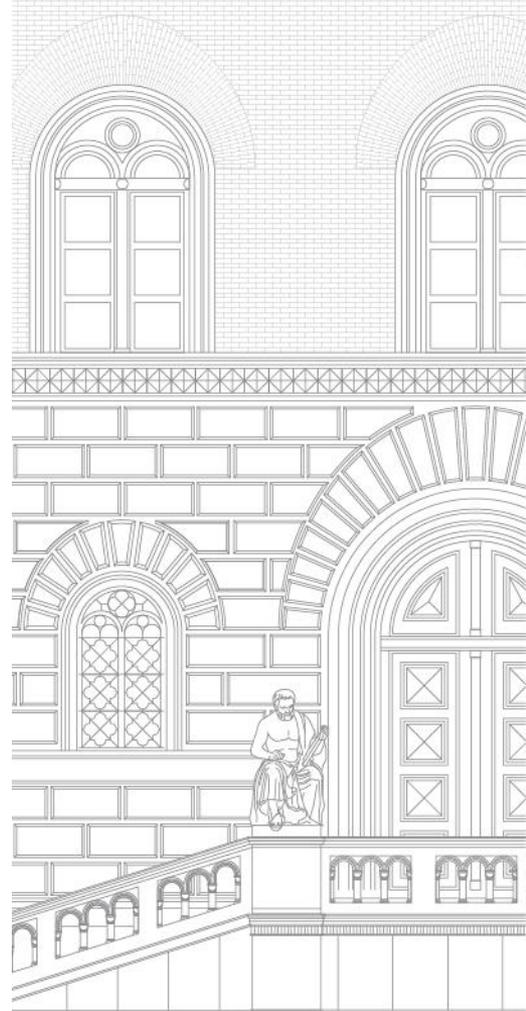


# Yewno Discover & Yewno Unearth

Ansatz, Ergebnisse, Erfahrungen und Schlussfolgerungen



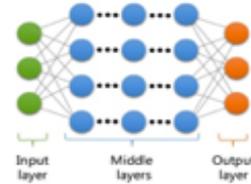
# Von Daten zum Wissen: Maschinelles Lernen + Computerlinguistik + Graphentheorie

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

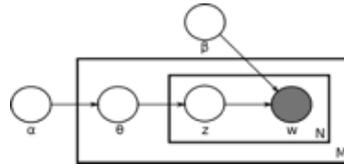


Datenerfassung →

(Deep) Neural Networks

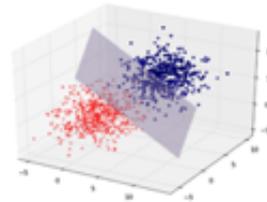


(Dynamic) Topic Models

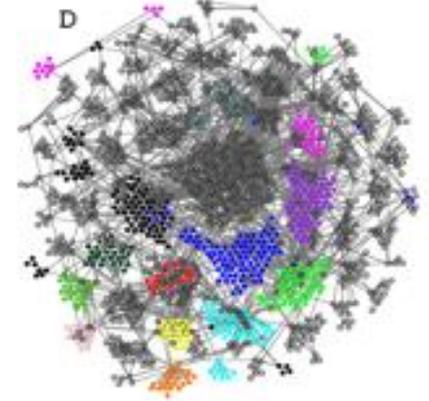


→ Extrahieren von Wissen

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·ˈni·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 | \ˈin·fə-ˈran(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 Maschinellem 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

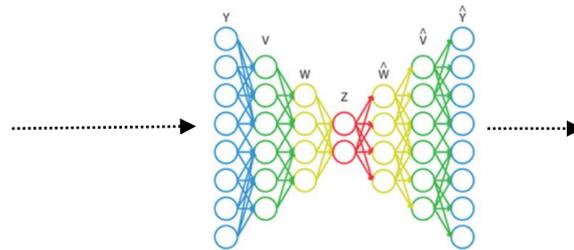
Konzept-Definition

**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.

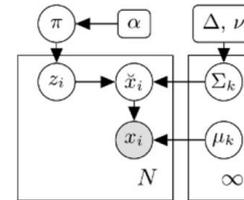
# Überführung von Konzepten in ein Netzwerk

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.

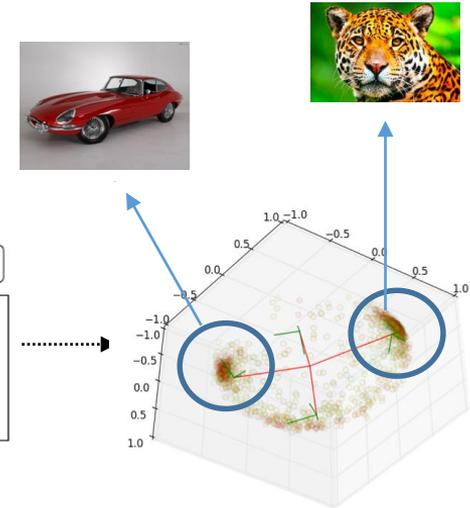
This turnaround led to parent-company Ford, and **Jaguar's** then chairman Wolfgang Reitzle, to set ambitious sales targets in the belief that **Jaguar** could become a major luxury car manufacturer in the mould of Audi, BMW, and Mercedes-Benz.



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, a magnifying glass labeled 'SEARCH', a person with a speech bubble labeled 'JOURNEY', a notepad labeled 'NOTEBOOK', a gear labeled 'SETTINGS', a question mark labeled 'HELP CENTER', and a person with a speech bubble labeled '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'. The right side of the interface features a network diagram with 'European Union' at the top and 'Treaty of Lisbon' at the bottom. Numerous nodes are connected by lines, representing relationships between concepts such as 'European Economic Community', 'European Parliament', 'European Commission', 'European Council', 'European Union law', 'European Union legislative procedure', 'European Council of the Regions', 'Council of the European Union', 'European Council', 'President of the European Council', 'Signing of the Treaty of Lisbon', 'Charter of Fundamental Rights of the European Union', 'Future enlargement of the European Union', 'Treaties of the European Union', 'European Communities', 'European integration', 'Languages of the European Union', 'Treaty of Accession 2003', 'Foreign relations of the European Union', 'Enhanced cooperation', 'Institutions of the European Union', 'Treaty establishing a Constitution for Europe', 'Amato Group', 'European Union legislative procedure', 'Voting in the Council of the European Union', and 'President of the European Commission'. The interface also includes a 'Back' button, a 'FAVORITE' star, and a 'Snapshot' button.

# Integration

2009 (3338)  
 >= 1879 (211730)

▼ Medienart

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

Mehr anzeigen ...

▼ Sprache

- Englisch (45190)
- Chinesisch (38181)
- Deutsch (17161)
- Unbekannt (5214)
- Französisch (2023)

Mehr anzeigen ...

▼ Autor / Hrsg

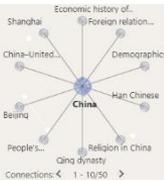
- Gu gong bo wu yuan (... (501)
- Zhongguo guo jia tu... (403)
- ProQuest (Firm) (280)
- Zhou, Jinfu (215)
- Chen, Jianhua (210)

Mehr anzeigen ...

Yewno Discovery Service

SEARCH TERM

**China**



Connections: < 1 - 10/50 >

3.  **China and the World in a Changing Context: Perspectives from Ambassadors to China**  
 Wang, Huiyao; 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; Rofel, Lisa; Rofel, Lisa; Silber, Cathy; White, Tyrene; White, Tyrene; Woo, Margaret Y. K.; Xiaojiao, Li; Xiaoxian, Gao; Xueping, Zhong; Yijun, Chen; Ziyun, 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, 1988-  
 London ; New York, 2022  
 Buch Eine Bestellung ist derzeit leider nicht möglich

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

civil rights movement

ERWEITERTE SUCHE

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 (14.399)

Science & Technology (16.254)

Civil Rights Movements (15.714)

Social Sciences (14.361)

Physical Sciences (10.335)

Weitere anzeigen

Yewno Discover



1.  **MEHRERE VERSIONEN**  
**The civil rights movement**  
 Newman, Mark  
 102 Versionen gefunden. Alle Versionen anzeigen >

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

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

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

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

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



# Unsere Erfahrungen

## Yewno Uearth

# 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	<b>Instructions</b>	
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	<b>Ranking</b>	
7	Relevant Precise Concept = 2	
8	Relevant Broad Concept = 1	
9	Irrelevant, Redundant, Ambiguous = 0	
10	Fragment, Error, Noise = -1	
11		<b>Score</b>
45		<b>11</b>
46	<b>Odnos države prema crkvi</b>	
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		<b>16</b>

# 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/>

