

Information Retrieval – Übung Query Expansion

Bitte bearbeiten Sie die Aufgaben **vor den Übungsterminen**. In den Online-Sitzungen arbeiten wir dann mit den Ergebnissen Ihrer Vorarbeiten. Es stellt jemand aus Ihren Reihen die vorbereitete Lösung vor. Die Lösung kann anschließend im Plenum diskutiert, erweitert und in Kontext gesetzt. Auch Peer-Reviews sind möglich.

Die Präsentation der Lösung kann beispielsweise über zuvor eingereichte Videos oder Bildschirmaufnahmen, Kurzreferate mit vorbereiteten Folien oder interaktiven Kollaborationswerkzeuge wie Google Jamboards erfolgen.

I Manuelle Query-Expansion

Sie wollen einen Artikel zum Thema „Jugendarbeitslosigkeit in Ostdeutschland“ schreiben. Sie möchten sich zunächst einen Überblick über das Fachgebiet verschaffen und suchen dazu im System **EconBiz** nach einschlägigen Dokumenten.

- Entwerfen Sie eine **Blockstrategie**, indem Sie in die folgende Tabelle Begriffe eintragen, die Sie im STW – Standard Thesaurus Wirtschaft finden und die Ihnen logisch erscheinen.
- Notieren Sie eine **Begründung** für Ihre Auswahl an Begriffen.
- Wie würde eine Anfrage zu diesen Suchbegriffen in einer **booleschen Anfrage** aussehen?
- Setzen Sie Ihre Suche in EconBiz entsprechend der Anleitung unter <https://www.econbiz.de/eb/de/hilfe-erweiterte-suche/> um.

Link zum Thesaurus: <http://zbweu.stw./versions/latest/about.de.html>

Link zum Suchformular in EconBiz: <https://www.econbiz.de/Search/Advanced>

	Begriff 1	Begriff 2
	Jugendarbeitslosigkeit	Ostdeutschland
Synonyme		
Verwandte Begriffe		

II Normalized Google Distance

Die Normalized Google Distance¹ (NGD) berechnet eine semantische Ähnlichkeit oder Nähe zwischen zwei Termen, basierend auf der Anzahl der Treffer, die Google (oder eine andere Suchmaschine) liefert. Details finden Sie u.a. im Wikipedia-Artikel oder in der Original-Quelle unter <https://arxiv.org/abs/cs/0412098>.

Im Folgenden soll die Ähnlichkeit von Begriffen anhand der Normalized Google Distance bewertet werden. Zur Auswertung steht Ihnen dazu ein Google Colab Notebook zur Verfügung.

Link zum Colab Notebook:

<https://colab.research.google.com/drive/1A7Lc7L7rTS9W6TkmKFewojQd7D068hkz?usp=sharing>

In den letzten beiden Zellen können Sie Begriffe spezifizieren, für welche dann die Ähnlichkeit ausgewertet wird. Für diese Aufgabe stehen Ihnen die Suchmaschinen Google und Wikipedia bereit.

- (1) Werten Sie die Ähnlichkeit der Begriffspaare aus, die Sie in die obige Tabelle eingetragen haben.
- (2) Welche Begriffe ähneln sich am stärksten? Wie wirkt sich die semantische Ähnlichkeit auf die Ergebnisse einer Suche aus?
- (3) Wie unterscheidet sich die Normalized Google Distance zwischen den verwendeten Suchmaschinen? Wodurch kommen die Unterschiede zwischen den Suchmaschinen zustande?

¹ https://en.wikipedia.org/wiki/Normalized_Google_distance