****

**本科生毕业设计任务书**

题目：京津冀安全大数据获取与分析平台

**学生姓名 李天鹏**

**学院名称 软件学院**

**专 业**

**学 号**

**指导教师 王文俊**

**职 称 教授**

一、原始依据（包括设计或论文的工作基础、研究条件、应用环境、工作目的等。）

随着科学、技术和工程的迅猛发展，近20年来，许多领域(如光学观测、光学监控、健康医护、传感器、用户数据、互联网和金融公司以及供应链系统)都产生了海量的数据。大数据的数据集大小以难以想象的速度增长，给数据处理带来了极大的挑战[1-3]。

**本课题的研究目的**：研究大数据获取及分析技术，爬取京津冀安全相关的开源数据，并在Mediawiki和MVC上实现这些事件的语义浏览与统计分析，实现大数据分析平台化，为京津冀的安全提供策略保障。

**本课题的工作基础**：随着大数据的发展，大数据分析平台构架也是层出不穷，这使得信息分析方法的探索不断有新的突破[4]，学者们对大数据时代的情报分析技术与方法[5]进行了创新，包括研究大数据预测建模[6]，构建基于大数据的数据可视化方法，开发处理大数据的高效和安全的云存储系统[7]。大数据告诉我们数据量大不是困难所在，信息分析的关键而是在于对海量、复杂、非结构化数据的分析，不借助于专业的分析工具很难在规定的时间内完成分析任务， 或者很难在较短的时间范围内更多地发现大数据里潜藏着的情报价值[8,9]。

二、参考文献

1. 程学旗，靳小龙，王元卓等。大数据系统和分析技术综述[J]. 软件学报, 2014(9):1889-1908.
2. 李学龙, 龚海刚. 大数据系统综述[J]. 中国科学:信息科学, 2015, 45(1):1-44.
3. 肖源, 郝杰, 刘莹,等. 信息分析视角下的大数据分析平台构架研究[J]. 情报科学, 2016, V34(9):83-89.
4. Talia D. Clouds for scalable big data analytics[J]. Computer, 2013,46(5):98-101 .
5. Agrawal D, Das S, El Abbadi A. Big data and cloud computing: current state and future opportunities [C]// EDBT 2011, International Conference on Ex⁃tending Database Technology, Uppsala, Sweden,2011:530-533.
6. 程学旗, 靳小龙, 王元卓. 大数据系统和分析技术综述[J]. 软件学报, 2014, (9):1889-1908.
7. Morales GDF. SAMOA: A platform for mining big data streams. In: Proc. of the 22th Int’ l World Wide Web Conf. (WWW 2013). Rio de Janeiro: ACM Press.[EB/OL]. <http://www.engineeringvillage.com/search/doc/detailed.url?SEARCHID=M3862b207144>
8. 顾君忠. 大数据与大数据分析[J]. 软件产业与工程, 2013, (4):17-21.
9. 李广建, 化柏林. 大数据分析与情报分析关系辨析[J]. 中国图书馆学报, 2014,(5):14-22.

三、设计（研究）内容和要求（包括设计或研究内容、主要指标与技术参数，并根据课题性质对学生提出具体要求。）

**本课题的研究内容**：研究大数据爬取技术和分析技术，获取京津冀安全相关的数据，对这些数据进行信息抽取并分析，具体按照以下几步实施：

1. 从网上爬取京津冀所有路网、地块、单位的数据；
2. 从网上爬取京津冀所有突发事件新闻及其舆情数据；
3. 对这些数据进行信息抽取，形成地域-事件-人-行为多层异质网络，并进行多层异质复杂网络分析，挖掘安全事件背后的驱动机制；
4. 在MVC与Mediawiki上实现这些事件的语义浏览与统计分析。

**具体要求**：

1. 熟悉大数据获取及分析技术及其过程，并能够从理论上理解消化；
2. 对京津冀安全大数据进行信息抽取，形成语义报告；
3. 对京津冀安全数据进行语义实证分析，并能够解释所体现的现象，最好能够形成高水平刊物期刊论文；
4. 在semantic mediawiki上实现京津冀安全相关数据爬取、数据处理、语义浏览、统计分析、可视化等功能。

指导教师（签字）

年 月 日

审题小组组长（签字）

年 月 日