

Title	Webでの文章入力時における情報補完
Author(s)	中村, 和正
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/3589">http://hdl.handle.net/10119/3589</a>
Rights	
Description	Supervisor:鳥澤 健太郎, 情報科学研究科, 修士

# Suggestion of referential information for writing documents on the web

Kazumasa Nakamura (410092)

School of Information Science,  
Japan Advanced Institute of Science and Technology

February 8, 2007

**Keywords:** Aid System for User Writing, AJAX, attribute, value.

In this thesis, we propose a system for assisting users to write a document by using referential information acquired from the web. The system automatically collects from the web information that is closely related to the topics of the sentences that the user writes. Recently, we often collect information related to topics of sentences we wrote from web.

For instance, when we describe an object (places we visited, goods we bought, events we joined) we experienced on that day on weblogs, we examine a formal name of the object, refer to an opinion of other bloggers on that object, or investigate an access to the object, the date of the object (event), specification of the objects, or other various attributes of the object, in order to improve the quality of our documents. When such *information needs* occur, we should stop writing to collect the referential information from the web by ourselves.

In this thesis, we attempt to find users' information needs using lexico-syntactic patterns[1], and to show an appropriate knowledge for that information needs.

When we show users the knowledge acquired from the web, we use AJAX to show them as seamless possible as in order not to avoid users' writing. In this thesis, we focus on the proper nouns (such as "Germany" and "Titanic"). And we show them by using early research by Yoshinaga and Torisawa[2].

## Flow of processing of our system

1. The user writes sentences on the editor screen of the client of our system.  
Example: ‘昨日訪れた真打の住所は、 (*The address in the Shin-uchi visited yesterday is* )’
2. The client constantly transmits users’ sentences as an input to the server of our system.
3. The server then detects user’s information needs from the data that the server received, and then acquire the related information by running the information extraction engine on the web.
4. When related information can be acquired, the server transmits the related information to the client.  
Example: ‘石川県金沢市片町 1-1 (*1-1 Katamachi Kanazawa-shi Ishikawa*)’
5. When the client receives related information, it displays it in the editor screen.
6. The user can continue writing the sentences by referring to displayed related information.

Let us assume that a user is writing a sentence ‘昨日はラーメン屋真打に行きました。真打の住所は、 (I went to the Ramen shop Shin-uchi yesterday. Address of the Shin-uchi is )’. Even if the user doesn’t know the address of the restaurant, the system helps the user to write the sentence by acquiring the information on the address of that restaurant from the web, and by showing that information to the user. Like this, because instead of the user, the system finds information that answers users’ information needs, the users can continue writing.

Our system takes advantage of Yoshinaga and Torisawa’s system[2] that extracts attribute-value information from the web. Our system first generates an input query to their system by detecting objects and their attributes using lexico-syntactic patterns from the given fragments of sentences, and then obtains attribute-value information for that object, which can be an answer to the user’s information needs.

We proposed a system that of method of helps the user to write documents by using AJAX and the existing system for extracting attribute-value information from the web. We performed experiments to examine whether referential information acquired by our system is useful for the user. Our system suggested correct attribute-value information for users' information needs at the of 50%. We also confirmed that about the half is an accuracy information.

We think that our system can support it enough because we can improve our system by using various kinds of lexico-syntactic patterns to find various types of information needs, and by using appropriately information extraction engines according to the types of users' information needs.

## References

- [1] Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In Proc. of COLING, pp.539 - 545, 1992.
- [2] Naoki Yoshinaga, and Kentaro Torisawa. Automatic Acquisition of Instances' Attribute-Value Information from the Web. In Proceedings of the 13th Annual Meeting of the Association for Natural Language Processing, Seta. March, 2007. in Japanese.