

# ESuPol Datensätze: Übersicht

Im Projekt ESuPol wird der Einfluss von Suchmaschinen, speziell von Query Suggestions, auf die politische Meinungsbildung erforscht. Dazu dienen Sammlungen von Suchanfragen mitsamt der von den Suchmaschinen zurückgelieferten Query Suggestions, die per Crawler erfasst und in SQL Datenbanken gespeichert werden.

Die Daten der verschiedenen Datensätze werden jeweils in Tabellen gespeichert, die über IDs verknüpft sind. Die für Sie relevanten Daten liegen in den “suggestions” Tabellen vor, in denen für jede Session (= hier: Suchanfrage) der Queryterm und die von der Suchmaschine vorgeschlagenen “suggestterms” mitsamt der korrespondierenden Position des Vorschlags festgehalten sind. Im folgenden aufbereiteten Ausschnitt des BTW17 Datensatzes wurde beispielsweise zum Suchterm “Angela Merkel” als erste Suggestion “Angela Merkel biography”, als zweite Suggestion “Angela Merkel husband” usw. vorgeschlagen:

session id	queryterm	datetime	suggestterm	position
2086154	Angela Merkel	2017-08-03 05:46:17	biography	0
2086154	Angela Merkel	2017-08-03 05:46:17	husband	1
2086154	Angela Merkel	2017-08-03 05:46:17	young	2
2086154	Angela Merkel	2017-08-03 05:46:17	trump	3
2086154	Angela Merkel	2017-08-03 05:46:17	meme	4
2086154	Angela Merkel	2017-08-03 05:46:17	eye roll	5
2086154	Angela Merkel	2017-08-03 05:46:17	twitter	6
2086154	Angela Merkel	2017-08-03 05:46:17	photo	7
2086154	Angela Merkel	2017-08-03 05:46:17	snl	8
2086154	Angela Merkel	2017-08-03 05:46:17	and donald trump	9
...	...	...	...	...

In dem Datensatz selbst liegen die Daten die sie abgebildet sehen in der Regel in der Spalte “raw\_data” als JSON vor.

Im Einzelnen sind die aktuellen Datenbanken im Projekt:

1. Bundestagswahl 2017 (BTW17)
  - a. Datensatz zur Bundestagswahl. Die Suchbegriffe bestehen aus Namen von deutschen Politikern (Mitglieder des Bundestags, der Landtage und wichtige

deutsche Politiker auf Europaebene) sowie Personen des öffentlichen Lebens mit einer gewissen politischen Relevanz (wichtige Persönlichkeiten der Wirtschaft, ehemalige Politiker). Die Suchen werden auf drei Suchmaschinen (Google, DuckDuckGo, Bing) durchgeführt.

## 2. European political Sphere (EU)

- a. Über Proxies verschiedener Länder wird zu einer großen Zahl europäischer Politiker in europäischen Ländern gesucht. Die Politiker sind sowohl wichtige Vertreter von Politikern auf nationaler Ebene (Angela Merkel, Nicolas Sarkozy, usw.) als auch amtierende und ehemalige Abgeordnete des EU Parlaments.

Die Suchterme wurden darüber hinaus noch um einige Begriffe rund um das Phänomen "Brexit" erweitert.

## 3. Minorities (minorities)

- a. Der Datensatz sammelt Query suggestions zu Search Terms, die Anfänge von Sätzen (Aussagen und Fragen) zu einer Reihe von ethnischen oder religiösen Gruppen und Nationalitäten darstellen. Beispielsweise wurde zur Bildung der Terme die Sätze "Warum haben alle" und "Warum sind alle" um die verschiedenen "minorities" (z.B. Moslems, Juden, Spanier, Polen, ...) erweitert und dann die suggestions analog zu den anderen Datensätzen erfasst.

In der Regel werden die Suggestions durchgehend zwei mal am Tag gesammelt, sodass auch die Veränderung über die Zeit beurteilt werden kann. Im GitHub finden Sie Auszüge der Datenbanken, um sich einen noch genaueren Eindruck verschaffen zu können.