Finale Präsentation
– Opinion Mining –

Vortrag im Rahmen der Vorlesung
Data Warehouse

Dozentin: Prof. Dr. Frey-Luxemburger
WS 2011/2012

HOCHSCHULE FURTWANGEN UNIVERSITY | HFU

Referent: Florian Kalisch (GR09)

Agenda

- Einleitung
- Rückblick
- Opinion Mining
  - Einführung
  - Theoretische Grundlagen
    - Basiseinheiten
    - Meinungsdefinition
    - Arten
- Projekt
  - Vorstellung der Tools
  - Praxisversuch / Prozess
- Fazit

Einleitung

- Projektziele:
  - Näherbringung des Forschungsbereichs Opinion Mining
  - Prototypische Implementierung

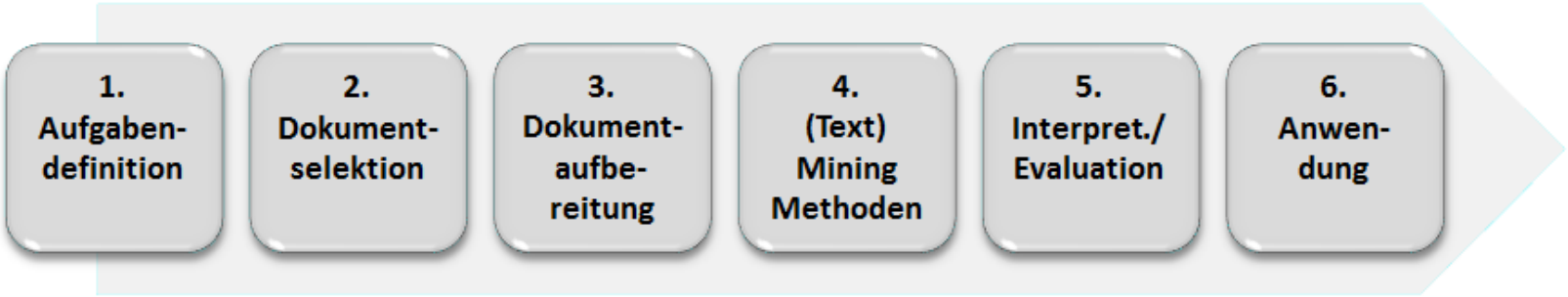
- Ziele dieser Abschlusspräsentation
  - Kurzer Rückblick auf die Meilensteinpräsentationen
  - Vertiefte Betrachtung von Opinion Mining

Einleitung

- Ziele dieser Abschlusspräsentation
  - Einblick in den praktischen Teil
    - Anforderungen
    - Vorstellung gewählter Tools
    - Umgesetzte und fehlende Teile des Prototypen
    - Zuordnung der Tools zu den Prozessphasen
    - Probleme
- Fazit

Rückblick

- Meilenstein 1
  - Schwerpunkt lag auf dem Text-Mining
  - Vorstellung des Text-Mining Prozesses



  - Verwendung des Prozesses beim Opinion Mining ist möglich

Rückblick

- Meilenstein 2
  - Grobe Vorstellung des Opinion Mining
  - Einblicke in die Praxis

  - Offene Arbeitspakete waren:
    - Recherche nach Tools
    - Auswahl eines Tools
    - Definition der Praxis-Anforderungen
    - Praktische Umsetzung

Opinion Mining / Einführung

- Definitionen
  - Bing Liu:

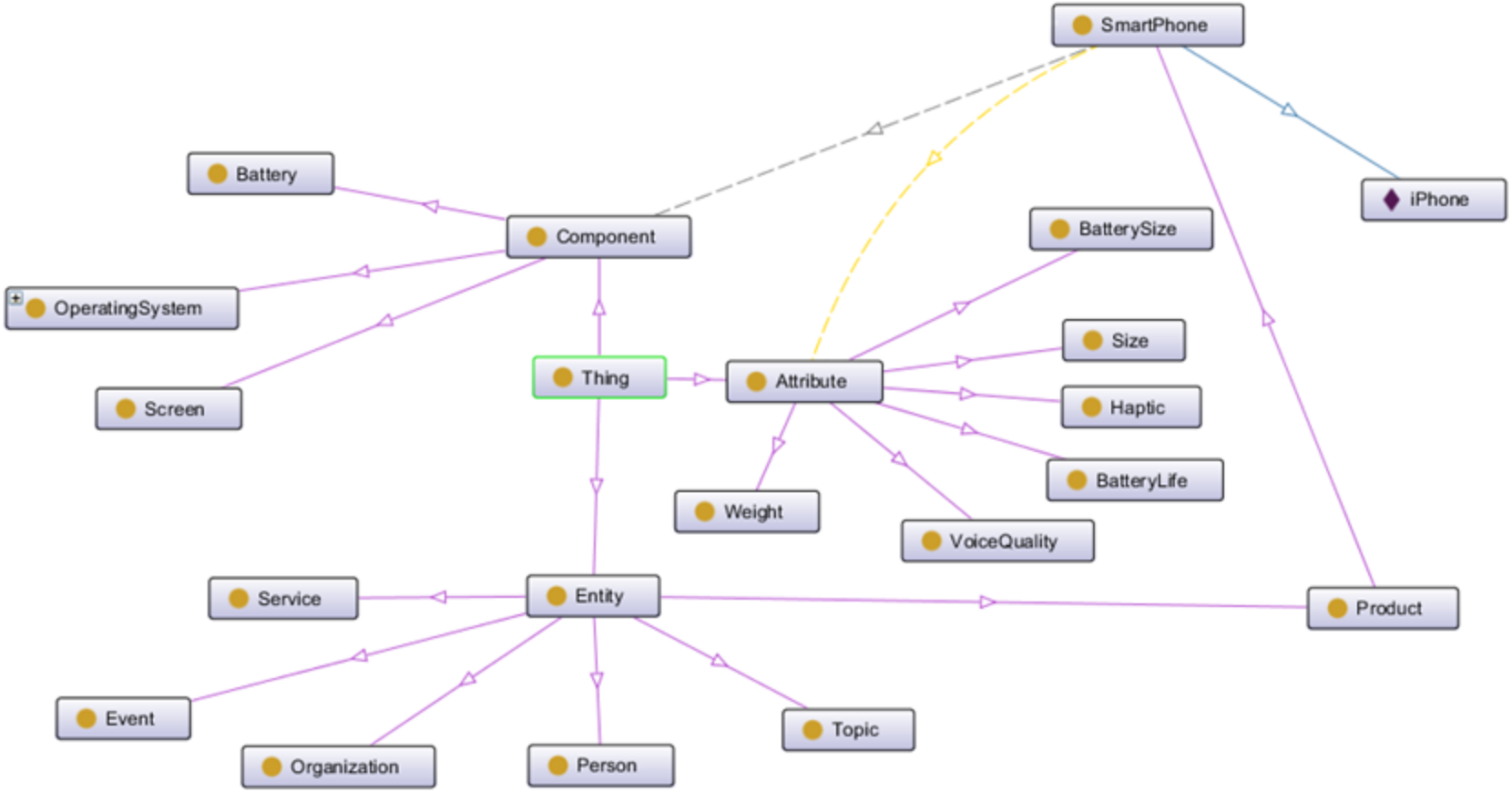
    „Given a set of evaluative text documents D that contain opinions (or sentiments) about an object, opinion mining aims to extract attributes and components of the object that have been commented on in each document d ∈ D and to determine whether the comments are positive, negative or neutral.“

  - Lee et al.:

    “The task of analyzing such data, collectively called customer feedback data, is known as opinion mining.”

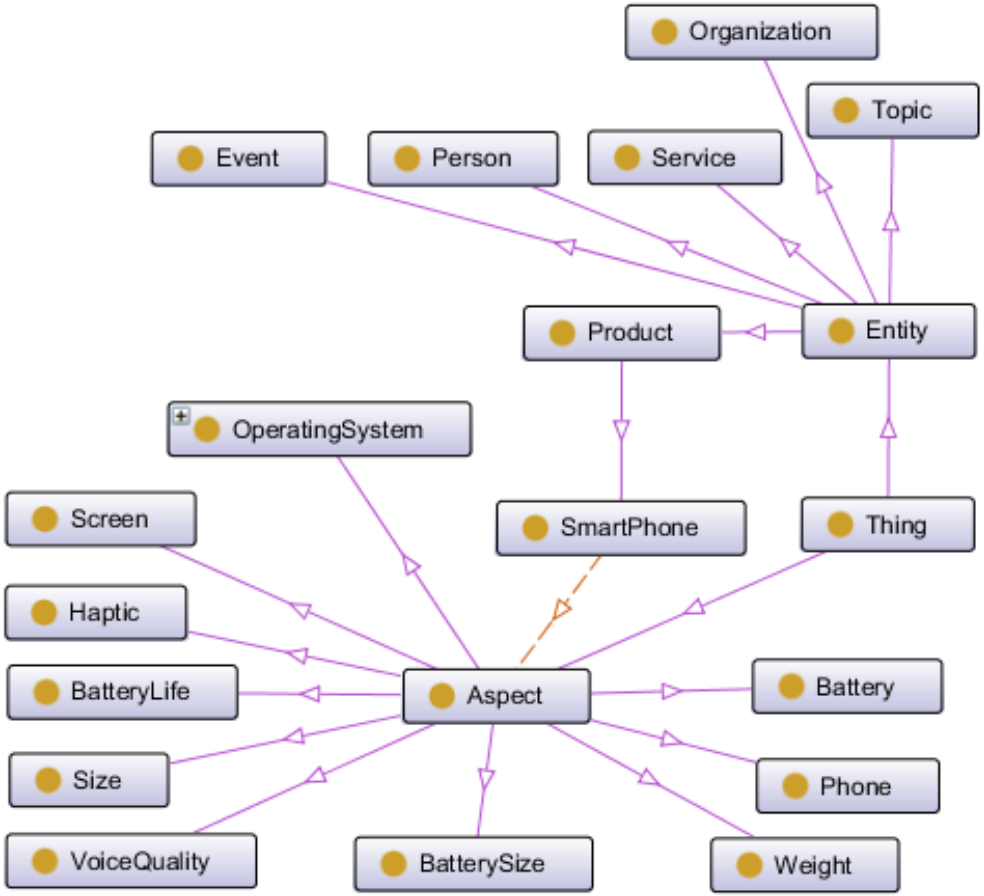
Opinion Mining / Grundlagen

- Basiseinheiten (komplexe Darstellung)



Opinion Mining / Grundlagen

- Basiseinheiten (vereinfachte Darstellung)



Opinion Mining / Grundlagen

- Aspekte und Entitäten besitzen
  - Verschiedene Ausdrucksweisen („expressions“)
    - Die Namen können in Texten unterschiedlich geschrieben werden
  - Sowie einen Namen
    - Gruppiert die Ausdrucksweisen zu einem einheitlichen Namen

Opinion Mining / Grundlagen

- Aspekt-Name:
  - Screen
- Aspekt-Ausdrucksweisen:
  - Touchscreen, Display

- Entitäten-Name:
  - Motorola
- Entitäten-Ausdrucksweise
  - Z.B. „Mot“ oder „Moto“

Opinion Mining / Grundlagen

- Definition der Meinung
  - Eine Meinung wird durch folgendes Quadrupel ausgedrückt:

$$(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$$

Entität	e_i
Aspekt	a_{ij}
Meinungsausrichtung	oo_{ijkl}
Meinungsäußerer	h_k
Zeitpunkt	t_l

Opinion Mining / Grundlagen

- Grundlegende Arten des Opinion Mining
  - Verschiedene Verfahren mit unterschiedlicher Zielsetzung
  - Vom Dokumenten-Fokus zum Aspekt-Fokus

1	Document Sentiment Classification
2	Sentence Subjectivity and Sentiment Classification
3	Aspect-Based Opinion Mining

Opinion Mining / Grundlagen

- Document Sentiment Classification
  - Klassifikation eines ganzen Dokumentes in Bezug auf die darin ausgedrückte Meinung
  - Zugrunde liegender Ansatz:

$$(e, GENERAL, oo, h, t)$$

Entität	e	Vorannahme: nur ein e
Aspekt	GENERAL	
Meinungsausrichtung	oo	Meinung bezieht sich direkt auf e
Meinungsäußerer	h	Vorannahme: nur ein h
Zeitpunkt	t	

Opinion Mining / Grundlagen

- Sentence Subjectivity and Sentiment Classification
  - Zwei Teilaufgaben
    - Subjectivity classification
    - Sentence-level sentiment classification
  - Kann als Zwischenschritt dienen
  - Das Wissen, dass ein einzelner Satz eine positive oder negative Meinung ausdrückt reicht oft nicht aus

Opinion Mining / Grundlagen

- Aspect-Based Opinion Mining
  - Auch wenn ein Dokument als negativ klassifiziert wurde, kann es positiv bewertete Aspekte darin geben
  - Dazu muss der Fokus auf die Aspekte gelegt werden
  - Im Gegensatz zur Behandlung als Klassifikationsproblem, müssen umfangreichere Methoden des Natural Language Processing eingesetzt werden

Opinion Mining / Grundlagen

- Ziel:
  - Ermittlung aller „Meinungs-Quadrupel“ in einer Dokumentensammlung D
- Schrittfolgen zur Problemlösung
  1. Extrahierung aller Entität-Ausdrücke und Gruppierung synonymer Ausdrücke in Cluster

     $=> (\boldsymbol{e_i}, a_{ij}, oo_{ijkl}, h_k, t_l)$
  2. Extrahierung aller Aspekt-Ausdrücke zu einer Entität und Gruppiere diese in Cluster

     $=> (e_i, \boldsymbol{a_{ij}}, oo_{ijkl}, h_k, t_l)$

Opinion Mining / Grundlagen

- Schrittfolgen
  3. Extrahierung von Meinungsäußerer und Zeitpunkt

     $=> (e_i, a_{ij}, oo_{ijkl}, \boldsymbol{h_k}, \boldsymbol{t_l})$
  4. Ermittlung, ob die geäußerte Meinung zu einem Aspekt positiv, neutral oder negativ ist

     $=> (e_i, a_{ij}, \boldsymbol{oo_{ijkl}}, h_k, t_l)$
  5. Erzeugung aller „Meinungs-Quadrupel“ in Dokument D, basierend auf den Schritten 1 bis 4

     $=> (\boldsymbol{e_i}, \boldsymbol{a_{ij}}, \boldsymbol{oo_{ijkl}}, \boldsymbol{h_k}, \boldsymbol{t_l})$

Opinion Mining / Grundlagen

- Beispiel zur Verdeutlichung
- Fiktives Review:

> Geschrieben von: SeelenPluecker am 16.01.2012
> Ich habe mir vor ein paar Tagen ein Motorola Tablet gekauft und meine Freundin sich ein Tablet von Apple. Als wir daheim waren, testeten wir beide unsere Geräte. Der Touchscreen meines Mot fühlte sich sehr träge an, während die Haptik jedoch sehr gut war. Meine Freundin war sehr zufrieden mit ihrem Äpfelchen und dem Screen. Ich möchte aber ein Tablet mit gutem Display. Wahrscheinlich tausche ich es um.

- Anwendung der Schrittfolge auf obiges Beispiel.

Opinion Mining / Grundlagen

1. Extrahierte Entitäten mit Ausdrücken:
   - Motorola {Mot}, Apple {Äpfelchen}
2. Extrahierte Aspekte mit Ausdrücken:
   - Display {Touchscreen, Screen}
   - Haptik {}
3. Extrahierte Meinungsäußerer und Datum
   - SeelenPluecker, Freundin_von_ SeelenPluecker
   - 16.02.2012

Opinion Mining / Grundlagen

4. Extrahierte Meinungen
   - „Der Touchscreen meines Mot fühlte sich sehr träge an, während die Haptik jedoch sehr gut war.“
     - Negativ: Display des Motorola
     - Positiv: Haptik des Motorola
   - „ Meine Freundin war sehr zufrieden mit ihrem Äpfelchen und dem Screen.“
     - Positive Äußerung auf gesamtes Gerät
     - Positiv: Display des Apple Tablets
5. Erzeugte „Meinungs-Quadrupel“
   - (Motorola, Display, Negativ, SeelenPluecker, 16.02.2012)
   - (Motorola, Haptik, Positiv, SeelenPluecker, 16.02.2012)
   - (Apple, Allgemein, Positiv, Freundin_SeelenPluecker, 16.02.2012)
   - (Apple, Display, Positiv, Freundin_SeelenPluecker, 16.02.2012)

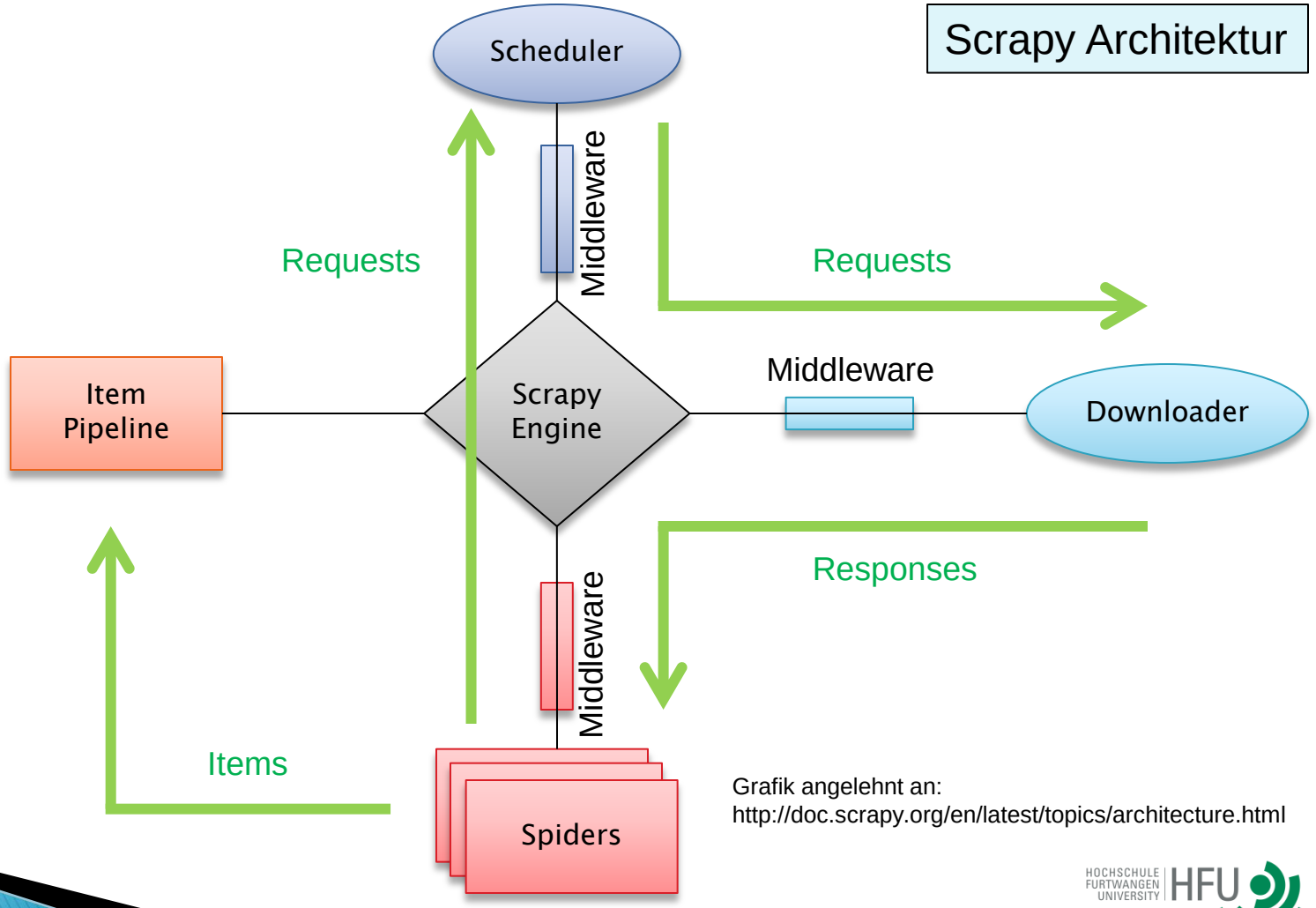
Opinion Mining / Projekt

Vorstellung der Tools

Opinion Mining / Projekt

- Scrapy
  - Framework für Web-Crawling und Screen-Scraping
  - Entwickelt in Python
  - Hauptaspekte
    - Einfach
    - Produktiv
    - Schnell
    - Erweiterbar

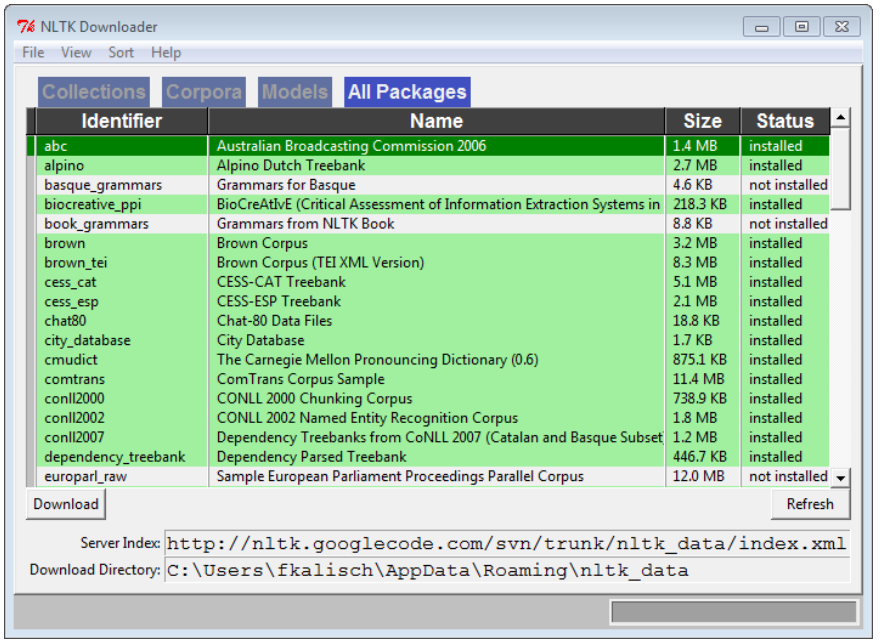
Opinion-Mining / Projekt



Grafik angelehnt an:
http://doc.scrapy.org/en/latest/topics/architecture.html

Opinion Mining / Projekt

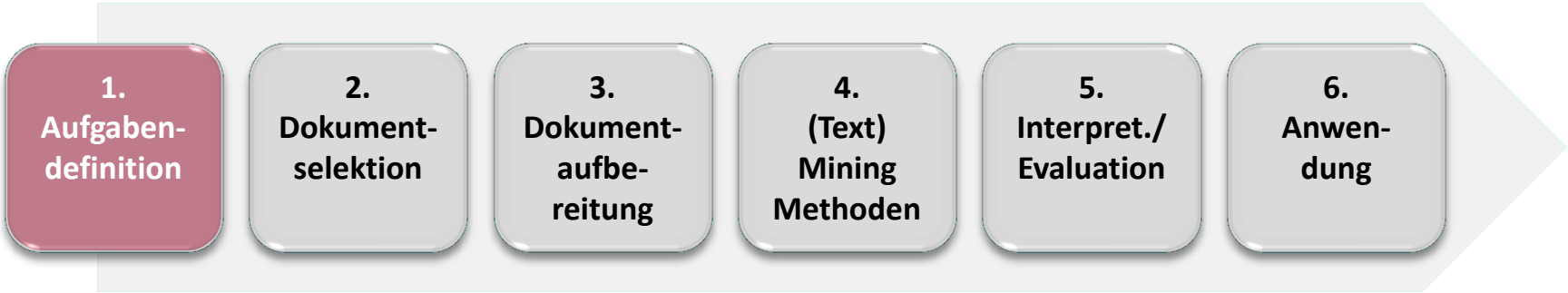
- Natural Language Toolkit
  - Bibliotheken zur Python-Erweiterung
  - Nutzbar in den Bereichen
    - Computerlinguistik
    - Künstliche Intelligenz
  - Umfangreiche Beispieldaten



Opinion Mining / Projekt

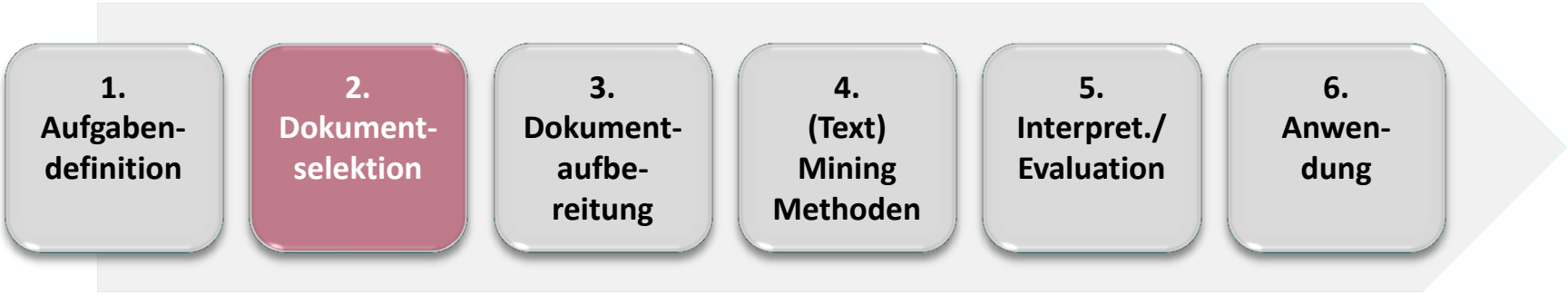
Vorstellung der Praxis
anhand des
Opinion Mining Prozesses

Opinion Mining / Projekt



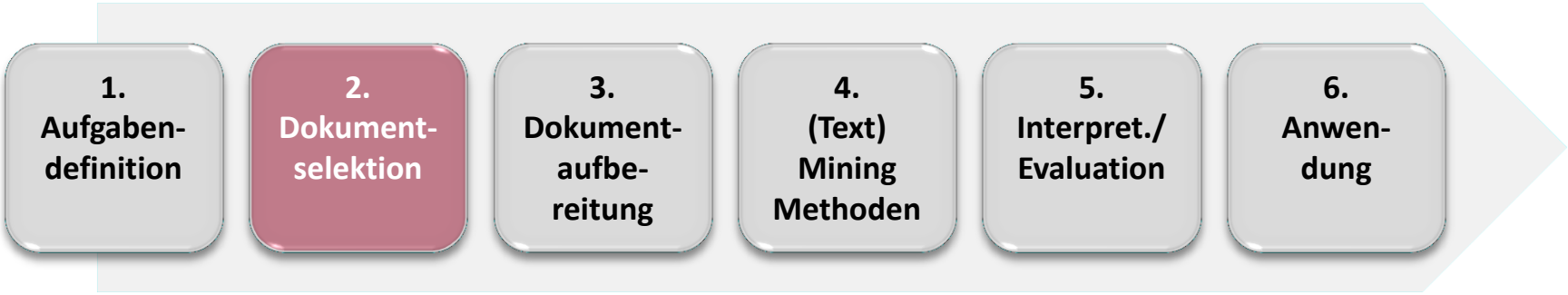
- 1. Aufgabendefinition
  - Halbautomatische Extraktion von englischen Kundenmeinungen zum iPhone 4
  - Vorbereitung der Daten für die Meinungsklassifikation auf Dokumentenebene
  - Anwendung des Opinion Mining

Opinion Mining / Projekt



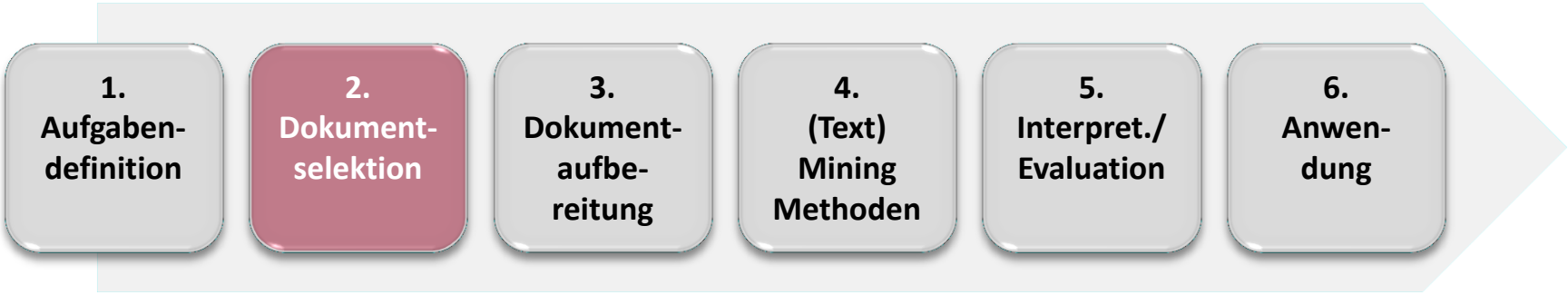
- 2. Dokumentselektion
  - Beschränkung auf die Kundenbewertungen von www.amazon.com
  - Erster Versuch:
    - Nutzung der Amazon API

Opinion Mining / Projekt



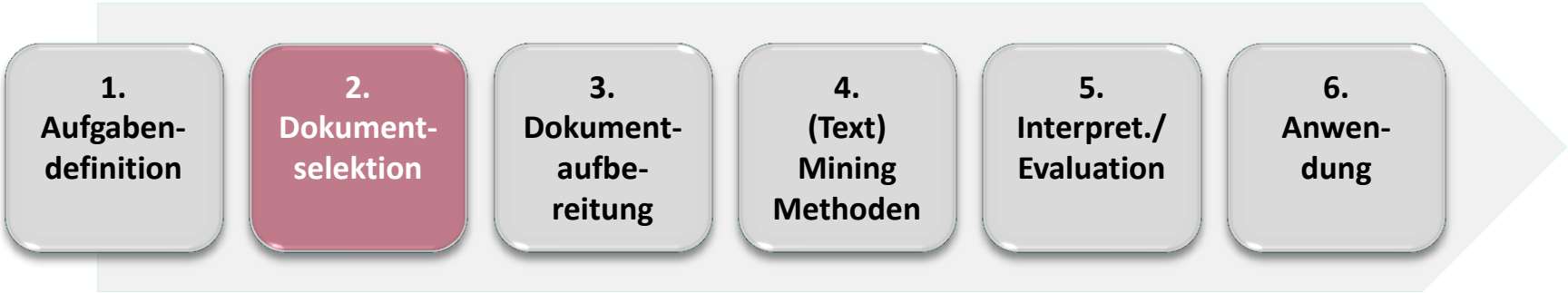
- 2. Dokumentselektion
  - Nutzung der Amazon API
    - Häufige Änderungen an den API-Schnittstellen
    - Unklare Dokumentation
    - Das direkte Auslesen der Reviews wird nicht mehr unterstützt
    - Amazon-Partner (mit Partner-ID) bekommen Zugriff auf HTML-Source einzelner Reviews

Opinion Mining / Projekt



- 2. Dokumentselektion
  - Nutzung der Amazon API
    - Keine weitere Verfolgung dieses Lösungsweges
  - Zweiter Versuch:
    - Nutzung des Web-Scraping Frameworks Scrapy

Opinion Mining / Projekt

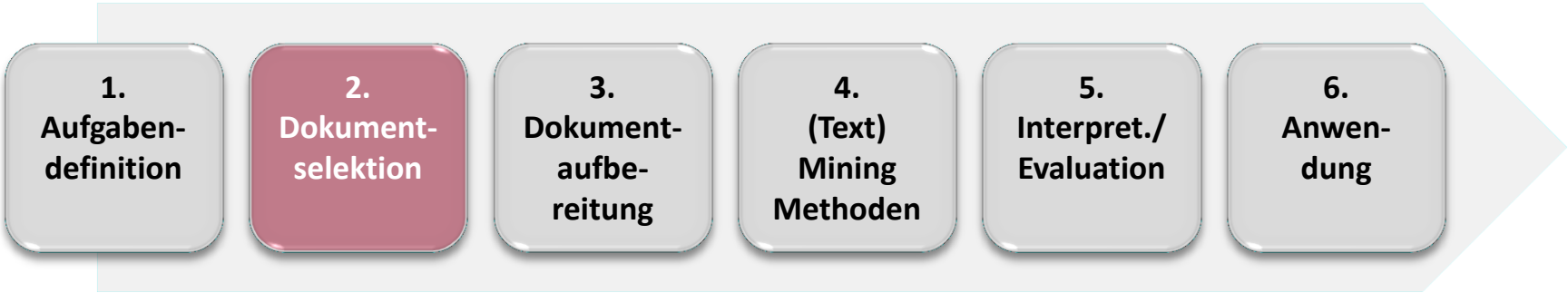


- 2. Dokumentselektion
  - Scrapy Item Definition (Ausschnitt):

```
from scrapy.item import Item, Field

class AmazonItem(Item):
        helpful = Field()
        rating = Field()
        title = Field()
        reviewDate = Field()
        author = Field()
        productToReview = Field()
        review = Field()
```

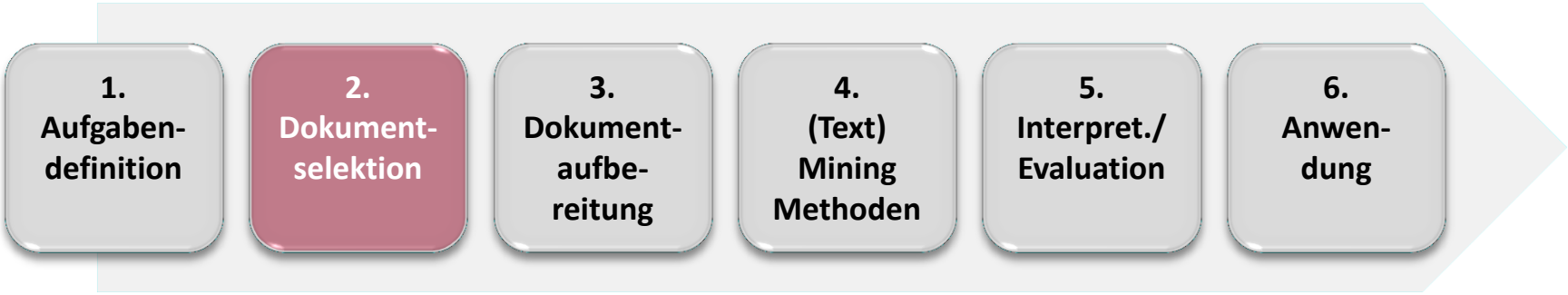
Opinion Mining / Projekt



- 2. Dokumentselektion
  - Scrapy parser (Ausschnitt):

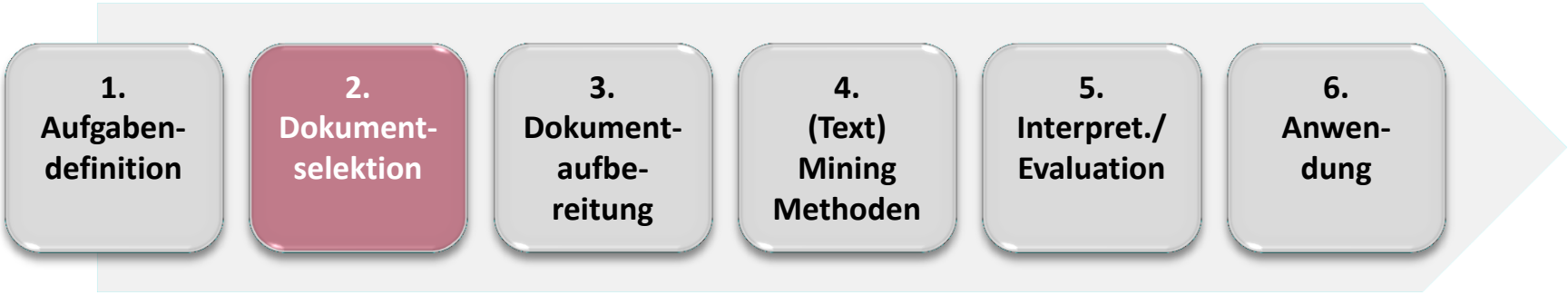
```
def parse(self, response):
    hxs = HtmlXPathSelector(response)
    items = []
    for i in range(1,11,1):
     item = AmazonItem()
     item['helpful'] = hxs.select('//html//body//table[@id=\'productReviews\']//tr//td[1]//div[' + str(i) + ']/div[1]/text()').extract()
     items.append(item)
    return items
```

Opinion Mining / Projekt



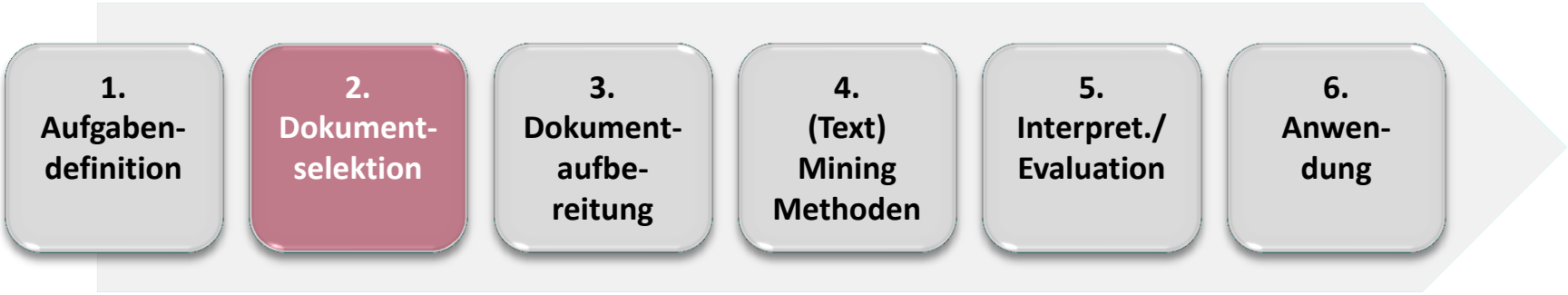
- 2. Dokumentselektion
  - Start des Web-Crawling
    - scrapy crawl amazon -o revitems.xml -t xml
    - Export der Items im XML-Format in die Datei revitems.xml

Opinion Mining / Projekt



- 2. Dokumentselektion
  - Web-Crawling
    - Nur teilweise erfolgreich, da die Amazon-Reviews ungenügende HTML-Struktur aufweisen
      - => Extraktion von nicht benötigtem Text
      - => Absolute XPATH-Angaben führen zu fehlerhaften Extraktionen, da erstes Tag unterschiedliches bedeuten kann

Opinion Mining / Projekt



- 2. Dokumentselektion
  - Dritter Versuch:
    - Manuelle Extraktion
    - 57 Kundenmeinungen ausgelesen

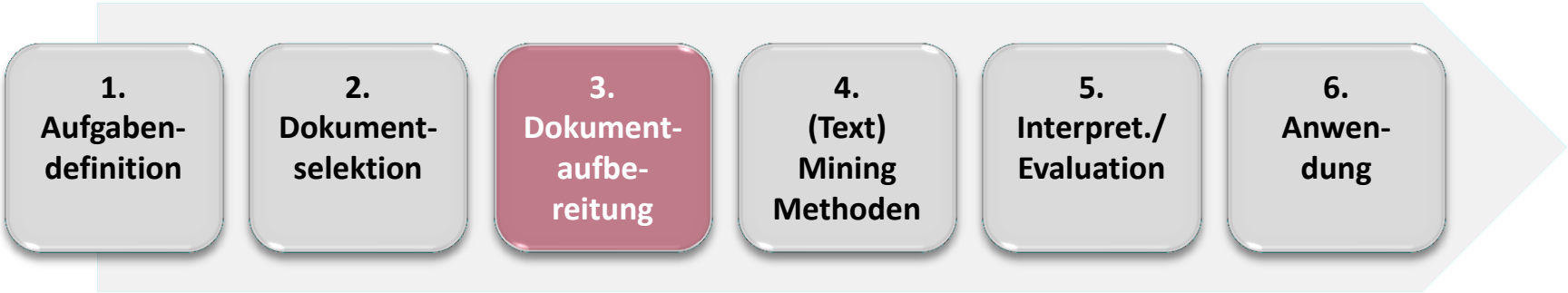
Opinion Mining / Projekt

1. Aufgaben-definition
2. Dokument-selektion
3. Dokument-aufbe-reitung
4. (Text) Mining Methoden
5. Interpret./ Evaluation
6. Anwen-dung

- 2. Dokumentselektion

	A	B	C	D	E	F
1	helpful	rating	title	reviewDate	author	review
2	83 of 94	5.0 out of 5 stars	The 16GB and 32GB are the same exact phone.	December 9, 2010	Ghenghis	Below is my review of the 16GB model that I purchas
3	24 of 25	5.0 out of 5 stars	WARNING TO ALL POTENTIAL BUYERS!	June 4, 2011	R. Lanham "LoveMyKindle"	If you are looking to buy an iPhone-whether 3, 3GS,
4	25 of 28	4.0 out of 5 stars	I can never go back!	February 7, 2011	C. Au "sophisticated"	The iPhone4 is the first smartphone I ever owned so
5	35 of 49	1.0 out of 5 stars	Drops Calls and has very poor signal strenght and reception!	February 16, 2011	Dman "Dman"	I loved this phone but had to return it because it dro
6	2 of 2	5.0 out of 5 stars	Great Device / Poor Network	May 11, 2011	CheapButConscientious	I have owned this product for 10 months now and th
7	2 of 2	5.0 out of 5 stars	Awesome phone!	March 6, 2011	MC "rebelcovehunter"	Best phone ever made, destroys the droid, I have thi
8	4 of 5	4.0 out of 5 stars	Don't leave home without it	January 8, 2011	Lisa Herbertson "Author of 'Swimming Along'"	I'm a traditionalist when it comes to mobile phones
9	10 of 14	5.0 out of 5 stars	Papa's Pocket 'Puter	January 8, 2011	John M. Ford "johnDC"	I've had my iPhone 4 for months, after an upgrade fr
10	3 of 4	3.0 out of 5 stars	Watch out for the latest operating system	October 27, 2011	K. Bradford "Kimmie92592"	Ive had this phone for over a year now and as far as a
11	7 of 10	4.0 out of 5 stars	MIXED BAG OF GOODNESS	December 9, 2010	Margaux Paschke	Let me start by noting that I am new to the iPhone u
12	2 of 3	3.0 out of 5 stars	Its ok	August 3, 2011	Michelle	I bought this phone before doing my research, later
13	2 of 3	4.0 out of 5 stars	Solid, but not perfect smartphone.	June 21, 2011	D. Gong	I have had my iPhone 4 for almost a year and before
14		5.0 out of 5 stars	Confessions of an iPhone Addict	January 3, 2012	David Ashlock	I admit it -- yes, I am an addict. I had never used an i
15		5.0 out of 5 stars	Excellent celphone	December 22, 2011	Quick	The condition is excellent and it is an Apple product
16		5.0 out of 5 stars	FANTASTIC product, AWESOME seller!!!!	December 21, 2011	veronica figueiredo	I didn't have a chance to rate this product and the se
17		4.0 out of 5 stars	Amazing but technical problems	December 20, 2011	S. Miller "Stillwater Traveler"	The Apple iPhone 4 is amazing and we are learning h
18		5.0 out of 5 stars	The most user friendly phone yet.	December 19, 2011	Bzom-b	As these iPhones progress they just keep getting be
19		5.0 out of 5 stars	Excellent!	December 1, 2011	Nice iPhone	the iPhone was in excellente conditions, it was ever
20		5.0 out of 5 stars	DO IT	November 28, 2011	sassafras	I bought this product about three weeks ago and it w
21		5.0 out of 5 stars	THE BEST!	November 27, 2011	Thinkman "SmartEnoughToKnowBetter"	The iPhone 4 has become like an extension of my ar
22		5.0 out of 5 stars	iPhone with Great Condition	October 31, 2011	techguy	Used iPhone with a great condition. LCD Lights leak t

Opinion Mining / Projekt

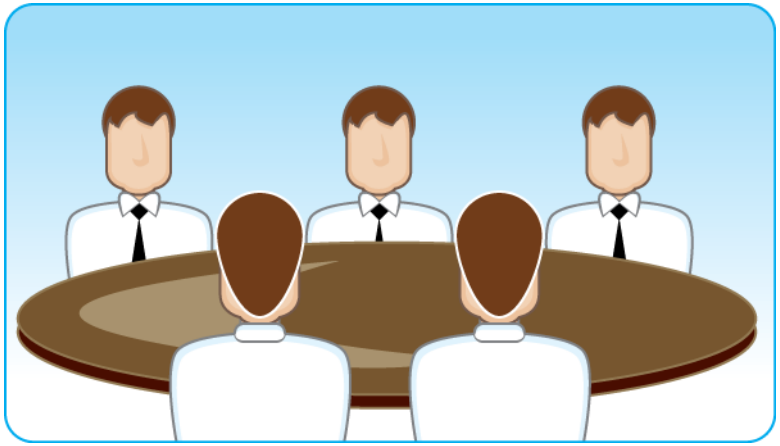


- 3. Dokumentaufbereitung
  - Abbruch der praktischen Umsetzung an dieser Stelle
  - Alternative:
    - Versucht, einen Demozugang zu ConsumberBase oder OpinionEQ zu erlangen
      - => war leider nicht möglich

Opinion Mining / Fazit

- Sehr interessanter Forschungsbereich
- Durch dieses Projekt einen guten Überblick über das Themengebiet bekommen
- Komplexität in Bezug auf die praktische Umsetzung unterschätzt
- Opinion Mining auf Ebene der Aspekte am aufwändigsten, birgt aber die größten Potenziale

Fragen? Anregungen?



Literatur

- [HamO10] Hammer, T.: Opinion und Relationship Mining in sozialen Netzwerken. Extraktion von Meinungen und Beziehungen mittels Textmining und sozialer Netzwerkanalyse. VDM Verlag Dr. Müller, 2010.
- [InaO10] Ina Kimmling: Opinion Mining, Koblenz, 2010.
- [IndH10] Indurkhya, N.; Damerau, F. J.: Handbook of natural language processing. Chapman & Hall/CRC, Boca Raton, FL, 2010.
- [LeeO08] Lee, D.; Jeong, O.-R.; Lee, S.-g.: Opinion mining of customer feedback data on the web: Proceedings of the 2nd international conference on Ubiquitous information management and communication. ACM, New York, NY, USA, 2008; S. 230-235.