

Einflüsse auf die Kriminalität in den Counties von England

Abschlusspräsentation

Semantic Web

Anne Matthes – 14MI-M

07.07.2015

Fragestellung

Einflüsse auf die Kriminalität in britischen Großstädten

Welche sozialen und lokalen Kriterien nehmen Einfluss auf die Häufigkeit von Kriminalitätsvorfällen in britischen Regionen? Es ist zu überprüfen, ob eine Abhängigkeit zur Arbeitslosigkeit und Bevölkerungszahl besteht.

(Fokus: Counties von England)

Fragestellung

Interpretation der Fragestellung

- Bestehen generell Abhängigkeiten zwischen den drei Zahlen
 - Kriminalitätsvorkommen
 - Bevölkerung
 - Arbeitslosigkeit

weiterführende Ansätze:

- Bestehen Abhängigkeiten zur Jahreszeit?
- Verändern sich die Abhängigkeiten über die Zeit und gibt es auffällige Unterschiede zwischen den einzelnen Orten?
- Anwendung durch Hinzunahme weiterer Quellen für Wales, Schottland und Nordirland
- Vergleich eines Counties zur Region oder zum gesamten England

Relevante Datenquellen

Daten zum Kriminalitätsvorkommen Englands

Link: <http://data.police.uk/data/>

Datenformat: CSV (in komplexer Ordnerstruktur)

Schnittstelle: HTTP, (früher: API)

Lizenz: Open Government Licence

Open Data: 

Relevante Datenquellen

Daten zur Bevölkerung in England

Link: <https://www.nomisweb.co.uk>

Datenformat: XLSX, CSV, TSV, HTML

Schnittstelle: HTTP, API

Lizenz: Open Government Licence

Open Data: 

Relevante Datenquellen

Daten zur Arbeitslosigkeit in England

Link: <http://data.london.gov.uk/dataset/unemployment-rate-region/resource/8a29ec0c-9de3-4777-832f-49ef8c2b4d14XLS>

Datenformat: XLS

Schnittstelle: HTTP

Lizenz: Open Government Licence

Open Data: 

Relevante Datenquellen

Daten zur Zuordnung der Counties zu den Regionen

Link: https://de.wikipedia.org/wiki/Verwaltungsgliederung_Englands

Datenformat: HTML

Schnittstelle: HTTP

Lizenz: GNU Free Documentation License

Open Data: 

Extraktion der Daten und Import in einen Triple Store

Eine Java Application für alle Datenquellen

1. Auslesen der Daten und Befüllen von `ArrayList<String[]>`

- Parsen von HTML-Seiten mittels *JSOUP*

```
Document doc = Jsoup.connect(res[0]).get();
splittedHtml = doc.body().toString().split("\n");
```

- Auslesen von Excel- oder CSV-Dateien mittels *Java Excel API*

```
public ArrayList<String[]> loadData(String res) {
    File inputWorkbook = new File(res);
    Workbook w;
    try {
        w = Workbook.getWorkbook(inputWorkbook);
        Sheet sheet = w.getSheet(0);

        for (int i = 8; i < sheet.getRows() - 2; i++) {
            Location = sheet.getCell(0, i).getContents().split(":")[1];

            for (int j = 2; j < sheet.getColumns(); j += 2) {

                if (sheet.getCell(j, i).getContents().equals("-"))
                    continue;
            }
        }
    }
}
```

- Navigation durch Ordnerstruktur

Extraktion der Daten und Import in einen Triple Store

2. Anpassen der Datenformate

■ Hilfs- und Formatierungsmethoden

- `formRegionName()` Bsp.: „South East England“ → „South_East“
- `formCountyName()` Bsp.: „County Durham“ → „Durham“
- `getLocationType()` Bsp.: nur „uacounty“ relevant
- `formatNumbers()` Bsp.: Kommas und Punkte entfernen
- `convertDate()` Bsp.: „Apr 15“ → „2015-04“
- ...

Extraktion der Daten und Import in einen Triple Store

3. Erstellen der RDF-Files mithilfe des *Jena Frameworks*

```
String uri = "http://www.imn.htwk-leipzig.de/~amatthes/semweb/schema#";

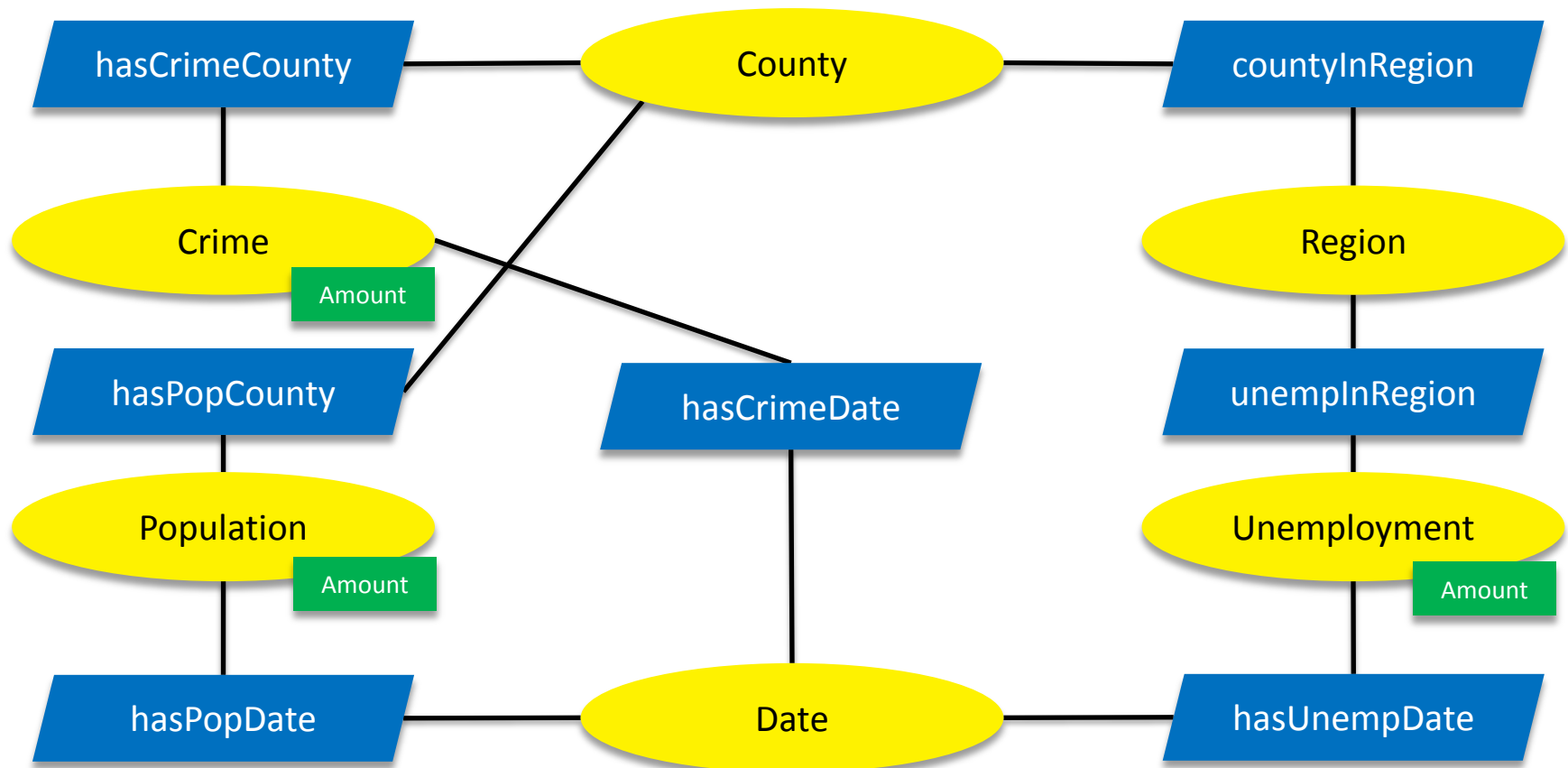
OntModel model = ModelFactory
    .createOntologyModel(OntModelSpec.OWL_DL_MEM_RULE_INF);
model.setNsPrefix("am", uri);

Resource crimeRes = model.createResource(uri + "Crime");
model.add(crimeRes, RDF.type, RDFS.Class);
Resource countRes = model.createResource(uri + "Count");
```

- Ausgabe im RDF-Format

Verlinkung von Ressourcen

3. Erstellen der RDF-Files mithilfe des *Jena Frameworks* und importieren der RDFs in den Triple Store *Apache Fuseki*



Verlinkung von Ressourcen

4. Verlinkung von Ressourcen

- Verbindung der Klassen Crime, Population und Unemployment über **Zeitpunkt** und **Ort**

Crime	Population	Unemployment
Date (Monate)	Date (Monate)	Date (Monate)
County ↴ Region ↴	County ↴ Region ↴	Region

Extraktion der Daten und Import in einen Triple Store

5. SPARQL-Anfrage an den Triple Store *Apache Fuseki*

- Vergleich der Gesamtwerte der Regionen – Anfrage:

```
1 PREFIX am: <http://www.imn.htwk-leipzig.de/~amatthes/semweb/schema#>
2
3 SELECT ?Region (SUM(?CrimeAmount) AS ?CrimeSum)
4 (SUM(?PopAmount) AS ?PopSum) (SUM(?UnempAmount) AS ?UnempSum)
5 WHERE{
6   ?crime am:hasCrimeDate ?Date .
7   ?pop am:hasPopDate ?Date .
8   ?unemp am:hasUnempDate ?Date .
9   ?crime am:hasCrimeCounty ?County .
10  ?pop am:hasPopCounty ?County .
11
12  ?unemp am:unempInRegion ?Region .
13  ?County am:countyInRegion ?Region .
14
15  ?crime am:hasCrimeAmount ?CrimeAmount .
16  ?pop am:hasPopAmount ?PopAmount .
17  ?unemp am:hasUnempAmount ?UnempAmount .
18 }
19 GROUP BY ?Region
20 ORDER BY ?CrimeSum
```

Extraktion der Daten und Import in einen Triple Store

5. SPARQL-Anfrage an den Triple Store *Apache Fuseki*

- Vergleich der Gesamtwerte der Regionen – Ergebnis:

	Region	CrimeSum	PopSum	UnempSum
1	am:West_Midlands	388328.0e0	6.31761E7	2.0359926692E10
2	am:North_West	453857.0e0	7.54734E7	2.355963389E10
3	am:South_West	541208.0e0	6.74808E7	1.9353152211E10
4	am:South_East	857673.0e0	1.78776E8	3.1897693128E10
5	am:East_Midlands	1046682.0e0	1.645527E8	3.5251563955E10

Extraktion der Daten und Import in einen Triple Store

5. SPARQL-Anfrage an den Triple Store *Apache Fuseki*

- Vergleich von Kriminalitätsvorkommen zur Bevölkerung – Abfrage:

```
1 PREFIX am: <http://www.imn.htwk-leipzig.de/~amatthes/semweb/schema#>
2
3 SELECT ?CrimeAmount ?PopAmount
4 WHERE{
5   ?crime am:hasCrimeDate ?Date .
6   ?pop am:hasPopDate ?Date .
7   ?unemp am:hasUnempDate ?Date .
8   ?crime am:hasCrimeCounty ?County .
9   ?pop am:hasPopCounty ?County .
10
11   ?unemp am:unempInRegion ?Region .
12   ?County am:countyInRegion ?Region .
13
14   ?crime am:hasCrimeAmount ?CrimeAmount .
15   ?pop am:hasPopAmount ?PopAmount .
16   ?unemp am:hasUnempAmount ?UnempAmount .
17 }
```

Extraktion der Daten und Import in einen Triple Store

5. SPARQL-Anfrage an den Triple Store *Apache Fuseki*

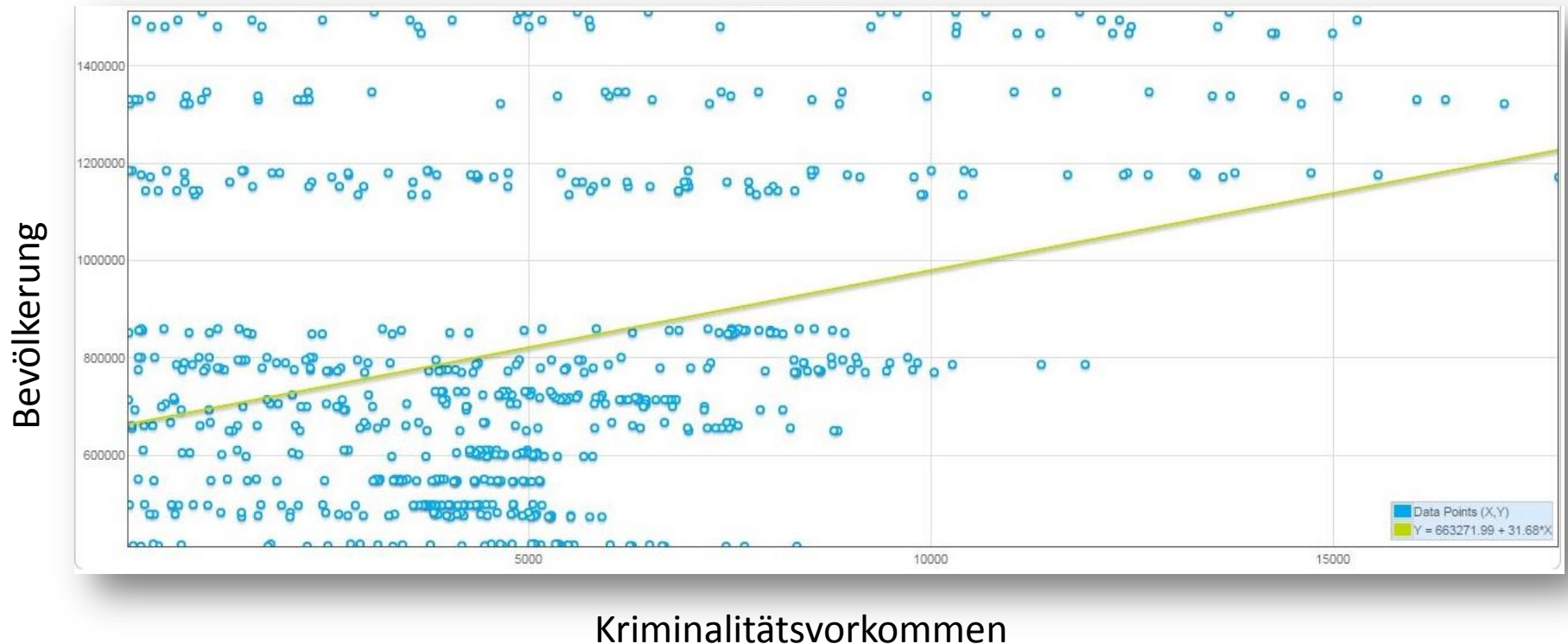
- Vergleich von Kriminalitätsvorkommen zur Bevölkerung – Ergebnis:

	CrimeAmount	PopAmount
1	8549.0	860200.0
2	6634.0	779800.0
3	606.0	714400.0
4	13701.0	1510400.0
5	908.0	801400.0
6	6689.0	667900.0
7	40.0	1184700.0
8	5334.0	418300.0

Extraktion der Daten und Import in einen Triple Store

5. SPARQL-Anfrage an den Triple Store *Apache Fuseki*

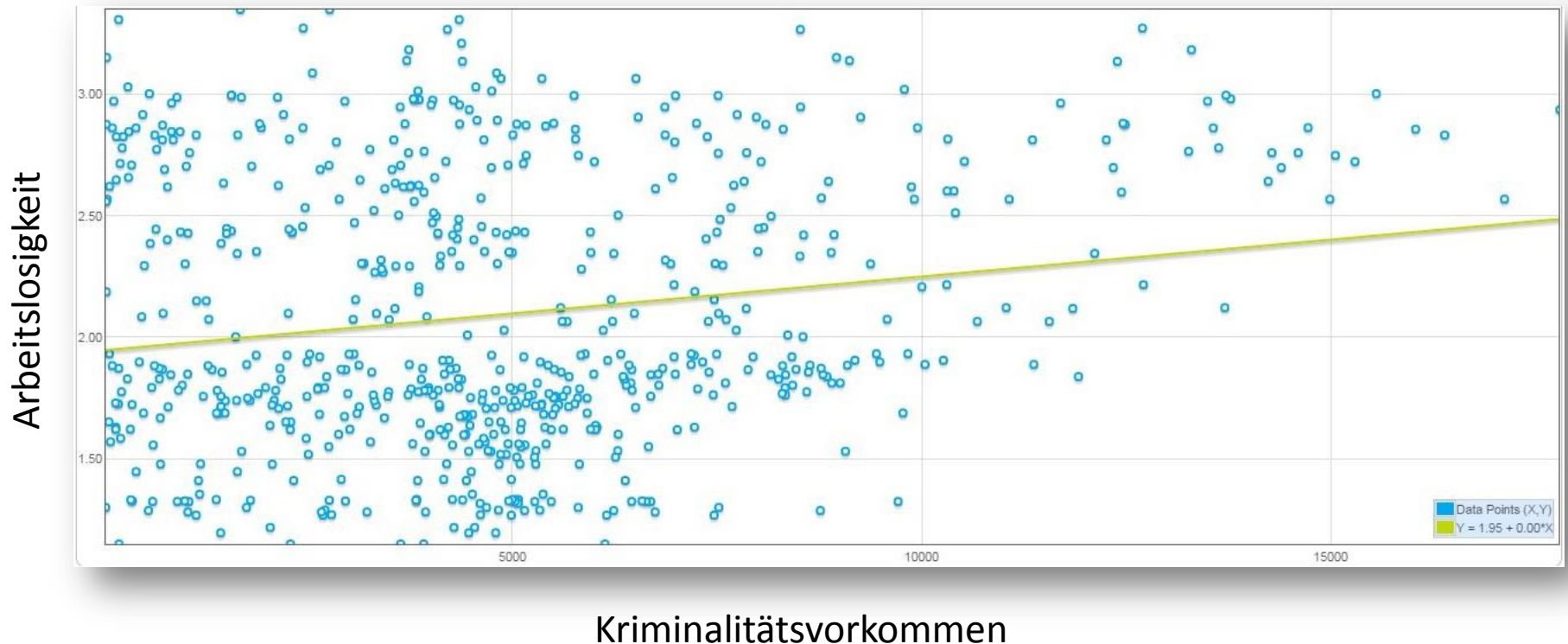
- Vergleich von Kriminalitätsvorkommen zur Bevölkerung – Ergebnis:



Extraktion der Daten und Import in einen Triple Store

5. SPARQL-Anfrage an den Triple Store *Apache Fuseki*

- Vergleich von Kriminalitätsvorkommen zur Arbeitslosigkeit – Ergebnis:



Extraktion der Daten und Import in einen Triple Store

Interpretation der Ergebnisse

Einfluss der Bevölkerung

- nur gering ableitbar:
 - Häufungen in zwei Bereichen (weniger Orte mit hoher Bevölkerung)
 - je höher die Bevölkerung, desto höher die Kriminalität
- Linienstruktur, da sich die Bevölkerungswerte (Jahresdurchschnittswert) pro Ort über den untersuchten Zeitraum nur gering ändern

Einfluss der Kriminalität

- Ableitung:
 - je höher die Arbeitslosigkeit, desto höher die Kriminalität
 - hohe Streuung

A graphic consisting of three overlapping teal rectangles. The central rectangle is the darkest and contains the text. It is flanked by two lighter teal rectangles, one on the left and one on the right, which overlap the central one.

Vielen Dank!