ISBN: 978-0-9775978-3-3
http://australia.vsmm.org

Generative Website: Visualising Possible Stories

Shigeki Amitani[1] and Ernest Edmonds[1]

[1] Creativity & Cognition Studios, Australasian CRC for Interaction Design,
University of Technology, Sydney, Australia
shigeki@shigekifactory.com, ernest@ernestedmonds.com

Abstract. This paper presents our on-going project called "Generative Website Project". We design and develop generative systems that visualise possible contexts, that is, possible sequences of information out of existing information in order to stimulate human creative thinking. Possible sequences of information segments, usually called storytelling, scenarios and narratives have been used as a tool for exploring and stimulating thinking about possible events, assumptions relating to these events, and courses of actions for a broad range of information designers from public audiences searching on the internet to analysts and policy makers. In this paper, we report our progress of the Generative Website Project and related technology.

Keywords: Generative system, Generative Website Project, visualisation, contexts, Term Dependency, Term Attractiveness, Main Topic Term Attractiveness, Knowledge Liquidization & Crystallization, Knowledge Nebula Crystallizer, knowledge creation

1 Introduction

The aim of this research is to develop a system that visualises possible sequences of information out of existing information so that information designers and audiences can discover possible contexts that otherwise could be missed.
Fig. 1 shows expected interactions between users and the system. First, information artefacts (existing ones and/or new pieces of information) are collected and stored (left in Fig. 1). The system (top in Fig. 1) segments the input information artefacts into small semantic units, then it restructures them into possible meaningful sequences of information shown in the right of Fig. 1. These outputs work in two ways: (1) as final products that a user (audience) can enjoy; and (2) as draft materials that a user (information designer) can modify (at the centre of Fig. 1).

A generative system visualises possible contexts, i.e. possible sequences of information as “stimulants” [9] regarding to a core theme that a user enters.

Suppose you are investigating “public art” on the Internet to write, for example, a report. There is a plenty of information regarding to interactive art, however, if you are writing a report, you have to decide from which viewpoint you are analysing public art (e.g. history of public art, types of public art, examples of public art, etc.), and also you have to compare which viewpoints yield interesting contexts for the report.

For this aim, you can repeatedly use a search engine to retrieve relevant information. However, as is often the case, enormous amount of web pages are retrieved that overwhelms users' capacities. Then you use more keywords to improve the search result. Problems here are: (1) the search results are limited by viewpoints that you have already in your mind; and (2) the search results with different keywords (viewpoints) actually have to be compared by jumping around multiple pages.

In this research, we develop a system to visualise possible sequences of information to challenge these problems. The concept of the system derived from Knowledge Liquidization & Crystallization [4, 15].

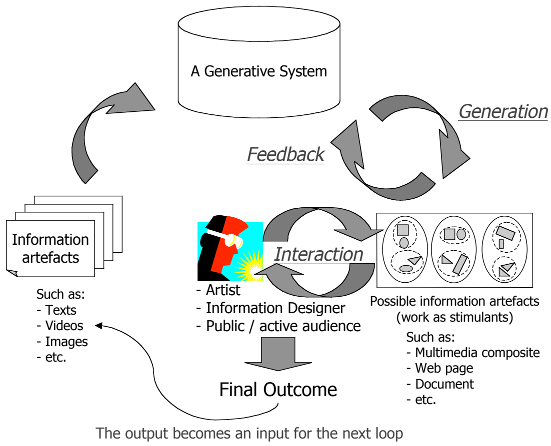

Fig. 1. Expected Interactions

2 Related Work

Humans naturally share knowledge by telling stories, sequences of information. This is a form of knowledge exchange and organise our experiences [11].

Possible sequences of information segments, usually called contexts, scenarios, narratives, or storytelling, have been used as a tool for exploring and stimulating thinking about possible events, assumptions relating to these events, and courses of actions for a broad range of information designers from public audiences searching on the internet to analysts and policy makers in various fields such as business [21-23] and military operations [7, 18, 21, 28, 30].

In 1990's, Nonaka et al. have pointed out that while research on *knowledge accumulation* has been reported, few studies have been conducted on *knowledge creation* [20]. Through their case studies, they identified features of knowledge: *knowledge is embedded in human practices and should be structured dynamically in* *accordance with the context of human practices*. Nonaka has claimed that knowledge includes the ability of observation on a phenomenon from various viewpoints and it is the time to establish methods to put theoretical frameworks of knowledge creation into practices [19]. That is, the process of knowledge creation includes producing actionable information.

This research takes the knowledge creation viewpoint. Hori et al. have proposed a framework of the process of knowledge creation called Knowledge Liquidization & Knowledge Crystallization in order to fill a gap between theoretical frameworks and human practices [15]. Knowledge Liquidization means dissolving knowledge into small granularity that have a core grounding to the real world and that preserve the local semantic relationships around the core. Knowledge Crystallization means restructuring the relationships among the granular units in accordance with a current context.

Suppose you are writing a paper, you refer to a number of relevant books and papers. You do not merely pile what is written in them as solid blocks, but you extract relevant parts (such as a paragraph, sentences, etc., i.e. semantic segments) from the books and the papers and fuse them into your paper.

This is the process of knowledge creation, and the Knowledge Liquidization & Knowledge Crystallization are correspondent to these processes. This framework is consistent with the perspective of the brain science, that is, human neocortex stores what a human being perceives as sequences of patterns, and recalls patterns auto-associatively [14].

Hori et al. also proposed a conceptual system called Knowledge Nebula Crystallizer that supports the process of Knowledge Liquidization and Knowledge Crystallization [15]. The essential functions are: (1) dissolving information into small semantic segments (corresponding to Knowledge Liquidization); and (2) restructuring relationships among these semantic segments (corresponding to Knowledge Crystallization). Actual systems have been implemented and applied to several domains (e.g., [3]).

Automated scenario, narratives, and storytelling generation have been explored (e.g. [16]). In military applications, the Campaign ontology was developed so that chains of commands and courses of actions can be compared with each other [18, 28].

While it is possible to establish ontology in areas where vocabularies for concepts and actions are almost fixed, creative activities do not always have their own established ontology or do not have a fixed ontology.

Akaishi et al. [2] have developed “Topic Tracer” that arranges segmented documents relevant to an input keyword (= a context) along with their time line. This is useful visualisation when a user knows in which context the user should analyse existing documents.

This research challenges the case (1) where no established ontology is available; and (2) where a user does not know what possible contexts exist in existing information repository.[1]

---

[1] We are aware that there are a number of related studies conducted in such fields as natural language processing, semantic web, text mining, etc., however, the related work concentrates on knowledge creation to contextualise this study and due to the space.

3 Design Rationales of Generative Systems

In the process of knowledge creation, it is possible to assume:

1. You have a certain topic in your mind
2. You are looking for contextually relevant pieces of information
3. You are trying to sequence the information pieces in a contextually correct way

In order for the generative system to support this process, the interaction should include:

1. A user enters or choose keywords
2. The system returns sequences of information pieces which are reasonably well-connected contexts as stimulants

Following computations support these interactions:

1. Segmenting information into pieces
2. Restructuring segments into sequences
3. Representing the generated sequences

In the following section, we discuss necessary computations.

3.1 Segmenting Information into Pieces

We adopted the criteria called Term Dependency (TD) and Term Attractiveness (TA) [2]. The ideas of these criteria are:

- If the value of Term Attractiveness of a term in a document is increasing as the document proceeds, then the term is becoming a main topic in this part of the document
- If the value of Term Attractiveness of the main topic term decreases, then the topic has changed to another

The topic terms are defined as words that have the highest value of Term Attractiveness as a document proceeds. Segmentation is conducted at the points where the value of Term Attractiveness of a main topic term decreases. Term Dependency is defined as follows:

$$td(t_1, t_2) = \frac{NumOfSentences(t_1 \cap t_2)}{NumOfSentences(t_1)}$$

where t_1 and t_2 are terms in a given document d, $td(t_1, t_2)$ is the value of Term Dependency of t_1 on t_2 in the document d, and $NumOfSentences(t)$ is the number of sentences that includes t. That is, Term Dependency is a conditional probability that t_2 is included in a sentence that also includes t_1.

Term Attractiveness of a term t in a document d denoted as $attr_d(t)$ is defined as sum of the values of Term Dependency on the term t:

$$attr(t_1) = \sum_t td(t, t_1) \quad (t \neq t_1)$$

By definition, the value of Term Attractiveness of a term t changes as sentences proceed. So the value of Term Attractiveness of a term t_j up to i-th sentence is denoted as $attr_i(t_j)$. When $attr_i(t_j)$ is the maximum value among all the attractiveness values, then the term t_j is regarded as a main topic term. Segmentation is conducted when another term becomes main topic term. As a result of this process, the system produces segments each of which is tagged with its main topic term.

3.2 Restructuring Segments into Sequences

We extend the idea of the Term Dependency to Main Topic Dependency. Main Topic Dependency is defined as follows:

$$mtd(mt_1, mt_2) = \frac{NumOfDocuments(mt_1 \cap mt_2)}{NumOfDocuments(mt_1)}$$

where mt_1 and mt_2 are main topic terms in given document set D, $mtd(mt_1, mt_2)$ is the value of Term Dependency of mt_1 on mt_2 in D, and $NumOfDocuments(mt)$ is the number of documents that include mt. That is, Main Topic Dependency is a conditional probability that mt_2 is included in a document that also includes mt_1.

This value is used for sequencing the segments. If a keyword that a user enter is contained in some segments, then they are retrieved as first segments of sequences. Sequences are generated based on the value of the Main Topic Dependency. Segments with the highest values of the Main Topic Dependency are retrieved and connected to the first segments. Then this process is repeated to make longer sequences.

4 An Application: a Generative Website

We are currently working on the Generative Website Project supported by Australasian CRC for Interaction Design (ACID), in collaboration with International Federation of Arts Councils and Culture Agencies (IFACCA). They are renewing their website[1] in order to provide more dynamic contents to audiences of the website.

The generative website is expected to be a place for following aims:

- Audiences of the website post their articles, comments, and video clips.
- The website provides dynamically woven sequences of information so that audiences can enjoy dynamic contents each time new information artefacts are added by the content provider and / or audiences.

We have implemented prototype systems that generate sequences of information with actual data. A prototype system has been developed in Java 1.4.2 to test the algorithms and its visualisation form. The segments are analysed again with TreeTagger [24] so that index word vectors can be developed for these segments. In the following sections, the implementation of the prototype system is explained.

The generated sequences are arranged in a two-dimensional space as the target users engage in information-intensive tasks [12, 17, 25]. A number of studies that retrieves information pieces containing designated keywords have been done. Information pieces are usually arranged close to each other according to their similarities. This representation supports users to see which information pieces are similar, which is useful for users to understand the information space [29]. However, it does not support users to grab how the retrieved information are inter-related with each other except for similarity. We consider that providing sequences of information, i.e. contexts, and visualising how they are inter-related with each other support users to restructure existing information.

We are developing two prototypes: one for textual information and another for videos. The system for textual information and its expected interactions are explained in this paper in order to communicate how our generative systems work. The prototype system for video information is also briefly explained in the discussion section.

In our example, 10 text data of articles on the website of the IFACCA website were used. The articles are actual reports from art conferences and news articles. The system segments them into 57 segments.

The prototype system is a stand alone application at the time when this paper is written. It has a simple user interface as shown in Fig. 2. It is as simple as Google's interface [13].

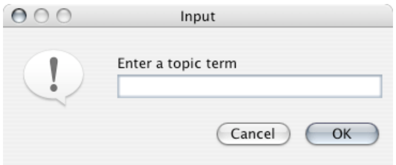

Fig. 2. The Initial Interface of the Prototype System

A usage scenario here is that a user is searching “stories” regarding to “musician” on the IFACCA website. The user enters the topic word "musician" in the text box in Fig. 2. Then the system retrieves segments and puts them into sequences. Fig. 3 shows the search result with keyword "musician".

A user enters a topic term that the user thinks is a main topic for what the user is searching. Topic words could be broad and general words such as words that people might choose first when they conduct a keyword search on a search engine.

The system searches segments containing the input keyword. Retrieved segments are regarded as first segments of sequences to be generated. Then the system retrieves segments that have the highest values of Main Topic Term Dependency to the first segments. These retrieved segments are the second segments in the sequences. This process is repeated until the system does not find segments with the value of Main Topic Term Dependency more than a certain threshold.

The circled segment in Fig. 3 (labelled "right0") is a segment containing the search term "musician". This segment is the first segment of possible sequences. The content of this segment is shown in the text area on the right side of Fig. 3. The segments connected with the first segment with lines in Fig. 3 (labelled with "presentation", "Organization", and "work") are suggested as the second segment of a sequence. That is, the user recognises that stories about "musician" is extended to the topics such as "presentation", "organisation" and "work".

By clicking segments, correspondent contents are shown up in the text area (Fig. 4). By pressing mouse button down on a segment, the segment is linked to the next candidates so that users can instantly see possible connections. By releasing the button the lines disappear. These clicking functions are applied to all the segments in the space.

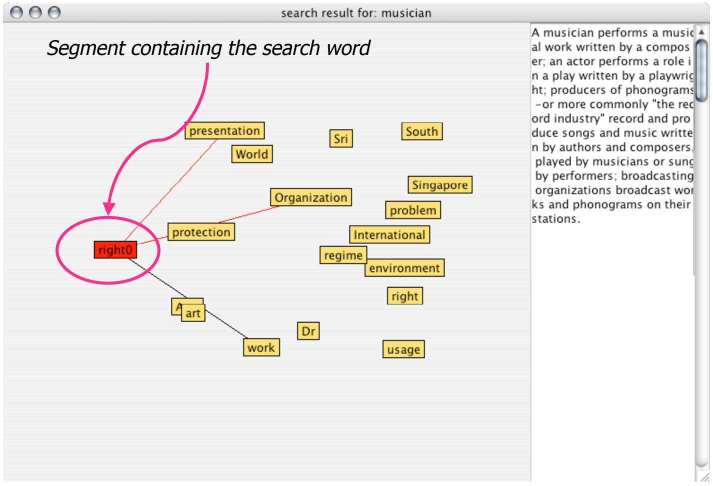

Fig. 3. An Initial Display of the Search Result

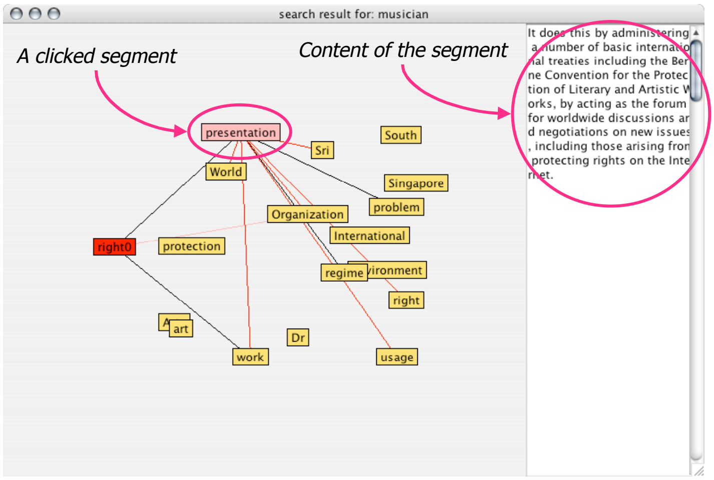

Fig. 4. A Clicked Segment Shows its Contents

By double clicking, the position of the target segment is fixed so that a user can fix his/her viewpoint (Fig. 5). Fixed segments are still movable by drag & drop. Users can formalise the information space incrementally [26].

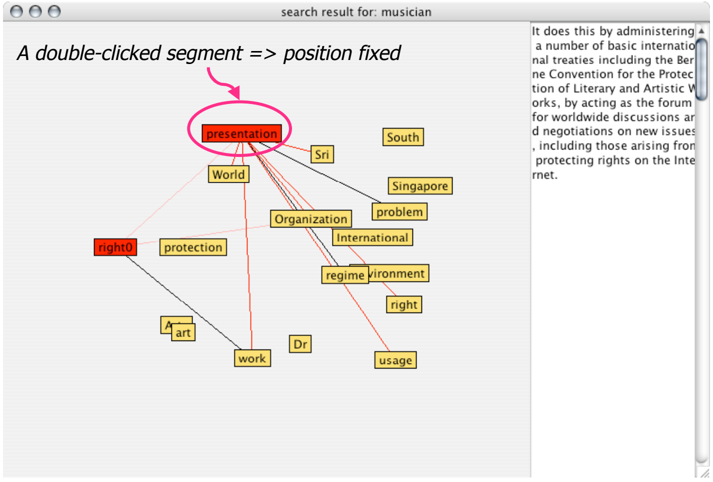

Fig. 5. Fixing Segment Position by Double-Clicking

These functions support to compare possible contexts simultaneously. For example, a user fixes two segments labelled with "presentation" and "Organization" as shown in Fig. 6 to see what these sequences are and which sequence is more interesting than the other for the user, and to see what segments are connected to these fixed segments to let sequences grow. This allows users to incrementally formalise suggested sequences.

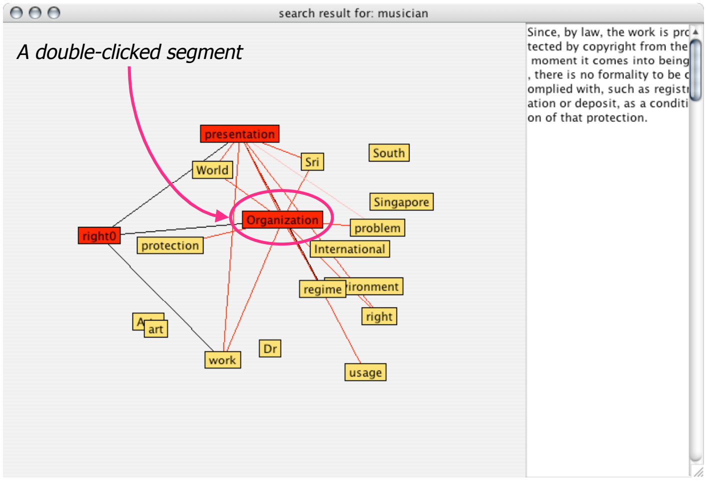

Fig. 6. Comparing Multiple Contexts Simultaneously

By double-clicking with the shift key, the system shows the entire selected sequences so that the user can actually read and understand the content of the sequence (Fig. 7). If a user shift-double-clicks a segment labelled "presentation" which is the second segment following the first segment labelled "right0" in this case, then the system shows a new text box that connects the content of the "right0" segment and that of the "presentation" segment (Fig. 7). Contents of the first segment up to the shift-double-clicked segment are connected into one sequence. Users can open as many text areas as they want so that they can compare multiple sequences with each other.

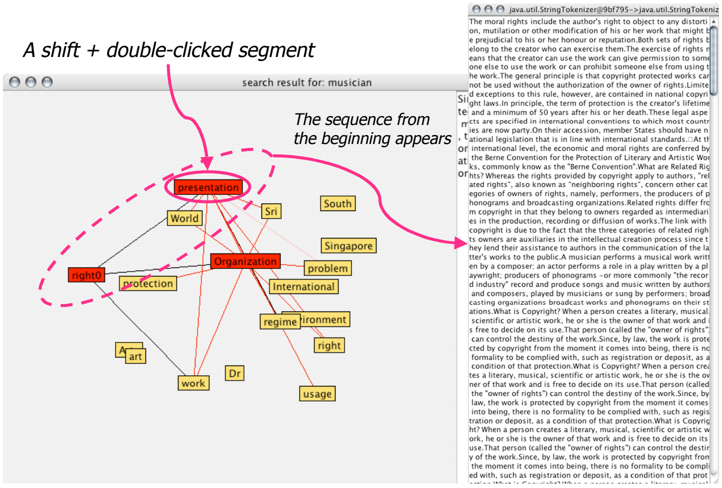

Fig. 7. Shift + Double-Clicking for Connected Content

5. Discussion

An example of generated sequence is shown in the appendix section. This result is composed of three segments whose main topics are "right", "Organization", and "protect" respectively, searched by the word "musician". The sequence is "right" => "Organization" => "protection"[2].

Our first impression is that the generated sequence is consistent and seems to be "a story". This sequence starts with general concepts of "author's right" in the context of creative activity in artistic fields. Then it moves to a more specific type of right "copyright". Then it tells the copyright in the musicians' context, and ends with practices of intellectual properties.

Our system aims at generating possible sequences that stimulate users' thinking based on the Knowledge Liquidization & Crystallization concept.

The system is an implementation of the KNC system. The computations, representations of information, and interactions with information are correspondent to what the KNC system has suggested [15]. It dissolves information artefacts (documents in this case) into small segments with preserving the local semantic relationships, and then it restructures the relationships among the segments in a contextual way.

---

[2] This generated sequence is available at: http://shigekifactory.com/research/GenerativeSystem/ExampleSequence.html

The generated sequences are "possible draft materials". The system supports this process with its representations and interactions: it indicates possible connections between segments in an incremental way so that a user can follow various possibilities of sequences and can formalise the information space incrementally. Also the system provides an overview of possibilities of sequencing information by the spatial representation, then it also provides detailed views by presenting sequenced texts. These representation and interactions support users essential processes in design, that is, going back and forth between overview and details, and whole and parts [27].

A number of studies have been done in the field of information retrieval to retrieve similar information pieces. Some studies of sequence generation have been conducted with using ontology or grammatical rules, however, little has been explored in such a bottom-up way as we conducted (e.g. [1]). We consider that a bottom-up approach is useful in order to establish a general method for generating stories across various domains.

We recognise a number of issues to challenge. For example, scalability is one of the most important issues that we have to consider, as the natural language processing is time-consuming when the amount of documents increases.

Also the quality of segmentations and generated contents need to be evaluated. We are going to conduct experiments to evaluate content quality. Interactions between users and the system should be further evaluated to improve the system. Our plan is to adopt the protocol analysis method [10] for this evaluation, as conducted in the authors' previous works ([3, 5, 6, 8]).

We are extending the mechanism to non-textual information such as videos. Fig. 8 shows a snapshot of a web-based system named VideoContextualiser.

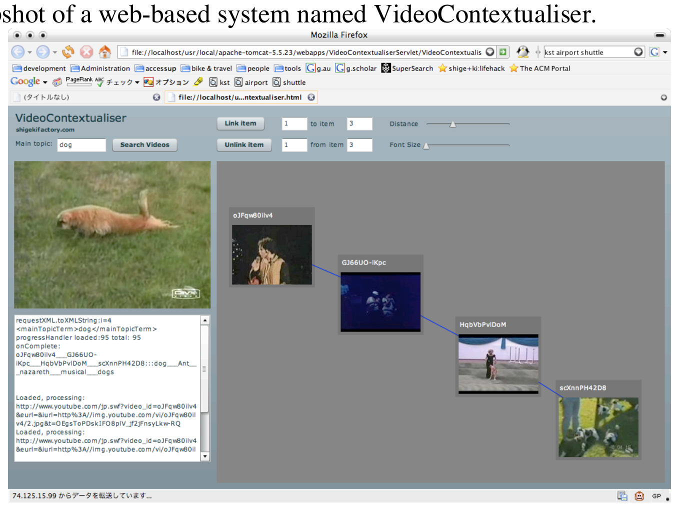

Fig. 8 A Snapshot of VideoContextualiser

Each of posted videos with tags is regarded as a sentence, and then the same computations described above are conducted so that videos are sequenced as a story.

We are going to develop this system as a significant part of the contents in the generative website[3].

6. Conclusion

In this paper, we presented design rationales of a generative system and prototype systems that visualise possible contexts out of existing information.

The criteria Term Dependency is extended to Main Topic Term Dependency so that contextualised sequence of information can be produced. While further evaluation is required, the preliminary evaluation of the generated content suggests the potential benefit of the system. The developed prototype system has been applied to an actual data set in our on-going project.

The preliminary evaluation of the generated content indicates the potential benefit of the system. We are going to conduct evaluations regarding to the quality of generated sequences and interactions to improve user experiences for their knowledge creation.

We are also extending this mechanism to videos. Annotated videos posted, for example, to YouTube by public audiences are sequenced into stories so that visitors to the website can enjoy dynamic and different experiences each time when they visit this website.

Acknowledgments. This research/work was partly conducted within the Australasian CRC for Interaction Design, which is established and supported under the Australian Government's Cooperative Research Centres Programme. The authors are also grateful for the support by Australian Centre for the Moving Images.

References

Akaishi, M., Hori, K. and Satoh, K., Topic Tracer: a Visualization Tool for Quick Reference of Stories Embedded in Document Set. in *International Conference on Information Visualization (IV2006)*, London, UK (2006), 101-106.

Akaishi, M., Hori, K. and Satoh, K., Topic Tracer: A Visualization Tool for Quick Reference of Stories Embedded in Document Set. in *International Conference on Information Visualisation 2006*, London, UK (2006), 101-106.

Amitani, S. A Method and a System for Supporting the Process of Knowledge Creation, Department of Advanced Interdisciplinary Studies, University of Tokyo (2004).

Amitani, S. and Hori, K., Knowledge Nebula Crystallizer for Knowledge Liquidization & Crystallization - from a Theory to a Methodology of Knowledge Management -. in *Expertise in Design*, University of Technology, Sydney, Australia, UTS Press (2003).

Bilda, Z., Costello, B. and Amitani, S. Collaborative analysis framework for evaluating interactive art experience. *CoDesign*, *2* (4) (2006) 225-238.

[3] The details about VideoContextualiser which is under development will be described in the near future and not in this paper as it is beyond the realm of this paper and due to lack of the space.

Candy, L., Amitani, S. and Bilda, Z. Practice-led strategies for interactive art research. *CoDesign*, *2* (4) (2006) 209-223.

Chick, S., Sánchez, P.J., Ferrin, D. and Morrice, D.J., Simulation Modeling and Analysis in Support of Brigade Assault Bridging Opration Planning. in *the 2003 Winter Simulation Conference* (2003).

Costello, B., Muller, L., Amitani, S. and Edmonds, E. Understanding the experience of interactive art: Iamascope in Beta_space. *Proceedings of the second Australasian conference on Interactive entertainment* (2005) 49-56.

Edmonds, E., Artists augmented by agents (invited speech). in *Proceedings of the 5th international conference on Intelligent user interfaces, New Orleans, Louisiana, United States* (2000), 68-73.

Ericsson, A. and Simon, H. *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press, (1993).

Garvey, C. *Play*. Open Books, London, (1977).

Golovchinsky, G. and Marshall, C.C. Hypertext interaction revisited. *Proceedings of the eleventh ACM on Hypertext and hypermedia* (2000) 171-179.

Google Inc.

Hawkins, J. and Blakeslee, S. *On Intelligence*. Owl Books, (2005).

Hori, K., Nakakoji, K., Yamamoto, Y. and Ostwald, J. Organic Perspectives of Knowledge Management: Knowledge Evolution through a Cycle of Knowledge Liquidization and Crystallization. *Journal of Universal Computer Science*, *10* (3) (2004) 252-261.

Kumar, D., Ramakrishnan, N., Helm, R.F. and Potts, M. Algorithms for storytelling. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (2006) 604-610.

Marshall, C. and Shipman, F. Spatial Hypertext: Designing for Change. *Communications of the ACM*, *38* (8) (1995) 88-97.

McKeever, W., Gilmour, D., Lehman, L., Stirtzinger, A. and Krause, L. Scenario management and automated scenario generation. *Proceedings of SPIE*, *6228* (2006) 62281A.

Nonaka, I. and Konno, N. *Methodology for Knowledge Creation (in Japanese)*. Toyo Keizai Inc., (2003).

Nonaka, I. and Takeuchi, H. *The Knowledge-Creating Company : How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, (1995).

Porter, M.E. *Competitive advantage*. Free Press New York, (1985).

Ringland, G. *Scenario planning : managing for the future*. John Wiley, New York, (1998).

Ringland, G. *Scenarios in business*. Wiley, Chichester, (2002).

Schmid, H. TreeTagger - a language independent part-of-speech tagger. *Institut fur Maschinelle Sprachverarbeitung, Universitat Stuttgart* (1995).

Shipman, F.M., Marshall, C.C. and Moran, T.P. Finding and Using Implicit Structure in Human-Organized Spatial Layouts of Information. *Proceedings of CHI'95* (1995) 346-353.

Shipman, F.M. and McCall, R.J. Incremental formalization with the hyper-object substrate. *ACM Transactions on Information Systems*, *17* (2) (1999) 199-227.

Snodgrass, A. and Coyne, R. Is Designing Hermeneutical? *Architectural Theory Review*, *1* (1) (1997) 65-97.

Stone, G.F. and McGinnis, M.L. Building scenarios in the next generation of simulations. *PROC IEEE INT CONF SYST MAN CYBERN*, *4* (1998) 3652-3657.

Sugimoto, M., Hori, K. and Ohsuga, S. An Application of Concept Formation Support System to Design Problems and a Model of Concept Formation Process. *Journal of Japanese Society for Artificial Intelligence*, *8* (5) (1993) 39-46.

Warren, R., Diller, D.E., Leung, A., Ferguson, W. and Sutton, J.L., Simulating Scenarios for Research on Culture & Cognition Using a Commercial Role-Play Game. in *Winter Simulation Conference, 2005 Proceedings of the* (2005), 1109-1117.