

Progetto Ingegneria Informatica (5 CFU)
2017/2018

Riccardo Poiani, Mattia Tibaldi, Tang-Tang Zhou
Politecnico di Milano
Mercurio

April 11, 2018

Contents

1	Introduction	4
2	Web scraping	4
2.1	Static sites	4
2.2	Dynamic sites	4
2.2.1	Infinite scroll websites	4
2.3	Deployment	4

1 Introduction

This is the introduction section

2 Web scraping

The first target that had to be accomplished to turn the Mercurio project and its analysis to an international version was the data collection of financial news from various sources by web scraping techniques. In particular, the sites that had been considered were:

- bloomberg.com
- nytimes.com
- thisismoney.co.uk
- money.cnn.com
- marketwatch.com
- reuters.com
- investing.com
- moneymorning.com
- 4-traders.com

An important distinction among the sources mentioned above is whether or not is required an ajax/javascript interaction with the user in the web page that arranges the intended articles. The ones that do not involve these interaction will be treated in the "static sites" section, the leftovers beneath "dynamic sites" section, with a focus on infinite scroll websites.

2.1 Static sites

This regards static sites.

2.2 Dynamic sites

This regards dynamic sites.

2.2.1 Infinite scroll websites

This regards infinite scrolls.

2.3 Deployment

This regards deployment.