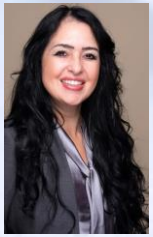
NLP SUMMIT HEALTHCARE

Transforming the Clinical Decision Support and Clinical Analytics Landscape Through NLP



Elia Lima-Walton, MD

Director Data Science & Clinical Analytics Health Elsevier

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Agenda

- Our mission, how we operate, our community
- Elsevier's Medical Taxonomy Introduction & Connection with Health Products
- Data Science, NLP, and Clinical Analytics Capabilities

NLP SUMMIT

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Our Mission

- Elsevier helps researchers, educators, and healthcare professionals advance science and improve health outcomes for the benefit of society.
- We do this by facilitating insights and critical decision-making for customers across the global research and health ecosystems.

Who we are
A global information analytics business specializing in science and health

What we do it
We help you solve your challenges, for the benefit of humanity

ELSEVIER

What we do
We help institutions and professionals progress science, advance healthcare and improve outcomes

A unique combination
Combine content with technology, supported by operational efficiency, to turn information into actionable knowledge

NLP SUMMIT

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Supporting Our Five Main Customer Segments

Clinicians	‘Consider this treatment for this patient’
Researchers	‘This article answers your questions’
Governments	‘This is the research to invest in’
Pharmaceutical companies	‘This is the novel treatment you should pursue’
Medical and Nursing Students	‘This is the area you need to improve to succeed’

NLP SUMMIT

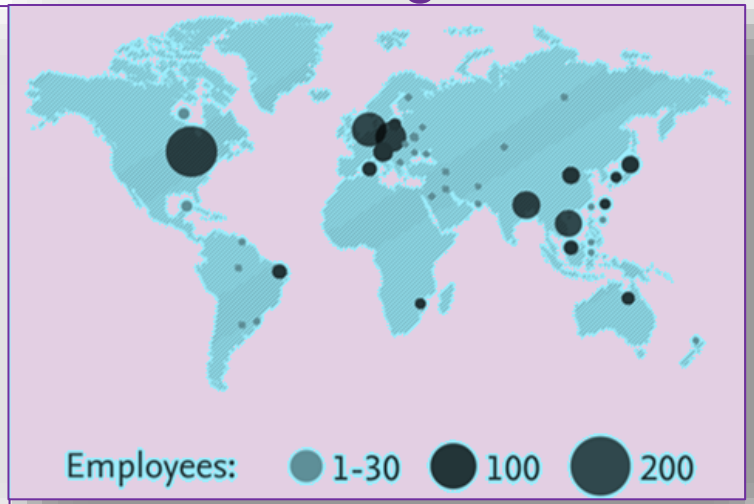
Elia Lima-Walton, MD
Elsevier
March 27, 2023

Elsevier Flagship Products

At Elsevier we facilitate insights and critical decision-making for our global customers.

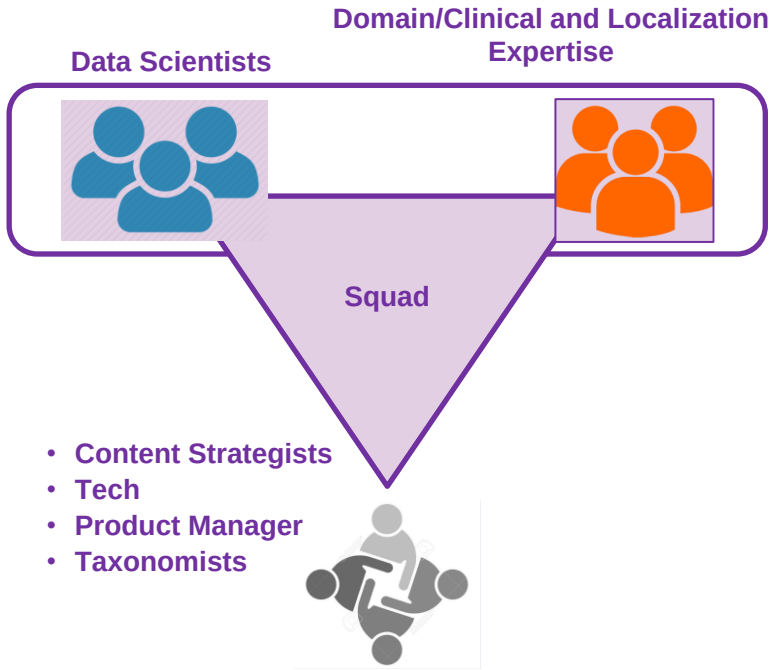
Elsevier's Data Science Community and Global Organization

- **250+ data scientists and clinician specialists**
- Several Global Collaborations with Universities in the United States, Europe, and India
- Main hubs in:
  - Amsterdam
  - London
  - Philadelphia
  - Chennai
  - Bengaluru
- Artificial Intelligence Lab in Amsterdam https://discoverylab.ai/



Elsevier has **8,700** employees and serves customers in **180+ countries**

Data Science Health, Who We Are and How We Operate



- We are organized locally and globally
- Support Health initiatives globally with data science, clinical, and localization expertise
- Manage extended bench of external contractors and suppliers
- Execute on innovative projects aligning with internal external collaborators

- Philadelphia
- Houston
- Dallas
- Knoxville
- Asheville

- Amsterdam
- Barcelona

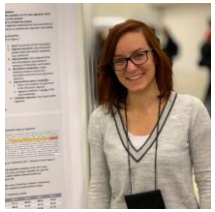
- Chennai
- Delhi/Gurgaon
- Bengaluru
- Mumbai

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Meet the team– Data Science & Clinical Analytics

Elia Lima-Walton, MD
Director Data Science & Clinical Analytics Health

Hina Nazir, MD
Clinical Analytics Manager



Dasha Herrmannova, PhD
NLP & Data Science Manager

Leah Livingston, RN
Clinical Data Scientist & Analytics Specialist

Sameer Chivukula, MSc
Data Science Manager

Sharvari Jadhav, MSc
Data Scientist

Harsh Sindhwa, MSc
Data Scientist

Charlotte Tesselaar, MSc
Sr. Localization Specialist

Pranita Mahajan, MEng.
Senior Data Scientist

Vidhyaa Rajee, MSc
Data Scientist

Roles and expertise

Multi-disciplinary department: Data Scientists, Clinical Analytics Specialists, Localization Specialist, Knowledge Management, Life Sciences

Data Science and Health expertise: Machine Learning, Natural Language Processing, Machine Translation, Optical recognition, Clinical Analytics, Medicine, Nursing, Hospital Admin, Pharmacy and Drug Development

Dwipjayeta Choudhury, RN
Sr. Clinical Data Analyst

Maria Jedrzejowska, MSc
Localization Specialist

Key Capabilities Developed in Data Science and Clinical Analytics

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Combining the Power of our Trusted Medical Content with Data Science

Key capabilities which we prioritize and develop:

1.)Search and Discoverability, NLP & ML

Concept/taxonomy alignment and information retrieval: Aligning medical taxonomy and ontology to content in order to power user search experience

2.)Standardization of metadata, normalization, new technologies: Developing innovative technologies to support integration, normalization, and generation of text, images, and multimedia assets.

3.) Globalization/Localization :*Focus on the global customer* by adapting products, content, and initiatives to various locales and cultures.

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Unique Assets and Capabilities to Extract Information from Clinical Records and Medical Curricula

Use Cases

- Deidentify text records
- Commercialize health data
- Risks, predictive analytics
- Coding and reimbursement
- Extract knowledge from literature
- Map insights to patients
- Mapping educational curricula

Problems to Solve

Convert unstructured into structured data from
- clinical notes
- radiology reports
- pathology reports
- lab reports
- biomarkers
- education alignment

→ NLP

Standardize data from multiple hospitals (SNOMED, ICD 10, CPT, LOINC, ATC, NDC, …)

→ taxonomies

Contextualize and interpret data

→ knowledge graph

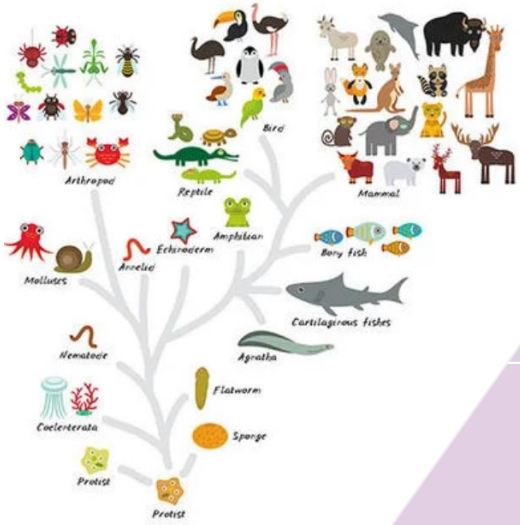
NLP SUMMIT

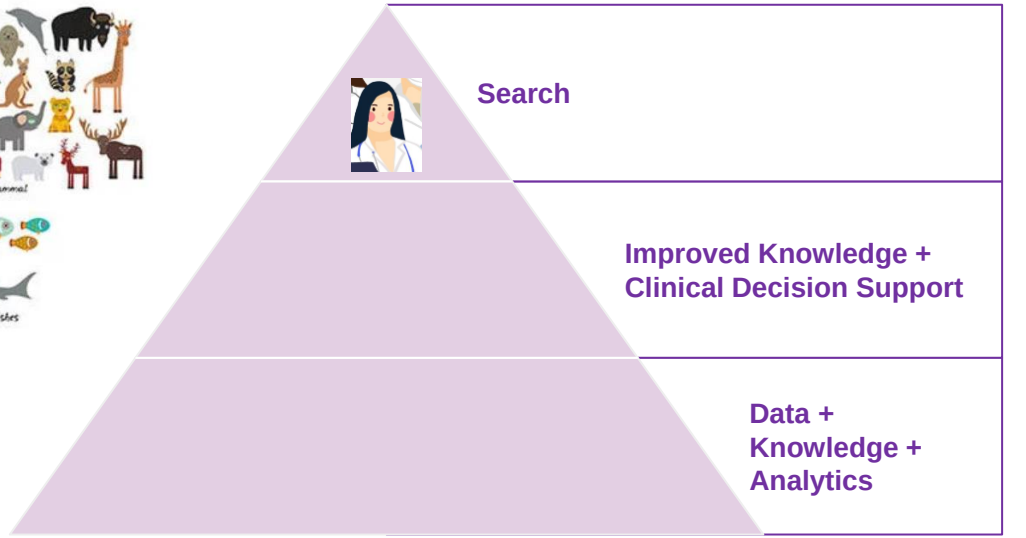
Taxonomy for synoptic content discoverability

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Taxonomy Powering Search and Product

Taxonomy: Science of identifying and naming species and arranging them into classification





Simple — Lists — Terminologies — Complex

Controlled Vocabularies — Ontologies

Objective
Clinicians, researchers, students, and educators need concise actionable answers. Identify and prioritize key concepts embedded in content to aide in improved information retrieval.

NLP SUMMIT

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Introducing Elsevier's Medical Taxonomy

- **Multilingual:** English, French, Spanish, German, Brazilian Portuguese
  - Arabic, Chinese, Somali, Nepali, Armenian, in pilot
- **Concept-based:** All terms, synonyms, translations, mappings are related to one unique identifier ("IMUI")
- **Ontology:** Provides semantic relationships between concepts (symptoms of a disease, treatment procedures of a disease, complications of a disease etc…)
- Explicitly mapped to **international medical standards** *(SNOMED-CT, ICD-10-CM, LOINC, CVX, etc.)* and Elsevier's internal vocabularies (Gold Standard, Dorland's).
- Continuously curated by team of clinical medical terminology experts, available via API
- Is at the heart of our knowledge graph, connecting concepts and relationships supported by evidence in content, unlocking knowledge through scalable, easily-navigable information services.
- Knowledge graph has **400K+** medical concepts, **900K SYN** with over **220K semantic relations**, **90K symptoms**, **46K drugs**, **75K diseases**, **63K procedures** and extensive mappings to industry vocabularies and
multi-language support.

Introducing Elsevier's Taxonomy and Knowledge Graph

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Parents

- Breast disorders (IMUI:2805667)
- Cancer of the thorax (IMUI: 8258010)
- Neoplasm of breast (IMUI: 8258030)
- Malignant neoplasm (IMUI: 2791003)

IMUI: 2790981

- Lang: En
- **Medical Name**
- Malignant neoplasm of the breast
- **Consumer-Friendly Name**
- Breast cancer
- **Synonyms**
- Malignant tumor of the breast
- Malignant breast neoplasm
- Breast CA
- **STY/STYGROUP**
- Neoplasm process / Disease
- **Mappings**
- UMLS – C0006142 (exact)
- ICD9-CM – 174.9 (exact)
- ICD10-CM – C50 (exact)
- SNOMED-CT – 254837009 (exact)
- MeSH – D001943 (exact)

Children

- Breast sarcoma (IMUI:8258036)
- HER2-positive breast cancer (IMUI: 8316103)
- Male breast cancer (IMUI: 8009640)
- Lobular carcinoma (IMUI: 3816091)

…

Semantic Relations

HAS PHYSICIAN SPECIALTY
Specificity
10 — 7

HAS CAUSE
Prevalence
10 — 7

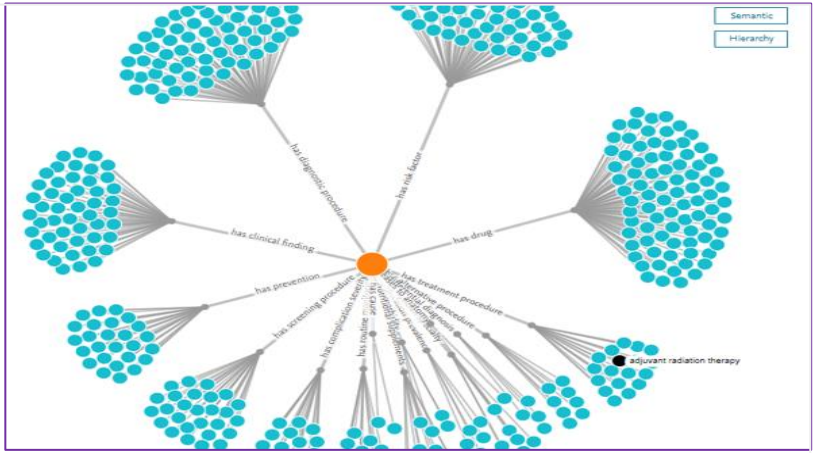
HAS CLINICAL FINDING
Prevalence
10 — 7

HAS DIAGNOSTIC PROC
Indication
10 — 7

HAS TREATMENT PROC
Indication
10 — 7

HAS DRUG
Indication
10 — 7

HAS COMPLICATION
Severity
10 — 7



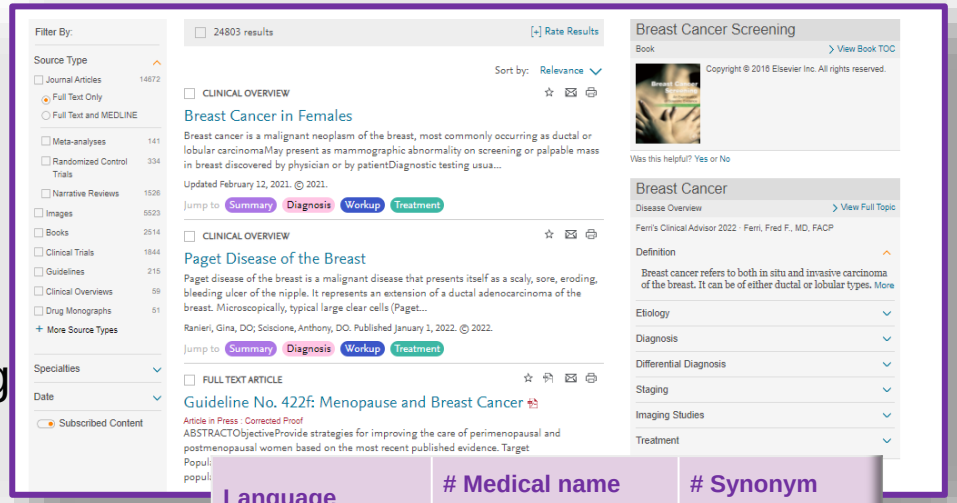
- What clinical findings are associated with **breast cancer?**
- How do you screen **breast cancer?**
- What is the diagnostic work-up for **breast cancer?**
- What is the gold standard treatment for **breast cancer?**
- What are common differential diagnoses for **breast cancer?**

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Helping Clinicians and Students Worldwide

- ClinicalKey US
- ~20 million content assets
- **Our taxonomy is used for:**
- Autocomplete search
- More accurate search results
- Faceted navigation (Related screening drugs)
- Query refinement

Localized versions for France, Spain, Brazil and Germany

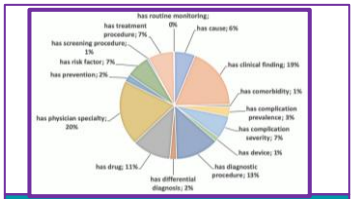


Language	# Medical name labels	# Synonym labels
English	400K	1M
Spanish	200K	400K
German	178K	167K
French	111K	210K
Brazilian Portuguese	150K	111K

Knowledge Extraction From “Noisy” Content

Elia Lima-Walton, MD
Elsevier
March 27, 2023

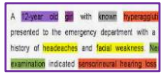
- Mining semantic patterns from our taxonomies
- Improving the quality of diverse data sets and annotations and search, extracting and identifying novel relationships for CDS support.
- Shared taxonomy and ontology used for breaking silos, connecting content, peoples, and different assets.



200K SME curated medical triples for 16 relation types

FPE | CM_2 | DL-NER

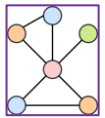
Medical Entity annotators configured with Elsevier’s medical taxonomy and UMLS



Medical and clinical text

Pattern- and DL–based Annotators

Medical Taxonomy



Knowledge Graph

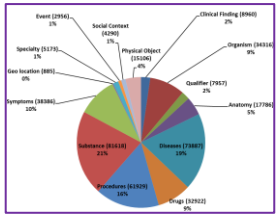
GOLD Set

SME Curated Datasets

Pattern(x)
,
Pattern(y)
,
filter(z)

Heuristics and Rules

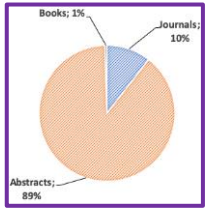
Sentences and relationships validated by clinical SME’s



400K medical entities classified under 16 semantic types

```
def isblacklisted():
def isSeqList():
def isDepList():
def isDepNeg():
```

Rules-based on syntax, sentence structure and SME’s knowledge

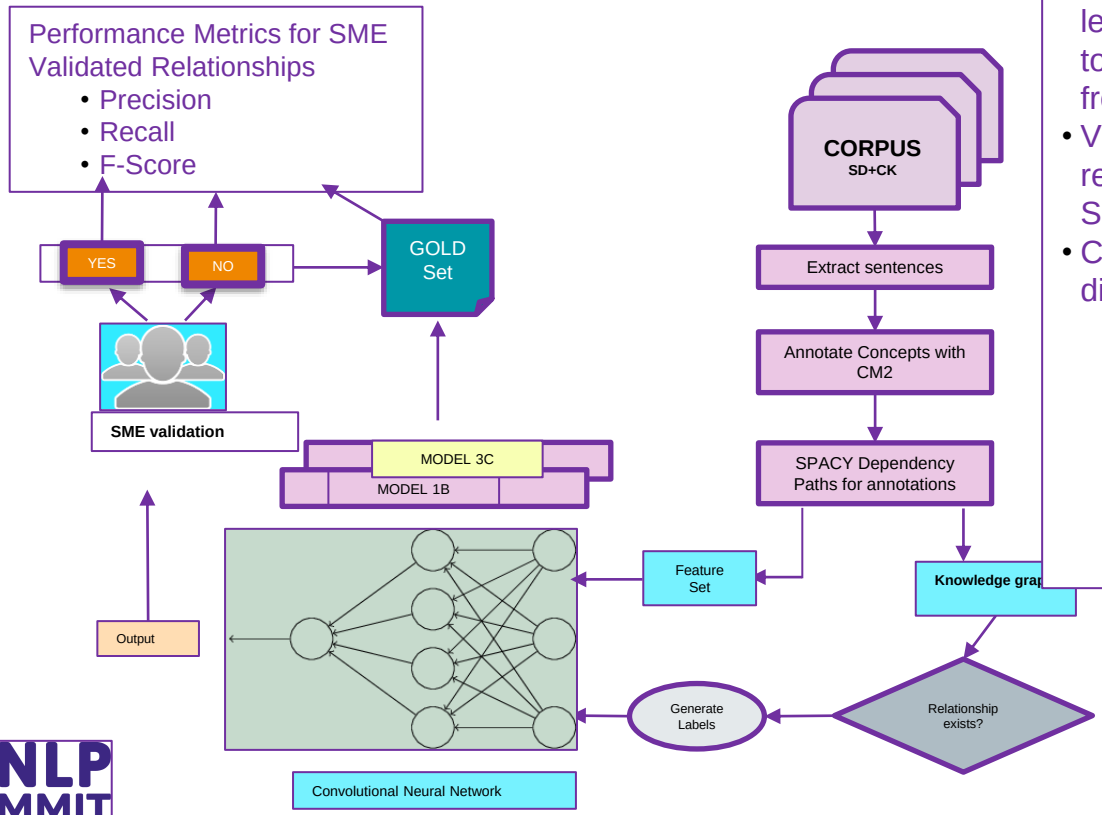


12.58M documents

- 1300 books
- 1.3M articles
- 11.2M abstracts

NLP SUMMIT

Triple Extraction: Symptoms, Treatment, Diagnostic Procedure, and Drug Relationships from Medical Content at Scale



- Modeling workflow utilizes weakly supervised learning via convolutional neural networks (CNN) to predict symptom to disease relationships from free text content
- Validation done on a Gold Set in which relationships are assessed as true or false by SMEs.
- Connecting products and improving discoverability at Point-of-Care.

  Add value by:

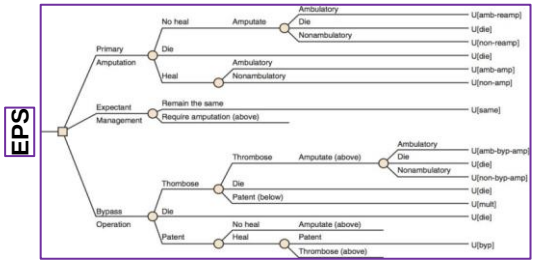
  1. **Improving accuracy of treatment**
  2. **Reducing patient cost of unnecessary tests/treatment**
  3. **Reduced malpractice litigation**

	Number of Documents
Books	1,320
Journals	1,317,321
Abstracts (Medline)	1,322,484
Total	2,641,125

> 4.2 million clinical relationships extracted from unique evidence

Elia Lima-Walton, MD
Elsevier
March 27, 2023

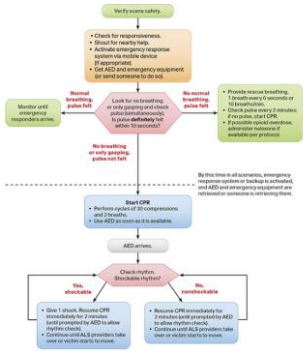
Automating Extraction of Medical Decision Trees Via Computer Vision



Flowchart or decision tree corpus is **heterogenous**

–

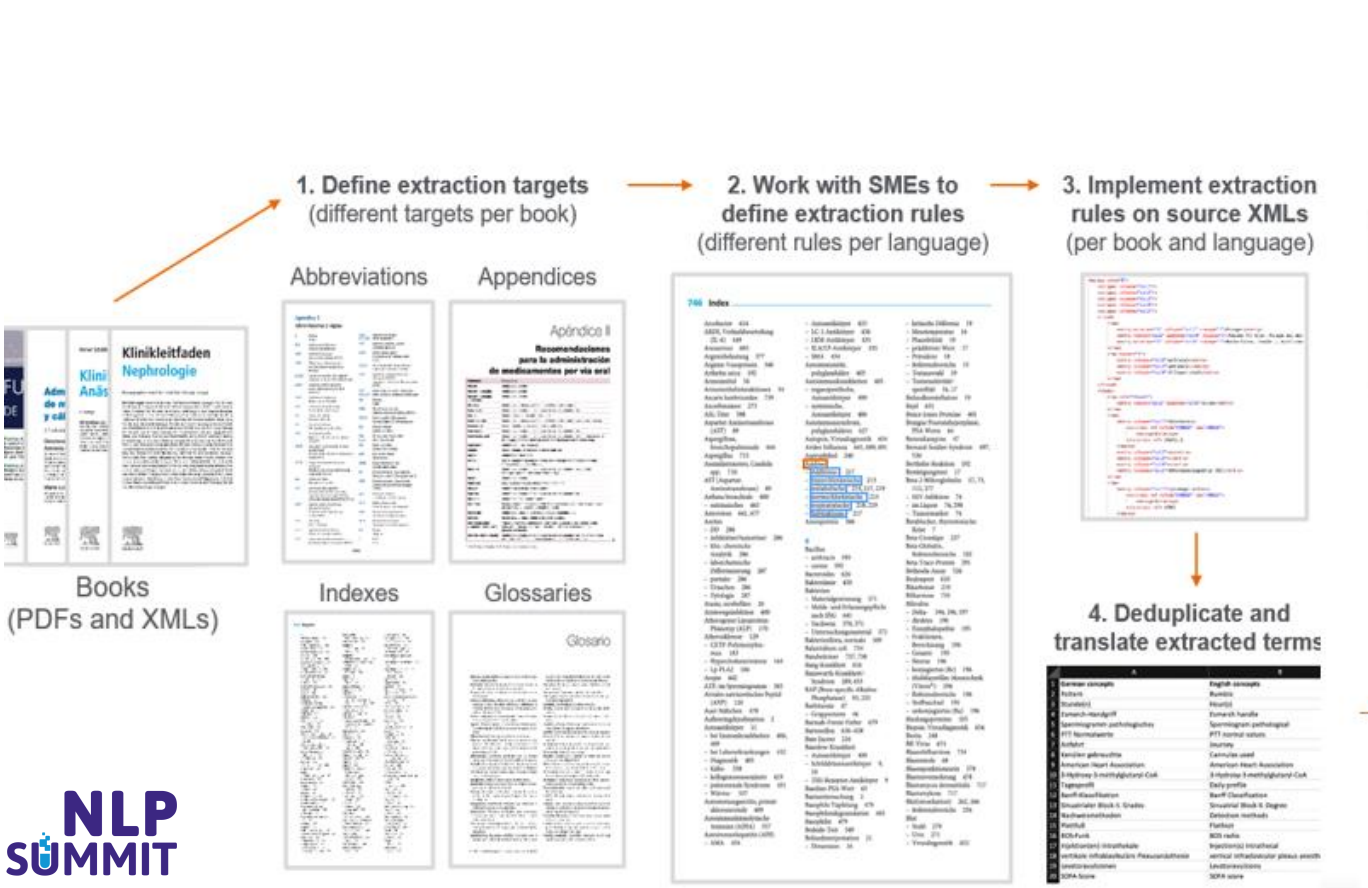
Different shapes, flow types, formats, colors, resolution, etc.



We developed a general, **computer vision-based AI system**
for **flowchart deconvolution** to be used at the **point-of-care**
and in **education**

Data Science Tools and Domain Expertise Applied for Automated Taxonomy Alignment for Localization & Search Improvement

Elia Lima-Walton, MD
Elsevier
March 27, 2023

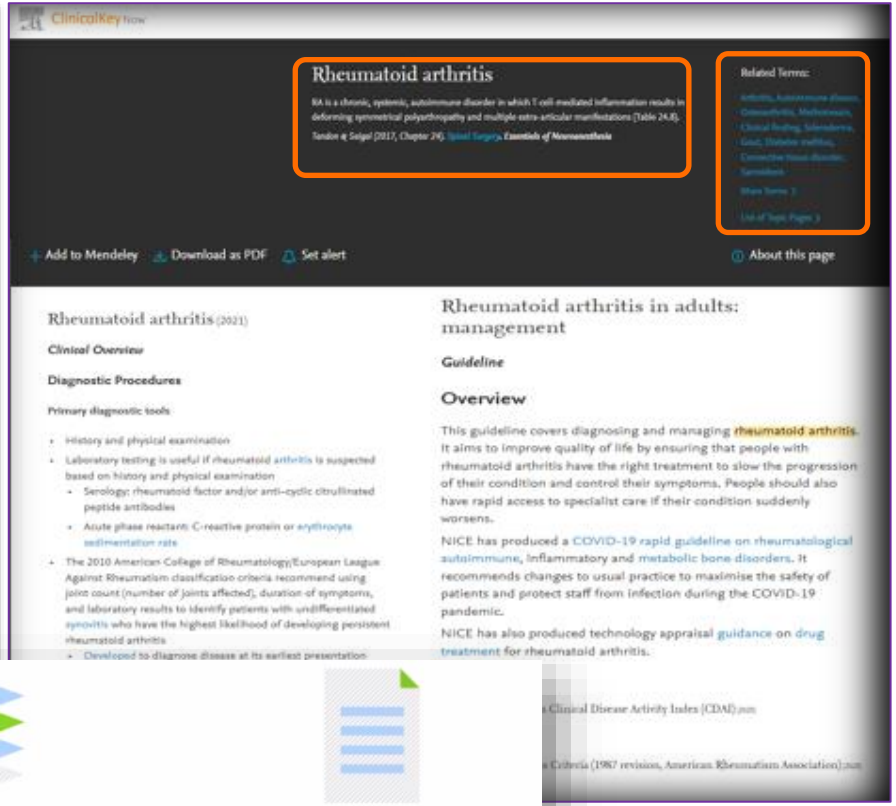


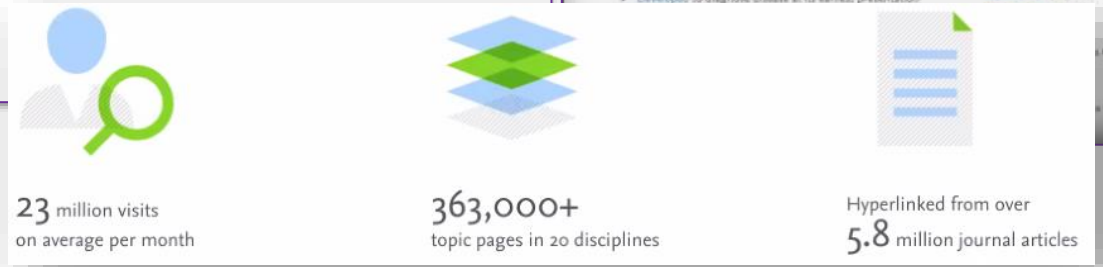
NLP SUMMIT

Empowering Clinicians and Students TO Discover Content

Elia Lima-Walton, MD
Elsevier
March 27, 2023

- Using deep learning and data indexing techniques to develop **Topic Pages** that **are indexable and searchable** on **Google** for our customers
- Provide definitions and related terms for each concept
- Provide synoptic snippets of high-quality, relevant, peer-reviewed content
- Offer hundreds of thousands of comprehensive and topical web pages
- Free to access through our Elsevier Platforms





Improving the User Search Experience

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Snippet evaluation

- Snippet images were reviewed by semantic group: Symptoms/findings, disease, drug, procedure
- Ranking system generated for image snippets

Validating and Labelling definitions sentences

- Sentences prioritized by token
- Semantic patterns identified in sentence structure
- Definitions if multiple for a term were classified as preferred (Yes) or not preferred (No)
- 5,000+ definitions analyzed

Images for Diffuse dermatitis

Fig 1 Clinical presentation on admission. Coalescent erythematous scaly papules in perioral area and periorbital edema are noted.

Lovett, Audrey, MD: Diffuse dermatitis: An unexpected initial presentation of cystic fibrosis. From: Journal of the

Fig 2 Diffuse erythematous scaly papules on trunk.

Lovett, Audrey, MD: Diffuse dermatitis: An unexpected initial presentation of cystic fibrosis. From: Journal of the American Academy of Dermatology,

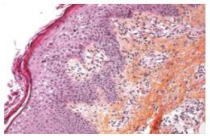


Fig 3 Specimen from first biopsy showing acanthosis, confluent parakeratosis, mild spongiosis, and mild lymphohistiocytic perivascular infiltrate with a few eosinophils. (Hematoxylin-eosin stain; original magnification: ×20.)

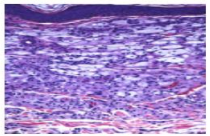


Fig 5 Later biopsy specimens showing predominately foamy histiocytes with rare Touton giant cells. (Original magnification: ×10.)

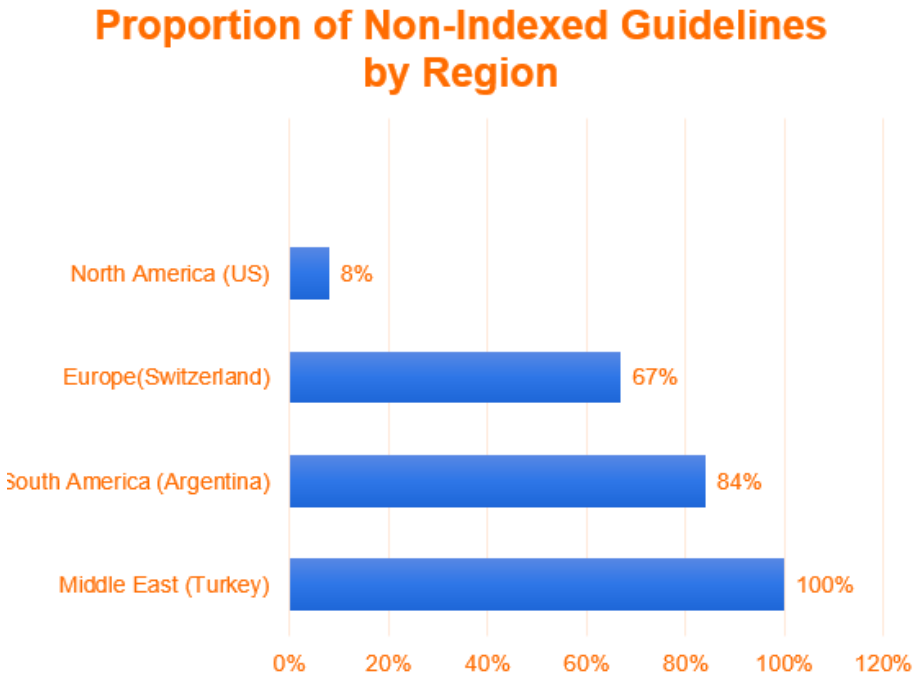
Satter, Elizabeth K., MD, MPH: Diffuse xanthogranulomatous dermatitis and systemic

sentence	label	probat	tokenl	concept	imui	Preferred
Constipation is defined symptomatically as the infrequent passage of hard stools, straining while passing a stool, or pain associated with the passage of a hard stool.	TRUE	0.89073	0	Constipation	2791117	Y
Contact lens wear is another form of corneal microtrauma that seems to be associated with keratoconus.	TRUE	0.58857	0	Contact lens	2791118	N
Febrile seizures (FSs) are the most common type of pediatric seizure and occur in patients between 6 months and 5 years of age associated with fever but without intracranial infection or defined cause.	TRUE	0.77412	0	Febrile seizures	2791122	Y
Febrile seizures are the most commonly reported neurologic adverse event after measles vaccination.268,726,729â733 Fever from any source lowers the threshold for seizures, and a febrile seizure is not a sign of central nervous system (CNS) infection or disease.	TRUE	0.57132	0	Febrile seizures	2791122	N
A cosmetic is any substance applied to the body for cleansing, beautifying, promoting attractiveness or altering the appearance.	TRUE	0.92267	1	cosmetic	2791125	Y
Unlabelled boxCosmeticsâCA cosmetic is a substance applied to the body for cleansing, to promote attractiveness or to alter the appearance.	TRUE	0.90877	2	cosmetic	2791125	N
Tetracosactide is a synthetic peptide analogue which consists of the active N-terminal amino acids 1â24 of the ACTH molecule.	TRUE	0.60257	0	Tetracosactide	2791126	Y
Craniosynostosis refers to the premature closure of one or more cranial sutures, which may occur during the prenatal period, early infancy, or childhood.	TRUE	0.81027	0	Craniosynostosis	2791127	N
Craniosynostosis is premature fusion of one or more cranial vault sutures with resultant skull deformity.	TRUE	0.7404	0	Craniosynostosis	2791127	N
Craniosynostosis is a congenital disorder that occurs due to premature intrauterine fusion of one or more of the six cranial sutures leading to an abnormal shape of skull.	TRUE	0.94814	0	Craniosynostosis	2791127	Y

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Most North American Guidelines are published in journals; Most international guidelines are published on websites

- Guidelines that are published in medical journals are usually indexed in literature databases such as MEDLINE
- A small but important proportion of North-American guidelines are published on websites rather than indexed MEDLINE journals
- A large proportion of international guidelines are not indexed on MEDLINE



https://smw.ch/article/doi/smw.2019.20134
https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-3-43
https://www.dovepress.com/assessing-the-quality-of-clinical-practice-guidelines-in-the-middle-ea-peer-reviewed-fulltext-article-JMDH#t0002

NLP SUMMIT

Smart Surveillance BOT for Mining Dynamic

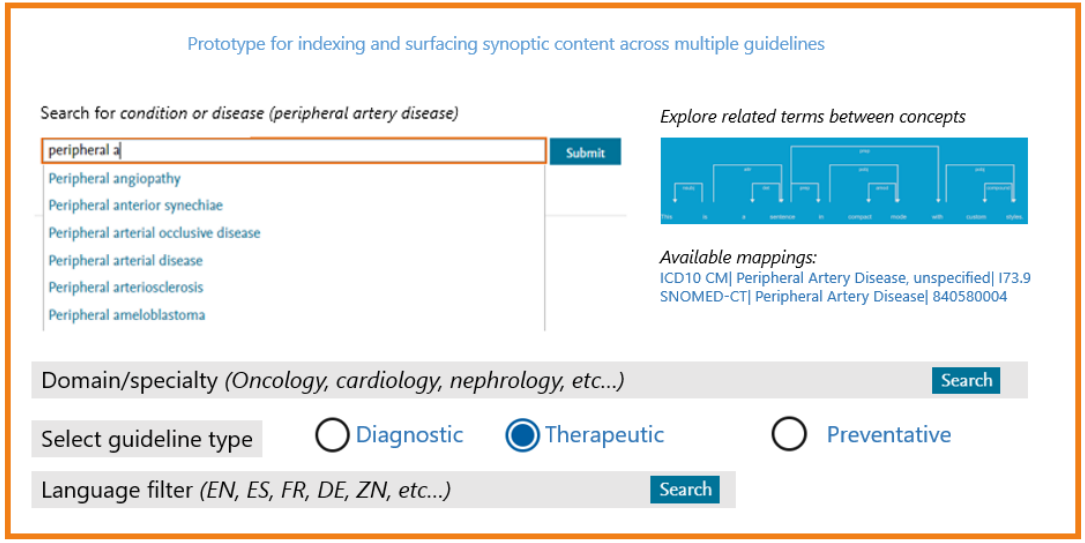
Elia Lima-Walton, MD
Elsevier
March 27, 2023

❑ Problem Statement

- Background: As the amount of content to review expands the importance of a smart literature surveillance BOT also increases.
- Bot provides dynamic streamlining by connecting a central repository guidelines that are easily searchable.

❑ Objective

- Creation of a surveillance BOT which can be run autonomously to gather dynamic changes in guidelines to improve evidence curation and decision support at the point of care.
- Empowering our clinician customers to better find how to diagnose and how to treat their patients.



NLP SUMMIT

Elia Lima-Walton, MD
Elsevier
March 27, 2023

AI Radiologic Image Interpretation and Report Summarization

❑ Problem Statement

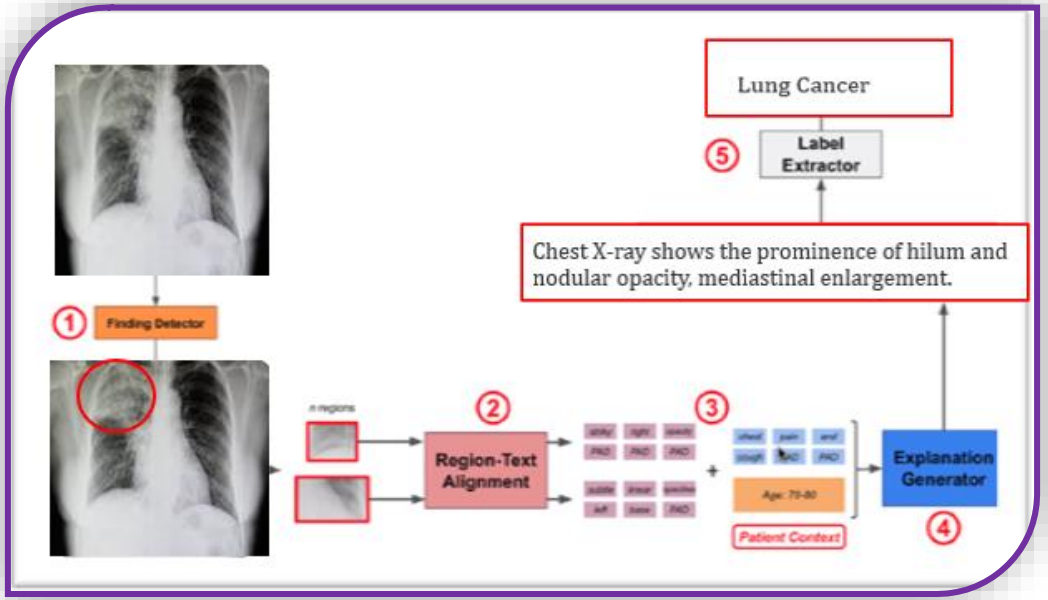
- There is a need to be able to provide predictive analytics at the point-of-care to improve patient outcomes by providing ***timely detection of deleterious signs*** in the hospital setting.

❑ Objective

1.Develop a predictive model for gaining insight into the patient journey and improving accuracy of diagnosis while decreasing the time to diagnose.

2.Algorithmically identify which features can be indicative of disease or prognostic outcome.

3.Provide support at the point of care.

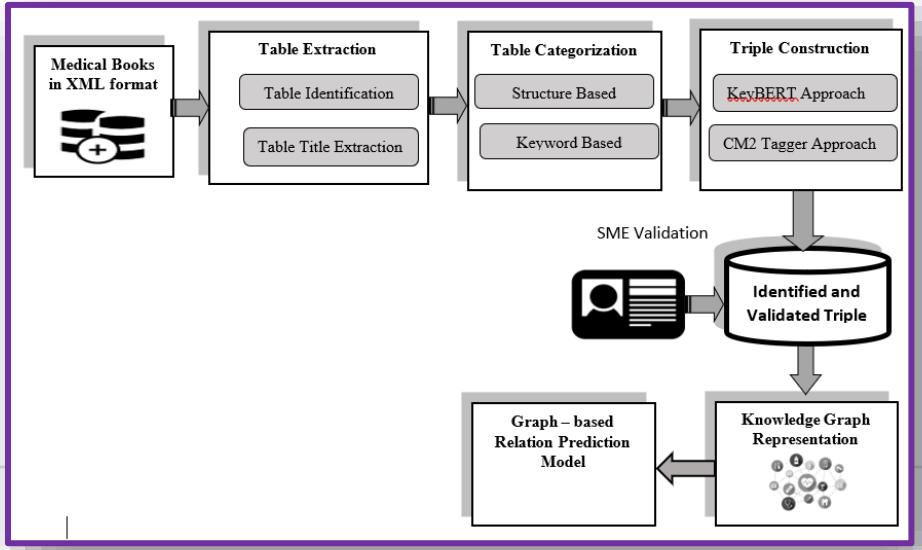


NLP SUMMIT

Elia Lima-Walton, MD
Elsevier
March 27, 2023

Semantic Table Extraction

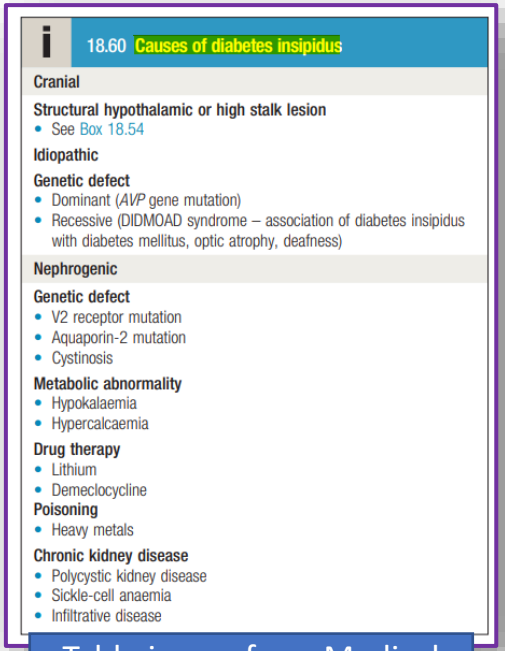
- As the amount of scientific literature continues to grow exponentially, so does the need and time to review vast amount of content increase for clinicians and medical authors.
- Goal: Develop and implement tool aiding in knowledge extraction, in tabular format for purposes of generated synoptic content in graph form which can locate and identify relevant data by semantic category and prepare it for further processing and transformation.
- Benefits: Are many, improved authoring of content, information retrieval, decision support, and business intelligence.



NLP SUMMIT

Semantic Table Extraction Pipeline

Elia Lima-Walton, MD
Elsevier
March 27, 2023



18.60 Causes of diabetes insipidus

Cranial

Structural hypothalamic or high stalk lesion
- See Box 18.54

Idiopathic

Genetic defect
- Dominant (*AVP* gene mutation)
- Recessive (DIDMOAD syndrome – association of diabetes insipidus with diabetes mellitus, optic atrophy, deafness)

Nephrogenic

Genetic defect
- V2 receptor mutation
- Aquaporin-2 mutation
- Cystinosis

Metabolic abnormality
- Hypokalaemia
- Hypercalcaemia

Drug therapy
- Lithium
- Demeclocycline

Poisoning
- Heavy metals

Chronic kidney disease
- Polycystic kidney disease
- Sickle-cell anaemia
- Infiltrative disease

Table image from Medical Book PDF

1. Prepend a column header to each concept
2. Detect row header – if any, prepend it with each concept from row
3. Generate a triple with {table_keywords, Column_header, Concept_Value}

Steps to extract data

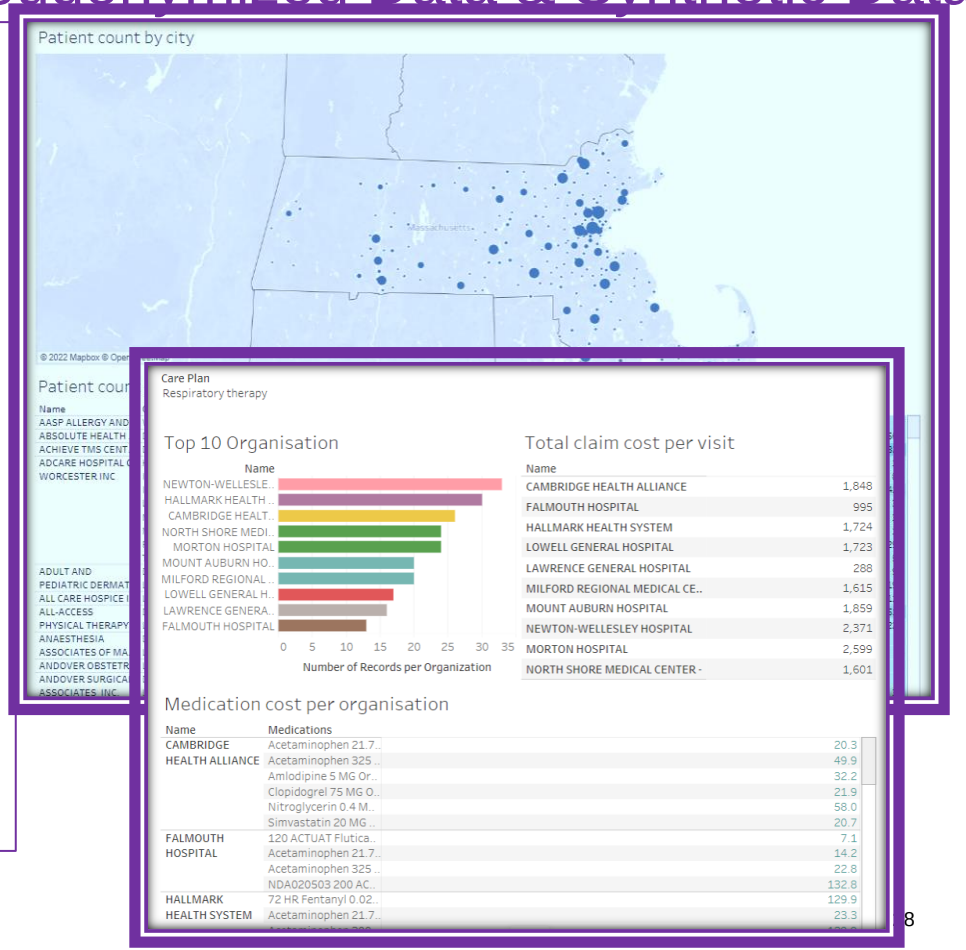
index	table_id	Table_name	Keywords	Predicate	Concept_list
0	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error studies']	Error category - No fault	'Examples - Missing information']
1	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error studies']	Error category - No fault	Examples - Unusual presentation of a disease',
2	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error studies']	Error category - System error	'Examples - Inadequate diagnostic support'
3	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error studies']	Error category - System error	'Examples - Results not available'
4	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error studies']	Error category - System error	'Examples - Error-prone processes'
5	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error studies']	Error category - System error	Examples - Poor supervision of inexperienced staff'
6	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error studies']	Error category - System error	'Examples - Poor team communication'
7	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error studies']	Error category - Human cognitive error	'Examples - Inadequate data-gathering'
8	tit0015	Root causes of diagnostic error in studies	['root causes', 'diagnostic error	Error category - Human cognitive	'Examples - Errors in reasoning']

Generated Triples

Elia Lima-Walton, MD
Elsevier
March 2, 2023

Predictive Analytics Using Pseudonymized Data & Synthetic Data

- Displays adherence of products in relation to data sets
  - Facility, location, date, condition
- Identifying Plan Adherence
- Automating Reporting
  - Viewer advantage of interactive Tableau dashboard vs. manually reviewing PDF's.
    - Patient readmission
    - Care team utilization
    - Nurse staffing
    - Hospital flow simulation etc.



Elia Lima-Walton, MD
Elsevier
March 27, 2023

Automated Patient-to-Trial Finder to Improve Access to Care

- Clinical trials are the cornerstone of evidence-based medicine, ensuring the availability of safe and effective treatments by studying their effects on human subjects.
- Matching eligible patients to clinical trials is essential for achieving statistically significant results; however, the recruitment of patients represents a bottleneck in clinical research.
- To better understand the limitations of our approach, we performed a manual evaluation of 600 topic-trial recommendations, demonstrating the importance of accounting for the trial exclusion criteria in retrieval.
- SOLUTION: Developed a prototype via application of transformer embeddings using MiniLMmodel.

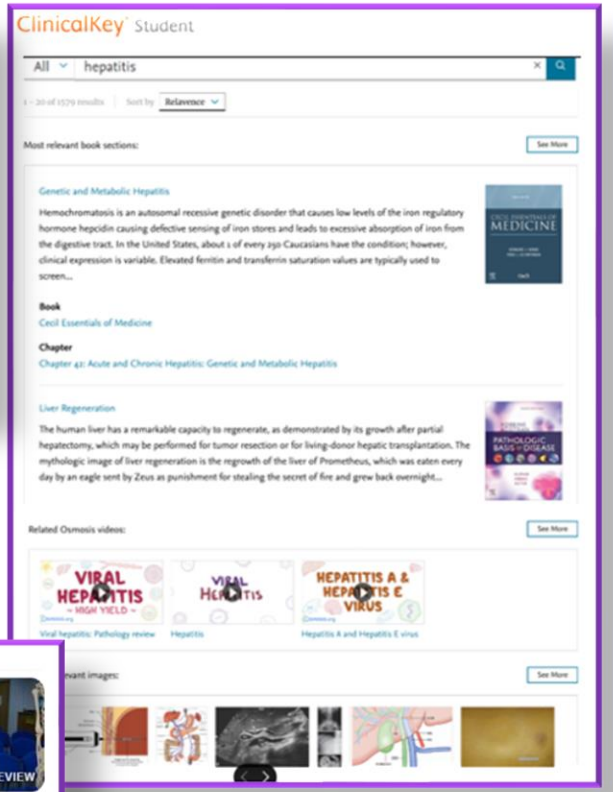
https://trec.nist.gov/pubs/trec31/papers/els_dshs.T.pdf

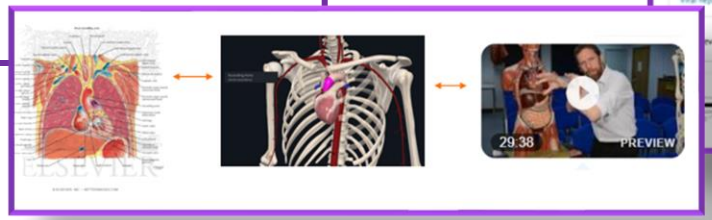
Data Science Improving Educational Offerings

Automated Transcript Generation and Metadata Tagging for Linking Elsevier Health Products for Medical Students

Elia Lima-Walton, MD
Elsevier
March 27, 2023

- Improved content indexation and interlinking between transcripts and clinical vignettes
- Facilitating navigation and interoperability across education products
- Used to create model which prioritizing key terms/meta data to be surfaced
- Additionally:
  - Automated annotation and parsing
  - 2D/3D image cross walks created
  - Automated prioritization of key to support various Products





NLP SUMMIT

Elia Lima-Walton, MD
Elsevier
March 27, 2023

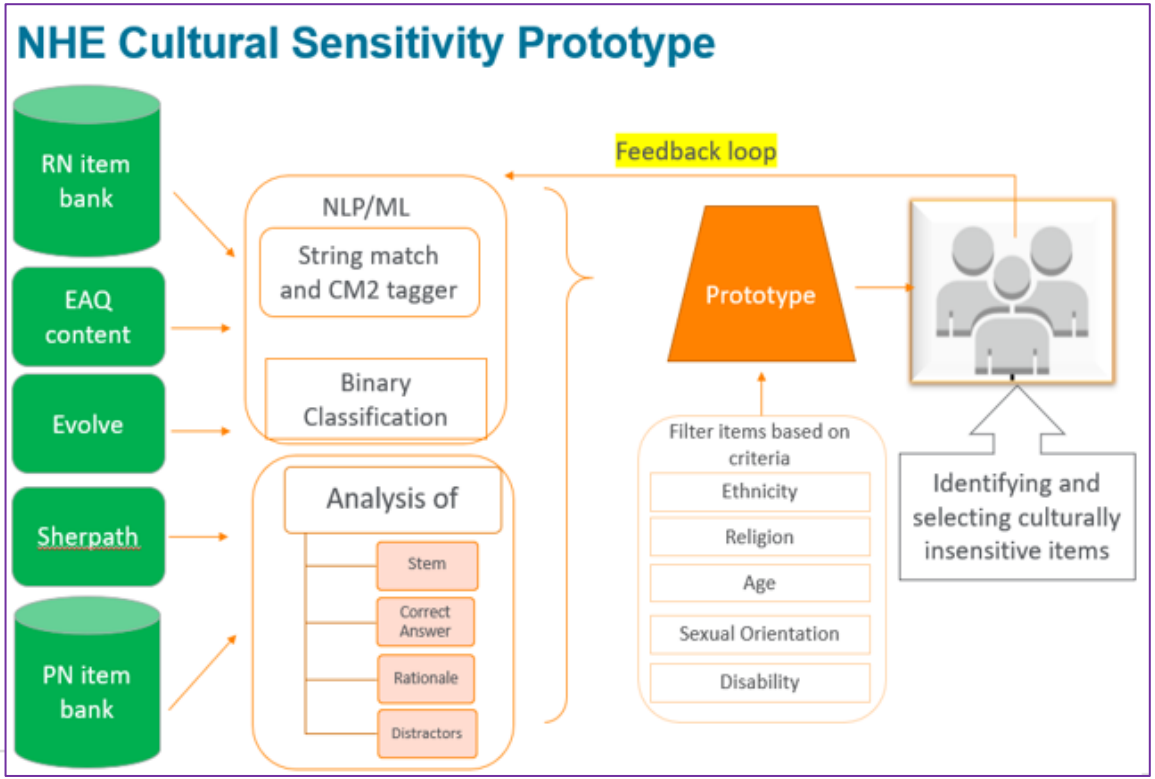
Cultural Sensitivity Prototype for Nursing Health Education

Objective

1.Investigate which items are culturally insensitive in nursing practice test items.
2. Provide analytics of topic area and banks containing culturally insensitive items.
3.Build a prototype using ML, NLP, & Analytics which identifies items from nursing banks and remediation material that is culturally insensitive.

Key Impact

Text extraction, ingestion, and binary classification to feed into algorithm and built prototype for NHE support.



Supporting Patient, Student, and Researcher's Journey

"At the heart of improving healthcare is effectively applying outcomes-driven analytics across the care continuum. I'm excited about the tremendous opportunity to combine the market-leading evidence based content and clinical reference and workflow solutions of Elsevier Health with our differentiated data and analytics capabilities to support healthcare providers in delivering better outcomes for patients."

Josh Schoeller, *President Clinical Solutions, Chief Executive Officer of the Health Care business of LexisNexis® Risk Solutions*

Thank you

Contact details:
e.lima@elsevier.com

NLP SUMMIT HEALTHCARE

Presented by Elia Lima-Walton, MD

John Snow LABS

www.nlpsummit.org