

## תרגיל 1 נתוני עתק

Harel Meir

205588940

סעיף א'

### **Given xpath:**

```
//span[@id[contains(., "Twin_towns")]]/..following::ul[1]/li/a/@href[starts-with(., "/wiki")]
```

### **Given url:**

[https://en.wikipedia.org/wiki/Ramat\\_Gan](https://en.wikipedia.org/wiki/Ramat_Gan)

### **Returns:**

/wiki/London\_Borough\_of\_Barnet

/wiki/Kassel

/wiki/Main-Kinzig-Kreis

/wiki/Penza

/wiki/Phoenix,\_Arizona

/wiki/Qingdao

/wiki/Rio\_de\_Janeiro

/wiki/San\_Borja\_District

/wiki/Shenyang

/wiki/Strasbourg

/wiki/Szombathely

/wiki/Taoyuan,\_Taiwan

/wiki/Weinheim

/wiki/Wroc%C5%82aw

## סעיף ב'

Xpath:

```
//table[@class="infobox ib-settlement vcard"]//div[@class="category"]//a[contains(text(), "city") or contains(text(), "City")]
```

Given URL:

[https://en.wikipedia.org/wiki/Ramat\\_Gan](https://en.wikipedia.org/wiki/Ramat_Gan)

given output:

non-empty(City)

## סעיף ד'

התוצאות הינן שונות, לא יצאו אותן ערים.

לדעתי, זה קורה כיוון שכאשר הטפלורט כבוי(כלומר תמיד עוקבים אחרי הלינק) אז ישנן קבוצת ערים שמובילות למעין מלכודת עכביש, כלומר לינקים שמייצרים צורה מעגלית עד וצוברים ניקוד דירוג, ועד שהגולש הרנדומי יוצא ממנו הדפים שמרכיבים את המעין מעגל הנ"ל צוברים ניקוד.

כאשר הוא דולק, הוכחנו בכתה שכאשר מוסיפים קפיצות רנדומיות, בהסתברות גבוהה אכן מגיעים להתכנסות. ולכן לדעתי מכאן נובע ההבדל.