TextToScene Fingerübung

Tim Fauerbach, 4631886

November 2020

Die Ausführung von fingeruebung.py zeigt trägt zunächst alle Daten der XML-Dateien aus dem Ordner Traning zusammmen in eine XML gemeinsam mit den von Spacy erzeugten PoS Daten auf dem Text. Danach wird die Auswertung ausgegeben, anschließend die Satzlänge mit der Häufigkeit geplottet und zum Schluss die Graphen für Bicycles.xml und Highlights_of_the_Prado_Museum.xml angezeigt.

1 Auswertung

Anzahl PoS Tags:
('SPACE', 825)
('PROPN', 2095)
('AUX', 1035)
('DET', 3203)
('NOUN', 5030)
('ADP', 3005)
('PUNCT', 3501)
('ADV', 1326)
('ADJ', 1781)
('VERB', 2721)
('PRON', 1362)
('NUM', 683)
('CCONJ', 825)
('SCONJ', 431)
('PART', 491)
('X', 30)
('SYM', 28)
('INTJ', 10)

Anzahl bestimmte Tags:
SpacialEntities: 1417
Places: 1852
Motions: 771
Signals: 1240

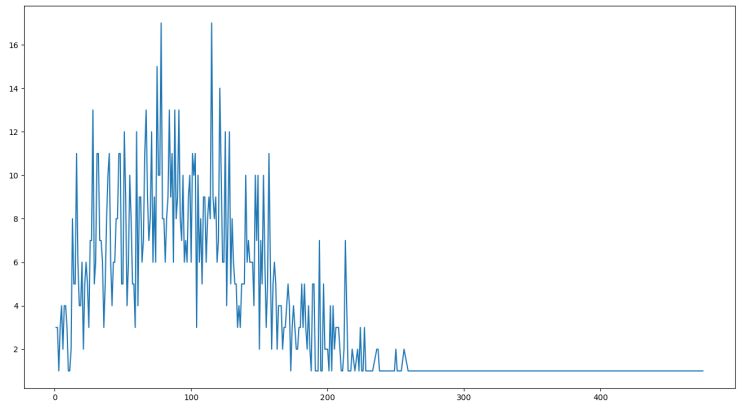

Figure 1: Satzlänge und Häufigkeit

QsLinks: 970
OLinks: 244

QsLink Typen:
('NTPP', 42)
('IN', 586)
('EC', 196)
('TPP', 53)
('EQ', 35)
('PO', 12)
('OUT', 3)
('DC', 41)
(", 2)

QSLINKS und OLINKS Trigger:

('NTPP', [", 'with', 'in', 'houses', 'inside', 'on', 'On', 'of', 'through'])
('IN', [", 'houses', 'in', 'on', 'contain', 'house', 'packed with', 'On', 'with', 'In', 'around', 'full of', 'where', 'through', 'at', 'inhabited', 'into', 'inside of', 'along', 'At', 'of', 'including', 'stocked', 'from', 'across', 'under', 'filled', 'contains', 'packed', 'has', 'inside', 'for', 'restricted'])
('EC', ['filled', 'on', '", 'with', 'adjacent to', 'surmounted', 'bordering', 'against', 'At', 'of', 'covered', 'Along', 'coiling up', 'about', 'on top of', 'On', 'where', 'around', 'at', 'outside', 'along', 'on top', 'next to', 'surrounded', 'between', 'atop', 'surrounding', 'line', 'directly beneath', 'in', 'up to', 'upon'])
('SURROUND', ['around', 'coiling up', 'surrounded', 'surrounding'])

('TPP', ['', 'with', 'At', 'part of', 'houses', 'atop', 'of', 'on'])
('EQ', ['', 'where', 'houses', 'in', 'at', 'Everywhere', 'of', 'covering'])
('SOUTH', ['south of', 'south', 'in that direction'])
('LEFT', ['', 'to the left', 'on your left'])
('TOP', ['on', 'atop', ''])
('WEST', ['West of', 'west', 'to the west from'])
('FRONT_FACE', ['facing', 'Facing'])
('PO', ['in', 'connects', 'to', 'on', ''])
('ACROSS', ['Across', 'across from', 'behind', 'in the direction to', '', 'over', 'between', 'across'])
('UP', ['up'])
('BESIDE', ['alongside', 'adjacent to', 'next door to', 'beside', 'along', 'Along', 'next to'])
('BEHIND', ['behind', 'backwards', 'on'])
('SOUTHEAST', ['Southeast of', 'Southeast of'])
('EAST', ['east of', 'east'])
('ABOVE', ['surmounted', 'on', 'covered', '', 'on top of', 'On', 'over', 'of', 'overlooking', 'on top', 'above', 'atop', 'line', 'in', 'upon', 'overlook'])
('OUT', ['', 'at'])
('BELOW', ['under', 'below', 'beneath', 'Down', 'On', 'directly beneath', 'of', 'up to'])
('NORTH', ['north of', 'north'])
('DC', ['to', 'beside', 'far from', 'away from', 'behind', 'apart from', 'next to', 'At', 'after', 'in front of', 'up', 'further', 'under', 'In front of', 'along', 'over', 'between', 'away', 'overlooking', 'at', 'surrounded', 'out of', 'afar', 'from', 'outside', 'apart', 'Down'])
('TOWARD', ['toward', 'to'])
('SOUTHWEST', ['southwest', 'to SW'])
('BOTTOM', ['up'])
('RIGHT', ['on the right'])
('FRONT', ['in front of', 'In front of'])
('NORTHEAST', ['northeast'])
('NEXT TO', ['neighboring'])
('ALONG', ['along', 'on'])
('DOWN', ['down', 'Down'])
('BETWEEN', ['between'])
('NEXT_TO', ['next to', 'along'])
('', ['on'])
('OVERLOOKING', ['overlooking'])
('AROUND', ['surrounding', 'surrounded'])
('UPSTREAM', ['upstream from'])

Bei den MOVEMENT Verben war ich unsicher, habe die 5 häufigsten MOTION verben genommen:

('biked', 22)
('visited', 21)
('biking', 16)

('bike', 15)
('go', 15)

2 Visualisierung

Ich habe das Python Paket NetworkX genutzt. Die Einträge mit dem Tag 'PLACE' habe ich als grüne Knoten und die Einträge mit dem Tag 'SPATIAL_ENTITY' als hellblaue Knoten dargestellt. QsLinks sind rot, OLinks sind lila.

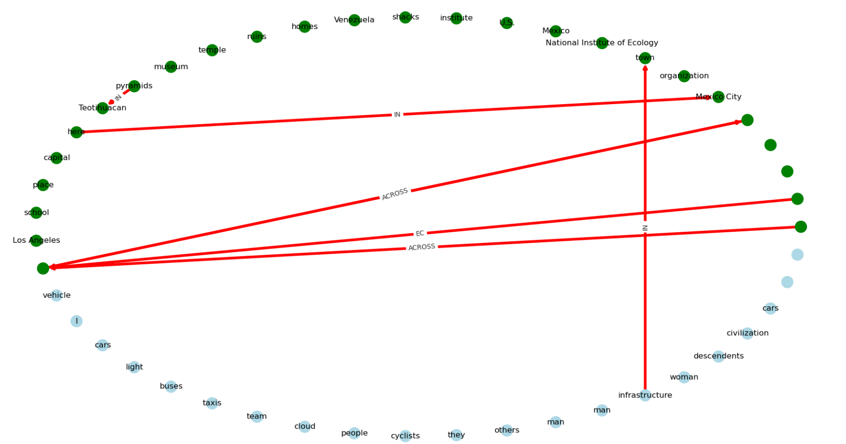

Figure 2: Bicycles.xml

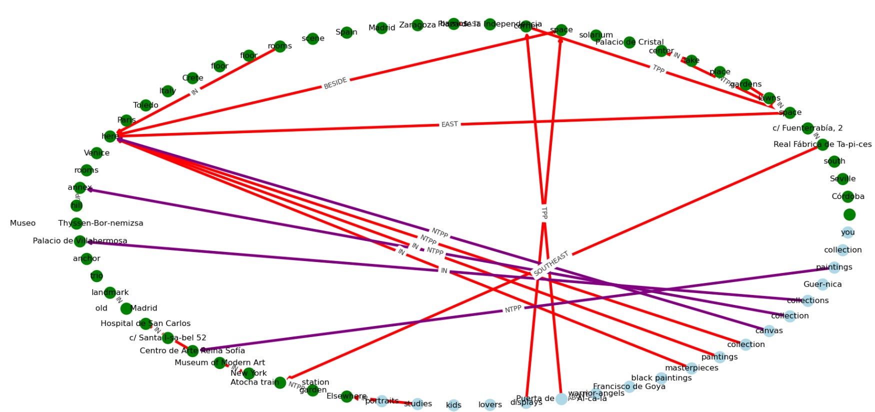

Figure 3: Highlights_of_the_Prado_Museum.xml