

Query Expansion Approaches for Image Retrieval in Argumentative Contexts

Nico Reichenbach

Universität Leipzig

18. Juni 2020

- 1 Einordnung
- 2 Aufbau des Prototypen
- 3 Query Expansion Methoden
- 4 Evaluation

Argumentative Bilder



- Erweitern von args um eine Bildersuche

- Erweitern von args um eine Bildersuche
- Suche in einem verschlagworteten Bilderindex

- Erweitern von args um eine Bildersuche
- Suche in einem verschlagworteten Bilderindex
- Annahme: Bilder mit positivem/negativem Bezug auf Websites mit positivem/negativem Bezug

- Erweitern von args um eine Bildersuche
- Suche in einem verschlagworteten Bilderindex
- Annahme: Bilder mit positivem/negativem Bezug auf Websites mit positivem/negativem Bezug
- Erweitern der Query durch positive/negative Begriffe

Suchergebnis als JSON

```
{
  "queryString": <query>,
  "results": [
    {
      "imageUrl": <imageUrl>,
      "thumbnailURL": <thumbnailURL>,
      "origin": <originURL>,
      "altText": <altText>
    }, ...
  ]
}
```


Methoden

Methoden

- reguläre Ausdrücke

Methoden

- reguläre Ausdrücke
- DOM-Traversierung

Methoden

- reguläre Ausdrücke
- DOM-Traversierung
- X-Path

Methoden

- reguläre Ausdrücke
- DOM-Traversierung
- X-Path

Probleme

Methoden

- reguläre Ausdrücke
- DOM-Traversierung
- X-Path

Probleme

- Request Rate Limitation/IP-Blocking

Methoden

- reguläre Ausdrücke
- DOM-Traversierung
- X-Path

Probleme

- Request Rate Limitation/IP-Blocking
- CAPTCHAs

Methoden

- reguläre Ausdrücke
- DOM-Traversierung
- X-Path

Probleme

- Request Rate Limitation/IP-Blocking
- CAPTCHAs
- User Agent Tests

Methoden

- reguläre Ausdrücke
- DOM-Traversierung
- X-Path

Probleme

- Request Rate Limitation/IP-Blocking
- CAPTCHAs
- User Agent Tests
- wechselnde DOM-Struktur

- startpage.com als Suchmaschine

- startpage.com als Suchmaschine
- REST-Applikation mit Spring Boot

- startpage.com als Suchmaschine
- REST-Applikation mit Spring Boot
- Selenium-Framework mit Chrome als Browser

- startpage.com als Suchmaschine
- REST-Applikation mit Spring Boot
- Selenium-Framework mit Chrome als Browser
- Docker-Container als Microservice

- A - Menge an Worten in Pro-Argumenten

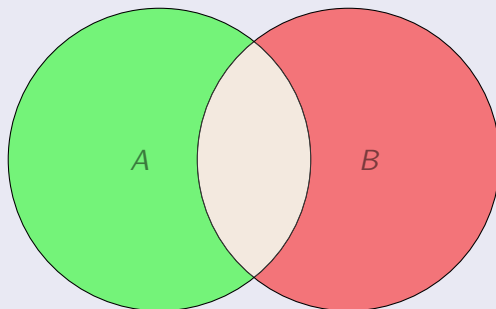
Query Expansion

- A - Menge an Worten in Pro-Argumenten
- B - Menge an Worten in Contra-Argumenten

Query Expansion

- A - Menge an Worten in Pro-Argumenten
- B - Menge an Worten in Contra-Argumenten

Pro und Contra Bag of Words



nuclear power

vote

spacecrafts

judge

tract

windfall

Wahrscheinlichkeitsverteilung von Worten

$$P_{Pro}(t) = \frac{freq_{t,Pro}}{\sum_{t' \in Pro} freq_{t',Pro}}$$

relative Häufigkeit von Worten in Pro-Argumenten

Wahrscheinlichkeitsverteilung von Worten in Argumenten

Wahrscheinlichkeitsverteilung von Worten

$$P_{Pro}(t) = \frac{freq_{t,Pro}}{\sum_{t' \in Pro} freq_{t',Pro}}$$

relative Häufigkeit von Worten in Pro-Argumenten

Frequenz von Worten in Contra-Argumenten

$$P_{Con}(t) = \frac{freq_{t,Con}}{\sum_{t' \in Con} freq_{t',Con}}$$

relative Häufigkeit von Worten in Contra-Argumenten

- Maß für Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen

Kullback-Leibler-Divergenz

- Maß für Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen
- Worte, die am meisten zur Divergenz beitragen

Kullback-Leibler-Divergenz

- Maß für Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen
- Worte, die am meisten zur Divergenz beitragen

Kullback-Leibler-Divergenz

- Maß für Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen
- Worte, die am meisten zur Divergenz beitragen

$$D_{KL}(P_{Pro}||P_{Con}) = \sum_{t \in Args} P_{Pro}(t) * \log \frac{P_{Pro}(t)}{P_{Con}(t)}$$

Kullback-Leibler-Divergenz

- Maß für Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen
- Worte, die am meisten zur Divergenz beitragen

- $$D_{KL}(P_{Pro}||P_{Con}) = \sum_{t \in Args} P_{Pro}(t) * \log \frac{P_{Pro}(t)}{P_{Con}(t)}$$

- $$D_{KL}(P_{Con}||P_{Pro}) = \sum_{t \in Args} P_{Con}(t) * \log \frac{P_{Con}(t)}{P_{Pro}(t)}$$

Beitrag zur Divergenz

Beitrag zur Divergenz

- $P_{Pro}(t) * \log \frac{P_{Pro}(t)}{P_{Con}(t)}$

Beitrag zur Divergenz

- $P_{Pro}(t) * \log \frac{P_{Pro}(t)}{P_{Con}(t)}$
- $P_{Con}(t) * \log \frac{P_{Con}(t)}{P_{Pro}(t)}$

Beitrag zur Divergenz

- $P_{Pro}(t) * \log \frac{P_{Pro}(t)}{P_{Con}(t)}$
- $P_{Con}(t) * \log \frac{P_{Con}(t)}{P_{Pro}(t)}$

Add-One-Smoothing

Beitrag zur Divergenz

- $P_{Pro}(t) * \log \frac{P_{Pro}(t)}{P_{Con}(t)}$
- $P_{Con}(t) * \log \frac{P_{Con}(t)}{P_{Pro}(t)}$

Add-One-Smoothing

- $$P_{Pro}(t) = \frac{freq_{t,Pro} + 1}{\sum_{t' \in Pro} freq_{t',Pro} + \sum_{t' \in Args} 1}$$

Beitrag zur Divergenz

- $P_{Pro}(t) * \log \frac{P_{Pro}(t)}{P_{Con}(t)}$
- $P_{Con}(t) * \log \frac{P_{Con}(t)}{P_{Pro}(t)}$

Add-One-Smoothing

- $P_{Pro}(t) = \frac{freq_{t,Pro} + 1}{\sum_{t' \in Pro} freq_{t',Pro} + \sum_{t' \in Args} 1}$
- $P_{Con}(t) = \frac{freq_{t,Con} + 1}{\sum_{t' \in Con} freq_{t',Con} + \sum_{t' \in Args} 1}$

Top 5 Pro-Worte für Topic nuclear Power

nuclear power

women
pro-choice
child
human
pregnancy

- Subjectivity Lexicon der University of Pittsburgh mit über 8000 Termen

- Subjectivity Lexicon der University of Pittsburgh mit über 8000 Termen
- Satzkoookkurrenzen bestimmen Relevanz der Terme

- Subjectivity Lexicon der University of Pittsburgh mit über 8000 Termen
- Satzkontexten bestimmen Relevanz der Terme
- Korpus des Wortschatz-Projekts nicht ausreichend groß

- Subjectivity Lexicon der University of Pittsburgh mit über 8000 Termen
- Satzkontexten bestimmen Relevanz der Terme
- Korpus des Wortschatz-Projekts nicht ausreichend groß
- Anfrage an chatnoir.eu





Maasen et. al. (2006)

Bilder als Diskurse - Bilddiskurse

Velbrück Wissenschaft



Maasen et. al. (2006)

Bilder als Diskurse - Bilddiskurse

Velbrück Wissenschaft



Heiskanen (2017)

Meme-ing Electoral Participation

European journal of American studies

-  Maasen et. al. (2006)
Bilder als Diskurse - Bilddiskurse
Velbrück Wissenschaft
-  Heiskanen (2017)
Meme-ing Electoral Participation
European journal of American studies
-  Wachsmuth et. al (2017)
Building an Argument Search Engine for the Web
Association for Computational Linguistics



Maasen et. al. (2006)

Bilder als Diskurse - Bilddiskurse

Velbrück Wissenschaft



Heiskanen (2017)

Meme-ing Electoral Participation

European journal of American studies



Wachsmuth et. al (2017)

Building an Argument Search Engine for the Web

Association for Computational Linguistics



Stab et. al (2018)

ArgumenText: Searching for Arguments in Heterogeneous Sources

Association for Computational Linguistics



Maasen et. al. (2006)

Bilder als Diskurse - Bilddiskurse

Velbrück Wissenschaft



Heiskanen (2017)

Meme-ing Electoral Participation

European journal of American studies



Wachsmuth et. al (2017)

Building an Argument Search Engine for the Web

Association for Computational Linguistics



Stab et. al (2018)

ArgumenText: Searching for Arguments in Heterogeneous Sources

Association for Computational Linguistics



Bar-Ilan (2001)

Data collection methods on the Web for informetric purposes A review and analysis

Kluwer Academic Publishers



Uyar et. al. (2017)

Investigating the precision of Web image search engines for popular and less popular entities

Journal of Information Science



Uyar et. al. (2017)

Investigating the precision of Web image search engines for popular and less popular entities

Journal of Information Science



Strathmann (2016)

Startpage und Ixquick: Zwei diskrete Suchmaschinen wachsen zusammen

Die Zeit



Uyar et. al. (2017)

Investigating the precision of Web image search engines for popular and less popular entities

Journal of Information Science



Strathmann (2016)

Startpage und Ixquick: Zwei diskrete Suchmaschinen wachsen zusammen

Die Zeit



Ormiston (2006)

Denial-of-Service Distributed Denial-of-Service on The Internet



Uyar et. al. (2017)

Investigating the precision of Web image search engines for popular and less popular entities

Journal of Information Science



Strathmann (2016)

Startpage und Ixquick: Zwei diskrete Suchmaschinen wachsen zusammen

Die Zeit



Ormiston (2006)

Denial-of-Service Distributed Denial-of-Service on The Internet



Zhao (2017)

Web Scraping

Springer International Publishing



Boleda (2020)

Distributional Semantics and Linguistic Theory

Annual Re- view of Linguistics



Boleda (2020)

Distributional Semantics and Linguistic Theory

Annual Review of Linguistics



Wilson et. al. (2005)

Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis

Danke für eure Aufmerksamkeit!