What do Mainers Search About?

Abigail Pitcairn, University of Southern Maine Department of Computer Science – abigail.pitcairn@maine.edu
Behrooz Mansouri, University of Southern Maine Department of Computer Science – behrooz.mansouri@maine.edu

Abstract

Local search leverages geographic data to refine search results according to the user's location, providing insights into region-specific interests and needs. Analyzing local search history has become valuable for understanding these regional interests, catering to localized needs, and enabling targeted decision-making [1].

Introduction

Prior research has demonstrated that examining web search engine query logs can help understand various facets of user search behavior [2, 3, 4]. This research aims to explore web search behavior in the state of Maine, a state with distinct socioeconomic and cultural characteristics.

Research Questions

- How similar are queries from Maine versus other regions of the United States (Texas and New York) and versus the United States as a whole?
- How similar are Maine queries from the last 12 months versus Maine queries from the last 5 years?

Methods

- Download Google Trends data from Maine, Texas, New York, and the United States from 15 categories from the 12-month and 5-year time periods
- Calculate cosine similarity using BERT model and Dice coefficient for similarity (Table 1):
  - Avg. similarity of queries by category
  - Avg. similarity across all categories
- Identify lowest similarity categories between regions
- Calculate query length by number of characters and number of words for each category for each region and compare (Table 2)
- Pull query examples from areas of interest and analyze text
- Identify possible explanations for low query similarity

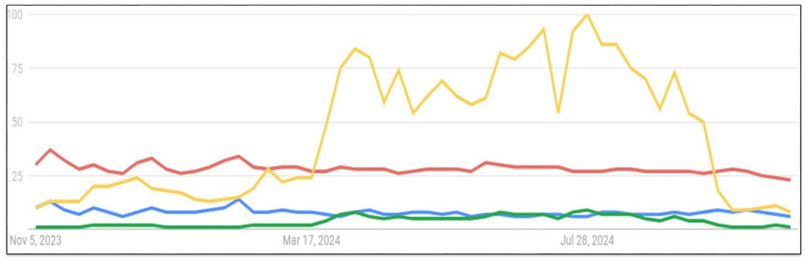

Figure 1: Interest over time on Google Trends for four queries over past 12 months: Las Vegas in United States (Red), Las Vegas in Maine (Blue), Red Sox in United States (Green), and Red Sox in Maine (Yellow)

Query Similarities from 12-month Data								
Category	Region							
	Maine vs. Texas		Maine vs. New York		Maine vs US		Maine 12mo vs. 5yr	
	Dice Co.	Cos. Sim.	Dice Co.	Cos. Sim.	Dice Co.	Cos. Sim.	Dice Co.	Cos. Sim.
Average Similarity	**0.54**	**0.78**	**0.58**	**0.80**	**0.61**	**0.81**	**0.77**	**0.89**
Arts & Entertainment	0.52	0.73	0.68	0.83	0.68	0.82	0.76	0.87
Beauty and Fitness	0.48	0.72	0.52	0.74	0.52	0.72	0.72	0.89
Books and Literature	0.64	0.84	0.68	0.86	0.68	0.88	0.76	0.88
Food and Drink	0.56	0.83	0.56	0.84	0.68	0.87	0.88	0.94
Health	0.44	0.69	0.48	0.74	0.48	0.71	0.60	0.76
Hobbies and Leisure	0.52	0.81	0.52	0.78	0.64	0.86	0.84	0.90
Jobs and Education	0.44	0.71	0.44	0.71	0.44	0.66	0.84	0.92
Shopping	0.64	0.84	0.60	0.82	0.64	0.84	0.8	0.90
Sports	0.52	0.83	0.60	0.89	0.60	0.88	0.88	0.96

Table 1: Similarity Scores for 12 Month Data

Average Query Length					
	TX 12mo	NY 12mo	US 12mo	ME 12mo	ME 5yr
Average # Words	**1.304**	**1.291**	**1.285**	**1.355**	**1.397**
Average # Characters	**7.63**	**7.51**	**7.73**	**8.11**	**8.52**
Business and Industry	7.44	7.08	7.44	7.72	7.72
Health	7.64	7.40	8.08	8.84	10.00
Jobs and Education	7.68	7.76	8.60	11.04	10.84
News	8.48	7.88	9.72	10.00	10.00
Online Communities	7.72	7.72	8.20	7.50	8.68
People and Society	7.68	7.96	7.26	8.28	8.56
Shopping	6.16	5.64	5.68	6.60	6.80

Table 2: Average Query Lengths

Results

Maine was most similar to the US region, then NY, then TX. Maine 12 mo. queries were 9.4% more similar to Maine 5yr. queries than to US 12 mo. queries.

The lowest similarity between regions was in the categories: *Health, Jobs and Education, Arts and Entertainment,* and *Beauty and Fitness.*

Highest similarity between regions was in the categories: *Shopping*, *Books and Literature, Sports,* and *Food and Drink*.

The average Maine query had about 4.9% more characters and 5.3% more words than the average US query. Most notable is the *Jobs and Education* category where Maine queries were about 32% longer than the average of query length from TX, NY, and US for 12-month data (Table 3).

Longest Queries from *Jobs and Education*			
ME 12mo		US 12mo	
Query	**Length**	**Query**	**Length**
University of Southern Maine	28 char.	*interview questions*	19 char.
University of Maine	19 char.	*google classroom*	16 char.

Table 3: Sample of Longest Queries

Applications and Future Work

By identifying trends in search queries from different regions, we can refine information retrieval systems to perform better for specific regions by filtering out non-applicable search results and prioritizing search results that align with query trends. Further work will explore popularity over time for top searches in Maine, similar to the data shown in Figure 1.

[1] Gan, Qingqing, et al. "Analysis of Geographic Queries in a Search Engine Log." Proceedings of the first international workshop on Location and the web. 2008.
[2] Craswell, Nick, et al. "Orcas: 18 million clicked query-document pairs for analyzing search." Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020.
[3] Mansouri, Behrooz, et al. "Online job search: Study of users' search behavior using search engine query logs." The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018.
[4] Scells, Harrisen, et al. "The Impact of Query Refinement on Systematic Review Literature Search: A Query Log Analysis." Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval. 2022.