

CSDN代码搜索系统的研发与技术实践

2016-7-20



目录

- **CSDN代码搜索系统起源**
- 系统架构
- 技术实践
- 未来方向
- CSDN 数据团队介绍

CSDN 致力于满足开发者需求的不断演进

技术认知阶段

学习互动阶段

实践参与阶段



知道



学习



尝试



创新



了解



帮助



使用



软件开发者对代码垂直领域的需求

软件开发初学者

1. 代码学习
2. 标准库使用
3. 算法、设计模式代码实例
4. ...



有经验的软件工程师

1. 功能实现实例
2. 如何解决开发难题
3. 代码移植
4. ...



开源项目贡献者

1. 代码共享
2. 代码收藏
3. ...

当前CSDN 代码搜索系统的数据来源

CSDN站内代码资源

- CSDN是面向IT专业人员的知识分享/学习平台，超过270万人开通博客，超过510万篇原创技术文章。代码资源丰富。
- 超过1180名业界专家博主，代码资源优质。
- CSDN博客专栏、知识库等优质栏目，提供体系完备的代码资源。
- 覆盖多领域的代码资源：移动开发、Java、C、web前端、云计算、数据库等。
- 活跃的开发社区论坛、代码平台提供多渠道的代码资源。



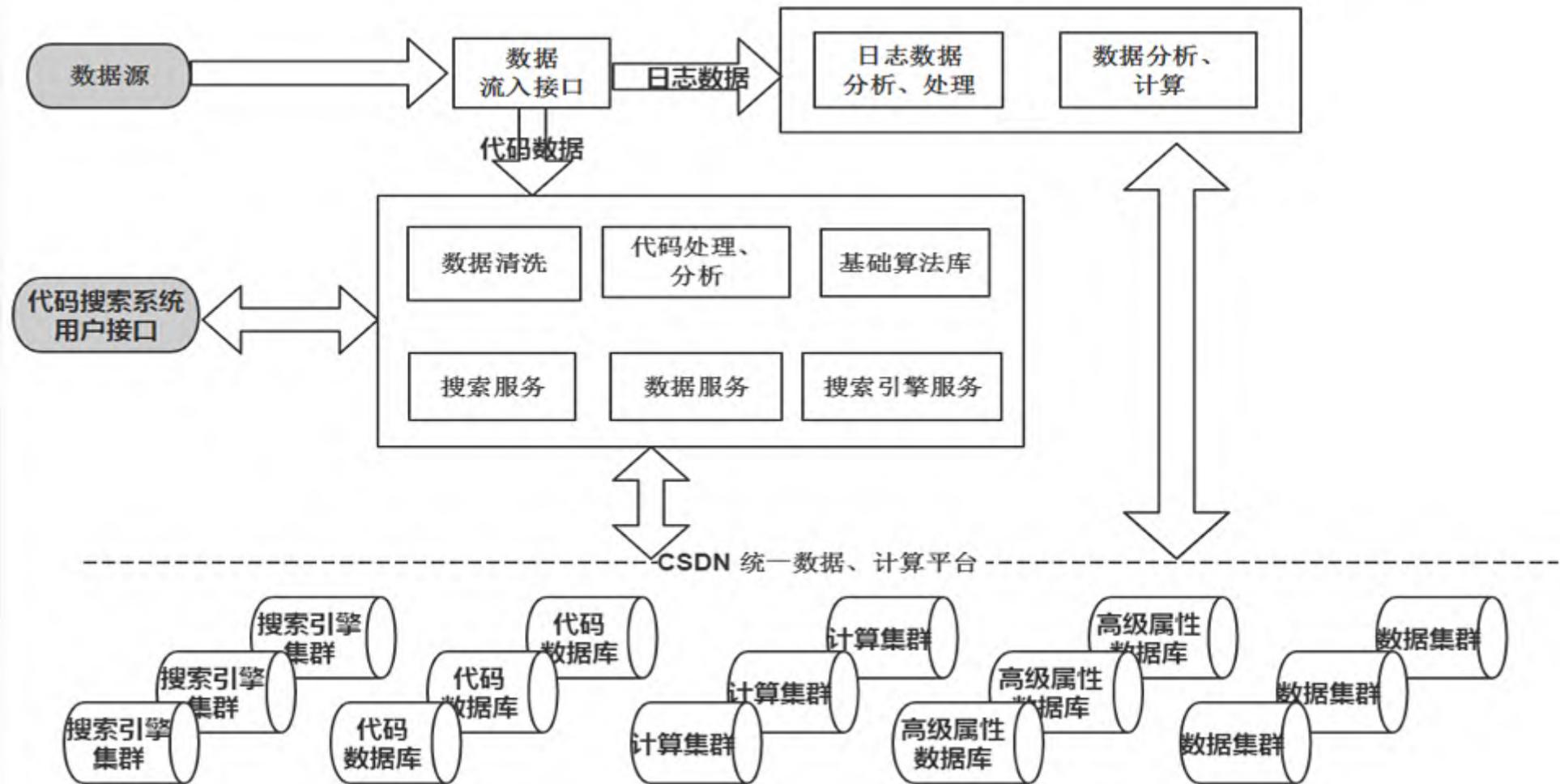
MySQL



目录

- CSDN代码搜索系统起源
- **系统架构**
- 技术实践
- 未来方向
- CSDN 数据团队介绍

系统架构





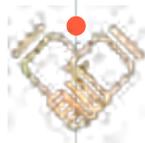
目录

- CSDN代码搜索系统起源
- 系统架构
- **技术实践**
- 未来方向
- CSDN 数据团队介绍

代码搜索效果如何最优？

挖掘你的数据

- 对代码分析的越全面、越准确，越能挖掘代码本身的价值

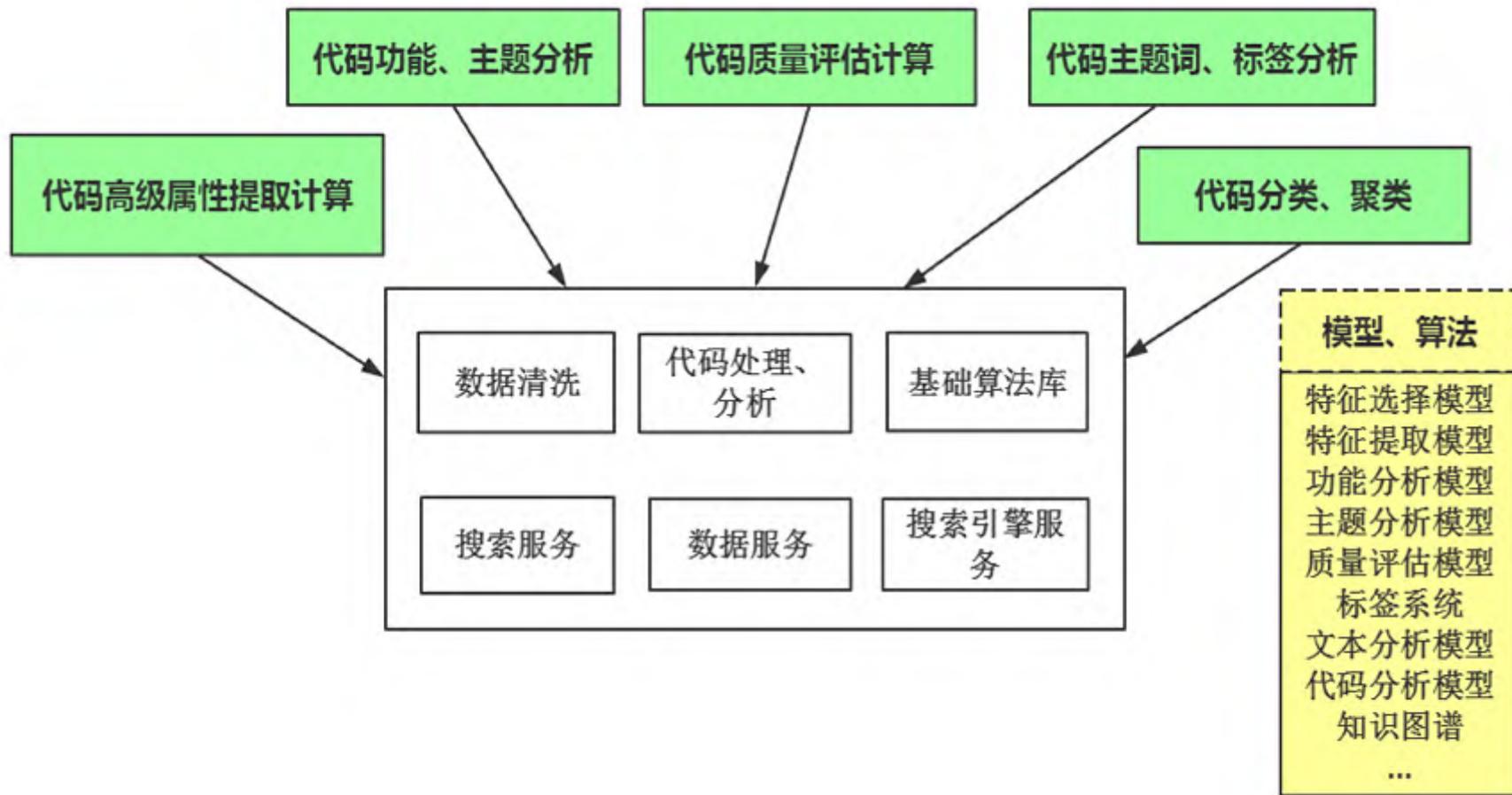


挖掘用户需求

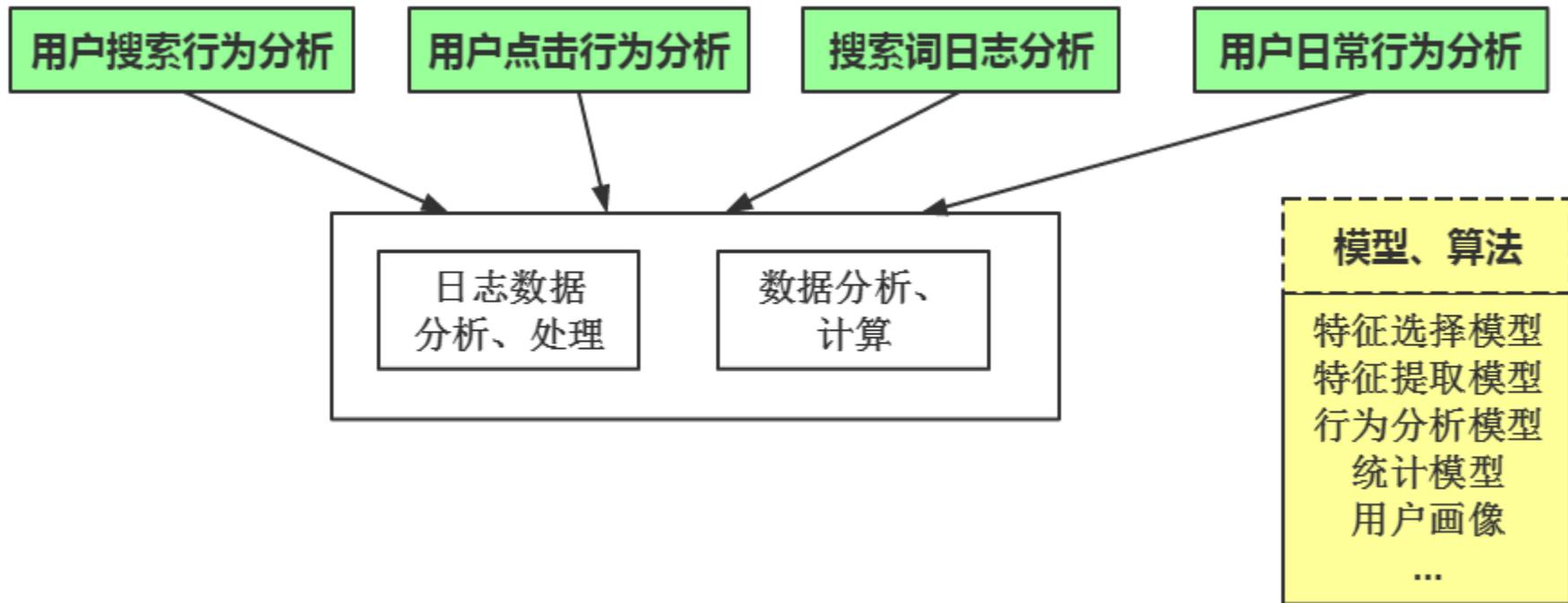
- 对用户需求、用户特征了解越全面、越准确，越能提供用户最满意的搜索结果



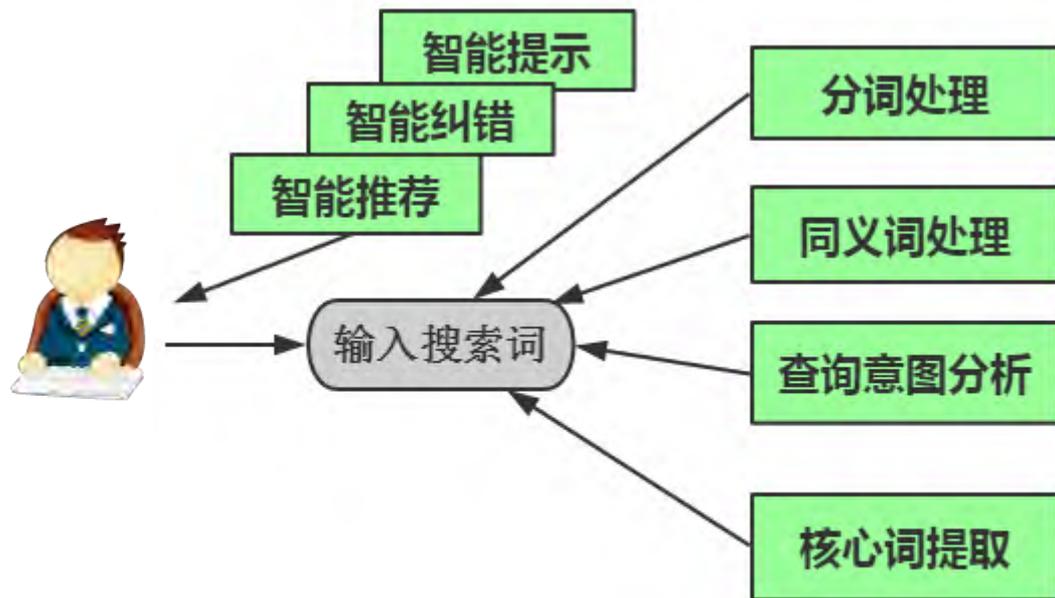
技术实践 – 代码数据处理



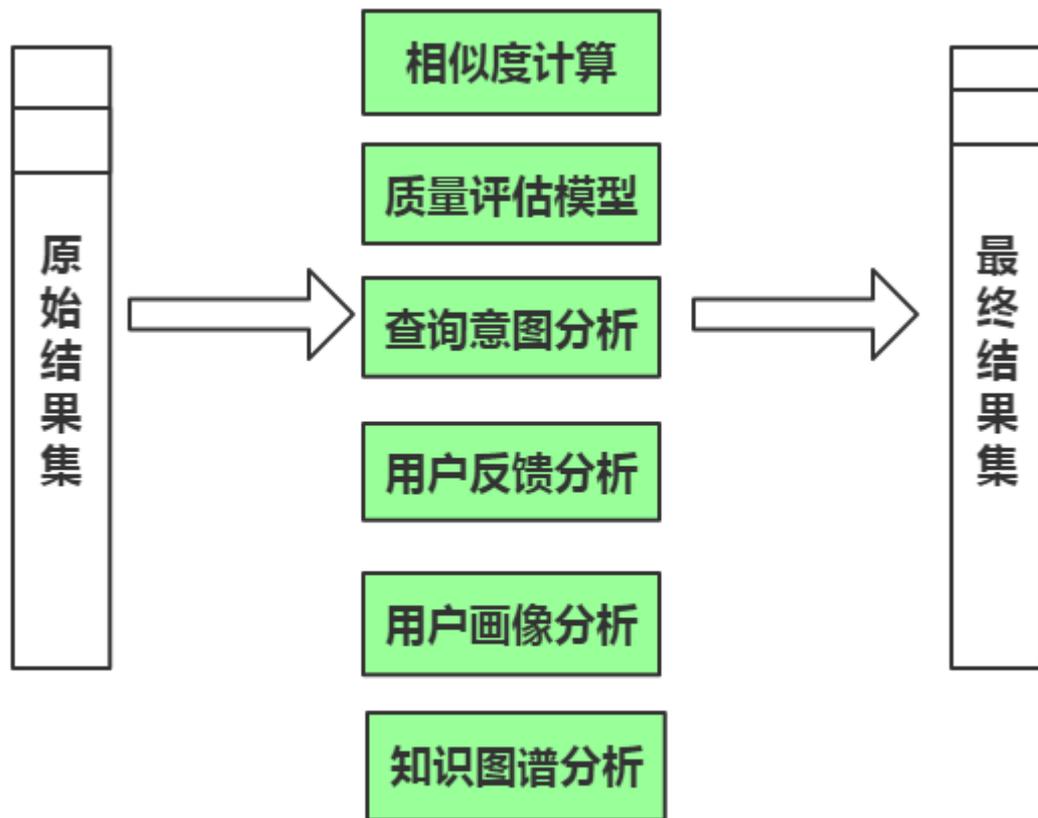
技术实践 – 日志分析



技术实践 - 用户搜索处理



技术实践 – 搜索排序



技术实践 – 效果评估

- 测试集验证准确率、召回率
- DCG评估
- AB test
- 用户日志分析



目录

- CSDN代码搜索系统起源
- 系统架构
- 技术实践
- **未来方向**
- CSDN 数据团队介绍

未来方向



更广数据源：GitHub、StackOverflow ...

更深层挖掘数据：深度学习、语义理解 ...

更深入理解用户需求：个性化搜索、推荐 ...





目录

- CSDN代码搜索系统起源
- 系统架构
- 技术实践
- 未来方向
- **CSDN 数据团队介绍**

CSDN 数据团队画像



行业背景：

- 互联网
- 通信、IT
- 多媒体
- 金融

70%以上硕士学历：

- 清华
- 北邮
- 北交大
- 西电
- 南京航空航天大学

成功商业案例：

- 世界500强公司海量数据分析及可视化项目
- 国内电商平台内容搜索、用户画像、个性化内容推荐&数据分析支撑系统
- 企业级应用系统开发

重点大学实验室、科研机构技术合作、交流

谢谢！



欢迎交流、加入！

chendong@csdn.net

The CSDN logo is rendered in a bold, white, sans-serif font. The letters 'C', 'S', and 'D' are connected, and the 'N' is slightly separated. The background is dark blue with scattered, colorful, feather-like shapes in shades of red, orange, and yellow, and thin white lines with dots.

为中国开发者创造价值

[http:// www.csdn.net](http://www.csdn.net)