

CSC 实验室数据整理及说明手册

(更新于 2018 年 4 月 27 日)

1 数据整理

序号	数据集	数据量(条 篇)	获取时间	负责人	存放位置
1	豆瓣电影数据	56084	2017.02	王玮玮	个人电脑
2	厦大新闻网访问日志	84989138	2004.11-2014.02	纪幼纯	个人电脑
3	享项项目数据	29998	2016.06	腾飞	享项服务器
4	享项新闻数据	5033	2016.01-2016.06	腾飞	享项服务器
5	厦门大学论文信息数据	26043	2006-2016	杞坚玮	个人电脑
6	金融新闻数据	310340	2008.01-2017.03	王辉强	公司服务器
7	上市公司基本数据	3170	2017.02	王辉强	公司服务器
8	上市公司高管信息数据	90596	2017.02	王辉强	公司服务器
9	上市公司主营业务构成数据	129832	2017.02	王辉强	公司服务器
10	法律文书	3617485	2018.03	胡志强	公司服务器
11	金融问答数据	219769	2018.04	胡志强	公司服务器

2 数据说明

2.1 豆瓣电影数据

(暂无)

2.2 厦大新闻网访问日志

(暂无)

2.3 享项项目数据

(暂无)

2.4 享项新闻数据

(暂无)

2.5 厦门大学论文信息数据

(暂无)

2.6 金融新闻数据

金融新闻数据(数据库: financial)			
数据表	数据来源	数据量(条)	数据样例及字段说明
financialNews	证券日报 (http://zqrb.ccstock.cn) 中证网-宏观经济 (http://www.cs.com.cn/xwzx/) 第一财经 (http://www.yicai.com/news/) 东方财富网 (http://finance.eastmoney.com.html) 经济观察网 (http://www.eeo.com.cn/) 网易财经-不动产 (http://quotes.money.163.com) 巨潮资讯网 (http://three.cninfo.com.cn) 腾讯财经 (http://finance.qq.com) 新浪财经 (http://finance.sina.com.cn) 财经时报 (http://ibtimes.com.cn) 投资者网 (http://www.investorchina.com.cn) 全景网	310340	数据样例: ID: 1 timestamp: 1484726160 timestampCrawl: 1484841618 title: 领读 郎永淳: 读书支撑着我一直往前跑下去 keywords: content: 在《经济观察报》举办的“中国最受尊敬企业”年会中, 郎永淳担任主持。身着西装, 系着黑色小领结, 整个人看上去有些消瘦, 但很精神。主持间隙, 我们做了个简单的采访, 聊了些他的境况以及家庭和读书的话题~~~ source: eeo url: http://www.eeo.com.cn/2017/0118/296491.shtml img: http://upload.eeo.com.cn/2016/1221/thumb_650_305_1482309625743.jpg author: 王雅洁 classes: 领读中国 views: 3717 entity: 哈佛,阿图·葛文德,约翰·克利斯朵夫,陆陆,公司,孙正义,达尔文,大学,郎永淳,国家 similarity: 427,429,12857,133594,133594 字段说明: ID: 新闻标识 timestamp: 新闻日期 timestampCrawl: 新闻时间 title: 新闻标题 keywords: 新闻摘要 content: 新闻正文 source: 新闻来源

金融新闻数据(数据库: financial)			
数据表	数据来源	数据量(条)	数据样例及字段说明
	(http://www.p5w.net)		url: 原始 URL img: 新闻图片 author: 作者 classes: 分类 views: 浏览次数 entity: 命名实体 similarity: 相似新闻

2.7 上市公司基本数据

上市公司基本数据(数据库: financial)			
数据表	数据来源	数据量 (条)	数据样例及字段说明
companySurvey	公司概况 (http://f10.eastmoney.com/f10_v2/CompanySurvey.aspx?code=)	3170	数据样例: ID: 1 code: 600939 name: 重庆建工集团股份有限公司 name_en: Chongqing Construction Engineering Group Co.,Ltd. name_old: -- A_code: 600939 A_name: 重庆建工 B_code: -- B_name: -- H_code: -- H_name: -- S_type: 上交所主板 A 股 industry: 工程建设 manager: 陈晓 legal_representative: 魏福生 deputy_secretaries: 窦波 chairman_of_the_board: 魏福生 securities_affairs_representative: -- independent_director: 张永水,陈箭宇,童文光 phone: 023-63511570 email: zqb@cceg.cn fax: 023-63525880 website: www.cceg.cn

上市公司基本数据(数据库: financial)			
数据表	数据来源	数据量 (条)	数据样例及字段说明
			<p>office_address: 重庆市经开区北区金开大道 1596 号 registered_address: 重庆市经开区北区金开大道 1596 号 area: 重庆 zip_code: 401122 registered_capital: 18.1 亿 business_registration: 9.15E+17 number_of_employees: 14719 number_of_managers: 23 law_office: 北京市金杜律师事务所 accounting_firm: 大信会计师事务所(特殊普通合伙)</p> <p>company_profile: 公司是以建安和路桥施工、市政建设为主业,集工程设计、机械制造、特许经营、物流配送等为一体的大型企业集团,具有房屋建筑工程施工总承包特级、公路工程施工总承包特级、市政公用工程总承包壹级为主的资质体系,涵盖桥梁、隧道、机电、钢结构工程专业承包壹级、轨道交通、设计和多个其它业务领域一级、甲级资质。集团注册资本 16.33 亿,资产总额 548.98 亿,下属全资、控股企业 29 家,拥有专业技术人才 1 万余名。</p> <p>business_scope: 房屋建筑工程施工总承包(特级),市政公用工程施工总承包(壹级),机电安装工程施工总承包(壹级),建筑装修装饰工程专业承包(壹级),钢结构工程专业承包(壹级),公路工程施工总承包(贰级),园林古建筑工程专业承包(贰级),预应力工程专业承包(贰级),城市轨道交通工程专业承包资质,隧道工程专业承包(贰级),预拌商品混凝土专业承包(贰级),土石方工程专业承包(壹级),消防设施工程专业承包(壹级),地质灾害治理工程施工甲级,工程设计;建筑行业(建筑工程)甲级,(以上经营范围按资质证书核定事项从事经营),承包本行业境外工程和境内国际招标工程,上述境外工程所需的设备、材料出口,对外派遣本行业工程、生产及服务的劳务人员,建筑科技咨询服务,仓储(不含危险品),国内航空、陆路货物运输代理,销售建筑材料(不含危险化学品),建筑设备租赁,授权范围内的国有资产经</p>

上市公司基本数据(数据库: financial)			
数据表	数据来源	数据量 (条)	数据样例及字段说明
			<p>营。</p> <p>date_of_establishment: 1998/4/21</p> <p>listing_date: 2017/2/21</p> <p>Issue_price_earnings_ratio: 22.92</p> <p>online_release_date: 2017/2/9</p> <p>distribution_method: 网上定价发行,网下询价配售,市值申购</p> <p>each_face_value: 1</p> <p>circulation: 1.82 亿</p> <p>per_issue_price: 3.12</p> <p>issue_cost: 2832 万</p> <p>total_market_value: 5.66 亿</p> <p>raised_net_funds: 5.38 亿</p> <p>first_day_opening_price: 3.74</p> <p>first_day_closing_price: 4.49</p> <p>first_day_turnover_rate: 0.03%</p> <p>highest_price_first_day: 4.49</p> <p>net_placement_rate: 0.02%</p> <p>pricing_success_rate: 0.09%</p> <p>字段说明:</p> <p>ID: 标识</p> <p>code: 股票代码</p> <p>name: 公司名称</p> <p>name_en: 英文名称</p> <p>name_old: 曾用名</p> <p>A_code: A 股代码</p>

上市公司基本数据(数据库: financial)			
数据表	数据来源	数据量 (条)	数据样例及字段说明
			A_name: A 股简称 B_code: B 股代码 B_name: B 股简称 H_code: H 股代码 H_name: H 股简称 S_type: 证券类别 Industry: 所属行业 Manager: 总经理 legal_representative: 法人代表 deputy_secretaries: 董秘 chairman_of_the_board: 董事长 securities_affairs_representative: 证券事务代表 independent_director: 独立董事 phone: 联系电话 email: 电子信箱 fax: 传真 website: 公司网址 office_address: 办公地址 registered_address: 注册地址 area: 区域 zip_code: 邮政编码 registered_capital: 注册资本(元) business_registration: 工商登记 number_of_employees: 雇员人数 number_of_managers: 管理人员人数

上市公司基本数据(数据库: financial)			
数据表	数据来源	数据量 (条)	数据样例及字段说明
			law_office: 律师事务所 accounting_firm: 会计师事务所 company_profile: 公司简介 business_scope: 经营范围 date_of_establishment: 成立日期 listing_date: 上市日期 Issue_price_earnings_ratio: 发行市盈率(倍) online_release_date: 网上发行日期 distribution_method: 发行方式 each_face_value: 每股面值 circulation: 发行量(股) per_issue_price: 每股发行价(元) issue_cost: 发行费用(元) total_market_value: 发行总市值(元) raised_net_funds: 募集资金净额(元) first_day_opening_price: 首日开盘价(元) first_day_closing_price: 首日收盘价(元) first_day_turnover_rate: 首日换手率 highest_price_first_day: 首日最高价(元) net_placement_rate: 网下配售中签率 pricing_success_rate: 定价中签率

2.8 上市公司高管信息数据

上市公司高管信息数据(数据库: financial)			
数据表	数据来源	数据量(条)	数据样例及字段说明
management	东方财富网 (http://emweb.securities.eastmoney.com/CompanyManagement/Index?type=web&code=sz300059#glcj-0)	90596	<p>数据样例: ID: 1 code: 603817 name: 汪璟 duties: -- date_of_service: 2015/10/30 date_of_departure: -- gender: 男 education: 大专 birthday: 1969 resume: 汪璟先生,1969 年 11 月出生,中国国籍,无境外永久居留权,大专学历,曾任中国建筑第六工程局四公司项目经理,北京桑德环保产业集团总裁助理,北控水务集团有限公司给水事业部副总经理、水务事业部副总经理、中原业务区总经理、.....</p> <p>字段说明: ID: 标识 code: 股票代码 name: 姓名 duties: 职务 date_of_service: 任职日期 date_of_departure: 离职日期 gender: 性别 education: 学历 birthday: 出生年份 resume: 个人简历</p>

2.9 上市公司主营业务构成数据

上市公司主营业务构成数据(数据库: financial)			
数据表	数据来源	数据量(条)	数据样例及字段说明
mainBusiness	主要业务构成 (http://f9.eastmoney.com/sh601899.html#cwzb)	129832	数据样例: ID: 1 code: 603817 btype: 按地区分类 post_date: 2016/6/30 income_composition: 福州地区 main_income: 1.61 亿 income_ratio: 99.74 main_costs: 7890 万 cost_ratio: 99.72 main_profit: 8257 万 profit_ratio: 99.77 gross_margin: 51.14 字段说明: ID: 标识 code: 股票编号 btype: 分类 post_date: 报告期日期 income_composition: 收入构成 main_income: 主营收入(元) income_ratio: 收入比例(%) main_costs: 主营成本(元) cost_ratio: 成本比例(%) main_profit: 主营利润(元) profit_ratio: 利润比例(%)

上市公司主营业务构成数据(数据库：financial)			
数据表	数据来源	数据量(条)	数据样例及字段说明
			gross_margin: 毛利率(%)

2.10 法律文书

数据表	数据量(篇)	数据样例及字段说明
机动车交通事故责任纠纷-案例.txt	668001	<p>数据样例:</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 1 1 葛印楼与孙果城、青县凯达汽车运输队机动车交通事故责任纠纷一审民事判决书</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 14 2 河北省唐山市丰南区人民法院</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 15 3 民 事 判 决 书</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 2 4 (2017)冀 0207 民初 511 号</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 3 5 原告：葛印楼，男，1954 年 9 月 27 日生，汉族，住河北省遵化市。</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 3 6 被告：孙果城，男，1978 年 9 月 27 日生，汉族，住河北省沧州市青县。</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 3 7 委托代理人：崔志学，丰南区丰南镇宏伟法律服务所法律工作者。</p> <p>.....</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 4 26 原告葛印楼与被告孙果城、青县凯达汽车运输队、南皮县明大汽车运输队、中国人民财产保险股份有限公司青县支公司、刘太胜、吴海强、李建平、边志永、中国人民财产保险股份有限公司鹤壁市分公司、华安财产保险股份有限公司邯郸中心支公司机动车交通事故责任纠纷一案，本院受理后，依法组成合议庭，公开开庭进行了审理。.....</p> <p>.....</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 6 28 孙果城辩称，我的车辆在被告青县人保和华安财险投保了机动车交强险和商业三者险，事故发生在保险期限内，原告的损失应当由以上保险公司优先履行给付义务，不足部分按照双方责任比例承担；已生效的判决中已经明确了我为所驾车辆的实际所有人。</p> <p>.....</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 7 33 当事人围绕诉讼请求依法提交了证据，本院经审理认定事实如下：.....</p> <p>.....</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 8 42 本院认为，依照《最高人民法院〈关于审理人身损害赔偿案件适用法律若干问题的解释〉》第十一条规定：.....</p> <p>.....</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 9 43 依照《中华人民共和国侵权责任法》第六条、第四十八条，《中华人民共和国道路交通安全法》第七十六条，《最高人民法院关于审理人身损害赔偿案件适用法律若干问题的解释》第十一条和有关法律、法规，</p>

数据表	数据量(篇)	数据样例及字段说明
		<p>.....</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 16 44 判决如下:</p> <p>.....</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 10 51 案件受理费人民币 1720 元, 由被告吴海强、孙果城各负担 860 元。</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 11 52 如不服本判决, 可在判决书送达之日起十五日内, 向本院递交上诉状, 并按对方当事人的人数提出副本, 按照河北省高级人民法院冀高法[2015]79 号意见提交支持上诉事由和上诉诉讼请求的证据等材料, 上诉于河北省唐山市中级人民法院。</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 12 53 审判长 王树宁</p> <p>.....</p> <p>00003076-119d-11e8-9a7d-172214d14ea3 99 58 特别提示: 根据冀高法[2015]79 号意见, 当事人不服一审判决、裁定, 在法律规定期间内上诉的, 通过原一审人民法院提出, 并提交以下材料:</p> <p>.....</p> <p>字段说明:</p> <p>第 1 列: 案例唯一编号(keyid)</p> <p>第 2 列: 分段标记编号(tid)</p> <p>第 3 列: 案例自然段序号(orderid)</p> <p>第 4 列: 文书自然段正文(section)</p> <p>注: txt 包含两列字段, 字段之间分隔符"tab"(Python/Java 语言中\t)</p>
机动车交通事故责任纠纷-争议焦点.txt	180222	<p>数据样例:</p> <p>00005279-296b-11e7-990f-17070a1d658f @hc 关于本案争议的焦点: 1、关于人寿保险公司辩称原告的起诉超过诉讼时效问题, 本案交通事故的发生时间是 2014 年 3 月 6 日, 因李如强涉嫌交通肇事, 被淮北市杜集区人民检察院提起公诉, 本院于 2014 年 9 月 4 日作出(2014)杜刑初字第 00056 号刑事判决。在刑事案件的审理过程中, 死亡受害人的亲属张长根作为刑事附带民事诉讼的原告人曾提起过附带民事诉讼, 后撤诉。现张祚环于 2015 年 5 月 28 日向本院提起民事赔偿诉讼, 未超过诉讼时效, 人寿保险公司的上述理由依法不能成立。2、.....</p> <p>字段说明:</p> <p>第 1 列: 案例唯一编号</p> <p>第 2 列: 案例争议焦点文本(正文中@hc 表示自然段分段符)</p>

数据表	数据量(篇)	数据样例及字段说明
		注：txt 包含两列字段，字段之间分隔符"tab"(Python/Java 语言中\t)
民间借贷纠纷-案例.txt	2002346	<p>数据样例：</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 1 1 江西远大保险设备实业集团有限公司与徐军民间借贷纠纷一审民事判决书</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 14 2 江西省樟树市人民法院</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 15 3 民 事 判 决 书</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 2 4 (2017)赣 0982 民初 322 号</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 3 5 原告：江西远大保险设备实业集团有限公司，住所地：江西省樟树市观上工业区。</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 3 6 法定代表人：徐小华，该公司董事长。</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 3 7 委托诉讼代理人：邹明才，男，该公司副总经理，代理权限：特别授权。</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 3 8 被告：徐军，男，1987 年 8 月 30 日出生，汉族，住樟树市，</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 4 9 原告江西远大保险设备实业集团有限公司与被告徐军民间借贷、买卖合同纠纷一案，本院于 2017 年 2 月 23 日立案后，依法适用普通程序，于 2017 年 4 月 17 日公开开庭进行了审理，原告委托诉讼代理人邹明才到庭参加诉讼，被告徐军经传票传唤无正当理由拒不到庭参加诉讼。本案现已审理终结。</p> <p>.....</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 7 12 当事人围绕诉讼请求依法提交了证据，本院组织当事人进行了证据交换和质证。对当事人无异议的证据，本院予以确认并在卷佐证。.....</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 8 13 本院认为，合法的借贷关系应受法律保护。被告向原告借款 100000 元后重新出具借条的事实，有被告出具的借条、银行电子回单等证据，被告未提出抗辩，本院依法予以认定。.....</p> <p>.....</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 9 15 《最高人民法院关于民事诉讼证据若干规定》第二条,《中华人民共和国民事诉讼法》第十三条、第一百四十四条</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 16 16 规定，判决如下：一、被告徐军本判决发生法律效力之日起十日内归还原告江西远大保险设备集团有限公司借款 100000 元及利息(自 2013 年 5 月 20 日起至还清全部借款之日止的利息，按年利率 24% 计算)；</p> <p>.....</p>

数据表	数据量(篇)	数据样例及字段说明
		<p>0000051a-11a4-11e8-9a7d-172214d14ea3 10 20 如果未按本判决指定的期间履行给付金钱义务，应当依照《中华人民共和国民事诉讼法》第二百五十三条之规定，加倍支付迟延履行期间的债务利息。</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 10 21 案件受理费 8257 元，由被告徐军负担。</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 11 22 如不服本判决，可在判决书送达之日起十五日内，向本院递交上诉状，并按对方当事人或者代表人的人数提出副本，上诉于宜春市中级人民法院。</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 12 23 审判长 肖青平</p> <p>.....</p> <p>0000051a-11a4-11e8-9a7d-172214d14ea3 12 27 书记员 熊文霞</p> <p>字段说明：</p> <p>第 1 列：案例唯一编号(keyid)</p> <p>第 2 列：分段标记编号(tid)</p> <p>第 3 列：案例自然段序号(orderid)</p> <p>第 4 列：文书自然段正文(section)</p> <p>注：txt 包含两列字段，字段之间分隔符"tab"(Python/Java 语言中\t)</p>
民间借贷纠纷-争议焦点.txt	766916	<p>数据样例：</p> <p>000022fe-1517-11e8-b05b-2bc0567c4793 本院认为，本案争议的焦点是：1.本案被告王春利是否应当对被告庞瑞芝的借款承担连带责任。根据原告提交的上述借条，被告是以“保证人”身份在该借条上签字确认，该借条上对于被告王春利的保证方式没有约定，故王春利应当按照连带责任保证承担保证责任。本院认为根据当事人的陈述及相关证据，能够认定被告庞瑞芝向原告魏锁安借款 15 万元的事实，现约定的借款时间已经届满，被告理应返还原告。原告主张被告王春利对上述给付义务承担连带责任于法有据，本院予以支持。@hc 综上所述，对于原告要求被告庞瑞芝返还借款 15 万元，被告王春利承担连带责任的诉讼请求，本院予以支持。</p> <p>字段说明：</p> <p>第 1 列：案例唯一编号</p> <p>第 2 列：案例争议焦点文本(正文中@hc 表示自然段分段符)</p>

分段标记编号(tid)详细说明			
分段标记编号(tid)	分段标记名称	文书审理程序（一审=1、二审=2、再审=3、 其它非判决文书=300）	排序字段
1	判决书名称	1	1
14	法院名称	1	2
15	文书类型	1	3
2	案号	1	4
3	当事人信息	1	5
4	案件审理信息	1	6
5	原告诉称	1	7
6	被告辩称	1	8
13	第三人述称	1	9
17	证据	1	10
7	法院认定事实	1	11
8	法院说理部分	1	12
9	裁判依据	1	13
16	裁判结果	1	14
10	案件受理费	1	15
11	上诉法院	1	16
12	文书尾部	1	17
101	判决书名称	2	1
119	法院名称	2	2
120	文书类型	2	3

102	案号	2	4
103	当事人信息	2	5
104	审理经过	2	6
122	原审原告诉称	2	7
123	原审被告辩称	2	8
128	原审第三人	2	9
105	一审法院查明事实	2	10
106	一审法院说理部分	2	11
107	一审法院裁判依据	2	12
118	一审裁判部分	2	13
108	上诉人诉称	2	14
109	被上述人辩称	2	15
129	二审第三人	2	16
124	二审证据	2	17
110	本院查明	2	18
111	本院认为	2	19
112	二审裁判依据	2	20
121	二审裁判结果	2	21
114	案件受理费	2	22
115	判决效力	2	23
117	文书尾部	2	24
201	判决书名称	3	1
219	法院名称	3	2

220	文书类型	3	3
202	案号	3	4
203	当事人信息	3	5
204	审理信息	3	6
222	一审原告诉称	3	7
223	一审被告辩称	3	8
205	一审法院查明	3	9
206	一审法院说理	3	10
207	一审裁判依据	3	11
218	一审裁判结果	3	12
208	二审上诉人诉称	3	13
209	二审被上诉人辩称	3	14
224	二审证据	3	15
210	二审法院查明	3	16
211	二审法院说理	3	17
212	二审裁判依据	3	18
221	二审裁判结果	3	19
229	再审申请人诉称	3	20
230	申诉理由（抗诉再审用）	3	21
231	被申请人辩称	3	22
225	再审法院事实	3	23
226	再审法院说理	3	24
227	再审裁判依据	3	25

228	再审裁判结果	3	26
214	案件受理费	3	27
215	上诉判决效力	3	28
217	落款	3	29
301	标题	300	1
314	法院名称	300	2
315	文书类型	300	3
302	案号	300	4
303	当事人信息	300	5
304	审理经过	300	6
312	文书尾部	300	7

2.11 金融问答数据

金融问答数据(数据库: finance_qa)			
数据表	数据来源	数据量 (条)	数据样例及字段说明
shse_qa_latest	上交所“最新答复” (http://sns.sseinfo.com/qa.do)	175118	数据样例: id: 9 timestamp: 2018-04-21 15:09:06 user_name: Naruto stock_name: 千禾味业 stock_code: 603027 question_time: 04 月 09 日 08:54 question_content: 如果中美贸易战真正开打, 大豆价格上升, 对贵公司的业绩有何影响? 公司将采取何种措施应对? answer_time: 昨天 17:24 answer_content: 尊敬的投资者您好, 感谢您对公司的关注。针对贸易战导致的可能的原料价格波动, 公司已强化供应链管理, 持续开发与公司共同成长的战略合作伙伴作为长期供应商, 采取战略性备料方式, 依订单需求分批进料, 降低采购成本。 字段说明: id: 数据编号 timestamp: 时间戳 user_name: 提问者用户名 stock_name: 股票名称 stock_code: 股票代码 question_time: 提问时间 question_content: 提问内容 answer_time: 回答时间 answer_content: 回答内容
shse_qa_hot	上交所“热门回答” (http://sns.sseinfo.com/qa.do)	1000	
shse_qa_hotpush	上交所“热推问答” (http://sns.sseinfo.com/qa.do)	671	
szse_qa_latest	深交所“最新答复” (http://irm.cninfo.com.cn/ircs/sse/sseSubIndex.do?)	500	
szse_qa_best	上交所“最满意答复” (http://irm.cninfo.com.cn/ircs/sse/sseSubIndex.do?)	499	
qjw_qa	全景网“问答” (http://rs.p5w.net/index/company/showQuestionPage.shtml)	38269	

金融问答数据(数据库: finance_qa)			
数据表	数据来源	数据量 (条)	数据样例及字段说明
ths_qa	同花顺“问答” (http://www.iwencai.com/school/ ask?qs=school_answer)	2347	数据样例: id: 9 timestamp: 2018-04-23 03:54:10 question_content: 市净高好还是低好 answer_content: 一般来说, 市净率低的股票投资价值比较高一些, 市净率低意味着投资风险小, 万一上市公司倒闭, 清偿的时候可以收回更多成本。 update_time: 2017-02-09 08:50:38 likes: 68 字段说明: id: 数据编号 timestamp: 时间戳 question_content: 提问内容 answer_content: 回答内容 update_time: 更新时间 likes: 点赞数量
ths_article	同花顺“文章” (http://www.iwencai.com/school/ article?qs=study_article)	1365	数据样例: id: 3 timestamp: 2018-04-23 04:18:17 category1: 股票 category2: 基础知识 url: http://www.iwencai.com/school/article-detail/id/4545 title: 股市什么时候开盘 content: 中国股市一天的开盘时间为: 每日 9: 15 开始参与交易, 11: 30-13: 00 为中午休息时间, 下午 13: 00 再开始, 15: 00 交易结束。

金融问答数据(数据库: finance_qa)			
数据表	数据来源	数据量 (条)	数据样例及字段说明
		 字段说明: id: 数据编号 timestamp: 时间戳 category1: 类别 1 category2: 类别 2 url: 文章链接 title: 文章标题 content: 文章内容