**Codebuch (für den Pretest)**

**Vorgehen bei Codierung:**

**Teil 1:**

Erhebung aller (verlinkten) Texte der Autoren auf achgut.com und auf anderen Webseiten im folgenden Untersuchungszeitraum: **01.01.2018 bis 31.01.2018**

1. Konzentration auf einen Autor der Website “achgut.com”
2. Google-Suche nach dem Namen des Autors in der Standard-Google-Suche und **zusätzlich** in der Rubrik “News”
3. Begrenzung der Suchergebnisse auf den Zeitraum 01.01.2018 bis 31.01.2018
4. Erfassung jeder einzelnen Verlinkung bzw. Veröffentlichung eines Textes des Autors auf anderen Seiten
   * zum Beispiel: Suche nach Autor “Archi W. Bechlenberg”, Erfassung eines Artikels des Autors auf epochtimes.de
5. Zusätzliche Suche in folgenden Kombinationen:
   * Autorname und “Journalist”
   * Autorname und “Überschrift Artikel” (hier wird nach allen Artikelüberschriften gesucht, die in diesem Monat auf der Achse des Guten von diesem Autor erschienen sind, z.B.: “Archi W. Bechlenberg Das Antidepressivum zum Sonntag”)

**Teil 2:**

Erhebung der biographischen Angaben der Autoren

1. Recherche über Kurzbiographie auf “achgut.com”
2. evtl. zusätzliche Recherche über Munzinger Personenarchiv

**Edgelist:**

In der Edgelist werden alle Verlinkungen der Texte eines Autors von “achgut.com” auf anderen Webseiten erfasst, die bei der Google-Suche zu finden sind (Vorgehen siehe oben). **Jede Verbindung wird in einer neuen Zeile erfasst, auch, wenn ein Autor auf einer Website mehrfach auftaucht** (z.B. fünf Artikel von Stefan Frank auf epochtimes.de → fünf separate Codierzeilen).

**Knoten 1: (from)**

Hier wird der Name des Autors erfasst (Schreibweise s.u.).

**Knoten 2: (to)**

Hier wird der Name der Website erfasst (Schreibweise s.u.)

**Beziehung (relation):**

Hier wird die Art der Verlinkung des Artikels erfasst. Wir unterscheiden zwischen

* Volltext (full text): Wird codiert, wenn der vollständige Artikel eines Autors auf der Seite abgebildet ist.
* Teaser (teaser): Wird codiert, wenn es eine Artikelvorschau gibt, die auf einen Artikel des Autors verweist und darauf verlinkt.
* Verlinkung (link): Wird codiert, wenn es sich nur um einen Verweis und eine Verlinkung auf einen Text des Autors handelt.

**Artikelname (article):**

Hier wird der Titel des Artikels erfasst.

**Link von Artikel (link):**

Zur besseren Nachvollziehbarkeit und Kontrolle wird die vollständige URL des gefundenen Artikels an dieser Stelle erfasst.

**date:**

Datum des Artikels, falls vorhanden. In folgender Schreibweise: tt.mm.jjjj (21.01.2018)

Wenn nicht vorhanden, wird na codiert.

**comments:**

Anzahl der Kommentare, die unter den Artikel geschrieben wurde. Wird nur codiert, wenn es sich um Kommentare zum Volltext des codierten Artikels handelt, wird nicht NICHT codiert, wenn sich die Kommentare auf einen anderen Artikel beziehen, unter dem bspw. der codierte Artikel als Teaser oder Link auftaucht.

**Beispiel Edgelist:**

knot 1 = from  
knot 2 = to

link = URL

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **from** | **to** | **relation** | **article** | **link** | **date** | **comments** |
| stefan\_frank | achgut | full text | Das Antidepressivum zum Sonntag | www.domain/abc-xyz.de |  |  |
| stefan\_frank | haloam | link |  |  |  |  |
| stefan\_frank | welt | teaser |  |  |  |  |

**Nodelist:**

In der Nodelist werden

* die biographischen Daten der Autoren
* die erfassten Websites

erfasst.

Zur Erfassung der biographischen Daten werden die Kurzbiographien auf “achgut.com” herangezogen. Als Ergänzung kann das Munzinger Personenarchiv und Wikipedia dienen.

**id:**

Die Spalte **id** ist die Nennung des Knoten in der Edgelist. **Die ids in der Nodelist müssen identisch mit denen der edgelist sein (siehe Abkürzungsregeln unten)!**

**label:**

ist die genaue Bezeichnung des Knotens aus ID. Diese wird z.B. auch eingesetzt, um den Knoten in der Visualisierung zu beschreiben. Da label als Text-String definiert ist können wir hier auch Umlaute oder komplexere Beschreibungen des Autors oder der Website verwenden.

Bei Autoren wird hier der Vor- und Nachname der Autoren mit Leerzeichen erfasst: Stefan Frank.

Bei Websites wird die komplette Standard-Domain-Bezeichnung erfasst.

Beispiel: “achgut.com” wird erfasst als “achgut.com” / “faktum-magazin.de” wird erfasst als “faktum-magazin.de”

**type:**

Codiert werden die Types 0 und 1, um die Kategorien in unserem two-mode-Netzwerk unterscheidbar zu machen.

0 = Website

1 = Autor

**sex:**

Hier wird das biologische Geschlecht der Autoren codiert.

0 = männlich

1 = weiblich

Bei keiner möglichen Angabe wird na codiert. Bei Websites wird na codiert.

**birth:**

Bei Autoren: umfasst das Geburtsjahr, aus dem sich das Lebensalter berechnen lässt.

Wenn keine Angabe möglich, wird na codiert. Bei Websites wird na codiert.

**profession:**

Hier wird die hauptsächliche Tätigkeit im Erhebungszeitraum erfasst, die der Autor neben seiner Schreibtätigkeit für achgut.com nachgeht.

Wenn der Autor neben seiner journalistischen Tätigkeit für “achgut.com” auch Journalist bei anderen Medien ist, wird das bei der Kategorie “profession” mit “Journalist” erfasst.

Übt der Autor sonst keine Tätigkeit aus oder ist bei der Recherche keine andere Tätigkeit zu finden, wird na codiert.

Bei Websites wird na codiert.

**media:**

Wenn in der Kurzbiographie eines Autors auf “achgut.com” andere Medien genannt werden, in denen der Autor auch publiziert, dann werden diese hier eingetragen. Wenn keine Medien genannt sind, wird na codiert.

Bei Websites wird na codiert.

**page impressions:**

Seiten zur Erfassung der page impressions:

<https://www.similarweb.com/>

→ erfasst wird die Anzahl der Klicks im Monat Januar 2018, die sich aus dem Graph ablesen lässt, wenn man mit dem Cursor über den Punkt im Monat Januar fährt

Gewichtung der page impressions zur Auswertung:

< 10.000 → 1

< 100.000 → 2

< 1.000.000 → 3

< 10.000.000 → 4

>= (größer oder gleich) 10.000.000 → 5

**traffic source social media:**

Tool zur Erfassung des Klickzahlen-Anteils, der von social media kommt

<https://www.similarweb.com/>

>>> Total visits, rank, traffic by countries, traffic sources,

Referrals, traffic from social media

Gewichtung des Traffics von social media zur Auswertung:

< 5 Prozent → 1

< 15 Prozent → 2

< 30 Prozent → 3

>= 30 Prozent → 4

**Qualitätsmedium?**

Handelt es sich bei der Website nach einer kurzen Einschätzung um ein Qualitätsmedium?

nein → 1 (Nein wird mit einer 1 codiert.)

ja → 2 (Ja wird mit einer 2 codiert.)

Bei Autoren wird na codiert.

**Beispiel Nodelist:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **id** | **label** | **type** | **sex** | **birth** | **profession** | **media** | **qualitatesmedium** |
| stefan\_frank | Stefan Frank | 1 | 1 | 1976 | Publizist | na | na |
| achgut | achgut.com | 0 | na | na | na | na | 1 |

**Abkürzungsregeln bei id-Schreibweise:**

* Umlaute in Namen werden zu ae, oe, ue
* Schreibweise der Autorennamen: alles klein geschrieben, Vor- und Nachname durch Unterstrich getrennt:
  + Stefan Frank → stefan\_frank
  + Bernd Müller → bernd\_mueller
* Schreibweise der Websites: alles klein geschrieben, Bindestriche und Punkte werden in Unterstriche übersetzt, Zahlen werden zu x, die Top-Level-Domain (.de, .com, .at…) wird weggelassen
  + achgut.com → achgut
  + faktum-magazin.de → faktum\_magazin
  + politik24.de → politikxx