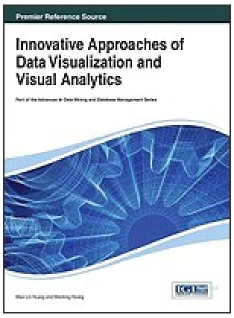

Innovative Approaches of Data Visualization and Visual Analytics

by Mao Lin Huang and Weidong Huang (eds)

Chapter 8: Understanding Collections and Their Implicit Structures Through Information Visualization

J. Alfredo Sánchez,
Universidad de las Américas Puebla
Mexico

ABSTRACT

This chapter discusses how various approaches to information visualization can be used to assist users in understanding large digital collections and discovering relationships among the entities involved explicitly or implicitly in their development including people, organizations, and documents. Our main postulate is that visualization schemes, such as *fisheye views, starfield displays*, or *self-organizing maps*, when integrated and coupled with semantic layouts of topic areas, can significantly facilitate the analysis and discovery of existing and potential relationships among a wide range of entities. A series of developments illustrates how users play a key role in determining advantages and limitations of information visualization schemes, as well as in finding opportunities for improvement and new application areas.

INTRODUCTION

Myriad relationships exist among people, objects, and practically among any entities or concepts. Many new connections are being created every instant and are somehow represented in the digital realm. Thus, for example, contact lists are being extended in social networks, hyperlinks are being created between web pages, and documents are being written that cite various sources. These are examples of explicit relationships that are forged by users or authors. Also, there are a very large number of relationships that may not be evident, and that only exist by virtue of the characteristics of entities, their activities, location or other attributes. Thus, for instance, friends of friends in a social network are indirectly related, users from the same geographical area could be grouped together, papers on the same topic or written by authors from the same institutions or countries can also be considered to be connected in some implicit way.

In order to deal with the volume, complexity and dynamism of this expanding information universe, it has become crucial for people to understand how large collections of digital entities are organized and how their elements are interrelated. Relationships determine structures that can be of interest for various users or perspectives. For example, books grouped according to their publisher may be of interest for booksellers and for librarians, but not necessarily for library patrons, who may be interested in hierarchical, general-to-specific views of the books' subjects. Similarly, roads that connect towns on a map and their travel times may be of interest for tourists, but manufacturers planning product distribution or government officials making budgetary or tax decisions may be more interested in various demographic layers and groupings of the same geographical locations.

Information visualization schemes play a key role in providing graphical representations of large collections and of the relationships among their elements. These visualization schemes, coupled with appropriate control mechanisms for parameters such as scale, attributes on display, or evolution over time, have an enormous potential to become user interfaces that will help users understand not only the attributes of large number of digital objects and their explicit inter-relationships, but also the implicit structures that result from considering multiple perspectives and implicit relationships. In this chapter, we focus on such information visualization schemes. In particular, we discuss the design and applications of user interfaces we have developed for supporting user activities that rely on the analysis and comprehension of very large data sets.

The main emphasis of the chapter is on the potential of three existing techniques for information visualization, namely *starfield displays*, *fisheye views,* and *self-organizing maps*, to help users in detecting and understanding relationships and structures among elements of large collections that are defined implicitly in terms of a number of relevant attributes. Though these techniques were originally devised or applied to the visualization of large collections and explicit relationships, we have worked on adaptations and applications for visualizing implicit relationships and structures.

The chapter is organized as follows: The following section introduces basic concepts and summarizes related work in areas such as applications and advances in starfield visualizations, fisheye views and graphical representation of ontologies, digital repositories, and collaboration networks. The next section discusses our work that focuses on the visualization of digital collections, our experiences with actual users, and the evolution from the use of basic starfields to their enhancement with the introduction of fisheye views. This will provide the basis for the core section on our current work, which integrates starfield visualizations, fisheye views, and lightweight ontologies, as well as its applications to visualizing collaboration networks. Observations from actual use of this integrated visualization scheme are presented and discussed. We then discuss the broader implications of our approach and our findings, and close the chapter by providing conclusions that can be derived from our work.

BACKGROUND

In this section we provide some basic definitions of the concepts involved in the chapter and we also review related work. We first refer to one of the structures we aim to visualize, namely collaboration networks. We also provide some background on ontologies, a notion we have used to classify and organize items to be visualized. Then we describe three major visualization mechanisms we have relied on for supporting collection understanding: starfields, fisheye views and self-organizing maps. Finally, we provide an overview of recent work being conducted in the area of visualization of abstractions and large document collections.

Structures

Collaboration networks are structures that represent interconnected groups of people who work together at various coupling levels. In academic settings, some collaboration networks can easily be determined from the lists of authors of publications held by a digital collection. Other, more implicit structures, on the other hand, can be hidden and can only be suggested as potential networks among authors of documents by inferring relationships from their metadata. Relationships in collaboration networks may vary in strength, as they may result from direct, evident connections (such as co-authorship) or from indirect ties determined, for example, by semantic overlaps in the subjects of their documents in a collection.

In the context of information systems, the literature presents varying definitions of *ontologies*. One that is generally accepted establishes that an ontology is an explicit specification of a conceptualization (Gruber, 1993). This definition has been extended by suggesting that the conceptualization must be shared (Borst, 1997). Lightweight and heavyweight ontologies are distinguished according to the degree of formality involved in their encoding (Lassila & McGuinness, 2001). Both for grouping items and for representing them in visualization interfaces, we have limited our work to the use of lightweight ontologies. These range from enumerations of terms to graphs or taxonomies of concepts with well-defined relationships among them, which provide a representation of an information space. The term lightweight indicates that the construction of ontologies does not involve domain experts. It also refers to tree-like structures where each node label is a language-independent propositional formula (Giunchiglia, Marchese, & Zaihrayeu, 2007).

Base Visualization Mechanisms

Though our initial explorations in the realm of information visualization focused on the advantages of three-dimensional representations (Amavizca, Sánchez, & Abascal, 1999; Proal, Sánchez, & Fernández, 2000), we soon adopted an approach that relies on simplified two-dimensional graphics that are enhanced by direct manipulation components. This combination of interface elements ensures control by the user to delimit information areas, filter needed or unwanted entities based on their attributes, and select specific information units.

One key representation we introduced in our research is that of a *starfield* (Ahlberg & Shneiderman, 1994) A starfield is a grid-based visualization method that highlights the intersections of two axes on a plane. Each axis typically is used to represent an attribute of the objects being displayed. Thus, for example, we can use the horizontal axis to refer to the subjects of publications held by digital collections, use the vertical axis to refer to authors of publications, and highlight their intersection if there is at least one publication on a given subject written by a given author.

As we experimented with starfield visualizations, initially of metadata associated with large physical collections (Silva, Sanchez, Proal, & Rebollar, 2003), we have been able to observe how users benefit from their potential, but also to obtain first-hand evidence of their shortcomings (Sánchez, Twidale, Nichols, & Silva, 2005). Keeping users oriented as they visualize starfield-based representations has been one of the challenges we decided to address in subsequent research. Thus, as we moved from visualizing physical, centralized collections to vast digital, distributed repositories, we also set out to investigate how fisheye views (Furnas, 1986, 1999, 2006) could be coupled with starfields to provide users with a focus-plus-context mechanism to prevent disorientation.

Fisheye views are the basis for a technique that allows users to explore in detail an area of interest while maintaining its context. The name refers to the distortion effect created around non-important elements while those that are important are magnified. Our research shows that coupling starfields with fisheye views represents an important evolution of basic starfields (Sánchez, Quintana, & Razo, 2007). By observing users, we also found that examining sub-areas of a starfield visualization did not always result in meaningful partitions, as most of our data orderings had been chronological or alphabetical. Clearly, more semantical groupings of attributes on both axes of the starfield were needed.

We have found that one technique that naturally exposes implicit relationships is that of Self-Organizing Maps (SOM). A Self-Organizing Map (SOM), or Kohonen map, is a neural network that competes by means of mutual lateral interaction (Kohonen, 1990). A SOM consist of neurons organized in a low-dimensional grid (typically two dimensions). Each neuron is represented by an n-dimensional weight vector (also known as prototype vector, codebook vector). The main difference between a self-organized network and a conventional one is that correct output cannot be defined *a priori*, therefore a SOM utilizes an unsupervised learning algorithm. A SOM consists of two neuron layers: the input layer (input vector) and the output layer (lattice). The algorithm used in a SOM classifies entities in collections (thesis, papers, etc.) by using their attributes or metadata (input vector for each document), and updates a map (lattice of neurons) so that each neuron represents a set of similar documents based on their characteristics. SOM-based visualization has become an important technique for our ongoing work in the area of collection understanding, a concept discussed in the following section.

Related Work

Proposals for visualizing complex networks have taken various forms (Herman, Melancon, & Marshall, 2000), the most popular of which have been based on the notion of nodes that are linked by edges. Node-link visualizations have been used in many different contexts. There are even US patent applications for registering implemented methods for handling node-link representations, such as the system proposed by (Yakowenko & Matange, 2004) for displaying nodes wherein the nodes have a hierarchical context. Positional information associated with multiple nodes is used to generate a display for the nodes in response to a change in focal position. The generated node display maintains hierarchical contextual information about the nodes.

Similarly, visualization schemes are being used to study the structure of social networks. A social network is a collection of people, each of whom is acquainted with some subset of the others (Barabasi, 2010). Research on this particular structure has been supported by graph-oriented techniques (Yang, Asur, Parthasarathy, & Mehta, 2008) and analysis tools (Smith et al., 2009). However, tools that combine exploration, statistics and visualization techniques have not been proposed. SocialAction (Perer & Shneiderman, 2008) is a tool designed for finding relationships among people by using data sets, and for analyzing the structure of the resulting graph. Based on a similar approach, a system called CrimeNet Explorer (Xu & Chen, 2005) supports police agencies in the exploration of criminal networks and in understanding their complex structure by using a graph representation.

A user interface for visualizing collaboration networks must be able to represent entities, relationships, item groupings and their instances, collaboration groups and various properties of these groups. Ontological representations offer helpful means for organizing relationships and suggesting potential collaboration networks. However, visualization of ontologies is not an easy task due mainly to volume, complexity and diversity of representation mechanisms. There have been many approaches for tackling this problem. Katifori, Halatsis, Lepouras, Vassilakis, and Giannopoulou (2007) report a very thorough survey of existing ontology visualization methods.

In order to create a visualization method adapted to the detection and visualization of potential collaboration networks to work on specific areas, we surveyed the area and characterized existing methods. Reported works were selected considering some of the criteria established by the classic framework for categorizing visualization methods (Shneiderman, 1996). We included the task topology (overview, zoom, filter, details-on-demand, viewing relationships, history) and other characteristics such as meta-data, 2D or 3D visualization, grouping (on a criteria or using an ontology) and context. Those tasks were used as evaluation dimensions applied to the context of collaboration networks visualization. Details of our survey can be found in (Ramos, Sánchez, & Hernández-Bolaños, 2010).

We observed that most existing approaches make it possible for the user to obtain a general overview of a data set and to view relationships among items, at least within a close perimeter. Other popular functions include the ability to zoom in and out of selected areas, as well as to obtain details of required items. We found it interesting that 2D visualizations remain popular. Even though some 3D representations have been attempted (as in [Yang, Asur, Parthasarathy, & Mehta, 2008]), the simplicity of 2D, both for the user and the developer, has generally outperformed the complexity that 3D brings about. Only few methods allow users to filter items dynamically or to group them semantically. Even fewer offer functionality either to maintain the context while exploring specific areas or to provide access to previously viewed scenarios.

Maintaining context is particularly important when visualizing collaboration networks, hence our decision to focus on features that make this possible. Also, grouping techniques, overview and zoom functionality, as well as context preservation and displaying relationships between entities were considered key features for a model oriented to the discovery and exploration of collaboration networks.

Information visualization has a great potential for assisting users understand what large collections of objects are about. *Collection understanding* has been suggested as an alternative to Information Retrieval (IR), which focuses on finding specific objects in collections by providing "terms," commonly thought of as values of metadata fields. Collection understanding promotes an exploratory approach by providing users with a general sense of collections (Chang et al., 2004). In order to accomplish this goal, collection understanding relies on visualization and filtering mechanisms. This notion has motivated work that produces collages from collections of images (Cunningham & Bennett, 2008; Chang et al., 2004), or uses tags and filters (Girgensohn, Shipman, Turner, & Wilcox, 2010) or explores the use of map-based visualizations (Buchel, 2011). We currently are working on producing visual representations of large, distributed collections based on hierarchical clusters that are produced by applying the notion of self-organizing maps.

DISCOVERING IMPLICIT STRUCTURES FOR UNDERSTANDING COLLECTIONS

Many digital collections are held in institutional repositories and digital libraries that make their contents available through metadata servers. In order to provide access to those collections, applications are needed that collect those metadata in response to specific queries from users. This is performed by a process referred to as harvesting, which contacts various collections and produces a unified view of metadata. If individual collections can be difficult to comprehend by the user, composite collections that result from metadata harvesting may be even more difficult to handle conceptually by users of all types. Visualization mechanisms can be of significant help for providing perspective and interfaces for exploring large and complex collections. However, making metadata available for generating meaningful and useful visual representations is not trivial, as discussed next.

From Disperse Collections to Visualizing Meaningful Structures

In order for data from multiple collections to be available for visualization, several stages need to be completed. These stages should allow objects in diverse collections to be organized, their hidden relationships to be revealed and their implicit structured to be discovered. We propose four general stages for this process: Harvesting, normalization, inference and visualization, as illustrated in Figure 1:

- **Harvesting:** Standard protocols for harvesting and uniform metadata descriptions are necessary for facilitating the process of gathering data from heterogeneous sources. We have focused most of our efforts to utilizing collections that comply with the Protocol for Metadata Harvesting of the Open Archives Initiative (OAI-PMH version 2.0). OAI-PMH requires that data providers comply at least with the Dublin Core (DC) metadata standard. Prior to looking for implicit relationships in data, we harvest metadata that is exposed by distributed collections and we build an intermediate virtual repository with relevant metadata.
- **Normalization:** Though digital libraries do represent valuable information, data coming from heterogeneous sources needs to undergo significant pre-processing so mining or inference can be performed with some level of effectiveness. Since OAI collections adhere to the Dublin Core (DC) metadata standard, uniformity is expected at least for naming attributes. However, DC is quite flexible and the usage of its elements typically differs from collection to collection. Even within the same collection, uniformity for key attributes such as author or institution names cannot be taken for granted. Thus, a normalization process examines metadata that results from the preceding harvesting stage and produces a version of the intermediate repository that contains uniform metadata.
- **Inference:** In this stage, normalized metadata is processed so as to produce representations that can be visualized meaningfully. For example, lightweight ontologies can be constructed to represent the classification of the topics and subtopics found in the collections, neural networks may be used to generate thematic clusters, and hierarchies can be generated that represent the affiliation of authors to their institutions and the countries to which each institution belongs. More importantly for collaboration networks, a representation can be produced for the potential relationships among authors of publications as determined from the collections harvested to assemble an intermediate repository.
- **Visualization:** Finally, graphical representations of the inferred data are produced and made accessible to users so they can visually

detect potential relationships and explore the suggested collaboration networks, clusters or other groupings.

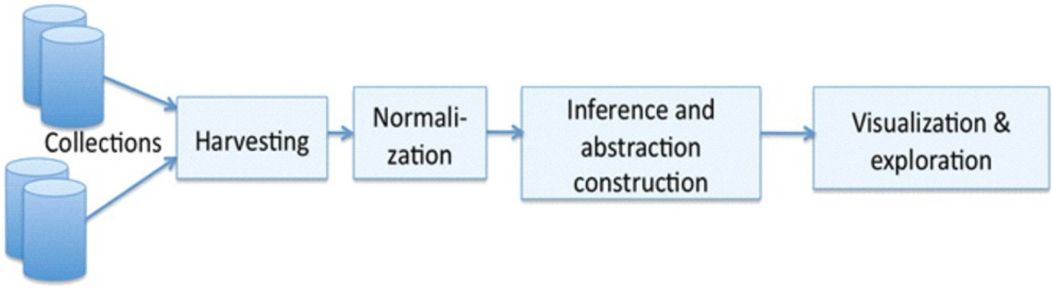

Figure 1: Stages from metadata collecting to visualization

Exposing implicit structures and relationships is the result of both the inference and visualization stages. Much of what can be inferred regarding hidden structures can only be perceived by the user if appropriate visualization mechanisms are available. But user interface controls for adjusting and filtering what is being displayed can significantly enhance the potential of information visualization means for discovering additional relationships.

Harvesting has become a stable process particularly for OAI-compliant collections. Normalization efforts require significant human intervention both for reaching agreements on how metadata standards should be used and for adapting non-compliant legacy metadata to newly defined standards. Though these are important tasks, they are not discussed with more depth, as our emphasis in this chapter in on the use of visualization mechanisms. However, given the close connection between our visualization approaches (discussed in the following section) with inference mechanisms, we do discuss the latter with some additional detail.

Inference and Ontology Construction

We take advantage of our previous work on lightweight ontologies for organizing collections of objects so they fall into one of various levels of topics and subtopics. Particularly for OAI collections, we have adapted the *Frequent Itemset-based Hierarchical Clustering* (FIHC) algorithm (Fung, Wang, & Ester, 2003) as the basis for a semi-automatic method for constructing lightweight ontologies of metadata records. Since we have focused on the classification of metadata records, Dublin Core attributes play an important role in this process. In particular, we have used the title, description and subject for each document in the collections. We have applied our method to specific participating collections and the resulting lightweight ontologies have been used as one of the main inputs for our visualization technique. Details of the construction of lightweight ontologies, which are referred to as "ontologies of records", can be found in (Sánchez, Medina, Starostenko, Benitez, & Domínguez, 2012).

VISUALIZATION APPROACHES FOR COLLECTION UNDERSTANDING

We have progressed through various stages in our quest to face the challenge of facilitating access and user comprehension of large digital collections and their explicit or implicit underlying structures. Though our initial explorations focused on the advantages of three-dimensional representations, particularly for navigating through explicit taxonomic classifications (Amavizca, Sánchez, & Abascal, 1999; Proal, Sánchez, & Fernández, 2000), we soon adopted an approach that relied on simplified two-dimensional graphics enhanced by direct manipulation components. This combination of interface elements ensures control by the user to delimit information areas, to filter needed or unwanted entities based on their attributes, and to select specific information units.

Using Starfields for Discovering Patterns

One of our early efforts took advantage of the concepts and tools of Visual Information Seeking (VIS), in which the user is presented, in a two-dimensional depiction, with multi-dimensional overviews of information spaces. Various filters may be applied to this information space by manipulating graphical sliders to zoom in and out of areas of interest until relevant data elements are found (Ahlberg, Williamson, & Shneiderman, 1992; Ahlberg & Shneiderman, 1994). VIS has been considered a promising approach to address issues in the design of interfaces for large repositories and to deal with information overload. However, it had been demonstrated only with relatively small collections of a few thousand items, which made it possible to keep entire collections in main memory and perform recalculations and rendering in real time as the user manipulated sliders and filters in an interactive fashion. Vast digital libraries, however, comprise collections that may include hundreds of thousands, millions or even more items. We developed EVA2D, a visualization environment that implemented and extended the notion of VIS to facilitate the exploration of large collections comprised by digital libraries. Its main interface components, as applied to a collection of over 192,000 library records, are illustrated in Figure 2.

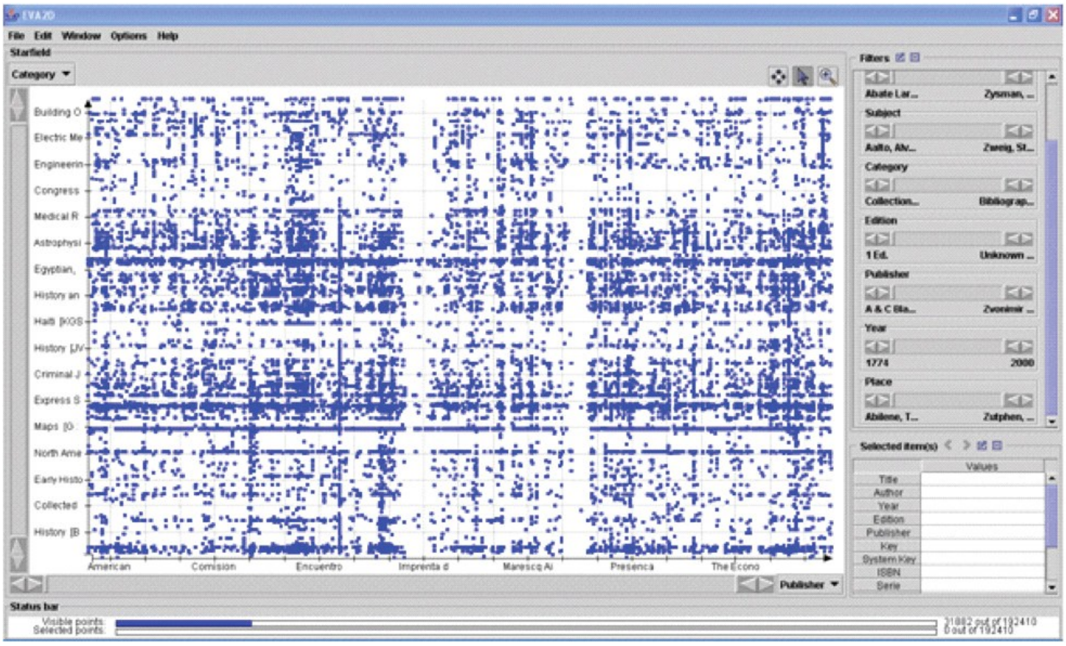

Figure 2: The EVA2D visualization interface

As observed in the figure, EVA2D presents the now classic scatter plot used in starfields, and also adds functionality to modify the user's perspective by changing the main attributes on display. In the figure, book categories are plotted against their publishers, but the axes may alternatively refer to attributes such as publication year, author name, or place of publication. Interface controls included a magnifying glass (see top right) and filtering sliders (right panel) that allow users to include or eliminate from the display value ranges for any of the collection's attributes. Details of items on display can be viewed on demand (bottom right corner) by clicking on their visual representations.

In order to deal with scale issues, EVA2D introduced four main features: (1) pre-computation of graphical data, which significantly reduced rendering time; (2) simultaneous bi-dimensional selection or direct zooming, which made it possible for users to visually select and work with subcollections; (3) precision filtering mechanisms, based on zoom bars (a variation of a scroll bar), which allows for refinement of the data in the starfield display according to values of additional attributes, and (4) quasi-immediate feedback, which implies that only the initial and final positions for zoom bars are recalculated and rendered. Implementation and experimental details regarding this work are reported in (Silva, Sánchez, Proal, & Rebollar, 2003).

For the purpose of exploring the ways in which EVA2D could support collection understanding, we conducted user studies that produced interesting results in terms of the interpretations given to the starfields. The evolution of collections over time or the abundance or scarcity of items in general subject areas were very easy to grasp for most users. But as users experimented with zooming and filtering functionality, they were able to spot patterns that helped them discover the emphases of collections. For example, some patterns made them conjecture whether a given collection supported specific curricula. More details on our experiments with EVA2D are reported in (Sánchez, Twidale, Nichols, & Silva, 2005).

Providing Help to Make the Most of Starfields

Also based on our user studies, shortcomings of our basic application of starfields became evident. Functionality suggested by users included visual cues to indicate various aspects of items in the collections being visualized, more accurate filtering options, and improved responsiveness to prevent confusing action sequences with interface components. One problem we observed was the need for users to maintain the context as they zoomed into areas of their interest. They sometimes became disoriented and were not able to easily determine their location with respect to the entire collection. In the next stage of our development, we addressed these issues.

StarFish is a visualization interface that continues to use starfields as its base representation for collections, but also incorporates fisheye viewing functionality. When the user selects a region of the starfield, a magnifying effect is applied to that region, whereas the rest of the starfield is scaled down. In essence, the entire object collection is permanently available on the interface, but a sub-area is shown in detail and its elements can be selected and viewed with further detail. Switching to a different area can be easily accomplished by clicking on its scaled down representation. We developed an operational prototype for this interface on top of a number of heterogeneous collections that make up the Open Network of Digital Libraries (ONeDL, or RABiD, after its initials in Spanish, http://www.rabid.org.mx). Digital theses, journal papers and ancient books are the main document types provided by member institutions of RABiD. Descriptors for documents in these distributed collections are exposed through metadata servers and collected via the OAI-PMH protocol described earlier.

Figure 3 illustrates the main components of the StarFish visualization interface. Noticeably, small, colored icons are used to represent collection items. The shape of each icon indicates the type of document, whereas colors are associated with institutions that provided the documents. Filtering functionality (on top of the starfield in this figure) allows users to easily include or exclude documents of specific types or provided by a given institution. By clicking on one of the grid regions, users can magnify its contents, obtain metadata or even the actual documents, whereas the rest of the regions become smaller but remain available for selection. StarFish has been effective both for providing a uniform overview of heterogeneous collections and for facilitating exploration and close examination of documents that are related regardless of their provider. Further information on StarFish can be found in (Sánchez, Quintana, & Razo, 2007).

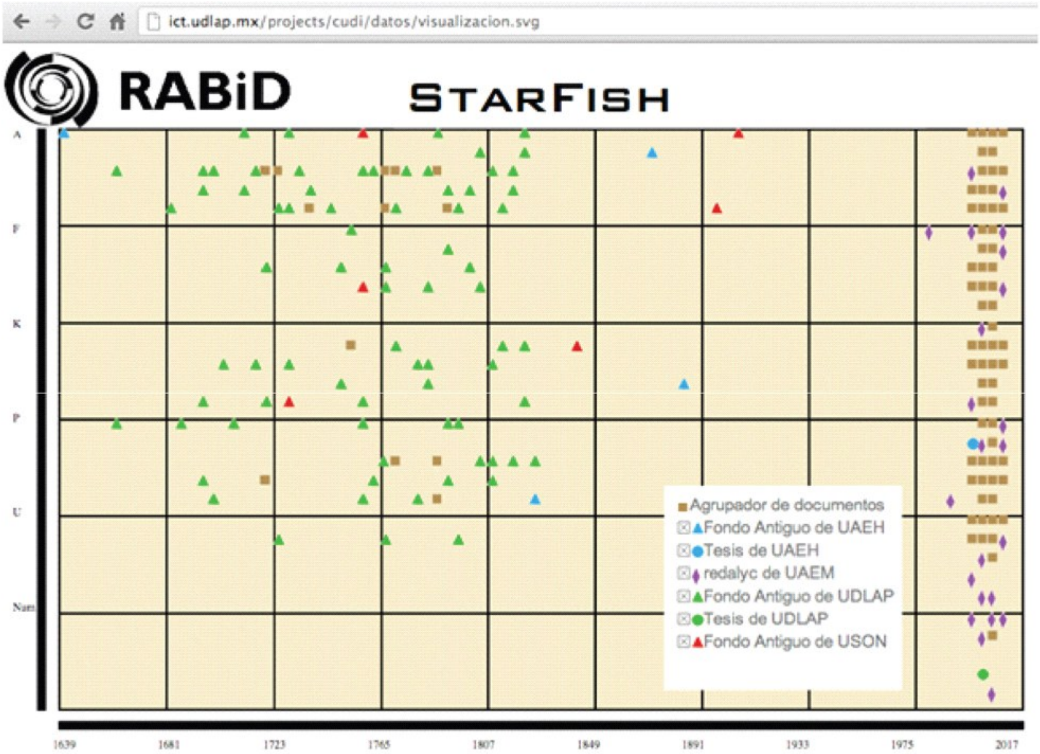

Figure 3: The StarFish visualization interface

Discovering Relationships and Collaboration Networks

For the next stage in our search for means to assist users in understanding collections, we focused on the discovery of implicit relationships among collection items and, more specifically, relationships among authors of documents that may result in potential collaboration networks. Implicitly, digital collections can provide access to knowledge regarding the contents and structure of their documents, as well as the communities of authors, institutions and users of all available information elements.

Finding collaboration networks can be helpful in many areas for at least three main reasons: (1) Researchers may be interested in learning about the global community that is producing scientific advances in their area; (2) Researchers need to find potential partners with similar or complementary interests so as to assemble multi-disciplinary, multi-institutional or multi-national teams required for funding opportunities; and (3) Funding agencies may want to become familiar with existing or potential collaboration networks so as to make sure calls for proposals will find appropriate audiences as well as to validate proposed research groups.

We have successfully applied the process illustrated in Figure 1 for harvesting collections and inferring relationships. In what follows, we discuss two approaches we have explored for visualizing and exploring such relationships so users are able to discover potential collaboration networks.

Egocentric Node-Link Visualizations

As noted earlier, node-link visualizations have been used frequently to represent relationships such as those occurring in social networks. Though generally very intuitive, node-link graphical representations do not scale well. For very large collections of documents, which are common in current settings, the number of relationships rapidly becomes unwieldy and a graphical representation of that complex web of nodes and links becomes impractical.

The feedback we received from user studies with low fidelity prototypes of node-link visualizations indicated this was the most natural representation for collaboration networks. Thus, we decided to tackle scale issues by presenting users with a so-called egocentric representation that shows, at its center, a person or institution for whom a collaboration network is needed, and on its periphery, people or institutions with various degrees of relatedness or collaboration potential. The collaboration potential between authors or institutions is estimated in terms of factors such as co-authorship and overlapping subjects of interest as derived from their publications and is part of the inference and abstraction construction process described earlier.

Even for a single author or institution, the number of potential collaborators can be unmanageably large. We thus have designed various graphical interface components that are used to easily filter out irrelevant relationships, as illustrated in Figure 4. First, the size and color of nodes are related to the strength of their relationship with the entity at the center of the display. Three sizes and colors are used according to thresholds defined experimentally for the estimated relatedness values. Colors can be selected by the user from palettes shown on the top right area of the interface. Second, the thickness and colors of links or edges are also related to the strength of the estimated relatedness among potential collaborators of the author or institutions at the center of the display. Finally, sliders at the top can be used to filter potential collaborators depending on the strength of their relatedness with respect with the entity at the center or among themselves. Text fields on the top left corner of the interface allow users to search the specific entity that will be placed at the center of the visualization.

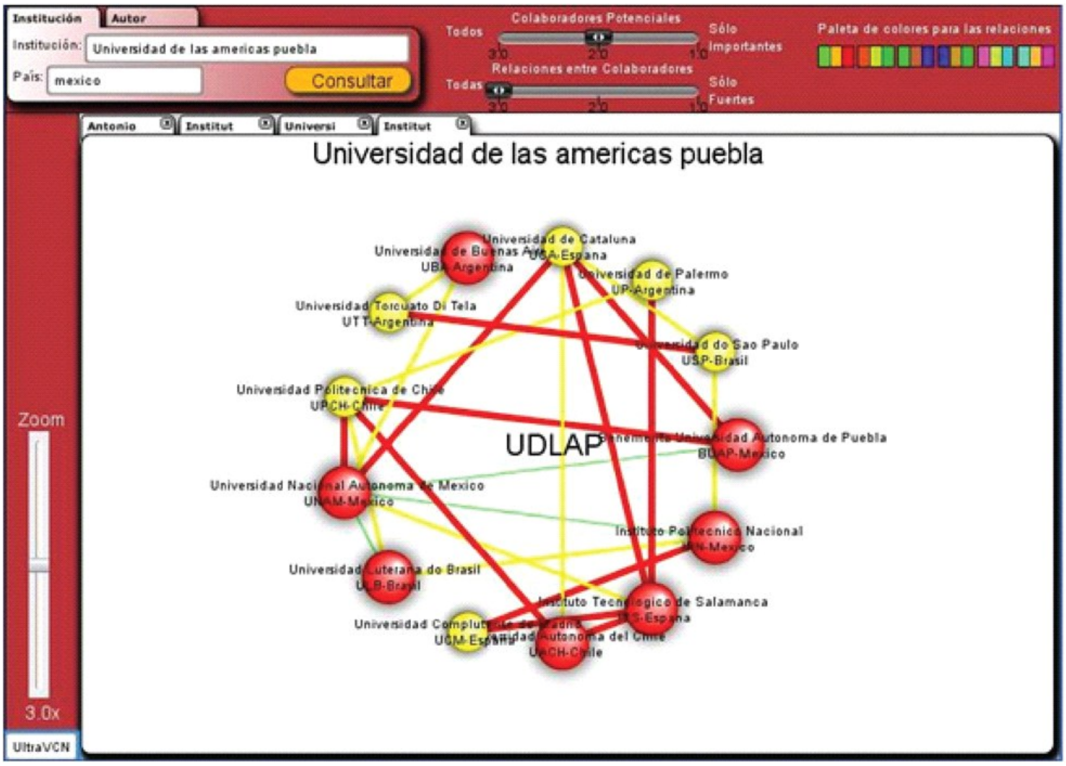

Figure 4: An egocentric node-link visualization of collaboration networks

User studies in which this visualization has been used have produced encouraging results. We designed exploration tasks for researchers in various domain areas and they were able to delimit nodes and links rapidly and suggest potential relationships and collaboration networks that would not have been evident solely from accessing publications in the underlying collections. Detailed results from user studies have been reported in (Hernández-Bolaños, 2010).

Discovering Collaboration Networks on Starfields

Whereas node-link visualizations have some advantages, an important drawback for collection understanding is that they do not facilitate maintaining the context for the user to have access to a general, clear overview of the entire set of relationships at any given time. We decided to explore alternative interfaces for collaboration networks on our previous experience, particularly with EVA2D and StarFish. We considered that starfields should provide a compact graphical representation that would include all possible relationships. Also, StarFish was a useful evolution of basic starfields but still posed the need for more semantic grouping of attributes on both axes, as generally the horizontal axis presented chronologically ordered data, whereas the vertical axis presented attributes that were ordered alphabetically. Particularly in the latter case, selecting sub-areas did not result in meaningful partitions.

Our next development, which we termed *OntoStarFish*, results from two key changes to the StarFish base scheme: Adding semantics to starfield axes, and introducing fisheye views through multiple lenses that allow for examination of several regions of the starfield simultaneously.

Figure 5 provides a graphical explanation of the concepts involved in the design of OntoStarFish. In contrast with the original starfields, in which axes display linearly a set of possible values for an attribute, *ontological axes* in OntoStarFish are not flat with respect to semantic groupings of attribute values. Instead, we can display attribute values that may be expanded or collapsed at various levels of abstraction, as selected by the user, corresponding to specializations or generalizations represented by the lightweight ontologies derived from the collections.

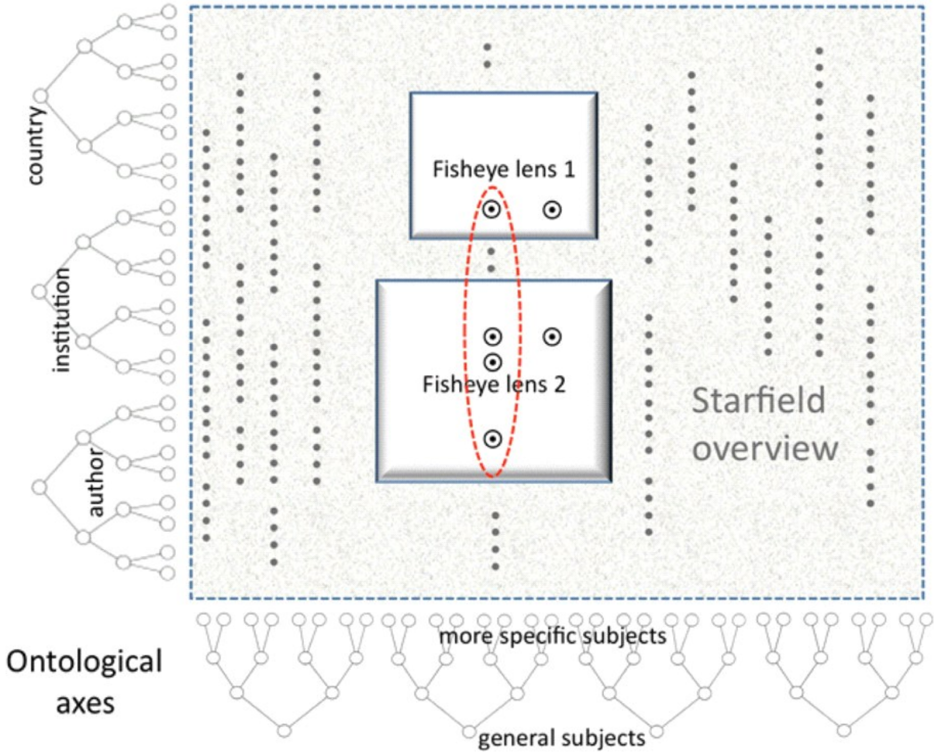

Figure 5: OntoStarFish elements: Ontological axes and multiple fisheye lenses

With the introduction of what we refer to as ontological axes, it becomes possible to explore countries, institutions and researchers, and look at the relationships each of these elements may have with a general or particular research subject, all simultaneously on the same display. As shown in Figure 5, the horizontal axis is used to display an ontological arrangement of subjects, which are derived from collections, whereas the vertical axis is used to display the countries, institutions or names of the authors of the publications. In ontological terms, subjects on the X-axis participate in "is-a" relationships, whereas people and places on the Y-axis play a role in "belongs-to" relationships. In both cases, it is the user who should decide the level of detail to be displayed. A dot (or a *star*) on the starfield indicates there are documents in the underlying collections on a given subject (or sub-subject), which have been authored in a particular country, at a given institution or by a specific author.

When using the OntoStarFish interface for exploring relationships, the levels of abstraction may vary even on the same axis. In a likely scenario, a user may need to visualize dots that refer to publications authored by specific researchers (author level) at, say, the University of Valparaíso (institution level) in Chile (country level) on a sub-topic such as "Fisheries Science", which may be classified under the more general topic of "Marine Biology." The same user should be able to explore, on a different area of the interface, dots that also indicate publications about "Fisheries science" but authored, say, in Argentina (displaying only the country level). Yet another area of the interface may show whether Brazil (at the country level) has work published in the general subject of "Ecology". Users should be able to expand or collapse any section of the ontological axes depending on their interests, and the starfield should be updated accordingly.

One advantage of fisheye views is that they allow for the exploration of details in a starfield while keeping the user oriented in the overall view of the collections. After observing users who performed tasks with low fidelity prototypes, we realized that a single fisheye view would not be sufficient for a thorough exploration and detection of relationships. The new ontological structure suggests that various collaboration networks can be displayed simultaneously, but they may not be evident, as the large number of relationships can hide them. A single fisheye lens may draw attention to relationships in a particular area, but those that are not contiguous still would remain out of sight. As illustrated in Figure 5, we introduced multiple fisheye lenses the user can move around to display details of various regions of the starfield. When two fisheye lenses are positioned in a way that they overlap vertically (as shown in the figure), potential collaboration networks at any level (researchers, countries or institutions that work on related subjects) correspond to coinciding columns in multiple fisheye views. This is exemplified in Figure 5 by the dots within the dashed ellipse. When the lenses overlap horizontally, they can be regarded as referring to potential collaboration networks within the same country or institution (on different subjects). Users should also be able to adjust the size of fisheye lenses so as to dynamically delimit areas of interest. We elaborate on how collaboration networks can be detected next, as we describe a prototypical implementation of OntoStarFish and provide further details of its user interface.

In order to demonstrate how the concepts involved in the design of OntoStarFish, we have implemented an operational prototype, which also takes advantage of the distributed collections provided by the RABiD initiative mentioned earlier. In what follows, we describe the resulting visualization through the use of scenarios that illustrate how OntoStarFish can be used to reveal collaboration networks. We have applied inference methods to find implicit relationships in test collections that involve nearly 120,000 documents about a large number of subjects. Participant data providers include institutions based in six countries from Latin America and Spain.

Finding the Locations of Related Researchers

OntoStarFish can be used to explore existing relationships and determine the affiliation of researchers who work on given subjects. Consider the scenario depicted by Figure 3, in which three fisheye lenses are placed over the starfield. The reddish dots on top of the "Agriculture" label indicate that five out of the six countries listed on the vertical axis have some published work on that subject. When the cursor is placed the cursor over a dot at the intersection of a country and the general subject of "Agriculture", a legend appears next to the cursor indicating the number of researchers in that country (nine in this example) who are potential collaborators in that field.

Each of the graphical elements in OntoStarFish is clickable. When the user clicks on a dot within a lens a the level displayed on Figure 6, both the corresponding country and general subject are expanded, as illustrated in Figure 7. This same effect could have been accomplished by clicking separately on the corresponding axis labels. On the vertical axis, acronyms for all universities or institutions in the country are displayed, whereas on the horizontal axis sub-subjects, such as Agronomy and Aquaculture, are displayed. The starfield is updated accordingly so as to make it possible to visualize relationships at that deeper level. It is worth noting that when an item is expanded on any axis, the others are collapsed but remain on the display to maintain context.

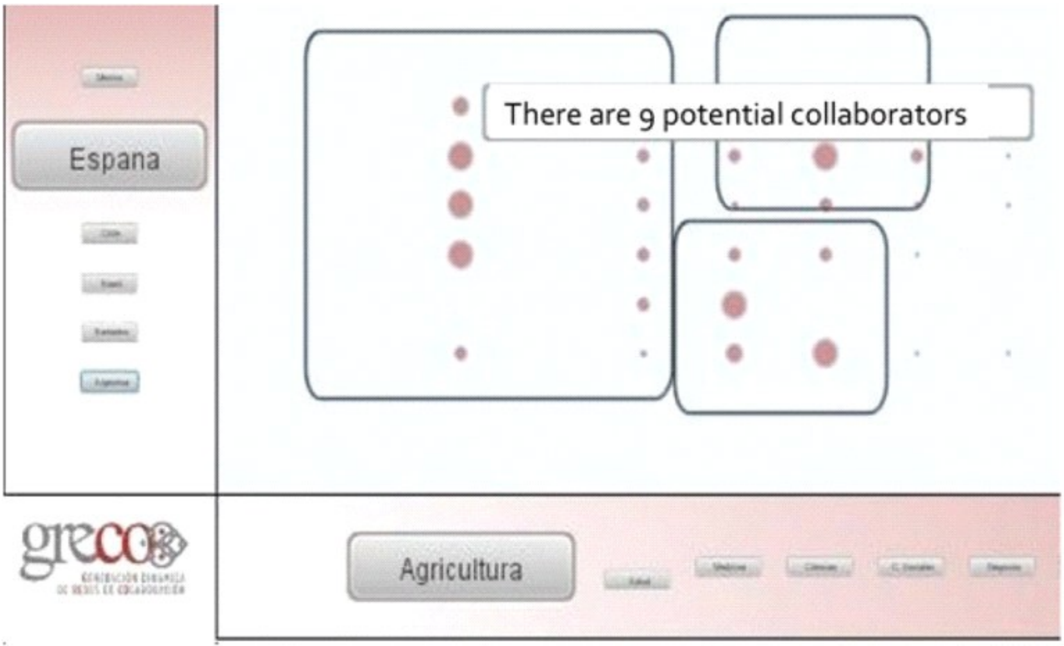

Figure 6: Visualization at the country and general subject levels

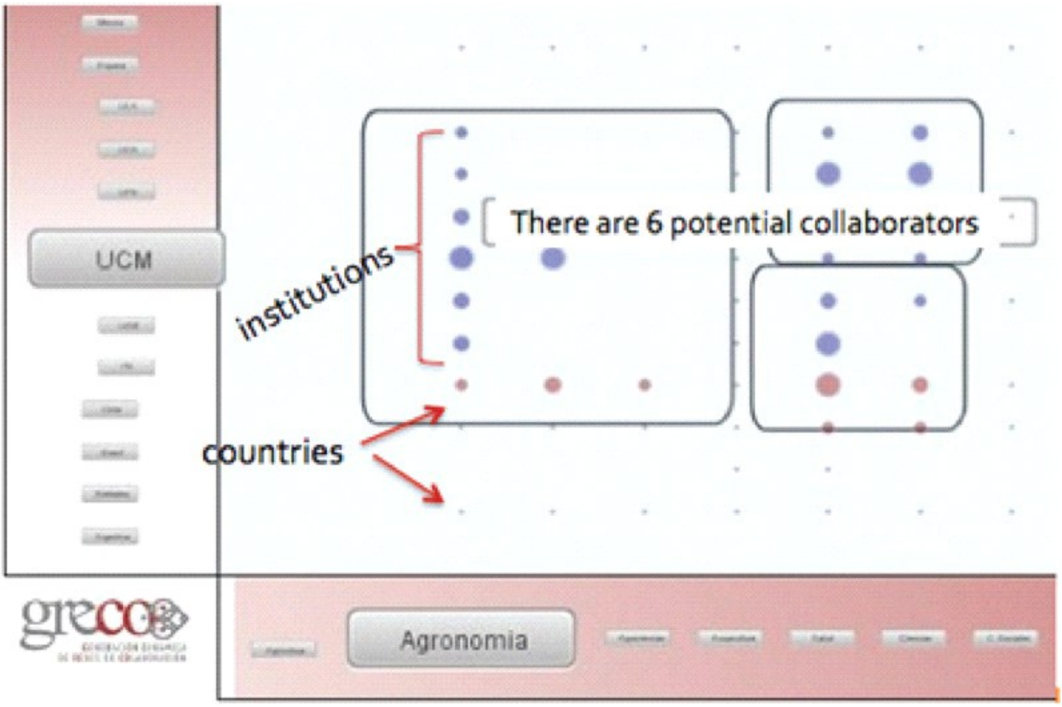

Figure 7: Visualization of relationships at the levels of institution and specific subjects

Three main interface components help users in coping with the increase of information on the starfield that results from expanding ontological elements into its sub-categories: Fisheye views, ontological layouts and color coding. Fisheye lenses allow users to explore specific areas by magnifying their contents as they are moved around the display. Ontological arrangement on both axes guarantee that items displayed contiguously also are semantically related. Each subject level is displayed on a different color. As highlighted by arrows and labels on top of the interface in Figure 7, blue dots help distinguish the new institution-subject relationships from the already existing country-subject relationships.

It is interesting to note that the starfield at this point displays universities within Spain that might have common interests and therefore could be in a position to collaborate, but also the countries with which potential collaboration networks could be established. By moving the lenses around the starfield these relationships may become more evident. In Figure 7, the user has clicked on the blue dot that indicates that the institution labeled UCM (actually "Universidad Complutense de Madrid") has authors who publish in the area of Agronomy, and a legend next to the cursor indicates six authors fall into that category. If the exploration continues and the user clicks on this relationship, both elements are expanded again, as illustrated in Figure 8 (one of the fisheye lenses also has been resized to include as many relationships as possible).

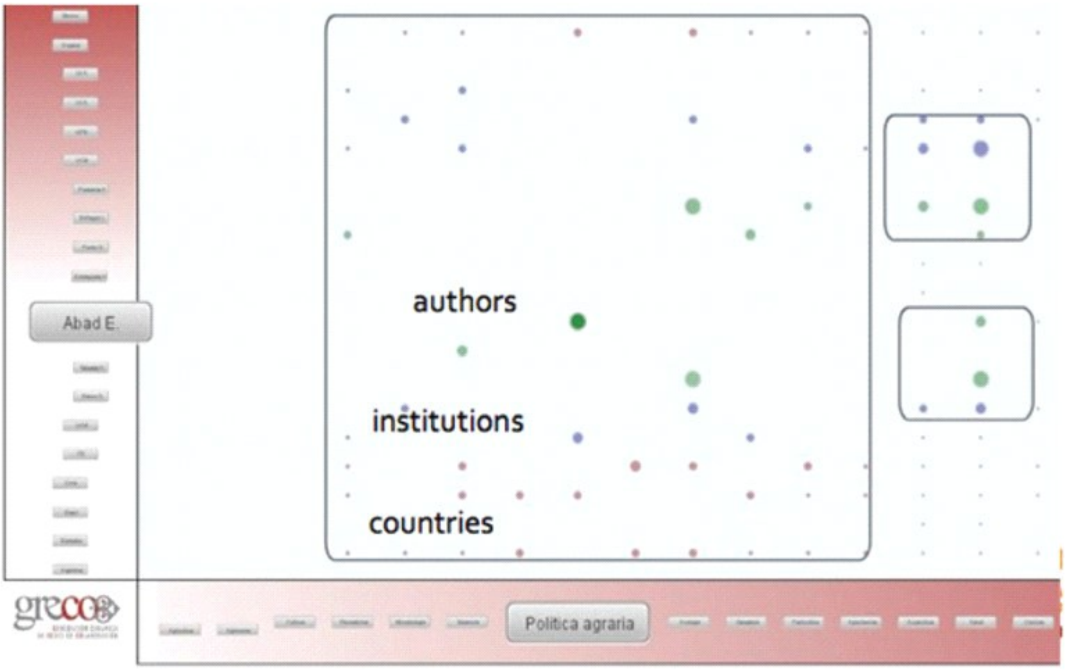

Figure 8: Visualization at the author level and more specific subjects

Figure 8 shows, on the vertical axis, all researchers within UCM, whereas more specific subjects within Agronomy are displayed on the horizontal axis. Relationships between authors and subjects are now indicated by green dots, which coexist on the display with blue and reddish dots that refer to institutions and countries, respectively. In this example, the relationship between an author named "Abad" and the specific subject of "Agrarian Policy" has been highlighted. As noted by the labels placed on top of the interface, relationships at all three levels of the vertical axis coexist on a single display, each colored differently: authors (green), institutions (blue) and countries (reddish). The starfield thus shows, by interpreting the same column, that the university (UCM) does not have other researchers who have published on the subject being examined, as there are no other green dots on the column, but the blue dot below the main author-subject relationship indicates there is one other university in the same country (Spain) who also has published on the subject, and that there are potential collaborators in two other countries (reddish dots on both ends of the column). It can also be noted that rows of the starfield provide helpful information as they indicate all subjects of interest for a given author, institution or country. This may be particularly helpful when collaborators are needed for multi-disciplinary projects.

Using Multiple Fisheye Lenses to Explore Relationships

Should we want to learn about potential collaborators for the researcher "Abad" in the two countries shown as red dots displayed in Figure 8, we would have to click on each of them to find the corresponding institutions and then on the blue dots to list the researchers. Given that their countries are on both ends of the vertical axis, if we used only one fisheye lens it would become difficult for the user to highlight the relationships of interest, as too many dots would be covered by one lens, and many relationships that are not relevant for the collaboration network would be included in the fisheye view.

The introduction of multiple fisheye lenses addresses these issues, as shown in Figure 9. In this case, the vertical axis includes items for institutions in three countries and researchers in only one university in each country. Fisheye lenses have been placed on green dots that form a column on top of the subject of interest (Agrarian Policy), thus allowing the relationships that are not contiguous to become evident for the user. The user may decide to add or remove fisheye lenses depending on the task at hand and the complexity of the starfield.

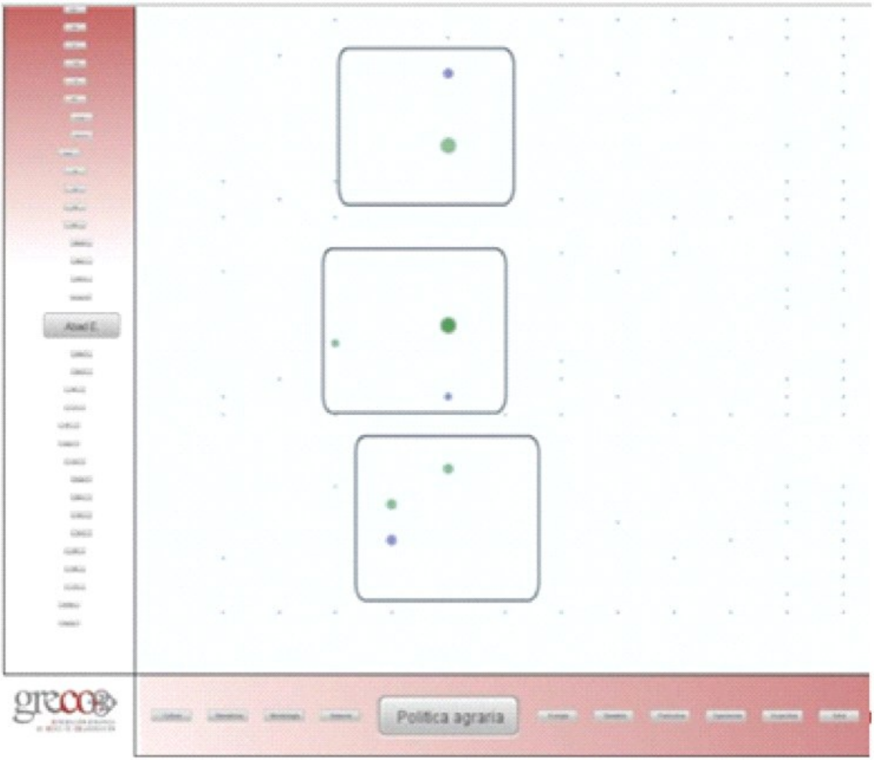

Figure 9: Using multiple fisheye views for exploring non-contiguous areas

Evaluation of OntoStarFish

OntoStarFish has been evaluated at various stages of its development. Formative evaluation based on the observation of two groups of users motivated various adjustments to the user interface and produced the version of the prototype discussed in this paper. We designed a series of tasks related to the characteristics and goals of OntoStarFish, particularly for determining its effectiveness for detecting collaboration networks. Overall, 28 users have been observed while using OntoStarFish in sessions lasting from 30 to 45 minutes. Participants have included university faculty and students with varying backgrounds including business, education, information technologies, industrial engineering and environmental engineering. Sessions were recorded and post-test questionnaires were applied to obtain explicit feedback.

The OntoStarFish visualization interface is very specialized and is far from conventional. Still, users provided positive feedback on its helpfulness for exploring collaboration networks. They were able to find potential collaborations in other institutions and other countries. Semantic grouping of data on the axes, which is a key feature, also was positively evaluated. Users were eager to explore the functionality afforded by clickable elements and expanded and collapsed data on the axes and the starfields at will. Some users even suggested applications for social networks in general and for managing relationships in business and enterprise settings.

We also obtained interesting feedback from open questions in post-test questionnaires. The most important aspect noted by users had to do with the use of multiple fisheyes. Users liked the possibility of exploring different areas on the same starfield and having access to different collaboration groups through multiple lenses. Also, the resizing capability feature of each fisheye was very useful for exploring larger (or smaller) areas of interest over the starfield. We also have observed that users rely on multiple fisheye views to explore several areas of the starfield. Typically, when they have located an area of interest they leave a fisheye lens fixed and start using another one to explore related data, especially when the axes have been expanded to a great level of detail.

The use of colors (red for countries, blue for institutions and green for researchers) made it easy to determine not only the ontological level users are exploring but also to focus on the kinds of groups users wanted to explore. Another important feature commented as positive was the indentation on each axis which helped users in maintaining their focus on the elements they were exploring. Zooming into the various levels of the subject ontology also contributed to find potential collaboration networks.

As for areas of improvement, users raised two main issues. They thought that the starfield-based representation of relationships was not always clear and that some training would be needed before researchers can use OntoStarFish comfortably. In spite of this, we observed that even novice users became familiar rapidly with the interface and were able to perform tasks with some fluency in less than fifteen minutes.

Based on feedback from users and further experimentation, we have produced an improved version of the OntoStarFish interface. Salient enhancements include: (1) a navigation bar to constantly provide contextual information regarding the levels dealt with on the vertical and horizontal axes, as well as on the number of potential collaborators to be displayed if the user clicks on a given dot on the starfield; (2) handy buttons for easily adding or removing fisheye lenses while exploring starfields; and (3) status messages to avoid frustration from clicking and not seeing instant results when the system is recalculating the starfield. Additionally, in the new version of the interface, dots resulting from expanding an ontological level remain highlighted even if the cursor moves to explore a different area. Users suggested this in order to provide visual clues that reduce disorientation or confusion when multiple actions are performed on the interface.

OntoStarFish highlights the importance of the representation mechanism used to support the visualization of collaboration networks, which is but one aspect of collection understanding. Metadata associated with documents in digital collections does not generally exhibit any evident pattern that describes such networks. The levels of detail provided by OntoStarFish seem to offer several advantages over the traditional node-link scheme, including the visualization of complex network structures, a detailed view of potential collaborative groups, and the visualization of multiple collaboration networks at the same time.

ONGOING WORK AND FUTURE RESEARCH DIRECTIONS

We plan to continue working on two-dimensional visual representations for supporting collection understanding. Our OntoStarFish interface leaves ample room for improvement and further research. For instance, users put forward the need for options for visualizing documents from which relationships are inferred, as researchers find it important to learn about the reasons why potential collaborations are suggested on specific areas. This implies adding new features to the interface, such as a document previewing at the last level of the subject ontology. Performance also is a very important feature of visualization methods. We also are considering the exploration of the alternative applications for OntoStarFish suggested by our users.

Also, we recently started working with a network of institutional repositories that will hold and disseminate the scientific production of a large number of universities. As part of our contribution to the network, we are developing visualization interfaces to assist users in understanding what the repositories are about by providing graphical representations of the collections. Our goal is to provide a visualization mechanism that will represent the subjects in the collections as a map in which regions are charted so their color and shape denote documents grouped according to thematic similarities. The map metaphor has shown significant promise in our initial studies with potential users. In the resulting visualization, map regions displayed with the same color refer to documents about very similar subjects. The shape of the regions is defined in such a way that they get as close as possible to other groups of documents. Some map regions may actually result in disjoint areas of the same color so as to reflect similarities with other thematic areas. In order to obtain meaningful groupings, our graphical representation is tightly coupled with the classification mechanism, which is based on the notion of Self-Organizing Maps (SOM), described earlier.

In order to maintain the number of groups manageable even for very large collections, thresholds are being defined so maps comprise only a relatively small number of groups. This implies that the SOM technique is applied successively, producing naturally a hierarchy of maps that can be used directly for navigation purposes. This hierarchy of maps is illustrated in Figure 10. Starting from the upper left corner, collections are visualized as a map that consists of ten regions, which actually refer to seven general groups of documents, since two of the groups are divided into two and three regions of the same color, respectively. Users will be able to select a region from the map, which will take them to a more specific grouping, and so on until actual documents are displayed. For the sake of illustration, the figure shows only three maps at the second level and one at the third level of detail.

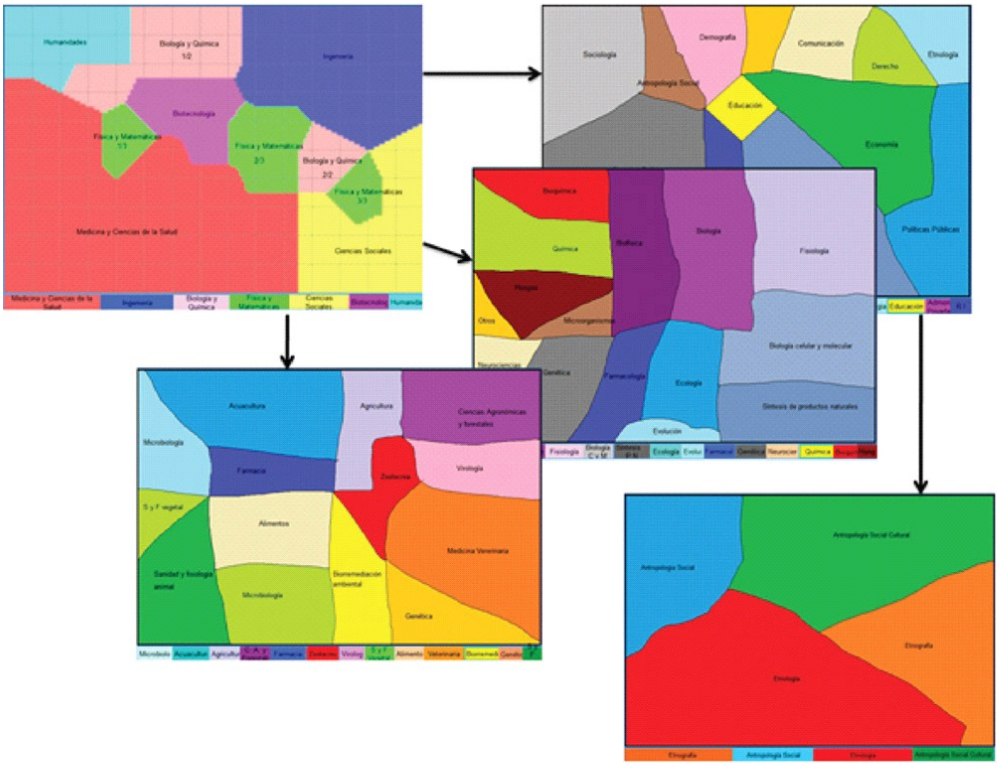

Figure 10: Distributed collections visualized as a hierarchy of maps

One additional feature that will be explored with SOM-based visualization is a timeline slider that will be placed at the bottom of the interface. This will provide users with further analysis tools, as they will be able to understand how a collection or a number of collections have evolved over time. Based on our current low-fidelity prototype, Figure 11 depicts the state of collections at three points in time. User studies indicate this should be an intuitive means for visually keeping track of the emergence, growth, decrease and generally the evolution of the subjects or any attributes related to documents (or other objects) held by digital collections. We expect to release a working prototype for this visualization during the first half of 2013.

Figure 11: Evolution of collections over time

CONCLUSION

This chapter has discussed the potential of information visualization techniques for supporting collection understanding, a notion that promotes an exploratory approach and provides users with a general sense of digital collections. We have emphasized the role of visual representations as means to assist the user in discovering patterns and structures that occur implicitly in collections. Through a series of developments, we have demonstrated how two-dimensional visualization techniques, such as starfields, fisheye views and self-organizing maps can be used to provide uniform interfaces for heterogeneous collections and to reveal hidden patterns, implicit structures and relationships, including potential collaboration networks. Throughout the process, we have relied on user studies to inform all our design decisions.

Although significant progress has been made, much work needs to be done in order to provide users with more intuitive, meaningful and responsive visualization interfaces that will help them understand the ever-growing digital collections at hand. Information visualization will continue to play a key role in this quest, but it will also be important to take advantage of new developments in natural user interfaces, including multi-touch interactive surfaces and gestural interfaces. We plan to work on versions of our visualization interfaces for these environments, which also should be produced by working closely with end users.

ACKNOWLEDGMENT

The concepts discussed in this chapter as well as the implementation and evaluation of prototype software that demonstrates their potential are the result of the collaboration of various generations of talented members of the Laboratory of Interactive and Cooperative Technologies at the Universidad de las Américas Puebla. Special recognition goes to Federico Hernández Bolaños, Nilda Galán, Ernesto Gutiérrez, María Auxilio Medina, Carlos Proal, Alfredo Ramos, Antonio Razo, Christian Rebollar, and Guadalupe Quintana. Yenny Méndez was instrumental in producing the final version of this document. Support for our projects has been provided by our home institution (UDLAP) as well as by Conacyt and the University Corporation for the Advancement of the Internet (CUDI) through its Program for Advanced Applications.

REFERENCES

Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 313–317. Boston, MA: ACM Press. doi:10.1145/191666.191775.

Ahlberg, C., Williamson, C., & Shneiderman, B. (1992). Dynamic queries for information exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 619–626. New York: ACM Press. doi:10.1145/142750.143054.

Amavizca, M., Sánchez, J., & Abascal, R. (1999). 3DTree: Visualization of large and complex information spaces in the floristic digital library. In *Proceedings of the Workshop on Computer Mediated Interaction, 2nd Mexican International Conference on Computer Science*. Pachuca, México: ACM Press.

Barabasi, A.-L. (2010). *Bursts: The hidden pattern behind everything we do.* Retrieved from http://dl.acm.org/citation.cfm?id=1941887.

Borst, W. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. (Dissertation). Enschede, The Netherlands, University of Twente.

Buchel, O. (2011). Designing map-based visualizations for collection understanding. In *Proceeding of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries.* New York: ACM Press. doi:10.1145/1998076.1998169.

Chang, M., Leggett, J. J., Furuta, R., Kerne, A., Williams, J. P., Burns, S. A., & Bias, R. G. (2004). Collection understanding. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries*, 334–342. New York: ACM. doi:10.1145/996350.996426.

Cunningham, S. J., & Bennett, E. (2008). Understanding collection understanding with collage. In *Proceedings of the 11th International Conference on Asian Digital Libraries: Universal and Ubiquitous Access to Information*, 367–370. Berlin: Springer-Verlag. doi:10.1007/978-3-540-89533-6_46.

Fung, B. C. M., Wang, K., & Ester, M. (2003). *Hierarchical document clustering. Encyclopedia of Data Warehousing and Mining*. Hershey, PA: IGI Global.

Girgensohn, A., Shipman, F., Turner, T., & Wilcox, L. (2010). Flexible access to photo libraries via time, place, tags, and visual features. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, 187–196. New York: ACM. doi:10.1145/1816123.1816151.

Giunchiglia, F., Marchese, M., & Zaihrayeu, I. (2007). Encoding classifications into lightweight ontologies. *Journal on Data Semantics VIII, 4380*(4380), 57–81. Retrieved from http://dx.doi.org/10.1007/978-3-540-70664-9_3.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. In H. Burkhardt & B. Smith (Eds.) Knowledge Creation Diffusion Utilization, *5*, 199–220. doi:doi:10.1006/knac.1993.1008.doi:10.1006/knac.1993.1008

Herman, I., Melancon, G., & Marshall, M. S. (2000). Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, *6*(1), 24–43. doi:10.1109/2945.841119

Hernández-Bolaños, F. (2010). *UltraVCN: Uso de filtros y características gráficas para la visualización de redes de colaboración*. (Unpublished B.Sc. Thesis). Puebla, Mexico, Universidad de las Américas.

Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., & Giannopoulou, E. (2007). Ontology visualization methods-A survey. *ACM Computing Surveys*, *39*(4), 10. doi:10.1145/1287620.1287621

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480. doi:10.1109/5.58325

Lassila, O., & McGuinness, D. (2001). The role of frame-based representation on the semantic web. *Linköping Electronic Articles in Computer and Information Science, 6*(5), 2001. Retrieved from http://www.ksl.stanford.edu/KSL_Abstracts/KSL-01-02. html.

Perer, A., & Shneiderman, B. (2008). Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis. In M. Burnett, M. F. Costabile, T. Catarci, B. DeRuyter, D. Tan, M. Czerwinski, & A. Lund (Eds.), Human-Computer Interaction (265–274). Florence, Italy: ACM Press. doi:doi:10.1145/1357054.1357101.doi:10.1145/1357054.1357101

Proal, C., Sánchez, J. A., & Fernández, L. (2000). UVA: 3D representations for visualizing digital collections. In *Proceedings of the Third International Conference on Visual Computing*, 185–192. Mexico City: IEEE Press. Retrieved from http://ict.udlap.mx/pubs/UVA2000.pdf.

Ramos, A., Sánchez, J. A., & Hernández-Bolaños, F. (2010). OntoStarFish: Visualization of collaboration networks using starfields, ontologies, and fisheye views. In *Proceedings of the 3rd Mexican Workshop on Human Computer Interaction*, 44–53. San Luis Potosí, México: Universidad Politécnia de San Luis Potosí. Retrieved from http://dl.acm.org/citation.cfm?id=1978702.1978713.

Sánchez, J. A., Medina, M., Starostenko, O., Benitez, A., & Domínguez, E. (2012). Organizing open archives via lightweight ontologies to facilitate the use of heterogeneous collections. In *Proceedings of ASLIB, 64*, 46–66. New York: Emerald Press.doi:10.1108/00012531211196701

Sánchez, J. A., Quintana, M., & Razo, A. (2007). Star-fish: Starfields+ fisheye visualization and its application to federated digital libraries. In *Proceedings of the 3rd Latin American Conference on Human-Computer Interaction*. Rio de Janeiro, Brazil: HCIR Press. Retrieved from http://clihc.org/2007/papers/StarFish_ID38_longpaper. pdf.

Sánchez, J. A., Twidale, M. B., Nichols, D. M., & Silva, N. N. (2005). Experiences with starfield visualizations for analysis of library collections. In *Proceedings of the Conference on Visualization and Data Analysis, 5669*, 215–225. San Jose, CA: SPIE Press.

Shneiderman, B. (1996). The eyes have it: A Task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages,* 336–343. Washington, DC: IEEE Press. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=545307.

Silva, N. N., Sanchez, J. A., Proal, C., & Rebollar, C. (2003). Visual exploration of large collections in digital libraries. In *Proceedings of the Latin American Conference on Human-Computer Interaction*, 147–157. New York: ACM Press. Retrieved from http://portal.acm.org/citation.cfm?id=944519.944535.

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., & Capone, T. (2009). Analyzing (social media) networks with NodeXL. In D. Hansen, B. Shneiderman, M. Smith, & R. Ackland (Eds.), *Proceedings of the Fourth International Conference on Communities and Technologies-CT 09* (255–264). New York: ACM Press. doi:10.1145/1556460.1556497.

Xu, J. J., & Chen, H. (2005). CrimeNet explorer: A framework for criminal network knowledge discovery. *ACM Transactions on Information Systems*, *23*(2), 201–226. doi:10.1145/1059981.1059984

Yakowenko, J., & Matange, S. (2004). Computer-implemented system and method for handling node-link representations. *U. S. Patent No. 7587409.* Retrieved from http://www.patents.com/us-7587409.html.

Yang, X., Asur, S., Parthasarathy, S., & Mehta, S. (2008). A visual-analytic toolkit for dynamic interaction graphs. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Las Vegas, NV: ACM Press. doi:10.1145/1401890.1402011.

ADDITIONAL READING

Ahn, J.-W., Taieb-Maimon, M., Sopan, A., Plaisant, C., & Shneiderman, B. (2011). Temporal visualization of social network dynamics: prototypes for nation of neighbors. In *Proceedings of the 4th International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, 309–316. Berlin: Springer-Verlag. Retrieved from http://dl.acm.org/citation.cfm?id=1964698.1964741.

Andronis, C., Sharma, A., Virvilis, V., Deftereos, S., & Persidis, A. (2011). Literature mining, ontologies, and information visualization for drug repurposing. *Briefings in Bioinformatics*, *12*(4), 357–368. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21712342 doi:10.1093/bib/bbr005

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Readings in information visualization: Using vision to think. In S. K. Card, J. D. Mackinlay, & B. Shneiderman (Eds.), *Information Display* (686). New York: Morgan Kaufmann. Retrieved from http://portal.acm.org/citation.cfm?id=300679.

Dörk, M., Carpendale, S., & Williamson, C. (2012). Fluid views: A zoomable search environment. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 233–240. New York: ACM. doi:10.1145/2254556.2254599.

Elmqvist, N., & Fekete, J.-D. (2010). Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, *16*(3), 439–454. doi:10.1109/TVCG.2009.84

Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, *10*(4), 289–309. doi:10.1177/1473871611416549

Graham, M., & Kennedy, J. (2010). A survey of multiple tree visualisation. *Information Visualization*, *9*(4), 235–252. doi:10.1057/ivs.2009.29

Grammel, L., Tory, M., & Storey, M.-A. (2010). How Information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 943–952. doi:10.1109/TVCG.2010.164

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Communications of the ACM*, *55*(4), 45–54. doi:10.1145/2133806.2133821

Keim, D. A., Mansmann, F., & Thomas, J. (2010). Visual analytics: how much visualization and how much analytics? *SIGKDD Exploration Newsletter*, *11*(2), 5–8. doi:10.1145/1809400.1809403

Kielman, J., Thomas, J., & May, R. (2009). Foundations and frontiers in visual analytics. *Information Visualization*, *8*(4), 239–246. doi:10.1057/ivs.2009.25

Liu, Z., & Stasko, J. (2010). Mental models, visual reasoning, and interaction in information visualization: A top-down perspective. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 999–1008. doi:10.1109/TVCG.2010.177

Lovejoy, M., Vesna, V., & Paul, C. (2011). *Context providers: Conditions of meaning in media arts*. Bristol, UK: Intellect Ltd.

Meiguins, B. S., Do Carmo, R. M. C., Almeida, L., Gonçalves, A. S., Pinheiro, S. C. V., De Brito Garcia, M., & Godinho, P. I. A. (2006). Multidimensional information visualization using augmented reality. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*, 391–394. New York: ACM. doi:10.1145/1128923.1128996.

Newman, M. E. J. (2000). The structure of scientific collaboration networks. In *Proceedings of the National Academy of Sciences of the United States of America*. Washington DC: National Academy of the Sciences. Retrieved from http://arxiv.org/abs/cond-mat/0007214.

Padia, K., AlNoamany, Y., & Weigle, M. C. (2012). Visualizing digital collections at archive-it. In *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries*, 15–18. New York: ACM. doi:10.1145/2232817.2232821.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148. doi:10.1109/TVCG.2010.179

Shneiderman, B. (2001). Inventing discovery tools: Combining Information visualization with data mining. In *Proceedings of the 4th International Conference on Discovery Science*, 17–28. London, UK: Springer-Verlag. Retrieved from http://dl.acm.org/citation.cfm?id=647858.738688.

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science American Society for Information Science*, *50*(9), 799–813. doi:10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Spence, R. (2007). *Information visualization: Design for interaction* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Stasko, J. (2000). An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies*, *53*(5), 663–694. doi:10.1006/ijhc.2000.0420

Steele, J., & Iliinsky, N. (2010). *Beautiful Visualization: Looking at Data through the Eyes of Experts* (1st ed.). Sebastapol, CA: O'Reilly Media, Inc.

Tufte, E. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.

Tufte, E. R. (1986). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tufte, E. R. (2001). *The visual display of quantitative information* (p. 200). Cockeysville, MD: Graphics PR.

Ward, M., Grinstein, G., & Keim, D. (2010). *Interactive data visualization: Foundations, techniques, and applications*. Natick, MA: A. K. Peters, Ltd.

KEY TERMS AND DEFINITIONS

Collaboration Networks: Structures that represent interconnected groups of people who work together at various coupling levels.

Collection Understanding: A technique intended to cope with the complexity of large digital collections that promotes an exploratory approach of repositories by providing users with a general sense of their contents.

Egocentric Network Visualization: A visual representation of a network that shows, at its center, a node of interest, and on its periphery, nodes with various degrees of relatedness.

Fisheye View: A visualization technique that allows users to explore in detail an area of interest while maintaining its context. Its name refers to the distortion effect created around non-important elements while those that are important are magnified.

Low-Fidelity Prototype: A low-cost, simplified representation of a design or concept, typically rendered on physical materials, used to gather user feedback at early stages of design.

Ontology: An explicit shared specification of a conceptualization. Lightweight and heavyweight ontologies can be distinguished according to the degree of formality involved in their encoding.

Starfield: A grid-based visualization method that highlights the intersections of two axes on a plane. Each axis typically is used to represent an attribute of the items being displayed.