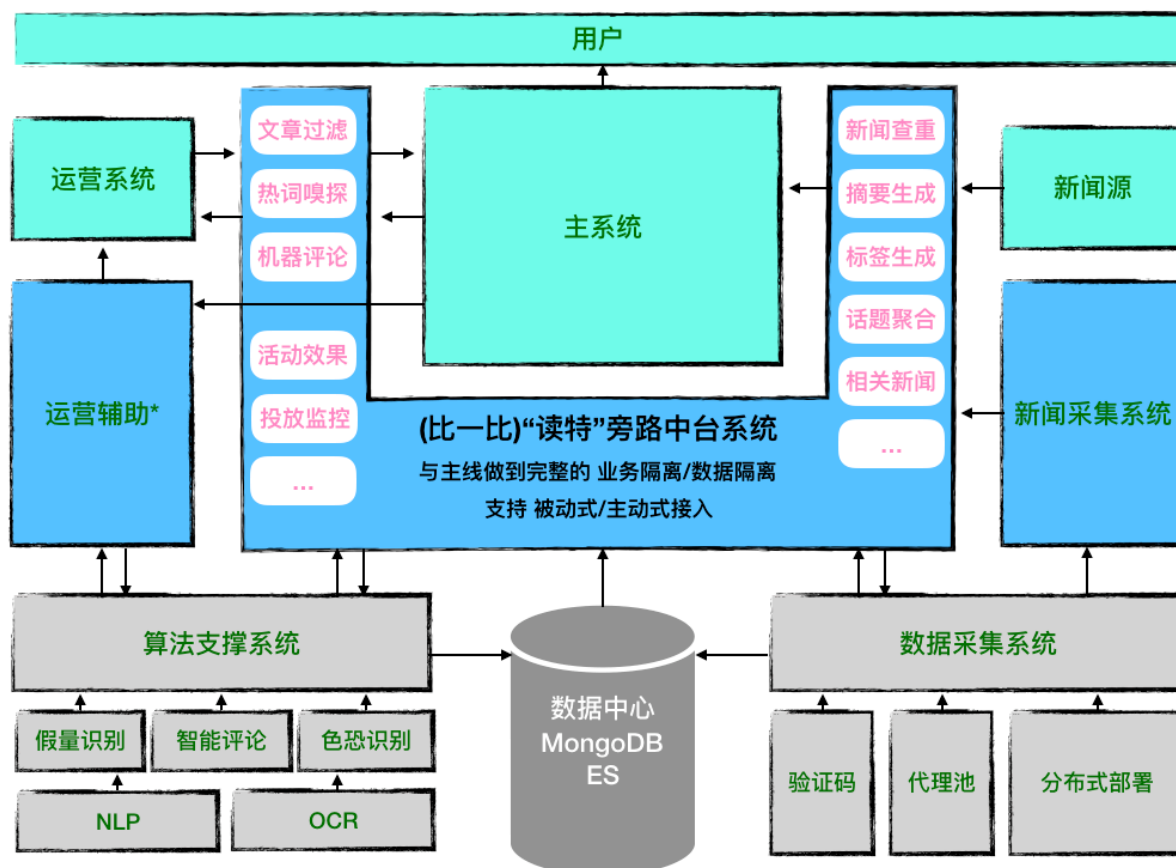


“独特”旁路中台系统技术架构图

方楠

- 架构图：



- 架构设计：

- 整体中台系统由两层构成，蓝色为应用层。灰色为基础服务层、粉色字为应用场景。
- 整体系统开发可以分为3个阶段完成。

阶段	时间	基本功能	涉及范围
1	1-2个月	数据采集系统+新闻采集系统+智能评论	2个应用场景
2	2-4个月	整体架构PaaS平台+40%算法支撑系统	3-5个应用场景
3	2-4个月	算法支撑系统+运营辅助系统	10个应用场景
4	2-4个月	用户画像+推荐系统+主系统优化	10+个应用场景

- 各个模块的使命和依赖

序号	模块	使命	依赖
1	数据中心	完成外部数据的存储	无
2	数据采集系统	完成外部数据源的采集	数据中心
3	新闻采集系统	完成对新闻场景的采集+清洗+去重	数据采集系统
4	算法支撑系统	完成旁路系统的算法支撑	主系统数据+数据采集系统+数据中心
5	运营辅助系统	完成对运营痛点的支持	算法支撑系统
*5	推荐系统	完成人物画像和协同推荐算法	主系统

- 需求分析与应答：

序号	需求点	应答方案	应答模块	应答阶段
1	新闻采集需求	通过对各个新闻渠道的采集，采集最新新闻。采集渠道可配置化，实时同步到主系统。	新闻采集系统	1
2	自动评论需求	一方面通过采集竞品的评论信息，通过评论自身的评论数点赞数等同步优秀评论。另一方面在第二阶段优化评论算法，能够自动筛选优秀评论，自动生成带交互评论（评论的评论）等。	新闻采集系统 算法支撑系统	1+2
3	假量识别需求	导入主系统的渠道流量数据，第二阶段通过一定的训练数据生成识别拦截模型。第三阶段建立自进化识别模型，通过机器学习的模式学习分析假渠道数据	算法支撑系统	2
4	智能运营平台	第一解决运营活动监测问题，能够对运营人员进行的活动通过对日活增长/留存率的分析从而对活动打分。第二接入假量识别模型和色情恐怖筛选过程中的模糊项，进行人工鉴别。第三根据运营中出现的问题支持纵向拓展。	运营辅助系统	2+3
5	模块服务供应需求	该需求是指接入第三方的一些本地服务，如天气或者政务。在本架构中未进行应答。	无	无
6	涉黄暴图片审核需求	通过成熟的涉黄涉暴识别OCR模型，自动识别暴露色情图片、涉及政治人物的图片、血腥恐怖等图片，并予以拦截。	算法支撑系统	3

序号	需求点	应答方案	应答模块	应答阶段
7	涉黄暴文字评论和文章筛查需求	通过成熟的涉黄涉暴识别NLP模型，自动对文本中的脏言脏语、地域黑人身攻击、敏感历史事件进行拦截。	算法支撑系统	2
8	首次发稿审核需求	通过对全渠道的新闻采集监控，在第三阶段进行首次发稿审核。通过文章指纹特征的对比，可以在500ms内完成对千万级资讯的相似度去重比对任务。	算法支撑系统 数据中心	3
9	版权侵权审核接口	在第三阶段数据采集系统已经积累了版权数据之后，对侵权风险的资料进行指纹化处理。自动对内容中的侵权风险进行高速识别（单资料的识别时间少于500ms）。预警侵权风险。	算法支撑系统 数据采集系统	3
10	用户画像和推荐算法需求	在第四阶段中，对主系统的发布流程进行修改。接入用户画像、用户标签、协同推荐等算法，通过千人千面的推荐算法，全方面提高资讯质量。	算法支撑系统 主系统	4