



Yewno

Transforming Information into Knowledge

Yewno Vicky Hampshire

BIC New Trends in Publishing Seminar 2018





Contents

- Our Common Challenge
- Yewno Solutions
- Yewno Unearth
- Yewno Unearth: Use Cases

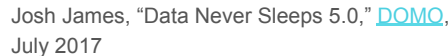


Our Common Challenge

Transforming Information into Knowledge



- ## Digital Data Alone



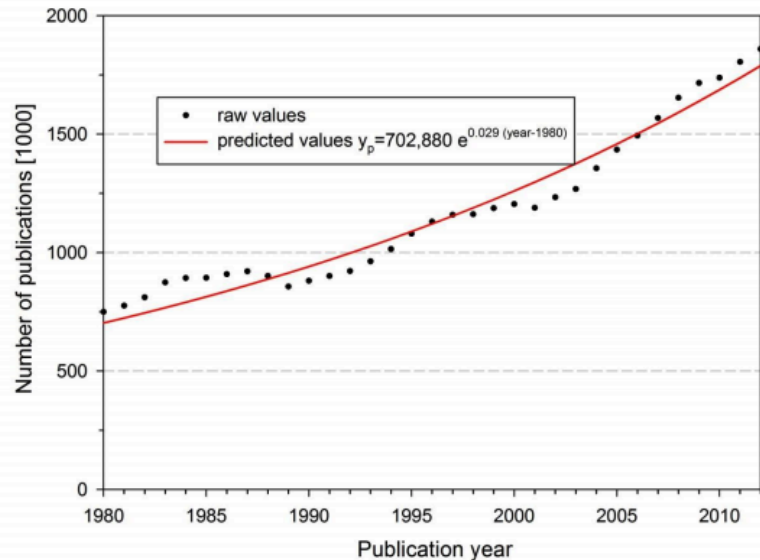


Our Common Challenge

Exponential Growth in Scientific Outputs

- Global scientific output growth rate
~8-9% per year
- Translates to a doubling every 9 years
- Growth phases, each of which led to growth rates tripling in comparison with the previous phase:
 - From less than 1% up to the middle of the 18th century
 - 2 to 3% up to the period between the two world wars
 - 8 to 9% to 2012

Scientific Output from 1980 to 2012



analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology, 66(11),. Available at: <http://dx.doi.org/10.1002/asi.23329>.



Yewno Solutions

A Profoundly New Approach to
Knowledge Discovery



Yewno Approach

Transforming Information into Decision-Making Knowledge

“What distinguishes knowledge from information is the way in which knowledge empowers actors with the capacity for intellectual or physical activity. Knowledge is a matter of cognitive capability and enables actors to do and reflect. Information, by contrast, is passive and meaningless to those without suitable knowledge. Knowledge provides the means by which information is interpreted and brought to life.*

Yewno transforms unstructured textual data into decision-making knowledge

- With little – or no – human supervision
- In real-time
- Efficiently

*Ian Brinkley, Rebecca Fauth, & Sotiria Theodoropoulou.
(2009). [A Knowledge Economy Programme Report](#).



Yewno Approach

Conceptualization of Information to Produce Knowledge

Yewno leverages machine learning, computational linguistics, and graph theory to achieve two goals:

- To identify and extract concepts from both structured and unstructured information
- To unearth significant knowledge via an inferential chain of connections between identified concepts

con·cept

/ˈkæn,sept/ 🔊

noun

an abstract idea; a general notion.
"structuralism is a difficult concept"
synonyms: **idea, notion, conception, abstraction; More**

- a plan or intention; a conception.
"the center has kept firmly to its original concept"
- an idea or invention to help sell or publicize a commodity.
"a new concept in corporate hospitality"

sig·nif·i·cance

/sigˈnɪfɪkəns/ 🔊

noun

1. the quality of being worthy of attention; importance.
"adolescent education was felt to be a social issue of some significance"
synonyms: **importance, import, consequence, seriousness, gravity, weight, magnitude,**
momentousness; **formal moment**
"a matter of considerable significance"

2. the meaning to be found in words or events.
"the significance of what was happening was clearer to me than to her"
synonyms: **meaning, sense, signification, import, thrust, drift, gist, implication, message, essence,**
substance, point
"the significance of his remarks"

in·fer·ence

/ˈɪnf(ə)rəns/ 🔊

noun

- a conclusion reached on the basis of evidence and reasoning.
synonyms: **deduction, conclusion, reasoning, conjecture, speculation, guess, presumption,**
assumption, supposition, reckoning, extrapolation
"there should be no inference drawn from the fact that he chooses not to be a witness"
- the process of inferring something.
"his emphasis on order and health, and by **inference** cleanliness"



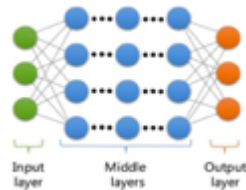
Yewno Technology

From Data to Knowledge

Machine Learning + Computational Linguistics + Graph Theory

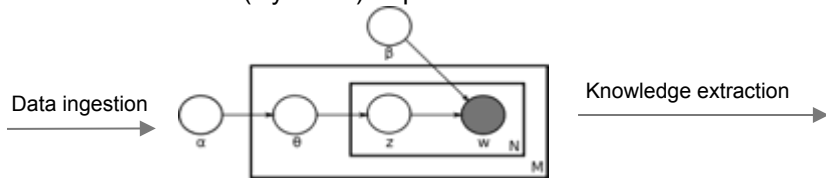
Computational linguistics analysis is undertaken on raw data to create a vast, multi-layered, multi-dimensional set of **concepts**

(Deep) Neural Nets

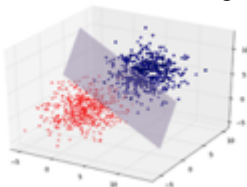


Concepts are projected onto a hierarchical **knowledge network** that represents their relations

(Dynamic) Topic Models



Stochastic Learning



Data-driven analysis is performed against the knowledge network to detect **emerging phenomena**



Yewno Approach

Concepts—not Keywords

Yewno hunts for concepts, not keywords, and identifies these as objects that carry a description and a significance

A concept is therefore not identified by its label

- What Jaguar is?
- What Quantitative Easing is?
- What Global Warming is?
- Etc., etc.



Jaguar (animal)



Jaguar (tank)



Jaguar (automobile)



Jaguar (software)



Jaguar (beverage)



Jaguar (athletic program)



Yewno Approach

What is a Concept?

A concept is an atomic unit of information, composed of:

- One or more definitions
- One or more topics



Jaguar (animal)

Definition(s)

Topic(s)

The **jaguar** S.p.A. is the largest cat in the Americas. The jaguar has a broad head and powerful jaws. Its coat is normally yellow and tan, but the color can vary from reddish brown to black.

Nature / Zoology
Nature / Wildlife

Jaguar is the luxury vehicle brand of Jaguar Land Rover, a British multinational car manufacturer with its headquarters in Whitley, Coventry, England, owned by the Indian company Tata Motors.

Topic(s)

Definition(s)



Jaguar (automobile)

Engineering / Car
Marketing /
Brands



Yewno Approach

Computational Linguistics

In using Computational Linguistics, and not its simpler relative NLP, we can take the meaning of concepts into account, which has the distinct benefits of overcoming the issues of:

- Polysemy: the coexistence of many possible meanings for a word or phrase (e.g., jaguar, java)

Yewno disambiguates very effectively by understanding the context in which words occur

- Synonymy: different terms with equivalent meanings (e.g., “Brexit”, UK “Withdrawal from the European Union”)

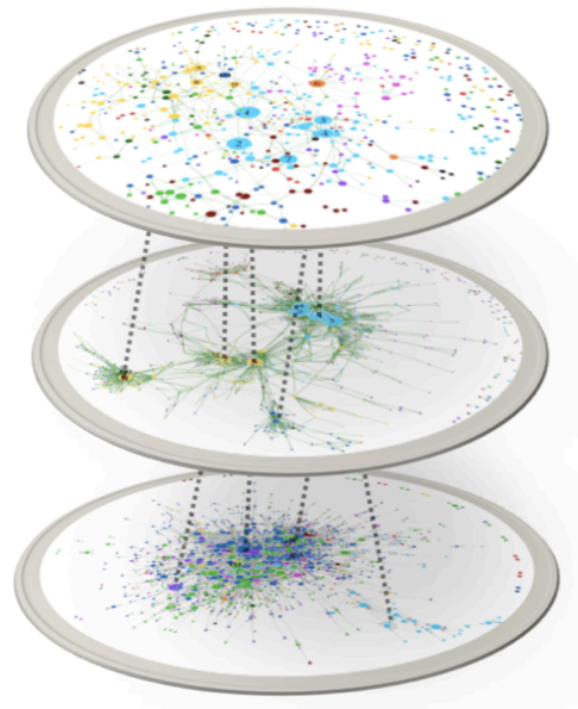
Yewno brings together terms under single concepts for effective searching



Yewno Technology

Graph of interconnected concepts

- Concepts do not occur in isolation: they form a network of interconnected concepts along several similarity dimensions
- We have developed a set of measures to assess the similarity between two concepts
- Each similarity measure implicitly induces a graph of connected concepts





Yewno Approach

A Topic-Model Approach

Yewno adopts topic models as a basic building block for advanced categorization

- Each information resource is represented as a combination of topics (categories) with different proportions
- Each topic, in turn, is represented as a weighted combination of topics

Topics are nested—a hierarchical approach

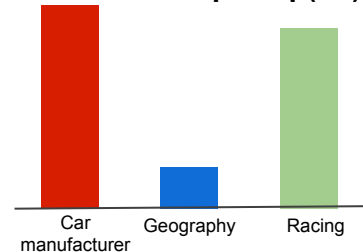
We have trained a dense hierarchical topic model

- 33 top-level classes
- 800 subclasses

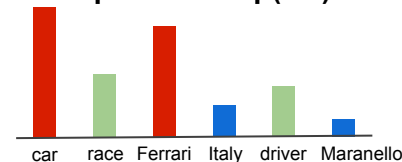
Ferrari

Ferrari S.p.A. is an Italian sports car manufacturer based in Maranello, Italy.... The company sponsored drivers and manufactured race cars. Throughout its history, the company has been noted for its continued participation in racing, especially in Formula One.

Document Topics: $p(t/d)$

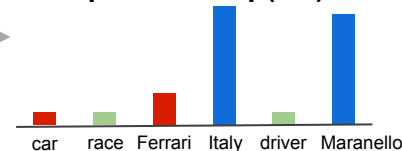


Topic Words: $p(w/t)$



Car manufacturer

Topic Words: $p(w/t)$



Geography



Yewno Unearth

A Toolkit for Publishers



Yewno Unearth Toolkit

Collection Builder

- Our unique AI tool looks across entire collections, helping expose the full value within holdings.
- We ingest full or part collections, and process the information against the 'knowledge' gained from hundreds of millions of items of content.
- Each 'item' of content is then associated with hundreds or thousands of concepts, topics and sub-topics.
- The Unearth collection builder is an easy to use web interface allowing users to search across their portfolio.

DEMO.....



Yewno Unearth Toolkit

Concept Metadata API (1 of 2)

The Yewno Concept Metadata API provides a comprehensive list of features extracted from any resource in a content holder's corpus.

- The response typically contains topics (typically 1–4), subtopics (number dependent on source data), and hundreds of concepts for each resource
- Topic and subtopic scores, indicating how central each is to the requested resource
- Normalized labels suitable for display to end users
- Can provide returns on discrete lists of resources
- Can be trained on proprietary taxonomies or specific domains



Yewno Unearth Toolkit

Concept Metadata API (2 of 2)

- Proprietary content: enhance metadata display on internal and external search platforms

The Amazon product page for 'Nanotechnology: The Future is Tiny' 1st Edition by Michael Berger. The page shows the book cover, a price of \$77.85, and various buying options. A red arrow points to the 'Add to Cart' button. The page also includes a 'More Buying Choices' section and a 'prime student' discount.

The Stanford Libraries SearchWorks article page for 'Nanotechnology: The Future is Tiny'. The page displays the book's title, author (Berger, Michael), publication date (2016-01-01), and language (English). It includes an abstract and a list of subjects, with 'Nanotechnology' highlighted by a red arrow. The details section shows the format as 'eBook' and the journal as 'Nanotechnology: The Future is Tiny'.

The Google Books search results for 'Nanotechnology: The Future is Tiny'. The results show the book's title, author (Michael Berger), and a brief description. A red arrow points to the 'Preview this book' link.

The Wiley Online Library page for 'Nanotechnology: The Future is Tiny'. The page displays the book's title, author (Michael Berger), and a brief description. It also includes a list of subjects and a 'Details' section.



Yewno Unearth Toolkit

Similarity API

The Yewno Similarity API identifies similar text resources based on overlapping concepts. The API delivers a list of conceptually similar resources for any English-language resource in a content holder's corpus.

- The response contains metadata that can be used to display the similar resources, as well as hundreds or even thousands of concepts that co-occur in the similar resources.
- The Yewno Similarity API can be deployed on both end-user sites and within internal tools.

The screenshot displays the Yewno Unearth Toolkit interface. At the top, a search bar contains the text 'brain medicine'. Below the search bar, the interface is divided into two main sections. The left section, titled 'SIMILAR DOCUMENTS', lists three documents: 'Software Engineering tutorial' (1980), 'Design Structure Matrix Methods and Applications' (2002), and 'Ultrawideband Phased Array Antenna Technology for Sensing and ...' (2002). Each document entry includes a small thumbnail image and a list of concepts with their percentages. The right section, titled 'Robot Ethics' (2013), shows the book cover and a list of topics: Technology & Engineering (65%), Robotics, Military Science, and Law (34%). Below the topics, an abstract is displayed, discussing the ethical concerns of robotics technology.

MIT Press

Search: brain medicine

Back to Results

Robot Ethics
2013
Patrick Lin, Keith Abney, George A. Bekey

Link: <https://mitpress.mit.edu/9780262299534>

ISBN: 9780262299534

TOPICS

Topic	Percentage
Technology & Engineering	65%
Robotics	
Military Science	
Law	34%

ABSTRACT

Robots today serve in many roles, from entertainer to educator to executioner. As robotics technology advances, ethical concerns become more pressing: Should robots be programmed to follow a code of ethics, if this is even possible? Are there risks in forming emotional bonds with robots? How might society -- and ethics -- change with robotics? This volume is the first book to bring together prominent scholars and experts from both science and the humanities to explore these and other questions in this emerging field. Starting with an overview of the issues and relevant ethical theories, the topics flow naturally from the possibility of programming robot ethics to the ethical use of military robots in war to legal and policy questions, including liability and privacy concerns.

Chapter 1: Title Title



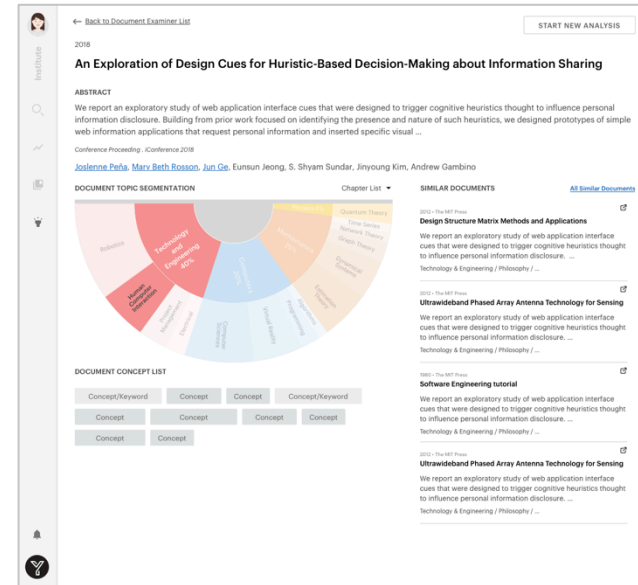
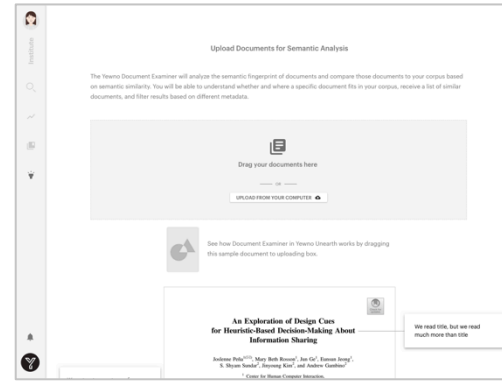
Yewno Unearth Toolkit

Document Examiner

The Yewno Document Examiner analyzes the semantic fingerprint of individual documents (journal manuscripts, book proposals, grant proposals) and compares those documents to a content holder's portfolio based on semantic similarity.

- The API facilitates understanding of whether and where a specific document fits in a corpus of content and provides a list of similar documents
- Results can be filtered based on a number of criteria

Benefits: Use for strategy and planning activities, understand how new content maps/overlaps to existing content; consider new items and their value; identify new areas not covered





Yewno Unearth Toolkit

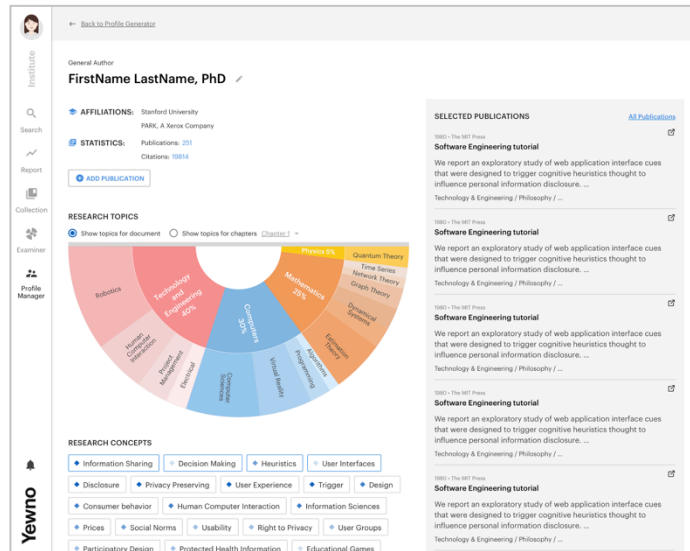
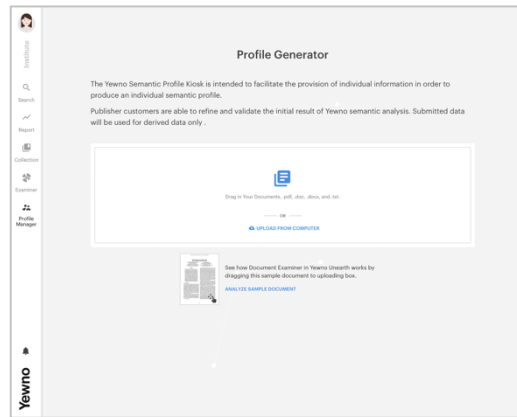
Semantic Profile Kiosk (in development)

The Yewno Semantic Profile Kiosk facilitates the provision of individual information (in the form of textual data suitable for semantic analysis) in order to produce an individual semantic profile

- Individual kiosk users can refine and validate the initial result of Yewno semantic analysis
- Submitted data will be used for derived data only (i.e., not mapped to Yewno schemas and never displayed)

Benefits: Fast, accurate, unbiased profiles kept up to date as further items are published

Strictly confidential, not for distribution without the prior consent of Yewno



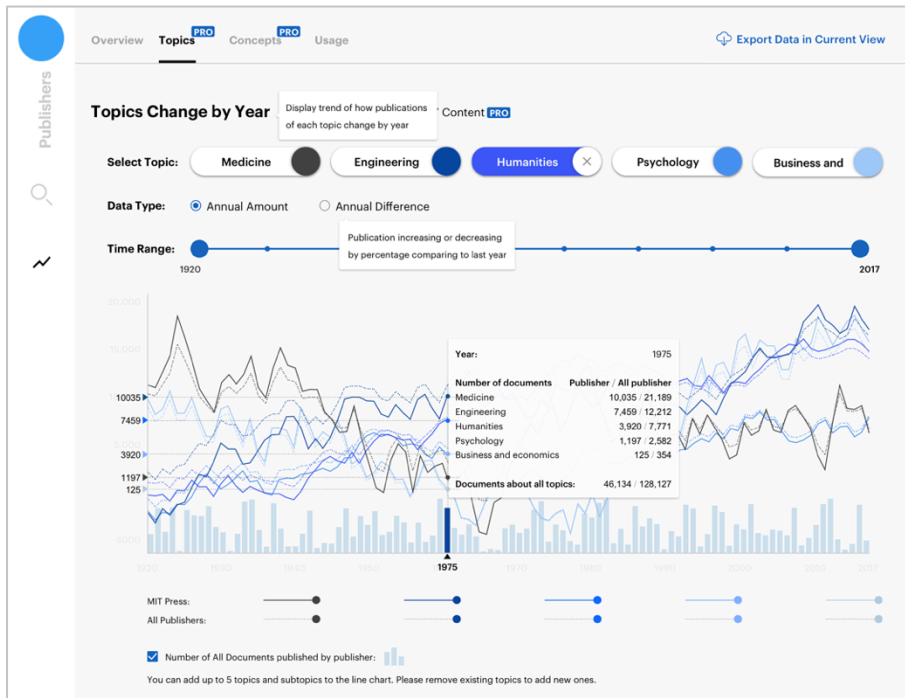


Yewno Unearth Toolkit

Roadmap: Semantic Distribution API

The Yewno Semantic Distribution API provides the semantic profile of a group of content-holder resources, based on topical selections

- Up to 5 topics can be used to generate the semantic profile
- Results can be displayed by annual amount or annual difference
- Results can be filtered by publication date
- Results can be compared to external sources (aggregated)





Yewno Unearth

Use Cases



Yewno Uneath

Use Cases and benefits

Yewno Uneath offers a wide range of use cases for publishers

Conceptual content categorization and topic hierarchy

- Granular, and dynamic content taxonomy across an entire portfolio or slices of a portfolio
- Escapes the limitations and expense of fixed taxonomies requiring manual curation
- Allows for more targeted sales and marketing
- Facilitates a more strategic approach to acquisitions
- Enables portfolio gap analysis
- Exposes metadata and cataloging errors

● Enhance end-user resource discovery

- Conceptual relatedness can be used for search-result display on e-commerce site
- Reduce turnaways, improve discovery, and increase sales on internal websites
- Improve discovery and sales on third-party e-commerce sites (Amazon, Barnes & Nobles, etc.)



Yewno Uneath

Use Cases and benefits

Yewno Uneath offers a wide range of use cases for publishers, aggregators, and organizations with unstructured content portfolios

- **Conceptual linking**
 - Document to document (e.g., books to books, journal article to book chapter)
 - Document to person (e.g., manuscript to reviewer)
 - Document to portfolio (e.g., manuscript to journal, book to series)
- **Trending concepts**
 - Expose the emergence of concepts over time
 - Produce a list of resources where the concept appears
- **Dynamic content collections**
 - Artificial intelligence + human curation
 - Target sales and marketing to specific clients

Thanks for listening!
Questions?

vicky@yewno.com



Yewno

Transforming Information to Knowledge