Time-Critical Search

Nina Mishra   Ryen W. White   Samuel Ieong   Eric Horvits
Microsoft Research

Google | electric fire oven

Web | Shopping | Images | Videos | Maps | More | Search tools

About 2,770,000 results (0.33 seconds)

Shop for electric fire oven on Google — Sponsored

Camp Chef Italia Artisan Portable Pizza Oven - PZ...
$319.99
ShoppersChoice.com

Ainslie Portable Indoor-Outdoor Fireplace
$168.63
Cymax Stores

Pacific Living Outdoor Pizza Oven - 430 Stainle...
$1,099.00
Woodland Direct

Fire in oven element - can I still use burners? - resolved | Ask ...
ask.metafilter.com/234149/**Fire**-in-**oven**-element-can-I-still-use-burners
Jan 28, 2013 - I was preheating my **electric oven** and when I opened the door, one of the elements popped, sparked and shot out an impressive flame .

Oven heating element just caught fire. - AR15.COM
www.ar15.com › General › General Discussion
Oct 22, 2010 - 44 posts - 26 authors
Once it's started, it doesn't need **electricity** to continue. It's like a magnesium **fire**, it gets so hot that nothing will put them out, they have to run ...

A (possible) emergency situation...

Time-critical health search example

An example of a time-critical health search situation drawn from the logs of the Microsoft Bing search engine.

8:08	heavy limbs fatigue
8:10	heavy limbs fatigue slurred speech
8:17	stroke in women
8:39	slurred speech heavy limbs
8:42	best emergency room in sugarland
8:44	best hospital in sugarland

Figure 1: Sample time-critical search session.

People frequently turn to search engines for assistance with urgent medical needs

- While people are waiting for emergency medical services (EMS)
  - Especially in remote areas
- To help people understand the urgency of situations. EMS underutilized because:
  - uncertainty about the relevance of symptoms
  - economic disadvantage
  - reliance on self-treatment

Some background...

Studies have shown that results returned by search systems for urgent health scenarios are irrelevant, inaccurate, and do not consider the influence of cognitive load on people facing emergency situations.

An initial survey conducted to better understand urgency in search settings

- Estimate the frequency with which people turned to search engines when faced with time-critical needs
- General circumstances of the situations
- Understand how the degree of urgency influenced the use of search engines
- Understand the types of urgent problems experienced among all users vs. those who used a search engine

Survey Details

- Employed Amazon’s Mechanical Turk
- The sample comprised of ATM Masters
- 133 respondents
- Only considered participants who had recalled being in an emergency situation within the past year

Two Preliminary Surveys

The first aimed at understanding general characteristics of how users behave in an emergency.

The second focused specifically on the use of search in emergencies.

Some results from the survey

12% of users turn to a Web search engine in an emergency.

Among those with a smartphone, 16% turned to search for additional assistance.

Participants categorize their time-critical needs using a classification available from the Medical Priority Dispatch System., a U.S. national standard for 911 response coding.

Table 1: Distribution of recent urgent problems recalled and those who had used a search engine for assistance.

All Respondents	Fall (11%), Breathing Problems (10%), Abdominal Pain (8%), Chest Pain (7%), Allergies (7%), Overdose (6%), Bleeding (6%), Other (6%), Back Pain (5%), Headache (4%), Psychiatric (4%), Eye (4%), Unconscious (3%), Pregnancy (3%), Heart (3%), Diabetes (2%), Convulsions (2%), Choking (2%), Cardiac (2%), Burns (2%), Animal (2%), Stroke (1%), Trauma (1%), Entrapment (1%), Assault (1%)
Search Respondents	Abdominal Pain (14%), Breathing Problems (10%), Other (10%), Allergies (9%), Chest Pain (7%), Fall (5%), Eye (5%), Overdose (5%), Heart Problems (4%), Headache (4%), Burns (4%), Bleeding (4%), Back Pain (4%), Psychiatric (3%), Pregnancy (3%), Diabetes (3%), Cardiac (2%), Animal (2%), Unconscious (1%), Convulsions (1%), Choking (1%), Trauma (1%), Heat Cold (1%), Electricity (1%)

Distribution of urgent problems reported by users

Some more findings from the initial survey on how people use search for emergency situations

- Online search precedes a visit to the emergency department.
- Search engines are used to guide searchers through the steps of the treatment process.
- Searches follow a call to 911.

83% of users were satisfied but some explained that they could not blame the search engine since their symptoms could mean many different things.

Some more findings from the initial survey on unsatisfied users

- The quality of displayed results questioned
- Too many advertisements and ‘wishy washy’ answers due to litigation possibilities
- Too many search results come up
- Some users were moved to anxiety by results that turned out to be erroneous

Summary of findings from initial surveys

- People rely on search engines for urgent needs
- Search is often used as a first response in emergency situations
- People are dissatisfied with search results in terms of both quality and quantity of returned results

What the search engine should do in urgent emergency situations

- Detect the time-criticality surrounding a query
- Retrieve appropriate content
- Identify, construct, or manually curate content designed for assisting with action in urgent situations

Focus of this paper...

The research of this paper is to employ supervised machine learning to develop a classifier of whether or not a search might have a user-perceived time-critical health information need.

Training Data

Human-generated labels for positive or negative time-critical sessions.

1. Identify sessions containing health queries
2. Remove sessions containing adult content
3. If a session contains a call/search for an emergency room and ends or the topic doesn’t shift away from health

Manually categorized training data

Table 2: Top categories of a random sample of positive and negative training data.

MeSH Positive Category (%)	Sample Queries from Positive Set
Bacterial Infections (18%)	[symptoms of appendicitis], [what do u do when u have strep]
Virus Diseases (12%)	[can you get shingles more than once], [pregnant flu symptoms]
Pathological Conditions (12%)	[sudden upper abdominal pain], [kidney stone signs]
Respiratory Tract Diseases (8%)	[pneumonia in the elderly], [pneumonia symptoms 1 year old]
Wounds and Injuries (6%)	[broken toe what to do], [sudden bruising in foot]
Pregnancy Complications (6%)	[having an iud and miscarrying symptims],[pregnancy calculator]
MeSH Negative Category(%)	**Sample Queries from Negative Sessions**
No Health Query (15%)	[what makes blood boil], [heart of glass], [protein shakes]
Reproductive Phenomena (10%)	[pregnancy calculator week by week], [can you ovulate twice in one month]
Organic Chemicals (8%)	[cexedrine], [azithromycin 250 mg], [divalproex er], [tylenol pm recall]
Pathological Conditions (7%)	[weight loss patch], [ingrown toenail]
Neoplasm (7%)	[esophageal cancer], [skin cancer], [signs of bone cancer in leg]
Heterocyclic Compounds (7%)	[what is melatonin usex for], [how long does hydrocodone stay in your system]

Features

- User characteristics
  - Likelihood of having an emergency over time
  - Length of queries
- Historical Query Statistics
  - Query and browsing behavior
  - Timing and clickthrough rate
- Query Words
  - Such as help, choking, chest pain, trouble breathing
  - Don’t include emergency room

Situated features

- Behavioral
  - Queries in the session so far
  - Average time between those queries
  - Number of search results visited by searchers
- Geospatial
  - Located at a place where likely to engage in physical activity
  - Place difficult for EMS to get to
    - Located where large number of people are likely to gather
    - On a boat or hiking/climbing
  - Current velocity or distance traveled
  - Current distance from nearest urgent care center

More situated features

- Temporal
  - Time of day
  - Day of the week
  - Hours of darkness
- Topical
  - Sudden emergence of new topics
  - Relationship between the topics that the user is interested in

Table 3: Features used in predicting urgent information needs

Feature Type	Description of Features [Number of Features]
Bag of Words	Words in the query (unigrams) [Undersampling: 2,132, Realistic: 56,429]
Historic Query Statistics	Number of times the query was issued, Average number of ad clicks per query instance, Average number of result clicks per instance, Average dwell time of result clicks, Average time to first result click, Average time to last result click, Average number of successful clicks (dwell $\geq$ 30s), Average number of switches to other engine [43 – includes additional if null features]
User-Related (preceding current session)	Number of queries from current user, Number of sessions from current user, Number of days with a query from current user, Number of days with a query from current user, Number of time-critical facility queries from current user, Number of calls or directions from current user, Average query length for current user [7 features]
Situated	[43 features]:
Behavioral	Number of queries, Average term overlap between consecutive queries, Average length of queries, Average time between queries, Number of queries without result clicks, Number of queries with result clicks, Number of result clicks, Number of unique Web domains for result clicks, Duration
Geospatial	Distance traveled, Average speed of travel, Altitude, Distance to nearest {urgent care facility, hospital, park, mall, school, pitch (soccer field), recreation ground, body of water, retail store, kindergarten, sports stadium, running track}
Temporal	Current query time in three hour bucket (e.g., 12PM-3PM), daylight (Between sunrise and sunset in the location of query at day of year), working hours (hospital open), weekend, Saturday, Sunday
Topical	Number of unique topics (based on top result for query), Number of topic changes between consecutive queries, Number of health queries, Number of new (not in user history) topics, Health query is new topic for user, Change in topic at health query (versus previous topic)

List of different features

Experiment Setup

- Collected 6 months of search activity from Bing mobile search spanning 9/2012 - 2/2013
- Found 536 positive examples (0.6%)
- Found 945,989 negative examples (99.4%)
- Train a boosted tree classifier
- 10-fold cross validation (repeat each 10 times)
- Results are reported in terms of precision and recall

Downsampling for first study

Sample smaller set from the negatives in order to have an equal number of positives and negatives.

Giving the performance of random guessing a 50% baseline.

Ability to asses each of the feature categories separately to understand their benefit.

Different feature subsets obtained via downsampling negative examples

Numbers are averages (standard deviation) over 10 runs.

Features	Accuracy	Positive Precision	Positive Recall	Negative Precision	Negative Recall	Area Under Curve
Bag of Words	0.665 (0.044)	0.706 (0.057)	0.574 (0.068)	0.639 (0.040)	0.757 (0.059)	0.712 (0.044)
Query Statistics	0.610 (0.047)	0.606 (0.055)	0.590 (0.069)	0.618 (0.044)	0.631 (0.081)	0.650 (0.054)
User-related	0.556 (0.044)	0.562 (0.045)	0.551 (0.066)	0.551 (0.045)	0.562 (0.066)	0.587 (0.047)
Situated	0.771 (0.042)	0.806 (0.057)	0.697 (0.057)	0.749 (0.038)	0.841 (0.056)	0.834 (0.040)
All	0.815 (0.042)	0.846 (0.059)	0.745 (0.068)	0.797 (0.044)	0.877 (0.054)	0.876 (0.040)

Second study in a realistic setting

- Train on sessions from January 2013
  - 139 positives, 230K negatives
- Evaluate on sessions from February 2013
  - 136 positives, 238K negatives

Using all features, compare ability to predict time criticality after first and after the last health query in a session. (If miss prediction after the first, should be able to predict after the last...)

Prediction results for realistic setting

Session-depth	Accuracy	Positive Precision	Positive Recall	Negative Precision	Negative Recall	Area Under Curve
First health query	0.999	0.875	0.883	0.999	0.999	0.999
Last health query	0.999	0.888	0.928	≈ 1	0.999	0.999

More session data means more opportunity to detect time criticality!

Discussion

Imperfect training data:

Users search for an emergency room and not be in a time-critical situation or don't search for an emergency room when they are.

Biased info from mechanical turk.

Not including people who do not have access to search engines. (Or mobile phones.)