2D4-04

Algorithms for Scenario Generation Systems

Algorithms for Scenario Generation Systems

Shigeki Amitani[*1] Ernest Edmonds[*1]
Shigeki Amitani Ernest Edmonds

[*1] Creativity & Cognition Studios
Australasian CRC for Interaction Design
University of Technology, Sydney, AUSTRALIA

Creativity & Cognition Studios, Australasian CRC for Interaction Design
University of Technology, Sydney, AUSTRALIA

This paper presents design rationales and prototype of a generative system for visualising possible sequences of information to stimulate human creative thinking. 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. Thanks to the search technology, it has become easier to obtain information that a person is looking for, however, it is a laborious task to grab an overview of information space so that an information designer can find contextually relevant information pieces and sequence them into contextually meaningful ways. In this research, we design and develop a generative system that visualises possible contexts.

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.

The concept of the system derived from Knowledge Liquidization & Crystallization [Amitani '04, Hori, et al. '04]. Figure 1 shows expected interactions between users and the system.

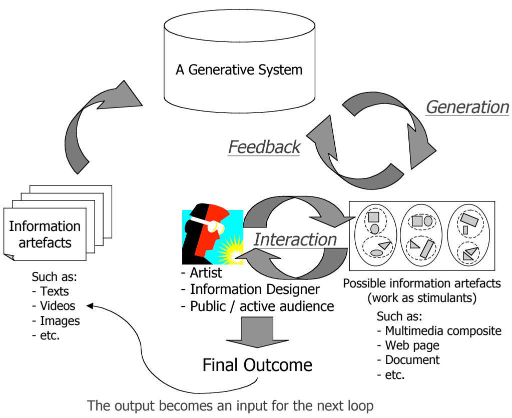

Figure 1 Expected Interactions

First, information artefacts (existing ones and/or new pieces of information) are collected and stored (left in Figure 1 Expected Interactions). The system (top in Figure 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 Figure 1.

These outputs work in two ways listed below:

- As final products that a user (audience) can enjoy
- As draft materials that a user (information designer) can modify (at the centre of Figure 1)

A generative system visualises possible contexts, i.e. possible sequences of information as "stimulants" [Edmonds '00] 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. As is often the case, enormous amount of web pages are retrieved. Then you use more keywords to improve the search result.

Problems here are:

- The search results are limited by viewpoints that you have already in your mind.
- 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.

2. Related Work

Humans naturally share knowledge by telling stories, sequences of information. This is a form of knowledge exchange and organise our experiences [Garvey '77].

Possible sequences of information segments, usually called contexts, scenarios, narratives, or storytelling, have been used as

Contact: Shigeki Amitani Creativity & Cognition Studios, University of Technology, Sydney, shigeki@shigekifactory.com

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 [Porter '85, Ringland '98, '02] and military operations [Chick, et al. '03, McKeever, et al. '06, Porter '85, Stone and McGinnis '98, Warren, et al. '05].

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* [Nonaka and Takeuchi '95]. 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 [Nonaka and Konno '03].

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 [Hori, et al. '04]. 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 [Hawkins and Blakeslee '05].

Hori et al. also proposed a conceptual system called Knowledge Nebula Crystallizer that supports the process of Knowledge Liquidization and Knowledge Crystallization [Hori, et al. '04]. 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., [Amitani '04]).

Automated scenario, narratives, and storytelling generation have been explored (e.g. [Kumar, et al. '06]). In military applications, the Campaign ontology was developed so that chains of commands and courses of actions can be compared with each other [McKeever, et al. '06, Stone and McGinnis '98].

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.

Our challenge is to design scenario generation systems used in domains where such established vocabulary sets are unavailable in advance.

Akaishi et al. [Akaishi, et al. '06] 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.

3. DESIGN RATIONALES

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 contextually correct ways

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

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) [Akaishi, et al. '06]. 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 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. 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. A PROTOTYPE SYSTEM

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 Cultural Agencies (IFACCA). They are renewing their web site[1] in order to provide more dynamic contents to audiences of the web site.

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

- Audiences of the web site can post their articles, comments, and video clips.
- The web site 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 a prototype system that generates sequences of information with actual data. The prototype system has been developed in Java 1.4.2 to test the algorithms and its visualisation form.

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

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 retrieved segments are arranged in a two-dimensional space. Each rectangle in a space represents a segment labeled with the main topic term of the segment. Segments are arranged in the space with a spring model [Theisel and Kreuseler '98]. The text area on the right side indicates the content of a selected segment.

The system retrieves segments and puts them into sequences. Figure 2 shows the search result with keyword "musician".

The circled segment in Figure 2 (labeled "right0" with the red background) 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 in Figure 2.

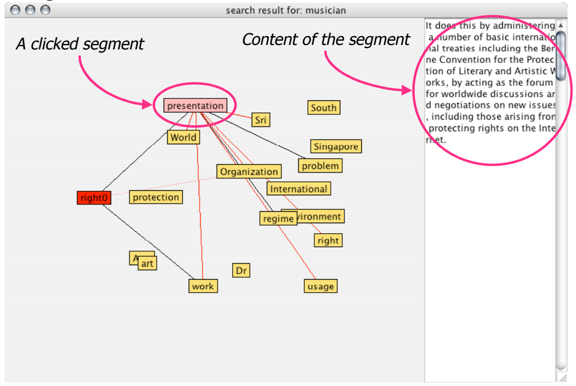

Figure 2 A Search Result for "musician"

The segments connected with the first segment with lines in Figure 2 (labeled 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 double-clicking with pressing down the shift key, the system shows the entire selected sequences so that the user can actually read and understand the content of the sequence (Figure 3). If a user shift-double-clicks a segment labeled "presentation" which is the second segment following the first segment labeled "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 (Figure 3).

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.

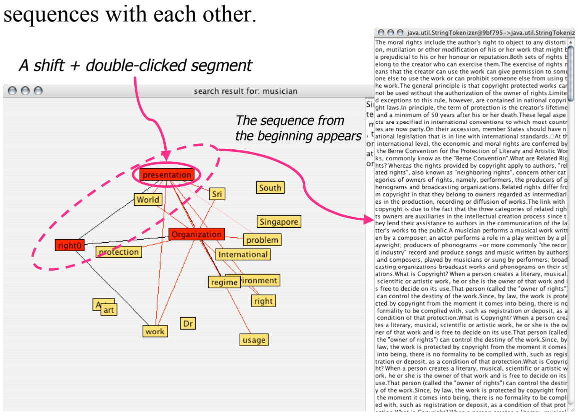

Figure 3 Shift+Double-Clicking for Connected Content

Our first impression is that the generated sequence is consistent and seems to be "a story"[1]. This result is composed of three segments whose main topics are "right", "Organization", and "protect" respectively, searched by the word "musician". The

[1] An example of generated sequence is available at http://shigekifactory.com/research/GenerativeSystem/ExampleSequence.html.

sequence is "right" => "Organization" => "protection". 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.

5. DISCUSSIONS

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 [Hori, et al. '04]. 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.

The system provides an overview of possibilities of sequencing information by the spatial representation, and 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 [Snodgrass and Coyne '97].

6. CONCLUSION

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

Main Topic Term Dependency was proposed as criteria for sequencing information segments as a consistent context. Also we presented a prototype system with a preliminary result. We are developing the system further to generate longer sequences.

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. Furthermore, while the algorithm presented in this paper is for text-based information, we are going to extend this algorithm for non-text-based information such as videos and music.

Acknowledgements

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.

References

[Akaishi, M., Hori, K. and Satoh, K. Year] 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. 2004] Amitani, S. A Method and a System for Supporting the Process of Knowledge Creation, Department of Advanced Interdisciplinary Studies, University of Tokyo, 2004.

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

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

[Garvey, C. 1977] Garvey, C. *Play*. Open Books, London, 1977.

[Hawkins, J. and Blakeslee, S. 2005] Hawkins, J. and Blakeslee, S. *On Intelligence*. Owl Books, 2005.

[Hori, K., Nakakoji, K., Yamamoto, Y. and Ostwald, J. 2004] 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). 252-261 (2004).

[Kumar, D., Ramakrishnan, N., Helm, R.F. and Potts, M. 2006] 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*. 604-610 (2006).

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

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

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

[Porter, M.E. 1985] Porter, M.E. *Competitive advantage*. Free Press New York, 1985.

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

[Ringland, G. 2002] Ringland, G. *Scenarios in business*. Wiley, Chichester, 2002.

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

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

[Theisel, H. and Kreuseler, M. 1998] Theisel, H. and Kreuseler, M. An Enhanced Spring Model for Information Visualization. *Computer Graphics Forum*, *17* (3). 335-344 (1998).

[Warren, R., Diller, D.E., Leung, A., Ferguson, W. and Sutton, J.L. Year] Warren, R., Diller, D.E., Leung, A., Ferguson, W. and Sutton, J.L., Simulating Scenarios for Research on Culture & Cognition Using a Commercial Role-