需求说明书

# 概述

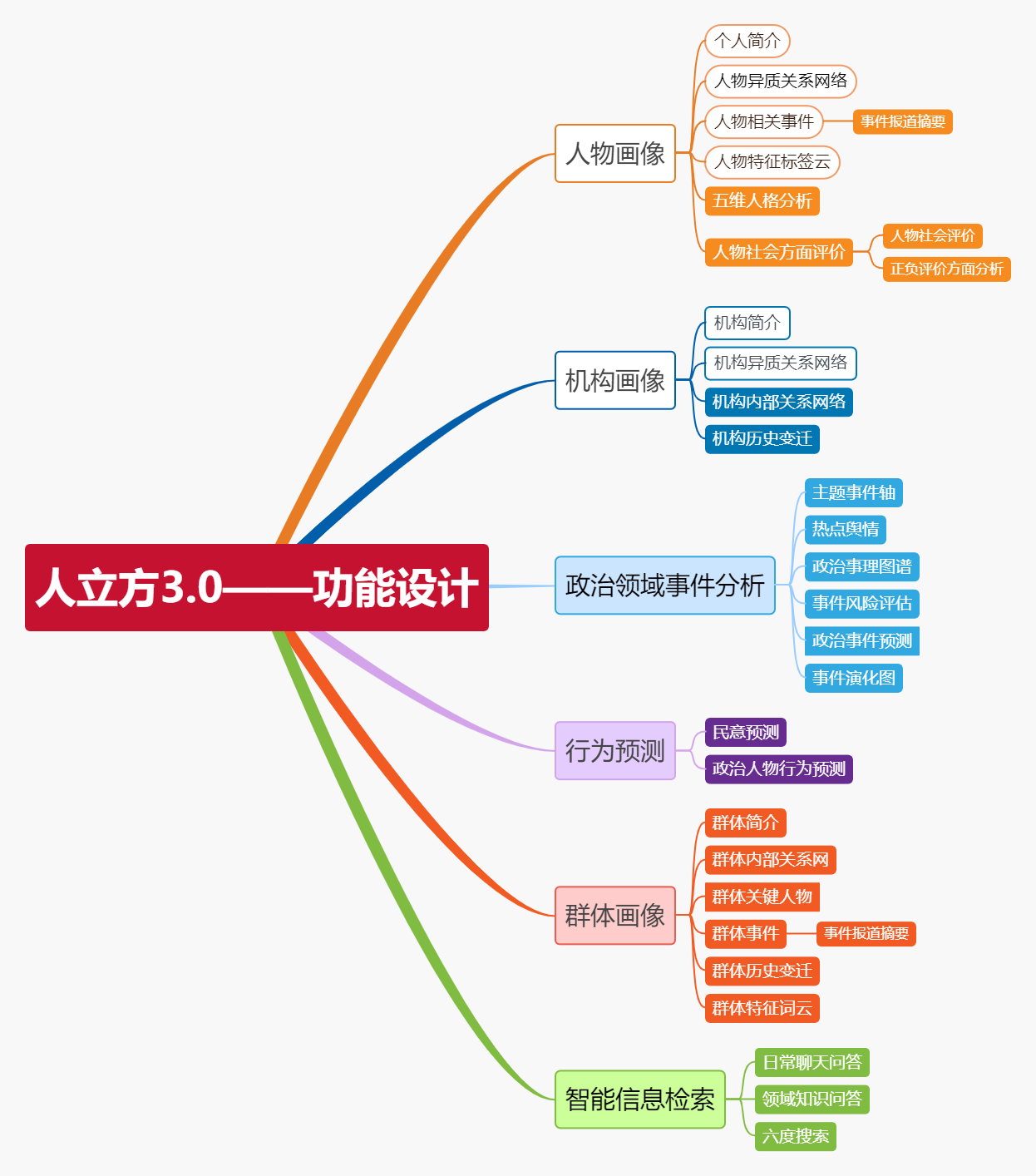
## 开发背景

基于公开数据源和领域知识的融合，通过实体识别，关系抽取等技术构建以人为核心的异构关系图谱，同时针对公众人物利用新闻、维基百科、推特、官方网站等文本数据提取人物特征摘要，收集和分析关联的热点事件及其演化，组成领域情报库。综合上述信息，结合实体和关联检索等功能，构建重点人物关系挖掘系统，并进一步地挖掘图谱网络中的特征。以帮助用户了解台湾问题中所涉及的重点人物、组织的基本情况及其相互之间的关联关系。

## 目标用户

人立方——台湾问题知识服务系统，面向国家和政府情报部门以及心系祖国统一大业的广大人民群众。

## 功能概览



1. 人立方3.0功能分层结构图（有色填充的标签为新增功能）



1. 人立方3.0功能概览图

人立方3.0在人立方2.0原有的功能基础之上增加4个一级功能块，21个二级功能和3个三级功能，组合成为以人物、机构画像为核心的台海领域知识服务系统，其系统结构如图1所示，有色填充的标签表示相对于人立方3.0新增的功能。

人立方台海人物关系挖掘系统的功能块主要分为六大部分，分别为人物画像、机构画像、政治领域事件分析、行为预测、群体画像、智能信息检索。其概览如图2所示。

上述功能中智能信息检索功能块包括的：智能问答系统、简单机器对话、六度搜索3项功能和行为预测模块包括的：政治人物行为预测1项功能为需要实时处理的功能，其余均为提前处理完成，用户查看时查询数据库交给前端展示的前端功能。

# 功能需求说明

## 人物画像功能模块



1. 人物画像功能模块概览

人物画像主要针对台湾问题中的重要公众人物。

人物画像展示的主要内容包括个人简介、异质关系网络、相关事件、特征标签云、五维人格分析、人物社会方面评价几个部分。这些信息将作为主页的主要部分，让用户可以快速地了解该人物的主要特征，并直接找到感兴趣的信息进行查看。

### 个人简介

爬取公众人物在维基百科的个人信息，根据预先设置的属性集对信息进行整理形成结构化的数据，属性集应该包括姓名、性别、籍贯、民族、信仰、党派、生卒年月、职位、生平等，然后与相关的领域数据进行融合得到最终的个人简介。

### 人物异质关系网络

异质关系网络的含义主要在于网络中节点类型的多样性，主要包括人物，地点、机构和热点事件。为了获取以人为核心的关系图谱，主要从两个方面获取实体(节点)和实体间的关联(边)。一方面，从公开网站中爬取的新闻、微博以及官方公众号文本集中通过文本处理的相关技术提取出有效的实体和关系，另一方面，通过分析领域数据库中的属性之间的关联性获取关系。基于上述获取的信息，构建异质关系网络。

### 人物相关事件

从文本数据集中提取热点事件，通过识别文章与人物的共现关联，构造出人物的相关事件集。根据特定的评估标准过滤重要的事件集合并以时间轴等形式展示人物相关的事件。点击事件时，可以立刻跳转到包含事件的文章列表页面。用户可以此快速掌握该人物近期发生的重要事件。

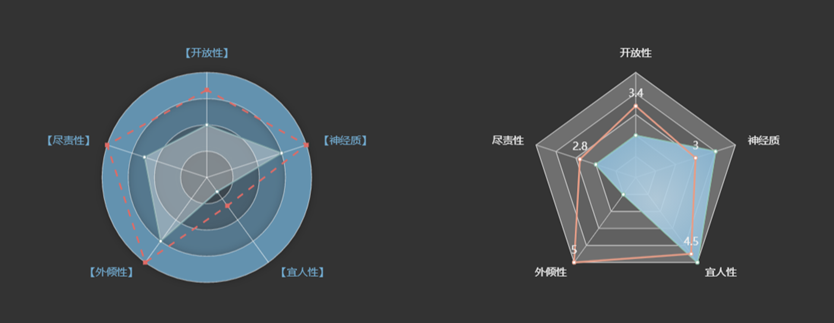
**事件报道摘要：**该功能基于人物相关事件功能，根据人物相关的主题事件模块的聚类结果，利用文本摘要算法抽取报道某一主题事件的全部新闻文章的摘要，对关于该事件的全部报道提取出精炼的简要信息。当用户将鼠标放置于某个事件时，系统即可展示有关该事件的报道摘要。

### 人物特征标签云

特征标签云的含义指利用一系列体现人物特征的名词组成标签云。特征的来源主要通过从人物相关的文本信息中提取重要的有代表性的关键词，然后进行词性过滤和频数统计获得最终的标签集。

### 五维人格分析

从新闻文本集中或社交平台动态文本中抽取相关人物相关的言论等，然后根据文本数据构建与已有人格图谱的关联，通过分层注意力网络学习人物的大五人格分值，其中，大五人格包括：开放性、尽责性、外倾性、宜人性和神经质性，其通过五种特质涵盖了人格描述的所有方面，获取了人格分值后，在人物画像下人物简介模块增添可视化组件，如下图所示，对五人格的分值进行展示：



1. 人格分析展示组件示意图

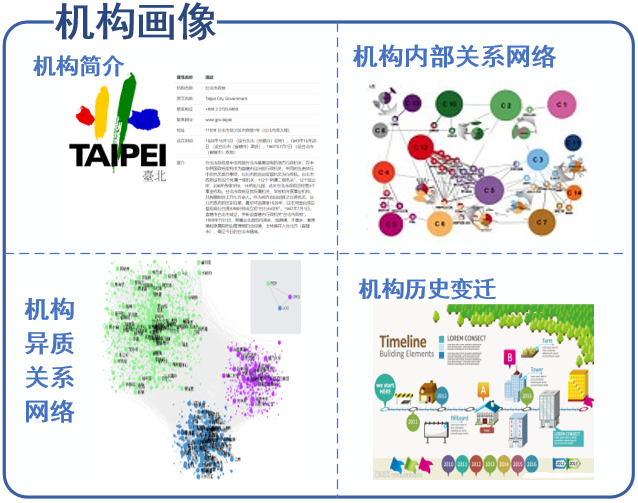
### 人物方面社会评价

从社交网站中提取关键人物主页下的社会大众的评论，用关键词或超级话题搜索与人物相关的大众推文，并对这些短文本进行情感分析，反映人物在社会中的评价。于此同时，识别出某一句评论针对的是人物的行为、言论或其它方面的评价。对情感相同的评价进行关键词抽取，产生类似于淘宝/美团对于商品/商家的正负评价的方面标签。

举例来说，一条针对蔡英文的评价：“又老又醜，做總統什麼事情都不做，整天就知道舔日本，發推都不用漢字。”可以从中识别出为负面评价，跟进一步地，识别出针对外貌的评价“又老又醜”，对行为的评价“做總統什麼事情都不做”，“整天就知道舔日本”，“發推都不用漢字”。对具有相同情感的多条评价进行关键词抽取，可能可以得到的评价方面标签为“丑”，“无能”，“汉奸”等。

人物的方面社会评价可以反映民众从道德层面上对于人物的所作所为的评价，那些行为能为他赢得民心，那些则会让他失去民心。于此同时所有人物的社会方面评价能够反映出台湾社会所普遍认同的道德标准。

## 机构画像功能模块



1. 机构画像功能模块概览

### 机构简介

从Wikipedia的机构词条和机构官网中获取的照片、LOGO和部分基本信息，包括机构名、总部所在、主要业务、成立年月、官方简介等。

### 机构异质关系网络

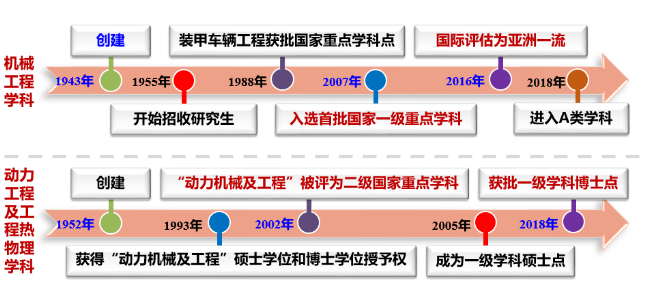
与人物异质关系网络类似，一方面，从公开网站中爬取的新闻、微博以及官方公众号文本集中通过文本处理的相关技术提取出有效的实体和关系；另一方面，通过分析领域数据库中的属性之间的关联性获取关系。基于上述获取的信息，构建异质关系网络。

### 机构内部关系网络

搜集与本机构存在所属关系的人物实体，展示这些人物实体之间所形成的关系网络。

### 机构历史变迁

通过分析机构官网和维基百科的数据，展示机构发展变化的过程，以下图为例，某学科的历史变迁。该部分存在原始数据一致性差的问题，实现难度较高。



1. 历史变迁（参考）

## 政治领域事件分析功能模块

事件是信息传播的基本单元，如何快速了解事件的发展过程，是事件分析模块的核心任务，该模块通过六个部分组成：主题事件轴、事理图谱、事件预测、事件演化图、事件风险评估和热点事件。可以帮助用户从多维度多方面了解事件的发展脉络。



1. 政治领域事件分析功能模块概览

### 主题舆情事件轴



1. 新冠肺炎主题舆情时间轴（仅参考）

从新闻网站上爬取政治领域相关的文章，然后根据主题和文章的内容对文章集合进行聚类分析。对不同主题舆论报道进行时序事件轴展示。点击某个标题时，可以立刻跳转到文章详情页面。用户可以从时间维度了解舆论的发展进程。本部分较难获得准确的舆论主题，实现难度高。

### 政治事理图谱

对事件发展的因果关系进行判别，形成事件发展的因果关系图。该功能模块为用户提供政治领域的脚本事件因果推理功能，从大量政治新闻文本数据中以事件间因果关系抽取、统计方法挖掘政治事件背后潜在的因果关联模式，并以清晰明了的有向图形式呈现给用户。政治事理图谱页面主要包括图谱主体、核心事件查询框、上下文展示区域三个主要部分。

事理图谱主体展示了由核心事件拓展出的事理图谱子图。用鼠标拖动图中空白部分可移动显示区域，拖动节点可以改变图的形状，使用滚轮可以调整图谱的缩放，双击节点可以由此节点继续向下拓展。将鼠标移动到节点上，将高亮显示该节点及其相邻节点，其他节点的显示会变淡。

如图9的事理图谱示意图图例说明如下：

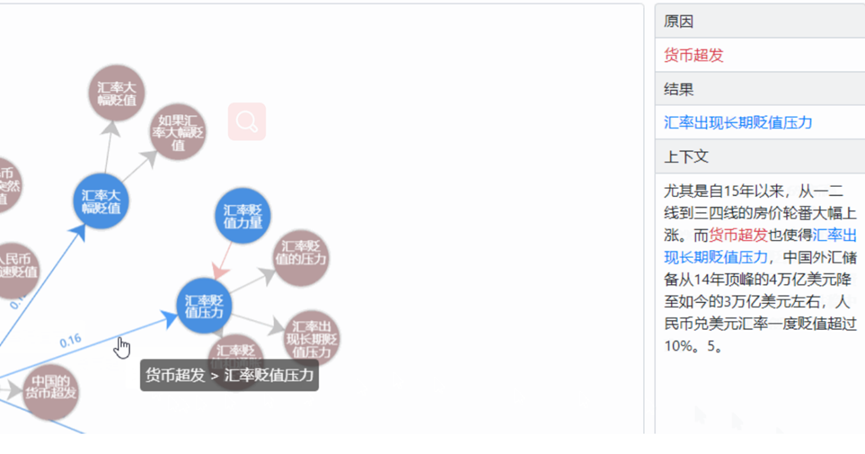
 蓝色节点：表示抽象事件

 棕色节点：表示具体事件

 红色节点：表示核心事件

 蓝色有向边：表示抽象因果关系

* 灰色有向边：表示事件泛化关系



1. 事理图谱（仅参考）

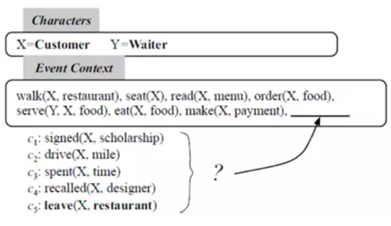
图谱主体上方为核心事件查询框，可以在此输入感兴趣的核心事件，点击“提交”查看由该核心事件拓展出的事理图谱子图。

图谱右侧为因果上下文展示区域，将鼠标移到图谱中因果边上时，此区域会显示抽取该因果边时的上下文。

### 政治事件预测

该功能模块为用户提供了核心事件预测功能（具体预测事件待定），如针对某地区蓝绿阵营变化进行事件预测、对民调支持率变化进行预测等。

脚本事件预测任务要求给定一个事件的上下文 ( context )，需要从候选列表中选出接下来最可能要发生的事件，如下图所示：



1. 事件预测示例

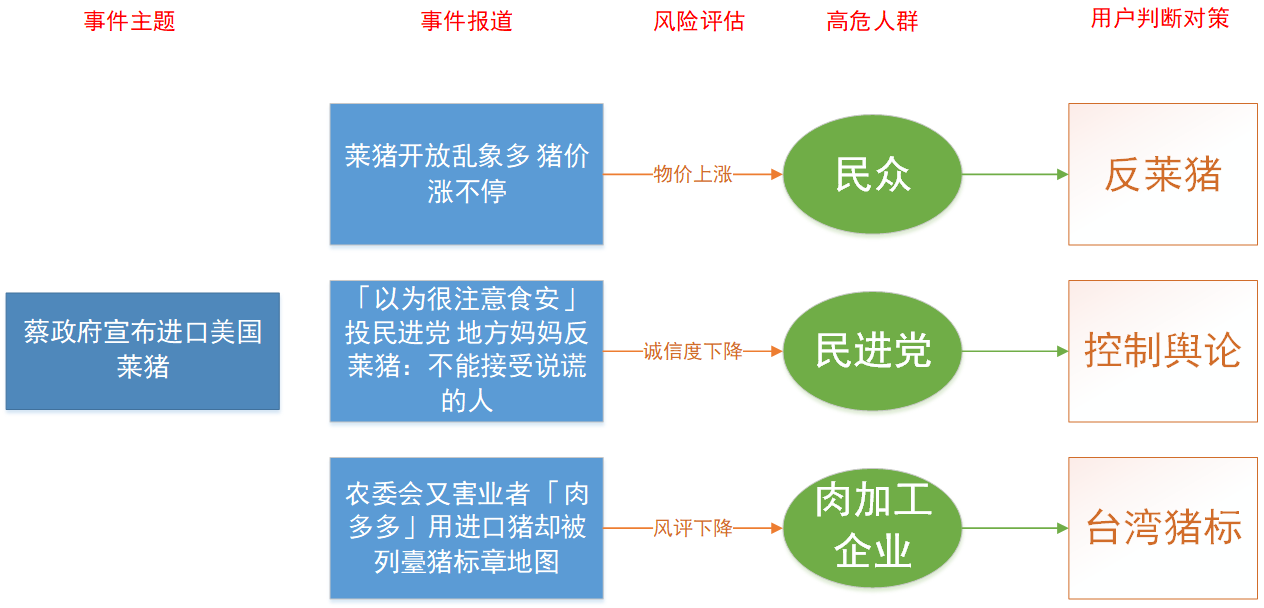
这是在餐馆场景下发生的一系列事件，根据已经发生的事件，5个候选事件中下一个最可能发生的事件就是离开餐馆( leave )。

参照上述脚本事件任务，将候选事件封闭地设定为用户所关心的问题（具体预测事件待定），依据大量新闻文本数据抽取出的文本摘要，将其抽象为事件作为模型的输入，从而得到最终事件预测结果。同时，用户也可以设定输入模型事件的时间跨度范围，模型根据不同输入得到不同的预测结果。

### 事件演化图

事件随时间的进行，整体的发展脉络，包括顺承、因果、演变分支、不同主题事件的发展交汇影响等过程。

### 事件风险评估



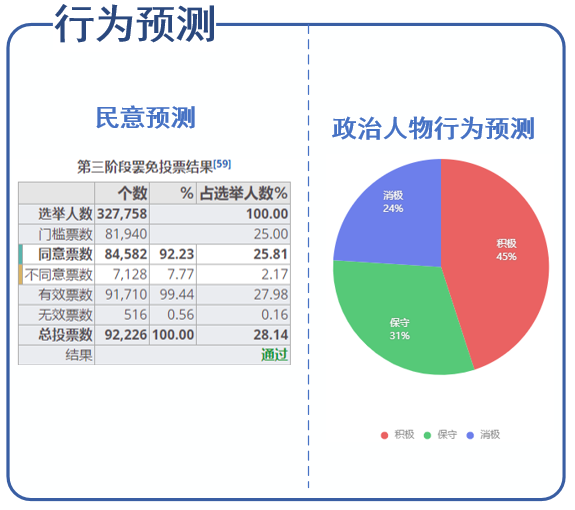
1. 事件风险评估示意图

通过分析近期的新闻报道和主题事件，结合事件因果、事件预测、事件演化的分析结果，预测近期发生的事件可能对社会、政治、经济所造成的不良影响，给出事件的风险评估值。如果可能，进一步寻找受事件负面影响最深重的高危人群。该功能可以用以辅助决策，当具有不良影响的事件发生时可以参考这些信息做出应对策略，尽可能控制影响的广度和深度。

### 热点事件

展示各个新闻网站中近一段时间点击量最高、受关注度最高的主题事件，点击对应主题后跳转到该主题的事件时间轴。

## 行为预测功能模块



1. 行为预测功能模块概览

### 民意预测

台湾的各个民调机构也都有着鲜明的政治立场，其民调结果大多数都相差甚大，甚至与最终的结果也都相去甚远。本功能将参考历史上的数次公投的民调、相关事件和最终结果，通过分析近期的事件和相关民调数据，结合事件预测功能，对于即将进行的人民公投、选举等活动的最终结果进行预测。

### 政治人物行为预测

政治人物具有比较大的社会影响力，其行为往往会引起一系列后续事件的发生。对于一个事件，政治人物做出的反应实质上是多方立场和利益权衡的结果，应该是可以预测的。本功能通过分析近期的舆情和相关民调数据以及人物的画像（主要是关系网络、性格信息），判断针对某个主题事件某个政治人物所采取的行动。目前将行动按照积极程度分为三类：

**积极：**通过包括声援、回应、引导舆论、资源投入、直接插手等方式，积极参与并尝试引导事件的发展。

**保守：**参与事件发展，但仅通过类似评论、表态等有限的方式进行参与，对于事件的发展持中立、观望态度。

**消极：**拒绝参与事件，且拒绝发表任何相关的言论，对于事件的发展持完全不关心、逃避、忽略的态度。

## 群体画像功能模块

在台湾选举中，由蓝绿阵营、党派、年龄段、职业等区分的不同群体对选举的倾向性具有很强的关联性，因此分析描述群体画像对台湾选举预测，具有非常重要的影响作用。

### 群体划分

1. **蓝绿政治族群**

台湾蓝绿阵营主要是由政治立场来区分，“蓝”表示以国民党为首的反对台独的台湾政治群体，“绿”表示以民进党为首的支持台独的台湾政治群体。在台湾选举中表示分别支持国民党和民进党的两大阵营。由于国民党的党旗颜色是蓝色为主，所以支持国民党的阵营有泛滥、深蓝、浅蓝三个群体。同样地，支持民进党的阵营有泛绿、深绿、浅绿三个群体。

**泛蓝**：主张台湾属于中国，也指台湾国民党和与国民党相同政治主张的党派及其拥护者，这不是一个正式的组织。主要包括中国国民党、亲民党和新党及其支持者。代表人物马英九、朱立伦、宋楚瑜等。

**泛绿**：主张台湾独立，也指台湾民进党和与民进党相同政治主张的党派及其拥护者，这不是一个正式的组织。主要包括民主进步党、台湾团结联盟、新国家连线及其支持者。代表人物蔡英文等。

**深蓝**：坚决支持一个中国原则，反对“台独”，承认“九二共识”。主张发展两岸关系。主要包括国民党亲中派、新党、中华统一促进党及其支持者。代表人物洪秀柱、赵少康、王建煊、张安乐等。

**深绿**：坚决主张“台独”的激进群体，主要包括民进党独派、台湾团结联盟、时代力量、建国党及其支持者们。代表人物李登辉、黄国昌、蔡英文等。

**浅蓝**：能够容忍现状再决定未来走向，反对极端台独主义

**浅绿**：本土派

1. **党派群体[[1]](#footnote-1)**

台湾的主要党派有国民党、民进党、亲民党等，对不同政党进行画像表述对选举预测非常有意义。

**中国国民党（泛滥）**：中国国民党成立于1894年，自1949年迁台后，一直以“执政党”身份掌控台湾政局。在2000年3月的台湾地区领导人选举中，因国民党本身分裂导致票源分散，而丧失“执政权”，沦为在野党。直至2008年3月22日，台湾地区领导人选举，中国国民党籍候选人马英九、萧万长获胜，国民党重新执政。在2016年1月的台湾地区领导人选举中败选，再次沦为在野党。国民党现任党主席为江启臣 。

**亲民党（泛滥）**：2000年3月31日，独立参选人宋楚瑜凭借在2000年台湾地区领导人选举中所赢取的466万票，正式宣布成立以“人民第一”为宗旨的亲民党，并逐步发展为国民党、民进党之外的台湾第3大党。亲民党现任党主席为宋楚瑜。

**民国党（摇摆泛滥）**：由前中国国民党籍民意代表（兼国民党中常委）徐欣莹等人于2015年1月发起组党，2015年3月13日备案正式成立。民国党自称成继孙文的理念，精神是博爱，宗旨是自由、平等、民主、法治、科学。

**新党（深蓝）**：1993年8月10日，赵少康、王建煊等7名国民党中生代“立委”因不满李登辉日益偏离“九二共识”，推行“台独”分裂路线而另组新党。认同孙中山先生的理念，追求民族统一、政治民主、民生均富的目标。新党现任党主席为吴成典。

**中华统一促进党（深蓝）**：2005年10月25日，台湾“光复”六十周年，与此同时，台湾第一百三十个政党，台湾保护中华大同盟党正式改名为中华统一促进党。该党总裁张安乐在接受记者专访时表示，“中华统一促进党是目前在台湾既能深入基层各个角落，又敢公开主张‘和平统一、一国两制’的政党”。

**民主进步党即民进党（泛绿）**：1986年9月28日，反对国民党专制独裁、追求“民主自由”、主张“台湾独立”的各股“党外”势力合组建成“民主进步党”(简称民进党)，打破了国民党长期“一党专制”的局面。民进党成立后，一直稳步发展，并在2000年3月的台湾地区领导人选举中，因国民党阵营的分裂而渔翁得利，一夕之间成为“执政党”。在2008年台湾地区领导人选举中败选，再次沦为在野党。在2016年1月的台湾地区领导人选举中胜选，再次执政。民进党现任党主席为蔡英文。

**台湾团结联盟（深绿）**：2001年8月12日，以李登辉为“精神领袖”，以延续“两国论”分裂路线为宗旨的“台湾团结联盟”正式成立，并在2001年底“立委”选举中获得13个席次，成为“立法院”第四政党。进入“立法院”后，为抢攻岛内“急独”票源，主导台湾政局走向、阻扰两岸关系的发展，“台联党”一再抛出各种“激进台独”主张，逐渐成为岛内“激进台独”势力的总代表。

**时代力量（深绿）**：现为台湾立法院第三大政党，现任党主席陈椒华。2015年1月25日成立，由林昶佐担任建党总队长，黄国昌于5月3日加入，并于7月2日成为代理总队长，9月13日当选为主席团执行主席。2016立法委员选举得5个席次，仅次于民进党与国民党，并拥有国会党团。代表人物黄国昌、林昶佐、洪慈庸、徐永明、高潞等，代表事件台湾青年学生“太阳花运动”也称“学运”。

**台湾民众党（非蓝绿）**：台湾民众党2019年08月06日开创党大会，大会之前的会前会，应到110人、实到72人，过半党员出席并通过章程，柯文哲以63票担任首任党主席。根据章程，党主席任期4年，得连选连任。根据2019年8月台湾民调，“台湾民众党”以9.1%的支持度，击败“时代力量”，进而成为第三大党。代表人物柯文哲。

1. **党内派系**

自孙文组建中国国民党起，其党内的派系就一直错综复杂，究其原因，是辛亥革命后组成革命势力的人来自各个地区、各个阶级，各个派系虽然都打着革命的旗号，但实际上仅有极少部分人是忠于革命的。蒋介石时期，国民党的党内建设一直未受到重视，虽然各路军阀再一次统一到了青天白日旗下，但依然是雄踞一方各怀鬼胎，即使是蒋介石的直系中，凡是有一定地位的人也都会结党营私，悄悄拉起自己的山头。这一现象一直持续至今，派系的斗争不仅造成了国民党的数次分裂，也成为了国民党在目前台湾的政治中屡屡败退，日渐式微的重要原因。

而台湾政坛的另一大党派民进党也并非铁板一块。民进党是由多个党外组织组合而成，因而也就演变出后来的诸多派系。尽管2006年7月23日，民进党“全代会”通过提案，解散党内所有派系，但是派系依然在民进党内变换为基金会、智库等形态存在。最开始的时候，主要是美丽岛系与新潮流系以及不结盟的政治人物，在民进党的成长过程中，不断经历派系实力的消长与重组。美丽岛系在经过数次路线之争以及新的年轻派系如正义连线、福利国连线等的挑战，已解散重组。民进党党内的派系众多，派系斗争始终激烈。

俗话说“百足之虫、死而不僵”，即使有一天祖国实现了统一，但若不能瓦解台湾旧有的政治体系，台湾也很可能会像香港一样，在统一多年后再陷入混乱。为此，研究和挖掘台湾各大党派内部的势力矛盾，通过分裂、肢解反动势力以实现社会的平稳过渡和长治久安也是具有较大政治意义的工作。

1. **不同年龄段群体**

台湾选票要求20岁以上所以分为20-25岁，25-30岁，30-40岁，50-60岁，60岁以上。据了解台湾年轻人多泛绿，即使在深蓝眷村中长大的年轻人也多为泛绿，而年长的人大多都认为台海会在未来统一。

通过对不同年龄层的fb数据和民调中不同年龄层支持的数据进行挖掘来描述不同年龄层群体的画像，可以打上“蓝绿”标签。

1. **不同职业群体**

教师、律师、新竹高科技人、鸿海（没想好不同职业人群对选举的影响，这个群体可以去掉）

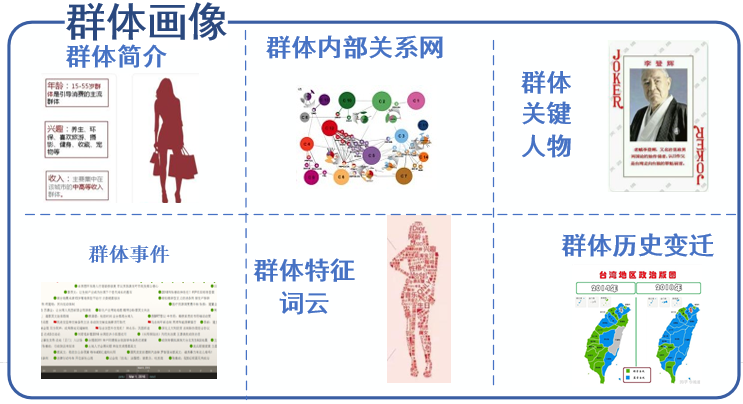
1. **不同地区群体**

桃园市、台北、新竹、台中、嘉仪、台南、高雄等眷村分布较多的地区大多是“蓝”营。对不同地区群体的蓝绿分析，对台湾选举预测具有重要价值。

划分方式（四）到（六）涉及民众的数据，有隐私权问题，目前难以实现。

### 功能设计

群体画像的主要功能有群体简介、群体内部关系网、群体关键人物、相关事件、群体历史变迁、群体特征词云等。



1. 群体画像功能模块概览

#### 群体简介

爬取群体在维基百科的信息，根据预先设置的属性集对信息进行整理形成结构化的数据，然后与相关的领域数据进行融合得到最终的群体简介。应包含群体的主要特征，群体内的代表人物（点击人物名字应能跳转到人物画像系统内）。



1. 群体简介（仅参考）

#### 群体内部关系网

蓝绿阵营、政党等群体都包含很多代表性重点人物，政党可以归属在机构中，也可以归属在群体内。挖掘出群体内人物的关系对群体的刻画也是至关重要的。另外，群体的拓扑关系可能也需要时序拓扑关系，因为有些关系是在特定时间存在，特定时间结束的。



1. 群体内部关系网（仅参考）

#### 群体关键人物

通过对群体内部关系网的分析和人物实体本身的异质关系网络，提取出在群体内部和社会中具有巨大影响力的若干个关键人物进行展示。

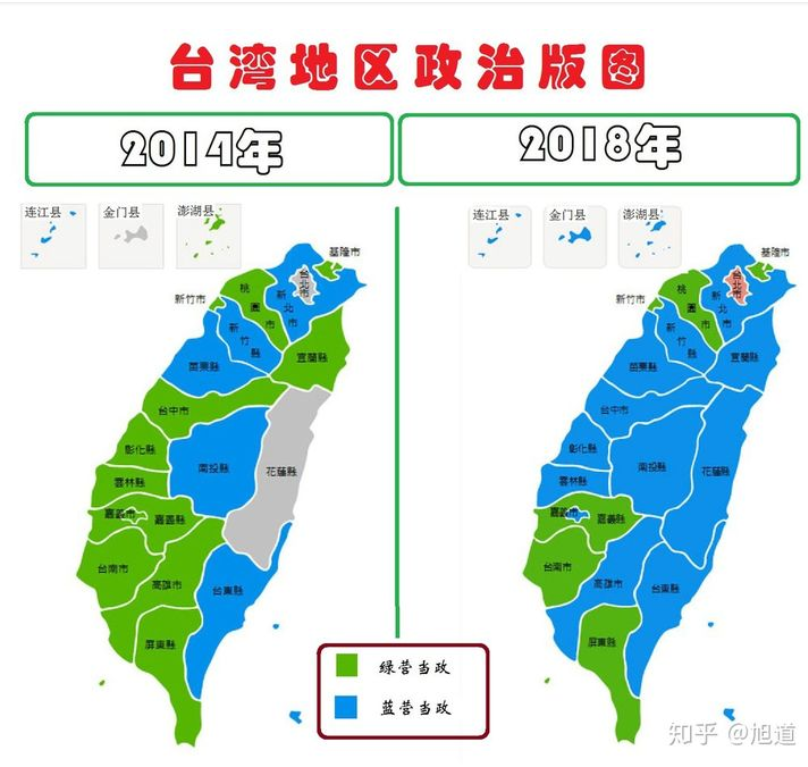
#### 相关事件

识别群体及群体的重点人物所关联的事件，获取在一段时间内该群体所关联的重点人物集。利用重点人物集，通过群体或人物的共现词对，从文本数据集中提取与群体相关的热点事件。根据特定的评估标准过滤重要的事件集合并以时间轴等形式展示群体或群体重点人物相关的事件。点击事件时，可以立刻跳转到包含事件的情报列表页面。用户可以此快速掌握该群体或群体内重点人物近期发生的重要事件

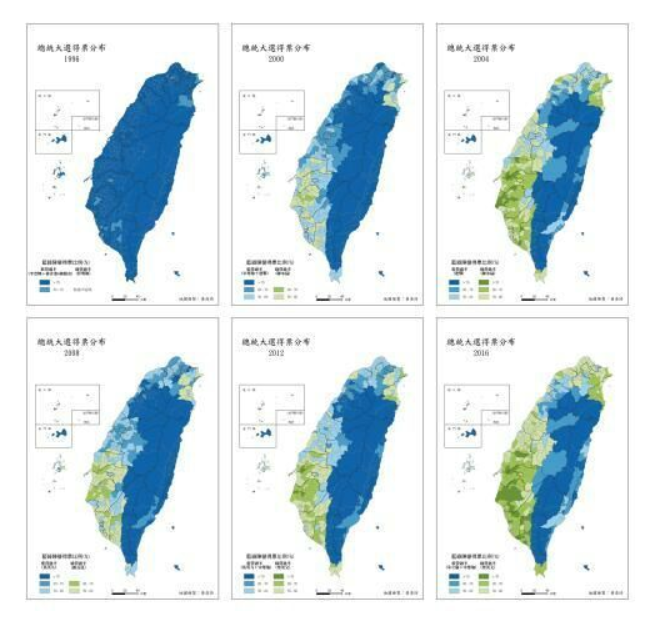
**事件报道摘要：**该功能基于人物相关事件功能，根据人物相关的热点事件模块的聚类结果，利用文本摘要算法抽取报道某一热点事件的全部新闻文章的摘要，对关于该事件的全部报道提取出精炼的简要信息。当用户将鼠标放置于某个事件时，系统即可展示有关该事件的报道摘要

#### 群体历史变迁

台湾蓝绿阵营历史



2018年12月24日，台湾九合一选举之后蓝绿阵营分布图



#### 群体特征词云（群体特征摘要）

特征标签云的含义指利用一系列体现群体特征的名词组成标签云。特征的来源主要通过从群体或群体内重点人物相关的文本信息中提取重要的有代表性的关键词，然后进行词性过滤和频数统计获得最终的标签集。

## 智能信息检索功能模块



1. 智能信息检索功能模块概览

### 领域知识问答

本系统从新闻及百科数据等垂直数据构建知识图谱，由此可以开发出一个基于知识图谱的问答机器人。通过对用户所输入的问题进行分析，系统将生成相应的数据库查询方法，将得到的结果展示出来供用户查看。

首先，搭建出一个基于传统规则方式的问答系统，即：设定好问题种类，对于用户提出的问题进行关键词匹配，用数据库查询语句进行查找，组织好结果语句后展示出来，供用户查看。

下面是一个医疗领域问答机器人的示例：



传统的问答机器人一般只能基于知识图谱的连边情况进行回答，如果由有限数据构建的知识图谱不够完善，就可能忽略真实世界中的实际连边。本模块还将支持多跳问答，可以挖掘实体间的潜在连边，提供给用户尽可能完善的结果。

### 日常聊天问答

面向通用领域的聊天问答功能，用户给出偏向日常聊天的问题，如询问姓名，天气等，系统给出一个在人类看来比较合理的回答，为领域知识问答功能所无法识别的问题给出合理的回答。

### 六度搜索

该功能块以人际关系网的六度理论为基础，即认为，地球上的任意两个人都可以通过其它的若干个人在六跳以内的关系而建立连接。从人物的关系网中查找两个目标实体的六跳以内的关系通路，并通过剪除其中不必要的实体从而获得两个人之间的重要关系通路，这有助于分析两个重要人物实体之间的公开的和潜在的关系。

# 与现有系统的对比

## 人物画像部分

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 人立方3.0 | 人立方2.0 | Aminer | Microsoft Academic Graph | 天眼查 | 知乎大V 人物画像 |
| 公开来源 |  |  | gct.aminer.cn | academic.microsoft.com | www.tianyancha.com | https://www.hanspub.org/journal/paperinformation.aspx?paperid=36030 |
| 领域 | 政治 | 政治 | 学术 | 学术 | 工商业 | 社交 |
| 画像目标人群 | 关键政治人物 | 关键政治人物 | 学者 | 学者 | 企业公司高管、股权人 | 知乎大V |
| 数据源 | 多网站新闻 推特 维基百科 | 单网站新闻 | 论文专著 | 论文专著 | 工商局官方信息 司法部门官方信息 新闻 | 知乎文章 |
| persona/profile | persona+profile | persona+profile | persona+profile | profile | persona+profile | profile |
| 主要刻画维度 | 基本属性 特征词云 异质关系网络 相关事件 人格分析 社会评价 | 基本属性 特征词云 异质关系网络 相关事件 | 基本属性 研究兴趣 同质关系网络 领域影响力预测 学术成果 | 研究兴趣 异质关系 领域影响力及发展历史 学术成果 | 基本属性 任职信息 曾任职信息 实际控制权 风险预警 异质关系网络 新闻舆情 | 知乎影响力 粉丝群体画像 |
| 特色维度 | 人格分析 社会评价 相关事件 | 相关事件 | 领域影响力预测 | 领域影响力及发展历史 | 任职信息 曾任职信息 实际控制权 风险预警 新闻舆情 | 粉丝群体画像 |
| 关键技术 | 实体识别 实体消歧 关系抽取 关系归一化 关键词发现 事件抽取 人格分析 情感分析 | 实体识别 实体消歧 关系抽取 关系归一化 事件抽取 | 学者信息抽取 兴趣关键词聚类 影响力分析和预测 | 论文作者实体发现 研究领域实体发现 论文集、活动和机构实体发现 | 数据聚合 实体抽取 实体消歧 关系抽取 风险分析 | 数据聚合 情感分析 关键词发现 |

## 机构画像部分

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 人立方3.0 | 人立方2.0 | Microsoft Academic Graph | 天眼查 | 天元数据 企业画像 |
| 公开来源 |  |  | academic.microsoft.com | www.tianyancha.com | http://qpx.tdata.cn/a/sy/sy |
| 领域 | 政治 | 政治 | 学术 | 工商业 | 工商业 |
| 画像目标机构 | 政党、政府单位、公司、企业 | 政党、政府单位、公司、企业 | 科研机构、高校 | 公司、企业 | 公司、企业 |
| 数据源 | 多网站新闻 官网 维基百科 | 单网站新闻 | 论文专著 | 工商局官方信息 司法部门官方信息 知识产权局官方信息 新闻 | 工商局官方信息 知识产权局官方信息 新闻 |
| persona/profile | persona+profile | persona+profile | profile | persona+profile | profile |
| 主要刻画维度 | 基本属性 特征词云 异质关系网络 机构内部关系网络 机构变迁历史 | 基本属性 特征词云 异质关系网络 | 主要研究方向 异质关系 领域影响力及发展历史 学术成果 | 公司背景 司法风险 经营风险 公司发展 经验状况 知识产权 新闻舆论 | 企业概况 工商登记信息 风险信息 经营信息 知识产权 公司发展 舆情信息 |
| 特色维度 | 机构内部关系网络 | 无 | 领域影响力及发展历史 | 公司背景 司法风险 经营风险 公司发展 经验状况 知识产权 新闻舆论 |  |
| 关键技术 | 实体识别 实体消歧 关系抽取 关系归一化 | 实体识别 实体消歧 关系抽取 关系归一化 | 论文作者实体发现 研究领域实体发现 论文集、活动和机构实体发现 | 数据聚合 实体抽取 实体消歧 关系抽取 深层关系挖掘 风险分析 | 数据聚合 实体抽取 实体消歧 关系抽取 深层关系挖掘 风险分析 |

## 事件分析部分

|  |  |  |  |
| --- | --- | --- | --- |
|  | 人立方3.0 | 学迹 | 北航Ring |
| 领域 | 政治 | 通用 | 通用 |
| 数据源 | 新闻 | 新闻 | 新闻 贴吧 |
| 主要功能 | 主题事件轴 热点子事件 事件因果图 事件演化图 事件预测 | 主题事件集 事件因果 产业影响 | 主题事件轴 热点事件集 事件演化 事件热度趋势 |
| 特色功能 | 事件预测 | 产业影响 | 事件热度趋势 |

1. 参考文章：

   https://www.sohu.com/a/425225597\_99958605

   https://www.zhihu.com/question/22660013/answer/837674142 [↑](#footnote-ref-1)