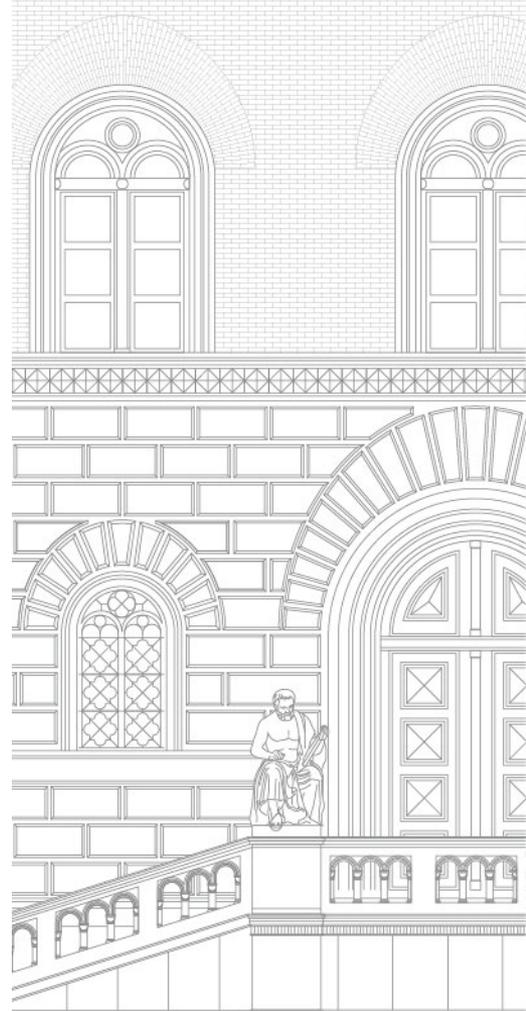


Yewno Discover & Yewno Unearth

Nutzung von Yewno Unearth zur
automatischen Inhaltserschließung von
Grauer Literatur aus Südosteuropa -
Projektziele und Workflow

Dr. Arnošt Štanzel, Ronny Sternecker



Die Firma Yewno

Yewno is an AI and data science company that aims to transform the world by turning vast amounts of information into actionable knowledge.

Yewno is an inference and analytics software as a service platform in the field of knowledge discovery

- Finanzen: Datenvisualisierung + KI für Investment-Strategien
- Bildung: Discover + Uneath
- Biomedizin: Research-Unterstützung, Modellierung komplexer Biosysteme



Yewno an der Bayerischen Staatsbibliothek

- Kooperation seit Anfang 2017
- vorerst nur Yewno Discover
- als dreimonatiger Test gestartet, dann Entwicklungspartnerschaft
- Inhalte vor allem englischsprachig, keine Vollständigkeit zu erwarten
- Ziel: erste Erfahrungen mit dieser Art von Rechercheinstrument sammeln
- Nutzung Yewno Discover eher mäßig
- Vertrag mittlerweile um Yewno Unearth und Widgets erweitert

Yewno Discover

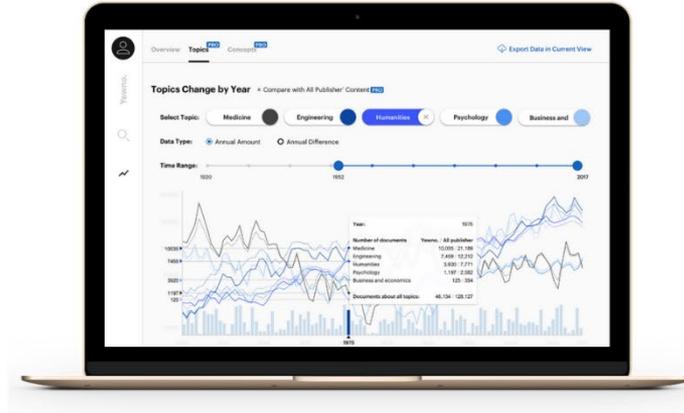
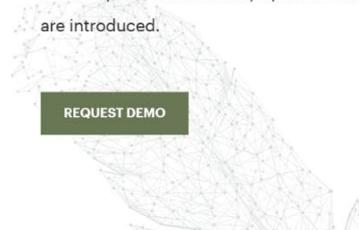
*Navigation in
„Wissensräumen“ auf
Grundlage von Konzepten
und deren Verbindungen
untereinander*

The screenshot displays the Yewno Discover application interface. On the left is a vertical navigation sidebar with icons for 'Do', 'SEARCH', 'JOURNEY', 'NOTEBOOK', 'SETTINGS', and 'HELP CENTER'. The main content area is titled 'Frankfurt School' and includes a 'Back' button, a 'FAVORITE' icon, and tabs for 'Overview', 'Concepts', and 'Documents'. Below the tabs is a portrait of a man and a 'DEFINITIONS' section with text: 'The Frankfurt School (German: Frankfurter Schule) is a school of social theory and philosophy associated in part with the Institute for Social Research at the Goethe University Frankfurt. Founded during the interwar period, the School'. A 'Source' link is visible below the text. The central part of the screen features a radial knowledge map with 'Frankfurt School' at the center, connected to various concepts: Max Horkheimer, Dialectic of Enlightenment, Character mask, Marxist schools of thought, György Lukács, Herbert Marcuse, Critical theory, Dialectical materialism, Georg Wilhelm Friedrich Hegel, Marxist philosophy, Theodor W. Adorno, and Jürgen Habermas. On the right side, there are utility icons: 'Snapshot', 'Generate Connections', 'Maps(1)', and zoom controls (+, -).

Yewno Uneath

Uneath

Yewno Uneath looks across entire collections, helping expose the full value within holdings. The system creates a topical hierarchy and automatically updates whenever new concepts are introduced.



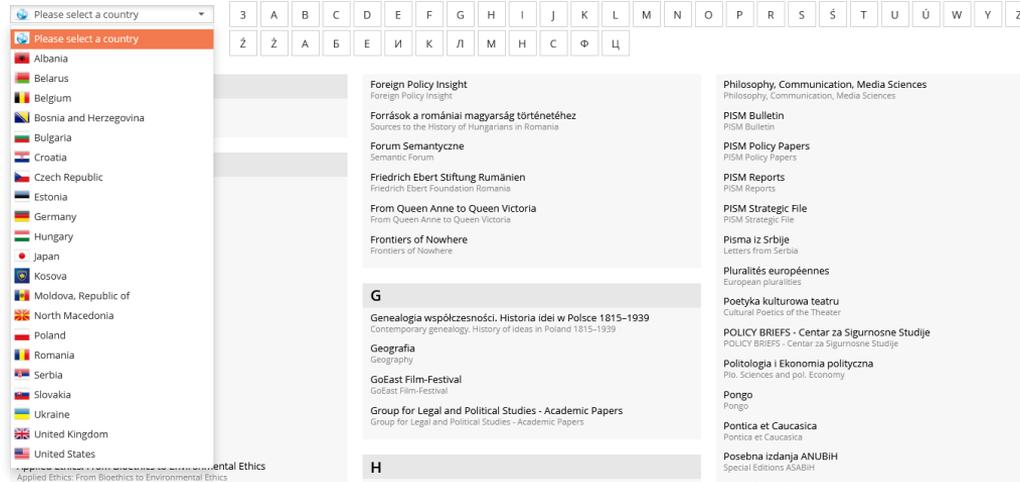
Bezeichnung für die Yewno-Technologie, mit der Konzepte in Texten extrahiert werden und die so Grundlage für den Wissensgraphen von Yewno Discover sind

Anwendungsbeispiel: Graue Literatur aus der Central and Eastern European Online Library



Browse by Grey Literature

This page is listing series of Grey Literature published by various types of institutions (NGOs, Think-Tanks, Research-Institutes...) without an ISSN-number. CEEOL is adopting the nomenclature (where exists) as preset by the publisher (as for example: „NGO1 Policy Papers“ / „NGO2 Reports“ / „Institute3 Analyses“)



Maschinelles Lernen zu Sacherschließung von Grauer Literatur

- Erwerbung Grauer Literatur aus Südosteuropa von der Central and Eastern European Online Library (CEEOL)
 - bislang wurden über 5000 Titel (PDFs mit Volltext) geliefert
 - die Texte liegen in verschiedenen Sprachen vor und haben einen Umfang von einer bis 300 Seiten
- Manuelle Verschlagwortung aus Kapazitätsgründen nicht machbar

Experiment: Nutzung der Yewno-Technologie zur automatischen Sacherschließung

- Ziel: Nutzung der Yewno-Konzepte als freie Schlagworte für Graue Literatur, sodass diese Titel für Forschende besser such- und nutzbar sind
- Werkzeug: Test des Dienstes Yewno Unearth -> Extraktion von ‚Concepts‘ ‚Topics‘ mit Hilfe von Verfahren Maschinellen Lernens



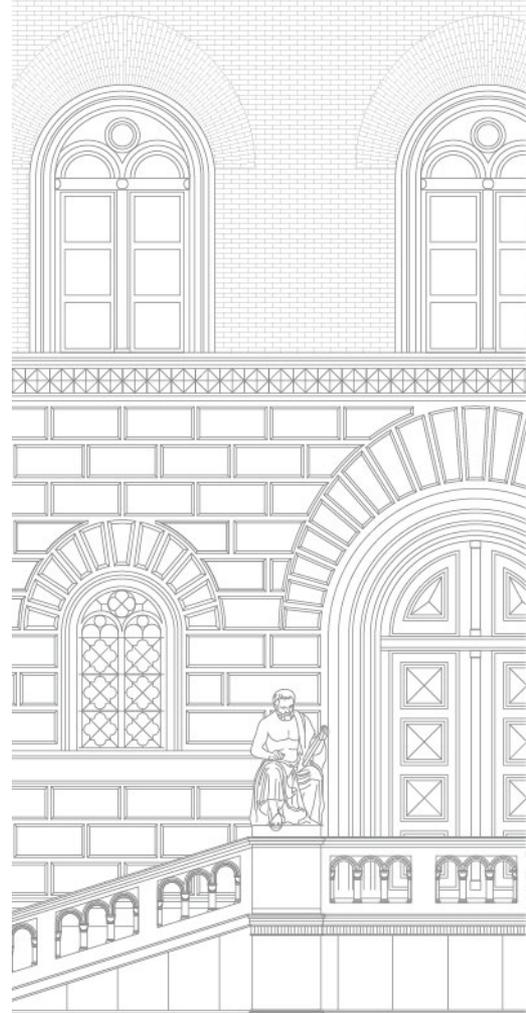
Der Ablauf

1. Transfer der PDFs samt Titeldaten im MARC-Format an Yewno
2. Automatische Übersetzung der Titel mit Google Translate
3. Analyse der Titel durch Yewno Unearth
4. Ergebnisse in Form von Konzepten und Topics werden als CSV-Datei geliefert
5. Import als Freie Schlagworte in den Verbundkatalog

	A	C	D
1	sourceId	Concept_1	Concept_2
2	ceeol527262	Privatization; 1	Restructuring; 0.310258398
3	ceeol828206	Individual ministerial responsibility; 1	Explanation; 0.8613573285
4	ceeol828029	France; 1	Printing; 0.7948882303
5	ceeol521522	Debt; 1	Credit (finance); 0.5940638536
6	ceeol698331	Montenegro; 1	Serbia; 0.6965655832
7	ceeol521378	Private sector; 1	Credit (finance); 0.6299266446
8	ceeol843824	Germany; 1	France; 0.9543726472
9	ceeol801732	Civil service; 1	Ombudsman; 0.8970227264

Yewno Discover & Yewno Unearth

Ansatz, Ergebnisse, Erfahrungen und
Schlussfolgerungen



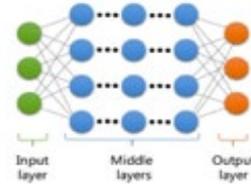
Von Daten zum Wissen: Maschinelles Lernen + Computerlinguistik + Graphentheorie

Computerlinguistische Analysen der Rohdaten generieren einen umfangreichen, vielschichtigen und mehrdimensionalen Satz von **Konzepten**

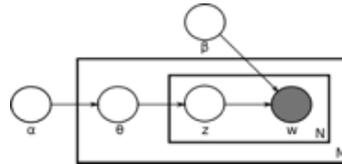


Datenerfassung →

(Deep) Neural Networks

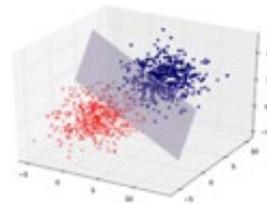


(Dynamic) Topic Models

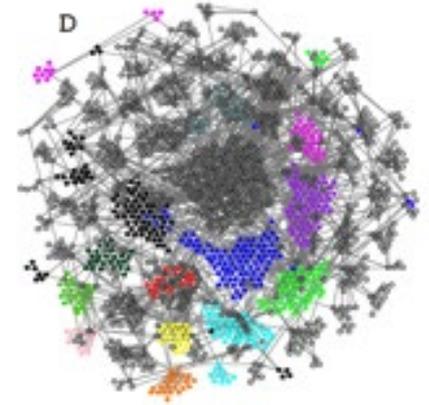


→ Extrahieren von Wissen

Stochastisches Lernen



Konzepte werden als hierarchisches Netzwerk mit Beziehungen dargestellt



Concepts, not keywords

Yewno nutzt Algorithmen neuronaler Netzwerke, Computerlinguistik und Graphentheorie mit zwei Zielen:

- um Konzepte zu identifizieren und zu extrahieren, sowohl aus unstrukturierten als auch strukturierten Daten
- um bedeutendes Wissen aufzudecken in Form von einer Kette von Verbindungen zwischen diesen Konzepten

Was ist ein Konzept?

concept 
noun | con·cept | \ˈkän·sept\
Simple Definition of CONCEPT Popularity: Top 10% of words
: an idea of what something is or how it works
Source: Merriam-Webster's Learner's Dictionary

significance 
noun | sig·nif·i·cance | \sig·ni·fi·kən(t)s\
Simple Definition of SIGNIFICANCE Popularity: Top 30% of words
: the quality of being important : the quality of having notable worth or influence
: the meaning of something

inference 
noun | in·fer·ence | \in·fə·ran(t)s, -fərn(t)s\
Simple Definition of INFERENCE Popularity: Top 1% of lookups
: the act or process of reaching a conclusion about something from known facts or evidence
: a conclusion or opinion that is formed because of known facts or evidence



Konzepte als kleinste Einheiten von Wissen

- In Kontrast zu Keyword-basierten NLP liegt der Ansatz bei Yewno in Computerlinguistik und Maschinellern Lernen (und ein klein wenig NLP)
- Umsetzung basiert auf der Idee, dass Wissen als ein Netzwerk von miteinander verbundenen kleinsten Einheiten, hier 'Konzepte' genannt, dargestellt werden kann
- Entspricht der Theorie der Begriffsräume (conceptual space) des Kognitionswissenschaftlers Peter Gärdenfors (geometrischer Ansatz für Bedeutung von Denken)



Extraktion von Konzepten

Neue Konzepte werden kontinuierlich aus einem hybrid-semantischen Korpus extrahiert

- Es werden Kandidaten für Konzepte über syntaktische Mustererkennung generiert
- Zur Weiterverarbeitung müssen diese Kandidaten beschrieben und kategorisiert werden
- Konzepte werden weiterverarbeitet, wenn diese verständlich im Text beschrieben werden können (ein Minimum an Beschreibungen mit ähnlichem Inhalt erforderlich)

Zufälliges Dokument

Quantitative easing (QE) is a type of **monetary policy** used by central banks to stimulate the economy when standard **monetary policy** has become ineffective. Quantitative easing, and **monetary policy** in general, can only be carried out if the central bank controls the currency used in the country

Extraktion eines Kandidaten

monetary_policy

Suche nach Beschreibung

Yewno syntagmatic Hub



Generierung der Beschreibung

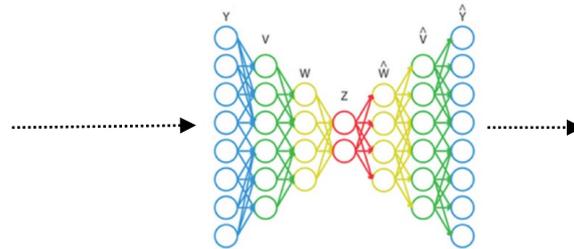
Konzept-Definition

Monetary policy is the process by which the monetary authority of a country controls the supply of money, often targeting an inflation rate or interest rate to ensure price stability and general trust in the currency.

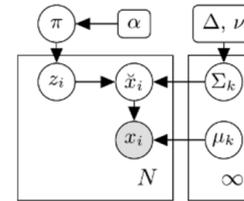
Überführung von Konzepten in ein Netzwerk

Yewno kann Konzepte in einem großen Korpus an unstrukturierten Daten erkennen und deren Bedeutung in ein Wissensnetzwerk überführen, in welchem Beziehungen untereinander erkannt und dargestellt werden.

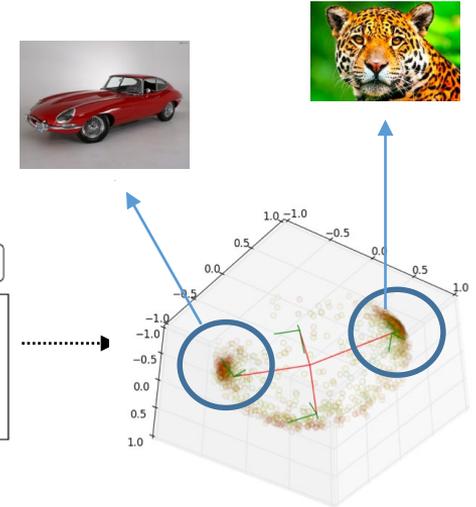
This turnaround led to parent-company Ford, and **Jaguar's** then chairman Wolfgang Reitzle, to set ambitious sales targets in the belief that **Jaguar** could become a major luxury car manufacturer in the mould of Audi, BMW, and Mercedes-Benz.



Auto-encoder
Deep Learning Network



Dirichlet
Prozess



Projektion in ein Netzwerk mit
Beziehungen

Integration – Yewno Discover

The screenshot displays the Yewno Discover interface. On the left, a sidebar contains navigation icons: a person icon (profile), a magnifying glass (SEARCH), a person with a speech bubble (JOURNEY), a notepad (NOTEBOOK), a gear (SETTINGS), a question mark (HELP CENTER), and a person with a checkmark (NOTES).

The main content area shows search results for "Treaty of Lisbon". The top result is "Treaties of the European Union", followed by "Treaty establishing Constitution for Europe", "European Union", "European Union legislative procedure", and "European Council".

On the right, a network diagram visualizes the relationships between concepts. The "Treaty of Lisbon" is at the bottom, and "European Union" is at the top. Numerous other concepts are connected to them, including "European Economic Community", "Languages of the European Union", "European integration", "Foreign relations of the European Union", "European Communities", "Treaty of Accession 2003", "European Union law", "Treaties of the European Union", "Future enlargement of the European Union", "Charter of Fundamental Rights of the European Union", "Council of the European Union", "President of the European Council", "Signing of the Treaty of Lisbon", "European Council", "Voting in the Council of the European Union", "European Union legislative procedure", "Amato Group", "European Committee of the Regions", "Institutions of the European Union", "Treaty establishing a Constitution for Europe", "Enhanced cooperation", "European Parliament", "President of the European Commission", and "European Commission".

Integration

2009 (3338)
 >= 1879 (211730)

▼ Medienart

- Druckschrift (102538)
- Monographie (102111)
- mit Illustrationen (23212)
- Computerdatei (16643)
- Online-Ressource (16606)

Mehr anzeigen ...

▼ Sprache

- Englisch (45190)
- Chinesisch (30181)
- Deutsch (17161)
- Unbekannt (5214)
- Französisch (2023)

Mehr anzeigen ...

▼ Autor / Hrsg

- Gu gong bo wu yuan (... (501)
- Zhongguo guo jia tu ... (403)
- ProQuest (Firm) (280)
- Zhou, Jinfu (215)
- Chen, Jianhua (210)

Mehr anzeigen ...

Yewno Discovery Service

SEARCH TERM

China



Connections: 1 - 10/30

3.  **China and the World in a Changing Context: Perspectives from Ambassadors to China**
 Wang, Huijiao; Miao, Lu
 Singapore, 2022
 Buch
 Online lesen

4.  **The earliest China**
 Xu, Hong, 1963-
 Singapore, [2022]
 Buch

5.  **Evaluation on government transparency index in China (2009-2016)**
 Lü, Yanbin, 1976-; Tian, He, 1956-
 Singapore, [2022], [2022]
 Buch
 Versand/Ausleihen

6.  **US-China relations: perilous past, uncertain future**
 Sutter, Robert G., 1943-
 Lanham ; Boulder ; New York ; London, [2022]
 Buch
 Eine Bestellung ist derzeit leider nicht möglich

7.  **Engendering China: Women, Culture, and the State**
 Barlow, Tani E., 1950-; Campbell, S. Katherine; Carlitz, Katherine; Furth, Charlotte; Gilmarin, Christina K.; Gilmarin, Christina K.; Hershatter, Gail; Hershatter, Gail; Hong, Zhu; Mann, Susan; Rofe, Lisa; Rofe, Lisa; Silber, Cathy; White, Tyrene; White, Tyrene; Woo, Margaret Y. K.; Xiaojiao, Li; Xiaoxian, Gao; Xueping, Zhong; Yiyun, Chen; Zyun, Li
 Cambridge, MA, [2022]
 Buch
 Online lesen

8.  **China's role in global governance**
 Zhang, Bin
 Singapore, [2022]
 Buch
 Eine Bestellung ist derzeit leider nicht möglich

9.  **Civil society in China: how society speaks to the state**
 Qiaoan, Runya, 1968-
 London ; New York, 2022
 Buch
 Eine Bestellung ist derzeit leider nicht möglich

China's freedom rally contributions: lessons from China's 1989 Tiananmen

civil rights movement

0 ausgewählt SEITE 71.183 Ergebnisse Personalisieren

Ergebnisse optimieren

Meine Ergebnisse erweitern

Sortieren nach Relevanz

Verfügbarkeit

Bereitgestellt von der Bibliothek

Online verfügbar

Open Access

Peer-überprüfte Zeitschriften

Thema (Schlagwort)

Civil Rights (14.299)

Science & Technology (16.254)

Civil Rights Movements (15.714)

Social Sciences (14.361)

Physical Sciences (10.335)

Weitere anzeigen

Yewno Discover



1.  **The civil rights movement**
 Newman, Mark
 Mehrere Versionen
 12 Versionen gefunden. Alle Versionen anzeigen >

2.  **The Civil Rights Movement**
 Newman, Mark
 2019
 This introduction to the Civil Rights Movement synthesizes its history, exploring its origins, development and results as well as historiographical debates...
 Online verfügbar >

3.  **The civil rights movement**
 Patterson, Charles
 1995
 Verfügbar bei Bayerische Staatsbibliothek (Library) Magazin Neu (95.4283) >

4.  **The civil rights movement**
 Levy, Peter B., 1956-
 1998
 Verfügbar bei Bayerische Staatsbibliothek (Library) Magazin Neu (98.28521) >

5.  **The Civil Rights movement**
 Verfügare Services überprüfen >

6.  **Northern protest: Martin Luther King, Jr., Chicago, and the civil rights movement**
 Ralph, James R. Jr. 1960-
 1993



Unsere Erfahrungen

Yewno Uneath

Vorarbeiten: Fine-Tuning der Yewno Datenanalysealgorithmen

- Erstes Sample mit knapp 70 Titeln in sechs verschiedenen Sprachen wird von Yewno analysiert und 15 Konzepte pro Titel geliefert
- Mitarbeiter*innen der Osteuropa-Abteilung der BSB haben geprüft, wie gut die Konzepte die Inhalte treffen
- Insgesamt drei Korrekturrunden für extrahierte Konzepte mit Hilfe einer Bewertungsmatrix von Yewno
- aufgrund von Testdaten Verbesserungen bei Konzepten zu Geografika und Personennamen erbeten

1	Instructions	
2	Please evaluate the concepts listed below the document title, apply	
3	ranking color (using the copy formatting paint brush tool) and enter the	
4	numeric score in the blue-shaded cell.	
5		
6	Ranking	
7	Relevant Precise Concept = 2	
8	Relevant Broad Concept = 1	
9	Irrelevant, Redundant, Ambiguous = 0	
10	Fragment, Error, Noise = -1	
11		Score
45		11
46	Odnos države prema crkvi	
47	Concordat	2
48	Catholic Church	2
49	God	-
50	Diocese	2
51	Archbishop	1
52	Seminary	-
53	Catholic Action	2
54	Lateran Treaty	1
55	Italy	1
56	Sovereignty	1
57	Canon law	2
58	Priesthood (Catholic Church)	(1)
59	Pope	2
60	Latvia	(1)
61	Clergy	2
62		16

Auswertung der ersten Testergebnisse

In einer ersten Lieferung wurden 970 Titel von Yewno analysiert

Pro Titel werden 20 Konzepte mit eigenem Relevanzscore geliefert sowie drei Topics

Derzeit weitere Feedbackschleife mit Rückmeldungen zu den extrahierten Konzepten

Positiv:

- größtenteils gute Beschreibung der Inhalte
- Es wurden teilweise Konzepte extrahiert, die den Inhalt gut beschreiben, die als Begriffe im Text selbst aber nicht vorkommen.
- 3 Topics (jeweils drei Sub-Topics) im Sinne einer inhaltlichen Klassifikation werden meist gut erkannt.



Auswertung der ersten Testergebnisse

Verbesserungswürdig:

- in wenigen Fällen werden völlig falsche Konzepte oder wenig aussagekräftige Abkürzungen wie „V“, „fig.“ oder „educ.“ extrahiert
- Fehlerquote bei extrahierten Konzepten liegt bei ca. fünf Prozent
 - "Agriculture (Chinese mythology)" in einem Text über Landwirtschaft in Serbien
- Basierend auf falschen Konzepten werden auch falsche Topics geliefert; bei nur drei Haupttopics schnell eine Fehlerquote von 33%
- Relevanzscores gelten pro Titel und sind nicht untereinander vergleichbar
- Kein Einblick in den Yewno-Algorithmus



Generelle Beobachtungen zu den Testergebnissen

- Konzepte zu Texten auf Englisch und Französisch sind treffender
- Zeitungen wegen Layout problematisch
- je älter die Texte desto schlechter die Ergebnisse
- Konzepte sind in ihrer Mehrzahl korrekt, allerdings fehlt Zuspitzung/Abstraktion; Topics ersetzen gut gewählte Schlagworte (noch **nicht**)
- Im Vergleich zur Sacherschließung an der BSB werden generell weniger Geographika und Personen von Yewno als Konzepte extrahiert.
- osteuropäische Geographika und Personen werden schlechter erkannt als westliche
- einige falsche Konzepte sind offensichtlich Resultat der nicht immer korrekten automatisierten Übersetzungen durch Google Translate



Diskussion

- Fehlerquote von 5% bei Konzepten: Ist das gut oder schlecht?
- Einsatz von „KI“ einer privatwirtschaftlichen Firma:
 - Von Methoden des maschinellen Lernens mit geringem Personalaufwand profitieren vs. fehlender Einblick in die zu Grunde liegenden Algorithmen.
- Was ist der beste Weg, um Nutzenden die Yewno-Konzepte im OPAC anzuzeigen und zu erläutern?
- Yewno-Widget: Nette Spielerei oder echte Hilfe für Nutzer*innen zur Knowledge Discovery?



Mehr Infos zu Yewno

Website der Bayerischen Staatsbibliothek:

<https://www.bsb-muenchen.de/suchen-und-finden/yewno/>

Publikationen:

Artificial Intelligence in Libraries and Publishing. June 2022. Edited by Ruth Pickering and Matthew Ismail. ISBN: TBC

Web Scale Discovery Services. Principles, applications, discovery tools and development hypotheses. March 2022. Roberto Raieli (Yewno: Chapter 2, par. 2.3, p. 52-54). Paperback ISBN: 9780323902984; eBook ISBN: 9780323902991

Publishing in the Digital Age. 2021. Michael Ross. (Chapter 7: Yewno). ISBN: 9781000454635, 1000454630

Text Mining for Information Professionals. November 2021. Manika Lamba and Margam Madhusudhan (Chapter 5 - Yewno Discover: A Next Generation Research Tool)

Unveil stock correlation via a new tensor-based decomposition method. arXiv.org. 2019. Gramatica, R., Di Matteo T.. [arXiv:1911.06126](https://arxiv.org/abs/1911.06126)

Random and frozen states in complex triangulations. Philosophical Magazine, 92(1-3). 2012. Aste, T., Gramatica, R. & Di Matteo, T. <http://dx.doi.org/10.1080/14786435.2011.613861>

Dynamical generalized Hurst exponent as a tool to monitor unstable periods in financial time series. Physica A: Statistical Mechanics and its Applications, 391(11). 2012. Morales, R., di Matteo, T., Gramatica, R., & Aste T. <http://dx.doi.org/10.1016/j.physa.2012.01.004>

Exploring complex networks via topological embedding on surfaces. Physical Review E, 86(3). 2012. Aste, T., Gramatica, R. & Di Matteo, T. <http://dx.doi.org/10.1103/physreve.86.036109>

Graph Theory Enables Drug Repurposing – How a Mathematical Model Can Drive the Discovery of Hidden Mechanisms of Action R. Lambiotte, ed. PLoS ONE, 9(1). 2014. Gramatica, R. et al. <http://dx.doi.org/10.1371/journal.pone.0084912>

Structuring data in a knowledge graph: <https://patents.google.com/patent/US10528871B1/>

Information network with linked information nodes: <https://patents.google.com/patent/US10025862B2/>

