

Configurations of the computer  
Operation system: Win7  
CPU: Intel i7 8th  
RAM: 12G  
ROM: 512G

Wisers Information Portal

WiseSearch

Ming Pao Instant News

News Sources

Document ID

DOCUMENT ID: 201809123221525

港聞 | 2018-11-01

Basic Info.

港灣道爆水管路面下陷 西行線封閉

Title of the water pipe burst news

港灣道25號一地下食水管爆裂，導致路面下陷，事件中2人受傷。

Main context

提取到的数据

Extracted news articles

#	字段1 Basic Info.	字段2 Title	字段3 Main context	字段4 Document ID	字段5 News Sources
1	A02   港聞   澳門...	澳門口岸又出醜 旅檢...	【靜態組報道】被指...	DOCUMENT ID: 201...	Oriental Daily News
2	E01   文化   2018...	【港情講趣】Sor9y...	【港情講趣】香港人...	DOCUMENT ID: 201...	Apple Daily
3	A06   港聞   2018...	灣仔爆水管 大家樂...	【本報訊】灣仔港灣...	DOCUMENT ID: 201...	Apple Daily
4	P24   港聞   2018...	無妄之災 貨車司機跟...	灣仔華潤大廈對開路...	DOCUMENT ID: 201...	Headline Daily
5	A16   港聞   2018...	爆水管貨車輪陷 膠籃...	【本報訊】一輛隸屬...	DOCUMENT ID: 201...	Oriental Daily News
6	港聞   2018-11-01	港灣道爆水管路面下...	港灣道25號一地下食...	DOCUMENT ID: 201...	Ming Pao Instant Ne...

Web-crawler for news collection

- 21 pieces of news per minute shown in Fig. S2(a)
- Around 134 mins for collecting 2828 news in this study

Performance of the web-crawler

Collected news: 105 pieces (1 repetition); The time used: 4 Mins 48 Seconds; Collection speed: 21 pieces of news per minute

已采集:105条(1条重复) 已用时: 4分钟48秒 平均速度:21条/分钟

(a) Time consumed by web-crawler for collecting news articles

PythonExercise

infrastructure.py

Recognition of infrastructure entities

189 secs for extracting 7064 infrastructure entities shown in Fig. S2(b)

Codes for recording time

```
endtime = datetime.datetime.now()
print(endtime)
print((endtime-starttime).seconds)
```

Named entity recognition of infrastructure entities

Start time

2020-03-13 21:28:45.028323

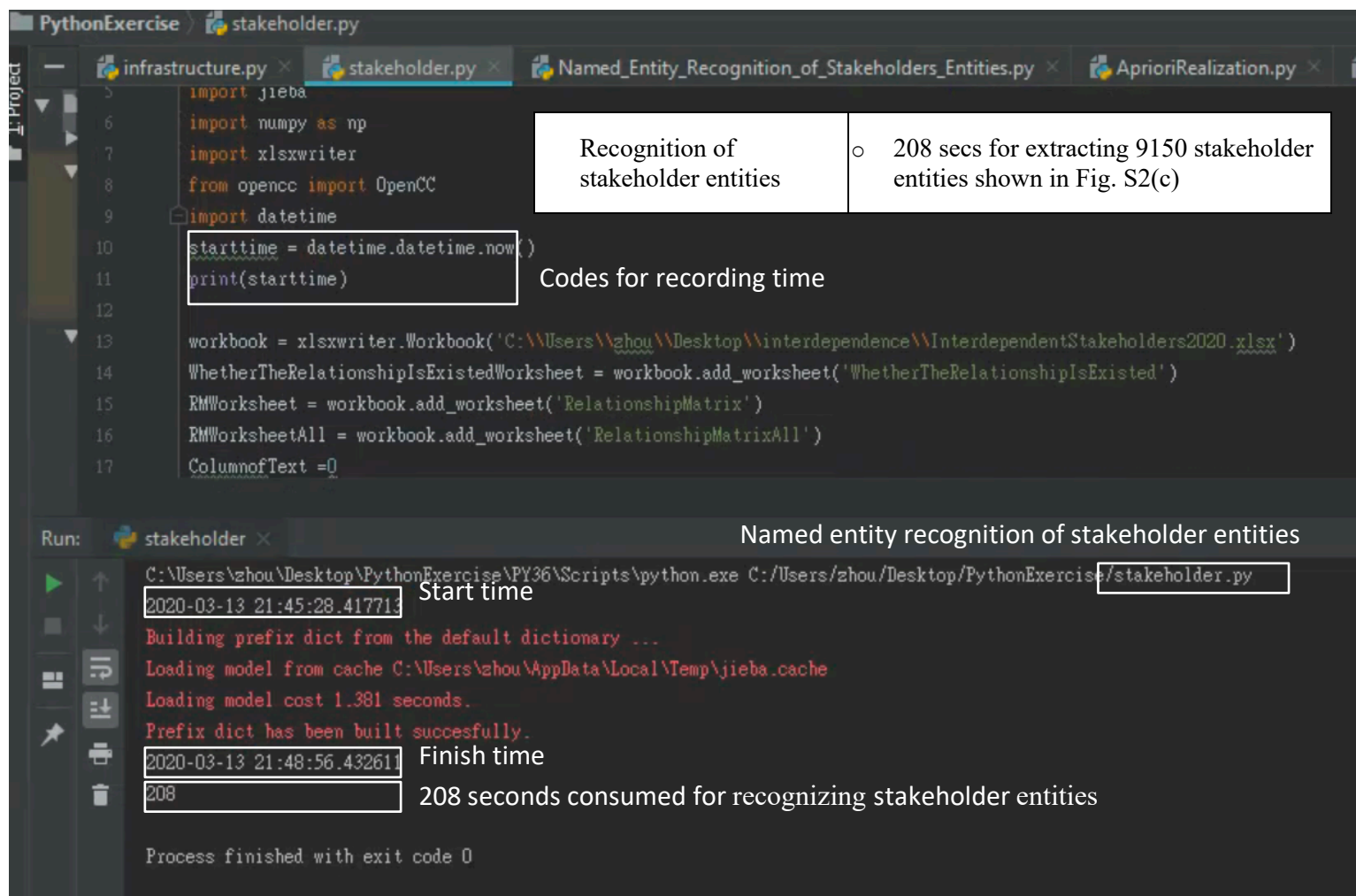
Finish time

2020-03-13 21:31:54.091137

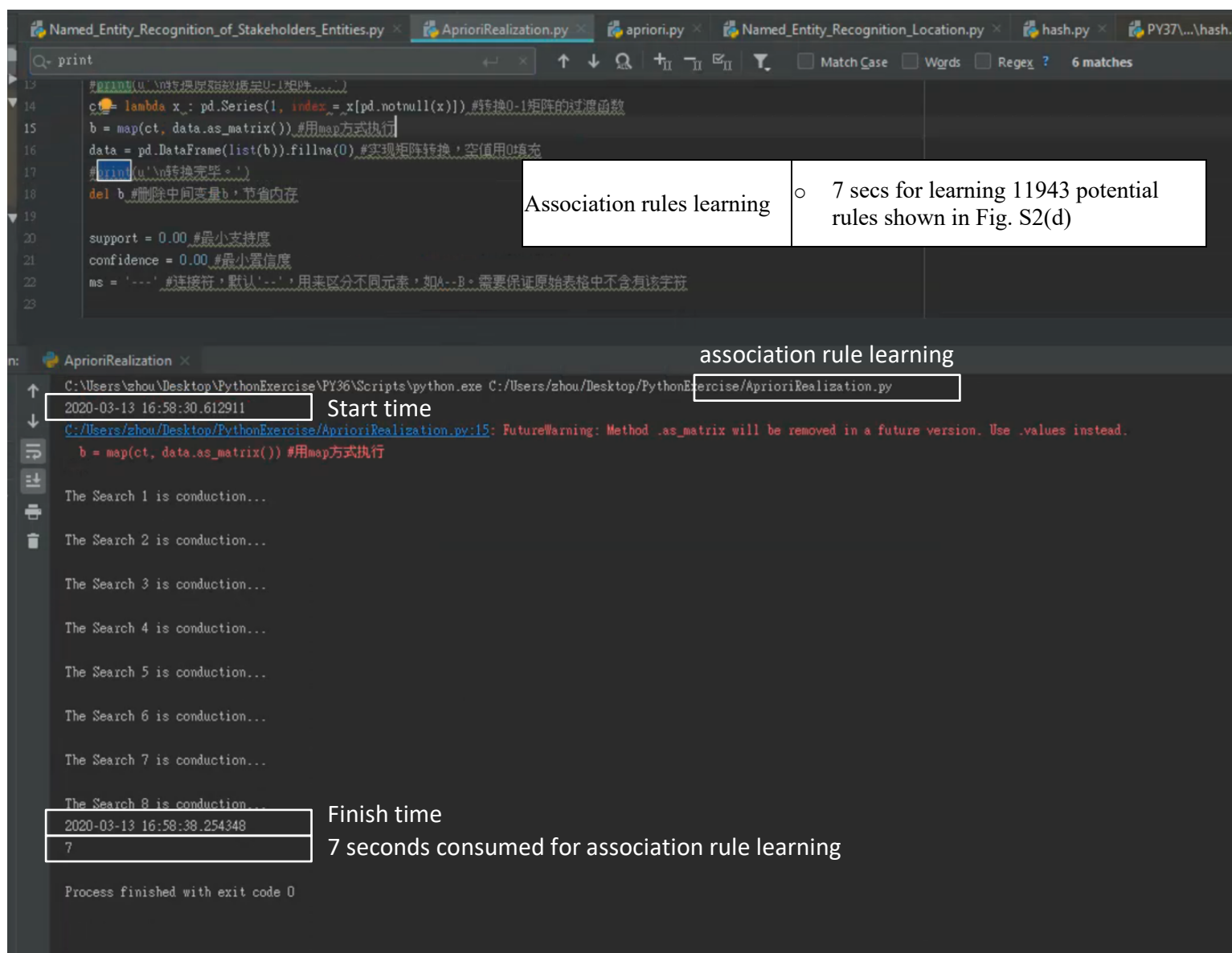
189

189 seconds consumed for recognizing infrastructure entities

(b) Time consumed by recognizing infrastructure entities



(c) Time consumed by recognizing infrastructure entities



(d) Time consumed by association rule learning

Fig. S2 Time consumptions of main steps in the text mining approach