Reviewing Crowdsourcing literature in User Interface Design - Using LISC package

Sanaz Nabavian

August 6, 2020

Abstract

The size and growing scale of the scientific literature result in a time-consuming manual process and limited scope literature summarization methods. These methods, including meta-analyses and systematic reviews, could be an excellent start to find the gap in any area. As a Ph.D. student, I should search for a vast amount of literature to find my topic of interest. I will use the package called LISC in this report for my research area, crowdsourcing, to find the recent two-year research focus. It takes the keywords you are aiming to search, using APIs(application programming interfaces) to collect data from different databases and save the results in defined objects.

1 Introduction

LISC is an interface using APIs connected to Pubmed database, providing access to collect and analyze biomedical literature, and to the OpenCitations database (Heibi,et al., 2019), providing access to citation data (Donoghue, 2019). LISC offers Three types of literature data collection: Counts, Words, and Citations(Donoghue, 2019). In this report, I will use this package to find a general understanding of the literature in my area of interest, designing a user interface for a crowdsourcing website.

2 Method

Supported API

Entrez Programming Utilities (E-utilities) It consists of nine server-side programs that provide an interface to the Entrez query and database system at the National Center for Biotechnology Information (NCBI). These server-side programs are: 1- EInfo (database statistics), 2- ESearch (text searches), 3- EPost (UID uploads), 4- ESummary (document summary downloads), 5- EFetch (data record downloads), 6-ELink (Entrez links), 7- EGQuery (global query), 8- ESpell (spelling suggestions), 9- ECitMatch (batch citation searching in PubMed)To access these data, a piece of software first posts an E-utility URL to NCBI Then, it retrieves the results of this posting, after which it processes the data as required. The software can thus use any computer language that can send a URL to the E-utilities server and interpret the XML response("National Center for Biotechnology Information", n.d.). Open Citations OpenCitations is an independent infrastructure organization for open scholarship dedicated to the publication of open bibliographic and citation data by the use of Semantic Web (Linked Data) technologies. It is also engaged in advocacy for open citations, particularly in its role as a key founding member of the Initiative for Open Citations (I4OC) ("Open Citations," n.d.).

Data collection

LISC currently offers the following types of literature data collection:
- Counts: tools to collect and analyze data on the co-occurrence of specified search terms
- Words: tools to collect and analyze text and meta-data from scientific articles
- Citations: tools to collect and analyze citation and reference data

Here, we will use counts object for analyzing the number of co-occurrence of the search keywords. Then, using the word object and analysis to see the new research from 2018. My research area is about crowdsourcing data, and how the user interface design can change the data quality. So I have entered these keywords and their synonyms as the input of this package.

$$terms_a = [['crowd','crowdsourcing','crowdsourced','Citizenscience'],['datacrowdsourcing','crowdsourceddata']] \quad (1)$$

$$terms_b = [['data','quality','dataquallity'],['design','userinterface','UIdesign','UI']] \quad (2)$$

3 Result

The first result shows count data for the word co-occurrence data between terms. Co-occurrence data is a matrix of numbers reflecting the relationship between terms.

```
[[     0 396426   2152    576]
 [396426      0    457     73]
 [  2152    457      0    576]
 [   576     73    576      0]]
```

LISC makes clustering analysis and visualizations, that attempt to find structure in the data, as shown in Figures 1 and 2.These are based in a matrix of numbers reflecting the relationship between terms.

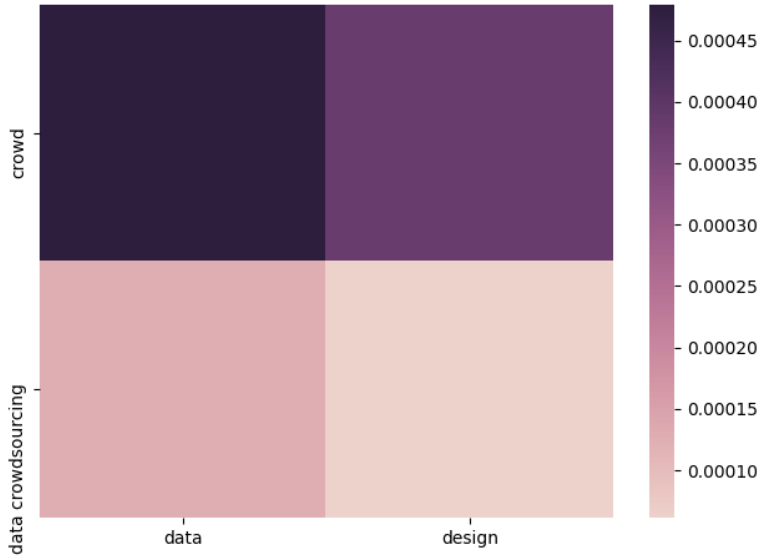

Figure 1: Figure 1.

The words object provides a word cloud of the collected data for the first term (crowdsourcing), as shown in Figure 3.

You can find more details about the articles' title, doi, and journal with this object. Below is a sample output showing the titles of articles amongst 100 articles found by the term 'crowdsourcing' since 2018.

```
Causal explanations of depression on perceptions of and likelihood to choose cog
nitive behavioural therapy and antidepressant medications as depression treatmen
ts.
NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization.
A framework based on deep neural networks to extract anatomy of mosquitoes from
images.
The motivation for citizens' involvement in life sciences research is predicted
by age and gender.
Individual differences in infancy research: Letting the baby stand out from the
crowd.
A Vision for Using Simulation & Virtual Coaching to Improve the Community Practi
ce of Orthopedic Trauma Surgery.
Big Data in Ophthalmology.
What does crowdsourced data tell us about bicycling injury? A case study in a mi
d-sized Canadian city.
Examining Change in Self-Reported Gambling Measures Over Time as Related to Soci
ally Desirable Responding Bias.
Assessment of volunteered geographic information for vegetation mapping.
Public awareness of and attitudes towards research biobanks in Latvia.
Understanding FAHFAs: From structure to metabolic regulation.
A Comparative Analysis of Surgical Scar Cosmesis Based on Operative Approach for
```

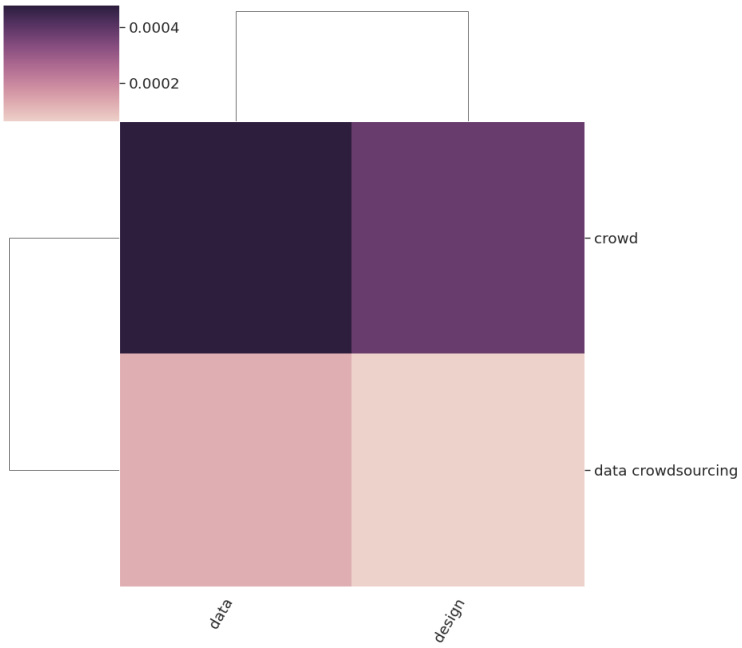

Figure 2: Figure 2.

Radical Prostatectomy.
When is SARS-CoV-2 in your shopping list?
New Coleoptera records from eastern Canada, with additions to the fauna of Manitoba, British Columbia, and Yukon Territory.
A GNSS-Based Crowd-Sensing Strategy for Specific Geographical Areas.
Breed differences in dog cognition associated with brain-expressed genes and neurological functions.
A citizen science approach estimating titanium dioxide released from personal care products.
A review of subtidal kelp forests in Ireland: From first descriptions to new habitat monitoring techniques.
Development of Brief Child Nutrition and Physical Activity Screening Questions for Electronic Health Record Use.
ACA Medicaid Expansion Associated With Increased Medicaid Participation and Improved Health Among Near-Elderly: Evidence From the Health and Retirement Study.
Crowded environment affects the activity and inhibition of the NS3/4A protease.
Sharing for science: high-resolution trophic interactions revealed rapidly by social media.
Community Assessment of the Predictability of Cancer Protein and Phosphoprotein Levels from Genomics and Transcriptomics.
A Smartphone Crowdsensing System Enabling Environmental Crowdsourcing for Municipality Resource Allocation with LSTM Stochastic Prediction.
Automated Bird Counting with Deep Learning for Regional Bird Distribution Mapping.
Real-time traffic accidents post-impact prediction: Based on crowdsourcing data.
Systematic Chemogenetic Library Assembly.
Crawling the German Health Web: Exploratory Study and Graph Analysis.
Intensive care unit (ICU) diaries and the experiences of patients' families: a grounded theory approach in a lower middle-income country (LMIC).
Peer crowd affiliations as predictors of prosocial and risky behaviors among college students.
Gender differences in attention to pain body postures in a social context: a nov

Figure 3: Figure 3.

```
el use of the bodies in the crowd task.
Research Mentorship Crowdsourcing Contest: Toward Global Health Good.
The Impact of Dermal Fillers on Perceived Personality Traits and Attractiveness.
When three is a crowd: Chaos from clusters of Kuramoto oscillators with inertia.
Outlining where humans live, the World Settlement Footprint 2015.
Venous thromboembolism research priorities: A scientific statement from the Amer
ican Heart Association and the International Society on Thrombosis and Haemostas
is.
How Do Older Adults Recruited Using MTurk Differ From Those in a National Probab
ility Sample?
Protistological science dissemination.
The Roses Ocean and Human Health Chair: A New Way to Engage the Public in Oceans
 and Human Health Challenges.
Addition of an Emotionally Stable Node in the SOSa-SPSa Model for Group Emotiona
l Contagion of Panic in Public Health Emergency: Implications for Epidemic Emerg
ency Responses.
A citation study of earth science projects in citizen science.
Gazing Without Eyes: A "Stare-in-the-Crowd" Effect Induced by Simple Geometric S
hapes.
The association between sleep quality and attenuated psychotic symptoms.
Outlier detection methods to improve the quality of citizen science data.
Nine years of mosquito monitoring in Germany, 2011-2019, with an updated invento
ry of German culicid species.
Swab and Send: a citizen science, antibiotic discovery project.
Pedestrian orientation dynamics from high-fidelity measurements.
Highlights in Medical Entomology, 2019: Familiar Foes and New Frontiers.
The extended Moral Foundations Dictionary (eMFD): Development and applications o
f a crowd-sourced approach to extracting moral intuitions from text.
```

4 References

Donoghue, (2019). LISC: A Python Package for Scientific Literature Collection and Analysis. *Journal of Open Source Software*, 4(41), 1674. https://doi.org/10.21105/joss.01674

OpenCitations. (n.d.). OpenCitations. https://opencitations.net/

Sayers, E. (2020, August 6). *A General Introduction to the E-utilities - Entrez Programming Utilities Help - NCBI Bookshelf*. The National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/books/NBK25497/