3D Visualization
of Temporal Changes in Bloggers' Activities and Interests

Masahiko ITOH   Naoki YOSHINAGA   Masashi TOYODA   Masaru KITSUREGAWA

Institute of Industrial Science, University of Tokyo

ABSTRACT

This paper presents a novel system for analyzing temporal changes in bloggers' activities and interests on a topic through a 3D visualization of dependency structures related to the topic. Having a dependency database built from a blog archive, our 3D visualization framework helps users to interactively exploring temporal changes in bloggers' activities and interests related to the topic.

1 INTRODUCTION

Social media such as blog and twitter enables users to easily and rapidly publish personal activities and interests. Now that they have been considered as an important diachronic data for marketing, linguistic, and sociology, which implies a growing demand to support their analysis via visualization [3].

The current systems for analyzing the blogosphere [1] focus on the topic detection in the blogosphere, and provide a little information that explains *why* (in what contexts) the topic becomes popular. Although some systems supplementarily provide a set of words (e.g., Hatoyama) related to the given topic (e.g., 'Obama'), they are insufficient; the users are forced to read the individual posts on the topic to understand the contexts between the related words.

This paper proposes a system for visualizing temporal changes in bloggers' activities and interest through dependency analysis of contents in a blog archive. Our system could directly relate a given topic with other related words via the *events* themselves (e.g., "Obama *meets* Hatoyama"). We can thereby gain a deeper insight into what is actually happening on the given topic. The intended users of our system include not only professional analysts who have a prior knowledge about the topic for marketing and/or social problems but also casual social media users who may find out their interesting topics using the other analysis systems to explore the detail of the topic.

2 DEPENDENCY ANALYSIS AND DATABASE

We construct a dependency database that consists of *events*, each of which is defined as a set of dependency relations on a particular verb, from the blog archive. Hereafter, we regard a set of dependency relations between a verb and its dependents (nouns) as a target event. The verb represents an action or a state of being (in which the bloggers may have an interest) and the nouns modifying the verb further characterizes the action/state. We can therefore grasp details of a particular event anchored by a verb, by collecting a set of dependency relations between the verb and its dependents.

We have collected Japanese blog feeds since Feb. 2006; currently the archive consists of over 2 million blogs with in total over 350 million posts. We first analyze the sentences in the blog archive by using a dependency parser[1] to collect dependency relations that represent events. We next build a dependency database for each

---

month. The dependency database allows us to search events that involve a given topic (noun). We preclude the details of the database due to space limitations.

To summarize events for the topic, we categorize each event based on a thematic role (e.g., agent, experiencer, theme and location) of the noun for the verb. In Japanese, we can exploit case particles (e.g., '-ga', '-wo', '-ni', '-to', and '-de') attached to the nouns to guess their thematic roles.

3 3D VISUALIZATION

3.1 Visualizing Events on a TimeSlice

We visualize retrieved set of events that are represented as a tree structure on a TimeSlice [2]. When a user inputs a topic noun, the system retrieves events from the dependency database. Each event is represented as a triplet (a noun, a case particle, and a verb) and a path between constituents of the triplet on a TimeSlice. A set of events are visualized as a tree. TimeSlice displays only top-n frequent events related to each case particle belonging to a selected month. TimeSlice changes the size of nodes representing verbs according to the frequency of events in the selected month.

In Figure 1 (a), the system visualizes events about 'swine flu' such as "catch swine flu (shingata-influenza-ni kansen-suru)". In this case, the topical noun 'swine flu' is centered, and case particles '-de', '-ni', and '-ga' are around the noun 'swine flu'. Verbs such as 'catch' are arranged around the topic noun with case a particles.

(a) Visualizing events related to "swine flu"   (b) Expanding nodes to show detail of the event "catch swine flu"
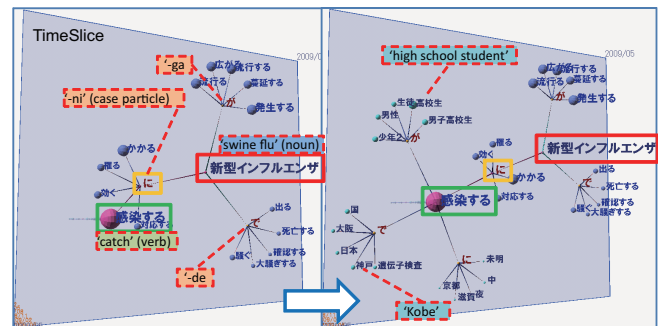
Figure 1: Visualization of events and detailed information about a selected event

3.2 Interface for Event Exploration

We provide an interface for drilling down detailed information after visualizing events about the particular topic on the TimeSlice.

In Figure 1, we select the event 'catch swine flu' by selecting the verb 'catch'. Then, the system shows more specific events such as "high school student catches swine flu" and "catch swine flu at Kobe", and then visualizes them as child nodes of the selected event.

3.3 Comparison of Multiple Topics

To compare events retrieved from different topics, we utilize a *split view* proposed by [2] to visualize two sets of events in different TimeSlices (the upper and lower part of Figure 2). The positions of the different TimeSlices along the timeline are synchronized with one another. The users can also add sets of TimeSlices having different time stamps to the timeline.

In the *split view*, the system treats the positions of the central noun, the same case particles, and the same verbs on different TimeSlices as the same nodes respectively, and arranges them at the same position on the different TimeSlices. Moreover, when we select nodes on one TimeSlice, the system automatically selects and highlights the same nodes on the different TimeSlices.
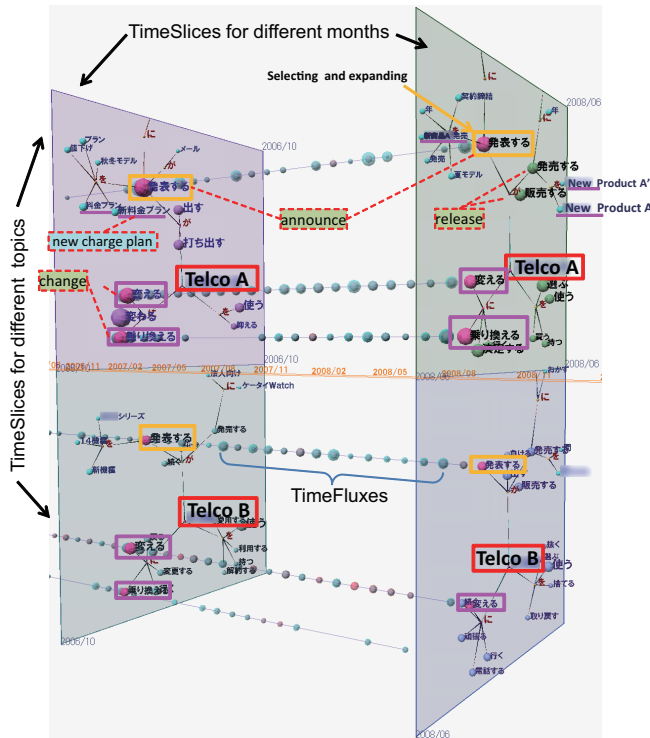

Figure 2: Multiple TimeSlices for comparing events in different months and for comparing events on different topics

3.4 Visualizing Temporal Changes in Event Tree Structure

A TimeSlice displays only events in a specified month, and can be dragged along the timeline to change visualized events. In addition, our system allows us to interactively add new TimeSlices in different positions along the timeline, and to compare multiple TimeSlices in the 3D environment (Figure 2). This enables us to compare events in different months.

The same events belonging to different timings, which are visualized in different TimeSlices, have the same position to easily understand the similarity and differences among the different months.

3.5 Visualizing Temporal Changes in Event Frequency

A TimeFlux is a line of spheres for visualizing the changes in the amount of information such as the number of events within a given period of time (Figure 2). They enable us to intuitively observe when a selected event attracted attention such as a bursting point, and periodicity of its trends.

For example, TimeFluxes in the upper part of Figure 3 show that events related to 'catch swine flu' explosively increase after May. 2009. On the other hand, TimeFluxes in the lower part of Figure 3 show that events related to 'catch seasonal flu' increase and decrease in an annual manner.
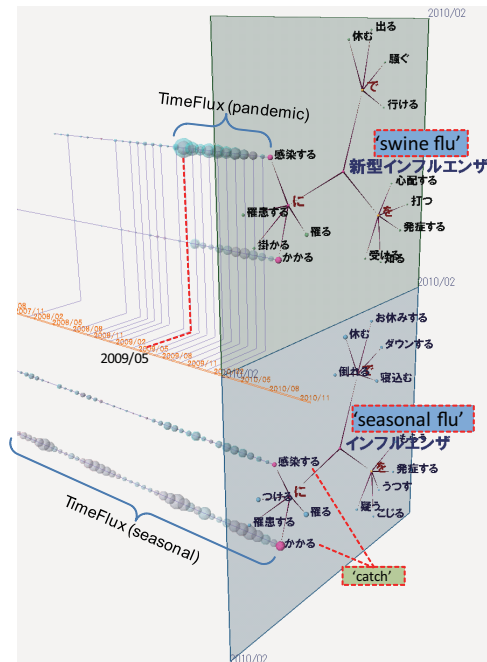

Figure 3: Comparison of temporal changes in frequency among events about different topics by using multiple TimeFluxes

4 CASE STUDIES

Figure 2 shows the visualization for a comparison of marketing effect about two telephone companies (Telco for short) in Japan. The upper TimeSlice shows a topic about 'Telco A', and the lower one shows a topic about 'Telco B'[2]. We can recognize many events related to 'change' appear in TimeSlice about 'Telco A' than 'Telco B'. When we show TimeFluxes about such events, we can observe that there are some peaks in the events 'change to Telco A' on the TimeSlice for 'Telco A'. In Figure 2, we put left TimeSlices to the position of the timing when the first peak came (Oct. 2006), right ones to the second peak (Jun. 2008). We can notice that 'Telco A' announced or released something because the size of nodes 'announce' and 'release' becomes large. Next, we expand these nodes to show the detail about announcement and released products, and then we find out that they announced a 'new charge plan' at the first peak, and a 'new product A' at the second peak. On the other hand, although 'Telco B' announced new products very well, there are a few peaks related to events 'change' about 'Telco B'. These results show that the marketing activities of 'Telco A' effectively give an impact to Japan than those of 'Telco B'.

REFERENCES

[1] BlogScope. http://www.blogscope.net/.
[2] M. Itoh, M. Toyoda, and M. Kitsuregawa. An Interactive Visualization Framework for Time-Series of Web Graphs in a 3D Environment. In *Proc. IV 2010*, pages 54–60, 2010.
[3] M. Kitsuregawa, T. Tamura, M. Toyoda, and N. Kaji. Socio-Sense: A System for Analysing the Societal Behavior from Long Term Web Archive. In *Proc. APWeb 2008*, pages 1–8, 2008.

---

[2]Companies' names are anonymized.