鹰眼社交化聆听系统需求设计说明书

1、引言

1.1、目的

社交网络的不断发展给人们的生活提供了诸多便利的同时也给了很多不法分子散步谣言和煽动性言论的便利，由于社交网络的虚拟性和隐蔽性，尤其是国外社交媒体，无法获取相关用户的具体信息，在发生重大群体性事件时，无法快速掌握群体情况和事件发展态势，给予相关部门在管理上造成了很大的困难。在此背景下，设计并研发了针对国内外社交媒体的社交化聆听系统。

1.2、背景

系统名称：Unicorn（鹰眼社交化聆听系统）

1.3、术语说明

2、项目概述

2.1、简介

通过定向批量采集国内外社交媒体的数据，多维度分析，对特定群体或者个人进行预警和管控。通过虚拟平行感知，将社交媒体中的虚拟身份与现实中的身份进行疑似关联。通过智能感应，自动发现同一个人同一个社交平台多个账号以及同一个人不同社交平台的多个账号。

2.2 、功能描述

数据源：定向智能数据采集系统

采集平台：暂时支持Twitter、SinaWeibo （Facebook、YouTube、Gmail等0.0.2版本）

分析模块：影响力活跃度分析、关系图谱分析、虚拟及现实账号关联分析、推文多维度分析。

2.3、用户特征和水平

最终用户属于业务员，无任何软件背景。

2.4、运行环境

Centos7.0 64位操作系统

2.5、假定和约束

运行本系统需要有能够连接访问国外网络的主机。

3、功能需求

3.1、功能划分

系统功能主要划分为三个大的部分，分别为定向智能数据采集、多维度分析、重点人员及群体管理。

3.2、功能描述

3.2.1、定向智能数据采集系统采集境内外社交媒体数据，采集的数据主要分为以下几种：

1、针对海外社交媒体，能够支持针对某种语言的账号信息抓取

2、能够支持针对昵称或者账号中含有某些特殊字符或者描述的账号信息抓取

3、账号信息涵盖用户的昵称、账号、注册时间、粉丝数、关注数、群组数、账号简介以及发表详细内容抓取。发表内容抓取部分要能够支持图片和视频等多媒体文件的抓取和展示。

3.2.2、 多维度分析模块功能分为以下几种：

1、通过账号相关基础信息建立模型，计算得出账号的活跃度和相关影响力依据

2、针对账号的社交关系建立关系图谱进行分析

3、根据昵称进行相近账号自动疑似关联功能

4、分析推文内容，自动关联所提到的敏感信息，主要包括姓名、邮箱、url、ip等

5、针对推文进行分词分析，通过文字云的形式进行展现，对推文进行文本情感分析

4、外部接口需求

4.1、用户界面

用户界面以简介为主，屏幕适配为宽屏。

4.2、硬件接口

硬件采用统一设备类型

4.3、软件接口

操作系统：Centos7

底层存储模块：HDFS、HBase、Elasticsearch、Mysql