

# TEXT2SCENE AUSWERTUNG

Paul Schnürer (6942535)

## Voraussetzungen

Um das Modul ausführen zu können braucht es:

- Python 3.8
- Anaconda oder Miniconda

## Installation

Zunächst klonst man das Repository mit

`git clone https://github.com/Patrotus/Text2SceneSS21-Uebung2.git` und öffnet den Ordner in einer Conda-Konsole.

Mit `conda env create -name <name> -f t2s.uebung.yml` erstellt man ein neues Environment und aktiviert dieses mit `activate <name>`. Das Modul `en-core-web-sm` muss mit `python -m spacy download en_core_web_sm` anschließend noch installiert werden.

## Ausführung

Um den Analyzer schließlich durchlaufen zu lassen muss nur `python run.py` im Repository-Ordner ausgeführt werden. Anschließend läuft das Programm und analysiert die Daten im `resources` Ordner.

## Auswertung

Die ausgewerteten Daten werden im Ordner `results` gespeichert. Im Ordner `csv` befinden sich Tabellen für die jeweiligen Auswertungen. Im darüberliegenden Ordner die Verteilung der Satzlängen. Pfad ausgehend vom Ordner `csv`:

- Wie oft kommen welche PoS-Tags vor?
  - `total-pos.csv`
- Wie viele [SpatialEntities, Places, Motions, Locations, Signals, QsLinks, OLinks] gibt es

- total\_tag.csv
- Wie oft kommen welche QsLink Typen vor? (DC,EC, ...)?
  - total\_qs\_type.csv
- Verteilung der Satzlänge graphisch darstellen (x: Satzlänge, y: Wie häufig)?
  - ..\sentence\_distribution.png
  - sentence\_distribution.csv
- Welche Links (QSLinks, OLinks) werden von welchen Präpositionen (markiert durch SPATIAL\_SIGNAL) getriggert?
  - total\_link\_QSLINK.csv
  - total\_link\_OLINK.csv
- Welches sind die fünf häufigsten „MOTION“ Verben (und wie oft kommen diese vor)?
  - top\_5\_verbs.csv

Auf den folgenden Seiten sind die Ergebnisse auch nochmal in drei Tabellen dargestellt. Ich war leider nicht in der Lage sie 'schön' zu formatieren.

Tabelle 1: Anzahl der Tags & Anzahl der PoS-Tags & Anzahl der QSLink-Typen

Tag	Anzahl	PoS	Anzahl	QSLINK-relType	Anzahl
PLACE	1852	NOUN	5030	IN	586
METALINK	1788	PUNCT	3501	EC	196
SPATIAL_ENTITY	1417	DET	3203	TPP	53
QSLINK	970	ADP	3005	NTPP	42
MOVELINK	803	VERB	2721	DC	41
MOTION	771	PROPN	2095	EQ	35
SPATIAL_SIGNAL	714	ADJ	1781	PO	12
MOTION_SIGNAL	526	PRON	1362	OUT	3
PATH	434	ADV	1326		2
NONMOTION_EVENT	341	AUX	1035		
OLINK	244	SPACE	825		
MEASURE	170	CCONJ	825		
MEASURELINK	93	NUM	683		
MLINK	42	PART	491		
CP	17	SCONJ	431		
URL	17	X	30		
		SYM	28		
		INTJ	10		

Tabelle 2: Anzahl der OLink-Trigger

OLINK-Trigger	Anzahl	OLINK-Trigger	Anzahl
on	45	facing	1
along	14	Across	1
next to	12	across from	1
between	8	alongside	1
over	7	Southeast of	1
up	5	adjacent to	1
covered	5	surmounted	1
down	5	west	1
south	4	Facing	1
below	4	next door to	1
beneath	4	toward	1
surrounded	4	southwest	1
above	4	beside	1
around	3	in the direction to	1
behind	3	Along	1
east of	3	to the west from	1
under	3	to SW	1
to	3	coiling up	1
of	3	on the right	1
overlooking	3	on top of	1
surrounding	3	backwards	1
south of	2	on your left	1
east	2	in that direction	1
north of	2	In front of	1
atop	2	northeast	1
Southeast of	2	neighboring	1
in front of	2	line	1
On	2	directly beneath	1
Down	2	across	1
north	2	up to	1
on top	2	upstream from	1
in	2	upon	1
to the left	1	overlook	1
West of	1		

Tabelle 3: Anzahl der QSLink-Trigger

QSLINK-Trigger	Anzahl	QSLINK-Trigger	Anzahl
in	228	house	1
on	75	adjacent to	1
where	68	packed with	1
at	54	surmounted	1
of	44	bordering	1
In	22	part of	1
with	15	against	1
along	14	beside	1
houses	11	behind	1
full of	10	Along	1
next to	10	apart from	1
At	6	inhabited	1
covered	5	coiling up	1
between	5	into	1
contain	4	about	1
inside	4	after	1
connects	4	on top of	1
On	4	inside of	1
around	4	in front of	1
through	4	up	1
away from	4	further	1
filled	3	In front of	1
far from	3	Everywhere	1
including	3	stocked	1
from	3	out of	1
surrounding	3	contains	1
atop	2	line	1
to	2	packed	1
under	2	afar	1
over	2	directly beneath	1
away	2	for	1
outside	2	apart	1
overlooking	2	Down	1
on top	2	up to	1
across	2	upon	1
surrounded	2	covering	1
has	2	restricted	1

## Visualisierung

Die beiden folgenden Graphiken zeigen die Visualisierungen für `Bicycles.xml` und `Highlights_of_the_Prado_Museum.xml`.

Beide sind im pdf-Format auch im Ordner `results/graph` zu finden.

Ich habe mich beschlossen auch die MOTION-Tags mit einzufügen, da es ansonsten Links mit Verweisen auf nicht existierende Tags geben würde.

Abbildung 1: Bicycles.xml

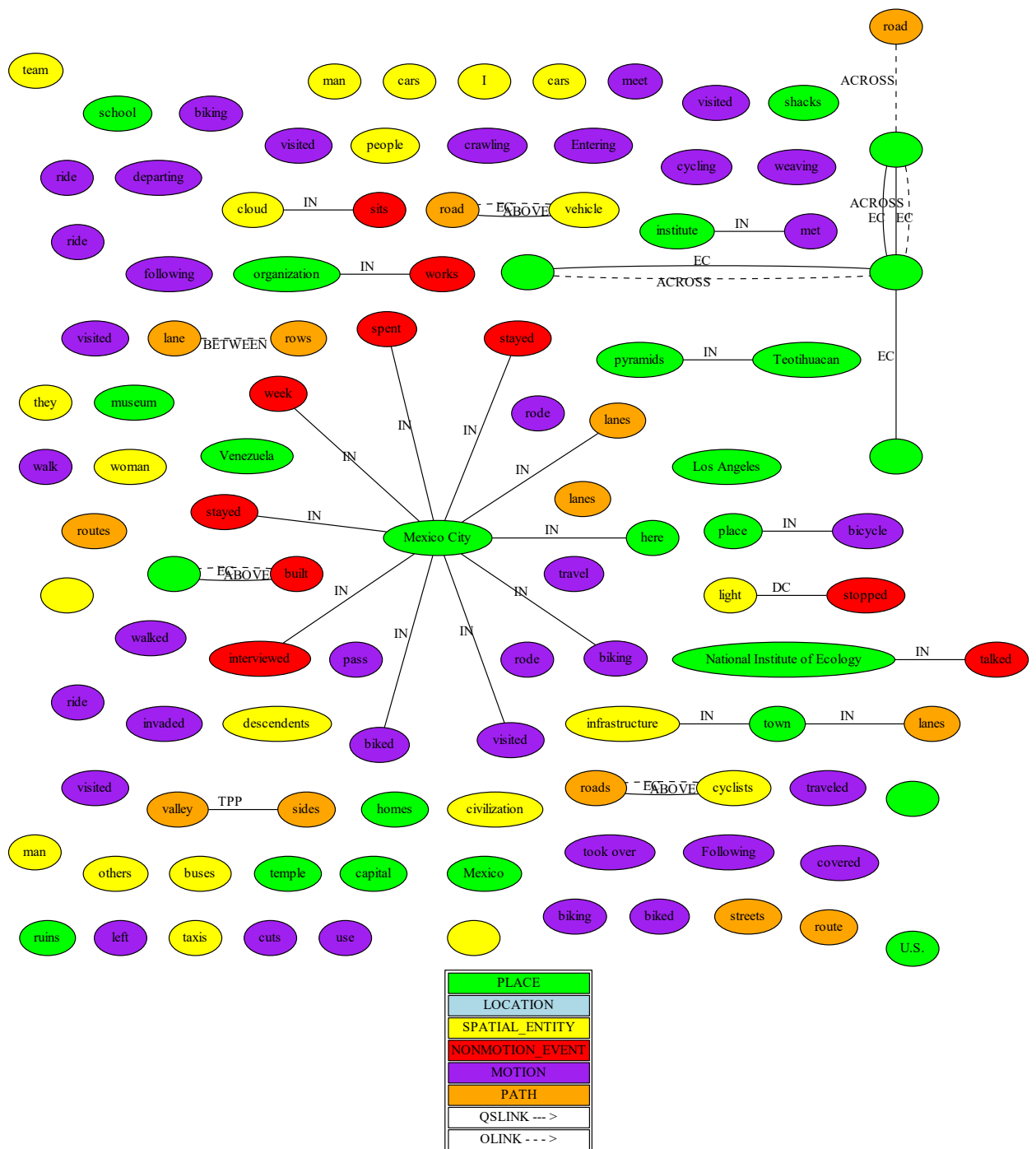


Abbildung 2: Highlights\_of\_the\_Prado\_Museum.xml

