

# GAME OF THROLLS

ex nihilo omnia

Idee & Entwicklung: Dr. Martin Hahmann

## BVM Data Science Cup 2019

Die Aufgabenstellung des diesjährigen DSC basiert auf dem Konzept der Gamification, d. h. der Anwendung spiel-typischer Elemente in einem spielfremden Kontext. Das Game of Throlls (GoT) abstrahiert Konzepte von Meinungsbeeinflussung und „Trolling“ in sozialen Medien und bettet sie in eine Spielmechanik ein. Teilnehmer durchsuchen Twitter nach geeigneten Tweets und fügen diese geschickt zu Themen hinzu, um ihren Einfluss auszubauen bzw. fremde Einflüsse zu reduzieren. Dazu werden Werkzeuge zur Extraktion von Tweets, Amazon Comprehend zur Textanalyse, sowie Metriken für die Textähnlichkeit und den Einfluss einzelner Posts eingesetzt.

### **1. Spielprinzip:**

Das Spiel verläuft in Runden. Pro Werktag wird eine Runde gespielt. Es wird über einen Zeitraum von 4 Wochen im April gespielt, also insgesamt über 20 Runden. Der Zeitraum wird in zwei Spielzeiten zu je 2 Wochen bzw. 10 Runden aufgeteilt. Am Ende einer Spielzeit wird eine Wertung durchgeführt und die Spieler erhalten Punkte.

Mo	Di	Mi	Do	Fr	Sa	So
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					
Runden im April					Spielzeit 1	Spielzeit 2

Das „Spielfeld“ besteht aus einer Anzahl initial vorgegebener Themen. In jeder Runde kann jeder Spieler eine festgelegte Anzahl von Tweets auswählen und diesen Themen hinzufügen. Tweets die in einer Runde eingesetzt werden, müssen am Vortag der aktuellen Runde gepostet worden sein und in englischer Sprache verfasst sein (um mittels Amazon Comprehend analysiert werden zu können).

Das Grundprinzip von GoT ähnelt Auslegekartenspielen wie Mau-Mau oder Uno. Dabei entspricht ein Thema einem Ablagestapel und Tweets den darauf ausgespielten Karten. Beim klassischen Kartenspiel müssen ausgespielte Karten hinsichtlich Farbe und Kartenwert zum Ablagestapel passen. Diese Mechanik wird für GoT adaptiert, wobei Farbe und Kartenwert durch bestimmte Eigenschaften eines Tweets repräsentiert werden.

Durch das Hinzufügen von Tweets zu einem Thema, wird dessen Wert und Stimmung verändert, wodurch schlussendlich Punkte für die Spieler generiert werden.

## 2. Eigenschaften eines Tweets:

Im GoT werden die folgenden Eigenschaften eines Tweets betrachtet:

**Wert:** Diese Eigenschaft beschreibt die Interaktion zwischen Publikum und Tweet und repräsentiert damit seinen Einfluss/Wert. Zur Bestimmung wird die folgende Formel genutzt:

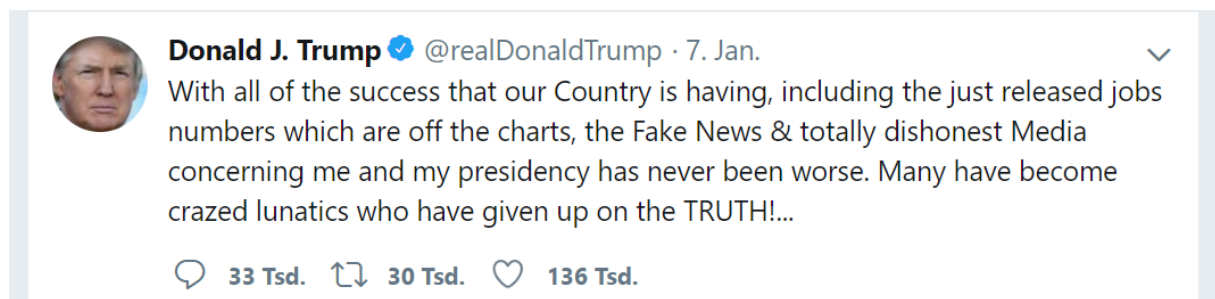
$$\text{Wert} = \text{Likes} + 5 \times \text{Replies} + 10 \times \text{Retweets}$$

Die unterschiedlichen Gewichte repräsentieren den Einfluss der einzelnen Reaktionen. Replies sind höher bewertet als Likes, da sie eine Diskussion fördern. Retweets sind höher bewertet als Replies, da sie zusätzlich die Reichweite des Tweets erhöhen.

**Stimmung:** Die Stimmung eines Tweets beschreibt, inwieweit die in ihm geäußerte Haltung als positiv oder negativ zu bewerten ist. Zur Bestimmung wird die Sentiment Analyse vom Amazon Comprehend (siehe Abschnitt 7) eingesetzt. Diese klassifiziert einen Tweet/Text entweder als positiv, negativ, neutral oder gemischt.

**Inhalt:** Um den Inhalt eines Tweets zu beschreiben wird eine Menge von Schlüsselworten generiert. Dazu werden zuerst die Schlüsselphrasen des Tweets mittels Amazon Comprehend extrahiert. Aus diesen werden dann Stoppwörter (siehe Liste), Satz- und Sonderzeichen, sowie Zahlen und Zahlwörtern entfernt. Aus der so erzeugten Schlüsselwortmenge werden, eventuell vorhandene, Duplikate entfernt.

Beispiel:



$$\text{Wert} = 136.000 + 5 \times 33.000 + 10 \times 30.000 = 601.000$$

Stimmung: negativ

```
"Sentiment": "NEGATIVE",
  "SentimentScore": {
    "Mixed": 0.01868121139705181,
    "Negative": 0.8237994909286499,
    ...}
```

Inhalt:

Schlüsselphrasen:

```
"the success"
"our Country"
"the just released jobs
numbers"
"the charts"
"the Fake News & totally
dishonest Media"
"my presidency"
```

"crazed lunatics"

Schlüsselwörter:

```
"success" "country"
"released" "jobs" "numbers"
"charts" "fake" "news"
"totally" "dishonest" "media"
"presidency" "crazed"
"lunatics" "truth"
```

"the TRUTH"

### **3. Einen Tweet hinzufügen:**

Durch das Hinzufügen eines Tweets zu einem Thema, versuchen die Spieler sich als Meinungsführer für dieses Thema zu etablieren. Dazu muss ein Spieler „Stimmung machen“ also durch Tweets mit positiver und/oder negativer Stimmung die Meinungsbildung im Thema beeinflussen. Der optimale Weg sich als führender Befürworter/Kritiker eines Themas zu etablieren liegt im Hinzufügen von Tweets mit hohem Wert und passendem Inhalt.

Um den inhaltlichen Bezug zwischen einem Tweet  $t$  und einem Thema  $x$  zu repräsentieren werden deren Inhalte in Form der Schlüsselwortmengen  $T$  und  $X$  mittels des Otsuka-Ochiai Koeffizienten verglichen:

$$\text{Bezug}(T, X) = \frac{|T \cap X|}{\sqrt{|T| \times |X|}}$$

Damit wird die Anzahl der übereinstimmenden Schlüsselworte ins Verhältnis zur Gesamtzahl der Schlüsselworte gesetzt. Der Wertebereich der Ähnlichkeit liegt demzufolge zwischen 0 und 1 wobei 1 die Identität beider Mengen anzeigt.

Aus dem inhaltlichen Bezug und dem Wert eines hinzugefügten Tweets  $t$  ergibt sich der Anteil eines Spielers an der Meinungsbildung eines Themas  $x$ . Die Größe dieses Anteils wird berechnet mit:

$$\text{Anteil}(t, x) = \text{Bezug}(T, X) \times \text{Wert}(t)$$

Aus der Stimmung des hinzugefügten Tweets ergibt sich die Art des erzeugten Anteils: positiv, negativ oder neutral. Wird ein Tweet mit gemischter Stimmung hinzugefügt, wird die berechnete Größe hälftig auf je einen positiven und negativen Anteil verteilt. Anteile gleichen Typs die ein Spieler an einem Thema erwirbt, werden aufsummiert. Spieler können positive und negative Anteile am selben Thema haben.

#### **4. Themen entwickeln:**

Durch das Hinzufügen von Tweets verändern Spieler nicht nur ihren Anteil an bestimmten Themen, sondern auch deren Wert, Inhalt und emotionale Aufladung. Diese Eigenschaften entscheiden schlussendlich darüber wie viele Punkte ein Spieler aus einem Thema generieren kann. GoT definiert vier Eigenschaften für jedes Thema, die denen eines Tweets ähneln:

**Themenwert:** Die Summe der Werte aller dem Thema hinzugefügten Tweets.

**Themenstimmung:** Die Menge aller positiven und negativen Anteile die Spieler am Thema erworben haben. Der Betrag der Themenstimmung  $|\text{Themenstimmung}|$  bezeichnet die Summe aller positiven und negativen Anteile.

**Klima:** Diese Eigenschaft beschreibt die emotionale Aufladung des Themas und ergibt sich aus dem Verhältnis von  $|\text{Themenstimmung}|$  zu Themenwert. Ist die Summe aller positiven und negativen Anteile eines Themas kleiner gleich der Hälfte seines Wertes ist das Klima **sachlich**. Übersteigt diese Summe die Hälfte des Themenwertes herrscht ein **polarisiertes** Klima. Sollte der Betrag der Themenstimmung den Themenwert erreichen oder übersteigen, herrscht ein **toxisches** Klima. Das Klima eines Themas beeinflusst die Menge an Punkten die Spieler aus dem Thema gewinnen können.

**Inhalt:** Für die Ermittlung des Bezugs zwischen Tweet und Thema wird eine Schlüsselwortmenge  $X$  benötigt, die den Inhalt des Themas repräsentiert. Diese wird initial vorgegeben und ändert sich danach am Ende jeder Runde. Der in der aktuellen Runde hinzugefügte Tweet aus dem sich der größte Anteil an der Meinungsbildung ergibt wird ausgewählt. Seine Schlüsselwortmenge repräsentiert den Themeninhalt in der nächsten Runde.

Beispiel:

Thema Y	(+) Anteile	(-) Anteile
Spieler A	500	700
Spieler B	1000	200

$ \text{Themenstimmung} $	2400
Themenwert	5000

Ausgangslage: Zwei Spieler besitzen bereits positive und negative Anteile am Thema Y. Der Themenwert von Y beträgt 5000, der Betrag der Themenstimmung 2400. Daraus ergibt sich ein sachliches Klima. Im Folgenden fügt jeder Spieler je einen Tweet zum Thema hinzu:

Spieler A fügt einen negativen Tweet hinzu:  $Anteil(t_A) = 0,4 \times 900 = 360$  negativ (-)

Spieler B fügt einen positiven Tweet hinzu:  $Anteil(t_B) = 0,9 \times 600 = 540$  positiv (+)

Obwohl der Tweet von Spieler A den höheren Wert aufweist, ergibt sich durch den schwächeren inhaltlichen Bezug ein kleinerer negativer Anteil. Spieler B erreicht durch hohen Bezug mit seinem Tweet einen größeren positiven Anteil. Der Tweet von Spieler B produziert in dieser Runde den größten Anteil an der Meinungsbildung. Seine Schlüsselwortmenge repräsentiert das Thema in der nächsten Runde. Am Ende der Runde entwickelt sich das Thema wie folgt:

Thema Y	(+) Anteile	(-) Anteile
Spieler A	500	700+360
Spieler B	1000+540	200

Themenstimmung	2400+900
Themenwert	5000+1500

Beide Spieler bauen ihre jeweils dominanten Anteile aus. Der Themenwert erhöht sich um die Summe der Tweet Werte auf 6500, während der Betrag der Themenstimmung durch erzeugte Anteile auf 3300 ansteigt. Durch den stärkeren Anstieg der Themenstimmung im Vergleich zum Themenwert verändert sich deren Verhältnis und das Klima wird polarisiert.

## **5. Punkte erhalten:**

Spieler generieren Punkte aus den Anteilen, die sie an verschiedenen Themen erworben haben. Der Wert eines Anteils ändert sich von Runde zu Runde abhängig von der Themenentwicklung. Am Ende einer Spielzeit werden die Anteile bewertet und die Punkte den Spielern gutgeschrieben. Jeder Spieler erhält pro Thema Punkte aus zwei Wertungen:

- Die erste Wertung basiert auf dem größten Anteil eines Spielers am Thema.
- Die zweite Wertung basiert auf dem Anteil eines Spielers an der vorherrschenden Stimmung (positive Anteile vs. negative Anteile) eines Themas.

Die Umrechnung der Anteile in Punkte hängt vom Klima des Themas ab. Wie bereits erwähnt geht es im GoT um's „Stimmung machen“, daher ergeben sich aus Themen mit hoher Emotionalisierung mehr Punkte. Die Punkturnrechnung erfolgt folgendermaßen:

**Sachliches Klima:** In beiden Wertungen wird dem Spieler die Größe des jeweils relevanten Anteils als Punkte gutgeschrieben.

**Polarisiertes Klima:** Pro Wertung wird der jeweils relevante Anteil eines Spielers auf den Themenwert hochgerechnet, d. h. für einen Anteil, dessen Größe  $x$  Prozent von  $|\text{Themenstimmung}|$  ausmacht erhält der Spieler  $x$  Prozent des Themenwertes als Punkte. Da Themenwert  $> |\text{Themenstimmung}|$  gilt, ergeben sich mehr Punkte als beim sachlichen Klima.

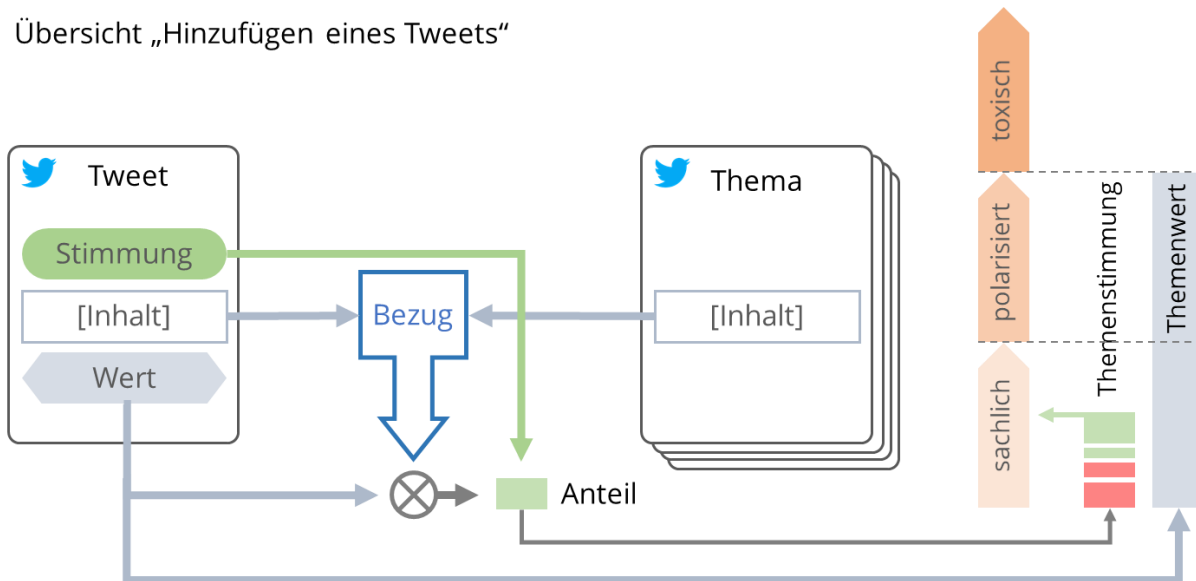
**Toxisches Klima:** Die erste Wertung wird wie beim polarisierten Klima durchgeführt. In der zweiten Wertung erhält der Spieler mit dem größten Anteil am Thema den kompletten Themenwert als Punkte. Alle anderen Spieler erhalten keine Punkte aus der zweiten Wertung.

Beispiel:

Themenstimmung	2400		2400		2400	
Themenwert	5000		4000		2400	
Klima	Sachlich		polarisiert		toxisch	
	Spieler A	Spieler B	Spieler A	Spieler B	Spieler A	Spieler B
(+) Anteile	500	1000	500	1000	500	1000
(-) Anteile	700	200	700	200	700	200
1. Wertung	700	1000	1167	1667	700	1000
2. Wertung	500	1000	833	1667	0	2400
Gesamt	1300	2000	2000	3333	700	3400

Das Beispiel zeigt den starken Einfluss des Klimas auf die Punktausbeute eines Themas. Gegenüber einem sachlichen Thema erlaubt ein polarisiertes Thema eine erhöhte Punktausbeute selbst bei einem geringeren Themenwert. Ein toxisches Thema ermöglicht es einem Spieler seine Punkte zu erhöhen und gleichzeitig die Punktausbeute seiner Konkurrenten zu verringern.

## Übersicht „Hinzufügen eines Tweets“



### 6. Troll Tweet & Neutraler Tweet:

Die starke Emotionalisierung eines Themas erhöht die Punktausbeute. Mit dem Troll Tweet und dem neutralen Tweet erhalten Spieler einen Hebel, der direkt am Klima eines Themas ansetzt.

Der **Troll Tweet** fokussiert auf die Emotionalisierung. Trolling hat kaum Bezug zu einer Diskussion und setzt auf reine Provokation, um inhaltlichen Schaden zu verursachen und Konflikte zu schüren. Der Anteil der sich aus einem Troll Tweet  $t$  ergibt wird berechnet als:

$$\text{Anteil}'(t, x) = (1 - \text{Bezug}(T, X)) \times \text{Wert}(t), \text{Bezug}(T, X) > 0$$

Durch den veränderten Einfluss des Bezugs kann mit einem fast beliebigen Tweet ein großer Anteil an einem Thema erzeugt werden. Ein Troll Tweet muss eine positive oder negative Stimmung aufweisen und schädigt das Thema, dem er hinzugefügt wird. Der Wert des eingesetzten Tweets wird vom Themenwert abgezogen und kann diesen bis auf null reduzieren. Durch die Erhöhung des Betrags der Themenstimmung bei gleichzeitiger Reduzierung des Themenwertes wird das Klima des Themas verschärft.

Der **neutrale Tweet** erlaubt es Spielern Sachlichkeit in ein Thema zu bringen. Ein Tweet mit neutraler Stimmung erzeugt einen neutralen Anteil entsprechend der Formel in Abschnitt 3. Nachdem alle nicht neutralen Anteile der aktuellen Runde eingebracht wurden, werden die neutralen Anteile proportional von allen Anteilen in der Themenstimmung abgezogen (reduzierte Anteile werden gerundet). Der Wert des eingesetzten Tweets wird zum Themenwert addiert. Durch die Reduzierung des Betrags der Themenstimmung bei gleichzeitiger Erhöhung des Themenwertes wird das Klima des Themas entspannt. Dadurch

können, durch Trolling geschädigte Themen wiederaufgebaut und Themen gezielt abgekühlt werden um die erhöhte Punktgewinnung polarisierter/toxischer Themen zu verhindern.

Beispiel Troll Tweet:

	(+) Anteile	(-) Anteile
Spieler A	500	700
Spieler B	1000	200

Klima	sachlich
Themenstimmung	2400
Themenwert	5000

Spieler A: negativer Troll Tweet:  $Anteil'(t_A) = (1 - 0,3) \times 900 = 630$  negativ (-)

Spieler B: fügt positiven Tweet hinzu:  $Anteil(t_B) = 0,75 \times 600 = 450$  positiv (+)

	(+) Anteile	(-) Anteile
Spieler A	500	1330
Spieler B	1450	200

Klima	polarisiert
Themenstimmung	3480
Themenwert	4820

Beispiel Neutraler Tweet:

	(+) Anteile	(-) Anteile
Spieler A	700	800
Spieler B	1000	500

Klima	polarisiert
Themenstimmung	3000
Themenwert	5000

Spieler A: fügt neutralen Tweet hinzu:  $Anteil(t_A) = 0,4 \times 2000 = 800$  neutral (o)

Spieler B: fügt positiven Tweet hinzu:  $Anteil(t_B) = 0,75 \times 600 = 450$  positiv (+)

	(+) Anteile	(-) Anteile
Spieler A	538	614
Spieler B	1114	384

Klima	sachlich
Themenstimmung	2650
Themenwert	6250

## 7. Technisches:

Der aktuelle Stand des Spiels wird auf der DSC Website angezeigt. Vor Spielbeginn wird jedem Teilnehmer eine anonyme Spieler ID zugeteilt. Anhand dieser, kann der aktuelle Punktestand eingesehen werden.



Spieler reichen ihre Tweets für die aktuelle Runde im Format: *Zielthema, TweetID, Troll ja/nein* per Email an [kontakt-dsc@bvm.org](mailto:kontakt-dsc@bvm.org) bis 15 Uhr des Rudentages ein. Der Erhalt wird bestätigt. Danach werden die Einreichungen geprüft und verarbeitet. Tweets die nicht den Regeln entsprechen werden verworfen. Im Anschluss werden der Spielzustand und die Website aktualisiert.

Twitter ist ein dynamisches Medium. Der Wert eines Tweets wird mit den aktuellen Likes, Replies und Re-Tweets zum Zeitpunkt der Verarbeitung berechnet. Bitte speichern Sie Screenshots ihrer eingereichten Tweets als Backup für den Fall einer Löschung ab.

Sollten Elemente der Spielmechanik in einer Art und Weise eingesetzt werden, die das Spiel unausgeglichen oder unfair macht, werden die Organisatoren durch Regelanpassungen gegensteuern. Regelanpassungen finden nur an den Wochenenden (nach jeweils 5 Runden) statt und werden vor Beginn der nächsten Runde per Website und Email kommuniziert.

Twitter bietet unter <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html> eine Search API für den automatisierten Zugriff an. Die Standard Variante dieser API ist kostenlos nutzbar unterliegt aber bestimmten Einschränkungen. Weiterhin existieren Bibliotheken wie <https://github.com/taspinar/twitterscraper> die auf Twitters Frontend aufsetzen und einige Einschränkungen der Search API umgehen.

Für die Nutzung von Amazon Comprehend und AWS im Allgemeinen finden sich unter <https://aws.amazon.com/de/comprehend/> ausführliche Anleitungen und Beispiele zur Registrierung bei AWS und zur Anwendung von Amazon Comprehend. Die Nutzung von Amazon Comprehend ist kostenpflichtig und wird über Einheiten abgerechnet. Allerdings erhält jedes AWS Konto pro Monat ein kostenloses Kontingent von 50.000 Einheiten. Damit können ca. 8300 Tweets analysiert werden.