InfoCrystal:

A visual tool for information retrieval & management

Anselm Spoerri

Center for Educational Computing Initiatives
Massachusetts Institute of Technology
Building E40-370, 1 Amherst Street, Cambridge, MA 02139
aspoerri@athena.mit.edu

ABSTRACT

This demonstration introduces the **InfoCrystal™** that can be used both as a **visualization tool** and a **visual query language** to help users search for information. The InfoCrystal visualizes all the possible binary as well as continuous relationships among N concepts. Users can assign relevance weights to the concepts and set a threshold to select relationships of interest. The InfoCrystal allows users to specify **Boolean** as well as **vector-space** queries graphically. Arbitrarily complex queries can be created by using the InfoCrystals as building blocks and organizing them in a hierarchical structure. The InfoCrystal enables users to explore and filter information in a flexible, dynamic and interactive way.

Keywords: Information visualization, visual query language, information retrieval, graphical user interface.

INTRODUCTION

Information is becoming available in ever growing quantities as the access possibilities to it proliferate. However, better methods are needed to filter the potentially unlimited influx of information [3]. Towards that end we have developed the InfoCrystal™ that has the following functionality: 1) Users can *explore* an information space along several dimensions simultaneously without having to abandon their sense of overview. 2) Users can *manipulate* the information by *creating useful abstractions*. 3) Similar to a spreadsheet, users can ask *"what-if"* questions and observe the effects without having to change the framework of a query. 4) Users receive *support* in the search process because they receive *dynamic visual feedback* on how to proceed. 5) Users can formulate queries *graphically*, and they have *flexibility* in terms of the particular methods used to retrieve the information.

INFOCRYSTAL

How can all the possible relationships among several search criteria be visualized in a two-dimensional display ? Venn diagrams can be used to visualize set relationships by intersecting geometric shapes. However, it is difficult to represent all the possible relationships among more than three concepts in a visually compact and simple way. Figure 1 shows how a Venn diagram can be transformed into an InfoCrystal. This transformation can be generalized so that all the possible relationships among N variables can be represented at same time in an elegant way. Users can use their visual reasoning skills to establish how the interior icons are related to the border icons. The following visual coding principles are used in a redundant way: shape, proximity, rank, color / texture, orientation and size coding.

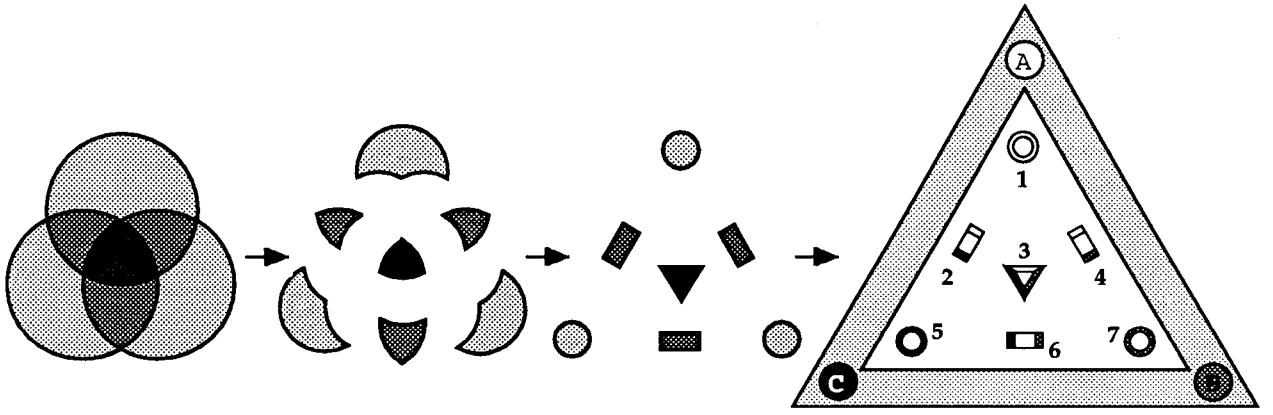

Figure 1: shows how we can transform a Venn diagram into an iconic display, called the *InfoCrystal*, which represents all the possible Boolean queries involving its inputs in disjunctive normal form. The interior icons have the following Boolean meanings: 1 = (A and (not (B or C), 2 = (A and C and (not B), 3 = (A and B and C), 4 = (A and B and (not C), 5 = (C and (not (A or B)), 6 = (B and C and (not A), 7 = (B and (not (A or C)).

Visual Query Language

The InfoCrystal has the desirable property that each of its interior icons represents a distinct Boolean relationship among the input criteria (see Figure 1). Hence, the InfoCrystal can be used by users to specify Boolean queries

by interacting with a direct manipulation interface. Users do not have to use logical operators and parentheses explicitly. Instead they need to recognize the relationships of interest and select them. An InfoCrystal partitions the space defined by its N inputs into $2^N - 1$ disjoint subsets or *constituents*. Any Boolean query can be represented by the union of a certain number of the constituents (i.e., the InfoCrystal represents all the possible queries in *disjunctive normal form*). Existing visual query languages allow users to formulate specific queries, whereas the proposed visual query language enables users to formulate a whole range of related queries by creating a single InfoCrystal [1, 5].

Creating Complex Queries

The InfoCrystals can be used as building blocks and organized in a hierarchical structure to create complex Boolean queries. First, an InfoCrystal has several inputs, which are represented by the border icons, and its output is defined by the selected interior icons. Second, the output of an InfoCrystal will be one of the inputs to an InfoCrystal one level up in the query hierarchy. Similar to a spreadsheet, users can ask "what-if" questions by changing which interior icons are selected in one InfoCrystal and observe how the contents of the dependent icons higher up in the query hierarchy change dynamically.

Interfacing with the Retrieval Engines

The atom or "leaf" nodes of the query structure represent the criteria that a retrieval engine will use to search for information in the selected database(s). A key feature of the InfoCrystal is its flexibility: 1) It works for any data type, provided its corresponding retrieval method returns unique data identifiers. 2) Any retrieval method can be used to initialize the query structure.

Relevance Weights & Thresholds

Users can interact with sliders to assign relevance weights between -1 and 1 to the inputs of an InfoCrystal to reflect their degree of importance. The assigned weights can be used to compute a relevance score for each interior icon. By setting a threshold, users can select only the interior icons whose relevance score is above the threshold.

Bull's-Eye Layout

Users can either choose a layout of the interior icons that ensures that the number of criteria satisfied increases towards the center of an InfoCrystal, or they can display the interior icons to reflect the current setting of the relevance weights. The latter mapping, called the *bull's-eye layout*, causes the relationships with a higher relevance score to be placed closer to the center. We use a novel polar representation to determine the placement of the interior icons.

Visualizing Vector-Space Queries

The InfoCrystal can be generalized to visualize the continuous case where documents can have a value between -1 and 1 to reflect the degree to which a criterion is (not) satisfied. This allows us to specify vector-space queries graphically, since the vector-space approach computes the relevance score of a document by taking the dot-product of the vectors of the index terms that represent the query and the document respectively [2]. We can apply the bull's-eye layout principle to visualize how the retrieved documents satisfy the input criteria to varying degrees. Documents with high relevance scores are displayed closer to the center of an InfoCrystal. The polar transform used to map the documents has the attractive feature that it not only visualizes the ranking of the relevance scores, but it also provides users with a qualitative sense of how the ranked documents are related to the input criteria. Hence, the proximity principle is preserved by the polar transform.

Query Outliner

This tool enables users to incrementally structure complex search queries by creating a hierarchical query structure. The *query outliner* has a similar functionality as the familiar outlining tool available in word-processing packages. The query outliner solves the problem of how users can easily annotate and summarize in a word or two what the nodes in a query hierarchy represent.

CONCLUSION

The *InfoCrystal™* can be used both as a *visualization tool* and a *visual query language*. It can visualize all the possible binary as well as continuous relationships among N concepts. In the binary case, the InfoCrystal uses proximity, rank, shape, color and size coding to enable users to see in a single view how a large information space relates to their interests. In the continuous case, a novel polar representation is used to visualize the relevance scores of the retrieved documents. Users can assign *relevance weights* to the concepts and formulate *weighted queries* by setting a *threshold*. Further, users can decide to visualize an InfoCrystal in such a way that the relationship satisfying the most criteria, called the *rank layout,* or the one with the largest relevance score, called the *bull's-eye layout*, will lie in its center, respectively. The InfoCrystal allows users to specify *Boolean* as well as *vector-space* queries graphically. Complex queries can be created by using the InfoCrystals as building blocks and organizing them in a hierarchical structure. Readers who require a more detailed discussion of the InfoCrystal should please consult [4].

Acknowledgments: Support for the research described in this paper has been generously provided by the UBILAB of the Union Bank of Switzerland (UBS). We would also like to thank the MIT Patent & Licensing Office.

REFERENCES

1. **Anick, P.; Brennan, J.; Flynn, R.; Hanssen, D.; Alvey, B. & Robbins, J.** (1990) "A Direct Manipulation Interface for Boolean Information Retrieval via Natural Language Query," *Proc. ACM SIGIR '90.*
2. **Belkin, N. & Croft, B.** (1992) "Information Filtering and Information Retrieval: Two Sides of the Same Coin" *Comm. of the ACM*, Dec., 1992.
3. **Card, S.; Robertson, G. & Mackinlay, J.** (1991) "The Information Visualizer, an information workspace," *Proc. CHI'91 Human Factors in Comp. Systems*, 1991.
4. **Spoerri, A.** (1993) "InfoCrystal: a visual tool for information retrieval & management," *Proc. ACM Information & Knowledge Management*, 1993.
5. **Young, D. & Shneiderman, B.** (1993) "A Graphical Filter/Flow Representation of Boolean Queries: A Prototype Implementation and Evaluation, " *Journal of American Society for Information Science*, 44:6, 1993.