A Platform for Mining and Visualizing Regional Collective Culture

Shin Ohno, Shinya Saito, and Mitsuyuki Inaba

Digital Humanities Center for Japanese Arts and Cultures,
Ritsumeikan University
56-1 Tojiin-Kitamachi, Kita-ku, Kyoto 603-8577, Japan
{shin,ken,inabam}@arc.ritsumei.ac.jp

Abstract. This paper proposes computational methods for mining and visualizing collective culture among the community members of a region. This paper first outlines a procedure to extract significant narratives with text mining technique and spatiotemporal analysis on the textual data transcribed from oral-history interviews with the regional community members. It also introduces the KACHINA-CUBE system that imports the narratives as contextualized fragments of sentences based on spatiotemporal information, visualizes them onto a virtual 3D space, and assist researchers to discover commonalities and diversities among them based on the trajectory equifinality model (TEM), which is a theoretical framework to clarify both the similarities and differences among the trajectories of individual life courses. At the end of this paper, we illustrate a test case on collective culture regarding the once-flourishing film industry in Kyoto.

Keywords: collective culture, personal culture, text mining, narrative analysis, qualitative GIS.

1 Introduction

A new interdisciplinary approach, Cultural Computing [9] focuses on applying computer technology for preserving, sharing, and analyzing tangible and intangible culture. Since the concept of culture is broad, complex, and ambiguous, it is important for researchers to recognize their perspective or scope for approaching to culture before they apply computing technology to cultural phenomena or cultural artifacts.

In this paper, we set our viewpoint on culture by adopting the concept of culture proposed by J. Valsiner [10]. A socio-cultural psychologist, he argues that the culture is a dynamically emerging phenomenon that is co-constructed by the interplay between "personal culture" and "collective culture." The former consists of personal sense, memory and knowledge, while the latter is considered as collection of the former. There exists dynamic interplay between personal and collective cultures, and they go through the process of "internalization" and "externalization." Through the process of communication with others, personal

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 188–199, 2010.

culture is externalized. In other words, shared within the community, the personal culture affects the collective counterpart. On the other hand, collective culture is internalized to the individual's sense and personal culture.

Since both personal and collective cultures have intangible, complex, and ambiguous characteristics, capturing them in socio-cultural contexts is one of challenging tasks for researchers. To understand this contextual feature of culture, we specifically focus on spatiotemporal information as a key to capture personal and collective cultures. When people recount their personal experience, they frequently give utterances directly or indirectly related to spatiotemporal information. We also found different people sometimes refer to the same location or time in their stories. Intersection of such stories could be considered as a clue to understand collective culture. Although personal culture may not be externalized directly, we could have a glimpse of the collective culture from collection of independent stories as a result of indirect externalization of personal culture.

In the following, we propose a method to extract the information for personal culture from transcribed oral histories. The extraction procedure is followed by the introduction of the KACHINA CUBE (KC) system, a web-based platform for visualizing and mining regional collective culture. We, then, test KC with a case of collective culture from three research participants, regarding the once-flourishing film industry in Kyoto. Finally, we evaluate our methods and the platform by comparing to the related works.

2 Mining Personal and Collective Cultures

This section elucidates the procedure of extracting information of personal and collective culture from the narrative data. In reality, the extraction process follows the research that involves collecting narratives from research participants. It should be also mentioned that the linguistic processing in this section is partially assuming the Japanese textual data transcribed from oral-history interviews with the regional community members in a certain area of Kyoto, although the main part of the following procedure would be applicable to other languages.

2.1 Extracting Personal Culture

We consider the spatiotemporal information appeared in narrative data would be a guide to split them into chunks related to personal culture. Fig.1 explains procedure to extract personal culture from the narrative data. The procedure contains morphological analysis, syntactical analysis, and spatiotemporal analysis. After obtaining the narrative data, researchers first perform morphological analysis on them, splitting them into morpheme units. Using both the basic and user dictionaries, we can detect morpheme units from the original transcripts. To explain this in more detail, the user dictionary covers miner or specific terms that are not registered in the basic dictionary. For example, "Uzumasa-Eigamura," or Uzumasa Movie Village, indicates one specific location, which the basic dictionary might not register as a proper noun. Such names of locations, events, persons, and buildings are considered as entries for the user dictionary.

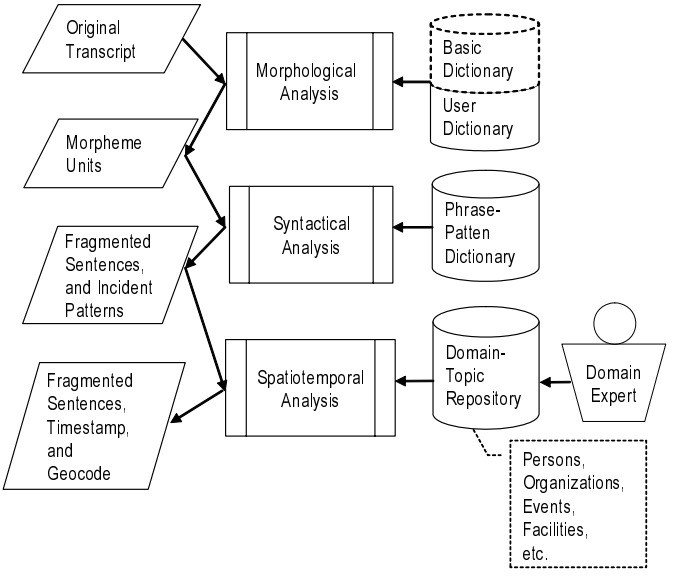

Fig. 1. The flow of extracting personal culture from the original transcripts. The whole process entails three phases of analyses, morphological, syntactical, and spatiotemporal.

The next step is syntactical analysis that examines the transcripts based on the phrase pattern dictionary that has frame sets of incidents. The analysis outputs fragmented sentences with incident patterns. For example, the sentence, "My uncle started Makino Cinema Theater at Kyoto Senbon area". would be detected by a syntactical analysis. In this case, the phrase pattern dictionary has one entry with temporal information, which is "starting Makino Cinema Theater," and another with spatial information of "Kyoto Senbon area".

Final process is spatiotemporal analysis that maps between spatiotemporal information and incident patterns. Meta information such as timestamp and geocode is added to the fragmented sentences. For this process, we propose the domain-topic repository, which is developed based on the information from domain experts in order to define the semantic associations between persons, organizations, events, facilities and so on. The linkages from the information to the spatiotemporal information are also stored in the repository. With the repository, it becomes possible for the researchers to traverse the terms or phrases in the narratives and add meta information to the fragmented sentences.

In short, our way of analyzing personal culture makes use of three different databases: the user dictionary for morphological analysis, the phrase-pattern dictionary for syntactical analysis, and the domain-topic repository for spatiotemporal analysis.

2.2 Bridging Personal Culture to Collective Culture

The databases we mentioned in the previous section are reusable for researching the same periods, the same areas, or both. The same spatiotemporal information,

such as names of buildings and events are more likely to turn up if referred locations are close enough. Since the dictionaries and domain-topic repository are reusable, the more pieces of information are stored to them, the easier the process gets to extract spatiotemporal information from other sources.

The collection of fragmented sentences based on the narrative data from the regional community members can be considered as a clue to understand collective culture in the region. As we discussed, even though the community members did not meet each other those days, and did not go through the externalization process as Valsiner proposes, it is highly likely that they share the similar experience, which means there exists the possibility of indirect externalization. For example, a research participant has some story about his/her experience in the elementary school, and another research participant went to the same elementary school in the similar time period. They might or might not have known each other then, but it is likely that they had some friends, teachers, and school events in common. Through such intermediates, each personal culture gets externalized and shared by the community. In the end, the collection of personal stories based on spatiotemporal information can be considered as collective culture like this case.

3 Visualizing Collective Culture

The following sections introduce the KC system as a platform for visualizing regional collective culture. It imports the contextualized fragments of sentences generated by the procedure above, visualizes them onto a virtual 3D space based on spatiotemporal information, and assists researchers to understand collective culture shared by the community members in a region.

3.1 User Interface

The KC system provides a web-based virtual 3D space to its user (see Fig. 2). The space includes a two dimensional map with timescale, and is used to store and plot the contextualized fragments of sentences extracted from the narrative data based in spatiotemporal information.

Fig. 3 shows a 3D space from two different angles. It consists of a map with timescale. The user can rotate the 3D space to browse the fragmented sentences or narratives from suitable angles.

3.2 Data Structure

Expecting KC to be sharable, we apply RDF (Resource Description Framework)/ OWL (Web Ontology Language) [12] to define the classes for storing narrative data. RDF and OWL are Semantic Web standards that provide a framework for sharing and reuse of data on the Web. Fig. 4 shows a part of the class definition in KC. Under the **Document** class, the following classes are defined.

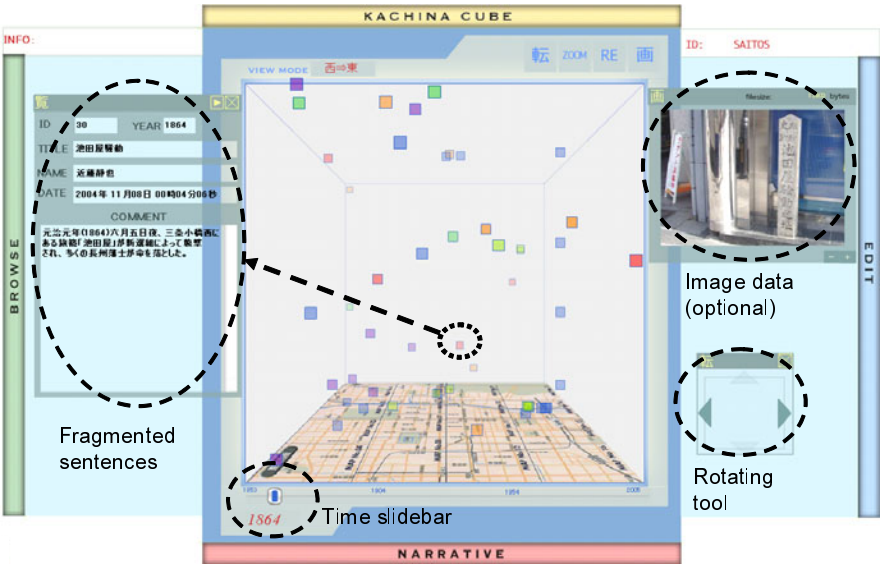

Fig. 2. User interface of the KC system. The central window shows a 3D space with a map and markers to fragmented sentences. The user can define links to images or video clips to the markers. The user also can turn around the 3D space with the rotating tool.

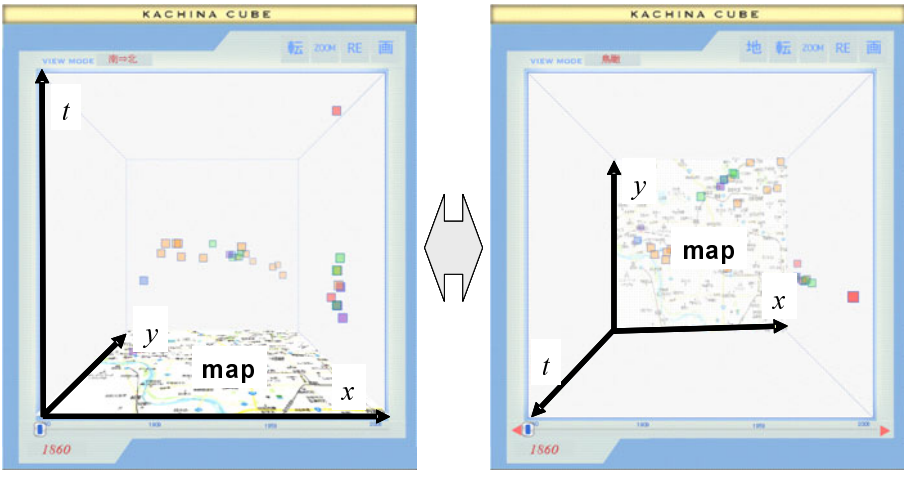

Fig. 3. A virtual 3D space from two different angles. It consists of a map (dimension x and y) with timescale (dimension t).

Fragment class:
Generic class for fragmented sentences
StoryFragment class:
Fragmented sentences from narrative data
HistoryFragment class:
Topics that are considered as historical fact
Storyline class:
Contextualized data consists of story/historical fragments

The **Attribute** class is used to represent the spatiotemporal information of the narrative data in KC. It has the followings two subclasses.

Geography class:
Geographical information of story and historical fragments
Temporal class:
Time related to the incidents in story or historical fragments

Since each fragment has both spatial and temporal information, the standard format is suitable for KC and motivates other researchers to access the data. It should be noted that this data format allows us to define semasiological relationship and class structure among fragmented sentences.

3.3 OPP Detector

Valsiner and Sato [11] proposed the concept of Trajectory Equifinality Model (TEM) as a framework to analyze personal experience with diverse and possible trajectories. One of the components to support TEM, OPP (Obligatory Passage Point) is the point which most of the people have to go through because of their socio-cultural contexts, institutions, and customs.

Implemented in KC, OPP detector can find fragmented sentences related to similar space or time in multiple narratives provided by research participants. Fig. 5 illustrates how KC finds OPP among the narratives. Narratives A to D in the figure go through similar space and time. OPP detector of KC recognizes the intersection of the narratives as OPP. In other words, OPP detector assists researchers' understanding on regional collective culture based on the analysis on multiple personal life courses.

4 Application

4.1 A Test Case of Implementation

To advocate the method, we implemented tools and databases to extract the information related to personal culture based on the method we discussed above, and imported the extracted data into the KC system.

As a test case, we used the narrative data of movie culture in Kyoto's Rakusai area, a.k.a., Japan's Hollywood. We utilized oral history data collected by [6], [7],

```
<owl:Class rdf:ID="Document"> </owl:Class>

<owl:Class rdf:ID="Fragment">
  <rdfs:subClassOf rdf:resource="#Document" />
</owl:Class>

<owl:Class rdf:ID="StoryFragment">
  <rdfs:subClassOf rdf:resource="#Fragment" />
</owl:Class>

<owl:Class rdf:ID="HistoryFragment">
  <rdfs:subClassOf rdf:resource="#Fragment" />
</owl:Class>

<owl:Class rdf:ID="StoryLine">
  <rdfs:subClassOf rdf:resource="#Document" />
</owl:Class>

<owl:ObjectProperty rdf:ID="consistOf">
  <owl:domain rdf:resource="#StoryLine" />
  <owl:range rdf:resource="#Fragment" />
</owl:ObjectProperty>

<owl:Class rdf:ID="Attribute"></owl:Class>

<owl:Class rdf:ID="Geography">
  <rdfs:subClassOf rdf:resource="#Attribute" />
</owl:Class>

<owl:Class rdf:ID="Temporal">
  <rdfs:subClassOf rdf:resource="#Attribute" />
</owl:Class>
```

Fig. 4. RDF/OWL representation of class definition in KC

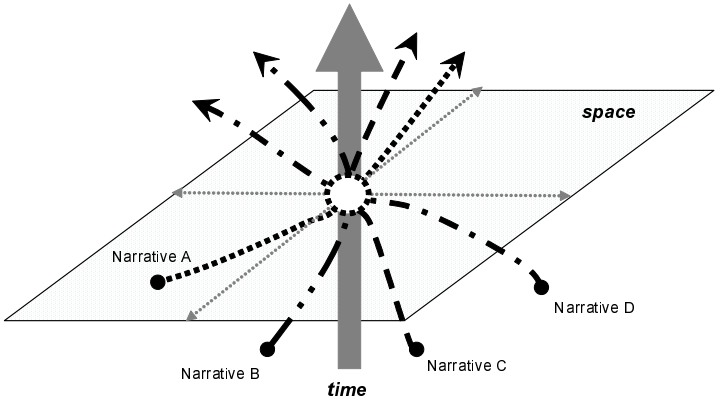

Fig. 5. Image of OPP as an intersection of the trajectories of multiple narratives provided by research participants

and [8]. Those oral history data were collected from three research participants who had involved in the movie industry in the area from the 1910s to the 1930s. Each research participant did not refer to each other in the data, but their experiences are considered as connected in some cases. Therefore, extracting personal cultures from all three research participants can give us some ideas about collective culture of that area. The step-by-step procedures to use the method are as follows:

1. Read the oral histories and collect direct spatiotemporal information from them.
2. Enter the direct spatiotemporal information to the user dictionary.
3. Enter the indirect spatiotemporal information such as names of persons, buildings, and events to the phrase-pattern dictionary.
4. Extract fragmented sentences from both morphological and syntactical analyses by using the dictionaries.
5. Add the domain-topic repository to map the spatiotemporal information to computer-readable data.
6. Add timestamp and geocode to the fragmented sentences.
7. Store the data to KC, and see how it works.

4.2 Analyzing Personal Culture

For morphological analysis, we add the user dictionary to the MeCab program [4]. Furthermore, to make the syntactical analysis easier, we also add the phrase-patterns to the program. After doing so, we prepare a tool for these two analyses to split text into lines which can go through the line-by-line analysis. As an experiment, we set the threshold parameter as three lines, which are treated as a unit of fragmented sentences. With the domain-topic repository, we add timestamp and geocode data to the fragmented sentences, extracted from the first two analyses.

4.3 Visualizing and Exploring Collective Culture

After the procedures above, we make instances of the StoryFragment class from the fragmented sentences with spatiotemporal information. The instances of Storyline class are generated based on the contextual information of the fragmented sentences. Fig. 6 shows a result of plotting those instances on KC.

We could get the narrative data from three research participants in this test case. Since some stories in their narratives were referring to a similar time and place, KC's OPP detector could automatically find the intersection point of their storylines in the virtual 3D space (see Fig. 7). With the assist of the OPP detector, we could know a focal point to start exploring the sharable experiences in the community, and get a clue to understand the indirect collective culture regarding movie industries in Kyoto.

In this test case, KC found one OPP where three storylines from three different research participants went through a similar time and place in the virtual 3D

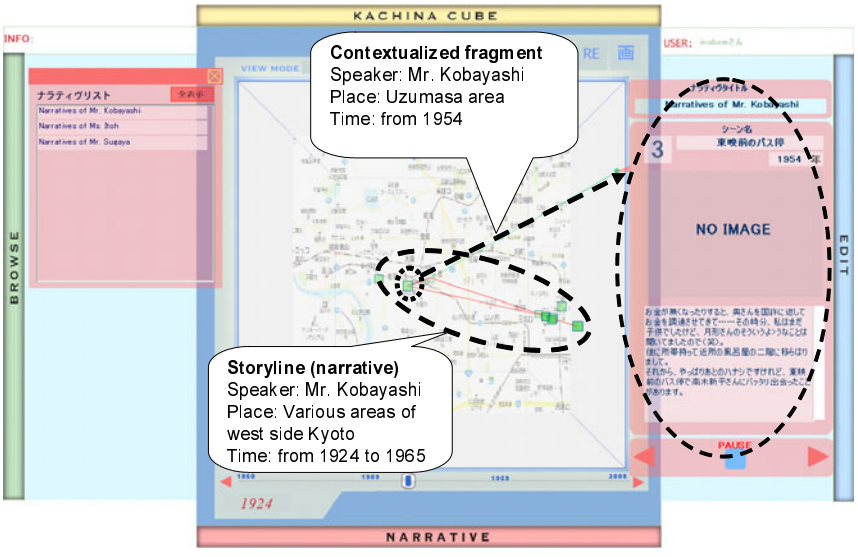

Fig. 6. Fragmented sentences and storylines plotted in the virtual 3D space of KC

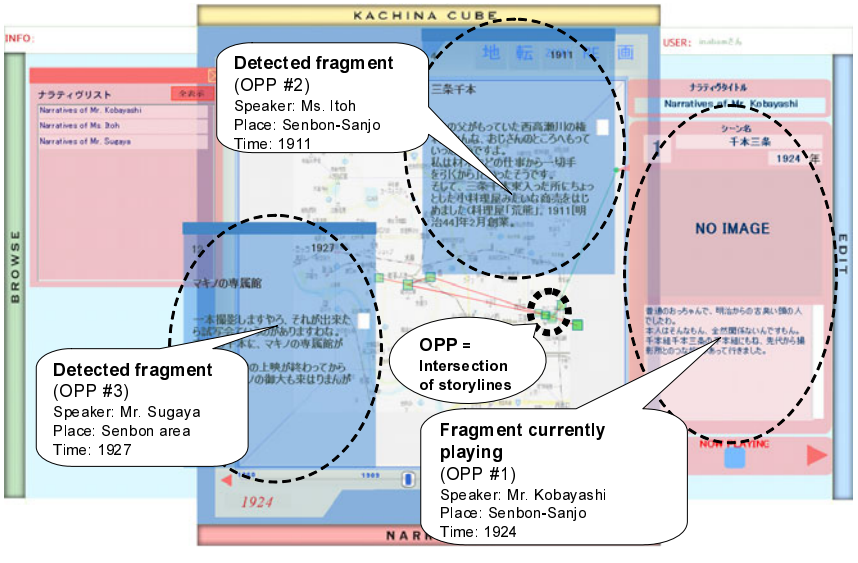

Fig. 7. OPP detector finds an intersection among storylines from three research participants

space. The point indicates the Senbon-Sanjo area in the central Kyoto from the 1911s to the 1927s. According to the storyFragments provided by research participants, there used to be some movie theaters and an enterprise related with movie industry in that area. Since there are no movie theaters or movie-related firms in the area at the present day, the place and time can be considered as a crucial point to explore or restore collective culture regarding the once-flourishing film industry in Kyoto.

5 Related Works

Our approach is closely related to that of GIS-based narrative analysis. For example, Kwan et al. [5] develop the method called "geo-narrative," which extends current GIS capabilities to analyze and interpret narrative materials such as oral histories, life histories, and biographies. In this method, researchers are supposed to go over narratives line by line, interpret sentences, and then plot them onto GIS space. Although it provides a precise way to examine narrative materials, its process of data plotting cannot exclude the researchers' subjective interpretation.

Other researchers are exploring the possibilities of automatic geocoding from text data. For example, Yao et al. (see [13], [14], and [15]) have been working on the fundamental research on "qualitative location (QL)" which refers to a spatial location using linguistic terms such as qualitative spatial descriptions. Their research proposes a way to process multiplicity and uncertainty in QL described in natural language. Angel et al. [1] also advocate a methodology for the semiautomatic geocoding of persistent Web pages in the form of collaborative human intervention to improve on automatic geocoding results. While their techniques are universally applicable, their approaches cannot handle various proper nouns.

CulturalSampo [3] is an integrated virtual museum with the combination of semantic web technology and web-based geographical system. It allows its user to browse cultural artifacts in a specific region in a certain time. The target domain of this system, however, is the public history or cultural artifacts which are regularly exhibited in real or virtual museums.

Different from the researches above, our approach uniquely combines text mining and semantic web technology to extract spatiotemporal information from the narrative data, and helps the researcher understand personal and collective cultures with the visualization of narratives in the virtual 3D space.

6 Conclusion

Taking Valsiner's socio-cultural definition of culture, this paper focused on spatiotemporal information to capture personal and collective cultures, since such information regularly appears in people's life stories. To analyze the culture, we propose a method to extract personal culture from their stories.

We analyzed oral histories, narrative data from three research participants in Kyoto's Rakusai area, and stored extracted spatiotemporal information to

KC in order to examine our method. As a result, KC demonstrates significant potentials for study of regional culture, based on spatiotemporal information. Reviewing similar researches in Cultural Computing, we evaluate our original method to tackle extracted culture.

As the next step of our research, we have to work on three major challenges. First, we have to enhance our user dictionary, pattern-phrase dictionary, and domain-topic repository to handle the wide range of stories. Since these databases are reusable for researchers interested in similar areas and times, collaborative development and sharing of such databases could contribute to enhance a variety of socio-cultural research activities. Secondly, we still have trouble dealing with ambiguous and fuzzy spatiotemporal information. As GIS researchers are experiencing similar problems, we should collaborate with each other to develop techniques or corpus to handle such ill-defined spatiotemporal information. Finally, we need to apply our method and platform to more cases for enhancing generalizability of our way on mining and visualizing collective culture.

Acknowledgments. The authors would like to thank Prof. Mika Tomita and her colleagues who provided the narrative data for our research. This research has been partially supported by Digital Humanities Center for Japanese Arts and Cultures, Ritsumeikan University.

References

1. Angel, A., Lontou, C., Pfoser, D., Efentakis, A.: Qualitative geocoding of persistent web pages. In: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, Irvine, California (2008)
2. Arpinar, I.B., Sheth, A., Ramakrishnan, C., Usery, E.L., Azami, M., Kwan, M.: Geospatial ontology development and semantic analytics. Transactions in GIS 10, 551–575 (2005)
3. Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen, T., Frosterus, M., Sinkkilä, R., Paakkarinen, P., Laitio, J., Nyberg, K.: CultureSampo – Finnish Culture on the Semantic Web 2.0: Thematic Perspectives for the End-user. In: Museums and the Web 2009, Indianapolis, USA (2009)
4. Kudo, T.: MeCab: Yet Another Part-of-Speech and Morphological Analyzer, `http://mecab.sourceforge.net/`
5. Kwan, M.-P., Ding, G.: Geo-Narrative: Extending Geographic Information Systems for Narrative Analysis in Qualitative and Mixed-Method Research. The Professional Geographer 60, 443–465 (2008)
6. Tomita, M., Itakura, F.: Voices from Kyoto: Interview with ITO Asako – An aspect of Japanese film history. Art Research 1, 127–138 (1999)
7. Tomita, M., Kamiya, M., Gondo, C.: Voices from Kyoto: Interview with KOBAYASHI Masanori – An aspect of Japanese film history. Art Research 2, 115–136 (2000)
8. Tomita, M., Kamiya, M., Oya, A.: Voices from Kyoto: Interview with SUGAIE Koyo - An aspect of Japanese film history. Art Research 3, 241–270 (2001)

9. Tosa, N., Matsuoka, S., Ellis, B., Ueda, H., Nakatsu, R.: Cultural Computing with Context-Aware Application: ZENetic Computer. In: Kishino, F., Kitamura, Y., Kato, H., Nagata, N. (eds.) ICEC 2005. LNCS, vol. 3711, pp. 13–23. Springer, Heidelberg (2005)
10. Valsiner, J.: Personal Culture And Conduct Of Value. Journal of Social, Evolutionary, and Cultural Psychology 1, 59–65 (2007)
11. Valsiner, J., Sato, T. (eds.): Historically Structured Sampling (HSS): How can psychology's methodology become tuned in to the reality of the historical nature of cultural psychology?. Transcript Verlag, Bielefeld (2006)
12. The World Wide Web Consortium: OWL Web Ontology Language Overview, http://www.w3.org/TR/owl2-overview/
13. Yao, X., Jiang, B.: Visualization of qualitative locations in geographic information systems. Cartographic and Geographic Information Science 32, 219–229 (2005)
14. Yao, X., Thill, J.-C.: Spatial queries with qualitative locations in spatial information systems. Computers, Environment and Urban Systems 30, 485–502 (2006)
15. Yao, X., Thill, J.-C.: Neuro-fuzzy modeling of contextcontingent proximity relations. Geographical Analysis 39, 169–194 (2007)