



北京大学

本科生毕业论文

国内外个人信息保护法政策内容、
媒体报道与公众舆论对比研究：

题目：以中国、韩国、欧盟为例

A comparative study of the policymaking,
media coverage and public debate about
personal information protection laws:
Evidence from China, Korea and the EU

姓 名：权九泽

学 号：1600093009

院 系：信息管理

专 业：信息管理与信息系统

导师姓名：闫蒲

二〇二二年五月

北京大学本科毕业论文导师评阅表

学生姓名	权九泽	学号	1600093009
院系	信息管理系	专业	信息管理与信息系统
指导教师	闫蒲	职称	助理教授
毕业论文题目	国内外个人信息保护法政策内容与公众舆论对比研究:以中国、韩国、欧盟为例		
导师是否同意参加 毕业论文答辩	是	建议成绩 (可选填)	90
导师评语	<p>该论文结合了定性的政策文本分析与定量的社会化媒体与主流媒体文本挖掘与分析,研究系统对比了三个国家个人信息保护法侧重点的异同,以及相关媒体与公众舆论讨论的差异。研究具有一定的研究价值与政策价值,论文结构清晰。</p> <p>导师签名: 闫蒲 2022 年 5 月 20 日</p>		

版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。

摘要

个人信息保护法保障个人信息使用、储存、分享的自由和权利，保护信息时代个体信息用户的尊严和价值。我们正在生活在第四次工业革命时代，即新技术高度发达的信息社会。现代社会广泛收集、分析、利用大量多样的信息。被恶意利用或遭到泄露的个人信息不仅会对数据所有者的个人生活造成很大的伤害，还会进一步危及公共安全。中国、韩国、欧盟等世界各国（地区）认识到现代社会的个人信息保护的重要性，正在加强相关政策制定、制度建设和技术保障。

各个国家的个人信息保护法制定背景不同，法规侧重点存在差异。本研究首先对各个国家和地区个人信息保护法的特点进行对比分析，其次，了解和分析各国主要媒体的舆论讨论，对比分析不同地区社会化媒体用户对个人信息保护法的公众舆论。本论文一方面为研究互联网治理公共政策提供跨国对比的实证研究支持，另一方面为完善个人信息保护制度提供切实可行的政策建议。

通过本研究得出的结论如下：第一，政策对比的研究发现，三国个人信息保护法侧重异同主要表现为 GDPR 比中国和韩国的政策更有严格的规定，更关注信息主体的同意和撤销权部分。中国和韩国的政策比 GDPR 更具有惩罚性；第二，通过对三国主流传统媒体与社会化媒体关于个人信息保护法相关舆论讨论的分析发现中国和欧盟的主流媒体集中于新技术与个人信息保护方面议题的讨论。在韩国的公共舆论上可以找到说明韩国政策与其他政策的不同点的话题。

关键词： 中国个人信息保护法，韩国个人信息保护法，GDPR，舆情挖掘与分析

ABSTRACT

The Personal Information Protection Act protects the freedom and rights of the individual and further ensures the dignity and value of individual users. We are living in the era of the fourth industrial revolution, with the information society driven by highly developed digital technologies. Technology companies in a digital society extensively collect, analyze and utilize a large and diverse amount of information. Personal information that is maliciously used or leaked not only leads to great harm to the personal lives of data owners, but further endangers public safety. China, South Korea, the European Union and other countries (regions) around the world recognize the importance of personal information protection in the information society and are strengthening internet policies, reinforcing institutional development and imposing technical safeguards.

The background of personal information protection law formulation differs from country to country, and there are differences in the focus of regulations. This study provides a comparative analysis of the characteristics of personal information protection laws in three different countries or regions, aiming to understand and analyze the public opinion around personal information protection laws in major media in each country. We also compare and analyze public opinion on personal information protection laws among social media users in different regions. This thesis provides empirical research support for cross-country comparisons to study public policy on Internet governance on the one hand, and practical policy recommendations for improving the policymaking of personal information protection laws.

The conclusions drawn from this study are as follows. First, the study of policy comparison found that the similarities and differences among the three countries' personal information protection laws are mainly in the form of regulations and penalties. the GDPR has stricter regulations than the policies of China and Korea, and is more concerned with the consent and revocation rights of information subjects. The policies of China and Korea are more punitive than the GDPR. Second, the analysis of mainstream traditional media and social media in the three countries reveals that the mainstream media in China and the EU focus on the discussion of new technologies and personal information protection issues. In the Korean public opinion, it is possible to find topics that illustrate the differences between the Korean policy and other policies.

KEY WORDS: [Personal Information Protection Law of the People's Republic of China, Personal Information Protection Act of the Republic of Korea, GDPR, Public opinion mining and analysis]

目 录

第一章 引言	1
1.1 研究背景	1
1.1.1 大数据时代：个人信息保护政策背景	1
1.1.2 各国个人信息保护的政策尝试	1
1.2 文献综述	2
1.3 研究方法	3
1.3.1 政策比较分析	3
1.3.2 舆论内容分析	4
1.3.3 自然语言处理分析方法	5
第二章 国内外个人信息保护法政策背景与政策要点	7
2.1 国内外个人信息保护法立法背景与过程	7
2.1.1 中华人民共和国个人信息保护法	7
2.1.2 大韩民国个人信息保护法 大韩民国个人信息保护法	7
2.1.3 欧盟通用数据保护条例(GDPR)	7
2.2 国内外个人信息保护法的政策内容比较分析	8
2.2.1 个人信息的定义	8
2.2.2 个人信息处理者	9
2.2.3 法规的适用范围	9
2.2.4 个人信息处理原则	10
2.2.5 个人信息处理的同意和撤销	10
2.2.6 与新技术应用相关的个人信息保护	13
2.2.7 个人信息的境外转移	14
2.2.8 个人信息主体的权利保障	15
2.2.9 违反法规时的法律责任	15
第三章 公众舆论分析	21
3.1 公众舆论分析比较	21
3.1.1 中华人民共和国公众舆论	21
3.1.2 大韩民国公众舆论	33

3.1.3 欧盟公众舆论	42
3.2 比较各国公众舆论	51
第四章 结论	51
4.1 研究结论	51
4.2 研究不足与下一步研究方向	52
参考文献	53
附录 A	55
致谢	62
北京大学学位论文原创性声明和使用授权说明	63

第一章 引言

1.1 研究背景

1.1.1 大数据时代：个人信息保护政策背景

在第四次工业革命时代，我们正生活在一个大数据、移动互联网、人工智能等技术高度发达的信息社会。通过大数据这一关键资源培育新产业正在成为一项重要的国家发展方略。孵化新兴技术产业需要利用人工智能、云计算、物联网等技术，并开发大数据资源。因此，各国均迫切需要建立数据安全使用的法律规范。在这种环境下，大量、多样的信息被广泛收集、分析、利用，个人信息也不可避免地包含其中。然而，在广泛而普遍的信息收集中，可能包含个人敏感信息或个人隐私数据，因此，在处理这些信息时，必须采取适当的法律保护措施和政策规范。在这种情况下，包括中国，欧盟、韩国等世界各国（地区）都认识到个人信息保护的重要性，并正在逐步加强这方面的法律规范、制度建设和技术保障。

根据 2015 年的《中国网络空间安全发展报告》，基于互联网电商、网络营销、电子支付等技术，具有极大的商业价值的网民的个人信息正在成为企业和不法分子的非法牟利的工具。因此，各国网民都面临的个人数据被互联网公司非法采集、交易的风险。由于互联网技术的快速发展，出现了法律制度，尤其是个人信息保护法的缺失。欠成熟的互联网法规让很多个人和企业对互联网个人信息的收集、存储和使用仍然依靠自我监管。同时，跨国公司带来的个人信息跨境转移也让个人信息保护的全球监管面临现实挑战。

1.1.2 各国个人信息保护的政策尝试

各国政府均着手针对科技公司侵犯个人信息、制造市场垄断制定惩罚性条款。比如，2016 年 3 月 20 日，德国联邦卡特尔局(Bundeskartellamt)以脸书公司(Facebook)在社交网络市场上对用户信息的基本设置（个人信息处理政策）相关服务条款滥用，及脸书的市场垄断行为，对该公司展开调查。2017 年 12 月 19 日，联邦卡特尔局公布了初步审查结果，认为 Facebook 通过第三方收集和使用个人信息的行为构成市场垄断。这里，“第三方”不仅包括 Facebook 集团的子公司 WhatsApp 和 Instagram，还包括在使用 Facebook 业务工具（如“Like 点赞”、“用 Facebook 登录”和“Facebook 分析”）时链接“Facebook API”的第三方网站。Facebook 要求其服务用户签署同意了“同意广泛使用个人信息”的相关条款，德国竞争主管部门将此判断为剥削性条件。因此，Facebook 的行为属于市场主导者强加的“不公平交易条件”，并违反《欧盟运作条约》(TFEU)第 102 条。这一问题也与违反《个人信息保护法》的行为密切相关，围绕着该问题，近期学术界和企业界对互联网反垄断与个人信息保护关系展开积极讨论，个人信息保护法的制定和落实引起国内外学术界密切关注^①。

从平台运营商的角度来看，随着人工智能时代的到来，利用大数据洞察市场需求、提高内容服务质量必须收集一定量的个人信息。但是，个人信息的收集需要确保用户的基本信息权利。各国政府在个人信息保护领域出台法律法规是网络治理发展的重要标志性事件之一，与此同时，公众对于个人信息保护的认识也有所提升。从全球视野看，不同地区、不同国家在个人信息保护方面都推出了行之有效的举措，已出台的法律法规，为全球化网络空间治理提供了明确的指引和参考。各国针对个人信息的收集和使用规定了用户的基本原则，而相关原则的界定由于各国的国情不同有所差异。与此同时，个人信息保护也出现了一系列国际上广泛达成共识的基本原则。

本研究在个人信息保护的重要背景下深入剖析各个国家在个人信息保护立法内容层面与公众舆论

^①Bundeskartellamt. Bundeskartellamt prohibits Facebook from combining user data from different sources Background information on the Bundeskartellamt's Facebook proceeding.[EB/OL]. https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Pressemitteilungen/2019/07_02_2019_Facebook_FAQs.pdf?blob=publicationFile&v=5, 2019-02-07/2022-05-12.

讨论方面的异同。该问题与科技反垄断、用户权利保护息息相关，正引发世界范围内学术研究、政策制定的广泛关注。对于中国、韩国和欧盟三个地区个人信息保护法律区别的分析，有助于我们更好的理解个人信息立法保护的异同。此外，对于三个地区的媒体和社会化媒体的舆论的分析，有助于理解主流媒体与公众对各个地区政策的想法。

1.2 文献综述

大数据时代个人信息已经成为重要的生产要素，个人数据信息具有量大、价值密度低、智能处理以及信息获得和其使用结果之间相关性弱等特征，使得个人无法按照自己的意志控制数据信息的产生、存储、转移和利用。随着大数据的发展，近年来国内外学者对个人信息保护政策进行了较为深入的研究。个人信息一词最早出现于1968年联合国“国际人权会议”中所提出的“数据保护”(data protection)。^①由于以智能手机为代表的移动设备快速更新换代和移动通信技术的发展，互联网进入高速发展阶段。而且随着云计算、物联网和以深度神经网络为代表的人工智能技术的发展，人们面临着个人信息被关联分析、过度开发利用和隐私泄露等问题。用户难以理解和控制与自己相关的个人信息的收集和使用。这一时期亚太地区的大部分国家意识到指定系统性的法律规范的重要性，在参考全球和各国既有隐私保护法律法规的基础上制定和颁布了本国针对数据时代个人信息保护的法律条例。

本研究从两个方向进行文献综述。第一，对于不同国家和地区个人信息保护政策的研究。第二，对于各国（地区）民众个人信息保护公众舆论的研究。

因为中国个人信息保护法的出现比其他国家地区的政策晚，所以各个学者针对的研究比较少。各个学者对中国个人信息保护法的评价为如下：于晓洋，何波（2022）说中国《个人信息保护法》是中国第一部在法律层面上的个人信息保护专项立法，为中国个人信息的深入保护、监督和管理提供了坚实有力的法律保障。王利明，丁晓东（2021）认为中国《个人信息保护法》有六个亮点：进一步扩大了个人信息的保护范围；全面规定了个人在信息处理中的权利；强化了个人信息处理者的保护义务；构建了名按个人信息的严格保护规则；规范了国家机关的个人信息处理行为；完善了个人信息的法律救济途径。

从对于不同国家和地区个人信息保护政策的研究方面看，学者们针对国际上其他国家的个人信息保护法的亮点进行点评。因为GDPR开始实施的时间比较早，分析了GDPR的先进法规，强调应该作为中国和韩国法律的一个很好的立法指南。然而，具体而言，对个人信息保护相关的法律的整体进行比较和分析较少。目前缺乏对韩国、中国和欧洲的所有个人信息保护法规进行系统比较和分析的实证研究。杨智博（2022）指出修订后的韩国《个人信息保护法》的主要变动有三点：第一，对“个人信息”的保护范围予以调整和明确；第二，缓和理解目的限制原则；第三，修改和增设不适用知情统一原则的具体情形。郭晓然，方元欣，曲东昕（2021）认为GDPR是“史上最严格的数据保护法”，将对全球数字经济产生深远的影响。刘云（2017）指出2016年发布的GDPR，从三个方面对个人信息保护行政执法机构的管理方式做了改革创新：第一、设立欧洲信息保护委员会并升级欧洲信息保护局工作战略；第二、建立欧盟内个人信息保护一站式管理服务模式；第三、实施风险等级差异化管理方法。Hyun-seok Yoon（2021）在其研究中发现，韩国和中国的《个人信息保护法》的异同在所辖部门、目的、个人信息的境外转移、处罚方面。Sung-hyun Kim, Chang-moo Lee（2018）分析了GDPR和韩国《个人信息保护法》对“个人信息”和“信息处理”的定义，而且指出境外转移、法律责任、适用范围方面的异同。盖立伟（2021）在其研究中指出GDPR较于中国个人信息保护政策主要异同有三个：个人信息保护规制对象不同、个人信息保护的行政监管法律责任不同、个人信息保护的行政监管主体和模式不同。

从对于各国（地区）民众个人信息保护公众舆论的研究方面看，Yuan, Feng, Danowski（2013）采用语义网络(semantic network analysis)分析来追寻微博上“隐私”的意义。他们发现微博所呈现的不仅是“隐私”话题对日常生活的重要影响，而且也讨论了新兴技术如何塑造我们思考“隐私”的方

^① 占南.国内外个人信息保护政策体系研究[J].图书情报知识,2019(05):120-129.

式。研究还发现微博提供了展示、实践和争论“隐私”的一个场所。新浪微博上关于“隐私”的讨论显示出公众对国家监控和商业利益入侵的日益敏感。Choi H S , Lee W S , Sohn S Y (2017)将 LDA(Latent Dirichlet Allocation)主题模型应用于 Scopus 数据库中 1972 年至 2015 年 9 月间出版的所有“个人信息隐私”文献的摘要。他们发现与隐私有关的领域越来越多样化,而且在脸书(Facebook)和 Orbit Showtime Network (OSN)的隐私研究一直在稳步增长,而法律领域的研究却在减少。

以往关于个人信息保护的政策和舆论研究有三个局限:第一,跨国跨地区的对比研究不足。即有实证研究往往基于一个国家或地区的政策或一个数据库的研究论文;第二,关于个人信息保护的政策研究不足,研究者指出与隐私有关政策的深入研究相对不足;第三,很少有实证研究把政策内容研究与公众舆论研究结合。在这个局限来看,分析三个地区的不同媒体对个人信息保护政策的舆论具有较高科学价值和政策价值。

个人信息等数据在企业和机构的立场上也被界定为可以创造附加价值的无形资产。但是一旦个人信息被恶意利用或遭到泄露,不仅会对数据所有者的个人生活造成很大的伤害,严重情况下还会危及公共安全(如医疗、社保等敏感数据遭到泄漏)。世界各国政策制定者认识到个人信息保护的重要性,并为此加强法律、制度、技术等方面的建设。与此同时,各国媒体与公众也开始意识到个人信息保护的重要性,并且围绕个人信息保护的立法实践和社会影响展开讨论。通过分析媒体和社会化媒体的舆论分析可以掌握关于个人信息保护法公众舆论讨论的主要主题、理解不同国家相关舆论的异同,并可以根据舆论挖掘与分析补充对各个地区个人信息保护法的宏观掌控。

本研究通过对比研究信息技术产业较为先进的三个国家和地区的《个人信息保护法》政策内容以及关于个人信息保护法的新闻媒体与社交媒体的舆论内容,比较分析三种法规在政策内容和公众影响两个维度的异同。论文融合了定性的政策分析法与定量的文本挖掘与分析法,同时兼顾政策制定者与公众两种视角。研究兼具科学意义和应用前景:一方面为研究互联网治理公共政策提供跨国对比的实证研究支持,另一方面为完善个人信息保护制度提供切实可行的政策建议。

本研究的目如下:

第一,理解个人信息保护法的政策重心,在系统梳理各个国家和地区个人信息保护法制定过程和主要内容后,对各个国家和地区个人信息保护法的特点进行对比分析。

第二,挖掘和分析各个国家和地区主要传统纸质媒体与电视媒体关于个人信息保护法相关报道,了解和分析各国主要媒体的舆论讨论。

第三,挖掘和分析各个国家和地区主要社会化媒体平台关于个人信息保护法的公众评论,对比分析不同地区社会化媒体用户对个人信息保护法的公众舆论。

1.3 研究方法

本研究将通过定性和定量方法的结合研究三个国家和地区个人信息保护法的政策内容和公众舆论差异。研究方法包括:1)以公共政策文本分析的方法比较各国个人信息保护法的立法背景和主要内容;2)以内容挖掘与应用自然语言处理相关方法定量分析对比各个国家和地区主流新闻媒体和社交网络平台关于个人信息保护法的媒体讨论与公共舆论。政策内容的分析有助于引导研究对比公众舆论在政策内容不同维度上的相关讨论,而对于公众舆论的分析也反过来促进研究者回顾政策内容中个人信息保护法立法中的国别差异,以及相关舆论对公众个人信息保护法态度的影响。具体研究方法的研究路线图参考图 1.1。

1.3.1 政策比较分析

虽然各国的个人信息保护法在内容上有相似的条款,但是由于各个国家的个人信息保护法制定背景不同,法规侧重点存在差异。因此本研究将围绕中国、欧盟、韩国三个国家或地区个人信息保护法的

立法背景，讨论个人信息保护法的立法环境。并系统对比各个国家和地区在个人信息保护法中关于个人信息的定义、个人信息处理者、法规的适用范围、个人信息处理原则、个人信息处理的同意和撤销、与新技术应用相关的个人信息保护、个人信息的境外转移、个人信息主体的权利保障、违反法规时的法律责任共九个方面。本研究将通过从 EUR-Lex、全国人民代表大会、Korea Internet & Security Agency（韩国 KISA）、韩国法律中心，获得关于个人信息法律法规的文本内容，整理了各个政策的特性后从九个维度进行内容分析和对比研究。

1.3.2 舆论内容分析

为了分析传统媒体与社交媒体平台关于个人信息保护法的相关舆论，作者将从中国、欧盟（以德国为代表）、韩国的新闻媒体爬取个人信息保护法相关的新闻数据，并在几个国家具有代表性的社会化媒体平台收集新媒体讨论的相关数据。爬取中国的数据时候，用“个人信息保护法”；爬取韩国的数据时候，用“개인정보보호법”；爬取欧盟的数据时候，用“GDPR”。

新闻媒体选定的标准为：在各个国家或地区选定以电视为代表的传统媒体和以报纸为代表的传统媒体。中国新闻媒体选定了中国中央电视台官网和人民日报官网；韩国新闻媒体选定了 SBS (Seoul Broadcasting System) 和朝鲜日报官网；欧盟新闻媒体选定了 EuroNews 和 DWnews 官网；。社会化媒体平台选定的标准如下：比较了国际上影响最大的媒体和在中国影响最大的社会化媒体。其中，推特和微博较有代表性：两个平台都要求用户在有限的字符数内传播核心信息，同时被用户用作参与公共讨论的场域场。

表 1.1 不同国家和提取代表性主流媒体媒体类型、数据集规模、数据时间范围

国家	媒体	媒体类型	初步爬取数据集规模	初步爬取数据时间范围	数据清理后数据集规模	数据清理后数据集时间范围
中国	人民日报	传统媒体：报纸	3269	19 年 01 月 02 日到 21 年 01 月 01 日	277	21 年 10 月 12 日到 21 年 12 月 30 日
	央视网	传统媒体：电视	300	12 年 08 月 30 日到 21 年 12 月 29 日	58	20 年 10 月 12 日到 21 年 12 月 29 日
	微博	社会化媒体	282	16 年 04 月 18 日到 22 年 01 月 05 日	259	21 年 10 月 31 日到 22 年 01 月 01 日
韩国	朝鲜日报	传统媒体：报纸	1553	93 年 03 月 17 日到 21 年 12 月 31 日	205	20 年 02 月 05 日到 21 年 12 月 31 日
	SBS	传统媒体：电视	1047	99 年 09 月 13 日到 21 年 12 月 31 日	129	20 年 02 月 03 日到 21 年 12 月 31 日
	推特	社会化媒体	2245	20 年 01 月 01 日到 21 年 12 月 31 日	2152	20 年 02 月 01 日到 21 年 12 月 31 日
欧盟 (德国)	DW news	传统媒体：报纸	53	18 年 04 月 11 日到 21 年 09 月 02 日	53	18 年 04 月 11 日到 21 年 09 月 02 日
	Euronews	传统媒体：电视	101	18 年 02 月 27 日到 21 年 12 月 16 日	98	18 年 02 月 27 日到 21 年 12 月 16 日
	推特	社会化媒体	4354	18 年 01 月 01 日到 22 年 01 月 01 日	3861	18 年 01 月 01 日到 21 年 12 月 31 日

初步爬取数据集规模指的是第一轮数据收集时的数据规模，数据清理后数据集规模指的是经过数据清理后的数据规模。初步爬取数据时间范围指的是第一轮数据收集的数据的发布期间范围，数据清理后数据集时间范围指的是经过数据清理后的数据的发布期间范围。

在收集方法上,除了人民日报官网以外,传统新闻媒体都利用 BeautifulSoup 和 Requests 进行内容爬取,人民日报官网、推特和微博利用了 Github 和 CSDN 提供的 API 和代码依据“个人信息保护法”和“GDPR”关键词分别进行了社交媒体数据爬取。推特本身提供的 API 只提供最近 7 日数据,因此找到了新的爬取工具“Scweet”爬取发布于数据收集时间 7 日前相关内容,本人在 2022 年 1 月 1 日收集数据。爬取人民日报内容的时候,由于相关数据量较大,考虑数据收集时间限制,只收集了前 100 页的新闻内容。

1.3.3 自然语言处理分析方法

研究使用 Python 作为程序语言清理、分析传统媒体与社交媒体中关于各国个人信息保护法的相关内容。自然语言分析的数据预处理使用了各语言提供的库。韩语是 Konlpy, 中文是 jieba, 英语是 nltk。首先删除冠词和助词等非常用词汇,并在单词上进行词性标记工作 (Part-of-Speech Tagging), 选择动词、名词、形容词作为文本分析对象。根据词组出现的频率,通过 Bigram 识别常见词组,比如网页和链接经常共同出现,可合并成“网页_链接”。研究利用 TF-IDF 进行关键词提取和 LDA 话题建模进行自然语言处理分析。TF-IDF 是一种关键词统计法。TF 指词频, IDF 指逆向文件频率。TF-IDF 的公式如下:

$$TF-IDF = \frac{\text{在某一类中词条}\omega\text{出现的次数}}{\text{该类中所有的词条数目}} * \log \left(\frac{\text{语料库的文档总数}}{\text{包含词条}\omega\text{的文档数} + 1} \right)$$

某个词的 TF-IDF 值越大,该词的重要度越高。本研究利用 TF-IDF 方式提取各个媒体讨论的关键词,并使用 TF-IDF 得分制作政策公众舆论讨论词云。普通频次分析的缺点是,在所有文件中出现频率过高的词会给出很高的数值,尽管它们实际上不是有意义的词,而 IDF 可以控制这个部分。

LDA 以事先了解的主题单词数分布为基础,通过分析在文本信息中发现的单词数分布,可以预测相关文件将被分类为什么样的主题。结果解释更容易,而且解决了过拟合 (overfitting), 因此有利于从庞大的非典型数据中得出各种主题。以提取的数据为基础进行 LDA 话题建模 (topic modelling), 分析了各国主流媒体和社交媒体平台的个人信息保护相关舆论。研究通过模拟不同主题数量 k 的 coherence value 得分,并且结合对不同主题数量 k 下主题内容的区分度,选择最佳主题数量。

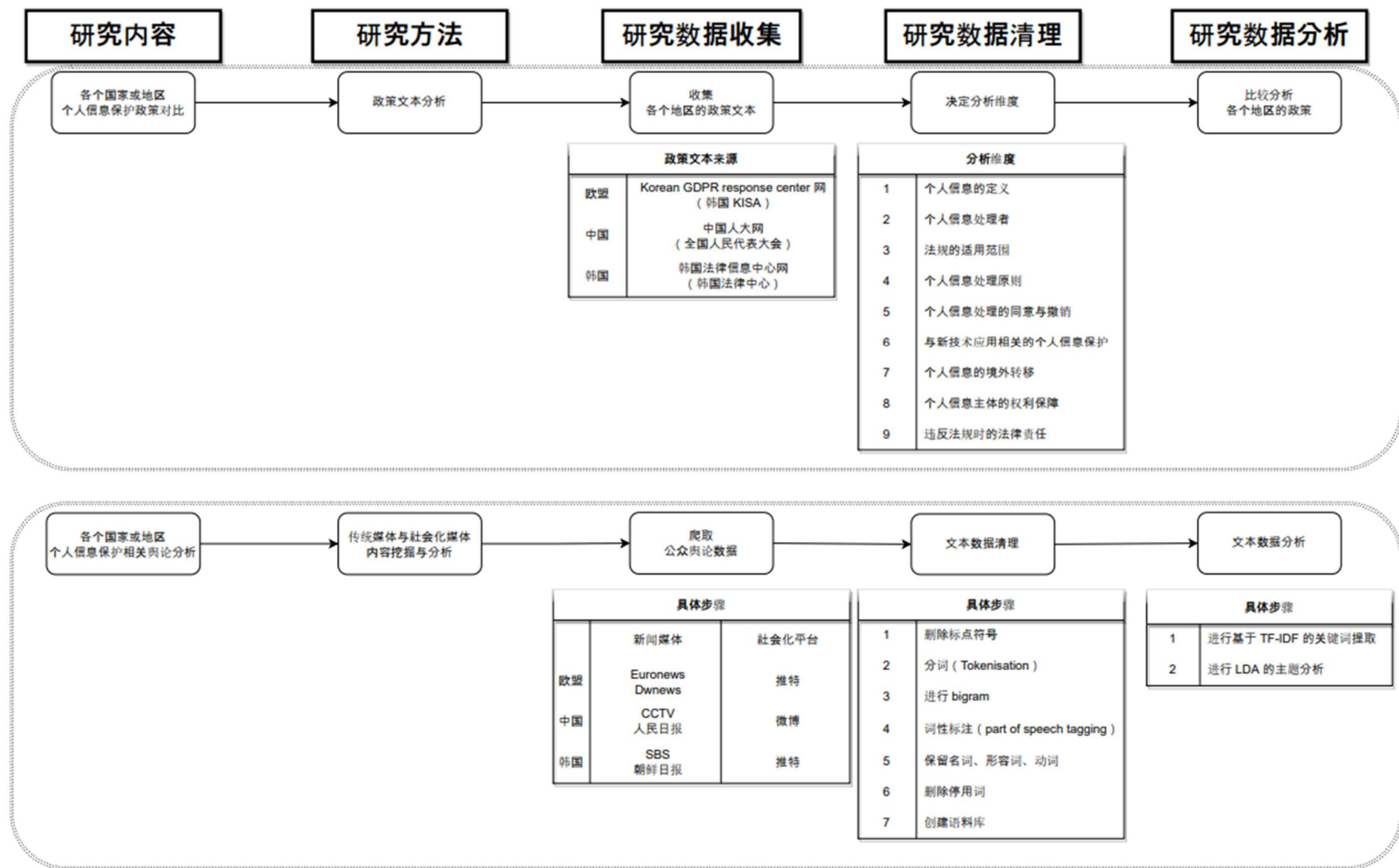


图 1.1 研究路线图

第二章 国内外个人信息保护法政策背景与政策要点

2.1 国内外个人信息保护法立法背景与过程

2.1.1 中华人民共和国个人信息保护法

中国通过完善个人信息保护法，保护其公民的个人信息，限制个人信息向国外转移。个人信息保护法草案提请十三届全国人大常委会第二十二次会议审议^①。十三届全国人大常委会第三十次会议表决通过《中华人民共和国个人信息保护法》，自 2021 年 11 月 1 日起施行。《个人信息保护法》确立了公益诉讼条款，加强了对个人信息侵权行为的法律规制，加大了对个人信息领域权益的保护力度^②。

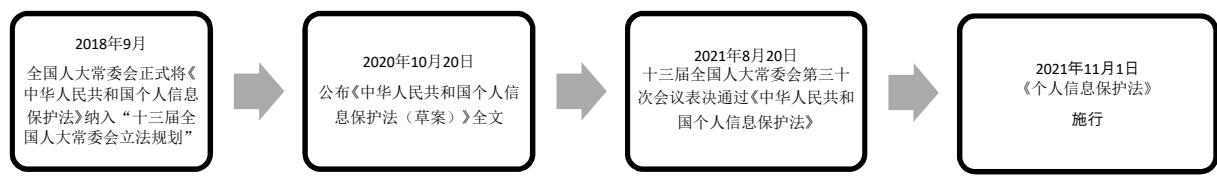


图 2.1 中国个人信息保护法推进时间轴

2.1.2 大韩民国个人信息保护法 大韩民国个人信息保护法

韩国从 2011 年开始制定并实施个人信息保护法。但是，现行法律上个人信息保护监督功能和保护相关法令不统一，相对比较分散。在为了培育新产业而支持数据利用灵活性方面感到了局限性。因此，为了促进个人信息保护与相关产业协调发展，2020 年系统地完善了个人信息保护相关法令。

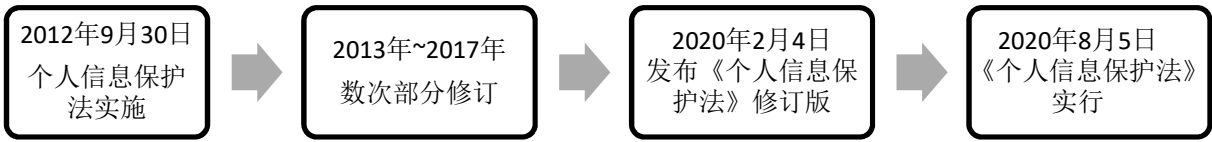


图 2.2 韩国个人信息保护法推进时间轴

2.1.3 欧盟通用数据保护条例(GDPR)

由于欧盟数据保护条例 95/46/EC 是一个“条例”性约束，具有一定局限性，因此很难引起欧盟成员国的积极配合及有力执行。此外，该保护条例显然没有跟上个人数据的收集、访问和使用方式的快速技术发展的转变，以及科技全球化的步伐。对于在欧盟多个国家经营的公司来说，要注意每个成员国的法律，存在着不便之处，但随着该条例的颁布，商业便利性将进一步加强。因此，个人信息保护的监管体系在整个欧盟并不统一，导致了个人信息保护处理中的法律不确定性。颁布将统一适用于所有欧盟成员国的《个人信息保护法》的动向需求日益显著。GDPR 草案于 2012 年公布，随后，通过欧洲议会、欧盟理事会和欧盟委员会的三方协商，于 2016 年 5 月 4 日在欧盟媒体上颁布，于 2018 年 5 月 25 日正式生效。

^① 夏红真 . 个人信息保护法草案首次亮相. [J/OL].

<http://www.npc.gov.cn/npc/c30834/202010/569490b5b76a49c292e64c416da8c994.shtml>, 2020-10-15/2022-05-12.

^② 张倩. 个人信息保护法施行两个多月 公益诉讼成为个人信息保护利器. [N/OL]. http://www.zgcsjs.org.cn/News-news_list-id-58042.aspx, 2022-01-11/2022-03-12.

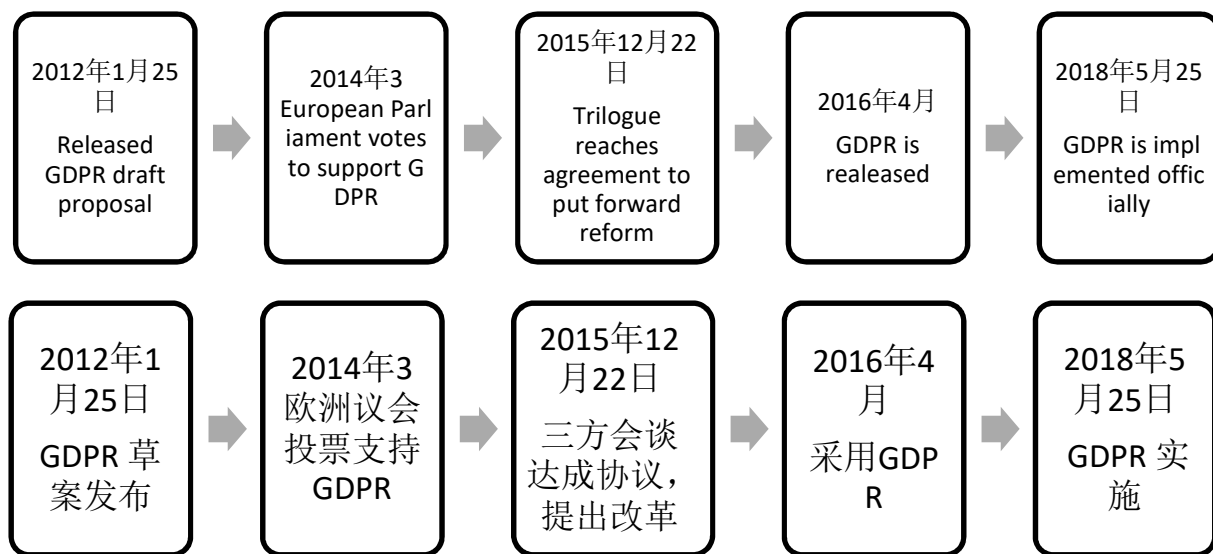


图 2.3 GDPR 推进时间轴

2.2 国内外个人信息保护法的政策内容比较分析

2.2.1 个人信息的定义

中国的《个人信息保护法》的定义与韩国《个人信息保护法》和 GDPR 的个人信息定义相比, 虽然很简单, 但是规定为“可以识别个人的信息”, 中国《个人信息保护法》和 GDPR 中的个人信息与韩国《个人信息保护法》中规定的个人信息相对应^①。而且, 对于敏感信息的适用对象也有一些差异。韩国和欧洲包括政治见解及特定集团加入与否、性相关信息, 欧洲还包括种族或民族血统信息。

在中国的《个人信息保护法》中, “个人信息”是以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息^②, 受法律保护。由于在信息特性上强调“识别性”, 不包括匿名化处理后的信息。在记录方式上, 规定为以电子方式或其他方式记录的信息, 所以对记录方式没有限制。个人信息的处理包括个人信息的收集、存储、使用、加工、传输、提供、公开、删除等^③。

韩国的《个人信息保护法》将“个人信息”定义为关于活着的个人的信息, 并规定“通过姓名、身份证号码、视频等可以识别的信息(仅凭该信息, 即使无法识别特定个人, 也包括与其他信息容易结合而可识别的信息)”^④。

在 GDPR 中, “个人信息”是指与已识别或可识别的自然人(“信息主体”)有关的任何信息。可识别的自然人是指可以直接或间接识别的人, 特别是通过参考诸如姓名、识别号码、位置数据、在线标识符等标识符或与该自然人的身体、生理、遗传、精神、经济、文化或社会身份有关的一个或多个具体因素^⑤。GDPR 更明确地规定个人信息。

敏感信息的界定也是各国个人信息保护法的要点, GDPR 原则上禁止处理特殊类别的个人信息, 即敏感信息。GDPR 在第 9 条(Article 9)中单独规定了一般敏感信息。韩国个人信息保护法将“敏感信息”规定为思想、信念、工会和政党退出、政治见解、健康、性生活等相关信息, 除此之外, 有可能明显侵犯信息主体私生活的个人信息^⑥。敏感信息也包括遗传信息和犯罪经历资料相关的信息。在中国《个人

^① [韩]박노형. 한국 개인정보보호법과의 비교를 통한 EU GDPR 의 이해[J]. 언론중재위원회 2018:62-71.

^② 《中国个人信息保护法》 第四条

^③ 《中国个人信息保护法》 第四条

^④ 《韩国个人信息保护法》 第二条 第一款

^⑤ 《GDPR》第四条 第一款

^⑥ 《韩国个人信息保护法》 第二十三条 第一款

信息保护法》中，敏感个人信息是一旦泄露或者非法使用，容易导致自然人的人格尊严受到侵害或者人身、财产安全受到危害的个人信息，包括生物识别、宗教信仰、特定身份、医疗健康、金融账户、行踪轨迹等信息^①。

2.2.2 个人信息处理者

与中国和韩国的《个人信息保护法》中所说的个人信息处理者相比，GDPR 中对于个人信息处理相关的主体进行了控制员(controller)、处理员(processor)、代理人(representative)、数据保护员(data protection officer, 简称 DPO)等更加细化的规定。

在中国的《个人信息保护法》中，“个人信息处理者”是指在个人信息处理活动中自主决定处理目的、处理方式的组织、个人^②。

在韩国的《个人信息保护法》中，“个人信息处理者”是指以业务为目的运用个人信息文件，通过自己或他人处理个人信息的公共机关、法人、团体及个人等^③。

在欧盟的 GDPR 中，“控制员(controller)”是指单独或与他人共同决定处理个人数据的目的和方式的自然人或法人、公共当局、机构或其他团体^④。“处理员(processor)”是指代表控制者处理个人数据的自然人或法人、公共当局、机构或其他机构^⑤。

在 GDPR 中，决定是否控制员的标准不是取决于个人信息处理行为，而是取决于谁有权做出决定。因此，这可能源于处理个人信息的目的和手段有关的法律责任或实际影响的存在或不存在。确定是否为控制员的另一个标准是不为其他公司处理个人信息，而是为自己的商业目的处理个人信息。控制员有责任遵守隐私原则，并且必须能够证明其遵守了这一原则。特别是，控制员还负责采取适当的技术和行政措施来保护个人信息，并考虑处理的性质、范围和目的以及对自然人的权利和自由的风险的可能性和严重性^⑥。

中国个人信息处理者与欧盟 GDPR 上的“控制员”、“处理员”在意义上略有差异，但更接近于“控制员”的意义。“控制员”的定义中“共同决定处理个人数据的目的和方式的”，这一部分接近于中国的信息处理者定义的“是指在个人信息处理活动中自主决定处理目的、处理方式的”。根据韩国的《个人信息保护法》，个人信息处理者可以看作是与 GDPR 控制员相应的概念，但很难看作是完全一致的概念。因为，如果个人信息处理者以个人信息处理的形态为中心，那么 GDPR 则侧重于确定处理的目的和手段^⑦。

2.2.3 法规的适用范围

相比于韩国法规，中国和欧盟的个人信息保护法考虑到了全球化时代个人数据的流动性，并且根据跨国数据流动制定了相应法规。中国的《个人信息保护法》和 GDPR 详细定义了法规的适用范围，还定义了境外使用个人信息是否适用。与中国和韩国的《个人信息保护法》和 GDPR 相比较，可以看出韩国的不足之处在于，它只是将法规的适用范围界定为个人信息处理者。

中国的个人信息保护法规定，不仅是在中国国内的个人信息处理活动，在中国国外处理中国国内收集的个人信息的情况下，也属于该法的适用对象^⑧。其特点是，在适用于中国国内个人信息处理活动的所谓“属地主义原则”的基础上，增加了“域外效力”。在中国境外处理个人信息的活动有下列情形：第一、以向境内自然人提供产品或者服务为目的；第二、分析、评估境内自然人的行为；第三、法律、

^① 《中国个人信息保护法》 第二十八条

^② 《中国个人信息保护法》 第八章 第七十三条 第一款

^③ 《韩国个人信息保护法》 第二条 第五项

^④ 《GDPR》 第四条 第七款

^⑤ 《GDPR》 第四条 第八款

^⑥ 《GDPR》 第二十五条和第三十二条

^⑦ [韩]박노형, 고환경, 구태언. EU 개인정보보호법-GDPR 을 중심으로[M]. 서울:박영사. 2017: 37.

^⑧ 《中国个人信息保护法》 第三条

行政法规规定的其他情形^①。比如,如果在韩国成立的游戏公司向中国的个人发布游戏或提供在线服务,将成为中国《个人信息保护法》适用对象。第三条规定的中国境外的个人信息处理者,应当在中国境内设立专门机构或者指定代表,负责处理个人信息保护相关事务,并将有关机构的名称或者代表的姓名、联系方式等报送履行个人信息保护职责的部门^②。

韩国的《个人信息保护法》与中国的《个人信息保护法》和 GDPR 不同,没有规定详细的适用范围。该法只以个人信息处理者规定为适用范围。

对于 GDPR 的领土适用范围,处理和利用欧盟区域内信息主体的个人信息时,将成为适用对象^③。在此,应注意本地区信息主体不是国籍标准,而是地理标准。和中国《个人信息保护法》一样,即使工作地点在欧盟之外,向位于区域内的消费者提供财物及服务时,也适用该规定。欧盟内进行信息收集及处理的所有企业(包括境外企业)必须“指定代理人(designate a representative)”并运用。对于定期收集公共机关、政府机关及敏感信息的企业,应任命“数据保护员(DPO)”

此外,即使个人信息在欧盟以外的地区被处理,同样适用 GDPR^④。欧盟区域内的企业通过提供次要服务(如销售或营销),如果欧盟地区以外的其他工作场所支持相关个人信息的处理,将成为充分的适用条件。如果韩国的分销公司向居住在欧盟地区的韩国人提供商品和服务,那么它就必须遵守这一规定。例如,当居住在欧洲的韩国人从韩国内的网上商城购买商品时,就是这种情况。

2.2.4 个人信息处理原则

在个人信息处理原则方面,三个政策的内容都相似。但不同点为中国的个人信息保护法里没有提到数据最小化的问题,韩国有最低限度个人信息,欧盟有数据最小化。

在中国的《个人信息保护法》中,处理个人信息应当遵循合法、正当、必要和诚信原则,应当具有明确、合理的目的,并应当与处理目的直接相关^⑤。而且,处理个人信息应当遵循公开、透明原则,公开个人信息处理规则,明示处理的目的、方式和范围,应当保证个人信息的质量、准确性^⑥。个人信息处理者应当对其个人信息处理活动负责,并采取必要措施保障所处理的个人信息的安全^⑦。

韩国的《个人信息保护法》中,处理个人信息应明确个人信息的处理目的,并在其目的所需的范围内,合法、正当地收集、处理、利用最低限度的个人信息,确保个人信息的准确性、完整性及最新性。此外,个人信息必须得到安全管理,信息主体的权利必须得到保障。在处理个人信息时,应尽量减少对信息主体隐私的侵犯,如果可以进行匿名处理,则应进行匿名处理,如果不能通过匿名处理达到目的,则应进行假名处理。最后,必须努力获得信息主体的信任^⑧。

GDPR 的个人信息处理原则也与中国和韩国的《个人信息保护法》相似。在 GDPR 中,处理个人信息应以合法、公平和透明的方式进行处理;为特定、明确和合法的目的而收集,并且不以与这些目的不相符的方式进一步处理。而且,GDPR 强调数据最小化(data minimisation)、准确性(accuracy)、完整性(integrity)、保密性(confidentiality)和责任感(accountability)^⑨。

2.2.5 个人信息处理的同意和撤销

GDPR 比中国和韩国的《个人信息保护法》有更严格的个人信息处理规定。特别是从企业的角度来看,由于保证了人力资源管理的便利性和合法性,中国和韩国的《个人信息保护法》比 GDPR 更具有企业亲和性。

^① 《中国个人信息保护法》 第三条 第二款

^② 《中国个人信息保护法》 第五十三条

^③ 《GDPR》 第二十七条

^④ 《GDPR》 第三条 第一款

^⑤ 《中国个人信息保护法》 第五条、第六条

^⑥ 《中国个人信息保护法》 第七条、第八条

^⑦ 《中国个人信息保护法》 第九条

^⑧ 《韩国个人信息保护法》 第三条

^⑨ 《GDPR》 第五条

中国的《个人信息保护法》规定，处理个人信息时必须得到相关个人的同意^①。然而，也有一些允许在未经个人同意的情况下处理个人信息的例外情况。在属于《个人信息保护法》第13条第2款至第7款的任何情况下，不需要个人同意^②。特别是“为订立、履行个人作为一方当事人的合同所必需，或者按照依法制定的劳动规章制度和依法签订的集体合同实施人力资源管理所必需；“，通过第二款的例外规定，规范了在履行合同和管理公司员工人事方面处理个人信息的合法性。”为应对突发公共卫生事件，或者紧急情况下为保护自然人的生命健康和财产安全所必需；“的情况下，也不需要个人同意。互联网经营者必须按照合法、正当的原则收集个人信息，公开其利用规则，明确个人信息收集的目的、方式和范围，并征得提供者的同意。

表 2.1 中国《个人信息保护法》的个人信息处理的例外情况

中国《个人信息保护法》的个人信息处理的例外情况
<p>(二) 为订立、履行个人作为一方当事人的合同所必需，或者按照依法制定的劳动规章制度和依法签订的集体合同实施人力资源管理所必需；</p> <p>(三) 为履行法定职责或者法定义务所必需；</p> <p>(四) 为应对突发公共卫生事件，或者紧急情况下为保护自然人的生命健康和财产安全所必需；</p> <p>(五) 为公共利益实施新闻报道、舆论监督等行为，在合理的范围内处理个人信息；</p> <p>(六) 依照本法规定在合理的范围内处理个人自行公开或者其他已经合法公开的个人信息；</p> <p>(七) 法律、行政法规规定的其他情形。</p>

中国的《个人信息保护法》还确保个人信息提供者的同意撤销权。基于个人同意处理个人信息的，个人有权撤回其同意。个人信息处理者应当提供便捷的撤回同意的方式^③。但是，撤回个人同意不会对撤回前基于个人同意的个人信息处理活动的效力产生影响。

其中，应注意“单独同意”。这一规定适用于个人信息处理者向他人提供个人信息的情况，单独同意必须明确如收集和处理方法，以及对个人的影响等，充分保证知情权，并允许个人决定是否同意。例如，如果通过在公共场所安装图像采集或个人身份识别设备获得的生物识别信息和身份信息等个人信息为公共安全目的向公共机构提供，则不需要获得“单独同意”。然而，如果它被用于其他私人目的，它必须采取向信息提供者征求单独同意的措施^④。

韩国的《个人信息保护法》在得到信息主体的同意的情况下，可以收集个人信息，并在其收集目的范围内使用，和中国一样增加了例外事项。

表 2.2 韩国《个人信息保护法》的个人信息处理的例外情况

韩国《个人信息保护法》的个人信息处理的例外情况 ^⑤
<p>2. 在法律有特殊规定的情况下，或必须遵守法律规定的义务的情况下。</p> <p>3. 在公共机构不可避免地要履行法规等规定的职责时。</p> <p>4. 在与信息主体签订和执行合同时不可避免地需要的情况。</p> <p>5. 在信息主体或其法定代表人无法表达其意愿或因地址不详等原因无法事先获得同意的情况</p>

^① 《中国个人信息保护法》第十三条 第一款

^② 《中国个人信息保护法》第十三条 第二款-第七款

^③ 《中国个人信息保护法》第十五条

^④ 《中国个人信息保护法》第二十六条

^⑤ 《韩国个人信息保护法》第十五条 第一款 第二项-第六项

下，并明确认为对信息主体或第三方的紧急生命、身体或财产利益是必要的情况下。

6. 为实现个人信息处理者的合法利益所必需，并且明显优先于信息主体的权利的情况。在这种情况下，仅限于与个人信息处理者的合法利益有重大关系的情况且不超过合理范围的情况。

个人信息处理者在得到第 1 款第 1 项同意时，他应将以下事项通知信息主体，即使以下任何事项发生变化，也应予以通知并获得同意：第一、收集和使用个人信息的目的；第二、收集的个人信息项目；第三、个人信息拥有、利用期间；第四、有权拒绝同意的事实及拒绝同意而造成的不利情况时，其不利的内容^①。

个人信息处理者在当初与收集目的合理相关的范围内，考虑到是否对信息主体造成不利影响、是否采取了加密等确保安全性的必要措施等，根据总统令的规定，个人信息处理者可以在未经信息主体同意的情况下使用个人信息。

韩国的《个人信息保护法》没有表示撤回对处理个人信息的同意，但表示要求停止处理个人信息。信息主体可以要求个人信息处理者停止处理本人的个人信息。当收到停止请求时，应毫不迟疑地停止对个人信息的全部处理或信息主体的部分请求^②。然而，增加了可以拒绝信息主体停止处理要求的例外条款。

表 2.3 韩国《个人信息保护法》的拒绝信息主题的听一直要求的例外情况

韩国《个人信息保护法》的拒绝信息主体的停止要求的例外情况 ^③
1. 在法律有特殊规定的情况下，或者为了遵守法律义务而不可避免的情况下。
2. 有可能伤害他人的生命和身体或不公平地侵犯他人的财产和其他利益时。
3. 如果公共机构不处理个人信息就无法履行其他法律规定的职责时。
4. 如果不处理个人信息，无法提供与信息主体约定的服务等难以履行合同的情况，以及信息主体没有明确表示终止合同的意向时。

当个人信息处理者根据第二款的内容拒绝停止处理请求时，他应毫不拖延地将原因告知信息主体。对于应信息主体要求暂停处理的个人信息，个人信息处理者应毫不迟疑地采取销毁相应个人信息等必要措施^④。

GDPR 也和个人信息保护法一样，规定企业只有得到信息主体的同意(Explicit consent)才能处理个人信息。信息主体的“同意”是指信息主体自由地、具体地、知情地和毫不含糊地表明其意愿，通过声明或明确的肯定行动(by clear affirmative action)，表示同意处理与他或她有关的个人信息^⑤。

知情同意(informed consent)是处理个人信息的合法理由之一，明确的同意被认为是极其重要的伦理规范，因为它成为处理敏感信息、包括分析(profiling)的自动决策或与个人信息境外转移有关的处理的合法基础。个人信息处理活动总是可以通过同意而合法化，但在使用同意时需要谨慎。这是因为获得同意可能不仅在商业上是不可能的，而且可能要承担法律风险。因此，如果处理活动可以通过其他法律理由合法化，在追求控制员或第三方的合法利益的情况下，处理活动必须通过其他法律理由合法化^⑥。

与中国和韩国的《个人信息保护法》有些不同的是，GDPR 只有在无法使用其他法律依据的情况下才需要同意。这是由于 GDPR 对得到有效同意规定了严格的条件，而且这种同意可以在任何时候被撤销。特别要注意的是，必须为每个目的分别获得同意。例如，如果要根据同意为多个目的处理个人信息，就必须提供一个带有多个复选框的表格，而不能作为一个整体获得同意。还需要有一个程序，以适当的方

^① 《韩国个人信息保护法》 第十五条 第二款

^② 《韩国个人信息保护法》 第三十七条 第一款

^③ 《韩国个人信息保护法》 第三十七条 第二款

^④ 《韩国个人信息保护法》 第三十七条 第三款-第四款

^⑤ 《GDPR》 第四条 第十一款

^⑥ 《GDPR》 第六条 第一款

式记录获得同意的事实。如果同意是通过电子方式获得的,应该可以通过准备解释这种方法来证明符合 GDPR 的规定。另一方面,在公共利益、法律诉讼以及信息提供者和信息处理者之间履行合同的情况下,与非同意有关的例外情况^①。

与撤销个人信息同意一样, GDPR 与中国和韩国的《个人信息保护法》一样,可以随时撤销对提供个人信息的同意。如果不同意,也有权删除发布在互联网上的个人信息。然而,当相关信息需要用于公共利益、科学或历史研究目的或统计目的时,会有例外^②。

2.2.6 与新技术应用相关的个人信息保护

所有法律都没有提到特定的新技术。中国的《个人信息保护法》将焦点放在新技术使用个人信息可能带来的不利影响,要求保证透明度和进行影响评估。欧洲的 GDPR 认为新技术所需的数据几乎是个人信息,要求通过匿名处理可得的安全性以及对目的变更的适当措施。韩国修订后变得模糊不清,是因为根据法律解释的不同,个人信息保护与否也会发生变化。

通过人工智能(AI)、大数据和算法等新 IT 技术处理个人信息,会给个人带来不便和不利。一些公司通过大数据或基于自动化决策算法的个人信息分析向消费者提供差异化的价格^③。大数据是一种利用大量个人信息的方法,以更好地了解企业,从现有的个人信息中得出有趣的结论。为了防止这种情况发生,中国《个人信息保护法》增加了“保证决策的透明度”、“不得实行不合理的差别待遇”等条款。第 24 条明确规定:“通过自动化决策方式向个人进行信息推送、商业营销,应当同时提供不针对其个人特征的选项,或者向个人提供便捷的拒绝方式^④。”

根据中国的《个人信息保护法》,个人信息处理者利用个人信息进行自动化决策时,不应针对交易价格等交易条件进行不合理的差别待遇,以自动化决策方式向个人进行特定信息的推送营销时,必须提供不针对其个人特征的选项或提供简单、便捷的拒绝接收方式^⑤。因此,企业想要将个人信息用于广告性推送营销等时,不仅要遵守上述规定,还要根据《个人信息保护法》第 55 条,履行事前个人信息保护影响评估。

修订前的韩国《个人信息保护法》在国际上被评价为是一部创新但强有力的个人信息保护法^⑥。《个人信息保护法》第 3 条规定,应明确处理个人信息的目的,只在该目的所需的范围内处理最少的个人信息。关于允许使用个人信息的目的以外的处理规定,首先,禁止使用和提供超出该法第 15 条和第 17 条范围的个人信息,但如果“为了制定统计及学术研究等目的,必要时以无法辨认特定个人的形式提供个人信息”,则允许例外。第二,《个人信息保护法》第 2 条将个人信息处理定义为“个人信息的收集、生成、连接、联动、记录、储存、拥有、加工、编辑、搜索、输出、更正、恢复、利用、提供、公开、废弃以及其他类似行为^⑦”,以及“个人信息的目的以外的利用、提供限制”的修订前的第 18 条仅限于这些处理中的利用和提供。

作为利用个人信息的最大障碍的是,《个人信息保护法》第 18 条第 2 款第 4 项只允许在目的之外的“利用和提供”。这一条款在 2020 年 2 月修订时被删除。从个人信息保护的角度来看,《个人信息保护法》中“为了制定统计及学术研究等目的的需要”这句话可以通过对“等”的解释来扩大。“等”的解释对于大数据分析技术的使用来说,可以说是可取的,但从保护信息主体的法律稳定性角度来看,就不可取了^⑧。

GDPR 由于规定的各种制约事项,在使用大数据时,有许多应注意的限制。首先, GDPR 认为用于利用

^① [韩] 황화경, 홍창표. 중국과 EU 의 개인정보보호 규정 비교와 시사점[J]. 중국지역연구. 2021 (8):215-239.

^② 《GDPR》 第十七条 第三款

^③ 史洪举. 消费者不是“韭菜”,大数据得讲“规矩”. [J/OL] 海外网 i. 2022-01-06/2202-03-12

^④ 《中国个人信息保护法》 第二十四条

^⑤ 《中国个人信息保护法》 第二十四条.

^⑥ Graham Greenleaf. Asian Data Privacy Laws: Trade & Human Rights Perspectives[M] OUP Oxford. 2014:122

^⑦ 《韩国个人信息保护法》 第二条

^⑧ [韩] 박노형, 정명현. 빅데이터 분석기술 활성화를 위한 개인정보보호법의 개선 방안 - EU GDPR 과의 비교 분석을 중심으로[J]. 고려법학. 2017:1-39 -

大数据的信息几乎都是个人信息。用于大数据分析的个人信息数量越多，具体个人信息之间的联系就越大，所以重新识别自然人的可能性也会增加。因此，在使用大数据时，使用匿名的个人信息是安全和可取的。由于对大数据的数据集进行适当的匿名处理是一个非常困难的问题，如果不确定是否能够实现这一目标，最好能够遵守 GDPR 的相关规定。

关于大数据和 GDPR 的一些常见的几个注意事项如下。第一，要注意向信息主体提供个人信息处理方法和相关信息^①。但是一般来说，企业在收集个人信息后，决定使用与大数据相关的个人信息。这意味着更改处理目的^②。在这种情况下，必须依法以透明的方式向信息主体提供关于该计划的信息。这可能需要与信息主体进行讨论。第二，如上所述，当个人信息被用于大数据时，处理的目的一般会发生变化^③，必须为这一特定目的确定其他法律依据。有时可以根据合法利益来处理个人信息^④。根据实际情况，如如何使用个人信息，重新识别信息主体的风险程度，个人信息的范畴和重要性，大数据可能需要其他方式的合法化（例如，单独取得信息主体的同意）。第三，可以查看个人信息的个人信息处理员（如服务提供商）经常参与大数据项目，为了让处理员参与，根据 GDPR 第 28 条，需要签订个人信息处理合同和其他保护措施。因此，在允许服务提供商查看个人信息之前，应该注意这一点。

2.2.7 个人信息的境外转移

与中国和韩国的《个人信息保护法》中所说的个人信息的境外转移方面相比，中国的《个人信息保护法》和欧盟的 GDPR 在境外转移需要严格的标准和确认程序。但韩国的《个人信息保护法》只要求信息主体的同意，并规定相关信息的通知。从这一点来看，韩国的《个人信息保护法》比其他政策对境外转移的规定不足。

在中国经营的外国公司特别关注的部分之一就是他们是否可以向海外提供（或转移）在中国获得的个人信息。在这方面，中国的《个人信息保护法》与 GDPR 一样，规定了个人信息向海外提供的内容。

根据中国的《个人信息保护法》，在中国获得的个人信息可以向海外提供。但是要满足以下三个条件才能实现。第一，‘海外’的个人信息保护水平不能低于中国^⑤。第二，一定要取得个人的单独同意^⑥。第三，应当通过国家网信部门组织的安全评估^⑦。

在中国国内获得的个人信息也会根据主体的性质和处理数量决定是否在中国储存。关键信息基础设施运营者和处理个人信息达到国家网信部门规定数量的个人信息处理者，应当将在中华人民共和国境内收集和产生的个人信息存储在境内。否则，中国国内的储存规定不适用，可以放在海外^⑧。另外，分析信息类型后，如果是与国家安全及公共安全、公共利益相关的敏感信息，必须放在中国境内。

不同的是，中国的《个人信息保护法》规定，关键信息基础设施运营者和个人信息处理数量达到国家网信部门规定数量的一般个人信息处理者必须遵守向国外转移个人信息的规定在将个人信息转移到境外时必须遵守的事项。一般个人信息处理者向境外转移个人信息时需要遵守的事项有：第一，境外接受方应采取必要措施，符合《个人信息保护法》规定的个人信息保护标准；第二，向个人告知个人信息海外转移相关内容（领取方的名称或姓名、联系方式、处理目的和方式等），取得个人的同意；第三，事先进行个人信息保护影响评估；第四，与境外接收方订立合同，约定双方的权利和义务行^⑨。

韩国的《个人信息保护法》只通过通知信息主体以下事项并征求其同意的话，可以进行海外转移：第一、转移个人信息的项目；第二、转移个人信息的国家、日期和方法；第三、个人信息接收方的姓

^① 《GDPR》 第十三条-第十四条

^② 《GDPR》 第六条 第四款

^③ 《GDPR》 第六条 第四款

^④ 《GDPR》 第六条 第一款 第 f 项

^⑤ 《中国个人信息保护法》 第三十八条

^⑥ 《中国个人信息保护法》 第三十九条

^⑦ 《中国个人信息保护法》 第四十条和第五十五条 第四项

^⑧ 《中国个人信息保护法》 第四十条

^⑨ 《中国个人信息保护法》 第三十八条、第三十九条、第五十五条

名（指公司名称和信息管理经理的联系信息）；第四、个人信息的使用目及拥有、利用期间^①。

GDPR 将个人信息转移到欧盟境外的第三者视为处理个人信息的行为。个人信息被转移到欧盟境外，即转移到个人信息保护措施不合适的国家时，GDPR 要求使用适当的替代方法来确保个人信息保护的最低标准。这意味着，欧盟国家公民个人信息的境外转移，只有在第三国具有与欧盟相应水平的信息保护体系并保证安全水平的情况下才能进行。

而 GDPR 规定了几种实现这一目标的方法。如果根据 GDPR 第 45 条的“充分性决议 (adequacy decision) ^②”，某个国家被认为是安全国家，那么在向该国转移个人信息时，除了 GDPR 的第五六条的措施外，没有必要采取其他措施。换句话说，根据规定，对个人信息保护的充分性进行评估，只对被认定为具有高保护体系的国家批准信息转移。此外，位于未经欧盟批准的第三国的公司只有在证明信息通过欧盟行为准则和遵守公司规则得到安全保护的情况下，才允许境外转移。总之，总是需要一个程序来检查是否有 GDPR 第 6 条规定的有效法律依据，以及是否符合 GDPR 与个人信息境外转移有关的规定的额外要求。

2.2.8 个人信息主体的权利保障

在个人信息主体的权利保障方面，三个政策都保障基本权利。

中国个人信息保护法第四章包含了个人在个人信息处理活动中拥有的权利相关内容。个人对其个人信息的处理享有知情权、决定权，有权限制或者拒绝他人对其个人信息进行处理^③，具有查阅权、复制权、转移权^④，有权请求删除、补充^⑤、删除^⑥。

韩国《个人信息保护法》第一章第四条载有关于信息主体在其个人信息处理方面的权利。信息主体有权获得有关处理个人信息的信息，有权选择和决定是否同意个人信息处理和同意范围等，有权确认个人信息处理与否和阅览个人信息，有权要求停止、纠正、删除和销毁个人信息。最后，信息主体有权有权按照迅速、公正的程序得到救济个人信息处理造成的损失^⑦。

GDPR 保障接收信息的权利 (Right to be informed)、阅览权 (Right of access)、更正权 (Right to rectification)、信息删除权 (Right to erasure)、个人信息移动权 (Right to data portability)、反对权 (Right to object)、处理限制权 (Right to restriction of processing)、自动化决定及分析相关权利 (Right to related to automated decision making and profiling) 等相关信息提供主体的权利^⑧。特别是新引进了现有方针 (DPD) 中没有的删除权、个人信息移动权、自动化决定 (包括分析) 等相关权利，进一步强化了个人信息主体的权利保障^⑨。

2.2.9 违反法规时的法律责任

在违反法规时的法律责任方面，中国和韩国的《个人信息保护法》与 GDPR 相似，但中国和韩国的《个人信息保护法》对违法行为的制裁程度更高。如果问题严重，GDPR 和中国的《个人信息保护法》允许罚款，最高可达企业销售额的 4-5%，在处以大额罚款方面是相似的，但中国还规定可以“吊销相关业务许可或吊销营业执照”。此外，韩国对违规行为处以不超过销售额 3% 的过徵金，但规定对违规

^① 《韩国个人信息保护法》第三十九条的 12

^② 到目前为止，欧盟委员会已经承认安道尔、阿根廷、加拿大、法罗群岛、根西岛、以色列、马恩岛、日本、泽西岛、新西兰、大韩民国、瑞士、在 GDPR 和 LED 下的英国，以及乌拉圭提供了充分的保护。

(https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en)

^③ 《中国个人信息保护法》第四章 第四十四条

^④ 《中国个人信息保护法》第四章 第四十五条

^⑤ 《中国个人信息保护法》第四章 第四十六条

^⑥ 《中国个人信息保护法》第四章 第四十七条

^⑦ 《韩国个人信息保护法》第一章 第四条

^⑧ 《GDPR》第十三条-第二十二條

^⑨ [韩]최경진. EU GDPR 의 분석 및 시사점 [J]. NAVER Privacy White Paper. 2016:137-179

行为可以判处监禁。

在中国的《个人信息保护法》中，处理个人信息时，如果违反本法规定或不诚实履行法律规定的个人信息保护义务，将根据相关规定承担法律责任。相关类型可能会受到相关部门的责令改正或警告，或者被没收违法所得。对于应用程序的话，可能会责令暂停或者终止提供服务。如果不采取纠正措施，法人将被处以一百万元以下的罚款，直接负责的主管人员和其他直接责任人员被处以一万元以上十万元以下罚款。如果情节严重，可对法人处五千万元以下或者上一年度营业额百分之五以下罚款，并可以责令暂停相关业务或者停业整顿、通报有关主管部门吊销相关业务许可或者吊销营业执照。对直接负责的主管人员和其他直接责任人员处十万元以上一百万元以下罚款，并可以决定禁止其在一定期限内担任相关企业的董事、监事、高级管理人员和个人信息保护负责人^①。

个人信息处理者违反本法规定处理个人信息，侵害众多个人的权益的，人民检察院、法律规定的消费者组织和由国家网信部门确定的组织可以依法向人民法院提起诉讼^②。中国《个人信息保护法》将危害公共利益的企业列入限制或者禁止个人信息提供清单，予以公告，并采取限制或者禁止向其提供个人信息等措施^③。

在韩国的《个人信息保护法》中，对于违反本法规定，可能会被处以罚款、过怠料及过徵金等多种经济制裁。过怠料一般可以说是违反行为进行行政制裁，维持秩序的处罚。它是为了维护行政秩序而采取的一种纪律处分，不考虑经济利益。因为与罚款不同，不具有刑罚性质，所以不会留下犯罪记录。过徵金是对违反行政法义务的行为进行制裁的一种金钱惩罚。如果在违反行政义务的情况下取得了不公平的收益，将剥夺不公平利润，并收集其利润，以防止同样的事情再次发生。过徵金与过怠料没有太大区别，但不同的是，其目的是为了收回不公平的利润。

韩国的《个人信息保护法》规定，个人信息处理者处理的身份证号码丢失、被盗、泄露、伪造、变造或损毁时，可以处以五亿韩元以下的过徵金^④。此外，对违法行为进行分类，并处以不超过相关销售额 3%的过徵金、不超过十年、五年、三年和两年的监禁，或处以最高一亿韩元、五千万韩元、三千万韩元和两千万韩元的罚款^⑤。此外，如果一个人犯了属于第 70 至 73 条的任何罪行，可以没收与相关违法行为有关的金钱或其他利润，如果不能没收，可以收取其价值。在这种情况下，没收或追缴可以附加到其他处罚中。此外，根据第 75 条规定，对属于特定行为的人处以不超过五千万韩元的过怠料^⑥。在适用第 75 条关于过怠料的规定时，对依照第 34 条之 2 征收罚款的行为，不得征收过怠料^⑦。

违反 GDPR 被处以最高两千万欧元的行政罚款，如果是企业，则最高为上一财政年度全球年总营业额的 4%，以较高者为准^⑧。罚款可能因个别案件而有所不同。在决定是否处以行政罚款和决定每个案件的行政罚款数额时，应适当考虑到以下因素^⑨。

表 2.4 决定每个案件的行政罚款数额的因素

(a) 侵权行为的性质、严重程度和持续时间，同时考虑到有关处理的性质、范围或目的，以及受影响的数据主体的数量和他们遭受的损害程度。
(b) 侵权行为的故意或过失性质。
(c) 控制者或处理者为减轻数据主体的损失而采取的任何行动。
(d) 考虑到控制者或处理者根据第 25 条和 32 条实施的技术和组织措施，控制者或处理者的责任程度。

① 《中国个人信息保护法》 第七章 第六十六条

② 《中国个人信息保护法》 第七章 第七十条

③ 《中国个人信息保护法》 第三章 第四十二条

④ 《韩国个人信息保护法》 第三十四条 第二项

⑤ 《韩国个人信息保护法》 第七十条-第七十三条

⑥ 《韩国个人信息保护法》 第七十五条

⑦ 《韩国个人信息保护法》 第七十六条

⑧ 《GDPR》 第八十三条 第五款

⑨ 《GDPR》 第八十三条 第二款

-
- (e) 控制者或处理者以前的任何相关违规行为。
 - (f) 与监管机构的合作程度，以纠正侵权行为和减轻侵权行为可能产生的不利影响。
 - (g) 受侵权行为影响的个人数据的类别。
 - (h) 监管机构了解侵权行为的方式，特别是控制者或处理者是否通知了侵权行为，如果通知了，程度如何。
 - i) 如果第 58(2) 条中提到的措施曾被命令针对有关控制者或处理者的同一主题事项，那么遵守这些措施的情况。
 - (j) 遵守第 40 条规定的经批准的行为守则或第 42 条规定的经批准的认证机制。
 - (k) 适用于案件情况的任何其他加重或减轻因素，如因侵权行为直接或间接获得的经济利益或避免的损失。

表 2.5 各个地区的政策比较表格

维度	中国 个人信息保护法	韩国 个人信息保护法	欧盟 《通用数据保护条例》(GDPR)
个人信息的定义	以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息	通过姓名、身份证号码、视频等可以识别的信息	指与已识别或可识别的自然人(“信息主体”)有关的任何信息。
个人信息处理者	在个人信息处理活动中自主决定处理目的、处理方式的组织、个人。	以业务为目的运用个人信息文件,通过自己或他人处理个人信息的公共机关、法人、团体及个人等。	控制员:指单独或与他人共同决定处理个人数据的目的和方式的自然人或法人、公共当局、机构或其他团体。 处理员:指代表控制者处理个人数据的自然人或法人、公共当局、机构或其他机构。
法规的适用范围	国内的个人信息处理活动,在国外处理国内收集的个人信息的情况	只以个人信息处理者规定为适用范围	处理和利用欧盟区域内信息主体的个人信息时
个人信息处理原则	<ol style="list-style-type: none"> 1. 遵循合法、正当、必要和诚信原则 2. 应当具有明确、合理的目的 3. 与处理目的直接相关 4. 处理个人信息应当遵循公开、透明原则,公开个人信息处理规则,明示处理的目的、方式和范围 5. 应当保证个人信息的质量、准确性 6. 对其个人信息处理活动负责,并采取必要措施保障所处理的个人信息的安全 	<ol style="list-style-type: none"> 1. 明确个人信息的处理目的,并在其目的所需的范围内,合法、正当地收集、处理、利用最低限度的个人信息 2. 确保个人信息的准确性、完整性及最新性。 3. 个人信息必须得到安全管理 4. 信息主体的权利必须得到保障。 5. 应尽量减少对信息主体隐私的侵犯 6. 必须努力获得信息主体的信任 	<ol style="list-style-type: none"> 1. 以合法、公平和透明的方式进行处理 2. 为特定、明确和合法的目的而收集,并且不以与这些目的不相符的方式进一步处理 3. 强调数据最小化、准确性、完整性、保密性和责任感
个人信息处理的同意与撤销	同意 <ol style="list-style-type: none"> 1. 处理个人信息时必须得到相关个人的同意 2. 有一些允许在未经个人同意的情况下处理个人信息的例外情况 3. 互联网经营者必须按照个人信息处理原则,并征得提供者的同意。 	<ol style="list-style-type: none"> 1. 得到信息主体的同意的情况下,可以收集个人信息,并在其收集目的范围内使用 2. 有一些允许在未经个人同意的情况下处理个人信息的例外情况 	<ol style="list-style-type: none"> 1. 只有得到信息主体的同意才能处理个人信息 2. 为每个目的分别获得同意

撤销	<ol style="list-style-type: none"> 1. 确保个人信息提供者的同意撤销权 2. 撤回个人同意不会对撤回前基于个人同意的个人信息处理活动的效力产生影响 	<ol style="list-style-type: none"> 1. 没有表示撤回对处理个人信息的同意，但表示要求停止处理个人信息 2. 有可以拒绝信息主体停止处理要求的例外条款 	<ol style="list-style-type: none"> 1. 可以随时撤销对提供个人信息的同意。 2. 有权删除发布在互联网上的个人信息。 3. 当相关信息需要用于公共利益、科学或历史研究目的或统计目的时，会有例外。
与新技术应用相关的个人信息保护	<ol style="list-style-type: none"> 1. 保证决策的透明度 2. 不得实行不合理的差别待遇 3. 通过自动化决策方式向个人进行信息推送、商业营销，应当同时提供不针对其个人特征的选项，或者向个人提供便捷的拒绝方式 4. 企业想要将个人信息用于广告性推送营销等时，履行事前个人信息保护影响评估 	与信息处理原则一样、这取决于信息处理者如何理解法律。	<ol style="list-style-type: none"> 1. 要注意向信息主体提供个人信息处理方法和相关信息 2. 当个人信息被用于大数据时，必须为这一变化的特定目的确定其他法律依据 3. 需要签订个人信息处理合同和其他保护措施
个人信息的境外转移	<ol style="list-style-type: none"> 1. 收集的中国居民个人信息保护不能低于中国境内 2. 一定要取得个人的单独同意 3. 应当通过国家网信部门组织的安全评估 4. 一般个人信息处理者时： <ol style="list-style-type: none"> A. 境外接受方应采取必要措施，符合《个人信息保护法》规定的个人信息保护标准 B. 向个人告知个人信息海外转移相关内容，取得个人的同意 C. 事先进行个人信息保护影响评估 D. 与境外接收方订立合同，约定双方的权利和义务 	<ol style="list-style-type: none"> 1. 通知以下事项并征求同意，可以进行海外转移： <ol style="list-style-type: none"> A. 转移个人信息的项目 B. 转移个人信息的国家、日期和方法 C. 个人信息接收方的姓名 D. 个人信息的使用目及拥有、利用期间 	<ol style="list-style-type: none"> 1. 只有在第三国具有与欧盟相应水平的信息保护体系并保证安全水平的情况下才能进行 2. 位于未经欧盟批准的第三国的公司只有在证明信息通过欧盟行为准则和遵守公司规则得到安全保护的情况下，才允许境外转移
个人信息主体的	<ol style="list-style-type: none"> 1. 个人对其个人信息的处理享有知 	<ol style="list-style-type: none"> 1. 有权获得有关处理个人信息的信息 	<ol style="list-style-type: none"> 1. 保障接收信息的权利、阅览权、更正权、信息删除权、

权利保障	情权、决定权，有权限制或者拒绝他人对其个人信息进行处理 2. 具有查阅权、复制权、转移权 3. 有权请求更正、补充、删除	2. 有权选择和决定是否同意个人信息处理和同意范围等 3. 有权确认个人信息处理与否和阅览个人信息 4. 有权要求停止、纠正、删除和销毁个人信息 5. 有权按照迅速、公正的程序得到救济个人信息处理造成的损失	个人信息移动权、反对权、处理限制权、自动化决定及分析相关权利等相关信息提供主体的权利 2. 新引进了删除权、个人信息移动权、自动化决定等相关权利
违反法规时的法律责任	1. 罚款 2. 吊销相关业务许可或吊销营业执照 3. 禁止担任高级管理人员或负责人	1. 罚款 2. 监禁	罚款

第三章 公众舆论分析

3.1 公众舆论分析比较

本章涉及公众舆论分析的结果。公共舆论分析结果有公共舆论关注度变化分析、基于 TF-IDF 的主题词提取分析、LDA 主题模型分析。在分析舆论关注度的变化时，将各国各地区媒体的各时期文件数量用图表进行了视觉化。通过分析不同时期发行量的差异来研究关注度的变化。在 TF-IDF 分析中，对各个媒体进行了 TF-IDF 分析，并将前 50 个关键词用词云和表格表示。在 LDA 主题模型分析中，对各国媒体进行分析，找出媒体数据的主要主题，并将每个主题的发行量以图表形式表示。

3.1.1 中华人民共和国公众舆论

公众舆论关注度变化

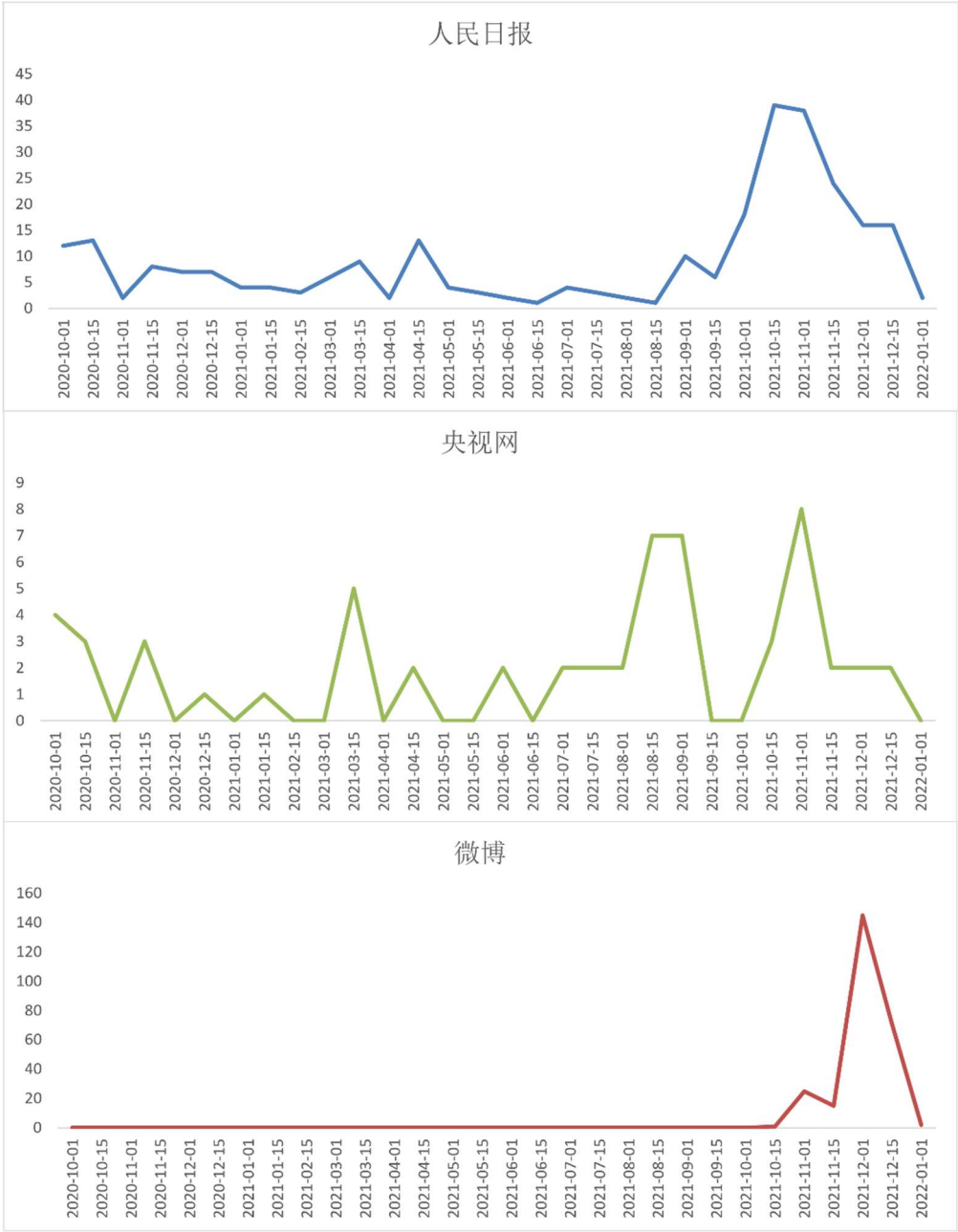


图 3.1 中国媒体文档数量变化图表

这张图显示了每个平台的文档数量随时间的变化。为了掌握共同点，我们可以看到主流新闻媒体，如央视、微博在 11 月达到高峰，微博在 12 月达到高峰。这是因为中国的《个人信息保护法》在 2021 年 11 月 1 日施行。微博的上升相对于其他媒体滞后一些是因为在中国的媒体环境内，官媒对于议程设置有一定的引领性作用。

可以发现三个平台的每日相关文档量级有差异，央视网的讨论少于微博和人民日报。而且可以看出，在 2021 年 11 月之前，央视网和人民日报一直有一定的文档数量。两个平台的不同之处在于，人民日报在 10 月急剧上升，在 11 月达到顶峰，而央视网在 8 月急剧上升并下降，在 11 月再次达到顶峰。人民日报的相关讨论在 11 月实施的政策之前已有发表。中央电视台在 8 月和 11 月达到高峰的原因是，2021 年 8 月十三届全国人大常委会第三十次会议表决通过《中华人民共和国个人信息保护法》，而 2021 年 11 月 1 日《个人信息保护法》施行。

基于 TF-IDF 的主题词提取

三个平台的共同点是三个平台的前 10 个关键词都包含“个人信息”、“保护”、“信息”、“App”。根据人民日报文档的分析结果，TF-IDF 的值最高的单词是“个人信息”。其次，有“信息”、“保护”、“数据”、“App”、“用户”、“网络安全”、“收集”、“人脸识别”、“消费者”等。这说明该媒体的文档注重于个人信息保护法的信息安全。根据央视网文档的分析结果，TF-IDF 的值最高的单词是“个人信息”。其次，有“保护”、“信息”、“同意”、“草案”、“个人信息保护法”、“App”、“法律”、“快递”和“收集”等。这说明该媒体的文档注重于个人信息保护法中“保护信息”的目的。根据微博文档的分析结果，TF-IDF 的值最高的单词是“个人信息保护法”。其次，有“个人信息”、“App”、“保护”、“信息”、“下架”、“数据”、“用户”、“正式”、“法律”等。这说明该媒体的文档注重于个人信息保护法施行后的社会影响。与该媒体文档关键词的词云和前 50 个词的 TF-IDF 值表显示下图 3.2、图 3.3、图 3.4 和附录 A.1、附录 A.2、附录 A.3。



图 3.2 人民日报词云



图 3.3 央视网词云



图 3.4 微博词云

LDA 主题模型

在选择 LDA 主题数量时,研究先从 2 个到 30 个,按两个单位的主题个数进行建模,看 coherence-value 值。对结果进行不断调试,而最后得到了最好的 LDA 主题数量。当分类主题数量为 10 时,Coherence score 值最好,而分类的边界清晰。如下表所示,可以知道 10 个主题的概率指数前 20 位的主题词。

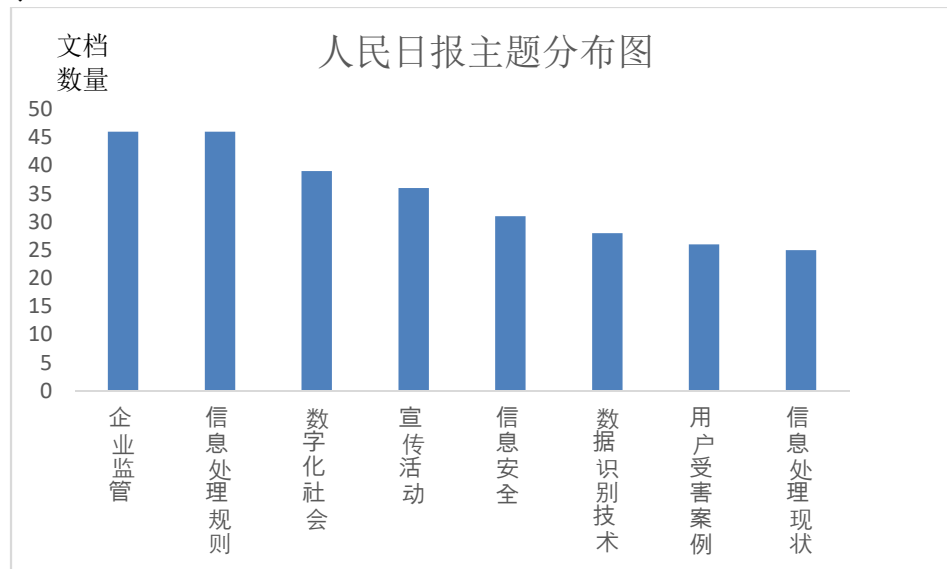


图 3.5 人民日报主题分布图

如上图,人民日报文档中导出的 10 个主题中占据最多比重的主题是“Topic 5: 企业监管 (16.61%)”和“Topic 8: 信息处理规则 (16.61%)”。Topic5 的主要内容是对数据企业、平台企业等的监管活动。该主题的一篇文章说国家计算机病毒应急处理中心(NCERT)最近通过互联网监测发现,17 款手机应用程序存在隐私违规行为,违反了《网络安全法》和《个人信息保护法》的相关规定,涉嫌收集超出范围的个人隐私信息。该机构以避免个人隐私信息泄露提醒几个点:第一,要谨慎下载和使用违规的移动应用程序;第二,注意仔细阅读应用的用户协议和隐私政策说明;第三,不得随意打开和同意不必要的隐私权限;第四,不得随意输入个人隐私信息;第五,并定期维护和清理相关数据⁷⁴。Topic8 的主要内容是信息处理时要遵守的规则。其中,一篇文章说社会生活各方面的个人信息处理和运行,如果按照传统的功能划分进行管理,很难有效防止个人信息权利受到侵害。该文章说草案强调了在处理个人敏感信息时保护个人信息的义务,包括严格限制处理个人敏感信息,以及在紧急情况下保护自然人的生命和健康,而且说明了草案的详细内容。其次,有“Topic 4: 数字化社会(14.08%)”、“Topic 7: 宣传活动(网络安全)(13.00%)”、“Topic 2: 信息安全(11.19%)”、“Topic 3: 数据识别技术 (10.11%)”、“Topic 6: 用户受害案例 (9.39%)”和“Topic 1: 信息处理现状 (9.03%)”。

⁷⁴ 张建新. 国家计算机病毒应急处理中心监测发现十七款违法移动应用 [J/OL].
<http://finance.people.com.cn/n1/2021/1128/c1004-32293738.html>, 2021-11-28/2022-05-01.

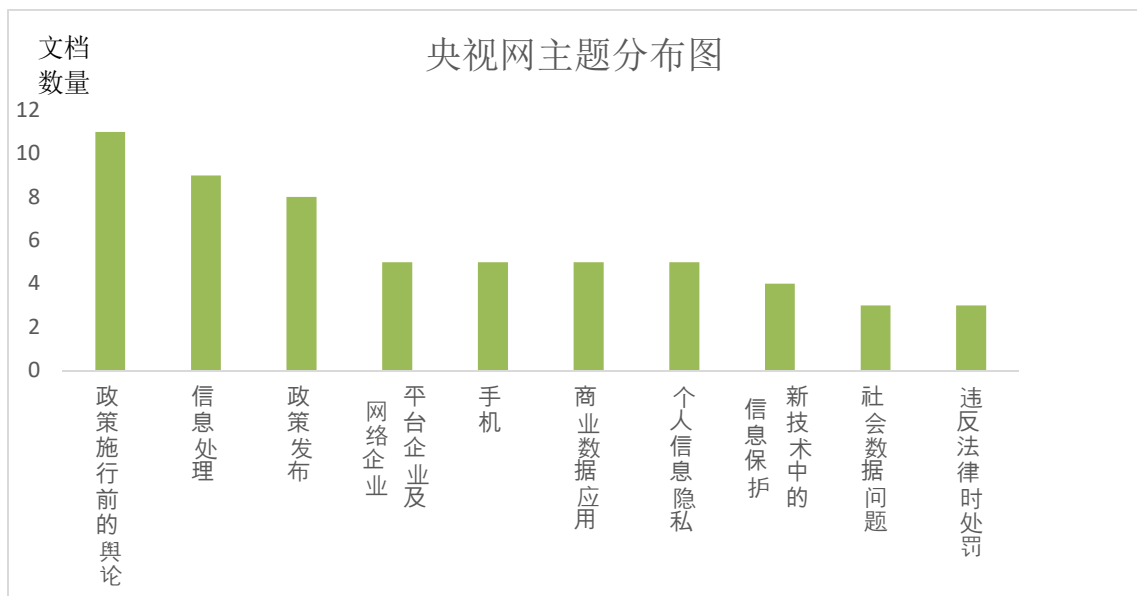


图 3.6 央视网主题分布图

如上图，央视网文档中导出的 10 个主题中占据最多比重的主题是“Topic 8：政策施行前的舆论”（18.97%）。央视网比其他平台文档不同的是，该法律施行前有一次达到文档数量高峰的时候。我认为 topic 8 说明这个现象。因为，topic8 的主要内容是说明《个人信息保护法》的草案或专家的意见。该主题文档之一说相关业内专家认为，中国的个人信息保护法律体系已经逐步建立，但仍难以适应信息技术的快速发展。因此，应在现有法律的基础上制定专门的法律，增强法律规范的系统性、针对性和可操作性。他们表示，草案将为建立更加完善的个人信息保护体系提供更有力的法律保障。由此可见，央视网作为重要的视频主流媒体平台，在法规推出前后具有重要的带领公众和专家共同讨论立法过程的作用。其次，有“Topic 1：信息处理（15.52%）”、“Topic 6：政策内容（13.79%）”、“Topic 3：平台企业及网络企业（8.62%）”、“Topic 5：手机（8.62%）”、“Topic 9：商业数据应用（8.62%）”、“Topic 10：个人信息隐私（8.62%）”、“Topic 2：新技术中的信息保护（6.90%）”、“Topic 4：社会数据问题（5.17%）”、“Topic 7：违反法律时处罚（5.17%）”。

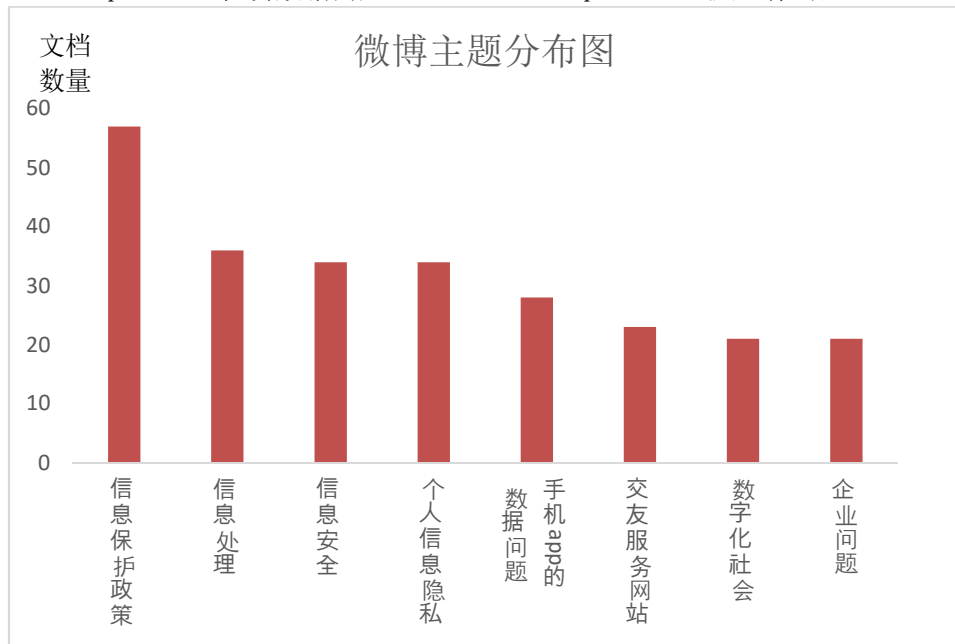


图 3.7 微博主题分布图

如上图，微博文档中导出的 10 个主题中占据最多比重的主题是“Topic 3：信息保护政策（22.44%）”。该主题的主要内容是浅析或讨论《个人信息保护法》的内容。其中一个微博通报了浅析个保法的实施与失联修复的合规性的论文发布。其次，有“Topic 2：信息处理（14.17%）”、“Topic

7: 信息安全 (13.39%)”、“Topic 8: 个人信息隐私 (13.39%)”、“Topic 5: 手机 app 的数据问题 (11.02%)”、“Topic 4: 交友服务网站 (9.06%)”、“Topic 1: 数字化社会 (8.27%)”和“Topic 6: 企业问题 (8.27%)”。值得注意的是, 社交媒体平台特变关注了不同场景(比如手机 app、交友网站)的个人信息保护, 这一点与主流媒体舆论有所差异。

人民日报的主要主题是企业监管和信息处理规则。央视网的主要主题是政策施行前的舆论, 微薄的主要主题是信息保护政策。主流媒体和社会化媒体都关注信息处理, 两个媒体的不同点是主流媒体采取的是宏观的方式, 如企业相关内容。社会化媒体有很多更贴近生活中的普通大众的内容, 如应用和社区。

表 3.1 人民日报 LDA 主题结果列表

topic 1 信息处理现状		topic 2 信息安全		topic 3 数据识别技术		topic 4 数字化社会	
数据	0.033	个人信息	0.028	信息	0.058	数据	0.056
国家	0.018	信息	0.027	人脸识别	0.037	发展	0.027
提供	0.017	泄露	0.021	人脸	0.027	保护	0.023
部门	0.016	诈骗	0.018	技术	0.025	中国	0.014
信息	0.014	网络	0.018	采集	0.020	数字	0.013
服务	0.014	公民	0.016	信息安全	0.009	技术	0.011
保护	0.014	保护	0.015	生物_识别	0.008	隐私	0.010
数据安全	0.013	电信	0.010	消费者	0.007	数据安全	0.010
征信	0.011	企业	0.010	风险	0.007	我国	0.010
互联网	0.011	相关	0.009	收集	0.007	数字_经济	0.010
个人信息	0.010	信息安全	0.009	相关	0.007	治理	0.009
相关	0.010	案件	0.008	识别	0.007	社会	0.009
网信	0.010	非法	0.007	原则	0.007	企业	0.009
机构	0.010	力度	0.007	支付	0.006	领域	0.008
数据处理	0.010	快递	0.007	泄露	0.006	体系	0.008
关键	0.009	打击	0.007	记者	0.005	平台	0.007
评估	0.009	犯罪	0.006	照片	0.005	监管	0.006
平台	0.008	加大	0.005	企业	0.005	国家	0.006
情况	0.008	导致	0.005	小区	0.005	风险	0.006
业务	0.008	建议	0.005	设备	0.005	标准	0.006
topic 5 企业监管		topic 6 用户受害案例		topic 7 宣传活动（网络安全）		topic 8 信息处理规则	
app	0.062	消费者	0.033	网络安全	0.052	个人信息	0.104

个人信息	0.046	平台	0.026	活动	0.024	保护	0.048
用户	0.041	用户	0.024	工作	0.018	法律	0.024
收集	0.037	信息	0.013	网络	0.017	同意	0.023
企业	0.016	商家	0.013	宣传	0.016	个人信息保护法	0.022
信息	0.014	个人信息	0.011	国家	0.013	草案	0.018
互联网	0.013	快递	0.010	法治	0.012	信息	0.015
监管	0.013	短信	0.010	群众	0.012	信息处理	0.014
服务	0.010	提供	0.008	组织	0.010	规则	0.011
保护	0.010	经营者	0.008	个人信息保护法	0.008	权益	0.011
隐私	0.010	面单	0.008	建设	0.007	敏感	0.010
权限	0.010	相关	0.007	防范	0.007	民法典	0.010
相关	0.009	隐私	0.007	意识	0.007	立法	0.010
违规	0.009	消费	0.007	知识	0.007	提供	0.009
权益	0.008	授权	0.007	宪法	0.007	收集	0.009
违法_违规	0.007	发现	0.007	维护	0.006	方式	0.008
信息安全	0.007	账号	0.007	服务	0.006	信息处理者	0.008
情况	0.007	方式	0.007	宣传周	0.006	利用	0.008
工信部	0.006	公司	0.007	全国	0.006	权利	0.007
发布	0.006	电商	0.006	主题	0.006	侵害	0.007

表 3.2 央视网 LDA 主题结果列表

topic 1 信息处理		topic 2 新技术中的信息保护		topic 3 平台企业及网络企业		topic 4 社会数据问题		topic 5 手机	
个人信息	0.116	信息	0.061	保护	0.038	个人信息	0.064	app	0.072
个人信息 保护法	0.062	人脸	0.054	发展	0.038	保护	0.045	个人信息	0.064
保护	0.042	人脸识别	0.048	互联网	0.035	法律	0.033	用户	0.053
权益	0.029	同意	0.036	网络	0.032	个人信息保护 法	0.026	收集	0.052
敏感	0.026	未成年人	0.028	监管	0.024	同意	0.022	信息	0.043
原则	0.022	技术	0.018	行业	0.020	信息	0.019	过度	0.019
信息处理者	0.016	信息处理者	0.016	中国	0.018	条款	0.015	手机	0.017
权利	0.015	方式	0.014	规范	0.018	相关	0.014	授权	0.013
完善	0.015	强制	0.013	数字	0.014	社会	0.012	建议	0.013
侵害	0.014	采集	0.013	方式	0.013	数据_杀熟	0.012	情况	0.012
制度	0.014	征得	0.011	发布	0.012	规则	0.012	服务	0.011
影响	0.013	自然人	0.011	企业	0.012	互联网	0.010	受访者	0.010
法律	0.013	保护	0.010	领域	0.012	规范	0.010	应用程序	0.010
实施	0.012	依法	0.010	提升	0.011	删除	0.009	我国	0.009
信息处理	0.012	服务	0.009	治理	0.010	用户	0.009	安装	0.009
规则	0.010	对此	0.009	推动	0.010	工作	0.009	协议	0.008
信息处理 _活动	0.010	权益	0.008	建设	0.010	告知	0.008	权限	0.008
目的	0.010	认定	0.008	平台	0.010	收集	0.007	下载	0.007
泄露	0.009	小区	0.008	相关	0.010	产品	0.007	生活	0.006
最小	0.009	单独	0.008	权益	0.009	app	0.007	涉及	0.006
topic 6		topic 7		topic 8		topic 9		topic 10	

政策发布		违反法律时处罚		政策施行前的舆论		商业数据应用		个人信息隐私	
消费者	0.055	信息	0.057	个人信息	0.081	用户	0.028	快递	0.097
部门	0.017	信息安全	0.026	同意	0.038	平台	0.025	信息	0.039
管理	0.016	收集	0.026	提供	0.037	数据	0.023	面单	0.031
组织	0.015	相关	0.024	草案	0.036	app	0.023	隐私	0.030
隐私	0.015	企业	0.022	保护	0.018	专家	0.020	泄露	0.024
社会	0.013	数据	0.022	法律	0.014	技术	0.014	平台	0.019
违法	0.012	采集	0.016	拒绝	0.013	发现	0.014	公民	0.018
服务	0.012	记者	0.015	相关	0.013	法律	0.011	公司	0.013
维护	0.012	处罚	0.014	国家	0.011	提供	0.010	个人信息	0.013
提出	0.011	进一步	0.013	目的	0.011	审稿	0.010	记者	0.012
实施	0.010	保障	0.013	告知	0.011	提出	0.010	系统	0.010
日起_施行	0.010	商家	0.013	个人信息 保护法	0.011	过程	0.010	账号	0.009
消费	0.009	违法	0.012	方式	0.010	收到	0.009	企业	0.009
技术	0.009	提高	0.012	核心	0.010	介绍	0.009	姓名	0.009
计算	0.009	时代	0.011	作出	0.010	画像	0.009	诈骗	0.008
条例	0.009	非法	0.011	包括	0.009	精准	0.009	圆通	0.008
央视网_消息	0.008	力度	0.011	服务	0.009	意见	0.009	客户	0.007
合规	0.008	立法	0.010	部门	0.009	价格	0.008	案件	0.007
未经	0.008	设备	0.009	罚款	0.008	数据安全	0.008	员工	0.007
涉及	0.008	显示	0.009	公开	0.008	广告	0.008	个人隐私	0.007

表 3.3 微博 LDA 主题结果列表

topic 1 数字化社会		topic 2 信息处理		topic 3 信息保护政策		topic 4 交友服务网站	
数据	0.053	个人信息	0.178	个人信息保护法	0.141	信息	0.050
数字	0.032	保护	0.081	实施	0.037	记者	0.026
数据安全	0.024	法律	0.048	施行	0.034	新闻	0.023
网络	0.024	收集	0.024	微博_视频	0.026	公司	0.019
国家	0.023	提供	0.021	正式	0.023	门店	0.016
发展	0.020	同意	0.021	民法典	0.011	权限	0.015
法治	0.019	我国	0.020	责任	0.010	施行	0.014
中国	0.019	中华人民共和国	0.018	宪法	0.010	销售	0.014
治理	0.016	违法	0.018	微博	0.009	情况	0.014
社会	0.015	正式	0.017	生效	0.009	裸奔	0.014
信息	0.013	非法	0.016	评论	0.008	后台	0.013
领域	0.010	过度	0.015	力度	0.008	卧底_世纪	0.012
生活	0.010	网页_链接	0.014	图解	0.007	个人信息保护法	0.012
时代	0.010	手机	0.013	民法	0.007	隐私	0.011
法律法规	0.009	信息处理者	0.013	热点	0.006	会员	0.011
体系	0.009	日起	0.012	普法	0.006	世纪_佳缘	0.010
带来	0.009	部门	0.011	保护法	0.006	发现	0.010
管理	0.008	敏感	0.009	颁布	0.006	佳缘	0.010
经济	0.008	服务	0.009	面试	0.006	机构	0.010
数据安全法	0.008	买卖	0.007	未成年人	0.006	关闭	0.009
topic 5 手机 app 的数据问题		topic 6 企业问题		topic 7 信息安全		topic 8 个人信息隐私	
app	0.112	消费者	0.035	个人信息保护法	0.038	个人信息保护法	0.044

用户	0.041	公民	0.024	平台	0.032	版本	0.031
下架	0.040	违反	0.020	用户	0.030	隐私	0.026
工信部	0.034	侵犯	0.020	泄露	0.027	合规	0.021
相关	0.034	小鹏_汽车	0.014	隐私	0.025	注销	0.020
收集	0.027	公开	0.014	信息	0.023	商店	0.015
豆瓣	0.022	人脸	0.014	快递	0.022	涉及	0.014
整改	0.022	银行	0.013	企业	0.019	账号	0.013
组织	0.020	方式	0.013	互联网	0.018	互联网	0.012
通报	0.017	账户	0.012	信息安全	0.017	涉嫌	0.012
服务	0.017	分析	0.012	影响	0.017	希望	0.012
权益	0.016	事件	0.011	数据	0.017	视频	0.012
违规	0.014	客户	0.010	面单	0.014	权利	0.010
法律	0.013	处罚	0.010	电话	0.014	个人隐私	0.010
发布	0.013	上海	0.009	监管	0.011	软件	0.010
腾讯	0.013	开户	0.008	杀熟	0.011	个人信息	0.010
网络安全	0.013	小区	0.008	利用	0.011	真的	0.009
侵害	0.011	搜集	0.008	抖音	0.010	老王	0.009
技术	0.010	公众	0.008	落地	0.009	条件	0.009
予以	0.010	催收	0.008	工作	0.009	不合理	0.009

3.1.2 大韩民国公众舆论

公众舆论关注度变化

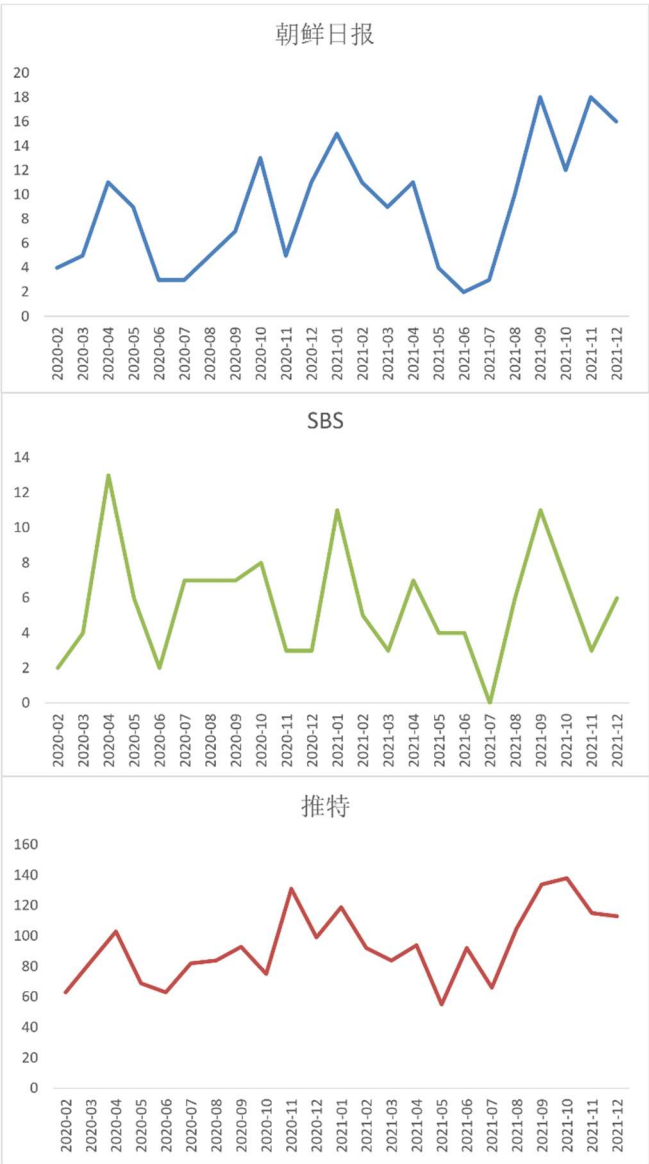


图 3.8 韩国媒体文档数量变化图表

我们可以看到的三个平台共同点是，每个平台的文档数量在 2020 年 4 月和 2021 年 9 月达到高峰，中间都有不规则的图形。因为“N 号房”^①事件在 2020 年 3 月开始成为热门话题，并在 4 月达到高峰。2021 年 9 月，作为总统候选人的“尹锡悦”和“李在明”过去的活动^②引起了争议，该行为是否违反了个人信息保护法成为热门话题。虽然有细微的差别，但这三个平台在 2021 年 1 月都处于上升期，因为被称为聊天机器人“Iruda”^③的人工智能对个人信息保护的侵犯成为了一个问题。韩国平台的一个主

^① “N 号房”事件在韩国成为了热门话题，是因为这个事件是作为制作和传播儿童和青少年及成人性剥削的案件。在案件的调查过程中，发现一名居民中心的工作人员将数据库里的个人信息交给了与案件有关的人，违反《个人信息保护法》的嫌疑也被适用。

^② 由于尹锡悦涉嫌过去唆使在野党议员告发的指责，在此过程中滥用职权、泄露官方机密、违反个人信息保护法等嫌疑成为争议焦点。向民主党候选人选拔的时候参加的选举团的市民发送了李在明的选举运动相关信息，相关市民认为竞选结束后应该废除的个人信息没有正常废除。因此是否违反《个人信息保护法》引起了争议。

^③ “Iruda”成为一个热门话题，因为它作为一个使用深度学习技术的人工智能聊天机器人，以类似人类的真实聊天方式走红。然而，由于性骚扰和少数人厌恶发言等的持续争议，该服务被暂停。在调查过程中，发现开发商的另一款应用程序“恋爱的科学”使用“KakaoTalk”用户的约 100 亿条数据进行学习，在此过程中是否遵守了《个人信息保护法》成为

LDA 主体模型

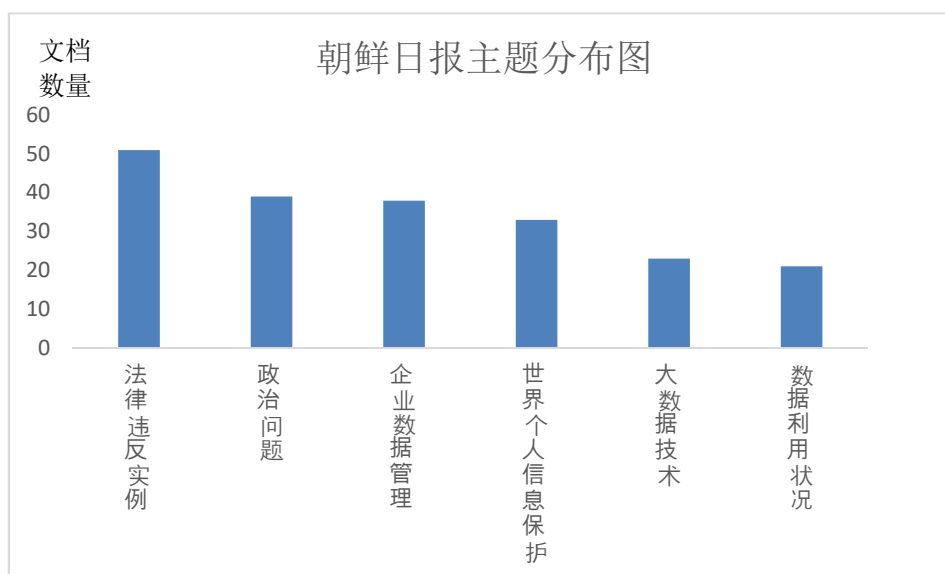


图 3.12 朝鲜日报主题分布图

如上图，朝鲜日报文档中导出的 10 个主题中占据最多比重的主题是“Topic 3：法律违反实例（24.88%）”。Topic3 的主要内容是违反法律的实例。该主题文档涉及违反《个人信息保护法》的案件，包括第“N 号房”的案件，以及是否对几个案件适用该法。比如，在一个谋杀案中，它涉及到对通过侦探机构带来受害者个人信息的犯罪者适用违反《个人信息保护法》的指控。其次，有“Topic 1：政治问题（19.02%）”、“Topic 5：企业数据管理（18.54%）”、“Topic 6：世界个人信息保护（16.10%）”、“Topic 4：大数据技术（11.22%）”、“Topic 2：数据利用状况（10.24%）”。法律违反实例和政治问题是比重最多的主题，这点与朝鲜日报文档数量高峰时期的原因密切相关。

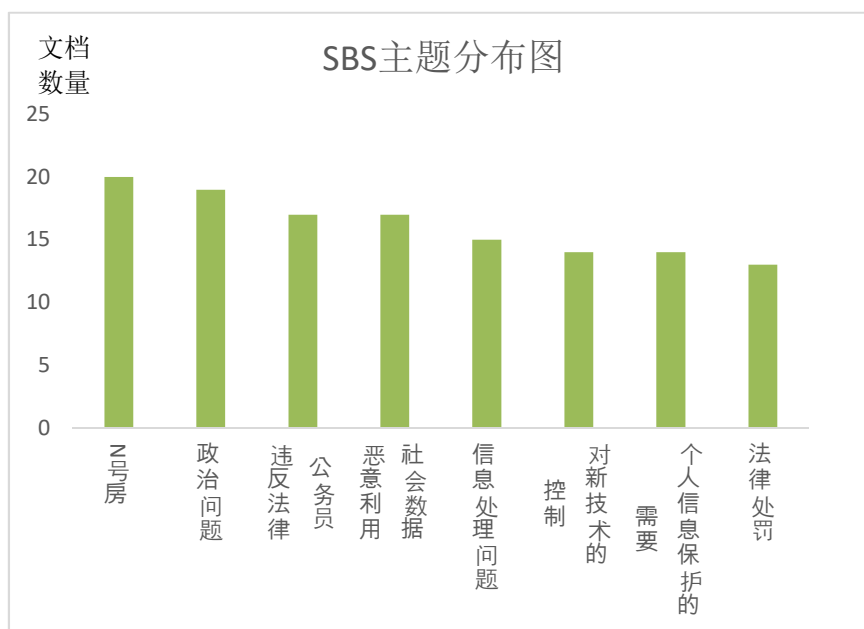


图 3.13 SBS 主题分布图

如上图，SBS 文档中导出的 8 个主题中占据最多比重的主题是“Topic 2：N 号房（15.50%）”。Topic 2 主要讨论韩国“N 号房”事件。该主题文档的大部分都是与事件有关的新闻，包括 N 号房事件的调查情况和损害情况等。其次，有“Topic 6：政治问题（14.73%）”、“Topic 1：公务员违反法律（13.18%）”、“Topic 7：社会数据恶意利用（13.18%）”、“Topic 5：信息处理问题（11.63%）”、“Topic 3：对新

技术的控制（10.85%）”、“Topic 4: 个人信息保护的需要（10.85%）”、“Topic 8: 法律处罚（10.08%）”。除了 Topic2 外，前两个内容是公务员与政治上的信息问题。这个主题与在 2021 年 9 月达到高峰的原因有密切相关。

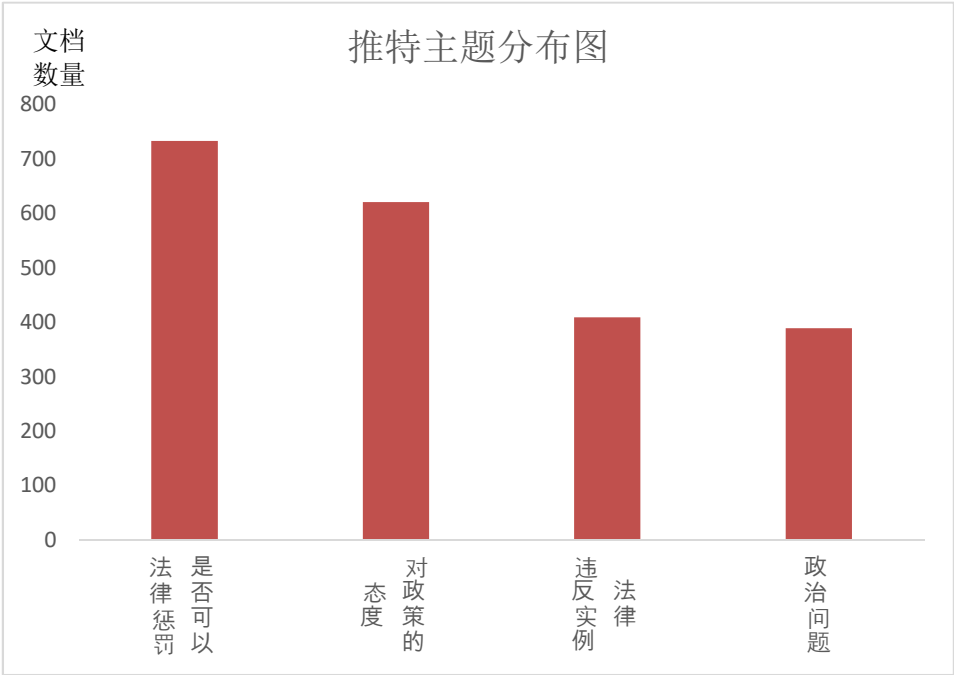


图 3.14 韩国推特主题分布图

如上图，推特文档中导出的 4 个主题中占据最多比重的主题是“Topic 4: 是否可以法律惩罚（34.06%）”。该主题的主要内容是对某一件事是否可以法律处罚。该主题的推特之一是问个人信息保护法，包括其他法律是否适用于 N 号房事件。另一个推特是一位律师发起了一个案件，并解释了是否适用该法律的一篇文章。其次，有“Topic 1: 对政策的态度(28.86%)”、“Topic 3: 政策违反实例(19.01%)”、“Topic 2: 社会个人信息保护问题（18.08%）”。

韩国主流媒体和社会化媒体都关注都关注违反法律的实例。两个媒体的不同点是主流媒体也关注数字化社会的现状以及个人信息保护方面的内容。社会化媒体更侧重于法律惩罚，如法律违反实例或是否可以惩罚等。

表 3.4 朝鲜日报 LDA 主题结果列表

topic 1			topic 2			topic 3		
政治问题			数据利用状况			法律违反实例		
의원	国会议员	0.026	데이터	数据	0.027	혐의	嫌疑	0.021
수사	侦查	0.025	기술	技术	0.017	경찰	警察	0.018
검사	剑士	0.024	기업	企业	0.015	장관	长官	0.013
검찰	检察	0.019	수집	收集	0.011	처벌	处罚	0.011
사건	事件	0.019	분석	分析	0.011	범죄	犯罪	0.011
의혹	疑惑	0.015	대표	代表	0.011	피해자	受害者	0.010
고발	举报	0.014	의료	医疗	0.010	내용	内容	0.009
혐의	嫌疑	0.014	사용	适用	0.009	주장	主张	0.008
조사	调查	0.014	개발	开发	0.009	행위	行为	0.008
총장	总账	0.009	보안	保安	0.009	위반	违反	0.008
감찰	监察	0.009	활용	或用	0.008	불법	非法	0.008
대검	大检察厅	0.008	기반	基底	0.008	직원	职员	0.007
청구	请求	0.008	시장	市场	0.007	공개	公开	0.007
판사	法官	0.008	사업	事业	0.007	개인정보보호법위반	违反 个人信息保护法	0.007
국회	国会议员	0.008	서비스	服务	0.006	교사	教师	0.007
개인정보보호법	违反	0.008	코로나	疫情	0.006	인정	承认	0.007
위반	个人信息保护法							
체포	逮捕	0.008	선수	选手	0.006	변호사	律师	0.007
법무부	法务部	0.007	핵심	核心	0.006	사건	事件	0.006
민주당	民主堂	0.007	정보	信息	0.006	선고	宣告	0.006
사주	唆使	0.006	스마트시티	智慧城市	0.006	이석준	李碩俊 (人名)	0.006
topic 4			topic 5			topic 6		
大数据技术			企业数据管理			世界个人信息保护		
데이터	数据	0.032	개인정보	个人信息	0.049	규제	控制	0.021

금융	金融	0.024	정보	信息	0.026	기업	企业	0.019
활용	活用	0.016	동의	同意	0.012	산업	产业	0.012
고객	顾客	0.011	조치	措施	0.009	정부	政府	0.012
신용평가	信用评价	0.011	유출	泄露	0.009	개인정보보호	个人信息保护	0.011
서비스	服务	0.010	페이스북	脸书 (Facebook)	0.009	한국	韩国	0.009
마이데이터	Mydata	0.009	이용자	用户	0.009	중국	中国	0.009
개인	个人	0.009	제공	提供	0.009	지원	支援	0.007
정보	信息	0.009	규정	规定	0.008	마련	安排	0.007
플랫폼	平台	0.009	개인정보보호법	个人信息保护法	0.008	추진	推动	0.007
기관	机关	0.008	라인	LINE (聊天软件)	0.008	분야	部门	0.007
시장	市场	0.008	이용	利用	0.008	디지털	数字	0.007
소비자	消费者	0.008	관리	管理	0.008	제도	制度	0.007
다양하다	多样	0.007	중국	中国	0.007	개인정보보호법	个人信息保护法	0.006
대출	出借	0.007	위반	违反	0.007	개정안	修订案	0.006
데이터사업	数据事业	0.007	공개	公开	0.007	미국	美国	0.006
사업자	事业家	0.007	부과	罚款	0.006	정책	政策	0.006
빅테크	大科技	0.007	정보주체	信息主体	0.006	과제	课题	0.006
중요하다	重要	0.007	운영	运营	0.006	요구	要求	0.006
필요하다	必要	0.006	개인정보유출	个人信息泄露	0.006	계획	计划	0.006

表 3.5 SBS LDA 主题结果列表

topic 1 公务员违反法律			topic 2 N 号房			topic 3 对新技术的控制			topic 4 个人信息保护的需要		
수사	侦查	0.046	혐의	嫌疑	0.058	조사	调查	0.018	정보	信息	0.039
의혹	疑惑	0.022	경찰	警察	0.035	사진	图片	0.017	탐정	侦探	0.019
검사	剑士	0.022	개인정보	个人信息	0.018	조치	措施	0.017	수집	收集	0.017
문건	文件	0.019	피해자	受害者	0.017	한국	韩国	0.013	서비스	服务	0.017
고발	举报	0.018	박사방	博士房	0.017	요구	要求	0.013	사기	欺诈	0.016
의원	国会议员	0.016	불법	非法	0.017	이용	利用	0.013	개인정보보호법	个人信息保护法	0.016
공무원	公务员	0.014	서울	首尔	0.016	정부	政府	0.012	주소	地址	0.012
개인정보보호법	违反	0.012	아동	儿童	0.014	투기	投机	0.011	개인	个人	0.012
위반	个人信息保护法										
작성	拟定	0.012	명의	名义	0.013	코로나	疫情	0.011	금융	金融	0.012
압수수색	扣押搜查	0.012	근무	工作	0.013	스캐터랩	SCATTER LAB	0.011	제도	制度	0.011
대상	对象	0.012	구속	拘留	0.012	관리	管理	0.011	필요하다	需要	0.010
진행	进行	0.012	조회	查询	0.012	공유	共享	0.011	활동	活动	0.008
기자	记者	0.011	유포	流传	0.012	포함	包含	0.010	법률	法律	0.008
전달	传输	0.011	주민센터	居民中心	0.011	거래	交易	0.010	지적	指出	0.008
총장	总长	0.010	공익	公益	0.011	개발	开发	0.010	활용	或用	0.008
소속	所属	0.010	적용	适用	0.010	국민	国民	0.009	삭제	删除	0.008
고발장	举报信	0.010	구속영장	逮捕令	0.010	데이터	数据	0.009	신용	信用	0.008
비밀	秘密	0.010	과약	把握	0.009	사례	案例	0.009	기업	企业	0.007
고위	高位	0.009	검찰	检察	0.009	본인	本人	0.008	고객	顾客	0.007
직원	职员	0.009	개인정보보호법	违反	0.009	이용자	用户	0.007	시행	施行	0.006
	위반		个人信息保护法								
topic 5 信息处理问题			topic 6 政治问题			topic 7 社会数据恶意利用			topic 8 法律处罚		

기자	记者	0.022	의원	国会议员	0.032	유출	泄露	0.040	검찰	检察	0.040
처벌	处罚	0.020	정보	信息	0.027	개인정보	个人信息	0.028	재판	审判	0.035
앵커	主播	0.018	공개	公开	0.025	경찰	警察	0.026	사건	事件	0.030
문자	短信	0.017	개인정보	个人信息	0.024	내용	内容	0.026	혐의	嫌疑	0.028
피해자	受害者	0.017	행위	行为	0.022	사건	事件	0.018	검사	剑士	0.027
피해	受害	0.016	목적	目的	0.018	이름	名字	0.017	감찰	监察	0.017
연락	联系	0.016	수집	收集	0.016	결과	结果	0.016	기소	起诉	0.015
개인정보보호법	违反	0.014	판단	判断	0.014	위반	违反	0.016	기록	记录	0.014
위반	个人信息保护法										
병원	医院	0.014	동의	同意	0.014	교사	教师	0.015	인사	人士	0.013
휴대전화	手机	0.013	정보처리	信息处理	0.014	관계자	有关人员	0.015	중인	众人	0.012
환자	病人	0.013	민주당	民主堂	0.013	학생	学生	0.015	장관	长官	0.011
범죄	犯罪	0.012	출석	出席	0.013	조사	调查	0.015	조사	调查	0.011
입장	立场	0.011	징계	惩戒	0.012	영상	视频	0.014	형사	刑事	0.011
전화번호	电话号码	0.011	범위	范围	0.012	학교	学校	0.013	진술	陈述	0.010
주장	主张	0.010	국회	国会议员	0.011	학부모	学生父母	0.012	수사	搜查	0.010
번호	号码	0.010	제기	提起	0.010	업체	企业	0.011	자료	资料	0.010
접근	接近	0.010	판사	法官	0.010	논란	论难	0.011	변호사	律师	0.010
사이트	网站	0.010	규정	规定	0.010	신고	申告	0.010	위반	违反	0.009
여성	女生	0.009	사유	理由	0.009	문제	问题	0.010	변호인	律师	0.009
식당	食堂	0.009	검찰	检察	0.009	요청	要求	0.009	증인	证人	0.009

表 3.6 韩国推特 LDA 主题结果列表

topic 1 对政策的态度			topic 2 政治问题			topic 3 法律违反实例			topic 4 是否可以法律惩罚		
개인정보 보호법	个人信息保护法	0.074	고발	举报	0.027	개인정보보호법	个人信息保护法	0.040	개인정보보호법 위반	违反 个人信息保护法	0.041
개인정보	个人信息	0.044	개인정보보호법 위반	违反 个人信息保护法	0.026	정보	信息	0.030	위반	违反	0.027
동의	同意	0.017	윤석열	尹锡悦	0.023	문건	文件	0.013	개인정보보호법	个人信息保护法	0.024
데이터	数据	0.013	검찰	检察	0.017	유출	泄露	0.012	수집	收集	0.017
개인	个人	0.011	혐의	嫌疑	0.015	교육	教育	0.012	처벌	处罚	0.011
정부	政府	0.009	수사	侦查	0.014	제공	提供	0.012	신고	申告	0.011
목적	目的	0.008	기자	记者	0.013	이하	以下	0.009	공개	公开	0.011
요구	要求	0.008	검사	剑士	0.011	한국	韩国	0.009	불법	非法	0.010
기업	企业	0.007	의혹	疑惑	0.011	이용	利用	0.009	변호사	律师	0.010
활용	活用	0.007	뉴스	新闻	0.011	사건	事件	0.008	내용	内容	0.009
법률	法律	0.007	의원	国会议员	0.011	산업	产业	0.007	행위	行为	0.008
국회	国会	0.007	국민	国民	0.010	사진	图片	0.007	업무	业务	0.008
정보주체	信息主体	0.006	위반혐의	违反嫌疑	0.010	장모	丈母	0.006	번호	号码	0.008
주의	注意	0.006	직권남용	滥用职权	0.010	관리	管理	0.006	처리	处理	0.007
개인정보보 호	个人信息保护	0.006	공무원	公务员	0.010	병원	议员	0.006	주민번호	身份证号码	0.007
영상	视频	0.006	경찰	警察	0.009	법정	法庭	0.006	허용	许可	0.007
제출	提交	0.006	범죄	犯罪	0.009	포함	包含	0.005	문자	短信	0.006
보호	保护	0.006	공개	公开	0.008	고객	顾客	0.005	사용	适用	0.006
위원회	委员会	0.006	총장	总账	0.008	단독	单独	0.005	전화번호	电话号码	0.006
설치	设置	0.006	사찰	查察	0.008	직원	职员	0.005	이름	名字	0.006

3.1.3 欧盟公众舆论

公众舆论关注度变化

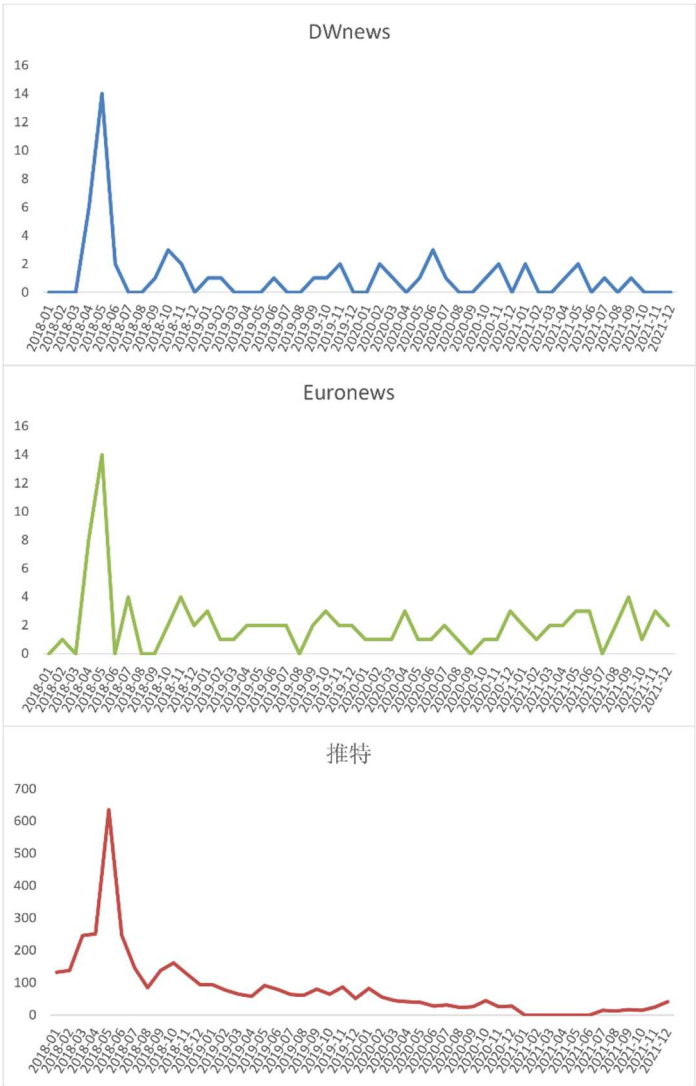


图 3.15 欧盟媒体文档数量变化图表

这张图显示了每个平台的文档数量随时间的变化。为了掌握共同点，我们可以看到所有媒体都在 2018 年 5 月达到最高峰。这是因为欧盟的 GDPR 在 2018 年 5 月实施。随着时间的流逝，所有媒体的文档数量均逐渐减少。

两个新闻媒体和推特的不同点是新闻媒体在达到高峰后迅速减少，其次是少量的文档数量。但推特呈现显著减少甚至个别时间节点没有相关讨论文档的趋势。这是因为推特是一种社会化媒体，肯定受到用户兴趣的影响，说明用户对 GDPR 政策本身的兴趣越来越减少。

基于 TF-IDF 的主题词提取

三个平台的共同点是三个平台的前 10 个关键词都包含“datum（数据）”，“protection（保护）”和“privacy（隐私）”。根据 DW news 文档的分析结果，TF-IDF 的值最高的单词是“datum”。其次，有“say”，“facebook”，“company”，“user”，“privacy”，“law”，“platform”，“protection”和“new”等。根据 Euronews 文档的分析结果，TF-IDF 的值最高的单词是“datum”。其次，有“say”，“user”，“company”，“facebook”，“privacy”，“protection”，“European”，“gdpr”和

“regulation”等。DW news 和 Euronews 的关键词说明该媒体的文档注重于在企业 and 用户方面看个人信息保护。两个媒体的关键词不同点是 Euronews 的关键词有更多的与 GDPR 直接相关的关键词。根据推特文档的分析结果，TF-IDF 的值最高的单词是“GDPR”。其次，有“datum”，“privacy”，“com”，“compliance”，“get”，“Data protection”，“protection”，“Compliant”和“new”等。这说明该媒体的文档注重于是否遵守 GDPR。与该媒体文档关键词的词云和前 50 个词的 TF-IDF 值表见图 3.16、图 3.17、图 3.18 和附录 A.7、附录 A.8、附录 A.9。



图 3.16 DW news 词云



图 3.17 Euronews 词云



图 3.18 欧盟推特词云

LDA 主体模型

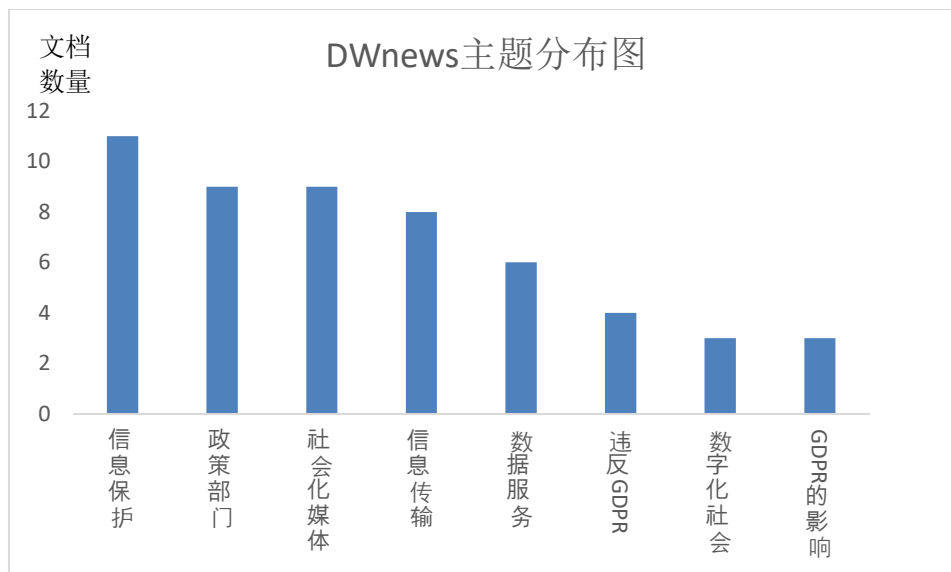


图 3.19 DW news 主题分布图

如上图, DWnews 文档中导出的 8 个主题中占据最多比重的主题是“Topic 7: 信息保护(20.75%)”。Topic7 的主要内容是数字化社会中的信息行为以及此过程中的信息保护。该主题的文档之一指出在我们的超级连接的数字世界中, 还有许多其他较小但同样有害的邪恶行为。文档里说, 有一些公司、机构和人被认为“以突出和持续的方式(in a prominent and sustained way)”滥用他人的隐私和数据。其次, 有“Topic 3: 政策部门 (16.98%)”、“Topic 4: 社交媒体 (16.98%)”、“Topic 8: 信息传输 (15.09%)”、“Topic 6: 数据服务 (11.32%)”、“Topic 1: 违反 GDPR (7.55%)”、“Topic 2: 数字化社会 (5.66%)”、“Topic 5: GDPR 的影响 (5.66%)”。

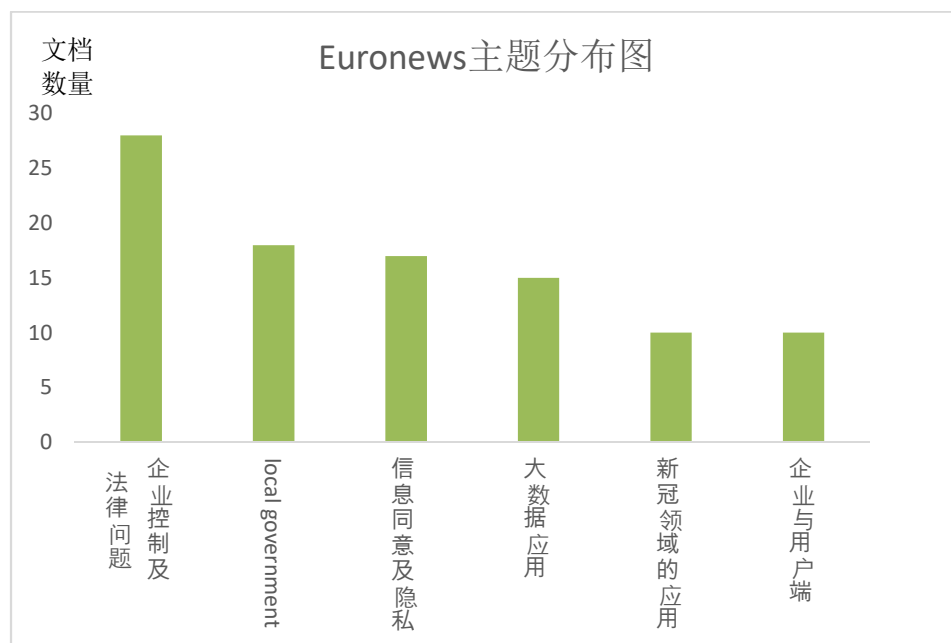


图 3.20 Euronews 主题分布图

如上图, Euronews 文档中导出的 6 个主题中占据最多比重的主题是“Topic 2: 企业控制及法律问题 (28.57%)”。Topic2 的主要内容是对企业的监管及企业行为的法律问题。该主题的主要文档之一的内容是根据欧盟最高法院裁定, Facebook、谷歌和 Twitter 等大型科技公司可以在数据隐私方面受到任何欧盟国家监管机构的质疑。欧盟法院的结论就是在某些条件下, 平台现在可以被欧

盟的任何数据保护机构起诉，而不仅仅是主导监管机构。还有，其他文档的内容是爱尔兰数据保护委员会（DPC）因 WhatsApp 违反欧盟关于用户隐私的规定而被罚款 2.25 亿欧元。其次，有“Topic 3: local government (18.37%)”、“Topic 6: 信息同意及隐私 (17.35%)”、“Topic 5: 大数据应用 (15.31%)”、“Topic 1: 新冠领域的应用 (10.20%)”、“Topic 4: 企业与用户端 (10.20%)”。

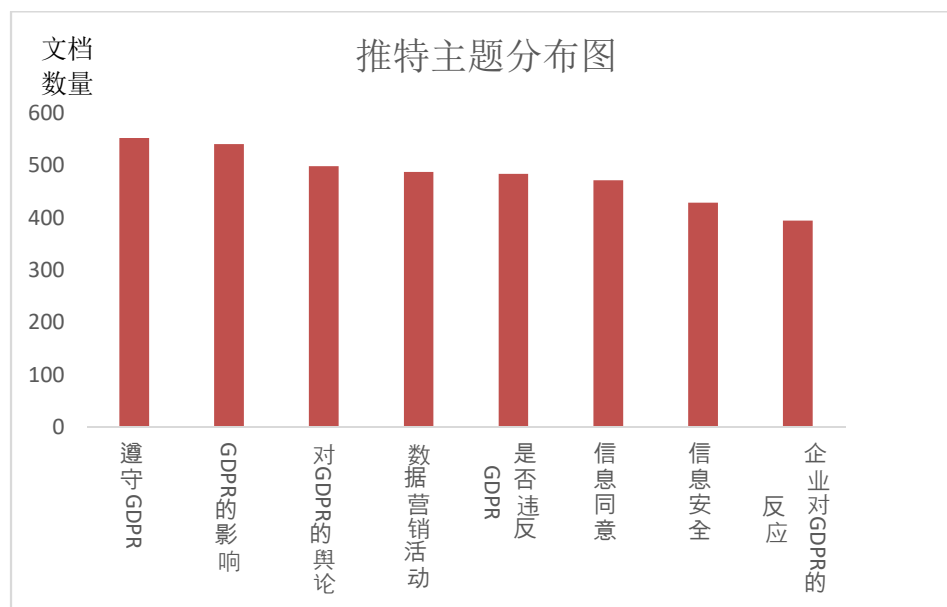


图 3.21 欧盟推特主题分布图

如上图，推特文档中导出的 8 个主题中占据最多比重的主题是“Topic 1: 适用 GDPR(14.32%)”。该主题的主要内容是适用 GDPR 后对数据处理活动或数据企业的监管等。该主题文档之一的内容是荷兰数据保护局 AP 对 Uber B.V. 和 Uber Techn 进行了 60 万欧元的罚款。其次，有“Topic 4: GDPR 的影响 (14.01%)”、“Topic 2: 对 GDPR 的舆论 (12.92%)”、“Topic 3: 数据营销活动 (12.64%)”、“Topic 6: 是否违反 GDPR (12.54%)”、“Topic 7: 信息同意 (12.22%)”、“Topic 5: 信息安全 (11.11%)”、“Topic 8: 企业对 GDPR 的反应 (10.23%)”。

表 3.7 DWnews LDA 主题结果列表

topic 1 违反 GDPR		topic 2 数字化社会		topic 3 政策部门		topic 4 社会化媒体	
company	0.061	call	0.023	datum	0.055	facebook	0.048
gdpr	0.028	digital	0.017	facebook	0.044	user	0.044
business	0.028	technology	0.015	personal	0.020	platform	0.030
whatsapp	0.018	system	0.015	user	0.016	social_media	0.021
fine	0.017	risk	0.015	firm	0.016	law	0.019
european_union	0.016	health	0.014	company	0.014	datum	0.017
big	0.015	school	0.013	election	0.013	european	0.016
app	0.014	datum	0.012	scandal	0.013	social	0.014
small	0.014	people	0.012	political	0.012	free	0.013
country	0.013	measure	0.011	claim	0.011	access	0.013
private	0.011	life	0.011	question	0.011	gdpr	0.012
german	0.010	potential	0.010	general	0.010	account	0.011
phone	0.010	time	0.010	include	0.010	court	0.011
follow	0.010	bias	0.010	leave	0.010	comply	0.011
share	0.010	important	0.010	message	0.009	network	0.011
customer	0.009	lack	0.009	breach	0.009	company	0.010
address	0.009	put	0.008	campaign	0.009	announce	0.010
set	0.009	public	0.008	executive	0.008	read	0.010
number	0.009	machine	0.008	member	0.008	consent	0.009
decision	0.009	spread	0.008	answer	0.008	instagram	0.009
topic 5 GDPR 的影响		topic 6 数据服务		topic 7 信息保护		topic 8 信息传输	
people	0.030	internet	0.028	datum	0.139	law	0.042

month	0.015	service	0.026	privacy	0.058	government	0.023
year	0.014	group	0.025	protection	0.040	personal	0.018
high	0.014	time	0.020	data	0.021	european	0.017
make	0.013	make	0.020	rule	0.018	state	0.017
app	0.011	online	0.019	company	0.015	force	0.016
world	0.011	year	0.016	information	0.015	global	0.015
leader	0.010	consumer	0.015	store	0.014	citizen	0.013
long	0.010	give	0.015	change	0.013	legal	0.013
start	0.010	digital	0.014	protect	0.013	information	0.013
economy	0.010	advertising	0.014	read	0.013	tax	0.012
strategy	0.009	user	0.013	work	0.012	report	0.012
week	0.009	website	0.013	protection_regulation	0.012	case	0.011
target	0.009	base	0.012	control	0.011	send	0.011
plan	0.009	find	0.012	add	0.011	process	0.011
remain	0.009	sign	0.011	policy	0.011	issue	0.011
room	0.009	market	0.011	post	0.011	transfer	0.010
hand	0.009	revenue	0.010	german	0.010	receive	0.010
freedom	0.008	major	0.009	make	0.010	legislation	0.010
end	0.008	ad	0.009	term	0.009	concern	0.010

表 3.8 Euronews LDA 主题结果列表

topic 1 新冠领域的应用		topic 2 企业控制及法律问题		topic 3 local government		topic 4 企业与用户端		topic 5 大数据应用		topic 6 信息同意及隐私	
technology	0.027	datum	0.115	government	0.032	company	0.086	datum	0.025	user	0.055
country	0.018	gdpr	0.046	datum	0.018	law	0.040	digital	0.023	privacy	0.028
system	0.014	protection	0.037	year	0.016	consumer	0.021	european	0.022	app	0.026
european	0.012	privacy	0.024	firm	0.014	tech	0.021	make	0.020	datum	0.021
risk	0.012	personal	0.021	user	0.013	internet	0.020	business	0.020	people	0.020
citizen	0.011	regulation	0.019	political	0.013	change	0.016	people	0.018	include	0.018
power	0.010	fine	0.018	report	0.013	online	0.014	regulation	0.015	information	0.017
government	0.009	rule	0.017	scandal	0.011	require	0.013	rule	0.015	share	0.013
trade	0.009	data	0.015	question	0.011	collect	0.013	work	0.013	website	0.012
time	0.009	information	0.015	platform	0.011	government	0.010	health	0.012	add	0.012
call	0.009	facebook	0.014	regulation	0.010	service	0.010	market	0.012	find	0.012
approach	0.008	authority	0.014	campaign	0.010	give	0.010	put	0.011	account	0.010
commission	0.007	general	0.013	issue	0.010	state	0.009	important	0.010	time	0.010
member_state	0.007	case	0.011	brexit	0.009	product	0.009	start	0.009	track	0.010
world	0.007	breach	0.011	address	0.009	euronew	0.009	good	0.008	accord	0.009
european_union	0.007	policy	0.010	giant	0.009	base	0.009	thing	0.008	concern	0.009
open	0.007	comply	0.010	election	0.009	group	0.008	year	0.008	month	0.009
surveillance	0.007	legal	0.009	target	0.009	ad	0.008	create	0.008	statement	0.009
base	0.007	regulator	0.009	end	0.008	ensure	0.008	set	0.008	phone	0.008
develop	0.007	decision	0.008	industry	0.008	content	0.007	global	0.008	sell	0.008

表 3.9 欧盟推特 LDA 主题结果列表

topic 1 适用 GDPR		topic 2 对 GDPR 的舆论		topic 3 数据营销活动		topic 4 GDPR 的影响	
gdpr	0.088	gdpr	0.129	gdpr	0.174	gdpr	0.149
compliant	0.034	protection	0.050	business	0.030	day	0.027
blockchain	0.022	regulation	0.043	good	0.029	great	0.022
service	0.021	law	0.037	time	0.029	policy	0.017
thing	0.020	data	0.029	email	0.026	update	0.015
event	0.015	issue	0.017	interesting	0.017	protect	0.015
internet	0.015	general	0.015	ready	0.014	start	0.014
week	0.014	global	0.012	marketing	0.010	apply	0.011
blog	0.013	citizen	0.010	show	0.010	base	0.011
solution	0.012	state	0.009	legal	0.009	gif	0.011
dsgvo	0.011	include	0.009	live	0.009	discuss	0.010
discussion	0.010	lead	0.008	approach	0.008	post	0.010
opportunity	0.010	month	0.008	support	0.008	mail	0.009
cloud	0.009	conference	0.008	point	0.008	effect	0.009
industry	0.009	art	0.008	hear	0.008	web	0.008
future	0.009	legislation	0.007	place	0.007	term	0.008
german	0.009	implement	0.007	real	0.007	session	0.008
join	0.009	high	0.007	affect	0.007	design	0.008
meet	0.008	due	0.007	answer	0.007	happen	0.007
store	0.008	enforcement	0.007	eprivacy	0.007	implementation	0.007
topic 5 信息安全		topic 6 是否违反 GDPR		topic 7 信息同意		topic 8 企业对 GDPR 的反应	
privacy	0.129	compliance	0.045	datum	0.199	gdpr	0.078

dataprotection	0.051	gdpr	0.033	personal	0.035	user	0.035
today	0.028	make	0.032	consent	0.030	facebook	0.027
news	0.026	fine	0.026	company	0.024	work	0.027
security	0.026	read	0.024	people	0.022	share	0.022
cybersecurity	0.021	article	0.023	information	0.022	big	0.018
breach	0.021	rule	0.020	give	0.021	change	0.014
company	0.020	year	0.020	customer	0.016	challenge	0.012
dataprivacy	0.018	website	0.019	case	0.015	lot	0.012
talk	0.018	learn	0.016	request	0.014	question	0.011
technology	0.016	european	0.015	twitter	0.013	problem	0.010
top	0.014	cookie	0.014	provide	0.013	risk	0.009
ai	0.014	consumer	0.013	process	0.013	public	0.009
tech	0.014	find	0.013	access	0.012	create	0.009
digital	0.014	comply	0.011	control	0.010	platform	0.009
report	0.012	check	0.011	legal	0.009	part	0.008
world	0.012	follow	0.011	open	0.009	content	0.007
stay	0.011	site	0.010	call	0.009	small	0.007
impact	0.010	ad	0.010	collect	0.009	private	0.007
iapp_europe	0.008	key	0.009	address	0.009	trust	0.007

3.2 比较各国公众舆论

从关注度变化方面看,中国和欧盟的主流媒体和社会化媒体表现出类似的趋势。然而,韩国的主流媒体和社会化媒体在趋势线上并没有表现出规律的趋势。此外,中国和欧盟的特点是在政策实施时达到顶峰。韩国在政策公布或实施时没有变化,往往对与《个人信息保护法》有关的案件表现出极大的兴趣。中国在其政策公布和实施之前的数字很少。在韩国,可以看出对政策相关的社会问题比对政策本身有更大的兴趣。在欧洲,自政策实施以来,兴趣逐渐减少。

从 TF-IDF 和 LDA 方面看各国各地区的关注议题的话,中国的主流媒体更注重政策内容,对信息安全和目的感兴趣,而社会化媒体则关注于政策对社会的影响。本人认为,这是因为对主流媒体来说,准确的信息传递很重要,所以提供客观的政策相关信息。由于社会化媒体可以写下个人的想法,所以相对主观的意见和风险感知相关讨论较多。从关注度变化可以看出,韩国媒体对个人信息泄露相关社会事件比对政策本身更感兴趣。韩国主流媒体注重的是已经违反法律的案例,社会化媒体则倾向于该政策是否适用于社会上发生的事件。而且,这里面都涉及的“国会议员”、“剑士”等单词说明关于政客讨论。这个现象的原因好像与上面阐述的原因相似。因为准确的信息传递很重要,主流媒体根据已经裁决或正式宣布的案件发表文章,但社会化媒体是一个空间可以在正式宣布或裁决之前可以分享案件是否适用于政策的意见。欧盟的主流媒体把政策作为一个整体来看,如从企业和用户的角度看政策本身,社会化媒体感兴趣的是是否遵守政策,而不是政策本身。而且,欧盟的媒体和公众都很关注大型跨国科技公司的个人信息保护法相关惩罚措施。

从关注点上,似乎三个国家地区的社会化媒体都非常关注违反个人信息保护法的案例以及该法规的社会影响,主流媒体均侧重政策细则。不同点是中国和欧盟表现出对政策内容和企业用户影响关系的兴趣,韩国对社会问题本身感兴趣。从媒体平台上,主流媒体有传递准确信息的特性,尤其在中国,主流媒体具有重要的引导社会讨论、邀请专家学者向公众解释个人信息保护法的作用。社会化媒体显示一个可以分享意见的平台特性,反映出较多与个人消费者息息相关的个人信息保护议题。

第四章 结论

4.1 研究结论

本研究对各国各地区的政策内容和新闻媒体、社交媒体的舆论内容进行了研究,并对三个国家和地区的政策内容和公众影响的异同进行了比较。本文为了考虑政策制定者和公众的视角,将定性的政策分析与定量的文本挖掘和分析相结合。该研究为互联网治理公共政策的跨国比较提供了实证研究支持,也为完善制度提供了实用的政策建议。综上,各个国家地区的政策都有自己的特点和差异。此外,在各个国家地区的主流媒体和社会化媒体上的公众舆论可以找到与这些特点有关的主题。

在政策内容上,各个政策都有各自的特点。在个人信息的定义上,各个政策对敏感信息的适用对象有一些差异。韩国和欧盟包括政治见解、特定集团加入与否和性相关信息,欧盟还包括种族或民族血统信息。在个人信息处理者方面,欧盟比中国和韩国更加细化地规定个人信息处理者,如控制员、处理员、代理人和数据保护员。在法规的适用范围上,中国和欧盟详细定义了范围,但韩国只将范围定为个人信息处理者。在个人信息处理原则和个人信息主题的权利上,三个政策的内容相似。在个人信息处理的同意和撤销,GDPR 比其他政策有更严格的规定。在新技术应用相关的内容上,三个政策都没有提到特定技术,中国的《个人信息保护法》的关注点是新技术可能带来的不利影响和需要保证透明度和进行影响评估。欧盟的 GDPR 认为新技术利用的数据就是个人信息,要求安全性和对目的变更的适当措施。韩国的《个人信息保护法》有模糊不清的内容,根据怎么样的解释内容会带来不同的结果。在个人信息的境外转移方面,中国和欧盟要求严格的标准和确认程序,但韩国只要求信息主体的同意。在法律责任方面,韩国和中国对违法行为的制裁程度更高。三个政策都有罚款,但中国还有吊销相关业务许可或营业执照,

禁止担任高级管理人员或负责人，韩国有监禁。

在公众舆论上，可以在三个方面得出结论。在文档数量变化方面，中国和欧盟的主流媒体和社会化媒体在讨论个人信息保护法时有时候会是同步的。特别是该国际地区的政策发布或施行的时候，达到顶峰。但在韩国相对非常不同步，韩国的文档数量与社会事件有密切关系。从关注点上，似乎三个国家地区的社会化媒体都非常关注违反个人信息保护法的案例以及该法规的社会影响，主流媒体均侧重政策细则。不同点是中国和欧盟表现出对政策内容和企业用户影响关系的兴趣，韩国对社会问题本身感兴趣。从媒体平台上，主流媒体有传递准确信息的特性，社会化媒体显示一个可以分享意见的平台特性。

在结合政策内容的结论和公众舆论的结论，可以得出四个结论。第一，在中国的主流媒体上有与新技术有关的主题，如数字识别技术、企业监管和新技术中的信息保护等。在欧盟的主流媒体上也有大数据应用、数据营销活动为主题，可以看出中国和欧盟集中于新技术与个人信息保护方面。第二，GDPR比中国和韩国的《个人信息保护法》更有严格的规定。欧盟的媒体上可以发现与信息同意有关的主题，如信息同意及隐私和信息同意等。可以知道欧盟关注信息主题的同意和撤销部分。第三，韩国比其他国家的每项政策都有自己的特点和差异。此外，在主要媒体或社会化媒体平台上可以找到与这些特点有关的话题。

本研究进行跨国跨地区的个人信息保护政策对比研究，而且把政策内容和公共舆论研究结合进行实证研究。在这些方面来看，本研究具有前沿研究没有的较高科学价值和政策价值。

4.2 研究不足与下一步研究方向

本研究通过对各个国家地区的不同政策内容的比较分析和利用各国地区的主流媒体和社会化媒体的自然语言处理分析，分析了各个政策的差异和各个媒体关注的主要主题和关键词，并将其可视化。本研究在进行跨国跨地区的对比研究和有实证研究把政策内容研究与公共舆论研究结合方面，有意义的。然而，本研究有以下局限性。

首先，根据用户在主流媒体和社会化媒体上使用的关键词进行自然语言处理的分析，是以研究者的主观判断进行的。其次，进行公共舆论分析的平台仅限于各地区的主流媒体和社会化媒体，不考虑各媒体对各地区的影响力。而且由于各政策的发布和施行时间不同，数据的时间范围也不尽相同，这就存在着概括性的局限。因此，在今后的研究中，判断有必要通过扩大时间范围和针对媒体进行补充研究。本人对未来研究的建议如下：第一，本人建议一项研究，对《个人信息保护法》以外的数据相关政策的内容进行比较和分析。在中国和韩国，除了《个人信息保护法》之外，与数据相关的政策也包括信息安全和信息保护的内容。因此，该分析将很容易理解国家对个人信息保护的立场。第二，本人建议增加另一项研究，通过社会网络分析来比较和分析公共舆论。因为通过 TF-IDF 或 LDA 提取的关键词的网络或相互关系的把握，可以进行更复杂精确的分析。

通过政策内容梳理，发现每个国家的个人信息保护有一些可以提升或者改进的地方。中国需要考虑将数据最小化，以减少对信息主体的损害。随着信息的海外转移数量的增加。中国可以参考 GDPR 的自动化决定及分析相关权利等相关信息提供主体的权利。因为该权利可以算法推荐治理的基础，比如上面提到的自动化决策和价格歧视的实例。韩国需要保护通过海外转移处理的韩国公民的个人信息，并需要建立一个可实际执行的个人信息海外转移政策，使韩国人的个人信息在国外不被滥用。因为数据在当今社会已经变得非常重要，欧盟需要考虑更严格的惩罚措施。

参考文献

- [1] 冯惠玲. 电子文件的双重鉴定——《拥有新记忆——电子文件管理研究》摘要之三[J]. 档案学通讯, 1998, 03:33-36.
- [2] 王众托. 企业信息化与管理变革——管理科学文库[M]. 中国人民大学出版社, 2001.
- [3] 王立荣. 考虑公平偏好的供应链决策与协调研究[D]. 哈尔滨工业大学, 2018.
- [4] Federico G . Feeding the World: An Economic History of Agriculture, 1800-2000[M]. 2010.
- [5] Roth A E . The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory[J]. Scholarly Articles, 1984, 92(6):991-1016.
- [6] 吴伟光. 大数据技术下个人数据信息私权保护论批判[J]. 政治与法律, 2016(7): 116-132.
- [7] 张平. 大数据时代无法拒绝个人信息被收集[EB / OL]. http://www.sohu.com/a/165144368_123753. [2018-07-15].
- [8] 占南. 国内外个人信息保护政策体系研究[J]. 图书情报知识, 2019(5): 120-129.
- [9] 王利明, 丁晓东. 论《个人信息保护法》的亮点, 特色与适用[J]. 法学家, 2021(6):17.
- [10] 杨智博. 韩国《个人信息保护法》的最新修正及其对我国之启示[J]. 华南理工大学学报:社会科学版, 2022, 24(1):10.
- [11] 刘云. 欧洲个人信息保护法的发展历程及其改革创新[J]. 暨南学报:哲学社会科学版, 2017, 39(2):13.
- [12] Hyun-seok Yoon. Trends in China's Personal Information Protection Legislation - Comparison with Korea's Personal Information Protection Law[J]. Journal of Law and Politics research, 2021, 21(1): 351-380.
- [13] Kim Sung Hyun, Lee Chang Moo. A Comparative Analysis of EU GDPR with Privacy Laws in South Korea[J]. Journal of Information and Security, 2018, 18(5):83-92.
- [14] 盖立伟. 基于 GDPR 的我国个人信息保护政策研究[D]. 黑龙江大学.
- [15] Yuan E J , Feng M , Da Nowski J A . "Privacy" in Semantic Networks on Chinese Social Media: The Case of Sina Weibo[J]. Journal of Communication, 2013(6):1011-1031.
- [16] Choi H S , Lee W S , Sohn S Y . Analyzing research trends in personal information privacy using topic modeling[J]. Computers & Security, 2017, 67(JUN.):244-253.

[18]于晓洋, 何波. 我国《个人信息保护法》立法背景与制度详解[J]. 大数据, 2022, 8(2):14.

[19]郭晓然, 方元欣, 曲东昕. 个人信息保护国际规制比较研究 [J]. 信息通信技术与政策, 2021, 47(10) : 49-55.

[20] 夏红真. 个人信息保护法草案首次亮相.[J/OL]. <http://www.npc.gov.cn/npc/c30834/202010/569490b5b76a49c292e64c416da8c994.shtml>, 2020-10-15/2022-05-12

[21]张倩. 个人信息保护法施行两个多月 公益诉讼成为个人信息保护利器. [N/OL]. http://www.zgcsjs.org.cn/News-news_list-id-58042.aspx, 2022-01-11/2022-05-12

[22][韩]박노형. 한국 개인정보보호법과의 비교를 통한 EU GDPR 의 이해[J]. 언론중재위원회 2018:62-71.

[23][韩]박노형, 고환경, 구태언. EU 개인정보보호법-GDPR 을 중심으로[M].首尔:박영사. 2017: 37.

[26][韩]황화경, 홍창표. 중국과 EU 의 개인정보보호 규정 비교와 시사점[J]. 중국지역연구. 2021(8):215-239.

[27]史洪举. 消费者不是“韭菜”，大数据得讲“规矩”. [J/OL]海外网 i. 2022-01-06/2022-03-12

[28]Graham Greenleaf. Asian Data Privacy Laws: Trade & Human Rights Perspectives[M] OUP Oxford. 2014:122

[29][韩] 박노형, 정명현. 빅데이터 분석기술 활성화를 위한 개인정보보호법의 개선 방안 - EU GDPR 과의 비교 분석을 중심으로[J]. 고려법학. 2017:1-39 -

[30][韩] 최경진. EU GDPR 의 분석 및 시사점[J]. NAVER Privacy White Paper. 2016:137-179

[31] 张建新. 国家计算机病毒应急处理中心监测发现十七款违法移动应用 [J/OL]. <http://finance.people.com.cn/n1/2021/1128/c1004-32293738.html>, 2021-11-28/2022-05-01.

[32] 占南. 国内外个人信息保护政策体系研究[J]. 图书情报知识, 2019(05):120-129.

[33] Bundeskartellamt. Bundeskartellamt prohibits Facebook from combining user data from different sources Background information on the Bundeskartellamt's Facebook proceeding.[EB/OL]. https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Pressemitteilungen/2019/07_02_2019_Facebook_FAQs.pdf?__blob=publicationFile&v=5, 2019-02-07/2022-05-12.

附录 A

附录 A.1：人民日报 TF-IDF 前 50 个词列表

	单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值
1	个人信息	31.482064	11	同意	8.440372	21	服务	6.6243546	31	公民	5.414583	41	工作	4.9412121
2	信息	19.223854	12	企业	8.1057463	22	个人信息保护法	6.4856029	32	信息处理	5.3724549	42	规范	4.936691
3	保护	17.837953	13	技术	8.0574481	23	发展	6.4840955	33	规则	5.2802957	43	采集	4.8651847
4	数据	15.510443	14	网络	7.4348291	24	互联网	6.3906295	34	数据安全	5.2735242	44	部门	4.8551003
5	app	14.905816	15	草案	7.4340551	25	提供	6.290943	35	社会	5.2047145	45	宣传	4.7704674
6	用户	12.910591	16	法律	7.4066215	26	活动	6.2745509	36	权益	5.2030588	46	风险	4.6757076
7	网络安全	11.622663	17	平台	7.3591967	27	隐私	6.2291932	37	方式	5.1634997	47	金融	4.5989749
8	收集	10.81183	18	人脸	7.300289	28	国家	6.1335313	38	诈骗	5.127357	48	治理	4.5589303
9	人脸识别	10.034271	19	相关	7.0792445	29	监管	6.0419006	39	快递	5.0736501	49	我国	4.445094
10	消费者	8.5545268	20	泄露	6.9053423	30	信息安全	5.9090183	40	数字	4.9694508	50	立法	4.3759382

附录 A.2：央视网 TF-IDF 前 50 个词列表

	单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值
1	个人信息	10.89836697	11	用户	2.654085	21	数据	1.863071	31	服务	1.4742664	41	发展	1.2725039
2	保护	5.013568313	12	提供	2.606789	22	规则	1.764316	32	制度	1.4296383	42	网络	1.2705919
3	信息	4.8335126	13	人脸	2.2927185	23	平台	1.6101347	33	完善	1.4231607	43	保障	1.2615021
4	同意	4.072262169	14	相关	2.285161	24	方式	1.5782143	34	原则	1.420556	44	中国	1.2509194
5	草案	3.384186456	15	人脸识别	2.0192794	25	隐私	1.5768631	35	企业	1.4108524	45	技术	1.2507114
6	个人信息保护法	3.369464644	16	敏感	1.9665839	26	泄露	1.5667013	36	告知	1.409925	46	规范	1.2472444
7	app	3.325145032	17	消费者	1.9302469	27	违法	1.5520337	37	社会	1.4083954	47	记者	1.2022203

8	法律	2.991794271	18	互联网	1.9017493	28	未成年人	1.5298714	38	制定	1.3625911	48	部门	1.1972085
9	快递	2.983962084	19	信息处理者	1.8735789	29	公民	1.5084286	39	目的	1.3303034	49	拒绝	1.1822461
10	收集	2.903851776	20	权益	1.866801	30	信息处理	1.481663	40	采集	1.2821746	50	实施	1.1716261

附录 A.3：微博 TF-IDF 前 50 个词列表

	单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值
1	个人信息保护法	13.68080858	11	实施	5.614684071	21	网页_链接	3.275305242	31	电话	2.910103749	41	服务	2.584273796
2	个人信息	12.43827404	12	施行	5.543883399	22	泄露	3.233092123	32	违反	2.866892278	42	微博	2.541946095
3	app	11.43895003	13	隐私	5.202567403	23	国家	3.207827805	33	整改	2.863407615	43	图解	2.527882378
4	保护	8.927729255	14	收集	4.920365282	24	平台	3.151728041	34	账号	2.850703476	44	非法	2.51737332
5	信息	7.0979835	15	微博_视频	4.538103472	25	注销	3.081355626	35	中华人民共和国	2.829785792	45	手机	2.49853095
6	下架	6.742369232	16	工信部	4.436224057	26	公民	3.053884117	36	日起	2.780003089	46	侵犯	2.449354103
7	数据	6.60911225	17	豆瓣	4.381374583	27	杀熟	3.050613543	37	我国	2.722256401	47	影响	2.393088344
8	用户	6.041578015	18	相关	3.689775987	28	消费者	3.042632681	38	组织	2.721146051	48	信息安全	2.386477287
9	正式	5.807496461	19	互联网	3.577620496	29	网络	3.035954568	39	过度	2.599365278	49	通报	2.385284124
10	法律	5.777936215	20	快递	3.276954478	30	裸奔	2.992314392	40	新闻	2.58779512	50	公开	2.380456053

附录 A.4: 朝鲜日报 TF-IDF 前 50 个词列表

	单词	中文	TF-IDF 值		单词	中文	TF-IDF 值		单词	中文	TF-IDF 值
1	데이터	数据	8.1861637	18	장관	长官	3.846817656	35	개인정보보호	个人信息 保护	2.734835654
2	개인정보	个人信息	7.827279299	19	의혹	疑惑	3.813423062	36	대검	大检察厅	2.694375922
3	의원	国会议员	6.84159647	20	규제	规制	3.681644046	37	주장	主张	2.66501879
4	혐의	嫌疑	6.163918921	21	총장	总长	3.300237425	38	한국	韩国	2.584603232
5	수사	侦查	6.00760629	22	피해자	受害者	3.238507404	39	서울	首尔	2.557118314
6	정보	信息	5.546057485	23	스캐터랩	SCATTER LAB	3.216191718	40	페이스북	脸书 (Facebook)	2.499408559
7	검사	剑士	5.510971152	24	위반	违反	3.215679705	41	관리	管理	2.472103937
8	기업	企业	5.359834126	25	이용자	用户	3.140259428	42	처벌	处罚	2.450528539
9	검찰	检察	4.881170603	26	산업	产业	3.135527171	43	이용	利用	2.444419346
10	금융	金融	4.444486534	27	공개	公开	3.077629355	44	개발	开发	2.435454937
11	사건	事件	4.376910457	28	감찰	监督	3.009139509	45	재판	审判	2.421573849
12	경찰	警察	4.348006677	29	동의	同意	2.96300708	46	개인정보보호법	个人信息 保护法	2.420569016
13	조사	调查	4.118751197	30	행위	行为	2.878724299	47	개인	个人	2.417725151
14	고발	举报	4.110325916	31	기술	技术	2.871841521	48	판사	法官	2.403698055
15	중국	中国	3.964405123	32	시장	市场	2.834537516	49	내용	内容	2.386105875
16	활용	活用	3.864954947	33	개인정보보호법위반	违反 个人信息保护法	2.764715823	50	마이데이터	Mydata	2.384271438
17	서비스	服务	3.860262992	34	제공	提供	2.757962596				

附录 A.5: SBS TF-IDF 前 50 个词列表

	单词	中文	TF-IDF 值		单词	中文	TF-IDF 值		单词	中文	TF-IDF 值
1	혐의	嫌疑	6.069119496	18	수집	收集	2.572277989	35	연합뉴스	联合新闻	2.020902097
2	의원	国会议员	5.234979623	19	의혹	疑惑	2.508723669	36	피해	受害	1.983739021
3	개인정보	个人信息	4.953044483	20	공무원	公务员	2.50682335	37	변호사	律师	1.968182288
4	수사	侦查	4.845013535	21	서울	首尔	2.401468637	38	내용	内容	1.964476778
5	경찰	警察	4.636705069	22	박사방	博士房	2.391356558	39	감찰	监察	1.95583333
6	검찰	监察	3.967365443	23	사진	图片	2.386117025	40	명의	名义	1.944699482
7	검사	剑士	3.829105381	24	위반	违反	2.336125948	41	기소	起诉	1.859084719
8	정보	信息	3.725558836	25	근무	执勤	2.323198067	42	작성	拟定	1.835984494
9	유출	泄露	3.723439172	26	직원	职员	2.319336607	43	행위	行为	1.800588536
10	사건	事件	3.45652164	27	불법	非法	2.312113588	44	동의	同意	1.781224115
11	피해자	受害者	3.302885235	28	조회	查询	2.246901829	45	출석	出席	1.754915231
12	재판	审判	3.062590376	29	여성	女生	2.244053339	46	적용	适用	1.754572117
13	조사	调查	2.890191376	30	공개	公开	2.19277903	47	스캐터랩	SCATTER LAB	1.751153052
14	문건	文件	2.701606189	31	병원	医院	2.164802608	48	구속	拘留	1.748062096
15	고발	举报	2.697889559	32	처벌	处罚	2.159780517	49	이용	利用	1.735939297
16	기자	记者	2.670945809	33	기록	记录	2.117995708	50	영상	视频	1.731794161
17	개인정보보호법위반	违反个人信息保护法	2.593601452	34	주민센터	居民中心	2.106078044				

附录 A.6: 韩国推特 TF-IDF 前 50 个词列表

	单词	中文	TF-IDF 值		单词	中文	TF-IDF 值		单词	中文	TF-IDF 值
1	개인정보보호법	个人信息保护法	153.3974093	18	총력	全力	23.91471426	35	범죄	犯罪	13.37002522
2	개인정보보호법위반	违反个人信息保护法	71.21313672	19	공개	公开	23.21319628	36	이용	利用	13.03795074
3	위반	违反	62.83290435	20	처벌	处罚	19.99078563	37	문건	文件	12.95296307
4	개인정보	个人信息	46.89817983	21	검찰	监察	18.83948686	38	목적	目的	12.94654531
5	수집	收集	34.76779519	22	개인	个人	17.88297556	39	위반혐의	违反嫌疑	12.74077366
6	정보	信息	30.58342154	23	데이터	数据	16.68367247	40	고소	起诉	12.6594528
7	고발	举报	28.61877819	24	혐의	嫌疑	16.65748997	41	기자	记者	12.42764712
8	주의	注意	26.93127724	25	번호	号码	15.40695421	42	의원	国会议员	12.244031
9	위반여부	是否违反	26.24034148	26	불법	非法	15.33227086	43	한국	韩国	12.24278753
10	당부	嘱咐	26.01720551	27	수사	侦查	15.2047457	44	정부	政府	12.18629343
11	대한의사협회	韩国医生协会	25.93071469	28	국민	国民	14.92471507	45	직권남용	滥用职权	12.02907987
12	주민번호	身份证号码	25.84805783	29	유출	泄露	14.90516434	46	사찰	查察	11.83688409
13	허용	许可	24.63804141	30	뉴스	新闻	14.49402949	47	행위	行为	11.78838849
14	윤석열	尹锡悦	24.39915505	31	경찰	警察	14.46427321	48	개인정보보호	个人信息保护	11.77978606
15	동의	同意	24.32350936	32	이하	以下	14.43624832	49	공무원	公务员	11.77951815
16	신고	申告	24.06495424	33	제공	提供	14.40258706	50	전화번호	电话号码	11.65989165
17	병협	韩国医院协会	23.91471426	34	검사	剑士	14.34420807				

附录 A.7: DWnews TF-IDF 前 50 个词列表

	单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值
1	datum	7.081550272	11	personal	1.827870222	21	whatsapp	1.501166428	31	year	1.241583407	41	app	1.115502295
2	say	4.771049786	12	european	1.793574487	22	social_media	1.438196849	32	control	1.239833105	42	include	1.106703524
3	facebook	4.409851379	13	read	1.633506322	23	consent	1.362654482	33	time	1.231835739	43	breach	1.105432288
4	company	3.745798974	14	make	1.623991573	24	fine	1.349732259	34	many	1.217736232	44	election	1.081054093
5	user	3.459115683	15	gdpr	1.616778597	25	come	1.344684772	35	german	1.176719351	45	online	1.07736722
6	privacy	2.513800157	16	people	1.586988587	26	service	1.32341561	36	scandal	1.176459892	46	citizen	1.075024757
7	law	2.38161397	17	information	1.535364333	27	internet	1.296219776	37	digital	1.170003782	47	access	1.07045688
8	platform	2.097358452	18	use	1.521232742	28	business	1.282673514	38	government	1.152848565	48	account	1.06841846
9	protection	1.983746827	19	rule	1.516179122	29	group	1.272183802	39	tax	1.138636985	49	regulation	1.065472983
10	new	1.949416188	20	right	1.505573151	30	take	1.257380566	40	firm	1.123447172	50	country	1.055435851

附录 A.8: Euronews TF-IDF 前 50 个词列表

	单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值
1	datum	11.884251	11	new	3.3845687	21	fine	2.379559	31	share	2.1774527	41	include	1.9830259
2	say	6.7266003	12	people	3.3348924	22	consumer	2.3636326	32	year	2.174501	42	find	1.9793406
3	user	6.0366434	13	rule	3.296149	23	need	2.3428811	33	citizen	2.1548186	43	know	1.9680506
4	company	5.6201529	14	app	3.2307314	24	right	2.3129678	34	internet	2.1321808	44	work	1.9676767
5	facebook	5.3460163	15	information	3.1779319	25	take	2.3069307	35	want	2.0826314	45	time	1.9195541
6	privacy	4.316589	16	law	3.1009441	26	see	2.2484777	36	go	2.0798672	46	collect	1.8827203
7	protection	3.9654718	17	personal	2.9689662	27	business	2.2315106	37	come	2.0730215	47	government	1.8778419
8	european	3.825021	18	use	2.8011438	28	website	2.2114747	38	health	2.0449973	48	data	1.8644458
9	gdpr	3.6126574	19	digital	2.612701	29	country	2.1985151	39	think	2.0109885	49	breach	1.850466
10	regulation	3.5262468	20	technology	2.4011679	30	make	2.1952736	40	authority	1.9975118	50	account	1.8397478

附录 A.9: 欧盟推特 TF-IDF 前 50 个词列表

	单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值		单词	TF-IDF 值
1	gdpr	246.5013646	11	regulation	42.7116471	21	user	35.55035631	31	time	30.28378984	41	thing	26.16425142
2	datum	126.2633113	12	company	41.862084	22	good	35.12538212	32	think	29.90110294	42	give	26.06936622
3	privacy	96.05024743	13	day	39.81643534	23	email	34.04682853	33	read	29.82117528	43	security	26.05258234
4	com	71.20555001	14	need	38.61287204	24	say	33.98148547	34	fine	29.36547867	44	blockchain	26.00445007
5	compliance	47.67977291	15	know	38.10286086	25	today	33.25204576	35	news	29.29151446	45	talk	25.51518131
6	get	47.52186138	16	law	37.14704173	26	see	32.46109613	36	right	28.3682134	46	information	25.46774045
7	Data protection	47.50114704	17	personal	36.54671334	27	come	31.42068175	37	people	28.31395858	47	share	25.4444443
8	protection	45.3956065	18	consent	36.22746762	28	work	31.17836471	38	take	28.05451756	48	great	25.34419867
9	compliant	44.1436599	19	go	36.12914537	29	data	30.90459792	39	facebook	27.87657497	49	use	25.28995936
10	new	43.35566138	20	make	35.88112097	30	business	30.64138687	40	thank	26.86160293	50	article	25.06431911

致谢

这些时间让我不得不怀着感恩之心。首先，感谢闫蒲老师在百忙之中担任我的学位论文的审查委员，并提出好的建议。非常感谢老师指导我的大部分研究,通过尖锐的评论给予我很大帮助。通过老师的指导，我知道了研究是什么，我的不足之处是什么，应该朝着什么方向发展。还有感谢我的周边人。

感谢一直相信我，默默支持我的亲爱的父母。我会努力做好，成为报答父母恩惠的自豪的儿子。

北京大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名： 权九洋

日期： 2022 年 05 月 24 日

学位论文使用授权说明

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

- 按照学校要求提交学位论文的印刷本和电子版本；
- 学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
- 学校可以采用影印、缩印、数字化或其它复制手段保存论文；

论文作者签名： 权九洋 导师签名： 闫蒲

日期： 2022 年 05 月 24 日