

V8.0 A类赛题答疑（截止至 20200212，标黄为本周更新）

【A01】基于绿色出行场景个人碳排放积分系统开发与运营【八维通】

1、碳积分运行系统。我们是单独做一个 app 或者微信小程序，还是在八维通的 APP 里面内嵌我们的系统模块呢？

答：内嵌的，可以考虑 h5 或者 sdk 的方式

2、在所需数据接口中共有六个接口，请问支付接口中的微信和支付宝代扣需要自己实现吗？

答：不用

3、在所需数据接口中共有六个接口，请问支付接口中的微信和支付宝代扣需要自己实现吗？

答：不用

4、赛题 demo 统一 api 接口访问所需的 appid 和账户有提供么？

答：api 地址已在群文件中公布，为了方便开发，没有设置验证，http post 请求调用即可，接口返回的也都是模拟数据。在模拟环境 API 准备时，为了简化工作，去掉了接口验证要求。现在所有参数都不用带，直接调用接口即可。

5、关于碳积分的定义，题目中的碳积分的定义指二氧化碳排放配额，企业或个人可通过购买碳积分消除碳足迹，是一种惩罚措施。而本题的碳积分是鼓励绿色出行的，绿色出行的效果减少碳排放，是不是可以定义为碳积分的量化是绿色出行带来的减碳量，而不是一般的碳排放量来量化的。

答：是的，通过绿色出行带来的碳的减排量获得奖励（政府或公益组织提供）碳积分，进而获得商业价值。可以通过碳积分交易，卖给有碳排放需求的用户变现其价值。

6、是否会提供商家入驻、支付等相关接口？

答：不提供。可以自己做一些模拟数据和操作，不用实际实现。

7、在 api 接口地址中，能查到行程信息，但是没有 user_id 和 user_name 等用户信息，返回 message: "Required String parameter 'user_id' is not present", path: "/v1.0/api/app/user/getUserInfo"，请问用户信息表是本来

就没有吗？

答：userid 值取 1，在请求头中赋值。见文档更新。

8、在提供的文档，“验签过程”中提出我们应该提交 json 格式的数据，但是现在给出的接口里面只能提交明文数据。如果提交明文的话，服务器端收到数据后，要按照什么格式转换为 JSON 进行签名校验，然后 null 字段要不要计算？各个字段按照什么顺序排列？是否需要换行？

答：已取消验签环节。见群文件接口文档更新。

9、如果是开发 sdk 的话，演示的时候是不是可以做个简易的 app 入口来仿八维通，在这个入口然后调用 sdk，传入相关数据进行演示，实现嵌入 app 效果？

答：可以，请自行设计，不做要求。

10、贵公司是否可以提供地铁或公交的实时到站信息或者我们是否可以自己找数据。

答：模拟环境接口暂时不通知。可自行模拟，加以说明即可。

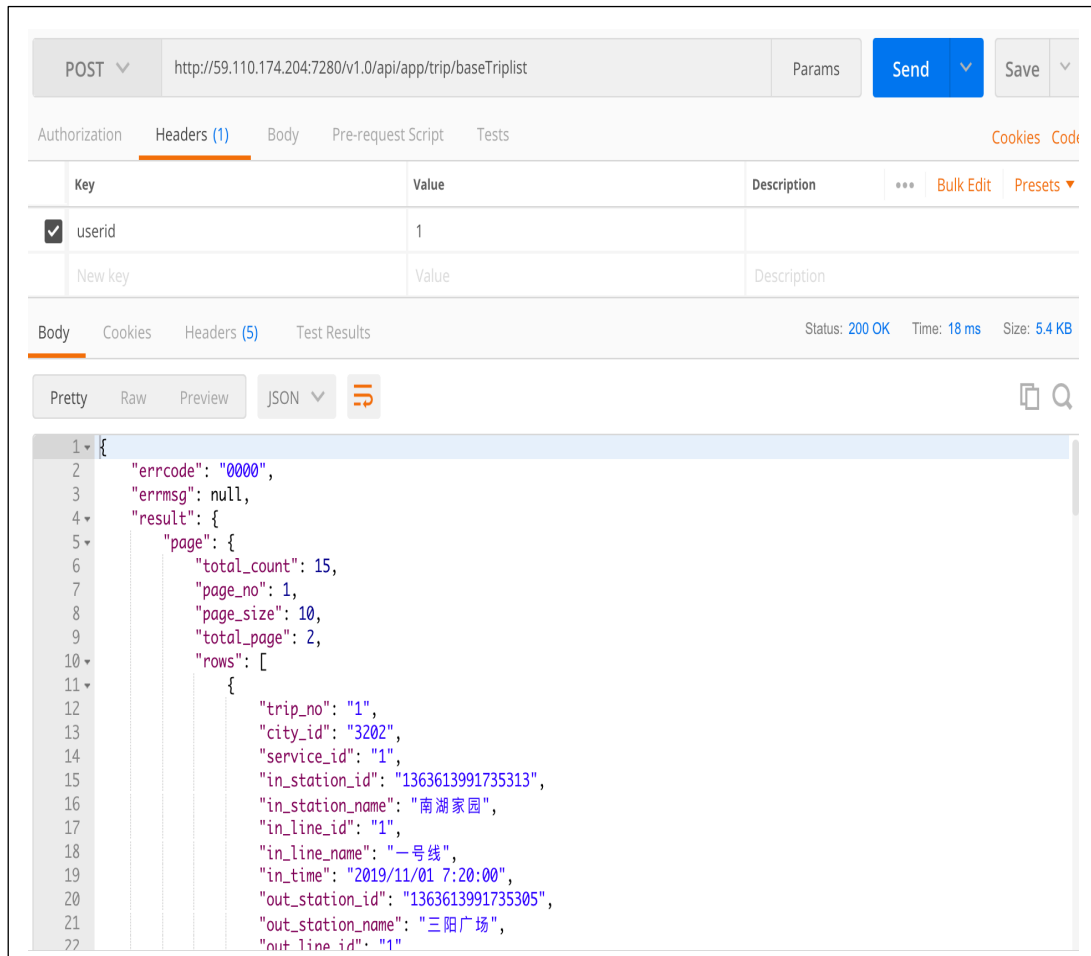
11、请问贵公司该项目比较看重的成果是偏向商业还是偏向技术

答：偏向商业设计。

12、行程列表接口调用不行，请求：

http://59.110.174.204:7280/v1.0/api/app/trip/baseTriplist 请求 body
service_id 00 trip_flag 00 page_no 1 page_size 10 userId 1 token 1111
响应内容： { "timestamp": "2020-01-08T02:44:51.883+0000", "status":
400, "error": "Bad Request", "message": "Missing request header
'userId' for method parameter of type String", "path":
"/v1.0/api/app/trip/baseTriplist" } 传了 userid 是 1 还是不行。工具用的
postman

答：header 中传userid，见截



13、碳积分对象是个人用户还是包括企业，对于企业是自愿性质的吗？怎么去监管，这个 APP 数据库需要自己完善吗？个人或企业通过购买碳积分消除碳足迹，企业为什么会购买碳积分？碳积分是怎么进行流通的，卖给有需求的客户可以举例吗？

答：不包括企业，主要围绕乘坐公共交通人群。App 数据库不需要自己维护，与八维通平台合作。碳排放积分可以到碳交易所交易变现。碳积分交易可以设计多种玩法：如对换乘车券，对换其他奖品。需要参赛选手广开思路，也是评分亮点！

14、界面设计是可以设计八维通的界面一起还是只需设计碳积分的界面？

答：只需设计碳积分相关页面

15、地铁只有无锡的，是只需要我们做出无锡的吗？还是其他城市需要我们自己找？

答：以无锡为例给出 demo 即可，不用考虑其他城市

16、怎么知道用户使用什么工具，需要相应的 API？还是直接调用定义的取值？

答：请按文档直接调用定义的取值，做出 demo 程序即可

17、公交站点和地铁站是一样的？

答：暂不考虑公交站点.

18、看到你们给出“碳交易所”，这个碳交易所是现有的，还是要我们自己写呢？碳排放量转换成碳积分的数值模型需要我们自己写吗？

答：是已有的。不需要考虑数值模型。

19、在已给的接口中没有步行数据，是否会给我们步行数据呢？

答：接口暂不提供，可以自行模拟设计

20、关于企业已经发布的码上行 app，是要求参赛团队根据已有的 app 来设计个人积分系统吗？参赛项目名称需要与 app 保持一致吗？

答：这个系统，最终会是码上行的一个相对独立的模块/部分，而不是一个 app。积分系统设计不用受现有 app 太多限制（当然业务场景是绑定的）。

21、请问绿色出行方式是单指无锡市地铁，还是包含公交、共享单车等所有绿色出行方式？

答：设计上，可以都包含。本次的接口中暂无相关数据，可以自行模拟，也可以只在设计上体现。

【A02】基于地铁出行场景的小游戏设计【八维通】

1、请问这个游戏最终是嵌入现有平台使用呢？还是单独做一个游戏？，如果是嵌入现有平台使用，游戏大小以及风格有无限制？

答：基于接口做独立的 h5 应用。

2、最后测试时，可以拿到 app 来先测试吗？

答：不用集成到 app 测，就基于接口测 demo 就可以。

3、请问一下，企业能否考虑提供 H5 开发平台？，介于题目要求做独立的 H5 运用，现阶段网上的各大 H5 开发平台均不是免费开放。

答：“独立 H5 应用”不等于要使用第三方平台。理解成是要做一个可以独立

运行、前端基于 H5 的 demo 应用系统，不要求在互联网上可访问。所以不提供，也不希望使用网上的开发平台。

4、demo 程序是仅供观看还是供试玩？提交 demo 程序材料时以什么格式提交？
请问程序开发到什么程度？

答：需要基本可试玩的 demo。需要可运行的程序。核心设计理念得到体现、核心流程可走通即可。

5、赛题中给出的考虑后台参数配置、数据统计需求，是需要做出单纯只是小游戏的系统参数配置还是需要整个系统的管理后台

答：小游戏相关的配置管理。前期进行考虑，最终给出相关设计即可，不要求必须实现，有能力的团队可尽量实现。

6、是否要求该小游戏只能是在地铁乘车的场景下游玩？

答：要求是紧密结合地铁场景，不会要求出了地铁不能玩，如果能全天一直玩当然更好:)

7、有个关于接口的问题想询问贵公司，是否能够通过已有的接口获取列车运行信息，如还有几分钟到站这样的信息数据。

答：模拟环境接口暂时不通知。可自行模拟，加以说明即可。

8、是否能获取到用户当前时刻行进到的地铁站台的信息？

答：接口暂时不提供，但可以自行模拟这个信息，并加以说明。

9、说明中称会提供的“会员等级积分接口”未在《A01+A02-八维通-赛题所需数据接口》中找到，是否需要自己搭建后端存储用户积分数据？此处的“积分”是否是指的游戏中的虚拟货币？

答：八维通会员等级积分可以认为是常见的用户积分，通过日常使用或活动积累，可用于积分商城兑换。不要求在 demo 程序中对接使用，可以在方案设计中给出未来怎么做结合。游戏中的虚拟货币体系可以自行设计，建议可以给出和八维通积分打通的设计方案。

10、接口中的居住地站点、工作地站点、常去站点信息是否为乘客乘车一段时间后才能得到的统计结果？是否一定准确、随时可用？

答：是的。不是每个用户都有这几个画像信息。

11、我们团队想基于已有的游戏引擎进行游戏的开发，这个引擎是开源免费的，

是否可以使用，或者必须使用原生开发？

答：可以使用。

12、因为目前还是没有提供对应的游戏程序的 API 接口，所以我想问一下到时初赛时提交的游戏 demo 有什么具体的要求？比如说是要什么格式，运行环境是如何啊，之类的。

答：要求是做一个 H5 的游戏 demo，环境没有要求

13、请问如果游戏的形式结合了地铁出行，但游戏的内容、背景和地铁关系不大，就是游戏如果不基于地铁出行可以玩但很难进行，但游戏的内容和地铁本身关系不大，差不多就游戏界面有地铁的图片，这样行不行？

答：可以，没有硬性要求。

14、集成 AR 是否是直接调用别人的功能，如导航，还是结合自己的应用来改变？

答：可以应用第三方方案，但如果是收费的方案应慎重并说明。

15、周边 POI 的获取，能否调用其他 App 的接口，例如百度地图等？

答：可以，应加以说明。

16、游戏中的素材（如人物形象等），能否使用别人现成的素材。如使用，算是抄袭吗？

答：demo 中可以，但建议尽量避免，并不要做为主要亮点。

17、要运用 AR 技术，最终游戏导出能否做成 SDK 的形式？对于 AR 设计的实现需要基于 H5 实现，还是可以导入到 app？

答：可以做成 SDK。

18、行程列表中，请求内容那块的数据宽度的类似 V4 F4 的含义？

答：代表数据类型及长度。模拟环境中可忽略。

19、对于基于 H5 的小游戏开发和 AR 可以通过集成 SDK 实现这两个点 可以在 App 连接吗？

答：只要 demo 是相对独立的模块，与 APP 本身尽量松耦合

20、在企业的介绍视频中，跳一跳这款游戏作为反例出现。那么假设我们将跳一跳这款游戏结合上地铁场景和相关激励机制，是否符合企业的要求？（当然我们最终不会做一款跳一跳，我们只是担心是不是所有和跳一跳游戏模式类似的游戏不能满足要求）

答：可以，但可能不是特别好，希望游戏本身跟场景能有比较大的关联。

21、在调用接口的过程中出现了跨域问题，报错 No 'Access-Control-Allow-Origin' header is present on the requested resource. 请问贵公司有添加相关的设置么？

答：接口设计是针对后端系统调用的，建议开发一个简单的后端，从后端调用我们的接口，就可以避开跨域问题。

22、请问可以写入用户游戏信息吗？接口文件中没有找到定义规范。

答：应该说是必需的，需要自行构建一个系统来支持。接口文件是我们提供的用户基础信息和辅助信息。

【A03】工业物联网——高性能边缘计算轻量级容器管理引擎设计【文思海辉】

1、文思海辉赛题讲解视频要求提交的作品是倾向于边缘计算的 web 容器管理工具，而赛题手册中赛题是要求设计一个应用容器引擎，请问以哪个为准提交作品？

答：最终交付以最后发的文档为准，web 的交付耗时，不存在技术上的难点，需要学生实现的是边缘计算应用容器的设计，可以没有 web。

2、根据我们团队调研发现，当前市场上的边缘计算网关基本都是商用的，不能提供开发环境，请问是否可以选择树莓派作为运行测试环境，如果不行，在硬件方面，贵公司有什么更好的建议吗？

答：关于边缘计算的硬件设备可以是 X86 或者 ARM 架构的，用树莓派也可以。最终交付的软件尽量与硬件平台无关。

3、赛题中说到实现容器引擎可参考 Docker 和 Rocket，请问是否可以在现有的 Docker 容器管理引擎的开源代码上进行进一步的修改开发，还是说需要从头开始重新编写一套管理引擎？

答：可以参考 Docker 和 Rocket 的实现，这两者目前在云计算架构已经较为成熟。在边缘计算设备上两者较为重量，可以参考两者进行裁剪，以适用于嵌入式设备。重新编写管理引擎具有挑战性，有团队希望较高难度的挑战可以尝试编

写管理引擎，这部分可以参考 Linux LXC LXK。

4、请问贵公司在该容器管理引擎的具体应用场景上有什么需求，我们团队希望在设计的时候将这一块因素考虑进去，更有针对性地满足贵公司的需求。

答：容器管理引擎的场景，解决同一设备的应用间的隔离，比如” AI 计算程序”，” 数据采集程序”，” 计划管理程序”，这些程序的厂商不同，需要通过容器隔离保证数据的隐私性。

易于维护性，边缘计算程序运行与容器中，对原系统的侵入性较小，易于更新与维护。

5、如果是要设计一个基于解决边缘计算问题的应用容器引擎，那么可能更多的注意力应该是放在如何解决边缘计算相关的问题上，但是这些具体的计算问题应该是由内置在容器内部的应用去解决，而不是应用容器引擎本身，照理说应用容器引擎本身只给各个容器提供一个运行调度等功能的平台，它面对的对象是容器，而解决问题的主体是容器，而不是应用容器引擎本身。如果赛题是要设计一个基于解决边缘计算问题的容器编排管理调度平台，那么感觉还是有思路的，但是赛题说的是设计应用容器引擎，这让我十分疑惑，希望命题组能够予以解惑，万分感谢！

答：可以理解为”基于解决边缘计算问题的容器编排管理调度平台”，本次赛题只需实现容器管理引擎即可。

6、赛题在“用户期望”中提到需实现“边缘网关应用程序框架”及“数据发布、接收、持久化接口”，但在“任务清单”中只提到了容器管理引擎的设计，请问贵公司希望项目主要集中在“基于容器化的边缘节点应用框架设计”上呢？还是“边缘节点容器管理引擎”上呢？如果是后者的话，请问贵公司认为项目需要在已有的开源项目（如 Docker）上增加何种改进呢？

答：聚焦于容器管理引擎上：

轻量化：可以去除远程仓库功能，优化 go 语言的编译镜像大小

安全性：考虑对镜像数字签名及签名认证（基于数字证书）

其他：优化容器管理性能等等。

7、本题是否要对不同的脏数据给出不同的处理方案，还是只用给出脏数据的大体解决方案就可以了？

答：重点在容器运行那一块，这部分可以不给出脏数据的处理方案。

【A04】工夹具全寿命智能管理系统（TMS）【捷普】

1、一个工夹具就是一套还是几个工夹具就是一套？

答：一个就是一套。

2、视频中分为上线夹具和下线夹具，请问有什么区别？

答：上线是夹具从仓库到生产线上（开始使用），下线是从生产线返回仓库（结束使用）。

3、点检是专门有一个人来负责的吗？是定时的吗？

答：点检由专人定期进行。

4、工夹具的定位是简单的用表展示在某个厂区某个区域某人使用即可？请问位置跟踪的具体要求。

答：只需显示厂区，区域，线别，工站。

5、移动端的最终样式是类似独立的功能 app 吗？有无具体要求。

答：移动端 app 是主体的补充，使用同一个数据库。

6、视频中点检记录只有上下线点检，是在出库之后就进行点检，再上生产线；在下生产线时，先点检再入库？那视频中的定时点检又是如何记录到现有的这张纸质表格中？

答：上下线的点检记录是这样的，夹具上线先要点检然后分配到生产线，夹具下线也要点检然后进入到仓库库位保存。至于定时点检的记录是在另外一份定时点检表格中的记录的。

7、现有的这张表格上下线夹具不是同一套，是否可理解为这一条生产线即将下线的一套夹具和即将上线的一套夹具记录在同一张表格中。

答：是的

8、在视频中有提到“需要临时领出去”，记录的表格是产线领退记本，领出去是领到哪里去？

答：领去的地方有很多，有可能去生产线临时需要用一下，有可能去一个工作区域临时做校验验证等，地点是不固定的，临时领取的夹具数量也是单一

的不是产品成套的。

9、另外夹具出入库,这个库是否就是夹具柜,临时领出去和这个出入库的区别?

这个过程在功能综述和流程概述中并没有十分明确的体现,是否需要实现?

答:库就是夹具柜,临时领出去和出入库其实性质是一样的,都是做出入库记录。不一样的地方是正常出入库一定是从仓库领去产线,产线退回仓库,数量是产品的成套夹具。而临时领出去就如上面讲的出去的地点不确定,数量不确定不是产品的成套夹具。临时出入库其实只要做一个简单的出入库记录功能就可以,便于查询这个夹具的去向。

10、 采购是否会出现一次采购多种夹具,并且夹具数量不为1的情况。采购之后夹具上的编号是何时印刷的(这个采购流程是否与我们系统相关,我们是否需要提供编号),另外可否提供一下采购单的照片。采购审核主要审核哪些方面?

答:会出现一次采购多种夹具的情况,每次采购数量不同肯定不为1。夹具上的编号在夹具制作的时候就已经生成,系统只要负责录入记录即可。系统可以不要采购审核的内容,系统只要录入已经核实好的信息,比如:夹具名称,夹具本身的编号,夹具的接受编号,夹具的接收日期,夹具属于哪个产品,夹具的PM频率,夹具的PM类型,夹具的库位信息,夹具的负责人等。

生产线、具体夹具是否只有一个所属的工作部门,对其他工作部门不可见?

答:可编辑和不可编辑,统一由管理员授权,系统通过公司个人的NTID登录。

11、 Family(所属大类) 是什么含义,也是要完全独立吗?

答:Family属于一个产品的家族大类,它下面有不同的型号,型号下面还有不同的频段号。

12、 前端框架是否可以用vue框架?

答:可以

13、 关于出入库:产线员工需不需要作为系统用户,如果是,那么产线员工与仓管员是否都是作为初级用户?仓管员是否作为一个独立于部门存在的职位?

答:产线员工指的是操作这个系统的人员,他们的主要职责是将夹具从仓库中领出然后发放到生产线上,并且在系统上做登记。还有就是临时领用做

登记，夹具归还做登记，夹具异常做登记，夹具保养做登记。而仓库管理员的权限在操作员上一个等级，他要夹具的基本信息录入到系统，夹具的库位信息的管理变更等。

- 14、系统以 workcell 为单位隔离数据，各个 workcell 之间数据完全独立互不可见，那么请问只有本部门的产线员工能够借所属的工夹具？个人理解是产线员工都是一个生产部门的

答：Workcell 里面由工程部，生产部，质量部，测试部门，而这个夹具管理的职能部门是工程部是辅助工程部对夹具有系统化可视化的管理，所以对这个系统有操作权限的都是工程部，其他部门只是可见基本信息即可。这里说的产线员工指的是工程部的实际操作这个系统的在线操作员工。

- 15、由于提供的数据缺乏真实性，按数据中推得一个工夹具对应一个采购单号，这与实际情况相悖

答：一个夹具有唯一的自身编号和 buyoff 编号，这里说的 buyoff 编号是经过各个部门验证这个夹具符合功能的接收编号，不是采购编号。

- 16、关于申请入库流程：请问贵公司的入库申请是在采购之前，申请之后经过审批再去购买，还是说已经购买回来了，只是需要更新仓库内工夹具信息

答：入库是在 buyoff 之后。

- 17、工夹具的哪些基础信息是固定的，比如说同一类的工夹具的共性。

答：固定的有夹具自身编号，夹具 buyoff 编号，夹具的 jbuyoff 日期，ower，使用的产品 family，型号。

- 18、有无用于预警预测的数据？预警模型的构建需要有大量的历史记录以及此工具夹在借用之后是否损坏，若损坏则对应的损坏程度如何。

答：目前尚无这方面的具体数据模型，有能力的团队可以充分发挥创造性思维去设计和开发。

- 19、前端是否不限制框架？如果限制，可用范围是？

答：前端框架原则上不作限制，但希望是 bootstrap/jquery 或 vue/element-ui 等主流框架

- 20、是否会有一个夹具属于两种不同的产品？

答：同一个产品会有多套不同的夹具，同一套夹具会用于不同的产品。

21、 公司有几个 WorkCell,每个 WorkCell 都有生产部、质量部等等? A WorkCell 里面定义的夹具类别 在 B WorkCell 中是否能够看到? 还是说完全屏蔽?

答:每个 workcell 是独立的系统,数据互不可见,用户和权限也是各自一套,互不相干。

22、 1、采购入库针对的采购还是入库? 2、采购入库申请是在购买前进行申请还是在检查无误后入库时进行申请? 3、我们没有在您提供的数据 excel 中找到 buyoff 编号。

答:系统中的采购入库仅指夹具第一次录入数据库,与实际采购行为(定价、合同、付款等)毫无无关系。

23、 请问工作部门是怎么划分的? 你的权限里有初级用户、高级用户、监管员、经理,这都是一个部门的,还是不同部门的? 这些不同的角色需要区分于不同的部门吗?

答:每个 workcell 都有自己独立的一套的用户和权限。

24、 在问题总结中,您指出临时领出去是数量不确定的,不是产品的成套夹具,请问这个非成套夹具指的是半成品吗? 因为您在第一个问题中指出一套夹具就是一个,所以这个非成套夹具指的是什么意思呢?

答:成套的意思是这个产品需要的所有夹具,比如说 A 产品需要 10 个夹具,这 10 个就叫成套。非成套是临时领取 这个产品的 2 个或者 3 个夹具,这是非成套的意思。

25、 请问夹具的使用记录与领用的信息记录有哪些区别? 或者说是不是等同于记录夹具的领用情况? 我们想详细了解一下夹具的使用记录需要记录哪些内容。

答:夹具的使用记录和领用记录,其实都是对夹具的进入库做一个登记,可以视为同种功能。记录的内容就是,时间,班次,生产产品型号,领用人,领出时间点,夹具使用地,备注等。

26、 请问系统的部署环境是怎么样的?

答:部署环境为 windows server 2016 + IIS 9。

27、 能否不止限于局域网,比如增加连接互联网使用微信等方式使用该系统?

答：夹具数据涉及公司各个客户的商业机密，本系统不可能接入外网。

28、 工具夹报修报修何时何地进行，工具夹报修时，应该先执行返库流程再报修，还是直接报修？

答：直接报修。

29、 请问线下工人和仓管员属于权限角色的哪一个，或者还是要单独列权限角色？

答：线下工人只是操作的角色，他的功能是执行夹具出入库登记，夹具损坏报修。仓管员的功能是夹具的信息管理，保养，报修的异常处理，等。

30、 贵公司提供的两张 excel 表格，其中一张名为夹具信息的表格中存在的信息与另一张表中的夹具定义表和夹具实体表中存在的信息均无对应关系（例如夹具信息表中有夹具编号而另外两张表格中并无此类信息），可否自行添加？

答：可以

31、 工夹具入库以后，工夹具的位置就固定在仓库的工作柜中的一个 bin 位，以后的进出库它的位置都不变吗？还是每次都进行改变？

答：bin 位需要可变。

32、 关于领走和退还功能，已知领走是夹具在生产线上被临时领走，那领走的具体操作是怎么样的，是产线上任何员工都可以进行领走产线确定吗？还有退还的时候是直接给该夹具的仓管员还是也是给任意一个产线员工呢？

答：领走不是从生产线领走的，是从仓库领走的。需要记录从哪边领走去到哪里谁领走的做登记记录。进库操作员可以登记，仓管员也可以操作。

33、 夹具在提交维护申请后，谁进行维护呢？这个是否在系统的范畴之类，如果在的话则这个维护是不是属于其他部门来进行维护而不是工程部，这个其他部门处理好后再反馈给该工程部的高级用户进行维护后的信息记录？

答：系统的夹具信息由仓管员维护，当然也可以设立一个管理员的角色来维护整个系统。

34、 一个班次有多少个仓管员，多个仓管员是联合起来一起管 50 个夹具柜不分你我，还是一个仓管员负责一部分的夹具柜？如果是一个人一部分的话需要对仓管进范围的圈限？

答：最多 2 个仓管员，仓管员负责整个 workcell 的所有夹具柜。

- 35、 关于点检人员级别的疑问。①点检人员的级别是 Operator I (初级人员) 还是 Operator II (高级人员)? ②夹具点检表由哪一级别的人员提交?

答：1 初级人员。2 初级人员。

- 36、 夹具的生产线流程是怎么样的? 具体步骤有哪些? 产线员工在去从申请到得到该夹具会是怎样的流程呢, 目前有如下理解 1. 产线工人去找管理员 (指定位置) -> 管理员在其办公处进行相关登记 -> 带他一起去拿夹具 2. 产线工人自己在手机上申请 (连接公司内网 WIFI) /附近公司提供的申请设备 (如连接至该系统的电脑) -> 管理审核完毕后 -> 通知到指定位置领用 还是说有其他方式? 对于扫描员工卡确认身份, 是已经有该保障了吗 (如有, 大致怎样扫描确认), 如果没有, 是希望该功能融入现有系统吗?

- 37、 答: 工夹具上线流程, 产线工 (也就是初级人员) 根据生产计划 (另外的系统) 确定该条生产线今天需要生产的产品型号, 然后工人通过手持电脑在系统中根据产品型号检索到该产品型号的所有夹具信息, 然后工人去对应的库位提取夹具, 然后在系统中登记出库记录夹具的去向及出库时间操作人员等 (每出一个夹具操作一次出库登记), 到这时一个夹具的出库流程完成了。人员的身份登记可以通过刷员工卡可以通过扫身份二维码。身份确认需要融入系统, 就是当这个工人登录系统后他的所有操作都是系统自动记录。

- 38、 初级用户操作进出库: 是指可以在界面选择自己想要借用和归还的工夹具; 还是指初级用户负责操作对借出的工夹具进行记录和归还的工夹具进行归还记录, 这个页面动作是记录信息还是操作借用和归还工夹具?

答: 这个页面既是操作界面也是记录界面。

- 39、 Model 夹具模组 (多个) 与 PartNo 夹具料号 (多个) 有什么作用, 查询的时候需要按照他们查询吗

答: 根据这个信息能快速检索定位到改夹具, 或者该型号的所有夹具。

- 40、 不同批次采购入库的同种工夹具的 code 是否是一样的 (比如说报废了一套, 又要买一套新的一样的夹具回来, code 会变吗)

答: 夹具的 code 在同一时间是唯一的, 但是当夹具报废后, 它所占用的 code 会被系统回收, 并赋予给新采购来的夹具。

41、在（3）流程概述中，进出库流程需要扫描员工卡确认身份，是扫描的什么类型的员工卡？是公司有硬件扫描磁条卡还是手机扫描二维码那种？员工的工号的规则是怎么样的，对于 Admin（系统管理员）的管理用户功能来说，是先知道员工的工号，再录入员工信息，还是先录入员工信息，再由系统自动生成编号？

答：员工卡为公司硬件磁条卡，员工信息需要从公司员工卡管理中心获取。

42、夹具出入库中，入缓冲区时间是什么意思？

答：入缓冲区的意思是当夹具从生产线上撤下来的时候，这些夹具没能及时归还到仓库，这个时候夹具需要暂放在一个地方这个地方就是缓冲区。时间在 12 小时以内这些夹具必须入库。

43、如果点检出现了问题，是希望系统自动总结提交报修信息还是员工手动提交？

答：点检出现问题则点检无法通过，员工手动在系统中提交故障报告，等修复完成后 Supervisor 在系统中标记故障已解决。在此期间系统会阻止此夹具被领用上线。

44、请问产线员工和线下员工算不算在初级用户里面？

答：凡是在本系统中有基本操作权限的员工才算用户，否则只作为访客。访客仅可对部分数据进行查询、察看、导出，关于用户权限划分我们在设计要求中的[6. 项目说明/用户期望/用户权限划分]已经有详细说明。

45、1. 一条生产线是固定只能生产一种产品吗？ 2. 一条生产线能同时生产不同的产品的吗？ 3. 同一条生产线，是不是能根据不同的工夹具生产出不同的产品？

答：一条生产线不是固定只生产一种产品。生产不同产品时需要换线，也就是夹具需要切换

46、请问贵公司的部门下，有多少个岗位？除了线下工人、点检员、仓管员，请问还有别的岗位吗？

答：初级用户、高级用户、监管员、Workcell 经理、系统管理员。请参考[6. 项目说明/用户期望/用户权限划分]

47、操作员、仓管员和初级用户、高级用户的关系是怎样的？赛题手册中说“初级用户”负责进出库操作、提交报修申请，“高级用户”负责入库申请、修改工夹具基本信息、处理报修申请、提交保费申请。但我们看到，上几次答疑中总是出现“操作员”和“仓管员”，意思是操作员就是负责进出库、提交报修申请

的初级用户吗？然后“仓管员”就是负责入库申请等任务的“高级用户”？

答：请参考赛题[6. 项目说明/用户期望/用户权限划分]

48、赛题手册的用户期望中说“通过数据挖掘模型设计鼓掌预警算法”在工夹具有较高概率故障前通过 email 发给相关人员预警，这里的相关人员是指那个层级的用户？

答：发送给监管员，CC 给 Workcell 经理

49、是不是同一个人先点检在领用？还是这两个过程是两个人来完成？

答：点检和领用之间不存在顺序或因果关系，也不一定由同一人进行

50、Family, Model, PartNo 三者的关系是怎样的？

Family, model, partNo 的关系是：Family>model>PartNo.

“创建和修改工夹具类别”是否只是修改工夹具的 Family？

答：不是，是修改夹具的类型，比如说是锁螺丝类夹具，刷胶类夹具，扶持类夹具等。

Family 是否和 Model 相互关联？修改 Family 的时候用不用修改 Model？

答：产品的 Family, Model 都是不变的，无需修改

51、我们原来的理解中，初级员工是对系统的操作人员，因为提供给我们的表格中有记录人和经手人两项，我们把记录人理解为初级员工，经手人理解为产线上的员工。但是好几次的答疑都把产线员工等同初级员工，那么经手人又是谁？不禁产生了产线员工完全可以不登记就拿走了夹具的疑问。

答：记录人是系统用户，经手人为产线员工，在出库操作中经手人即是“领用人”，在还库操作中经手人即是“归还人”，经手人未必是本系统的用户。

52、夹具借出的时间是否有上限，比如我借 A 夹具在生产线上能放多久？还有平均来说夹具借出与归还的时间差是多少？

答：没有时间限制，某些夹具可能会在线上持续使用几个月，但系统应该可以显示每套夹具已经上线使用多久了。

53、不同员工之间有不同权限，那权限等级高的员工，能否做一些权限等级低的事情，比如低级用户是负责“提交报修申请”的，高级用户负责处理申请，那高级用户是否也可以“提交报修申请”？

答：权限向下兼容，高权限用户可以进行低权限用户的操作。

54、Workcell 里面有工程部，生产部，质量部，测试部门”是指，整个公司就以上四个部门，每个部门都是一个“Workcell”，还是说一个“Workcell”由这四个部门组成，而公司有很多个“Workcell”。

答：本系统只对各 workcell 负责，不需要考虑公司的其它部门，对于夹具而言，每个 workcell 都有自己的仓库、质检、测试等职能人员。

55、请问贵公司机床上是否装有可以监控夹具参数的传感器？如果有，监控的是夹具的哪些参数？

答：没有

56、高级用户的“修改工夹具基础信息”功能是指修改“夹具定义”的信息还是“夹具体”的信息？

答：两者都可以，但一般而言，修改夹具信息指修改夹具定义，因为夹具体的绝大部分信息都是根据夹具定义内容自动产生的。

57、.NET Framework 平台是否可以使用微软推出的开源且多平台的.NET Core？

答：捷普的服务器全部使用 windows 系统，请使用.NET Framework 进行开发。

58、原有工夹具的基本信息录入以及新采购的工具夹的基本信息由谁录入？高级用户为什么要对工具夹基础信息进行修改？具体修改在哪一方面？监管员在什么场景下会进行工具夹类别的创建和修改？

答：原信息我们会导出成 excel 文件，系统的初始化工具需要具备导入数据库的功能。工夹具基础信息的修改并不常用，通常在录入信息有误，或公司产品参数发生变更使才会修改。夹具类别的创建和修改也是系统常用功能的一部分。

59、由于表格中的夹具故障维护信息十分有限，仅有四条数据，要以此建立模型十分困难，最后的偏差可能非常的大，那么这种偏差是不是不影响评审，或是说贵公司能提供更多的维护数据来分析，例如在哪条产线上损坏的夹具较多等等，仅仅四条数据比较有限。

答：数据可以自己模拟。

60、项目详细方案中需要在其中体现出可行性分析吗？比如市场分析，用户分析，市场趋势等，考虑到这是一个内部的局域网系统不需要面向社会的用户，或者说

希望给出能够将这种服务,这种 tms 系统卖给不仅仅是夹具,但是仍在纸质记录的公司的分析吗?还是说仅仅需要针对捷普公司夹具这一块就好呢?

答:每个公司的具体业务需求千差万别,tms 为根据捷普情况出题。

61、请具体解释一下 Family、Model、PartNo、工夹具类别、产品之间的关系。请具体介绍他们能来干什么,通过这些字段我们能得到什么信息。比如生产一个产品所需要用到的哪些夹具,这个信息我们需要怎么太确认,这个信息是否存 在于提供给我们的两个表格中。特别是工具夹类别有什么用,能不能给我们具体描述一下您们的工作流程,或流程中运用到这些数据的情况。

答:Family 是指产品的大类别,比较胶 NGR FU。Model 是产品的型号,它在 Family 下面,比如 NGR FU B1. PartNo 是产品的代码它在 model 之下,比如 KRF3457890/1R2C. 工夹具类别是这个工夹具属于哪一个种类,比如 锁螺丝夹具,辅助支撑类,刷胶类等,是夹具本身的特性。

Family>Model>PartNo,

工夹具的类别的目的是类别不一样 PM 的周期不一样,PM 的 check list 不一样,所有夹具要区分类别。不同类别的夹具使用不同的 PM check list。

62、之前的问答中说到“夹具的 code 在同一时间是唯一的,但是当夹具报废后,它所占用的 code 会被系统回收,并赋予给新采购来的夹具。”,我们想确定的是,Code 是属于夹具定义的,夹具体体才有 SeqID,夹具报废回收的应该是 SeqID,所以问题来了,夹具定义会被删除,并且回收他的 Code?

答:夹具定义中的 code 和夹具体体的 code 是不同的概念,前者用来区别不同的定义,后者区别不同的夹具,但二者在格式上有显而易见的关联性,例如夹具定义 EF008G,其下有 3 个夹具体体,EF008G-1,EF008G-2,EF008G-3,后三者中的 EF008G-1 报废后,EF008G-1 这个代码并不会随之永久作废,而是将被一副新采购会来的夹具继续使用。

63、之前答疑中在点检和领用关系上说法相互矛盾,请明确点检和领用间是否有顺序之间的关系?

答:没有顺序关系,可以独立存在

64、要求实现“对大量历史数据的分析和筛选,对工夹具的故障概率进行较为准确的预测”,但是所提供的历史数据过少,难以有效实现功能和检验效果。故,能否提供更多的历史数据?

答:没有过多数据,自由发挥

65、要求实现定位功能,该功能需要一定的硬件支持(需要获得夹具的位置信息,如 GPS 坐标),那么硬件部分需要我方团队自己实现还是由甲方提供?还是说我方团队只需要模拟(虚构数据),然后开放接口即可?

答;乙方提供可行性方案

66、夹具定义里的责任人是指什么人?是监管员还是 Workcell 经理?

答:夹具责任人是产品工程师

67、在 excel 文档表述中,有夹具代码,如:EF2189,也有夹具编号,如:LM2132-3,夹具编号是否可以理解为夹具代码和夹具序列号的合成?如:LM2132-3 表示夹具代码为 LM2132,序列号为 3,LM2132 唯一确定一个夹具体。

答:举例中的 LM2132 是夹具的代码。-3 是第三个相同的夹具。所有有-的都是有多套相同的夹具

68、手册中提到的流程概述是每个 workcell 都包括吗?简单点就是各个 workcell 的具体工作分工,比如说手册中提到的采购入库流程是只需要生产部进行操作吗?保修流程是只需要测试部门进行监测吗?

69、一个 workcell 是不是独立管理不同的生成线,一套工夹具是不是只能属于一个 workcell,能不能在不同的 workcell 中流通?

答:Workcell 是一个个独立的整体,它包含生产部,工程部,测试部,质量部,每个部门在 Workcell 中运行,一套夹具只属于这个 workcell。其它 workcell 无法查看本 workcell 的数据。每个 workcell 的数据都是独立的。

70、工夹具对应的类型需要在开发过程中预留什么样的格式?工夹具有哪几层类型划分?每个工夹具都有自己独立的编号吗?可以举一个工夹具的编码作为例子说明吗?

答:目前类型可只使用“夹具”一种,这是为了系统今后的扩充性考虑。每个夹具都有独立编号,编码规则由各 Workcell 根据自己的习惯自行定义,开发者不应予以干涉,约定俗成的基本规则(非强制性)为:编码由 A-Z、0-9、中划线/下划线组成,编码以 A-Z 开头。例如 EF008G-1。

71、关于员工磁卡中信息以及员工编号格式问题。问题一:想问一下贵公司的员工的磁卡读取后其中的内容是什么?现在有两种想法:①读取磁卡后是特定的字符串,而这个字符串需要在数据库/服务器中对应着具体信息②读取磁卡后是具体的信息内容,如:姓名,编号,部门等问题二:贵公司对员工编号的格式是什么?是否含有下划线(_)、美元(\$)等特殊符号问题三:贵公司的员工的磁卡读取后会以什么样的文件格式返回?

答:员工磁卡读出信息为 10 位字符串(下称员工卡号),全部由数字组成[0-9],公司有员工信息数据库,员工卡号为数据库主键,数据库中存储有对应员工的具体信息,包括姓名,工号(不同于员工卡号),所属部门,费用中心等等。员工信息公司可以提供 Web API 接口,格式如下:

`http://服务器名/mesvc01/api/employee/query/{员工卡号}`

上述 Web API 返回 json 格式数据:

`<CardNo>0004975569</CardNo>`

`<CostCenter>6310225</CostCenter>`

`<Department>ME</Department>`

`<EmployeeID>W3187</EmployeeID>`

`<Name>Tony Smith</Name>`

72、扫描员工证件这个操作是不是只是用来确认员工信息,员工信息是不是有另外一个接口需要被调用?

答:是否指员工卡扫描?如是,请参考问题 71.

73、请问临时被借走是在仓库中被借走还是线上被临时借走领出去呢?

答：是在仓库中被借走

74、请问贵公司的生产车间是已经布置好了局域网还是需要各团队来分析局域网的组建方案呢？现在车间里每一个员工都已经配备了各自的个人电脑吗？贵公司所主要使用的浏览器是什么？

答：车间已部署局域网，每个车间都有若干台电脑，员工用自己的账号各自登录。主要浏览器为 Google Chrome。

75、请问仓库对夹具存储的原则是什么呢？是对仓库进行分区，每一种夹具一般会固定地存储在那几个库位中吗？比如按类型分区，在类型分区下再按 Code 分区？贵公司是有提前为每种夹具规划多少的存储库位吗？每一个库位中含有的 bin 位的个数一样吗？还是根据不同种类夹具的具体情况划分不同个数的 bin 位？

答：夹具存储按照夹具类型分区放置。

76、夹具报修，初级用户申请，高级用户处理，那是高级用户维修吗，也就是维修员是属于高级用户？维修好之后要不要提交维修结果呢？

答：维修员就是仓管属于高级用户，维修记录需要有记录

77、夹具借出与归还和夹具出入库是否必须经过仓管员？

答：不是必须，初级用户即可在系统内登记记录。

78、不同产线需要的夹具不同，那夹具定义表中 UPL 即该夹具在每条产线上需要配备的数量是怎么来的，不是唯一吧？

答：UPL 是初始维护信息的时候填入的，是可变数据。

79、请问工夹具报废时是以套为单位报废吗，举例来说，一套工夹具下有一个夹具废了，整套夹具即为报废？还是说以夹具体表下的单个夹具为单位报废，这涉及到报废操作中的物品寿命计数，是记录哪个的寿命次数

答：报废是以单个夹具报废。

80、请问各权限的人对于申请的只是单纯的录入信息还是说可以选择同意或者不同意该申请。例如：出入库，报修，报废，采购等申请

答：信息录入，夹具报废，需要申请批准，其它出入库，报修 只要登记即可，有记录可查询追踪。

81、企业提供的出入库操作流程中涉及到身份验证，如果在登录的时候通过技术手段已经完成身份验证以及确认跟踪记录，那么在出入库时还需要进行二次验证

吗？

答：需要。

82、同 Family，同 model，同 PartNo 下的夹具是一模一样的吗？，一个夹具料号对应一个夹具吗，还是对应同一种夹具？

答：一个料号对应一个夹具

83、关于维修期间和维修之后夹具的去路。Q1: 原来的流程中能否保证夹具点检后才能上线？能否保证只有点检之后才能入库？ Q2: 系统的期望是否是点检后才能上线/入库？ 上线后夹具遇到故障直接报修？ 假如是，维修之后夹具是直接入库还是重新上线？ 假如维修之后可以重新上线是否需要上线点检？ 维修期间是否存在别的同型号的夹具出库来顶？ 维修之后是否需要点检才能入库？

答：夹具维修后需要点检确认后即可上线入库。

84、系统设计过程中的一些问题。1. 报修等流程的内容中，物品代码是否是 code, family, 以及序列号这几类代码的统称？

答：物品代码是夹具的料号。 2. 类别代码是对应产品的哪一种代码？是 family 还是其他的？ 答：类别代码是夹具的类型，比如说是“锁螺丝类夹具”“刷胶类夹具”“辅助扶持类夹具” 3. 对于忘记密码的用户，是否可以通过密保手机等方式修改密码，还是为了安全考虑，只能通过管理员进行账户重置？ 答：管理员重置。 4. 如果需要新增产线，family 之类的信息，是由哪一级用户进行添加的？

答：高级用户。 5. 既然高一级用户可以拥有低级用户的权限，那么高级用户是不是可以自己申请自己审核？ 答：可以， 6. 高级用户对于报修申请的处理，是否是在维修完成后才进行的？ 答：是。 7. 采购入库的流程是否为：高级用户确认要采购的夹具并提交申请，申请通过后进行采购，然后再将采购的商品进行入库处理？ 答：无需采购流程，只用入库登记流程。 8. 库位的编码规则是怎样的？ 一个夹具柜可以存放多少夹具？ 答：一个库位放一个夹具，一个夹具柜子有多个库位。

85、之前企业回答说仓管员还要负责报修异常处理，那么具体的报修异常包括哪些内容呢？

答：报修异常由初级用户报修的，报修的内容有不固定，比如“夹具定位螺丝松动”“夹具挡片掉落”“钢网破裂”等，那仓管员看到报修的内容将异常的夹具修

复后重新维护好系统内信息然后夹具入库再次可以使用。

86、监察员的工夹具类别字段会记录在表里面吗？我在夹具定义表，夹具实体表中都没有找到这个字段，所以夹具类别的目的是什么？是为了便于搜索吗？我们要自己新建这个字段进夹具定义表吗？为什么类别是可以修改的？

答：夹具的类别的目的是便于夹具的 PM 管理，不同类别的夹具 PM 的内容不一样，比如“锁螺丝类夹具”“刷胶类夹具”“辅助扶持类夹具”这些类别的夹具 PM 内容不一样。类别修改的是说这个夹具的类别可以修改，比如今天是属于“锁螺丝类夹具”明天可能需要修改为“刷胶类夹具”。

87、数据库里的表的表头是否需要与贵公司给出的 Excel 的表头一致，可不可以新加一列或者将一列参数在数据库拆成两列？

答：可以

88、之前的答疑里有提到一个夹具有唯一的自身编号，那么这个唯一的自身编号指的是什么？

答：EFxxxx

89、怎么确定夹具在生产线下线后在 12 个小时内有没有归还？

答：系统可做时间定时提醒

90、之前疑点答复中说 model 是产品型号，partno 是产品代码，可之前的 Excel 中说 model 是模组，partno 是料号。请解释一下这到底是怎么回事？

答：是一个意思。

【A05】基于人工智能的视觉识别技术【文思海辉】

1、赛题中提到，不能使用商用模块，具体是指什么？

答：只能用 Apache、mit、bsd 授权的开源库。

2、现场测试的图片集是和题目给的数据集类似的吗？

答：是。

3、有数据清洗标准吗？xm 内的 l 字段能否给出详细解释？

答：清洗没有标准，按给出的数据集实现效果就行；xml 是 voc 数据集标注的标准格式，学生团队可自行查询。

4、题目中说要实现实时检测，而且要开发可以部署到上机位的系统，对这个系

统有什么要求呢？

答：没有具体要求 windows linux 都可以

5、最后的模型，赛方会验证吗，如果验证怎样的验证形式呢（视频，图片，还是...）

答：提供模型相关指标和代码 以及演示视频即可

6、最后提交的作品需要做成 APP 吗？还是基于 TensorFlow 之类的深度学习平台，只要能实现检测头盔的目的即可？

答：需要一个简单的演示界面，能够从监控视频中圈出没有合法佩戴安全帽的人员，输入可以是实时的监控视频流或者本地的视频文件。

7、这道题的一个最后有个要求是：要求开发可以部署到上位机的系统。我可以开发出一个单独的 App 吗

答：可以。

8、请问可以使用基于 LGPL 的 QT5 吗？

答：可以。

9、请问上位机需要能够控制摄像头吗？比如利用海康威视 sdk 二次开发，来控制摄像头的左右移动等等。还是说上位机只要获取数据流就可以了呢？

答：不用，读取数据流就行，不用控制摄像头。

10、可以使用 FFmpeg 吗？不使用其中 GPL 部分，不修改其源代码，仅使用

答：不建议使用 GPL 协议的开源代码（不知道这种比赛算不算商业行为），推荐使用 opencv 来处理视频。

11、开发使用的深度学习框架只能是 tensorflow 或 pytorch 吗？可以使用其他框架吗

答：可以，框架不作限制。

12、题目中涉及应用场景的词语有“车间”和“生产环境”，所以应用场景是车间吗？如果应用场景是车间的话，具体是哪种车间呢？有没有什么应用场景的示例？能不能描述一下具体的生产车间的工作环境？比如工作流程、会有哪些类别的人、车间工人从进入车间到离开车间的整个流程包含哪些行为？

答：这里涉及的车间和生产环境指的是必须佩戴安全帽的工作场景，没有特殊限制，比如轧钢车间、建筑工地等，都可视为该系统的应用场景，学生团队自己选择一种生产环境作为应用场景即可，不作限制。

13、赛题中显示说通过摄像头采集并识别职工在生产车间是否佩戴安全帽的图像数据，之前答疑写的是输入可以是实时的监控视频流或者本地的视频文件，这两类处理方式对打分是否有影响？视频中包含的内容是否有要求？可不可以任何场地只要能够检测是否戴安全帽？还是说必要是类似的车间工作场景？

答：同上题，最终演示视频，学生可以模拟或者实际拍摄任一生产环境（建筑工地、轧钢车间、机加工车间等必须佩戴安全帽的生产环境都可）的视频，作为系统输入，这里不做限制，能演示安全帽识别的功能即可。

14、请问最终要提供模型的哪些指标呢？一些和测试集相关的指标（如 mAP）我们该如何给出呢？一些指标的值是和用于测试的数据相关的，数据的不同会影响指标的可比性，这样的指标如何处理呢？

答：给出模型训练的指标作参考即可，不另设测试集，该题当做项目题完成，以实现功能为主。

15、具体需要完成的内容，只需要一个可以部署并能够从监控视频中圈出没有合法佩戴安全帽的人员的系统？虽然在文档要求中赛题方要求了实现目标探测，以及实现一个可部署到上位机的系统。但并未给出这个可部署到上位机的系统需要完成的除了目标探测以外的功能，而且在之前的答疑中赛题方回答说：“需要一个简单的演示界面，能够从监控视频中圈出没有合法佩戴安全帽的人员，输入可以是实时的监控视频流或者本地的视频文件”所以整个赛题需要除此外的自由发挥吗，比如对收集到的数据进行统计显示之类的功能。此外赛题方评判成绩的主要指标是什么？是模型的探测能力还是综合系统功能内容的丰富度。也就是这个赛题是一个“算法题”还是一个“项目题”？

答：为降低难度，当做项目题完成，以实现安全帽检测功能为主，其他可自由发挥，注意安全帽有三种情况（工人佩戴安全帽，工人未佩戴安全帽，安全帽静置置物架上）

17、买来的摄像头带人脸识别 SDK 可以用吗，算商用模块吗？

答：不可以。

18、可以接入免费的 sdk 吗？

答：License 是开源的可以，否则不行

19、题目中说“开发的系统需要能够部署到上位机”，请问是否需要上位机(软件)与远程服务器之间的通信？深度学习的计算和可视化界面能否都部署在同一台电脑上？还是上位机(软件)、下位机(计算)要分离呢？

答：部署在同一台计算机即可。

20、关于考核标准，问题一： 请问具体的考核指标是识别准确度与模型性能，还是除识别功能以外自行自由发挥的其他扩展配套功能？ 问题二： 若考核指标以识别准确度与识别功能的实现为主，那么以此为基础的“自由发挥”扩展功能是否为加分项？ 问题三： 在历史答疑中有提到过“为降低难度，题目当作项目题，以实现检测功能为主”，那么我可否围绕“算法检测精度“进行识别算法的改进与实现，而不侧重于除识别功能外的其他扩展功能？如此答题是否离题？

答：指标指的是模型准确度和性能，“自由发挥”的扩展功能不作加分项，以实现功能为主不离题。

21、可否使用 face++的免费人脸识别 API？ 是否可以使用 dlib 库？ opencv 自带的级联分类器可以用吗？ Facebook 开源的 detection 平台算商用模块吗？

答：这个需要选手自己查询第三方库的 license face++的免费库应该不能用

【A06】交通时空大数据分析挖掘系统【东软】

1、是否可以使用第三方数据集？是否可以使用第三方接口，或者开源工具？

答：开源工具无限制，第三方数据集及接口如果是为了实现模块任务，在现有数据基础上进行丰富也是可以的。

2、在企业提供的数据库外，学生可以自己添加数据吗？

答：数据格式为 csv，查看数据用 Notepad++等工具请同学熟悉赛题以及数据后再作决定。我们企业提供的数据库可以支撑赛题中所涉及的所有模块。

3、(timestamp) 信息记录开始时间为什么比 (timestamp1) 信息记录结束时间要大？

答：两个时间取一个既可。

4、企业提供的数据是真实的还是模拟数据？

答：真实业务场景，真实数据。

5、(timestamp) 信息记录开始时间为什么比 (timestamp1) 信息记录结束时间要大？

答：两个时间取一个既可。

6、赛题中提到的实时热力图，一共只有 10000 多条数据 100 多台机器 如果要实现实时热力图 请问 `实时` 要求的是 小时级别的 还是分钟级别的 还是秒级的？

答：时间的数据格式是精确到秒的 建议最好能达到秒级，分钟也可

7、请问给出的是静态数据，如何展示实时的热力图？

答：给出的数据是可实现模块内容的，这个问题涉及到模块的实现，请学生回去好好思考，不要再咨询如何设计模块等问题

8、出行方式静态数据中，已经明确了出行方式。所谓标签化如何理解？

答：将每个人标签化，定位每个人的出行方式

9、数据说明中提到 imsi 是用户唯一 id，而原始数据中，学生观察到大部分都是同一个用户，也就是说原始数据中都是相同的一两个人的数据，这个是否有问题？

答：请学生自行查找`信令数据`方向的相关文章，这块不难理解，原始数据中是几百人的数据

10、1. 原始数据集中“nid -信息类别 1 ， npid -信息类别 2”，分别代表什么？

答：参考数据说明，其他清洗后的字段不影响各模块设计。

11、出行方式判断结果只能从已给出行方式（公交和地铁）中选择吗？还是说类似于驾车步行等出行方式需要自己进行判断？ 另外，请问基本清洗后数据量已不足万条，是否可以提供更多数据？

答：出行方式可拓展，可自行判断，不能提供更多数据

12、是否需要考虑设计相关 APP（或小程序），为用户推荐出行方式？比如根据“非实时的”原始数据进行分析，下午五点三台子地铁站人口密集，相关 APP 倾向于不推荐用户前往此地铁站乘坐地铁； 又比如根据目前的“实时的”原始数

据分析，莆田路地铁站人数异常突增，是否需要在 APP 上提醒用户或者在要求的前端页面上标注异常？

答：有能力的团队可以考虑更多的合理设计。

13、本系统面向的用户是居民百姓，还是政府或商业机构？

答：系统面向政府。

14、企业只提供了一天的数据，能否给出一个时间段的数据

答：现有数据可支撑赛题要求，企业不提供其他数据。

15、如何才算是人群的驻留，在工作地的停留算驻留吗？

答：分布式环境

16、题目是要求搭建伪分布式环境还是单机环境

答：伪分布式环境

17、想请问数据出行方式静态数据中的经纬度，指的是起点或者终点的经纬度，还是地铁站的经纬度呢？因为我们查找发现，该表经纬度并不是地铁站所在位置的经纬度。您在上次的回答中说数据是真实存在的，这里我们存在疑问。

答：出行方式静态数据中的经纬度指地铁及附近一定区域

18、我们只有一天的数据，异常聚集和高峰时段都是人数量的陡增，这样要怎么辨别呢，想问一下企业怎么定义这两个。赛题中提到的拥挤程度，请问有什么指标吗？比如到什么数的时候，算哪个拥挤程度？

答：拥挤程度指标及其他定义，请同学自行定义

19、企业提供的基站的数据仅包含了 Id 还有经纬度，请问每个基站是三家运营商（电信、联通、移动）共用的还是只属于一个运营商？

答：只属于一个运营商

20、大数据平台搭建的操作系统的选择？Windows 和 Linux 操作系统都可以使用吗？

答：Linux

21、原始数据中的 phone 字段中的手机号码为什么是 12 位？例如，其中一个 phone 数据是：86137666647316，尾数多了一位。项目需要利用手机号来获取归属地信息。

答：手机号为敏感信息，数据为脱敏数据。

22、请问原始数据表中 nid npid 这两个字段是否可以拿来扩展数据，若是可以，请问这两个字段是什么意思呢？

答：可以扩展，可自行定义。

23、基站的经纬度和出行方式静态分析中的经纬度有什么关联呢？

答：举例，出行方式静态数据中的‘地铁经纬度’指该地铁站及其附近一定区域。关联性请自行定义。

24、面向政府的具体哪些部门？是否包含城市管理行政执法局，交通局，公安局等部门？是否是面向多个政府机关的，还是单纯的交通局。

答：可设想面向政府多部门

25、赛题给出的经纬度是哪一种坐标系 WGS84、GCJ02、BD09？

答：WGS-84

26、请问信令数据采集的规律是什么？是固定间隔采集一次数据，还是当用户所在基站变化时采集一次数据，亦或者用户进行某些特定操作时采集一次数据？

答：理解数据很重要，从某人一天的数据中寻找规律。

27、我们设计的任务目的是设计一个显示历史交通数据的程序还是一个可以实时显示当前交通状况的程序？

答：都可以做

28、能否描述一下再实际应用场景，需要处理的数据量大约为多少？

答：TB 级别

29、赛题上要求自行搭建伪分布式（单机）环境，请问完全分布式环境是否可以？

是否要求所有框架都要伪分布式环境？kafka 使用单机环境是否可行？

答：有条件的话，搭建分布式环境

不是所有框架都要伪分布式环境

kafka 可使用单机环境

30、目前给的数据集是固定的，请问赛题要求批处理还是流式处理？流式处理的意思的能接受源源不断的数据，在短时间计算后显示；批处理的意思是只能通过提交数据计算后产生结果，对实时性要求较差。

答：数据可做批处理和流式处理，多思考。

31、同一个经纬度可以对应多个基站经纬度信息吗？我发现里面的同一经纬度有多个基站信息。 123.3417969 41.66085815 16386-67529483 123.3417969 41.66085815 16386-67529484 123.3417969 41.66085815 16386-67529523 123.3417969 41.66085815 16386-67529533 123.3417969 41.66085815 16386-67529543

答：同一个经纬度可以对应多个基站经纬度信息

32、测试数据中给出的手机号码是脱敏的，那生产环境中手机号码数据是否也是脱敏的？

答：应用的业务场景不同，数据不同

33、请问，原始数据集是经过处理的日志信息还是原始的日志信息？因为要实现实时效果，必然是需要一个实时的数据流。可否为原有数据进行处理，模拟出实时数据流的环境？

答：可理解为原始的日志信息；可以

34、企业所给的数据是2018年10月3号，也就是十一黄金周时期。给这样特殊时间段的数据就可能会影响产品模块的设计，比如早高峰、晚高峰等特征就不明显。所以我们想问，企业给出这样的数据是不是有意为之（即分析黄金周时期的交通）？如果不是，是否可以考虑更换数据？

答：同学自行理解，多思考；数据不能更换。

【A07】酒店视觉 AI 解决方案【虹软】

1、是否只能使用虹软提供的人脸识别接口

答：推荐使用虹软的人脸识别

2、是否可以引入外部的一些开源项目，比如 OpenCV 等，对项目的部分功能起到辅助作用，但人脸识别、活体检验等部分仍由虹软官方的 SDK 进行识别。

答：对开源项目的引用无限制

3、酒店类型众多，主要需要考虑哪种类型的酒店？

答：酒店类型没有限制，关键是给酒店带来价值，给客人带来便利

4、请问关于赛题设计范围，设计是要贯穿酒店多个环节还是可以只针对某个或者几个特定场景来深度挖掘？

答：不限制场景，以真正能够给酒店客人提供便利和给酒店带来价值为核心，当然如果能整套的解决方案更佳

5、只做网页还是要做网页，app，小程序都要做

答：只做网页可能无法满足需求，建议加上 APP 等

6、硬件方面是要做实际的小模型出来吗还是应用在实物上？

答：做模型或实物都可以，可以根据团队自身情况决定

7、有对微信小程序的 sdk 吗

答：没有专门针对小程序的 SDK，目前的 SDK 并不限制应用到小程序中，可以以服务或则其他的形式应用

8、人证核验是否一定需要二代身份证读卡器？

答：并非一定，人证核验实际应用时用读卡器较多

9、关于人脸识别涉及到个人隐私问题，除客房以外的场所如果经常捕捉酒店顾客的人脸信息来完成一些服务，可能会影响酒店顾客的个人隐私，请问我们能使用吗？

答：隐私的关键是是否会保存或泄漏相关图片，SDK 本身并不会泄漏隐私，但设计该系统时请酌情考虑。

10、酒店视觉 AI 解决方案中提到的人证验证不仅需要摄像头，还需要用到二代身份证读卡器，那么是不是我们也要自己去购买一整套的硬件设备，实现相应的功能，还是说可以本地模拟一个 json 数据。或者说可以不需要解决这个问题呢？如果这样的话虹软提供的 SDK 中的人证校验的就没用到了。

答：隐私的关键是是否会保存或泄漏相关图片，SDK 本身并不会泄漏隐私，但设计该系统时请酌情考虑。

11、公司能否提供硬件 demo 搭建的示例，以及在开发者平台上提供新版 SDK 的示例 demo？

答：虹软官网有一些开发套件、应用套件可以下载，也有一些应用方案可以参考。同时虹软的开放社区也有一些网友上传的方案，可以参考。

<https://ai.arcsoft.com.cn/bbs/portal.php>。

12、这一版的 SDK 能否同时支持 Windows 桌面应用程序以及 web 端的调用

答:SDK 提供 Android/IOS/Windows/Linux4 个平台版本,可以根据需要选用。

13、当解决方案关系到酒店内灯光/门/家电等多物件物联控制,需要连同物联硬件一并制作吗?

答:如涉及硬件控制,请根据自己的产品设计方向和时间酌情考虑。

14、人脸识别(ArcFace)Windows(x64)版本 demo 工程运行问题。版本号:

V3.0,语言 C++ 在 VS 工程中无法在 Debug 模式下运行,会报 0x000000007b 的错误,只能在 Release 模式下运行。运行后无法注册人脸,一直显示注册人脸失败。

答:可以先检查下用到的库版本是否正确。

15、人脸数据库可以自制吗?

答:根据自己团队设计的产品出发考虑。

16、如果想实现车牌识别功能只能基于 opecv,yolo3 此类开源项目为基础实现,还是可以使用第三方公开的 API 实现?

答:没有限制。

17、朋友来访这个场景是让我们更加智能化的为朋友开门,还是如果认证了此人为朋友,那他就能自由进入房间?

答:这个场景建议从安全性和便捷性两方面考虑。

18、题面是酒店服务,需要不要实现酒店人员信息管理?

答:没有特殊限制,但如果是场景实现需要,建议考虑。

19、是否要开发面向酒店,以及面向顾客的两者 app,还有就是不同酒店的特色服务是否要在 app 上不同表现?

答:可以根据设计的功能需要开展,从便捷上来讲,适当的 APP 会有帮助。

20、人脸识别返回的参数具体是什么,文档里描写的并不详细。

答:不同的语言版本会有些不同,比如 C 的版本中开发文档和头文件中都会有。

【A08】智·会学习——今目标“智慧校园”场景挖掘与实践应用【今目标】

1、是否允许自己开发应用？

答：关于赛题的延伸，我们可以让有能力的团队在完成基础赛题后进行基于“教师管理场景”的深入思考，开发一套今目标外的小程序用于教师工资的自动核算与发布。

2、我们需要对题目中所列举的几个场景都找到管理难点给出解决方案，还是通过自己调研所总结的难点给出解决方案？

答：关于场景选择，可以根据团队的时间和精力，选择多个场景，可以是题目中罗列的，也可以是其他的场景。题目中的场景只是抛砖引玉，同学们可以先通过调研和走访找到几个具体痛点，然后再通过今目标，搭建解决方案。

3、可以给今目标移动端设计 UI 界面么？

答：设计 UI 界面跟本次赛题关联性不强，不作为评分参考项。

4、该怎样展现我们的成品，因为业务搭建是在 app 中进行，那我们该怎么说明我们是如何搭建的，是需要将每一步文字化然后配合截图放入在提交的详细方案里吗？详细方案里具体我们需要写些什么？演示视频是指整个模拟场景的展示还是指操作今目标的整个流程？

答：1、是的，在 app 中搭建后，可以通过截图+文字或者示意图的形式放到提交的方案中。2、详细方案中可以包括：痛点描述、痛点分析、解决思路、具体解决方案、实践效果等。3、演示视频可以是两者的结合，一方面是场景的模拟，同时加入操作界面的演示，另外可以配合制作人的介绍、目标客群的采访（原有工作模式与使用系统后的对比）等，内容题材不做硬性要求。

5、目前已加入今目标所给的企业号相关组织，但因为没管理权限我们很多操作无法开展，所以想问作品提交的时候是用此企业号搭建相关业务，还是说此企业号只是为了让我们更快熟悉和方便交流，而提交的作品是需在另创的一个新的模拟组织中来完成业务搭建？

答：请参赛团队及时注册企业账号（邀请企业号：86110797，注册链接 <https://web.jingoal.com/user/account.html#/join>，邀请码定期将在官方群里更新），组委会将以邮件的形式收集团队的模拟组织的企业号、组员人数，交于企业进行账号升级。请各团队队长关注组委会邮件（fwwbds@niso.edu.cn）。

10、利用信息化技术，帮助学校食堂一方智能管理食堂。这个应用算不算在校园场景内，可否参加这个赛题？

答：可以。

11、我们企业里只有我们团队的成员，有些功能需要普通的员工（僵尸号）来实现，能不能创建僵尸号？怎么创建僵尸号？

答：进入今目标网页端，顶部菜单栏“系统管理”，组织结构中可以新增员工。

12、关于购买专业版，在实际情况中，是企业有多少人就要购买几人份的专业版吗，还是只要几个管理人员购买即可，比如说一个组织，有人事部，技术部，宣传部等部门，每个部门五个人以上，是每个人都要算在购买专业版的人数里面吗？或者说，购买专业版中的人是作为一家企业中的什么角色，人数的意义是？不是专业版的用户，能不能加入我们企业？

答：目前今目标官方定义一个组织在 20 人以下可以使用今目标免费版，免费版会有些功能限制；20 人以上需要购买专业版账号，一个人一个账号，560 元/年/人。一个组织只会存在专业版和免费版两种状态，不会出现一个组织里既有专业版成员账号也有免费版成员账号。

13、企业补充资料：

<http://mg.jingaoal.com/mgw/collegemgw/index/jin.html?from=singlemessage&isappinstalled=0>，“今课堂”这个链接里面有平台内各个应用的趣味教学视频，可以让有需要的同学们观看，更容易上手操作。

【A09】僵尸企业画像及分类【科创信息】

1、最后放出的用于评估模型的测试集的特征属性？

答：与已发布数据特征一致，只是不同的数据集。可能在建模的时候不同的竞赛单位有特征组合，但是最后的评判还是输入原来的这些特征。

2、题目中判断僵尸企业的条件是.csv 文件中的 flag 列吗？若是 flag，题目叙述的 2000 家僵尸企业为什么在测试集中会有近 9000 个僵尸企业判定，若不是，如何通过训练判断其是僵尸企业？

答：训练集和测试集是分开的，僵尸企业数量不一样没什么问题，主要看学生怎么去用数据。另，每个企业对应多条数据。

3、项目运行中，数据模型提前载入内存会极大提升运行速度，但相对成本较高，如果每次执行任务的时候载入则成本相对较低，但运行速度相对较慢，想问问企业更偏向于成本控制还是运行速度？

答：可以考虑运行速度为主。

4、测试集和验证集必须按赛方给定的比例使用吗？请问在训练模型时必须按照赛方给定的测试集/验证集比例来训练模型吗？是否允许选手自行调整测试集验证集比例，如：将一部分赛方给定的验证集数据用于模型训练。

答：可以。

5、为什么训练集中会有一万多 flag 为 NA 的数据，是给错了吗？

答：提供给参赛队伍的数据集包括：训练数据集（5000 例僵尸企业，10000 例标签不确定的企业，包含僵尸企业数据和非僵尸企业数据）；验证数据集（30000 例企业数据，僵尸企业数据和非僵尸企业数据比例 3:7）。出题方保留的标准测试数据集：10000 例有标签的企业数据。

6、僵尸画像的用户是哪个政府部门？

答：市场监督管理局。

7、题中给出数据集只有六个省份，企业保留数据集亦是如此？

答：是。

8、视频说训练集共 15000 个样本，我下载数据发现有 15050 个样本？验证集说是 30000 多出一部分 884 个没有标签样本，是否可以将其补充到训练集或直接删除？

答：数目稍有差异，不会影响后续研究

9、可视化 web 端开发的开发语言有限制或推荐么

答：没有

10、这个系统在具体实施的时候面向的用户群体有多大？需要考虑很高的并发量吗？

答：暂时不用考虑目标群体大小，开放式的命题主要是考察分类规则的可解释性以及结果的准确率。

11、baseline 在验证集上结果已经达到 99%以上的准确率，这样的话这道赛题的数据集会不会太简单了（光是使用线性回归模型就能在验证集上达到 91%的准确率），还是说除了模型的精确性和高效性有什么别的需求是这道题的重点？

答：简单与否不是参赛方应该考虑的问题，作为一个目前没有确定标准的问题，可以作为经济领域一种现象的有益的探索。

12、赛题给的训练集中没有控制人 ID 这一属性，而验证集存在，请问再最后结果的测试集上是否会有这一特征？

答：测试时不会加入这个特征做评判。

13、关于净利润问题，本人非金融专业，但也觉得这净利润计算有问题，比如 id28 号数据，其利润总额为 7216，纳税总额为 0，那么根据公式，净利润则为利润总额 7216-纳税总额 0=7216，而数据给出的结果却是-7216

答：由于工商登记的时候没有认真审核，也可能有一些未知的原因导致误差，因此，利润总额这个特征的处理可以更加灵活一些。这也是目前大数据的一种困境，数据收集难，收集到的数据不一定都有价值。

14、请问 Web 端的用户输入，是要求用户输入所有的维度特征，而后进行实时的模型验证，还是只需要用户输入企业的 ID 编号，在预先分类完成的数据中进行查询？

答：验证的时候需要输入维度特征。

15、完成这一算法需要要用到什么软件？

答：web 客户端不做要求，模型也仅仅建议 Python 和 R 语言实现。

16、按照公式：净利润=利润总额-纳税总额，但是给出的数据完全不满足这一规律，这里牵扯到净利润、利润总额、纳税总额这三个特征，为何出题方只说“利润总额”可以灵活处理？

答：所有特征包括净利润、利润总额，纳税总额均可灵活处理。

17、本项目是否需要关注用户权限和数据安全？

答：作为竞赛题目，不关注用户权限和数据安全

18、数据接口文件 year_report_train_sum.scv 里面的净利润都是营业总额的-0.1 倍，这是巧合吗？

答：参赛方应该将目标放在提升模型的预测准确性上。

19、1、可视化 web 端开发，支持单个与批量输入查询分类结果，可以运用何种方法去达到这样的效果 2、可视化 web 端开发是否就是 web 数据的可视化 3、开发出的 web 端需要区分前端和后端吗 4、完全利用 Python 语言进行开发可行吗 5、可视化 web 端的开发大致有哪些阶段或者过程呢 因为网络上对这些问题的回答不太统一，所以想向老师们询问一下，希望能够尽早得到解答！

答：可视化 web 端开发主要目的是方便出题方测试参赛方得到的模型，如果在结果展示方面方面有亮眼的表现，还是可以加分的。至于代码开发方面，出题方没有限制和特殊的要求。

20、出题方是否注重项目盈利能力的考察？

答：作为一个开放的竞赛题目，暂时不考虑商业方面的因素。

21、请问出题方给定的数据中是否可能出现某些列几乎所有样本都是干扰数据的情况？还是说干扰数据在所有列中都仅占很小的比例。

答：最终的目的是考察模型预测的准确性，是否干扰数据，或者干扰数据的比例由模型自身来决定。

21、关于公司训练集年度财报的税收特征问题：有的企业的税收特征为 0 的话是偷税漏税企业吗？3 年为 0 就是三年没交税？

答：可以这么理解。

22、关于净利润，纳税总额和利润总额不匹配问题：根据净利润 = 利润总额 - 纳税总额公式训练集中挺多企业不符合该公式的是不是数据给错了？？？比如 id 5970001, id5970008, id5970014 等等

答：数据本身存在误差难以避免。

23、我们使用非常简单的模型即可在验证集上达到接近 100%的准确率，在与一些其他团队交流后发现它们也是如此。而贵公司对赛题的主要考察标准是模型的分类准确率，请问当参赛团队普遍达到极高的准确率时，贵公司要如何对团队进行区分评价？还是说除了模型准确率外，贵公司另有考察的侧重点？

答：考察的主要标准为模型分类识别的精确性与高效性，还有方案本身，数据处理流程等相关的内容作为参考。

24、关于赛题定位问题：参赛者是只要根据自行建立的僵尸企业画像对僵尸企业进行多维度的分类和分析就好，还是在后续要对于企业在不同的应用场景而言再继续建立一个的市场信用管理体系和风险评估模型。

答：市场信用管理体系和风险评估模型是企业画像的重点应用方向，目前赛题定位仅限于多维度分类和分析。

25、请问 money report train sum 那个表里面的四种融资方式，为啥每个公司只有一种融资方式？

答：选取了一种主要的融资方式。

26、请问给出数据集里面没有给出的 NA，是什么意思呢？

答：数据值不确定。

27、训练集 label 只有 1 和 na，没有 0，半监督模型貌似不能运行？

答：划分僵尸企业没有确定的标准，目前的划分只是主管方的一种判断。

28、所给数据中有的利润总额为正，但净利润为负，请问所给净利润是如何计算出来的？

答：三个特征之间应该是有关联的，但是数据本身有偏差。

29、要求中有这么一句话：“可视化 web 端开发，支持单个与批量输入查询分类结果”。请问这里的查询输入，仅是已给数据集中的数据吗？存在给入新数据，然后根据算法模型即时判断的情况吗？另外，对输出有什么要求吗？只判断是否即可，还是需根据企业 ID 给出企业的全部数据？希望能有输入输出的范例

答：以训练集同样形式的新数据单个或者批量输入算法模型进行测试。

30、批量查询结果是否可以理解为多个单个企业同时进行查询，结果仍然是一一对一的结果？批量输入是手工输入多个企业的维度特征还是导入.csv 文件？

答：每个企业的数据作为一条记录，批量输入为一次导入多条记录。

31、请问贵公司提供的年报数据中的资产总额指的是 企业所得税年报中的平均值还是工商年报中的期末值？

答：资产总额作为期末值的参考。

32、训练集共约 15000 条数据中，有约 2/3 的数据是不带标签的，并且剩余数据的标签全为 1。在没有 0 标签数据的情况下训练，无法得到良好的结果。请问为什么会有这么多不带标签数据？

答：答案同第 27 题。

33、注意到给出的测试集，flag 存在 0、1、NA 三个值，那么我们判断的结果是只需要输出 0、1，还是说考虑到有些因为公司刚创立不满三年等因素，还存在着 NA 的输出。另外每个连续变量的上下限、离散变量的所有值方便提供一下吗？

答：训练集和测试集可以综合考虑，关于变量的阈值或者变量的分布需要参赛方进一步分析挖掘。

34、在我们对赛题数据进行分析时，发现 money_report_train_sum 表和 money_information_verify 表中的各项融资额度、融资成本等数据大部分为 0(有 90%左右的数据都是 0)。我们团队分析认为这与实际情况下一般公司的真实融资情况存在较大差异，想请问企业是不是赛题所给数据有所缺失，所以都用 0 填充了？还是说数据本身就是这样的？

答：目前主管方的信息采集途径还不完善，统计结果也会有缺失和偏差。

35、出题方是否会下发测试集以供我们来验证结果？

答：不会再下发测试集，企业发布的验证集数据已经足够模型的反复验证。

【A10】金融科技服务平台企业数据的无监督分类系统【浪潮】

1、最终模型建完后，一个企业可以有多个标签还是一个标签？

答：多个标签

2、测试集数据的形式是怎么样的？或者说题目中要求：最终以完整系统的形式接受企业的信息输入。这个企业信息输入的格式是怎么样的？企业的信息输入，如果是 excel 的格式，那么数据是不是和给的训练集一样，分成这么多不同的表格，还是在一个表格里面。会不会有些表格因为没有数据，企业就不输入？

答：最终完整系统的实现接收的企业输入数据只是企业名称或者统一信用代

码或者序号之类的简单信息，和给的训练集是不一样的。

3、这道题是构建无监督系统，既然是无监督打标，那在评价指标里面有准确率、召回率等，那是不是就说明在给的测试集里面的数据是有标签的？如果有标签，那是不是对于无监督分类又是一个导向？还是在“技术要求与指标”的（3）模型评估指标只是一个参考？

答：训练数据是无标签的，模型评估指标只是一个参考。

4、能不能给出几份测试数据样例？具体是以怎么样的一种格式输入系统测试

答：输入企业名称或者企业名称对应的序号（通过训练样例将企业名称和序号对应，输入序号）。

5、技术要求与指标中有模型预测新企业样本所需的时间，是不是意味着我们需要设计输入新的企业样本信息，然后输出新输入的企业信息的分类和标签结果

答：可以选择训练样例中的某一家企业当做新企业进行预测。

6、是不是得把企业给的训练集按一定比例划分成训练集，测试集，样本集？答：是的，通常机器学习训练都要分成训练集、测试集进行交叉验证。

7、给定的训练数据集中，company_baseinfo 表中加密后的 entname 出现了重复。

答：重复的代表同一家公司，数据预处理过程中可以根据情况筛选。

8、最终测试模型所用到的企业名称，所代表的企业及其相关信息，是否在训练集中参与模型训练？如果该企业不在训练集中，那么仅仅凭借一个企业名称应该无法找到对应的企业簇。

答：(1) 所提供样本可以人为划分为训练集和测试集，测试模型所用企业可以选择训练集中数据，也可以选择测试集中数据；(2) 如果选择企业不在训练集中，如若能够搜集到企业相关的特征字段，也可以进行测试，如果只能搜集到企业名称，其他特征字段全为空，预测实现代码中应该考虑这种特殊情况，代码中作相应处理。

9、预测新企业样本数据，是指向模型输入一个企业的标识，然后给出该企业的标签吗？如果是这样，则不太合理。还是说输入大量新的企业样本数据，对这些数据进行聚类，输出结果。

答：(1) 预测新企业样本数据时，可以输入企业名称或企业对应的序号（和企业的

标识不同),然后输出企业所在的聚类簇,输出企业对应的标签。(2) 输入大量新的企业样本数据也可以,预测时则需要针对单个企业进行处理,输出大量企业样本数据的结果,也可以接收输入一个企业,然后输出单一企业的预测结果。

15、要求将企业进行分类是指:比如先进行聚类得到ABC类,聚类后对每类提取特征,总结出A类的特征是成立时间长欠款少,B类成立时间短欠款少,C类成立时间短欠款多,就可以据此对企业们打上相应的标签,把标签视为类别,然后做一个分类系统吗?

答:目标是实现企业数据的无监督分类,且每类抽取有效标签,标签合理可解释,实现可视化、交互,形成完整系统。

16、要求我们对现在所有的数据进行聚类分析,只需要在网站中输入主办方给的数据表格中已经有的企业名称,输出聚类结果给出的标签。(也就是只做聚类),还是要求我们将数据分为训练集和测试集,用训练集训练数据,然后用测试集进行调整参数,使预测更加完美。网站中则输入新企业(企业现在给的表格中并没有这个企业)的名称和相关变量,网站给出新企业的标签。

答:目标是实现企业的有效分类,可以根据聚类效果进行参数的调整,不限制采用的具体方法,可以借助分类的交叉验证去验证聚类效果,只是一种思路。

17、答疑中说输入值为企业名称(编号),而未输入企业其他数据如企业类型等,那么若输入企业属于训练集,web端系统就只是展示模型做好的聚类结果;若企业不属于训练集,那么无其他企业的变量则无法判断企业的类别,所以出现这种情况时(即企业不属于训练集时),可否要求额外输入企业的其他变量来辅助判断类别呢?

答:企业不属于训练集时,可以要求额外输入企业的其他变量,但需要考虑没有任何输入变量的情况的特殊处理,以实现最终企业的标签结果为目标。

18、关于之前答疑的是否要把训练集进行分块。企业回答:“是的,通常机器学习训练都要分成训练集、测试集进行交叉验证。”如果要进行交叉验证,那就需要带有标签的数据。但是赛题不是无监督分类嘛,那这个数据的标签是什么?或者说要验证什么?没有标签怎么判断准确性?

答:为达到无监督分类较好的分类效果,可以借助分类方法去验证聚类效果,只是一种思路,也可以采用其他的方法去调整聚类效果。

19、可不可以理解赛题最核心的需求是对企业进行合理、有效的无监督分类，任何考究算法的合理性和结果的准确性？最后，在借助无监督分类系统，进行交互、展示？

答：是的，赛题核心需求是对企业进行合理、有效的无监督分类，最终实现交互、展示。

【A11】多方数据安全计算平台的设计与实现【浪潮】

1、多方安全计算的政府管理员和企业管理员分别负责什么？图中说的计算节点管理具体是什么？

答：政府管理员负责监管数据的使用情况，企业管理员负责查看计算结果，节点管理是管理个计算节点的运行情况。

2、平台可否使用第三方开源工具，评判标准除了加密解密功能还有哪些是相对重要的？

答：可以用开源工具。没有其他附加的评判标准。

3、题中给出风控模型公式 $\text{Math. max}(\text{Math. min}(\text{Math. max}(A * 0.8, B * 15) * 10, 150000), 50000)$ 没有涉及到加法，是否需要实现全同态加密算法？也就是加法同态和乘法同态都必须实现？

答：加减乘除根据团队能力做

4、想问下贵公司是否对输入风控模型的数据集来源有提供？如果不提供是否有格式要求？

答：不提供；没要求。

5、关于赛题答疑中企业管理员负责查看计算结果，是谁在发起请求和接收查询结果，整个平台的用户角色构成除政府管理员与浪潮管理员之外还有什么角色。

答：角色有企业用户、政府管理员与浪潮管理员。

6、政府源 A，政府源 B 所提供信息是否有重叠，例如：情况 1：A 提供待评估人的公积金月缴额信息，B 提供社保缴费基数 情况 2：公积金和社保缴费信息只在政府 A 或政府 B 存储 题目要求的数据源是哪种情况？

答：不会有重叠。

7、一个计算节点是指整个风控模型公式还是公式中的某一步呢？

答：请参赛团队自行设计。

8、针对计算公式的函数和常数可否使用公钥进行修改

答：有能力的团队可自行设计。

9、对密文结果进行解密，该解密后的明文应该输出到哪里？

答：输出给调用方。

10、节点管理是由哪个管理员负责，属于政府还是企业？节点管理的内容具体是哪些？能否控制节点的启动或停止？

答：企业负责，节点管理具体功能根据自己能力开发功能，没有硬性需求。

11、赛题中由政府端输入的数据，经过模型端计算后的结果存于何处？是由各个模型端自行存储（即是说管理端不需要处理反向流量）、由管理端持久化到中心数据库，还是返回到政府端？

答：存储到数据库供后续查询。

【A12】基于算法的配送路线优化系统【华云+中诺思】

1、是否应该考虑配送人员的费用问题：题目假设中提到了工作人员每日时长是8h,那么是否可以认为一辆货车无法配送二次,但是人员可以多次参与配送。2、是否派送点只能经过一次：每个派送点只能由一辆车服务一次,是否意味着默认派送点只被经过一次,还是要考虑经过多次,但只派送一次的情况。3、要求中满载率的定义是否为 $\text{单辆车实际所载货物重量} / \text{该辆车最大载重量}$:题目要求满载率最大化,是指所有车辆满载率之和最大化,还是指单辆车满载率最大化?

答：1、可以不用考虑；2、派送点只经过一次,经过多次这个属于设计问题,我们只考察一个点由一辆车派送一次,不重复派送；3、单个车辆的满载率。学生团队在能力允许的范围内,可以做更优化的设计。

2、请问最终,我们最终测试的数据是按照发出来 excel 表中的数据还是题目中给出的数据呢?还有就是题目中的连线代表是只能这样走吗?能不能 A 走到 C、D、E 等等这些点呢?

答：题目中的数据中根据线的连接走，没有连接的没有给出两点的距离，不能直接走，但是表格中的数据因为是给出的坐标或距离，任何两点都可以走，可以以题目中的做为最终的，表格中的是验证你题目中的设计是否是最优，可以通过不同的数据来验证。

3、题中所提到限定时间是所有货物运输完的时间，还是每辆车的限定工作时间，例如：A车要在1小时完成货物配送，本来的路径需要1个半小时，加上限定条件后这个环境就不满足，就要重新规划路线？

答：可以先不用考虑时间。

4、题目中提到每个员工工作时间为八小时，请问这个八小时是指类似于上下班的固定时间段的八小时，还是其他情况？

答：暂时不考虑其它情况，这个工作时间也是个不是必要考虑的因素，固定的几个因素配送中心的客户量，每个客户的需求量，每辆车的满载率，车辆一次最大运行的最大距离，这些因素是必须考虑的，学生可以先不考虑工作人员的时间。

5、在内容说明给的样例中，题目让用算法求出车辆数量，而在下面的开发需求中需要用户设置车辆数目，能否给出更加详细的要求？

答：在没有给定车辆数的情况下，根据车辆的承载量计算需要车辆，还有一种情况就是给出了车辆数和承载，怎么去规划。

6、是只需要考虑车辆的满载率还是需要考虑车辆的耗油量的服务成本？请问关于车辆的服务成本有没有一些参考数据？

答：只考虑车辆的承载量。学生给出的成果中只要考虑了客户数，客户与配送中心的距离，客户需求量，车辆配载这些就可以，工作人员时间和车辆时间做为非必选项。

7、题目的具体数据以 excel 表格的数据集为准还是命题文档中的为准？

答：以命题文档中为准，表格中的可作为测试数据

8、最优路径的判断标准是否只有总路程，在官方数据集中只提到总路程最短？但是题目中需要输出配送完成时间 配送完成时间是否指的是配送路线中的需要完成时间最多的那一辆车的完成时间(所有配送车在同一时间从配送点出发) 那么就有可能出现配送路线中某一条路线的总路程小于另一条。而配送完成时间却大于另一条 是否需要我们确定一个权重。

答：需要路程最优，不一定是最短，还有车辆的使用情况需要考虑进去，不是车辆越多越好，完成时间只作为参考指标，根据配送量和车辆的满载率去考虑路径最优

9、在官方数据集中，要求我们需要随机生成各种数据，只明确了配送量和位置信息需要随机生成，请问其他数据是需要提供数据接口去设置吗？例如，车辆有哪几种类型，载重量等等。

答：车辆的载重量，数量需要能够配置

10、对于车辆速度，员工的工作时间等等，它是一个定量吗 能规范出哪些是定量吗

答：对题目的考核暂时可以不用考虑时间问题，这个后续可以作为拓展内容学生去研究。如果学生能在这方面有所创新，想法当然更好

11、系统部署指什么？是要封装出可供调用的 API？还是要求做出一个网站一样的东西吗？有必要根据作品，结合生活场景，如快递公司做出一个应用系统吗？如果是对界面有要求吗？

答：开发作品需要能够安装到计算机上，通过简单的界面，设置一些参数后可以看到结果

12、把算法应用到实际场景中，可以用其他公司如百度,阿里的 API 吗？

答：这个没要求

13、题目中求解是按照给定的工作人员工作时间，在工作时间之内将选定的个个位置的派送路线计算出来,并且耗时最短,花费最少。是按照这样的要求求解吗？

答：耗时，花费是整体的，总共那么多客户的量，完成所有客户的配送整体的路线，耗时。

14、时间复杂度会是算法的一个评判标准吗？

答：时间仅作为一个参考的评判。

15、员工的 8 小时工作时间是否包含送完最后一个配送点后的回程时间？

答：整个试题是从配送中心出发，配送完成后再回到配送中心，正常的 8 小时包含最后回到配送点的时间。

16、问题描述中的成本包含哪些内容？如：工资（员工工作时长）、邮费（行驶路程）等？能否给出一个定量的描述公式？

答：成本中目前可以只考虑车辆的选择，比如车辆装载量不同，对应的成本会不同。

17、题目描述中说要订单最大化，但同时又说明每个配送点需要的配送量，按后面这句话的意思是要满足每个配送点的需求，那么总订单应该是给定的，又如何实现订单最大化？

答：订单最大化一是指相同时间内配送订单的量最多，成本小。另一方面也指需求相同的情况下，用时最少，线路最优。

18、三个目标分别为：成本最小化，配送订单最大化，满载率最大化。此三个最优化目标是否具有同等重要程度？

答：我们重点要考察的是线路最优，车辆配置合理。成本，订单，满载率是过程中需要考虑的，成本和满载率涉及到车辆的配置，订单派送的不合理会影响线路的选择。

19、配送点数量有多少？能否给出一个模糊的范围，涉及到算法的选取，问题规模比较重要，还请告知。

答：每个配送中心需要配送的客户数量不等，在客户数量多的时候，可以根据配送量的多少配置多的车辆。我们考察的是同样的配送中心，同样的客户，给出同样的车辆，最后得出这个配送中心的最优方案。不同的配送中心不具有可比性。

20、A、每辆车只能服务一条路线是什么意思？一条路线如何定义？

B、一辆车可以先运完其中几个点，回到P拿取货物后，能再服务另几个点吗？

C、一辆车走过一些点后，但不满足其中订货点，其他车辆可以经过这一路线吗？

答：A、一条路线是指比如客户1，客户2归A车配送，客户3，客户4归B车进行配送。那么，客户1，2就属于一条线路，那么A就不能再给客户3，4进行配送。

B、这个可以，这个考察到车辆的满载率。

C、可以经过，但是实际上这样走是达不到最优路线的结果。

21、A、第一个样本算例中，配送车辆的基本大配送距离就是最大配送距离吗？如果是最大配送距离的话，这个是指配送车最远能到达的地方吗？但是下表的数据中有4个配送节点已经离0点超过了10KM了，这是怎么回事呢？

答：配送车辆的基本最大配送距离是指从一个客户到另一个客户之间的距离。节点与节点中距离有超过10KM的，说明不适合这两点放在紧挨的两个派送点。

B、第二个样本算例中，物流中心的坐标为（1415KM,1310KM）但是客户点的坐标都是十几或几的数量级，最大行驶距离才50KM，能到吗？是否是数据出错了？最后配送距离=行驶距离吗？

答：数据错误，数据中配送总里程指的就是车辆走过的距离。

22、由题意小车大车的行驶速度都是10公里每小时，且卸货时间相同，也未规定数量比得知两种车除载货量不同以外并无其他优劣性。所以我认为大车对于小车有绝对的优势，如果在算法中只考虑优先使用大车会影响评比得分吗？

答：不会影响，比如大车满载率达不到的情况下，可以考虑使用小车进行配送。

23、网点的含义是什么？只考虑单个配送中心还是需要考虑配送中心的增加和减少？每辆派送车的开工成本和每单位里程的运输成本是否需要考虑？Excel表中的第二个算例中物流中心的坐标为什么和客户点的坐标相差这么大？

答：网点的含义举个例子，顺丰在深圳有个中转场，所有从全国各地到达深圳的快件先到达中转场，再由中转场送至各个区管理中心，由各个区再发送至各个私人经营点，再由经营点送至客户手中，网点只是说法不同，各个区管理中心可以作为大的网点，私人经营点可以作为小的网点，题目及测试中我们只考虑题中经出的，也就是要送到的客户作为网点。

目前此次可以只考虑单个配送中心，减少复杂性，运输成本暂时可以不用考虑。

注：第二个案例数据有问题，已做修改，见群文件数据更新。

24、配送时长由什么因素决定？

答：此题的主要考察点在路线优化上，配送时长一般根据配送路线长短，客户数量决定，因为题目中的车辆速度，所有客户卸货时间都一样，所以平均都是一样的。

25、装货时间算不算在派送人员的工作时间里面？

答：不考虑

26、配送中心是只有一个还是需要多个，是不是可以设置1个或多个？

答：配送中心设置为1个，网点为多个

27、我看了给的测试数据，那我们在设计输入的时候，我们应该按照客户的坐标点来输入呢，还是直接输入邻接矩阵的形式呢？

答：给出的坐标是为了让学生计算出距离，最终是以距离计算的，

28、请问作品是只用解决数据集里的两个问题（无工作时长限制）还是以此为基础进一步实现命题赛题手册案例（有工作时长限制）呢？

答：只需解决无工作时长限制，只考察路线最优。

29、答疑中说是二种情况，那到底是按照车辆数量作为输入参数，还是车辆数量作为输出参数呢，请出题方给出明确答案。

答：输入参数比如为10辆，但实际实用只有5辆，输入和输出是有区别，题目中给出的车辆数不一定全部都投入配送使用。

30、对于配送点的数量，能否给出大体范围或者是其数量级？

这个无法进行定量，如果配送点多了，配送车辆的装载量是一定的，配送的车辆就会相对较多，

31、我们需要完成的是双车型，还是多车型的问题呢？

车辆只考虑数量和装载量的问题

32、题目中说限制车辆一次运行路线距离不超过35公里，这个35公里是指车辆从配送中心出发配送完成后再回到配送中心的总距离还是指两个配送点之间的距离 还是35公里只包含配送货物的总路程不包括返回配送中心的总路程。

答：一次运输的距离，从配送点出发再回到配送点的距离

33、题目中的客户点指的是送货上门客户的家庭住址，还是指的是类似与自提点中转站一类的概念，这关系到我们项目文档中市场需求分析方面的书写。

答：这个概念是一样的，题目可以是一个配送点给客户配送，也可以是一个中转

站给各下下属配送网点的配送

34、最后提交的成品作品可以是手机 APP 吗？题目中要求的是可以在电脑上安装的软件，但智能手机也属于计算机，并且 APP 也可以在电脑上运行，请问手机 APP 符合要求吗？

答：电脑运行即可，APP 不做要求

35、开发要求中提及考虑网点和各配送点之间的距离，我们原本理解网点就是客户，配送点也是客户，我们是否理解错误？

答：比如顺丰公司在将深圳的快递统一送到深圳中转站，深圳的中转站再给各个区分拨，这是一种。还有一种比如深圳南山的顺丰总点再给各个街道顺丰点的配送。也可以理解为街道的配送点送到最终客户的手里，道理都是一样的

36、车辆能否有剩余货物，如果有剩余货物，满载率计算可以完全让每辆车的载货量最大化？

答：如果有剩余货物的话，那就会造成运输的重复性，原则上应该是给一条线路进行配送，货物是可以送完的，可以有空，有剩余的话不能算做满载率。

37、请问第一个案例中的基本大配送距只存在于第一组测试值，还是说该约束针对全部数据恒久存在。因为最初发布的 A12 原题的约束条件中并未提到有车辆基本大配送距，该约束只在案例一中出现，案例二中亦无，是否可以理解为该约束（最大配送距）只针对于在样例一，而非作用于全部数据；还是说该约束（最大配送距）位同于原题中提及约束，要求针对全部数值，且约束数值可修改。

答：这个约束只为参考。

38、需求分析是对项目市场需求做分析还是对设计的系统做功能需求分析？

答：对系统功能的分析，类似于一个产品的功能要求。简单即可

39、请问数据输入具体是以怎样的方式，是.xls 或者.csv 这类直接输入还是挨个数据手动输入？

答：最好能导入数据

40、对于满载分别为 2 吨、5 吨的车辆，运输成本该如何计算？

答：用 5 吨的车运送 2 吨的货物成本肯定比 2 吨的车运 2 吨的货物。

41、请问配送中心是单一的还是可能有多个？

答：这道题只考虑单一的配送中心

42、请问之后还会给算例嘛？算例的规模大概为多少？（因为算例的规模涉及到算法的选择问题）

答：如果需要可以提供

43、如果拿总路径最小化和满载率最大化进行衡量，那么他们的权重是几比几？

答：我们出题的本意主要是考虑线路优化的，满载率是其中一个考虑的因素，权重不太好说是多少

44、作品能否采用 bs 架构？请问提交的作品是否有应用类型限制？可否是 web 应用？程序开发的形式可以以微信小程序的形式呈现么？

答：不限形式

【A13】运用文本相似度实现主观题自动评阅【华云+中诺思】

1、赛题中所说的测试样例在哪获取？

答：物流专业的简答题基本都是名词解释和概念释义类的题目，语句结构都比较简单，网络上有很多物流相关专业的简答题文档，学生团队可以案例题的样式自己准备数据集。

2、赛题要求将标准答案与学生答案进行对比分析，请问我们所获得的学生答案是否已经经过了 OCR 文本识别？

答：无须经过 OCR 文本识别，目前的场景是在线答题，而非扫描的纸质答卷。

3、题目没有告知评价模型的标准，那我们做完之后，该如何判断自己的模型好不好？

答：题目里给出的标准是对比人工评分，意即模型评分与人工评分的结果一致度越高，模型越好；如学生答案人工评分是 7 分，系统评分为 3 分，就偏差比较大。这里的人工评分标准在于语义是否正确，以及是否正确命中得分点验证的话，可以在网络上找几道物流仓储作业相关的简答题，模拟不同的答题情景让系统评分，然后自己对比参考答案给分，来判断模型是否基本满足需求。基本上物流相关的简答题都是名词解释，释义类的，答案都是比较好理解的。

4、是否需要按照类似人工的给分方式，将答案拆分为各小点分点给分？（例：有ABC三个点，每点3分）还是只需要最后模型给出的总分和人工阅卷的给分一致，不需要按点给分，这样就可以了？

答：如果可以按点给分也能实现总分与人工评分一致，也是可以的

5、企业后续是否还会提供数据？

答：企业将于1月10日前提供新的数据集，请关注大赛官网和官方QQ群文件更新。

6、测试集是会给出评分标准吗？就是回答怎么样给多少分，还是说只给标准答案、关键词和关键词的描述，通过关键词和关键词的描述是否匹配以及匹配的个数给分数。

答：测试集会给出标准答案、关键词等，但不限定输入参数，最终测试集我们也会依据学生提交的算法进行微调，保证测试结果准确性，算法合理性也会对比人工对测试集的评分。

7、主观题的类型有明确的指向吗？题目是只需限制在物流专业方向的主观题还是不限专业方向？就比如是围绕在物流专业方向的题库进行还是任意专业的主观题都能进行批改？在赛题提供的样例中，大多是询问考生“是什么”的题型，而且涉及的大都是物流管理专业，对于其他询问“为什么”、“怎么做？”的题型该考虑吗？

答：我们需要至少兼容物流专业方向题目，至少兼容名词解释类、概念释义类的简答题，但若同学们的解决方案能有更多更好的应用场景，我们也会酌情加分。

8、能否给出答案的关键词语料库，分好关键词的答案。关于数据集没有对应的题目，后续的数据集可以给出一开始的3个例子的数据集一样的类型吗？

答：后续会提供语料库，具体时间等待组委会通知。

9、请问用不用写出网页，学生在线答题的界面？

答：可以写网页，也可以不写(语言酌情考量)

10、测试集的格式是什么样的？测试集中题目、标准答案、学生答案、关键字等数据是以什么形式、什么格式给出的？对于每道题给出分数的反馈形式有一定要求吗？

答：题目 标准答案 学生答案 关键字 评分 五个维度进行展开

题目 a b c d 5 类似这种，由 a, b, c, d 得到这个测试得到 5 分。我们的评分会有一个标准的范围的。

11、人工评分是由出题方提供还是我们个人定义？

答：人工评分是有我们这边提供的，我们会根据我们这边的测试集给学生进行评分。

12、咱们这道题需要设计算法给主观题自动打分，看打出的分与人工打分最接近的是最优的算法。那么请问人工打分的标准是什么？

答：人工打分的标准就是答题的几个得分点是否匹配上，目前的简答题都是“客观”简答题，可以理解为人的作用在于识别，没有带有主观情绪的打分，不是语文大作文这种评分；分值可以简化为按得分点平均。

13、对用户期望中的第四点有疑问——能够针对常见的解释性语句做词语及其解释之间的匹配。能够针对常见的解释性语句做词语及其解释之间的匹配，这里匹配要求是什么？是输入一个词语，然后输出其解释吗？如：输入——地球，输出——一个天体；还是说在解析文本的过程中，同步自动找出解释性语句，并输出词语及其解释？

答：第二种理解，这里的解释性语句指的是解释型的句式，如地球是一个天体这句话，需要程序能解析出一个天体是地球的解释，如果答案是地球不是一个天体，那即便识别出地球和一个天体，这题也不能算对。

14、关于计算分数模型的输入，是只输入标准答案和学生答案用于计算，还是输入标准答案、学生答案和关键字用于计算。

答：只需输入标准答案和学生答案即可。至于关键字，模型算法里面会作处理。

【A14】金融领域公司实体消歧系统构建【恒生电子】

1、可以实时提交测试集的实体消歧结果，还是只能自己利用训练集判断模型好坏呢？假设我的模型可以得出测试集的实体消歧结果，我是否可以提交结果得到实时反馈，以判断我的模型在测试集上的效果呢？

答：验证集上可以充分验证结果。

2、待测试集中需要消歧的实体是只在待消歧实体对应表

company_2_code_sub.txt 中出现还是也会在金融领域公司实体对应表

company_2_code_full.txt 中出现

答：测试数据都包括在 sub 里面

3、能否利用一些额外的数据，比如搜索引擎检索数据利用？

答：可以利用额外数据，但尽量说明利用外部数据的原因。

4、一个简称有可能对应多个公司吗？如一品红广东有这个公司，上海也有这个公司，对于 mention 的结果是多分类还是二分类？

答：测试数据中所有待消歧实体都在 company_2_code_sub.txt 表中，你可以看一下情况。我们真实场景是存在这种可能的。

5、训练集里 text_id 为 1 的上海电力是否标注有误，这个上海电力指的不是公司吗？

答案：标注数据可能存在一定错误，允许一定噪声存在。数据是人工标注的，错误数据极少。具体情况，参赛选手需要理解数据。

6、最后提交的结果中 confidence 是必须要有的吗？

答案：关键字段可以留着，在没有具体数值的情况下，可给默认值。文档中描述过，它可能成为最终方案评价的加分项。

7、若一句话中存在两个金融公司实体，针对该句子进行实体链接时该如何分辨是哪个实体。10834 的此句话："作为炒股软件起家的大智慧的发展模式，可是偏偏又重组失败。有人这样形容大智慧的发展历程：跟东方财富和同花顺相比，大智慧财经信息终端是起了个大早赶了个晚集，张长虹没摆脱股评的路径依赖，加之投巨资的采编业" 这句话中实体识别的 groundtruth 是同花顺，但是经分析此句话的重点应该大智慧，是不是 groundtruth 中出现了标签标错的现象？另外，如果测试集中也有类似的语句，那么该如何对其中不止一个的实体进行取舍？还是都实体识别出来统统链接？

答：groundtruth 不区分文本描述主体，只关注该文本中是否存在待消歧的内容，只要是我们关注的待消歧实体，我们都要对它进行链接。而测试集中待消歧实体在 company_2_code_sub.txt 表中，训练集及验证集的待消歧实体参赛选手可通过标注数据直接统计获得。

8、命题文档中的 F1 是否有问题？P 和 R 的计算公式中，求和号的下标是否应该为“Q 属于 Dataset”（或者把 ME 和 ME' 的下标改为 q）？P 和 R 的计算，是否把 kb_id 预测为 -1 的样本也计入计算，即，文档中的原句“Q 中 N 个实体 mentions、位置及其链接到是知识库的实体 id”中的“实体 id”是否包含 -1？

答：下标有误，希望没有影响理解，改成 q 更合适。提交结果有相关描述，

注：提交的结果需包含该句中所有可能歧义子串的消歧结果（相同子串，位置不同，需分别给出消歧结果）。例如，

数据形式在文档中也有相应描述的，没有正确链接的待消歧实体用 -1 表示。

```
{
  "team_name": "ABC",
  "submit_result": [
    {
      "text_id": 0,
      "text": "太阳能上半年净利润同比增长3%；一块去获批挂牌新三板",
      "mention_result": [
        {
          "mention": "太阳能",
          "offset": 0,
          "kb_id": 123,
          "confidence": 0.80685
        },
        {
          "mention": "一块去",
          "offset": 15,
          "kb_id": 254,
          "confidence": 0.783856
        }
      ]
    },
    {
      "text_id": 1,
      "text": "各领域营销的能人智士一块去探讨、探索，共同努力",
      "mention_result": [
        {
          "mention": "一块去",
          "offset": 10,
          "kb_id": -1,
          "confidence": 1
        }
      ]
    }
  ]
}
```