

Suchergebnisdiversifizierung durch Anfragesegmentierung

Aufgaben einer Suchmaschine

Aufgaben einer Suchmaschine

- Anfragen entgegennehmen

Aufgaben einer Suchmaschine

- Anfragen entgegennehmen
- Anfragen verarbeiten

Aufgaben einer Suchmaschine

- Anfragen entgegennehmen
- Anfragen verarbeiten
- Ergebnisliste liefern

Ambige Anfragen

Ambige Anfragen

jaguar

Ambige Anfragen

jaguar

queen

Ambige Anfragen

jaguar source of the nile
queen

Ambige Anfragen

jaguar source of the nile

chinese democracy queen

Ambige Anfragen

new york times square dance
jaguar source of the nile
chinese democracy queen

Ambige Anfragen

new york times square dance
jaguar source of the nile
chinese democracy queen
family sports center weimar

Ambige Anfragen

new york times square dance
jaguar source of the nile
scientific present ideas
chinese democracy queen
family sports center weimar

Diversifizierung von Suchergebnissen

Diversifizierung von Suchergebnissen

- Anfragen entgegennehmen

Diversifizierung von Suchergebnissen

- Anfragen entgegennehmen
- Intentionen ermitteln

Diversifizierung von Suchergebnissen

- Anfragen entgegennehmen
- Intentionen ermitteln
- Anfragen entsprechend Intentionen verarbeiten

Diversifizierung von Suchergebnissen

- Anfragen entgegennehmen
- Intentionen ermitteln
- Anfragen entsprechend Intentionen verarbeiten
- Gemischte Ergebnisliste liefern

Ambige Anfragen

new york times square dance
jaguar source of the nile
scientific present ideas
chinese democracy queen
family sports center weimar

Ambige Anfragen

new york times square dance

jaguar source of the nile

scientific present ideas

chinese democracy queen

family sports center weimar

Ambige Anfragen

new york times square dance

jaguar source of the nile

scientific present ideas

chinese democracy queen

family sports center weimar

Anfragesegmentierung

new york times square dance

Anfragesegmentierung

"new york times" "square dance"

Anfragesegmentierung

"new york" "times square" "dance"

Idee

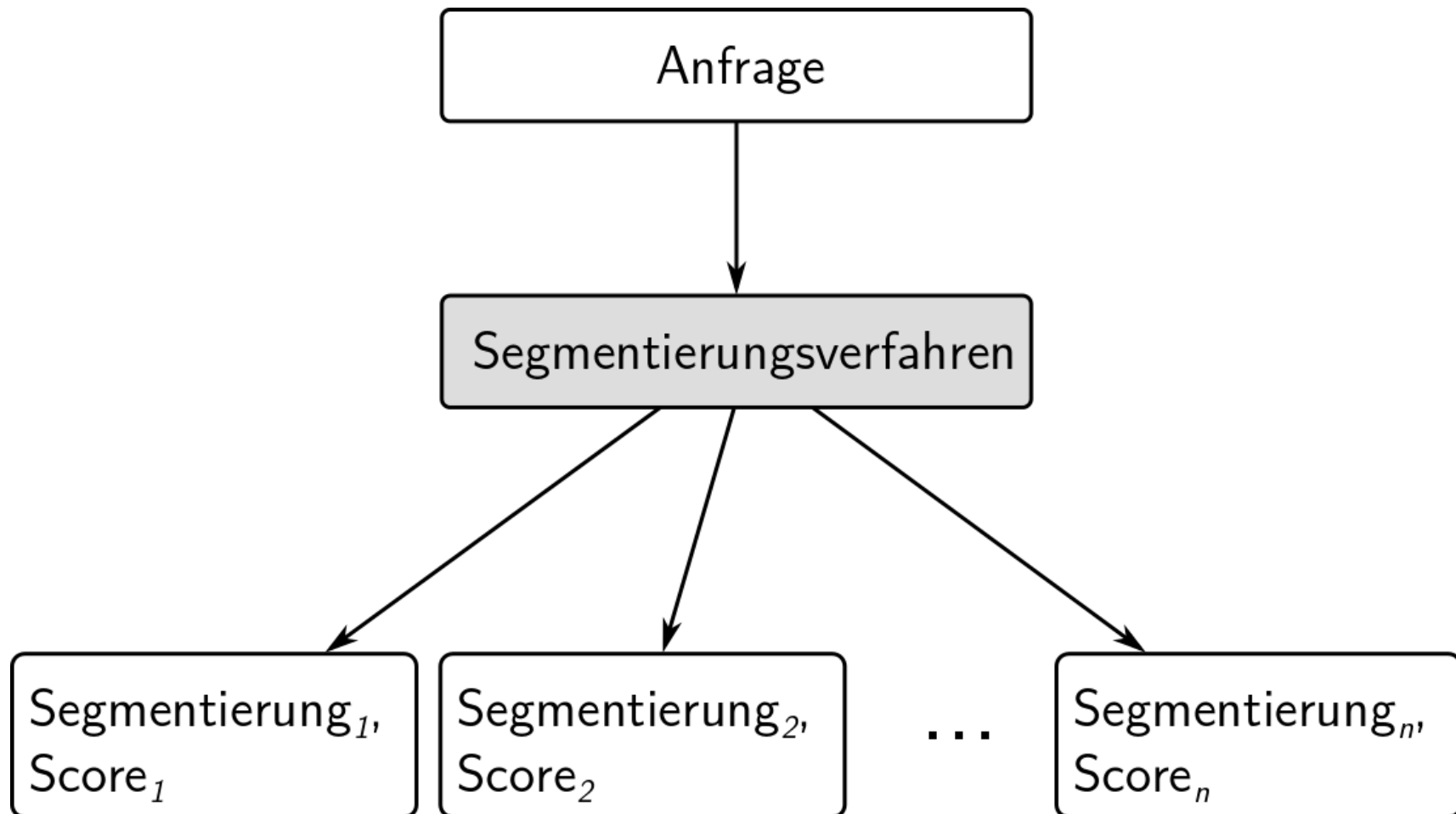
Idee

- Lange Anfragen automatisch segmentieren

Idee

- Lange Anfragen automatisch segmentieren
- Ambiguität auf Grundlage der Segmentierungen ermitteln

Automatische Segmentierung



Entwickeltes Diversifizierungsverfahren

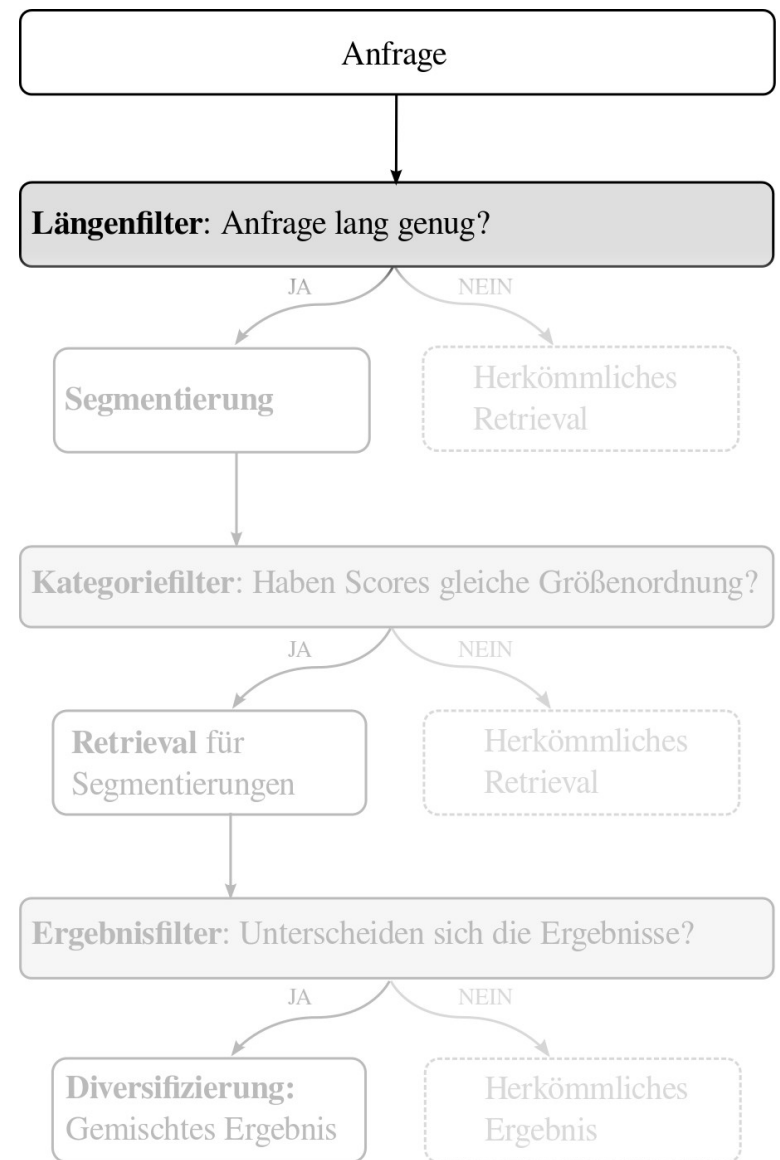
Entwickeltes Diversifizierungsverfahren

- Zwei Hauptaufgaben:
 1. Lange ambige Anfragen erkennen
 2. Suchergebnisse diversifizieren

Entwickeltes Diversifizierungsverfahren

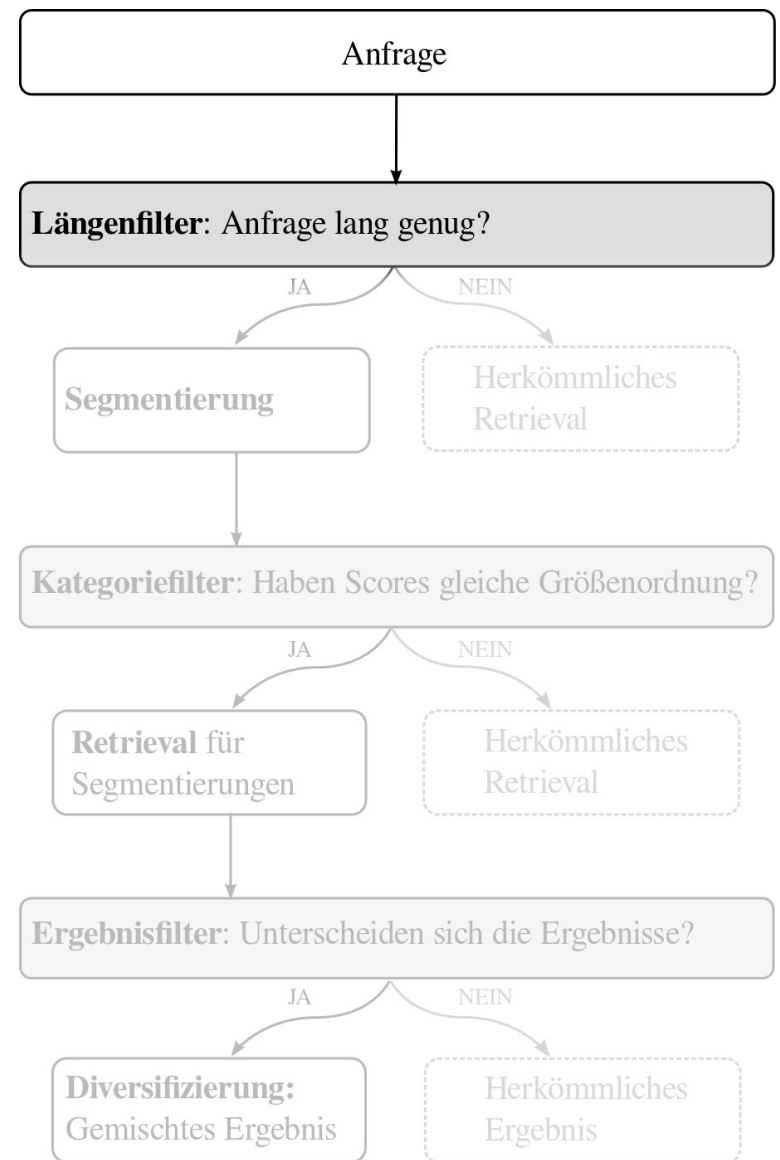
- Zwei Hauptaufgaben:
 1. Lange ambige Anfragen erkennen
 2. Suchergebnisse diversifizieren
- Verschiedene Filterschritte

Längenfilter



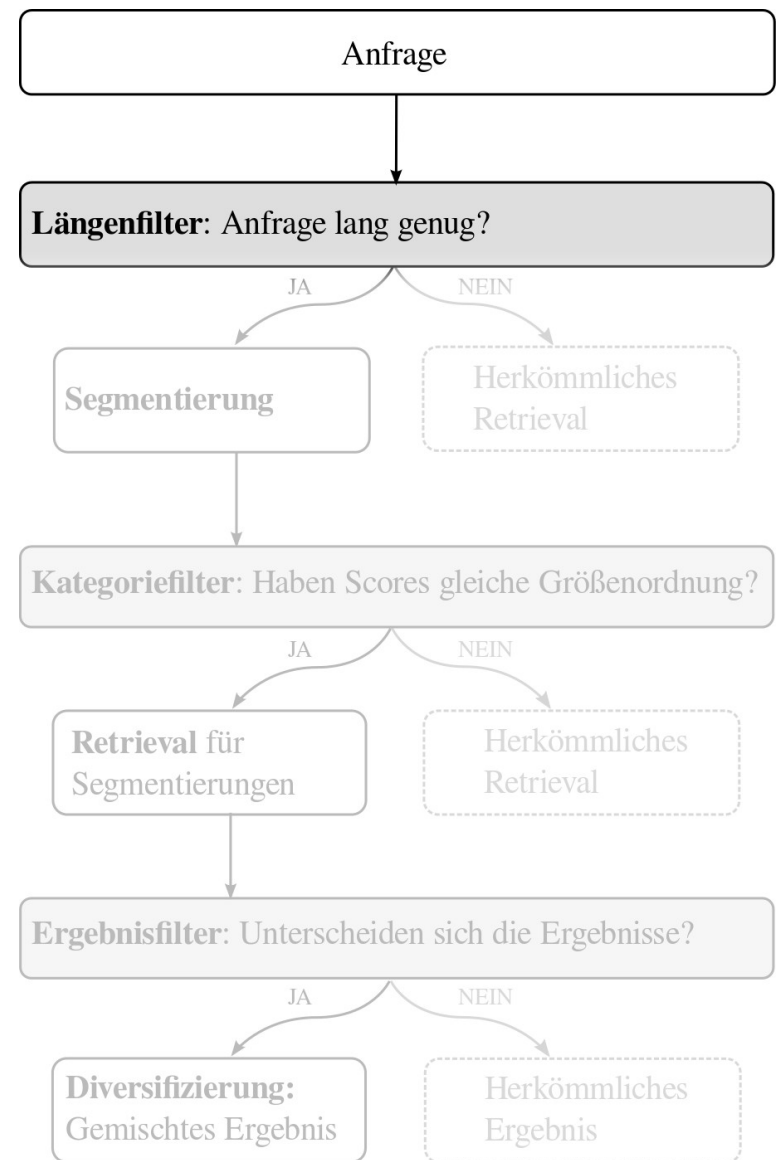
Längenfilter

- Kurze Anfragen ausgeschlossen

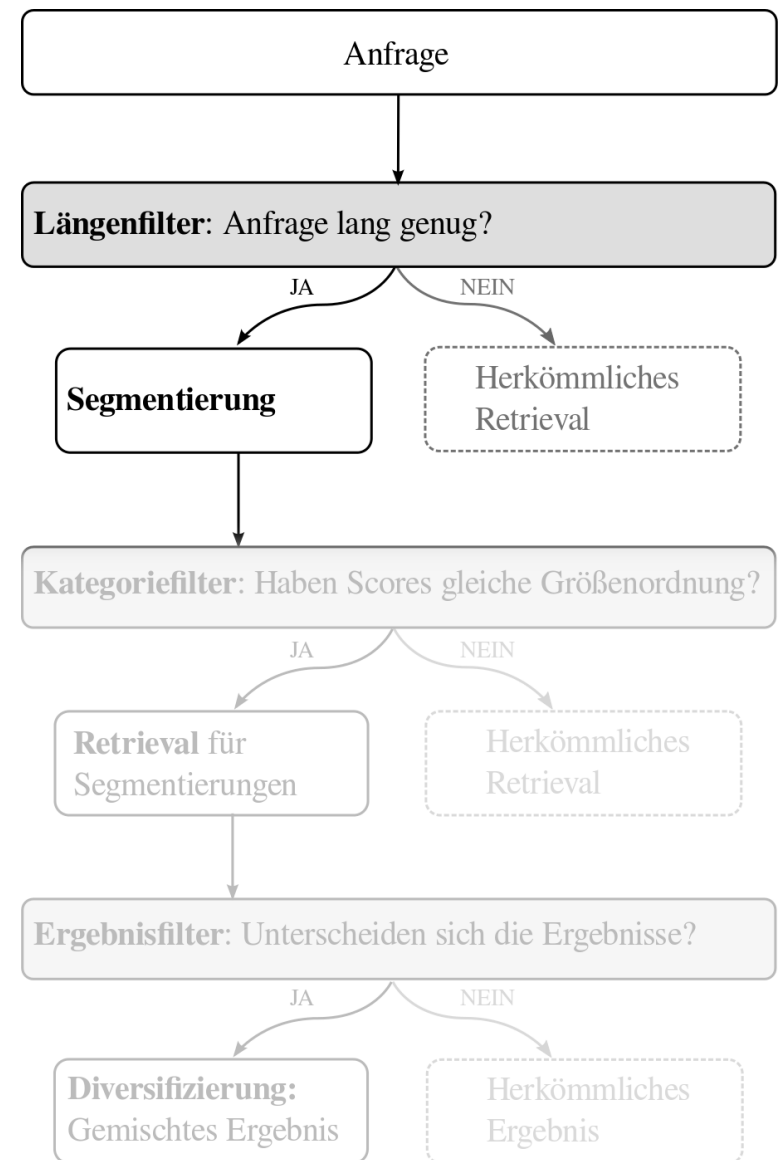


Längenfilter

- Kurze Anfragen ausgeschlossen
- Lange Anfragen in den nächsten Schritt

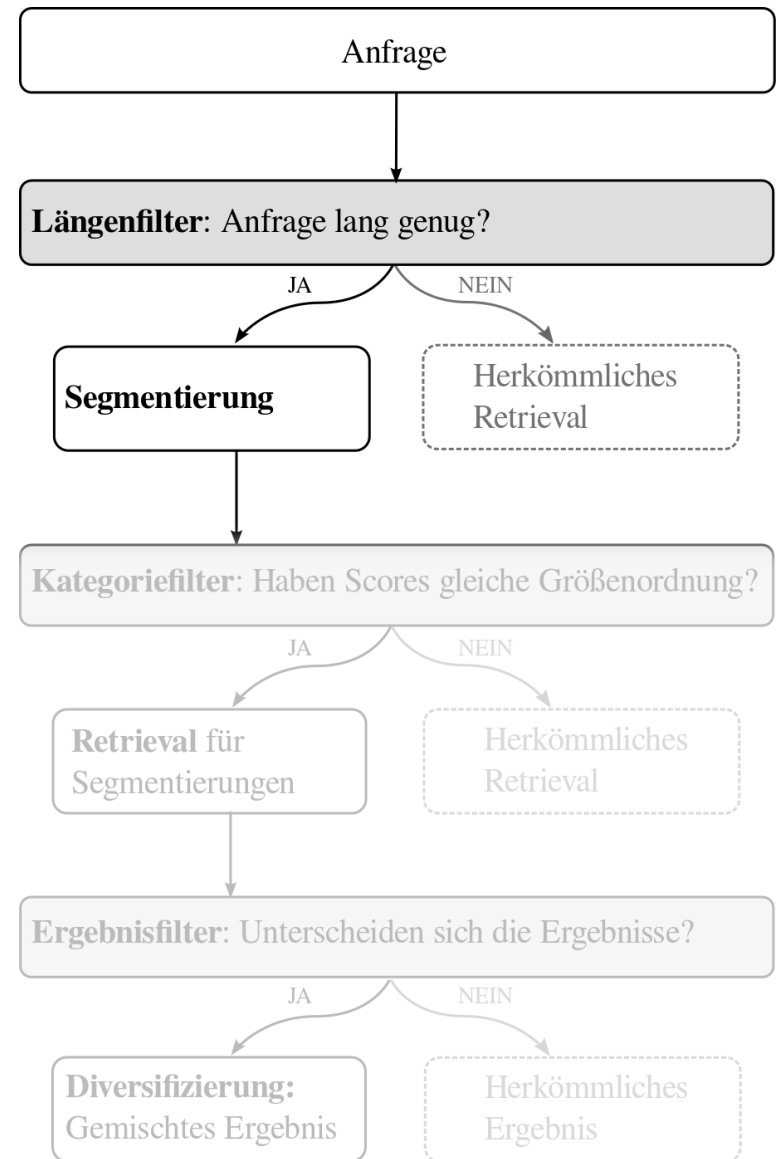


Segmentierungsschritt



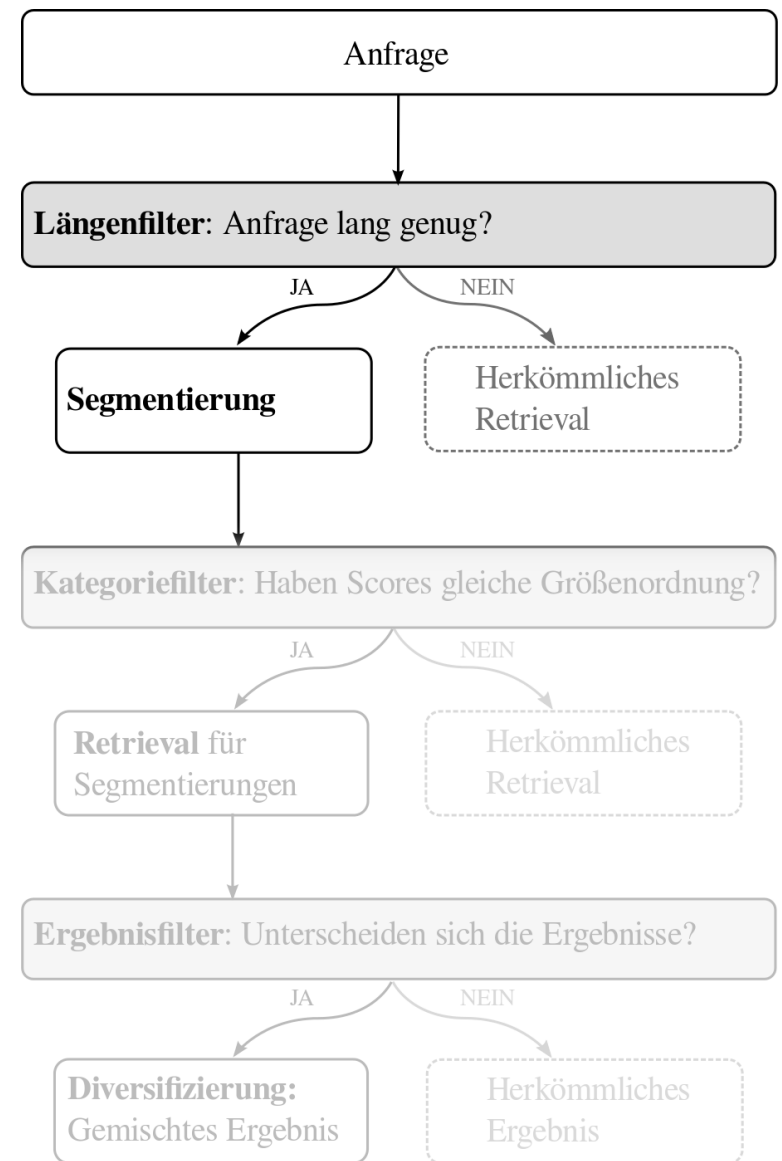
Segmentierungsschritt

- Automatisches Segmentierungsverfahren

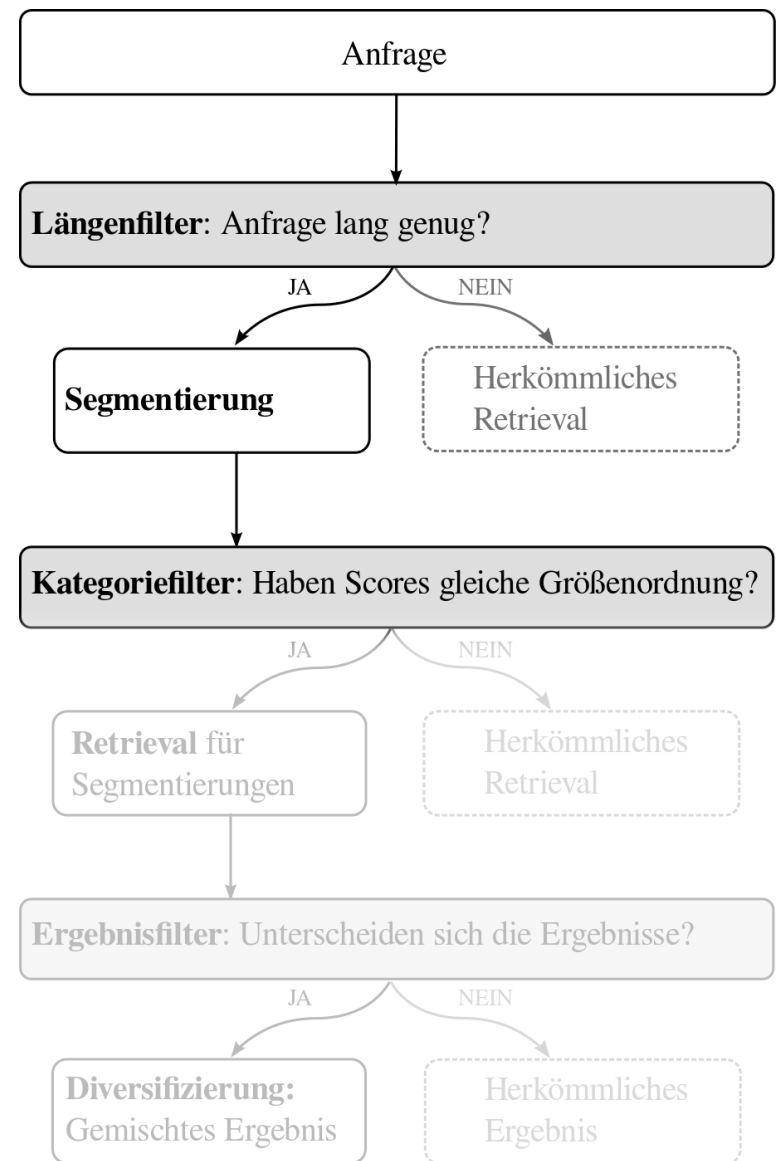


Segmentierungsschritt

- Automatisches Segmentierungsverfahren
- Ermitteln der beiden „besten“ Segmentierungen

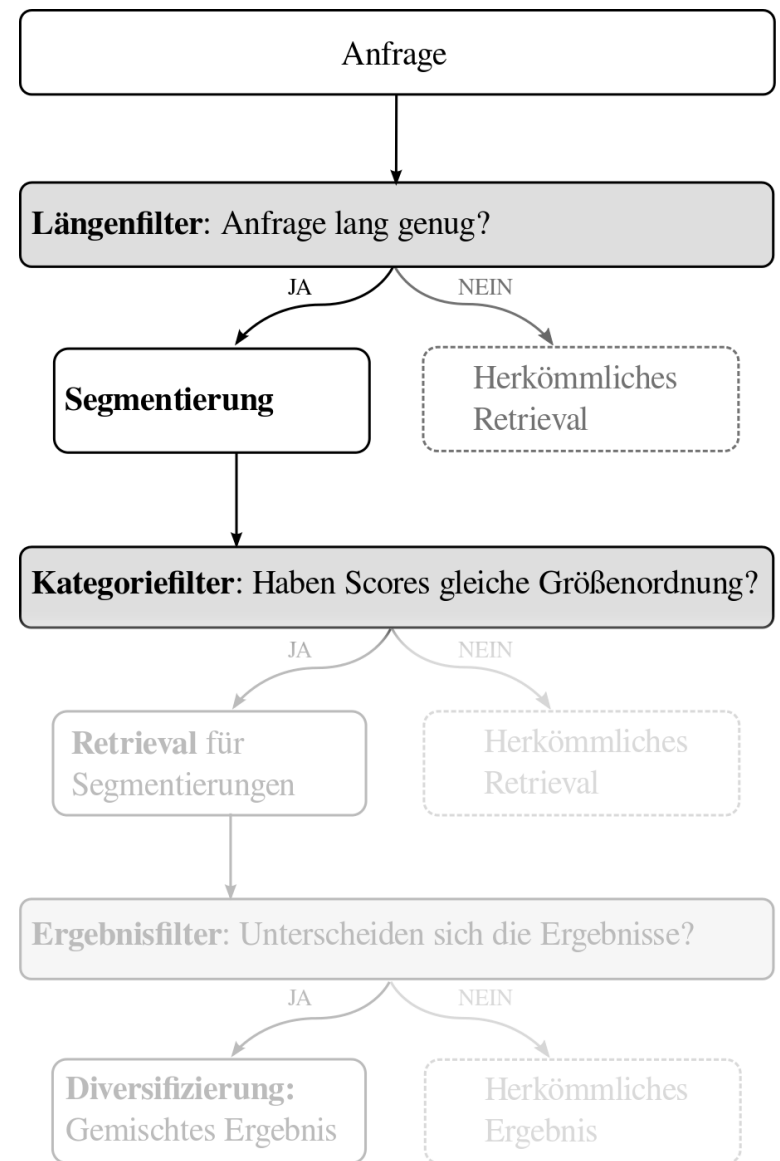


Kategoriefilter



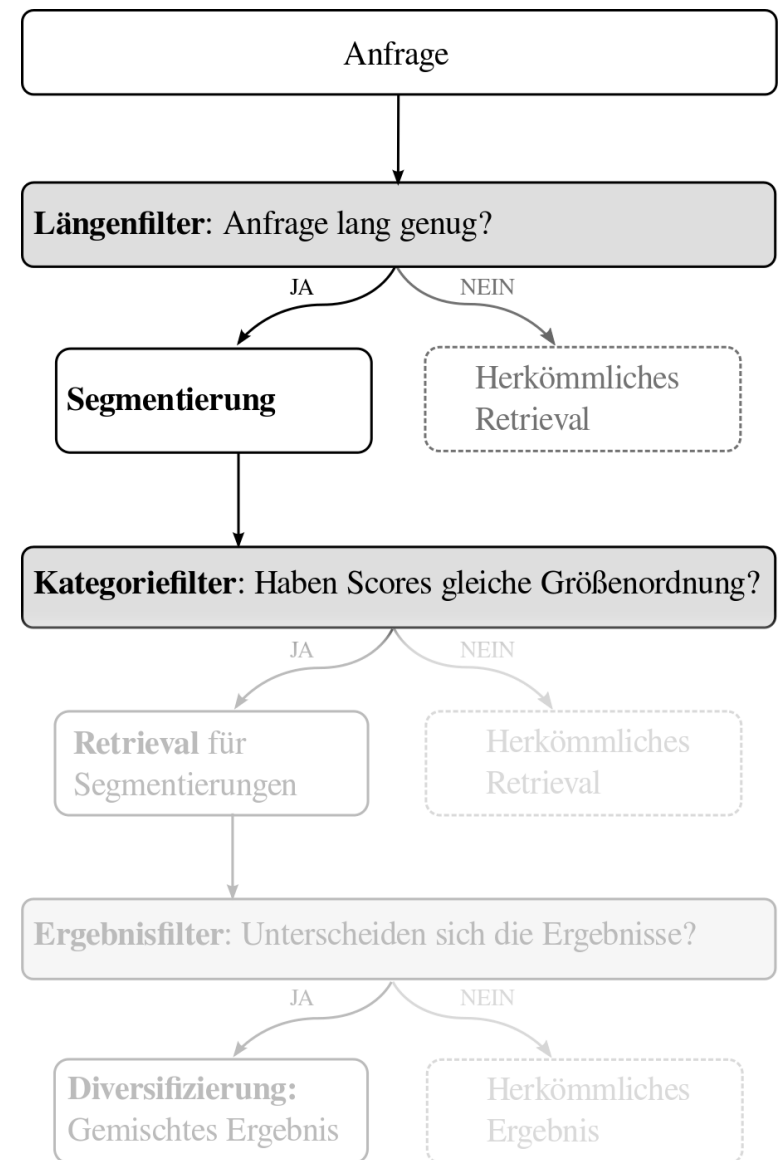
Kategoriefilter

- Kategorisierung der Anfragen

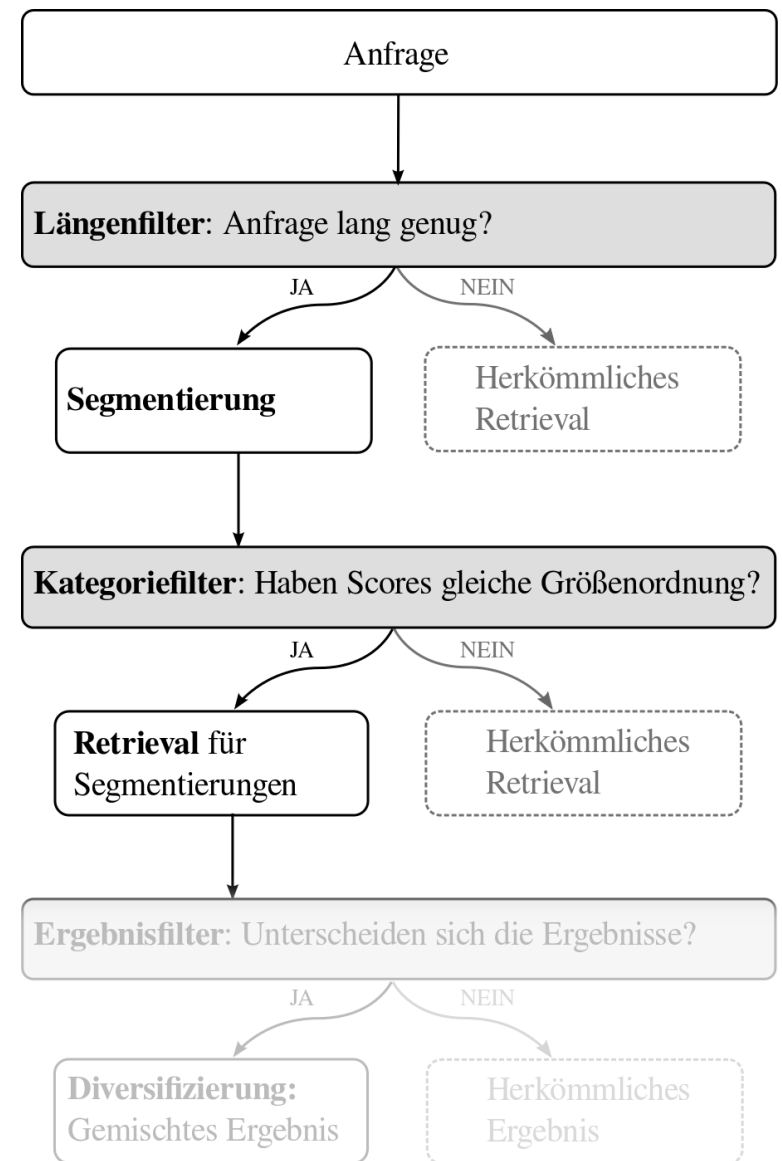


Kategoriefilter

- Kategorisierung der Anfragen
- „Unsichere“ Anfragen in den nächsten Schritt

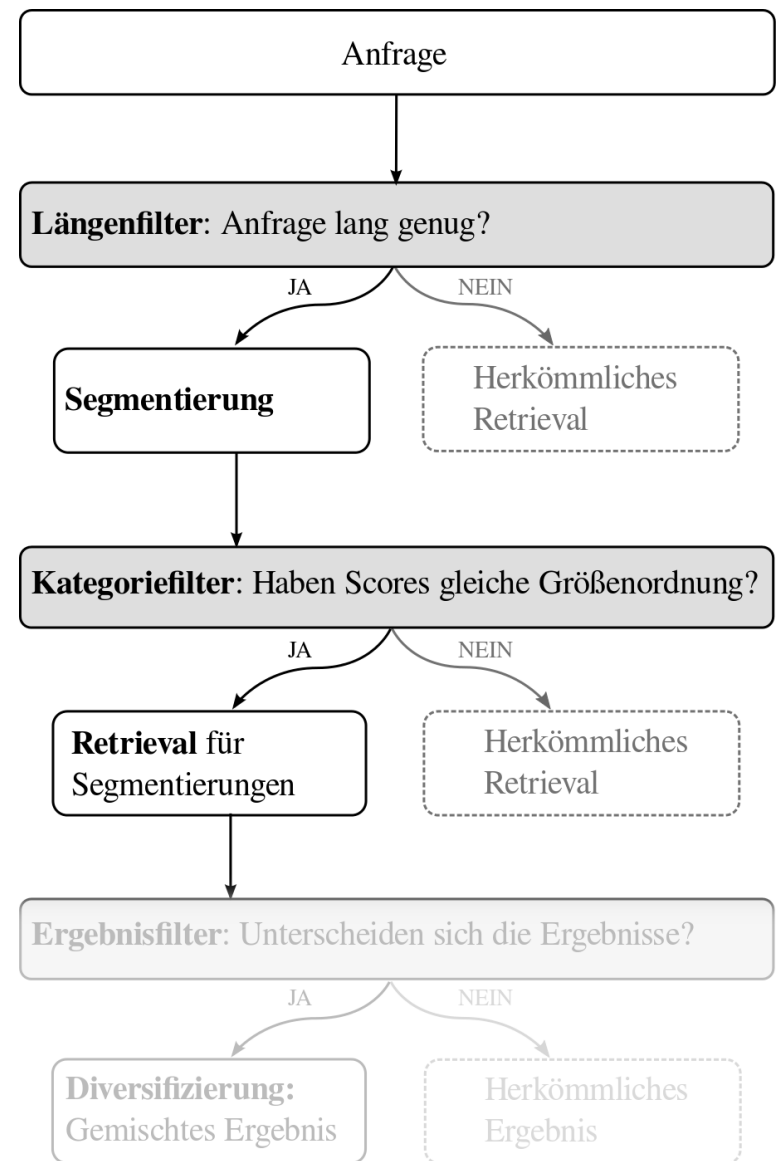


Retrievalschritt



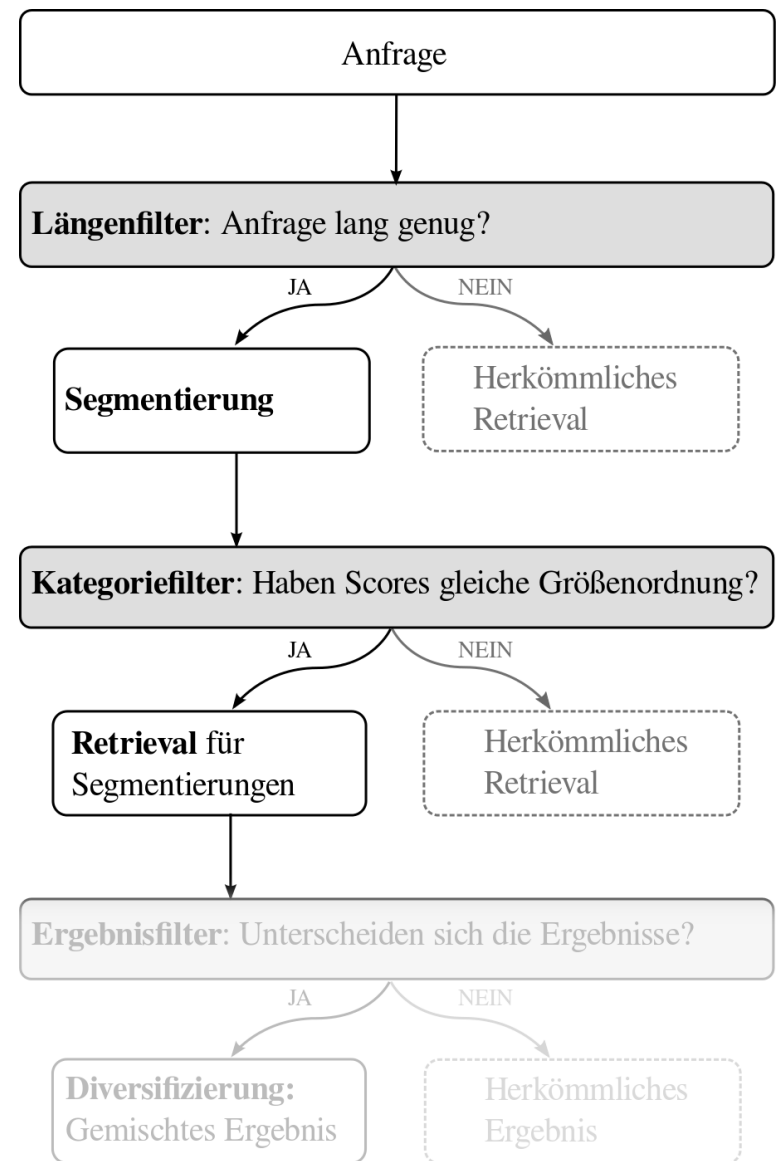
Retrievalschritt

- Suchergebnisse der beiden „besten“ Segmentierungen

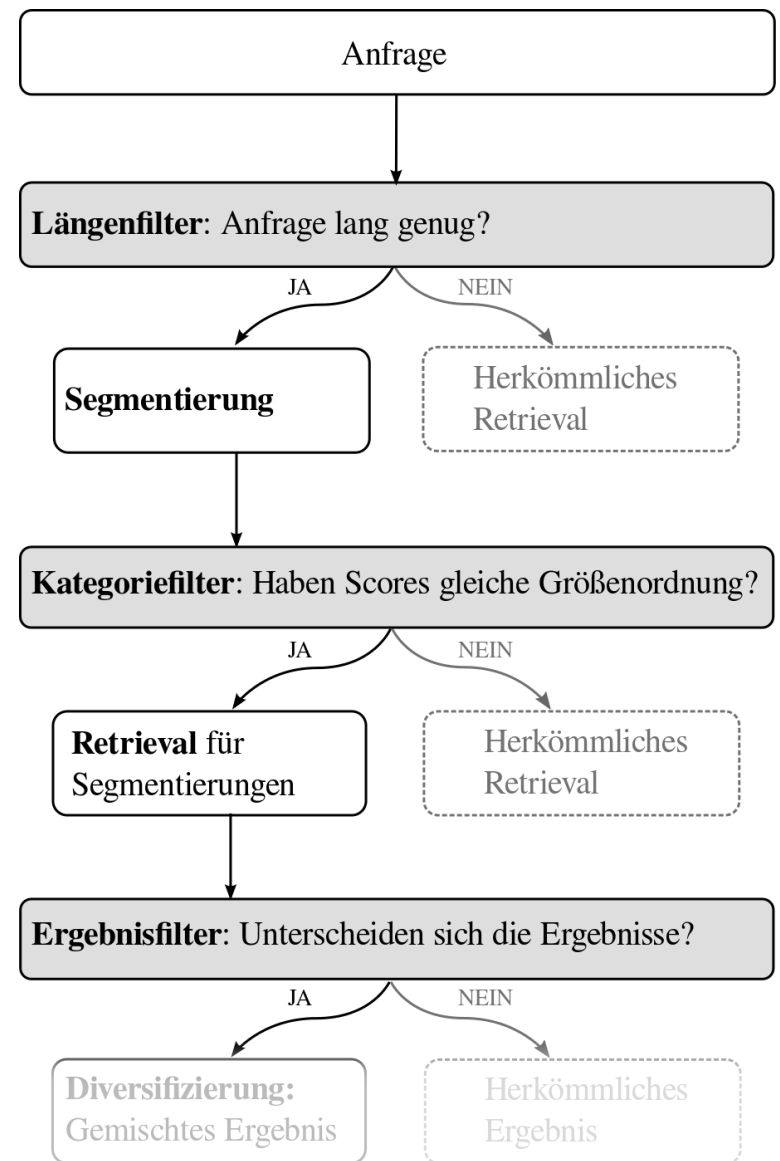


Retrievalschritt

- Suchergebnisse der beiden „besten“ Segmentierungen
- Ergebnislisten zur weiteren Verarbeitung

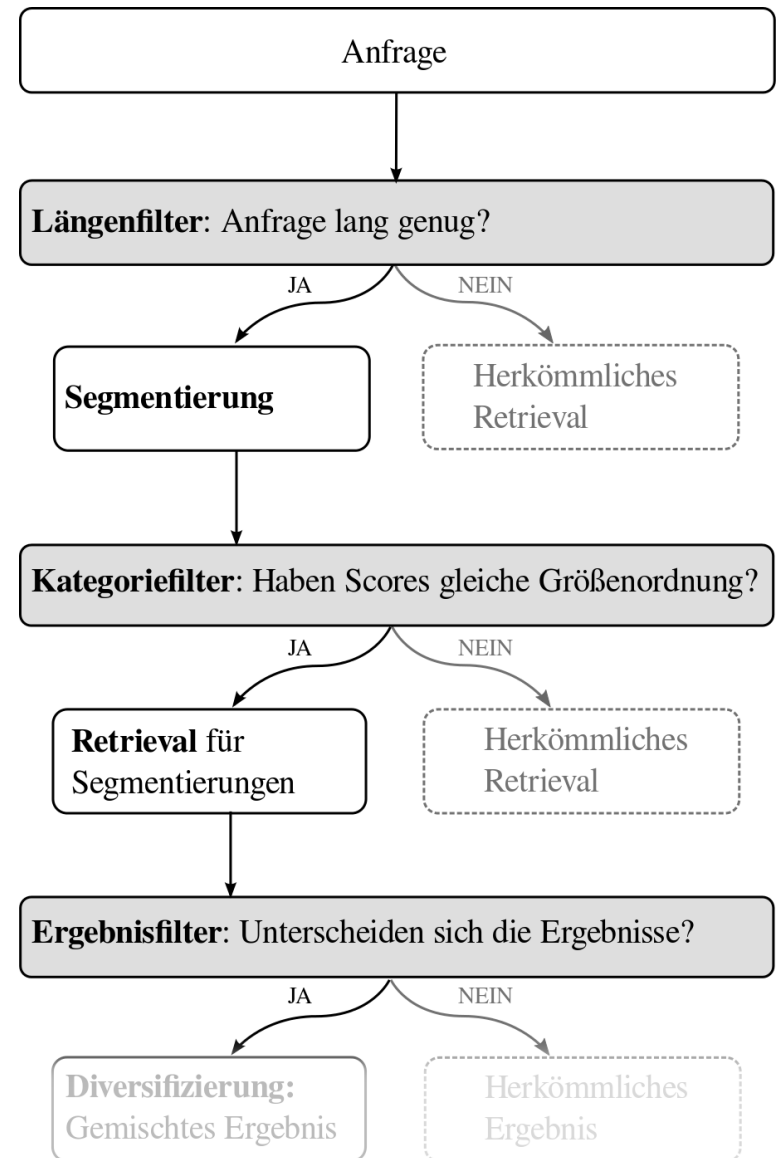


Ergebnisfilter



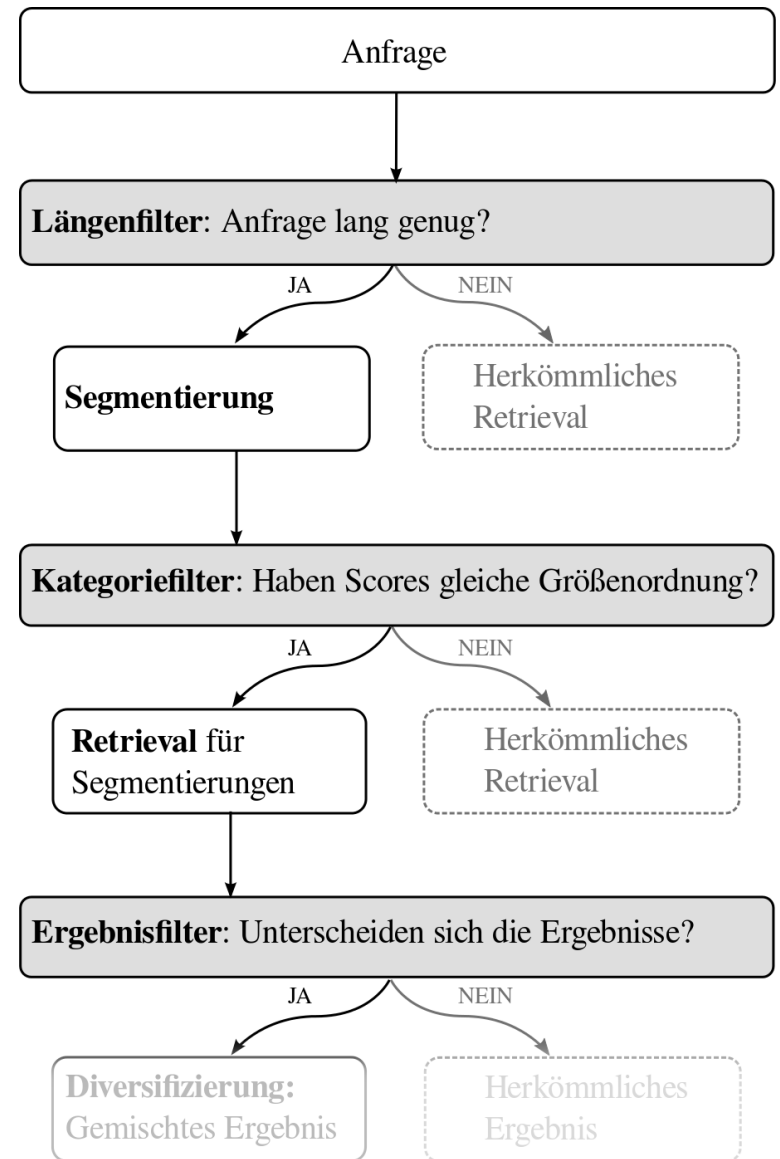
Ergebnisfilter

- Mischen der beiden Listen



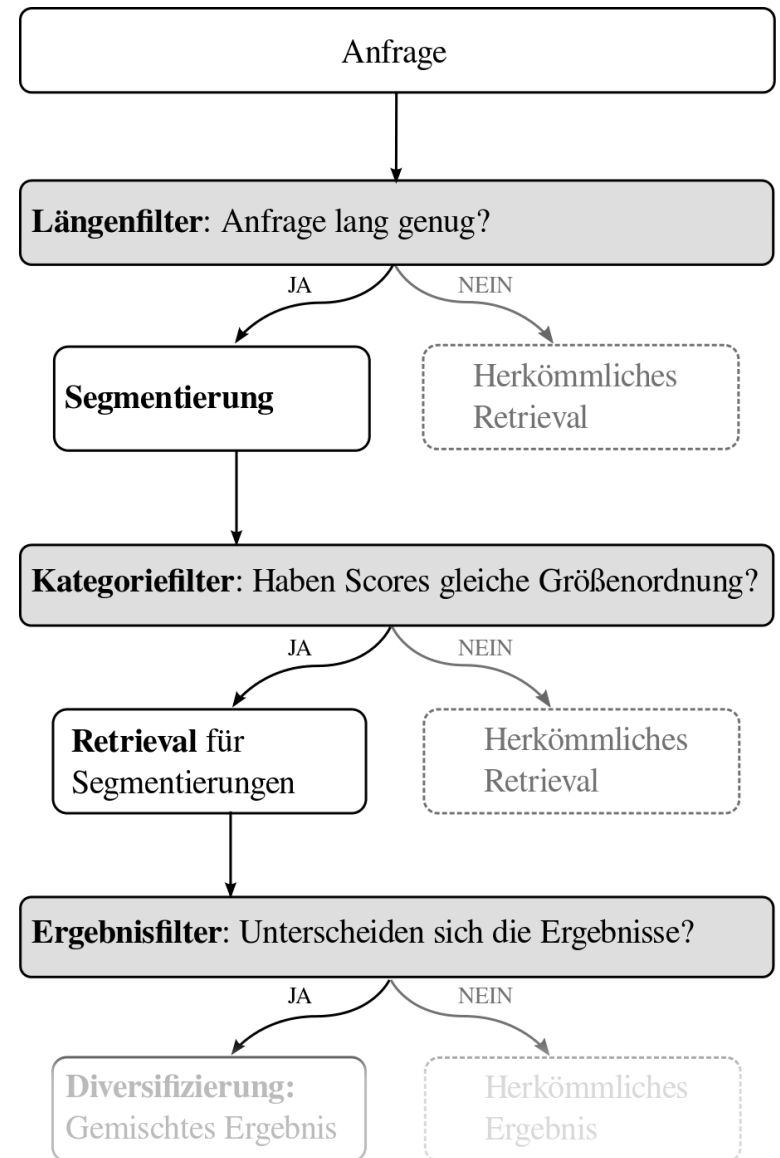
Ergebnisfilter

- Mischen der beiden Listen
- Vergleich der Ergebnisse

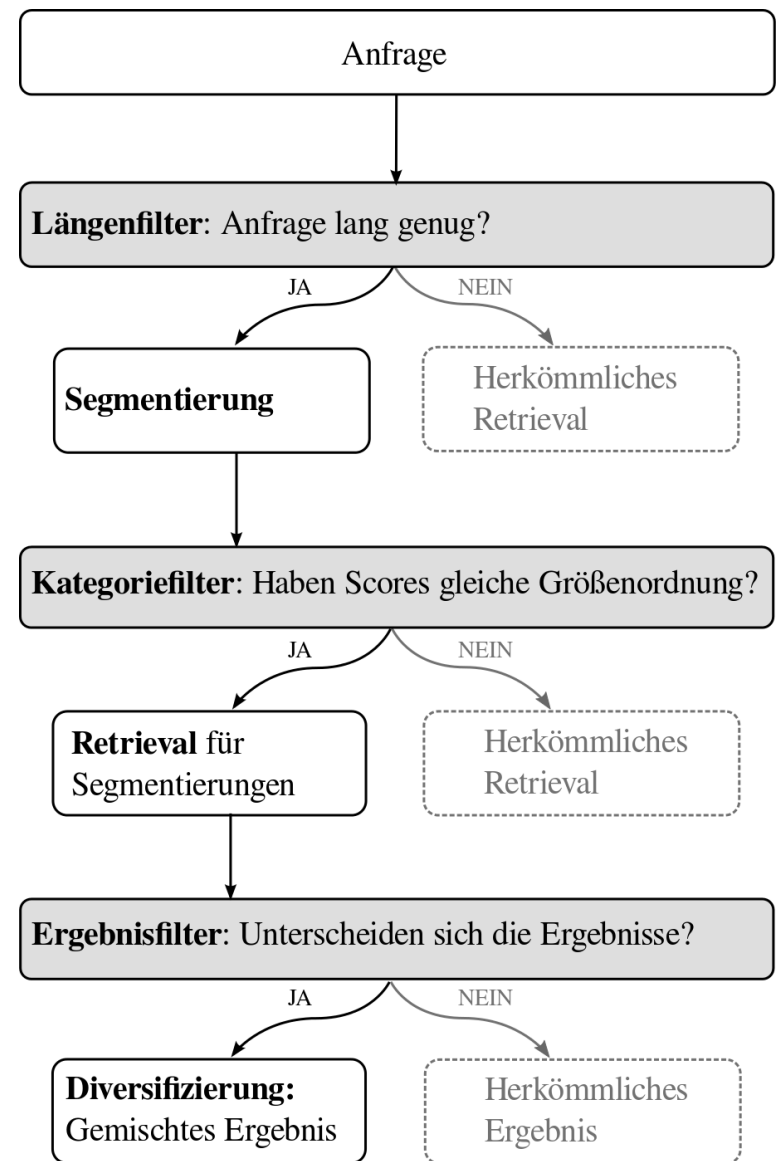


Ergebnisfilter

- Mischen der beiden Listen
- Vergleich der Ergebnisse
- Falls unterschiedlich, weiter verarbeiten

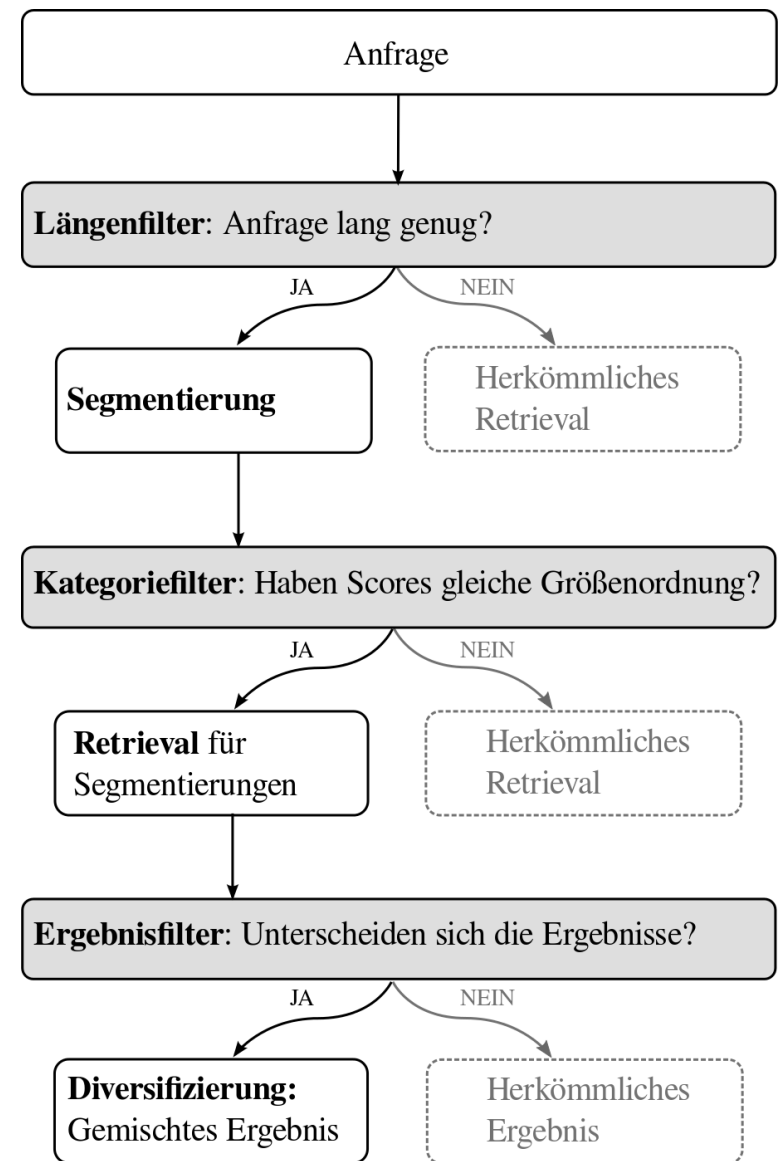


Diversifizierungsschritt



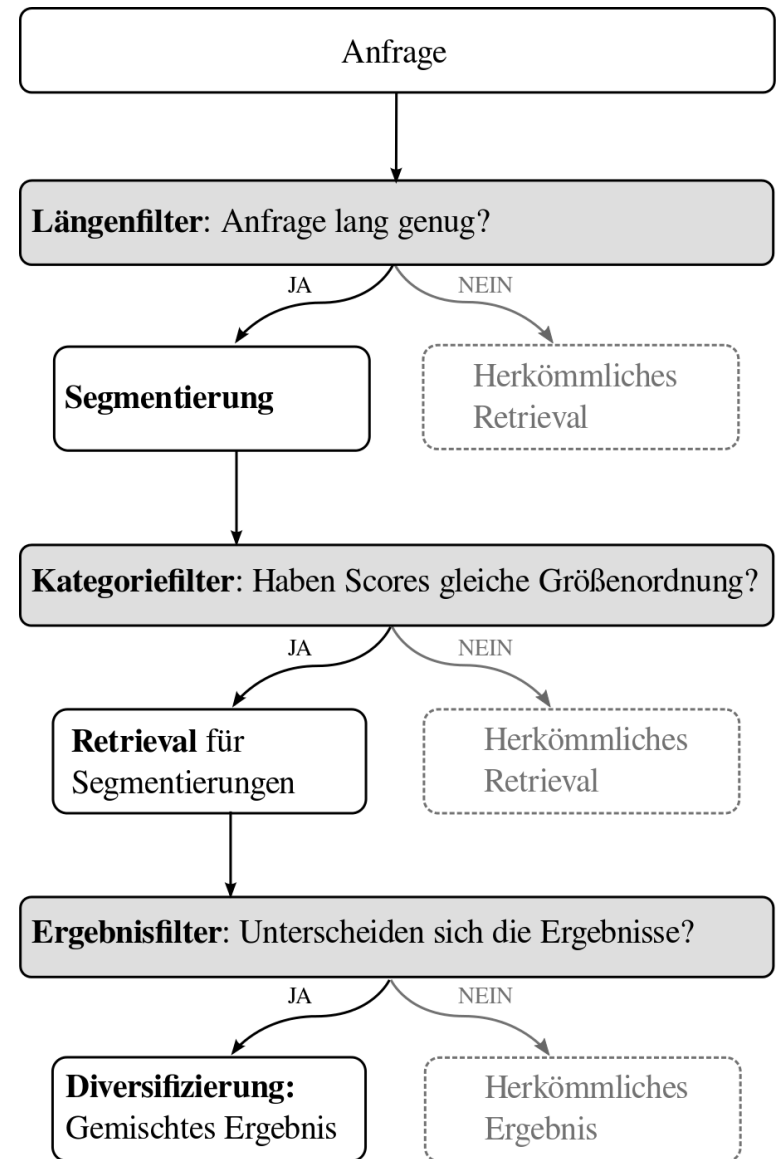
Diversifizierungsschritt

- Diversifizierte Liste als Ergebnis



Diversifizierungsschritt

- Diversifizierte Liste als Ergebnis
- Ende des Verfahrens



Evaluierung

Werden durch Verwendung des Verfahrens bessere
Suchergebnisse erzielt?

Evaluierung

Evaluierung

- Korpus aus 500 Anfragen und 14 171 zugehörigen Dokumenten [Roy et al., SIGIR 2012]

Evaluierung

- Korpus aus 500 Anfragen und 14 171 zugehörigen Dokumenten [Roy et al., SIGIR 2012]
- Anfragen zwischen 5 und 8 Wörter lang

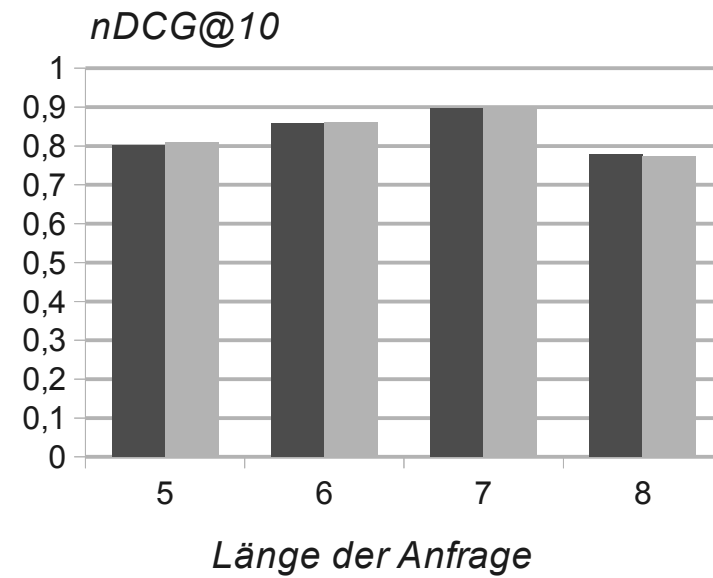
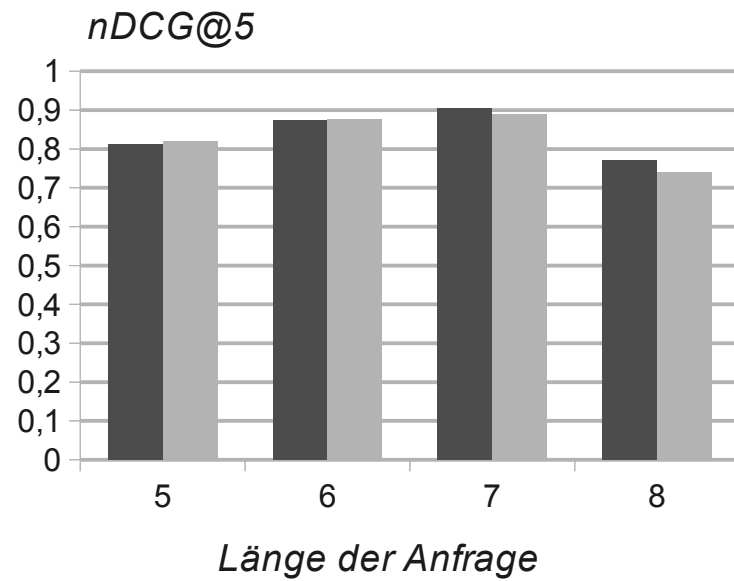
Evaluierung

- Korpus aus 500 Anfragen und 14 171 zugehörigen Dokumenten [Roy et al., SIGIR 2012]
- Anfragen zwischen 5 und 8 Wörter lang
- Relevanzbewertungen für jedes Dokument

Evaluierung

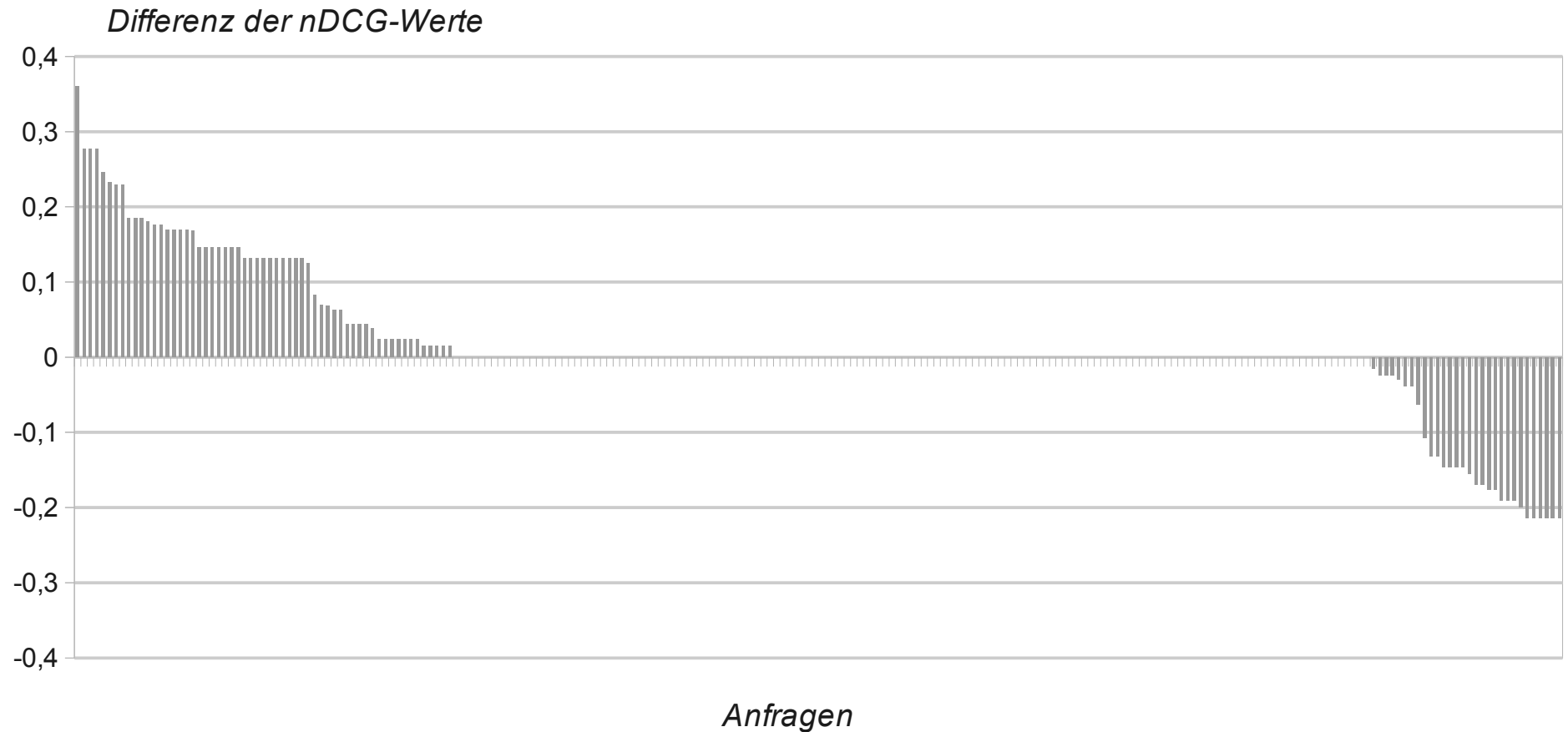
- Korpus aus 500 Anfragen und 14 171 zugehörigen Dokumenten [Roy et al., SIGIR 2012]
- Anfragen zwischen 5 und 8 Wörter lang
- Relevanzbewertungen für jedes Dokument
- nDCG als Bewertungsmetrik

Evaluierung



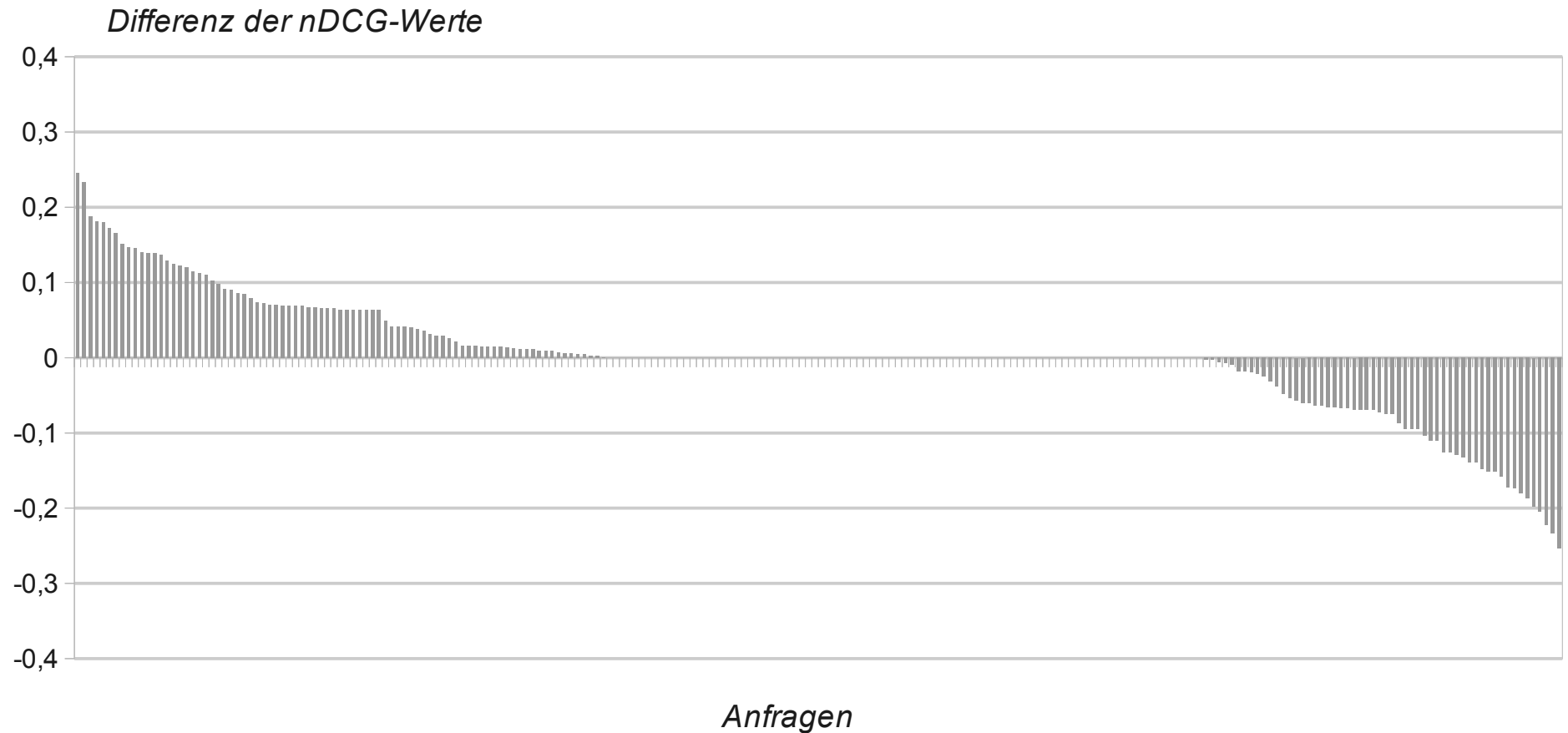
- Diversifizierte Ergebnisliste
- Ergebnisliste der besten Segmentierung

Evaluierung



Vergleich der ersten 5 Ergebnisse

Evaluierung



Vergleich der ersten 10 Ergebnisse

Evaluierung

Evaluierung

- Diversifizierte Ergebnisse sind leicht schlechter

Evaluierung

- Diversifizierte Ergebnisse sind leicht schlechter
- Zahlreiche Faktoren, die verbessert werden können:
 - Evaluierungskorpus
 - Segmentierungsauswahl
 - Diversifizierungsverfahren

Zusammenfassung

Zusammenfassung

- Pipeline als Grundlage

Zusammenfassung

- Pipeline als Grundlage
- Segmentierungen können beim Erkennen ambiger Anfragen helfen

Zusammenfassung

- Pipeline als Grundlage
- Segmentierungen können beim Erkennen ambiger Anfragen helfen
- Ergebnisse könnten verbessert werden

Zusammenfassung

- Pipeline als Grundlage
- Segmentierungen können beim Erkennen ambiger Anfragen helfen
- Ergebnisse könnten verbessert werden

Vielen Dank für Ihre Aufmerksamkeit!