

DIPAS NLP API Docs v0.6

API Docs (this): <https://docs.google.com/document/d/1vU1L5mbTOGJ5S9cNLYdmwqVUUq699ee6jIL0MXu3SJ8/>

NLP Slides: <https://docs.google.com/presentation/d/1E8W6u7My4FWKOlgTf5wUyLFa1XSu5jT6JK-nhkQUrF8/>

NLP Server Information	2
Current configuration (IP & Ports)	2
Microservice Architecture	3
Keyword-service (Fortiss)	4
Keyword-service (CSL)	5
Score-service (CSL)	8
Wordcloud	11
LDA Topic Modelling (used for Clustering)	13
Summary	16
Topicmaps	17
Licenses	17
User Stories	17
Listenansicht	18
Wordcloud	18
Cluster	18
Heatmap	18
Further Development	19
Backlog	19
Document History	19

NLP Server Information

The NLP Server is currently hosted and available until February 2023 on the HCU server data Infrastructure. The keyword service is currently available under

<http://194.95.76.31:10001/keywords>

Test the access from the browser <http://194.95.76.31:10001/> -> Hello, World!

Current configuration (IP & Ports)

Keywords Service (fortiss)

<http://analytics.hcu-hamburg.de:10001/keywords>

Legacy <http://194.95.76.31:10001/keywords>

Keywords Service (csl)

<http://analytics.hcu-hamburg.de:10005/keywords-csl>

Legacy <http://194.95.76.31:10005/keywords-csl>

Clustering Service

<http://analytics.hcu-hamburg.de:10002/clustering>

Legacy <http://194.95.76.31:10002/clustering>

Score Service

<http://analytics.hcu-hamburg.de:10003/scores>

Legacy <http://194.95.76.31:10003/scores>

Wordcloud Service

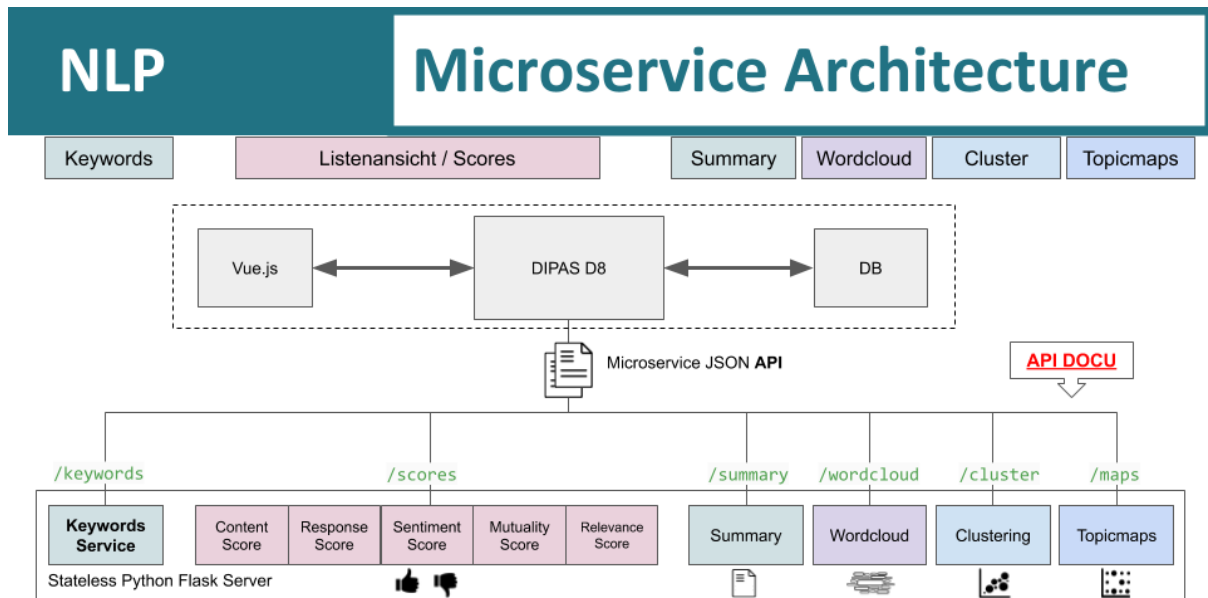
<http://analytics.hcu-hamburg.de:10004/wordcloud>

Legacy <http://194.95.76.31:10004/wordcloud>

These ports are open to the internet

Legacy for the following IPs (LGV), **Prod.: 51.5.149.160, Test.: 51.5.151.173**

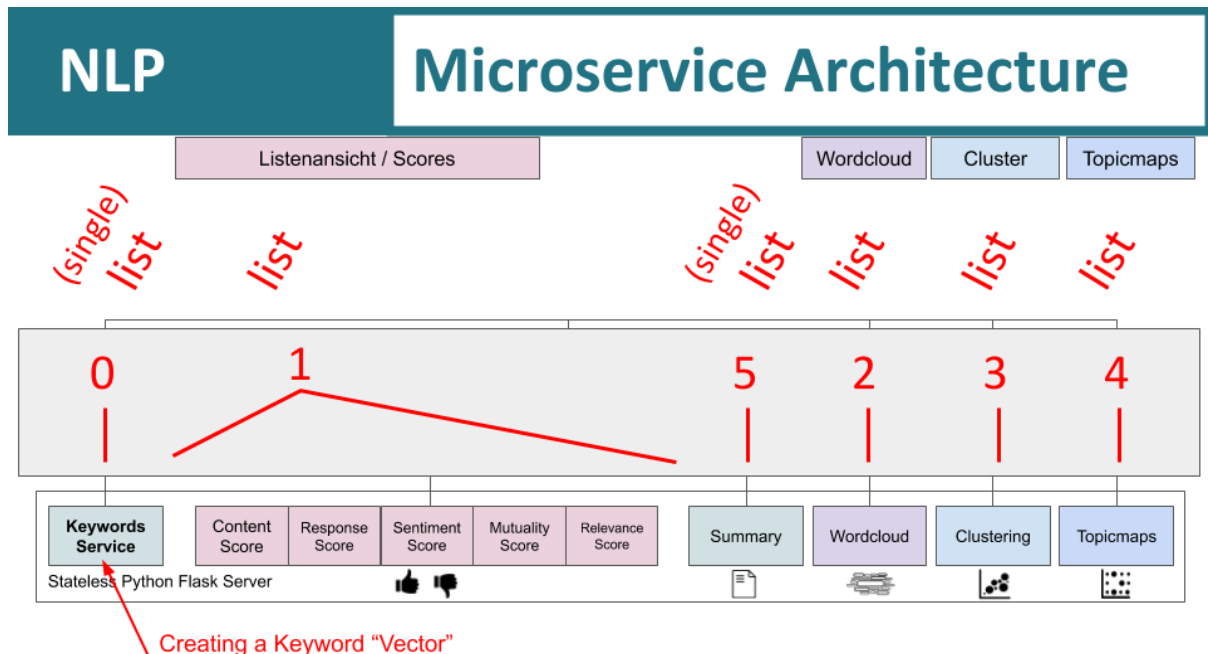
Microservice Architecture



Modular approach:

- all submodules are accessible individually
- Wrapping API allows access via single point via modular *request template*

<http://194.95.76.31:10001/scores>



NLP Services will be provided in a microservice architecture. That means that all functions are going to be exposed as APIs at the NLP Flask web server.

Keyword-service (Fortiss)

<https://git.fortiss.org/civitas-digitalis/CivDig-DigitalAssistant/ai-services/keyword-service>

The keyword service is a micro-service that suggests keywords to ideas using external services like Wortschatz Leipzig and DBPedia.

POST request to `/keywords`

Parameters:

description: Body of text that the keywords are being generated from

mode:

- internal (default, keywords are generated locally, no external service used),
- external (forwarding to external keyword service: DBpedia, Keyword Leipzig. no longer supported, because unreliable in terms of availability and response time)

externalService:

- DBPEDIA
- LEIPZIG

url: `http://keywordset.net`

num: 0-25 - number of max keywords returned

```
{
  "description": "Eine Erhol-Oase mit einem zentralen Brunnen und
  einer steinernen Bank drumrum, auf welcher man verweilen kann. Außerdem
  gibt es ein bachartiges Gewässer außenrum und Steinplatten, über die man
  die Wasserflächen queren kann. In den Wasserbecken soll es Fische und
  Wasserpflanzen geben.",
  "mode": "external",
  "externalService": "DBPedia",
  "url": "http://keywordset.net",
  "num": 6
}
```

Result:

```
[
  "geben",
  "Gewässer",
  "Brunnen",
  "Bank",
  "Wasserpflanzen",
```

```
"zentralen"  
]
```

Keyword-service (CSL)

Source: <https://github.com/hlgrprng/keyword-service>

The keyword service is a micro-service that suggests keywords, based on part-of-speech and entity-recognition of spacy.io (<https://spacy.io/>)

Usage:

POST request to `/keywords-cs1`

Parameters:

configuration: contains the configuration for the wrapper API

keywords:

- OFF (no keywords are generated)
- internal (default, keywords are generated locally)
- DBPEDIA (not implemented, unreliable)
- LEIPZIG (not implemented, unreliable)

externalUrl: <http://keywordset.net>

language: de, (en, es / not supported yet)

num: 0-25 - number of max keywords returned

pos: List of part-of-speech that will be returned as keywords
NOUN, ADJ, VERB, ...
Reference: <https://universaldependencies.org/u/pos/>

entities: List of entitie recognition that will be returned as keywords
PERSON, ORG, GPE, LOC, ...
<https://spacy.io/api/annotation#named-entities>

sentiments: OFF, SentiWS_v2.0, (Spacy compatible sentiments)

stoplist: Add specific stopwords, which are not indexed

ordered: Returns keywords in a specific order
(read more in *Keyword ordering*)

dictionary: OFF, alphabetic, invers
Generates a "dictionary" at the end of the results

spellchecking: 0-10, performs spell check and eliminates wrong spelling
(<http://hunspell.github.io/>)

documents: id, title, contributionText of contributionOriginal Object (PDS)

Body:

```
{
  "configuration": {
    "keywords": "internal",
    "externalUrl": "",
    "language": "de",
    "num": 10,
    "pos": ["NOUNS", "ADJ"],
    "entities": ["PERSON", "LOC"],
    "sentiments": "SentiWS_v2.0",
    "stoplist": ["Grasbrook", "Hamburg"],
    "ordered": "OFF",
    "dictionary": "invers",
    "spellchecking": 4
  },
  "documents": [{
    "id": "82734",
    "title": "Eine Erhol-Oase",
    "body": "Eine Erhol-Oase mit einem zentralen Brunnen
und einer steinernen Bank drumrum, auf welcher man verweilen kann.
Außerdem gibt es ein bachartiges Gewässer außenrum und Steinplatten,
über die man die Wasserflächen queren kann. In den Wasserbecken soll es
Fische und Wasserpflanzen geben."
  },
  {
    "id": "82745",
    "title": "Die Ampeln verbessern",
    "body": "Die Fahrbahnen nehmen hier einen gewaltigen
Raum ein und dazwischen sind Grünanlagen, die man nicht betreten kann.
Das ist doch Platzverschwendung. Fußgänger müssen sich umständlich über
mehrere nicht aufeinander abgestimmte Ampeln über die Kreuzung quälen.
Kann das nicht so umgestaltet werden, dass mehr Raum für Fußgänger und
Radfahrer entsteht und das Überqueren der Straße leichter wird"
  }
]
}
```

Result Parameter:

results: list of results per document - JSON objects (id, tags)

id: unique id, referring to

tags: list of keywords / tags

dictionary: dict (dict with all the keywords : number of appearance)

Result:

```
{
  "results": [{
    "id": "82734",
    "tags": [
      "Wasserpflanzen",
      "Wasserflächen",
      "Wasserbecken",
      "Steinplatten",
      "Gewässer",
      "Fische",
      "Erhol-Oase",
      "Brunnen",
      "Bank"
    ]
  },
  {
    "id": "82745",
    "tags": [
      "Raum",
      "Fußgänger",
      "Überqueren",
      "Straße",
      "Radfahrer",
      "Platzverschwendung",
      "Kreuzung",
      "Grünanlagen",
      "Fahrbahnen",
      "Ampeln"
    ]
  }
],
  "dictionary": {
    "Raum": 2,
    "Fußgänger": 2,
    "Überqueren": 1,
    "Wasserpflanzen": 1,
    "Wasserflächen": 1,
    "Wasserbecken": 1,
    "Straße": 1,
    "Steinplatten": 1,
```

```

        "Radfahrer": 1,
        "Platzverschwendung": 1,
        "Kreuzung": 1,
        "Grünanlagen": 1,
        "Gewässer": 1,
        "Fische": 1,
        "Fahrbahnen": 1,
        "Erhol-Oase": 1,
        "Brunnen": 1,
        "Bank": 1,
        "Ampeln": 1
    }
}

```

Score-service (CSL)

- **Content Score**
 - Anzahl Schlagworte absolut (keine Doppelzählungen)
- **Response Score** („Hot right now“, „Trending“)
 - Aggregat aus Kommentaren und Likes & Dislikes
- **Mutuality Score**
 - Anzahl der häufig im Verfahren genannten Schlagworte
- **Sentimentscore** (Polaritätsfaktor)
 - Wort-Ebene Skala +1 - -1; Betragsebene: Gesamtwert
 - Verteilung Likes & Dislikes? Kontrovers
- **Relevance Score**
 - (Content Score 25 %, Response Score 50 %, Mutuality Score 25%)

Usage:

POST request to `/scores`

Parameters:

configuration: contains the configuration for the wrapper API (optional)

contentScore:	normalized (returns 0-100, default), absolute, OFF
responseScore:	normalized (returns 0-100, default), absolute, OFF
mutualityScore:	normalized (returns 0-100, default), absolute, OFF
relevanceScore:	normalized (returns 0-100, default), absolute, OFF
sentimentScore:	normalized (returns 0-100, default), absolute, OFF
stoplist:	Add specific stopwords, which are not indexed
language:	de, (en, es / not supported yet)

documents:

id: unique id of document

title: title of the document

commentsNumber: number of comments to this document

votingPro: number of positive votes

votingContra: number of negative votes

tags: tags or contributionObject (if not existing fallback to description)

body: text of contributionOriginal Object

Body:

```
{
  "configuration": {
    "contentScore": "normalized",
    "responseScore": "normalized",
    "mutualityScore": "normalized",
    "relevanceScore": "normalized",
    "sentimentScore": "normalized",
    "stoplist": ["Grasbrook", "Hamburg"],
    "language": "de"
  },
  "documents": [{
    "id": "82734",
    "title": "Eine Erhol-Oase",
    "commentsNumber": 5,
    "votingPro": 2,
    "votingContra": 1,
    "tags": [
      "Wasserpflanzen",
      "Wasserflächen",
      "Wasserbecken",
      "Steinplatten",
      "Gewässer",
      "Fische",
      "Erhol-Oase",
      "Brunnen",
      "Bank"
    ],
    "body": "Eine Erhol-Oase mit einem zentralen Brunnen
und einer steinernen Bank drumrum, auf welcher man verweilen kann.
Außerdem gibt es ein bachartiges Gewässer außenrum und Steinplatten,
```

über die man die Wasserflächen queren kann. In den Wasserbecken soll es Fische und Wasserpflanzen geben."

```
    },
    {
      "id": "82745",
      "title": "Die Ampeln verbessern",
      "commentsNumber": 2,
      "votingPro": 1,
      "votingContra": 0,
      "tags": [
        "Raum",
        "Fußgänger",
        "Überqueren",
        "Straße",
        "Radfahrer",
        "Platzverschwendung",
        "Kreuzung",
        "Grünanlagen",
        "Fahrbahnen",
        "Ampeln"
      ],
      "body": "Die Fahrbahnen nehmen hier einen gewaltigen Raum ein und dazwischen sind Grünanlagen, die man nicht betreten kann. Das ist doch Platzverschwendung. Fußgänger müssen sich umständlich über mehrere nicht aufeinander abgestimmte Ampeln über die Kreuzung quälen. Kann das nicht so umgestaltet werden, dass mehr Raum für Fußgänger und Radfahrer entsteht und das Überqueren der Straße leichter wird"
    }
  ]
}
```

Result Parameter:

results: list of results per document - JSON objects (id, tags)

id: unique id, referring to document

scores dict: dictionary (score name : number 0-100)

```
{
  "results": [{
    "id": "82734",
    "scores": {
      "content": 48,
      "response": 23,
      "mutuality": 43,
      "relevance": 22,
      "sentiment": 80
    }
  ]
}
```

```

    },
    {
      "id": "82745",
      "scores": {
        "content": 48,
        "response": 23,
        "mutuality": 43,
        "relevance": 22,
        "sentiment": 80
      }
    }
  ]
}

```

Wordcloud

- ...

Usage:

POST request to `/wordcloud`

Parameters:

configuration: contains the configuration for the wrapper API (optional)

language:	de, (en, es / not supported yet)
num:	0-100 - number of max keywords in wordcloud
stoplist:	Add specific stopwords, which are not indexed
dictionary:	OFF, alphabetic, invers Generates a "dictionary" at the end of the results

documents:

id:	unique id of document
title:	title of the document
tags:	tags or contributionObject (if not existing fallback to description)
body:	text of contributionOriginal Object

Body:

```

{
  "configuration": {
    "language": "de",
    "num": 20,
    "stoplist": ["Grasbrook", "Hamburg"],
    "dictionary": "OFF"
  },
  "documents": [{
    "id": "82734",
    "title": "Eine Erhol-Oase",
    "tags": [
      "Wasserpflanzen",
      "Wasserflächen",
      "Wasserbecken",
      "Steinplatten",
      "Gewässer",
      "Fische",
      "Erhol-Oase",
      "Brunnen",
      "Bank"
    ],
    "body": "Eine Erhol-Oase mit einem zentralen Brunnen und einer steinernen Bank drumrum, auf welcher man verweilen kann. Außerdem gibt es ein bachartiges Gewässer außenrum und Steinplatten, über die man die Wasserflächen queren kann. In den Wasserbecken soll es Fische und Wasserpflanzen geben."
  },
  {
    "id": "82745",
    "title": "Die Ampeln verbessern",
    "tags": [
      "Raum",
      "Fußgänger",
      "Überqueren",
      "Straße",
      "Radfahrer",
      "Platzverschwendung",
      "Kreuzung",
      "Grünanlagen",
      "Fahrbahnen",
      "Ampeln"
    ],
    "body": "Die Fahrbahnen nehmen hier einen gewaltigen Raum ein und dazwischen sind Grünanlagen, die man nicht betreten kann. Das ist doch Platzverschwendung. Fußgänger müssen sich umständlich über mehrere nicht aufeinander abgestimmte Ampeln über die Kreuzung quälen. Kann das nicht so umgestaltet werden, dass mehr Raum für Fußgänger und Radfahrer entsteht und das Überqueren der Straße leichter wird"
  }
}

```

```
]
}
```

Result Parameter:

results: list of results per document - JSON objects (id, tags)

id: unique id, referring to document

scores dict: dictionary (score name : number 0-100)

```
{
  "url":
  "https://mfltricks.files.wordpress.com/2012/04/tagxedo_1.png"
}
```

LDA Topic Modelling (used for Clustering)

Source: <https://github.com/hlgrprng/nlp-services>

The [Latent Dirichlet Allocation \(LDA\)](#) is used to classify the contributions to a particular topic. It builds a topic per contribution and words per topic model, modeled as Dirichlet distributions.

Topic models are widely used statistical models for reducing the dimensionality of textual data. Although the approach is quantitative in nature, model selection and validation of topic model results still require qualitative inspection and validation.

- Each contribution is modeled as a multinomial distribution of topics and each topic is modeled as a multinomial distribution of words.
- LDA assumes that the every chunk of text will contain words that are somehow related. Therefore choosing the right corpus of data is crucial.
- It also assumes contributions are produced from a mixture of topics. Those topics then generate words based on their probability distribution, which are used for to give a title to the topic.

Parameters

- **Number** of Topics (Cluster)

Usage:

POST request

Parameters:

configuration: contains the configuration for the wrapper API (optional)

clusterCount: 3-10, number of clusters

stoplist: Add specific stopwords, which are not indexed

language: de, (en, es / not supported yet)

documents:

id: unique id of document

title: title of the document

tags: tags or contributionObject (if not existing fallback to body)

body: text of contributionOriginal Object

Body:

```
{
  "configuration": {
    "clusterCount": "4",
    "stoplist": ["Grasbrook", "Hamburg"],
    "language": "de"
  },
  "documents": [{
    "id": "82734",
    "title": "Eine Erhol-Oase",
    "tags": [
      "Wasserpflanzen",
      "Wasserflächen",
      "Wasserbecken",
      "Steinplatten",
      "Gewässer",
      "Fische",
      "Erhol-Oase",
      "Brunnen",
      "Bank"
    ],
    "body": "Eine Erhol-Oase mit einem zentralen Brunnen und einer steinernen Bank drumrum, auf welcher man verweilen kann. Außerdem gibt es ein bachartiges Gewässer außenrum und Steinplatten, über die man die Wasserflächen queren kann. In den Wasserbecken soll es Fische und Wasserpflanzen geben."
  },
  {
    "id": "82745",
    "title": "Die Ampeln verbessern",
    "tags": [
```

```

        "Raum",
        "Fußgänger",
        "Überqueren",
        "Straße",
        "Radfahrer",
        "Platzverschwendung",
        "Kreuzung",
        "Grünanlagen",
        "Fahrbahnen",
        "Ampeln"
    ],
    "body": "Die Fahrbahnen nehmen hier einen gewaltigen
Raum ein und dazwischen sind Grünanlagen, die man nicht betreten kann.
Das ist doch Platzverschwendung. Fußgänger müssen sich umständlich über
mehrere nicht aufeinander abgestimmte Ampeln über die Kreuzung quälen.
Kann das nicht so umgestaltet werden, dass mehr Raum für Fußgänger und
Radfahrer entsteht und das Überqueren der Straße leichter wird"
    }
}

```

Result Parameter:

results: list of results per document - JSON objects (id, tags)

title: title of cluster

ids: list of ids of items in the cluster

```

{
  "results": [{
    "title": "Fahrradverkehr Strasse Kreuzung ",
    "ids": ["253", "4335", "223", "2524"]
  },
  {
    "title": "Offentlicher Nahverkehr Ampel",
    "ids": ["253", "252", "25364"]
  },
  {
    "title": "Anbindung Ubahn Innenstadt",
    "ids": ["283", "435", "2233", "22533", "25534"]
  }
]
}

```

Summary

- ...

Usage:

POST request to `/summary`

Parameters:

configuration: contains the configuration for the wrapper API (optional)

length: length of summary max char

language: de, (en, es / not supported yet)

documents:

id: unique id of document

title: title of the document

tags: tags or contributionObject (if not existing fallback to description)

body: text of contributionOriginal Object

Body:

```
{
  "configuration": {
    "language": "de",
    "length": 140
  },
  "documents": [{
    "id": "82734",
    "title": "Eine Erhol-Oase",
    "body": "Eine Erhol-Oase mit einem zentralen Brunnen
und einer steinernen Bank drumrum, auf welcher man verweilen kann.
Außerdem gibt es ein bachartiges Gewässer außenrum und Steinplatten,
über die man die Wasserflächen queren kann. In den Wasserbecken soll es
Fische und Wasserpflanzen geben."
  },
  {
    "id": "82745",
    "title": "Die Ampeln verbessern",
    "body": "Die Fahrbahnen nehmen hier einen gewaltigen
```

Raum ein und dazwischen sind Grünanlagen, die man nicht betreten kann. Das ist doch Platzverschwendung. Fußgänger müssen sich umständlich über mehrere nicht aufeinander abgestimmte Ampeln über die Kreuzung quälen. Kann das nicht so umgestaltet werden, dass mehr Raum für Fußgänger und Radfahrer entsteht und das Überqueren der Straße leichter wird"

```
}  
]  
}
```

Result Parameter:

results: list of results per document - JSON objects (id, tags)

id: unique id, referring to document

summary: string of summary

```
{  
  "results": [{  
    "id": "82734",  
    "summary": "Eine Erhol-Oase mit einem zentralen  
Brunnen und einer steinernen Bank drumrum"  
  },  
  {  
    "id": "82745",  
    "summary": "Die Fahrbahnen nehmen hier einen  
gewaltigen Raum ein und dazwischen sind Grünanlagen, die man nicht  
betreten kann. Das ist doch Platzverschwendung."  
  }  
]  
}
```

Topicmaps

Licenses

[MIT](#)

User Stories

Generelle Anmerkung: In der GUI sollte ersichtlich werden, dass es sich um Beta Funktionalitäten handelt. Darüber hinaus sollten wir die Möglichkeit vorsehen, kurze

Erklärungstexte für die jeweiligen Funktionen vorzusehen, um den Projektadmins die Funktionen zu erläutern. Bitte im Entwurf erst mal Lorem Ipsum. Auf Grundlage der Entwürfe wird die Stadtwerkstatt Texte liefern.

Listenansicht

Als Projektadmin möchte ich mir ein Listenansicht einblenden können, in der ich mir gefiltert nach ...

1. Response Score (Bewertungen quantitativ!)
2. Content Score
3. **Relevance Score**
 - a. Rechenergebnis aus
 - i. Content Score 25 %
 - ii. Response Score 50 %
 - iii. Mutuality Score 25%

... alle Beiträge mit ihren entsprechenden Zuordnungen zum jeweiligen Score in Listenform anzeigen lassen kann. Die Beiträge sollen verlinkt sein.

Wordcloud

Als Projektadmin möchte ich mir für jede Beitragskategorie eine Wordcloud generieren lassen auf Basis des Beitragstextes und der jeweiligen Beitragskommentare.



Cluster

Als Projektadmin möchte ich mir für jede Beitragskategorie die jeweiligen Cluster anzeigen lassen (max. 5). Darüber hinaus möchte ich die Möglichkeit haben, die zu den jeweiligen Clustern zugehörigen Beiträge sehen und anklicken zu können. Die Reihenfolge der Beiträge erfolgt anhand des Relevance Scores.

['Radweg', 'Verbindung', 'Radfahrer']
['Verkehr', 'Bahnhof', 'Innenstadt']
['Oberbillwerder', 'Verkehr', 'Anbindung']
['S2', 'S21', 'Straße']

Heatmap

Als Projektadmin möchte ich mir pro Beitragskategorie die Verteilung der Cluster auf einer Karte anzeigen lassen können (Heatmap). Die Zuordnung soll über die Farbigkeit der Flächen auf der Karte passieren (eine Farbe pro Cluster).



Further Development

API DOC Examples

<https://openrouteservice.org/dev/#/api-docs>

<https://developers.shopware.com/developers-guide/rest-api/>

Studenten Projekt

- <https://docs.google.com/document/d/1DUsg-KuJu7n1pUk1YDXkdovt8Ozu348NqL-j2xZAzo4/edit>

Git Repositories

- https://github.com/dremerb/hamburg_beteiligungsverfahren_helper
- https://github.com/NFiedler/KWKM_Project_1b
- <https://git.informatik.uni-hamburg.de/6roeber/kwkm>

Freitag Fragen

- <https://docs.google.com/document/d/1JkoovamwxWSywlqaf3hOL5CcJO--AfEpdPh6YiXbHU/edit>
- <https://docs.google.com/document/d/1Rb6fyme19kG5cxFXa9bYftPa4V9I6exNFZeGK3OFJg8/edit>

Daten API Dipas:

<https://beteiligung.hamburg/jungfernstieg/drupal/de/masterportal/layer/contributions?SERVICE=GeoJSON&REQUEST=GetCapabilities>

Backlog

- Docker Template
- Swagger Interface
- API Documentation
- DIPAS Community Dataset / Open Data Access

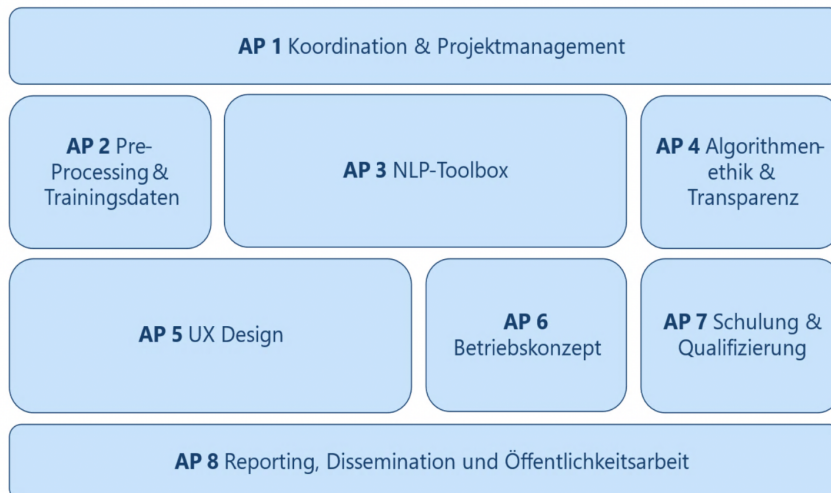
Document History

Version	Update	Date
v0.1	Structure	08.04.2020
v0.2	Keyword, List view update	29.04.2020
v0.3	Parameter description	13.05.2020

v0.4	Summary, Wordcloud description added	09.06.2020
v0.5	IP & Port configuration updated	09.02.2021
v0.6	Clustering Service Description	27.04.2021

DIPAS Weiterentwicklung

- Einleitung / Vorstellungsrunde
 - Claudius Lieven, Michael Fischer, Mateusz Lendzinski, Bianca Lüders, Helen Winter, Antonie Casper, Rico Herzog
- Presentation Claudius Lieven
- Conceptboard
- Aufgabe
 - NLP Core Capabilities
 - User-Interface



- **Geplante Projektlaufzeit: Zwei Jahre**
- **Projektbeteiligte:**
 - **BSW/Stadtwerkstatt (Federführung)**
 - **LGV/G13, G33**
 - **HCU/ CityScienceLab**
 - **?Dataport?**