A Method for Visualising Possible Contexts

Shigeki Amitani and Ernest Edmonds

Creativity & Cognition Studios,
Australasian CRC for Interaction Design,
University of Technology, Sydney
PO Box 123 Broadway NSW 2007, AUSTRALIA

Abstract. This paper presents a method for visualising possible sequences of information. Possible sequences of information segments, called contexts, scenarios, narratives, or storytellings, have been used as a tool for exploring and stimulating thinking about possible events, assumptions relating to these events, and courses of actions, i.e. chances 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 series of generative systems that visualises possible contexts.

Key words: Generative systems, context, scenario generation

1 Introduction

We are currently developing a system that visualises possible sequences of information out of existing information in order to increase "chances" for information designers to discover possible contexts that otherwise could be missed.

The concept of the system derived from Knowledge Liquidization & Crystallization[1][2]. Figure 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 Figure 1). The system (top in Figure 1) segments the input information artefacts into small semantic units, then it generates restructure them into possible meaningful sequences of information shown in the right of Figure 1.

These outputs work in two ways such as:

1. As final products that a user can enjoy
2. As draft materials that a user can modify (at the centre of 1)

   A generative system visualises possible contexts, i.e. possible sequences of information as "stimulants" [3] regarding to a core theme that a user inputs.
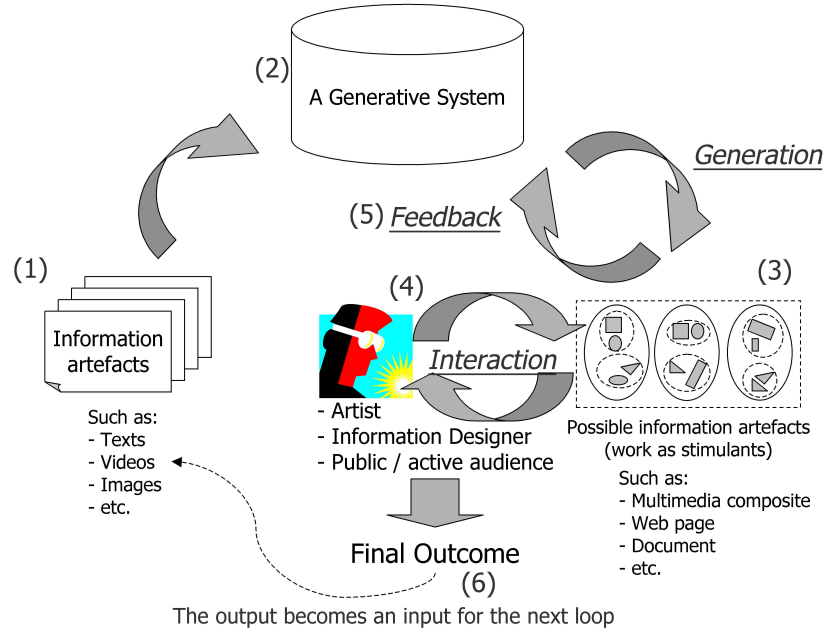
Fig. 1. How a Generative System Works

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: (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 propose a method to visualise possible sequences of information to challenge these problems.

2   Related Work

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 [1]. 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 this process.

They also proposed a conceptual system called Knowledge Nebula Crystallizer that supports the process of Knowledge Liquidization and Knowledge Crystallization [1]. 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., [2][4][5]).

Akaishi et al. [6] 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 where a user explores possible contexts.

3   Design Rationales for 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 be:

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

Following computations support the interaction:

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) [6]. 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:

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

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.

3.3   Representing the generated sequences

The generated sequences are arranged in a two-dimensional space. A number of studies that retrieves information pieces containing designated keywords have been done. The retrieved are usually arranged close to each other to show they are relevant.

This interaction, as implemented in most of other applications, provides an opportunity for users to explore information space, however, it does not support how the retrieved information are inter-related with each other, hence it does not support the process of restructuring existing information so that the user can produce new information artefacts. The user has to actually remember what he/she read in order to articulate the relationships among them.

We propose a representation shown in Figure 2. Each context is represented as a sequence of generated segments. If a document is shared by two contexts, as circled in Figure 2, that means that two contexts are inter-related with each other. We expect that crossing multiple contexts via the document yields a new context.

4   An Application: Generative Website

We are currently working on the Generative Website Project supported by the Australasian CRC for Interaction Design (ACID). This is a collaborative work with the International Federation of Arts Councils and Culture 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 followings:

– 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 grab the overview of the contents

We have implemented the algorithms listed above and applied to actual information. The web site of IFACCA has a number of PDF files which are mainly composed of reports from art conferences and news articles. As a first test of
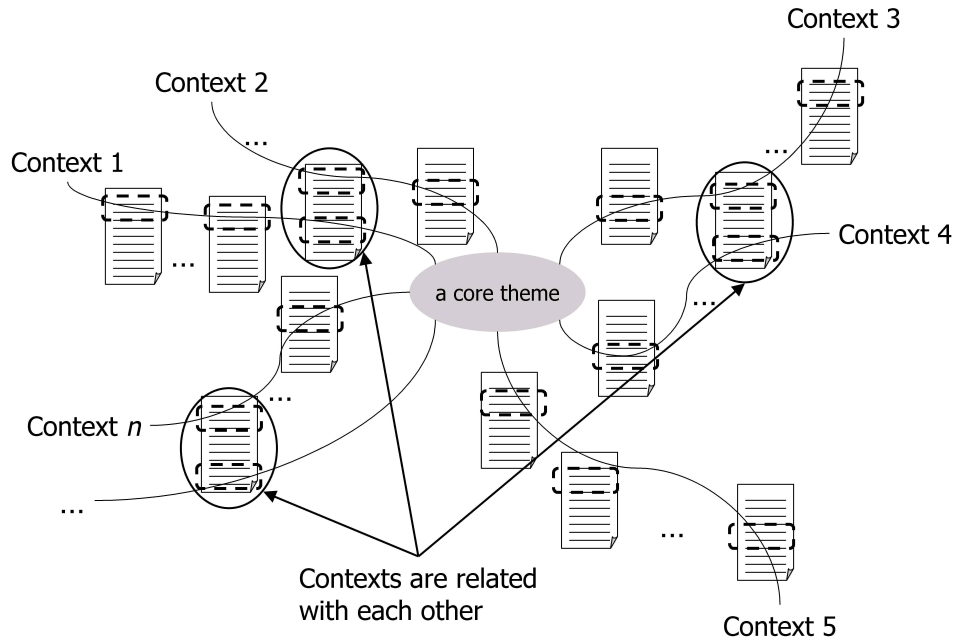
---

[1] http://www.ifacca.org/

Fig. 2. Each one have to be read to know which is what

the prototype system, text data from 10 PDF files was excerpted and converted into 10 text files. The system segments them into 57 segments.

A scenario here is that a user is searching "a story" about "musician" on the IFACCA web site. The result is shown in Figure 3. Note that at the time of writing this paper, the system does not have a graphical user interface. The image of this result is shown in Figure 4.

The user inputs a keyword "musician", then the system returns two segments that contain the keyword. The main topics of the retrieved segments are "minister" and "right". Then the system searches other main topic terms "Culture" and "work" on which the two retrieved main topic terms are dependent most. The generated sequences show that this web site discusses "musicians" in two contexts[2]:

– Government's actions toward cultural affairs
– Intellectual property issues

Although this result is preliminary that has only two steps and has not been evaluated, our impression is that the system has a potential to connect segments in meaningful ways.

---

[2] Due to the page limitation, the generated texts in this preliminary test are available at: http://shigekifactory.com/research/GenerativeSystem/PreliminaryTestResult.html

```
Search term: musician
foundMainTopicVector.elementAt(0): minister        → Retrieved Main Topic Terms
foundSegmentFileNameVector.elementAt(0):
/Users/amitani/shigeki/ScenarioGeneration/IFACCA/IFACCA10text/3Theroleofcultureinforeign
affairshughstephensb/3Theroleofcultureinforeignaffairshughstephensb-minister.segtt
foundMainTopicVector.elementAt(1): right
foundSegmentFileNameVector.elementAt(1):
/Users/amitani/shigeki/ScenarioGeneration/IFACCA/IFACCA10text/9WIPOGeoffreyYuENG9/9WIPOG
eoffreyYuENG9-right0.segtt
SegmentWordVector.java: done.
maxMainTopicTermVector.elementAt(0): Culture          → Retrieved next main topic terms
Next segment file is:                                 → Retrieved next segments
/Users/amitani/shigeki/ScenarioGeneration/IFACCA/IFACCA10text/3Theroleofcultureinforeign     for "Culture"
affairshughstephensb/3Theroleofcultureinforeignaffairshughstephensb-Culture.segtt
maxMainTopicTermVector.elementAt(0): work
Next segment file is:
/Users/amitani/shigeki/ScenarioGeneration/IFACCA/IFACCA10text/7IntellectualPropertyMahin
daRa___ENG7/7IntellectualPropertyMahindaRa___ENG7-work.segtt
Next segment file is:
/Users/amitani/shigeki/ScenarioGeneration/IFACCA/IFACCA10text/7IntellectualPropertyMahin
daRa___ENG7/7IntellectualPropertyMahindaRa___ENG7-work0.segtt
Next segment file is:
/Users/amitani/shigeki/ScenarioGeneration/IFACCA/IFACCA10text/9WIPOGeoffreyYuENG9/9WIPOG
eoffreyYuENG9-work.segtt
Next segment file is:
/Users/amitani/shigeki/ScenarioGeneration/IFACCA/IFACCA10text/9WIPOGeoffreyYuENG9/9WIPOG
eoffreyYuENG9-work0.segtt
Next segment file is:
/Users/amitani/shigeki/ScenarioGeneration/IFACCA/IFACCA10text/9WIPOGeoffreyYuENG9/9WIPOG
eoffreyYuENG9-work00.segtt
MainTopicTermDependency.java: done
```

Retrieved next segments for "World"

Fig. 3. A Result

5   Conclusion

In this paper, we presented a method of for visualising 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.

At the time when this paper is written, as the algorithms for the system has just been implemented, it does not have a graphical representation[3]. The interface will be developed for actual user studies. Also we are going to evaluate the algorithms and results through user studies to see the effectiveness of the system. We recognise that this algorithm is still argueable and that other sequencing algorithms (e.g. [7]) could produce semantically more meaningful sequences. These are the future works.

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

---

[3] At the time when this camera-ready version is written, a prototype system has been developed. However, describing the prototype and its interactions are not explained in this paper as it is beyond the accepted contents.
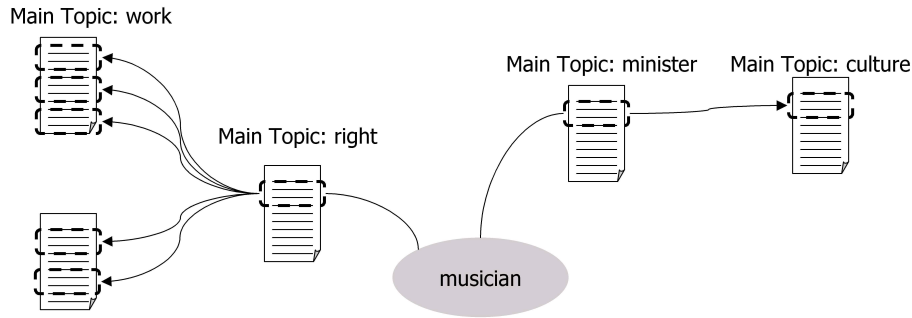
Fig. 4. An Image of the Result

Acknowledgements

References

1. Koichi Hori, Kumiyo Nakakoji, Yasuhiro Yamamoto, and Jonathan Ostwald. Organic perspecives of knowledge management: Knowledge evolution through a cycle of knowledge liquidization and crystallization. *Jounal of Universal Computer Science*, Vol. 10, No. 3, pp. 252–261, 2004.
2. Shigeki Amitani and Koichi Hori. Knowledge nebula crystallizer for knowledge lequidization & crystallization - from a theory to a methodology of knowledge management. *Proceedings of Expertise In Design, Design Thinking Research Symposium 6*, 2003.
3. Ernest Edmonds. Artists augmented by agents (invited speech). *Proceedings of the 5th international conference on Intelligent user interfaces New Orleans Louisiana United States*, pp. 68–73, 2000.
4. Shigeki Amitani. *A Method and a System for Supporting the Process of Knowledge Creation*. PhD thesis, Department of Advanced Interdisciplinary Studies University of Tokyo, 2004.
5. Shigeki Amitani and Ernest Edmonds. Knowledge nebula crystallizer for time-based information. In *Proceedings of Creativity & Cognition 2005, London, England, 12-16 April*, 2005.
6. Mina Akaishi, Koichi Hori, and Ken Satoh. Topic tracer: A visualization tool for quick reference of stories embedded in document set. In D. A. Keim, F. Mansmann, J. Schneidewind, H. Ziegler, C. Tominski, J. Abello, F. van Ham, H. Schumann, M. Jern, and J. Franz, editors, *International Conference on Information Visualisation 2006*, pp. 101–106, London, UK, 2006.
7. Yoko Nishihara, Wataru Sunayama, and Masahiko Yachida. Information acquiring support system based on keywords' continuity and informational difficulty. In *International Conference on Human-Computer Interaction*, Las Vegas, US, 2005.