

Gesamtpunktzahl: 30

Abgabe der Lösungen bis zum 16.01.2017

Anhand der folgenden Aufgaben sollen Sie lernen, eine etwas komplexere Aufgabenstellung mit den Mitteln der Logikprogrammierung zu lösen. Begründen und dokumentieren Sie in jedem Fall Ihre Annahmen und Designentscheidungen und diskutieren Sie die Vor- und Nachteile Ihrer Lösung im Vergleich zu den ggf. vorhandenen Implementationsalternativen.

**Hinweis:** Wenn Prolog die Ausgabe von Listen mit ... abkürzt, können Sie durch drücken von `w` und `p` bei der Wahl einer Alternative die vollständige Ausgabe ein- und ausschalten. Dasselbe erreichen Sie durch die Eingabe `set_prolog_flag(answer_write_options, [quoted(true), spacing(next_argument)])`.

**Aufgabe 1:** Datenimport und Formatierung

6 Punkte

maximale Bearbeitungszeit: 45 Minuten

Auf [hamburg.luftmessnetz.de](http://hamburg.luftmessnetz.de) stellt die Stadt Hamburg Statistiken zur gemessenen Luftverschmutzung in Hamburg bereit; die Daten sind im CSV-Format abrufbar.<sup>1</sup>

1. Betrachten Sie die Dateien zum Feinstaub in der Habichtstraße (Station 68HB) Anfang 2015, Ende 2015 und Anfang 2016 (2015-w01, 2015-w50, 2016-w01). Interpretieren Sie die in den CSV-Dateien enthaltene Information und stellen Sie die Verknüpfung her zu den Inhalten der entsprechenden Prolog-Dateien (XY-facts.pl).
2. Implementieren Sie ein Prädikat, das eine CSV-Datei in das Format überführt, das in `68HB-2015-w01-facts.pl` deklariert ist und verwendet wird.

**Hinweise:** Das Prolog-Modul zur Verarbeitung von CSV laden Sie mit `:- use_module(library(csv)).` Nutzen Sie `csv_read_file/3` zum Einlesen der Datei und setzen Sie die Optionen im dritten Parameter passend. `parse_time/2`

<sup>1</sup> Beispielsweise können die wöchentlichen Feinstaubdaten an der Kieler Straße unter <http://hamburg.luftmessnetz.de/station/64KS.csv?componentgroup=pollution&componentperiod=24hg&searchperiod=custom&searchfrom=START&searchuntil=ENDE> abgerufen werden, wobei START und ENDE Zeitangaben sind (zum Beispiel 01.01.2017+00:00); beachten Sie, dass mit jeder Anfrage höchstens Daten für 1 Woche geliefert werden.

kann nicht direkt mit dem genutzten Datumsformat umgehen, nutzen Sie deshalb zunächst `atom.chars/2` um die Bestandteile des Datums ins ISO-Format umzusortieren.

Ihr Prädikat sollte sowohl für Stationen mit PM10-Daten (z. B. Max-Brauer-Allee), PM2,5-Daten (Kieler Straße), als auch für Stationen die beide Messwerte liefern (Habichtstraße) funktionieren, also mit zwei- und drei-spaltigen CSVs umgehen können.

Bonus (2 Punkte): Erstellen Sie ein Prädikat, welches die Daten direkt aus dem Internet herunterlädt. Nutzen Sie hierfür das `http_client`-Modul.

Bonus (3 Punkte): Erweitern Sie Ihr Prädikat, sodass es für eine Liste von CSV-Dateinamen `pollution_data/6`-Fakten erzeugt, sodass nur pro Station und Typ ein Fakt erzeugt wird das alle vorhandenen Messwerte in einer einzigen Liste enthält. Achten Sie darauf, nicht zwei Werte zu einem Zeitpunkt (also doppelt) zu speichern und fügen Sie ggfs. leere Atome (``) ein, wenn zwischendurch keine Daten verfügbar sind.

## Aufgabe 2: Zeitreihenverarbeitung

15 Punkte

maximale Bearbeitungszeit: 180 Minuten

Die Zeitreihen in `feinstaub-2015-facts.pl` enthalten stündliche Werte (eine Granularität von  $1/(3600 \text{ Sekunden})$ ); für einen besseren Überblick wäre eine tageweise Darstellung (also Werte alle  $24 \times 3600 = 86400$  Sekunden) nützlich.

1. Implementieren Sie ein Prädikat, welches die Granularität der Daten reduziert indem aufeinanderfolgende Werte (24 Werte für Stunden  $\rightarrow$  Tage) durch ihren Mittelwert ersetzt werden; bei fehlenden Messwerten (``) bilden sie den Mittelwert aus den übrigen Messwerten, ist keinerlei Messwert verfügbar, so setzen Sie `` ein. Ihr Prädikat soll eine `pollution_data/6`-Zeitreihe sowie die neue Granularität (z. B. 86400 für einen Tag) entgegennehmen und die resultierende `pollution_data/6`-Zeitreihe der neuen Granularität produzieren.
2. Ermitteln Sie die durchschnittliche Feinstaubbelastung an allen Stationen in 2015 indem Sie die Granularität der Daten auf  $60 \times 60 \times 24 \times 365 = 31536000$  setzen. Bleibt das Jahresmittel überall im Rahmen der EU-Feinstaub-Richtlinie ( $\text{PM-10} \leq 40 \mu\text{g}/\text{m}^3$ ) und der WHO-Luftgüte-Empfehlung ( $\text{PM-10} \leq 20 \mu\text{g}/\text{m}^3$  sowie  $\text{PM-2.5} \leq 10 \mu\text{g}/\text{m}^3$ )?
3. Die EU-Feinstaub-Richtlinie erlaubt die Überschreitung des PM10-Tagesmittelwerts von  $50 \mu\text{g}/\text{m}^3$  an höchstens 35 Tagen im Jahr. Reduzieren Sie die Granularität der PM10-Daten auf Tage und zählen Sie, für die Stationen die Überschreitungen im Jahr 2015.

Die bisher genutzte Zeitreihenrepräsentation ist ‘zeilenbasiert’ für jede Messstation getrennt. Eine Mittelwertbildung für die Belastung in ganz Hamburg ist allerdings in einer ‘spaltenbasierten’ Darstellung deutlich einfacher.

4. Implementieren Sie ein Prädikat zur Matrixumformung, das eine M-elementige Liste von N-elementigen Listen in eine N-elementige Liste von M-elementigen Listen überführt:

$$\begin{bmatrix} [1, & 2, & 3, & 4, & 5] \\ [5, & 4, & 3, & 2, & 1] \\ [4, & 4, & 4, & 4, & 4] \end{bmatrix} \rightarrow \begin{bmatrix} 1, \\ 5, \\ 4 \end{bmatrix} \begin{bmatrix} 2, \\ 4, \\ 4 \end{bmatrix} \begin{bmatrix} 3, \\ 3, \\ 4 \end{bmatrix} \begin{bmatrix} 4, \\ 2, \\ 4 \end{bmatrix} \begin{bmatrix} 5, \\ 1, \\ 4 \end{bmatrix}$$

Verzichten Sie auf die Verwendung eingebauter Prädikate in Ihrer Prädikatsdefinition! Beschreiben und testen Sie Ihre Implementierung, auch im Vergleich zum eingebauten Prädikat `transpose/2`.

5. Erstellen Sie ein Prädikat, das die mittlere Luftbelastung in Hamburg als Zeitreihe (mit Bezeichnung ‘Hamburg-Mittelwert’) liefert. Nutzen Sie `findall` um die Einzelzeitreihen aus der Datenbasis in eine Liste von Listen zu lesen, und die Lösung aus 4 (oder `transpose`), mitteln Sie dann die einzelnen Spalten (behandeln Sie fehlende Messwerte (‘ ’) wie zuvor).
6. An wievielen Tagen wurde 2015 hamburgweit der PM10-Grenzwert gerissen?

### Aufgabe 3: Korrelation

9 Punkte

maximale Bearbeitungszeit: 75 Minuten

Die Korrelation ist nützlich um die Ähnlichkeit zweier Zeitreihen zu beschreiben. Der Pearson’sche Korrelationskoeffizient nimmt Werte zwischen  $-1$  und  $+1$  an, wobei Werte nahe  $0$  für keinen Zusammenhang, Werte nahe  $+1$  für eine starke Übereinstimmung und Werte nahe  $-1$  für einen starken Gegensatz stehen. Für zwei Vektoren  $\vec{x}$  und  $\vec{y}$  der Länge  $n$  berechnet er sich wie folgt<sup>2</sup>:

$$r = r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Die Korrelationsanalyse zwischen einzelnen Zeitreihen ist zum Beispiel nützlich um festzustellen, ob Verschmutzungsarten miteinander zusammenhängen.

1. Implementieren Sie ein Prädikat zur Berechnung des Pearson’schen Korrelationskoeffizienten für zwei gleichlange Zeitreihen. Wenn einer (oder beide) Werte eines Paares fehlt (‘ ’), so überspringen sie das Paar in der Berechnung.<sup>3</sup>

<sup>2</sup>Siehe [http://en.wikipedia.org/wiki/Pearson\\_product-moment\\_correlation\\_coefficient#For\\_a\\_sample](http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient#For_a_sample).

<sup>3</sup>Vgl. [http://dept.stat.lsa.umich.edu/~kshedden/Python-Workshop/correlation\\_missing\\_data.html](http://dept.stat.lsa.umich.edu/~kshedden/Python-Workshop/correlation_missing_data.html) und dort *Complete Case Analysis*.

2. Mit welcher Anfrage finden Sie die Messstation welche 2015 die mittlere Hamburger (PM10-)Feinstaubbelastung am besten repräsentiert?

Frau W. und Herr B. betreiben seit dem Herbst einen eigenen Feinstaubsensor. Dieser hängt an ihrer Hauswand und misst (ungefähr) minütlich die PM<sub>2,5</sub>- und PM<sub>10</sub>-Belastung der Luft. Der Sensor ist nicht geeicht und möglicherweise weniger verlässlich als die amtlichen Werte (er kostet dafür als Bausatz auch nur 30 Euro).<sup>4</sup>

Die minütlichen Daten des Feinstaubsensors seit Herbst (sowie die stündlichen Werte des offiziellen Messnetzes im selben Zeitraum) finden Sie in der Datei `feinstaub-herbst-2016-facts.pl`.

3. Implementieren Sie ein Prädikat, welches die minütlichen Daten des Feinstaubsensors (*mode* = '1min') zum gleitenden 24h-Mittel umrechnet indem Sie alle in den vergangenen 24h liegenden Datenpunkte (soweit vorhanden) zusammenfassen und den Mittelwert bilden ('24hg'). Im Falle fehlender Daten (') verfahren Sie wie oben (d.h. berechnen aus weniger Daten bzw. tragen '' ein wenn keinerlei Datenpunkte verfügbar sind).
4. Reduzieren Sie mittels `coarsen_granularity/3` die minütliche Zeitreihe für 'zu Hause' auf stündliche Messwerte und berechnen Sie, mit welchem der Hamburger Messpunkte der private Feinstaubsensor am besten korreliert. Ist die Korrelation mäßig (> 0.4), gut (> 0.6) oder sogar sehr gut (> 0.8)?

### Freie Bonusaufgabe:

maximal 10 Bonuspunkte

Suchen Sie sich eine geeignete weiterführende Aufgabe im Bereich Luftverschmutzung, zum Beispiel:

- Benutzen Sie `display.pl` (oder Alternativen) zur Anzeige von Zeitreihen.
- Explorieren Sie die JSON-API des Luftmessnetzes: <http://luftmessnetz.de/api/doc>.
- Links zu anderen Schadstoffdaten (und ggfs. Wetterwerte) finden Sie über "Stationsdaten" auf <http://hamburg.luftmessnetz.de> für unterschiedliche 'Messperioden'.
- Die Daten sehr vieler freier Feinstaubsensoren finden Sie auf <http://archive.luftdaten.info/> und fertig visualisiert auf <http://opendata-stuttgart.github.io/feinstaub-map/>.

Die Bepunktung richtet sich nach der Schwierigkeit und der Kreativität der Bearbeitung.

---

<sup>4</sup>Vgl. <http://luftdaten.info/> und <http://www.madavi.de/sensor/csvfiles.php?sensor=esp8266-761330>.