

Extending the model of the Database of Known Fakes into the Realm of Disinformation Spread and Campaigns

Andrey Tagarev 22.02.2023

## **Base Model: Claim Review**



A Facebook post shared in Kenya claims to show footage of high-rise towers collapsing during the Turkey-Syria earthquake.

The video is old and unrelated. While the 7.8-magnitude earthquake caused massive devastation, collapsing buildings and killing thousands of people, the video shows controlled demolitions in China...



#### **Claim Review Extended**

A Facebook post shared in Kenya claims to show footage of high-rise towers collapsing during the Turkey-Syria earthquake.

https://weibo.com/1962117985/JrlzTzGsi

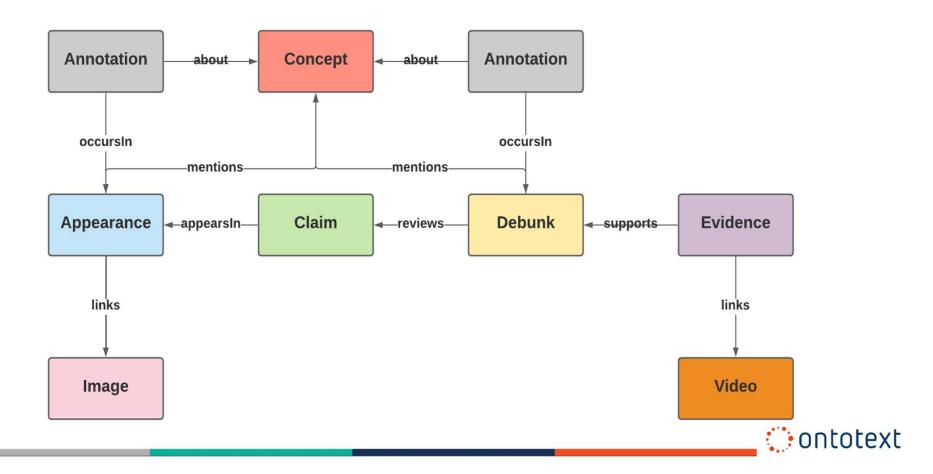


https://www.facebook.com/mary.makau/videos/573432234663224/

The video is old and unrelated. While the 7.8-magnitude earthquake caused massive devastation, collapsing buildings and killing thousands of people, the video shows controlled demolitions in China...

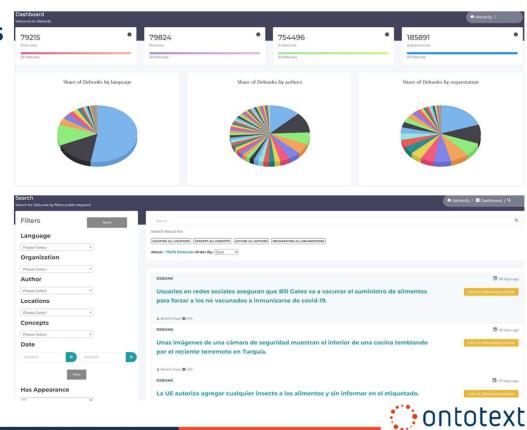


## **Claim Review Enriched**



#### The Database Of Known Fakes

- Over 80k claims and debunks
- Over 175k appearance links
- Over 750k evidence links
- Over 30 fact checking organisations
- Various search methods:
  - Faceted
  - Multilingual
  - Visual similarity
  - o Multimodal



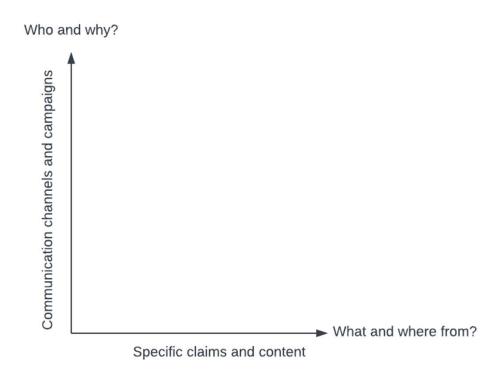
#### The Database Of Known Fakes: Enrichments

- Multilingual concept tagging and linking
- Event detection within text
- Archiving of appearances
- Identification and retrieval of relevant visual content
- Modeling of disinformation campaigns and spread



#### The structure of disinformation

- Two distinct dimensions
- The more complex focuses on the hidden actors and their goals
- The more accessible to automation focuses on the topics and content reuse





# **Cluster Example: Hot Water and Lemon**



ЛЕЧЕНИЕ при Корона вирус и други вируси

от Зор Алеф

- 1.Спираме да се храним две денонощия. Когато се храним 80 процента от имунитета отива за неутрализиране токсините от храната. Когато гладуваме 100 процента от имунитета ни отива за битка с вируса.
- 2. Ако вдигнем висока температура, но не по висока от 39.5 градуса, в никакъв случай не вземайте медикаменти да я смъкнете.
- 3. Пийте чай от лайка и липа.
- 4. Всеки ден пийте сока на 2-3 лимона, с 2 чаени лъжички мед.
- 5. Изолирайте се в къщи на топло и пестете енергията си. Не се натоварвайте физически.
- 6. Сутрин и вечер пийте по 30 до 40 капки спиртен извлек на прополис. Прополисът помага на имунитета, като убива доста вредни патогени.
- 7. Вечер преди лягане пийте от една до десет капки кислородна вода. Започнете от една капка и увеличавайте дозата докато стигнете 10 капки. Кислородната вода помага на организма доставяйки атомен кислород в червата да се възпроизведе силен имунитет.

#### CHECKYOURFACT 🗸



8:24 PM 03/26/2020 Brad Sylvester | Fact Check Editor

¥:

A  $\ensuremath{\mathsf{Facebook}}$  post claims that drinking hot water with sliced lemon can kill the new coronavirus.

Boire de l'eau avec ; je suis Laila Ahmadi de Chine, F de Zanjan. Le virus Corona arrivi fait aucun doute que dou équipement de dia Coronavirus ? illiser autant que possiliser ayutant que possiliser système immunitaire. Ne vous inquiét

beaucoup, et ça va.





"Slices of lemon in a cup of hot water can save your life. The hot lemon can kill the proliferation of" the novel coronavirus.

FACEBOOK FACT-CHECKS CORONAVIRUS & CHAIN MESSAGE





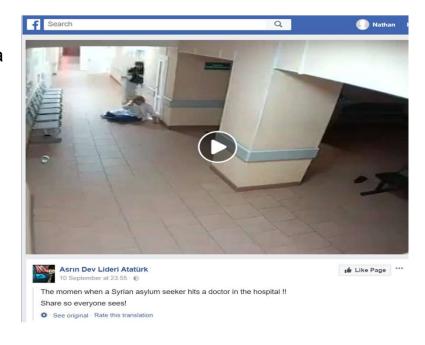
# **Cluster Example: Immigrant Attacking Nurses**

A video of a man assaulting nurses in Novgorod, Russia, was disseminated by French far-right Facebook pages as if it was a migrant assaulting hospital staff in France; in Spain, as if it was a Muslim doing it in a Spanish health centre; and in Turkey as if it was a Syrian in Turkey. This reveals who the main scapegoats were in the different local contexts

# No, internet, this man hitting a nurse is not a migrant

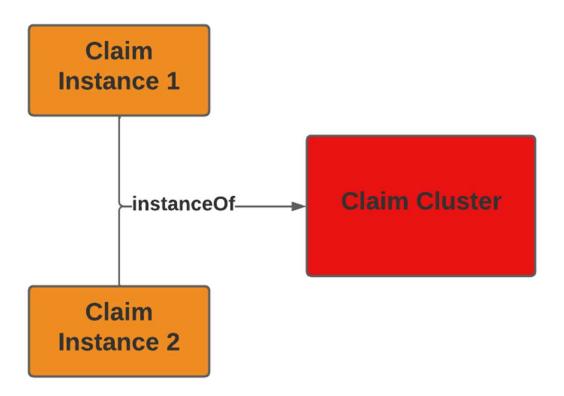






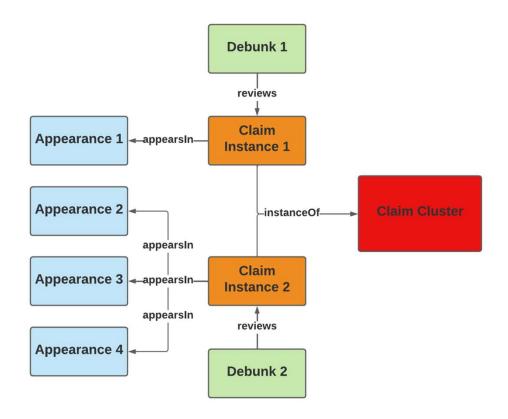


## **Disinformation Reuse in Claims 1**



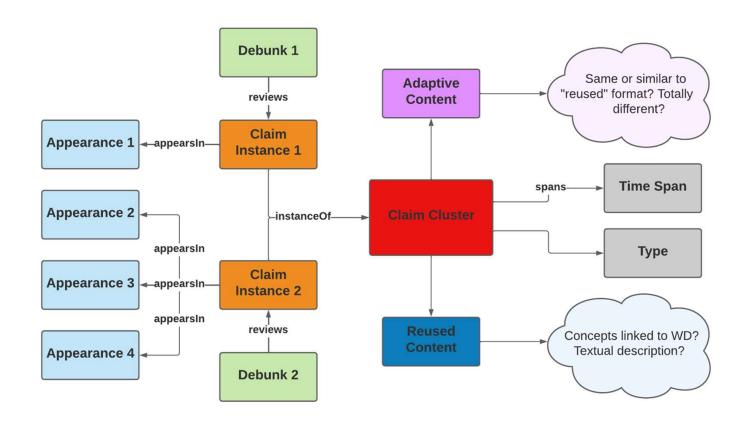


## **Disinformation Reuse in Claims 2**



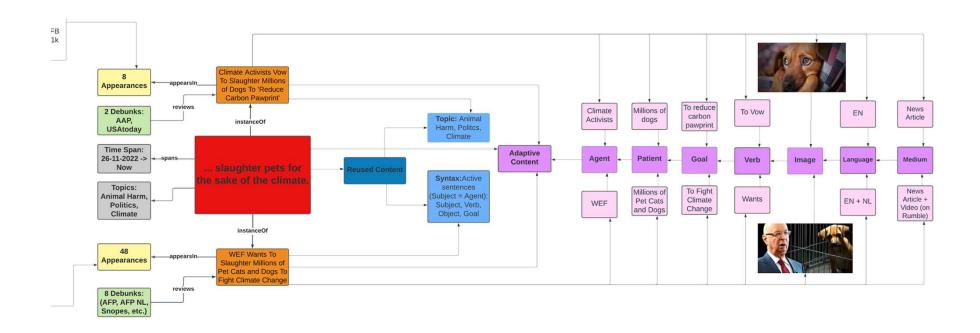


## **Disinformation Reuse in Claims 3**





# **Reuse Examples: Climate Change and Pets**





## **Open Questions: Scope**

Example vaccine disinformation from throughout 2021:

- A 2012 study did not use mRNA vaccines or result in animals dying from disease
- Staff administering COVID-19 vaccines are not war criminals
- COVID-19 vaccines did have clinical trials
- COVID-19 vaccines will not genetically modify humans
- Available mRNA vaccines do not target syncytin-1, a protein vital to successful pregnancies
- COVID-19 vaccines do not contain live immortal creatures
- Vaccine production was not halted due to deaths caused by COVID vaccines



# **Open Questions: Scope**

- Simple clusters but of what size?
- Hierarchical clusters but of what depth?
- No explicit clusters but rather claim similarity?



## **Open Questions: Preparation or Intervention**

- Is this kind of model or semi-automated detection useful?
- What is the best way to make it accessible and to whom?
- Better to react quickly once disinformation appears locally or try to prepare before it has even appeared?

