

A Web-Based Visualization of Large Document Collections in a Collaborative Environment

Brittany Kondo • Christopher Collins • Visualization for Information Analysis Lab • Faculty of Science • UOIT



Research Goals

Create an interactive visualization of large document collections with visual cues representative of user activity to encourage collaboration.

Specific design goals include:

- Selecting an appropriate layout for the document collection that is well organized with an ordering that is easily recognizable
- Incorporating collaborative awareness within the visualization
- Showing correlation between documents
- Creating filter functions to ease the search for specific documents

Inter-Document Similarity

Correlations between documents are shown through calculated similarity scores. The scores were calculated on a pair-wise basis using cosine similarity measures with terms extracted from the text of the documents. In the visualization, these similarity scores are represented by varying shades of blue. Higher scores are darker and lower scores are lighter (Figure 1).

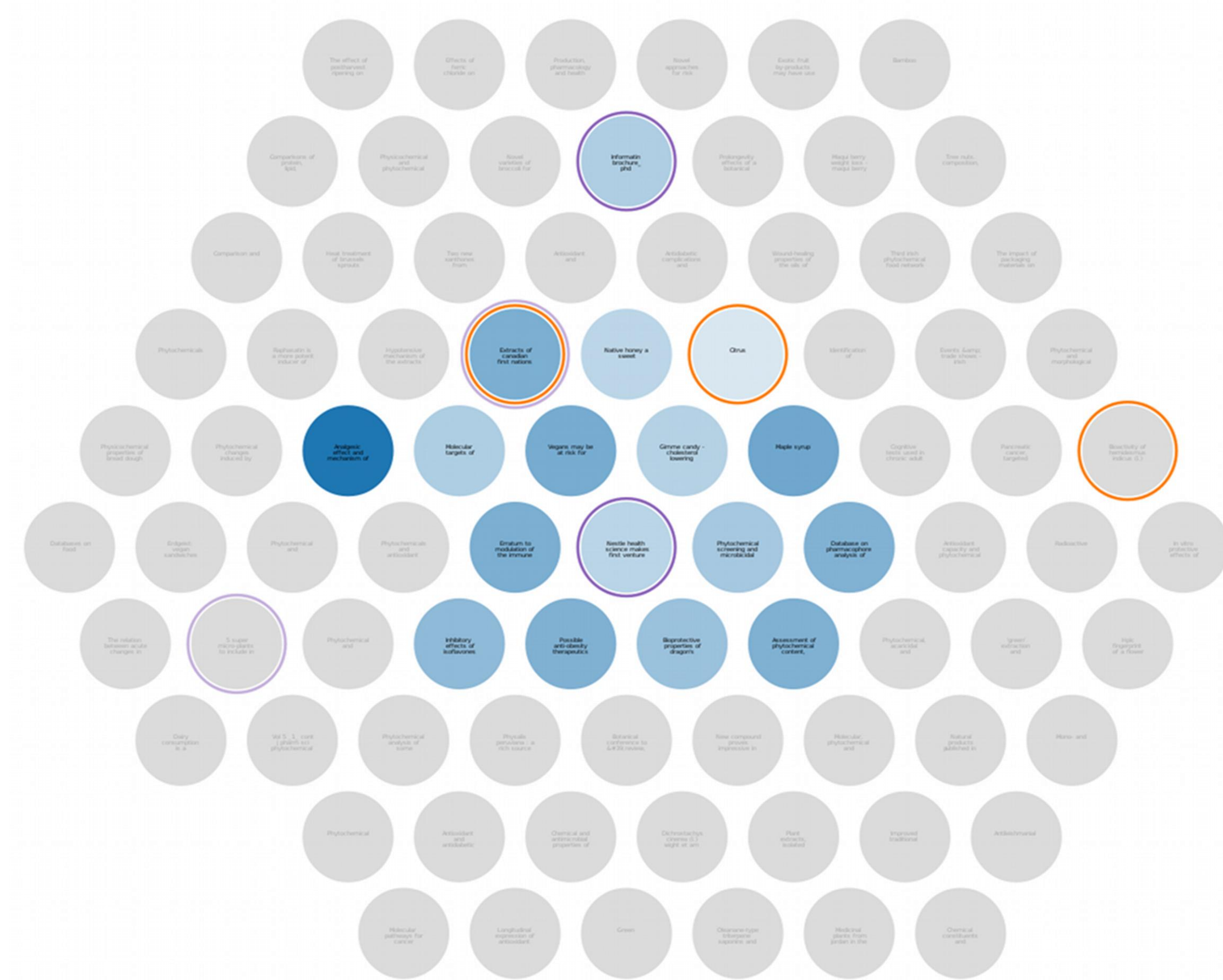


Figure 1: The darkest blue icon is right-clicked.

Collaborative Awareness

Collaboration cues are integrated into the visualization by adding rings around the document icons. These rings are representative of user activity related to the document, such as commenting on a document. Each colour represents a different type of user activity (Figures 2 and 3) and the differences in transparency among these colours demonstrate more or less user activity, respectively.



Figure 2: A scored document

Completed Implementations

The visualization was created using Protovis [1], a visualization toolkit written in Javascript. The circle packing layout available in Protovis was used to arrange the document icons on the screen. The design follows an overview and 'details-on-demand' structure. Various interaction techniques have been implemented to achieve this. Left-clicking a document icon displays the full text of the document. Right-clicking a document icon will re-colour the surrounding icons to show corresponding similarity scores (Figure 1). Hovering the mouse over each document icon as well as the enclosing rings will reveal additional details pertaining to the document and its amount of user activity. The user may also use the mouse wheel to zoom in on an area of interest and click and drag the mouse to pan the screen.

EmailAddress MedicalCondition Country
 ompanyPublishedMedium Organization
 sStation MusicGroup Continent
 ProgrammingLanguageFacility
 umber OperatingSystem
 MedicalTreatment Holiday
 umber MusicAlbum Position
 Movie MarketIndex IndustryTerm
 City TVStation Person
 Product URLNaturalFeature
 Technology

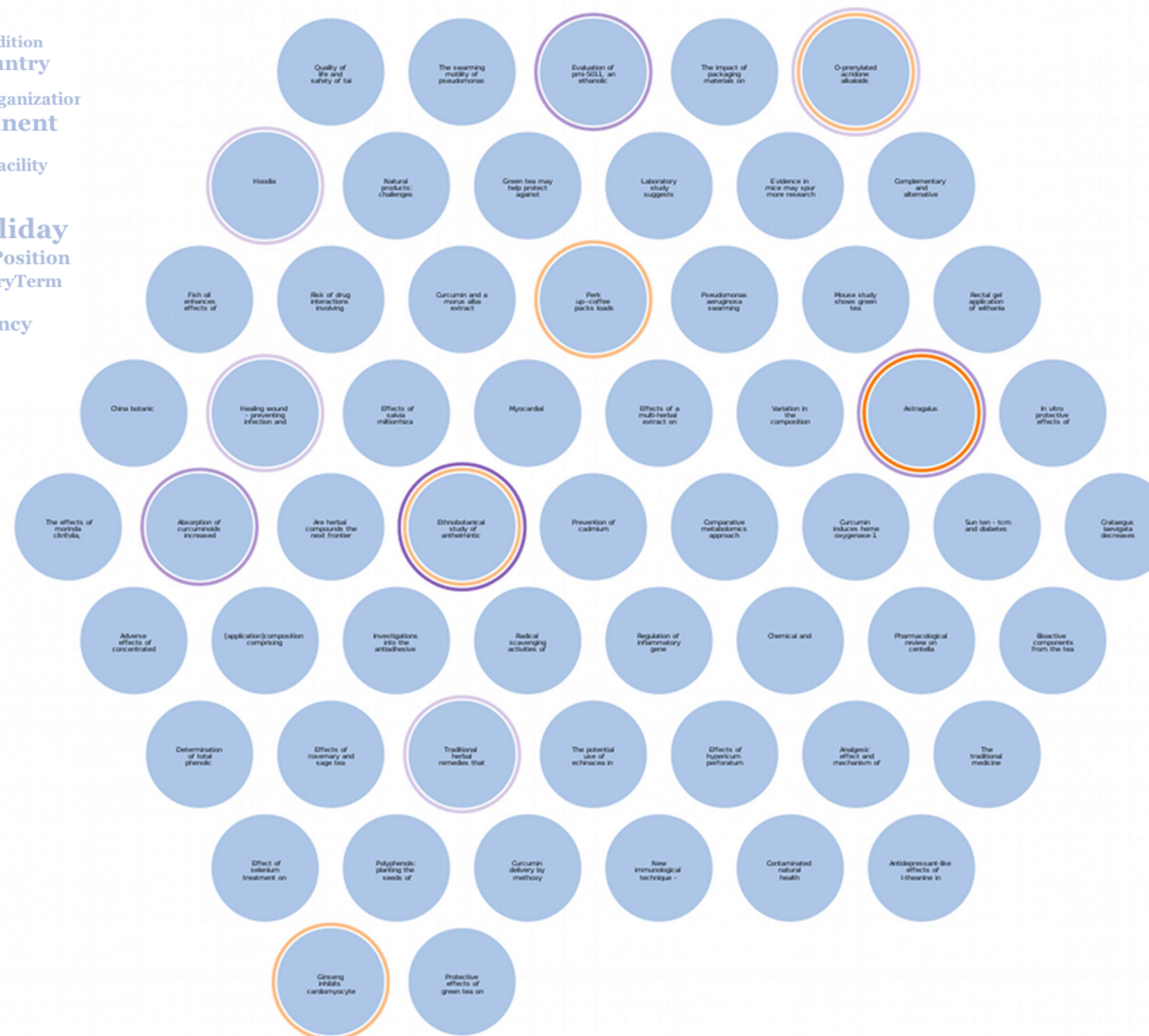


Figure 4: An overview of a document collection

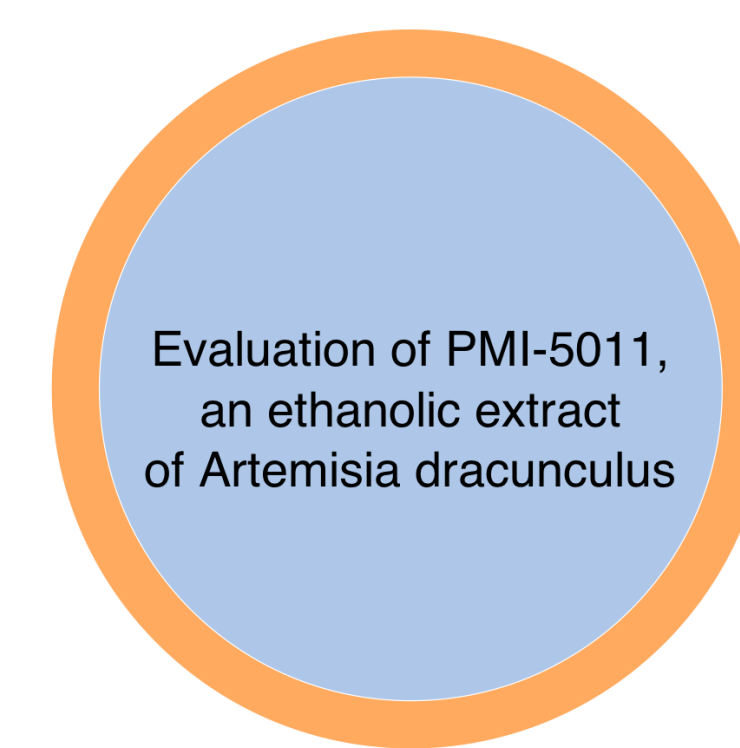
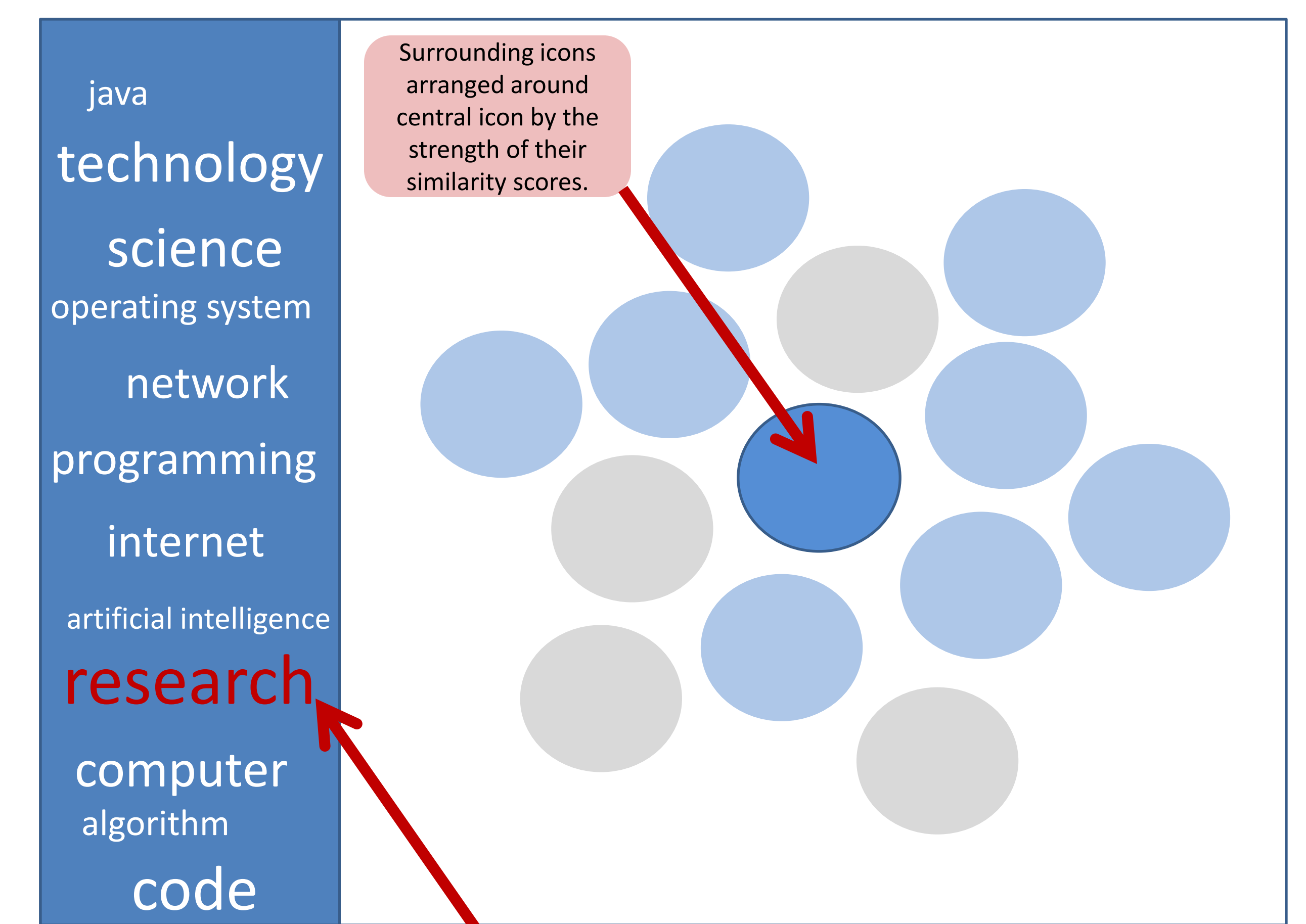


Figure 3: A commented document

Work in Progress

A tag cloud visualization will be added to work synchronously with the currently implemented visualization. A tag cloud is a visualization technique used for representing text, commonly single words. The tag cloud for this interface will contain various key words selected from the documents. These key words may be clicked by the user to filter the document icons on the screen which will aid the search for specific documents. Another design goal that was not achieved is having an effective ordering of document icons. Some potential ideas for ordering documents include document recency, document popularity (most visited) or even the similarity scores.



Tag cloud navigation tool allows the user to click any keyword to filter the documents in the visualization. Here, blue icons contain the word and grey icons do not.

[1] Bostock, Michael and Heer, Jeffrey. "Protovis: A Graphical Toolkit for Visualization," IEEE Transactions on Visualization and Computer Graphics, pp. 1121-1128, November/December, 2009