

利用领域本体提高信息对称性的研究

李雨轩¹, 黄奇², 陈雪¹, 郑姝雅¹, 张戈³

(1. 南京大学信息管理学院, 南京 210023; 2. 南京大学国家信息资源管理南京研究基地, 南京 210023;
3. 南京大学工程管理学院, 南京 210093)

摘要 本文针对现实生活中的信息不对称问题, 提出了利用领域本体缓解信息不对称现象的一种模型, 给出了这个模型在购物和在线医疗咨询领域中的应用场景, 并依托购物场景对模型进行实证研究, 最后依据实证研究结果对模型进行了改进。实证研究表明, 该模型可以为信息不对称的双方提供结构化的知识和可视化的表示方式, 能够在一定程度上缓解信息不对称现象。

关键词 信息不对称; 领域本体; 信息资源共享

A Study of Information Symmetry from the Perspective of Domain Ontologies

Li Yuxuan¹, Huang Qi², Chen Xue¹, Zheng Shuya¹ and Zhang Ge³

(1. School of Information Management, Nanjing University, Nanjing 210023; 2. National Information Resource Management Nanjing Research Base of Nanjing University, Nanjing 210023;
3. School of Management & Engineering, Nanjing University, Nanjing 210093)

Abstract: This paper aims to examine information asymmetry in public life. It proposes a model that adopts domain ontologies in applications to solve this problem. The paper conducts an empirical study of how the model is applied to shopping and online medical consultation and improves the model according to the study results. The research indicates that, to some extent, this model can help counter information asymmetry by providing structural knowledge and visual rendering for both sellers and buyers.

Key words: information asymmetry; domain ontology; information resources sharing

1 引言

大数据的快速发展带来了信息量的激增, 面对各类庞大、复杂的信息, 人们在信息获取的过程中, 对于所需信息的获取途径、获取方法与辨别分析往往感知能力低下。大数据在某种程度上并没有使人

们变得“更有知识”, 反而使人们感受到一种新的知识匮乏, 这种信息过载引发的知识匮乏包括用户“知道自己不知道什么”和“不知道自己不知道什么”两种情形。这种匮乏受到了个体现有知识结构、信息检索、信息分析能力等信息素养的影响。

信息不对称是信息的基本特性, 是在相互对应

收稿日期: 2018-03-14; 修回日期: 2018-06-14

基金项目: 国家社会科学基金重点项目“基于知识组织的产品分类本体研究”(13ATQ005); 南京大学人文社科双一流建设第二批“百层次”科研项目“本体特征分析及其应用研究”。

作者简介: 李雨轩, 女, 1995年生, 在读研究生, 主要研究方向为网络信息资源的语义化、电子商务; 黄奇, 男, 1961年生, 博士, 教授, 主要研究方向为网络信息资源的语义化、电子商务, E-mail: huangq@nju.edu.cn; 陈雪, 女, 1993年生, 在读研究生, 主要研究方向为网络信息资源的语义化、电子商务; 郑姝雅, 女, 1995年生, 在读研究生, 主要研究方向为网络信息资源的语义化、电子商务; 张戈, 男, 1995年生, 在读研究生, 主要研究方向为网络信息资源的语义化、电子商务。

的经济人之间有关某些事件的信息或知识呈不对称分布。在经济行为决策中,非对称信息是指经济主体间因掌握的信息量不同而导致信息的不对称性、不均衡性^[1]。信息不对称包含两种情形:一种是信息劣势方知道自己与信息优势方之间存在怎样的不对称,可以通过信息获取等手段缓解不对称性;还有一种是信息劣势方自己也不清楚是怎样的不对称,这就增加了信息获取和缓解信息不对称的难度。信息不对称现象普遍存在于人类经济生活的各个领域。在电子商务领域中,由于买家和卖家的信息不对称,造成买卖双方在电子交易中的信任缺失,在市场交易中出现逆向选择和道德风险,影响市场机制正常发挥作用;在医疗卫生领域,医患关系紧张也在一定程度上受到病人和医生的信息不对称影响。

面对上述这些现实问题,本文提出了在应用层面上利用领域本体缓解信息不对称现象的一种模型,并给出这个模型在购物和在线医疗咨询领域中的应用场景,最后依托购物场景对这个模型进行了实证研究,并依据实证研究结果对模型进行了改进。

2 提高信息对称性的模型

诺贝尔经济学奖得主 G. Akerlof 教授通过对美国二手车交易市场的研究,发表了论文《柠檬市场:质量的不确定性与市场机制》并提出了逆向选择理论^[2]。在二手车市场中,买家和卖家关于二手车质量信息的不对称现象会造成交易中二手车的平均价格与平均质量轮番下降,以致把好车逐出市场而使坏车成交的逆向选择过程。信息不对称是市场不完全性的体现,会导致很多不确定性因素,提高交易成本^[3]。由于信息不对称的存在,市场会阻碍互利交易顺利进行。由此可见,买卖双方的信息不对称对于市场均衡水平起到决定作用,严重的信息不对称对

买卖双方的利益均有损害,因此需要用非市场方法弥补因信息不对称造成的缺陷^[4]。

为了弥补信息不对称所造成的缺陷,有学者认为可以通过设计激励机制来均衡信息不对称环境^[5],有学者提出应建立有效的政府干预机制来防止和纠正信息市场失灵^[6],还有学者指出良好的信息资源共享市场规制的方案应该包括合理的市场结构、有效的价格规制和良好的共享秩序^[7]。本文试图通过引入领域本体,提出了提高信息对称性的一种模型,至今未发现相同的研究。

由于诺贝尔经济学奖得主 G. Akerlof 教授已通过实验指出严重的信息不对称对买卖双方的利益均有损害,影响市场机制正常发挥作用。在现有的互联网经济中,卖家也越来越重视和买家的沟通交流、重视买家的评价与反馈,愿意在信息服务中付出时间精力,在这种现实基础与趋势下,本文假定信息优势方出于维护自身利益的考虑,愿意去缓解信息不对称性。本文的研究重点在于如何提高信息对称性来缓解市场交易中的不公平与信用缺失问题,而不在于研究信息优势方提供给信息劣势方的信息意愿。

本体是一种形式化的、对于共享概念体系的明确而又详细的说明,是关于特定知识领域内各种对象、对象特性和对象间可能存在关系的理论,也是一种知识重用、共享和建模的工具^[8-9]。领域本体(domain ontologies)描述的是特定定义域的知识,提供该定义域特定的概念的定义以及概念之间的关系,可以在定义域范围内共享和重用。由于本体具有形式化、结构化、可以通过机器推理查询等优点,是解决语义层次上 Web 信息共享和交换的基础,有助于知识共享与重用。因此本文将领域本体引入,提出了借助领域本体提高信息对称性的一种模型,如图 1 所示。

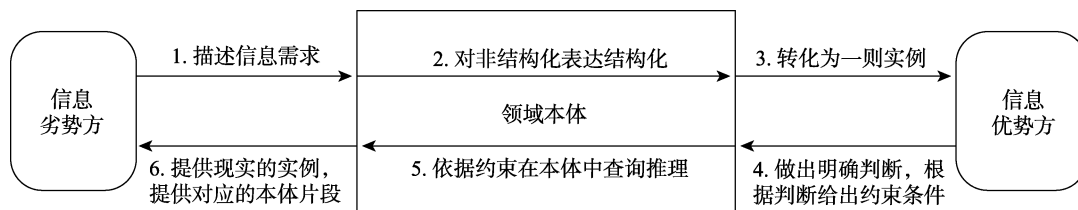


图 1 提高信息对称性的一种模型

图 1 是由信息劣势方、信息优势方和基于领域本体的第三方平台这 3 个对象组成的模型。在该模型中,信息劣势方向第三方平台描述信息需求,平台将信息劣势方的非结构化表达转化为结构化表

达,进而转化为一则实例提供给信息优势方,相比信息劣势方最初的信息需求文字描述,这种可视化结构化的实例表达更易信息优势方理解;信息优势方依据这则实例中的信息在平台中查询推理,并结

合自己已有的知识和经验,提供出几则符合信息劣势方需求且匹配现实的实例,并将相应内容的本体提供给信息劣势方。

由于图 1 提高信息对称性的概念模型较为抽象,结合现实场景中购物场景的特点,形成了如图 2 所示的在购物场景中提高信息对称性的一种应用模型。本文在购物场景中的信息不对称是指关于商品信息的不对称,买家是信息劣势方,卖家是信息优势方。

图 2 所示的购物场景中的电子交易平台是基于领域本体架构的。例如,买家需要采购一批用于办公的扫描仪,但苦于对这一领域并不了解,不知该如何挑选。而在如图 2 所示的模型中,买家首先在平台中通过类目查询或关键词检索找到所需的商品类目“扫描仪”,平台会弹出一个包含“扫描仪”类目的子类目和属性界面,买家在页面中选择扫描仪各个属性(如“价位”、“品牌”)的属性值范围,形成了一

份买家购买意愿清单。平台将买家填好的这份购买意愿表单转化为一则实例提供给卖家,卖家依据买家的“购买意愿实例”利用本体进行查询和推理。由于买家的“购买意愿实例”有可能是过于理想化的,所以需要卖家结合已有的知识和经验,提供出几则匹配现实的、最能达到买家要求的商品实例,买家将同时获得几则符合需求的商品实例和“扫描仪”领域的相应本体。在这场交易中,买家的信息变得更加完备,买家和卖家的信息不对称性得以缓解。

除了购物场景外,这个模型还可以应用于在线医疗咨询场景。现有的在线医疗咨询模式如图 3 所示,病人在平台中对病情进行文字描述,医生依据描述诊断病情,并通过在线咨询平台反馈给病人。

基于本文提出的提高信息对称性的一种模型(图 1),结合在线医疗领域的特点,形成如图 4 所示的在线医疗咨询领域中提高信息对称性的一种模型。

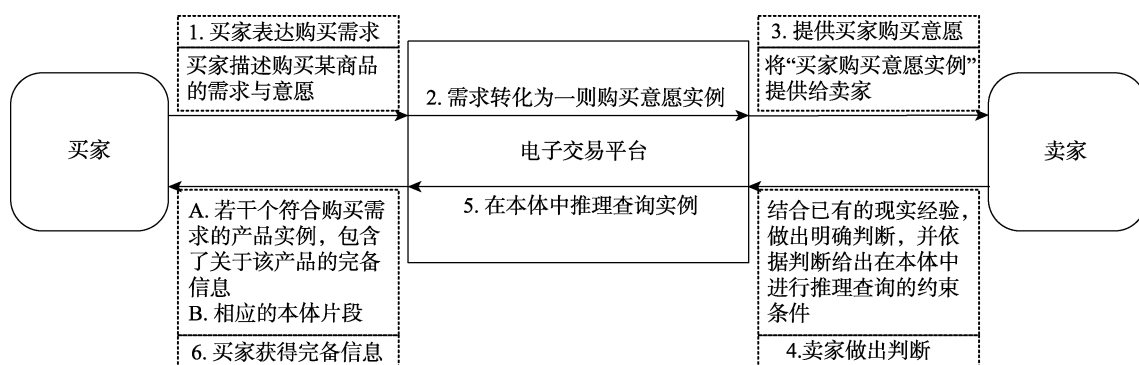


图 2 购物场景中提高信息对称性的一种模型

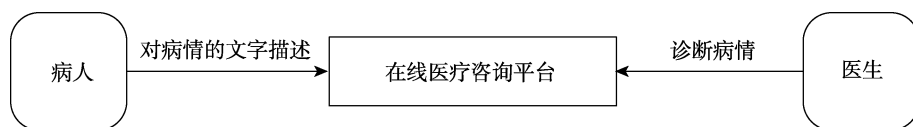


图 3 现有的在线医疗咨询模式

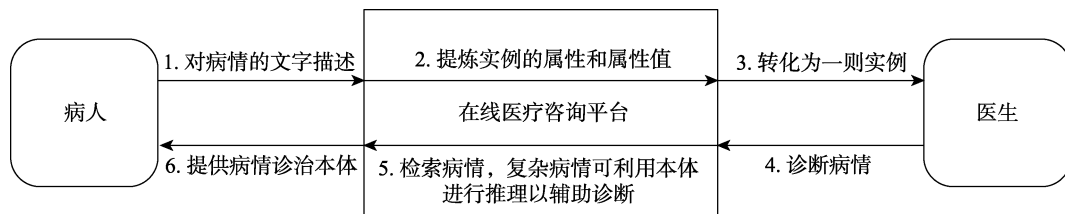


图 4 在线医疗咨询中提高信息对称性的一种模型

病人首先对病情进行文字描述,例如,病人 A 自述病情:“我今年 25 岁,女,以前没得过大病。我发现我的手指头会经常抖动,不受自己控制。我

有时会感觉心慌不舒服,肚子特别容易饿,饭量比朋友都大。甲状腺六项功能检查显示我的促甲状腺激素非常低,游离甲状腺素、游离三碘甲状腺原氨

酸、促甲状腺素受体抗体都大大地超出正常值。我还经常出汗。”依据这段病人自述，平台对这段非结构化表述进行结构化处理，提炼出实例的属性值和属性值，形成实例 A，如图 5 所示。

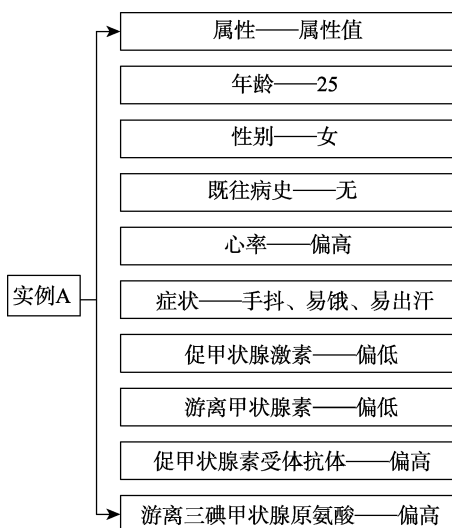


图 5 实例 A

相比于文字描述，这种可视化、结构化的实例表现更易被医生理解。医生依据实例 A 的属性和属性值，诊断结果为甲亢，将该结果反馈给在线医疗咨询平台。平台在本体中检索甲亢，得到“甲状腺功能亢进症”本体，将该结果反馈给病人，如图 6 所示。

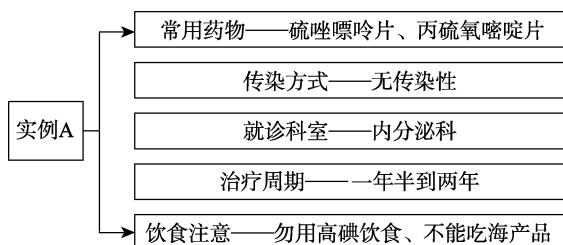


图 6 甲亢诊治本体

通过图 4 所示模型，病人既得到了有效的病情诊断，又对自己所患的病情有了结构化、规范化的了解和认知，病人和医生之间的信息对称性得以提高。

综合图 1 提高信息对称性的一种通用模型、图 2 购物场景中提高信息对称性的一种模型和图 4 在线医疗咨询中提高信息对称性的一种模型来看，可以发现：

(1) 领域本体的引入有助于提高信息双方的对称性，改进语义层次上 Web 信息共享和交换。由于信息劣势方提供的表述是非专业性的自然语言表

述，基于本体的第三方平台将非专业的表述映射到规范化、明确的本体语言中，将规范后的表达提供给信息优势方，有助于减少信息二义性的产生。

(2) 本体为信息劣势方和信息优势方均提供了结构化的知识和可视化的表示方式，可以更好地帮助双方理解信息。

(3) 本体具有概念正确、语义完备的特点，但并不代表实例完备。由于信息需求方需要的是解决现实问题，所以信息需求方提出信息需求后，还需要信息提供方提供恰当的匹配现实的实例。因此，模型中的三个对象都有其存在的意义。

3 依托购物场景的实证研究

为了验证上文提出的模型在现实生活中是否可以提高信息对称性，本文以其中的购物场景中提高信息对称性的一种模型为例，以钢笔作为购买对象进行实证研究。由于实验场景和成本的限制，本文将模型拆解为 3 个部分进行研究：

(1) 在购买某一商品时，买家和卖家是否存在信息不对称现象？若存在，是怎样的信息不对称？即购物场景中的信息不对称现状。

(2) 领域本体的引入是否有助于提高买家对于商品的了解和认知，改善其信息劣势现状？即引入领域本体对于改善信息不对称现象的作用。

(3) 相比于文字描述等表现形式，领域本体是否可以促进信息双方更好地理解信息？即领域本体对于促进信息双方理解沟通的作用。

第一个问题是对模型的前提——信息不对称现状进行了验证，第二个问题对于在模型中引入本体的作用进行研究，第三个问题则是对在模型中引入领域本体的优势进行研究。

由于购物场景中买家和卖家的信息不对称中主要来自于对产品指标即产品属性的信息不对称，本文以产品属性作为信息量的主要衡量参数。

3.1 购物场景中的信息不对称现状

3.1.1 卖家的信息量

本文以日用品钢笔为例，在淘宝网上搜索“钢笔”，随机挑选了 40 个卖家销售的不同钢笔，钢笔的价位从 3.5 元到 938000 元不等，对卖家提供的钢笔属性进行了调研与统计，40 位卖家提供的钢笔属性汇总情况如表 1 所示。

表 1 40 位卖家提供的钢笔属性汇总情况

| 钢笔属性汇总 | |
|--------|------|
| 笔杆材质 | 价格 |
| 笔尖嵌入方式 | 墨水材质 |
| 笔尖种类 | 品牌 |
| 笔身纹理 | 上墨方式 |
| 产品名称 | 生产企业 |
| 钢笔笔尖成分 | 适用场景 |
| 国家地区 | 书写粗细 |
| 货号 | 型号 |
| 颜色分类 | |

3.1.2 买家的初始信息量

为了获取买家在购买钢笔时所考虑的钢笔属性,本文随机挑选了 40 位买家进行调研,对其在挑选钢笔时会考虑的钢笔属性进行统计,40 位买家初始考虑的钢笔属性汇总情况如表 2 所示。

表 2 40 位买家初始考虑的钢笔属性汇总情况

| 钢笔属性汇总 | | |
|---------|-----------|----------|
| 重量 | 容纳墨水量 | 粗细 |
| 质感 | 品质 / 质量 | 出水率 |
| 长度 | 品牌 / 牌子 | 出水流畅度 |
| 颜值 | 耐用性 / 耐用度 | 出水粗细 |
| 颜色 | 用途 | 产地 |
| 形状 | 耐摔性 | 材料 |
| 写字质感 | 墨水 | 笔芯构造 |
| 写字粗细 | 墨囊类型 | 笔围 |
| 吸墨水方式 | 墨胆类型 | 笔头形状 |
| 吸墨水的量 | 款式 | 笔头的样子 |
| 握杆光滑度 | 口碑 | 笔头 |
| 握笔的感觉 | 壳子的材质 | 笔尖长短 |
| 外观 / 样式 | 结构 | 笔尖类型 |
| 书写顺滑性 | 价格 / 价位 | 笔尖接触纸的触感 |
| 书写是否流利 | 滑不滑 | 笔尖的调调 |
| 书写流畅度 | 好不好用 | 笔尖粗细 |
| 书写感觉 | 好不好写 | 笔尖材质 |
| 手感 | 好不好看 | 笔杆颜色 |
| 是否顺畅 | 钢笔长短 | 笔盖与笔身的比例 |
| 是否便携 | 服务态度 | 保养难易 |

从表 2 可以看出,买家对于钢笔属性的描述存在如下三个特点:

(1) 买家表达口语化、不规范,表达存在多义性。如买家描述的“滑不滑”并未指代清楚是指笔杆、

握位还是笔芯的光滑度。除此之外还有“手感”、“痕迹粗细”等非规范的表达。

(2) 买家描述的属性更偏重商品封装后的属性和使用体验上的属性。如“是否舒适”的属性值是主观、因人而异的;“书写流畅度”、“手感”、“出墨均匀”、“质感”、“外观”、“质量”等属性对应的属性值是无法量化的,卖家与供应商也不会把这些作为商品属性,这些实际对应的应为“材质”、“人体工程学设计”、“颜色”、“笔尖材质”、“笔尖形状”、“上墨系统”等基本属性。

(3) 其他属性。如口碑、服务态度等属于附加服务的东西,不是商品本身的属性。

对比表 1 卖家提供的钢笔属性和表 2 买家初始考虑的钢笔属性可以发现,买家和卖家的信息不对称现状是存在的,具体体现在:①买家选购商品时有关商品自身基本属性的有效信息量小于卖家的信息量;②买家在选购商品时更看重商品封装后的属性和使用体验的属性,而卖家的信息量主要为商品生产、供应角度的属性,二者存在不平衡性;③买家的表达较为口语化,存在不规范和二义性现象,一定程度上影响到买卖双方的交流。

3.2 领域本体对于改善信息不对称现象的作用

为了验证领域本体的引入是否有助于提高买家对于商品的了解和认知,提升其有效信息量,本文基于国际通用产品分类 eCl@ss 构建的产品分类本体中的子本体“钢笔”,以可视化的方式呈现给 40 位买家,如图 7 所示。

接着调研在有领域本体参考的情况下,这 40 位买家在挑选钢笔时会考虑的属性,汇总表如表 3 所示。

对比引入钢笔本体前后的买家所考虑钢笔属性信息发现,当引入领域本体后,买家在挑选钢笔时考虑的属性数量有了大幅度增长,说明买家所考虑的属性数量受到其信息量的制约,因此本文以属性数量作为信息量的衡量是合理的。同时,买家的表达规范性和专业性有了大幅度提高,对于钢笔属性的描述更侧重生产和供应的角度,对应的属性值是科学可测量的,表达的二义性有了明显的改善。因此,领域本体的引入是有助于提高买家对于商品的了解和认知,改善其信息劣势现状的。

3.3 领域本体对于促进信息双方理解沟通的作用

本体可以为用户提供图形化的概念和关系描述,

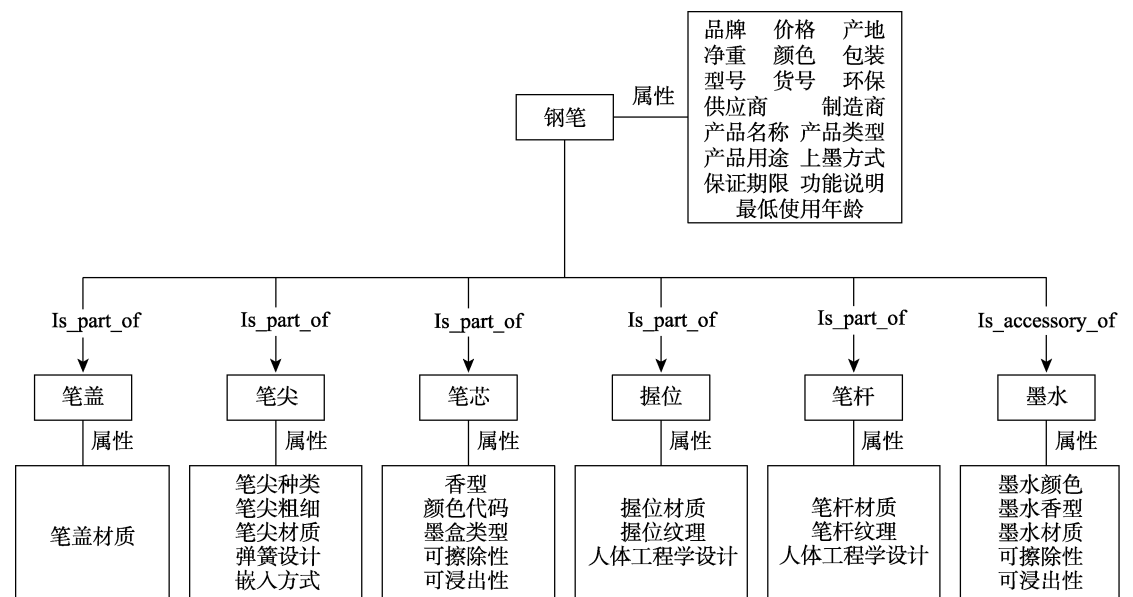


图 7 钢笔本体

表 3 领域本体引入后 40 位买家考虑的钢笔属性汇总情况

| 钢笔属性汇总 | |
|-------------------------------------|---|
| 最低使用年限 | 可擦除性 |
| 制造商 | 净重 |
| 造型是否特别 | 价格 |
| 颜色 | 环保 |
| 型号 | 供应商 |
| 握位（纹理 / 材质 / 人体工程学设计） | 功能说明 |
| 外观 | 弹簧设计 |
| 是否限量版 | 产品用途 |
| 上墨方式 | 产品名称 |
| 人体工程学设计 | 产地 |
| 嵌入方式 | 材质 |
| 品牌 | 笔芯（颜色代码 / 香型 / 可浸出性 / 可擦除性） |
| 墨水（颜色 / 香型 / 品牌 / 材质 / 可浸出性 / 可擦除性） | 笔尖（种类 / 粗细 / 材质 / 嵌入方式 / 可浸出性） |
| 墨盒类型 | 笔杆（颜色 / 纹理 / 材质 / 设计 / 舒适度 / 美观度 / 人体工程学设计） |
| 可浸出性 | 笔盖（样式 / 材质） |
| 包装 | |

视觉信息对于了解一个领域知识被证明是非常有用的^[10-11]；在知识门户中拥有一个本体，可以帮助销售人员了解领域中哪些概念是重要的、是如何关联的，帮助他们找出有用的关键词来搜索知识库，并帮助他们了解所检索的文档的含义。视觉本体有助于帮助用户学习知识领域的相关概念和关系，以更专业的方式搜索知识库。

本体与信息可视化二者的结合即为知识可视化，展示的是经过提炼、组织、检索后的与用户查

询语义匹配度较高的知识元素，而非海量信息的简单罗列和展示^[12]。可视化的领域本体对概念间关系的表达呈现深层化和多样化的特点，拥有良好的知识检索接口，具有丰富的类目显示、全面的信息展示、易用性与可扩展性等优势^[13]。

对于第三个问题，由于已有国内外学者做了大量的研究与论证工作，本文就不再重复论证。可以发现，一方面，用户的认知过程是从非结构化走向结构化的，而领域本体便是以结构化的方式描述领

域知识;另一方面,已有研究证明可视化的本体有助于促进用户对于信息的理解。因此将领域本体引入模型中是有助于促进信息双方的更好理解与交流的。

4 购物场景中提高信息对称性的模型改进

通过实证研究发现,虽然信息优势方相比信息

劣势方拥有更多的信息,但二者对于该领域的知识都不是完备集。由于卖家有关商品的信息也是不完备的,所以卖家在商品采购时也可以参照本体进行调整。除此之外,虽然领域本体是概念正确且语义完备,但不代表实例完备。由于产品更新换代很快,实例是动态变化的,当有新的实例出现时,卖家可以将其及时补充入本体库中。基于以上两点,本文对购物场景中的模型进行改进与修正,如图 8 所示。

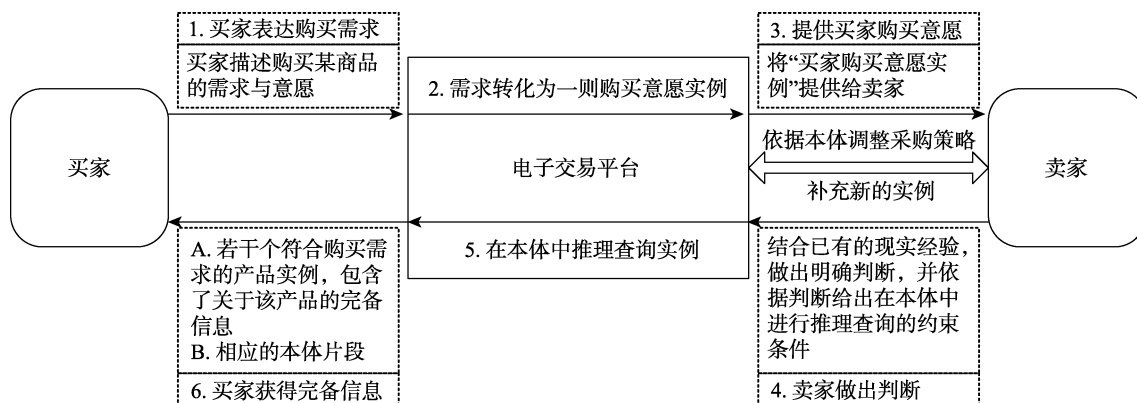


图 8 购物场景中提高信息对称性的改进模型

5 结束语

本文针对现实生活中普遍存在的信息不对称问题,提出了在应用层面上利用领域本体缓解信息不对称现象的一种模型,给出这个模型在购物和在线医疗咨询中的应用场景,依托购物场景对这个模型进行了实证研究,并依据实证研究结果对模型进行了改进。通过实证研究发现:

(1) 购物场景中,买家关于商品自身属性的有效信息量小于卖家的信息量,信息不对称现状主要体现在买家更看重商品封装后的属性和使用体验的属性,而卖家的信息量主要为商品生产、供应角度的属性,二者存在不平衡性;同时,由于买家的表达较为口语化,存在不规范和二义性现象,一定程度上影响到买卖双方的交流。

(2) 引入领域本体后,买家和卖家间的信息不对称现象有了明显改善,买家在挑选钢笔时考虑的属性数量有了大幅度增长,同时买家的表达规范性和专业性有了大幅度提高,表达的二义性有了明显的改善,结构化、可视化的本体可以促进信息双方的理解与交流,有助于改进语义层次上的 Web 信息共享和交换。

(3) 虽然信息优势方相比信息劣势方拥有更多的信息,但二者对于该领域的知识都不是完备集。

虽然领域本体是概念正确且语义完备,但不代表实例完备。由于产品更新换代很快,实例是动态变化的,当有新的实例出现时,卖家可以将其及时补充入本体库中。因此,本文对购物场景中的模型进行改进与修正。

本文在实证研究中使用的“钢笔”本体是基于国际通用产品分类 ecl@ss 构建的,ecl@ss 给出的是钢笔的描述,但是没有直观性和逻辑性,非专业人士具有一定的理解难度。而本体理清了它的逻辑关系,以图形化、直观化、通俗化的方式提供给用户,有助于用户对于知识的理解与获取。本文现有的属性池是依据 ecl@ss 得到的,未来还可以通过用户生成内容对属性的科学性进行进一步验证。

参 考 文 献

- [1] 陈建斌. 信息经济学[M]. 北京: 清华大学出版社, 2010: 35-37.
- [2] Akerlof G A. The market for “Lemons”: Quality uncertainty and the market mechanism[J]. Quarterly Journal of Economics, 1970, 84(3): 488-500.
- [3] 卢小宾. 我国信息技术创新的市场学分析[J]. 情报学报, 2002, 21(3): 373-377.
- [4] 乌家培. 信息经济学与信息管理[M]. 北京: 方志出版社, 2004: 89-91.
- [5] 辛琳. 信息不对称理论研究[J]. 嘉兴学院学报, 2001(3): 38-42.
- [6] 查先进. 论信息市场失灵与政府干预[J]. 中国图书馆学报,

- 2000(4): 27-29, 40.
- [7] 马费成, 裴雷. 信息资源共享的市场规制[J]. 中国图书馆学报, 2004(3): 5-10.
- [8] Gruber T R. A translation approach to portable ontology specifications[J]. Knowledge Acquisition, 1993, 5(2): 199-220.
- [9] 李金海, 马云蕾, 孙玲芳, 等. 基于语义一致性的多层本体元模型构建方法研究[J]. 情报学报, 2017, 36(5): 494-502.
- [10] Mayer R E, Heiser J, Lonn S. Cognitive constraints on multimedia learning: When presenting more material results in less understanding[J]. Journal of Educational Psychology, 2001, 93(1): 187-198.
- [11] Turetken O, Sharda R. Visualization of web spaces: state of the art and future directions[J]. ACM Sigmis Database, 2007, 38(3): 51-81.
- [12] 董慧, 王超. 本体应用可视化研究[J]. 情报理论与实践, 2009, 32(12): 116-120.
- [13] 张玉峰, 周磊, 王志芳, 等. 领域本体构建与可视化展示研究[J]. 情报理论与实践, 2012, 35(10): 95-98, 128.

(责任编辑 马 兰)