

## AN ANNOTATION APPROACH TO CONTEXTUAL ADVERTISING FOR ONLINE ADS

Yen-Liang Chen  
Department of Information Management  
National Central University  
Chung-Li, Taiwan 32001, R.O.C.  
[ylchen@mgt.ncu.edu.tw](mailto:ylchen@mgt.ncu.edu.tw)

Zong-Cheng Chen  
Department of Information Management  
National Central University  
Chung-Li, Taiwan 32001, R.O.C.  
[954003201@cc.ncu.edu.tw](mailto:954003201@cc.ncu.edu.tw)

Yu-Ting Chiu  
Information & Communication Security Laboratory  
Chunghwa Telecom Laboratories  
Chung-Li, Taiwan 32601, R.O.C.  
[g.chiu07@gmail.com](mailto:g.chiu07@gmail.com)

Chia-Ling Chang  
Department of Information Management  
National Central University  
Chung-Li, Taiwan 32001, R.O.C.  
[vena92207451@gmail.com](mailto:vena92207451@gmail.com)

### ABSTRACT

In recent years, contextual advertising has been widely applied to text-based web advertising. Because contextual advertising relies on text to determine the advertising context, it cannot be applied to multimedia advertisements that have no text. Therefore, this paper proposes an annotation-based contextual advertising method for matching multimedia advertisements and web pages. In the proposed method, a term vector represents a web page and an advertisement is represented by a small set of keyword tags that are either annotated by experts or extracted automatically from the text of the advertisement. This study then developed a matching mechanism for computing the similarity between a web vector and the advertisement tags. Experiments and evaluations were performed for demonstrating the performance of the proposed method. The results showed that the proposed method demonstrated a more favorable performance than traditional advertising methods.

Keywords: Contextual advertising; Tag; Recommendation; Online advertisements; Annotation

### 1. Introduction

Web advertising has recently become one of the most commonly used marketing channels. Mu and Galletta reported that web advertising is one of the top three advertising mediums worldwide [2007]. Contextual advertising [Anagnostopoulous et al. 2007; Broder et al., 2007; Ciaramita et al., 2008] is one of the major approaches to text-based web advertising; contextual advertising is a form of targeted advertising in which the content of an advertisement is directly correlated to the content of the web page a user is viewing. An example includes a user visiting a website concerning traveling in Asia and then seeing an advertisement pop-up offering a special price on a flight to Taiwan. Contextual advertising is also called “In-Text” advertising or “In-Context” technology. The essence of this technique is to place an advertisement on a website on the basis of content similarity between advertisements and web pages. The success of contextual advertising depends on the ability of a system to determine which advertisements are most relevant to the content of a web page.

In the past, two main methods have been used for matching advertisements and web pages, and these are the vector space model and keyword-based model. In the vector space model, the advertisements and pages are represented as term vectors in a vector space, and the matching process is typically based on the similarity or correlation between an advertisement vector and a page vector. In the keyword-based model, the advertisements and pages are represented as a set of phrases or keywords, and the matching process is typically based on crossover between an advertisement keyword set and a page keyword set.

Multimedia advertisements have become increasingly popular because advertisers prefer vivid multimedia content with no text—such as a video, animation, image, or sound—rather than advertisements with long text, which do not draw the attention of users. Traditional approaches to contextual advertising have the following drawbacks when addressing multimedia advertisements.

- (1) Although the vector space model is suitable for representing web pages, it is not ideal for advertisements; this phenomenon is due to multimedia advertisements having no text. The context of an advertisement cannot be determined without text.
- (2) The keyword-based model cannot represent web pages efficiently because only a few keywords are selected to represent the page, rather than a full vector. The potential information contained in a web page is not fully used, thus resulting in a decreased performance.

A research question arising immediately is whether a new model that is suitable for contextual advertising with multimedia advertisements can be developed. Therefore, this study combined the strengths of both the vector space and keyword-based models to develop a new model for multimedia advertisements; in other words, we represented web pages and advertisements by using the vector space and keyword-based models, respectively. Because advertisers understand their advertisements clearly, this study assumed that advertisers assume responsibility for annotating advertisements. Annotating multimedia resources is a commonly used operation in numerous current e-commerce sites, such as Flickr, Twitter, and YouTube; therefore, annotating the advertisements should not be a difficult task for advertisers. The proposed approach comprises the following parts:

- (1) Multimedia advertisements, such as images or videos, are distinguished by annotating keywords or phrases as tags.
- (2) Web pages are represented as vectors.
- (3) A mechanism is proposed to match the advertisement keywords with page vectors.

The proposed approach has the following advantages. (1) The model is simple and easy to use; advertisers can easily annotate their advertisements. (2) The model can determine the context of multimedia advertisements, even without text. (3) The proposed model was developed by combining the strengths of two traditional approaches to contextual advertising.

To increase the capabilities of the proposed model further, we also propose an extended method for extracting representative tags from advertisements containing sufficient text information. By using this extension, the proposed method can be applied to both multimedia-based and text-based advertisements. Advertisers are not required to annotate the advertisements when the advertisements contain text.

In summary, the contributions of this paper are as follows:

- (1) We propose a new hybrid model to represent web pages and advertisements. The model integrates the vector space and keyword-based models.
- (2) We designed a new matching mechanism for computing the similarity between an advertisement keyword set and a page vector.
- (3) We designed a tag-assigning mechanism to annotate advertisements with text automatically.

## 2. Related Work

In this section, online advertising is reviewed first, and existing approaches for matching web pages and advertisements are subsequently introduced.

### 2.1. Online Advertising

The Internet has considerably changed the behavior of people because of the tremendous amount of information it contains. Therefore, Internet advertising has become critical for companies. According to the Internet Advertising Revenue Report [IAB, 2011], online advertising revenue increased by 23.2% and reached 14.9 billion in the second quarter of 2011. Early Internet advertising was similar to offline advertising, and users were more likely to remember advertisements that were graphical and less text-based [Plessis, 2005]. HotWired first introduced online advertising, in 1994, in the form of a graphical banner [Chowdhury, 2007]. Athanasiadis and Mitropoulos [2010] proposed a prototype system that provided personalized advertisements on digital interactive televisions via the Internet. Through the adoption of web and search engine technology, methods of online advertising have varied over time.

However, online advertisements are mostly provided through text-based advertising methods, similar to traditional approaches. Two main categories of text-based advertising have been developed: sponsored searches and contextual advertising [Anagnostopoulous et al., 2007; Broder et al., 2007]. In a sponsored search, advertisements are derived from the query on a web search engine and are placed on the web page of the search engine beside the search results [Lin and Hung, 2009; Jansen and Schuster, 2011; Xiao et al., 2009]. In other words, numerous sponsored links are displayed on the right side of the screen in the list returned by the search engine; this is called “sponsored search advertising.” All major web search engines (e.g., Google, Microsoft, and Yahoo! Search) support sponsored advertisements and function simultaneously as a web search and advertisement search engine.

A sponsored search comprises six items [Fain and Pedersen, 2006]: (1) advertiser-provided content: a set of advertiser hyperlinks annotated with keyword tags, titles, and descriptions; (2) advertiser-provided bids that value traffic on specified concepts or keywords; (3) combining a manual and automated review process to ensure that the content of an advertisement is relevant to the target keywords; (4) matching advertisement content to user queries as they are received by the search engine; (5) displaying advertisement content in a specific order alongside other algorithmic (i.e., nonsponsored) search engine content; and (6) gathering data, meter clicks, and charging advertisers based on the number of consumer clicks on their displayed content.

In contextual advertising (or Internet display advertising), however, commercial advertisements are placed on any given web page. In other words, advertisers reach a target Internet audience through a visual or multimedia-based advertisement. Almost all for-profit, nontransactional websites currently rely on advertising revenue to a certain extent. Content matching supports sites ranging from individual bloggers and small niche communities to large publishers, such as major newspapers. The web would be a lot smaller without this model.

Contextual advertising comprises four main entities that interact with each other: the publisher, advertiser, ad-network, and users [Anagnostopoulous et al., 2011; Broder et al., 2007; Niu et al., 2009]. Similarly, Li and Jhang-Li [2009] reported that the four main members are the publishers, advertisers, channel providers, and visitors. Publishers offer advertisements with web page content, thus attracting visitors. Publishers rent space on their web pages in return for revenue; value is typically measured according to the amount of revenue and user retention. Advertisers provide a pool of advertisements for different web pages with the aim of increasing product sales. Users’ interest in the advertisers’ products is increased by viewing attractive advertising.

The ad-network (channel provider) serves as an intermediary between publishers and advertisers; the ad-network places advertisements on websites according to the content of the web page and shares revenue with the publisher. Specifically, an ad-network helps advertisers publish or post their advertisements on appropriate web pages. Most advertisers attempt to reach users who visit the publishers’ websites.

Advertisers target users who surf the Internet; they expect an advertisement to pique a user’s attention and for the user to click on it. In addition, because an advertisement is an online recommendation [Mohanraj et al., 2012], it captures user intention and can suggest further information for that user. The publisher earns money through users’ clicks. Publishers use numerous models to create profit [Mahdian and Tomak, 2007; Nazerzadeh et al., 2008]: pay-per-click, where advertisers pay for every click on their advertisements; pay-per-impression, where advertisers are charged for the number of advertisement exposures; and pay-per-action, where advertisers pay for every transaction that leads to a sale.

## 2.2. Existing Approaches for Matching Pages and Advertisements

There are two main existing approaches for matching an advertisement and a web page: the keyword-based model and vector space model. The keyword-based model extracts phrases from the page and matches them with the bid phrases of the advertisements. Turney [2000] proposed Turney’s GenEx system, one of the most well-known programs for keyword extraction. Carrasco et al. [2003] proposed a clustering method involving bipartite advertiser keyword graphs for suggesting keywords and identifying advertisers’ groups. Yih et al. [2006] proposed a system for training the features and term frequency of keywords to extract keywords from unseen web pages. Jang et al. [2007] established relationships with different keywords in ontologies that comprised two parts: one part was used for keywords and the other part was used for advertisements. They used a Keyword Extraction Module to extract keywords from web pages and advertisements, and subsequently used the Apriori algorithm to determine the relationships among various keywords. Li et al. [2007] attempted to determine language patterns from online content, such as web pages, by using language pattern mining and keyword extraction. Their results showed that the performance of the keyword extraction algorithm they proposed was superior to that of the N-Gram-based method. Liu et al. [2014] also proposed a method for extracting and suggesting keywords. They designed an approach by considering Part-of-Speech (POS) and named-entities tagging. Khan et al., [2009] extracted optimal keywords by performing term frequency and inverse document frequency ( $tf*idf$ ) on web pages. They also extracted keywords from various parts of the advertisement, and subsequently matched the extracted keywords by using a matching function. After matching, the most relevant advertisement was displayed on the target web page. Liu et al. [2010]

created a text network for a single web page and applied the PageRank algorithm to determine the highest-ranked keywords. Dave and Varma [2010] proposed systems that were designed based on the chunking method. A naive Bayes classifier was used to train the systems through web pages annotated with advertising keywords. The systems subsequently determined keywords from unseen web pages.

The second approach is the vector space model. Ribeiro-Neto et al. [2005] matched web pages and advertisements by using several strategies. They presented the advertisements and pages as vectors in a vector space, and then matched them according to the cosine of the angles between the vectors. Broder et al. [2007] considered topical proximity and discovered that the standard string matching approach could be improved according to the taxonomy of topics. They classified the pages and advertisements through taxonomy, and then calculated the similarity of the pages and advertisements. The taxonomy similarity scores were based on vector cosines, and they reported a 25% improvement for midrange recalls of the vector cosine model. Lee et al. [2013] established a topic taxonomy hierarchy for classifying web pages and advertisements. They also ranked the advertisements according to the topical relevance. Anagnostopoulos et al. [2007] presented Just-in-Time contextual advertising by extracting words and classification features; they then represented them as vectors and calculated the similarity scores between pages and advertisements by using both the words and classification features. For suggesting advertisements more quickly, Anagnostopoulos et al. [2011] proposed a method for summarizing advertisements for further matching. Fan and Chang [2011] used the vector model to evaluate the similarities between web pages and advertisements. They preprocessed the terms on the page and advertisement by using a sentiment detection process, and then matched the terms by using cosine similarity functions. Vargiu et al. [2013] introduced a hybrid method that involves a combination of collaborative filtering and inlink (if the topic of web page  $a$  is related to web page  $b$ , there is an inlink between pages  $a$  and  $b$ ) information for classifying web pages and suggesting suitable advertisements. Pak and Chung [2010] used Wikipedia for improving the precision of contextual advertising. They used Wikipedia articles, web pages, and advertisements as vectors, and then used cosine similarity to determine the shortest distance among the articles, pages, and advertisements.

According to the foregoing literature review, the vector space model is suitable for matching advertisements with webs if the advertisements contain a high amount of text. This is because the vector space model is an algebraic model for representing text documents as vectors of index terms extracted from data. Therefore, if the advertisements contain a high amount of text, the vector space model can represent advertisements adequately by transforming each advertisement into a corresponding vector of index terms appearing in the advertisement. However, when the advertisements have only a low amount of text, previous studies have relied on the keyword-based model to represent the advertisements. The keyword-based model, however, has the limitation of being unable to represent web pages efficiently because only a few keywords, rather than a full vector, are selected to represent the page; the potential information contained in a web page is not fully utilized, thus resulting in decreased performance.

In summary, the vector space model is not suitable for representing advertisements containing a low amount of text, and the keyword-based model is not ideal for representing web pages containing a high amount of text. These two drawbacks indicate a research gap that can be filled by developing a new model that involves integrating the strengths and simultaneously avoiding the limitations of both models. Therefore, we propose a hybrid model that represents web pages and advertisements by using the vector space and keyword-based models, respectively. In addition, we also developed a new mechanism based on the WordNet lexical database for matching web pages with advertisements, which are represented in separate models. To our knowledge, this is the first study to propose a hybrid method involving a vector space model and keyword-based model for resolving the advertisement matching problem.

### 3. Research Design

As shown in Figure 1, the proposed approach comprises three main phases, and these are (1) the advertisement tag addition phase, (2) vector transforming phase, and (3) similarity matching phase. In the first phase, experts select appropriate keywords as tags and manually add them to advertisements without text. For advertisements containing text, we propose a specific approach for recommending tags (Section 3.3). In the second phase, each web page, which consists of several terms, is represented as a vector (Section 3.1). In the final phase, a vector is generated for each advertisement tag and the tag vectors are subsequently aggregated into a new advertisement vector (Section 3.2). The similarity between an advertisement and a web page can then be calculated through cosine similarity. The three phases are not presented in chronological order. The first phase is presented last because it contains supplementary material regarding how tags can be assigned when advertisements contain text.

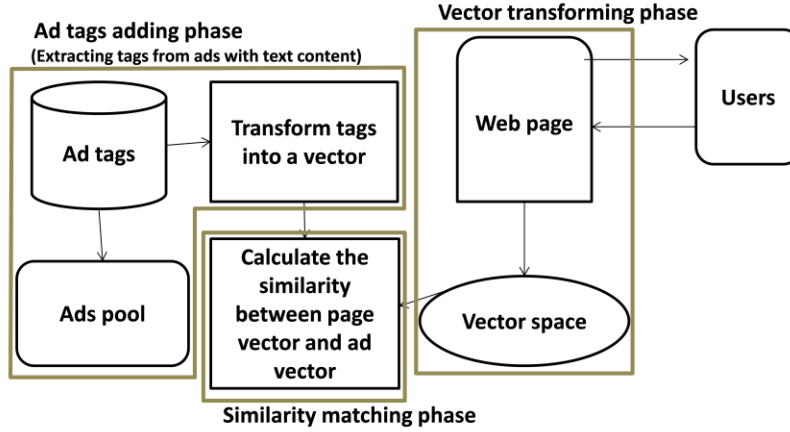


Figure 1: Architecture of the system

3.1. Vector Transforming Phase

The vector transforming phase consists of two main steps: the preprocessing steps and feature selection (Figure 2).

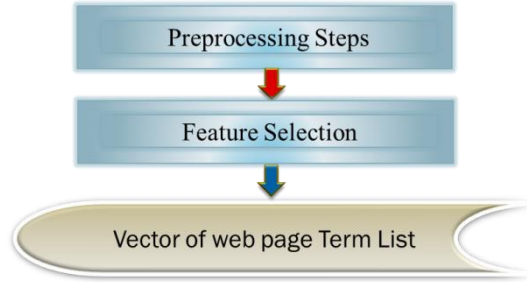


Figure 2: Vector transforming phase

The preprocessing part of the vector transforming phase comprises three steps. First, we extracted terms and keywords through a word segmentation process. Because we used an English-language data set in the experiment, a word is the smallest independent, meaningful element. We used an English lexical analysis tool, the Stanford Log-linear POS Tagger, for extracting terms. This tool reads text and assigns a POS, such as a noun, verb, or adjective, to each word (and other tokens). We used this tool to generate word tokens with POS tags for the content of the web pages [Toutanova and Manning, 2000; Toutanova et al., 2003].

Second, we conducted a stop-word elimination process because not all the word tokens generated by the Stanford Log-linear POS Tagger were meaningful. The stop-words were removed based on an English stop-word list.

Even after removing the stop-words, numerous noise terms remained. Therefore, we used POS tags to determine meaningful candidate terms. Because most documents can be represented by nouns, for reducing the noise, we did not consider adjective and verb terms.

After the preprocessing steps, the second part of the vector transforming phase involved feature selection. Because we used the vector space model to represent web pages, and because there are numerous crucial feature terms, we represented a web page by using a page term set  $\{t_1, t_2, \dots, t_n\}$ , where  $n$  denotes the number of terms that occur in the web page. A weight  $w_{i,j}$  is assigned to term  $t_i$  in web page  $p_j$  to represent the relevance between  $t_i$  and  $p_j$ . We assigned the weights by using a weighting scheme  $tf_{i,j} * idf_i$ .

$$w_{i,j} = tf_{i,j} \times idf_i$$

where  $tf_{i,j}$  denotes the normalized frequency of term  $t_i$  in web page  $p_j$ . The variable  $idf_i = \log(m/m_i)$  represents the inverse document frequency of term  $t_i$ , where  $m$  is the total number of web pages and  $m_i$  denotes the number of web pages containing term  $t_i$ .

The web page vectors were obtained after completing this step. Matrix  $W$  records the weight  $w_{i,j}$  for every term  $t_i$  in web page  $p_j$ .

### 3.2. Similarity Matching Phase

In the similarity matching phase, we attempted to calculate the similarity between a web page and an advertisement. Because the advertisement tags are not vectors, the tag keywords must be transformed into vectors that are subsequently aggregated into a new advertisement vector. Finally, the cosine similarity between a web page and the advertisements was calculated for determining relevant advertisements for the web page.

We used a well-known method, WordNet::Similarity on WordNet, for transforming the tags. WordNet is an online lexical reference system, the design of which was inspired by current psycholinguistic theories of human lexical memory [Miller et al., 1990]. English nouns, verbs, and adjectives are organized into synonym sets, with each set representing one underlying lexical concept. Different relations link the synonym sets. WordNet::Similarity is a freely available software package that enables measuring the semantic similarity and relatedness between a pair of concepts (or “synsets”). This software package provides six measures of similarity and three measures of relatedness based on the WordNet lexical database [Pedersen et al., 2004].

Three of the six similarity measures are based on the content of information of the least common subsumer of Concepts A and B. These three measures are res [Resnik, 1995], lin [Lin, 1998], and jcn [Jiang and Conrath, 1997]. The other three similarity measures—lch [Leacock and Chodorow, 1998], wup [Wu and Palmer, 1994], and path—are based on the path lengths between a pair of concepts. Relatedness measures are more general, in that they can be performed across the boundaries of POS, and they are not limited to considering is-a relationships. There are three such measures in the software package: hso [Hirst and St-Onge, 1998], lesk [Banerjee and Pedersen, 2003], and vector [Patwardhan, 2003].

The first step in the similarity matching phase (Figure 3) involves generating a vector for each tag associated with an advertisement by using the WordNet::Similarity method. A tag  $g_i$  can be related to the term list  $t_{1,j}, t_{2,j}, \dots, t_{n,j}$  in a web page  $p_j$  to various degrees. The new vector  $(v_{1,j,i}, \dots, v_{n,j,i})$  is an n-dimensional vector that represents the similarity between tag  $g_i$  and the term list of page  $p_j$ .



Figure 3: Advertisement term vector transformation

For advertisements containing text, tags are automatically assigned through the tag recommending approach. However, for advertisements without text, tags are recommended by experts. A value  $gw_i$  is then assigned to tag  $g_i$ , where the value represents the importance of  $g_i$  to the advertisement. Equal values are assigned to the tags if they are unordered. However, if the tags are ordered, higher values are assigned to tags that have a higher priority. The similarity value  $v'_{k,j,i}$  between tag term  $g_i$  and term  $t_k$  in web page  $p_j$  is calculated using WordNet::Similarity, as follows:

$$v'_{k,j,i} = 0 \quad \text{if } \text{WordNet::Similarity}(t_{k,j}, g_i) < \gamma,$$

$$\text{otherwise, } v'_{k,j,i} = \text{WordNet::Similarity}(t_{k,j}, g_i)$$

where  $\gamma$  denotes a threshold value, and  $\text{WordNet::Similarity}(t_{k,j}, g_i)$  represents the similarity value between term  $t_{k,j}$  and tag term  $g_i$ .

The  $ratio_{k,j,i}$  is subsequently obtained by dividing  $v'_{k,j,i}$  by the sum of  $v'_{k,j,i}$  over all web pages.

$$ratio_{k,j,i} = \frac{v'_{k,j,i}}{\sum_{k=1}^n v'_{k,j,i}}$$

Finally, the weight of  $gw_i$  was distributed to web page terms  $p_j$  according to  $ratio_{k,j,i}$  to obtain  $v_{k,j,i}$ .

$$v_{k,j,i} = gw_i \times ratio_{k,j,i}$$

In the following example (Figure 4), a tag  $g_i$  can be represented by the vectors of  $t_{1,j}$ ,  $t_{2,j}$ , and  $t_{3,j}$ .

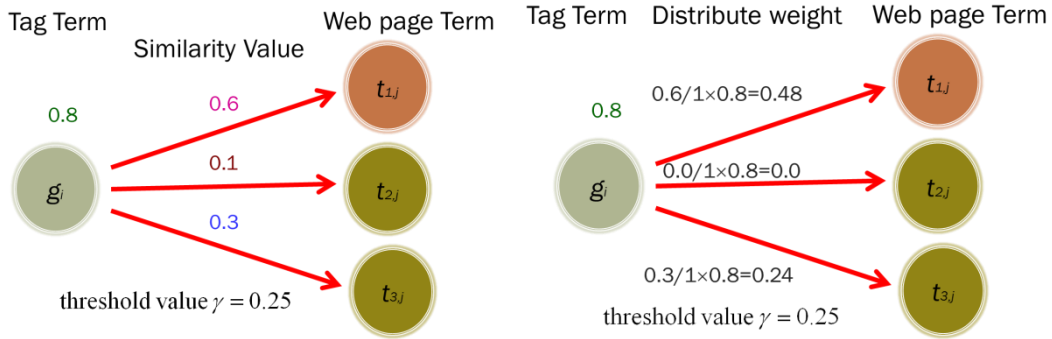


Figure 4: Example of advertisement vector transformation

After transforming every advertisement tag into a vector of terms, we summed all the vectors into a new vector that represents the entire advertisement. The weight  $v_{k,j,i}$  of each tag term was summed according to the following formula, as shown in Figure 5.

$$a_{k,j} = \sum_{i=1}^m v_{k,j,i}$$

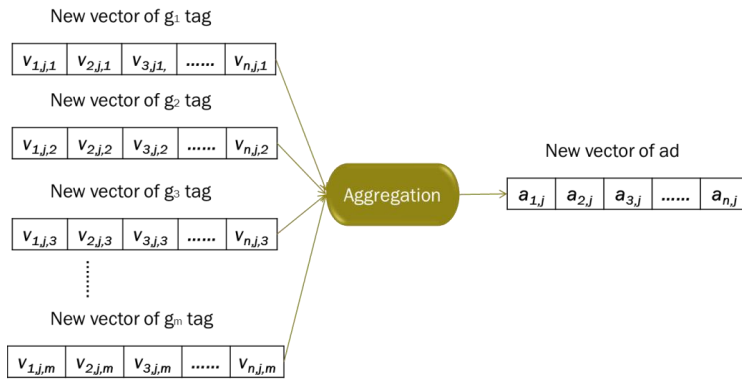


Figure 5: Advertisement vector aggregation

In the final step of the similarity matching phase, we used the similarity between the new advertisement vector and vector of the web page term list to rank the advertisements (Figure 6). We computed the similarity value between the two vectors by using cosine similarity. A cosine similarity function measures how closely an advertisement is related to a web page  $p_j$ , on a scale of 0 to 1.

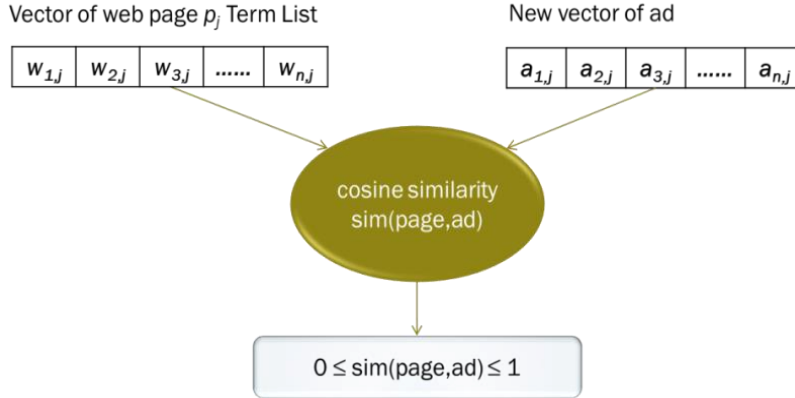


Figure 6: Similarity between a web page and an advertisement

$$\text{sim}(p_j, ad) = \frac{\vec{p}_j \cdot \vec{ad}}{|\vec{p}_j| |\vec{ad}|} = \frac{\sum_{k=1}^n w_{k,j} \times a_{k,j}}{\sqrt{\sum_{k=1}^n w_{k,j}^2} \times \sqrt{\sum_{k=1}^n a_{k,j}^2}} \quad 0 \leq \text{sim}(p_j, ad) \leq 1$$

where  $p_j$  denotes the web page,  $ad$  represents the new advertisement vector,  $w_{k,j}$  is the weight of term  $k$  on page  $p_j$ , and  $a_{k,j}$  denotes the weight of term  $k$  in  $ad$ . A value of zero indicates that the advertisement and web page are completely dissimilar. A higher similarity value indicates a higher correlation between the advertisement and web page. Accordingly, we obtained the top  $r$  recommended advertisements.

### 3.3. Extracting Tags from Advertisements Containing Text

Some advertisements contain a high amount of text. In these cases, we propose a tag recommending approach to extract automatically crucial keywords and add them to each advertisement as tags.

PageRank is a link analysis algorithm [Wikipedia, 2012] named after Larry Page and used by the Google search engine. This algorithm assigns a numerical weight to each element of a hyperlinked set of documents, such as the World Wide Web, to measure the relative importance of the element within the set. This algorithm can be applied to any collection of entities containing reciprocal quotations and references. The numerical weight that this algorithm assigns to any given element  $E$  is called the PageRank of  $E$  and denoted by  $PR(E)$ .

We developed a PageRank-like method for selecting tags from advertisements. The difference between our approach and PageRank is that we considered PageRank scores as well as tf\*idf values in determining the highest-ranked keywords. We considered both factors because (1) the tags used to represent an advertisement must have close relationships with other keywords in the advertisement, and (2) the tags must be representative words in the advertisement. Criterion (1) implies the necessity of including PageRank scores [Wikipedia, 2012] and Criterion (2) implies the necessity of including tf\*idf values.

In the next step, we defined the text network of a web page. First, we extracted terms from advertisements and calculated the tf\*idf weights of each term. We then established the text network by calculating the WordNet::Similarity values among the advertisement terms. The network was denoted  $\text{Graph} = \{V, E, W\}$ , where  $V$  represents a set of nodes,  $E$  denotes a collection of edges, and  $W$  is the weights of edges. In addition, node  $V_i$  represents a term in the advertisement with an initial value, edge  $E_{i,j}$  denotes a link connecting two nodes, and weight  $W_{i,j}$ , which is defined in the following equation, is the linking score from node  $i$  to node  $j$ .

$$W_{i,j} = \frac{D_{i,j}}{\sum_{i \text{ and } i \neq j} D_{i,j}}$$

where  $D_{i,j}$  denotes the similarity between  $V_i$  and  $V_j$  determined using WordNet::Similarity, and  $W_{i,j}$  represents the weight of the edge from  $V_i$  to  $V_j$ .

The text network was then constructed and the weights between two nodes was calculated. Figure 7 presents an example where weight  $W_{1,5}$  for  $V_1$  to  $V_5$  is 0.25.



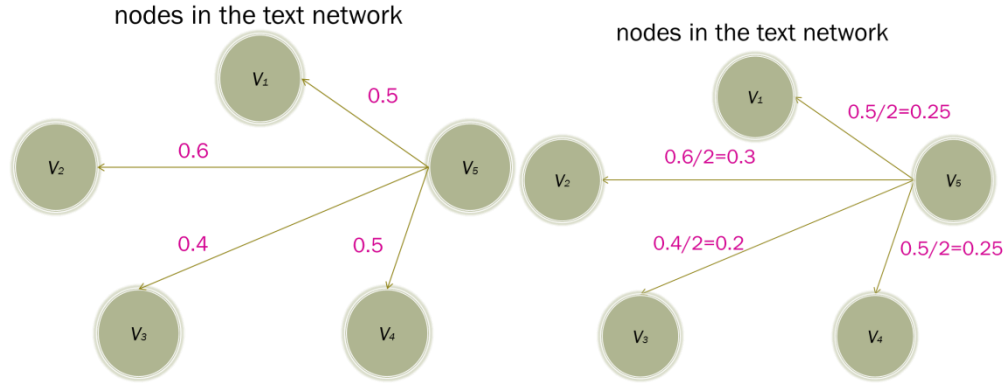


Figure 7: Example of text network and weight computation

After creating the text network and edge weights, we used the PageRank algorithm for calculating the PageRank score of vertex  $V_i$ , as follows:

$$S(V_i) = (1 - d) + d \times \sum_{j \in C(V_i)} \frac{W_{j,i} \times S(V_j)}{N(V_j)}$$

where  $V_i$  denotes the  $i$ -th node of the text network,  $S(V_j)$  represents the score of  $V_j$  with an initial value,  $C(V_i)$  is the set of vertices that have arcs entering  $V_i$ ,  $N(V_j)$  denotes the sum of the weights of outbound arcs of  $V_j$ ,  $W_{ji}$  represents the weight of edge  $E_{ji}$ , and  $d$  is a damping factor that can be set between 0 and 1 (typically set at 0.85).

Subsequently, we applied the PageRank algorithm to the text network iteratively until it converged. We then multiplied the PageRank score of vertex  $V_i$  with the  $tf \cdot idf$  value of vertex  $V_i$  to obtain the relevance of vertex  $V_i$ :  $I(V_i)$ .

$$I(V_i) = S(V_i) \times (tf_{i,j} * idf_i)$$

where  $S(V_i)$  denotes the PageRank score of  $V_i$ , and  $(tf_{i,j} * idf_i)$  represents the  $tf \cdot idf$  value of  $V_i$  in the  $j$ -th advertisement. The value  $I(V_i)$  indicates the relevance of a term; this is because a term with a high  $tf \cdot idf$  value means this term is a crucial and representative word in the advertisement, and a term with a high PageRank value means it is closely related to several terms in the advertisement. In other words, a term with a high importance value is a term that is representative of the advertisement and also closely related to other words in the advertisement. Finally, we recommended the top  $k$  vertices that registered higher  $I(V_i)$  scores as the advertisement tags.

#### 4. Experiments and Evaluation

We conducted two experiments, Experiment 1 and Experiment 2, for evaluating the proposed approach. The advertisements in Experiment 1 contained text, whereas those in Experiment 2 comprised only images. Accordingly, the tags in Experiment 1 were extracted from the text automatically, whereas those in Experiment 2 were provided by experts. Because advertisements that contained text were used in Experiment 1, we compared the proposed method with the traditional approach, which involves transforming advertisements into term vectors by using the  $tf \cdot idf$  formula and then recommending advertisements that demonstrate the highest similarities to the web vector. Because advertisements that did not contain text were used in Experiment 2, we traced those advertisements in our advertisement pool to determine the pages they were linked to; we then collected noun words manually from the titles of these linked pages for each advertisement. In other words, we selected nouns appearing on the linked pages for each advertisement and then used the traditional method for obtaining the most relevant advertisements for each web page. Next, we compared the proposed method with the traditional approach. Multimedia advertisements had no text/information on them; therefore, the only feasible approach involved gathering terms from the title on the out-linked pages as keywords.

During the evaluation stage, we invited 45 users to participate in our experiments. Each user provided satisfaction scores ranging from 1 to 5, with a higher number indicating a higher level of satisfaction.

The satisfaction score indicates the relevance and relatedness between a web page and the suggested advertisements. Because the web pages and advertisements used in the experiments were randomly collected from the Yahoo! News website and Internet, respectively, the degree of satisfaction was not related to the information on the web page or design of the advertisements. If participants perceived that the content of a web page was highly relevant to the content of the advertisements, they expressed satisfaction by providing a high score. However, if the advertisements were unrelated to the current web page, the participants provided a low score.

We subsequently used the statistical tool, paired  $t$  tests, to compare the scores of the two approaches and test for significant differences. If significant differences existed, the approach that registered a higher average score was more favorable.

#### 4.1. Experiment 1: Advertisements Containing Text

In Experiment 1, we randomly collected 30 articles from news web pages and 30 advertisements containing text from the Yahoo! website. These web pages are evenly divided into four main categories (Entertainment, Sports, Technology, and Health) with 10 topics, including entertainment movies, entertainment music, entertainment fashion, entertainment books, sports, tech Internet, tech social media, entertainment TV, tech gaming, and health. In the traditional method, the  $tf*idf$  formula was used to recommend the top one, three, and five advertisements for each web page. In the proposed method, we also recommended the top one, three, and five advertisements for each web page. Table 1 lists all the parameters and their descriptions.

Table 1: Parameters of Experiment 1

Parameters	Description
$D$	A damping factor of PageRank that can be set between 0 and 1.
$S(V_j)$	The score of vertex $V_j$ .
$\Gamma$	The threshold value of similarity value between tag and term.
Lin measure	The basic measure of WordNet::Similarity measures.
$gw_i$	The importance of $g_i$ for the ad.

We set the damping factor ( $d$ ) of PageRank to 0.85, and the initial score of vertex  $S(V_j)$  to 1. We tested the threshold value  $\gamma$  iteratively in the experiments, and determined the optimal threshold as 0.75. Because WordNet::Similarity provides six measures of semantic processing, we used the basic measure—the Lin measure—as the WordNet::Similarity measure. We also considered the tags as a term set, meaning the weights of the tags were all the same. Therefore, we set the relevance  $gw_i$  value to 1 for all the tags.

During the evaluation stage, the participants provided satisfaction scores regarding the top one, three, and five advertisements recommended by the proposed method and traditional approach, respectively. Table 2 shows the average satisfaction scores for the top one, three, and five recommended advertisements in both approaches (i.e., our\_approach and Trad\_approach), indicating that the scores for the proposed approach are all higher than those of the traditional  $tf*idf$  approach.

Table 2: Average satisfaction scores for both approaches (Experiment 1)

	One Sample Mean			
	Observation	Mean	Standard deviation	Standard error of the mean
our_approach_top1	45	2.8096	0.5419	0.0120
Trad_approach_top1	45	2.7037	0.6450	0.0143
our_approach_top3	45	3.0400	0.6674	0.0148
Trad_approach_top3	45	2.9405	0.6768	0.0150
our_approach_top5	45	3.0333	0.6073	0.0135
Trad_approach_top5	45	2.9511	0.6470	0.0144

As shown in Table 3, the  $p$  values for the one-tailed and two-tailed paired  $t$  tests between the two approaches were less than 0.05. Significant differences were observed between the two approaches (i.e., the proposed method and traditional approach) regarding the top one, three, and five advertisements. Thus, the proposed approach significantly outperformed the traditional approach.

Table 3: Paired *t*-test between Our Approach and tf\*idf Approach

Paired <i>t</i> -Test						
Experiment 1	our_top1	Trad_top1	our_top3	Trad_top3	our_top5	Trad_top5
Mean	2.8096	2.7037	3.0400	2.9405	3.0333	2.9511
Variance	0.2937	0.4160	0.4454	0.4580	0.3688	0.4186
Observation	45	45	45	45	45	45
Degree of freedom		44		44		44
<i>t</i>		2.0699		2.4520		2.1571
$p(T \leq t)$ One-tailed		0.0222		0.0091		0.0183
Critical value: One-tailed		1.6802		1.6802		1.6802
$p(T \leq t)$ Two-tailed		0.0444		0.0182		0.0365
Critical value: Two-tailed		2.0154		2.0154		2.0154

Because the scores for the top three recommended advertisements were higher than those for both the top one and top five recommendations (Table 2), we used paired *t* tests to determine whether the differences between the top one and three, and top three and five advertisements were significant. Table 4 shows the results of the paired *t* tests of the top three advertisements (and other advertisements) for the proposed method and traditional tf\*idf approach. The results indicated that recommending the top three advertisements demonstrated the best result among these three options, implying that providing excessive or insufficient advertisements is not beneficial. For maximizing advertising effects, advertisers must consider how to provide users with the most appropriate amount of information.

Table 4: Paired *t*-test of two recommended advertisement situations

Paired <i>t</i> -Test								
Experiment 1	our_top1	our_top3	our_top3	our_top5	Trad_top1	Trad_top3	Trad_top3	Trad_top5
Mean	2.8096	3.0400	3.0400	3.0333	2.7037	2.9405	2.9405	2.9511
Variance	0.2937	0.4454	0.4454	0.3688	0.4160	0.4580	0.4580	0.4186
Observation	45	45	45	45	45	45	45	45
Degree of freedom		44		44		44		44
<i>t</i>		-4.2291		0.1592		-4.8501		-0.2128
$p(T \leq t)$ One-tailed		0.0001		0.4371		0.0000		0.4162
Critical value: One-tailed		1.6802		1.6802		1.6802		1.6802
$p(T \leq t)$ Two-tailed		0.0001		0.8742		0.0000		0.8325
Critical value: Two-tailed		2.0154		2.0154		2.0154		2.0154

#### 4.2. Experiment 2: Advertisements without Text

The same 30 articles collected in Experiment 1 were used in Experiment 2. Moreover, we collected 30 advertisements without text from the Yahoo! portal. Two experts assigned five tags to each advertisement. During the coding process, if the two experts reach a consensus on the selected terms for an advertisement, then the consensual terms represented the advertisement. However, if the two experts have no consensus, then the experts took turns to provide a term to represent the advertisement until they determined and suggested a total of five terms/tags. We used a statistical measure, Cohen's kappa in R, for estimating the intercoder reliability score of the tagged terms. After the assessment process, we obtained a kappa value of 0.714, thus indicating substantial agreement.

We next used the proposed approach to recommend the top one, three, and five advertisements. In addition, noun keywords discovered by tracing the links of the advertisement were used to apply the traditional vector space model approach, using the tf\*idf formula. Therefore, we recommended the top one, three, and five advertisements for each web page as a comparative approach (called traditional approach).

Regarding the parameters, we tested the threshold value  $\gamma$  repeatedly, and determined the optimal threshold as 0.75. Because WordNet::Similarity provides six measures of semantic processing, we used the basic measure, the Lin measure, as the WordNet::Similarity measure. We also considered the tags as a term set, meaning the weights of the tags were all the same. Therefore, we set all the relevance values of the tags to 1.

During the evaluation stage, the invited users provided scores for the top one, three, and five recommended advertisements in the proposed approach as well as the traditional approach. Table 5 shows the average satisfaction

scores for the top one, three, and five recommended advertisements in both approaches, indicating that all scores for the proposed approach are higher than those for the traditional approach.

Table 5: Average satisfaction scores in both approaches (Experiment 2)

<b>One Sample Mean</b>				
	Observation	Mean	Standard deviation	Standard error of the mean
our_approach_top1	45	2.6696	0.6324	0.0141
Trad_approach_top1	45	2.5037	0.7131	0.0158
our_approach_top3	45	2.7089	0.7337	0.0163
Trad_approach_top3	45	2.3570	0.6950	0.0154
our_approach_top5	45	2.7333	0.7549	0.0168
Trad_approach_top5	45	2.3707	0.7094	0.0158

As shown in Table 6, the  $p$  values for the one-tailed and two-tailed paired  $t$  tests between the two approaches were less than 0.05. We observed significant differences between the two approaches regarding the top one, three, and five advertisements. The proposed approach thus significantly outperformed the traditional method.

Table 6: Paired  $t$ -test between the proposed approach and traditional approach

<b>Paired <math>t</math>-Test</b>						
Experiment 2	our_top1	Trad_top1	our_top3	Trad_top3	our_top5	Trad_top5
Mean	2.6696	2.5037	2.7089	2.3570	2.7333	2.3707
Variance	0.3999	0.5085	0.5383	0.4831	0.5699	0.5033
Observation	45	45	45	45	45	45
Degree of freedom		44		44		44
$t$		2.0598		7.4951		6.8736
$p(T \leq t)$ One-tailed		0.0227		0.0000		0.0000
Critical value: One-tailed		1.6802		1.6802		1.6802
$p(T \leq t)$ Two-tailed		0.0454		0.0000		0.0000
Critical value: Two-tailed		2.0154		2.0154		2.0154

#### 4.3. Summary of the Experiments

The results obtained in the experiments showed that the proposed approach significantly outperformed the traditional approach. This is because, in the traditional approach, the problem of semantic gaps between web pages and advertisements is not considered, whereas, in the proposed approach, WordNet::Similarity was used to mitigate the semantic gaps between the advertisement tags and web page terms. This provides the proposed approach with a high-quality performance, and also solves the problem of multimedia content in contextual advertising.

We also observed the performance of the proposed approach regarding the number of recommended advertisements by using various approaches. The top three recommended advertisements typically registered the more favorable satisfaction scores. Three advertisements may be the appropriate amount of information for users to absorb, not too much and not too little. Users may perceive one advertisement as insufficient information, but may feel overloaded by five advertisements.

## 5. Conclusions

Diverse formats exist for web advertising; advertisements can be text-based or multimedia-based. Because of this diversity, matching advertisements and web pages by using only the vector space model or keyword-based model is inadequate. Furthermore, the semantic gap problem between the two models renders them incompatible. This study resolved the semantic gap problem by using WordNet::Similarity and also combined the advantages of the vector space and keyword-based models to establish an efficient model. In the proposed approach, we used web pages and advertisements as vectors and tags, respectively. We propose a mapping method for computing the similarities between the advertisement tags and web page vectors. We used this approach to solve the problem of multimedia advertisements in contextual advertising. In addition, we propose a tag recommending method that involves recommending tags for advertisements containing text. In this recommending method, we used the PageRank algorithm with WordNet::Similarity for recommending keyword tags for advertisements.

In Experiments 1 and 2, the results showed that the proposed approach demonstrated a more favorable performance compared with the traditional approach. Statistical examinations were performed and the results

indicated that the advertisements recommended for web pages by using the proposed approach were significantly more favorable than those recommended using the traditional approach. Therefore, the proposed approach is an effective and satisfactory technique for matching web pages and advertisements.

For the experiments, we collected 30 articles from news web pages and used them as the testing data points. Each participant was required to answer 360 questions in the questionnaire (two experiments in total, each experiment involved 30 web pages for participants to read, each web page comprised six sets of proposed advertisements for participants to read). This might have exhausted the participants, and possibly induced them to provide careless responses to the questions. For remedying this problem, the entire evaluation process can be divided into several sessions interleaved with break times. Doing so can reduce the load on the participants and improve the validity of the experimental results.

In this research, we designed the advertisement algorithms by assuming that the web pages contain a high amount of text, advertisements contain no text, and advertisers can provide tags for advertisements. Furthermore, another limitation of this research is that the collected web pages should belong to a single topic because a hub page containing numerous topics and links—such as the Yahoo! homepage (<https://tw.yahoo.com/>) or Amazon online shopping homepage (<http://www.amazon.com/>)—is not the main target of the proposed method. These requirements restrict the scope of applicability of the proposed methods, thus constituting a limitation of this research. However, the limitations also indicate possible future research directions. First, because the idf values may be different when the data set is changed or when the context topic drifts, an alternative process is required for addressing this situation. A possible solution involves designing an automatic adjustment approach for consistently altering the idf values according to the changed data set or new context. This is a possible future research direction. Second, to reduce the annotation load of the advertisers, tag-recommendation algorithms can be developed for advertisements without text. A possible solution to this problem involves using the text of the hyperlinks linking to or linked by these advertisements. Third, we determined whether an advertisement is suitable for a web page according to only the contextual similarity. Other factors, such as user browsing history, personal preference, and user profile, can also be considered. Future studies can incorporate these factors into the proposed contextual advertisement model.

### Acknowledgement

The authors are grateful to the anonymous referees and associate editor for their helpful comments and valuable suggestions for improving an earlier version of this article. This study was partially funded by the National Science Council of Taiwan under Grant NSC 101-2410-H-008 -008 -MY3.

### REFERENCES

- Anagnostopoulous, A., A. Z. Broder, E. Gabrilovich, V. Josifovski, and L. Riedei, “Just-in-time Contextual Advertising,” *Proceedings of the 16th International Conference on Information and Knowledge Management*, 331–340, 2007.
- Anagnostopoulous, A., A. Z. Broder, E. Gabrilovich, V. Josifovski, and L. Riedei, “Web Page Summarization for Just-in-Time Contextual Advertising,” *ACM Transactions on Intelligent Systems and Technology*, Vol. 3, No. 1: 14:1–14:32, 2011.
- Athanasiadis, E. and S. Mitropoulos, “A Distributed Platform for Personalized Advertising in Digital Interactive TV Environments,” *Journal of Systems and Software*, Vol. 83:1453–1469, 2010.
- Banerjee, S. and T. Pedersen, “Extended Gloss Overlaps as a Measure of Semantic Relatedness,” *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, 805–810, 2003.
- Broder, A., M. Fontoura, V. Josifovski, and L. Riedel, “A Semantic Approach to Contextual Advertising,” *Proceedings of the 30th International Conference on Research and Development in Information Retrieval*, 559–566, 2007.
- Carrasco, J.J., D. C. Fain, K. J. Lang, and L. Zhukov, “Clustering of Bipartite Advertiser-Keyword Graph,” *International Conference on Data Mining Workshop*, 2003.
- Chowdhury, N., “A Survey of Search Advertising,” <http://www.mosharaf.com/wp-content/uploads/nmmkchow-webads-report-fall07.pdf> (Accessed on 2012/06/17).
- Ciaramita, M., V. Murdock, and V. Plachouras, “Semantic Associations for Contextual Advertising,” *Journal of Electronic Commerce Research*, Vol. 9, No. 1:1-15, 2008.
- Dave, K.S. and V. Varma, “Pattern Based Keyword Extraction for Contextual Advertising,” *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 1885-1888, 2010
- Fain, D.C. and J. O. Pedersen, “Sponsored Search: a Brief History,” *Bulletin of the American Society for Information Science and Technology*, Vol. 32, No.2:12-13, 2006.

- Fan, T.K. and C. H. Chang, "Blogger-Centric Contextual Advertising," *Expert Systems with Applications*, Vol. 38, No. 3:1777-1788, 2011.
- Hirst, G. and D. St-Onge, "Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms," In C. Fellbaum, editor, *Wordnet: an Electronic Lexical Database*, 305–332, 1988.
- IAB, 2011. Internet Advertising Revenue Report. <http://www.iab.net/media/file/IAB-HY-2011-Report-Final.pdf>.
- Jang, Y., T. Lee, K. Kim, W. Lee, D. Ann, and S. Chung, "Keyword Management System based on Ontology for Contextual Advertising," *Proceedings of the Sixth International Conference on Advanced Language Processing and Web Information Technology*, 440-445, 2007.
- Jansen, B. J. and S. Schuster, "Bidding on the Buying Funnel for Sponsored Search and Keyword Advertising," *Journal of Electronic Commerce Research*, Vol. 12, No. 1:1-18, 2011.
- Jiang J.J. and D.W. Conrath, "Semantic Similarity based on Corpus Statistics and Lexical Taxonomy," *Proceedings on International Conference on Research in Computational Linguistics*, 19–33, 1997.
- Khan, S., Q. M. Ilyas, and W. Anwar, "Contextual Advertising using Keyword Extraction through Collocation," *Proceedings of the 7th International Conference on Frontiers of Information Technology*, 69-69, 2009.
- Leacock, C. and M. Chodorow, "Combining Local Context and Wordnet Similarity for Word Sense Identification," In C. Fellbaum, editor, *Wordnet: an Electronic Lexical Database*, 265–283, 1998.
- Lee, J.H., J. Ha, J.Y. Jung, and S. Lee, "Semantic Contextual Advertising Based on the Open Directory Project," *ACM Transactions on the Web*, Vol. 7, No. 4:24:1–24:22, 2013.
- Li, H., D. Zhang, J. Hu, H. J. Zeng, and Z. Chen, "Finding Keyword from Online Broadcasting Content for Targeted Advertising," *Proceedings of the 1st International Workshop on Data Mining and Audience Intelligence for Advertising*, 55-62, 2007.
- Li, Y.M. and J. H. Jhang-Li, "Pricing Display Ads and Contextual Ads: Competition, Acquisition, and Investment," *Electronic Commerce Research and Applications*, Vol. 8:16-27, 2009.
- Lin, D., "An Information-Theoretic Definition of Similarity," *Proceedings of the Fifteenth International Conference on Machine Learning*, 296-304, 1998.
- Lin, F. H. and Y. F. Hung, "The Value of and Attitude Toward Sponsored Links for Internet Information Searchers," *Journal of Electronic Commerce Research*, Vol. 10, No. 4:235-251, 2009.
- Liu, J., C. Wang, Z. Liu, and W. Yao, "Advertising Keywords Extraction from Web Pages," *Proceedings of the International Conference on Web Information Systems and Mining*, 336-343, 2010.
- Liu, P., J. Azimi, and R. Zhang, "Automatic Keywords Generation for Contextual Advertising," *Proceedings of the companion publication of the 23rd international conference on World Wide Web companion*, 345-346, 2014.
- Mahdian, M. and K. Tomak, "Pay-per-action Model for Online Advertising," *Proceedings of the 1st International Workshop on Data Mining and Audience Intelligence for Advertising*, 1-6, 2007.
- Miller, G.A., R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to WordNet: An On-Line Lexical Database," *International Journal of Lexicography*, Vol. 3, No. 4:235-244, 1990.
- Mohanraja, V., M. Chandrasekaranb, J. Senthikumara, S. Arumugamc and Y. Suresha, "Ontology Driven Bee'S Foraging Approach Based Self Adaptive Online Recommendation System," *Journal of Systems and Software*, Vol. 85:2439–2450, 2012.
- Mu, E. and D. F. Galletta, "The Effects of the Meaningfulness of Salient Brand and Product-related Text and Graphics on Web site Recognition," *Journal of Electronic Commerce Research*, Vol. 8, No. 2:115-127, 2007.
- Nazerzadeh, H., A. Saberi, and R. Vohra, "Dynamic Cost- Per-Action Mechanisms and Applications to Online Advertising," *Proceedings of the 17th International Conference on World Wide Web*, 179-188, 2008.
- Niu, X., J. Ma, and D. Zhang, "A Survey of Contextual Advertising," *Proceedings of the 6th International Conference on Fuzzy Systems and Knowledge Discovery*, Vol. 7:505-509, 2009.
- Pak, A.N. and C. W. Chung, "A Wikipedia Matching Approach to Contextual Advertising," *World Wide Web*, Vol. 13, No. 3:251-274, 2010.
- Patwardhan, S., *Incorporating Dictionary and Corpus Information into a Context Vector Measure of Semantic Relatedness*. Master's Thesis, University of Minnesota, Duluth, 2003.
- Pedersen, T., S. Patwardhan, and J. Michelizzi, "Wordnet:: Similarity: Measuring the Relatedness of Concepts," *Proceedings of HLT-NAACL*, 38-41, 2004.
- Plessis, E. D., *The Advertised Mind*. Kogan Page, 2005.
- Resnik, P., "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 448–453, 1995.
- Ribeiro-Neto, B., M. Cristo, P. Golgher, and E. S. D. Moura, "Impedance Coupling in Content-Targeted Advertising," *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 496-503, 2005.

- Toutanova, K. and C. D. Manning, "Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger," *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 63-70, 2000.
- Toutanova, K., D. Klein, C. D. Manning, and Y. Singer, "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics*, 173-180, 2003.
- Turney, P.D., "Learning Algorithms for Keyphrase Extraction," *Information Retrieval*, Vol. 2, No. 4:303-336, 2000.
- Vargiu, E., A. Giuliani, and G. Armano, "Improving Contextual Advertising by Adopting Collaborative Filtering," *ACM Transactions on the Web*, Vol. 7, No. 3: 13:1-13:22, 2013.
- Wikipedia, 2012. PageRank. <http://en.wikipedia.org/wiki/PageRank>.
- Wu, Z. and M. Palmer, "Verb Semantics and Lexical Selection," *Proceedings of 32nd Annual Meeting of the Association for Computational Linguistics*, 133-138, 1994.
- Xiao, B., W. Yang, and J. Li, "Optimal Reserve Price for the Generalized Second-price Auction in Sponsored Search Advertising," *Journal of Electronic Commerce Research*, Vol. 10, No. 3:114-129, 2009.
- Yih, W., J. Goodman, and V.R. Carvalho, "Finding Advertising Keywords on Web Pages," *Proceedings of the 15th International Conference on World Wide Web*, 213-222, 2006.