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

港聞 | 2018-11-01 — Basic Info.

Document ID — DOCUMENT ID: 201809123221525

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

提取到的數據 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
o 21 pieces of news per minute shown in Fig. S2(a)
o 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

< 1 ··· 9 10 11 >

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

File Edit View Navigate Code Refactor Run Tools VCS Window Help

PythonExercise > infrastructure.py

infrastructure.py ×　Named_Entity_Rec…

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

```
165    for ColumnofRM in range(len(
166        RMWorksheet.write_string(RowofRM, ColumnofRM, str(RM[RowofRM][ColumnofRM]))
167        RMWorksheetAll.write_string(ColumnofRM, RowofRM, str(RM[RowofRM][ColumnofRM]))
168    # worksheet.write(NumberofRow, NumberofColumn, WhetherTheRelationshipIsExisted[m][index])
169    workbook.close()
170    endtime = datetime.datetime.now()
171    print(endtime)
172    print((endtime-starttime).seconds)
```

Codes for recording time

Named entity recognition of infrastructure entities

C:\Users\zhou\Desktop\PythonExercise\PY36\Scripts\python.exe C:/Users/zhou/Desktop/PythonExercise/infrastructure.py

Start time
2020-03-13 21:28:45.028323
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\zhou\AppData\Local\Temp\jieba.cache
Loading model cost 0.805 seconds.
Prefix dict has been built successfully.

Finish time
2020-03-13 21:31:54.091137
189   189 seconds consumed for recognizing infrastructure entities

Process finished with exit code 0

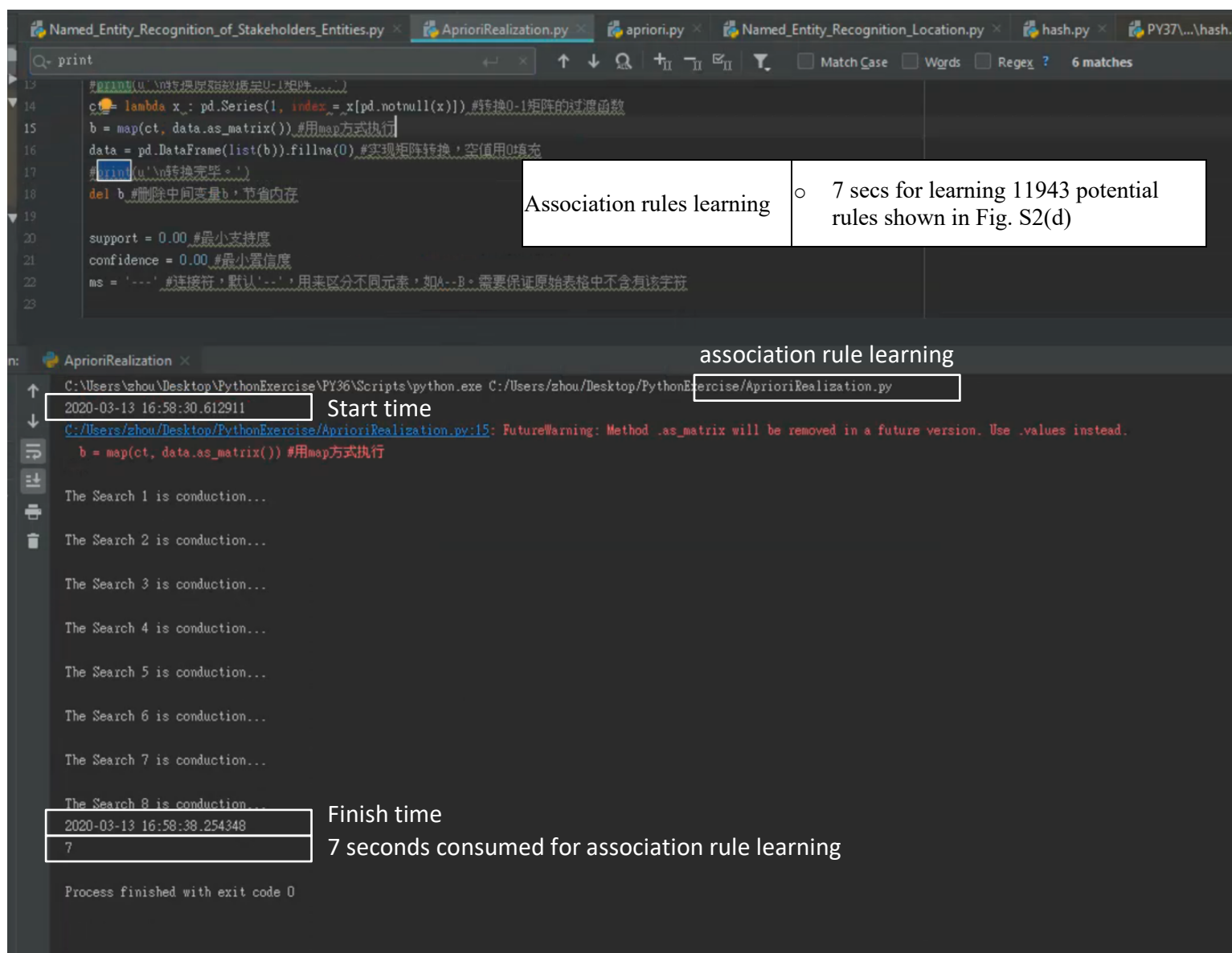(b) Time consumed by recognizing infrastructure entities
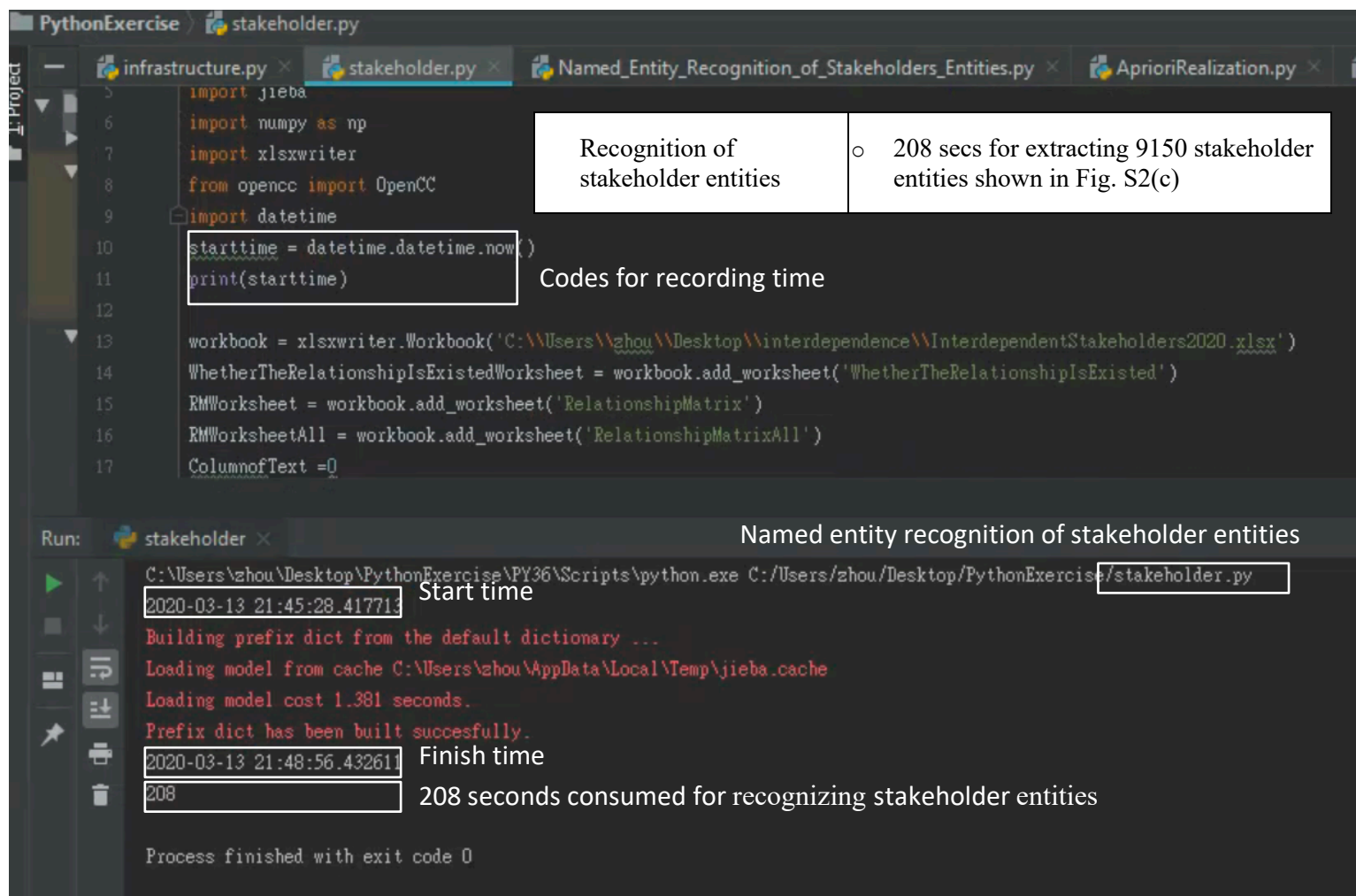
(c) Time consumed by recognizing infrastructure entities

(d) Time consumed by association rule learning

Fig. S4-1 Time consumptions of main steps in the text mining approach