HYPOSTATIZING DATA COLLECTIONS, ESPECIALLY BIBLIOGRAPHIC ABSTRACTIONS, REPRESENTATIONS, SENSUALIZATIONS, ADAPTATIONS/PERSOINALIZATIONS, . . .

Roland Hjerppe

In one word, Leibnitz intellectualized phenomena, just as Locke, in his system of noology (if I may be allowed to make use of such expressions), sensualized the conceptions of the understanding, that is to say, declared them to be nothing more than empirical or abstract conceptions of reflection. Instead of seeking in the understanding and sensibility two different sources of representations, which, however, can present us with objective judgments of things only in conjunction, each of these great men recognized but one of these faculties, which, in their opinion, applied immediately to things in themselves, the other having no duty but that of confusing or arranging the representations of the former.

The Critique of Pure Reason by Immanuel Kant;
http://www.knuten.liu.se/~bjoch509/works/kant/cr_pure_reason.txt

The universe is information and we are stationary in it, not three-dimensional and not in space or time. The information fed to us we hypostatize into the phenomenal world.

Philip K. Dick ("Valis")

ABSTRACT

Data, originating in abstractions of realities, and denoting selected aspects, at specific times and places, for particular purposes, are for storage and/or transfer made manifest through representations. Some of the representations also serve as sensualizations (presentations) or vice
versa. The use of computers is, however, based on a separation of representation from presentation, necessitating and enabling intentional designs of each, separately or conjointly. Every sensualization of a collection of data, be it as a list or as multimodal interaction, involves hypostatizations—i.e., treating or regarding concepts, ideas, etc. as distinct substances or realities. Such sensualizations, hitherto usually static visual presentations, have been based on, and employed, metaphors, with Cartesian analytic geometry as the dominant metaphraser. The sensualization of collections of bibliographic data for access to documents, although facilitated by a range of new tools and techniques, is impeded both by existing legacy meta-data structures that grew out of a focus on individual static documents and by the lack of methodologies for the generation and description of collections (documents or subjects) and their dynamics.

INTRODUCTION

The theme for this 34th Annual Clinic on Library Applications of Data Processing, Visualizing Subject Access for 21st Century Information Resources, is equivocal in several ways. There are at least two different interpretations possible for the first part—i.e., Imagining/Envisioning Subject Access ..., and Visualization (as a means) for Subject Access. I have chosen the second interpretation as the basis for this discussion. Any presentation that appeals to our visual sense is of course a visualization. Visualization is, however, customarily interpreted as a range of techniques and/or methods for alternative visual presentation of data sets. Alternative indicates that, instead of presenting a "raw" data set (usually numbers, as such—in the traditional form of tables or graphs—something else), an image or animation is presented. There has, naturally, to be some kind of correspondence between the data set and that which is presented.

Visualization is, however, just one of the forms possible for alternative presentation. Instead of visualization, we could use auralization or tactilization. We could in fact appeal to any or several or all of our senses. There are categories of people for whom presentation types other than visual are important or even essential—e.g., people with visual impairments.

Generalizing alternative presentation of information or data to involve any, or several, of our senses leads to the term "sensualization." Sensualization would then be the process of making something available to our senses that was not so before. Presentation switches, which are not generally available today, will probably be used in the near future to select the type of presentation—e.g., for a text whether it should be aural (by voice) or tactile (Braille) or visual (on screen/paper). Hence the use of the word sensualization in the title of this presentation.
HYPOSTATIZATION?

Definitions of the word *hypostatization* can be found in dictionaries, on the World Wide Web—e.g., Webster or Merriam-Webster:

*Webster (Hypertext Webster Interface, http://gs213.sp.cs.cmu.edu/prog/webster)*
definition for “hypostatize”
hy*pos*ta*tize \hi-'pa:s-t*-.ti-z\ vt [Gk hypostatos substantially existing, fr. hyphistasthai]

: to construe [a conceptual entity] as a real existent : REIFY

*Merriam-Webster Dictionary (http://www.m-w.com/netdict.htm)*
Main Entry: hy•pos•ta•tize
Pronunciation: hl-'pas-t&-”tlz
Function: transitive verb
Inflected Form(s): -tized; -tiz•ing
Etymology: Greek hypostatos substantially existing, from hyphistasthai
Date: 1829

: to attribute real identity to (a concept)
—hy•pos•ta•ti•za•tion /-"päst&t&-'zA-sh&n/ noun

Hypostatization is thus the treatment or view of a concept or idea as a distinct substance or reality, as something existing and having an identity. Reification—to regard something abstract as a material thing—is a close synonym but with different connotations. Whereas HotBot, AltaVista, and Infoseek returned (in February 1997) only ninety-six, eighty-nine, and seventy-seven matches for the search term “hypostatization,” there are 1,777, 1,724, and 1,032 matches returned for the term “reification.” The term reification has a number of more specific uses; it is, for example, used in object orientation as the reverse of abstraction.

The deeper issues of what is real, the nature of reality, and so on, which are discussed in the branch of philosophy called ontology, will not be delved into here. The position behind the discussion that follows is one of what might be called phenomenology-based naive realism—i.e., what we experience gives us direct access to reality.

HYPOSTATIZATION, FALLACIES, AND METAPHORS

In logic, the term “fallacy” is generally used for a form of technically incorrect argument—especially if the argument appears valid or convincing. A fallacy is thus a logical argument which looks correct, but which can be seen to be incorrect when examined more carefully. Overviews of the most common types of fallacies in argumentation and logic are available on the World Wide Web.

Hypostatization is thus the fallacy of taking something which has only a conceptual or imaginary existence and treating it as if it were a simple con-
crete reality. Hypostatizations are common. They occur all the time but are very rarely deliberate or intentional. Typical examples are utterances like: "History teaches us that...." "Science knows...." "Medicine has found...."

Hypostatization is, however, not necessarily always an error, wrong, or undesirable. Philosophy, theology, and literary criticism are some of the areas in which the term hypostatization is used, not necessarily as a verdict of a fallacy, but, for example, as an indication of a reification. In Christian theology, the term hypostases are used for the three manifestations of God—the Trinity.

Hypostatization can be regarded as a natural consequence of abstraction and is thus also closely related to metaphor as well as to cathexis—i.e., the investment of emotional significance in an activity, object, or idea. Why the concern with hypostatization? Simply put, visualization/sensualization entails hypostatizations.

**HOW?**

A data set is, or can be regarded as, a representation of a reality originating in an abstraction of that reality. It denotes selected aspects, at specific times and places, for particular purposes. People engaged in conceptual modeling—e.g., for databases—are continuously reminded of this. Data sets are for storage and/or transfer made manifest through representations. Some of the representations also serve as sensualizations (presentations) or vice versa. Sensualizations are likewise (indirect) representations of data.

**DATA AND COMPUTERS**

The use of computers is, however, based on a separation of representation from presentation necessitating and enabling intentional designs of each separately or conjointly. The letter "a" has in the computer a specific representation of zeroes and ones, but on the screen the letter can have many different forms—e.g., glyphs which are separate from the internal representation. Another representation is then, of course, needed to indicate which of the glyphs to use in each case.

Every sensualization of a collection of data—be it as a list, table, or as multimodal interaction—involves hypostatizations—treating or regarding concepts, ideas, and so on as distinct substances or realities in order to make these presentations.

**MULTI-MODAL SENSUALIZATION**

Sensualization does not have to be uni-modal—i.e., involving only one of the senses. Multi-modal sensualization involves several of the senses, either presenting the same data to several senses or different (related) data to each sense.
VIRTUAL REALITIES

The ultimate in sensualization is probably virtual realities. The virtual realities available today are still far from realistic, but that does not necessarily have to be the case. What is important is whether they function for the intended purpose, whether the "suspension of disbelief" by the user is accomplished. In virtual reality applications that are intended to visualize data, "all around you the dance of biz, information interacting, data made flesh in the mazes of the black market" (from Neuromancer, my addition of italics), we thus get an inversion—i.e., we go from data as representation of reality to reality as representation of data.

TRADING REALITIES

Sensualization involves a trade: realities (invented/design) are used as/for representations of realities (other). We start with a reality from which we get the data, then we take that data and use it to build another reality. One reality is traded for another to emphasize some aspects and to minimize or eliminate all others. In other words, we get metaphors built on hypostatization.

WHY SHOULD WE SENSUALIZE?

Sensualization is unavoidable: the indented list is a visualization as is layout and typography, paragraphs and sentences, punctuation, etc.; the analog thermometer is a visualization. The question is therefore not why we should do sensualization but rather when should we do what kind of sensualization?

INFORMATION HIDING

Why should we, as professional librarians and information specialists, be concerned with sensualization? The answer to that follows from the assertions that most databases and information systems are advanced exercises in information hiding and, hence, almost all OPACs are opaque.

INFORMATION REVELATION

A real library is designed for semi-direct access, exploration and inspection, being in the collection, assisting in building a (mental) model of the collection. A library is thus transparent, revealing to the user in its midst its local content, and to the explorer its structure and design. That does not, however, imply that the proficient use of all libraries is self-evident.
TOOLS NEEDED

What we need are tools that enable us to approach information resources the same way we use libraries and other things in our everyday life—i.e., using our motor-perceptual capabilities and our experience of being in the world.

We have, all of us, a very long training period—up to twenty years—of being in the world and our society before we start making use of it. We need a very long training period. It is the most comprehensive education everyone gets, and it is adapted to, and develops, our motor-perceptual capabilities. How can we make use of that? The obvious claim here is that sensualization is one of the ways to capitalize on these capabilities that have been developed.

INTERACTION: SENSUALIZATION—MANIPULATION

Presentation (sensualization) is, nevertheless, not enough. Users also want to act/react—in order to interact. We need to relate the representations and the modes of presentation to the modes of action and interaction. The modes of interaction are as important as the modes of sensualization.

WHAT DO WE HAVE?

The sensualizations and interactions we (can) provide are partly dependent on the data sets available and their characteristics.

THE BASES AND THEIR SHADOWS

The bases, the starting points, for all the deliberations at this clinic are collections of objects. In our case, the primary objects are documents of various kinds; these include not only the physical books and serials but also the documents on the Internet.

One of the things we do in libraries for each of the documents is to also generate a thin silhouette, a representation, a catalog record. Each collection thus has a shadow collection of representations.

THE LIGHT SOURCE

Shadows are cast on a background by a light source that illuminates the objects. The characteristics of the light source, its vantage point in relation to the objects, and the background determine the appearance of the shadow.

Unfortunately, very little of an object’s features, revealed by the illumination, except the outline, remains in the shadow. Very little of the book remains in the catalog record.
WHAT CAN WE DO WITH A SHADOW COLLECTION?

What can we then do with a collection of representations? The things that can be done with the book collections can be seen in any library.

An Example—TopLib

TopLib was a demonstration system we did in 1990 to enable simple interactive visualizations of the contents of a library catalog or, rather, the distributions for some of the data in the records in a catalog database. A brief description of TopLib is available in Hjerrpe’s (1992) “Database Visualization.”

TopLib Data

The data we pulled out of the catalog, with the assistance of the library, were: ISBN/ISSN, type of document, branch library code, classification code (SAB), language code, country of publication code, year of publication, and year of acquisition. These are variables that have values that can easily be counted and that have a limited number of values—implying that they can be fit onto an axis, and which are meaningful to a librarian. The number of different author names or words in a title are too numerous to be useful in the kind of presentation we wanted to provide.

TopLib Visualizations

The basic metaphor we used was a landscape seen from above and to the side; a topographic presentation in which height, the z-axis, always corresponds to frequency of occurrence, and the variables of the other two axes, x and y, depend on the interactive selections by the user from the variables available.

The interesting part of this project is that when we showed the system to the librarians, they could see both their conceptions of the libraries verified and anomalies of which they were not aware. They would say: “Yes, that is how it is.” and “I wonder about that thing over there. How can it be explained?”

One of the most interesting things was that until then they had not thought about the collection in this multi-dimensional manner and that the tools were unavailable that provided an overview and the ability to analyze the collection easily from different points of view and zoom in on details.

We included the ISBNs so that it should, in principle, be possible to zoom in on individual items and identify them. In practice, one would have needed to link from the ISBN to the library catalog to get the full details, and this was not implemented in the demonstration system. The point is that, with simple data extraction and simple visualization techniques, it is possible to provide powerful tools for the analysis of collections.
WHAT ELSE DO WE HAVE?

In addition to the shadow collection, we also have various SKOs (Systems for Knowledge Organization), and Diane Vizine-Goetz will discuss this in her presentation. Others will talk about SKOs as mirrored in collections—i.e., the instantiations of SKOs in collections and collections as seen through the lenses of SKOs. We also have structures, explicit and implicit, links/relations, citations, URLs and structures generated by/from the links/relations.

The people working with citation analysis, bibliometrics, scientometrics, and so on are, of course, making use of these structures and relations. So far, little has been done with these approaches considering the wealth of material on the Internet although proposals have been made.

Structures provide a natural base for visualization through the generation of graphs. Some of the hypertext systems—NoteCards, gIBIS, StorySpace, and Intermedia—had the capabilities to generate graphs and maps of the structures of a specific hypertext document, and the user could set several parameters for the generation of these graphs.

NoteCards Browser Examples

SKOs usually have a (more or less) hierarchical structure—sometimes a forest of small linked trees—and could thus be treated as a special case of structures for specific visualization purposes. One of the problems is that if the hypertext document is large then the overview will not fit on the screen. Some systems provide zoom-in and zoom-out operations and/or scrolling windows, but scrolling windows are a nuisance, and an overview of a large network becomes cluttered and unreadable. One of the proposals to solve this problem is a dynamically changing fish-eye view of the structure. The details of the network are enlarged only in the immediate neighborhood of the node that is currently displayed. The remainder of the network is filtered in some way to show only the important details. The lack of insight as to how to determine what is important for the overall structure is the main reason why most systems do not provide a fish-eye view.

In discussing structures, it is important to make a distinction between micro and macro structures; in our case, for example, micro-structures being those internal to a document and macro-structures those of collections of documents.

IS THERE ANYTHING LEFT?

Is there anything left in terms of the data sets available to us for visualization? So far we have not considered time. The discussion has implicitly been focused on static data sets.
Time

Time is a dimension for everything we do, and all data have more or less explicit time attributes; they are created at specific times and pertain to situations at points or intervals in time. We could thus consider sensualizing. Status at a specific time or the dynamics of evolution of a collection of documents, status at a specific time or the dynamics of evolution of an SKO, status at a specific time or the dynamics of evolution of structures, status at a specific time or the dynamics of evolution of... 

EVOLUTION, STATIC PRESENTATION

Some very simple examples (first done in 1991 and updated in 1993) of a static presentation of an evolution of a collection (of collections) with respect to specific terms (multi-media or multimedia) follow.

Example 1. The evolution of the occurrence of the terms “multimedia” or “multi-media” in the set of databases available through Dialog. The databases are ordered according to their numbering by Dialog.

Example 2. Same data as in Example 1 but with a logarithmic frequency scale to focus on the higher frequencies.

Example 3. Same data as in Example 1 but with the databases ordered by frequency of occurrence. Note the “exponential” distributions in time and productivity.

Example 4. Same data as in Example 1 but with the databases ordered by frequency of occurrence and with logarithmic frequency scale.

EVOLUTION, DYNAMIC PRESENTATION

A dynamic visualization is akin to an animation or a movie. A dynamic sensualization would hence be akin to an experience in the corresponding sensory domain, in the general multi-modal case to an experience of a reality—a virtual reality.

Space

Space is likewise a dimension for everything we do and all of our data also have more or less explicit associated spatial location attributes. They are created at specific times and pertain to objects at places or regions in space.

Space is perhaps the most neglected of all dimensions in our area. Only recently have the approaches inherent in, and the capabilities of, Geographical Information Systems (GIS) been considered.
Space-Time

The travels through space-time of each physical book and each collection forms a unique trajectory that intersects with other space-time trajectories. Although SKOs also evolve, it is only the aggregate and individual manifestations in physical objects that have been assigned classes/terms that travel through space-time.

Associated/Related Data

Space and time are, when they are not explicitly available, also examples of associated/related data—data that we could get and/or attach to the data we already have using some common denominator.

Some of that associated/related data that is available concerns the use and the users. There is also data available on the creation/production/distribution and the creators/producers/distributors, and such data are already employed in, for example, scientometrics.

Meta-data

Finally, we have meta-data—data about the data—data structures, data about generation/production, and so on of data, ownership, etc. We can, of course, also have meta-meta-data and so on. I want to stress here that very little has been done with respect to the use of meta-data in our domain.

Users need meta-data to use data properly. We thus need to consider things such as how to present data structures. On the Internet, we see a renewed and growing interest in meta-data to get handles on the sprawl.

ADAPTATIONS AND PERSONALIZATIONS

Each approach to sensualization has to be adapted to the collections at hand and, conversely, each collection has also to be adapted to the approach chosen. There will therefore always be a need for adaptations of various kinds.

On the Internet, we are no longer bound to the exact presentation determined by someone (an author/editor) prior to our reception of the material. We are ourselves, to some extent, in control of the presentations generated by our Internet browsers. We have a choice both of what to experience next (within the boundaries of what is available) and of some of the visual aspects such as the type face and its size.

Personalization will become more common. Sensualizations can be pre-determined, or under the control of the user, or usually a mixture. Too much personalization leads to communication problems. If I have one type of sensualization and the person I’m talking with has or has had another type, then we will have difficulties in agreeing on what we are talking about.
INTERNET, LIBRARY CATALOGS, AND COLLECTION DESCRIPTION

Through the Internet, we can access a large number of library catalogs, but currently we do not provide any tools to users that could assist them in determining which library or which catalog to use for a specific purpose.

How do we compare and sensualize the differences of, for example, the holdings of a large number of libraries? How do we do it in a way that is relevant to “ordinary” users? The sensualization of collections of bibliographic and related data is impeded by existing legacy meta-data structures that grew out of a focus on individual static documents and the lack of methodologies for the generation and description of collections (of documents, or SKOs, or structures, etc.) and their dynamics.

We have AACR2 for the description of documents and SKOs for the description of their contents—what they are about—but we have nothing for describing collections, what they are about, and their evolution.

Nothing is available today that could assist users in determining which collections to use for a specific purpose except very vague and general verbal statements.

One of the main differences between our collections of documents and the actual documents is that “a library is a growing organism”—collections evolve and they are dynamic. That is, of course, the case also with the documents on the Internet. The documents we have handled until now have been static, and so we have been content with providing static descriptions.

CONCLUSION

How can we combine what we have with sensualizations? Many tools and methodologies are available. Much research is ongoing, here and in many other places, and we will hear about that during the rest of this conference. There is much that can be done and will be, but the important questions (to me) are: What is useful? for whom? How do we determine that? How do we get it done? Do we do it? Will we have it done to us? What is needed in terms of systems? in terms of change of focus? How do we deliver it to the end user?

REFERENCES