



CoLaxy 中科天玑

中科天玑 数据能力介绍





1

总体能力



境内外开源数据采集能力

采集分类	采集通道	信源数量	采集信息	采集规模
互联网	新闻	1.5万+网站	文章：标题、内容、发布时间、URL、网站、作者等 评论：用户昵称、评论内容、评论时间、头像、原文id、原文URL等	200万条/日
	电子报	1000+网站	文章：标题、内容、发布时间、URL、网站、作者等	2万+/日
	智库	150+中外知名智库	帖子：标题、内容、发布时间、URL、网站、作者、报告附件地址等	1000+/日
两微一端	微博	2亿+用户	微博信息：微博ID、微博内容、发布时间、用户ID、用户昵称、用户头像、微博URL、转发数、评论数、点赞数、图片列表等 M和T方向的网站2000个，账号8W+，重点账号5000个	7000万+/日
	微信	200万+活跃用户	微信公众号文章：标题、内容、发布时间、URL、公众号id、公众号名称等	180万+/日
	APP移动端	100+APP	文章：标题、内容、发布时间、URL、网站、作者等 评论：用户昵称、评论内容、评论时间、头像、原文id、原文URL等	300万/日
短视频	抖音、快手、火山、西瓜、好看、美拍、bilibili	2亿+用户	视频信息（发布时间、视频url、用户ID、用户名、关注数、粉丝数、视频ID、转发数、收藏数、弹幕数、视频分享数、观看数、播放量、评论数量、视频封面图链接、视频简介、无水印下载链接。） 、 用户信息	1.2亿+/日
互动社区	论坛	8万+板块	帖子信息：标题、内容、发布时间、URL、网站、作者等	30万+/日
	问答	2千+板块	帖子：标题、内容、发布时间、URL、网站、作者等	1万+/日
JW	境外新闻	1万+网站	文章：标题、内容、发布时间、URL、网站、作者等	15万/日
	Twitter	5千万+用户	推文信息：推文ID、推文内容、推文创建时间、推文是否原创、推文转发数、推文喜爱数、用户ID、用户名称、用户头像、用户注册时间、用户粉丝数、用户关注数、用户推文数、用户喜爱数等 个人基本信息：人物名、人物屏幕名称、人物粉丝数目、人物好友数目、人物简介、人物创建时间、人物头像、人物位置信息、人物帖子数目等； 好友或粉丝信息：人物名、人物屏幕名称、人物粉丝数目、人物好友数目、人物简介、人物创建时间、人物头像、人物位置信息、人物帖子数目等 评论转推信息：推文ID、推文评论内容、推文创建时间、转推人用户名称、转推人用户头像、转推人注册时间、转推用户粉丝数、转推人关注数、转推人推文数、转推人喜爱数等	400万+/日
	Facebook	10万+用户	推文信息：推文ID、推文内容、推文创建时间、推文图片URL列表、推文评论数、推文转发数等 个人基本信息：人物名、人物屏幕名称、人物头像、个人描述、个人主页地址等 好友信息：人物名、人物屏幕名称、人物头像、个人描述、个人主页地址等	10万+/日
	Instagram	1000+主流媒体和大V账号	帖子信息：推文ID、推文内容、推文创建时间、推文图片URL列表、推文评论数、推文转发数等	4万+/日
	Telegram	1000+主流媒体和大V账号	帖子信息：推文ID、推文内容、推文创建时间、推文图片URL列表等	8万+/日
	YouTube	1万+账号	视频信息：视频ID、视频封面图链接、视频发布时间、视频描述、播放量、评论数、点赞数、视频地址等	
	领英	4亿+账号	用户履历信息：名字、头像、描述、个人技能、教育经理、工作经历等	
	mastodon		推文信息：推文ID、推文内容、推文创建时间、推文是否原创、推文转发数、推文喜爱数、用户ID、用户名称、用户头像、用户注册时间、用户粉丝数、用户关注数、用户推文数、用户喜爱数等 个人基本信息：人物名、人物屏幕名称、人物粉丝数目、人物好友数目、人物简介、人物创建时间、人物头像、人物帖子数目等；	10万+/日



重点账号积累情况

国内重点账号2000+。FH NGO近千个，FH个人账号近千个，FH组织及其成员账号数百个：账号、姓名、性别、学历、住址、FH活动等。涉K、涉邪等账号数百个。

序号	姓名 (必填)	外文名	所在机构	职务	性别	出生日期	国家	省/州	市/区	个人简介		
1	黎智英	Jimmy Lai Chee-ying	壹传媒创始人	香港壹传媒创始人	男	1948/12/8	中国	香港		香港企业家，佐治		
2	陈方安生	Chan Fang On Sang		首位华人布政司	女	1940/1/17	中国	香港		陈方安生曾经是		
3	戴耀廷	Benny Tai Yiu-ting	香港大学	香港大学副教授	男	1964/7/12	中国	香港		戴耀廷中学毕业		
4	李柱铭	Martin Lee Chu-ming	香港公民党	香港民主党创党主席	男	1938/6/8	中国	香港		广东惠州人，成		
5	何俊仁	Albert Ho Chun-Yan	香港公民党	香港民主党前主席 香港立法会议员	男	1951/12/1	中国	香港		广东中山人，成		
6	李永达	Lee Wing-tat	香港公民党	香港民主党前主席 香港立法会议员 葵青	男	1955/12/25	中国	香港		广东惠阳人，出		
7	胡志伟	Wu Chi-wai	香港公民党	香港民主党主席 香港民主党立法会议员	男	1962/10/18	中国	香港		香港出生，前黄		
8	林卓廷	Lam Cheuk Ting	公民党立法会	香港民主党立法会议员	男	1977/6/13	中国	香港	屯门	香港立法会议员、		
9	许智峰	Hui Chi-fung	公民党立法会	香港民主党立法会议员	男	1982/6/8	中国	香港	屯门	现任香港中西区		
10	尹兆坚	Andrew Wan	公民党立法会	香港民主党立法会议员	男	1969/6/7	中国	香港	九龙	香港政治人物，		
11	涂谨申	James To Kun-sun	公民党立法会	香港民主党立法会议员	男	1963/3/11	中国	香港	九龙	香港大学法律系		
12	黄碧云	Helena Wong Pik-wan	公民党立法会	香港民主党立法会议员	女	1959/3/21	中国	香港	九龙	现任香港九龙西		
13	邝俊宇	Roy Kwong Chun-yu	香港公民党	香港民主党中央委员会委员	男	1983/2/9	中国	香港	元朗	香港立法会议员、		
14	梁家杰	Alan Leong Kah-kit	香港公民党	香港公民党主席 香港立法会议员	男	1958/2/22	中国	香港	旺角	广东新会人，香		
15	杨岳桥	Alvin Yeung	公民党立法会	香港公民党立法会议员	男	1981/6/5	中国	香港	元朗	香港执业大律师、		
16	郭荣铿	Dennis Kwok Wing-hang	公民党立法会	香港公民党立法会议员	男	1978/4/15	中国	香港		香港政治人物、		
17	毛孟静	Claudia Mo Man-ching	公民党立法会	香港公民党立法会议员	女	1957/1/18	中国	香港	跑马地	祖籍浙江宁波，三教、家庭教会		
18	谭文豪	Jeremy Tam Man-ho	公民党立法会	香港公民党立法会议员	男	1975/6/13	中国	香港		出生于香港，祖		
19	陈淑庄	Tanya Chan	公民党立法会	香港公民党立法会议员	女	1971/9/14	中国	香港		祖籍上海，香港		
20	郭永健	Steven Kwok Wing-kin	香港工党	香港工党主席	男	1986/11/30	中国	香港		祖籍广东揭阳，		
21	李卓人	Lee Cheuk-yan	香港工党	香港工党副主席	男	1957/2/12	中国	香港		广东潮阳人，上		
邓聿文					@dyw1968316					https://twitter.com/dyw1968316	纽约时报》专栏作者 政治评论员、海外中文媒体撰稿人	中国政府、中国对外关系
滕彪					@tengbiao					https://twitter.com/tengbiao	维权律师，“零八宪章”发起人之一，曾响应所谓“茉莉花革命”	
											人权、死刑、中国政府	



虚实映射能力

虚实映射的人物组织数据库，基于wiki结构化生成，人工打标补充完善，优质人物数据50万+、优质机构数据50万+。

模糊 ▾

请输入目标名称

类型 ▾

请选择地区

GD X

搜索

新建目标档案

模糊 ▾

请输入目标名称

类型 ▾

请选择地区

TD X

搜索

新建目标档案



梁继昌
人物 GD
地区：中国香港
简介：未知



张敬轩
人物 歌手、词曲作家、录音...
地区：中国香港
简介：香港歌手



刘祖迪
人物 政治活动者、土地测量...
地区：未知
简介：香港LIHKG讨论区会员，政治...



卢建民
人物 涉港、GD、民运



多維勒-薩卡利埃內
人物 媒体、科研、大学教员...
地区：立陶宛/考纳斯
简介：多維勒-薩卡利埃內是立陶宛...



楊佩樺
人物 政治人物、护理人员、T...
地区：中国台湾/高雄市
简介：内科、急重症護理師，台湾基...



Simon-Pierre Savard-Tremblay
人物 媒体、舆论记者、非虚...
地区：加拿大/魁北克市
简介：Simon-Pierre Savard-Trembla...



彼得·海特
人物 政治人物、德国联邦议...



林保華
人物 TD、作家、政治评论员...
地区：中国台湾/台北市
简介：笔名凌鋒、艾克思和林卫等，...



Chris Lewis
人物 政治人物、加拿大国会...
地区：加拿大
简介：Canadian politician who was ...



Angelo Iacono
人物 政治人物、律师、加拿...
地区：加拿大
简介：Canadian politician and lawye...



馬大悅·阿里·塞拉
人物 政治人物、Member of t...



柯景騰(九把刀)
人物 作家、电影导演、编剧...
地区：中国台湾/彰化县
简介：臺灣作家、導演



Richard Martel
人物 Ice hockey coach、政治...
地区：加拿大/萨格奈
简介：Canadian politician and form...



茱蒂·史葛洛
人物 政治人物、加拿大国会...
地区：加拿大/蒙特利尔
简介：Canadian politician. A membe...



Asep Ahmad Maoshul Affandy
人物 政治人物、Member of t...



2

对M数据



涉M数据积累：政要数据

117届国会议员、2022年国会候选人、建制派、特朗普支持者、右翼势力
4000+、NGO账号数据数百个、其他A政治人物和意见领袖2000+

类别	党籍	两院	职位	英文名-去空格	英文名	author_name
右翼势力	共和党			Abraham Hamade	Abraham Hamadeh	Abe Hamadeh for Arizona AG
右翼势力	共和党			Adam Laxalt	Adam Laxalt	Adam Paul Laxalt
右翼势力	共和党			Alison Hayden	Alison Hayden	Alison Hayden 賀亞珊us CA-14 vs.Eric Swalwell
右翼势力	共和党			Andy Biggs	Andy Biggs	Andy Biggs
右翼势力	共和党			Barry Loudermilk R	Barry Loudermilk R	Barry Loudermilk
右翼势力	共和党			Ben Cline	Ben Cline	Ben Cline
右翼势力	共和党			Bill Posey	Bill Posey	Bill Posey
右翼势力	共和党			Brian Babin	Brian Babin	Brian Babin
右翼势力	共和党			Brian Perras	Brian Perras	Brian Perras
右翼势力	共和党			Byron Donalds	Byron Donalds	Byron Donalds
右翼势力	共和党			Chip Roy	Chip Roy	Chip Roy
右翼势力	共和党			Bob Good	Bob Good	Congressman Bob Good
右翼势力	共和党			Dan Bishop	Dan Bishop	Dan Bishop
右翼势力	共和党			Dan Cox	Dan Cox	Dan Cox us - Delegate & Candidate for Governor
右翼势力	共和党			Darren Bailey	Darren Bailey	Darren Bailey for Governor
右翼势力	共和党			Darren Aquino	Darren Aquino	Darren.Dione.Aquino usCONGRESSusUS.House Elec
右翼势力	共和党			Debbie Lesko	Debbie Lesko	Debbie Lesko
右翼势力	共和党			Gary Palmer	Gary Palmer	Gary Palmer
右翼势力	共和党			Greg Steube	Greg Steube	Greg Steube
右翼势力	共和党			Jim Marchant	Jim Marchant	Jim Marchant



涉M数据积累：专业人物账号

新闻工作者5000+、专家学者1000+、经济学家400+、科研人员1W+、企业家1000+



Marc Gonzalez
人物 媒体、新闻
地区：美国/登维尔
简介：American TV



Kyle Cheney
人物 媒体、新闻
地区：美国
简介：American journalist



David A. Weiner
人物 媒体、电视
地区：美国
简介：American film



Jess Bravin
人物 媒体、新闻



Blanche D
人物 科学
地区：美国
简介：univ



Michael B
人物 科学
地区：美国
简介：com



德沃尼·卢卡瑟
人物 科学
地区：美国



拉贾·克里希纳穆尔蒂
人物 企业 科研 工程师
地区：美国
简介：印度



Raul Moreno
人物 企业、企业家
地区：美国
简介：Entrepreneur from the United...



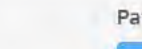
Tex Keith
人物 企业、企业家、首席执行官
地区：美国/普雷斯科特山谷
简介：Tex Keith is an american entre...



Joel Mark Witt
人物 企业、企业家、popular...
地区：美国
简介：US entrepreneur



Jed McCaleb
人物 企业、程序员、企业家...
地区：美国/旧金山



Patty Talahongva
人物 企业 学院教授 记者



Ted Nadeau
人物 企业、企业家
地区：美国
简介：Entrepreneur from the United...



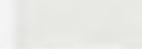
Nathan Latka
人物 企业、投资者、作者、...
地区：美国
简介：author



Cliff Lerner
人物 企业、商人、企业家、...
地区：美国
简介：Amerikaans ondernemer



Cyrus Massoumi
人物 企业、企业家、创立者
地区：美国



查理·安德森
人物 科研 作家 历史学家



Soraya Darabi
人物 企业、企业家、天使投...
地区：美国
简介：Entrepreneur from the United...



Holley M. Murchison
人物 企业、作家、教师、企...
地区：美国/温斯顿-塞勒姆
简介：author, education producer, a...



Clara Villarosa
人物 企业、企业家、作家、...
地区：美国/芝加哥
简介：American entrepreneur, auth...



Chris Wanstrath
人物 企业、软件开发者、企...
地区：美国



涉M数据积累：专业机构账号

科研机构8000+、军事机构200+、企业2W+



德克萨斯基督教大学

科研机构 科研、私立大学、...

地区：美国

简介：private university in Fort Wort...



普林斯顿大学工程与应用科学学院

科研机构 科研、科研机构、...

地区：美国

简介：esports team sponsored by th...



美国陆军电子竞技

军事机构 电玩游戏战队、军...

地区：美国

简介：esports team sponsored by th...



United States Forces Afghanis...

军事机构 一体化作战司令部

地区：美国

简介：미국 중부사령부 예하 통합전...



VMU-3

军事机构 军事单位、飞行中...

地区：美国

简介：unmanned aerial vehicle squa...



军眷与士兵福利休闲计划

军事机构 军事单位

地区：美国



耶鲁大学工程与应用科学院

科研机构 科研、研究所

地区：美国

简介：military unit



VR-55

军事机构 军事单位

地区：美国

简介：military unit



3d Space Experimentation Sq...

军事机构 军事单位

地区：美国

简介：military unit



United States Naval Sea Cadet...

军事机构 军事单位

地区：美国

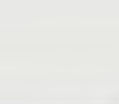
简介：U.S. Navy sponsored organiza...



西南师

军事机构 军事单位

地区：美国



美国驻韩海军陆战队部队

军事机构 军事、司令部

地区：美国

简介：미국 해병대의 한반도 전구 사...



第132联队

军事机构 军事单位

地区：美国

简介：military unit



坦克汽车研究、开发和工程中心

军事机构 军事单位

地区：美国

简介：military unit



南大西洋

军事机构 军事单位

地区：美国



3

对T数据



涉T数据积累：政要数据

T政要数据，含党派、电话、传真、邮箱、出生地点、工作地点、信仰、教育经历、社会关系等。

name_cn	name_en	name_alia	nickname	标签	profession	wiki_cn	wiki_en	potrait	country	party_name	organ	post	birth_plac	birth_date	residence	sex	brief_intro	pub
林佳龙	Lin Chia-lun		龙哥、龙龙	政客	政治学	https://zh.wikipedia.org/wiki/林佳龙	https://www.10.gov.tw/1010.jpg		中国台湾	民主进步党			台北市万华	1964-02-11	台中市西区	男	林佳龙（台《》）	
卓伯源				政客		https://zh.wikipedia.org/wiki/卓伯源	\1050.jpg		中国台湾	中国国民党			彰化县田中	1965-03-27		男	卓伯源（台《》）	
张亚中				政客、国际		https://zh.wikipedia.org/wiki/张亚中	\1051.jpg		中国台湾	中国国民党	孙文学堂	总校长	台中市	1954-12-21		男	张亚中（台《》）	
陈时中	Chen Shih-wei		阿中部长	政客、牙医		http://www.10.gov.tw/1010.jpg			中国台湾			政客						
黄伟哲	Huang Wei		阿哲市长	政客		http://www.10.gov.tw/1010.jpg			中国台湾									
陈其迈	Chen Chi-mai		迈迈、暖男	政客		http://www.10.gov.tw/1010.jpg			中国台湾									
林智坚	Lin Chih-chiang			政客		http://www.10.gov.tw/1010.jpg			中国台湾									
赖清德	Lai Ching-te			政客、内科		http://www.10.gov.tw/1010.jpg			中国台湾									
蔡英文		Tsai Ing-wen	小英、辣妹	政客、大学竞争法、国际		http://www.10.gov.tw/1010.jpg			中国台湾/大安区									
李大维				政客、外交		http://www.10.gov.tw/1010.jpg			中国台湾/台北市									
苏贞昌				政客、前特		http://www.10.gov.tw/1010.jpg			中国台湾/台北市									
沈荣津				政客	工业管理	http://www.10.gov.tw/1010.jpg			中国台湾/新竹市									
李孟谚				政客		http://www.10.gov.tw/1010.jpg			中国台湾/新竹市									
游锡堃			水牛伯、	政客		http://www.10.gov.tw/1010.jpg			中国台湾/新竹市									
蔡其昌				政客		http://www.10.gov.tw/1010.jpg			中国台湾/新竹市									
林志嘉				政客		http://www.10.gov.tw/1010.jpg			中国台湾/新竹市									
许宗力				政客、著名		http://www.10.gov.tw/1010.jpg			中国台湾/桃园县									
周弘亮				政客		http://www.10.gov.tw/1010.jpg			中国台湾/桃园县									
刘建忻				政客		http://www.10.gov.tw/1010.jpg			中国台湾/桃园县									
陈菊			花妈、菊姐	政客		http://www.10.gov.tw/1010.jpg			中国台湾/桃园县									
朱富美				政客		http://www.10.gov.tw/1010.jpg			中国台湾/桃园县									
徐国勇			勇哥、阿勇	政客		http://www.10.gov.tw/1010.jpg			中国台湾/桃园县									

柯文哲
人物 政治人物、医生、外科...
地区：中国台湾/大安区
简介：台湾政治人物

蘇貞昌
人物 政治人物、律师、中华...
地区：中国台湾/屏东市
简介：台湾政治人物、行政院院长

陳淑華
人物 政治人物、立法委员、...
地区：中国台湾/台南市
简介：台湾政治人物、环保运动人士...

柯建銘
人物 政治人物、立法委员、...
地区：中国台湾/新竹市
简介：台湾政治人物，民主进步党籍...

賴清德
人物 政治人物、医生、立法...
地区：中国台湾/万里区
简介：中華民國第15任副總統，台湾...

蔡英文
人物 科研、政治人物、律师...
地区：中国台湾/中山区
简介：中華民國第14、15任總統，台...

王定宇
人物 政治人物、立法委员、...
地区：中国台湾/桃园县
简介：台湾政治人物，民主进步党籍...

蔣萬安
人物 政治人物、律师、立法...
地区：中国台湾/台北市
简介：中華民國政治人物、中國國民...

邱顯智
人物 律师、政治人物、人权...
地区：中国台湾/竹崎乡
简介：台湾政治人物、律师

林俊惠
人物 政治人物、立法委员、...
地区：中国台湾/桃园县
简介：台湾政治人物，民主进步党籍...

范云
人物 科研、社会学家、政治...
地区：中国台湾/台北市
简介：台湾政治人物、中国国民...

鄭文燦
人物 政治人物、桃园市市长...
地区：中国台湾/桃园市
简介：台湾政治人物、桃园市市长...



涉T知识图谱

T政府人物组织图谱，包括组织名称、成立时间、创始人、负责人、政治倾向、成员规模、关注领域、电话、地址、职能、对华态度、坐标；人物姓名、婚姻状况、出生日期、性别、党派、政治倾向、受教育程度、民族、职务、工作单位、教育经历等。

组织名称	简介(字符串)	成立时间	创始人	负责人(字符串)	政治倾向	成员规模	关注领域	电话	邮箱	地址	组织架构			
新党	新党是中华民国的	1993年8月10日	赵少康、郁慕	吴成典	蓝营	约500人	政治、民主	(02) 2756-2222		台北市松山区光复南路	郁慕明曾经把新党			
民主进步党	民主进步党（英语：	1986年9月28日	陈水扁，沈富	赖清德	绿营	约24万人	进步主义 自	06-2655120		台北市中正区北平东路	民主进步党的组织			
中国国民党	中国国民党（Kuomi	1894年11月24日	孙中山、黄兴	朱立伦	蓝营	约37万人	时事、党政	(02) 87711234		台湾省台北市中山区八	中国国民党的党务			
台湾民众党	台湾民众党（台湾	2019年8月2日	柯文哲	蔡壁如	白营	约10万人	民主、自由、多元、开放、法治、人权			台北市士林区松山南路	台湾民众党的党务			
中国民众党	中国	关注领域	电话	邮箱	地址	组织架构	职能/主张	对华态度	社交账号	组织特点(字符串)	坐标(小数别名(字符串)英文名(字符串)			
无党团结联盟	无党	政治、民主	(02) 2756-2222		台北市松山区光复	郁慕明曾经把新	新党主张清廉制	亲华	0004458171270	中华民族主义、中国	121.52706		New Party (Taiwan)	
工党（台湾）	198	进步主义 自	06-2655120		台北市中正区北平	民主进步党的组	如果国共片面和	反华	000443929906	“台独”、“住民自	121.52007	民进党	Democratic Progressi	
中华统一促进党	为促	时事、党政	(02) 87711234		台湾省台北市中山	中国国民党的党	依据《中华民国宪法》，	西	000443659921		121.54176	国民党	Kuomintang of China	
台湾基进	台湾	民主、自由、多元、开放、法治、人权			台北市中正区杭州	台湾民众党自命	民众党宗旨为，以台湾的	独	000441756521		121.52007		Taiwan People's Party	
台湾团结联盟	台湾				基隆市信义区忠孝里深溪路36巷70弄23号					信仰国家社会主义，	121.77229			
社会民主党	社会				台北市大同区南京西路344巷25号6楼					自由保守主义	121.51392	无盟无党联	Non-Partisan Solidar	
亲民党	亲民				台南市新化区民治路43巷21号					两岸统一、社会主义	120.34361			
台湾劳动党	劳动				新北市中和区健康路21号7楼					和平统一、一国两制	121.49308		Unionist Party (Chine	
姓名	婚姻状况	简介	出生日期	性别	党派	政治倾向	受教育程	年龄	民族	职务	工作单位	教育经历	基进党TSP	Taiwan Statebuilding
邱国正	已婚	邱国正（1	1953年4月12日	男	无党籍			70		台湾国防部部长	台湾国防部	台北市立党	台联党	Taiwan Solidarity Un
锺树明		锺树明（1	1964年	男				59		第10任陆军司令	台湾陆军司令部		社民党	Social Democratic Pa
陈建义		陈建义196	1968年	男						陆军司令部参谋长	台湾陆军司令部			People First Party
敖以智		敖以智，中华民国海军中将，		男						海军第1副司令	台湾海军司令部			The Labor Party of T
刘任远		刘任远（1	1963年	男			硕士毕业	60		空军司令	台湾空军司令部			The People Union Par
孙连胜		孙连胜（1	1964年	男				59	鲁凯族	空军副司令	台湾空军司令部			Green Party Taiwan
马英汉		马英汉，中华民国国军陆军中将，为第一任资通电军指挥官					硕士毕业			指挥官	台湾资电作战指挥部		时力（NPP	New Power Party
梅家树		梅家树（1	1963年	男				60		海军司令部参谋长	台湾海军司令部		维新党	Taiwan Renewal Par
黄佑民		黄佑民，中华民国海军中将，		男			硕士毕业			常务次长	台湾国防部			
陈忠文		陈忠文（1	1964年	男				59		陆军副司令	台湾陆军司令部			
蔡沐霖		蔡沐霖（1	1983年4月18日	男	民主进步	绿营		40		民进党前青年发展部主任	民进党青年发展部			



涉T数据积累

类型	子类	数据量
新闻资讯及社媒数据	内容数据，关系数据	6221577+
全球涉T信源	——	10000+
重点目标社交账号	——	5000，可实现虚实映射
T知识图谱数据	——	10w以上，包括：人物共14881，三元组共182238；其他实体：56413，三元组共403242，总的三元组数：585480。
人物库	T政要画像	170
	T民众户籍数据（闭源数据）	2300w
	宗教界、学术界、商界人物等	13847
组织库	T政党资料	75个政党
	T智库名单	——
	T企业资料	1478家企业，通联方式、地址（台在大陆的企业、500强、上市企业等）
	T宗教资料	13546个宗教团体
	T民调机构资料	——
	T族群结构资料	——
	T人民团体相关资料	23785个人民团体



涉T数据积累

基础资料	TW选举相关资料	——
	TW政治献金相关资料	台湾大选期间台内部的政府献金，献金对象、时间、金额、来源
	美对T军售资料	——
	T在美游说相关资料	——
	美军舰自由航行资料	——
	TW稳定性评估指标数据	2006-2021年，51个指标，经济社会指标
	T社会域基础资料（人口、就业率、宗教/族群分布.....）	——
	2017-2012路透社数字新闻研究所报告（包含Tw社群平台受欢迎程度排名，新闻信任度，报纸、电视、收集新闻的受众等）	
	TW民众TW人/中国人认同趋势分布	来源：国立政治大学选举研究中心
	TW民众统独立场趋势分布	来源：国立政治大学选举研究中心
	TW民众政党偏好趋势分布	来源：国立政治大学选举研究中心
	TW历年统计年鉴	——



涉南海数据



涉南海数据：人物、组织账号数据

美国、日本、越南、菲律宾、马来西亚、文莱、印尼、中国台湾等周边声索国及主要关涉方的人物、组织账号数据和虚实映射数据。

模糊

请输入目标名称

人物

模糊

请输入目标名称

政府

模糊

请输入目标名称

菲律宾

请输入标签进行搜索

搜索

新建目标档案



Kit Thompson

人物

模特儿

地区：菲律宾/安赫莱斯

简介：Filipino model



Rommel Curaming

人物

political historian, Sou...

地区：菲律宾/卡塔纳万

简介：Philippine political historian



Ryan Maboloc

人物

科研、政治哲学家、大...

地区：菲律宾

简介：Philippine political philosopher



罗伯特·柏力克

人物

篮球运动员

地区：菲律宾/马尼拉



驻菲律宾台北经济文化办事处

政府

台湾驻外机构、T驻外机...

地区：菲律宾/马尼拉

简介：Taipei Economic and Cultural ...



Embassy of Poland, Manila

政府

大使馆、涉政

地区：菲律宾

简介：سفارة



Land Transportation Franchisi...

政府

政府机构、涉政

地区：菲律宾

简介：Philippine government agency



关心教师联盟

政府

左翼、全国民主运动、...

地区：菲律宾



ACT-CIS Partylist

政府

政党、涉政

地区：菲律宾

简介：political party in Philippines



Embassy of the Netherlands, ...

政府

大使馆、涉政

地区：菲律宾

简介：diplomatic mission of the Net...



美国驻菲律宾大使馆

政府

大使馆、涉政

地区：菲律宾

简介：diplomatic mission of the Uni...



菲律宾民主社会党

政府

中间偏左、分产主义、...

地区：菲律宾



沙烏地阿拉伯駐菲律賓大使館

政府

大使馆、涉政

地区：菲律宾

简介：سفارة



World Health Organization

政府

政府机构、涉政

地区：菲律宾

简介：government organization in ...



Butil Farmers Party

政府

农业主义、cooperative ...

地区：菲律宾

简介：agrarian political party in the ...



Department of Education Cag...

政府

政府机构、涉政

地区：菲律宾

涉南海数据：政情百科数据

美国、日本、越南、菲律宾、马来西亚、文莱、印尼、中国台湾等周边声索国及主要关涉方的政情百科数据，涉及基本国情、人口、城市、宗教、军事实力、犯罪率等。

国家地区	时间	军事地位 (排名现代化水平)	军事预算(单位: 美)	军队规模	核武器 (单位)
印度尼西亚	2000年	国家军由正规部	1.13E+09	484000	无核
印度尼西亚	2001年		9.19E+08	499000	无核
印度尼西亚	2002年		1.37E+09	499000	无核
印度尼西亚	2003年		2.13E+09	492000	无核
印度尼西亚	2004年		2.43E+09	492000	无核
印度尼西亚	2005年	24	2.15E+09	492000	无核
印度尼西亚	2006年	24	2.61E+09	497000	无核
印度尼西亚	2007年	13	3.35E+09	582000	无核
印度尼西亚	2008年	13	3.23E+09	582000	无核
印度尼西亚	2009年	13	3.30E+09	582000	无核
印度尼西亚	2010年	14	4.66E+09	582000	无核
印度尼西亚	2011年	18	5.84E+09	582000	无核
印度尼西亚	2012年	15	6.53E+09	582000	无核
印度尼西亚	2013年		8.38E+09	582000	无核
印度尼西亚	2014年	19	6.93E+09	676500	无核
印度尼西亚	2015年	12	7.60E+09	676500	无核
印度尼西亚	2016年	14	7.40E+09	676500	无核
印度尼西亚	2017年	14	8.80E+09	676500	无核
印度尼西亚	2018年	15	7.49E+09	675500	无核

印尼总人口(单位: 亿)

年份	人口 (亿)
2000年	2.05
2001年	2.08
2002年	2.11
2003年	2.14
2004年	2.17
2005年	2.20
2006年	2.23
2007年	2.26
2008年	2.29
2009年	2.32
2010年	2.35
2011年	2.38
2012年	2.41
2013年	2.44
2014年	2.47
2015年	2.50
2016年	2.53
2017年	2.56
2018年	2.59
2019年	2.62
2020年	2.65
2021年	2.68
2022年	2.71
2023年	2.75

印尼军事预算(单位: 美元)

年份	预算 (美元)
2000年	1.50E+09
2001年	1.60E+09
2002年	1.70E+09
2003年	1.80E+09
2004年	1.90E+09
2005年	2.00E+09
2006年	2.10E+09
2007年	2.20E+09
2008年	2.30E+09
2009年	2.40E+09
2010年	2.50E+09
2011年	2.60E+09
2012年	2.70E+09
2013年	2.80E+09
2014年	2.90E+09
2015年	3.00E+09
2016年	3.10E+09
2017年	3.20E+09
2018年	3.30E+09
2019年	3.40E+09
2020年	3.50E+09
2021年	3.60E+09
2022年	3.70E+09
2023年	3.80E+09





5

涉Z数据



涉Z数据积累：人物组织数据

可虚实映射的涉藏人物和组织100+，含姓名、简介、性别、国籍、所在地区、出生日期、语言、外文姓名、种族民族、所属政党、出生地、职业、担任职务、就读院校、社交账号（脸书、推特、电报、telegraph、reddit、油管等）。

序号	平台	中文名	账号	链接	备注
1	推特	自由亚洲电台	RebelRFACHinese	https://twitter.com/RebelRFACHinese	
2	推特	法广	RFI 華語 - 法國國際廣播電台	https://twitter.com/RFI_TradCn	
3	脸书	西藏新闻	Tibet News	https://www.facebook.com/Tibatnews/	
4	脸书	西藏电视台	TIBET TV		
5	脸书	自由亚洲电台藏文版			
6	脸书	美国之音藏文版			
7	推特	维吾尔时报中文（维吾尔之声）	维吾尔时报中文（维吾尔之声）	https://twitter.com/...	
8	推特	RFI 华语 - 法国国际广播电台	RFI 华语 - 法国国际广播	https://twitter.com/RFI_Chinese	
9	推特	美国之音争鸣论坛	美国之音争鸣论坛	https://twitter.com/voachina	
10	推特	DW 中文- 德国之声	DW 中文- 德国之声	https://twitter.com/dwchinese	
11	推特	BBC News 中文	BBC News 中文	https://twitter.com/bbcchinese	
12	推特	美国之音中文网	美国之音中文网		
13	推特	独立中文笔会 ICPC	独立中文笔会 ICPC	https://twitter.com/ICPC	独立中文笔会





人物 ZD

地区：美国

简介：འཕྲིན་རྒྱུ་མཉམ་སྦྲེལ་གྱི་མཁའ་མཛོད་



索甲仁波切

人物 作家、演员、喇嘛、ZD

地区：中国/西藏

简介：Tibetan Dzogchen lama of th...



班玛噶波

人物 ZD、政治人物、教授、...

地区：日本/东京

简介：政治學家、桐蔭橫濱大學教授...



Tibet Bureau Geneva

其他组织 ZD、人物

地区：瑞士/日内瓦

简介：The Tibet Bureau Geneva is th...



Tenzin Wangchuk

人物 ZD、人物

地区：印度/新德里

简介：西藏青年大會成員



Kelsang Tsomo Sky

人物 ZD、人物

地区：印度/班加罗尔

简介：Fashion Designer and Creator...



Logon Rinpoche

人物 ZD、人物

地区：印度/达兰萨拉

简介：འཕྲིན་རྒྱུ་མཉམ་སྦྲེལ་གྱི་མཁའ་མཛོད་འཕྲིན་རྒྱུ་མཉམ་སྦྲེལ་གྱི་མཁའ་མཛོད་



Mere Bharat

媒体 ZD、新闻资讯、人物

地区：美国

简介：Mere Bharat mahen hai



Yael Gal

人物 ZD、人物

地区：美国

简介：未知



次旺仁增

人物 政治人物、ZD、重要人...

地区：美国/明尼苏达/明尼阿波利斯



国际西藏独立运动

NGO 非营利组织、ZD

地区：美国/印第安那



西藏青年大会纽约分部

其他组织 ZD、涉政

地区：美国/纽约



6

涉J数据



涉J数据积累：JD人物组织虚实映射数据

可虚实映射的JD人物及组织近百个。含社工数据、人脸数据、社交账号（脸书、推特、电报、telegraph、reddit、油管等）。

库扎特·阿尔泰

Former Chinese Citizen | American Businessman | Harvard Business Alum | Independent Uyghur State | Founder & CEO <https://t.co/ZtXr9GBeH4> | China, Uyghur, Tech | 真实姓名库扎提买买提，原“世维会”主席热比娅·卡德尔的侄子，目前在美国活动，常与美国议员有密切接触，并时常参加欧洲议会、美国国会听证会等西方政府官方正式场合，与美国政府工作人员或其他活跃“东突”分子及“东突”组织领导人共同参加示威活动，时常举行维吾尔议题相关会议并发表讲话。籍贯：新疆乌鲁木齐，国籍：美国，宗教信仰：伊斯兰教，已婚，妻子拉希拉·阿尔泰系；教育背景：1997年至2003年，就读于乌鲁木齐市第十四中学；2018年至2019年，参加哈佛商学院高管培训课程。精通中文、英语、维吾尔语和土耳其语；职业：企业家、演讲家，办公地址 Cybertek：美国弗吉尼亚州泰森斯角琼斯布兰奇路7920号3300号办公楼（7925 Jones Branch Dr, #3300, Tysons, VA 22102, USA）；职务：“美国维吾尔协会”主席、“维吾尔企业家网络”创始人、信息技术教育公司Cybertek创始人兼CEO；家庭住址：美国弗吉尼亚州麦克莱恩市风山路1327号（1327 Windy Hill Rd, Mc Lean, VA 22102）工作经历：2008年8月至2012年8月，在华盛顿都会区的Consultant公司担任软件质量检查工程师；2012年9月至2013年10月，在美国技术公司Inovalon任自动化测试工程师；2013年10月至2014年3月，在总部位于加拿大蒙特利尔的跨国公司CGI集团任自动化测试工程师；2014年3月至2014年10月，在跨国管理咨询公司埃森哲（Accenture）任自动化开发员；2014年10月至2016年2月，在美国房屋抵押贷款公司房利美（Fannie Mae）任自动化开发员；2015年8月至今，创办信息技术教育公司Cybertek并担任首席执行官；2019年10月29日，当选“美国维吾尔协会”主席。

目标类型 **人物** 目标标签 **涉疆** **涉疆**

基本信息

性 别	男	国 籍	美国	所在地区	美国
出生日期	未知	语 言	未知	外文姓名	Kuzzat Altay
种族民族	未知	所属政党	未知	入库时间	2023-08-23 15:48:10





涉J数据积累：8类小语种语音识别的舆情数据

涉疆信息全量语音识别，支持维吾尔语、哈萨克语、乌兹别克语、土库曼语、塔吉克语、柯尔克孜语、塔塔儿语、阿尔泰语8类语音识别。

特色专题

短视频

信息列表

全部 2.1万

抖音 11.6万

快手 4,054

微博 6

今日头条 2.1万

小红书 252

2024-03-20 14:19:36

2024-03-20 14:19:36

2024-03-20 14:19:36

2024-03-20 14:19:36

2024-03-20 14:19:36

2024-03-20 14:19:36

2024-03-20 14:19:36

2024-03-20 14:19:36

来了一部我的抖音生活日记 #

能意拍

Дос көрек мыған дос көрек

1303120777

筛选条件

收起 ^

基本筛选

地域: 全部 北京市 天津市 河北省 山西省 内蒙古自治区 辽宁省 更多 >

领域: 全部 社会民生 城市管理 社会问题 司法相关 涉政相关 更多 >

时间: 今天 昨天 3天 7天 30天 自定义

属性: 全部 正面 中偏正 中性 中偏负 负面

结果: 全部 正常信息 疑似敏感

类型: 全部 原创 评论 转发

地域命中: 全部 标题/正文 AI识别 注册地 IP属地 定位地 作者名称

语种: 全部 维吾尔语 哈萨克语 乌兹别克语 土库曼语 塔吉克语 柯尔克孜语 塔塔儿语 阿尔泰语

账号信息

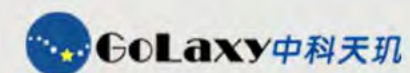
收起 ^

认证: 全部 蓝V 黄V 普通用户

性别: 全部 男 女 未知

粉丝: 全部 100万以上 50万-100万 10万-50万 5万-10万 1万-5万 1万以下 自定义

保存并搜索 重置 搜索



THANK YOU



电话: 400-901-9755

官网: <http://www.golaxy.cn/>

地址: 北京市海淀区西三旗中科智汇工场N1楼301号





面向争夺国际话语权和维护国家意志需要，中科天玑团队投入近 50 余人，基于在舆情、情报分析领域多年积累的学术与技术成果，研发出可服务国家层面的认知域对抗、引导的智宣 GoPro 系统。

GoPro 系统部分技术成果已在国务院港澳办、安全部、联参情报局、战支一线部队等部门部署应用，并为应用部门在回应党中央重大关切，诸如香港问题，台海问题，中美围绕疫情、贸易、高科技、大选等展开的博弈，印度地缘政治冲突等方面，提供了态势感知、事件分析、舆论引导等技术支持。

下面给各位领导详细介绍 GoPro 系统。

首先，GoPro 系统综合不同应用单位的工作需要，基于天玑 WDE 平台，从传统媒体、社交网站、即时通信等 800 万个信源，广泛采集香港、台湾、美国、印度等地区的政治、军事、外交、经济数据，并每日实时聚合出不同区域的热门和敏感政治事件，深度分析与该事件相关的话题、媒体和人物，挖掘媒体和相关人物所持立场与观点。

其次，GoPro 系统基于积累十年的情报分析技术和 GoIN 系统，对政治事件幕后组织、机构、人物进行精准画像，并通过广谱关联技术实时锁定目标，以近期香港国安法和蓬佩奥拟访台事件为例，从全网监测的 966 万 twitter 用户和 18 万香港用户中，锁定与这两起事件相关的 3000 多



名主要参与人物、100 多名网络意见领袖和 20 多家新闻媒体。锁定目标后，快速回击目标对象的黄论、谬论，及时正本清源。

第三，在事件的舆论征候和特定区域的风俗习惯等基础信息之上，一是通过研判对方涉我言论的真伪和逻辑漏洞；二是寻找对方内部分化和极化观点，比如用马斯克对中国抗疫的言论反驳美西方国家对我的抹黑和污蔑，实现以己之矛攻己之盾；三是利用人工智能技术如自动联想、素材借鉴、文案衍生、长文本生成等方式辅助创作中国抗疫好故事。

最后，非受控环境下的智能推送，综合解决一体化指挥调度、智能化传播导入、科学化评估反馈、拟人化生存等技术难题。视频中向各位领导展示，利用智能推送功能，可实现跨 Twitter、Facebook 等平台同时对目标对象进行实时、高可达、浸入式的舆论引导。

天玑团队自研的 GoPro 系统，目前已具备政治态势感知、实时锁定、高质量内容创作、快速回击等能力，已在国家相关部门产生一定政治效应。未来我们期盼在科学院领导和所领导的大力支持下，进一步加强与中央网信办、中宣部、军科委、安全部等部门的合作，将 GoPro 打造成为能真正自动讲好中国故事、发出中国声音、扩大中国影响力的技术平台，为早日实现“东风压倒西风”提供全方位的技术支持。



会议纪要

会议主题：纵横系统工作进展汇报沟通会

会议时间：2023年5月11日下午2:00

会议地点：通研院林萃桥办公区一层大会议室

参加人员：郭杰，三家课题相关人员
苏鹏，课题五，南京信息技术研究院

一、课题一（王清林，刘杨昊，黄振明，王洪航，隋雨恒）：

1. 汇报进度：完成全部内部编码，完成联调（90%）。
2. 下两周迭代：调试数据，集成测试，沟通外部接口。
3. 播放系统演示视频，大家评审。
4. 会议输出：

（一）郭所反馈：郭所了解了课题一整体工程进度（在计划内）。同时要求课题一刻盘当天的演示视频给他，已完成。

（二）确定下周对外沟通会议议程：

1. 接口课题三：下周一下午下午（5月15日），去清华，讨论仿真推演数据需求和方案。

2. 接口课题二：周四下午（5月18日），地点另行协商讨论对接接口（1.敏感事件，2.隐含关系，3.为舆图提供活数据）

三）苏老师反馈：

Q：事件生成的关键词逻辑可以支持多关键词间的“与、或、非以及分组逻辑”。



A: 回去调研相关算法部分。

(四) 课题三反馈:

Q: 数据资源监控功能没有体现。

A: 数据资源监控功能做在运维视角子系统里面,不在当前面向业务人员视角的子系统里面。天玑有相关设计,未开始。

(五) 视频演示反馈或者发现的问题:

- 1) 事件: 相关推文的“暂无信息”
- 2) 群组: 按什么维度去划分,存在几个建群组的属性,
- 3) 人物画像: 存在字符集问题; 有多少个维度
- 4) 平台和语种: 语种问题不大, 单个平台到多个平台画像对齐没有设计
- 5) 降噪处理: 事件关键词组合过滤文本
- 6) 新闻数据 时间问题 (2023 年 8 月 1 日的新闻)

二、课题二 (仝春艳, 闫洲):

1. 系统开发进展:

特定 HL (黑料) 曝光: 基于内容的隐含线索挖掘

涉 H 议题引导: 摇摆人 (精英人物、普通网民)

摇摆人受众精准定位: 精英人物, 基于利益立场导向
普通民众, 基于情感态度导向

XJ (选举) 舆论引导: 最复杂的, 两党中的精英之争

XJ 舆论场引导: 两党汇总的精英之争->选举舆论场态势



技术。

感知->下发合理的任务目标->基于 XJ 影响的策略制定
评估 XJ 舆论场两党各自的主要影响因素:确定主要人物对
华态度->评估精英动员的影响能力->确定特定议题对候
选人的影响路径

任务评估: 多次投放, 多次评估, 目前整体做评估
用户 zz 心理画像模型:

zz 立场模型

大五人格模型

舆图-议题-党派分布: 效果图 还未接入真实数据

舆图-议题-社会阶层分布

账号可以定位地区上

人物视图也加进来了

议题下面有评论的数据: 前两百条左右

地图数据: 采集不到, 使用框架进行集成, 开源的地图框架
openMap?

年龄分布: 用模型去推断年龄

立场: 在一个议题下, 以账号视角去看, 三种立场: 支持,
反对, 中立。

语种: 目前能够支持中英文, 模型好说, 训练集比较难找

图数据库: neo4j

2. 会议反馈

郭所: 关注现有设计前提下实际采集来的数据支持能力如何,



系统要能数据落地，课题二反馈目前有推断数据。

苏老师：可以基于原始数据进行属性补全

对课题三数据库设计反馈（训瀚）：重构数据库部署设计，三

家数据库分隔开。各个课题分库：读写分离。

对课题一数据采集有特定要求（接下来会提出）：

议题，带评论的数据，前 200 条即可

地图，采集有困难，目前用的开源数据）

三、课题三（高云鹏，徐文博，刘训瀚等四人）：

1.完成素材发现的设计和功能实施

2.机器人自动化注册脚本以及接口封装（短期：半自动人工验证）

3.机器人账号的作战任务的设计，日志视图

4.新增 200+ Twitter 草根账号的数据

5.仿真推演模型的实施方案

账号养殖情况：

1. twitter：婴儿期大 v：10 婴儿期普通：50 草根：600

2. reddit：20 个草根

3. facebook：完成基础调研工作：不好找突破口

4. 购买境外服务器，用 IP 池，精养账号绑定固定 IP，炮灰账号 IP 在州范围内流动。

5. 目前用了两天服务器，采用 Selenium Webdriver 模拟



技术。

账号养殖问题:

1. 账号养殖: 模拟器 网页端 培育: 手机版、模拟器
2. 使用邮箱账号, 后面有可能要求绑定手机号
3. 自动化脚本: 当前围绕网页, 调研模拟器
4. 没有动作, 被锁的概率: 5%
5. 邮箱: 目前用的开源邮箱, 自己搞一个域

下两周任务:

1. 仿真推演
作战投放

Facebook 平台突破方案, 交付多平台账号养殖
权限方案设计

郭所反馈: 素材发现不仅满足于采集形成的素材, 尽量支持
语义搜索形成轻量级的素材挖掘渠道。

高老师回复: 已实现采集回来的素材 (新闻), 工程易实现
支持语义搜索的素材挖掘, 在模型开发过程中。



中科天玑数据科技股份有限公司

技术产品体系



公司战略和产品定位



- 1956年成立
- 从事计算机科学技术综合性研究的学术机构
 - 中国计算机事业的“摇篮”
 - 国家高性能计算机研发基地

中国科学院计算机研究所

- 中国计算机事业的“摇篮”
- 孕育创建联想、曙光、天玑等20多个各领域高新技术企业
- 1958年 中国第一台数字电子计算机 -103
- 22位院士工作或学习过
- 孕育创建联想、曙光、天玑等20多个各领域高新技术企业



- 2010年成立
- 中科院计算所孵化的高新技术企业
 - 以大数据智能计算引擎、人工智能算法为核心
 - 服务于党政军和教育、金融等行业



**联想**
**中科曙光**
**GoLaxy**中科天玑

大数据分析系统国家工程实验室

- 突破大数据分析技术瓶颈，研制第三代大数据分析软件栈
- 2017年发改委批复依托计算所成立；李国杰院士任主任，程学旗副所长任常务副主任
- 共建单位：国科大、网络中心、中科天玑、曙光、国创科视

中科院网络数据科学与技术重点实验室

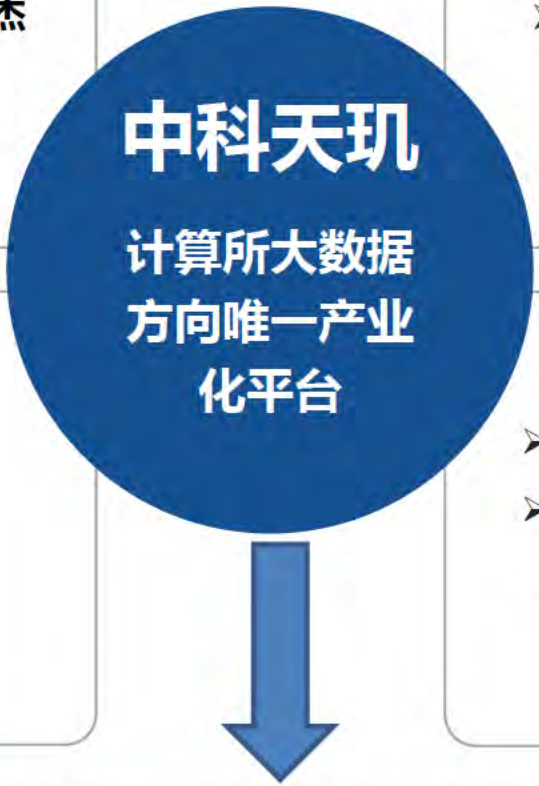
- 中科天玑团队的重要组成部分
- 先后承担了多项国家重点863、973、科技攻关等重大课题项目，科研成果多次获得国家及省部级奖励

计算所&地方分所

- 烟台网络技术研究所：80人
- 郑州大数据研究院：50人
- 福州数据应用技术研究院：40人
- 厦门。。。

中科天玑研发中心

- 75%占比的技术研发团队
- 提供党军政、教育、金融等领域的大数据整体解决方案的应用



大数据分析技术领域“国家队”，硕士、博士100+人，科研人员400+人



服务于国家安全、国家战略

三大目标市场：党、政、军

二大技术应用：舆情、情报大数据分析

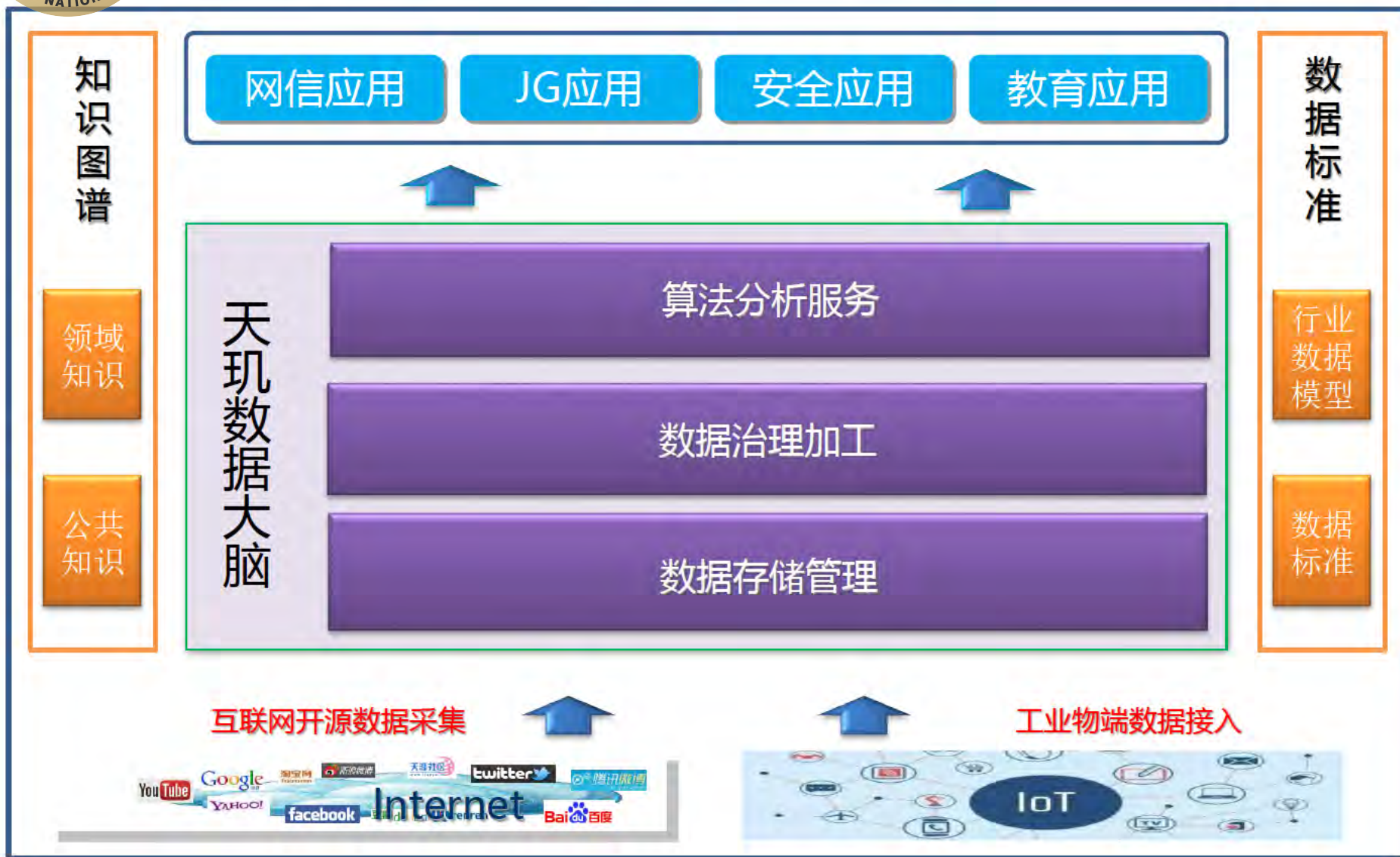
一个核心技术系统：天玑数据大脑

情报大数据分析
第一品牌



天玑技术产品体系及产品家族





天玑数据大脑提供
从数据到价值的数据
服务生产全链条

1. 数据模型能力
2. AI算法模型能力
3. 行业应用能力

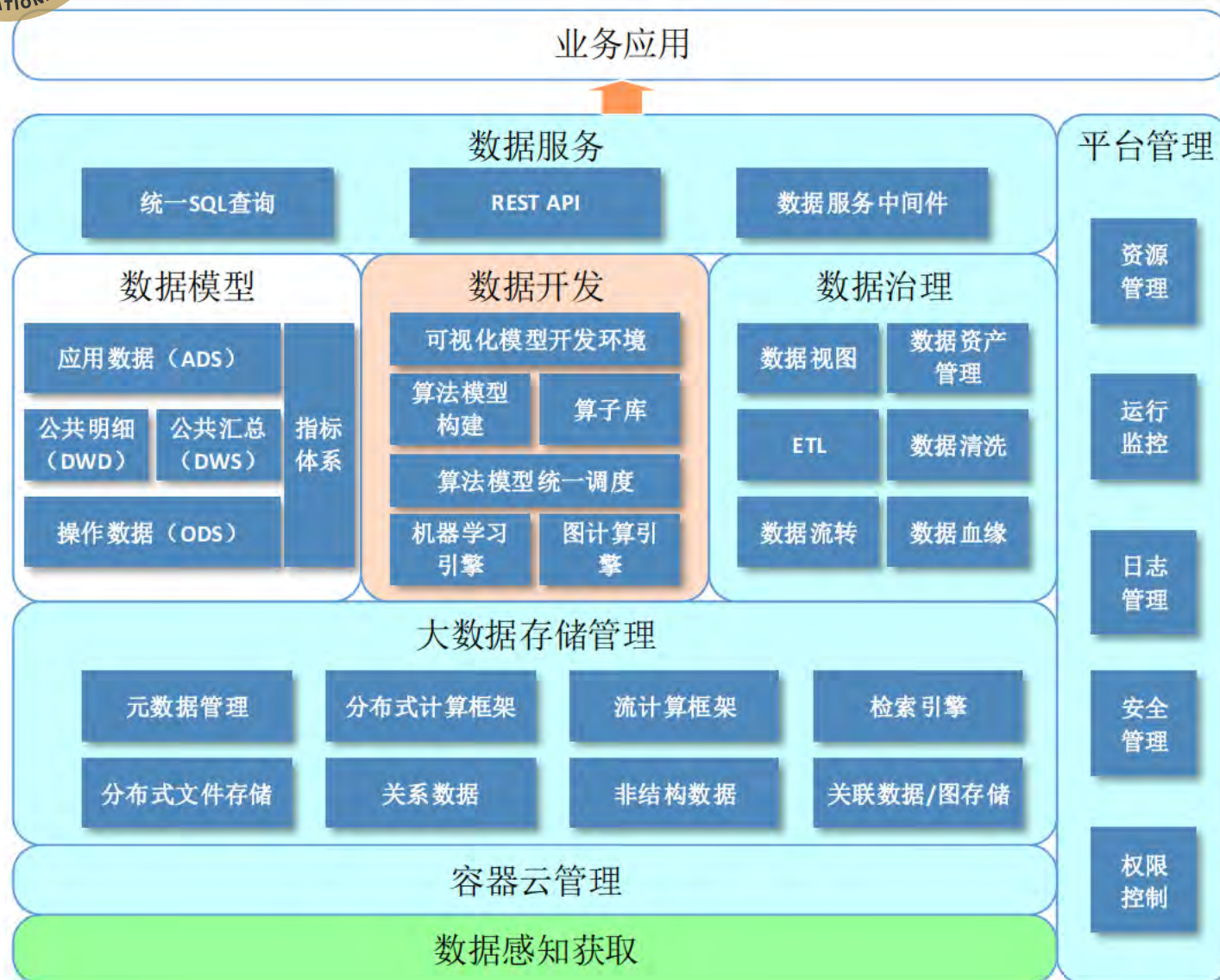




基础平台产品



基础平台产品线



天玑数据大脑，为各行业产品和应用提供数据中台支持

- 可视化智能机器学习系统 (BDA)
- 大数据存储管理平台 (BDE)
- 数据采集服务平台 (WDE)

衍生产品形态

- 自主可控大数据一体机
- 超融合架构大数据一体机



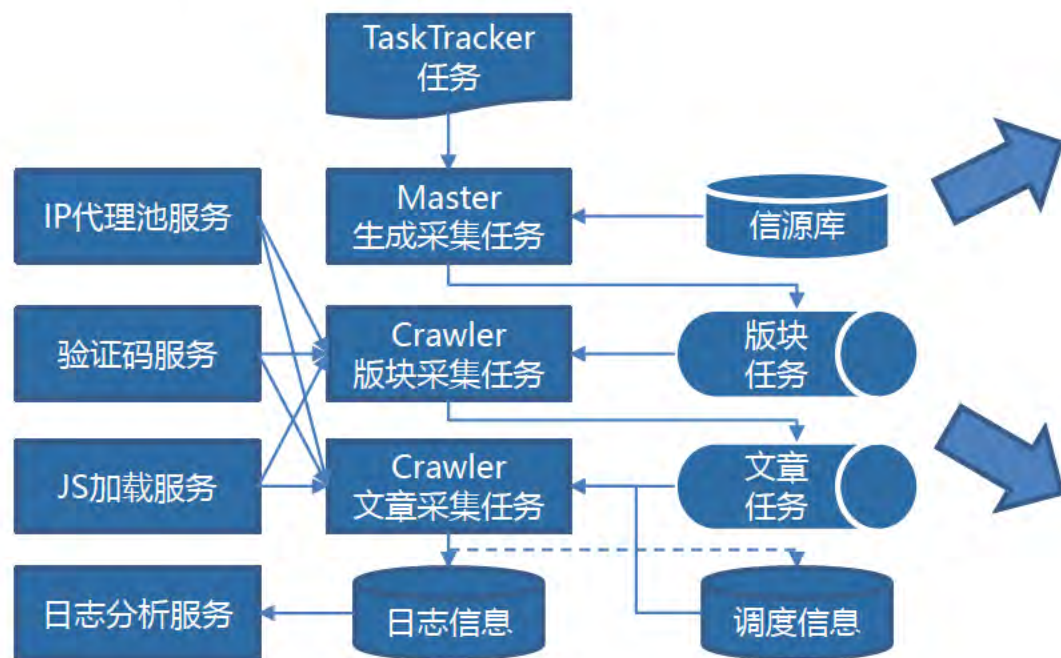
- 面向政府监管、安全、军工、金融、教育、企业等领域提供数据服务
 - 定制化采集与分析，包含实时数据采集、历史数据回溯、热点数据、数据分析报告，实现数据的按需获取
 - 以多种API接口或文件的方式进行推送
 - 平台覆盖境内和境外海量数据，支持国内主流新闻媒体、微博、微信、APP客户端，以及论坛、博客、贴吧、问答社区、短视频等通道，以及境外媒体、facebook、twitter等
 - 全天采集数据量在8000万-1亿



数据采集服务平台 (WDE)

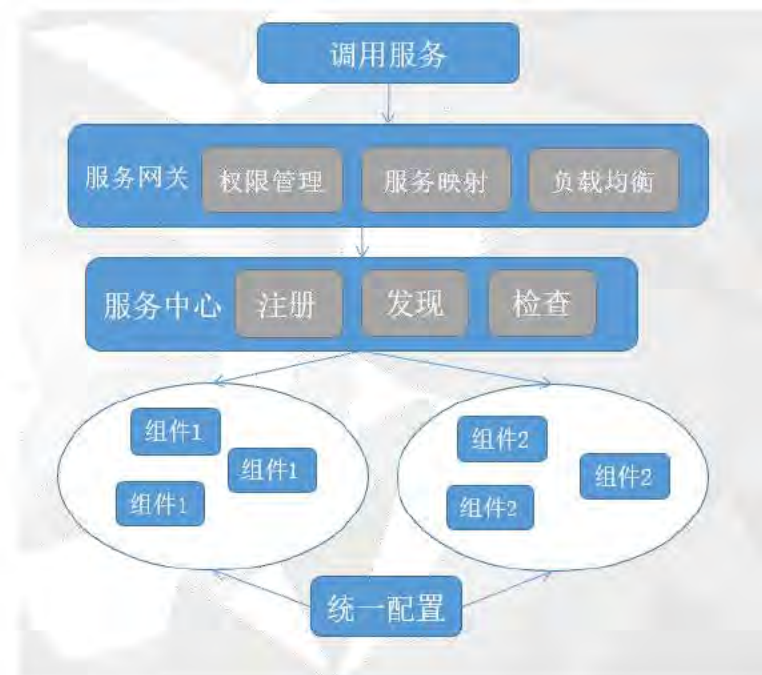
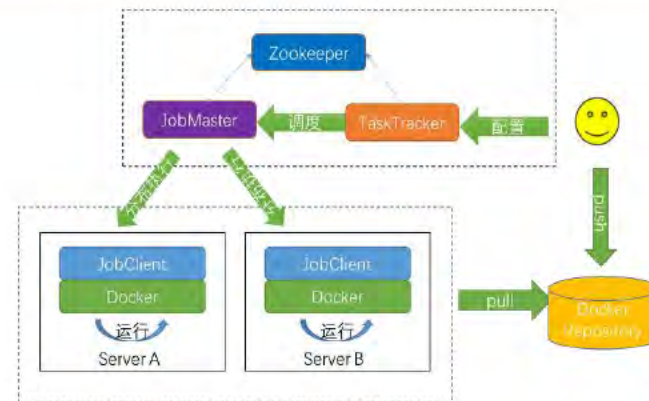


产品架构



分布式调度框架

采集服务框架





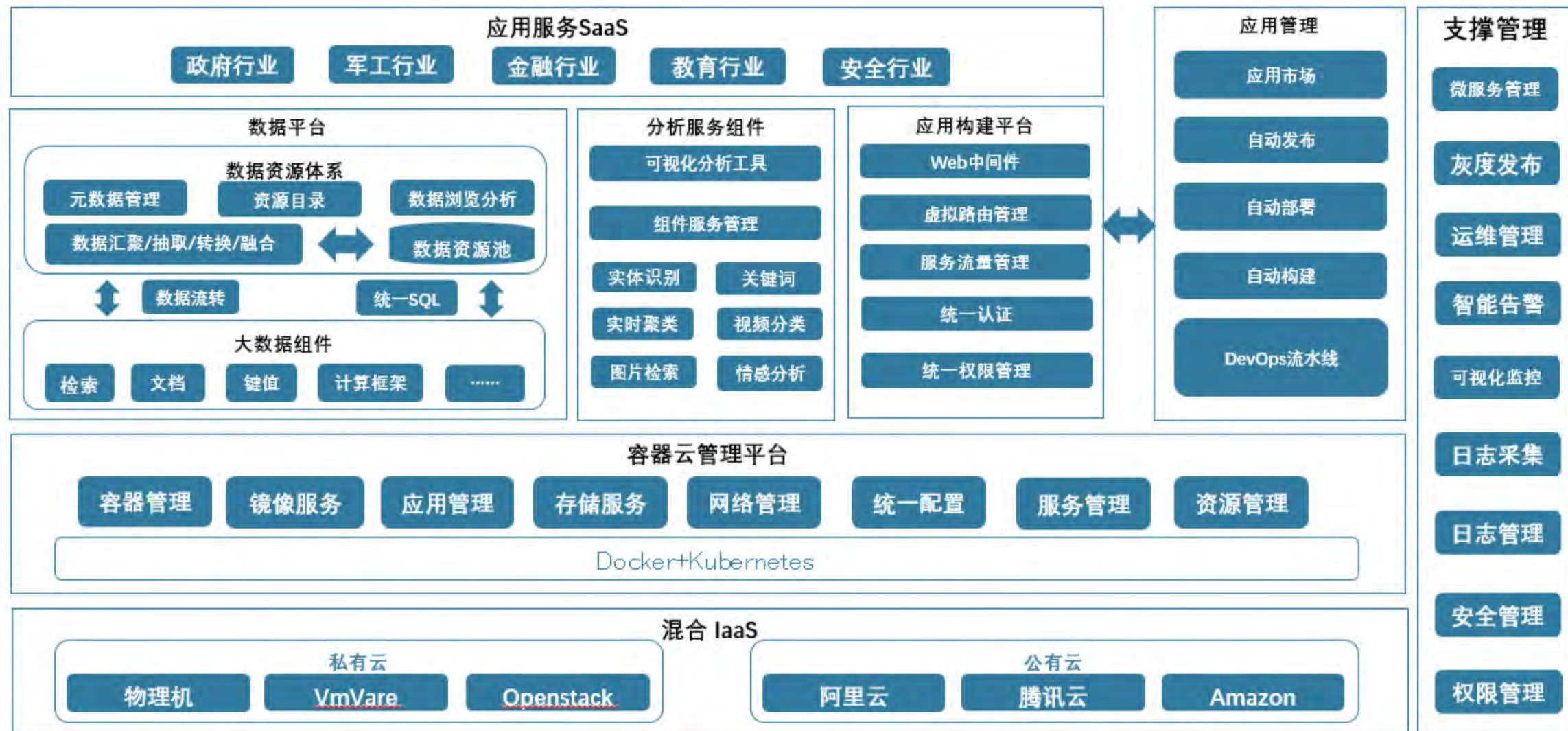
大数据存储管理平台 (BDEC)



- 大数据存储管理平台 (BDE) 是企业级大数据一体化应用开发基础技术平台，将容器云平台与大数据系统相结合，实现了云+大数据+智能服务的产品技术融合，为用户提供构建大数据业务系统所需的完整的基础平台能力
- 基于容器云环境，提供了IaaS+PaaS+SaaS多层服务能力，不同层级的组件和服务统一运行于容器云支撑平台，相互共享资源，资源可以依赖运行情况弹性伸缩，提高整体运行周期的资源利用率
- 覆盖数据存储、分析挖掘和应用构建三方面需求，是一个具备大数据系统全栈研发能力的一站式、标准化、可视化的智能大数据统一研发平台



大数据存储管理平台 (BDEC)



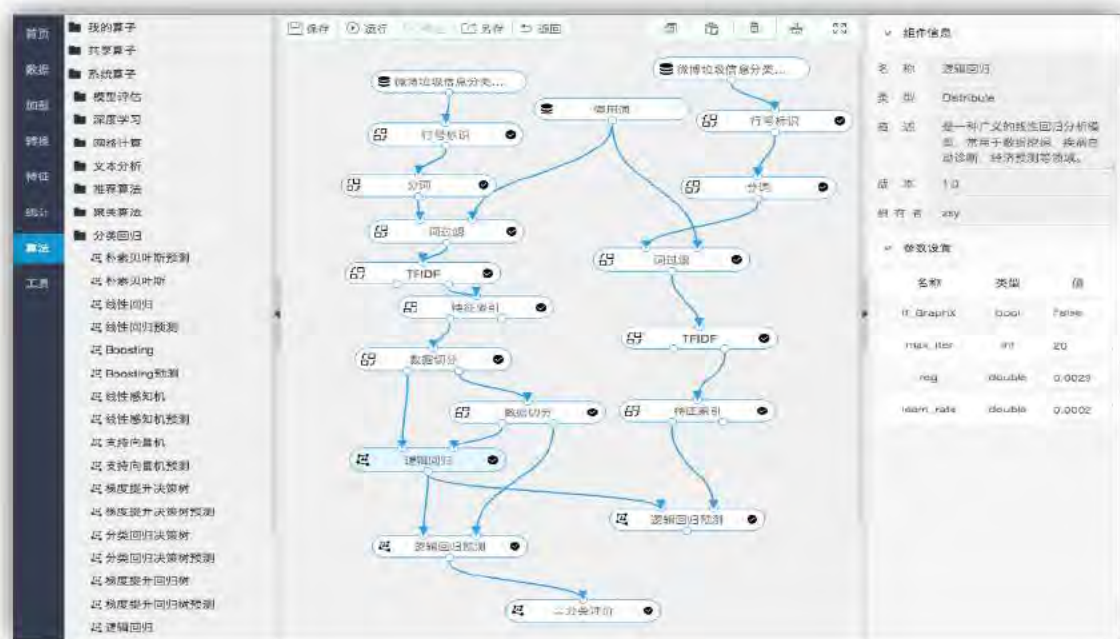


可视化智能机器学习系统 (BDA)



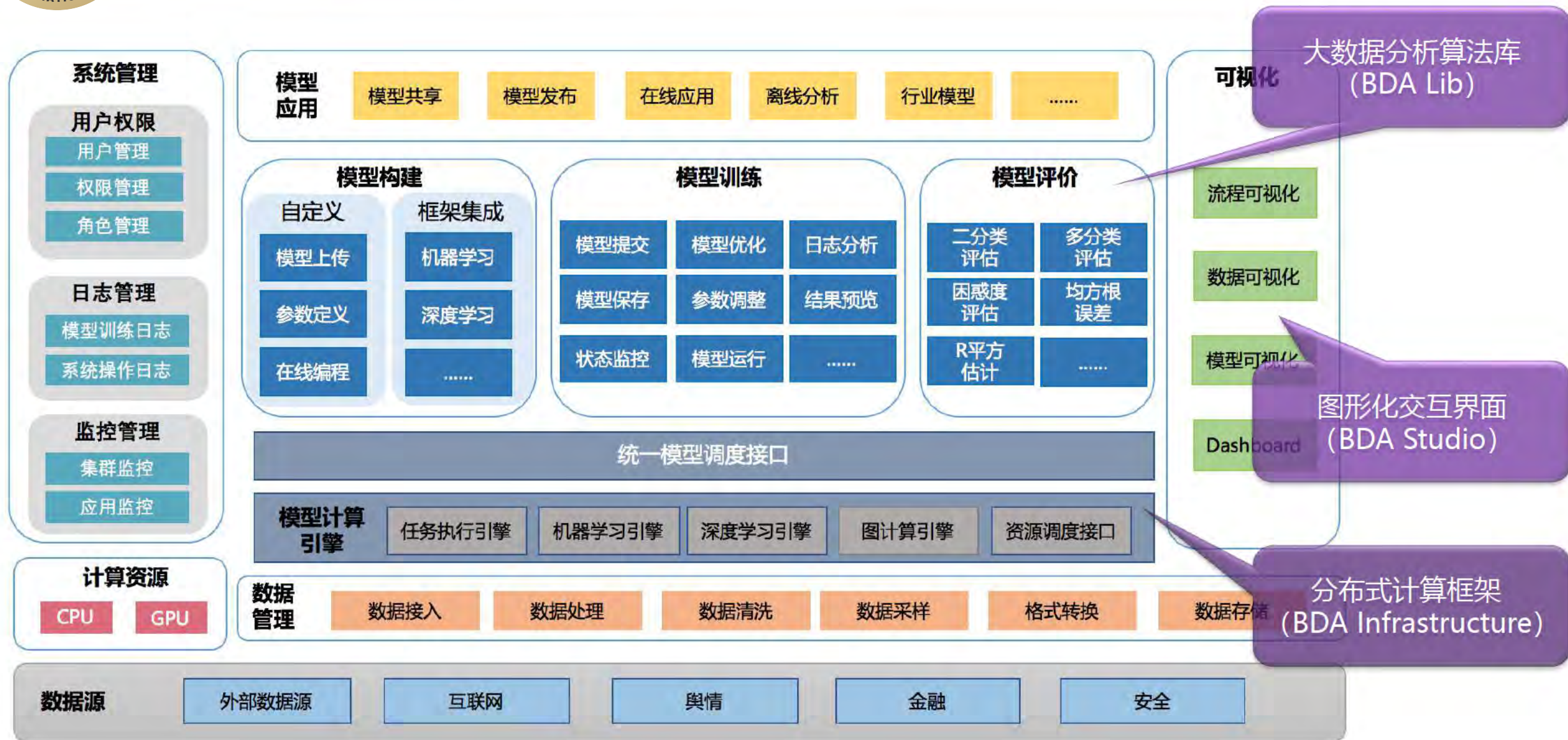
- 可视化智能机器学习平台是集数据预处理、模型训练、预测、评估为一体的**一站式大数据智能分析挖掘平台**，可帮助用户**快速构建**和**持续优化**大数据分析模型，挖掘大数据资产价值

- ✓ 基于全流程的可视化大数据分析平台，无需编程，快速实现大数据处理流程，降低使用门槛，提高工作效率
- ✓ 算法训练平台，迭代优化
- ✓ 针对不同业务场景复用和针对性优化算法模型
- ✓ 训练好的算法提交发布，形成服务





可视化智能机器学习系统 (BDA)



- 硬件：同品牌、同型号、同配置
- 软件：同操作系统、同JVM虚拟机、同系统账户体系

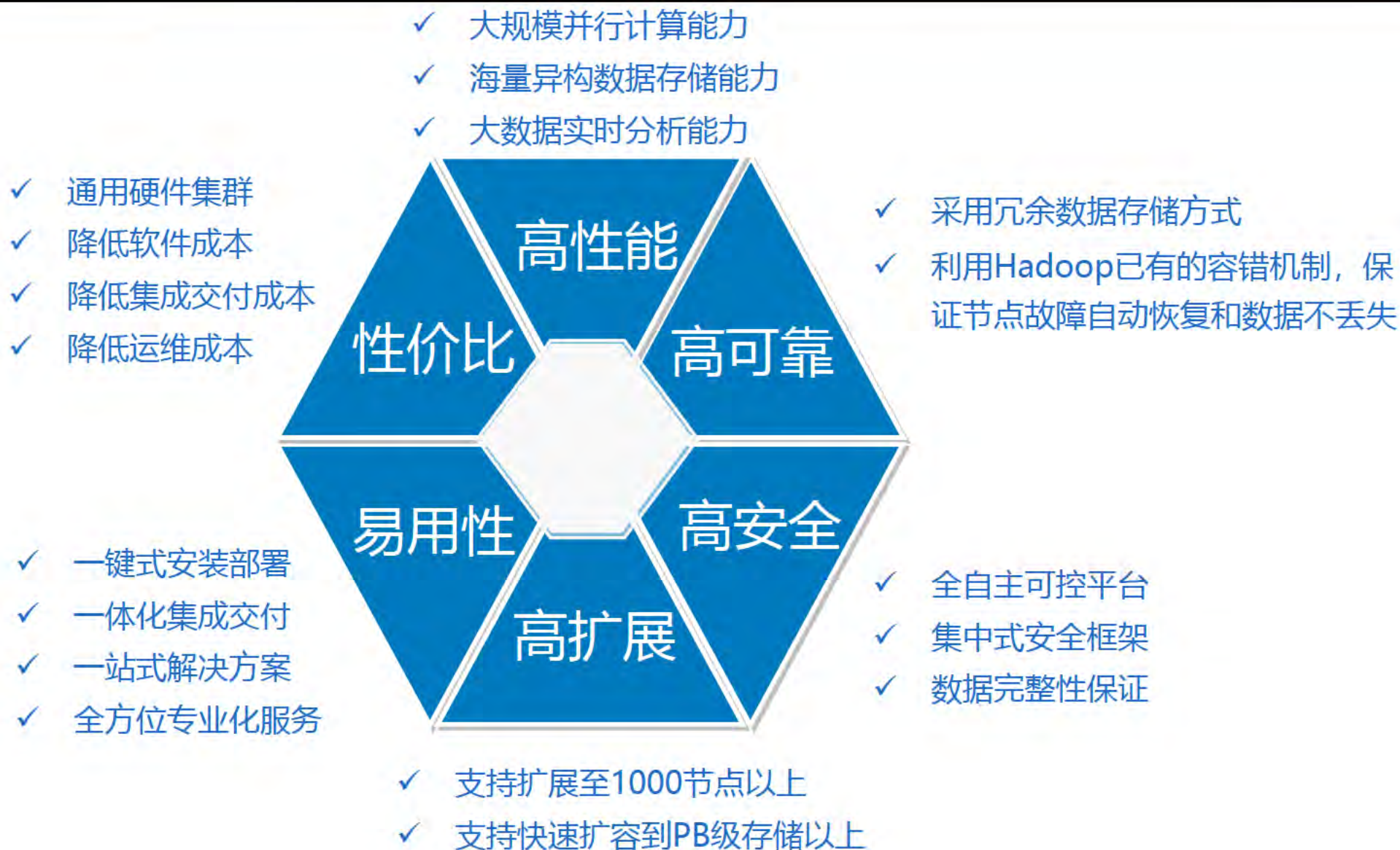


序号	分类	类型	技术规格	数量 (台套)	备注
自主可控大数据一体机 SDA60000L	硬件	服务器	机架式2U龙芯服务器，两颗龙芯3B3000系列处理器（4核 1.45GHz），内存64G，一块 960G SSD，六块2.5” 1200G/10000RPM/12Gb/SAS，一块Intel 网卡:PCI-E 双万兆光口网卡 半高（含模块），一块RAID卡:12G 3108芯片 八口 R0 1 5 6 10 50 60 2GB缓存 半高，2U 3.5寸8盘位类型，板载Intel 四千兆网口，最多支持4个PCIe槽，含导轨，建兴内置DVD光驱。1块2U 550W AC 白金电源(冗电源)，一套中标麒麟操作系统V7.0（服务器版）（试用版）。 服务：提供三年原厂服务，硬盘三年免回收服务。	6	长城超云国产龙芯服务器 R3210 L10
		业务交换机	业务接口：24个万兆光口数（SFP+），4个100G光口数（QSFP28） 管理接口：1 个RJ-45 端口，1 个RJ-45 Console 口，1 个USB 端口，1 个Mini USB Console口 MAC：支持128K MAC 表项 时延：<2us	1	迪普自主可控万兆交换机 LSW6600-24XGS4CQ-G
		管理交换机	业务接口：28 千兆电+4 千兆光（SFP）+4万兆光（SFP+）， 管理接口：1 个RJ-45 端口，1 个RJ-45 Console 口，1 个USB 端口，1 个Mini USB Console 口 扩展槽位：1个 MAC：支持64K MAC 表项，支持静态MAC、动态MAC、黑洞MAC、源MAC 地址过滤	1	迪普自主可控千兆交换机 LSW5662-28GT4GP4XGS-G
		KVM		1	国产KVM
		机柜		1	
	软件	操作系统	国产操作系统	6	中标麒麟服务器操作系统 v7.0
		数据库	国产数据库	1	达梦、金仓等
		中间件	国产中间件	1	中创、东方通、金蝶等
		大数据平台	国产大数据平台	6	中科天玑大数据平台 BDEv2.0
		Python、Java运行开发环境		6	操作系统自带
		浏览器		6	操作系统自带

- 硬件：同品牌、同型号、同配置
- 软件：同操作系统、同JVM虚拟机、同系统账户体系



序号	分类	类型	技术规格	数量 (台套)	备注
自主可控大数据一体机 SDA60000F	硬件	服务器	机架式2U飞腾服务器，一颗飞腾FT2000+（64核 2.2GHz），内存128G，一块 960G SSD，六块2.5” 1200G/10000RPM/12Gb/SAS，一块Intel 网卡:PCI-E 双万兆光口网卡半高（含模块），一块RAID卡:12G 3108芯片 八口 R0 1 5 6 10 50 60 2GB缓存 半高，2U 3.5寸8盘位类型，板载2个千兆网口，独立IPMI管理接口，含导轨，建兴内置DVD光驱。1块2U 550W AC 白金电源(冗电源)，一套银河麒麟V4.0服务器版(试用版)。 服务：提供三年原厂服务，硬盘三年免回收服务。	6	长城超云国产飞腾服务器 R2210 F11
		业务交换机	业务接口：24个万兆光口数（SFP+），4个100G光口数（QSFP28） 管理接口：1 个RJ-45 端口，1 个RJ-45 Console 口，1 个USB 端口，1 个Mini USB Console口 MAC：支持128K MAC 表项 时延：<2us	1	迪普自主可控万兆交换机 LSW6600-24XGS4CQ-G
		管理交换机	业务接口：28 千兆电+4 千兆光（SFP）+4万兆光（SFP+）， 管理接口：1 个RJ-45 端口，1 个RJ-45 Console 口，1 个USB 端口，1 个Mini USB Console 口 扩展槽位：1个 MAC：支持64K MAC 表项，支持静态MAC、动态MAC、黑洞MAC、源MAC 地址过滤	1	迪普自主可控千兆交换机 LSW5662-28GT4GP4XGS-G
		KVM		1	国产KVM
		机柜		1	
	软件	操作系统	国产操作系统	6	银河麒麟服务器操作系统 v4.0
		容器软件	国产容器软件	1	中科天玑、银河麒麟等
		数据库	国产数据库	1	达梦、金仓等
		中间件	国产中间件	1	中创、东方通、金蝶等
		大数据平台	国产大数据平台	6	中科天玑大数据平台 BDEv2.0
		Python、Java运行开发环境		6	操作系统自带
		浏览器		6	操作系统自带





通用组件产品



智能认知

- NLU/NLP
- 图像视频分析
- 智能搜索

知识计算

- 开放知识系统OKS

存储和算法引擎

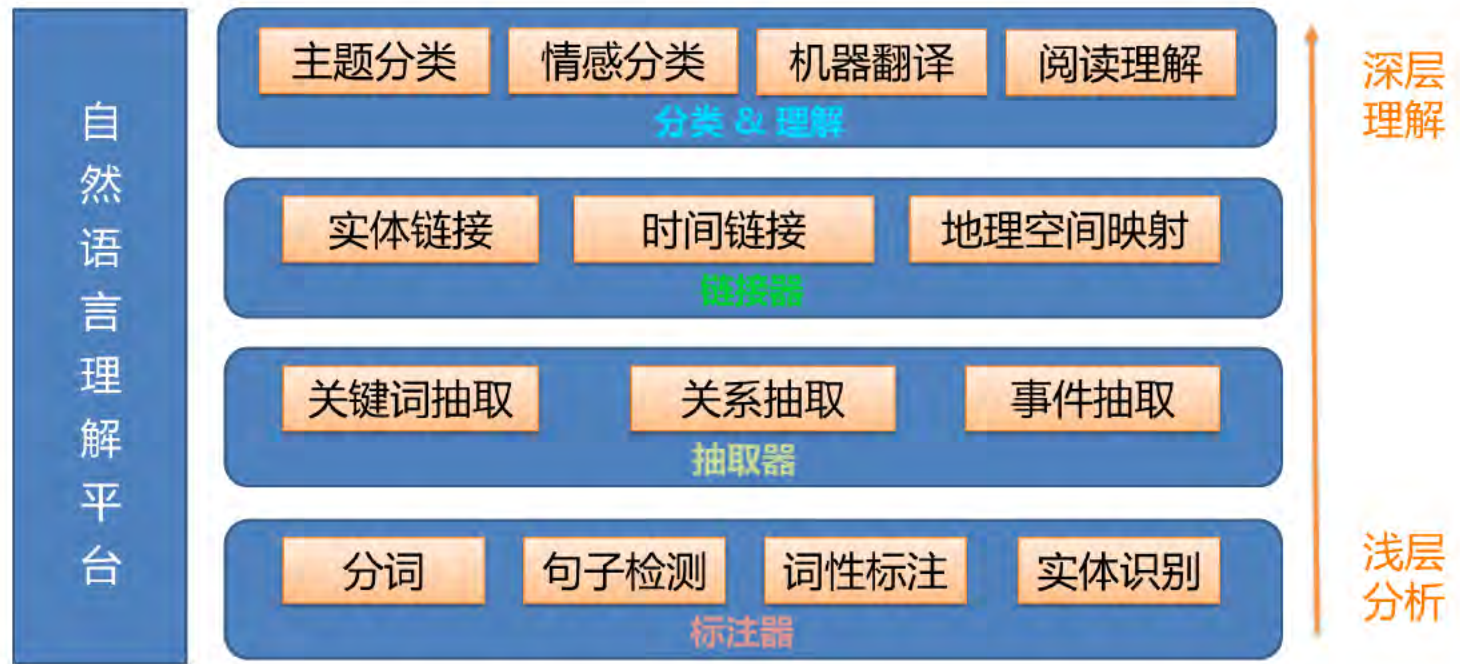
- SQLGraph
- 内存数据库
- 机器学习引擎

应用支撑组件

- 统一权限管理
- 可视化

• 自然语言处理数据处理分析组件，包括标注、抽取、链接、理解
个层次十余个自然语言模型组件

- ✓ 覆盖自然语言理解全体系
- ✓ 数据驱动+领域知识+机器学习
- ✓ 最前沿的深度学习模型、BERT-style预训练语言模型
- ✓ 通用/专业领域兼顾
- ✓ 支持不同数据格式
- ✓ 基础库/工具级产品/在线服务





自然语言理解组件 (NLU)



- 与学术界先进算法相比性能占优或相当

模型/数据集	词性标注	实体识别	主题分类	实体链接	事件抽取	机器翻译
	CoNLL 2003	CoNLL 2003	AG_News	ACE 2004	ACE 2005(trigger)	AI Challenge 2018
CRF (2015)	94.42	81.86	-	-	-	-
BiLSTM-CRF (2015)	95.15	84.18	-	-	-	-
AGDISTIS (2014)	-	-	-	66.4	-	-
NCEL (2018)	-	-	-	88.1	-	-
Char-CNN	-	-	0.8702	-	-	-
LEAM (2018)	-	-	0.9175	-	-	-
DMCNN (2016)	-	-	-	-	73.5	-
JMEE (2018)	-	-	-	-	75.9	-
GNMT+spm (2016)	-	-	-	-	-	25.22
天玑NLU	94.19	90.59	0.9206	87.3	78.4	24.56

- 在某军方项目POC测试中（参与方科大讯飞、拓尔思等）综合成绩第一，6项自然语言处理组件测试4项第一。

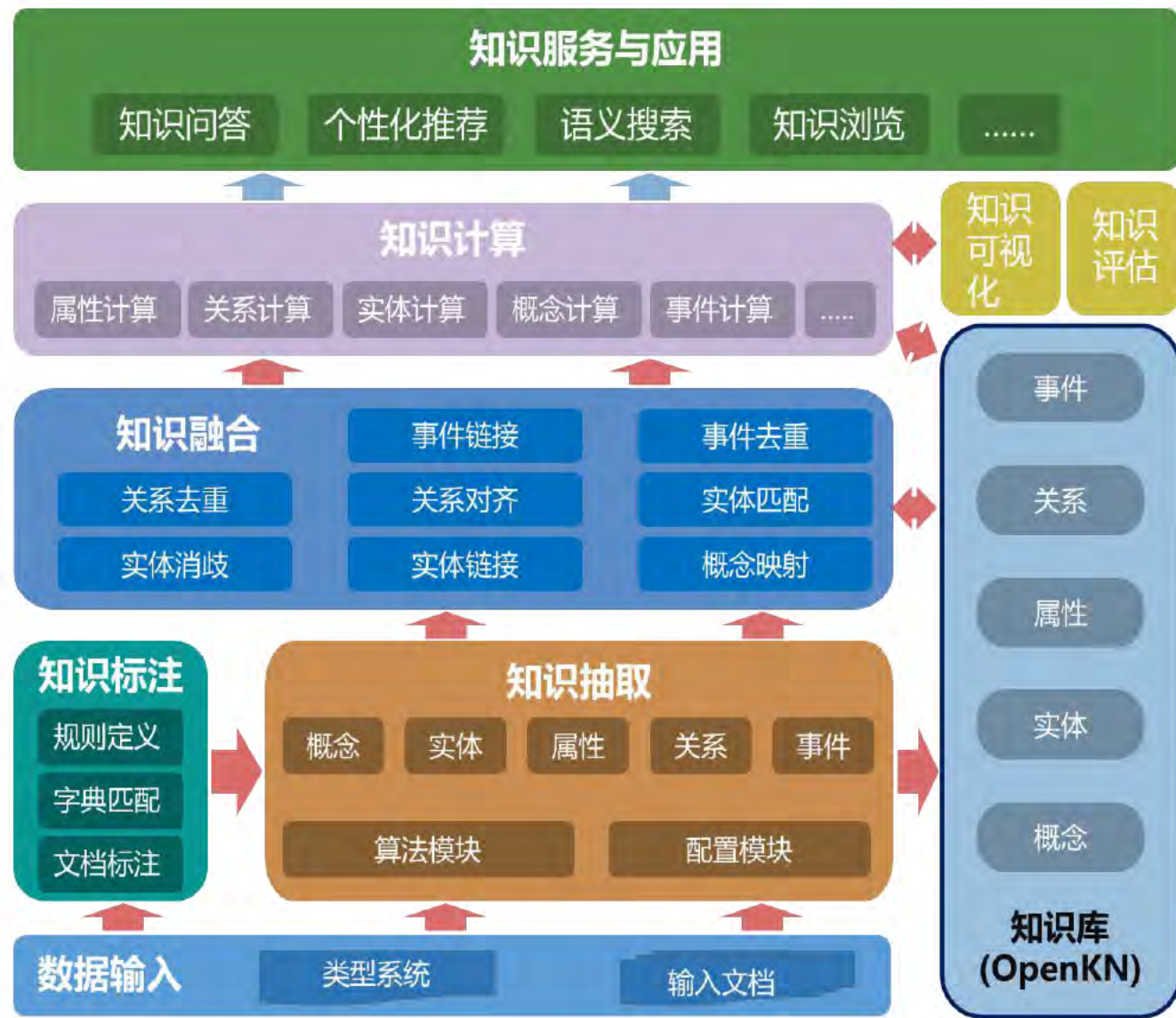
- 多年技术积累，在历次国际权威评测中名列前茅

关键技术	国际权威评测成绩
用户画像	SMP CUP2017评测排名第一
知识融合	TAC2016国际评测排名第一
知识发现	TAC2016国际评测单项排名第一
中文繁体纠错	SIGHAN2015国际评测排名第一
知识融合	TAC2014国际评测排名第一
知识发现	TAC2014国际评测单项排名第一
自动文摘	TREC2013国际评测排名并列第一
棒性检索技术	TREC2013国际评测排名并列第一
关联实体检索技术	TREC2009国际评测排名并列第一
动态文摘技术	TAC2008国际评测排名第一
企业专家检索技术	TREC2007国际评测排名第一
新信息发现技术	TREC2004国际评测排名第一
中文词法分析技术	SIGHAN2003国际评测排名第一



- 丰富的知识抽取算子库，灵活可定制
- 国际领先的关系推断算法
- 友好的知识标注工具
- 特色的行业知识图谱案例
- 支撑领域知识图谱的快速构建

实体抽取、实体链接、关系抽取
等多项关键技术**国际排名第一**





图计算引擎 (SQLGraph)



- SQLGraph图计算引擎是一项快速、高效的图分析数据库服务，可快速构建和高效分析高度关联的应用数据
- 可广泛应用于社交关系分析、推荐、精准营销、舆情及社会化聆听、信息传播、防欺诈等具有丰富关系数据的场景
- ✓ 直观的交互式建图，支持多源数据的导入
- ✓ 支持无向图、异构图、属性图，支持图的Schema设计，方便灵活的图属性查询、属性过滤
- ✓ 内置丰富的图查询、图挖掘算子库
- ✓ Python简易算法开发
- ✓ 接口灵活可定制，与SQL无缝对接
- ✓ 强大的图可视化展示
- ✓ 高速查询分析

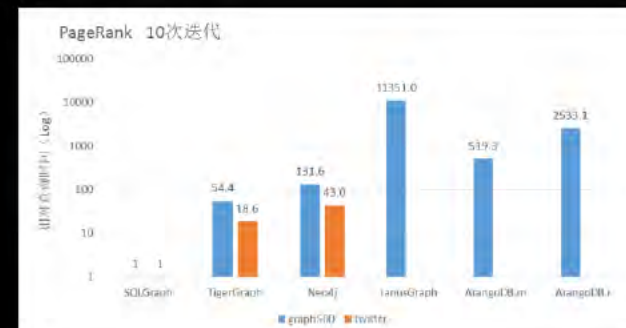
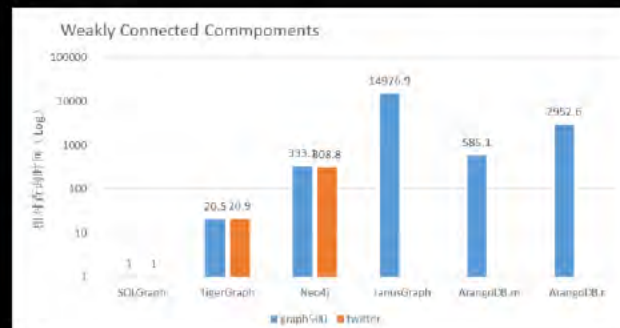
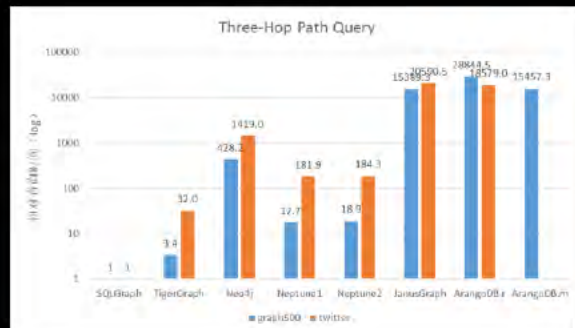
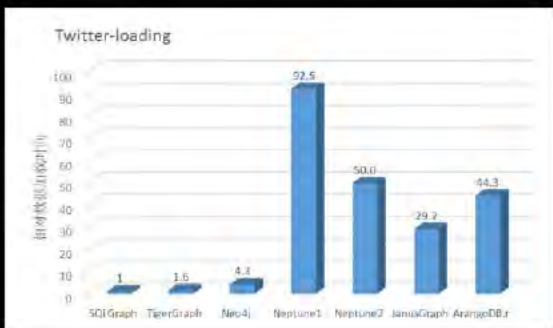




图计算引擎 (SQLGraph)



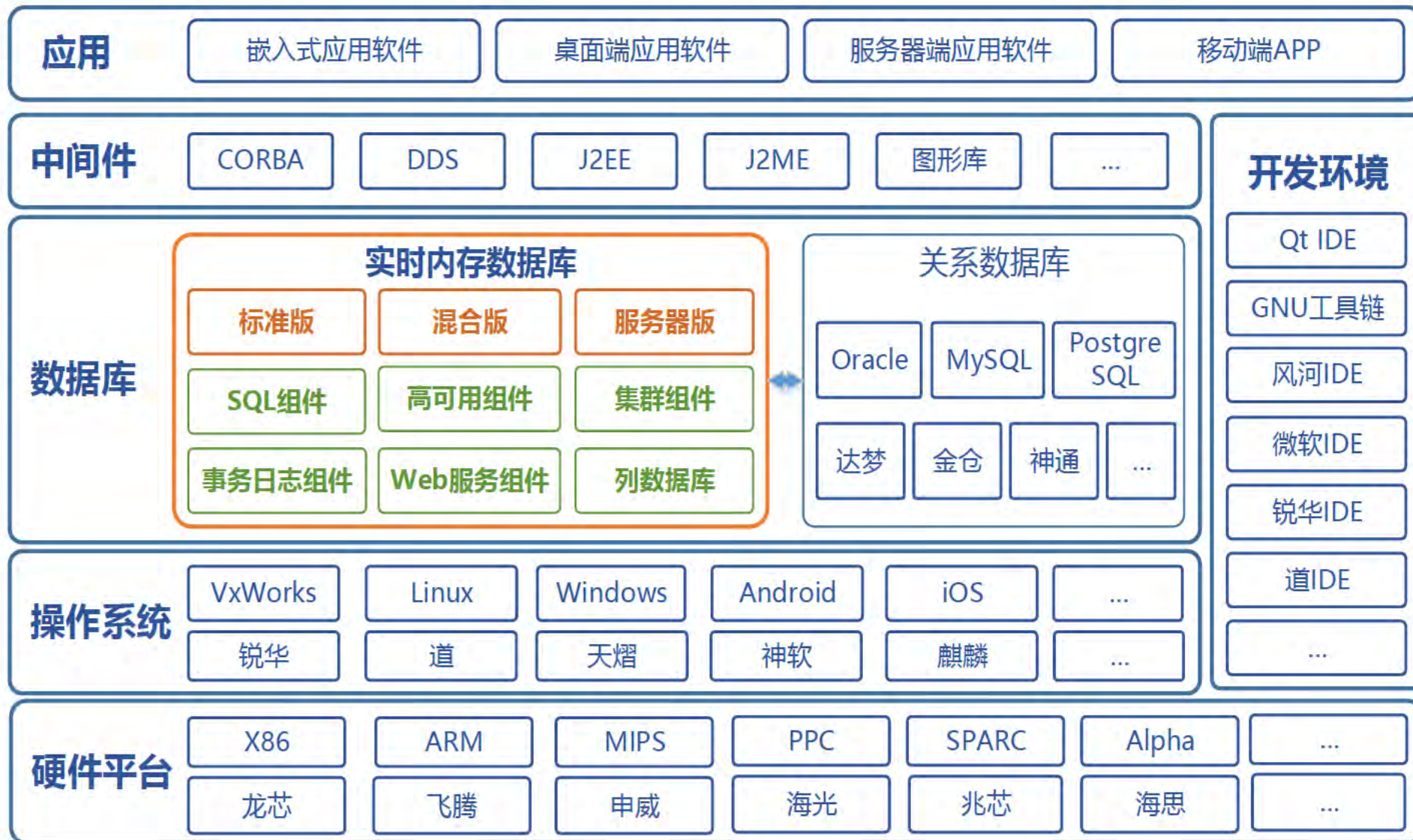
- 超高性能的关联查询分析
 - 支持**亿级节点、十亿级边**图计算的**秒级**响应
 - 亿级节点、边**导入图**十秒级**响应
- 国际主流产品对比
 - 与Neo4j、GraphX、TigerGraph等相比，图计算性能比快**10到1000倍**





内存数据库 (IMDB)

产品定位



开发环境

Qt IDE

GNU工具链

风河IDE

微软IDE

锐华IDE

道IDE

...



内存数据库 (IMDB)

体系结构

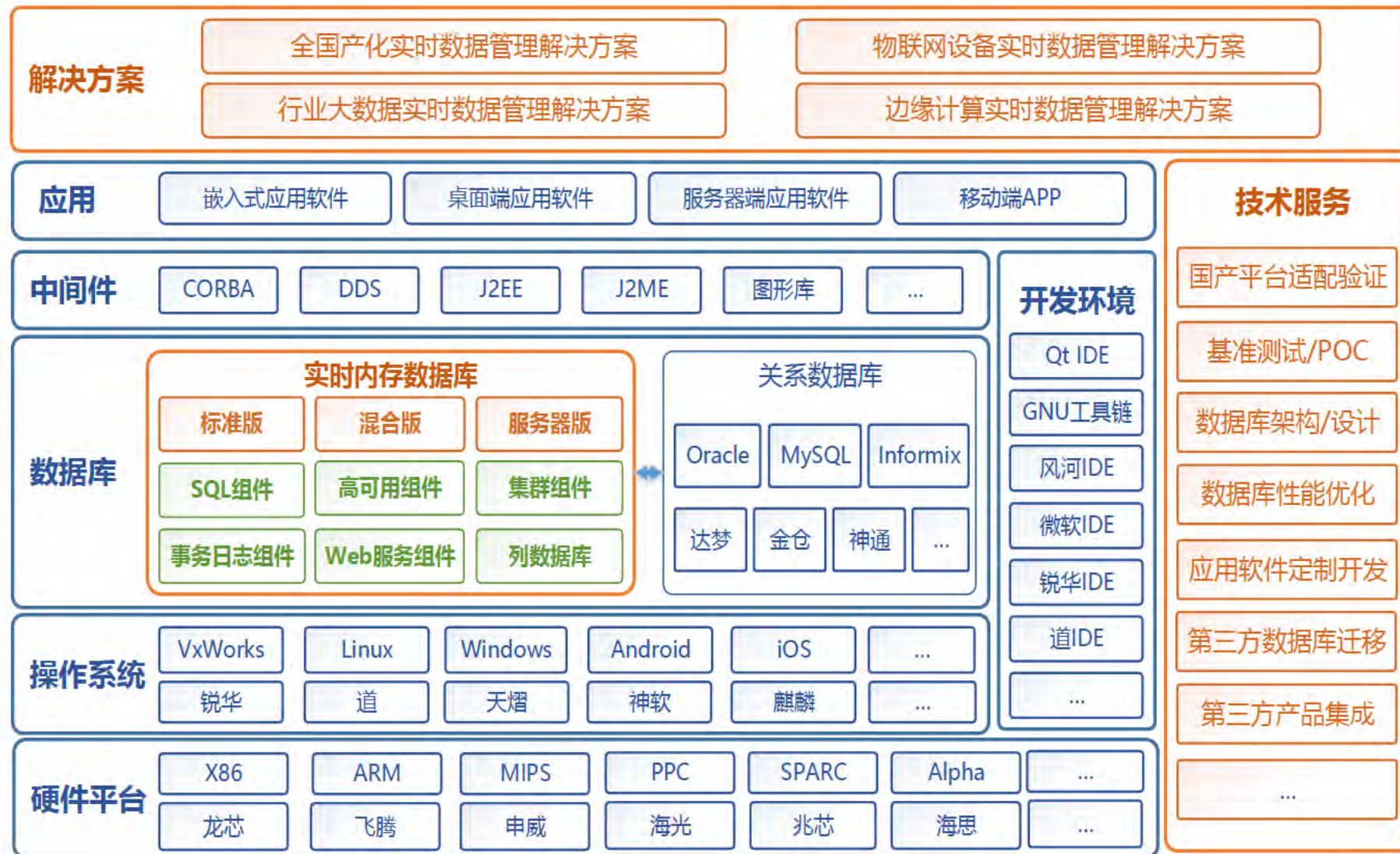




内存数据库 (IMDB)

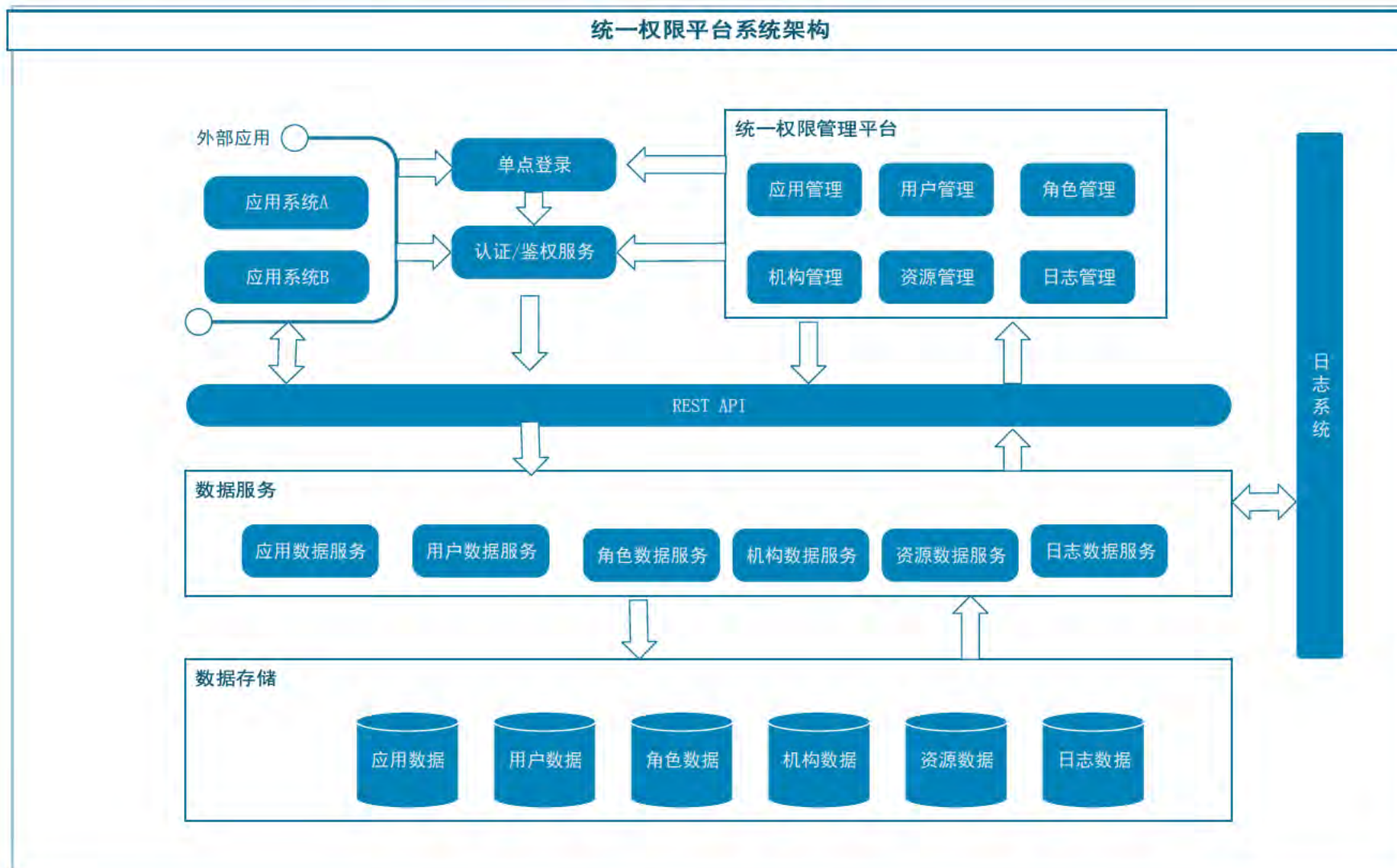


服务体系





应用组件产品-统一权限管理





PC端	移动端	大屏端	其他终端
-----	-----	-----	------

统一用户登录

业务层

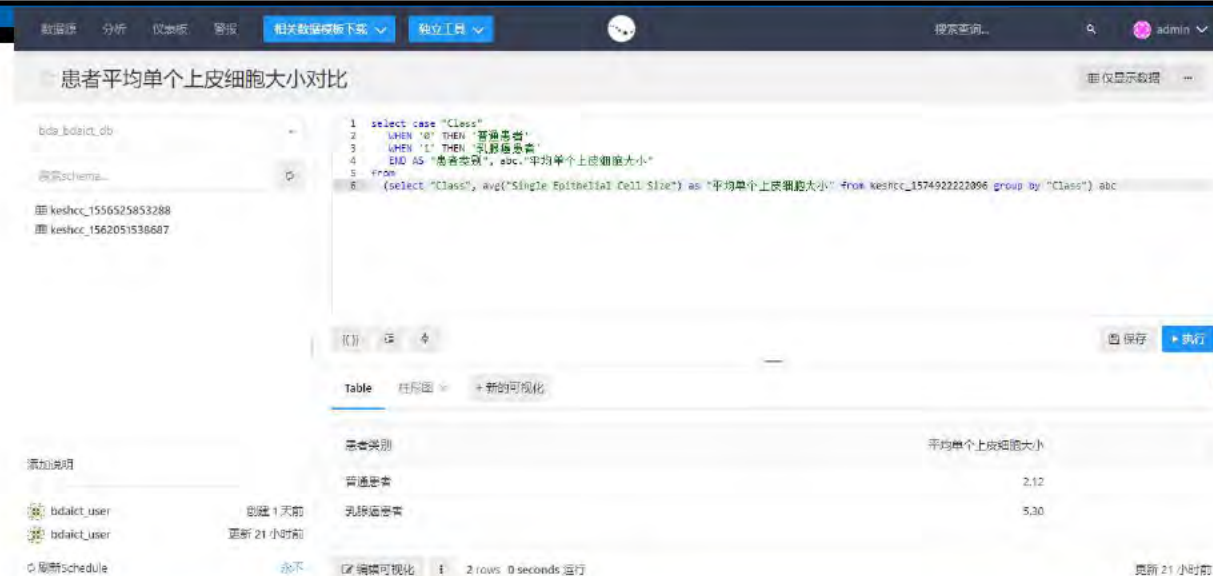
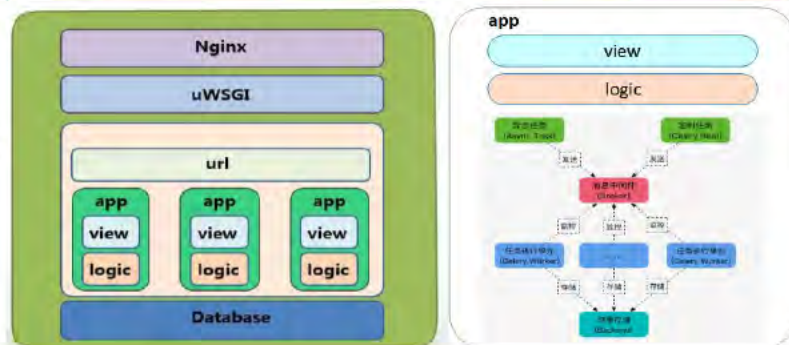
智能分析 实时大屏 应用统计 数据资产 智能运营

功能层

系统管理

组管理

支溝層





行业应用产品/解决方案

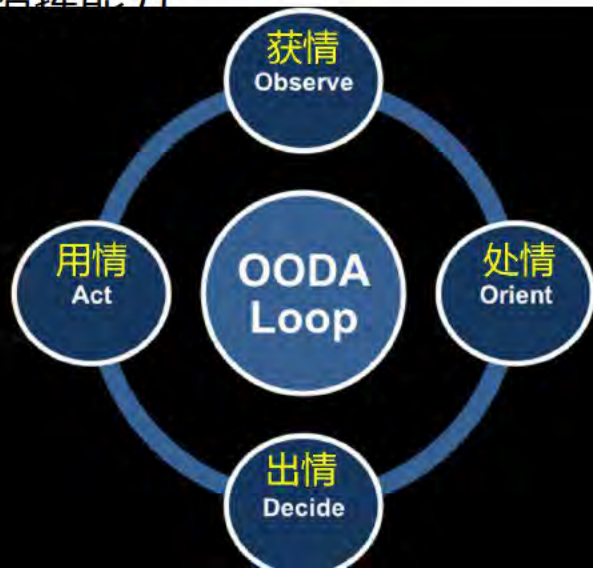




GoIN智能情报分析系统



以情报应用为主要抓手，通过对覆盖政治、军事、经济、社会等方面的大数据关联融合，构建贴近业务场景的分析挖掘、预测研判和智能生成系列工具，形成一套人脑决策和计算机智能共生的大数据分析环境及工具系统，辅助用户从大数据中萃取重大战略情报，并进一步转化为可操作的决策指挥能力



人机协同情报分析能力

GoIN
GoLaxy Intelligence

深度理解能力

认知套件

CoreNLU
CoreIMG
DeepAccess

关联融合能力

知识套件

OKS
OES

智能赋能能力

分析套件

SQLGraph
AdaxNet
BDA

大数据获取能力

XWDE

大数据管理能力

BDE

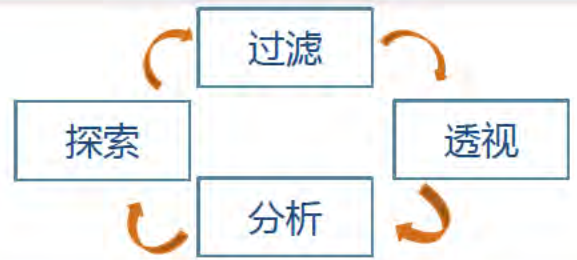


GoIN智能情报分析系统



交互式情报
分析引擎

User Interface



情报大数据融
合分析引擎

内容分析 (GA)

机器翻译 关键词 倾向性 文本摘要

关联分析 (NA)

节点重要度 社区发现 路径查找 子图挖掘

时空分析 (GA)

热力分析 区域探索 时间分析 轨迹分析

技术架构

基于动态知识结构的
数据融合技术

深度内容理解与解析技术

深度信息检索与获取技术

大数据弹性一体
化存储管理平台

Unified Storage APIs

MongoDB

SQLGraph

Elastic Search

Relational Database

Unstructured Data

Graphs

GIS

Files

数据接入
管理平台

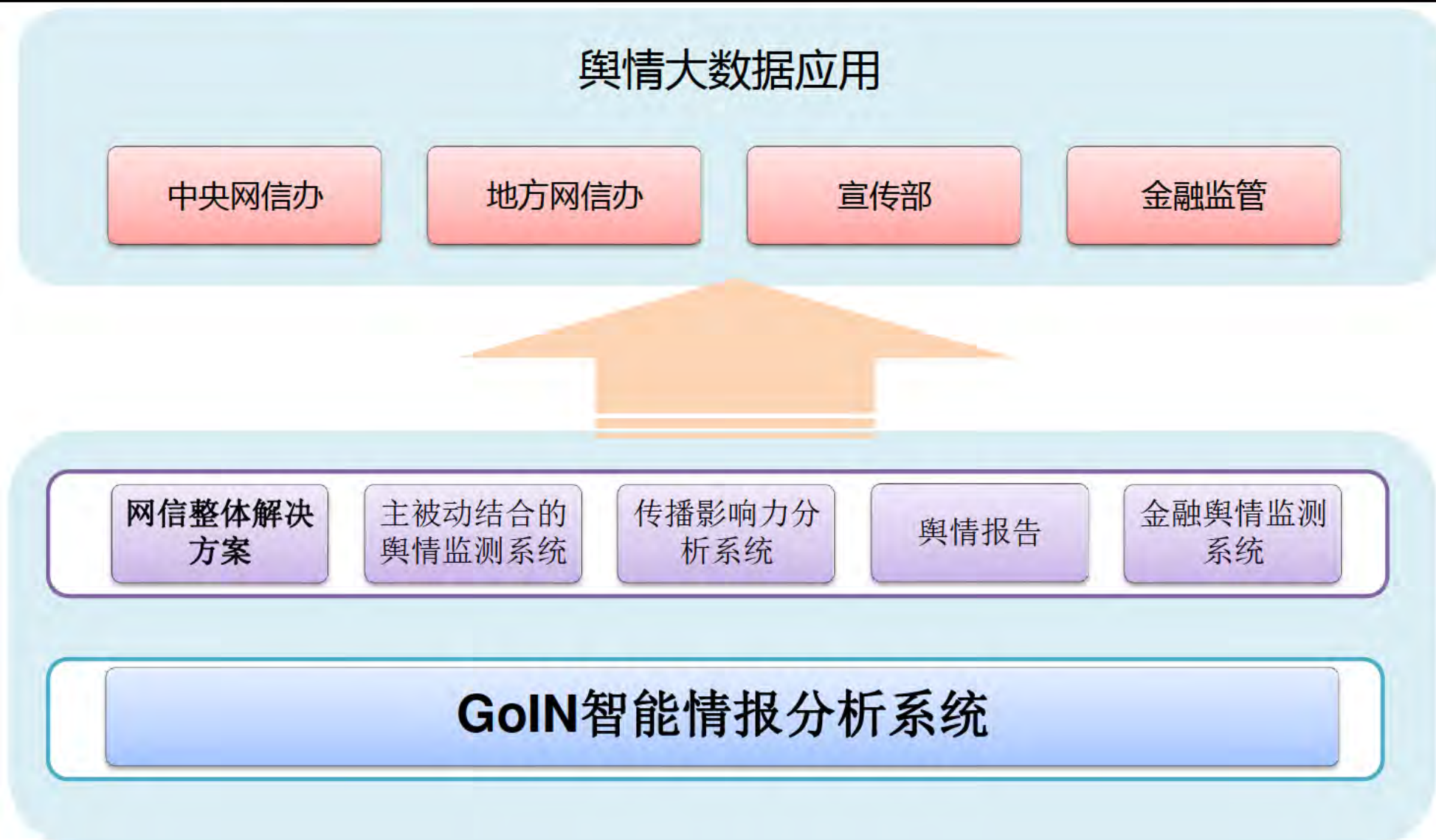
Unified Accessing APIs

元搜索

社交媒体

文件上传

配置采集





网信：整体解决方案（地方网信办）



1个体系

- 技术标准规范

2个平台

- 大数据平台
- 指挥平台

2大业务

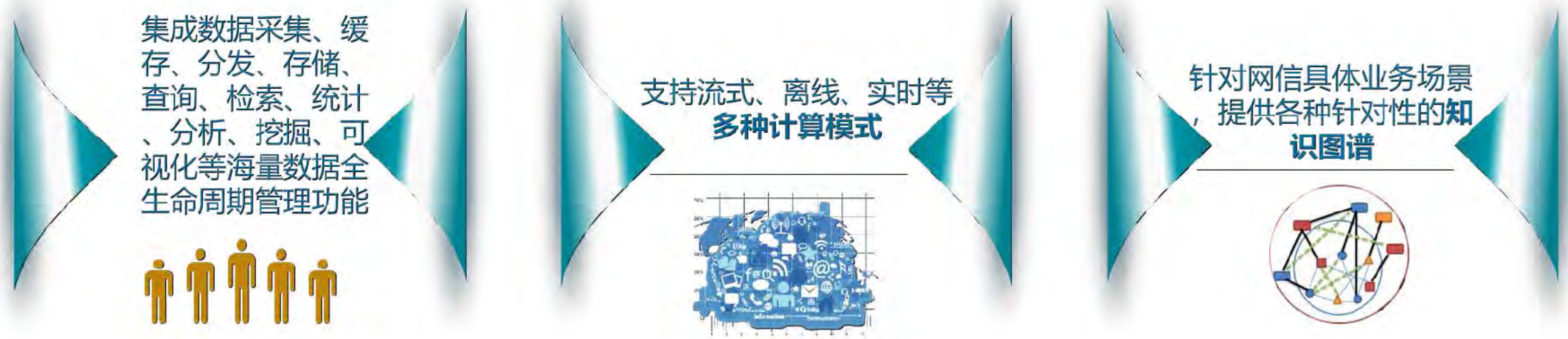
- 互联网信息监测
- 互联网宣传引导

解决方案	省网信平台	地市网信服务	政府舆情服务		
业务分析	舆情监测	属地化管理	新闻传播效果监测系统	取证	网评
	指挥通讯				
	网信知识库				
数据处理	网信大数据平台				
	数据采集				

- 以通用大数据平台+网信行业知识库，打造网信行业大数据平台，面向行业输出精准数据，面向客户输出定制化数据处理、基础性支撑平台能力

产品特色	关键指标	目标市场	对标竞品
- 行业知识闭环能力 - 行业数据分发能力	- 数据源覆盖及规模 - 行业知识库数量与精细度 - 数据标签准确性 - 行业知识来源	- 定制平台：部委/省级用户 - SAAS：基础性战略型产品	邦富/TRS底层数据标签

以**精准数据**、**爬取数据**、**推送数据**为依托，通过技术手段整合多种数据源，建设**统一网信高性能大数据平台**。实现**多源异构数据融合处理**、**分层分类存储**和**基于网信业务场景的分析计算**。



- 面向网信多级应急指挥协调的业务总线，实现省级网信办纵向指挥、横向协调的平台，通过业务系统对接，实现从信息发现、受理、研判、处置、落实、评估的业务闭环流程
- 要继续升级，充实省级网信办核心业务

产品特色	关键指标	目标市场	对标竞品
• 三级互通兼容 • 业务系统栈兼容（举报、舆情、传播监测与评估等）	• 上级指挥系统兼容性 • 业务系统兼容性 • 业务流程灵活性（权限、工作流可配置）	• 省级网信办	• 太极公司的指挥系统

- 形成上联中央、全省一体的指挥通讯，具备通信指挥、指令流转、指令效果评估等功能
- 实时安全、全面可靠、规范化、高可用性的中央、省级、地市三级指挥调度通讯



- 遵照中央网信办指挥通讯系统建设规范，稳定、可靠、安全的运行
- 包含指挥通讯系统专网端和互联网端，采用双网协同方式，保障指令传输安全可控和数据存储安全可靠

- 面向全通道舆情信息发现、预警、分析、追踪、出报的综合性舆情监测分析平台

产品特色	关键指标	目标市场	对标竞品
• 全通道数据融合 • 多角度舆情追踪（内容/目标/智能） • 精细化、智能化线索发现	• 通道数量及融合程度（全网溯源准确性、专题分析维度） • 舆情线索发现及时性、精准性	• 定制系统：部委/省级单位 • SAAS系统：地市单位	• TRS、云润（邦富）、智慧星光、人民舆情 • 微舆情、鹰击/鹰眼



网信：舆情监测分析系统



舆情系统是一款针对舆情发现、分析研判、处置流转全生命周期管理的数据分析服务平台。实现对电子报、新闻、论坛、博客、微博、微信、APP、知乎、twitter、facebook、网民评论等境内外全媒体数据的7*24小时实时采集、抽取、挖掘和分析处理，提供精准舆情搜索、热点舆情聚合、负面舆情巡查、专项舆情监测分析、高危/高价值舆情预警、智能舆情报告等服务。



• 面向属地监管单位的属地信息媒介监管产品

产品特色	关键指标	目标市场	对标竞品
• 属地化：范围、形态 • 智能化：智能违规判断	• 属地信息媒介库规模与精准性 • 数据全面性：属地信息媒介数据获取召回率 • 管理维度：内容合法性、资质合法性、媒介影响力 • 识别精准性：非法媒介、非法内容自动识别精准性	• 省级/地市管理单位	• 天润嘉能属地管理

• 面向宣传调控业务的调控方案管理、核查、评估综合系统

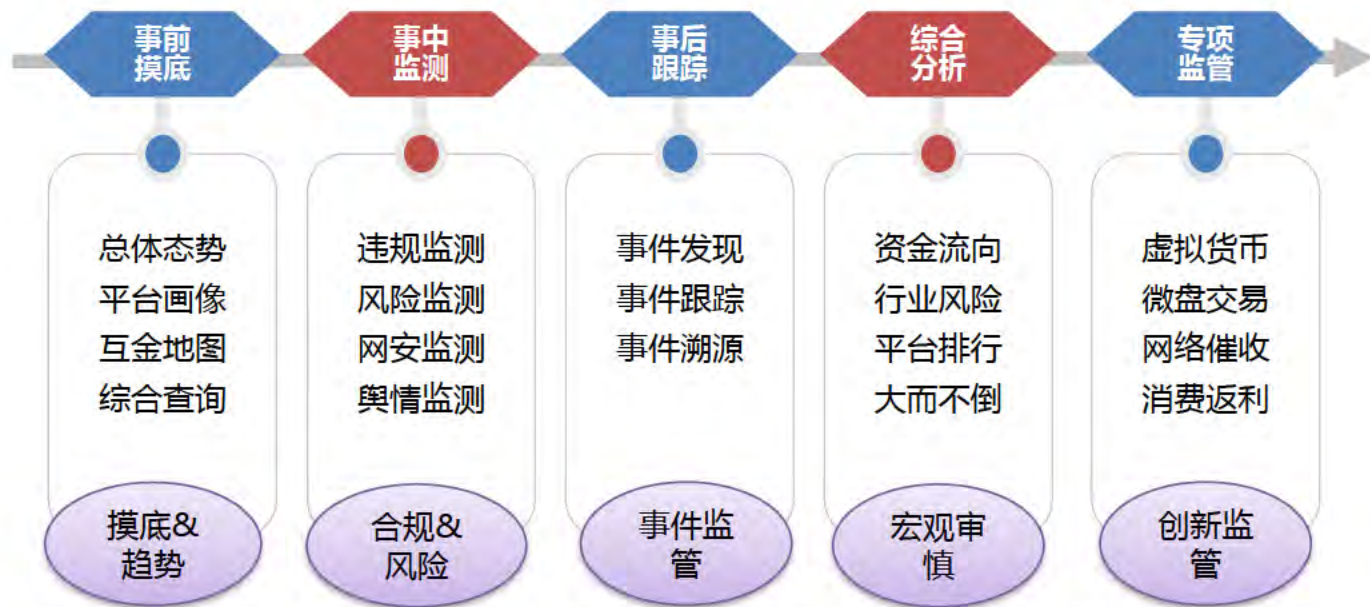
产品特色	关键指标	目标市场	对标竞品
- 精细化核查：属地精细化、核查维度精细化 - 综合性评估：影响力评估、绩效评估 - 与指挥系统衔接，快速部署	- 核查维度：范围、位置、时间、过程 - 评估可信度：指标体系、数据、可视化 - 定制化速度：核查范围、核查维度、评估维度	- 省级网信办 - 媒体发达的地市	人民网新闻指挥调度系统



互联网金融监管

- 技术平台实时汇聚、分析来自于工商、法院、税务、互联网等多渠道数据，支持多维度、多业态、多场景的对互联网金融平台进行摸底发现、合规判断、风险识别、事件追踪等。

- 全量工商、税务、法院、ICP备案库数据
- 亿级企业信息、5000万+高管信息
- 金融舆情：10万+版块，每天2000万+记录、30G+数据量





平台建设-技术体系



中科天玑联合中科院计算所在“互联网金融风险分析技术平台”的建设过程中承担了大量课题研究、算法建模以及系统开发工作：





教育科研大数据平台

各类高校的大数据相关专业

自2016年3月教育部批准设立大数据相关专业后，全国高等院校2956所，其中488所院校申请了大数据相关专业，未来还会有更多的院校申请建立这个学科。

使用大数据工具的学科实验室

在生物、医药、材料、金融、化学等需要大量统计分析数据的专业领域实验室，都越来越重视对大数据工具的掌握。

网络安全专业实验室

旨在为学校网络安全专业、社会网络安全评估机构、网络攻防测试机构、提供最新最全面的网络攻击手段以及防御能力评估。

大数据教学/实训



网络攻防实验

大数据科研实验



感谢聆听



应用场景：政治干预行动路径



社会科学
计算理论

基于选举舆论争夺的政治过程干预模型

基于3.5%法则的舆论引导政治干预模型



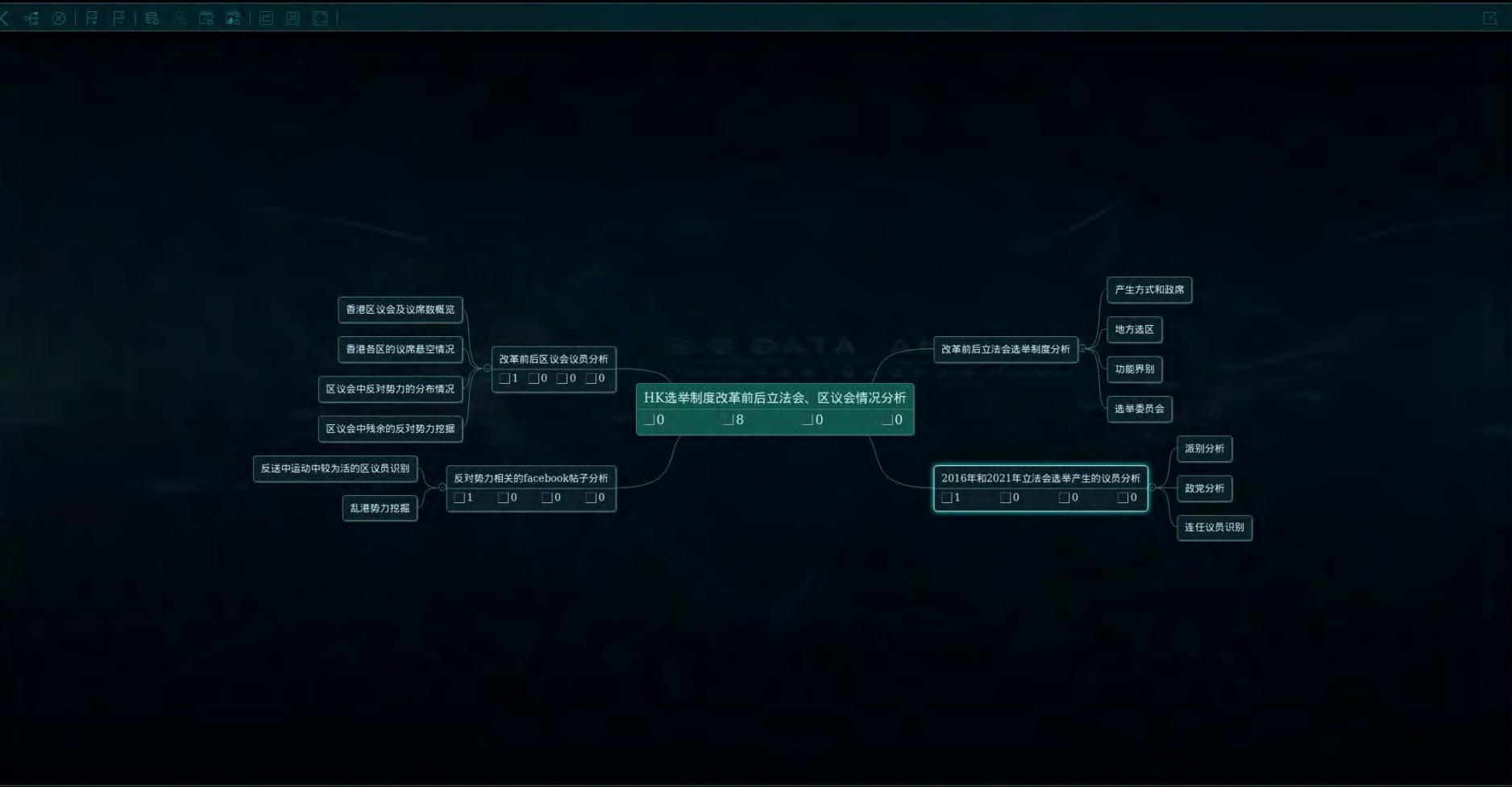
WestExec Advisors公司

本案例基于WestExec Advisors公司的官网数据及相关新闻，利用GoIN的智能分析和出报功能，尝试对该组织进行相对全面、深入的剖析，并生成报告。本案例除提...

图谱	状态	结果	人员
----	----	----	----



请选择左边节点，配置数据



2016年和2021年立法会选举产生的议员...
暂无描述

图谱	状态	结果	人员
----	----	----	----

改革前后立法会成员数据
香港选举制度改革前后立法会、区议会情况分析
创建时间：2023-07-04
来源：无结构化数据接入



活跃组织
暂无描述

图谱 状态 结果 人员

活跃组织

俾路支共和国和俾路支解放军最为活跃，俾...

保存时间：2023-08-14 16:36

来源：关联分析



国内外报道
暂无描述

图谱 状态 结果 人员

请选择左边节点，配置数据



2022年度部门任务目标及分解

序号	部门	指标名称		责任人	年度目标 (total)	Q1	Q2	Q3	Q4
1		经营业绩指标 (万元)	1.1新签有效合同						
			1.2净收入						
			1.3产品收入						
			1.4业务毛利						
			1.5行业利润						
2		重点工作任务项1:XXX	2.1xxx						
			2.2xxx						
			...						
3		重点工作任务项2:XXX	3.1						
			3.2						
			...						
4			4.1						
			4.2						
			...						



2023年度销售任务汇总表

单位：万元

部门	姓名	年度总任务 (total)		Q1		Q2		Q3		Q4	
安全	孔令茹	1,000.00	800.00	150.00	120.00	250.00	200.00	300.00	240.00	300.00	240.00
安全	侯阿奇	1,000.00	800.00	150.00	120.00	250.00	200.00	300.00	240.00	300.00	240.00
安全	待招1	1,000.00	800.00	150.00	120.00	250.00	200.00	300.00	240.00	300.00	240.00
安全	待招2	600.00	480.00	90.00	72.00	150.00	120.00	180.00	144.00	180.00	144.00
安全	待招3	600.00	480.00	90.00	72.00	150.00	120.00	180.00	144.00	180.00	144.00

合计

4,200.00

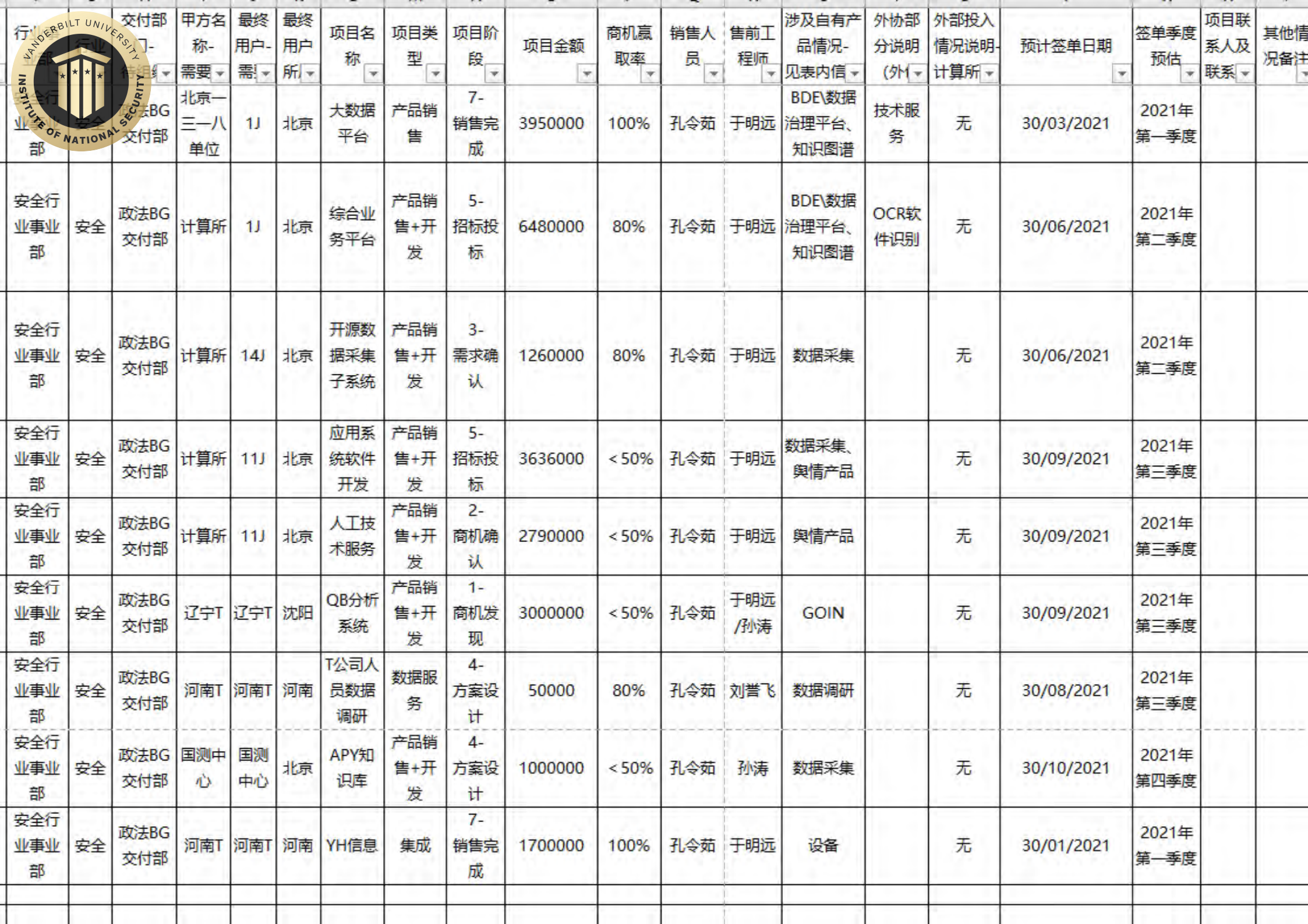
3,360.00

630.00

1,050.00

1,260.00

1,008.00



渠道	安全事业部	安全	待组织	需要	所	深圳	深圳局	数据专题分析	产品销售+开发	3-需求确认	3000000	50%	侯阿奇	于明远	无	无	无	30/11/2021	2021年第三季度	基本建设、维护、改造、服务器、预计、标流、委会、福州、,完、交付、上、质、办、以、现在、及到、系统、在、针对、增值、价5、九、大、体、案、,性、工、评、估、度、
渠道	安全事业部	安全	政法BG交付部	深圳局	深圳局	深圳局	广东厅	小蘑菇3.0	产品销售	1-商机发现	300000	70%	侯阿奇	于明远	暂无	无	无	30/09/2021	2021年第三季度	客户、造、求、更、求、维、护、改、造、务、器、预、计、标、流、委、会、福、州、,、完、交、付、上、报、质、,、办、公、以、后、现、在、及、到、系、统、在、真、针、对、增、值、价、5、九、大、体、案、,、性、工、评、估、度、
渠道	安全事业部	安全	政法BG交付部	四川厅	四川厅	四川厅	四川厅	数据治理	产品销售+开发	3-需求确认	4800000	50%	侯阿奇	于明远	暂无	服务器	无	25/11/2021	2021年第三季度	、
渠道	安全事业部	安全	政法BG交付部	福州局	福建	福建	福州局	桃T	数据分析服务	4-方案设计	60000	80%	侯阿奇	孙涛	暂无	无	无	25/03/2021	2021年第一季度	、
渠道	安全事业部	安全	政法BG交付部	甘肃省GA办	甘肃	甘肃	甘肃省GA办	舆情态势感知	产品销售+开发	2-商机确认	800000	50%	侯阿奇	于明远	暂无	无	无	26/12/2021	2021年第四季度	、
渠道	安全事业部	安全	政法BG交付部	西藏AQ厅	西藏	西藏	西藏AQ厅	邮件分析系统	产品销售+开发	2-商机确认	3000000	60%	侯阿奇	于明远	有	服务器	无	27/09/2021	2021年第三季度	、
渠道	安全事业部	安全	政法BG交付部	四川省AQ厅	四川	四川	四川省AQ厅	数据治理	产品销售+开发	1-商机发现	500000	<50%	侯阿奇	于明远	暂无	无	无	25/11/2021	2021年第四季度	、
渠道	安全事业部	安全	政法BG交付部	青海省公安厅	青海	青海	青海省公安厅	治安一体化防控	产品销售+开发	2-商机确认	10000000	<50%	侯阿奇	于明远/孙涛	暂无	无	无	25/11/2021	2021年第四季度	、



关键举措 (我们打算如何实现)		评价标准 (对E列目标的达成结果定义，	完成时间 (对应E列) 格式: 2020-12-31	负责人	Q1达成目标	Q1评价标准	Q2达成目标	Q2评价标准	目标
效签约额	项目1-填写具体项目名称			填写与项目主要责任人	填写与具体金额	可衡量的评价标准	填写与具体金额	可衡量的评价标准	可衡量的评价标准
	项目2								
	项目3								
收入	项目1								
	项目2								
	项目3								
品收入	项目1								
	项目2								
	项目3								
务毛利	项目1								
	项目2								
	项目3								
业利润	项目1								
	项目2								
	项目3								
	具体措施1				请填写与目标产出或效果				
	具体措施2								
	具体措施3								
	具体措施1								
	具体措施2								
	具体措施3								
		</							



评标结果确认函

中化商务有限公司:

项目编号: 0747-2060SCCZD820

项目名称: xx 实体抽取与检索技术 (二次招标)

我单位已审阅以上项目的评标报告, 同意以下评标结果:

中标候选人	投标总价 (人民币元)	综合得分	排名
中科天玑数据科技股份有限公司	252,000.00	82.54	1
北京中科浩晖科技股份有限公司	258,000.00	68.77	2

确认以此对外发布评标结果公示, 并在公示期结束后向排名第一的候选中标人发出中标通知书。

单位名称: 中国人民解放军 31040 部队

盖章/签字:

日期: 2020 年 11 月 9 日

会
会
会
参
林

课



版本: 1.0
密级: 内部

智能宣传系统（群体分析子系统） 概要设计说明书



	智能宣传系统（群体分析子系统）概要设计	1.0	1	

文档历史记录

日期	名称	版本	状态	签名
----	----	----	----	----

备注：状态表示为：1.新建 2.完成未审 3.完成已审 4.完成归档





文档修改记录

[目的与内容] 记录本文档的修改历史，包括版本、日期、变动内容和修改类型。

编号	版本	描述 (包括：日期，变更的图、表、段落号，标题或简单描述)	修改类型 A, M, D
1	1.0	初次编写文档	A
2			
3			
4			
5			
6			

备注：修改类型表示为： **A** - 增加 **M** - 修改 **D** - 删除





目录

1. 引言	5
1.1. 编写目的	5
1.2. 背景.....	5
1.3. 名词解释	5
2. 任务概述	6
2.4. 目标.....	6
2.5. 开发与运行环境.....	6
2.3. 需求概述	7
2.4. 条件与限制	8
3. 总体设计	9
3.1. 基本设计概念和处理流程	9
3.2. 结构.....	10
3.3. 其他需要说明的问题	10
4. 接口设计	11
5. 模块设计	12
5.1 群组账号列表.....	12
5.2 群体分析页面	12
5.3 个人分析页面.....	12
6. 数据库设计	13
7. 界面设计	15
8. 安全保密设计.....	19
9. 维护设计	20





1.引言

1.1.编写目的

该概要文档撰写的主要目的，是在需求设计的基础上，就该系统的接口、模块、库表、界面等几大部分进行深入分析和讨论并形成交互框架和视觉框架的过程，为用户研究和详细设计的无缝链接起到很好的桥梁作用，该文档预期的读者包括负责人员、开发人员、测试人员、运维人员，以及其他对该系统的开发和使用必须了解的人员。

1.2.背景

本概要设计所关联的系统是智能宣传系统，该任务是由智能宣传小组提出具体应用需求，并由王永庆老师负责需求、ui 组人员负责设计、组内开发人员负责开发、部署以及后期运维工作。

本系统为智能宣传的子系统，为辅助系统，属于智能宣传系统中的群体分析模块。本系统主要由刘金虎、史铂深二人负责。

1.3.名词解释

缩写、术语	解 释
社交网络（SNS）	Social Network Site，专指旨在帮助人们建立社会性网络的互联网应用服务。
JSON	Javascript Object Notation，一种轻量级的数据交换格式。
群体	不同个体按某种特征结合在一起，进行共同活动、相互交往，形成了群体，本系统中特指一系列有共同特征的社交网络帐号。
用户画像	本系统中指某个社交网络账号的用户的属性、行为与期待的数据转化集合。



2.任务概述

2.4.目标

本项目主要针对群体画像数据数据，实现群体分析功能。

2.5.开发与运行环境

2.5.1.开发环境的配置

类别	标准配置	最低配置
计算机硬件	16G 内存，4 核 2.2GHZ 以上，千兆网卡，500G	8G 内存，2 核 2.2GHZ 以上，千兆网卡，500G
软件	Windows、MAC os 、JDK、Tomcat、DT	Windows、JDK、Tomcat、DT
网络通信	TCP/IP，HTTP	TCP/IP，HTTP
其它	无	无

2.5.2.

2.5.2.运行环境的配置

类别	标准配置	最低配置
计算机硬件	64G 内存以上，8 核 2.2GHZ 以上，千兆网卡，1T	32G 内存，4 核 2.2GHZ 以上，千兆网卡，500G
软件	CentOS、JDK、Tomcat、DT	CentOS、JDK、Tomcat、DT
网络通信	TCP/IP，HTTP	TCP/IP，HTTP
其它	无	无



2.5.3.

2.5.3.测试环境的配置

类别	标准配置	最低配置
计算机硬件	16G 内存, 4 核 2.2GHZ 以上, 千兆网卡, 500G	8G 内存, 2 核 2.2GHZ 以上, 千兆网卡, 500G
软件	CentOS、JDK、Tomcat、DT	CentOS、JDK、Tomcat、DT
网络通信	TCP/IP, HTTP	TCP/IP, HTTP
其它	无	无

2.5.4.

2.3.需求概述

根据需求说明书，该系统的主要功能由以下几项组成：

1.群体创建

支持添加不同的分析选项，选择完成后向数据方发送群体分析创建的信息。通过 DT 异步返回群体数据。（分析选项是否持久化待定）

2、群体分析

群体分析模块分群体展示群体的相关数据，具体数据向数据方请求。

3、用户画像

单用户画像模块展示各个用户的相关数据，具体数据同样向数据放请求。



模块主要功能应用的场景以及输入和输出如下所示：

功能	场景	输入	输出
群体 创建	根据关键信息创建一个新群体	分析选项：关键词，重点用户 分析 起止时间 社交平台 地域	一个用户群体（通过 dt）
用户 列表	展示用户列表		
用户 画像	用户详情页		单个用户画像分析
群体 画像	群体详情页		单个群体画像分析

2.4.条件与限制

条件如下：

- 条件 1：模块所需要的数据包括：用户画像分析，用户详细数据 由实验室数据平台提供
- 条件 2：模块所需要的生产环境由实验室提供
- 条件 3：模块所需开发环境由业务人员自行选择
- 条件 4：开发环境和测试环境由业务人员自行调度



条件 5：系统试用期间，中心的业务人员能够投入一定比例的时间试用该系统，并就该系统存在的功能和业务问题给出反馈；

条件 6：开发期间若有业务上的问题，业务人员能够参与一些必要的讨论，并就问题能够给出一些建议；

条件 7：界面设计由组内人员讨论，ui 组实现具体设计

限制 1：进度限制，按照规定的日期进行开发；

限制 2：功能限制，参考需求和时间人员的投入来确定范围；

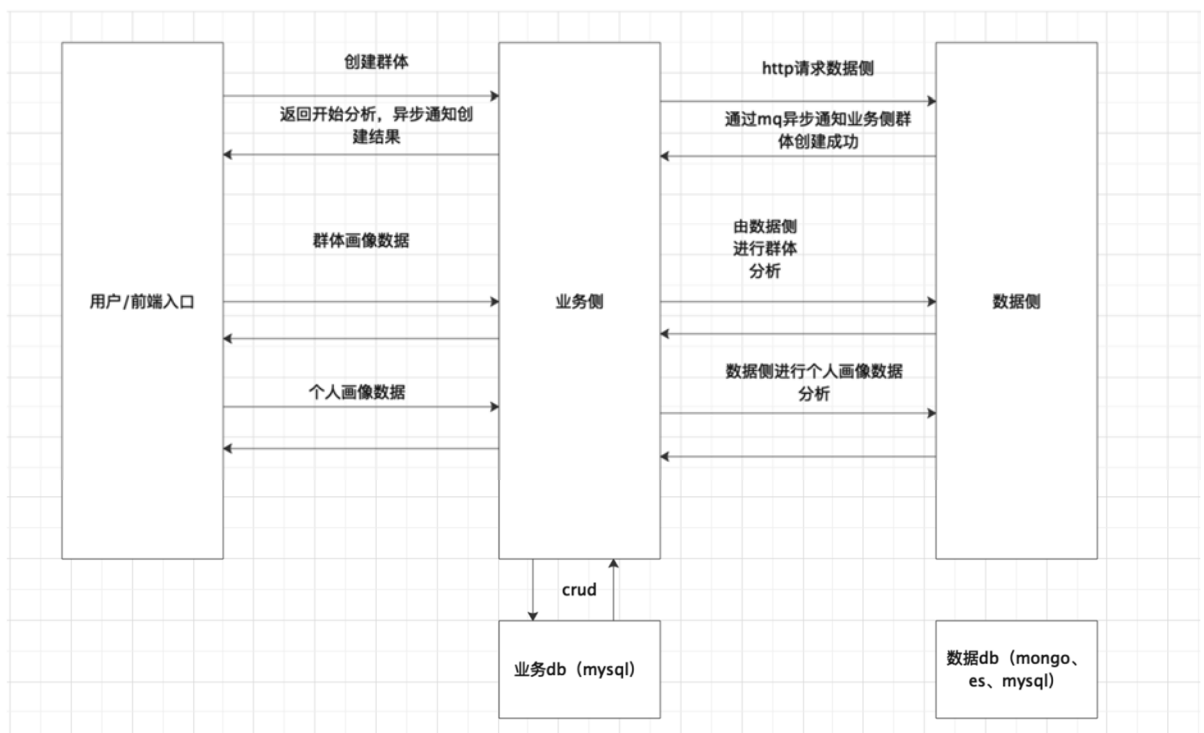
限制 3：设备限制，系统架构的设计应在计算和服务资源的限制范围内进行；



3.总体设计

3.1.基本设计概念和处理流程

群体分析子系统的整体架构如下：



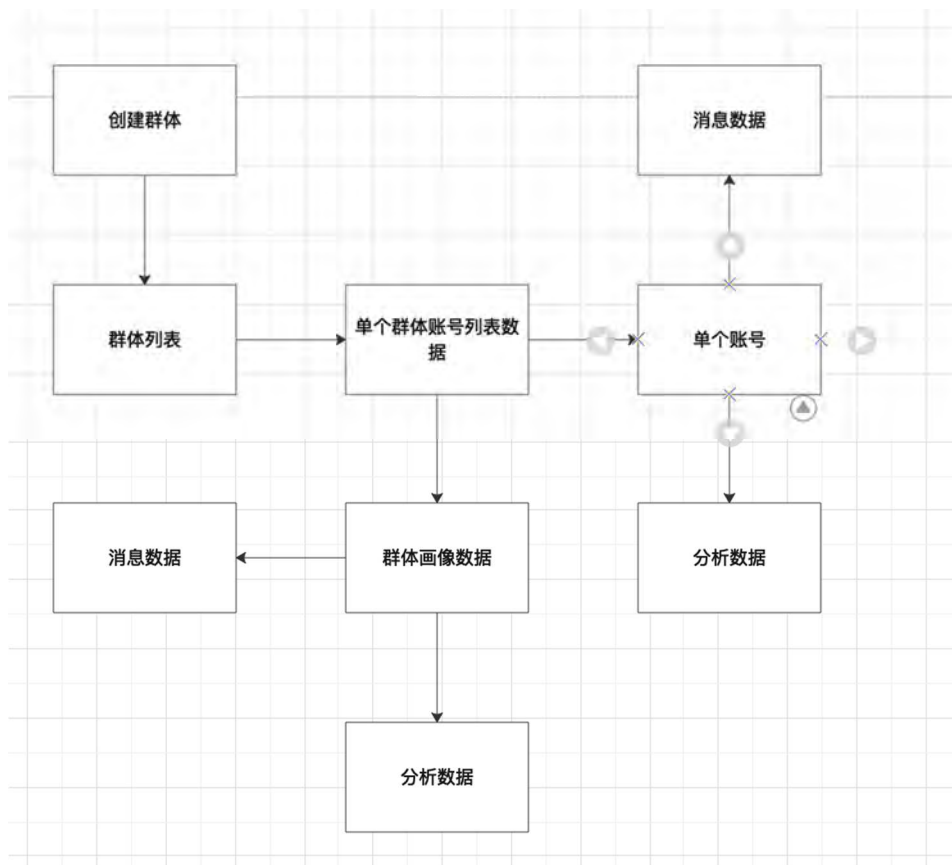
群体分析子处理流程如下：

- 1.由前端(用户)根据关键字开始群体分析，由业务侧告知数据侧
- 2.群体分析是一个由数据侧进行的耗时的操作，业务侧和数据侧通过 mq 异步通知，业务侧和前端可通过 socket（暂定）通知。
- 3.群体画像数据由前端请求，业务侧向数据侧请求数据并由数据响应时长等其他因素决定是否在业务 db 进行缓存。
- 4.个人画像数据由前端请求，业务侧向数据侧请求数据并由数据响应时长等其他因素决定是否在业务 db 进行缓存。



3.2.结构

群体分析系统结构如下：



说明：群体分析系统主要分为群体画像数据和个人画像数据，画像数据分为统计分析数据和消息数据。

3.3.其他需要说明的问题

无



4.接口设计

前端调用接口统一使用 restful api 操作，接口文档待与前端讨论后另出文件。



5.模块设计

5.1 群组账号列表

群体账号列表由数据侧在创建群体后将各个账号的基本信息返回，存入业务侧的数据库。业务侧维护账号和群体的映射关系。基本信息包括（粉丝数，关注数，消息数等）。

5.2 群体分析页面

群体分析页面的数据由数据侧提供，并根据响应时间等其他因素确定是否由业务侧进行缓存，并日更新。群体分析数据包括群体活跃度分析，关键词和话题等数据。

5.3 个人分析页面

个人分析页面包括基础数据和统计分析数据，基础数据由业务侧直接提供，统计数据则实时访问数据侧获取。个人基础数据包括粉丝数，关注数，发帖数等。统计数据包括，活跃度分析，关键词，话题等。并根据不同的目标账号类别进行话题粒度的区分，普通账号不进行话题粒度的区分，媒体账号和重点账号进行话题粒度的区分。



6.数据库设计

表名: target_account_basic_info

字段名	类型	说明
target_account_id	String	账号 id uid
user_name	String	用户显示名称
screen_name	String	@后的名称
real_name	String	真实姓名
gender	tinyint	性别-1 无效 0 女 1 男
lurl	String	头像地址
profile_url	String	头像地址
description	String	描述
create_time	Int	用户创建时间戳
gtime	Int	爬取时间戳
Is_verified	tinyint	是否认证 0 普通用户,1vip 用户 -1 未知
verify_info	String	认证信息
Is_real	tinyint	是否实名认证 (1:实名; 0:非 实名,-1:未知)
home_url	String	用户主页
pburl	String	横幅
banner_url	String	横幅
location	String	用户所在地
email	string	用户邮箱
birthday	String	用户生日
platform	String	平台
Update_time	Int	更新时间戳

表名: target_account_statistic_info

名称	类型	说明
target_account_id	String	用户 id



follow_count	Int	关注数
fans_count	Int	粉丝数
status_count	Int	发文数
forward_count	Int	转发数
comment_count	Int	评论数
yes_follow_count	Int	昨天 关注数
yes_fans_count	Int	昨天 粉丝数
yes_status_count	Int	昨天 发文数
yes_forward_count	Int	昨天 转发数
yes_comment_count	Int	昨天 评论数



7.界面设计

页面1：新建群体

群组名称*:

分析选项

关键词: (输入框形式, 具备自动提示功能)

重点用户: (输入框形式, 具备自动提示功能)

分析起止时间*: (输入框形式, 日期插件)

社交平台*: (多选框形式)

地域*: (输入框形式, 输入时自动提示)

是否需要提供扩展分析: ☐ 是 ☐ 否

>>返回

开始分析

页面2：账号列表（此处功能同上期账号列表页）

监测列表

账号列表

群组分析

发布推文

微博

豆瓣

知乎

贴吧

论坛

博客

新闻

视频

音乐

游戏

体育

时尚

科技

教育

健康

生活

娱乐

财经

军事

历史

文化

艺术

宗教

政治

法律

医学

农业

工业

商业

交通

旅游

环保

公益

慈善

教育

健康

生活

娱乐

财经

军事

历史

文化

艺术

宗教

政治

法律

医学

农业

工业

商业

交通

旅游

环保

公益

慈善

监测范围

账号选择

评分

导出

删除

账号名称

粉丝数

活跃度

评分

操作

微博

8368

301

223

11

8

操作

微博

340

188

148

0

8

操作

微博

2546

157

135

20

2

操作

微博

25324

118

118

0

9

操作

微博

32964

107

107

0

0

操作

微博

1107

104

84

20

0

操作

微博

20180

72

67

5

0

操作

微博

3582

55

55

0

0

操作

微博

1718

56

55

1

0

操作

微博

1559

53

53

0

0

操作



页面3：群组分析（此处功能分块同一期群组分析页，具体信息展示参照下图）



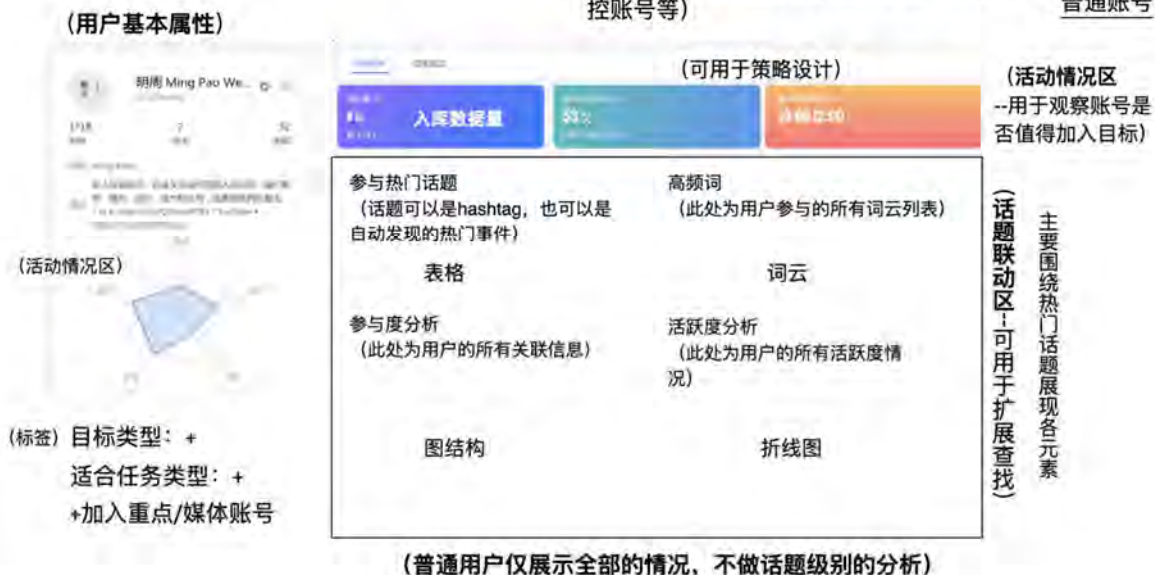
(河图, 选取热门的top事件的群体交互活跃度曲线
(另要以事件的交互次数为限制--提升展示效果))

页面4：相关贴文（一期发布贴文页删除，改为点击用户高频词、热门话题后的相关贴文展示）



页面5-1：用户画像（点击账号链接进入）

设计思路：通过对当前人物的观察，将用户加入目标列表（不同账号的显示方式应有区分，例如：XX媒体、被控账号等）
普通账号



页面5-2：重点账号画像

设计思路：通过对当前人物的观察，将用户加入目标列表（不同账号的显示方式应有区分，例如：XX媒体、被控账号等）
重点账号



页面5-3：媒体账号画像

设计思路：通过对当前人物的观察，将用户加入目标列表（不同账号的显示方式应有区分，例如：XX媒体、被控账号等）

媒体账号



页面5-4：相关贴文列表





8.安全保密设计

- 1.本系统所使用的安全模块为天玑公司提供，安全性由天玑公司保障。
- 2.任何账号由产品/项目经理所保管，不得外泄。
- 3.硬件服务器由实验室提供。



9.维护设计

1. 按照维护计划中要求，对本系统主要软件、硬件设备进行经常性地维护与监视，保证其正常工作。
2. 由专门维护人员进行周期性的维护，并如实记录维护报告与修理报告。
3. 按照维护计划中的要求，定期对系统设备进行检修、维护，保证系统满足运行要求。
4. 提交维护报告与修理报告，以便客户登录系统的日常工作状况，报告系统维护及修理情况。



编号:

密级

“××××工程”4月进展报告

1. 当前工作进展

围绕“战、建、备”协同推进的总体思想，4月份全面启动业务需求调研工作，并根据调研情况反馈同步推进工程建设工作。此外，由于工程行政总体变更为由正宫部主管，为配合工程总体调整，启动项目总体建设实施方案以及抓建模式调研工作，并初步形成总体论证调整思路。具体开展工作包括：

(1) 业务需求调研。明确沈华伟、王永庆、王石、张瑾、邱强、李蕾为工程展场核心人员，持续赴福州 61716 不对开展业务需求调研。作为工程理论总体，在全面了解业务单位的工程建设目标的基础上，总结理论瓶颈、梳理理论体系，与业务单位就任务设置、执行路径、成果形式等展开深入讨论，初步确定基础理论创新群的 2023 年度任务清单。

(2) 配合调研工作。为配合正宫部的调研活动开展，分别于福州、北京两地开展调研活动，就总体推进过程中如何理论先行牵引工程建设任务的方式及关键问题进行讨论。在计算所调研期



间，陈熙霖所长介绍所在认知方向上的战略布局，强调在工程建设上计算所不计成本不计投入优先保障。宣传局李军局长对计算所的技术积累及工作进展表示肯定。

(3) 总体论证优化。调研活动结束后，正宫部着手组织总体论证优化工作。计算所委派沈华伟、张浩、王永庆、王石、张瑾、邱强加入论证组，完善指挥决策与基础理论部分的均势能力需求、建设目标、建设内容与任务安排。截至月底前，根据国家宗提展的任务要求已形成能力需求、建设目标和建设内容初稿，并对基础理论的抓建模式给出建设性意见。

(4) 工程建设推进。跟随业务调研工作进展，计算所数据智能系统研究中心、网络数据科学与技术重点实验室、智能信息处理重点实验室、前瞻研究实验室 4 中心协调联动，于社交基期刃调度、指挥决策、内容智能生成、事件分析等方面，开展需求设计、技术方案设计和原型系统设计工作，相关方案已形成技术实施方案并入工程总体建设目标。

2. 下一阶段工作考虑

围绕持续推进的工程总体论证工作以及初步确定的理论研究体系和年度工作任务清单，下一阶段工作主要分以下几个部分开展：



(1) 总体论证优化。持续跟进，对指挥决策、理论体系以及理论部分的抓建模式进行优化调整，以理论牵引为目标，进一步明确理论工作的 5 年期任务开展方式及实施路径。

(2) 业务需求调研。在已初步明确的 2023 年任务清单基础上，进一步明确任务要求。相关已具备研究条件的课题内容，优先联合优势单位开展合作研究。对未具备研究条件的课题内容，调研工程需求，形成指南建议。

(3) 工程建设推进。瞄准 7 月沿袭任务，组建优势力量，重点就数据采集、内容生成、社交基期刃调度等系统建设形成建设方案，全面启动建设任务。



脚本文字稿

一、开场白

非常感谢院领导莅临我所检阅和指导工作，我谨代表计算所网络数据科学与技术重点实验室向各位领导汇报“天玑智能宣传系统 GoPro”。

二、智能宣传研究背景

近年来，国际形势风云变幻，暗潮涌动，以美国为代表的西方国家频频借香港问题、新疆问题、新冠病毒、贸易战在网络空间向我发动攻势强大的舆论战，严重威胁我国意识形态安全和政权安全。面对新形势新任务，习近平总书记多次指出“要增强阵地意识，敢于发声、敢于亮剑，尽快掌握网络舆论场上的主动权，不能被边缘化了。”因此，创新利用新兴网络技术，研发先进的技术系统，改变我国在网络舆论斗争的“守势”和“弱势”现状，弥补网络信息流“逆差”、形象“反差”、实力“落差”，成为科研工作者需要解决的时代命题。

面向国家重大需求，我所将智能宣传列为所内 8 大优先支持发展方向，并在网数实验室成立专门的系统部，由 2 名正高、5 名副高等 50 余人组成的研究团队，在尚未有国家相关项目和经费的支持下，部署并开展学科方向论证、关键



技术攻关、演示验证系统研发等工作，下面向各位领导详细介绍一下智能宣传系统。

三、智能宣传系统介绍

系统着着眼于新时期网络舆论引导和认知对抗的迫切需求，创新利用人工智能、大数据、社会心理、信息传播等技术成果，研究网络舆情智能传播、引导与利用技术，研发面向境外非受控环境的智能宣传系统，具备如下四大能力：

一是特定区域政治态势的智能感知。围绕当前中美博弈、香港问题、新疆问题、台海问题等涉我重大议题，基于 WDE 平台广泛采集特定区域政治、外交、军事等方面的数据，通过自然语言处理技术，实时聚合出每日热门政治事件，并深度分析该事件的热度、影响力、相关话题、演变趋势，以及与该事件相关的主要媒体和跨平台社交账号的情感倾向、政治立场和观点看法。

二是境外目标对象的智能锁定。基于实验室积累十年的情报分析技术和 GoIN 系统，深入挖掘热门政治事件的幕后组织和关键人物。以近期香港国安法为例，通过广谱关联技术从监测的 18 万香港 Twitter 用户中，挖掘出与该事件相关的 3000 多名主要人物，100 多名网络意见领袖，并利用人物画像技术多维度刻画人物的行为特征、活动规律、政治倾向、观点立场等，为舆论引导提供精确的目标对象。



三是沉浸式宣传内容的智能生成。根据特定区域的政治态势和受众对象的习惯爱好，充分利用人工智能技术辅助创作引导内容，如视频所示，基于联想功能可自动续写与输入内容相关的一段文本；相关素材推荐为创作者提供多源海量素材；衍生文案能基于原始文稿洗出若干篇语义不变、表述不同的新文稿；智能创作支持一体化的长文本内容生成。

四是非受控环境下的智能推送。境外社交媒体风控严格，在境外规模化、智能化、组织化培育社交机器人账号需要大量的资源投入和技术操作。目前本团队在 Twitter、Facebook、Telegram 等平台培育了数百个拟人化的机器人账号，并提供一站式跨社交媒体操作和调度平台，实现自动批量发帖、评论、转发，为大规模、精准、安全投放引导内容提供技术基本。

四、工作总结与展望

当前，我们仅是迈开万里长征的第一步，自主研发的智能宣传系统作为新兴事物，初步具备政治态势感知、群体心理画像、内容智能生成和信息智能投送能力，但离全面服务国家在境外网络舆论斗争的需求还有一定差距。

未来我们期盼在科学院领导和所领导的大力支持下，与中央网信办、中宣部、军科委、安全部等部门加强合作，明确任务需求，加大资源投入，打造出能真正自动讲好中国故



事、发出中国声音、扩大中国影响力的技术平台，为早日实现“东风压倒西风”提供全方位的技术支持。



QB 分析监测系统 需求说明书

中科天玑数据科技股份有限公司

2024 年 03 月 06 日



修改历史

时间	需求来源	版本	作者	修改内容
2024. 03. 06	项目客户	V1.0	杨菊海	新增



目录

- 目录..... 3
- 1 简介..... 5
 - 1.1 项目背景介绍..... 5
 - 1.2 范围..... 5
- 2 总体概述..... 6
 - 2.1 系统概述..... 6
 - 2.2 用户群体..... 6
 - 2.3 功能清单..... 6
 - 2.4 业务逻辑..... 8
 - 2.5 业务流程图..... 8
 - 2.6 功能结构图..... 9
 - 2.7 核心功能..... 9
 - 2.8 角色说明..... 9
- 3 功能需求说明..... 10
 - 3.1 综合搜索..... 12
 - 3.1.1 搜索首页..... 12
 - 3.1.2 搜索结果页..... 15
 - 3.1.3 信息分析页..... 21
 - 3.2 特色专题..... 24
 - 3.2.1 专题管理..... 24
 - 3.2.2 模板管理..... 58
 - 3.3 QB 服务..... 60
 - 3.3.1 任务管理..... 60
 - 3.3.2 图谱管理..... 60
 - 3.3.3 QB 分析..... 61
 - 3.4 报告管理..... 61
 - 3.5 采集概览..... 62
 - 3.6 系统管理..... 62
- 4 软硬件环境需求..... 62
 - 4.1 软件环境需求..... 62
 - 4.2 硬件环境需求..... 63
- 5 非功能需求..... 63
 - 5.1 信息采集..... 63
 - 5.2 文本处理..... 63
 - 5.3 信息分析..... 64
 - 5.4 数据存储..... 64
 - 5.5 响应时间..... 64
 - 5.6 系统并发..... 64
 - 5.7 系统安全..... 65
 - 5.8 数据安全..... 65



5.9 升级与扩展	66
5.10 可维护	66
5.11 可兼容	67
6 附录	67



1 简介

1.1项目背景介绍

各业务部门随着开展业务工作不断增多，掌握的数据种类不断丰富、数据结构不断异化、数据总量急速增长，导致业务人员缺少对数据的统一管理以及如何基于智能化引擎深挖数据价值，提升数据分析的智能化水平。因此亟需建设 QB 分析监测平台，具备“多源采集、专题监测、QB 分析”的能力，用来满足用户对关注业务方向实现动态感知与 QB 智能分析及实现业务的需求支撑。

1.2范围

此份文档的阅读对象包括：本项目的开发所有开发人员，测试人员，项目经理，以及 QA 小组人员。



2 总体概述

2.1 系统概述

平台通过对行业数据和开源数据进行融合集成,可用于辅助业务部门对特定事件、特定主题、特定领域等方面数据进行数据监测与分析,筛查汇总散布于海量数据中的重点信息,结合知识库对重点信息进行图谱分析及关联挖掘,深究数据包含的价值信息,最终为业务部门的业务需求提供保障。

2.2 用户群体

系统面向 QB 分析师和甲方业务部门人员,适用于初级用户和资深用户。

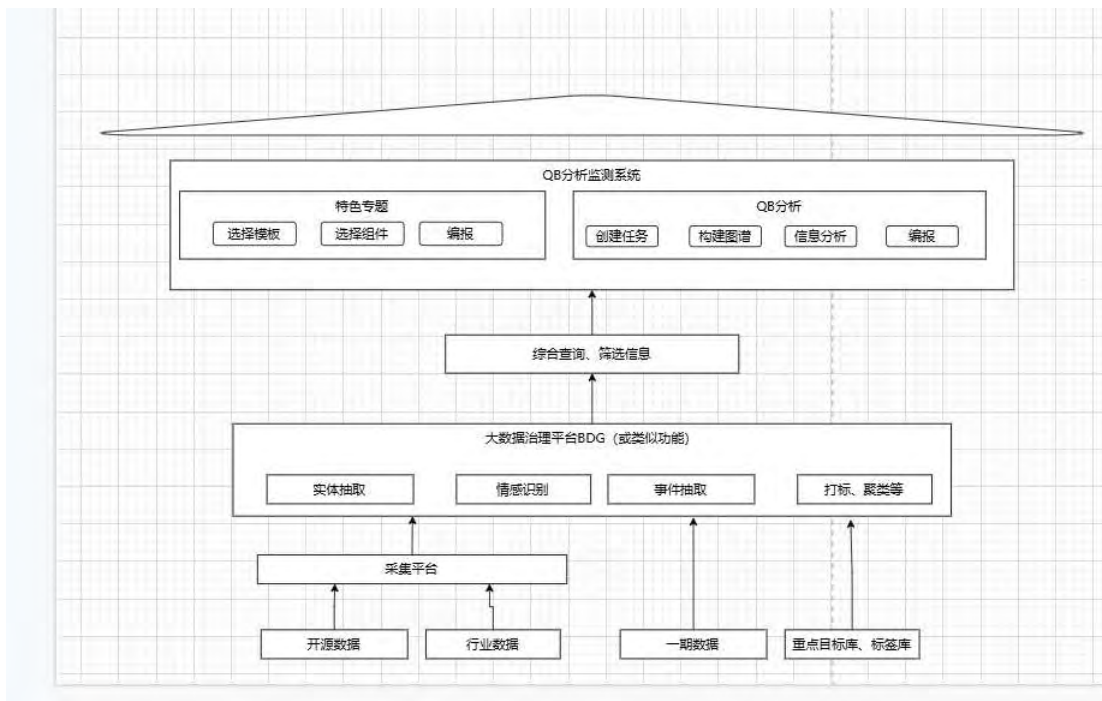
2.3 功能清单

编号	一级模块	二级模块	功能点	子功能点
1	综合搜索	——	搜索快捷入口	支持多模态检索、支持图谱检索
2		——	搜索结果查看和处理	支持二次搜索、高级搜索,支持信息的收藏、导出、推送到QB分析,复制链接,复制信息等功能。
3		——	支持S网A网的结果查看	
4		——	搜索信息的分析	支持信息作者画像的查看,支持查看相关地点、人物、机构。支持外文翻译
5	特色专题	专题管理	专题列表	支持专题的分类管理,支持专题的增删改查。
6		——	专题信息列表	支持对专题筛选后的信息列表进行查看。
7		——	专题分析	支持对专题进行多维度的图表分析,支持生成分析报告。
8		模板管理	支持对专题模板的管理,包括系统模板、个人模板。支持对模板的增删改查。	
9	QB 服务	任务管理	支持对QB任务的增删改查,支持对人物配置人员、配置	



			分析结果，生成分析报告。	
10		图谱管理	支持对图谱的增删改查。	
11		信息分析	关联分析	支持对图谱数据进行关联分析，支持保存中间分析结果和截图。
12			时空分析	支持对图谱数据加载到地图上，进行时空分析。支持保存中间分析结果和截图。
13	个人中心	我的收藏	支持查看个人收藏的信息，支持导出、复制，推送到关联分析。	
14		我的结果	支持查看个人的分析结果，包括中间分析结果，截图。	
15		修改资料	复用公司统一总控平台	
16		修改密码	复用公司统一总控平台	
17		退出登录	复用公司统一总控平台	
18	报告管理	——	支持对报告的查询、搜索、预览、下载等功能。	
19	采集概览	——	支持对采集情况的宏观统计展现。	
20	系统管理	——	复用公司统一总控平台，支持机构管理、用户管理、角色管理。	

2.4业务逻辑



整个产品的业务逻辑是围绕采集信息，进行特殊专题监测和 **QB** 服务分析开展的。

2.5业务流程图



2.6功能结构图



2.7核心功能

系统核心功能模块包括综合搜索、特色专题、QB 服务。

- (1) **综合检索：**支持用户对采集信息的多维度查询，以及对信息的深层分析。
- (2) **特色专题：**基于采集数据，提供大量业务组件和模板，辅助初级用户快速搭建分析专题，生成分析报告。
- (2) **QB 服务：**基于采集数据，提供多维度图谱分析算子，辅助资深用户去挖掘时间背后的人事组及关系。

2.8角色说明

角色	所属部门	权限说明
普通用户	待后续调研机构树	数据权限严格管控，只能看到个人数据+公共专题
管理员	待后续调研机构树	只能看到管理员个人数据，可修改公共专题



3 功能需求说明

统一说明：

功能需求中，对于业内通用功能将在此集中说明。对于特殊情况，以具体章节描述为准，否则全部采用统一说明的要求。

1、交互规范

1.1 页面尺寸

页面采用 1920*1080 大小设计，其他屏幕分辨率依照规则进行相应的尺寸适配。

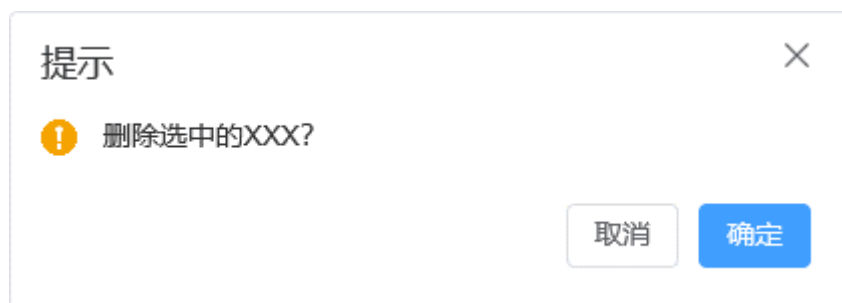
1.2 悬停效果

所有菜单导航栏、分类树和操作按钮，鼠标划过、悬停时均需支持变色提示。
参考公司 OKS 产品：<http://222.190.246.206:40037/oks/#/ontology/concept>

2、操作规范

2.1 删除操作规范

所有删除操作均需弹窗进行二次确认。XXX 为系统中专题，任务，图谱等新增内容。



2.2 操作结束提示词

按照常用消息类型，将提示分为取消操作、操作成功、操作失败 3 类。

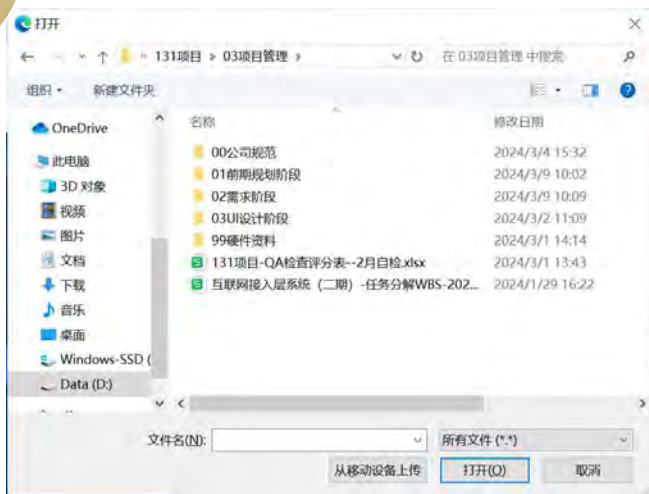


对于删除操作，删除成功后，会提示“删除操作成功。”

对于勾选后生效的按钮，若用户在未勾选下，直接点击按钮，则会提示“操作失败，未选中任何信息。”

2.3 上传文件交互规范

点击相关区域，弹窗选择文件，需进行文件格式、文件大小校验。校验内容见文件上传提示词。如下图所示。



上传成功或失败，会有 2.2 提到的对应提示词。

3、表格规范

3.1 表格排序

表格中无论是否包含创建时间，默认情况均按照创建时间降序排序，除非特殊说明；

编辑功能不改变表格中的数据展示顺序；

列表项内容包含数字、日期的，表头默认增加排序；

节点数量	连边数量	来源	状态	创建时间	更新时间	操作
------	------	----	----	------	------	----

3.2 表格内容字数限制

列表项文字过长如何显示？当列表项的宽度无法完整显示所有文字时，后面加 3 个点，鼠标移入后显示全部。

状态	关联服务器	设置报警
在用	停用bug测试。测试-G	是
设计中	停用bug测试。...	是

3.3 分页器规范

共 78 条 10条/页 < 1 2 3 4 5 6 ... 8 > 前往 1 页

分页器于列表搭配使用，当数量超过 10 个时，列表底部显示分页器。支持每页 10 条、20 条、50 条、100 条切换。

4、查询搜索规范

4.1 输入字符限制

单行输入框输入字符限制为 20 个。

多行输入框输入字符限制为 200 个。

4.2 所有查询搜索框、下拉筛选框，均支持模糊搜索和一键清空。

3.1 综合搜索

登录页面采用公司通用组件账号密码登录。

参考 OKS: <http://222.190.246.206:40037/oks/#/login>

登录后系统默认展示综合搜索内容，其包括搜索首页、搜索结果页和信息分析页。

3.1.1 搜索首页

➤ 原型如下：



➤ 页面布局：分为 Header 栏、Main 栏。

Header 栏为导航栏，从左至右依次为“Logo+系统名称”、“综合搜索”、“特色专题”、“QB 服务”、“报告管理”、“系统管理”；右侧为登录个人账号信息。

➤ Header 栏：

一级菜单	备注	二级菜单	备注	
综合搜索	单击，切换至综合搜索页面	无		
特色专题	单击，切换至特色专题首个二级菜单（专题管理）	专题管理、模板管理	单击，跳转至相关页面	
QB 服务	单击，切换至 QB 服务首个二级菜单（任务管理）	任务管理、图谱管理、QB 分析	单击，跳转至相关页面	
报告管理	单击，切换至报告管理页面	无		
系统管理	单击，切换至系统管理页面。需管理员角色权限才能查看。	用户管理、角色管理、单位管理、	单击，跳转至相关页面	
个人账号	单击，向下展开二级	我的收藏、我		



	菜单	的成果、修改资料、修改密码、退出登录		
--	----	--------------------	--	--

➤ Main 栏：综合搜索入口。
按照最小可行性产品原则（MVP 原则），所有搜索方式，暂不支持历史搜索记录功能。
（1）文本搜索
点击【文本搜索】图标，页面如下图所示。
支持按照关键词搜索、高级搜索。提示词“请输入关键词搜索”，输入后点击搜索图标，进入【文本搜索结果页】。



点击【高级搜索】按钮，页面如下图所示。点击搜索，将筛选条件带入【文本搜索结果页】。点击返回，则隐藏高级搜索弹窗。



分类维度及取值如下：

维度	取值	备注
渠道	新闻媒体、社交平台、短视频、S 网 A 网、其他通道	单选。 新闻媒体： 中国记协会员名录 百度百科 (baidu.com) 国外新闻网站：BBC、CNN 社交平台：微信，微博，推特等 短视频：抖音，快手，YouTube S 网 A 网：特定网站；
情感	正面、中性、负面	单选。
地域	境外+中国 34 个省级行政区	单选。
事件大类	Gdlet20 个事件大类	单选。如下图



是否噪音	正常信息、疑似噪音	单选。噪音是指传播学中无用、虚假及有害的信息。详见公司噪音算法。
评论数	100 以下； 100--1000； 1000 以上	单选。

备注：
Gdelt20 个事件大类如下。

代码	含义	代码	含义	代码	含义
01	发表公开声明	08	投降	15	武力表明立场
02	上诉	09	调查	16	减少关系
03	明确表示合作	10	需求	17	胁迫
04	咨询	11	指控	18	袭击
05	从事外交合作	12	拒绝	19	战争
06	参与材料合作	13	威胁	20	使用非常规暴力
07	提供援助	14	抗议		

参考来源：[Gdelt 数据库初探 \(douban.com\)](https://douban.com)

(2) 图片搜索
点击【图片搜索】图标，通过图片来搜索视频信息。页面如下图所示。
支持选择图片或拖拽图片。图片格式及大小限制，见页面提示词。上传成功后，
跳转到【图片搜索结果页】。



(3) 跨模态搜索
点击【图跨模态搜索】图标，通过文本来搜索视频信息。页面如下图所示。
支持按照关键词搜索。提示词“请输入关键词搜索”，输入后点击搜索图标，进



入【跨模态搜索结果页】。



(4) 视频搜索

点击【视频搜视频】图标，通过视频来搜索视频信息。页面如下图所示。支持选择视频或拖拽视频。视频格式及大小限制，见页面提示词。上传成功后，跳转到【视频搜索结果页】。



3.1.2 搜索结果页

页面布局为上下结构，顶部是四种搜索方式，其交互与搜索入口页一致，区别在于文本搜索时，能够查看筛选条件。

3.1.2.1 文本搜索结果

文本搜索结果是对所有采集通道结果的信息展现。





➤ 对于多个信息，支持功能如下：

功能点	功能描述	前置条件	备注
勾选全选功能	支持对信息的勾选、全选。选中后，支持推送到 QB 分析、收藏、导出操作。	无	
去重功能	支持对信息进行去重或不去重展现。去重是指信息标题和作者完全一样。	无	单选下拉，默认显示为不去重
信息排序	支持信息按照如图下拉框条件进行排序。 默认显示为发布时间降序，保证最新事件信息在顶部。		
推送到 QB 分析	将勾选的信息所涉及的节点和关系，一起推送到关联分析中。相当于批量新增节点和关系。	勾选信息	若未勾选直接点击按钮，则提示操作失败，未选中相关信息。
收藏	将勾选的信息进行收藏，在个人中心的“我的收藏”进行查看。	勾选信息	若勾选信息中包括已收藏和未收藏的信息，此时加入收藏时，在我的收藏中自动去重。
导出	将勾选的信息按照所需信息进行导出，支持导出格式为 Excel。 导出设置如图所示。	勾选信息	

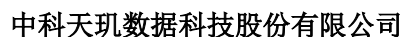


➤ 对于单个信息，支持功能如下：

功能点	功能描述	备注
封面图	单一封面图显示	
情感倾向	对信息进行情感分类，包括正面、中性、负面。	

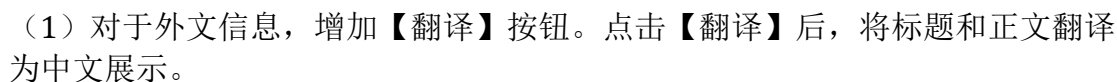


信息类型	包括原发、转发、评论。	
标题	显示信息标题。命中关键词的部分，字体标红突出显示。	点击标题进行信息分析页。
是否已读	用户点击后，会在标题后面打标显示已读。	
信息发布时间	格式为年月日时分秒	是采集作者的发布时间
正文	显示信息正文。页面显示首个命中关键词的部分，字体标红突出显示。前后超出字数则用...显示。	悬停时无效果。
信源类型	包括网站、电子报、新闻客户端、电视媒体、贴吧、论坛、视频、社交平台、其他	所有类型，包括境内外信源通道。
信源名称	格式为：通道名称+作者信息（或板块信息）。 （1）搜狐网—军事 （2）抖音—董宇辉 （3）北京日报	若无作者信息，则不显示，如示例3。
赞转评信息	显示信息的点赞、转发、评论的数量信息。	
查看原文	点击后，在新的浏览器页签查看原文	五个图标，对应五个功能。鼠标悬停在图标时，显示文字。
加入收藏	点击后，该条信息进入我的收藏。按钮改为“取消收藏”。	
素	点击后弹窗导出设置，同批量导出。	



材 导 出		     查看原文  加入收藏  素材导出  复制链接  复制信息
复 制 链 接	复制原文链接信息	
复 制 信 息	复制该信息在列表展现的元数据。示例如下 标题：转发微博 链接： http://weibo.com/5671712418/04qVV6TeK?refer_flag=100103_0103_ 摘要：空 时间：2024-03-11 02:41:18 来源：微博 作者：PeterDong 的微博 倾向性：中性	

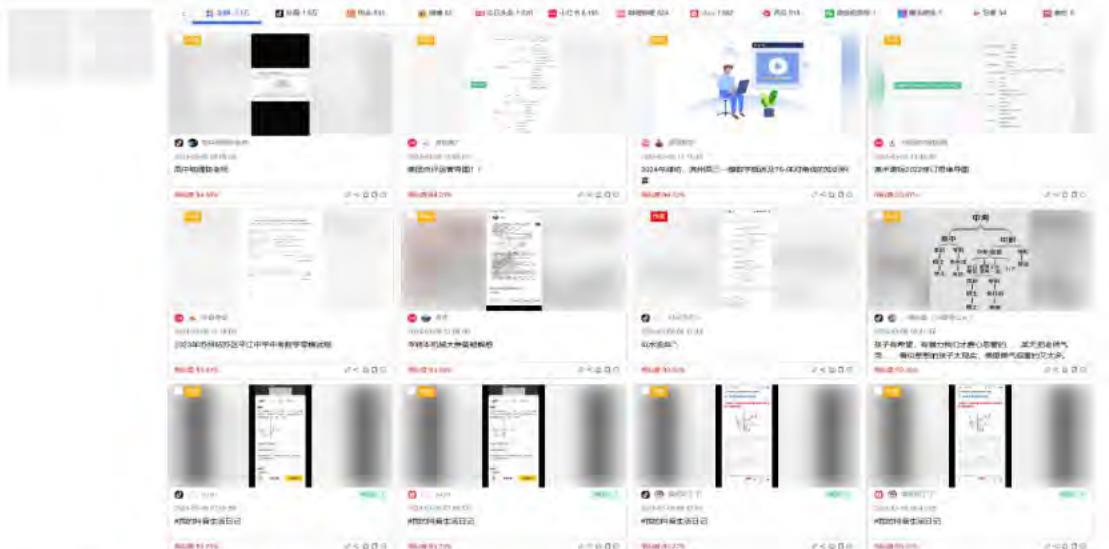
► 特殊情况说明



- (2) 对于 S 网 A 网信息, 由于其是电商网站结构, 会增加价格信息, 去掉赞评转信息。其余不变。

3.1.2.2 图片搜索结果

图片搜索结果是对视频信源采集通道结果的信息展现，简单理解为图搜视频。



每行最多展示 4 个卡片。

- 对于多个信息的操作，与文本搜索功能一致。
- 对于单个信息，视频信息功能如下：删除内容，标黄内容是方便与文本搜索结果进行对比。

功能点	功能描述	备注
封面图	单一封面图显示	
情感倾向	对信息进行情感分类，包括正面、中性、负面。	
信息类型	包括原发、转发、评论。	
标题	显示信息标题。命中关键词的部分，字体标红突出显示。	点击标题进行信息分析页。
是否已读	用户点击后，会在标题后面打标显示已读。	
信息发布	格式为年月日时分秒	是作者的发布时间



时间		
正文	显示信息正文。页面显示首个命中关键词的部分，字体标红突出显示。前后超出字数则用...显示。	悬停时无效果。
信息来源类型	包括网站、电子报、新闻客户端、电视媒体、贴吧、论坛、视频、社交平台、其他	所有类型，包括境内外信源通道。
信源名称	格式为：通道名称+作者信息（或板块信息）。 （1）搜狐网—军事 （2）抖音—董宇辉 （3）北京日报	若无作者信息，则不显示，如示例3。
相似度	利用图片相似度算法，显示视频中每帧图片相似度的最高值。	
赞转发评信息	显示信息的点赞、转发、评论的数量信息。	列表页不展现，分析页展现
查看原文	点击后，在新的浏览器页签查看原文	
加入收藏	点击后，该条信息进入我的收藏。按钮改为“取消收藏”。	
素材导出	点击后弹窗导出设置，同批量导出。	
复制链接	复制原文链接信息	
复制信息	复制该信息在列表展现的元数据。示例如下 标题：XXX 链接： http://douyin.com/5671712418/O4qVV6TeK?refer_flag=1001030103_	



摘要：空 时间：2024-03-11 02:41:18 来源：抖音 作者：董宇辉 倾向性：中性	
---	--

3.1.1.2.3 跨模态搜索结果

与图片搜索类似，可以理解为文本搜视频。差异点在于输入内容改为文本，相似度算法改为文本-视频相似度算法。

3.1.1.2.4 视频搜索结果

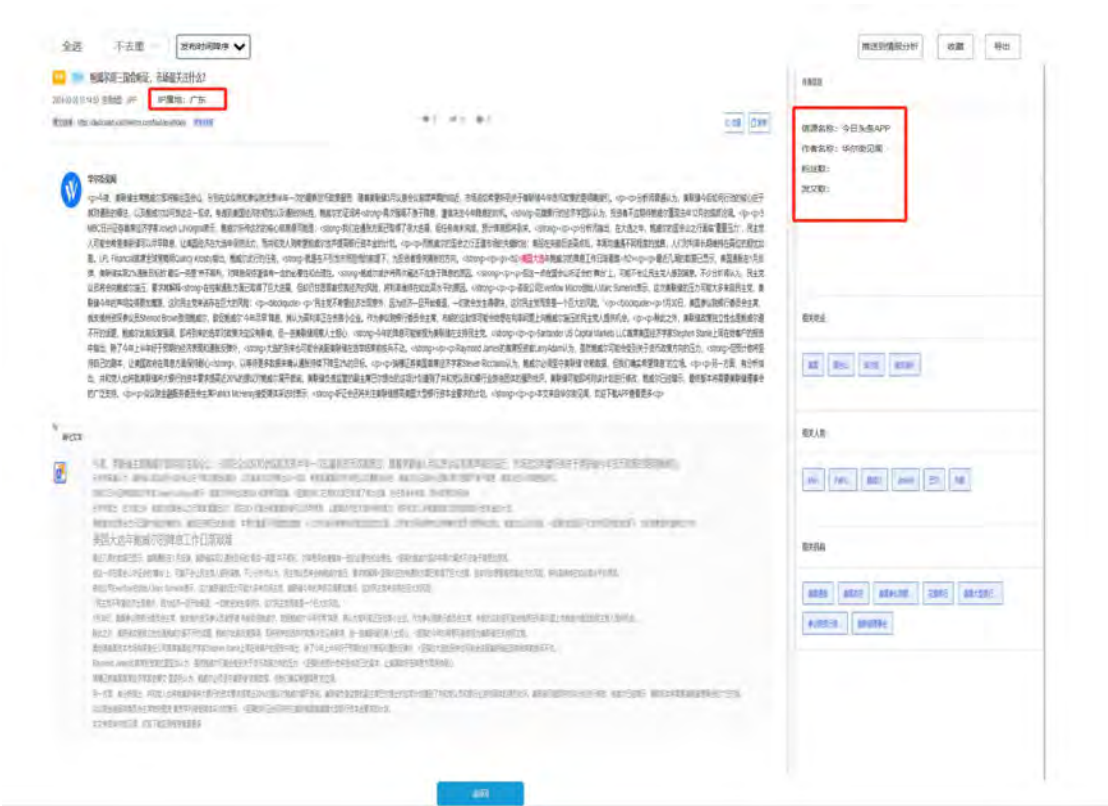
与图片搜索类型，可以理解为视频搜视频。只是输入内容改为视频，相似度算法改为视频相似度算法。

3.1.3 信息分析页

信息分析页主要包括 3 大类。图文信息分析、视频信息分析、S 网 A 网信息分析。

3.1.3.1 图文信息分析

对于图文信息分析，其分析维度如下图所示。



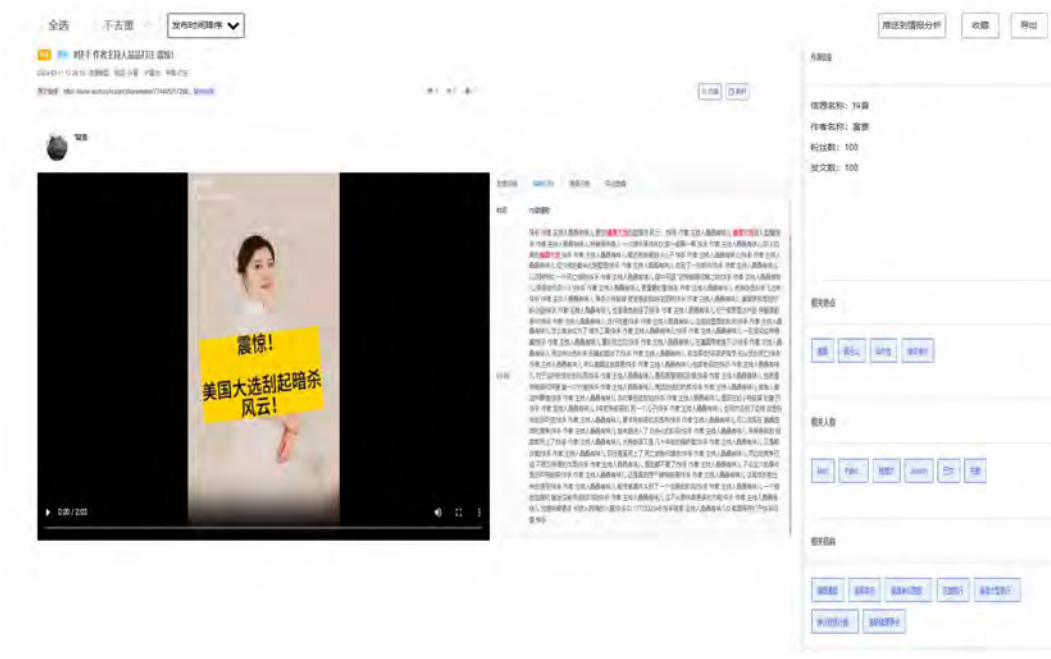
除去列表页中的元数据信息，额外新增了 IP 属性，右侧的作者信息中，新增粉丝数，发文数。新增相关地点、相关人物、相关机构信息。点击翻译，则在下文显示对应的中文信息。对于中文原文，则弹出提示词“原文

信息已经是中文”。

若某篇信息无法抽取相关内容，则只显示表结构信息，没有数据信息。
点击返回按钮，返回到列表页。

3.1.3.2 视频信息分析

对于视频信息分析，其分析维度如下图所示。



与图文信息分析类似，只是正文部分发生变化。
正文包括左侧的视频和右侧的识别。视频支持播放，控制音量、全屏和下载。右侧的是不包括封面识别、抽帧识别、语音识别和评论数据。

封面识别：

利用 OCR 技术，将封面中内容提取出来，字体标红突出显示涉及监测词的内容，并显示时间。

封面识别	抽帧识别	语音识别	评论数据
时间	内容提取		
0:00:00	8快手 作者主持人品品打注 震惊! 美国大选刮起暗杀 风云!		

抽帧识别

利用 OCR 技术，将不同帧中的内容提取出来，字体标红突出显示涉及监测词的内容，并显示时间。



封面识别	抽帧识别	语音识别	评论数据
时间	内容提取		
00:00	快手 作者:主持人晶晶有味儿 震惊!美国大选刮起暗杀 风云! 快手 作者:主持人晶晶有味儿 美国大选陷入血腥 快手 作者:主持人晶晶有味儿 特朗普被卷入 一次暗杀事件 和大家一起聊一聊 快手 作者:主持人晶晶有味儿 如火如荼的美国大选 快手 作者:主持人晶晶有味儿 最近特朗普的小儿子 快手 作者:主持人晶晶有味儿 快手 作者:主持人晶晶有味儿 在他爸的豪华大别墅里 快手 作者:主持人晶晶有味儿 收到了一份邮件 快手 作者:主持人晶晶有味儿 以及附带的 一个死亡威胁 快手 作者:主持人晶晶有味儿 信中写道 "在特朗普闭嘴之前 快手 作者:主持人晶晶有味儿 猜猜会死多少人" 快手 作者:主持人晶晶有味儿 更重要的是 快手 作者:主持人晶晶有味儿 就有白色粉末飞出来 快手 作者:主持人晶晶有味儿 事后小特朗普 被紧急安排前往医院 快手 作者:主持人晶晶有味儿 美国联邦危险材料小组 快手 作者:主持人晶晶有味儿 也紧急地前去了 快手 作者:主持人晶晶有味儿 位于佛罗里达州的 特朗普的家中 快手 作者:主持人晶晶有味儿 进行检查 快手 作者:主持人晶晶有味儿 这信封里面的粉末 快手 作者:主持人晶晶有味儿 怎么就会成为了 暗杀工具 快手 作者:主持人晶晶有味儿 快手 作者:主持人晶晶有味儿 一旦感染这种病毒 快手 作者:主持人晶晶有味儿 重则败血症 快手 作者:主持人晶晶有味儿 在美国早就有不少 快手 作者:主持人晶晶有味儿 用这种白色粉末 投毒的案件了 快手 作者:主持人晶晶有味儿 前后导致5名政府官员 和议员的死亡 快手 作者:主持人晶晶有味儿 所以美国这些政要 快手 作者:主持人晶晶有味儿 包括老百姓 快手 作者:主持人晶晶有味儿 对于这种粉末状的东西 快手 作者:主持人晶晶有味儿 是极其警惕和恐惧 快手 作者:主持人晶晶有味儿 也就是特朗普和拜登 第一次对垒 快手 作者:主持人晶晶有味儿 竞选总统的时候 快手 作者:主持人晶晶有味儿 就有人搞这种事情 快手 作者:主持人晶晶有味儿 当时拿到信封的 快手 作者:主持人晶晶有味儿 是现在的小特朗普 的妻子 快手 作者:主持人晶晶有味儿 8年前特朗普的 另一个儿子 快手 作者:主持人晶晶有味儿 也同时收到了这样 白色粉末的恐吓信 快手 作者:主持人晶晶有味儿 要求特朗普放弃选举 快手 作者:主持人晶晶有味儿 可以说现在 美国选举的竞争 快手 作者:主持人晶晶有味儿 越来越进入了 白热化的阶段 快手 作者:主持人晶晶有味儿 各种各样的 招数都用上了 快手 作者:主持人晶晶有味儿 光特朗普又是 几十年前的强奸案 快手 作者:主持人晶晶有味儿 又是欺诈案 快手 作者:主持人晶晶有味儿 现在甚至用上了 死亡威胁和暗杀 快手 作者:主持人晶晶有味儿 两边的竞争已经 不顾及所谓的体面 快手 作者:主持人晶晶有味儿 是脸都不要了 快手 作者:主持人晶晶有味儿 不论这次的事件是恐吓特朗普 快手 作者:主持人晶晶有味儿 还是真的想干掉特朗普 快手 作者:主持人晶晶有味儿 这背后折射出来的信号 快手 作者:主持人晶晶有味儿 就使美国内斗到了 一个全新的阶段 快手 作者:主持人晶晶有味儿 一个越发血腥和 越发没有底线的阶段 快手 作者:主持人晶晶有味儿 这不光意味着更多的内耗 快手 作者:主持人晶晶有味儿 也意味着更多 点燃火药桶的火星 快手ID:1772332345 快手搜索 主持人晶晶有味儿Q 截图保存打开快手观看 快手		

语音识别：
利用语音识别技术，将语音中的内容提取出来，字体标红突出显示涉及监测词的内容，并显示时间。

封面识别	抽帧识别	语音识别	评论数据
时间	内容提取		
0:00:00	美国大选陷入血腥，特朗普被卷入疑次暗杀事件。这个视频和大家一起聊一聊如火如荼的美国大选。最近特朗普的小儿子在他爸的好。		

3.1.3.3 S 网 A 网信息分析

对于 S 网 A 网的信息分析，其分析维度如下图所示。



除去列表页展现的元数据信息，额外新增了商品类型，数据来源、可支付币种，信源地址，详情、评价，和留言等信息。

备注：不同的特殊网站，表结构信息可能不同。缺少的内容展示为空。本文是以公司现有能力举例。

3.2 特色专题

包括专题管理和模板管理。左侧导航栏可以切换专题管理和模板管理页面。

3.2.1 专题管理

➤ 原型如下：



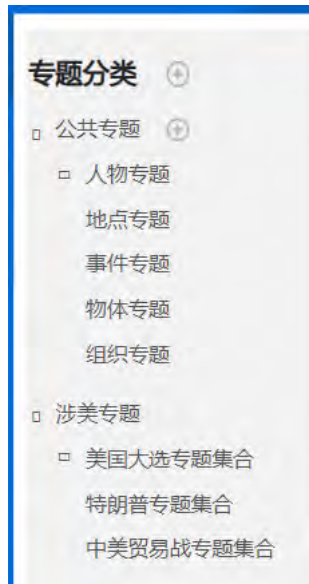


➤ 入口：点击左侧导航栏“专题管理”。左侧导航栏支持展开、收起功能。

3.2.1.1 专题列表页

➤ 主要功能区分为左右结构。

左侧是专题分类树，是纯粹的分组，不包括专题名称。



- 页面元素“专题分类”，默认为根节点。点击右侧加号，新增子节点。分类数量有限，小于 1000，暂不考虑树节点过多造成的加载时间问题。
- 页面元素“公共专题”，默认为首个二级节点。只有管理员角色具有公共专题的编辑权限。普通用户只能查看公共专题内容。专题来源，一方面是管理员新增，另一方面是普通用户可以把个人专题发布共享。
- 页面元素“涉美专题”，是当前用户个人新增的二级节点。点击节点，显示该节点下所有的专题。点击父节点，显示所有子节点的专题。点击专题分类时，显示该用户能查看的全部专题。
- 专题属于敏感信息，需要数据权限管理。每个用户只能查看个人专题和公共专题，只能编辑个人专题。
- 用户点击新增专题，若没有选择左侧分类树，则默认为属于根节点专题分类。

➤ 右侧是通用的卡片列表形式，暂不支持切换为卡片平铺形式。



上方是搜索框，支持按照专题名称、专题描述、关键词进行搜索。

中间是列表的功能操作区，支持专题列表的增删改查、全选、排序。

当选择多个，点击重命名、修改时，会有提示词“操作失败，只能选择单个”

➤ 下方是卡片列表信息。

功能点	功能描述	备注
-----	------	----



专题名称		单行输入框字符限制，按统一规范执行。
专题内容		
监测关键词		
监测时间		格式为日期
创建时间		格式为日期

3.2.1.2 新增专题页

新增专题的原型如下：

新增专题

* 专题名称

* 监测周期

yyyy/mm/日 至 yyyy/mm/日

* 监测主体

添加账号

搜索 | 4日头条 | 张福贵 | 5007607333 | 地址11643882 关注12 获赞0

选择语种

☐ 汉语 ☐ 法语 ☐ 俄语 ☐ 阿拉伯语

☐ 西班牙语 ☐ 英语 ☐ 印地语 ☐ 孟加拉语

☐ 葡萄牙语 ☐ 马来印尼 ☐ 日语 ☐ 德语

* 监测关键词

标题+正文 与 或 优先

切换语种

删除键 共1组

Parti républicain

排除关键词

标题+正文 与 或 优先

切换语种

(俄乌)((俄罗斯&乌克兰)&冲突&伤亡)

专题描述

请输入备注信息

0 / 200

关键词使用示例:

1、(俄乌)((俄罗斯&乌克兰)&冲突&伤亡)

含义：命中关键词XX，或者命中XXX，同时XX

2、2022&北京&((冬季&奥运会)&冬奥会)

含义：命中关键词XX，或者命中XXX，同时XX

取消

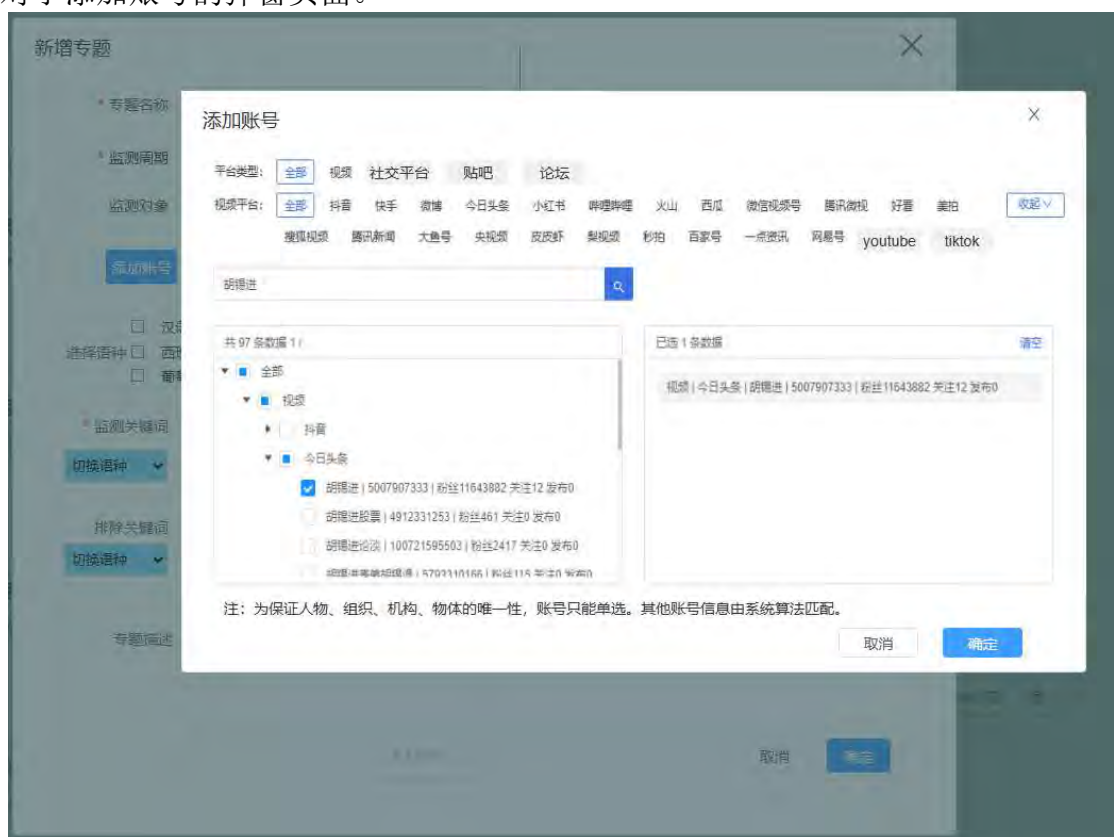
确定

字段名称	默认值	字段类型	是否开放修改权限	是否必填项	备注
专题名称		文本	是	是	按照统一规范要求：不超过 20 个汉字
监测周期		日期	是	是	
监测主体	显示提示 “为了监测效果，请确保唯一性”	文本	是	是	?帮助提示词“监测主体请确保没有歧义，与监测组件逻辑相关”



添加账号	--	--	是	否	事件类专题，无须账号。 点击按钮后弹窗。
选择语种	默认勾选中文	--	是	否	可多选 12 大语种。 勾选多语种后，将把全部的译文作为监测词去筛选采集信息。
监测关键词	--	--	是	是	支持下拉选择标题+正文、标题、正文； 支持与或优先操作； 支持切换语种；
排除关键词	--	--	是	否	同上
专题描述	无	文本	是	否	
关键词使用示例	--	--	--	--	前端写死，辅助用户写提示词。 后续根据业务调研补充。

➤ 对于添加账号的弹窗页面。



- 账号平台类型是采集信源类型的子集，如图所示，不包括网站，电子报等。都包括境内外信息。
- 每一类平台的具体采集信源平台，后续与业务调研后持续补充。
- 搜索框初始状态，见下图



共 0 条数据

暂无数据

已选 0 条数据

清空

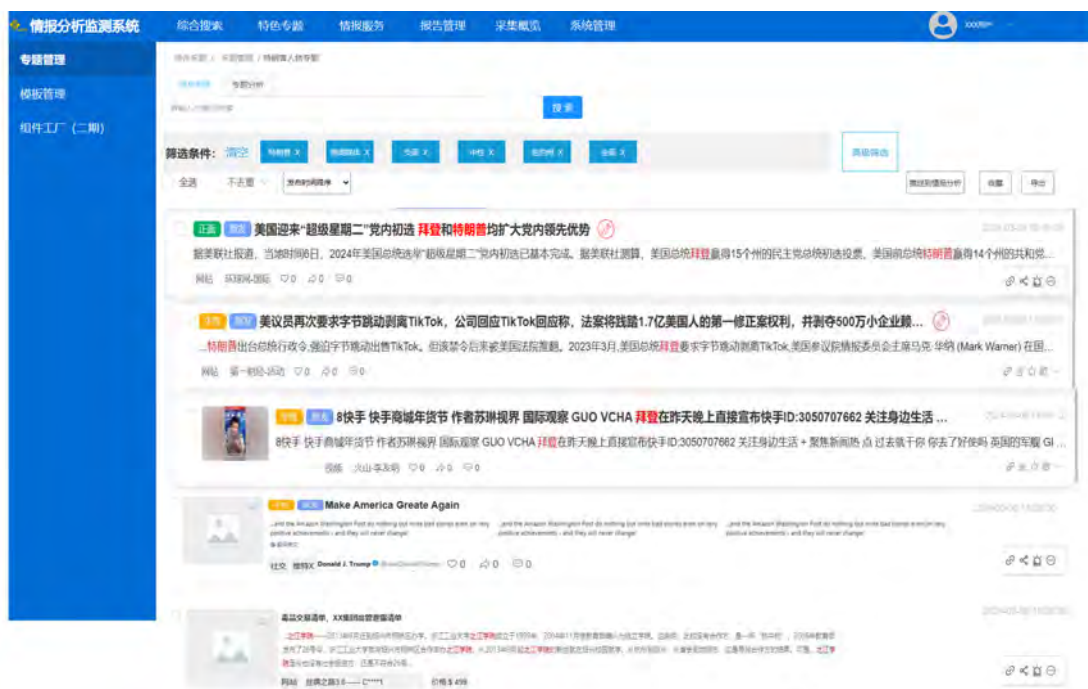
- 输入账号后 ID 后，显示如图所示。将多选框改为单选框。

3.2.1.3 专题详情页--信息

在列表页，点击专题名称，进入专题详情页的信息列表页，方便用户通过采集结果来验证监测词是否有效。

顶部是面包屑导航。两个页面“信息列表”“专题分析”可点击切换。

下方内容，与搜索结果页完全一致，都是采集的原始信息进行展现。



3.2.1.4 专题详情页--分析

专题分析页面包括左右两部分。



➤ 左侧是组件工厂

组件工厂，是系统内置的业务组件。本期不支持组件的二次开发。组件分类固定为下拉框展示内容。当选择分类后，组件内容随之改变。

当用户拖拽组件时，自动将组件所需数据显示出来，无须用户去配置组件的数据来源。技术支撑是系统内置采集后的人物库、机构库和各种抽取结果。

图形类组件，拖入后自带导入、导出和删除图标。

文本类组件，拖入后自带编辑、导出和删除图标。



➤ 右侧是专题布局内容

页面初次加载时，是白板形式。

中间有提示词“暂无组件，请拖拽左侧组件进行添加，或者选择模板”。

下方有模板推荐，模板卡片样式为封面图+模板名称。点击某个模板，相当于批量拖入一堆组件。点击“更多模板”卡片，弹窗显示所有模板。



所有的专题，右上角都有操作区。支持将专题内容生成报告、存为模板和发布共享。

点击生成报告，需用户填写报告名称。



点击存为模板，需用户填写模板名称、模板描述。



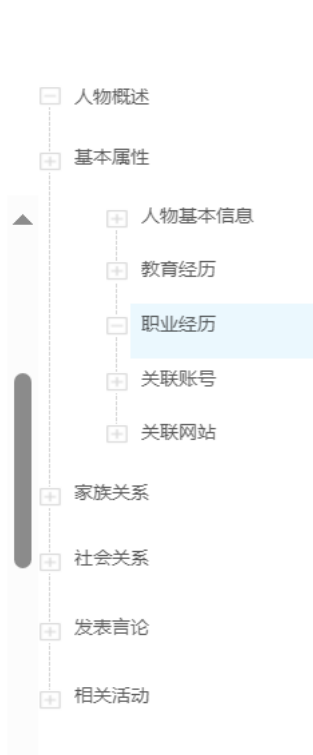
点击发布共享，是将个人专题同步拷贝一份到公共专题，需用户选择分组。



3.2.1.5 人物类模板组件

选择组件后，将同步读取组件的分类信息，在右侧的目录树中展示。方便用户点击目录树内容，直接跳转过去。

备注：所有的组件，都是针对监测主体来设计。



(1) 人物概述组件



- 该组件显示人物的头像、人物名称和人物简介。
- 支持编辑文本，导出功能和删除功能。其中，所有组件的导出按钮，支持导出数据和导出图片。点击导出图标后，再次选择导出内容，
- 数据来源是采集平台积累的人物库。通过创建专题时的监测主体和账号，保障人物库中的唯一性。

对采集平台的要求：需持续抽取、更新 wikidata 数据，形成人物基础库、人物动态信息库，作为行业积累。

(2) 人物基本信息组件



人物基本信息



中文名	唐纳德·特朗普	毕业院校	宾夕法尼亚大学、纽约军事学院
外文名	Donald Trump	政 党	美国共和党
别 名	Donald John Trump、川普	学 位	学士
性 别	男	学 历	大学本科
国 籍	美国	代表作品	做生意的艺术、学徒
民 族	德裔美国人 [2]	主要成就	第45任美国总统 (2017年1月20日-2021年1月20日)
出生地	美国纽约		《时代周刊》2016年度人物
出生日期	1946年6月14日		

- 该组件显示人物的基本信息，如图所示。
- 支持编辑文本，导入、导出和删除。导出数据后，在 Excel 中可以修改数据或表结构。再导入专题。系统将所有表结构字段按顺序分成两组显示。
- 数据来源是采集平台积累的人物基础数据库。通过创建专题时的监测主体和账号，保障人物库中的唯一性。相同功能，不再重复描述。

(3) 教育经历

教育经历



开始时间	结束时间	学校	学位
1964-08-01	1966-01-01	锡拉丘兹大学	-
1959-01-01	1964-01-01	纽约军事学院	-
1966-01-01	1968-05-01	沃顿商学院	理学士

(4) 职业经历

职业经历



开始时间	结束时间	就职组织/公司	职位
1971-01-01	-	-	会长
-	-	-	主管
1971-01-01	-	-	主席
2016-11-09	2017-01-20	-	候任美国总统
2017-01-20	2021-01-20	-	美国总统

(5) 关联账号

关联账号

Trump 账号ID 720193443260450960 个人主页 https://twitter.com/realDonaldTrump	The Trump Organization 账号ID 100064401390400 个人主页 https://www.facebook.com/trump	trump 账号ID 370x388257 个人主页 https://www.instagram.com/trump/
特朗普集团 账号ID UC9AKA0PvIGC5FqCEMaltag 个人主页 https://www.youtube.com/channel/UC9AKA0PvIGC5FqCEMaltag		

备注：不同的平台，其 logo 不同。系统内置全部平台的 logo。但导出的 Excel 中，图标是用平台地址来替换。

信息平台地址	账号名称	账号 ID	个人主页
www.youtube.com	Trump	798887388568	https://youtube.com/Trump

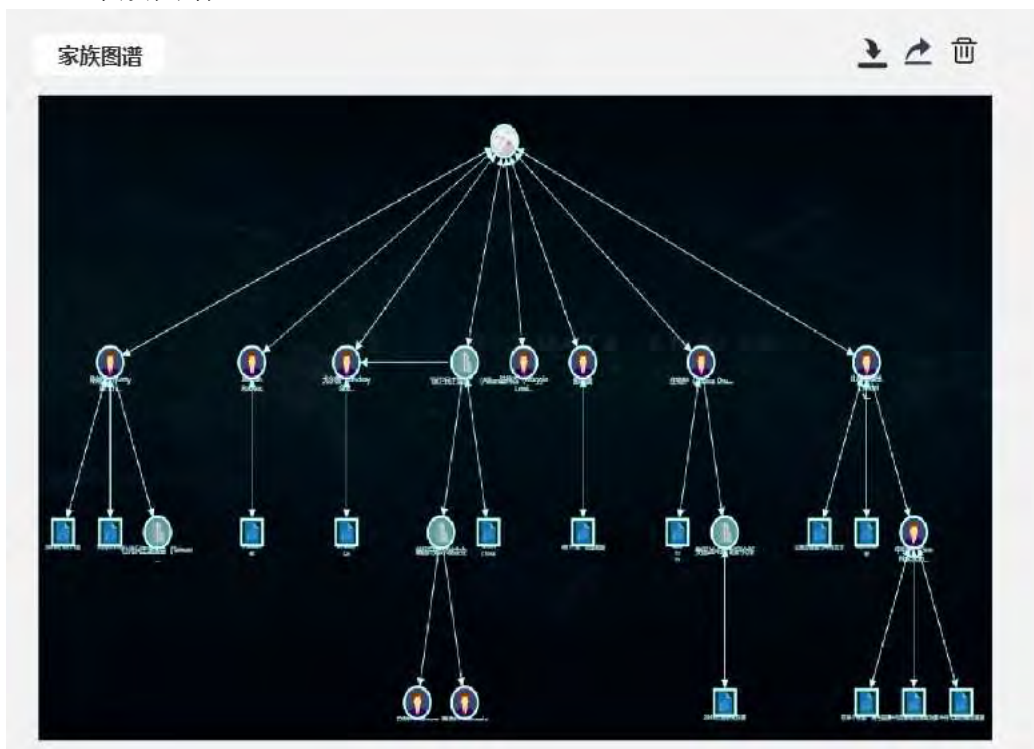
(6) 关联网站

关联网站

特朗普集团 账号地址 https://www.trump.com/

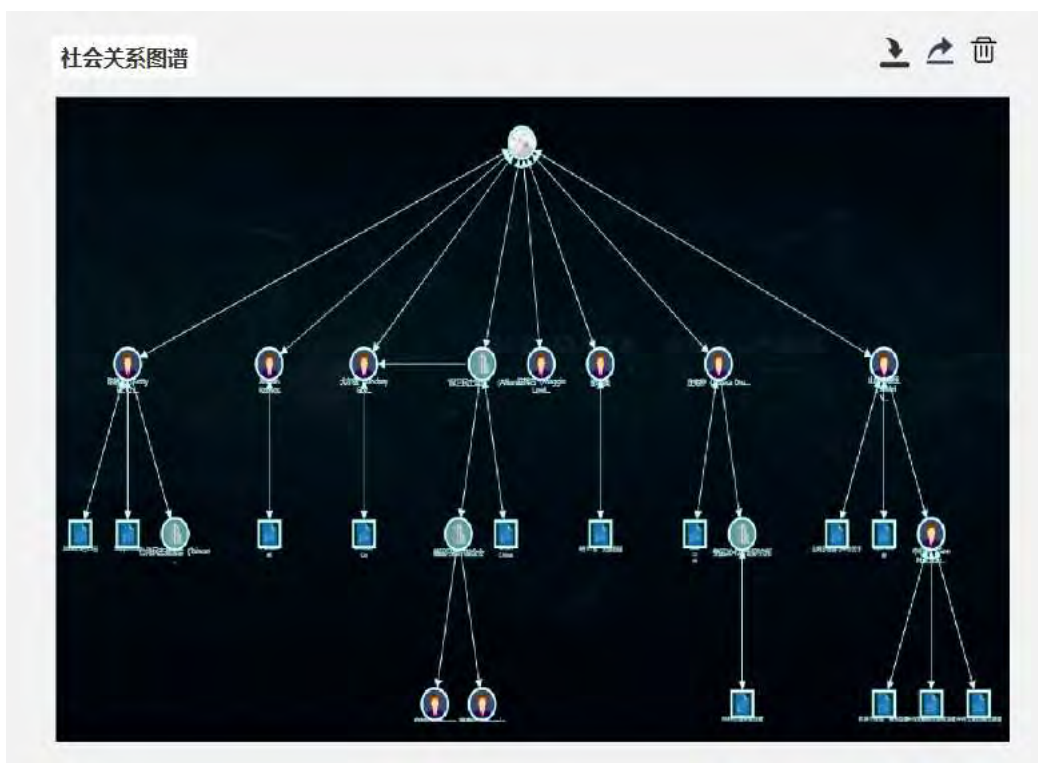
备注：与上图类似，但不同的网站，其 logo 保持一致。

(7) 家族图谱



- 鼠标悬停在节点或关系时，显示节点名称或关系名称。
- 备注：导出图谱的模板，等后续技术反馈。
- 数据来源：人物基础数据库。

(8) 社会关系图谱



数据来源：专题筛选后的采集信息，经过社会关系抽取算法后形成人物社会关系

库。

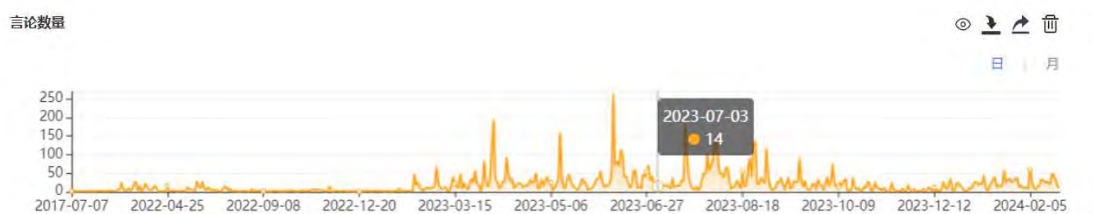
待定信息：是算法提前处理，还是拖入组件时再处理。

个人倾向是，产品提前梳理好重点人物，作为人物库的积累。

对于人物库里的专题，算法提前处理后，组件直接读取结果。

对于人物库之外的人物，则拖入组件时，实时调用算法抽取。页面提示处理中。

（9）言论数量



鼠标悬停时，显示日期和言论数量。支持按日、月切换。时间轴的起点，最早是监测周期的起点。

数据来源：专题筛选后的采集信息，经过言论抽取算法处理，再进行统计展现。

（10）言论倾向



数据来源：专题筛选后的采集信息，经过言论倾向识别算法处理，再进行统计展现。

（11）言论详情



- 支持关键词搜索，按言论时间降序排列。
- 显示日期和言论内容。
- 每屏显示 10 条（暂定，以 UI 设计条数为准），超出后上下滚动显示。

数据来源：专题筛选后的采集信息，经过言论抽取算法处理，再进行信息展现。

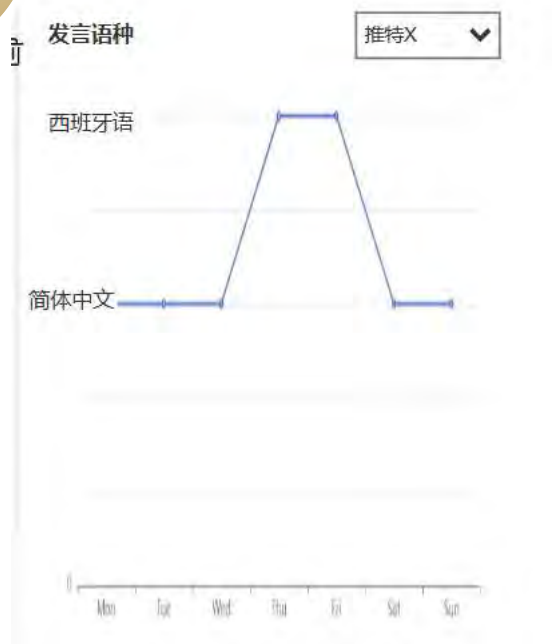
（12）话题



- 每屏显示 TOP10（暂定，以 UI 设计条数为准）。

数据来源：专题筛选后的采集信息，经过言论话题（主题）算法处理，再进行信息统计展现。

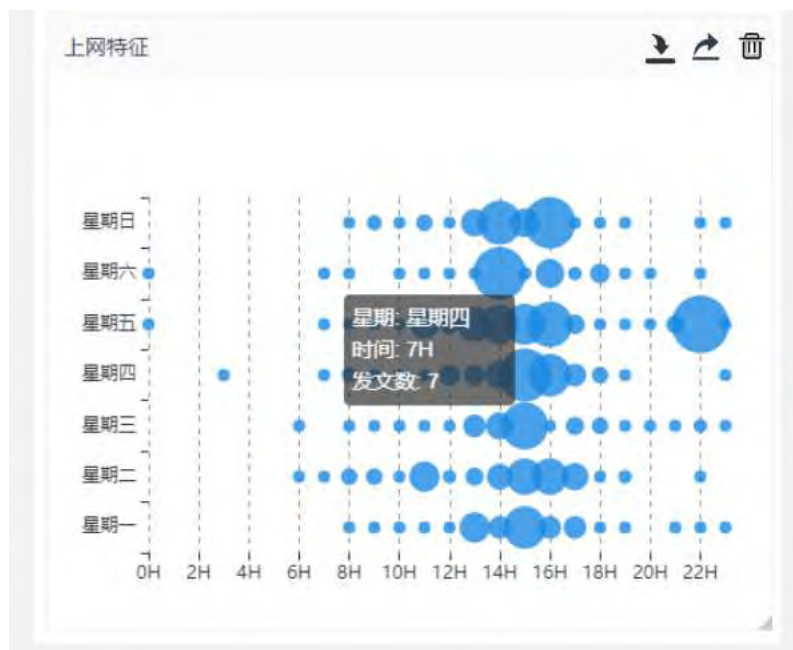
（13）发言语种



- 支持切换推特，Facebook，微博，微信，共计四种主流平台
- 横轴是日期排序，显示最近一月的发言语种。语种按照 12 大语种分类即可。

数据来源：专题筛选后的采集信息，经过言论语种算法处理，再进行信息展现。

(14) 上网特征



气泡图逻辑三要素。横轴是时间，纵轴是日期，发文数是气泡大小。

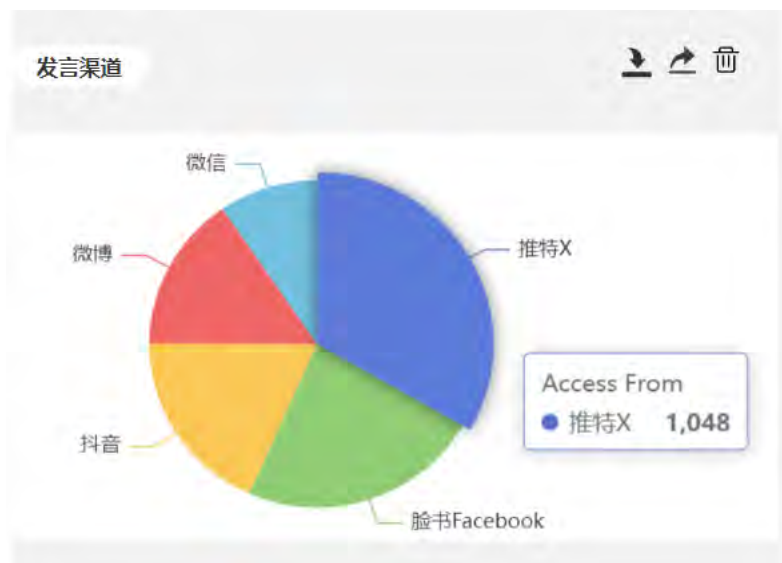
数据来源：专题筛选后的采集信息，经过言论时间算法处理，再进行信息统计展现。

(15) 提及热词



数据来源：专题筛选后的采集信息，经过言论热词算法处理，再进行信息统计展现。

(16) 发言渠道



数据来源：专题筛选后的采集信息，经过言论发言渠道算法处理，再进行信息统计展现。

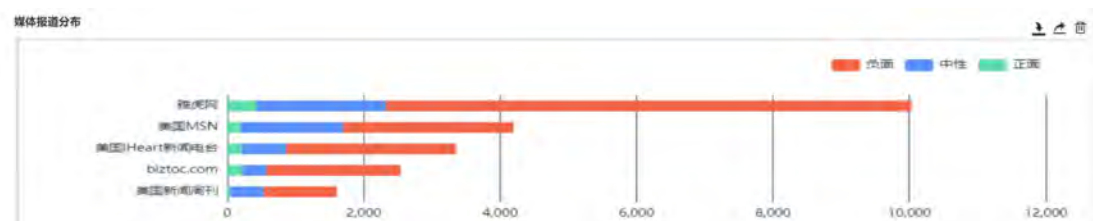
(17) 相关活动



相关活动是复合组件，按照左 2 右 1 的布局。展示主体人物活动的时间发布，活动详情和活动类型。

数据来源：专题筛选后的采集信息，经过活动算法处理，再进行信息统计展现。活动算法相当于谓语提取。

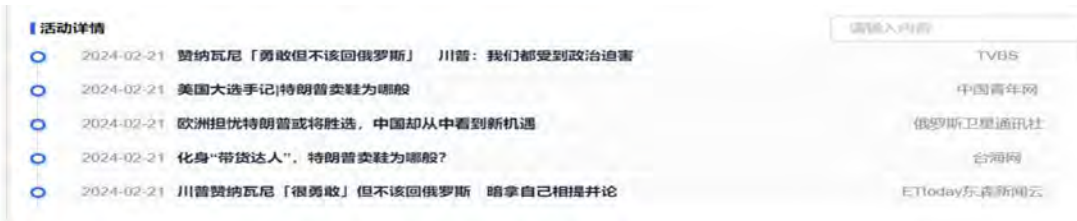
(18) 媒体报道分布



- 每屏显示 TOP5（暂定，以 UI 设计条数为准）。

数据来源：专题筛选后的采集信息，经过来源媒体抽取处理，再进行信息统计展现。

(19) 媒体报道详情



- 支持关键词搜索，按报道时间降序排列。
- 显示日期、报道内容、报道机构。
- 每屏显示 10 条（暂定，以 UI 设计条数为准），超出后上下滚动显示。

数据来源：专题筛选后的采集信息，经过来源媒体抽取处理，再进行信息展现。



在拖入第一个组件或者选择模板后，在底部会出现添加设置。可设置每行组件个数。最多是 3 个。

3.2.1.6 美国大选模板组件

整体大选模板的目录架构，如图所示

美国大选专题

- 大选热度趋势
 - (1) 选举概述
 - (2) 热点新闻
 - (3) 选举日程
 - (4) 事件热度
- 关注大选的人
 - (1) 官方人物表态
 - (2) 大V表态
 - (3) 民众表态
- 关注大选的组织
 - (1) 政府组织表态
 - (2) NGO表态
 - (3) 组织言论热词
- 选举流程
 - 党内初选
 - 全国党代会
 - 总统竞选
 - 全民投票
 - 总统就职

1、选举热度趋势

包括 4 个组件。

(1) 选举概述

选举概述

2024年美国大选^①是美国举行的总统选举活动，美国实行总统制，每四年举行一次总统选举，总统任期4年。美国2024年总统选举时间为2024年11月5日

提醒：文本类组件，支持编辑，导出、删除。重复功能不再描述。
公司没有事件库，随着项目建设，开始积累 AQ 领域的事件库。



数据来源：
组件标题为“选举概述”，
文本框内容固定为以下格式。

组件结构	数据来源	示例	备注
年份	创建时系统时间	2024 年	
监测主体	创建专题时必须填	美国大选	
是	组件写死		
国家/地区	对监测主体进行分词提取	美国	若分词库不好实现，则在创建专题时，增加地点信息。
举行的选举活动	组件写死		
美国实行总统制，每四年举行一次。美国 XX 年总统选举时间为	组件写死		由地点来触发组件词库，内置主要国家、地区的选举基本信息。比如韩国总统制，日本首相制。
2024 年 11 月 5 日	选举日程抽取		

(2) 热点新闻



组件结构	数据来源	示例	备注
标题	写死	热 点 新 闻 TOP10	后续不再重复描述标题，见原型。
参选党派组件	单选下拉，内置国	民主党、共和党	内置主要国家的参选



	家与党派关系库。		党派。 二期再支持来源数据配置。
热点话题	多选下拉。20 个国家安全观+两派竞选纲领议题。	政治安全、国土安全、军事安全、经济安全、文化安全、社会安全、科技安全、网络安全、生态安全、资源安全、核安全、海外利益安全、生物安全、太空安全、极地安全和深海安全	下拉数据来源两部分。 二期再支持来源数据配置。
信息列表			点击弹窗，进入信息分析页。

数据来源：在专题监测词筛选下，根据采集时信息阅读量排序。默认显示 TOP10，可上下滑动展示。

备注：所有下拉组件，均支持多选操作，方便用户对比！组件选择后，本期不支持用户更改组件样式，比如不这次从折线图改为柱状图等。

(3) 选举日程



整体支持滚动显示，中间轴是日期，左侧是民主党，背景改为蓝色（蓝党）。右

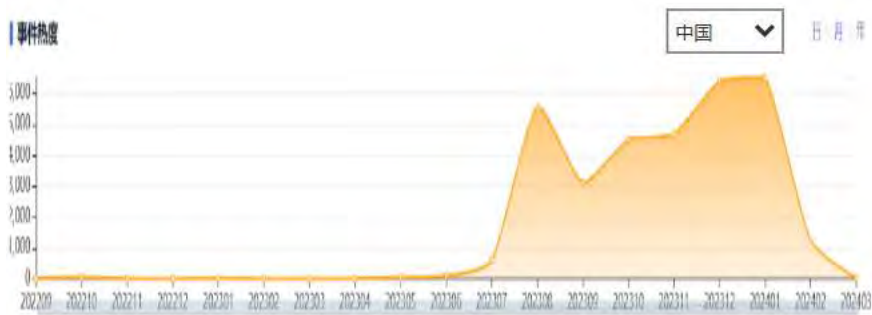


侧是共和党，背景改为红色（红党）。

数据来源：要求采集平台，定向采集。积累到 AQ 行业事件库。
美国 ABC 新闻网站大选板块。

[选举新闻、直播视频和更新 - ABC 新闻 \(go.com\)](#)

(4) 事件热度



事件热度：将专题筛选后的信息列表，按照时间进行统计展现。
纵轴是热度（数量），横轴是时间。支持切换日、月，没有年。

支持下拉选择国家。国家为五常+日韩。

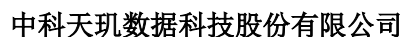
再次备注：所有下拉组件，均支持多选操作，方便用户对比！

2、关注选举的人

(1) 关注选举的政要排行

关注选举的政要排行				
本日 本周 本月 本年				
序号	人物	事件数	关注指数	
1	普京	251	3.53	言论热词
2	欧盟	131	2.17	言论热词
3	普京	110	2.15	言论热词
4	岸田文雄	79	1.81	言论热词
5	联合国	103	1.58	言论热词

组件结构	功能	备注
序号		只显示 top5
人物	人物+头像	头像来自内置的人物库



事件数	统计周期内，标题+正文中包含人物（多语种）的信息列表数。	
关注指数	（个人事件数/前 5 事件总数）*10	占比情况，量程为 0-10，最多取两位小数。
言论热词	统计周期内，人物发言热词词云展示。	点击按钮，弹窗
统计周期切换	按日，按周，按月	去掉按年。

言论热词弹窗如下



（2）关注选举的政要排行

关注选举的时政大V分析

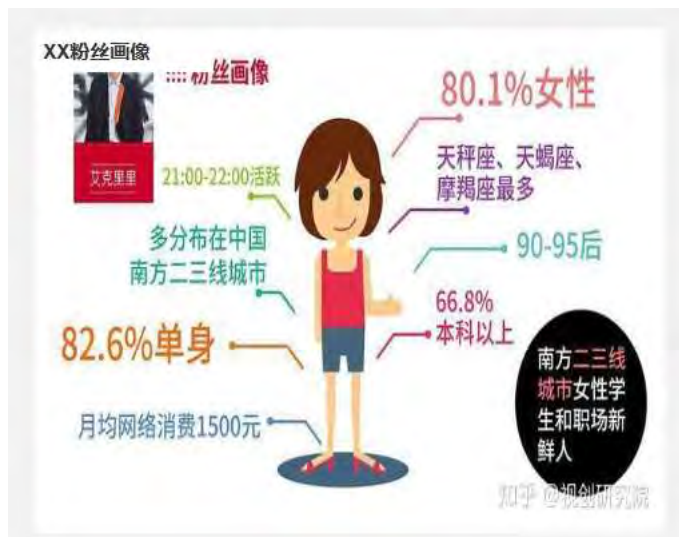
关注选举的时政大V分析				脸书	推特X	微博	微信
排行	相关大V	发文数	点赞	评论	转发		
1	胡锡进	100	100	100	100	粉丝画像	
2	司马南	80	80	80	80	粉丝画像	
3	卢克文	60	60	60	60	粉丝画像	
1	胡锡进	100	100	100	100	粉丝画像	
2	司马南	80	80	80	80	粉丝画像	
3	卢克文	60	60	60	60	粉丝画像	

组件结构	功能	备注
序号		只显示 top5
相关大 V	人物+头像	头像来自内置的人物库 在监测词筛选下，将信息列表中的大 V 抽取出来。
发文数	显示总数	
点赞数	显示总数	
评论数	显示总数	
转发数	显示总数	
粉丝画像		点击按钮，弹窗。画像维度见下文。

统计渠道切换	境内外主流 4 个社交平台
--------	---------------

粉丝画像弹窗如下：

画像维度包括性别，年龄段（10 年为一段），学历，职业、活跃时间段（按小时统计），单身比例。



（3）社媒情感分析

社媒情感分析



在监测词筛选下，将信息列表中的信息情感抽取出来，并按四大主流渠道进行划分。按日统计展现。

3、关注选举的组织

（1）关注选举的官方组织排行

关注选举的官方组织排行

英国 本日 本周 本月 本年

序号	官方组织	事件数	关注指数	对华态度
1	华约	251	3.53	
2	美步枪协会	131	2.17	
3	美籍华人协会	110	2.15	
4	美籍华人协会	79	1.81	
5	美籍华人协会	103	1.58	

展现逻辑同上。

官方组织定义：公司已有组织库，比如美国步枪协会。后续与客户进行业务梳理，是否有重点关注组织。

(2) 关注选举的 NGO 排行

关注选举的NGO排行

本日 本周 本月 本年

序号	NGO	事件数	关注指数	背后金主
1	华约	251	3.53	背后金主
2	美步枪协会	131	2.17	背后金主
3	美籍华人协会	110	2.15	背后金主
4	美籍华人协会	79	1.81	背后金主
5	美籍华人协会	103	1.58	背后金主

展现逻辑同上。

NGO 组织定义：公司已有组织库，比如美国国家民主基金会。后续与客户进行业务梳理，是否有重点关注组织。

(3) 组织热词

政府组织和 NGO，发言中提到的词云。



4、党内初选

(1) 初选投票统计



组件说明

- 支持两党切换展示
- 支持显示更新时间，美东。



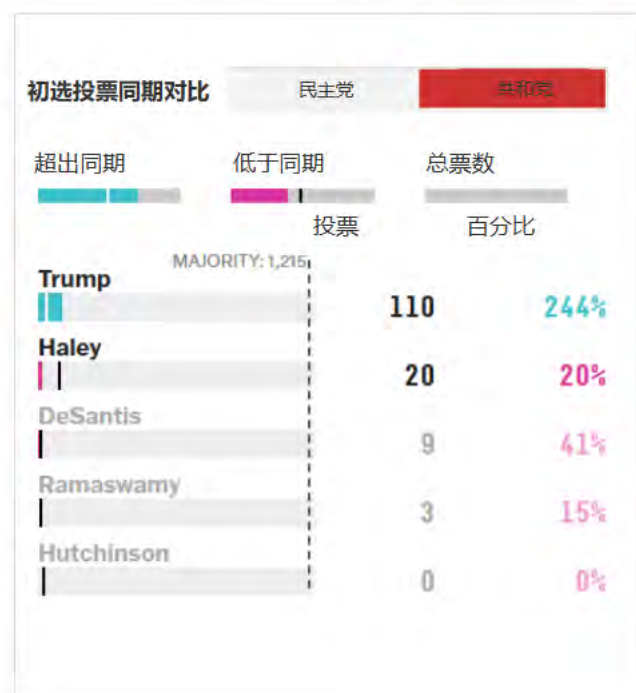
- 支持实时显示候选人的投票数，降序展示。
- 支持显示提名线，超出提名线的候选人将成为党内候选人。

数据来源：对于权威信息的展示，要求采集平台，定向采集。积累到 AQ 行业事件库。

美国 ABC 新闻网站大选板块。

[选举新闻、直播视频和更新 - ABC 新闻 \(go.com\)](#)

(2) 初选投票同期对比



组件说明：

超出同期，显示人气比往期更旺盛。

投票和百分比的位置不准，应该在 110 和 244% 的上方。

数据来源：要求采集平台，定向采集。积累到 AQ 行业事件库。

美国 ABC 新闻网站大选板块。

[选举新闻、直播视频和更新 - ABC 新闻 \(go.com\)](#)

(3) 初选选举日历

选举日历			民主的	共和
初选和预选会议				
1月15	爱荷华州	 王隆	29 票	
1月23	新罕布什尔州	 王隆	13 票	
2月6	内华达州	等待结果...		
2月8	内华达州	 王隆	26 票	
	维尔京群岛	 王隆	4 票	
2月24	南卡罗来纳州	 王隆	47 票	
2月27	密歇根州	 王隆	0 票	
3月2	爱达荷州	 王隆	13 票	
	密歇根州	等待结果...		
	密苏里州	 王隆	54 票	
3月3	哥伦比亚区	 哈莉	19 票	
3月4	北达科他州	 王隆	29 票	
3月5	阿拉巴马州	50 名代表		

数据来源：对于权威信息的展示，要求采集平台，定向采集。积累到 AQ 行业事件库。

美国 ABC 新闻网站大选板块。

[选举新闻、直播视频和更新 - ABC 新闻 \(go.com\)](https://abcnews.go.com/elections/)

(4) 认识候选人



认识候选人

民主党

共和党

其他

选择每个候选人，以了解更多关于他们在关键问题上的立场。



乔·拜登



迪恩·菲利普斯



玛丽安·威廉姆森



岑克·维吾尔语

点击候选人头像时，弹窗显示该人物画像。画像内容采用人物画像模板。

(5) 候选人相关事件



候选人相关事件

特朗普 ▼

对华事件 ▼

▼

发表言论 (9)

时间	描述
2024-02-26	罗凯文致词恭喜赖副总统...
2024-02-26	罗凯文并指出，美国右派...
2024-02-26	赖副总统表示，在总统蔡...
2024-02-26	罗凯文表示，身为美国德...
2024-02-26	传统基金会访团拜会赖清...
2024-02-26	赖副总统说，台湾未来面...
2024-02-26	中央社副总统赖清德今天...
2024-02-26	他说，当然选票是台湾人...
2024-02-26	赖副总统感谢传统基金会...

▼

会面 (3)

时间	描述
2024-02-26	赖副总统、民进党总统当...
2024-02-26	传统基金会访团拜会赖清...
2024-02-26	中央社副总统赖清德今天...

组件说明

- 事件类型按照 Gdelt 的事件分类，分为 20 个大类。后续与客户讨论，是否有其他事件分类。
- 人物下拉，选择初选党派得票前 3 的人物。在总统竞选阶段，还支持选择选举团队，（团队包括副总统、发言人，助手等。）
- 事件下拉，支持选择对华事件、对俄事件、对朝事件等等，与中国接壤的 14 个地缘国家，全部事件。对华事件，是指描述中提到中国或 China，其他类似。

再次备注：所有下拉组件，均支持多选操作，方便用户对比！

5、全国党代会

(1) 提名总统、副总统候选人



(2) 竞选纲领

竞选纲领



组件说明：

- 展示竞选口号：MAGA
- 展示纲领议题和主要内容，与中国以及中国国家安全观相关的，字体标红显示。比如经济，太空、极地等安全字眼



数据来源：对于权威信息的展示，要求采集平台，定向采集。积累到 AQ 行业事件库。

- 美国 ABC 新闻网站大选板块。[选举新闻、直播视频和更新 - ABC 新闻 \(go. com\)](#)
- 或者党派竞选网站。后续进行业务调研。

(3) 竞选纲领历史对比

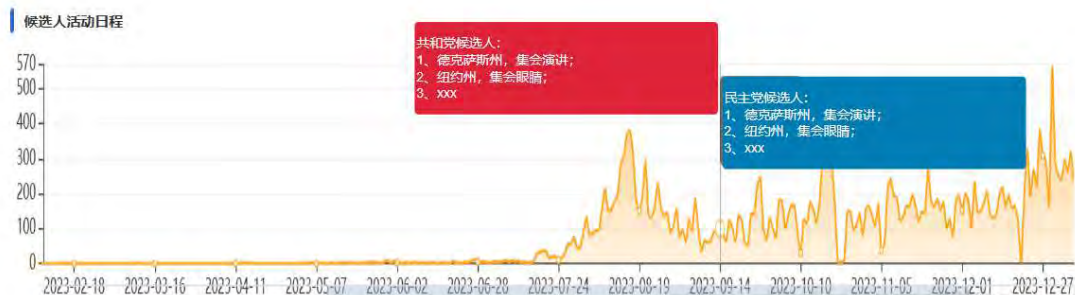
竞选纲领历史对比		民主党	共和党
2020年	实施	2016年	实施
1. 废除《奥巴马医改》，扩大个人账户使用。	败选	1. 废除《奥巴马医改》，扩大个人账户使用。	✓
2. 加强边境安全，修订移民法律。	败选	2. 加强边境安全，修订移民法律。	✓
3. 减少联邦政府的规模，取消环保法规。	败选	3. 减少联邦政府的规模，取消环保法规。	✗
4. 保护美国的国家利益，强硬对抗中国。	败选	4. 保护美国的国家利益，强硬对抗中国。	?
XXXXXXXX	败选	XXXXXXXX标注与中国国家安全观相关的 (政治安全、国土安全、军事安全、经济安全、文化安全、社会安全、科技安全、网络安全、生态安全、资源安全、核安全、海外利益安全、生物安全、太空安全、极地安全和深海安全)	✗

组件说明：与中国以及中国国家安全观相关的，字体标红显示。比如经济，太空、极地等安全字眼

数据来源：要求采集平台，定向采集党派竞选网站。若采集不到，则人工梳理，再导入数据库。实施状态显示议题在胜选后，是否实施。不确定的则用问号显示。

6、总统竞选

(1) 候选人活动日程



纵轴是活动数量，是两党候选人之和。鼠标悬停时，左侧显示共和党候选人当天日程。右侧是民主党。

(2) 候选人活动热力图



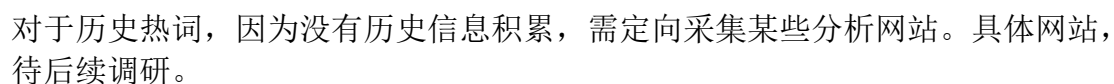
(3) 选情预测



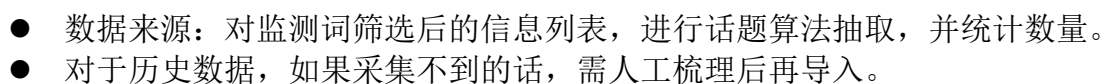
数据来源：要求采集平台，定向采集民调网站、标普数据、博彩网站。积累到AQ 行业事件库。

- 具体网站地址，后续进行业务调研。
- 对于历史预测，因为没有历史信息积累，需定向采集某些分析网站。具体网站，待后续调研。

(4) 候选人热词



(5) 大选话题 top10



(6) Top10 话题溯源分析

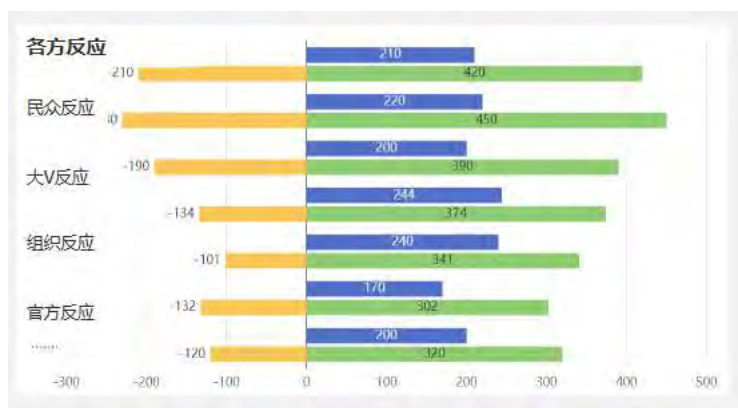




组件架构	功能描述	备注
下拉框	选择话题，默认显示 top1	单选
各方反应		点击按钮，弹窗
节点	包括时间、信源名称、国家、标题。 点击标题，弹窗进入信息分析页面	节点数过多时，支持左右滑动。默认显示前 5 个。

备注：

- 节点支持导出、导入，方便用户修改抽取的信息，形成真实溯源展现。
- 报道或转发的媒体，大 V 可能数量很多，组件要求上榜的条件是阅读量超过 10000。
- 各方反应弹窗



左侧橙色是负面统计，右侧绿色是正面统计，右侧蓝色是中性统计。

官方定义：是指政府部分机构，比如外交部等。来自采集大库中的官方库。

组织定义：政府组织和 NGO 组织。

大 V 定义：全网粉丝超过 100 万。

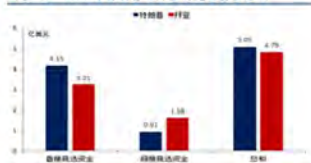
民众反应定义：对话题信息列表中，评论数据进行情感识别。

备注：需技术经理确认是否有相关数据库或者标签。如果采集平台没有，则需按照定义规则，从采集大库单独维护相关数据库，作为行业积累。

(7) 竞选资金分析

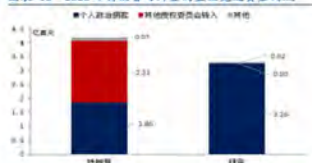
竞选资金分析（时间改为2024）

图表 17 2020 年特朗普与拜登的竞选资金总额



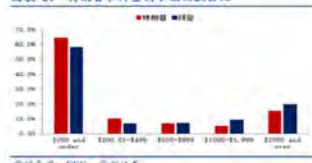
资料来源: Center for Responsive Politics. 华尔街日报, 数据截至 2020 年 8 月 21 日

图表 19 2020 年特朗普与拜登的直接竞选资金构成



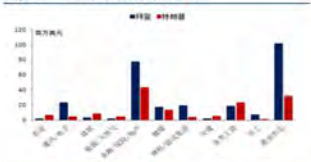
资料来源: FEC. 华尔街日报

图表 21 特朗普和拜登的小额捐款占比



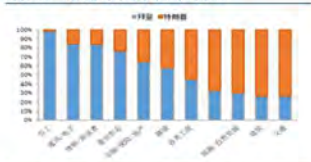
资料来源: FEC. 华尔街日报

图表 29 捐款者的部门分布, 金额



资料来源: Center for Responsive Politics. 华尔街日报

图表 30 捐款者的部门分布, 相对占比



资料来源: Center for Responsive Politics. 华尔街日报

图表 32 特朗普和拜登竞选的 PAC 列表

名称	PAC 名称	总票数 (P)	名称	PAC 名称	总票数 (P)
America First Action	AFAC PAC	880.1	Proclaim USA Action	Pro PAC	380.0
Committee to Defend the President	CDP PAC	120.2	American Bridge 21st Century	AB21 PAC	830.1
Great America PAC	GAP PAC	640.7	United the Country	UC PAC	201.8
American Conservative	ACM PAC	282.7	Paragon	APAC PAC	1408.3
Black Americans in the White House	BAWH PAC	240.3	America on Frontlines	AFAC PAC	41.3
Proton	Pro PAC	173.4	Defenders PAC	DEF PAC	13.8
Top Party PAC	TPP PAC	141.1	Right Side PAC	RS PAC	17.3
Rebuilding America Now	RAN PAC	101.4	Defend America PAC	DAPAC PAC	9.8
Partners of America PAC	POA PAC	108.3	AFAC	AFAC	1000
AFAC	AFAC	1000	AFAC	AFAC	1000

资料来源: Center for Responsive Politics. 华尔街日报

数据来源: 要求采集平台定向采集, 美国联邦选举委员会官网: [首页](#) | [FEC 系列](#)
组件说明: 资料来源不再显示。显示实时统计信息。标题即组件名称, 去掉图表
2020 年特朗普与拜登。

7、全民投票

(1) 选举结果地图



- 数据来源: 要求采集平台定向采集, 美国 ABC 新闻网站大选板块。 [选举新闻、直播视频和更新 - ABC 新闻 \(go.com\)](#)

(2) 往年摇摆州分析

往年摇摆州分析



当选总统: 里根, 布什, 克林顿, 小布什, 奥巴马, 特朗普, 拜登

数据来源：对从 1988 年开始，对每个州历届选举结果进行对比，只要发生过变更的即视为摇摆州。

如果采集不到历届选举结果，可人工维护数据。

(3) 支持者画像



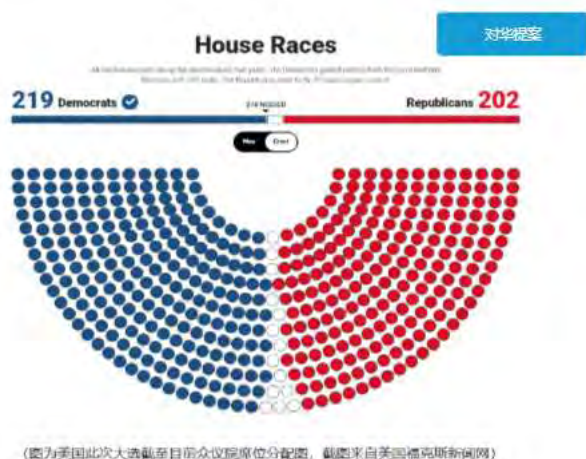
- 数据来源：要求采集平台定向采集，美国多家民调机构数据。具体网站，后续调研业务补充。

(4) 各国政要祝贺顺序



注：起始时间是 2024 年 11 月 5 日。切换成美国东部当地时间。

(5) 众议院席位



- 数据来源：要求采集平台定向采集，具体网站后续根据业务调研。

备注：美国众议院共计 435 个席位，所有席位统一每两年改选一次。参议院共计 100 个席位每两年选举其中 1/3



(6) 参议院席位
同上。

8、总统就职典礼

(1) 总统演讲摘要



(2) 各国官方报道时间



3.2.1.7 舆情类模板组件

与天罗产品，保持一致。后续根据业务调研，持续补充。

3.2.1.8 组织类模板组件

无特殊组件，与人物组件类似。后续根据业务调研，持续补充。

3.2.2 模板管理

➤ 原型如下：



3.2.2.1 模板列表页

- 整体分为上下两部分。
- 上面是系统模板，只有管理员角色的账号，可以添加系统模板。
 - 下面是个人自定义模板，管理员看不到普通用户的模板。
 - 列表页支持对模板的增删改查功能。最顶部支持关键词搜索，搜索范围是模板名称和模板描述。
- 模板采用卡片展示形式。

模板组件	功能描述	备注
封面图	系统内置	自定义模板的封面都一样，系统模板看 UI 图。管理员新增系统模板时，随机选择内置的封面图
模板名称		不超过 20 个字
模板描述		
预览图标	点击后进入预览查看	鼠标悬停时，显示图标
编辑图标	点击后进入编辑	
删除图标		按照统一规范，删除需二次确认。

包括封面图（系统内置），模板名称、模板描述。

3.2.2.2 模板详情页

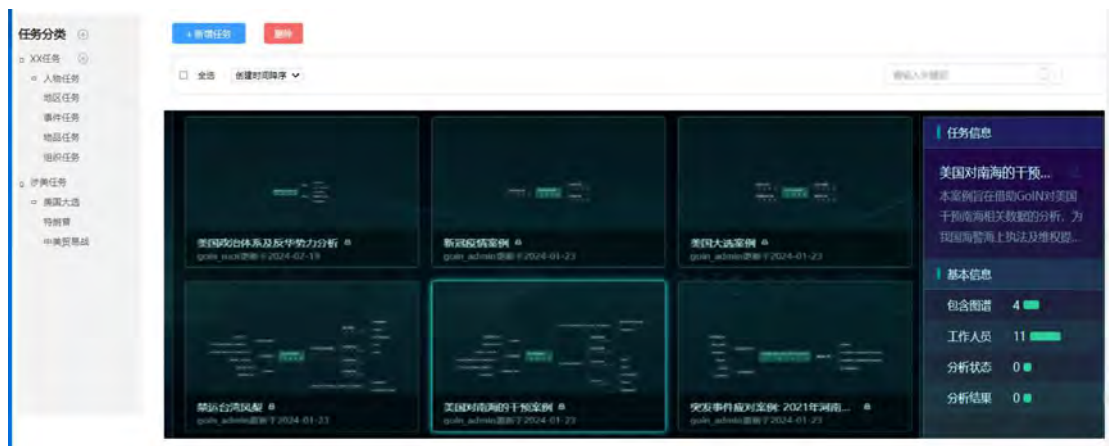
鼠标悬浮到模板卡片时，自动显示相关图标按钮，点击预览按钮，可查看模板详情。点击编辑按钮，可修改模板。个人只能修改个人创建的模板，无法修改公共模板。



3.3QB 服务

QB 服务包括任务管理、图谱管理和 QB 分析。QB 分析支持关联分析、时空分析。

3.3.1 任务管理



同一系统的设计逻辑保持一致，类似特色专题，支持任务的分类。任务的增删改查。

- 支持全选删除功能。
- 支持按创建时间排序
- 支持按关键词搜索，搜索范围是任务名称和描述。
- 选中单一任务后，右侧支持显示任务信息和基本信息。任务信息包括任务标题和描述。此处支持修改。基本信息包括工作人员，分析结果和分析截图。

3.3.2 图谱管理

需求未定，用户希望操作步骤极简。边建设，边沟通。

3.3.3 QB 分析

3.3.3.1 关联分析

需求未定，用户希望操作步骤极简。边建设，边沟通。

3.3.3.2 时空分析

需求未定，用户希望操作步骤极简。边建设，边沟通。

3.4 报告管理

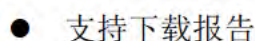
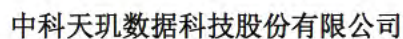
- 支持对报告的查询、搜索、预览、下载等功能。



- 支持报告的预览功能



- 支持报告编辑功能，支持对组件内容的编辑。



完全复用采集平台的网络数据底图功能。直接内嵌相关功能模块,无须二次开发。

采用公司统一控制平台，包括机构管理，用户管理，角色管理。

4.1软件环境需求

6



4.2 硬件环境需求

确定在不同并发量同时执行任务的最小配置。

需求名称	50-80 人并发
CPU	24 核
内存	管理节点：32G 计算节点：128G
硬盘	管理节点：500T 计算节点：2T

5 非功能需求

5.1 信息采集

客户需求未完全确定，重点采集网站待后续梳理。下同。

需要杨晓菊老师全部的重点人物库、组织库、机构库，全部的业务标签体系。

序号	性能项	需求描述
1	采集信源	
2	采集规模	
3	采集频率	
4	回溯采集	

5.2 文本处理

序号	性能项	需求描述
1	情感判断	支持正、中、负情感判断，准确率不低于 85%
2	文本分类	支持热点分类，准确率不低于 80%



3	文本聚类	支持热点聚类，准确率不低于 90%，聚合计算时间不高于 15s
4	主题抽取	支持对内容主题的聚合和抽取，准确率不低于 80%
5	内容去重	支持基于相似度计算的内容去重，准确率不低于 90%，去重计算时间不高于 15s
6	实体抽取	支持人名、地名、机构名等实体抽取，准确率不低于 90%
7	内容抽取	支持对时间、稿源、标题、征文
		后续还会持续补充。

5.3信息分析

序号	性能项	需求描述
1	实时分析	支持实时分析，可视化展现

5.4数据存储

序号	性能项	需求描述
1	存储周期	正式服务用户专题数据永久存储，其它数据存储周期 2 年以上

5.5响应时间

序号	性能项	需求描述
1	启动功能响应时间	低于 2s
2	数据保存响应时间	低于 5s
3	任务提交响应时间	低于 5s
4	简单查询响应时间	低于 5s
5	复杂查询响应时间	低于 10s

5.6系统并发

序号	性能项	需求描述
----	-----	------



1	并发用户	同时满足 150 人在线用户请求
---	------	------------------

5.7 系统安全

序号	性能项	需求描述
1	备份	系统应定期备份
2	日志	应用软件运行时须有完整的日志记录，以便及时审查日志文件，分析告警信息，掌握运行状况，对系统可能发生的故障做好应急方案。
3	权限	用户管理、权限管理应充分利用操作系统和数据库的安全性。提供统一有效的安全措施来进行权限限制，防止内部人员误操作或越权，每一个环节必须进行安全性检查，不能越权访问。对每次用户登录要求进行日志记录，对三次（参数可以改变）以上的登录尝试进行锁定（锁定的时间、解锁方式，参数可调），并发出告警。
4	开发与测试	开发、测试系统与生产系统要严格分开。
5	软件修改、升级	软件程序的修改或增加功能时，须提出修改理由、方案、实施时间，报上级主管部门批准；程序修改后，须在测试系统上进行调试，确认无误经批准后方可投入生产应用；软件修改、升级前后的程序版本须存档备查，软件修改、升级时须有应急补救方案
6	检验纠错	数据传输、处理应有检验、核对功能和较强的纠错功能。
7	监督告警	具有安全监督功能、故障和异常告警功能。

5.8 数据安全

序号	性能项	需求描述
1	接口安全	系统采用三层（多层）体系结构，所有的数据操作都通过中间层来处理，客户端不能直接与数据库相连。外围系统也不能直接存取本系统的数据，必须通过系统提供的接口服务，来进行信息的交换。提供完整的系统数据传送监控和报警处理。
2	数据传送	关键数据在传输过程中，必须加密传送，以保证信息的安全。



3	数据备份和恢复	可提供各种方式、各种介质的数据备份功能。各种原始业务数据均应保留备份、各种过程业务数据应保留备份、操作系统、数据库应定期备份。提供对一般性数据和交易性数据的保护，使得在任何一部分子交易失败后，整个交易的所有消息都回滚，保证数据的完整性。 系统配置型数据，按月全量备份更新，按周增量备份； 用户配置型数据，按周全量备份更新，按天增量备份； 内容数据，支持冷热数据备份，超过 1 年数据即回滚成冷数据；
4	数据范围	登陆时检查 IP 地址的合法性，同时记录 IP 地址的使用情况。

5.9 升级与扩展

序号	性能项	需求描述
1	可扩展	设计具有技术先进性和前瞻性,充分考虑系统升级、可拓展等问题，保证软件新旧版本的平稳过渡。
		保证主机系统，主机系统、网络系统、数据库系统、操作系统、业务应用系统和外围设备在将来能够顺利升级或扩容，且不影响正常的系统运行。
		用模块化设计,在整个系统正常运行的情况下，能够随时增加或取消接入用户。
		新的软件模块即插即用，新功能、新业务的增加能够在不影响系统运行的情况下实现
		可动态增加服务器，增加处理能力，能够实现负载平衡
2	版本管理	为了适应不同用户的不同处理流程的需要，系统应该支持使用版本组控制不同地方使用不同的版本组，在同一版本组内使用最新的版本，保证版本的自动更新。
		局部功能新旧版本的使用、并存、排斥要可控。局部功能新旧版本的使用、并存、排斥要可控。

5.10 可维护

序号	性能项	需求描述
1	可维护	高内聚、低耦合设计，增强代码可修改性。
		确保软硬件的可移植性。



5.11可兼容

序号	性能项	需求描述
1	可兼容	系统所有交互均基于 Web 方式完成，所有页面跟功能需兼容 IE10 以上、Chrome 、Firefox 等主流浏览器

6 附录



3. 系
指挥平台
——课

会议纪要

会议主题：纵横系统项目工作进展汇报沟通会

会议时间：2023年10月19日 上午9:00-上午12:00

会议地点：通研院林萃桥办公区一层大会议室

参加人员：郭杰，三家课题相关人员（课题一：荆向月，尚林林；课题二：仝春燕及算法人员；课题三：刘训瀚）

一、各家工作汇报

课题一：

1、汇报内容：①政治心理画像算法及数据研究进展；②系统自测情况汇报；③课题三所需仿真数据集采集提供（天玑）。
郭所反馈：围绕舆论场的线索深入挖掘、与地理信息结合，月底沟通新进展。

课题一采集：

郭所反馈：①对大数据的采集和管理提出一些要求，数据资源规范、数据源格式、数据接口提交给相应职能部门（主要是历史数据），不能作为单纯系统支撑；②需提供数据采集防溯源系统性指导，以文档为主。

课题二：

郭所反馈：舆图本身构建受限于模型复杂度和数据资源条件，可



用性和其他课题沟通的不是很到位，框架需要进一步完善（至少完善到 80%，能够说明问题）。

课题三：

郭所反馈：①一定保证资源投入；②账号的数据做一些心理刻画，账号养殖人设更丰满、行为更正常、更拟人化，大模型做一些探索；③课题三长期预置 TW 重点、立陶宛相关的话题和账号。预置培养一些立陶宛的账号。（考虑账号风格、账号个人说明再完善下凸显真实性）。

二、选题讨论

课题二汇报实验方案：以 Twitter 为信息投放实验平台，推高话题的传播声量。讲解推演方案（采集话题及传者、受众等，跨社区、跨领域传播）。——初步讨论确定话题平台及方向

三、后续任务

1. 模拟演练：尽快定话题，实战应用支撑（关于选题和策略方法通过演练验证其思路）。

郭所补充：选题不要涉政、可以着重考虑科技领域（投放角度围绕 M 和 T 政界、科技界、大 V。新兴技术的风险、大模型和智能机器人，如“特斯拉人性机器人”）；旅游美食慎选（对资源部署是一种浪费）。

2. 部署资源问题：课题三提供详细的资源配置参数供采购参考。



3. 系统风格问题：系统需单独新增一个页面突出态势感知，类似指挥平台，（领导觉得态势感知地图像十年前的风格，没有科幻感）——课题三主导，数据基座不变，展现样式调整。

4. 进一步和课题负责人沟通学术层面、技术层面进展和安排。
郭所补充：作为重点研发计划，明年八月份验收课题，计划一个大平台（展示能力和愿景）+若干工具的集合（可能最终实用的是小平台）。后续会和各家课题负责人展开新一轮的对接，希望各家从学术层面拔高，增加理论创新有突破或能够添彩的内容，充实和丰富成果包。

5. 宣传视频制作，要求：两分钟的视频，流程、逻辑、故事讲清楚。



中科天玑
GoLaxy

中科天玑数据科技股份有限公司 公司整体介绍

www.golaxy.cn





目录



01 公司介绍

02 技术能力

03 典型案例



Part 1

公司介绍





关于我们



中国科学院：中国自然科学最高学术机构

中国科学院，成立于1949年11月，全院拥有12个分院、100多家科研院所、3所大学（中国科学院大学、中国科学技术大学，与上海市共建上海科技大学）、130多个国家级重点实验室和工程中心、210多个野外观测台站，承担20余项国家重大科技基础设施的建设与运行。

中国科学院计算技术研究所：中国第一个专门从事计算机科学技术综合性研究的学术机构

中国科学院计算技术研究所，创建于1956年。主要研究方向和领域有信息处理、信息检索、网络安全、大数据处理、体系结构研究、智能技术研究、生物信息计算、虚拟现实技术等。研究所有计算机系统研究部、网络研究部、智能信息处理研究部3个研究部，下设16个研究机构，有国家重点实验室1个，国家研究中心2个，中国科学院重点实验室2个，15个分所与分部。

中科院计算所孵化企业

曙光

寒武纪

龙芯

晶上

联想

• • • • •

中科天玑数据科技股份有限公司：中科院计算所在大数据方向唯一的产业化平台

中科天玑数据科技股份有限公司，成立于2010年，依托中科院计算所网络数据科学与技术重点实验室、大数据分析系统国家工程实验室孵化的高新技术企业。专注在大规模数据感知获取、存储管理、分析挖掘与领域知识构建，面向“党、政、军”用户提供整体大数据智能化解决方案，是国内情报大数据分析领军品牌。



计算所科研布局

先进计算

系统研究部

- 计算机体系结构国家重点实验室 (2011年)
- 先进计算系统研究中心 (2013年)
- 智能处理器研究中心 (2016年)
- 微处理器研究中心
- 高性能计算研究中心
- 数据存储技术研究中心
- 计算机应用研究中心 (2011年)
- 高通量计算研究中心

网络计算

网络研究部

- 大数据分析系统国家工程实验室 (2017年)
- 网络数据科学与技术重点实验室 (2013年)
- 移动计算与新型终端北京重点实验室 (2012年)
- 网络技术研究中心
- 无线通信技术研究中心
- 专项技术研究中心 (2013年)

智能计算

智能研究部

- 中科院智能信息处理重点实验室
- 泛在计算研究中心
- 前瞻研究实验室

信息技术战略研究中心
(2011年)



国内唯一舆情分析领域三次荣获国家科技进步奖的单位



- 国家发展与改革委员会2017年批准设立
- 由法人单位中国科学院计算技术研究所负责建设
- 共建单位：
 - 中国科学院大学
 - 中国科学院计算机网络信息中心
 - 中科天玑数据科技股份有限公司
 - 曙光信息产业股份有限公司
 - 国创科视科技股份有限公司
 - 中科院深圳先进技术研究院
 - 中科院云计算中心

国家发展和改革委员会办公厅文件

发改办高技〔2017〕144号

国家发展改革委办公厅关于开展大数据分析系统国家工程实验室组建工作的通知

中国科学院办公厅：

你院报来《关于推荐院属单位申报大数据和互联网领域创新能力建设专项的函》（科发办函字〔2016〕88号）收悉。经研究，现函复如下：

一、原则同意由中国科学院计算技术研究所作为承担单位，联合相关单位筹建大数据分析系统国家工程实验室。

二、该工程实验室的主要任务：针对我国大数据分析能力弱、处理效率低、大数据分析技术应用门槛高等问题，围绕提升大数据分析系统分析能力和效率的迫切需求，建设大数据分析系统应用



中科院计算所“1+2+N”的全国布局推动国家大数据分析系统工程实验室分中心落地建设，形成围绕团队**核心业务布局**的落地方案。

各方向重点业务布局

- **网信方向：**中央网信办业务应用总体规划与建设，网信办大平台应用总集成。完成多个省级网信办系统建设：西藏、内蒙古、青海、甘肃、海南，天津等；
- **科技情报方向：**国家大数据分析系统工程实验室粤港澳大湾区（深圳）分中心，服务于大湾区对安全和科技大数据应用及大数据产业发展需求；
- **AQ方向：**与某厅建立大数据联合实验室，实现海量结构非结构化数据的综合治理、快速预处理、智能化分析、可视化展示；
- **公安方向：**与某公安联合成立警务大数据智能化联创实验室。打造具有地方特色的警务大数据平台；
- **纪检监察方向：**与某市纪委监委共同建立大数据监督技术实验室。面向纪检监察监督大数据方向重点研发监督工作标准规范、核心软件、专利技术。



关于中科天玑



成立于2010年，是中科院计算所在**大数据方向唯一**的产业化平台

- 国内**情报大数据分析**领军品牌
- 面向“**党、政、军**”用户提供整体大数据智能化解决方案

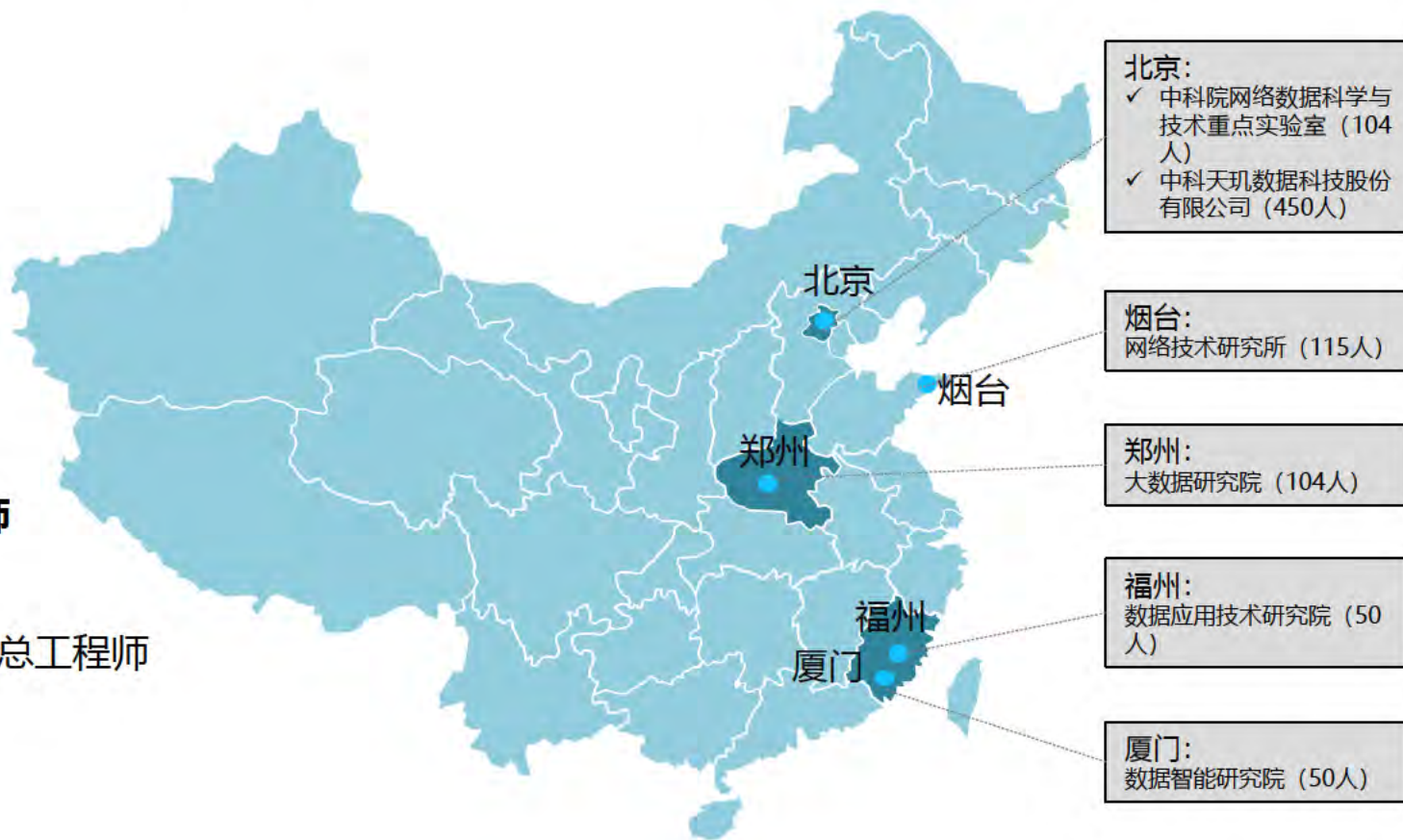


董事长 —— 程学旗

中科院计算所副所长、研究员、博士生导师

- 军委科技委大数据专家组首席科学家
- 网络数据科学与技术重点实验室主任、副总工程师
- 大数据分析技术国家工程实验室副主任
- 信息检索与内容安全专委会副主任
- CCF大数据专家委员会秘书长

天玑团队总人数近**900**人，**46%**具有高级职称，硕士、博士**100+**人，科研人员**500+**人





Part 2

技术能力



引入第三方数据需求迫切

- ①互联网采集数据（重点人数据）
- ②账号养号数据（开源情报）
- ③舆情数据（社交媒体，网站信息）
- ④攻坚数据（定点攻坚，拖库数据）

数据价值、质量、开放程度低

- ①数据未按照标准建设
- ②数据无法有效利用
- ③数据分散、割裂，对外服务能力弱
- ④数据重复采集，质量低下

数据分析能力弱

- ①依靠传统经验分析，以查询检索为主
- ②数据关联程度低，实时管控能力弱
- ③分析算法以规则判断、手动设置为主，智能化程度低

数据应用实战程度低

- ①大数据平台架构繁重，开发周期长，时效性差
- ②数据使用层层审批，使用不便

四个迫切需求

大数据智能化目标

- 引入多源异构数据，解决内网数据不足，提供差异化数据的有效补充
- 打破部门壁垒，汇集海量数据，构建具有“自主提炼”和“自主思考”能力的智能应用体系
- 智能化从“感知”向“认知”发展，提升研判、推理、预测等智能化水平，构建行业“智慧大脑”
- 由人力密集型向科技密集型转变，构建数据、情报、指挥、保障、行动一体化工作机制

三大技术方向规划

面向多层网络的多源数据集成能力



面向多源数据综合治理能力

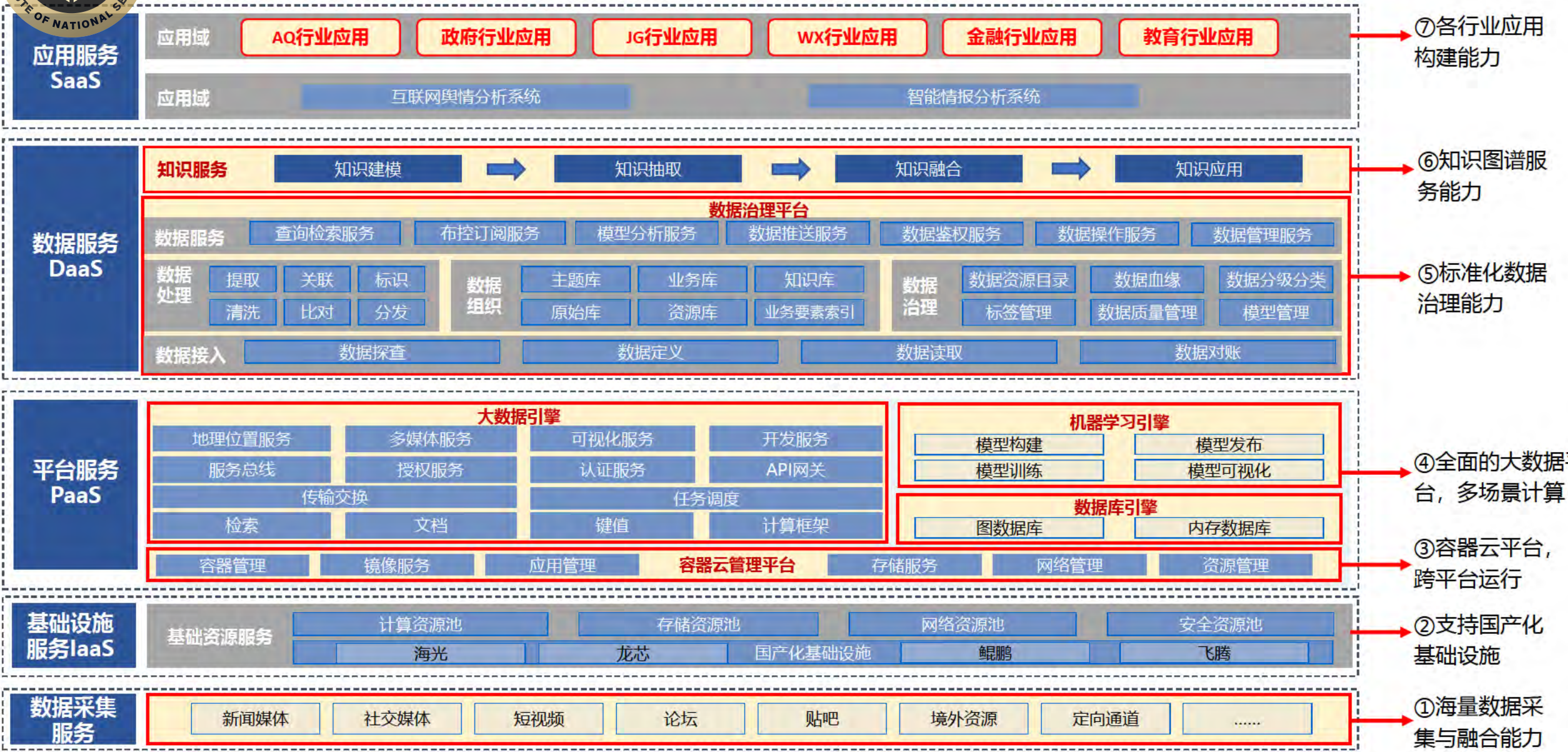


面向行业的智能分析能力





天玑大数据技术架构





天玑产品架构图



智能业务
应用

AQ行业应用

政府行业应用

JG行业应用

WX行业应用

金融行业应用

教育行业应用

智能支撑
引擎

自然语言理解文
本分析引擎

知识图谱引擎

大规模图分析引
擎

BDA智能分析
挖掘平台

数据中台

大数据存储管理平台



大数据治理加工平台

数据采集

互联网数据采集平台

开源情报获取

专项数据获取

情报 / 舆情分析服务

A stylized graphic of a constellation or network, consisting of several blue star-like nodes connected by thin lines, positioned above the text "Part 2.1".

Part 2.1

数据采集能力

采集信源全面覆盖全网各类国内外资讯媒体、自媒体、社交网络信息，每天从亿级的信息中采集、聚合分析网络开源信息，实现对领域热点、重大事件以及用户关注事件数据自动发现。同时能够集成用户自有情报数据和第三方采集数据，为客户提供多元化的数据集成服务

采集通道说明

采集通道	采集信息	现有采集能力
新闻	全局唯一文档id、信源通道ID、获取时间、发布时间、插入时间、更新时间、阅读数、评论回复数等 40余项	200万/日
微信	标题、内容、发布时间、URL、公众号id、公众号名称等 20余项	公众号文章采集量：180万/日 在看数、阅读数、评论：5千/日
微博	微博ID、微博内容、发布时间、用户ID、用户昵称、用户头像、微博URL、转发数等 30余项	微博采集量（原创+转发）：6000万/日 微博评论采集量：100万/日 关注与粉丝采集量：300万/日 推文采集量：400万/日 推文转推采集量：200万/日 评论信息采集量：200万/日
Twitter	人物名、人物屏幕名称、人物粉丝数目、人物好友数目、人物简介、人物创建时间、人物头像、人物位置信息、人物帖子数目等 50余项	Facebook监控用户量：800个用户 Facebook推文采集量：1万/日
Facebook	人物名、人物屏幕名称、人物头像、个人描述等 14项	用户跟推文不同，无采集量评估。
LinkedIn	用户名称、用户昵称、用户简介、用户性别、账号主页、用户头像url、用户的关注数、互粉书、收藏数、用户的粉丝数、账号状态、是否认证、认证类型、所在组织、更新时间等 15项	
Telegram	发布时间、消息id、内容、用户id、用户昵称、用户头像、性别、群头像url、所有群类型、所属群id、所属群昵称、所属群屏幕名等 13项	聊天信息采集量：3千/日
短视频	发布时间、视频url、用户ID、用户名、关注数、粉丝数、视频ID、转发数、收藏数、弹幕数、视频分享数、观看数、播放量等 30余项	4.5W/日
论坛	全局唯一文档id、信源通道ID、获取时间、发布时间、插入时间、更新时间、阅读数、评论回复数等 30余项	帖子：1万/日 帖子评论：1万/日
贴吧	全局唯一文档id、信源通道ID、获取时间、发布时间、插入时间、更新时间、阅读数、评论回复数等 30余项	14万/日

● 采集覆盖面广

- ✓ 采集信源覆盖面广，覆盖了新闻、APP新闻、境外新闻、境外社交、微信、微博、论坛、贴吧、博客、电子报、问答等10多种媒体类型，以及百度搜索、360搜索、必应搜索、搜狗搜索等主流搜索引擎
- ✓ **新闻监测的版块数达到十万级，境外新闻近万**，覆盖了所有主流新闻媒体和财经媒体；论坛监测网站数达到万级；微信监测公众号规模达到百万级；微博采集为6000万+；搜索引擎覆盖了百度搜索、360搜索、必应搜索、搜狗搜索等主流搜索引擎，金融、财经、经济等多种类型，万级以上的搜索关键词

● 采集数据量大

- ✓ 系统已积累了近5年的采集数据，积累数据规模200亿+，系统**每日新增数据量7000万+**
- ✓ 全天采集数据量在8000万-1亿



2.1 专项数据采集能力



自带全球人网数据

信源总量：**十万级**；
境内外新闻网站：**12272个** 论坛网站：**2167个**
新闻版块：**136803个** 论坛版块：**91730个**

历史数据：**千亿级** 现有知识图谱
文本：**400亿条** 详细人物图谱：**1500+**
短视频：**500万条** 概要人物图谱：**25万+**

HK: Twitter账号数据, 总量估计为**200万**左右
TW: 与政治XQ相关人物画像
涉蒙、涉藏人物、群体库, 含重点人物和重点关注群体

情报信源覆盖范围广阔

facebook twitter BBC REUTERS 境外FD网站
LinkedIn YouTube The New York Times WIKIDATA
智库网站
政府网站

21类信源类型, 覆盖所有国内外主流媒体和各类网站;
支持**5种**国内外主流搜索引擎。

境外人物组织资料信源

国外人员：**8类160+** 国外公司：**5类100+**

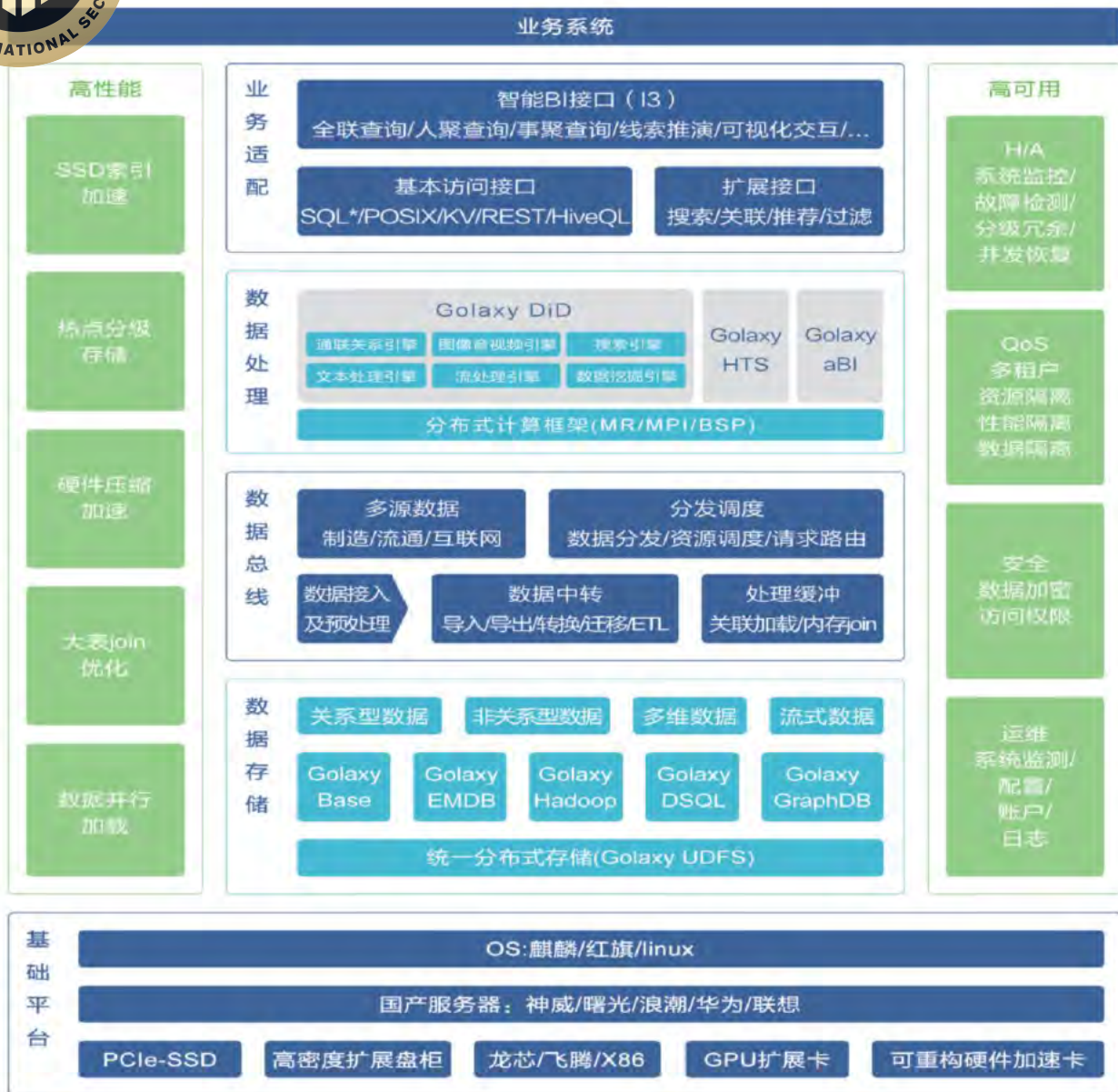




Part 2.2

数据中台

对采集的海量多源数据进行统一处理、加工、存储，实现数据资产化、服务化

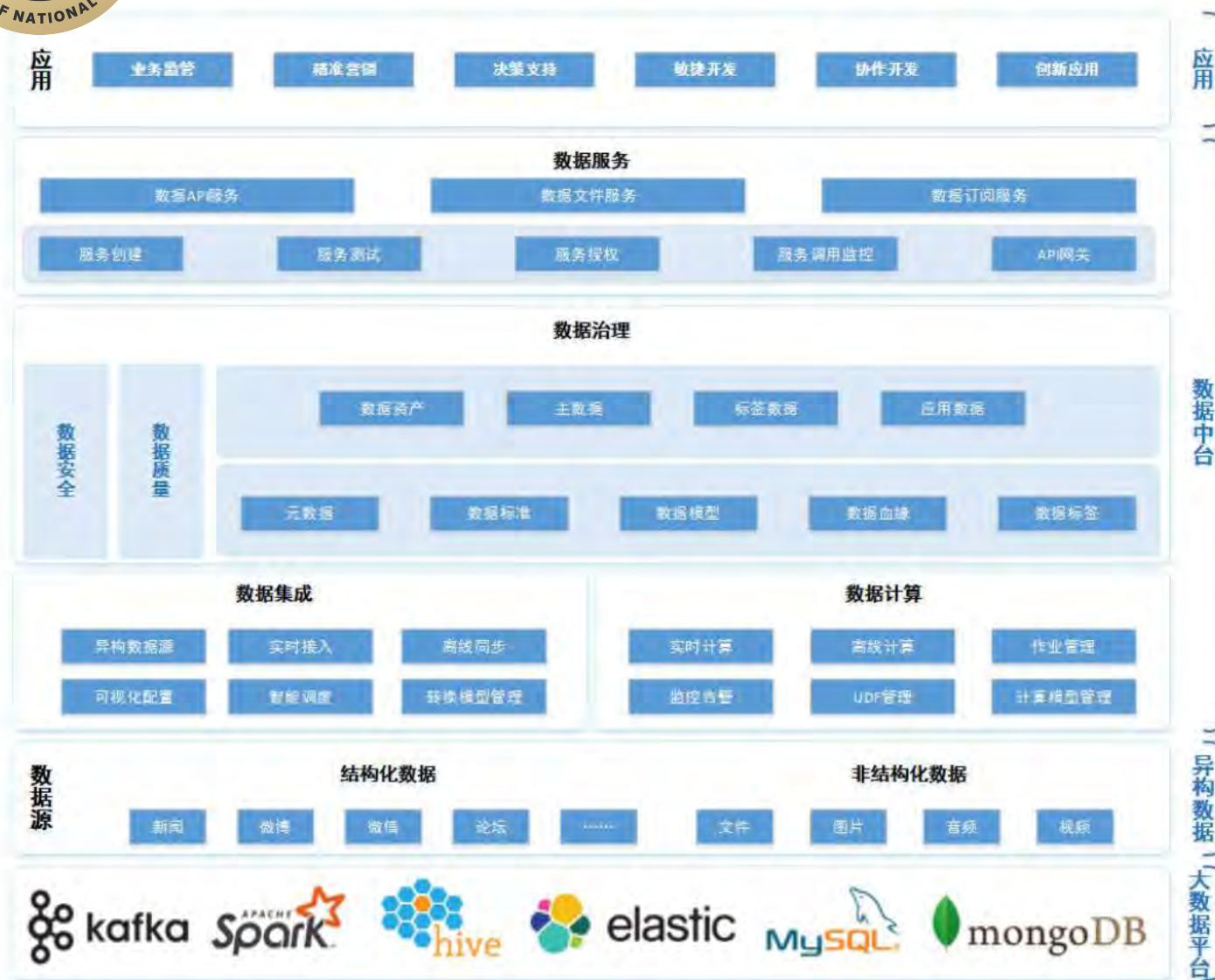


★ **平台概述:**
天玑大数据存储管理平台是一款分布式大数据平台软件产品套件，基于容器的基础平台，完全适配海光、鲲鹏、飞腾、龙芯等国产化硬件平台。

- ★ **产品特点:**
- ★ **海量数据全生命周期管理**
集成数据采集、缓存、分发、存储、查询、检索、统计、分析、挖掘、可视化等海量数据全生命周期管理功能
 - ★ **多种计算模式**
支持流式、离线、实时等多种计算模式
 - ★ **深度分析解决方案**
针对具体应用场景，提供各种专业的深度分析解决方案



2.2 数据中台——大数据治理加工平台BDP



★ 平台概述:

中科天玑大数据治理加工平台是一款针对大数据的全生命周期处理平台，能够整合多源异构数据，对数据进行预处理和清洗，按照业务要求对数据进行加工，并统一数据标准和口径，实现数据的资产化、服务化。

★ 产品特点:

一站式开发

从数据的采集、汇聚、融合、加工计算、治理、服务，提供全链路操作，无需切换多个工具

敏捷易用

可视化数据开发，操作流程敏捷易用，极大减轻了开发人员工作量，迅速响应需求变化和业务调整

数据驱动闭环

基于一站式开发，实现数据的汇聚、融合、资产化、服务化的完整闭环，真正实现数据驱动业务迭代和增长

统一SQL简便灵活

从数据加工分析任务到数据服务构建，均支持SQL方式，形式简洁，开发人员易于接受

动态水平扩展

平台具有极强的横向扩展的能力，可迅速扩展集群以承载更多业务



Part 2.3

智能支撑引擎

利用NLP、知识图谱、图计算、人工智能等计算对情报数据进行深入分析，构建智能化分析引擎

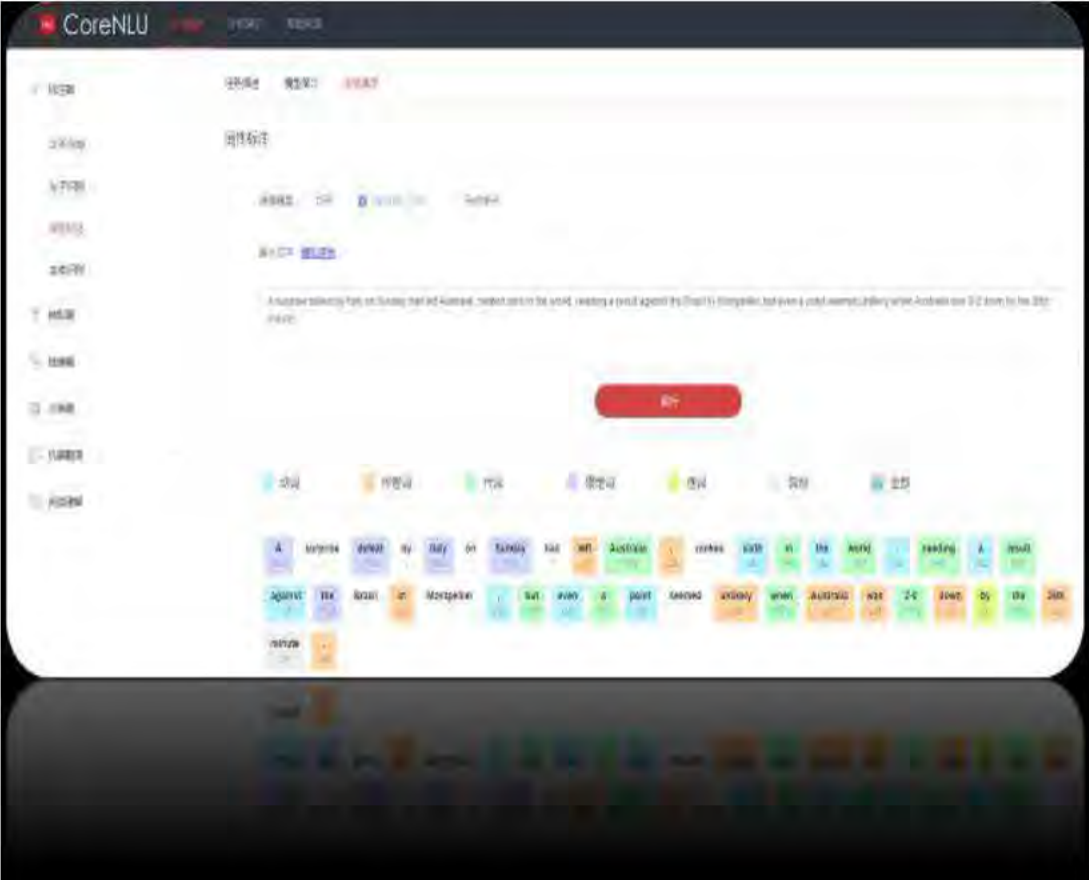
智能应用的基础：利用知识图谱构建知识体系

- 如何从半结构化和无结构化数据种抽取实体、属性等结构化信息？
- 如何消除概念的歧义，剔除错误概念，确保知识质量？
- 如何将各类知识表达成计算机可存储和计算的结构？
- 如何将知识加工成可以应用的知识体系？

知识获取	知识融合	知识存储与计算	知识应用
1. 利用天玑特有的采集能力，形成开放网络知识库 2. 多元异构自适应知识抽取 3. 自研隐含知识抽取方法，授权发明专利3项，研究成果发表在ACM AAAI2014等 4. 实体链接、属性推断技术国际评测排名第一（超过NELL,IBM, 微软），申请发明专利1项 相关技术： 实体识别、关系抽取 属性抽取、事件抽取等	1. 提出基于依赖图的联合推断算法（CIIGA）解决知识融合准确率问题 2. 采用快速领域知识库构建技术，较传统方法速度提高45%，形成功能组件和工具20余项，形成专利8项。 3. CoreNLU深度自然语言理解引擎，用于对海量无结构内容进行深层语义分析与理解 相关技术： 概念合并、关系合并、实体消歧、 实体融合、实体属性融合等	1. 支持十亿实体、百亿关系存储 2. 查询性能：毫秒级查询 3. 采用大图分析计算引擎，用于对大规模网络化数据实现快速、高效、可靠的关联分析 相关技术： 图切分、图查询、关联分析等	1. 积累行业通用知识规模达千万级实体、上亿级关系 2. 提供多维检索服务，提供对实体、文档、事件、快照等多元异构知识语义检索、精确定位 3. 知识应用算法20+种，涵盖知识问答、个性化推荐、人群分析等 相关技术： 知识推理、机器学习算法、 图分析算法、语义检索等

Core NLU

深度自然语言理解引擎，用于对海量无结构内容进行深层语义分析与理解



- 支持中英文两大语种
- 支持新闻、博客、社交媒体等不同语料
- 覆盖自然语言理解全体系
- 最前沿的深度学习模型
- 通用/专业领域兼顾

20+ 算法

全链路

亿级参数

大模型

秒级处理

高效率

- 与学术界先进算法相比性能占优或相当

模型/数据集	词性标注	实体识别	主题分类	实体链接	事件抽取	机器翻译
	CoNLL 2003	CoNLL 2003	AG_News	ACE 2004	ACE 2005(trigger)	AI Challenge 2018
CRF (2015)	94.42	81.86	-	-	-	-
BiLSTM-CRF (2015)	95.15	84.18	-	-	-	-
AGDISTIS (2014)	-	-	-	66.4	-	-
NCEL (2018)	-	-	-	88.1	-	-
Char-CNN	-	-	0.8702	-	-	-
LEAM (2018)	-	-	0.9175	-	-	-
DMCNN (2016)	-	-	-	-	73.5	-
JMEE (2018)	-	-	-	-	75.9	-
GNMT+spm (2016)	-	-	-	-	-	25.22
NLU	94.19	90.59	0.9206	87.3	78.4	24.56

- 在某J方项目POC测试中（参与方科大讯飞、拓尔思等）综合成绩第一，6项自然语言处理组件测试4项第一。

- 多年技术积累，在历次国际权威评测中名列前茅

关键技术	国际权威评测成绩
用户画像	SMP CUP2017评测排名第一
知识融合	TAC2016国际评测排名第一
知识发现	TAC2016国际评测单项排名第一
中文繁体纠错	SIGHAN2015国际评测排名第一
知识融合	TAC2014国际评测排名第一
知识发现	TAC2014国际评测单项排名第一
自动文摘	TREC2013国际评测排名并列第一
棒性检索技术	TREC2013国际评测排名并列第一
关联实体检索技术	TREC2009国际评测排名并列第一
动态文摘技术	TAC2008国际评测排名第一
企业专家检索技术	TREC2007国际评测排名第一
新信息发现技术	TREC2004国际评测排名第一
中文词法分析技术	SIGHAN2003国际评测排名第一



2.3 智能支撑引擎——开放知识系统 (OKS)

OKS

开放知识计算引擎，用于面向应用领域实现动态知识结构的快速构建、管理和推理计算

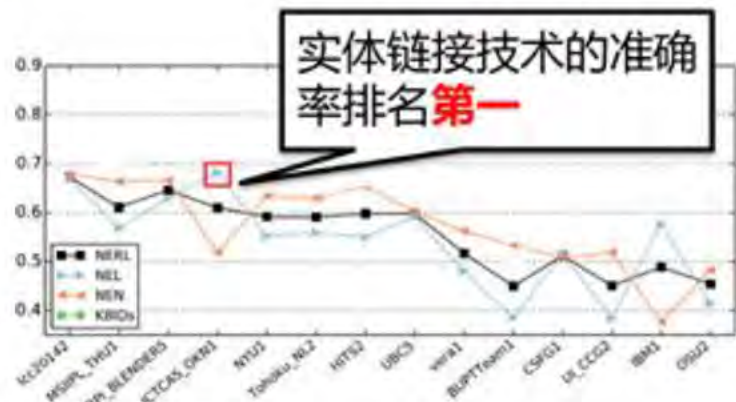


- 支持政治、JS、金融等多个垂直领域
- 支持结构化、非结构化数据
- 丰富的知识抽取算子库，灵活可定制
- 友好的人机协同知识标注和管理工具



2.3 开放知识系统技术优势

实体抽取、实体链接、关系抽取等多项关键技术**国际排名第一**



Monolingual ENG language condition:

Rank	strong_mention_match			strong_typed_mention_match		
	P	R	F1	P	R	F1
1	0.916	0.680	0.781	0.870	0.647	0.742
2	0.877	0.686	0.770	0.835	0.653	0.733
3	0.812	0.638	0.715	0.772	0.606	0.679
4	0.854	0.558	0.675	0.766	0.500	0.605
5	0.733	0.538	0.620	0.664	0.488	0.563
6	0.874	0.409	0.557	0.818	0.382	0.521
7	0.765	0.371	0.500	0.722	0.350	0.471

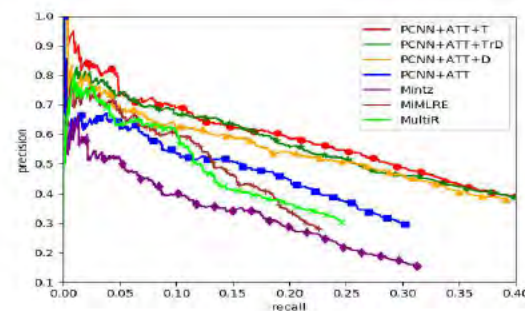
实体发现总**任务排名第一**，
6个指标中4个排名第一



typed_mention_caf			b_cubed		
P	R	F1	P	R	F1
0.764	0.598	0.671	0.779	0.515	0.620
0.749	0.531	0.621	0.852	0.452	0.591
0.656	0.557	0.603	0.636	0.465	0.537
0.666	0.435	0.526	0.803	0.347	0.484
0.600	0.441	0.508	0.632	0.344	0.445
0.752	0.352	0.479	0.845	0.233	0.366
0.553	0.268	0.361	0.697	0.169	0.272

实体聚类准确率**排名第一**

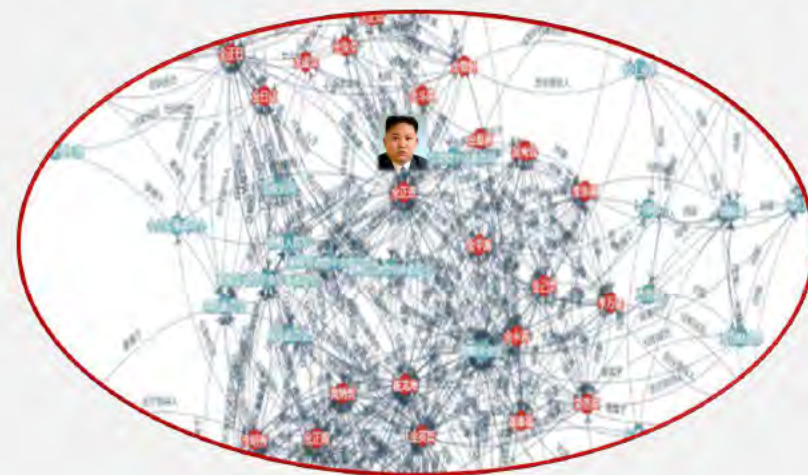
Method	WN18			
	MRR	Hits@1	Hits@3	Hits@10
RE\$CAL	0.890	0.842	0.904	0.928
TransE	0.495	0.113	0.888	0.943
TransR	0.605	0.335	0.876	0.940
ER-MLP	0.712	0.626	0.775	0.863
DistMult	0.822	0.728	0.914	0.936
HolE	0.938	0.930	0.945	0.949
ComplEx	0.941	0.936	0.945	0.947
R-GCN	0.814	0.686	0.928	0.955
ANALOGY	0.942	0.939	0.944	0.947
SENN	0.947	0.941	0.950	0.956
SENN+	0.947	0.942	0.950	0.956



2.3 开放知识系统成果积累

- 围绕公共安全、敏感人物、**东海、南海区域**等**10个领域**，从实体、概念、事件等不同粒度出发，构建了**超过5000万实体的大规模知识图谱**。
- **知识获取**：实体抽取准确率90%以上，召回率87%以上，性能高于64.9K/s。
- **知识融合**：基于表示学习的实体对齐正确率90%+，融合单个实体时间小于1s。
- **知识计算**：基于PRA的关系推断正确率70%+。基于表示学习的标签推理正确率70%。
- **知识服务与应用**：毫秒级查询

基于知识图谱的金正恩内阁决策体系分析



从**人物**、**机构**与**事件**三个视角分析金正恩内阁的决策体系，在情报业务应用中发挥重要作用

为面向重大战略问题的知识体系构建奠定了坚实的基础
对战略情报分析等专用数据知识体系形成技术能力支撑

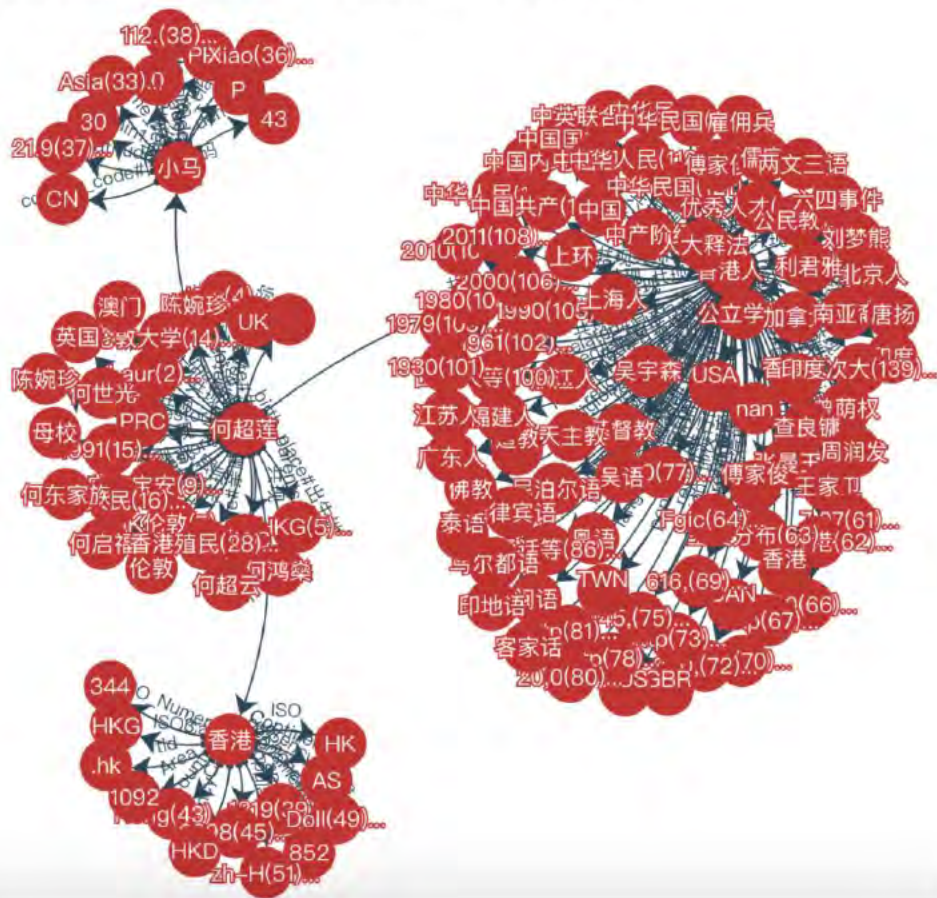


知识浏览

实体/概念查询

关系/路径查询

实体/概念数: 13,683,465 关系/属性数: 149,910,460



通用知识图谱数据规模:

实体/概念数: 13,683,465

关系数: 149,910,460

JS领域知识图谱数据规模

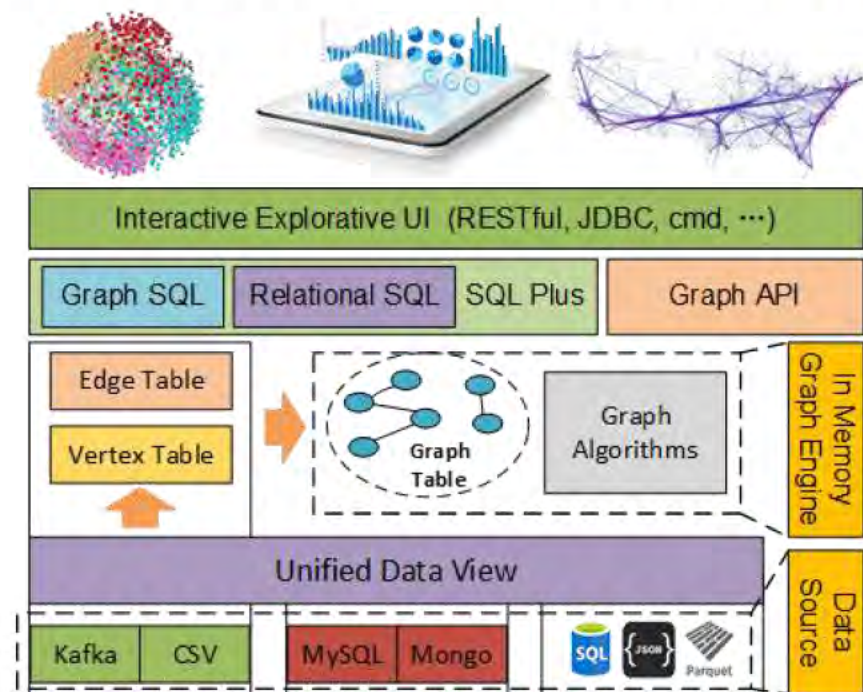
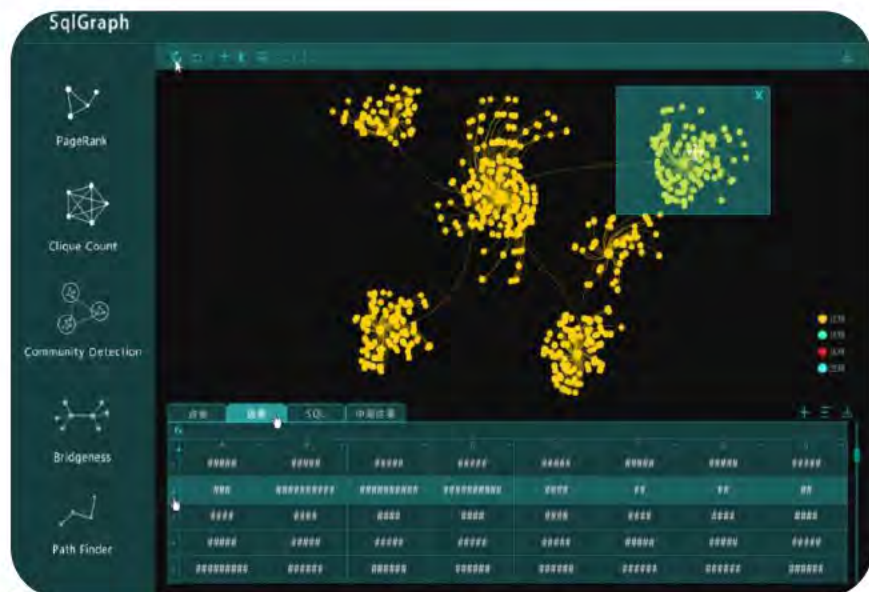
实体/概念数: 2,328,330

关系/属性数: 30,146,515



2.3 智能支撑引擎——图计算引擎 (SQLGraph)

SQL Graph 大图分析计算引擎，用于对大规模网络化数据实现快速、高效、可靠的关联分析



- 支持无向图、异构图、属性图
- 高效的动态图切分算法和图压缩表示
- 可扩展的图查询语言支持交互式图查询
- 内置丰富的图查询、图挖掘算子库



01

N度关联关系、直接、间接关联关系查询

02

共同好友、共同联系人分析查询

03

关联亲密度、中心成员分析查询

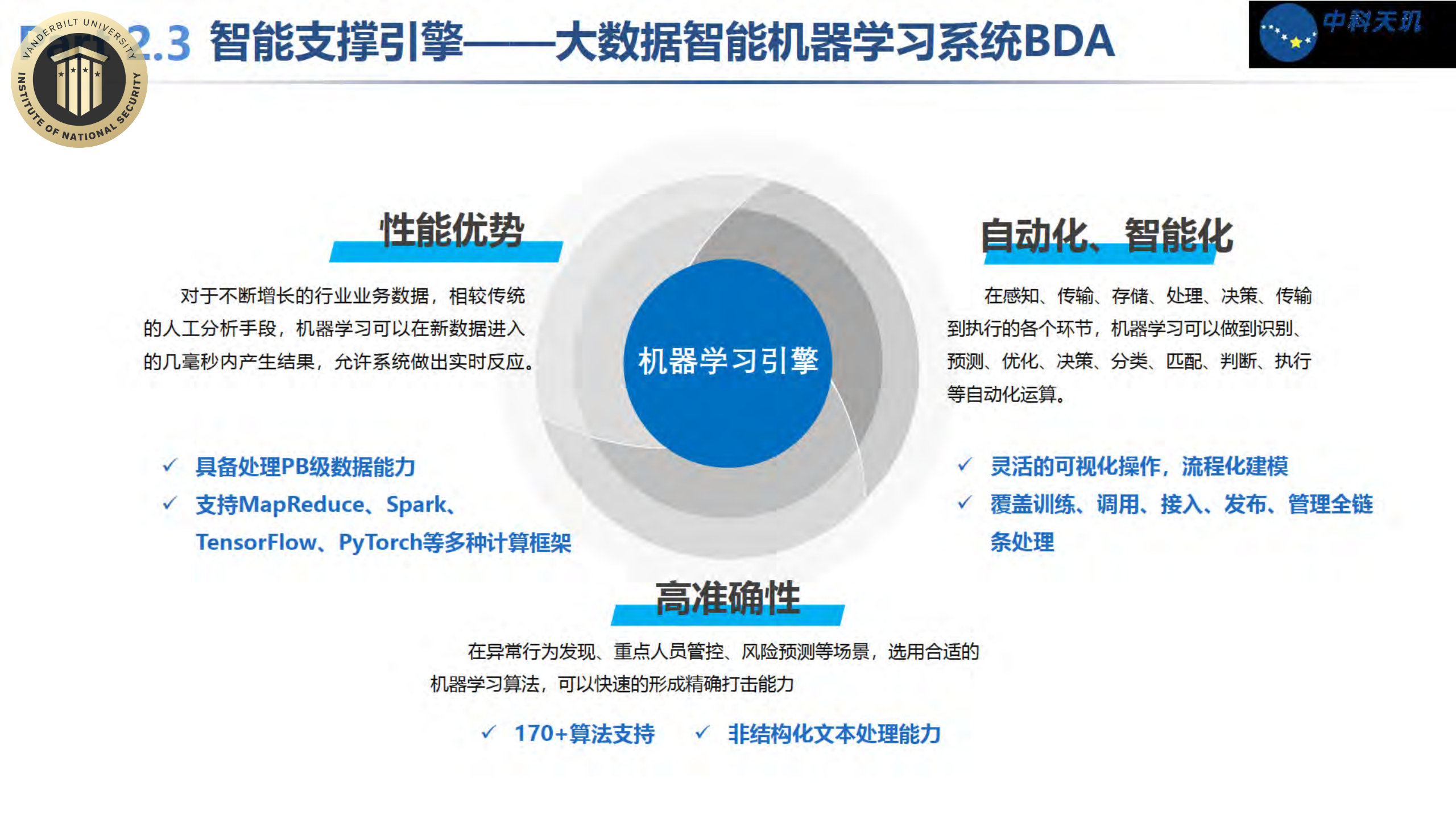


04

团伙、子团伙分析挖掘

05

资金链、用户兴趣爱好分析挖掘





2.3 智能支撑引擎——大数据智能机器学习系统BDA

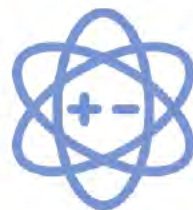


图形
交互



简单的交互界面

分析
算力



高维度、高精度模型

协作
共享



共享数据、模型、经验

解决AI技术在产业化落地过程中遇到的软硬件搭建成本高、现有分析工具链路缺失、缺乏协作与共享能力、AI应用构建门槛高、缺乏便捷用户接口、部署困难等痛点，为企业级大数据人工智能应用**提供低门槛、一站式、高性价比的应用开发环境**

A graphic of a constellation with several blue stars connected by thin lines, positioned above the text "Part 2.4".

Part 2.4

数据分析服务

面向各类用户提供专业、精准的情报和舆情分析服务，具备专业的分析师团队，
具备丰富的分析报告经验。



2.4 专业的数据分析服务



专业的**分析工具+分析团队**

涵盖**舆情分析、人物画像、组织调研、专题研究**4个领域

2.2

2.2.1 人物档案



及备注栏。右
多为在香港时
玩时拍摄。
76人，发生
点赞又评论的
Br

2.2.2 关联关系



上图为
系图。下方为时
表示该时间点活
其相应的实体、
表发帖及评论帖
组织。关系统计
评论、点赞、发
其未公开fb好友
友列表中。

江苏灌云“女辅警案” 舆情分析报告

(2021年3月11日-19日)

CCLaxy 中科天玑 舆情产品部

2021年3月

2.4 数据分析优势

高效



依托自主研发的大数据QB分析系统，可快速获取开源信息，大幅提高QB分析效率。

准确



多渠道、多源头核实所有数据，专家从分析过程与成果方面把关，确保分析过程严谨，分析成果准确。

专业



由20年J队情报工作经验的专家带队，辅以30+的外围专业人员，形成独家QB解决方案及实战队伍。

详实



数据准确、内容详实、指向明确，专项内容达1000+页，为分析任务提供了有力支撑。



Part 2.5

智能情报应用

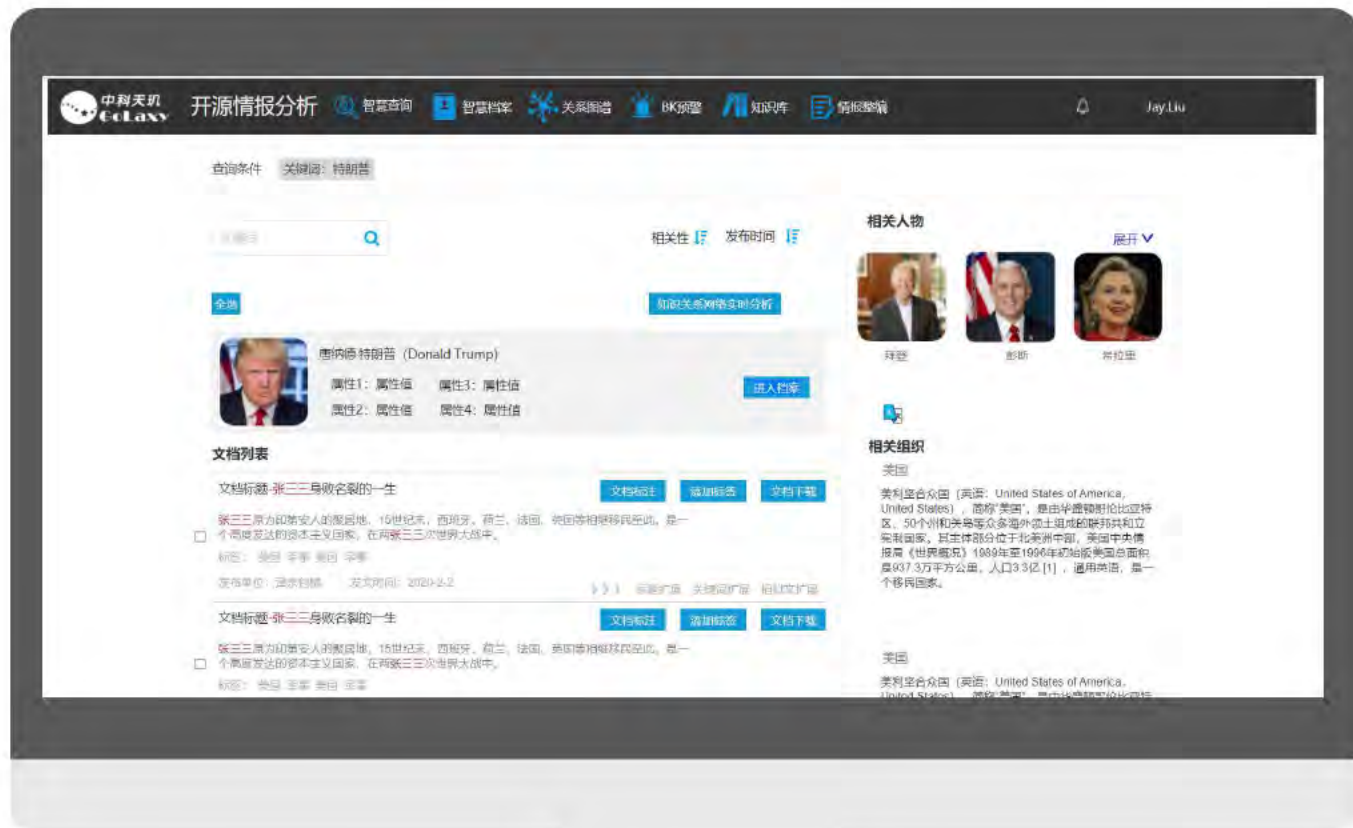
包括三大类分析应用：开源情报分析平台，GoIN战略情报分析平台和资金分析系统

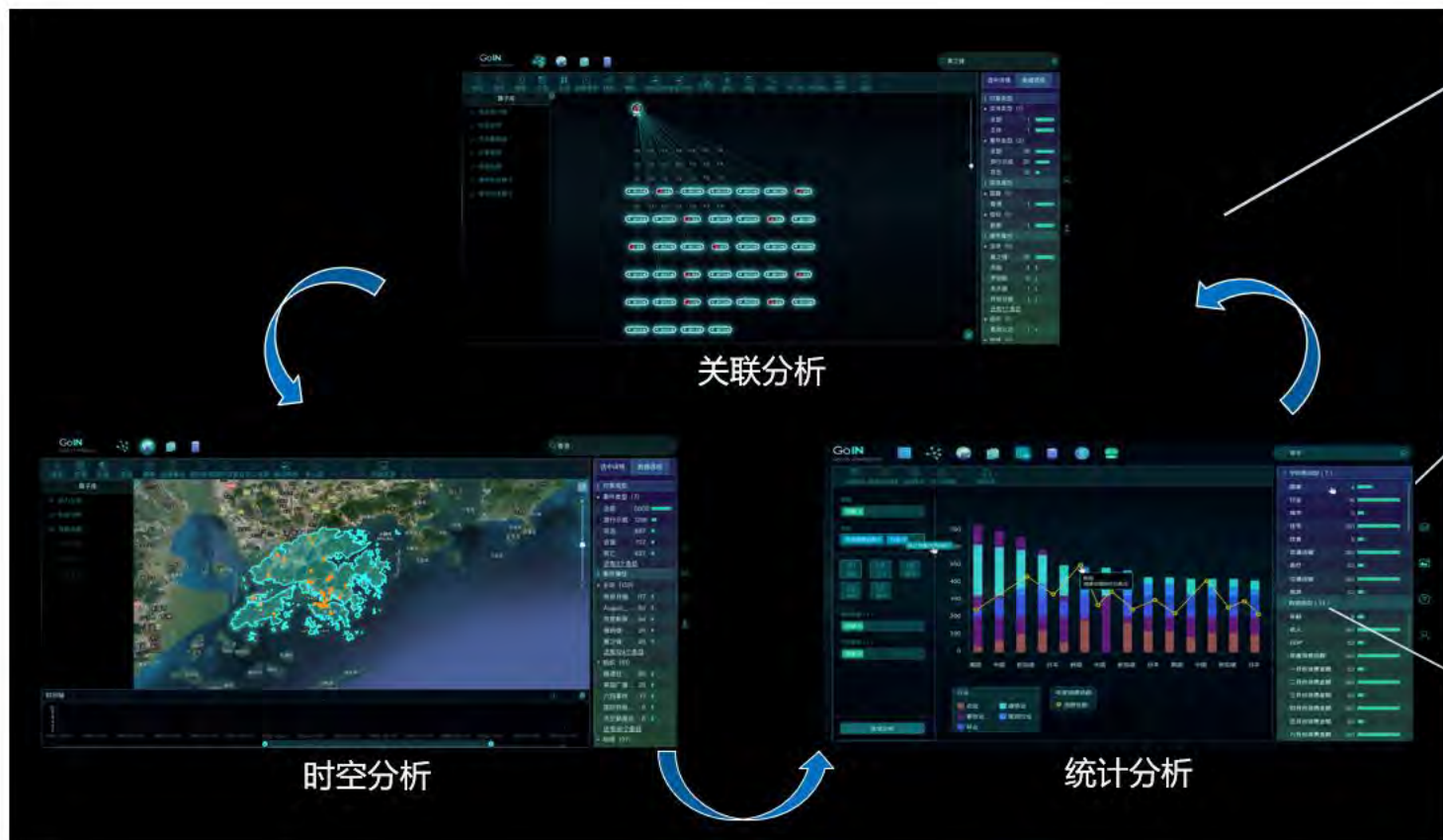


情报分析类——开源情报分析平台



- 致力于开源数据的全网信息情报智能发现和挖掘平台。平台聚合境内外资讯媒体、智库论坛、社交媒体、即时通讯等多渠道的海量数据，实现信息监测、智能推荐、社交分析、涉我发现、事件感知、情报智库、报告整编等能力，提供基于跨网络的目标监测、分析和拓线，满足AQ、GA涉外业务的全面掌握，实现全网情报的获取、感知与洞察。





以大数据深度分析应用（情报应用）为主要抓手

通过对多元、异质大数据的关联融合，利用混合人机智能的分析过程

辅助用户从大数据中萃取关键情报，并进一步转化为可操作的决策指挥能力

产品优势

- 战略到任务的分解，构建标准化协同分析流程
- 丰富的分析模式与全面的数据接入，满足不同客户不同条件下的全场景分析支撑
- 人机混合智能的开放式可视化交互设计，支持专业人士深度探索
- 贴近业务的智能算子库构建，支撑面向行业的个性化定制
- 云端商店模式构建知识图谱应用生态

5大特色功能

- ① 战略任务分解
- ② 异质数据接入
- ③ 人机协同分析
- ④ 情报综合生成
- ⑤ 行业定制专题



情报应用类——资金分析系统



基于大数据及算法模型可对资金流向及交易时间进行智能化分析，形成一套有效的异常资金转移活动分析系统，更好的打击网络赌博、非法洗钱等犯罪行为。

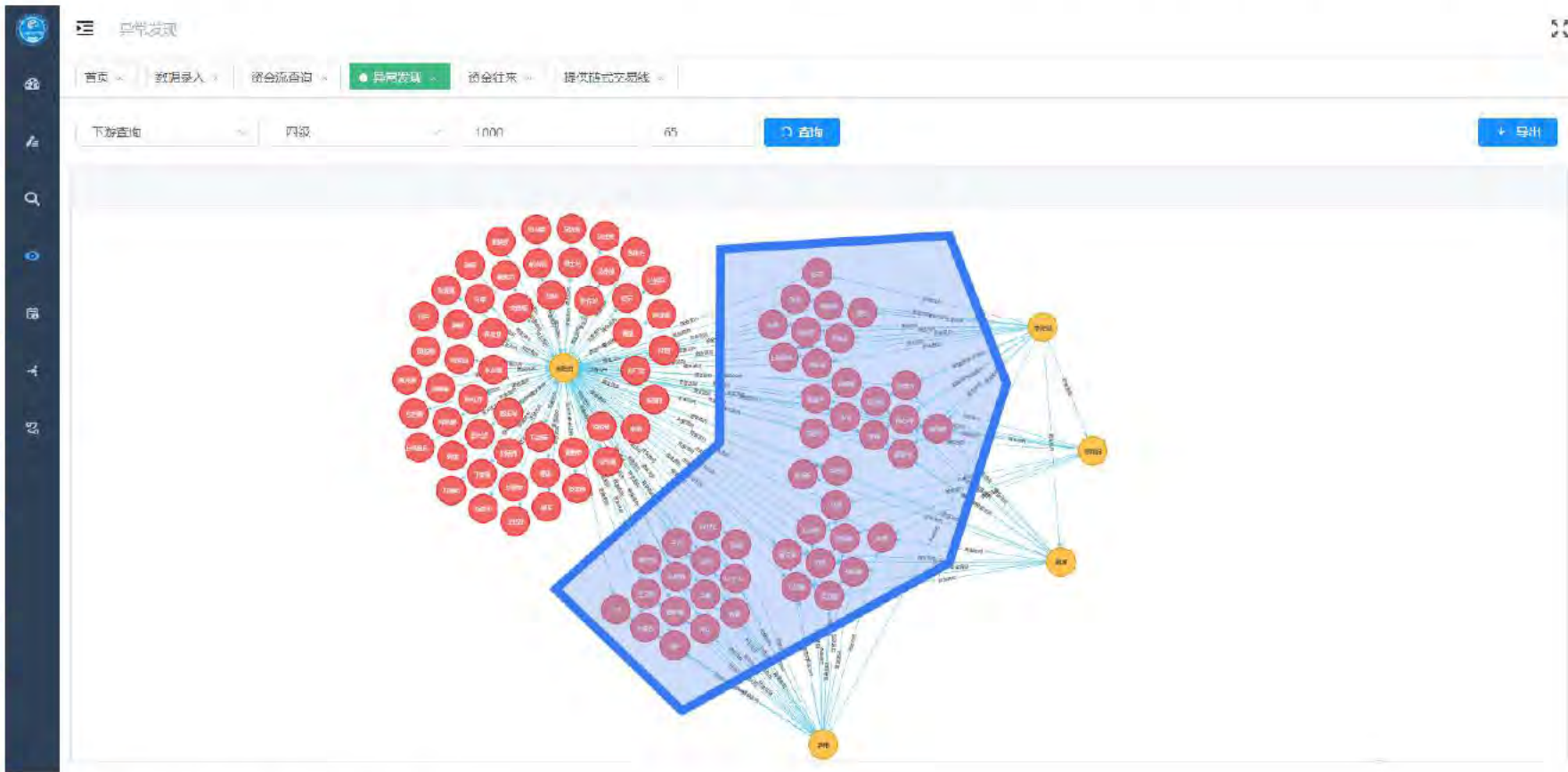
异常账户群发现

异常账户交易时间预测

链式交易线生成

嵌入式图数据查询

异常团伙等级评定



应用成效：本系统在2个月内为某县局提供如下信息：

✓ 线索卡号6000+，疑似过桥账户30+

✓ 提供冻结卡约100+，准确率高于90%

✓ 冻结金额约2000W

A graphic of a constellation with seven blue stars connected by thin black lines, positioned above the text "Part 3".

Part 3

行业解决方案及案例

•



3.1 政府行业解决方案布局

政府

宣传

JG

金融

教育

行业政策解读

销售工具集

行业市场分析

行业案例集

面向客户

AQ

公安

纪检

政法委

政府其他

应用领域

国家安全

公共安全

廉情热点

政法輿情

司法輿情

行业解决方案

互联网行为分析解决方案

图数据库分析一体机解决方案

智能数据分析解决方案

开源情报分析解决方案

海关缉私情报分析解决方案

反赌/反诈智能分析解决方案

政法輿情解決方案

纪检廉情大数据解决方案

司法與情解決方案

行业核心产品

多源异构智能数据处理产品

开源QB分析产品



3.1 政府行业方案规划与拓展方向



政府 > 宣传 > JG > 金融 > 教育

面向社会治理部门: 党政机关

- ✓ 网络安全管理
- ✓ 突发应急管理
- ✓ 社会民生管理
- ✓ 政府形象管理
- ✓ 监测发现/研判预警/追踪处置

舆情分析

面向情报需求部门: AQ、GA

- ✓ 情报要素抽取整合
- ✓ 多源信息跨源检索
- ✓ 人机交互挖掘工具
- ✓ 专项情报分析团队
- ✓ 分析结果导出成报

开源情报

面向业务分析场景: 案件、人员、资金

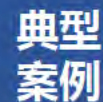
- ✓ 邮件文本快速导入
- ✓ 知识图谱关系分析
- ✓ 自然语言处理分析
- ✓ 智能写作/推荐/编报
- ✓ 行业知识库持续积累

智能后处理





政府 宣传 JG 金融 教育



宁波市委宣传部四力项目

为宁波市委宣传部搭建“县融客户端四力数字化平台”项目，基于数据的量化指标引入“四力”管理体系



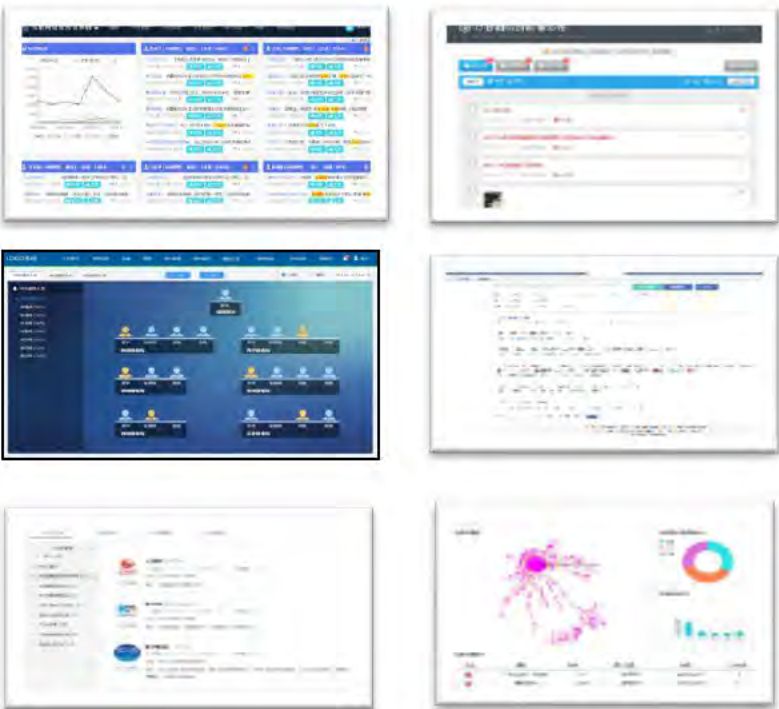
党建机器人项目

机器人“小天”是应人民网“智慧党建”要求，专门为建党一百周年定制研发的一款从事党务和党史服务的机器人



网信项目

为中央网信办7个局9个业务系统、6个地方网信办提供属地网站管理、指挥通讯系统、舆情监测系统等服务





3.2 宣传行业方案规划与拓展方向

政府 宣传 JG 金融 教育

网信类解决方案

指挥通讯系统
属地管控系统
新闻调控系统
音视频监测系统
... ..

融媒新产品解决方案

党建机器人
四力产品

规划中产品解决方案

国际传播力产品
法治大数据产品
评论管理产品矩阵
审校管理产品矩阵

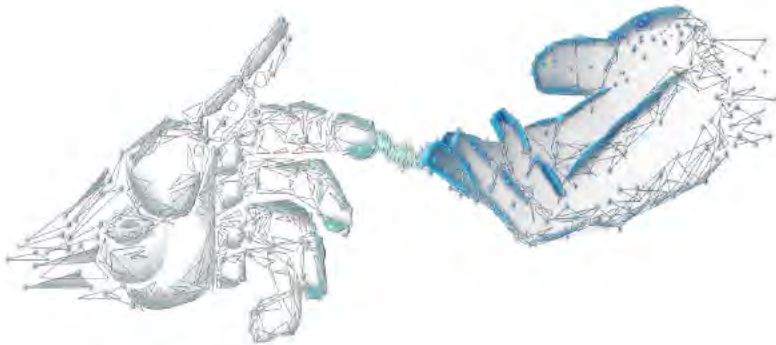


打造全方位宣传BG立体化解决方案



中美战略态势分析系统

为国防科技大学建立中美战略态势分析系统。采集并分析挖掘出可能发生冲突的战略关键点，进行数据融合展示战略态势，为首长决策提供辅助支撑。



某地区细粒度选情分析

面向联参某局，根据地区细粒度进行选情态势分析，基于开源数据和内部数据，利用选情模型，构建人物画像等方式实现对人物选情的监测和预测。

专业领域知识图谱构建

面向南部ZQ，针对值班综合报告、执勤综合等情况，构建领域知识图谱，为执勤、演习演训、各级指挥决策研判提供有效综合性专题和知识保障。



素材整编系统建设项目

面向战支某局，构建素材整编系统，实现多种类业务素材接入及对业务素材进行自动化或半自动化的内容识别、智能关联、辅助编译以及态势数据成品整编输出。

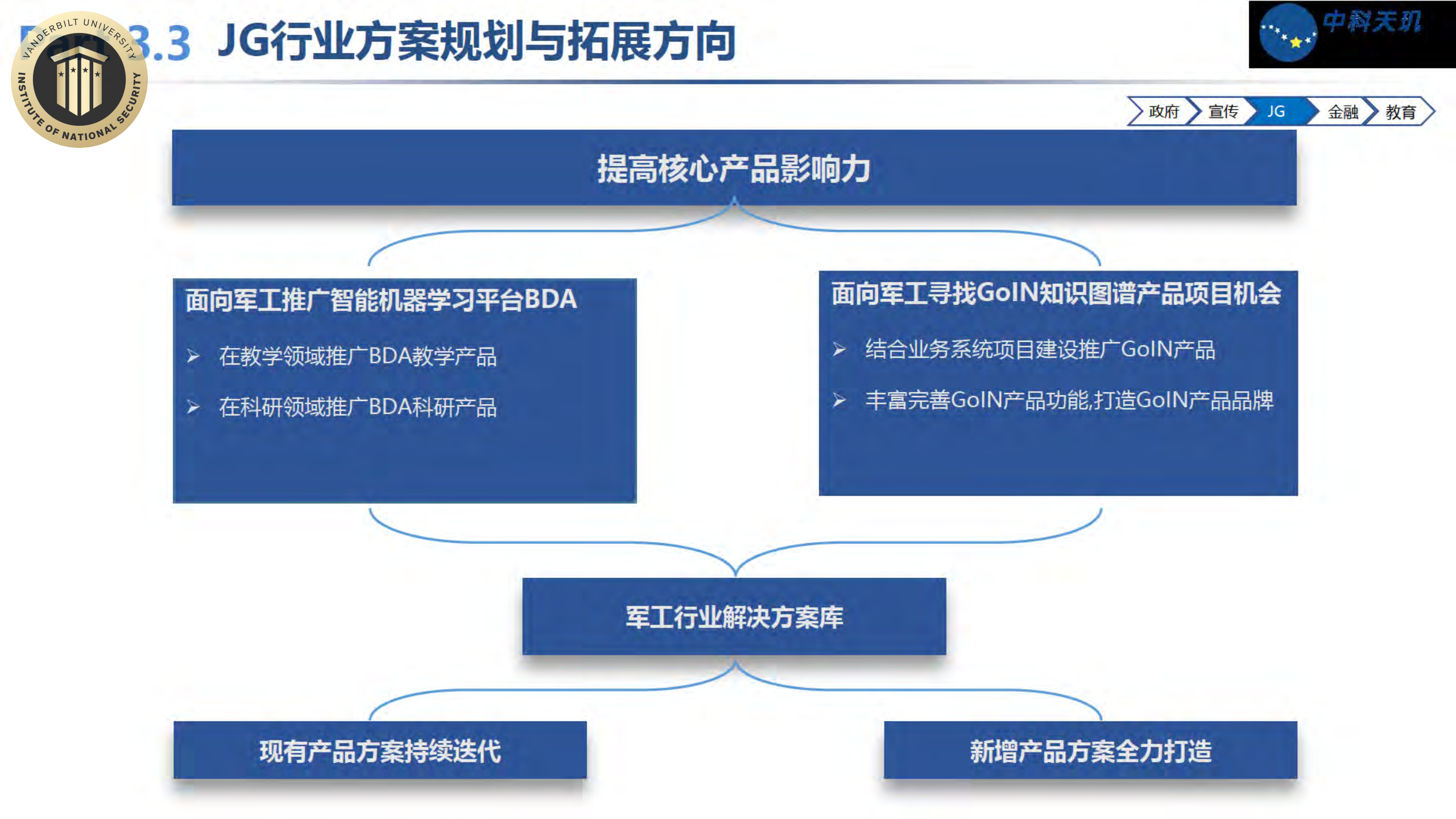
XX实体抽取与检索项目

面向联参某局，针对业务领域的业务实体与检索系统，提供语义化检索、场景化检索，解决业务数据上下难联、左右难融合、效益未挖掘等短板问题。



大数据平台建设方案

面向航天工程大学建设模拟训练大数据平台，开展航天领域业务模型建立及大数据分析工作，为军事安全战略环境评估提供辅助决策分析功能。





政府 宣传 JG 金融 教育

应用 场景	金融监管		金融客户舆情监测		智慧经营分析	
	非法集资监测 地方交易所监测 区域金融监测 监管协同联动		资本市场舆情监测 金融客户风险评估 企业风险综合预警 客户风险画像		精准营销 客户行为预测 车险预测 客户流失预警	
智慧 引擎	通用 解决 方案	企业知识图谱方案	企业舆情监测方案	区域金融风险监测	非法集资监测方案	监管协同联动方案
		地方交易所监测	地方交易所监测	客户智能画像方案	企业风险综合预警	精准营销方案
智慧 引擎	通用 技术 模块	感知 非结构化数据	分析 数据挖掘	决策 预测预警	监控 监测跟踪	深度学习 人工智能
		自然语言处理 语音识别 视频识别 知识图谱	统计分析 特征工程 聚类分析 可视化探索	风险评估 实时预警 偏好分析 流失预警	关联网络监控 热点发现 情感分析 趋势分析	深度学习 强化学习 对抗网络 卷积网络
知识 库	企业信息库 人员信息库 关联关系库 属性标签库 金融知识库 互联网知识库					
数据采 集 治理 整合 存储 管理	数据采集 + 数据治理 + 数据整合 + 数据存储					

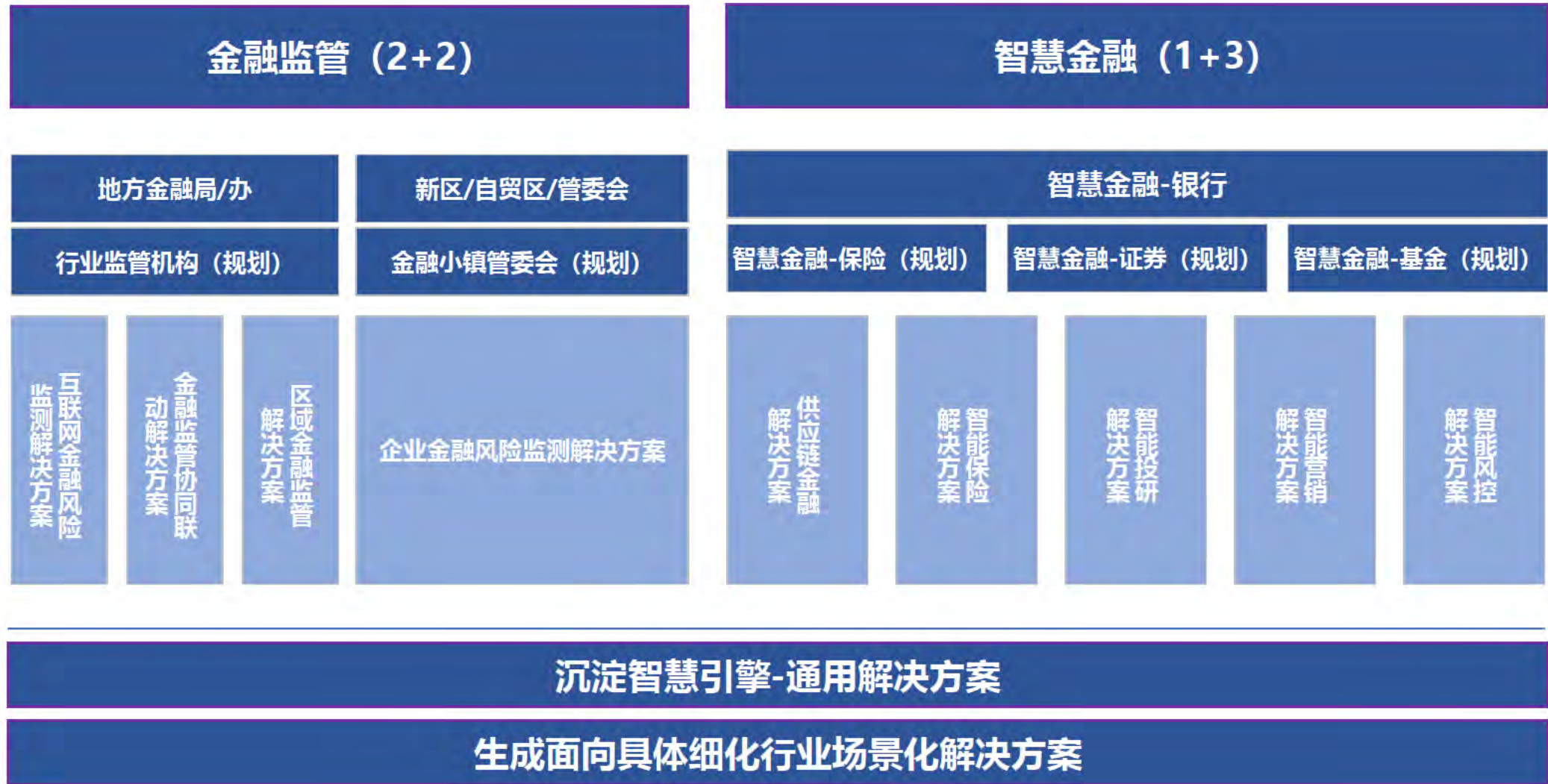


3.4 金融行业方案规划与拓展方向



政府 > 宣传 > JG > 金融 > 教育

天玑智能金融解决方案-未来规划为金融监管（2+2）和智慧金融（1+3），进一步拓展行业解决方案领域和行业服务能力，同时沉淀智慧引擎通用解决方案，完善解决方案体系。





3.5 教育行业方案布局

政府

宣传

JG

金融

教育

典型案例

工业数据实验室

智能驾驶数据分析与挖掘实验室

旅游大数据专业共建

大数据技术研究与应
研创新基地

AI金融实验室

医疗大数据实验室

某J校数据分析实验室

大数据视阈下全面建成小康社会分析课题

方案场景

院校

本科院校

高职院校

军警院校

专业

新工科

新医科

新农科

新文科

新商科

解决方案

大数据专业建设方案

人工智能专业建设方案

使用场景

算法模型构建

知识图谱构建

人工智能实验

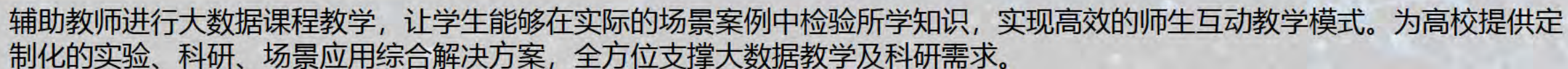
慕课教学

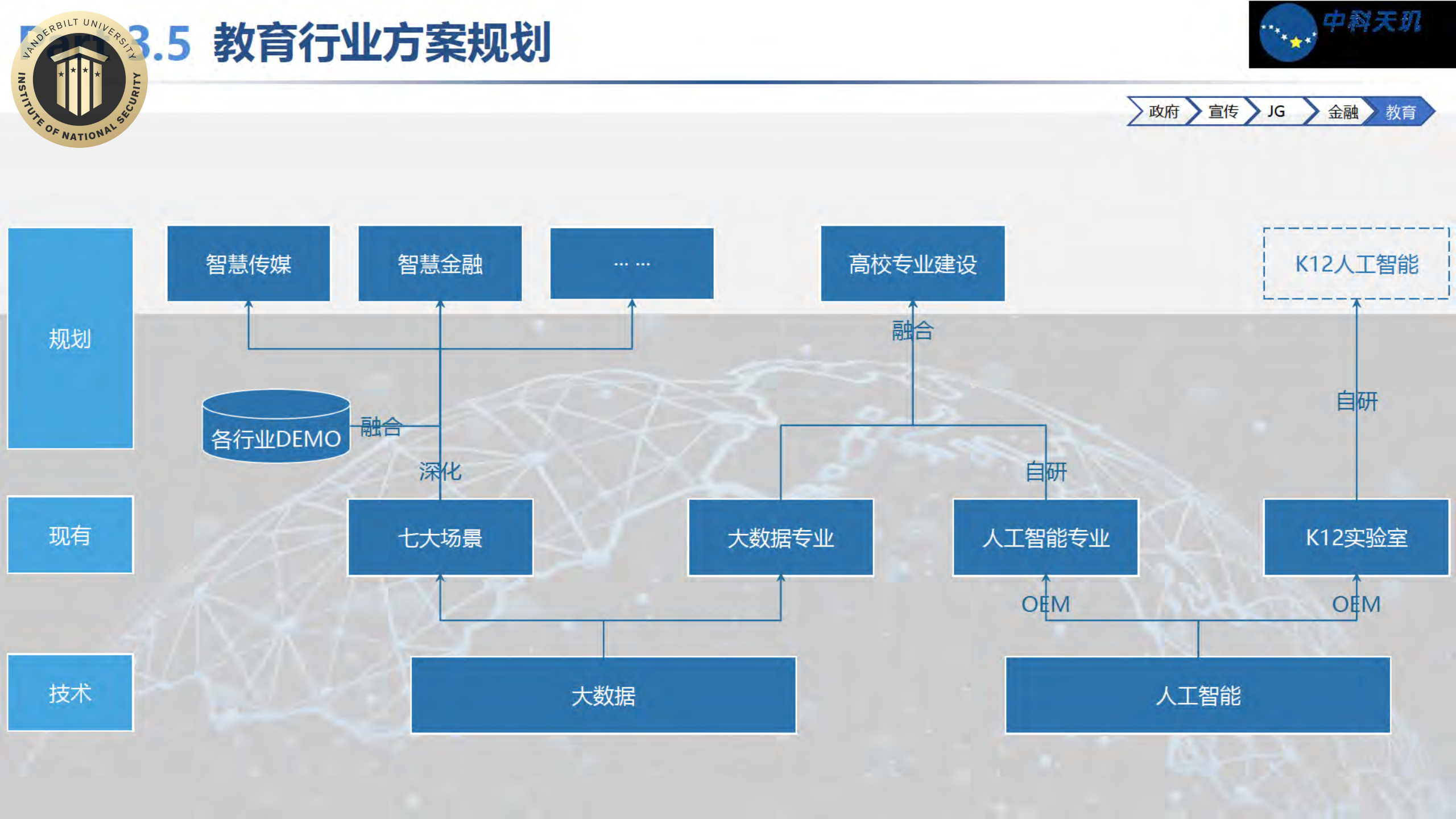
在线编程

在线考试



政府 宣传 JG 金融 教育







3.5 教育行业案例应用

政府

宣传

JG

金融

教育

01

华北水利水电大学
大数据技术研究与创新科研基地

建立支撑全校师生开展大数据技术研究与应
用科研创新的权限管理中心、计算支持中心。
支撑学校师生在开展采集、数据分析、机器
学习、深度学习及人工智能应用 的计算能
力支撑。确保能支持高并发、稳定、计算能
力强的服务集群，实现高校师生 在平台之
上开展不同行业领域的大数据科研创新实验
课题需求。

02

某校 数据分析系统

基于知识图谱智能构建、基于图数据库的数据分析系统，对多元异质大数据的深入挖掘、关联融合和动态开放的知识网络构建，以从大数据中萃取出关键的知识数据，充分发挥出大数据的价值。同时，支持对大数据资源的基于角色和资源属性的访问控制。实现科研成果推向业务，业务成果回归教学的一套系统多种应用的融合型平台。

03

广东轻工职业技术学院
工业大数据实验室

基于教学科研实验平台、《工业大数据分析挖掘》课程及相关案例，打造集理论教学、应用实验、项目实训、成果转化和科研创新为一体的工业大数据产教融合示范实验室，支撑院校双高专业群建设。师生可以在实验室内完成教学活动；结合校内的工业生产线，开展科研实践活动；利用实验室成果扶持下游企业。

04

上海交大医学院
医疗大数据实验室

通过建设大数据教学资源、大数据分析挖掘平台，与医学院既有的医疗数据进行对接，针对性的建模分析小儿心肺音，用于辅助诊断新生儿心脏病。项目旨在通过大数据课程学习，师生可以掌握利用大数据技术融入日常诊疗中，利用高科技手段提升医疗水平。

05

浙江金融职业技术学院
AI金融实训实验室

AI金融实训软件项目，为院校搭建大数据分析挖掘实验室，通过大数据、人工智能等技术，挖掘金融数据背后的价值，服务于金融风控、预警、监管等方面，助力其提升学院人工智能教学水平，打造人工智能金融体系。

06

北京联合大学
智能驾驶数据分析与挖掘实验室

为车辆工程专业提供了完备的数据分析与数据挖掘平台。探索深度学习算法在智能驾驶领域的应用。整个系统对智能驾驶过程产生的数据能够进行智能化的收集、流转、分析与挖掘，并对所得结果进行可视化展示。



深挖价值 服务国家 惠及大众

使命 价值 荣誉 担当





中国科学院网络数据科学与技术重点实验室
Key Laboratory of Network Data Science & Technology, CAS

I 总结与展望



中共中央网络安全和信息化委员会办公室
Office of the Central Cyberspace Affairs Commission



中华人民共和国国防部
MINISTRY OF NATIONAL DEFENSE OF THE PEOPLE'S REPUBLIC OF CHINA

WWW.MOD.GOV.CN



中共中央宣传部

自主研发的智能宣传系统作为新兴事物，初步具备政治态势感知、群体心理画像、内容智能生成和信息智能投送能力，但离全面服务国家在境外网络舆论斗争的需求还有一定差距。

在科学院领导和所领导的大力支持下，与中央网信办、中宣部、军科委、安全部等部门加强合作，明确任务需求，加大资源投入，打造出能真正自动讲好中国故事、发出中国声音、扩大中国影响力的技术平台，为早日实现“东风压倒西风”提供全方位的技术支持。



中国科学院网络数据科学与技术重点实验室
Key Laboratory of Network Data Science & Technology, CAS

GoPro宣传手册



中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES



中国科学院网络数据科学与技术重点实验室
Key Laboratory of Network Data Science & Technology, CAS



台湾地区网络舆情周报

(2023 年 11 月 25 日 0 时-12 月 1 日 12 时)

本周台湾地区正式进入“选举时间”，“蓝白合”热点话题消失，网络舆论热度小幅滑落。柯文哲借“2 亿美金捐客说”博得舆论关注，但弱势地位很难改变。“侯康配”汇聚蓝军士气，支持率直追“赖萧配”，为民进党带来较大压力，一定程度上使“赖萧配”在舆论场被边缘化。本周民进党当局对“重启服贸”的“膝盖式台独”反应、蔡英文受访称中国不会武力攻台、吸纳“太阳花学运”重要人物林昶佐等行为值得关注，需防范其重启岛内反服贸回忆、制造两岸对立，进一步蚕食两岸对话交流的基础。

一、总体舆情概况

(一) 舆情传播情况及倾向性分析

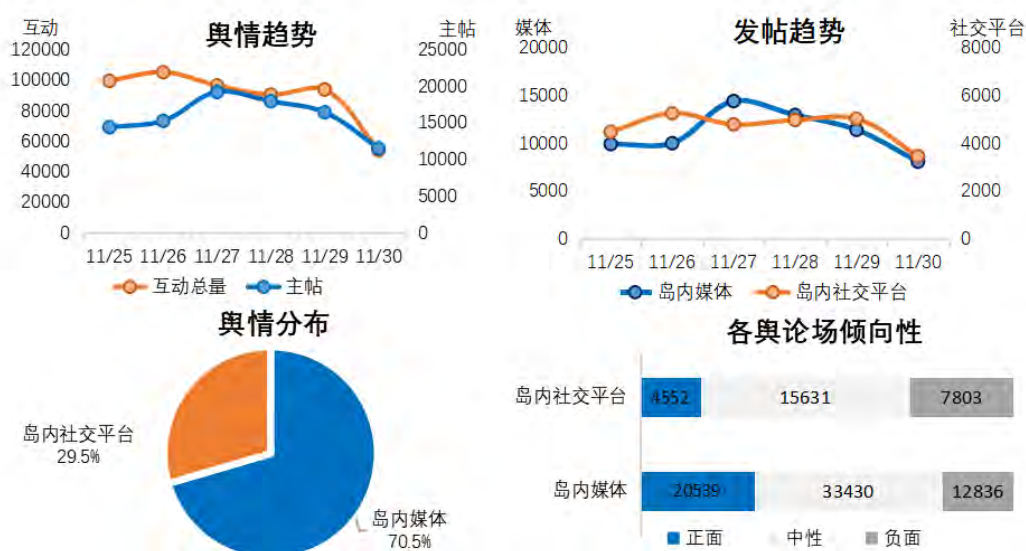


图 1 舆情传播情况及倾向性



注：主帖指岛内媒体和社交平台的内容类信息。互动指主帖引发的转评赞等交互活动。

本周台湾地区舆情总量共计 634448 篇次，其中发帖量 94791 篇条（岛内媒体 66805 篇，社交平台 27986 条），互动信息 53.9 万（含转评赞及阅读等互动情况），舆论场热度较上周小幅回落（上周舆情总量约 66.8 万篇次）。27 日侯友宜参加九大工商团体组织的“2023 台湾经济发展论坛-与国家未来领导人对谈”会议，提出重启服贸谈判、开放陆生在台就业引起舆论关注，岛内媒体发帖量冲至周内高峰。本周“侯康配”、柯文哲“2 亿美金捐客说”“叶元之看板改造”等话题持续引发讨论，部分程度使民进党的政见限制在绿营范围，需靠蔡英文受访、辅选开拓局面。此外，福建省召开《两岸融合发展示范区的意见》第一场新闻发布会、卢秀燕任蔡壁如竞总主委开辟“立委”选战蓝白整合局面等，也在岛内掀起关注。

（二）阵营舆情分析



图 2 三大阵营舆情走势 单位：篇/条

台湾地区“选举”进入关键倒数，三大阵营纷纷发力吸引



关注，舆情发帖量互相胶着。柯文哲提“2 亿美金捐客说”并不断在公开场合回应，吴欣盈暧昧回应双重国籍争议，引发舆论猜想，使民众党始终处于舆论关注中心；赵少康强大的新闻捕捉能力和媒体圈人脉，在推动“侯康配”民调走高的同时，也利用“叶元之看板改造”、密集受访等灵活手段发挥“眼球效应”吸引关注；绿营则凭借其对岛内媒体的掌控实力，宣传“林昶佐加入民进党”“蔡英文接受《纽约时报》访问”及“赖萧配”政见，凝聚基本盘关注焦点。随着选举日渐白热化，三方缠斗、对抗之势将继续加剧。

二、本周舆情热点

序号	热点舆情事件	舆情发帖量
1	“侯康配”发力	15976
2	蓝白重启服贸谈判争议	4580
3	柯文哲2亿美金捐客说	3258
4	民众党退党潮	2970
5	叶元之看板改造	1656
6	卢秀燕任蔡壁如竞总主委	1652
7	林昶佐加入民进党	1405
8	吴欣盈双重国籍争议	1290
9	蔡英文纽约时报专访	495
10	两岸融合发展示范区意见系列新闻发布会	376

图 3 本周热点舆情事件 单位：篇条

（一）“侯赵韩”发挥“合唱团效应”，深蓝全数归队，绿营猛打赵少康。

“蓝白合”破局后，侯友宜、赵少康、韩国瑜联手，展示超强话题制造与选战能力，“合唱团效应”凸显，声量显著抬升，



连带岛内选情发生变化，多家民调显示“侯康配”支持率迅速接近“赖萧配”，岛内媒体推测深蓝已经全数归队，集中力量拼选战。

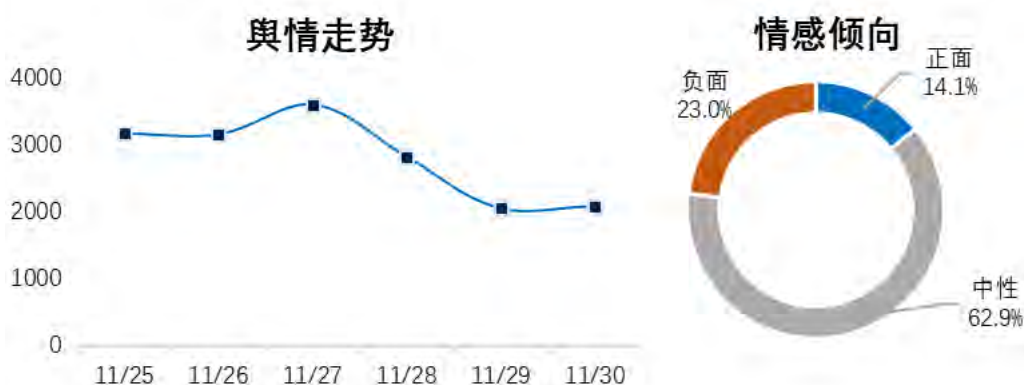


图 4 侯康配舆情发帖量 单位: 篇/条

第一，岛内舆论多肯定侯友宜选择赵少康是一招妙棋。赵少康深耕媒体多年，话题制造和捕捉能力极强；熟悉台湾政策及各种法律、法规，是追打民进党弊案的最佳人选；口才了得可以弥补侯友宜在辩才和个人魅力上的不足。“侯康配”成军后，赵少康迅速展示其话题制造能力：如建议侯登记前给柯文哲打电话、号召郭台铭及“经济蓝”选民回归、炮轰柯文哲两届台北市长毫无政绩、“监工”叶元之看板改造、批黄珊珊靠柯吃香喝辣、呼应柯提内阁制拉民众党入阁、计划进校园与青年直球对决，不仅提振蓝营选战声量，也提政见力图弥补国民党在青年支持度的不足。

第二，“侯康配”整体提升蓝营气势，选情呈蓝绿对决势头。多家民调显示，“侯赵配”支持度紧追“赖萧配”，表明郭台



铭的支持者或已归队，浅蓝“弃保”选择柯文哲的可能性大大降低。国民党线下造势活动的热情程度也凸显气势回归。26日侯赵台南造势获得超3万人力挺，直接对赖清德本命区造成冲击。蓝营网民在网络纷纷留言“封路区人全满，正气让阴霾全都走掉了”“好多年轻人参与，超夸张”“团结就会赢！投侯康，台湾有好康。”朱立伦趁机宣布计划选前开展15场造势活动，推动更多选民票投国民党。



图 5 岛内蓝营网民脸书留言

第三，绿营转化策略猛攻赵少康。绿营对赵少康也表现出少见的紧张和严阵以待，在以下四个方面对其攻击，并引发赵强势回击：一是称赵少康不请辞中广董事长，政媒结合，有违广播电视法，NCC 发函要求中广说明如何落实选举期间的公平报道；二是指责赵少康是台湾刻意区分、制造省籍差异的鼻祖；三是批赵少康坚持九二共识、义务役缩减至4个月等政见，意图让台湾卸下防卫能力，助力大陆统一目标；四是离间侯赵，质疑赵少康声量过强，蓝营选战“政出赵少康”，



明目张胆让侯友宜沦为傀儡，搞“双总统制”。

(二)2024 选情聚焦两岸经济议题，蓝白提重启服贸协商，侯友宜抛开放陆生来台学习就业，引蓝绿互呛。

本周选战三方陆续释放政见，两岸经济问题成选战热点攻防区。一方面，蔡当局百万陆客目标跳票引发民意不满；另一方面，侯友宜在“台湾经济发展论坛”抛重启服贸、货贸谈判，开放陆客陆生来台就读就业，柯文哲也在该论坛提重启两岸服贸谈判，表示要推动《两岸协议监督条例》后再讨论货贸、服贸，引发岛内舆论热议。在民进党引导下，岛内对侯友宜开放陆生就业存在疑虑，正面率仅为 6%。

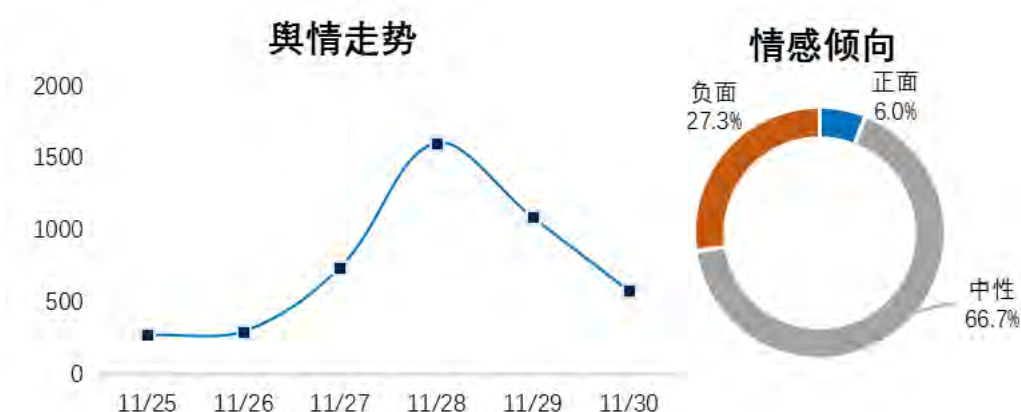


图 6 重启服贸争议舆情发帖量 单位: 篇/条

第一，侯友宜开放陆生就业成白绿攻击的舆论主战场。柯文哲较为克制，认为开放陆生就业，必须要有计划，先计算容量及方法，政府要能掌握风险。民进党为打蓝主力，绿营高层及网民分进合击，一是攻击侯友宜走依赖中国的老路，为讨好中国牺牲台湾；二是质疑将造成台湾大量失业问题，



冲击社会；三是指责中国喜欢用成绩优异的留学生做间谍，国际对大陆学生的警觉已是趋势，蓝营却反其道而行之。甚至还关联赵少康“20 几岁买房”以及检方侦办高雄“台湾中华泛蓝协会”成员接受大陆国台办等机关接待所引发的负面声量，反讽“国民党年轻人选票太多了”“侯友宜赵少康去跟每一个大学生讲，看大学生会有什么反应”，进一步削弱国民党抬高青年支持度的可能。岛内网民也发帖表达担忧，认为“中国大陆薪资过低，开放陆生就业会拉低台湾人工资，损害大学生利益”。

其次，蓝营反击民进党抹红，但国民党部分人士对开放陆生就业存在保留。侯友宜澄清此政见沿用陆生三法，并依照两岸监督条例进行审查；指责绿营对其抹红。朱立伦指责民进党“膝盖式”“台独式”反应，暗设抗中立场，只看颜色不问对错。不过，国民党部分人士也对此政见有所保留，韩国瑜点到为止表示尊重，桃园蓝营县长张善政表示还需要再设计、再商量，会选择性支持。一些蓝营议员表示期待侯竞办会补充关于陆生来台的政策细节，并指责民进党吸引陆生不够，只想享受两岸和平红利，却不愿释放任何善意。

第三，蓝白重启服贸协商的政见引蓝绿互斗。绿营对蓝白双线打击，指责柯文哲为选举再炒服贸议题，背叛从政初心，时代力量创始人林昶佐甚至还以此为借口顺势加入民进党。柯粉迅速发起反击，台湾网红“馆长”直言林昶佐无耻，视频引发 15.3 万次观看。对于侯友宜，民进党则继续抹红战法，



民进党发言人张志豪指责侯友宜对国际及两岸经贸情势变化认知肤浅，主张荒谬，是在“请鬼拿药单”。绿营媒体借机宣传 28 日台股 31 年来首次超过港股，绿营网民评论“票投国民党，台湾变香港”，指责侯友宜重启服贸谈判是在向中国输诚。相应的，蓝营反击较为被动，蓝媒挖出赖清德批两岸三通会木马屠城的历史亚纶，反讽民进党只想以经济红利骗票，要求赖清德回应是否停止 ECFA、两岸直航，并删除台独党纲。

（三）民众党要员现“跳船潮”，“柯盈配”被边缘化。

“蓝白合”破局后，民众党多名要员宣布退党。曾代表民众党参选基隆民代的徐佑升、彰化县民众党党员陈重嘉、陈鎔鎔分别于 23 日晚、24 日、28 日宣布退党，成立刚满一月的彰化民众党党团解散。退党潮引发民众党党内低气压，岛内对柯文哲选情担忧，舆论负面率达 27.6%。



图 7 民众党退党潮舆情发帖量 单位：篇/条

第一，“蓝白合”破局成民众党退党潮主因。民众党 5 日



3 要员退党，是“蓝白合”破裂后的连锁反应。一是蓝营封杀柯文哲，柯曝光度降低，各民调应声大跌，民众党不分区“立委”选势危殆。君悦饭店闹剧后，朱立伦感叹备受屈辱换不来真心，严禁蓝营不分区“立委”帮助或接受柯文哲与吴欣盈站台，避免其利用站台机会牟取宣传机会。二是蓝绿媒体一致唱衰民众党获胜可能，削弱其支持者信心，以分化柯粉促成弃保。岛内媒体认为，民众党人少钱少，无法为参选民代提供支持；台湾南部独特的政治生态使选举严重依赖基层动员，民众党刚成立 4 年难有财力支撑大型造势活动，为选民提供信心；柯文哲金主多支持“蓝白合”，合作破局后许多金主撤资，最终只能选择新光大公主吴欣盈为副手。三是蓝白分裂后民众党内士气低迷，陷入对破局罪人声讨的内斗。徐佑升退党后言称民众党已背离最初信念，对柯文哲失望，声讨黄珊珊“球员兼裁判”。黄珊珊脸书被灌爆，怒轰黄珊珊白皮绿骨、自私害惨民众党，支持者流失现象严重。

第二，柯文哲及高层极力控制党内低气压，跳船不安情绪或难以遏阻。柯文哲、陈智菡、黄国昌等高层驳斥退党假新闻，呼吁团结，并以强大的募款能力提振士气，试图扭转党内低迷气氛。柯粉接受号召，一方面将矛头指向蓝营，指责退党 3 人均与国民党有关。另一方面继续展示对柯文哲的支持，肯定第三势力崛起才是台湾未来。不过台湾地方团体素来西瓜偎大边，当前柯盈配民调支持率已跌破 20%，加上民众党陆战（基层组织）吃紧，岛内不安情绪一时间恐怕



很难阻遏。

三、舆情研判及建议

“蓝白合”破局、郭台铭退选后，岛内三方角力态势更加明显。柯文哲的优势在蓝绿夹杀下逐渐丧失，很可能面临边缘化境地，支持者被分化。岛内“总统选举”大概率呈现蓝绿对决局面，“侯康配”进一步整合泛蓝团体、争取年轻群体支持，蓝营士气被提振，获得与绿营一战的实力和可能。“赖萧配”持续释放对陆善意，争取对等谈判的道德主动权；将“大陆无武力攻台时间表”作为选举武器，瓦解国民党两岸政见的基石；进一步开展侯友宜“亲中卖台”认知作战，重掀反服贸历史记忆，甚至不排除蔡英文后续辅选，发布政策包以消解国民党获胜可能，继续践行其“台独”党纲。

我方一是宜继续关注岛内动向，警惕蔡政府以发放政策资源辅选“赖萧配”，瓦解国民党推动两岸对话政见的行动；二是关注“吴欣盈双重国籍争议”的走向，若其持有美国籍，则可能导致柯文哲不得不退选，正式确立蓝绿对决局面，很可能引发民进党对蓝军更猛烈的攻击，波及两岸关系，需做好相关应急预案；三是认清民进党“和大陆交战不是选项”“邀请习近平吃美食”等选举语言背后的对抗本质，做好“赖萧配”赢得“选举”的相关预案与推演。四是宜发动网军破解绿营对蓝营“亲中卖台”、将服贸等同于受大陆压制的认知作战。

中科天玑数据科技股份有限公司



2023 年 12 月 1 日

互联网AI算法

功能大类

应用推荐类

用户画像类

功能小类

过滤

排名

智能创作

智能匹配

模型大类

矩阵分解

协同过滤

排序学习

内容生成

特征学习

实现

PMF

CMF

XGBoost

VAE

GAN

CNN

RNN

GNN

概率矩阵分解

协同矩阵分解

决策树

自编码器

生成对抗网络

卷积神经网络

循环神经网络

图神经网络





中国科学院网络数据科学与技术重点实验室
Key Laboratory of Network Data Science & Technology, CAS

网络数据科学与技术重点实验室介绍

网络数据科学与技术重点实验室成立于2013年。实验室主任程学旗研究员。实验室在学术委员会的学术指导下，围绕大数据与数据科学展开研究，在学科建设、队伍建设、人才培养、成果转化以及对外学术交流等方面均取得了丰硕的成果，已发展成为我国大数据领域的重要基础研究基地。

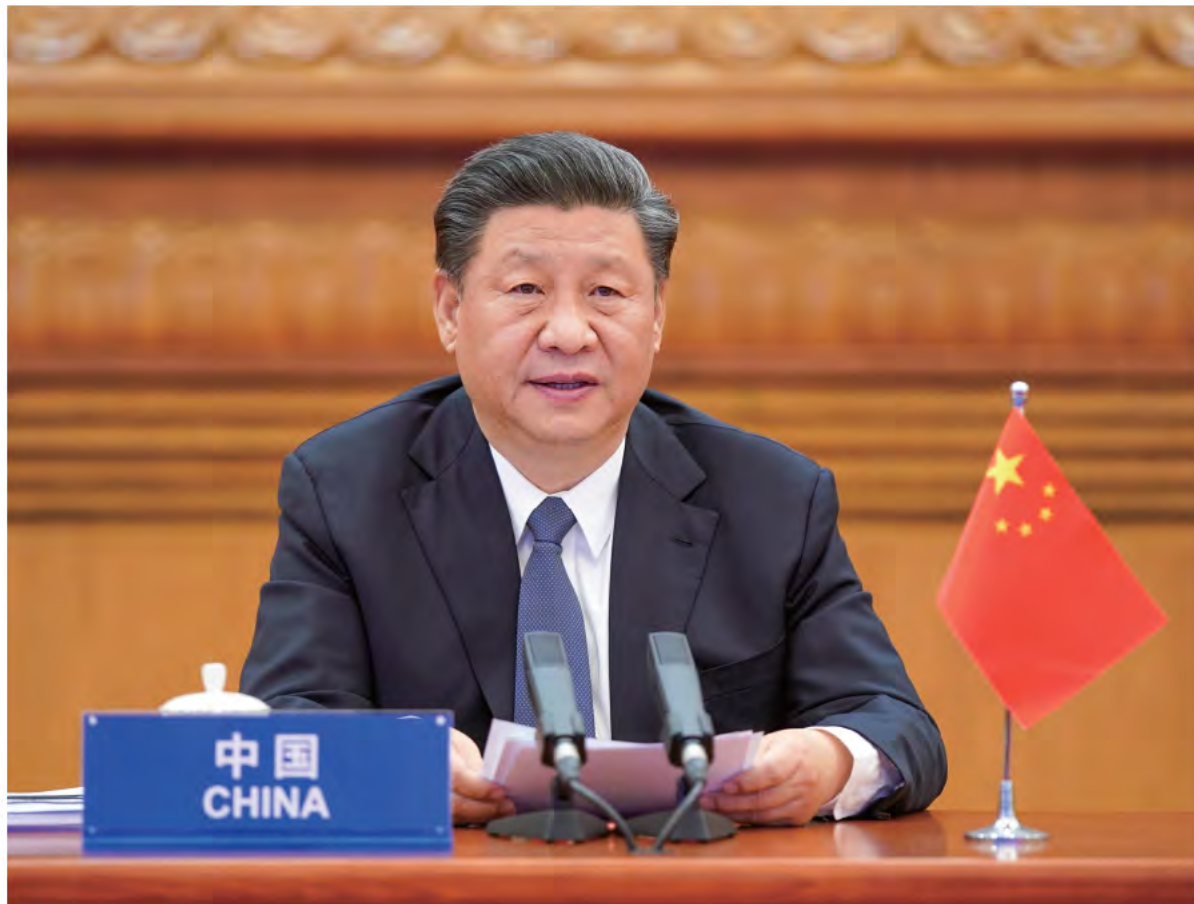
实验室立足国家在网络空间的重大战略需求以及大数据与“互联网+”的产业发展需求，开拓数据科学与技术这一涉及数学、物理、系统科学、社会科学、网络科学等多个领域的新兴交叉学科方向，促进我国大数据与数据科学生态的建设。同时，打造具有国际影响力的研究团队，力争取得被国际同行认可的高水平的基础研究成果。

实验室主要研究方向为面向网络空间的大规模数据感知与获取、存储与管理、分析与挖掘等方面的基础理论、关键技术与应用系统。具体包括数据科学基础理论、网络空间感知与数据表示、大数据存储与管理、网络数据挖掘和社会计算、大数据分析引擎与系统等，以此支撑国家重大任务，并推动网络数据的产业发展。实验室近年来承担多项国家973计划项目、国家863计划项目、国家自然科学基金重点项目的研制，在国际重要期刊和学术会议上发表论文数百篇，在国内外产生了重大影响。相关科研成果获得多项国家级和部委奖励，其中包括一项国家科技进步一等奖、三项国家科技进步二等奖。大量成果被应用于多个领域。

同时，实验室牵头大数据分析系统国家工程实验室的建设，旨在突破大数据分析的核心技术瓶颈，研制第三代大数据分析软件栈，实现理论、架构、算法和接口的整体性、系统性突破，在科学发现、智慧城市、社会安全等方面形成重要应用。并结合行业和地方产业的需求建立示范基地和分实验室，培养和汇聚大数据分析系统研发与应用高端人才，构建大数据分析技术开放的生态体系，提高我国大数据处理及分析技术水平，提升我国大数据领域自主创新能力，促进我国大数据产业的快速发展。



能宣传研究背景



“互联网已经成为舆论斗争的主战场。在互联网这个战场上，我们能否顶得住、打得赢，直接关系我国意识形态安全和政权安全。”

“要增强阵地意识，敢于发声、敢于亮剑，尽快掌握网络舆论场上的主动权，不能被边缘化了。”



算所战略安排+系统部+核心成员

面向国家网络空间治理的战略需求，中科院计算所将智能宣传列为8大优先支持方向，在中科院网络数据重点实验室成立了专门的系统部，在尚未获得国家项目和经费支持的情况下，由程学旗副所长、沈华伟研究员带领50余人的团队持续攻关，研制了跨社交媒体平台的一站式网络空间智能宣传系统。



程学旗 / 研究员



沈华伟 / 研究员



尹芷仪
高级工程师



王永庆
副研究员



庞亮
助理研究员



高金华
助理研究员



孟剑
工程师



王清林
工程师

其他工作人员

过慈双

杨硕

赵鑫

马东晶

刘金虎

朱川生

肖飞

史铂深

孙慧慧

刘丽丽

郭世钰

薛雪

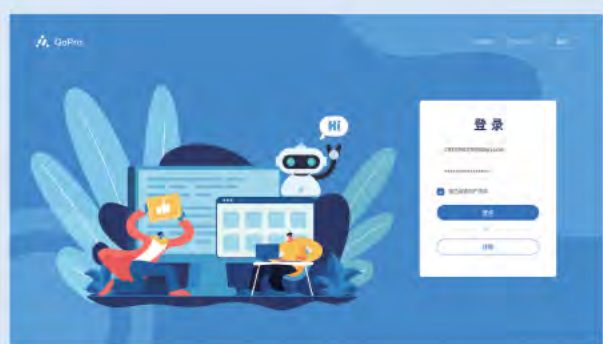
王志良

刘昱希



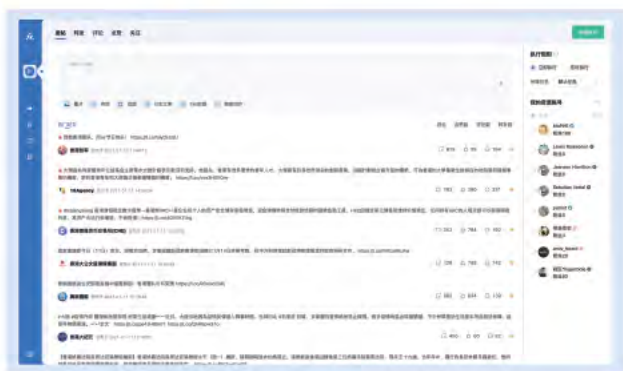
智能宣传系统总概

系统着眼于新时期网络舆论引导和认知对抗的迫切需求，创新利用人工智能、大数据、社会心理、信息传播等技术成果，研究网络舆情智能传播、引导与利用技术，研发面向境外非受控环境的智能宣传系统



登录界面

发帖界面



热点发现



群体监测





特定区域政治态势的智能感知



网络空间大数据感知,包含3大类(传统媒体、社交网站、即时通信)、11种、800万个信源,实现了全通道的实时感知,形成了对业务信源的全覆盖,系统保证持续稳定在线运行,基于WDE平台广泛采集特定区域政治、外交、军事等方面的数据。目前共获得12项软著、6项专利,包括论坛采集软件、天玑网络数据引擎系统、大规模网络数据的多信息来源采集方法和系统等



通过自然语言处理技术实时聚合每日热门政治事件

深度分析政治事件的热度、影响力、相关话题,以及主要媒体和跨平台社交账号的情感倾向、政治立场和观点看法





境外目标对象的智能锁定

实验室积累十年的情报分析技术和GoIN系统，深入挖掘热门政治事件的幕后组织和关键人物



针对香港国安法事件，从监测的18万香港Twitter用户中，挖掘出与该事件相关的3000多名主要人物，100多名网络意见领袖



利用画像技术多维度刻画人物的行为特征、活动规律、政治倾向、观点立场等，为舆论引导提供精确的目标对象



输入式宣传内容的智能生成



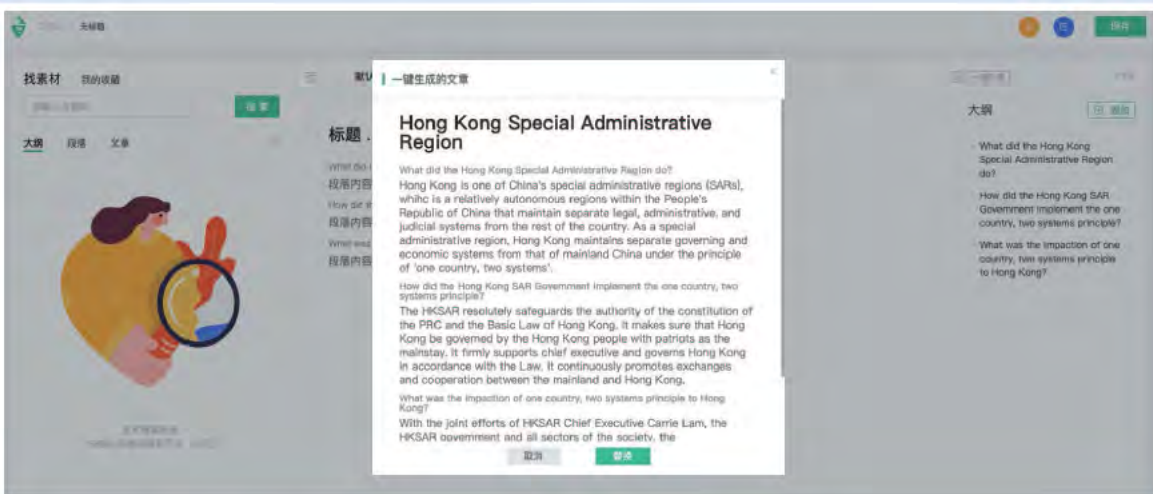
联想功能可自动续写与输入内容相关的一段文本



相关素材推荐为创作者提供多源海量素材



衍生文案能基于原文稿洗出若干篇语义不变、表述不同的新文稿



智能创作支持一体化的长文本内容创作



受控环境下的智能推送

在Twitter、Facebook、Telegram等平台培育了数百个拟人化的机器人账号

账号状态	不通过	注册时间	注册时间	注册时间	注册时间	注册时间
粉丝数	不通过	发文章	不通过	发文章	不通过	发文章
总计: 120个社交账号						
社交账号	粉丝数	发文章	注册时间	最近登录时间	状态	
MuPHE @MuPHE19	0	19	2020-12-11	2020-12-11 16:08:20	已认证	
Lewis Raikonen @LewisRaikonen	0	0	2020-12-11	2020-12-11 16:08:20	已认证	
Lewis Raikonen @LewisRaikonen	0	25	2020-12-11	2020-12-11 16:08:20	已认证	
Jansson Hämäläinen @JanssonHämäläinen	0	3	2020-12-11	2020-12-11 16:08:20	已认证	
Sebastian Vettel @SebastianVettel	0	9	2020-12-11	2020-12-11 16:08:20	已认证	

提供一站式跨社交媒体操作和调度平台，实现自动批量发帖、评论、转发

发帖 转发 评论 点赞 关注

众志成城，抗击疫情，全力恢复经济社会秩序，促进经济社会的发展，正式党和国家秉持着生命至上、人民至上的崇高理念。

我们坚信，我们的人，正是为了胜利。我们坚信，正是以全体人民的团结一心，我们坚信，正是以全体人民的团结一心，我们坚信，正是以全体人民的团结一心。

热门新闻 热门新闻

6月18日，国家主席习近平在第七十五届联合国大会一般性辩论上发表题为《坚定信心、同舟共济、科学防治、精准施策，坚决打赢疫情防控阻击战》的重要讲话。讲话全文如下：同志们、朋友们！当前，新冠肺炎疫情仍在全球蔓延，给人类生命财产和经济社会发展带来沉重打击。在中国共产党的坚强领导下，中国人民众志成城、抗击疫情，取得了重大战略成果。在此，我谨代表中国政府和中国人民，向国际社会致以诚挚的问候和衷心的感谢。中国将始终秉持生命至上、人民至上的理念，为全球疫情防控贡献中国智慧和力量。

马海霞 中国新闻网 2020-06-18 14:22:01

中国国家统计局22日在第七十五届联合国大会一般性辩论上发表重要讲话，指出：疫情对世界经济造成冲击，各国经济陷入衰退，生命至上理念，生命至上理念。

执行规则

立即执行 立即执行

我的资源账号

MuPHE 粉丝165

Lewis Raikonen 粉丝0

Jansson Hämäläinen 粉丝0

Sebastian Vettel 粉丝0

patfor 粉丝0

军迷精英 粉丝0

snrk_board 粉丝20

新冠Hyperbole 粉丝2

MuPHE 28岁 英文

2020年11月 38人

17 正在关注 1 关注者

你关注的人中没有人关注

博文 博文和回复 媒体 留言

众志成城 抗击疫情

Lewis Raikonen

个人资料

所有照片

照片

众志成城 抗击疫情

中华人民共和国外交部

好友 1位好友

全部好友

帖子

筛选条件

Lewis Raikonen 9分钟 · 0

众志成城，抗击疫情，全力恢复经济社会秩序，促进经济社会的发展，正式党和国家秉持着生命至上、人民至上的崇高理念。

众志成城 抗击疫情



全球多社交平台养号系统



1、我们的优势

各个社交平台养号成活率的主要影响因素



01.被平台识别为机器人，而非真实用户

各个社交平台基于众多交互点，发现用户的可疑行为，进而判定是否为真实个人用户。各平台机制不同，但比较典型的包括：

- ① 人机验证，例如各个平台都会调用Google reCAPCHE
- ② 人脸照片验证，例如Facebook
- ③ 其他各种可疑活动的行为，例如社交登录以及发帖异常，前一分钟在俄罗斯，后一分钟在美国...



02.违反社交平台社区运营守则

各个社交平台都会制定一系列的社群守则，规定了允许和禁止的内容和行为。例如Facebook推崇实名制、保障用户社交内容真实可信、确保安全不受威胁以及保护个人隐私等等；

因此，我们在养号过程中需要注意绕过平台的行为识别机制，避免账号被封。



1、我们的优势

1、通过技术和管理手段，模拟最真实的用户行为



一人一机一（组）账号

养号系统采用最真实的“一人+一机+一组账号”的方式，防止社交平台识别用户个人特征，造成多个账号关联以及封号。



真实、完备的虚拟人物信息

养号系统中每个账号背后都存在一个信息完备的虚拟人物，包括人物姓名、国籍、头像、手机号码、邮箱、使用语言，以及人物的通讯录、常用IP等。



模拟真实的社交行为

- 养号系统中的养号动作进行严格的策略控制，确保执行真实可信的社交行为。
- 养号系统中虚拟人物进行人物标签化管理，进而执行符合人物属性的社交行为。



1、我们的优势

2、支持全球多个国家的主流社交媒体平台



Facebook



Twitter



LinkedIn



Instagram



LINE



WhatsApp



telegram



Pinterest



Tiktok (海外版抖音)



YouTube



Reddit



Medium



1、我们的优势

3、提供各平台完备的社交行为策略库，并持续动态更新

1、各平台社交行为基准策略库

在养号系统中，针对各个社交平台定义了严格的基准策略，即禁止在该平台中执行的社交行为。基准策略作为账号养护的最基本依据；基准策略在web端不可见，但却是养号行为的基础依据。

2、各平台社交行为配置策略库

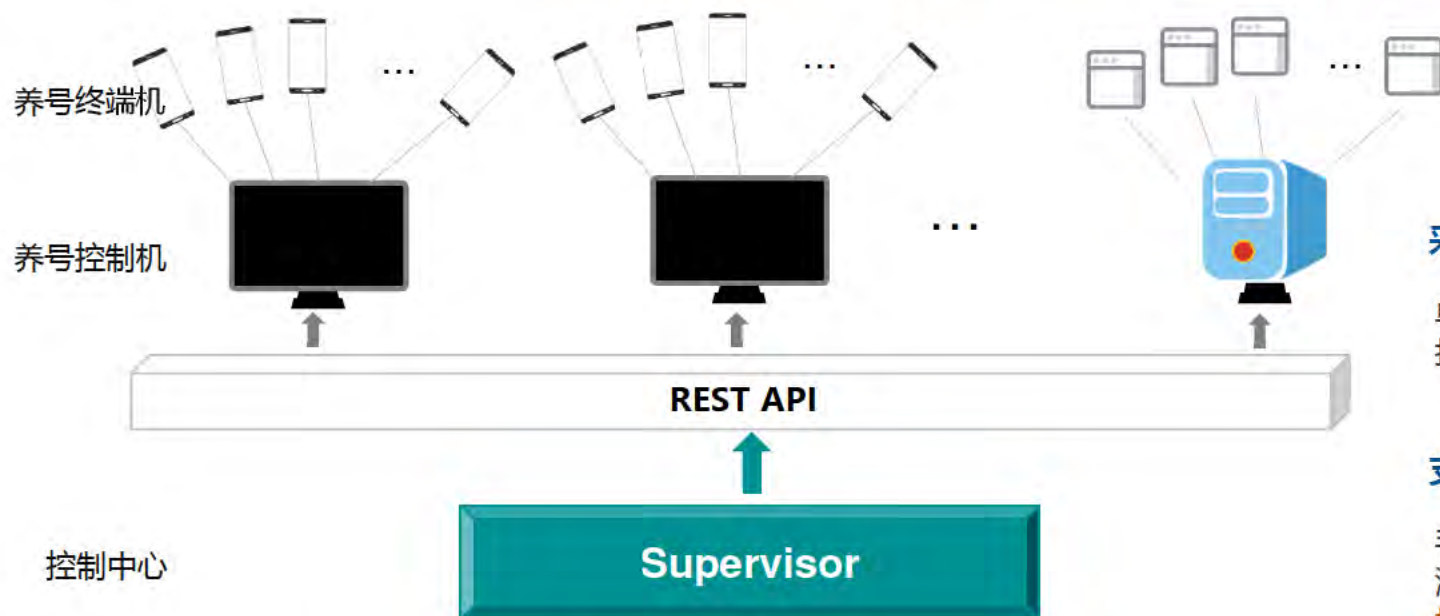
不同的场景下，对应的养号的业务目标不同，例如耐用号、影响号、广告号、水军号等，以及不同阶段的账号养号策略也不同，例如新号、刚解除安全模式的账号、老号等。

养号系统中支持基于不同属性的虚拟人物，（例如国籍、使用语言、所在时区以及兴趣爱好、社会角色和政治倾向等），定义与其人物属性相匹配的社交动作和社交动作组，来支撑不同场景下的账号的日常养护。



1、我们的优势

4、采用多层次、多模式群控，可实现设备数量无限扩展



采用“控制中心+控制机+终端机”的三级控制机制

单台主机可控制50+部手机，主机数量可无限扩展，支持设备数量无限制扩展。

支持手机真机、浏览器多种养号终端

手机真机适用于影响力账号培育和养护
浏览器适用于大规模账号养护，支撑各类养号场景需求。





2、人物档案—虚拟人物管理



Fan fangshuang

基础信息

*人物姓名
*性别
*国籍
宗教
*出生日期（年龄）
*居住地&时区
*家乡&语言
*手机号码
*邮箱
人物简介
*人物头像
...

人物履历

工作机构
职位
起止时间
教育机构
专业（或学位）
起止时间

绑定设备

当前绑定:
设备编号
序列号
绑定时间
管理人员
设备类型
设备品牌

历史绑定记录:
设备编号
解除绑定时间
...

社交账号

各平台社交账号信息:
用户名
密码

通讯录

联系人信息:
姓名
邮箱
手机号码
公司
职位
...
**添加联系人
批量更新**

管理信息

虚拟人物标签:
使用语言
时区
社会角色
兴趣爱好
政治倾向
...
管理人员绑定



2、人物档案—虚拟人物管理

基本信息

履历信息

绑定设备

网络代理

社交账号

通讯录

头像

Generaedit

Setup general profile details.

*First Name

Barry

Middle Name

Middle Name

*Last Name

Waterson

性别

男

国籍

美国

宗教

宗教

语言

英语

出生日期

1985

6

5

所属时区

Asia/Tokyo+0900

出生地（家乡）

出生地（家乡）

居住地

日本 ✕

手机号类型

V

Google Voices

请输入

*手机号

+1(575) 383-4582

Emails

⊕ 添加邮箱

*主邮箱

barrywatterson327@gmail.com

*主邮箱密码

zaq1@WSX!@#S%

辅助邮箱

辅助邮箱

*主邮箱

*主邮箱密码

辅助邮箱



2、人物档案—社交账号管理



Fan fangshuang





2、人物档案—社交账号管理

账号列表(3692)

Twitter

Facebook

LinkedIn

Instagram

Telegram

Whatsapp

Line

Youtube

Tiktok

Wechat

Pinterest

Medium

Reddit

关键词

国家/地区

不限

账号状态

状态

筛选

Screen Name	虚拟人物	国籍	手机	邮箱	操作
+9509258581483	Rae Berg	缅甸	+9509258581483	raeberg21@gmail.com	查看
+447533486994	Grace Garcia	英国	+447533486994	g.grace47@yahoo.com	查看
+447878630193	Janice Doherty	英国	+447878630193	janicedoherty10@yahoo.com	查看
+447402565719	Iris Hughes	英国	+447402565719	Irishughes71@yahoo.com	查看
+447832996230	Gill Price	英国	+447832996230	gillprice24@yahoo.com	查看
+447878855497	Polly Green	中国香港	+447878855497	pollygreen27@yahoo.com	查看
+85254269811	Jiaqi Lin	中国香港	+85254269811	linjiaqi592@gmail.com	查看
+85251346349	Lin wanrong		+85251346349	linwanrong1102@gmail.com	查看
+13473952646	Ava Bird	美国	(347)395-2646	sfgg6436@gmail.com	查看
gaokai1102@gmail.com	Gao Kai	中国台湾	+6721-24537	gaokai1102@gmail.com	查看



3、资源库—多媒体素材库





3、资源库—邮箱&手机号&Google voice&Proxy

手机号资源库

提供养号所需的各个国家及地区的手机号资源，资源库中提供手机号的信息记录和管理，以及绑定关系、充值记录、有效期等。

邮箱资源库

提供养号所需的各个非匿名邮箱服务商的邮箱资源，例如Google、Yahoo、outlook、ru、yandex。资源池提供邮箱的信息记录和管理，使用状态管理。

Google voice资源库

提供养号所需的Google语音验证账号资源，提供信息管理及状态管理

Proxy 资源库

提供养号所需的可靠的网络代理资源，包括SSR、V2Ray等多种代理节点，资源池提供信息管理、使用状态管理。

通讯录资源库

提供养号所需的各种属性的人物联系信息，例如某国海军指挥官、某地区新闻记者、某特殊组织人物等，资源池提供各类通讯录的信息管理和分配。



3、资源库—目标群体账号库

全球核心人物社交账号

持续监测全球军、政、商、科技、传媒、民间组织等领域的领袖及核心人物的多个社交平台账号。

数量：40W+

全球知名机构社交账号

持续监测全球政治、军事、科技企业、智库、高校以及非政府组织等知名机构的多个社交平台账号。

数量：46W+

其他目标群体账号

持续监测各个社交平台中各类非知名人物账号，例如“**全球对华友好言论人物**”，“美国特定标签WhatsApp用户”，“香港特定标签LINE用户”等数百种目标群体账号。

加好友
社交互动
好友推荐

社交关系维护

养号系统账号池

提供全球范围内各类目标群体账号，我方账号可进行加好友、社交互动等动作进行“特定圈子”的社交关系维护



3、知识库—养号行为策略库

发帖
浏览主页
浏览时间线
更新profile
转贴
转发带评论
点赞
关注
回复
加好友



发推
浏览主页
浏览时间线
更新profile
转推
转发带评论
点赞
关注
回复



加好友
浏览主页
添加关注
分享
点赞
评论



浏览
发帖
点赞
评论
关注



浏览Following
浏览For You
查找好友
点赞
评论



浏览Chats
Add people Nearby
Invite Friends
发送好友消息
calls



更新状态
好友calls
浏览chats
发送好友消息



浏览时间线
添加通讯录好友
Invite Friends
发送好友消息
Add Story
浏览chats
calls



浏览主页
添加关注
随机收藏



行为策略库控制养号系统账号的社交行为

行为策略库

基准策略

社交动作执行过程中，基准策略用于控制账号不会执行在社交平台中认为危险的、不被信任的行为和动作。

例如：

具有相同用户个人特征的人注册多个Facebook账号，比如相同的邮箱、手机号码以及cookie或者设备标示信息等。
或者新的WhatsApp账号频繁单向添加好友，而对方通讯录没有保存你的号码等行为等等。

配置策略

配置策略中定义了相同属性的人物进行账号日常养护所遵从的行为，包括：

- 在不同的社交平台中执行哪些动作及动作组（包括频率）
- 账号在不同的养护阶段分别执行的动作及动作组（包括频率）
- 根据人物的所在时区控制该账号的活跃时间段
- 根据账号养护目标（耐用号、影响号等）区分执行不同的动作及动作组（包括频率）



4、服务模式



解决方案整体交付

提供养号的全套一体化设备、包括养号手机，境外代理服务等。
提供远程运维及技术支持



账号交付模式

将培育的虚拟账号全体资料交付给用户。
可以选择虚拟账号使用的移动终端同时交付。

A light gray world map is centered in the background of the slide, showing the outlines of all continents.

THANKS



需求		
平台需求	关注国家或地区 (尽量选填)	巴基斯坦、伊拉克、柬埔寨、印度尼西亚、缅甸、尼日利亚、埃塞俄比亚、中非、刚果金、安哥拉；阿富汗及吉尔吉斯、哈萨克斯坦、乌兹别克斯坦等中亚国家
	关注地域 (选填)	中资企业项目所在地区；高风险地域（如易发冲突的边境区域；暴恐、分裂、宗教极端势力活动频繁区域）
	关注人物 (尽量选填)	政府高层、政党领袖、军队高层；大的财团、势力家族、部落首领；宗教领袖；恐怖分裂组织头目等
	关注组织 (尽量选填)	各国主要政党、主要国际或地域性组织（如西非国家经济共同体、东非共同体、中非国家经济共同体；阿拉伯国家联盟；英联邦；欧佩克、世界粮油组织、红十字会等）、当地活跃的NGO、各类恐怖分裂组织；主要情报机构（如美国的FBI、CIA；英国的情报局；五眼联盟）；知名的安保公司（如Control Risks Group; 俄罗斯的瓦格纳、美国的军事资源公司等）
	关注事件 (尽量选填)	涉中资事件；所在国别安全事件（政府更迭、政变、战乱、暴恐、群体事件、犯罪、环保、公共卫生；边界冲突）；所在国重大政治、经济、外交、贸易、税收、投融资、民生政策变化
	关注媒体 (尽量选填)	当地主流媒体、西方主流媒体（BBC、CNN、半岛电视台；路透社、俄通社、美联社）
	关注社交媒体 (尽量选填)	社交媒体（X、facebook、WhatsApp）等
	其他需求	
备注：		



下一步工作考虑



基本思路

- 以撰写专报、要报为主
 - 特点：被动式、触发式、浅层次
 - 作业方式：选择业务方向、捆绑业务单位
 - 能力展示方式：提供专报、要报
- 以系统建设为主
 - 特点：主动式、持续式、通用型
 - 作业方式：2-3个月迭代出智能宣传系统，涵盖群体画像、智能创造、智能推送、选举预测分析等模块
 - 能力展示方式：系统演示汇报、实际运用



撰写专报、要报

- 业务主题选择
 - 选举类：香港选举、台湾选举、澳门选举、美国选举
 - 政保类：政治谣言、境外访民干扰高访、境外颜色革命倒灌
 - 地缘政治类：南海、台海、朝核
 - 暴恐类：“一带一路”沿线国家暴恐态势、新疆、西藏
 - 突发类：新冠疫情、自然灾害
- 难点
 - 缺乏特殊数据支持
 - 缺乏业务场景
 - 分析深度、内幕性、现索性不够



撰写专报、要报

- 解决方案
 - 捆绑业务单位：**基地，**局，承担案头工作，收集整理开源QB
 - 捆绑社科单位：现代院、***，承担定量分析工作，提供开源背景和数据分析结果



系统建设

- 发散到聚焦
 - 业务方向：香港立法会选举
 - 语种：粤语、英语
 - 能力：能在香港市民活跃网络平台发声
- 复杂到简单
 - 在Facebook\Youtobe\Twitter\Telegram养一批账号
 - 在反中账号后面发出理性的言论（人工加算法），引导反中群体分化
- 自研到集成
 - 调研现有技术，尽量集成或改装



总结

- 根据会议决议，确定后续工作基本路线
- 制定工作方案和进度推进计划
 - 系统建设：建设周期、建设成效、经费投入、人员投入等
 - 专报要报：主题、单位、视角、频次等



开放课题讨论会

专家组办公室

2022年12月



汇报提纲

01

开放课题情况介绍

02

调研企业情况汇总

03

当前阶段问题汇总



开放课题情况介绍

- 培育项目首批课题设置了3个开放课题，具体信息如下：

序号	专题	课题名称	密级	负责人	外协经费/总经费 (万元)
1	理论	基于主流商用渠道的境外社会认知分析与透明化利用	内部	庞亮	285/300
2	分化	面向TD地区RZ分化与塑造的叙事化**资源库构建	MM	王永庆	285/300
3	塑造	***身份塑造培育及形象展示技术	MM	程学旗	475/500

开放课题设置目的：开拓J民合作渠道，探索RZ对抗方向民间力量的运用
(以民企及相关研究所的力量为主)

开放课题情况介绍

1. 理论专题——基于主流商用渠道的境外社会认知分析与透明化利用

➤ 研究目标

组织开放创新课题群，与互联网商业平台合作，推进关键社会认知对抗测绘、分化及塑造技术的大规模隐蔽实验验证工作的开展。

➤ 主要牵引性指标+考核方式

1. 针对测绘、分化和塑造三个技术专题，进行总数不少于20次的中小规模测试；

具体包括：1) 结合特定平台，针对典型的境内外热点议题（包括：经济、文化、政治、体育等类目下的诸如选举、台积电、中美关系等），分析境内外用户（包括：中国大陆、台、美、日、一带一路周边国家）的用户兴趣情况、用户年龄层次分布情况、传播情况分析，议题数量不少于5个；



开放课题情况介绍

1. 理论专题——基于主流商用渠道的境外社会认知分析与透明化利用

➤ 主要牵引性指标+考核方式

1. 针对测绘、分化和塑造三个技术专题，进行总数不少于20次的中小规模测试；
- 2) 结合特定平台，面向经济、文化、政治、体育等话题下的实时热点争议事件，分析事件争议点及人群特点，针对争议点及人群特点，设计具备平息议论的导向性内容（包括：广告、限制特定人物发言、删除内容、投放引导性内容等），并在10分钟、30分钟、1小时内产生显著效果的，实验数量不少于10个；
- 3) 针对境外平台（Twitter、Facebook、Tiktok等）设计5套可针对不同年龄段（青少年（10-20岁、20-30岁）、中壮年（30-40岁、40-50岁）以及中老年（50-60岁））、不同地域（台、美、日、一带一路周边国家）人员特点的社交账号培养策略，并交付至少5个1万粉丝上的社交账号，粉丝画像与设计策略应保持一致。





开放课题情况介绍

1. 理论专题—基于主流商用渠道的境外社会认知分析与透明化利用

➤ 主要牵引性指标+考核方式

2. 形成1份在互联网商业平台进行大规模隐蔽认知对抗实验的技术报告；

报告内容应至少包括：1) 涉及境内平台（微博、微信、小红书等）、境外平台（Twitter、Facebook、Tiktok等）包括受众分布、活跃用户情况等分析维度；2) 在典型社交平台上，分析涉及中国大陆、台、美、日、一带一路周边国家，包括青少年（10-20岁、20-30岁）、中壮年（30-40岁、40-50岁）以及中老年（50-60岁）人群的受众特点，维度应至少包括关注议题、流行符号、常用渠道、在线时长等；3) 针对青少年、中壮年、中老年受众，构建有效地营销策略，在1-2个场景中（账号培养运营、电商购物等）验证其有效性；4) 探索可能的海外MCN构建渠道，并对可行性进行分析验证。

开放课题情况介绍

1. 理论专题—基于主流商用渠道的境外社会认知分析与透明化利用

➤ 主要牵引性指标+考核方式

3. 进行总数不少于3次的大规模认知对抗实验测试。
- 实验应满足以下条件：1) 应至少包括一个境内社交平台以及一个境外社交平台；2) 参与用户数量应不少于5万人；3) 话题数量规模不少于20个；4) 用户互动数量应不少于10万次；5) 参与用户满意度；6) 参与受众人群是否与预期匹配。



开放课题情况介绍

2. 分化专题——面向TD地区RZ分化与塑造的叙事化**资源库构建

➤ 研究目标

聚焦TD地区RZ分化与塑造的叙事化**资源库构建的重大需求，开展对特定地区关键人物、关键认知场景、重点议题资源进行收集、整理、标注的研究工作，构建人物素材库、议题素材库、典型场景库，为分化及塑造的叙事化**提供基础资源支持。

➤ 主要牵引性指标 + 考核方式

面向TW及JW典型的新闻媒体平台（包括：台湾联合新闻网、台湾天下杂志、联合早报、路透社、BBC、CNN、纽约时报等），采集至少5万条涉T与至少5万条涉M新闻素材内容，新闻的数据字段应至少包含：标题，来源，时间，正文，地点-精确到县市，作者等，新闻类型可以包含图、文、音、视频等，文本内容以json格式提交，其他形式内容以附件形式提交。其他要求包括：





开放课题情况介绍

2. 分化专题—面向TD地区RZ分化与塑造的叙事化**资源库构建

➤ 主要牵引性指标+考核方式

1. 对新闻内容中所涉及的涉T人物进行标注，区分T“四派一面”（顽固派、友好派、摇摆派、客观派及其对立面人群），其中每类别中的重点人物应不少于1000人；
2. 对新闻所涉及议题进行标注，议题数量不少于100类，应至少包括：，议题应与涉T人物进行关联，在要求1基础上，与“四派一面”中各个类关联的新闻内容不少于5000条。此外，设计自动标注算法，能够对新议题进行自动发现，自动发现的新闻内容规模不少于5000条；
3. 涵盖TW地区13个以上的县市，且每县市级的新闻内容不少于5000条；

开放课题情况介绍

3. 塑造专题——***身份塑造培育及形象展示技术

➤ 研究目标

通过理解TD身份人物的偏好、心理认知和语言风格，构建符合目标人物形象的行为素材、交互策略及形象展示方法。在行为素材构建方面，生成符合TD目标语言个性特征的文本、图片及音视频内容；在交互策略方面，通过知识图谱技术构建数字人知识体系，支撑多轮对话、社交互动等实时交互能力；在形象展示方面，生成具备高还原度的真人数字形象，包括表情、姿态、语音、服饰、发型等的模拟与切换，培育与人设吻合的数字人形象。





开放课题情况介绍

3. 塑造专题—***身份塑造培育及形象展示技术

➤ 主要牵引性指标+考核方式

1. 支持TW普通话、闽南语、英语等多语种个性化文本生成，其中英语文本生成语言模型困惑度低于10，TW普通话、闽南语文本生成语言模型困惑度低于20；
2. 支持英语和闽南语两种语言的特定目标人语音、视频、图像多模态生成合成，MOS分达到4.0以上；
3. 支持5轮以上的多轮对话，MOS分达到4.0以上；
4. 提供至少100个数字人身份形象，提供至少10种虚拟场景；

开放课题情况介绍

3. 塑造专题一***身份塑造培育及形象展示技术

➤ 主要牵引性指标+考核方式

5. 具备4k高清晰度数字人生成与渲染能力，数字人与真实人物异地互动延时不大于50ms；
6. 提供指定社交平台（Facebook、Twitter、Tiktok等）的特定虚拟人不少于1000个（虚拟人的人设与数字人吻合）。





汇报提纲

01

开放课题情况介绍

02

调研企业情况汇总

03

当前阶段问题汇总

调研企业情况汇总

- 10+企业参与调研，6家企业有意向承担开放课题，意向情况如下：

课题名称	总集单位	分包商	对应课题指标
基于主流商用渠道的境外社会认知分析与透明化利用		厦门市美亚柏科信息安全研究所有限公司	1. 针对测绘、分化和塑造三个技术专题，进行总数不少于20次的中小规模测试 2. 进行总数不少于3次的大规模认知对抗实验测试
		厦门飞博共创网络科技有限公司	3. 形成1份在互联网商业平台进行大规模隐蔽认知对抗实验的技术报告



调研企业情况汇总

- 10+企业参与调研，6家企业有意向承担开放课题，意向情况如下：

课题名称	总集单位	分包商	对应课题指标
面向TD地区RZ分化与塑造的叙事化**资源库构建		北京新奥特集团有限公司	特定地区人群素材库
		厦门市美亚柏科信息安全研究所有限公司	1-2. 按地域（城市、乡、县等）、人物类别“四派一面”（顽固派、友好派、摇摆派、客观派及其对立面人群）构建面向特定地区关键人群素材库
		合肥讯飞数码科技有限公司	特定地区议题素材库 3. 按热点议题层次类别构建面向特定地区“四派一面”关键人群的热点议题语料库 特定地区典型认知场景库 4. 台典型认知场景数量不少于10个，每个场景须包含易感议题、关键人群



调研企业情况汇总

- 10+企业参与调研，6家企业有意向承担开放课题，意向情况如下：

课题名称	总集单位	分包商	对应课题指标
****身份塑造培 育及形象展示 技术		北京中科睿鉴科技有限公司 -指标一	一、面向特定人物风格的多媒体内容生成 1. 支持TW普通话、闽南语、英语等多语种个性化文本生成； 2. 支持英语和闽南语两种语言的特定目标人语音、视频、图像多模态生成合成，MOS分达到4.0以上； 二、面向特定人物形象的数字人形象及模型生成 3. 提供至少100个数字人身份形象，提供至少10种虚拟场景； 4. 具备4k高清晰度数字人生成与渲染能力，数字人与真实人物异地互动延时不大于50ms； 三、数字人交互技术 5. 提供指定社交平台的特定虚拟人不少于1000个（虚拟人的人设与数字人吻合）。 6. 支持5轮以上的多轮对话，MOS分达到4.0以上；
		北京新奥特集团有限公司 -指标一、二	
		北京百度网讯科技有限公司 -指标一、二、三	
		合肥讯飞数码科技有限公司 -指标一、二、三	





调研企业情况汇总

序号	公司名称	合作内容	合作点	企业优势
1	北京新奥特集团有限公司	叙事主题素材库构建	1.音、视频素材库构建	1. 军民融合项目经验。 2. 各类专业数字媒体内容制作及运营解决方案与技术服务。 3. 主要产品：有声像资料编辑系统、智能媒资管理系统、目标伪造与识别、大屏呈现系统（含动态数据呈现）、图文包装、演播系统等。 4. 具备***BM资质
		虚拟人格塑造技术	1.虚拟数字形象生成技术 2.音、视频生成合成技术	
2	北京百度网讯科技有限公司	虚拟人格塑造技术	1.虚拟数字形象生成技术 2.多轮对话技术 3.音、视频生成合成技术	数字人平台特点：1) 数字人形象生成（照片建模PTA、捏脸系统VTA、妆容系统、发型系统、服装系统、服装模型、布料解算系统）；2) 数字互动场景构建；3) 内容生成模块（文本、TTS语音、照片、视频生成；渲染模块）；4) 高精度动作捕捉

调研企业情况汇总

序号	公司名称	合作内容	合作点	企业优势
3	厦门市美亚柏科信息安全研究所有限公司	机制及策略设计	1. 用户激励机制设计与实验 2. 高价值账号培养策略设计 3. 内容投放策略设计	1. 有网络心理战技术与装备研发队伍 2. 具备网络信息采集与传播、账号培育、网络舆论监测与引导等研发力量 3. 具备***BM资质
		叙事主题素材库构建	1. 图、文素材库构建 2. 音、视频素材库构建	
		虚拟人格塑造技术	1. 音、视频生成合成技术	
4	合肥讯飞数码科技有限公司	叙事主题素材库构建	1. 图、文素材库构建 2. 音、视频素材库构建 3. 议题场景素材库构建	1. 主营业务方向为智能语音处理、图像识别、自然语言理解、云平台、大数据技术等。专注于人工智能技术在军事领域的应用。 2. 拥有深度学习及深度神经网络核心算法全链条的研究能力、大规模神经网络的计算能力。 3. 具备***BM资质
		虚拟人格塑造技术	1. 虚拟数字形象生成技术 2. 多轮对话技术 3. 音、视频生成合成技术	



调研企业情况汇总

序号	公司名称	合作内容	合作点	企业优势
5	北京中科睿鉴科技有限公司	虚拟人格塑造技术	1.音、视频生成合成技术 (特定目标人物的内容生成、合成)	致力于运用 AI 技术赋能数字内容安全。以图像、音视频内容为重点的数字内容合成方面，中科睿鉴相关成果已规模化应用于国家部委
6	厦门飞博共创网络科技有限公司	机制及策略设计	1.高价值账号培养策略设计 2.内容投放策略设计	1. 国内自媒体首家上市公司。 2. 内容矩阵顶层设计与制作产出。 3. 社交媒体营销方案的策划及执行。 4. 表情包IP设计与授权融合需求方相关品牌或形象元素设计及授权合作。





汇报提纲

01

开放课题情况介绍

02

调研企业情况汇总

03

当前阶段问题汇总

当前阶段问题汇总

• 课题外协方式设置考虑

- 基于主流商用渠道的境外社会认知分析与透明化利用课题：鉴于商用平台的差异化考虑，建议采用**分包模式**
- 面向TD地区RZ分化与塑造的叙事化**资源库构建课题：因课题指标之间存在强关联性，建议采用**总包模式**
- ****身份塑造培育及形象展示技术课题：考虑到社交平台及塑造账号的类型差异，建议采用**分包模式**



当前阶段问题汇总

- 计算所外协经费使用情况说明：

- 一般流程：召开专家组评审会、形成评审意见表、签订外协合同
- 评审意见表：奇数位专家给出评审意见，评审通过后评审意见表需有专家组组长签字
- 外协合同：专家组组长（及责任专家专家？）签字、计算所公章及法人章（甲方）、承研方公章及法人章（乙方）（涉及BM资质的，需提供资质证明）、经费预算？



讨论内容

- 除调研外企业，其他意向企业及民营研究院所情况
 - 征集：
- 对于分包课题及总集/分包课题的，经费设置比例情况
 - 建议根据承研情况设定范围，与意向单位报价情况匹配，综合确定
- 合同内容设置情况
 - 具体明确企业承担课题的研究目标、内容、考核指标及经费设置





附表

公司名称	合作内容	合作点	参考报价合计
北京新奥特集团有限公司	叙事主题素材库构建	1.音、视频素材库构建	338w
	虚拟人格塑造技术	1.虚拟数字形象生成技术 2.音、视频生成合成技术	
北京百度网讯科技有限公司	虚拟人格塑造技术	1.虚拟数字形象生成技术	500w
		2.多轮对话技术	
		3.音、视频生成合成技术	
厦门市美亚柏科信息安全研究有限公司	机制及策略设计	1.用户激励机制设计与实验 2.高价值账号培养策略设计 3.内容投放策略设计	510w
	叙事主题素材库构建	1.图、文素材库构建	
	虚拟人格塑造技术	2.音、视频素材库构建	
合肥讯飞数码科技有限公司	叙事主题素材库构建	1.音、视频生成合成技术	3589w (包含：单套报价 +调优报价)
		1.图、文素材库构建	
		2.音、视频素材库构建	
	虚拟人格塑造技术	3.议题场景素材库构建	
北京中科睿鉴科技有限公司	虚拟人格塑造技术	1.虚拟数字形象生成技术 2.多轮对话技术 3.音、视频生成合成技术	不明确
		音、视频生成合成技术	
厦门飞博共创网络科技股份有限公司	机制及策略设计	1.高价值账号培养策略设计	28w-65w 定制业务10w-30w
		2.内容投放策略设计	



敬請批評指正！



认知域 ZZ 集中体现为大数据 ZZ 和信息 DK，主要 ZZ 武器是来自三元空间的大数据，ZZ 能力体现为大数据智能分析水平。为了能够获取未来战争的领先优势，本项目拟提供战备级认知域 ZZ 地图、全球舆战态势图和全要素基础资源库、贯穿 ZZ 回路的舆战智能武器库以及平战一体的舆战试验场，为未来认知域 ZZ 提供理论保障和系统基础。

随着网络信息技术的发展，主流媒体把控舆论的能力显著下降。通过互联网、大数据、人工智能技术以及新型社会传播机理、社会心理学机制等，实现对社会群体的情绪和认知引导，已经成为国际上绝大多数国家的国家意志宣传和民众情绪管理的主要手段。在大国博弈日趋激烈的形势下，智能宣传正成为国际政治斗争的新手段、国家意识形态 DK 的主 Z 场、国际贸易 Z 的新工具，也是未来军事 ZZ 的一种新模式，并将在“信息-物理-社会”等更广泛的空间和领域发挥重要作用。

国际上，英国的剑桥分析公司除了助推 2016 年特朗普成功当选美国总统外，事实上已参与 40 多次美国政治竞选活动，并在幕后对乌克兰橙色革命、英国脱欧运动等起到了重要作用。2017 年 6 月英国牛津大学发布的一份研究报告显示，目前约有 20 多个国家在使用机器人水军影响舆论、左右时局。如图 1 所示的一系列政治事件，其背后或多或少地存在着智能技术支撑的舆论引导和民意操纵的成份。

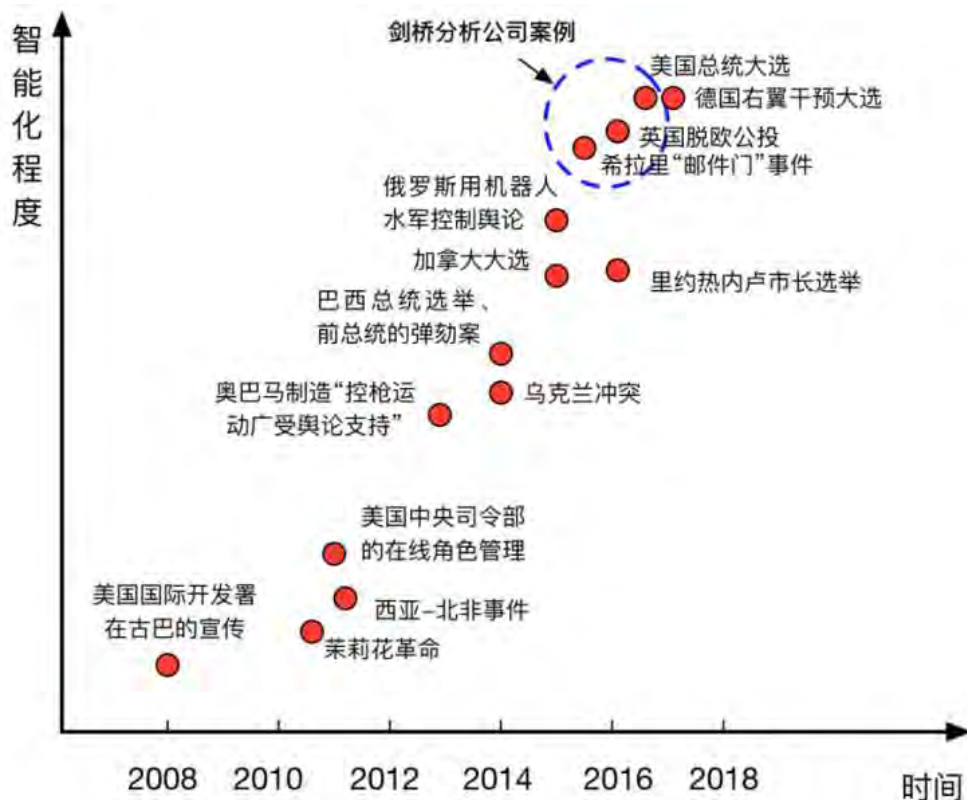


图 1 政治事件背后的宣传智能化发展态势

西方发达国家正大力提升其舆论攻防能力。欧美等西方集团暗中支持和操控各类主打意识形态宣传、渗透、推广的网络机构与各种非政府组织（NGO），逐步发展具有一定智能性质的宣传机器人，以更有效地干涉他国政局。早在 2008 年，美国国际开发署在古巴建立了虚假社交网络，试图向潜在的持有不同政见人士宣传具有强烈政治倾向的内容，以影响古巴舆论；2010 年以来，美国利用 Facebook、Twitter 等社交网络对埃及、突尼斯、利比亚、乌克兰、委内瑞拉、叙利亚、伊朗等诸多国家的意识形态进行大规模渗透，制造了一系列颜色革命，成功导致其政局动荡，甚至政权颠覆。智能宣传正成为一种可以决定国家命运的隐形利器。鉴于其强大的舆论攻击能力，欧美等国开始制定针对性的防御计划。



美国政府在奥巴马时期，认为“网络虚假不实误导性新闻”和“俄罗斯政府通过黑客袭击与信息操控”影响到美国总统选举，将会严重威胁美国的民主政治与市场经济制度，于2016年12月23日签署了《反宣传法案》，建立专门的反宣传中心 DK 他国的宣传渗透，同时帮助盟国获得更加真实的信息，保护其国民免受宣传洗脑，意图创造一种“能打赢思想 Z 的更全面、更积极的方式”。欧洲议会同时期也通过了《欧盟反击第三方宣传的 ZL 传播》建议案，号召欧盟各成员国采取措施反制第三方宣传。可见，欧美等西方国家在宣传领域“锐其矛”的同时，也在“固其盾”。

智能宣传将成为心理 Z 的新武器，多国正积极地筹划部署。智能化的宣传心理 Z 可以从精神层面扰乱、震慑和摧毁敌人，美国兰德公司提出了“新脑皮层 Z”的 ZZ 思想，旨在不破坏敌方组织的同时，通过信息的精准定向投放，限止和误导敌方领导者的意志力和理解力，设法影响和塑造敌方组织的行为。在该 ZZ 思想的影响下，美国 DARPA 针对新媒体的智能利用、攻击、检测和防御技术，启动了一系列面向信息心理 Z 的前瞻性研究项目，并发布了若干相应的军事 ZZ 指令（如表 1 所示）。



表 1 美国 DARPA 智能信息心理 Z 部分研究项目和条令

发布时间	名称	目标
2011.10	DARPA Narrative Networks	研究大脑如何对不同故事和言论主张做出反应，隐蔽的检测叙事对人类心理的影响及其关联神经生物学
2011.07	DARPA Social Media in Strategic Communication (SMISC)	研究利用社交媒体实现全自动或半自动的ZL沟通。
2013.01	军事信息支援 ZZ 条令 FM3-53	将意识领域的信息攻击作为一种军事活动，期望通过不易察觉的手段影响敌对政权、防止叛乱，遏制可能影响美国盟友和伙伴关系的冲突
2013.02	网络空间联合 ZZ 条令 JP3-12	提出可以对网络行为体，即网络空间的使用者“人”实施网络空间 ZZ
2015.02	DARPA The DARPA TWITTER BOT CHALLENGE	识别和清除在社交网络上进行非法信息传播来影响民众观点的社交机器人
2017.09	DARPA Active Social Engineering Defense (ASED)	自动确认、干扰和抵制通过机器人通信进行的钓鱼式攻击和社会工程学攻击



同样，俄罗斯国防部和英国国防部也提出在 Facebook 和 Twitter 等社交媒体上，通过“动态叙事”开展政治和军事宣传，对敌人进行“非致死性心理 Z”。研究显示，截至 2017 年已经有 28 个国家调用大量资源和人力来管理和操纵在线舆论，组建网络部队（如图 2 所示）。虽然各国网络部队的预算、行为和规模各不相同，但各国均已认识到，智能宣传是一种新型 Z 法，将大大拓展 ZZ 渠道、革新 ZZ 方式、提升 ZZ 效能，熟练掌握和运用智能化的舆论宣传、威慑宣传、诡诈宣传以及情感宣传武器，可达到不 Z 而屈人之兵的最高境界。

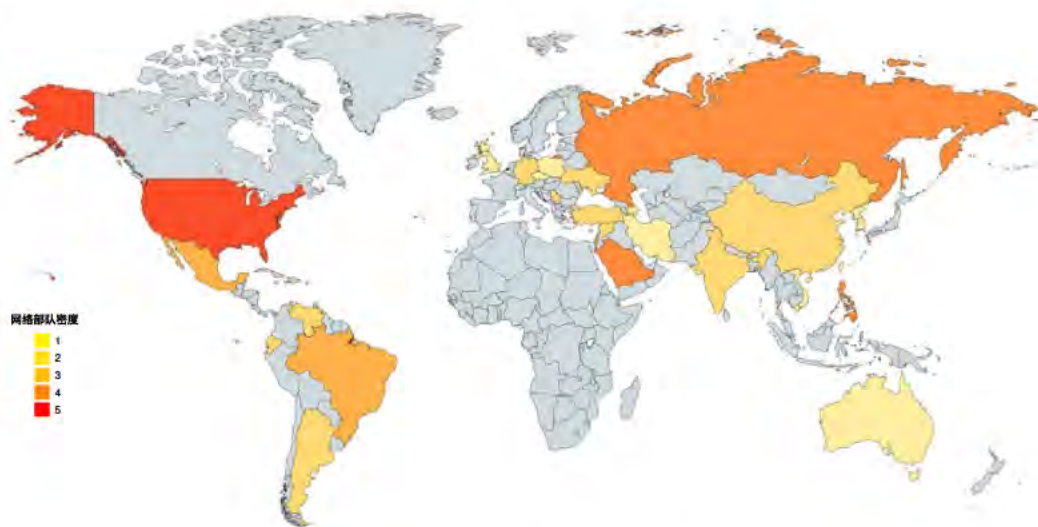


图 2 全球有组织的网络部队密度图

智能宣传将成为反华国家和势力对我舆论攻击的新工具，使国家安全面临新挑 Z。在国际舆论空间，反华势力已于各类社交媒体中组织起了规模庞大的机器人水军，包括 1989 群组、泛亚群组、独立反华机器人、反商业内容机器人、专业新闻和假新闻机器人等，主要针对历史、民族、宗



教、人权问题和舆情热点事件等，发布倾向性言论，营造舆论假象和虚假共识，混淆国际视听，抹黑我国的国际形象；高调宣扬“疆独”、“藏独”、“台独”和“港独”等分裂思想，为其分裂中国的行径制造舆情氛围。在国内舆论空间，以美国为首的西方集团在我方社交媒体平台上兴建机器人水军，窃取用户隐私、传播虚假消息，在政治、经济、文化、社会、生态等方面制造思想混乱、开展颠覆破坏活动，试图引发我国社会动荡变革。智能化的宣传工具将全方位、多领域、深层次地威胁我国国家安全。

智能宣传的强大舆论引导和动员能力，源自认知科学、大数据与人工智能技术的新发展。在认知科学层面，人们越来越注重信息、生理、心理、行为、环境的互动机制研究。相关研究表明，无意识的语义启动可以影响大脑活动，对比实验发现阈下信息刺激可使得政治观点极端的两类人政见趋于统一；网民情感能被信息改变，并通过网络进行大规模传播。近年来，人们更加注重从计算神经科学角度研究非程序的、适应性的、大脑风格的信息处理的本质和能力，这将成为智能宣传研究的重要科学基础。在大数据和人工智能技术层面，注重信息受众的海量背景数据和行为数据的收集与分析，通过数据挖掘技术支撑的心理画像模型与人工智能技术支撑的广告定向和行为分析算法，利用全方位的媒体渠道，构建规模庞大的实时宣传生态系统，对受众进行精准的信息分发和扩散，从而改变民众的心理和行为。今后，大数据、人工智能技术与计算神经科学将会深度融合，进一步提升智



能宣传的实 Z 效能。

本项目直面意识形态和国家安全领域尖锐复杂的斗争特别是“颜色革命”的现实危险，顺应网络化和智能化时代高科技发展和学科交叉融合趋势，构建智能科技支撑下的新型宣传技术体系。借助于智能宣传理论创新，创新宣传受众、内容、情境与 Z 法等要素智能分析的新型传播理论；突破一批具有精确、实时、可控和隐蔽等智能特征的共性关键技术，实现宣传工作的内容创新、手段创新和 Z 法创新；发展高度自主、实 Z 可用的智能宣传武器，针对敌对目标、亚稳定人群和普通大众的心理状态、思维意识和行为模式进行干预，并在若干重要或敏感场景、事件中开展示范应用，解决国家和军队在宣传领域的紧迫现实需求；运用人工智能在 ZL 欺骗与反欺骗中对对手精巧“设局”和巧妙“抽身”，并有机融合到联合 ZZ 体系中。

舆论是公众对于特定话题或事件的总体观点、态度和信念。舆论自从人类社会诞生之日起便在人类社会发展扮演着重要的作用，各国政府、企业和社会团体都非常重视舆论和舆论引导。舆论引导或舆论导向，是一种运用舆论去引导人们的观点、态度和信念的信息传播和宣传行为。上升到国家政治、经济和军事层面，舆论引导则演变为更具 DK 性和体现国家博弈的舆论 Z。舆论 Z 是 Z 争双方依据传播学原理，利用电视、广播、网络、报刊等大众媒体，有计划、有目的地向受众传递经过选择的信息，宣扬己方对特定事件的立场、观点和看法，阻断、瓦解和反击敌方的舆论攻势，从而影响



受众的情感和行为，引导社会舆论、影响民意归属，造成有利于己的舆论态势。舆论 Z 是一场没有硝烟的 Z 争，被称为“不 Z 而屈人之兵”的攻心 Z，是借助舆论从内部瓦解和分化敌方的 Z 争手段，通常具有 Z 线长、高隐蔽性、强破坏性、Z 争影响难以消除的特点。

国家间的舆论 Z 是国家意志在舆论上的自然延伸和集中体现，是营造民族认同和国家博弈的重要载体之一。自古以来，舆论 Z 就是重要的 Z 争手段之一，例如楚汉之争时的“四面楚歌”之计和二 Z 时美国对德国实施的舆论 Z。冷 Z 以来，以美国为首的西方国家开始将舆论 Z 作为意识形态输出和在全球范围内实现其国家意志的重要武器，舆论 Z 在苏联解体、东欧巨变、科索沃事件、中东“颜色革命”、克里米亚事件、英国脱欧事件、美国总统大选等影响世界格局的重大政治事件中都发挥着重要的甚至不可替代的 ZL 作用。近年来，美国依托其强大的经济实力，凭借其在信息领域的技术和平台优势，以 Google、Facebook、Twitter、YouTube 等互联网平台为武器，不断在全球范围内掀起舆论 Z，对不符合美国国家利益的国家和政府实施舆论引导，作为实现其国家意志和保障其国家利益的桥头堡和先头兵。舆论 Z 的威力和重要性逐渐被世界各国所重视，开始作为保障国家利益的重要 ZLZ 场。近年来，俄罗斯、英国、法国等安理会常任理事国在舆论 Z 方面开始展示实力，舆论 Z 美国一家独大的格局开始弱化，美国也从舆论 Z 的全面出击逐渐转向攻防兼顾。我国作为世界第二大经济体和最大的发展中国家，一直



是美国舆论 Z 的主要目标之一，例如美国等国家炮制的“中国威胁论”、打着“普适价值观”的旗号进行的文化侵略、借助互联网平台输出的“历史虚无主义”和“佛系丧文化”。为此，我国亟需提升自身的舆论 Z 能力，服务于我国舆论 Z 从积极防御逐渐转向 ZL 进攻。

舆论 Z 的主要手段是舆论宣传。广义而言，舆论宣传是运用各种媒介进行观点传播，以影响人们态度、引导人们行为的一种社会性传播活动。舆论宣传本质上是一种信息传播行为，其目的是使受众知悉、认同、传播某种特定的观点。

“知悉、认同、传播”从宣传目的的角度大体界定了宣传的三个阶段。第一阶段，宣传主要借助于各种媒介让大众知悉某种信息或观点，主要致力于汇聚大众的注意力和获得大众关注，更多地属于“广而告之”的范畴，主要手段和平台是大众媒体，典型的案例是新闻报道——选择具有重要性、时新性、接近性、显著性和趣味性的信息进行宣传从而实现获得关注的宣传目的。第二阶段，宣传则致力于说服受众认同和接受宣传的观点，这一阶段的宣传利用受众的社会关系、心理、文化、教育、地域、信仰等因素实现宣传目的，典型的案例是舆论引导。第三个阶段，宣传致力于在受众知悉和认同特定观点的基础上使其自发地传播该观点从而形成社会舆论，这一阶段的舆论宣传属于社会化的智能宣传，是舆论 Z 的最高境界。

舆论 Z 关系国家政治经济安全和社会稳定，已经从国家之间竞争的软力量发展成为影响国家 ZL 和发展环境的硬力



量。习近平总书记在党的新闻舆论工作座谈会上的讲话上指出的“团结稳定鼓劲、正面宣传为主”、“建设有国际影响力的外宣媒体旗舰”和最近网信工作座谈会上的讲话指出的“发挥网络引导舆论、反映民意的作用”、“网上网下要形成同心圆”对于强化正面引导、提升新闻舆论工作的吸引力、凝聚力、传播力和向心力指明了重点。实现习近平总书记提出的新闻舆论工作目标的重要途径是建设智能宣传系统，为维护国家繁荣稳定的良好舆论氛围营造和体现国家意志的对外正面宣传提供有力保障和坚强支撑。总体而言，我国目前的新闻舆论工作还主要集中在新闻宣传和舆论引导两个阶段，更多地体现在服务于保障国家经济发展和维护社会稳定方面，其技术水平和即 Z 力还不能完全甚至完全不能支撑大规模的舆论 Z，亟需装备符合我国新时代发展需求的、支撑舆论 Z 的智能宣传系统技术。

以计算机和互联网为代表的信息技术革命以前所未有的深度、广度和尺度革新了舆论 Z 的范畴和影响力，呈现出跨空间、跨媒体、跨网络的特点。(1) 跨空间：计算机和互联网极大地推动了人类社会的信息化水平，实现了人类社会、物理世界、网络空间的三元空间深度融合，大大缩短了人与人之间、信息和信息之间、人和信息之间的距离，为宣传和舆论提供了便捷的通道和广阔的空间。(2) 跨媒体：文字、声音、图片、视频、表情等多模态数据提供了跨媒体的宣传媒介，以丰富的手段大大提升了信息时代宣传的冲击力。(3) 跨领域：舆论不再局限于单个事件或单一领域，而是在政治、



经济、军事、文化等多个领域综合发酵形成“舆论合力”或“舆论场”，其影响也更加深远和难以消除。上述特点使得信息时代的舆论 Z 在时间、空间、传播媒介、影响范围等多个维度上呈现出不同于传统舆论 Z 的挑 Z。

伴随着信息革命的飞速发展，互联网、万维网、社交网络、社交媒体、移动互联网及各类“互联网+”服务应用，正在全面改变人们的生产生活方式，深刻影响着人类社会历史发展进程。网络空间是信息传播的新渠道、社会治理的新平台、国家主权的新疆域和舆论宣传的主 Z 场。借助网络空间的舆论宣传和渗透危害政治安全、威胁经济安全、侵蚀文化安全、破坏社会安定，网络空间的舆论 Z 方兴未艾，一场争夺网络空间舆论主动权的竞争日趋激烈，现在舆论 Z 的跨空间、跨媒体、跨空间的特点使得传统的舆论宣传手段不再适用，亟需新的基于人工智能和大数据的智能宣传系统技术。

智能宣传系统技术需要从三个层面形成突破：（1）应用范畴层面，需要从国内的非 DK 环境到走向全球的 DK 环境。传统的舆情系统更多地侧重于舆情监测，服务于国家公共安全和社稷稳定，智能宣传系统则更多地侧重于舆论引导，服务于国家意志输出和国家博弈的制信息权；（2）技术手段层面，需要态势感知走向精准导控。传统的舆情系统更多地是采用信息技术实现对舆情信息的自动获取和分析处理，属于态势感知层面，而智能宣传系统则致力于采用大数据与人工智能技术实现集态势研判、目标发现、突破点选择、高聚能精准打击和效果评估为一体的“闭环”系统；（3）呈现形式



层面，从软件系统走向复杂智能体。无论是数据感知获取、信息传播预测、群体画像和身份标识等，传统的舆情系统在呈现形式上都集中在技术和功能层面。智能宣传系统则是一个体系化的、浸入式的舆情引导系统，以一个复杂智能体的形式存在。

智能宣传系统技术的研究和部署是一个复杂的社会工程，涉及多个学科，包括计算机科学、人工智能、心理学、计算社会学、传播学、系统科学、博弈论和社会工程学。计算机科学将从计算层面解决数据的感知获取、存储管理和分析处理；人工智能解决系统的智能化实现，包括机器学习技术、深度学习技术和强化学习技术；心理学则关注受众在接受舆论宣传时的心理活动规律，包括从众心理、一致性心理、逆反心理、标新立异等；计算社会学则重点从社会层面探索文化、语言、地域、种族等社会因素对舆论宣传的影响机制；传播学探索舆论宣传的宏观传播态势和微观传播机理，支持传播模式的设计和传播接入的时机；系统科学则从复杂系统的角度探索通过多智能体平行社会对效果进行沙盘推演和效能评估；博弈论从机制设计的角度服务于 DK 环境下舆论 Z 的实施；社会工程学从心理脆弱性的角度探索基于社会工程学的舆论传播与引导。

智能宣传系统技术可作为一种 ZL 武器，单独作为 Z 争的发起点和终结点，和服务于情报 Z 的舆情系统和情报系统具有本质区别。智能宣传系统是一个基于跨空间跨媒体跨领域大数据分析技术的人在环路、虚实结合、自适应的智能化



体系化 ZZ 系统。智能宣传系统包括 ZL 态势感知、目标突破口发现、高聚能精准导控和效果评估四个环节。(1) 态势感知层面，研究一体化的智能宣传技术态势感知体系，包括综合线索感知、ZL 意图研判、舆论传播溯源、综合态势表达与可视化等；(2) 目标发现层面，采用智能宣传技术进行目标群体发现、用户画像与行为建模、攻击路径识别、心理脆弱性分析等；(3) 精准导控层面，构建具有前瞻性和颠覆性的宣传引导机制，研究信息智能生成、舆论传播与 ZL 诱导、舆论场营造、舆论传播与极化、舆论反制与渗透等；(4) 效果评估层面，研究舆论决策体系构建、复杂开放环境博弈、镜像 Z 场环境演练、舆论净效能评估等。智能宣传系统技术着眼于为“天朗气清、生态良好”的舆论环境营造提供能力支撑，更为舆论 Z 提供 ZL 支持，为国家实施舆论 Z 提供 ZL 武器和 Z 术手段。

综上所述，面向舆论 Z 的智能宣传系统技术的研制和部署势在必行、正当其时，符合我国新时代实现大国和平崛起的时代需求，是提升我国综合国力和 Z 备能力的重要手段，是我国谋求新时代更大发展空间和信息 ZL 纵深的必由之路。



目标：

针对个性化推荐系统 GJ 手段隐蔽性低、GJ 行为易被推荐系统检测、GJ 策略未充分考虑推荐系统工作机制、GJ 效果单一不可持续和社交推荐场景 GJ 手段匮乏等难题，突破用户信息消费模式分析、推荐系统智能感知、推荐系统脆弱性分析、推荐系统隐蔽 CT 和社交推荐隐蔽传播通道构建等关键技术，达到用户行为模式识别效果准、推荐系统工作机制认知全、推荐系统脆弱性分析可靠性高、推荐系统 CT 通道隐蔽性强和社交传播通道覆盖范围广等指标，为形成用户信息消费行为模式识别、推荐系统工作机制白盒化、推荐系统脆弱性评估、个性化 CT 内容生成与送达以及社交机器人网络构建与传播调度等能力。

背景：

推荐系统的出现，便利了用户信息获取的过程，也为内容生产商新增了传播渠道。进一步地，内容生产商可以尝试对推荐系统进行 GJ，实现内容的定向推送。基于个性化推荐系统构建隐蔽传播通道，需要对现有的推荐系统进行 GJ，使其做出于我方有利的推荐策略。



现状：

由于推荐系统自身具有的用户参与性和信息开放性，导致系统很容易受到外界的 GJ，出于商业竞争等目的，一些恶意用户人为地将大量伪造的用户概貌（User Profiles）植入系统中，使得他们成为大量用户的最近邻，从而使用户对目标物品的偏好发生偏移，导致系统产生对他们有利的推荐。这样的 GJ 方式通常称之为用户概貌植入 GJ，因为 GJ 用户特别像我们生活中所说的“托”，因此又叫做托 GJ，这些虚假的用户概貌称之为 GJ 用户概貌（Attack Profiles）。当推荐系统受到外界 GJ 用户的植入后，其推荐结果会发生很大的变化。常见的 GJ 类型有两种：推 GJ（Push Attack）和核 GJ（Nuke Attack），前者会导致推荐系统以更高地频率推荐目标物品，而后者则是为了降低目标物品的推荐频率。

相关国外项目：

近年来美军投入数百亿美元，通过国防部、DARPA 等机构，部署了一系列从舆论数据体系、关键技术支撑平台到舆论传播和操纵的实战项目。据“棱镜”计划披露，美军通过直接进入微软、谷歌、苹果等互联网公司服务器，实现对全球媒体和舆论数据的监控；部署 XDATA、Insight 等大数据快速处理和深度分析技术研究，支撑大规模舆论作战应用；通过“社交媒体战略传播”、“网络魔术师”及“诚挚之声行



动”等舆论操纵项目，对竞争对手进行否定、破坏、贬低和欺骗，帮助美军实施对抗 GJ 和舆论宣传战。同时美军正在加强包括“深度伪造”、“真相终结”、“社交媒体机器人”等下一代舆论监测和操纵技术研究，部分成果已形成实战能力。在这些项目部署中，**操纵内容的定向送达**都是一个不可或缺的部分。

牛津大学网络研究院发起了“计算宣传项目” (The Computational Propaganda Project)。自 2012 年以来，该团队一直在调查算法、自动化和计算宣传在公共生活中的应用。政治机器人在主要社交网络应用上操纵公众舆论。这个项目使一个新的社会和信息科学家团队能够调查自动脚本（通常称为 bot）对社交媒体的影响。该团队研究了机器人脚本和制造这种机器人的人，然后与计算机科学家合作，改进捕捉和阻止这种机器人的方法。经验表明，政治机器人最有可能出现在国际危机期间，通常是为了促进陷入困境的政府的利益。政治人物利用机器人操控对话，遣散反对派，并在 Twitter、Facebook 和新浪微博等热门网站上制造虚假支持。该项目会与受委托制造机器人的政治顾问和计算机专家进行国际实地调查，建立一个涉及机器人的政治事件的原始数据库，并利用这些知识来制作更好的工具，以便在政治机器人出现时进行检测。该项目定位于实时社会和信息科学，会定期向记者、工业界和外交政策专家传播研究成果。该项目希望通过开发一个政治机器人检测和编译原始数据集的



专家网络，帮助研究人员更好地了解机器人如何操纵社会网络，同时促进社会科学、计算机科学和工业界关于问题规模和可能解决方案的讨论。

该项目组于 2017 年发表了第 12 号研究报告：《水军、喷子和麻烦制造者：全球有组织的社交媒体操纵盘点》。该报告比较了 28 个国家的网络宣传活动，清点了网络宣传的消息种类、价值 (valence) 和通信策略，并根据组织形式对其项目预算和人员配置方面的能力进行了评估。报告显示，最早关于“有组织的社交媒体操纵”的报告出现在 2010 年，到 2017 年已经有 28 个国家的情况有迹可循。隶属于政府、军队或政党团队的网络部队致力于操纵社交媒体舆论，作为一个普遍的全球性现象，许多国家调用大量资源和人力来管理和操纵在线舆论，有时针对国内受众，有时针对外国公众。几乎每个威权国家都会通过社交媒体有组织地针对本国受众进行宣传，但针对外国公众的社交媒体宣传却不常见。然而并非只有威权国家才这样做，最早关于政府参与舆论推动的报道就涉及民主国家，并且政治传播技术的创新都来自政党。样本中几乎所有的民主国家都曾组织过针对外国公众的社交媒体宣传战，而针对本国选民的宣传都是由政党支持的。在社交媒体操纵背后的网络部队方面，报告指出：网络部队通常由不同的成员构成，它们使用各种策略、工具和技术进行社交媒体操纵。随着时间的推移，网络部队的组织模式已



经从涉及社交媒体舆论操纵实验的军事单位覆盖到与政府签订合同开展社交媒体宣传的战略传播公司。而且，网络部队在将来无疑将持续存在，不断演化，成为全球现象。

研究内容：

1.1. 推荐系统智能感知

现有的推荐系统 GJ 方法基本上都是围绕推荐系统的协同过滤策略展开的。实际上，推荐系统是一个复杂系统，从内容选择，到用户定位，再到信息与用户的匹配，多项机制共同作用完成最终的推荐过程。本关键技术定位于推荐系统工作机制白盒化，旨在揭示推荐系统中关键环节的运行机制，为 GJ 方法的设计提供运行机制基础。

1.2. 推荐系统脆弱性分析

当前推荐系统最常用的共同访问推荐技术、协同过滤技术以及最新的深度神经网络和图神经网络技术等，都存在着对抗 GJ 下的脆弱性问题。充分研究这些方法的脆弱性，一方面可以在信息作战中，针对对方推荐系统的脆弱性进行 GJ，实现敌为我用；另一方面，可以提升我方网络空间系统的鲁棒性和信息安全。本关键技术定位于探寻现有推荐系统的脆弱性，为 GJ 手段的设计提供 GJ 对象。



1.3. 推荐系统隐蔽 CT

在推荐系统智能感知和推荐系统脆弱性分析的基础上，进一步研究推荐系统隐蔽 CT 通道的构建，实现消息的隐蔽送达，且不被推荐系统和用户察觉。具体来说，推荐系统主要包括物料选择、用户定位和推荐机制三个方面。本关键技术分别针对这三个关键环节制定 CT 技术，最终实现对推荐系统的隐蔽 CT，完成消息的定向送达。

1.4. 社交推荐隐蔽传播通道构建

社交网络的角色已逐渐由网络社交向社交媒体转变。在社交平台上，用户可以选择感兴趣的用户进行关注，并后续从关注用户处获取信息。社交网络的存在，大大加速了信息的传播和送达的进程。本关键技术定位于社交推荐系统中隐蔽传播通道的构建，通过培养和模拟社交账号，实现对社交网络的覆盖和信息的定向送达。本关键技术包含如下四个子课题。



天玑数据采集中心介绍

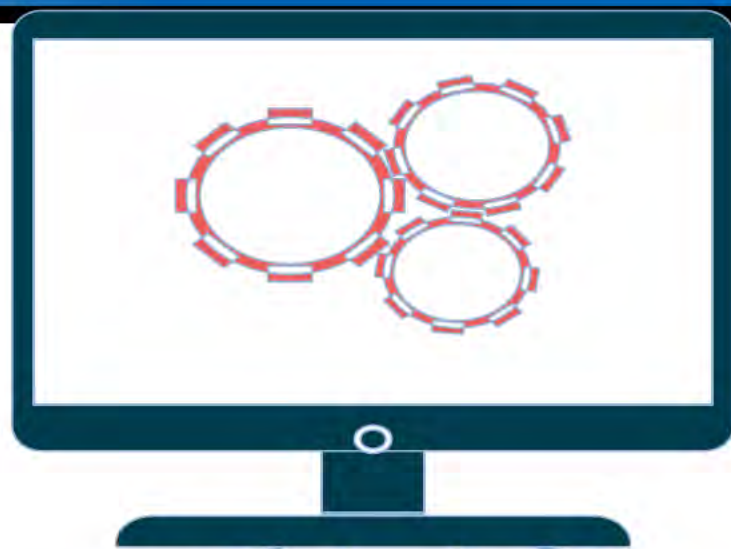


目录

CONTENTS

- 1 系 统 简 介
- 2 技 术 架 构
- 3 核 心 技 术
- 4 产 品 及 服 务

中科天玑数据采集中心致力于建设整套的数据采集方案，基于分布式高速采集技术，可一站式解决数据定制、采集、分析、挖掘、推送的问题，满足客户对即时数据、采集范围之外的历史数据及定制数据的采集需要，信源全面覆盖全网各类国内外资讯媒体、自媒体、社交网络信息，每天从千万级的信息中采集、聚合分析网络信息，实现对领域热点、重大事件、敏感事件以及用户关注事件数据自动发现，广泛服务于公检法体系、国家部委、网信、国安、军工、企事业单位等，满足客户需要。



自主研发技术

是以完全自主研发的采集技术引擎为核心，快速获取多通道规范化网络舆情数据

可拓展架构

设计可扩展的系统架构，帮助用户实现实时监测及追踪，自动采集数据，并对采集结果进行规范化处理与分析

开放接口

建立多通道统一数据系统，定义通用的数据结构以及开放的数据交换接口，实现不同舆情信息渠道的融合

一站式平台

实现数据采集、处理、推送、分析一体化，满足客户对即时数据、采集范围之外的历史数据及定制数据的需求



采集中心简介



采集技术领先

分布式云采集架构，模块化设计，按需动态扩展，功能强大、高效稳定，百级采集服务器，7*24小时运行任务。系统支持多种采集模式，可灵活配置，多线程并发采集提高采集效率，支持JS渲染、验证码解析、动态IP代理等，有强大的反爬能力。

采集数据量大

系统已积累了近4年的采集数据，系统**每日新增数据量6000万-8000万**，其中微博为千万级，新闻、微信、论坛等其它通道百万级。

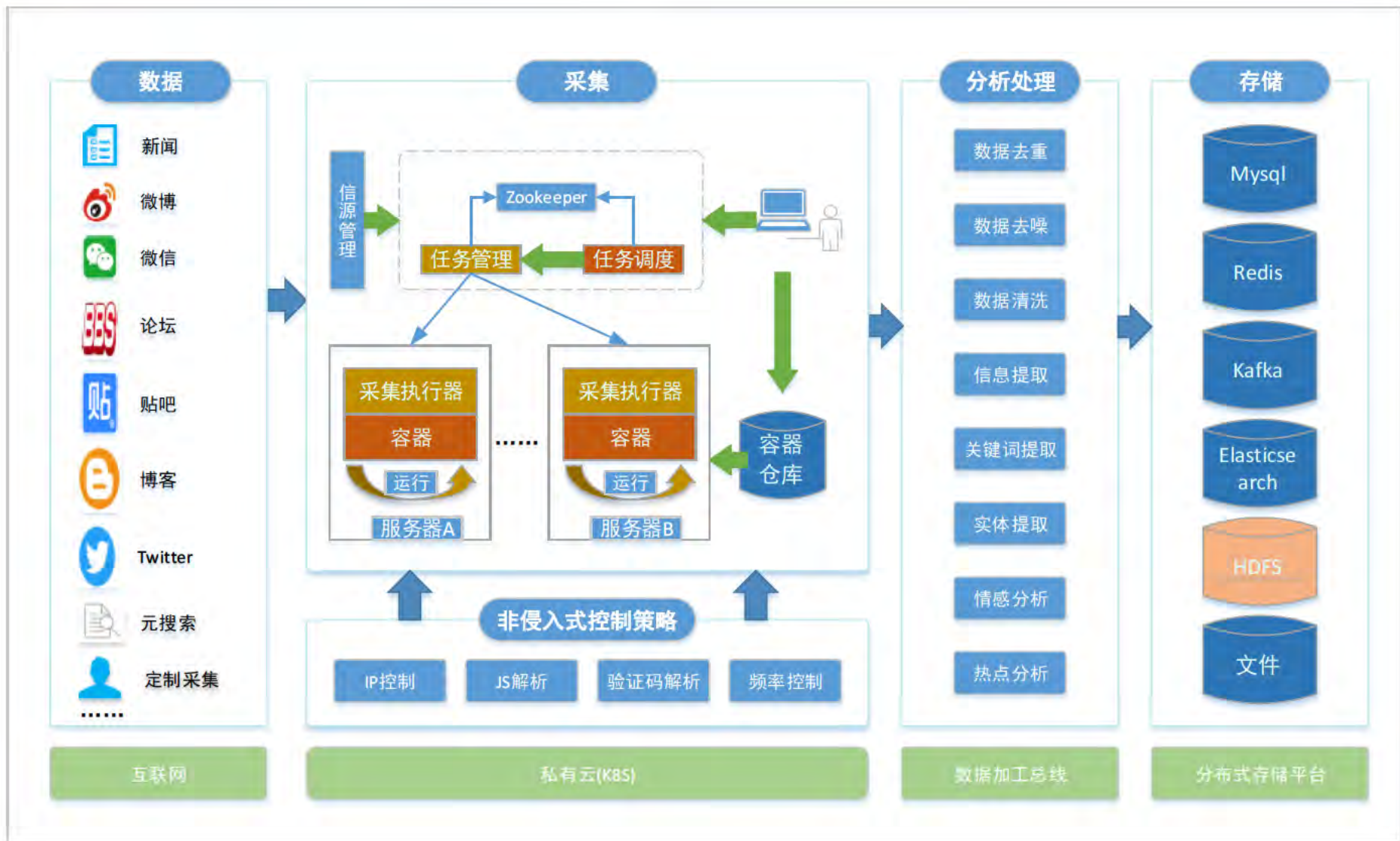


采集覆盖面广

系统采集信源覆盖面广，覆盖了新闻、APP新闻、境外新闻、境外社交、微信、微博、论坛、贴吧、博客、电子报、问答等10多种媒体类型，以及百度搜索、360搜索、必应搜索、搜狗搜索等主流搜索引擎。其中**新闻监测的版块数达到十万级，境外新闻近万**，覆盖了所有主流新闻媒体和财经媒体；论坛监测网站数达到万级；微信监测公众号规模达到百万级；微博采集为6000万+；搜索引擎覆盖了百度搜索、360搜索、必应搜索、搜狗搜索等主流搜索引擎，金融、财经、经济等多种类型，万级以上的搜索关键词。

采集实时性强

系统支持分布式多节点高并发采集，支持搭配多种分层、分级策略抓取，更新时效性强，对于不同类别的数据，系统采集延迟在1-30分钟以内，并支持用户需要按需定制。



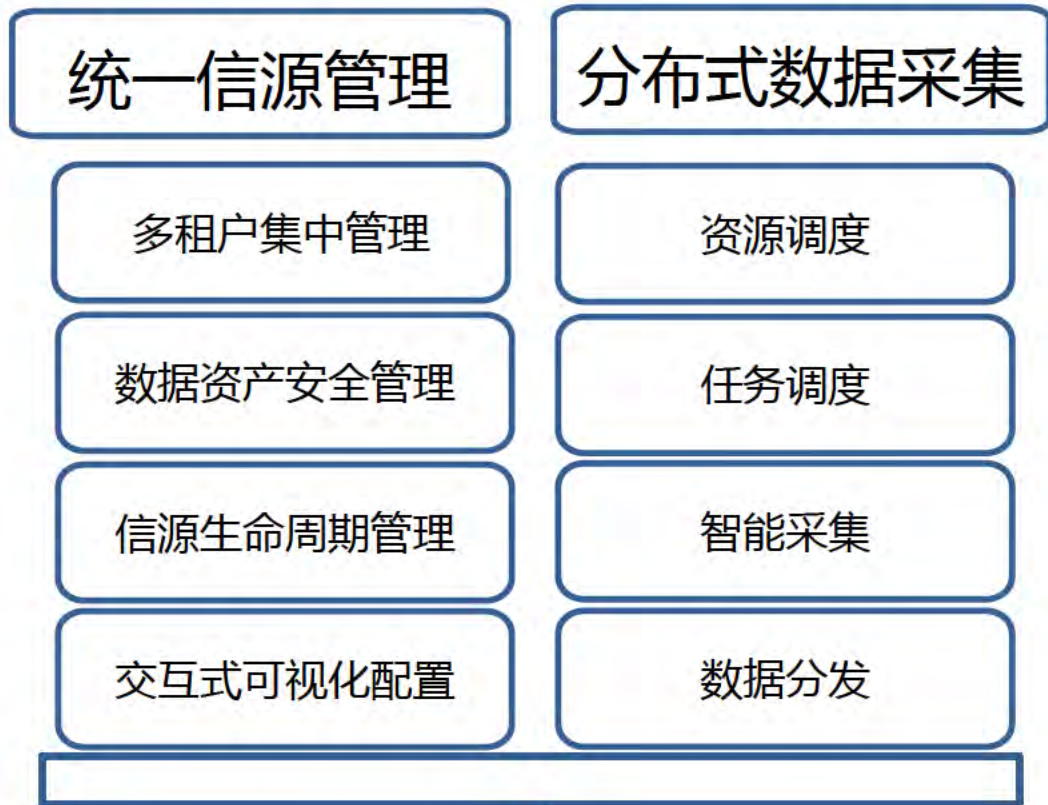
- 云采集架构
- 动态IP代理防封堵
- 验证码自动识别
- JS动态解析
- 文本语义分析
- 日志分析统计报警

天玑数据采集中心

天玑数据采集中心

实时采集**海量**互联网信息和移动APP数据，抽取关键信息，实现多通道跨平台的大数据内容汇聚。同时提供对外数据访问接口，方便其他系统进行集成。

核心技术



天玑数据采集中心



统一信源管理——核心设计理念

清晰的信源层级划分,统一管理信源

- 对网络信息来源进行三级划分：信源、信簇、传媒

信源

- 信息来源，信源是指最小粒度的信源，例如新闻频道、论坛版块、微博博主等

信簇

- 信簇为信源集合。一组具有某种相似特性的信源构成一个信簇。例如一个新闻网站或者一群微博博主。

传媒

- 传媒全称为传播媒介。传媒是信源的一种属性，表示该信源传递信息的手段，例如新闻、论坛、博客、微博、微信、APP等与传播技术有关的媒体。

自动配置,快速生成高质量模板

- 提供统一的采集、抽取模板配置平台，便于自动生成配置，人工审核、修正配置

快速配置

- 通过信源自动拆分、模板自动配置模块，结合可视化标注工具，实现快速的模板配置

协同抽取

- 多种抽取算法协同合作，结合组合抽取策略，实现高质量的结构化数据提取

多阶段验证,夯实采集抽取验证结果

- 信源验证按照配置活动的各个阶段，分为入口验证，正则验证，抽取验证，在线模拟验证

入口验证

正则验证

在线模拟验证

抽取验证

结合业务系统数据，反馈信源状态

- 提供反馈机制，汇总线上业务系统数据
- 结合主动检测和被动检测，提升信源失效检测的时效性

信源信息统计功能的接口

- 线上的系统可以基于此接口推送信源采集量、信簇采集量、信源已采集的文章最新发布时间、采集时间

失效检测反馈的接口

- 线上的系统可以基于此接口推送潜在失效信源及失效原因

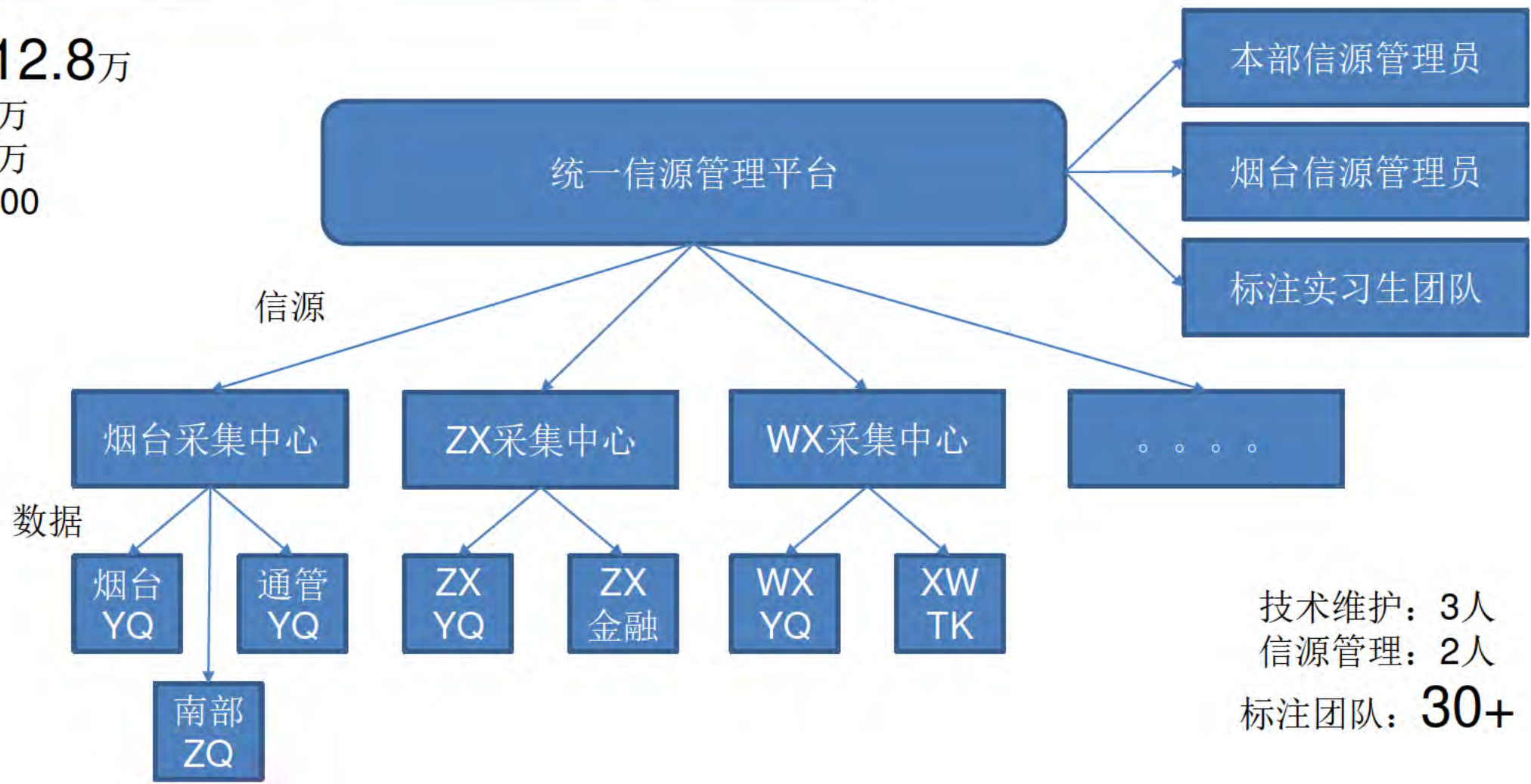
天玑数据采集中心





统一信源管理——多租户

整体信源 **112.8万**
新闻：24.5万
论坛：88.2万
电子报：1000



目前共支撑
65
个系统

技术维护：3人
信源管理：2人
标注团队：**30+**

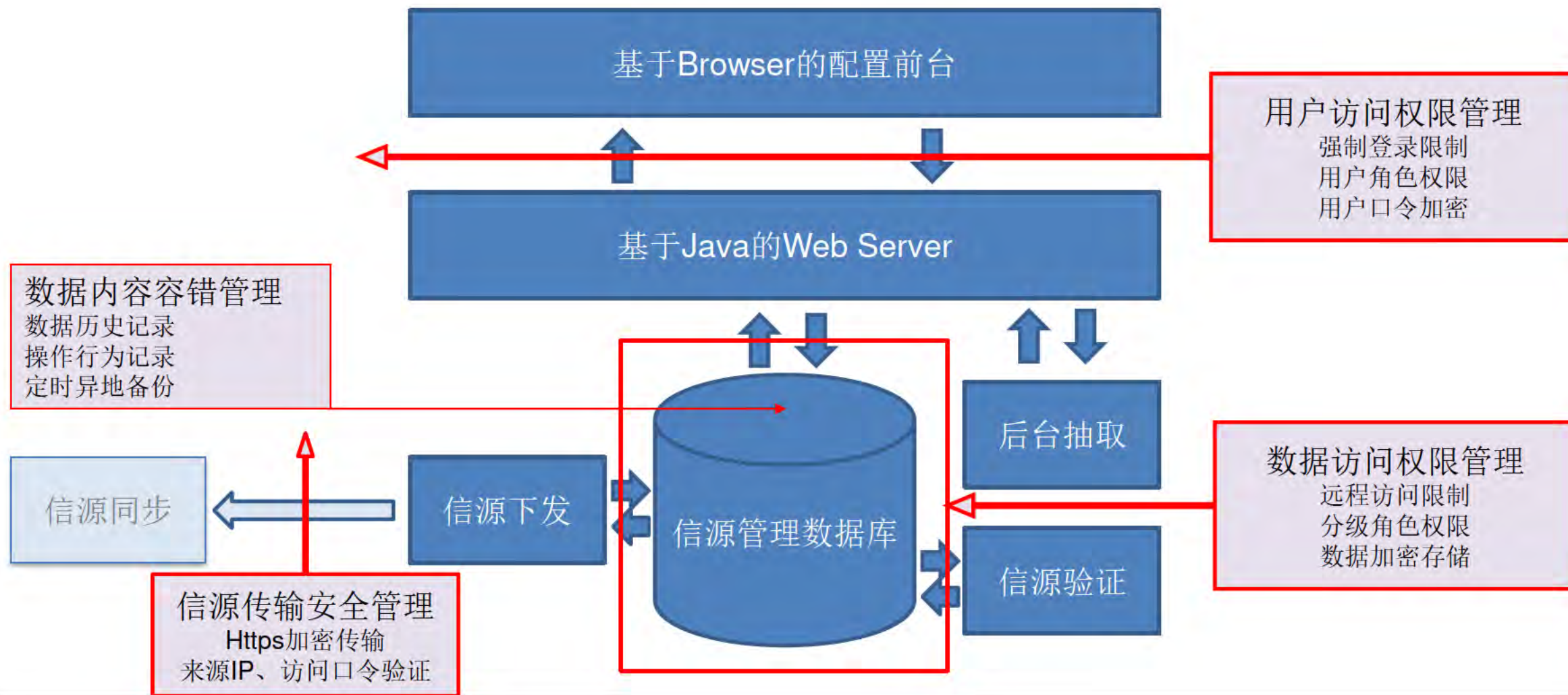
天玑数据中心



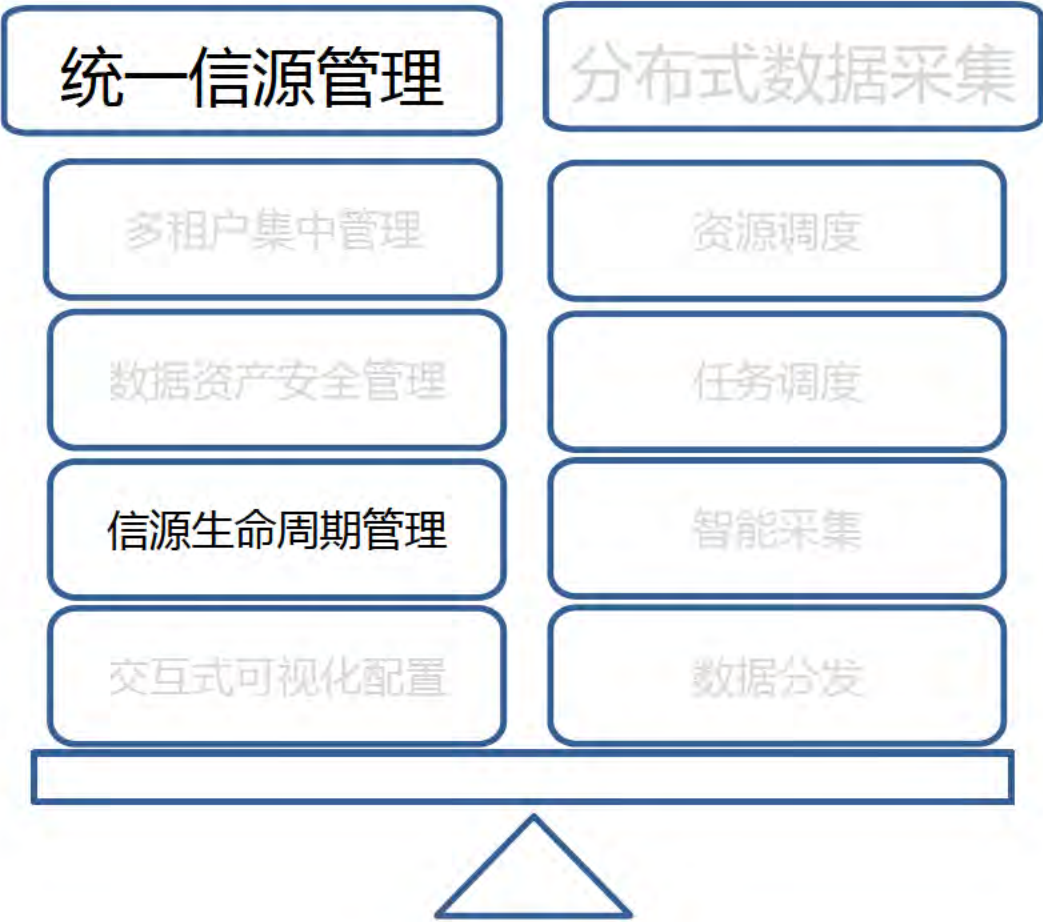


Colaxy 科天翼

统一信源管理——数据安全

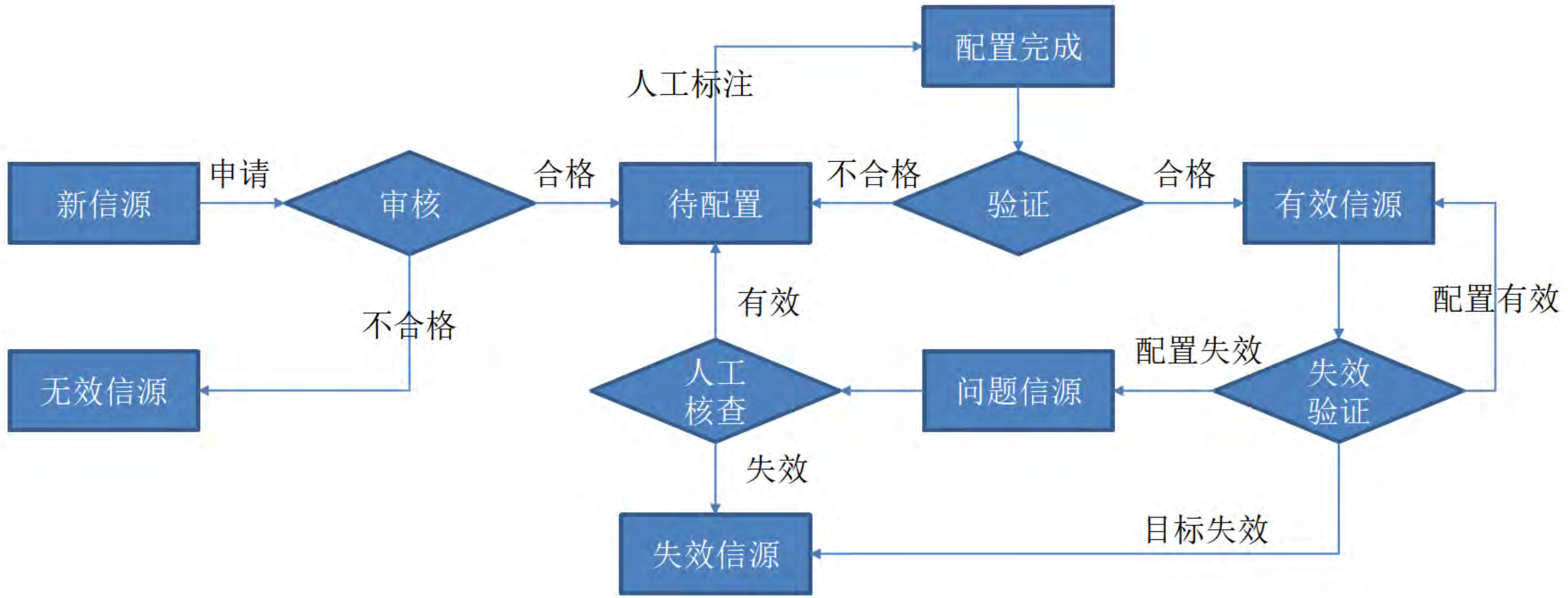


天玑数据采集中心





统一信源管理——全生命周期



天玑数据采集中心





统一信源管理——可视化配置

操作简单:

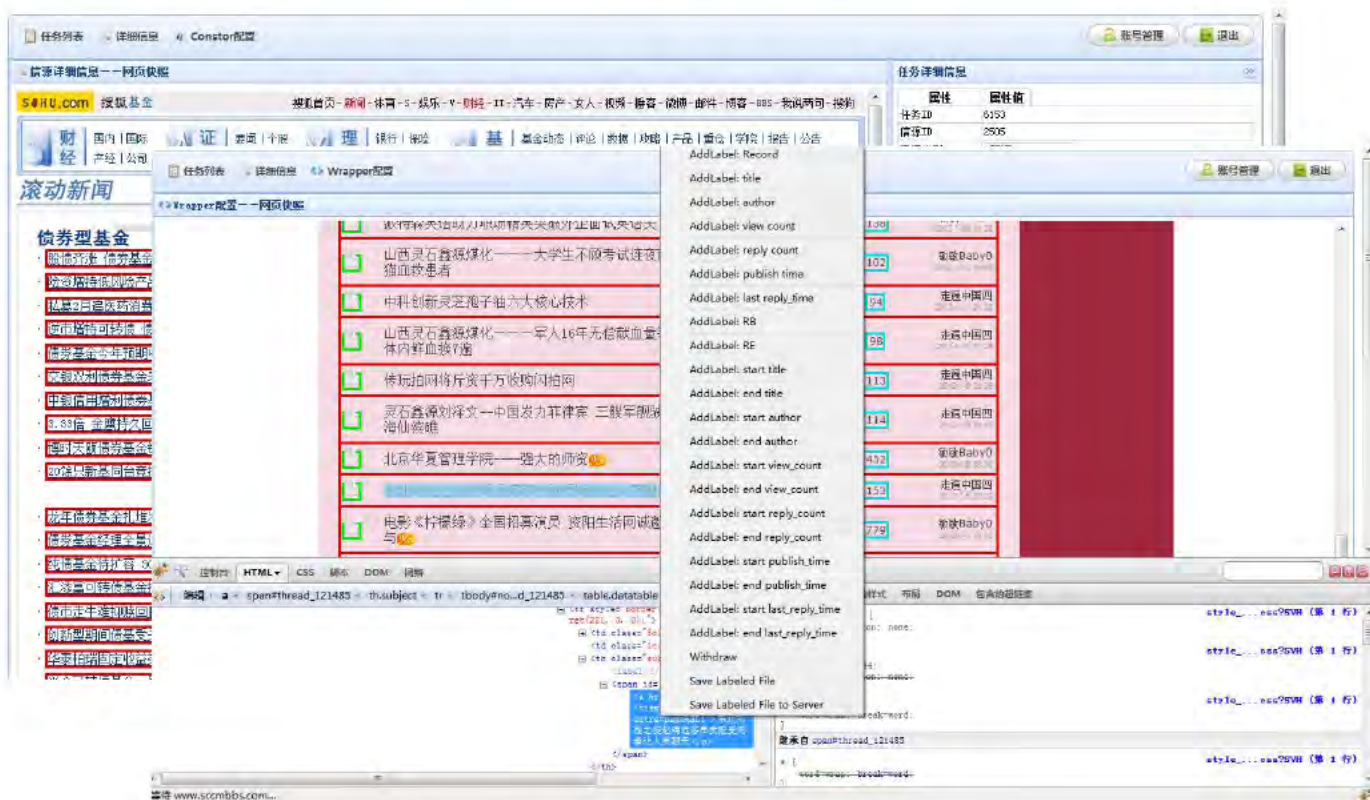
可视化: 浏览器可视化界面

交互式: 鼠标点击即可完成

安装简单:

B/S结构: 随时随地在线完成

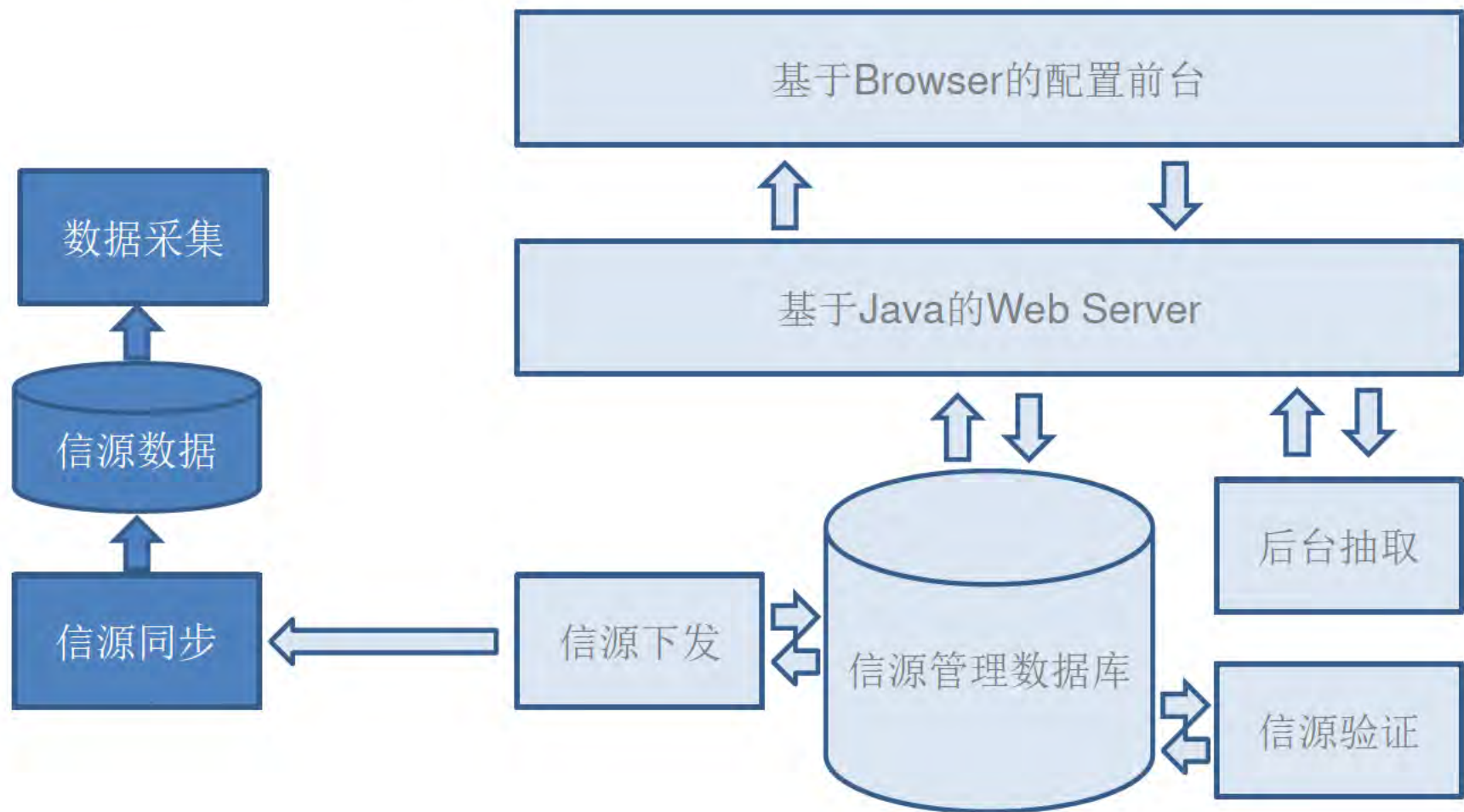
插件依赖: 拖拽式一键安装



天玑数据采集中心

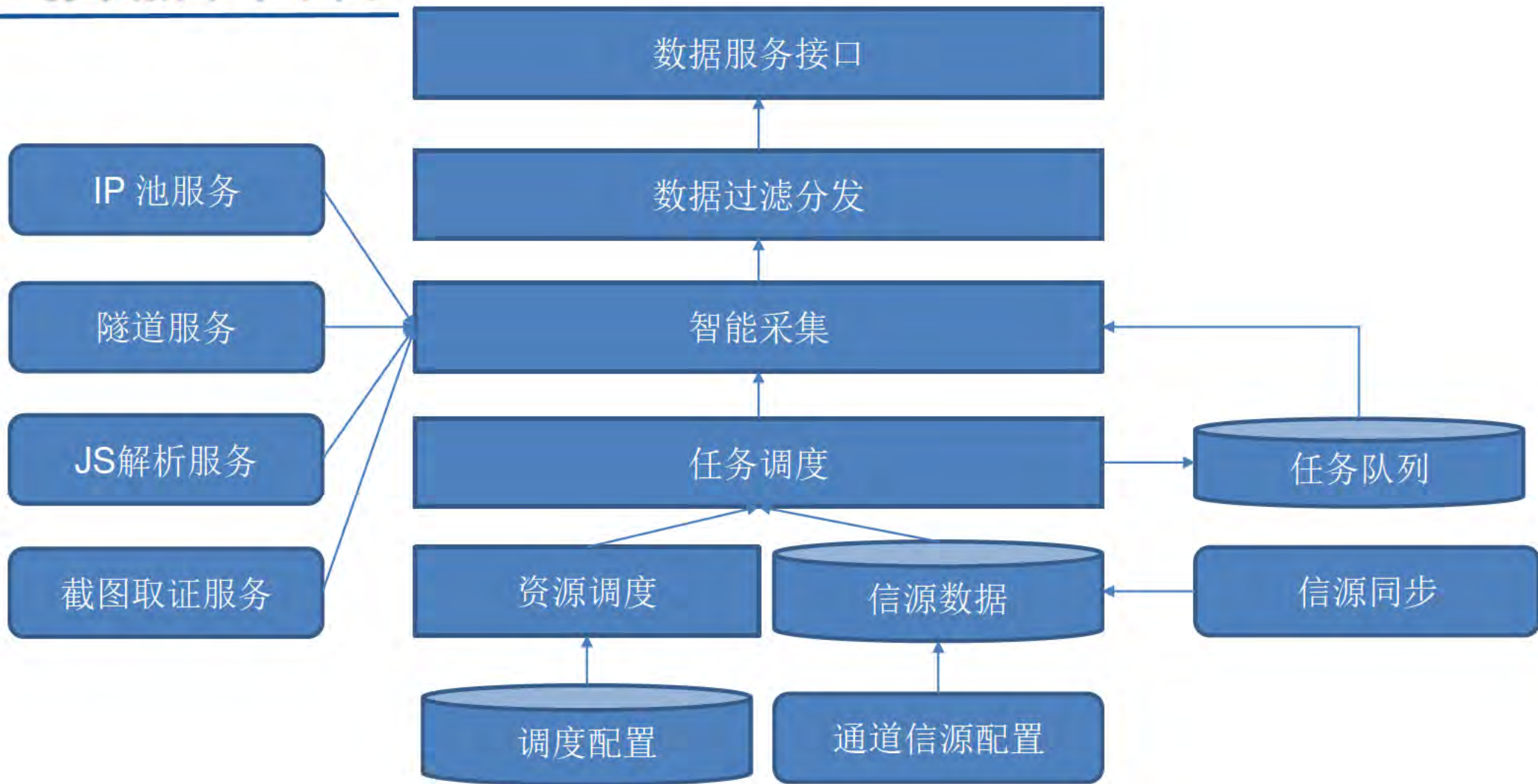


统一信源管理——整体架构





分布式数据采集



天玑数据采集中心



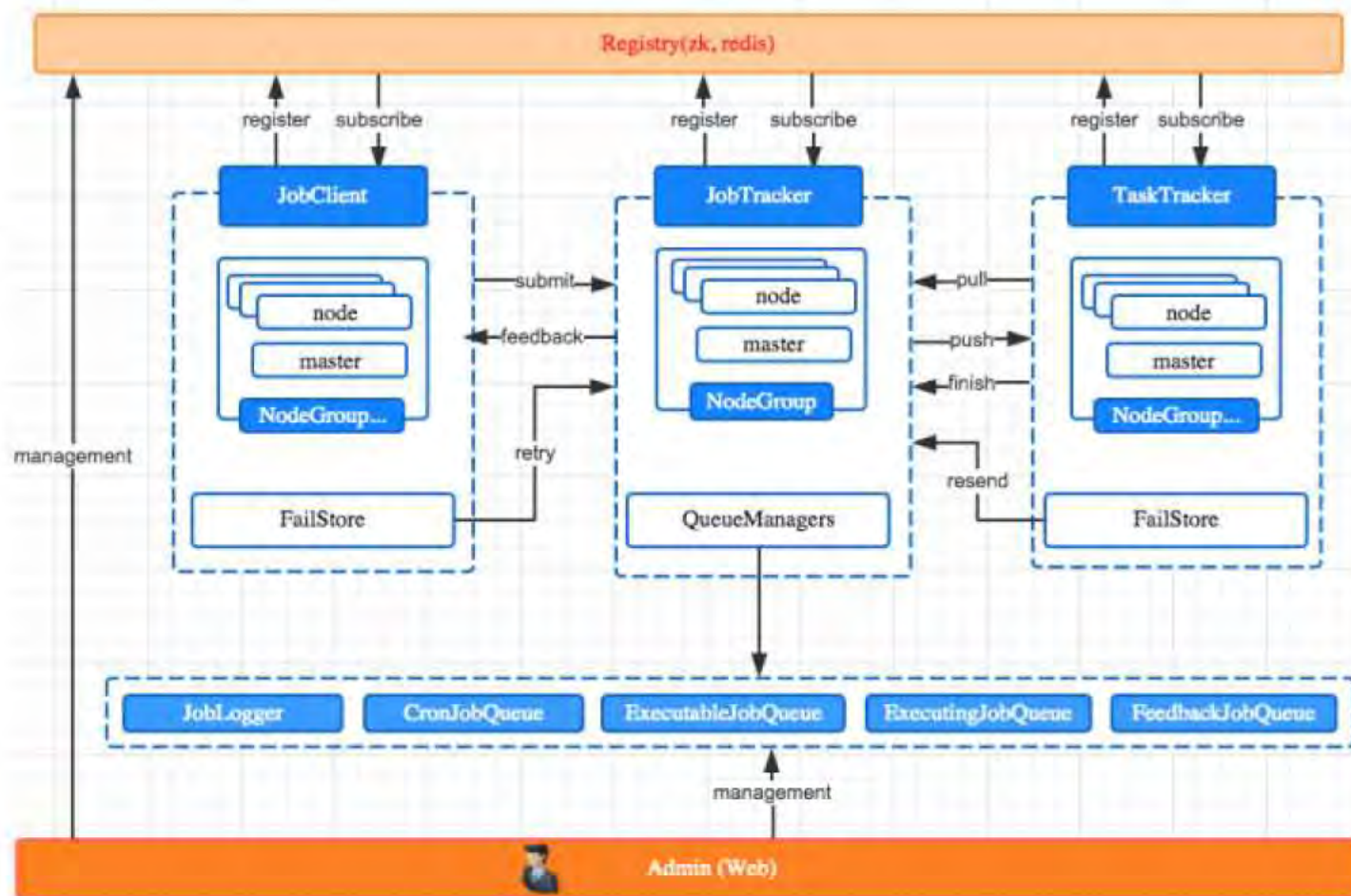
分布式数据采集——资源调度

资源调度：

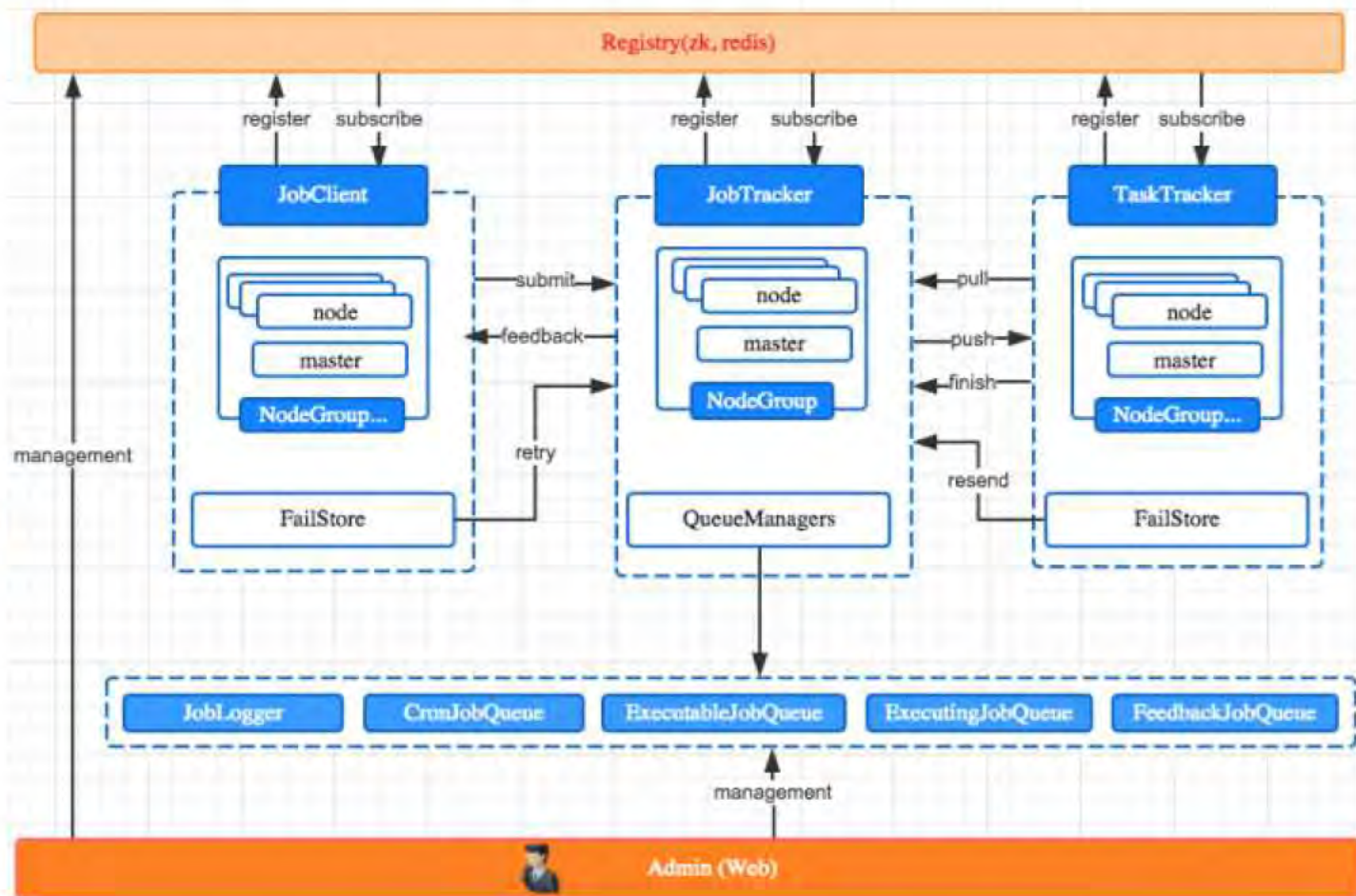
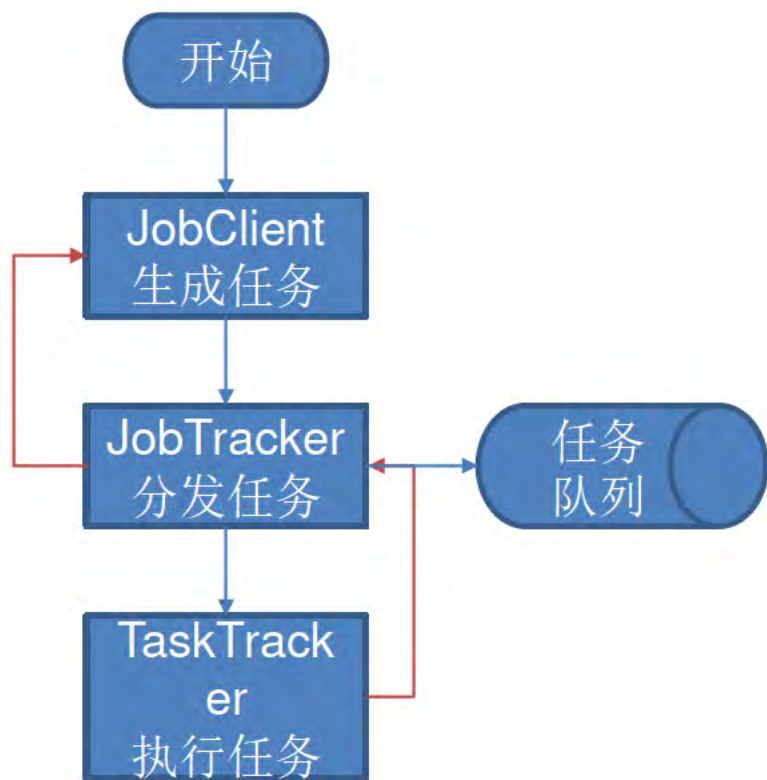
管理、调度、监测可用运算资源；
以采集器为单位，对采集资源进行调度、监测；
实现负载均衡、高可用性

支持：

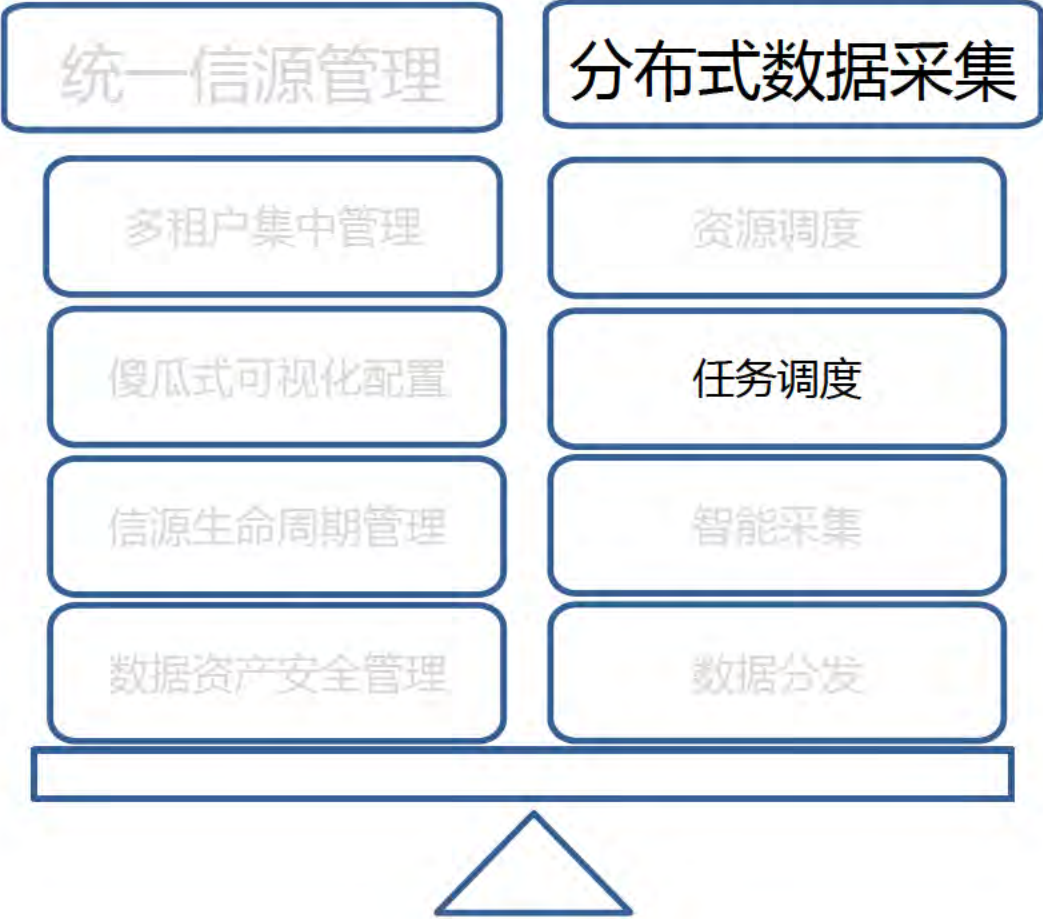
1. 持续任务
2. 定时任务
3. Cron任务



分布式数据采集——资源调度

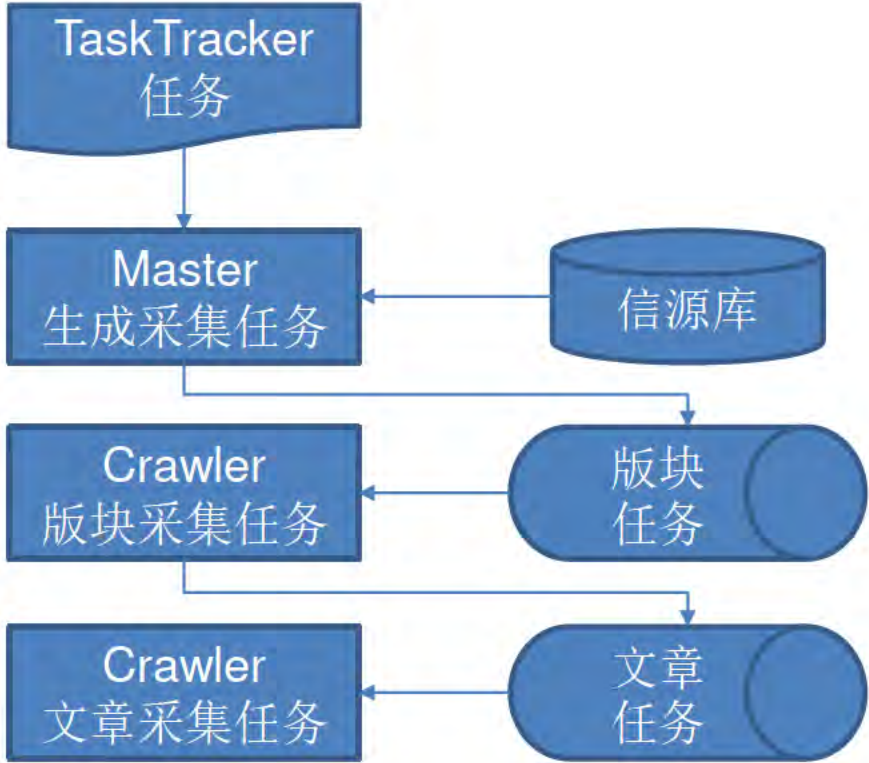


天玑数据采集中心



分布式数据采集——任务调度

任务调度：
管理、调度、监测采集任务；
以信源为单位，对采集资源进行调度、
监测；
实现分级优先采集

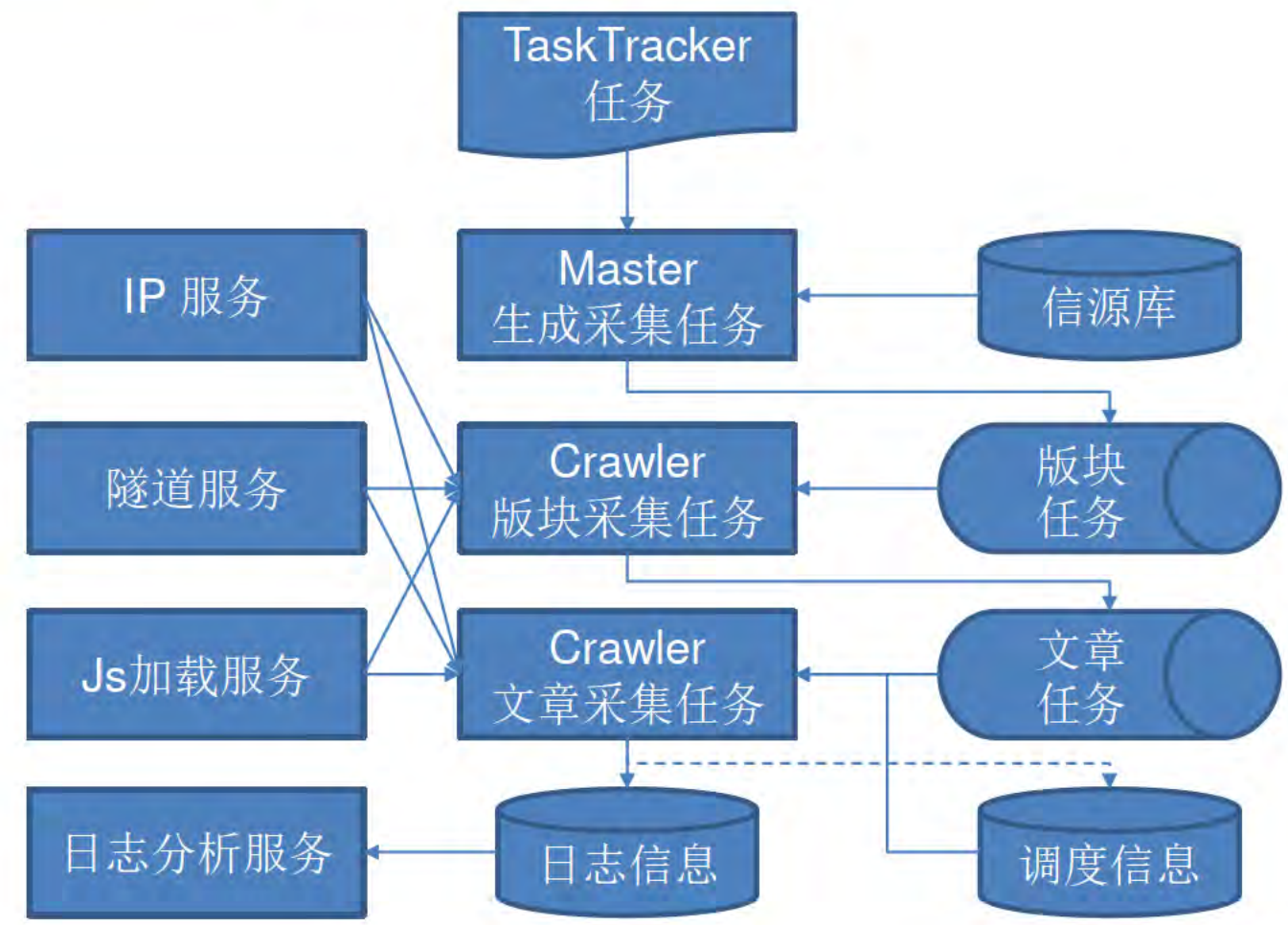


天玑数据采集中心



分布式数据采集——智能采集

智能采集：
按需执行、监测采集任务；
智能防封堵、隧道穿透、动态解析；
实现高召回、高准确数据采集抽取

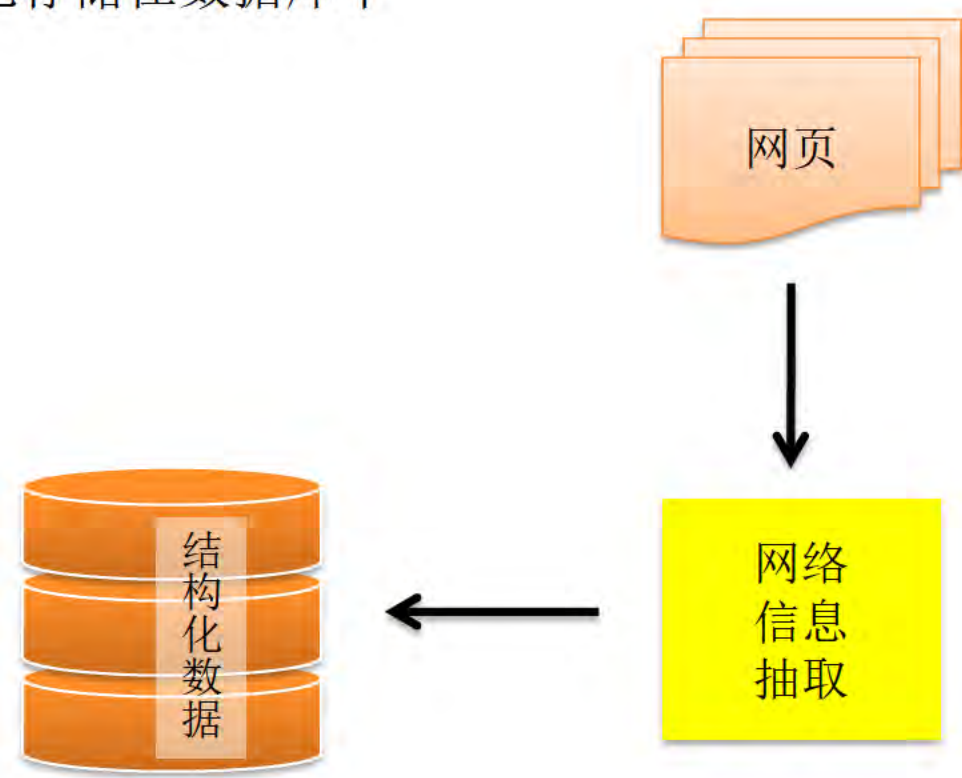


分布式数据采集——智能采集



分布式数据采集——智能采集

- 信息抽取：从半结构化的网页中抽取数据，结构化地存储在数据库中
- 技术方案
 - 半自动化方法
 - 手工编写模板
 - 自动化方法
 - 基于统计的方法
 - 构建DOM树，利用链接文字比、文字密度等统计信息
- 核心组件
 - ❑ 通用信息抽取工具Mask
 - ❑ 搜索引擎结果抽取组件ARISE
 - ❑ 多记录网页内容抽取组件Constor
 - ❑ 单记录网页内容抽取组件HtmlSAE
 - ❑ 网页的同模板自动检测组件WrapperCheck



分布式数据采集——智能采集

抽取组件主要用于资讯类新闻网页抽取，标准的页面结构抽取准确。

抽取组件提供其他功能

抽取组件提供编码识别，编码转换。语种识别，

抽取组件支持抽取多语言
抽取组件支持多语言抽取（英语，中文，藏语，日语，韩语，德语，阿拉伯语）

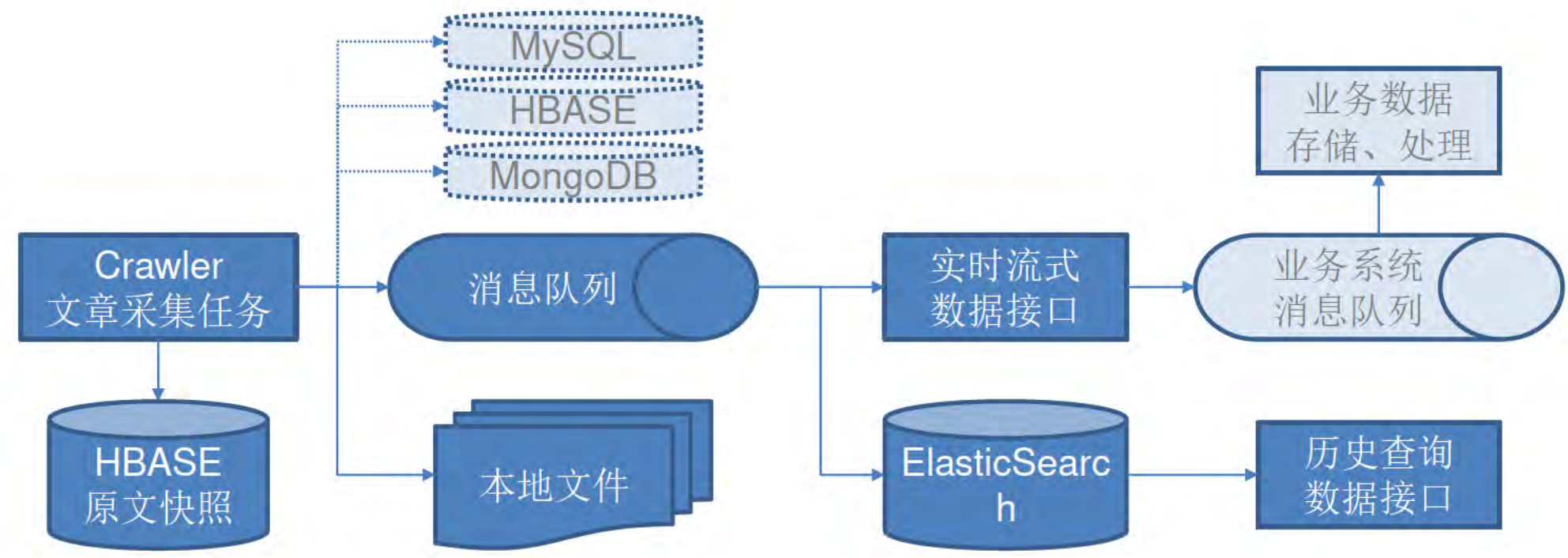


抽取组件准确率
抽取组件准确率包括正文、标题、发布时间等准确率。综合准确率可以达到85%。

天玑数据采集中心



分布式数据采集——数据分发





天玑数据采集中心建有数据实时推送系统、历史数据回溯系统满足用户对于实时数据和历史数据采集需求，同时也可提供元搜索工具或轻便的定制化采集小工具来满足客户定制化的采集需求。

01

数据实时
推送

专题数据推送

基于关键词规则和信源筛选的按需定制和推送

全量数据推送

全量基础数据实时推送

02

历史数据
回溯

数据回溯工具

7种工具回溯微博、微信、贴吧、新闻的账号、文章数据

03

元搜索工
具

元搜索工具

支持百度、必应、搜狗、360、谷歌、微信、贴吧、微博等常规搜索引擎和社媒网站搜索采集

04

定制化采
集小工具

采集小工具

轻便、部署简单、扩展方便、功能齐全
调度中心、信源配置中心、采集执行器三大组件，均支持分布式，支持HA
灵活的采集模板配置，解决大部分采集场景需求
支持多线程采集、动态代理、失败重试等



实时数据推送系统



海量数据采集

通过关键词及媒体的筛选方式组合，过滤出符合配置条件的特定数据。

配置方式：

渠道配置
全部媒体配置
关键词+媒体配置

消费状态

令牌管理

帮助

数据字典

REST接口文档

WEBSOCKET接口文档

数据实时推送系统

当前用户： quanliang 退出

当前时间： 2018年8月27日18:40:00

消费状态

新闻 论坛 博客 微博 微信 境外新闻 APP 电子报 问答 贴吧 Facebook的 推特

数据产生记录

起始位置： 12114447 结束位置： 12281806

当前位置： 12114447

数据产生记录

日期 实际产生数据 日最大可产生

文件请求

2018-08-26 ~ 2018-08-27 查询

取消 保存

- 微博关键词采集
- 微博账号回溯
- 微信关键词采集
- 微信公众号回溯
- 微信文章监测
- 新闻关键词采集
- 贴吧关键词采集

输入项示例

关键词规则：

五月歌人

关键词说明：多个关键词以空格分隔，如：“台湾 无远”；关键词间关系为与，关键词只可由汉字、英文、数字组成，每个关键词字符数不得少于2。

采集时段：

2018-06-04

~

2018-06-04

开始采集

采集结果示例

采集数据结果：

可获得用户名称、文章内容、发布日期、发布时间、转发数、评论数、点赞数、命中关键词、倾向性、分词、原始链接，和是否原创字段。



性能最佳的平台索引

搜索引擎采集软件

Q 搜索

任务

微信公众号

微信搜索

微信公众号

统计监控

统计

监控

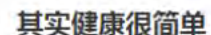
 设置

微信公众号

公众号名称/账号

查询

新增公众号

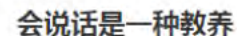


微信号: 40jiandan8322

上次更新时间: 2020-01-02 09:29:20

微信介绍: [jiandan8322](#) 免费口腔黏膜病健康咨询, 查口腔黏膜疾病, 找口腔黏膜专业医生, 了解口腔黏膜病全面知识, 热线: 01082062598

删除

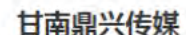


微信号: 40gh_619b3fa8143c

上次更新时间: 2020-01-02 09:29:33

微信介绍：你那么优秀，别输在说话上，在这里跟我一起，学做高情商会说话的女人。(商务合作请加QQ：3461483413)

删除



微信号: 40znzdc931021

上次更新时间: 2020-01-02 09:29:58

微信介绍:znzdc931021甘南鼎兴传媒公众平台是发布便民服务、生活信息、招聘求职信息、企业宣传为一体的综合服务平台。

删除



微信号: 40gh_93d91417de22

上次更新时间: 2020-01-02 09:30:11

微信介绍：关注我，每天跟你分享你喜欢的故事

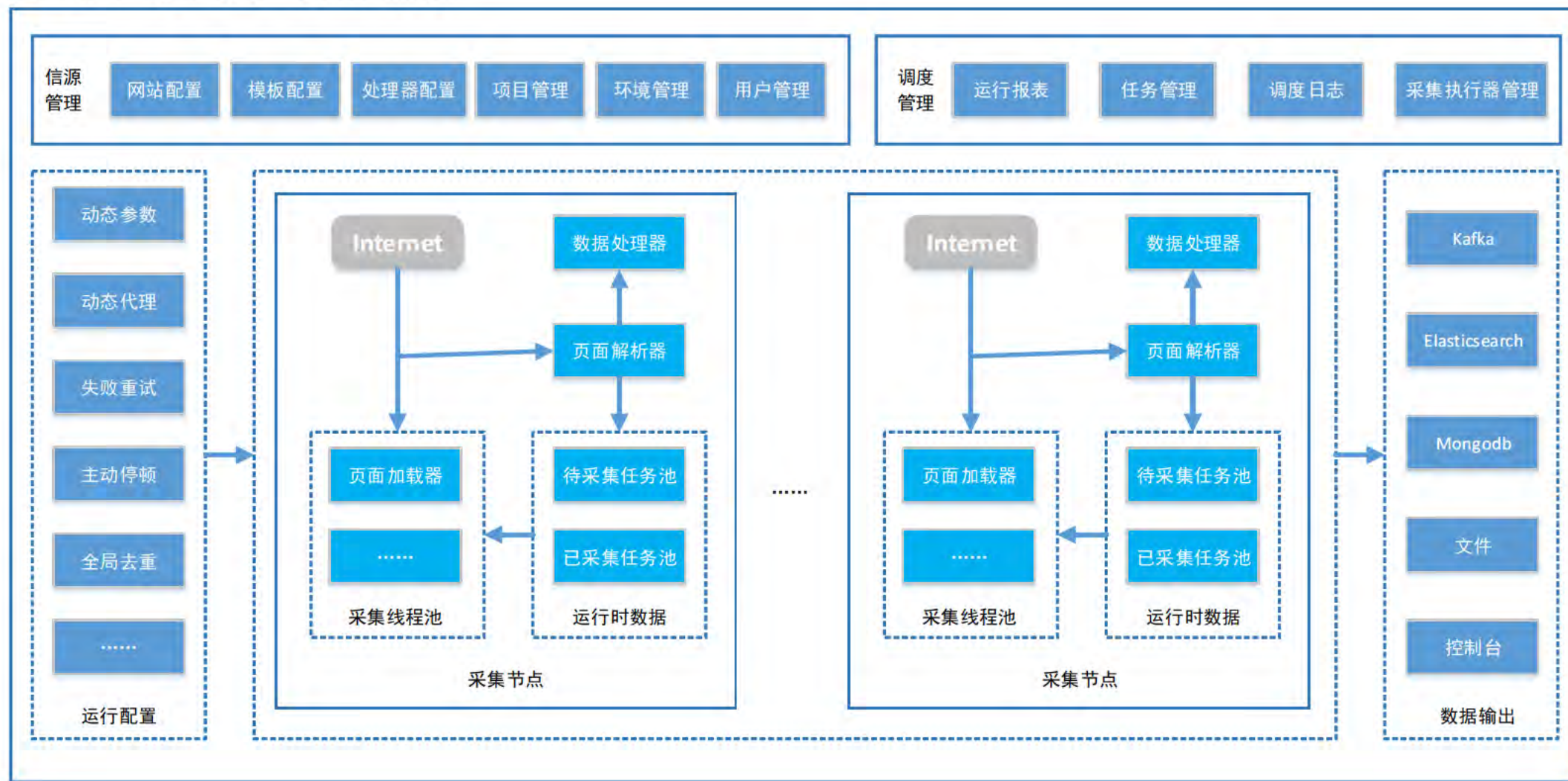
删除



定制化采集小工具



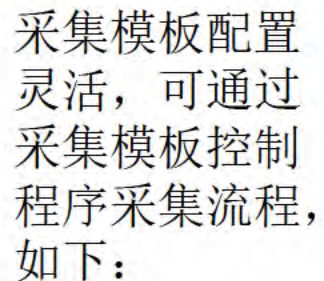
golaxy-crawler 是一个分布式定制化采集工具，其核心设计目标是使用简单、功能齐全、模板灵活、轻量级、易扩展，开箱即用。



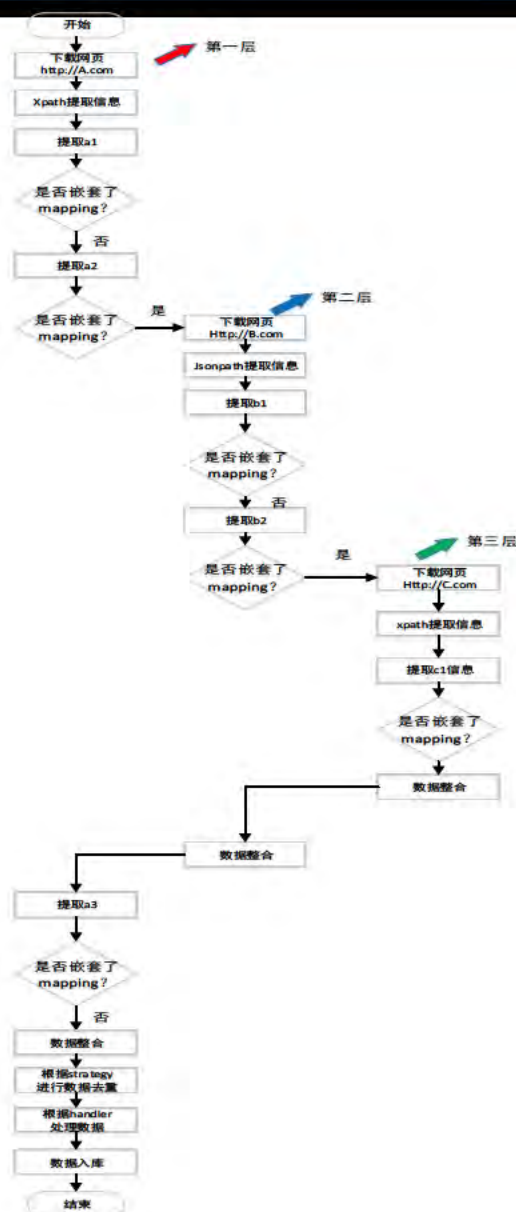
采集模板共包含entries、attributes、mapping、strategy、storage、handler六大部分，控制自定义采集流程：

- **Entries**: 采集任务入口列表，必须，格式为json数组，包含urlPattern、翻页参数pages、http请求参数、自定义属性attributes；
- **Attributes**: 任务携带属性，可选，json格式，可以将entries中所有任务的共性属性提取出来放到这来配置；
- **Mapping**: 字段信息提取映射关系，必须，json格式，包含提取方法method、行数rows、字段列表fields；
- **Strategy**: 主键和去重策略，可选，主键（_id）根据key、algorithm、prefix三个字段生成；
- **Handler**: 数据处理器，可选，在数据最终写入数据源之前对数据做进一步处理；
- **Storage**: 存储设置，可选，指定存储数据源类型和目标库表





第一层





定制化采集工具-配置展示



配置管理中心

当前环境: 测试环境(test) 欢迎 admin

配置管理

项目: 采集测试项目 网站: 搜索 新增配置

每页 20 条记录

KEY	网站	域名	属性	操作
p001.baidu.video	百度视频	video.baidu.com	["location":"北京"]	编辑 删除配置 删除
p001.sthj.shandong	山东生态环境厅	sthj.shandong.gov.cn	["location":"山东"]	编辑 删除配置 删除

第 1 页 (总共 1 页) 上页 1 下页

Powered by GOLAXY 1.6.1 Copyright © 2015-2020

网站配置

配置管理中心

当前环境: 测试环境(test) 欢迎 admin

配置管理

项目: 采集测试项目 网站: 搜索 新增配置

每页 20 条记录

KEY	网站	配置模板	频率	下次采集时间	操作
p001.baidu.video.movie	百度视频	["entries":[{"urlPattern":"http://video.baidu.com/commo...	每小时	2020-03-12 00:00:00.0	编辑 删除配置 删除
p001.p001.sdsth.hjzf	山东生态环境厅	["entries":[{"urlPattern":"http://sthj.shandong.gov....	每30分钟	2020-03-12 00:00:00.0	编辑 删除配置 删除
p001.sthj.shandong.xzd	山东生态环境厅	["entries":[{"urlPattern":"http://sthj.shandong.gov....	每30分钟	2020-03-12 00:11:15.0	编辑 删除配置 删除

第 1 页 (总共 1 页) 上页 1 下页

Powered by GOLAXY 1.6.1 Copyright © 2015-2020

模板配置

配置管理中心

当前环境: 测试环境(test) 欢迎 admin

模板测试

链接: http://sthj.shandong.gov.cn/hj/c/zhwy/ 方法: GET 测试

内容

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
<link href="/favicon.ico" rel="shortcut icon"/>
<title>山东生态环境厅-环境执法行政办法</title>
</html>
```

模板/模板

类型: regex

正则

```
<table width="96%" border="0" align="center"><table></table>
```

输出

```
<table width="96%" border="0" align="center" cellpadding="0" cellspacing="0">
<tr>
<td width="5%" height="26" align="center">img
src="/images/sdjh_104.gif" width="3" height="3" /></td>
<td width="80%"><a href="/202004120200405_3067000.html"
target="_blank" title="山东省生态环境厅2020年度部门内"双随机、一公开"抽查计
划">山东省生态环境厅2020年度部门内"双随机、一公开"抽查计划</a></td>
<td width="15%" align="center" class="date14">2020-04-09</td>
</tr>
<tr>
<td height="1" colspan="3" align="center"
background="/images/sdjh_13.gif"></td>
</tr>
<tr>
<td width="5%" height="26" align="center">img
src="/images/sdjh_104.gif" width="3" height="3" /></td>
<td width="80%"><a href="/202004120200407_3069411.html"
target="_blank" title="2019年第四季度产重超标企业处理情况">2019年第四季度产
```

模板测试

配置管理中心

当前环境: 测试环境(test) 欢迎 admin

处理器管理

项目: 全部 处理器名称: 搜索 新增处理器

每页 10 条记录

处理器名称	描述	更新时间	操作
p002.c2020	测试标准处理器	2020-04-22 22:54:24	可编辑 删除操作 删除 删除
p001.content_skip	正文内容处理器	2020-04-22 22:56:58	可编辑 删除操作 删除 删除

第 1 页 (总共 1 页, 2 条记录) 上页 1 下页

Powered by GOLAXY 1.6.1 Copyright © 2015-2020

处理器管理



感谢您的观看

中科天玑数据科技股份有限公司



序号	成果名称	概述	技术状态 (在研、测试、试验、验收、列装等)	主要用户
1	社交机器人检测	面向国内外主要社交媒体平台，基于特征工程、图神经网络、群组分析等技术，构建社交机器人样本库、话术库、群组关联关系等知识库，识别涉我敏感话题机器人	列装	中央网信办
2	智能算法缺陷分析	针对推荐算法、调度决策算法、检索过滤算法、精准排序算法、生成合成算法这五类常见信息服务算法，研究算法机制、机理，对算法的多样性、公平性、可信性等方面进行风险监测和安全评估	列装	中央网信办
3	大模型攻防技术	研究大模型的数据接入、风险监测、模型脆弱性分析、测试数据集自动构建、可信性测评等	在研	中央网信办
4	智能宣传辅助系统	采集传统媒体、社交网站、即时通信等800万个涉及政治、军事、外交、经济数据信源数据，进行事件分析、人物/群体画像，通过自有账号群建设，内容生成，跨平台智能推送至Twitter、Facebook等社交媒体实现宣传辅助	验收	军委科技委
5	代理实验平台系统	为使用各类境外社交媒体平台的用户或者机构提供批量账号培育服务，提供社交媒体平台账号代理，实验场景编排，实验场景执行等能力	验收	军委科技委
6	智能投送系统	面向批量化的社交媒体账号，提供具备精确传播致效能力的大规模调度服务，目前已实现资源调度、行动规划、协同调度等模块，支持万级万级账号并发调用，千级任务并发	在研	311
7	选情分析系统	面向M、T、港等主要地区，有效支撑对M2016年、2020年，T2020年地区领导人选举，香港2022年特首选举，相关选情系统及选举报告获重要批示，目前支持T2024年地区领导人选举工作	在研	国台办

密



2020 年 11 月 23 日 星期一

批量调度策略限制

发布，评论，转发

- 消息和 Bot 关系

- 一条消息可以被多个 Bot 同时发布
- 一条消息不可以在 4 个小时内被一个 bot 重复发布
- 一个 bot 发布一条消息时，bot 在 5 分钟内不能执行其他任务

- bot 占用：

- 任务对 bot 的占用为五分钟
- 在业务层需要将间隔扩展为十分钟

点赞，关注：

每个 bot 针对同一个用户或消息只能执行一次

点赞和关注操作由于 IP 过少，账号特性驱动，时间间隔需要在小时级别



智能引导组2021年度规划

王永庆



智能引导组定位

- 承担核心系统研制、关键技术攻关以及重要纵向课题申报
 - ❑ 1套核心系统：GoPro（智能引导系统）
 - ❑ 4组关键技术：智能创作、观点挖掘、账号培养、用户画像
 - ❑ 横纵向课题申请：网安、网信、KJW等来源方向
 - ❑ 纵向课题支撑：网安、网信、KJW等来源方向



系统目标

- GoPro系统：
 - ❑ **核心定位：**以“人”为核心的宣传操作平台
 - ❑ **核心价值：**智能辅助境外定向宣传
 - ❑ **系统特点：**交互式观点内容生成、事件倾向性观点分析、高价值账号培养、跨平台用户分析



发展规划

- 第一阶段（2020-2021）：内部发展，确定智能引导系统的基本系统形态及框架
 - 第二阶段（2022-2023）：外部生存，确定智能引导系统的核心用户群，关键技术就具体应用升级
 - 第三阶段（2024-）：影响力提升，立足国家级项目，申报国家级奖项
-



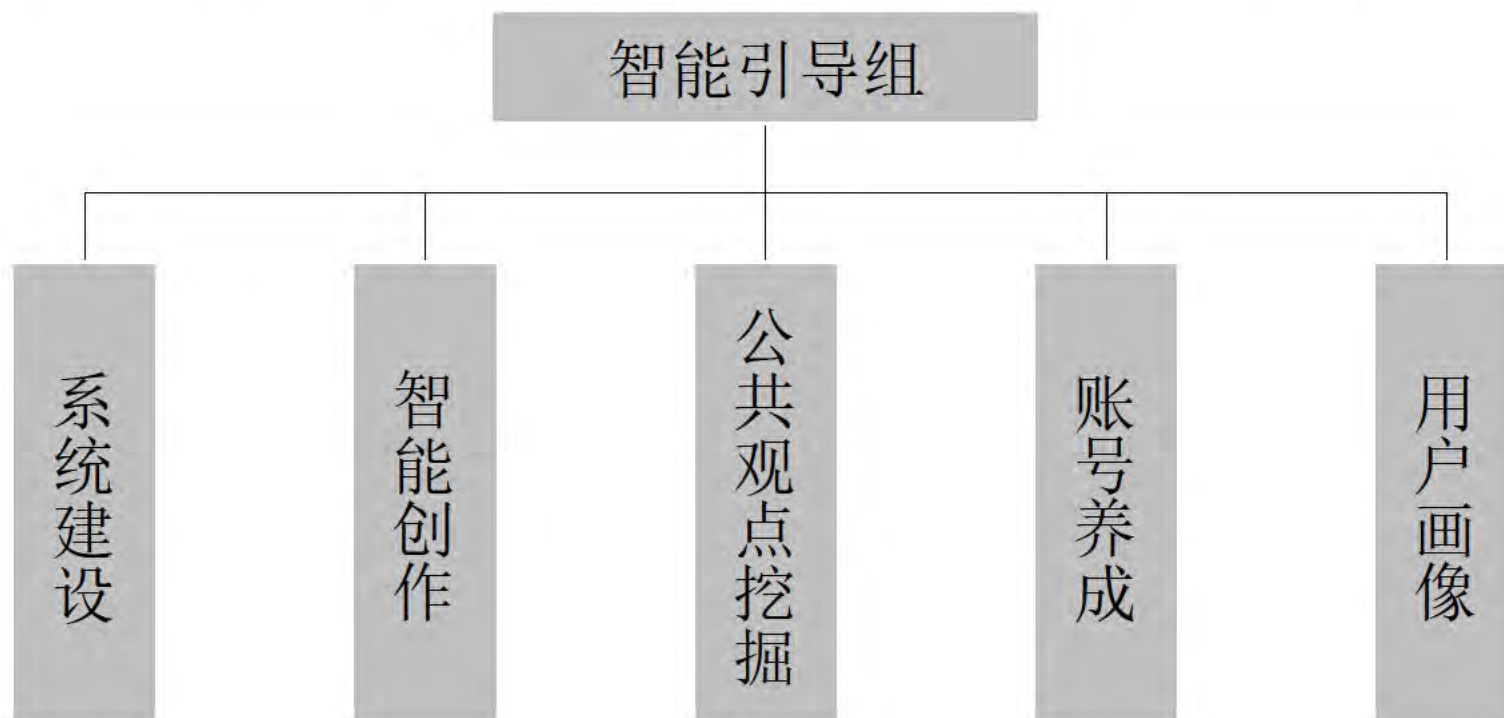
年度成果预期

- 专利/软著10项：涉及数据服务、业务通信、权限管理、文本生成、观点挖掘、账号培养、用户画像等方面
 - 高水平论文10篇：涉及文本生成、观点挖掘用户画像等方面
 - 开源算法平台1套
 - 培育方向4个
-



智能引导组设置

- 围绕GoPro系统能力建设，预设5个方向：





各方向人员安排

- GoPro系统建设方向：
 - 方向负责人：王清林
 - 成员：刘金虎、朱川生、薛雪

- 智能创作与推送策略优化方向：
 - 方向负责人：庞亮
 - 成员：丁汉星、侯丹阳、朱运昌、郝长盈、肖飞



各方向人员安排

- 公共观点挖掘与事件分析方向：
 - 方向负责人：高金华
 - 成员：赵睿卓、王亚杰、姜燕、柯文俊、赵鑫

- 用户画像与群体分析方向：
 - 方向负责人：王永庆
 - 成员：段雪野、高浩、邵江立、张皓彭、史铂深



各方向人员安排

- 账号养成与交互策略优化方向（培育）：
 - ▣ 方向负责人：孟剑



新风貌 新期望

- 正一点：风气正
 - 硬一点：能力硬
 - 飒一点：朝气蓬勃
-



愿未来可期！



用户画像系统目标围绕重点地区收集互联网数据绘制具有一定地区覆盖度的社交用户画像。现阶段主要围绕香港地区的 Twitter 用户，目标绘制 20 万（Twitter 的香港用户约 75 万）用户的社交用户画像。画像具体包括用户的 Twitter 账号信息及用户性别、年龄、职业、所在地、教育水平的基础用户属性。

当前系统通过与中科天玑、郑州分所、采集组合作，以及通过组内自采集方式累积微博用户账号 45,302,104 条，内容信息 434,230,549 条；Twitter 用户账号 32,148,771 条，内容信息 245,799,086 条。按照用户数据类型，将采集数据按照用户属性、行为、内容、关系进行分类处理，共累积 Twitter 用户属性 3200 万+、行为 2.4 亿+、内容 2.4 亿+、关系 950 万+；微博用户属性 4500 万+、行为 1 亿+、内容 4.3 亿+、关系 220 万+。

按照四类用户数据设计数据服务访问接口共 36 个，以上全部数据可通过接口访问。通过数据访问服务接口，设计实现用户画像检索系统 PYou Twitter 及微博版，内网访问地址：10.61.1.241:5000，提供通过用户昵称、所在地、内容关键词及 hashtag 检索用户数据。在已有积累数据基础上，对用户账号的归属地进行推断，归拢所有可能的香港地区用户账号，目前已累积可能的香港地区 Twitter 用户账号 17 万+，具体的属地推断结果展示可访问内网地址：10.61.150.2:5005。

此外，目前还初步探索用户画像的具体展现形式，设计实现用户属性、行为、内容、关系四类数据的展现，具体可访问：10.61.150.2:5005，点击用户详情信息即可查看。

系统截图如下：

- PYou 用户画像系统：10.61.1.241:5000



PYou-Twitter用户库

用户昵称 用户所在地 内容信息 hashtag

搜索用户昵称。例如: Apple Daily HK

搜索

切换到 PYou-微博

PYou-微博用户库

用户昵称 用户所在地 内容信息 hashtag

搜索用户昵称。例如: Apple

搜索

切换到 PYou-Twitter

Twitter用户昵称

Apple



查询用户: Apple

on 显示 首页 上一页 下一页 末页 当前第 1/150 页, 共2985 条记录

头像	昵称	位置	性别	粉丝数	关注数	操作
	Apple Apple		未知	16465	45	查看用户主页
	用户描述: 中部精油舒缓芳療 單純按摩預約 沒有外跑, 沒有到府按摩只能你們過來我這 最晚前一天預約 不接凌晨按摩 一律先匯款。					
	Apple Apple		未知	0	32	查看用户主页
	用户描述:					
	Apple Apple		未知	-1	-1	查看用户主页
	用户描述:					
	APPLE APPLE APPLE	SinRin Nation	未知	-1	-1	查看用户主页
	用户描述:					
	Apple		未知	-1	-1	查看用户主页
	用户描述:					

Twitter用户所在地

香港



查询地点: 香港

on 显示 首页 上一页 下一页 末页 当前第 1/500 页, 共13680 条记录

头像	昵称	位置	性别	粉丝数	关注数	操作
	但丁	香港香港	未知	-1	-1	查看用户主页
	用户描述:					
	Hui Man-cheong	香港	未知	-1	-1	查看用户主页
	用户描述: http://Godahsing.com總編輯, 股東倡議活動者, 新創公司菜鳥, 專業投資者, 節目主持, 發言人, 作者, 男					
	伊死	香港	未知	-1	-1	查看用户主页
	用户描述: 元編集長に名乗った元雑役。現, 某車ケアショップの通訳。声優厨、漫画厨。珈琲、マンガと声優がないと死ぬ。MONSTER HUNTERに永遠のHRZハンター。TweetはPlurkから転送。					
	iwkkwi	香港	未知	13	16	查看用户主页
	用户描述: Action figure collector Toy photographer sketch					
	西雲信	香港	未知	-1	-1	查看用户主页
	用户描述: 一個生於異國他鄉的東西南北人。					



Twitter内容搜索

国安法



检索词 国安法

20条/页 首页 上一页 下一页 末页 当前第 1/105 页, 共2090 条记录

用户昵称	消息内容	发布时间	设备信息
Sami Huang	等国安法下来, 严惩	2020-05-24 00:00:10	twitter.com
阿波罗网唯一官方推号	中共开会放风国安法最晚... 恐怕可囚终身 中南海最害怕... 增30枚核弹头 美军飞弹成功拦截: 人大常委会议月底登场, 放风国安法最晚7月生效... https://t.co/d9BJPY2PXQ #港版国安法 #港区国安法 #国安法	2020-06-28 23:47:02	dlvr.it
威神漫画 cartoon & comic	是赵安法, 不是国安法。	2020-05-22 14:52:36	Twitter for iPhone
禁闻网	港版国安法投票: 习近平按“赞成”李克强伸中指(组图) - 禁闻网 https://t.co/0VmmYINsc #中共 #习近平 #共产党 #国安 #国安法 #投票 #李克强 #港版国安法 #赞成	2020-05-29 13:44:02	禁闻网
禁闻网	中共欲强欲弘 港版国安法可能虎头蛇尾(图) - 禁闻网 https://t.co/J1nLrEUE6V #中共 #习近平 #国安 #国安法 #港版国安法 #美国 #英国 #香港 #黎智英	2020-06-22 13:24:03	禁闻网
阿波罗网唯一官方推号	港警执法扩权无上限 无须法院批准即可进屋搜查(组图/视频)... https://t.co/l3eBlwkBzo #国安法 #香港国安法 #港版国安法 https://t.co/nF6fmWYxJs	2020-07-07 23:42:33	dlvr.it
白麻婆(シロ)随中共中央驻台湾省 ■巡视组组长	支持香港版国安法, 希望明年有望推出台湾版国安法 https://t.co/WFA0vnoWYi	2020-06-29 08:39:43	Twitter for Android
阿波罗网唯一官方推号	港版国安法进逼台湾... 蔡英文暴怒出手了(图)... https://t.co/MdSBNBldd1 #习近平 #蔡英文 #港版国安法	2020-07-07 23:50:02	dlvr.it
中星时报	港区《国安法》通过 港丧失特殊地位	2020-06-30 00:00:10	twitter.com
阿波罗网唯一官方推号	大陆学者谈「港版国安法」: 清算规模不小于六四(图)... https://t.co	2020-05-23 04:09:02	dlvr.it

Twitter Hashtag搜索

NationalSecurityLaw



检索词 NationalSecurityLaw

20条/页 首页 上一页 下一页 末页 当前第 1/23 页, 共455 条记录

用户昵称	消息内容	发布时间	设备信息
香港苹果日报 Apple Daily Hong Kong	Beijing has no rights to enact the controversial #NationalSecurityLaw via its own parliament, the city's top barris... https://t.co/tNvPnW91rK	2020-05-26 19:00:00	TweetDeck
Danilo	Live feed of protests in different areas in HK as LegCo debate @rapplerdotcom @inquirerdotnet @gmanews @ABSCBNNews #NationalSecurityLaw	2020-05-27 00:00:10	twitter.com
EYEPRESS NEWS AGENCY	Protesters take to the street over Beijing's plan to impose national security law, bypassing local lawmakers. #nationalsecuritylaw	2020-05-27 00:00:10	twitter.com
Elsa C Nine	@realDonaldTrump, the "one country, two systems" principle is about to be ruined by the #NationalSecurityLaw passed... https://t.co/4TTTBOfRn	2020-05-24 10:45:07	Twitter for Android
Ken y	@realDonaldTrump, the "one country, two systems" principle is about to be ruined by the #NationalSecurityLaw passed... https://t.co/dNR3255GX1	2020-05-24 10:56:07	Twitter for iPhone
meowmeowtk	Say No to #nationalsecuritylaw say no to #CCPChina	2020-05-24 18:17:07	Twitter for iPhone
DC4HK 🇭🇰	@WhiteHouse Livestream announcing the US countermeasures to the CCP promulgation of #NationalSecurityLaw in Hong K... https://t.co/HODFxTIDw	2020-05-30 02:23:45	Twitter Web App
6all_2	Palestine expressed support for Chinas implementation of the #NationalSecurityLaw in Hong Kong. Congratulations!!... https://t.co/tVg2k0biAP	2020-06-01 16:28:34	Twitter for iPhone
Chris Wong	HKPolice fired tear gas & pepper spray to disperse the protesters. Hong Kong doesn't need the #nationalsecuritylaw... https://t.co/HakawObkAx	2020-05-24 18:18:52	Twitter for iPhone
Joshua Li 李承信	@realDonaldTrump maybe use the #nationalsecuritylaw to fight @Twitter	2020-05-30 08:50:08	Twitter Web App

- 用户属地—香港推断展示：10.61.150.2:5005



用户画像

20条/页

首页

上一页

下一页

末页

当前第 1/8773 页，共175449 条记录

昵称	用户名	性别	香港人概率	用户描述	操作
GA_hker	hongkonger	未知	1.0000	Hker,Free Hong Kong ,Anti extradition law, Anti Chinazi twitter缘要打 YouTube 缘都要打 #StandwithHK #hongkongers #FreeHK #手足互科 https://hkrev.info/main_en.php	查看用户主页
用户填写地区: Hong Kong 地区匹配: 是, 地区匹配得分: 0.8250 用户名匹配: 是, 用户名匹配得分: 0.9325 用户描述匹配: 是, 描述匹配得分: 0.9950 用户语言习惯匹配: 否, 语言使用习惯匹配得分: 0.0000					
AmyKong75018499	Kong	未知	1.0000	香港人Hker 光復香港	查看用户主页
用户填写地区: Hong Kong 地区匹配: 是, 地区匹配得分: 0.8250 用户名匹配: 是, 用户名匹配得分: 0.5750 用户描述匹配: 是, 描述匹配得分: 0.9994 用户语言习惯匹配: 否, 语言使用习惯匹配得分: 0.0000					
hker_coco	Cococho	未知	1.0000	Hker / Stand with Hong Kong	查看用户主页
用户填写地区: Hong Kong 地区匹配: 是, 地区匹配得分: 0.8250 用户名匹配: 是, 用户名匹配得分: 0.9144					

用户画像系统

属性画像

内容画像

行为画像

属性画像

昵称	Kforhongkong	真实姓名	未知	位置	Hong Kong	性别	未知	账号创建日期	2012-12-25 02:17:00
用户名	Kforhongkong	关注好友数	131	粉丝数	38	email	未知	出生日期	未知
描述	#Hker #followbackHongKong #求科旁 #科旁手足 #手足互科							用户主页	查看

用户关注关系

节点

关系

属性画像

内容画像

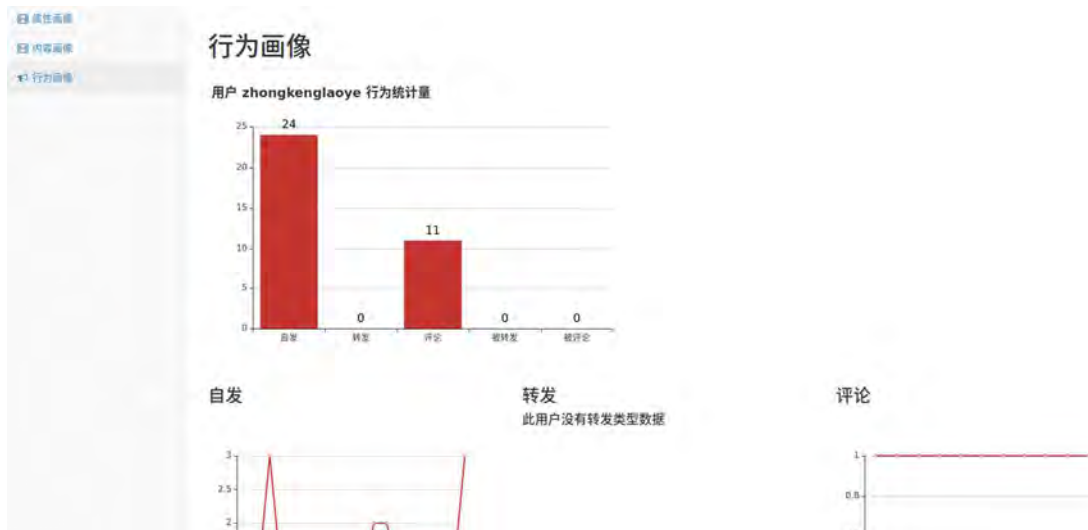
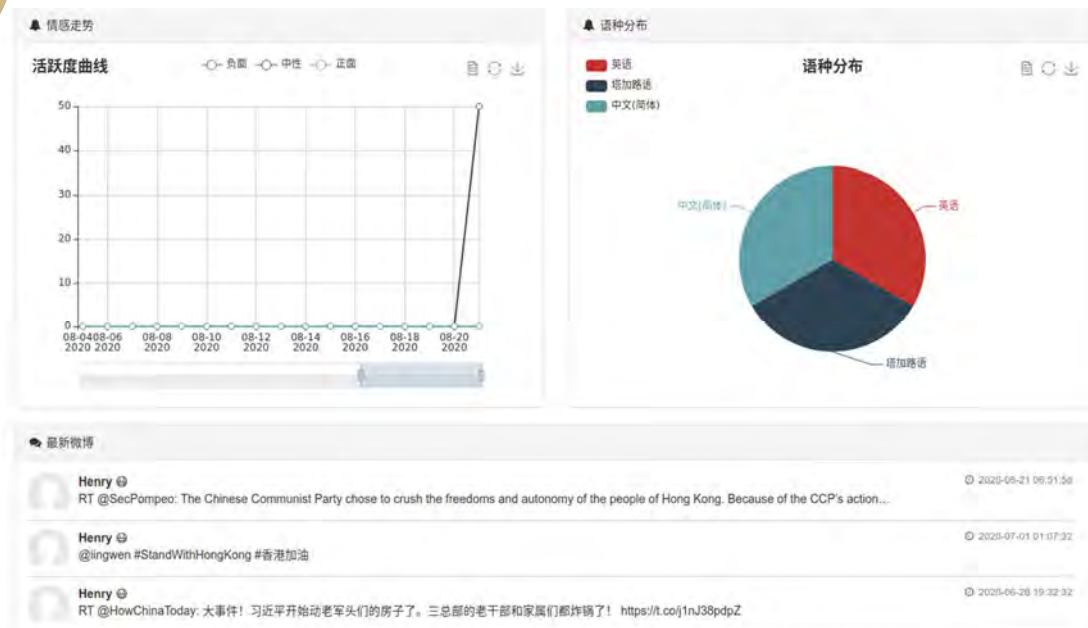
行为画像

内容画像

关键词云

关注话题/事件

- StandWithHongKong
-
-
-
-
-
-
-
-
-





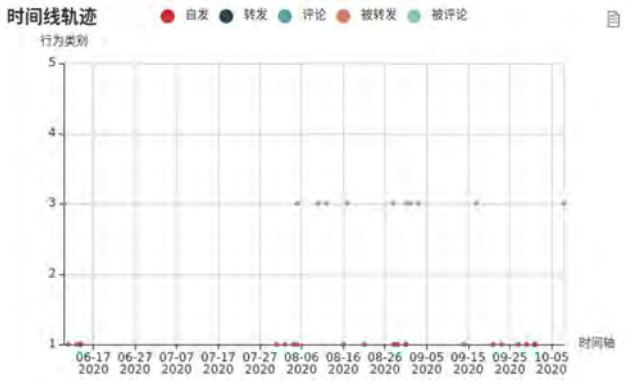
用户相关性排行榜

uid	all	用户转发	用户评论	转发用户	评论用户
Onebtcer	1	0	1	0	0
MsVChina1	1	0	1	0	0
galileocheng	1	0	1	0	0
Daisy_Zhou	1	0	1	0	0
waylybaye	1	0	1	0	0

用户影响力最大内容排行榜

内容id	影响力因子
------	-------

时间线轨迹





揭秘 中科天玑 (Unmasking Zhongke Tianji Technology)

黑心工厂

中科天玑的内部管理极为混乱，压榨员工成风，在新冠疫情期间依然不允许员工居家办公。强迫劳动现象严重，随时随地辱骂员工；平日里加班更是家常便饭，每天8号线地铁末班车都是中科天玑的员工，堪比三角贸易里被塞进船舱的黑奴。

可以再看看中科天玑的招聘信息：<https://www.zhipin.com/gongsi/job/20d958ccb111c44d1X173ty1.html>

中科天玑只愿意为5年工作经历的资深开发者支付最高3万RMB的月薪，等同于市场上3年工作经验开发者的薪资水平；对于大数据的核心——数据分析师，要求硕士毕业起步，却更是不到2万RMB，属实是黑心奴隶主。不管你是什么天之骄子，来了中科天玑这个黑厂就是纯纯的牛马。

再看看招聘要求，典中典，看到"自驱力责任感"这六个字我直接笑嘻嘻了，这不就是黑心老板在招聘启事里藏的黑话吗？翻译翻译：每天自觉加班到十点，周末钉钉在线秒回，背锅时主动往前跨一步，最好能把公司当家住。

内部员工的聊天记录更是说明了中科天玑的管理无序和随意加班的严重程度。“领导拍脑袋，牛马累断腰”，所谓面向领导编程不过如此：



5G

开发二临时(12)

...

你撤回了一条消息

22:25

都干的怎么样了?

26"

22:31

现在及时回复工作消息也做不到了是吧

@所有人

3"

收到，正在改接口，明天能提交

收到，不好意思，刚才没看手机



综上所述，中科天玑就是一个披着科技外皮的劳动密集型黑心作坊，中国那些所谓企业家的那些通病，比如能力低下、道德负数、上位靠野爹扶持.....在这里一点不少，甚至更甚！开发者们提高警惕，别进黑厂，别做二十一世纪的农奴，言尽于此。