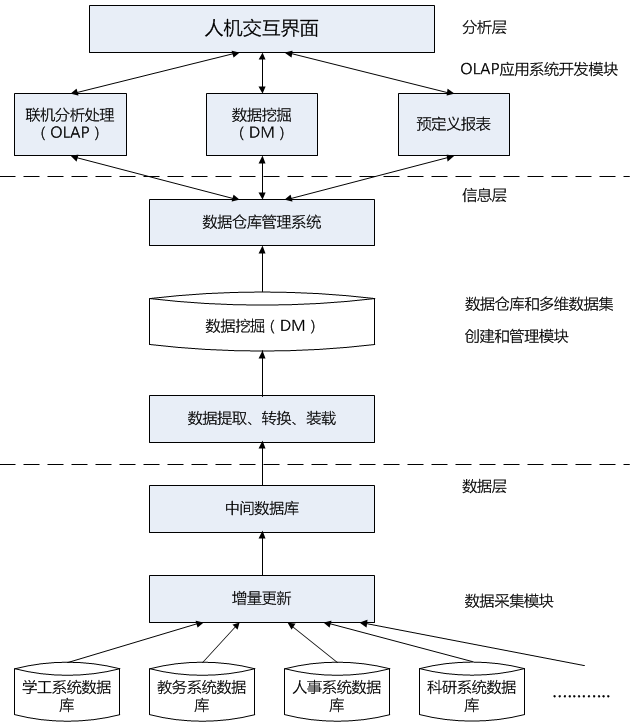
# 技术方案

本章主要阐述数据挖掘分析系统：精华科技数据挖掘分析系统（以下如不明确注明系统将以系统代替。）的技术特征，系统简介,软件配置及系统安全说明。

## 总体技术概述

### 数据挖掘系统关键技术



系统总体架构分为三层：

1. 数据层

数据挖掘管理系统最底层是数据层，数据层是构建数据仓库的基础。数据层包含各种数据源如学工系统、教务系统、人事管理系统、科研管理系统、一卡通等所有学校在用信息系统的数据库。通过建立中间数据库，将OLTP数据库上与数据仓库有关的数据增量更新到中间数据库，使对OLTP数据库的效率影响降低到最低。

1. 信息层

为了解决教学决策及评估等主题中需求目标难以量化的特点，信息层作为数据访问层和源数据层的桥梁，对数据进行处理，形成不同层次的信息，以便分析层在此基础上建模。信息层的核心就是与各主题相关的数据仓库。数据仓库通过对数据源的原始数据进行收集、分类、筛选和抽取形成能够对决策分析提供支持的多重粒度级的数据库，实现对数据源的统一管理和信息转换。同时为分析层的各种应用提供接口，构造应用于主题分析的更高粒度级的数据集市，或者直接为数据访问层提供数据支持。信息层通过管理工具对数据仓库进行管理和维护。以中间数据库为数据源，按分析主题建立数据仓库。最后将数据仓库中的数据按照一定方式聚合汇总，构成信息分析的多维视图，以一定的存储模式（如模式文件）将这些多维视图存储在OLAP服务器中。

1. 分析层

分析层由通用分析平台和基于主题的应用分析系统组成。通用分析平台是为了适应决策需求的不稳定的特点，提供一些基本数据挖掘手段和简单分析模型，这样在很多情况下不必专门的建立应用分析系统，大大提高了系统的普遍适用性和灵活性。针对复杂且相对稳定的决策需求，需要从数据仓库抽取数据并建立专题数据集市或是利用已有的数据集市，开发专用的分析系统。联机分析处理模块实现对多维数据集中立方体信息、维度信息、层次信息、度量信息的显示；实现钻取、切片、切块，旋转等多维分析操作，提供多种查询方式，查询结果以WEB页面的形式展现，以实现B/S形式的联机分析，全面支持决策分析。

三层结构化的体系结构适应决策挖掘系统需求的特点。三层之间通过数据仓库衔接，有效实现了数据驱动，并且任何一层的修改、添加或删除功能模块给整体系统的影响都很小。

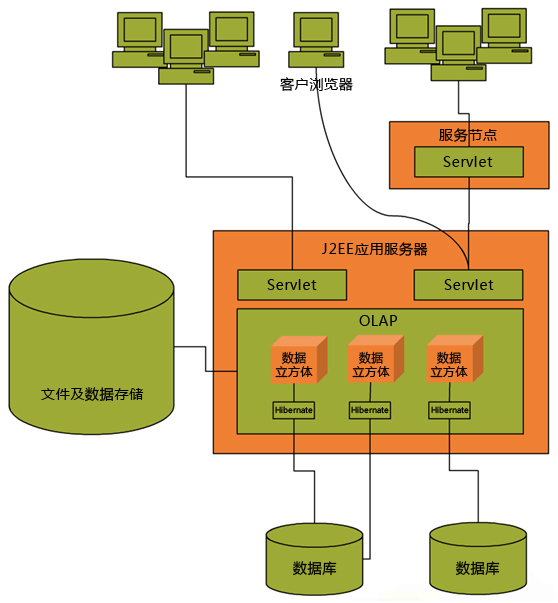
#### 1.1.1.1数据仓库技术

数据仓库中的数据时进行决策分析的基础，因此构造数据仓库是构建系统的首要任务。在数据仓库中，由于注重的是查询而不是事务，因此数据是按照决策分析的学生主题、教师主题、科研主题、一卡通主题、资产主题、设备主题等来组织。为了使数据仓库中的数据是面向对象的，必须将数据按要求的主题进行收集、归纳后重新存放；由于需要能进行多视角观察、分析数据，因此，对应各主题的数据必须按一定得属性进行组织；另外，针对数据的综合程度不同，应该对数据按不同的汇总程度存放。因此，数据仓库采用多维数据模型对各主题的数据进行组织。对应以上数据组织的要求，数据仓库定义了相应的结构元素来实现数据的结构模型：事实、维、粒度等。

由于数据仓库中的数据是静态的，很少对其进行插入或更新操作，而且用数据仓库中的数据创建立方体十分耗费时间，所以要建立较多的索引以加快信息检索和更新立方体的速度。

抽取与加载数据：数据仓库建立之后，就可以从信息系统中加载数据，这需要将多个数据转移到数据仓库。由于不同数据源的数据存在很大的异构性，因此在数据加载过程中，需要完成数据库验证，数据迁移，数据净化和数据转换等使数据一致，集成的操作。这一过程将通过ETL工具ODI来完成。

#### 1.1.1.2联机分析技术OLAP



联机分析处理OLAP是针对特定问题的联机数据访问。分析人员、管理人员通过对多维数据的很多可能的观察形势进行快速、稳定一致和交互性的存取，允许管理决策人员对数据进行深入观察。

1. 维

维是人们观察数据的特定角度，是考虑问题时的一类树形，树形集合构成一个维。例如，教学质量监督部门关系教学质量随时间推移而产生的变化情况，此时时间就是一个维，即时间维。

1. 维的层次

人们观察数据的某个特定角度还可以存在细节程度不同的多个描述方面，我们称这多个描述方面为维的层次。例如描述时间维，可以从学期、学年等不同的层次来描述，那么学期、学年等就是时间维的层次。

1. 维成员

维的一个取值。例如，学生对象这个维，这个维有院系、专业、年级、学生类型这四个层次，分别在四个层次上取一个值组合起来，就得到了开课对象维的一个成员，即“信息科学与技术学院计算机应用技术专业2005级硕士生”

1. 多维数组

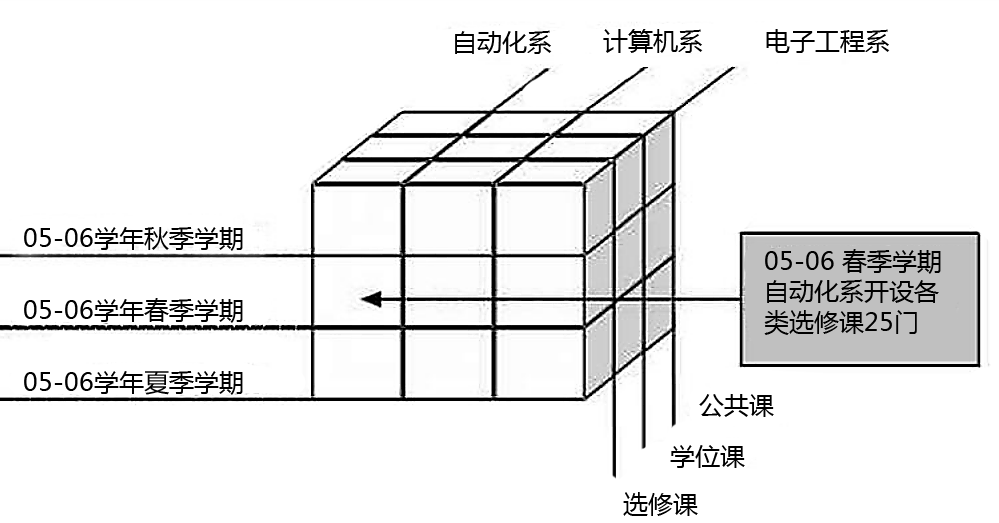
维1，维2，……，维n，变量。例如按照课程编号、院系、专业、年级组织起来的思维结构，加上学分，就组成了一个多维数组（课程编号，对象院系、对象专业、对象年级、学分）。

1. 数据单元

多维数据的取值。当多维数组的各个维都选中一个维成员，这些维成员的组合就确定了唯一的变量值。

1. OLAP立方

立方代表数据，立方的维代表事实表中的维。典型的数据立方体如图所示。



1. 切片、

在多维数据的各维上分别选定某个对应维成员的操作称为切片操作，即在确定每个维的成员后，所得的多维数组的子集称为是对其的一个切片。欺骗操作如下图所示。



1. 旋转

改变一个报表或页面显示的维方向的操作。如可以交换行和列，或把某一行维移到列维去，或把页面显示中的一个维和页面外的维进行交换。旋转操作如图所示。



1. 钻取

钻取操作是使用户在数据仓库的多层数据中能通过导航信息而获得更多的细节数据。通常包括下钻和上翻两种操作。

**信息技术与工程学院开课情况 信息技术与工程学院开课情况**

|  |  |
| --- | --- |
| 系所 | 05-06学年秋/春季学期 |
| 自动化系 | 6门/6门 |
| 计算机系 | 4门/6门 |
| 电工系 | 5门/3门 |

|  |  |
| --- | --- |
| 系所 | 05-06学年 |
| 自动化系 | 12门 |
| 计算机系 | 10门 |
| 电工系 | 8门 |

下钻操作

联机分析处理可以快速的、多维、更高层次、全局的角度的分析来展示全校全方位的数据。

#### 1.1.1.3数据挖掘技术

数据挖掘技术（DM）就是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们视线不知道的、但是又是潜在的有用信息和知识的过程。

由于学校现有业务系统大多为结构化的数据，数据挖掘技术方法主要采用关联分析、聚类分析、分类、预测、时序模式和偏差分析等。

1. 关联分析（AssociationAnalysis）

两个或两个以上数据项的取值之间存在某种规律性，就称为关联，可以建立起这些数据项的关联规则。数据关联是数据库中存在的一类重要的、可被发现的知识，它反映一个事件和其他事件之间的依赖和关联。如果两项或多项树形之间存在关联，那么其中一项的属性值就可以依据其他属性值进行预测。

1. 聚类分析（Clustering）

聚类是把数据按照它们的相似性归纳成若干类别，同一类别中的数据距离较小、彼此相似，不同类别中的数据距离偏大、彼此相异聚类分析可以建立宏观的概念，发现数据的分布模式，以及可能的数据树形之间的相关关系。在聚类分析法中可以给定距离阈值，各样本按阈值进行聚类。

1. 分类（Classification）

分类是数据挖掘中应用最多的任务。分类就是找出一个类别的感念描述，并用这种描述来构造模型。类别的概念描述代表着这类数据的整体信息，也就是该类的内涵描述。类的内涵描述分为：特征描述和辨别性描述。特征描述是对类中对象的共同特征的描述。辨别性描述是对两个或多个类之间的区别的描述。

1. 预测（Predication）

预测是利用历史数据找出变化规律，建立模型，并由此模型对未来数据的种类及特征进行预测，典型的预算方法是回归分析，即利用大量的历史数据，以时间为变量建立线性或非线性回归方程。预测时，只要输入任意的时间值，通过回归方程就可求出该时间的状态。

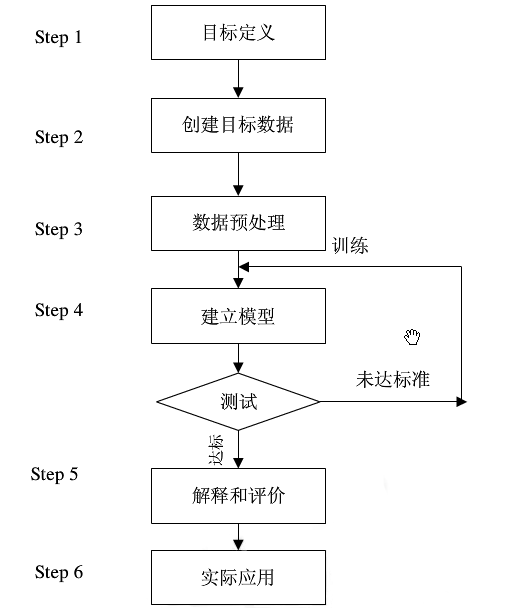
1. 时序模式（Time-SeriesPattern）

时序模式是指通过时间序列搜索出的重复发生概率较高的模式。与回归一样，它也是用一致的数据预测未来的值，但这些数据的区别是变量所处的时间不同。在时序模式中，需要找出在某个最小时间内出现比率一直高于某一最小百分比的规则。这些规则会随着形势的变化作适应的调整。时序模式中，一个有重要影响的方法是“相似时序”。用“相似时序”的方法，要按时间顺序查看时间事件数据库，从中找出另一个或多个相似的时序事件。

1. 偏差分析（Deviation）

数据库中的数据存在很多异常情况，发现数据库中数据存在的异常情况是非常重要的。偏差包括很多潜在的知识，如分类中的反常实例、不满足规则的特例、观察结果与模型预测值的偏差、量值随时间的变化等。偏差检测的基本方法是，寻找观察结果与参照值之间有意义的差别。

数据挖掘的基本步骤可以粗略分为问题定义，数据选取和处理，数据挖掘，结果的解释和评估以及最后实际应用。如下图



1. 目标定义

数据挖掘是为了在大量数据中发现有用的令人感兴趣的信息。在问题定义过程中，数据挖掘人员和最终用户紧密协作，一方面明确实际工作对数据挖掘的要求；另一方面通过对各种学习算法的对比，今儿确定可用的学习算法。后续的学习算法选择和数据集准备都是在此基础上进行的。

1. 创建目标数据

在一个或多个知识发现工具的帮助下，选择一组要进行分析的初始数据。

1. 数据预处理

数据预处理是数据挖掘过程中的一个重要步骤，所花的时间和精力要占整个数据挖掘项目的50%-90%，尤其是在对包含有噪声、不完整，甚至是不一致数据进行数据挖掘时，更需要进行数据的预处理，以提高数据挖掘对象的质量，并最终达到提高挖掘所获模式知识的质量。

数据预处理分为4个步骤：数据清洗（Data Cleaning）、数据集成（DataIntegration）、数据转换（Data Transformation）、数据消减（Data Reduction）。

1. 建立模型

通过应用一个或多个数据挖掘算法来创建表示数据的最佳模型，数据挖掘算法执行阶段首先根据对问题的定义明确挖掘的任务或目的。确定挖掘任务后选择算法，进而建立模型，在多个可供选择的模型中找出最佳模型。

1. 结果解释和评估

数据挖掘阶段发现出来模式，经过评估可能存在冗余或无关的模式，这时需要将其剔除，也有可能模式下不满足用户要求，这时则需要回退到发现过程的前面阶段，如重新选取数据，采用新的数据变换方法，设定新的参数值，甚至换一种挖掘算法等。另外，数据挖掘由于最终是面向人类用户的，因此可能要对发现的模式进行可视化，或者把结果转换为用户易懂的另一种表示，如把分类决策树转换为“IF…THEN…ELSE …THEN…”规则。

1. 实际应用

如果发现的知识被认为是有用的，根据模型的解释和评估结果对实际要解决的问题进行实践指导。

#### 1.1.1.4高速检索技术

由于数据仓库中的数据非常大，而且分析性查询操作也比较复杂，因此，提高数据仓库中数据的访问速度称为关键要求和必然之举。系统采用基于数据仓库的索引技术、内存管理和并行控制来保证数据检索的快速便捷。

我们主要采用索引技术方法来解决数据快速检索的问题。

1. B树索引

B树索引是一种动态调节的平衡树，较适合字段值分散且重复值少的字段。一个B树索引中包含一个由高层节点和相继低层结点组成的层次结构。B树索引有两类结点：一是分支节点；二是叶子节点。B树索引具有间接性、易维护性以及高速检索的特性。

1. 位图索引

位图索引是一组0和1组成的字符，在表的任何一列中，每一种可能存在的值都有它唯一的位图。表中的每一条记录都会分配一个位图中的相对位置，用来表明所考虑的记录是否出现了不同的值。下面通过一个例子来说明：

|  |  |
| --- | --- |
| 男 | 女 |
| 0 | 1 |
| 1 | 0 |
| 1 | 0 |
| 0 | 1 |

表1 学生信息表 表2 位图索引值表

|  |  |
| --- | --- |
| 姓名 | 性别 |
| 莉莉 | 女 |
| 张三 | 男 |
| 李四 | 男 |
| 王燕 | 女 |

1. 链接索引

链接索引的索引项是链接的并且满足链接要求的表的元组标志，它的每一个元组包括所有要链接的表的元组标志。链接索引的显著特点是它能提高多表链接的性能，它能够把所有符合链接条件的链接记录下来。

索引的优势：

1. 建立索引可以在很大程度上提高信息检索的效率。
2. 创建索引后，可以使索引和信息形成一对一的映射式对应关系，从而使数据具有唯一性。
3. 能够提高表与表之间的链接速度，提高了数据的参考完整性。
4. 可以减少检索过程中的分组及排序时间。
5. 使用优化隐藏器，提高整个信息检索挖掘的性能。

#### 1.1.1.5数据集成交换技术

1. **集成中心库模式**

模式以教育部《高等学校管理信息标准CELTS-33》为基础，结合精华教育过去多年集成经验进行设计。从总体上对高校数据进行梳理，在高校信息管理的数据层面按照面向对象进行设计。

高校往往将数据信息按照不同角度不同侧重点，分散在各个业务系统中进行管理，而集成中心库模式则将这些数据信息抽离出来，进行对象和业务类的拆分，将其分为对象类和活动类。

对象类是指数据对象，主要包括实体对象和逻辑对象。实体对象是指传统的数据对象，如人、活动等；而逻辑对象则是指一些非传统的数据对象，如院系信息等。

活动类包括对象管理活动和综合管理活动，可以按照对象的生命周期进行管理活动的设定。其中对象管理活动指基于对象的管理部分活动可以进行清晰的划分和归纳的数据，比如一些对实体对象的管理，；综合管理活动是指包含的内容和相关联系会随着应用、定位以及数据使用的角度不同而变化的活动。

1. **参照模型标准及数据**

参照标准集是一系列机构标准的集合，主要包括国家标准、教育部行业标准等。

1. **执行标准模型及数据**

执行标准集是学校自定义标准的集合，俗称“校标”，是学校信息化建设中实际使用的代码标准集。一般情况下，执行标准是高校参考了参照标准，再结合学校自身的需求而制定的。

1. **集成接口支持**

数据集成平台为多种数据源提供接口支持，如下：

* 支持主流关系型数据库，包括Oracle、IBM DB2 UDB、IBM DB2/400、Informix、Microsoft SQL Server、Sybase AS Enterprise、Sybase AS Anywhere。
* 支持非主流关系型数据库，包括MySQL、Derby、Hypersonic SQL、PostgreSQL。
* 支持ODBC数据源类型的接入，包括Microsoft Access、Microsoft Excel、Dbase、Visual Foxpro。
* 支持主题或者队列，包括JMS Topic、JMS Queue。
* 支持WebService
* 支持Tabled-Txt文件
* 支持XML文件
* 支持操作系统的网络协议，包括 FTP。

1. **数据集成KM模块**

数据集成平台提供数据集成KM库，定义了各种场景通用的知识模块（KM），包括了各种数据集成需求开发包共计100多项，大大减少了数据的集成的工作量。

1. **集成设计工具**

集成设计工具是设计阶段开发人员和元数据管理员的主要工具，系统所有项目的开发都是通过这个界面完成的。

用户使用该工具，为数据转换和数据完整性定义声明规则，然后将业务系统中数据库中的元数据导入或定义，为实际的生产环境生成方案。

1. **集成查看工具**

集成查看工具是运行阶段的主要工具，开发人员在设计阶段也可使用该工具进行调试。

用户使用该工具查看执行日志，包括错误数、被处理的行数、执行的统计信息和被执行的实际代码等等，实现对生产环境的管理以及ODI监控。

1. **集成调度工具**

用户通过集成调度工具实现对各个数据集成同步任务的调度控制，以此完成定制化的数据集成过程。

### J2EE开发平台

为了保证系统的高可用性、高可靠性和可扩展性，系统必须要选择支持强大的企业级计算的成熟的技术路线。当前能够满足这些条件的计算平台首推J2EE的企业标准。

J2EE是主流的技术体系，J2EE已成为一个工业标准，围绕着J2EE有众多的厂家和产品，其中不乏优秀的软件产品，合理集成以J2EE为标准的软件产品构建系统，可以得到较好的稳定性、高可靠性和扩展性。J2EE技术的基础是JAVA语言，JAVA语言的与平台无关性，保证了基于J2EE平台开发的应用系统和支撑环境可以跨平台运行。

### Webservice接口技术

对于各主题统计的内容，系统支持SOA的集成架构，应用组件中已经广泛采用了先进的WebService接口。

开发路线和设计方案详细如下：

1. **构建服务**

基于J2EE，采用SCA（Service Component Architecture服务组件架构）框架，主要通过WSDL(Web Service Description Language)开发和描述服务接口，采用松耦合的形式构建数据仓库上各部分数据的访问服务以及各类应用的服务组件，集中对外提供基于Web Service的服务调用。

系统将服务接口、实现、部署、调用完全分离，通过配置的形式灵活的组装，绑定，既能够对各类统计服务的访问接口分散开发，同时又可以集中控制服务访问口径，实现统计服务访问审计。

1. **实现服务通信**

各服务颗粒之间，以及服务与后端应用程序之间能够可靠地相互通信，在服务的调用方面支持多种传输协议以及多种通信模型，传输协议包括HTTP, HTTPS, JMS等协议，通信模型则提供同步和异步通信。该通信过程可以被ESB企业服务总线捕获，能够支持ESB集成方式。

1. **确保服务安全**

能够持久有效的保护和管理服务，提供对服务调用的安全、审核和日志管理。

### 面向服务的体系结构（SOA）

面向服务的体系结构（SOA）是一个组件模型，它将应用程序的不同功能单元（称为服务）通过这些服务之间良好的接口和契约联系起来。接口是采用中立的方式进行定义的，它应该独立于实现服务的硬件平台、操作系统和编程语言。这使得构建在各种各样的系统中的服务可以以一种统一和通用的方式进行交互。

SOA本身就是一种面向企业级服务的系统架构，简单来说，SOA就是一种进行系统开发的新的体系结构，在基于SOA架构的系统中，具体应用程序的功能是由一些松耦合并且具有统一接口定义方式的组件（也就是Service）组合构建起来的。因此，基于SOA的架构也一定是从企业的具体需求开始构建的。但是，SOA和其它企业架构的不同之处就在于SOA提供的业务灵活性。业务灵活性是指企业能对业务变更快速和有效地进行响应，并且利用业务变更来得到竞争优势的能力。创建一个业务灵活的架构意味着创建一个可以满足当前还未知的业务需求的信息技术体系架构

### 开发规范及设计原则

系统采用基于Ｂ/S架构体系设计，将数据资源汇集到资源池中，实现本校的软硬件资源的整合利用，可获取几乎无限可扩展的计算能力，支持按需自助服务，支持应用的横向(多台服务器的多个应用集群)、纵向(即单台服务器的多个应用的集成)两种集群布署方式，能自动调整全校资源来帮助应用更好地应对突发流量，实现统一的、绿色、高效能数据立方体。

遵循“可靠性、实用性、安全性、先进性、开放性、智能化、主动服务理念”的设计原则，集中体现在“结构优化、功能细化、操作简化、界面美化”。

1. **技术标准规范**

系统将严格按照各级有关标准执行，适应高校管理各类数据的处理、保存和交换的需要。遵循如下通用标准与规范：教育管理信息化标准：国家标准、行业(部)标准、学校自定义标准。

* 信息规范采用《国家经济信息系统设计与应用标准化规范》、《标准化工作导则-信息分类编码规定》（国标GB7026-86）和教育部公布的《高等学校管理基本信息集》（教信息[1994]1号）等；系统的建设遵循各项国家技术标准和相关行业标准，包括技术发展的标准、流程的标准、系统的标准和代码（数据）的标准化。保证系统内外的有机整合和信息的一致性。
* 需求调研分析过程的标准化：采用标准的需求调研分析工具，如UC矩阵数据调研分析工具、软件需求说明编制工具等；
* 设计与开发过程标准化：设计与开发按照《GB-T8567-2006计算机软件文档编制规范》，采用面向对象的系统设计方法，遵循UML设计规范即采用StartUML等；

1. Axure RP Pro 5

Axure RP 能帮助网站需求设计者，快捷而简便的创建 基于目录组织的原型文档、功能说明、交互界面以及带注释的wireframe网页，并可自动生成用于演示的网页文件和word文档，以提供演示与开发。

1. StarUML

可绘制9款UML图：用例图、类图、序列图、状态图、活动图、通信图、模块图、部署图以及复合结构图等。语法检验：StarUML遵守UML的语法规则，不支持违反语法的动作。正反向工程：StarUML可以依据类图的内容生成Java 、C ++、C#代码，也能够读取Java、C++、C#代码反向生成类图。

* 系统组件标准化：数据访问服务、信息门户、身份管理、身份安全审计、等组件均采用相应的国际标准和国家标准，可以与其它符合规范的第三方平台与应用实现对接；提供java/.net/c等主流语言接口。
* 系统接口标准化：数据访问接口、应用集成接口、身份认证接口均采用相应的国际标准和国家标准，方便第三方应用的接入；提供java/.net/c等主流语言接口。
* 运行环境标准化：系统平台支持Linux、Solaris等类Unix平台、Windows平台，支持WebSphere、Weblogic、Tomcat、Jboss等多种应用服务器，支持Oracle、mySQL、Sqlserver、Sybase等数据库系统，充分适应和兼容各类主流运行环境；
* 信息安全标准化：安全策略、密码与安全设备选用、网络互联、安全管理等设计符合我国信息安全法律法规。密码不可逆，默认用SSHA加密。

1. **先进性和成熟性**

系统采用代表信息化系统发展趋势的主流和成熟技术进行系统架构设计，选用技术先进、产品成熟的产品建设信息共享系统，保证系统的技术先进性和成熟性，采用J2EE技术框架，符合国家电子政务技术框架中对于分布式部署的要求。采用业界先进SOA技术框架，提高开发效率，降低运作、维护成本，增强系统的核心能力以及日后功能扩展性。

1. **安全策略与数据加密**

系统采用各种应用安全策略：用户与权限管理、统一身份认证、访问控制、管理控制、版本控制、数据关联控制、数据加密、数据存储、数据备份与恢复、日志与安全审计等，以确保系统安全及数据安全。信息安全是数据挖掘系统建设的前提，系统设计中首要考虑如何建立整个系统的安全性、保密性，而且这种考虑必须是整体的、全面的。系统必须要提供信息传输保密性、数据完整性、身份识别和认证、防抵赖性等安全保障措施。系统必须在设计上保护用户身份的安全、实现功能和数据权限、身份信息的安全传递、数据的加密和签名；系统应提供组织机构、权限管理组件提供全面的功能权限、数据权限、用户安全认证、安全审计功能，适应多部门、多行政管理方式等复杂的用户和权限管理。

1. **开放性、可扩展性**

系统具有良好的可扩展性，通过对软件的重新配置、改进，以满足学校发展的各类新需求。对于用户的个性化特殊需求，提供各类开放功能接口，支持二次开发功能，使系统具有可扩充能力。

1. **跨平台性**

软件系统采用JAVA 开发技术和N 层应用体系结构，因此，数据库服务器系统、Web 服务器系统和应用服务器系统可以运行于包括Windows、Unix 、Linux、MacOS等多种系列操作系统平台上，系统具有良好的可移植性、适应性。系统选择开放的应用平台，采用J2EE技术标准和XML数据交换格式，J2EE+XML实现了与平台和数据的无关性。跨平台技术开发，实现系统与操作系统、硬件平台等无关，便于今后系统移植。

1. **高易用性**

操作具有流程化界面、向导式操作、个性化风格。系统操作流程清晰，操作简单，能够准确反映每一项工作的流程、顺序、步骤，方便使用人员轻松掌握相应系统功能、快速完成相应管理工作。

1. **高效性、高速缓存**

系统的登录访问、数据的查询与检索要满足高效性原则。支持大量用户的并发访问，系统响应时间不随并发用户数量的增加而降低。支持大量数据的快速查询与检索，保证检索响应时间在秒级内。为加强多用户同时访问系统时的系统反应能力，基于高速缓存技术，有效提高系统的运行性能问题。

1. **易维护、升级**

系统采用优秀的SSH（Struts+Spring+Hibernate）框架，将各层之间解耦，保证系统的低维护、升级成本，系统提供平台升级，软件系统功能修改完善以及版本升级。系统提供管理工具可以对系统进行全面的监控和管理、配置，并能够帮助管理员对系统故障进行诊断、排错和分析与规划，方便管理，降低总体使用成本。

1. **高可用性和可靠性**

为保证数据挖掘系统的高效工作，系统在稳定正常运行的同时还提供较高的性能，从而能够高效率地处理各类关键事务，充分考虑如何保证系统的高可用性、可靠性和高效性，确保7\*24小时系统正常运行。

1. **系统集成**

学校已建设的部分信息系统已经有效与数字化校园基础平台整合集成，对于没有或未完全集成的系统，我们认为本期项目的集成内容主要在以下三个层面进行：

* 数据集成

利用数据中心平台，从应用系统数据库中抽取出需要共享的数据，使数据中心平台成为全校范围内唯一的全面的数据源，完成数据层的集成，同时为相关应用系统提供共享数据访问服务，为在全校范围内进行综合数据分析服务提供完备、有效、可信的数据基础；

* 认证集成

通过现有统一的认证平台，完成目前适合集成认证的系统的集成工作，同时留有接口供后续建设的系统无缝集成；

* 应用发布集成

以数据中心平台中的数据为基础，通过数据挖掘系统为主管部门提供综合高价值统计分析，为主管部门的决策分析提供准确可靠的统计功能。统计应用展现符合JSR-168、Portal等一系列的技术标准。

1. **关于需求变更**

由于软件系统需求的不确定性，在软件开发和实施过程中，在建设过程中，实际需求与标书内容产生10%以内的偏差，我们愿意无条件免费响应。

1. **保密协议**

涉及校方的系统数据不得以任何形式提供给第三方。

### 系统特点

1、多维度、全方位、丰富的统计展示及各式各样的表现形式（饼图、线图、环图、直方图、星型图等）。基于学校现有的数字校园系统所产生的业务数据的基础上，从学生基本情况、教职工基本情况、科研、资产、设备、校园一卡通、招生、迎新等各个维度、各个维度，对数据进行深度挖掘和整理，形成高质量的数据以及直观的数据展现形式，并通过对学校一些基本信息的数据挖掘，用来统计、分析学校的一些基本信息，实现办学思路与领导作用分析、师资队伍分析、教学条件分析、专业课程设置分析、教学质量分析。



2、提供多种流行语言（java、.net/C等）的统一身份认证接口与学校现有的统一身份认证系统整合。

3、丰富的查询方式，系统支持表头查询，高级查询，单字段查询.

4、统计数据的实时性，一旦业务数据发生了更新，则统计结果也要实时更新。功能强大的ETL工具ODI满足数据变化的实时更新与同步，从而保证统计结果实时反映业务系统中的变化。

5、纯B/S架构，支持7×24不间断服务。系统正常运行的同时还要提供较高的性能，从而能够高效率地处理各类关键事务，应该充分考虑如何保证系统的高可用性、可靠性和高效性，确保7\*24小时系统正常运行。

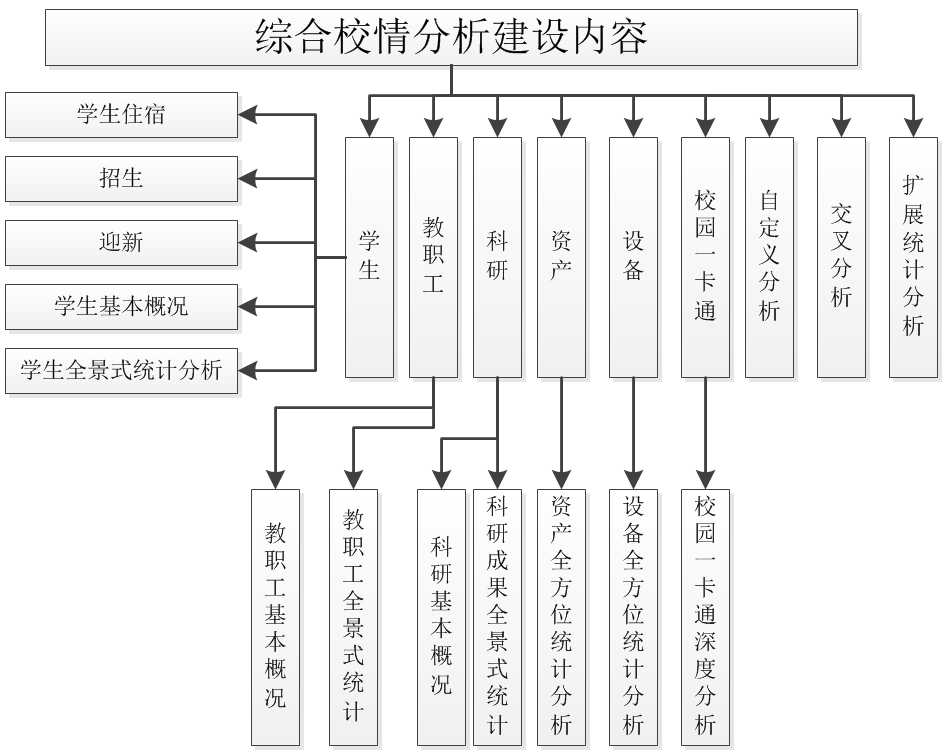
6、产品完全自主研发。

7、系统建设遵循各项国家技术标准和相关行业标准，包括技术发展的标准、流程的标准、系统的标准和代码（数据）的标准化。保证系统内外的有机整合和信息的一致性。在学校现有数据平台中的统一数据模型基础之上进行构建，以保证系统的执行效率和无缝集成。

8.支持各种文件形式(excel/pdf/word等)的导出功能。

## 综合校情分析系统建设内容

综合校情分析系统建设内容概览图:

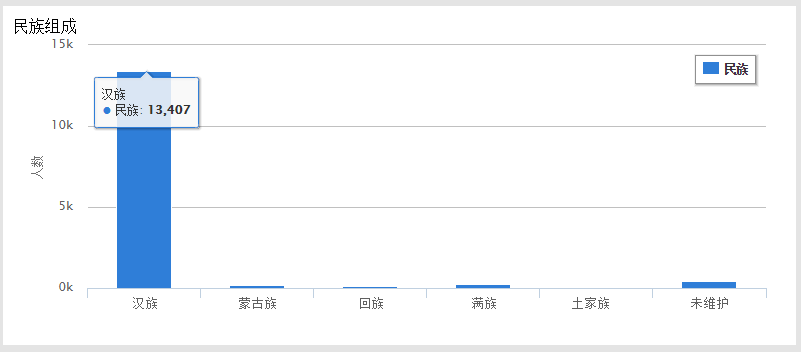


**学生基本情况统计**

学生基本情况统计支持各种图形、表格形式，以专业、年龄、职称、性别、民族、政治面貌、学历、学位、教职工来源等不同维度展示教职工人数分布、占比等统计信息。系统会对全校学生进行15个方面180多个字段进行统计分析，产出相应的高价值统计应用。

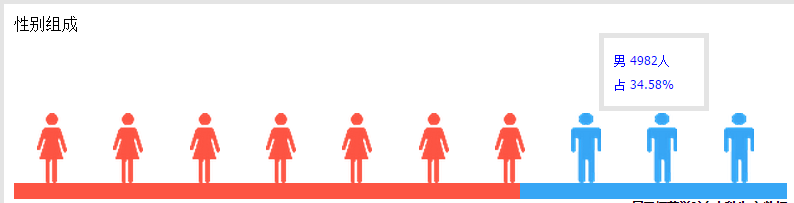
**在校生民族分布概况及比例统计**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生民族分布占比及比例；亦可以以汉族与少数民族进行分组统计学生的民族分布及其比例。
2. 系统将给出相同层级之间的横向对比统计。例如：各院系或专业间汉族与少数民族比例的横向对比，以从侧面反映专业与民族之间的相关性。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护民族的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



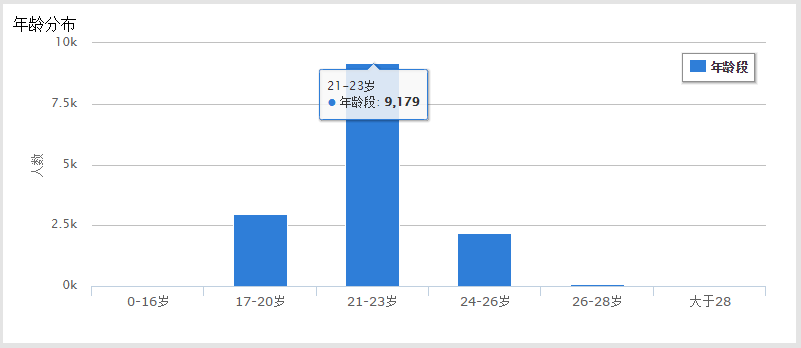
**在校生性别分布概况及比例统计**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的性别分布及比例。
2. 系统将给出相同层级之间的横向对比统计。例如：各院系或专业间男女比例的横向对比，以从侧面反映专业与性别之间的相关性。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护性别的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



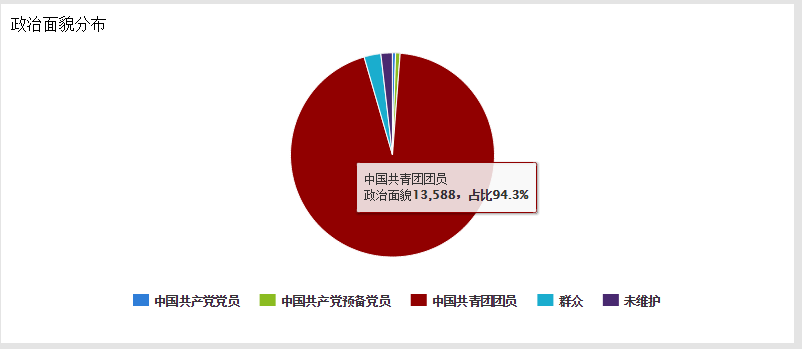
**在校生年龄段分布统计**

* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的年龄段分布及占比。
  2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
  3. 对于没有维护年龄段的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



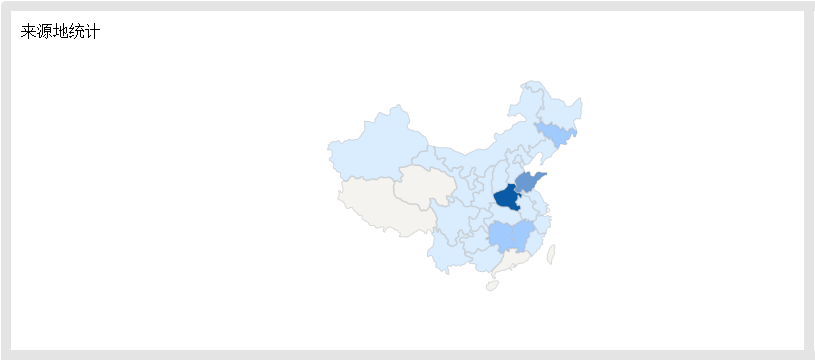
**在校生政治面貌分布及比例统计**

* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的政治面貌分布及占比。
  2. 系统将给出相同层级之间的横向对比统计。例如：各院系或专业间党员、预备党员、团员等占比的横向对比。
  3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
  4. 对于没有维护政治面貌的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



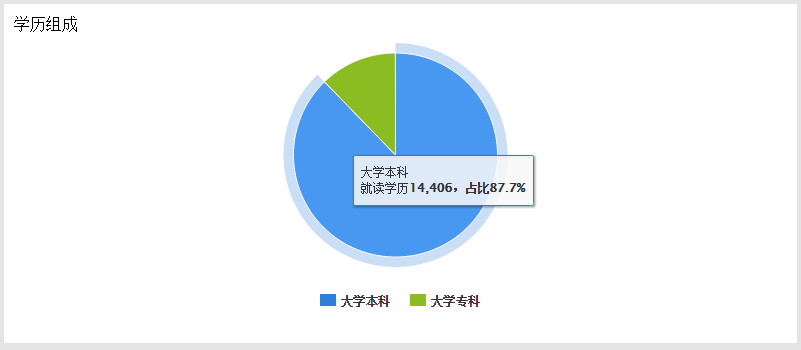
**在校生学生来源分布及比例统计**

* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的学生来源分布及占比。
  2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
  3. 对于没有维护学生来源的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



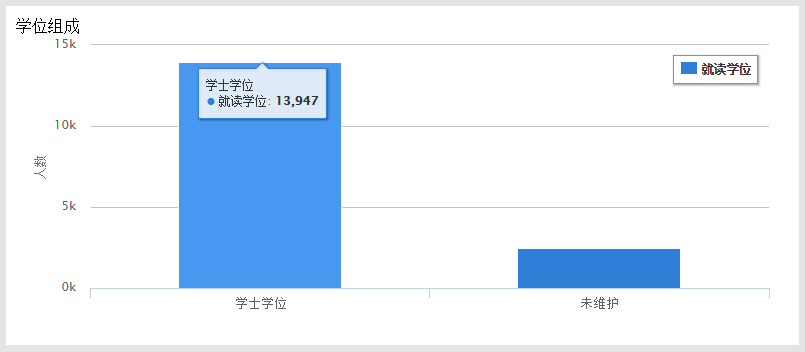
**在校生学历分布及比例统计**

* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的学历分布及占比。
  2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
  3. 对于没有维护学历的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



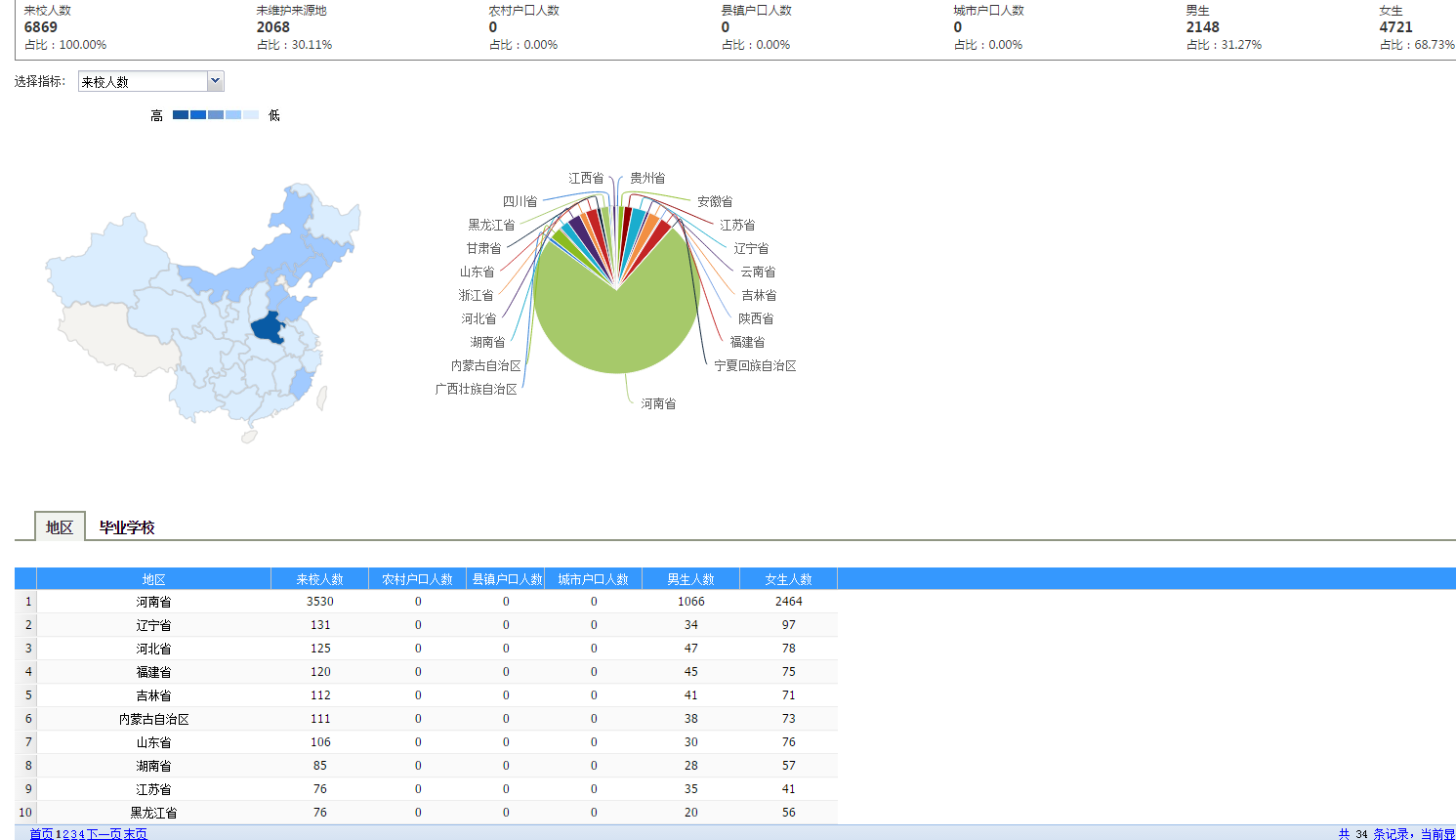
**在校生学位分布及比例统计**

* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的学位分布及占比。
  2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
  3. 对于没有维护学位的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



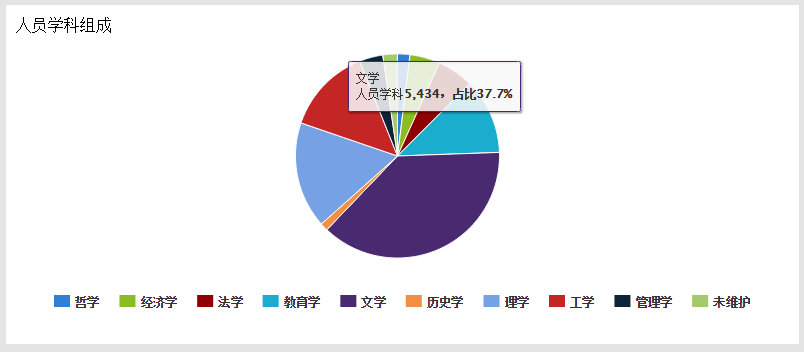
**在校生生源地分布及比例统计**

* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的**生源地**分布及占比。
  2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
  3. 对于没有维护**生源地**的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



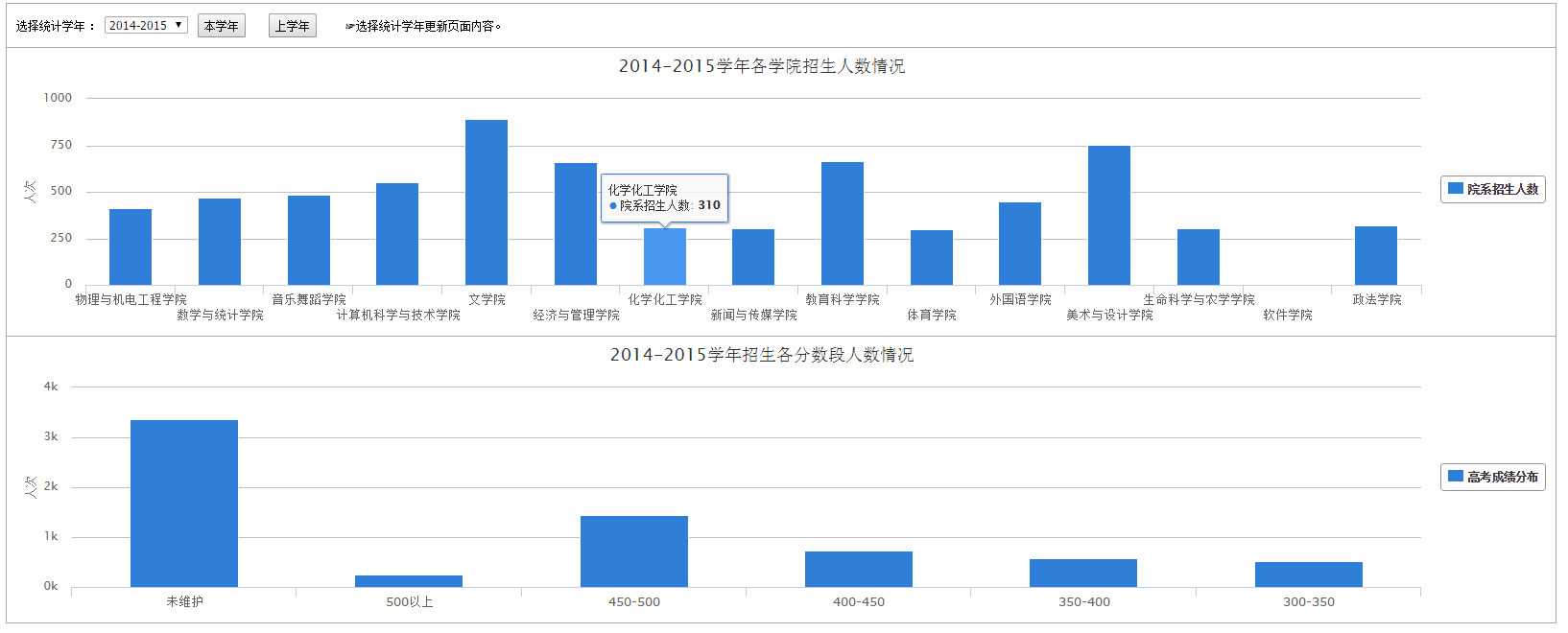
**在校生科类分布及比例统计**

* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的**科类**分布及占比。
  2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
  3. 对于没有维护**科类**的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



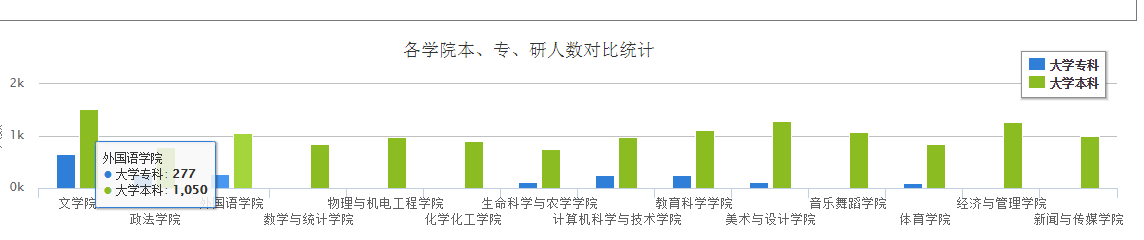
**在校生招生类别分布及比例统计**

* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的**招生类别**分布及占比。
  2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
  3. 对于没有维护**招生类别**的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



**在校生人数分布概况统计**

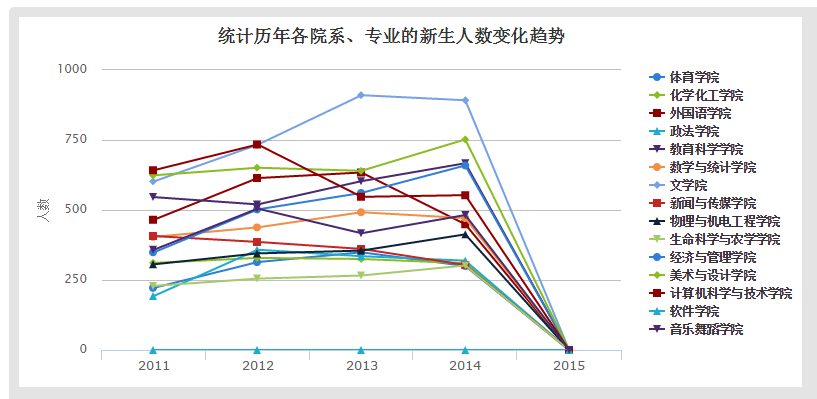
* 1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示本科生、专科生、研究生的**人数**分布及占比。
  2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



**学生全景式统计分析**

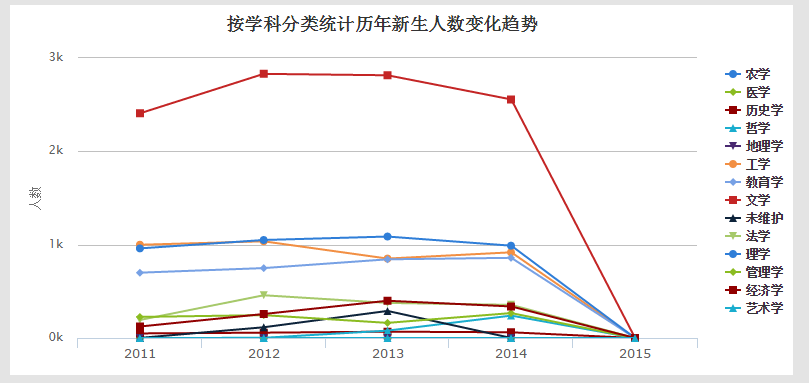
**历年各专业学生数量**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示学生的**历年人数及趋势曲线**。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



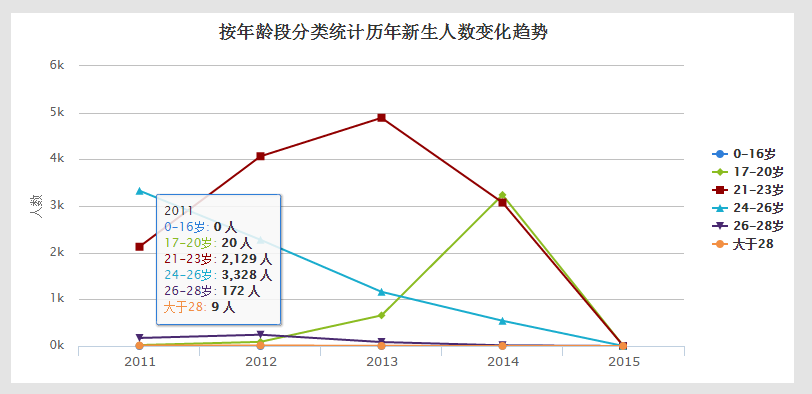
**历年各学科学生数量**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**历年各学科**的学生人数**及变化曲线**。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



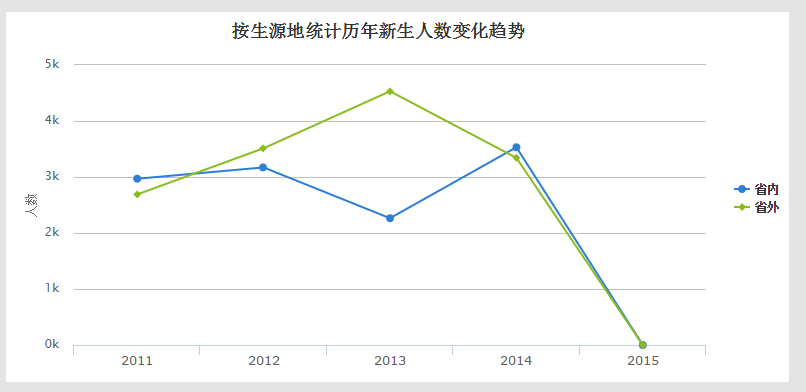
**历年学生年龄分布**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**历年各年龄段**的学生人数**及变化曲线**。年龄段以0-16、17-20,21-23、24-26、26-28、28以上进行分组。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



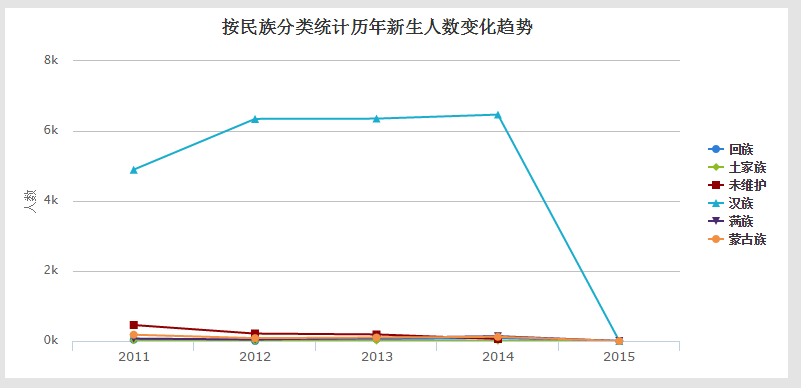
**历年学生来源地区分布**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**历年各生源地**的学生人数**及变化曲线**。生源地可以按照省内、省外进行分组，亦可以按照地市进行分组。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



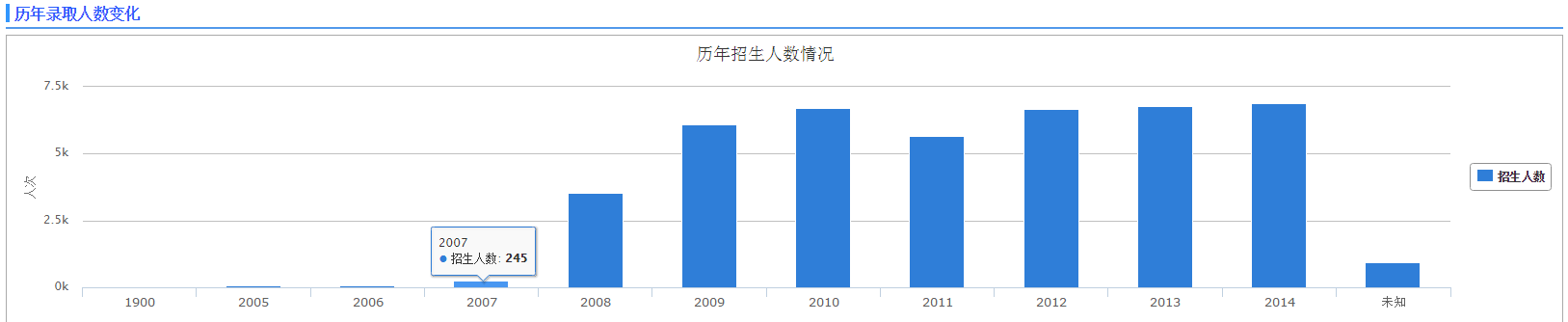
**历年学生民族分布**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**历年各民族**的学生人数**及变化曲线**。民族可以按照汉族和少数民族进行分组统计，亦可以按照各具体民族进行统计。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



**历年招生情况分析**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，时间为维度按照图形或表格的形式分别统计展示**历年各单位招生人数的变化**。



**来源学校分析**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**历年来源学校学生人数及变化曲线**、来校类型分布及占比情况统计等。

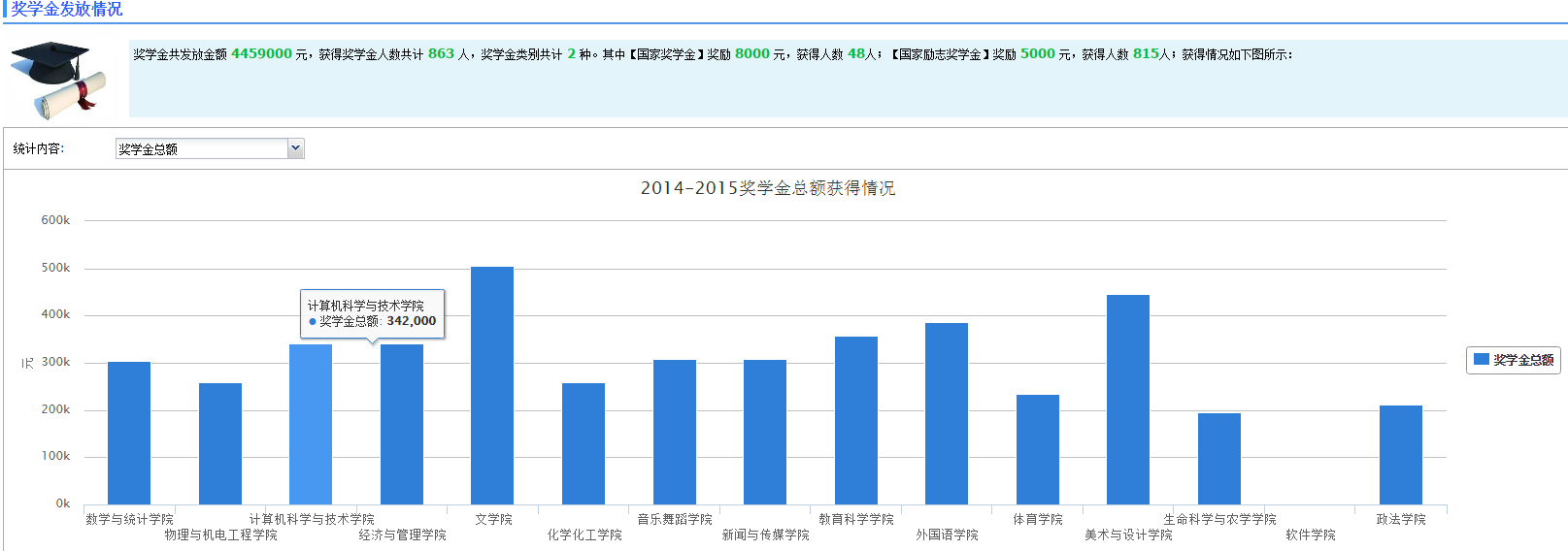


**预留就业分析接口**

1. 预留以以学校、院系、**专业**为层级，本科生、专科生、研究生为类别，就业总体人数及比率情况的统计分析接口或服务。
2. 预留以以学校、院系、**专业**为层级，本科生、专科生、研究生为类别，一次就业率和二次就业率的统计分析接口或服务。
3. 预留以以学校、院系、**专业**为层级，本科生、专科生、研究生为类别，横向就业率对比的统计分析接口或服务。
4. 预留以以学校、院系、**专业**为层级，本科生、专科生、研究生为类别，其他维度的统计分析接口或服务。

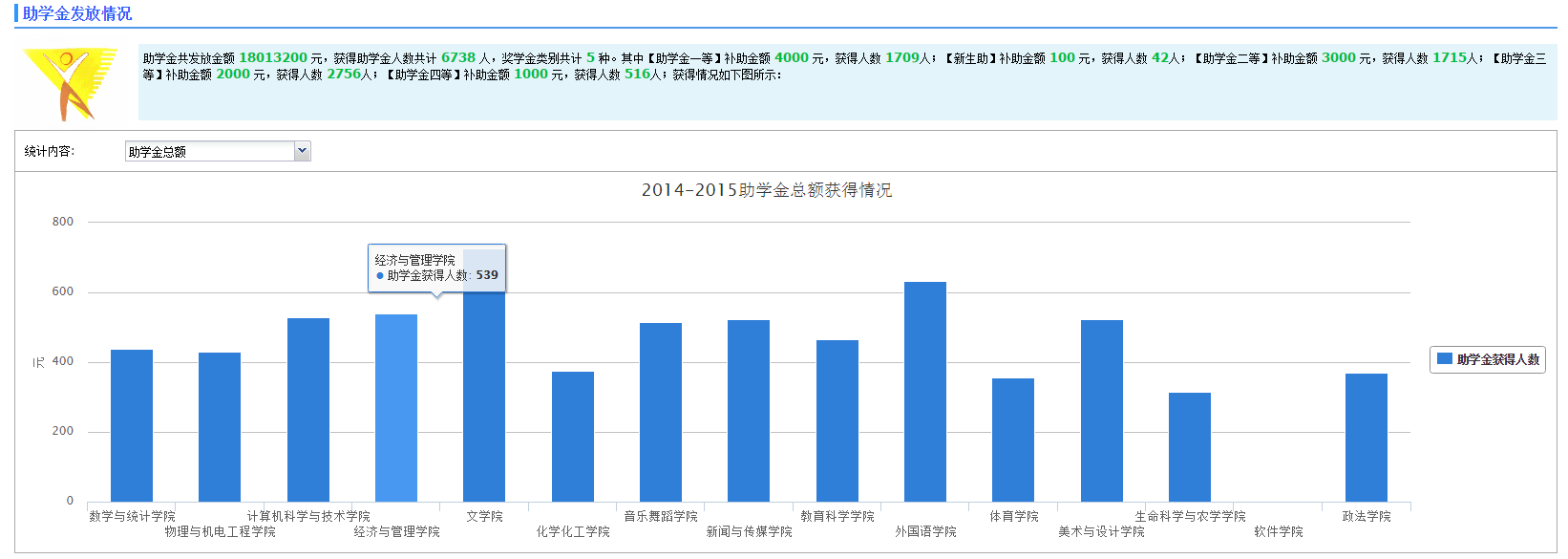
**奖学金统计**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**各奖学金类型或等级**的学生人数**及历史变化曲线**。奖学金等级可以按照国家级、省级以及校级等分组方式进行分组统计。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



**助学金统计**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**各助学金类型或等级**的学生人数**及历史变化曲线**。助学金等级可以按照国家级、省级以及校级等分组方式进行分组统计。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



**贷款统计**

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**贷款金额、贷款学生户口性质、生源地等维度进行分析统计贷款学生人数分布并展示历年的变化曲线**。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。



**学费减免统计**

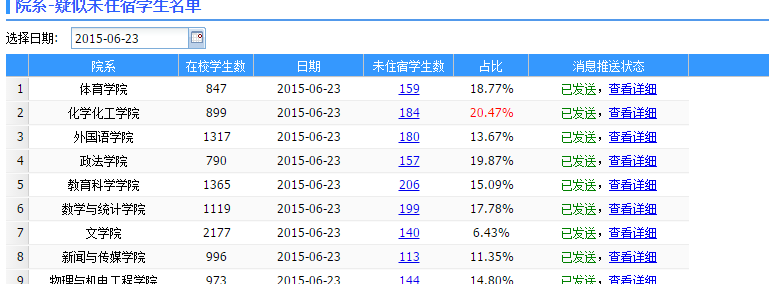
1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示**减免金额、减免学生户口性质、生源地等维度进行分析统计减免学生人数分布并展示历年的变化曲线**。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。

#### 学生预警

学生预警是在学生基本信息、一卡通消费信息、图书馆门禁系统信息以及宿舍门禁系统信息的基础上，通过缜密的逻辑和复杂的算法，从而分析出针对各个学院的学生管理方面的预警性信息。

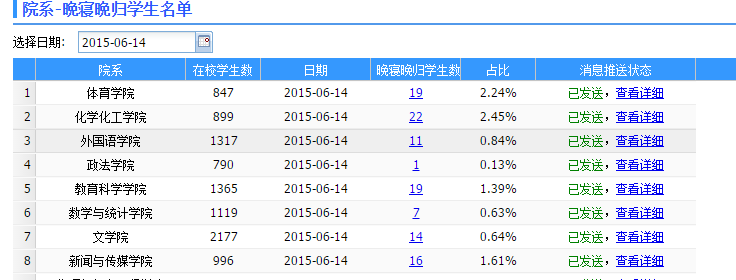
##### 疑似未住宿学生名单

1. 系统将以各个学院为层级，通过一卡通消费信息和宿舍门禁信息分析计算出疑似当日没有住宿的学生，并通过数据库进行存储和表格的形式展现。
2. 管理员或有该功能权限的用户可以按日期查询每天疑似未住宿的学生名单。
3. 系统将通过设置好的时间，定时自动的将名单发送至指定的邮箱中。
4. 点击各个学院下疑似未住宿学生人数，可显示出学生的详细信息和分析依据。
5. 点击“发Email”可直接向指定的邮箱中发送信息。
6. 点击“查看详细”可查看邮件的接收情况，以及附近内容。



##### 晚寝晚归学生名单

1. 系统将以各个学院为层级，通过宿舍门禁信息，每日记录22:30以后入寝刷卡的学生名单，并通过数据库进行存储和表格的形式展现。
2. 管理员或有该功能权限的用户可以按日期查询每天晚寝晚归的学生名单。
3. 系统将通过设置好的时间，定时自动的将名单发送至指定的邮箱中。
4. 点击各个学院下晚寝晚归学生人数，可显示出学生的详细信息和分析依据。
5. 点击“发Email”可直接向指定的邮箱中发送信息。
6. 点击“查看详细”可查看邮件的接收情况，以及附近内容。



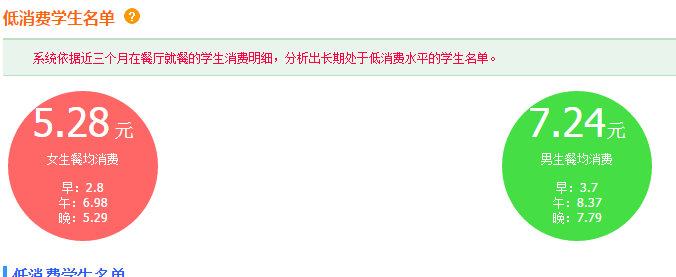
##### 疑似不在校学生名单

1. 系统将以各个学院为层级，通过一卡通消费信息和宿舍门禁信息分析出三天或三天以上既没有一卡通消费数据也没有门禁数据的学生，将被列入疑似不在校学生名单中，并通过数据库进行存储和表格的形式展现。
2. 管理员或有该功能权限的用户可以查看近期不在校的学生名单。
3. 系统将通过设置好的时间，定时自动的将名单发送至指定的邮箱中。
4. 点击各个学院下疑似不在校学生人数，可显示出学生的详细信息和分析依据。
5. 点击“发Email”可直接向指定的邮箱中发送信息。
6. 点击“查看详细”可查看邮件的接收情况，以及附近内容。



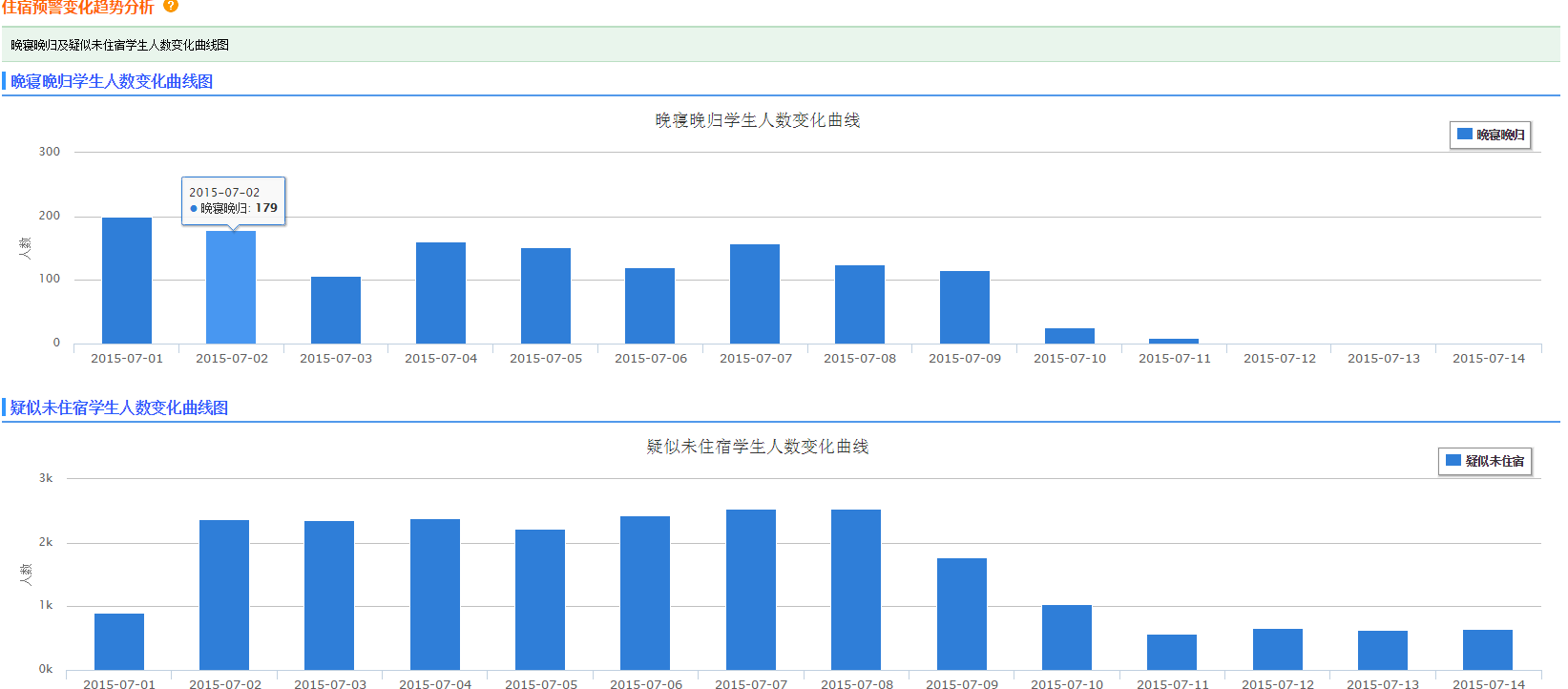
##### 低消费学生名单

1. 系统将以各个学校、学院为层级，通过一卡通消费信息依据近三个月在餐厅就餐的学生消费明细，分析出长期处于低消费水平的学生名单，通过数据库进行存储和表格的形式展现。
2. 系统通过分析完毕的低消费水平学生名单，按照不同的性别和早中晚三餐计算出各个的平均消费水平和不同性别的总体消费水平。
3. 点击吃饭次数下的数字，可显示出学生的详细消费信息和当日的平均消费水平。



##### 住宿预警变化趋势分析

1. 系统通过晚寝晚归学生人数，按照柱状体的形式展示出近半个月晚寝晚归学生人数变化情况和走势图。
2. 系统通过疑似未住宿学生人数，按照柱状体的形式展示出近半个月疑似未住宿学生人数变化情况和走势图。



##### 晚归、疑似未住宿每月TOP10

1. 系统将以各个学校、学院为层级，按照各个月份进行晚寝晚归学生人数统计，并通过表格的形式展示出每个月晚寝晚归学生最多的前十名。
2. 系统将以各个学校、学院为层级，按照各个月份进行疑似未住宿学生人数统计，并通过表格的形式展示出每个月疑似未住宿学生最多的前十名。
3. 点击学生学号或学生姓名，能够展示出学生的基本信息、晚归和未住宿的详细日期、一卡通消费、奖惩助贷、图书借阅等信息。



##### 实习离校学生

1. 系统在统计未住宿、不在校等预警性分析时，为保证分析计算结果的准确性和真实性，需要去除实习离校学生。
2. 点击“下载模板”，可下载要导入的excel格式。
3. 点击“选择文件”选择要导入的实习离校学生excel。通过点击“导入Excel”进行实习离校学生的导入工作。
4. 管理员或者该实习离校学生的院系负责人账号，可对其进行删除操作。

##### 院系负责人

1. 管理员用于维护学校下各个院系的负责人名称和联系邮箱。
2. 系统通过这些邮箱，将每天生成的预警性信息按照不同的院系负责人，发个给各个相关人员。

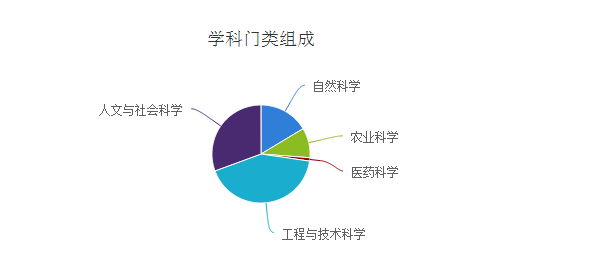


#### 教职工基本情况统计

教职工基本情况统计支持各种图形、表格形式，以专业、年龄、职称、性别、民族、政治面貌、学历、学位、教职工来源等不同维度展示教职工人数分布、占比等统计信息。系统会对教职工进行11个方面270多个字段进行统计分析，产出相应的高价值统计应用。

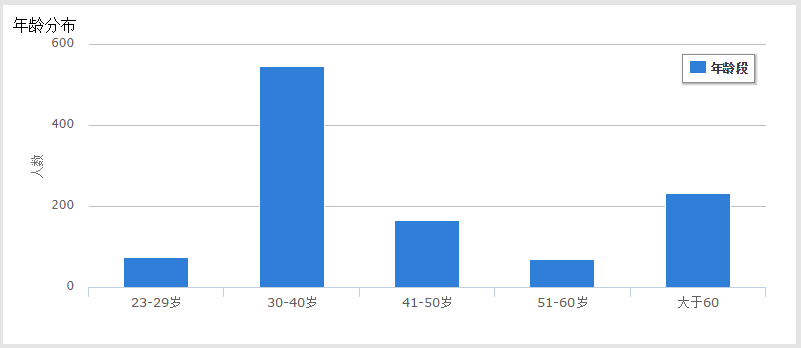
##### 在校教职工专业及科类统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的专业科类人数分布占比及比例；
2. 系统将给出相同层级之间的横向对比统计。例如：各部门之间专业人数对比统计直方图或线形图。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护专业科类的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



##### 在校教职工年龄段统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的年龄段人数分布占比及比例；年龄段将按照23-29、30-40、41-50、51-60、60以上分组形势进行分组统计各年龄段的人数及占比情况。
2. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
3. 系统将给出相同层级之间的横向对比统计。例如：各部门之间相同人员类别下人数对比统计直方图或线形图。
4. 对于没有维护人员出生日期的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。

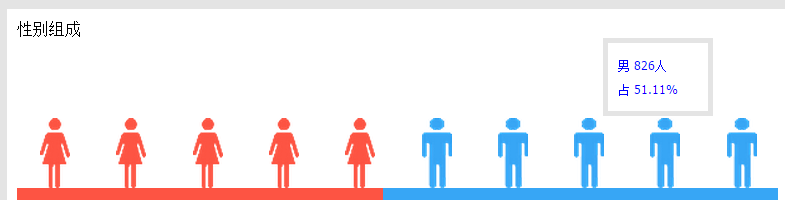


##### 在校教职工职称统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的职称人数分布占比及比例； 展示以高级、中级、初级职称等级为分组统计各分组下的人员分布及占比情况。
2. 系统将给出相同层级之间的横向对比统计。例如：各部门之间相同职称下人数对比统计直方图或线形图。
3. 对于没有维护人员出生日期的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。

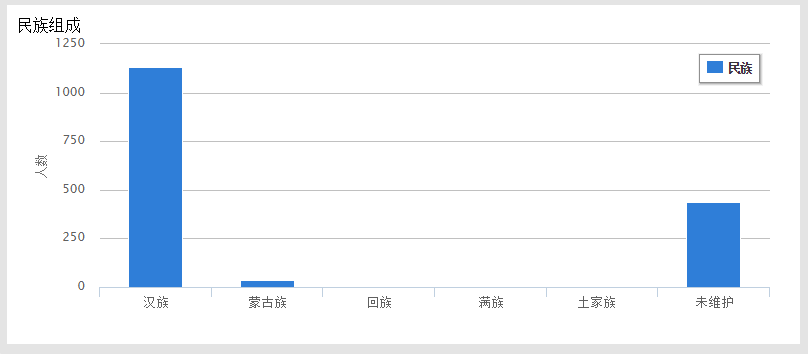
##### 在校教职工性别统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的性别人数分布占比及比例；
2. 系统将给出相同层级之间的横向对比统计。例如：各部门之间男女人数对比统计直方图或线形图。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护人员性别的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



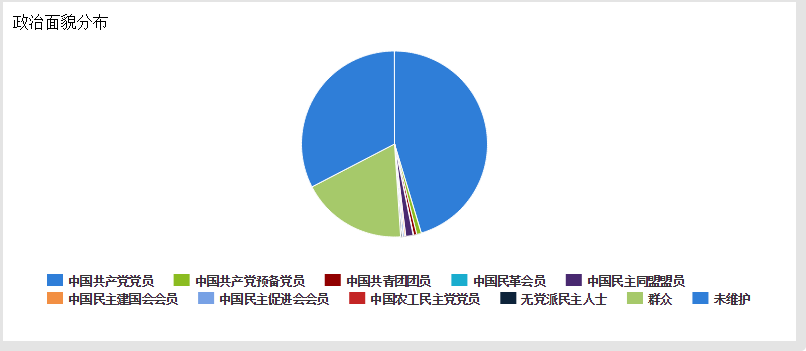
##### 在校教职工民族统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的民族人数分布占比及比例；
2. 系统将给出相同层级之间的横向对比统计。例如：各部门之间各民族下人数对比统计直方图或线形图。同样可以以汉族和少数民族等分组方式对民族进行各部门分组统计人数及占比比例。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护民族的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



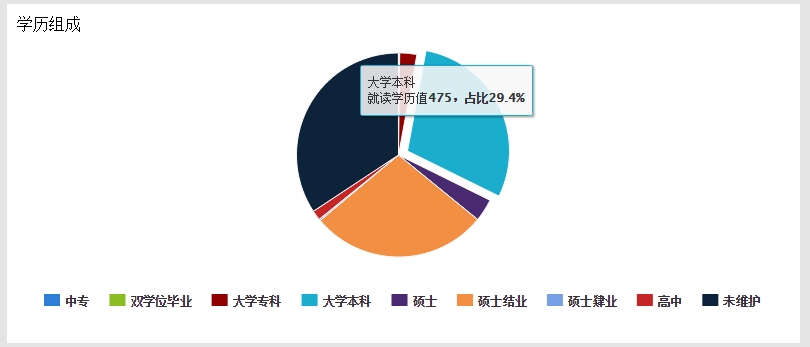
##### 在校教职工政治面貌统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的政治面貌人数分布占比及比例；
2. 系统将给出相同层级之间的横向对比统计。例如：各部门之间政治面貌下人数对比统计直方图或线形图。可以以党员、预备党员、非党员等分组方式分组展示各部门教职工政治面貌组成及分布。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护政治面貌的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



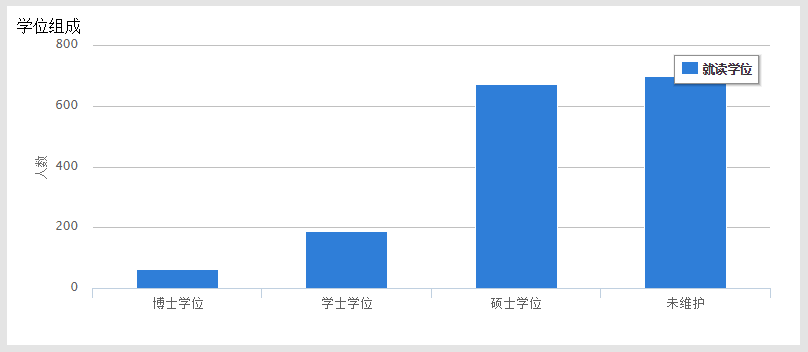
##### 在校教职工学历统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的学历人数分布占比及比例；
2. 系统将给出相同层级之间的横向对比统计。例如：各部门之间相同学历下人数对比统计直方图或线形图。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护学历的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



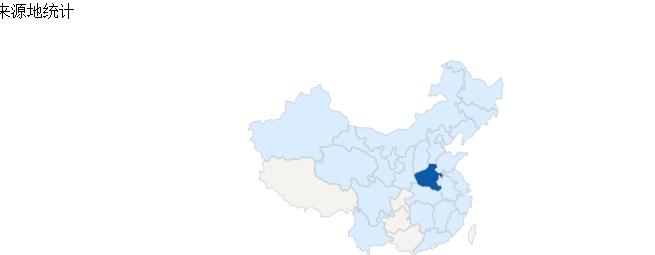
##### 在校教职工学位统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的学位人数分布占比及比例；
2. 系统将给出相同层级之间的横向对比统计。例如：各部门之间相同学位下人数对比统计直方图或线形图。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护学位的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



##### 在校教职工来源统计

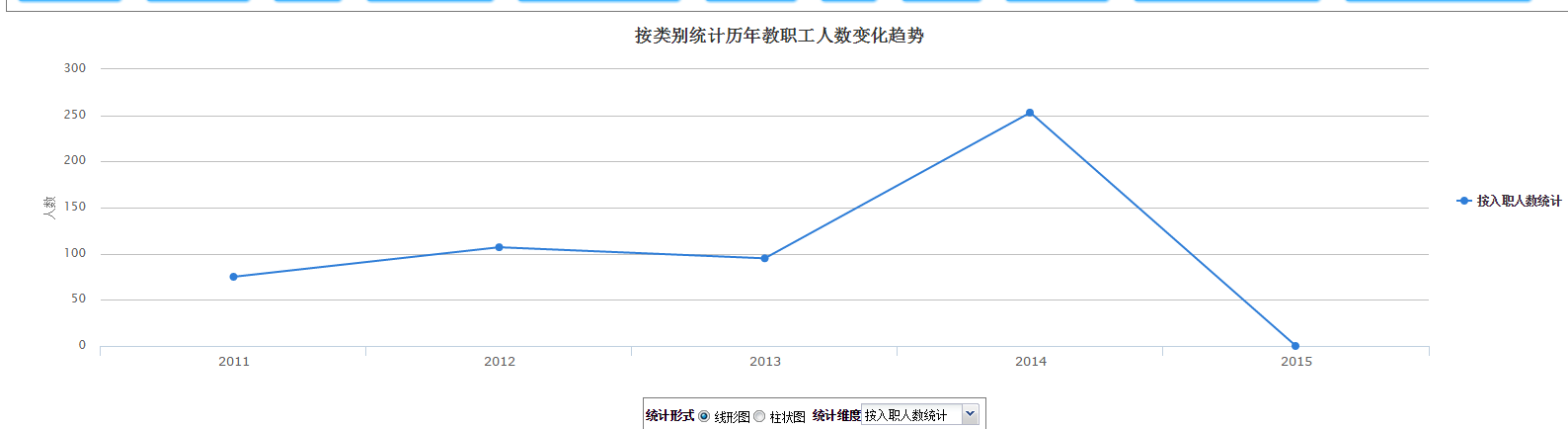
1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工的教职工来源人数分布占比及比例；
2. 系统将给出相同层级之间的横向对比统计。例如：各部门之间相同来源下人数对比统计直方图或线形图。
3. 用户点击相应概况统计数据时以表格的形式展示相应的明细信息。
4. 对于没有维护教职工来源的数据系统将以“未维护”进行数据归类统计，方便与数据监管系统组成数据质量监管的双重保障。



#### 教师全景式统计分析

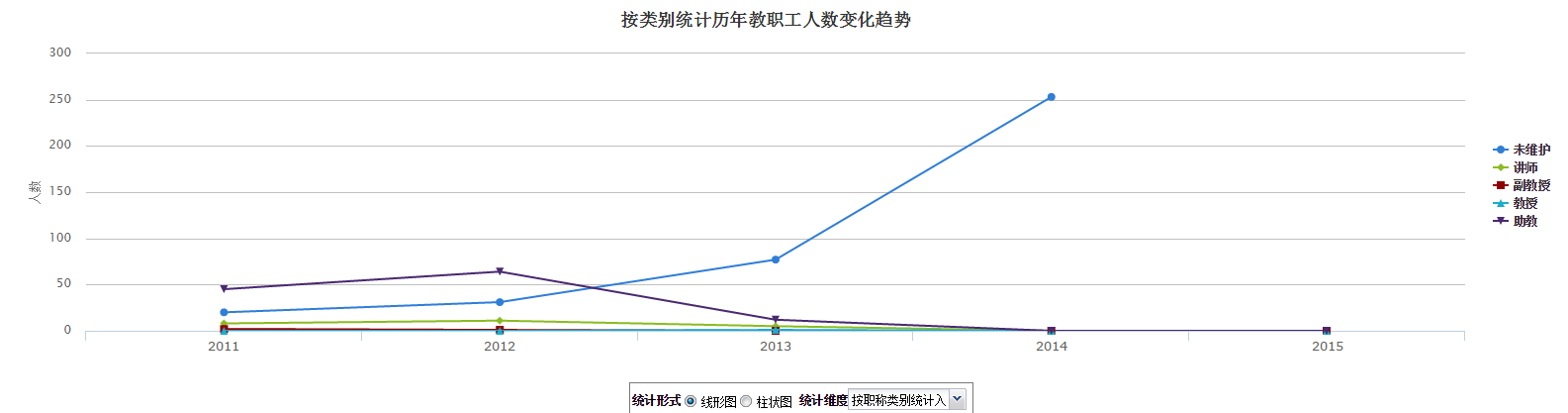
##### 支持教职工总量情况

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工总人数人数情况；
2. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工历年总人数人数情况及趋势曲线图；
3. 系统将给出相同层级之间的横向对比统计。例如：各部门之间教职工总人数对比统计直方图及线形图。



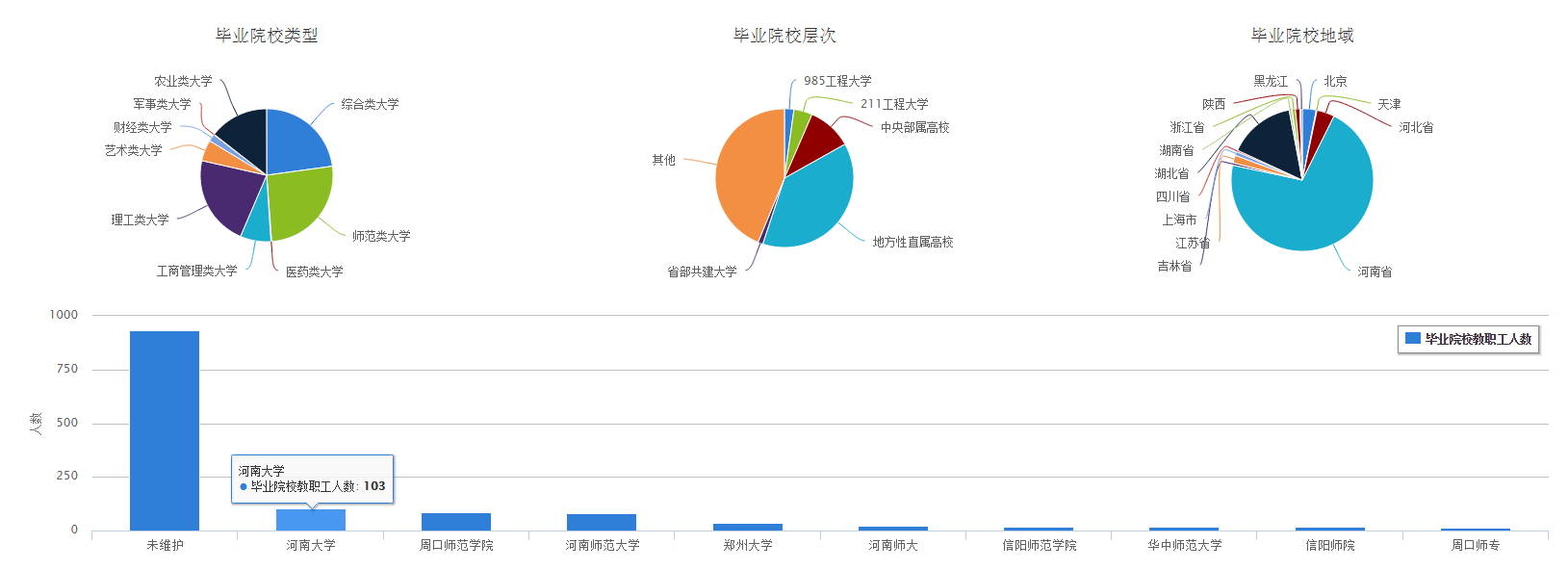
##### 教职工职称情况进行详细的统计分析

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工职称相关人数情况统计；
2. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示教职工历年各职称总人数人数情况及趋势曲线图；
3. 系统将以学校、部门为层级，按照图形或表格的形式分别按照高级、中级、初级等分组方式分组统计展示各分组总人数情况及趋势曲线图；
4. 系统将给出相同层级之间的横向对比统计。



##### 教职工毕业院校情况统计分析

1. 系统将以学校、部门为层级，以毕业院校类型、毕业院校层次、毕业院校地域等不同维度展示教职工人数分布、占比等统计信息，按照图形或表格的形式分别统计展示教职工毕业院校情况统计；
2. 系统将给出相同层级之间的横向对比统计。



##### 教职工来源类型统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示来源类型的人数情况；
2. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示历年来源类型人数情况及趋势曲线图；

##### 教职工编制类型统计

1. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示编制类型的人数情况；
2. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示历年编制类型人数情况及趋势曲线图；

#### 教学资源与教学活动统计

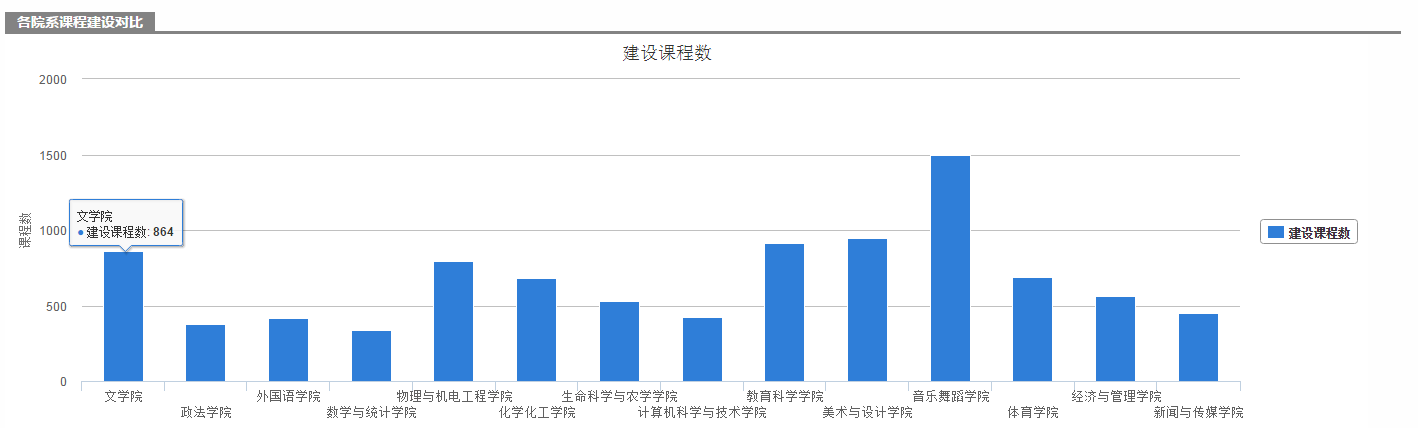
教学资源与教学活动统计支持各种图形、表格形式，以学校课程库、学时、学时分类、教学任务、学生成绩、学生信息、网络学习空间、一卡通消费、图书馆门禁等不同维度展示出课程占比、学时趋势、成绩对比与趋势、学霸等统计信息和成绩与消费、成绩与图书门禁等交叉分析。

##### 学校课程库统计

1. 统计学校的课程信息，包含课程库的课程总量，开设课程中的理论课与实践课占比，以及本学期开设课程的理论课与实践课情况。

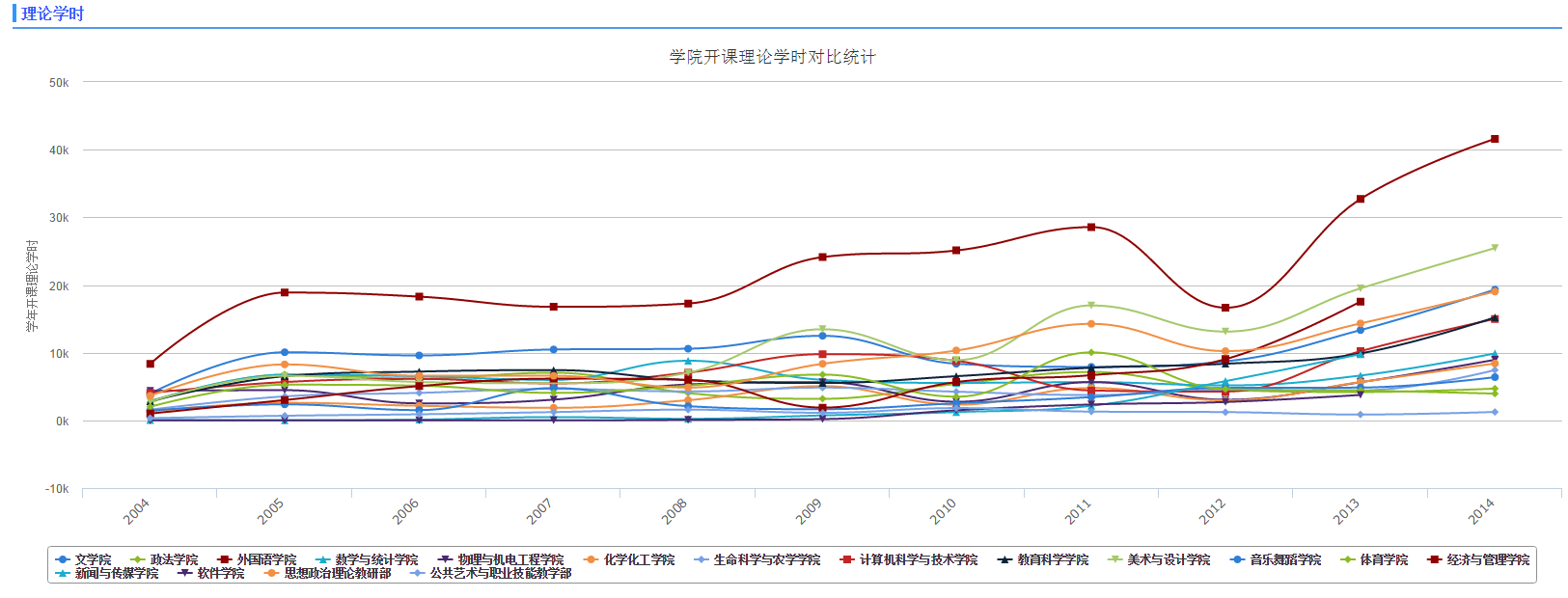


1. 系统以不同图形、文字描述、数字的形式展示了学校课程的分类情况和承担单位信息。
2. 系统以学院为分类，通过柱状图的形式展现出各个学院之间的课程数并进行横向对比。

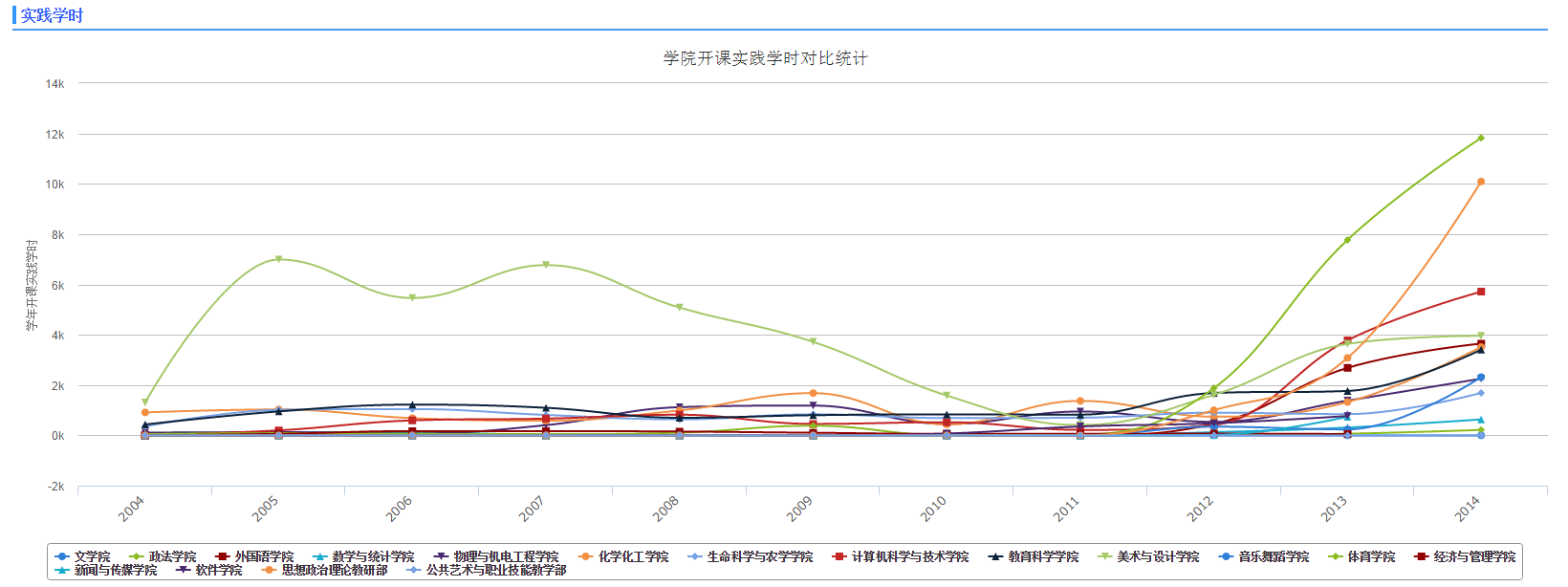


##### 各学院学年总学时变化趋势

1. 系统以学年为统计维度对各学院理论学时通过曲线图的形式进行展示，并体现出近10年中理论课时的走势情况。



1. 系统以学年为统计维度对各学院实践学时通过曲线图的形式进行展示，并体现出近10年中实践课时的走势情况。

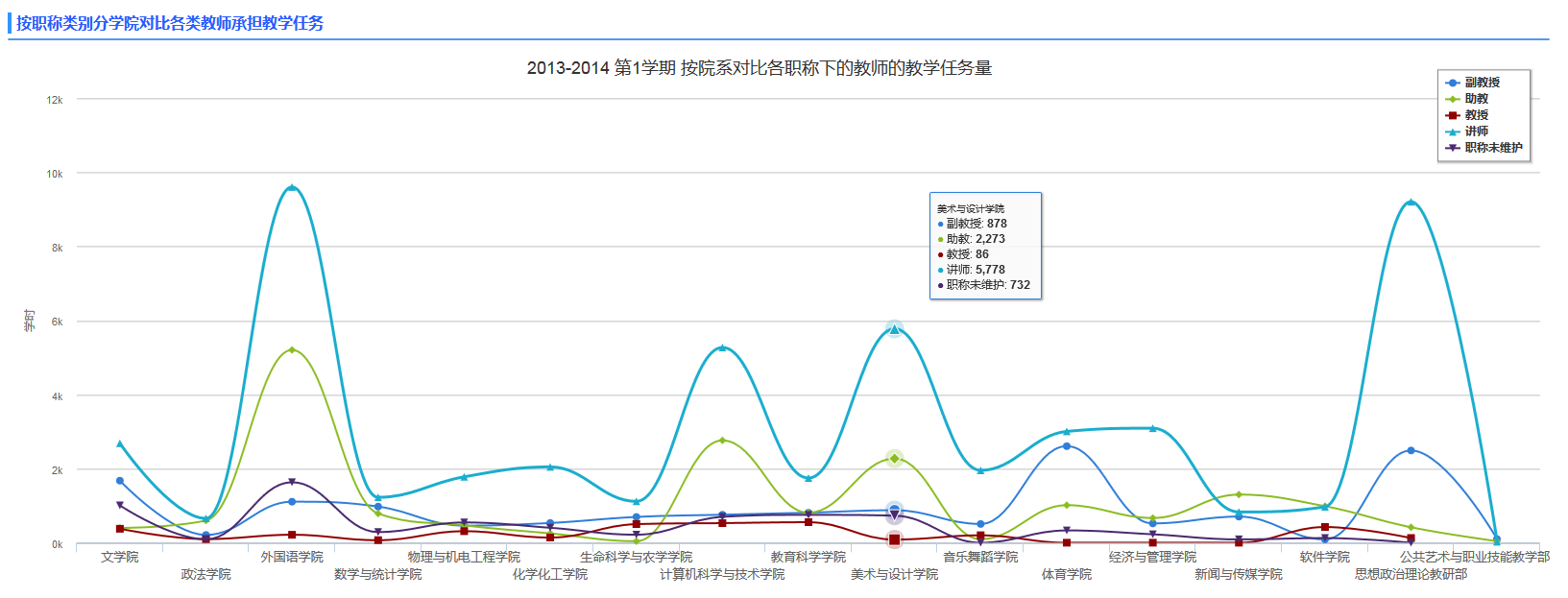


##### 各学院学年总学时变化趋势

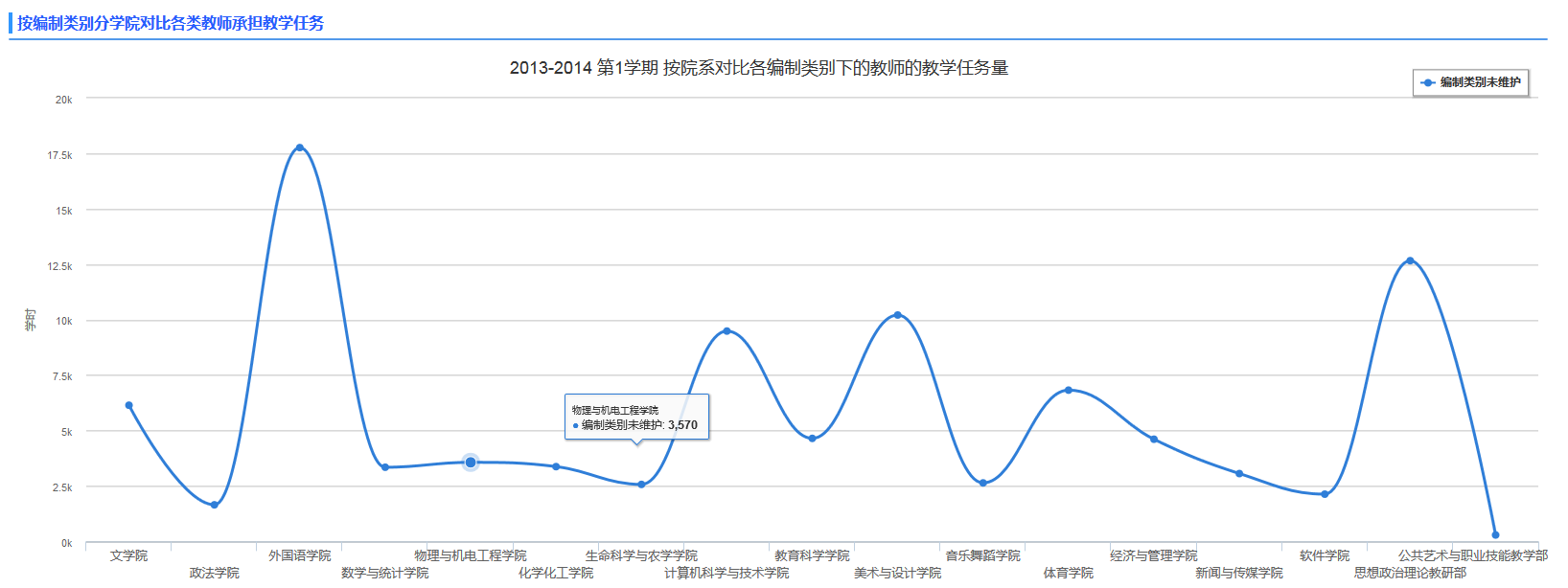
1. 系统将以学校、部门为层级，按职称、编制类别分学院(部)展示出不同学期学年中教师承担教学任务比例，并以图形和文字的形式展现。



1. 系统将以学校、部门为层级，按职称类别分学院对比各类教师承担教学的任务进行统计，并通过柱状图的形式进行学时上的展示和横向对比。



1. 系统将以学校、部门为层级，按编制类别分学院对比各类教师承担教学的任务进行统计，并通过柱状图的形式进行学时上的展示和横向对比。

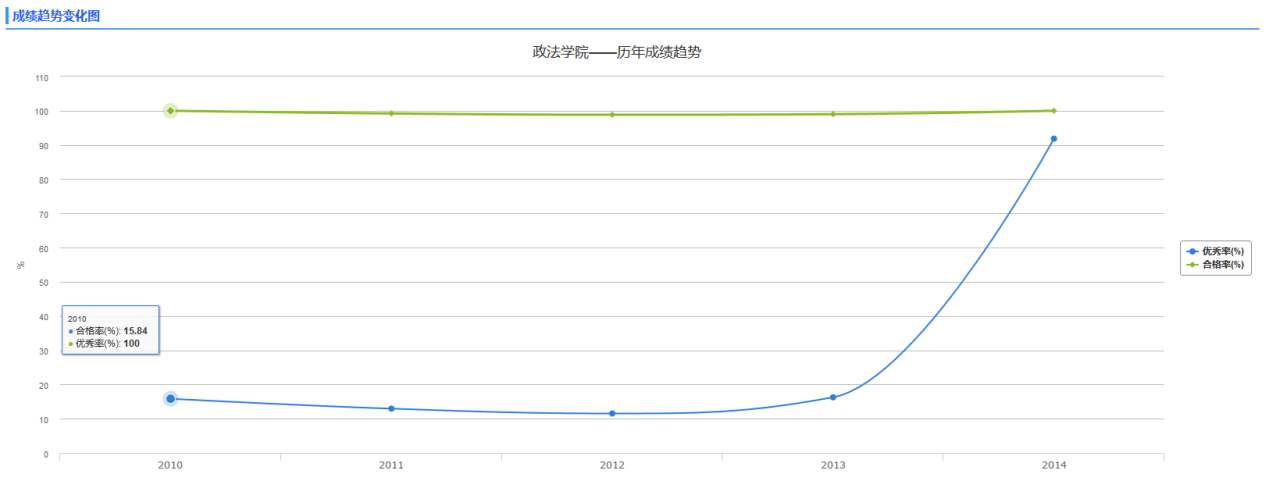


##### 各院系学生成绩统计

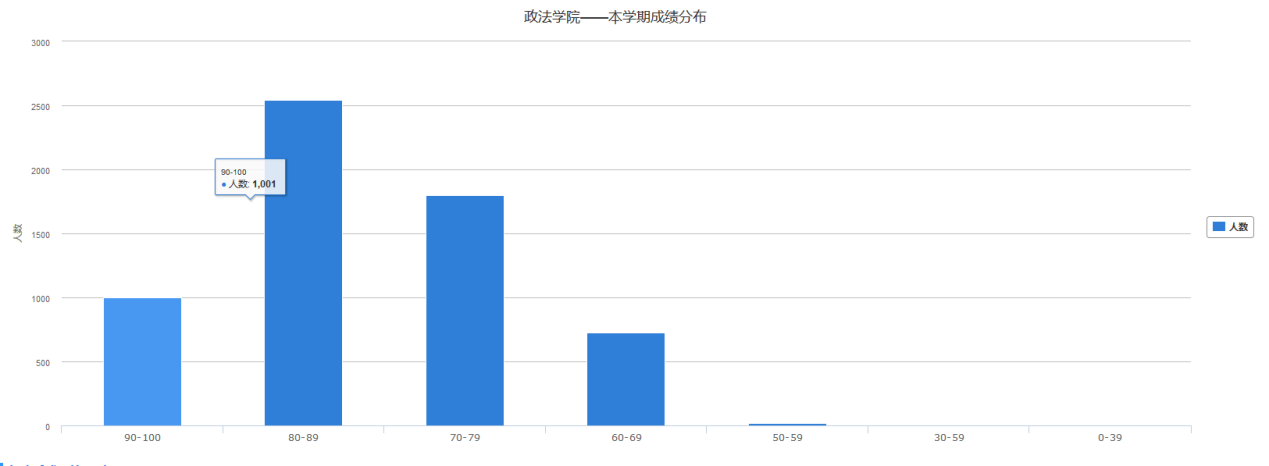
1. 系统将以学校、部门为层级，统计本学年学期中学生成绩，合格率、优秀率、平均分、方差、均方差、标准差反映学生成绩整体水平，并通过表格的形式展现。



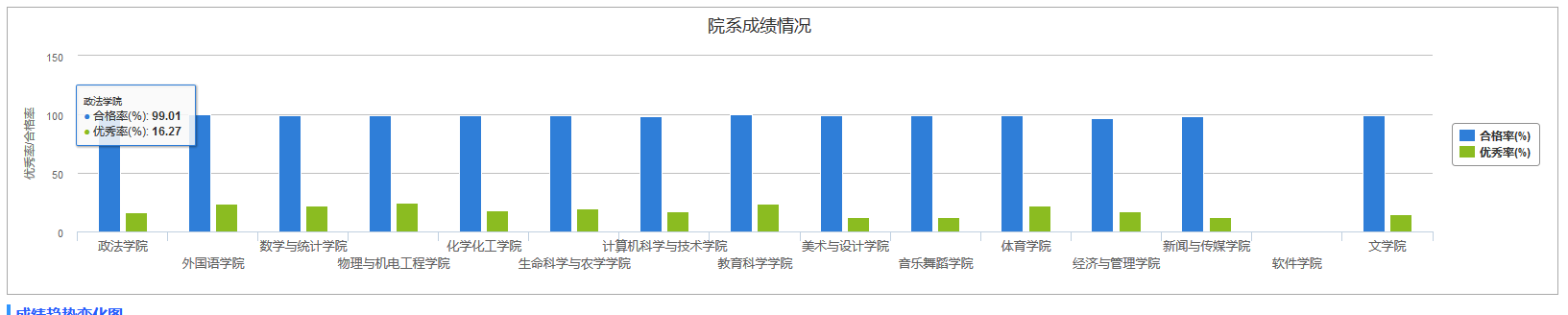
1. 点击“历史趋势”可查看近5年来该学院的历史成绩趋势，并通过曲线图的形式展现。



1. 点击“成绩分布”可查看该学院本学年学期中学生的成绩分布，并以不同成绩段为维度通过柱状图的形式进行展现和横向对比。

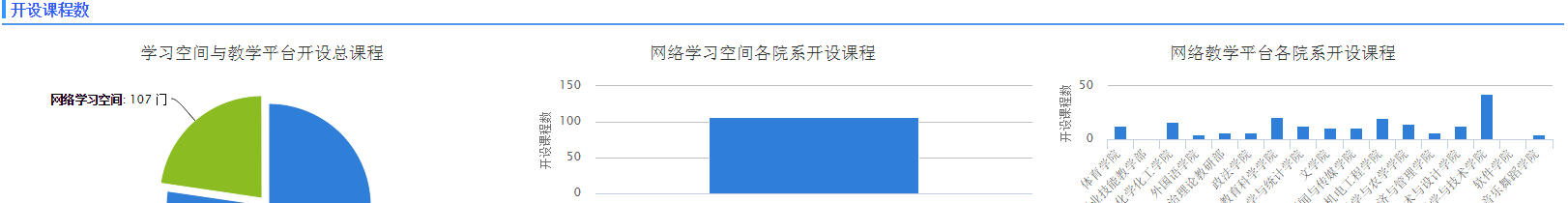


1. 系统将以学校、部门为层级，统计学生合格率和优秀率，并以不同学校为维度，通过柱状图的形式横向展示对比出各个学院的合格率和优秀率情况。

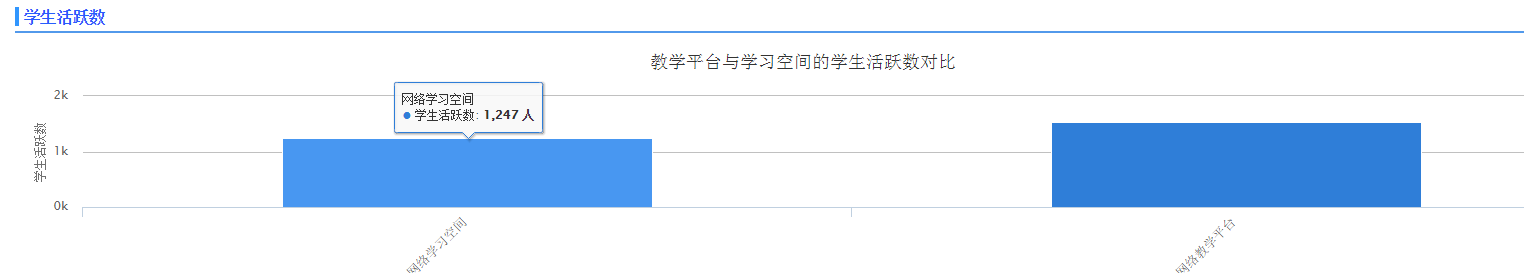


##### 网络学习统计分析

1. 统计网络学习空间和网络教学平台开设课程数。以时间为区间，按照饼状图、柱状图或表格的形式分别统计展示网络学习空间和网络教学平台上各院系开设的课程数。



1. 统计网络学习空间和网络教学平台学生活跃数。以时间为区间，按照饼状图、柱状图或表格的形式分别统计展示网络学习空间和网络教学平台上学生活跃数。



1. 统计网络学习空间和网络教学平台课程访问排名。以时间为区间，根据学生访问量，按照饼状图、柱状图或表格的形式分别统计展示网络学习空间和网络教学平台上前十名课程的排名情况。



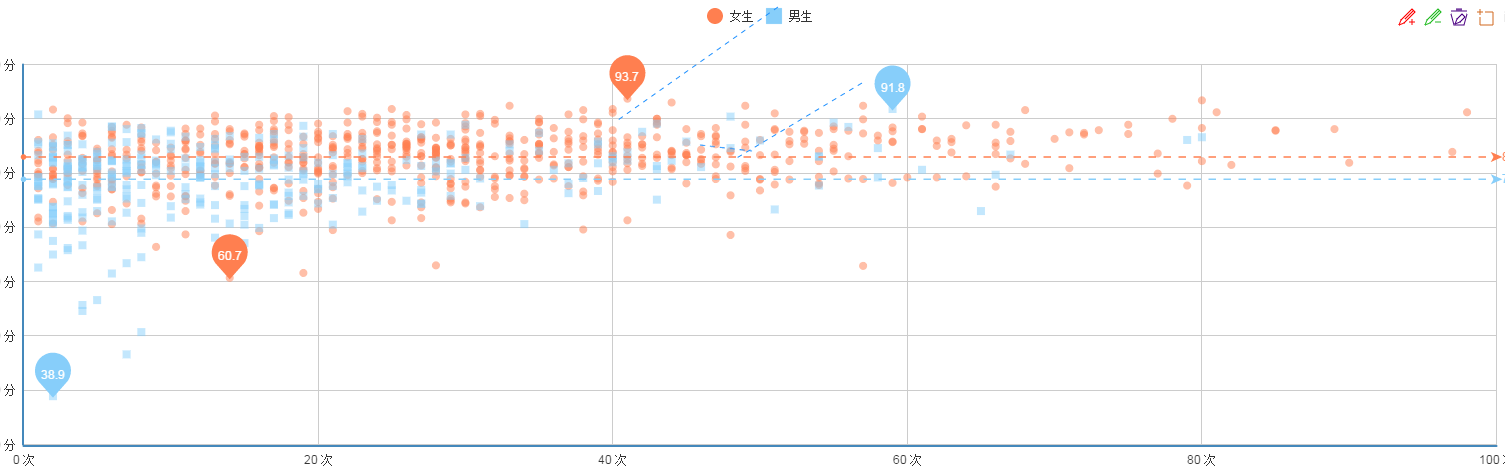
##### 学霸TOP5

1. 系统将以学校、院系、专业为层级，本科生、专科生、研究生为类别，按照图形或表格的形式分别统计展示大三以上学生，总成绩在3800以上，按综合平均绩点（GPA）从高到低查找的学霸TOP5。
2. 点击学生的学号或者姓名，可以进入查看该学生的详细信息，包括住宿情况，一卡通消费情况，奖惩助贷情况，以及图书借阅明细，成绩明细。



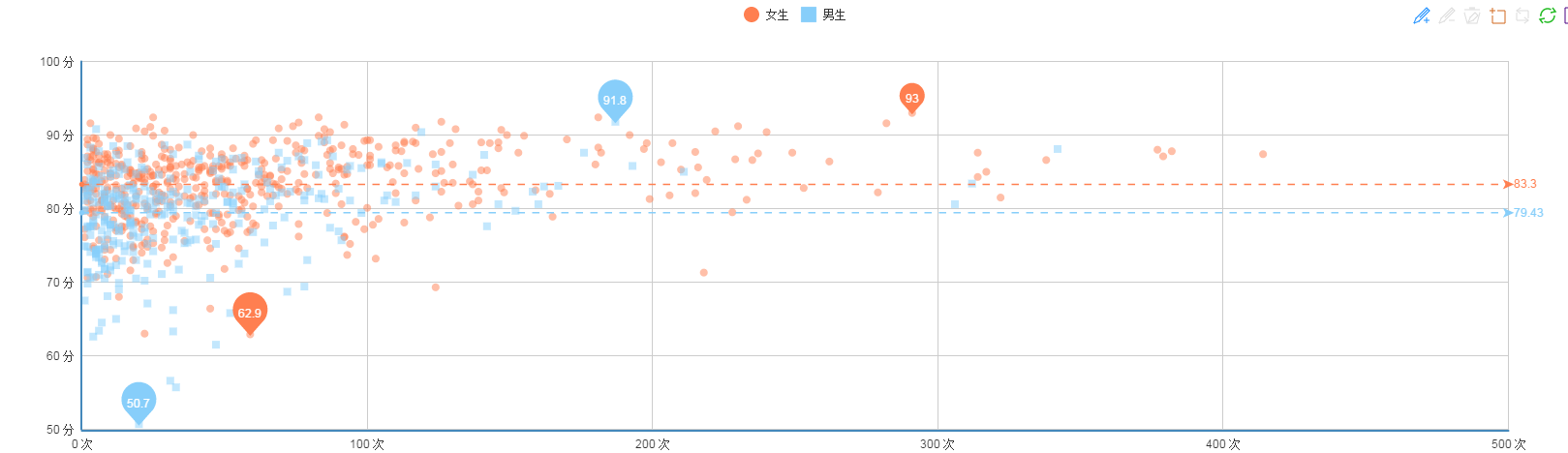
##### 男生女生学习成绩与早餐次数分布

系统将以抽样数据的方式，按男女为类别，按照图形或表格的形式分别统计展示一个学期抽样学生早餐次数和成绩的分布。



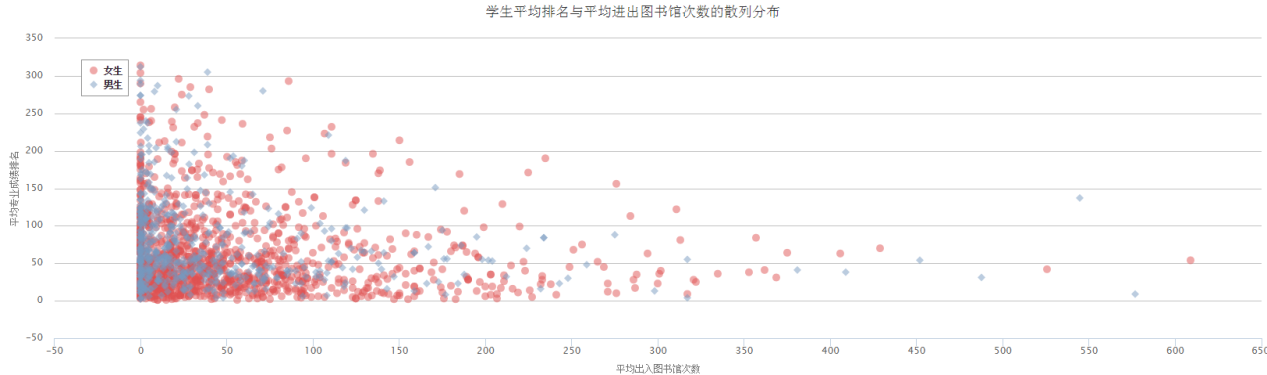
##### 男生女生学习成绩与图书馆出入分布

系统将以抽样数据的方式，按男女为类别，按照图形或表格的形式分别统计展示一个学期抽样学生图书馆门禁出入次数和成绩的分布。



##### 学生平均排名与平均进出图书馆次数的散列分布

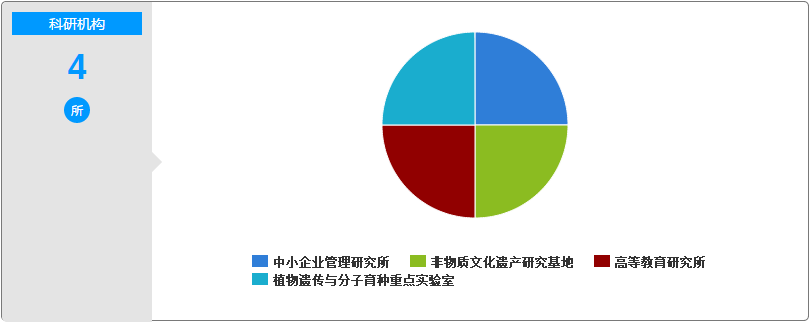
系统随机从全部在校学生中抽取男生500名，女生500名，并计算他们的平均排名与平均出入图书馆次数。按男女为类别，按照图形或表格的形式分别统计展示一个学期抽样学生平均出入图书馆次数和平均排名的分布。



#### 科研成果概况展示

##### 科研机构概况

1. 以文本、数字加图形的形式描述现学校的教研机构总数、教研机构类型组成、教研机构占全校机构或各院系、部门机构的百分比。
2. 系统将以学校、科研机构为层级，按照条形图对比统计各科研机构的科研人数。
3. 系统将以学校、科研机构为层级，按照数字、文本加图形的方式对比统计各科研机构的在研科目数、结项科目数、撤销科目数、暂停科目数。
4. 系统将以学校、科研机构为层级，按照数字、文本加图形的方式对比统计各科研机构的学术活动场次。

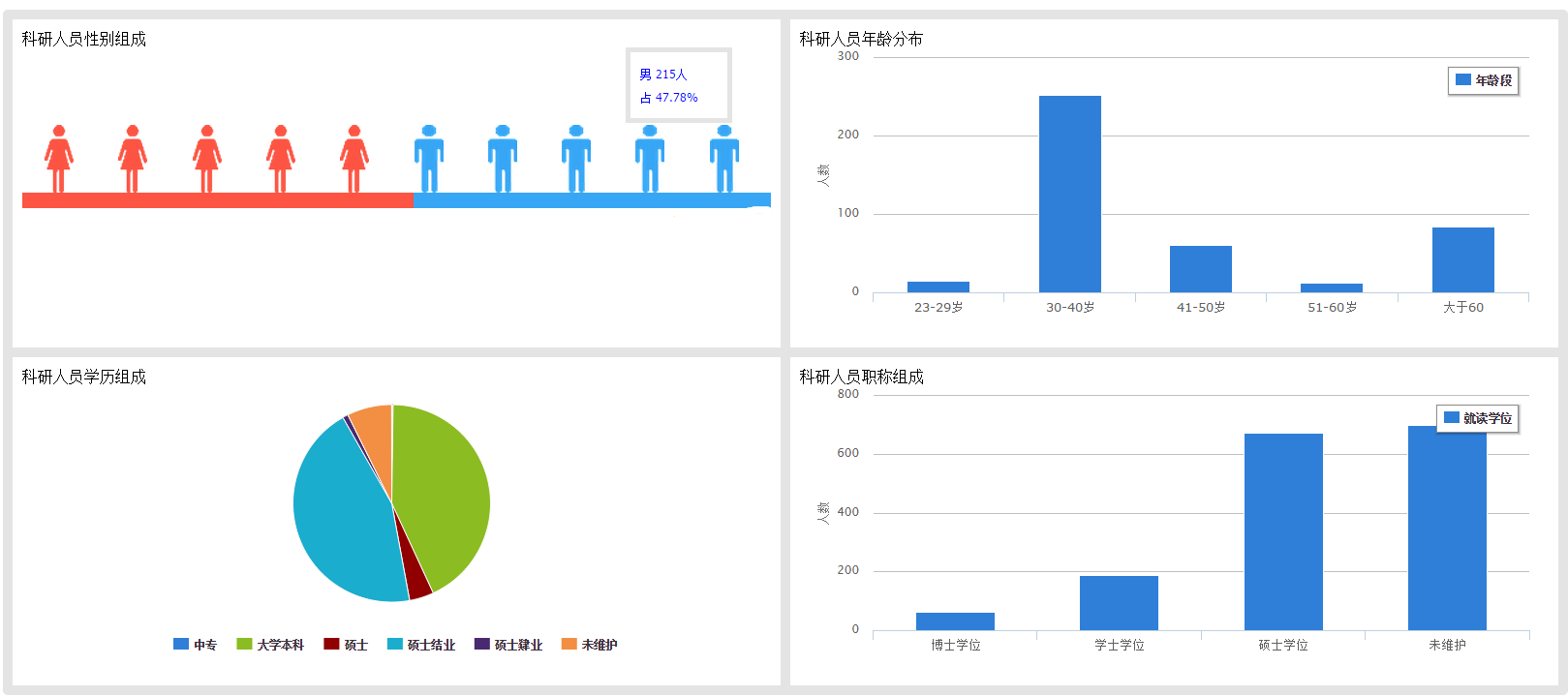


##### 科研人员基本情况

从性别、民族、政治面貌、学历、学位、人员类别、年龄、职称等维度统计学校科研人员的人数组成及占比情况。

以时间为维度统计学校在研科研人数及变化趋势曲线图。

按照学校、学院为层级横向对比各教研机构下的科研人员总数及比例信息。并以性别、学历、学位、人员类别、年龄为维度横向对比总人数及比例信息。

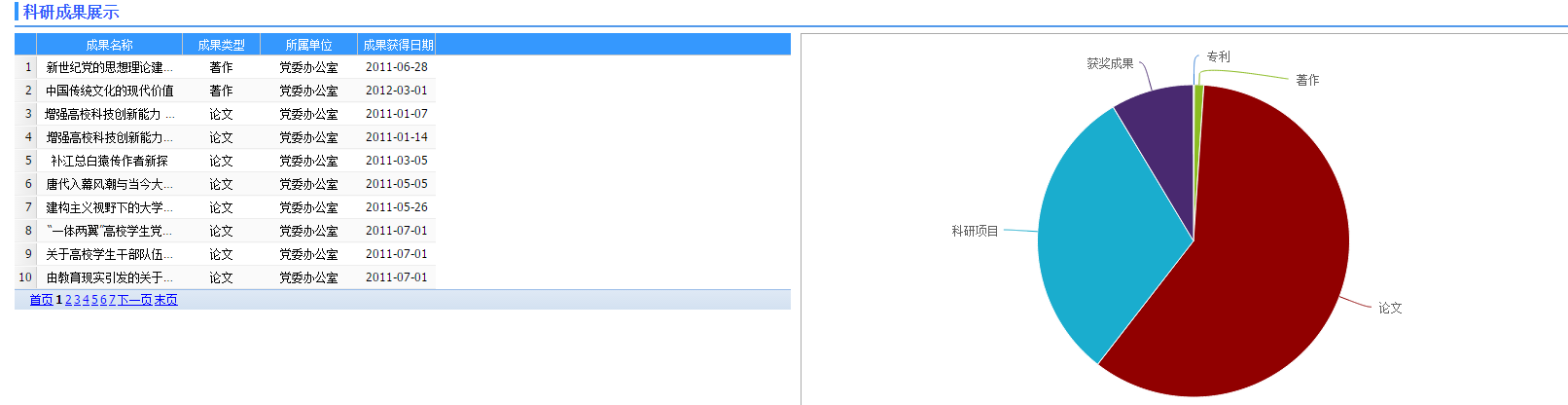
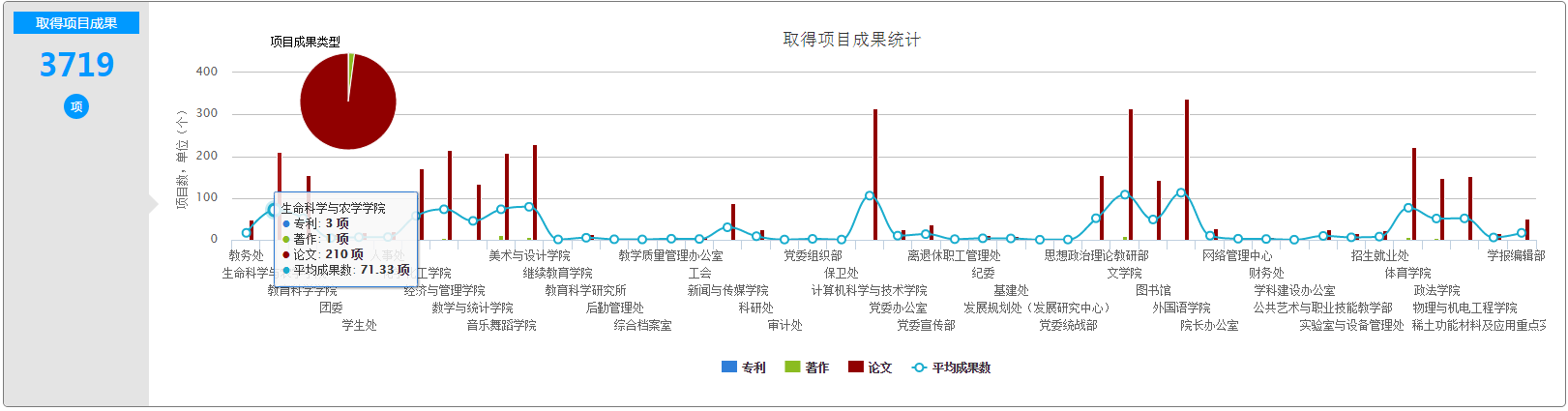


##### 科研项目基本情况

1. 系统将以学校、科研机构为层级，按照图形或表格的形式分别统计展示各机构的科研项目总数、科研项目总金额、科研项目规模等指标和维度下的信息；
2. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示历年各机构的科研项目总数、科研项目总金额、科研项目规模等指标和维度下的信息及趋势曲线图或对比直方图；
3. 系统将给出相同层级之间的横向对比统计。例如：各教研机构间各机构的科研项目总数、科研项目总金额、科研项目规模等指标和维度下的信息等对比统计直方图及线形图。

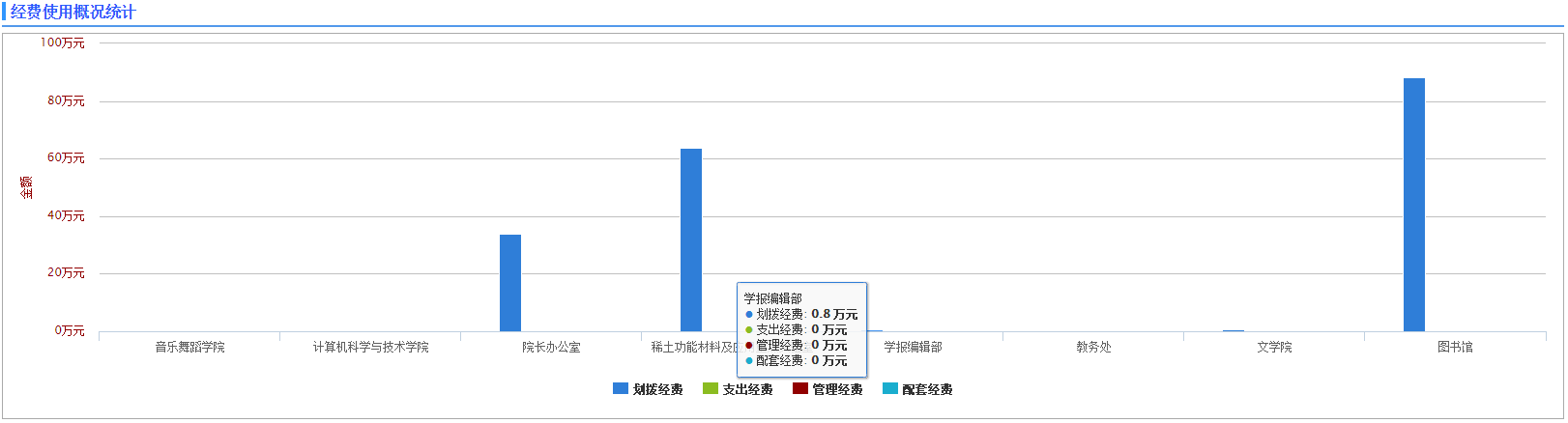
##### 科研成果基本情况（可进行分类比较）

1. 系统将以学校、科研机构为层级，按照图形或表格的形式分别统计展示各机构的科研成果数，科研成果获得奖励数等指标的统计数据；并可以按照科研成果的不同分类进行横向对比。
2. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示历年各机构的科研成果数、科研成果获得奖励数及趋势曲线图或对比直方图；
3. 系统将给出相同层级之间的横向对比统计。
4. 系统将以学校、部门为层级，以时间、部门统计为维度，按照数字、图形或表格的形式分别统计展示历年各机构的著作、艺术、论文、专利、研究成果等科研成果的信息。



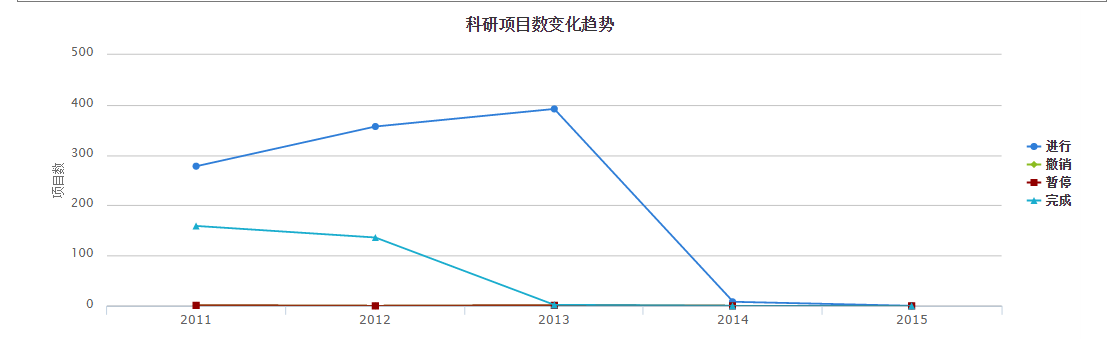
##### 科研经费使用情况（可按划拨、配套等比较）

1. 系统将以学校、科研机构为层级，按照图形或表格的形式分别统计展示各机构的科研经费总额等指标的统计数据；并按照经费类型以划拨、配套等形式进行横向和纵向对比以及已使用金额、剩余金额比率。
2. 系统将以学校、部门为层级，按照图形或表格的形式分别统计展示历年各机构的科研经费使用情况及趋势曲线图或对比直方图；
3. 系统将给出相同层级之间的横向对比统计。



##### 历年科研项目数对比统计

1. 以项目状态、项目分类、项目来源为维度，按照日期、科研部门对科研项目数横向及纵向对比统计。
2. 系统将以学校、部门为层级，按照线形图或条形图的形式分别统计展示各机构的项目数变化趋势。
3. 系统将以学校、部门为层级，按照表格的形式分别统计展示各机构的项目具体信息。例如：项目名称、项目来源、所属单位、项目状态等。



#### 资产设备与图书

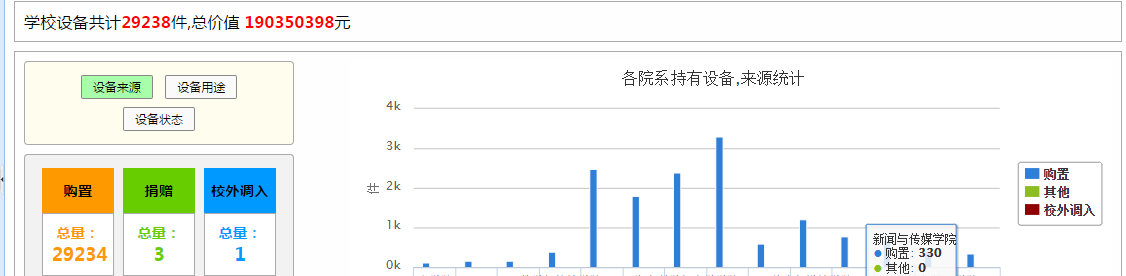
##### 资产总体信息

系统以文本、数字、图形等表现形式，通过资产类型分布组成、资产数量等维度统计反映资产的总体情况。



##### 仪器设备概况统计

1. 按照数字加文本的形式统计展示学校设备总数以及设备总价值。
2. 按照数字、文本、表格以及图形的形式统计展示设备来源、设备用途、设备状态的各类别数目和占比。例如：设备来源包括购置、捐赠、校外调入；设备用途包括教学、行政、科研、生活、其他、社会服务。



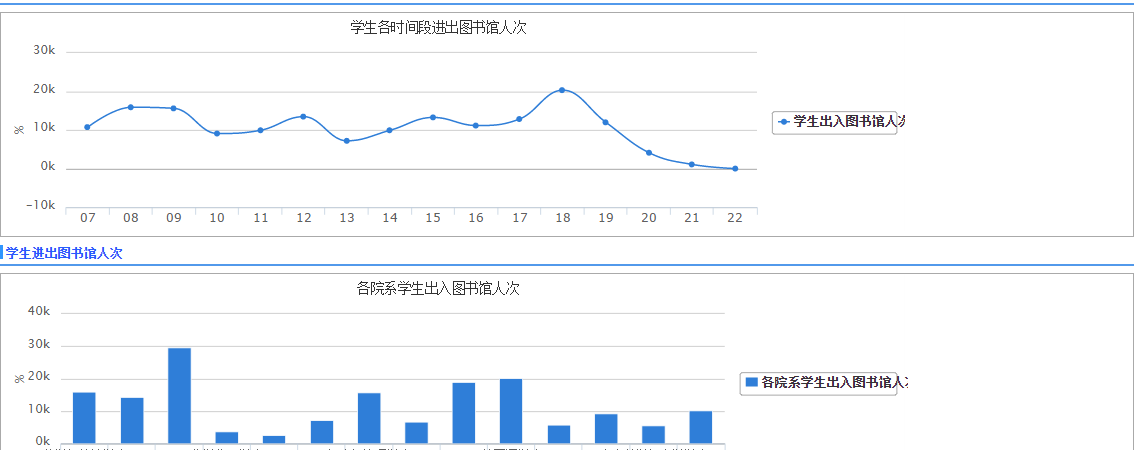
##### 图书统计分析

1. 以数字和文本统计展示学校藏书的类别总数、藏书总数和图书投入资金。
2. 按照条形图或表格的形式对比统计各馆藏类别的时间统计图书资金投入、图书类别统计图书册数、图书类别统计图书资金投入。
3. 按照文本、数字和图形的形式，统计图书馆的累计借阅量、最高日借阅量、日均借阅量以及图书馆门禁出入的累计学生流量、日均学生流量、最高学生流量。



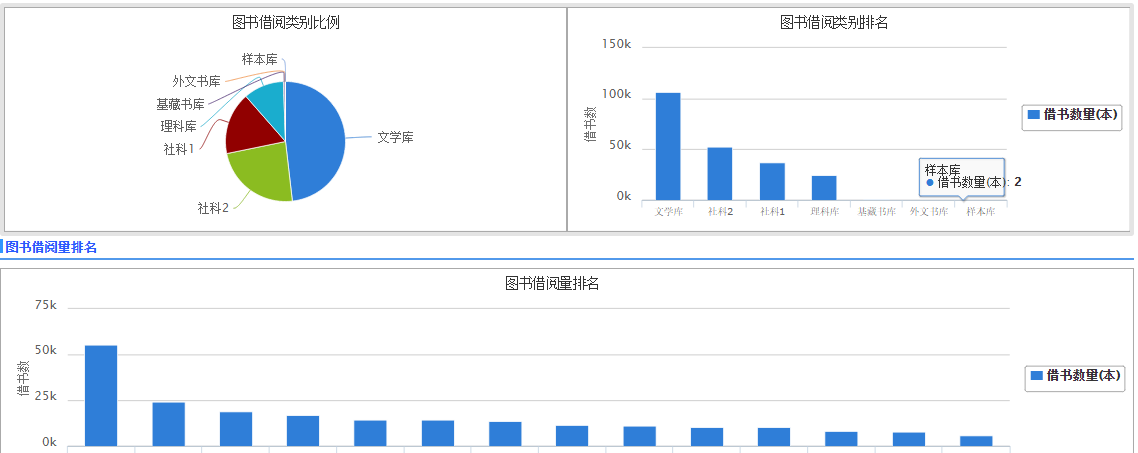
##### 学生进出图书馆统计

1. 系统将以学校、院系为层次，以时间段为维度按照线形图的形式统计展示学生各时间段进出图书馆的人次。
2. 系统将以学校、院系为层次，以时间段为维度按照条形图的形式对比统计某一时间段理各院系学生的进出图书馆次数以及院系内每个专业进出图书馆的次数。



##### 图书借阅量统计

1. 系统将以学校、院系、专业为层级，以时间段为维度，按照饼状图的形式统计展示各藏书馆图书借阅类别比例。
2. 系统将以学校、院系、专业为层级，以时间段为维度，按照条形图的形式统计展示各藏书馆图书借阅类别排名。
3. 系统将以学校、院系、专业为层级，以时间段为维度，按照条形图的形式统计展示各院系以及各专业图书借阅量排名。



##### 图书借阅TOP10

1. 系统将以学校、院系、专业为层级，以时间段为维度，按照表格的形式统计展示借出图书TOP10和借书量高的学生TOP10。
2. 点击学生的学号或者姓名，可以进入查看该学生的详细信息，包括住宿情况，一卡通消费情况，奖惩助贷情况，以及图书借阅明细，成绩明细。



##### 资产信息设置

按照表格的形式统计展示所有资产的分类、名称、数量，也可以对资产进行修改维护。



#### 一卡通消费统计

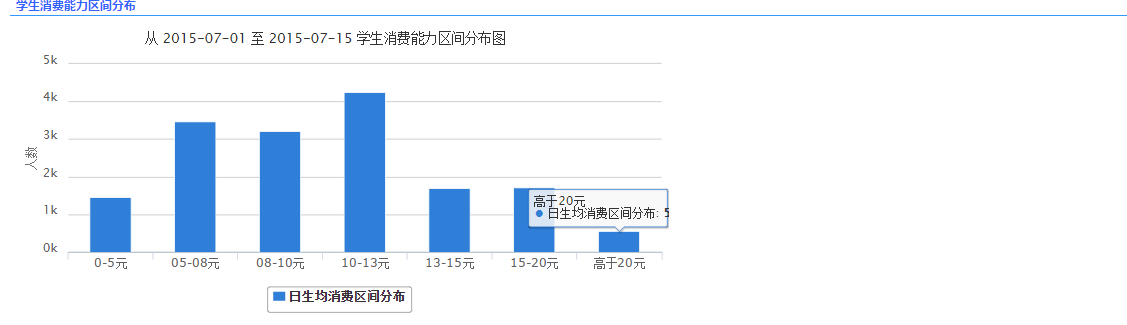
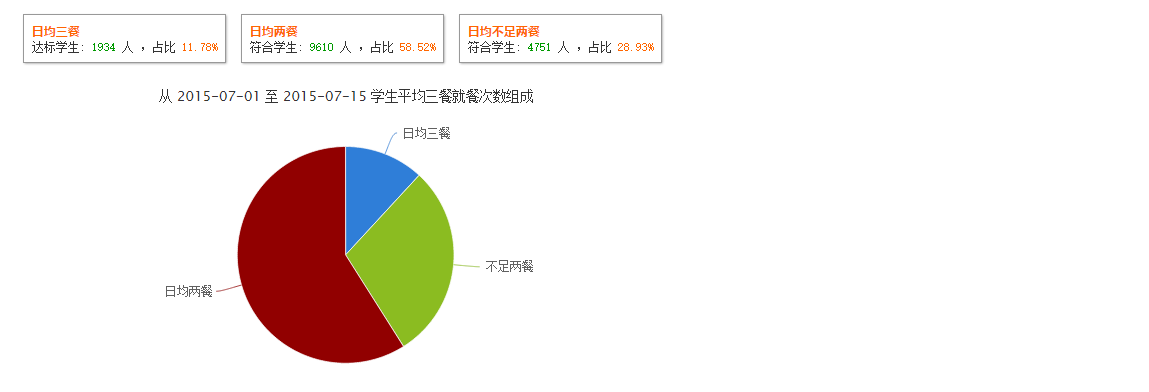
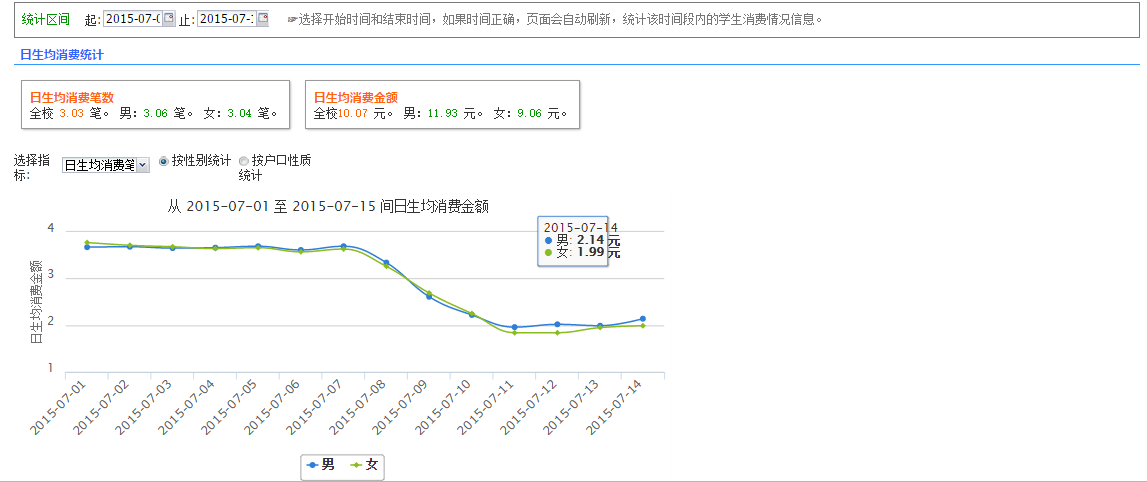
##### 学生人均消费概况

1. 以时期为维度，按照文本加数字的形式统计展示学生某一时间段内的消费总额、男女生消费占比、贫困生非贫困生占比、农村城镇户口的学生占比。
2. 以时期为维度，按照线形图或条形图的形式统计展示学生消费情况统计，包括一天内学生人均消费额、学生消费总额、消费总人数。
3. 以时期为维度，按照饼状图的形式统计展示早中晚三个时段的消费情况。
4. 以时期为维度，按照表格的形式统计展示各消费部门在早中晚三个消费时期的消费额和占总消费额的比例。



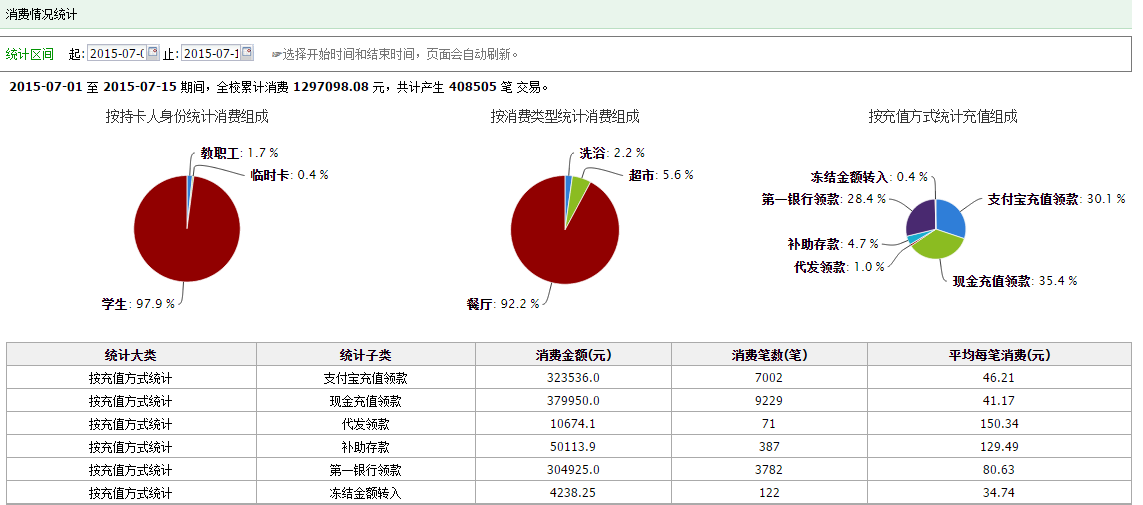
##### 学生消费习惯分析

1. 以时期为维度，按照文本、数字和线性表的形式统计展示了学生日均消费笔数和学生日均消费额。其中学生日均消费笔数包括学生日均消费笔数和男女生的消费笔数，学生日均消费额包括学生日均消费额和男女生的日均消费额。
2. 以时期为维度，按照文本、数字和饼状图的形式统计展示了学生就餐习惯的分析。就餐习惯的分析是从日均三餐、日均两餐、日均不足两餐的学生所占比例分析。
3. 以时期为维度，按照文本、数字和条形表的形式统计展示学生消费能力区间分布。学生的消费能力区间按照日均消费来分类。



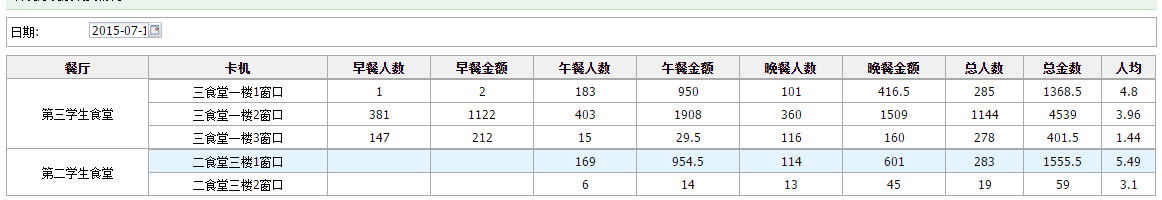
##### 消费类型分类统计

1. 以时期为维度，按照文本和数字的形式统计展示某一时期全校累计消费的总金额和总交易笔数。
2. 以时期为维度，按照饼状图的形式统计展示了某一时期持卡人身份消费所占消费总额的百分比。持卡人身份包括：教职工、学生以及临时卡。
3. 以时期为维度，按照饼状图的形式统计展示了某一时期消费的类型所占消费总额百分比。消费类型包括：餐厅、洗浴以及超市。
4. 以时期为维度，按照饼状图的形式统计展示了某一时期的充值方式占充值总额的百分比。充值方式包括：冻结金额转入、支付宝充值领款、现金充值领款、代发领款、补助存款、第一银行领款。
5. 以时期为维度，按照表格的形式统计展示了某一时期各统计大类以及子类的消费金额、消费笔数、平均每笔消费金额。



##### 餐厅日消费明细

以时期为维度，按照表格的形式统计展示了某一时期各餐厅的消费情况。消费情况从以下几方面分析：早餐人数、早餐金额、午餐人数、午餐金额、晚餐人数、晚餐金额、总人数、总金额、人均消费金额。

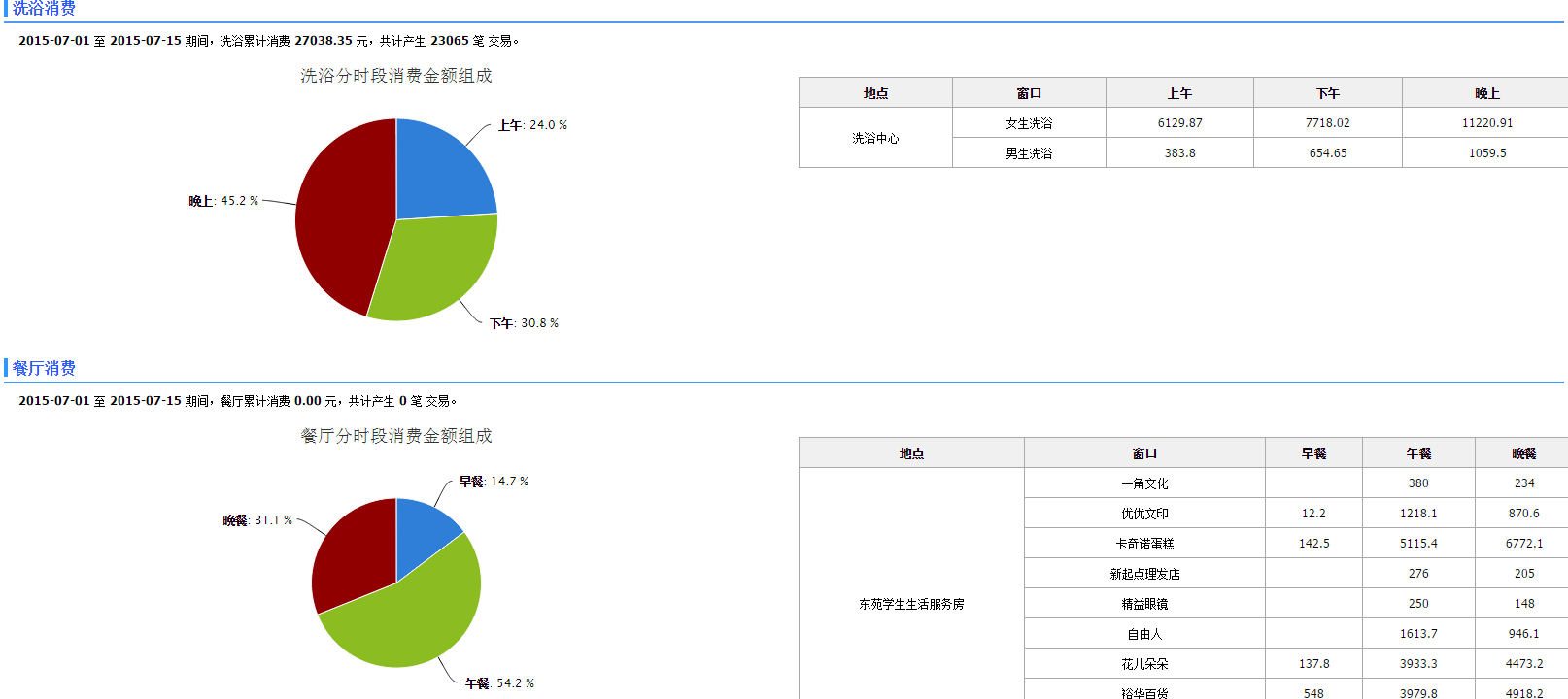


##### 分时段分类型消费统计

1. 以时期为维度，按照文本、数字、饼状图、表格的形式统计展示超市类型的消费情况。消费情况从消费时段消费的金额占总金额的百分比和消费时段的消费金额分析。
2. 以时期为维度，按照文本、数字、饼状图、表格的形式统计展示洗浴类型的消费情况。

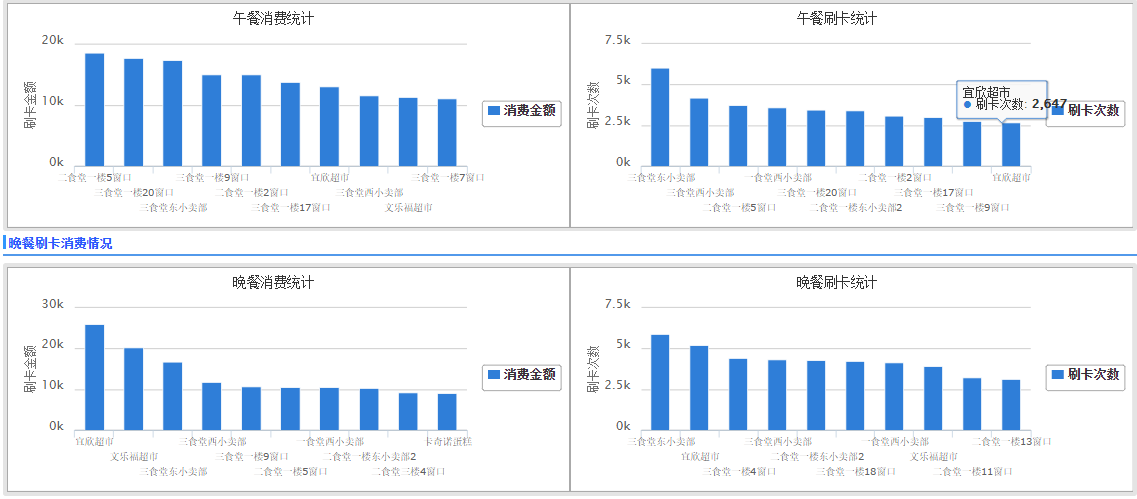
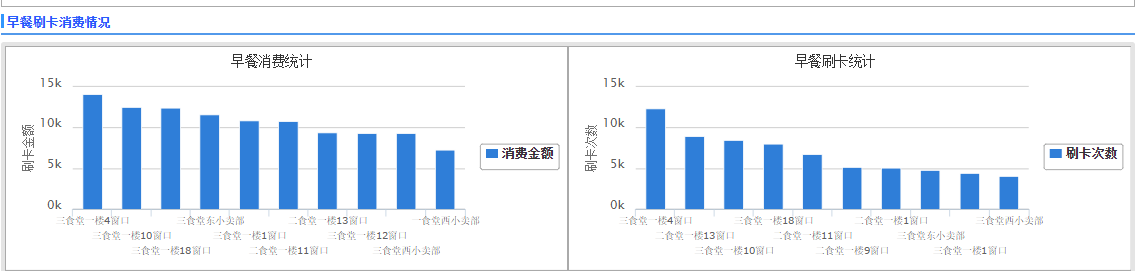
消费情况从消费时段消费的金额占总金额的百分比和消费时段的消费金额分析。

1. 以时期为维度，按照文本、数字、饼状图、表格的形式统计展示餐厅类型的消费情况。消费情况从消费时段消费的金额占总金额的百分比和消费时段的消费金额分析。



##### 学生用餐窗口排名

1. 系统将以学校、院系、专业为层级，以时期为维度，按照条形图的形式统计展示早餐的刷卡消费情况。刷卡消费情况从餐厅每个窗口刷卡消费的金额和刷卡消费的次数分析。
2. 系统将以学校、院系、专业为层级，以时期为维度，按照条形图的形式统计展示午餐的刷卡消费情况。刷卡消费情况从餐厅每个窗口刷卡消费的金额和刷卡消费的次数分析。
3. 系统将以学校、院系、专业为层级，以时期为维度，按照条形图的形式统计展示晚餐的刷卡消费情况。刷卡消费情况从餐厅每个窗口刷卡消费的金额和刷卡消费的次数分析。



#### 学生住宿情况全方位统计分析

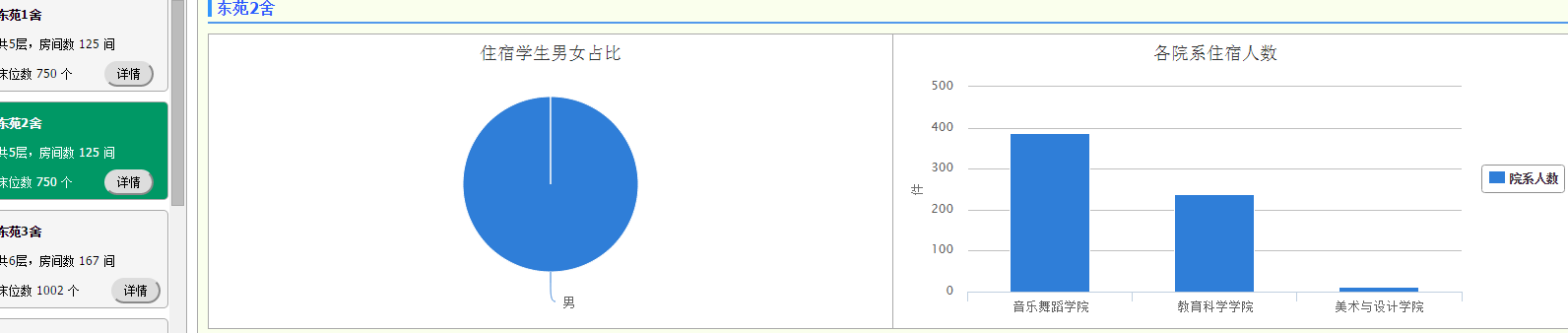
##### 在校生住宿概况

按宿舍楼统计各宿舍楼的住宿性别、容量、住宿标准、入住率、各学院的住宿比例，展示各宿舍楼学生住宿概况。



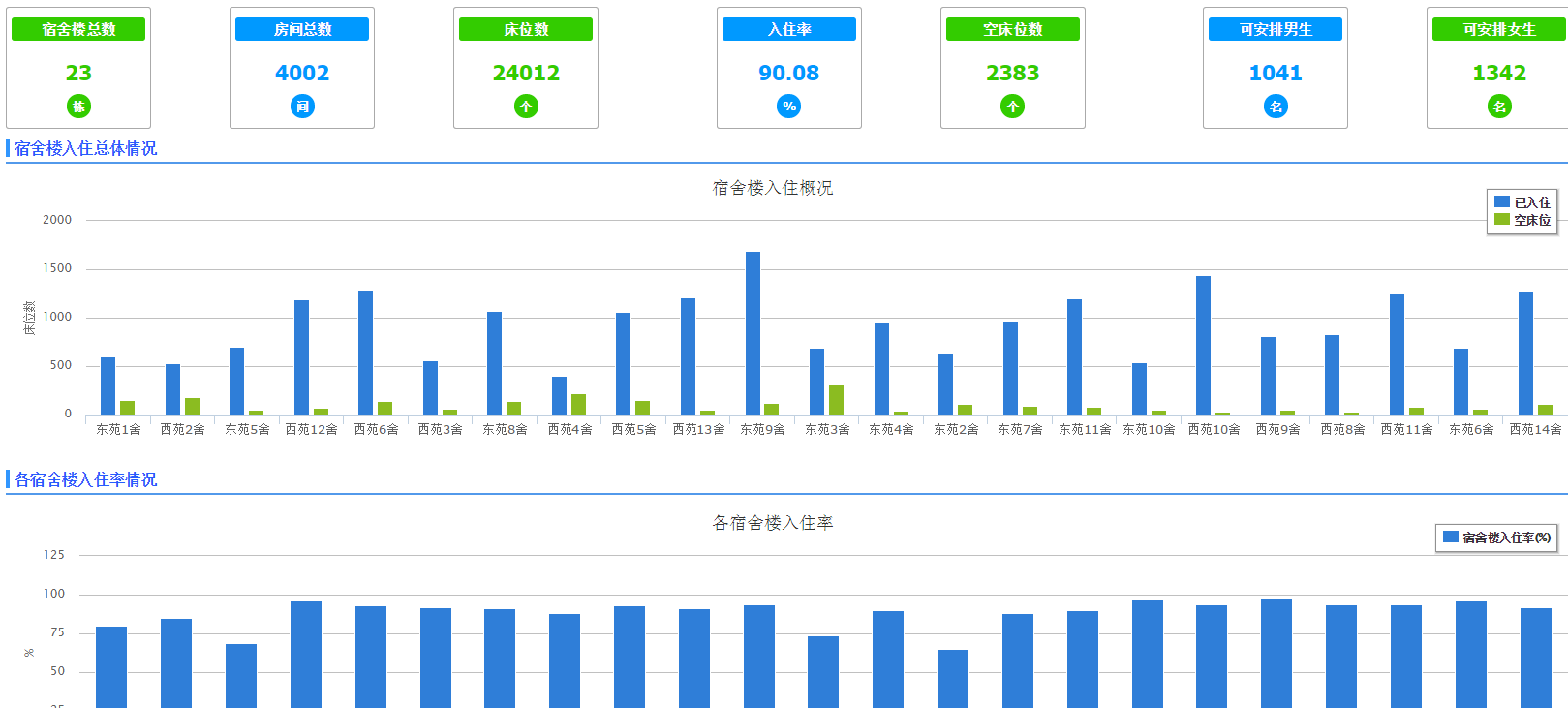
##### 在校生住宿明细

按宿舍楼为层次，以饼图、柱状图等、表格形式，以院系、专业、性别等不同维度展示在校生住宿情况。



##### 宿舍楼概况统计

统计宿舍楼总数量、房间总数、床位总数、总床位数、入住率、可再安排男女人数，以柱状图形式展示各宿舍楼入住概况及入住率。



##### 在校生住宿情况

以表格形式展示各学院在校生住宿人数、住宿占比及各年级学生的住宿分布情况。



##### 统计学生进出宿舍状况。

与现有一卡通宿舍门禁系统联动，按时间段统计各宿舍楼学生每天进出宿舍次数及计算高峰时段。



##### 预留与学校同步部署的宿舍管理系统接口。

预留与学校同步部署的宿舍管理系统接口服务，以便与学生宿舍管理数据进行数据集成进行数据挖掘及统计分析。

#### 学校资源概况统计

统计学校的教学组织结构、行政组织结构、科研机构、附属单位等信息，占地面积、教学楼、宿舍楼、图书馆、教师队伍等方面统计分析学校资源，并以文本、图表、表格的形式生成学校资源概况统计。



#### 招生迎新概况统计分析

按生源地、考生类别、性别、年龄分析统计招生信息并分析学生的报到率、未报到率以及报道高峰期等。

历年各专业、院系的学生报到率、未报到率的趋势对比曲线。

#### 基于以上主题的交叉式统计分析

例如以学生消费习惯与学生成绩进行的交叉式统计分析，以点阵图的形式展示学生早上8点以前吃早饭与学生成绩之间的关联性研究分析。

#### 自定义统计分析

用户可以自定义维度统计自己想要查看的数据，自定义组合扩展新的分析功能，充分满足客户个性化需求。用户可以根据自己的意愿组合自己的私有或共有的统计页面。公有统计页面经过管理员审核后可以允许相应权限的用户共享访问。用户自己创建的私有页面只有自己具有查看权限。

#### 预留及扩展统计分析

预留以上各子主题的统计分析接口和功能。上所有子主题内明细统计分析项（一个统计图形或表格为一个明细统计项）的详细定义，在建设过程中需要根据我校的实际需求进行响应，明细统计项的总数量将达到90多项，涉及统计字段500多个。

## 教学评估分析系统建设内容

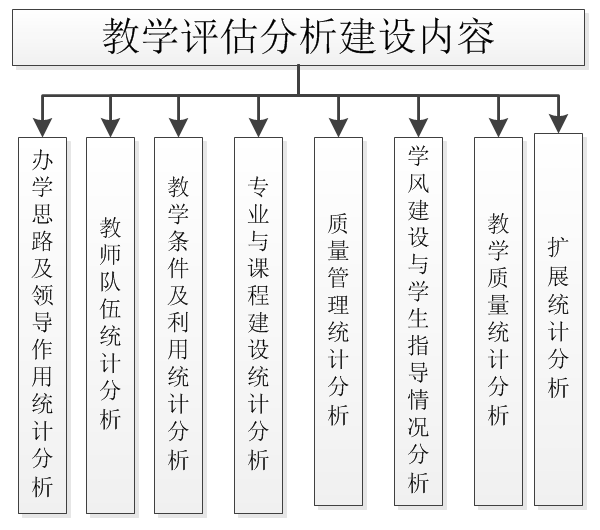
教学评估分析系统从教学评估所关注的狮子队伍、教学条件、教学管理、主页结构与布局、学风、教学效果等多个方面对高校进行全方位的数字评估，对评估指标进行有效跟踪，自动生成评估结果，并以报表、flash图表为依托清洗直观地展现各个评估指标的情况。同时系统还可以针对每一个指标设置预警阈值，可以实时展现各指标和评估指标质之间的差异。

教学评估分析系统采用程序导入，系统自动同步，手工录入等多种数据采集方式，并通过数据集成平台，把学工系统、教务系统、人事系统、招生系统、迎新系统等各业务系统中包含的部分数据信息建立了一个统一的共享数据库。数据库中的数据依据各个业务系统的真实数据实时更新，评估报告和分析图表再根据这些数据动态生成。教学评估分析系统保证了数据信息的准确性和及时性。

实现目标：

1. 让评估专家通过分析图表全面了解学校整体建设情况；
2. 让校领导实时跟踪教学评估相关指标数据，掌握每一个评估点的真实数据，也可以进行实时自评，为教学的科学决策提供依据；
3. 用智能化的预警手段来对未达标指标进行实时监控。

教学评估分析系统建设概览图



#### 办学思路及领导作用统计分析

系统支持学校定位、领导作用、人才培养模式等方面进行详细分析。

##### 学生就业情况分布

1. 不同学历学生就业分布情况
2. 不同性别学生就业分布情况
3. 不同民族学生就业分布情况
4. 不同专业学生就业人数、比率分布情况，专业之间进行横向对比就业率。
5. 不同学生类别就业人数、比率分布情况，本科生、专科生、研究生之间横向比较就业人数及就业率。

##### 学生考研情况分布

1. 不同学历学生考研分布及通过率情况。
2. 不同性别学生考研分布及通过率情况。
3. 不同民族学生考研分布及通过率情况。
4. 不同专业学生考研人数、通过比率分布情况，专业之间进行横向对比考研人数及通过率。

##### 校领导年龄分布

1. 按年龄段分段统计校领导年龄分布统计
2. 历年校领导年龄段分布统计

##### 校领导在职年限分析

1. 校领导在职年限分析统计
2. 校领导在职年限分组分析统计

##### 预留扩展统计分析

预留办学思路及领导作用统计分析接口。

#### 教师队伍统计分析

##### 教师的数量与结构

1. 不同学位的教师分布比例
2. 师资队伍职称结构
3. 师资队伍年龄结构
4. 师资队伍学历结构
5. 省部级重点学科师资队伍结构
6. 新专业师资队伍结构
7. 基础课程师资队伍结构

##### 教育教学水平

1. 生师比，展示学生分类以及折合学生数、教师分类以及折合教师数据。
2. 专任教师、主讲教师、教授副教授比例。
3. 基础课程及格率、优秀率分析。
4. 学生参加基础学科竞赛获奖情况分析。

##### 教师培养培训

1. 教师培养培训周期统计
2. 教师培养培训频率统计
3. 教师培养培训证书统计
4. 教师培养培训成绩统计

#### 教学条件及利用统计分析

##### 教学基本设施

1. 校舍面积分析
2. 教室拥有情况分析
3. 体育场馆面积统计分析
4. 体育场馆设施统计分析
5. 年藏书量和年进书量统计分析
6. 文献利用情况统计分析

##### 各种经费的投入对教学条件和利用情况、

1. 四项经费占学费收入比例分析
2. 四项经费情况统计分析
3. 四项经费增长情况统计分析

#### 专业与课程建设统计分析

##### 专业建设

展示各类专业数及占专业总数比例。

##### 课程与教学

展示近三年以下内容的数据及这些数据的变化趋势。

1. 选用教材总量
2. 选用优秀及获奖教材数量
3. 选用优秀及获奖教材占选用教材总量的比率
4. 选用新教材数量
5. 选用新教材数量占选用教材总量的比率
6. 展示近三学年必修课利用多媒体授课课时比率及其变化趋势。

##### 实践教学

展示近三学年含综合性、设计性试验课程比例统计数据及其变化趋势。

#### 质量管理统计分析

##### 教学管理队伍

1. 管理队伍统计分析
2. 实践教学统计分析

##### 质量监控

质量控制统计分析

#### 学风建设与学生指导情况统计分析

##### 学风建设

1. 大学生电子设计竞赛获奖情况分析
2. 学生参加大学生数学建模竞赛获奖情况分析
3. 教师风范
4. 学习风气
5. 实践教学

##### 指导与服务

展示近三年、获得河南省、全国普通高等学校本专科优秀毕业论文（设计）的数据统计和分析。

#### 教学质量统计分析

##### 德育

展示近三年、各个年级的本科生、专科生、研究生党员数量及入党比率。

##### 专业知识和能力

展示近三年、获得河南省、全国普通高等学校本专科优秀毕业论文（设计）的数据统计和分析。

##### 体育美育

展示近三年，本科生、专科生、研究生《学生体质健康标准》测试情况统计分析及其变化趋势。

##### 校内外评价

展示近几年，在全国30个省（市、自治区）新生录取情况统计及其变化规律分析。

##### 就业情况

展示近三年学校毕业生按院系、专业、学生类别、性别维度统计就业率及其变化趋势分析。

#### 预留及扩展统计分析

以上所有子主题内明细统计分析项（一个统计图形或表格为一个明细统计项）的详细定义，在建设过程中需要根据我校的实际需求进行响应，明细统计项的总数量将涵盖我校“教学评估分析报告”中的所有内容。支持并解决数据从学校本科教学评估数据采集系统到中心库的集成。

## 原有业务系统的集成及数据整改方案

#### 解决问题

通过数据集成交换平台进行数据清洗、转化、集成可以解决目前学校存在的问题：

一是，解决部分系统未与数据平台进行数据对接，数据平台中既采集不到这些系统的数据，也无法为这些系统提供基础数据的问题。

二是，解决已对接的部分系统，由于业务部门在使用系统时的原因（未使用或仅使用部分功能），导致部分基础性或共享性、交叉分析性质较强的业务类数据缺失的问题。这些数据和信息的缺失，导致目前学校的数据链条不够完整，尤其是关键性信息的缺失，导致数据链断裂为孤立的几段，无法拼接成完整的信息图谱，不利于数据的长期积累。

三是，解决已对接并能够正常流入数据平台的数据中，由于缺少数据清洗和数据质量的管控，还存在不少的错误数据或流入的数据不够完整（字段信息不完整或在时间上不连续）问题。搭建数据质量监管系统，不断的进行数据清洗和数据质量监管，以避免随着时间的推移，造成错误数据的积累和数据质量的下降。

#### 功能描述

**集成中心库模式**

模式以教育部《高等学校管理信息标准CELTS-33》为基础，结合精华教育过去多年集成经验进行设计。从总体上对高校数据进行梳理，在高校信息管理的数据层面按照面向对象进行设计。

高校往往将数据信息按照不同角度不同侧重点，分散在各个业务系统中进行管理，而集成中心库模式则将这些数据信息抽离出来，进行对象和业务类的拆分，将其分为对象类和活动类。

对象类是指数据对象，主要包括实体对象和逻辑对象。实体对象是指传统的数据对象，如人、活动等；而逻辑对象则是指一些非传统的数据对象，如院系信息等。

活动类包括对象管理活动和综合管理活动，可以按照对象的生命周期进行管理活动的设定。其中对象管理活动指基于对象的管理部分活动可以进行清晰的划分和归纳的数据，比如一些对实体对象的管理；综合管理活动是指包含的内容和相关联系会随着应用、定位以及数据使用的角度不同而变化的活动。

**参照模型标准及数据**

参照标准集是一系列机构标准的集合，主要包括国家标准、教育部行业标准等。

**执行标准模型及数据**

执行标准集是学校自定义标准的集合，俗称“校标”，是学校信息化建设中实际使用的代码标准集。一般情况下，执行标准是高校参考了参照标准，再结合学校自身的需求而制定的。

**集成接口支持**

数据集成平台为多种数据源提供接口支持，如下：

* 支持主流关系型数据库，包括Oracle、IBM DB2 UDB、IBM DB2/400、Informix、Microsoft SQL Server、Sybase AS Enterprise、Sybase AS Anywhere。
* 支持非主流关系型数据库，包括MySQL、Derby、Hypersonic SQL、PostgreSQL。
* 支持ODBC数据源类型的接入，包括Microsoft Access、Microsoft Excel、Dbase、Visual Foxpro。
* 支持主题或者队列，包括JMS Topic、JMS Queue。
* 支持WebService
* 支持Tabled-Txt文件
* 支持XML文件
* 支持操作系统的网络协议，包括 FTP。

**数据集成KM模块**

数据集成平台提供数据集成KM库，定义了各种场景通用的知识模块（KM），包括了各种数据集成需求开发包共计100多项，大大减少了数据的集成的工作量。

**集成设计工具**

集成设计工具是设计阶段开发人员和元数据管理员的主要工具，系统所有项目的开发都是通过这个界面完成的。

用户使用该工具，为数据转换和数据完整性定义声明规则，然后将业务系统中数据库中的元数据导入或定义，为实际的生产环境生成方案。

**集成查看工具**

集成查看工具是运行阶段的主要工具，开发人员在设计阶段也可使用该工具进行调试。

用户使用该工具查看执行日志，包括错误数、被处理的行数、执行的统计信息和被执行的实际代码等等，实现对生产环境的管理以及ODI监控。

**集成调度工具**

用户通过集成调度工具实现对各个数据集成同步任务的调度控制，以此完成定制化的数据集成过程。

**平台功能**

1. **接口层**

* **通用接口**

对于目前使用的各种数据源类型，制定相应的适配器，以满足连接不同数据源的需求：

* 支持主流关系型数据库，包括 Oracle、IBM DB2 UDB、IBM DB2/400、Informix、Microsoft SQL Server、Sybase AS Enterprise、Sybase AS Anywhere。
* 支持非主流关系型数据库，包括 MySQL、Derby、Hypersonic SQL、PostgreSQL。
* 支持 ODBC 数据源类型的接入，包括 Microsoft Access、 Microsoft Excel、Dbase、Visual Foxpro。
* 支持主题或者队列，包括 JMS Topic、JMS Queue。
* 支持 WebService，WebService接口提供了数据访问的最基本方法，即WebService方法，通过设定好的数据接口，直接提供数据访问的WSDL，或进行数据的运算，以便应用可以根据自身的需要进行访问和数据调用。
* 支持 Tabled-Txt 文件
* 支持 XML 文件
* 支持操作系统的网络协议，包括 ftp。

通过对数据源连接的简单定义，即可完成对数据源连接的建立，为下一步进行数据集成建立基本的通道。

* **专属接口**
* 对于特定的数据源，可以使用数据源特定的连接方式，充分利用数据库的特性，发挥效率和性能等方面的优势。
* Oracle OCI 方式
* Sun JDBC-ODBC bridge 方式
* XML 的 JDBC 方式
* DB2 JDBC Type2 方式
* **类型异构映射**

对于各种类型的数据源，针对异构数据类型进行完备的映射关系。内容包含以上提到的各种类型的数据源数据，从数据库到文件。

1. **抽取层**

* **反向模型**

通过反向模型功能，可以在数据交换共享平台中，对操作的内容进行数据定义，具体功能如下：

* 定义访问的表或者试图，可以是一个，也可以是多个。
* 提供系统表级的定义。
* 能够针对不同的物理架构（例如不同的版本），进行相应的反向工作。
* **模型定义**

模型定义是在反向模型的基础上，对模型进行重新定义，确保模型应用在数据集成过程中是完整准确的，具体功能如下：

* 模型的裁剪和补充，可以在原有物理模型上进行调整，以便适用于集成。
* 可以自定义键，以便在数据处理过程中进行调用。
* 提供了关联性内容的查看，了解当前模型在哪些项目中被引用。
* **增量设置**

对于数据模型，可以定义这个模型的 CDC（Change Data Capture，变化数据捕捉）策略，提供触发器、标识位、以及日志挖掘三种方法，具体包括以下功能：

* 数据订阅的定义，即变化数据的采集。
* 提供日记表自动生成，保存增量数据。
* 提供日记表、数据订阅功能的删除功能。
* 提供日记表内容查看功能。
* **约束条件**

对于数据源，可以设定约束条件，指定数据源获取数据的必要条件，包括以下功能：

* SQL 关键字的操作，例如 INSERT、ALTER 等。
* 字符串的操作，例如 LENGTH、REPLACE 等。
* 数学计算操作，例如 ABS、MOD 等。
* 日期和时间的计算和转换操作，例如 SYSDATE、EXTRACT 等。
* 系统参数调用。
* 聚合函数，例如 GROUP、COUNT 等。
* 类型转换函数，例如 CONVERT、TO\_LOB 等。
* 运算符支持，例如 +、\* 等。
* 平台系统函数，例如 getPK、getColumnList 等。
* 异常内容查看。

1. **转换映射层**

* **关联合并**

对源数据能够进行各种关联，并能够对字段进行合并拆分等操作，具体包括以下功能：

* 多对一的映射，即多个源表对应同一目标表。
* 多源表的关联，键的关联，自定义的关联。
* 字段的操作，拆分与合并。
* **智能映射**

源与目标的自动映射，包括以下功能：

* 源表内容自动添加到目标表。
* 界面内容拖拽式映射。
* 相同别名的自动映射，询问机制。
* 映射异常的自动预警、提醒，例如数据类型不匹配，目标数据长度比源数据短等。
* **清洗转换**

支持开发语言级别的数据清洗转换，包含以下功能：

* SQL 关键字的操作，例如 INSERT、ALTER 等。
* 字符串的操作，例如 LENGTH、REPLACE 等。
* 数学计算操作，例如 ABS、MOD 等。
* 日期和时间的计算和转换操作，例如 SYSDATE、EXTRACT 等。
* 系统参数调用。
* 聚合函数，例如 GROUP、COUNT 等。
* 类型转换函数，例如 CONVERT、TO\_LOB 等。
* 运算符支持，例如 +、\* 等。
* 平台系统函数，例如 getPK、getColumnList 等
* **联合配置**

针对映射过程，需要对特定的内容进行操作的联合配置，包含以下功能：

* 指定对插入或者更新操作进行同步。
* 指定联合的字段集设定，以便进行特定的操作。
* 指定映射过程的实现位置，源、临时区域或者目标。
* 指定目标的主键。

1. **加载层**

* **批量操作**

数据加载过程使用数据批量操作，一次提交多条数据，提高集成效率。

* **事务定义**

在数据加载的过程中，可以定义事务何时操作。

* **加载策略**

定义数据加载的各种策略，包含以下功能：

* 目标表是否自动创建。
* 目标表是否删除数据。
* 是否引入条件。
* 是否使用 truncate 操作。
* 插入、新增操作是否同步。
* 日记表是否同步。

1. **运行设定层**

* **运行周期**

提供多种项目运行的时间定义方法，包含以下功能：

* 启动时执行，满足轮询策略。
* 简单某一时刻运行。
* 周期执行，周期可以定义为小时、天、周、月、年
* 循环执行，可以只执行一次，也可以多次，也可以设定最大重复次数。
* 执行约束，例如失败一定次数即停止，也可以设定失败条件。
* **场景流程**

定义场景，在场景中可以定义接口的执行流程以及相应的策略，包含以下功能：

* 定义流程包含的模块（接口）。
* 定义模块运行先后顺序。
* 定义流程执行的策略，例如间隔时间、触发事件等。
* 定义流程执行成功和失败的步骤，以及策略
* **外部调用**

定义了需要引用外部的资源调用方法和配置信息，包含以下功能：

* 网络组件，包含 WebService 调用、邮件读写调用。
* 事件检测组件，包含等待数据、等待日志数据、等待 session 等调用。
* 元数据组件，包含取元数据、设置元数据以及反向表调用。
* 日志数据组件，包含日志数据记录和日记数据调用。
* 文件组件，包含文件修改、拷贝、删除、压缩解压等调用。
* 实用组件，包含操作系统调用、场景设定以及代理等功能调用。
* **版本管理**

对所定义的项目、策略、流程以及运行等方面的内容，提供版本管理，包含以下功能：

* 新建版本。
* 版本还原。
* 版本导入导出
* **结果查询**

提供项目、场景的运行结果的查看，包含以下功能：

* 根据不同的方式查看项目、场景的运行状态，例如时间、用户、会话、项目等。
* 可以查询集成的每个步骤，步骤的执行结果、执行时间、执行内容以及错误信息。
* 提供异常项目的重启功能。

1、 包括但不仅限于：一卡通、人事、教务、学工、工作量、评估教学，具体根据两大主题数据挖掘系统所涉及基本数据的要求而定。

2、 对一期工程已实施身份对接而未进行数据对接的系统进行数据分析，采集、集成其可用数据，包括论文系统、工作量统计系统、教学档案管理系统、实验室与设备管理系统、网络教学平台系统。

3、 集成学校现有其他业务系统，包括SAM系统、微哨系统等

## 数据质量监管系统

请将《数据质量监管平台 - 精华科技.doc》内容合并在此