EUROGRAPHICS 2002 / I. Navazo Alvaro and Ph. Slusallek
(Guest Editors)

Short Presentations

Multi-dimensional and Multi-scale Visualizer of Large XML Documents

C. Jacquemin and M. Jardino

LIMSI-CNRS and University of Paris 11, BP 133, 91403 Orsay Cedex, France
{Christian.Jacquemin,Michele.Jardino}@limsi.fr

Abstract

3D-XV (3-Dimensional XML Visualizer) is an interactive graphical interface for accessing large structured documents. It relies on a geometrical model that combines a sequential organization and a hierarchical structure. In the overview mode, the user can access information through natural language queries. In the browsing mode, the content of the paragraphs is displayed in a moving walkway fashion. The model is compared with previous studies on document visualization for passage retrieval. The merits of interactive 3D manipulation through the intuitively oriented 3D-XV geometrical model are outlined.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [User Interfaces]: Graphical user interfaces (GUI)

1. Interactive GUI for XML-Encoded Documents

Many information access and visualization activities are concerned with small and possibly interconnected documents such as networks of Web pages or large collections of journal articles. Such documents only represent a subset of the available on-line textual data: they do not encompass large and structured documents such as technical manuals, scientific reports, or administrative regulations. Due to their size and their internal structure, these types of documents require specific visualization interfaces that take into account their hierarchical organization and provide efficient means for spotting and accessing relevant passages.

This work presents *3D-XV* (3-Dimensional XML Visualizer), an interactive Graphical User Interface (GUI) for visualizing and accessing large XML documents. We chose to focus on XML as framework for describing text documents because of its increasing use and its universality. XML[3] is now becoming a widespread standard for encoding various types of semi-structured data including text, multi- or hyper-media documents. Because of its flexibility, XML can be used as a pivot representation of structured (hyper-)documents, and can then be exported to various proprietary formats. The data model of *3D-XV* relies on the tree-structured organization of XML documents. We use the ISO 12083 Book Document Type Description (DTD)[1] which provides a large range of structures for the description of books, manuals, theses, and corporate documentation. In *3D-XV*, XML documents are converted into a geometrical model that combines the sequentiality of the text and the hierarchical structure of parts, chapters, and sections.

We first recall previous studies on large document visualization combined with clustering and dynamic queries (Section 2). Then the geometrical model of *3D-XV*is presented in Section 3. In order to facilitate long distance similarity detection, the geometrical model is combined with unsupervised passage clustering for highlighting thematic similarities through block coloring (Section 4). Interactive natural language querying and small-scale passage visualization allow for fine information access (Section 5). Last, in Section 6, we refer to previous studies and compare the relative merits of 2D and 3D representations for large structured document visualization.

2. Visual Interfaces for Passage-Based Information Access

When accessing large documents, it is necessary to divide the text into small chunks of information, called *passages*. These passages then serve as basic units for information display. For query-based document access, the *TileBars*[4] interface associates each document with a passage relevance indicator. Each passage is represented by a column of gray squares, the darkness of the *n*th square denotes the relevance of the *n*th query term for this passage. Most relevant passages correspond to columns of dark squares, and most relevant documents to rows of dark columns.

An alternative to querying is clustering that consists in building classes of documents based on a measure of similarity. The measures used for calculating the relevance of a passage with respect to a user query in the preceding example can also be used for computing inter-passage similarity. Salton et al.[9] propose a circular representation of documents in which passages are materialized by arcs and passage similarities by lines.

The two aforementioned techniques for graphical passage-based information access cannot scale to large documents. In *TileBars* it would result in an unbrowsable list of passage relevances; in Salton's work, it would lead to an unmanageable tangle of short or long distance passage similarities. In contrast, *3D-XV* combines a two-scale approach to document browsing for easier drill-down selection, with fisheye visualization for cognitively lighter browsing of large data. The passage similarity dimension is reified by block-color similarities. The relevance associated with dynamic querying is materialized by histogram-like representations of passages.

At the highest scale level, fisheye distortion is obtained from a curvature of the document, combined with a perspective view of the curve's sharpest segment very similar in mind to the *Perspective Wall*[6]. At lower scale, the same document curve is used for text distortion, and results in lower text definition for out-of-focus passages as in the *Table Lens*[8]. In order to avoid displaying unreadable reduced fonts, the full-text content of contextual passages is replaced by salient text extracts obtained from linguistically-motivated text filtering.

3. Geometrical Model of *3D-XV*

Because of its structure and its subdivision into thematically coherent passages, an XML document is represented in *3D-XV* as a set of geometrical blocks (frustums) organized along two dimensions representing the reading order and the hierarchical structure:

- **Linearly**: the blocks are placed in the same order as they appear in the document. The length of a block is proportional to the number of words of the passage it stands for.
- **Hierarchically**: the top blocks represent passages while lower layers represent deeper structures: subsections, sections, and chapters. As for passages, geometrical length is proportional to text length.

Figure 1 represents two sequential chapters C_1 and C_2. The second chapter is made of three subsections $S_{2.1}$, $S_{2.2}$ and $S_{2.3}$ located on top of it.

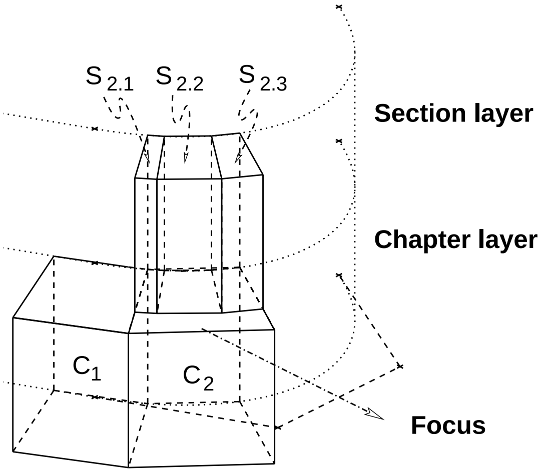


Figure 1: *The geometrical model of 3D-XV.*

The passages are linearly organized along a 3-piece curve that produces a fisheye effect through perspective viewing. The curve is made of two half-lines D_0 and D_1 (the previous and following contexts) and a Bezier curve C (the focus zone). The tangents at the extreme points of the curve are aligned with the lines in order to ensure geometric continuity. The sharpness of the curve is controlled by three parameters *D*, *d*, and *h* that define the locations of its four control points $\{K_0, K_1, K_2, K_3\}$.

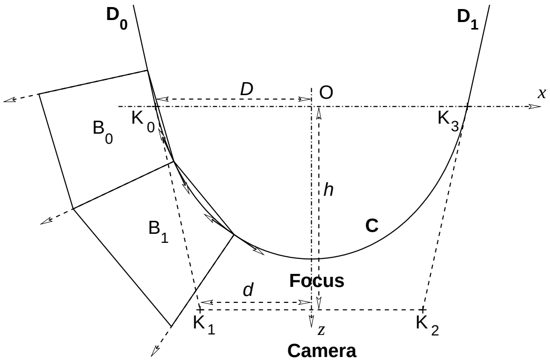


Figure 2: *The curvilinear organization of 3D-XV.*

Sharpness control

Since we assume that the curve is symmetric with respect to the O*z* axis, the coordinates of the control points are defined by *d* and *D* the front and rear semi-width of the curve and its height *h* (see Figure 2). In order to minimize the impact of sharpness modification on the linear position of blocks, it is necessary that the change of sharpness does not significantly modifies the length *l* of the Bezier curve. For this purpose, the three measures controlling the positions of $\{\mathsf{K}_0, \mathsf{K}_1, \mathsf{K}_2, \mathsf{K}_3\}$ are correlated and satisfy the following equations:

$$\begin{cases} d = D_{min} - D_{min} \times \frac{(D - D_{min})}{(D_{max} - D_{min})} \\ \quad \text{with} \quad D_{min} = 0.3 \text{ and } D_{max} = 0.6 \\ h = 2.0 \times d \end{cases} \quad (1)$$

The sharpness is modified by changing the value of D and by reflecting this change on the values of d and h according to the preceding equations.

Three different shapes are illustrated by Figure 3. The sharpest curve corresponds to values of D, d, and h such that $\{\mathsf{K}_0, \mathsf{K}_1, \mathsf{K}_2, \mathsf{K}_3\}$ are the vertices of a square, and D_0 and D_1 are parallel. The widest shape corresponds to a flat curve for which K_0, K_1, K_2, K_3, D_0, and D_1 are aligned. All the intermediate shapes are obtained for values of D between D_{min} (0.3) and D_{max} (0.6).

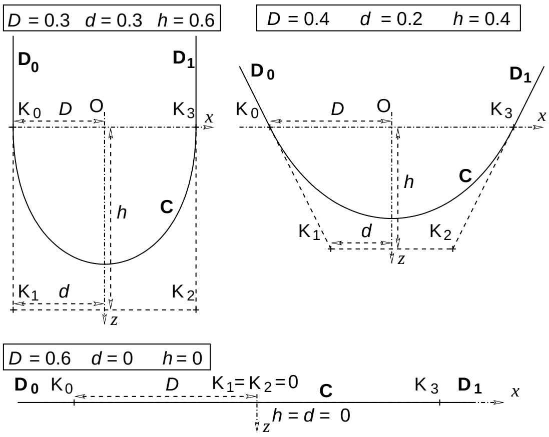


Figure 3: *Focus sharpness*

Focus and length control

Each point of the curve is associated with a parameter t. The points on C correspond to the values of t in [0,1], the points on D_0 to the negative values, and the points on D_1 to values greater than 1. K_0 corresponds to $t = 0$, and K_3 to $t = 1$. The points associated with values of t outside [0,1] are chosen so that the length associated with a parameter variation Δt are identical on the lines and on the Bezier curve. Let l be the length of C, and $\overline{\mathsf{K}_1\mathsf{K}_0}$ and $\overline{\mathsf{K}_2\mathsf{K}_3}$ the normalized vectors collinear to the vectors $\mathsf{K}_1\mathsf{K}_0$ and $\mathsf{K}_2\mathsf{K}_3$, the mapping between a value of t and a point P on D_0, C, or D_1 is computed as follows:

$$\begin{cases} t \leq 0 & \mathsf{P}_t \in \mathsf{D}_0 \wedge \mathsf{P}_t = \mathsf{K}_0 - lt \times \overline{\mathsf{K}_1\mathsf{K}_0} \\ 0 \leq t \leq 1 & \mathsf{P}_t \in \mathsf{C} \wedge \mathsf{P}_t = (1-t)^3\mathsf{K}_0 + 3t(1-t)^2\mathsf{K}_1 \\ & \qquad + 3t^2(1-t)\mathsf{K}_2 + t^3\mathsf{K}_3 \\ t \geq 1 & \mathsf{P}_t \in \mathsf{D}_1 \wedge \mathsf{P}_t = \mathsf{K}_3 + l(t-1) \times \overline{\mathsf{K}_2\mathsf{K}_3} \end{cases}$$

The parameters t_0 and t_1 associated with the anchor points P_{t_0} and P_{t_1} of a block B_0 are calculated from r_0 and r_1 the ranks of its initial and final words, L the curvilinear length of the document, N the total number of words in the document, and t_i the parameter associated with P_{t_i} the beginning point of the document:

$$t_0 = t_i + L\frac{r_0}{N} \quad \text{and} \quad t_1 = t_i + L\frac{r_1}{N} \quad (2)$$

Therefrom, the parameter t_f associated with P_{t_f} the end point of the document is $t_i + L$.

As indicated by Formula (2) above, the locations of the blocks on the curve depend on the values of their private variables (the ranks r_0 and r_1 of their initial and final words) and on the values of document variables (the parameter of the beginning point t_i and the curvilinear length of the document L). These last two variables define the focus of the document and the details with which blocks are viewed. A modification of their values can be used to change the focus of the document or to stretch or shrink the document blocks.

Focus change The focus of the curve is the median part of C, the point $\mathsf{P}_{0.5}$ whose parameter t is 0.5. If the values of the initial and final parameters of the document (t_i and $t_i + L$) are such that they contain the focus parameter 0.5 ($t_i \leq 0.5 \leq t_i + L$), there is a unique *focus word* whose rank $r_{0.5}$ is obtained from Equation (2):

$$r_{0.5} = \left\lceil \frac{N}{L}(0.5 - t_i) \right\rceil \quad (3)$$

The preceding formula shows that a modification of t_i, the initial parameter, results in a change of focus.

Stretching The length of a block is proportional to the curvilinear length of the document L. From Equation (2), we obtain that the parameter span $t_1 - t_0$, the difference between the values of the initial and final parameters of a block, is proportional to the length of the document:

$$t_1 - t_0 = \frac{L}{N}(r_1 - r_0) \quad (4)$$

Thus a modification of L, the document length, implies a proportional modification of the length of any block. The stretching of the document is associated with a simultaneous modification of the parameter t_i in order to avoid focus change and user disorientation. Stretching without change of focus requires to distribute the stretching on both sides of the document: the part between the initial parameter ($t = t_i$) and the focus ($t = 0.5$) and the part between the focus and the end parameter ($t = t_i + L$).

Let $k = L'/L$ be the ratio of L', the new document length, to L, its initial length. The new initial parameter value t'_i such that document stretching does not modify focus is obtained from Equation (3):

$$\begin{aligned} & (0.5 - t'_i)\tfrac{N}{L'} = (0.5 - t_i)\tfrac{N}{L} \\ \Leftrightarrow \quad & t'_i = \tfrac{1-k}{2} + k \times t_i \end{aligned} \quad (5)$$

4. Clustering and colorization

The geometrical model reflects the logical organization of the document (reading order and structure). For information access purposes, it is complemented with two additional dimensions: thematic clustering and dynamic querying. Thematic clustering can be used for similarity-based browsing since close colors correspond to thematically close passages or structures. Dynamic querying presented in the next section serves the purpose of targeted information access through natural language queries.

There are two main modes for text classification: *supervised* or *unsupervised clustering* [7]. Supervised clustering requires a training set which associates document or passage instances with human evaluations. In addition to their human cost, such algorithms cannot cope with documents on domains which have not yet been manually tagged. In an open environment, it is desirable to be able to accept any document on any domain without requiring a preliminary tagging task. For this reason, we have chosen an unsupervised algorithm which offers the additional facility of letting the user decide the final number of classes.

Term clustering

The computation of the similarity between paragraphs is made in three steps: first, paragraphs are represented as bag of terms (words or lemmas obtained from a tagging algorithm), second terms are grouped into K clusters through the following clustering algorithm, and third the paragraphs are projected on these K dimensions.

The algorithm used to cluster terms is an iterative optimization process of *K-means clustering* based on the *Kulback-Leibler* distance[2]. It is a dynamic process whose goal is to produce an optimal partition of the terms into K disjoint clusters. The optimal partition is obtained when the sum of the distances between the terms and the center of their clusters is smallest. An initial mapping is chosen, the terms are then moved from one cluster to another one, until there is no more distance improvement. The Kulback-Leibler distance is based on entropy, a global measure and makes the algorithm insensitive to the initial conditions.

The graphic representations presented in this article rely on a Professoral Thesis that contains 1,664 paragraphs and 160,000 words with 12,280 different word forms. Four levels of term classification are produced by the algorithm respectively made of 3, 6, 12, and 24 clusters. The algorithm is designed in order to ensure color continuity between the different levels: the clusters obtained for n clusters serve as initial mapping for building the $2 \times n$ clusters.

Paragraph colorization

In order to associate colors with paragraphs projected on the n clusters of terms, the coordinates of paragraphs in the n-dimensional hyperspace generated by these clusters must be converted into coordinates in a color space. First each dimension is associated with a unique saturated color, and then the colors associated with the dimensions are used to compute the paragraph colors.

The mapping between dimensions and saturated colors is performed in two steps. Since only saturated colors are associated with the K dimensions, they must first be projected on a 1-dimensional space. We first make a *projection* p of the K dimensions on the unit circle C_1 by associating each dimension d with a unique angle $p(d) = \alpha_d \in [0..2\Pi]$. Then a *color mapping col* defines a correspondence between an angle α_d on the unit circle and an angle $\alpha'_d = col(\alpha_d)$ on the HSV color circle (see Figure 4). The final mapping between dimensions and colors is the forward composition of p and col.

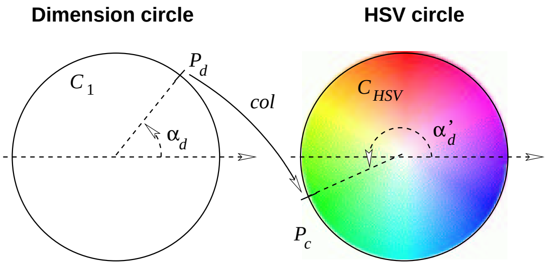


Figure 4: *Mapping between dimensions and saturated colors*

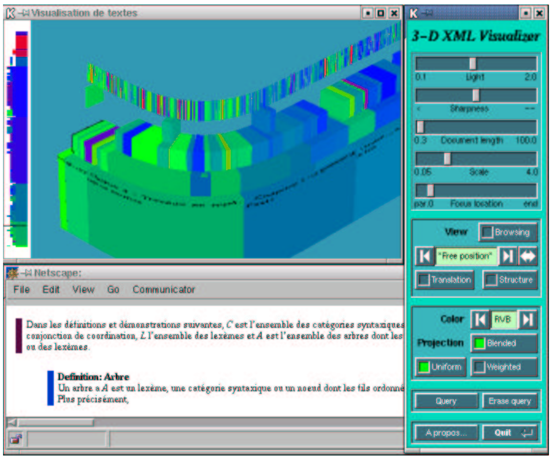
Two color mappings *col* are retained: a *uniform mapping* that distributes the dimensions uniformly on the color circle, and an *identical mapping* that preserves the angles. For each mapping the resulting colors vary depending on how the base colors (e.g. RGB) are placed on the color circle.

The colors of the paragraphs are based on the colors associated with each dimension. In the *blending mode*, the color of each paragraph is obtained through a weighted sum of the dimensions' colors, in which the weights are the coordinates of the paragraph. In the *clamping mode*, the color of a paragraph is the color of the dimension corresponding to its highest coordinate.

5. *3D-XV* GUI

The *3D-XV* GUI combines the spatial layout resulting from the geometrical model and colors produced by the preceding algorithms into a single representation of the document (see Figure 5). The interface is composed of a graphic and a monitoring window. The graphic window contains a perspective view of the colorized documents and a leftmost slider. The slider is a front view of the flat document and a position bar that indicates the position of the current focus. The monitoring window is used to modify the focus location, the viewing parameters, and the parameters of the geometrical model. In addition, the application is associated with an

HTML browser for direct access to the textual content of the document by clicking on a geometrical block.



Dans les définitions et démonstrations suivantes, *C* est l'ensemble des catégories syntaxiques
conjonction de coordination, *L* l'ensemble des lexèmes et *A* est l'ensemble des arbres dont les
ou des lexèmes.

Definition: Arbre
Un arbre a *A* est un lexème, une catégorie syntaxique ou un noeud dont les fils ordonné
Plus précisément,

Figure 5: *3D-XV GUI*

Large scale visualization and querying

Two visualization means are offered by the *3D-XV* graphic window: *overview* and *browsing* modes. The overview mode, illustrated by Figures 5 and 6, outlines the structure of the document. The titles of the chapter are the only textual data displayed in this mode. The perspective view provides a fisheye distortion with a focus on the center of the curve.

In the overview mode, the user can enter natural language queries. These queries are parsed by a natural language processing component and compared with an inverted file built from the words and lemmas in the paragraphs of the document. For each paragraph a relevance score is computed through a measure combining traditional information retrieval query-document relevancy and statistics about linguistic similarity. Each relevance score is displayed as a white bar on top of the front face of the corresponding block (Figure 6). As for color, relevance is percolated from the paragraphs to the upper structures.

Detailed browsing

In order to facilitate information access without using an associated HTML browser, *3D-XV* offers a browsing mode for reading the textual content of the paragraphs in the graphic window. The same curve is used as in the overview mode, the document is rotated 90 degrees, and the geometrical appearance of the blocks is modified. All the blocks have the same width, and only one face is displayed: a horizontal face for the paragraphs and a vertical face for the structure blocks (see Figure 7). As a result the text rolls like a moving walkway while titles of chapters and sections appear vertically in the topmost part of the screen.

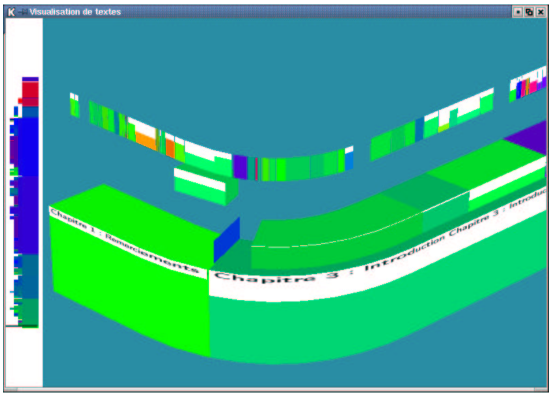
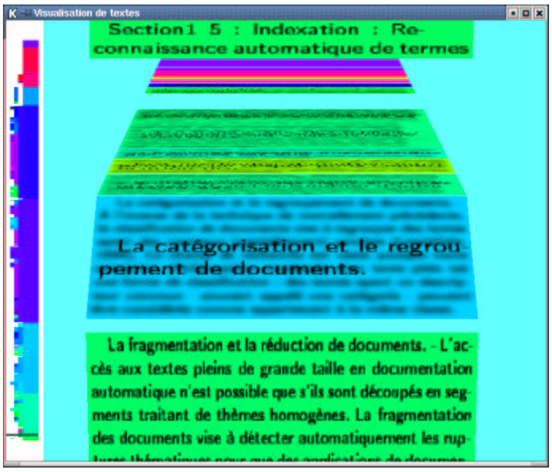


Figure 6: *3D-XV query relevance feedback*



Section1 5 : Indexation : Reconnaissance automatique de termes

La catégorisation et le regroupement de documents.

La fragmentation et la réduction de documents. – L'accès aux textes pleins de grande taille en documentation automatique n'est possible que s'ils sont découpés en segments traitant de thèmes homogènes. La fragmentation des documents vise à détecter automatiquement les ruptures thématiques pour que des applications de documen

Figure 7: *3D-XV paragraph browsing*

The text content of the paragraph can be displayed at two levels of detail. At the detailed level, the complete text of each paragraph is shown. In the skimming mode, the blurred text appears in the background and only a few significant words are displayed in the foreground. The display mode of the paragraphs automatically shifts from the skimming mode to the detailed level when the paragraphs come close to the focus point (see Figure 7). As a result, the graphic window is roughly divided into three zones from top to bottom: section and chapter titles, skimmed paragraphs, and detailed paragraphs. The leftmost slider is preserved in order to facilitate the user's orientation.

6. Discussion and comparison with previous studies

The *3D-XV* GUI integrates the results of previous studies on visualization of passage clustering or passage relevance. The use of a white bar indicating the relevance of passages is inspired from *TileBars*[4] and the coding of paragraph similarity while maintaining paragraph at their original location in the document is taken from Salton's work[9]. Contrary to these works, our approach is scalable to very large documents through the use of perspective distortion as in the *Perspective Wall*. Integration of clustering and structure through color coding is certainly a unique feature of our work.

Our approach differs from Web-based information retrieval systems such as *Lighthouse*[5], because we do not navigate in a network of interconnected documents but in a strongly structured geometrical model. The geometry of our model relies on the logical structure of the document. It makes it more easily acceptable by the users than the abstract geometry of a network.

The relevance of 3D interfaces for information access is debated in the information visualization community. For instance, Sebrechts et al [10] provide a controlled comparison of text, 2D, and 3D approaches to information retrieval. Their results show better response times with text and 2D interfaces than with a 3D spheric model. The 3D interface is less efficient for novice users, but the response times for the three modes of visualization are similar when the users are more experienced with the system. Their study also indicates that color coding is considered by the users as a useful tool.

The geometrical model of Sebrechts et al is a sphere that has the drawback of provoking user's disorientation because of its full symmetry. In his guideline #8, Vinson[12] suggests to add objects that break the symmetry of symmetrical models such as pine trees. Another solution is to use a model with a "natural" orientation because of its lack of symmetry. In the case of *3D-XV*, the geometrical model is clearly oriented, and easy to learn because it can be intuitively assimilated with a kind of building in which chapters would make the first floor and paragraphs the top floor. Isodirectional representations are better-suited for multi-user interfaces, in which all participants play similar roles[11]. In such a context, a circular version of *3D-XV* can easily be obtained by replacing the 3-piece curve by a circle.

7. Conclusion

3D-XV offers a new mode for accessing large structured documents. It combines the facility of visual display and direct document manipulation. Future user studies will provide results about the efficiency of the GUI in information access tasks. Future work also includes improvements such as

- automatic selection of relevant words for skimming visualization of paragraphs, and addition of iconic information corresponding to the rough semantics of the text,
- integration of the geometrical model in an immersive 3D virtual environment,
- extension of the model to document edition,
- application to domains such as computer-aided tutoring or information access in digital libraries.

Acknowledgements

We are grateful to Helka Folch (LIMSI-CNRS) for her help with the Book DTD, and to Jean-Philippe Merienne (LIMSI-CNRS) for his suggestions about text images.

References

1. Book DTD ISO-12083, http://www.xmlxperts.com/bookdtd.htm. 1
2. T. Cover and J. Thomas. *Elements of Information Theory*. Wiley & sons, New York, 1991. 4
3. Extensible Markup Language (XML), http://www.w3.org/XML/. 1
4. M. A. Hearst. Tilebars: Visualization of term distribution information in full text information access. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 59–66, Denver, CO, 1995. 1, 5
5. A. Leuski and J. Allan. Lighthouse: Showing the way to relevant information. In *Proceedings of IEEE Symposium on Information Visualization 2000 (InfoVis 2000)*, pages 125–130, 2000. 6
6. J. D. Mackinlay, G. G. Robertson, and S. K. Card. The perspective wall: Detail and context smoothly integrated. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 173–179, New Orleans, LA, 1991. 2
7. C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, 1999. 4
8. G. Robertson and J. D. Mackinlay. The document lens. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST'93)*, pages 101–108, 1993. 2
9. G. Salton, C. Buckley, and A. Singhal. Automatic analysis. theme generation and summarization of machine-readable texts. *Science*, 264:1421–1426, 1994. 2, 6
10. M. M. Sebrechts, J. Cugini, S. J. Laskowski, J. Vasilakis, and M. S. Miller. Visualization of search results: A comparative evaluation of text, 2D, and 3D interfaces. In *Proceedings, 22th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99)*, pages 3–10, 1999. 6
11. F. Vernier, N. Lesh, and C. Shen. Visualization techniques for circular tabletop interfaces. In *Proceedings Advanced Visual Interfaces 2002*, 2002. 6
12. N. G. Vinson. Design guidelines for landmarks to support navigation in virtual environments. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 278–285, 1999. 6