

Topic Extraction and Query Generation From Websites

Kardelen Bilir

University of Duisburg-Essen, Department of Data Science

kardelen.bilir@stud.uni-due.de

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Introduction

In recent years, the rise of conspiracy theories and extremist rhetoric on the internet has garnered attention from researchers and policymakers.[1] This study focuses on developing a system that extracts and tracks emerging vocabulary from online communities, including conspiracy, far-right, and far-left websites. The system is designed to extract relevant keywords and key phrases from website content, which can later be used to query search engines and measure how a website performs compared to competing websites. By analyzing newly coined terms and monitoring trends, the project aims to understand how new ideologies spread through language. This is crucial for addressing misinformation, combating radicalization, and identifying the spread of harmful narratives that can impact public perception and social stability.

Social Relevance

1. Combating Conspiracy Theories: Searches like "Google chemtrails" [2] indicate the spread of conspiracy theories. Analyzing such queries helps researchers develop strategies to counter misinformation and educate the public.
2. Monitoring Extremist Rhetoric: Terms like "Remigration" are used in far-right circles. Early detection allows authorities to intervene before radical movements gain momentum.

Research Questions

1. How effective is the developed system at extracting relevant keywords and detecting newly emerging terms that are not present in existing dictionaries?
2. What societal trends can be identified by analyzing the vocabulary of selected websites?
3. To what extent can the system contribute to identifying and tracking new ideological movements within specific online communities?

1 Methods

1.1 Data Collection

Data collection has been achieved through a custom web-scraping tool developed using both Scrapy and BeautifulSoup. Currently, the tool is used to crawl and extract content from the website sezeession.de, which is known for its content related to conspiracy theories and extremist rhetoric. The website is crawled daily to capture newly published articles. BeautifulSoup is used to parse the HTML and extract relevant content, which is stored in a MongoDB database.

1.2 Data Processing

The collected articles are analyzed using Natural Language Processing (NLP) techniques. The texts are preprocessed using Part-of-Speech (POS) tagging with spaCy to extract key word types such as nouns, verbs, and adjectives. In addition to individual words, multi-word phrases (N-grams) are extracted, specifically focusing on N-grams where the second word is a noun and the first word is an adjective or verb. This allows for capturing meaningful phrases like "ausgeprägter Deutschenhaß" or "türkische Jugendclans."

These extracted words and phrases are then compared against a reference dictionary to identify new or uncommon terms that are not part of the existing vocabulary. New terms are stored in a vocabulary database, and a daily summary is generated to track the frequency and recurrence of these terms over time.

Expected Outcomes

- The system should accurately extract relevant keywords and phrases in both English and German, while identifying new, potentially significant terms not found in standard dictionaries.
- By analyzing selected websites, the system should recognize and document important societal issues and trends.
- The system is expected to function consistently well across different thematic website focuses.
- The system should help reveal the emergence and spread of new ideological movements by detecting recurring use of new keywords across various websites.
- Extracted keywords will be useful for evaluating the positioning of observed websites in search engine results and comparing them with competing websites.

Preliminary Results

```
._id: ObjectId('66e75d7437888dddf71d8ea')
title: "Der verhausschweinte Streikposten"
url: "https://sezeession.de/300/verhausschweintung"
full_text: "Götz Kubitschek Manche Begriffe, die man leichthin verwendet, sind so ..."

._id: ObjectId('66e75d7637888dddf71d8eb')
title: "Briefwechsel Carl Schmitt/Hans-Dietrich Sander"
url: "https://sezeession.de/361/briefwechsel-carl-schmitt-hans-dietrich-sander"
full_text: "Götz Kubitschek Ich kann eine Neuerscheinung meines Verlags anzeigen: ..."

._id: ObjectId('66e75d7837888dddf71d8ec')
title: "Toleranz - Die 9. Todsünde der zivilisierten Menschheit"
url: "https://sezeession.de/392/toleranz-die-9-todsuende-der-zivilisierten-me..."
full_text: "Götz Kubitschek / 16 Kommentare pdf der Druckfassung aus Sezeession 28 ..."
```

Figure 1: Raw Articles

```
._id: ObjectId('66ed94b51eac69a0284f544')
title: "Rassismus gegen Deutsche"
url: "https://sezeession.de/5035/rassismus-gegen-deutsche"
full_text: "Gastbeitrag pdf der Druckfassung aus Sezeession 17/April 2007 von Danie..."
ranked_keywords: Array (56)
new_phrases: Array (28)
0: "Jugendclans"
1: "Deutschenhaß"
2: "Vortragen"
3: "multikulturalistischen"
4: "Xenophilie"
5: "multikulturalismuskritische"
6: "Integrationsunfähigkeit"
7: "traditionalistischislamischer"
8: "Einwanderergesellschaften"
9: "weltflüchtigen"
10: "Hofjuden"
11: "Abkapselung"
12: "Kommunitarismus"
13: "Supraethnische"
14: "rechtsnihilistischen"
15: "sauvageons"
16: "Migrantenkids"
17: "türkische Jugendclans"
18: "ausgeprägter Deutschenhaß"
19: "schienen Vortragen"
20: "multikulturalistischen Xenophilie"
21: "multikulturalismuskritische Konservative"
22: "offenkundigen Integrationsunfähigkeit"
23: "traditionalistischislamischer Einwanderergesellschaften"
24: "weltflüchtigen Deutschen"
first_processed: "2024-09-20"
last_processed: "2024-09-20"
```

Figure 2: Processed Articles

```
._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "Xenophilie"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "ausgeprägter Deutschenhaß"
first_seen: "2024-09-20"
last_seen: "2024-09-20"
occurrences: 1

._id: ObjectId('66ed94b51eac69a0284f544')
word: "türkische Jugendclans"
```