

# 1 Main-Algorithm

## Data:

$TV_0$ : bestehende, normalisierte Liste von Tokenvektoren

$Tweets$ : Liste von Tweets

$Stopwords$ : Liste von Stoppwörtern

## Result:

$TV_1$ : verbesserte, normalisierte Liste von Tokenvektoren

```
1  $TV_1 \leftarrow \emptyset$ ;  
2 foreach  $Tweet$  in  $Tweets$  do  
3    $Tweet_V \leftarrow (0, 0, 0, 0, 0, 0, 0)$ ;  
4   forall the  $Token$  in  $Tweet$  do  
5     if  $Token \notin Stopwords$  then  
6        $Tweet_V \leftarrow Tweet_V + TV_0(Tweet)$ ;  
7     end  
8   end  
9   forall the  $Token$  in  $Tweet$  do  
10    if  $Token \notin Stopwords$  then  
11       $TV_1(Tweet) \leftarrow TV_1(Tweet) + Tweet_V$ ;  
12    end  
13  end  
14 end  
15 foreach  $Token_V$  in  $TV_1$  do  
16    $Token_V \leftarrow \text{normalize}(Token_V)$ ;  
17 end  
18 if  $\text{difference}(TV_0, TV_1) = \text{significant}$  then  
19    $TV_0 \leftarrow TV_1$ ;  
20   Goto Line 2;  
21 else  
22   return  $TV_1$ ;  
23 end
```

**Algorithm 1:** Tweegion Main-Algorithm

## 2 Geo-Algorithm

**Data:**

*Tweets*: Liste von Tweets mit Geo-Informationen

*Stopwords*: Liste von Stoppwörtern

**Result:**

*TV*: normalisierte Liste von Tokenvektoren

```
1  $TV \leftarrow \emptyset$ ;  
2 foreach Tweet in Tweets do  
3    $n \leftarrow \text{classify}(\textit{Tweet})$ ;  
4    $\textit{Tweet}_V \leftarrow \text{createVector}(n)$ ;  
5   forall the Token in Tweet do  
6     if Token  $\notin$  Stopwords then  
7        $TV(\textit{Tweet}) \leftarrow TV(\textit{Tweet}) + \textit{Tweet}_V$ ;  
8     end  
9   end  
10 end  
11 foreach  $\textit{Token}_V$  in TV do  
12    $\textit{Token}_V \leftarrow \text{normalize}(\textit{Token}_V)$ ;  
13 end  
14 return TV;
```

**Algorithm 2:** Tweegion Geo-Algorithm