

# A Knowledge Graph for Curriculum System

Ying-bo DONG and Xia HOU\*

Computer School, Beijing Information Science and Technology University, Beijing, China.

\*Corresponding author

**Keywords:** Knowledge Graph, Graph database, Curriculum system.

**Abstract.** In view of the large number of college courses, complex content and complex system context, it brings a lot of inconvenience to the course arrangement and analysis. A method of constructing a curriculum system based on knowledge Graph is proposed on This paper, by analyzing the attributes of the course and the pre-requisition of each other, extract the relationship between the precursor and the successor of course, we create a curriculum system based on knowledge graph, it use the graph database Neo4j for data storage and display. Use the system we can analyze the relationship between analytic course clearly, get some information that hard to draw in the artificial way, it contributes to the carding the syllabus and the needs of the quick query course context information.

## 一种课程体系知识图谱的设计与实现

董颖博, 侯霞\*

北京信息科技大学, 计算机学院, 北京, 中国

\*通讯作者

**关键词:** 知识图谱; 图数据库; 课程体系

**摘要:** 针对目前大学课程数量多, 内容杂, 体系脉络较为复杂, 为课程安排和分析工作带来诸多不便的问题, 提出一种基于知识图谱的课程体系构建方法, 通过对课程的属性和彼此间的前置要求进行分析, 从中提取前驱和后继之间的关系, 旨在构建课程体系的知识图谱, 利用图数据库Neo4j进行数据存储和展示。基于该图谱可清晰地解析课程之间的脉络关系, 获得一些人工很难得出的信息, 有助于课程大纲的梳理以及快速查询课程脉络信息的需要。

### 1. 引言

随着现代大学课程体系的不断发展和完善, 各个专业类别下不同课程之间的联系变得越来越密切, 因此, 设计课程体系知识图谱将有效促进课程体系研究与专业的发展, 方便师生查阅和了解相关课程及其脉络。

与其他领域的资源体系不同, 课程体系设计必须考虑课程的种类、课程之间的联系等, 通过人工整理或者利用 SQL 语句实现数据的统计、分析及预测, 对课程体系的关系进行梳理和统计, 需要耗费大量人力, 知识图谱是解决这些问题的有效办法。构建课程体系知识图谱的研究对于解决以下问题具有重要意义:

(1)利用可视化的图形界面可以有效地检测出课程大纲中的录入错误和编排错误

课程大纲的编写往往由人工完成, 在这个过程中, 可能会因为编写者和录入者的疏忽而出现一些各种各样的错误, 知识图谱的构建可以解决课程大纲中出现的录入错误和编排错误, 为课程体系知识图谱的构建和后续的分析工作提供有效支持。

## (2)可以有效地提高搜索课程及梳理其脉络关系的效率

课程体系的构建涉及到大量的课程和它们之间的脉络关系,知识图谱可以有效地把这些信息组合在一起,将混乱的信息源转化为有序且易于使用的知识源,通过对图谱中语义关系的描述,将课程大纲中散乱的信息进行整合和管理,然后发现、选择和组织有用的信息和知识,将其传递给需要的人或者系统,从而实现信息的有效分析。

本文提出一种基于知识图谱的课程体系构建方案,将以北京信息科技大学计算机大类的课程大纲为对象,对课程体系知识图谱的设计与实施进行研究,从课程的“先修课程”入手,系统地阐述了课程体系知识图谱的构建,并在实验数据的基础上说明如何进一步发现和发掘课程大纲中的存在的问题,并对其脉络体系进行梳理和分析,对特定课程节点的前驱后继进行查询并展示,提高课程体系知识图谱的使用效率。

## 2. 相关工作

知识图谱的概念最早由谷歌公司提出[1],其本质上是一种结构化的语义知识库,能够以便于用户理解的图形式描述现实世界中存在的实体及其关系,构成网状的知识结构[2]。对于揭示知识领域的动态发展规律,描述实体之间的逻辑关系,和预测领域的未来发展趋势,提供切实、重要的参考,具有更直观、易推断的优点。目前已被应用于体育、影视等多个领域[3, 4]。然而,其目前在课程体系构建领域的应用还不多见。

文献[5]通过对美国多所大学的游憩学本科课程体系进行梳理研究,通过词频统计和共词分析,构建了美国游憩学的知识图谱,并对其知识结构进行分析,对其他国家的相关专业设置提供了参考,但因为仅通过课程名称进行分析,无法完全体现各概念之间的真正联系,可能会把一些孤立的比较重要的课程排除在外;文献[6]通过对网络上的课程数字化资源进行模块化的语义分析和关键词提取,分析资源实体之间的关系,并最终绘制出该领域的知识图谱,实现了基于知识图谱的推荐工具,能够很好地为学习者推荐相关联的学习资源,为研究推荐方法的教育者们提供了一种新思路。

本文提出一种课程体系构建方案,基于知识图谱来清晰地描述课程间的逻辑关系,可以更有效地发现课程体系中的潜在编写错误和编排错误,利于使用者梳理课程之间的脉络关系。用户不必大量搜索相关知识点信息,而是通过知识图谱中的指定节点查看该课程的相关信息,省去繁琐的查找步骤,大大提高搜索效率,为现代化教学提供便利。

## 3. 课程体系知识图谱构建方法

本文研究的主要内容是:

### (1)构建课程体系知识图谱模式

根据课程名称和关系,构建课程体系知识图谱,需要从结构化的数据中获得课程的属性和相关信息,通过对这些信息的获取和梳理,定义了类型和实体。从而展开对课程体系知识图谱的表达、存储和查询,以及建立在知识图谱基础上的纠错问题。

### (2)获取多源实体知识

本项目创建知识图谱的数据来源主要有:多门课程的课程大纲,某一既定专业的专业培养方案等。这些数据中大部分都是结构化数据,本文将研究如何实现对这些数据精细准确的提取,并分析其中包含的结构化信息,实现多元实体知识的获取。

### (3)实体链接

知识图谱的创建过程如图 1 所示:

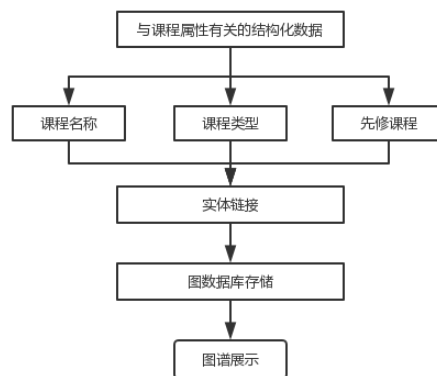


图1 知识图谱的创建过程

创建课程体系知识图谱，需要考虑课程本身的属性以及课程与课程之间的各种关系，通过对课程大纲的分析和归纳，课程属性或关系可以分为以下几种形式：

- 1)课程名称：课程的基础属性，也是最为关键的属性，例如“编译原理”。
- 2)课程类别：课程的性质，分为“必修”和“选修”。
- 3)课程学时：课程所占学时。
- 4)适用专业：课程使用的专业，例如“软件工程”。
- 5)课程学分：课程所占学分。
- 6)执笔人：课程大纲的撰写者。
- 7)审核人：课程大纲的审核人。
- 8)先修课程：课程的先修课程，例如“C 语言程序设计”。

利用知识图谱对课程体系的脉络进行梳理，可以更好的展开对课程的分析研究和编排工作。课程之间的先修关系是获取课程体系脉络的重要内容，同时也是课程之间最为紧密复杂的外部关联，本文将以北京信息科技大学计算机专业的课程大纲和专业培养方案为数据来源，提取其中相关的课程及其关系，从而使课程之间的关联条理化，并建立课程体系的知识图谱。

#### (4)绘制知识图谱

1)数据预处理：课程体系的脉络主要由两个部分组成，每门课程的课程大纲和专业的培养方案，其中课程大纲往往是以 **WORD** 文档的形式存在，**WORD** 文档中的数据包括表格形式的课程属性和课程之间的先修关系，而 **Excel** 文档中的数据包括各个专业的培养方案和课程学分分布，两者皆为结构化的数据存储格式，因此，可根据该特点来制定相关的内容提取规则。本文将课程名称和先修课程作为最后需要提取的实体内容。

2)数据特征分析：课程大纲的内容主要包括课程名称，课程类型，使用专业，先修课程等，本文选取先修课程为主要特征，对课程实体之间的关系进行抽取。

3)知识图谱绘制：经过上述步骤，课程体系中所有实体和实体间的关系都已经确定，接下来需要将得到的内容采用适当的工具绘制为相应的知识图谱。知识图谱的绘制工具有 **Ucinet**、**SPSS** 等，本文采用的是目前较为流行的图数据库 **Neo4j** 来完成课程体系知识图谱的绘制。采用 **Neo4j** 绘制课程体系的知识图谱，首先要将得到的数据存储 **Neo4j** 图数据库中，这里根据数据类型的不同将数据节点分为不同的颜色，以方便查找节点之间的关系。因为图数据库的特殊性，在存储过程中要特别注意中英文区别和节点以及关系的存在问题。

图2给出了课程体系知识图谱的片段，其中，“计算机组成原理”、“计算机网络”等节点代表相应的课程实体，而课程与课程之间的脉络关系则由图中边的属性可以看出。例如“计算机组成原理”这个课程的先修课程是“计算机电路基础”，而其本身又是“汇编语言与微机接口技术”、“CPU 设计”等课程的先修课程。其他部分的图谱形成与上述情况类似。

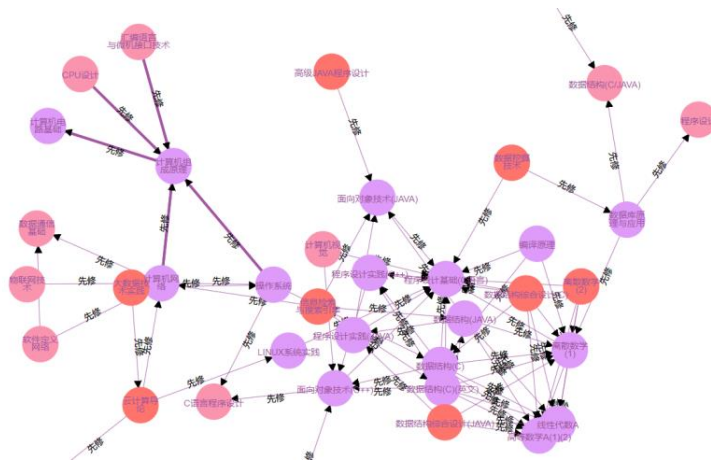


图2 部分课程体系知识图谱

#### 4. 基于课程体系知识图谱的数据分析

本文以北京信息科技大学计算机学院大类招生培养方案和 47 门课程的课程大纲为实验数据，按照前述定义和规则，设计了课程体系知识图谱，其功能主要包括：查询功能，显示功能，纠错功能。其目标用户包括：计算机专业教师和学生、教育行业和知识图谱领域的研究者，该系统用中文显示，适用于熟悉中文的用户。

##### A. 查询功能

查询能力是课程体系知识图谱的关键功能，是保证用户成功查找课程和梳理课程脉络的关键。通过在前期创建课程体系知识图谱，我们得到了一个规范的、完整的课程体系图形数据库，目前，根据已有的课程体系数据库，设计了课程查询功能。

该功能旨在帮助用户快速查找课程及其前驱后继关系，查询可分为查询单独一门课程的前驱后继和查询两门课程之间的脉络关系，通常，在用户输入一门指定课程以后，系统可以在课程体系知识图谱中进行匹配，找到该课程的先修课程或者以该课程为先修课程的课程。当用户输入两门课程的时候，系统可以找出该两门课程之间存在的线性关系。

##### B. 纠错功能

在数据导入的过程中，当课程大纲中出现错误，例如某门课程名称出现录入错误，或该课程的先修课程在课程大纲中并不存在，或者出现其他与文档编写和课程脉络编排有关的问题，系统会提醒用户，并将出现错误的条目列出，在用户进行检查和纠正以后，才能继续执行导入操作，生成新的课程体系知识图谱。

##### C. 显示功能

显示功能是显示当前查询结果，包括课程名称、先修课程、课程类型等信息，利用这些信息可以从中挖掘出课程与课程之间的关系，并据此进行进一步的数据分析，得到更多有价值的信息，这些信息对于知识图谱的使用者具有重要的参考价值。

图 3 和图 4 分别表示指定课程的先修课程、两门指定课程之间的脉络关系。

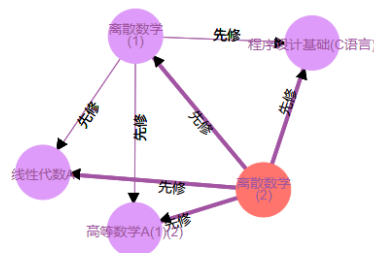


图3 指定课程的先修课程



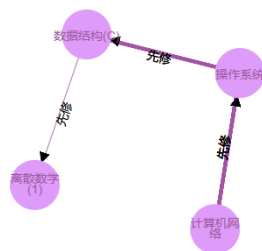


图4 两门指定课程之间的脉络关系

## 5. 结束语

本文针对课程体系显示不够直观、脉络体现不够清晰、人工检查课程大纲编写中的错误费时费力等问题，在基于知识图谱的技术下，通过提取课程大纲的前驱和后继关系构造了基于课程体系的知识图谱，从而使学生和教师对课程体系的脉络结构和错误信息有了条理化的认识和掌握，进而将极大的有助于其整体方面的数据分析和错误检查，为日后课程体系的编排和调整工作提供了新的思路。

本文的工作是我们正在进行的研究工作的一部分，只是对课程的前驱和后继两部分关系进行处理，数据量也不是很大，构建的知识图谱仅限于北京信息科技大学计算机大类课程体系的部分知识图谱，信息不是很全面，未来将课程的其他内容考虑进来并构建更大规模的知识图谱，从而实现课程体系更加全面的关联分析和关系挖掘，并将其应用到更加宽泛的领域。

## 6. 致谢

软件工程“一流专业”建设的研究与实践（项目号2017JGZB01）

## References

- [1] Singhal, Amit. Introducing the Knowledge Graph: Things, Not Strings. Official Blog (of Google). May 16, 2012
- [2] LIU Qiao, LI Yang, DUAN hong, et al. Knowledge Graph Construction Techniques[J]. Computer Research and Development, 2016, 53(3):582-600.(in Chinese).
- [3] WANG Jinli. The visualization analysis of international sport event based on mapping of knowledge[D]. Wuhan: Central China Normal University, 2015(in Chinese).
- [4] WANG Weiwei, WANG Zhigang, PAN Liangming, et al. Research on the construction of bilingual movie knowledge graph[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2016,52(1):25-34(in Chinese).
- [5] ZHAO Jianan. Research on the method of digital resource recommendation based on knowledge map[D]. Beijing University of Posts and Telecommunications, 2015(in Chinese).
- [6] ZHANG Jingru, ZHANG Xinmei, et al. A Knowledge Mapping Analysis on Curriculum System of Recreation Studies in USA, Tourism Science, 2015, 29(1):76-87(in Chinese).