# **论文阅读总结**

## **论文基本情况**

《大数据背景下微博热搜的新闻阅读服务功能》是2017年吉林大学硕士研究生毛贺祺的毕业论文。该论文共分为四部分，分别为绪论、新浪微博热搜的特征、新浪微博热搜的数据化思维、传统媒体对微博热搜功能的利用。其中，传统媒体对微博热搜功能的利用是本论文的重点。本论文的落脚点是传统媒体如何运用微博热搜，挖掘读者阅读习惯，提升新闻服务质量。

## **研究方法**

作者统计整理了2016年12月4日—12月24日每天8：00、12：00、18：00、22：00 四个时间段的热搜内容，使用了数据对比、内容分析等研究方法。首先，作者将热搜内容分为五类，分别是政治新闻、社会新闻、娱乐新闻、经济新闻、生活信息，然后观察这五类新闻的总体占比，对比了五类新闻在不同时间段内的发展趋势。同时，作者通过追踪特定热门事件信息，分析了热搜话题之间的联系。

## **基本观点**

通过分析，作者得出如下观点：

**从总体占比方面看，娱乐新闻的占比超过一半，经济新闻搜索频次最低。**微博作为一个社交平台，其娱乐属性仍为最为关注内容，与用户日常生活关系密切的社会新闻及生活话题关注度也相对比较高。

**从不同时间段的发展来看，不同新闻类型在各时间段内的传播特点各有不同。**政治新闻和经济新闻在各个时间段的关注度一直处于较为平稳的状态；社会新闻的搜索热度随着时间的变化而发生变化，社会新闻的发布和传播主要集中于一天中的上午时段，而生活话题类的信息一般在晚间休闲时间内得到的关注度较高；娱乐新闻的传播则与其他新闻类型都不同，在 8:00 的时间段达到一天内关注度最高的时间，8:00-12:00 处于陡降的趋势，然后趋于平稳。

**从热搜话题之间的联系来看，相关新闻助推热搜榜单新闻更新频率。**通过新浪微博热搜榜，可以及时的对热门事件的信息进行追踪，同时还可以牵引出许多其他相关的话题，促使事态进一步蔓延。一条独立的新闻很难在微博热搜上持续的存在下去，如果要推动一条新闻得到更多用户的关注，就要不断创造新的衍生关键词，才能一直维持新闻热度。

## **存在的问题**

作者的分析主要有以下问题：

1. 样本量小，时间短，一天只选择了四个时间点进行观测，缺乏连续性。作者在时间框架的选择上费了一番考量，例如选择 20 天为一个周期，是因为爆炸性新闻的持续关注周期最长不会超过 2 个星期；选择 8：00、12：00、18：00、22：00 四个时间段，是因为 8：00 和 18：00 是上下班的高峰期，12:00 和 22：00 则选择在一天中两个休息的时间点。但是选取的理由仍然不够充分，首先，没有任何研究结果表明热点事件持续周期最长不会超过 2 个星期，其次，无论是上下班高峰还是休息时间，都应该是一个时间段，而非时间点，因此这样的选择较为主观，不够可靠。
2. 将热搜内容划分为五类，划分标准不够明确，例如社会新闻和生活信息区别较小，部分热搜不完全属于这五大类，分类太广泛，笼统的判断对主观依附太重，有失客观性。
3. 缺乏理论基础。论文中理论引用不明确，部分已经被理论证明的内容重复解释。
4. 结论与分析分离，关联度小。论文最后的落脚点放在传统媒体如何利用热搜榜上，所提出的方法与前文所分析的热搜特点毫无关系。

## **启示**

通过分析已发表的多篇论文，我们的研究首先解决了样本量小，时间短的问题。例如可以在每隔5分钟抓取一次的情况下，抓取一个月的热搜进行分析。在分析框架上，我认为我们可以从以下几个方面进行分析：

1. 从热搜类型的角度出发，判断各类热搜比重。在热搜类型划分上，直接选择抓取到的分类作为热搜类型，然后分析监测期内各个类型所占比重，以及不同类型在一天不同时间段的变化。同时考虑这种变化有无普适性，是否受到节假日等因素的影响。
2. 从话题贡献者的角度出发，观察各类话题贡献者的参与程度。哪些账号频繁参与热搜话题，央媒对于热搜话题有无贡献，是否参与了其中的议程设置。根据目前几个样本来看，人民日报、新华社等央媒极少成为话题贡献者。
3. 从热搜内容的角度出发，观察哪些词汇出现的频率更高，哪些词汇的共现次数更多。
4. 从热门话题的角度出发，一个热门话题往往会延伸出不同的内容，一个热点事件不会以完全一样的关键词出现在热搜榜上，新出现的关键词可能是事件的后续发展，也可能是与之相关的其他事件。在计算热点持续时间时，如果把关联热搜词计算进去，对某一热点事件的分析会更全面。