



SKOS OF THE 1910 LIBRARY OF CONGRESS SUBJECT HEADING  
FOR THE TRANSFORMATION OF THE KEYWORDS TO  
CONTROLLED VOCABULARY OF THE NINETEENTH-CENTURY  
ENCYCLOPEDIA BRITANNICA

SONIA PASCUA, JANE GREENBERG, PETER LOGAN, AND JOAN BOONE

# OUTLINE



Introduction



Background and  
Motivation



Statement of the  
Problem



Methodology



Results and  
Findings



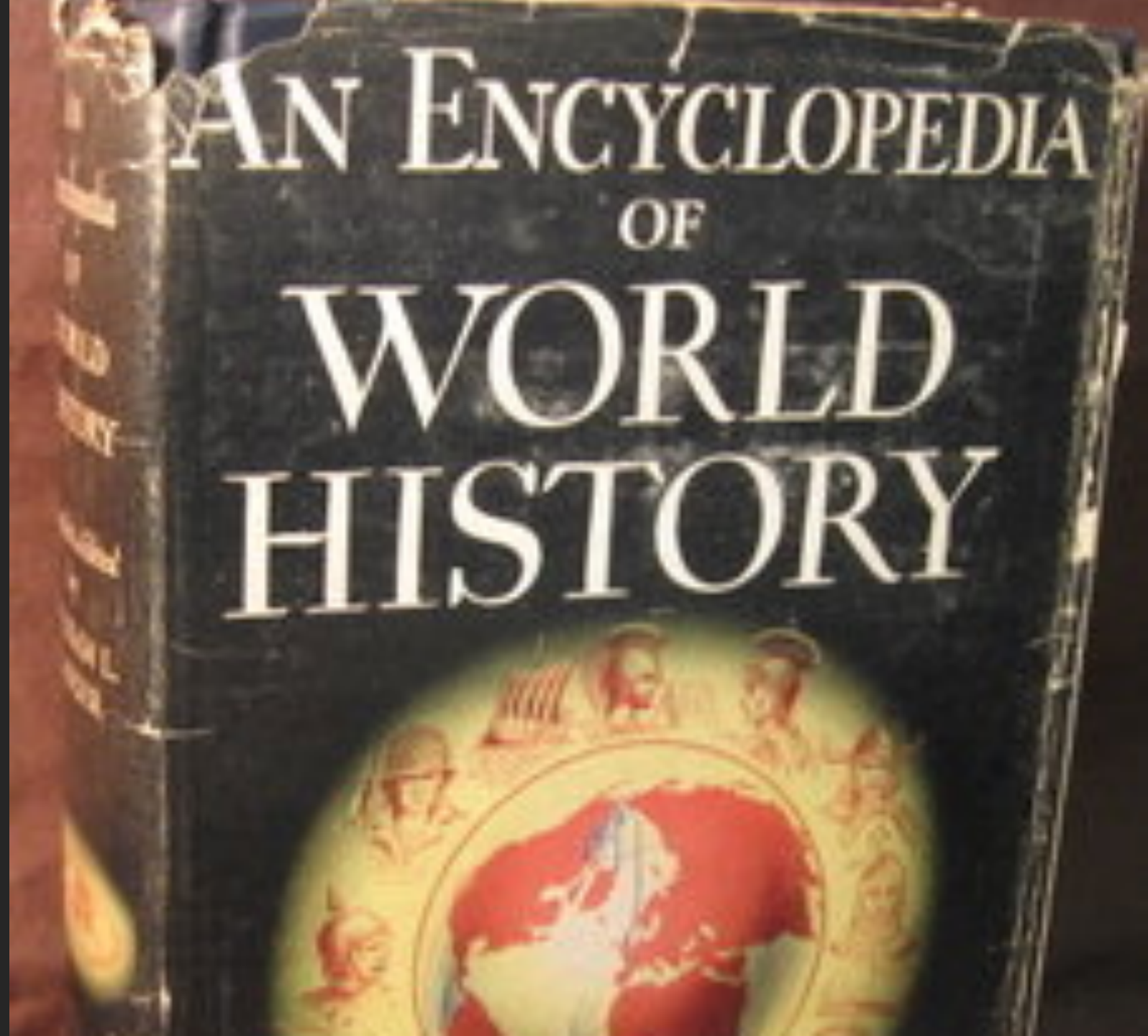
Conclusion



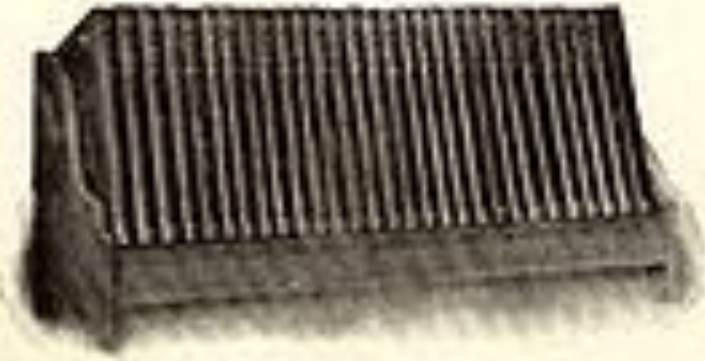
---

# INTRODUCTION

“ENCYCLOPEDIA,” AS A GENRE, IS A SIGNIFICANT RESOURCE FOR RESEARCHERS ACROSS THE HUMANITIES, ARTS, AND SCIENCES.



WHEN IN DOUBT—"LOOK IT UP" IN  
The  
*Encyclopaedia Britannica*



(New 11th Edition issued 1910-11 by the  
CAMBRIDGE UNIVERSITY PRESS (England))

The Sum of Human  
Knowledge

29 volumes, 28,150 pages,  
44,000,000 words of text.  
Printed on thin, but strong  
opaque India paper, each  
volume but one inch in  
thickness.

THE BOOK TO ASK QUESTIONS OF

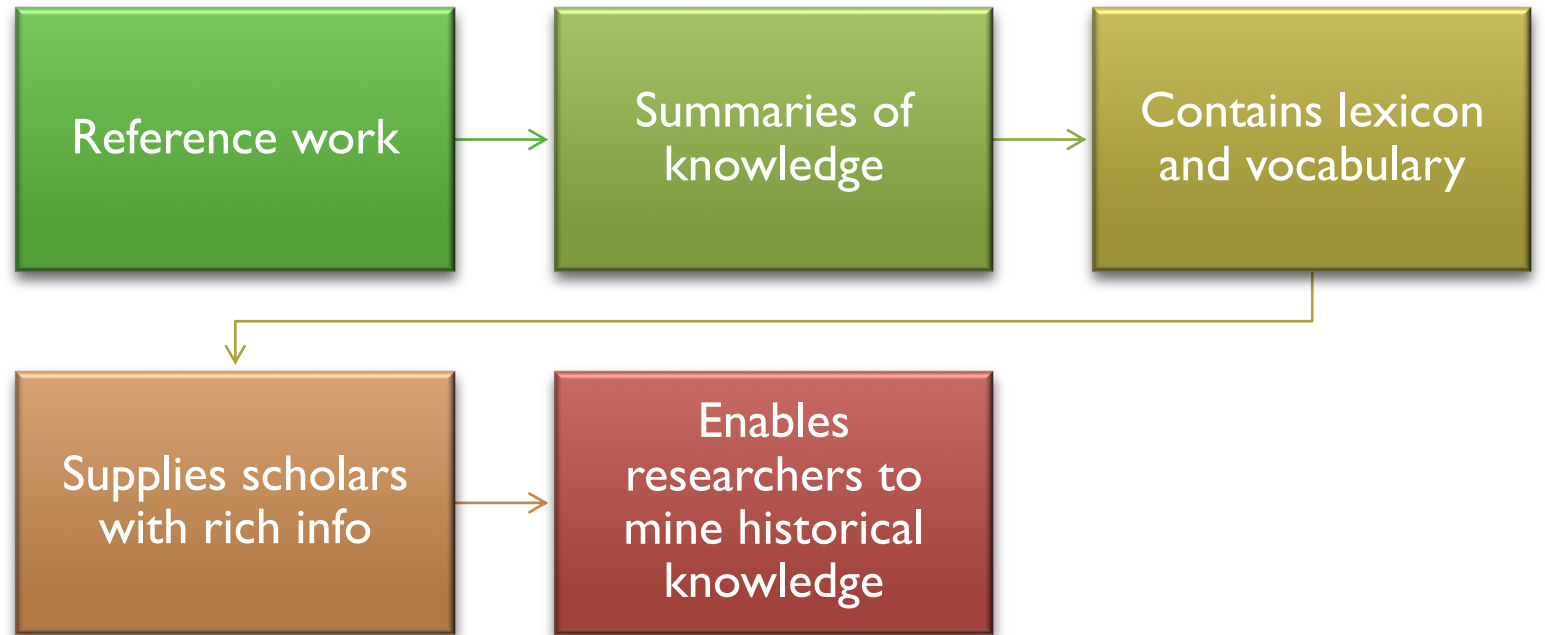
FOR READING OR FOR STUDY

This Photo by Unknown author is licensed under [CC BY-SA](#).

## INTRODUCTION

RESEARCHERS CAN USE  
ENCYCLOPEDIAS PRODUCED  
DURING EARLIER TIME-PERIODS  
FOR STUDY BECAUSE THEY  
PROVIDE ACCOUNTS NOT  
ONLY OF HISTORY BUT ALSO  
EVIDENCE OF HOW  
KNOWLEDGE WAS PERCEIVED  
AND ORGANIZED.

# INTRODUCTION - ENCYCLOPEDIA



This Photo by Unknown author is licensed under [CC BY-SA-NC](https://creativecommons.org/licenses/by-sa/4.0/).



## BACKGROUND OF THE STUDY

The 19th Century Knowledge Project is digitizing historical editions of the Encyclopedia Britannica (1797-1911)

This initiative will offer one of the most extensive, open, digital collections available today for studying the structure of 19th-century knowledge and its transformation

The metadata activities are being pursued through a collaboration with the Metadata Research Center (MRC) at Drexel University, supported by the U.S. National Endowment for the Humanities (NEH).



Digital Scholarship Center (DSC) of Temple University Libraries



MRC researchers are pursuing automatic methods of providing subject access for the Encyclopedia Britannica entries

## WHY SKOS?

- Simple Knowledge Organization System is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web.

## WHY SKOS?

- SKOS & RDF
- SKOS provides a standard way to represent knowledge organization systems using the Resource Description Framework (RDF). Encoding this information in RDF allows it to be passed between computer applications in an interoperable way.
- Using RDF also allows knowledge organization systems to be used in distributed, decentralised metadata applications. Decentralised metadata is becoming a typical scenario, where service providers want to add value to metadata harvested from multiple sources.





# Conceptual Model for the Encyclopedia Britannica Controlled Vocabulary



<b>Digitization</b> — OCR DocX	<b>SKOS</b> — XML RDF	<b>Hive Integration</b> — Python Rake
---	--------------------------------	--

**METHODOLOGY**

**SKOS of 1910 Library of Congress  
Subject Heading**

Automatic  
Indexing



Subject  
Cataloging



**Articles of 19<sup>th</sup>  
Century Encyclopedia  
Britannica**

**SKOS of the 1910 LCHS for the Transformation of Keywords  
to Controlled Vocabulary of the 19th Century Encyclopedia  
Britannica**

**Encyclopedia  
Britannica  
Controlled  
Vocabulary**

## APPROACH

Exploration – the current data set (in Docx) is explored and converted to different format – RDF/XML, relational tables, TEI for testing and evaluation

Testing – different platform is tried for evaluation of alternative solution

Evaluation – based from the results of different implementations, best solution is chosen

Documentation – Technical manual, paper and codes are prepared as deliverables



## METHODS

- Digitization – The 1910 LCSH provided in the project was in Docx and TEI version
- Encoding – Codes in Python were written including parsers to convert the TEI to RDF/XML format and def function to connect and create the 1910 LCSH database and insert records that built the 1910 LCSH schema.
- Programming
  - Characterizing the states and nature of the entries to define the objects
  - Enumerating the patterns for composition of the conditional statement
  - Data set was generated raw from the OCR process, thus the pattern was hardly identified for logic formulation.
- Digitalization – (Application Profiles) MultiTes and Python Program
  - MultiTes usage which was manual in process but yields 98% accuracy in terms of representation
  - Building of a program (Python) to automate the SKOS creation from TEI format to RDF/XML format. This yielded higher percentage of error which were identified from the inconsistencies found in the evaluation conducted when the control structures of the program were constructed. Further investigation could verify the percent error yield once compared to MultiTes version of SKOS RDF/XML.
- Metadata – Representation of SKOS elements to respective fields of the HIVE database. Tables are confined to BROADERS, RELATED and CONCEPT which has the following fields: ConceptURI, PrefLabel, AltLabel and ScopeNotes. USE, USE FOR, BT and NT are not represented because HIVE database has no provision for them.

Abacus.

QA135

Abandoned children. *See* Children—*Charities, protection, etc.*;  
Foundlings; Orphans and orphan-asylums.

Abattoirs. *See* Slaughtering and slaughter-houses.

**Abbeys.**

NA4800–6113 (Architecture)

*See also* Cathedrals; Convents; Monasteries.

Abbots.

Abbreviations.

Z111 (Paleography)

*See also* Cipher; Shorthand.

Abbreviations, English, [French, Hebrew, etc.]

# RESULTS AND FINDINGS

DOCX AND TEI FORMAT OF  
THE 1910 LIBRARY OF  
CONGRESS SUBJECT  
HEADINGS

```

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
>

<skos:Concept rdf:about="http://mysite.com/#Abacus"><prefLabel>Abacus</prefLabel>
<skos:notes> QA135
</skos:notes>
</skos:Concept>
<skos:Concept rdf:about="http://mysite.com/#Abandoned children"><prefLabel>Abandoned children</prefLabel>
<USE>Abandoned children.</USE>
</skos:Concept>
<skos:Concept rdf:about="http://mysite.com/#Abattoirs"><prefLabel>Abattoirs</prefLabel>
<USE>Abattoirs.</USE>
</skos:Concept>
<skos:Concept rdf:about="http://mysite.com/#Abbeys"><prefLabel>Abbeys</prefLabel>
<skos:notes> NA4800-6113 (Architecture)
</skos:notes>
<RT>Cathedrals ; Convents ; Monasteries.</RT>
</skos:Concept>
<skos:Concept rdf:about="http://mysite.com/#Abbots"><prefLabel>Abbots</prefLabel>
</skos:Concept>
<skos:Concept rdf:about="http://mysite.com/#Abbreviations"><prefLabel>Abbreviations</prefLabel>
<skos:notes> Z111 (Paleography)
</skos:notes>
<RT>Cipher ; Shorthand.</RT>
</skos:Concept>
<skos:Concept rdf:about="http://mysite.com/#Abbreviations, English, [French, Hebrew, etc]"><prefLabel>
Abbreviations, English, [French, Hebrew, etc</prefLabel>
</skos:Concept>
<skos:Concept rdf:about="http://mysite.com/#Abdomen"><prefLabel>Abdomen</prefLabel>
<skos:notes> QM543 (Regional anatomy)
</skos:notes>
<RT>Groin; Intestines; Kidneys; Liver; Peritoneum; Stomach.</RT>
</skos:Concept>
<skos:Concept rdf:about="http://mysite.com/#Abdomen-Diseases"><prefLabel>Abdomen-Diseases</prefLabel>
<skos:notes> RC941
</skos:notes>
</skos:Concept>

```

# SKOS OF THE 1910 LIBRARY OF CONGRESS SUBJECT HEADINGS

RDF / XML

MACHINE-READABLE FORMAT



## SoniaTest

Report generator

### Welcome to the Online Thesaurus System

To find a term, click on one of the letters below.

If you have comments or would like to discuss a term, click on the "feedback" link in the lower part of the term's page.

Thank you,  
Thesaurus creator.

A C E F G I K L M O P S

Send comments to [youremail@yourdomain.com](mailto:youremail@yourdomain.com) or visit our page at [www.yourdomain.com](http://www.yourdomain.com)

Created with MultiTes Pro

SKOS OF THE 1910  
LIBRARY OF  
CONGRESS SUBJECT  
HEADINGS

HTML VERSION

## SoniaTest

Report generator

### A

- [Abacus](#)
- [Abandoned children](#)
- [Abattoirs](#)
- [Abbeys](#)
- [Abbots](#)
- [Abbreviations](#)
- [Abbreviations, English \[French, Hebrew, etc.\]](#)
- [Abdomen](#)
- [Abdomen--Disease](#)
- [Abdomen--Surgery](#)
- [Abdomen--Tumors](#)
- [Abduction](#)
- [Abelian equations](#)
- [Abelian functions](#)
- [Abelian funstions](#)
- [Aberdeen-Angus cattle](#)
- [Aberration](#)
- [Aberration, Chromatic and spherical](#)
- [Ability](#)
- [Ability, Distribution of](#)
- [Abipone, Indians](#)
- [Abnaki language](#)
- [Abnaki, Indians](#)
- [Abnormal children](#)
- [Achromatism](#)
- [Amerescogin Indians](#)

## SoniaTest

Report generator

<< [Abbreviations, English \[French, Hebrew, etc.\]](#) | [Abdomen](#) | [Abdomen--Disease](#) >>

Back to: "A"

### Abdomen

- RT:** [Groin](#)  
[Intestines](#)  
[Kidneys](#)  
[Liver](#)  
[Peritoneum](#)  
[Stomach](#)

**Personal note:** [QM543 \(Regional anatomy\)](#)

[Search in Google](#) | [Search in Yahoo](#) | [Feedback](#)

Send comments to [youremail@yourdomain.com](mailto:youremail@yourdomain.com) or visit our page at [www.yourdomain.com](http://www.yourdomain.com)

Created with MultiTes Pro

# EVALUATION

Impeding question and looking at the solution for the project.

- How will the Docx format of 1910 LCHS be converted to RDF automatically?
- How will the Docx format of 1910 LCHS be loaded to HIVE DB automatically?

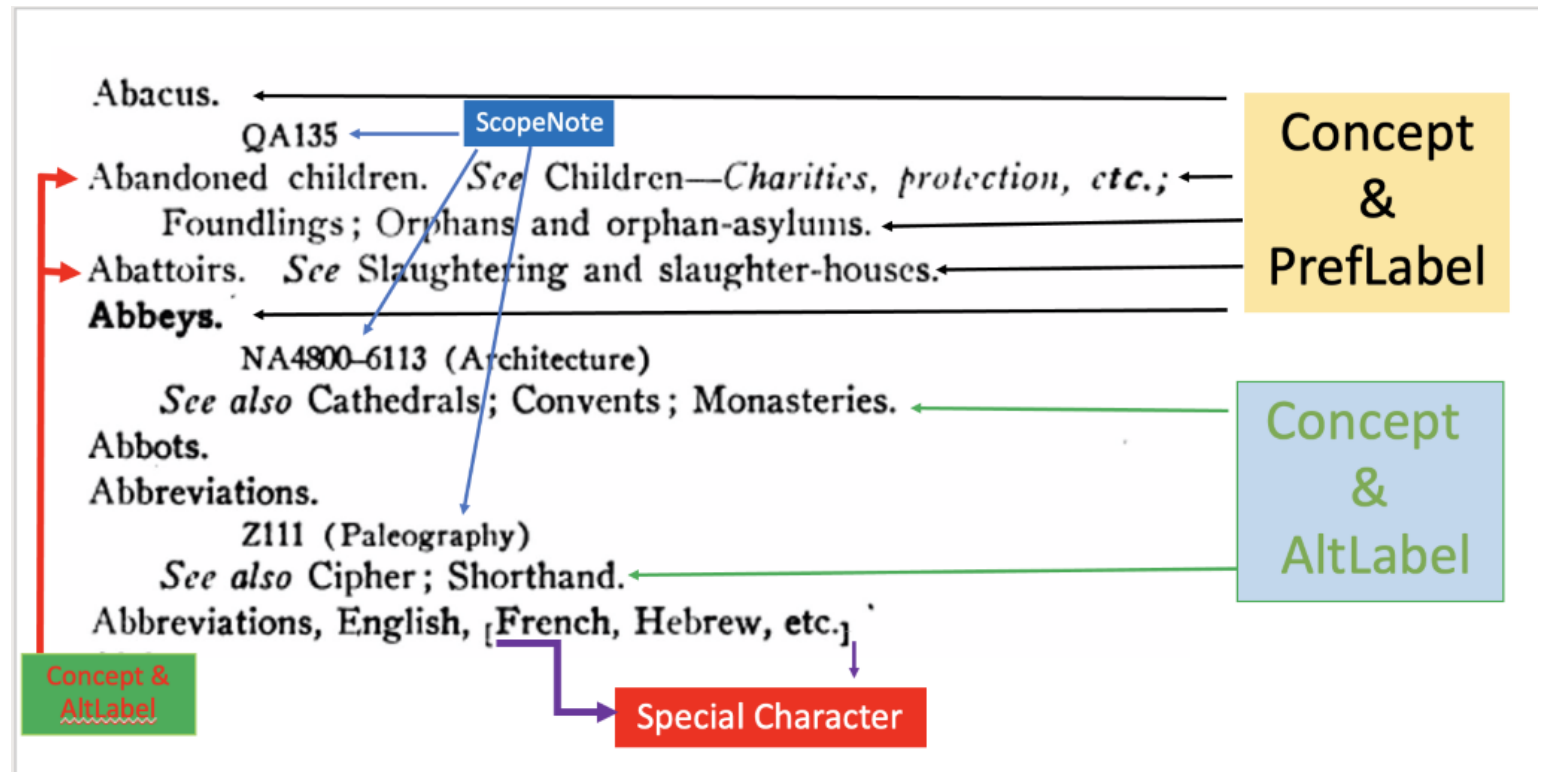
Concerns / Issues / Risks

- Which solution to take given the limited time
- SKOS in HIVE have limited elements of the standard SKOS

Pending action item

- To explore MultiTes in the automation of converting 1910 LCHS Doc to RDF
- To explore other tools in the automation of converting 1910 LCHS Doc to RDF
- To explore the HIVE code in the automation of loading 1910 LCHS DOC to HIVE db

# MAPPING



# MAPPING

- LCSH ruling and mapping to SKOS - RDF/XML and HIVE DB

LCSH Entries	SKOS - RDF/XML	HIVE DB
Subject entry	< <a href="#">skos:Concept</a> <a href="#">rdf:http://drexel.edu/lcsh1910/</a> #[Subject entry]>	<a href="#">ConceptURI</a>
Subject entry	< <a href="#">PrefLabel</a> >	<a href="#">PrefLabel</a>
SEE entry	< <a href="#">skos:Concept</a> <a href="#">rdf:http://drexel.edu/lcsh1910/</a> #[See entry]>	<a href="#">ConceptURI</a>
SEE entry	< <a href="#">PrefLabel</a> >	<a href="#">PrefLabel</a>
See also entry	< <a href="#">skos:Concept</a> <a href="#">rdf:http://drexel.edu/lcsh1910/</a> #[See also entry]>	<a href="#">ConceptURI</a>
See also entry	< <a href="#">PrefLabel</a> >	<a href="#">PrefLabel</a>
See also entry	< <a href="#">AltLabel</a> >	<a href="#">AltLabel</a>
Call numbers	< <a href="#">skos:Notes</a> >	<a href="#">ScopeNotes</a>
USE		
RT		
		<a href="#">TopConcept</a>
		<a href="#">NormPrefLabel</a>
	Other elements of SKOS	



Helping Interdisciplinary Vocabulary Engineering



Vocabularies Search Index

Vocabulary	Short name	Concepts	Last Updated
<a href="#">Asthma Ontology</a>	Asthma	289	03/02/2016
<a href="#">Cardiology</a>	Cardiology	155	07/25/2018
<a href="#">Combined Consumer Health Vocabulary</a>	CCHV	210153	08/08/2019
<a href="#">Diabetes Mellitus Diagnostic Ontology</a>	Diabetes	6439	12/20/2015
<a href="#">Ephraim Chambers Cyclopaedia</a>	Chambers	57	05/05/2019
<a href="#">Food and Agriculture Organization</a>	AGROVOC	35542	08/17/2018
<a href="#">Gastroenterology</a>	Gastroenterology	112	07/25/2018
<a href="#">Library of Congress Subject Heading 1910</a>	LCSH1910	23707	08/29/2019
<a href="#">Library of Congress Subject Headings</a>	LCSH	421572	07/26/2018
<a href="#">Medical Subject Headings</a>	MeSH	377824	08/25/2018
<a href="#">Metals</a>	Metals	44	01/01/2016
<a href="#">OCHV</a>	OCHV	87879	07/09/2019
<a href="#">Oncology</a>	Oncology	132	07/25/2018
<a href="#">Pediatrics</a>	Pediatrics	450	06/18/2018
<a href="#">Radiation Oncology</a>	ROO	1183	07/07/2015
<a href="#">Radiology Lexicon</a>	RADLEX	45471	11/16/2016
<a href="#">Respiratory</a>	Respiratory	142	07/25/2018
<a href="#">Smart Appliances REference ontology</a>	SAREF	112	02/10/2015
<a href="#">US Geological Survey</a>	USGS	968	01/01/2016
<a href="#">Unified Astronomy Thesaurus</a>	UAT	1843	01/31/2017

# 1910 LIBRARY OF CONGRESS SUBJECT HEADING IN HIVE



Alert Changes | Open Project | Save Project | Attach Database | Close Database

data | Edit Pragas | Execute SQL

New Record | Delete Record

AltLabel	ScopeNotes	TopConcept	NormPrefLabel
Filter	Filter	Filter	Filter
	QA135	0	Abacus
			Abandoned chil...
			Abattoirs
	NA4800-61...	0	Abbeys
			Abbots
	Z111 (Paleog...	0	Abbreviations
, E...			Abbreviations, E...
	QM543 (Regi...	0	Abdomen
ise...	RC941	0	Abdomen—Dise...
urg...	RD540-7	0	Abdomen—Surg...
im...	RD667	0	Abdomen—Tum...
	HV6571-4	0	Abduction
ions			Abelian equations
ions			Abelian functions
gu...	SF199	0	Aberdeen-Angu...
	QC671 (Phys...	0	Aberration
hr...			Aberration, Chr...
			Ability
out...			Ability, Distribut...
ins			Abipone Indians
s			Abnaki Indians
			E99

Go to: 1

Edit Database Cell

Mode: Text | Import | Export | Set as NULL

rd:about=http://www.drexel.edu/lchs1910/#Abacus|

Type of data currently in cell: Text / Numeric  
48 char(s)

Apply

SQL Log

Show SQL submitted by Application | Clear

```

1 PRAGMA foreign_keys = '1';
2 PRAGMA database_list;
3 SELECT type,name,sql,tbl_name FROM "main".sqlite_master;
4 PRAGMA encoding
5 SELECT COUNT(*) FROM (SELECT "_rowid_","* FROM "main"."BROADERS" ORDER B
6 SELECT "_rowid_","* FROM "main"."BROADERS" ORDER BY "_rowid_" ASC LIMIT 0, 4
7 SELECT COUNT(*) FROM (SELECT "_rowid_","* FROM "main"."CONCEPT" ORDER BY
8 SELECT "_rowid_","* FROM "main"."CONCEPT" ORDER BY "_rowid_" ASC LIMIT 0, 4
9

```

SQL Log | Plot | Remote

# 1910 LIBRARY OF CONGRESS SUBJECT HEADING IN HIVE

DATABASE



Helping Interdisciplinary  
Vocabulary Engineering



Vocabularies Search Index

Search one or more vocabularies for a term or phrase.

1 Select vocabularies

- AGROVOC
- Diabetes
- Metals
- ROO
- Asthma
- Gastroenterology
- OCHV
- Respiratory
- CCHV
- LCSH
- Oncology
- SAREF
- Cardiology
- LCSH1910**
- Pediatrics
- UAT
- Chambers
- MeSH
- RADLEX
- USGS

2 Enter search term

Search

**LCSH1910**

[Abattoirs](#)

List

JSON-LD

SKOS RDF/XML

Dublin Core

XML

**Preferred label** [Abattoirs](#)

**URI** [rdf:about=http://www.drexel.edu/lchs1910/#Abattoirs](http://www.drexel.edu/lchs1910/#Abattoirs)

**Alternate label**

**Notes label**

**Broader** No broader concepts

**Narrower** No narrower concepts

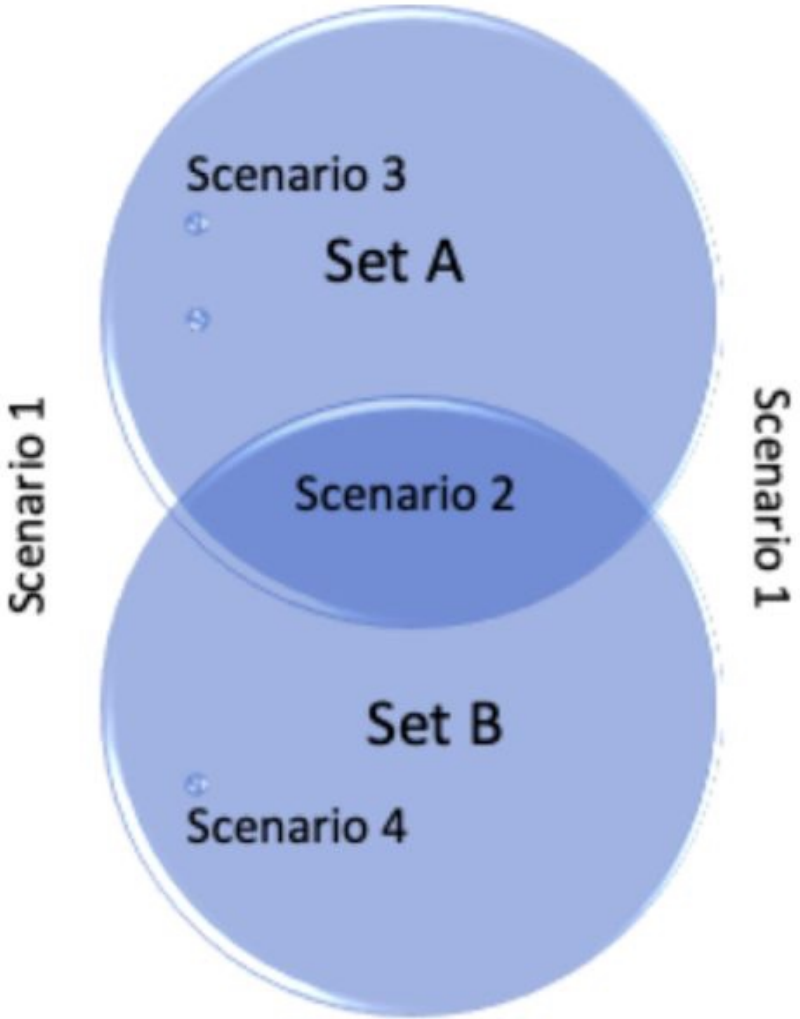
**Related** No related concepts

1910 LIBRARY OF  
CONGRESS  
SUBJECT  
HEADING IN HIVE

DATA SETS SPECIFICATIONS:

SET A: THE 19<sup>TH</sup> CENTURY  
ENCYCLOPEDIA BRITANNICA ARTICLES  
ARE INDEXED USING THE 1910 LCHS

SET B: THE 19<sup>TH</sup> CENTURY  
ENCYCLOPEDIA BRITANNICA ARTICLES  
ARE INDEXED USING 2016 LCHS



U  
S  
E  
C  
A  
S  
E  
S

Scenario 1: Union of Set A and Set B ( $A \cup B$ ) gives all the keywords found in the NCEB using both versions of LCHS. These keywords are all used in the transformation of the keywords to the controlled vocabulary of NCEB.

Scenario 2: Intersection of Set A and Set B ( $A \cap B$ ) gives the keywords found which are common only to both versions of LCHS. This data represents the keywords that are still being used from the 19th century until 2016.

Scenario 3: Difference of Set A and Set B ( $A - B$ ) gives the keywords found in the 1910 LCHS and could be deprecated already in the 2016 LCHS version or changed through time and has new terminology that replaces it.

Scenario 4: Difference of Set A and Set B ( $B - A$ ) gives the keywords found in the 2016 LCHS and but could be non-existent yet in the 1910 LCHS version. This could also be the new keywords used in the later century.

# SUBJECT ANALYSIS

I. The text format of the article sun from the 19<sup>th</sup> Century Encyclopedia Britannica was subjected to text analysis using TagCrowd.

TagCrowd

Create your own word cloud from any text to visualize word frequency.

[Start Over](#) [Help](#) [News](#) [Contact](#) [Commercial Use](#)

Showing top 50 of 1224 possible words

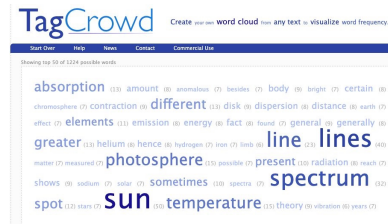
absorption (13) amount (8) anomalous (7) besides (7) body (9) bright (7) certain (8)  
chromosphere (7) contraction (9) different (13) disk (9) dispersion (8) distance (8) earth (7)  
effect (7) elements (11) emission (8) energy (8) fact (8) found (7) general (9) generally (8)  
greater (13) helium (8) hence (8) hydrogen (7) iron (7) limb (6) line (23) lines (40)  
matter (7) measured (7) photosphere (15) possible (7) present (10) radiation (8) reach (7)  
shows (9) sodium (7) solar (7) sometimes (10) spectra (7) spectrum (32)  
spot (12) stars (7) sun (50) temperature (15) theory (9) vibration (6) years (7)



# SUBJECT ANALYSIS

1. The text format of the article sun from the 19<sup>th</sup> Century Encyclopedia Britannica was subjected to text analysis using TagCrowd.

2. To apply the “Aboutness” approach to subject analysis, the following words were determined as the keywords. 16 out of 51 descriptors were selected as the keywords. They were spotted to reflect both the richness of the topic being discussed in the article and the expressiveness they contain in relation to the word “SUN”.



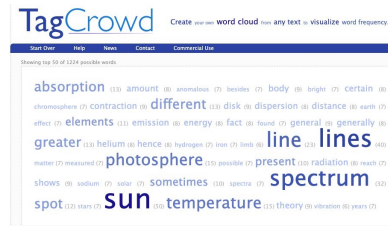
Article (19th Century Encyclopedia Britannica)	Tag Crowd	Word Frequency	Aboutness
Sun	Sun	50	Sun
	Lines	40	Lines
	Spectrum	32	Spectrum
	Line	23	
	Photosphere	15	Photosphere
	Temperature	15	Temperature
	Absorption	13	Absorption
	Different	13	
	Greater	13	
	Spot	12	Spot
	Elements	11	Elements
	Present	10	
	Sometimes	10	
	Body	9	Body
	Contraction	9	Contraction
	Disk	9	Disk
	General	9	
	Shows	9	
	Theory	9	
	Amount	8	
	Certain	8	
	Dispersion	8	Dispersion
	Distance	8	Distance
	Emission	8	Emission
	Energy	8	Energy

# SUBJECT ANALYSIS

1. The text format of the article sun from the NCEB was subjected to text analysis using TagCrowd.

2. To apply the “Aboutness” approach to subject analysis, the following words were determined as the keywords. 16 out of 51 descriptors were selected as the keywords. They were spotted to reflect both the richness of the topic being discussed in the article and the expressiveness they contain in relation to the word “SUN”.

3. For the purpose of simulation, manual indexing using 1910 LCSH (OCR version) was performed capturing the structure, subdivisions and subject headings for the index of the article “SUN”



Article (19th Century Encyclopedia Britannica)	Tag Crowd	Word Frequency	Aboutness
Sun	Sun	50	Sun
	Lines	40	Lines
	Spectrum	32	Spectrum
	Line	23	
	Photosphere	15	Photosphere
	Temperature	15	Temperature
	Absorption	13	Absorption
	Different	13	
	Greater	13	
	Spot	12	Spot
	Elements	11	Elements
	Present	10	
	Sometimes	10	
	Body	9	Body
	Contraction	9	Contraction
	Disk	9	Disk
	General	9	
	Shows	9	
	Theory	9	
	Amount	8	
	Certain	8	
	Dispersion	8	Dispersion
	Distance	8	Distance
	Emission	8	Emission
	Energy	8	Energy

**Absorption (Physiology)**

**QP88**

*See also* Digestion ; Fat ; Lymphatics ; Osmosis ; Skin.

*Absorption, Atmospheric. See* Solar radiation.

**Absorption of light.**

**QC437**

**Absorption spectra.**

**QC437**

*See also* Heat—Radiation and absorption; Spectrum analysis.

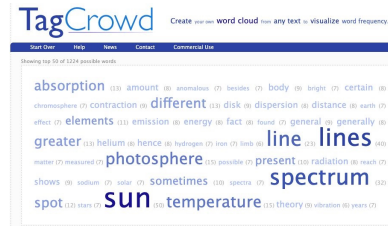
# SUBJECT ANALYSIS

1. The text format of the article sun from the NCEB was subjected to text analysis using TagCrowd.

2. To apply the “Aboutness” approach to subject analysis, the following words were determined as the keywords. 16 out of 51 descriptors were selected as the keywords. They were spotted to reflect both the richness of the topic being discussed in the article and the expressiveness they contain in relation to the word “SUN”.

3. For the purpose of simulation, manual indexing using 1910 LCSH (OCR version) was performed capturing the structure, subdivisions and subject headings for the index of the article “SUN”

4. The same article was uploaded to HIVE for automatic indexing, generating the 2016 LCSH version of the keyword “Absorption”



Article (19th Century Encyclopedia Britannica)	Tag Crowd	Word Frequency	Aboutness
Sun	Sun	50	Sun
	Lines	40	Lines
	Spectrum	32	Spectrum
	Line	23	
	Photosphere	15	Photosphere
	Temperature	15	Temperature
	Absorption	13	Absorption
	Different	13	
	Greater	13	
	Spot	12	Spot
	Elements	11	Elements
	Presents	10	
	Sometimes	10	
	Body	9	Body
	Contraction	9	Contraction
	Disk	9	Disk
	General	9	
	Shows	9	
	Theory	9	
	Amount	8	
	Certain	8	
	Dispersion	8	Dispersion
	Distance	8	Distance
	Emission	8	Emission
	Energy	8	Energy

Absorption (Physiology)

QP88

*See also* Digestion ; Fat ; Lymphatics ; Osmosis ; Skin.

Absorption, Atmospheric. *See* Solar radiation.

Absorption of light.

QC437

Absorption spectra.

QC437

*See also* Heat—Radiation and absorption ; Spectrum analysis.



Helping Interdisciplinary Vocabulary Engineering



DREXEL UNIVERSITY  
Metadata Research Center  
College of Computing & Informatics

Vocabularies Search Index

HIVE automatically extracts concepts from a file, or URL, using selected vocabularies.

1 Select vocabularies

- AGROVOC
- Diabetes
- Metals
- ROO
- USGS
- Asthma
- Gastroenterology
- Oncology
- Respiratory
- Cardiology
- LCSH
- Pediatrics
- SAREF
- Chambers
- MeSH
- RADLEX
- UAT

2 Enter a URL, or select a file, to index Indexing...

URL

or

Sun.docx

3 Select indexing filters (optional)

LCSH

Sun


Temperature Temperature Contracts Contraction Sodium Founding Bodies Helium Case

Cases Cases Case Iron Hydrogenation Hydrogen Measurement Measurement Distances Calcium Leveling Levellers Atmosphere Atmospheric Absorption Stars Red Dues Regionalism Regions Waves Light Lighting Lighting Titanium Losses Heat Heating Heating Sources Sources Scandium Vibration Vibrators Vibration Inference Magnesium Chromium Oxygen Oxygenators

 Cloud View

 Rank Order

 List View

 Alpha Order

**LCSH**

**Sun** Temperature

Temperature Contracts

Contraction Sodium Founding

Bodies Helium Case Cases

Cases Case Iron

Hydrogenation Hydrogen

Measurement Measurement

Distances Calcium Leveling

Levellers Atmosphere Atmospherics

**Absorption** Stars Red Dues

Regionalism Regions Waves Light

Lighting Lighting Titanium Losses

List

JSON-LD

SKOS RDF/XML

Dublin Core

XML

**Preferred label** Absorption

**URI** <http://id.loc.gov/authorities/subjects/sh85000245>

**Alternate label** Sorption;

**Notes label** Not provided

**Broader**

Chemistry, Physical and theoretical  
Packed towers

**Narrower**

Gases--Absorption and adsorption  
Light absorption  
Photoabsorption  
Sorbents

**Related** No related concepts

RESULTS FROM THE  
2016 LIBRARY OF CONGRESS SUBJECT  
HEADING IN HIVE



Article (19th Century Encyclopedia Britannica)	Tag Crowd	Word Frequency	Aboutness	1910 LCHS Manual Subject Indexing	Output Controlled Vocabulary	LCHS in Hive	Scenario
Sun	Sun	50	Sun				
	Absorption	13	Absorption	*Absorption (Physiology) *Absorption, Atmospheric See Solar -- Radiation Absorption of light *Absorption spectra See also Heat -- Radiation a	Absorption of light	<b>Preferred label: Absorption</b>  <b>URI <a href="http://id.loc.gov/authorities/subjects/sh85000245">http://id.loc.gov/authorities/subjects/sh85000245</a></b>  <b>Alternate label Sorption</b>  <b>Notes label Not provided</b>  <b>Broader</b> Chemistry, Physical and theoretical Packed towers  <b>Narrower</b> Gases--Absorption and adsorption Light absorption Photoabsorption Sorbents  <b>Related No related concepts</b>	Scenario 2

## WORD ANALYSIS MATRIX

Table above is a word analysis matrix for the descriptor “Absorption” that shows the result of the subject analysis conducted in the article “SUN”. The simulation of the word – Absorption fell in scenario 2 of the use cases. This means that the word “Absorption” intersects both data sets, thus it exists from 1910 till 2016.

Article (19th Century Encyclopedia Britannica)	Tag Crowd	Word Frequency	Aboutness	1910 LCHS Manual Subject Indexing	Output Controlled Vocabulary	LCHS in Hive	Scenario
Sun	Sun	50	Sun				
	Absorption	13	Absorption	*Absorption (Physiology) *Absorption, Atmospheric See Solar -- Radiation Absorption of light *Absorption spectra See also Heat -- Radiation a	Absorption of light	<b>Preferred label: Absorption</b> URI <a href="http://id.loc.gov/authorities/subjects/sh85000245">http://id.loc.gov/authorities/subjects/sh85000245</a> <b>Alternate label</b> Sorption <b>Notes label</b> Not provided <b>Broader</b> Chemistry, Physical and theoretical Packed towers <b>Narrower</b> Gases--Absorption and adsorption Light absorption Photoabsorption Sorbents <b>Related</b> No related concepts	Scenario 2

☁ Cloud View
★ Rank Order

☰ List View
⬇ Alpha Order

**LCSH**

Sun

Temperature

Temperature Contracts

Contraction Sodium Founding

Bodies Helium Case Cases

Cases Case Iron

Hydrogenation Hydrogen

Measurement Measurement

Distances Calcium Leveling

Levellers Atmosphere Atmospheric

Absorption Stars Red Dues

Regionalism Regions Waves Light

Lighting Lighting Titanium Losses

Heat Heating Heating Sources

Sources Scandium Vibration

Vibrators Vibration Inference

Magnesium Chromium Oxygen

Oxygenators

**Preferred label** [Absorption](#)

**URI** <http://id.loc.gov/authorities/subjects/sh85000245>

**Alternate label** Sorption;

**Notes label** Not provided

**Broader**

Chemistry, Physical and theoretical  
Packed towers

**Narrower**

Gases--Absorption and adsorption  
Light absorption  
Photoabsorption  
Sorbents

**Related** No related concepts

# WORD ANALYSIS MATRIX

Table above is a word analysis matrix for the descriptor “Absorption” that shows the result of the subject analysis conducted in the article “SUN”. The simulation of the word – Absorption fell in scenario 2 of the use cases. This means that the word “Absorption” intersects both data sets, thus it exists from 1910 till 2016.

## CONCLUSIONS

The use of both versions of the LCHS, 1910 and 2016, as a controlled vocabulary to extract keywords from the articles of 19<sup>TH</sup> Century Encyclopedia Britannica purposively rendered a wider consideration in the establishment of the 19<sup>TH</sup> Century Encyclopedia Britannica controlled vocabulary. Use cases showed the inclusive generation of keywords producing substantial vocabularies.

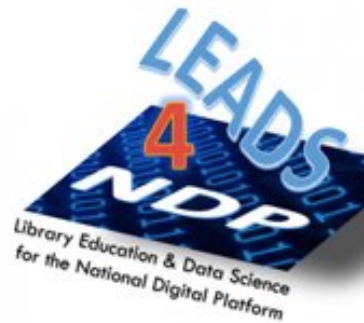
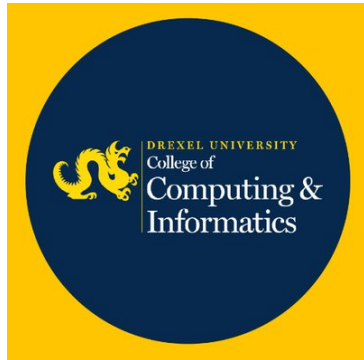
Scenarios put forward the impositions of different inferences of the data.

- Scenario 1 was the integration of all descriptors that matched to the vocabulary of the 2 versions of LCHS.
- Scenario 2 captured all the descriptors that commonly matched with both versions of LCHS. This meant that these words were used in the 19<sup>th</sup> century until 2016.
- Scenario 3 represented the words that were not existent anymore in the present day since it's only found as a vocabulary in the 1910 LCHS.
- On the other hand, scenario 4 is telling us that there were new terms formulated by men in the next centuries that came. Thus using the old LCHS or the appropriate version of LCHS would yield higher percent of accuracy in terms of completeness of 19<sup>th</sup> Century Encyclopedia Britannica keywords to be transformed to a controlled vocabulary. The goal is to take simulations of all scenarios to be able to justify the hypothesis and to conduct manual indexing using the current LCHS (2016) which serve another use case.

# FUTURE WORKS

GRAPH  
DATABASE OF  
HIVE



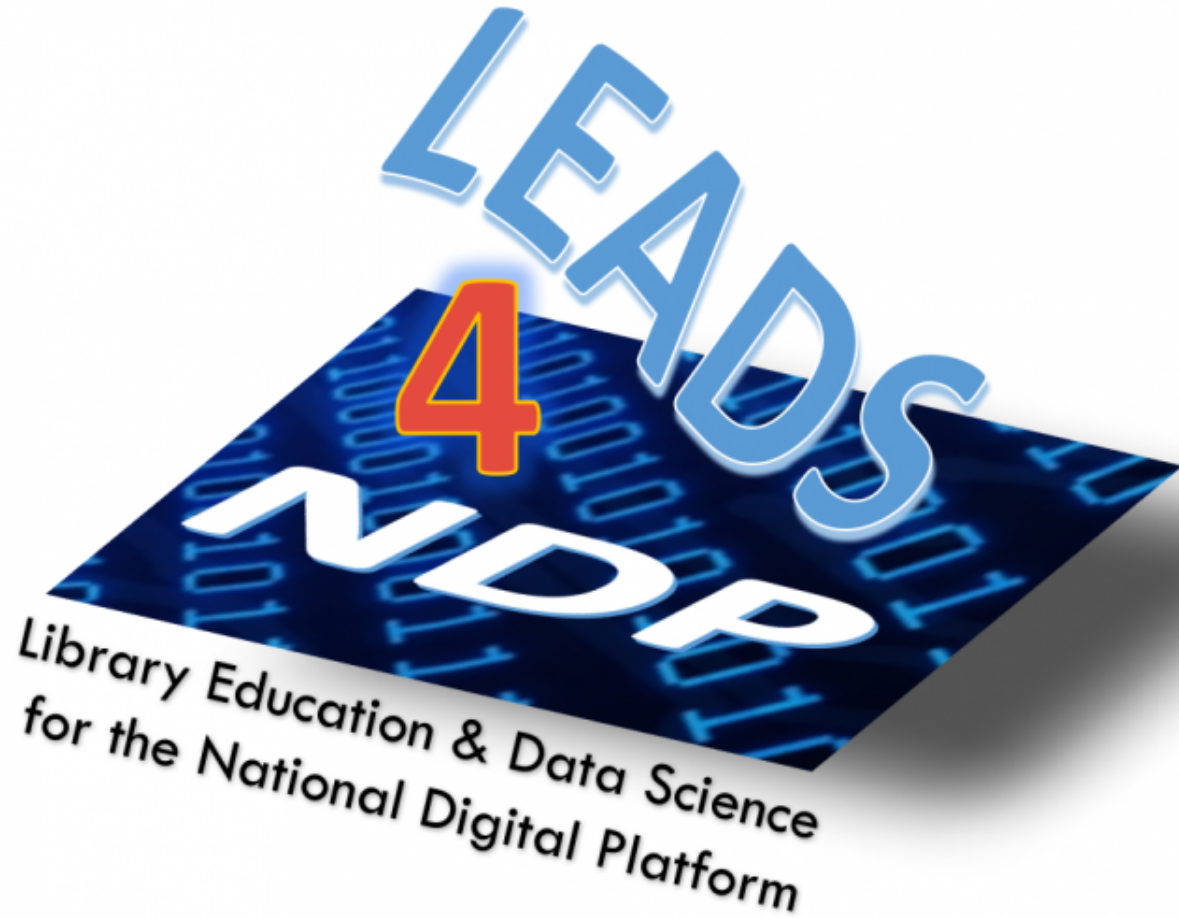


## REFERENCES

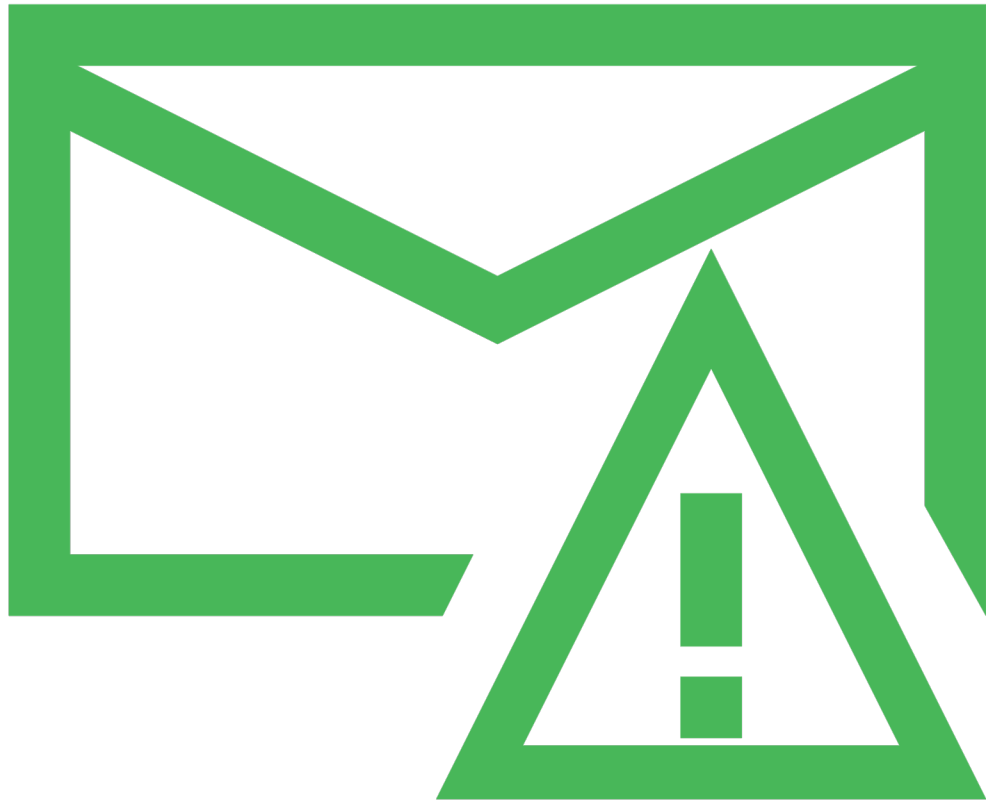
- Encyclopedia. (2019, June 14). Retrieved June 16, 2019, from <https://en.wikipedia.org/wiki/Encyclopedia> HIVE2. (n.d.). Retrieved June 16, 2019, from <http://hive2.cci.drexel.edu:8080/> Greenberg, J., Losee, R., Pérez Agüera, J. R., Scherle, R., White, H., & Willis, C. (2011). HIVE:
- Helping Interdisciplinary Vocabulary Engineering. *Bulletin of the American Society for Information Science and Technology*, 37(4): 23-26.
- Library of Congress Subject Headings. (2019, January 27). Retrieved June 16, 2019, from [https://en.wikipedia.org/wiki/Library\\_of\\_Congress\\_Subject\\_Headings](https://en.wikipedia.org/wiki/Library_of_Congress_Subject_Headings)
- Library of Congress Subject Headings. (n.d.). Retrieved June 16, 2019, from <https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html>
- Logan, P.M. Nineteenth-Century Knowledge Project. Retrieved June 16, 2019, from <https://tu-plogan.github.io/>
- Logan, P. M., Greenberg, J., & Grabus, S. (In Press). Knowledge Representation: Old, New, and Automated Indexing. In *Proceedings of Digital Humanities Conference 2019*, Utrecht, The Netherlands.

## ACKNOWLEDGEMENT

- LIS Education and Data Science for the National Digital Platform (LEADS-4-NDP)







---

SONIA PASCUA,  
DREXEL UNIVERSITY  
[SMP458@DREXEL.EDU](mailto:SMP458@DREXEL.EDU)

JANE GREENBERG  
DREXEL UNIVERSITY  
[JG3243@DREXEL.EDU](mailto:JG3243@DREXEL.EDU)

PETER LOGAN  
TEMPLE UNIVERSITY  
[PLOGAN@TEMPLE.EDU](mailto:PLOGAN@TEMPLE.EDU)

JOAN BOONE  
UNIVERSITY OF NORTH  
CAROLINA AT CHAPEL HILL  
[JPBOONE@AD.UNC.EDU](mailto:JPBOONE@AD.UNC.EDU)