

Mr. Yin

Address: Nanjing, China

OBJECTIVE

Crawling Engineer

WORK EXPERIENCE

07/2017

Java Crawling Engineer by Nine Powers Big Data, Nanjing, China

Responsibilities:

For the collection and cleaning of designated websites (government websites, tripartite websites, enterprise search, mobile app, WeChat articles, OTA travel sites - Tuniu, Ctrip, Tongcheng, Ma Honeycomb, where to go, etc.).

Responsibilities: Build a crawler system to implement massive data capture, cleaning, and storage. There is some research on solving verification codes, IP blocking and cookie analysis. Tools used: Selenium, Jsoup, HttpClient, regular expressions, and other open source software.

07/2016 – 02/2017

Java Crawling Engineer by Jiangsu Si Kerui Data Technology Co., Ltd., Nanjing, China

Responsibilities:

1. Data capture for designated websites, as well as data cleaning (such as China Judgment Document Network, China Construction Qualification Network, China Patent Network, etc.);

2. Commercial analysts, some government project leaders connected with the company, and analysis of foreign countries Website and related products abroad.

EDUCATION

2009 – 2013

Nanjing Normal University, China

Communication Engineering , Bachelor degree

2013 – 2015

University of Limerick, Irland

Computer and communication system, Master degree

Self Introduction

Proficiency in the use of eclipse, IDEA, SVN, fiddler and other development capture tools, familiar with HTTP protocol, HTML pages, AJAX response. Skilled in using open source technologies such as jsoup, htmlparser, regular expressions. There are a large number of web pages, mobile client, WeChat article data crawling experience. Good at analyzing the structure of the website and designing the corresponding

crawling strategy. Have a certain understanding and experience in solving website anti-crawling.
