

# awesome software engineering beginners

---

## 简介

---

本仓库主要存放了北京大学孙艳春老师领导的基于大数据分析的智能服务研究小组近三年的系列研究工作，主要是智慧城市服务和智慧学习服务的研究。

智慧城市服务主要基于微博数据、POI数据、公共交通数据等各类公开数据，结合多种机器学习模型和数据挖掘方法，从城市区域的功能、交通便利度、人口密度分布、建筑物功能等多个角度来综合性地描述城市地理语义信息；智慧学习服务主要基于GitHub、StackOverflow、CSDN、博客园等开源社区的公开数据，进行用户相关、仓库、API、代码等不同类型的数据分析与挖掘，挖掘并建立用户学习和开发的需求模型，并基于深度学习、知识图谱、NLP等技术提供面向学习者个性化学习需求的学习内容自动生成和推荐服务。

本仓库存储关于软件开发初学者的研究，主要包括论文, 专利, 竞赛, 代码, 开源系统, 以及分析过程中用到的数据。这些项目的作者包含了已毕业和尚未毕业的研究生。未发表的工作不在此列。

## 仓库列表

---

- <https://github.com/delavet/S0workspace>
  - 作者: 尹航
  - 论文标题: An API learning Service designed for inexperienced developers. Powered by StackOverflow, Wikipedia, Javadoc, etc
  - 会议: ICWS 2021
- <https://github.com/zhouyugui1492/GHTRec-Service>
  - 作者: 周彧祺
  - 论文标题: GHTRec: A Personalized Service to Recommend GitHub Trending Repositories for Developers
  - 描述: 仓库是对论文中服务的实现，用于推荐用户可能链接到自己仓库的话题。
  - 会议: ICWS 2021
- <https://github.com/Vremold/LinkSO>
  - 作者: 武家伟
  - 描述: 一种将 Stack Overflow 帖子链接到编程语言教科书的弱监督方法。主要依赖于 3 个提取特征中的匹配: 关键词、语义和代码元素
  - 年份: (2022~, 尚未发表)
- <https://github.com/Vremold/DLS/>
  - 作者: 武家伟
  - 论文标题: An Open-source Repository Retrieval Service Using Functional Semantics for Software Developers
  - 描述: 这项工作主要是从现有的前端开源库中挖掘结构化和非结构化的功能语义信息，并在此基础上为用户查询提供检索服务。
  - 会议: ICSS 2022 (最佳论文)
  - 比赛: NCSC 2022
  - 项目前端仓库: <https://github.com/quackson/Semantical-Library-Retrieval>
- <https://github.com/Git-JK/A-Retireval-System-Based-on-the-Functional-Semantics-of-Stack-Overflow-Posts>
  - 作者: 金珂

- 描述：使用动词短语和短语模式来提取问答帖标题中包含的显式功能语义，并使用基于Bert的bertlets模型来提取问题描述和答案中包含的隐式功能语义并回答帖子，并匹配待查询问题的显式和隐式函数语义，从而推荐StackOverflow的问答帖子。
- 年份：2022
- [https://github.com/Harris-pku/graduation\\_design](https://github.com/Harris-pku/graduation_design)
  - 作者：郑元昊
  - 描述：中文问答社区检索系统
  - 年份：2022
- [https://github.com/quackson/cn\\_blogs-analysis](https://github.com/quackson/cn_blogs-analysis)
  - 作者：徐子涵
  - 这个仓库是对中国技术社区 cnblog 的研究。使用图的方法对中文博客进行建模和分类
  - 年份：2021
- <https://github.com/quackson/springbootApiUsageCase>
  - 作者：张佳琪
  - 描述：本研究主要是针对springboot初学者的案例搜索研究，数据来源于GitHub开源数据。使用的算法主要是图核和谐聚类。
  - 比赛：NCSC 2021

## 论文列表

---

1. Jiawei Wu, Yanchun Sun (Corresponding Author), and Jiaqi Zhang. **An Open-source Repository Retrieval Service Using Functional Semantics for Software Developers**. CCF 15th International Conference on Service Science (ICSS 2022), May 13-15, 2022 (Best Paper Award).
2. Yanchun Sun, Hang Yin, Jiu Wen, Zhiyu Sun. **Urban Region Function Mining Service Based on Social Media Text Analysis[J]**. International Journal of Software Engineering and Knowledge Engineering, 2021, 31(4): 563-586.
3. Hang Yin, Yuanhao Zheng, Yanchun Sun (Corresponding Author), Gang Huang. **An API Learning Service for Inexperienced Developers Based on API Knowledge Graph**. In Proceedings of IEEE International Conference on Web Services (ICWS 2021), online virtual congress, September 5-10, 2021.
4. Yuqi Zhou, Jiawei Wu, Yanchun Sun (Corresponding Author). **GHTRec: A Personalized Service to Recommend GitHub Trending Repositories for Developers**. In Proceedings of IEEE International Conference on Web Services (ICWS 2021), online virtual congress, September 5-10, 2021.
5. Hang Yin, Zhiyu Sun, Yanchun Sun (Corresponding Author), Gang Huang. **Automatic Learning Path Recommendation for Open Source Projects Using Deep Learning on Knowledge Graphs**. In Proceedings of 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC 2021), all-virtual, July 12-16, 2021.
6. 张佳琪, 孙艳春 (通讯作者), 黄罡. **基于开源社区分析的API案例推荐服务[J]**. 计算机应用, DOI: 10.11772/j.issn.1001-9081.2021122070.
7. Yanchun Sun, Hang Yin, Jiu Wen, Zhiyu Sun. **Urban Region Function Mining Service Based on Social Media Text Analysis**. Proceedings of the 13th International Conference on Service Science (ICSS 2020), Qinghai, China.
8. 武家伟, 孙艳春 (通讯作者). **融合知识图谱和深度学习方法的问诊推荐系统[J]**. 计算机科学与探索, 2021, 15(8): 1432-1440.
9. Zhiyu Sun, Fang Peng, Junrui Guan, Yanchun Sun (Corresponding Author). **Learning path recommendation based on machine learning**. In Proceedings of ACM Internetwork 2019, Fukuoka, Japan, October 28-29, 2019.

10. Hang Yin, Zhiyu Sun, Yanchun Sun(Corresponding Author), Wenpin Jiao. **A Question-Driven Source Code Recommendation Service Based on Stack Overflow**. In Proceedings of IEEE SERVICES 2019, Milan, Italy, July 8-13, 2019.
11. Zhiyu Sun, Chao Xin, Yanchun Sun(Corresponding Author). **An Automatic Semantic Code Repair Service Based on Deep Learning for Programs with Single Error**. In Proceedings of IEEE SERVICES 2019, Milan, Italy, July 8-13, 2019.
12. Jiao Wenpin, Sun Yanchun. **A values-driven self-organization mechanism for automating multiagent coordination[J]**. Computational Intelligence, 2018, 34(2):635-678.
13. Jialun Shao, Yanchun Sun(Corresponding Author). **A Recommendation Service for Programming Study Based on Stack Overflow**. In Proceedings of IEEE SERVICES 2018, San Francisco, USA, July 2-7, 2018.
14. Jiu Wen, Yanchun Sun(Corresponding Author). **A Map-Matching Service Designed for Courier Trajectories**. In Proceedings of the 24th IEEE ICWS 2017, Hawaii, USA, June 25-30, 2017.
15. Yanchun Sun, Chao Xin. **Using Coursera Clickstream Data to Improve Online Education for Software Engineering**. In proceedings of ACM TUR-C '17, Shanghai, China, May 12-14, 2017.
16. Yanchun Sun, Kui Wei, ZijianQiao, Jiu Wen, Tianyuan Jiang. **A Personalized Service for Scheduling Express Delivery Using Courier Trajectories**. In Proceedings of the 23rd IEEE International Conference on Web Services (ICWS 2016) , San Francisco, USA, June 27 - July 2, 2016.
17. Yanchun Sun, ZijianQiao, Dejian Chen, Chao Xin, Wenpin Jiao. **An Approach to Using Existing Online Education Tools to Support Practical Education on MOOCs**. In Proceedings of the 40th IEEE Computer Society International Conference on Computers, Software & Applications (COMPSAC 2016) , Atlanta, Georgia, USA, June 10-14, 2016.
18. Wenpin Jiao, Yanchun Sun. **Self-adaptation of multi-agent systems in dynamic environments based on experience exchanges[J]**, Journal of Systems and Software, Volume 122, December 2016, Pages 165-179.

## 专利列表

---

1. 一种城市地理语义知识图谱的构建方法
  - 专利号: ZL201910725275.4
  - 完成人: 孙艳春、黄罡、林刘子轩, 孙志玉。
  - 授权日: 2022.3.25
2. 一种基于多元大数据的综合城市地理语义挖掘方法
  - 专利号: ZL201910701692.5
  - 完成人: 孙艳春、黄罡、刘瑜、温九
  - 授权日: 2022.1.28
3. 一种开源项目知识图谱的优化方法和系统
  - 专利号: ZL202010643010.2
  - 完成人: 孙艳春、黄罡、孙志玉
  - 授权日: 2021.6.25
4. 一种支持开源项目碎片化学习的推荐方法和系统
  - 专利号: ZL202010643009.X
  - 完成人: 孙艳春、黄罡、孙志玉
  - 授权日: 2021.5.28
5. 一种基于基因芯片数据的疾病影响因素的挖掘方法
  - 专利号: 201710595987X

- 完成人：孙艳春，乔子健
- 授权日：2020.3.20
- 6. 一种移动健康与医疗的饮食控制方法及其控制系统
  - 专利号：ZL 201610681391.7
  - 完成人：孙艳春、祖凌云、孙志玉、邵嘉伦
  - 授权日：2018.12.04
- 7. 一种基于情景感知的向移动用户推荐最优联系方式的方法
  - 专利号：ZL 201410709710.1
  - 完成人：孙艳春、庄希威、魏奎、黄罡
  - 授权日：2018.02.09
- 8. 一种在线矢量图建模过程的描述与重现方法
  - 专利号：201410662733.1
  - 完成人：孙艳春 陈德健 黄罡
  - 授权日：2017.11.06

## 竞赛获奖列表

1. Jiawei Wu, Yanchun Sun, and Jiaqi Zhang. An Open-source Repository Retrieval Service Using Functional Semantics for Software Developers. CCF 15th International Conference on Service Science (ICSS 2022), May 13-15, 2022 (**Best Paper Award**).
2. 武家伟, 张佳琪, “基于功能语义的开源仓库搜索服务”, 获得2022中国数字服务大会软件服务创新大赛一等奖。(指导教师: 孙艳春)
3. 周彧祺, 武家伟. 一种面向软件开发者的个性化流行软件仓库推荐服务. 获得2021中国数字服务大会软件服务创新大赛一等奖。(指导教师: 孙艳春)
4. 尹航. 基于Stack Overflow的问题驱动的java代码的检索服务. 获得2018中国计算机学会服务计算学术会议软件服务创新大赛二等奖。(指导教师: 孙艳春)
5. 张乐聪, 温九. 面向位置语义的智慧地图服务. 获得2018中国计算机学会服务计算学术会议软件服务创新大赛三等奖。(指导教师: 孙艳春)