

Projektdokumentation im Modul Semantic Web

Ursachen für chronisch entzündliche Darmerkrankungen

Philipp Anders

31.07.2016

Recherchefragestellung: Was hat einen größeren Einfluss auf die Neuerkrankungsrate von chronisch entzündlichen Darmerkrankungen in Europa: die Ernährung oder die Psyche?

1 Inhaltliche Interpretation der Fragestellung

Die Beantwortung dieser Fragestellung befindet sich im Kontext der epidemiologischen Forschung im Bereich der Rheumatologie. In diesem Forschungsfeld werden die Verbreitung, mögliche Ursachen sowie Folgen chronisch entzündlicher Erkrankungen studiert. Der konkrete Untersuchungsgegenstand dieser Arbeit ist die **chronisch entzündliche Darmerkrankung (CED)**; engl. inflammatory bowel disease **IBD**), deren Verbreitung in den Industrieländern während der letzten 50 Jahren enorm zugenommen hat. Das Erkrankungsbild ist von episodischen Entzündungszuständen des Verdauungstraktes geprägt, welche sich in einer Fehlregulation der Immunantwort begründet. Es existieren bisher zwar lindernde, jedoch keine heilenden Therapien. Die Erkrankung ist nicht akut lebensbedrohlich, stellt durch Folgeerscheinungen wie Darmkrebs, Osteoporose und die vielschichtige psychische Belastung der Patienten dennoch einen ernstzunehmenden Untersuchungsgegenstand dar (Fal15).

Es werden zwei Erkrankungsformen der CED unterschieden: **Morbus Crohn (MC)**, engl. crohn's disease **CD**) und **Colitis ulcerosa (CU)**, engl. ulcerative colitis **UC**). Während bei CU lediglich der Dickdarm betroffen ist, wirkt sich die Erkrankung bei MC auf den gesamten Verdauungstrakt aus. Sowohl im therapeutischen als auch epidemiologischen Bereich werden die beiden Erkrankungen separat betrachtet. Momentaner Kenntnisstand

bezüglich der Erkrankungsursachen sind Mutationen des NOD2-Gens in Verbindung mit bestimmten, bis jetzt unbekannten Umweltfaktoren. Zu den wohl wahrscheinlichsten dieser Faktoren zählen die Ernährungsgewohnheiten des Patienten sowie psychische Belastung und Stress (Fal15).

Ziel der Arbeit ist die Untersuchung, welcher dieser beiden Umweltfaktoren den größeren Einfluss auf die Erkrankung an CED besitzt. Hierzu wird eine Übersicht mit Kennzahlen zur jeweiligen Verbreitung von CED, der Ernährungsgewohnheiten sowie dem Grad der mentalen Gesundheit in den Ländern Europas erstellt. Auftretende Korrelationen können dann zur Interpretationen von Zusammenhängen herangezogen werden.

2 Relevante Datenquellen

Im Beobachtungsraum Europa müssen für die einzelnen Länder quantifizierte Kennwerte für die *CED-Verbreitung*, die *Ernährungsgewohnheiten* sowie dem Grad der *mental*en *Gesundheit* gefunden werden. Vor der Besprechung der möglichen Datenquellen werden zunächst einige konzeptionelle Vorüberlegungen angestellt.

Zur Darstellung der Verbreitung von Krankheiten existiert in der Epidemiologie ein statistischer Kennwert, der als *Inzidenz* bezeichnet wird. Dieser drückt die Wahrscheinlichkeit des Auftretens einer Erkrankung aus. Die *Inzidenzrate* ist hierbei eine Spezialform der Inzidenz, welche diesen Wert auf die „Anzahl neuer Krankheitsfälle pro 100.000 Einwohner pro Jahr“ normiert. Zur Quantifizierung der CED-Verbreitung sowie der mentalen Gesundheit eignet sich somit die Ermittlung der Inzidenzraten von Krankheitsbildern, die mit diesen Faktoren in Verbindung stehen.

Die Ernährungskultur hingegen ist nicht durch die Auftretenswahrscheinlichkeit bestimmter Ereignisse zu messen. Unter der Annahme, das Fast-Food als „ungesund“ einzustufen ist, birgt ein hoher Anteil von Fast-Food Restaurants in einem Land Potential für eine ungesunde Ernährungskultur. Unter den Gesetzen des freien Marktes kann ein solches Ernährungsangebot nur bei entsprechender Nachfrage bestehen. Analog zur Inzidenzrate wird die Ernährungskultur im Rahmen dieser Arbeit durch die Angabe der „Fast-Food Restaurants pro 100.000 Einwohner“ quantifiziert. Dieser Wert wird im Folgenden als **Dichte der Fast-Food Restaurants** bezeichnet.

2.1 Liste aller europäischen Städte

Die populäre Online-Enzyklopädie Wikipedia enthält über 35 Millionen Artikel in 286 verschiedenen Sprachen. Im Kontext der Fragestellung ist der englischsprachige Artikel über Europa relevant. Er enthält eine tabellarische Aufstellung aller europäischen Länder

mit deren jeweiligen Einwohnerzahlen, welche später zur Relativierung der Anzahl der Fast-Food Restaurants relevant werden.

Link	https://en.wikipedia.org/wiki/Europe
Datenformat	HTML
Schnittstelle	HTTP
Lizenz	CC BY-SA 3.0
Open Data	★★

2.2 Daten zur CED-Verbreitung

ScienceDirect ist eine Volltextdatenbank für wissenschaftliche Veröffentlichungen, die durch den niederländischen Verlag *Elsevier* betrieben wird. Sie enthält über 14 Millionen Publikationen aus über 3.800 begutachteten Journalen und 35.000 Büchern. Über 250.000 dieser Artikel sind dabei frei zugänglich (Els16).

Das „Journal for Crohn’s and Colitis“ ist spezialisiert auf Veröffentlichungen zum Thema CED und Teil des Open-Access Bereichs von ScienceDirect. 2013 ist darin die Metastudie „The burden of inflammatory bowel disease in Europe“ erschienen, welche die Ergebnisse der unterschiedlichen Datenerhebungen zur CED-Inzidenz in Europa zusammenträgt. Die einzelnen Studien werden in einer Tabelle subsummiert, aus welcher die MC- sowie CU-Inzidenzraten der jeweiligen Länder Europas unter Angabe des Beobachtungszeitraums abgelesen werden können. Die Tabelle kann als .CSV-Datei heruntergeladen werden.

Link	http://www.sciencedirect.com/science/article/pii/S1873994613000305
Datenformat	HTML / CSV
Schnittstelle	HTTP / CSV-Parser
Lizenz	CC BY-NC-ND 4.0
Open Data	★★

2.3 Daten zur Repräsentation der mentalen Gesundheit

Die World Health Organization (WHO) ist die Sonderorganisation der Vereinten Nationen, welche für das internationale öffentliche Gesundheitswesen zuständig ist. Zu ihren Aufgaben zählt unter anderem die Erhebung und Analyse von Krankheitsdaten der 194 Mitgliedstaaten (Wor16b). Die WHO bietet mit „Athena“ einen eigene API für den maschinellen Zugriff auf eine Auswahl der statistischen Veröffentlichungen an. Dabei sind vielfältige Anpassungen bezüglich der Repräsentationen der Ergebnisse möglich (Datenformat, Länderfilter, etc.). Die Daten dürfen für den privaten Gebrauch sowie für Bildungszwecke ohne Genehmigung genutzt werden (Wor16a).

In der momentanen Version 1.0 ist unter <http://apps.who.int/gho/athena/api/GHO> eine Übersicht aller über die API abrufbaren Studienergebnisse im XML-Format erhältlich. In der Kategorie „Substance use and mental health“ werden dabei Veröffentlichungen gelistet, die als Datengrundlage zur Repräsentation der mentalen Gesundheit eines Landes im Kontext der Fragestellung zur Auswahl stehen. Im Rahmen dieser Arbeit wird sich hierbei für die **Suizid-Inzidenz** entschieden, da sich die anderen Studien in diesem Gebiet vor allem mit psychischen Krankheiten im Zusammenhang mit Drogenmissbrauch beschäftigen. Die Suizid-Inzidenz ist über den Endpunkt nach Ländern und Geschlecht aufgeschlüsselt. Die Daten wurden 2012 veröffentlicht.

Link	http://apps.who.int/gho/athena/api/GHO/MH_12?profile=simple
Datenformat	XML / JSON / CSV / XLS
Schnittstelle	Rest-API
Lizenz	eigene
Open Data	★ ★ ★

2.4 Daten zur Repräsentation der Ernährungskultur

Open Street Map (OSM) sammelt frei nutzbare Geodaten aus der ganzen Welt und macht diese über ein Nutzerinterface zugänglich, welches mit dem von Google Maps vergleichbar ist. Die verzeichneten Orte werden in Knoten, Wege und Relationen unterschieden. Sie können mit sogenannten Tags ausgezeichnet werden, welche Metadaten als Schlüssel-Wert Paare repräsentieren. Die Tag-spezifische Suche kann entsprechend definierter Bounding-Boxes auf bestimmte Bereiche eingegrenzt werden.

Das Overpass-API Projekt stellt eine Read-Only Webschnittstelle mit eigener Query-Language zum maschinellen Auslesen der OSM-Daten bereit. Die Schnittstelle ermöglicht die automatisierte Abfrage von Orten mit bestimmten Tags. Zur Repräsentation der Ernährungskultur werden Fast-Food Restaurants gesucht. Orte, die Fast-Food verkaufen, werden bei OSM mit dem Tag „amenity=fast_food“ ausgezeichnet. Somit kann eine Liste aller Fast-Food Restaurants im Bereich der Staatsgrenzen der jeweiligen Länder Europas abgefragt werden.

Link	http://www.overpass-api.de/
Datenformat	JSON
Schnittstelle	Rest-API
Lizenz	Open Database License
Open Data	★ ★ ★ ★

3 Extraktion relevanter Daten und Import in einen Triplestore

Zur Extraktion der Daten wurde eine Java-Anwendung unter Verwendung von Bibliotheken zum Parsen der unterschiedlichen Austauschformate entwickelt. Daraufhin wurden die Daten mit Hilfe des *Apache Jena* Frameworks in ein RDF-Format konvertiert und in den Triplestore *Jena Fuseki* importiert. Das Projekt mit den resultierenden Daten ist unter <https://github.com/philderfuchs/semantic-web> einsehbar.

3.1 Extraktion der europäischen Länder

Zum Parsen der Wikipedia-Seite mit den Ländern Europas wurde die Java-Bibliothek *JSOUP* verwendet. Via Iteration über die einzelnen Zeilen werden die relevanten Daten extrahiert. Hierzu müssen die Bedeutungen der jeweiligen Spalten zuvor manuell ermittelt werden. Listing 1 zeigt auszugsweise, wie die Tabelleninformationen genutzt werden, um eine Liste von **Country**-Objekten zu erstellen. Diese werden im Anschluss für den Import in den Triplestore ins RDF-Format konvertiert.

Listing 1: Auszug des Codes zur Extraktion der Länder

```
ArrayList<Country> countries = new ArrayList<>();
Document doc = Jsoup.connect("https://en.wikipedia.org/wiki/Europe").get();
Element countryTable = doc.getElementsByTag("tbody").get(1);
for (Element row : countryTable.getElementsByTag("tr")) {
    // [...]

    countries.add(new Country(row.child(2).text().split("\\s\\[") [0],
        Integer.parseInt(row.child(4).text().replace(", ", ""))));
}
```

Hierfür wurde das terminologische (d.h. Schema-)Wissen über die jeweiligen Länder zunächst in Form einer Ontologie mit Hilfe des *Apache Jena* Frameworks modelliert. Daraufhin wurde das assertionale (d.h. Fakten-)Wissen durch Instanziierung der Länderressourcen

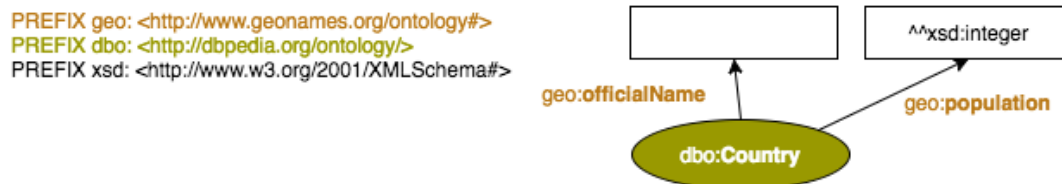


Abbildung 1: Ontologie für Instanzen der Länder Europas

entsprechend der Informationen aus der **Country**-Liste hinzugefügt. Abbildung 1 zeigt die konkrete Modellierung. Als zugrundeliegende Klasse wurde **Country** aus dem DBpedia-Namespace verwendet. Die Datatype-Properties **officialName** sowie **population** wurden der Geonames-Ontology entnommen, welche ein Vokabular zur Beschreibung ortsbezogener Daten bereitstellt. Die Standardvokabulare wurden zur kontextübergreifenden semantischen Auszeichnung der Daten verwendet, jedoch nicht in den Triplestore importiert. Als ID der **dbo:Country**-Instanzen wurde die URI des jeweiligen Wikipedia-Artikels des Landes verwendet.

3.2 Extraktion der CED-Inzidenzen

Zur Extraktion der CED-Inzidenzen wurde die heruntergeladene *.CSV*-Datei mittels der Java **Scanner**-Klasse eingelesen und verarbeitet. Der Prozess ist analog zur Extraktion der Länder Europas. Durch die zeilenweise Iteration und Kenntnis der Bedeutung der Spalten werden die relevanten Informationen ausgelesen, in einer Liste zwischengespeichert und daraufhin ins RDF-Format konvertiert.

Jede Tabellenzeile stellt eine Studie zur MC- bzw. CU-Inzidenzrate in einer bestimmten Region eines Landes und einem bestimmten Zeitintervall dar (Start- und Endjahr der Datenerhebung gegeben). Dabei sind einige Länder Europas mehrmals, andere gar nicht vertreten. Die Tabelle enthält auch separat gelistete Studien aus dem gleichen Erhebungsintervall, jedoch unterschiedlichen Regionen des gleichen Landes. Dieser Sachverhalt wird bei der SPARQL-Anfrage in Kapitel 5.1 berücksichtigt.

Abbildung 2 visualisiert die Modellierung des terminologischen Wissens über die CED-Studien. Da nicht alle Studien sowohl Daten zur MC- als auch UC-Inzidenz besitzen, wurde entschieden, die Datenerhebungen der beiden Krankheitsbilder als separate Subtypen **pa:CdStat** (Crohn's disease) bzw. **pa:UcStat** (Ulcerative colitis) des gemeinsamen Super-

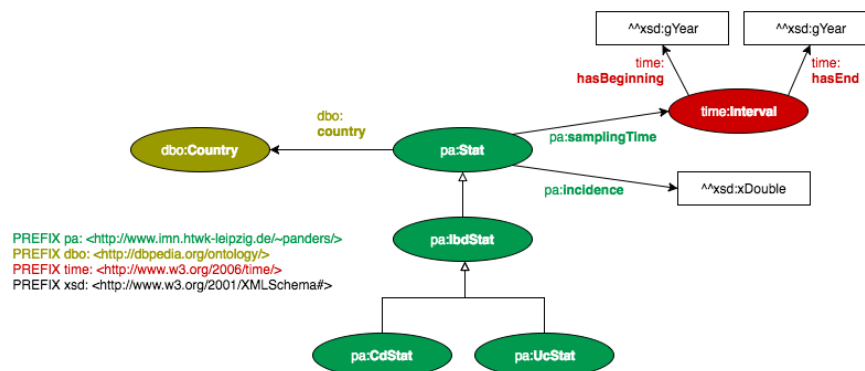


Abbildung 2: Ontologie für Instanzen der CED-Studien

typs `pa:IbdStat` (Inflammatory bowel disease) zu modellieren. Dieser leitet wiederum vom Supertyp `pa:Stat` ab, welcher bei der Extraktion der Suizid-Statistiken nochmals relevant wird. Jede Statistik besitzt das Erhebungsland als Objecttype-Property `dbo:country` und die Inzidenzrate als Datatype-Property `pa:incidence`. Das Erhebungsintervall wurde entsprechend der Empfehlungen der `time`-Ontologie des W3C zum Ausdrücken zeitlicher Bezüge modelliert. Jedoch wurden die externen Ontologien auch hier nicht zusätzlich in den Triple-Store importiert.

3.3 Extraktion der Suizid-Inzidenzen

Zur Verarbeitung der XML-Antwort der WHO *Athena*-API wurde die *JavaX-XML* Bibliothek verwendet. Jede Datenerhebung zur Suizid-Inzidenz eines Landes wird dabei unter jeweils einem `<Fact>`-Knoten gelistet. Listing 2 zeigt einen Ausschnitt des Codes, bei welchem die Extraktion durch Iteration über alle `<Fact>`-Knoten gelöst wird. Dabei wurden nur die Studien über beide Geschlechter berücksichtigt, da bei den CED-Studien keine geschlechterspezifische Auswertung vorliegt.

Listing 2: Auszug des Codes zur Extraktion der Suizid-Inzidenzen

```
ArrayList<SuicideRateStudy> suicideRateStudies = new ArrayList<>();
DocumentBuilder dBuilder =
    DocumentBuilderFactory.newInstance().newDocumentBuilder();
Document doc =
    dBuilder.parse("http://apps.who.int/gho/athena/api/GHO/MH_12?profile=simple");
doc.getDocumentElement().normalize();
NodeList nList = doc.getElementsByTagName("Fact");

for (int i = 0; i < nList.getLength(); i++) {
    Node nNode = nList.item(i);
    if (nNode.getNodeType() == Node.ELEMENT_NODE) {
        Element eElement = (Element) nNode;
        String sex =
            eElement.getElementsByTagName("SEX").item(0).getTextContent();
        if (sex.equals("Both sexes")) {
            suicideRateStudies.add(
                new SuicideRateStudy(
                    eElement.getElementsByTagName("COUNTRY").item(0).getTextContent(),
                    Double.parseDouble(eElement.getElementsByTagName("Numeric").item(0).getTextContent())
                ));
        }
    }
}
```

Abbildung 3 zeigt die Modellierung des terminologischen Wissens über die Suizidstatistiken. Analog zu `pa:IbdStat` bildet `pa:ScdStat` eine Ableitung der Oberklasse `pa:Stat`.

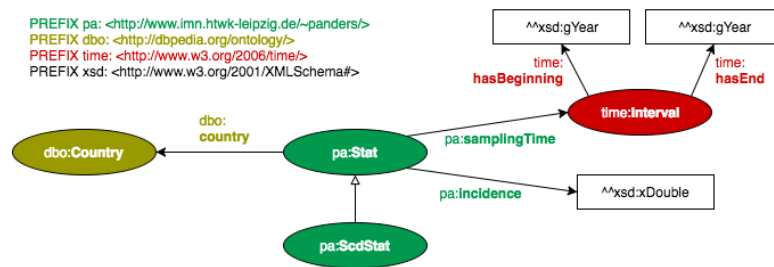


Abbildung 3: Ontologie für Instanzen der Suizid-Studien

Da die individuellen Erhebungszeiträume nicht veröffentlicht wurden, wurde das Erhebungsintervall für alle Suizid-Inzidenzen entsprechend des Jahres der Datenpublikation standardmäßig auf 2012 bis 2012 festgesetzt.

3.4 Extraktion der Fast-Food-Restaurants

Listing 3 zeigt die Anfrage an die Overpass-Schnittstelle in Form der overpass-spezifischen Query-Language. Dabei werden alle Orte angefordert, die mit dem Tag „amenity=fast_food“ ausgezeichnet sind. Der Code muss URL-escaped an die Adresse „http://www.overpass-api.de/api/interpreter?“ angehängt werden. An der Stelle <GEOCODE> in der Query ist auf der Overpass-Weboberfläche die Eingabe der englischen Bezeichnung des gewünschten Suchbereichs vorgesehen, welche dann intern auf Clientseite zur ID der korrespondierenden Bounding-Box (eine Relation von Linien) aufgelöst wird. Bei der programmatischen Anfrage via URI muss diese ID jedoch zuvor ermittelt werden, indem in einer extra Abfrage die ID derjenigen Relation ermittelt wird, die den Namen des jeweiligen Landes trägt.

Listing 3: Query zur Abfrage aller Fast-Food Restaurants eines Landes

```
[out:json][timeout:25];
{{geocodeArea:<GEOCODE>}}->.searchArea;
( node["amenity"="fast_food"](area.searchArea); );
out body;
```

Die Antwort im JSON-Format repräsentiert die Fast-Food Restaurants als Array von JSON-Objekten. Jedes enthält neben der ID und den Geo-Koordinaten alle Tags des jeweiligen Ortes. Die Daten wurden mit Hilfe der *Google Gson*-Bibliothek verarbeitet und anschließend ins RDF-Format konvertiert. Abbildung 4 zeigt die Modellierung des terminologischen Wissens über die Fast-Food Restaurants. Abermals wurden Standardvokabulare zur kontextübergreifenden Auszeichnung der Daten verwendet. Die gigantische Meta-Wissensbasis *yago* des Max Planck-Instituts, die Wissen aus DBpedia und anderen



Abbildung 4: Ontologie für Instanzen der Fast-Food Restaurants

Quellen vereint, besitzt hierbei bereits eine Standardklasse für Fast-Food Restaurants, welche als Basis der Modellierung genutzt wurde (Max16).

4 Verlinkung von Ressourcen

Abbildung 5 zeigt die Integration der separaten Wissensbasen zu einem gemeinsamen RDF-Graphen, der das gesamte Schema- sowie Instanzwissen in 265.956 Tripeln beinhaltet. Die Verlinkung geschieht über das gemeinsame Objecttype-Property `dbo:Country`. Die Tripel, die den Zusammenhang herstellen, sind hierbei hellrot umrandet. Der konsistente Einsatz der `dbo:Country`-Ressourcen wurde hierbei bereits zum Zeitpunkt des RDF-Exports sichergestellt. Uneinheitliche Bezeichnungen wurden vorausschauend identifiziert und mittels einer selbstgeschriebenen Klasse behandelt, die programmatisch die Umwandlung zu Standardbezeichnungen realisiert. Hierbei mussten nur wenige Landesnamen angepasst werden (z.B. „UK“ zu „United Kingdom“).

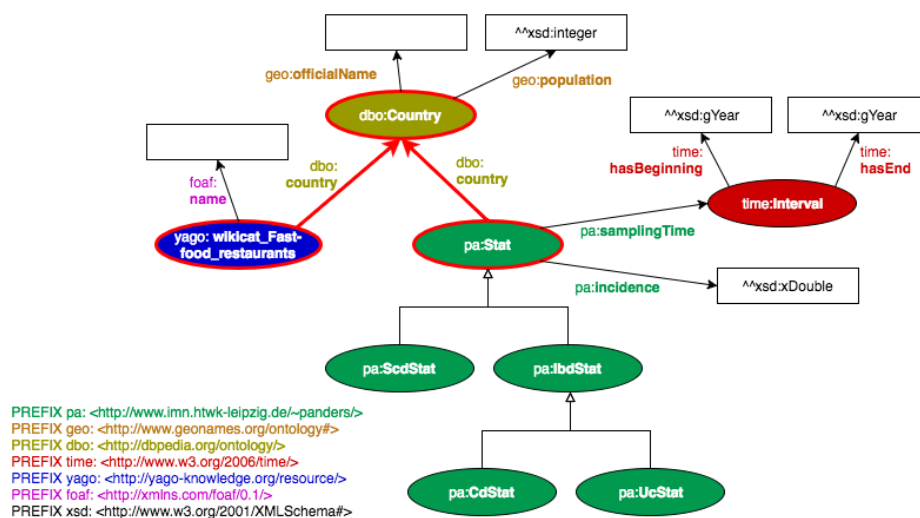


Abbildung 5: Verlinkung der separaten Wissensbasen

5 Anfrage an die Forschungswissensbasis

5.1 SPARQL-Anfrage

Listing 4 zeigt die SPARQL-Anfrage, welche im Triple-Store *Jena Fuseki* durchgeführt wurde. Besonders ist dabei die Behandlung der CED-Studien aus unterschiedlichen Erhebungszeiträumen. Zu einigen Ländern existieren aus der Metastudie aktuelle sowie ältere Inzidenzdaten. Um eine möglichst gute Vergleichbarkeit mit den aktuellen Suizid- sowie Fast-Food-Daten zu gewährleisten, werden nur die neuesten MC- bzw. CU-Studien bezüglich des Endjahres der jeweiligen Datenerhebung berücksichtigt. Sollten mehrere Studien in einem Land mit dem gleichen Endjahr vorliegen, wird der Durchschnitt der entsprechenden Inzidenzwerte ermittelt. Die resultierenden MC- sowie CU-Mittelwerte werden zur Bildung einer finalen CED-Inzidenzrate addiert, um der ursprünglichen Fragestellung gerecht zu werden. Zur Bildung der Dichte der Fast-Food Restaurants wird deren relative Anzahl pro 100.000 Einwohner gebildet, in dem die Summe aller Fast-Food Restaurants eines Landes durch deren Bevölkerungszahl geteilt und anschließend mit 100.000 multipliziert wird.

Listing 4: SPARQL-Anfrage an die Wissensbasis.

```
PREFIX pa: <http://www.imn.htwk-leipzig.de/~panders/>
PREFIX geo: <http://www.geonames.org/ontology#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX time: <http://www.w3.org/2006/time/>
PREFIX yago: <http://yago-knowledge.org/resource/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?n ((?countOfFFVenues / ?pop) * 100000 as ?Fast-Food_Restaurants) (?sR as
    ?Suizid_Inzidenz) ((?cdAverage + ?ucAverage) as ?CED_Inzidenz)
WHERE {
    ?c geo:population ?pop.
    ?c geo:officialName ?n.
    ?s a pa:ScdStat.
    ?s dbo:country ?c.
    ?s pa:incidence ?sR.
    {
        SELECT ?c (count(?s) as ?countOfFFVenues) ?cdAverage ?ucAverage
        WHERE {
            ?s a yago:wikicat_Fast-food_restaurants.
            ?s dbo:country ?c.
            {
                SELECT ?c (AVG(?cdR) as ?cdAverage) (AVG(?ucR) as ?ucAverage)
                WHERE {
                    ?i a pa:CdStat.
                    ?i dbo:country ?c.
                    ?i pa:samplingTime ?cdSt.
                    ?cdSt time:hasEnd ?cdLatest.
```

```

        ?i pa:incidence ?cdR.

        ?j a pa:UcStat.
        ?j dbo:country ?c.
        ?j pa:samplingTime ?ucSt.
        ?ucSt time:hasEnd ?ucLatest.
        ?j pa:incidence ?ucR.
        {
            SELECT ?c (MAX(?cdEnd) as ?cdLatest) (MAX(?ucEnd) as ?ucLatest)
            WHERE {
                ?i a pa:CdStat.
                ?i dbo:country ?c.
                ?i pa:samplingTime ?cdSt.
                ?cdSt time:hasEnd ?cdEnd.

                ?j a pa:UcStat.
                ?j dbo:country ?c.
                ?j pa:samplingTime ?ucSt.
                ?ucSt time:hasEnd ?ucEnd.
            }
            GROUP BY ?c
        }
    }
    GROUP BY ?c
}
GROUP BY ?c ?cdAverage ?ucAverage
}
}
ORDER BY DESC(?IbdRate)

```

5.2 Ergebnis der Anfrage

Ergebnis der Anfrage ist eine Tabelle, die für jedes Land Europas dessen jeweilige CED-Inzidenz pro Jahr pro 100.000 Einwohner (**CED-Inzidenzrate**), Suizid-Inzidenz pro Jahr pro 100.000 Einwohner (**Suizid-Inzidenzrate**) sowie die Anzahl der Fast-Food-Restaurants pro 100.000 Einwohner (**Dichte der Fast-Food Restaurants**) angibt. Dabei werden nur Länder aufgeführt, für die eine sinnvolle CED-Inzidenzrate gebildet werden konnte (sowohl Studien zur MC- als auch zur CU-Inzidenzrate vorhanden). Abbildung 6 visualisiert das Ergebnis geordnet nach aufsteigender CED-Inzidenzrate. Die genauen Werte sind Tabelle 1 im Anhang zu entnehmen.

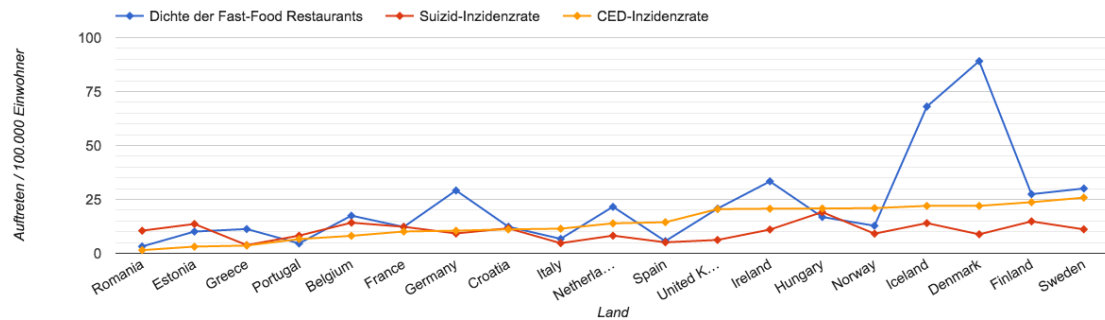


Abbildung 6: Ergebnisdarstellung der SPARQL-Anfrage (unkonsolidiert)

6 Interpretation und Zusammenfassung

Augenscheinlich ist keine Korrelation zwischen der CED-Inzidenz, der Suizid-Inzidenz sowie der Dichte der Fast-Food Restaurants pro 100.000 Einwohner feststellbar. Dies wird durch eine Korrelation als Ähnlichkeitsmaß von lediglich 0.23 bezüglich der CED-Suizid-Korrelation sowie 0.58 bezüglich der CED-Fast-Food-Korrelation bestätigt (Werte gegen 1.0 bzw. -1.0 stellen interpretierbare Korrelationen dar). Auffällig sind die Ausreißer bezüglich der Dichte der Fast-Food Restaurants von Island und Dänemark. Die Konsolidierung der Datenbasis durch Bildung des Konfidenzintervalls aus Mittelwert und 2.5-facher Standardabweichung entfernt beide Länder aus der Betrachtung und führt zu einer CED-Suizid-Korrelation von 0.22 und einer CED-Fast-Food-Korrelation von 0.66. Somit spricht das Ergebnis auf Grundlage dieser Datenbasis aus Perspektive der statistischen Kennwerte eher für einen Zusammenhang zwischen der Ernährungskultur und CED, auch, wenn dieser nicht augenscheinlich ist. Dabei sei hervorgehoben, dass aus einer Korrelation nicht zwangsläufig auf eine Kausalität geschlossen werden kann.

Auf Grundlage dieser Datenbasis konnte somit keine eindeutige Korrelation festgestellt werden. Ursache hierfür könnte die falsche Wahl der Einflussfaktoren (Ernährung und Psyche), die falsche Wahl deren Repräsentation (Suizidrate und Fast-Food Restaurants) und/oder nicht vertrauenswürdige Datenbasen sein. Beispielsweise basiert die Auszeichnung der Fast-Food Restaurants bei Open Street Map auf der lokalen Aktivität sowie der Einschätzung der Community und ist somit in ihrer Repräsentativität anzweifelbar. Auch sind einige Differenzen der Erhebungszeiträume von CED- sowie Suizid-Inzidenzen mit bis zu 20 Jahren relativ hoch, obwohl nur die aktuellsten Studien betrachtet wurden.

Zusammenfassend kann im Rahmen dieser Arbeit bezüglich eines möglichen ursächlichen Zusammenhangs keine Aussage gemacht werden, wobei jedoch eine Tendenz zur Ernährungskultur argumentierbar ist. Dies bestätigt der Stand der Forschung, bei dem die Psyche als beeinträchtigender Faktor für den Krankheitsverlauf gilt, dieser jedoch keine ursächliche Rolle zur Entwicklung des Krankheitsbildes zugesagt wird (Fal15).

Literatur

- [Els16] Elsevier B.V. About sciencedirect, 2016. Online: <https://www.elsevier.com/solutions/sciencedirect>.
- [Fal15] Falk Foundation e.V. Colitis ulcerosa und morbus crohn. eine Übersicht über die krankheitsbilder und ihre behandlung., 2015. Online: https://www.drfalkpharma.de/uploads/tx_tocfpshoperw/S80_56-2-15.pdf.
- [Max16] Max Planck Institute for Informatics. Yago: A high-quality knowledge base., 2016. Online: <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>.
- [Wor16a] World Health Organization. About who. permissions and licensing., 2016. Online: <http://www.who.int/about/licensing/en/>.
- [Wor16b] World Health Organization. About who. what we do., 2016. Online: <http://www.who.int/about/what-we-do/en/>.

Anhang

Tabelle 1: Ergebnis der SPARQL-Abfrage

Land	CED-Inzidenzrate	Suizid-Inzidenzrate	Dichte der Fast-Food Restaurants
Romania	1.47	10.50	3.18
Estonia	3.10	13.60	10.07
Greece	3.60	3.80	11.26
Portugal	6.60	8.20	4.59
Belgium	8.10	14.20	17.47
France	10.10	12.30	12.23
Germany	10.50	9.20	29.15
Croatia	11.10	11.60	12.33
Italy	11.43	4.70	6.72
Netherlands	13.90	8.20	21.59
Spain	14.40	5.10	5.64
United Kingdom	20.50	6.20	20.74
Ireland	20.70	11.00	33.34
Hungary	20.80	19.10	16.85
Norway	20.95	9.10	12.77
Iceland	22.00	14.00	68.02
Denmark	22.00	8.80	89.11
Finland	23.70	14.80	27.42
Sweden	25.80	11.10	30.05