UNSILO White Paper

Comparing UNSILO concept extraction to leading NLP cloud solutions

By Mario Juric, Head of R&D at UNSILO, and Mads Rydahl, CVO at UNSILO.

Machine learning and artificial intelligence tools are promoted as solutions to some of mankind's hardest challenges. But in reality, novel technology solutions are often either accepted or rejected based on highly subjective judgement, either because they are "black box" solutions that provide little transparency, or because no one has the resources to properly evaluate the quality or accuracy of a proposed solution. Machine Learning can be applied to the same problem in many ways, and two different service providers may even apply the same methods and return different results. How can we meaningfully compare the results of machine learning tools from different providers? In this paper we provide an overview of the machine learning techniques used by UNSILO, and compare the output of the UNSILO Concept Extraction Service to that of other leading concept extraction tools.

Background

Although machine learning and artificial intelligence tools can be used to solve a number of different tasks that were previously the exclusive domain of Subject Matter Experts (SMEs), they do not "understand" knowledge like a human expert, and much less are these algorithms capable of qualitative evaluation of subject matter without significant human engineering effort towards extrapolating a coherent model of all possible contributing factors.

Like most natural-language analytics providers, UNSILO uses a combination of probabilistic Natural Language Processing (NLP), structured knowledge in the form of dictionaries, ontologies, and thesauri, hard-coded rule sets, and adaptive machine learning to determine the most important elements in text, and facilitate the development of features like document similarity, reader interest profiles, and trend analysis.

Methodology

For this White Paper, the UNSILO Concept Extraction API was compared with the most widely adopted concept extraction services available today; the Google Cloud Natural Language API, The Microsoft Cognitive Services Text Analytics API, and the IBM Watson Alchemy Language API. To test performance across a variety of different subjects and terminologies, we randomly selected four scholarly articles from four different domains: Nanotechnology, Biomedical Science, Computer Science, and Food Science and Nutrition.

The full text of each article was submitted to each of the four designated API services, and from each service, the top 20 concepts were examined according to a set of qualitative criteria and categorized into four classes, each with a different scoring weight. The qualitative criteria were: a) Relevance to the subject matter of the article, b) Specificity and unambiguity, c) Syntactic completeness, and finally d) Uniqueness; whether a concept is a synonym of another concept in the same set.

Based on these qualitative criteria, each concept was assigned to one of four classes, and a corresponding point score was awarded, leading to an aggregated document evaluation score, calculated as the sum total of the class score of the top 20 concepts. For example, correctly identified one-gram ontology terms like "KNN" and "Vitamin D" were classified as "Relevant broad Concepts", which contribute one point to the the document evaluation score, while longer phrases with an unambiguous meaning that are in common use within the domain were classified as "Relevant Precise Concepts". Duplicate concepts and concepts that were deemed synonymous with another concept in the same set were classified as "Irrelevant or Redundant", as were concepts with no connection to the subject matter, including names and geolocations of authors and sponsoring organizations, which should be provided as metadata properties instead.

Relevant Precise Concept	2 points
Relevant Broad Concept	1 point
Irrelevant, Redundant, Ambiguous	0 point
Fragment, Error, Noise	-1 point

Caution: In contrast to all other services, the publicly available Microsoft Cognitive Services Text Analytics API only parses the first 5K of each document. Perhaps counterintuitively, this may favour the Microsoft service, since the documents used were scholarly articles starting with an abstract of approximately 5K, in which almost any concept or phrase mentioned may be assumed to be highly relevant to the subject matter of the whole article.

Results and Analysis

Results clearly show that the UNSILO Concept Extraction API does a much better job in identifying relevant concepts in every tested domain, scoring on average 33.0 points per article compared to 14.8 points for all other services across all domains. This corresponds to an average score 122% higher than the competition. The second best score was obtained by the Microsoft Cognitive Services API, which averaged 22.5 points across all domains. See the performance summary in Table 1 below.

Most of the competing services return numerous broad ontology concepts like "HIV" or "Ceramics" or ambiguous concepts like "Study" or "Feature", which have little descriptive value, and are unsuited for classification or fingerprinting of documents.

The service output and classifications can be viewed in detail in Table 2. In particular, it should be noted that Google almost exclusively returns single terms, such as "phase", "group", and "model", while Microsoft includes more speculative phrases, like "Significant percent of HIV infected individuals" but Microsoft also has the highest percentage of noise ratio at 10% compared to zero for UNSILO.

Table 1: Concept Quality Across Academic Domains

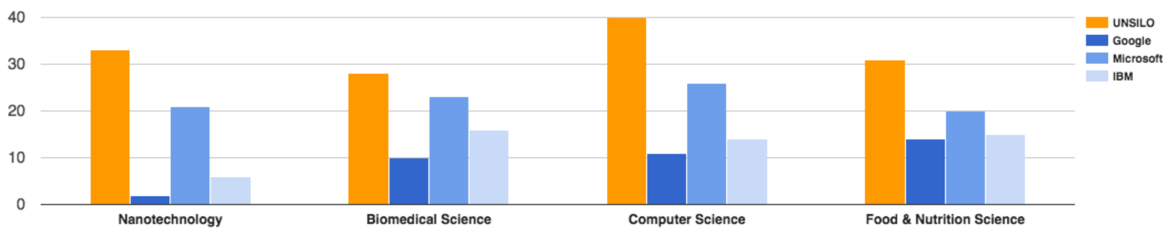

Table 2: Service Output and Classification of Concepts found in Scholarly Articles

A: Enhanced piezoelectric properties in vanadium-modified lead-free (K0.485Na0.5Li0.015)(Nb0.88Ta0.1V0.02)O3 ceramics prepared from nanopowders

http://www.sciencedirect.com/science/article/pii/S0925838814027820

UNSILO Concept Extraction API		
Common form	Score	Eval
KNN	1.00	1
KNN Ceramic	0.58	2
KNN System	0.46	2
PZT	0.44	1
Room Temperature Dielectric Constant	0.42	2
Piezoelectric Property	0.41	1
Ferroelectric	0.40	1
Piezoelectric	0.40	1
Pure KNN	0.39	2
KNN Crystal	0.37	2
High Tetragonality	0.36	2
Li-doped KNN	0.35	2
Perovskite	0.34	2
KNbO3	0.34	2
Electromechanical Coupling Factor	0.34	2
Dense KNN	0.34	2
Piezoelectric Ceramic	0.34	2
O3 Ceramic	0.33	2
Good Piezoelectric Property	0.33	0
Electromechanical Coupling	0.33	2
		33

Google Cloud Natural Language API		
Common form	Score	Eval
ceramics	1.00	1
KNNV3	1.00	-1
KNN	0.80	1
KNNV2	0.60	-1
phase	0.40	0
properties	0.40	0
ceramics	0.40	0
ceramics	0.40	0
K0	0.40	-1
Vx) O3	0.40	-1
properties	0.20	0
microstrain	0.20	1
study	0.20	0
O3 ceramic	0.20	2
stress	0.20	0
ceramics	0.20	0
materials	0.20	0
balls	0.20	0
system	0.20	0
PZT	0.20	1
		2

Microsoft Cognitive Services Text Analytics API		
Common form	Score	Eval
Enhanced piezoelectric properties	NA	2
good piezoelectric properties	NA	0
superior piezoelectric properties	NA	0
enhancement of piezoelectric properties	NA	0
piezoelectric properties of lead-free piezocerar	NA	2
alternative lead-free piezoelectric ceramics	NA	2
O3 exhibits high piezoelectric properties	NA	-1
piezoelectric applications	NA	2
O3 ceramics	NA	2
study of KNN ceramics	NA	0
Pure KNN ceramics	NA	2
sinterability of KNN ceramics	NA	2
sensitivity of properties	NA	-1
properties comparable	NA	-1
Pb-free piezoelectric materials	NA	2
piezoelectric charge constant	NA	2
NaO	NA	1
conventional sintering of KNN	NA	2
K0	NA	1
maximum piezoelectric coefficient	NA	2
		21

IBM Watson AlchemyLanguage API		
Common form	Score	Eval
KNN	1.00	1
KNNV2	0.93	-1
KNNV	0.59	-1
KNNV	0.57	-1
PZT	0.48	1
KNN	0.48	0
nanopowders	0.47	1
KNN	0.45	0
Li	0.43	1
edge technologies	0.42	0
Gibbs	0.42	0
Ps	0.40	-1
Williamson – Hall	0.38	-1
Department of Science and Technology of Indi	0.36	2
Wayne Kerr	0.35	1
Japan	0.35	0
India	0.35	1
Kakimoto	0.35	1
X-ray	0.34	1
non-stoichiometry	0.34	1
		6

B: Effects of vitamin D supplementation on the bone specific biomarkers in HIV infected individuals under treatment with efavirenz

https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-5-204

UNSILO Concept Extraction API		
Common form	Score	Eval
Vitamin D	1.00	1
Bone Mineral Density	0.35	2
Vitamin D Receptor	0.23	2
Serum CTx Concentration	0.21	2
HIV Positive Individual	0.19	2
Bone Formation Marker	0.17	2
Bone Resorption Marker	0.16	2
Bone Formation	0.16	1
Bone Formation Biomarker	0.16	0
Serum OC Level	0.16	2
Bone Biomarker	0.14	2
Serum Vitamin	0.12	0
HIV-infected Patient	0.12	1
HIV Negative Individual	0.12	2
Collagen	0.11	1
Efavirenz	0.11	1
Bone Resorption	0.11	2
Osteocalcin Concentration	0.11	2
25-OH Vitamin	0.10	1
Hepatitis C	0.10	1
		28

https://cloud.google.com/natural-language/

Google Cloud Natural Language API		
Common form	Score	Eval
HIV	1.00	1
drugs	0.12	0
individuals	0.09	0
patients	0.06	0
CTx	0.06	1
patients	0.03	0
vitamin d	0.03	1
concentrations	0.03	0
treatment	0.03	1
vitamin D	0.03	0
effects	0.03	0
study	0.03	0
supplementation	0.03	0
bone biomarkers	0.03	1
efavirenz	0.03	1
OC	0.03	1
vitamin D supplementation	0.03	1
bone biomarkers	0.03	1
individuals	0.00	0
bone	0.00	1
		10

Microsoft Cognitive Services Text Analytics API		
Common form	Score	Eval
supplementation of vitamin D	NA	2
Vitamin D deficiency	NA	2
bone biomarkers	NA	1
vitamin D deficient HIV positive individuals	NA	2
HIV positive patients	NA	2
HIV-infected patients	NA	0
catabolism of vitamin D	NA	2
IU vitamin D	NA	-1
Effects of vitamin D supplementation	NA	2
HIV Efavirenz Vitamin D Findings Background	NA	-1
HIV negative individuals	NA	2
duration of HIV infection	NA	2
bone disorders	NA	2
months of vitamin D administration	NA	0
HIV viral load	NA	2
Significant percent of HIV infected individuals	NA	0
early diagnosis of HIV infection	NA	2
bone specific biomarkers including osteocalcin	NA	0
favorable bone formation	NA	0
indicator of bone resorption	NA	2
		23

https://alchemy-language-demo.mybluemix.net/

IBM Watson AlchemyLanguage API		
Common form	Score	Eval
vitamin D	1.00	1
Vitamin D deficiency	0.66	2
HIV infection	0.63	2
efavirenz	0.46	1
bone fracture	0.30	1
ALP	0.26	1
HIV virus	0.23	1
HIV Clinic of Iranian HIV/AIDS Research Cent	0.22	0
Chronic hepatitis B	0.20	1
OC	0.19	1
Chronic hepatitis C	0.19	1
Tehran University of Medical Sciences	0.19	0
hypogonadism	0.18	1
Chicago	0.18	0
Tehran	0.18	0
drug abuse	0.18	1
adrenal insufficiency	0.18	1
USA	0.17	0
Iran	0.17	0
Osteoblast	0.17	1
		16

C: A novel density-based clustering method using word embedding features for dialogue intention recognition

https://rd.springer.com/article/10.1007/s10586-016-0649-7

UNSILO Concept Extraction API		
Common form	Score	Eval
Dialogue Act	1.00	2
Latent Dirichlet Allocation	0.55	2
Deep Neural Network	0.55	2
Word Embedding	0.48	2
Latent Semantic Analysis	0.42	2
Lexical Feature	0.37	2
Emotion Classification	0.36	2
Support Vector Machine	0.36	2
Smart Home System	0.36	2
Density-based Cluster Method	0.32	2
Non-convex Cluster	0.32	2
Density-based Cluster	0.32	2
Feature Selection Method	0.30	2
Stochastic Neighbor Embedding	0.30	2
Word Frequency Distribution	0.29	2
Natural Language Understand	0.28	2
User Utterance	0.26	2
Cluster Feature	0.26	2
Maximum Entropy Model	0.25	2
Social Network Analysis	0.25	2
		40

Google Cloud Natural Language API		
Common form	Score	Eval
features	1.00	1
features	0.56	0
clustering	0.44	1
model	0.22	0
SVM	0.22	1
corpora	0.11	1
model	0.11	0
features	0.11	0
feature	0.11	0
emotion classification	0.11	1
models	0.11	0
methods	0.11	1
words	0.11	1
corpus	0.11	0
clustering method	0.11	1
dialogue acts	0.11	1
results	0.11	0
word frequency distribution	0.11	1
words	0.00	0
Korean	0.00	1
		11

Microsoft Cognitive Services Text Analytics API		
Common form	Score	Eval
dialogue acts	NA	2
dialogue intention recognition	NA	2
word similarity	NA	2
word embedding features	NA	2
dialogue act classification	NA	2
dialogue systems	NA	2
word embedding model	NA	2
user intention analysis	NA	2
word embedding methods	NA	1
understanding user utterances	NA	2
previous classification models	NA	0
emotion classification	NA	2
Various classification models	NA	0
problem of data sparseness	NA	0
data sparseness problem	NA	0
word frequency distribution	NA	2
analysis of embedding features	NA	1
sufficient training data	NA	2
emotion recognition	NA	0
extensive amounts of training data	NA	0
		26

IBM Watson AlchemyLanguage API		
Common form	Score	Eval
natural language	1.00	2
Representative	0.80	0
Kim	0.67	0
SVM	0.62	1
fl	0.55	0
application domain	0.54	2
core point	0.52	0
Paul Ekman	0.48	0
International Organization	0.47	0
Sect.	0.46	0
Latent Semantic Analysis	0.46	2
SVM	0.45	1
knowledge base	0.43	2
modi	0.43	0
Twitter	0.43	1
Lee	0.43	0
Hasegawa	0.43	0
neural network	0.42	2
CRF	0.42	1
Xu	0.42	0
		14

D: Replacing carbohydrate with protein and fat in prediabetes or type-2 diabetes: greater effect on metabolites in PBMC than plasma

https://rd.springer.com/article/10.1186/s12986-016-0063-4

UNSILO Concept Extraction API		
Common form	Score	Eval
Plasma Lp-PLA2 Activity	1.00	2
Lp-PLA2 Activity	0.74	0
High Lp-PLA2 Activity	0.40	2
Plasma Lp-PLA2	0.39	0
Plasma ox-LDL	0.35	1
12-week Dietary Intervention	0.32	1
Impaired Fasting Glucose	0.31	2
Score Scatter Plot	0.28	0
Basal Metabolic Rate	0.25	2
PBMC Gene Expression	0.22	2
Glycemic Control	0.21	2
Type-2 Diabetes	0.20	1
High Lp-PLA2	0.20	2
ox-LDL Level	0.19	2
Total Energy Expenditure	0.19	2
lysoPC Level	0.19	2
THP-1 Monocyte	0.18	2
Plasma Metabolite	0.18	2
Wallac Victor2 Multilabel Counter	0.16	2
Dietary Fiber Intake	0.15	2
		31

Google Cloud Natural Language API		
Common form	Score	Eval
group	1.00	0
PBMC	0.96	1
PBMC	0.30	0
activity	0.13	0
Ox-LDL	0.13	1
Lp-PLA2	0.09	1
Lp-PLA2	0.09	0
grains	0.09	1
PLS-DA	0.09	1
groups	0.04	0
blood cells	0.04	2
UPLC-LTQ-Orbitrap MS	0.04	1
components	0.00	0
intervention	0.00	0
epidemic metabolic disorder	0.00	2
plasma	0.00	1
metabolites	0.00	1
rice	0.00	1
T2D	0.00	1
subjects	0.00	0
		14

Microsoft Cognitive Services Text Analytics API		
Common form	Score	Eval
Lp-PLA2 activity	NA	2
intervention diet	NA	0
Dietary intervention	NA	2
release of Lp-PLA2	NA	2
plasma Lp-PLA2 activities	NA	2
blood cells	NA	2
week intervention study	NA	-1
B cells	NA	2
usual refined-rice diet	NA	0
usual diet	NA	1
T cells	NA	2
natural killer cells	NA	2
PBMC metabolites	NA	2
week intervention phase	NA	-1
replacement of refined rice	NA	0
blood glucose	NA	2
cooked refined rice	NA	0
typical diet	NA	2
Nonobese subjects aged	NA	-1
Assessment of dietary intake	NA	0
		20

IBM Watson AlchemyLanguage API		
Common form	Score	Eval
PBMC	1.00	1
MDA	0.50	1
oleamide	0.45	1
blood glucose	0.44	1
IFG	0.43	1
Korea	0.43	0
PBMC	0.43	1
BMI	0.41	1
IGT	0.40	1
Lp-PLA2	0.40	1
PBMC oleamide	0.40	1
Hitachi Ltd.	0.38	0
Diabetes	0.37	1
ox-LDL	0.37	1
Thermo Fisher Scientific	0.36	0
Tokyo	0.35	0
Japan	0.35	0
fatty acids	0.35	1
immune system	0.34	1
metabolic rate	0.34	1
		15