

Health Co-Inquiry WebCrawler Guidebook

June-2018

Lauren S. Seifert

Malone University

Author Note

Please, address correspondence to Lauren S. Seifert, Ph.D., Malone University, 2600 Cleveland Ave., NW, Canton, Ohio 44709. Email: LSEIFERT@malone.edu phone: 330-471-8558. This project was initiated under a sub-award related to an institutional grant from the National Science Foundation, CC-NIE Award Number #1541342. Special thanks go to M. Adam Klemman, Jim Shaffer, Jason Courter, and Shawn Campbell, recipients of the grant who have utilized it to help enhance our data collection. Additional thanks go to Camden Baucke, J. "Harley" Ferrel, Tess Schmucker, Matthew Beechy, Rosa Cleaver, Tyler Bowman, and Kathryn Bash, for their excellent assistance searching for URLs and composing dictionaries for specific health conditions. Special thanks go to Cory Edwards for his programming efforts.

Health Co-Inquiry WebCrawler Guidebook

This is a guide and set of instructions to accompany the 2018 Health Co-Inquiry WebCrawler. The primary purpose of the WebCrawler is to search URLs related to chronic health conditions. WebCrawler activity yields frequency data about the total number of words, number of health-related terms, global health terms, and condition-specific words that appear on those pages. In addition, the WebCrawler produces a graph to show links to the original “seed” URL. The inclusion criteria for a URL are that it must have more than 40 characters.

The next page of this Guidebook is a single-page, poster-style summary of the Health Co-Inquiry Project. Please, study it carefully to comprehend the nature of this research study.

MALONE UNIVERSITY, CANTON, OHIO, USA 44709

Technology in Health Co-Inquiry: Online Resources and Data



Background: With rapid growth of the internet in recent years, it has become an important but little used source of qualitative data about the experiences of stakeholders in management of chronic health conditions. We developed a bifurcated method to better utilize online information posted by patients/clients, caregivers, and health providers. Our technique may clarify our understanding of stakeholders in Health Co-inquiry.

Health Co-inquiry is a process whereby stakeholders in management of chronic conditions collaborate, with efforts toward stakeholder activation, person-centredness, evidence-based practice, and integrated care. With dramatic growth of internet use by the general population in the new millennium, stakeholders in management of chronic conditions may be using the internet more to seek information, share advice, and give feedback. Analysis of narratives from publicly available blogs/websites related to chronic conditions, and (2) development of a computer program that crawls the same websites (URLs) to locate and count the frequencies of general health terms and condition-specific terms and provide pictorial representations of the relative frequencies of those terms and of the networks of links between URLs. One goal is to compare results from the two branches of the bifurcated method to find convergent and divergent results. Safeguards include removing stakeholder monitors from quoted material, using only publicly available sites, never posting on sites from which we derived data, and obtaining IRB approval.

Managing chronic health conditions is a worldwide challenge, with tremendous suffering and expense being sustained (e.g., in days lost to disability, caregiver time and energy, treatment costs, and health and mental health provider effort; World Health Organization, 2011). Typically, such conditions involve health changes and management over a long period of time (WHO, 2014). So, planning, action, and evaluation of actions are essential for patient well-being. Some negative outcomes might be prevented when communication and collaboration improve between most stakeholders (patients, clinicians, and health care organizations). For example, the use of patient-centered care, which emphasizes the preferred modes of fully informed patient participation and caregivers' responsiveness in understanding the patient's health, planning and implementation of care plans (Fletcher, Linn, Dave, & Corbin-Smith, 2016). Health Co-Inquiry (Selinger, 2015) is a highly cooperative and multi-method approach. It includes stakeholder activation, person-centeredness, evidence-based practice, integrated care, and openness to subjective inquiry. Through Health Co-Inquiry, stakeholder collaboration can lead to improved outcomes.

[In healthcare, the person with a health condition is generally referred to as a "patient", and in mental health practice, s/he is called a "client". So, here we will refer to the individual as a patient/client in order to honor each discipline's preferred term.]

Presumably, the same sorts of activation might benefit caregivers and providers, too, and when all stakeholders are activated toward cooperation and positive outcomes, they can work together to achieve compromises.

Collaborating toward co-impact is vital to Health Co-Inquiry. A vibrant concept in healthcare today, *person-centredness* originated in the 1950s in psychology (Rogers, 1957). Yet, it is now in use in many fields and disciplines. Essentially, a *person-centred* approach takes into account the wants and needs of the individual, and while it is typically associated with the wants and needs of the patient/client in mental health and health care practice, a Health Co-Inquiry approach also links it to the wants and needs of other stakeholders. Like caregivers and providers. When “reciprocal person-centredness” is observed, each stakeholder can feel free to communicate about her or his wishes and needs in an open and accepting exchange—fostering mutuality and balance (Seltzer, Flaherty, & Trill, 2013; Valloir, 1966).

Having access to expertise and knowing the exact literature related to a specific health condition is absolutely necessary in the context of managing chronic illness. Usually, one's health or mental health provider is presumed to be the access point to such expertise. However, in Co-Inquiry, each stakeholder is highly regarded for his or her expertise – i.e., what knowledge the stakeholder brings to collaborative inquiry that is relevant to managing the chronic health condition. For example, a provider might know about the newest, approved treatments, a caregiver might have insights about what is possible within the patient's/client's ecology, and the person with the chronic condition might be aware of subtle nuances that are very individualized (e.g., personal habits and preferences). So, each stakeholder brings a kind of expertise that can be shared during treatment planning and evaluation. Each stakeholder's unique *value* should be treated as an *advantage*...as another piece of the puzzle, which might lead the collaborative team to assemble the whole thing.

As a fourth component, health Co-inquiry, integrated care fits the spirit of the foregoing concepts and practices. The best models of integrated care include descriptions of stakeholder activation, person-centredness, and evidence-based practices, while incorporating expertise across different types of providers (e.g., physicians, mental health professionals) in order to optimize patient/client outcomes (Jones & Ky, 2015).

This can foster patient activation and aid the cooperative endeavor between patients/nurses, their caregivers, and their health and mental health providers. One factor that seems to predict patient/clinician participation in health support online is an "intrinsic health orientation" (i.e., "wanting health information and seeking prevention"; Dutta & Feng, 2007, p. 181). Being at risk for, or being diagnosed with, a particular health condition seems to further predict an increase in seeking online support and information related to that condition (Dutta & Feng, 2007).

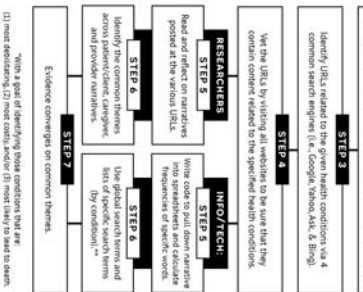
In Health Co-inquiry, persons with health conditions, their caregivers, and their providers collaborate toward positive outcomes (Selfiet, 2015). As mentioned previously, part of the collaborative endeavor is person activation, and the Internet can play a vital role. For example, persons with a cancer diagnosis who self-reported seeking support online also acknowledged that the greatest benefit of their health-related Internet use was to equip them with knowledge (Seccin, 2011). For family caregivers of persons with pediatric cancer, Coulson and Greenwood (2011) found that the highest levels of perceived support from online activity were related to information gathering and emotional buttressing (see also, Gupta & Sun, 1992). So, caregivers, too, can be activated toward and report benefits from Internet use.

STEP 1

Develop list of chronic health conditions* using WHO and US data about top 10 disabling and top 10 fatal ones.

STEP 2

Develop a list of global, health-related search terms and develop a condition-specific dictionary for each health condition.



grant sub-award from an institutional grant from the National Science Foundation to Malone University, CC-NIE Award Number #15413242. Special thanks go to M. Adam Klemman, Jim Sharley, Jason Courter, and Shawn Campbell, recipients of the grant who have utilized it to help enhance our data collection. Additional thanks go to J. "Harley" Ferrell, Jess Schumaker, Matthew Beechey, Ross Cleaver, Tyler Bowman, Camden Baucke, and Kathryn Bash, for their excellent assistance searching for URIs and composing dictionaries for specific health conditions.

Coulson, S.M., & Greenwood, N. (2011). Families affected by childhood cancer: An analysis of the provision of social support within families impacted by cancer. *Cancer Care, Health and Care Research*, 3(2), 360-370.

Dawson, C.E., & Sallis, J.F. (1997). Correlates of stressful events and satisfaction with spouse support behaviors. *Communication Research*, 19, 154-176.

Derry, K., Schoenbaum, S.C., & Aulet, A. M. (2020). A review of patient-centered primary care. *Journal of General Internal Medicine*, 35(10), 953-957.

Driscoll, C. (2000). Health Communication: A review of the predictors of social and health support. *Health Communication*, 15(2), 187-198.

Frederick, L., Choi, K.H., Dawe, G., & Corbie-Smith, B. (2012). Integrating social and community-based participatory action research to achieve health equity. *American Journal of Public Health*, 102(7), 1153-122.

Holmes, R. (2007). Integrating a clinical stage and employing a "weak steps" approach to patient and management. *Journal of Ambulatory Care Management*, 30(1), 2-8.

International Association for Participatory Public Research. (2016). About us. Retrieved from <http://www.iap2r.org/>

Internet Users. (2017). I share the same fears. Retrieved from <http://www.fox.com/>

Johnson, E. K., & Kline, D. (2015). Stress and health inequalities: are community health centers in the United States. *American journal of Public Health*, 105(10), 2028-2034.

Oh, H., & Lee, B. (2012). The effect of computer-mediated social support in online communication on patient empowerment and doctor-patient communication. *Health Communication*, 27, 34-40. doi:10.1080/10810739.2011.557449

Rege, C. K. (1981). *Death deferred: therapy, its current practice, implications, and theory*. New York: Basic Books.

Seale, C. (2011). Emotional and decisional empowerment in palliative health support communities: Initial exploratory validation of the Cyber to Decisional Empowerment Scale (CDES) and preliminary data from administration of the scale. *Support Care Oncol*, 19, 2057-2061. doi:10.1007/s00520-011-1249-y

Shelley, L. S. (2015, November). *Is there a room for traditional science and action research in aging studies?* Presented at the Fifth Annual Conference of the Society for Aging Studies, Ft. Lauderdale, FL.

Shelley, L. S., Finkbeiner, K., & Kline, C. (2013). Beyond prison-oriented dementia care: The Silver Spring Center for Day-to-Day Dementia Care. *Activities Director Quarterly for Alzheimer and Other Dementia Patients*, 42(2), 15-31.

Valletti, M.C. (1966). Existentialism: a philosophy of commitment. *The American Journal of Nursing*, 66(3), 500-505. Retrieved from [http://www.ajic.org/S0361-3617\(2\)31972-0](http://www.ajic.org/S0361-3617(2)31972-0)

World Health Organization. (2001). *International classification of functioning, disability, and health* (ICF). Geneva, Switzerland: WHO.

World Health Organization. (2014). *Profilaxis: Non-communicable diseases*. Retrieved from http://www.who.int/rates/factfiles/noncommunicable_diseases/en/

**Researchers used a list of URLs for a given condition that was verified by visiting each site to assure that it was a forum, newsgroup, or informational page that was accessible without a membership/password. If a person could enter as a guest on more than one occasion, then the URL was included. If a user was required to log in, password, email address, or other identifying information had to be e-mailed to the site before access was granted. The search engines used were Google, HotBot, Excite, and MSN. Only English speaking US patients, family/friends of patients, health providers, and/or mental health providers in order to be included.

About the global and condition-specific dictionaries, they were created by accessing our institutional library database for books and eBooks (Ovidline), a database that provides access to books from around the globe as well as a leading network of Ohio, USA, public, private, and university libraries. Using the condition name as search term, the first ten books for which we could access tables of contents were used. In the case of the global dictionary (which is applied to all health conditions) the search term was "chronic illness," but the investigators also added terms to the global dictionary at their discretion, based on their knowledge of chronic illness. The condition-specific dictionaries were developed by searching for the condition name in the global dictionary and adding relevant terms. In addition, some terms were removed, and all entries were made to book titles to avoid confusion. For example, the condition "Diabetes" was removed, and all entries were made to "book on diabetes." The condition "cognitive decline" was removed, and all entries were made to "book on cognitive decline." A few terms remained unchanged. It is assumed that, so that the two researchers can work with the condition outlines as a team, remaindained unchanged. It is assumed that,

Logging In to the WebCrawler (June-2018)

At Malone University, we use the National Science Foundation (NSF) Applications main page in order to log-in to the Health Co-Inquiry Project.

If you are from Malone University, you may click the Malone University logo in order to proceed; if you are not, then type in your institution's name.

Login using your institutional user name and password.

At any time, you may click "Options" in the upper right, in order to look at tasks, available files, or to get back to the NSF Applications page.

Once you have done this, you will see the project options that are available to you. As of June-2018, the project option for this WebCrawler is just one: the Health Co-Inquiry Project.

The "Main Page" or "Front Page"

What follows are screen shots of the WebCrawler's Main Page (Front Page). In order to change the appearance of the page so that it is black text on a white background with colors enables, go to the bottom right and click the button the "Toggle inverted colors".

Below is a screen shot of the Main Page (or front page) of the health Co-Inquiry WebCrawler. In this view, the left-most tab at the top of the central box (called the "Big Data Box") has been clicked to show "URL results". Notice that the Big Data Box lists data for all URLs selected in the "Seed URL Box" to the left of the central box. [As will be described in more detail below, condition specific and global dictionary terms can be selected or de-selected, as well as URLs.]

Condition:
Alzheimer's disease

Seed URL:
http://www.alzforum.org/
https://patient.info/forums/discuss/browse

Condition dictionary:
18F
18F-AV-45
18F-FDG-PET imaging
APB11
AD-133
Alzheimer

Globel dictionary:
accident
Ache
Aggravate
Ange
Angry
Bad

Retrieve crawl data

10

Get URLs with highest score

URL results

URL	Score	Total number of words	Seed	Number of fails	
http://www.alzforum.org/	54	1139	true	0	Extra
https://patient.info/forums/discuss/browse/alzheimer-s-disease-67	153	3165	true	0	Extra

Export retrieved data

dictionary word cloud

Filter frequency:
Alzheimer:37
Mother:11
dementia:9
sign:9
ill:8
diagnosis:7
stage:6
ill:6
Medical:6
symptom:5
Doctor:5
Symptom:5
disease:4
onset:4
work:4
Onset:4
neurodegeneration:4
Wife:4
Son:4
genetic:4
immunotherapy:4
early onset:3
traopathy:3
Trauma:3
husband:3
Current:3
Crime:3
good:2
best:2
Pain:2

Toggle inverted color

On the Main Page, there are some important features of which to take note:

Condition Selection Box: on the upper left side, this box permits the user to select a specific chronic health condition for data collection

ADVISORY: Before adding Seed URLs for a selected condition or topic, add all dictionary terms (global and condition specific). The system requires much less time to add dictionary terms than to add Seed URLs. This is because adding a Seed URL required the WebCrawler to go to that URL and to search it for links, calculate word frequencies of dictionary terms, and yield search output.

Seed URL Box: below the Condition Selection Box on the left side of the Main Page, this box allows the user to look at what seed URLs are available in the WebCrawler for the selected condition. Note that the availability of seed URLs is contingent upon whether they have been entered into the system, a task for which ONLY Health Co-Inquiry Co-PIs and the WebCrawler programmer have clearance.

The Seed URL Box shows available, searchable seed URLs. Any user can click the “edit” button that is immediately to the right of the Seed URL Box. Check or un-check URLs in order to select specific ones for evaluation.

A Seed URL is a “parent” for the two tiers of URLs that are derived from the WebCrawler search. Seed URLs are specified by the researchers, determined through our method of using 4 search engines (GoogleTM, AskTM, YahooTM and BingTM) to find websites that are related to one or more search terms (with a search term being specified by the topic under study, e.g., Alzheimer’s disease, diseases of the lung, stroke, or other). Seed URLs are entered into the WebCrawler for a given Health Condition or topic. Those Seed URLs are then searched by the WebCrawler. Currently, the WebCrawler “goes out” two tiers to find children of the Seed URLs (i.e., sites that are accessed via links at the Seed URL) and grandchildren of the Seed URLs (i.e., children of the Seed URLs children).

ADVISORY: Note that the oversight institution for the WebCrawler sometimes turns off this function, because it can hang up the crawler. As of this writing, the function is working and is turned on (June-2018).

Condition Specific Dictionary: below the Seed URL Box on the left side of the Main Page, this box’s contents will change, depending on what condition has been selected in the Condition Selection Box (above). For those who have administrative privileges, **cut and paste** a list of dictionary terms from a Word document by using the find ^p and replace \$\$ function in Word. It will replace all hard returns within your list of dictionary terms in Word with \$. This should allow you to cut and paste the entire list of terms into the “add” box in the edit menu for the dictionary. The WebCrawler will tell you whether the terms have been added successfully. Click “OK” in the information box. Then, scroll down and click “Accept”. This should take you back to the main page, and when you refresh the main page, the dictionary terms should appear in the dictionary.

Global Dictionary: below the “Condition Specific Dictionary” on the left side of the Main Page, this is a list of terms that relate to chronic illness and chronic health conditions. Every time a specific condition’s

URLs are searched, the global dictionary is applied in order to filter the terms and show their frequencies. Users should not edit the Global Dictionary; it is a pre-determined list that was generated using a specified method (see p. 3 of this Guidebook). Use the **cut and paste** instructions that are above (in the section on the Condition Specific Dictionary) in order to add global dictionary terms (for those who have administrative privileges).

Crawler Specified Dictionary Term: any word that is present in one or more WebCrawler Dictionaries (Global and/or Condition Specific).

Action Buttons for WebCrawling URLs:

a) Retrieve Crawl Data Button: below the Global Dictionary on the left side of the Main Page, users may ask the WebCrawler to retrieve data for the given condition and seed URLs that s/he has selected (instructions, as above).

b) Get URLs with Highest Scores Button: below the Retrieve Crawl Data Button on the lower left of the Main Page. The default webcrawl finds the TOP scoring URLs (among all those which have been found by the WebCrawler). To the left of the Get URLs with Highest Scores Button is a box to specify how many of the highest scoring URLs one wants to retrieve. The default is to get the 50 highest scoring URLs, but a user can change that number to get a different quantity.

ADVISORY: A user must click on one of the above two buttons to Retrieve Crawl Data or to get URLs with Highest Scores before trying to Export Data to a spreadsheet via the button at the bottom, center of the main page.

Export Data Button: at the bottom center of the Main Page, once the WebCrawler has completed a crawl and loaded the Main Page with data (e.g., in the scores and filter boxes, as described below), then click the Export Data Button in order to export the evidence to an Excel spreadsheet.

Big Data Box: central to the main screen, this box is empty until one has selected a condition with seed URLs and asked the WebCrawler to crawl. Once those actions have been taken, the WebCrawler populates the “Big Data Box” with scores (e.g., how many dictionary words from the combined global and condition specific dictionaries have been found for each URL, how many words [total] are at each URL).

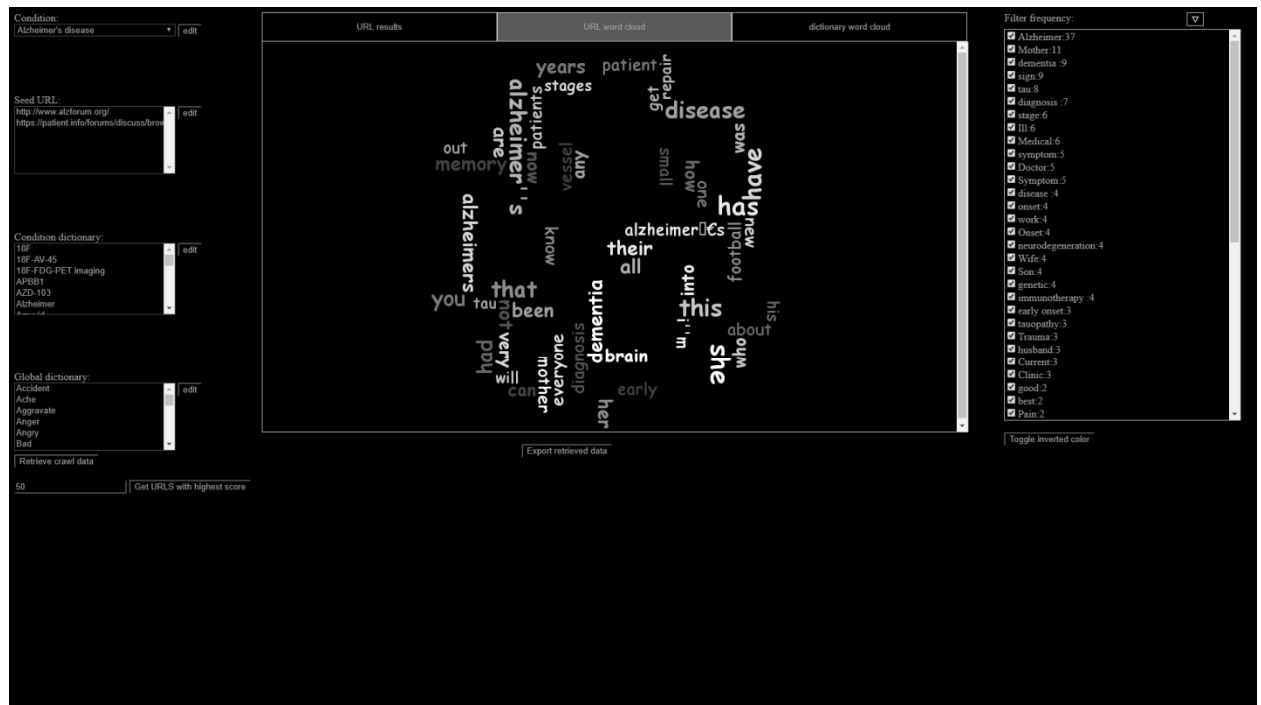
URL Results Tab: at the top left of the Big Data Box, it is also set as the default for the Big Data Box. When a user clicks “Retrieve Crawl Data” or “Get URLs with Highest Scores” the URL view of those data (with their scores) will be generated and shown in the Big Data Box. Click this tab after using one of the other two tabs, in order to return to the URL scores.

URL Word Cloud Tab: center, top of the Big Data Box. Click this to see the relative frequencies of all words among all the selected URLs. The size of each word in the cloud shows its frequency relative to the other words in the cloud.

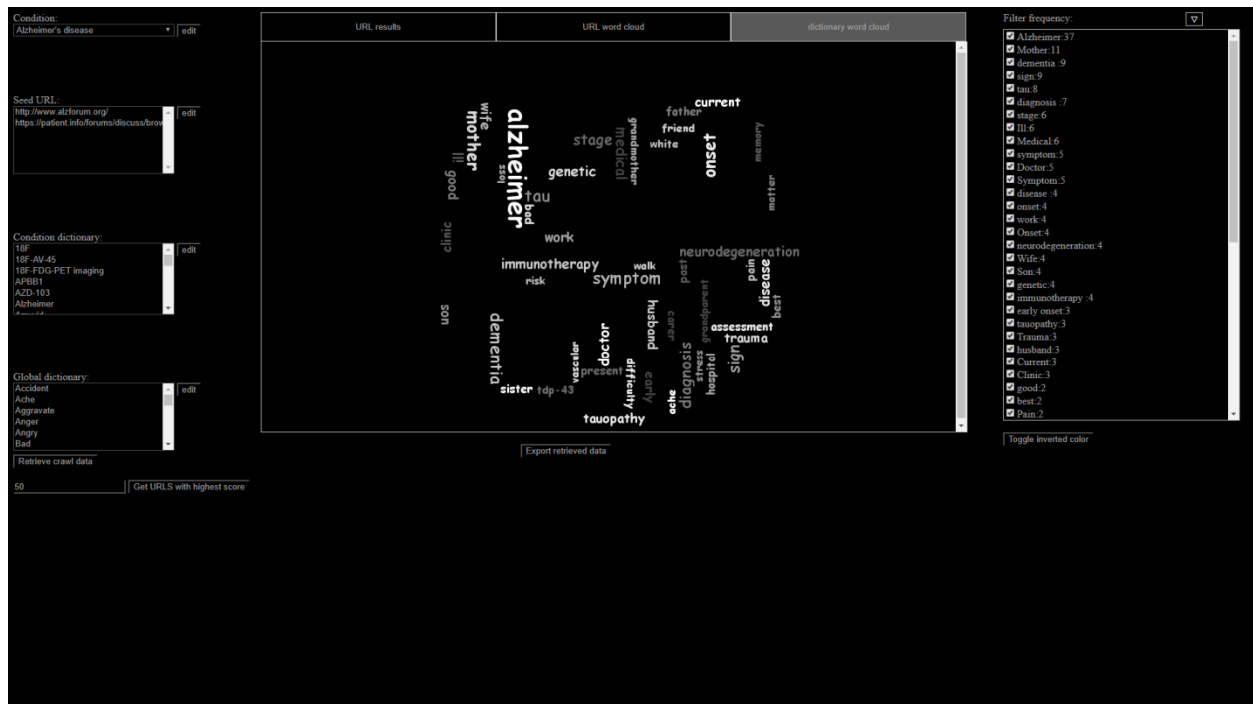
Dictionary Word Cloud Tab: at the top right of the Big Data Box. Click this tab to see the relative frequencies of global and condition specific dictionary terms among the selected URLs. The size of each word in the cloud shows its frequency relative to the other words in the cloud.

Filter Frequency Box: on the right side of the screen, the Filter Frequency Box shows all global and condition specific terms that appear on the crawled URLs. Next to each term is its frequency on the URLs which have been crawled. The dictionary terms that show up in the Filter Frequency Box also show up in the Dictionary Word Cloud (as above).

Below is a screen shot of the Main Page (or front page) of the Health Co-Inquiry WebCrawler. In *Figure 1*, the central tab at the top of the central box (called the “Big Data Box”) has been clicked to show the “JRL Word Cloud” results. This indicates the relative frequencies of ALL words at the site. As noted above, the size of each word shows its relative frequency (with more frequent words being larger).



Below is a screen shot of the Main Page (or front page) of the health Co-Inquiry WebCrawler. In this view, the right-most tab at the top of the central box (called the “Big Data Box”) has been clicked to show “Dictionary Word Cloud” results. This indicates the relative frequencies of all global and condition specific dictionary words at the site. As mentioned above, the size of each word shows its relative frequency (with more frequent words being larger).



The “Individual URL Page”

From the Main Page, a user can navigate to the “Individual URL Page” to get data for a specific URL. Do this by going to the central Big Data Box on the Main Page. Choose a URL for further analysis on the WebCrawler Main Page. Then, scroll across to the right of that individual URL. Click the “Extra” button for that URL. This will navigate the user to the individual URL page for the chosen URL. There is a screen shot of an individual URL page below.

In the view below, the “URL Text Tab” at the top left of the central box has been clicked. It shows the text on the page at the URL.

Notice that the individual URL for which this page is retrieving and displaying data is indicated at the top of the page. Click that web address to navigate out of the WebCrawler and to the actual individual URL site.

The screenshot displays the 'Individual URL Page' for the URL <http://www.alzforum.org/>. The page is divided into two main sections: a central content area and a right-hand sidebar.

Central Content Area:

- URL:** <http://www.alzforum.org/>
- Navigation Tabs:** url text (selected), URL word cloud, dictionary word cloud, url connections.
- Articles:**
 - Ring Around the Vessel: Enlarged Spaces Signal Vascular Disease**
Gaps surrounding blood vessels in the brain may predict cognitive decline and vascular dementia.
 - Revised Guidelines for Diagnosing Progressive Supranuclear Palsy**
The updated, expanded manual helps [Doctors](#) diagnose this rare [disease](#) and its variants earlier, in hopes that more patients can take advantage of [clinical](#) trials.
 - Large Majority of Football Players in BU Brain Bank Have CTE**
The prevalence of chronic [traumatic](#) encephalopathy supports a link between multiple concussions and this degenerative [disease](#), though the sample is self-selected.
 - sAPP Binds GABA Receptor, and More News on APP**
An astrocytic [Aβ](#) protease, destabilization of γ -secretase by [BACE](#) mutations, and an sAPP receptor all made their debuts at Heidelberg conference.
 - C9ORF72 Throws a Wrench into DNA Repair Machinery**
Run on repeats not only break DNA, they thwart the crucial repair pathways needed to sew the strands back together.
 - TMEM106B and Programulin Duke It Out at the Lysosome**
 - Knockout mice suggest opposing effects on lysosomal enzymes and [neurodegeneration](#) in frontotemporal dementia, implicating an ATPase**
 - Analysis of brain tissue from [Alzheimer's](#) patients offers a glimpse of proteomic changes in the disease**
 - Barring Export of Expanded Repeat RNAs Tames Their Toxicity**
 - Researchers identified nuclear protein SRSF1 as the culprit that allows aberrant C9ORF72 RNA to escape into cytoplasm and give rise to**
- Export retrieved data**

Right-Hand Sidebar:

- Filter frequency:**
 - ☒ tau:5
 - ☒ Alzheimer:7
 - ☒ immunotherapy:4
 - ☒ genetic:3
 - ☒ neurodegeneration:3
 - ☒ neuropathy:3
 - ☒ Clinic:3
 - ☒ diagnosis:2
 - ☒ Present:2
 - ☒ Trauma:2
 - ☒ sign:2
 - ☒ TDP-43:1
 - ☒ disease:1
 - ☒ dementia:1
 - ☒ symptom:1
 - ☒ Sost:1
 - ☒ onset:1
 - ☒ III:1
 - ☒ Medical:1
 - ☒ FAD:1
 - ☒ Doctor:1
 - ☒ Symptom:1
 - ☒ risk:1
 - ☒ Aβ:1
 - ☒ Onset:1
 - ☒ stage:1
- Process Date:** 2017-07-27
- Condition:** [Alzheimer's disease](#)
- Toggle inverted color**

URL: <http://www.alzforum.org/>

Export retrieved data

URL: <http://www.alzforum.org/>

Export retrieved data

Filter frequency:

- ☒ tau: 8
- ☒ Alzheimer: 7
- ☒ immunotherapy: 4
- ☒ genetic: 3
- ☒ neurodegeneration: 3
- ☒ tauopathy: 3
- ☒ Clinic: 2
- ☒ diagnosis: 2
- ☒ Present: 2
- ☒ Trauma: 2
- ☒ sign: 2
- ☒ TDP-43: 1
- ☒ disease: 1
- ☒ dementia: 1
- ☒ symptom: 1
- ☒ Sen: 1
- ☒ onset: 1
- ☒ Ill: 1
- ☒ Medical: 1
- ☒ FAD: 1
- ☒ Doctor: 1
- ☒ Symptom: 1
- ☒ risk: 1
- ☒ Aβ: 1
- ☒ Onset: 1
- ☒ stage: 1

Process Date: 2017-07-27

Condition: Alzheimer's disease

Toggle inverted color



In order to get Filter Frequency Box scores for a URL other than the one that is at the top of the page, the user should sweep over the URL in the URL Nodular Graph and click on it, or take the ID number from the upper right box that is within the central box. Then, add that ID number into the web address in the browser URL address at the top of the screen (after deleting the existing ID number that is up there).

Individual URL Data Box (within the central data box, upper right). Then, backspace to get rid of the ID number for the URL currently being displayed on the Individual URL Page. Type the new ID number (without the letter) into the web address at the top of the browser. Then, hit “Enter”. This should bring up an Individual URL Page for the new URL, although this function currently freezes the program.

Filter Frequency Box: on the right side of the screen, the Filter Frequency Box shows all global and condition specific terms that appear on the URL which is listed atop the Individual URL Page. Next to each term is its frequency on the specified URL page. The displayed data are for all condition specific and global dictionary terms that were selected by the user.

URL Text Tab: (in the central box at the top left). In order to read the text on the page at the URL listed at the top of the Individual URL Page, click “URL text” at the top left of the central box. In order to be displayed, the text at the URL must be more than 40 characters in length.

URL Word Cloud Tab: In order to see the relative frequencies of all words at the URL atop the Individual URL Page, click the URL Word Cloud Tab (center, left atop the central box on the Individual URL Page).. Words of higher relative frequency will appear larger; those of lower relative frequency will appear smaller. This tab yields data about ALL words at the URL page, not just dictionary-specific terms. Note: There is an issue with special characters. For example, sometimes apostrophes show up as a little box with a script C next to it which has 2 horizontal lines through it; likewise, Greek letters are not represented correctly in word clouds.

Note that the WebCrawler sometimes hangs up on the URL Word Cloud function (but not on the Dictionary URL Word Cloud). It is owing to the large numbers of words being counted when all words (and not just dictionary terms) are counted. If the crawler has not produced a URL Word Cloud within 24 hours (via a check of the task page), then it probably will not be able to do so.

Dictionary URL Cloud Tab: To view only the relative frequencies of Crawler Specified Dictionary Terms at the URL listed atop the Individual URL Page, click the Dictionary URL Cloud Tab (center, right atop the central box on the Individual URL Page). Words from the Global and Condition Specific Dictionaries that are present on the URL site will appear. Words of higher relative frequency will appear larger; those of lower relative frequency will appear smaller. There is an issue with special characters. For example, sometimes apostrophes show up as a little box with a script C next to it which has 2 horizontal lines through it; likewise, Greek letters are not represented correctly in word clouds.

Click and unclick words in the Filter Frequency Box in order to put them into the word cloud or take them out.

URL Nodular “Node” Graph/Diagram: To generate a diagram of a URL and all of its links, click this tab. It will produce a picture with “blossoms”. Each blossom represents a URL that is linked to the parent/seed URL. The parent/seed URL is also included in this diagram. Hover over a blossom/node and click on it. This will show you what URL it represents. The WebCrawler node graph function is limited to the first 500 links to a seed URL.

If you would like to save a picture of a word cloud or node graph, right click on the image and select “Save Image”. This will allow you to save the image as a file (like a .jpg).

Note that the oversight institution for the WebCrawler sometimes turns off this function, because it can hang up the crawler. As of this writing, the function is working and is turned on (June-2018).

URL Connections Tab: To generate the URL Nodular Graph, click the URL Connections Tab at the top right of the central box on the Individual URL Page. The URL Nodular Graph shows the seed URL (i.e., one of the URLs listed for a given condition on the WebCrawler Main Page. It also shows URLs that are linked to the seed URL as nodules or “blossoms”. The size of each “blossom” can be linked to word frequency scores or the number of links a URL has. (As above, see Score Button and Link Button).

Score Button: (under the URL text tab, top right, central box). Click the Score Button to make the size of each “blossom” in the URL Nodular Graph represent the dictionary word frequency score for the URL at the center of that blossom/nodule. Sweep over a specific URL to populate the Individual URL Data Box (upper right of central box) with information about the score and links to that URL.

The “Score” for a URL is the word frequency count.

Links Button: (to the right of the Score Button, top right, central box). Click the Links Button to make the size of each “blossom” in the URL Nodular Graph represent the number of links to the URL at the center of that blossom/nodule. Sweep over a specific URL to populate the Individual URL Data Box (upper right of central box) with information about the score and links to that URL.

The “Links Score” for a URL is the number of in- and out-bound links for that URL.

Process Date Drop-Down Menu: To the far right of the Individual URL Page, click a date to see the crawl results for the specified URL on that date.

Export Data Button: (bottom, center of Individual URL Page) functions in the same way as the Export Data Button on the Main Page. The importance of this button on the Individual URL Page is that it will yield an Excel spreadsheet with one line of data: scores for the URL listed atop the Individual URL Page. Thus, if a user wants to see scores for only one URL page and save them, the Export Data Button on the Individual URL Page should be used.

Toggle Inverted Color Button: at bottom right on each page, use this button to change the screen to have a white background.

Conducting Thematic Analyses of URL Narrative/Text

The original Health Co-Inquiry Team at Malone University (2016-2017) includes a mental health practitioner, medical provider, and a research methodologist in psychology/psychobiology, *with each holding a doctorate and with the former two maintaining licensure in their clinical fields (clinical counseling and nursing, respectively)*. Four undergraduates in psychology, one graduate student in clinical counseling, and two graduate students in nursing were added to the team during 2017. Additional graduate students and undergraduates joined the team in 2018.

A linchpin of the Bifurcated Method (described earlier in this document) is the thematic analysis of each health condition that is conducted by the Health Co-Inquiry Team. The WebCrawler produces frequency data for URLs regarding dictionary terms and their occurrences on those pages, but it is the thematic analyses by Health Co-Inquiry Project members that helps clarify how those terms interact at a URL in order to represent themes in the narratives of those who post on the URLs. The process whereby Health Co-Inquiry Team members work to uncover themes is as follows:

1) A random number generator is used to create a list of 10 numbers. That list is used to select just 10 URLs from among all the seed URLs for a specific chronic health condition. For information about how search engines are used to create lists of seed URLs, see the Health Co-Inquiry methods described earlier in this document (i.e., on the page that looks like a single PowerPoint slide or “research poster”).

2) All researchers receive the list of 10 randomly selected URLs and each one (separately) evaluates the URLs for the following:

- a. What appears to be the purpose of this URL page?
- b. For an individual blog/post, what appears to be its purpose(s)?
- c. For the entire URL (if applicable) and/or for a given blog/post (as specified in b., above), what appear(s) to be the overt theme(s)?
- d. For the entire URL (if applicable) and/or for a given blog/post (as specified in b., above), what appear(s) to be the underlying (not necessarily explicitly stated) theme(s)?

3) Each researcher creates a spreadsheet with his/her responses to Item #2 (above). Then, spreadsheets are combined in order to show common and/or disparate points of view among the researchers regarding the 10 URLs. It is the inductive analysis method of Hatch (2002) that is most commonly used in our determination of themes. **The researchers’ frames of analysis are individual postings at a URL. However, they also consider the entire page when remarking in their individual reflection spreadsheets (as when responding to the questions listed above). The WebCrawler processes a URL by counting chunks of text that are more than 40 characters. Thus, tabs, navigation trees, headings, and buttons (like the “reply” button for a blog post) will not be counted among the word frequencies on the page if they are 40 or fewer characters. Thus, this presents a limitation of the Bifurcated Method, as the researchers regard all words on a page when conducting their thematic analyses—even headings, tabs, and action buttons. Given that the researchers are aware of the “40+**

character stipulation” on the crawler, they can pay attention to headings in order to discern whether they are important indicators of themes.

Domains are stakeholder categories, e.g., patients/clients; caregivers, and health and mental health providers. Researchers and agencies/organizations are other categories of stakeholders; however, they rarely post personal narratives online. Instead, their posts are usually research articles or descriptions of their (or other researchers’) work. An additional domain at which we have looked is “apparent purpose of a post” (e.g., seeking information/support, giving information/support). Once we have searched within domains, we work to determine the themes and relationships that exist within and between domains.

Reference

Hatch, J.A. (2002). *Doing qualitative research in education settings*. Albany, NY: State University of New York Press.

Limitations

Current WebCrawler limitations include:

- 1) that a user must be logged in to Malone and be on campus to use it;
- 2) that the WebCrawler cannot crawl URLs that rely on cookies, JavaScript, or that re-direct it beyond one URL;
- 3) that the WebCrawler cannot “detect language” and may sometimes return data from non-English sites;
- 4) that the filter feature is “absolute” and “literal”. Dictionary terms are searched exactly as they are typed (except for hyphens). Thus, “beta amyloid” and “ β amyloid” are counted separately by the WebCrawler filter;
- 5) that the WebCrawler only filters seed URL data in the Big Data Box on the Main Page;
- 6) that the filtered words (Filter Frequency Box) on the Individual URL Page will not change when a user sweeps over a new URL on the page in order to see its score. A user should PAY ATTENTION to what URL is listed at the top of the Individual URL Page. This indicates what URL is being filtered over in the right side. Even if a user sweeps over a URL in the node graph and clicks “score” (Score or Link Button) in order to see the scores for that new URL, the new URL’s word frequencies won’t show up in the Filter Frequency Box. A user would have to cut and paste the ID for the new URL into the browser (as the suffix of the web address) in order to get filter (word frequency) data for that new URL.

Screen Shots in Reverse Colors

Main Page: URL Results Tab is clicked.

The screenshot displays the Health Co-Inquiry Webcrawler interface. The main section is titled "URL results" and contains a table with the following data:

URL	Score	Total number of words	Seed	Number of fails	
http://www.alzforum.org/	54	1139	true	0	Extra
https://patient.info/forums/discuss/browse/alzheimer-s-disease-67	153	3165	true	0	Extra

Below the table is a button labeled "Export retrieved data".

On the left side, there are several input fields and buttons:

- Condition:** A dropdown menu showing "Alzheimer's disease" with an "edit" button.
- Seed URL:** A text area containing "http://www.alzforum.org/" and "https://patient.info/forums/discuss/browse" with an "edit" button.
- Condition dictionary:** A list of terms including "18F", "18F-AV-45", "18F-PDG-PET imaging", "AP1801", "AZD-103", and "Alzheimer" with an "edit" button.
- Global dictionary:** A list of terms including "Accident", "Ache", "Aggravate", "Anger", "Angry", and "Bad" with an "edit" button.
- A "Retrieve crawl data" button.
- A "Get URLs with highest score" button.

On the right side, there is a "Filter frequency" section with a list of terms and their counts, such as "Alzheimer: 37", "Mother: 11", "dementia: 9", "sign: 9", "tau: 8", "diagnosis: 7", "stage: 6", "III: 6", "Medical: 6", "symptom: 5", "Doctor: 5", "Symptom: 5", "disease: 4", "onset: 4", "work: 4", "Onset: 4", "neurodegeneration: 4", "Wife: 4", "Son: 4", "genetic: 4", "immunotherapy: 4", "early onset: 3", "tauopathy: 3", "Trauma: 3", "husband: 3", "Current: 3", "Clinic: 3", "good: 2", "best: 2", and "Pain: 2". There is also a "Toggle inverted color" button.

Main Page: URL Word Cloud Tab is clicked.

The screenshot displays the Health Co-Inquiry Webcrawler interface. The central panel, titled "URL word cloud", shows a word cloud of terms extracted from URLs. The most prominent words are "alzheimer's", "disease", "years", "patient", "stages", "memory", "out", "are", "patients", "vessel", "any", "small", "home", "was", "have", "tau", "diagnosis", "stage", "ill", "medical", "symptom", "doctor", "onset", "work", "onset", "neurodegeneration", "wife", "son", "genetic", "immunotherapy", "early onset", "tauopathy", "trauma", "husband", "current", "clinic", "good", "best", and "pain".

On the left side, there are three dictionaries for filtering: "Condition dictionary" (Alzheimer's disease), "Seed URL" (http://www.alzforum.org/, https://patient.info/forums/discuss/browse), and "Global dictionary" (Accident, Ache, Aggravate, Anger, Anxious, Bad). Below these is a "Retrieve crawl data" button and a "Get URLs with highest score" button with a value of 50.

On the right side, there is a "Filter frequency" list showing the frequency of each word in the cloud. The list includes: Alzheimer:37, Mother:11, dementia:9, sign:9, tau:8, diagnosis:7, stage:6, ill:6, Medical:6, symptom:5, Doctor:5, Symptom:5, disease:4, onset:4, work:4, Onset:4, neurodegeneration:4, Wife:4, Son:4, genetic:4, immunotherapy:4, early onset:3, tauopathy:3, Trauma:3, husband:3, Current:3, Clinic:3, good:2, best:2, Pain:2. A "Toggle inverted color" button is located at the bottom right.

Main Page: Dictionary Word Cloud Tab is clicked.

The screenshot displays the Health Co-Inquiry WebCrawler interface. The main window is divided into three tabs: "URL results", "URL word cloud", and "dictionary word cloud". The "dictionary word cloud" tab is active, showing a word cloud with terms related to Alzheimer's disease. The word cloud includes terms such as "alzheimer", "dementia", "symptom", "onset", "genetic", "stage", "work", "risk", "immunotherapy", "walk", "past", "pain", "best", "assessment", "trauma", "sign", "diagnosis", "hospital", "stress", "early", "diffusely", "tauopathy", "sister", "1dp-43", "vessel", "present", "husband", "cerebr", "grandparent", "friend", "white", "memory", "matter", "current", "father", "wife", "mother", "ill", "poor", "clinic", "son", "work", "genetic", "stage", "diagnosis", "disease", "onset", "work", "onset", "neurodegeneration", "Wife", "Son", "genetic", "immunotherapy", "early onset", "tauopathy", "Trauma", "husband", "Current", "Clinic", "good", "best", "Pain".

On the left side, there are several input fields and buttons:

- Condition:** A dropdown menu showing "Alzheimer's disease" with an "edit" button.
- Seed URL:** A text area containing "http://www.alzforum.org/" and "https://patient.info/forums/discuss/forov" with an "edit" button.
- Condition dictionary:** A dropdown menu showing "18F", "18F-AV-45", "18F-FDG-PET imaging", "APBB1", "A2D-103", and "Alzheimer" with an "edit" button.
- Global dictionary:** A dropdown menu showing "Accident", "Ache", "Aggravate", "Anger", "Angry", and "Bad" with an "edit" button.
- Retrieve crawl data** button.
- 50** (a numeric input field) and **Get URLs with highest score** button.

On the right side, there is a **Filter frequency** section with a list of terms and their frequencies, each with a checkbox:

- ☒ Alzheimer:37
- ☒ Mother:11
- ☒ dementia:9
- ☒ sign:9
- ☒ tau:8
- ☒ diagnosis:7
- ☒ stage:6
- ☒ Ill:6
- ☒ Medical:6
- ☒ symptom:5
- ☒ Doctor:5
- ☒ Symptom:5
- ☒ disease:4
- ☒ onset:4
- ☒ work:4
- ☒ Onset:4
- ☒ neurodegeneration:4
- ☒ Wife:4
- ☒ Son:4
- ☒ genetic:4
- ☒ immunotherapy:4
- ☒ early onset:3
- ☒ tauopathy:3
- ☒ Trauma:3
- ☒ husband:3
- ☒ Current:3
- ☒ Clinic:3
- ☒ good:2
- ☒ best:2
- ☒ Pain:2

At the bottom right, there is a **Toggle inverted color** button.

Individual URL Page: URL Text Tab has been clicked.

URL:<http://www.alzforum.org/>

url text

URL word cloud

dictionary word cloud

url connections

Ring Around the Vessel: Enlarged Spaces **signal** Vascular Disease

Gaps surrounding blood vessels in the brain may predict cognitive decline and vascular dementia.

Revised Guidelines for Diagnosing Progressive Supranuclear Palsy

The updated, expanded manual helps **Doctors** diagnose this rare **tau**-pathy and its variants earlier, in hopes that more patients can take advantage of **Clinical** trials.

Large Majority of Football Players in BU Brain Bank Have CTE

The prevalence of chronic **Trauma**-tic encephalopathy supports a link between multiple concussions and this degenerative **tau**-pathy, though the sample is self-selected.

sAPP Binds GABA Receptor, and More News on APP

An astrocytic **Aβ** protease, destabilization of γ-secretase by **FAD** mutations, and an sAPP receptor all made their debuts at Heidelberg conference.

C9ORF72 Throws a Wrench into DNA Repair Machinery

Run-on repeats not only break DNA, they thwart the crucial repair pathways needed to sew the strands back together.

TMEM106B and Programulin Duke It Out at the Lysosome

Knockout mice suggest opposing effects on lysosomal enzymes and **neurodegeneration** in frontotemporal dementia, implicating an ATPase.

Analysis of brain tissue from **Alzheimer** patients offers a glimpse of proteomic changes in the disease.

Barring Export of Expanded Repeat RNAs Tames Their Toxicity

Researchers identified nuclear protein SRSF1 as the culprit that allows aberrant C9ORF72 RNA to escape into cytoplasm and give rise to

Export retrieved data

Filter frequency:

tau:5

Alzheimer:7

immunotherapy:4

genetic:3

neurodegeneration:3

tauopathy:3

Clinic:3

diagnosis:2

Present:2

Trauma:2

sign:2

TDP-43:1

disease:1

dementia:1

symptom:1

Son:1

inset:1

III:1

Medical:1

FAD:1

Doctor:1

symptom:1

risk:1

Aβ:1

Onset:1

stage:1

Process Date: 2017-07-27

Condition: Alzheimer's disease

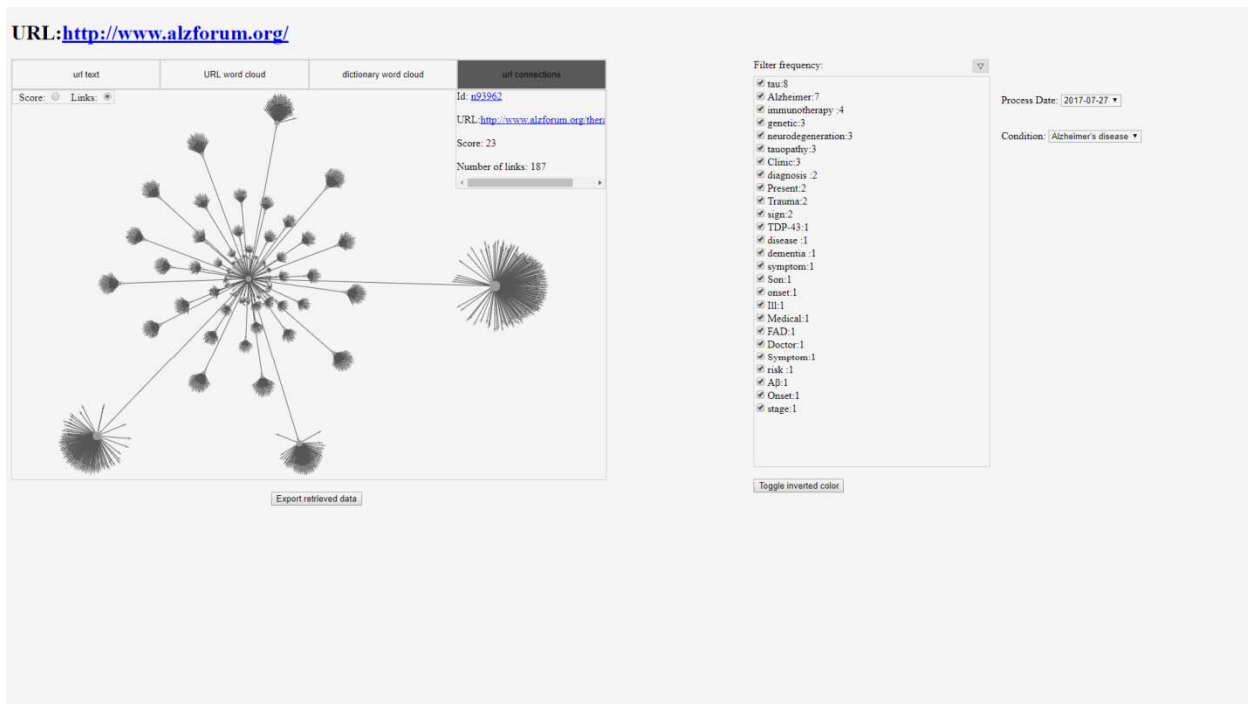
Toggle inverted color

[illegible]

URL: <http://www.alzforum.org/>

The figure displays a word cloud visualization of terms extracted from the URL <http://www.alzforum.org/>. The words are arranged in a circular pattern, with their size indicating frequency or importance. Key terms include "tau", "dementia", "onset", "risk", "clinic", "neurodegeneration", "genetic", "trauma", "sign", "stage", "pathway", "diagnosis", "present", "treatment", "drug", "therapy", "clinical", "research", "progress", "breakthrough", "hope", "future", "promise", "potential", "possibility", "chance", "opportunity", "benefit", "advantage", "merit", "virtue", "quality", "excellence", "superiority", "distinction", "reputation", "prestige", "fame", "glory", "honor", "respect", "admiration", "praise", "applause", "acclaim", "recognition", "acknowledgment", "appreciation", "gratitude", "thanks", "credit", "pride", "satisfaction", "contentment", "well-being", "happiness", "joy", "pleasure", "enjoyment", "fun", "amusement", "entertainment", "recreation", "leisure", "pastime", "hobby", "interest", "passion", "dedication", "commitment", "devotion", "loyalty", "faithfulness", "trustworthiness", "integrity", "honesty", "transparency", "openness", "vulnerability", "emotional", "intellectual", "physical", "spiritual", "psychological", "social", "cultural", "historical", "geographical", "temporal", "spatial", "quantitative", "qualitative", "numerical", "verbal", "written", "spoken", "visual", "auditory", "olfactory", "gustatory", "tactile", "kinesthetic", "proprioceptive", "vestibular", "auditory", "visual", "somatosensory", "motor", "cognitive", "perceptual", "memory", "attention", "executive function", "decision making", "problem solving", "critical thinking", "creativity", "innovation", "invention", "discovery", "exploration", "experimentation", "analysis", "synthesis", "evaluation", "comparison", "contrast", "classification", "organization", "management", "administration", "coordination", "collaboration", "communication", "interpersonal skills", "teamwork", "leadership".

Export retrieved data

Individual URL Page: URL Connections Tab has been clicked.

Tasks and Job Completion Times

There are a number of tasks that the WebCrawler can complete. They include, retrieving crawl data for selected URLs (see Main Page information, above), providing individual URL data (see Individual URL Page information, above), and adding URLs and adding dictionary terms. None of these takes place instantaneously. They are all tasks that modify the data base. As a result, they consume time.

When asking the WebCrawler to retrieve crawl data, to retrieve individual URL data, to add a URL to the Main Page, or to add condition specific or global dictionary terms, the system might require 30 minutes or more to make changes to the WebCrawler.

Time on task for the system might be 30 minutes or more, depending on the extent to which the request modifies the WebCrawler. For individual URL searches, if the system seems to get “hung up”, try to refresh the browser, and make your request again, if it seems that the WebCrawler did not complete the task. When adding URLs or dictionary terms, please, keep in mind that it might take 30 minutes per term or URL. Adding URLs is a particularly cumbersome process for the system, because it must go through and re-calculate the relative positions of all URLs relative to the one that is being added.

ADVISORY: Before adding Seed URLs for a selected condition or topic, add all dictionary terms (global and condition specific). The system requires much less time to add dictionary terms than to add Seed URLs. This is because adding a Seed URL required the WebCrawler to go to that URL and to search it for links, calculate word frequencies of dictionary terms, and yield search output.

Note that the oversight institution for the WebCrawler sometimes turns off this function, because it can hang up the crawler. As of this writing, the function is working and is turned on (June-2018).