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** |
| 2016-05-17 | 10:30~12:00 | 创牌南601A | 690PSI | 胡艺友 |  |
| **出席人**：徐美霖，赵明国，卢瑾，杨鸣婵，李鑫，胡艺友，黄裔 | | | | | |
| **发送**：卢瑾，李鑫，黄裔，胡艺友  **抄送**：邹晓慧，张玉丽，Shoping | | | | | |

1. **会议主题**

|  |
| --- |
| **主题**：调研外部微博、论坛数据、以及官网社区等自建平台数据 |
| **目的**（拟解决的问题或产出的内容）： |
| **议程：** |

1. **会议纪要**

| No. | 纪要描述 |
| --- | --- |
| **外部数据调研（北大方正：微博、论坛、网站等数据源）** | |
|  | 目前官网运营平台主要负责除电商之外的第三方数据的抓取及分析，例如，微博、论坛，贴吧。  电商数据主要存储在云识云图平台。 |
|  | 数据逻辑来源：  Step1：北大方正对40,000+网站、论坛、贴吧进行数据漫爬  Step2：海尔提供家电相关关键词，从北大方正获取约150万/天的数据  Step3：海尔对此150万数据进一步通过关键词过滤掉无用数据，最终每日在系统中存储约50万条/天  数据存储约60M/天。海尔将信息以文字的形式进行保存，图片只保留URL。 |
|  | 数据的清洗、使用及分析：  近两三年海尔已经对这些数据进行了文本挖掘，通过聚类等算法得到的结论主要是噪音太大、能耗太高之类信息。  即使不进行挖掘，产业的人也知道这些事情，但是很难改变。  {建议} 通过U+的数据获取用户的使用数据、机器运转数据，以便对产品能耗方面做出一些改善。 |
|  | 之前发送过数据给产业，产业请尼尔森做过分析，但是没有得到太多有用结论。 |
|  | 数据以半结构化的方式进行存储，借口已经开放给DTS吴晓亮  如果需要可以完全开放给众创汇项目  但是只能提供与海尔产品相关数据，如果要获得其他方面数据需要向北大方正额外进行费用支付。 |
|  | 690微博目前由杨琦负责运营 |
| **海尔社区** | |
|  | 海尔社区论坛（bbs.haier.com）目前尚未进行文本挖掘分析。 |
|  | 注册信息为手机号、邮箱、用户名，部分数据与海尔会员信息已经打通 |

**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 |
| --- | --- | --- | --- | --- | --- |
|  | 数据源筛选 | 提供数据源筛选关键词 | 徐美霖 | 5月27日 |  |

**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,………