

Information Retrieval – Übung Boolean Retrieval

I Mengen und Sets

Gegeben ist das folgende Vokabular:

{Auto, Straße, Schild, Ampel, Verkehr, Lärm, Benzin, Diesel, Umwelt}.

Zusätzlich seien folgende Dokument gegeben:

Dok1 = {Auto, Schild, Verkehr, Benzin, Umwelt}

Dok2 = {Straße, Ampel, Lärm, Diesel}

Dok3 = {Schild, Verkehr}

Bestimmen Sie folgende Kombinationen der Bag-of-words:

1. Dok1 AND Dok2
2. Dok1 OR Dok2
3. Dok1 AND Dok3
4. Dok1 NOT Dok3
5. Dok2 OR Dok3
6. Dok2 XOR Dok3

II Term-Dokument-Matrix

Sie haben einen Dokumentkorporus, der aus vier Dokumenten besteht:

Dokument 1: population below the poverty line

Dokument 2: calculation of the poverty gap index

Dokument 3: population, poverty and economic growth

Dokument 4: the impact of economic growth on poverty

1. Erstellen Sie die Term-Dokument-Matrix (auf Papier, mit Excel, Python, egal).
2. Was sind die Ergebnisse für die folgenden beiden Anfragen? Zeigen Sie jeweils die einzelnen Rechenschritte und nicht nur das Endergebnis!
 - ~ poverty **AND** population
 - ~ the **AND NOT** (economic **OR** poverty)