

# 戚志锋

性别：男

电话：13818430422

邮箱：[337850904@qq.com](mailto:337850904@qq.com)

民族：汉

户籍：上海

在职，正在看机会

出生日期：1985.4



## 自我评价

7 年以上团队管理经验，10 年以上搜索/架构的实施经验。利用 elasticsearch、springboot、kafka、hbase、flink、pytorch、tornado 等主流框架，负责搜索中台的架构和算法从 0-1 的建设和持续优化工作和平台化建设。持续 深度挖掘业务，与业务方展开合作，提供业务数据指导和高可用的平台化建设。把握行业技术发展方向，结合 NLP 算法、个性化排序算法，充分理解和挖掘用户，并持续为平台的每一个流量带来价值增长。

主要开发语言 java、python

熟悉 elasticsearch、solr、springboot、redis、canal 等基础框架

熟悉 mysql、mongodb、redis、hbase

熟练使用 spark、flink、kafka 、hive 等大数据处理框架

熟悉 pytorch 框架、相关 word2vec、albert、xgboost、lightgbm、DNN 等模型



## 工作经历

### 泛微网络

2022.02 — 至今

技术部 · 搜索模块负责人

OA办公系统

下属人数：5

汇报对象：技术总监

### 公司简介

泛微网络股份有限公司是国内OA（办公自动化）行业唯一一家上交所主板上市公司，旗下有诸多的产品可供用户选择，为用户提供了智能、协同、信创的全程数字化办公解决方案，帮助客户驱动组织数字化转型步伐，提升用户数字化办公水平

### 工作职责

- 新一代ecology10搜索相关的产品、技术研发
- 上一代ecology9的搜索产品的迭代、维护、优化
- 协助完善、解决各模块海量数据下的各种搜索场景
- 新产品功能开发，智能问答库、图像（以图搜图）、文档附件、复杂权限等
- 智能化升级，结合NLP、推荐、智能问答提供搜索准确率
- 负责搜索业务线的日常需求对接与开发
- 团队建设与迭代计划制定

项目和业绩

◆ 搜索产品模块

负责全面产品化设计搜索模块，以灵活支持本系统内部数据及外部数据的配置化检索需求。该产品化功能涵盖搜索页面的用户界面设计、搜索功能的详细规划、严格的权限管理体系、全面的数据安全保护、便捷的搜索管理后台、实时的搜索监控平台，以及个性化的扩展能力。此外，我们还实现了整个产品的一体化安装流程，以确保用户能够轻松部署并快速投入使用。

◆ 附件解析系统

该系统具备对PDF、Word、PPT、Excel等多种格式附件内容的深度解析能力，并且还支持OCR技术以解析图像中的文本信息。经过精细处理后的文本内容，将被无缝整合进搜索系统中，从而实现高效、准确的文本搜索功能。

◆ 图像搜索

我们融合了Elasticsearch与CLIP及OCR技术，以及Milvus搭配CLIP与COR的架构，精心打造了以图搜图的高效功能。此功能全面适配不同规模的数据集，让用户能够凭借文字描述或图像样本，迅速精准地定位到所需的文档数据。

◆ 知识库检索

在用户自定义的知识库中，我们对检索功能进行了全面升级。针对用户输入的query，我们实现了技术改写，涵盖文本关键字的有效提取、query的标准化处理（归一）、智能纠错以及query的扩展等多样功能。这些改进确保了无论是长文件还是短文件的搜索，系统都能应对自如。

◆ 知识推荐系统

采用ner模型+用户行为数据的预测模型，对用户本次搜索的意图加以识别，通过elasticsearch进行多模块的数据召回，并结合规则+推荐模型进行数据推荐

◆ 搜索系统管理平台

该后台系统对搜索集群的全方位管理进行了整合，包括索引管理、日志监控、安全保障、报警响应以及个性化设置等核心模块。它依据功能特性精心设计了多个可视化的管理界面，借助直观的图表展示、详尽的列表数据以及便捷的操作功能，确保了搜索系统的日常运维高效且顺畅。此外，系统还赋予了客户高度的灵活性，允许他们根据自身的特定搜索场景，灵活配置搜索条件，并轻松添加自定义字段与搜索规则，从而精准满足其个性化的搜索需求。

识货

技术部 · 搜索技术专家

下属人数：9

2019.12 — 至今

电商

汇报对象：CTO

公司简介

识货是虎扑体育旗下的导购 app，专注于鞋、服、美妆等垂直类目，整合全网商品资源，为用户提供价廉物美的正品商品。识货的运动装备建立了专业的正品审核机制，为年轻人提供了专业的网购决策指导。

工作职责

- 升级和建设搜索架构体系，提供召回+排序，持续迭代升级优化搜索框架，并保证系统稳定。
- 分析数据，结合算法模型提升搜索体验、提升搜索的转化率
- 内部业务场景挖掘，拓宽业务场景，持续提升搜索的转化率
- 搜索数据分析，为运营提供指导性运营策略和挖掘新优化点
- 提供 NLP、图像识别能力，持续挖掘商品的价值
- 负责搜索业务线的日常需求对接与开发
- 团队建设与迭代计划制定

## 项目和业绩

### ◆ 大搜索系统

承接识货 app 的主搜、社区、类目导航等场景。提供基于 ES+springboot、faiss 等多路召回系统，并提供个性化和非个性化排序。系统经 618、双 11 考验，稳定运行连续 4 季度 0 故障。

搜索相关性上 basecase 大幅降低。整体单次查询相应小于 200ms。至今为止整体搜索转化率提升 71%，大盘占比提升绝对值 7%，UV 价值提升 89%

### ◆ 搜索原子词库系统

提供行业原子词库，包括同义词、纠错词、违禁词、词性等自动入库、人工干预修改功能。系统上线后，同义词相对增加 845%，纠错词相对增加 1290%，词库扩大 8 倍。

### ◆ 搜索意图识别

提供基于 albert 的 ner 模型+用户行为数据的预测模型，对用户搜索 query 的类目、品牌的意图预测。用于筛选导航展示策略判断、卡片格式等场景。目前类目准确率达 94.8%、品牌准确率 99%、精确词准确率 97%。

### ◆ 搜索运营后台

支持词库维护、搜索词跳转落地页、搜索词下的排行榜/广告/商品等运营，个别/批量商品提权等功能。

### ◆ 搜索预警系统

搜索下商品的排序实时检测，搜索次流量实时变化，商品 ctr 暴涨等异常监控预警。可实时提供给运营并指导其进行补货、商品提权等方案指导。

### ◆ 流量分发系统

根据业务规则和场景提供不同的路由功能，如列表展示样式、卡片样式、banner 位、商品信息补足，坑位调整、提权策略、ABtest、等自动化功能。

### ◆ 新品赛马服务

长尾商品的扶持策略，针对不同的新品池，提供不同的曝光梯度设置。系统上线后，新品曝光占比提升 265%，列表 gmv 提升 4%。

### ◆ 搜索多路召回系统

提供基于 ES、向量、意图、关键词降级、i2i、商品系列等多维度结果召回，并结合粗排+业务策略进行筛选，整体召回率显著，badcase 大幅度降低。uvctr 提升绝对值 4%。

### ◆ 搜索个性化排序系统

基于 lightbgm、dnn 的 ECPM 等多个模型的个性化排序模型，实现按用户时间的排序回放/T+1 自动化训练，T+1 自动化部署等功能。整体搜索人均 UV 价值提升 10+%，uvctr 提升绝对值 3%。

## ◆ 搜索平台系统

提供给予 ES 的 web 可配置化系统，日常后台搜索功能在页面配置即可实现搜索功能，0 代码开发。整体开

发周期大幅提升，1 个需求 1 个工作日即可交付测试。

#### ◆ 识货口碑评论系统

对评论进行情感分析，并进行正、负向标签分类，辅助用户进行购买决策。并对重复图片、广告文本、广告图片等进行检测，并提供不同的排序策略，大幅提升用户体验。

#### ◆ 文本知识图谱

基于用户搜索行为信息和商品文本信息，挖掘关键词之间的关联关系。并应用于商品集标题自动生成，商品集生成，同义词挖掘等领域。辅助提升搜索的转化率和点击率。

#### ◆ 新词发现系统

基于互信息和左右熵的新词发现系统，对平台新商品标题进行自动化新词挖掘，并建立有效的审核机制。

相关技术：java、python、elasticsearch、springboot、mybatis、hbase、flink、lightbpm、dnn、faiss 等

## 淘集集

2018.09 — 2019.12

技术部 · 搜推技术专家

电商

下属人数：8

汇报对象：CTO

### 公司简介

淘集集是一个社交电商 app，专注于电商下沉市场，类目几乎覆盖全网类目，台上一线城市用户占比 4.46%，二线城市用户占比 32.95%，三四线城市及县乡村用户占比 62.68%。用户增长迅速，日活高峰达到 300W，在下沉电商市场排行第二。

### 工作职责

- 从 0-1 构建升级整体搜索架构和推荐架构，并日常维护并保持其稳定
- 负责淘集集 APP 的搜索推荐整个业务线业务的开发
- 理解商品、用户，深度挖掘搜推场景，持续提升转化率
- 负责算法的研发和落地，为其他团队提供算法支持
- 团队日常管理与计划制定

### 项目和业绩

#### ◆ 搜索和推荐的架构体系

从 0 到 1 建设搜索/推荐的架构系统并参与 app 端埋点规范，实现数据实时和 t+1 自动化入库。并增加基于 itemcf/faiss 的多路召回功能，精排实现个性化和非个性化排序。整体搜索搜索 CTR 相对提升 34%，转化率提升 89%，推荐 CTR 相对提升 45%，转化率提升 23%

#### ◆ 特征工程

基于用户行为，挖掘商品文本信息，建设商品/用户的特征体系，实现 t+1 的静态特征和基于 flink 的实时特征。

#### ◆ 词库建设、文本敏感词库建设

通过对平台商品文本的挖掘和三方相关服务的信息整合，建立自由的违禁词和词库系统，提高搜索精度。

- ◆ 基于 bi-lstm-crf 的命名实体识别模型，识别商品标题中的性别、材质、品类、品牌等此行信息。该模型为意图模型和分词提供服务。
- ◆ 商品比价平台，使用爬虫抓取相关友商的公开信息。历史 NLP 识别相似商品并进行价格的比对，为运营提供定价策略。
- ◆ 基于 lr 粗排模型和基于 LTR 的精排模型开发，实现 t+1 自动训练和部署
- ◆ 搜索推荐数据决策平台。提供各业务线定制和非定制的数据报表，辅助业务方进行决策。
- ◆ 基于分桶策略的多路 ABTest 框架

对流量进行不同的划分，以用户体验影响最小的代价，结合数据决策平台，对平台上线的不同方案进行快速验证，根据结果对方案进行收益最大化评估，

#### ◆ 搜索平台系统

业务方自动化配置平台，对复杂的后台查询 sql 进行数据扁平化操作，结合 canal 等工具进行近实时更新。并同时提供原生 ES 语法、类 mysql 语法等功能。

#### ◆ 文本知识图谱

对文本信息建立图数据库。从关系网络中，辅助从中挖掘出相关的商品特征、用户特征、近义词等信息。

相关技术：java、python、elasticsearch、springboot、mybatis、hbase、flink、kafka、deepFM、itemCF 等

## 平安好医生

2016.03 — 2018.09

技术部 · 高级搜索架构师

互联网医疗

下属人数：3

汇报对象：技术中台总监

### 公司简介

平安健康是平安健康医疗科技有限公司旗下的 APP，平安好医生，集“家庭医生、名医问诊、健康社区、健康评测、健康习惯、健康档案”六大特色服务于一体，为用户提供一站式健康咨询及健康管理服务。

2014 年上线，2018 年在港股 IPO。现 app 更名为平安健康。

### 工作职责

- 负责公司所有 app 端、web 端等业务搜索入口。
- 负责平安好医生所有业务搜索，针对医疗线、商城线的业务需求提供搜索解决方案。
- 搭建基于 solr、elasticsearch 的搜索集群，并对其进行日常的运维工作
- 针对不同的业务场景，提供不同的推荐算法，使用技术手段提供搜索转化率

### 项目和业绩

- ◆ 对公司所有产品线的搜索框架进行重构开发，100%覆盖和解决所有历史业务遗留问题，使日后的搜索业务能稳定健康发展。
- ◆ 建立针对不同业务线的分类标签系统，对医疗、商城、头条等数据进行清洗，解决标签过多、数据不准等问
- ◆ 基于 LR 的排序模型研发，医疗线转化率相对提升 10%，商城转化率相对提升 16%。
- ◆ 针对日常的搜索数据进行挖掘，建立基于搜索的用户推荐系统、并把该策略同时应用于提示词、热门推荐词

等业务场景。

- ◆ 建立一个完整的搜索平台，使用户能够可视化的接入自己的搜索库
- ◆ 建立搜索监控平台，近实时查看集群健康、各业务线接口的搜索情况。
- ◆ 解决实时地理位置搜索和多边形区域搜索

相关技术：java、python、elasticsearch、solr、springboot、mybatis、kafka、dubbo、hbase、flink、相关机器学习算法

## 优酷土豆

2014.07 — 2016.03

技术部 · 高级搜索工程师

下属人数：3

汇报对象：技术经理

### 公司简介

优酷和土豆两大视频网站合并后，合称优酷土豆。产品已覆盖 PC、电视、移动三大终端，兼具影视、综艺、资讯三大内容形态，贯通视频内容制作、播出、发行三大环节，成为真正意义的互联网电视媒体，引领视频行业格局及全媒时代的大格局。

### 工作职责

- 搭建基于 solr 的搜索引擎，对其他部门的搜索业务进行支持。升级并维护土豆之前的搜索集群。
- 将优土的分散数据进行整合统一由搜索引擎提供。
- 对新词库新词的扩展、维护。
- 并对现在有视频打标、用户打标、建立视频与用户、用户与用户之间的关系。
- 完成其他的开发工作。

### 项目和业绩

- ◆ 搭建基于 solr 的搜索引擎集群，对原 lucene 的架构进行升级更新替换，性能相对提升 34%
- ◆ 搭建 kafka 的离线消息队列系统，对接搜索集群，实现近实时搜索。
- ◆ 建立统一搜索平台，可对现有的集群进行监控、接口测试、数据展示。
- ◆ 筛选用户、视频优质数据到不同的搜索集群，以提供产品团队参考。
- ◆ 建立新词库体系，对原视频标引进行统一整改。
- ◆ 对旧词库进行改造，建立词分类属性体系，建立词与词之间关系、视频与视频、视频与用户、用户与用户之间的关系，并对视频的质量进行打分，搜索准确率相对提升 4.7%。
- ◆ 首次使用 LR 模型改进排序结果，有效播放率相对提升 6.3%
- ◆ 所负责搜索接口提供土豆 40%的数据出口。
- ◆ 视频指纹中间层开发，视频重复度检测准确率提升至 94%，相对提升 2.4%，性能提升 8%

相关技术：java、solr、kafka、springMVC、ibatis、hadoop、hive 等



## 公司简介

丁丁网专注于本地生活，解决城市居民的生活问题，定位于本地生活搜索专家。公司自主研发位置引擎和搜索技术。主要的产品有丁丁地图、丁丁优惠、丁丁生活、丁丁签到、券酷等产品。

## 工作职责

- 保证搜索引擎正常运行，并进行技术框架的改进、性能的优化。对新需求进行支持开发。
- 搜索词库建设、业务分词规则改写，增加排序策略，转化率相对提升 13%。
- 对大数据统计团队提供支持。
- 对新技术的研究，并保证在正式环境中稳定运行。
- 培训、指导新员工使用新技术。
- 内部系统的 UI 整体设计与开发。
- 文档的制定和技术规范的编写。
- 完成部门经理交付的其他任务。

## 项目和业绩

- ◆ 升级 solr 搜索引擎项目版本，并对索引机器进行迁移，并全新开发 cmcg 系统的搜索引擎，搜索整体性能相对提升 25%，qps360/台，各项服务稳定运行，无重大事故。
- ◆ 自研分词器开发、词库维护
- ◆ 改进搜索的策略排序，转化率相对提升 13%
- ◆ 独立定制新的 Android、IOS 系统的消息推送平台，保证推送的稳定、并提供相关的推送统计信息。并支持 CRM 平台也接入推送服务。
- ◆ 搭建新的 mongodb 数据库用于日志的存储，并独立创建了 idss 用户行为分析系统。
- ◆ 独立研发性能监控平台 iMonitor、并提供被监控应用相关的优化解决方案。
- ◆ 重构了原数据统计后台的 UI 框架、性能优化、加入动态统计图形。
- ◆ 负责研究部的后续其他项目的 UI 框架架设。

相关技术：java、solr、kafka、springMVC、extjs、jsp、redis、ibatis 等

## 上海携宁计算机科技有限公司

**2010.03 — 2012.08**

技术部 · java 开发工程师

下属人数：0

汇报对象：技术主管

## 公司简介

是一家专门从事金融领域 IT 服务的高新技术软件企业，主要为基金公司提供三方定制软件服务，如报表自动化、股市预测、自动化 workflow 后台等产品。



工作职责

- 参与公司新框架、产品模块的设计与开发。
- 编写相关的开发文档，数据库设计文档，产品使用手册等。
- 指导、培训初级程序员解决技术难题。
- 现场开发时，与客户业务的沟通和现场项目实施。
- 每日整理项目进度、项目需求、开发文档的更新。

项目和业绩

- ◆ 完成中金阳光私募、易方达场外系统重要项目
- ◆ 基于 microsoft word 的自动化报告系统
- ◆ 基于不同的数据自动化绘制数据图形（折线图、雷达图、柱状图等）
- ◆ 基于 droolsflow 的可配置规则引擎并在公司推广该模块

相关技术： java、ssh、extjs、jsp、jfreechar、oracle 等



教育经历

同济大学（非全日制）	本科	行政管理	2012.09 — 2014.07
上海大学（全日制）	专科	应用电子	2004.09 — 2007.07