

# Online Search Engine Advertising Data Visualization Tool

Project Proposal

Yingsai Dong

dysalbert@gmail.com

Department of Computer Science  
University of British Columbia  
CPSC 547 Information Visualization  
October 31, 2014

## Brief Introduction

At this time when people can buy almost everything on the Internet, online advertising has become so important if you want to bring more attention to your website. Especially when you own an online store, advertise your services or demonstrate your art works and a lot other different types of website, better online advertising means more profits actually. It's obviously the most efficient advertising skill and tool nowadays compared to all the other traditional media platforms.

I want to focus on the search engine advertising. Everyone spending sometimes online are familiar with what happen when you put a query/keyword in the search engine. The moment you hit the search button you will see some ads related to that keyword around your organic search results (non-ads search results related to your keyword). Then people usually will find what they want in the first one or two pages and click on them and then come to the landing pages.

What I want to do is to create a visualization tool to help advertisers discover profitable keywords, get other useful information and eventually improve their ranking in the organic search and update their search engine advertisement optimizing strategy. It would include two parts: domain analysis and keyword research. I have a paid account of SEMrush.com which collect online advertising data from Google in different local regions. It is huge dataset, providing updated data with sufficient information for this project.

# Personal Expertise

I have been doing research about ad auction for almost a year at UBC as a graduate student. However, it's quite theoretic. My work is mainly about equilibrium analysis for several specific ad auction mechanisms.

I did an internship at Google with a DoubleClick display ads reporting team. What I did there with my internship project is mainly about ads users log data process for generating campaign report through different pipelines. Though it's for display ads rather than search engine ads, I did get some experience of handling online ads data.

# Proposed InfoVis Solution

As you can see from the following two pictures, all the data we can get from this dataset are just plain data in a table. No matter the keywords for a domain name or reports for a specific keyword, they are in the format of rows of data with different associated information across columns. It's very inefficient for the users to look up and different to have insights.

PAID SEARCH POSITIONS 1 - 100 (340,338) ⓘ

Filter by keyword  Filters Export

Ad	Keyword	Pos	Block	Volume	CPC	URL	Traffic %	Costs %	Com.	Results	Trend	SERP source	Last update
	<a href="#">tripadvisor</a>	1 (1)		1,830,000	0.09	<a href="#">www.tripadvisor.com/</a>	15.31	0.98	0.12	78,800,000			10 hr ago
	<a href="#">trip advisor</a>	1 (1)		368,000	0.12	<a href="#">www.tripadvisor.com/</a>	3.07	0.26	0.11	118,000,000			10 hr ago
	<a href="#">florida keys</a>	1 (1)		74,000	0.80	<a href="#">www.tripadvisor.com/</a>	0.61	0.35	0.20	1,390,000			10 hr ago
	<a href="#">outer banks</a>	1 (2)		60,500	1.08	<a href="#">www.tripadvisor.com/Banks</a>	0.50	0.38	0.50	1,490,000			10 hr ago
	<a href="#">jekyll island</a>	1 (1)		49,500	1.55	<a href="#">www.tripadvisor.com/Island</a>	0.41	0.45	0.68	243,000			10 hr ago
	<a href="#">travel advisor</a>	1 (1)		40,500	0.34	<a href="#">www.tripadvisor.com/</a>	0.33	0.08	0.18	62,300,000			10 hr ago
	<a href="#">tulum mexico</a>	1 (3)		40,500	1.07	<a href="#">tulum.tripadvisor.com/</a>	0.33	0.25	0.28	11,400,000			10 hr ago
	<a href="#">turks and caicos resorts</a>	1 (1)		40,500	2.47	<a href="#">www.tripadvisor.com/Caicos</a>	0.33	0.59	0.87	33,900,000			10 hr ago
	<a href="#">niagara falls canada</a>	1 (1)		40,500	1.32	<a href="#">www.tripadvisor.com/Falls</a>	0.33	0.31	0.62	46,800,000			10 hr ago
	<a href="#">cabo san lucas</a>	2 (2)		135,000	1.45	<a href="#">www.tripadvisor.com/Lucas</a>	0.31	0.32	0.66	1,120,000			10 hr ago
	<a href="#">big bear cabins</a>	1 (1)		33,100	3.55	<a href="#">www.tripadvisor.com/Lake</a>	0.27	0.69	0.94	1,040,000			10 hr ago
	<a href="#">tripadvisor.com</a>	1 (1)		33,100	0.15	<a href="#">www.tripadvisor.com/</a>	0.27	0.02	0.09	838,000,000			10 hr ago

Figure. 1 Paid keywords for tripadvisor.com

Phrase match report				LIVE UPDATE Related keywords report			
Keyword	Volume	CPC	SERP source	Keyword	Volume	CPC	SERP source
flowers	673,000	7.71		proflowers	301,000	0.23	
flowers in the attic	201,000	7.50		1800flowers	301,000	0.17	
flower	135,000	5.07		fid	135,000	0.19	
flower delivery	135,000	6.39		flower	135,000	5.07	
1800 flowers	110,000	0.23		flower delivery	135,000	6.39	
flower girl dresses	90,500	0.89		1800 flowers	110,000	0.23	
1-800-flowers	90,500	0.18		1-800-flowers	90,500	0.18	
lotus flower	74,000	1.14		pro flowers	49,500	0.60	
pro flowers	49,500	0.60		florist	49,500	5.60	
flower tattoos	49,500	0.32		flowers.com	40,500	5.07	

Figure. 2 Reports for keyword “flowers”

Currently the InfoVis solution I am planning to create will let users dynamically filter the results by different multiple dimensions. The filtered results should simultaneously be shown to the users. There are a lot different dimensions in a domain name report like figure 1 and a keyword report in figure 2. The dimensions could be filtered should be selected with caution. And they should just be part of the dimensions, not all of them. The rest should be shown separately. The ideal interface should be roughly like this:

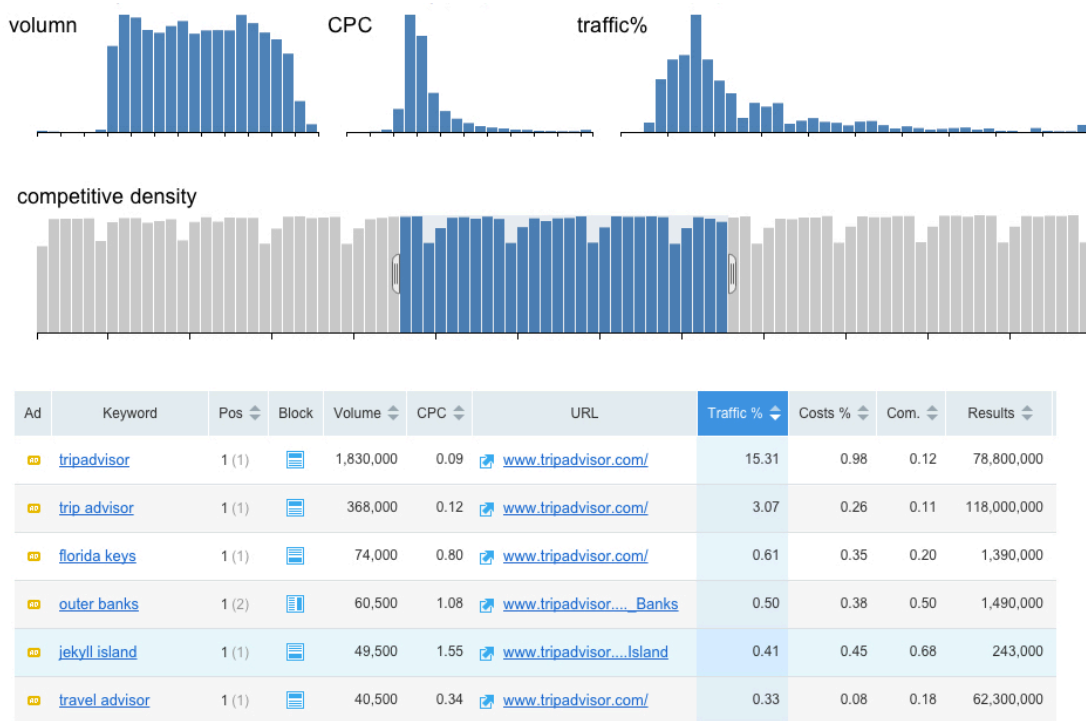


Figure. 3 Demo interface

## Scenario of Application

This visualization tool or interface could be efficiently as a part of the online advertising research report. It would effectively improve the user experience of people doing the analysis when they want to improve their search engine advertising optimization based on data from resources like SEMrush.com. My proposed interface will include user-friendly interactions for users to explore data that want and easily get insights.

## Proposed Implementation Approach

//You don't need lots of detail, just high-level things like which language and platform(s) you will use, and whether you will build on any pre-existing software or toolkits.

This project is a programming project. The technique I want to use is d3.js especially the CrossFilter library. Web application is a trend recently. D3.js is based on HTML5 and JavaScript. The application built with it could be easily used on most browsers on any platform. You never need to independently install any software on my machine. You can do everything on your browser. This minimizes the requirements for users. I have experiences using HTML4 and I have taking online courses for HTML5. JavaScript is something I have TAed for lower-level undergraduate students. I just need to check out some of the new features. The d3.js framework and Crossfilter library are definitely what I need to learn. I have started working on that.

## Milestones and Schedule

- Proposal (2014/10/31)
- Interface and feature design in details (2014/11/7)
- HTML5 and JavaScripts (2014/11/14)
- D3.js and CrossFilter (2014/11/21)
- Implementation for domain report (2014/11/28)
- Implementation for keyword report (2014/12/01)
- Wrapping up for entire implementation(2014/12/5)
- Testing (2014/12/8)
- Preparation for final presentation (2014/12/11)

## **Previous Work**

I have designed and implemented several website using HTML4, JavaScript and CSS. However I have to get used to the new features of HTML5 which has a huge gap with HTML4.

I have done some theoretic research work on online ads auction.

I have done an internship at Google in summer 2014 with a display ads reporting team.