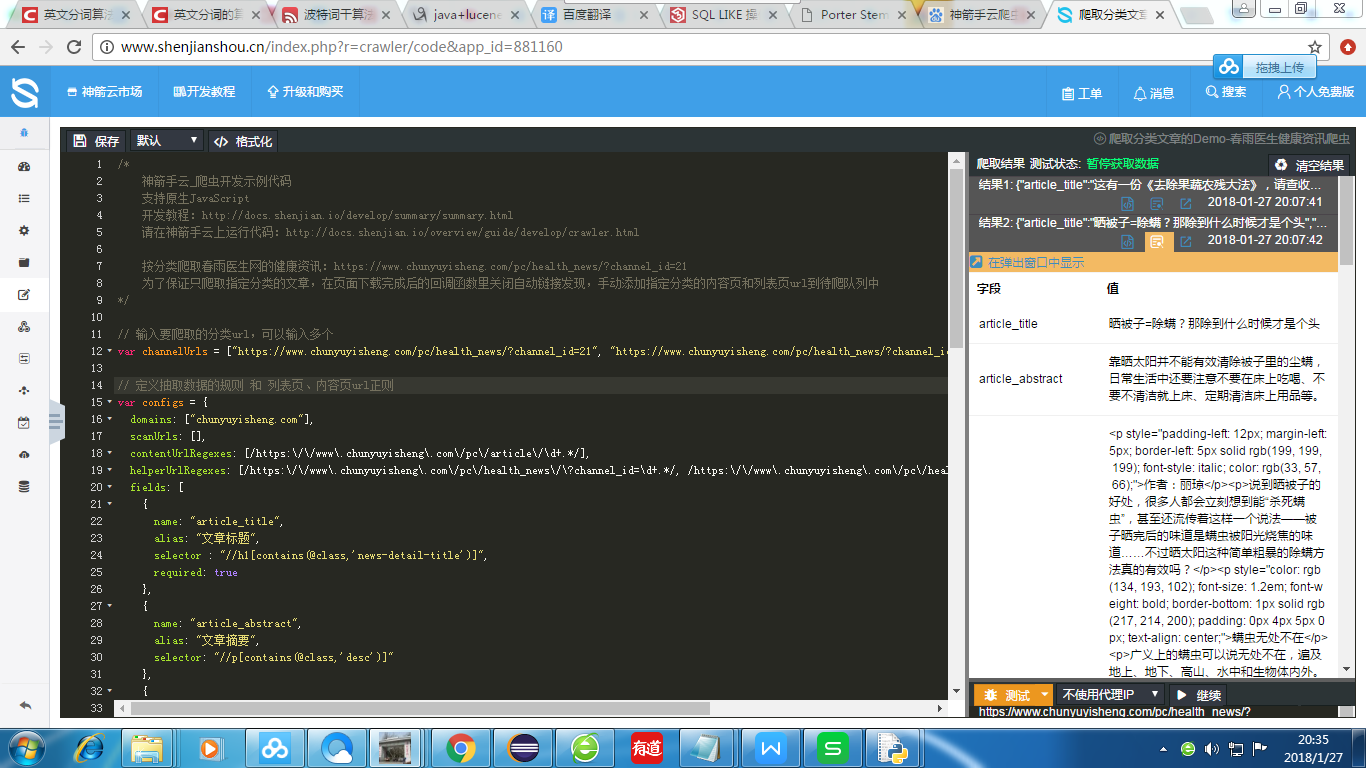
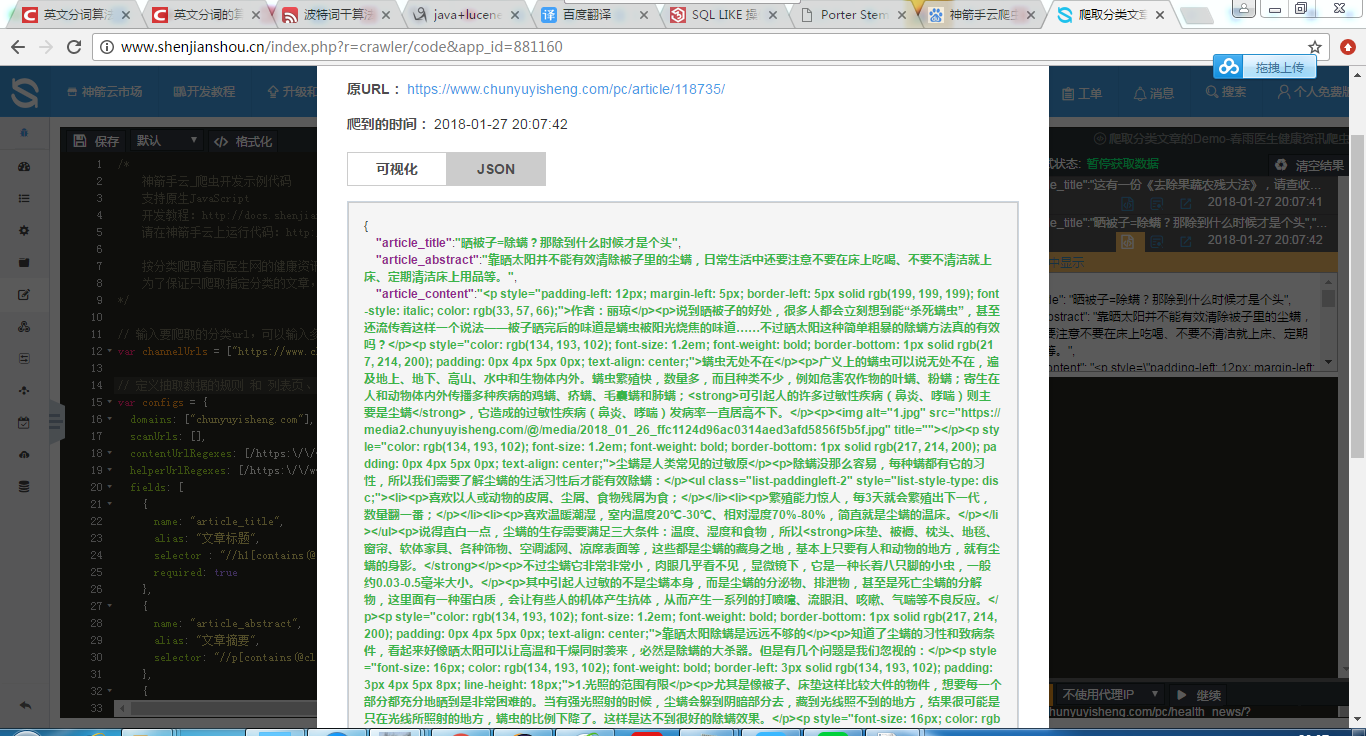


We use the online platform has been configured crawler environment, we only need to pay attention to the crawler code changes and improvements can be, do not need to pay attention to the underlying configuration, and to help manage the team members can modify and manage code online, Help code maintenance and update, very convenient.

The left side of the screen shows the specific crawler code, the top right shows the results crawled, the bottom right shows the progress of the crawler.

In the crawler's improvement, we have changed the data format we need on the basis of the original, we need to get the relevant health information articles, so we set the article title, the article abstract, the article content, the release time, the article classification Five fields, and limit the url of the target link to ensure that we crawl process, there will be no crawler access to irrelevant information, the other crawler URL link can have more than one, all support for different web crawler format.



The left side of the figure shows the field of one of the crawled data and the corresponding field value.

This is the style we crawl through Json, which is the key-value pair.

This is the visual data corresponding to Json



This is the original page information