UNSILO

Combining human and machine approaches to classification

Michael Upshall
STM Digital Publishing, Dec 5 2017

UNSILO

Finding relevant knowledge from text

- Founded 2012
- Based Denmark
- Machine Learning, Text Analytics and NLP

insulin insensitivity in obese children
=
overweight girls may have reduced hormone responses

SPRINGER NATURE

KARGER
Medical and Scientific Publishers

Taylor & Francis Group
an informa business

OECD
BETTER POLICIES FOR BETTER LIVES

Classification

Document classification

The task is to assign a document to one or more classes or categories.

Limitations of taxonomies?

1. They change
2. They grow
3. They impose a world view
4. New ones come along
5. They are endless

Imposing a world view

London Street Lighting   *use*: **Street Lighting England London**

Thermoregulation   *use*: **Body temperature regulation**

Permafrost   *use*: **Frozen ground**

"the easiest way to find and use LCSH is to start with a 'keyword' search and then look at the Subject Headings to locate other related material."

Endless growth

New taxonomies come along

Taxonomies are endless
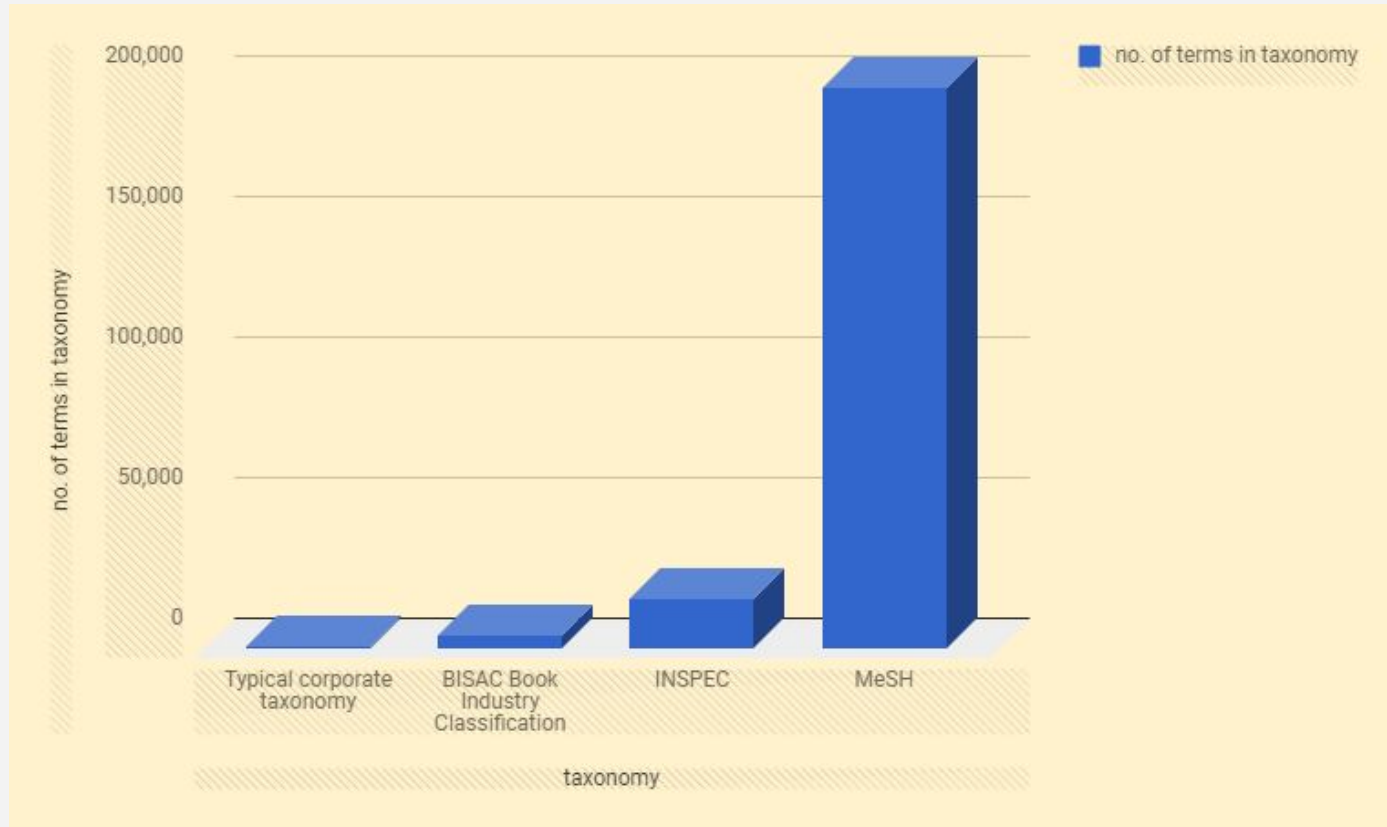
Endless Labour

Aye, and I saw Sisyphus in violent torment, seeking to raise a monstrous stone with both his hands

Homer, *Odyssey*, Book 11
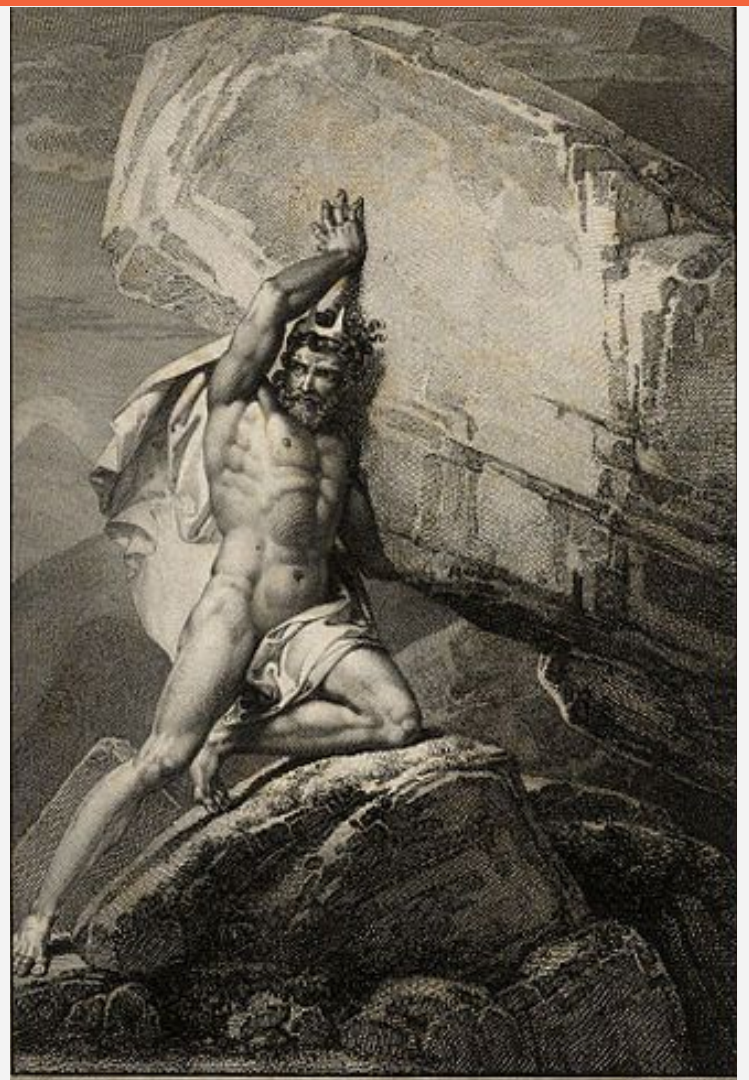
UNSILO Concept Extraction

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.

Granularity
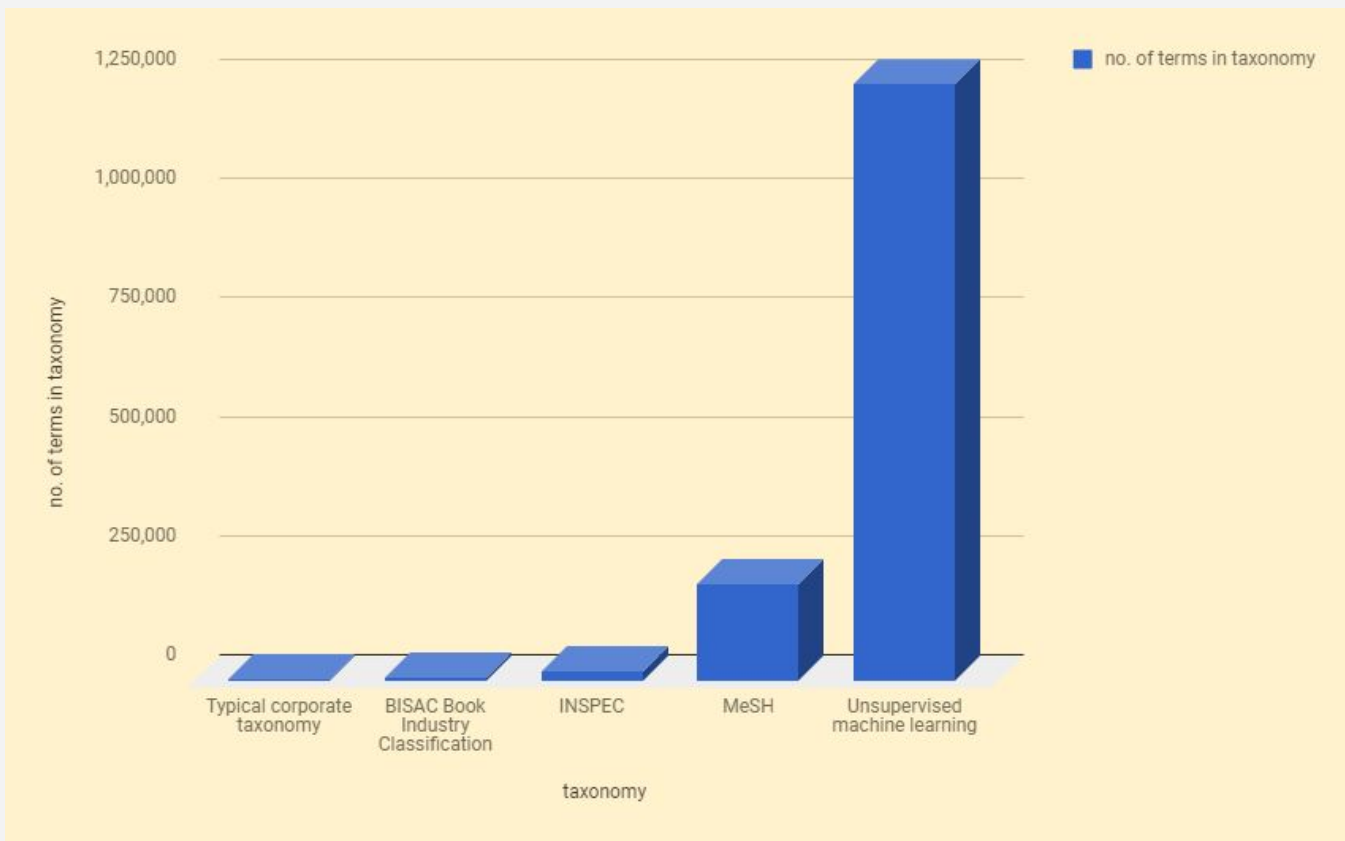
Classification methods compared

	Manual taxonomy	Rules	Unsupervised machine learning
Manual labour to maintain	high	fair	zero
Requires cataloguing expertise	yes	yes	no
Requires IT expertise	no	yes	no
Add new term after completion	slow	quick	quick
Speed of updating	slow	fair	quick
Can the system learn?	no	no	potentially
Human intervention?	yes	cumbersome	no
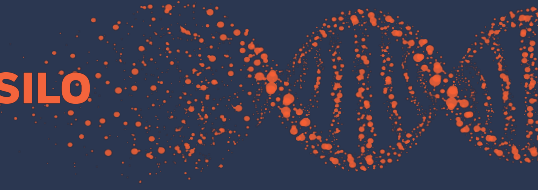
Combining human and machine capabilities

Human-machine interaction

Diabetes

A package about Diabetes with articles from past 5 years

Pending Articles (185)   Approved Articles (326)

Inspection Filter ⓘ

Adjust the level of manual inspection needed for articles that match concepts of this package.

| Auto discard | Inspect manually | Auto approve |

30% **50%**

0   10   20   30   40   50   60   70   80   90   100
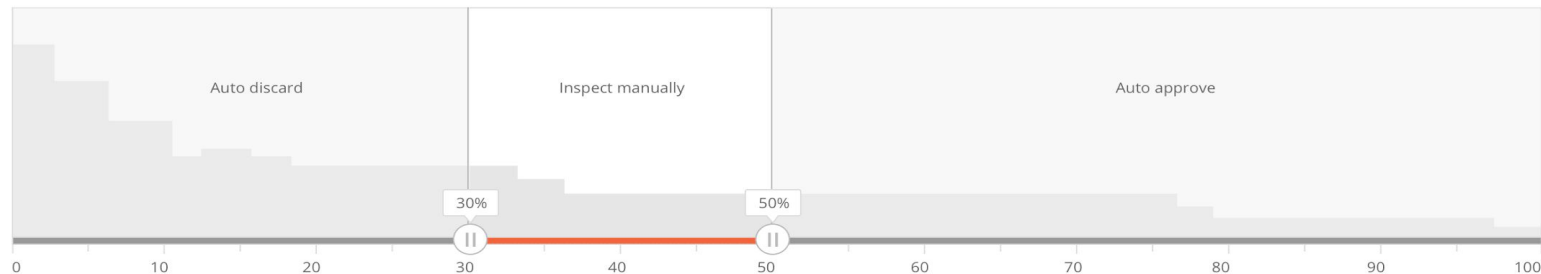
Articles ⓘ   ☐ Show entire abstract

Discarded (2765)   **Inspect manually (185)**   Auto approved (278)

☆
Mark as
Gold Article

Approve for
package?

[Yes] [No]

Diabetes mellitus and tuberculosis: programmatic management issues

Salis O, Bedir A, Ozdemir T, Okuyucu A, Alacam H in PLoS ONE Journal (2015)

The Diabetes Interactive Atlas is a recently released Web-based collection of maps that allows users to view geographic patterns and examine trends in diabetes and its risk factors over time across the United States and...

Top concepts of the article

[trifunctional antibody] [elispot assay] [peritoneal relapse] [epcam gene expression] [epcam protein]

☆
Mark as
Gold Article

A Rare Case of Congenital Diabetes Insipidus

Salis O, Bedir A, Ozdemir T, Okuyucu A, Alacam H in Frontiers in Medicine (2015)

The Diabetes Interactive Atlas is a recently released Web-based collection of maps that allows users to view geographic patterns and examine trends in diabetes and its risk factors over time across the United States and...

New collection

Search for a class, subclass or collection

Taxonomy

⊕ Anatomy

⊕ Organisms

⊕ Diseases

⊕ Chemicals and Drugs

⊕ Analytical, Diagnostic and Therapeutic...

⊕ Psychiatry and Psychology

⊕ Phenomena and Processes

⊕ Disciplines and Occupations

⊕ Humanities

Type class name

Setup new class

① Select starting point　② Add input　③ Add filters

Concepts　Add concept

Type 1 Diabetes Mellitus ×

Elispot Assay ×

Peritoneal Relapse ×

Epcam Gene Expression ×

Nitric OxideIndomethacin ×

Wall Shear Stress ×

Diabetic Rat ×

Vascular Disease ×

BH4 Deficiency ×

BH4 Level ×

Trifunctional Antibody ×

Gold Documents　Add document

Comparison of skin microvascular reactivity with hemostatic markers of endothelial dysfunction and damage in type 2 diabetes

Pollock, Jonathan P. · Patel, Hardikkumar M. · Randolph, Brittney J., et al. in Physiological Reports (2014)

Effective medications are a cornerstone of prevention and disease treatment, yet only about half of patients take their medications as prescribed, resulting in a common and costly public health challenge for the US health care system...

Back

Next

New collection

Search for a class, subclass or collection

⊖ **Taxonomy**

⊕ Anatomy

⊕ Organisms

⊕ Diseases

⊕ Chemicals and Drugs

⊕ Analytical, Diagnostic and Therapeutic...

⊕ Psychiatry and Psychology

⊕ Phenomena and Processes

⊕ Disciplines and Occupations

⊕ Humanities

Type class name

Setup new class

① Select starting point   ② Add input   ③ Add filters

Paste pararaph

Developmental dyslexia affects almost 10 % of school-aged children and represents a significant public health problem. Its etiology is unknown. The consistent presence of phonological difficulties combined with an inability to manipulate language sounds and the grapheme – phoneme conversion is widely acknowledged. Numerous scientific studies have also documented the presence of eye movement anomalies and deficits of perception of low contrast, low spatial frequency, and high frequency temporal visual information in dyslexics. Anomalies of visual attention with short visual attention spans have also been demonstrated in a large number of cases. Spatial orientation is also affected in dyslexics who manifest a preference for spatial attention to the right. This asymmetry may be so pronounced that it leads to a veritable neglect of space on the left side. The evaluation of treatments proposed to dyslexics whether speech or oriented towards the visual anomalies remains fragmentary. The advent of new explanatory theories, notably cerebellar, magnocellular, or proprioceptive, is an incentive for ophthalmologists to enter the world of multimodal cognition given the importance of the eye 's visual input.
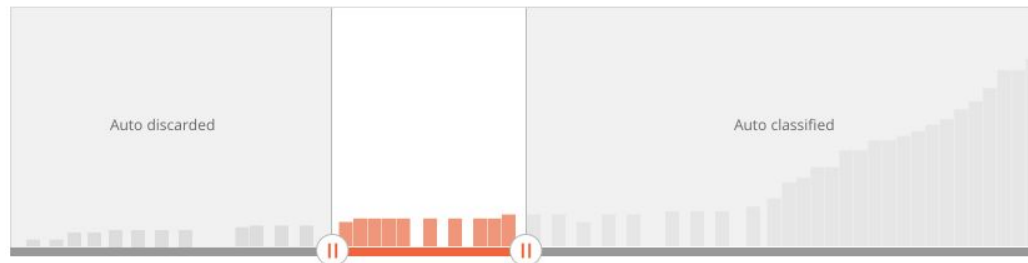
Back

Next

New collection

Taxonomy

Chemistry classes

Subjects on website

Book themes

Add collection

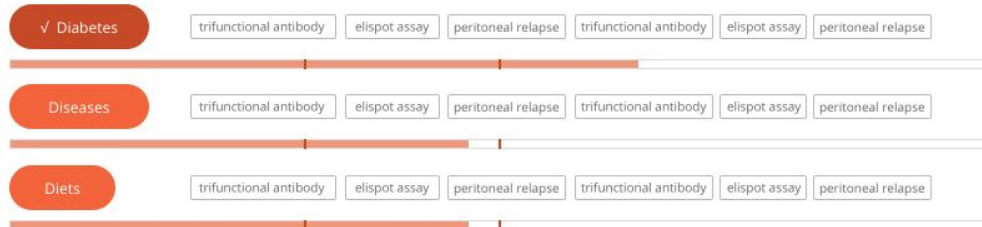## Adjust collection settings ⓘ

Auto discarded

Auto classified

Documents that need manual inspection (10)

Improving diabetes medication adherence: successful, scalable interventions

Salis O, Bedir A, Ozdemir T, Okuyucu A, Alacam H in PLoS ONE Journal (2015)

Effective medications are a cornerstone of prevention and disease treatment, yet only about half of patients take their medications as prescribed, resulting in a common and costly public health challenge for the US health care system...

Most likely belongs to one of these classes

√ Diabetes	trifunctional antibody	elispot assay	peritoneal relapse	trifunctional antibody	elispot assay	peritoneal relapse

Diseases	trifunctional antibody	elispot assay	peritoneal relapse	trifunctional antibody	elispot assay	peritoneal relapse

Diets	trifunctional antibody	elispot assay	peritoneal relapse	trifunctional antibody	elispot assay	peritoneal relapse

UNSILO

Finding relevant knowledge from text