

# Ontology-Based Interactive Visualization of Patient-Generated Research Questions

David Borland<sup>1</sup> Laura Christopherson<sup>1</sup> Charles Schmitt<sup>2</sup>

<sup>1</sup>RENCI, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States

<sup>2</sup>National Institute of Environmental Health Sciences, Durham, North Carolina, United States

**Address for correspondence** David Borland, PhD, RENCi, University of North Carolina at Chapel Hill, 100 Europa Drive, Suite 540, Chapel Hill, NC 27517 (e-mail: borland@renci.org).

Appl Clin Inform 2019;10:377–386.

## Abstract

**Background** Crohn's disease and colitis are chronic conditions that affect every facet of patients' lives (e.g., social interaction, family, work, diet, and sleep). Thus, treatment consists largely of disease management. The University of North Carolina at Chapel Hill chapter of the Crohn's and Colitis Foundation—IBD Partners—has created an interactive website that, in addition to providing helpful information and disease management tools, provides a discussion forum for patients to talk about their experiences and suggest new lines of research into Crohn's disease and colitis.

**Objectives** The primary objective of this work is to enable researchers to more effectively browse the forum content. Researchers wish to identify important/popular patient-suggested research topics, appreciate the full breadth of the research topics, and see connections between them, in order to more effectively prioritize research agendas.

**Methods** To help structure the forum content we have developed an ontology describing the major themes in the discussion forum. We have also created a prototype interactive visualization tool that leverages the ontology to help researchers identify common themes and related patient-generated research topics via linked views of (1) the ontology, (2) a research topic overview clustered by relevant ontology terms, and (3) a detailed view of the discussion forum content.

**Results** We discuss visualizations and interactions enabled by the visualization tool, provide an example scenario using the tool, and discuss limitations and future work based on feedback from potential users.

**Conclusion** The integration of a user-community specific ontology with an interactive visualization tool is a promising approach for enabling researchers to more effectively study user-generated research questions.

## Keywords

- ▶ data visualization
- ▶ ontologies
- ▶ Crohn's disease
- ▶ colitis

## Background and Significance

Crohn's disease is an inflammatory bowel disease (IBD) with symptoms that can include diarrhea, inflammation of both the gut and other parts of the body, fatigue, abdominal pain, and weight loss, among others. Colitis refers to inflammation of the inner lining of the colon, and commonly co-occurs with

Crohn's disease. There is no known cure for either condition, although certain therapies can help treat their symptoms, sometimes bringing about long-term remission. Thus, treatment largely consists of disease management. Given the varied ways in which these conditions can present themselves in different patients, and their chronic nature that affects every facet of patients' lives (e.g., social interaction,

received  
December 12, 2018  
accepted after revision  
April 17, 2019

© 2019 Georg Thieme Verlag KG  
Stuttgart · New York

DOI <https://doi.org/10.1055/s-0039-1688938>.  
ISSN 1869-0327.

family, work, diet, and sleep), researchers in the University of North Carolina at Chapel Hill chapter of the Crohn's and Colitis Foundation—IBD Partners (formerly Crohn's and Colitis Foundation of America [CCFA] Partners)—are interested in engaging patients to aid them in disease management and to collect information useful for researching potential treatments. To this end, they have created an interactive website that provides a discussion forum for patients to talk about their experiences, suggest and discuss new lines of research into their conditions, and vote on promising research topics.<sup>1</sup>

Although such a forum can be invaluable for generating and prioritizing research questions based on patient experiences, it can be time and labor intensive sifting through all of the questions and comments on the discussion forum, trying to effectively interpret such a large volume of text. IBD Partners is interested in developing more efficient approaches for identifying common themes and determining which research questions are most frequently discussed by patients. Interactive visualization offers a potential solution to help clinicians and researchers explore the data and identify the salient questions and needs of the patients.

Interactive visualization has proven to be a useful method for analyzing datasets across a wide range of disciplines, including in the health care domain,<sup>2,3</sup> and holds great promise for advancing the state-of-the-art in health care. Many visualization tools for health care applications operate on a wide variety of structured data.<sup>4-7</sup> Prior work in visualizing structured data from patients with various types of abdominal pain includes that of Rao et al, which involves the extraction of diagnostic paths from electronic health record (EHR) data.<sup>8</sup> While such work can be very effective, much of it is not directly applicable to the visualization of largely unstructured text from an online patient forum. Sorbello et al present the utility of using structured text—MeSH terms—with a visual analytics interface for pharmacovigilance; however, they do not directly address the problem of extracting structured information from free text.<sup>9</sup> Tools such as Jigsaw<sup>10</sup> enable the interactive extraction and visualization of named entities and their co-occurrence from document collections, and the work of Sampathkumar et al shows the utility of an ontology-based approach for visualizing data from online health forums.<sup>11</sup>

Ontologies are controlled vocabularies that represent knowledge about a domain of interest.<sup>12</sup> They offer richer representations than other controlled vocabularies (e.g., taxonomies, thesauri) because they enable relationships beyond hierarchical and synonymous. Ontologies have a long history in medicine and biological research,<sup>13-15</sup> and are used for a variety of purposes, e.g., classifying literature for information retrieval, mapping and integrating diverse data sources, aggregating/clustering information, and natural language processing applications.<sup>16,17</sup> Biomedical ontologies tend to focus on representing encyclopedic knowledge about a given domain. For example, UBERON<sup>1</sup> contains approximately 20,000 concepts ranging from very granular anatomy (e.g., cell membrane) to larger systems (e.g., digestive system).

<sup>1</sup> <http://www.obofoundry.org/ontology/uberon.html>

We have created an ontology to help organize the content of the IBD Partners forum and make it more suitable for computational analysis, and developed an interactive visualization prototype utilizing the ontology to enable the interactive exploration of the patient-generated forum content.

## Objectives

The primary objective of this work is to enable researchers to more effectively browse the forum content. Researchers wish to identify important/popular patient-suggested research topics, appreciate the full breadth of the research topics, and see connections between them, in order to more effectively prioritize research agendas. More broadly, this work aims to help physicians better understand how patients think about their condition.

To support these objectives, we have developed an initial ontology from the forum content. This ontology serves to help organize the concepts discussed in the forum, and can serve as the basis for the development of future analysis tools. We have also developed an interactive visualization prototype that uses the ontology to enable researchers to (1) explore the ontology and identify frequently used concepts from the ontology and links between these concepts, (2) identify similar research topics, as defined by shared ontology terms, and (3) quickly navigate the relevant forum text. Based on feedback from potential users, we identify future work using this prototype as a framework for the development of specialized interfaces for different user populations.

## Methods

### Forum Data

The data snapshot used when creating the CCFA forum ontology consists of 97 *research topics* (i.e., user posts consisting of a proposed research question and a description of the question), and 121 user comments made by fellow patients on proposed questions, for a total of 17,322 words. An example research topic post is the following:

#### Question:

Nicotine has shown to be effective for UC [ulcerative colitis] in some individuals, both prior- and nonsmokers. What is the mechanism? Does nicotine affect the microbiome, the immune system, or both?

#### Description:

Big Pharma will not take on the role of studying nicotine as there is no \$\$\$ in it. Few studies with small sample sizes have been done but more research is needed.

Each research topic also has an anonymized user ID (400 unique users), the number of votes for each topic (1,246 total votes), and one of nine predefined categories (*diet, medications, procedures and testing, environment, alternative therapies, lifestyle, genetics, exercise, and other*) selected by the topic creator.

### Ontology Creation

We initially performed an analysis of the forum data using some basic linguistic processing on the text, such as

calculating word and phrase frequencies. However, the results did not effectively capture the forum conversation. Phrases such as *inflammatory bowel*, *controlled trial*, and *disease activity* appeared frequently, which simply confirmed the obvious: patients were discussing IBD and its research. These frequencies did not capture the nuance of specific lines of research the patients were interested in.

We therefore created an ontology of the forum conversation to provide a structure that more effectively captures the depth and breadth of research topics in which the patients were interested. To create the ontology, we first conducted an in-depth, manual exploration of the forum text. Specifically, we applied content analysis to the forum text, sifting through manifest content (i.e., what is seen directly in the text, such as the occurrence of a particular word) to find latent content (i.e., underlying meaning, connotation, nuance). According to Wildemuth, “An example of latent content is the level of research anxiety present in user narratives about their experiences at the library.”<sup>18</sup> In other words, a user may not directly state, “I am so anxious.” Instead, the anxiety may be implied, e.g., “My heart won’t stop beating so fast” or “I wish I could relax.” Wildemuth notes, “Sometimes there is no existing theory or research on your message populations; you may not know what the important variables are. The only way to discover them is to explore the content.”<sup>18</sup> In other words, it may be impossible to identify themes without first immersing one’s self in the text, allowing the themes to be revealed as one becomes more intimate with the conversation. This is reflected in the fact that the most common predefined category assigned by users to their proposed research topic was *other* (34 out of 97, over 40%), implying that the categories did not fully capture the breadth of their interests and discussions.

The manual analysis was performed by a single member of the research team with significant experience in content analysis. Spreadsheet software with manual entry was used to keep track of the analysis. After completing the content analysis, it was clear that no existing ontology would adequately represent the patient conversations. Most biomedical ontologies provide encyclopedic objective knowledge about a particular subject, whereas the CCFA forum text describes personal patient experiences, emotions, and desires. The goal of CCFA physicians and researchers is to understand their patients’ needs and wants, and an effective ontology needs to reflect this goal to help bridge the gap between how clinical practitioners, researchers, and patients view their conditions.

Our ontology structured and classified the raw information in the forum. Concepts (e.g., *medication*, *surgery*, *diet*, and *symptom*) discussed in the forum became “classes” in the ontology. Although relationships beyond hierarchical (e.g., medication *treats* symptom) are possible, our ontology does not currently include such relationships, and thus functions primarily as a taxonomy—a hierarchical grouping of terms. As the ontology is expanded, other types of ontological relationships (such as medication *treats* symptom) will be added. Additional concepts (primarily from the CCFA website to align with their approach to care) were included in anticipation of future forum conversations. Where applic-

able, classes from two pre-existing ontologies, the Ontology for Adverse Events (OAE)<sup>2</sup> and the Disease Ontology (DO)<sup>3</sup>, were used. In total, 165 classes from the OAE and 36 from the DO were included. During the ontology creation, we consulted with the IBD Partners team to ensure that the ontology structure seemed appropriate. The resulting ontology describes a hierarchy of 337 total classes, with seven top-level classes: *comorbidity*, *diagnosis/monitoring method*, *IBD course*, *quality of life*, *risk factor*, *symptom*, and *treatment method*, and a maximum depth of 6. The ontology was created using Protégé,<sup>19</sup> exported in OWL format<sup>4</sup>, and converted to an OBO Graph<sup>5</sup> using ROBOT<sup>6</sup> for easy ingestion into our visualization tool. Based on the content analysis, each research topic was labeled with one or more terms from the ontology. Although the initial content analysis was conducted based on the research topic question, description, and comments, only the question and description were used for labeling. A chart showing the frequency of each top-level ontology term when labeling the research topics, along with immediate children for top-level terms that have a child with frequency greater than 1, is shown in **Fig. 1**. The ontology structure and linkage to the research topics enable the interactive visualization described in the next section.

### CCFA Explorer

The CCFA Explorer is a browser-based tool developed using the D3 visualization library<sup>20</sup> that consists of three different interactive visualizations: (1) the CCFA forum ontology, (2) an overview of the patient-generated research topics, and (3) a detailed view of the forum text and other information about each research topic (**Fig. 2**). The ontology visualization enables researchers to understand the structure of the ontology, see which areas of the ontology were more frequently discussed by the forum users, and see how frequently different ontology terms were discussed together in the same research topic. The research topic overview enables researchers to quickly identify clusters of research topics that discuss similar ontology terms, and the detailed view enables the researcher to read the forum text in-depth. In order to understand relationships between ontology terms and research topics, users can select visual elements representing ontology terms or research topics in each view. All three views are linked to automatically highlight relationships from the various visual elements in each view to the selected items. These linked views enable the researcher to, for example, quickly examine the forum text associated with an ontology term of interest, or determine which ontology terms are related to a cluster of research topics. To develop effective interactive visualizations, Shneiderman’s visual information seeking mantra—*overview first, zoom and filter, then details on demand*—has been adopted by a wide range of data visualization tools.<sup>21</sup> We adopt this approach, providing overviews of the CCFA ontology and

<sup>2</sup> <http://www.oae-ontology.org/>

<sup>3</sup> <http://disease-ontology.org/>

<sup>4</sup> <https://www.w3.org/OWL/>

<sup>5</sup> <https://github.com/geneontology/obographs>

<sup>6</sup> <http://robot.obolibrary.org/>

REPRESENTATIVE CLASSES	POSTS
Comorbidity	11
Diagnosis/monitoring method	7
IBD course	39
Pre-diagnosis time period	1
Diagnosis event	5
Post-diagnosis time period	31
Quality of life	8
Risk factor	58
Demographic factor	7
Environmental factor	18
Lifestyle factor	20
Physiological factor	28
Psychological factor	5
Symptom	36
Gastrointestinal manifestation	12
Extra-gastrointestinal manifestation	3
Treatment method	50
Alternative therapy	7
Holistic treatment	12
Medication	28
Surgery	13

**Fig. 1** Ontology class frequencies as used to label patient-generated research topics. Top-level classes, and immediate children for classes with a child with a frequency of at least 2, are shown.

forum content, along with the ability to filter and obtain detailed forum content based on patterns and relationships discovered from interacting with the overviews.

### Ontology Visualization

Hierarchies are a specific form of ontology, in which each node may have at most one parent, and multiple children. Visualization techniques for hierarchies include tree maps,<sup>22</sup> icicle plots,<sup>23</sup> and tree diagrams (e.g., tidy trees<sup>24</sup>). Although such visualization techniques are effective for showing hierarchical structure, they are not designed to show other types of ontological relationships. Network diagrams offer the

ability to encode different types of relationships via different styles of links in the diagram. Due to this flexibility we adopted this approach, although the current version of the ontology contains only hierarchical “is a” relationships. Kamdar et al present research analyzing user interactions with biomedical ontologies for different visualization types, including network diagrams, and show that different users interact with ontologies differently.<sup>25</sup> Such research suggests that a suite of ontological visualization approaches may be useful, especially when dealing with different user populations, which will help inform our future work.

The CCFA Explorer force-directed network shows the ontology structure and indicates the most prominent ontology terms (→Fig. 2, left). Each ontology term is represented by a node in the visualization, and links (i.e., arrows connecting nodes) indicate “is a” hierarchical relationships (e.g., *medicine* is a *treatment method*). Node radius is proportional to the number of research topics labeled with that ontology term. For any given ontology term, if a research topic has been labeled with that term, the research topic is labeled with all ancestors of that ontology term. Thus no child node will ever be larger than its parent. When the visualization initially loads, node labels for top-level terms in the ontology are visible. Labels for other nodes appear when the user hovers over a node, or upon user selection as described in the Interactive Selection and Highlighting section.

Many researchers may already have an idea of what ontology terms they are interested in. To facilitate rapid identification of predetermined areas of interest, the ontology visualization includes a search box. The user can begin typing into the search box, which shows suggestions for all matching ontology terms. Node labels for all matching ontology terms will be shown and highlighted in red, enabling the user to investigate nodes of interest.

### Topic Overview

The topic overview uses t-SNE,<sup>26</sup> via the t-SNE.js library<sup>7</sup>, to lay out circular glyphs representing each research topic (→Fig. 2, middle). t-SNE is a technique for dimensionality reduction that can be used to lay out objects (e.g., research topics) in two dimensions based on their similarity across a large number of dimensions (e.g., labeled ontology terms). We use t-SNE to place research topics labeled with similar sets of ontology terms closer together, which enables the user to visibly identify clusters of research topics labeled with similar sets of ontology terms. The radius of each glyph is proportional to the number of forum comments made in response to that research topic, and the outline thickness is proportional to the number of user votes for that topic, enabling the user to identify popular topics. The glyph color represents which of the nine predefined categories was chosen by the research topic creator.

We introduce three modifications to the standard t-SNE layout to enable more effective visualization of the CCFA forum data. (1) Because two or more research topics may be labeled with similar sets of ontology terms, glyphs may overlap and occlude each other. Such overplotting can make it difficult to

<sup>7</sup> <https://github.com/scienceai/tsne-js>

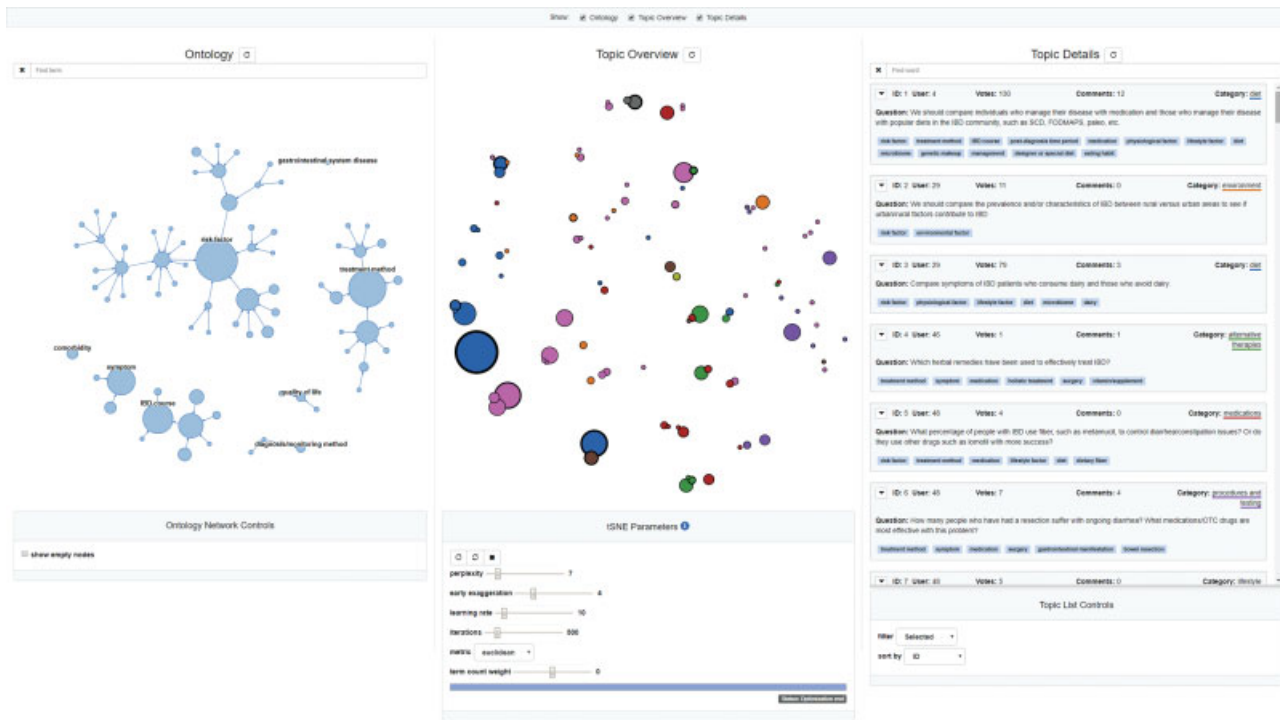


Fig. 2 The CCFA Explorer interface: ontology (left), topic overview (middle), and topic details (right).

see cluster sizes for very similar research topics, and to see and select individual topics. We therefore apply a force-directed layout for overlapping glyphs that separates the centers of each glyph while maintaining some overlap to indicate closely related clusters (→ Fig. 3A). (2) Due to the hierarchical nature of the ontology, we enable weighting of higher-level (closer to the root) or lower-level (closer to the leaves) ontology terms to determine at which level in the hierarchy research topic glyphs are clustered. Weighting higher-level ontology terms results in fewer clusters based on more general terms (→ Fig. 3B), and weighting lower-level terms results in more clusters based on more specific terms (→ Fig. 3C). (3) Greater weights can be applied to the currently selected ontology terms, resulting in clusters reflecting combinations of the selected terms. For example, → Fig. 3D shows a layout emphasizing two selected ontology terms, with three clusters indicating the presence of only the first term, only the second term, or both terms. This

feature enables, for example, easy selection of all research topics with a given set of ontology terms.

**Topic Details**

The topic details view is a scrollable list of panels for each research topic in the forum. Each research topic panel contains the research question, description, and comments for that topic, along with additional information such as the number of user votes, color-coded user-selected category, and tags indicating the ontology terms labeling that topic (→ Fig. 4). Users may select three different levels of details to display each research topic’s text: (1) question only, (2) question and description, and (3) question, description, and comments. The list of research topics can be sorted by *topic ID*, *user ID*, *number of votes*, *number of comments*, and *category*. The list can also be filtered based on currently selected research topics or ontology terms, as described in

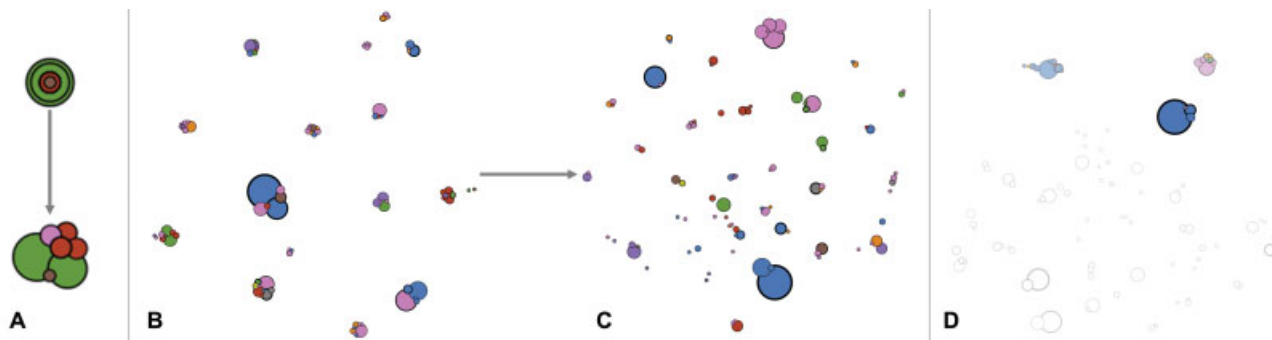


Fig. 3 Modifications to the standard t-SNE layout: (A) force-directed layout of overlapping glyphs to increase cluster legibility, (B and C) differential weighting of ontology terms emphasizing (B) higher-level terms resulting in fewer, more general clusters and (C) lower-level terms resulting in a larger number of more specific clusters, and (D) emphasizing the currently selected ontology terms for clustering.

**ID:** 83 **User:** 2560 **Votes:** 3 **Comments:** 3 **Category:** [procedures and testing](#)

**Question:** Creation of a test to verify or deny IBD in persons struggling with symptoms

**Description:** It took 16 years to verify my Cohn's, from age 13-29 years. Had all sorts of tests and was treated like I was crazy. Spent a month in treatment due to anorexia because eating was so painful. Diagnosed by having exploratory surgery. Because of that I had malnutrition, lost all my teeth and suffered way too long.

**Comments**

- User: 2777** There is a blood test for this.
- User: 2777** There is a blood test for this. It was only developed a few years ago.
- User: 727** Do you know the name of the test?

[treatment method](#) [IBD course](#) [symptom](#) [surgery](#) [diagnosis/monitoring method](#) [diagnosis event](#) [pre-diagnosis time period](#)

**Fig. 4** An example research topic in the topic details view.

the Interactive Selection and Highlighting section. In addition, the user can search for text in the search box, with the matching text highlighted in red in each research topic panel.

### Interactive Selection and Highlighting

The user can interactively select visual elements representing ontology terms or research topics in any of the three views, and all views will be automatically updated to highlight relationships to the selected items. These linked views enable the researcher to perform actions such as finding all research topics labeled with a selected set of ontology terms, or determining which ontology terms a selected cluster of research topics share in common.

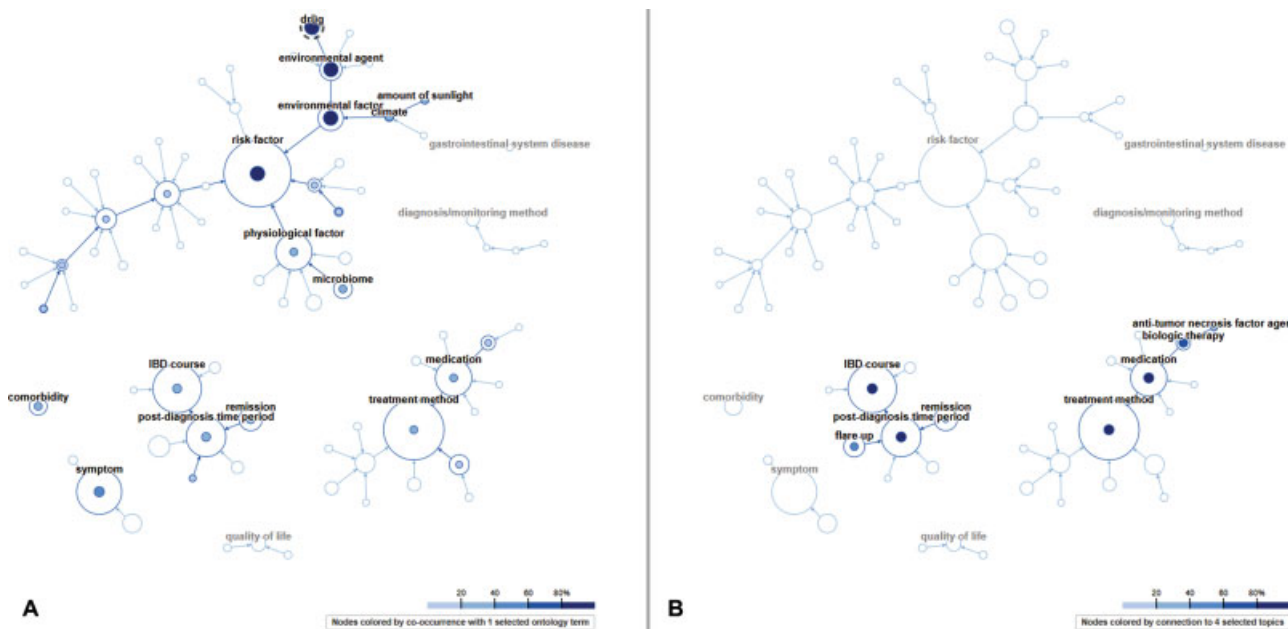
We define three types of possible relationships between ontology terms and research topics: (1) the *co-occurrence* between two ontology terms is the number of research topics that have been labeled with both terms, and therefore is an indication of which ontology terms are discussed together by the forum users. For multiple selected ontology terms, the co-occurrence between a term and the selection is the size of the union of the common research topics. (2) The *association* between two research topics is the number of ontology terms that the two topics share in common, and is an indication of how closely related the two topics are. For multiple selected research topics, the association between a research topic and the selection is the size of the union of the ontology terms they have in common. (3) The *connection* between an ontology term and a research topic is 1 if the topic is labeled with that term, and 0 otherwise. For multiple selected ontology terms or research topics, the connection is the sum of each individual connection.

In the ontology visualization, ontology terms can be selected by clicking on the node for that term. In the topic overview, research topics can be selected by clicking on the glyph for that topic. In the topic details view, research topics can be selected by clicking on the panel for that topic, and ontology terms can be selected by clicking on the tag for that term in any given topic. In all views, selected visual elements are represented by dashed outlines for consistent represen-

tation of selections. Selection in any view results in highlighting in all three views.

In the ontology visualization, the co-occurrence with any currently selected ontology terms is represented by an inset circle for each node, with size proportional to the co-occurrence and color proportional to the percent co-occurrence (co-occurrence divided by total number of research topics connected to the selected ontology terms  $\times 100$ ) with the selected ontology terms (**Fig. 5A**). Similarly, the association with any currently selected research topics is represented by an inset circle with radius proportional to the association, and color proportional to the percent association (association divided by total number of selected research topics  $\times 100$ ) with the selected research topics (**Fig. 5B**). In both cases, labels are displayed for any nodes with a percent co-occurrence/association of at least 25%. In the case of selected ontology terms and selected research topics, highlighting research topic connections takes precedence in the ontology visualization. Whenever there is a current selection being used for highlighting, a label is shown in the visualization indicating what is currently being highlighted, e.g., “nodes colored by co-occurrence with two selected ontology terms” or “nodes colored by connection to three selected topics.” Automatic highlighting of the ontology visualization enables the user to quickly find ontology terms that are discussed in the same research topics, and which ontology terms are related to the selected group of topics.

In the topic overview, the connection with any currently selected ontology terms is mapped to glyph color saturation, normalized by the total number of selected ontology terms (**Fig. 3D**). Similarly, the association with any currently selected research topics is also mapped to glyph color saturation, normalized by the by total number of ontology terms for that glyph's research topic (such that any selected topic will be fully saturated). In the case of selected ontology terms and selected research topics, highlighting ontology term connections takes precedence in the topic overview. Whenever there is a current selection being used for highlighting, a label is



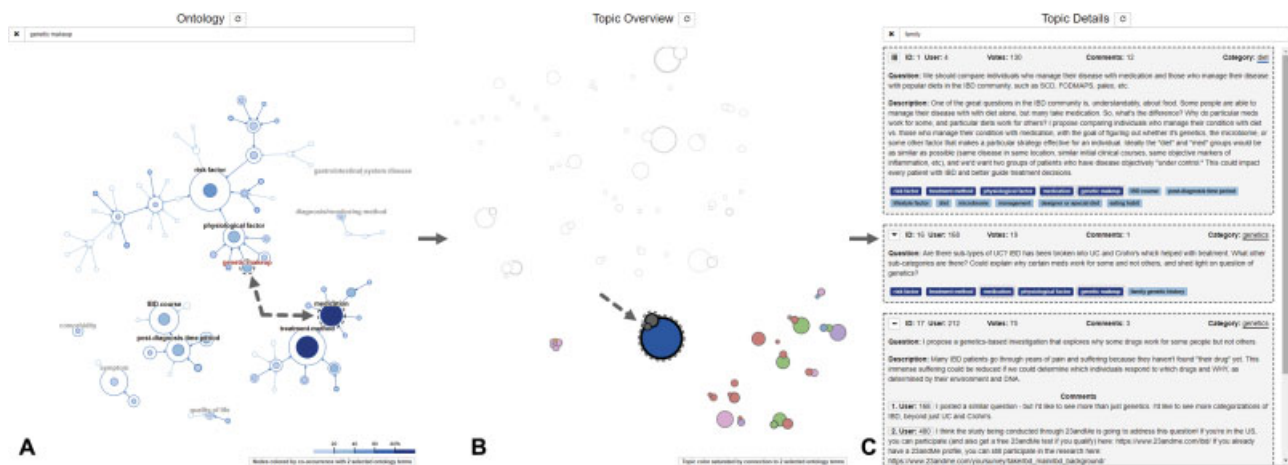
**Fig. 5** Interactive highlighting of the ontology visualization, enabling (A) highlighting of co-occurrences with a selected ontology term (*drug*), and (B) highlighting of connections to research topics selected in one of the other views.

shown in the topic overview indicating what is currently being highlighted, e.g., “topic color saturated by association with four selected topics” or “topic color saturated by connection to one selected ontology term.” Automatic highlighting of the topic overview enables the user to quickly find research topics related to ontology terms of interest, and discover research topics with ontology terms in common.

In the topic details view, research topics can be optionally filtered by *selected* or *connected*. For *selected*, if any research topics are selected, only those topics will be shown. For *connected*, if there are any selected ontology terms or research topics, only topics with a nonzero connection or association will be shown. In this manner, the user can quickly drill down to see the forum text related to ontology terms or research topics of interest. In addition, the same color map applied to the ontology nodes during highlighting is applied to the ontology term tags for each research topic (–Fig. 6C).

**Results**

–Fig. 6 illustrates how a researcher might explore the forum data. The researcher is interested in how the forum users discuss genetics, and whether there are any other concepts or themes that are discussed along with genetics. They begin typing “genetics” in the ontology term search box, which shows all ontology terms with matching text as they type, and select the ontology term “genetic makeup,” which is then highlighted in red (–Fig. 6A). The researcher selects the “genetic makeup” node, which highlights the co-occurrences with that ontology term in the other ontology nodes. They notice that “medication” has a relatively high co-occurrence with “genetic makeup,” indicating that forum users often discuss those concepts in the same research topics, and add “medication” to the ontology term selection. The researcher then re-runs the t-SNE in the topic overview to cluster research topics primarily by these two



**Fig. 6** Example use case with selection of ontology terms (A), selection of research topics related to those terms (B), and detailed inspection of selected research topics (C).

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

selected ontology terms (→ Fig. 6B). The researcher notices a cluster of three research topic glyphs, including one very large glyph (indicating a popular research topic with many comments), and so selects those three research topic glyphs for closer inspection in the topic details view, which is filtered to show only the three selected research topics (→ Fig. 6C). The researcher is then able to inspect the full text and comments of these research topics to answer various questions such as, “Are these patient-generated research questions really asking the same thing, or are they distinct,” “Are these questions created by the same user, or different users,” “Are there any shared misconceptions across the proposed research topics that should be addressed,” etc. Various other exploratory work flows are also enabled by the tool, based on the research focus of the user.

## Discussion

After presenting the CCFA Explorer tool to members of the IBD Partners team, we received useful feedback that will help inform our future work. In general, they thought that the tool was a useful way to explore the CCFA forum data, and made it possible to quickly identify major themes and popular research topics; however, they felt that effective use of some of the tool’s features may be too complex for users—both researchers and others—who are unfamiliar with advanced interactive visual interfaces. Two themes in particular that were identified to address this issue were (1) the utility of a simplified patient-facing interface focused on helping forum users find similar patients and more easily identify research topics relevant to them, and (2) a researcher-facing interface focused on helping researchers in specific domains quickly identify information related to their research area and generate summaries of relevant information that can be easily presented to stakeholders. To this end, we intend to refine our tool in various ways. For example, the ontology visualization, while effective at showing the overall structure of the ontology and highlighting relationships with the ontology terms, is not very well suited for navigation to find ontology terms of interest. We therefore plan to redesign our ontology visualization to make navigation easier, while incorporating some of our current work in interactive highlighting. We also plan to explore the use of text summarization techniques to include in a summary panel that will present an infographic-like view of any currently selected ontology terms or research topics.

Nelson et al present a useful rapid-prototyping model for refining user requirements for dashboards in a health care setting that will help inform our work as we adapt the interface for these specific user populations.<sup>27</sup> To aid in usability evaluation, we will also incorporate Dowding and Merrill’s dashboard visualization heuristics, designed for evaluation of information visualizations in a medical setting.<sup>28</sup>

Another important line of future research will involve expanding the ontology to include a wider variety of relationships than the strict hierarchical relationships currently present. In addition, we will explore automatic and semiautomatic methods to analyze the forum text and classify research topics

based on existing ontology terms, or by expanding the current ontology. This will enable more rapid ingestion of additional research topics, as well as labeling the full forum conversation via comments. In addition, it may be fruitful to combine the visualization of unstructured data, as presented here, with structured data from an EHR, such as the work of Rao et al.<sup>8</sup>

Although the current version of the tool was developed with data specific to the structure of the CCFA forum content (e.g., questions, descriptions, and categories), much of the structure should be generalizable across a wide range of online discussion forums (e.g., questions, descriptions, and comments can map to discussion threads, and user IDs are typically associated with discussion thread content). It may therefore be useful to employ an abstracted forum-content structure, enabling the investigation of these techniques across a wider range of discussion forums involving patient-generated content. For example, previous work has applied ontologies to the analysis of self-help forums for chronic kidney disease.<sup>29</sup> Combining such ontology-based text mining approaches with the interactive visualization techniques described in this article could enable more effective exploration, analysis, and dissemination of online forum data across a wide variety of patient populations.

## Conclusion

We have presented an interactive visualization tool that enables users to explore patient-generated research questions from a forum for individuals suffering from Crohn’s disease and colitis. We described the development of an ontology created from the forum text to help structure the forum content, enabling more effective analysis, visualization, and exploration of the data. To our knowledge, this is the first such ontology incorporating concepts of how patients actually talk about their own conditions. Using linked views that automatically highlight relationships between selected ontology terms and research topics, the researcher can gain insights into concepts of importance to the forum participants. Future work will further refine the tool for specific user populations, such as patients, or researchers with different analytical needs.

## Clinical Relevance Statement

The online forum for patients with Crohn’s disease and colitis created by IBD Partners, where patients can discuss their symptoms and propose potential research questions, is an invaluable source of information for researchers with a patient-centered research agenda. Our approach of combining an ontology with interactive visualization enables the investigation of important concepts and related themes in the forum content. Such an approach can serve as a model for future research into patient-generated content.

## Multiple Choice Questions

1. Ontologies are often organized primarily as hierarchies, and therefore hierarchical visualization techniques can be



used to visualize them. Although such visualization techniques are effective for showing the hierarchical structure, they are not designed to easily integrate other types of ontological relationships. Which of the visualization techniques below makes it easier to show relationships beyond hierarchical parent-child relationships?

- a. Icicle plot.
- b. Force-directed network.
- c. Treemap.
- d. Tidy tree.

**Correct Answer:** The correct answer is option b. Icicle plots and treemaps are both space-filling techniques that effectively show hierarchical relationships via spatial layout, but do not directly enable the display of other types of relationships. Tidy trees, and other tree layouts, show relationships between nodes via links, but the layout is optimized to show hierarchical information. Force-directed networks also show relationships between nodes via links, but the layout is more flexible, enabling different types of relationships to more easily be shown via different types of links.

2. Controlled vocabularies are used for a variety of computerized tasks, e.g., search, metadata and description of information artifacts, classification of documents, etc. A vocabulary is considered “controlled” if it is planned, developed, and maintained by humans over the life of the vocabulary. Humans ensure that duplicate terms are not added, misspellings are corrected, and new terms are added as needed. There are a variety of controlled vocabularies, many of which you are familiar with even if you are new to the term “controlled vocabulary.” For example, there are thesauri, glossaries, taxonomies, subject headings, etc. Each of these vocabularies differs in terms of complexity, typical usage, level of detail, etc. What is the one way in which an ontology differs from some of these other controlled vocabularies?

- a. An ontology is limited to hierarchical parent-child relationships.
- b. An ontology is typically unstructured, and difficult for machines to read.
- c. An ontology permits a wider variety of relationships, and the development of customized relationships.
- d. Ontologies are limited to synonymous/anonymous relationships.

**Correct Answer:** The correct answer is option c. Taxonomies, thesauri, ontologies, etc. allow for relationships between terms (e.g., parent/child, whole/part), but unlike its counterparts, an ontology permits wider variation of relationships and the development of customized relationships. Groups of people develop all different kinds of controlled vocabularies (e.g., ontologies, taxonomies, thesauri, etc.), often because the development effort is substantial and more heads help to reduce bias. Controlled vocabularies of all kinds are used for a variety of purposes; ontologies, taxonomies, thesauri are all used for classifying knowledge, metadata

and description, improving search, text mining, machine learning, etc. Ontologies are also frequently used for annotation of knowledge, as in the case of the gene ontology. Ontologies, and some other controlled vocabularies, are machine readable, making it easy to use for computation and programming.

#### Protection of Human and Animal Subjects

This research was conducted with de-identified data from the CCFA Partners (now IBD Partners) Internet Cohort, made available via a Data Use Agreement with the University of North Carolina at Chapel Hill for a research project approved by the CCFA Partners Research Team.

#### Conflict of Interest

None declared.

#### Acknowledgments

We would like to thank IBD Partners for their help in obtaining the data, refining the ontology, and providing feedback on the visualization prototype.

#### References

- 1 Chung AE, Sandler RS, Long MD, et al. Harnessing person-generated health data to accelerate patient-centered outcomes research: the Crohn's and Colitis Foundation of America PCORnet Patient Powered Research Network (CCFA Partners). *J Am Med Inform Assoc* 2016;23(03):485–490
- 2 Gotz D, Borland D. Data-driven healthcare: Challenges and opportunities for interactive visualization. *IEEE Comput Graph Appl* 2016;36(03):90–96
- 3 Shneiderman B, Plaisant C, Hesse BW. Improving healthcare with interactive visualization. *Computer* 2013;46(05):58–66
- 4 Simpao AF, Ahumada LM, Larru Martinez B, et al. Design and Implementation of a visual analytics electronic antibiogram within an electronic health record system at a tertiary pediatric hospital. *Appl Clin Inform* 2018;9(01):37–45
- 5 Radhakrishnan K, Monsen KA, Bae SH, Zhang W; Clinical Relevance for Quality Improvement. Visual analytics for pattern discovery in home care. *Appl Clin Inform* 2016;7(03):711–730
- 6 Wanderer JP, Gruss CL, Ehrenfeld JM. Using visual analytics to determine the utilization of preoperative anesthesia assessments. *Appl Clin Inform* 2015;6(04):629–637
- 7 Karami M, Safdari R. From information management to information visualization: development of radiology dashboards. *Appl Clin Inform* 2016;7(02):308–329
- 8 Rao G, Kirley K, Epner P, et al. Identifying, analyzing, and visualizing diagnostic paths for patients with nonspecific abdominal pain. *Appl Clin Inform* 2018;9(04):905–913
- 9 Sorbello A, Ripple A, Tønning J, et al. Harnessing scientific literature reports for pharmacovigilance. Prototype software analytical tool development and usability testing. *Appl Clin Inform* 2017;8(01):291–305
- 10 Stasko J, Gorg C, Liu Z, Singhal K. Jigsaw: Supporting Investigative Analysis through Interactive Visualization. In: 2007 IEEE Symposium on Visual Analytics Science and Technology; 2007:131–138
- 11 Sampathkumar H, Chen XW, Luo B. Ontology-based visualization of healthcare data mined from online healthcare forums. Paper presented at: 2015 International Conference on Healthcare Informatics; Dallas, TX; IEEE, 2015:325–334

- 12 Harpring P. Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works. Los Angeles, CA: Getty Publications; 2010
- 13 Ashburner M, Ball CA, Blake JA, et al; The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;25(01):25–29
- 14 The Gene Ontology Consortium. Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res* 2017;45(D1):D331–D338
- 15 Köhler S, Vasilevsky NA, Engelstad M, et al. The human phenotype ontology in 2017. *Nucleic Acids Res* 2017;45(D1):D865–D876
- 16 Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. *Brief Bioinform* 2008;9(01):75–90
- 17 Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform* 2008(Suppl 1):67–79
- 18 Wildemuth BM. Applications of Social Research Methods to Questions in Information and Library Science. Westport, CT: Libraries Unlimited; 2009
- 19 Musen MA; Protégé Team. The Protégé Project: a look back and a look forward. *AI Matters* 2015;1(04):4–12
- 20 Bostock M, Ogievetsky V, Heer J. D<sup>3</sup>: data-driven documents. *IEEE Trans Vis Comput Graph* 2011;17(12):2301–2309
- 21 Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. Paper presented at: Proceedings 1996 IEEE Symposium on Visual Languages; 1996:336–343
- 22 Shneiderman B. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans Graph* 1992;11(01):92–99
- 23 Kruskal JB, Landwehr JM. Icicle plots: better displays for hierarchical clustering. *Am Stat* 1983;37(02):162–168
- 24 Reingold EM, Tilford JS. Tidier drawings of trees. *IEEE Trans Softw Eng* 1981;7(02):223–228
- 25 Kamdar MR, Walk S, Tudorache T, Musen MA. Analyzing user interactions with biomedical ontologies: a visual perspective. *J Web Semant* 2018;49:16–30
- 26 van der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* 2008;9:2579–2605
- 27 Nelson SD, Del Fiore G, Hanseler H, Crouch BI, Cummins MR. Software prototyping: a case report of refining user requirements for a health information exchange dashboard. *Appl Clin Inform* 2016;7(01):22–32
- 28 Dowding D, Merrill JA. The development of heuristics for evaluation of dashboard visualizations. *Appl Clin Inform* 2018;9(03):511–518
- 29 Burckhardt P, Padman R. Analyzing self-help forums with ontology-based text mining: an exploration in kidney space. *AMIA Annu Symp Proc* 2015;2015:1821–1830