

## 1. Overview

We address two questions:

- Do users have a sense of how useful any given query will be? and
- Can they anticipate the effectiveness of alternative queries for the same information need?

We asked crowd workers to **predict relative query quality** and compared their evaluations to a retrieval systems result.

In the process we **employed two distinct interfaces** and **measured the agreement between workers** within and across each interface.

## 2. Preliminary Task

In each of two different interfaces workers were first presented with a retrieval need expressed as a backstory and asked to respond to three initial statements via a five-point Likert item from “strongly disagree” to “strongly agree”:

- I am familiar with this topic;
- this topic is interesting to me; and
- it should be easy to find relevant information for this topic.

The users were next requested to write a query that they would use to search for information on that topic.

## 3. Main Task

After the worker completed the first part, they were presented with the main task and asked to either *rank* or *rate* queries related to the topic.

**Backstory:**

A tooth at the back of your jaw is giving you a lot of pain - you think it might be an abscess. What treatments are available for it?

**Backstory:**

A tooth at the back of your jaw is giving you a lot of pain - you think it might be an abscess. What treatments are available for it?

**Reminder:** The best query is the one most likely to generate useful results.

**Reminder:** A good query is one that is likely to generate useful results.

1 star: very bad query 5 stars: very good query

**Queries**

abscess treatment

tooth abscess treatment

treatment for jaw pain

Rank the best query on top

1 tooth pain treatment

2 will tooth fairy help with pain

3 treatment for tooth abscess

tooth pain treatment ★★★★★ 4

treatment for jaw pain ★★★★★ 3

tooth abscess treatment ★★★★★ 5

abscess treatment ☆☆☆☆☆

will tooth fairy help with pain ★☆☆☆☆ 1

treatment for tooth abscess ☆☆☆☆☆

(a) Partially completed query ranking task

(b) Partially completed query rating task

## 4. Data Collection Results

- Data was collected for 12 topics (from the UQV100 collection);
- Five queries were evaluated by users for each topic;
- Overall 100 workers completed each HIT;
- Each worker worked only on one interface type;
- We approved 93.8% of the HITs for payment;
- After additional post filtering 86.3% (518/600) of the HITs left; and
- A total of 267 workers participated.

## 5. Worker Consistency

Means and confidence intervals of worker responses, with queries presented in order of decreasing mean ranking.

95% Confidence Intervals were computed using a bootstrap method with 1,000 iterations for each interval.

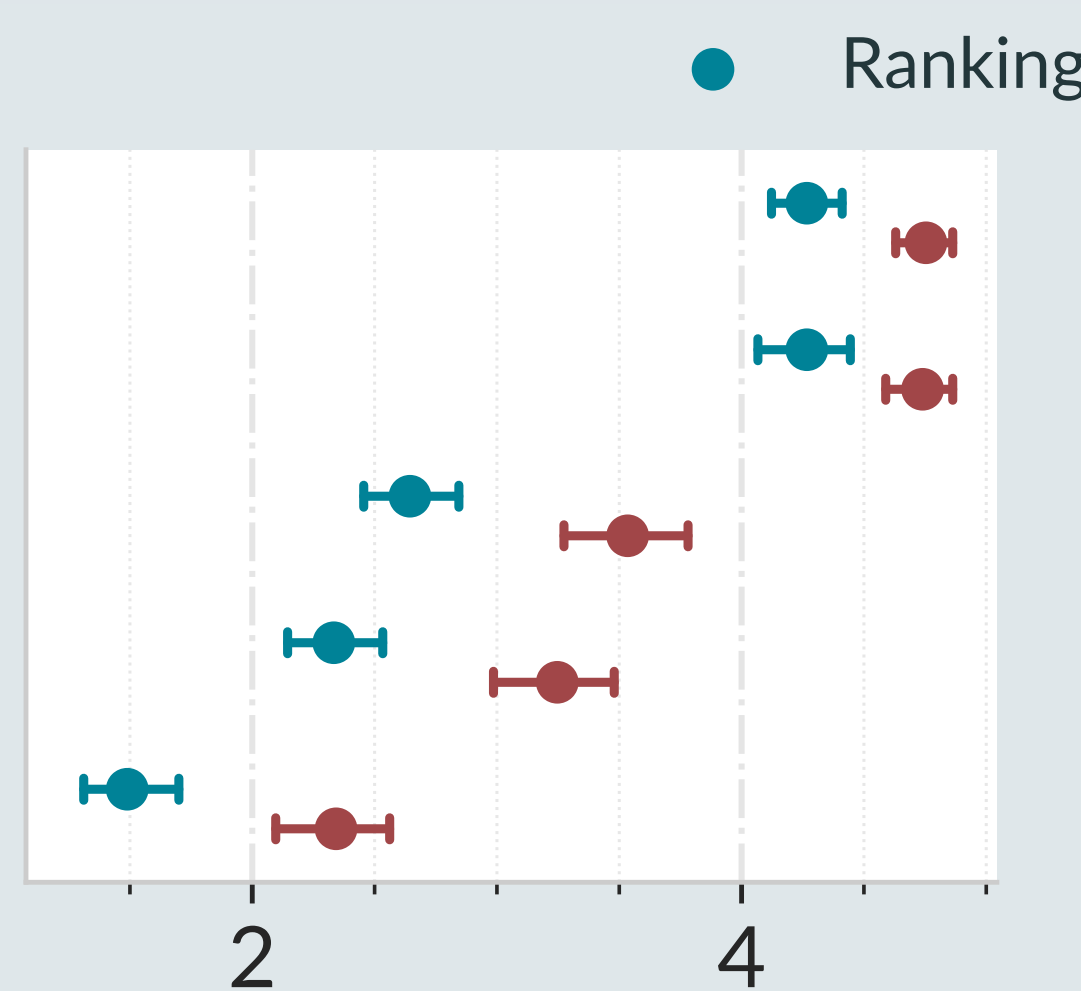
tooth abscess treatment

treatment for tooth abscess

abscess treatment

tooth pain treatment

treatment for jaw pain



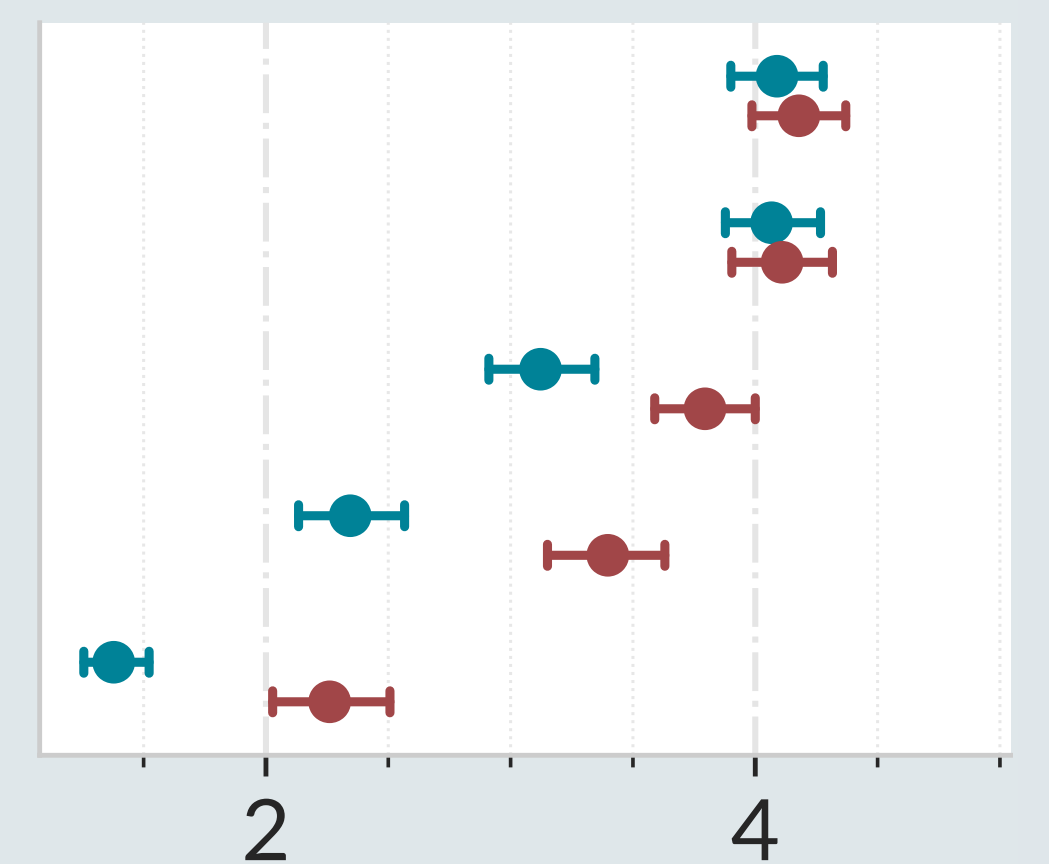
carpal tunnel syndrome

what is carpal tunnel syndrome

carpal tunnel syndrome wiki

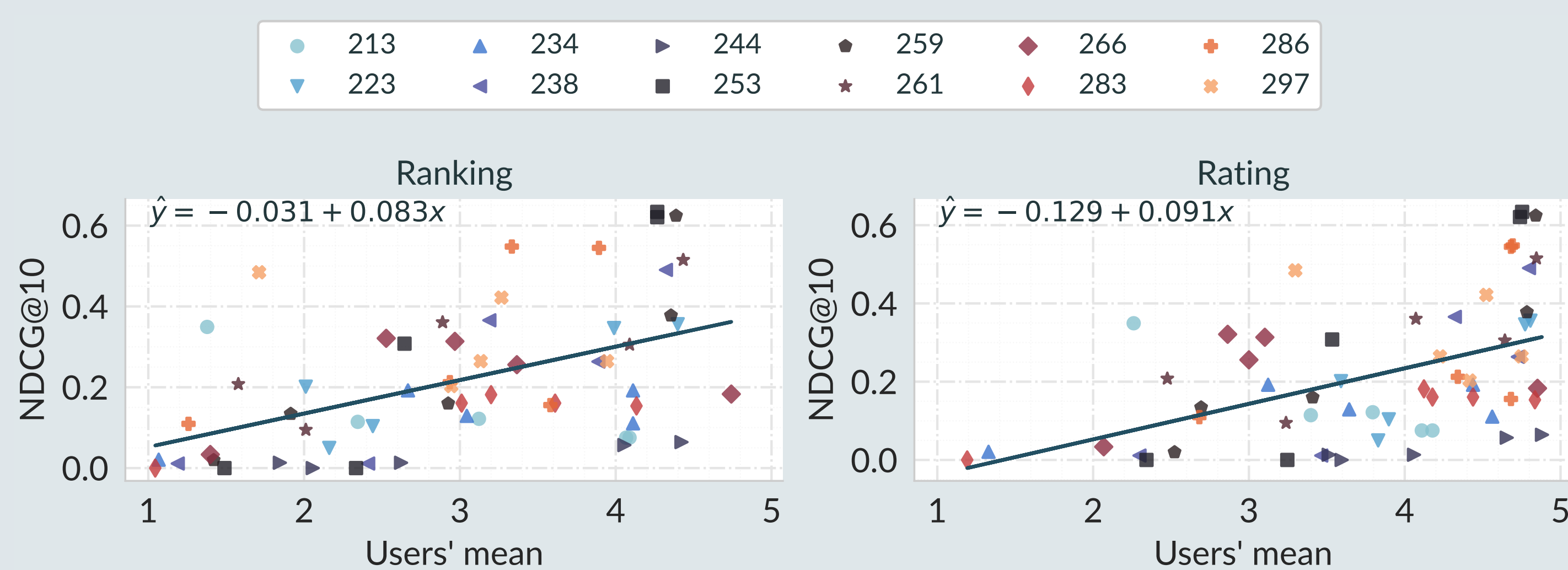
carpal tunnel

cts syndrome



## 6. Worker Predictions

Scatter-plot of NDCG@10 scores (vertical axes) against crowd worker opinion in regard to query usefulness (horizontal axes), using rankings (left pane) and ratings (right pane). The line of best fit is overlaid for each set of 60 data points.



The Pearson correlation coefficients all with significance  $p < 0.001$ :

	Ranking - User Mean	Rating - User Mean
NDCG@10	0.50	0.49
Rating - User Mean	0.90	1.00

## 7. Summary

Our experiments revealed three interesting relationships:

- Crowd workers are consistent in their evaluation of query quality;
- The “rating” and “ranking” interfaces yielded consistent outcomes; and
- Crowd worker evaluations of query usefulness correlated with actual effectiveness from a typical retrieval system when SERPs were evaluated using NDCG@10.

## Paper and Data

The paper is available at: <https://doi.org/10.1145/3477495.3531893>.  
Code and data are available from: <https://github.com/Zendelo/cs-qp>.

## Funding

This work was supported by the Australian Research Council’s *Discovery Projects Scheme* (grant DP190101113). The conference registration was funded by a SIGIR student travel grant.

