1. **会议信息**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **会议日期** | **会议时间** | **会议地点** | **发起部门** | **记录人** | **联系电话** |
| 2016-05-19 | 09:30 | S座306 | 690PSI | 胡艺友 |  |
| **出席人：**杨雪莹、卢瑾、许学平、胡艺友、黄裔 | | | | | |
| **发送：** | | | | | |

1. **会议主题**

|  |
| --- |
| **主题：云识云图系统介绍** |
| **目的**（拟解决的问题或产出的内容）**：**  了解云识云图系统当前业务及数据内容，分析众创汇项目未来数据来源 |
| **议程：** |

1. **会议纪要**

| No. | 纪要描述 |
| --- | --- |
|  | **云识**  后台数据分析系统，可以根据用户提出的话题进行 |
|  | **云图整体介绍**  目前海量数据是系统的服务供应商，海量每日会从互联网抓取各行业数据，海尔每日根据特定限制条件从海量的系统中抓取URL及对应内容。  数据内容：家电行业电商8大平台评论、标题等信息  新闻、贴吧、口碑、热词、话题、双微等信息  品牌：海尔及30个大中品牌竞品 |
|  | **云图电商数据**  标题、评论、URL、时间、昵称、区域、来源（e.g.京东、淘宝）、型号~标题解析（与海尔主数据不完全匹配）、  评论位置、规格参数全量爬取（未清洗）--、只有最新价格（不保留历史价格 & 打折机制）、销量不准确  数据年限：3年，可以抓取网络公开的历史数据  系统可以对评论进行情感分析（好、中、差）  数据量：2000万+条， 220G+，约80万/月  **现状**1：由于一个产品标题会对应多个评论(e.g. 标题：ASBK1123空调~1匹 & 2匹两个型号)，导致评论内容无法与正确的单一型号进行匹配。  **现状**2：规格数据为清洗  **现状**3：价格不保留历史信息，无从得知是否打折等  **现状**4：销量数据不准确  **现状**5：标题解析的型号与海尔主数据不完全匹配 |
|  | **舆情**  可以按照关键词、网站来源等范围在海尔/ 海量的数据库中抓取数据 |
|  | **云图指数：**  汇总结果，没有细化到访问用户 |
|  | **方正系统：微博合作🡪云图**  云图从方正系统获取微博方面数据  方正也有论坛等方面的数据，但是云图尚未从方正抓取此方面数据 |
|  | **产业的人会对抓取的评论信息进行业务处理** |

**四、本次会议调度及决议**

1、决议状态：完成（按计划完成）；进行中（尚未到达计划完成日期任务执行中或任务的计划延期且经过同意）；未完成（未按计划完成）。

2、各参会人员按照决议要求及时落实解决，并请按决议到位时间反馈闭环结果。

| No | 调度及决议项 | 描述 | 责任人 | 到位时间 | 闭环结果 |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**滚动待关闭决议**

| No | 调度及决议项 | 描述 | 责任人 | 到位  时间 | 闭环结果 |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

**五、评价方式(根据具体使用范围,输出可执行和落地的机制)**

1.决议拖期反馈, …… 2.决议未按进度要求完成,………