

构建知识图谱，衡量高校课程体系与产业需求契合度——以软件工程为例^{*}

马春燕 郑江滨 张涛

西北工业大学软件学院，陕西 710129

摘要 针对高校课程体系是否满足产业高速发展需求的问题，目前多采用人工分析的方法进行判断，这存在时间和人力投入大，准确性、及时性和可靠性不足等问题，我们借助知识图谱和自然语言处理等现代信息技术，通过可视化展示和知识自动对比的方式，帮助高校科学、客观和及时衡量课程体系中学生知识和能力的培养是否存在落后于产业需求的问题。我们从知识图谱理论出发，深入探讨构建课程与产业需求等两类知识图谱的方法，以及如何运用这两类知识图谱的对比分析，客观地衡量课程体系与产业需求的契合度及目前课程体系存在的问题。最后以西北工业大学软件工程专业为例，展示了知识图谱在衡量高校课程体系与产业需求契合度中的有效性。

关键字 课程体系，产业需求，知识图谱

Constructing a Knowledge Graph to Measure the Fit between College Curriculum System and Industry Needs--Taking Software Engineering as an Example

Ma Chunyan Zheng Jiangbin Zhang Tao

School of Software Northwestern Polytechnical University, Shanxi 710129, China

Abstract—To address the issue of whether the curriculum system of colleges and universities meets the needs of high-speed development of industry, artificial analysis is mostly used to judge, which has the problems of large investment in time and manpower, and insufficient accuracy, timeliness and reliability, etc. With the help of modern information technology such as knowledge mapping and natural language processing, we can help colleges and universities to scientifically, objectively and timely measure whether there is any problem of lagging behind the industrial needs in the cultivation of students' knowledge and competence in the curriculum system, by means of visualization demonstration and automatic comparison of knowledge. Whether the cultivation of students' knowledge and ability is lagging behind the industrial demand. Starting from the theory of knowledge mapping, we discuss the method of constructing two types of knowledge maps, such as curriculum and industrial demand, and how to use the comparative analysis of these two types of knowledge maps to objectively measure the degree of fit between the curriculum system and the industrial demand as well as the problems existing in the current curriculum system. Finally, the software engineering program of Northwestern Polytechnical University is taken as an example to demonstrate the effectiveness of knowledge mapping in measuring the fit between the curriculum system of universities and industrial demands.

Keywords—Curriculum system, Industry demand, Knowledge graph

1 引言

随着新工科理念的深入贯彻，以及一流学科建设的不断推进，以“专注产业、创新引领”为核心导向的人才培养模式要求我们从更深的层次，更宽的角度去审视和调整高校的专业课程体系和教学内容^[1]。这意味着高校应深入研究行业发展趋势，找出高校课程体系与产业需求的差距，定期更新教学内容，以保持与产业前沿的同步。目前衡量高校课程体系与产业需

求契合度主要通过人工分析的方式，首先，学校专家详细审查高校的课程和学习目标，确定它们与特定行业的相关性；其次，通过对企业专家进行函询或企业调研获得相应数据；然后，比较课程内容与产业需求，找出二者的差距；最后调整专业课程设置^[2]。然而，这种依赖人工分析的方法存在时间和人力投入大，其准确性（人工收集数据缺乏全面性）、及时性（信息收集周期较长），及可靠性（分析中人为因素占主导，存在主观性）不足等问题。幸而，现代新兴信息技术的发展为教育环境的变革和创新注入了新活力，为解决上述困境带来了更高效、准确的途径。

^{*} **基金资助：** 本文得到西北工业大学教育教学改革重点研究项目“课程体系&教学内容与产业需求匹配度研究和实践 ——以软件工程为例”的支持

采用信息化技术赋能教育教学也是我国教育的重要发展趋势和要求^[3]。在党中央、国务院和教育部密集出台《关于一流本科课程建设的实施意见》《教育信息化 2.0 行动计划》、《“十四五” 数字经济发展规划》、《中国教育现代化 2035》等文件中，均强调信息技术赋能教育教学的建设任务。鉴于此，本文以此为理论指导，将课程教学大纲和产业招聘需求文档为研究起点，将知识图谱作为信息技术手段，通过构建和分析课程知识图谱和产业需求知识图谱，自动对比高等教育人才培养中课程体系与产业需求的衔接问题，可以系统地、全面地比较和对接课程体系和产业需求，为高等教育的课程体系改革与优化提供依据。

2 课程和产业需求知识图谱的构建

知识图谱可以将分散在各个来源的知识进行整合、归纳和组织，使得知识的管理和检索更加高效和有序，

并且为自动化分析提供了基础^[4]。课程知识图谱是可视化表示课程知识的结构化模型，它通过将课程相关的概念、主题、实体和关系连接起来，构建一个有向图，帮助学生、教师和研究者更好地理解 and 导航复杂的课程内容，也有助于在此之上进行一些自动化分析^[5]。产业需求知识图谱是一种用于描述和分析各个产业需求和关联关系的知识表示方法。它可以帮助我们更好地理解产业现状、发展趋势和需求，以便作出更明智的决策。

2.1 课程和产业需求知识图谱的定义

本文从模式层和数据层对课程和产业需求知识图谱进行定义。

根据课程教学大纲和企业招聘信息的文本结构特点，将知识图谱模式层定义为如图 1 所示，知识图谱模式层由本体概念的集合、关系的集合组成。

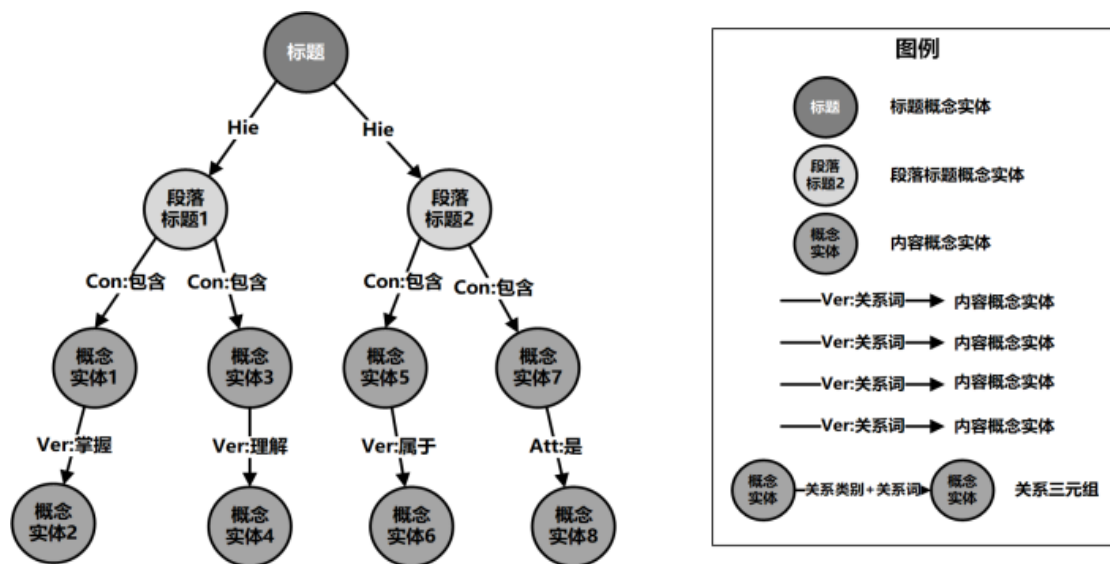


图 1 教学大纲和产业需求知识图谱模式层示意图

本体概念包括文档标题、段落标题，内容。概念之间的关系包括动词关系、包含关系、属性关系、层次关系。教学大纲和产业需求的知识图谱数据层，以“实体-关系-实体”的三元组存储，形成一个图状知识库^[6]。其中，实体是知识图谱的基本元素，指具体的标题名称、段落标题名称、内容中的概念实体等。关系是两个实体之间的语义关系，是模式层SCM所定义的动词关系、包含关系、属性关系以及层次关系的实例，例如“掌握”、“运用”、“理解”等属于动词关系Ver；“包括”、“属于”等属于包含关系Con；实体和对应属性的关系属于属性关系Att；“父类关系”、“子类关系”和“兄弟关系”等属于层次关系Hie。

2.2 课程和产业需求知识图谱的构建

本文结合自然语言处理技术，对课程体系对应的教学大纲，和企业招聘信息进行分析，以构建课程和产业需求对应的知识图谱，并进行可视化展示。

在教学大纲的知识图谱构建中，概念实体是构成知识图谱的基本元素之一，也是大纲中的重要信息单位，从教学大纲中提取概念实体是构建模型的重要环节。我们采用基于规则的技术提取概念实体，首先通过bert 预训练模型和Stanford CoreNLP 工具对教学大纲进行预处理，然后根据教学大纲文本特征设计并实现 4 种规则生成概念实体集合，最后从理论知识、技术

能力、开发工具和职业素养 4 个维度对教学大纲概念实体进行标记。具体方法如下：

(1) 教学大纲的预处理。具体步骤包括：

① 从输入的教学大纲文档中提取文本，使用 Stanford CoreNLP 对教学大纲文本进行指代消解，消除其中的指示代词；

② 基于标点符号（“ . ? ! ; ”）进行分句生成文本语句集合，并标记语句位置，包括大纲标题、段落标题和段落内容；

③ 基于 bert 预训练模型训练得到教学大纲文本描述二分类器，判断文本语句是否与课程要求或内容相关，教学大纲中可能会包含一些与课程要求或内容无关的内容，如开课学期等，为避免这些信息的干扰，本文人工构建数据集训练分类器来识别未包含有效课程信息的语句；

④ 通过分类器筛选所有与课程要求和内容相关的文本语句，通过 Stanford CoreNLP 对语句分词、词性标记后，加入课程要求和内容相关语句集合。教学大纲预处理结束，输出课程要求和内容相关语句集合。

(2) 基于筛选规则的概念实体生成方法。在教学大纲中，概念实体的词性通常为名词，多以名词短语的形式出现。相较于单个名词，名词短语在语义描述上更加准确和完整，因此采用名词短语描述概念实体。从名词短语抽取得到的候选实体集合中寻找概念实体的过程实际上是从所有名词短语集合中过滤非概念实体名词短语的过程。本文总结了以下 4 条筛选规则：

① 规则 R1 消除教学大纲对应课程无关名词或名词短语；

② 规则 R2 大纲标题或段落标题中作为主语或宾语的名词或名词短语为实体；

③ 规则 R3 将表示某种特定对象的名词或名词短语用泛指该类事物的名词代替；

④ 规则 R4 课程指示名词或名词短语用该课程全名替换。

(3) 教学大纲概念实体间关系的提取方法。本文设计了两种关系三元组抽取策略：基于句法特征的抽取策略与基于最短距离的抽取策略。基于句法特征的概念实体关系三元组抽取策略。本文通过 Stanford CoreNLP 进行依存句法分析以获得句法特征。依存句法分析的主要作用在于对输入的自然语言文本语句进行依存关系分析，从而揭示语句结构信息。在依存句法分析中，依存关系是指文本语句中词与词间的关系。若词与词之间存在依存关系，则其中一个词作为支配者，另一个词作为被支配者。基于最短距离的实体关

系三元组抽取策略。根据语料分析，组成实体关系三元组的实体对和关系指示词联系较为紧密，这种紧密表现在两个方面：一方面，实体对和关系指示词在依存句法关系图中的最短连通距离较小；另一方面，实体对和关系指示词在句子中的序列距离也较小。基于上述两个特点，本文首先筛选出具有两个及两个以上实体和关系指示词的语句，计算句子中任意两实体和关系指示词的序列距离和依存距离。将两个结果加权和作为总距离。最后选择总距离最短的实体对和关系指示词作为抽取结果。

总之，通过上述方法，构建出教学大纲的知识图谱，为教育资源的整合、教学内容的改进等应用场景提供支持。

在构建产业需求时，首先，本文通过爬虫技术自动抓取相关招聘信息，并进行整合预处理，形成岗位需求文本。其次，本文采用自然语言处理技术对产业需求进行分析，构建产业需求知识图谱。在分析岗位需求描述文本时，其描述语句结构与教学大纲中的语句结构相似，在某些细节处理上，它们在构建知识图谱的过程中存在以下两处差异：

(1) 预处理过程的差异：相较于教学大纲，岗位需求描述中的干扰信息较多。为此，我们需要完善文本提取方法，并更新文本筛选分类器的训练数据。接着，重新训练一个分类器，以判断岗位需求描述文本语句是否为有效的岗位需求描述。

(2) 概念实体集合筛选规则的差异：在提取产业需求概念实体时，我们将规则 R4 更改为“用岗位全称替换岗位指示名词或名词短语”。鉴于产业需求描述通常是动宾结构的简单句形式呈现，并在多语义环境下默认当前岗位作为主语而直接省略，我们将规则 R2 更新为“使用岗位全称补充动宾结构语句的主语”。具体来说，当基于句法分析发现句子呈现动宾结构且省略主语时，我们使用岗位全称来补充语句主语。

3 课程体系与产业需求契合度衡量

本文通过对比分析课程体系知识图谱与产业需求知识图谱的差异性，量化衡量课程体系与产业需求的差距，并以关键词的方式给出课程体系存在的问题。

由于知识图谱是一种图结构的数据，可以理解为多个从根节点不断拓展子图的结构，所以知识图谱的相似度可以转化为模型中所有子图的相似度。本文提出计算两个知识图谱间匹配度的方法，计算两个知识图谱间的差异性与相关性，并有效挖掘知识图谱中的知识，实现知识图谱间的比较，以衡量知识图谱的课程体系与产业需求契合度。对于知识图谱 A 和 B 相似

度 $sim_M(A, B)$ 的计算公式如公式(1)所示,称每个连通子图为一个知识图谱单元。

$$sim_M(A, B) = \frac{1}{|A|} \sum_{A_i \in A} \max_{B_j \in B} \{sim_U(A_i, B_j)\} \quad (1)$$

其中, $sim_U(A_i, B_j)$ 表示知识图谱A中某知识图谱单元 A_i 与知识图谱B中某知识图谱单元 B_j 的相似度。将知识图谱A中任意知识图谱单元分别和知识图谱B中所有的概念单元计算相似度 $sim_U(A_i, B_j)$,然后取其最大值作为知识图谱A中当前语句与知识图谱B的相似度。重复执行上述操作,最后将知识图谱A中所有知识图谱单元和知识图谱B的相似度求算术平均值,作为知识图谱A和知识图谱B的匹配度 $sim_M(A, B)$ 。

知识图谱单元 c_Q 与 c_R 间的相似度 $sim_U(c_Q, c_R)$ 计算如公式(2)所示.该计算是一个递归计算子图相似度的过程,为避免在计算时出现无限递归情况,在知识图谱单元匹配时指定模型中心词作为入口节点,帮助在整个匹配过程中能够有效地检索并获取到候选的部分知识图谱。

$$sim_U(c_Q, c_R) = w(c_Q, c) \cdot sim_c(c_Q, c_R) + \max_{each\ relation} \left[\sum_j w(c_Q, j) \cdot sim_r(r_Q^j, r_R^j) \cdot sim_U(c_Q^{r_Q^j}, c_R^{r_R^j}) \right] \quad (2)$$

$$w(c_Q, c) + \sum_j w(c_Q, j) = 1 \quad (3)$$

其中, sim_c 表示概念实体的相似度, sim_r 表示关系的相似度。 r_Q^j 表示知识图谱单元 c_Q 的第 j 条关系, r_R^j 表示知识图谱单元 c_R 的第 j 条关系, $c_Q^{r_Q^j}$ 和 $c_R^{r_R^j}$ 分别表示与 r_Q^j 和 r_R^j 有关的模型部分的入口节点。 $w(c_Q, c)$ 表示入口节点的权重, $w(c_Q, j)$ 则表示与入口节点相关的第 j 条关系的权重。 $\max_{each\ relation}$ 表示在所有关系下求解表达式的最大值。和入口节点相连的每一条关系的部分知识图谱可能有多个组合,在这些组合中我们取相似度的最大值作为结果。

综上,从理论知识、技术能力、开发工具和职业素养4个维度,基于改进的HowNet和扩展同义词词林的词语匹配技术,本文提出一种递归匹配子图的课程教学大纲和产业需求知识图谱匹配度计算方法,并导出报告,标识课程体系与产业需求的存在差距地方,衡量的结果可以帮助专业负责人和课程负责人实时、准确、客观掌握课程体系、课程内容与产业需求的差距,改变课程体系和课程教学内容落后的现状,从而快速提高学校教育的质量。

4 方法在软件工程专业中的应用

本文选取西北工业大学软件工程课程体系及内容中的20门课程教学大纲,以及2022年中国500强企业高峰论坛发布的“中国企业500强”名单中,涉及软件工程专业相关岗位的20家公司的招聘信息作为实验对象。通过构建包含1043个实体与1175个关系的课程体系及内容知识图谱,以及包含1566个实体与1712个关系的产业人才需求知识图谱,本文对教学大纲与岗位需求两类知识图谱的相似度进行计算,以获取每个课程与岗位要求的匹配度。

在对未匹配的内容进行整理和分析后,本文得出了33项课程教学内容空缺的分析结果。最终,18位高等教育教师和21位企业专家对分析结果的认同度达到了93.2%。

4.1 课程教学大纲和产业需求的知识图谱

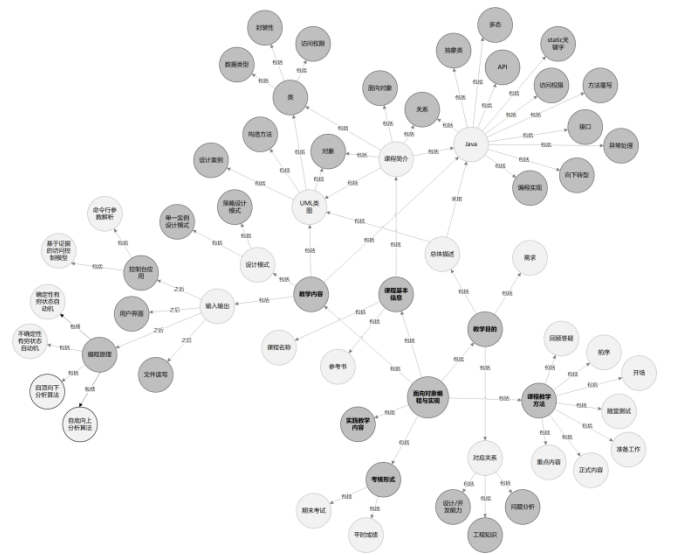


图2 《面向对象编程与设计》知识图谱示意图

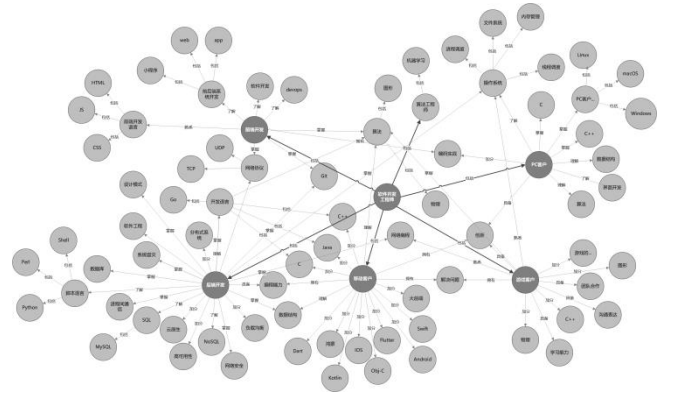


图3 产业需求知识图谱图示

以《面向对象编程与设计》课程教学大纲作为本文案例，构建的知识图谱利用 Neo4j 图数据库进行可视化展示，效果如图 2 所示。生成的产业需求知识图谱，存储在 Neo4j 图数据库中，并以可视化图形的方式展示，如图 3 所示。

表 1 岗位类别及匹配度结果

编号	岗位类别	课程教学大纲匹配度	
1	后端开发工程师	《面向对象编程与设计》	0.86
		《网络与分布式计算》	0.83
		《计算机网络》	0.78
		《数据库系统》	0.73
		《数据结构》	0.71
2	软件测试工程师	《软件测试》	0.88
		《数据库系统》	0.71
3	云计算工程师	《面向对象编程与设计》	0.73
		《计算机网络》	0.68
		《计算机操作系统》	0.63
4	客户端开发工程师	《面向对象编程与设计》	0.85
		《数据结构》	0.78
		《计算机网络》	0.65
		《计算机操作系统》	0.64
		《编译原理》	0.61
		《Python程序设计》	0.84
5	运维工程师	《计算机操作系统》	0.76
		《数据库系统》	0.69
		《数据结构》	0.54
6	前端开发工程师	《网页设计》	0.88
		《计算机网络》	0.73
		《计算机操作系统》	0.52
7	算法工程师	《Python程序设计》	0.85
		《数据结构》	0.81
		《算法设计与分析》	0.77
8	数据挖掘	《Python程序设计》	0.91
		《数据挖掘》	0.87
		《数据库系统》	0.72

4.2 衡量课程体系与产业需求契合度

从软件工程专业相关岗位中选取 8 类岗位的职业需求信息作为实验语料，依据本文提出的方法，得到课程与产业需求匹配度的部分实验结果如表 1 所示。

软件工程相关岗位技能需求以不同的匹配度体现在不同的课程教学大纲中。此外，对于教学大纲知识图谱中未覆盖的岗位要求进行定性分析，整理分析了 33 项课程教学内容空缺等分析结果，各岗位对应的教学大纲中部分空缺内容如表 2 所示。在教学大纲内容中存在未涵盖产业需求的内容，例如，hadoop 分布框架、Redis 缓存数据库、docker 容器技术、React 前端框架等。这些信息是高校专业课程体系逐步优化的有益参考，其将教学内容和就业市场需求更加紧密地联系在一起，并在实质层面揭示了二者之间的关系。

在就业市场需求中也有与当前教学内容关系较弱的要素。例如，所有相关岗位的需求中都要求“有较好的产品意识”、“有强烈的求知欲、好奇心和进取心，能及时关注和学习业界最新技术”等自身素质要求，这些信息可以作为素质教育内容对专业教学内容进行补充^[7]。

表 2 课程教学内容的空缺项

编号	岗位类别	课程教学内容空缺项
1	后端开发工程师	hadoop 分布框架
		缓存数据库
		高并发技术
		云边协同计算
2	软件测试工程师	自动化测试框架
3	云计算工程师	分布式架构
		缓存数据库
4	运维工程师	docker 容器技术
5	前端开发工程师	DevOps
		MVVM 框架
6	算法工程师	React 前端框架
		强化学习
		知识图谱
		计算机视觉

在就业市场需求中也有与当前教学内容关系较弱的要素。例如，所有相关岗位的需求中都要求“有较好的产品意识”、“有强烈的求知欲、好奇心和进取心，能及时关注和学习业界最新技术”等自身素质要求，这些信息可以作为素质教育内容对专业教学内容进行补充^[7]。

4.3 专家对分析结果的认可度调查

本文通过设计调查问卷，在提供相应培养方案和 20 门课程的教学大纲的前提下，邀请 5 类不同的从业单位中 18 位高等教育教师、21 位企业专家（见表 3），对 33 项课程的内容空缺进行查阅和评价，平均认同度达 93.2%。但经专家会同研讨，认为 33 项课程内容空缺项是否对某专业具有参考价值，应该进一步根据本专业的专业方向和培养目标，进一步进行确认。针对西北工业大学软件学院专业，本文将 33 项课程内容空缺项进行分析，得出如下 3 类结论。

表 3 调查问卷填写的专家信息

从业单位性质	人数	约占比 (%)
国有企业	3	8%
民营企业	14	36%
三资企业	2	5%
高等院校（事业单位）	18	46%
科研设计单位（事业单位）	2	5%
合计	39	

(1) 18 个空缺项代表的知识点应该融入课程内容, 某些课程内容需要更新, 该空缺项的分析对课程优化具有参考意义;

(2) 8 个空缺项代表的知识点无法融入相关课程, 应该开设 2 门前沿技术对应的新课程, 该空缺项的分析对课程体系优化具有参考意义;

(3) 7 个空缺项代表的知识点与本专业的专业方向及培养目标偏离较大, 对相应专业培养方案优化不具有参考意义。

5 结束语

课程体系是衡量高校人才培养供给侧与产业需求侧在结构、质量、水平上是否适应的关键载体, 其不断调整和优化是新时代教育改革的重要研究课题。我们采用知识图谱、自然语言处理技术以及可视化等现代信息技术手段, 实现课程体系与教学内容对接产业需求的契合度方面的自动化评估, 通过自动比较课程内容与产业需求, 找出二者的差距, 将分析结果反馈给学校, 以便学校调整课程内容, 更好地满足产业需求。与传统的人工分析相比, 该方法在时效性、准确性和客观性方面都有明显提升, 为培养方案的整体优

化和教学内容改革提供了新的途径, 并为教育教学改革提供了有力的决策支持。在今后的研究中, 可以针对不同领域和专业进行定制化分析, 以满足各个产业和学科的特殊需求。

参考文献

- [1] 刘国买, 姜哲. 以产业为要 推进人才培养改革创新[J]. 中国大学教学. 2022, (11): 12-16.
- [2] 胡淼, 吴迪. 大学计算机基础课程实践教学模式探索[J]. 计算机技术与教育学报. 2022, 10(3): 77-80.
- [3] 高松, 李正, 项聪. 信息化促进高等教育高质量公平——以东西部高校课程共享联盟实践为例[J]. 中国大学教学. 2022, (3): 4-10.
- [4] 吕洁华, 刘思彤, 刘艳迪. 多维度专业与就业匹配的度量方法探析——基于高校涉林专业的调查数据分析[J]. 教育探索. 2020, (7): 50-52.
- [5] 涂建华, 肖珺怡, 姜广峰. 构建微积分知识图谱 助推一流课程建设[J]. 中国大学教学. 2020, (11): 33-37.
- [6] 傅湘玲, 唐轶, 侯党, 等. 多计算机专业课程体系知识图谱构建及分析[J]. 计算机技术与教育学报. 2022, 10(3): 65-69.
- [7] 刘音, 程卫民, 刘震, 等. 专业学位课程体系与高级工程师岗位能力匹配度实证研究[J]. 大学教育. 2019, (10): 47-50.