

ENTWICKLUNG DES BACKENDS

Julien Hofer (Backend-Entwickler)

Landesbetrieb Geoinformation und
Vermessung / Urban Data Analytics



Konferenzzentrum BSW | 15.09.2025



Hamburg

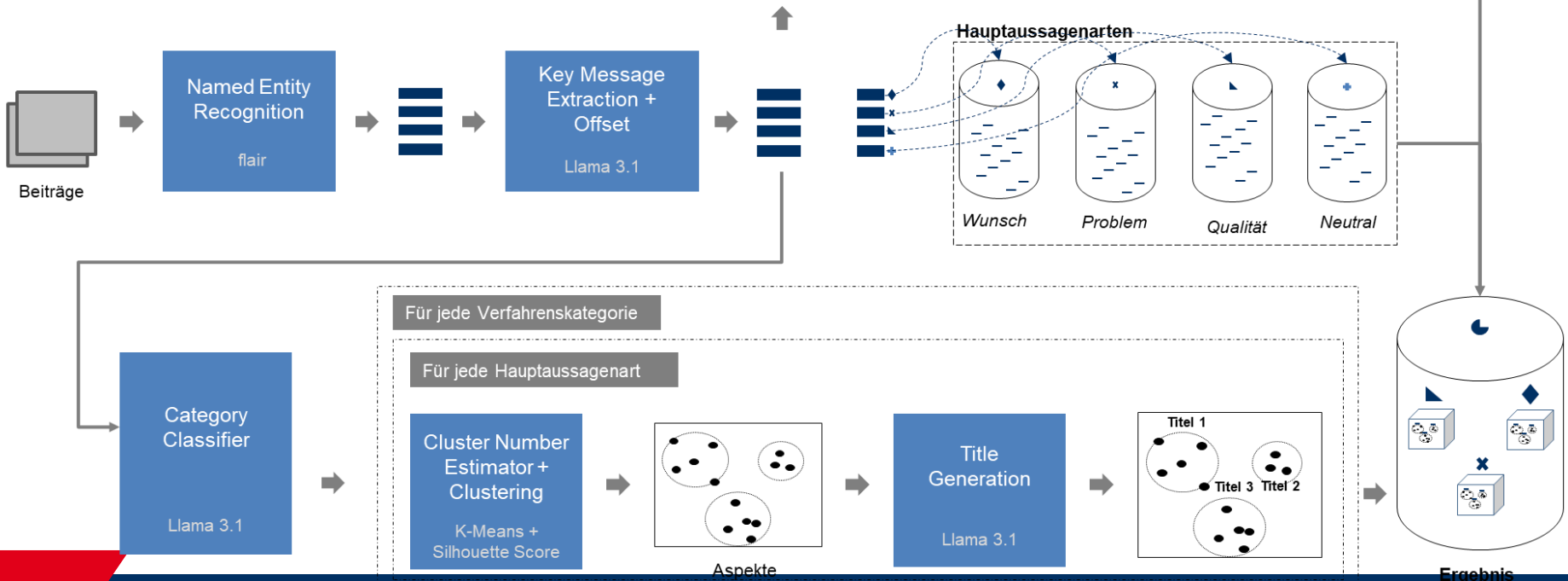
Landesbetrieb
Geoinformation
und Vermessung



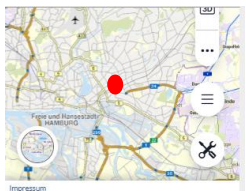
Hamburg

Technischer Workflow

Für jedes Verfahren



Beitragsanalyse: Erforderliche NLP-Services



Beitrag

Extraktion der Hauptaussagen, ihres Typs und ihrer Ursprungssätze

[Key Message Extraction]

Relevanter Satz

Key Message

„Abgesenkte Bürgersteige werden zugeparkt“

Key Message

„Zugeparkte Kurven versperren die Sicht“

Key Message

„Schnell kann ein Unfall passieren“

„Es ist nur eine Frage der Zeit, bis auch an dieser Ecke etwas passiert.“

Übersicht der NLP-Services

- Generatives LLM erforderlich: Einsatz von **Llama 3.1**

Beitrags-
analyse

**Key
Message
Extraction**

Statement
Labeling

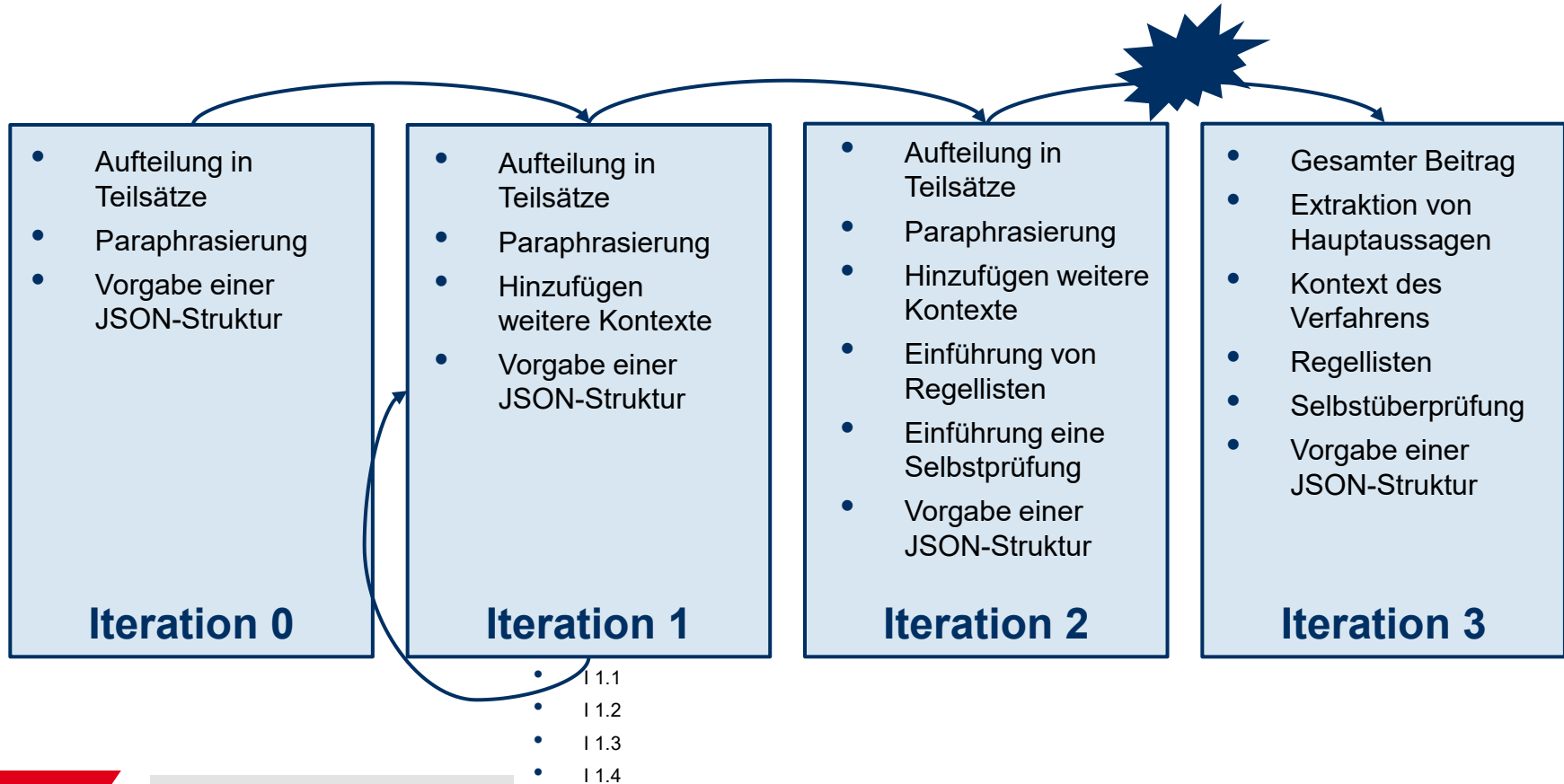
Category
Assignment

Beitrags-
vernetzung

Clustering

Title
Generation

Prompting-Historie: Methodik



Prompting-Historie: Ergebnis

Ich gebe dir jeweils einen Beitrag. Extrahiere die Hauptaussagen aus diesem Beitrag. Der Beitrag stammt aus einem Beteiligungsverfahren. Dieses Beteiligungsverfahren wird wie folgt beschrieben: {proceeding_description}

Beachte dabei folgende Regeln:

1. Verarbeite den gesamten Beitrag von Anfang bis Ende, ohne Auslassungen.
2. Eine Hauptaussage besteht mindestens aus einem Nomen und einem Verb.
3. Eine Hauptaussage besteht in der Regel aus vier bis sieben Wörtern.
4. Konkrete und spezifische Ortsbezüge sind für Hauptaussagen nicht relevant und sollen von dir nicht beachtet werden.
5. Es gibt nur 4 Arten von Hauptaussagen (Hauptaussagenart):
 - a) Wunsch: Beschreibt eine geforderte Veränderung für die Zukunft. Leite Wünsche nicht aus reinen Beschreibungen ab.
 - b) Problem: Beschreibt einen gegenwärtigen negativen Zustand.
 - c) Qualität: Beschreibt einen gegenwärtigen positiven Zustand.
 - d) Zustand: Beschreibt neutral einen gegenwärtigen Zustand.
6. Sätze, die einen negativen Zustand beschreiben, gehören in die Kategorie "Problem".
7. Nur wenn ein Wunsch explizit erkennbar ist, darf der Satz als "Wunsch" formuliert werden.
8. Bei der Formulierung der Hauptaussagen darf eine Negation nicht einfach ins Gegenteil umgewandelt werden, es sei denn, dies wird durch die Satzbedeutung eindeutig unterstützt.
9. Ordne die Sätze aus dem Beitrag deinen Hauptaussagen zu.
10. Füge für jede Hauptaussage eine kurze Begründung hinzu.
11. Formuliere deine Antwort je nach Hauptaussagenart wie folgt:
 - a) Wunsch: Ich wünsche mir in Zukunft, dass <Hauptaussage>.
 - b) Problem: Ein Problem aktuell ist, dass <Hauptaussage>.
 - c) Qualität: Gut finde ich aktuell, dass <Hauptaussage>.
 - d) Zustand: Aktuell ist <Hauptaussage>.

Prüfe: ...

- Gesamter Beitrag
- **Extraktion von Hauptaussagen**
- Kontext des Verfahrens
- Regellisten
- Selbstüberprüfung
- Vorgabe einer JSON-Struktur

Iteration 3

Prompting-Historie: Ergebnis

Ich gebe dir jeweils einen Beitrag. Extrahiere die Hauptaussagen aus diesem Beitrag. Der Beitrag stammt aus einem Beteiligungsverfahren. Dieses Beteiligungsverfahren wird wie folgt beschrieben: {proceeding_description}

Beachte dabei folgende Regeln:

1. Verarbeite den gesamten Beitrag von Anfang bis Ende, ohne Auslassungen.
2. Eine Hauptaussage besteht mindestens aus einem Nomen und einem Verb.
3. Eine Hauptaussage besteht in der Regel aus vier bis sieben Wörtern.
4. Konkrete und spezifische Ortsbezüge sind für Hauptaussagen nicht relevant und sollen von dir nicht beachtet werden.
5. Es gibt nur 4 Arten von Hauptaussagen (Hauptaussagenart):
 - a) Wunsch: Beschreibt eine geforderte Veränderung für die Zukunft. Leite Wünsche nicht aus reinen Beschreibungen ab.
 - b) Problem: Beschreibt einen gegenwärtigen negativen Zustand.
 - c) Qualität: Beschreibt einen gegenwärtigen positiven Zustand.
 - d) Zustand: Beschreibt neutral einen gegenwärtigen Zustand.
6. Sätze, die einen negativen Zustand beschreiben, gehören in die Kategorie "Problem".
7. Nur wenn ein Wunsch explizit erkennbar ist, darf der Satz als "Wunsch" formuliert werden.
8. Bei der Formulierung der Hauptaussagen darf eine Negation nicht einfach ins Gegenteil umgewandelt werden, es sei denn, dies wird durch die Satzbedeutung eindeutig unterstützt.
9. Ordne die Sätze aus dem Beitrag deinen Hauptaussagen zu.
10. Füge für jede Hauptaussage eine kurze Begründung hinzu.
11. Formuliere deine Antwort je nach Hauptaussagenart wie folgt:
 - a) Wunsch: Ich wünsche mir in Zukunft, dass <Hauptaussage>.
 - b) Problem: Ein Problem aktuell ist, dass <Hauptaussage>.
 - c) Qualität: Gut finde ich aktuell, dass <Hauptaussage>.
 - d) Zustand: Aktuell ist <Hauptaussage>.

Prüfe: ...

- Gesamter Beitrag
- Extraktion von Hauptaussagen
- **Kontext des Verfahrens**
- Regellisten
- Selbstüberprüfung
- Vorgabe einer JSON-Struktur

Iteration 3

Prompting-Historie: Ergebnis

Ich gebe dir jeweils einen Beitrag. Extrahiere die Hauptaussagen aus diesem Beitrag. Der Beitrag stammt aus einem Beteiligungsverfahren. Dieses Beteiligungsverfahren wird wie folgt beschrieben: {proceeding_description}

Beachte dabei folgende Regeln:

1. Verarbeite den gesamten Beitrag von Anfang bis Ende, ohne Auslassungen.
2. Eine Hauptaussage besteht mindestens aus einem Nomen und einem Verb.
3. Eine Hauptaussage besteht in der Regel aus vier bis sieben Wörtern.
4. Konkrete und spezifische Ortsbezüge sind für Hauptaussagen nicht relevant und sollen von dir nicht beachtet werden.
5. Es gibt nur 4 Arten von Hauptaussagen (Hauptaussagenart):
 - a) Wunsch: Beschreibt eine geforderte Veränderung für die Zukunft. Leite Wünsche nicht aus reinen Beschreibungen ab.
 - b) Problem: Beschreibt einen gegenwärtigen negativen Zustand.
 - c) Qualität: Beschreibt einen gegenwärtigen positiven Zustand.
 - d) Zustand: Beschreibt neutral einen gegenwärtigen Zustand.
6. Sätze, die einen negativen Zustand beschreiben, gehören in die Kategorie "Problem".
7. Nur wenn ein Wunsch explizit erkennbar ist, darf der Satz als "Wunsch" formuliert werden.
8. Bei der Formulierung der Hauptaussagen darf eine Negation nicht einfach ins Gegenteil umgewandelt werden, es sei denn, dies wird durch die Satzbedeutung eindeutig unterstützt.
9. Ordne die Sätze aus dem Beitrag deinen Hauptaussagen zu.
10. Füge für jede Hauptaussage eine kurze Begründung hinzu.
11. Formuliere deine Antwort je nach Hauptaussagenart wie folgt:
 - a) Wunsch: Ich wünsche mir in Zukunft, dass <Hauptaussage>.
 - b) Problem: Ein Problem aktuell ist, dass <Hauptaussage>.
 - c) Qualität: Gut finde ich aktuell, dass <Hauptaussage>.
 - d) Zustand: Aktuell ist <Hauptaussage>.

- Gesamter Beitrag
- Extraktion von Hauptaussagen
- Kontext des Verfahrens
- **Regellisten**
- Selbstüberprüfung
- Vorgabe einer JSON-Struktur

Iteration 3

Prüfe: ...

Prompting-Historie: Ergebnis

Beachte dabei folgende Regeln [...]

Prüfe:

1. Dass die Hauptaussagen strukturell sinnvoll und klar formuliert sind.
2. Die Hauptaussagen nicht mehr als sieben Wörter beinhalten.
3. Die Hauptaussagen keine konkreten und spezifischen Ortsbezüge enthalten.
4. Eine Hauptaussage darf die Bedeutung des Originalsatzes nicht umkehren oder verändern. Negative Zustände (z. B. "fehlen") müssen als Problem formuliert werden und dürfen nicht als Wunsch oder positive Aussage dargestellt werden.
5. Die Hauptaussagen die wichtigsten Punkte des Beitrages widerspiegeln und die Relevanz für die Stadtplanung verdeutlichen.
6. Ein Satz, der einen aktuellen negativen Zustand beschreibt, darf nicht zu einem Wunsch umformuliert werden, sofern dies im Originalsatz nicht ausdrücklich als Wunsch erkennbar ist.
7. Hauptaussagen die eine Konjunktion (Bindewort) enthalten in mehrere Hauptaussagen aufgeteilt werden.
8. Die Sätze (Satz) aus dem Beitrag in ihrem Original wiedergegeben werden.
9. Die Hauptaussagen je nach Hauptaussagenart korrekt formuliert sind.
10. Dass du für die Sätze (Satz) keine Umformulierungen oder Ergänzungen und keine Rechtschreib- oder Grammatikkorrektur gemacht hast.
11. Dass du für die Sätze (Satz) keine Korrekturen von Steuerzeichen oder Formatierungen gemacht hast.
12. Wenn eine dieser Regeln nicht eingehalten wird. Beginne von vorne.
Der Beitrag lautet: '{contribution_text}'

Strukturiere...

- **Gesamter Beitrag**
- Extraktion von Hauptaussagen
- Kontext des Verfahrens
- Regellisten
- **Selbstüberprüfung**
- Vorgabe einer JSON-Struktur

Iteration 3

Prompting-Historie: Ergebnis

Prüfe: [...]

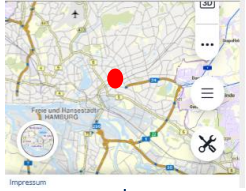
Strukturiere die Antwort wie in dem folgenden durch drei Backticks abgegrenztem JSON-Format (RFC 8259). Gib nicht die drei Backticks mit zurück.:

```
```  
 {{
 "key_messages":
 [
 {{
 "key_message_phrase": "<Antwort>",
 "key_message": "<Hauptaussage>",
 "key_message_type": "<Hauptaussagenart>",
 "related_sentences":
 [
 "<Satz>",
 ...
],
 "explanation": "<Begründung>"
 }}
]
 }}
  ```
```

- Gesamter Beitrag
- Extraktion von Hauptaussagen
- Kontext des Verfahrens
- Regellisten
- Selbstüberprüfung
- **Vorgabe einer JSON-Struktur**

Iteration 3

Beitragsanalyse: Erforderliche NLP-Services



Beitrag

Relevanter Satz

„Durch Gegenverkehr wird umständlich rangiert, zugeparkte Kurven versperren die Sicht und abgesenkte Bürgersteige werden zugeparkt.“

Relevanter Satz

„Es ist nur eine Frage der Zeit, bis auch an dieser Ecke etwas passiert.“

Labeling der Ursprungssätze
[Statement Labeling]

Key Message

„Abgesenkte Bürgersteige werden zugeparkt“

Key Message

„Zugeparkte Kurven versperren die Sicht“

Key Message

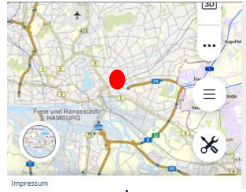
„Schnell kann ein Unfall passieren“

Übersicht der NLP-Services



- Generatives LLM zur Extraktion der Hauptaussagen erforderlich: Einsatz von **Llama 3.1**
- Modell für das Statement Labeling: RoBERTa (Robustly Optimised BERT Approach), WECHSEL-Variante (übertragen auf die deutsche Sprache), 125 M Parameter, trainiert auf ~ 1000 manuell gelabelten Sätzen

Beitragsanalyse: Erforderliche NLP-Services



Beitrag

Relevanter Satz

„Durch Gegenverkehr wird umständlich rangiert, zugeparkte Kurven versperren die Sicht und abgesenkte Bürgersteige werden zugeparkt.“

Relevanter Satz

„Es ist nur eine Frage der Zeit, bis auch an dieser Ecke etwas passiert.“

Key Message

„Abgesenkte Bürgersteige werden zugeparkt“

Key Message

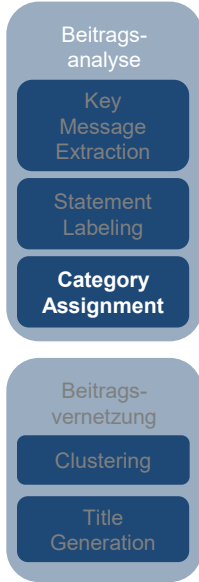
„Zugeparkte Kurven versperren die Sicht“

Key Message

„Schnell kann ein Unfall passieren“

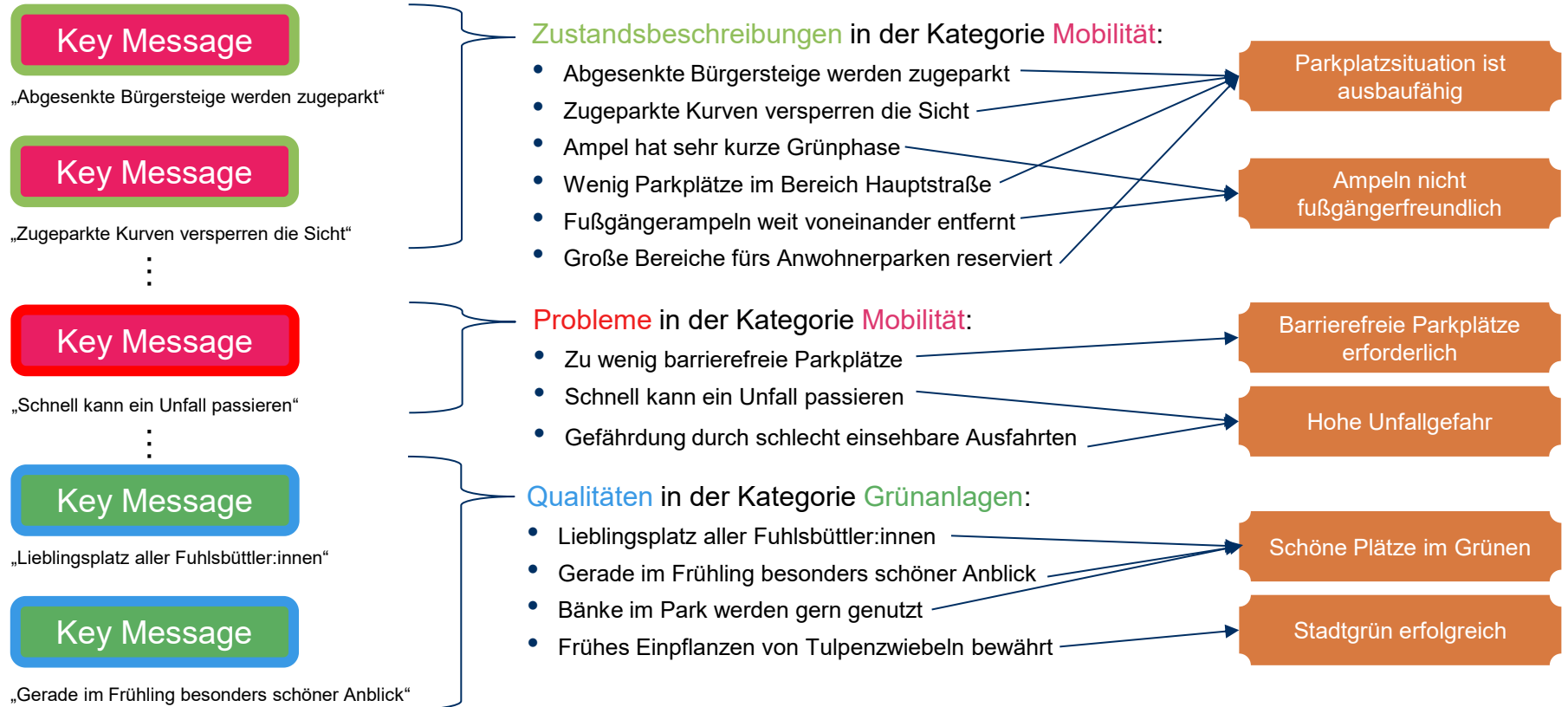
Zuweisung einer Kategorie zu den Key Messages
[Category Assignment]

Übersicht der NLP-Services

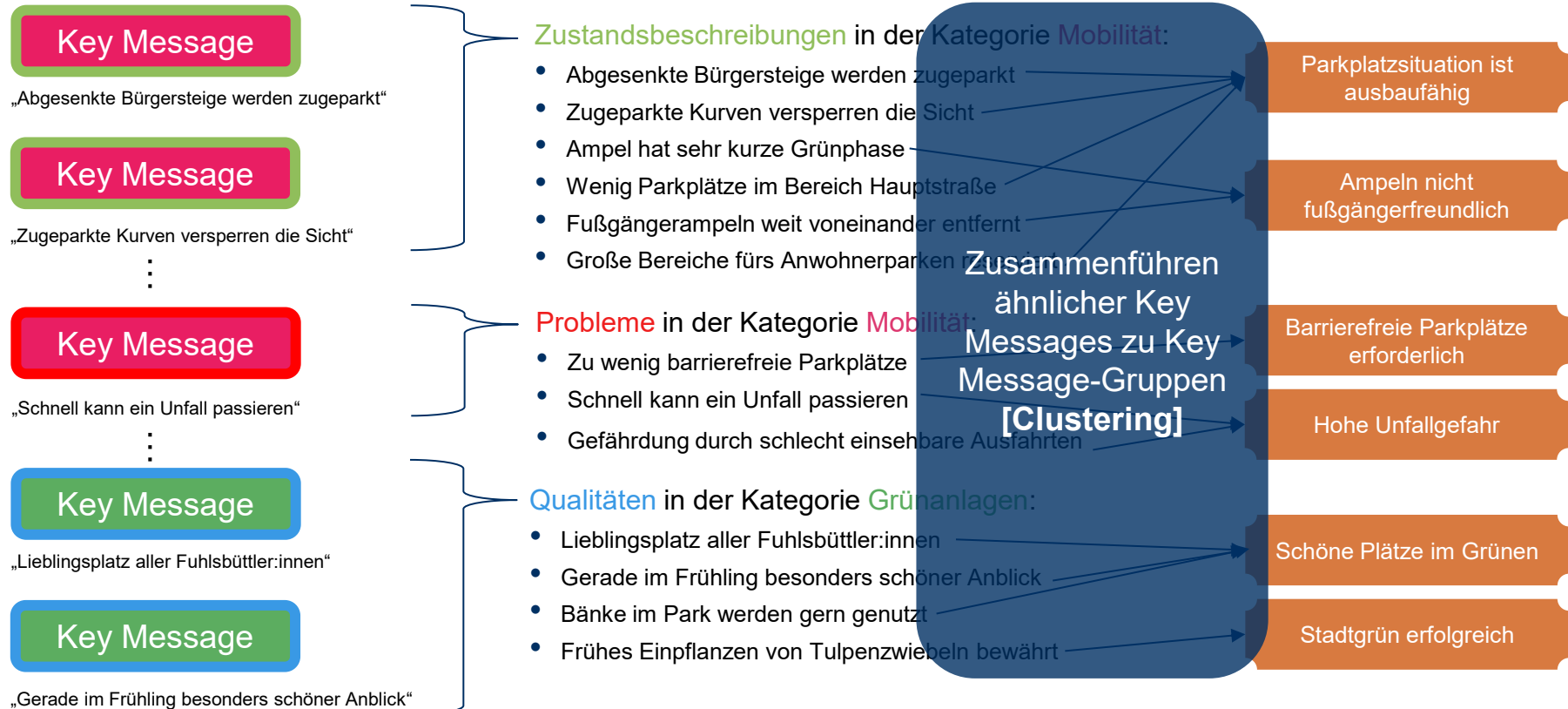


- Generatives LLM zur Extraktion der Hauptaussagen erforderlich: Einsatz von **Llama 3.1**
- Modell für das Statement Labeling: RoBERTa (Robustly Optimised BERT Approach), WECHSEL-Variante (übertragen auf die deutsche Sprache), 125 M Parameter, trainiert auf ~ 1000 manuell gelabelten Sätzen
- Keine Trainingsdaten für das Category Assignment – Wechsel von fine-tuning hin zu einem *instruct-based* Model: Einsatz von **Llama 3.1**

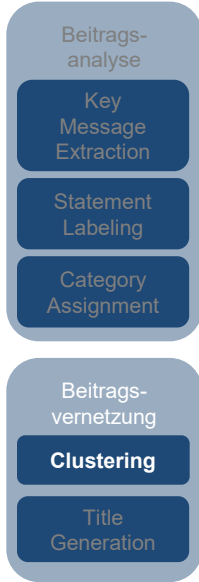
Beitragsvernetzung: Verbindung der Beiträge über ihre Key Messages zu *Aspekten*



Beitragsvernetzung: Erforderliche NLP-Services

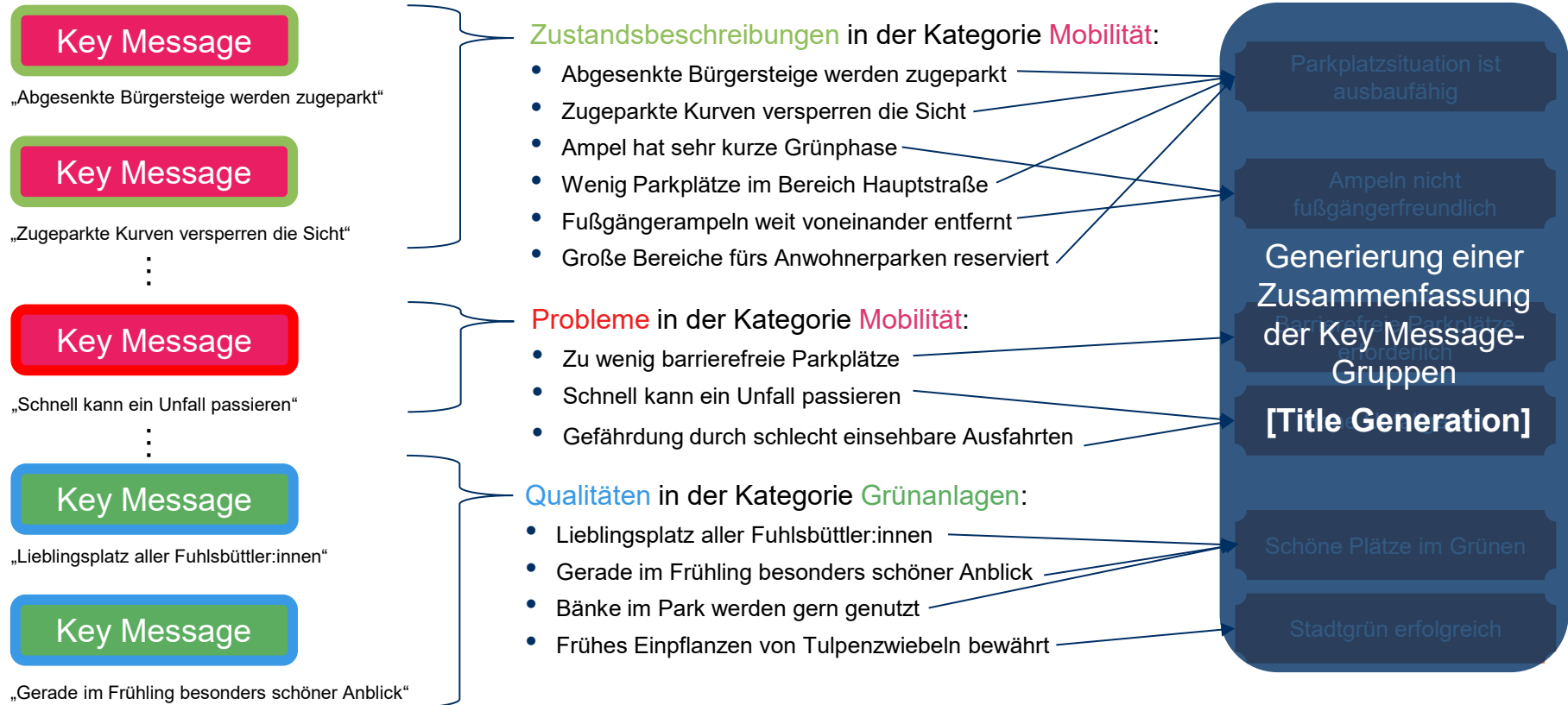


Übersicht der NLP-Services

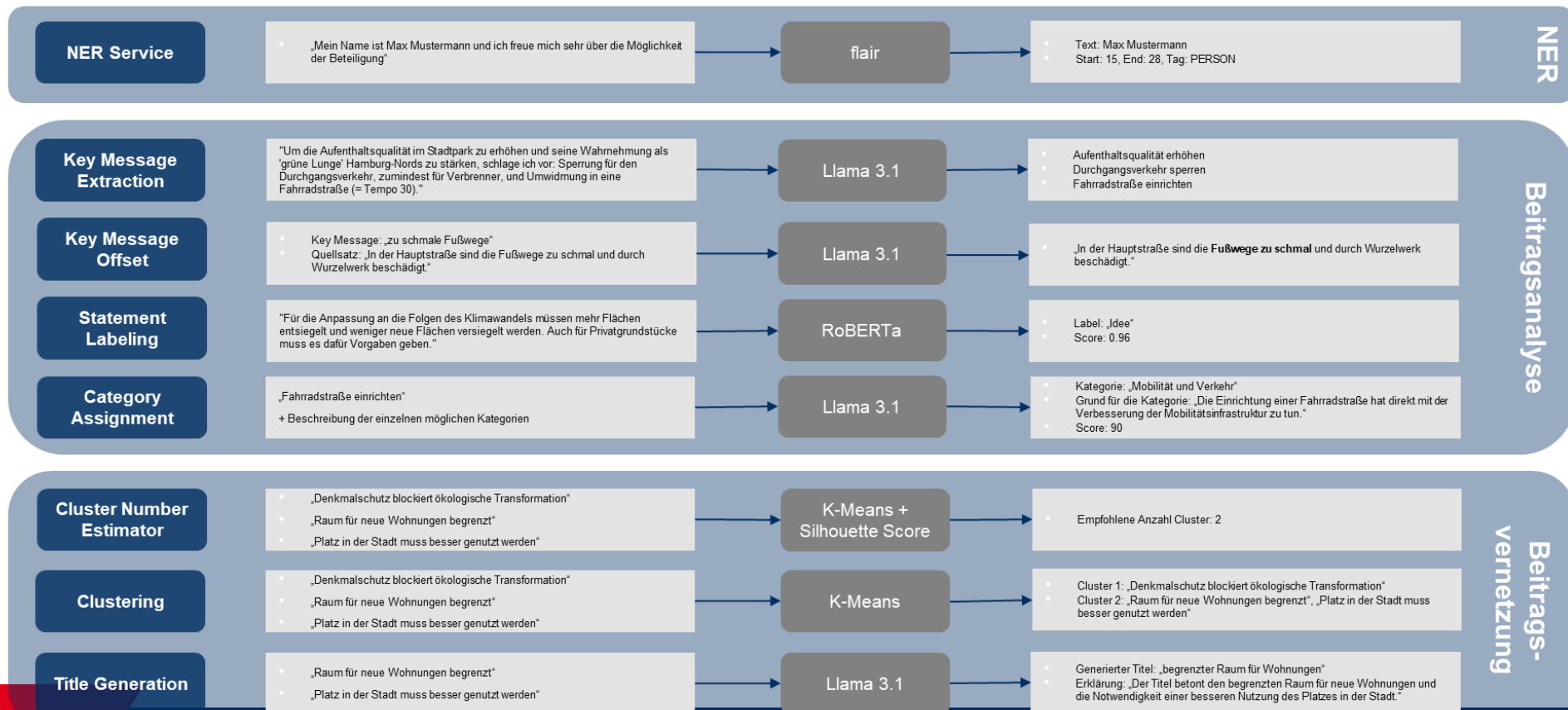


- Generatives LLM zur Extraktion der Hauptaussagen erforderlich: Einsatz von **Llama 3.1**
- Modell für das Statement Labeling: RoBERTa (Robustly Optimised BERT Approach), WECHSEL-Variante (übertragen auf die deutsche Sprache), 125 M Parameter, trainiert auf ~ 1000 manuell gelabelten Sätzen
- Keine Trainingsdaten für das Category Assignment – Wechsel von fine-tuning hin zu einem *instruct-based* Model: Einsatz von **Llama 3.1**
- Im Clustering transparente Herangehensweise mit Embedding und k-Means-Clustering

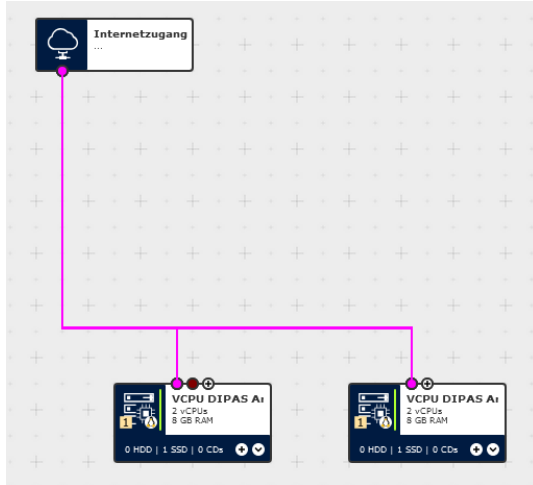
Beitragsvernetzung: Erforderliche NLP-Services



Die 8 NLP-Services und Prozessoren im Überblick



Betrieb bei IONOS



- Backend besteht aus zwei Teilen:
 - **Beitragsanalyse:** Erweiterung der einzelnen Beiträge eines Verfahrens um z.B. Hauptaussagen, deren thematische Kategorie, und die Teile des Beitrags, auf denen sie beruhen
 - **Beitragsvernetzung:** Verbindung der Beiträge eines Verfahrens untereinander anhand ihrer Hauptaussagen
- IONOS über Rahmenvertrag beauftragt, da dieser den **datenschutzkonformen AI Model Hub** anbietet
- Hierfür zwei VPS bei Ionos gebucht:
 - **NLP Services:** bietet alle erforderlichen modularen NLP-Services
 - **Orchestrator:** verbindet in zwei Prozessoren die NLP-Services zu Beitragsanalyse und -vernetzung