Automatic Generation of Listing Ads by Reusing Promotional Texts

Atsushi Fujita‡, Katsuhiro Ikushima†*, Satoshi Sato†,
Ryo Kamite⋆, Ko Ishiyama⋆, Osamu Tamachi⋆
‡Future University Hakodate, 116-2, Kameda-nakano-cho, Hakodate, 041-8655, Japan
†Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan
⋆Recruit Co., Ltd., GranTokyo South Tower, 1-9-2, Marunouchi, Chuo-ku, Tokyo, 100-6640, Japan
fujita@fun.ac.jp, ssato@nuee.nagoya-u.ac.jp
{kamite,koishiyama,otamachi}@r.recruit.co.jp

ABSTRACT

This paper describes a system that automatically generates shop-specific listing ads by reusing textual data promoting each shop. As the textual data available for this research are primarily created for use on a restaurant portal site on the Web, we applied three methods for making it usable for the descriptive texts of ads. The only manual task is creating a couple of domain-specific patterns. Subjective evaluation showed that our system can generate ads with sufficiently high precision and coverage. A one-month experiment using Overture Sponsored Search showed that a number of the automatically generated ads had higher CTRs than the template-based baseline ads. This indicates that automatically generated ads can promote shops more effectively than template-based ads.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing; K.4.4 [**Computers and Society**]: Electronic Commerce; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing, Linguistic Processing

General Terms

Algorithms, Experimentation, Economics

Keywords

sponsored search, automatic ad generation

1. INTRODUCTION

The amount of money spent for online advertising has been increasing rapidly. Despite the worldwide recession in 2008-2009, in Japan, the amount grew 1.2% in 2009 and exceeded ¥707B ($7.6B), while that spent for all other media including TV, newspapers, magazines, and even specialized media such as billboards and direct mail dropped more than 10% [5]. Textual advertising has become a key component of online advertising. For example, ¥193B ($2.1B) of the total above was spent for ads displayed alongside the Web search results (henceforth, *listing ads*) provided by several ad services, such as Google AdWords[1] and Overture Sponsored Search[2].

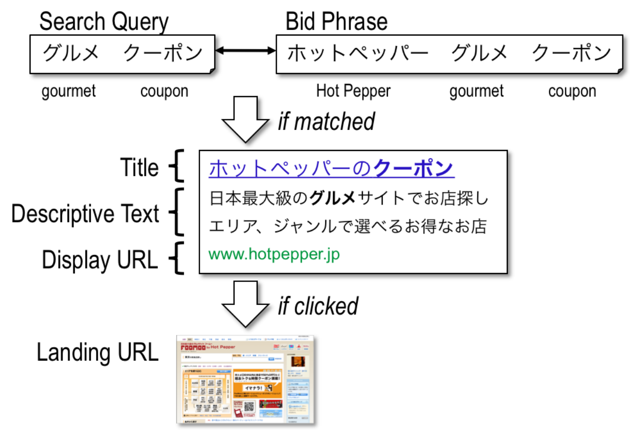

Figure 1: Listing ad and user flow.

Figure 1 shows an example of a listing ad and the user flow. When a user enters a search query into a search engine, the search engine's ad service retrieves and displays ads by comparing the query with the *bid phrase* for each ad. If the query is matched to a bid phrase and the ad is judged more suitable for the query than competing ads by the ad service, a tuple comprising the *title*, *descriptive text*, and *display URL* of the ad is displayed on the search result page. Clicking on the displayed tuple takes the user to the Web site indicated by the *landing URL* of the ad.

The model for such ad services is called pay-per-click (PPC) and is illustrated in Figure 2. Rather than pay for having an ad shown along with the Web search results, advertisers pay only if a user clicks on the ad. Due to the cost efficiency of the PPC model and the platforms available to advertisers for easily conducting advertising campaigns, the volume of such advertising is expected to further increase.

*A part of this work was done while the second author was a master course student of Nagoya University.

[1] http://adwords.google.com/
[2] http://listing.yahoo.co.jp/service/srch/

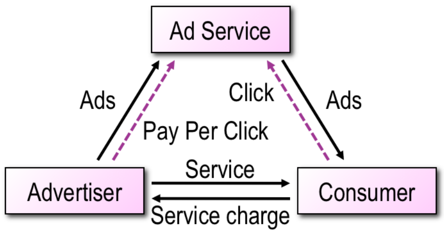

Figure 2: Pay-per-click model.

An advertiser's goal is to increase the number of specific actions, such as purchasing a product, opening an account, and registering for a mailing list, taken by users while visiting the advertiser's Web site. Therefore, advertisers use PPC-based ads to attract users to their Web site. There are two promising approaches to increasing the effectiveness of a PPC-based ad. The first is to improve the bid phrase so that the ad will be displayed in Web search results more frequently. The second is to improve the corresponding title and descriptive text so that more users click on it. Both approaches entail generating candidates and selecting the most effective ones. Many researchers, in particular those working for ad services, have focused on the second issue. Broder et al. [2] investigated methods for improving the accuracy of matching between user search queries and bid phrases. Because the task can be interpreted as a kind of information retrieval, techniques developed in the research area, such as latent semantic indexing and pseudo-relevance feedback, have yielded promising results. For ad services, the expected *click-through-rate* (CTR) for unseen (newly registered) ads has been another important topic [13, 14]. The CTR of ad a is defined by

$$CTR(a) = \frac{\text{\# of times } a \text{ is clicked}}{\text{\# of times } a \text{ is displayed}}. \quad (1)$$

Advertisers also use the CTR to measure the performance of PPC-based ads.

In contrast, there has been less investigation of the first issue, the candidate generation. While the titles and descriptive texts of ads play important roles in connecting user's search intentions to the products and services promoted by the ads [7], the focus of the research has so far been on the bid phrases [8, 1, 4, 12][3]. Because of the diverse search intentions of users and the diverse characteristics of long-tail content, improving an ad's title and descriptive text should increase its CTR. However, templates and fixed texts are typically used to produce the title and descriptive text due to their cost efficiency even though their use tends to result in ads that are poor at attracting users.

Aiming at recommending thousands of products and services to a wider range of consumers more effectively and at reducing the labor and other costs spent for promoting them, we have been investigating methods for exploiting textual data associated with a particular type of content. As part of our investigation, we have developed an automatic ad generator tailored for the restaurant domain. Our system generates shop-specific listing ads, exploiting both domain-specific findings obtained through our previous market research and textual data promoting each shop created for a restaurant portal site on the Web.

This paper is organized as follows. Section 2 describes the model underlying our research. Section 3 explains the resources available for generating ads. Section 4 presents our ad generator. Because our present goal is to clarify the effectiveness of using existing textual data for generating ads, two types of evaluation were conducted. Section 5 reports on the performance of our ad generator in terms of precision and coverage. Section 6 presents the performance of automatically generated ads evaluated on the basis of CTR. Finally, Section 7 summarizes key points and mentions future work.

[3]https://adwords.google.com/select/KeywordToolExternal

2. DOMAIN AND MODEL

The research reported here was conducted using Hot Pepper Foomoo[4] (henceforth simply Hot Pepper), a Japanese portal service for the restaurant domain.

2.1 Model

To bring consumers to advertisers, in this case shops, Hot Pepper has so far been collecting information about participating shops and providing it to consumers, as shown on the left side of Figure 3. Hot Pepper started providing information to consumers through its free paper distributed monthly in various areas in Japan; it then extended its coverage to a Web portal site. In this model, consumers can obtain special services, such as discounts, extra dishes, and secret menus, by making reservations through Hot Pepper's portal site or by presenting coupons clipped from the free paper or printed Web page at the shop.

The information about each shop is roughly divided into four categories:

- general information
- photographs with captions
- promotional text about the shop in general
- promotional text about limited sales campaigns.

The general information includes the shop name, physical data (address, phone number, etc.), and features (capacity, non-smoking seats, private rooms, etc.). The photographs with captions include photographs of the shop itself and of selected menu items. Other portal services in this domain provide similar information.

In contrast, the other two types of information, i.e., promotional text about the shop in general and about limited sales campaigns, are unique to Hot Pepper. This information is obtained through discussions between the Hot Pepper sales staff and the participating shops. The sales people study shop operations and suggest not only catch phrases and advertising copy, but also sales campaigns and coupons. Hence, the resultant texts comprehensively illustrate various aspects of the shop.

Since starting its portal site in April 2005, HotPepper has been working with an increasing number of shops: for example, 25,815 shops were promoted in October 2009. The portal site provides several functions, such as condition-based search and free-word search, that help users search for shops. This also increased the number of users.

2.2 Extension to ad services

The emergence of ad services has enabled Hot Pepper to extend its model as shown on the right side of Figure 3. However, the textual data described above cannot be directly used for ads because

[4]http://www.hotpepper.jp/

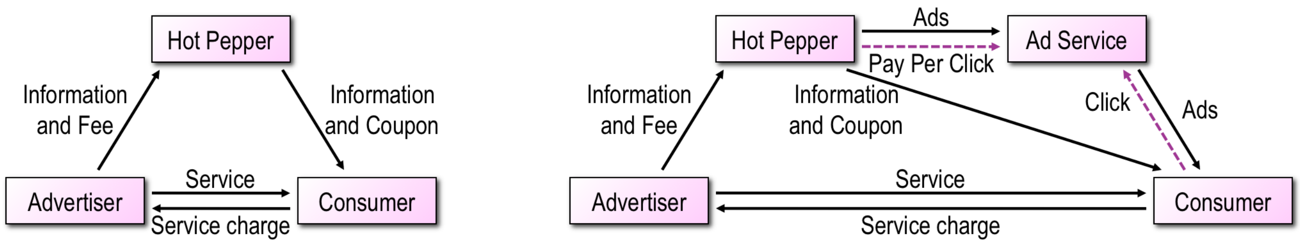

Figure 3: Hot Pepper's business models (left: basic, right: pay-per-click extension).

Table 1: Examples of special feature sites.

Feature Type Feature name	URL
Franchise name	
牛角 (*Gyukaku*)	gyukaku-navi.net
坐・和民 (*Za Watami*)	zawatami-navi.net
サンマルク (Saint-Marc)	sanmaruku-navi.net
Shop type	
居酒屋 (pub)	izakaya-jyouhou.net
スイーツ (sweets)	sui-tu-navi.net
バー (bar)	bar-navi.net
Specialty	
日本酒 (Japanese *sake*)	nihonsyu-navi.net
カレー (curry)	kare-navi.net
ホルモン (organ meat)	horumon-navi.net
Special occasions	
誕生日 (birthday)	tanjyoubi-navi.net
二次会 (second party)	nijikai-navi.net
デート (date)	date-jyouhou.net
Features	
ソファー (sofa)	sofa-navi.net
ダーツ (dart)	darts-navi.net
食べ放題 (all-you-can-eat)	tabehoudai-jyouhou.net

they were primarily created for use on the Web portal site. That is, they do not necessarily fulfill the criteria prescribed by the ad services, such as length of text and available characters. Manual ad creation is impractical given the large number of restaurants promoted. Therefore, Hot Pepper has been using heuristically handcrafted template-based ads although they do not precisely describe the information about each shop.

This motivated us to investigate methods for automating ad creation by reusing the textual data.

2.3 Special feature sites

The search queries and transitions of users at Hot Pepper's portal site exhibit several typical search intentions, such as the nearest branch of a particular franchise, shops that serve particular brands of *sake*, and shops that celebrate consumers' birthdays. Assuming that the search intentions at the portal site reflect the natural distribution of search intentions, Hot Pepper started providing special feature sites in September 2009. Each special feature site is comprises a subset of shops that are retrieved from the portal site by specifying several conditions that approximate typical search intentions. At a special feature site, users can easily find shops that match their needs.

Currently, Hot Pepper operates 200 special feature sites, including ones for popular franchises, various shop types, specialties, special occasions, and features, etc., such as those listed in Table 1. One aim is to increase the exposure of each shop and the number of channels to the portal site. Another is to improve the branding of HotPepper and consequently improve the conversion rate of PPC-based ads.

3. AVAILABLE RESOURCES

Our advertising target consists of shops on special feature sites. We generate ads for them using three types of information available from Hot Pepper:

- textual data promoting each shop
- information for each special feature site
- search query logs for the portal site

3.1 Textual data promoting each shop

Table 2 shows example textual data for a shop. As they were originally created for promoting the shop on the Web portal site, they contain attractive, impressive, and informative expressions, as exemplified by:

(1) a. 味・触感・香と３拍子揃ったうまい肉
(Beef with triple-threat of taste, texture, and flavor)
[from advertising copy]

b. 黒毛和牛の飲放付コースが*¥4515* →*¥3000*
(Discounts for Japanese beef course with free drink)
[from caption of representative photo]

c. 全席堀りごたつだから楽ちんだね☆
(Comfortable with foot warmers set in floor recesses)
[from captions of photo showing recommended dishes]

d. 毎月 *29* 日は何かが起こる！
(Something will happen on the 29th of every month)
[from campaign catch phrases]

However, the textual data cannot be used directly for ads. First, most of them are longer than the limit imposed by ad services. Second, the text other than the general information text tends to contain many symbols such as "★" (star) and "♪" (musical note), which are not allowed in ads. We therefore need to tailor the textual data to make it usable for ads. The elements marked with a "⋆" are common to a number of shops. Thus, finding shop-specific expression is another issue.

We call the set of elements marked with a "†" in Table 2, i.e., catch phrase, advertising copy, representative photo caption, and campaign catch phrases, "*basic text*", and regard it as the primary source of information for generating descriptive text. Other elements are used as alternative sources or as other parts of an ad.

3.2 Information for each special feature site

For each special feature site, two types of information are manually specified, as exemplified in Table 1. The first is the *feature name* of

Table 2: Textual data for a shop.

Category & Element name	Example
General	
Shop ID	J0000abcde
Shop name	焼肉 XXX YYY 店 (Grill XXX YYY-branch)
Area (large)*	愛知 (Aichi prefecture)
Area (middle)*	名古屋駅周辺 (around Nagoya station)
Area (small)*	名古屋駅西口 (West exit of Nagoya station)
Text used for top page of the shop in the portal site	
Genre catch phrase*	食べ放題＆飲み放題の焼肉
Shop catch phrase†	当日 OK ★焼肉食べ放題！ 名駅徒歩 30 秒！
Advertising copy†	★満足な宴会は焼肉で決まり★　　【焼肉】食べ放題＆【ドリンク】飲み放題＝クーポン利用で３０００・４０００円・５０００円がございます。　コラーゲン付！国産牛しゃぶしゃぶの【食べ放題】＆ドリンク【飲み放題】もクーポン利用で３５００円で楽しめます。　贅沢な食材で至福の時を満喫してください。　全席掘りごたつで、カップルはもちろん宴会にピッタリの空間をご用意 !!　又、創業３０年、味自慢の XXX では、カルビが絶対オススメ !!　是非、味・食感・香と３拍子揃ったうまい肉 をお召し上がり下さい !! 皆様の御来店を、スタッフ一同心よりお待ちしております
Access guide	名駅　西口（新幹線乗り場）を左にスグ！
Course names (up to 5 records)	(a) 当日 OK ★焼肉Ａ【食放】＋【飲放】¥3990 → ¥3000 ※+¥500 でタン＋サラダをご提供！ (b) [1] 当日 OK ★焼肉Ｂ【食放】【飲放】¥5040 →¥4000[2] 黒毛和牛★納得コース【飲放】¥4515 →¥3000 (c) [1] 当日 OK ★焼肉Ｃ【食放】【飲放】¥6090 →¥5000[2] 黒毛和牛★極上コース【飲放】¥5565 →¥4000 (d) 当日 OK ★黒毛和牛のしゃぶしゃぶ or すきやき【食放】【飲放】¥4515 →¥3500
Photo captions	
Representative photo†	名駅新幹線口、徒歩 30 秒★ 2 名〜当日 OK の焼肉食放＆飲放充実!!毎月 29 日は何かが起こる!
Recommended dishes (up to 3 records)	(a) 2 名〜当日 OK の新登場メニュー★コラーゲンボール入り♪すき焼きの食べ放題でスタミナバッチリ♪♪ (b) 2 名〜当日 OK ★全 10 品のお気軽コースは、学生さん・サラリーマンにも大人気！ (c) 2 名〜当日 OK ★ちょっと贅沢に会社宴会にも…「牛タン」も食べ放題♪
Atmosphere (up to 3 records)	(a) 全席掘りごたつだから楽ちんだね☆ (b) カップルで落ち着いた一時を♪ (c) なんてったって店内がキレイだから気持ちいいよね !!
Campaigns	
Campaign catch phrases† (up to 6 records)	(a) 名駅徒歩 30 秒★ガッツリ焼肉食べ放題♪飲み放題♪ナント 3000 円〜 (b) スタミナ◎黒毛和牛のすきやき食べ放題＆飲み放題が 3500 円！ (c) 毎月【肉の日＝ 29 日】は XXX の日★毎月 29 日は、何かが起こる…!! (d) 名古屋駅新幹線口徒歩 30 秒★掘りごたつで宴会、最大 60 名迄 OK！ (e) 黒毛和牛の飲放付コースが 4515 → 3000 円 ★ 5565 → 4000 円もご用意！ (f) 名駅 30 秒★ワイワイ食べ放題＆飲み放題で宴会！最大 60 名迄！
Campaign types* (up to 6 records)	(a) 食べ放題プランのあるお店 (b) お手頃！4000 円以下の飲み放題付コース (c) おいしいお肉が食べたい！ (d) 50 人以上の宴会が可能なお店 (e) 3000 円以下で楽しめるコース料理 (f) 20 人以上の宴会が可能なお店

the site, which is the original Japanese spellings of the URL of the special feature site. The second is *feature type*. We manually classified the special feature sites into five categories, as shown in the table: (a) franchise name, (b) shop type, (c) specialty, (d) special occasions, and (e) features.

3.3 Search query logs for portal site

Search query logs for Hot Pepper's portal site are also available. However, due to several restrictions, we can use only 482,700 lines from the logs[5]. Each line consists of the query string and the type specified by the user (shop name, station/area, food type, etc.).

4. AD GENERATOR

Given the ID of a shop on a special feature site, our ad generator produces multiple ads for the shop (Figure 4).

Advertisers may expect the system to generate ads for any given content: not only Web pages or other types of text, but also specification sheets for products and services and other non-textual data, such as pictures and videos. Ravi et al. [12] modeled the process of ad creation as first bid phrase and then title and descriptive text. In contrast, our primary concern is using the textual data for each shop created for the portal site for creating ads. Therefore, as illustrated in Figure 4, the current version of our system generates descriptive text separately from the pairs of bid phrase and title. Bid phrases and titles are heuristically generated so that they reflect typical user intentions in this domain as well as the findings of previous studies. In contrast, the descriptive texts are composed of information specific to a shop. Since the original textual data was not created for ads, they are trimmed and summarized.

We implemented our system assuming the use of Overture Sponsored Search as the ad service. Thus, several parts of our system, such as the maximum lengths of titles and descriptive texts (15 and 33 characters, respectively) and the available/prohibited characters, were adjusted to meet Overture's criteria.

4.1 Generating bid phrases

The generation and selection of bid phrases are common topics in computational advertising research. Generally speaking, the more frequently a phrase is queried on a search engine, the higher the bidding for that phrase because there is a greater number of com-

[5]They were collected in August, 2009.

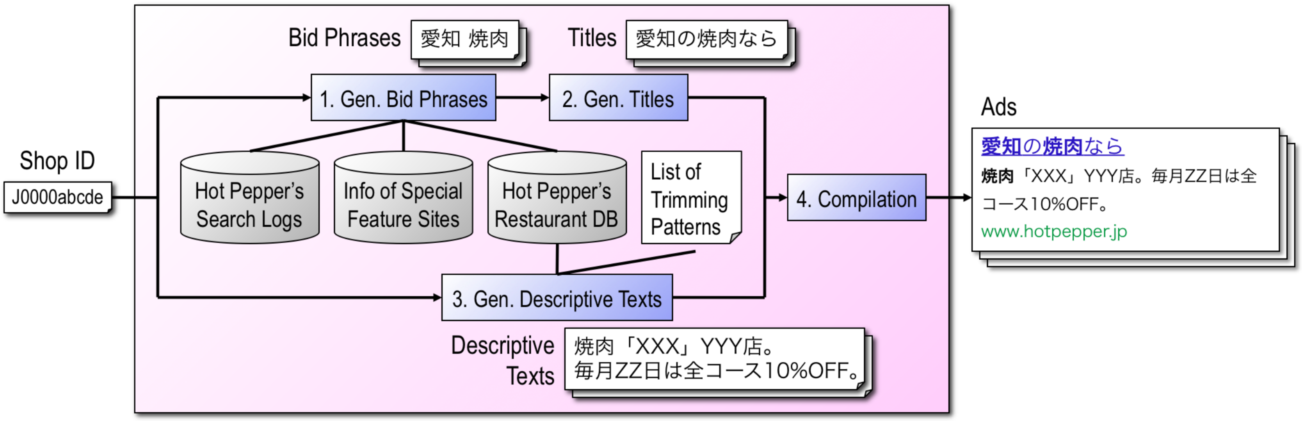

Figure 4: Overview of our ad generator.

peting ads. Previous studies have therefore focused on identifying semantically related phrases in order to obtain the same volume of clicks at a lower cost [8, 1]. However, using less common expressions as bid phrases is less effective because the corresponding ads will be shown less often, so the volume of clicks will be lower. Bid phrase are more effective if they reflect the user's intention [3, 12]. An ambiguous or vague search query is likely to produce poor results, which will motivate users to add another word or expression to the query and search again instead of clicking on an ad with that ambiguous or vague query as the bid phrase.

A user's intention in this domain is, in most cases, either to find a shop or to get information about a specific shop. For the first intention, users tend to use place names and/or genre in the search query. For the second, the shop name should be sufficient. However, people tend to also use place names to specify the area or branch of a franchise. Likewise, the genre of the shop may also be used as a part of the query to disambiguate different shops coincidentally having the same name.

Given these considerations, we use a pair of place name and shop genre as a bid phrase. Given the ID of a shop on a special feature site, our system generates a list of place names through the following procedure.

1. Retrieve area names (large, middle, and small) of the shop from the database (Section 3.1). For example, given shop ID "J0000abcde," "愛知" (Aichi prefecture), "名古屋駅周辺" (around Nagoya station), and "名古屋駅西口" (West exit of Nagoya station) are retrieved.
2. Trim and split each area name into a set of atomic place names using 19 handcrafted patterns such as those shown in Table 3.
3. Filter place names that have been used as a search query at Hot Pepper's portal site (Section 3.3).

The search query log is referred to in the third step due to the assumption that strings used for searching for shops at the portal site are also used for the same purpose on search engines. The system compiles a bid phrase using a place name on the list. This means that multiple bid phrases (up to the number of place names on the list) can be generated. The feature name of the special feature site is used as the shop genre (Section 3.2).

Table 3: Examples of patterns for trimming and extracting place names.

Condition		Result
"XXX (YYY · ZZZ)"	→	"XXX" "YYY" "ZZZ"
"XXX, YYY, その他" ("XXX, YYY, etc.")	→	"XXX" "YYY"
"XXX 郊外" ("suburbs of XXX")	→	"XXX"
"XXX 周辺" ("around XXX")	→	"XXX"

4.2 Generating titles

A certain portion of people determine whether to click on a displayed ads by evaluating its relevance to the query [7]. Ad services emphasize the relevance by showing parts of the search query in an ad's title and descriptive text in bold face. Using the bid phrase for the ad title and/or descriptive text is thus one way to connect the user's search intention with the advertising target.

Our system generates ad titles by completing the template below with a given bid phrase, i.e., a pair of place name (k_p) and shop genre (k_g).

(2) k_p-の k_g-なら
k_p-GEN k_g-in case
In case (you search for) k_g in k_p . . .

Template-based titles have been prevalently used due to their cost efficiency. Another reason for using a template is to meet Overture's criterion for the use of place/area names for bid phrases[6].

4.3 Generating descriptive texts

The descriptive texts generated by the system comprise the shop name and a promotional expression. Shop names are used because they are obviously the main way by which consumers identify the advertising targets. Another reason is the difficulty of generating long expressions: the longer an automatically generated expression, the more likely it is ungrammatical.

[6]When a place/area name is used in the bid phrase, the title or descriptive text must contain place/area names equivalent to or more specific than that used in the bid phrase and must clarify the relevance between the content shown in the landing page and the indicated place/area.

S_s $2(36+20)+3(20+11)+0(30+11)+3(11+14)+2(14+24)$
S_n 36 20 30 11 14 24
S_e 2 3 0 3 2

旬-の	食材-を	ふんだんに	使っ-た	料理-が	好評-です
season-GEN	foodstuff-ACC	in plenty	use-PAST	dish-NOM	popular-COP

Dishes with plenty of seasonal foodstuff are popular here.

Figure 5: Example of dependency tree of a sentence and its score.

The shop name is first determined by trimming the original string in the database (Section 4.3.1). The procedure for generating the second half of the descriptive text is invoked only when a shop name can be specified. The maximum length of the second half is determined on the basis of the Overture's criterion (33 characters) and the length of the specified shop name. Three summarization methods are separately applied to the basic text of the shop to generate a promotional expression (Sections 4.3.3, 4.3.4, and 4.3.5). The simplest method (method 1) is also applied to other elements within the textual data that illustrate each shop from various points of view.

In the field of text summarization, user-focused and query-biased techniques have been investigated [9]. In contrast, our system generates a shop-specific descriptive texts using textual data specific to the shop. While supervised machine learning techniques have achieved significant performance for summarization [10, 6], pairs of source data and ideal ads are unavailable. We therefore use a heuristic scoring function previously described in [11].

4.3.1 Shop name extraction

We limit the length of the shop name to 15 characters so that it fits on the first line of descriptive text and attach a period to its tail. The name is available from the shop database; however, it is sometimes longer than the limit. In such cases, a trimming process is invoked according to the feature type of the special feature site to which the shop belongs.

If the feature type of a given shop is "franchise name," the name of the shop in the shop database tends to be composed of decorative expressions, franchise name, and branch name. The example below consists of 21 characters including white spaces between elements.

(3) *AAA* 居酒屋 *ZZZ* 料理 *XXX* *YYY* 店
pub in AAA ZZZ cuisine franchise name YYY-branch

Although decorative expressions are useful for promotion, in this case, franchise name and branch name are prioritized as elements of shop name. Therefore, we emphasize the franchise name by using quotation marks and by deleting the decorative expressions from the head of the string until the string becomes shorter than the limit. The finally determined shop name is composed of 14 characters in this example as shown in (4).

(4) *ZZZ* 料理 「*XXX*」 *YYY* 店
ZZZ cuisine "franchise name" YYY-branch

Shop names for other than franchises are not regularly formatted. Several samples demonstrated that the shop name tends to be located at the head of the string, unlike the placement for franchises; therefore, components of the shop name are deleted from the tail of the string in these cases.

Table 4: Score table for dependency edges.

S_e	Class of dependency relation (Example $\langle n_c, n_p \rangle$)
3	Verb with an argument "⟨ 個室で, 楽しむ⟩" (to enjoy in a private room) Deverbal noun with an argument "⟨ 食材を, 使用⟩" (to use foodstuff) Adnominal modification "⟨ くつろげる, 個室⟩" (private room for relaxing)
2	Adjectival modification "⟨ 新鮮な, 魚⟩" (fresh fish) Adjectival predicate "⟨ 気遣いが, うれしい⟩" (be pleased with one's care) Nominal (except coordination) "⟨ 旬の, 食材⟩" (seasonal foodstuff)
1	Nominal coordination "⟨ お造りと, 煮付け⟩" (sliced raw fish and boiled fish)
0	Otherwise "⟨ ふんだんに, 使った⟩" (with plenty of)

4.3.2 Scoring function

The scoring function proposed by [11] is used in the three summarization methods described below to measure the appropriateness of an expression for promotion.

Given sentence s within the textual data for shop x, its score $S_s(x,s)$ is calculated using

$$S_s(x,s) = \sum_{\langle n_c,n_p\rangle \in T(s)} S_r(x,n_c,n_p), \quad (2)$$

$$S_r(x,n_c,n_p) = S_e(n_c,n_p)\big(S_n(x,n_c)+S_n(x,n_p)\big), \quad (3)$$

where $T(s)$ stands for the set of child node (n_c) and parent node (n_p) pairs in the dependency tree of the sentence and $S_r(x,n_c,n_p)$ stands for the score of a pair of nodes. $S_e(n_c,n_p)$ and $S_n(x,n)$ are the scores of an edge and a node, respectively. Figure 5 illustrates an example dependency tree for a sentence and its score.

The score of an edge $S_e(n_c,n_p)$ is defined on the basis of its class, as shown in Table 4. Predicates and their arguments are preferred because they are more important in composing a sentence than attributive elements, such as adjectives and adverbs.

The score of node $S_n(x,n)$ is computed using

$$S_n(x,n) = \max_{w \in CW(n)} tf(x,w)\left[\log\Big(\frac{N}{df(w)}\Big)+1\right],$$

where $CW(n)$ indicates the set of content words contained in the node n, $tf(x,w)$ stands for the frequency of the content word w within the basic text of the shop x, and $df(w)$ stands for the number of shops for which the basic text contains the content word w

Table 5: Examples of patterns for trimming texts.

n_c	n_p
X の	{こだわり, 食材, 数々,. . .}
{pride and joy of, foodstuff of, variety of, . . .} X	
X が	{おすすめ, 豊富, 人気,. . .}
X is {recommended, abundant, popular, . . .}	
X を	{提供, 楽しめる, 感じる,. . .}
X is {serve, gustable, felt, . . .}	
X に	{こだわる, どうぞ, 合う,. . .}
{proud of, please for, suit for, . . .} X	

among N shops ($N = 25,815$ in the experiment reported here). In other words, the score of each content word is defined on the basis of tf-idf (term frequency and inverse document frequency) used in information retrieval to give higher scores to shop-specific words.

4.3.3 Method 1: sentence & phrase extraction

Given the basic text of a shop, the first method simply outputs a sentence by:

1. splitting the text into a list of sentences,
2. filtering out sentences longer than the limit, and
3. selecting the one that has the highest score as calculated using Eq. (2).

To evaluate the effectiveness of using textual data, we also applied this method to six other types of elements conveying domain-specific information, i.e., access guide, course names, photo captions of recommended dishes, photos illustrating atmosphere, genre catch phrase, and campaign type (see Table 2). Note that we use only the first line of a campaign types because it (and genre catch phrase) is not specific to a shop.

4.3.4 Method 2: sentence trimming

A set of trimming patterns is applied to the sentences longer than the limit. Given the basic text of a shop, this method first generates a list of candidates using two types of trimming techniques:

1. Split the text into a list of sentences.
2. Parse each sentence and make a list of dependency trees.
3. For each dependency tree, find new roots using seven handcrafted patterns, such as those shown in Table 5. Each pattern represents a dependency edge whose parent node can be regarded as a root expression for ads. The content words of the parent node of patterns are lexically conditioned. The determined subtrees are added to the candidate list.
4. For each dependency tree, enumerate all subtrees that contain the root node [10]. For example, Figure 6 shows the dependency subtrees extracted from the dependency tree in Figure 5.
5. Filter out subtrees longer than the limit.
6. Select the candidate subtree with the highest score as calculated using Eq. (2).
7. Generate an expression from the resulting subtree.

By analyzing samples, we have identified 67 words for the parent node of patterns.

Table 6: Inventory of generated descriptive texts.

Type	Source	Method
d_1		1
d_2	Basic text	2
d_3		3
d_4	Access guide	1
d_5	Course names	1
d_6	Captions of photos recommending dishes	1
d_7	Captions of photos describing atmosphere	1
d_8	Catch phrase of genre*	1
d_9	Campaign types*	1

4.3.5 Method 3: sentence reconstruction

Method 2 has two restrictions: it always remains the root node of the sentence, and it requires handcrafted patterns. The third method takes a more aggressive approach: it reconstructs a sentence from the set of dependency edges in the basic text.

Given the basic text of a shop, this method generates a candidate expression using a six-step algorithm.

1. Split the text into a list of sentences.
2. Parse each sentence and extract the dependency edges $E = \bigcup_s T(s)$.
3. Select the core edge from E with the highest score as calculated using Eq. (3) and remove it from E.
4. Filter out from E those edges extracted from the other sentences, those edges with a score of zero, those edges with a child node containing a verb, formal noun, or adverbial noun, and those edges with nodes containing a pronoun.
5. Adjoin the remaining edges E in descending order of their scores as long as the length of the linearized text does not exceed the length limit.
6. Generate an expression from the resulting subtree.

The conditions in step 4 may seem overly strict; however, they are necessary to prevent generating ungrammatical expressions.

4.4 Compilation of ads

Finally, our ad generator complies a set of ads for a given shop using k pairs of bid phrase and title and up to nine types of descriptive texts, as summarized in Table 6. Two other components are specified:

- Display URL: Hot Pepper's portal site (www.hotpepper.jp)
- Landing URL: Shop page at Hot Pepper's portal site (http://www.hotpepper.jp/strShopID/). "ShopID" indicates the ID of a given shop.

5. EVALUATION 1: PRECISION AND COVERAGE

Our first evaluation focused on how many ads our system generates and on their appropriateness. First, the numbers of ad components (bid phrase, title, and descriptive text) were calculated. Then, the precision and coverage of the resulting ads were evaluated subjectively. Note that we have left the evaluation of the quality of component combinations for future work although titles and descriptive texts may coincidentally share some expressions.

旬-の 食材-を ふんだんに 使っ-た 料理-が 好評-です
season-GEN foodstuff-ACC in plenty use-PAST dish-NOM popular-COP
Dishes with plenty of seasonal foodstuff are popular here.

食材-を ふんだんに 使っ-た 料理-が 好評-です
foodstuff-ACC in plenty use-PAST dish-NOM popular-COP
Dishes with plenty of foodstuff are popular here.

ふんだんに 使っ-た 料理-が 好評-です
in plenty use-PAST dish-NOM popular-COP
*Dishes with plenty are popular here.

旬-の 食材-を 使っ-た 料理-が 好評-です
season-GEN foodstuff-ACC use-PAST dish-NOM popular-COP
Dishes with seasonal foodstuff are popular here.

食材-を 使っ-た 料理-が 好評-です
foodstuff-ACC use-PAST dish-NOM popular-COP
Dishes with foodstuff are popular here.

使っ-た 料理-が 好評-です
use-PAST dish-NOM popular-COP
*Dishes with are popular here.

料理-が 好評-です
dish-NOM popular-COP
Dishes are popular here.

好評-です
popular-COP
be popular here.

Figure 6: Dependency subtrees containing root node of original dependency tree in Figure 5.

5.1 Ad generation

First, 20 special feature sites were randomly selected. For each special feature site, up to ten shops were sampled for each of nine districts[7] in Japan. There were a total of 575 shops because less than ten shops were registered in some countryside districts.

The elements of the ads for each shop were then generated in three steps (see also Figure 4).

1. For a given shop, two bid phrases were generated separately using two randomly selected place names.
2. For each bid phrase, a title was generated using the template shown in Section 4.2.
3. For each shop, nine types of descriptive texts were generated independently of generating the bid phrase (step 1) and title (step 2).

5.2 Results

For all of the sampled shops, we were able to obtain two place names and to generate bid phrases. None of the titles generated using the place name and feature name of the special site exceeded Overture's length limit (15 characters).

Then, the system, using the three summarization methods, attempted to generate descriptive texts for the shops. It was able to find a shop name for 513 of the 575 shops (89%). That is, it failed to find a shop name that met the length limit for 62 shops. It then generated the second half of the descriptive texts for the 513 shops. As the three methods sometimes output the same descriptive text, there were 3,775 unique descriptive texts in total. The numbers of each type of descriptive texts are shown in Table 7. All three methods generated a descriptive text from the basic text for 513 shops. Only method 1 was used to generate descriptive texts from the other element texts. Except for the genre catch phrase element text (d_8), it was unable to generate a descriptive text for all 513 shops from the other element texts. In particular, it generated descriptive texts for only 43% (219/513) of the shops from the access guide text (d_4) and 57% (290/513) from the course names text (d_5) because these texts tend to contain many symbols prohibited for use in ads, such as "→" (arrow) and parentheses.

Table 7: Precision and coverage of descriptive texts.

Type	Generated	Appropriate	Precision	Coverage
d_1	513	496	0.97	0.86
d_2	513	428	0.83	0.74
d_3	513	400	0.78	0.70
d_4	219	219	1.00	0.38
d_5	290	234	0.81	0.41
d_6	470	450	0.96	0.78
d_7	487	479	0.98	0.83
d_8	513	513	1.00	0.89
d_9	469	469	1.00	0.82
Total	3,775	3,476	0.92	0.89

The generated descriptive texts were subjectively evaluated by a person with 5-years expertise in creating ads. Of the 3,775 unique descriptive texts, 299 were judged inappropriate. Three examples of inappropriate descriptive texts (second half) are shown below.

(5) a. 付き選べる鍋コース *4780* 円最大
*Selectable pot meal course with (something), ¥4780, up to (something). . . . ungrammatical

b. *2500* 円、飲み放題+*500* 円
¥2500, +¥500 for free drink. . . . content is unclear

c. まずはこれを食べてほしい。
We'd like you to eat this first. . . . no referent

[7] *Hokkaido*, *Tohoku*, *Kanto*, *Hokuriku-Koshin'etsu*, *Tokai*, *Kansai*, *Chugoku*, *Shikoku*, and *Kyushu*.

Finally, precision and coverage were calculated as follows:

$$Precision = \frac{\text{\# of appropriate descriptive texts}}{\text{\# of generated descriptive texts}},$$
$$Coverage = \frac{\text{\# of shops appropriate descriptive texts cover}}{\text{\# of target shops (575)}}.$$

The calculated precision and coverage are summarized in Table 7. The overall precision was 92%. This indicates that the ad generation requires a post-process to filter out inappropriate ones, particularly for d_3 type. Since the labeled descriptive texts can be used to train a classifier to filter out such ads, precision can be further improved. The overall coverage was 89% and the average number of appropriate descriptive texts generated for a shop was 6.8 (3,476/513). A wider variety of descriptive texts can be obtained by letting the system seek the second and third best candidates and/or considering the use of elements other than those in Table 2.

These results demonstrate the feasibility of automatically generating ads by using the textual data created for the Web portal site. A number of promotional expressions can be discovered at a moderately high rate and a fairly high rate of them are appropriate for descriptive texts of ads.

6. EVALUATION 2: CTR

Our second evaluation focused on the effectiveness of the automatically generated ads on the basis of the CTR calculated using Eq. (1). To obtain reliable CTRs, we registered ads with Overture Sponsored Search and counted the numbers of impressions and clicks for them over a month period[8].

6.1 Experimental settings

We randomly selected five special feature sites among those used for the first experiment and corresponding 105 shops. For those shops, the system has generated 709 unique descriptive texts and two pairs of bid phrase and title. To evaluate the effectiveness of the ads, particularly the automatically generated descriptive texts, we created a template-based descriptive text, d_0, for each of the 105 shops as a baseline for the second half of the descriptive texts. In this template, shown in (6), k_g indicates the feature name of the special feature site to which the shop belongs.

(6) k_g ナビで お店探し
k_g-navi (site name)-with search for shops
Search for shops with k_g-navi.

By combining the resultant 814 unique descriptions with two pairs of bid phrase and title, we obtained 1,628 unique ads, and registered them at Overture Sponsored Search.

The number of impressions for an ad greatly depends on its bid phrase. To precisely evaluate the effectiveness of each ad, we placed ads having the same pair of bid phrase and landing page into a unique *campaign group* (henceforth simply group), with 210 groups in total (2 for each shop). In the PPC-based model, at most one ad within a group is displayed when a user's query is matched to a bid phrase in the group. An ad with a CTR lower than that of the others in the same group will be displayed less frequently. Alternatively, an ad with a higher CTR is displayed more frequently. Consequently, the CTR for a group can be improved if at least one automatically generated ad in the group obtains a CTR higher than that of the baseline ad in the group.

[8]From November 18th to December 26th, 2009.

Table 8: Average click-through-rate.

Type	# of ads	Avg. CTR (%)
a_0	210	1.02
a_1	210	0.89
a_2	158	0.89
a_3	164	0.97
a_4	80	0.66
a_5	94	0.83
a_6	184	0.93
a_7	196	0.84
a_8	210	0.91
a_9	196	0.83
Total	1,628	0.89

Table 9: Pairwise comparison of CTRs.

Pair of ads	N	# of a_i superior to a_0 n.s.	# of a_i superior to a_0 $p < 0.10$
a_0 vs a_1	210	73 (35%)	12 (6%)
a_0 vs a_2	158	53 (34%)	8 (5%)
a_0 vs a_3	164	73 (45%)	10 (6%)
a_0 vs a_4	80	26 (33%)	4 (5%)
a_0 vs a_5	94	31 (33%)	5 (5%)
a_0 vs a_6	184	79 (43%)	6 (3%)
a_0 vs a_7	196	71 (36%)	11 (6%)
a_0 vs a_8	210	79 (38%)	15 (7%)
a_0 vs a_9	196	76 (39%)	9 (5%)

Each ad was evaluated by comparing its CTR with that of the baseline ad in the same group. The difference in CTRs for a pair of ads was assessed using the two-sample test for equality of proportions (with continuity correction, one-sided). We used $p < 0.10$ as the significance level simply to determine whether this approach is promising.

6.2 Results

Table 8 shows the average CTR for each type of ad, where a_i denotes an ad with a descriptive text type of d_i. This macro view shows that the baseline ads, a_0, had the highest CTR among the ten ad types. In other words, the ads automatically generated from promotional texts were not necessarily superior to template-based ads.

Table 9 summarizes the pairwise comparison of CTRs. Although the a_3 type ads tended to be ungrammatical, as described in Section 5, the manually selected grammatical ads obtained a higher CTR than the baseline ads and thus contributed to improving the CTR for a number of groups comparable to the other types of ads generated from basic texts. The ads generated from other elements, except types a_4 and a_5, also improved the CTR for a number of groups. Even a_8 and a_9 type ads, which were based on elements common to a number of shops, were useful for attracting user attention. Examples (7) and (8) are ads that improved the CTR of groups the most; both were based on basic texts.

(7) k_g バイキング (buffet-style)
d_1 *XXX*。*MED* のディナーコースが女性に大好評！
XXX (shop name). Mediterranean dinner is highly commended by women!

(8) k_g 日本酒 (Japanese *sake*)
d_3 *XXX*。刺身桶盛りは、豪快さに圧巻！
XXX (shop name). Raw fish served in a bucket will impress you with its magnificence!

Table 10: **Number ratio of ads, groups, and shops that had a higher CTR than that of baseline ad.**

Unit	N	n.s.	$p < 0.10$
Ads	1,418	532 (39%)	77 (5%)
Groups	210	168 (80%)	38 (18%)
Shops	105	100 (95%)	33 (31%)

Table 10 shows the number ratio of ads, groups, and shops for which at least one of the automatically generated ads had a CTR higher than that of a_0. In total, 532 unique ads out of 1,418 (39%) had a CTR higher than the corresponding baseline ad, and the difference was significant for 77 of them. Although the ratio of ads was not particularly high, that for groups and shops was quite high. This means that our automatic generation approach using textual data also contributes to attracting consumers more effectively.

The average CTR for the 210 ads that had the highest CTR in a group was 1.5 times that for the baseline ads (1.57%/1.02%). Our next task is to determine such ads in a short-term preliminary investigation.

7. CONCLUSION

Our automatic ad generator, based on domain-specific findings obtained through previous market research and tailored for the restaurant domain, is aimed at promoting thousands of shops to a wider range of consumers more effectively and at reducing the labor, time, and money spent for promoting those shops. Our system generates shop-specific listing ads by using textual data created for promoting shops at a restaurant portal site on the Web. The only manual task is creating a couple of domain-specific patterns. Subjective evaluation showed that our system can generate ads with sufficiently high precision and coverage. A one-month experiment using Overture Sponsored Search showed that a number of the automatically generated ads had higher CTRs than the template-based baseline ads. This indicates that automatically generated ads can promote shops more effectively than template-based ads.

Motivated by these results, we plan to extend this system to all the shops on all the special feature sites. Future research includes investigation of the portability of our system to domains other than the restaurant one. The current version of our system requires textual data specific to each of thousands of contents and a couple of domain-specific patterns. Therefore, we plan to apply our system to the portal sites for which huge volumes of textual data are also available. Suitable candidates are, for example, Jalan (hotels)[9] and SUUMO (houses for rent)[10]. We will also investigate the usefulness of consumer-generated media, such as reviews collected at Web portal sites, as word-of-mouth advertising is effective in promotion.

8. ACKNOWLEDGMENTS

We are deeply grateful to the former members of our research team, Mutsumi Yokokawa, Yoshinori Iwamoto, and Ryo Kataoka, for their valuable comments and supports.

9. REFERENCES

[1] V. Abhishek and K. Hosanagar. Keyword generation for search engine advertising using semantic similarity between terms. In *Proceedings of the 9th International Conference on Electronic Commerce (ICEC)*, pp. 89–94, 2007.

[2] A. Broder, P. Ciccolo, and E. Gabrilovich. Online expansion of rare queries for sponsored search. In *Proceedings of the 18th International World Wide Web Conference (WWW)*, pp. 511–520, 2009.

[3] W. Chang, P. Pantel, A.-M. Popescu, and E. Gabrilovich. Towards intent-driven bidterm suggestion. In *Proceedings of the 18th International World Wide Web Conference (WWW)*, pp. 1093–1094, 2009.

[4] Y. Chen, G.-R. Xue, and Y. Yu. Advertising keyword suggestion based on concept hierarchy. In *Proceedings of the 1st ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 251–260, 2008.

[5] Dentsu Inc. The advertisement expenses in 2009 in Japan, 2010. http://www.dentsu.co.jp/news/release/.

[6] K. Filippova and M. Strube. Dependency tree based sentence compression. In *Proceedings of the 5th International Natural Language Generation Conference (INLG)*, pp. 25–32, 2008.

[7] B. J. Jansen and M. Resnick. Examining searcher perceptions of and interactions with sponsored results. In *Proceedings of Workshop on Sponsored Search Auctions, the 6th ACM Conference on Electronic Commerce (EC)*, 2005.

[8] A. Joshi and R. Motwani. Keyword generation for search engine advertising. In *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM) Workshop*, pp. 490–496, 2006.

[9] I. Mani and E. Bloedorn. Machine learning of generic and user-focused summarization. In *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI)*, pp. 821–826, 1998.

[10] T. Nomoto. A generic sentence trimmer with CRFs. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 299–307, 2008.

[11] M. Oka, T. Koyama, and Y. Ueda. Phrase construction method for phrase-represented summarization. In *Information Processing Society of Japan SIG Notes, NL-129-15*, pp. 101–108, 1999. (in Japanese).

[12] S. Ravi, A. Broder, E. Gabrilovich, V. Josifovski, S. Pandey, and B. Pang. Automatic generation of bid phrases for online advertising. In *Profeedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 201–210, 2010.

[13] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th International World Wide Web Conference (WWW)*, pp. 521–529, 2007.

[14] X. Wang, A. Broder, M. Fontoura, and V. Josifovski. A search-based method for forecasting ad impression in contextual advertising. In *Proceedings of the 18th International World Wide Web Conference (WWW)*, pp. 491–450, 2009.

[9] http://www.jalan.net/

[10] http://suumo.jp/