意类设计竞赛大多是对衣服设计、LED 灯设计等作品进行征集,并将作品以简单直观的方式展示给社会公众。对参赛作品的评价主要是根据公众的主观感受投票选出优秀的创意作品,评价维度较为单一。通过考察数字人文领域创意类开放数据竞赛的规则,笔者发现主办方鼓励或要求参赛者以互联网应用的形式(如 APP、微信小程序、WAP 网站)提交作品。因此,传统创意类设计竞赛作品的评价方式较为简单,并不适用于创意类开放数据竞赛作品的评价和遴选。优秀的创意类开放数据竞赛作品不仅需具备一定的创新性和新颖性,其功能、界面、系统等方面均需纳入综合考量范围。

目前,针对开放数据竞赛作品评价指标的相关研究较少。Brunswicker 等指出开放数据竞赛作品必须满足功能性、感知有用性和最低技术标准三个要求[13]。王丹丹以 Linked Up Project 为例构建了竞赛作品的评价框架,包括教育创新、可用性、性能、数据、合法性、用户六个维度共 20 个具体指标[14]。从研究对象的类型而言,这两项工作的研究对象均是挑战类开放数据竞赛,较为注重作品性能和技术水平。然而,挑战类和创意类开放数据竞赛的目的具有较大的差异。相比于挑战类开放数据竞赛作品,创意类作品的评价指标需凸显作品的创意性和可推广性,因此挑战类开放数据竞赛的作品评价体系也并不完全适用于创意类竞赛。

1.2 数字人文领域互联网应用评价的研究进展

由于数字人文领域创意类开放数据竞赛作品最终需以互联网应用的形式提交,故笔者认为,创意类开放数据竞赛作品的评价指标体系构建可以适当参考数字人文领域互联网应用评价指标体系的相关研究。在此基础上,结合创意类开放数据竞赛的目的和数字人文的领域特色,有针对性地构建创意类开放数据竞赛的评价指标体系。目前国外数字人文领域互联网应用评价相关文献主要是对博物馆网站、美术馆网站、文化馆场馆导览应用等进行可用性评价或构建相应的指标体系。Shneiderman 等通过实验研究发现,在博物馆网站中嵌入式超文本系统的导航效果优于传统索引[15]。Stoica 等通过对比不同情境下文字类和图像类游戏的完成情况,评估历史文化场馆移动应用的可用性及其对博物馆展览效果的影响[16]。Tesoriero 等通过设计移动端美术馆辅助展览应用,指出移动端应用的用户界面需有一致的设计风格和丰富的交互元素,并提供准确的定位,以便于用户的理解和使用[17]。Kabassi 通过考察近 30 年的文献,将学者对博物馆网站的评价角度概括为可用性和功能性两个维度,可用性维度包含一致性、可访问性、用户界面等 16 个指标,功能性维度包含多媒体功能、隐私规范标准等 11 个指标[18]。

我国数字人文领域互联网应用的评价角度较为分散,主要研究方向有移动端场馆导览类评价指标体系构建、馆藏展览类网站可用性评价等。杜华等针对场馆 APP 构建了以界面、技术、内容、功能、用户体验五个维度为准的三层评价指标体系,并指出色彩协调性、界面合理性、信息内容个性化、对用户的感知与定位等指标具有较高的权重[19]。张瑞等指出,导航条的清楚易用,信息内容的有效性、易检索性、一致性、针对不同用户的区别性,以及交互操作的及时反馈与符合用户习惯,是数字博物馆网站可用性评价的几个主要指标[20]。总的来说,系统的可访问性与稳定性、界面的交互性与设计风格、信息的内容与组织以及应用提供的特色服务是数字人文领域互联网应用评价研究的重要构成指标。然而,数字人文领域互联网应用的评价体系仍然存在一些问题。其一,评价体系缺少理论层面的规划与分析,存在评价维度不全面、指标含义宽泛等问题。其二,评价维度间

的独立性有待进一步改进,如“用户体验”实则由技术、功能、内容等多个维度共同影响,指标界定较为模糊。其三,上述工作的研究客体尽管为数字人文领域的网站或移动应用,但并未从真正意义上突出并强化其中的人文要素,也没有清晰的学术标签去指代数字人文的领域特色,从而缺乏一定的专指性。

1.3 “全评价”分析框架及应用

“全评价”分析框架于2010年由叶继元教授提出,该框架包含评价目的、评价客体、评价主体、评价方法、评价标准及指标、评价制度六大要素,以及形式、内容、效用三个维度[21]。“全评价”分析框架不仅能够对评价体系的建立及实施的全过程予以有效指导,并且提供了较为全面的视角对评价客体进行评价。其中,形式评价是指评价主体对评价客体的外部特征进行评价;内容评价是指评价主体对评价客体所含数据、信息和知识等的评价;效用评价是指经过实践或体验后对评价客体作用及价值的评价。叶继元教授指出形式评价、内容评价和效用评价既可以作为评价方法,又可以作为评价标准。

在提出“全评价”分析框架后,叶继元教授及其团队进一步验证了该框架的科学性和有效性,在评价期刊质量、学术/科技网站开发与建设等方面做出了示范性应用。譬如,基于该分析框架对国内图书馆学的期刊质量进行验证性分析,并对图书馆学优秀期刊进行竞争性评价[22]。范佳佳和叶继元运用“全评价”分析框架对21世纪以来国际学术界网站评价的相关研究进行梳理,指出未来的评价研究应注重评价主体的多元化、评价客体类型的选择、评价目的的制约性、评价方法的综合性、评价标准的合理性、评价制度的保障性[23]。随后,范佳佳等基于“全评价”体系框架从内容评价与效用评价两个角度构建科技网站的信息质量评价框架,并利用结构方程模型进行实证研究[24]。卢文辉和叶继元基于“全评价”分析框架对现有图书馆网站的评价现状进行分析,指出图书馆网站评价体系所存在的不足并提出了相应的对策建议[25]。陈铭和叶继元将“全评价”分析框架应用于构建开放存取仓储形式评价体系[26]。叶继元和陈铭在对中国科技论文在线优秀期刊进行评价时引入了“全评价”分析框架并从形式、内容、效用三个方面构建了相应的评价标准[27]。综上所述,“全评价”分析框架在评价体系确立及实施中具有较强的合理性和科学性。本文将参考“全评价”分析框架来构建与测定创意类开放数据竞赛中作品的评价指标体系。

2 评价指标的初步构建

本文的评价客体为数字人文领域创意类开放数据竞赛作品。构建评价指标体系的目的在于对创意类开放数据竞赛作品的质量从多个方面进行综合性评价,从而帮助竞赛举办方遴选出较为优秀的作品,并为用户提供优质的产品服务。通过解构评价目标及评价客体的关键信息,笔者认为评价标准及指标需从四个视角进行综合考察。一是“创意作品”视角,创意类竞赛作品通过互联网应用的具体形式来表现,即该评价指标体系归根结底是对一个创意类应用程序进行评价,因此评价指标中需包含互联网应用评价的若干基础内容以及作品的创意性。二是“数字人文”视角,数字人文视域下的创意作品与其余垂直领域作品之区别主要在于对人文资源和人文要素的精良组织和创意呈现,由于数字人文要素需要通过作品所提供的信息资源与其功能和界面设计体现,本研究拟将数字人文视角的评价内容作为创意作品视角指标的补充。三是“开放数据”视角,开放数据的利用是本文评价客体的重要内容,创意作品对开放数据的开发利用情况,在开放数据竞赛评审规则中占有相当大的比重。四是“竞赛”视角,遵守竞赛规则是保证竞赛公平公正的基础,是否符合竞赛规则是创意作品是否合格的重要评判依据。

2.1 “创意作品”与“数字人文”视角的评价内容

学界对互联网应用的评价,已形成了一些较为固定的角度,根据评价客体和评价目的的差异,研究者对指标的表述与其所属的一级指标的归类也有所不同。为确保本研究中指标设置的科学性,笔者拟参考 Delone 和 McLean 提出的信息系统成功模型(D&M 模型)进行基本评价维度的筛选和构建[28]。

D&M 模型自提出以来,已广泛应用于信息系统的评估检验和指导系统的设计与开发。初始 D&M 模型以系统质量和信息质量作为影响信息系统成功的原始变量[28],后期又对 D&M 模型进行了修正和完善[29],关系到信息系统成功的 6 类变量分别为信息质量、系统质量、服务质量、使用(意愿)、用户满意和净收益,并以信息质量(语义层面)、系统质量(技术层面)和服务质量(功能服务层面)作为评估信息系统运行质量的三个基本变量[30]。三个基本变量与系统所提供的信息与服务的质量息息相关,较为全面地涵盖了系统表现的多个层面,其间内涵彼此独立,但并不包含界面设计要素。然而,界面设计在互联网应用评价中是重要的考察维度。因此,本研究对创意作品评价的维度,将在 D&M 模型三大基本变量的基础上,补充界面这一变量,从信息、功能、界面、系统四个角度展开,进行“创意作品”视角的总体评价。

信息资源是信息系统提供信息服务的基础,在各类计算机和互联网应用评价指标体系中都处于重要地位,并分为定性指标和定量指标两大类。定性指标的具体评价内容通常有信息的全面性与丰富性[18]、信息内容范围(或称深度与广度)[19,31]、个性化[19,29]、准确性和权威性[32]、时效性(或称信息流通情况)[33]、新颖性/独特性/原创性[34]等;定量指标多与网络计量有关,包括链接数量(分为内部和外部链接)[34]、网络影响因子(分为内部和外部影响因子)[31]等,由于本文评价客体的形式不以网站网页为主,因此暂不考虑该类指标。参考张洋等对信息内容评价标准类型的划分[35],上述指标可以分为两类:一类反映信息范围,一类反映信息质量。另外,根据数字人文领域互联网应用的特征,除文本信息外,还利用图片、视频、音频等多媒体来提供人文资源,并加以展示和解说。因此,信息呈现的数字化方式也需要作为信息丰富度的评价内容[18]。数字图书馆的特色馆藏资源,是信息内容维度的常用考察内容[36,37]。相应地在竞赛情境下,如何选取具有一定特色和创意的作品主题,或在有限的主题下挖掘重构出新颖独特的信息内容,是团队创意的重要考察点。综合以上考量,笔者初步选取了一级指标“信息内容”下的具体评价指标,如表 1 所示。

表 1 一级指标“信息内容”下的指标初步构建结果

二级指标	三级指标	指标说明
信息丰富度	信息全面性	信息内容全面完整。
	信息个性化	呈现或推送个性化的信息内容。
	呈现方式丰富度	信息内容以多种形式(如文本、音频、视频等)进行全方面呈现。
信息质量	准确性	参赛作品提供的信息内容准确。
	权威性	参赛作品提供的信息内容有可靠的来源。
特色人文资源	主题独特性	作品的主题具有一定的特色和创意。
	内容独特性	作品提供的信息内容具有一定的特色和创意。

创意作品向用户提供的服务,需要通过开发者设计的功能来实现。Garibay 等将数字图书馆用户需求转译成功能需求,研究结果显示用户对数字图书馆的网站搜索以及资源覆盖度的

关注度较高,相应地,数字图书馆需提供基本的信息搜索功能以及数字资源可获取功能[38]。熊伟等在研究虚拟旅游网站功能设计评价指标时提出,可以从界面功能和产品服务功能两类指标进行评价,其中界面功能包括检索、导航、信息更新等基础功能,而产品服务功能则属于与网站类型相关的特色功能[39]。结合以上观点,本研究设置"功能设计"这个一级指标来考察创意作品的服务质量维度,并将其分解为"基础功能"与"特色功能"两个二级指标。其中,基础功能指普适于互联网应用的、满足用户对信息的浏览查询需求的基础功能,并以信息构建思想[40,41]指导指标说明的确立;特色功能指开发者在对人文资源如何组织和展现进行深入思考后,所提供的具有独特性和趣味性的可用性特色功能,具体如表2所示。

表2 一级指标"功能设计"下的指标初步构建结果

二级指标	三级指标	指标说明
基础功能	基础功能全面性	作品设计了能满足浏览、查询等用户基本信息需求的功能(如导航、组织、分类标识、检索等)。
	基础功能可用性	作品的基础功能流畅可用。
特色功能	功能独特性	作品设计了具有一定人文特色的功能。
	功能趣味性	特色功能新颖有趣,可吸引用户参与并使用。
	特色功能可用性	作品的特色功能流畅可用。

创意作品界面的美观程度决定了评分专家和用户对作品的第一印象。前人研究中通常包含图像/图标设计[42]、色彩[43]、布局[44]、交互[33]等可具象化的指标和强调主观感受的视觉吸引力[45]等元素,也有部分学者提出字体及其大小同样会影响用户体验[43,46]。上述指标中,图像/图标设计、色彩、字体三类与视觉美观程度关系最为紧密,布局影响到用户界面的整体结构和层次。由于用户与创意作品之间的交互是一个动态的过程,其与系统运行的稳定程度更为相关,故不作为"界面设计"角度下的评价指标。综合前人对产品界面评价的研究与创意作品的开发情况,本研究将一级指标"界面设计"的评价内容分解为色彩、字体、图案、界面布局四个二级指标并强调各个元素视觉上的协调统一[47]。另外,值得重视的是,由于开放数据竞赛是在数字人文领域下开展的,所以创意作品的图案设计需体现人文元素,从而展现其人文艺术价值,具体如表3所示。

表3 一级指标"界面设计"下的指标初步构建结果

二级指标	三级指标	指标说明
色彩设计	色彩协调性	作品界面色彩协调统一。
	色彩与主题匹配性	界面主体色彩与作品主题相匹配。
字体设计	字体	字体美观、大小适中(或可调整)。
	文字排版	文字的排版美观、便于浏览。
图案设计	人文元素	作品界面所包含的图案含有明显的人文元素。
	图案与主题匹配性	界面所包含的主要图案与作品主题相匹配。
界面布局	模块布局合理	各模块位置布局合理、主次模块清晰分明。
	模块分布一致	相同或相似模块在界面中的布局位置、视觉风格一致。

系统的性能是创意作品流畅稳定运行的基础,包含可连接性[33]、加载速度/响应时间[45]、隐私保护、身份鉴别与用户类型划分[20]等具体评价内容。在互联网应用的日常使用过程中,常出现闪退、空链接或死链接、服务器崩溃等现象[48],导致无法正常使用,因此笔者将无闪退或无崩溃现象、链接有效、资源稳定提供三个指标作为"系统稳定性"下的三个评价内容。由于部分创意作品需记录用户操作从而向不同(类别)用户提供个性化的服务,故用户的"身份鉴别"与"隐私保护"也需要纳入评价指标之中,属于"安全与隐私"的评价内容。系统对用户需求的响应时间反映了网络的传输速度和服务器的处理性能,会对用户体验造成极大的影响,因此需要将"响应时间"作为一项二级指标。此外,交互操作在各个评价指标中都具有一定的权重。在本研究中,由于交互操作的多样性和流畅性,更多地反映了后台系统的技术性能以及开发者的创意和对应用开发的认知程度,因此笔者将"交互体验"作为系统性能下的二级指标,以"交互丰富"和"交互友好"考察系统的易用程度,以"交互个性化"考察开发者的创意,具体指标如表 4 所示。

表 4 一级指标"系统性能"下的指标初步构建结果

二级指标	三级指标	指标说明
交互体验	交互丰富	交互方式丰富多样。
	交互友好	交互操作符合使用习惯,友好易用。
	交互个性化	作品提供个性化或创新性的交互方式。
系统稳定性	无闪退或无崩溃现象	作品打开和使用期间不会出现突然退出中断的异常现象。
	链接有效	作品中的链接均可指向正常页面。
	资源稳定提供	作品中所提供的资源是可稳定获取的。
安全与隐私	用户鉴别	系统可鉴别不同的用户。
	隐私保护	系统可对用户的信息加以保护,防止泄露。
响应时间	响应时间可接受	提交需求后,系统的响应时间在用户可接受的范围内。
	响应时间稳定	面对同类需求,系统响应时间的变化幅度不大。

2.2 "开放数据"视角的评价内容

竞赛主办方所提供的开放数据和多方共享的公开开放数据,是开放数据竞赛的数据资源基础。开放数据的利用情况,是各大竞赛对参赛作品进行评审的关键指标之一。笔者在考察现有开放数据竞赛规则后发现,主办方多以"开放数据利用程度"作为开放数据维度下的评审规则①②,但关于"利用程度"包含的细则和评价的标准,则表意模糊。因此,本研究将开放数据视角的评价指标细化,并给出具体的指标说明。

根据对主办方人员的访谈和竞赛中数据集的浏览,笔者有以下发现:开放数据竞赛的原始数据,通常由各个主办、协办单位提供,彼此独立存储于各单位的特色数据库中,包括细粒度的元数据和经过初步挖掘和组织的数据。以 2018 年上海图书馆开放数据竞赛为例,主办方提供了来自于上海图书馆、CADAL 项目管理中心、哈佛大学计量社会学系等机构的数百万条

① 上海开放数据创新应用大赛. 大赛规则[EB/OL]. [2019-02-26]. http://shanghai.sodachallenges.com/rules.html.

② 重庆开放数据创新应用大赛. 大赛规则[EB/OL]. [2019-02-26]. http://www.coda2017.com/rules.html.

数据,涵盖家谱、古籍、人物传记、诗词用韵等内容①,具有很高的开发价值。参赛者如何将来自不同数据集相互独立的信息资源进行关联融合,生产出具有一定深度的二次信息,是开放数据开发程度的考核要素。因此,本研究将“数据来源丰富度”和“数据开发深度”作为开放数据视角下的评价指标。另一方面,在提供的数据之外,互联网上存在着数量繁多的开放数据,能在原有数据集基础上起到重要的补充作用,主办方也鼓励参赛者进行外部数据的关联挖掘,并将其作为评审指标之一。开发者亦可通过超文本链接的形式[15],为用户提供这部分的信息资源,具体指标如表 5 所示。

表 5 一级指标“开放数据开发利用情况”下的指标初步构建结果

二级指标	三级指标	指标说明
数据利用率	内部数据利用率	作品对主办方提供的开放数据的利用率。
	外部数据利用率	作品对多方共享的公开开放数据的利用率。
数据来源丰富度	数据集丰富度	作品是否综合利用了多个来源渠道的数据。
	外链设置情况	作品通过一定的外链,为用户提供更为丰富的数据资源。
数据开发深度	数据的开发深度	数据是否进行了深度的挖掘和关联,以有所创新。

需要说明的是,“信息内容”维度下的指标,仅考察当前创意作品所提供的信息资源本身,而“开放数据利用情况”视角下的指标,将作品利用的数据置于主办方允许使用的全部数据集下,考察其利用率和开发程度。因此,两类指标不可合并为同一维度。

2.3 “竞赛”视角的评价内容

竞赛规则是赛事的要素之一,在创意作品的开发与提交流程满足竞赛规则的前提下,作品评审结果才具有公正性和公平性。笔者通过浏览各开放数据竞赛官网关于竞赛规则的表述,发现不同竞赛的规则各有特色,但大多包含以下几个共同点:①规定了参赛作品(在不同阶段)的提交形式;②要求参赛作品具有原创性和独特性;③明确了(各阶段)提交参赛作品的时间节点。作品原创性与独特性需通过作品的信息内容、功能设计、界面与交互等维度的指标加以体现,在 2.1 节已经进行构建。另外,部分竞赛在初始阶段,会要求参赛者提交作品创意或规划说明,但通过对主办方人员的访谈,笔者了解到部分创意作品至提交时,未能实现初期的规划目标。因此本部分的指标,将着重考察创意作品的提交形式、规划内容完成情况和提交时间。综上,“竞赛规则符合程度”视角下的评价指标初步构建结果如表 6 所示。

表 6 一级指标“竞赛规则符合程度”下的指标初步构建结果

二级指标	三级指标	指标说明
完成质量	作品形式	最终提交完整作品,且形式符合竞赛规则,为 APP、小程序、微站中的一种。
	目标达成度	最终提交的作品对初期规划目标的达成情况。
完成时间		在规定时间内提交作品。

① 开放数据应用开发竞赛 2018. 开放数据内容[EB/OL]. [2019-02-26]. http://opendata.library.sh.cn/.

本文从研究客体所涉及的“创意作品”“数字人文”“开放数据”“竞赛”四个视角初步构建了创意类开放数据竞赛作品评价指标体系，包含“信息内容”“功能设计”“界面设计”“系统性能”“开放数据开发利用情况”和“竞赛规则符合程度”6个一级指标。其中，“界面设计”和“竞赛规则符合程度”分别是对创意类开放数据竞赛作品的外部特征（包括色彩、字体、图案及界面）及作品完成的基本信息（包括作品形式、提交时间以及目标达成度）进行的形式评价。“信息内容”和“开放数据开发利用情况”是对作品的内容评价。“信息内容”聚焦于信息的丰富度、质量及特色人文资源内容及主题的独特性，是对作品所提供信息情况的内容评价。“开放数据开发利用情况”是对内外部数据利用率、数据来源丰富度及开发深度的内容评价。“功能设计”和“系统性能”是对作品的功能及其交互情况予以效用评价。虽然“功能设计”和“系统性能”指标在其可用性方面略有交叉，但是这两个指标的侧重点具有较大的差异。“功能设计”更侧重于创意作品功能的全面完整性，旨在提供可用的功能，从而满足用户的信息需求。“系统性能”侧重于用户与系统之间的交互体验及系统的稳定性和安全性，旨在为用户提供良好的使用体验。总的来说，初步构建的评价指标体系能够从形式、内容、效用全方位地对创意类开放数据竞赛作品进行综合评价，且指标间具有较强的独立性。值得注意的是，由于评价指标体系是评价体系的子体系，所以评价指标体系的设计需要突出“评价目的制约论”[21]。鉴于此，在综合性评价的基础上，下文将从评价主体、方法和标准选择三个方面进一步规范评价指标体系的构成。

3 评价指标体系的确定

本研究拟采用德尔菲法向专家征询关于指标选择（及指标说明）的意见和建议，并逐步收集专家对指标的评分，从而确定最终的评价指标并统计各指标得分情况。本研究采用问卷的形式收集专家意见，指标得分分为5个等级，5表示该指标非常重要，1表示非常不重要，在要求专家给指标重要性进行打分以外，还询问专家是否需要在各维度增添新的指标。由于构建评价体系的评价目标在于评选出较为优秀的作品，所以需要对作品质量进行打分排序。层次分析法在评价指标权重确定方面具有较强的科学性及可行性，并且在信息管理领域有较多应用[49,50]。因此，本文选择层次分析法来确定各指标的权重。

3.1 专家选择

本文所构建的评价体系的评价目标是选出优秀的创意类开放数据竞赛作品。据此，笔者邀请上海图书馆、复旦大学计算机科学学院、复旦大学人文社科大数据研究所、华东师范大学商学院、南京大学信息管理学院、南京理工大学信息管理系、上海大学图书情报系、上海师范大学人文学院等机构专家共14人，这些专家均对数字人文领域有所涉猎，进行过计算机和互联网应用的评价或熟悉应用开发的原则，并且对创意类开放数据竞赛规则具有一定的了解。

3.2 问卷结果统计与评价内容确定

根据每轮问卷回收情况，本研究问卷共发放三轮。

第一轮共回收14份问卷，除了“开放数据开发利用情况”的重要性认可度较高以外，其余维度的重要性在专家中均有一定分歧。在本轮中，一位专家补充两个三级指标，分别为“界面设计”维度的“图案美观性”（界面的图案设计是否有一定的美学欣赏价值）和“开放数据开发利用情况”维度的“数据精益化程度”（是否能细化、低成本、高效率地进行开放数据的开发）。两个指标的内容均未包含在原有指标体系中。同时，根据补充的新指标，

将二级指标“开放数据开发深度”的指标说明“数据是否进行了深度的挖掘和关联,以有所创新”更正为“数据是否进行了深度的挖掘和关联,并进行粒度的细化”,在第二轮问卷中进行反馈。

第二轮共回收 14 份问卷,没有新的指标补充入体系,但专家给出了若干完善建议。譬如,现有指标对技术因素的考察力度不够,可通过修改已有指标说明,突出对使用技术的考量。考虑到创意作品的开发技术可通过“功能设计”“系统性能”与“开放数据开发利用情况”三个维度的指标共同反映,参考专家意见,做出如下修改:①将“系统性能”维度下三级指标“交互友好”的指标说明改为“交互操作符合使用习惯,高效易用”;②将一级指标“开放数据开发利用情况”下的二级指标“数据开发深度”改为“数据开发技术”。

第三轮共回收 14 份问卷,反馈结果中没有新指标及其他意见建议的提出,且专家打分趋于统一,可认为专家已达成一致意见,无需再进行下一轮问卷发放。根据 Parente 等的观点[51],当专家组人数为 10—30 人时,可保证研究的信度和效度,因此问卷的最终回收情况符合科学性的要求。最终确立的评价层次结构如图 1 所示,包括 6 个一级指标,18 个二级指标,39 个三级指标,各指标的满分频度如表 7 所示(百分数保留到小数点后一位)。

3.3 权重计算与结果分析

3.3.1 权重计算结果

利用基于指数标度 a^n 的层次分析法,将各指标的满分频度视为指标的相对重要性,将指标间满分频度之差转换为两两比例标度[52],分别构建三个层级的判断矩阵。由于目前市面上通用的层次分析法软件大多只支持 1—9 标度矩阵的权重计算,故笔者采用和积法,利用 Excel 软件进行指标权重计算[53]和矩阵一致性检验[54],最终求得各指标权重及三级指标的合成权重如表 8 所示(保留到小数点后四位),各判断矩阵均通过一致性检验。

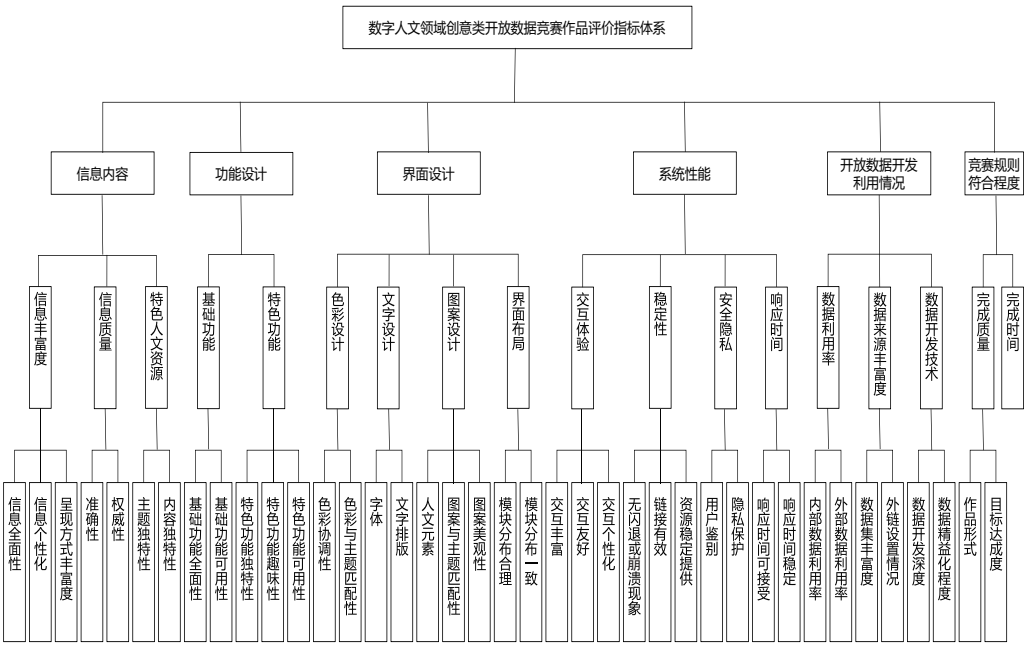

图 1 评价指标体系层次结构

表 7　各级指标最终满分频度

指标层级	指标名称	满分频度(%)	指标名称	满分频度(%)	指标名称	满分频度(%)
一级指标	信息内容	71.4	界面设计	14.3	开放数据开发利用情况	92.9
	功能设计	28.6	系统性能	71.4	竞赛规则符合程度	85.7
二级指标	信息丰富度	7.1	文字设计	14.3	响应时间	71.4
	信息质量	71.4	图案设计	14.3	数据利用率	92.9
	特色人文资源	28.6	界面布局	78.6	数据来源丰富度	21.4
	基础功能	71.4	交互体验	71.4	数据开发技术	78.6
	特色功能	64.3	稳定性	78.6	完成质量	78.6
	色彩设计	14.3	安全隐私	57.1	完成时间	57.1
三级指标	信息全面性	57.1	色彩与主题匹配性	42.9	资源稳定提供	85.7
	信息个性化	14.3	字体	21.4	用户鉴别	14.3
	呈现方式丰富度	50.0	文字排版	35.7	隐私保护	64.3
	准确性	78.6	人文元素	28.6	响应时间可接受	85.7
	权威性	50.0	图案与主题匹配性	14.3	响应时间稳定	78.6
	主题独特性	78.6	图案美观性	7.1	内部数据利用率	92.9
	内容独特性	57.1	模块布局合理	85.7	外部数据利用率	21.4
	基础功能全面性	64.3	模块分布一致	28.6	数据集丰富度	42.9
	基础功能可用性	85.7	交互丰富	21.4	外链设置情况	28.6
	特色功能独特性	35.7	交互友好	85.7	数据开发深度	71.4
	特色功能趣味性	21.4	交互个性化	7.1	数据精益化程度	28.6
	特色功能可用性	71.4	无闪退或崩溃现象	92.9	作品形式	71.4
	色彩协调性	14.3	链接有效	92.9	目标达成度	71.4

表 8　数字人文领域创意类开放数据竞赛作品评价指标体系及权重

一级指标	二级指标	三级指标	合成权重
C1 信息内容 0.177 1	C11 信息丰富度 0.149 0	C111 信息全面性 0.445 2	0.011 7
		C112 信息个性化 0.173 9	0.004 6
		C113 呈现方式丰富度 0.380 9	0.010 1
	C12 信息质量 0.612 0	C121 准确性 0.652 1	0.070 7
		C122 权威性 0.347 9	0.037 7
	C13 特色人文资源 0.239 0	C131 主题独特性 0.615 9	0.026 1
		C132 内容独特性 0.384 1	0.016 3
C2 功能设计 0.069 1	C21 基础功能 0.538 9	C211 基础功能全面性 0.384 6	0.014 3
		C212 基础功能可用性 0.615 4	0.022 9
	C22 特色功能 0.461 1	C221 特色功能独特性 0.255 0	0.008 1
		C222 特色功能趣味性 0.186 3	0.005 9
		C223 特色功能可用性 0.558 7	0.017 8

续表

一级指标	二级指标	三级指标	合成权重
C3 界面设计 0.050 5	C31 色彩设计 0.140 7	C311 色彩协调性 0.347 9	0.002 5
		C312 色彩与主题匹配性 0.652 1	0.004 6
	C32 文字设计 0.140 7	C321 字体 0.422 1	0.003 0
		C322 文字排版 0.578 9	0.004 1
	C33 图案设计 0.140 7	C331 人文元素 0.424 8	0.003 0
		C332 图案与主题匹配性 0.310 3	0.002 2
		C333 图案美观性 0.264 8	0.001 9
	C34 界面布局 0.577 8	C341 模块布局合理 0.778 1	0.022 7
		C342 模块分布一致 0.221 9	0.006 5
C4 系统性能 0.177 1	C41 交互体验 0.256 3	C411 交互丰富 0.171 3	0.007 8
		C412 交互友好 0.703 5	0.031 9
		C413 交互个性化 0.125 1	0.005 7
	C42 稳定性 0.300 2	C421 无闪退或崩溃现象 0.350 4	0.018 6
		C422 链接有效 0.350 4	0.018 6
		C423 资源稳定提供 0.299 2	0.015 9
	C43 安全隐私 0.187 2	C431 用户鉴别 0.250 0	0.008 3
		C432 隐私保护 0.750 0	0.024 9
	C44 响应时间 0.256 3	C441 响应时间可接受 0.538 9	0.024 5
		C442 响应时间稳定 0.461 1	0.020 9
C5 开放数据开发利用情况 0.270 0	C51 数据利用率 0.515 9	C511 内部数据利用率 0.827 9	0.115 3
		C512 外部数据利用率 0.172 1	0.024 0
	C52 数据来源丰富度 0.107 3	C521 数据集丰富度 0.577 9	0.016 7
		C522 外链设置情况 0.422 1	0.012 2
	C53 数据开发技术 0.376 8	C531 数据开发深度 0.719 1	0.073 2
		C52 数据精益化程度 0.280 9	0.028 6
C6 竞赛规则符合程度 0.256 2	C61 完成质量 0.615 9	C611 作品形式 0.500 0	0.078 9
		C612 目标达成度 0.500 0	0.078 9
	C62 完成时间 0.384 1	—	0.098 4

3.3.2 指标重要性分析

一级指标的权重从高到低排列,分别为“开放数据开发利用情况”(0.2700)、“竞赛规则符合程度”(0.2562)、“信息内容”(0.1771)、“系统性能”(0.1771)、“功能设计”(0.0691)、“界面设计”(0.0505)。可以看到,“开放数据开发利用情况”和“竞赛规则符合程度”两大维度,权重之和相加超过0.5,这两大维度为特定竞赛背景下的参赛创意作品与其他计算机互联网应用的根本性区别所在。重要性排在第二梯度的为“信息内容”和“系统性能”两大维度,信息内容为创意作品的数据资源基础,系统性能为创意作品运行的技术基础,二者共同保证向用户提供稳定的信息服务,也是D&M模型初始版本中直接影响系统使用和用户满意度的两大维度。接下来是“功能维度”,创意作品具体提供怎样的信息服务,是通过功能的设计规划来实现的,在创意作品评价中占有一定的地位,但重要性

不高。最后为“界面设计”维度,界面影响了用户对创意作品的第一印象,但在持续深入使用的过程中,并非重要的考量维度。三级指标合成权重大小升序排列如图 2 所示。

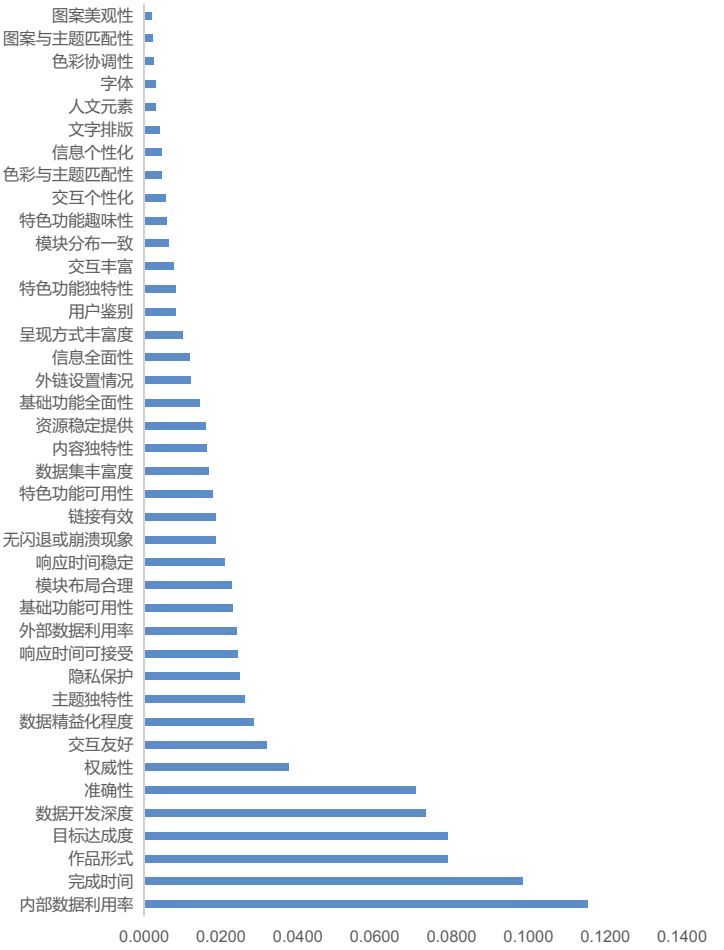

图 2　三级指标合成权重

“开放数据开发利用情况”与“竞赛规则符合程度”两大维度下的 9 个三级指标中,共 8 个排在前 50%,其中,“内部数据利用率”权重最高,表明专家在评判创意作品时,最看重参赛者是否能充分融合和利用竞赛主办方提供的开放数据;“竞赛规则符合程度”下的三个具体评价内容分别排在二、三、四位,表明专家注重竞赛公平性,只有符合竞赛规则、按时提交并完整性良好的作品,才能给出公正的评判。

“信息内容”维度下的 7 个三级指标中,“准确性”和“权威性”具有最高权重,这两个指标在全体指标中也名列前十,说明在“信息内容”维度下,内容准确权威最为重要,它影响到用户是否能获取可靠的知识,这也启示参赛者,尤其在利用外部数据时,要特别甄别其真实性。同时,“主题独特性”的权重也较高,参赛者在进行作品设计时需要体现作品的特色和创意。有趣的是,尽管个性化推荐已经成为泛在意义上互联网应用的重要评价指标,“信息个性化”指标的权重却很低,表明个性化的信息服务在竞赛作品中尚未受到广泛的重视。

“系统性能”维度下的 10 个三级指标中,“交互友好”占据最高权重,相对而言“交互丰富”和“交互个性化”权重较低,表明专家看重创意作品的交互操作是否符合用户习惯、高效流畅,但对交互方式的多样性和个性化并不注重。在移动端应用日益受市场青睐的背景下,交互设计可更多考虑移动设备特有的形式,力求操作便利流畅。“响应时间”相关指标也具有较高的权重,说明专家更注重系统能否快速响应需求和稳定连接服务器。因此参赛者在开发作品时,需着力提升系统运算能力和网络连接速度。

“功能设计”维度下,“特色功能”和“基础功能”的可用性权重较高,随之是“基础功能全面性”“特色功能独特性”和“特色功能趣味性”指标。并且,从整体上看“基础功能”评价指标的权重略高于“特色功能”评价指标。可见,作品需要在确保基础功能能够满足用户基本信息需求的前提下,再将富有创意和趣味的功能融入作品中。

“界面设计”维度下,相对色彩、文字和图案设计,“界面布局”这一指标的权重较高,反映出现阶段专家在创意作品中更为注重作品模块布局是否合理以及模块分布是否一致,对视觉元素美观性相对要求较低。合理一致的模块分布可提升作品的使用效率,从而保证用户良好的使用体验。色彩、字体、图案作为界面设计的一部分,依托于界面布局而存在。参赛者需根据界面布局并结合作品的信息内容和功能设计较为合理的字体、文字排版、色彩、图案等,并将人文元素体现在作品的界面设计中。

综上所述,本研究构建的指标体系反映了专家对创意类开放数据竞赛作品考察维度相对重要性的判断,可用于指导后续创意类开放数据竞赛作品评审规则的设置和权重分配。同时,也可以启示竞赛参与者,首先需要考虑如何全面深入地进行内外部开放数据的挖掘利用,开发出符合竞赛要求的创意作品;其次考虑作为一款互联网应用,作品如何确保信息内容准确性、增强交互操作友好程度、缩短响应时间等;接下来考虑在信息、功能、界面、系统四个维度的其他层面,尽可能完善创意类开放数据竞赛作品的性能和用户体验。

4 评价指标体系的测定:以上海图书馆开放数据竞赛为例

4.1 上海图书馆开放数据竞赛简介

上海图书馆在过去的十多年间逐步整合了大量富有特色的数字人文馆藏资源,也致力于开放数据的服务模式和技术的研究与推广。在此基础上,上海图书馆联合多家拥有特色人文数据的机构与企业,从 2016 年起,已连续举办三届开放数据竞赛,并吸引了众多来自企业和高校的数字人文研究者、移动应用开发者、历史文化爱好者参赛。上海图书馆以应用程序接口(API)的形式向参赛者提供数据,以避免数据集更新后参赛者需要重新处理数据或修改程序的问题,并鼓励融合主办方所提供数据集之外的其他开放数据。

2018 年上海图书馆开放数据竞赛,联合哈佛大学计量社会学系、广州搜韵文化发展有限公司等机构,补充中国历代人物传记资料库、古人诗例用韵参考等数据库,向参赛者提供开放数据。比赛共交出 30 份创意作品,最终评选出一等奖作品 1 份,二等奖作品 2 份,三等奖作品 3 份,优秀奖作品 4 份①,均以微站、APP、小程序等形式向用户提供服务。综合考虑研究实施的可行性、作品的可获取性、奖项覆盖率和三种应用形式覆盖率,本研究选取原印鉴章(微信小程序)、树人者(微站 http://edu.swapo.net/#)、风华诗韵(APP)、时迹(APP)、山大诗友(微信小程序)以及中文诗歌知识图谱(微站 www.freekg.cn)共 6 份创意作品作为待评价对象。

2018 年上海图书馆开放数据竞赛评审规则分为书面评审和答辩评审两个环节,书面评审包含创意、数据利用度、外部数据关联程度、技

① 开放数据应用开发竞赛 2018. 获奖作品[EB/OL]. [2019-03-04]. http://opendata.library.sh.cn/.

术先进性、工作量五个指标，答辩评审包括主题创新性、功能创新性、交互体验、开放数据利用度四个指标。笔者对评审规则进行分析并与相关专家讨论后发现一些问题。首先是两个阶段的评价内容有所重叠，如第二阶段的创新性指标，考察内容涵盖于第一阶段的创意指标中。其次是指标间的部分逻辑关系有待进一步梳理，如对交互体验的指标描述中包含对功能全面性的考察。三是评价指标仅包含一个层级，因此每项指标内涵较广，有专家反映在具体的评价实施过程中不便于细化评分。鉴于此，本研究构建了层次更为丰富、维度更为全面、粒度更为多元的评价指标体系，以期在一定程度上弥补原有评审规则的局限性，从而为评价的科学性和可操作性提供参考依据。

4.2 评价指标体系测定及作品得分排序

为保证评价结果的合理性和权威性，笔者从参与本研究的14位专家中邀请5人对待测评作品进行评分。在评分前，5位专家对上海图书馆开放数据竞赛所提供的数据集及竞赛规则，以及待测评作品的开发技术、目标完成度、提交时间等情况已有所了解，并且专家真实地体验过6份作品。评分采取问卷形式，使用1—5标度进行打分，5分表示打分者对该作品在该指标上的表现非常满意，1代表很不满意。5位专家均独立地对6份参赛作品分别进行评分，并不知晓其他专家的评分情况，确保了评价结果的有效性。在获取5位专家对6份作品在各指标上表现的得分后，笔者使用逼近理想解排序法[55]进行作品排序，最终得到每个作品与正理想解的距离（D_j^+）、与负理想解的距离（D_j^-）及与理想解的相对接近度（C_j^+），具体如表9所示（结果保留到小数点后四位）。作品的最终从高到低排名为风华诗韵、时迹、山大诗友、原印鉴章、中文诗歌知识图谱、树人者。

表9 作品与正负理想解的距离和相对接近度

作品	D_j^+	D_j^-	C_j^+
风华诗韵	0.006 4	0.028 7	0.818 2
时迹	0.011 0	0.025 1	0.695 6
山大诗友	0.012 3	0.024 9	0.669 3
原印鉴章	0.015 7	0.019 0	0.548 1
中文诗歌知识图谱	0.023 0	0.015 6	0.404 9
树人者	0.026 7	0.009 9	0.271 2

4.3 结果分析

排名结果显示，风华诗韵APP综合用户体验最好，获得最高排名。时迹APP综合利用了丰富的内部开放数据和来自中国气象局的外部开放数据，开发了具有创意的功能，但系统性能较弱，从而影响了用户体验。原印鉴章小程序在内容的特色和个性化方面表现突出，但响应时间很不稳定，初期规划功能未有效实现，这些因素共同影响了其最终得分和排名。树人者微站的主题为20世纪教育名家，具有一定的创意，但其数据利用率较低，功能设计过于简单，导致排名最低。

三种移动端应用形式中，APP、小程序、微站的排名逐次降低。该排名可部分说明，在移动端场景下，APP和小程序的生态对用户较为友好，例如，原印鉴章小程序可通过实时拍照识图的方式进行信息检索，体现了移动端互联网应用在移动视觉搜索方面的优势；而微站由PC端网站形式拓展而来，对移动硬件的适配及交互的设计易产生不符合新的使用场景的问题。另

外,6 份作品在“用户鉴别”“隐私保护”和“外链设置情况”三个指标下得分较低,其原因在于,在系统安全方面,这些创意作品不提供面向不同用户的个性化服务,无用户属性区分和隐私保护的硬性需求;在“数据集丰富度”方面,这些创意作品对信息内容拓展的需求不高,因此极少通过外链的设置向用户提供更多信息。

通过对作品的得分和排名结果分析可见,该评价指标体系能够从“信息内容”“功能设计”“界面设计”“系统性能”“开放数据开发利用情况”和“竞赛规则符合程度”六个维度较为全面地对数字人文领域下基于开放数据的创意作品进行评价。同时,笔者得出如下启示。

(1)对数字人文领域参赛者及互联网应用开发者来说,针对不同终端使用场景的特点进行特色功能的创意设计和交互方式的改进,能有效提升用户满意度。对于以 APP 或小程序等适应移动端的应用的形式所呈现的作品,可以通过充分挖掘开放数据资源的价值,设计更多适用于移动场景的特色创意功能,并以便捷的移动端交互方式优化用户体验。在微站的开发中,应当参考 APP 和小程序的交互设计和模块布局等,且在创意作品开发时使用多屏幕自适应的编码方式,使微站的使用更符合移动端的用户习惯。

(2)后续创意作品在开发时,若考虑向不同类型的用户提供更有针对性的信息服务,则必须在信息内容准确丰富的基础上,引导用户注册,做好用户属性区分、用户兴趣分析和隐私保护工作,而用户兴趣的分析和个性化信息服务的提供,也可反过来促进数据间的关联融合和开发深度。小程序形式的作品可借助微信平台,使用微信账号进行登录,在免除因注册流程复杂而导致用户流失这一问题之外,还可以更便捷地为用户提供个性化信息、进行通知推送及其他定制服务。

(3)为拓展创意作品提供的信息内容,设置外链并非唯一途径,但在页面上的文本、图片等元素的信息组织和标引之处,以外链(或其变化形式,如弹窗等)等形式进行信息补充,可提升用户的信息获取效率,加深其对当前浏览内容的接受和理解。

5 结语

5.1 结论与启示

本研究构建的数字人文领域创意类开放数据竞赛作品评价指标体系,在参考已有互联网应用评价研究成果和国内外主要开放数据竞赛情况的基础上,借鉴“全评价”分析框架,通过德尔菲法和层次分析法,确立“信息内容”“功能设计”“界面设计”“系统性能”“开放数据开发利用”“竞赛规则符合程度”六个维度为一级指标,共抽取了 40 个具体的评价指标,评价内容细致详尽,具有较强的科学性和可操作性。其中,“开放数据开发利用情况”和“竞赛规则符合程度”两个维度的权重,显著高于其他四个维度,而这两个维度也是本研究与其他互联网应用评价指标体系研究最大的不同之处。在所确立的评价指标体系的基础上,笔者利用逼近理想解排序法邀请专家对 2018 年上海图书馆开放数据竞赛的部分创意作品进行打分排序,证实了该体系在实际应用上的可行性。研究成果为数字人文领域下竞赛创意作品评审规则的确立、创意作品对开放数据的开发利用和创意作品本身的设计研发以及数字人文领域互联网产品的开发与升级方向,提供了一定的思路。

对数字人文领域创意类开放数据竞赛主办方来说,本研究构建的评价指标体系,以及对指标体系的测定方案,可直接用于后续创意作品评审和排名。主办方亦可以参考研究结果,对原有评审规则进行修正,对不同维度指标的权重高低进行更科学的规划。部分组织赛前培训的竞赛主办方,可参考研究成果对创意作品的开发提出指导。

对于数字人文领域创意类开放数据竞赛参赛者来说,本文给出了创意作品开发的方向。首先,作品需通过对各类开放数据的利用和关联,保证提供准确、丰富、富有人文特色和创意

的信息内容,开发技术应尽量精细高效;最终作品的表现形式和提交展示,应严格遵守竞赛规则。其次,开发者需保证创意作品具有作为应用程序的功能可用性。对于提供个性化功能服务的创意产品,还应注重用户信息安全的保护。再次,在满足用户基础信息需求的基础上,再将富有特色和创意的功能融入其中。最后,作品界面需在保证以上维度表现质量的情况下,提升视觉吸引力,营造人文氛围。对于具体的移动端作品形式,开发时也有不同的侧重点。例如以微信小程序形式提交创意作品,可借助微信平台的便利直接让用户使用微信账号登录并提供个性化的信息与功能;以微站形式提交的创意作品则需要重视创意作品的模块布局、交互体验设计从而更加适配移动端的使用场景。

此外,数字人文领域互联网应用的开发者也可参考该评价指标体系,有针对性地对创意产品的信息内容、系统性能、功能设计和界面设计等层面的规划和升级做出优先级的划分,以此更高效地开发和完善产品,并开展市场孵化和应用推广。

5.2 局限与未来展望

本研究中尚存在一些不足之处。其一,评价体系中缺乏有效的定量指标。在本研究所设立的指标中,诸如"内部数据利用率""目标完成度"等,更准确的测评方法是将其作为定量指标,通过具体比率数值反映作品表现优劣,但受限于研究条件,最终也只能依靠专家的主观判断。另外,由于本研究中的作品形式并不局限于APP,且截至评审阶段部分作品尚未公开上线,故针对移动端互联网应用的评价指标中的评分、下载量、评论量等基于应用商城的定量指标并不适用。其二,尽管研究团队尝试从不同学科和领域遴选了相关专家进行作品的定性评价,并在这一选择过程中反复与国家文化和旅游部公共服务研究上海图书馆基地的专家进行讨论,然而囿于时间和资金等约束条件,无法全面体现"全评价"分析框架中倡导的评价过程监督机制和评价结果反馈制度,这一空缺将由后续研究加以弥补和完善。

在未来的研究中,笔者将结合开放数据竞赛的特点,进一步挖掘定量指标。在使用该指标体系进行作品评审时,可通过记录最终完成功能数量、开发利用的内部数据量、运用的数据开发技术的数量等,对指标给出精确的分值,提升评分的客观性。在未来的数字人文领域开放数据竞赛中,可以应用该评价指标体系来对参赛作品进行考量,并通过用户的使用情况及满意度情况对该评价体系予以进一步验证和补充。

致谢

本研究在开展的过程中受到国家文化和旅游部公共服务研究上海图书馆基地各位专家的支持和帮助,在此一并向他们表达我们最诚挚的谢意!

参考文献

[1] 刘炜,谢蓉,张磊,等. 面向人文研究的国家数据基础设施建设[J]. 中国图书馆学报,2016,42(5):29-39. (Liu Wei,Xie Rong, Zhang Lei, et al. Towards a national data infrastructure for digital humanities[J]. Journal of Library Science in China, 2016,42(5):29-39.)

[2] 张磊, 夏翠娟. 面向数字人文的图书馆开放数据服务研究——以上海图书馆开放数据应用开发竞赛为例[J]. 图书馆杂志, 2018, 37(3): 33-48. (Zhang Lei, Xia Cuijuan. A study on open data service of library for digital humanities:taking Shanghai Library's open data contest as an example[J]. Library Journal, 2018, 37(3): 33-48.)

[3] Hutter K ,Hautz J ,Füller J, et al. Communitition: the tension between competition and collaboration in community-based design contests[J]. Creativity & Innovation Management, 2011, 20(1):3-21.

[4] Hoffmann A B, Neyer A K, Rass M, et al. Community-based innovation contests: where competition meets cooperation[J]. Creativity & Innovation Management, 2010, 19(3):290-303.

[5] Wooten J O, Ulrich K T. Idea generation and the role of feedback: evidence from field experiments with innovation tournaments[J]. Production and Operations Management, 2016, 26(1):80-99.

[6] Zhu H, Kock A, Wentker M, et al. How does online interaction affect idea quality? The effect of feedback in firm-internal idea competitions[J]. Journal of Product Innovation Management, 2019, 36(1):24-40.

[7] Hofstetter R, Zhang J Z, Herrmann A. Successive open innovation contests and incentives: winner-take-all or multiple prizes?[J]. Journal of Product Innovation Management, 2018, 35(4):492-517.

[8] Füller J, Hutter K, Faullant R. Why co-creation experience matters? Creative experience and its impact on the quantity and quality of creative contributions[J]. R & D Management, 2011, 41(3):259-273.

[9] Füller J, Hutter K, Hautz J, et al. The role of professionalism in innovation contest communities[J]. Long Range Planning, 2017,50(2):243-259.

[10] Poetz M K, Schreier M. The value of crowdsourcing: can users really compete with professionals in generating new product ideas?[J]. Journal of Product Innovation Management, 2012, 29(2):245-256.

[11] Cui Z, Kumar S P M, Gonçalves D. Scoring vs. ranking: an experimental study of idea evaluation processes[J]. Production and Operations Management, 2019, 28(1): 176-188.

[12] Özaygen A, Balagué C. Idea evaluation in innovation contest platforms: a network perspective[J]. Decision Support Systems, 2018, 112:15-22.

[13] Brunswicker S, Jensen B, Song Z, et al. Transparency as design choice of open data contests[J]. Journal of the Association for Information Science and Technology, 2018, 69(10): 1205-1222.

[14] 王丹丹. 面向教育的关联网络数据项目之全面开放机制研究[J]. 数字图书馆论坛, 2015(7): 44-48. (Wang Dandan. Research on the open mechanism of linking web data for education[J]. Digital Library Forum, 2015(7): 44-48.)

[15] Shneiderman B, Brethauer D, Plaisant C, et al. Evaluating three museum installations of a hypertext system[J]. Journal of the American Society for Information Science, 1989, 40(3): 172-182.

[16] Stoica A, Fiotakis G, Cabera J S, et al. Usability evaluation of handheld devices: a case study for a museum application[C]//Proceedings of PCI, Volos. Greece, 2005(11): 32-42.

[17] Tesoriero R, Gallud J A, Lozano M, et al. Enhancing visitors' experience in art museums using mobile technologies[J]. Information Systems Frontiers, 2014, 16(2): 303-327.

[18] Kabassi K. Evaluating websites of museums: state of the art[J]. Journal of Cultural Heritage, 2017, 24: 184-196.

[19] 杜华, 杨晓宏. 移动学习工具评价指标体系研究——以场馆 APP 评价为例[J]. 中国远程教育, 2018(8): 63-69. (Du Hua, Yang Xiaohong. Research on evaluation index system of mobile learning tools: taking the venues app evaluation as an example[J]. Distance Education in China, 2018(8): 63-69.)

[20] 张瑞. 数字博物馆的可用性评估[D]. 济南:山东大学, 2005. (Zhang Rui. Usability evaluation of digital museums[D]. Jinan: Shandong University, 2005.)

[21] 叶继元. 人文社会科学评价体系探讨[J]. 南京大学学报(哲学·人文科学·社会科学), 2010,47(1):97-110. (Ye Jiyuan. Approaching evaluation system in humanities and social sciences[J]. Journal of Nanjing University(Philosophy, Humanities and Social Sciences), 2010, 47(1):97-110.)

[22] 叶继元. 图书馆学期刊质量"全评价"探讨及启示[J]. 中国图书馆学报,2013,39(4):83-92. (Ye Jiyuan. Research and evaluation of the "All-round Evaluation" of the quality of journals in library sciences[J]. Journal of Library Science in China, 2013,39(4):83-92.)

[23] 范佳佳,叶继元. 21世纪以来网站评价研究进展与趋势[J]. 中国图书馆学报,2014,40(2):92-106. (Fan Jiajia, Ye Jiyuan. Research progress and trends of website evaluation since the 21st century[J]. Journal of Library Science in China, 2014,40(2):92-106.)

[24] 范佳佳,叶继元. 基于结构方程的科技网站信息质量评价模型构建及应用[J]. 图书馆杂志, 2016,35(9):66-75. (Fan Jiajia, Ye Jiyuan. Construction and application of science and technology website information quality evaluation model based on structural equation[J]. Library Journal, 2016,35(9):66-75.)

[25] 卢文辉,叶继元. "全评价"分析框架下的图书馆网站评价研究[J]. 图书馆杂志,2018,37(10):31-42,60. (Lu Wenhui,Ye Jiyuan. Evaluation of library websites under the analytical framework of "Full Evaluation"[J]. Library Journal,2018,37(10):31-42,60.)

[26] 陈铭,叶继元. 基于"全评价"分析框架的开放存取仓储评价体系研究[J]. 图书馆论坛,2014,34(8):40-47. (Chen Ming, Ye Jiyuan. The evaluation of open access repositories based on all-round evaluation system [J]. Library Tribune, 2014,34(8):40-47.)

[27] 叶继元,陈铭. 开放存取期刊学术质量"全评价"体系研究——以"中国科技论文在线优秀期刊"为例[J]. 图书与情报, 2013(2):81-87. (Ye Jiyuan, Chen Ming. The evaluation of open access journals based on all-around evaluation system: take the outstanding journals in science paper online for example[J]. Library & Information, 2013(2):81-87.)

[28] Delone W H, McLean E R.. Information systems success: the quest for the dependent variable[J]. Information Systems Research, 1992, 3(1): 60-95.

[29] Delone W H, McLean E R.. The DeLone and McLean model of information systems success: a ten-year update [J]. Journal of Management Information Systems, 2003, 19(4): 9-30.

[30] 王文韬, 谢阳群,谢笑. 关于 D&M 信息系统成功模型演化和进展的研究[J]. 情报理论与实践, 2014, 37(6): 73-76. (Wang Wentao, Xie Yangqun, Xie Xiao. Research on the evolution and progress of D&M information system success model[J]. Information Studies: Theory & Application, 2014, 37(6): 73-76.)

[31] 王知津, 李明珍. 网站评价指标体系的构建方法与过程[J]. 图书与情报, 2006(3): 45-52. (Wang Zhijin, Li Mingzhen. The methods and process of indicator system-building in websites evaluation[J]. Library & Information, 2006(3): 45-52.)

[32] Dragulanescu N G. Website quality evaluations: criteria and tools[J]. The International Information & Library Review, 2002, 34(3): 247-254.

[33] Kim P, Eng T R, Deering M J, et al. Published criteria for evaluating health related web sites: review[J]. British Medical Journal, 1999, 318(7184):647-649.

[34] 陈雅, 郑建明. 网站评价指标体系研究[J]. 中国图书馆学报, 2002, 28(5): 57-60. (Chen Ya, Zheng Jianming. Research on website evaluation index system[J]. Journal of Library Science in China, 2002,28(5): 57-60.)

[35] 张洋, 张磊. 网络信息资源评价研究综述[J]. 中国图书馆学报, 2010, 36(5): 75-89. (Zhang Yang, Zhang Lei. Review of web information resources evaluation[J]. Journal of Library Science in China, 2010, 36(5): 75-89.)

[36] 王尊新. 高校图书馆网站评价指标体系研究[J]. 现代图书情报技术, 2005(3): 60-62. (Wang Zunxin. Study of evaluation indicators system for college library website[J]. New Technology of Library and Information Service, 2005(3): 60-62.)

[37] Kim S, Decoster E. Organizational schemes of information resources in top 50 academic business library websites [J]. The Journal of Academic Librarianship, 2011, 37(2): 137-144.

[38] Garibay C, Gutiérrez H, Figueroa A. Evaluation of a digital library by means of quality function deployment (QFD) and the Kano model[J]. The Journal of Academic Librarianship, 2010, 36(2): 125-132.

[39] 熊伟, 叶淋玮. 我国虚拟旅游网站的功能评价研究[J]. 人文地理, 2011(2): 154-160. (Xiong Wei, Ye Linwei. A study on function evaluation of virtual tour websites in China[J]. Human Geography, 2011(2): 154-160.)

[40] 马费成, 姜婷婷. 信息构建对当代情报学发展的影响[J]. 图书馆论坛, 2003, 23(6): 20-25. (Ma Feicheng, Jiang Tingting. Impacts of information architecture on the development of contemporary information sci-

ence[J]. Library Tribune, 2003, 23(6): 20-25.)

[41] 周晓英. 信息构建的基本原理研究[J]. 图书情报工作,2004(6):5-7. (Zhou Xiaoying. The basic principles of information architecture[J]. Library and Information Service,2004(6):5-7.)

[42] Othman M K, Sulaiman M N S, Aman S. Heuristic evaluation: comparing generic and specific usability heuristics for identification of usability problems in a living museum mobile guide app[J]. Advances in Human-Computer Interaction, 2018:1-13.

[43] Chao H. Assessing the quality of academic libraries on the web: the development and testing of criteria[J]. Library & Information Science Research, 2002, 24(2): 169-194.

[44] Stoyanov S R, Hides L, Kavanagh D J, et al. Mobile app rating scale: a new tool for assessing the quality of health mobile apps[J]. JMIRM Health and UHealth, 2015, 3(1): 1-9.

[45] Loiacono E T, Watson R T, Goodhue D L. WebQual: an instrument for consumer evaluation of web sites[J]. International Journal of Electronic Commerce, 2007, 11(3):51-87.

[46] Tesoriero R, Gallud J A, Lozano M D, et al. Enhancing visitors' experience in art museums using mobile technologies[J]. Information Systems Frontiers, 2014, 16(2):303-327.

[47] 黄炜, 李总苛, 黄建桥. 移动应用程序(APP)评价指标体系研究[J]. 图书与情报, 2016(3): 110-117. (Huang Wei, Li Zongke, Huang Jianqiao. Research on evaluation index system of APP[J]. Library & Information, 2016(3): 110-117.)

[48] Smith A G. Applying evaluation criteria to New Zealand government websites[J]. International Journal of Information Management, 2001, 21(2):137-149.

[49] 赵宇翔,张轩慧. 基于层次分析法的弹幕视频网站信息构建评价[J]. 情报资料工作, 2017(3):42-51. (Zhao Yuxiang, Zhang Xuanhui. Evaluation study of danmaku video website information construction based on analytic hierarchy process[J]. Information and Documentation Services, 2017(3):42-51.)

[50] 郭顺利,张向先,李中梅. 高校图书馆微信公众平台传播影响力评价体系研究[J]. 图书情报工作,2016,60(4):29-36,43. (Guo Shunli, Zhang Xiangxian, Li Zhongmei. Research on the evaluation system of the WeChat official account admin platform of university libraries[J]. Library and Information Service, 2016,60(4):29-36,43.)

[51] Parente F J, Anderson J K, Myers P, et al. An examination of factors contributing to Delphi accuracy[J]. Journal of Forecasting, 1984, 3(2): 173-182.

[52] 朱庆华. 网络信息资源评价指标体系的建立与测定[M]. 北京:商务印书馆, 2012: 115-119. (Zhu Qinghua. Construction and measurement of evaluation index system of network information resources[M]. Beijing:The Commercial Press, 2012: 115-119.)

[53] 张炳江. 层次分析法及其应用案例[M]. 北京:电子工业出版社, 2014: 30-33. (Zhang Bingjiang. Analytic hierarchy process and its application cases[M]. Beijing:Publishing House of Electronics Industry, 2014: 30-33.)

[54] 吕跃进. 指数标度判断矩阵的一致性检验方法[J]. 统计与决策, 2006(9): 31-32. (Lv Yuejin. Consistency testing method for exponential scale judgment matrix[J]. Statistics & Decision, 2006(9): 31-32.)

[55] Opricovic S, Tzeng G H. Compromise solution by MCDM methods:a comparative analysis of VIKOR and TOPSIS [J]. European Journal of Operational Research, 2004, 156(2): 445-455.

赵宇翔 南京理工大学经济管理学院信息管理系教授,博士生导师。江苏 南京 210094。

刘周颖 南京大学信息管理学院博士研究生。江苏 南京 210043。

刘 炜 上海图书馆副馆长,研究员。上海 200031。

高胜寒 南京大学信息管理学院硕士研究生。江苏 南京 210043。

张 磊 上海图书馆系统网络中心开发部主任。上海 200031。

(收稿日期:2019-06-16;修回日期:2019-08-22)