

# 个人情况

---

姓名：董一峰

性别：男

籍贯：河南省安阳市

联系方式：17090079753

邮箱：[519100798@qq.com](mailto:519100798@qq.com)

个人博客：<https://dongyifeng.github.io/>

# 求职意向

---

岗位：Java 研发工程师（最好是搜索工程方向）

期望薪水：面议

到岗时间：随时

# 技能描述

---

1. 熟悉数据结构与算法，对跳表、前缀树、AC 自动机、动态规划、树形 DP 等都有深入了解，并在项目中使用前缀树实现搜索引擎关键字智能提示系统。
2. 熟悉 Java 核心知识，多年一线研发经验，具备良好的编程能力，并熟练使用设计模式。
3. 熟悉 Java 并发编程，对 Java 的各种锁机制、线程池机制、AQS 等都有深入了解，并在项目中熟练使用。
4. 熟悉 JVM，对 GC 算法、垃圾回收器都有深入的理解，并且参与过线上项目的 JVM 调优工作。
5. 熟悉 spring、springmvc、mybatis、sprintboot 等核心框架，可以根据需求快速搭建项目，并阅读过核心源码（看过循环依赖的核心代码、bean 的生命周期、springboot 自动装配原理、内嵌 tomcat 源码、springboot 启动流程）。
6. 熟悉 springcloud、sprintcloud aibaba、dubbo 等微服务框架、堆 nacos、sentinel 等组件有深入理解，堆服务划分、服务治理、服务分层等都有深入理解，并具备线上项目经验。
7. 熟悉常用的分布式解决方案：分布式 ID、分布式锁、分布式事务。
8. 熟悉 Mysql，对 Mysql 的锁、事务、索引等都有深入研究，并参与过线上项目的调优工作。
9. 熟悉 Zookeeper 底层原理和实现机制，了解 paxos、zab 算法，对 CAP 有深入了解。
10. 熟悉 Redis，了解持久化机制、主从同步机制、Redis 事务，在工作中能熟练使用。
11. 熟悉 Kafka，了解核心原理，对底层刷盘机制、集群分片机制都有深入理解，并对消息丢失和重复消费有对应的上线解决方案。
12. 熟悉 Elasticsearch，了解索引的写流程、更新流程、Search 流程、master 选举机制，在工作中开发过自定义的分词插件。
13. 了解朴素贝叶斯算法、Logistic Regression、K-Means 等经典机器学习算法，在工作中使用过 LR 对 sug 服务进行模型排序，使用 K-Means 对 Query 做过聚类分析。

# 工作经历

公司名称	职位	时间
雪球财经	Java 资深研发工程师	2014 年 4 月至 2022 年 11 月
汽车之家	C# 研发工程师	2011 年 10 月至 2014 年 4 月
北大法意	C# 研发工程师	2010 年 6 月至 2011 年 10 月

# 个人评价

12+ 年的一线开发经验，其中 3 年以上大型系统架构设计经验，具有分布式、高并发、高可用的系统机构设计以及开发经验。在雪球负责注册用户 6 千万，135万日活的搜索系统的架构和研发。在雪球搜索引擎关键字智能提示系统中，创新性提出**加权 Trie 树**和**下探剪枝策略**，解决系统性能问题。对朴素贝叶斯算法、**Logistic Regression**、**K-Means** 等经典机器学习算法有一定的了解。平时喜欢研究数据结构和算法。热爱工作，沟通能力强，责任心强。

# 项目

项目名称：雪球搜索引擎关键字智能提示系统（sug 服务）

项目描述：搜索关键字智能提示是一个搜索应用的标配，主要作用是避免用户输入错误的搜索词，并将用户引导到相应的关键词上，以提升用户使用体验。雪球 sug 服务基于**定制化加权 Trie 树 + 模型排序** 实现，并创新性设计出**加权 Trie 树下探时的剪枝策略**。sug 服务抽象出索引数据，并与业务数据进行隔离，使得 sug 基础服务（Trie 树检索部分）非常稳定。该系统服务于雪球千万级别用户，工作日 QPS 峰值为 1100，耗时 P99 在 1ms 以内。分享地址：<https://mp.weixin.qq.com/s/GCX3mPJ2giUlvIJCpByF4Q>

技术架构：加权 Trie 树、Logistic Regression、Redis、MyBatis、Zookeeper、Kafka、RabbitMQ、FTP、Pinyin4j、HanLP 等。

职责描述：

- 1. 参与核心模块需求评审，架构规划，设计以及技术选型，系统重构，性能优化，攻克系统技术难题，构建高可用、高性能、高并发、高扩展的系统架构。
- 2. 负责加权 Trie 树的架构设计和开发。
- 3. 通过加权和剪枝策略攻克 Trie 树前缀树下探性能问题，提高系统的稳定性。
- 4. 负责设计 Trie 树的数据抽象，抽象出检索数据和业务数据并进行隔离，使得 sug 基础服务非常稳定，不收业务影响。
- 5. 负责系统高并发多级缓存架构的设计和开发。
- 6. 参与平台分布式实时数据同步中间件的架构设计和开发。
- 7. 负责 sug 服务效果提升，选用 Logistic Regression 模型打分排序、设计特征、训练模型、模型上线及效果跟踪。**第一位点击率相对提升 35%**。
- 8. 负责拼音、简拼、数字、多音字、纠错搜索的设计开发。
- 9. 负责设计并开发 Trie 树维护系统，接入业务的 MQ 消息，更新 sug 服务中的 Trie 树的数据。

10. 负责设计 Trie 树的数据持久化方案，采用快照 + AOF 混合持久化的方案，提高加载效率的同时保全了数据的完整性。
11. 负责股票关键词维护、清洗，设计关键词的挖掘方案。
12. 负责推广 sug 服务为其他业务系统赋能：在 Query 理解项目中，用于识别 Query 中的股票。在索引系统中，分析帖子关联股票。
13. 负责维护在 sug 在线服务 Trie 树数据正确性，对线上 Trie 树的数据进行检查和修复。
14. 设计并开了 sug 服务效果对比系统，为新功能上线保驾护航。

## 项目名称：雪球索引分布式重建平台

**项目描述：**索引是搜索引擎的基石。在索引新建字段时，在算法调整时，在股票 symbol 变更时，在索引数据出现错误时...，都需要大批量重建 Elasticsearch 索引。因此我们设计开发了索引重建平台。该平台具备高可用、高性能、高吞吐量的特性。该平台包含任务管理和任务执行两个模块，任务管理模块接收到请求，会将大任务切分成小任务，然后分发给多台机器去执行。平台包括：任务分发、断点续传、失败重试等核心功能。

**技术架构：**Elasticsearch、Redis、MyBatis、HDFS、ThreadPool、分布式锁

### 职责描述：

1. 参与核心模块需求评审，架构规划，设计以及技术选型，性能优化，攻克系统技术难题，构建高可用、高性能、高并发的系统架构。
2. 负责设计分布式架构，以及任务分发方案。接受请求的 worker，将任务切分后放入 Redis，所有 worker 从 redis 中争抢任务。
3. 负责设计并开发任务管理模块，该模块包括：任务启停、任务分发、断点续传、失败记录重试等核心功能。
4. 负责设计并开发任务执行模块，该模块包括：任务解析、数据查询、控速、创建索引等核心功能。
5. 负责设计代码结构，采用策略模式，增强了代码的扩展性、复用性、可读性。
6. 负责实现索引重建策略：根据时间范围修复，根据股票修复，根据作者修复，根据关键词修复，从 HDFS 上拉去分词结果修复等。

## 项目名称：雪球搜索 Query 聚类分析

**项目描述：**在搜索系统中要了解用户意图、制定优化方向，都需要用到 Query 聚类分析。通过定制化 K-means + Word2vec 的方案，进行 Query 聚类。重新定义 Query 与 Query 的相似度，使用的特征有：Word2vec 计算词与词的相似度、词性、核心词、动宾结构、句法结构等。创新性提出使用实质心来计算聚类结果并更新实质心。分享地址：<https://mp.weixin.qq.com/s/8tl59REdytjW1JuKymcoWg>

**技术架构：**K-means、Word2vec、HanLP、Sug、Python

### 职责描述：

1. 参与需求评审，设计以及算法选型，攻克系统技术难题。
2. 负责收集训练 Word2vec 模型的语料，包括：雪球帖子数据：1.3G，维基百科数据：1.2 G，智齿系统用户反馈问题数据：544 条。
3. 负责开发加载 Word2vec 模型的程序，并可以计算词与词的相似度的方法。
4. 负责收集训练 K-means 数据，通过 HanLP 进行分词，获得单词、词性、依存句法分析、股票分析等。
5. 负责开发 K-means 模型训练的代码，优化初始质心的选择，初始质心尽可能远。
6. 负责设计 K-means 距离公式方案，将 K-means 使用的欧式距离改为用 Word2vec 计算 Query 与 Query 的相似度。自定义的相似度包含：Word2vec 计算词与词的相似度、词性、核心词、动宾结构、句法结构等。
7. 负责设计 K-means 的聚类算法与实质心更新算法的方案。

- 8. 负责设计多轮 K-means 的聚类算法方案。
- 9. 负责使用 K-means 对 Query 聚类的最终效果。

项目名称：雪球预加载系统

项目描述：在雪球搜索场景下，Query 是股票名称占总搜索的 80%，这样的 Query 只有 2.5万个，因此非常适合预加载搜索结果。提升搜索服务性能和稳定性。预加载系统还承担更多的职责：1. 搜索可以尝试更复杂的排序模型，尝试更多特征。2. 当召回服务或者排序服务异常时，预加载以系统可以为搜索兜底。3. 支持更多实验场景：对比雪球搜索结果与百度的搜索结果。

技术架构：Redis、MyBatis

职责描述：

- 1. 负责架构规划，技术选型，系统重构，性能优化，构建高可用、高性能、高并发的系统架构。
- 2. 预加载系统是我发起的，整个项目独自设计开发。
- 3. 采用 RabbitMQ 异步处理加载需要的数据，降低系统的 P99。
- 4. 负责系统高并发多级缓存架构的设计和开发。
- 5. 负责设计预加载系统支持切流上线的方案。

教育背景

学校	学历	专业	毕业时间
郑州电力高等专科学校	大专	计算机应用技术	2010 年 6 月