Using a Landscape Metaphor to Represent a Corpus of Documents

Matthew Chalmers

Rank Xerox EuroPARC, 61 Regent St., Cambridge CB2 1AB, U.K.
chalmers@europarc.xerox.com

Abstract. In information retrieval, sets of documents are stored and categorised in order to allow for search and retrieval. The complexity of the basic information is high, with representations involving thousands of dimensions. Traditional interaction techniques therefore hide much of the complexity and structure of the modelled information, and offer access of the information by means of isolated queries and word searches. Bead is a system which takes a complementary approach, as it builds and displays an approximate model of the document corpus in the form of a map or landscape constructed from the patterns of similarity and dissimilarity of the documents making up the corpus. In this paper, emphasis is given to the influences on and principles behind the design of the landscape model and the abandonment of a 'point cloud' model used in an earlier version of the system, rather than the more mathematical aspects of model construction.

1 Introduction

Bead is a prototype system for the graphically–based exploration of information. The underlying notion of the system is one of our most familiar metaphors: spatial proximity to represent similarity in some more abstract interpretive framework [8]. We represent the relationships between articles in a bibliography by their relative spatial positions. We attempt to place similar articles close to one another and dissimilar ones further apart. The emergent structure is a model of the corpus: a landscape or map of the information within the document set. This 3D scene can be used to visualise patterns in the high–dimensional information space. The aim is to make interaction with a database of information more graphically–oriented, and to move away from interaction styles requiring knowledge of query languages and the database material itself. This allows people to move from cognitive problem–solving to more natural sensorimotor strategies, and to support more exploratory and cumulative modes of use.

The modelling techniques involved were the main focus of a description of an earlier version of Bead, published as [4]. That paper describes the kind of numerical techniques used and their combination with text analysis techniques to make 3D 'point clouds' of graphical data. An example point cloud is shown in Figure 1. This type of graphical structure is common in statistical graphics [5]. The raw material was a small

In Proc. European Conf. on Spatial Information Theory, Elba, September 1993, published as *Spatial Information Theory*, A. Frank & I. Caspari (eds.), Springer Verlag LNCS 716, pp. 377–390, 1993.

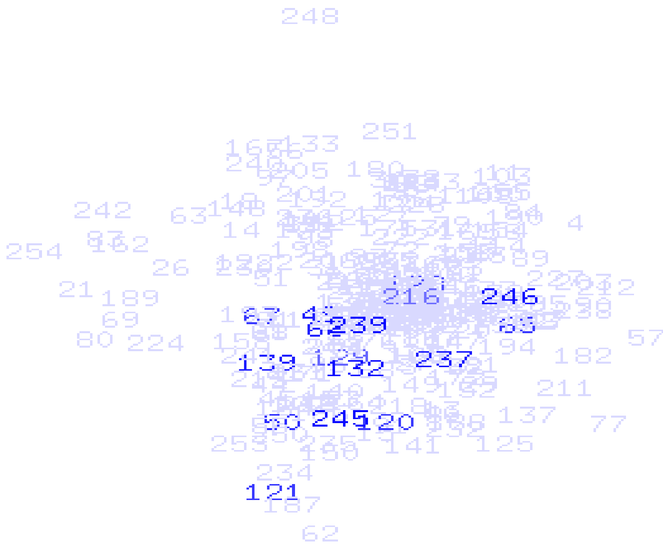

Figure 1. A point cloud constructed from articles in a human-computer interaction conference, CHI91. A search on the words ‘information retrieval’ leads to some of the document numbers being highlighted. Documents mostly lie in a central cluster, but documents matching the search are mixed throughout the corpus. This is more apparent in an animated display, but occlusion and overall scene complexity still inhibit corpus comprehension.

bibliography of articles on HCI. Since then, the interface style has progressed to the use of a more accessible graphical structure, and the issues underlying this progress are the subject of this paper.

Low dimensionality is in accord with our everyday experience. We are used to a space of three physical dimensions wherein we perceive individual characteristics of objects and also their patterns and interrelationships. Given our lives on the surface of the earth, our experience is of a world with greater extent in the horizontal than the vertical: one might even call our everyday world ‘2.1–dimensional’. Physical spaces of high dimensionality are unfamiliar to most of us, and it is generally more difficult to present, perceive and remember patterns and structures within them. If our activities depend on judgements based on both the individual characteristics of documents and their relative properties, then we will gain by employing a representation which shows both in a space of familiar dimensionality.

In initial experimentation within EuroPARC this was found to be a problem of the ‘point cloud’ representation which, even with only three dimensions, did not easily afford overview of the entire set of documents. Occlusion of distant information, the lack of a fundamental ground plane, and weakness with regard to other bases of everyday perception meant that it was difficult to become familiar with a significantly–sized corpus. The features that make graphical display most effective were not being brought to bear in making the information design effective and useful. Users found it difficult to orient themselves and navigate within the space, and consequently did not build up a useful mental model of the corpus. Instead they found occasional items of interesting data but had difficulty in assessing their relevance or significance in the wider context of the corpus. This seemed to echo a point forcefully put in [1]:

> Items of data do not supply the information necessary for decision–making. What must be seen are the relationships which emerge from consideration of the entire set of data. *In decision–making the useful information is drawn from the overall relationships of the entire set*. [Author's italics]

A goal of Bead is to represent a corpus of documents in a way which helps with tasks which rely on the relationships of the entire set of documents as much as the properties of individual members. A document must therefore play a dual role: it must act as an autonomous unit and it must also play a role as a component within a higher–level structure. We wish the corpus to have a layout which represents the more abstract patterns within it. The model of a corpus used in Bead emphasises patterns of thematic similarity as estimated by similarities in word usage. Individual documents take their places in these patterns by dint of the words used within them, but they also have important characteristics not used in constructing the overall layout. Labelling Bead as an information retrieval system is to take a narrow view. In a wider, more general sense, it is a system intended to aid in tasks which rely on consideration of the overall relationships of the themes and words active in a corpus, as well as the individual elements.

There is a point of view which holds that most information retrieval tasks are better performed if the overall relationships within a corpus are presented. As this view has spread, a more general model of the use of information systems than that traditionally used in Information Retrieval has arisen. The word 'retrieval' suggests an action by some agent to find and bring back information to a somehow detached or uninvolved user. The agent is given a specification of what is wanted, and the user waits for the results to be returned. However, it is becoming more accepted that people may not be able to express what they want to access in a corpus in a query language or even a natural language, for that matter, because they may not know exactly what they are looking for. Instead, they may be able to do what they want to do if they can begin by finding out roughly what is available in the corpus and, in an exploratory manner, refine and adapt their enquiries.

An individual document may initially appear 'relevant' but later be discarded if other documents better serve the interests of the user. There may be a great number of relevant documents, or there may be none: both are fair and on occasion appropriate results. There may be documents that are relevant in different ways, and this may lead to a continuation and adaptation of work as these different associations are assessed. Initially known documents may be dispersed among other unknown but potentially relevant ones. These examples stress the relative judgements that drive an adaptive and exploratory style of use that contrasts with and complements the less interactive ‘retrieval’ approach. This approach has been labelled information *access*, and to some extent reflects the increasing awareness of the importance of exploration and dynamism in perception and model construction [7, 14]. These ideas have also been influential upon the designers of other information interfaces, e.g. the work of the User Interface Research and Intelligent Information Access groups at PARC [3, 10, 12, 13]. The following section goes slightly further into these and other areas of work that have been influential in attempts to make the design of Bead better suited to information access.

2 Influences and Comparisons

In this section, some of the issues, techniques and concepts behind the Bead system are presented and discussed. From the basic infrastructure of information retrieval techniques to influential examples and metaphors, a wide range of areas of study have had their effect on the choice of information displayed and the structure of the display: collectively, the *information design*.

2.1 Information Retrieval

As in most Information Retrieval systems[1], a document is essentially represented as a list of the words which occur within it, and some numeric measure of the relative frequency of each occurring word. The most simple measure is a ‘one’ for a word occurring in a document, and a ‘zero’ for non–occurrence. For reasons which include efficiency, not every word is actually used: common ‘noise words’ such as ‘the’ and ‘or’ are discarded.

This representation has a geometric interpretation. It is as though each document is a point in space. The large number of words means a large number of dimensions to this space: each unique word in a corpus of documents defines one dimension, and the frequency of occurrence for each word is a coordinate for that dimension. Word frequencies can be further weighted to take account of extra information. ‘Noise words’ are effectively given zero weight. In the case of the bibliographic data used in Bead, words in the ‘keywords’ and ‘title’ sections of a bibliography entry are given more weight than those from the abstract. Documents are close in this type of high-dimensional space if they have roughly the same words occurring with roughly the same weights. The assumption — generic and fundamental to all of Information Retrieval — is that if this is so then the documents are most likely to be similar in themes and topics, i.e. spatial distance corresponds to some degree with thematic similarity.

Since the number of words in a corpus of documents can easily run into thousands, the number of dimensions is too large to display directly. We can get some information

1. [15] is one introduction to the basics of Information Retrieval techniques and systems.

about what documents are close to each other by the more traditional means of information retrieval: in effect we get a list of the documents near to one or more documents known *a priori*. The system can then find the nearest (most relevant) documents and return them to the user, usually in order of spatial distance (relevance ranking). A simple view of a query to an information retrieval system is that we make a fake, temporary document with just the keywords we give and with chosen (presumably high) weights, find where that document is placed in the high–dimensional space, and then return a list of nearby documents. Again, we get an idea about documents close to the item of information we placed into the space, but we obtain little global information. We find out about the locality near to the query, but we gain no idea about documents further away (involving different words) and how they relate to each other.

Note that we have to start off with some clear, *a priori* knowledge of what we want to find out. Unfortunately this is not so often the case, for example when we want to explore and browse our way towards whatever it is we come to decide is of interest or relevance. We can only do this by many repeated samples of localised regions, and in ourselves building up some sort of cognitive model of what there is in the corpus and how they fit together. Defining such samples means either having many well–understood documents in the corpus from which one can work outwards, or knowing how to put the words together in a query (i.e. how to use a query language) to sample in the region you want to know more about. It would be better for those who do such sampling less often (or who know less about the corpus or the query language) if some of the cognitive load could be taken on by the representation of the corpus.

Another approach using the same underlying representation is to partition the space into some number of regions which share roughly the same words and weights. We can then show representative members or the highly weighted words which typify each region. We therefore obtain a concise but more approximate representation which we can show in a list. We now get some idea of the overall range of the documents in the corpus, but our accuracy is limited because we can only write so much about each one on a screen or page. We could then choose one or two of these selected regions to look at more closely, perhaps by gathering in the members of the regions and then trying to spread them out again into another list of regions to select from, scatter out again and so on with ever more refined choices of documents. This is a rough description of the Scatter/Gather technique described in [6].

This approach is better for someone browsing the corpus because at each stage they get some idea of the overall contents of the corpus presented to them, and they need not know a query language. They choose one or more relevant members of a list, and in so doing move themselves closer to their goal. By reducing dependence on initial knowledge of keywords and query languages, Scatter/Gather is intended to favour information access more than information retrieval.

The model of the document corpus in Bead is similar to Scatter/Gather in its focus on the perception of the global structures and relationships of documents rather than the techniques associated with a hidden, high–dimensional representation. The model of the individual document is one which creates these larger–scale aggregate structures. In Bead, each individual document has as the dominant factor in its behaviour the resolution of the similarities and differences with all other documents. By making visible the setting of each document within the larger–scale structure of the corpus, browsing and exploration is aided. For this to work, the layout must impart to the user relationships and structures which 'make sense' and yet it must also let users orient themselves, navigate and examine the corpus. In other words, Bead should make the corpus 'imageable'.

2.2 Design of Complex Spatial Structures

This term is drawn from a work on the theory of a certain type of complex spatial structure of varied use and interpretation, namely the city. In *The Image of the City* [9], Lynch described properties of cities which led to people being able to orient themselves and navigate within the city so as to carry out tasks related to the spatial structure, e.g. finding the way to some location or sketching out a rough city map. Lynch carried out surveys of people performing such tasks, and collected and analysed the results to come up with some ideas about the characteristics needed for imageability. These included: landmarks visible from most of the region, which allow for orientation; a delineated border (and perhaps borders of subregions) which serve as reference points; viewpoints, so that parts of the city region can be overviewed before travelling into the more local neighbourhoods; a skeleton of routes into and through the region which one can use to go to particular local areas and from which one's knowledge of local detail can grow and be fleshed out; and consistency of local texture (e.g. building styles) so that one can determine from a street's local detail information about the subregion one is in.

Another study of the interrelationship of the design of spatial structures with ease of use is published as [2]. This study looked at situations where the design of public buildings was such that signposting was necessary to overcome the difficulties users had in use and navigation. He suggests that "in many cases signposting is an admission of design failure. It reflects the fact that there are many situations in which the designer cannot rely upon the knowledge or experience of the user in finding the way" and in assessing various buildings in use, "those buildings with many short, winding corridors came out worst, whereas... the building which consists essentially of a large, open, partly differentiated space came out the best." In another situation, an airport terminal, Canter noted how the linear sequence of functions a traveller undergoes (check-in, customs, boarding, etc.) allowed a building design which was easily understood and used. He found that when the pattern of use of the building was simple in structure then the building could conform to this pattern, but for more varied or general use it was best for the structure to be largely open and visible, allowing users to gain an overview and plan their routes so as to suit their tasks.

Romedi Passini is one author who has noted how planning and wayfinding can be a dynamic activity, and many aspects of architectural design do influence our ability to find our way through buildings and landscapes [11]. While problems of this sort may often be broken down into subgoals, various types of information available may be used to progress towards a desired goal. In the simplest situations, sensory information is enough to guide us towards our goal, e.g. what we seek is directly visible. In slightly more complex situations, some degree of memory use is needed in combination with sensory information in order to achieve a goal. In the most complex situations, the sensory and memory information need to be used and manipulated to create new information and achieve the goal. As memory and cognitive load increases, the task is made more difficult. All three levels may interact in complex and dynamic ways, but if we can design the framework wherein goals and activities take place so that more reliance is made on sensory than memory and manipulated information, then we will make the information tasks of users less difficult.

In many situations, however, we cannot determine our goal *a priori*, and we must rely on our skills in the dynamic interpretation of information. We continually orient ourselves, comparing expected and visible information, and so reassess our plans and expectations as we move towards a final goal. Another significant point is the danger of information overload, where information is presented in such volume and with such low discrimination of importance that one is swamped, and cannot find one's way through to a final goal. Even though the pertinent information may be available, it is lost amongst the clutter of other useless or distracting information.

2.3 Information Design and Perception

This latter point of visual clutter is one of the main issues put forward by Edward Tufte in [16]. Too many information designs are counter–productive in that attempts to enrich the information presented to a user lead to distracting or confusing results. In the case of the point cloud used in the first version of Bead, it was often the case that the complexity of the 3D patterns meant that the data cluttered itself, and even small amounts of detail presented at each data point led to a great morass of visual clutter. To improve the information design, such clutter should be reduced or tailored so that one is more selective in the way that one displays detail as one examines different areas of an information display.

In the studies of perception, there is a significant body of work that explores the dynamism that is an essential and everyday characteristic of our behaviour [7]. Gibson argues that direct perception, memory and interpretation act together as we explore our environment, build up our mental models, and plan uses and actions within that environment. We resolve visual ambiguities, direct our attentional focus, add areas of the environment to our memory, and perceive shapes and uses all in relation to our body and its movement.

When we look at one area close up, the detail there has as its context the neighbouring regions which extend on out to the boundary or horizon. Continuity of movement over a landscape, coupled with perspective viewing, allows one to incrementally refine one's attentional focus down to more local areas, while smoothly adding more information to the context or periphery of view. We select areas to examine in detail by moving closer, while distant areas are seen in a less detailed (or more abstract) way. There is continuity between the close and detailed, and the far and abstract. Reference is continually made to the ground plane of the landscape, as perceived by the variations in size and texture of objects as well as larger-scale features such as the horizon and directionality of light. Although areas of the environment become hidden as we move, we are experienced in maintaining a mental model which lets us return to or otherwise use such areas.

Note that in the case of point clouds and other 'strongly 3D' structure, the environment is more complexly structured with many occlusions and obstructions of view. Without references such as a horizon and a consistent ground plane, information gained by overview and exploration is more difficult to come by. Our skills in perception mental model–making, as honed on our everyday '2.1D' world, become more difficult to employ.

In this section the fundamental data to be accessed has been described, along with the basic notion of using a landscape metaphor for data representation, and some issues relevant to the design and perception of such landscapes have been presented. The next section describes some ways in which these different threads have been, or are being, woven together to improve the information design.

3 Improving the Information Design

In coming to advance the design of Bead's information display, the collective effect of the issues pointed out in the preceding section suggested a move away from strongly 3D structures and towards map-like (2D) or landscape-like ('2.1D') structures. The decision was made to sacrifice the greater exactitude of relative distances that can be gained in full 3D, and the modelling process of Bead was changed accordingly. An attempt is under way to use the greater accessibility and familiarity of the landscape metaphor to display a corpus of documents. A grayscale snapshot from the colour display is shown in Figure 2.

In the type of visualisation system common to 3D graphics, viewpoints are easily available since one can place one's eye arbitrarily in space so as to get an overview of any region of interest. Similarly, one need not really have 'routes' in or through the corpus, as one is free to move above the landscape. In trying to introduce other useful characteristics, though, the structure and appearance of the modelled data must serve to provide such features as landmarks and borders, and the positions of documents must be done so that it is apparent that there is local consistency of themes and topics. Individ-

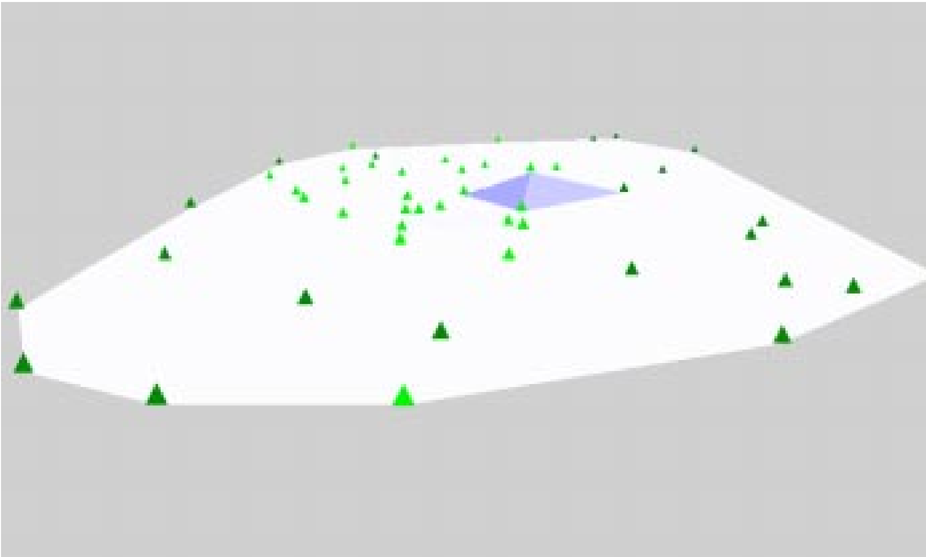

Figure 2. An overview of a landscape constructed from articles in an HCI conference, CHI91. Documents matching the word ‘interface’ are marked in light green, while the remainder are black. Matching documents mostly lie in a central cluster, befitting the frequency of use and significance of this theme in the corpus. Documents less central in this regard lie more on the periphery of the ‘island’ of documents.

ual documents are shown as coloured markers placed within the setting of the landscape and they consequently produce collective patterns of density and locality.

The open landscape affords an overview of the patterns and structures of the corpus. Perspective viewing heightens the visual effect, emphasising nearer documents but also fitting in with our everyday framework of vision and perception. The lack of more literal (and constraining) ‘ways in’ such as routes or other surface features, however, suggests that more abstract and flexible means should be provided. This also means that one should be aware of some of the variety of means by which people initially approach such bodies of information.

A small and informal pilot study was performed at EuroPARC as part of an initial assessment of the information design. Two years’ articles of CHI were laid out using three different sets of layout rules, making six layouts in all. For the purposes of the study, layouts were printed out on large sheets of paper, with titles and keywords shown for all articles. The polygons and shading of the animated window-based interface — the standard tool for interaction in Bead — were left out of the printed ‘maps’.

The first layout rule used only the keywords of the articles, and the consequent patterns produced separated, dense clusters of sometimes naively-grouped documents, e.g. the various senses of a relatively common word such as ‘design’ would cause all related documents to be close together. The second rule used abstracts of articles as well as keywords, and so disambiguated some word uses but led to relatively uniformly dispersed layout patterns with few clusters or gaps. The last layout rule was close to random positioning, although a few pairwise distances were set exactly, so as to offer some local instances of ‘sense’. These layouts were presented in turn to each of seven EuroPARC researchers who were familiar with the journal involved. They were asked to describe the qualities (or lack of them) in the layouts, and to rank the layouts in order of overall preference.

In the absence of a genuine task or ‘way into’ the information, general features such as dense clusters, exceptionally close pairs and the overall centre of the layout were initially examined. Other initial activities included the location of familiar or relevant articles and scanning for consistency of documents in randomly-chosen subareas of the layout.

Overall, the keyword-based layouts were most favoured, while abstract-based and near-random layouts were less favoured to a roughly equal degree. Keyword-based layouts were said to be best because the clumps and gaps served as reference points to access the layout, to return to areas previously examined, and to assess distances between clumps of related documents. While the random layouts had some of this useful ‘texture’, the overall sense of document positioning was weak. Overall sense of positioning and disambiguation of word usage was better in the abstract-based layouts, but the lack of reference features (such as clumps and gaps) meant that the subjects found it difficult to put together an overall model of the corpus.

Apart from physical features such as patterns of local density, word searching provides perhaps the most significant ‘way in’ to the landscape of documents, affording access to its basic patterns and themes, and providing the initial experiences which support later browsing and exploration. When a search for a keyword is made in the animated Bead interface, matching documents have their colour changed (see Figure 3). The resulting patterns of colour show the distribution of matching documents in the corpus of documents. The patterns show how discriminating the search is and how matching documents are distributed throughout the corpus. The latter point tends to show the areas where there are different uses or aspects of the words searched for.

The landscape model means that one can immediately obtain an overview of the entire set which is the basis for such judgements. More information such as the title and keyword list are shown when an individual document (matching or non–matching) is selected with the mouse.

From such initial searches and selections, one gains basic knowledge which one can use to browse and explore other nearby documents. These might be close to those previously inspected but may not have matched one’s initial search. One can find documents which are relevant but do not contain familar keywords, and one can move beyond one’s initial search by using interesting words found as one explores. One can

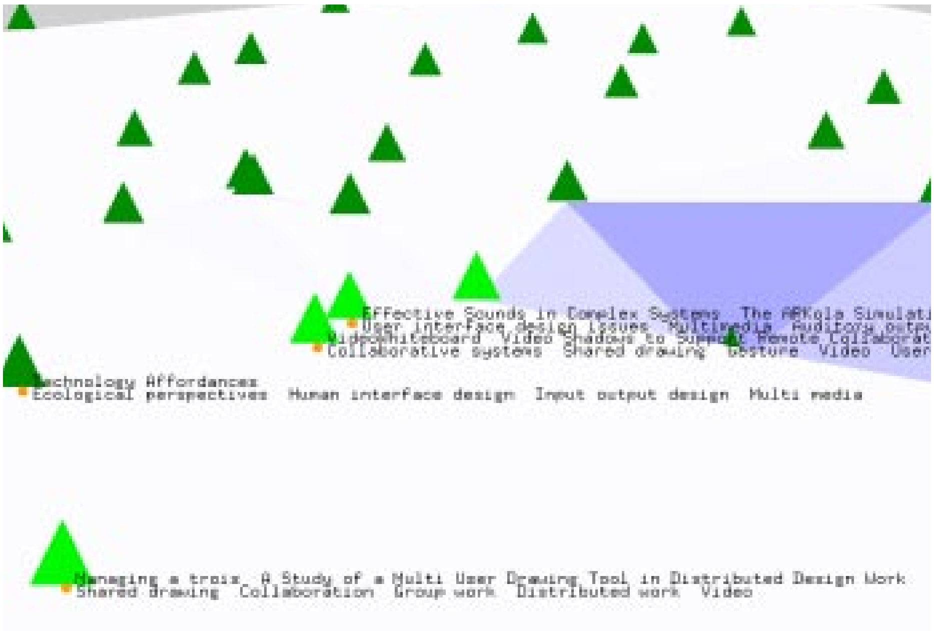

Figure 3. After a search for the word ‘collaborative’, some documents near to the ‘valley’ region at the centre of the corpus have been highlighted. Browsing with the mouse reveals the title of a nonmatching document (Technology Affordances) that is near to one of the matching papers (Effective Sounds in Complex Systems) and may be of interest as the two papers have more general issues in common as well as an author, Bill Gaver.

shift smoothly from browsing to searching and back again, all the while maintaining continuity and context by reference to the static landscape framework.

The Bead ‘Viewer’ program displays a scene showing a corpus landscape, and commands for movement (e.g. zoom in) and searching can be invoked either by typing in commands or by using a set of buttons arrayed above the display. The same Viewer commands can also be invoked via messages sent from other concurrently executing programs. (System architecture issues are discussed in more detail in [4].) Movement, viewing and mouse selection, along with simple searches for word occurrences, make up the means of access to the model. This small set seems straightforward to use; the limitations on use are pushed over more towards the model itself. Given a good corpus model, it should not be necessary to be an expert in query languages or on the database itself — the ideal is that this load should be taken on by the system.

The modelling process of Bead sometimes leads to peaks and valleys, and there is always a surrounding contour or ‘shore’. These serve as natural reference points (or, more literally, landmarks) and are important in orientation and navigation. Slopes are shaded in accordance with their gradient, so that steep slopes are darker then gentler slopes and flat areas. The peaks and valleys also show areas where the 'fit' of the documents is rough. This may mean that the system could not find a good layout in 2D, although such a layout may exist. Alternatively, it could be that there is an inherent conflict in the set of desired mutual distances of documents, and therefore no planar layout can be found. In either case, the roughness of fit is an important property which should be conveyed to the user. This type of metalevel information is not often shown in an information display, although it does convey an extra dimension of the modelled information which could be influential in judgements based on proximity and implied similarity of documents.

The shore delimits the corpus and is usually made up of documents which are less strongly associated with any central theme (or themes) of the corpus. In the modelling process, such documents are pushed out to the physical periphery, which adds to the consonance of the model. Lastly, areas of density show clusters of strongly related documents, serve as landmarks and reference points, and offer bases for initial exploration of the corpus model.

Although more realistic shading and texturing could make the landscape more 'naturalistic', such colours and textures might be better put towards informational content rather than the 'framework' of the landscaping. Indeed, there is a danger that adding such detail might produce distracting visual clutter, obscuring information content. In all information design, one must choose which dimensions of the display should be used for information content (e.g. the position and colour of individual documents), which for a framework which conveys the basic style of interpretation of the information (e.g the shading of the slopes of the hills and valleys), and which should be unused so as to avoid clutter. Adding more information to the landscape design in this way is one of the topics of current work discussed in the next section.

4 Conclusion

The system described attempts to give access to a body of complex, high–dimensional information by using a model or metaphor of a landscape. The display design is directed towards a more exploratory and dynamic style of use than that of most traditional information retrieval systems, and it tries to take advantage of our natural sensorimotor skills by presenting a corpus as a mostly open landscape. This model supports overview and browsing, as well as searches for keywords represented by the colouring of individual documents. It is intended that the dynamics of movement, selection and searching can be combined in a way which favours information access rather than information retrieval.

Current work focuses on the modelling algorithms which construct the landscape, and how low-level issues such as word weighting affect the accessibility and quality of the resulting model. Apart from pairwise properties such as relative distance of documents, the emergent properties of the layout such as patterns of density and overall 'texture' are significant features in the perception and use of the modelled corpus. A related topic is the extension of the construction techniques to more robustly and efficiently handle larger–sized corpora. Other lines of interest are directed towards the enrichment of the information content of the display while avoiding excessive and counter–productive visual clutter, making the 'information space' able to be shared by people working on different workstations, and improving the tools for movement and navigation within the space.

Ideally, the modelled information space should reflect the documents and document–related activities of the people using the system. It should help to guide and should be guided by their work with other systems and in other media. In this way, Bead might offer new possibilities for computer–based support of both individual and collaborative work in a semantically rich virtual environment, and might take a useful place amongst the more general environment of everyday work.

5 Acknowledgements

Thanks are due to my fellow researchers at EuroPARC and PARC for discussion and comment, in particular Bob Anderson, Victoria Bellotti, Bill Gaver, Gifford Louie, Diane McKerlie, Tom Moran, Abi Sellen and Andreas Weigend.

This is an expanded (by ~50%) version of a paper titled 'Visualisation of Complex Information' submitted for presentation at the East-West Human-Computer Interaction Conference, to be held in Moscow in August 1993.

6 References

1. J. Bertin: Graphics and Graphic Information Processing, Walter de Gruyter, Berlin, 1981.
2. D. Canter: Wayfinding and Signposting: Penance or Prosthesis? In: Proc. NATO Conf. on Visual Presentation of Information. Published as: Information Design, Easterby & Zwaga (eds.), Wiley, 1984, pp. 245–264.
3. S.K. Card, G.G. Robertson & J.D. MacKinlay: The Information Visualizer, an Information Workspace. Proc. CHI'91 (New Orleans, Louisiana, 28 April – 2 May, 1991), ACM, New York, pp. 181–188.
4. M. Chalmers & P. Chitson, Bead: Explorations in Information Visualisation. In: Proc. SIGIR'92, published as a special issue of SIGIR Forum, ACM Press, pp. 330–337, June 1992.
5. W.S. Cleveland & M.E. McGill (eds.): Dynamic Graphics for Statistics, Wadsworth & Brooks/Cole statistics/probability series, Belmont, CA, 1988.
6. D.R. Cutting, J.O. Pedersen, D. Karger & J.W. Tukey: Scatter/Gather: A Cluster–Based Approach to Browsing Large Document Collections. Proc. SIGIR'91, published as a special issue of SIGIR Forum, ACM Press, pp. 318–329.
7. J.J. Gibson: The Ecological Approach to Visual Perception, Lawrence Erlbaum, 1979.
8. G. Lakoff & M. Johnson: Metaphors We Live By, University of Chicago Press, 1980.
9. K. Lynch: The Image of the City, MIT Press, 1960.
10. J.D. MacKinlay, G.G. Robertson & S.K. Card: The Perspective Wall: Detail and Context Smoothly Integrated. Proc. CHI'91 (New Orleans, Louisiana, 28 April – 2 May, 1991), ACM, New York, pp. 173–180.

11. R. Passini: Wayfinding in Architecture, Van Nostrand Reinhold, New York, 1992.
12. R. Rao, S.K. Card, H.D. Jellinek, J.D. MacKinlay & G. Robertson: The Information Grid: A Framework for Information Retrieval and Retrieval–Centred Applications. Proc. UIST'92 (Monterey, California, November 1992), ACM, New York, pp. 23–32.
13. G.G. Robertson, J.D. MacKinlay & S.K. Card: Cone Trees: Animated 3D Visualizations of Hierarchical Information. Proc. CHI'91 (New Orleans, Louisiana, 28 April – 2 May, 1991), ACM, New York, pp. 189–194.
14. D.M. Russell, M.J. Stefik, P. Pirolli & S.K. Card: The Cost Structure of Sensemaking. Proc. InterCHI'93 (Amsterdam, April 93), ACM, New York, pp. 269–276.
15. G. Salton: Automatic Text Processing, Addison–Wesley 1989.
16. E. Tufte: Envisioning Information, Graphics Press, 1990.