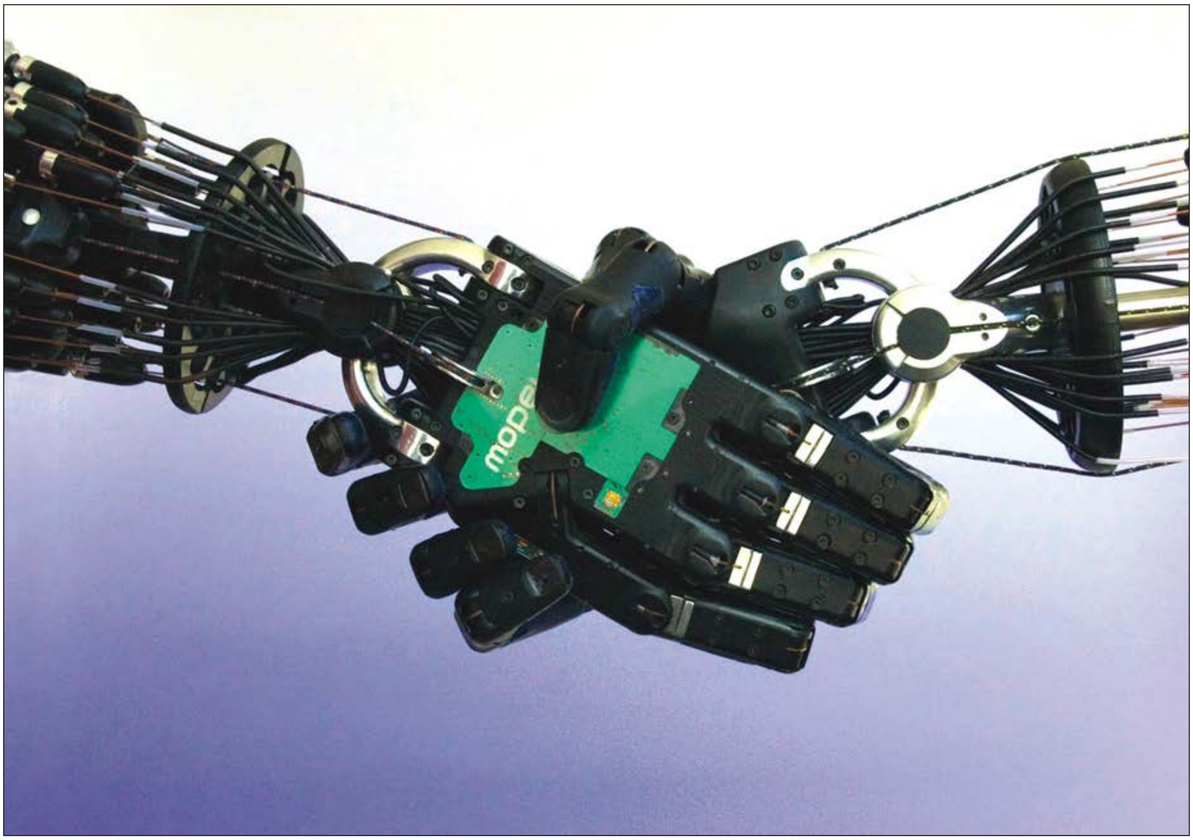

Journal of Artificial Intelligence

ISSN 1994-5450

Journal of Artificial Intelligence 1 (1): 53-60, 2008
ISSN 1994-5450

Story Summary Visualizer Using L Systems

T. Mala and T.V. Geetha
Department of Computer Science and Engineering, College of Engineering,
Guindy, Anna University, Chennai, India

Abstract: Visualization is an art where abstract data is graphically depicted to analyze the underlying knowledge inherent in the data. Documents can be considered as abstract data and document visualization is one area where the unbounded hidden knowledge in the documents can be clearly revealed using various visual techniques. Here in this study, we describe a Story Summary Visualizer which uses a novel method of biological story summary visualization based on Gestalt perception law. The visualizer takes children stories as input and generates summary of the stories using Relevance model. Later this summary information was used to model biological plants. Lindenmayer systems called as L systems which are fractal based techniques were used to model the plants applying gestalt similarity perception principle. The plants which are mathematically generated uses a set of iterative rules called productions applied over an axiom. This repetitive principle of applying productions results in the generation of biological plants which have the self inherent similarity within itself. This plant modeling technique uses the weighted sentence value of all the sentences in the story obtained from relevance model for generating the graftal based plant. Two plants are generated one story plant based on all the sentence weights and the summary plant is a child of the story which is extracted from the parent story plant. The summary plant helps in visualizing the summary sentence locations within the story and the summary sentence weights with respect to all story sentence weightages. This visualizer also helps in visual story analysis to know about spatial organization of the theme of the story and story summary orientation.

Key words: Information visualization, fractals, L systems, gestalt laws

INTRODUCTION

Document Visualization is an information visualization task that visualizes abstract data to facilitate comprehension, memory, communication and inference. Various methods were used to visualize the text as a whole or the attributes associated with text such as category information visualization, author attribution visualization, Ontology visualization and summary information visualization. Overview and detail method and distortion based visualization techniques are usually used to visualize document with their attributes. Category visualization techniques use self organizing maps for clustering (Yang *et al*., 2002), galaxy visualization using proximity to show relationships between documents (Hetzler *et al*., 1998), interactive time line viewer to show variants among documents (Monroy *et al*., 2002) and visualizing category weight information of news items using spirals using gestalt perception based models based on a novel SRP categorizing algorithm (Mala and Geetha, 2005, 2007a). Author attribution visualization helps in visualizing the author write prints using principal component analysis technique (Abbasi *et al*., 2006) and Modeling author style features using blobby objects (Mala and Geetha, 2007b). Ontology visualization is done using many tools like

Corresponding Author: T. Mala, Department of Computer Science and Engineering, College of Engineering, Guindy, Anna University, Chennai, India

Jambalya, Ontoviz, TG Viz, Ezowl, Ontoedit and Visualizer. These tools use network, hierarchical, neighborhood and hyperbolic view schemes to visualize the ontology concept and relational information (Bosca1 *et al*., 2005). Here in this study we propose a new technique of summary information visualization. The summary information is extracted from stories using Relevance model. The summaries thus generated are taken and they are then visualized using plant models. The visual plant models are generated using L systems which are Fractal based systems.

Summary Visualization

Summary information of text is also visualized using various techniques. Ineats an information extraction based multidocument summarization method uses content selection, Content filtering and Content presentation for summarizing multidocuments. It has a location based map visualizer for visualizing the location words and number of occurrences of location words. It also visualizes the summary sentences in the document using different colours (Leuski *et al*., 2003). A multi document summary system using centroid, title, location and query method was developed and the summary sentences visualized in an interface window which displays the summary and the original document (Lopez *et al*., 2004). Fractal based summarization technique was proposed where the self similarity inherent in the structure of a text material was used to summarize the text at various abstraction levels and the result is also visualized onto a mobile screen (Yang and Wang, 2003). An interactive document summary visualization system is built where Information extraction, Filtering, Summarization and Analysis are involved. This system uses filtering technique to filter the phrases and summarize the extracted information and uses contingency tables for visualization over the phrases and the document attribute information like author information, dates and references (Mothe and Dkaki, 1998). All the above summary generation techniques visualize the summary information as text based visualization. Here in our work a novel summary visualization technique based on L system has been used to generate plant structures. This uses similarity principle of gestalt perception law to analyze the document summary based on both location and thematic information. According to Gestalt similarity principle similar looking objects tend to group together (Mala and Geetha, 2007a). Similarity principle is also used here for generating self-similar fractals.

Fractal Based Visualization

Use of phyllotactic patterns to create tree layouts (Carpendale and Agarawala, 2004) is an interesting application of fractal based tree visualization. In the study however only the tree structure has been designed but no mapping of information has been performed. Vector force field used with L system provides interaction with the environment and animations can be created (Noser, 2002). The animations created only represent natural scenarios but does not visualize information as such. Binary fractal tree and Van Koch snow flake fractals are used to visualize hierarchical information of directory structures and categorization of meta data in digital collections (Ong *et al*., 2005). Fractal rooms are techniques to display the result of queries to display multimedia data (Xu and Sergeant, 2007). In this study, a novel approach of modeling summary information for plants have been developed where the inherent hierarchical information based on sentence weight has been taken into account.

Story Summary Generation Based on Relevance Model

Children stories are considered and the summary generated using relevance measure (Gong and Liu, 2001). The relevance measure is the inner dot product of individual sentence term vector in the story to the overall story term vector. The sentence with highest relevance score is the most thematic sentence. This sentence is then eliminated from entire story for further sentence score calculation. While calculating the dot product for each sentence the already selected highest score

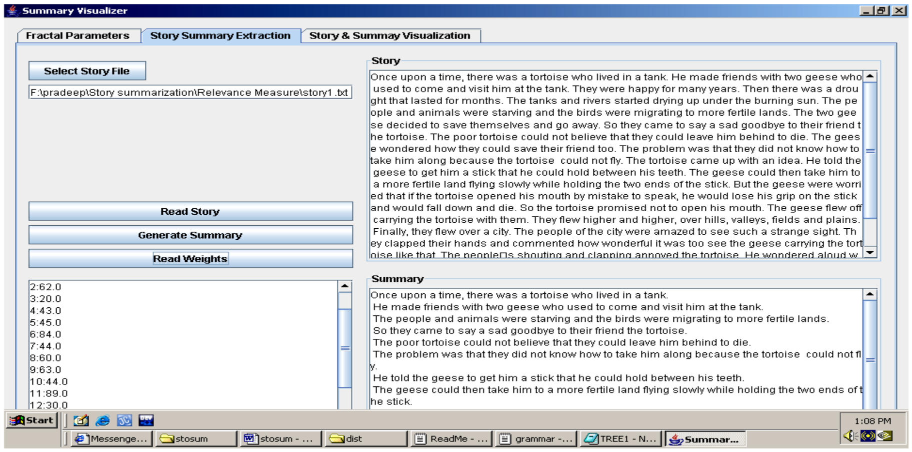

Fig. 1: Story summary generation

sentences are eliminated and the story term vector also changes for each iteration. The terms present in the highest score sentences are also eliminated from the story for further calculations. This is to reduce the occurrence of redundant sentences in the summary. As the number of sentences needed in summary is generated the process is stopped. This gives weightages for summary sentences and other sentences in story. These weights are mapped onto plant fractals for plant generation (Fig. 1).

Visual Plant Modeling Using L-Systems

The visual plant modeling is performed using L systems. The L systems are fractal based systems which are iteratively developed based on an axiom and productions. Here we used a simplest class of L systems the deterministic and context free L Systems called as DOL systems (Prusinkiewicz and Lindenmayer, 1990). Formal definitions describing DOL system is as below:

Let V denote an alphabet, V^* the set of all words over V and $V+$ the set of all non empty words over V.

A string OL system is an ordered triplet $G = < V, \omega, P>$ where V is an alphabet of the system, $\omega \; \varepsilon \; V+$ is a non empty word called the axiom and $P \subset V \; X \; V^*$ is a finite set of productions.

A production $(a,X) \; \varepsilon \; P$ is written as $a \rightarrow X$.

The letter a and word X are called the predecessor and successor of the production.

An OL system is deterministic if and only if for each $a \; \varepsilon \; V$ there is exactly one $X \; \varepsilon \; V^*$ such that $a \rightarrow X$.

To model higher plants using the above DOL system a graphical interpretation of L system is needed. The idea of the graphical interpretation using turtle is given below. A state of the turtle is defined as a triplet (x, y, α) where the Cartesian coordinates (x, y) represent the turtles position and the angle α called the heading is interpreted as the direction in which the turtle is facing. Given the step size d and the angle increment δ, the turtle can respond to commands represented by the following symbols:

(F)-> Draw Straight line in Current Direction for the specified length d. The state of the turtle changes to (x', y', α) where $x' = x + d \cos \alpha$ and $y' = y + d \sin \alpha$. A line segment between points (x,y) and (x',y') is drawn.

(+)-> Turn clockwise by specified angle δ. The next state of the turtle is $(x, y, \alpha - \delta)$.

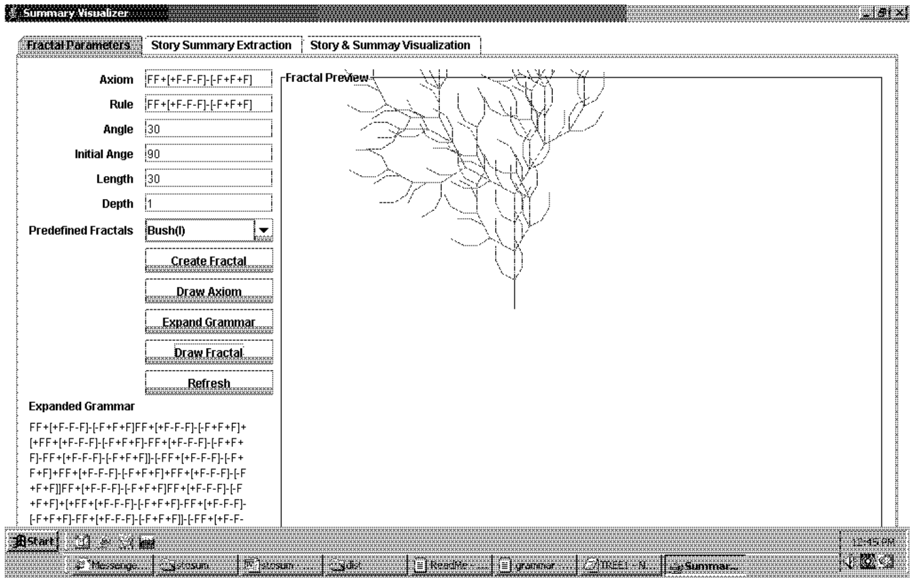

Fig. 2: Tree generation using L System

(-)-> Turn anticlockwise by specified angle δ. The next state of the turtle is (x, y, α+δ).
([)→ Save the current (x,y, α) in the stack.
(])→ Restore the current (x,y, α) from the stack.

Given a string V, the initial state of the turtle (x_0, y_0, α_0) and fixed parameters d and δ, the turtle interpretation of V is the figure drawn by turtle in response to the string V. Figure 2 with d = 30, α = 90 and δ = 30 is obtained by interpreting strings generated by the following L system. The axiom is given as FF+[+F-F-F]-[-F+F+F] and the production is FF+[+F-F-F]-[-F+F+F]. The tree so generated shows the iterative application of the production over the axiom and the expanded grammar being generated. These L systems are used for biological plant modeling and for generating inflorescences, flowers and patterns. Here L Systems are used for story summary plant generations.

The story sentences weights are calculated using relevance model as stated in the previous section. These weights are mapped onto the length or d value. This d value is used for drawing a line in the current direction with length d which is the mapped sentence weight. The initial angle α value is taken as 90 degrees which gives the direction in which the turtle is facing and δ value considered as 60 degrees, which gives the angle turned either clock wise or anti clock wise with the specified angle value. The axiom is given as FF+[+F]-[-F] and the production is FF+[+F]-[-F]. With these values the entire document is drawn as a plant with the sentence weightages as shown in Fig. 3.

The number of sentences to be present in the summary can be given by user. The summary sentences alone are shown in Fig. 4. The user has selected half the number of sentences for summary in Fig. 4. The plant models are generated based on the sentence weights generated by relevance model as said above. The weight of the sentences are mapped onto the length parameter d. Whenever symbol F occurs in the axiom or production, it is replaced by drawing a straight line in the current direction with the specified length d. Thereby the length of the branch depends on the sentence weightage. The location of the summary sentences can also be well understood by the increase in length of the summary sentences. The visualizer also maps perceptually to Gestalt perception principles.

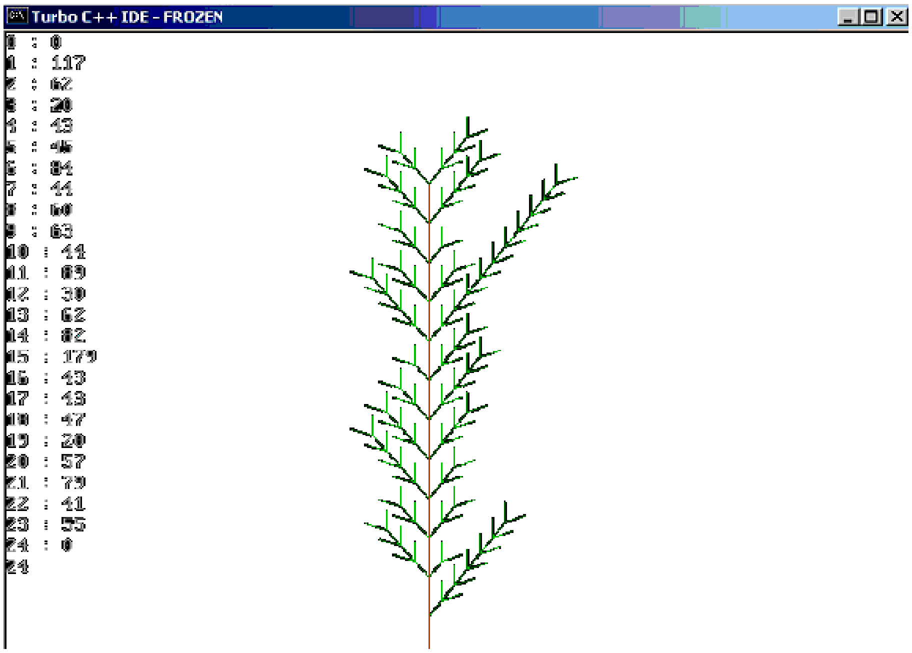

Fig. 3: Story summary visualizer

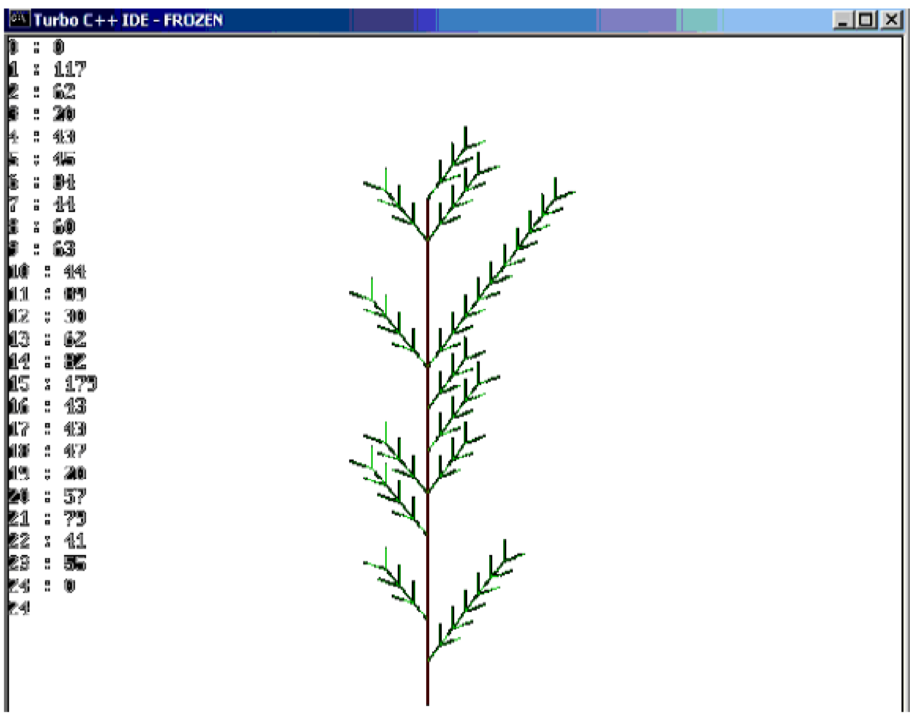

Fig. 4: Visualization of summary sentences alone

Gestalt Perception Model Applied on Visual Plant Models

The Summary Visualizer is based on Similarity principle of Gestalt perception model. By Gestalt similarity principle similar looking objects form a group. Here Self Similarity is inherently present as it is a fractal based plant model. Summary sentences have greater weights and thereby the length of the branch is more when compared to normal sentences. This helps in similar looking lengthy branches to be considered as summary sentences. Proximity principle is also mapped with the spatial nearness of the sentences in the documents. Here the spatial proximity is maintained where sentences lying near each other in the story will also be placed nearer by the visualizer. Thereby Gestalt principles of Similarity and Proximity are mapped onto visual plant models in the visualizer.

EXPERIMENTAL RESULTS

Experiments were conducted using four stories. The stories were given to two humans and the summarizer for evaluation. Let S_{man}, S_{sum} be the set of sentences selected by human evaluators and the Summarizer, respectively. The standard definitions for Precision(P) and Recall(R) are given as follows:

$R = S_{man} \wedge S_{sum}/S_{man}$
$P = S_{man} \wedge S_{sum}/S_{sum}$

Precision gives the number of correct sentences marked by system to the total number of sentences marked by the system while Recall gives the number of correct sentences marked by system to the total number of correct sentences marked by the human. Precision is a measure of the usefulness of a summary while Recall is a measure of the completeness of the summary. The key in building better summarizer is to increase precision without sacrificing recall. The calculated Precision and Recall are shown in Table 1 for four different stories for one user.

The Table 1 shows that for three stories out of four the Precision is more than Recall which means the summary sentences retrieved by the summarizer are valid sentences and provides a useful summary. In story 1 Recall is more than Precision, which means the summarizer provided summary sentences which matches many sentences in human summary which gives a more complete summary. This proves the efficiency of the Summarizer. The visual output is also evaluated based on four factors namely Data density, Number of simultaneous dimensions displayed, Occlusion percentage and Number of identifiable points (Bertini and Santucci, 2004). Data density is given as the ratio of the number of data points to the number of pixels. This gives how clearly the data can be displayed. Number of simultaneous dimensions displayed gives the number of maximum dimensions that can be displayed in a visual. Occlusion percentage gives the number of broken elements in the visual output and Number of identifiable points tells the clarity in the image. The parameter percentages were calculated. The calculated values clearly shows that the clarity of the image is good since identifiable points is 100% and the image is clearly visible without any breakages which makes the occlusion percentage to be 0. Data density value is calculated to be 40% and the number of simultaneous data dimension displayed is 2 X number of sentences in the story which shows the dimensions that can be seen clearly from the image at a time.

Table 1: Calculated precision and recall using relevance model

Precision	Recall
0.45	0.58
0.46	0.39
0.65	0.55
0.53	0.47

CONCLUSION AND FUTURE WORKS

The proposed system was developed and implemented. The results proved to be satisfactory. The system helps in generating the summary of children stories and also helps in Visualizing the stories. Relevance score model proved to be efficient in extracting the summary. The story summary visualizer helps users in analysing the summary sentence weights to that of normal sentence weights. The plant models generated using L Systems also helps in understanding the summary weightages, Summary sentence spatial location and total sentences in the story with their weights as a whole and the location of the story theme within the story. Gestalt perception principles when mapped onto the system help in clear and easy understanding of the visual. The visualizer can be improved by generating three dimensional plant models using L Systems. The summarizer can also be improved by extracting summary from multiple documents or stories.

REFERENCES

Abbasi, A., H. Chen and S. Mehrotra *et al*., 2006. Visualizing authorship for identification. LNCS 3975, pp: 60-71.

Bertini, E. and G. Santucci, 2004. Quality metrics for 2D scatterplot graphics: Automatically reducing visual clutter. Smart Graphics. 4th International Symposium on Smart Graphics, Banff Centre, Canada, May 23-25, pp: 77-89.

Bosca1, A., D. Bonino1 and P. Pellegrino1, 2005. OntoSphere: More than a 3D ontology visualization tool. SWAP 2005, 2nd Italian Semantic Web Workshop, Trento, Italy, December 14-16.

Carpendale, S. and A. Agarawala, 2004. PhylloTrees: Harnessing Nature's Phyllotactic Patterns for Tree Layout. Extended Abstracts of IEEE Conference on Information Visualization (Poster), IEEE Press.

Gong, Y. and X. Liu, 2001. Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. SIGIR'01, September 9-12, New Orleans, Louisiana, USA.

Hetzler, B., W.M. Harris, S. Havre and P. Whitney, 1998. Visualizing the full spectrum of document relationships. 5th International Soc. for Knowledge Organization (ISKO) Conference, pp: 168-175.

Leuski, A., C. Lin and E. Hovy, 2003. iNeATS: Interactive multi-document summarization. 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003), Interactive Posters and Demos Session, Sapporo, Japan, July 7-12.

Lopez, M., M.D. Buenaga and J. Gomez-Hidalgo, 2004. Multidocument summarization: An added value to clustering in interactive retrieval. ACM Trans. Inform. Syst., 22 (2): 215-241.

Mala, T. and T.V. Geetha, 2005. Visualization of biological patterns in event detection and tracking based on srp algorithm. 9th International Conference on Information Visualization (IEEE), London, 4: 587-592.

Mala, T. and T.V. Geetha, 2007a. Multi level categorization visualization based on Gestalt perception model. 15th Scientific Convention Relations and Structures, Developments of Gestalt Theory in Psychology and Adjacent Fields. May 25-27, Marcerata, Italy.

Mala, T. and T.V. Geetha, 2007b. Visualizing author attribution using blobby objects. 4th International Conference Computer Graphics. Imaging and Visualization 14: 15-17 August, Bangkok, Thailand.

Monroy, C., R. Kochumman, R. Furuta and E. Urbina, 2002. Interactive timeline viewer (ItLv): A Tool to visualize variants among documents. Lecture Notes Comput. Sci., 2539: 39-49.

Mothe, J. and T. Dkaki, 1998. Interactive multidimensional document visualization. SIGIR'98, Melbourne, Australia.

Noser, H., 2002. Lworld: An animation system based on rewriting. Proceedings of the 10th Pacific Conference on Computer Graphics and Applications (PG'02).

Ong, T., J.J. Leggett and U. Yun, 2005. Visualizing hierarchies and collection structures with fractal trees. Proceedings of the 6th Mexican International Conference on Computer Science (ENC'05).

Prusinkiewicz, P. and A. Lindenmayer, 1990. The Algorithmic Beauty Of Plants. Springer Verlag.

Xu, Y. and D. Sergeant, 2007. An Overview of Document Visualization. EVIE Project, EDV-08.

Yang, C.C., H. Chen and K. Hong, 2002. Exploring the world wide web with self-organizing map. Proceedings of International World Wide Web Conference, Honolulu, Hawaii, May 7-11.

Yang, C.C. and F. Wang, 2003. Fractal summarization for mobile devices to access large documents on the Web. WWW2003, May 20-24, 2003, Budapest, Hungary, ACM 1-58113-680-3/03/0005.