EMERSE: A user-friendly text searching and natural language processing tool for electronic health record notes

Presentation for
Biomedical & Health Informatics Seminar
University of Iowa

December 11, 2025

Web: project-emerse.org
Email: hanauer@umich.edu
emerse-team@umich.edu

David Hanauer, MD, MS
Dept of Learning Health Sciences
University of Michigan

http://project-emerse.org/presentations.html

If you want to take notes or visit links

these slides can be found at:

this link will be
on most slides

Disclosures

Disclosures

Funding: NIH (NCI-ITCR, NCATS-CTSA)

U of Michigan Royalties/Licensing: "Synonyms" dataset–optional "plugin" for EMERSE.

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE
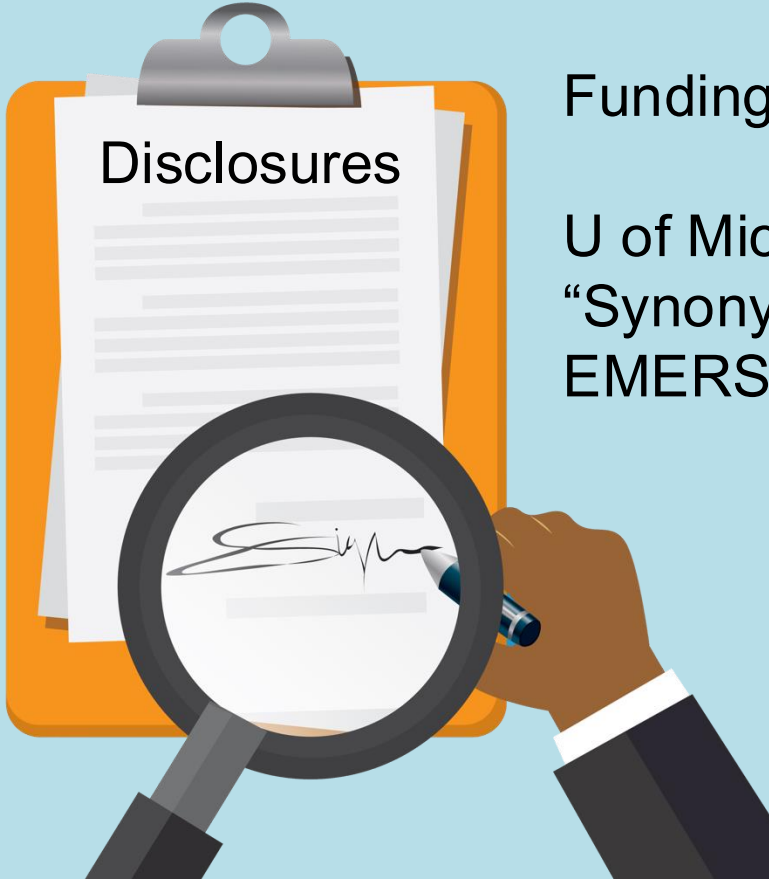
I am here to tell you about EMERSE

2021 study out of UC Irvine: *Design, Implementation, and Usability of the Electronic Medical Record Search Engine (EMERSE) Tool*
https://escholarship.org/uc/item/44p23878

"Users unanimously responded that they would recommend the system to others, and…for a tool they found so useful, they believed that **far too few people both within and outside of their network knew about the tool's existence**."

Unstructured vs Structured Data

EMERSE is for this…	…not this
Unstructured Data (free-text)	*Structured Data*
Mrs. Jones is a 56 year old female with a history of HTN, hypercholesterolemia, and T2DM who comes to the clinic today with a 3 day h/o dizziness and severe headache on the left side.	WBC: 5.6 Total cholesterol: 182 Weight: 67.4 AST: 30 ALT: 52

80% of EHR data are in unstructured free-text

structured data

unstructured data/free text

Most medical centers lack tools for free-text

And the ones that do exist aren't great

2024 study on EMR usability:

"The same three items (integration into workflow, **finding information**, and usability of alerts) received the highest number of 'poor' ratings among hospital and practice physicians."

https://www.nature.com/articles/s41746-025-01657-4

Do we even need search anymore?

Yes!

Perspective

Search still matters: information retrieval in the era of generative AI

William Hersh ID **, MD***

https://pmc.ncbi.nlm.nih.gov/articles/PMC11339511/pdf/ocae014.pdf
2024

The EMERSE solution

- A system "for the people"
- Users search the notes on their own
  - No need to wait in a queue for an analyst or a data scientist
- Easy-to-use for non-technical users
- Unlike with some EHRs, EMERSE can search across all notes and all patients at once
- Continuous refinements for almost 19 years

The EMERSE solution

- A system "for the admins"

- Enterprise grade, easy to support

- Configurable with granular roles/privileges to control access (e.g, all patients vs. some patients)

- Data are kept secure within a centralized, audited system *at your own site*

  – No need to download*/store data elsewhere

EMERSE is a useful tool

- It will be one of many you need to build/support your project(s)

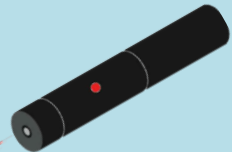https://doi.org/10.1093/jamia/ocad236

EMERSE is a mature tool

- EMERSE turns 20 tomorrow!

We think EMERSE is really good

- We're proud of what we created
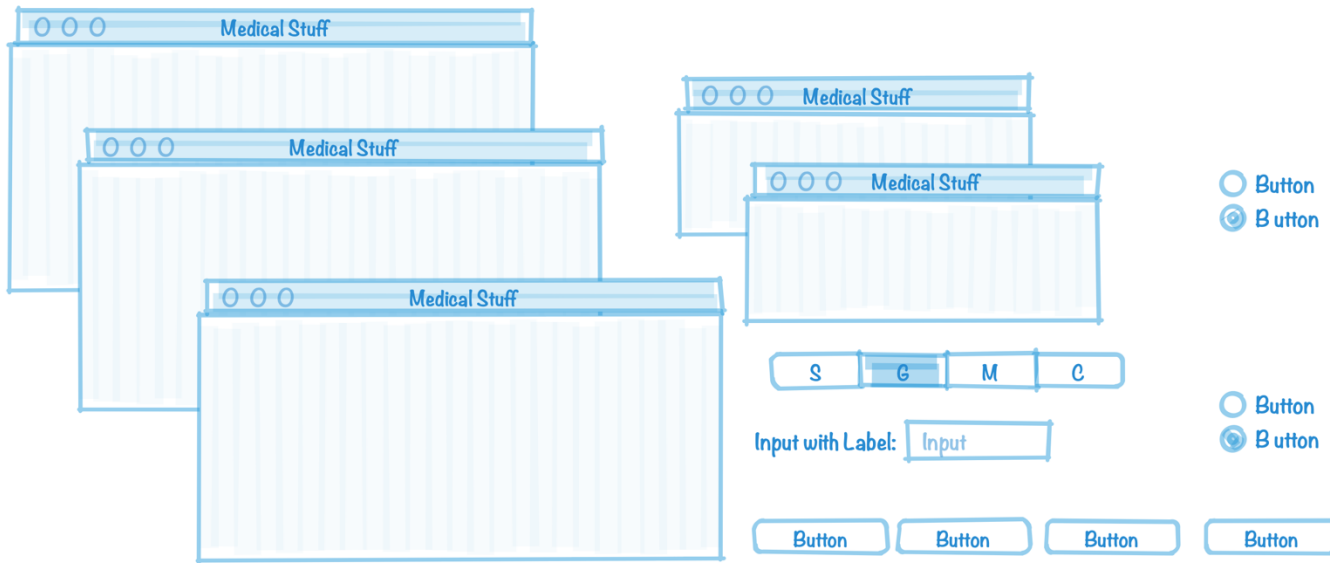- We think it's better than other software tools that cost $$$

emerse
ELECTRONIC MEDICAL RECORD   SEARCH ENGINE

http://project-emerse.org/presentations.html

Why do we think it's better?
We've been laser-focused on it

EHRs have thousands of features

ACME EHR

Search 🔍

Medical Stuff

Medical Stuff

Medical Stuff

Medical Stuff

Medical Stuff

○ Button
◉ Button

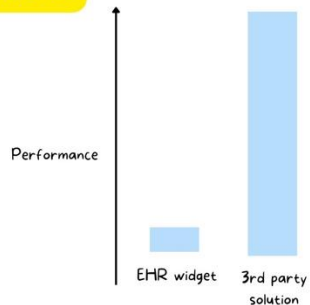| S | G | M | C |

Input with Label: Input

○ Button
◉ Button

Button   Button   Button   Button

The EMERSE team has invested $millions and 19 years on just this one feature (search) because it's so important

em⊇rse
ELECTRONIC MEDICAL RECORD   SEARCH ENGINE

Why do we think it's better?
It <u>has</u> to be better

It's not a health system's fault for hoping an EHR can provide everything.

3rd party solutions must be 10x better on a differentiator that truly matters.

Performance

EHR widget   3rd party solution

10/10

@joshuapliu

EMERSE is a "best of breed" product and must stay ahead of integrated EHR tools to remain relevant

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

Why do we think it's better?
We've studied it

- "new users of the EMERSE system are able to complete basic but critical workflow tasks in the system with a high rate of success…are **highly satisfied with the interface**, and have highly positive perceptions of its expected utility in their work."

- "survey results indicate **very high ratings of usability and satisfaction** with the EMERSE system by new users after only one session of use."

https://project-emerse.org/documents/reyes_masters_thesis_2021.pdf

Why do we think it's better?
Our users tell us

Thank you for…this important tool which is **proving extremely valuable** in enhancing patient safety and quality of care delivered at Michigan Medicine.

..continued gratitude for the availability of this powerful research tool. As always, **thank you for this innovation**!

EMERSE is **working out great for casefinding**...have found quite a bit of cases that we would have missed otherwise. **Thank you and everyone on your team!**

To say "**it is the most useful tool I use in my job**" would be an understatement, so **thank you** for creating and sharing it with others!!

EMERSE is **an absolute gem** @umichmedicine. The functionality is very friendly and it saved hours of time during the data collection process. **Thank you to the team** that created this powerful research tool!

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

Why do we think it's better?
Researchers mention it in their publications

- "Reviewers used the EMERSE search tool to **ensure thorough review** of the available documentation..." [PMID 36119396]

- "the **tool avoids the pitfalls of diagnostic inaccuracy** seen with tools querying on ICD and billing codes…" [PMID 36114099]

- "[EMERSE] provides software features to **comprehensively scan all clinical documents**…for keywords and phrases to ensure that even rarely mentioned events are detected." [PMID 36550198]

- "…information [was] captured via EMERSE…in order to **obtain the most accurate and complete information** per patient."[PMID 36752520]

Why do we think it's better?
Top-tier medical centers use it

Site	Site
U of Michigan	Case Western Reserve U
Harvard U – Dana Farber Cancer Center	Utrecht University, Netherlands
Columbia U Cancer Center	U of Iowa
U of North Carolina – Chapel Hill	Weill Cornell Medical Center
U of California – San Francisco	U of Virginia
U of Kentucky	Moffitt Cancer Center, Tampa, FL*
U of Cincinnati	Children's Health Orange County, Calif.*

IOWA

*Currently installing

What can EMERSE do?

Lots of things!

Cardiac Surgery Research

"It allows us to basically target exactly what we need to know."

--Linda Farhat, Clinical Research Coordinator

Quality Analytics

"I can't imagine having to do what I do without having EMERSE."

--Andrew Heiler, Quality Analytics Coordinator

Infection Prevention

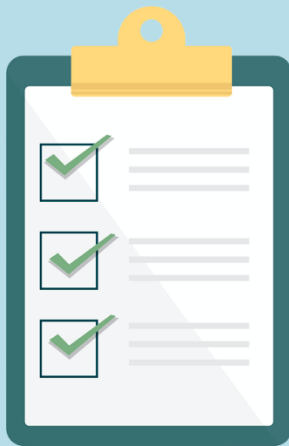"The value of EMERSE is immeasurable."

--Jayna Berger Heiler, Infection Prevention Coach

Watch our videos: https://project-emerse.org/use_cases.html

Find cohorts

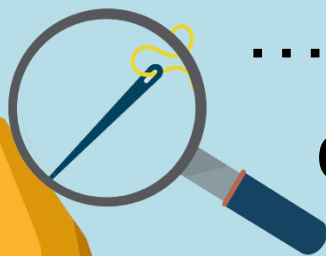EMERSE allows you to find cohorts based on things mentioned in the notes

- diseases

- drugs

- symptoms

- anything*

*if it is mentioned

Find cohorts

It's perfect for finding rare things…

…like rare cancers such as cutaneous leiomyosarcoma

See this talk for more details:
https://vimeo.com/677482835
"Using EMERSE to Improve Research Involving Rare Cancers"

Enroll patients on trials

- **We were able to enroll patients!**

- **Note:**
  While CUMC rolled out the EMERSE workflow for FORTE, we were also simultaneously testing reaching out to EMERSE identified FORTE patients over MyChart to introduce the FORTE study to patients
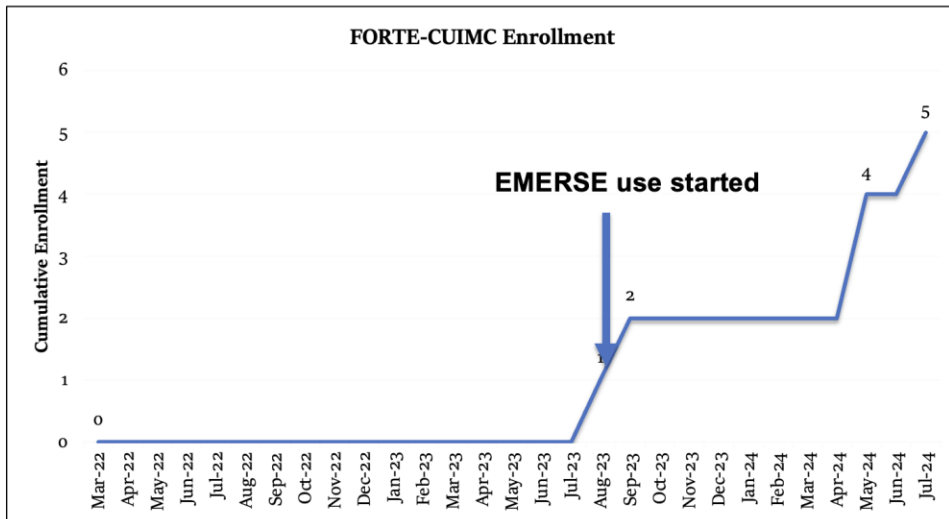
See this talk for more details:
https://vimeo.com/1084845123/
EMERSE Community Meeting May 2025

Highlight documents
for chart review

Thoracocentesis confirmed the recurrence of mantle cell lymphoma. Disease restaging work-up revealed multicompartment lymphadenopathy in the neck, mediastinal, retrocrural, retroperitoneal and pelvic regions. Bone marrow was also involved. The patient was treated with a total of six cycles of rituximab, cyclophosphamide, vincristine, doxorubicin and dexamethasone (R-HyperCVAD) completed in January 2007. That treatment led to complete remission that lasted until October 2008, when the disease was found to have recurred in the left pleural space and retroperitoneum without bone marrow involvement.

emerse
ELECTRONIC MEDICAL RECORD   SEARCH ENGINE

Our philosophy

It's important to view the terms/concepts in the context of the original text to truly understand the clinical meaning.

Meta Mapping (1000):
 1000 *^patient (Patients) [Patient or Disabled Group]
Meta Mapping (702):
 742 Presented (Presentation) [Idea or Concept]
 742 PALPABLE (Palpable) [Qualitative Concept]
 784 right Breast mass (Lump in right breast) [Finding]
Meta Mapping (1000):
 1000 Clinical Laboratory (Clinical Laboratory Services) [Health Care Activity]
Meta Mapping (1000):
 1000 Clinical Laboratory (Laboratories, Clinical) [Health Care Related Organization,Manufactured Object]
Meta Mapping (947):
 947 Mammography finding [Finding]
Meta Mapping (743):
 715 CARCINOMA OF BREAST (Breast Carcinoma) [Neoplastic Process]
Meta Mapping (1000):
 1000 *^patient (Patients) [Patient or Disabled Group]
Meta Mapping (1000):
 1000 LUMPECTOMY (Lumpectomy of breast) [Therapeutic or Preventive Procedure]
Meta Mapping (1000):
 1000 SENTINEL LYMPH NODE BIOPSY (Sentinel Lymph Node Biopsy) [Diagnostic Procedure]
Meta Mapping (745):
 806 Frozen Section (Frozen Sections) [Tissue]
 593 Tumour (Neoplasms) [Neoplastic Process]
Meta Mapping (745):
 806 Frozen Section (Frozen Sections) [Tissue]
 593 Tumor (Tumor tissue sample) [Tissue]
Meta Mapping (745):
 806 Frozen Section (Frozen Sections) [Tissue]
 593 Tumor (Specimen Source Codes - tumor) [Intellectual Product]
Meta Mapping (745):
 806 Frozen Section (Frozen Sections) [Tissue]
 593 Tumor (Tumor Mass) [Finding]
Meta Mapping (745):
 806 Frozen Section (Cryoultramicrotomy) [Laboratory Procedure]
 593 Tumour (Neoplasms) [Neoplastic Process]
Meta Mapping (745):
 806 Frozen Section (Cryoultramicrotomy) [Laboratory Procedure]
 593 Tumor (Tumor tissue sample) [Tissue]
Meta Mapping (745):
 806 Frozen Section (Cryoultramicrotomy) [Laboratory Procedure]
 593 Tumor (Specimen Source Codes - tumor) [Intellectual Product]

Meta Mapping (745):
 806 Frozen Section (Cryoultramicrotomy) [Laboratory Procedure]
 593 Tumor (Tumor Mass) [Finding]
Meta Mapping (1000):
 1000 Sentinel node (Sentinel node (disorder)) [Disease or Syndrome]
Meta Mapping (1000):
 1000 Sentinel Node (Sentinel Lymph Node) [Body Part, Organ, or Organ Component]
Meta Mapping (1000):
 1000 Revealed [Qualitative Concept]
Meta Mapping (1000):
 1000 Granulomatous Inflammation (Granulomatous inflammation) [Pathologic Function]
Meta Mapping (1000):
 1000 Gross examination (Sample macroscopy) [Laboratory Procedure]
Meta Mapping (1000):
 1000 confirmed (Confirmed by) [Qualitative Concept]
Meta Mapping (1000):
 1000 Confirmed (Confirmation) [Finding]
Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis) [Diagnostic Procedure]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 MASTITIS (Mastitis) [Disease or Syndrome]
Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis) [Diagnostic Procedure]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammatory disorder of breast) [Disease or Syndrome]
Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis) [Diagnostic Procedure]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammation of non-human mammary gland) [Disease or Syndrome]
Meta Mapping (696):
 760 Diagnosis (Disease) [Disease or Syndrome]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 MASTITIS (Mastitis) [Disease or Syndrome]
Meta Mapping (696):
 760 Diagnosis (Disease) [Disease or Syndrome]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammatory disorder of breast) [Disease or Syndrome]
Meta Mapping (696):
 760 Diagnosis (Disease) [Disease or Syndrome]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammation of non-human mammary gland) [Disease or Syndrome]

Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis Code) [Intellectual Product]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 MASTITIS (Mastitis) [Disease or Syndrome]
Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis Code) [Intellectual Product]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammatory disorder of breast) [Disease or Syndrome]
Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis Code) [Intellectual Product]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammation of non-human mammary gland) [Disease or Syndrome]
Meta Mapping (696):
 760 diagnosis (diagnosis aspect) [Functional Concept]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 MASTITIS (Mastitis) [Disease or Syndrome]
Meta Mapping (696):
 760 diagnosis (diagnosis aspect) [Functional Concept]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammatory disorder of breast) [Disease or Syndrome]
Meta Mapping (696):
 760 diagnosis (diagnosis aspect) [Functional Concept]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammation of non-human mammary gland) [Disease or Syndrome]
Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis Study) [Research Activity]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 MASTITIS (Mastitis) [Disease or Syndrome]
Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis Study) [Research Activity]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammatory disorder of breast) [Disease or Syndrome]
Meta Mapping (696):
 760 DIAGNOSIS (Diagnosis Study) [Research Activity]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammation of non-human mammary gland) [Disease or Syndrome]
Meta Mapping (696):
 760 Diagnosis (Date of diagnosis) [Temporal Concept]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 MASTITIS (Mastitis) [Disease or Syndrome]
Meta Mapping (696):
 760 Diagnosis (Date of diagnosis) [Temporal Concept]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammatory disorder of breast) [Disease or Syndrome]

Syndrome]
Meta Mapping (696):
 760 Diagnosis (Date of diagnosis) [Temporal Concept]
 593 Tuberculous (Tuberculosis) [Disease or Syndrome]
 593 Mastitis (Inflammation of non-human mammary gland) [Disease or Syndrome]
Meta Mapping (1000):
 1000 *^patient (Patients) [Patient or Disabled Group]
Meta Mapping (1000):
 1000 RECEIVED (Receive) [Qualitative Concept]
Meta Mapping (719):
 753 TUBERCULOSIS (Tuberculosis) [Disease or Syndrome]
 753 therapy (therapeutic aspects) [Functional Concept]
 623 Six months [Temporal Concept]
Meta Mapping (719):
 753 TUBERCULOSIS (Tuberculosis) [Disease or Syndrome]
 753 Therapy (Therapeutic procedure) [Therapeutic or Preventive Procedure]
 623 Six months [Temporal Concept]
Meta Mapping (719):
 753 TUBERCULOSIS (Tuberculosis) [Disease or Syndrome]
 753 Therapy (Therapy Object (animal model)) [Finding]
 623 Six months [Temporal Concept]
Meta Mapping (1000):
 1000 side effects (aspects of adverse effects) [Functional Concept]
Meta Mapping (1000):
 1000 N side effects (Adverse event) [Pathologic Function]
Meta Mapping (1000):
 1000 N Side effects (Adverse effects) [Pathologic Function]
Meta Mapping (888):
 694 Further [Spatial Concept]
 861 N Complications (Complication) [Pathologic Function]
Meta Mapping (888):
 694 Further [Spatial Concept]
 861 complications (Complication Aspects) [Functional Concept]

 emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

Location_Start	Location_End	Semantic	Entity
28	63	problem	a palpable mass of the right breast
51	63	bodyloc	right breast
130	136	bodyloc	breast
130	146	problem	breast carcinoma
170	180	treatment	lumpectomy
185	211	test	sentinel lymph node biopsy
213	240	treatment	Frozen section of the tumor
249	262	bodyloc	sentinel node
273	299	problem	granulomatous inflammation
308	325	test	gross examination
353	373	problem	tuberculous mastitis
396	421	treatment	anti-tuberculosis therapy
426	436	temporal	six months
442	444	negation	no
445	457	problem	side effects
461	486	problem	any further complications

CASE PRESENTATION

The patient presented with a palpable mass of the right breast with clinical, laboratory and mammographic findings indicative of breast carcinoma. The patient underwent lumpectomy and sentinel lymph node biopsy. Frozen section of the tumor and the sentinel node revealed "granulomatous inflammation", while gross examination confirmed the diagnosis of tuberculous mastitis. The patient received anti-tuberculosis therapy for six months with no side effects or any further complications.
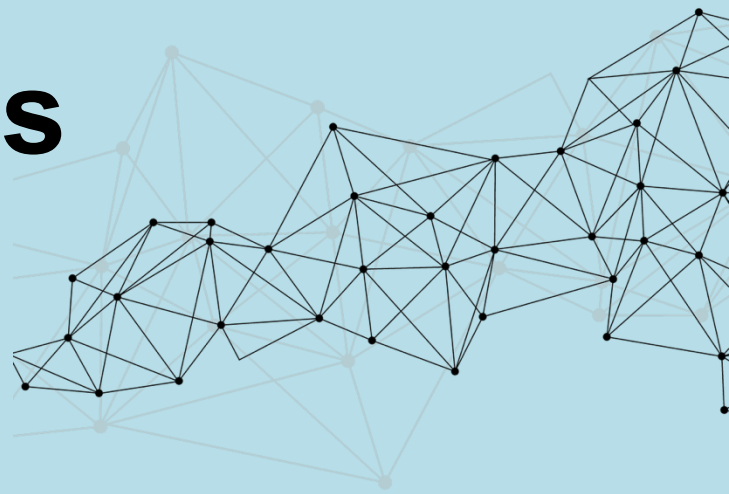
EMERSE is ⟹*fast*

Query to identify all patients with the following	Reporting DB time (s)	EMERSE time (s)	EMERSE advantage
cavernous hemangioma	14,652	2	7,320x
gray platelet syndrome	14,940	2	7,470x
inferior lingular segment of the left upper lobe	17,784	9	1,980x

…enabling real-time querying

Synonyms

- Used for query expansions
- User-controlled
- Multiple datasets can be included
- EMERSE Synonyms
  - acronyms, abbreviations, professional/consumer terms, misspellings, trade/generic drug names, species, chemo regimens, phrase variations, malapropisms, idioms, neologisms, organizations, companies, & more
  - 3 million unique entries

Synonyms for "smoker"

Click individual terms to highlight or de-highlight.

Synonyms (126)   SORT A TO Z ≡   HIGHLIGHT ALL   **HIGHLIGHT NONE**

abuse nicotine abuses nicotine addicted to nicotine addicted to tobacco addiction to nicotine

addiction to tobacco ash tray ash trays ashtray ashtrays chain smoker chain smokers chain smoking

chainsmoker chainsmokers chainsmoking cigar cigar day cigar per day cigarette day cigarette pack day

cigarette packs each day cigarette packs every day cigarette packs per day cigarette per day

cigarette smoke cigarette smoker cigarette smokers cigarette smoking cigarettes cigarettes day

cigarettes per day cigars cigars days cigars per day continue to smoke continued to smoke

continues to smoke continuing to smoke current every day smoker current everyday smoker current smoker

currently smokes dependence on cigarettes dependence on nicotine dependence on tobacco

dependent on cigarettes dependent on nicotine dependent on tobacco nicotine nicotine abuse

nicotine abuser nicotine abusers nicotine addiction nicotine addictions nicotine dependence

nicotine dependent pack a day pack day pack each day pack history pack per day pack year

pack year smoker pack years pack yr pack yrs packs packs a day packs day packs each day

packs of cigarette day packs of cigarette each day packs of cigarette per day packs of cigarettes day

packs of cigarettes each day packs of cigarettes every day packs of cigarettes per day packs per day

packs per year packyear packyears packyr packyrs pipe pipes pk year pk years pk yr pk yrs

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

Synonyms for "social determinant of health"   ✕

Click individual terms to highlight or de-highlight.

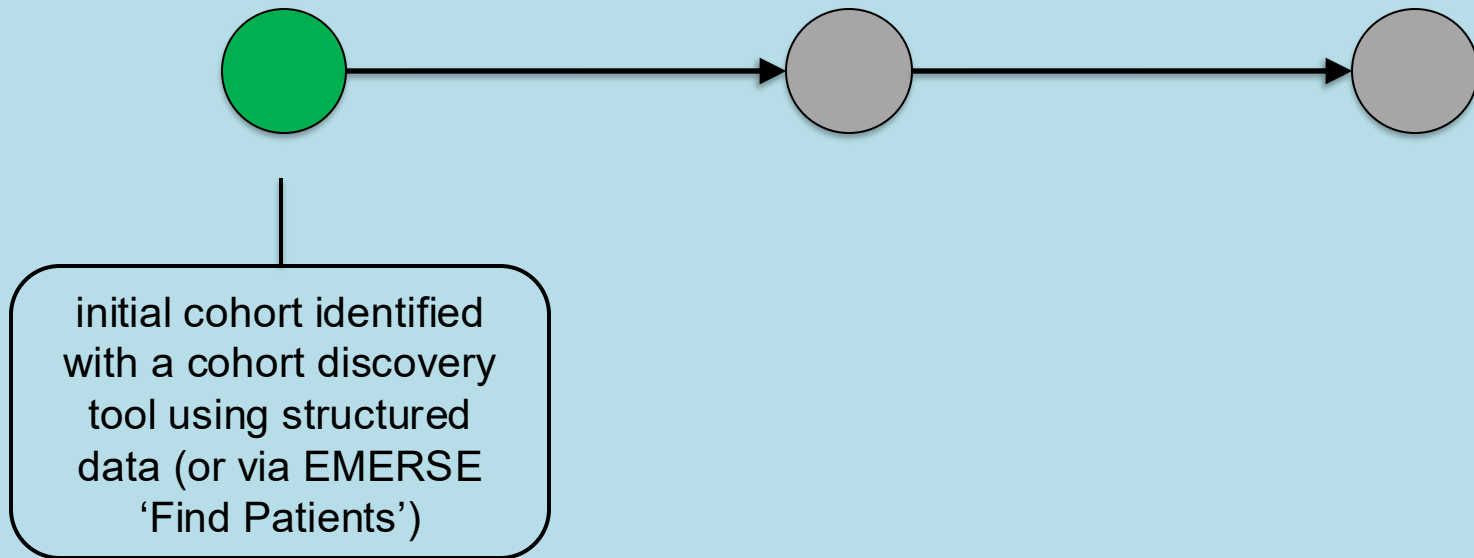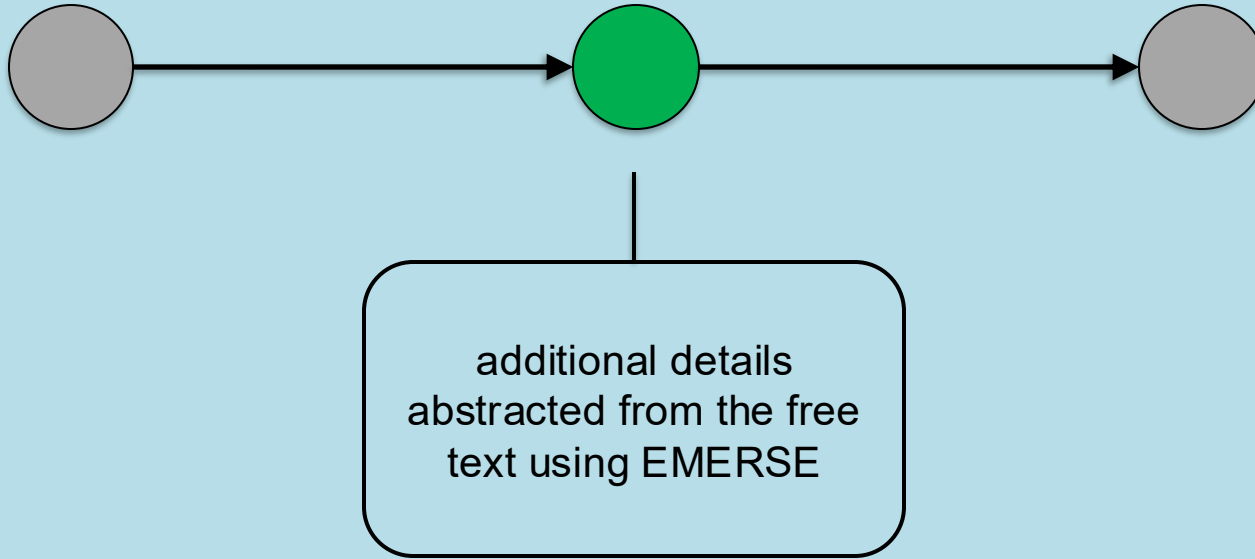**Synonyms (364)**   SORT A TO Z ≡   HIGHLIGHT ALL   **HIGHLIGHT NONE**

abused | academic problem | adoptive parent | afford | affordable medication | affordable prescription

alcoholism | amphetamine abuse | antisocial | appears depressed | appears sad | appears to be depressed

appears to be sad | assisted living facilit | assisted living facility | assistive device | bankruptcy | banned

bath bench | bath not on first floor | bed and bath are not on the first floor | boarding house | burglary

care giver support | cared for by a friend | cared for by a relative | charity care | cheaper | childhood abuse

chipped paint | cigarettes | cluttered | co pay | cocaine abuse | common law partnership | community support

computer illiterate | condominium | copay | correctional facility | court date | court mandated | court ordered

crime victim | death of family member | death of friend | death of partner | dementia | detox

difficult working conditions | difficulty maintaining employment | disability income | disabled partner

disabled partners | disabled relative | disorderly | divorced | drug seeking | drunk | dysthymia | dysthymic

economic problems | education level | education problem | education problems | elderly parents

employment difficulties | employment income insecurity | employment insecurity | estranged | evict | evicted

experiencing discrimination | exposed to violence | extended care facility | extended tub bench | feeling down

feels down | felony | filthy living | finances | financial concerns | financial constraints | financial difficulties

financial issues | financial strain | financial stress | financial stressors | financially | fixed income | flooded

food bank | food insecure | food insecurity | food pantries | food stamps | foreclosed | forgery | forging

foster care | foster family | fruit and vegetable intake | gender identity | general financial insecurity

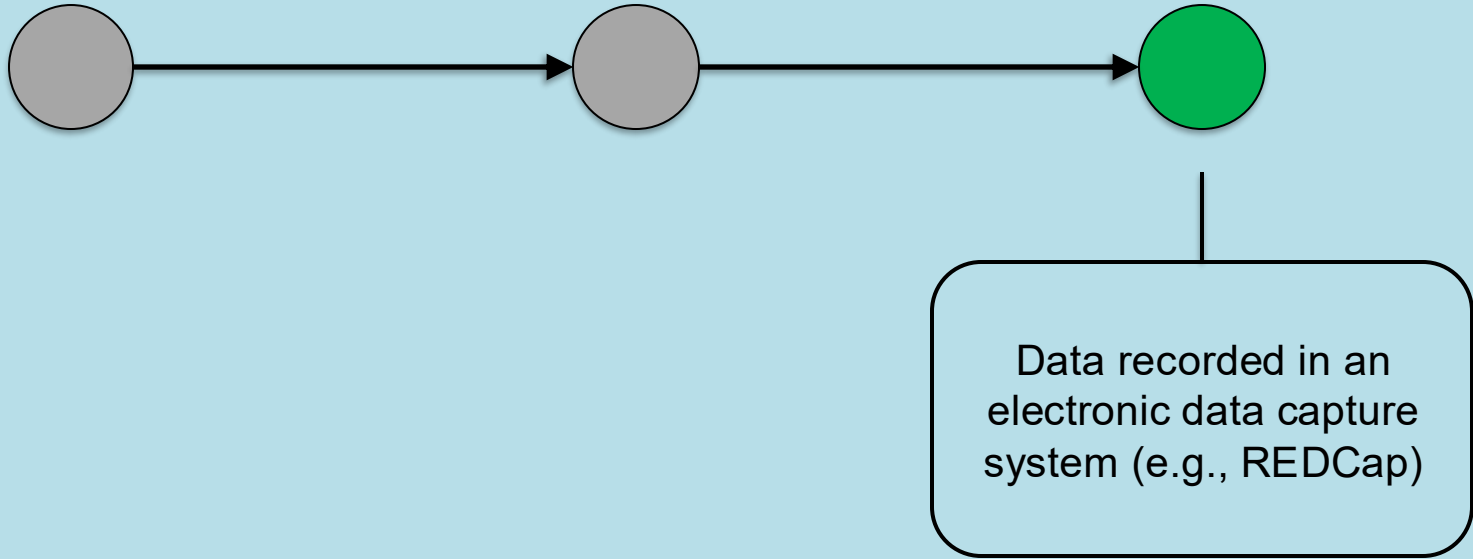grab bars present | group home | gunpoint | hand held shower | health insurance | healthcare problem

emerse
ELECTRONIC MEDICAL RECORD   SEARCH ENGINE

Typical workflow

initial cohort identified with a cohort discovery tool using structured data (or via EMERSE 'Find Patients')

Cohort discovery tools:
i2b2/ENACT, Leaf, etc.

Typical workflow

additional details abstracted from the free text using EMERSE

Typical workflow

Data recorded in an electronic data capture system (e.g., REDCap)

Publications using EMERSE

737

papers, abstracts,

theses, etc.

Full list at:

http://project-emerse.org/publications.html

Recently released…

NEW RELEASE

Incorporation of NLP features

- negation
- uncertainty
- subject (patient vs other)
- history of
- named entity recognition/mapping to ontologies

If you don't like our built-in NLP, use your own

emerse
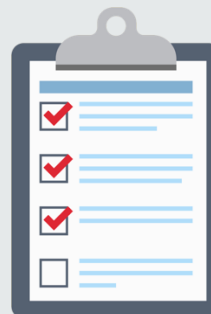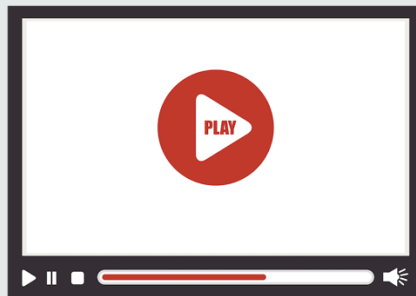ELECTRONIC MEDICAL RECORD   SEARCH ENGINE

Future Work
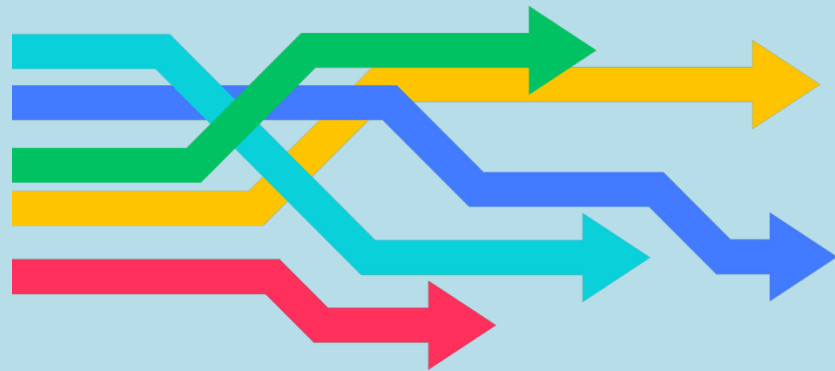
Data extraction from templated notes

Integration with ChatGPT or similar tools

& more…

Interested in EMERSE?

Contact us to schedule a time with your team for:

- Discussions about usage strategies

- Training

- Live demonstrations (abstractors, IT teams, etc)

Community Meetings
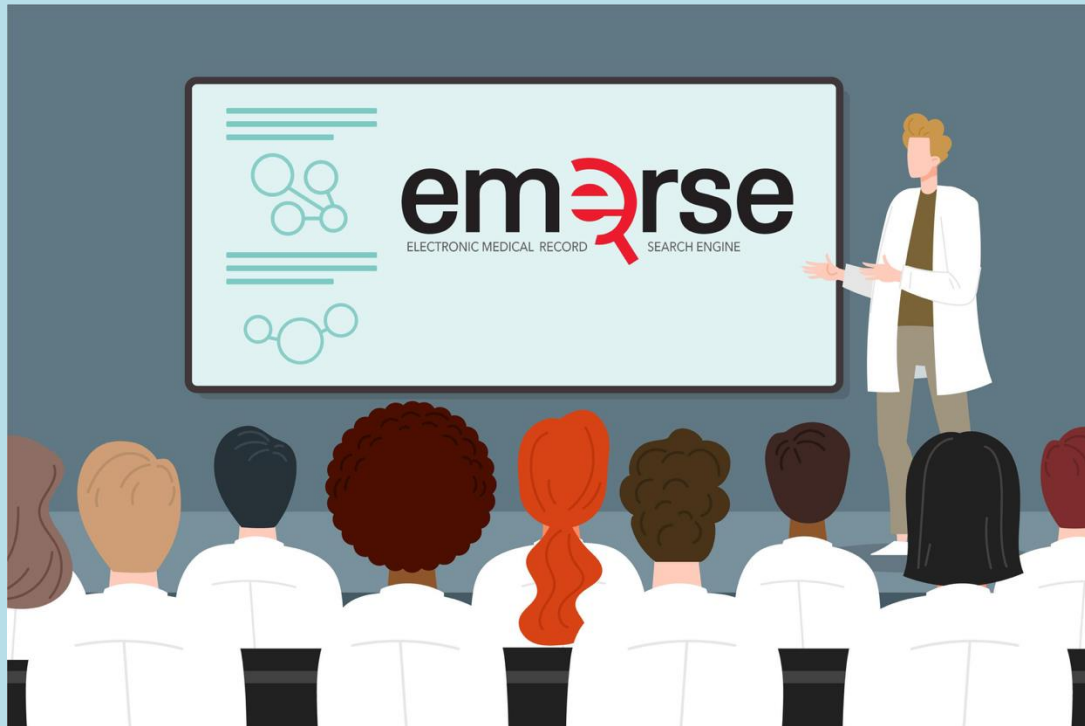

Twice per year

Last meeting November 2025

Next meeting May 2026

EMERSE-team@umich.edu

Lisa Ferguson

David Hanauer

Kellen McClain

Guan Wang

Live Demo

- No real names
- No PHI
- Publicly available
- Abstracts & case reports