

Design and Implementation of Python Learning Platform

Sixia Lian^{1,a}, Shangyu Sun^{1,b}, Zerui He^{1,c}, Jiali Xu^{1,d} and Qian Yin^{1,e*}

¹College of Information Science and Technology, Beijing Normal University, Beijing, China

^alsx@mail.bnu.edu.cn, ^b201511210121@mail.edu.cn, ^c201511210135@mail.bnu.edu.cn,
^dbnuxujiali@mail.bnu.edu.cn, ^eyinqian@bnu.edu.cn

Keywords: Python; Mvc; Nutch; Search Engine; Hanlp; Knowledge Gragh

Abstract. Since Python has become a popular programming language and has wide applications in many areas, such as machine learning, data mining, and web development, it is essential for us to establish a convenient and efficient Python learning platform. The whole platform is based on a typical JSP + servlet + JavaBean model, including MVC as the framework, MySQL as the database, Tomcat as the server and JSP as a main front-end technology. Six primary modules including personal data, online classroom, knowledge base, online programming, blog publishing and real-time news have been implemented based on two excellent tools, Nutch and hanLP, and they are closely linked to provide users with a good and efficient python learning platform.

Introduction

In this era of information explosion, the Internet and IT industries have been developed with an unprecedented speed, even causing a third industrial revolution. Specifically, some popular fields, such as artificial intelligence and data mining are developing rapidly, and among these, Python has taken on an indispensable role. In the *IEEE Spectrum*'s fifth annual interactive ranking of the top programming languages(2018), Python burned up the charts and stays on top[1]. So having a good command of Python is one of the necessary skills for computer-related researchers or professional students. There are some learning resources mainly obtained from official website, private blogs and books, but it also exists some flaws and deficiencies as follows:

1. For official or private website[2], we can easily get access to accumulation of knowledge points and documents, but it is too boring to lose interest for a beginner during learning serious grammar.
2. For educational videos, for example, in MOOC, identifying the pros and cons of resources is complex and difficult, more often, there will be demand for expensive costs to purchase a high-quality course.

In a word, when a beginner plans to learn Python, he can only get the uneven and scattered learning materials on the Internet, which brings a lot of unexpected troubles and reduces the learning efficiency. So we are committed to building an open Python learning platform to improve the current situation. Web application based on a typical JSP + servlet + JavaBean model is mature and commonly used in the website construction nowadays[3,4]. Search engine[5,6] based on knowledge graphs[7] can provide related prior knowledge or advanced knowledge of search terms, which is necessary for us to learn a programming language systematically and search relevant information accurately. We will introduce detailly the design and implementation of our Python learning platform implemented with appropriate technologies and tools in this paper.

System Design

Architecture Design. Firstly, the basic framework of this system is MVC, which is a classic framework commonly used in the software industry, and we use one of the most typical MVC framework, JSP + servlet + JavaBean, in the Python learning platform to realize separation of database, client page, and the logic dealing with operations. When it comes to databases, we choose MySQL for the reason that it is a very practical database that can satisfy the use of small and medium-sized websites and high concurrent access. Although compared with Oracle, SQL Server,

DB2 and other databases, MySQL has certain shortcomings, but for our system, the services provided by MySQL are more than enough for now at least, and another reason is that we are deeply attracted by the characteristics of open-source and free. In addition, Tomcat is also a free and open-source server with relatively stable performance and can withstand the intensity of our project. As shown in Figure 1., resources from server, such as Html, CSS, and JavaScript, are rendered by browser rendering engine and vivid webpages will be presented to users, users interacts with the platform through the interface, interaction information and events will be processed by the "controller" in MVC framework. Based on different users' behavior, "controller" will directly choose corresponding "view" to update the interface or modify the information in the "model" and finally feedback changes to user interface. We utilize the JDBC API and MySQL JDBC driver to facilitate conversation with database. All in all, we aim to provide an user-friendly, perceivable and interactive learning platform using web programming techniques from front end to back end as Java, SQL, JSP and JavaScript.

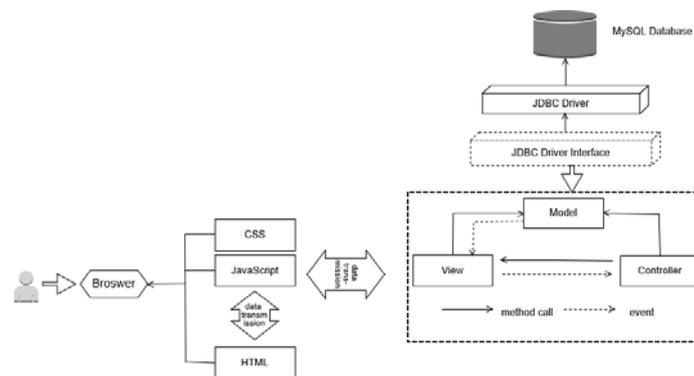


Figure. 1 Overall architecture design of the Python learning platform

Functional Design. Through the investigation of existing learning resources, we find that basic grammar, programming exercises, teaching video and highly relevant searches are indispensable in a sound learning platform. For a Python beginner, high-quality and systematic learning documentation is necessary to build the solid foundation. Secondly, many people felt from previous study experiences that online classes can help them comprehend and cement knowledge. However, the most important thing to learn programming language is coding, but for a beginner, the installation of Python and the configuration of environment variables are a very cumbersome process. What's worse, when using different computers, you must reconfigure the programming environment laboriously. Therefore, online programming is particularly important, which means that just coding in the corresponding module and the corresponding result will show immediately. In addition, the acquisition of information is also a very important part in the learning process. However, mainstream search engine provides a huge amount of too extensive rather than specialized information, meanwhile, advertising and malicious plugins are also headaches.

As shown in Figure 2., the system is mainly composed into six functional modules: personal data, online classroom, knowledge base, online programming, blog publishing and real-time news. The search engine is embedded on the platform to search all the information of these six parts, and then recommend relevant search results according to the knowledge graphs. Online classroom is mainly a collection of various Python learning videos, and no doubt, this module is realized as dynamic webpages. To ensure the quality of videos, only the administrator has the right to add or delete videos after strict selection. This part of the knowledge base is also the integration of Python knowledge, mainly including some authoritative documents, from the shallower to the deeper, from building a development environment to actual combat. Moreover, functions as personal data, blog publishing and real-time will enable our learning platform to be more complete. After all, browsing a few industries news and publishing blog to record own experience are more interesting than studying all along.

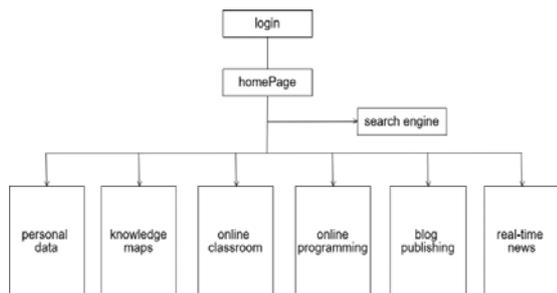


Figure. 2 Our platform can be divided into six modules according to functions.

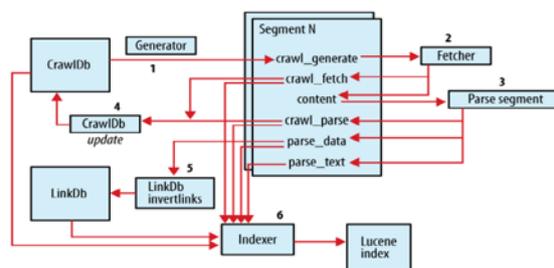


Figure. 3 Working principle of embedded search engine.

Main Implementation

In this section, we'll look at the main implementation for the system. It's worth mentioning that implementation of some common functions as blog publishing or online classroom is omitted (specific interface implementation can be seen in Figure 4., we have implemented some practical successively, including blog publishing, online programming, real-time news and online classroom), and we will focus on a few difficulties to establish a search engine dedicated to python knowledge, which are the core of the whole system, including embedded search engine, establishment of knowledge graphs and words segmentation.



Figure. 4 This Python learning platform provides friendly interface to users.

Embedded Search Engine. Apache Nutch[15] is an open-source web search engine (or crawler) based on Java. We can find links in assigned websites automatically, assign depth to implement broad first search (DFS) to obtain relative content in website, establish indexes based on these content, and find assigned data according to data indexes, as shown in Figure 3. Nutch is based on Lucene (a full text retrieval toolbox based on java providing index and search function rather than a complete search application), and Lucene provides text retrieval and search API for Nutch. Nowadays, Lucene is not only an open-source project from Apache Jakarta family but also most popular full text retrieval toolbox based on java. It can establish indexes for text type data, so just transfer data to text type data, and Lucene will index and search your text. But after Nutch 1.2 version, Apache Nutch project has changed his own direction and Nutch has evolved Internet worm, it will gain content effect by using with Solr. At the same time, Apache Nutch has been divided two series, 1.x and 2.x. These two versions main differences are bottom storage difference. Series 1.x are based on Hadoop frame, bottom storage is HDFS, and Series 2.x is Apache Gora, which makes Nutch is accessible to HBase, ccumulo, Cassandra, MySQL, DataFileAvroStore and other NoSQL databases. Solr is full text retrieval server based on Lucene with a high performance. But Solr extends Lucene to provides more fluent search language, optimize search performance and provide a complete function management interface. Solr is a very splendid full text retrieval, which mainly concludes: high efficiency, flexible cache, vertical search and etc.

Nutch-1.9 + Solr4 .6.1 + tomcat8.5.29 + win10 is used in this system. After downloading the source code for Nutch-1.9 , compile with ant and then start configuring the relevant information, such as Nutch's Clawer information, which requires the identity of the mission Clawer and the type of protocol for the data obtained. We need to create a new file to store root URLs of the webpage to be crawled by Clawer of Nutch and change Clawer's filtering rules, including URL type and dynamic link, according to your own situation.

Establishment of Knowledge Graphs. The knowledge graph is also known as scientific knowledge graph, which is graphs to reflect development process and a series of structural relationships between various knowledge. In short, the knowledge graph is the set of relationships between different knowledge points and the connections between all knowledge points are showed clearly. The search results from search engines, such as Baidu and Google, always contain recommendations from knowledge graphs, no matter what kind of connection between search keywords and results is. Many researchers attach great importance to the study of knowledge graphs technology and we also apply it in our python learning platform. Knowledge graphs for Python learning is far less than the huge amount that Baidu or Google has built for searching. For Python learners, the number of knowledge points that need to be mastered is limited. As shown in Figure 5., common knowledge points are listed and then we can draw an inclusion relationship diagram of python knowledge points according to the inclusion relationship of all points. Finally, directed graphs are used to represent topological relationships among knowledge points, in which we can easily find the precursor node as prerequisite and the successor node as advanced knowledge of a python knowledge point. In this way, python knowledge graphs have become a directly manipulated object and in the retrieval of a certain knowledge point, the relevant points will also be retrieved. What's more, the construction of python knowledge graphs is an iterative process that needs to be manually maintained and updated, because knowledge points need to be comprehensive and python is constantly evolving.

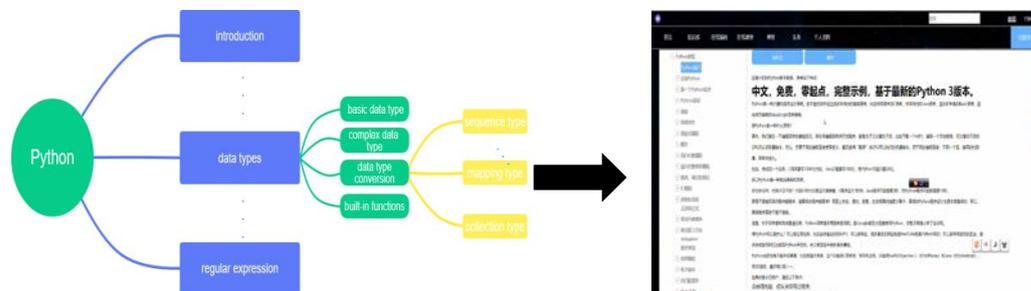


Figure. 5 A knowledge graph example shows connections between some knowledge points at left and the final interface is as shown at right. Users can easily study or review several knowledge points.

Word Segmentation. The search engine is strongly dependent on the establishment of knowledge graphs to provide related prior knowledge or advanced knowledge of search terms. In the search process, keyword extraction is one of the most steps and we use hanLP (Han Language Processing)[16], an NLP toolkit consisting of a series of models and algorithms, to improve the quality of search results. The hanLP toolkit not only provides general word segmentation, but also can extract keywords and even phrases, so that the relevant knowledge points are accurately retrieved at the time. For example, when the user searches for "how to define a function", the general word segmentation package will divide this sentence into "how to \ define \ a \ function", but hanLP can extract the key phrase "definition function", so knowledge points of "definition function" and related points can be accurately fed back to the user. In short, hanLP can greatly improve the accuracy of search results. In addition, many word segmentation packages is not professional enough to contain professional vocabulary. Therefore, the error is very great when part of the word segmentation contains python knowledge points. For example, when users search for "balanced binary tree", the result of many word segmentation will be "balanced \ binary \ tree",

because the word binary tree is very rarely used in ordinary life. Luckily, hanLP can define user's own vocabulary. Therefore, we have added a considerable amount of vocabulary commonly used in Python, so that "balanced binary tree" will be recognized as a whole and relevant knowledge points will be retrieved. Combining the basic functions of hanLP(extracting the keyword, extracting the phrase function) and professional vocabulary, the accuracy of identifying the knowledge points searched by the user can be greatly improved.

Summary

Based on the MVC framework, a Python learning platform is designed and implemented for the public to learn and use, with MySQL as the database and JSP + servlet + JavaBean as the model. A search engine is embedded in this platform, which helps users to retrieve the information they are looking for in an accurate and timely manner. The construction of the knowledge graphs is to perfect the search results and user needs. We have not only implemented many powerful and practical functions, but also spent a lot of time on interface design to provide friendly and elegant interface. Therefore, the entire platform is convenient, fast, and suitable for users to learn Python.

Acknowledgement

The research work described in this paper was fully supported by the National Natural Science Foundation of China (Project No.61472043), the Beijing Normal University Postgraduate Training-Method Course Construction Foundation(2017). Prof. Qian Yin is the author to whom all correspondence should be addressed.

References

- [1] Information on <https://spectrum.ieee.org/at-work/innovation/the-2018-top-programming-languages>.
- [2] Information on <https://www.liaoxuefeng.com>.
- [3] Y.Q. Zhang, C. Chen and H.E. Mingde: Computer Engineering, (2001).
- [4] W.U.Q. Chen and Z.H. Rong: Computer Engineering, (2001).
- [5] X.L. Yu and J.Z. Zhao: Journal of Beijing Institute of Technology(2003).
- [6] E.M. Jose, M.M. Michael and D.D. Silva: International Conference on Supercomputing(2007), p.3.
- [7] K. Zhang and Q. Xue: International Conference on Future Computer and Communication(2010), Vol. 1, p.850.
- [8] W.W. Fu, R.F. Li, Y.F. Liu and S.L. Huang: Telecommunications Science(2011)
- [9] M.F. Sanner: Journal of Molecular Graphics & Modelling, Vol. 17(1999) No.1, p.57
- [10] G. Lindstrom: It Professional, Vol. 7(2016) No.5, p.10
- [11] J.L. Gordon: Knowledge-Based Systems, Vol. 13(2000) No.2, p.71.
- [12] J. H. Lee, and A. Segev: Computers & Education, Vol. 59 (2012) No.2, p.0.
- [13] D. P. Pop and A Altar: Annals of Daaam & Proceedings, Vol. 69(2014) No.1, p.1172.
- [14] Information on <http://nutch.apache.org/>.
- [15] Information on <http://hanlp.linrunsoft.com/>.