Chapter 11 Benchmarks and Evaluation Criteria for Information Retrieval Visualization

Information retrieval visualization has more than several decades of history. Theoretical visualization models, pilot visual retrieval systems, and commercial visualization retrieval software packages have burgeoned. It is understandable that researchers and developers have paid more attention to innovative visualization retrieval technique development and system implementation, and less attention to research on evaluation of these systems and models in the early phase. That is because system evaluation usually lags behind system development and implementation. In the initial phase, the priority is model design and system development. Without available models/systems, it is impossible to conduct system evaluation. However, as the techniques and theories of information retrieval visualization mature and the commercialization of information retrieval visualization systems proliferate, evaluation of these systems and models is becoming a pressing issue in the field. The situation calls for better metrics and benchmark repositories to evaluate and examine these tools.

11.1 Information retrieval visualization evaluation

Unlike scientific visualization, information retrieval visualization as a branch of information visualization does not have a clearly defined inherent physical structure to visualize in a visual space. It leads to the diversity of information retrieval visualization models which are used to reveal and reflect abstract, invisible, semantic relationships among data in a data set. For instance, in a vector-based information system the spatial characteristic based visualization models, the multiple reference point based visualization models, the self-organizing map visualization models, the multi-dimensional scaling visualization models, etc. can be employed to describe and visualize the same dataset. Each model or environment demonstrates unique perspectives of the dataset. On the other hand, this diversity also increases the difficulty to evaluate these information retrieval visualization models/systems due to the lack of objective comparison standards. Furthermore, the richness of database types, multiformity of the described objects in a database, in conjunction with the complexity of information retrieval in the visualization environment, make the evaluation of information retrieval visualization an intriguing and challenging task. There are many database types available for visualization, ranging from a vector-based information model, to the Boolean based model, hierarchical information organization model, hyperlink-based data model, etc. Each possesses its own intrinsic data structures, characteristics, and data processing. Visualized objects in the same dataset may be quite different, let alone the visualized objects in the different datasets. As an information visualization environment changes, the ways of both information presentation and the corresponding retrieval would change. All these play a role in evaluation of information retrieval visualization. Users search information in an information retrieval visualization environment quite differently from a traditional retrieval system. In a traditional search environment, users usually enter a text-based query, select other search restrictions, choose the presentation structure of search results such as alphabetical ranking, chronological ranking, or relevance ranking, and make a relevance judgment. In a retrieval visualization environment, users may have to visually convert and "spatialize" their information needs in a visual space; understand the framework of a visual presentation, icons, and metaphors; interpret the visual display of projected documents or objects; and manipulate the display and interact with it. The search process in a visual retrieval environment is more complex than in a traditional retrieval environment. The evaluation of a retrieval visualization environment is more difficult than a traditional retrieval system.

In fact, the evaluation for information retrieval visualization has twofold: retrieval result evaluation and retrieval environment/interface evaluation. Recall and precision are two primary criteria for retrieval result evaluation while retrieval environment/interface evaluation has a different criterion system.

Recall and precision are widely recognized as evaluation criteria for traditional information retrieval systems. These criteria are no longer competitive enough for information retrieval visualization. In a study, traditional information retrieval with visualization was compared with information retrieval without visualization against the proposed criteria like documents saved per search, interactive task precision, and interactive user precision. The authors found that these precision-based criteria failed to handle the complex visualization situations (Veerasamy and Belkin, 1996). Notice that both recall and precision are basically designed to evaluate retrieved individual documents in a traditional retrieval system. In a retrieval visualization environment users not only retrieve individual documents at the micro-level but also retrieve aggregate information at the macro-level thanks to the visual configurations. Unfortunately, the latter cannot be measured by neither recall nor precision. Cugini (2005) addressed the performance metrics for presenting search results in a visual space. He examined the performance from the following perspectives: percentage of relevant documents found within a given time, relative error of response, relevance score of a selected document, time taken to find a relevant document, and time taken to answer a specific question.

People are aware of the importance of information visualization evaluation and have made efforts to solve the problem. One of the pioneering studies in information visualization evaluation was done by Shneiderman (1996). The author presented seven well-defined general criteria, which are: gaining an overview of an entire database, zooming on objects of interest, filtering out irrelevant objects, choosing a set of interest objects to get details if necessary, viewing relationships among objects, keeping a history of previous users' activities, and extracting a subset of a collection. The evaluation criteria are supposed to apply to all information visualization environments. After these evaluation criteria were adopted, they were used to evaluate four 3D information visualization designs (Wiss et al., 1998). In a study (Freitas et al., 2005), cognitive complexity, spatial organization, information coding, and state transition were identified as evaluation criteria for visual representation of information visualization techniques. Orientation and help, navigation and querying, and data set reduction were also brought out to examine the interaction mechanism of information visualization. These criteria were applied to evaluate an information visualization application, *Bifocal Browser*, in their study. Others try to analyze information visualization from the goal and task point of view. Visualization technique evaluation principles were presented along this line (Winckler et al., 2004). They specified the users' goals and verified whether they can reach these goals with an information visualization application. And then they identified interaction mechanisms that can accomplish the task, and graphic rendering function to show information and relating these goals. These can be summarized as four task levels: goal, generic tasks, interactions, and visual presentation. From the data mining angle, a study came up with the following criteria for information visualization: scalability, expressing domain knowledge, dealing with incorrect data, ease of classification and categorization, high dimensionality, visualization flexibility, query functionality, and summary of results (Grinstein et al., 2005).

Komlodi et al., (2004) conducted a survey to summarize information visualization evaluation experiments. After analyzing the natures and designs of fifty information visualization experimental studies, the authors classified the experimental studies and generalized four thematic groups of information visualization evaluations. They are controlled experiments comparing design features of an application, usability studies for an information visualization application, controlled experiments comparing multiple tools, and case studies of an application.

A methodology for testing a novel information retrieval visualization system was introduced (Morse and Lewis, 2002). Instead of testing all static and dynamic features of as information retrieval visualization system, some non-significant features are disabled or "de-featured" and only basic features are studied. The benefits of this strategy include focusing on the visual display, reducing the influence from context variables, simplifying experimental procedure, and using a larger number of subjects.

In order to offer a common evaluation testing environment similar to *TREC*, researchers in the information visualization field have set up a sample dataset, aiming to initiate the development of the evaluation benchmarks, to provide a common test environment available to the public, and to establish a forum to promote various evaluation methods. For each of sub-datasets, application domain was described, and open-ended domain specific tasks were provided. It was found that it was difficult to compare systems even with specific datasets and tasks

(Plaisant, 2004). A special journal issue about empirical evaluation of information visualizations was organized to address the growing concern about information visualization evaluation (Chen, 2000).

Existing research primarily gravitates around the evaluation of information visualization which has a much larger scope than evaluation of information retrieval visualization and a different emphasis. Although information retrieval visualization may be regarded as a special area of information visualization, it has its own unique features and distinctiveness. These unique features and distinctiveness must be integrated and reflected in its evaluation system. The former concentrates more on information visual representation and information expression in a visual environment while the latter concentrates more on information retrieval in addition to information visual representation in a visual context. There are interactive activities in information visualization environments, but there may not be information retrieval activities. However, it is crystal clear that information retrieval visualization has a natural connection to information visual representation. People cannot address information retrieval in a visualization environment without mentioning information visual representation. In fact, information visual representation is the foundation of information retrieval visualization. The characteristics and structures of information visual representation have a strong impact on the characteristics and features of an information visualization environment. Due to these differences, the evaluation criteria for information retrieval visualization should be different from those of information visualization. The evaluation for visual information retrieval should combine both information retrieval and information visualization.

Developing widely accepted and sound evaluation criteria for information retrieval visualization is an important and urgent research topic. Such an evaluation system can contribute to both theoretical research of information retrieval visualization and practical information retrieval visualization system development. The study would benefit researchers, designers, system developers, and end-users as well. An evaluation system would guide and steer researchers, developers, and designers toward optimal information retrieval visualization solutions, models and theories. Evaluation systems help them to discover potential features, to identify potential weaknesses of a visualization tool, and to avoid design loopholes. It can also be used by ordinary customers to select information retrieval visualization software among rival products. This would maximize their efforts to improve information retrieval visualization and encourage a more widespread adoption of information retrieval visualization.

The evaluation criteria should be valid, universal, fair and applicable to every kind of information retrieval visualization environments, offering a standard of comparative evaluation for across information retrieval visualization tools/models.

11.2 Benchmarks and evaluation standards

11.2.1 Factors affecting evaluation standards

It has been shown that developing a benchmark and evaluation system for information retrieval visualization is not a simple task because it is affected by various factors and variables. It is necessary to address the factors that actually play a role in information retrieval visualization evaluation. Identifying such factors would give people a better understanding of the later proposed evaluation benchmarks.

The first is information visualization task and data. Task and data should be discussed together because the task usually is intertwined with the data. The nature of data usually determines the task of a system. Information visualization tasks and data vary widely; which makes a unified evaluation methodology difficult to create. It is clear that no one general set of visualization tools will be suitable to address all problems (Grinstein et al., 2005). Information visualization systems usually are designed to target toward a specific problem and support tasks are very well associated with this problem. Therefore they behave differently when they are used for visualizing different datasets (Wiss et al., 1998). The true quality of an information visualization system can only be measured in the context of a particular purpose or task (Rushmeier et al., 1995).

The second factor is the interactivity of the information search process in a visual environment. A search process is a complicated process and needs a series of interactions between users and an information retrieval visualization system. Users navigate an information space, discover and explore relevant information based upon their information needs. These may comprise a variety of interactive activities. Weherend and Lewis (1990) categorized potential operations users may conduct in visual environments. These include locating an item from a known entry, identifying a set of unknown items, distinguishing objects from a presented set, categorizing objects described by users, clustering linked and grouped objects, distribution of specified categories, ranking a interest objects, comparing entities with different attributes, comparing within and between relations interest object sets, associating objects displayed, and correlating shared attributes between objects. A visual search environment provides users with an intuitive, interactive, and convenient platform for information retrieval and enriches their search activities. However, it is this interactivity that makes the evaluation of it more complex.

The third factor is dynamic information seeking in the context of a visual environment. Unlike traditional information retrieval systems, information retrieval visualization makes internal objects and relationships among documents/objects transparent to users. A search process in a visual environment, in fact, is a complex decision-making process about information relevance judgment. A search process is a process of information discovery in a dynamic and information-rich visual environment. This sophisticated process may involve users' learning ability, spatial orientation ability, perception ability, and a cognitive aspect as well.

The fourth factor is the diversity of information retrieval visualization tools and models. The diversity reflects dimensionality of a visual space which can be two-dimensional, three-dimensional, or virtual reality; semantic frameworks of information representations which can be a subject directory, neural networks, hierarchy structure, or subject map; projected objects in the visual space which can be documents, Web pages, stack information, information flow, or traffic information in a server; semantic relationships among objects which can be visible hyperlinks, bibliographic citation, or invisible semantic similarities; and ways of illustrating these relationships such as the hyperbolic technique. Each of these variables can make a significant contribution to information retrieval visualization evaluation.

It is clear that some of these factors for evaluating the effectiveness of information retrieval visualization are by nature more subjective and task-oriented. Therefore, it is difficult to find and generalize their characteristics. It is challenging to present an evaluation benchmark system and to define a measurable metrics system to measure them.

11.2.2 Principles for developing evaluation benchmarks

The proposed evaluation benchmarks and criteria should be comprehensive and exhaustive. All of the effectiveness characteristics of information retrieval visualization should be included in such an evaluation system. This ranges from visual information representation, controllability for interactivity between users and visual information systems, to information searching and information browsing. The proposed criteria should be applicable to all data types and tasks of information retrieval visualization models/systems. The criteria should be measurable. In other words, each of the benchmarks and criteria can be managed in terms of measurement. However, in reality, due to the nature of information retrieval visualization, it is extremely difficult to come up measurable criteria for each of the proposed benchmarks.

11.2.3 Four proposed categories for evaluation criteria

Information retrieval can basically be classified into two categories based upon its search nature and purpose. The first is a search for detailed information of a known item. For instance, users search for works of a known author, a full-text of a given title, or patent of a particular patent number. The other is a search for uncertain information within a defined interest topic. In the latter case, searchers know the subject/topic they are looking for but they do not know exactly which concrete items they are looking for. In reality, the majority of users' searches fall in the second category. Unfortunately, traditional information retrieval systems which are built upon a query search mechanism like Boolean based information retrieval systems are more suitable for the first category than the second category. The beauty of information retrieval visualization lies in its capacity for information browsing in a visual information space. Due to its unique 2D or 3D nature of information space, users can engage in information discovery, data mining, and data harvesting by browsing in the visual space. It is totally different from a query search. The process of information browsing, in fact, is also a process of their need clarifications and need definitions. Information retrieval visualization really changes the way that people search for information. It is the interactivity, flexibility, and multi-dimensional nature of a visualization environment that makes visualization information retrieval more competitive when dealing with the second category of searches. In other words, in a visualization environment, users equipped with a variety of interactive control mechanisms can effectively browse information, navigate a visual information space, find relevant information, and discover new information. Therefore, the proposed benchmarks system should include not only evaluation for query search which is an indispensable perspective, but also evaluation from an information browsing perspective.

One of the most prominent characteristics of information retrieval visualization is its visual space. Within a visual space, a semantic framework is presented, visual data/objects are projected onto the framework, logical and semantic relationships in the context of the framework are illustrated, and various interactions are carried out. Visual data, the framework, and the way that visual data is presented within the framework are defined as visual information representation. Visual information representation is fundamental and essential to information retrieval visualization. To a large degree, the visual information representation determines the foundation, features, functionality, and characteristics of information retrieval visualization. That is, whether visual information representation is successful or not would decide the success of information retrieval visualization. The proposed evaluation system should include it due to this reason.

As we know, a visual information environment offers an ideal and intuitive interface for end-users to interact with. The environment is a window that enables users to communicate with systems. It is a place that users exchange information with visual information systems. Through various interactions between users and systems, users may browse information, submit queries, navigate visual space, make information cluster analysis, customize a local information space based upon their interests, drill down to details of an interesting object, and so on. Information retrieval visualization must provide users with control mechanisms to manipulate information, to participate in decision making, and to complete their tasks. Controllability for information retrieval visualization interaction should be considered in the evaluation metrics.

Information browsing, querying, visual information representation, and controllability for interactions are four primary categories within the proposed evaluation system. Querying and information browsing reflect evaluation requirements of information retrieval. Information presentation addresses the way that visual data is organized and presented. It provides a platform for users to control and retrieve information. Controllability for interaction emphasizes interactions between an information retrieval visualization system and its users. They are integrated as a whole and they are dependent upon each other in the visual space.

11.2.4 Descriptions of proposed benchmarks

It is evident that the four categories are too broad to measure and examine information retrieval visualization. But they present a structural framework which can guide people to develop more detailed benchmarks within each of the four categories.

Information browsing

The first criterion within this category is guidance. Users navigate in a two-dimensional or three-dimensional visual space to search for relevant information. Due to the multi-dimensional nature of a visualization environment, users need a guidance mechanism to orient themselves in the visualization environment during navigation. This is similar to a compass for a tourist traveling in an unfamiliar territory. This guidance should not only orient users in a visual space but also lead users to appropriate and desired locations. Some information retrieval visualization systems integrate a subject hierarchy structure to facilitate users browsing and locating information (*STRETCH*, 2005; *SPACETREE*, 2005; Beaudoin et al., 2005). Displaying information about the area surrounding of a focus area would help users to make a decision about the next browsing step. At each of navigation points, providing users with available and appropriate information discovery means and disabling non-appropriate features would decrease possible disorientation for users. Finally, a well-designed and user-friendly help file which includes explanation for all features and functions would be useful for guidance.

The second criterion is exploration. Information retrieval visualization should enable users to overview the entire information space which usually is set as a starting point of navigation. More importantly, a local information space should be generated and presented to users upon request. Detailed information about an object should be provided if that object is selected during browsing. The detail degree of a browsed area should be controlled by users. The local information space should also be easily and smoothly shifted to the entire information space, and jumping from the overview of the visual information presentation to a local view should be allowed. An overview of an entire area, a local view of information space, the control over the detail degree of a browsed area, and detail of a required object are the basic elements of information exploration.

The third criterion is the dimensionality of a displayed object in the visual space. The dimensionality of a displayed object in the visual space refers to the degree to which a displayed object offers information about itself in both depth and width. A displayed object in a visual space is usually an icon which tells users information about the object it represents. The design of an icon should be concise, meaningful, and intuitive. Colors, size, and shapes, or their combinations are employed to represent multiple meanings of represented data. For instance, the size of an icon represents the relevance degree. The shape of an icon can represent the type of an object. The color of an icon can represent the status of an object. In *Visual Net* (Belmont Abbey College North Carolina, 2005), for instance, a holding item icon consists of a circle, several concentric rings, and associated arrows. A red center circle indicates that the holding is for printed material while a blue center circle means that the holding is an electronic material. The thickness of the white concentric ring indicates how large the holding is and the thickness of the green concentric ring shows how new the holding is. When the blue arrow on the outer ring appears, it suggests the holding is in a foreign language, and when a blue arrow on the outer ring appears, it implies the holding is a reference item.

The fourth criterion is connections or relationships of a displayed object to others in a visual space. When an object is displayed in a visual space, it is not isolated or disconnected. In other words, when it is presented in a visual space, its relationships with other objects are also illustrated. What relationships are illustrated and how they are illustrated need to be evaluated. In some systems, the connections are visible and in others the connections are invisible. Some connections may be visible only after users make such a request. Sometimes the relationships are connected by links, like hyperbolic-technique-based visualization systems (*Visual Thesaurus*, 2005; *Inxight*, 2005), adjacent orders (*Map of the Market*, 2005), distances and directions such as *TOFIR* and *DARE*. *VIBE* can use length of connected line between two related objects to represent the connection degree of the two objects (Olsen et al., 1993). In a subject tree structure, sibling relationships and parent-child relationships may be shown.

Querying

A query search feature is indispensable and necessary for information retrieval visualization. This feature distinguishes it from other information visualization models/systems. The method of querying in information retrieval visualization varies in different visual environments and is primarily affected by visual information representation and nature of visualized data.

The first one is a simple query search. Information retrieval visualization should accept a search query formed by search terms. Unlike a traditional information retrieval system, it maps the matched results onto its visual environment and visually illustrates them for users by highlighting them. In the visual context, users can observe results, results distribution, and relationships between the query and retrieved objects. These matched objects are colored differently in the visual space so that users can easily distinguish them from other un-matched objects. Basically, information retrieval visualization in this case does not visualize the internal matching processing and only visualizes the matched results. In some systems (*Visual Thesaurus*, 2005; *Inxight*, 2005), for instance, search query windows are offered, search results are color differently from unmatched objects, and visual presentations are adjusted and regenerated so that search results are emphasized based on new users needs in the visual spaces. In *VIBE* the relevance degree between a query and a result object is presented by using different colors.

This query search mechanism should be embedded into two levels: global and local. The former refers to the idea that querying is carried out within the entire database while the latter refers to the idea that the querying is within a specified local area. The former is a global search while the latter is a local search. The latter is useful when users navigate into a specific local space such as a sub-branch of a subject tree structure or a browsed sub-map area, and they may want to search only within that local area.

The second criterion is the information retrieval model visualization. In a broad sense, information retrieval is not simple keyword matching. It includes using powerful information retrieval models wherein users may control and manipulate the retrieved results. There are various information retrieval models such as the Boolean retrieval model, cosine model, conjunction model, disjunction model, distance model, ellipse model, and so on. Visualizing these information retrieval models in a visual environment is more challenging than just visualizing the results of a search query. That is because visualizing an information retrieval model is, in fact, visualizing internal information retrieval processing. Users can manipulate both the information retrieval process and information retrieval results. This makes both information representation transparent and information retrieval processing transparent to end-users. As we know an ellipse model can determine a hyper-ellipse contour in a high dimensional vector space which can not be observed by users. The contour is invisible in a high dimensional space. The location of the hyper-ellipse contour is determined by two users' information interest points, (User interest point is a broader concept of user query. It can include users' background, reading habits, previous queries, and so on. A query is presented by multiple interest points in an information space). The objects within the contour are regarded as retrieved objects. Users can control the size of the contour to change the size of retrieved objects, or they can change the position of the contour if their interests change. In *DARE*, the ellipse information retrieval model can be visualized as follows. The ellipse contour in the high dimensional space is mapped onto a low two-dimensional space which can be observed by people. After it is converted to the low dimensional space, it no longer preserves its ellipse shape in the high dimensional space. Instead, it becomes a wave-like curve in the visual space. After this conversion, the invisible hyper-ellipse contour in a high dimensional space becomes visible. The most important and exciting aspect of this conversion is that users can control and manipulate the concrete and visible contour to control information retrieval in the low visual space at will. Another example is *Filter/Flow* (Young and Shneiderman, 1993). In *Filter/Flow*, documents in a database are defined as water flow and Boolean logic operators such as logic OR and logic AND are defined as valves to control the water flow (documents). Users can add valves to the water control system to include relevant documents and exclude irrelevant documents. Users can observe the flow change in the visual space.

The difference between the previous scenario and this scenario is that the former only visualizes the final results of a search while the latter visualizes both internal search processing and final search results.

The third criterion of a query search feature is query reformulation. As we know, the information search process is a dynamic one. The information search process may be affected by the degree of information need understanding, familiarity with the information retrieval system, and the searchers' background and experiences. For these reasons, a multiple-step search is needed to adjust the search strategy and make the search more accurate. In other words, users need to reformulate their queries based upon initial search results. It is necessary that information retrieval visualization provides users with a feedback mechanism to adjust search queries. Some systems such as *DARE* and *GUIDO* allow users to pick up any documents or their combinations in a visual space to replace a current query, or add them to a current query, or revise them. *DARE* allows users to shift the role of the involved reference points to change the retrieval emphasis. In most multiple user interest point based environments, the content of a user interest point can be dynamically changed or redefined based upon one's needs.

Visual information representation

Visual information representation is essential for information retrieval visualization. It is the foundation of information retrieval visualization. Within this category there are seven criteria concerning information retrieval visualization.

The first criterion is the dimensionality of the visual space. Visual information space can be two-dimensional, three-dimensional, or virtual reality. Users definitely behave differently in a 2D environment versus a 3D environment (Sebrechts et al., 1999). It is believed that a three-dimensional visual space offers an extra dimension to represent more information. Adding an extra dimension to a two dimensional space is not as easy as "2 + 1 = 3". The impact of the joined third dimension on information retrieval visualization may be larger than people imagine. Because of an additional dimension, presented information may be richer, illustrated semantic relationships among objects may be more complicated and sophisticated, visual information representation may be more intuitive and natural, and presented information may be more informative. On the other hand, adding an additional dimension to the visual space would increase technical difficulties when systems are implemented and also the operational complexity for users.

The second criterion of visual information representation is the semantic attribute revelation. Semantic attribute revelation defines the visual space to some degree. It is evident that an object can have multiple attributes while these attributes define characteristics of the object. In a visual environment, not all attributes of an object are identified and utilized to represent that object. Useful, meaningful, salient, and necessary attributes are selected and preserved while others may be sacrificed and excluded in the construction of a visual environment. The identification and revelation of object attributes has a significant and direct impact on visual information representation. Selected attributes may be assigned to the *X*-axis, *Y*-axis, or *Z*-axis of a visual space respectively. Attributes can also be expressed in other different ways. These selected attributes lay a foundation for their visual frameworks. For instance, both distance and direction attributes of an object in *DARE*, direction attributes in *TOFIR*, distance attributes in *GUIDO,* hierarchy attributes in *CHEOPS*, similarity ratio in *VIBE*, and time attribute and subject attributes in *GRIDL*(GRaphical Interface for Digital Libraries, 2005) are identified and represented in their visual spaces. In the two dimensional *GRIDL* space, attributes in the visual space can even be redefined and replaced by a pool of attributes such as classification, publishing years, author, title, physical location, classification and conference place.

The third one is the semantic framework of the visual space. A semantic framework is usually associated with the revealed attributes from objects. A semantic framework, where objects are projected onto, defines the structure of a visual space. Semantic frameworks range from a grid, hierarchy, map, network, to circle, triangle, rectangle, etc. A semantic frame should be meaningful in terms of information retrieval, concise in terms of structure, and aesthetic in terms of visual presentation such as symmetry.

The fourth criterion is the intuitiveness of visual information representation. Intuitiveness includes both easy information expression and easy understanding of visual information. The visual information presentation should be expressed in a straight forward manner so that users can easily adapt to the environment. Unfortunately, in reality, due to the complexity of a database, when certain attribute characteristics of data in the database must be preserved and presented and high dimensionality must be reduced, it is difficult to find a simple and intuitive way for the visual information representation. Without a doubt, users prefer an intuitive visual information representation and are more comfortable and willing to interact with an intuitive interface. Researchers and designers of information visualization have been searching for appropriate and applicable metaphors which may be embedded into the semantic frameworks of visual information representation. Familiar concepts, objects, or environments from the real world would facilitate users to understand the visual information representation, decrease users' learning time for the systems, reduce users' anxiousness, and therefore increase effectiveness and efficiency. It is easy to find the systems which employs metaphors, for instance, a water fluid metaphor(*Filter/Flow*), a solar system metaphor (*WebStar* (Zhang and Nguyen, 2005)), geographic map metaphor (*WebMap*, 2003), Fisheye (Fisheye menu, 2005), and so on.

The fifth aspect of visual information representation is clustering and categorizing. As we know, a displayed object in a visual environment is not isolated or semantically independent of other displayed objects. The displayed objects are semantically connected and associated in some sense. Object location in the context of a semantic framework implies and indicates something. Objects projected onto a close neighborhood suggest that they share similar characteristics because they are projected onto the same spot according to the same projection algorithm. This phenomenon can help users to make a clustering and categorizing analysis in visual environments. This analysis can answer such questions as: How many objects are grouped in a cluster? What are the relationships among different object clusters? Basically, attributes identified and employed from an object to construct a visual space decide the nature of the clustering and categorizing. Clustering and categorizing can be used to support search feedback, perform object similarity analysis, understand overview distribution of documents in a database, and other purposes. For instance, all objects are clustered as a group in *DARE* if they share similar distances and angles against defined interest points. Semantically relevant objects are clustered and related subjects are adjacent in a semantic map.

The sixth criterion is visual information representation customization. Information space for a database should illustrate all data perspectives. However, users' interests usually concentrate on a limited topics/subjects compared to entire coverage of a database. During a search users may change their topics/subjects. It suggests that information retrieval visualization should support both an overview of entire information space and also a customized local view of interests. Upon request from users, it should offer a detailed and customized local view. It is clear that a local view based upon users' interests is dynamic. It varies in users and even varies in different steps of a search for the same user. This visualization information representation customization is different from simple zoom in/out feature in an interface. Views generated by a series of zoom in/out operation preserve hierarchy relationships while visualization information representation customization does not necessarily follow the same principle. In some situations where a high dimensional information space is converted to a low dimensional visual space, visual information representation is more complicated. The same local area in a high dimensional space may correspond to multiple visual presentations which emphasize different perspectives of the local area.

The last criterion within this category is the disambiguation mechanism. Ambiguity is a unique phenomenon of information visualization. Ambiguity happens when a high dimensional information space is converted to a low dimensional visual space. Ambiguity refers to the idea that objects that are located in different places in a high dimensional information space are projected onto the same spot in a low visual information space. It is apparent that projection ambiguity can mislead users because objects that are located in different places in a high dimensional information space should be projected onto different spots in a low visual information space. Notice that mathematical projection ambiguity is inevitable when a space with a high dimensionality is reduced to a space with a low dimensionality. When data is processed and projected onto a visual space in a certain way, the data must be customized, some attributes are preserved, some attributes are eliminated, and some attributes are altered or "distorted" after projection. The point is that if this happens, information retrieval visualization should provide a disambiguation mechanism to solve the problem. For example, in *DARE, TOFIR, GUIDO*, and *VIBE,* which are built on a vector-based document space, a spot in visual space can correspond to multiple documents which may be far away from each other in the vector space. Revising user interest point(s), repositioning the affected user interest points, or adding/discarding interest points in visual space can effectively disambiguate the phenomena in these systems.

Controllability

The first criterion of controllability is the ability to zoom in/out. The original zooming definition refers to the metaphorical operations of a camera that can scan across a scene, move in for a closer observation, or back away to get a wider view. The concept is incorporated into information visualization to allow for exploration of information at both specific level and general level. Toward this aim, all of the display data should be organized and categorized in terms of the detail degree. Users should be able to zoom in/out on interest areas or objects at will. Narrower, more detailed and specific information becomes available as users zoom in. Broader and more general information becomes available as users zoom out.

When zooming, it is important to keep the zooming path and global context. This helps users avoid the possibility of disorientation and improves zoom operation controllability. The way to zoom in/out and the zoom detail level should also be considered.

The second criterion of controllability is the activity history preservation and display. Unlike a traditional information retrieval system where interactions with the system are relatively simple, interactions with information retrieval visualization are richer, more diverse and complicated. They range from query search, navigation, browsing, disambiguation, to visual representation customization. All conducted interaction activities with information retrieval visualization should be preserved in some way such as a reverse order. Upon request from users, previous activities should be traced back to allow for replay or a revisit. This is necessary because information exploration process in a visual space, sometimes, is a process of trial and error. Users make a correct decision or reach satisfactory results by trying out various means or features until mistakes are sufficiently reduced or minimized. Activity history preservation and display would reduce the users' burden of recalling all past activities.

The third criterion is filtering. When users navigate into a local area of interest, a visual information representation is customized, search results are displayed in an area, or an object cluster is observed in a visual environment, users may be interested in the visual contexts and some of the objects in the contexts. In other words, some inappropriate or unwanted "noise" should be filtered out while the contexts are kept. For example, certain types of objects, objects within certain time periods, objects with certain attributes, etc. are filtered from the context. In fact, filtering processing is the process of data refinement.

The fourth criterion is selection. Selection includes selecting an interest object and an interest area in a visual space. Selection is important for users to navigate in the visual space. Selection enables users to examine a focus object or area, investigate the content of a focus object or area, distinguish possible overlapping objects, and make relevance decision about the involved objects. After objects are selected, the detailed information about the selected objects should be demonstrated and associated operations should be provided.

All detailed evaluation criteria or benchmarks discussed above are summarized as the follows:

- *Information browsing*
  - *Guidance*
  - *Exploration*
  - *Dimensionality of a displayed object*
  - *Connections of a displayed object to others*
- *Querying*
  - *Simple search and visual display of search results*
  - *Information retrieval model visualization*
  - *Query reformulation*
- *Visual information representation*
  - *Dimensionality of a visual space*

Semantic attribute revelation
Semantic framework
Intuitiveness of visual information representation
Clustering/categorizing
Visual information representation customization
Disambiguation mechanism

Controllability

Zooming in/out
Activity history preservation and display
Filtering
Selection

Regarding retrieval result evaluation, recall, precision, and other criteria used for individual result evaluation at the micro-level can still be used in retrieval result evaluation for information retrieval visualization. Without a doubt browsing in a visualization environment would definitely make a positive contribution to retrieving relevant individual objects. On the other hand browsing requires not only efforts but also time. Both two retrieved results and time factors should be considered in measurement of browsing. Therefore, the ratio of the number of retrieved relevant objects/documents to the time sent in browsing in a visual space can be used to measure the quality of browsing.

11.3 Summary

Attention to information visualization has increased significantly. More and more research information retrieval visualization models, pilot systems, and commercial applications are available. However, there are still a limited number of studies regarding information visualization evaluation, let alone information retrieval visualization evaluation. There are no widely accepted, reliable, accurate, effective evaluation benchmarks, evaluation criteria, or metrics systems available to test the effectiveness and efficiency of information retrieval visualization. What are the metrics and benchmarks suitable for information retrieval visualization? From which perspectives can the quality of a visual information retrieval be measured? How can similar visual information retrieval tools/models be compared? Researchers, developer, designers, or users of information retrieval visualization want the answers to these questions.

A benchmark system and evaluation criteria for information retrieval visualization are presented. Affecting factors from both information retrieval and information visualization are considered in the system. This evaluation standard consists of four categories: information browsing, querying, visual information representation, and controllability for interactivity. Each of the four categories emphasizes a different perspective of information retrieval visualization. Both information browsing and querying reflect information retrieval fundamental natures and characteristics, visual information representation considers visual space characteristics, the essential part of information retrieval visualization, and finally, controllability for interactivity addresses indispensable interaction between users and an information retrieval visualization environment. It is clear that both information browsing and querying are more associated with tasks, visual information representation is more related to data, and controllability for interactivity is more connected to users. These four categories are dependent upon and affect each other.

When people conduct an experience study to examine and test an information retrieval visualization system, they should be aware of the prototype effect problem. Since novel information retrieval visualization models are usually first introduced in the form of a proof-of-theory prototype, examining and testing such a system may bring a concern caused by the fact that its interface may be not user-friendly, the system may not be not robust because of undetected glitches or bugs, and features may be immature and incomplete due to the nature of a prototype system. It may have an unexpected impact on experimental study results. People should be especially cautious when an information retrieval visualization prototype system is compared with a commercial information retrieval visualization system, or when an information retrieval visualization prototype system is compared with a commercial traditional information retrieval visualization system.