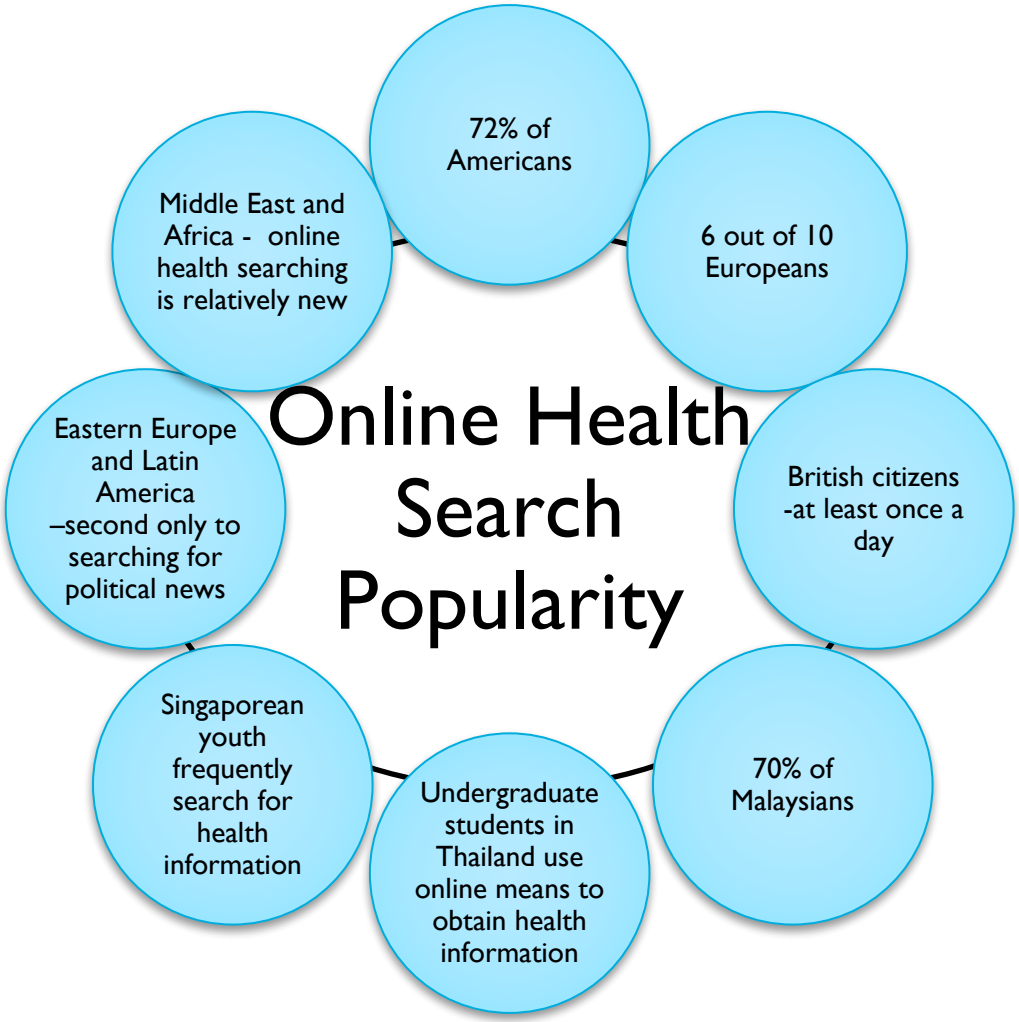
A Perspective on Human Computer Interaction and Information Retrieval – Health Information Searching

Anushia Inthiran
Department of Accounting and Information Systems
University of Canterbury

Agenda

- Online Health Search & Evolution
- Online Search Patterns & Challenges
- Research Questions & Experiment Methodology
- Results
- Design Suggestions
- Limitations
- Future Work

Online Health Search

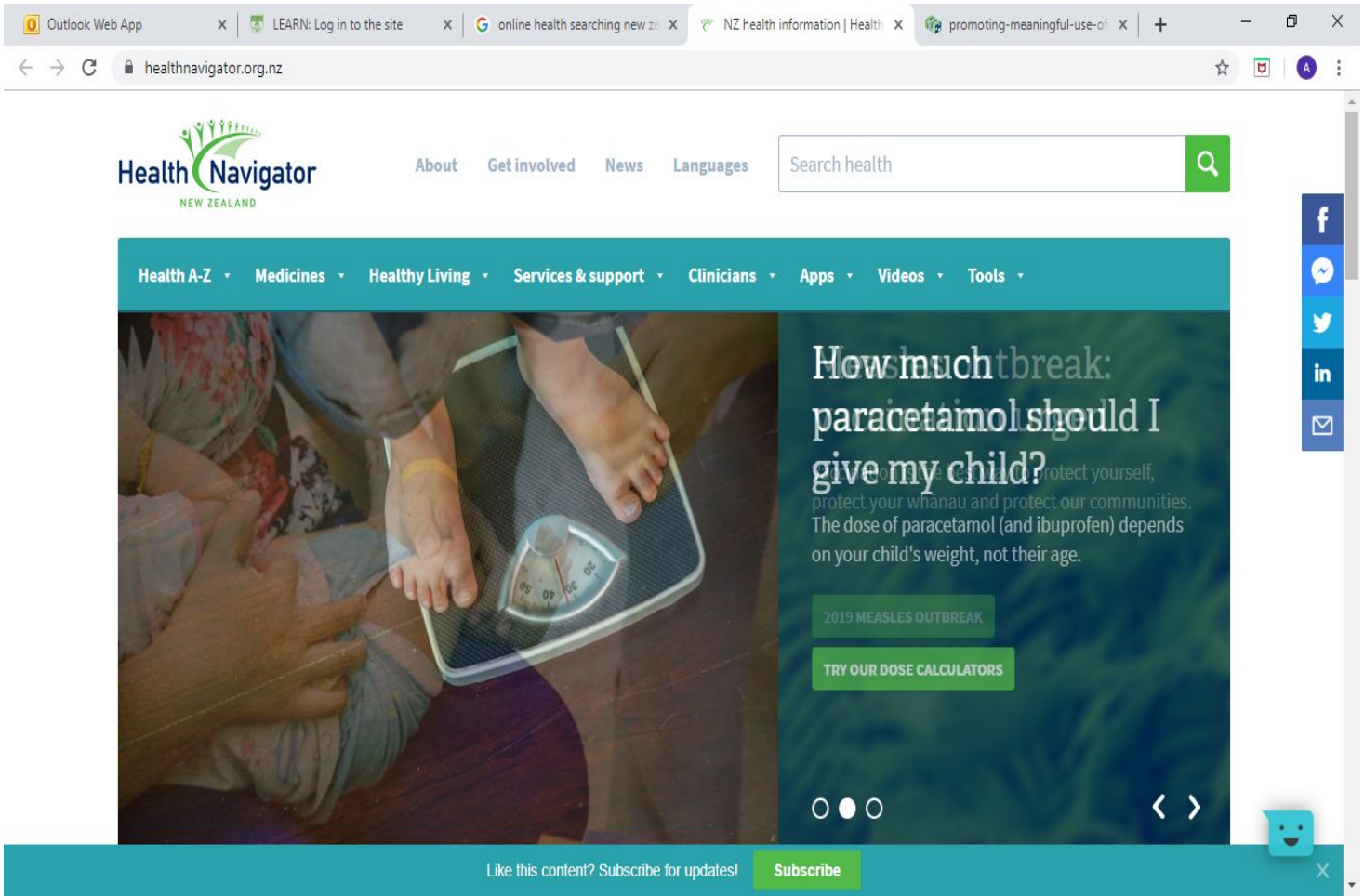


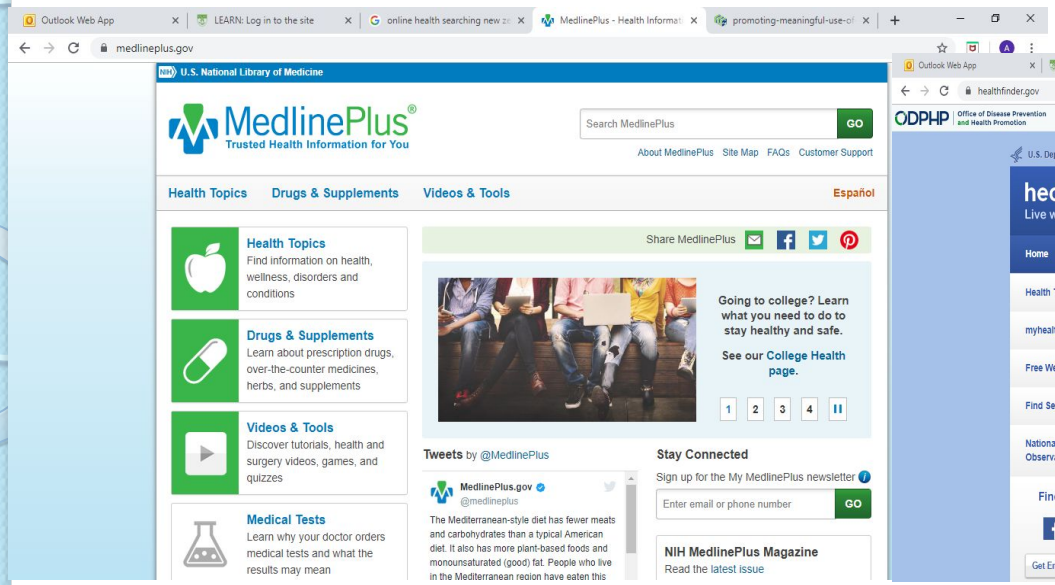
Online Health Search – New Zealand

- New Zealanders use Google to answer questions about cancer
- "breast cancer NZ"
- Third most preferred source of health information - Internet
- Two thirds of participants had used the Internet to access health information
- Little awareness or use of existing New Zealand-based online health resources

Evolution (1)

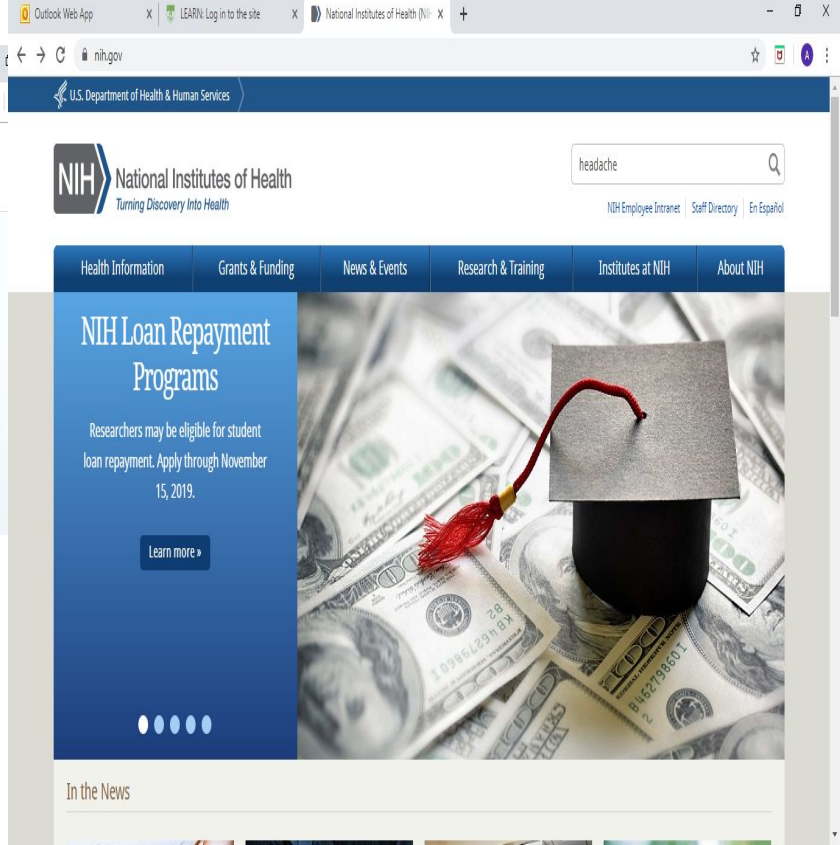
- Vertical search engines



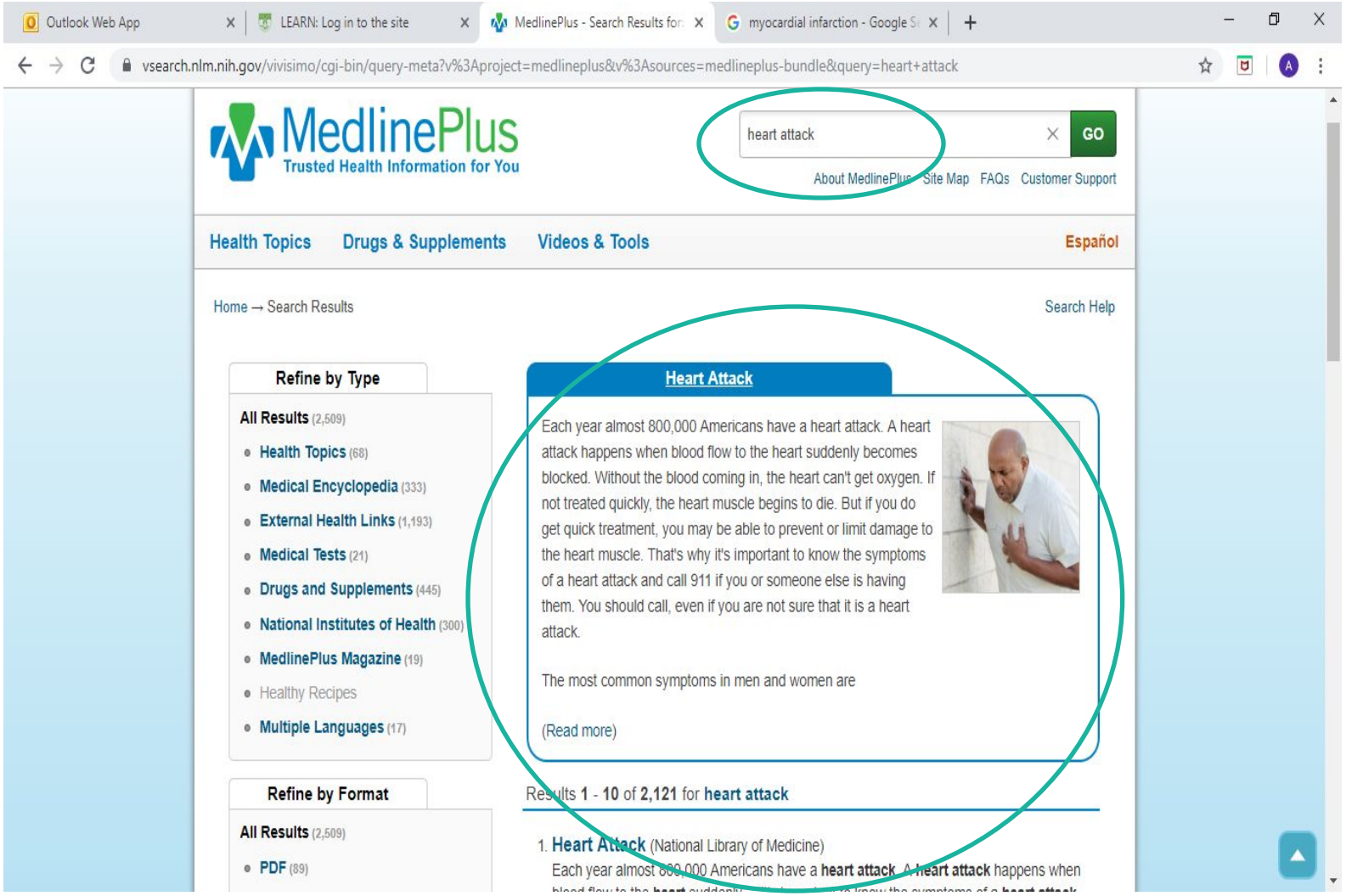

medlineplus.gov
U.S. National Library of Medicine
MedlinePlus®
Trusted Health Information for You
Search MedlinePlus
GO
About MedlinePlus Site Map FAQs Customer Support
Health Topics Drugs & Supplements Videos & Tools Español
Health Topics
Find information on health, wellness, disorders and conditions
Drugs & Supplements
Learn about prescription drugs, over-the-counter medicines, herbs, and supplements
Videos & Tools
Discover tutorials, health and surgery videos, games, and quizzes
Medical Tests
Learn why your doctor orders medical tests and what the results may mean
Share MedlinePlus
Going to college? Learn what you need to do to stay healthy and safe.
See our College Health page.
1 2 3 4
Tweets by @MedlinePlus
MedlinePlus.gov
@medlineplus
The Mediterranean-style diet has fewer meats and carbohydrates than a typical American diet. It also has more plant-based foods and monounsaturated (good) fat. People who live in the Mediterranean region have eaten this
Stay Connected
Sign up for the My MedlinePlus newsletter
Enter email or phone number
GO
NIH MedlinePlus Magazine
Read the latest issue

healthfinder.gov
ODPHP Office of Disease Prevention and Health Promotion
health.gov
healthfinder.gov
HealthyPeople.gov
U.S. Department of Health and Human Services
healthfinder.gov
Live well. Learn how.
En español
Search healthfinder.gov:
Go
Home
Health Topics A to Z
myhealthfinder
Free Web Content
Find Services Near You
National Health Observances
Find us on:
Get Email Updates
Protect Your Newborn Baby
If you're pregnant, talk with your doctor or midwife about newborn screening before your baby is born.
Read More
Health Topics
Get information to help you and your loved ones stay healthy.
Health Conditions and Diseases
Nutrition and Physical Activity
Pregnancy
myhealthfinder
See which preventive services you or a loved one may need this year.
Age:
Sex: Female

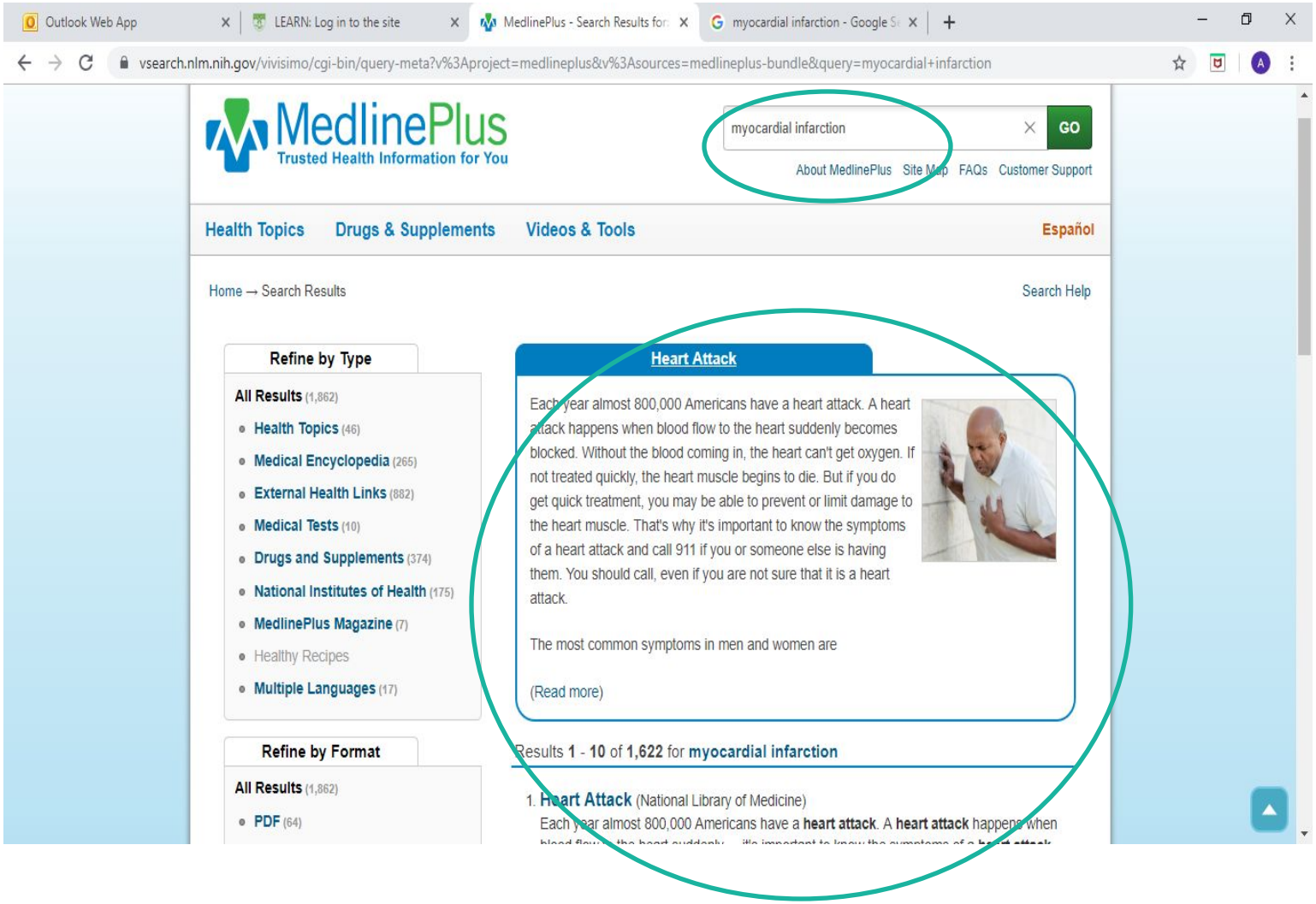
tanyadok.com
tanyadok.com
Arsip Konsultasi Artikel Kesehatan Cari Dokter Tanya Dokter Login
Apa masalah kesehatan sobat?
Tanya jawab dengan dokter ahli disini
Tanya Dokter
Profil Dokter TanyaDok
Psikolog Irma Gustiana: Pentingnya
Dokter Lady Dhita Alfara, M.Gizi, SpGK: Pendekatan
Dokter Bram Pradipta, SpOG: Dokter Bertalenta
Dokter Amanda: Ketahui Autisme Sejak Dini


nih.gov
U.S. Department of Health & Human Services
NIH National Institutes of Health
Turning Discovery Into Health
headache
NIH Employee Intranet Staff Directory En Español
Health Information Grants & Funding News & Events Research & Training Institutes at NIH About NIH
NIH Loan Repayment Programs
Researchers may be eligible for student loan repayment. Apply through November 15, 2019.
Learn more »
In the News

Vertical Search Engines

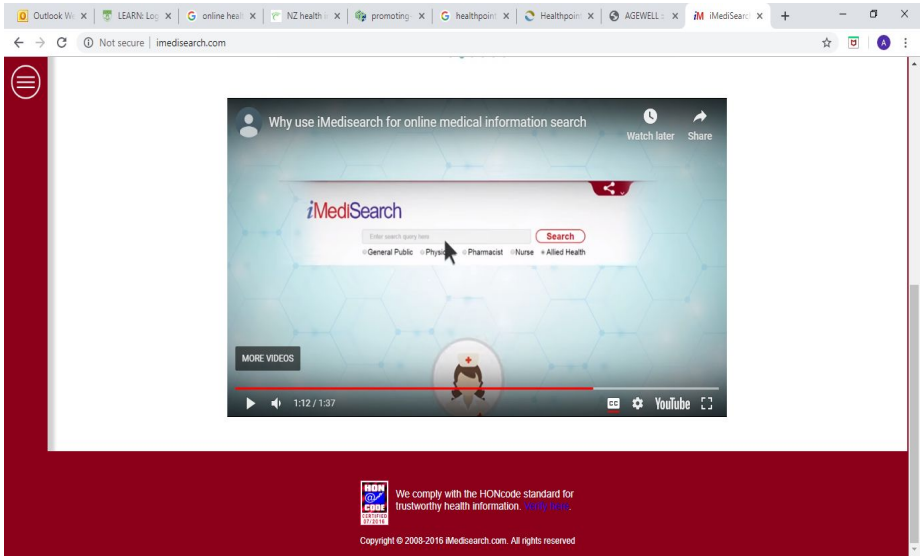


Vertical Search Engines



Evolution (2)

- Next generation sources
  - imedisearch
  - Pogofrog
  - HON Foundation



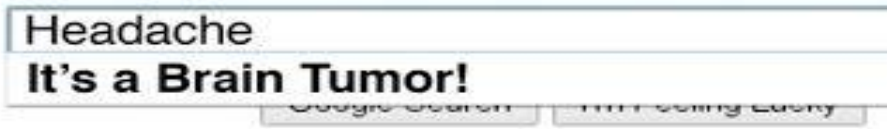
Evolution (3)

- Web 2.0 applications
  - health support groups (QuitNet)
  - consumer health portals (PatientsLikeMe, AfterTheInjury.org)
- track and share health indicators

Search Experience

How dangerous is Dr. Google?

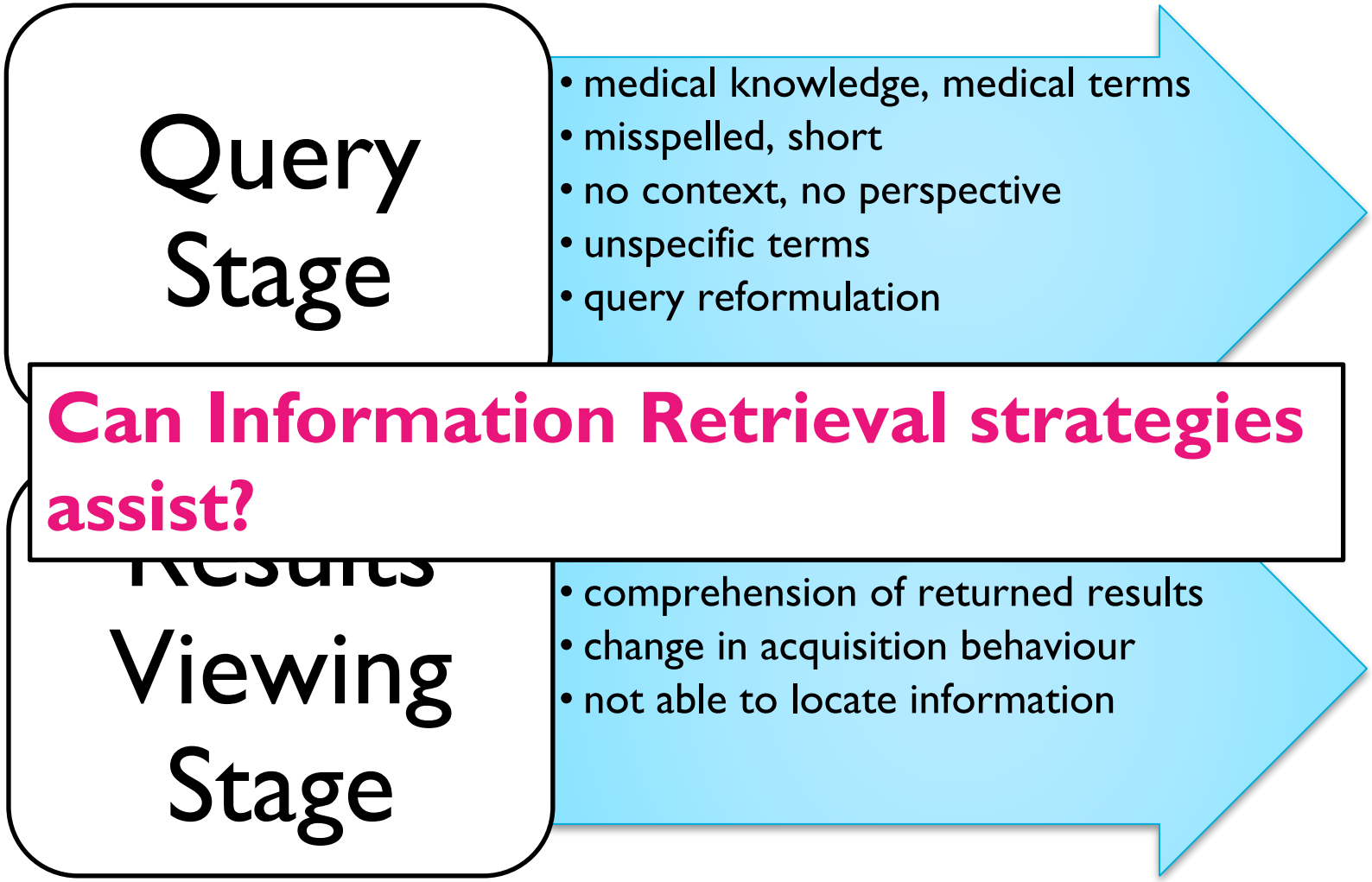
Not dangerous! It's easier to search diagnoses than symptoms. But google does not do too badly on symptoms either.



Online Search Patterns

- Average Query Length
  - 3.3 terms per query, simple and short (2002-2007)
  - 6.27 terms per query (2011)
  - natural language type question queries (2011-2012)
  - query reformulation (2014)
- Average Number of Queries
  - 2.34 queries (2002-2007)
  - 4-5 queries (2012)
- Task Completion Time
  - low, average 3 minutes (2002-2007)
  - high, average 15 minutes (2011-2013)
- High Search Failure (2002-2012)
- Sequence of Searching & Browsing (2012-2014)

Search Challenges



Research Question

- Describe the typology of patterns demonstrated when searching across tasks of varying levels of difficulty

Experiment Methodology

- Interactive Information Retrieval
  - User Centric
  - Information Retrieval System
- Instruments
  - Socio Demographic Questionnaire
  - Simulated Situations
  - Keylogging Software
  - Post Search Task Difficulty Interview
- Search Domain
  - MedlinePlus

Simulated Situations

- short cover story that describes a situation that leads to the use of an information retrieval system
- two parts: simulated work task scenario (cover story) and indicative request (search goal suggestion)
- substitute for real world information needs
- study and compare search behaviour using a standardised scenario
- invokes a common information need

Simulated Situations

Simulated work task scenario: Today morning after getting out of bed you noticed that you could not move your neck. You could not move it to the left or right. There is swelling on the left side of your neck. The swelling seems to be near lymphatic nodes. You are also experiencing some pain near the neck area. You want to find out what is wrong.

Indicative Request: Find for an instance, information to inform you of your condition, what can be done and why are you experiencing this pain at the neck.

Simulated Situations

Simulated work task scenario: Your colleague had just undergone a health test and found out that his kidney is enlarged and there is a stricture. A procedure called URS&RPG was performed. After this procedure, he experienced urine retention. He is in pain and has been told that surgery is required. You are concerned for your colleague and would like to use MedlinePlus to provide him with some information.

Indicative Request: Find for an instance, information to provide to your colleague to inform him of his condition, why is he experiencing this condition? Try to help him locate possible alternatives to treat his condition. Is surgery the only option?

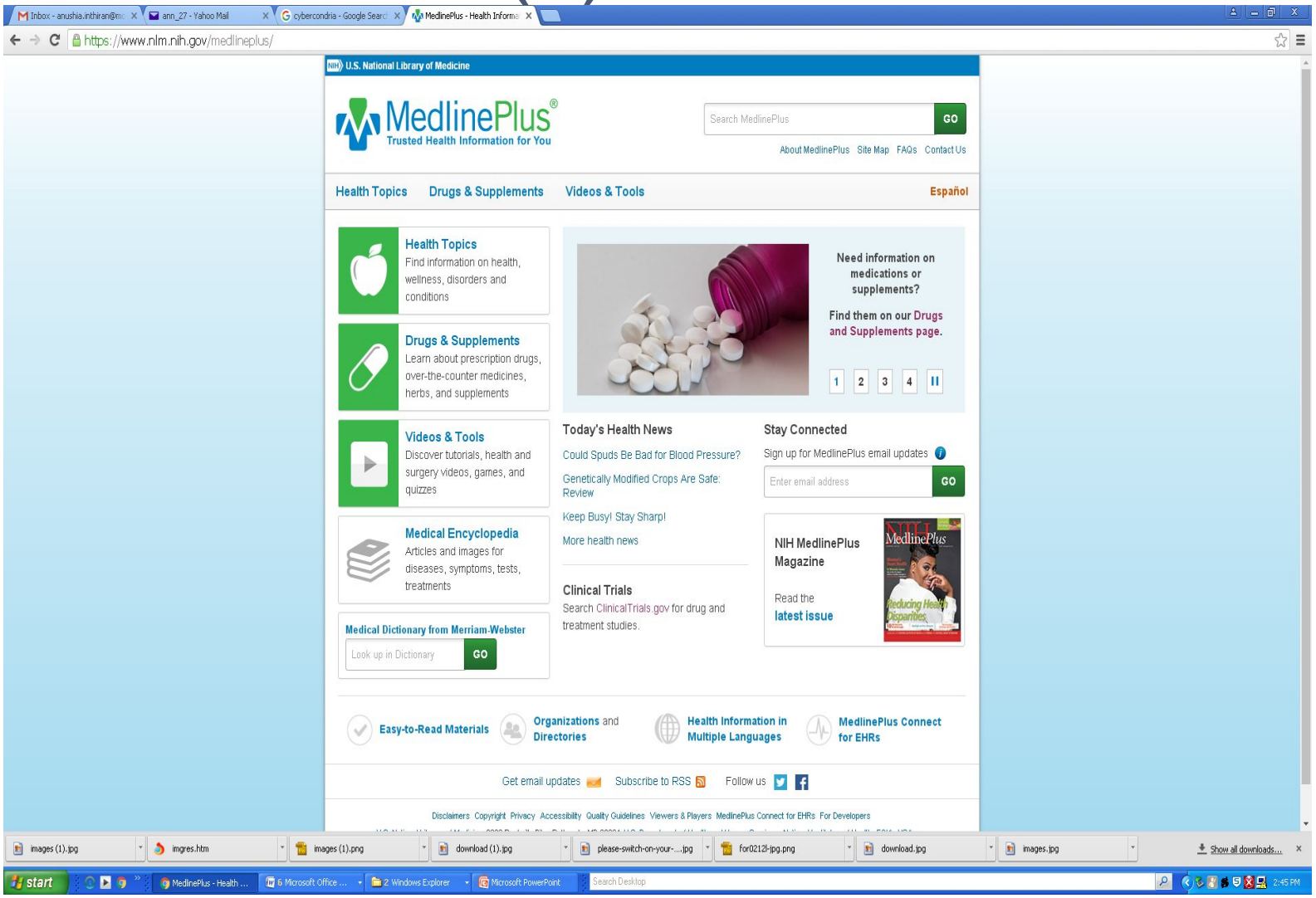
Simulated Situations

Simulated work task scenario: Yesterday at the mall you witnessed a child about the age of five having fits/seizure. The child's parents who were nearby were in a state of shock. On arriving home, you were curious about the situation at the mall and would like to find out more about the childs condition. Use MedlinePlus to help you find out what could have been done to assist the child.
Indicative Request: Find for an instance, information advising you of what first aid measures you could have taken to assist the child and what are some possible reasons for the child to be experiencing this condition.

Search Domain - MedlinePlus

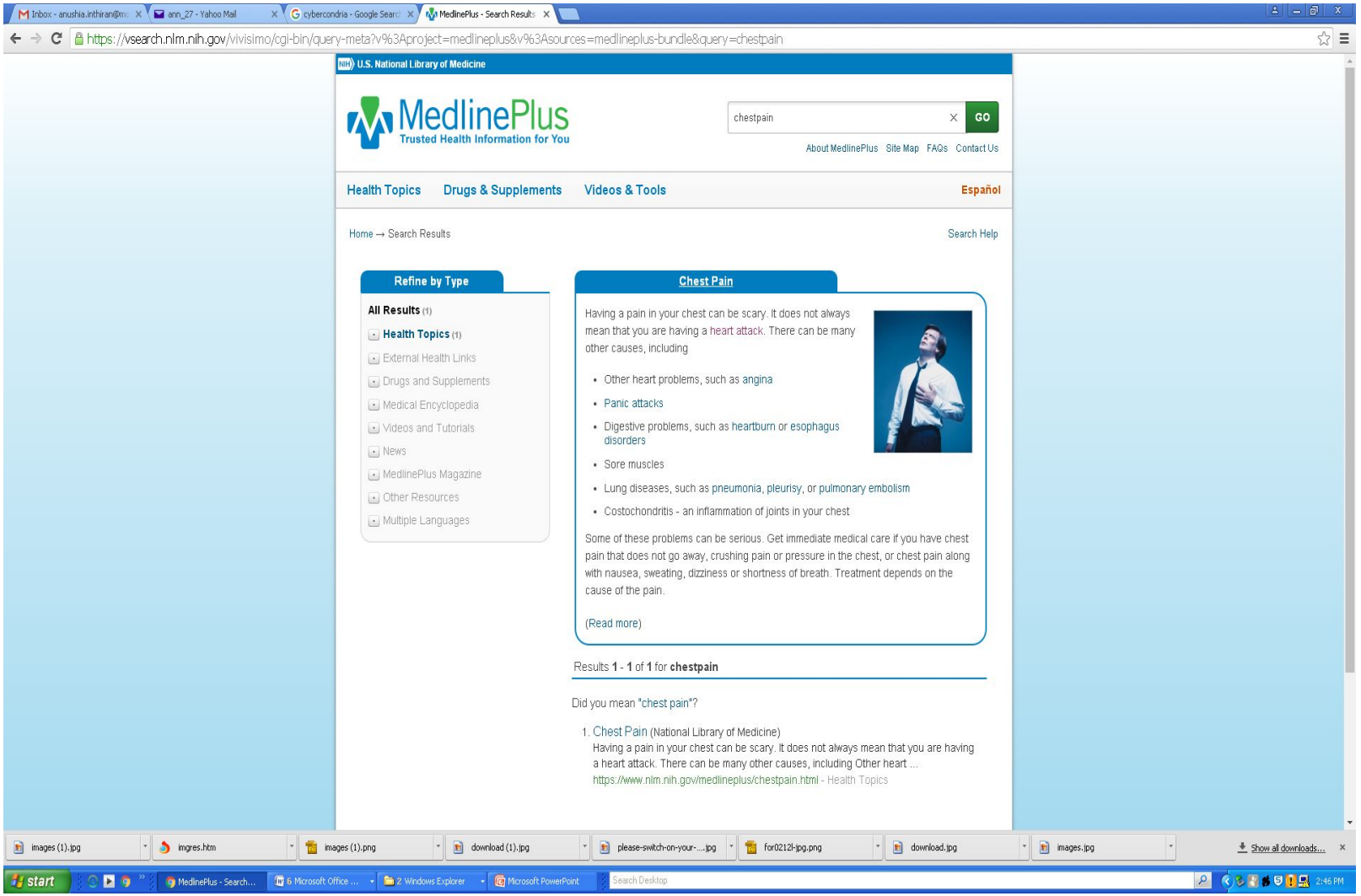
- free and publicly available clinical medical domain managed by the American National Library of Medicine and the National Institute of Health
- interface is similar to other publicly available medical search domains on the Internet
- retrieval strategy employed by MedlinePlus is basic text matching
- search assisting features must be explicitly invoked

MedlinePlus (1)



https://www.nlm.nih.gov/medlineplus

MedlinePlus (2)



Socio Demographic Details (1)

Category	Details			
Gender	Male	Female		
	38	22		
Use of the English Language (%)	1st Language	2nd Language	3rd Language	
	55.0	25.0	20.0	
Age (years)	Max	Min	Mean	SD
	60	18	33.7	9.6

Category	Education Type	% Value
Undergraduate Students (20.0% of total participants)	Business	10.0
	Information Technology	10.0
Post-Grad Students (25.0% of total participants)	Business	8.0
	Information Technology	9.0
	Arts and Social Sciences	5.0
	Engineering	3.0
Occupation Details		
Working Professionals (55.0% of total participants)	University Lecturers	18.0
	Administrators	10.0
	Engineers	10.0
	Managers	10.0
	Consultants	7.0

Socio Demographic Details (2)

Category	Details			
General Search Experience (years)	Max	Min	Mean	SD
	16.0	2.0	10.3	3.5
Medical Search Experience (years)	10.0	0	6.2	3.5
Medical Search Experience On General Search Engines Only (75.0% of total participants) [years]	8.0	2.0	9.9	3.5
Medical Search Experience On Medical Search Engines Only (10.0% of total participants) [years]	8.0	1.0	6.2	3.5
Medical Search Experience on Medical Search Engines and General Search Engines (15.0% of total participants) [years]	5.0	1.0	5.0	2.0

Test Statistics Methods

- Task Difficulty
  - Chi Square and Standardised Residuals
- Search Behaviour
  - Kruskall-Wallis and Dunn's Post Hoc Test

Perception of Task Difficulty

- Easy

  Simulated work task scenario: Your colleague had just undergone a health test and found out that his kidney is enlarged .....

- Difficult

  Simulated work task scenario: Today morning after getting out of bed you noticed that you could not move your

Criteria\Task		A1 (D) (N=60)	A2 (E) (N=60)	A3 (N=60)
# of queries issued per search session	Mean	2.8	3.7	2.6
	SD	1.5	1.9	1.8
	Median	3	4	2
	Min	1	1	1
	Max	8	8	10
	Test Statistics Mean # of Queries H=21.87, p<0.001	111.6	131.5	105.4
Query length per search session	Mean	8.5	8.5	6.5
	SD	5.9	6.5	5.3
	Median	7	7	5
	Min	1	2	1
	Max	26	32	21
	Test Statistics Mean Query Length H=15.94 p=0.0012	115.6	126.3	103.2
% of queries using query operator		15.0	22.0	10.0
% of queries re-issued		5.0	0	2.0
% medical queries		1.0	14.0	0
% queries without clicks		20.0	25.0	22.0
% natural language		0	0	2.0
% of ineffective queries	Spelling Error	2.0	2.0	0
	Query Too Specific	11.0	42.0	0
Querying Assistance				
% of usage Health Topics Tab		13.0	10.0	0
% of usage Videos & Cool Tools Tab		3.0	3.0	6.0
% of usage Drugs & Supplements Tab		16.0	3.0	6.0
% of usage Interactive Tutorial		3.0	3.0	6.0

Results Viewing Behaviour

Criteria\Task		A1 (D) (N=60)	A2 (E) (N=60)	A3 (N=60)
# of search results clicked per search session	Mean	2.0	1.8	1.8
	SD	1.4	1.7	0.7
	Median	2	2	1
	Min	1	1	1
	Max	8	10	8
# of sub-links clicked per search session	Mean	1.8	1.4	2.3
	SD	1.3	0.6	1
	Median	1	2	2
	Min	1	3	1
	Max	4	3	4
% of usage Control-F (find)		19.0	12.0	4.0
% of participants clicked beyond the 1st results		21.0	3.0	10.0
% unsuccessful search sessions		15.0	35.0	3.0
Refining Search Results Assistance				
% of usage Refine by Keyword		5.0	25.0	5.0
% of usage Refine by Topic		0	2.0	2.0
% of usage Refine by Type		7.0	8.0	0

Comparison and Task Completion Time

Criteria\Task		**A1 (D) (N=60)**	**A2 (E) (N=60)**	**A3 (N=60)**
Total Querying versus Clicking Behaviour		173 vs 164	195 vs 127	139 vs 131
Task Completion Time Per Search Session (mins)	Mean	5.8	7.6	4.7
	SD	3.6	5.3	3.9
	Median	5	6	3.5
	Min	1	2	1
	Max	15	30	12
	Test Statistics Mean Task Completion Time H=15.92 p=0.0012	117.7	128.8	90.7

Search Behaviour - Easy

Search Behaviour Name	**Unbalanced**		
Sub-Classification	Active Technical Query Issuers	Slow and Unproductive	Critical Results Viewers
Features	Most number of queries	Most unsuccessful search sessions	Most results refining assistance usage
	Longest queries	High task completion time	Most queries without clicks
	Most query operators	Most ineffective queries	
	Most medical queries	Uneven querying versus clicking behaviour	

Search Behaviour - Difficult

Search Behaviour Name	**Motivated**	
Sub-Classification	Problematic Query Issuers	Active Results Viewers
Features	Most query re-issues	Most viewing results beyond the 1st page
	Most usage of querying assistance	

Design Suggestions

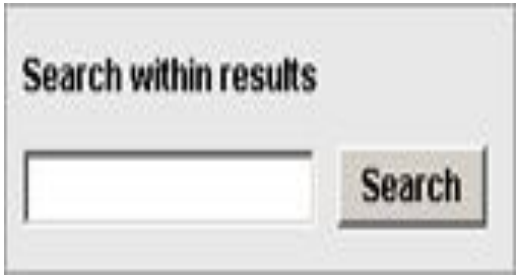
TFIDF

For a term i in document j:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

tf_{ij} = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

- Profiles
  - typology of search behavior
  - intelligent agent to identify search patterns
- Summary Snippet Suggestions
  - relevance score based on weighted calculation
  - relevance score based on user voting
- Query and Search Results Refining Suggestions
  - query suggestion
  - search within results - specific keywords or terms
  - clustering - reduce the number of pages from the same site
  - page content identification

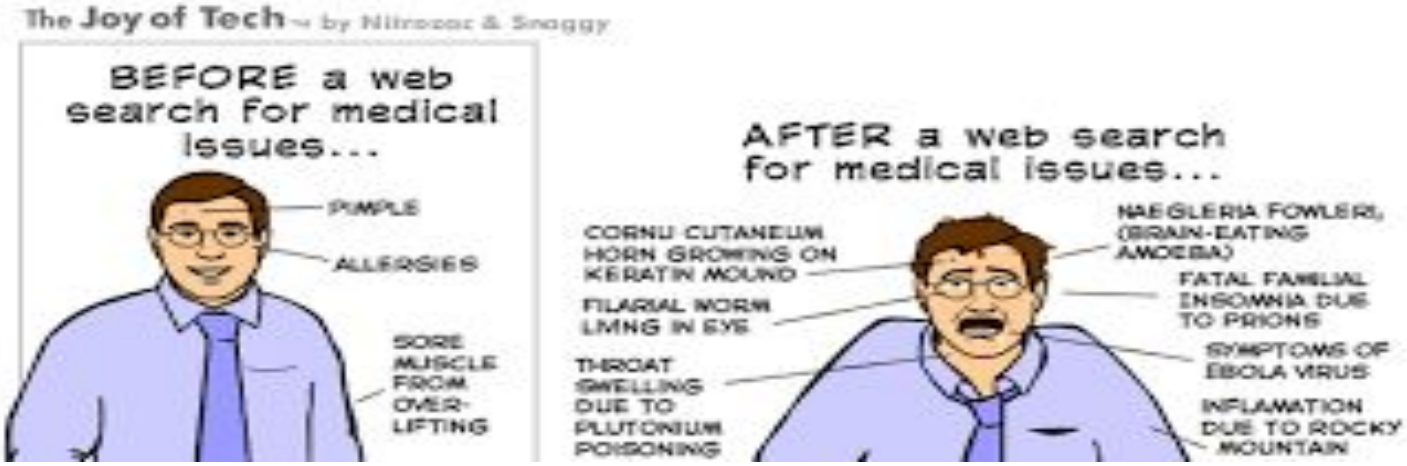


Limitations

- results cannot be generalised
- explicit reasons as to why search sessions were perceived as easy or difficult
- pre and post task difficulty changes
- implement the think aloud protocol
- controlled search session

Future Work

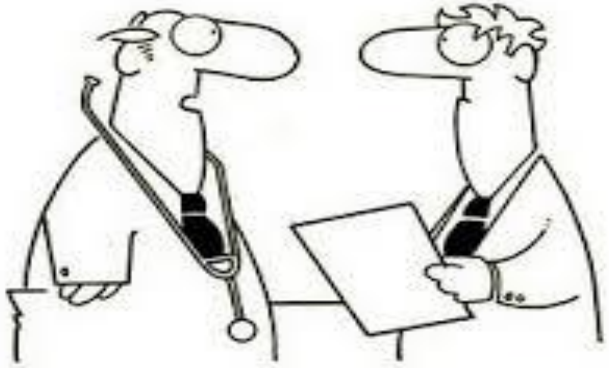
- analyse data from a qualitative aspect
- examine search patterns of successful and unsuccessful search sessions
- prototype design suggestions



Thank you for your attention

Any Questions or Comments?

anushia.inthiran@canterbury.ac.nz



"More and more patients are going to the Internet for medical advice. To keep my practice going, I changed my name to Dr. Google."

Web MD

(proper noun)

Something that makes a mild cold into a deadly disease that will kill you within the next 24 hours.