**Project 5.43. Web scraping – VC/PE investment events**

**Background**

**Cyzone** is a platform for start-ups and investors (Venture Capital/Private Equity) in China. Their [website](http://www.cyzone.cn/event/list-764-0-1-0-0-0-0/) provides a database of financing events as shown below, around 20000 events in total. These data can be used to analyze the latest trend of technological & social development of China, the investment preference of VC/PE funds and their cooperation, and also potential companies to go IPO.



**Requirement**: submit a code to realize the following function:

1. Use web-scraping to retrieve the financing events data with the following information:
   1. Company name
   2. Amount of Financing
   3. Round of Financing
   4. Investor
   5. Industry
   6. Date
   7. Details link
2. Users can an enter filter condition(s) in the **search box,** for example——
3. amount of financing: >50 million
4. round of financing: B round
5. industry: healthcare
6. date: from 2015 to 2016

Then return a filtered csv file according to the above condition(s).

1. **[Bonus]** Picture a VC/PE network using Gephi. The picture can reveal the following information among VC/PE:
   1. Top investors
   2. Cooperation (lead-investor / co-investor)
   3. Competition (a cluster of investors vs. another cluster of investors)

**Codes will be graded based on:**

1. The format of the codes
2. The organization of the csv file
3. The speed of the codes
4. The extendibility of the codes

Estimated completion time: 40 hours. Time consumption depends on the codes’ quality.

Suggestive points:

Deadline:

Bonus points:

1. Work completion one day before the deadline will generate an extra 20% bonus.

Penalty:

1. Late delivery/missing deadline will incur -10% of finally determined points of the whole project per missing day.
2. Serious mistakes that made the submitted results not useful will incur -120% of the whole project.