

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.

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

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

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 , 1290, 5
 1290 , 0 ,20
 5 , 20, 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.

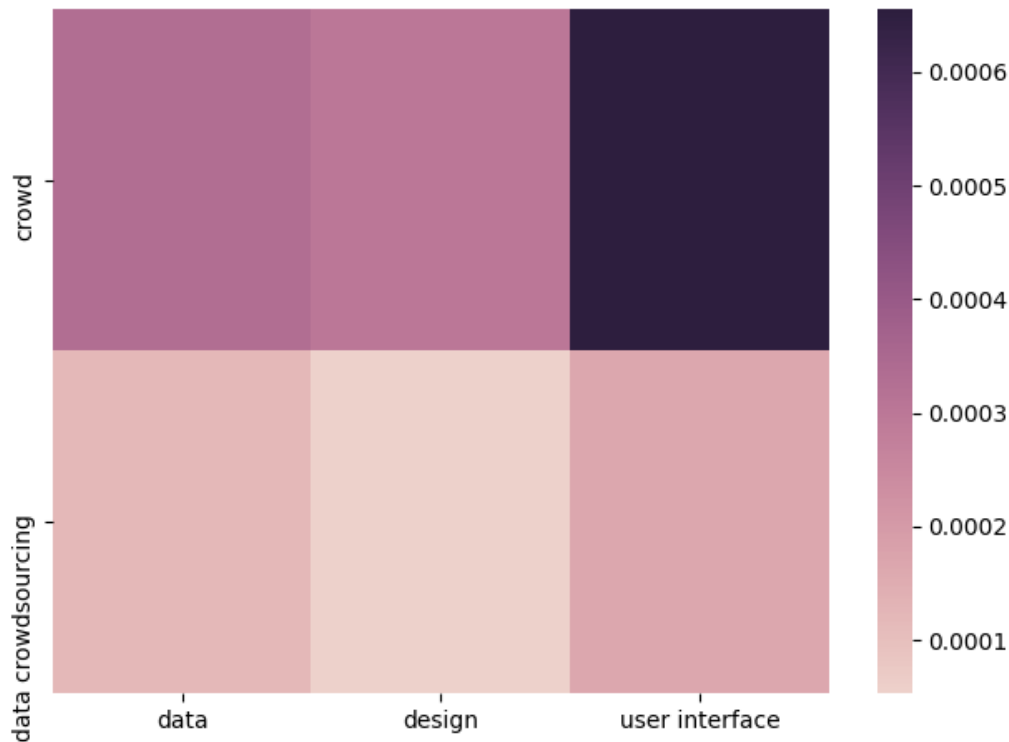


Figure 1: Figure 1.

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.

- 1 Citizen Neuroscience: Brain-Computer Interface Researcher Perspectives on Do-It-Yourself Brain Research
- 2 Evaluating the data quality of iNaturalist termite records.
- 3 Mechanical Turk data collection in addiction research: utility, concerns and best practices.
- 4 The CrowdWater game: A playful way to improve the accuracy of crowdsourced water level class data.
- 5 Applying citizen science to gene, drug and disease relationship extraction from biomedical abstracts.
- 6 The SMART Framework: Integration of Citizen Science, Community-Based Participatory Research, and System
- 7 How Comprehensive Are Media-Based Data on Police Officer-Involved Shootings?
- 8 Evaluating the ability of citizen scientists to identify bumblebee (Bombus) species.
- 9 On the impact of Citizen Science-derived data quality on deep learning based classification in marine i
- 10 Combining Citizen Science and Deep Learning to Amplify Expertise in Neuroimaging.
- 11 Connecting data and expertise: a new alliance for biodiversity knowledge.
- 12 Producing knowledge by admitting ignorance: Enhancing data quality through an "I don't know" option in
- 13 Hoping for optimality or designing for inclusion: Persistence, learning, and the social network of citi
- 14 Designing online species identification tools for biological recording: the impact on data quality and

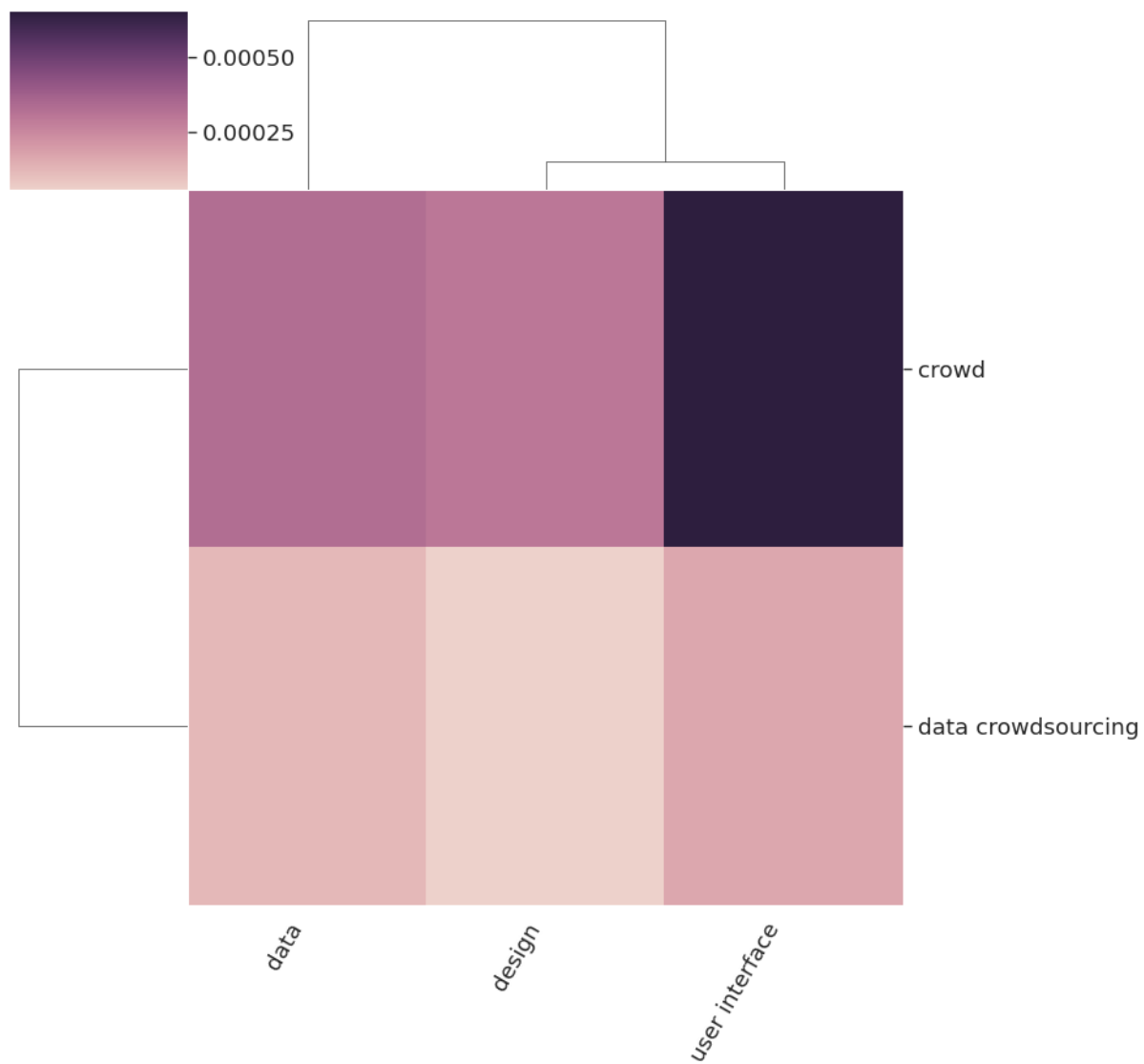


Figure 2: Figure 2.



Figure 3: Figure 3.

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/>