



**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



# 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

USE Relationship quality  
**Quality of life** (*May Subd Geog*)  
[HN25]  
UF Life, Quality of  
BT Economic history  
Human ecology  
Life  
Social history  
RT Basic needs  
Human comfort  
Social accounting  
Work-life balance  
NT Cost and standard of living  
Economic indicators  
Health status indicators  
Quality of work life  
Relationship quality  
Social indicators  
Well-being

## —Religious aspects

### —Buddhism, [Christianity, etc.]

Quality of life damages  
USE Hedonic damages  
Quality of marine sediments  
USE Marine sediments—Quality  
Quality of marriage  
USE Marital quality  
**Quality of products** (*May Subd Geog*)  
[HF5415.157]  
UF Product quality  
BT Consumer protection  
RT Quality control  
Workmanship  
SA *subdivision* Quality under individual animals  
and groups of animals and individual plants  
and groups of plants, e.g. Cattle—Quality;  
Corn—Quality  
NT Beef—Quality  
Dried fruit—Quality  
Eggs—Quality  
Feeds—Quality  
Food—Quality  
Frozen fish—Quality  
Fuelwood—Quality  
Goat meat—Quality  
Horticultural products—Quality  
Kangaroo meat—Quality  
Machine-readable bibliographic data—Quality  
Meat—Quality  
Milk—Quality  
Online library catalogs—Quality  
Pork—Quality  
Poultry products—Quality  
Product obsolescence  
Wood—Quality

USE Camassia quamash  
Quamash root  
USE Camassia quamash  
Quamasia  
USE Camases  
Quamo Post Office Building (Averill Park, N.Y.)  
USE Major George Quamo Post Office Building  
(Averill Park, N.Y.)  
Quamoclit  
USE Ipomoea  
Quần đảo Đông Sa  
USE Pratas Islands  
Quần Đảo Hoàng Sa  
USE Paracel Islands  
Quan Dao Trờng Sa  
USE Spratty Islands  
**Quan zhen jiao** (*May Subd Geog*)  
[BL1943.C55]  
UF Ch'üan chen chiao [*Former heading*]  
Complete Realization school (Sect)  
Complete Truth school (Sect)  
BT Sects—China  
**Quan zhen jiao in art** (*Not Subd Geog*)  
**Quannah Parker Lake Dam (Okla.)**  
UF Parker Lake Dam (Okla.)  
BT Dams—Oklahoma  
Quannapowitt Lake (Mass.)  
USE Quannapowitt, Lake (Mass.)  
Quannapowitt Pond (Mass.)  
USE Quannapowitt, Lake (Mass.)  
**Quander family** (*Not Subd Geog*)  
UF Amkwandoh family  
Kwandoh family  
Quanders family  
Quando family  
Quandoe family  
Quonder family  
Quanders family  
USE Quander family  
Quando family  
USE Quander family  
Quandoe family  
USE Quander family  
**Quandt family** (*Not Subd Geog*)  
UF Quant family  
Quante family  
Quantz family  
Quandt-Haus (Bad Homburg von der Höhe, Germany)  
USE Herbert-Quandt-Haus (Bad Homburg von der  
Höhe, Germany)  
Quangen  
USE Kwongan  
Quangtri (Southeast Asian people)  
USE Bru (Southeast Asian people)  
Quangtri Văn Kieu language

Quashing on risk management, cost of capital, and  
sales projections.  
UF Analysts, Quantitative  
Quants (Quantitative analysts)  
BT Computer scientists  
Mathematicians  
Quantitative business analysis  
USE Industrial management—Mathematical models  
Management science  
Quantitative chemical analysis  
USE Chemistry, Analytic—Quantitative  
**Quantitative easing (Monetary policy)** (*May Subd Geog*)  
UF QE (Monetary policy)  
Queassing (Monetary policy)  
BT Banks and banking, Central  
Monetary policy  
**Quantitative genetics**  
[QH452.7]  
UF Quantitative inheritance  
BT Genetics  
Quantitative inheritance  
USE Quantitative genetics  
**Quantitative magnetic resonance**  
[QH324.9.Q35 (*Biology (General)*)]  
UF QMR (Quantitative magnetic resonance)  
BT Body composition—Measurement—Technique  
Nuclear magnetic resonance  
Quantitative methods (Research)  
USE Quantitative research  
Quantitative microscopy  
USE Stereology  
**Quantitative research** (*May Subd Geog*)  
UF Data analysis (Quantitative research)  
Exploratory data analysis (Quantitative  
research)  
Quantitative analysis (Research)  
Quantitative methods (Research)  
BT Research  
NT Webometrics  
Quantitative structure-activity relationships  
(Biochemistry)  
USE QSAR (Biochemistry)  
Quantitative taxonomy  
USE Numerical taxonomy  
Quantities, Absolute  
USE Physical constants  
Quantities, Imaginary  
USE Numbers, Complex  
Quantities, Physical  
USE Physical constants  
**Quantity (Philosophy)**  
[B105.Q34]  
BT Categories (Philosophy)  
Philosophy  
Quantity (Phonetics)

# New taxonomies come along

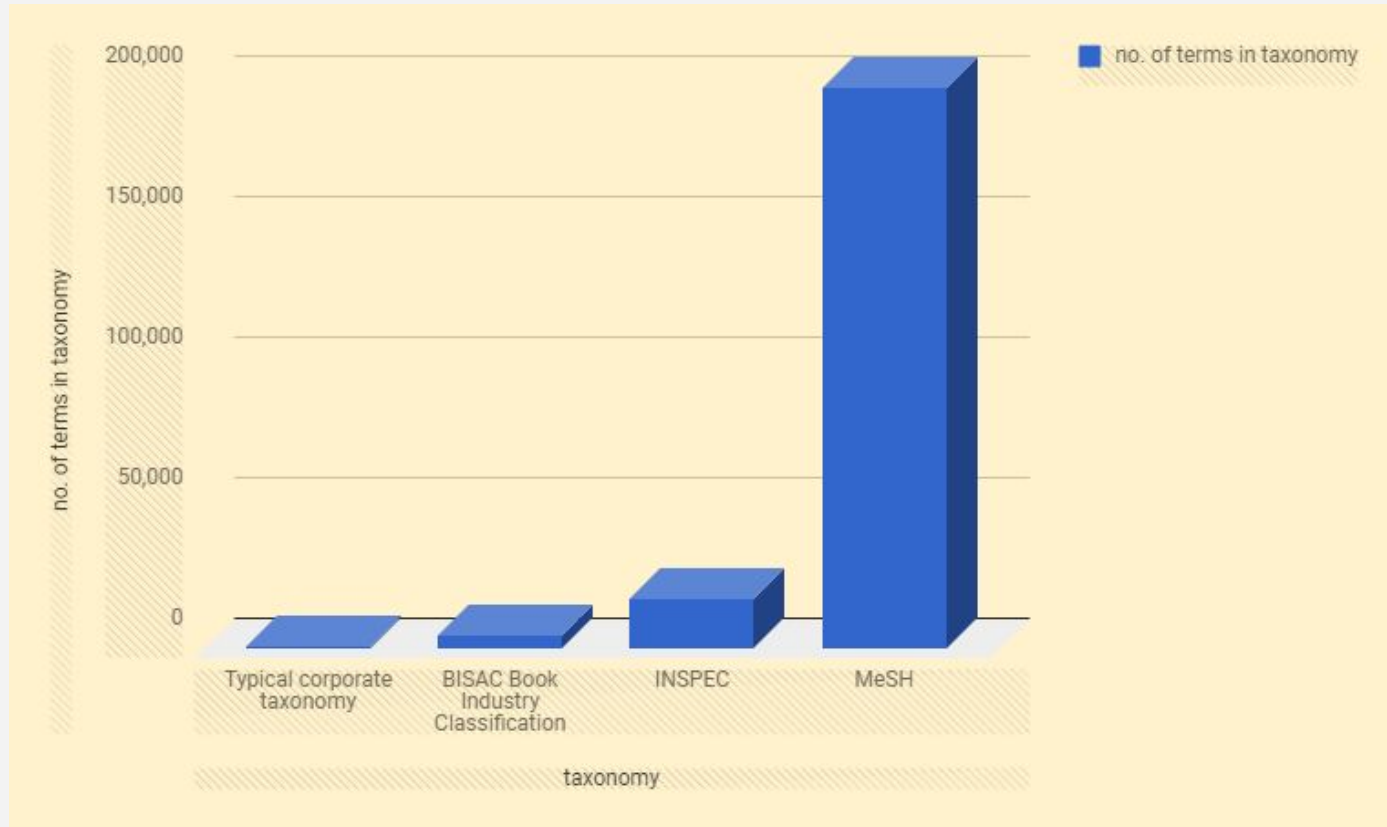


**SUSTAINABLE DEVELOPMENT GOALS**

**17 GOALS TO TRANSFORM OUR WORLD**



# Taxonomies are endless

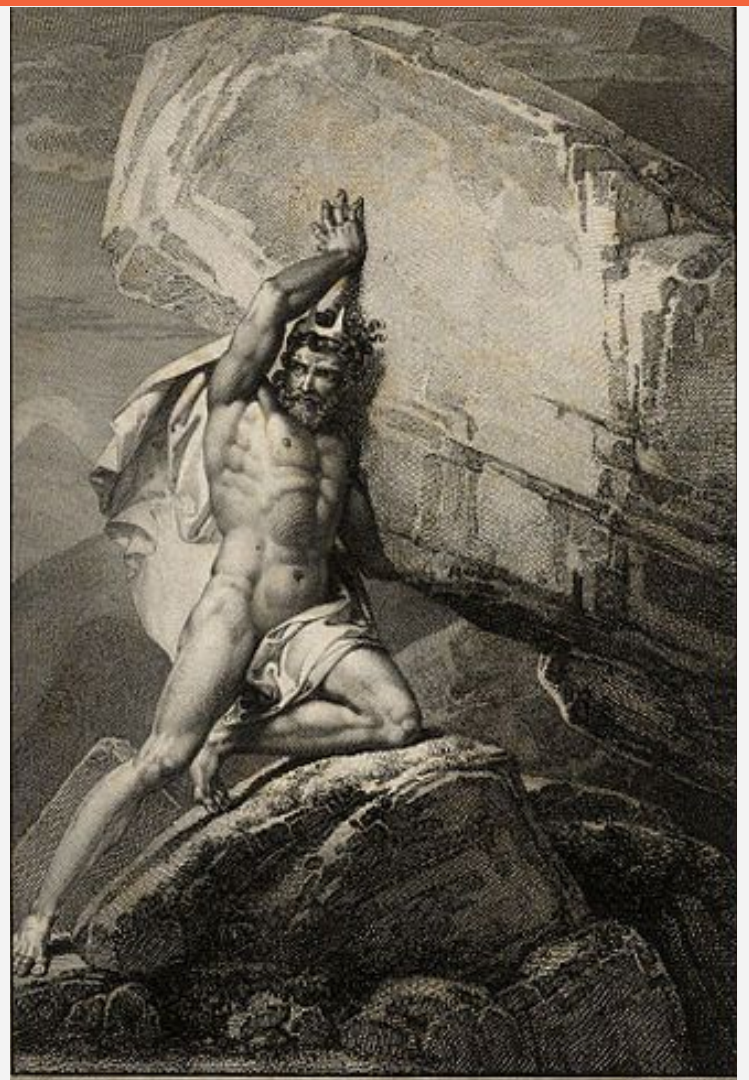




# Endless Labour

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?**

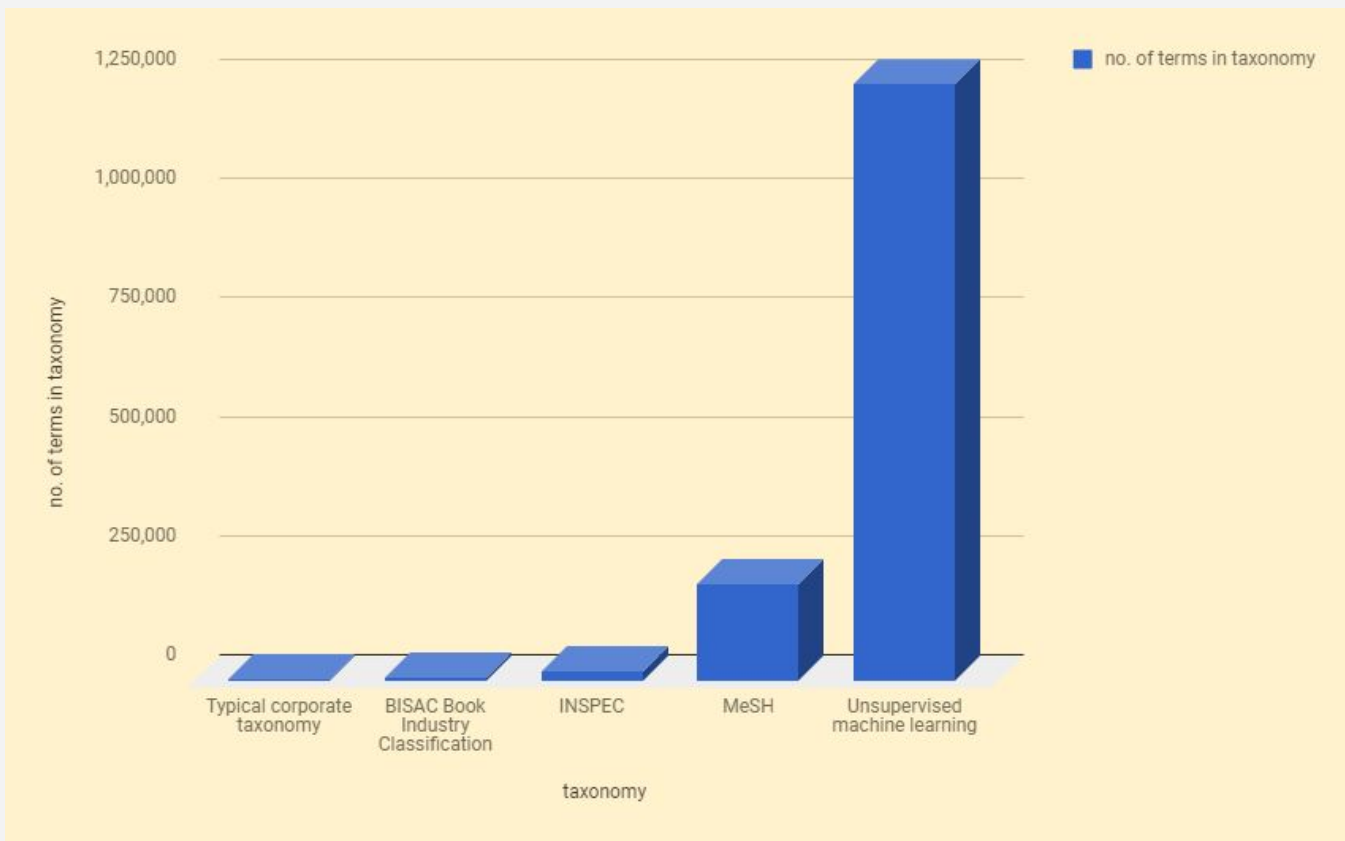
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3894530/>

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.

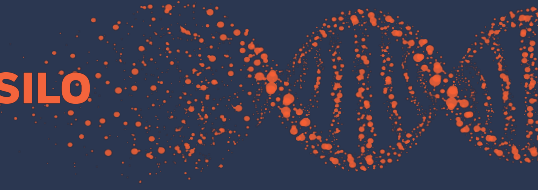
# Granularity



# Classification methods compared

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

UNSILO



# Combining human and machine capabilities

# Human-machine interaction



# Diabetes

A package about Diabetes with articles from past 5 years

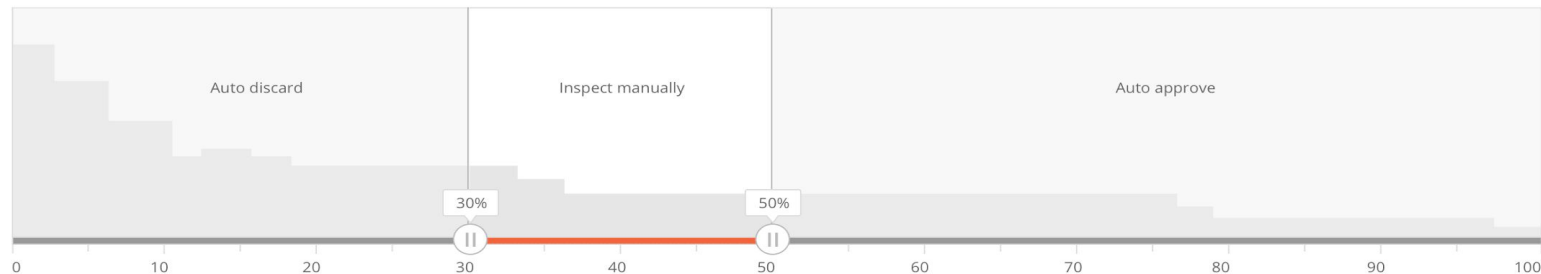
Edit package settings

Pending Articles (185)

Approved Articles (326)

## Inspection Filter ?

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



## 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...



- Taxonomy

+ Anatomy

+ Organisms

+ Diseases

+ Chemicals and Drugs

+ Analytical, Diagnostic and Therapeutic...

+ Psychiatry and Psychology

+ Phenomena and Processes

+ Disciplines and Occupations

+ Humanities

## 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 Oxide/Endomethacin ✕
- Wall Shear Stress ✕
- Diabetic Rat ✕
- Vascular Disease ✕
- BH4 Deficiency ✕
- BH4 Level ✕
- Trifunctional Antibody ✕

Back

## 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...

Next





- Taxonomy

+ Anatomy

+ Organisms

+ Diseases

+ Chemicals and Drugs

+ Analytical, Diagnostic and Therapeutic...

+ Psychiatry and Psychology

+ Phenomena and Processes

+ Disciplines and Occupations

+ Humanities

## Setup new class



1

Select starting point

2

Add input

3

Add filters

## Paste paragraph

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

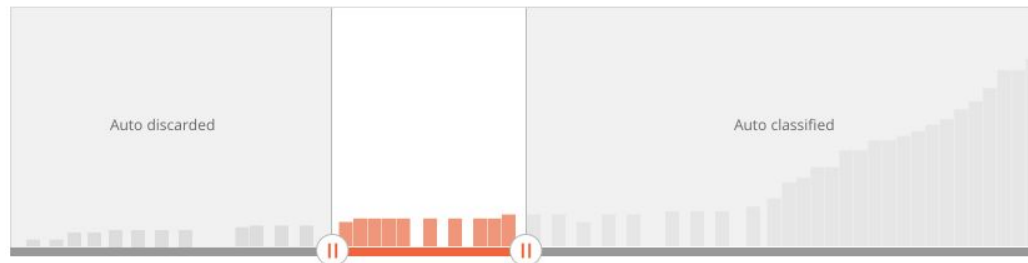
Next



[+ Chemistry classes](#)[+ Subjects on website](#)[+ Book themes](#)[+ Add collection](#)

## Taxonomy

Adjust collection settings ⓘ



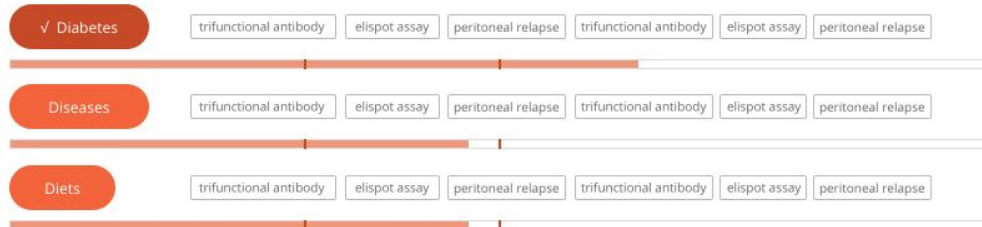
### Documents that need manual inspection (10)

#### 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





**UNSILO**

**Finding relevant knowledge from  
text**

STM Digital Publishing 2017

Michael Upshall [mu@unsilo.ai](mailto:mu@unsilo.ai)