

中国高校计算机大赛

2022 中国高校计算机大赛 —— 微信大数据挑战赛

通知

2016年，教育部高等学校计算机专业教学指导委员会、教育部高等学校工程类专业教学指导委员会、教育部高等学校大学计算机教学指导委员会、全国高等学校计算机教育学会合办了“中国高校计算机大”（China Collegiate Computing Contest, C4），前“中国高校计算机大”由全国高等学校计算机教育学会主办。大数据挑战是其中的一事，在2018-2021年均入选全国普通高校学排榜，得会各界的度关注和广泛好。

2022中国高校计算机大——微信大数据挑战（以下“大”）是由清华大学和微信事业合举办，云提供平台和源支持，以企业实场景和实敏数据为基，向全球开放的法。大旨在技的方式，提升人们对数据分析与处理的法与技术应用力，探大数据的核心学与技术，尝创新大数据技术，推动大数据的产学用。

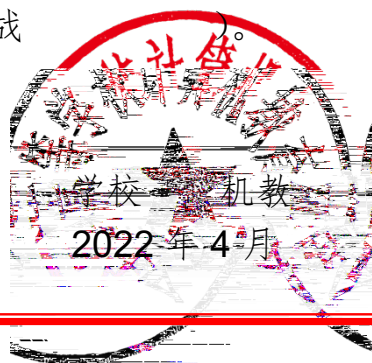
本次大向全球开放，不年国，校在校学生（包括专、本生、生）以及机构和企业从业人员均可报名参。参伍根据求应的法数据分析和处理，比果按照指定的价指标使用在测数据测和排名，得分最优。

各学校极合，按照和大做好宣传和作，为在校生和毕业生参与提供必的条件和支持。

情件（2022大数据挑战

全国高等学校计算机教育学会

2022年4月



2022 中国高校计算机大赛——微信大数据挑战赛 竞赛规程

2016

Computing Contest C4 China Collegiate
2018-2021

2022 —

一、参赛对象

- &
-

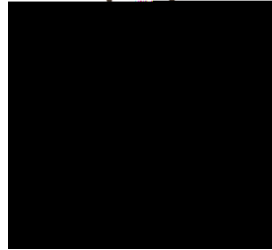
-
-
-
-

10

二、赛制说明

&

1. 报名&组队 (4月26日 - 6月21日)



大 名小 序
“微信大 ”

2022 4 26 10:00

2022 6 21 12:00

-
- 2-3
- <https://algo.weixin.qq.com/>
-
-
- data@tsinghua.edu.cn
- QQ 762146461 / 901317172

2. 初赛 (5月20日 - 6月22日)

5 20 10:00

- AB
- A 5 20 10:00 - 6 21 20:00
3
- B 6 22 10:00-20:00 6 21 21:00
2

TOP110

6 22 20:00

B

6 26

12:00

75

3. 复赛(7月1日-8月5日)

docker

“ ”

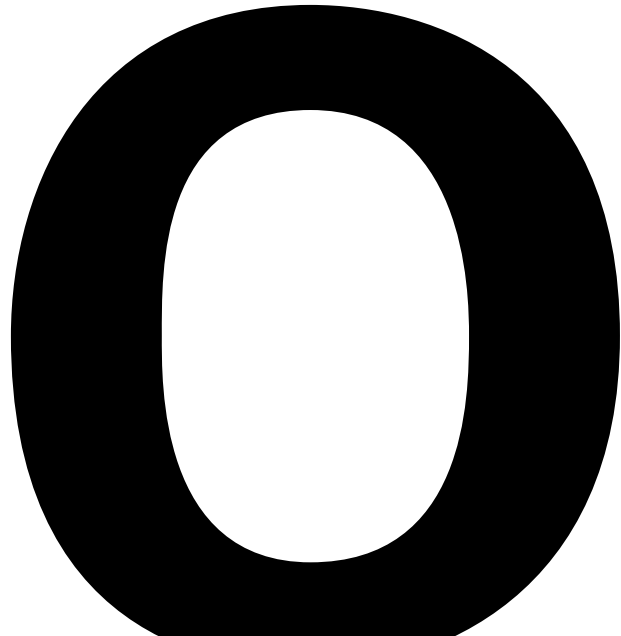
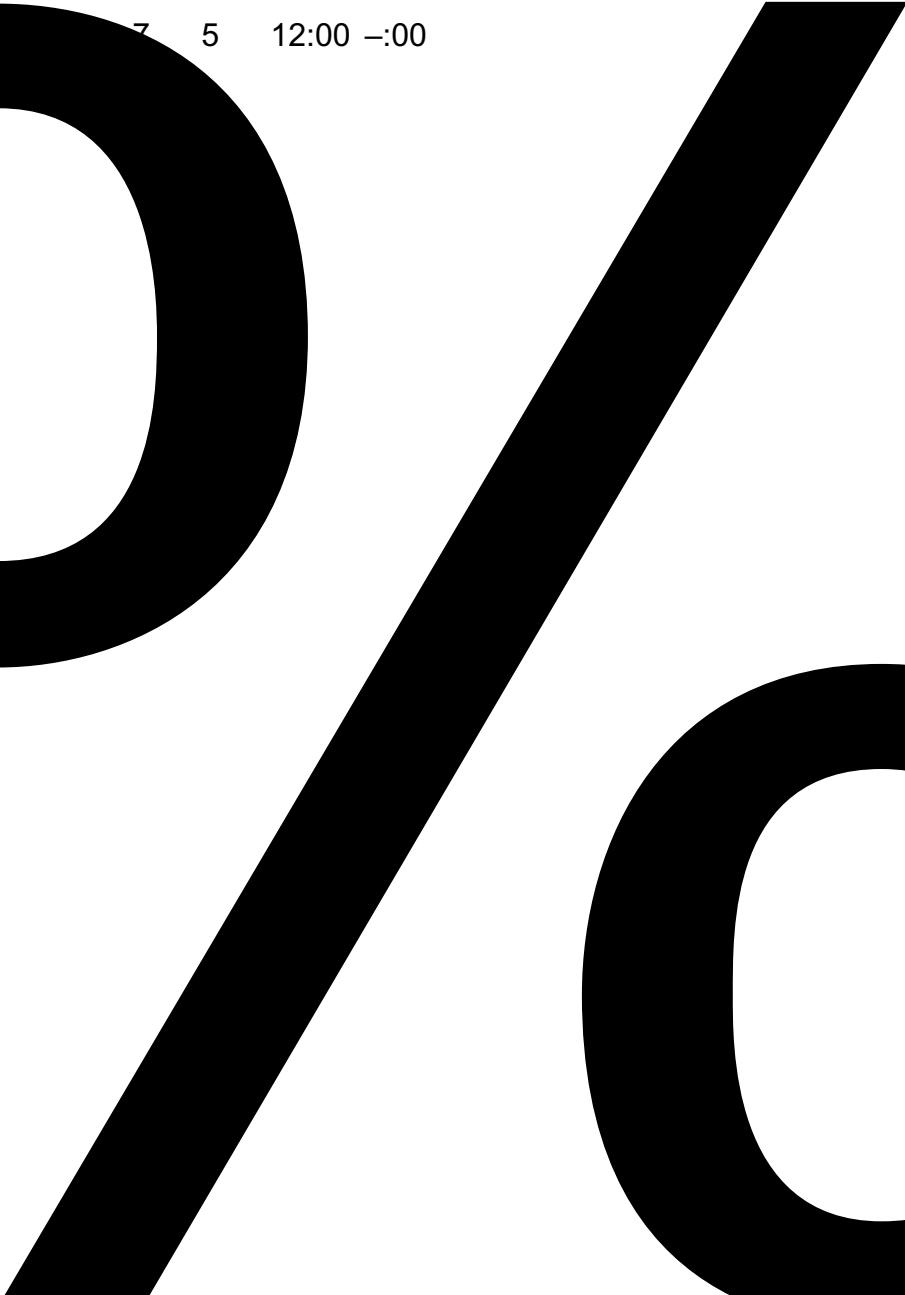
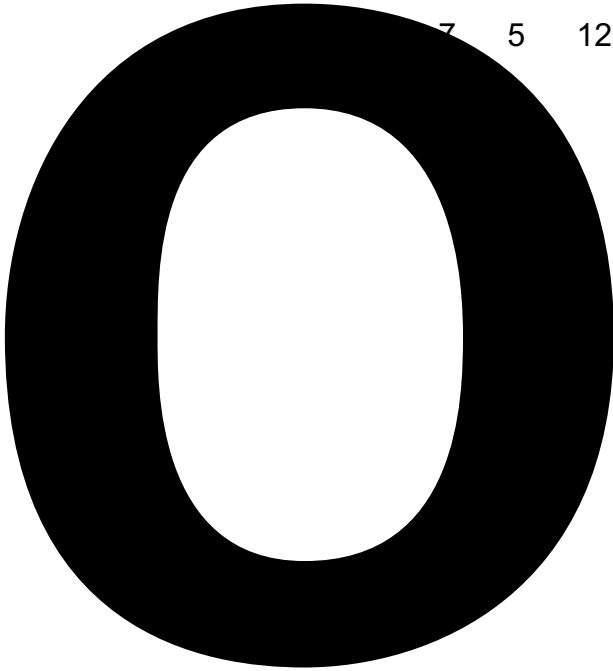
7 1 12:00 - 7 4 12:00

5

7 5 12:00 - 8 5 22:00

AB

7 5 12:00 - :00



三、奖项设置

1. 初赛奖项

TOP110

2. 复赛与决赛奖项

56

| | | |
|-------|----|----|
| | | |
| 1 | 1 | 30 |
| 2 | 1 | 10 |
| 3 | 1 | 6 |
| 4-6 | 3 | 2 |
| 7-10 | 4 | 1 |
| 11-30 | 20 | |

3. 在校学生队伍奖项

2022 4 26

| | | |
|--|----|-------|
| | | |
| | 10 | 1-10 |
| | 20 | 11-30 |
| | N | 31 |

4. 周周星

在 ， 奖励。从 开 ， 以 一中 12 排 为准，取前 名参 伍 ； 会 一个名 ， 以 保 伍中 一个在 伍 一个在 人员 伍。对 前 得 伍，不 ， 名 名 。

四、违规处理

参 应 实、公平 度参 ， 如在以下 ， 大 委 员会（ “ 委会”） 取 参 所在 伍 参 ， 报参 所在 其 。

1. : 参 保 信息 实性和 性， 参 人 ； 参 参 ， 一参 不可 个 、 分 ； 如 据 为参 在 ， 委会可以单 大 平 。

2. :

- 参 伍 。 如不 参 伍 度 ， 定 在 为 ， 委会 取 关参 伍 参 ， 关参 。
- 参 应保 其在 中所产 所 产 、 业 及其 合

5.

五、申诉与仲裁

1.

2.

3.

5

六、其他说明

1.

2.

! " # \$ % & ' — () * + , - . /

多模态短 分类是 理 域的基础技术之 , 在安全审核 推 内容搜索等 域 有着十分 常广泛的应用 条短 中 常包含有三种模态信息, 即文本 , 它们在 不同 义层 的分类体系中发挥着相互促 和 充的 作用 微信产品的内容生态繁 , 创作 盖 围大, 导 短 数据中普 存在着模态缺失 相关性弱 分类标签分布不均 等 , 是 实 应用中 着 决的技术 点 本 求参 基于微信 号短 数据以及对应的 分类标签标注, 用合理的机器学习技术对指定的测 短 分类 测

一、竞赛数据

比 分为初 和复 两个 段: 初 段提供百万 级的无标注数据和十万 级的有标注数据 用于 练; 复 段 练数据和初 相同, **主要区别是初赛阶段只提供视频抽帧特征, 而复赛阶段 提供视频抽帧原始图像** 初 段所有 练数据对参 开放下 ; 复 段的 练数据为 源 数据, 参 在[腾讯云 TI-ONE 平台](#)完成 练

1. 数据格式

| 字段名 | 类型 | 举例 | 说明 | 备注 |
|-------------|--------|----------------------|---------|----|
| id | string | 13655102198344648800 | 唯 ID | |
| category_id | string | 2117 | 人工 分 ID | |

2. 数据集

比赛提供的数据集有三个类别：无标注数据集、有标注数据集、测试数据集。各类数据集具体包含字段如下所示

| 字段 | 初赛 | | 复赛 | |
|----|-------|-------|-------|-------|
| | 训练数据集 | 测试数据集 | 训练数据集 | 测试数据集 |

三、其他说明

1. 本次比赛全程不允许使用外 数据 ；
2. 允许使用开源的 典 embedding 和 练模型，以上数据和模型 在复 开始前开源，且 件的形式报备开源 接地址和 md5，报备 箱为 wechat_algo@tencent.com