

Query Log Analysis of an Electronic Health Record Search Engine

Lei Yang¹, Qiaozhu Mei^{1,2}, Kai Zheng^{1,3}, David A. Hanauer^{4,5}

¹School of Information

²Department of EECS

³School of Public Health Department of Health Management and Policy

⁴Comprehensive Cancer Center

⁵Department of Pediatrics

University of Michigan

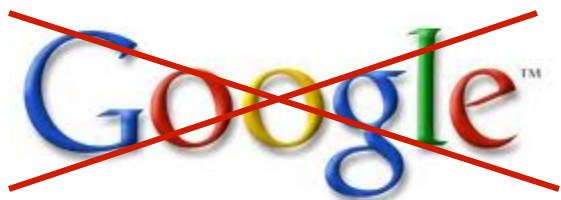
Motivation of the Study

- There is a significant need to conduct **full-text search in medical records** (e.g., by clinicians, researchers, administrators)
- **Very few electronic health records (EHR)** have full-text search functionality (Natarajan et al., IJMI 2010)
- An improved understanding of **end-user search behavior** will help design such functionalities.

Information Retrieval in EHR

- Retrieving Electronic Health Records

- Specialized language
- Rich, implicit intra/inter document structures
- Deep NLP/Text Mining is necessary
- Complicated information needs



What are We Going to Do?

- By studying query log of a EHR search engine
 - What are users of electronic health record search engine **looking for**?
 - How **different** is this search task from Web search?
 - How to **apply the learning** to build effective EHR search engines?

EMERSE

- EMERSE - **Electronic Medical Record Search Engine**
- Full-text search engine
- Widely used in UMHS since 2005 (and VA)
- Boolean keyword queries
- Semi-automatic query suggestion
- Collaborative search (Zheng et al., JAMIA 2011)
 - Users save queries as “search bundles”
 - Create short names for bundles
 - Share bundles with other users



http://141.214.189.34:8081/emerse/selectAbstractingPackages.jsp

Search Terms

Username: hanauer Search Terms: "Dupuytren's contracture" ophthalmologic

Search Terms: [History]

Please choose a Bundled Search Terms package below or enter search terms, separated by spaces, above.

Advanced Options

Bundle Name ↓	Description	Owner	Modified	View	Edit	Duplicate
Asthma Exclusions	Terms that may exclude pts. from Asthma registry. A pt. with a confirmed exclusion diagnosis will be excluded from the Asthma Registry.	mitchel	03/30/2006			
Asthma Search Terms	Terms with potential to a new positive diagnosis of asthma. Reviewer decision is required to confirm/reject asthma diagnosis.	mitchel	03/22/2006			
breast chemo	for qopi study section 7	miela	03/23/2006			
Cancer Staging Terms	This list contains a variety of terminology related to cancer staging.	hanauer	03/17/2006			
colorectal ca	colon & rectum ca for qopi	miela	03/28/2006			
Essential HTN search terms	For finding essential HTN	est	03/30/2006			
GVHD Terms	This includes a list of terms useful in searching for Graft Versus Host	hanauer	09/27/2006			

Example: the Search Interface

Log Out

Help

Preferences

Search Terms

Enter CPIs

Patient List

Full Search

4 of 5

Demographics

Problem Summ.

Documents

Pathology

Rad/Nuc

Other Results

55 of 116

Dictated Medical Information (DMI): Single Note

Username: hanauer

Bundled Search Terms: DRR - depression [\[View all\]](#)

Case Date	Doc Type	Clinician	Service	Department	Admit	Discharge	Signed
	LET/NOTE-RV		DFP	FP			Y

Subjective: The patient comes in to talk about her **depression**. She has been having ~~counseling~~ for many many years and finally agreed to try some **antidepressants**. She did very well with Celexa but got a reaction to it and had to stop it, although I am not sure exactly what she had. Then she was switched to sertraline 25 mg and has done wonderfully on this. Her symptoms, besides feeling **depressed**, was also she was almost obsessive compulsive in many ways. She said all of those things are much better. Her husband and her children have also noted that she is doing much better

Objective: Mental status with appropriate thought processes, content, behavior. She is not suicidal, no hallucinations, and judgment is good. Her appearance is normal.

Assessment: **Depression**/obsessive compulsive controlled with sertraline.

Plan: Discussed. Refilled. Of note is that the patient also talks about menopausal-type symptoms such as hot flashes and change in her periods. I suggested that she schedule to have this looked into. I did order an LH, FSH, and a fasting cholesterol if we can do this before she comes in, if not, we will do it when she comes in.

Example: a Retrieved Document

EMERSE: Electronic Medical Record Search Engine

https://emerse.med.umich.edu/viewPackageDetails.jsp?packageName=pk... Google

Bundled Search Terms Details

Username: hanauer Bundled Search Terms: **DRR - depression** [View all](#)

Bundle Details

Name: DRR - depression

Owner: hanauer

Created: 06/18/2010

Last Modified: 06/18/2010

Description: For finding patients with depression who have diabetes for the diabetes research registry (DRR)

Phrases to highlight [Highlight]	Phrases to ignore
"anti-depressant"	"-ST depression"
"anti depressant"	"-ST segment depression"
"anti-depression"	
"anti depression"	
"mood disorder"	
depression	
depressed	

Example: a Search Bundle

EMERSE: Electronic Medical Record Search Engine

http://141.214.189.30:8081/emerse/editAbstractingPackage.jsp?package Google

Edit Search Terms Bundle

Username: hanauer Bundled Search Terms: **Breast cancer terms** [View all](#)

This bundle can be viewed / used by:

All users
 A subset of users

Share a Bundle Publically/Privatey

Full user list

- andrewcw
- angelass
- arthrama
- asedman
- bradbury
- brehmer
- cathymah
- cgadegbe
- cjmartin
- ckitko

Users for this bundle

- hanauer
- sungchoi

Add to list →

← Remove from list

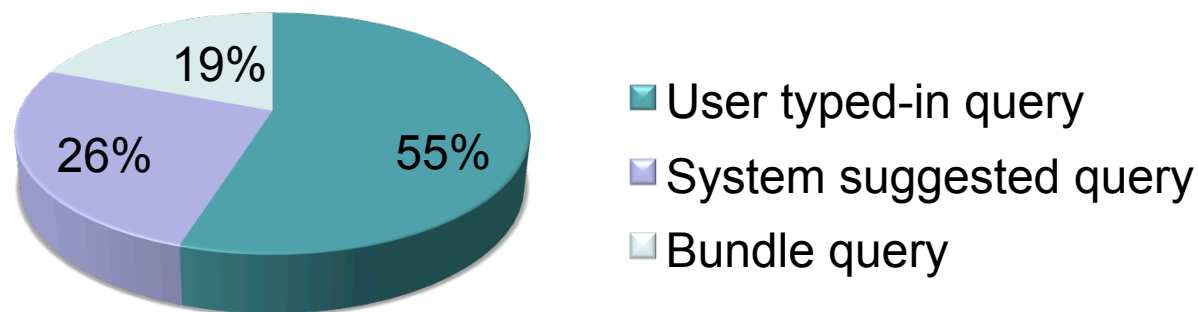
To select more than one user, control-click (Windows) or command-click (Macintosh)

Delete Cancel Edits Save Changes

Descriptive Statistics of the Log Data

- 202,905 queries collected over 4 years (2006 - 2010)
- 533 users (medical professionals in UMHS)
- Each query log recorded query string, user, and time stamp.
- We employed effective heuristics to classify queries as three types.

Types of Queries



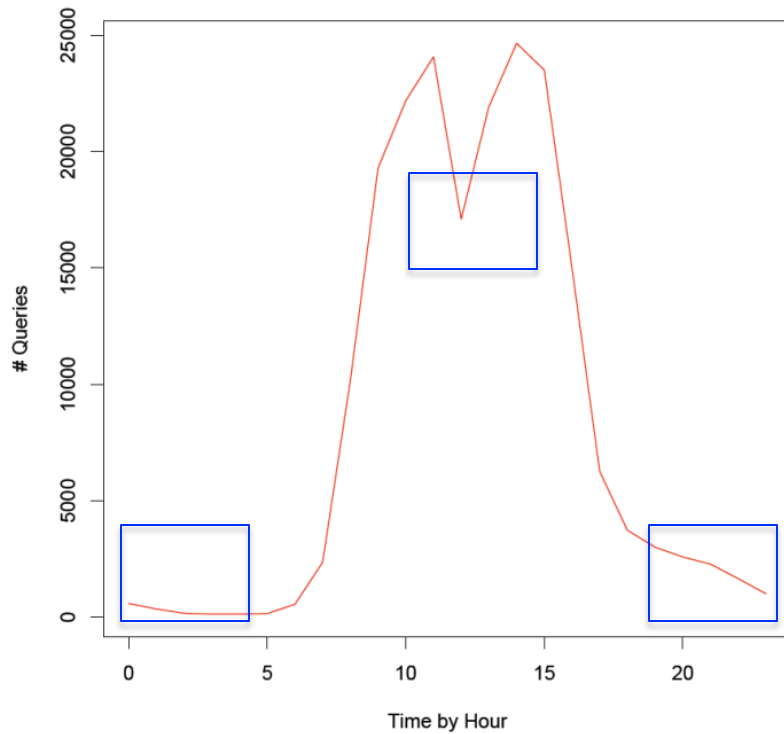
Examples of Query Log

- Non-bundle query could be either a user typed-in query or a system suggested query.

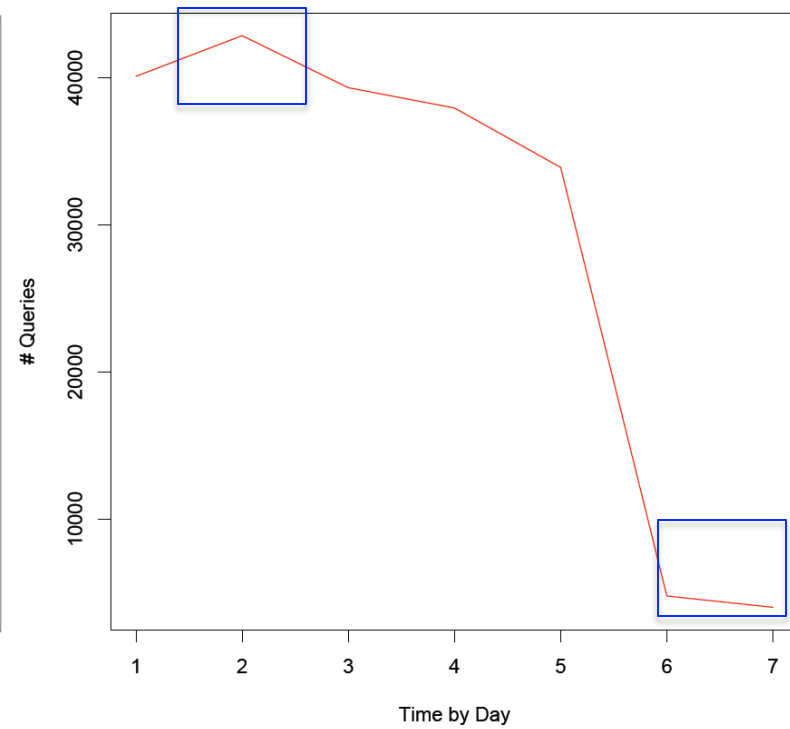
User	Query	Type	Time	IP Address
A	Chemotherapy Terms	Bundle	10/06/2006 14:34:29	10.20.1.31
B	infection	Non-bundle	09/19/2006 12:02:00	10.20.97.60

Temporal Patterns

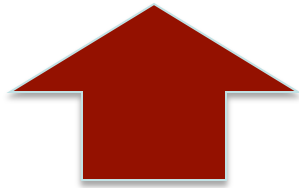
People in University of Michigan are working really hard !!!



Hours of a day



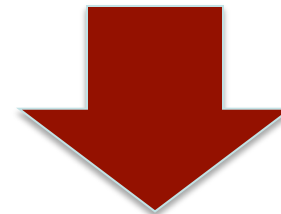
Days of a week (Mon - Sun)



Ok, that's the query log

Question One:

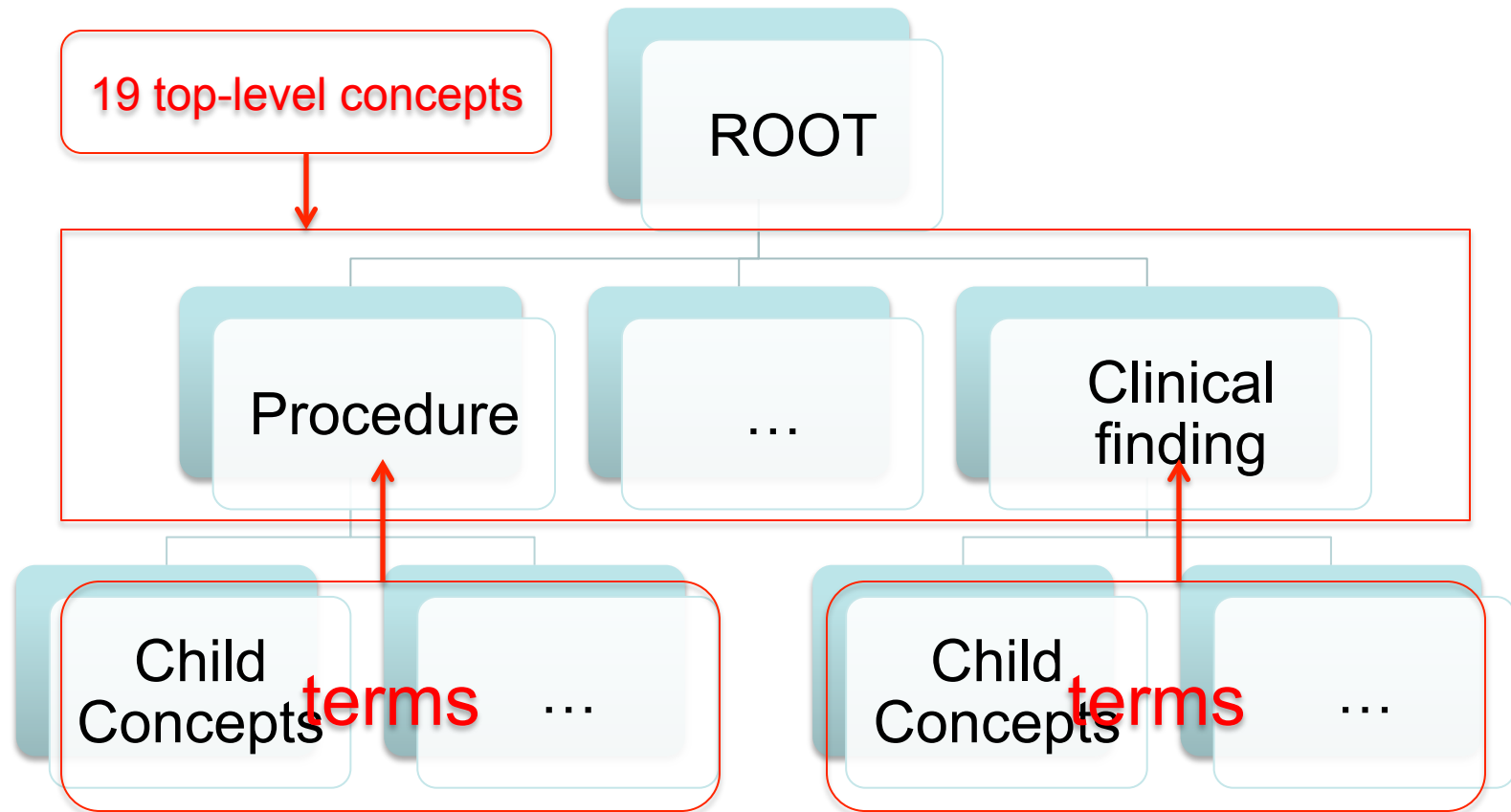
What are users looking for?



To Describe Users' Information Needs

- Extract **a category list** from SNOMED Clinical Terms, the comprehensive clinical terminology.
- Categorize **each query** into one of the categories in the list.
- Use the **query category distribution** to describe users' information needs.

SNOMED CT Ontology Structure



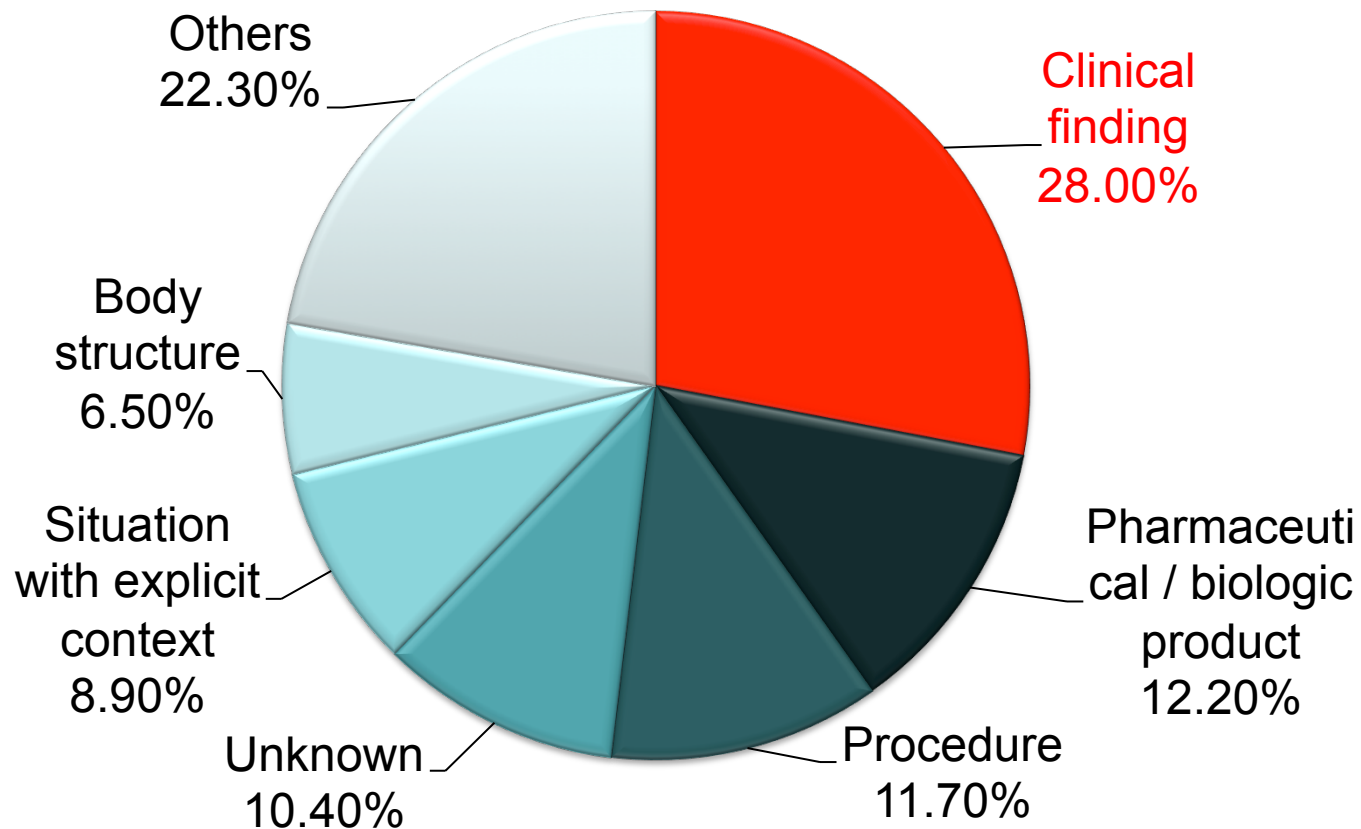
An Example

- Concept ***procedure*** = {t1, t2, t3, t4, t5}
- Concept ***clinical_finding*** = {t1, t6, t7, t8, t9}
- Query ***q*** = {t1, t6, t7, t8}

- $|\mathit{procedure} \cap \mathit{q}| = 1$
- $|\mathit{clinical_finding} \cap \mathit{q}| = 4$

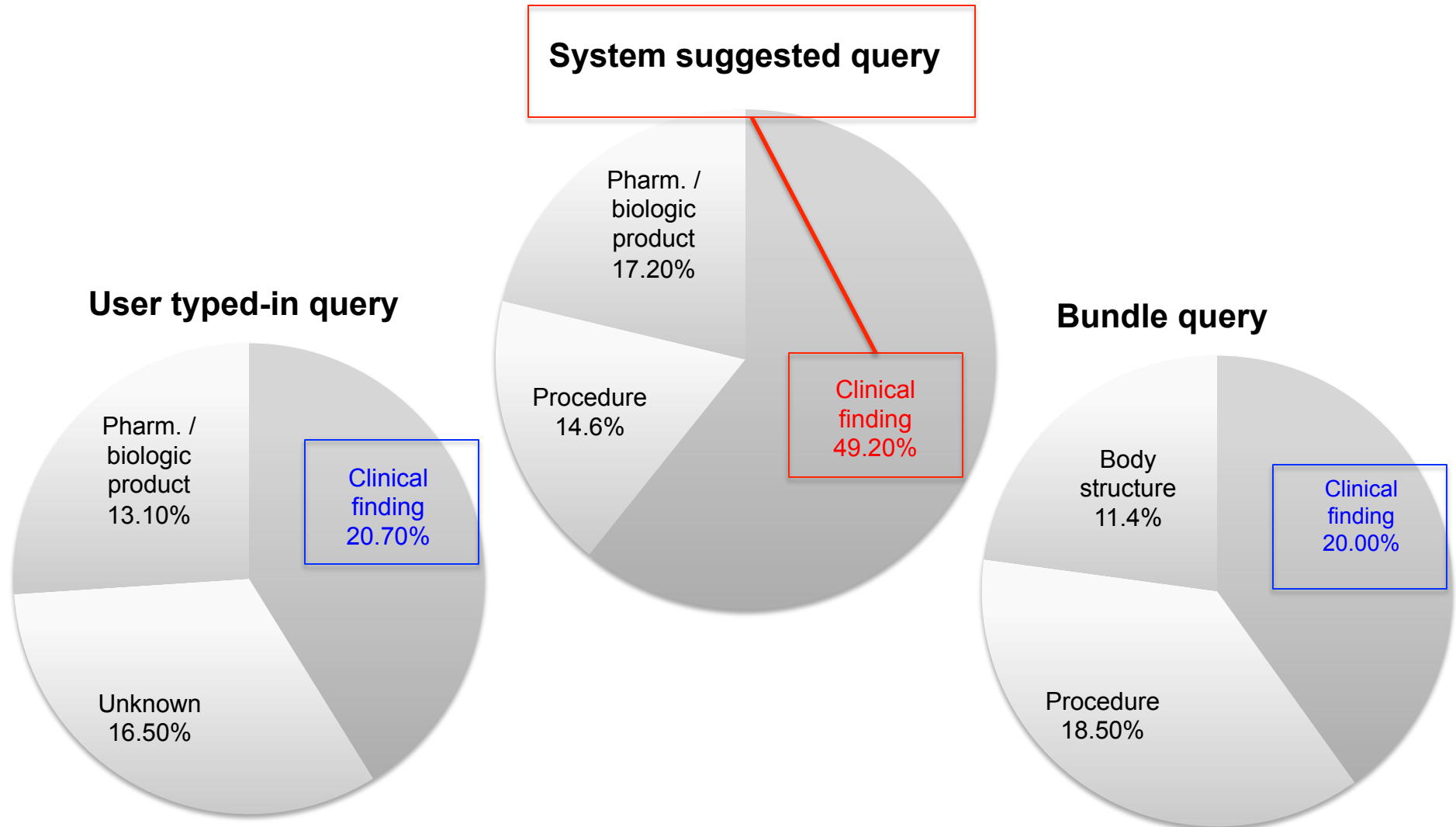
Categorize query ***q*** as ***clinical finding***.

Distribution of Query Category



An example of UNKNOWN query: **vicoden** **vicodin**

Top 3 Categories of Each Type of Query

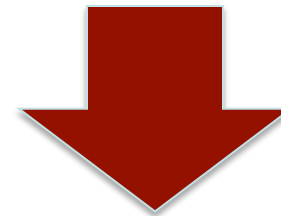




Diversified queries reflect diversified information need.

Question Two:

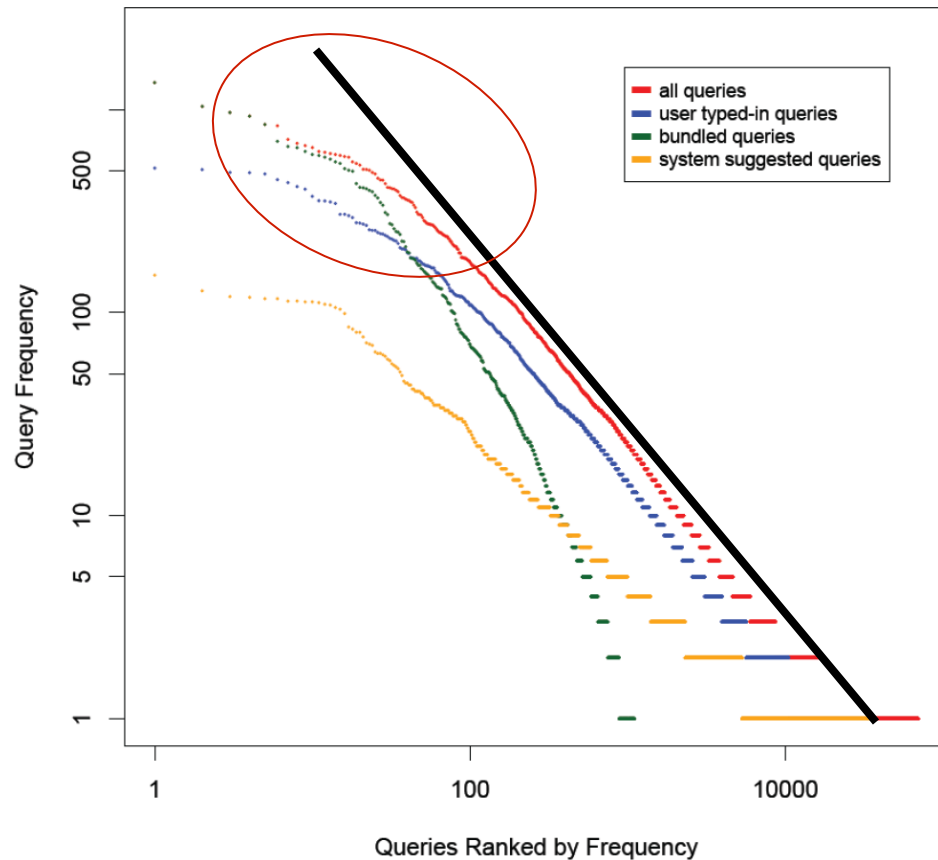
How different is EHR search from Web search?



Questions

- How are queries distributed?
 - Are **navigational queries** also popular in EHR search?
 - What are the **most frequently used** queries?
- How hard is the search task?
 - Are EHR queries more **complex to construct**?
 - How many **query terms** are outside the scope of formal biomedical vocabularies?
 - What is the average length of **search sessions**?
- ...

Query Distribution



Long tail but no fat head

Conclusion:

Fewer **navigational queries** compared to Web search (e.g., “facebook”, “amazon”)

Top Queries

Rank	Query	Type	Rank	Query	Type
1	Cancer Staging Terms	Bundle	11	PEP risk factors	Bundle
2	Chemotherapy Terms	Bundle	12	Cigarette/alcohol use	Bundle
3	Radiation Therapy Terms	Bundle	13	comorbidities	Bundle
4	Hormone Therapy Terms	Bundle	14	Radiation	Typed-in
5	CTC Toxicities	Bundle	15	lupus	Bundle
6	GVHD Terms	Bundle	16	Deep Tissue Injuries	Bundle
7	Performance Status	Bundle	17	Life Threatening Infections	Bundle
8	Growth Factors	Bundle	18	CHF	Bundle
9	Other Toxicity Modifying Regimen	Bundle	19	relapse	Typed-in
10	Advanced lupus search	Bundle	20	anemia	Typed-in

Query Length

The average length of user typed-in queries (1.7 terms) is less than that of Web search queries (2.35 terms).

Conclusion: information need of EHR users is much more difficult to be represented than that of Web search.

All	User Typed-in	System Suggested	Bundle Name	Bundle Query
5.0 terms	1.7 terms	13.8 terms	2.6 terms	58.9 terms

Dictionary Coverage

Web search query term coverage: 85% - 90%

Unfortunately, **even the meta-dictionary** can only cover 68.0%. (e.g., adsr01->?, anemaia->anemia)

Conclusion: This low coverage indicates more challenging to handle in EHR search engine.

Dictionary	Size of Vocabulary (terms)	Query Term Coverage
SNOMED CT	30,455	41.6%
UMLS	797,941	63.6%
Meta-dictionary	934,400	68.0%

Session Analysis

- Definition
 - A search session: a basic mission of a user to achieve **a single information need**.

User	Time	Query
A	10:29:04	warifin
A	10:29:08	warfrin
A	10:29:11	warfarin
A	10:29:14	warfarin Coumadin

- Using a **30-minute timeout**, we managed to identify 35,928 search sessions.

Session Analysis

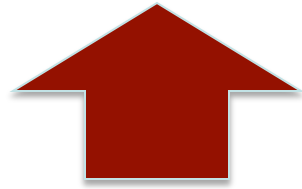
- Observations

The average length of a session is 14.8 minutes, which is **considerably longer** than a session of Web search.

A search task in EHR is generally more difficult than a task in Web search.

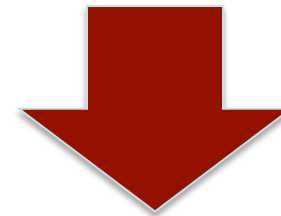
27.7% sessions end with a search bundle compared to the overall bundle usage of **19.2%**.

Many users seek for bundles after a few unsuccessful attempts to formulate queries.



Query log analysis

What have we learned?



Lessons Learned

- EHR search is **much more challenging** than Web search
 - More complex information need
 - Longer queries, more noise
- Users have **substantial difficulty** to formulate their queries
 - Longer search sessions
 - High adoption rate of system generated queries
- Collaborative Search might be the **solution**
 - The effectiveness of collaborative search encourage to further explore social-information-foraging-techniques.

Future Work

- Query suggestion using large scale, heterogeneous language networks
- Useful presentation of search results
- Support result analysis and decision making
- Enhance collaborative search in EHR search engines
- ...



References

- [1] Natarajan K, Stein D, Jain S, Elhadad N. An analysis of clinical queries in an electronic health record search utility. *Int J Med Inform.* 2010 Jul;79(7):515-22. Epub 2010 Apr 24.
- [2] Zheng K, Mei Q, Hanauer DA. Collaborative search in electronic health records. *Journal of the American Medical Informatics Association.* 2011;18(3):282–91.

Thanks!



- This work is supported by the National Library of Medicine,
HHSN276201000032C : Developing an Intelligent and Socially Oriented Search Query Recommendation
Service for Facilitating Information Retrieval in Electronic Health Records PI: Zheng