UNSILO

Why has the take-up of AI been so limited in academic publishing?

Michael Upshall, UNSILO
Researcher 2 Reader Conference, 25 Feb 2020

UNSILO: identifying knowledge from text

- Founded in 2012
- Specializing in Text Analytics, NLP, ML
- Working with publishers and information services
- Acquired by Cactus Communications Jan 2020

"Using UNSILO's fully automated content enrichment technology, we can identify the most descriptive concepts and phrases within any document in our content portfolio, and provide more valuable reading suggestions, even across domains with a highly variable terminology."

Jan-Erik de Boer, CIO, Springer Nature
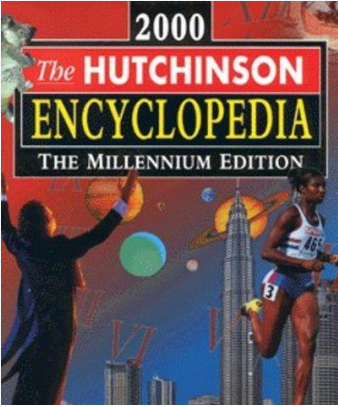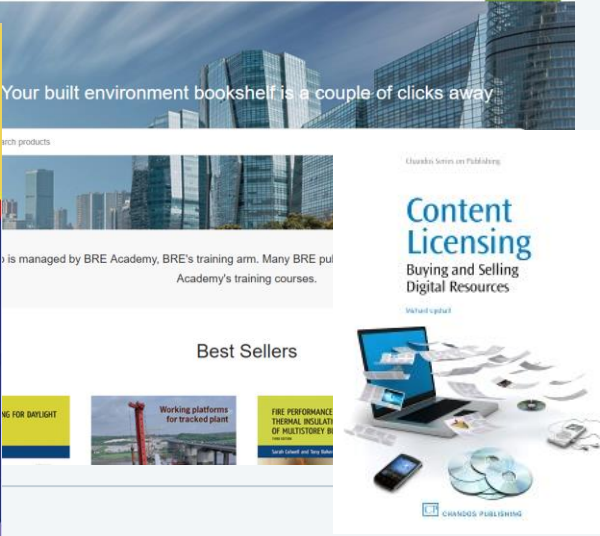
SPRINGER NATURE

Taylor & Francis Group
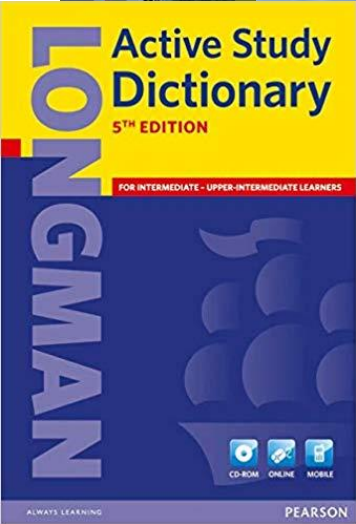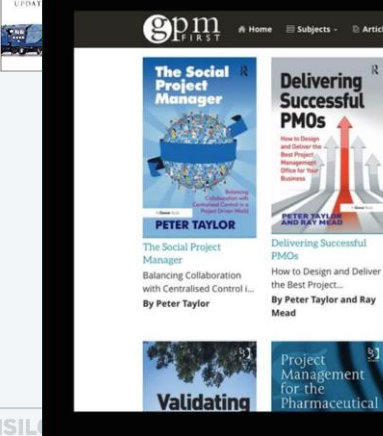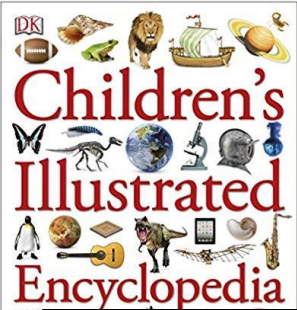an *informa* business

OECD

Infosys®

KARGER

Michael Upshall

Outline

1. Common complaints about AI
2. The technology
3. The Academic Publishing Workflow
4. Taxonomies and Classification systems: human vs machine indexing
5. Using AI for peer review
6. Using AI to build subject collections
7. Summary: How best to implement AI

Common complaints about AI

Complaints

- I don't know how it works
- It's biased
- It's not accurate enough
- I can't control it
- It might put me out of work
- It's not flexible enough

What kind of AI are we talking about?

Full Text Abstract

Pseudohyponatremia: Does It Matter in Current Clinical Practice?

Serum consists of water (93% of serum volume) and nonaqueous components, mainly lipids and proteins (7% of serum volume). Sodium is restricted to serum water. In states of hyperproteinemia or hyperlipidemia, there is an increased mass of the nonaqueous components of serum and a concomitant decrease in the proportion of serum composed of water. Thus, pseudohyponatremia results because the flame photometry method measures sodium concentration in whole plasma. A sodium-selective electrode gives the true, physiologically pertinent sodium concentration because it measures sodium activity in serum water. Whereas the serum sample is diluted in indirect potentiometry, the sample is not diluted in direct potentiometry. Because only direct reading gives an accurate concentration, we suspect that indirect potentiometry which many hospital laboratories are now using may mislead us to confusion in interpreting the serum sodium data. However, it seems that indirect potentiometry very rarely gives us discernibly low serum sodium levels in cases with hyperproteinemia and hyperlipidemia. As long as small margins of errors are kept in mind of clinicians when serum sodium is measured from the patients with hyperproteinemia or hyperlipidemia, the present methods for measuring sodium concentration in serum by indirect sodium-selective electrode potentiometry could be maintained in the clinical practice.

Applying human indexing criteria

Pseudohyponatremia: Does It Matter in Current Clinical Practice?
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3894530/
doi: 10.5049/EBP.2006.4.2.77

Key: Chemical Technique Anatomy Disease Species

Serum consists of water (93% of serum volume) and nonaqueous components, mainly lipids and proteins (7% of serum volume). Sodium is restricted to serum water. In states of hyperproteinemia or hyperlipidemia, there is an increased mass of the nonaqueous components of serum and a concomitant decrease in the proportion of serum composed of water. Thus, pseudohyponatremia results because the flame photometry method measures sodium concentration in whole plasma. A sodium-selective electrode gives the true, physiologically pertinent sodium concentration because it measures sodium activity in serum water. Whereas the serum sample is diluted in indirect potentiometry, the sample is not diluted in direct potentiometry. Because only direct reading gives an accurate concentration, we suspect that indirect potentiometry which many hospital laboratories are now using may mislead us to confusion in interpreting the serum sodium data. However, it seems that indirect potentiometry very rarely gives us discernibly low serum sodium levels in cases with hyperproteinemia and hyperlipidemia. As long as small margins of errors are kept in mind of clinicians when serum sodium is measured from the patients with hyperproteinemia or hyperlipidemia, the present methods for measuring sodium concentration in serum by indirect sodium-selective electrode potentiometry could be maintained in the clinical practice.

Unsupervised concept matching

Pseudohyponatremia: Does It Matter in Current Clinical Practice?
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3894530/
doi: 10.5049/EBP.2006.4.2.77

Key: Chemical Anatomy Disease Species

Serum consists of water (93% of serum volume) and nonaqueous components, mainly lipids and proteins (7% of serum volume). Sodium is restricted to serum water. In states of hyperproteinemia or hyperlipidemia, there is an increased mass of the nonaqueous components of serum and a concomitant decrease in the proportion of serum composed of water. Thus, pseudohyponatremia results because the flame photometry method measures sodium concentration in whole plasma. A sodium-selective electrode gives the true, physiologically pertinent sodium concentration because it measures sodium activity in serum water. Whereas the serum sample is diluted in indirect potentiometry, the sample is not diluted in direct potentiometry. Because only direct reading gives an accurate concentration, we suspect that indirect potentiometry which many hospital laboratories are now using may mislead us to confusion in interpreting the serum sodium data. However, it seems that indirect potentiometry very rarely gives us discernibly low serum sodium levels in cases with hyperproteinemia and hyperlipidemia. As long as small margins of errors are kept in mind of clinicians when serum sodium is measured from the patients with hyperproteinemia or hyperlipidemia, the present methods for measuring sodium concentration in serum by indirect sodium-selective electrode potentiometry could be maintained in the clinical practice.

UNSILO

Permuted index (keyword in context)

therefore both direct potentiometry and **indirect potentiometry** are currently used for Na+ testing in blood
remain normal In severely ill patients **indirect potentiometry** commonly leads to relevant errors in Na+ estimation
show a disagreement between direct and **indirect potentiometry** which is ≥4 mmol/L mostly spuriously elevated Na+
awareness of the poor performance of **indirect potentiometry** in some clinical settings is crucial for the
Whereas only two flame photometry and **indirect potentiometry** of the three current methods available for measuring
Levels of DNa in blood gas panels or **indirect potentiometry** INa in metabolic panels run on chemistry analyzers
second kind Experimental results for the **indirect potentiometry** of a number of amino acids and proteins are then
there are differences between direct and **indirect potentiometry** Moreover the activities measured cannot be extrapolated
and severe hyponatremia as measured by **indirect potentiometry** She was initially intensively treated for hyponatremia
rare modifications are monitored by an **indirect potentiometry** method The method is based on adding an electron
serum sodium 139 mmol/L as measured by **indirect potentiometry** This case shows that extreme hypercholesterolemia
Cl- concentration was measured with **indirect potentiometry** All analyses were completed by personnel blinded
from 3 mm DBS punches were measured via **indirect potentiometry** using the Roche Cobas 8000 routine chemistry analyzer
one single electrolyte exclusion effect **indirect potentiometry** ion-selective electrode monoclonal gammopathy multiple
in levels of potassium were measured by **indirect potentiometry** RESULTS In the total sample there was no association
is determined by flame photometry or **indirect potentiometry** but not when determined by ultracentrifugation and
meter against plasma [K+] determined by **indirect potentiometry** revealed a linear relationship that was almost
routine chemical analysis typically use **indirect potentiometry** involving the dilution of samples to measure sodium
comparison of the results obtained by **indirect potentiometry** with those obtained by other techniques we can
urine samples at baseline 1997-1998 by **indirect potentiometry** UCaE was treated as a continuous variable and a
requiring little maintenance The use of **indirect potentiometry** offers results consistent with those obtained with
determination in aqueous samples using **indirect potentiometry** and the results obtained are compared with those
especially in hypernatremic patients **indirect potentiometry** with the Technicon RA-1000 yielded intermediate
chloride concentration was determined by **indirect potentiometry** using a potassium ion selective electrode
four routine analytical methods/systems **indirect potentiometry** direct potentiometry and enzymatic assay were compared
flouride and chloride were determined by **indirect potentiometry** whereas copper magnesium and phosphate were determined

The five tribes of AI
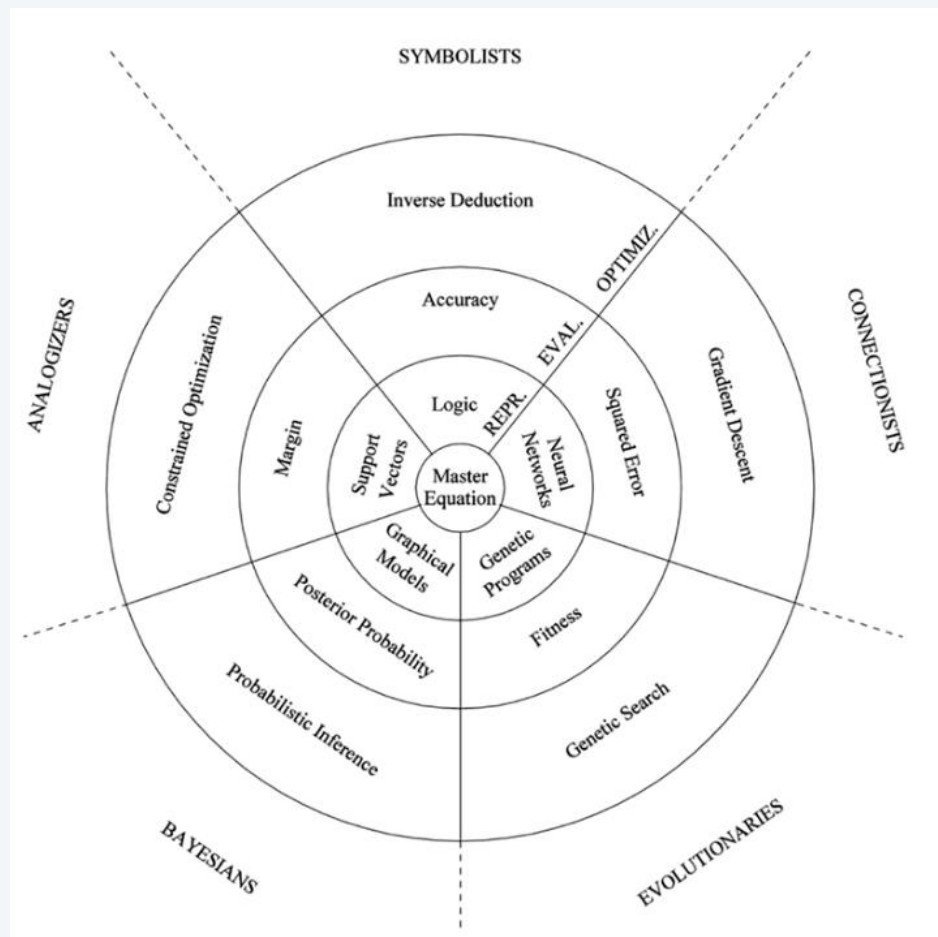
Domingos, *The Master Algorithm*, 2015

Supervised AI: Using a training set

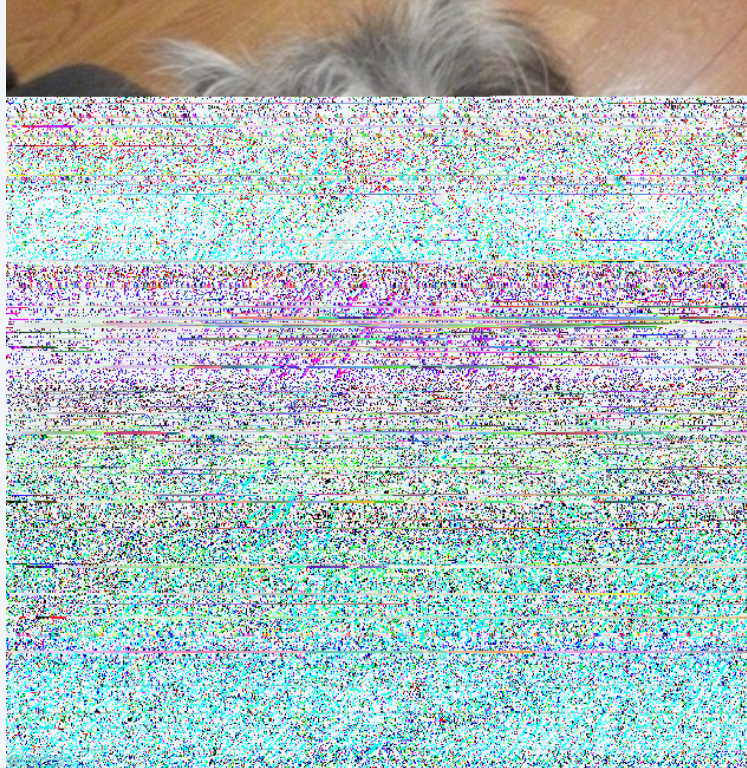

Training data:

$x_1 =$ $x_2 =$ $x_3 =$

Training identity: $y_1 = [0,1,0]$ $y_2 = [0,0,1]$ $y_3 = [1,0,0]$

Test data:

Example output: $\hat{y}_1 = [0.04, 0.82, 0.14]$ $\hat{y}_2 = [0.39, 0.26, 0.35]$ $\hat{y}_3 = [0.68, 0.09, 0.23]$

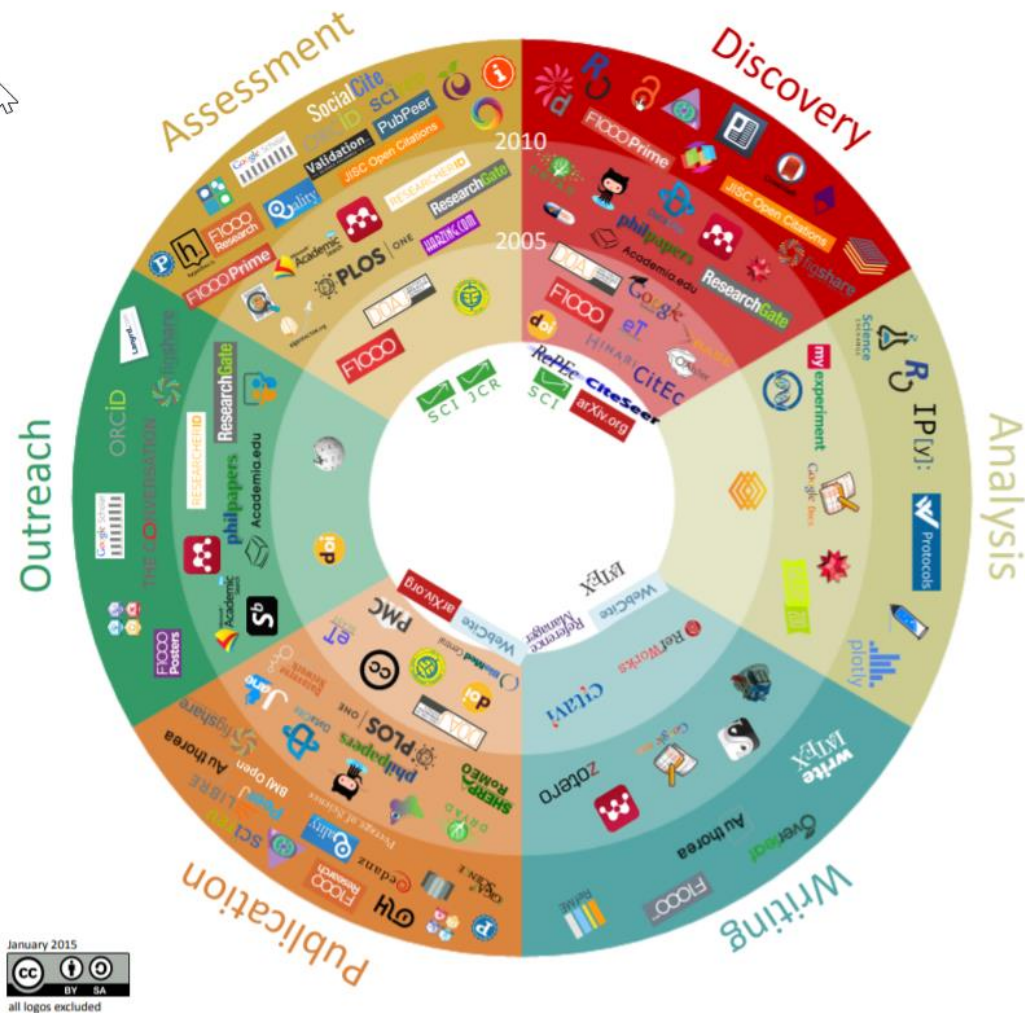$y = [p_{cat}, p_{dog}, p_{goat}]$ where p_{cat} is the probability the image shows a cat, etc...

Limitations of training sets

The Academic Publishing Workflow

The wheel of scholarship

Kramer and Boesman, 2015

Scholarly publishing needs a scalable solution

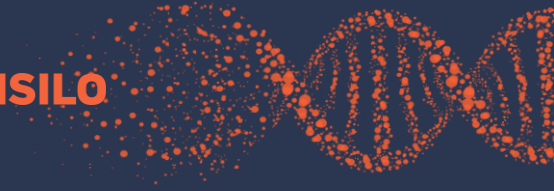**50m**   Scholarly science articles (1665-2009)

3,000   New science articles published per day

24,000   Science journals

5,000   Papers presented at the 2018 ASCO Conference

Vocabularies, Taxonomies, Ontologies

Controlled Vocabulary lists
- LOCSH, MeSH

Taxonomies
- Linnaean Classification

Ontologies
- SNOMED, UMLS

Standards Frameworks to link all the above

Hand-curated subject headings (MeSH)

Lists of synonyms

MeSH ▾   [MeSH ▾]   heart attack

Create alert   Limits   Advanced

Full ▾

Myocardial Infarction
NECROSIS of the MYOCARDIUM caused by an obstruction of the blood supply to the heart (CORONARY CIRCULATION).
Year introduced: 1979

PubMed search builder options
Subheadings:

☐ analysis
☐ anatomy and histology
☐ blood
☐ cerebrospinal fluid
☐ chemically induced
☐ chemistry
☐ classification
☐ complications
☐ congenital
☐ diagnosis
☐ diagnostic imaging
☐ diet therapy
☐ drug therapy
☐ economics

☐ embryology
☐ enzymology
☐ epidemiology
☐ ethnology
☐ etiology
☐ genetics
☐ history
☐ immunology
☐ legislation and jurisprudence
☐ metabolism
☐ microbiology
☐ mortality
☐ nursing
☐ organization and administration

☐ parasitology
☐ pathology
☐ physiology
☐ physiopathology
☐ prevention and control
☐ psychology
☐ radiotherapy
☐ rehabilitation
☐ statistics and numerical data
☐ surgery
☐ therapy
☐ urine
☐ veterinary
☐ virology

☐ Restrict to MeSH Major Topic.
☐ Do not include MeSH terms found below this term in the MeSH hierarchy.

Tree Number(s): C14.280.647.500, C14.907.585.500
MeSH Unique ID: D009203
Entry Terms:

- Infarction, Myocardial
- Infarctions, Myocardial
- Myocardial Infarctions
- Cardiovascular Stroke
- Cardiovascular Strokes
- Stroke, Cardiovascular
- Strokes, Cardiovascular
- Heart Attack
- Heart Attacks
- Myocardial Infarct
- Infarct, Myocardial
- Infarcts, Myocardial
- Myocardial Infarcts

UNSILO

Limitations of taxonomies

- Expert curator required
- Humans don't agree
- Not granular enough
- Require constant maintenance

Automatic concept expansion

Class settings for heart disease

Cancel   Save

General   Concepts (1)   Gold Documents   Filters

All active concepts ▾   Add concept...

Latest   Alphabetical

↑   Heart Disease ✕

Heart Disease   ✕

Related concepts

Cardiac Disease +   Ischemic Heart Disease +   Ischemic Heart Disease +

Valvular Heart Disease +   Congestive Heart Failure +

Atherosclerotic Heart Disease +   Cardiac Failure +

Rheumatic Valvular Heart Disease +   Structural Congenital Heart Disease +

Cardiac Problems +   Cardiac Disorders +   Congestive Heart Disease +
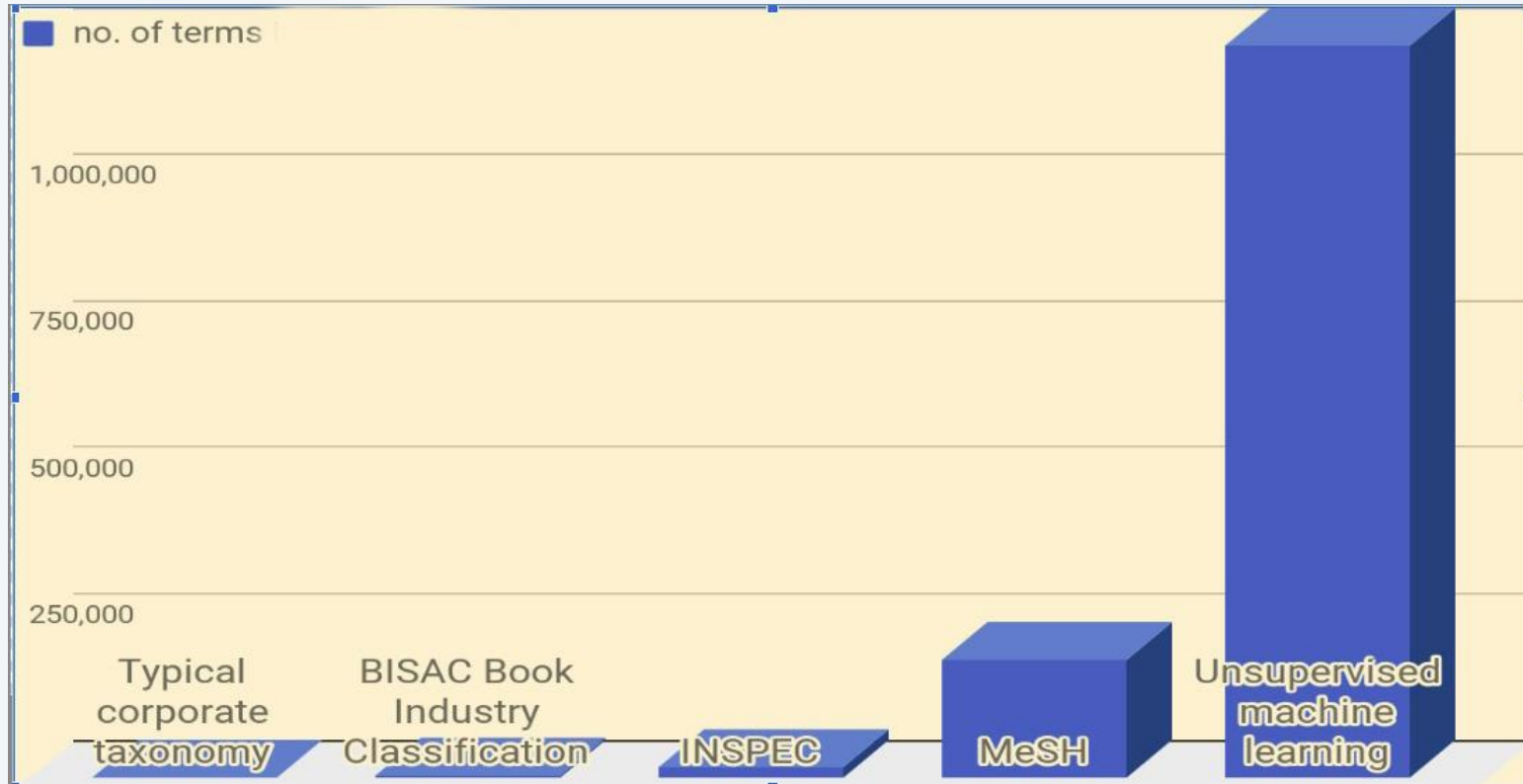
Structural Cardiac Abnormalities +   Heart Conditions +

Myocardial Insufficiency +   Myocardial Disorders +

Rheumatic Valvular Disease +   Noncardiac Disease +   Cardiac Impairment +

Pulmonary Heart Disease +   ‹ SHOW LESS

1.65.0

Automatic expansion of concepts based on the corpus

Number of concepts identified

Bar chart titled "no. of terms" with y-axis values 250,000, 500,000, 750,000, and 1,000,000. Categories: Typical corporate taxonomy, BISAC Book Industry Classification, INSPEC, MeSH, Unsupervised machine learning.

Humans vs machines: Strengths and weaknesses

	Manual taxonomy	Rule-based systems	Unsupervised machine learning
Updating the system	slow	fair	immediate
Requires cataloguing expertise	yes	yes	no
Requires AI/IT expertise	no	yes	no
Add a new term	slow	quick	Real-time
Learns from new content	no	no	yes

Which approach to use?

"More recent [AI-based] content analysis approaches use more statistical and grammatical analysis, rather than … a taxonomy … There are cases where the use of a taxonomy or ontology are still appropriate, but this should no longer be the assumed starting point."

STM Association Overview of Scientific and Scholarly Publishing 2018 Report
https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf

Finding Peer Reviewers

Current peer review

- Mainly keyword-based
- In-house reviewer database risks bias
- Much of the workflow is manual
- Often 5+ reviewers contacted for each acceptance

Finding a peer reviewer

26% of US academics contacted for peer review in 2016 declined because the paper was outside their subject area

[Wiley, presentation to the ALPSP Annual Conference 2017]

Selecting a peer reviewer

Reviewer Finder ⓘ   **10 candidates**

SEARCH SETTINGS >

Main areas found in your manuscript

> Cardiorespiratory Fitness **PRIORITIZED**
> Brain Glucose Metabolism **PRIORITIZED**
> Exercise Training
> Alzheimer 's Disease
> Physical Activity
> Executive Function

Bo Hyun Kim   20   14+ 28
pons71@hanmail.net   H   🌐   📋

🏛 *Biomedical Research Institute, Pusan National University Hospital, Busan, Korea*

Cardiorespiratory Fitness   ↑
Brain Glucose Metabolism   ↑
Exercise Training

Evaluate another manuscript

Brain Glucose Metabolism, Cognition, and Cardiorespiratory Fitness Following Exercise Training in Adults at Risk for Alzheimer 's Disease

Issue title: Exercise Effects on Cognitive Function in Humans

Guest editors: **Ozioma Okonkwo and Henriette van Praag**
Article type: Research Article
Authors: Gaitán, **Julian M.a** | Boots, **Elizabeth A.a**; 1 | Dougherty, **Ryan J.a**; b; c | Oh, **Jennifer M.a**; d | Ma, Yuea | Edwards, **Dorothy F.a**; d; e | **Christian, Bradley T.a**; f | Cook, **Dane B.b**; c | Okonkwo, Ozioma C.a; d; e; *

Affiliations: [a] Wisconsin Alzheimer 's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA | [b] Department of Kinesiology, University of Wisconsin School of Education, Madison, WI, USA | [c] William S. Middleton Memorial Veterans Hospital, Madison, WI, USA | [d] Geriatric Research Education and Clinical Center, William S. Middleton Memorial Veterans Hospital, Madison, WI, USA | [e] Wisconsin Alzheimer 's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA | [f] Department of Medical Physics, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA Correspondence: [*] Correspondence to: Ozioma Okonkwo, PhD, Department of Medicine and Alzheimer 's Disease Research Center, University of Wisconsin School of Medicine and Public Health, 600 Highland Ave. J5/1M CSC MC2420, Madison, ‖ WI 53792, USA. Tel.: +1 608 265 4479; Fax: +1 608 265 3091; E-mail: ozioma@medicine.wisc.edu. Note: [1] Present address: Department of Psychology, University of Illinois at

Case study: subject collections

UN Sustainable development goals

Building collections by using concepts

Key topics and analytical content

SDG 5 GENDER EQUALITY

Uncover key analysis on gender equality

Achieve gender equality and empower all women and girls

Filters

☐ Hide country specific content

International organisations
☑ ITU   ☑ NCM   ☑ OECD
☑ CW   ☑ UN   ☑ WTO

Female Genital Mutilation
Physical Violence
Unpaid Care Work
Reproductive Rights
Family Planning
Gender Gap
Unpaid Work
Gender Stereotypes
Gender Discrimination

CW	Introduction: Violence Against Women and Girl...
OECD	Gender Inequality in West African Social Inst...
UN	Gender equality in The Sustainable Develo...
CW	Domestic Violence in Judicial Resource Bo...
UN	Everyone has the right to choose
UN	Eliminating all forms of violence against wom...
OECD	Towards greater gender equality
UN	Institutions and funding to ensure rights and...
UN	The obstacle underlying all others
UN	Realizing rights and choices for all: If not ...
OECD	Access to politics and gender - sensitive leg...
OECD	Restricted physical integrity in SIGI 201...
OECD	SIGI 2019 results in SIGI 2019 Global Rep...
OECD	Gender inequality in unpaid work
UN	Sustainable Development Goal 5: Gender equali...
CW	Definitions of Violence Against Women
UN	Overview in State of World Population 201...
UN	Gender equality in Towards Gender-Equitab...
UN	Women in Yearbook of the United Nations 2...

Sexual Harassment
Sexual Violence
Domestic Violence
Child Marriage
Reproductive Health
Gender Equality
Gender Mainstreaming
Women Candidates
Access To Technology

https://sdg-pathfinder.org/

Building collections by using concepts

https://sdg-pathfinder.org/
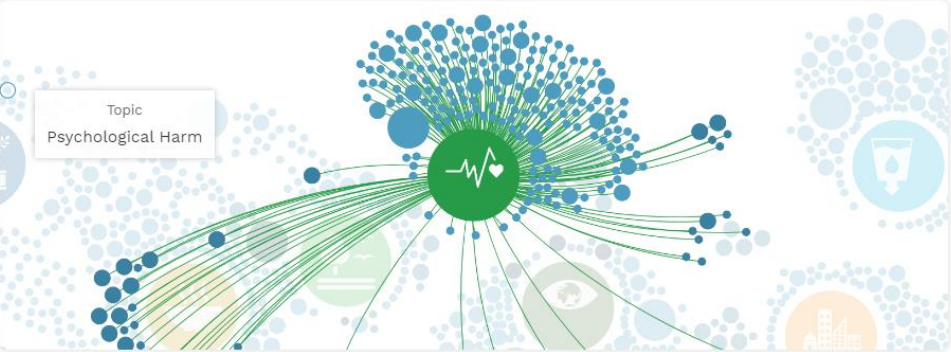
Subject Collections

Recent cardiology articles from our journals

Heart failure and cardiomyopathies
The association of ischaemic stroke in patients with heart failure without atrial flutter / fibrillation
3 October 2019   Heart

Cardiovascular medicine
Benefits and harms of lower blood pressure treatment targets: systematic review and meta-analysis of randomised placebo - controlled trials   30 September 2019   BMJ Open

Heartbeat
Heartbeat: oxygen transport close to and far from the ventricle in heart failure   26 September 2019
Heart

Review
Blood pressure and the brain: the neurology of hypertension   26 September 2019   Practical Neurology

Arrhythmias and sudden death
Long - term follow - up of normal and structural heart ventricular tachycardia catheter ablation: real - world experience from a UK tertiary centre   24 September 2019   Open Heart

https://topics.bmj.com/cardiology/

Bias
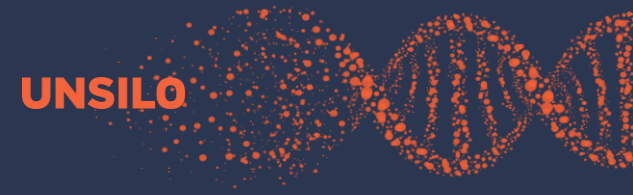
Summary

How to implement AI tools successfully

- Start with a business use case
  - I want to find related articles to my submission so that I can find an expert who can peer review it.
  - I want to find the most relevant journal for this manuscript article so I can submit it.
- Choose the most appropriate tool
  - Manual classification or automatic concept extraction?
  - Does it need technical skill and knowledge to implement?
  - Can humans contribute to the outcome?
- Identify an appropriate mant", etric
  - Citations, time, cost
- Advise on how to use and evaluate
  - Check for bias in the corpus

UNSILO

Comments and questions

Michael Upshall
mu@unsilo.ai
Head of Sales and Business Development